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



塑料齿轮研究的进展和方向

石照耀　　辛　栋

3 | 2026
Vol.52

北京航空航天大学学报

第 52 卷　第 3 期　（总第 397 期）　2026 年 3 月

目　　次

塑料齿轮研究的进展和方向 …… 石照耀，辛栋（627）

融合地形参数的 CYGNSS 青藏高原土壤湿度反演 …… 张云，师丽云，杨树瑚，潘海燕，韩彦岭，洪中华（643）

基于双参数的 GNSS/SINS 故障检测及抗差自适应算法 …… 赵桂玲，王金宝，姜子昊，高帅（655）

增强上下文特征交互的实时无人机影像分割算法 …… 李云红，张富星，苏雪平，李丽敏，王梅，梁成名（668）

液压作动器差动连接下双唇级联靴形往复密封特性研究 …… 彭超，谭爽，叶强，苗建明，欧阳小平（678）

基于 R 藤 Copula-DBN 齿轮箱故障诊断 …… 王进花，刘正奇，曹洁，刘昀强，陈莉（687）

横向非均匀积水对飞机滑水偏出跑道行为影响 …… 李岳，张耀华，蔡靖，任怡楠（698）

一种无锚旋转框遥感图像船只目标检测算法 …… 李露，陈科研，刘辰阳，史振威（706）

基于 SCACGAN 的小样本齿轮箱故障诊断 …… 王进花，刘秦玮，曹洁，陈莉（713）

基于信息熵的客机倾斜姿态应急撤离评价 …… 程明，韦柯宇（724）

飞机复合材料结构载荷测量的温度修正方法 …… 李俊，李文龙，高腾龙，李亚南，刘敬礼（736）

反舰弹道导弹杀伤效能的不确定性研究 …… 王志辉，相志宁，高萍（744）

双连孔力传感器柔度建模与应变解析 …… 李立建，唐守乾，姚建涛，王迎佳（753）

多因素飞行冲突探测中的认知负荷评估方法 …… 杨越，马博凯，程龙（763）

矩形高层建筑顺风向脉动风荷载空间相关性 …… 袁家辉，陈水福，夏俞超，刘奕（772）

基于相似性加权的复杂系统费用测算模型 …… 常菊，刘晓东，何荧（784）

半倾转构型旋翼飞行器纵向操纵策略研究 …… 田洪源，鲁森魁，武上景（791）

基于小波变换的压力脉动信号分析方法 …… 覃子宇，韩猛，韩啸，赵硕，姜宇涵，林宇震（801）

涵道式冲压发电涡轮部件级性能模型建模和参数设计方法 …… 张华波，郭迎清，李桂财，赵万里，叶鹏（809）

一种定转速涡桨发动机的简化建模方法 …… 杨阳，戴树岭，雷小永（818）

非均匀加热下超临界燃油在并联双管中质量流量分配特性 …… 董昊，赵超凡，朱剑琴，程泽源（826）

基于 2D-FFT 与 2D-CFAR 的调频引信快速高精度信号处理方法 …… 张安琪，曹荣刚，周宇，李佳武，曹宇曦，于永滨（837）

强夯作用下夯沉量实用计算方法 …… 许国一，冯锦艳，OCHIENG J. OBONGO，于卓奇，童朝霞（846）

双置信度下多特征自适应融合的红外弱小目标实时跟踪 …… 陈家俊，李响，宋延嵩，董小娜（853）
超声速盘收口盘缝带伞不同收口比下气动性能 …… 代雨柔，李健，荣伟，薛晓鹏，黄明星，方康（864）
丙烯工质平板环路热管的稳态建模与特性分析 ……
…… 贾志超，毕涵礼，彭正，李国广，吴琪，张红星，苗建印（874）
生物航煤燃烧低压模化数值模拟 …… 何彦潼，邓甜，王伟（885）
红外遥感弱小船舶目标检测算法 …… 谢友晨，徐天阳，汤张泳，吴小俊（895）
基于近似计算技术的逻辑电路 HDL 级面积优化 …… 于宗源，王伦耀，储著飞，夏银水（908）
SMA 驱动多状态变后缘弯度机翼的设计、分析与实验 ……
…… 王子昂，卢志荣，李鸿濠，周文雅，王晓明（917）
MRAM 芯片单粒子功能中断机制研究 …… 刘羽寒，马英起，李世行，王菡滨，王碧（926）
融合邻域信息的民间文学知识图谱补全 …… 高方正，王笳辉，段亮，岳昆，李忠斌（935）
城市环境改进点云配准的 LiDAR 辅助 GNSS/IMU 定位算法 …… 黄明明，孙蕊（945）
小直径 C 形金属封严环多道次液压成形技术 …… 刘春雷，孟宝，朱宇，闫彬宇，万敏（955）
基于 FastSAM 辅助表示增强的自监督单目深度估计 …… 张冬冬，王春平，付强（964）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 346 * zh * P * ￥50.00 * 400 * 35 * 2026-03

（**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. 52 No. 3 (Sum 397) March 2026

CONTENTS

Plastic gear's research progress and direction
…… SHI Zhaoyao, XIN Dong (627)

Integrating topography parameters for soil moisture retrieval using CYGNSS on the Qinghai-Tibet Plateau
…… ZHANG Yun, SHI Liyun, YANG Shuhu, PAN Haiyan, HAN Yanling, HONG Zhonghua (643)

A GNSS/SINS fault detection and robust adaptive algorithm based on two parameters
…… ZHAO Guiling, WANG Jinbao, JIANG Zihao, GAO Shuai (655)

Real-time UAV image segmentation algorithm with enhanced contextual feature interaction
…… LI Yunhong, ZHANG Fuxing, SU Xueping, LI Limin, WANG Mei, LIANG Chengming (668)

Investigation into characteristics of double-lip cascade boot-shaped reciprocating seals in differential-connection working conditions of hydraulic actuators …… PENG Chao, TAN Shuang, YE Qiang, MIAO Jianming, OUYANG Xiaoping (678)

Gearbox fault diagnosis based on R-vine Copula-DBN
…… WANG Jinhua, LIU Zhengqi, CAO Jie, LIU Yunqiang, CHEN Li (687)

Influence on veer-off behavior of aircraft hydroplaning due to transverse uneven water accumulation
…… LI Yue, ZHANG Yaohua, CAI Jing, REN Yinan (698)

An anchor-free rotated box remote sensing image ship object detection algorithm
…… LI Lu, CHEN Keyan, LIU Chenyang, SHI Zhenwei (706)

Fault diagnosis of gearbox with small-sample based on SCACGAN
…… WANG Jinhua, LIU Qinwei, CAO Jie, CHEN Li (713)

Emergency evacuation assessment of aircraft tilting attitude based on information entropy
…… CHENG Ming, WEI Keyu (724)

Temperature correction method for load measurement of aircraft composite structures
…… LI Jun, LI Wenlong, GAO Tenglong, LI Yanan, LIU Jingli (736)

Research on uncertainty in kill effectiveness of anti-ship ballistic missiles
…… WANG Zhihui, XIANG Zhining, GAO Ping (744)

Compliance modeling and strain analysis of double-hole force sensor
…… LI Lijian, TANG Shouqian, YAO Jiantao, WANG Yingjia (753)

Cognitive load assessment method for multifactorial flight conflict detection
…… YANG Yue, MA Bokai, CHENG Long (763)

Spatial correlation of along-wind fluctuating wind loads on rectangular high-rise buildings
…… YUAN Jiahui, CHEN Shuifu, XIA Yuchao, LIU Yi (772)

Complex equipment cost estimation model based on similarity weight
…… CHANG Ju, LIU Xiaodong, HE Ying (784)

Research on longitudinal manipulation strategy of partial tilting aircraft
…… TIAN Hongyuan, LU Senkui, WU Shangjing (791)

Method for analyzing pressure fluctuation signals based on wavelet transform
…… QIN Ziyu, HAN Meng, HAN Xiao, ZHAO Shuo, JIANG Yuhan, LIN Yuzhen (801)

Modeling and parameter design methodology for component-level performance model of ducted ram air generation turbine
…… ZHANG Huabo, GUO Yingqing, LI Guicai, ZHAO Wanli, YE Peng (809)

A simplified modeling method for a constant speed turboprop engine
…… YANG Yang, DAI Shuling, LEI Xiaoyong (818)

Mass flow distribution characteristics of supercritical fuel oil in parallel tubes under non-uniform heating conditions
…… DONG Hao, ZHAO Chaofan, ZHU Jianqin, CHENG Zeyuan (826)

Fast and high precision signal processing method for frequency modulation fuze based on 2D-FFT and 2D-CFAR
…… ZHANG Anqi, CAO Ronggang, ZHOU Yu, LI Jiawu, CAO Yuxi, YU Yongbin (837)

Practical calculation method for settlement under dynamic compaction
…… XU Guoyi，FENG Jinyan，OCHIENG J. OBONGO，YU Zhuoqi，TONG Zhaoxia （846）

Real-time tracking of infrared dim-small target with multi-feature adaptive fusion under double confidence
…… CHEN Jiajun，LI Xiang，SONG Yansong，DONG Xiaona （853）

Aerodynamic characteristics of supersonic mid-gore reefing disk-gap-band parachute under different reefing ratios
…… DAI Yurou，LI Jian，RONG Wei，XUE Xiaopeng，HUANG Mingxing，FANG Kang （864）

Steady state modeling and characteristic analysis of propylene flat loop heat pipes
…… JIA Zhichao，BI Hanli，PENG Zheng，LI Guoguang，WU Qi，ZHANG Hongxing，MIAO Jianyin （874）

Numerical simulation of low-pressure modeling of bio-jet coal combustion
…… HE Yantong，DENG Tian，WANG Wei （885）

Detecting small ship targets via infrared remote sensing
…… XIE Youchen，XU Tianyang，TANG Zhangyong，WU Xiaojun （895）

Area optimization of logic circuits in HDL-level using approximate computing
…… YU Zongyuan，WANG Lunyao，CHU Zhufei，XIA Yinshui （908）

Design, analysis, and experimentation of SMA-driven multi-state variable camber wing
…… WANG Ziang，LU Zhirong，LI Honghao，ZHOU Wenya，WANG Xiaoming （917）

Research on single-event function interruption mechanism of MRAM chips
…… LIU Yuhan，MA Yingqi，LI Shixing，WANG Hanbin，WANG Bi （926）

Folklore knowledge graph completion by fusing neighborhood information
…… GAO Fangzheng，WANG Jiahui，DUAN Liang，YUE Kun，LI Zhongbin （935）

LiDAR aided GNSS/IMU positioning algorithm based on improved point cloud registration in urban environment
…… HUANG Mingming，SUN Rui （945）

Multi-steps hydroforming technology of small diameter C-shaped metallic seal ring
…… LIU Chunlei，MENG Bao，ZHU Yu，YAN Binyu，WAN Min （955）

FastSAM-assisted representation enhancement for self-supervised monocular depth estimation
…… ZHANG Dongdong，WANG Chunping，FU Qiang （964）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0788

塑料齿轮研究的进展和方向

石照耀[1,*]，辛栋[1,2]

(1. 北京工业大学 北京市精密测控技术与仪器工程技术研究中心，北京 100124；2. 深圳市兆威机电股份有限公司，深圳 518105)

摘 要：塑料齿轮以其独特的优势在智能家居、汽车等领域得到广泛应用，国内外研究人员对塑料齿轮的结构设计、材料特性及服役性能进行了大量研究，初步形成了塑料齿轮标准。为此，简要回顾了塑料齿轮的发展历程，介绍了塑料齿轮的特性和局限、齿形和结构的创新设计、新材料的改性和应用、成型工艺的优化、精度测量新方法、性能测试新技术等方面的研究进展，以及塑料齿轮标准化的情况，总结出塑料齿轮研究尚处于重要发展阶段，未来的研究方向在于通过设计创新和新材料应用提升其综合性能，开展更加系统全面的实验研究，准确理解服役行为，并评估其性能。

关 键 词：齿轮；塑料；注射成型；承载力；标准化
中图分类号：TH132.41
文献标志码：A **文章编号**：1001-5965(2026)03-0627-16

齿轮作为机械传动的基础零件，随着材料科学和制造工艺的进步而不断发展。自20世纪30年代，塑料开始在齿轮上应用，塑料齿轮以其独特的性能，得到越来越广泛的应用。塑料齿轮的应用向医疗仪器、汽车电子、航空航天等领域拓展，也对塑料齿轮性能的要求日益提高，带动了塑料齿轮的研究不断深入。受限于塑料材料本身的局限，“以塑代钢”仍需进一步提升塑料齿轮的性能。综合性能的改善是一个综合复杂的过程，需要设计、制造、测试等多方面的研究，新材料开发和改性这一传统的途径可以弥补塑料的一部分短板，结合应用场景的灵活设计也取得了一定成效。温度、压力、速度等成型工艺对塑料齿轮精度和性能的影响是十分复杂的，塑料齿轮失效机理和服役行为尚待揭示，这些问题的探究可以使人们更加清晰认识和深入理解塑料齿轮。

本文回顾了塑料齿轮的国内外研究进展和标准化情况，指明了塑料齿轮的研究方向，以期为相关研究提供借鉴和参考。

1 塑料齿轮概述

1.1 塑料齿轮的发展历史

古希腊的Antikythera装置是国外现存的最早的金属齿轮实物。在中国，早在公元前3世纪的战国时期就使用了铸铁齿轮。2021年，汉文帝霸陵考古中出土了青铜齿轮。无论国外还是国内，金属齿轮都是机械传动的典型代表。

20世纪30年代出现了塑料齿轮，20世纪60年代中国报道了利用FX-501纤维增强酚醛树脂制作小模数齿轮，1976年，浙江生产的水表计数器齿轮采用塑料材质解决了铜齿轮锈蚀问题，1994年，某仪表塑料齿轮厂就实现了1.5亿件塑料齿轮产量[1]。据国外机构报告，2019年，全球塑料齿轮市场规模已超230亿元人民币。塑料齿轮逐渐成为齿轮传动中重要的组成部分。

1.2 塑料齿轮的特性

塑料是有机高分子材料，由大量小分子经聚合反应形成链状或网状结构，与具有低分子量晶态结

收稿日期：2023-12-04；**录用日期**：2023-12-29；**网络出版时间**：2024-01-16 16:25
网络出版地址：link.cnki.net/urlid/11.2625.V.20240116.0925.001
基金项目：“塑料齿轮承载能力计算”国家标准计划(20221815-T-469)；深圳市技术攻关重点项目(SGG20220831104001002)
* **通信作者**. E-mail：shizhaoyao@bjut.edu.cn

引用格式：石照耀，辛栋. 塑料齿轮研究的进展和方向[J]. 北京航空航天大学学报，2026，52(3)：627-642.
SHI Z Y，XIN D. Plastic gear's research progress and direction[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：627-642 (in Chinese).

构的金属齿轮相比，塑料齿轮密度小、质量轻。塑料的微观作用力以分子间作用力为主，分子链长且柔顺性好，表现出较大的阻尼，齿轮啮合中的冲击可通过分子链变形转为内能，使塑料齿轮具有良好的吸振和低噪声特性。塑料齿轮的微观变形还会在接触面产生转化层，使摩擦系数和磨损速率降低，表现为自润滑和耐磨损的优点。宏观上看，塑料齿轮在负载时产生的局部变形使接触面积变大，接触应力相应降低。

塑料齿轮大量采用注射成型工艺，具有批量大、产品一致性好、制造成本低的优势，对于具有复杂结构的齿轮，其在简化工艺流程、节约材料方面的优势格外突出。例如，可将齿轮与金属轴嵌件一体成型；双联齿轮、人字形齿轮可在一套模具上通过一次成型制得；采用一模多腔的平衡充填技术（见图1），可将一个成型周期内的齿轮产品数量提升数倍，结合自动化上料、出模和分拣实现大规模的连续生产。

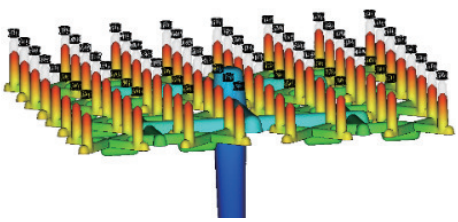


图1 平衡充填的一模64腔齿轮模具

Fig. 1 A gear mold with 64 balance-filled cavities

塑料材料普遍具有良好的耐化学磨蚀性能，不仅不会在潮湿空气中生锈，而且能够承受一定的化学腐蚀。塑料分子主要通过共价键连接，缺少自由移动的电子，是电和磁的不良导体，使塑料齿轮在数码、通信等产品的应用上具有优势。

1.3 塑料齿轮的局限

塑料齿轮的弹性模量较小，负载变形较大，使其应用局限在轻载工况。研究表明，塑料齿轮不适用于大负载的动力传动。塑料齿轮的极限许用应力通常在100 MPa以内，单位体积可传递的扭矩远低于金属齿轮。直接用相同尺寸的塑料齿轮代替金属齿轮在工程应用中通常是行不通的，单纯依靠增加尺寸来提高塑料齿轮的承载能力也是不可取的。小模数塑料齿轮传递较大扭矩时的变形量相比金属齿轮更大，对啮入啮出和传动平稳性的不利影响更加明显。

塑料齿轮的机械性能与温度具有显著的相关性。一方面，齿面摩擦生热和迟滞生热使啮合过程中齿面温度升高；另一方面，材料的弹性模量随温度的升高而降低。因此，评估塑料齿轮的承载能力必须考虑温度的影响。温度高、转速高、负载大等工况下，塑料齿轮齿面温度成为限制其承载能力的重要因素。塑料的热膨胀系数较大，对于小模数塑料齿轮来说，温度差异导致的尺寸变化相对于中心距的比例不容忽视，在温度变化大的环境下使用塑料齿轮时需仔细校核。

成型工艺对塑料齿轮的尺寸、精度和性能具有显著影响。图2展示了塑料齿轮成型中常见的问题[2-4]。塑料齿轮在注射成型过程中，由于厚度不均、温度不均和材料取向等因素导致其各部分的收缩不一致，进而产生缩腰、大小头等变形，影响塑料

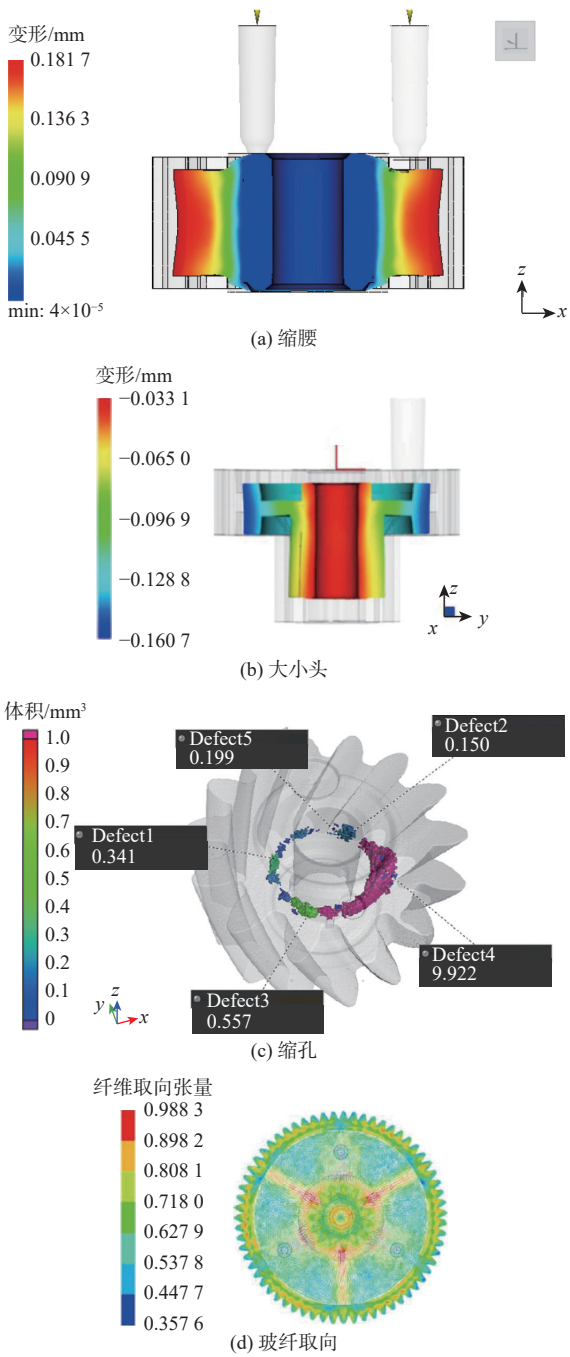


图2 塑料齿轮成型中存在的问题[2-4]

Fig. 2 Defects in plastic gear molding[2-4]

齿轮的制造精度, 限制了其在精密传动中的应用。POM 齿轮在成型中存在不同程度的缩孔现象, 借助工业 CT 可以检测到缩孔的大小和分布, 缩孔的存在会降低齿轮的承载能力。

塑料材料存在不同程度的各向异性, 特别是纤维增强材料沿纤维方向与垂直纤维方向的差异十分明显。在注射成型中, 材料沿流动方向的取向使塑料齿轮的尺寸和性能存在各向异性, 这一特性使塑料齿轮性能的表征和预测变得更加复杂。此外, 聚酰胺类材料的吸湿性会导致齿轮尺寸随环境中水分的变化而变化, 影响齿轮的尺寸稳定性。

2　塑料齿轮的创新设计

塑料齿轮的设计具有更大的自由度。参与啮合的齿面齿形、不参与啮合的齿根过渡曲线及齿轮本体的结构, 均可结合塑料齿轮的特点进行针对性的优化设计。

2.1　塑料齿轮的齿形设计

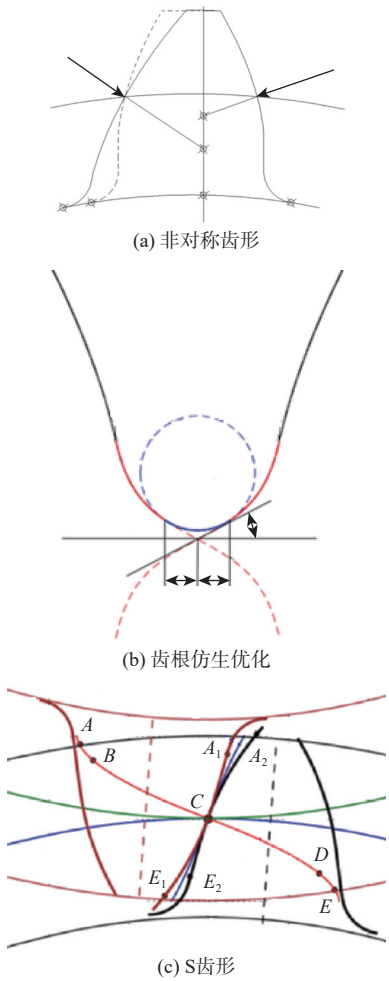
塑料齿轮的制造工艺不受传统展成加工原理的限制, 最常见的制造工艺是采用模具注射成型。模具型腔齿形采用线切割、电火花等方法加工, 可轻松实现非对称齿形、自定义齿根曲线和非渐开线齿形等设计, 如图 3 所示[5-9]。

左右齿面设计不同压力角的非对称齿形, 会大大增加金属齿轮的制造难度和成本, 但采用注塑成型时与对称齿形并无明显差异。大压力角齿形的径向分力相对较大, 轮齿的承载能力更大[5], 即使左右不对称的齿根圆角也可以减小 7% 的齿根应力[6]。2022 年, Kassem 等采用全圆弧、椭圆、仿生等不同方法对渐开线塑料齿轮齿根过渡曲线进行优化, 研究表明, 各方法的最优解对齿根承载能力均有提升效果, 且差异不大, 但有玻纤填充和无玻纤填充的塑料齿轮需采用不同方法才能达到最佳效果[7]。塑料齿轮负载变形会导致齿轮啮合长度增加, 使未修形塑料齿轮在啮入和啮出时存在冲击。2016 年, Duhovnik 等提出了 S 齿形, 通过应力分析和耐久试验等证实了在大负载下 S 齿形比渐开线齿形更有优势[8]。2022 年, Becker 等在交错轴传动中应用了 S 齿形, 分析结果显示, S 齿形与圆弧齿形的组合可提升承载能力 30%, 减小滑动距离 60%, 对减小磨损有改善效果[9]。

2.2　塑料齿轮的结构设计

针对塑料齿轮材料和工艺的特殊性, 通过合理优化轮缘、辐板、轴毂和加强筋的尺寸比例, 保证齿轮强度的同时实现轻量化设计。

由于金属与塑料在强度上存在巨大差异, 在设计与金属蜗杆配对的塑料齿轮时, 人为增加塑料齿轮齿厚并减薄金属蜗杆齿厚的方法是提高塑料齿轮承载能力的有效途径。2022 年, Oberle 等采用“以塑代钢”对汽车发动机旁的平衡轴齿轮进行减重和降噪, 塑料齿轮采用 2K 结构设计(见图 4(a)), 使用无玻纤塑料制造轮缘、玻纤增强塑料制造本体、金属材料制造轮毂, 无玻纤塑料的高阻尼实现了降噪, 玻纤增强塑料解决了尺寸稳定性、残余应力的问题[10]。2019 年, Fernandes 等在塑料齿轮中添加了金属嵌件(见图 4(b)), 通过优化金属嵌件的材料和形状来改善塑料齿轮的散热[11]。如图 4(c)所示, Düzcükoğlu 通过在塑料齿轮上添加气孔来改善空气流动和提高散热效率, 对温升有一定的改善效果, 但产生的应力集中问题不容忽视[12]。



(a) 非对称齿形

(b) 齿根仿生优化

(c) S齿形

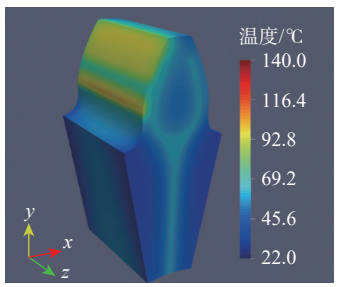
图 3　塑料齿轮的非标齿形[5-9]

Fig. 3　None standard profile of plastic gears[5-9]

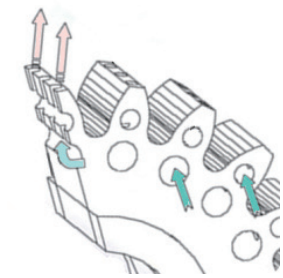
齿形设计创新可改善大负载下塑料齿轮的齿根和齿面应力, 散热设计可改善塑料齿轮的温升, 结构设计的创新和优化在弥补塑料齿轮性能上的局限可见一斑; 发挥塑料齿轮设计自由度高的优势, 通过设计创新将塑料低噪声、自润滑特点与金属高强度优势合理组合, 弥补塑料齿轮局限, 是塑料齿轮设计研究的重要方向。



(a) 2K结构



(b) 金属嵌件



(c) 散热孔

图 4 塑料齿轮的新结构[10-12]

Fig. 4 Creative structures of plastic gear[10-12]

3 塑料齿轮的材料研究

3.1 新材料的开发和应用

塑料齿轮的性能主要取决于材料特性。塑料材料的种类繁多，按结晶行为可分为非结晶和半结晶2类，按性能可分为通用塑料（聚丙烯（PP）、聚丙烯腈-丁二烯-苯乙烯（ABS）、聚乙烯（PE）等）、工程塑料（聚甲醛（POM）、聚酰胺（PA）、聚对苯二甲酸丁二酯（PBT）、聚碳酸酯（PC）等）和高性能工程塑料（聚酰亚胺（PI）、聚醚醚酮（PEEK）、聚苯硫醚（PPS）等），如图 5 所示，多数无填充塑料已进行了齿轮应用的尝试和研究[13-15]。受力学特性、耐温性的限制，PE 等通用塑料制造的齿轮主要应用于室温环境下的运动传递。目前，POM 和 PA 仍是应用和研究最多的塑料齿轮材料，其机械强度能够满足轻载工况的动力传递，可在 90℃ 下持续工作。随着汽车工业中塑料齿轮使用量的持续增长，对塑料齿轮耐温性和机械性能的要求进一步提高，PPS、PEEK 等高性能工程塑料的齿轮研究和应用逐年升温。

国外在新型工程塑料的应用上取得了较快的进展。在家用电器、食品机械、电动自行车等产品

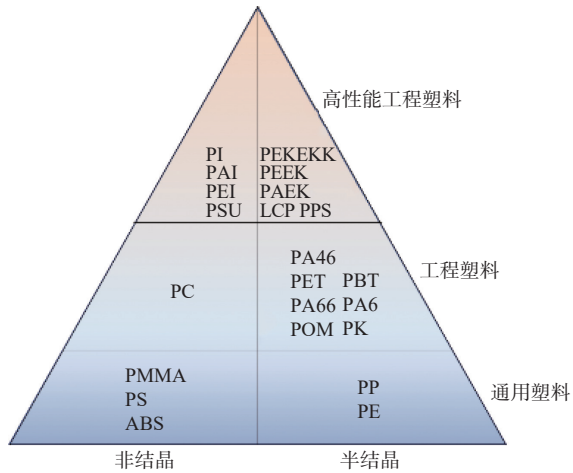


图 5 常用塑料材料

Fig. 5 Common gear plastics

均开展了塑料齿轮应用研究。Stahl 团队开展了 PEEK 齿轮代替金属齿轮的理论和实验研究，为塑料齿轮拓展城市轻量化电动汽车领域进行了有益尝试[16-18]。日本 Kurary 在汽车电子执行器上使用 PA9T 进行减重降噪的试验也取得了成功[19]。

3.2 塑料齿轮材料的改性

面对塑料与金属材质在性能上的巨大差异，为进一步提高塑料齿轮的性能，研究人员将焦点转向了塑料填充物的研究。

不同填充物、不同添加比例对塑料齿轮的性能具有不同的影响。玻璃纤维、碳纤维、矿物粉等无机物添加到塑料中，可不同程度地提高塑料齿轮的机械性能[20]。在小负载下，无玻纤塑料齿轮寿命较长，虽然含玻纤塑料齿轮的温度略高，但其耐温性更好，在中、大负载下，玻纤增强塑料齿轮寿命更有优势[21]。

石墨、二硫化钼（MoS_2）、聚四氟乙烯（PTFE）、碳纤维等添加剂具有减磨效果，适当含量可保持塑料齿轮机械性能的同时改善摩擦学性能，但其机理和最佳含量各不相同。碳纤维类似物理交联的刚硬支撑作用和磨损后的表面富积现象，是减小摩擦系数、提高耐磨性、改善热传递性的主要原因。PTFE 在齿面形成自润滑薄膜使摩擦系数变小，但受自身脆性的影响，过多的 PTFE 会导致冲击韧性降低。碳纤维的最佳含量为 20%，PTFE 的最佳含量为 10%[20]。20% 的石墨含量具有最佳的综合效果，此时拉伸强度、断裂伸长率最大，摩擦系数和磨损速率最小[22]。纳米碳酸钙填充可降低 PBT 齿轮的本体温度和磨损速率，6% 的含量可使寿命提升 63% 以上[23]。

将具有润滑效果的材料以涂层形式覆盖在塑料齿轮表面，对减少啮合过程中的摩擦生热和磨损量、提升塑料齿轮寿命具有一定的效果，其中，PTFE 涂层效果最佳[24]。类金刚石（DLC）涂层的厚度对

油膜厚度和压力的影响很小, 即使很薄的 DLC 涂层也可以提高塑料齿轮的散热, 并通过隔热来保护塑料齿轮[25]。但塑料齿轮本体的刚度也有重要影响, 有限元仿真研究发现, 2 μm 的涂层厚度受变形影响效果不佳[26]。固体润滑剂涂层会在应力作用下破碎并快速损失, 进而限制了减磨效果。

新型工程塑料可大幅扩大塑料齿轮的应用, 但研究需要较长的周期; 通过添加剂和表面涂层等改善塑料齿轮性能已取得一些成果, 但填充后强度和耐磨性的平衡、综合性能的提高、涂层厚度优化等仍待进一步研究。

4　塑料齿轮的成型工艺研究

注塑成型工艺作为一种成熟的批量制造工艺, 是塑料齿轮最主要的制造工艺, 奠定了塑料齿轮低成本制造的优势。塑料齿轮与普通塑料制品相比, 对尺寸和性能都有更高的要求。虽然塑料齿轮的性能从根本上是由塑料材料决定的, 但对于同一牌号的材料, 不同的成型工艺条件对塑料齿轮的性能具有重要影响。针对各工艺参数, 采用科学的实验或仿真方法, 研究其影响规律并进行工艺优化是塑料齿轮研究中的重要内容, 如图 6 所示。

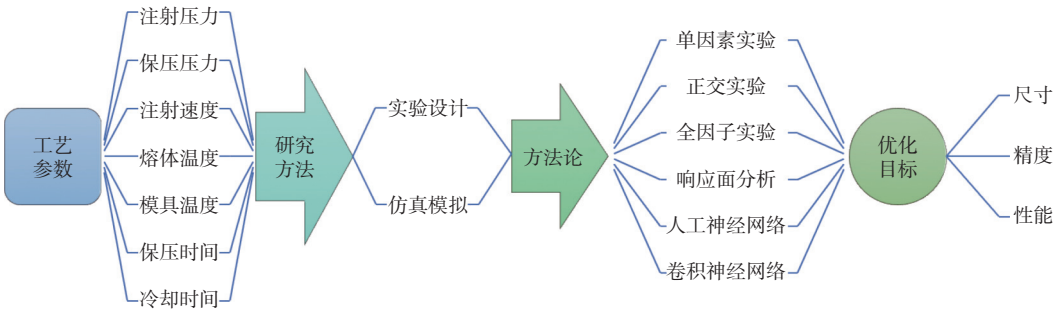


图 6　工艺参数优化的流程

Fig. 6　Flow chat of molding parameter optimization

4.1　工艺研究的实验设计

注塑成型工艺可大致分为注射+保压+冷却等 3 个基本的成型阶段, 各阶段对塑料齿轮品质影响的因素可分为温度、压力和时间 3 类, 包括注射阶段的注射压力、注射时间和 V/P 切换点, 保压阶段的保压压力、保压时间, 冷却阶段的冷却时间, 以及熔体温度和模具温度等。此外, 成型前备料阶段的烘料温度、烘料时间及出模后处理阶段退火温度等也会影响到塑料齿轮的成型。塑料材料的流变特性及成型过程中温度、压力在空间和时间上分布的不均匀性导致塑料齿轮在注塑成型过程中表现出明显的各向异性, 对塑料齿轮在啮合过程中的传动平稳性和长期服役时的可靠性造成不利影响。

合理设置注塑成型工艺参数以取得最佳的塑料齿轮精度和性能是一个求解非线性、多变量问题最优解的过程, 针对塑料齿轮的工艺优化, 需要开展大量实验研究。注塑工艺优化实验消耗大量材料、人力和时间, 借助实验设计方法, 通过有限的实验即可掌握各工艺参数取值对塑料齿轮的影响规律, 常用方法有单因素实验、正交实验、全因子实验等。

采用单因素实验方法可以对注塑中熔体温度、模具温度、注射速度和注射压力等单个参数的影响进行快速定性评价, 该方法是实践生产中应用最多的方法[27-28]。考虑到多个参数的共同影响, 正交实验方法是目前研究塑料齿轮注塑工艺中多因素作用的主要方法。姚文龙采用 6 因素 4 水平的正交实验, 对 PC 材料塑料齿轮的翘曲变形进行了研究, 取得了良好的效果[28]。注塑成型中, 温度、压力、速度等相互关联, 各参数间可能存在交互作用。例如, 提高注射速度使塑料熔体产生更大的剪切热, 进而使熔体温度升高。采用全因子的实验设计, 可以在确定单个工艺参数对塑料齿轮影响的同时, 考虑不同工艺参数的相互作用。周标等对塑料齿轮收缩变形的研究发现, POM 和超高分子量聚乙烯(UHMWPE)在注塑中模具温度和保压时间之间存在交互作用, 而 PA66 在注塑中注射压力和注射时间之间存在交互作用[29]。2022 年, 马文博等开展了 POM 齿轮缩孔改善的研究, 通过全因子实验发现, 保压压力与熔体温度之间不存在交互作用, 并通过分别优化 2 个参数使塑料齿轮缩孔占比显著降低[30]。

4.2　工艺研究的数值计算技术

充分利用计算机仿真技术可以模拟注塑成型过程中的温度、压力分布及材料的取向情况, 能够更好地理解注塑工艺。模流分析技术在塑料齿轮工艺优化中的应用节约了时间和经济成本, 通过实际注塑实验验证和反向修正方法可以取得良好的

成型效果，解决部分工艺参数优化无法解决的成型问题。在正交实验的基础上，2021年，武亚坤借助DSC分析发现了工艺参数对POM结晶度的影响规律，通过提高模具温度和注射速度使塑料齿轮弯曲强度得到提高[31]。沈亚强等借用MOLDFLOW软件分析了PA46齿轮在成型时的温度分布，发现中部与两端存在70℃的温度差异，考虑到工艺优化的局限性，采用反向修正的方法将齿轮精度从DIN 9级提高到DIN 5级，并采用相同方法解决了POM双联齿轮大小头的问题[3]。

采用神经网络技术建立注塑工艺参数与目标值之间的关联关系模型，进行工艺优化是国内外前沿的研究方法之一。徐铖和刘月云等分别借助人工神经网络技术对PA66塑料齿轮进行了研究，在较少的样本下完成了注塑工艺的优化[32-33]。

塑料齿轮的工艺参数优化应在借鉴同类研究的基础上结合具体产品和情况开展针对性分析，塑料齿轮的尺寸和形状不同、模具结构、注塑机、辅助设备等存在差异及受其他因素的影响，不能照搬现有研究的实验结果。计算机仿真CAE、神经网络、机器学习、人工智能等数值计算技术的研究对解决工艺优化问题具有积极作用，塑料齿轮工艺研究可在工艺过程大数据、AI大模型等方向开展研究。

5 塑料齿轮的精度测量研究

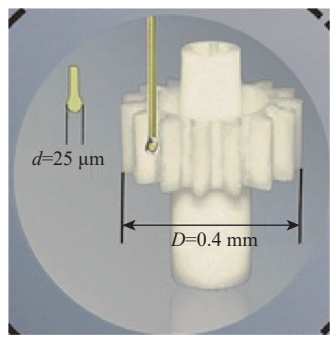
5.1 塑料齿轮测量的特点

塑料齿轮的测量具有精度低、速度快、测力小等特点。①塑料齿轮的精度低于常规金属齿轮。虽然研究中可制造出较高精度的塑料齿轮，受收缩不均、材料取向等影响，同一个产品不同轮齿位置、不同轴向位置的精度存在明显差异；同时，受模具结构、成型工艺等的影响，保持批量生产中精度的一致性较为困难，国内塑料齿轮批量制造的精度为7～11级，其中，直齿轮精度为7～9级，斜齿轮精度为9～11级。②塑料齿轮注射成型工艺的生产批量大，过程中变化因素多，对测量的速度要求高。不同批次的成型工艺、模具上不同模穴位置、不同齿号的位置等都影响精度测量结果，因此，要求在有限时间内进行大量塑料齿轮的检查。③塑料材料强度低、模数小、易变形，测量中对接触力敏感。测量力控制不当时，不仅影响测量的精度，还可能出现损坏探针的情况。

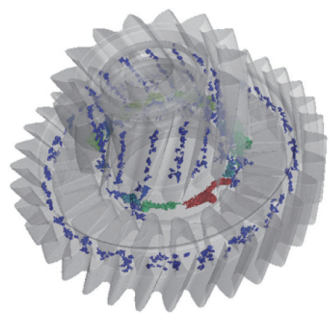
5.2 塑料齿轮的测量方法

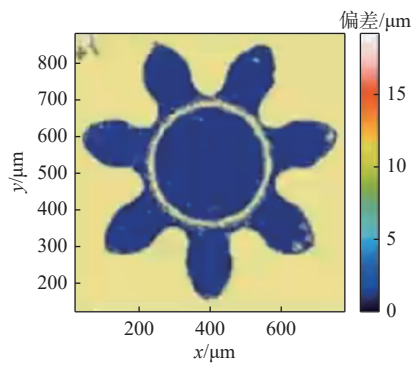
塑料齿轮常用的测量方法有双面啮合仪的综合测量、齿轮检测中心的分析式测量及在线检测的视觉测量。

综合测量的操作便捷、速度快、精度好，是塑料齿轮生产中大量采用的成熟测量方法，21世纪初国内就开展了相关的理论研究，并在塑料齿轮测量中进行应用。视觉测量具有更快的测量效率，可同时进行尺寸和误差的检查，也是应用较为广泛的测量方法。2007年起，国内就开始了计算机视觉在线测量的研究，包括照明系统、测量基准求解、图像处理和边缘求解算法等，近年来，YOLO网络、深度学习等新方法使视觉测量效果不断提升。2020年，周佳艺实现了0.005 mm的170个/min的快速视觉检测[34]。采用激光探针的齿轮检测中心对于提高分析式测量速度具有一定帮助。针对微小塑料齿轮的测量，国外研制了直径为0.025 mm的光纤探针(见图7(a))，测量了0.12 mm模数注塑齿轮的精度[35]，国内研究人员尝试了白光干涉法的非接触式测量[36]，如图7(c)所示。此外，图7(b)是CT技术扫描得到的塑料齿轮形貌，工业CT在检测塑料齿



(a) 光纤探针测量


(b) 工业CT扫描



(c) 白光干涉测量

图7 塑料齿轮测量的新方法

Fig. 7 New measurement methods for plastic gear

轮缩孔缺陷的同时, 也可用于齿轮精度的评价, 但其测量精度目前仅能达到 0.01 mm[37]。

GB/T 38192—2019 标准的实施使国内塑料齿轮测量和评价得到规范, 针对齿轮径向综合偏差测量时偏心和浇口问题、单面啮合综合测量的滤波和傅里叶变换问题、微小齿轮检测方法和偏差值等都提供了有益的参考和指导[38]。

综合测量和视觉测量日趋成熟, 分析式测量的研究仍待深入。4.2 节中反向修正方法是实现更高精度塑料齿轮制造的途径之一, 分析式测量如何快速、准确、全面地反映塑料齿轮的精度是该方法的基础, 也是塑料齿轮测量的重要课题。线激光、白光干涉、工业 CT 等已展现其优越性, 提高测量精度和实用性是非接触式测量的主要研究方向。微小齿轮测量和微小探针等需求逐步上升, 是国内学者值得关注的研究方向。

6　塑料齿轮的性能研究

早期的塑料齿轮主要应用于运动传递方面, 对几何形状和轮齿精度具有一定要求, 其负载很小, 对强度评估的需求不多。塑料齿轮的优势在于大批量和低成本, 大量的计算和实验验证势必造成开发周期延长和成本的上升。因此, 早期的塑料齿轮采用简易模具试制, 仅通过样机实验来验证是适宜的。

随着塑料齿轮承载能力提升和在动力传递应用方面的增长, 对塑料齿轮的强度进行计算, 评估其服役性能成为重要研究课题。以汽车工业为例, 塑料齿轮应用越来越多, 其失效带来的影响越来越大, 针对驻车制动系统等安全相关的塑料齿轮进行强度和寿命评估成为开发中必不可少的关键环节。

6.1　塑料齿轮的失效形式

塑料齿轮的失效形式与金属齿轮具有一定的相似性, 但也有其特殊性。塑料齿轮的失效形式主要有轮齿折断、齿面磨损、齿面点蚀、轮齿变形、齿面剪切及齿面熔融等。影响塑料齿轮失效的因素除工况、材料、润滑条件和环境条件等, 还包括成型工艺和填充剂等。

在开式传动中, 塑料齿轮的失效以磨损、疲劳折断为主; 在闭式传动中, 低速条件下失效与开式传动相似, 但中高速时以疲劳折断和齿面熔融为主[39]。

玻纤填充的塑料齿轮通常以磨损和断齿失效为主, 但在大负载工况下也会出现齿面熔融的热失效, 齿面温度的变化速度决定了齿轮失效形式, 当温升较慢时表现为齿面熔融, 当温升很快时表现为轮齿折断。2021 年, Hriberšek 等在 POM 和 PA 齿轮热行为和耐久实验中得到该结论, 为大负载下塑料齿轮失效形式的判定提供了参考[40]。陆松等对塑料齿轮热失效提出了添加填充物提升耐温性和强度性能、优化结构提高散热等改善措施[41]。

注塑工艺过程与塑料齿轮失效存在密切的关系, 刘光新通过注塑工艺优化改善了 POM 塑料齿轮强度和寿命[42]。

不同应用中塑料齿轮的缺陷和失效形式各不相同[43], 其失效机理也不尽相同。设计人员应对塑料齿轮的失效形式有明确的预期, 并进行针对性的评估和优化。

6.2　塑料齿轮的齿根弯曲应力

塑料齿轮在齿根弯曲应力作用下出现折断是常见的严重失效形式。将轮齿看作悬臂梁, 计算弯矩作用下齿根弯曲应力是塑料齿轮弯曲强度校核的基本方法。

现有研究中, 塑料齿轮的弯曲强度计算方法有: ①尼曼公式; ②VDI 2545: 1981 塑料齿轮弯曲强度计算方法[44]; ③VDI 2736-2: 2014 塑料齿轮弯曲强度校核方法[45]; ④宝理齿轮弯曲强度评估方法; ⑤JIS B 1759: 2019 塑料齿轮弯曲强度校核方法[46]。

VDI 2736-2: 2014 作为应用最广泛的塑料齿轮弯曲强度评估方法, 仍存在不足和修订的必要。2017 年, Hasl 等验证了 VDI 2736-2: 2014 在计算 POM 齿轮弯曲疲劳强度时的准确性, 提出了考虑负载变形对重合度影响的 ACORARS 修正公式, 采用 FZG 齿轮几何进行的台架实验证实了修正后计算公式的优越性[47]。

VDI 2736-2: 2014 与 JIS B 1759: 2019 是目前仅有的 2 个弯曲强度计算标准, 两者形式上不同且各有优劣。在塑料齿轮全球化设计与制造的背景下, 建立具有普遍共识的国际标准十分必要[48]。

6.3　塑料齿轮的齿面接触应力

点蚀是塑料齿轮的初期失效形式, 随着点蚀的发展, 可转化为齿面磨损或轮齿折断等其他失效形式, 使点蚀失效不易观察到。

由于塑料相比金属更软, 在齿面接触受力时塑料易变形, 导致接触面积更大, 接触应力较小, 不易表现出由齿面应力导致的失效。同时, 塑料材料的迟滞效果导致齿面发热较大, 塑料齿轮在连续工作时温升明显, 首先表现出热失效。只有 PEEK、PA 等部分材料与金属齿轮配对、润滑充分且齿面应力较高的条件下才表现出点蚀失效。

Düzcükoğlu 在研究 PA6 与钢齿轮的承载能力时, 通过扫描电子显微镜观察了齿面裂纹扩展后形成层状材料脱落的现象[12]。2020 年, Yu 等发现中

等负载条件下油润滑的PEEK齿轮齿面出现点蚀，如图8所示，初期，仅个别点蚀出现在齿顶区域，随着齿轮啮合次数的增多，点蚀从齿顶移到节线以下区域且点蚀数量随之增加，最终，形成齿面裂纹并导致塑料齿轮在节线位置断齿[49]。2022年，Stahl团队同样在油润滑PEEK-钢齿轮副的试验研究中观察到了齿面点蚀，以及点蚀扩展导致的齿面磨损和齿厚减薄[18]。

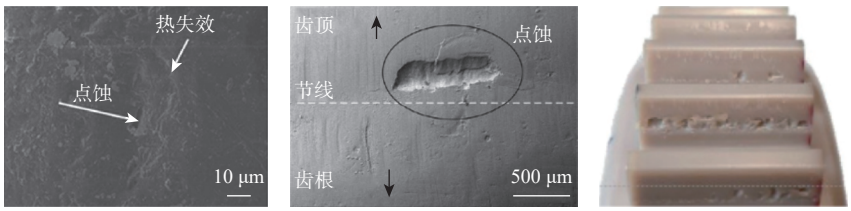


图8 PEEK齿轮齿面点蚀失效[12,18,49]

Fig. 8 Pitting failure of PEEK gear flank[12,18,49]

6.4 塑料齿轮的温升

1）塑料齿轮温升的重要性。塑料齿轮的尺寸、弹性模量、摩擦系数等都与温度具有相关性[13]，且许用应力随温度的变化十分明显。确定塑料齿轮的齿根温度是准确计算弯曲强度的前提，齿面局部温升是校核熔融失效的重要依据。塑料齿轮热膨胀系数是钢齿轮的10倍，热传导系数是后者的1/300，热量不能及时扩散导致温度持续升高，热膨胀还可能会导致侧隙不足进而发生双侧齿面接触，因此，塑料齿轮的温升研究比金属齿轮更加重要。

2）塑料齿轮的复杂性。塑料齿轮在啮合中存在摩擦生热和迟滞发热2种热源，使塑料齿轮的温升比金属齿轮更加复杂。考虑摩擦迟滞双热源及温度-模量耦合效应的情况下，采用恰当的有限元模型能够更准确地预测塑料齿轮的温升。1991年，陈其泰在塑料齿轮温升研究中采用解析数学模型对滞后温升进行计算[50]。随着有限元仿真方法的发展和应用，使塑料齿轮温升计算更加准确。2007年，陆松等采用黏弹性和超弹性相结合的本构模型模拟了塑料齿轮的迟滞现象[51]。2014年，郝一舒等采用非线性结构分析与热分析的联合仿真，发现考虑黏弹性后POM与钢齿轮的啮合温度比未考虑黏弹性时啮合温度高5%～10%[52]。2022年，余国达等指出赫兹理论在塑料齿轮接触和热分析中存在缺陷，基于黏弹性模型对POM齿轮进行有限元仿真，通过与干摩擦实验的稳态温升对比，发现仿真与实验的差异约5%，还发现模数较大时塑料齿轮的温升以摩擦为主、滞后生热轻微[53]。

3）影响塑料齿轮温升的因素。转速和转矩是影响塑料齿轮温升的重要因素，研究人员采用有限元仿真方法对两者的影响进行了研究。随着转矩和转速的增加，齿面温升呈非线性增大趋势[53]，其中，负载的影响大于转速的影响[54]。塑料齿轮转速升高，滑动速度相应升高，可采用压应力和相对滑动速度乘积PV计算其温升，而塑料的PV值可通过双圆柱摩擦实验取得[55]。低速下塑料齿轮PV值更高，即承载能力更大[56]，高转速下塑料齿轮承载能力相对较弱。塑料齿轮的温升还与模数、配料配对、材料填充物、润滑方式等其他因素有关。模数小的塑料齿轮，其齿面温度比模数大的齿轮低[57]。模数较小时，迟滞发热相对明显，模数较大时，迟滞热影响不大[14]。迟滞发热随无机填料比例增加而降低，填料改性可提高塑料齿轮热承载能力[15]。润滑对塑料齿轮温升的影响巨大，干摩擦下齿面温度高、波动大，而油润滑下温度低且平稳，即使负载提高7倍，油润滑下齿面温度仍低于干摩擦[58]。

6.5 塑料齿轮的磨损

磨损是塑料齿轮最常见的失效形式之一。

1）磨损试验形式和表征。考虑到塑料齿轮磨损的复杂和耗时，可通过简化试验和加载试验进行研究。简化试验形式包括销-盘、销-环、销-板、销-鼓及2个圆环等（见图9）；简化试验摩擦形式单一，齿面上曲率半径、转速、压力和滑移等都更加复杂，采用齿轮副进行磨损研究更接近实际情况[59]。齿轮副阶梯加载试验是一种快速试验方法。Mao等采用该方法将数周的恒定转矩疲劳试验缩短到4 h[60]，且结果与恒定转速转矩下的结果相当。

塑料齿轮磨损可以通过尺寸或质量等的变化

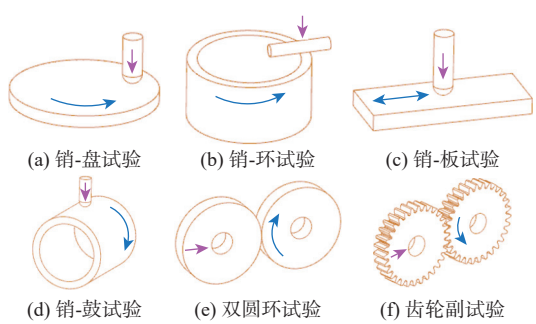


图9 简化摩擦试验形式

Fig. 9 Forms of simplified tribological test

进行表征。由于齿面的磨损是非均匀分布的, 测量塑料齿轮质量是一种常见且便捷的做法。

2) 影响磨损的因素和机理。塑料齿轮的磨损历程可分为初期跑合、稳定磨损、急剧磨损等 3 个阶段。影响塑料齿轮磨损的因素除转矩、转速、环境温度等工况条件外, 基体材料、材料配对、添加剂及齿轮安装误差等都有重要影响。

随着负载转矩的增加、转速的升高, 塑料齿轮的磨损速度加大, 磨损曲线的斜率增大[61], 该变化并不是线性的, 当达到临界转矩后, 磨损速度发生突变[62], 相同工况下, 不同材料的临界转矩从小到大依次为 POM<PA66<PEEK[63], 转速升高后, 临界转矩随之降低[64]。材料的不同组合表现出不同的磨损现象。对于 POM-PA、POM-PEEK 组合, 相同材料配对的磨损大于不同材料配对的磨损[62-63]。玻璃纤维常用来提高塑料齿轮强度, 其对塑料齿轮的磨损也有改善作用[61-62]。添加外部润滑对减小塑料齿轮磨损作用显著, PEEK-PEEK 齿轮副在有外部润滑时的承载能力比干摩擦时提高 40%[63]。

安装误差会导致塑料齿轮不能在理想的状态下啮合, 进而加速磨损。POM 齿轮在轴向误差、径向误差、面外倾斜和面内倾斜下表现出不同的磨损曲线、磨损速度、齿面磨痕及磨屑形貌[65](见图 10), 这些现象对于分析塑料齿轮的磨损失效具有指导意义。

Mao 团队将 POM 齿轮磨损与齿面瞬时温度相关联, 指出热失效是 POM 齿轮磨损速度在临界转矩时急剧升高的根本原因[61]。然而, 对于 PA、PEEK 磨损机理尚待研究[63]。

为了较好地估算塑料齿轮的性能, VDI 2736-2: 2014[45] 中提出了依次进行齿面温升、齿根应力、齿面应力、齿面磨损和轮齿变形计算的流程。大量实验研究发现, 不同失效间存在相互作用, 使得塑料齿轮的计算成为复杂的迭代过程。2021 年, Tavčar 等尝试开发了多准则函数和软件, 并通过计算机迭代计算达到最优的综合效果[66]。该方法虽有许多局限, 却是一次有益的尝试。

塑料齿轮的性能研究虽然取得了一些成果, 但仍处于早期阶段。塑料齿轮失效的机理及不同失效间的转换关系尚不清楚, 在不同尺度上开展微观形貌和结构的分析, 是深入理解塑料齿轮失效机理和不同失效间转换关系的重要途径。齿根应力、齿面应力温升和磨损等具体性能的评估目前在部分材料和部分工况下取得进展, 仍需要大量材料和试验的基础数据才能逐步建立起具有普遍性和准确性的性能计算方法。只有更全面准确地评估塑料齿轮的性能, 才能使塑料齿轮得到更恰当的应用。塑料齿轮性能的评估仍需将实验、计算、仿真等多种方法相结合进行进一步的深入研究。

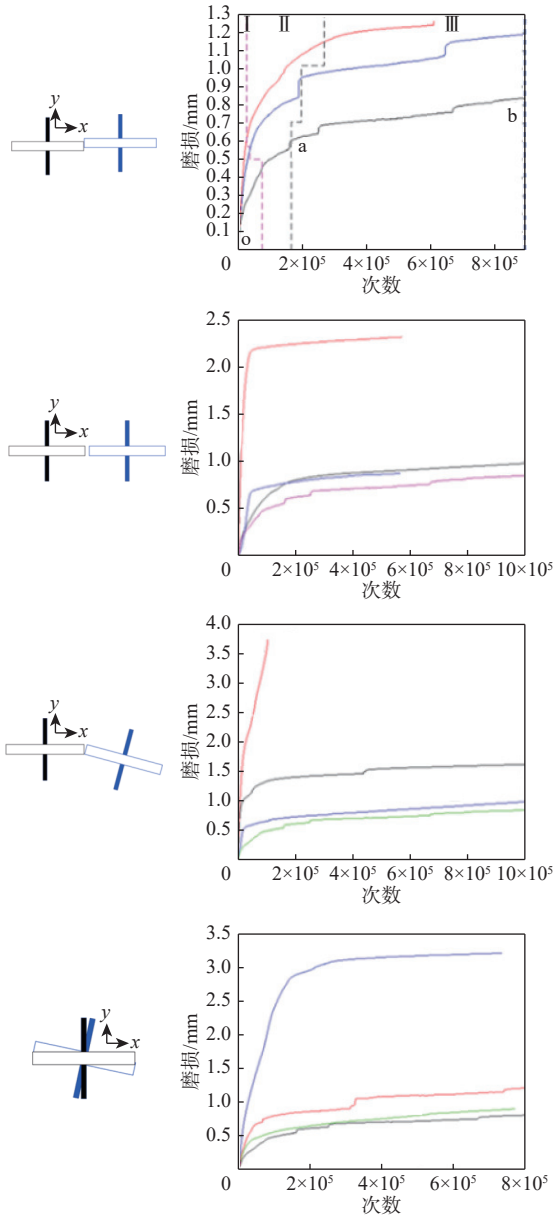


图 10　POM 齿轮不同安装误差下的磨损[65]

Fig. 10　Wear of POM gear with different alignment errors[65]

7　塑料齿轮试验台研究

塑料齿轮试验台是开展实验研究的前提, 国内研究相对有限。结合塑料齿轮试验特点深入研究, 才能开发更具实用性的试验台。

7.1　塑料齿轮试验台的特点

塑料齿轮与金属齿轮性能的差异使其试验设备具有如下特点: ①从应用领域和使用数量来看, 塑料齿轮以小模数齿轮为主, 集中在 0.2～2.0 之间, 其外形尺寸相对较小, 结构设计要求更加紧凑。②塑料齿轮仍主要用于轻载工况, 其传递的功

率远小于 1 kW，测试转矩小于金属齿轮试验台量程的 10%，采用金属齿轮试验台的测试结果准确性无法保证，需要更小的传感器。③塑料齿轮的性能随温度变化大，需要准确控制测试温度，金属齿轮常用的油温控制方法也不适用于塑料齿轮。④塑料齿轮的失效形式与金属齿轮不同，需要试验台进行针对性的监测方法和失效判定依据。

7.2 塑料齿轮试验台的布置

从加载方式上看，由于塑料齿轮传递功率小，采用开放式试验台，其功率消耗也不大，现有针对塑料齿轮研究的试验台较多采用功率消耗式，其加载装置包括磁粉制动器、伺服电机、液压装置和直流电机等。采用 FZG 试验台可进行部分塑料齿轮的试验[49]。

从空间上看，塑料齿轮试验台的常见布置有头对头和肩并肩 2 种，如图 11 所示。头对头布置中，测试齿轮位于中部，输入轴和输出轴分别位于齿轮两侧；该布置中，电机和加载装置可直接与轴相连，可在很小的中心距下进行试验。肩并肩的布置中，测试齿轮位于一端，输入和输出位于齿轮的同侧并排布置。肩并肩布置中，温度传感器等附件的安装和实时监测更加方便[67]。头对头布置中，2 个齿轮轴之间没有足够空间放置红外摄像机。采用抛光镜片将齿轮端面的温度分布情况反射到红外照相机中，即可观测到沿齿高方向上温度的分布情况[68]。

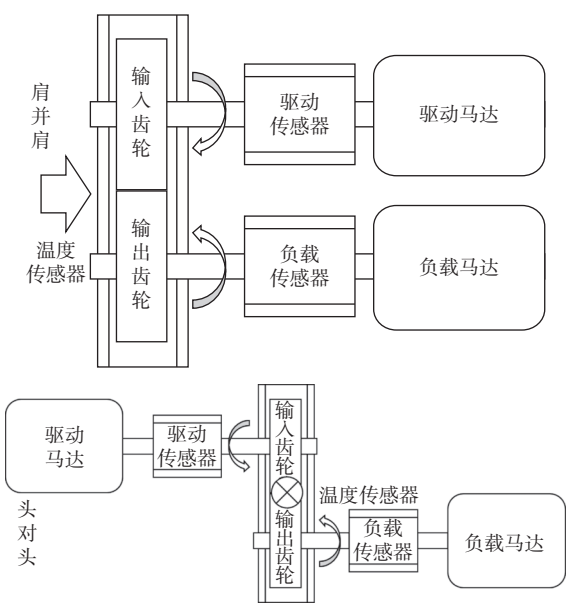
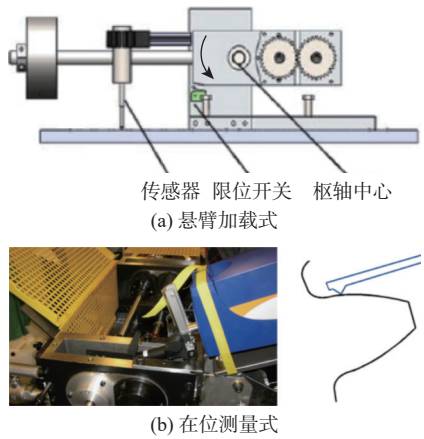


图 11 塑料齿轮试验台的 2 种布置

Fig. 11 Two type layouts of plastic gear tester

塑料齿轮试验过程中，磨损导致齿厚显著减薄，传统试验台无法维持恒定的试验负载转矩。Mao 等在机械封闭齿轮试验台上进行了塑料齿轮性能的试验，利用悬臂加载[64]（见图 12(a)），悬臂在齿轮磨损后可绕枢轴中心旋转，通过传感器测量旋转角度进而得到试验过程塑料齿轮的磨损量。Andersson 等在 FZG 的机械封闭式试验台上添加了轮廓度仪（见图 12(b)），通过对比磨损前后齿廓的变化实现了试验过程中在位的齿面磨损测量[69]。



(a) 悬臂加载式

(b) 在位测量式

图 12 塑料齿轮试验中不同齿厚测量方式[64,69]

Fig. 12 Different methods for plastic gear tooth thickness measurement during test[64,69]

7.3 塑料齿轮试验台的国内研究

国内针对塑料齿轮试验台的研究不足 10 年，相关研究有限。为满足塑料齿轮的恒定温度扭矩小转速高的工况特点及尺寸小的结构特点，塑料齿轮试验台在结构设计上更加紧凑，塑料齿轮的安装和连接是主要难点之一。

济南大学王砚军团队针对塑料齿轮快速磨损导致齿轮侧隙增大、扭矩剧烈波动的问题，研制了单支撑的伺服杠杆加载系统，在开展特种工程塑料齿轮摩擦磨损实验中确立了误差校正的方法[70]。

北京工业大学石照耀团队针对塑料齿轮的试验台进行了大量研究。2019 年，研制了塑料齿轮强度性能试验台[71]，采用立式结构提高了试样安装和实验操作的便利性，通过转矩-转角曲线实现了齿轮变形和断裂的识别。2021 年，基于恒定应力加速疲劳寿命测试原理研制了功率消耗式的塑料齿轮寿命试验机[72]（见图 13），提出了基于传动效率变化的



图 13 塑料齿轮加速疲劳寿命试验台[72]

Fig. 13 Tester of plastic gear accelerate fatigue test[72]

失效检测方法, 实现了塑料齿轮的准确快速测试。

塑料齿轮试验的小负载、小体积特点, 要求试验台在紧凑的空间内实现转速、转矩、温度等的准确监控。现有研究解决了塑料齿轮试验中温控加速、磨损对转矩影响等问题。轮齿变形监测、初期失效的识别等问题在塑料齿轮试验台研制中仍待解决。

8　塑料齿轮的标准化

从塑料首次被用来制造齿轮到现在, 塑料齿轮已经经历了 50 余年的发展, 然而尚未形成一个取得广泛共识的塑料齿轮国际标准。塑料齿轮的材料、制造工艺和应用领域仍在发生着持续的变化, 从事塑料齿轮制造、应用和研究的单位在不断增加, 产品检验、技术交流等对塑料齿轮的标准化提出了迫切的需求。

8.1　国外塑料齿轮的标准化进展

德国是形成塑料齿轮规范文件较早的国家。1981 年, 德国发布了 VDI 2545: 1981[44], 规定了塑料齿轮名词及参数、精度和偏差允许值、弯曲强度、齿面强度、变形量计算方法及结构设计建议等内容, 该文件于 1996 年废止。在此基础上发布了 VDI 2736 系列规范文件, 该系列规范分为 4 部分[45,73-75](见图 14), 这是目前取得最广泛认可和应用的塑料齿轮规范。

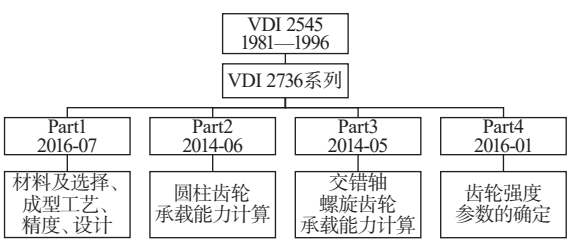


图 14　德国塑料齿轮的标准文件[45,73-75]

Fig. 14　Plastic gear standards in German[45,73-75]

英国的塑料齿轮标准 BS 6168 于 1982 年发布, 1987 年更新, 其中, 规定了塑料齿轮的术语和符号、基本齿条的齿形、模数、制造方法、精度等级、测量方法、齿轮偏差和尺寸公差、轮毂轮缘、热膨胀和吸湿膨胀及图面应标注信息等内容。该标准在 1987 年进行了补充[76], 增加了温升、弯曲强度和齿面强度等计算方法及材料极限应力的推荐数值。

美国涉及塑料齿轮的标准发布于 1997 年, 在 AGMA 917-B97 中包含了大量塑料材料的信息, 此后 20 多年, 美国塑料齿轮的标准体系逐步建立起来[77-82], 如图 15 所示。1997 年定义了塑料齿轮的齿形比例, 2001 年发布了材料标准并于 2015 年更新, 2006 年对技术要求做了规定, 对测量做了规定, 2021 年发布了测试方法标准。

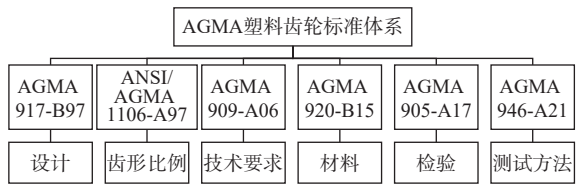


图 15　美国塑料齿轮的标准体系[77-82]

Fig. 15　Plastic gear standard system in USA[77-82]

日本的塑料齿轮标准于 2008 年发布, 将塑料齿轮精度作为 JIS B 1702 的第 3 部分[83]。塑料圆柱齿轮弯曲强度评估方法标准 JIS B 1759 于 2013 年发布, 2019 年更新, 该标准仅用于弯曲强度计算, 也未包含材料参数和极限应力值[46], 实用性欠佳。

纵观国外塑料齿轮的标准化历程, 其与塑料齿轮的行业发展具有密切联系。德国塑料齿轮标准是建立在 DSM、IMS、TUM 及大量汽车企业的基础上, 涵盖了材料厂、设备商、生产商、研究所和应用企业的完整产业链。美国塑料齿轮标准除了本国企业的参与, 还吸收了他国的经验, 形成了较为完善的标准体系。20 世纪 90 年代, 日本电子产品畅销全球, 塑料齿轮大量应用, 为塑料齿轮标准化奠定了良好的技术和产业基础。

8.2　中国塑料齿轮标准的进展

继德国、英国、美国、日本之后, 中国也制订了塑料齿轮国家标准。笔者牵头起草的中国塑料齿轮标准 GB/T 38192—2019 于 2019 年正式发布, 对塑料齿轮的精度做出了详细规定[38]。从图 16 可知, 中国塑料齿轮的标准化起步较晚, 但有后发优势, 该标准对接了最新 ISO 齿轮标准。紧随其后, 由笔者牵头制订的“塑料齿轮承载能力计算”国家

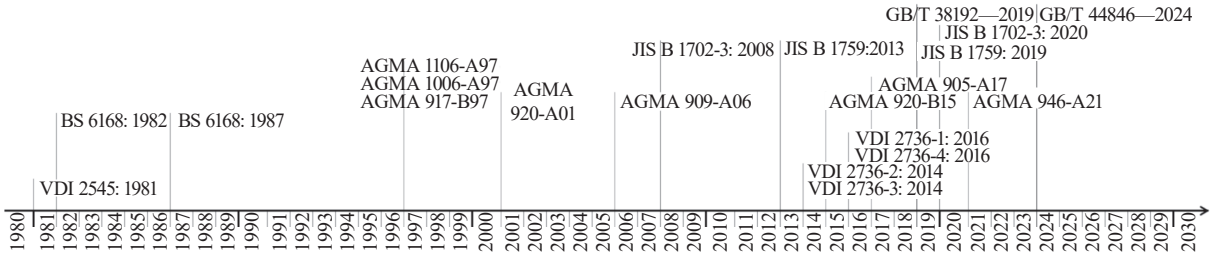


图 16　塑料齿轮标准的时间线

Fig. 16　Timeline of plastic gear standards

标准已于2024年发布实施[84]。

目前，中国塑料齿轮的标准化正迎来重要的发展机遇。①中国新能源汽车等领域快速发展，国际市场中显现出优势，使高附加值塑料齿轮不断增长，塑料齿轮的行业影响力也得到提升。②中国塑料齿轮在汽车零件全球化制造中具有成本优势，跨国合作开发给塑料齿轮标准化带来了机遇。③政策对制造业和双碳经济的重视，对塑料齿轮发展和标准化具有积极的作用。

8.3 塑料齿轮标准的内容

GB/T 38192—2019[38]仅解决了塑料齿轮精度评价的问题，塑料齿轮的标准化还应包括以下内容：①名词术语、符号等的规范和统一。塑料齿轮相关的专用名词、符号和量的规范化有利于塑料齿轮的技术交流，发挥工艺灵活性。②塑料齿轮强度评估方法的规范。正在制订的塑料齿轮承载能力计算标准，对提升塑料齿轮性能具有重要意义。③塑料齿轮测试试样、测试设备和测试方法的规范。④塑料材料和注塑工艺是塑料齿轮标准化中必不可少的部分。由于大量使用国外的原料和设备，此部分标准面临不小挑战。⑤塑料齿轮设计的指导。塑料齿轮特性、结构设计和模具设计的指导原则对提升技术水平具有积极作用，以技术指导文件的形式在标准体系中呈现是一个备选方案。

与金属齿轮悠久历史相比，塑料齿轮尚处于发展阶段。国外塑料齿轮标准化初具雏形，中国塑料齿轮的标准化工作已然起步，专有术语、强度计算、性能测试和设计原则等有待规范，标准化体系有待建立。产业发展和政策利好带来了重要机遇，有望使塑料齿轮标准化进一步发展。

9 结束语

塑料齿轮以其独特的性能得到了广泛应用，回顾塑料齿轮的发展历程和最新研究情况，可以得到以下结论：

1）塑料齿轮与金属齿轮相比，其发展历史短，具有制造优势，也有性能局限；但结构和齿形的创新设计、新型工程塑料的开发应用，使塑料齿轮的性能得到了提升。

2）塑料齿轮的精度和性能受制造工艺、轮齿温升等多因素影响，存在多种失效，其演变过程和机理虽部分探明，但温升、承载能力、磨损的定量研究尚存在不足。

3）中国塑料齿轮标准化相对国外起步晚，但面临发展机遇。

塑料齿轮研究尚处于重要的发展阶段，随着汽车电子、机器人等领域应用拓展，对塑料齿轮承载能力、可靠性等提出更高的需求。塑料齿轮进一步的研究包括：

1）通过新材料的应用和结构的创新设计提升塑料齿轮的综合性能。

2）基于大数据和AI技术探明成型工艺对塑料齿轮精度和性能的普遍影响规律。

3）完善试验台和测试规范，开展更加系统全面的实验研究，深入理解塑料齿轮的热、机械、摩擦行为。

4）快捷准确地预测塑料齿轮的失效形式和服役寿命。

参考文献（References）

[1] 欧阳志喜, 石照耀. 塑料齿轮设计与制造[M]. 北京: 化学工业出版社, 2011.
OUYANG Z X, SHI Z Y. Design and manufacturing of plastic gears[M]. Beijing: Chemical Industry Press, 2011(in Chinese).

[2] 沈亚强, 辛栋, 赵凯, 等. 塑料齿轮缩腰的机理分析及解决方法[J]. 模具工业, 2019, 45(2): 1-5.
SHEN Y Q, XIN D, ZHAO K, et al. Mechanism analysis and solution for middle shrinkage of plastic gear[J]. Die & Mould Industry, 2019, 45(2): 1-5(in Chinese).

[3] 沈亚强, 赵凯, 李岩, 等. 塑料直齿轮大小头形成机理分析及解决方法[J]. 模具工业, 2021, 47(2): 60-64.
SHEN Y Q, ZHAO K, LI Y, et al. Mechanism analysis and solution for taper plastic spur gear[J]. Die & Mould Industry, 2021, 47(2): 60-64(in Chinese).

[4] 沈亚强, 张杨, 赵凯, 等. 浇口对玻纤增强塑料齿轮精度及力学性能的影响[J]. 模具工业, 2021, 47(8): 46-49.
SHEN Y Q, ZHANG Y, ZHAO K, et al. Effect of gate design on precision and mechanical property of glass fiber-reinforced plastic gear[J]. Die & Mould Industry, 2021, 47(8): 46-49(in Chinese).

[5] MOHAN N A, SENTHILVELAN S. Preliminary bending fatigue performance evaluation of asymmetric composite gears[J]. Mechanism and Machine Theory, 2014, 78: 92-104.

[6] BIRK S, STOPPEL E, LABRIOLA V. Influence of plastic-metal-pairings on the lifetime of gears[C]//Proceedings of the International Conference on Gears. Munich: VDI Verlag, 2023: 1535-1546.

[7] KASSEM W, OEHLER M, KOCH O. Effect of tooth root fillet design on tooth root stress in short fiber reinforced plastic gears[M]. Alexandria: AGMA, 2022: 1-20.

[8] DUHOVNIK J, ZORKO D, SEDEJ L. The effect of the teeth profile shape on polymer gear pair properties[J]. Tehnickivjesnik-Technical Gazette, 2016, 23(1): 199-207.

[9] BECKER L, VILL D, TENBERGE P. Optimized crossed helical gearboxes using non-involute transverse sections[C]//Proceedings of the International Conference on Gears. Munich: VDI Verlag, 2022: 1197-1208.

[10] OBERLE S, LABRIOLA V, MELNIKOV E, et al. High performance plastic gear solution in the engine environment[C]//Proceedings of the International Conference on Gears. Munich: VDI Verlag,

2022: 1033-1044.

[11] FERNANDES C M C G, ROCHA D M P, MARTINS R C, et al. Hybrid polymer gear concepts to improve thermal behavior[J]. Journal of Tribology, 2019, 141(3): 032201.

[12] DÜZCÜKOĞLU H. Study on development of polyamide gears for improvement of load-carrying capacity[J]. Tribology International, 2009, 42(8): 1146-1153.

[13] 陈其泰, 李惠彬. 塑料齿轮的温度与材料性能之关系的探讨[J]. 机械设计, 1993, 10(4): 13-16.
CHEN Q T, LI H B. Research into the relation between the temperature of the plastic gear and the material properties[J]. Journal of Machine Design, 1993, 10(4): 13-16(in Chinese).

[14] 刘福林, 姜弘, 朱均. 干摩擦条件下聚甲醛塑料动力齿轮传动啮合效率的分析研究与计算[J]. 摩擦学学报, 1995, 15(3): 263-270.
LIU F L, JIANG H, ZHU J. Study and calculation of gearing efficiency for polyoxymethylene plastic-steel gear pair under dry frictional condition[J]. Tribology, 1995, 15(3): 263-270(in Chinese).

[15] 刘广建, 孙海峰. 工程塑料的粘弹性滞后能耗产生机理研究[J]. 塑料工业, 2003, 31(2): 39-41.
LIU G J, SUN H F. Study on occurring mechanism of viscous elasticity delayed energy consumption of engineering plastics[J]. China Plastics Industry, 2003, 31(2): 39-41(in Chinese).

[16] LAGIER F, BISCHOFBERGER A, BAUSE K, et al. Suitability of additive manufactured PEEK gears as a means of gear sampling for automotive high-speed gearboxes[C]//Proceedings of the International Conference on Gears. Munich: VDI Verlag, 2022: 1045-1060.

[17] REITSCHUSTER S, ILLENBERGER C M, TOBIE T, et al. Application of high performance polymer gears in light urban electric vehicle powertrains[C]//Proceedings of the 4th International Conference on High Performance Plastic Gears. Munich: VDI Verlag, 2022: 1181-1196.

[18] ILLENBERGER C M, TOBIE T, STAHL K. Operating behavior and performance of oil-lubricated plastic gears[C]//Proceedings of the 4th International Conference on High Performance Plastic Gears. Munich: VDI Verlag, 2022: 1117-1132.

[19] YANAGISAWA K, WEINMANN A. Weight and cost reduction potentials with high performance polymers for gears[C]//Proceedings of the 4th International Conference on High Performance Plastic Gears. Munich: VDI Verlag, 2022: 1019-1032.

[20] 陈战, 王家序, 秦大同. 塑料齿轮材料的改性研究[J]. 机械工程材料, 2003, 27(3): 42-44.
CHEN Z, WANG J X, QIN D T. Modification of plastic gear materials[J]. Materials for Mechanical Engineering, 2003, 27(3): 42-44(in Chinese).

[21] HRIBERŠEK M, KULOVEC S. Thermal and durability characterization of polyacetal and polyamide gear pairs[J]. Journal of Mechanical Science and Technology, 2021, 35(8): 3389-3394.

[22] KUMAR S S, KANAGARAJ G. Investigation on mechanical and tribological behaviors of PA6 and graphite-reinforced PA6 polymer composites[J]. Arabian Journal for Science and Engineering, 2016, 41(11): 4347-4357.

[23] SOUDMAND B H, SHELESH-NEZHAD K. Experimental investigation on the durability and failure modes of polybutylene terephthalate/calcium carbonate nanocomposite gears[J]. Engineering Failure Analysis, 2021, 120: 105113.

[24] DEARN K D, HOSKINS T J, PETROV D G, et al. Applications of dry film lubricants for polymer gears[J]. Wear, 2013, 298: 99-108.

[25] ZIEGLTRUM A, MAIER E, LOHNER T, et al. A numerical study on thermal elastohydrodynamic lubrication of coated polymers[J]. Tribology Letters, 2020, 68(2): 71.

[26] BAE S M, SEO K J, KIM D E. Effect of friction on the contact stress of a coated polymer gear[J]. Friction, 2020, 8(6): 1169-1177.

[27] 潘秀石, 陈堂敏. 基于 Pro/E 的塑料锥齿轮注塑成型工艺参数设计与优化[J]. 塑料, 2014, 43(5): 113-116.
PAN X S, CHEN T M. Optimization and design of plastic injection molding gear cone based on Pro/E[J]. Plastics, 2014, 43(5): 113-116 (in Chinese).

[28] 姚文龙. 基于 CAE 塑料齿轮注塑成型工艺参数优化研究[D]. 长沙: 中南林业科技大学, 2015.
YAO W L. Study on optimization of injection molding process parameters of plastic gear based on CAE[D]. Changsha: Central South University of Forestry & Technology, 2015(in Chinese).

[29] 周标, 邓小雷. 注塑工艺参数交互作用对塑料齿轮收缩变形的影响[J]. 轻工机械, 2008, 26(5): 23-27.
ZHOU B, DENG X L. Influence of the injection molding process parameters interaction on plastic gear shrinkage deformation[J]. Light Industry Machinery, 2008, 26(5): 23-27(in Chinese).

[30] 马文博, 曹伟, 沈亚强, 等. 抑制聚甲醛齿轮塑件缩孔的工艺参数研究[J]. 模具工业, 2022, 48(3): 42-46.
MA W B, CAO W, SHEN Y Q, et al. Research on processing parameters of suppressing voids in POM injection molding gears[J]. Die & Mould Industry, 2022, 48(3): 42-46(in Chinese).

[31] 武亚坤. 聚甲醛齿轮弯曲强度测试及其影响因素分析[D]. 郑州: 郑州大学, 2021.
WU Y K. Bending strength test of polyoxymethylene gear and analysis of its influencing factors[D]. Zhengzhou: Zhengzhou University, 2021(in Chinese).

[32] 徐铖, 程智虎, 陆金桂. 聚酰胺 66 齿轮注塑工艺参数优化设计方法研究[J]. 机械传动, 2018, 42(1): 69-72.
XU C, CHENG Z H, LU J G. Study on optimization design method of PA66 gear injection molding process parameter[J]. Journal of Mechanical Transmission, 2018, 42(1): 69-72(in Chinese).

[33] 刘月云, 刘碧俊. 基于 BP 神经网络及 PSO 算法的食品输送齿轮注塑工艺参数优化研究[J]. 塑料科技, 2018, 46(1): 80-84.
LIU Y Y, LIU B J. Research on optimization of injection molding process parameters of food transportation gear based on BP neural network and PSO algorithm[J]. Plastics Science and Technology, 2018, 46(1): 80-84(in Chinese).

[34] 周佳艺. 面向生产现场的注塑齿轮快速视觉检测技术研究[D]. 北京: 北京工业大学, 2020.
ZHOU J Y. Research on rapid visual inspection technology of injection-molded gear for production site[D]. Beijing: Beijing University of Technology, 2020(in Chinese).

[35] WOLFGANG K. When other measurement methods fail[J]. Mikroproduktion, 2015, 2: 16-19.

[36] 石照耀, 宋辉旭, 于渤. 基于白光干涉仪的微小齿轮齿距偏差和齿廓偏差高精度测量方法: CN110940283A[P]. 2020-03-31.
SHI Z Y, SONG H X, YU B. High precision measurement method of pitch deviation and profile deviation of micro gear based on white light interferometer: CN110940283A[P]. 2020-03-31(in Chi-

nese).

[37] 石照耀, 于渤, 宋辉旭, 等. 20年来齿轮测量技术的发展[J]. 中国机械工程, 2022, 33(9): 1009-1024.
SHI Z Y, YU B, SONG H X, et al. Development of gear measurement technology during last 20 years[J]. China Mechanical Engineering, 2022, 33(9): 1009-1024(in Chinese).

[38] 国家市场监督管理总局, 中国国家标准化管理委员会. 注射成型塑料圆柱齿轮精度制 轮齿同侧齿面偏差和径向综合偏差的定义和允许值: GB/T 38192—2019[S]. 北京: 中国标准出版社, 2019.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. Injection molding plastic cylindrical gears tolerance classification system—Definitions and allowable values of deviations relevant to flanks of gear teeth & double flank radial composite test: GB/T 38192—2019[S]. Beijing: Standards Press of China, 2019(in Chinese).

[39] 刘广建, 徐俊艳. 塑料齿轮失效形式及机理分析[J]. 工程塑料应用, 1996, 24(1): 31-33.
LIU G J, XU J Y. The invalidation form and the mechanism analysis of the plastic gears[J]. Engineering Plastics Application, 1996, 24(1): 31-33(in Chinese).

[40] HRIBERŠEK M, KULOVEC S. Study of the glass fibres and internal lubricants influence in a polyamide 66 matrix on the wear evolution of polyacetal and polyamide 66 based gears in a meshing process[J]. Engineering Failure Analysis, 2022, 134: 106071.

[41] 陆松, 孟惠荣. 塑料齿轮的热分析[J]. 塑料, 2003, 32(5): 58-60.
LU S, MENG H R. Thermal analysis of plastic gears[J]. Plastics, 2003, 32(5): 58-60(in Chinese).

[42] 刘光新. 影响塑料齿轮使用寿命的因素分析[J]. 常州信息职业技术学院学报, 2005, 4(3): 43-45.
LIU G X. Analysis on the factors influence plastic gear's service life[J]. Journal of Changzhou Vocational College of Information Technology, 2005, 4(3): 43-45(in Chinese).

[43] ZORKO D, KULOVEC S, DUHOVNIK J, et al. Durability and design parameters of a Steel/PEEK gear pair[J]. Mechanism and Machine Theory, 2019, 140: 825-846.

[44] VDI. Gear wheels made from thermoplastics: VDI 2545: 1981[S]. Dusseldorf: VDI-GME, 1981.

[45] VDI. Thermoplastic gear wheels-cylindrical gears-calculation of the load-carrying capacity: VDI 2736-2: 2014[S]. Dusseldorf: VDI-GPP, 2014.

[46] Japanese Standards Association. Estimation of tooth bending strength of cylindrical plastic gears: JIS B 1759: 2019[S]. Tokyo: Japanese Standards Association, 2019.

[47] HASL C, OSTER P, TOBIE T, et al. Bending strength of oil-lubricated cylindrical plastic gears[J]. Forschung Im Ingenieurwesen, 2017, 81(2): 349-355.

[48] BAE I, KISSLING U. Comparison of strength ratings of plastic gears by VDI 2736 and JIS B 1759[C]//Proceedings of the International Conference on Gears. Munich: VDI Verlag, 2019: 1267-1278.

[49] YU G D, LIU H J, MAO K, et al. An experimental investigation on the wear of lubricated steel against PEEK gears[J]. Journal of Tribology, 2020, 142(4): 041702.

[50] 陈其泰. 塑料齿轮齿面接触应力计算及其变位系数选择[J]. 中国矿业大学学报, 1991, 20(4): 35-44.
CHEN Q T. Calculation of contact stress and selection of addendum modification of plastic gears[J]. Journal of China University of Mining & Technology, 1991, 20(4): 35-44(in Chinese).

[51] 郑晓雯, 陆松, 孟惠荣. 基于黏弹性的超高分子量聚乙烯齿轮计算分析[J]. 中国矿业大学学报, 2007, 36(5): 641-646.
ZHENG X W, LU S, MENG H R. Computation and analysis of ultra high molecular weight polyethylene gear based on viscoelasticity[J]. Journal of China University of Mining & Technology, 2007, 36(5): 641-646(in Chinese).

[52] 郝一舒, 祝标, 赵星龙. 基于Marc对塑料-金属斜齿轮副黏弹性的有限元分析[J]. 机械传动, 2014, 38(1): 67-70.
HAO Y S, ZHU B, ZHAO X L. Finite element analysis of viscoelasticity of plastic-steel helical gear pair based on Marc[J]. Journal of Mechanical Transmission, 2014, 38(1): 67-70(in Chinese).

[53] 余国达, 刘怀举, 卢泽华, 等. 脂润滑条件下塑料齿轮稳态温度场仿真与试验研究[J]. 中国机械工程, 2022, 33(8): 890-898.
YU G D, LIU H J, LU Z H, et al. Simulation and experimental study of steady-state temperature field of plastic gears under grease lubrication conditions[J]. China Mechanical Engineering, 2022, 33(8): 890-898(in Chinese).

[54] 苗天峰, 王砚军, 赵中轩, 等. 高分子材料齿轮的热力耦合模型及接触界面瞬态温度场数值分析[J]. 机械传动, 2019, 43(2): 94-99.
MIAO T F, WANG Y J, ZHAO Z X, et al. Thermal-mechanical coupling model of polymer gear and numerical analysis of transient temperature field at its contact interface[J]. Journal of Mechanical Transmission, 2019, 43(2): 94-99(in Chinese).

[55] 黄文贵, 程中元. 轻纺机械塑料齿轮及其承载能力的计算[J]. 天津纺织工学院学报, 1988(Z1): 205-209.
HUANG W G, CHENG Z Y. Calculation of light textile mechanical plastic gear and its bearing capacity[J]. Journal of Tianjin Textile Engineering Institute, 1988(Z1): 205-209(in Chinese).

[56] 徐佩弦. 塑料齿轮抗弯疲劳强度与摩擦磨损[J]. 电子工艺技术, 1990, 11(3): 45-48.
XU P X. Defending fatigue strength and friction wear of plastic gear[J]. Electronics Process Technology, 1990, 11(3): 45-48(in Chinese).

[57] 叶丹. 尼龙塑料齿轮的强度设计[J]. 机械制造, 1991, 29(9): 6-7.
YE D. Strength design of nylon plastic gear[J]. Machinery, 1991, 29(9): 6-7(in Chinese).

[58] 卢泽华, 刘怀举, 朱才朝, 等. 润滑和载荷状态对聚甲醛齿轮服役性能的影响[J]. 中国机械工程, 2021, 32(17): 2047-2054.
LU Z H, LIU H J, ZHU C C, et al. Effects of lubrication and loading levels on POM gear durability performance[J]. China Mechanical Engineering, 2021, 32(17): 2047-2054(in Chinese).

[59] MERTENS A J, SENTHILVELAN S. Durability of polymer gear-paired with steel gear manufactured by wire cut electric discharge machining and hobbing[J]. International Journal of Precision Engineering and Manufacturing, 2016, 17(2): 181-188.

[60] MAO K, LANGLOIS P, HU Z, et al. The wear and thermal mechanical contact behaviour of machine cut polymer gears[J]. Wear, 2015, 332: 822-826.

[61] MAO K. A new approach for polymer composite gear design[J]. Wear, 2007, 262(3-4): 432-441.

[62] MAO K, LI W, HOOKE C J, et al. Friction and wear behaviour of acetal and nylon gears[J]. Wear, 2009, 267(1-4): 639-645.

[63] MAO K, LANGLOIS P, MADHAV N. A comparative study of

polymer gears made of five materials[J]. Gear Technology, 2019, 11: 68-72.

[64] HOOKE C J, MAO K, WALTON D, et al. Measurement and prediction of the surface temperature in polymer gears and its relationship to gear wear[J]. Journal of Tribology, 1993, 115(1): 119-124.

[65] HU Z D, MAO K. An investigation of misalignment effects on the performance of acetal gears[J]. Tribology International, 2017, 116: 394-402.

[66] TAVČAR J, ČERNE B, DUHOVNIK J, et al. A multicriteria function for polymer gear design optimization[J]. Journal of Computational Design and Engineering, 2021, 8(2): 581-599.

[67] HRIBERSEK M, ERJAVEC M, HLEBANJA G, et al. Durability testing and characterization of POM gears[J]. Engineering Failure Analysis, 2021, 124: 105377.

[68] LETZELTER E, GUINGAND M, DE VAUJANY J P, et al. A new experimental approach for measuring thermal behaviour in the case of nylon 6/6 cylindrical gears[J]. Polymer Testing, 2010, 29(8): 1041-1051.

[69] ANDERSSON M, SOSA M, OLOFSSON U. The effect of running-in on the efficiency of superfinished gears[J]. Tribology International, 2016, 93: 71-77.

[70] 张琪. 塑料齿轮磨损试验台加载系统设计及其特性分析[D]. 济南: 济南大学, 2017.
ZHANG Q. Design and characteristic analysis of the loading system of plastic gear wear test bench[D]. Jinan: University of Jinan, 2017(in Chinese).

[71] 郝雷, 于渤, 石照耀, 等. 微小型行星减速器强度性能试验台的研制[J]. 工具技术, 2019, 53(12): 87-91.
HAO L, YU B, SHI Z Y, et al. Development of atrength performance test bench for micro planetary reducer[J]. Tool Engineering, 2019, 53(12): 87-91(in Chinese).

[72] 石照耀, 王伟, 于渤, 等. 塑料齿轮加速疲劳寿命试验机研制及试验研究[J]. 机械传动, 2021, 45(12): 130-135.
SHI Z Y, WANG W, YU B, et al. Design and experiment research of accelerated fatigue life tester for plastic gear[J]. Journal of Mechanical Transmission, 2021, 45(12): 130-135(in Chinese).

[73] VDI. Thermoplastic gear wheels-materials, material selection, production methods, production tolerances, form design: VDI 2736-1: 2016[S]. Dusseldorf: VDI-GPP, 2016.

[74] VDI. Thermoplastic gear wheels-crossed helical gears-mating cylindrical worm with helical gear-calculation of the load-carrying capacity: VDI 2736-3: 2014[S]. Dusseldorf: VDI-GPP, 2014.

[75] VDI. Thermoplastic gear wheels-Determination of strength parameters on gears: VDI 2736-4: 2016[S]. Dusseldorf: VDI-GPP, 2016.

[76] British Standards Institution. Specification for non-metallic spur gears: BS 6168: 1987[S]. London: British Standards Institution, 1987.

[77] American Gear Manufacturers Association. American gear manufacturers association. design manual for parallel shaft fine-pitch gearing: AGMA 917-B97[S]. Alexandria: American Gear Manufacturers Association, 1997.

[78] American Gear Manufacturers Association. Tooth proportions for plastic gears: ANSI/AGMA 1106-A97[S]. Alexandria: American Gear Manufacturers Association, 1997.

[79] American Gear Manufacturers Association. Materials for plastic gears: AGMA 920-B15[S]. Alexandria: American Gear Manufacturers Association, 2015.

[80] American Gear Manufacturers Association. Specifications for molded plastic gears: AGMA 909-A06[S]. Alexandria: American Gear Manufacturers Association, 2006.

[81] American Gear Manufacturers Association. Inspection of molded plastic gears: AGMA 905-A17[S]. Alexandria: American Gear Manufacturers Association, 2017.

[82] American Gear Manufacturers Association. Test methods for plastic gears: AGMA 946-A21[S]. Alexandria: American Gear Manufacturers Association, 2021.

[83] Japanese Standards Association. Cylindrical gears-the accuracy. Part 3: definitions and allowable values of deviations relevant to corresponding flanks of the gear teeth and radial composite deviations of injection molding plastic gears: JIS B 1702-3: 2008[S]. Tokyo: Japanese Standards Association, 2008.

[84] 国家市场监督管理总局,国家标准化管理委员会. 塑料齿轮承载能力计算: GB/T 44846—2024[S]. 北京: 中国标准出版社, 2024.
State Administration for Market Regulation, National Standardization Administration. Calculation of load capacity for plastic gear: GB/T 44846—2024[S]. Beijing: Standards Press of China, 2024(in Chinese).

Plastic gear's research progress and direction

SHI Zhaoyao[1, *], XIN Dong[1, 2]

(1. Beijing Engineering Research Center of Precision Measurement Technology and Instruments, Beijing University of Technology, Beijing 100124, China;
2. Shenzhen Zhaowei Machinery & Electronics Co., Ltd., Shenzhen 518105, China)

Abstract: Plastic gears have many unique advantages, and have been widely used in smart household, automobile and various fields, domestic and foreign researchers have carried out many studies on the structural design, material characteristics and performances of plastic gears, and formed some standards. A brief overview of the history of plastic gears was given, and the most recent advancements in plastic gear research were presented. These included the pros and cons of plastic gears, creative profile and structure designs, material modifications and applications, injection molding process optimization, new measurement techniques, performance testing and evaluation, and the creation of plastic gear standardization. It's concluded that plastic gear is at an important stage of development, comprehensive performance improvement by new structure design and material applications well as carrying out more deep experimental researches to accurately grasp their behavior and performance are the directions of future plastic gear researches.

Keywords: gear; plastic; injection molding; load capacity; standardization

Received: 2023-12-04; **Accepted**: 2023-12-29; **Published Online**: 2024-01-16 16:25

URL: link.cnki.net/urlid/11.2625.V.20240116.0925.001

Foundation items: "Calculation of Load Capacity for Plastic Gear" National Standard Plan (20221815-T-469); Shenzhen Technical Key Project (SGG20220831104001002)

* **Corresponding author**. E-mail: shizhaoyao@bjut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0789

融合地形参数的CYGNSS青藏高原土壤湿度反演

张云[1,2]，师丽云[1]，杨树瑚[1,2,*]，潘海燕[1,2]，韩彦岭[1,2]，洪中华[1,2]

(1. 上海海洋大学 信息学院，上海 201306；2. 上海海洋大学 上海海洋智能信息与导航遥感工程技术研究中心，上海 201306)

摘 要：青藏高原的土壤湿度是影响全球大气环流和气候变化的重要因素，旋风全球导航卫星系统(CYGNSS)使用全球导航卫星系统反射信号技术(GNSS-R)，为监测青藏高原的土壤湿度提供了新的手段，但高原复杂的地形环境使得CYGNSS的反射率难以被直接用于土壤湿度的反演。基于此，提出了融合修正后CYGNSS反射率、CYGNSS入射角、地形参数（海拔、坡度和地表粗糙度）等5个特征参数的星载GNSS-R土壤湿度机器学习反演模型。对CYGNSS反射率进行发射功率的系统误差修正、地表植被和地表粗糙度衰减误差修正；将修正后反射率等5个参数作为输入特征量，SMAP土壤湿度作为验证数据，使用2020年解冻期（6—9月）数据按5∶5随机分为训练集和验证集，分别建立随机森林(RF)和人工神经网络(ANN)青藏高原土壤湿度反演模型，以2021年解冻期的数据作为测试集，考察模型的泛化能力。随机森林模型的结果优于人工神经网络模型，测试集上反演结果的均方根误差(RMSE)为0.058 6 cm^3/cm^3，皮尔逊相关系数为0.703 3，模型具有较好的泛化性能，且反演得到的青藏高原土壤湿度的空间变化和降水的空间变化趋势相吻合。将CYGNSS土壤湿度与那曲实测土壤湿度进行对比，均方根误差为0.070 cm^3/cm^3，具有较高的精度。研究结果表明，融合修正后CYGNSS反射率、CYGNSS入射角和地形参数建立的反演模型能够较为准确地反演青藏高原大范围的土壤湿度。

关 键 词：土壤湿度；青藏高原；旋风全球导航卫星系统；反射率修正；机器学习

中图分类号：P715.6

文献标志码：A **文章编号**：1001-5965(2026)03-0643-12

青藏高原面积和海拔是全球之最[1]，土壤湿度在青藏高原的地-气能量和水循环过程中起着重要作用，影响着全球大气环流和气候变化[2-3]，但青藏高原地理环境极端、气候条件恶劣，只能进行小空间尺度和短时间的现场测量，难以满足青藏高原气候变化研究的需要[4]。近年来，学者们利用遥感数据对青藏高原地区进行土壤湿度研究。2017年，杨婷等[4]采用神经网络模型建立基于中分辨率成像光谱仪(moderate-resolution imaging spectroradiometer, MODIS)植被指数、先进微波扫描辐射计-地球观测系统(advanced microwave scanning radiometer-Earth observing system, AMSR-E)微波亮温数据和辅助气象数据的土壤湿度反演模型，对青藏高原地区那曲、玛曲和阿里3个观测网进行土壤湿度反演，与这3个站点实测数据相比，均方根误差(root mean square error, RMSE)为0.031～0.041 cm^3/cm^3。2020年，程渊[5]采用集成学习模型，通过MODIS产品的光谱指数、植被覆盖度、叶面积指数和地表温度等参数，对2016年4—7月青藏高原地区3个观测网(那曲、玛曲和阿里)进行土壤湿度反演，均方根误差为0.041 cm^3/cm^3。

全球导航卫星系统反射信号技术(global navigation satellite system reflectometry, GNSS-R)通过接收GNSS卫星的反射信号对地面进行探测，可以实现

收稿日期：2023-12-04；录用日期：2024-02-23；网络出版时间：2024-03-21 08:38
网络出版地址：link.cnki.net/urlid/11.2625.V.20240320.0902.001
基金项目：国家自然科学基金(42176175,42271335)；国家重点研发计划(2019YFD0900805)
*通信作者. E-mail：shyang@shou.edu.cn

引用格式：张云，师丽云，杨树瑚，等. 融合地形参数的CYGNSS青藏高原土壤湿度反演[J]. 北京航空航天大学学报，2026，52(3)：643-654.
ZHANG Y，SHI L Y，YANG S H，et al. Integrating topography parameters for soil moisture retrieval using CYGNSS on the Qinghai-Tibet Plateau[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：643-654 (in Chinese).

远程感知[6]、土壤湿度[7]、海面风速[8]、海面高度[9]等地球物理信息。GNSS-R接收机的质量和功耗都较小，有利于小卫星组网观测，且星载GNSS-R不受云雾干扰[10]，可以用低成本的方式对全球进行高时空分辨率的连续观测，为长时间的土壤湿度监测提供了新的机遇。目前，基于星载GNSS-R的土壤湿度研究使用的数据大都来自2014年萨里卫星技术有限公司发射的TechDemoSat-1卫星（TDS-1）和2016年美国国家航空航天局（National Aeronautics and Space Administration, NASA）发射的旋风全球导航卫星系统（cyclone global navigation satellite system, CYGNSS）。2016年，Chew等[11]利用TDS-1数据验证了星载GNSS-R土壤湿度反演的可行性；2019年，Clarizia等[12]利用CYGNSS数据计算得到反射率，结合SMAP项目提供的辅助数据，建立反射率-植被-粗糙度多元线性回归算法，在全球范围内的反演结果与SMAP的土壤湿度产品的均方根误差为0.07 cm^3/cm^3；万玮等[13]使用2017—2019年的CYGNSS数据进行两步标定修正反射率误差，反演得到全球的土壤湿度与SMAP土壤湿度的皮尔逊相关系数从0.46提升到了0.74；2022年，张云等[7]综合多种误差修正模型，对2020年10月至2021年9月中国东部地区的CYGNSS数据进行反演，均方根误差为0.026 4 cm^3/cm^3。随着机器学习算法的不断成熟，其逐渐被运用在土壤湿度反演领域。2020年，Senyurek等[14]提出了使用随机森林（random forest, RF）、人工神经网络（artificial neural network, ANN）和支持向量机（support vector machine, SVM）3种机器学习算法，结合2017—2019年的CYGNSS数据、MODIS的归一化植被指数（normalized difference vegetation index, NDVI）、植被含水量（vegetation water content, VWC）、全球数字土壤制图系统SoilGrids的土壤质地数据和国际土壤水分网络ISMN实测站点的土壤湿度数据建立回归模型，3种算法的均方根误差分别为0.052, 0.061, 0.065 cm^3/cm^3。

TDS-1能够覆盖全球，但其只有1颗卫星，数据量少且目前已经停用；CYGNSS有8颗卫星，数据量多，其遥感的纬度范围为±38°，可以覆盖青藏高原大部分地区；同时，2017年12月之后，CYGNSS开始接收海拔大于600 m的反射率信号[15]，为星载GNSS-R青藏高原土壤湿度反演研究提供了可能。青藏高原地形问题导致反射率无法直接被用来反演土壤湿度。2021年，胡羽丰等[16]通过反射率结合CYGNSS星载数据中的经纬度和时间、SMAP的植被归一化指数、地表粗糙度和土地覆盖类型等参数，使用ANN模型克服了青藏高原地形对CYGNSS的影响，实现了对青藏高原的川藏交通廊道沿线（经度：90°E～105°E，纬度：28°N～33°N）土壤湿度的反演，与SMAP土壤湿度相比，外推结果的偏差为0.010 cm^3/cm^3，均方根误差为0.034 cm^3/cm^3，皮尔逊相关系数为0.743。目前，基于星载GNSS-R的青藏高原土壤湿度的研究较为缺乏，且使用机器学习模型时需要输入大量的辅助参数，影响了模型的建立和反演速度。

本文以青藏高原（经度：82°E～102°E，纬度：28°N～36°N）为实验区域，青藏高原解冻期（6—9月）为实验的时间范围，提出修正后的CYGNSS反射率融合青藏高原地形参数的星载GNSS-R土壤湿度机器学习反演模型。首先，对CYGNSS反射率进行发射功率的系统误差修正、地表植被和地表粗糙度对信号的衰减误差修正；然后，在修正后的反射率、入射角的基础上仅加入海拔、坡度和地表粗糙度等3个地形相关参数作为输入特征量，以SMAP土壤湿度为验证数据，将青藏高原2020年解冻期的数据集随机5∶5划分为训练集和验证集，分别建立RF和ANN青藏高原土壤湿度反演模型并进行性能评估，以2021年解冻期的青藏高原数据验证模型在青藏高原大范围内的泛化性和鲁棒性；最后，与那曲土壤湿度观测网实测数据对比，验证CYGNSS青藏高原土壤湿度的可靠性。研究结果表明，本文提出的基于星载GNSS-R的少参数机器学习反演模型能够对青藏高原大范围的土壤湿度进行长时序反演。

1 数据来源

1.1 CYGNSS数据

CYGNSS星座由NASA于2016年12月发射，其由8颗卫星组成，轨道高度约510 km。每颗卫星都搭载了空间GNSS遥感接收机仪器（space GNSS receiver remote sensing instrument, SGR-ReSI）接收地表的反射信号。本文使用CYGNSS的2.1版本的Level 1数据，以NetCDF格式存储了时延多普勒图（delay Doppler map, DDM）、入射角、天线增益、接收机位置等相关信息（可通过 https://podaac.jpl.nasa.gov 下载），使用的数据时间范围为2020年6月1日至9月30日和2021年6月1日至9月30日，经度范围为82°E～102°E、纬度范围为28°N～36°N。

1.2 GTOPO30数据

GTOPO30是全球数字高程模型数据，涵盖全球范围（下载地址为 http://earthexplorer.usgs.gov/）。该数据空间分辨率为30″，海拔值为−407～8 752 m，海洋区域被屏蔽为“无数据”，并被指定为−9 999[17]。

1.3　SMAP 数据

2015 年，NASA 发射的 SMAP 卫星主要用于观测全球土壤湿度，该卫星同时搭载了主动雷达和被动辐射计 2 种传感器。研究结果表明[18]，SMAP 的土壤湿度产品与实测土壤湿度结果更为接近，精度较好。本文使用 2020 年 6 月 1 日至 9 月 31 日和 2021 年 6 月 1 日至 9 月 31 日的 SMAP L3 级土壤湿度产品作为土壤湿度的真值，其空间分辨率为 36 km×36 km、时间分辨率为 1 d（下载地址为 https://nsidc.org/data）。该产品包含的特征量有土壤湿度、植被含水量、地表粗糙度等，其中，土壤湿度作为验证数据，反映了 0～5 cm 土壤平均含水量，植被含水量和地表粗糙度作为反射率修正模型的输入参数，地表粗糙度同时又作为机器学习输入特征。

1.4　那曲土壤湿度观测网数据

那曲土壤温湿度观测网（经度：91.5°E～92.5°E，纬度：31°N～32°N）位于青藏高原中部半湿润区 100 km×100 km 的空间范围，平均海拔 4 650 m。该网络部署区域的地形平坦，生物量低，是评价卫星土壤湿度反演较为理想的区域。该观测网部署了 57 个观测站点，每个站点有一个传感器倾斜插入 0～5 cm 的表层土中，这与 L 波段无源微波传感器（CYGNSS 和 SMAP）的穿透深度相当，其他 3 个传感器分别以 10 cm、20 cm 和 40 cm 的深度水平插入，测量时间间隔为 30 min。该观测网提供 3 个空间尺度的土壤湿度数据：1°×1°（对应全球气候模型（global climate model, GCM）网格尺度）、0.3°×0.3°（对应被动微波卫星象元尺度）、0.1°×0.1°（对应主被动融合微波卫星象元尺度）[19]。本文选用观测深度为 5 cm、空间观测尺度为 0.3°×0.3°的实测数据。

2　反演原理与方法

2.1　反演基本原理

全球定位系统（global positioning system, GPS）发射的信号到达地球表面，被土壤表面反射，土壤的介电性能、粗糙度及植被等性质决定了反射率的大小，而土壤湿度又是影响土壤介电性能的重要因素，从而可以通过地表反射率来实现土壤湿度的反演。在土壤表面没有地形起伏且粗糙度很低，同时植被覆盖不厚的情况下，可以基于双基地散射雷达公式计算陆地表面反射率。本文采用的反射率计算式为[20]

$$\Gamma = \frac{P_r (4\pi)^2 (R_{ts} + R_{sr})^2}{\lambda^2 G^r P^t G^t} \tag{1}$$

式中：Γ为反射信号的地面反射率；P_r为 DDM 的峰值；R_{sr}为 CYGNSS 卫星到镜面反射点之间的距离；R_{ts}为 GPS 卫星到镜面反射点之间的距离；G^r为 CYGNSS 的天线增益；P^t为 GPS 信号的发射功率；G^t为 GPS 的天线增益；λ为 GPS 的 L1 信号波长（0.19 m）。这些参数在 CYGNSS 的数据中均有提供，计算时将参数的单位均转换为 dB。

2.2　反射率修正

实际的观测环境较为复杂，反射率会受到环境的影响偏离真实值，因而需要对其进行修正。综合文献中的误差修正模型，本文对反射率进行了 3 种修正：GPS 发射功率的系统误差修正[21]、地表植被[22]和地表粗糙度[23]对信号的衰减误差修正。

2.2.1　GPS 发射功率系统误差修正

2020 年，Chew 等[21]通过对土壤湿度变化极小且植被覆盖极少的撒哈拉地区的研究发现，GPS 卫星的发射功率存在伪随机噪声（pseudo random noise, PRN）码相关的误差，可通过撒哈拉地区进行土地定标，得到基于 PRN 码分类的发射功率误差经验值，利用误差经验值对 GPS 卫星的发射功率进行修正，从而修正反射率。本文通过 2020 年和 2021 年 CYGNSS 全年的撒哈拉数据获得了发射功率误差经验值，并将误差经验值应用于研究区域，通过修正公式（2）得到修正发射功率误差后的反射率Γ_1：

$$\Gamma_1 = \Gamma - B_p \tag{2}$$

式中：Γ_1为经过发射功率误差修正后的反射率；B_p为发射功率误差经验值。

2.2.2　地表植被衰减误差修正

2001 年，Bindlish 和 Barros[22]提出了参数化水云模型修正地表反射率，消除植被含水量对地表反射率的影响。水云模型修正公式如下：

$$\Gamma_2 = \left(\frac{\Gamma_1^2 - R_{veg}}{\gamma^2} \right)^2 \tag{3}$$

$$\gamma = \exp(-BW_v \sec(\theta)) \tag{4}$$

$$R_{veg} = AW_v(1-\gamma^2)\cos\theta \tag{5}$$

式中：Γ_2为经过植被衰减误差修正后的反射率；γ为植被透射率；R_{veg}为由植被引起的散射系数；W_v为 SMAP 提供的植被含水量，kg/m^2；θ为 CYGNSS 提供的入射角；A 为植被密度相关的参数，B 为与双向传播路径相关的参数，本文采用所有土地都适用的取值，即 A=0.001 2，B=0.091[21]。

2.2.3　地表粗糙度衰减误差修正

1979 年，Choudhury 等[23]研究证明了地表粗糙度是影响相关功率的环境因素特征量之一，其造成的衰减可以表示为

$$S_r = 4k^2h^2\cos\theta \tag{6}$$

式中：k为波数；h为SMAP提供的地表粗糙度。

经过地表粗糙度衰减误差修正后的反射率为

$$\Gamma_3=\Gamma_2\exp(S_r) \tag{7}$$

式中：Γ_3为经过粗糙度衰减误差修正后的反射率，是用于建立机器学习模型的参量。

图1为2020年6月1日随机500个反射率修正前后的对比。可以看出，修正前后反射率变化较明显。

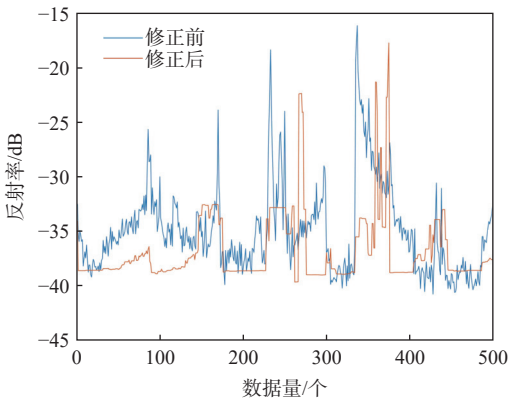


图1 修正前后的反射率对比

Fig. 1 Comparison of reflectivity before and after correction

2.3 特征量

本文采用2类共5个特征量建立了机器学习青藏高原土壤湿度反演模型：①CYGNSS特征量，包括修正后的CYGNSS反射率和入射角等2个特征量；②地形特征量，包括海拔、坡度和地表粗糙度等3个特征量。

2.3.1 CYGNSS特征量

修正后的反射率Γ_3是CYGNSS对地球不断变化的土壤湿度和表面信息的接收结果，是土壤湿度反演的关键特征量。

GPS信号入射角θ是入射信号与镜面反射点处法线的夹角，介电常数、入射角和反射率存在函数关系，而土壤湿度影响介电常数，因此，入射角与土壤湿度存在函数关系。卫星高度角、介电常数和反射率的关系如下[10]：

$$\Gamma=\frac{(\varepsilon-1)^2(\sin\theta_1)^2\left[\varepsilon-(\cos\theta_1)^2\right]}{\left[\sin\theta_1+\sqrt{\varepsilon-(\cos\theta_1)^2}\right]^2\left[\varepsilon\sin\theta_1+\sqrt{\varepsilon-(\cos\theta_1)^2}\right]^2} \tag{8}$$

式中：θ_1为GPS卫星的高度角，θ_1=90°−θ；ε为相对介电常数。

相对介电常数ε与土壤湿度m_v的关系如下[10]：

$$\varepsilon=3.03+9.3m_v+146.0m_v^2-76.7m_v^2 \tag{9}$$

式中：m_v为土壤湿度，cm^3/cm^3。

利用式(8)和式(9)可以得到不同土壤湿度下反射率和入射角的关系，如图2所示。可以看出，在不同的土壤湿度条件下，随着入射角的不断增大，反射率均不断减小，因而选择入射角θ作为反演模型的特征量。

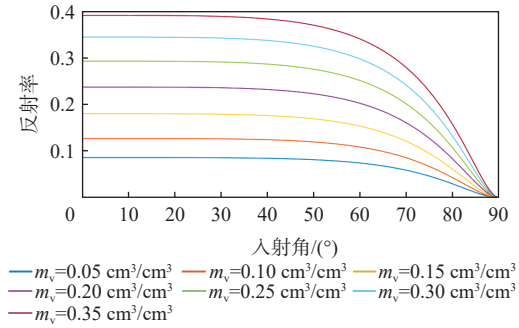


图2 入射角与反射率的关系

Fig. 2 Relationship between incident angle and reflectivity

2.3.2 地形特征量

当反射面相对平滑时，反射信号DDM功率[24]可表示为

$$\langle|Y_r(b,f)|^2\rangle=|\Lambda(b)S(f)|^2**|\sigma(b,f)|^2 \tag{10}$$

式中：Λ为PRN码的自相关函数；S为GPS和CYGNSS相对运动产生的多普勒模糊函数；“**”表示对延迟和多普勒二维卷积；σ为双基地散射系数；b和f分别为时间和多普勒频率。

当反射面粗糙时，反射信号DDM功率[24]可表示为

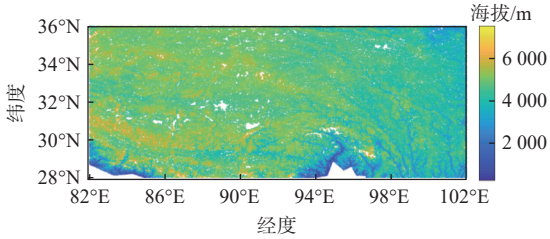
$$\langle|Y_{r,\mathrm{topo}}(b,f)|^2\rangle=\langle|Y_r(b,f)|^2\rangle**D(b,f) \tag{11}$$

式中：D为散射点的平均密度。

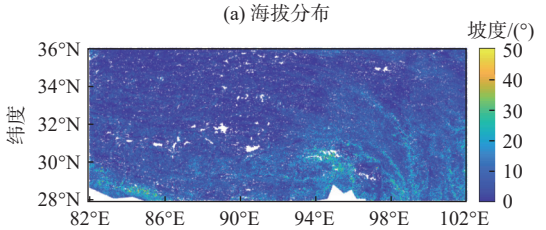
本文选择海拔、坡度和地表粗糙度表示青藏高原的地形特征。青藏高原地区由盆地、山地、峡谷、高原等多种地形组成，使用GTOPO30可获得该地区的海拔并可以计算获得对应坡度。从图3(a)可以看出，研究区域的大部分海拔分布在3 000～5 000 m，从图3(b)可以看出，坡度分布在0°～50°。图4为2020年青藏高原地表粗糙度分布。可以看到，研究区域的粗糙度大部分在0.15～0.16之间。

2.4 机器学习反演算法

土壤湿度是影响反射率的主要因素，同时植被覆盖、地形和土壤质地等环境因素也会影响接收机接收到的GNSS反射信号的强度，因此，土壤湿度反演过程是一个复杂的非线性过程。本文将机器学习算法与星载GNSS-R技术相结合进行青藏高原土壤湿度的反演，是一个多元回归问题。RF和ANN机器学习算法在多元回归问题上有较好的效果[25]。RF是一种多棵决策树集成的算法，其泛化能力强且训练速度快，图5为本文的RF算法流程；ANN是一种自学习和自适应的算法，善于处理复杂非线性问题且鲁棒性较高，图6为本文的ANN网络模



(a) 海拔分布



(b) 坡度分布

图 3　青藏高原的海拔和坡度分布

Fig. 3　Elevation and slope distribution of the Qinghai-Tibet Plateau

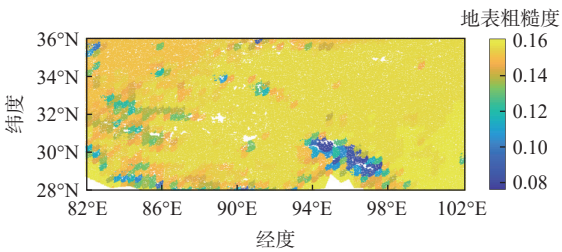


图 4　2020 年青藏高原地表粗糙度分布

Fig. 4　Surface roughness distribution of the Qinghai-Tibet Plateau in 2020

型。经过多次模型参数实验, 综合考虑拟合精度和模型复杂度, 本文确定了 RF 模型决策树个数为 100, 最大特征数为 0.6; ANN 有 2 层隐藏层, 每层由 16 个节点构成。通过对比 RF 和 ANN 的反演精度, 选择最适合青藏高原土壤湿度反演的机器学习算法。

2.5　模型评价指标

选用偏差 E_{b}、均方根误差 E_{rms} 和皮尔逊相关系数 R 作为模型评价指标, 计算公式如下:

$$E_{\mathrm{b}}=\frac{\sum_{a=1}^{n}(Y_a-y_a)}{n} \tag{12}$$

$$E_{\mathrm{rms}}=\sqrt{\frac{\sum_{a=1}^{n}(Y_a-y_a)^2}{n}} \tag{13}$$

$$R=\frac{\sum_{a=1}^{n}(Y_a-\overline{Y})(y_a-\overline{y})}{\sqrt{\sum_{a=1}^{n}(Y_a-\overline{Y})^2\sum_{a=1}^{n}(y_a-\overline{y})^2}} \tag{14}$$

式中: Y_a 为模型反演值; y_a 为 SMAP 土壤湿度值; n 为样本数量; $\overline{Y}$ 为模型反演值均值; $\overline{y}$ 为 SMAP 土壤湿度均值。

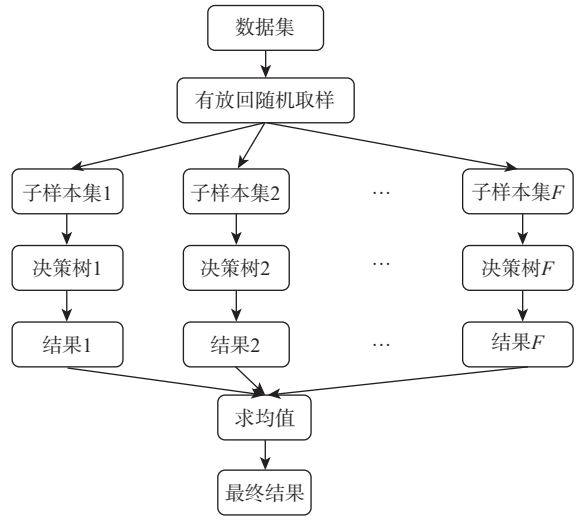


图 5　RF 算法流程

Fig. 5　Flowchart of RF algorithm

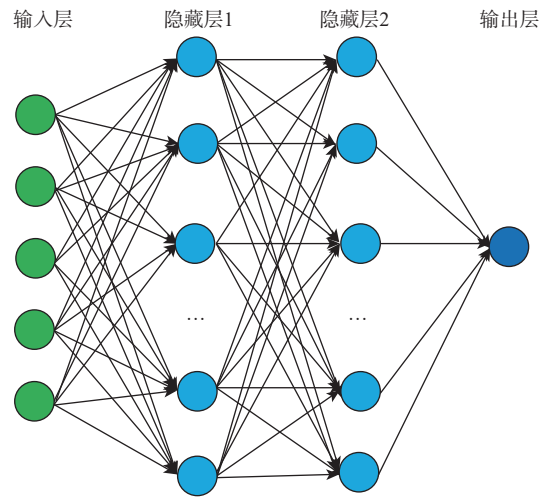


图 6　ANN 网络模型

Fig. 6　ANN network model

模型的偏差和均方根误差越小, 皮尔逊相关系数越大, 说明模型反演效果越好。

3　实验与结果分析

3.1　数据预处理

数据本身的质量会影响后续的反演精度, 因而需要对数据进行预处理。首先, 通过标准质量控制筛除部分 CYGNSS 观测数据[15]; 然后, 按表 1 所示条件再次筛选 CYGNSS 数据。对 GTOPO30 高程数据和 SMAP 产品, 剔除其无效值。数据预处理后, 采用时空最邻近方法为 CYGNSS 的每个镜面反射点匹配地形特征量和 SMAP 参数。CYGNSS 与 SMAP 和 GTOPO30 匹配的空间距离分别小于 36 km 和 1 km, 与 SMAP 时间距离小于 6 h, 与那曲

表1 CYGNSS数据筛选条件

Table 1 CYGNSS data filtering criteria

数据类型	筛选条件
入射角/(°)	0～65
天线增益/dB	>0
信噪比/dB	>0
DDM峰值时延位置	第7～10个

土壤湿度观测网站点匹配的空间距离和时间距离分别小于15 km和小于0.5 h。

3.2 土壤湿度反演流程

本文青藏高原土壤湿度反演的流程如图7所示。首先,进行数据预处理;然后,对CYGNSS的反射率进行修正;最后,以青藏高原2020年解冻期CYGNSS修正后的反射率、入射角、海拔、坡度和地表粗糙度作为输入特征量,SMAP土壤湿度作为验证数据,分别建立RF和ANN土壤湿度反演模型并评估性能,选择内推结果精度最好的模型外推青藏高原2021年解冻期的土壤湿度,并用那曲实测土壤湿度与反演结果进行对比验证。

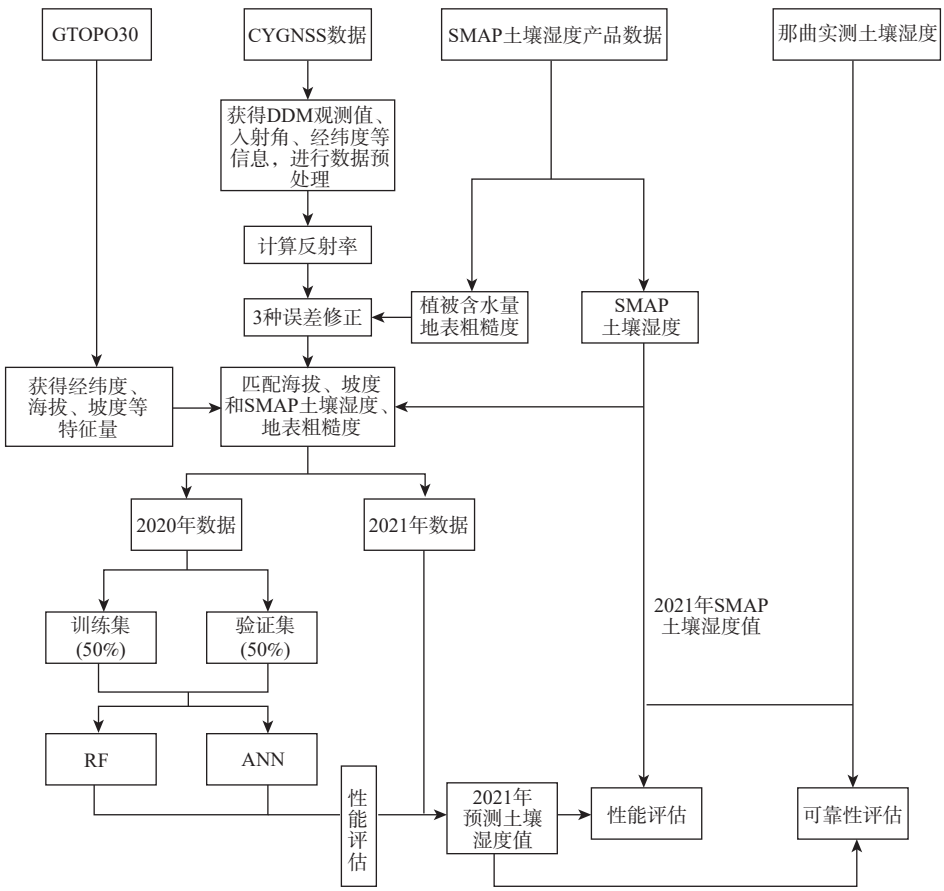


图7 青藏高原土壤湿度反演流程

Fig. 7 Flow chart of soil moisture inversion of the Qinghai-Tibet Plateau

3.3 CYGNSS数据反演青藏高原土壤湿度的可行性验证

为验证反射率修正的效果,本文使用2020年CYGNSS数据(50%为训练集,50%为验证集),仅以CYGNSS反射率和入射角作为输入特征量进行2组实验:①输入未修正的反射率Γ和入射角;②输入修正后的反射率Γ_3和入射角。都分别使用RF和ANN算法。图8、图9分别为RF、ANN反演结果的密度散点图,斜线为等值线,具体反演结果如表2所示。从表2、图8和图9可以看出,RF模型在2组实验中的结果都优于ANN模型。虽然通过未修正的反射率Γ和入射角获得的青藏高原土壤湿度精确度尚可,但和SMAP土壤湿度的皮尔逊相关系数较小,整体趋势与真实土壤湿度相差较大。修正后的反射率Γ_3和入射角的RF模型的反演结果更加准确,均方根误差提高了28%,皮尔逊相关系数提高到了0.690 1。修正后的CYGNSS反射率数据能够明显提高青藏高原的土壤湿度反演结果,说明反射率修正是有效的,后续反演模型均以修正后的反射率Γ_3作为特征量。

3.4 融合地形特征量模型的结果

在GYGNSS参数(修正后的反射率Γ_3和入射角)的基础上仅加入3个地形特征量(海拔、坡向和地表粗糙度),使用2020年解冻期的数据(50%训

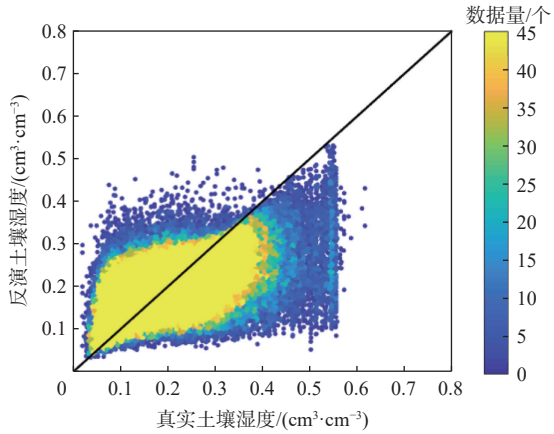


图 8　RF 反演结果的密度散点图

Fig. 8　Density scatter plot of RF inversion results

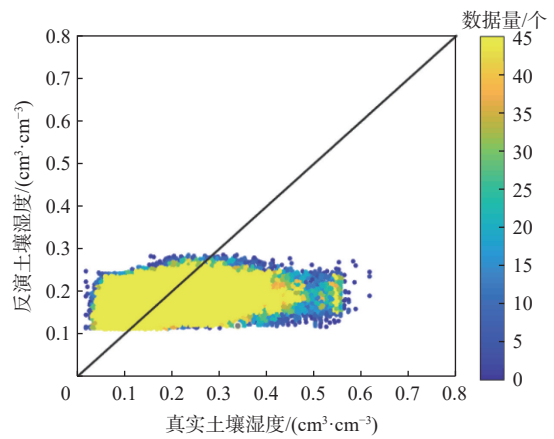


图 9　ANN 反演结果的密度散点图

Fig. 9　Density scatter plot of ANN inversion results

表 2　2020 年反射率修正前后模型反演结果

Table 2　Model inversion results before and after reflectivity correction in 2020

模型	输入特征量	E_b/(cm³·cm⁻³)	E_{rms}/(cm³·cm⁻³)	R
RF	未修正的反射率Γ和入射角	−0.000 2	0.094 1	0.094 4
	修正后的反射率Γ_3和入射角	0.000 1	0.067 5	0.690 1
ANN	未修正的反射率Γ和入射角	−0.000 1	0.145 0	0.161 2
	修正后的反射率Γ_3和入射角	−0.001 1	0.081 4	0.395 2

练集, 50% 验证集) 分别构建 RF 和 ANN 青藏高原土壤湿度反演模型。表 3 为 RF 和 ANN 模型与文献 [16] 的反演结果。可知, 本文的 RF 模型内推偏差为−0.000 7 cm³/cm³, 均方根误差为 0.045 4 cm³/cm³, 皮尔逊相关系数为 0.861 4, ANN 模型内推偏差为−0.000 9 cm³/cm³, 均方根误差为 0.058 4 cm³/cm³, 皮尔逊相关系数为 0.751 7, RF 模型具有更好的反演性能。与文献 [16] 在青藏高原的川藏交通廊道沿线地表的土壤湿度反演结果相比, 本文 RF 模型的偏差绝对值仅为其 7%, 均方根误差高 0.015 4 cm³/cm³, 皮尔逊相关系数与之相当。

此外, 还使用本文研究区域的数据进行了少参

表 3　RF、ANN 模型与文献 [16] 的反演结果

Table 3　Inversion results of RF、ANN models and Ref.[16]

模型	CYGNSS特征量	融合特征量	E_b/(cm³·cm⁻³)	E_{rms}/(cm³·cm⁻³)	R
RF	修正后的反射率、入射角	海拔、坡度、地表粗糙度	−0.000 7	0.045 4	0.861 4
ANN	修正后的反射率、入射角	海拔、坡度、地表粗糙度	−0.000 9	0.058 4	0.751 7
文献[16]	反射率、经度、纬度、时间	地表粗糙度、NDVI、高程、地物类型	0.010	0.030	0.857

数模型(本文的 5 参数)和多参数模型(文献 [16] 的参数)花费时间对比实验。在模型训练部分, 与多参数 RF、ANN 模型相比, 本文的少参数 RF、ANN 模型训练时间减少了 29%、16%。在 RF 模型外推部分, 与多参数 RF 模型相比, 本文的少参数 RF 模型训练时间减少了 13%。不论是模型训练部分还是模型外推部分, 本文提出的少参数模型的运行效率均有一定的提升。

图 10 为 2020 年解冻期 RF 模型对比密度散点图, 斜线为等值线。可知, 散点主要集中在等值线附近, 说明本文基于 CYGNSS 数据融合地形参数反演的土壤湿度与 SMAP 有较好的一致性。与未加入地形参数相比, 均方根误差提高了 32%, 皮尔逊相关系数提高了 25%, 证明融合地形特征量能够有效提升反演精度。

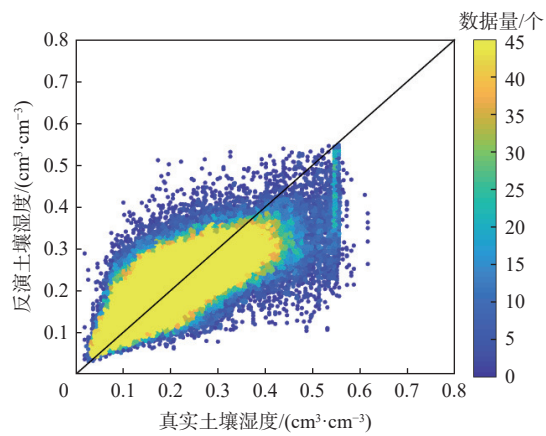


图 10　2020 年青藏高原解冻期 RF 模型对比密度散点图

Fig. 10　Density scatter plot of RF model comparison during thawing period of the Qinghai-Tibet Plateau in 2020

图 11 为 2020 年解冻期土壤湿度分布。可知, 青藏高原解冻期土壤湿度大多在 0.05～0.3 cm³/cm³ 之间, 0.4 cm³/cm³ 以上的数据量较少, 因而导致了模型对 0.4 cm³/cm³ 以上土壤湿度的低估。

3.5　RF 青藏高原土壤湿度反演模型的泛化性

使用 3.4 节构建的 RF 青藏高原土壤湿度反演模型外推 2021 年解冻期青藏高原的土壤湿度, 验证模型的泛化性。表 4 为 RF 模型外推 2021 年与文献 [16] 外推 2019 年的反演结果。可知, 本文 RF 反

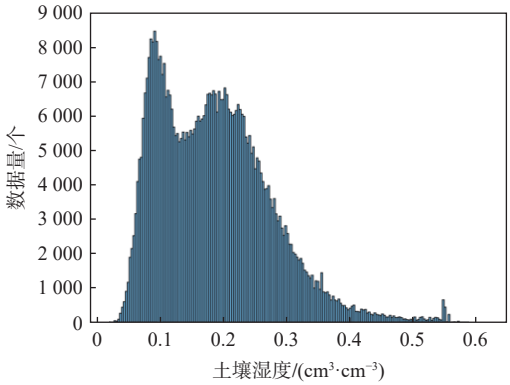


图 11 2020年青藏高原解冻期土壤湿度分布

Fig. 11 Soil moisture distribution during thawing period of the Qinghai-Tibet Plateau in 2020

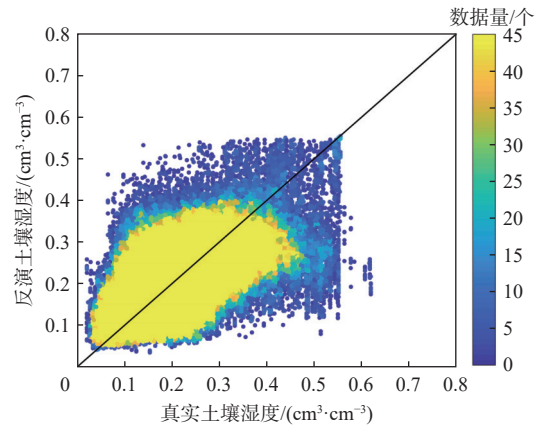


图 12 2021年青藏高原解冻期RF模型对比密度散点图

Fig. 12 Density scatter plot of RF model comparison during thawing period of the Qinghai-Tibet Plateau in 2021

表 4 RF模型2021年解冻期泛化结果与文献[16]2019年泛化结果

Table 4 Generalization results of RF model for 2021 thawing period and Ref. [16] for 2019

时间	E_b/(cm³·cm⁻³)	E_{rms}/(cm³·cm⁻³)	R
2021年6月	−0.012	0.058 9	0.703 0
2021年7月	0.002 1	0.058 3	0.679 2
2021年8月	0.007 9	0.061 8	0.683 2
2021年9月	−0.009 6	0.055 1	0.761 2
2021年6—9月	−0.003 2	0.058 6	0.703 3
2019年(文献[16])	0.010	0.034	0.743

演模型解冻期的偏差在−0.012～0.007 9 cm³/cm³之间，均方根误差在0.055 1～0.061 8 cm³/cm³之间，皮尔逊相关系数在0.679 2～0.761 2之间，6—9月整体的偏差为−0.003 2 cm³/cm³，均方根误差为0.058 6 cm³/cm³，皮尔逊相关系数为0.703 3。与内推结果类似，RF模型外推结果的偏差绝对值为文献[16]的32%，均方根误比文献[16]高0.024 6 cm³/cm³，皮尔逊相关系数与之相当。这一结果可能是因为本文的研究区域更大，包含的地表情况更加复杂多样，不同时空的偏差有正有负，因而最终的偏差相对较小；同时，模型是针对青藏高原整体建立，因而均方根误差相对文献[16]较差。外推结果说明使用CYGNSS融合地形特征量建立的RF青藏高原土壤湿度反演模型具有较好的泛化性，能获得精度较高的土壤湿度，可用于对青藏高原大范围长时序的土壤湿度分析。

图12为2021年解冻期RF模型对比密度散点图，斜线为等值线。可知，散点大部分都集中在等值线附近，0.4 cm³/cm³以上的预测结果同样也存在低估的现象，可能与建立模型时0.4 cm³/cm³以上的训练集样本量较少有关。

如表5所示，本文从0开始，以0.1为步长将2021年解冻期的土壤湿度分为7个区间，给出了模型在每个区间的泛化精度及对应的数据量。从表5可知，训练集数据大部分分布在1、2、3区间，模型在这3个区间精度最好，但相差不大；6、5、4、1这4个区间的训练集数据量逐渐增加，对应的均方根误差也逐渐变好；7区间的训练集数据量仅个位数，且当土壤湿度大于0.65 cm³/cm³时[15]，CYGNSS将不再适用于土壤湿度反演，该区间的反演基本没有意义。从表5的结果可知，机器学习模型的泛化能力极大依赖于训练集的数据量，当训练集数据量太少时(6、5区间)，模型无法从中学习到足够多的信息，在泛化数据集上的表现就较差；随着训练集数据量的增加(6、5、4、1区间)，模型能够学习到更多信息，从而有更好的泛化能力；当模型能够获得足够的信息后，性能逐渐稳定，数据量的增加对模型泛化能力的影响降低(1、3、2区间)。

表 5 不同区间的泛化精度与对应的数据量

Table 5 Generalization accuracy and corresponding data volume of different intervals

区间号	土壤湿度范围/(cm³·cm⁻³)	E_{rms}/(cm³·cm⁻³)	训练集数量	测试集数量
1	(0, 0.1]	0.053 3	48 729	85 804
2	(0.1, 0.2]	0.052 2	102 098	251 642
3	(0.2, 0.3]	0.053 4	79 370	167 619
4	(0.3, 0.4]	0.081 7	20 349	36 311
5	(0.4, 0.5]	0.157 3	3 775	6 055
6	(0.5, 0.6]	0.235 3	1 567	1 393
7	(0.6, 0.7]	0.397 1	3	32

图13为2021年6—9月的土壤湿度反演结果。可以看到，青藏高原土壤湿度整体趋势自西北向东南逐渐升高，与文献[26]研究结果一致，符合青藏高原不同地区的气候区空间分布；青藏高原在解冻期的降水从西北向东南递增，与土壤湿度大小的空间分布相似。同时，从图13也可以看到，从

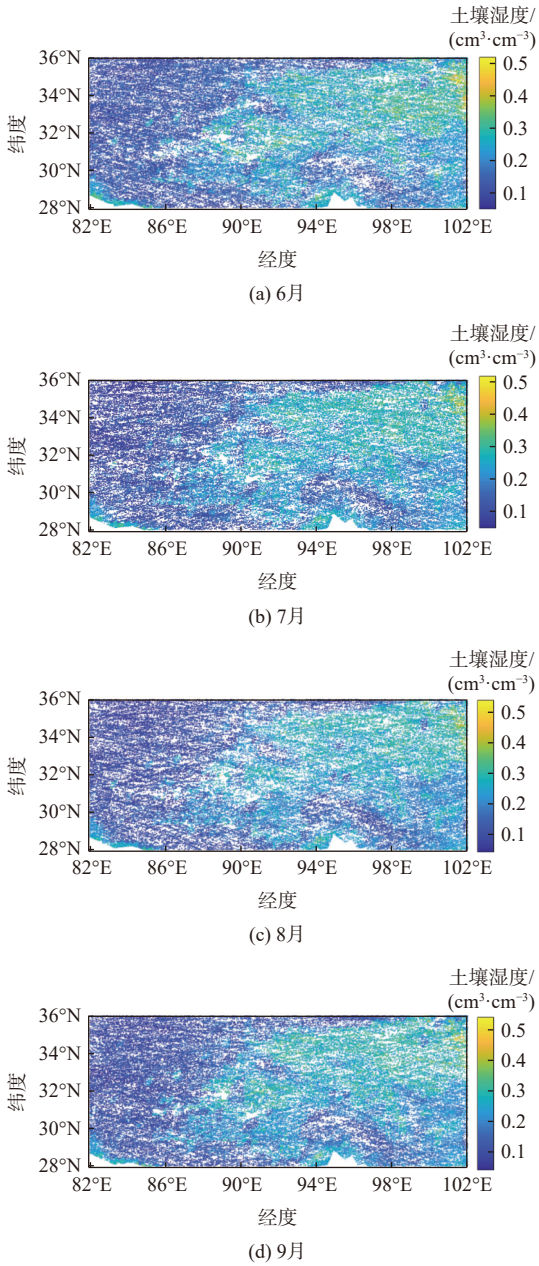


图 13　2021 年 6—9 月的土壤湿度反演结果

Fig. 13　Inversion results of soil moisture in June, July, August, and September, 2021

6 月到 9 月，青藏高原东部土壤湿度较西部变化大，西部土壤湿度整体较低，这可能是因为东部夏季受亚洲季风气候影响降水较多[27]。

3.6　与实测数据对比验证

国家青藏高原科学数据中心网站提供了到 2021 年 9 月 15 日为止的那曲土壤湿度观测网实测土壤湿度数据，因此，本文对 2021 年 6 月 1 日至 9 月 15 日的 CYGNSS 土壤湿度、SMAP 土壤湿度和那曲实测数据进行对比分析。3.5 节中已匹配好 CYGNSS 镜面反射点和 SMAP 土壤湿度数据，再通过时空最邻近的方式与那曲土壤湿度观测网站点数据匹配，将到 CYGNSS 镜面反射点空间距离小于 15 km，同时时间间隔小于 30 min（那曲土壤湿度观测网站点数据每 30 min 记录一次）的所有那曲实测土壤湿度的平均值作为匹配的那曲实测数据，用来评估 CYGNSS 和 SMAP 土壤湿度的结果。图 14 为 2021 年 6 月 1 日至 9 月 15 日 CYGNSS 土壤湿度、SMAP 土壤湿度和那曲土壤湿度观测网实测土壤湿度时序图。可以看出，CYGNSS 对那曲土壤湿度观测网的土壤湿度大都低估，SMAP 则高估。CYGNSS 土壤湿度与那曲实测土壤湿度的偏差、均方根误差和皮尔逊相关系数分别为−0.059 cm^3/cm^3、0.070 cm^3/cm^3和 0.122；SMAP 土壤湿度与那曲实测土壤湿度的偏差、均方根误差和皮尔逊相关系数分别为 0.099 cm^3/cm^3、0.133 cm^3/cm^3和 0.062，CYGNSS 的结果优于 SMAP。

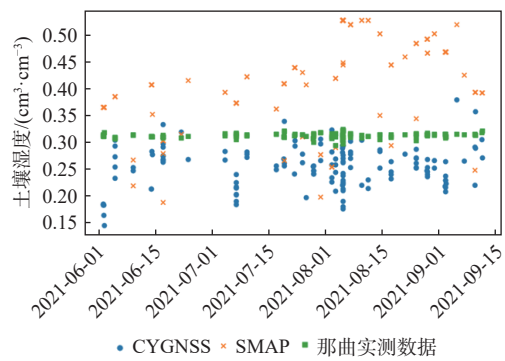


图 14　2021 年 6 月 1 日至 9 月 15 日 CYGNSS 土壤湿度、SMAP 土壤湿度和那曲实测土壤湿度时序图

Fig. 14　Time series of CYGNSS soil moisture, SMAP soil moisture, and measured soil moisture from the Naqu observation network from June 1 to September 15, 2021

4　结　论

本文以 2020 年青藏高原解冻期 CYGNSS 数据为基础，融合少量地形特征量，将 SMAP 土壤湿度作为对照数据，将数据集 5∶5 分为训练集和验证集，分别使用 RF 和 ANN 算法建立了青藏高原土壤湿度反演模型并验证其在验证集上的精度，并对 2021 年解冻期青藏高原（经度：82°E～102°E，纬度：28°N～36°N）的土壤湿度进行了长时序的反演，与那曲土壤湿度观测网站点实测数据对比，得出如下结论：

1）修正后的反射率显著提高了反演精度，修正后的 CYGNSS 反射率能够用于青藏高原土壤湿度的反演。通过仅输入 CYGNSS 的反射率和入射角等 2 个 CYGNSS 特征量，对比反射率修正前后的反演结果，输入修正后的反射率的 RF 模型的偏差为 0.000 1 cm^3/cm^3，均方根误差为 0.067 5cm^3/cm^3，相较于未修正的实验，均方根误差提高了 28%，皮尔

逊相关系数由0.094 4提高到了0.690 1，输入修正后的反射率的ANN模型的偏差为−0.001 1 cm^3/cm^3，均方根误差为0.081 4 cm^3/cm^3，相较于未修正的实验，均方根误差改进了43%，皮尔逊相关系数由0.161 2提高到了0.395 2。RF和ANN的结果均证明了反射率修正的可行性。

2）修正后的反射率和入射角仅融合海拔、坡度和粗糙度等3个地形特征量后，RF模型反演结果的偏差为−0.000 7 cm^3/cm^3，均方根误差为0.045 4 cm^3/cm^3，皮尔逊相关系数为0.861 4；ANN模型反演结果的偏差为−0.000 9 cm^3/cm^3，均方根误差为0.058 4 cm^3/cm^3，皮尔逊相关系数为0.751 7，RF和ANN模型在验证集上的反演精度都得到较大提高，RF模型的结果优于ANN模型；与未加入地形参数相比，RF模型的均方根误差改进了32%，皮尔逊相关系数提高了25%，实现了仅需融合少量地形参数的青藏高原大范围、较高精度的土壤湿度反演。

3）本文建立的RF青藏高原土壤湿度反演模型具有较好的泛化性能。使用RF模型对2021年解冻期的数据反演，反演结果的偏差为−0.003 2 cm^3/cm^3，均方根误差为0.058 6 cm^3/cm^3，皮尔逊相关系数为0.703 3，证明了该模型可以较为准确地对青藏高原大范围土壤湿度进行长时序分析，反演结果的空间分布与研究区域的降水空间分布具有一致性，同时土壤湿度的变化也与气候具有一致性。

4）将2021年青藏高原解冻期的CYGNSS土壤湿度、SMAP土壤湿度分别与那曲土壤湿度观测网实测土壤湿度对比，CYGNSS土壤湿度偏差为−0.059 cm^3/cm^3，均方根误差为0.070 cm^3/cm^3，SMAP土壤湿度偏差为0.099 cm^3/cm^3，均方根误差为0.133 cm^3/cm^3，CYGNSS反演的土壤湿度能够更为准确地反映青藏高原那曲观测网的土壤湿度。

本文实现了基于少参数的机器学习星载GNSS-R的青藏高原大范围及长时序解冻期土壤湿度反演。但由于CYGNSS轨道的限制，不能探测到青藏高原38°N以上的地区，在未来的研究中将结合全球导航卫星掩星探测仪-Ⅱ型（global navigation satellite system occultation sounder Ⅱ，GNOSⅡ）等数据来解决该问题；同时在未来的研究中将通过增加大土壤湿度的训练集数据量作为提高模型精度的方法。

致谢 感谢NASA网站（https://podaac.jpl.nasa.gov）提供的CYGNSS数据，NSIDC网站（https://nsidc.org/data）提供的SMAP数据，USGSEDC网站提供的GTOPO30数据，阳坤等在国家青藏高原科学数据中心网站（https://doi.org/10.11888/Terre.tpdc.271918）提供的青藏高原中部土壤温湿度多尺度观测网数据集（2010—2021年）。

参考文献（References）

[1] 樊威伟, 胡泽勇, 荀学义, 等. 青藏高原季风演变及其气候效应综述[J]. 高原气象, 2021, 40(6): 1294-1303.
FAN W W, HU Z Y, XUN X Y, et al. Review of Qinghai-Xizang Plateau monsoon's evolution and climatic effects[J]. Plateau Meteorology, 2021, 40(6): 1294-1303(in Chinese).

[2] DENTE L, VEKERDY Z, WEN J, et al. Maqu network for validation of satellite-derived soil moisture products[J]. International Journal of Applied Earth Observation and Geoinformation, 2012, 17: 55-65.

[3] KARTHIKEYAN L, PAN M, WANDERS N, et al. Four decades of microwave satellite soil moisture observations. Part 1. A review of retrieval algorithms[J]. Advances in Water Resources, 2017, 109: 106-120.

[4] 杨婷, 陈秀万, 万玮, 等. 基于光学与被动微波遥感的青藏高原地区土壤水分反演[J]. 地球物理学报, 2017, 60(7): 2556-2567.
YANG T, CHEN X W, WAN W, et al. Soil moisture retrieval in the Tibetan Plateau using optical and passive microwave remote sensing data[J]. Chinese Journal of Geophysics, 2017, 60(7): 2556-2567 (in Chinese).

[5] 程渊. 基于集成学习的土壤含水量遥感反演研究[D]. 成都: 电子科技大学, 2020: 1-80.
CHENG Y. Study on remote sensing inversion of soil moisture based on ensemble learning[D]. Chengdu: University of Electronic Science and Technology of China, 2020: 1-80(in Chinese).

[6] GARRISON J L, KURUM M, NOLD B, et al. Remote sensing of root-zone soil moisture using I- and P-band signals of opportunity: instrument validation studies[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2018: 8305-8308.

[7] 张云, 张丹丹, 孟婉婷, 等. CYGNSS/SMAP数据融合半经验模型的土壤湿度反演[J]. 北京航空航天大学学报, 2023, 49(11): 2873-2882.
ZHANG Y, ZHANG D D, MENG W T, et al. Soil moisture retrieval using CYGNSS/SMAP data fusion semi-empirical model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(11): 2873-2882(in Chinese).

[8] ZHANG Y, YIN J W, YANG S H, et al. High wind speed inversion model of CYGNSS sea surface data based on machine learning[J]. Remote Sensing, 2021, 13(16): 3324.

[9] ZHANG Y, LU Q, JIN Q, et al. Global sea surface height measurement from CYGNSS based on machine learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 841-852.

[10] 涂晋升. 利用GNSS-R观测数据反演土壤湿度[D]. 成都: 西南交通大学, 2019: 10-30.
TU J S. Inversion of soil moisture using GNSS-R observation data[D]. Chengdu: Southwest Jiaotong University, 2019: 10-30(in Chinese).

[11] CHEW C, SHAH R, ZUFFADA C, et al. Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1

satellite mission[J]. Geophysical Research Letters, 2016, 43(7): 3317-3324.

[12] CLARIZIA M P, PIERDICCA N, COSTANTINI F, et al. Analysis of CYGNSS data for soil moisture retrieval[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(7): 2227-2235.

[13] WAN W, JI R, LIU B J, et al. A two-step method to calibrate CYGNSS-derived land surface reflectivity for accurate soil moisture estimations[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 2500405.

[14] SENYUREK V, LEI F N, BOYD D, et al. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS[J]. Remote Sensing, 2020, 12(7): 1168.

[15] CHEW C C, SMALL E E. Soil moisture sensing using spaceborne GNSS reflections: comparison of CYGNSS reflectivity to SMAP soil moisture[J]. Geophysical Research Letters, 2018, 45(9): 4049-4057.

[16] 胡羽丰, 汪吉, 李振洪, 等. 基于星载 GNSS-R 获取川藏交通廊道沿线地表土壤湿度[J]. 地球科学, 2022, 47(6): 2058-2068.
HU Y F, WANG J, LI Z H, et al. Land surface soil moisture along Sichuan-Tibet traffic corridor retrieved by spaceborne global navigation satellite system reflectometry[J]. Earth Science, 2022, 47(6): 2058-2068(in Chinese).

[17] 马龙, 李颖. 从 GTOPO30 到 SRTM DEM 精度研究: 以西藏为例[J]. 水土保持通报, 2006, 26(5): 71-74.
MA L, LI Y. Study on accuracy of GTOPO30 and SRTM DEM: a case study of Tibet[J]. Bulletin of Soil and Water Conservation, 2006, 26(5): 71-74(in Chinese).

[18] 荆琛琳. SMAP、CCI 和 CLDAS 土壤湿度产品在青藏高原典型区域的比较验证[J]. 亚热带资源与环境学报, 2020, 15(1): 85-94.
JING C L. Comparative evaluation of SMAP & CCI & CLDAS soil moisture products in typical region of Qinghai-Tibet Plateau[J]. Journal of Subtropical Resources and Environment, 2020, 15(1): 85-94(in Chinese).

[19] 阳坤, 陈莹莹, 赵龙, 等. 青藏高原中部土壤温湿度多尺度观测网数据集 (2010—2021)[DB/OL]. (2021-12-27)[2023-12-01]. https://data.tpdc.ac.cn/en/data/b6269aeb-8b44-4d03-b514-2c804c2cfc26/.
YANG K, CHEN Y Y, ZHAO L, et al. The multiscale observation network of soil temperature and moisture on the central Tibetan Plateau (2010—2021) [DB/OL]. (2021-12-27)[2023-12-01]. https://data.tpdc.ac.cn/en/data/b6269aeb-8b44-4d03-b514-2c804c2cfc26/(in Chinese).

[20] STILLA D, ZRIBI M, PIERDICCA N, et al. Desert roughness retrieval using CYGNSS GNSS-R data[J]. Remote Sensing, 2020, 12(4): 743.

[21] CHEW C, SMALL E. Description of the UCAR/CU soil moisture product[J]. Remote Sensing, 2020, 12(10): 1558.

[22] BINDLISH R, BARROS A P. Parameterization of vegetation backscatter in radar-based, soil moisture estimation[J]. Remote Sensing of Environment, 2001, 76(1): 130-137.

[23] CHOUDHURY B J, SCHMUGGE T J, CHANG A, et al. Effect of surface roughness on the microwave emission from soils[J]. Journal of Geophysical Research: Oceans, 1979, 84(C9): 5699-5706.

[24] CARRENO-LUENGO H, LUZI G, CROSETTO M. First evaluation of topography on GNSS-R: an empirical study based on a digital elevation model[J]. Remote Sensing, 2019, 11(21): 2556.

[25] 乔莹莹. 基于数值预测的机器学习相关算法综述[J]. 安阳工学院学报, 2017, 16(4): 71-74.
QIAO Y Y. Summarization of machine learning based on numerical prediction[J]. Journal of Anyang Institute of Technology, 2017, 16(4): 71-74(in Chinese).

[26] 王静. 青藏高原土壤湿度的时空特征及对长江流域夏季降水的影响[D]. 南京: 南京信息工程大学, 2018: 25-40.
WANG J. Temporal and spatial characteristics of soil moisture in Qinghai-Tibet Plateau and its influence on summer precipitation in Yangtze River basin[D]. Nanjing: Nanjing University of Information Science & Technology, 2018: 25-40(in Chinese).

[27] 黄凌昕, 陈婕, 阳坤, 等. 现代青藏高原亚洲夏季风气候北界及其西风区和季风区划分[J]. 中国科学: 地球科学, 2023, 53(4): 866-878.
HUANG L X, CHEN J, YANG K, et al. The northern boundary of the Asian summer monsoon and division of westerlies and monsoon regimes over the Tibetan Plateau in present-day[J]. Scientia Sinica (Terrae), 2023, 53(4): 866-878(in Chinese).

Integrating topography parameters for soil moisture retrieval using CYGNSS on the Qinghai-Tibet Plateau

ZHANG Yun[1, 2], SHI Liyun[1], YANG Shuhu[1, 2, *], PAN Haiyan[1, 2], HAN Yanling[1, 2], HONG Zhonghua[1, 2]

(1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;
2. Shanghai Engineering Research Center of Marine Intelligent Information and Navigation Remote Sensing, Shanghai Ocean University, Shanghai 201306, China)

Abstract: The soil moisture of the Qinghai-Tibet Plateau plays a crucial role in global atmospheric circulation and climate change. The cyclone global navigation satellite system (CYGNSS), utilizing global navigation satellite system reflectometry (GNSS-R), provides a novel method to monitor soil moisture on the Qinghai-Tibet Plateau; however, the complex topographic environment of the plateau hinders the direct use of CYGNSS reflectivity for soil moisture retrieval. This paper proposes a spaceborne GNSS-R soil moisture machine learning inversion model, which integrates five characteristic parameters: corrected CYGNSS reflectivity, CYGNSS incident angle, and terrain parameters (elevation, slope, surface roughness). First, the CYGNSS reflectance is corrected for two aspects: systematic errors in transmit power, and attenuation induced by surface vegetation and surface roughness. Then, the corrected reflectivity (along with the other four aforementioned parameters) is adopted as input feature quantities, and SMAP soil moisture data is used for model verification. For data partitioning, the 2020 thaw period (June–September) data are randomly split into a training set and a verification set at a 5∶5 ratio. On this basis, two soil moisture inversion models (random forest (RF) and artificial neural network (ANN)) are established specifically for the Qinghai-Tibet Plateau. Use the data from the 2021 thaw period as a test set to examine the generalization ability of the model. The results of the random forest model are better than the artificial neural network model, the inversion result yielding a root mean square error (RMSE) of 0.058 6 cm^3/cm^3 and Pearson correlation coefficient of 0.703 3 on the test set. The model exhibits strong generalization performance: the spatial variation of the inverted soil moisture is consistent with the spatial variation trend of precipitation over the Qinghai-Tibet Plateau. Finally, a comparison between the CYGNSS-derived soil moisture and the in-situ measured soil moisture (from Naqu) shows high accuracy, with a RMSE of 0.070 cm^3/cm^3. The research results show that the inversion model, which integrates corrected CYGNSS reflectance, CYGNSS incident angle, and topography parameters, achieves a more accurate invert of soil moisture in a large range of the Qinghai-Tibet Plateau.

Keywords: soil moisture; Qinghai-Tibet Plateau; cyclone global navigation satellite system; reflectance correction; machine learning

Received: 2023-12-04; **Accepted**: 2024-02-23; **Published Online**: 2024-03-21 08:38

URL: link.cnki.net/urlid/11.2625.V.20240320.0902.001

Foundation items: National Natural Science Foundation of China (42176175,42271335); National Key Research and Development Program of China (2019YFD0900805)

* **Corresponding author**. E-mail: shyang@shou.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0822

基于双参数的GNSS/SINS故障检测及抗差自适应算法

赵桂玲*，王金宝，姜子昊，高帅
(辽宁工程技术大学 测绘与地理科学学院，阜新 123000)

摘 要：全球导航卫星系统(GNSS)信号易受到干扰产生故障，导致滤波器性能降低，影响全球导航卫星系统/捷联惯性导航系统(GNSS/SINS)组合导航系统输出精度。针对该问题，提出一种基于双参数的GNSS/SINS故障检测及抗差自适应算法。算法基于新息残差对系统量测数据进行评估，设计基于平滑有界层宽度的故障检测函数进行故障精确检测；利用故障检测函数值构造抗差协因数矩阵进行误差实时修正，提高状态估计的精度与鲁棒性。阶跃故障和缓变故障2类典型的GNSS故障实验结果表明：基于双参数的GNSS/SINS故障检测及抗差自适应算法与传统的基于残差卡方故障检测的抗差自适应算法相比，速度故障检测率提高7.6%～23.2%，位置故障检测率提高3.2%～12.3%；速度精度提高20.7%～27.1%，位置精度提高22.2%～34.6%。所提算法有效提高了GNSS/SINS组合导航系统的精度与鲁棒性。

关 键 词：组合导航；故障检测；抗差自适应算法；新息残差；平滑有界层宽度
中图分类号：V249.3; P228
文献标志码：A **文章编号**：1001-5965(2026)03-0655-13

全球导航卫星系统(global navigation satellite system, GNSS)和捷联惯性导航系统(strapdown inertial navigation system,SINS)凭借其强大的互补性和高可靠性成为组合导航系统的发展主流，提高了复杂环境下的定位精度，在无人系统、车载导航等多个领域得到广泛应用[1-4]。但在GNSS/SINS组合导航系统中，GNSS信号极易受房屋、树木、电磁及其他因素的干扰产生故障，出现信息异常，从而影响滤波器的性能，降低系统的定位精度和鲁棒性[5-7]。

为对GNSS信号故障进行精准检测并降低故障对系统性能的影响，有学者提出将故障检测理论与抗差自适应理论相结合的算法解决这一问题[8-10]。文献[11]将残差卡方故障检测算法与Sage-Husa抗差自适应算法相结合，先利用残差卡方故障检测算法进行故障检测，再利用Sage-Husa抗差自适应算法对系统噪声进行实时补偿，使系统保持一定的自适应能力。文献[12]将残差卡方故障检测与鲁棒估计相结合，在对故障进行检测的同时，利用故障检测函数值构建抗差协因数矩阵对误差进行实时修正，提升系统的鲁棒性。将故障检测理论与抗差自适应理论相结合的算法在组合导航领域虽得到广泛应用，但对于小阶跃故障或初期故障数值较小的缓变故障无法进行有效检测，从而无法对故障进行有效限制。为此，很多学者将序贯概率比检验算法引入GNSS故障检测。文献[13]将序贯概率比检验算法与隔离抗差自适应算法相结合，通过对数似然比估计故障的发生概率，利用隔离滤波器将GNSS信号故障数据进行抗差自适应处理，提升故障的检测能力，降低故障对系统性能的影响。文献

收稿日期：2023-12-18；**录用日期**：2024-02-23；**网络出版时间**：2024-03-11 17:09
网络出版地址：link.cnki.net/urlid/11.2625.V.20240311.0959.001
基金项目：辽宁省自然科学基金(2020-MS-303)；辽宁省教育厅基本科研项目(JYTMS20230807)
* **通信作者**. E-mail：zhaoguiling@lntu.edu.cn

引用格式：赵桂玲，王金宝，姜子昊，等. 基于双参数的GNSS/SINS故障检测及抗差自适应算法[J]. 北京航空航天大学学报，2026，52(3)：655-667. ZHAO G L，WANG J B，JIANG Z H，et al. A GNSS/SINS fault detection and robust adaptive algorithm based on two parameters[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：655-667 (in Chinese).

[14]在序贯概率比故障检测基础上引入了渐消因子，解决了无法检测故障结束的问题，使算法对故障的检测能力进一步增强。但是，序贯概率比检验算法对故障结束时刻的检测存在较长延迟，且随着故障时间增长，单独惯导解算误差会积累的越来越大，导致系统估计精度降低。

近年来，平滑变结构滤波(smooth variable structure filter, SVSF)算法逐渐应用到组合导航领域[15-18]，其平滑有界层宽度能够体现组合导航系统估计的平均精度，反映滤波器受到的干扰。基于此，本文对平滑有界层宽度模型进行研究，提出一种基于双参数的GNSS/SINS故障检测及抗差自适应算法。与传统基于故障检测的抗差自适应算法相比，本文算法基于新息残差对系统量测数据进行先验质量评估，选取一部分无故障时段数据解算平滑有界层宽度的总体标准差，建立基于平滑有界层宽度的故障检测函数，从而对GNSS信号故障进行精确检测，提高了故障的检测效率；利用故障检测函数值构造抗差协因数矩阵，结合抗差估计，对误差进行实时修正，降低故障对滤波器的影响，提高系统状态估计的精度与鲁棒性。

1 GNSS/SINS组合导航系统模型

由于GNSS/SINS松组合具有结构简单、算法易实现等优点[19-21]，本文采用松组合的方式构建GNSS/SINS组合导航系统滤波模型。选取的导航坐标系为东-北-天。选取SINS输出的导航参数误差$\boldsymbol{\phi}$、$\partial\boldsymbol{v}$、$\partial\boldsymbol{r}$和惯性传感器的零偏$\boldsymbol{\varepsilon}$、$\boldsymbol{\omega}$作为GNSS/SINS组合导航系统的状态向量。

$$\boldsymbol{X}=\left[\boldsymbol{\phi}^{\mathrm{T}},(\partial\boldsymbol{v})^{\mathrm{T}},(\partial\boldsymbol{r})^{\mathrm{T}},\boldsymbol{\varepsilon}^{\mathrm{T}},\boldsymbol{\omega}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1}$$

构建的GNSS/SINS组合导航系统滤波模型状态方程为

$$\boldsymbol{X}_k=\boldsymbol{\Phi}_{k/k-1}\boldsymbol{X}_{k-1}+\boldsymbol{G}_{k-1}\boldsymbol{W}_{k-1} \tag{2}$$

式中：$\boldsymbol{X}_k$为k时刻的状态向量；$\boldsymbol{G}_{k-1}$为噪声转移矩阵；$\boldsymbol{W}_{k-1}$为具有零均值特性的系统噪声矩阵；$\boldsymbol{\Phi}_{k/k-1}$为状态转移矩阵。

选取SINS与GNSS的位置之差和速度之差作为量测量，构建的GNSS/SINS组合导航系统滤波模型量测方程为

$$\boldsymbol{Z}=\begin{bmatrix}\boldsymbol{r}_{\mathrm{SINS}}-\boldsymbol{r}_{\mathrm{GNSS}}\\ \boldsymbol{v}_{\mathrm{SINS}}-\boldsymbol{v}_{\mathrm{GNSS}}\end{bmatrix}=\boldsymbol{H}_k\boldsymbol{X}_k+\boldsymbol{V}_k \tag{3}$$

式中：$\boldsymbol{r}_{\mathrm{SINS}}$、$\boldsymbol{r}_{\mathrm{GNSS}}$和$\boldsymbol{v}_{\mathrm{SINS}}$、$\boldsymbol{v}_{\mathrm{GNSS}}$分别为SINS和GNSS在导航坐标系下的位置和速度；$\boldsymbol{H}_k$为量测矩阵；$\boldsymbol{V}_k$为量测噪声矩阵。

2 平滑有界层宽度模型

SVSF算法中的平滑有界层宽度能够体现组合导航系统估计的平均精度，反映滤波器受到的干扰。但SVSF算法中的平滑有界层宽度只包含了一个变量形式φ_{11}，单一的平滑有界层宽度对应多维量测误差，并未充分考虑量测误差之间可能存在的耦合关系[22]。为使每个量测误差都有相应的平滑有界层宽度与之对应，定义一个全平滑有界层宽度矩阵$\boldsymbol{\varphi}$。

$$\boldsymbol{\varphi}=\begin{bmatrix}\varphi_{11}&\varphi_{12}&\cdots&\varphi_{1n}\\ \varphi_{21}&\varphi_{22}&\cdots&\varphi_{2n}\\ \vdots&\vdots&&\vdots\\ \varphi_{n1}&\varphi_{n2}&\cdots&\varphi_{nn}\end{bmatrix} \tag{4}$$

考虑全平滑有界层宽度矩阵中各个平滑有界层宽度之间的耦合关系，状态估计协方差矩阵$\boldsymbol{P}_k$和全平滑有界层宽度矩阵$\boldsymbol{\varphi}$之间的函数关系为

$$\frac{\partial(\mathrm{tr}[\boldsymbol{P}_k])}{\partial\boldsymbol{\varphi}}=\boldsymbol{0} \tag{5}$$

由于全平滑有界层宽度矩阵是一个n维方阵，对SVSF算法[15]增益矩阵$\boldsymbol{K}_k$进行改进：

$$\boldsymbol{K}_k=\boldsymbol{H}_k^{-1}\left[\mathrm{diag}\left(\left|\boldsymbol{e}_{k/k-1}\right|+\gamma\left|\boldsymbol{e}_{k-1/k-1}\right|\right)\circ \mathrm{sat}(\boldsymbol{\varphi}_k^{-1}\mathrm{diag}(\boldsymbol{e}_{k/k-1}))\right]\left[\mathrm{diag}\left(\boldsymbol{e}_{k/k-1}\right)\right]^{-1} \tag{6}$$

式中：γ为记忆因子，且$0<\gamma<1$[18]；$\boldsymbol{e}_{k/k-1}$为k时刻的先验新息残差；$\boldsymbol{e}_{k-1/k-1}$为$k-1$时刻的后验新息残差；“$\circ$”表示Hadmard乘积；$\boldsymbol{\varphi}_k$为k时刻的平滑有界层宽度矩阵；sat()表示对“()”中的列向量各个元素进行饱和运算[23]。

将式(6)的形式进行简化，令

$$\tilde{\boldsymbol{C}}=\mathrm{diag}\left(\left|\boldsymbol{e}_{k/k-1}\right|+\gamma\left|\boldsymbol{e}_{k-1/k-1}\right|\right) \tag{7}$$

则式(6)中的增益矩阵$\boldsymbol{K}_k$为

$$\boldsymbol{K}_k=\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1} \tag{8}$$

状态估计协方差矩阵$\boldsymbol{P}_k$为

$$\begin{aligned}\boldsymbol{P}_k=&(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k/k-1}(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)^{\mathrm{T}}+\boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}}=\\ &\boldsymbol{P}_{k/k-1}-\boldsymbol{K}_k\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}-\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{K}_k^{\mathrm{T}}+\\ &\boldsymbol{K}_k\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{K}_k^{\mathrm{T}}+\boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}}\end{aligned} \tag{9}$$

式中：$\boldsymbol{R}_k$为量测噪声协方差矩阵。

新息残差协方差矩阵[24]为

$$\boldsymbol{D}_k=\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k \tag{10}$$

将式(8)和式(10)代入式(9)得

$$\begin{aligned}\boldsymbol{P}_k=&\boldsymbol{P}_{k/k-1}-\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}-\\ &\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\right)^{\mathrm{T}}+\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\boldsymbol{D}_k\left(\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\right)^{\mathrm{T}}\end{aligned} \tag{11}$$

将式(11)代入式(5)得

$$\frac{\partial(\mathrm{tr}[\boldsymbol{P}_k])}{\partial\boldsymbol{\varphi}}=\frac{\partial\left(\mathrm{tr}\left[\boldsymbol{P}_{k/k-1}\right]\right)}{\partial\boldsymbol{\varphi}}-\frac{\partial\left(\mathrm{tr}\left[\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\right]\right)}{\partial\boldsymbol{\varphi}}-\frac{\partial\left(\mathrm{tr}\left[\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\right)^{\mathrm{T}}\right]\right)}{\partial\boldsymbol{\varphi}}+\frac{\partial\left(\mathrm{tr}\left[\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\boldsymbol{D}_k\left(\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\right)^{\mathrm{T}}\right]\right)}{\partial\boldsymbol{\varphi}}=\boldsymbol{0}\tag{12}$$

将式 (12) 中各部分偏导数进行解算，整理可得

$$\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}-\boldsymbol{H}_k^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\varphi}_k^{-1}\boldsymbol{D}_k=\boldsymbol{0}\tag{13}$$

求解式(13)可得

$$\boldsymbol{\varphi}_k=(\tilde{\boldsymbol{C}}^{-1}\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{D}_k^{-1})^{-1}\tag{14}$$

将式 (7) 代入式 (14)，得到与多维量测误差一一对应的平滑有界层宽度矩阵$\boldsymbol{\varphi}_k$。

$$\boldsymbol{\varphi}_k=\left\{\left[\mathrm{diag}\left(\left|\boldsymbol{e}_{k/k-1}\right|+\gamma\left|\boldsymbol{e}_{k-1/k-1}\right|\right)\right]^{-1}\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{D}_k^{-1}\right\}^{-1}\tag{15}$$

从式 (15) 可以看出，平滑有界层宽度是由当前时刻的先验新息残差$\boldsymbol{e}_{k/k-1}$、量测矩阵$\boldsymbol{H}_k$、先验状态估计协方差矩阵$\boldsymbol{P}_{k/k-1}$、新息残差协方差矩阵$\boldsymbol{D}_k$和前一时刻后验新息残差$\boldsymbol{e}_{k-1/k-1}$所构成的函数，平滑有界层宽度与系统模型所估计的系统噪声和量测噪声有着紧密的关系，能够反映滤波器受到的干扰。一般认为系统模型和系统噪声具有良好的稳定性，量测噪声容易受到外界环境的影响，且对滤波器的影响较为明显，因此，平滑有界层宽度为组合导航故障检测函数的构造提供了新思路。

3　算法设计

3.1　基于双参数的 GNSS/SINS 故障检测算法

当 GNSS 量测数据良好、系统无故障时，新息残差$\boldsymbol{e}_k$符合均值为零的高斯分布，当 GNSS 量测数据发生异常、系统出现故障时，可认为此时新息残差$\boldsymbol{e}_k$不符合均值为零的高斯分布[25]。对新息残差$\boldsymbol{e}_k$做二元假设，构成的假设检验问题可表示为

$$\begin{cases}H_0:\ \boldsymbol{e}_k\sim N(0,\boldsymbol{D}_k)\\ H_1:\ \boldsymbol{e}_k\sim N(\emptyset,\boldsymbol{D}_k)\end{cases}\tag{16}$$

式中：H_0代表无故障发生，H_1代表有故障发生；$\emptyset$表示不为 0。

将新息残差$\boldsymbol{e}_k$与新息残差协方差矩阵$\boldsymbol{D}_k$归一化可得$\boldsymbol{T}_k$：

$$\boldsymbol{T}_k=\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{D}_k^{-1}\boldsymbol{e}_k\tag{17}$$

式中：$\boldsymbol{T}_k$服从自由度为τ的卡方分布[26]。

利用卡方检验法，设置阈值T_{D}，判断$\boldsymbol{T}_k$与阈值T_{D}的大小关系，初步筛选出系统无故障时段。选取一部分系统无故障时段数据，求得选取系统无故障时段数据对应平滑有界层宽度的标准差δ_i：

$$\delta_i=\sqrt{\frac{\sum_{i=1}^{6}(|\boldsymbol{\varphi}_1|-\overline{\varphi}_i)^2(|\boldsymbol{\varphi}_2|-\overline{\varphi}_i)^2\cdots(|\boldsymbol{\varphi}_6|-\overline{\varphi}_i)^2}{6}}\tag{18}$$

式中：i为维数。

对平滑有界层宽度的标准差δ_i进行总体标准差求解，得到每一维的总体标准差$\tilde{\delta}_i$为

$$\tilde{\delta}_i=\sqrt{\frac{\sum_{t=1}^{m}(\delta_1-\overline{\delta})^2(\delta_2-\overline{\delta})^2\cdots(\delta_m-\overline{\delta})^2}{m}}\tag{19}$$

式中：$\overline{\delta}$为平均值；m为选取无故障时段数据个数，本文m取 300。

在滤波过程中，结合总体标准差$\tilde{\delta}_i$构建的实时故障检测函数为

$$\lambda_{ki}=\delta_{ki}/\tilde{\delta}_i\tag{20}$$

式中：λ_{ki}为第k时刻第i维量测信息对应的故障检测函数；δ_{ki}为第k时刻第i维平滑有界层宽度的标准差。

选取阈值θ，将故障检测函数λ_{ki}与阈值θ进行比较。当故障检测函数λ_{ki}小于阈值θ时，表明量测数据正常，系统无故障存在；当故障检测函数λ_{ki}大于或等于阈值θ时，表明量测数据异常，系统出现故障。

$$\begin{cases}\lambda_{ki}<\theta & \text{量测数据正常}\\ \lambda_{ki}\geqslant\theta & \text{量测数据异常}\end{cases}\tag{21}$$

式中：θ为计算所得的常数，一般为 3.5～4。

由于平滑有界层宽度与模型所估计的系统噪声和量测噪声有着紧密的关系，由平滑有界层宽度标准差构成的故障检测函数λ_{ki}对故障有很高的灵敏度，可以有效提高故障检测效率。

3.2　基于双参数的 GNSS/SINS 抗差自适应算法

当量测信息发生异常，系统检测出故障后，采用故障检测函数值构造抗差协因数矩阵$\boldsymbol{\beta}_k$：

$$\beta_{ki}=\begin{cases}1 & \lambda_{ki}<\theta\\ \dfrac{\theta}{\lambda_{ki}} & \lambda_{ki}\geqslant\theta\end{cases}\tag{22}$$

式中：$\boldsymbol{\beta}_k=\mathrm{diag}\left[\beta_{k1},\beta_{k2},\cdots,\beta_{km}\right]$。

结合抗差估计理论，通过抗差协因数矩阵实时调节卡尔曼滤波增益矩阵$\boldsymbol{K}_k$中量测噪声的权重，降低故障对系统精度的影响。

$$\boldsymbol{K}_k=\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k/k-1}\boldsymbol{H}_k^{\mathrm{T}}+\beta_{ki}^{-1}\boldsymbol{R}_k)^{-1}\tag{23}$$

$$\boldsymbol{X}_k=\boldsymbol{X}_{k/k-1}+\boldsymbol{K}_k(\boldsymbol{Z}_k-\boldsymbol{H}_k\boldsymbol{X}_{k/k-1})\tag{24}$$

相应的，状态估计协方差矩阵$\boldsymbol{P}_k$为

$$\boldsymbol{P}_k=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k/k-1}(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)^{\mathrm{T}}+\boldsymbol{K}_k\beta_{ki}^{-1}\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}}\tag{25}$$

当量测信息发生异常、系统检测出故障时，通过抗差协因数矩阵$\boldsymbol{\beta}_k$自适应调节量测噪声协方差

矩阵$\boldsymbol{R}_k$，减小观测值增益，提高组合导航系统的精度。综上，本文提出的基于双参数的GNSS/SINS故障检测及抗差自适应算法流程如图1所示。

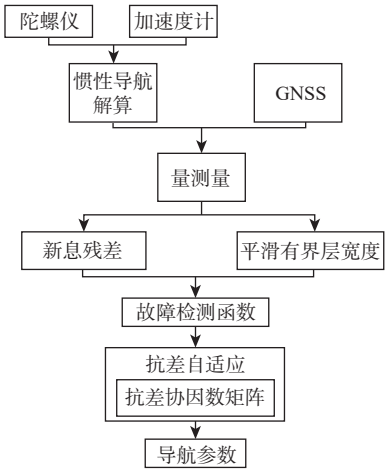


图1 本文算法流程

Fig. 1 Flow chart of the proposed algorithm

4 实验验证与分析

为验证本文算法的先进性与可靠性，设计阶跃故障、缓变故障2类GNSS典型故障进行车载组合导航实验验证与分析。为进一步对比2种算法的检测效果并避免实验结果的偶然性，每类故障实验进行100次。以商用组合后处理软件IE解算的GNSSRTK/INS紧组合平滑结果作为参考值。将选取的车载组合导航实验数据分别加入大阶跃故障、小阶跃故障和缓变故障，采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的GNSS/SINS故障检测及抗差自适应算法进行处理，分别对比2种算法的故障检测效果和抗差效果。为准确体现速度、位置的故障情况，残差卡方故障检测函数采用序贯滤波形式，服从自由度为1的卡方分布，根据卡方分布置信区间求得阈值为6.63，基于双参数的GNSS/SINS故障检测及抗差自适应算法阈值为4。为使检测效果对比更加明显，利用二分法，将有故障时段统计量数值设为1，无故障时段统计量数值设为0。车载组合导航实验设备参数如表1所示。车载组合导航实验设备运行轨迹如图2所示。

表1 GNSS和SINS设备参数

Table 1 Equipment parameters of GNSS and SINS

设备	陀螺仪		加速度计	
	零偏/((°)·h^{-1})	频率/Hz	零偏/((°)·h^{-1})	频率/Hz
SINS	0.8	200	100	200
设备	**速度**		**位置**	
	误差/(m·s^{-1})	频率/Hz	零偏/(m·s^{-1})	频率/Hz
GNSS	0.1	1	[10,10,30]	1

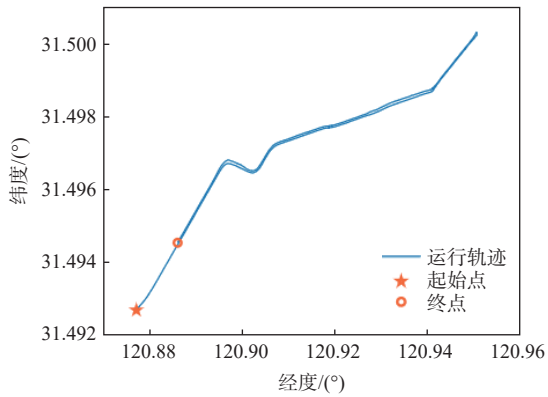


图2 车载组合导航实验设备运行轨迹

Fig. 2 Running trajectory of vehicle integrated navigation experimental equipment

4.1 阶跃故障实验

4.1.1 大阶跃故障

在GNSS中，阶跃故障是一种常见的数据故障类型，假定在301～600 s的时间段内发生了大阶跃故障，速度故障大小为2 m/s，位置故障大小为200 m，分别采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的GNSS/SINS故障检测及抗差自适应算法进行故障检测，得到2种算法的故障检测函数曲线如图3所示。

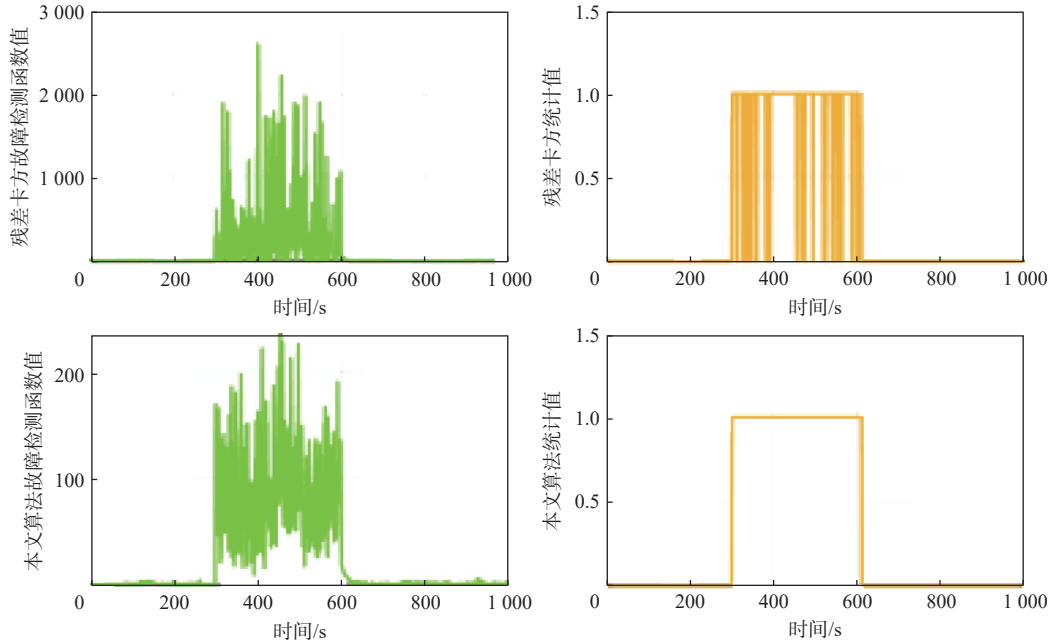
为进一步对比2种算法的检测效果并消除偶然性的影响，共进行100次大阶跃故障实验，得到2种算法速度、位置故障的漏警时长对比曲线，如图4所示。

将上述大阶跃故障实验结果进行统计，得到2种算法的首次故障检测时间和平均漏警时长，如表2所示。

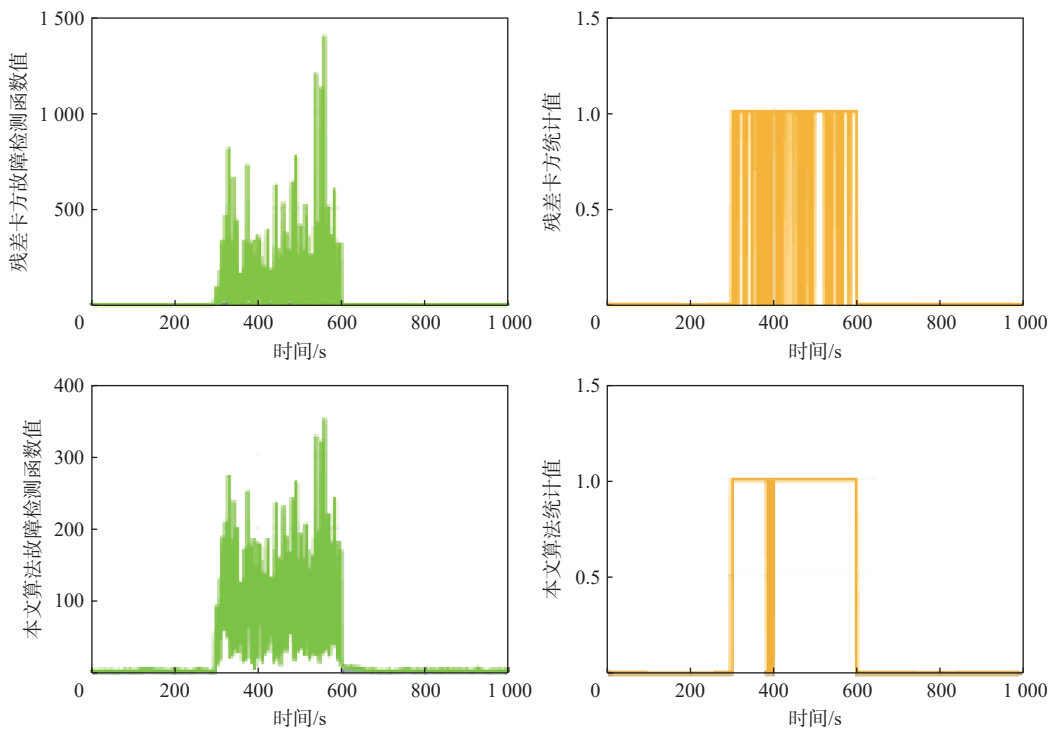
由图3、表2可知，在大阶跃故障时段，2种算法在速度和位置方面的首次故障检测时间均为301 s，均第一时间检测到大阶跃故障的存在。由图4、表2可知，在大阶跃故障时段，残差卡方算法在速度和位置方面的平均漏警时长分别为23.6 s和10.0 s，故障检测率分别为92.1%和96.7%；本文算法在速度和位置方面的平均漏警时长分别为0.8 s和0.3 s，故障检测率分别为99.7%和99.9%；与残差卡方算法相比，本文算法的漏警时长明显减少，故障检测率在速度和位置方面分别提高约7.6%和3.2%，具有更高的灵敏度和更好故障检测性能。

采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的GNSS/SINS故障检测及抗差自适应算法进行处理，得到2种算法的速度、位置误差曲线，如图5所示；2种算法的速度、位置均方根误差的统计结果如表3所示。

由图5、表3可知，当GNSS发生大阶跃故障时，本文算法估计的速度误差和位置误差更小，得到的



(a) 速度故障检测函数



(b) 位置故障检测函数

图 3　大阶跃故障下 2 种算法的故障检测函数

Fig. 3　Fault detection functions of two algorithms under large step fault

速度精度和位置精度更高，与残差卡方算法相比，速度精度和位置精度分别提升了约 27.1% 和 30.4%，提高了 GNSS/SINS 组合导航系统的精度和鲁棒性。

4.1.2　小阶跃故障

假定在 301～600 s 的时间段内发生了小阶跃故障，速度故障大小为 0.5 m/s，位置故障大小为 50 m，分别采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法进行故障检测，得到 2 种算法的故障检测函数曲线，如图 6 所示。

为进一步对 2 种算法的检测效果进行对比并消除偶然性的影响，共进行 100 次小阶跃故障实验，得到 2 种算法速度、位置故障漏警时长对比曲线，如图 7 所示。

将上述小阶跃故障实验结果进行统计，得到 2 种算法的首次故障检测时间和平均漏警时长，如表 4 所示。

由图 6、表 4 可知，在小阶跃故障时段，2 种算

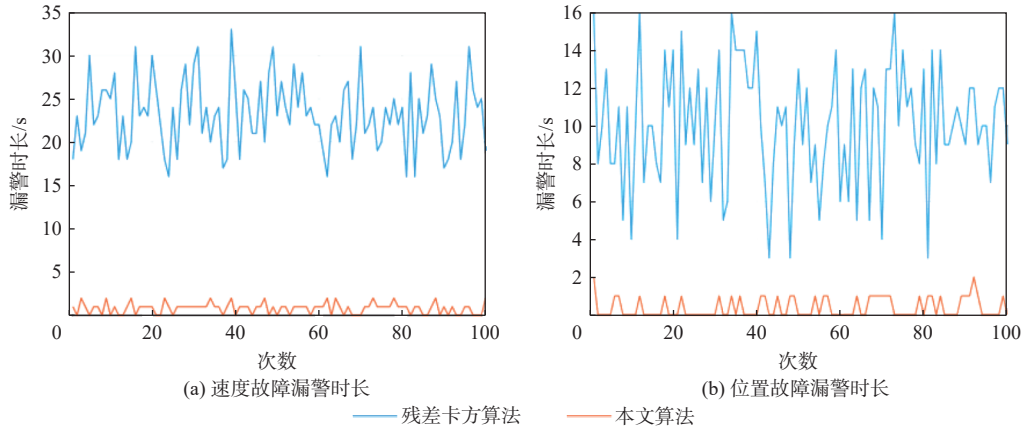


图 4 大阶跃故障下 2 种算法速度、位置故障的漏警时长

Fig. 4 Missed alarm duration of velocity, position fault of two algorithms under large step fault

表 2 大阶跃故障下 2 种算法的检测效果对比

Table 2 Comparison of two fault detection algorithms under large step fault

算法	故障时段/s	首次故障检测时间/s		平均漏警时长/s		故障检测率/%		故障检测率提高比例/%	
		速度	位置	速度	位置	速度	位置	速度	位置
残差卡方	301～600	301	301	23.6	10.0	92.1	96.7		
本文算法	301～600	301	301	0.8	0.3	99.7	99.9	7.6	3.2

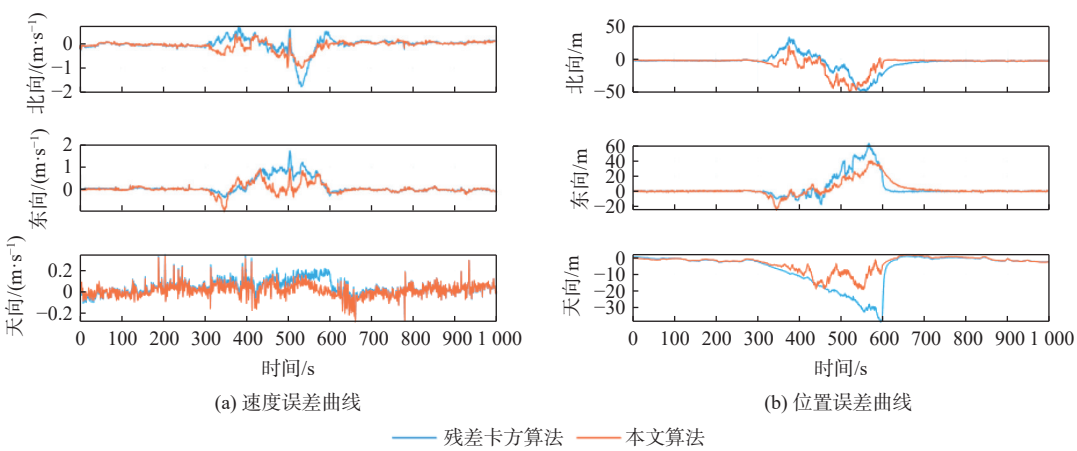


图 5 大阶跃故障下 2 种算法的速度、位置误差曲线

Fig. 5 Velocity and position error curves of two algorithms under large step fault

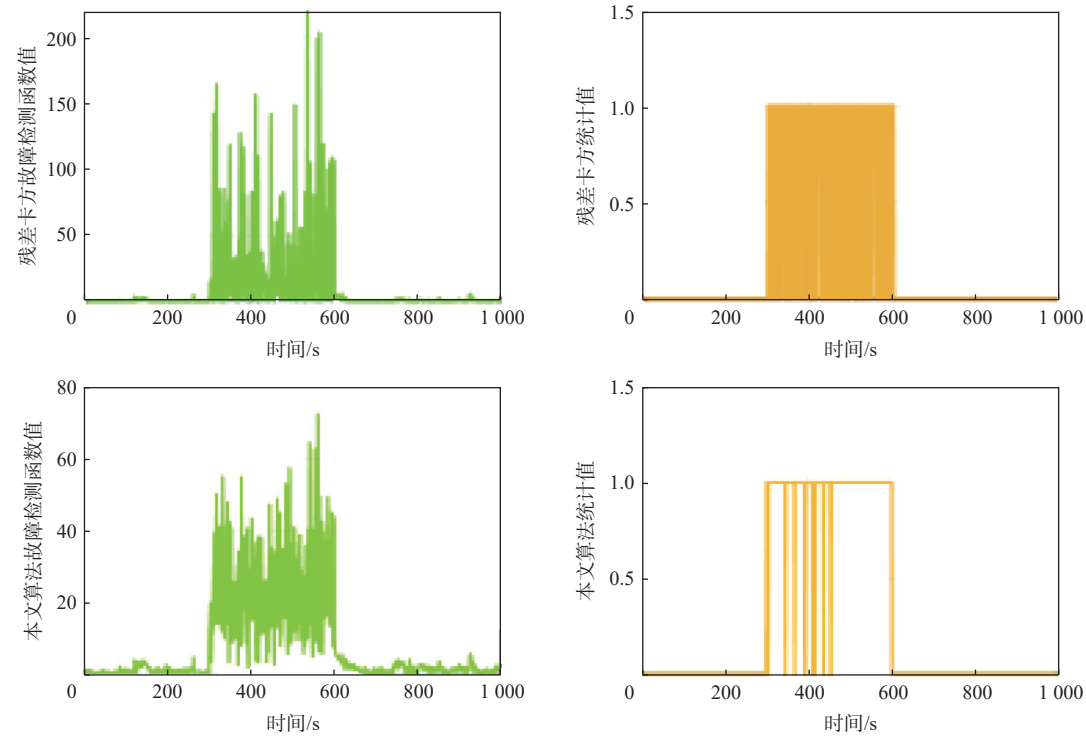
表 3 大阶跃故障下 2 种算法的速度、位置均方根误差

Table 3 Root mean square error of velocity and position of two algorithms under large step fault

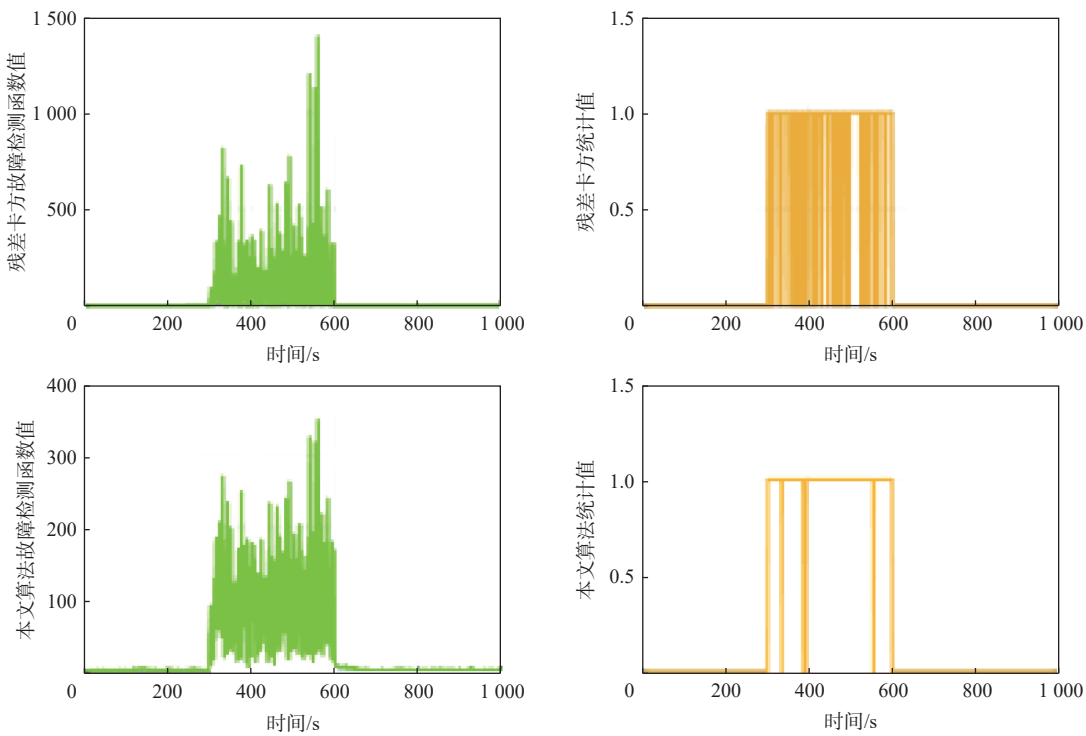
算法类型	速度均方根误差/(m·s^{-1})			位置均方根误差/m			精度提高比例/%	
	北向	东向	天向	北向	东向	天向	速度	位置
残差卡方	0.532 9	0.606 9	0.124 0	22.164 8	25.619 6	19.661 8		
本文算法	0.499 8	0.384 0	0.076 5	20.397 2	16.943 1	9.936 6	27.1	30.4

法在速度和位置方面的首次故障检测时间均为301 s，2 种算法均第一时间检测到小阶跃故障的存在。由图 7、表 4 可知，在小阶跃故障时段，残差卡方算法在速度和位置方面的平均漏警时长分别为86.2 s 和 40.9 s，故障检测率分别为 71.3% 和 86.4%；本文算法在速度和位置方面的平均漏警时长分别为 16.6 s 和 4.0 s，故障检测率分别为 94.5% 和 98.7%；与残差卡方算法相比，本文算法的漏警时长大幅减少，故障检测率在速度和位置方面分别提高约 23.2% 和 12.3%，灵敏度和故障检测性能大幅度提高。

采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法进行处理，得到的 2 种算法的速度、位置误差曲线如图 8 所示；2 种算法的速度、



(a) 速度故障检测函数



(b) 位置故障检测函数

图 6　小阶跃故障下 2 种算法的故障检测函数

Fig. 6　Fault detection functions of two algorithms under small step fault

位置误差的统计结果如表 5 所示。

由图 8、表 5 可知，当 GNSS 发生小阶跃故障时，本文算法估计的速度误差和位置误差更小，得到的速度精度和位置精度更高，与残差卡方算法相比，速度精度和位置精度分别提升了约 25.9% 和 34.6%，提高了 SINS/GNSS 组合导航系统的精度和鲁棒性。

4.2　缓变故障实验

在 GNSS 中，缓变故障是一种常见的数据故障类型，假定在 301～600 s 的时间段内发生了缓变故障，速度缓变规律为 $\sigma_v \sin[2\pi(t-300)/100]$，$\sigma_v$= 1 m/s，位置缓变规律为 $\sigma_r \sin[2\pi(t-300)/100]$，$\sigma_r$=

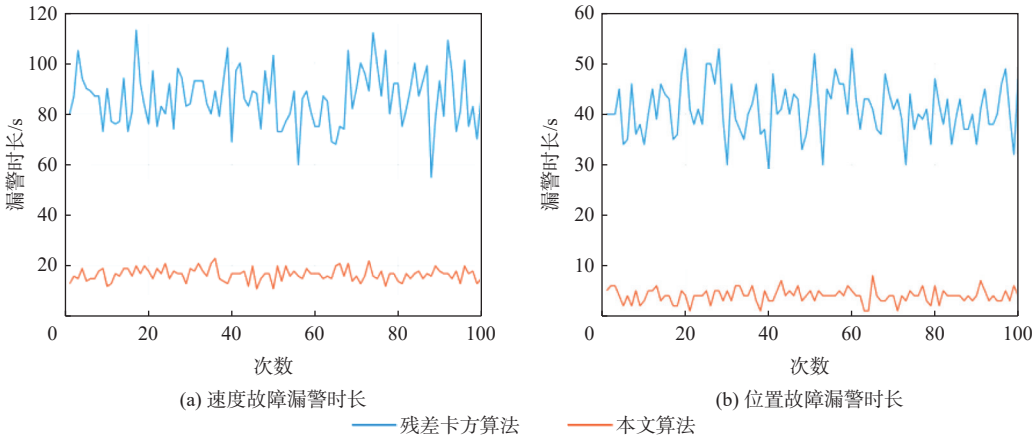


图 7 小阶跃故障下 2 种算法速度、位置故障的漏警时长

Fig. 7 Missed alarm duration of velocity, position fault of two algorithms under small step fault

表 4 小阶跃故障下 2 种算法的检测效果对比

Table 4 Comparison of two fault detection algorithms under small step fault

算法	故障时段/s	首次故障检测时间/s		平均漏警时长/s		故障检测率/%		故障检测率提高比例/%	
		速度	位置	速度	位置	速度	位置	速度	位置
残差卡方	301～600	301	301	86.2	40.9	71.3	86.4		
本文算法	301～600	301	301	16.6	4.0	94.5	98.7	23.2	12.3

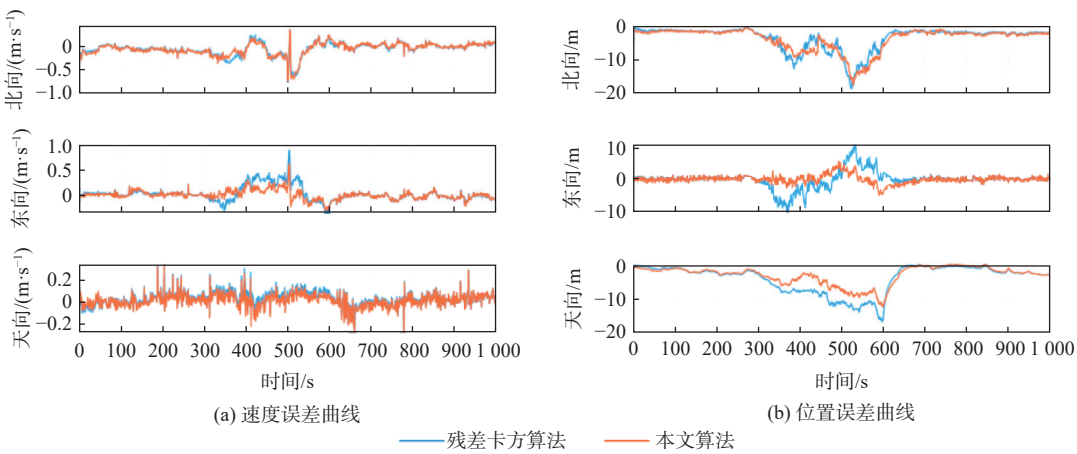


图 8 小阶跃故障下 2 种算法的速度、位置误差曲线

Fig. 8 Velocity and position error curves of two algorithms under small step fault

表 5 小阶跃故障下 2 种算法的速度、位置均方根误差

Table 5 Root mean square error of velocity and position of two algorithms under small step fault

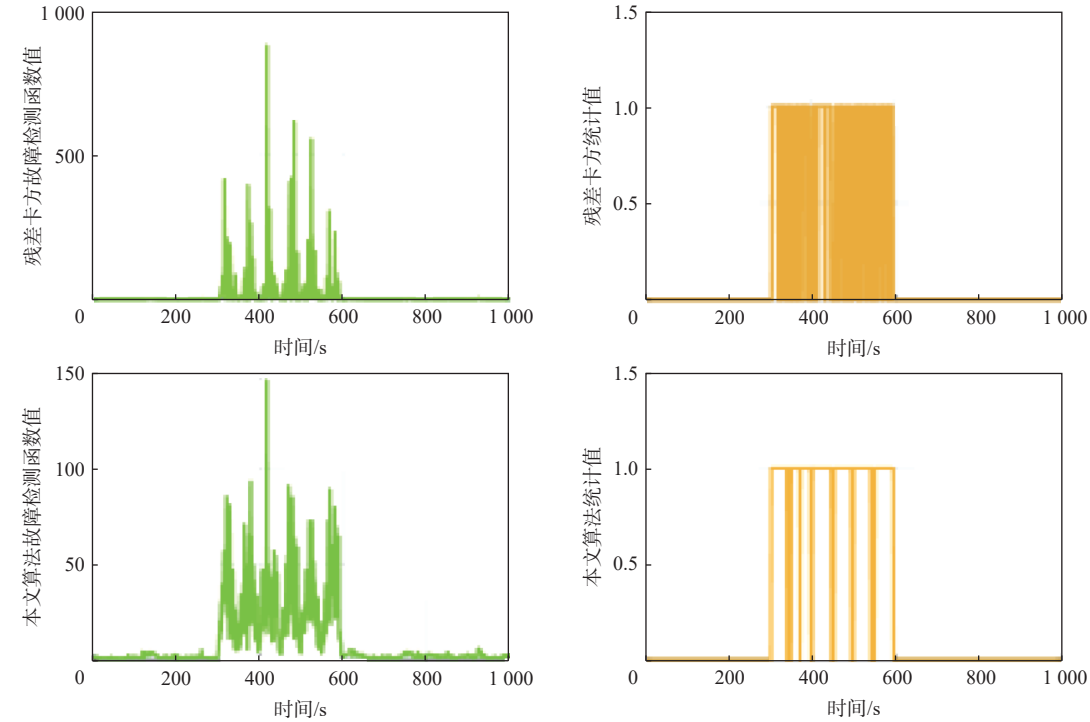
算法	速度均方根误差/(m·s⁻¹)			位置均方根误差/m			精度提高比例/%	
	北向	东向	天向	北向	东向	天向	速度	位置
残差卡方	0.231 1	0.258 1	0.090 4	8.604 8	4.807 2	9.714 6		
本文算法	0.202 5	0.145 4	0.070 9	7.701 4	2.034 2	6.261 0	25.9	34.6

100 m, 分别采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法进行检测, 得到 2 种算法的故障检测函数曲线, 如图 9 所示。

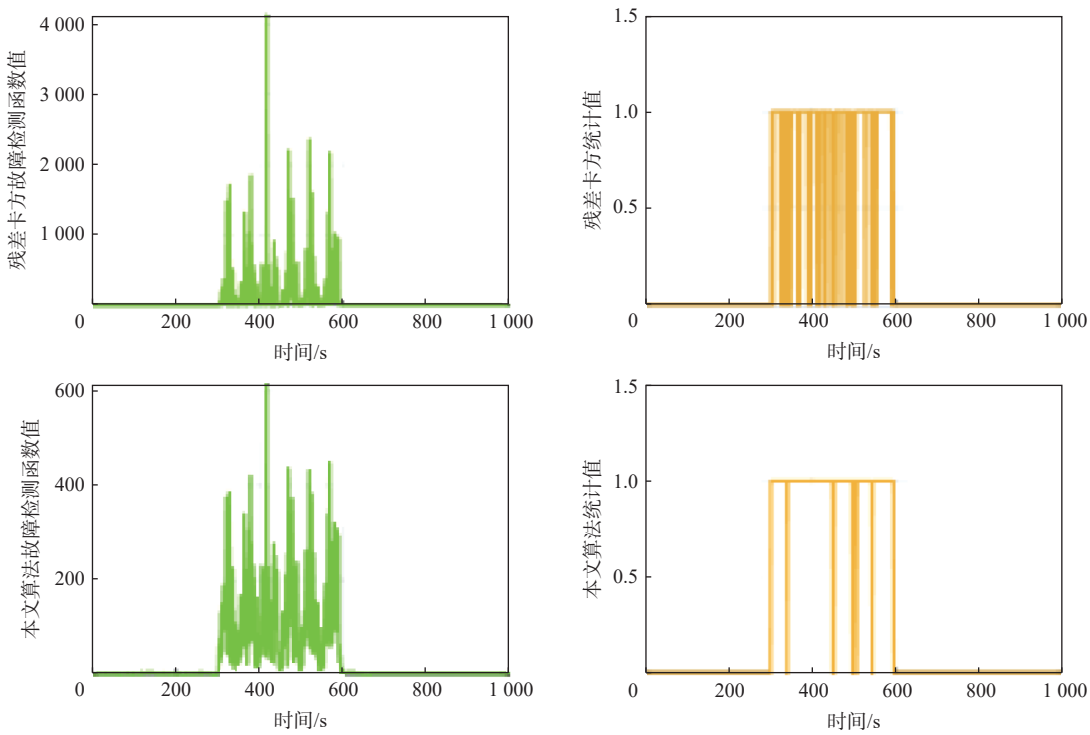
为进一步对 2 种算法的检测效果进行对比并消除偶然性的影响, 共进行 100 次缓变故障实验, 得到 2 种算法速度、位置故障漏警时长对比曲线, 如图 10 所示。

将上述缓变故障实验结果进行统计, 得到 2 种算法的首次故障检测时间和平均漏警时长, 如表 6 所示。

由图 9、表 6 可知, 在缓变故障时段, 残差卡方算法在速度和位置方面的首次故障检测时间为 304 s 和 303 s, 未第一时间检测到缓变故障存在, 故障检



(a) 速度故障检测函数



(b) 位置故障检测函数

图 9　缓变故障下 2 种算法的故障检测函数

Fig. 9　Fault detection functions of two algorithms under slowly-varying fault

测存在时延；本文算法在速度和位置方面的首次故障检测时间均为 301 s，第一时间检测到缓变故障存在。由图 10、表 6 可知，在 301～600 s 缓变故障时段，残差卡方算法在速度和位置方面的平均漏警时长分别为 80.3 s 和 42.2 s，故障检测率分别为 73.2% 和 85.9%；本文算法在速度和位置方面的平均漏警时长分别为 37.7 s 和 12.4 s，故障检测率分别为 87.6% 和 95.7%；与残差卡方算法相比，本文算法的漏警时长大幅减少，故障检测率在速度和位置方面分别提高约 14.4% 和 9.8%，灵敏度和故障检测性能大幅度提高。

由表 2、表 4、表 6 可知，与残差卡方算法相比，本文算法在大阶跃故障下检测率提高的数值远小于小阶跃故障和缓变故障下的故障检测率提高的

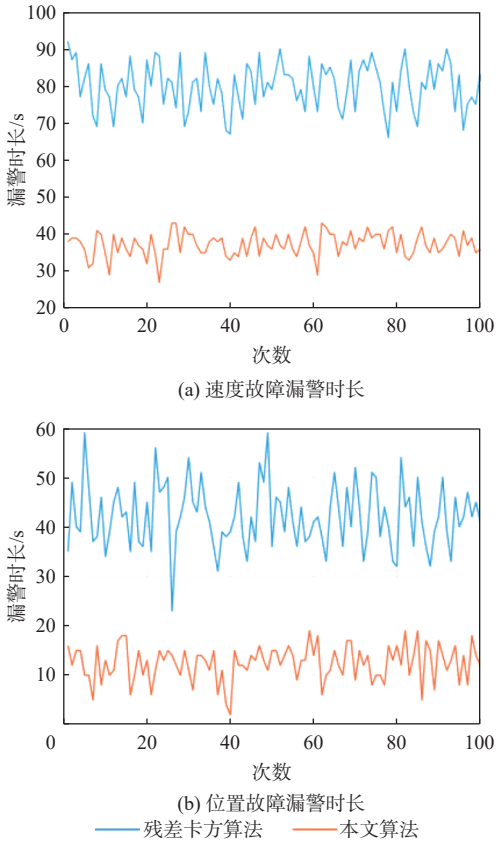


图 10 缓变故障下 2 种算法速度、位置故障的漏警时长

Fig. 10 Missed alarm duration of velocity, position fault of two algorithms under slowly-varying fault

数值，主要原因在于：残差卡方算法采用当前时刻的新息残差及其协方差构建故障检测函数，检测样本只有当前时刻的残差，对大阶跃故障具有较高的灵敏度，但对于小阶跃故障和初期误差数值较小的缓变故障灵敏度较差，无法有效识别，只有当故障增大到一定程度，才能做出正确反应。而本文算法采用新息残差和平滑有界层宽度设计了一种新的故障检测函数，函数构造中既利用当前时刻的先验新息残差及其协方差，又利用前一时刻的后验新息残差，扩大了检测样本，与残差卡方算法相比，对于小阶跃故障及缓变故障检测的灵敏度和时效性大大提高。

采用基于残差卡方故障检测的抗差自适应算法和本文提出的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法进行处理，得到 2 种算法的速度、位置误差曲线，如图 11 所示；2 种算法的速度、位置误差的统计结果如表 7 所示。

由图 5、图 8、图 11 可知，在 GNSS 典型故障时段，残差卡方算法和本文算法均能对检测的故障进行自适应调整，限制故障对系统精度的影响，但与残差卡方算法相比，本文算法对故障限制性更好，误差曲线更加舒缓，导航估计精度更高。当故障结束后，残差卡方算法和本文算法均能使误差曲线迅速收敛，提高了速度和位置估计精度和误差曲线的平稳性。

由图 11、表 7 可知，当 GNSS 发生小阶跃故障时，本文算法估计的速度误差和速度误差更小，得到的速度精度和位置精度更高，与残差卡方算法相比，速度精度和位置精度分别提升了约 20.7% 和 22.2%，提高了 GNSS/SINS 组合导航系统的精度和

表 6 缓变故障下 2 种算法的检测效果对比

Table 6 Comparison of two fault detection algorithms under slowly-varying fault

算法	故障时段/s	首次故障检测时间/s		平均漏警时长/s		故障检测率/%		故障检测率提高比例/%	
		速度	位置	速度	位置	速度	位置	速度	位置
残差卡方	301～600	304	303	80.3	42.2	73.2	85.9		
本文算法	301～600	301	301	37.7	12.4	87.6	95.7	14.4	9.8

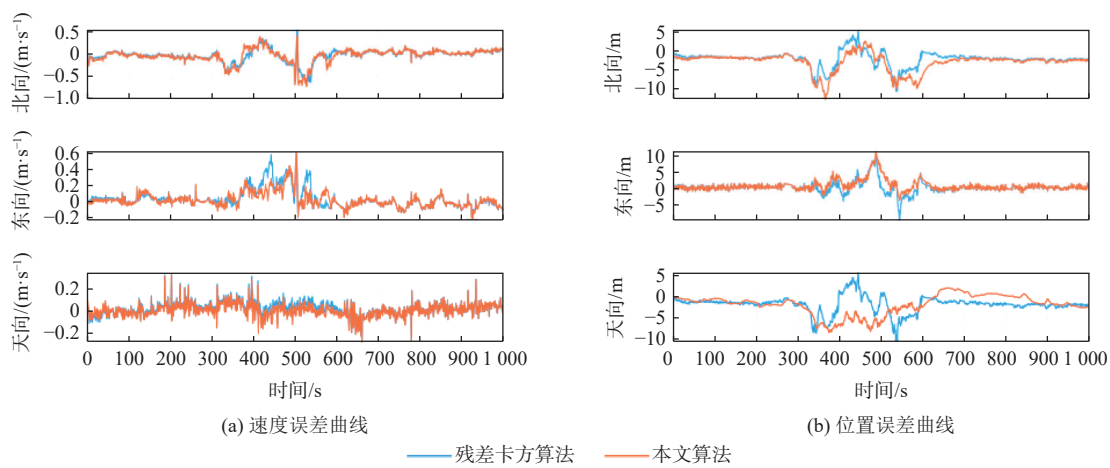


图 11 缓变故障下 2 种算法的速度、位置误差曲线

Fig. 11 Velocity and position error curves of two algorithms under slowly-varying fault

表 7　缓变故障下 2 种算法的速度、位置均方根误差

Table 7　Root mean square error of velocity and position of two algorithms under slowly-varying fault

算法	速度均方根误差/(m·s^{-1})			位置均方根误差/m			精度提高比例/%	
	北向	东向	天向	北向	东向	天向	速度	位置
残差卡方	0.267 9	0.207 9	0.081 7	5.727 3	3.264 2	8.184 4		
本文算法	0.224 9	0.150 0	0.066 9	4.351 4	3.052 1	5.221 7	20.7	22.2

鲁棒性。

5　结　论

针对 GNSS 信号易受到干扰发生故障, 导致滤波器性能降低, 影响 GNSS/SINS 组合导航系统输出精度问题, 本文提出了一种基于双参数的 GNSS/SINS 故障检测及抗差自适应算法。

1) 通过建立平滑有界层宽度模型, 得出平滑有界层宽度与模型所估计的系统噪声和量测噪声有着紧密的关系, 能够反映滤波器受到的干扰, 为组合导航故障检测函数的构造提供了新思路。

2) 通过与传统的基于残差卡方故障检测的抗差自适应算法相比, 利用新息残差和平滑有界层宽度构造的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法能够对 GNSS 典型故障做出快速反应, 提高速度故障和位置故障的检测率。

3) 利用新息残差和平滑有界层宽度构造的基于双参数的 GNSS/SINS 故障检测及抗差自适应算法与传统的基于残差卡方故障检测的抗差自适应算法相比, 组合导航系统的速度精度和位置精度提高 20% 以上, 且平稳性得到了一定的提升。

参考文献（References）

[1] 李增科, 孙耀文, 陈昭冰, 等. 顾及组合导航闭环反馈的变分贝叶斯自适应滤波优化算法[J]. 测绘学报, 2023, 52(5): 725-737.
LI Z K, SUN Y W, CHEN Z B, et al. Optimization of variational Bayesian-based adaptive filter for closed-loop feedback in integrated navigation[J]. Acta Geodaetica et Cartographica Sinica, 2023, 52(5): 725-737(in Chinese).

[2] YANG L, LI Y, WU Y L, et al. An enhanced MEMS-INS/GNSS integrated system with fault detection and exclusion capability for land vehicle navigation in urban areas[J]. GPS Solutions, 2014, 18(4): 593-603.

[3] ZHANG G H, XU P H, XU H S, et al. Prediction on the urban GNSS measurement uncertainty based on deep learning networks with long short-term memory[J]. IEEE Sensors Journal, 2021, 21(18): 20563-20577.

[4] SUN R, XU L T, XUE D B, et al. GPS signal reception classification using adaptive neuro-fuzzy inference system[J]. Journal of Navigation, 2019, 72(3): 685-701.

[5] THUMIGER N, DEGHAT M. A multi-agent deep reinforcement learning approach for practical decentralized UAV collision avoidance[J]. IEEE Control Systems Letters, 2021, 6: 2174-2179.

[6] 王鼎杰, 孟德利, 李朝阳, 等. 抗野值自适应卫星/微惯性组合导航方法[J]. 仪器仪表学报, 2017, 38(12): 2952-2958.
WANG D J, MENG D L, LI Z Y, et al. Adaptively outlier-restrained GNSS/MEMS-INS integrated navigation method[J]. Chinese Journal of Scientific Instrument, 2017, 38(12): 2952-2958(in Chinese).

[7] NARASIMHAPPA M, MAHINDRAKAR A D, GUIZILINI V C, et al. MEMS-based IMU drift minimization: sage husa adaptive robust Kalman filtering[J]. IEEE Sensors Journal, 2020, 20(1): 250-260.

[8] 刘子龙, 楼家祺. 基于正交性原理的组合导航系统故障检测[J]. 控制工程, 2024, 31(10): 1841-1848.
LIU Z L, LOU J Q. Fault detection of integrated navigation system based on orthogonality principle[J]. Control Engineering of China, 2024, 31(10): 1841-1848(in Chinese).

[9] 杨元喜, 任夏, 许艳. 自适应抗差滤波理论及应用的主要进展[J]. 导航定位学报, 2013, 1(1): 9-15.
YANG Y X, REN X, XU Y. Main progress of adaptively robust filter with applications in navigation[J]. Journal of Navigation and Positioning, 2013, 1(1): 9-15(in Chinese).

[10] 刘韬, 徐爱功, 隋心. 基于自适应抗差卡尔曼滤波的 UWB 室内定位[J]. 传感技术学报, 2018, 31(4): 567-572.
LIU T, XU A G, SUI X. Adaptive robust Kalman filtering for UWB indoor positioning[J]. Chinese Journal of Sensors and Actuators, 2018, 31(4): 567-572(in Chinese).

[11] 陈志高, 柳锐, 吴子豪, 等. 基于改进 Sage-Husa 自适应滤波的 ADCP 船速基准精确获取[J]. 哈尔滨工程大学学报, 2023, 44(10): 1732-1740.
CHEN Z G, LIU R, WU Z H, et al. Obtaining a reference velocity of ADCP measurement on the basis of an improved adaptive Sage-Husa filtering algorithm[J]. Journal of Harbin Engineering University, 2023, 44(10): 1732-1740(in Chinese).

[12] 严恭敏, 刘璠, 李梓阳, 等. 一种基于软卡方检测的自适应 Kalman 滤波方法[J]. 导航定位与授时, 2023, 10(4): 81-86.
YAN G M, LIU F, LI Z Y, et al. An adaptive Kalman filter based on soft Chi-square test[J]. Navigation Positioning and Timing, 2023, 10(4): 81-86(in Chinese).

[13] PRABHU G R, BHASHYAM S, GOPALAN A, et al. Sequential multi-hypothesis testing in multi-armed bandit problems: an approach for asymptotic optimality[J]. IEEE Transactions on Information Theory, 2022, 68(7): 4790-4817.

[14] 高广乐, 高社生, 彭旭, 等. 渐消 SPRT 在 SINS/CNS/SRS 导航系统软故障诊断中的应用[J]. 中国惯性技术学报, 2020, 28(6): 834-840.
GAO G L, GAO S S, PENG X, et al. Fading SPRT method for soft fault diagnosis in SINS/CNS/SRS integrated navigation system[J]. Journal of Chinese Inertial Technology, 2020, 28(6): 834-840(in

Chinese).

[15] HABIBI S. The smooth variable structure filter[J]. Proceedings of the IEEE, 2007, 95(5): 1026-1059.

[16] 施常勇, 张丽敏. 平滑变结构滤波在航天器组合导航中的应用[J]. 中国惯性技术学报, 2017, 25(4): 478-482.
SHI C Y, ZHANG L M. Application of SVSF in spacecraft integrated navigation system[J]. Journal of Chinese Inertial Technology, 2017, 25(4): 478-482(in Chinese).

[17] LI Y W, LI G, LIU Y, et al. A novel smooth variable structure filter for target tracking under model uncertainty[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 5823-5839.

[18] 赵宾, 曾庆化, 刘建业, 等. 基于平滑有界层的故障检测及抗差自适应滤波算法[J]. 中国惯性技术学报, 2023, 31(3): 245-253.
ZHAO B, ZENG Q H, LIU J Y, et al. Fault detection and robust adaptive filter algorithm based on smooth bounded layer[J]. Journal of Chinese Inertial Technology, 2023, 31(3): 245-253(in Chinese).

[19] 顾明星, 刘卫, 胡媛, 等. 陀螺罗经和计程仪辅助的 GNSS/SINS 松组合导航系统[J]. 航天控制, 2021, 39(1): 8-14.
GU M X, LIU W, HU Y, et al. A loosely coupled GNSS/SINS integrated navigation system assisted by gyroscope and vialog[J]. Aerospace Control, 2021, 39(1): 8-14(in Chinese).

[20] 张景伟. INS/GPS/SAR 组合导航系统关键问题研究[D]. 西安: 西北工业大学, 2003: 12-18.
ZHANG J W. Research on key problems of INS/GPS/SAR integrated navigation system[D]. Xi'an: Northwestern Polytechnical University, 2003: 12-18(in Chinese).

[21] 张超, 吕志伟, 柯晔, 等. 欺骗干扰影响下 GNSS/INS 松组合与紧组合比较[J]. 导航定位学报, 2022, 10(5): 18-24.
ZHANG C, LYU Z W, KE Y, et al. Comparison of GNSS/INS loosely coupled and tightly coupled systems under spoofing interference[J]. Journal of Navigation and Positioning, 2022, 10(5): 18-24 (in Chinese).

[22] 陈帅. 平滑变结构滤波算法研究及其在惯性导航系统初始对准中的应用[D]. 哈尔滨: 哈尔滨工程大学, 2017: 20-30.
CHEN S. Research on smoothing variable structure filtering algorithm and its application in initial alignment of inertial navigation system[D]. Harbin: Harbin Engineering University, 2017: 20-30 (in Chinese).

[23] 陈宇. 平滑变结构滤波方法及其在目标跟踪中的应用[D]. 西安: 西安电子科技大学, 2022: 18-23.
CHEN Y. Smooth variable structure filtering methods and its application in target tracking[D]. Xi'an: Xidian University, 2022: 18-23 (in Chinese).

[24] 胡晓梅, 潘新龙, 刘成铭. GNSS/SINS 组合导航的改进 Sage-Husa 自适应滤波算法[J]. 大地测量与地球动力学, 2023, 43(8): 791-794.
HU X M, PAN X L, LIU C M. An improved Sage-Husa adaptive filtering algorithm for GNSS/SINS integrated navigation[J]. Journal of Geodesy and Geodynamics, 2023, 43(8): 791-794(in Chinese).

[25] 刘正午, 孙蕊, 蒋磊. 基于伪距残差和新息的 GNSS/IMU 抗差自适应定位算法[J]. 北京航空航天大学学报, 2024, 50(4): 1316-1324.
LIU Z W, SUN R, JIANG L. Robust adaptive position algorithm for GNSS/IMU based on pseudorange residual and innovation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(4): 1316-1324(in Chinese).

[26] 苗岳旺, 周巍, 田亮, 等. 基于新息 χ^2 检测的扩展抗差卡尔曼滤波及其应用[J]. 武汉大学学报 (信息科学版), 2016, 41(2): 269-273.
MIAO Y W, ZHOU W, TIAN L, et al. Extended robust Kalman filter based on innovation Chi-square test algorithm and its application[J]. Geomatics and Information Science of Wuhan University, 2016, 41(2): 269-273(in Chinese).

A GNSS/SINS fault detection and robust adaptive algorithm based on two parameters

ZHAO Guiling*，WANG Jinbao，JIANG Zihao，GAO Shuai

(School of Geomatics，Liaoning Technical University，Fuxin 123000，China)

Abstract：The global navigation satellite system (GNSS) signal is susceptible to interference, which leads to a decrease in filter performance and affects the output accuracy of the global navigation satellite system/strapdown inertial navigation system (GNSS/SINS) integrated navigation system. For this problem, this paper proposes a GNSS/SINS fault detection and robust adaptive algorithm based on two parameters. For precise fault detection, the algorithm creates a fault detection function based on the breadth of the smooth bounded layer and assesses the system measurement data using the innovation residual. Using fault detection function values to construct the robust cofactor matrix for real-time error correction to improve the accuracy and robustness of state estimation. The experimental findings of two common GNSS faults, step fault and slow fault, demonstrate that: the robust adaptive algorithm and the GNSS/SINS fault detection method based on two parameters are contrasted with the conventional robust adaptive algorithm based on residual chi-square fault detection techniques. The velocity fault detection rate is increased by 7.6%−23.2%, and the position fault detection rate is increased by 3.2%−12.3%. The velocity accuracy is increased by 20.7%−27.1%, and the position accuracy is increased by 22.2%−34.6%. The algorithm effectively improves the accuracy and robustness of the GNSS/SINS integrated navigation system.

Keywords：integrated navigation；fault detection；robust adaptive algorithm；innovation residual；width of smooth bounded layer

Received：2023-12-18；**Accepted**：2024-02-23；**Published Online**：2024-03-11 17：09

URL：link.cnki.net/urlid/11.2625.V.20240311.0959.001

Foundation items：Natural Science Foundation of Liaoning Province (2020-MS-303); Basic Research Projects of the Education Department of Liaoning Province, China (JYTMS20230807)

* **Corresponding author**. E-mail：zhaoguiling@lntu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0830

增强上下文特征交互的实时无人机影像分割算法

李云红*，张富星，苏雪平，李丽敏，王梅，梁成名
(西安工程大学 电子信息学院，西安 710048)

摘 要：针对无人机影像语义分割任务中轻量级算法缺乏全局信息交互导致分割结果中目标漏检与不完整问题，提出了一种增强上下文特征交互的实时无人机影像分割算法。算法采用双分支结构，利用不同方向的全局平均池化对通道和空间信息进行编码，保留了精准的位置信息，并增强了对图像中局部细节信息的关注；利用位置感知循环卷积和空间加权构建全局感知提取模块，实现了全局上下文信息捕获；对不同尺度特征采用加权操作进行融合，降低了融合过程中的信息损失与算法的计算量。在UAVid和AeroScapes数据集上对所提算法进行验证，结果显示：平均交并比(mIoU)分别达到66.5%和63.0%，相比BiSeNet V2提升了2.6%和2.2%，分割速度分别达到79.9帧/s和71.4帧/s，相比BiSeNet V2提升了8.3帧/s和6.9帧/s，在保证实时分割速度的同时取得了较好的分割精度。

关 键 词：语义分割；无人机影像；卷积神经网络；上下文信息；特征融合
中图分类号：TP391.41
文献标志码：A **文章编号**：1001-5965(2026)03-0668-10

随着科技的快速发展，无人机技术已经在许多领域引起了广泛的关注和应用。无人机影像语义分割[1-2]是将计算机视觉和深度学习领域的最新技术与无人机相结合的典型示范，该技术具有无限的潜力，可应用于农业、城市规划、自然资源管理、环境监测和军事等各个领域。

无人机影像具有分辨率高、细节信息与小目标较多、场景繁杂等特点，同时无人机有较低的算力[3]。因此，如何在实现模型轻量化的同时，检测到更多的细节信息，是目前面临的问题。自2015年以来，各种轻量化的语义分割网络层出不穷[4-6]。U-Net[7]网络的提出促进了医学领域分割的发展，其具有结构简单、参数量小的特点，目前依然是医学分割领域的主流网络。为增大模型感受野，Deeplab[8]系列网络运用空洞卷积来扩大感受野，并提出空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)结构，并行使用多个不同空洞率的空洞卷积，以同时捕获图像中多个不同尺度的信息，虽然空洞卷积可以在一定程度上增大模型感受野，但不能实现全局空间信息交互。高效神经网络(efficient neural network, ENet)[9]是基于SegNet[10]改进的实时分割的轻量化模型，其降低了输入特征图的尺寸，将$n\times n$的卷积核拆分为$n\times 1$和$1\times n$的卷积核，从而有效减少了参数量和计算量。图像级联网络(image cascade network, ICNet)[11]将不同分辨率的图像经过不同深度的网络，分辨率高的图像经过较浅的网络结构，分辨率较低的图像经过较深的网络结构，达到了降低计算量的目的。上下文网格网络(context grided network, CGNet)[12]提出了CG模块，该模块中先后提取局部特征和联合特征，并用上下

收稿日期：2023-12-22；**录用日期**：2024-03-29；**网络出版时间**：2024-04-18 17:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20240417.1132.001
基金项目：国家自然科学基金(62203344)；陕西省自然科学基础研究计划重点项目(2022JZ-35)；陕西高校青年创新团队；西安市“科学家+工程师”队伍建设项目(25KGYB00029)
*通信作者. E-mail：hitliyunhong@163.com

引用格式：李云红，张富星，苏雪平，等. 增强上下文特征交互的实时无人机影像分割算法[J]. 北京航空航天大学学报，2026，52(3)：668-677.
LI Y H，ZHANG F X，SU X P，et al. Real-time UAV image segmentation algorithm with enhanced contextual feature interaction[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：668-677 (in Chinese).

文信息来指导联合特征，从而达到突出更加重要部分的作用。双边分割网络（bilateral segmentation network, BiSeNet）[13] 采用双分支结构，分为细节分支与语义分支，细节分支能够为语义分支提供缺失的细节信息。根据诸多研究经验来看，应用多个支路获取不同尺度的特征可以增强算法效果[14]。BiSeNet V2[15] 进一步改进了 BiSeNet，采用新的融合方式，在降低参数量的同时，提升分割速度。虽然以上网络模型在实时性方面取得了进展，但普通卷积的局部感受野在一定程度上限制了网络的分割性能，导致对无人机影像中目标物体分割效果差，容易出现漏检等情况。针对该问题，Yi 等[16] 提出了 UAVformer，采用 Transformer 模块作为基础特征提取模块，增强对无人机影像中小目标与边界的分割，但 Transformer 较大的计算量导致分割速度较慢，并不适于实时分割任务。

为在保证实时分割速度的同时，增强对无人机影像中目标的分割，本文进一步改进 BiSeNet V2，提出了增强上下文特征交互的实时无人机影像分割算法，在保证较快分割速度的同时，增强了对图像上下文信息的提取，有效提高了分割精度，在小目标物体及其他细节方面均取得了较好的分割效果。

1　原理与算法

算法主要分为 3 个部分，即主干网络、特征融合和上采样，整体结构如图 1 所示。主干网络由 2 个分支组成，分别为细节分支与语义分支。细节分支采用 3 层 VGG 网络，进行 1/8 倍下采样，尽可能减少计算量。语义分支在 1/8、1/16、1/32 下采样之前加入特征增强模块（feature enhancement module, FEM），有效实现了空间和通道特征的关注，增强了特征的表征能力。此外，在 1/16、1/32 下采样之后采用全局感知提取（global perception extraction, GPE）模块来提取特征，替换了原网络中不进行下采样的聚集和展开（gather-and-expansion, GE）层，扩大了特征提取时的感受野，其中，1/32 下采样之后通过使用 3 个 GPE 模块，能够在保证模型计算量适量的同时，结合全局信息提取更加复杂和抽象的深层次语义特征，进一步提高模型对图像中复杂场景的理解，并增强对图像中小目标的分割。特征融合部分使用重新设计的特征融合模块（feature fusion module,FFM），能够在较少参数的情况下实现 2 个分支信息的融合并减少信息损失。上采样部分为图 1 中下部绿色虚框内及右侧的黄色框分割头，其采用原网络结构，作用是将特征图恢复原始大小。

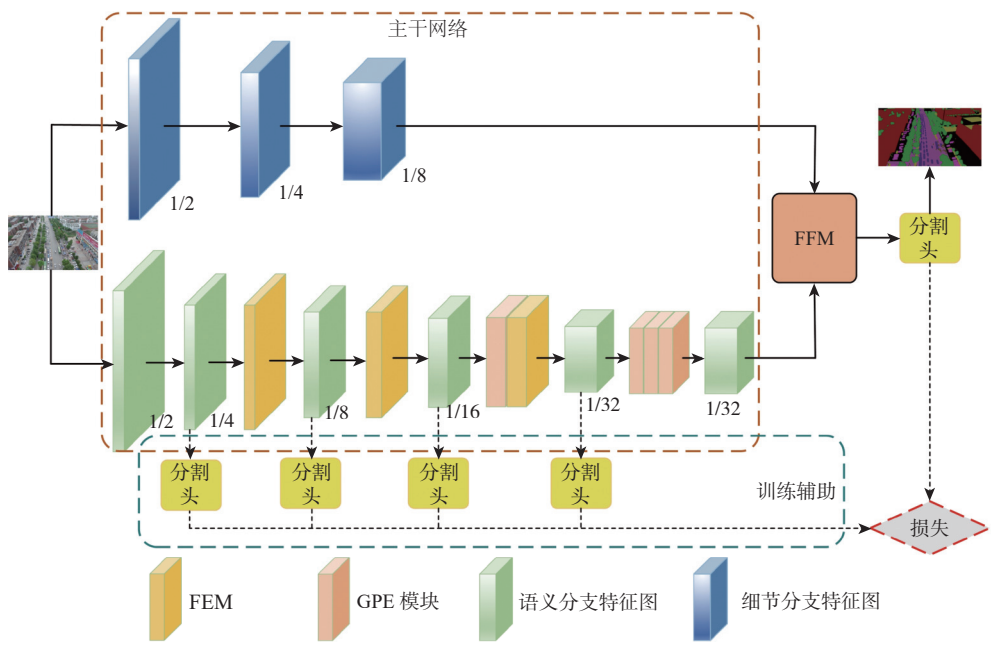


图 1　模型整体结构

Fig. 1　Overall structure of model

1.1　特征增强模块

为进一步增强语义分支对无人机图像中细节特征的提取，提出了 FEM，其详细结构如图 2 所示。该模块能够将精确的位置信息嵌入到通道中，并在空间上捕捉长距离的相互作用，增强特征表征能力，同时对特征通道进行加权，实现了对重要特征的关注。总的来说，FEM 能够在减少计算量的同时，增强模型对图像中关键信息的提取能力。

在 F_1 部分中，设输入张量为 $\boldsymbol{X} \in \mathbf{R}^{C\times H\times W}$（$C$ 为特征图的通道数，H 为特征图的高，W 为特征图的

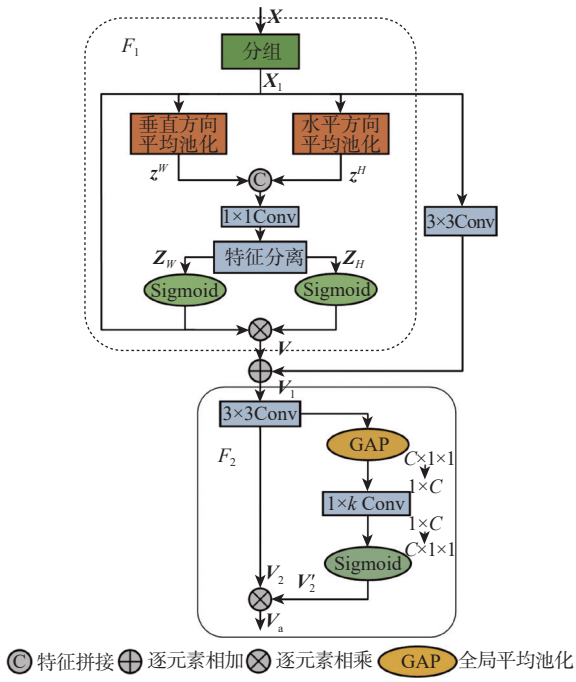


图2 特征增强模块

Fig. 2 Feature enhancement module

宽)，为提高模型的表征能力和泛化能力，同时减少计算量，$\boldsymbol{X}$经过通道分组，得到输出$\boldsymbol{X}_1 \in \mathbf{R}^{C/G\times H\times W}$（$G$为要分的组数)，分别经过垂直方向平均池化和水平方向平均池化，二者原理可以表示为

$$z_a^W(W)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_a(j,W) \tag{1}$$

$$z_a^H(H)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_a(H,i) \tag{2}$$

式中：x_a为第a个通道处的输入特征。

经过垂直方向平均池化和水平方向平均池化后，分别捕获了特征图在垂直和水平方向上的长程依赖关系，将所得到的特征进行拼接得到$z_c \in \mathbf{R}^{C/G\times 1\times (H+W)}$，并经过一个1×1卷积操作，1×1内核使模型能够捕捉局部跨通道交互特征[17]，并在通道之间共享相似性。将1×1卷积的输出分解为2个并行的1D特征编码向量[18]，记为$\boldsymbol{Z}_W$、$\boldsymbol{Z}_H$，并在每个并行路径中分别加一个非线性Sigmoid函数，与$\boldsymbol{X}_1$相乘得到输出结果$\boldsymbol{V}$。以上过程表示如下：

$$z_c=z^W\|z^H \tag{3}$$

$$\boldsymbol{Z}_W,\boldsymbol{Z}_H=F_{\text{split}}(\text{Conv}_{1\times 1}(z_c)) \tag{4}$$

$$\boldsymbol{V}=\boldsymbol{X}_1\otimes F_{\text{permute}}(\delta(\boldsymbol{Z}_W))\otimes\delta(\boldsymbol{Z}_H) \tag{5}$$

式中：“$\|$”表示特征拼接；$\text{Conv}_{1\times 1}$表示卷积核为1×1大小的卷积运算；F_{split}表示特征分解；δ表示Sigmoid函数；F_{permute}表示维度位置调整；“$\otimes$”表示元素级相乘。

虽然F_1部分沿不同维度方向在空间上捕获了长距离的相互作用，但并未考虑整个空间位置之间的相互作用，因此，增加一个分支，只用一个3×3大小的卷积层，用来实现在整个空间上信息的相互作用，将这一分支的输出与F_1部分的输出进行相加融合，得到输出$\boldsymbol{V}_1$。

$\boldsymbol{V}_1$作为F_2部分的输入，经过一个3×3的卷积得到$\boldsymbol{V}_2$，$\boldsymbol{V}_2$经过一个全局平均池化来聚合空间信息，对张量的维度进行调整，转变为1×C大小，经过一个1×k大小的一维卷积获得不同通道的权重，经过Sigmoid函数获得权重占比$\boldsymbol{V}_2'$，$\boldsymbol{V}_2'$与$\boldsymbol{V}_2$相乘得到最终的输出$\boldsymbol{V}_a$。上述过程表示如下：

$$\boldsymbol{V}_2=\text{Conv}_{3\times 3}(\boldsymbol{V}_1) \tag{6}$$

$$\boldsymbol{V}_2'=\delta(T_{C\times 1\times 1}(\text{Conv}_{1\times k}(T_{1\times C}(F_{\text{GAP}}(\boldsymbol{V}_2))))) \tag{7}$$

$$\boldsymbol{V}_a=\boldsymbol{V}_2'\otimes\boldsymbol{V}_2 \tag{8}$$

式中：$\text{Conv}_{3\times 3}$表示卷积核大小为3×3的卷积操作；$\text{Conv}_{1\times k}$表示卷积核大小为1×k的卷积操作；F_{GAP}表示全局平均池化；$T_{1\times C}$表示将张量调整为1×C大小；$T_{C\times 1\times 1}$表示将张量调整为C×1×1大小。

1.2 全局感知提取模块

由于获取的图像语义信息丰富与否非常依赖提取特征时的感受野，为此，采用空洞卷积、大的卷积核或叠加卷积层[19]，虽然可在一定程度上增加感受野，但不能实现全局信息的交互。针对该问题，提出了GPE模块，详细结构如图3所示。GPE模

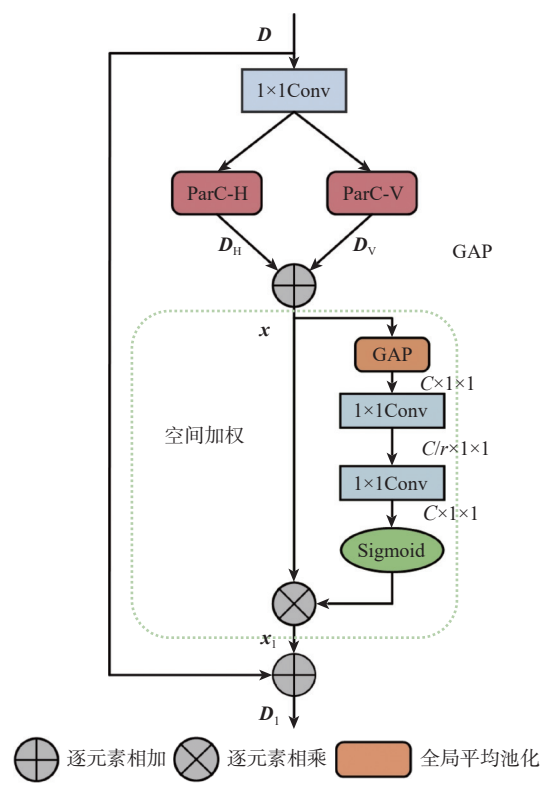


图3 全局感知提取模块

Fig. 3 Global perception extraction block

块采用位置感知循环卷积(position aware circular convolution, ParC)[20]来进行特征提取，最大限度提升感受野，实现了全局空间信息交互。此外，设计了空间加权模块，给予特征不同通道不同的权重，实现了特征通道之间的信息交互。空间加权操作相较 1×1 卷积有更少的参数量和计算量，在一定程度上降低了算法复杂度。

将 $\boldsymbol{D}\in\mathbf{R}^{C\times H\times W}$ 作为模块的输入，位置感知循环卷积有 2 种类型，即水平方向和垂直方向。在 GPE 模块中，$\boldsymbol{D}$经过一个 1×1 卷积操作进行通道信息交互，分别通过水平方向的位置感知循环卷积(ParC-H)和垂直方向的位置感知循环卷积(ParC-V)，来获得 2 个方向的全局信息交互特征，将 2 个方向获得的特征进行相加，得到融合特征 $\boldsymbol{x}$。上述过程表示如下：

$$\boldsymbol{D}_{\mathrm{H}}=F_{\text{ParC-H}}(\mathrm{Conv}_{1\times1}(\boldsymbol{D})) \tag{9}$$

$$\boldsymbol{D}_{\mathrm{V}}=F_{\text{ParC-V}}(\mathrm{Conv}_{1\times1}(\boldsymbol{D})) \tag{10}$$

$$\boldsymbol{x}=\boldsymbol{D}_{\mathrm{H}}\oplus\boldsymbol{D}_{\mathrm{V}} \tag{11}$$

式中：$F_{\text{ParC-H}}$和$F_{\text{ParC-V}}$分别表示水平方向的位置感知循环卷积和垂直方向的位置感知循环卷积；“⊕”表示元素级相加。

得到融合特征 $\boldsymbol{x}$后，经过空间加权模块进行通道信息的交互。首先，$\boldsymbol{x}$通过一个全局平均池化来对空间信息进行压缩，获得 C×1×1 大小的特征图；然后，经过 2 个 1×1 卷积层来降低信息冗余和引入非线性变换，从而增强模型的表示能力，获得所需权值大小，获得的权值经过 Sigmoid 函数得到权值所占比例，与$\boldsymbol{x}$相乘得到$\boldsymbol{x}_1$，实现不同通道不同的占比，从而达到通道信息交互的目的；最后，$\boldsymbol{x}_1$与$\boldsymbol{D}$相加得到最终的输出$\boldsymbol{D}_1$。上述过程表示如下：

$$\boldsymbol{x}_1=\delta(\mathrm{Conv}_{1\times1}(\mathrm{Conv}_{1\times1}(F_{\mathrm{GAP}}(\boldsymbol{x}))))\otimes\boldsymbol{x} \tag{12}$$

$$\boldsymbol{D}_1=\boldsymbol{x}_1\oplus\boldsymbol{D} \tag{13}$$

1.3　特征融合模块

图 4 给出了原 FFM 和改进 FFM 对比。图中：DWConv 表示深度可分离卷积。原模型的特征融合方式中采用了较多的 3×3 卷积，一定程度上增加了较多参数量和计算量，此外，3×3 卷积对于细节特征的处理损失了少部分的细节信息，语义特征与细节特征的相乘并不能提供足够的语义理解，因此，新的 FFM 对于细节分支的处理并未采用卷积操作(见图 4(b))，保留了完整的细节信息并降低了参数量。同时，在特征融合过程中采用拼接交互加权的方式，不仅能够降低参数量和计算量，而且利用交互之后的特征来指导细节特征和语义特征，可以加强模型对细节信息和语义信息的理解，增强融合效果。通过计算，原 FFM 参数量为 0.48×10^6，新 FFM 参数量为 0.40×10^6，降低了 0.08×10^6。

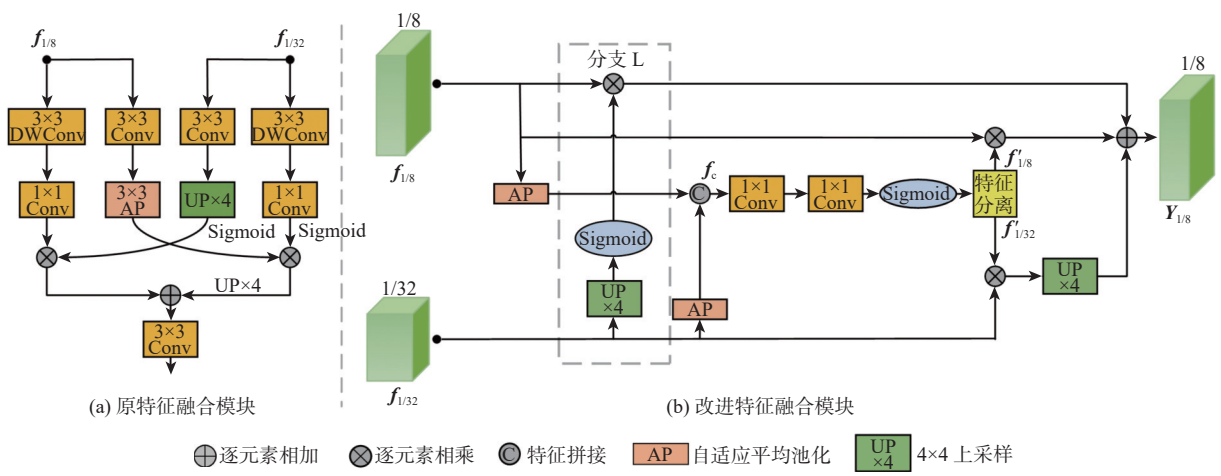


图 4　原特征融合模块和改进特征融合模块的对比

Fig. 4　Comparison of original feature fusion module and improved feature fusion module

图 4(b)中，将细节分支和语义分支所得到的 1/8 和 1/32 的特征图经过自适应平均池化调整到 1/16 大小，并进行通道方向的拼接，经过 2 个 1×1 卷积层来实现 2 个分支信息的融合，再通过 Sigmoid 函数和沿通道维度的特征分解操作，得到 2 部分的特征权重$\boldsymbol{f}'_{1/8}$、$\boldsymbol{f}'_{1/32}$。将 2 部分权重分别与原始特征图相乘，得到信息交互后的特征，将二者进行相加融合得到$\boldsymbol{y}$。设 2 个输入分别为$\boldsymbol{f}_{1/8}$、$\boldsymbol{f}_{1/32}$，上述过程表示如下：

$$\boldsymbol{f}_{\mathrm{c}}=F_{\mathrm{AP}}(\boldsymbol{f}_{1/8})\|F_{\mathrm{AP}}(\boldsymbol{f}_{1/32}) \tag{14}$$

$$\boldsymbol{f}'_{1/8},\boldsymbol{f}'_{1/32}=F_{\mathrm{split}}(\delta(\mathrm{Conv}_{1\times1}(\mathrm{Conv}_{1\times1}(\boldsymbol{f}_{\mathrm{c}})))) \tag{15}$$

$$\boldsymbol{y}=(\boldsymbol{f}'_{1/8}\otimes\boldsymbol{f}_{1/8})\oplus F_{\mathrm{bilinear}}(\boldsymbol{f}'_{1/32}\otimes\boldsymbol{f}_{1/32}) \tag{16}$$

式中：F_{AP}表示自适应平均池化；F_{bilinear}表示双线性上采样。

为进一步增强细节特征的检测，增加一条支路L，将语义分支所得到的1/32倍的特征图$\boldsymbol{f}_{1/32}$进行4倍上采样，经过Sigmoid函数后与细节分支输出相乘，得到这一分支的输出，其保留了大部分的细节信息，能够很好地为输出$\boldsymbol{y}$补充缺失的细节信息。将这一输出与$\boldsymbol{y}$进行相加融合，得到融合模块最终的输出$\boldsymbol{Y}_{1/8}$。

2 实验结果与分析

2.1 数据集

实验采用2个无人机影像数据集来验证本文算法的有效性，分别为UAVid和AeroScapes数据集。

UAVid数据集是针对城市场景语义分割任务的无人机遥感图像语义分割数据集，包含42个序列，每个序列包括10张图像，共420张图像，其中，20个训练集序列，7个验证集序列，15个测试集序列。即训练集200张图像，验证集70张图像。图像分辨率为3 840像素×2 160像素。数据集包含8种地物类别，即建筑、公路、静态车、树、低植被、人、移动车、杂乱背景。

AeroScapes数据集中的数据是使用商业无人机从5～50 m高度范围内拍摄的，包含3 269张1 280像素×720像素的真实图像和语义标签。其中，2 321张图像为训练集，648张图像为验证集，300张图像为测试集。数据集场景包括农田、湖泊、桥梁等不同场景，其中的对象包含行人、自行车、汽车、无人机、动物、障碍物、船、建筑、植被、道路、天空等11个类别。

2.2 实验条件

实验基于PyTorch1.11.0深度学习框架和CUDA-11.3库，Python版本为3.8，在Linux操作系统上使用一块NVIDIA GeForce GTX 3090进行模型训练，使用一块NVIDIA GeForce GTX 1080Ti进行模型验证。训练中采用AdamW优化器进行优化，BatchSize为8，初始学习率为0.001，权重衰减为0.000 1，Epoch为300。训练过程中采用数据增强，包括随机水平翻转和随机垂直翻转，并将图像随机裁剪为512像素×1 024像素大小，此外，还采用了随机缩放，缩放量在0.75～2之间。训练损失函数采用交叉熵损失函数。

2.3 性能评估指标

采用平均交并比（mean intersection over union，mIoU）来衡量算法预测精度，浮点运算速度来评估算法的复杂度，分割速度来评估所提模型在设备上的计算速度。

mIoU计算全部类别真实值与预测值交集和并集之比的平均值，计算公式为

$$M_{\text{iou}}=\frac{1}{m+1}\sum_{i=0}^{m}\left(\frac{p_{ii}}{\sum_{j=0}^{m}p_{ij}+\sum_{j=0}^{m}p_{ji}-p_{ii}}\right) \tag{17}$$

式中：m为像素的种类数量；p_{ij}表示真实种类为i类却预测为j类的像素数量；p_{ii}表示真实种类为i类预测也是i类的像素数量。

分割速度计算公式为

$$f_{\text{FPS}}=N\Big/\sum_{j=1}^{N}T_j \tag{18}$$

式中：N为图像序列的帧数；T_j为处理j帧图像的时间。

2.4 消融实验

2.4.1 在UAVid数据集上的消融实验

在UAVid数据集上对模块进行不同组合的功能验证，结果如表1所示，其中，Baseline表示BiSeNet V2。可以看出，每个模块都给Baseline模型带来了一定程度上mIoU的增长，所提3个模块整体加入后，mIoU达到66.5%，相较Baseline模型提升

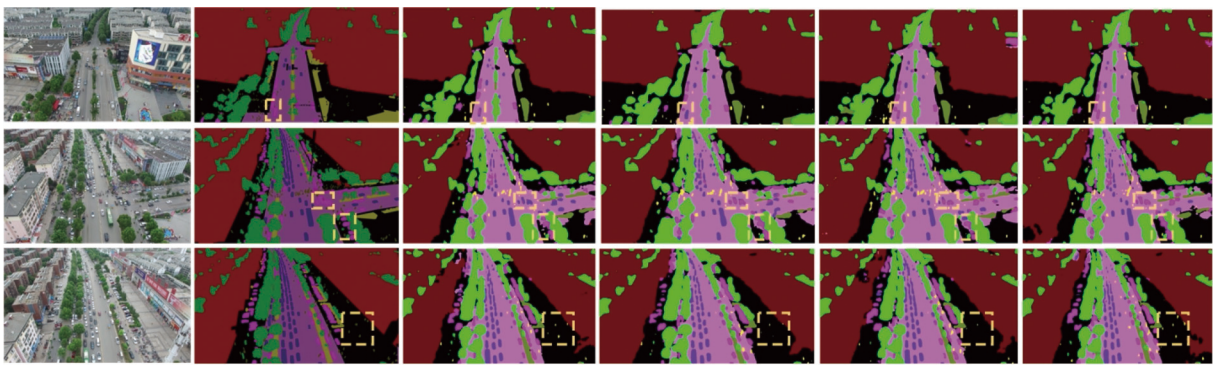
表1 UAVid数据集上各模块消融实验数据

Table 1 Data from ablation experiment of modules on UAVid dataset

Baseline	GPE模块	FEM	FFM	Swin-Block	mIoU/%	浮点运算速度/10^9 s^{-1}	分割速度/(帧·s^{-1})
√					63.9	14.78	71.6
√	√				64.8	12.82	69.0
√		√			65.2	11.58	74.8
√			√		64.4	10.24	82.2
√	√	√			65.7	11.79	74.1
√	√		√		65.0	12.50	74.2
√		√	√		65.4	11.24	81.7
√	√	√	√		66.5	12.93	79.9
√				√	65.3	21.65	38.3
√		√		√	66.2	20.07	42.1
√		√	√	√	66.9	19.84	45.6

了 2.6%。GPE 模块单独加入模型后, 分割速度相较 Baseline 降低了 2.6 帧/s, 这是由于位置感知循环卷积相比普通卷积参数量有一定的增长。为解决该问题, FEM 和 FFM 均极大限度地进行了参数量和计算量的缩减, 同时, 为更进一步降低模型整体计算量, 算法将原始分割头中提取特征的数量减少了一半, 降低了计算冗余。从表 1 中可以看出, FEM 和 FFM 较好地实现了降低计算量的同时提升模型性能的目的。最终模型相较 Baseline, 在浮点运算速度上降低了 1.85×10^9 s^{-1}, 在分割速度上提升了 8.3 帧/s。此外, 表 1 中还展示了在 GPE 模块相同位置加入 Swin-Block[21] 的实验结果, 虽然 mIoU 最高达到了 66.9%, 相较 GPE 模块高 0.4%, 但在浮点运算速度上增加了 6.91×10^9 s^{-1}, 在分割速度上降低了 34.3 帧/s, 在计算量与速度方面存在一定的不足。

图 5 给出了 UAVid 数据集上消融实验视觉对比结果, 其中, 对部分差异进行了标注。可以看出, 加入所提模块后, 增强了模型对细节和小目标的分割。例如, 从第 1 行和第 2 行分割结果来看, 加入模块后, 模型能够精准地分割出车辆的类型, 并且保持了车辆的完整性; 分割结果中, 人用黄色来表示, 从第 3 行分割结果中能够看出, 加入 3 个模块后分割出了更多的人。实验结果表明了所提模块的有效性。



(a) 原始图像　(b) 真实标签　(c) BiSeNet V2　(d) BiSeNet V2+GPE　(e) BiSeNet V2+GPE+FEM　(f) BiSeNet V2+GPE+FEM+FFM

图 5　UAVid 数据集上消融实验视觉对比结果

Fig. 5　Visual comparison results of ablation experiments on UAVid dataset

2.4.2　在 AeroScapes 数据集上的消融实验

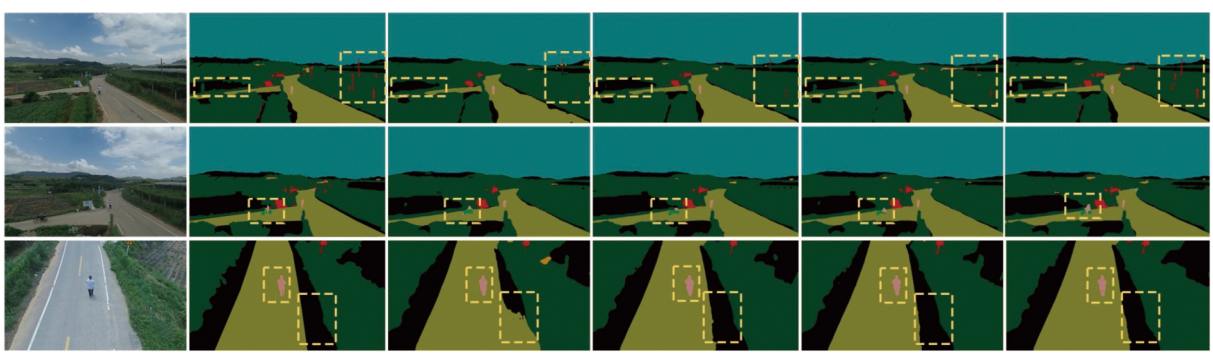
在 AeroScapes 数据集上对模块进行不同组合的功能验证, 结果如表 2 所示, 其中, Baseline 表示 BiSeNet V2。可以看出, 在 AeroScapes 数据集上, 各模块依然展示了其有效性, 模块单独或不同组合加入 Baseline 都给 mIoU 带来了一定程度的增长, 同时加入 3 个模块后, mIoU 达到 63.0%, 分割速度达到 71.4 帧/s, 相较 Baseline 分别提升了 2.2% 和 6.9 帧/s。

图 6 展示了在 AeroScapes 数据集上消融实验视觉对比结果。从结果可知, 改进模型增强了对小目标

表 2　AeroScapes 数据集上各模块消融实验数据

Table 2　Data from ablation experiments of modules on AeroScapes dataset

Baseline	GPE模块	FEM	FFM	mIoU/%	分割速度/(帧·s^{-1})
√				60.8	64.5
√	√			61.1	59.2
√		√		61.8	66.2
√			√	61.3	77.4
√	√	√		61.9	64.6
√	√		√	62.5	66.8
√		√	√	62.1	72.5
√	√	√	√	63.0	71.4



(a) 原始图像　(b) 真实标签　(c) BiSeNet V2　(d) BiSeNet V2+GPE　(e) BiSeNet V2+GPE+FEM　(f) BiSeNet V2+GPE+FEM+FFM

图 6　AeroScapes 数据集上消融实验视觉对比结果

Fig. 6　Visual comparison results of ablation experiments on AeroScape dataset

与细节的检测。例如,在第 1 行结果中,改进模型将密集小障碍物较为清晰地分割出来,从第 2 行和第 3 行结果来看,改进模型分割结果中,人物边界更为清晰,形状更为完整,更加接近真实标签;在道路与草丛的交界处也能够较好地分割出道路。结果证明了本文中各模块的泛化性,有效提升了模型的分割精度与分割速度。

2.5 对比实验

2.5.1 在 UAVid 数据集上的结果对比

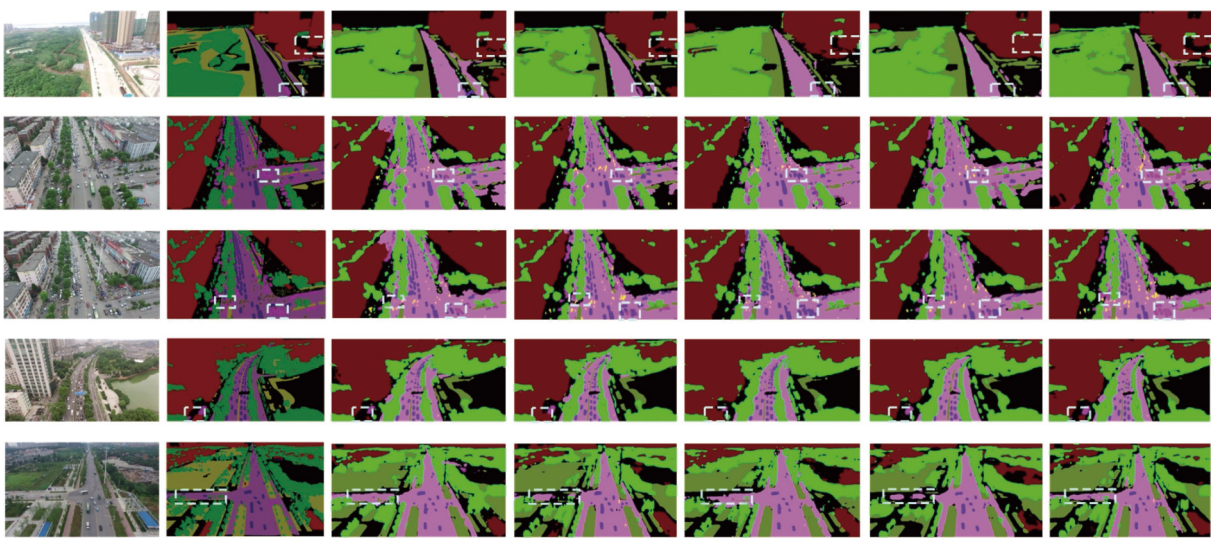
为验证本文算法,在 UAVid 数据集上与其他模型进行对比,如 SegNet[10]、U-Net[7]、SwiftNet[17]、ICNet[11]、Deeplabv3+[8]、BiSeNet[13]、SFNet[22]、BiSeNet V2[15]、HyperSeg-S[23]、PIDNet[24] 等,对比结果如表 3 所示。可以看出,本文算法在 mIoU 上达到了最高值 66.5%,均优于其他模型;在模型复杂度上,浮点运算速度达到了 12.93×10^9 s^{-1},低于大部分模型,虽然比 BiSeNet[13]、HyperSeg-S[23]、PIDNet[24] 略高,但在 mIoU 和分割速度上,都占据较大优势;在分割速度方面,达到了 79.9 帧/s,均在一定程度上高于其他模型。总的来看,相较近年来的算法,本文算法在分割精度、模型复杂度和分割速度上均取得了一定的优势,综合性能优越,在一定程度上实现了模型分割精度和复杂度的平衡。

表 3 不同算法在 UAVid 数据集上实验数据比较

Table 3 Comparison of experimental data of different algorithms on UAVid dataset

模型	Backbone	mIoU/%	浮点运算速度/10^9 s^{-1}	分割速度/(帧·s^{-1})
SegNet[10]		55.3	112.63	11.5
U-Net[7]		58.4	64.01	12.6
SwiftNet[17]	Resnet18	62.1	13.42	48.9
ICNet[11]	PSPNet50	61.9	25.91	73.2
Deeplabv3+[8]	MobileNetV2	61.4	20.47	71.2
BiSeNet[13]	Resnet18	61.8	12.64	74.1
SFNet[22]	Resnet18	65.3	18.52	62.4
BiSeNet V2[15]		63.9	14.78	71.6
HyperSeg-S[23]	EfficientNet-B1	64.3	10.12	27.5
PIDNet[24]		62.0	5.07	74.8
本文		66.5	12.93	79.9

图 7 给出了部分模型分割结果可视化对比,其中,人用黄色来表示,图中对部分差异进行了标注。可以看出,相较其他模型,本文模型的分割结果图更接近真实标签。此外,小目标(如车和人)的分割更为完整,轮廓更为清晰,漏检问题也得到了改善。在一些其他连接部分,如在一些道路与背景连接处,能够很好地区别出道路与背景,保证了道路分割的完整性。



(a) 原始图像 (b) 真实标签 (c) SFNet[22] (d) BiSeNet V2[15] (e) HyperSeg-S[23] (f) PIDNet[24] (g) 本文

图 7 部分算法在 UAVid 数据集上预测结果比较

Fig. 7 Comparison of prediction results of some selected algorithms on UAVid dataset

2.5.2　在 AeroScapes 数据集上的结果对比

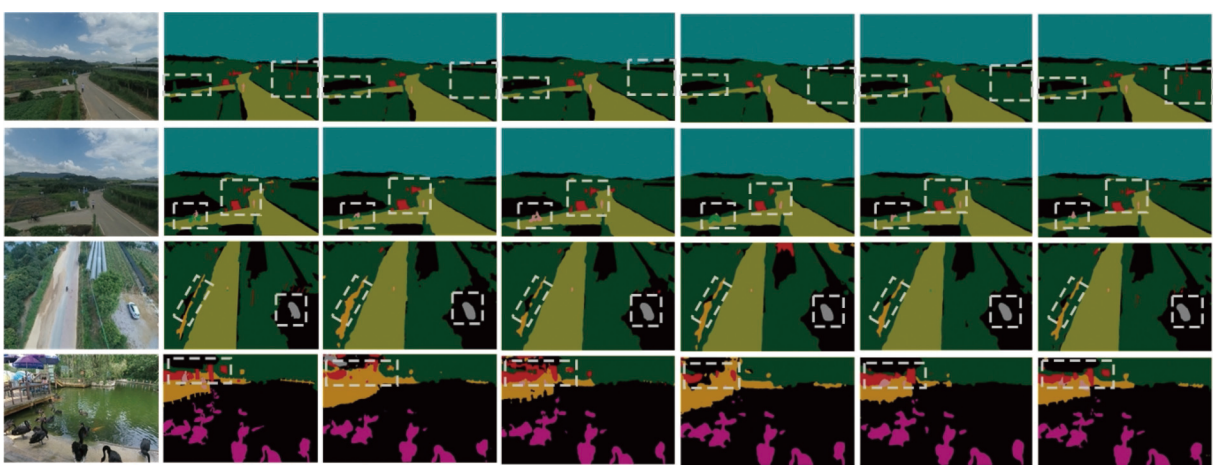
为进一步验证本文算法的有效性，在 AeroScapes 数据集上与 SegNet[10]、U-Net[7]、SwiftNet[17]、ICNet[11]、Deeplabv3+[8]、BiSeNet[13]、SFNet[22]、BiSeNet V2[15]、HyperSeg-S[23]、PIDNet[24] 等模型进行了对比，结果如表 4 所示。可以看出，本文算法在 mIoU 上达到了 63.0%，在分割速度上达到了 71.4 帧/s，都处于最高值，虽然在参数量上略高于 Deeplabv3+[8] 和 BiSeNet V2[15]，但在 mIoU 和检测速度都有较大优势，较好地平衡了模型的参数量和分割精度。

表 4　不同算法在 AeroScapes 数据集上实验数据比较

Table 4　Comparison of experimental data of different algorithms on AeroScapes dataset

模型	Backbone	参数量	mIoU/%	分割速度/(帧·s^{-1})
SegNet[10]		15.62×10^6	50.6	11.8
U-Net[7]		17.26×10^6	51.4	11.1
SwiftNet[17]	Resnet18	12.07×10^6	60.1	51.1
ICNet[11]	PSPNet50	23.15×10^6	59.5	67.2
Deeplabv3+[8]	MobileNetV2	5.82×10^6	59.9	61.8
BiSeNet[13]	Resnet18	13.62×10^6	58.2	66.9
SFNet[22]	Resnet18	12.72×10^6	62.1	64.0
BiSeNet V2[15]		5.21×10^6	60.8	64.5
HyperSeg-S[23]	EfficientNet-B1	10.24×10^6	59.4	27.7
PIDNet[24]		7.72×10^6	60.9	66.5
本文		6.44×10^6	63.0	71.4

图 8 展示了在 AeroScapes 数据集上 Deeplabv3+[8]、SFNet[22]、BiSeNet V2[15]、PIDNet[24] 和本文算法预测结果对比，并在图中对部分差异进行了标注。可以看出，本文模型增强了细节方面的检测。从第 3 行检测结果中的车来看，很明显，改进后的模型检测结果边缘轮廓更为完整，线条更为清晰流畅，更接近真实标签。从其他结果来看，小目标（如人、摩托车、植被）分割更为完整，实现了对小目标检测的增强。综上所述，结果展现了本文算法的优越性。



(a) 原始图像　(b) 真实标签　(c) Deeplabv3+[8]　(d) SFNet[22]　(e) BiSeNet V2[15]　(f) PIDNet[24]　(g) 本文

图 8　部分算法在 AeroScapes 数据集上预测结果比较

Fig. 8　Comparison of prediction results of some selected algorithms on AeroScapes dataset

3　结　论

1）提出了一种用于无人机影像的轻量级语义分割算法，提高了无人机影像分割任务中的分割速度，同时保证了分割结果中目标的完整性。

2）设计了特征增强模块，提高目标特征的显著性，并建模了特征之间的长距离依赖关系，帮助模型更好地提取到小目标特征。

3）针对普通卷积提取特征效果差的问题，提出了全局感知提取模块，增强了提取图像特征时的感受野，有效兼顾了全局上下文信息。

4）设计了特征融合模块，更好地融合细节信息与语义信息，增强了不同尺度特征的交互，同时也降低了模型的计算量，加快了分割速度。

5）在UAVid数据集上，mIoU达到66.5%，分割速度达到79.9帧/s，在AeroScapes数据集上，mIoU达到63.0%，分割速度达到71.4帧/s。从对比实验结果来看，相较大部分对比模型均具有一定的优势，较好实现了精度与速度的平衡。

在今后的工作中，将继续优化模型架构，进一步降低模型参数量与计算量，提升算法分割精度分割和速度。

参考文献（References）

[1] 宝音图, 刘伟, 李润生, 等. 遥感图像语义分割的空间增强注意力U型网络[J]. 北京航空航天大学学报, 2023, 49(7): 1828-1837.
BAO Y T, LIU W, LI R S, et al. Semantic segmentation of remote sensing images based on U-shaped network combined with spatial enhance attention[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1828-1837(in Chinese).

[2] 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).

[3] LIU S Y, CHENG J, LIANG L K, et al. Light-weight semantic segmentation network for UAV remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 8287-8296.

[4] ZHANG G, LI Z, TANG C, et al. CEDNet: a cascade encoder-decoder network for dense prediction[J]. Pattern Recognition, 2025, 158: 111072.

[5] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2023: 1-5.

[6] YU C Q, XIAO B, GAO C X, et al. Lite-HRNet: a lightweight high-resolution network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 10435-10445.

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

[8] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 833-851.

[9] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. (2016-06-07)[2023-12-01]. https://arxiv.org/abs/1606.02147.

[10] 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.

[11] ZHAO H S, QI X J, SHEN X Y, et al. ICNet for real-time semantic segmentation on high-resolution images[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 418-434.

[12] WU T Y, TANG S, ZHANG R, et al. CGNet: a light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.

[13] YU C Q, WANG J B, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 334-349.

[14] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style ConvNets great again[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13728-13737.

[15] YU C Q, GAO C X, WANG J B, et al. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(11): 3051-3068.

[16] YI S, LIU X, LI J J, et al. UAVformer: a composite transformer network for urban scene segmentation of UAV images[J]. Pattern Recognition, 2023, 133: 109019.

[17] WOO S, DEBNATH S, HU R H, et al. Convnext v2: co-designing and scaling ConvNets with masked autoencoders[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 16133-16142.

[18] ZHANG H, WU C R, ZHANG Z Y, et al. ResNeSt: split-attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 2735-2745.

[19] ORŠIC M, KREŠO I, BEVANDIC P, et al. In defense of pre-trained ImageNet architectures for real-time semantic segmentation of road-driving images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 12599-12608.

[20] ZHANG H K, HU W Z, WANG X Y. ParC-net: position aware circular convolution with merits from ConvNets and Transformer[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2022: 613-630.

[21] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.

[22] LI X T, YOU A S, ZHU Z, et al. Semantic flow for fast and accurate scene parsing[C]//Proceedings of the 16th European Conference on Computer Vision. Berlin: Springer, 2020: 775-793.

[23] NIRKIN Y, WOLF L, HASSNER T. HyperSeg: patch-wise hypernetwork for real-time semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 4060-4069.

[24] XU J C, XIONG Z X, BHATTACHARYYA S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 19529-19539.

Real-time UAV image segmentation algorithm with enhanced contextual feature interaction

LI Yunhong*，ZHANG Fuxing，SU Xueping，LI Limin，WANG Mei，LIANG Chengming

(School of Electronics and Information，Xi'an Polytechnic University，Xi'an 710048，China)

Abstract：A real-time UAV image segmentation algorithm with enhanced contextual feature interaction is proposed to address the problem of target omission and incompleteness in segmentation results due to the lack of global information interaction in lightweight algorithms for semantic segmentation tasks of UAV images. The approach uses a two-branch structure. To encode the channel and spatial information, global average pooling in various directions was used. This preserves the correct position information and increases the attention to the local detail information in the image. Secondly, a global perceptual extraction module was constructed by using the position-aware circular convolution and spatial weighting, which achieves the global contextual information capture; Finally, the weighting operation is applied to the features of different scales for the fusion, which reduces the information loss in the fusion process and the computation of the algorithm. The UAVid and AeroScapes datasets are used to validate the algorithm. The results indicate that the mean intersection over union (mIoU) achieved 66.5% and 63.0%, respectively, marking a 2.6% and 2.2% improvement over BiSeNet V2. The segmentation speeds reached 79.9 and 71.4 frames per second, respectively, showing an increase of 8.3 and 6.9 frames per second compared to BiSeNet V2. This method ensures real-time segmentation speed while delivering satisfactory segmentation accuracy.

Keywords：semantic segmentation；UAV image；convolutional neural network；context information；feature fusion

Received：2023-12-22；**Accepted**：2024-03-29；**Published Online**：2024-04-18 17：34

URL：link.cnki.net/urlid/11.2625.V.20240417.1132.001

Foundation items：National Natural Science Foundation of China (62203344); Key Projects of Natural Science Basic Research Program of Shaanxi (2022JZ-35); The Youth Innovation Team of Shaanxi Universities; Xi'an City "Scientist+Engineering" Team Construction Project (25KGYB00029)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0779

液压作动器差动连接下双唇级联靴形往复密封特性研究

彭超[1]，谭爽[1]，叶强[2]，苗建明[1,*]，欧阳小平[3]

(1. 中山大学 海洋工程与技术学院，珠海 519082；2. 国家油气钻井装备工程技术研究中心，宝鸡 721002；
3. 浙江大学 流体动力与机电系统国家重点实验室，杭州 310027)

摘 要：液压作动器差动连接方式使活塞杆处往复密封工况更加恶劣，摩擦力增大，发生倾覆几率大幅度提高。常规单唇口往复密封无法满足高可靠密封要求，以双唇级联靴形往复密封为研究对象，基于软弹流理论，融合有限元软件与数值编程迭代方法，利用流体运输速率守恒准则迭代更新唇间压力，研究差动连接工况下双唇级联靴形往复密封的界面特性，对比双唇级联靴形往复密封与传统VL密封在差动连接工况下的摩擦与泄漏性能。结果表明：双唇级联靴形往复密封的主唇口主要起防止油液泄漏的密封作用，副唇口主要起支撑作用，增强密封结构稳定性；唇间压力有助于减小接触力和摩擦力；差动连接工况条件下，保持相同密封性能（泄漏量），与传统VL密封相比，双唇级联靴形往复密封主唇口的摩擦力下降了36.87%，能够有效减轻唇口磨损，减少差动连接工况中发生倾覆的几率。

关 键 词：往复密封；双唇口；差动连接；唇间压力；密封特性
中图分类号：V245.1；TB42
文献标志码：A **文章编号**：1001-5965(2026)03-0678-09

往复密封件作为液压系统中的核心部件，广泛应用于航空、深海、石油等工业领域。据统计，作动器失效情形中，90% 是由于往复密封故障引起的[1]。往复密封一旦失效，轻者导致内部介质泄漏，影响周围环境，降低液压系统效率，重则导致液压结构元件的工作失效，危及液压系统安全[2]。

为提高工作效率，作动器在无负载的快进出行程中，一般采用差动连接的工作方式，其是在不增加泵源流量的前提下实现更快速度工作的有效方法。但相比正常出行程的工作方式，差动连接出行程的工作方式使得活塞杆往复密封所处工况更加复杂。作动器处于简单连接出行程时，有杆腔内压力接近零；而差动连接出行程时，有杆腔压力为系统高压，此时密封所受到的腔体内压力推动密封向外，活塞杆伸出所产生的摩擦力也会拖拽密封向外。同时，在出行程中，受空化效应影响，与同样使有杆腔为系统高压的入行程相比，密封在高压出行程的摩擦力大于相同压力下入行程的摩擦力[3-4]。因此，作动器差动连接使活塞杆处密封工况更加恶劣，增大密封扭转的发生，仅采用常规的往复密封（如斯特封等）无法满足高可靠性的要求，需要更稳定的往复密封结构。

为提高往复密封性能，更稳定的双唇密封结构被逐渐应用到密封系统中[5-9]，如车式齿形密封、Plus Seal 等。双唇密封结构中的副唇口不仅保护主唇口不受外界污染物的影响，而且起到辅助支撑的

收稿日期：2023-11-30；录用日期：2023-12-22；网络出版时间：2024-01-26 18:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20240126.0953.001
基金项目：国家重点研发计划 (2021YFC2800700)；国家自然科学基金 (52305083)；广东省基础与应用基础研究基金 (2023A1515012042)
* 通信作者. E-mail：miaojm@mail.sysu.edu.cn

引用格式：彭超，谭爽，叶强，等. 液压作动器差动连接下双唇级联靴形往复密封特性研究 [J]. 北京航空航天大学学报，2026，52（3）：678-686.
PENG C，TAN S，YE Q，et al. Investigation into characteristics of double-lip cascade boot-shaped reciprocating seals in differential-connection working conditions of hydraulic actuators [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：678-686（in Chinese）.

作用，大大增强了密封件的稳定性。双唇密封结构与组合密封形式相结合，利用组合密封聚四氟乙烯(PTFE)环部分摩擦力较低的特点来避免往复运动中摩擦力增大的问题，具有较为出色的密封性能[10-11]。国内外围绕双唇密封结构也展开了相关研究。Yang 和 Salant[6-7] 在双唇密封方面做了一些开创性的工作，发现油液会在唇间区聚集，形成一定的压力，并提出压力迭代的方法来获得唇间压力。Huang 等[12] 发现唇间区聚集的油液可以减少乏油润滑的发生。张付英等[13] 研究发现，粗糙度是影响双唇密封性能的重要因素，随着粗糙度的增加，存在临界粗糙度使泄漏方向发生改变。Gu 等[14] 建立了多唇式弹流润滑模型，研究了双作用多唇密封的密封特性。除上述双唇单一密封的性能研究外，Peng 等[15] 还对一种双唇组合密封进行性能研究，发现其唇间压力随着速度的降低而降低，随着粗糙度的增加，唇间压力先增加后减少。

本文以结构稳定性强的双唇组合密封为研究对象，研究其在差动连接下的密封性能。基于软弹流理论，融合有限元软件和数值编程迭代方法，综合考虑界面流体域方程、微观接触理论、变形及流-固耦合理论，利用流体运输速率守恒和唇间压力更新来耦合 2 个密封唇口的特性，深入研究差动连接工况条件下双唇组合往复密封的特性，并与传统 VL 密封进行性能对比，揭示差动连接下双唇组合密封的摩擦泄漏特性。

1　建模方法

本文以一种带有靴形双唇的组合密封为研究对象，如图 1 所示。整个密封件由靴形 PTFE 环和丁腈橡胶 (NBR) O 形圈 2 个部分组成，融合了橡胶材料超弹性和 PTFE 低摩擦系数、高抗磨性等优点，具有良好的密封性能和寿命。靴形 PTFE 环有 2 个密封唇口：位于腔体侧的主唇口和位于空气侧的副唇口。当密封件安装到槽内时，主唇口和副唇口与活塞杆接触形成 2 个密封区域，2 个唇口之间有一个间隙，称为唇间区。

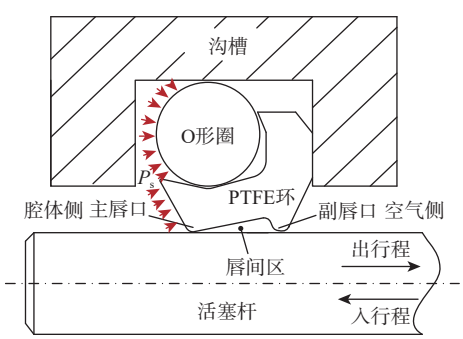


图 1　双唇组合密封示意图

Fig. 1　Schematic of seal with double lips combination

往复运动时，活塞杆与密封件之间存在一层油膜，其油膜特性会对油液泄漏和摩擦磨损特性产生深刻影响。大量理论和试验研究表明，往复密封界面多为混合润滑状态[16]，混合润滑理论模型包括界面流体域模型、微观接触模型和变形方程。

1.1　界面流体域模型

随着活塞杆的往复运动，密封唇口与杆之间存在一层油膜。界面流体域模型采用修正后的雷诺方程(见式 (1))。引入流体因子 ϕ_{xx}、$\phi_{s.c.x}$，以考虑表面微凸体对流体的裹挟、阻碍作用[17]。同时，借助 JFO 模型考虑密封区域的空化效应，变量 Φ 在流体区域表示无量纲油膜压力，在空化区域则表示无量纲等效密度，利用有限体积法和 TDMA 算法求解雷诺方程。

$$\frac{\partial}{\partial x}\left[\frac{\phi_{xx}\rho h^3}{\mu\dfrac{\partial}{\partial x}(F\Phi)}\right]=6u\frac{\partial}{\partial x}\left[\rho+(1-F\Phi)\right](h_{\mathrm{T}}+R_{\mathrm{a}}\phi_{\mathrm{s.c.}x}) \tag{1}$$

式中：ρ 为流体密度；u 为往复运动速度；μ 为流体黏度；F 为空化开关，在流体域时等于 1，在空化域时等于 0；h 为流体油膜厚度；h_{T} 为截断膜厚；R_{a} 为轮廓算数平均偏差。

密封界面流体黏度随油膜压力更新而变化，本文用 Roeland 方程表征压力-黏度效应。

$$\mu=\mu_0\exp\left[(\ln\mu_0+9.67)\left(1+\frac{p_{\mathrm{f}}}{p_{\mathrm{L}}}\right)^{0.68}-1\right] \tag{2}$$

式中：p_{L} 为参考压力，p_{L}=0.196 GPa；p_{f} 为油膜压力；μ_0 为大气条件下的初始黏度。

1.2　微观接触模型

因密封区域为混合润滑状态，除油膜之外，在密封区域也会发生微凸体接触。本文采用 Greenwood-Williamson 接触模型计算微凸体接触压力。假定凸度服从高斯分布，根据 Greenwood-Williamson 模型计算微凸体接触压力 p_{c}，如下：

$$p_{\mathrm{c}}=\frac{4}{3}\eta E R_{\mathrm{ar}}^{1/2}\int_h^{\infty}\phi(z)(z-h)^{\frac{3}{2}}\mathrm{d}z \tag{3a}$$

$$\phi(z)=\frac{1}{\sqrt{2\pi}R_{\mathrm{a}}}\mathrm{e}^{\frac{-z^2}{2R_{\mathrm{a}}^2}} \tag{3b}$$

式中：η 为微凸体密度；R_{ar} 为平均粗糙度半径；E 为等效弹性模量；$\phi(z)$ 为表面微凸体概率密度分布函数。

1.3　变形方程

往复运动时，微凸体接触压力与油膜压力之和不同于静态接触压力 p_{static}(见式 (4))。密封环被轻微抬起，但由于油膜厚度非常小，假定密封环的轻

微变形与压力差成正比。

$$\Delta p = p_{\mathrm{f}} + p_{\mathrm{c}} - p_{\mathrm{static}} \tag{4}$$

变形计算采用半无限空间法。流体动压效应引起的油膜变形 Δh 可由式 (5) 得到:

$$\Delta h(x) = -\frac{1-\nu^2}{\pi E}\int_0^{x_0} \Delta p(\xi)\ln(\xi - x)^2 \mathrm{d}\xi \tag{5}$$

式中: ξ为沿密封长度的附加坐标; ν 为泊松比; x_0 为轴向密封宽度。

油膜厚度可表示为

$$h = h_{\mathrm{int}} + \Delta h \tag{6}$$

式中: h_{int} 为初始静态油膜厚度, 可以通过已知静态接触压力的情况下, 将式 (3a) 反向计算获得。

1.4 载荷收敛准则

上述往复密封各子模型间存在强耦合关系, 模型将不断迭代, 直到达到收敛条件。计算中采用载荷平衡法(见式 (7a) 和式(7b)); 同时, 当整个计算达到收敛时, 压差趋于稳定, 最小油膜厚度收敛(见式 (7c))。数值迭代收敛时需要同时满足上述 2 个条件。

$$b_{\mathrm{res}} = \frac{\left|\sum (p_{\mathrm{static}}(x,y)) - \sum (p_{\mathrm{c}}(x,y) + p_{\mathrm{f}}(x,y))\mathrm{d}x\right|}{\sum (p_{\mathrm{static}}(x,y))\mathrm{d}x} \tag{7a}$$

$$\frac{\left|b_{\mathrm{res}}^{\mathrm{new}} - b_{\mathrm{res}}^{\mathrm{old}}\right|}{b_{\mathrm{res}}^{\mathrm{old}}} \leqslant \beta \tag{7b}$$

$$\frac{\left|h_{\mathrm{min}}^{\mathrm{new}} - h_{\mathrm{min}}^{\mathrm{old}}\right|}{h_{\mathrm{min}}^{\mathrm{old}}} \leqslant \varepsilon \tag{7c}$$

式中: β 和 ε 分别为载荷平衡条件下的载荷收敛容差和最小油膜厚度收敛容差, 在迭代计算中, 2 个容差均设置为 0.01。

1.5 摩擦力和泄漏量计算

在密封界面油膜分布完成计算后, 获得了密封区域流体压力分布和微观接触压力分布。由于密封区域属于混合润滑状态, 摩擦力 F_{t} 既有微凸体接触摩擦造成的库伦摩擦力 F_{a}, 又包含流体黏性剪切力 F_{f}。

密封区域微凸体接触引起的库伦摩擦力表示为

$$F_{\mathrm{a}}^i = \pi D_{\mathrm{rod}} f p_{\mathrm{c}}^i \tag{8}$$

式中: D_{rod} 为活塞杆直径; f 为摩擦系数。

密封区域各点的流体黏性剪切力表示为

$$F_{\mathrm{f}}^i = \pi D_{\mathrm{rod}}\left[\frac{-\mu u}{h^i}(\phi_{\mathrm{f}} + \phi_{\mathrm{fs}}) + \phi_{\mathrm{fp}}\frac{h^i}{2}\cdot\frac{\mathrm{d}p_{\mathrm{f}}^i}{\mathrm{d}x}\right] \tag{9}$$

式中: 流体因子 ϕ_{f}、ϕ_{fs} 和 ϕ_{fp} 分别用于修正速度、表面粗糙度和流体压力对摩擦力的影响。

往复作动中, 总的摩擦力为库伦摩擦力与流体黏性剪切力之和, 如下:

$$F_{\mathrm{t}} = \sum \left(F_{\mathrm{a}}^i + F_{\mathrm{f}}^i\right)\mathrm{d}x \tag{10}$$

流体运输速率由式 (11) 计算, 其由泊肃叶流动部分和库埃特流动部分组成。

$$q = \pi D_{\mathrm{rod}}\frac{1}{12}\cdot \left\{\underbrace{-\phi_{\mathrm{cc}}\frac{h^3}{\mu}\cdot\frac{\mathrm{d}p_{\mathrm{f}}}{\mathrm{d}x}}_{\text{泊肃叶流动}} + \underbrace{6u[1+(1-F)\Phi]\left[h_{\mathrm{T}} + R_{\mathrm{a}}\phi_{\mathrm{s.c.}x}\right]}_{\text{库埃特流动}}\right\} \tag{11}$$

式中: ϕ_{cc}为压力流体因子。

1.6 计算流程

结合上述公式, 计算油膜厚度分布的整个过程如图 2 所示。首先, 通过在 ANSYS 软件中建立宏观有限元分析 (finite element analysis, FEA) 模型, 得

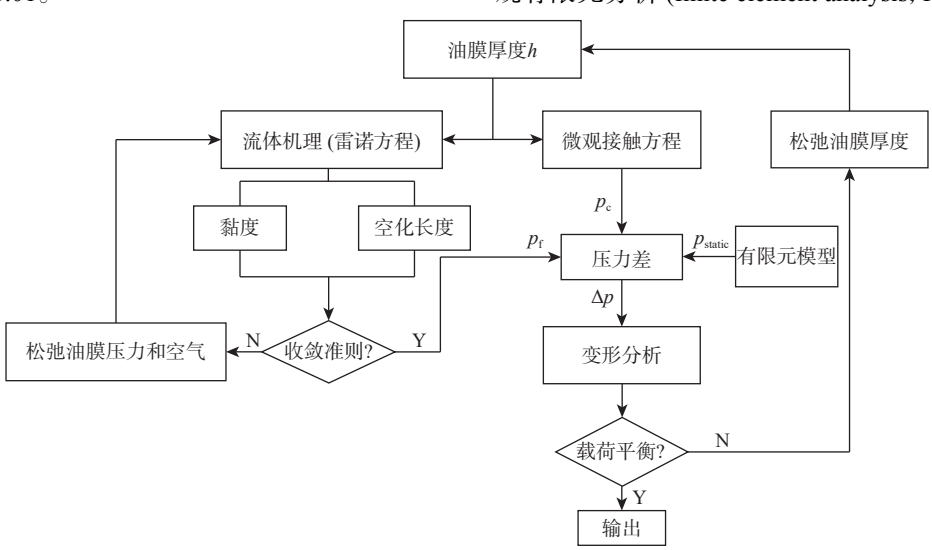


图 2 油膜厚度计算流程

Fig. 2 Procedure of oil film thickness calculation

到静态接触压力 P_{static}，利用静态接触压力分布可获得初始油膜厚度和密封长度；而界面流体模型被用于计算往复作动中流体动压效应，同时考虑黏压影响，整个密封界面流体模型迭代直到黏度和空化长度完成收敛。然后，由微观接触模型计算混合润滑条件下的微凸体接触压力；将计算所得微凸体接触压力和油膜压力之和同静态接触压力 p_{static} 相比较，其差值导致油膜厚度变化，更新整个油膜厚度后，再次计算界面油膜压力；整个循环不断迭代直到收敛。密封界面油膜厚度计算完成后，通过式(11)辅助计算流体运输速率。但是，由于双唇密封中有 2 个密封唇口，2 个密封唇口的流体运输速率不一定相同，并且随着唇口两侧的压力差改变而发生变化。如流体输送能力较低的唇口位于整个密封的下游方向时，则油液会被该唇口阻碍，从而在唇间区积聚，唇间压力逐渐增大。唇间压力的不断增大又将导致 2 个唇口两侧的压力差发生变化，进而影响 2 个唇口的流体运输速率。因此，唇间区唇间压力迭代的整体计算如图 3 所示，通过油膜厚度计算流程，分别计算 2 个唇口流体运输速率，根据 2 个唇口的流体运输速率迭代更新唇间压力，直到主唇口的流体运输速率与副唇口的流体运输速率相等。

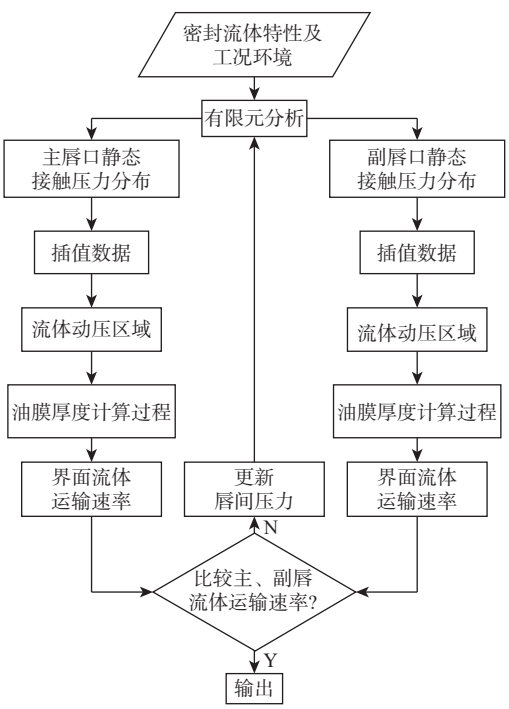


图 3　双唇密封迭代计算流程

Fig. 3　Procedure of iterative calculation of seal with double lips

2　仿真设置

靴形双唇组合密封中，O 形圈部分由橡胶制成，考虑到橡胶本身结构的复杂性及应变-应力非线性，选用式 (12) 所示的 Mooney-Rivlin 模型来描述橡胶的应变-应力关系：

$$W = C_{10}(I_1 - 3) + C_{01}(I_2 - 3) \tag{12}$$

式中：C_{10} 和 C_{01} 为 Mooney-Rivlin 参数；I_1 和 I_2 为 Green-Lagrange 应变[18-19]。

往复密封在正常情况下处于轴对称状态，因此，将密封截面作为研究对象，密封简化为二维轴对称模型，如图 4 所示。整个模型在 ANSYS 软件中建立，L 形圈和 O 形圈被活塞杆的运动挤压，利用有限元模型可以得到密封的内部应力和静态接触压力。有限元模型的单元类型采用 PLANE183 单元，并选用 8 节点四边形单元，考虑到橡胶材料的超弹性特性，在单元公式中采用混合 U/P 法。靴形双唇组合密封属性参数如表 1～表 3 所示，其中，L 形圈材料为 PTFE，O 形圈材料为 NBR，其材料参数通过万能材料试验机进行压缩试验获得。

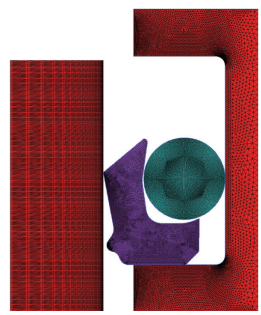


图 4　双唇组合密封宏观有限元模型

Fig. 4　Macroscopic FEA model for combined seals with double lips

表 1　O 形圈材料参数

Table 1　Parameters of O-ring material

参数	数值
材料	丁腈橡胶(NCT83)
截面直径/mm	2.62
内径/mm	26.64
常温(25℃)邵氏硬度	75
Mooney-Rivlin系数/MPa	C_{10}=−0.11, C_{01}=1.38

表 2　L 形圈材料参数

Table 2　Parameters of L-ring material

参数	数值
材料	聚四氟乙烯(Turcon®M30)
泊松比	0.42
弹性模量/MPa	221
摩擦系数(25℃)	0.045

表 3　活塞杆和沟槽材料参数

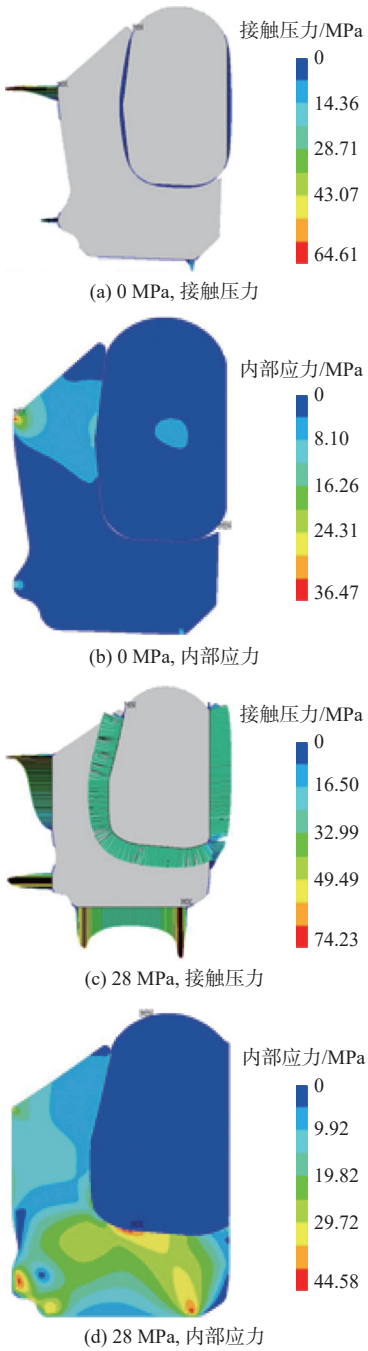
Table 3　Parameters of rod and groove material

参数	数值
材料	合金钢(30CrMnSiA)
活塞杆直径/mm	25.35
弹性模量/MPa	2×10^5
衬套底径/mm	31.50

3 靴形双唇组合密封特性分析

3.1 初始阶段下的特性分析

双唇组合密封在初始阶段下系统压力为0 MPa和28 MPa时的接触压力和内部应力分布如图5所示。当密封件被安装到沟槽内后(0 MPa条件下),由于主唇口直接被O形圈挤压,其接触压力和内部应力远大于副唇口,副唇口同活塞杆接触可起到稳定整个密封件的作用,防止密封件的倾覆。在施加28 MPa系统压力后,副唇口也被挤压,其接触压力



(a) 0 MPa, 接触压力

(b) 0 MPa, 内部应力

(c) 28 MPa, 接触压力

(d) 28 MPa, 内部应力

图5 初始阶段下的接触压力和内部应力分布

Fig. 5 Distribution of contact pressure and internal stress in initial stage

与内部应力增大,2个唇口间的间隙减小。在往复作动过程中,该间隙区域可作为储油槽,储油槽中的油液起到润滑主、副唇口的作用。在28 MPa系统压力下,挤压变形后的储油槽截面积约为0.034 mm^2,其存储油液量约为2.70 mm^3。

双唇密封在初始阶段下系统压力为0 MPa和28 MPa时的静态接触压力曲线如图6所示。在系统压力为0 MPa下,相较于主唇口,副唇口的接触压力较小,对整个密封件起到了主要支撑作用。随着系统压力增大到28 MPa,主唇口处的接触压力峰值略微增大,副唇口处的接触压力峰值大幅度增大。同时,随着系统压力升高,2个唇口的密封区域宽度均大幅度增加。由于密封唇口的形状,主唇口的接触压力分布呈斜三角形,副唇口的接触压力分布呈抛物线形,其中,斜三角形的接触压力分布有助于防止往复运动中油液泄漏[20]。

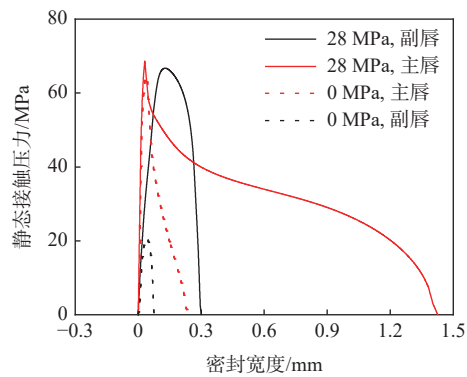


图6 静态接触压力分布曲线

Fig. 6 Distribution curves of static contact pressure

3.2 差动连接工况下的特性分析

图7为正常连接与差动连接的液压油路图。当换向阀1和换向阀2均处于左侧位置时,作动器处于差动连接工作状态。作动器的进油和回油连

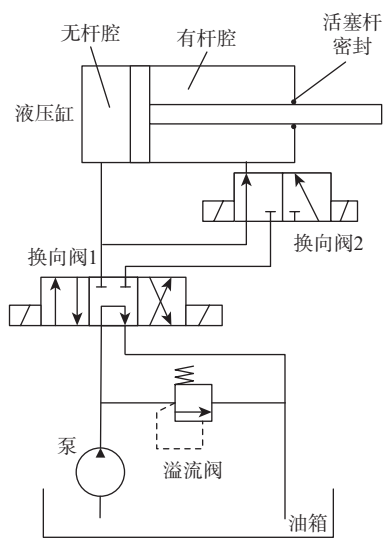


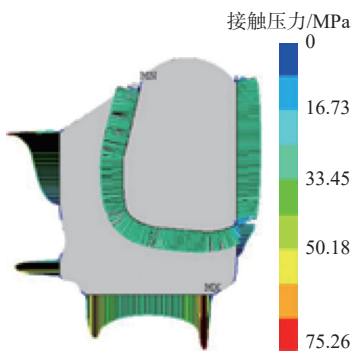
图7 作动器在不同状态下的液压回路

Fig. 7 Hydraulic circuit for cylinder in different working states

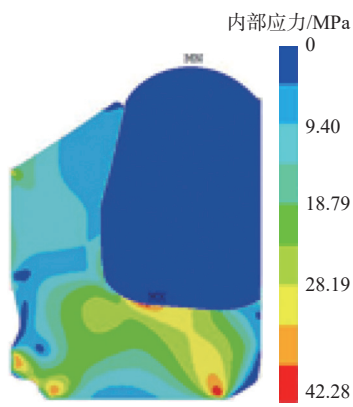
接在一起，如忽略管道和换向阀 2 等的压力损失，作动器的无杆腔和有杆腔的压力相等，均处于高压状态。本文中，差动连接时系统压力为 28 MPa。此时，由于受到高压的作用，主唇口界面的油液泄漏速度大于副唇口油液泄漏速度，油液在 2 个唇口间隙中持续积累，在唇间区会逐渐形成一定的唇间压力，形成困油现象。当主唇口泄漏的油液量同副唇口处泄漏的油液量一致时，该间隙的唇间压力达到稳定。

在差动连接工作状态下的初始阶段时，唇间区的油液还未聚集，唇间压力为零，此时主唇口的流体运输速率大于副唇口，使唇间油液越来越多，压力逐渐升高。在迭代计算中，按照唇间油液逐渐聚集的规律，根据主、副唇口流体运输速率守恒原则，不断迭代 2 个唇口的唇间压力值，当唇间压力达到 10.5 MPa 时，最终实现 2 个唇口流体运输速率的平衡，此时主、副唇口的流体运输速率为 2.878 mm^3/s。唇间压力稳定后，密封界面接触压力和内部应力分布如图 8 所示。整体分布情况同初始阶段下 28 MPa 时的分布情况类似，其中，接触压力的最大值大于初始阶段，内部应力的最大值小于初始阶段。

双唇组合密封唇间压力建立前后主、副唇口的界面特性如图 9 所示。图 9(a) 为宏观接触压力分布，相比于高压环境密封初始状态的接触压力，在唇间压力作用下，主唇口的接触压力峰值增大，而副唇口的接触压力峰值减小，同时 2 个唇口逐渐被挤向两侧，两者的密封区域宽度均小于初始阶段下



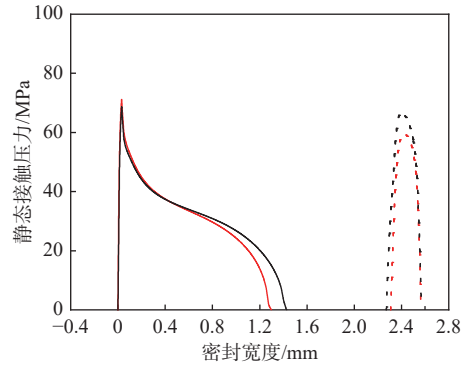
(a) 接触压力



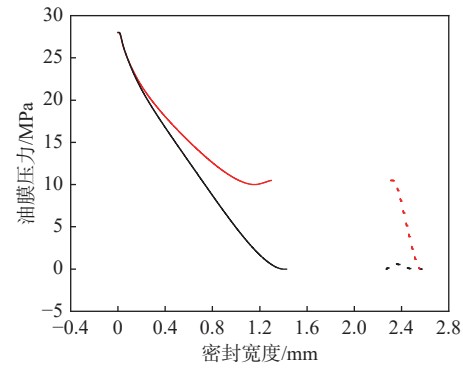
(b) 内部应力

图 8　唇间压力稳定后的接触压力和内部应力分布

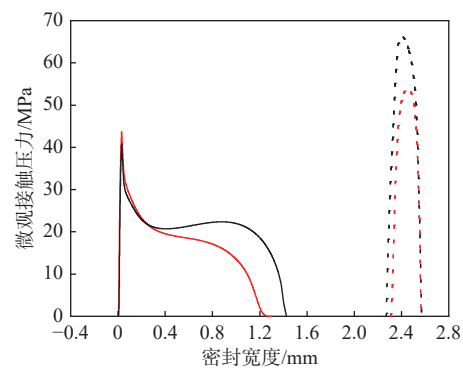
Fig. 8　Distribution of contact pressure and internal stress after inter-lip pressure is stabilized



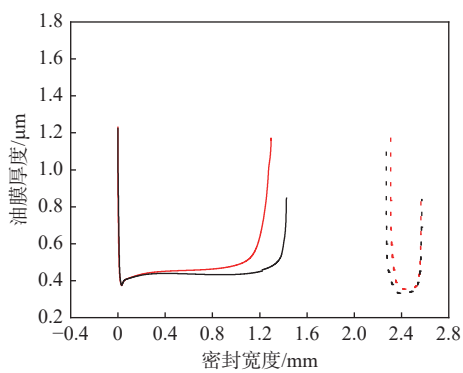
(a) 静态接触压力



(b) 油膜压力



(c) 微观接触压力



(d) 油膜厚度

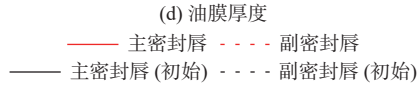


图 9　唇间压力稳定后的双唇密封特性

Fig. 9　Characteristics of seal with double lips after inter-lip pressure is stabilized

的密封区域宽度。

图 9(b) 为油膜压力分布，在双唇密封处于初始阶段时，由于唇间区的唇间压力还未建立，主唇口密封区域的油膜压力单调减小到 0 MPa，而副唇口则在间隙侧产生较小的油膜压力后，空气侧大面积发生空化。当唇间压力稳定后，主唇口密封区域的油膜压力减小至唇间压力 10.5 MPa，但此过程中，油膜压力并非单调下降。由于空气侧油膜厚度的急剧升高，靠近唇间区的主唇口这一狭小密封区域的油膜压力低于唇间压力，随后回升到唇间压力；副唇口密封区域的油膜压力从 10.5 MPa 开始单调减小至 0 MPa。此外，主、副唇口密封区域油膜厚度的增大还使密封件被轻微抬起，从而造成 2 个唇口的微观接触压力减小，如图 9(c)、图 9(d) 所示。

唇间压力的建立减小了主唇口两侧的压力差，从而降低了泊肃叶流动，减小了主唇口的流体运输速率；对于副唇口，唇间压力的建立增大了副唇口两侧的压力差，提高了副唇口的流体运输速率。2 个唇口处的流体运输速率一减一增，最终达到平衡。

4 双唇组合密封与传统 VL 密封在差动连接工况中的性能对比

本文针对双唇级联靴形往复密封结构与传统 VL 密封结构进行对比分析。VL 密封是 Trelleborg 公司的密封形式，广泛应用于航空作动器的轴用密封中。如图 10 所示，VL 密封也由 PTEF 环和 O 形圈 2 个部分组成，其唇口形式同双唇级联靴形往复密封的主密封唇口形式类似，区别在于 VL 密封的 PTFE 环为 L 形，只有一个密封唇口。

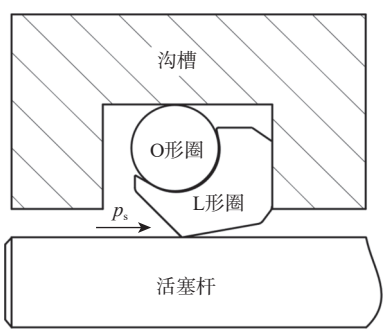


图 10 VL 密封示意图

Fig. 10 Schematic of VL seal

差动连接工况下(28 MPa)，VL 密封接触界面特性如图 11 所示。经过与图 9 的对比可知，在相同的差动工况下，VL 密封的界面润滑特性与初始阶段下双唇组合密封主唇口界面润滑特性相似。2 个密封区域的油膜压力都是沿密封方向单调减小至 0 MPa。最小油膜厚度临近腔体侧，从最小油膜位置开始，

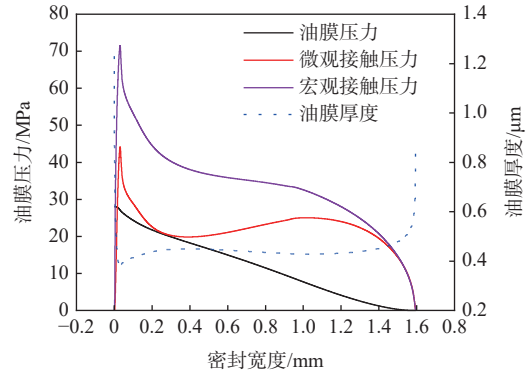


图 11 差动连接工况下 VL 密封特性

Fig. 11 Characteristics of VL seal in differential conditions

沿密封长度方向，油膜厚度都是按照先增大后减小随后再增大的分布规律。油膜厚度能直接反映微观接触压力大小，当油膜厚度增大时，微观接触压力减小。

差动连接工况(28 MPa)下，VL 密封与双唇组合密封的主唇口摩擦与泄漏特性对比如图 12 所示。在 28 MPa 系统压力下，VL 密封和靴形双唇密封的差动连接工况下泄漏量均约为 2.88 mm^3/s，在相同泄漏水平下，相比于 VL 密封摩擦力 125.62 N，当唇间压力建立后，双唇组合密封的主唇口摩擦力下降了 36.87%，有效保护了主密封唇口，从而延长密封寿命。

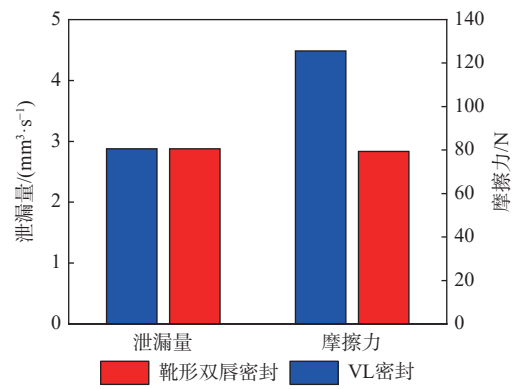


图 12 差动连接工况下 VL 密封与双唇组合密封的主唇口摩擦与泄漏特性对比

Fig. 12 Comparison of friction and leakage characteristics between VL seal and primary lip of combined seal with double lips in differential conditions

5 结 论

本文分析了差动连接工况下的双唇级联靴形密封界面润滑特性，并在相同差动连接工况条件下与传统 VL 密封进行性能对比，得到以下结论：

1) 双唇密封结构中，主唇口的接触压力分布为斜三角形，有助于防止往复运动中油液的泄漏；副唇口有助于支撑密封件防止其扭转，增强密封结构

稳定性。

2）双唇组合密封的唇间压力有助于支撑密封件并减小接触力和摩擦力。

3）在相同泄漏水平下, 当唇间压力建立后, 靴形双唇组合密封的主唇口摩擦力相比于 VL 密封下降了 36.87%, 可有效减少主唇口磨损, 延长密封寿命。

参考文献（References）

[1] OLSEN R B. Dynamic seals for advanced hydraulic systems[R]. Dallas: Vought Corp, 1981.

[2] 马琳. 液压往复密封的摩擦特性分析[J]. 液压与气动, 2021, 45(12): 122-128.
MA L. Friction characteristics analysis for hydraulic reciprocating seal[J]. Chinese Hydraulics & Pneumatics, 2021, 45(12): 122-128 (in Chinese).

[3] PENG C, GUO S R, OUYANG X P, et al. Mixed lubrication modeling of reciprocating seals based on a developed multiple-grid method[J]. Tribology Transactions, 2018, 61(6): 1151-1161.

[4] PENG C, OUYANG X, SCHMITZ K, et al. Numerical and experimental study on combined seals with the consideration of stretching effects[J]. Journal of Tribology, 2021, 143(6): 062301.

[5] 车恒德. 车恒德^{R}密封及其应用[J]. 机床与液压, 2002, 30(6): 35-38.
CHE H D. Chehengde ^{R} seal and its application[J]. Machine Tool & Hydraulics, 2002, 30(6): 35-38(in Chinese).

[6] YANG B. Elastohydrodynamic model of reciprocating hydraulic rod seals[D]. Atlanta: Georgia Institute of Technology, 2010.

[7] YANG B, SALANT R F. A numerical model of a reciprocating rod seal with a secondary lip[J]. Tribology Transactions, 2008, 51(2): 119-127.

[8] 彭超. 多工况往复密封跨尺度建模及试验研究[D]. 杭州: 浙江大学;2018: 135-145.
PENG C. Trans-dimensional modelling and experimental study for reciprocating seals under multi-working conditions[D]. Hangzhou: Zhejiang University, 2018: 135-145(in Chinese).

[9] 张付英, 初宏怡, 贺佘燕. 双唇 Y 形拉杆密封的密封性能研究[J]润滑与密封, 2021, 46(9): 21-25.
ZHANG F Y, CHU H Y, HE S Y. Study on sealing performance of double lip Y-type pull rod seal [J]. Lubrication Engineering, 2021, 46(9): 21-25(in Chinese).

[10] PENG C, OUYANG X, SCHMITZ K, et al. Investigation of the tribological performance of reciprocating seals in a wide temperature range[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2021, 235(11): 2396-2414.

[11] PENG C, OUYANG X P, GONG G F, et al. Investigation into the performance of the VL seal based on the 3D model[C]//Proceedings of the ASME/BATH 2017 Symposium on Fluid Power and Motion Control. New York: ASME, 2017: V001T01A017.

[12] HUANG Y L, SALANT R F. Numerical analysis of a hydraulic rod seal: flooded vs. starved conditions[J]. Tribology International, 2015, 92: 577-584.

[13] 张付英, 初宏怡, 贺佘燕. 基于田口试验方法的双唇 Y 形拉杆封参数优化研究[J]. 润滑与密封, 2021, 46(12): 105-111.
ZHANG F Y, CHU H Y, HE S Y. Optimization of structural parameters of double lip Y-type pull rod seal based on taguchi test method[J]. Lubrication Engineering, 2021, 46(12): 105-111(in Chinese).

[14] CHENG D H, GU L C, SUN Y. Mixed lubrication modeling of multi-lip reciprocating seals based on elastohydrodynamic lubrication theory[J]. Machines, 2022, 10(6): 483.

[15] PENG C, MIAO J M, BAUER N, et al. Investigation into the inter-lip characteristics of combined seals with double lips in different working conditions[J]. Tribology International, 2023, 178: 108036.

[16] SCHMIDT T, ANDRÉ M, POLL G. A transient 2D-finite-element approach for the simulation of mixed lubrication effects of reciprocating hydraulic rod seals[J]. Tribology International, 2010, 43(10): 1775-1785.

[17] PATIR N, CHENG H S. An average flow model for determining effects of three-dimensional roughness on partial hydrodynamic lubrication[J]. Journal of Lubrication Technology, 1978, 100(1): 12-17.

[18] MOONEY M. A theory of large elastic deformation[J]. Journal of Applied Physics, 1940, 11(9): 582-592.

[19] RIVLIN R S. Large elastic deformations of isotropic materials. IV. Further developments of the general theory[J]. Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences, 1948, 241(835): 379-397.

[20] MULLER H. Fluid sealing technology: principles and applications [M]. London: Routledge, 2019.

Investigation into characteristics of double-lip cascade boot-shaped reciprocating seals in differential-connection working conditions of hydraulic actuators

PENG Chao[1]，TAN Shuang[1]，YE Qiang[2]，MIAO Jianming[1,*]，OUYANG Xiaoping[3]

(1. School of Ocean Engineering and Technology，Sun Yat-Sen University，Zhuhai 519082，China；
2. National Engineering Research Center for Oil and Gas Drilling Equipment，Baoji 721002，China；
3. State Key Laboratory of Fluid Power and Mechatronic Systems，Zhejiang University，Hangzhou 310027，China)

Abstract：The differential connection of the hydraulic actuator deteriorates the working conditions of reciprocating seals at the piston rod, increasing the friction, as well as the chance of overturning. The conventional single-lip reciprocating seal can not meet the requirements of high-reliability seal. The double-lip cascade boot-shaped reciprocating seal was taken as the research object in this paper. This study used the fluid transport velocity conservation criterion to iteratively update the inter-lip pressure and investigate the interface characteristics of the double-lip cascade boot-shaped reciprocating seal under differential working conditions. It was based on the soft elastic flow theory and connected the numerical programming iterative method with the finite element software. The performance of the double-lip cascade boot-shaped reciprocating seal were was compared with that of the traditional VL seal in the same differential working condition. According to the findings, the primary lip of the double-lip cascade boot-shaped reciprocating seal primarily serves to seal and stop oil leaks, while the secondary lip primarily supports and strengthens the sealing structure. The inter-lip pressure helps to reduce the contact force and friction. Under the same differential working conditions, compared with that of the traditional VL seal, the friction of the primary lip of the double-lip cascade boot reciprocating seal is reduced by 36.87%, which can effectively reduce the wear of the lip and the chance of overturning in differential working conditions.

Keywords：reciprocating seal; double lips; differential connection; inter-lip pressure; sealing characteristics

Received：2023-11-30；**Accepted**：2023-12-22；**Published Online**：2024-01-26 18：34

URL：link.cnki.net/urlid/11.2625.V.20240126.0953.001

Foundation items：National Key Research and Development Program of China (2021YFC2800700); National Natural Science Foundation of China (52305083); Guangdong Basic and Applied Basic Research Foundation (2023A1515012042)

* **Corresponding author**. E-mail：miaojm@mail.sysu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0777

基于R藤Copula-DBN齿轮箱故障诊断

王进花[1,*]，刘正奇[1]，曹洁[1,2,3]，刘昀强[1]，陈莉[2]

(1. 兰州理工大学 电气工程与信息工程学院，兰州 730050；2. 兰州城市学院 信息工程学院，兰州 730050；
3. 甘肃省制造信息工程研究中心，兰州 730050)

摘 要：传统动态贝叶斯初始网络在多维数据下进行结构学习时，需搜索的有向无环图空间大，难以获得最优结构，导致故障诊断精度低。为此，提出一种R藤Copula模型与动态贝叶斯网络(DBN)相结合的故障诊断方法。采用结构预测模型对所提取的特征进行筛选，得到相关性较强的节点，减小网络结构空间的大小；采用R藤Copula模型第一层树结构结合传递熵方法构建动态贝叶斯初始网络，将初始网络在时间序列按马尔可夫过程展开构建DBN进行故障诊断，解决在多特征下网络构建难以获得最优结构的问题。采用东南大学齿轮箱数据进行验证，与其他方法对比，结果表明，所提方法能够更好地进行DBN结构学习，且数据与模型的拟合度较高，在故障诊断时能够取得良好的诊断结果。

关 键 词：故障诊断；动态贝叶斯网络；R藤Copula；结构预测模型；齿轮箱
中图分类号：TP277; TH133.33
文献标志码：A **文章编号**：1001-5965(2026)03-0687-11

齿轮箱是工业领域中广泛使用的机械装置，承担着将高速低扭矩动力转换为低速高扭矩输出的功能[1]。由于长期运转和缺乏维护等原因，齿轮箱部件容易出现各种故障，如齿面疲劳、齿面磨损和断齿等，这些故障如果不能及时发现和处理，将导致设备停机，给生产带来巨大的损失[2]。因此，对齿轮箱进行故障诊断和维护尤为重要。

故障耦合性和不确定性是齿轮箱故障诊断研究中面临的难点之一。Chen等[3]针对行星齿轮箱基于单一故障信号故障诊断结果具有高度不确定性的问题，提出一种基于多源信息深度融合的行星齿轮箱故障诊断方法，有效降低了故障诊断的不确定性。Gu等[4]为降低齿轮箱中不确定因素对故障诊断结果的影响，提出一种改进的动态不确定因果关系图故障诊断方法，可准确识别根故障。Guo等[5]针对齿轮箱发生多部件耦合故障时诊断精度低的问题，提出一种基于多任务并行卷积神经网络的齿轮箱耦合故障诊断方法，实现了同时对多个部件耦合故障的诊断。Huang等[6]针对齿轮箱在实际工况下故障耦合性高、故障数据不充分的问题，提出一种由数据驱动的改进分类器生成对抗网络，保证了生成数据的质量，提高了模型在实际工况场景下的泛化能力。

前人通过不同方法克服故障耦合性和不确定性，但仍存在一些不足：基于多源信息深度融合的方法需要大量训练数据，数据的获取是一个挑战。基于动态不确定因果图的方法，在复杂齿轮箱系统中选择恰当的变量和合理的权重具有困难。基于多任务并行卷积神经网络的方法采用强化输入和多任务并行，导致模型复杂度增加，限制了实际应用的可行性。

基于深度学习的方法在进行故障诊断时存在过拟合现象，导致诊断精度不高[7]，动态贝叶斯网络[8](dynamic Bayesian network, DBN)通过概率分布的

收稿日期：2023-11-30；**录用日期**：2024-03-12；**网络出版时间**：2024-04-25 13:50
网络出版地址：link.cnki.net/urlid/11.2625.V.20240425.0834.001
基金项目：国家自然科学基金(62063020)；国家重点研发计划(2020YFB1713600)；甘肃省自然科学基金(20JR5RA463)
*通信作者. E-mail：wjh0615@lut.edu.cn

引用格式：王进花，刘正奇，曹洁，等. 基于R藤Copula-DBN齿轮箱故障诊断[J]. 北京航空航天大学学报，2026，52(3)：687-697.
WANG J H，LIU Z Q，CAO J，et al. Gearbox fault diagnosis based on R-vine Copula-DBN[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：687-697 (in Chinese).

形式对不确定性进行建模，建立系统状态模型和故障模型来处理故障耦合性。DBN将系统状态模型和故障模型相结合，实现对系统状态和故障的联合推理。但是，在进行多维特征数据故障诊断时，由于存在大量变量节点，导致结构搜索空间庞大，增加了结构学习的计算复杂性和搜索时间，DBN的结构学习变得更加困难，前人对此进行了一些研究。孙美婷等[9]针对DBN结构学习是NP难的问题，提出基于改进遗传算法的DBN结构自适应学习方法，将最大互信息与遗传算法相结合，限制了DBN的结构搜索空间，并提高了遗传算法的全局搜索能力，但该方法容易陷入局部最优解。冷翠平等[10]基于变量间的基本依赖关系和依赖分析方法进行DBN结构学习，建立变量的关系草图，通过条件独立性除去多余的边，利用条件相对预测能力确定边的方向，但该方法无法动态调整结构搜索空间，导致潜在的最优解丢失。上述方法虽在一定程度上减小了网络的搜索空间，但在进行多维特征数据的网络构建时，存在易陷入局部最优解且难以动态调整网络结构搜索空间的问题。

R藤模型是一种用于多变量联合分布建模的统计方法，是基于Copula模型的一种扩展，常用于多维数据的依赖结构建模。现有Copula模型主要有C藤模型[11]和D藤模型[12]，上述模型虽然在一定程度上能够建立数据之间的依赖关系，但需对变量间相关性大小做出严格假设，因此，主要适用于数据量较少的场景。R藤Copula模型通过树形结构能够更灵活地建模多维特征数据变量之间的依赖关系，具有更好的可解释性和可拓展性。Liu等[13]提出一种基于R藤Copula模型的多传感器数据下发动机剩余使用寿命预测方法，采用R藤Copula模型对多传感器退化数据进行建模，提高了发动机剩余使用寿命的预测精度。Amin等[14]针对核电厂故障检测中非线性和非高斯性多维数据建模诊断精度较低的问题，提出一种采用R藤Copula模型的核电厂故障诊断多元概率框架，可提供更好的监测机制，提高故障诊断精度。涂青宇等[15]针对电网安全稳定运行中存在的风电功率不确定性问题，建立了动态化的R藤Copula模型，提出了区域风电集群功率区间预测方法，所提模型预测结果具有良好的可靠性。综上所述，R藤Copula模型在多维特征数据下能够对特征依赖性准确建模，可采用R藤Copula模型构建DBN结构，解决DBN多维数据下结构构建困难的问题，提高诊断性能。

基于以上分析，本文提出一种基于R藤Copula模型构建DBN的方法，主要贡献如下：

1）针对多维数据下有向无环图空间大、特征筛选难以保证候选父节点的全面性问题，结合互信息(mutual information, MI)、皮尔逊相关系数(Pearson correlation coefficient, PCC)和最大信息系数(maximal information coefficient, MIC) 3种相关性模型优点构造结构预测模型，从不同尺度度量数据特征的相关性，去除弱相关性特征，在保证候选父节点选取全面性的同时，降低有向无环图的空间大小。

2）针对多维数据下难以获得最优DBN结构的问题，在构建DBN的过程中引入R藤Copula模型，获得最优初始网络结构，在时间序列上展开获得最优DBN，提高了模型诊断性能。

1 DBN基本原理

静态贝叶斯网络(Bayesian network, BN)[16]的定义为：$\psi=(G,\theta)$，其中，G为随机变量的有向无环图，θ为网络的参数。模型表达了多维随机变量的联合概率分布。

$$P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P(X_i\mid \mathrm{Pa}(X_i)) \tag{1}$$

式中：$\mathrm{Pa}(X_i)$表示随机变量X_i的父节点。

DBN是一种概率图模型，在静态BN基础上引入时间信息，形成一种可描述动态过程的模型。DBN模型可以看作由若干静态BN时间片组成，每一时刻的状态保存在当前时间片中，每个时间片包含一组随机变量且所含随机变量属于同一子集。设$X=\{X_1,X_2,\cdots,X_n\}$为时变随机变量的集合，$X_i[t]$表示变量X_i在时刻t对应的随机变量，$X[t]$表示变量X在时刻t对应的随机变量集合。DBN结构可以表示为$B=(B_0,B_{\rightarrow})$，如图1所示。其中，B_0表示初始网络结构，定义DBN初始时刻的概率分布$P(X[0])$，$B_{\rightarrow}$表示时间片内和跨时间片连接的转移网络，定义从时刻$t-1$到时刻t的变量转移概率$P(X[t]|X[t-1])$。对于给定的DBN：$B=(B_0,B_{\rightarrow})$，该网络在$X[1],X[2],\cdots,X[T]$的概率分布可以表示为

$$P(X[1],\cdots,X[T])=\begin{cases}p_{B_0}(x[0]) & t=0\\ p_{B_{\rightarrow}}(x[t]|x[t-1]) & 0<t\leqslant T\end{cases} \tag{2}$$

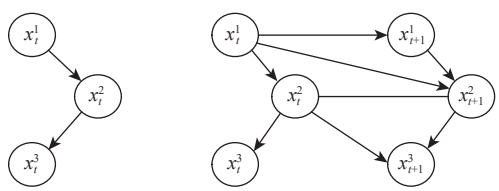


图1 DBN的初始网络和转移网络

Fig. 1 Initial network and transfer network of DBN network

2　结构预测模型

互信息[17]可用来衡量变量之间的相关性。对于离散变量, 熵表示每个变量的平均信息量$H(X)$, 可以表示为

$$H(X)=-\sum_{x\in X}p(x)\ln p(x) \tag{3}$$

式中: $p(x)$为变量x的边际概率函数。

X和Y的联合熵可以表示为

$$H(X,Y)=-\sum_{x\in X,y\in Y}p(x,y)\ln p(x,y) \tag{4}$$

式中: $p(x,y)$为x和y的联合概率密度函数。

互信息以熵的形式表示为

$$I(X;Y)=H(X)+H(Y)-H(X,Y) \tag{5}$$

因此, 2 个离散随机变量X和Y的互信息表达式为

$$I(X;Y)=\sum_{x\in X,y\in Y}p(x,y)\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{6}$$

PCC[18]是衡量 2 个变量X和Y之间线性相关性的方法, 取值为[−1,1]。当值接近 1 或者−1 时, 2 个变量X和Y的相关性越高, 当值越接近 0 时, 2 个变量之间相关性越低。PCC 表示为

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}} \tag{7}$$

式中: $\overline{X}$和$\overline{Y}$分别表示变量X和Y的平均值。

MIC[19]用于测量 2 个连续变量 X 和 Y 的线性或非线性相关强度。MIC 使用网格划分法计算 2 个变量之间的相关性。将变量 X 的范围划分为 a段, 变量Y的范围划分为b段, 将划分得到的网格定位表示为 L, 由于$a\times b$有多种划分方式, 采用不同划分方式下的最大值作为互信息值, 其表达式为

$$I_{\mathrm{MIC}}(X;Y)=\max I(X;Y\mid L) \tag{8}$$

式中: $\max I(X;Y|L)$为不同划分下的最大互信息值。

对不同的划分正则化后, 选择最大值作为 MIC 值, 2 个变量的 MIC 表示为

$$\mathrm{MIC}(X;Y)=\max_{a\times b<A}\frac{I_{\mathrm{MIC}}(X;Y)}{\log_2\min(a,b)} \tag{9}$$

式中: A为$a\times b$网格数的上限。

为更好地描述不同变量间的相关性, 本文采用加权平均方法来综合 3 种度量指标, 对互信息、PCC 和 MIC 分别进行归一化处理。为每个度量指标分配权重系数, 对每个特征计算互信息、PCC 和 MIC, 并将归一化值乘以对应的权重系数, 将所得值相加得到该特征的综合度量指标。综合度量指标 R 表示为

$$R=w_1I(X;Y)+w_2r+w_3\mathrm{MIC}(X;Y) \tag{10}$$

式中: w_1、w_2和w_3分别为 3 种相关性度量的权重系数。

由于齿轮箱故障数据中存在大量非线性数据, 结合先验知识与故障数据特点, 令互信息权重 w_1 为 0.4。PCC 适用于衡量变量之间的线性关系, 在齿轮箱故障数据中, 变量之间呈线性关系数据较少, 因此, PCC 权重系数 w_2 为 0.3。MIC 可发现齿轮箱故障数据中任何形式的相关性, 可使综合相似度指标对变量相关性分析更加全面, 确保不会错过重要的相关关系, 结合齿轮箱故障数据特点, 令其权重 w_3 为 0.3。根据综合度量指标的大小对特征进行排序, 去除相关性较弱的节点得到网络结构的候选特征。

3　R 藤模型

3.1　Copula 函数

Copula 函数[20]是一种考虑单变量边缘分布和联合分布的函数, 可对多变量联合概率分布进行组合。假设向量$\boldsymbol{\upsilon}=[x_1,x_2,\cdots,x_m]$包含了$m$个随机变量, 基于 Sklar 定理可得, 存在一个唯一的函数$C(u_1,u_2,\cdots,u_n)$, 满足[21]:

$$F(x_1,x_2,\cdots,x_m)=C(F_1(x_1),F_2(x_2),\cdots,F_m(x_m))=C(u_1,u_2,\cdots,u_n) \tag{11}$$

式中: $F(x_1,x_2,\cdots,x_m)$为$\boldsymbol{\upsilon}$的联合累计密度函数; $u_i=F_i(x_i)$, $F_i(x_i)$为对随机变量进行估计的边缘累计密度函数, 可表示为

$$F_i(x_i)=\int_0^{x_i}f_i(x_i)\mathrm{d}x_i \tag{12}$$

其中: $f_i(x_j)$为边缘概率密度函数, $i=1,2,\cdots,m$为随机变量的个数。

联合概率密度函数由 2 个不同部分构成: 单变量边缘概率密度空间和 Copula 密度区域。联合概率密度函数表示为

$$\xi(x_1,x_2,\cdots,x_m)=c(F_1(x_1),F_2(x_2),\cdots,F_m(x_m))\prod_{i=1}^{m}f_i(x_i) \tag{13}$$

式中: ξ为x的联合概率密度函数; c 表示 Copula 的概率密度函数。

采用 Copula 函数对变量之间的相关性进行参数估计, 定义如下:

$$c(F_1(x_1),F_2(x_2),\cdots,F_m(x_m))=\frac{\partial^m C(F_1(x_1),F_2(x_2),\cdots,F_m(x_m))}{\partial F_1(x_1),\partial F_2(x_2),\cdots,\partial F_m(x_m)} \tag{14}$$

3.2 R藤 Copula 模型

R藤模型将多维随机变量相关性问题转化为多个二维随机变量的相关性问题。d维随机变量的R藤 Copula 模型由$d-1$层树构成，用e表示2个随机变量相连的边，则第n层树边的集合为e_n，节点集合为N_n。$d=5$的R藤 Copula 模型结构如图2所示。

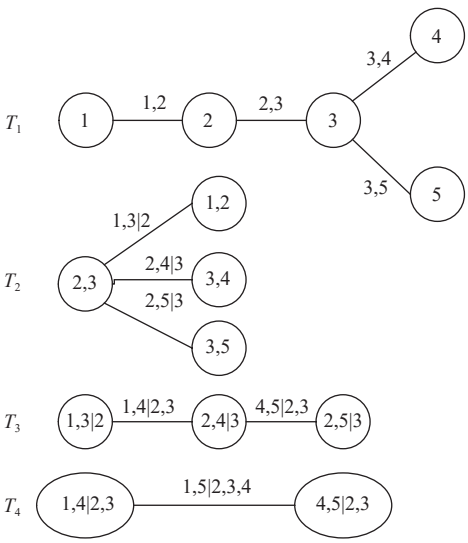


图2 5维随机变量的R藤 Copula 模型

Fig. 2 R-vine Copula model of 5-dimensional random variables

当d维随机变量R藤 Copula 模型结构确定后，则树形结构的联合概率密度为

$$f(x_1,x_2,\cdots,x_d)=\prod_{k=1}^{d}f_k(x_k)\cdot \prod_{l=1}^{d-1}\prod_{(i,j|D(e))\in e_n}c_{i,j|D(e)}\left(F\left(x_i\mid x_{D(e)}\right),F\left(x_j\mid x_{D(e)}\right)\right) \tag{15}$$

式中：$D(e)=\{c_1,c_2,\cdots,c_n\}$为条件值的集合，对应变量集合$x_D=\{x_{c_1},x_{c_2},\cdots,x_{c_n}\}$；$F(x_i|x_{D(e)})$表示已知$x_{D(e)}$的变量分布时，$x_i$服从的条件概率分布；$(i,j|D(e))$表示$e_n$中的边，$i$和$j$分别对应求解的变量$x_i$和$x_j$，其满足以下性质：

1）$n=1$时，$D(e)$为空集，i和j分别为边e的两端节点。

2）$n\neq 1$时，假设第$d-1$层树中相邻元素分别构成的集合为S_1和S_2，则$D(e)=S_1\cap S_2$。

R藤 Copula 模型通过式(14)来描述变量之间的相关性。在构建R藤 Copula 模型时，通过计算二元 Copula 函数对应的 Kendall 秩相关系数，将其作为最大生成树的权重，结合 Copula 概率密度函数对故障诊断变量之间相关性进行分析来确定第一层树结构。

Copula 概率密度函数公式为

$$c(u,v)=\int_0^1\int_0^1 C(u,v)\mathrm{d}u\mathrm{d}v \tag{16}$$

Kendall 秩相关系数τ为

$$\tau=4\int_0^1\int_0^1 C(u,v)\mathrm{d}C(u,v)-1 \tag{17}$$

式中：$C(u,v)$为 Copula 函数；$u=K(z),v=L(y)$，$K(z)$和$L(y)$为随机变量的分布函数。随机变量z和y发生变化时，式(17)计算同时间段下的变量z和y的秩相关系数。

Kendall 秩相关系数τ表示二元随机变量(X_1,X_2)和 Copula 函数是否一致的概率差异度量，$\tau\in[-1,1]$，$\tau=-1$表示完全负相关，$\tau=1$表示变量完全正相关，$\tau=0$表示变量之间不相关，τ越大，随机变量的相关性越强。

确定第一层树结构后，可根据性质1)和性质2)确定其他层树的结构，从而构建出R藤 Copula 模型。R藤 Copula 构建的网络是无向图，缺少对变量间因果关系的描述。本文采用传递熵方法获得变量间因果关系构建B_0结构。R藤 Copula 树模型构建过程如图3所示。

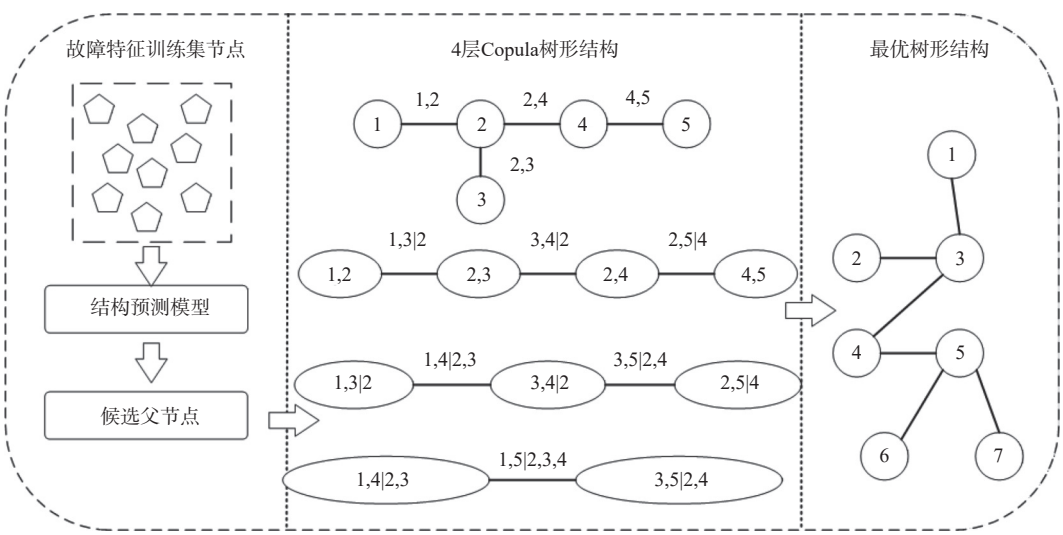


图3 R藤 Copula 树模型构建过程

Fig. 3 Building process of R vine Copula tree model

4　R 藤 Copula-DBN 网络的构建

1) 建立初始B_0网络。采用生成 R 藤 Copula 结构的首层树结构作为B_0结构。采用传递熵确定变量之间的因果关系。信息熵能够度量单个变量的不确定度, 被观测的变量不确定性越高, 信息熵越大, 不确定性越小, 信息熵越小。其公式表达式如式(3)所示。

传递熵可表示 2 个变量之间的信息传递, 传递熵表达式为

$$T_{Y\to X}=H(X_t\mid X_{t-1})-H(X_t\mid X_{t-1},Y_{t-1}) \tag{18}$$

$$T_{X\to Y}=H(Y_t\mid Y_{t-1})-H(Y_t\mid Y_{t-1},X_{t-1}) \tag{19}$$

式中: $T_{Y\to X}$为 Y到 X的传递熵, 即 Y能够对 X传递信息的大小; $T_{X\to Y}$为 X到 Y的传递熵, 即 X能够对 Y传递信息的大小; $H(X_t\mid X_{t-1})$、$H(Y_t\mid Y_{t-1})$、$H(X_t\mid X_{t-1},Y_{t-1})$和$H(Y_t\mid Y_{t-1},X_{t-1})$为对应变量的条件熵, 表示已知变量先验信息的条件下, 下一时刻变量的不确定性。若 $T_{Y\to X}>T_{X\to Y}$, 则 Y为 X的父节点, X为 Y的子节点, 确定出网络结构的因果关系, 构建出完整的初始网络B_0。

2) 建立转移网络 $B_{\to}$。转移网络 $B_{\to}$是构建 DBN 结构过程中的重要一环, 通过$B_{\to}$对不同时间片之间节点状态进行连接。转移网络$B_{\to}$服从马尔可夫过程, 即状态空间 Q 所描述故障诊断过程中涉及的所有随机变量、时间步长和随机变量相关的条件概率$P(Q_{t+1}|Q_t)$只与当前时刻 t 时的随机变量有关, 而与所有过去的状态无关。通过将初始网络B_0按历史数据时间序列进行扩展, 并将相邻时刻的状态变量相互连接后按时间先后顺序连接为有向边, 建立故障诊断网络的转移网络$B_{\to}$。

5　齿轮箱故障诊断

本文所提基于 R 藤 Copula-DBN 的齿轮箱故障诊断方法流程如图 4 所示。

具体步骤如下:

步骤 1　利用振动传感器采集齿轮箱故障的振动信号, 进行预处理, 按 3∶1 划分为测试集和训练集。

步骤 2　提取训练集振动信号的时域特征。

步骤 3　采用结构预测模型 (式 (6)、式 (7) 和式 (9)) 去除所提取特征中综合相关性较低的特征, 将获得的特征属性进行等宽离散。

步骤 4　通过式 (16) 和式 (17) 计算 Copula 函数与 Kendall 秩相关系数, 将计算结果作为最大生成树的权重。由式 (15) 获得 R 藤 Copula 模型的树形结构, 采用树形结构的第一层作为初始网络B_0的无向无环图。

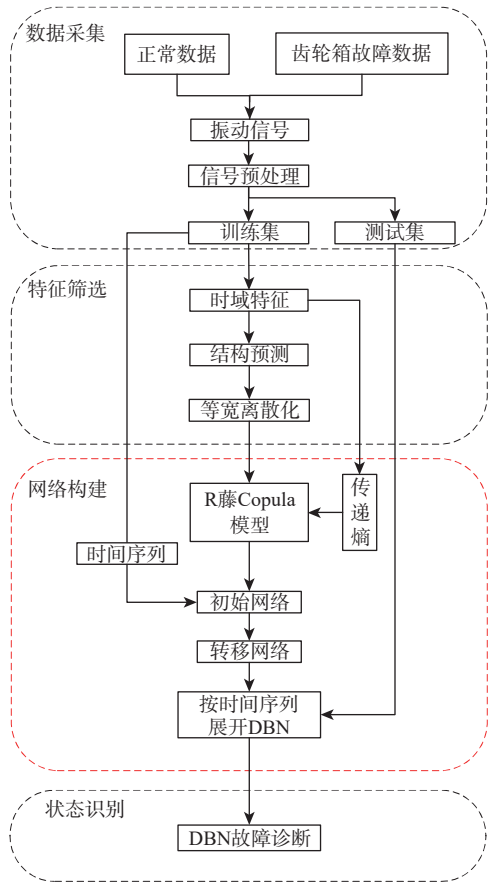


图 4　齿轮箱故障诊断方法流程

Fig. 4　Flow chart of gearbox fault diagnosis method

步骤 5　通过式 (18) 和式 (19) 计算特征之间传递熵, 确定节点之间的因果关系, 获得初始网络B_0结构。

步骤 6　根据马尔可夫过程将初始网络B_0在时间序列延拓, 将相邻时刻变量之间按时间顺序连接为有向边, 建立转移网络$B_{\to}$, 形成 DBN。

步骤 7　获得 DBN 网络展开拓扑结构后, 采用 EM 算法[22] 计算网络结构的参数θ, 得到完整的 DBN, 由前向后向算法[23] 推断发生故障的节点。

6　实验结果与分析

6.1　数 据 集

为验证本文所提 R 藤 Copula-DBN 方法的有效性, 采用东南大学齿轮箱数据集[24] 进行实验, 实验平台如图 5 所示。

该数据集通过传动系统动力学模拟器 DDS 对振动信号进行实时采集, 信号以 20 000 Hz 采样频率采样, 每种状态采样 1 048 560 个数据点。设置了转速负载为 20 Hz-0 V 和 30 Hz-2 V 2 种工况。该实验平台主要由电机控制器、电机、行星齿轮箱、制动器、制动控制器组成。数据集包含 5 种故障状态, 其故障描述如表 1 所示。

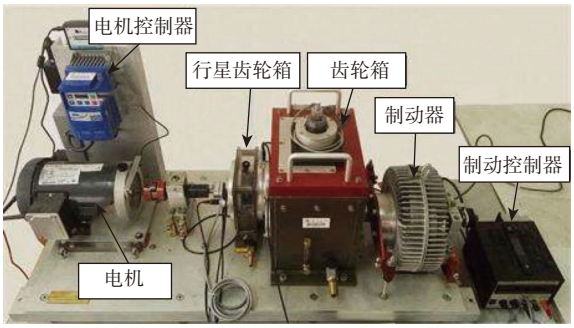


图 5 实验模拟平台

Fig. 5 Experimental simulation platform

表 1 故障编码及描述

Table 1 Fault coding and description

故障类型	故障描述	故障编码
正常	健康运行状态	1
缺损	齿轮出现裂纹	2
断齿	齿轮上出现断齿	3
齿根磨损	齿根处出现裂纹	4
齿面磨损	齿轮表面出现磨损	5

6.2 特征提取及网络构建

表 2 列出了正常状态和 4 种故障状态的部分时域特征，将信号时域特征用于齿轮箱故障的诊断和分类。

表 2 部分信号时域特征参数

Table 2 Part of signal time domain characteristic parameters

故障类型	峰值/(m·s^{-2})	峰峰值/(m·s^{-2})	均值/(m·s^{-2})	方差/(m·s^{-2})	标准差/(m·s^{-2})	峭度	均方根/(m·s^{-2})	偏度
正常	0.447 80	0.434 04	−0.337 16	0.000 89	0.029 80	5.600 92	0.142 87	−0.290 32
缺损	0.180 89	0.032 82	−0.165 83	0.000 02	0.004 54	3.072 51	0.642 12	−0.047 03
断齿	0.690 75	0.094 66	−0.641 82	0.000 38	0.019 60	1.991 41	0.642 12	−0.047 03
齿根磨损	0.451 62	0.095 55	−0.403 61	0.000 32	0.018 01	2.246 31	0.642 12	−0.047 03
齿面磨损	0.267 83	0.081 84	−0.228 25	0.000 17	0.013 20	2.738 07	0.122 38	0.057 45

故障类型	平均幅值/(m·s^{-2})	方根幅值/(m·s^{-2})	波形因子	峰值因子	脉冲因子	裕度因子	余隙因子
正常	0.336 36	0.337 48	1.003 89	1.322 99	1.328 15	1.330 83	3.908 61
缺损	0.167 83	0.165 71	1.000 37	1.094 01	1.090 81	1.091 02	6.572 73
断齿	0.541 82	0.641 67	1.000 46	1.075 73	1.076 23	1.076 48	1.075 26
齿根磨损	0.404 61	0.403 41	1.000 99	1.117 81	1.118 93	1.119 48	2.766 74
齿面磨损	0.218 06	0.228 06	1.001 67	1.174 17	1.173 43	1.174 41	5.123 47

对结构预测模型特征筛选有效性进行验证，采用 PCC、互信息和最大相关最小冗余方法与结构预测方法对特征筛选进行对比。采用贝叶斯信息准则 (Bayesian information criterion, BIC)、搜索空间和特征数量作为对比指标。BIC 准则是一种常用于模型选择的评价指标，采用 BIC 准则来评估不同特征数量下所得网络结构对数据的解释能力，其值越小，表明所得模型结构越好。实验结果如表 3 所示。表中：MRMR 为最大相关最小冗余，SP 为结构预测。

表 3 特征筛选有效性验证

Table 3 Validity verification of feature screening

特征筛选方法	BIC	搜索空间	特征数量
无筛选	$-1.886\,7\times10^{4}$	2^{201}	15
互信息	$-2.145\,2\times10^{4}$	2^{102}	10
PCC	$-2.037\,5\times10^{4}$	2^{154}	13
MRMR	$-1.943\,1\times10^{4}$	2^{175}	14
SP	$-2.610\,2\times10^{4}$	2^{123}	11

结构预测模型与未使用筛选方法相比，有效降低了特征数量和搜索空间大小。与其他特征筛选方法相比，结构预测模型在搜索空间和特征数量上略高于互信息方法，但获得了最小的 BIC 分数，原因是：互信息在特征筛选期间出现了误删有效特征的现象。综合 3 种指标，结构预测模型在有效降低搜索空间和特征数量的前提下获得了较好的网络结构，具有更全面的特征筛选能力。

通过不同综合相似度 R 筛选特征，比较所得特征构建结构的 BIC 分数，对综合相似度强弱进行判定。为使结果清晰，采用 BIC 分数的绝对值作为指标，所得结果如图 6 所示。当以综合相似

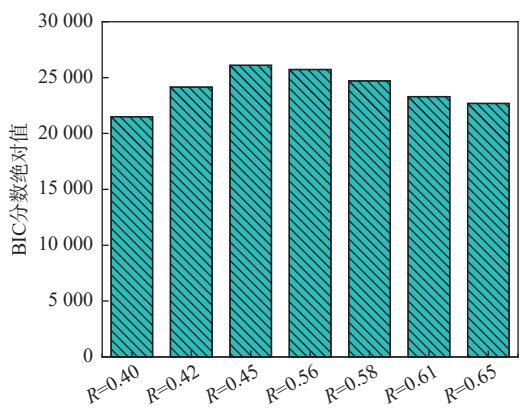


图 6 不同综合相似度特征筛选所得结构 BIC 分数绝对值

Fig. 6 Structure BIC score absolute values obtained by screening different comprehensive similarity features

度为 0.45 筛选特征时, 所得结构的 BIC 分数绝对值最高, 即网络结构最好。因此, 判定综合相似度 0.45 以下为弱相关特征, 大于 0.45 为强相关特征。

通过计算故障样本的 15 种时域特征, 采用结构预测模型进行特征选取, 分别用式 (6)、式 (7) 和式 (9) 计算特征变量的相关性, 采用综合相似度指标将节点的相关性大小进行排序。特征节点的相关性如图 7 所示。

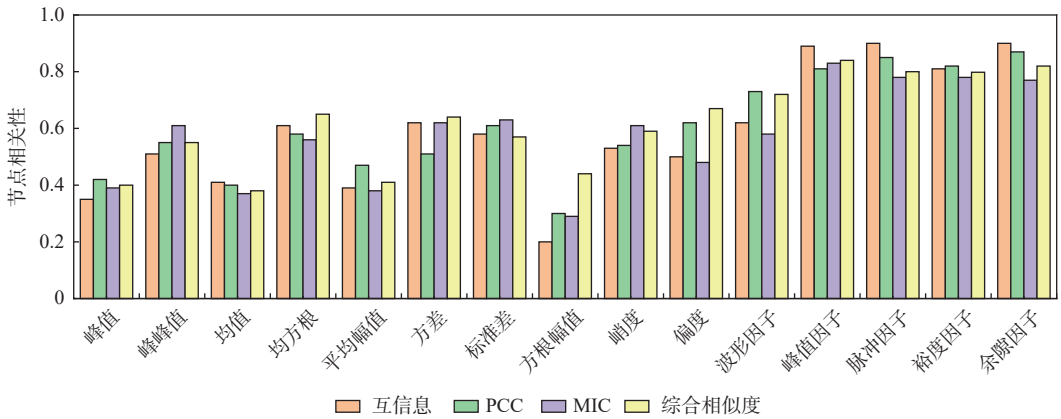


图 7　特征节点相关性

Fig. 7　Feature node correlation

通过综合相似度对特征进行筛选后, 保留的特征属性为峰峰值、均方根、方差、标准差、峭度、偏度、波形因子、峰值因子、脉冲因子、裕度因子、余隙因子。采用以上特征计算 Kendell 秩相关系数作为最大生成树的权重构建最优 R 藤树形结构。所构建的 R 藤 Copula-DBN 结构如图 8 所示。

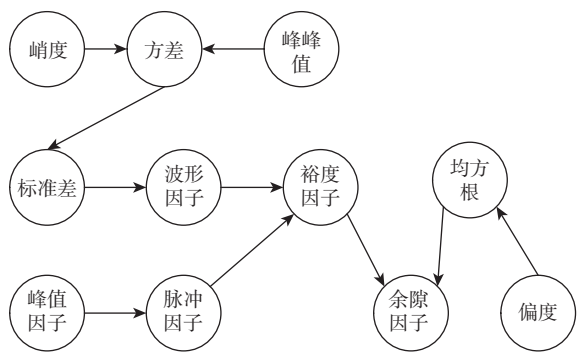


图 8　R 藤 Copula-DBN 模型第一层结构

Fig. 8　The first layer structure of R vine Copula-DBN model

由于所构建 R 藤 Copula-DBN 模型第一层结构为无向无环图, 采用传递熵计算 R 藤 Copula-DBN 结构因果关系, 确定初始网络结构 B_0, 并将初始网络结构在时间序列上展开构成 DBN, 通过采样频率 20 000 Hz 和采样周期 10 s 将故障数据分为 10 个时间片, 构造的 DBN 如图 9 所示。图中: 变量之间的箭头表示 2 个节点之间的因果关系, 节点之间的弧表示与下个时间片之间连接的转移概率。

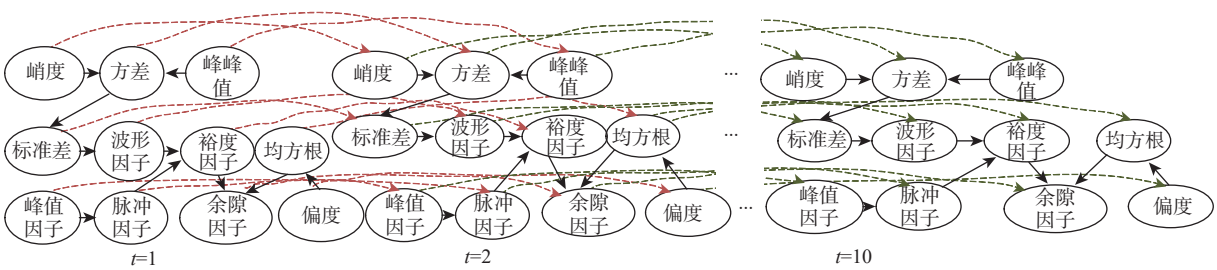


图 9　齿轮箱 DBN

Fig. 9　Gearbox DBN

6.3　故障诊断与分析

采用混淆矩阵对本文构建的 R 藤 Copula-DBN 故障诊断模型进行验证, 并与 DBN 模型进行对比。混淆矩阵的横轴代表故障预测的分类, 纵轴代表故障的实际分类, 对角线为每类故障的诊断精度。由图 10 和图 11 的混淆矩阵可以看出, 由 R 藤 Copula 构建初始网络的 DBN 对齿轮箱故障有较好的分类效果。

将本文方法与 BN、一阶马尔可夫 DBN(first-order Markov DBN, FM-DBN)、故障树 DBN(fault tree-DBN, FT-DBN)、遗传神经网络(genetic neural network, GNN)、深度对抗卷积神经网络(deep adversarial convolutional neural network, DACNN)进行实

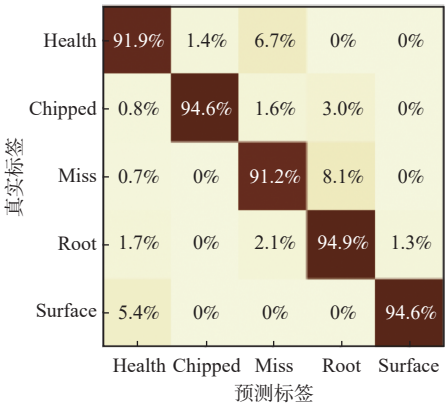


图 10 R 藤 Copula-DBN 混淆矩阵

Fig. 10 Confusion matrix of R vine Copula DBN

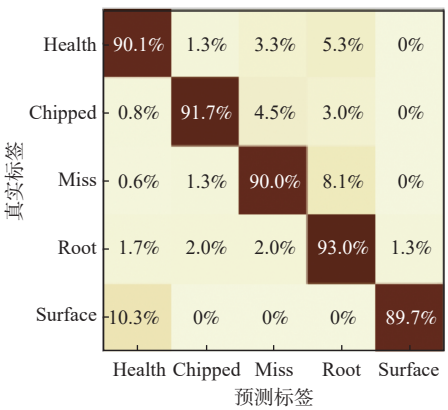


图 11 DBN 混淆矩阵

Fig. 11 Confusion matrix of DBN

验对比。

采用准确率、F_1、召回率及模型运行时间作为方法评判指标。准确率是所有正确分类样本占总样本数的比例，用于评估模型整体性能，其值越高，表示模型预测中正确样本数量越多，表明模型能够正确分类大部分样本。F_1 是用于衡量分类模型的指标，综合考虑准确率和召回率，表达模型的准确性和完整性，以 F_1 作为评判指标，F_1 值越高，表示模型在高精确率和高召回率之间取得了较好的平衡，模型具有更好的分类性能。召回率是评估分类模型识别真正例的指标，其值越高，表示模型在捕捉正例方面表现越好。每种方法采用 10 次实验平均值作为诊断指标结果，所得结果如表 4 所示。

$$A=\frac{Q+W}{S} \tag{20}$$

$$F_\beta=\frac{(\beta^2+1)\rho\varphi}{\beta^2\rho+\varphi} \tag{21}$$

$$F_1=\frac{2\rho\varphi}{\rho+\varphi} \tag{22}$$

$$\varphi=\frac{Q}{Q+N} \tag{23}$$

表 4 1 000 样本下各模型的诊断结果及运行时间对比

Table 4 Comparison of diagnostic results and running time of each model under 1 000 samples

模型	准确率/%	F_1/%	召回率/%	运行时间/s
R藤Copula-DBN	93.6	93.4	93.3	8.3
DBN	90.7	90.4	90.2	8.1
BN	88.5	88.4	88.1	8.0
FM-DBN	91.7	90.3	91.2	8.4
FT-DBN	90.6	91.2	89.2	8.6
GNN	91.6	90.8	89.9	6.7
DACNN	92.8	91.6	90.8	7.0

式中：A 为准确率；Q 为真正例数量；W 为真负例数量；S 为总的样本数量；F_β 为加权 F 分数；ρ 表示预测为正例的样本中实际为正例的比例；φ 为召回率，即实际为正例的样本模型预测为正例的比例；β 为准确率和召回率的调和率；N 为假反例的数量。

选取 5 000 个采样点作为一个训练样本，每种状态采集 200 个独立样本。由表 4 可得，采用 R 藤 Copula-DBN 模型取得了最高的准确率、F_1 值和召回率，表明在分类任务中具有较高的准确性和较好的综合性能。R 藤 Copula-DBN 模型相比于传统 DBN 建模，分类准确率、F_1 值、召回率分别提高了 2.9%、3%、3.1%。由于受到异步性处理的限制和高斯核函数假设的限制，R 藤 Copula-DBN 与 FM-DBN 相比，分类指标分别提高了 1.9%、3.1%、2.1%。由于 FT-DBN 依赖一系列的建模假设，当面对复杂动态系统时，模型分类准确率下降，R 藤 Copula-DBN 与 FT-DBN 相比分类准确率提高了 3%、2.2%、4.1%。由于 GNN 在数据量较大时难以搜索最佳网络结构和参数，R 藤 Copula-DBN 比 GNN 的分类指标分别高 2%、2.6%、3.4%。由于 DACNN 易受数据不确定性影响，不确定性会干扰模型的决策，R 藤 Copula-DBN 比 DACNN 分类指标高 0.8%、1.8%、2.5%。由模型运行时间可以看出，在采用 R 藤 Copula-DBN 模型构建 DBN 结构时，运行速度相比于 DBN 慢 0.2 s，由于 R 藤 Copula-DBN 在进行构建时相比 BN 结构更加复杂，所涉及推理和运算远高于 BN，运行时间相比于 BN 慢 0.3 s。

分别采用 1 200、800 和 600 个样本验证所提模型在不同数据量大小下的诊断性能，所得实验结果如表 5～表 7 所示。由结果可知，当数据量下降时，R 藤 Copula-DBN 模型的诊断准确率有所上升，在小数据量下，相比于其他 6 种模型，R 藤 Copula-DBN 模型的诊断准确率最高。当数据量增大时，在 DBN 模型中引入 R 藤 Copula 模型能使 DBN 网络更精准地构建网络结构，但其对故障诊断的准确

表 5　1 200 样本下各模型的诊断结果及运行时间对比

Table 5　Comparison of diagnostic results and running time of each model under 1 200 samples

模型	准确率/%	F_1/%	召回率/%	运行时间/s
R藤Copula-DBN	93.8	93.7	93.4	9.73
DBN	89.1	87.3	89.2	9.47
BN	87.1	86.2	86.1	8.43
FT-DBN	91.3	89.4	90.6	9.74
FM-DBN	89.5	90.7	88.3	9.6
GNN	94.0	93.8	93.6	7.93
DACNN	94.6	94.3	94.1	8.3

表 6　800 样本下各模型的诊断结果及运行时间对比

Table 6　Comparison of diagnostic results and running time of each model under 800 samples

模型	准确率/%	F_1/%	召回率/%	运行时间/s
R藤Copula-DBN	95.4	94.2	93.4	7.8
DBN	91.2	92.1	90.2	7.6
BN	87.4	88.3	88.5	7.3
FT-DBN	91.3	92.4	90.4	8.8
FM-DBN	91.5	92.2	91.7	9.0
GNN	89.8	90.1	91.3	6.4
DACNN	90.3	91.1	92.3	6.9

表 7　600 样本下各模型的诊断结果及运行时间对比

Table 7　Comparison of diagnostic results and running time of each model under 600 samples

模型	准确率/%	F_1/%	召回率/%	运行时间/s
R藤Copula-DBN	95.6	94.5	93.6	7.4
DBN	93.3	92.1	91.6	7.3
BN	88.5	87.1	86.7	7.0
FT-DBN	92.1	91.2	90.1	8.1
FM-DBN	92.6	92.4	91.8	8.8
GNN	88.2	87.8	86.3	6.2
DACNN	89.2	90.3	89.2	6.5

率略低于 GNN 和 DACNN 模型。由于 R 藤 Copula-DBN 模型在处理复杂依赖和动态时间序列需要更多的计算步骤, 增加了运行时间, 导致 R 藤 Copula-DBN 模型运行速度低于 GNN 和 DACNN 模型。但在处理小数量下具有不确定性数据时, GNN 和 DACNN 模型的准确率和召回率低于 R 藤 Copula-DBN 模型, R 藤 Copula-DBN 在增加数据量时, 与其他贝叶斯网络模型相比, 依然能够保证模型诊断的准确率。

为验证模型结构与数据的拟合性, 本文取 10 次独立仿真实验的平均结构汉明距离 (average structural Hamming distance, ASHD) 进行验证, 值越小, 表示网络的结构性能越好, 所得实验结果如图 12 所示。在 5 种模型中, R 藤 Copula-DBN 的平均结构汉明距离最小, 所获得的网络结构性能最优。由于 R 藤 Copula-DBN 采用结构预测模型对故障特征节点进行了筛选, 所得网络平均结构汉明距离低于 DBN 和 FM-DBN。FT-DBN 在进行结构搭建时较为依赖先验知识, 易受主观因素影响, 因此, 平均结构汉明距离远高于 R 藤 Copula-DBN。由于 BN 难以处理动态系统, 无法搭建动态系统准确的网络结构, R 藤 Copula-DBN 平均结构汉明距离远低于 BN。综上所述, R 藤 Copula-DBN 网络所搭建的网络结构与数据的拟合性最优。

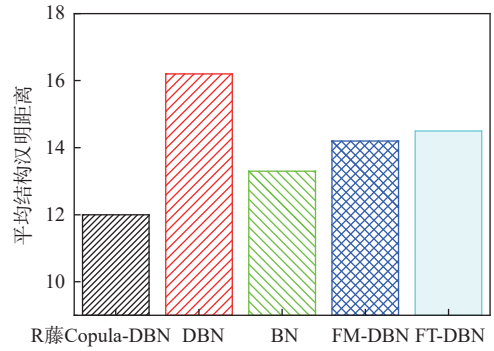


图 12　平均结构汉明距离

Fig. 12　Average structural Hamming distance

7　结　论

本文通过结构预测模型对数据特征进行筛选, 减小网络的构建规模, 采用 R 藤 Copula 模型构建 DBN 的初始网络结构。由齿轮箱故障数据进行实验验证, 本文方法对多维数据下齿轮箱故障诊断有较好的表现。

1) 本文方法在不同数据量下具有较好的准确率。1 000 数据量下, 准确率为 93.6%, F_1 为 93.4%, 召回率为 93.3%; 800 数据量下, 准确率为 95.4%, F_1 为 94.2%, 召回率为 93.4%; 600 数据量下, 准确率为 95.6%, F_1 为 94.5%, 召回率为 93.6%。

2) 与其他 DBN 方法相比, 本文方法构建的网络结构和数据的拟合度更高。

3) 所构建的 DBN 结构, 通过结构预测模型对故障特征进行筛选, 减小了网络的构建规模, 但由于 R 藤 Copula 模型的引入, 导致故障诊断时间增加。

参考文献 (References)

[1] BHARDWAJ U, TEIXEIRA A P, SOARES C G. Reliability prediction of an offshore wind turbine gearbox[J]. Renewable Energy, 2019, 141: 693-706.

[2] 魏莎, 韩勤锴, 褚福磊. 考虑不确定性因素的齿轮系统动力学研究综述[J]. 机械工程学报, 2016, 52(1): 1-19.
WEI S, HAN Q K, CHU F L. Review on research of uncertain dynamics for gear system[J]. Journal of Mechanical Engineering,

2016, 52(1): 1-19(in Chinese).

[3] CHEN R, HUANG X, HU X, et al. Planetary gearbox fault diagnosis technique based on multi-source information deep fusion[J]. Journal of Vibration Engineering, 2020, 33(5): 1094-1102.

[4] GU Y K, ZHANG M, ZHOU X Q. Fault diagnosis of gearbox based on improved DUCG with combination weighting method[J]. IEEE Access, 2019, 7: 92955-92967.

[5] GUO S, YANG T, HUA H C, et al. Coupling fault diagnosis of wind turbine gearbox based on multitask parallel convolutional neural networks with overall information[J]. Renewable Energy, 2021, 178: 639-650.

[6] HUANG N T, CHEN Q Z, CAI G W, et al. Fault diagnosis of bearing in wind turbine gearbox under actual operating conditions driven by limited data with noise labels[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 3502510.

[7] 王进花, 岳亮辉, 曹洁, 等. 基于随机变分推理贝叶斯神经网络的发电机轴承故障诊断[J]. 控制与决策, 2023, 38(4): 1015-1021.
WANG J H, YUE L H, CAO J, et al. Fault diagnosis of generator bearing based on stochastic variational inference Bayesian neural network[J]. Control and Decision, 2023, 38(4): 1015-1021(in Chinese).

[8] DEAN T, KANAZAWA K. A model for reasoning about persistence and causation[J]. Computational Intelligence, 1989, 5(2): 142-150.

[9] 孙美婷, 刘彬. 基于改进遗传算法的 DBN 结构自适应学习算法[J]. 计量学报, 2021, 42(1): 91-99.
SUN M T, LIU B. DBN structure adaptive learning algorithm based on improved genetic algorithms[J]. Acta Metrologica Sinica, 2021, 42(1): 91-99(in Chinese).

[10] 冷翠平, 王双成, 王辉. DBN 结构学习的依赖分析方法研究[J]. 计算机工程与应用, 2011, 47(3): 51-53.
LENG C P, WANG S C, WANG H. Study on dependency analysis method for learning dynamic Bayesian network structure[J]. Computer Engineering and Applications, 2011, 47(3): 51-53(in Chinese).

[11] FANG C, XU Y L, LI Y L. Optimized C-vine Copula and environmental contour of joint wind-wave environment for sea-crossing bridges[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2022, 225: 104989.

[12] SUN F Q, FU F Y, LIAO H T, et al. Analysis of multivariate dependent accelerated degradation data using a random-effect general Wiener process and D-vine Copula[J]. Reliability Engineering & System Safety, 2020, 204: 107168.

[13] LIU S J, JIANG H. Engine remaining useful life prediction model based on R-vine Copula with multi-sensor data[J]. Heliyon, 2023, 9(6): e17118.

[14] AMIN M T, YAO Y T, YU J, et al. Probabilistic monitoring of nuclear plants using R-vine Copula[J]. Annals of Nuclear Energy, 2023, 190: 109867.

[15] 涂青宇, 苗世洪, 林毓军, 等. 基于动态 R 藤 Copula 模型的区域风电集群超短期功率区间预测方法[J]. 高电压技术, 2022, 48(2): 456-470.
TU Q Y, MIAO S H, LIN Y J, et al. Ultra-short-term interval forecasting method for regional wind farms based on dynamic R-vine Copula model[J]. High Voltage Engineering, 2022, 48(2): 456-470 (in Chinese).

[16] MALONE B, KANGAS K, JÄRVISALO M, et al. Empirical hardness of finding optimal Bayesian network structures: algorithm selection and runtime prediction[J]. Machine Learning, 2018, 107(1): 247-283.

[17] PANINSKI L. Estimation of entropy and mutual information[J]. Neural Computation, 2003, 15(6): 1191-1253.

[18] MU Y S, LIU X D, WANG L D. A Pearson's correlation coefficient based decision tree and its parallel implementation[J]. Information Sciences, 2018, 435: 40-58.

[19] ZHANG Y H, ZHANG W S, XIE Y. Improved heuristic equivalent search algorithm based on maximal information coefficient for Bayesian network structure learning[J]. Neurocomputing, 2013, 117: 186-195.

[20] YU D M, GHADIMI N. Reliability constraint stochastic UC by considering the correlation of random variables with Copula theory[J]. IET Renewable Power Generation, 2019, 13(14): 2587-2593.

[21] AMIN M T, KHAN F, AHMED S, et al. Risk-based fault detection and diagnosis for nonlinear and non-Gaussian process systems using R-vine Copula[J]. Process Safety and Environmental Protection, 2021, 150: 123-136.

[22] WANG L L, XIN T, YANLOU L. An improved parameter-estimating method in Bayesian networks applied for cognitive diagnosis assessment[J]. Frontiers in Psychology, 2021, 12: 665441.

[23] CHEN H, REN J. Structure-variable hybrid dynamic Bayesian networks and its inference algorithm[C]//Proceedings of the 24th Chinese Control and Decision Conference. Piscataway: IEEE Press, 2012: 2815-2820.

[24] SHAO S Y, MCALEER S, YAN R Q, et al. Highly accurate machine fault diagnosis using deep transfer learning[J]. IEEE Transactions on Industrial Informatics, 2018, 15(4): 2446-2455.

Gearbox fault diagnosis based on R-vine Copula-DBN

WANG Jinhua[1, *]，LIU Zhengqi[1]，CAO Jie[1, 2, 3]，LIU Yunqiang[1]，CHEN Li[2]

(1. School of Electrical Engineering and Information Engineering，Lanzhou University of Technology，Lanzhou 730050，China；
2. School of Information Engineering，Lanzhou City University，Lanzhou 730050，China；
3. Gansu Manufacturing Information Engineering Research Center，Lanzhou 730050，China)

Abstract：Low diagnostic accuracy results from the wide set of directed acyclic graphs that must be searched when doing structure learning on dynamic Bayesian starting networks under multidimensional input. Conventional approaches find it challenging to find the best structure. In this paper, a method is proposed to combine the R-vine Copula model with a dynamic Bayesian network (DBN) for fault diagnosis. First, the network structure space is made smaller by using the structure prediction model to filter the retrieved features and identify nodes with high correlation. Then, the first-layer tree structure of the R-vine Copula model is used combined with the transfer entropy method to construct the initial network of dynamic Bayesian network, and the DBN of the initial network is built according to the Markov process in time series for fault diagnosis, which solves the problem that it is difficult to obtain the optimal structure in the network construction under multiple features. The gearbox data of Southeast University is used for verification, and the comparison results show that the method can better learn the DBN structure, and the fit between the data and the model is high, and good diagnostic results can be obtained in fault diagnosis.

Keywords：fault diagnosis；dynamic Bayesian network；R-vine Copula；structure prediction model；gearbox

Received：2023-11-30；**Accepted**：2024-03-12；**Published Online**：2024-04-25 13：50

URL：link.cnki.net/urlid/11.2625.V.20240425.0834.001

Foundation items：National Natural Science Foundation of China (62063020); National Key Research and Development Program of China (2020YFB1713600); Natural Science Foundation of Gansu Province (20JR5RA463)

* **Corresponding author**. E-mail：wjh0615@lut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0845

横向非均匀积水对飞机滑水偏出跑道行为影响

李岳[1]，张耀华[1]，蔡靖[1, 2, *]，任怡楠[1]

(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：机场道面积水受多重因素影响，非均匀分布特征显著，可导致飞机偏出跑道事故，仅以单一积水厚度指标表征能力不足。对此，建立飞机滑水偏出跑道行为理论模型，基于美国足尺加载试验开展飞机轮组滑水仿真分析，分段量化机场道面横向非均匀积水分布，对比两侧飞机轮组摩阻作用差异。提出一种非均匀积水临界长度评价指标，通过ADAMS整机运动学分析探讨飞机滑行轨迹及速度变化与评价指标关联，验证其合理性与适用性。结果表明：飞机在横向非均匀积水机场道面着陆时，两侧飞机轮组所受摩阻作用存在明显时程差异；飞机处于非对称偏转力交错影响之下，偏出跑道风险贯穿着陆滑行全过程；3组案例分析所得非均匀积水临界长度与理论值相差在6.6%以下，该指标可考察由横向非均匀积水引发的偏转作用累积影响，为起降安全保障及机场道面积水控制提供参考。

关 键 词：飞机轮组；轮胎滑水；非均匀积水；偏出跑道；ADAMS
中图分类号：V351.11; U8
文献标志码：A **文章编号**：1001-5965(2026)03-0698-08

降水导致机场道面积水湿滑，飞机在此类污染道面条件下起降时发生航空事故的概率约为理想干燥道面条件的8倍[1]。积水道面起降安全保障问题受到民航界的普遍重视。

由飞机轮胎滑水引起的航空事故可分为2类：冲出跑道(over-run, OR)事故和偏出跑道(veer-off, VO)事故。以往研究大多围绕OR事故问题展开[2-5]，主要目标是求解临界滑水速度指标v_p，量化保障飞机安全起降滑行速度区间范围。对此，《航空承运人湿跑道和污染跑道运行管理规定》中考察OR类事故滑水风险，污染道面条件飞机滑行距离可延长15%，以13 mm作为允许起降道面最大积水厚度条件[6]。需指出，上述结论建立在两侧主起落架所受道面摩阻力作用对称的前提下，在实际运行中难以充分保证。单一积水厚度指标对实际机场道面积水分布情况表征能力有限，不利于机场道面积水的科学控制。

当前，民航机型主起落架间距多在6 m以上，非均匀积水道面两侧飞机轮组摩阻力差异显著，飞机可在偏转力作用下产生横向运动趋势[7]，引发VO类滑水事故。在该领域中，研究工作主要针对车辆行驶安全问题展开。Fwa等[8]采用计算流体力学方法进行轮胎与水流的三维模拟，分析路面纵向和横向变形对v_p影响，提出以v_p和制动距离作为积水分布评价依据。季天剑等[9]基于人工神经网络方法建立路面水膜厚度预测模型，仿真计算路面附着系数、行车速度与积水厚度因素关联。徐进等[10]认为汽车两侧轮胎因滑水造成的速度差异是汽车偏出路面事故的主要原因，积水路面行驶时间越长，事故发生几率越高，应考察路面积水长度对车辆偏出行为的影响。考虑到飞机与车辆在轮组构型、轮胎规格、行驶速度及运动学特征方面的差异，上述结论难以适用于飞机起降安全技术领域。

对此，本文建立了飞机滑水偏出跑道行为理论

收稿日期：2024-01-02；**录用日期**：2024-02-23；**网络出版时间**：2024-03-09 13:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20240307.1151.001
基金项目：国家自然科学基金面上项目(52472369)；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-2025-01)
* **通信作者**. E-mail：caijing75@163.com

引用格式：李岳，张耀华，蔡靖，等. 横向非均匀积水对飞机滑水偏出跑道行为影响[J]. 北京航空航天大学学报，2026，52(3)：698-705.
LI Y，ZHANG Y H，CAI J，et al. Influence on veer-off behavior of aircraft hydroplaning due to transverse uneven water accumulation[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：698-705 (in Chinese).

模型, 基于美国足尺加载试验开展飞机轮组滑水仿真分析, 分段量化机场道面横向非均匀积水分布, 对比非均匀积水条件下两侧飞机轮组摩阻作用时程差异。在此基础上, 提出了非均匀积水临界长度评价指标, 通过 ADAMS 整机运动学分析探讨飞机滑行轨迹及速度变化与评价指标的关联, 为起降安全保障及机场道面积水控制提供分析基础。

1　飞机滑水偏出跑道行为理论模型

理想滑行条件下, 飞机两侧轮组受力同步且对称, 机身自重 $\boldsymbol{G}$、机翼升力 $\boldsymbol{L}$ 与道面支撑力 $\boldsymbol{R}$ 在竖向达到动态平衡。水平方向上, 作用力坐标系由机身惯性力、空气阻力 $\boldsymbol{K}$、广义摩阻力 $\boldsymbol{F}$ 和发动机反推力 $\boldsymbol{T}$ 组成, 飞机在其共同作用下完成减速滑行直至脱离跑道。由此建立如图 1 所示的飞机着陆滑行分析模型[11]。

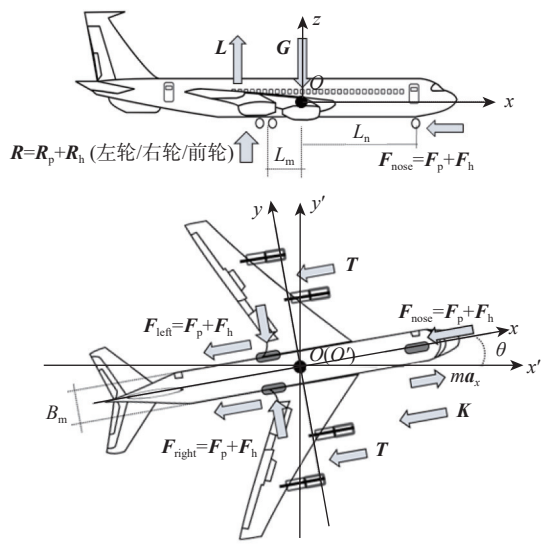


图 1　飞机着陆滑行分析模型

Fig. 1　Analytical model of aircraft landing and taxing

图 1 中, xOy 坐标系基于飞机机身建立, 属局部坐标系, 原点 O 与飞机质心重合, x 轴与飞机滑行方向保持一致, y 轴垂直于飞机滑行方向; $x'O'y'$坐标系基于跑道长度方向建立, 属整体坐标系, O'点为接地滑行起点飞机质心位置, x'轴与跑道中线重合, y'轴垂直于跑道中线。当飞机因机场道面积水发生偏转后, x 轴与 x'轴方向不再一致, 两者之间形成如图 1 所示的夹角(偏转角为θ)。飞机为当前坐标系内发生平面运动的三自由度体系, 另 2 个自由度为沿 x'轴方向平动分量 $\boldsymbol{X}$ 和沿 y'轴方向平动分量 $\boldsymbol{Y}$。

理想平衡滑行条件苛刻, 实际运行中, 易受外部扰动影响引发滑行姿态偏转, 当θ超出飞行控制极限时, 发生 VO 类事故几率增加。为突出分析重点, 简化分析条件, 作如下假定: ①滑行过程中, 飞机近似为刚体, 忽略机体变形影响; ②机场道面位于同一平面内, 忽略地球曲率及地形导致的纵坡度影响; ③由非均匀积水引发的不平衡摩阻是造成飞机偏转的主要原因, 无侧风扰动且前起落架轮胎无转角。飞机地面滑行基本运动学方程为

$$\begin{cases} \boldsymbol{R}_{\text{left}}+\boldsymbol{R}_{\text{right}}+\boldsymbol{R}_{\text{nose}}=\boldsymbol{G}-\boldsymbol{L} \\ \boldsymbol{K}+\boldsymbol{F}_{\text{left}}+\boldsymbol{F}_{\text{right}}+\boldsymbol{F}_{\text{nose}}=m\boldsymbol{a}_x-2\boldsymbol{T} \end{cases} \tag{1}$$

式中: 下标 left、right、nose 分别表示左、右两侧主起落架和前起落架; $\boldsymbol{a}_x$ 为沿 x 轴的加速度分量; 竖直方向上, $\boldsymbol{R}$ 由道面支撑力分量 $\boldsymbol{R}_{\text{p}}$ 和动水抬升力分量 $\boldsymbol{R}_{\text{h}}$ 这 2 个部分组成; $\boldsymbol{G}$ 按飞机着陆重量代表值 m 计算, 故$|\boldsymbol{G}|=mg$, g=9.8 N/kg。

由伯努利方程计算飞机机翼升力为

$$|\boldsymbol{L}|=C_{\text{lift}}\frac{1}{2}\rho_{\text{air}}|\boldsymbol{v}|^2S_{\text{wing}} \tag{2}$$

式中: C_{lift} 为无量纲升力系数; ρ_{air} 为空气密度, 非高原地区近似取 1.293 kg/m^3; S_{wing} 为机翼面积, 由查表得到; $\boldsymbol{L}$ 为滑行速度 $\boldsymbol{v}$ 的二次函数, 着陆时$|\boldsymbol{v}|$在 250～90 km/h 范围内连续改变[12]。

水平方向上, 理想平衡滑行时仅有 x 轴方向加速度分量 $\boldsymbol{a}_x$, 惯性力项为 $m\boldsymbol{a}_x$; 单个发动机反推力 $\boldsymbol{T}$ 与 $\boldsymbol{v}$ 相关, 可根据文献 [13] 确定反推力曲线; $\boldsymbol{F}$ 主要由湿摩擦力分量 $\boldsymbol{F}_{\text{p}}$ 和积水附加阻力分量 $\boldsymbol{F}_{\text{h}}$ 这 2 个部分组成。$\boldsymbol{F}_{\text{p}}$ 项通过轮胎附着系数φ与 $\boldsymbol{R}_{\text{p}}$ 相乘得到; $\boldsymbol{F}_{\text{h}}$ 项与轮胎滑水状态相关且在滑水临界时刻达到峰值[14], 其计算式为

$$|\boldsymbol{F}_{\text{h}}|=nC_{\text{resist}}\frac{1}{2}\rho_{\text{water}}|\boldsymbol{v}|^2S_{\text{tire}} \tag{3}$$

式中: n 为轮组系数; ρ_{water}为水体密度, ρ_{water}=1.0 kg/m^3; C_{resist} 为阻力系数, 通过有限元仿真得到; S_{tire} 为轮胎前缘积水接触面积, 可按文献 [15] 计算。

考虑到地面滑行时空气阻力系数仅为 0.09, 忽略 $\boldsymbol{K}$ 项及其他气动力影响, 推导 $\boldsymbol{a}_x$ 的表达式为

$$\boldsymbol{a}_x=\frac{\boldsymbol{F}_{\text{left}}+\boldsymbol{F}_{\text{right}}+\boldsymbol{F}_{\text{nose}}+2\boldsymbol{T}}{m} \tag{4}$$

矢量 $\boldsymbol{a}_x$ 与滑行运动方向相反, 通过积分可得任意时刻瞬时速度 $\boldsymbol{v}_x$ 为

$$|\boldsymbol{v}_x|=|\boldsymbol{v}_0|+\int|\boldsymbol{a}_x|\mathrm{d}t \tag{5}$$

式中: $\boldsymbol{v}_0$ 为接地初速度。

减速至 $\boldsymbol{v}_x$ 时, 总滑行距离 S_x 为

$$S_x=\int|\boldsymbol{v}_x|\mathrm{d}t \tag{6}$$

考虑到飞机着陆为连续变减速过程, 为便于计算, 采用多段匀减速运动替代, 当分段数量 k 足够多时, 计算结果趋于稳定。进而推导滑行速度与滑行距离计算式为

$$\begin{cases} |\boldsymbol{v}_x| = |\boldsymbol{v}_0| + \sum_{i=1}^{k} |\boldsymbol{a}_i| t \\ S_x = \sum_{i=1}^{k} \dfrac{|\boldsymbol{v}_i|^2 - |\boldsymbol{v}_{i-1}|^2}{2|\boldsymbol{a}_i|t} \end{cases} \tag{7}$$

式中：$|\boldsymbol{a}_i|$为分段内加速度分量。

若飞机因机场道面非均匀积水发生偏转，可由飞机设计手册查得两侧主起落架间距 B_{m}、主起落架到质心距离 L_{m} 及前起落架到质心距离 L_{n}。根据动量矩定理，各项水平力对质心取矩，两侧飞机轮组对称项相互抵消，无人为操纵干预条件下，$\boldsymbol{F}_{\mathrm{nose}}$ 分量始终通过飞机质心 O 点，当前分段内绕 z 轴转动扭矩 $M_z^{(i)}$ 为

$$M_z^{(i)} = \frac{1}{2}B_{\mathrm{m}}F_{\mathrm{h,left}}^{(i)} - \frac{1}{2}B_{\mathrm{m}}F_{\mathrm{h,right}}^{(i)} + L_{\mathrm{m}}\varphi_{\mathrm{left}}R_{\mathrm{p,left}}^{(i)} - L_{\mathrm{m}}\varphi_{\mathrm{right}}R_{\mathrm{p,right}}^{(i)} \tag{8}$$

转动角速度 $\omega_z^{(i)}$ 为

$$\omega_z^{(i)} = \int \frac{M_z^{(i)}}{I_z}\mathrm{d}t \tag{9}$$

式中：I_z 为机身转动惯性矩。

通过积分得到当前分段偏转角 $\theta^{(i)}$ 的计算式为

$$\theta^{(i)} = \frac{M_z^{(i)}}{2I_z}t^2 \tag{10}$$

飞机减速滑行全过程偏转角 θ 为

$$\theta = \sum_{i=1}^{k}\theta^{(i)} \tag{11}$$

显然，对于特定机型，其质量及刚度特征为定值，此时 θ 主要由 $\boldsymbol{R}_{\mathrm{p}}$、$\boldsymbol{F}_{\mathrm{h}}$ 作用力及非均匀积水区域滑行时长 t 共同决定。

2 积水道面摩阻作用仿真

机场道面积水分布受横坡度和轮辙共同影响，极为复杂，以往获取道面变形数据方式主要有基于真实道面的 3D 点云表面扫描或室内缩尺道面轮辙试验。前者扫描数据量可达 TB 级，大范围应用计算成本陡增；后者受缩尺比例及加载条件限制，路面变形与实际情况仍有较大差异。

美国机场道面试验中心 2012 年实施了编号 CC-5 的足尺柔性道面循环加载试验，目的是获取道面变形全过程数据。CC-5 试验采用可模拟轮载作用位置横向偏移的飞机轮组加载装置，与室内缩尺道面轮辙试验相比，所得道面变形特征与实际一致性更好。由于 CC-5 试验在室内稳定温湿度条件下进行，与真实道面 3D 点云扫描结果相比，受环境因素影响较小且数据量可控，为获取代表性的道面变形模型提供了研究支撑[16]。据此，开展非均匀积水评价与飞机轮组滑水仿真，获取积水道面摩阻特征，对比两侧飞机轮组滑水时程差异。

2.1 机场道面非均匀积水分布

以 CC-5 试验编号 LFC1-NW-SW 结果为例，机场道面横断面分布形式如图 2 所示。机场道面经过约 660 次循环加载作用后，最大变形深度达 9.8 mm，对照《民用机场道面评价管理技术规范》[17]，其为修复前最不利积水条件。此时，道面变形总影响宽度约 18.0 m，覆盖跑道中线两侧飞机轮组高频驶过区域，积水非均匀分布特征明显，当飞机于不同位置接地时，可存在较大的滑水状态差异。

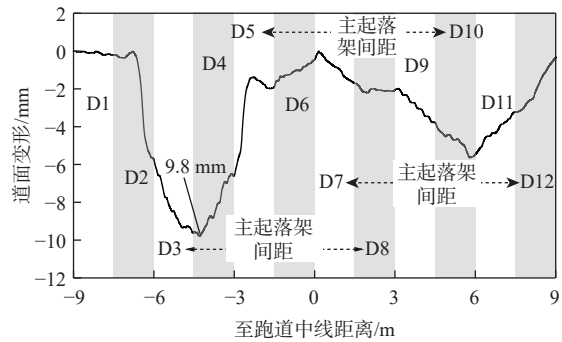


图 2 机场道面横断面变形曲线

Fig. 2 Deformation curve of pavement transverse profile

将该横断面划分为 12 个分段编号（见图 2），分段宽度约 1.5 m，与飞机轮组轮隙宽度及轮胎接地面宽度之和对应；对分段内最大变形点取积水深度代表值 W_i，忽略横坡度引起积水差异；共考察机场道面因降雨整体覆盖 3 mm（轻微积水）、7 mm（中等积水）或 13 mm（重度积水）3 种情况，整理得到各分段积水厚度结果，如表 1 所示。中等积水条件下有 5 个分段积水超出允许起降条件[6]，重度积水时，超限情况更加严重。

2.2 飞机轮组滑水仿真建模

采用 CEL 算法开展飞机轮组滑水流固耦合仿真，获取横向非均匀积水条件下 $\boldsymbol{R}_{\mathrm{p}}$ 和 $\boldsymbol{F}_{\mathrm{h}}$ 的变化规律。基于 ABAQUS 软件建立的仿真模型如图 3 所示，考察的滑水影响因素包括轮胎规格（尺寸、胎压、胎纹等）、轮组构型、积水厚度、滑行速度等。各部分离散网格及单元类型已标注，分析精度经过前期充分验证应用，限于篇幅不再展开[18-20]。

仿真开始后，水流从左侧入口端以速度 $\boldsymbol{v}_0$ 覆盖流域，模拟与滚动轮胎相对运动过程，并由右侧出口端流出，冲击速度按 $\boldsymbol{a}_x$ 连续改变；轮胎承受动水压力冲击，在两侧形成溅水喷流，提取轮胎前缘 $\boldsymbol{F}_{\mathrm{h}}$ 及轮胎底部 $\boldsymbol{R}_{\mathrm{p}}$ 时程结果，表征轮胎达到滑水临界状态直至脱离滑水全过程；机场道面积水层厚度依据飞机轮组接地位置所属分段确定，中心转动角速度 $\boldsymbol{\Phi}$ 与水流速度匹配，滑移率保持为 1.0。

表 1　机场道面分段积水分布结果

Table 1　Water film distribution results of independent pavement sections

道面分段	分段变形量/mm	轻微积水时W_j/mm	中等积水时W_j/mm	重度积水时W_j/mm
D1	0.2	3.2	7.2	13.2
D2	6.2	9.2	13.2	19.2
D3	9.6	12.6	16.6	22.6
D4	9.8	12.8	16.8	22.8
D5	6.3	9.3	13.3	19.3
D6	1.6	4.6	8.6	14.6
D7	1.7	4.7	8.7	14.7
D8	2.2	5.2	9.2	15.2
D9	3.8	6.8	10.8	16.8
D10	6.6	9.6	13.6	19.6
D11	5.4	8.4	12.4	18.4
D12	3.3	6.6	10.3	16.3

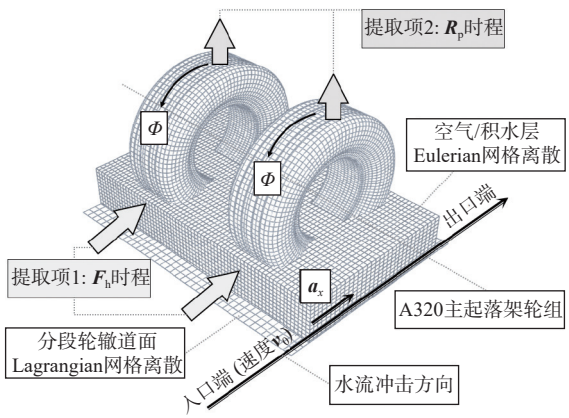


图 3　飞机轮组滑水模型

Fig. 3　Hydroplaning model of aircraft tire group

2.3　轮辙积水道面摩阻特征

图 4 对比了 $\boldsymbol{F}_h$ 和 $\boldsymbol{R}_p$ 分量时程变化曲线, 叠加道面变形后, D1 和 D4 分段分别对应表 1 中最小和最大积水厚度情况。如图 4 所示, 飞机减速滑行过程中, $\boldsymbol{R}_p$ 振荡增大, $\boldsymbol{F}_h$ 先增大后减小, 参考文献 [21] 可知, $\boldsymbol{R}_p$ 恢复初值时刻速度即为当前滑行条件下 $\boldsymbol{v}_p$ 结果, 与 $\boldsymbol{F}_h$ 曲线拐点位置对应, 两者共同验证了滑水结果的可靠性; D1 分段, $|\boldsymbol{v}_p|$大于 300 km/h, 远高出常规飞机着陆接地速度, 滑水几率极低; D4 分段, $|\boldsymbol{v}_p|$仅为 210 km/h, 落在图 4 中灰色接地速度区间内, 积水超限导致轮胎滑水风险增加; 临界滑水状态前后, D4 分段 $\boldsymbol{F}_h$ 变化幅度显著, 对飞机滑行姿态影响不容忽视。

考虑到两侧飞机轮组摩阻力差异是造成飞机偏转的主要原因, 定义差值分量$|\Delta\boldsymbol{F}_h|$和$|\Delta\boldsymbol{R}_p|$, 公式如下:

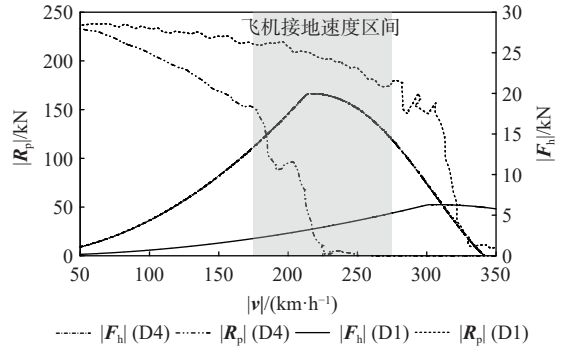


图 4　$|\boldsymbol{R}_p|$和$|\boldsymbol{F}_h|$时程曲线

Fig. 4　Time history curves of $|\boldsymbol{R}_p|$ and $|\boldsymbol{F}_h|$

$$\begin{cases} |\Delta\boldsymbol{F}_h| = F_{hi} - F_{hj} & i < j \\ |\Delta\boldsymbol{R}_p| = R_{pi} - R_{pj} & i < j \end{cases} \tag{12}$$

式中: 脚标 i 和 j 为分段编号, 与两侧飞机轮组间距和飞机接地位置横向偏移量有关。

对照 A320 主起落架间距 8.9 m, 将表 1 中分段两两编组, 共得出 7 组接地位置组合。以 D3 和 D8 分段组合为例, 图 5 依次给出 3 种积水条件下

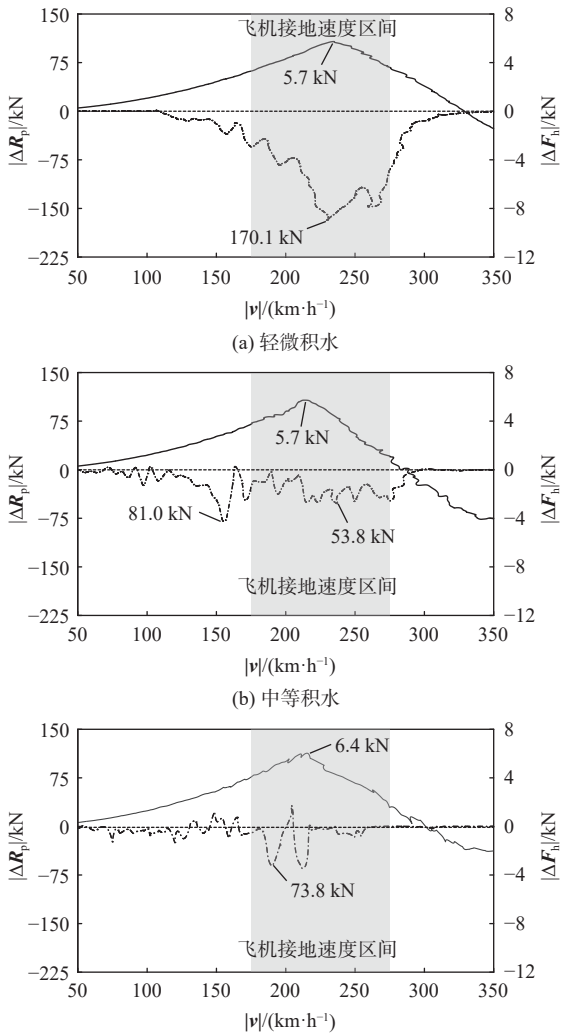


图 5　$|\Delta\boldsymbol{R}_p|$和$|\Delta\boldsymbol{F}_h|$时程曲线

Fig. 5　Time history curves of $|\Delta\boldsymbol{R}_p|$ and $|\Delta\boldsymbol{F}_h|$

$\Delta \boldsymbol{F}_{\mathrm{h}}$ 和 $\Delta \boldsymbol{R}_{\mathrm{p}}$ 分量时程曲线结果，此时为两侧飞机轮组积水厚度相差最大。

可以看出：①D3分段，由于积水较厚，$\Delta \boldsymbol{F}_{\mathrm{h}}$ 结果在300 km/h以下速度区间内均为正值；水膜层的隔绝作用使得D3分段道面支撑力更低，$\Delta \boldsymbol{R}_{\mathrm{p}}$ 结果基本保持为负，表明该侧 $\boldsymbol{F}_{\mathrm{p}}$ 低于D8分段。②$\Delta \boldsymbol{F}_{\mathrm{h}}$ 呈先递增后递减规律，峰值点位于飞机接地速度范围内，对高速着陆初始阶段影响突出；轻微积水时，$\Delta \boldsymbol{R}_{\mathrm{p}}$ 先振荡上升后逐步下降，峰值点出现时刻与 $\Delta \boldsymbol{F}_{\mathrm{h}}$ 对应；随着机场道面积水增厚，图5(b)、(c)中 $\Delta \boldsymbol{R}_{\mathrm{p}}$ 变化幅度减小。③参照文献[21]，取附着系数代表值φ=0.262，轻微积水时，$\Delta \boldsymbol{R}_{\mathrm{p}}$ 引起的最大偏转力为44.5 kN，约为 $\Delta \boldsymbol{F}_{\mathrm{h}}$ 最大值的7.8倍，两者部分相互抵消后，对飞机形成绕D8分段偏转运动趋势；重度积水时，$\Delta \boldsymbol{R}_{\mathrm{p}}$ 影响明显减弱，最大偏转力仅为 $\Delta \boldsymbol{F}_{\mathrm{h}}$ 峰值的1/5，$\Delta \boldsymbol{F}_{\mathrm{h}}$ 持续影响高速滑行过程，形成绕D3分段偏转运动趋势。

综合来看，非均匀积水道面条件下，两侧飞机轮组间摩阻力作用差异始终存在，飞机处于非对称偏转力的交错影响之下，VO类事故风险贯穿高速滑行全过程。

3 积水评价指标与整机运动学分析

飞机在横向非均匀积水区域滑行时长 t 较短时，轮组间摩阻力差异不足以构成威胁；但随着滑行距离的延长，上述偏转力持续累积直至引发VO类事故。显然，仅以积水厚度差值无法表征机场道面积水对整机滑行运动的影响，本节提出一种非均匀积水特征量化指标。以飞机滑水偏出跑道行为理论模型为基础，运用ADAMS平台开展非对称偏转力作用下的整机运动学分析。

3.1 非均匀积水临界长度评价指标

依据文献[22-23]，机轮接地滑行初始速度大，不可控刹车阶段安全风险高，当侧向偏转角不超过1°时，飞机滑行姿态较为稳定。当有侧风作用于着陆过程时，飞机多采用偏流法进近，且最大偏转角控制在5°以内。考虑到实际运行中的复杂情况，本文以θ=1°作为飞机偏转控制极限条件。推导非均匀积水临界长度评价指标 L_{cw}，公式如下：

$$L_{\mathrm{cw}}=\sqrt{\frac{|\boldsymbol{M}_z|}{2I_z}}\sum_{i=1}^{k}\frac{|\boldsymbol{v}_i|^2-|\boldsymbol{v}_{i-1}|^2}{2|\boldsymbol{a}_i|} \tag{13}$$

结合式(8)可以看出，L_{cw} 与飞机物理力学性能和道面摩阻力分布密切相关，可针对不同机场跑道起降条件开展精细化分析，其物理意义是量化非均匀积水道面偏转力作用累积情况，可作为机场道面积水控制依据。

3.2 基于ADAMS整机运动学模型

基于ADAMS平台建立图6所示的整机运动学分析模型。依据A320结构手册[24]，采用CATIA完成机身结构建模并导入装配；在ADAMS中定义机身对应的运动副结构、零部件对应的转动副、移动副、圆柱副、胡克副和球副连接关系，限于篇幅，此处不展开说明[25]。进而，将2.3节求解36组 $\boldsymbol{R}_{\mathrm{p}}$ 和 $\boldsymbol{F}_{\mathrm{h}}$ 样条数据导入模型，通过AKISPL和IF函数对两侧飞机轮组分别定义摩阻作用变化量；调用STEP阶跃函数，添加发动机推力曲线[26]，模拟着陆滑跑过程中双发反推作用。

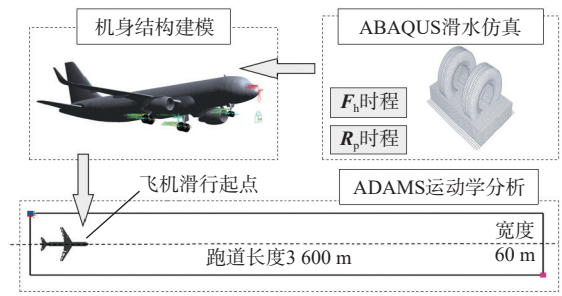


图6 A320建模与ADAMS运动学模型

Fig. 6 A320 modeling and ADAMS-based kinematic model

计算开始后，飞机以 $\boldsymbol{v}_0$ 沿跑道中线滑行，在机身惯性力、$\boldsymbol{R}_{\mathrm{p}}$、$\boldsymbol{F}_{\mathrm{h}}$ 和 $\boldsymbol{T}$ 共同组成力系作用下进行变减速运动，以飞机偏出跑道边缘或滑行速度降至50 km/h以下作为结束分析条件。假定机场道面积水完全均匀且飞机接地位置对中，飞机着陆滑行轨迹及速度变化曲线如图7所示。

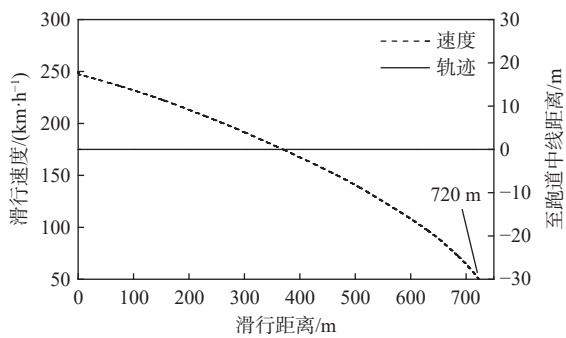


图7 均匀积水道面滑行轨迹及速度曲线

Fig. 7 Taxing track and speed curves under even distributed water film pavement condition

如图7所示，滑行速度由250 km/h降至50 km/h，总滑行距离约720 m，符合《波音和空客系列飞机飞行品质监控项目规范》[27]规定的着陆距离范围(300～750 m)，满足转入滑行道条件。由于两侧飞机轮组所受摩阻力对称均等，质心轨迹始终保持在跑道中线上。上述模型能较好地再现A320着陆滑行过程，可用于后续分析。

3.3　飞机滑水偏出跑道行为案例分析

为验证 L_{cw} 指标的适用性, 开展飞机滑水偏出跑道行为案例分析, 表 2 列举了 3 组滑行参数组合。图 8 对比了不同案例间着陆滑行轨迹及速度变化曲线差异。

表 2　飞机滑水偏出跑道行为案例参数组合

Table 2　Parameters combination of aircraft hydroplaning veer-off behavior cases

案例编号	积水条件	左轮组接地分段	右轮组接地分段	积水厚度差/mm	非均匀积水长度/m	L_{cw}理论值/m
案例1	中等积水	D3	D8	7.4	50	137
案例2	中等积水	D3	D8	7.4	750	137
案例3	中等积水	D1	D6	1.4	750	408

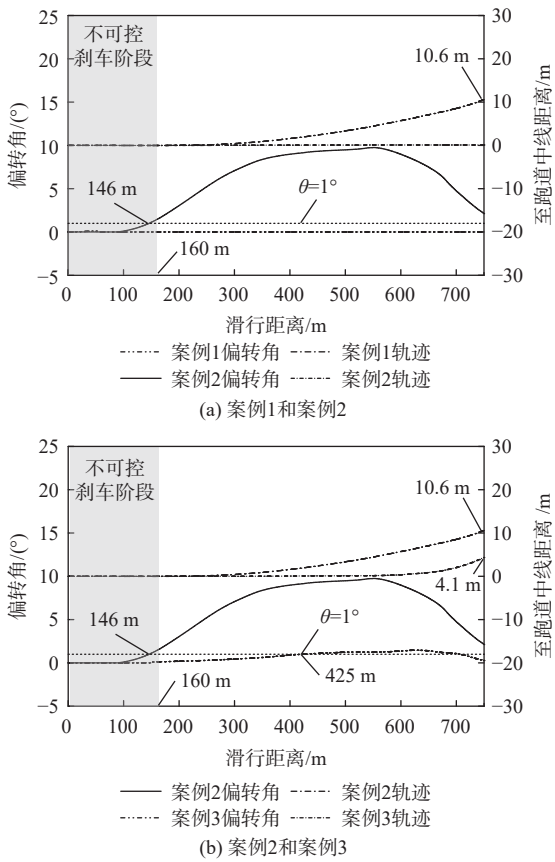


图 8　非均匀积水道面滑行轨迹及速度曲线

Fig. 8　Taxing track and speed curves under uneven distributed water film pavement condition

由图 8 可知: ①案例 1 中, 非均匀积水区域处于不可控刹车阶段, 由于其长度远低于 L_{cw} 理论值, 累积偏转作用有限, 飞机在惯性力主导下基本保持直线滑行的运动轨迹, 质心横向偏移量略有波动后迅速归零, 未见明显偏转角发生。②案例 2 中, 非均匀积水长度贯穿滑行全过程, 不可控刹车阶段 θ显著增加, 经过 146 m 滑行后, 即达到飞行控制极限条件, 滑行距离与 L_{cw} 理论值仅相差 6.6%; 假定当前无人为干预飞机经过 750 m 滑行后, 最大横向偏移量已达 10.6 m, 此时飞机已偏离常规接地中心位置区域, 发生 VO 类事故几率显著增大。③当前积水条件下, $\Delta \boldsymbol{R}_p$ 与 $\Delta \boldsymbol{F}_h$ 出现峰值点时对应滑行速度不一致(见图 5(b)), 两者交替主导飞机偏转运动趋势, 案例 2 中, 偏转角呈现先增大后减小的变化规律, 最大偏转角达 9.7°。④飞机于 D1～D6 分段接地时, 两侧积水厚度差较前 2 组案例显著减小, 案例 3 中, 不可控刹车阶段未见明显偏转, 完成 750 m 滑行后, 最大横向偏移量仅 4.1 m, 不足案例 2 的 1/2; 最大偏转角略高于 1°, 首次超出控制极限条件位置出现在滑行距离 425 m 处, 与 L_{cw} 理论值相差 4.2%。

综上, L_{cw} 指标可量化偏转作用累积程度, 对常规积水厚度指标形成有效补充; 两侧飞机轮组积水厚度差越大时, L_{cw} 理论值越低, 对机场道面积水控制要求越严格; 由 ADAMS 运动学分析得出, L_{cw} 结果与理论值一致性较好。

4　结　论

1) 机场道面积水受横坡度及轮辙共同影响, 非均匀分布特征显著, 可导致飞机偏出跑道事故, 单一积水厚度指标表征能力不足, 对此, 本文提出非均匀积水临界长度评价指标, 考察由横向非均匀积水引发的累积偏转作用影响。

2) 基于美国足尺加载试验开展飞机轮组滑水仿真, 结果表明, 机场道面非均匀积水会导致飞机于不同位置接地时, 两侧轮组所受摩阻作用存在明显时程差异; 飞机处于非对称偏转力交错影响之下, 偏出跑道风险贯穿着陆滑行全过程。

3) 基于飞机滑水偏出跑道行为理论推导, 进行 ADAMS 整机运动学分析, 结果表明, 该模型能较好地预测不同积水条件下飞机滑行轨迹及速度变化过程, 所得非均匀积水临界长度与理论值相差在 6.6% 以下, 验证了该指标的合理性与适用性。

多维度表征道面非均匀积水分布, 对预测飞机滑水偏出跑道行为及预防 VO 类事故是必要的; 考虑到不同机型起降性能的差异, 后续将针对 L_{cw} 指标的参数敏感性作深入讨论, 为起降安全保障及道面积水控制提供量化依据。

参考文献 (References)

[1] JIANG Y. Review on flight performance certification standard for wet and contaminated runway[J]. Procedia Engineering, 2011, 17: 7-12.

[2] YAN J, ZHANG H W, HUI B. Driving safety analysis using grid-based water-filled rut depth distribution[J]. Advances in Materials Science and Engineering, 2021, 2021: 5568949.

[3] 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).

[4] FWA T F, ONG G P. Wet-pavement hydroplaning risk and skid resistance: analysis[J]. Journal of Transportation Engineering, 2008, 134(5): 182-190.

[5] TANG T C, ANUPAM K, KASBERGEN C, et al. A finite element study of rain intensity on skid resistance for permeable asphalt concrete mixes[J]. Construction and Building Materials, 2019, 220: 464-475.

[6] 中国民用航空局飞行标准司. 航空承运人湿跑道和污染跑道运行管理规定: AC-121-FS-2009-33[S]. 北京: 中国民用航空局飞行标准司, 2009.
Flight Standards Division, Civil Aviation Administration of China. Regulations for the management of wet runway and contaminated runway operation by air carriers: AC-121-FS-2009-33[S]. Beijing: Flight Standards Division, Civil Aviation Administration of China, 2009(in Chinese).

[7] 燕姣, 李岩, 惠冰. 车辙断面非均匀积水对车辆安全性影响的仿真分析[J]. 汽车技术, 2019(11): 47-51.
YAN J, LI Y, HUI B. Simulation analysis of rutting section unevenness with respect to vehicle driving safety[J]. Automobile Technology, 2019(11): 47-51(in Chinese).

[8] 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.

[9] 季天剑, 高玉峰, 陈荣生. 轿车轮胎动力滑水分析[J]. 交通运输工程学报, 2010, 10(5): 57-60.
JI T J, GAO Y F, CHEN R S. Dynamic hydroplaning analysis of car tire[J]. Journal of Traffic and Transportation Engineering, 2010, 10(5): 57-60(in Chinese).

[10] 徐进, 彭其渊, 邵毅明. 直线路段积水路面车辆事故产生机理分析[J]. 中国公路学报, 2009, 22(1): 97-103.
XU J, PENG Q Y, SHAO Y M. Mechanism analysis of vehicle accident on surface gathered water in straight sections[J]. China Journal of Highway and Transport, 2009, 22(1): 97-103(in Chinese).

[11] 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).

[12] 宗辉杭. 基于轮辙道面飞机轮组滑水的积水评估指标研究[D]. 天津: 中国民航大学, 2022.
ZONG H H. Study on evaluation index of water accumulation in aircraft wheelset water skiing based on frog road surface[D]. Tianjin: Civil Aviation University of China, 2022(in Chinese).

[13] 黄文静, 吴密翠, 周晓飞. 使用反推力时飞机着陆滑跑距离的换算[J]. 飞行力学, 2003, 21(3): 56-58.
HUANG W J, WU M C, ZHOU X F. Conversion method of landing distance under reverse-thrust conditions[J]. Flight Dynamics, 2003, 21(3): 56-58(in Chinese).

[14] 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).

[15] 徐长群, 陶超. 大型客机积水跑道起降附加阻力评估[J]. 民用飞机设计与研究, 2018(1): 98-103.
XU C Q, TAO C. Additional drag assessment on standing water runways for large civil aircraft[J]. Civil Aircraft Design & Research, 2018(1): 98-103(in Chinese).

[16] FAA Airport Technology Research & Development Branch. National airport pavement test facility[EB/OL]. (2020-04-12)[2023-12-27]. http://www.airport-tech.tc.faa.gov/naptf.

[17] 中国民用航空局. 民用机场道面评价管理技术规范: MH/T 5024—2019[S]. 北京: 中国民航出版社, 2019: 39-67.
Civil Aviation Administration of China. Specification for pavement evaluation and management of civil airports: MH/T 5024—2019[S]. Beijing: China Civil Aviation Publishing House, 2019: 39-67(in Chinese) .

[18] 赵夫朋. 基于足尺轮辙试验的柔性道面力学性能评价指标研究[D]. 天津: 中国民航大学, 2020.
ZHAO F P. Study on evaluation index of mechanical properties of flexible pavement based on full-scale frog test[D]. Tianjin: Civil Aviation University of China, 2020(in Chinese).

[19] 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[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).

[20] 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).

[21] 宗辉杭, 李岳, 蔡靖, 等. 基于飞机滑水安全沥青道面积水评价指标[J]. 北京航空航天大学学报, 2024, 50(3): 765-773.
ZONG H H, LI Y, CAI J, et al. Evaluation index of asphalt road area water based on aircraft water skiing safety[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 765-773(in Chinese).

[22] 李少杰. 考虑不平衡摩阻的湿滑污染道面飞机着陆安全问题研究[D]. 天津: 中国民航大学, 2020.
LI S J. Study on landing safety of aircraft with wet and polluted road surface considering unbalanced friction[D]. Tianjin: Civil Aviation University of China, 2020(in Chinese).

[23] 中国民用航空局飞行标准司. 民用航空飞行签派员执照和训练机构管理规则[S]. 北京: 中国民用航空局, 2009.
Department of Flight Standards, Civil Aviation Administration of China. Management rules on licenses of civil aviation flight dispatchers and training organizations[S]. Beijing: Civil Aviation Administration of China, 2009(in Chinese).

[24] 王裕, 刘卫, 张嘉振. 面向增材制造技术的民用飞机典型铰链结构拓扑优化设计[C]//第八届中国航空学会青年科技论坛. 北京:

中航出版传媒有限责任公司, 2018: 549-554.
WANG Y, LIU W, ZHANG J Z. Topology optimization design of typical hinge structure of civil aircraft for additive manufacturing technology[C]//Proceedings of the 8th Youth Science and Technology Forum of the Chinese Society of Aeronautics and Astronautics. Beijing: AVIC Publishing and Media Co. , Ltd. , 2018: 549-554(in Chinese).
[25] 侯志礼, 王钢林. 某无人机着陆滑跑动态仿真分析[J]. 飞机设计, 2011, 31(5): 9-11.
HOU Z L, WANG G L. Dynamic simulation of landing and rolling for a certain UAV[J]. Aircraft Design, 2011, 31(5): 9-11(in Chinese).
[26] 黄敬杰, 张鑫, 孟超, 等. 反推力装置对民航飞机减速性能影响[J]. 航空动力学报, 2023, 38(3): 728-734.
HUANG J J, ZHANG X, MENG C, et al. Effect of thrust reverser on the deceleration performance of civil aircraft[J]. Journal of Aerospace Power, 2023, 38(3): 728-734(in Chinese).
[27] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 波音和空客系列飞机飞行品质监控项目规范: GB/T 26763—2011[S]. 北京: 中国标准出版社, 2012.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Flight operational quality assurance events and deviation limits for Boeing and Airbus aircraft: GB/T 26763—2011[S]. Beijing: Standards Press of China, 2012(in Chinese) .

Influence on veer-off behavior of aircraft hydroplaning due to transverse uneven water accumulation

LI Yue[1], ZHANG Yaohua[1], CAI Jing[1, 2, *], REN Yinan[1]

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

Abstract: Accumulated water film on the pavement is unevenly distributed due to several factors, which may lead to an aircraft veer-off accident. The single index of water film thickness lacks the ability to characterize the water film distribution condition. In this paper, a theoretical model of the deviation behavior of aircraft hydroplaning was established. A hydroplaning simulation of an aircraft wheel group was carried out on the basis of the American full-scale test result of pavement loading. The transverse water film uneven distribution was quantified section by section. The differences in the frictional effect between both sides of aircraft wheel group were compared. It was suggested to use a new type of evaluation index called the critical length of unevenly distributed water film. The rationality and applicability of such an index were verified through a correlation study between critical length and aircraft taxing performance variation by using ADAMS-based kinematic analysis. The result indicated that notable differences exist between the time-history curve of the frictional effect between both sides of the wheel group in the case that the aircraft lands on the pavement with a transversely unevenly distributed water film. The entire landing procedure raised the likelihood of a veer-off mishap because asymmetric deviation forces constantly impact the aircraft's taxing performance. The critical length of uneven distributed water film derived from three study cases were merely 6.6% differed from theoretical results. The proposed evaluation index may take the accumulative impact of deviation forces due to transversely unevenly distributed water film into consideration, which may provide useful reference for take-off and landing safety protection and water distribution control on the pavement.

Keywords: aircraft wheel group; tire hydroplaning; uneven water accumulation; veer-off; ADAMS

Received: 2024-01-02; **Accepted**: 2024-02-23; **Published Online**: 2024-03-09 13:24
URL: link.cnki.net/urlid/11.2625.V.20240307.1151.001
Foundation items: General Program of the National Natural Science Foundation of China (52472369); Open Topic for the Research Center of Airport Intelligent Construction and Industrialization Engineering Technology of Civil Aviation (MHJGKFKT-2025-01)
* **Corresponding author**. E-mail: caijing75@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0852

一种无锚旋转框遥感图像船只目标检测算法

李露，陈科研，刘辰阳，史振威*
(北京航空航天大学 宇航学院，北京 100191)

摘 要：精准高效的船只目标检测在确保海洋利益和构建海洋强国中发挥着重要作用，其现实价值显著。然而，现有的基于可见光遥感图像的船只检测算法大多依赖于锚框，计算量大、超参数多且算法的泛化能力有限。尽管自然图像目标检测算法通过采用无锚的方法改进了这些问题，但大部分只能实现水平框目标检测，无法应对船只目标形状狭长、角度分布不同和排列紧密等特点。针对这些问题，设计一种无锚旋转框遥感图像船只目标检测算法。在CenterNet的基础上，针对遥感图像船只检测任务，提出一种基于分布先验的置信度系数预测分支来产生更高质量的正样本；一种基于双曲激活的角度预测分支来限制角度的输出空间，得到更准确度旋转角度表示；同时，该算法采用柔性正负样本标签分配策略，以提供动态的精细化的监督信息，加速网络的收敛。在HRSC2016数据集上的实验验证了所设计算法相比与其他先进对比算法的优越性，并证实了所设计的各模块的有效性。

关 键 词：遥感图像；光学遥感图像；船只目标检测；无锚检测器；旋转框目标检测
中图分类号：V443+.5; TP751
文献标志码：A **文章编号**：1001-5965(2026)03-0706-07

近年来，随着科学技术的进步，遥感对地观测技术得到了迅速发展。其中，利用遥感技术对海面状况进行有效监控，是保障海上利益、构建海洋强国的重要基础。船只目标的精准检测是海面状况有效监控的关键保障，具有极高的现实价值。在国民经济领域，精准的船只目标检测可以协助管理海运交通、打击非法捕捞；在海洋安全领域，可以辅助海上救援、打击贩毒与非法移民；在军事部署领域，可以帮助掌握敌舰的动态和作战意图、对敌舰实施精准打击[1]。

光学遥感图像船只目标检测是指利用特定算法从可见光遥感图像中搜索并标记出船只目标所在位置[2]。常见的标记方式包括水平框标记、旋转框标记及像素级别的实例分割标记。由于船只目标常常呈现多角度、密集排列的状态，采用水平框标记会导致目标框中出现大量背景信息，且目标框之间存在大量重叠区域，无法准确反映目标状态。虽然像素级别的实例分割标记可以解决上述问题，但由于分割任务需要大量像素级标注，标注任务量大，限制了其在实际任务中的应用。相比之下，旋转框标记可以更准确地描述目标的状态，反映目标的长宽比、角度等信息，且在标注难度上远小于像素级标注。因此，采用旋转框对船只目标进行标记更加有效、合理。

基于深度学习的遥感图像目标检测算法主要分为两阶段算法和单阶段算法，其中，单阶段算法因其结构简单、运算效率高等特点逐渐成为主流目标检测算法，具体包括基于锚框的算法及基于无锚框的算法。尽管锚框已经成为现有主流检测算法的标配，但其存在的问题仍然明显：①计算量大；②超参数多；③算法泛化能力差。此外，在解决遥感图像目标旋转框检测问题上，基于锚框的算法需要在特征图上设置多个角度的参考框，这进一步增加了计算量[3]。

无锚检测器主要分为两类：一类以CornerNet[4]

收稿日期：2024-01-03；**录用日期**：2024-04-23；**网络出版时间**：2024-05-10 13:46
网络出版地址：link.cnki.net/urlid/11.2625.V.20240510.1036.001
基金项目：国家自然科学基金(62271018)；北京航空航天大学教改项目：航天信息工程专业综合实验教学改革(2023-15-02-08)
* **通信作者**. E-mail：shizhenwei@buaa.edu.cn

引用格式：李露，陈科研，刘辰阳，等. 一种无锚旋转框遥感图像船只目标检测算法[J]. 北京航空航天大学学报，2026，52(3)：706-712.
LI L，CHEN K Y，LIU C Y，et al. An anchor-free rotated box remote sensing image ship object detection algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：706-712 (in Chinese).

为代表, 通过直接回归关键点并进行匹配来完成目标检测; 另一类以 CenterNet[5] 为代表, 在定位目标关键点后, 直接预测边界框的相关参数。尽管面向自然图像的无锚检测器有效缓解了锚框设计引入的诸多问题, 但受限于其无法生成旋转边界框, 在面对形状狭长、方向任意且排列密集的船只目标时, 检测效果并不理想。

因此, 本文设计一种无锚旋转框遥感图像船只目标检测算法, 旨在解决现有主流基于锚框的船只检测算法存在的问题, 实现对船只目标的旋转框检测。具体来讲, 本文在 CenterNet 的基础上, 针对遥感图像船只检测任务, 设计了一个基于分布先验的置信度系数预测分支和基于双曲激活的角度预测分支, 进一步提高船只目标的检测精度。同时, 提出了一种柔性正负样本标签分配策略, 提供更精细的监督信息。

1　本文算法

本文从光学遥感图像船只检测任务的难点出发, 提出一种无锚旋转框遥感图像船只目标检测算法。该算法摒弃了设置锚框的主流算法, 改善了其存在的计算量大、超参数设置复杂、算法设计泛化能力差等问题, 从而提高了船只目标的检测精度和效率。

在算法设计上, 本文提出带权重的损失函数及柔性划分正负样本的策略, 并设计了一种基于分布先验的置信度系数预测分支和基于双曲激活的角度预测分支。基于分布先验的置信度系数预测分支采用更符合船只目标特性的置信度系数先验分布图, 可以有效剔除低质量的检测框。基于双曲激活的角度预测分支通过双曲非线性激活函数限制了角度预测值的输出范围, 优化了模型的学习空间, 从而获得了更准确的角度预测值。为验证本文算法的优势, 在船只检测数据集—HRSC2016[6] 数据集上与现有算法进行了对比。

1.1　整体网络结构

本文算法主要包括主干网络和无锚检测头, 其网络结构如图 1 所示。

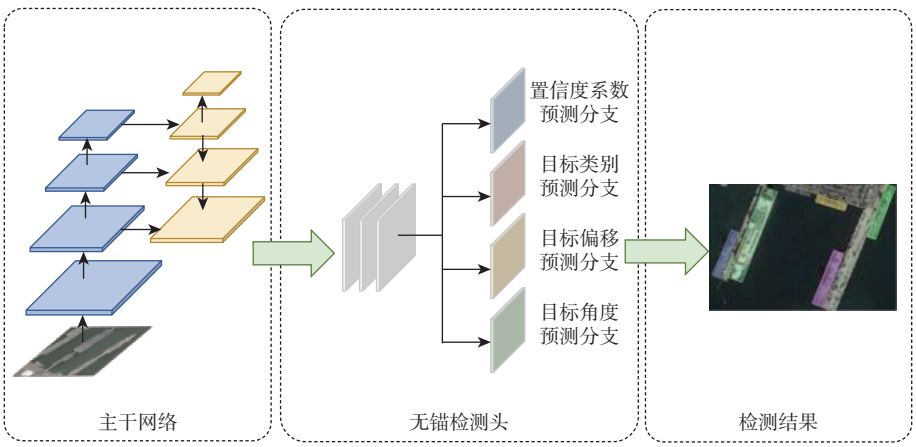


图 1　本文算法网络结构

Fig. 1　Network structure of the proposed algorithm

主干网络通过逐层卷积提取遥感图像的高层语义特征, 同时通过构建特征金字塔得到不同语义层次的特征图。特征图的层数越深, 对应原图中的感受野越大, 反之越小。多层次的特征图有助于检测出不同尺度的目标。无锚检测头包括目标置信度系数预测分支、目标类别预测分支、目标相对中心点的偏移预测分支和目标角度预测分支。通过目标类别预测分支完成对该中心点的分类, 通过目标相对中心点的偏移预测分支获取目标旋转框距离中心点的距离, 通过目标角度预测分支获取该旋转检测框的角度, 通过目标置信度系数预测分支获取该中心点属于前景目标的概率。

1.2　主干网络

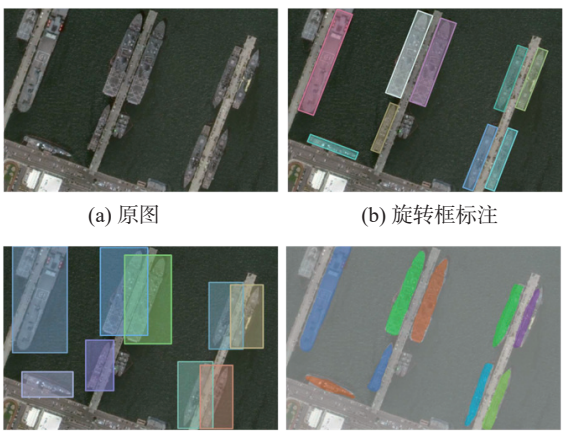
主干网络由特征提取器和多尺度特征金字塔两部分构成。在特征提取器中, 常见的网络包括 ResNet[7]、ResNeXt[8]、VGG(visual geometry group)[9] 等。尽管结构复杂的特征提取网络可能带来更优秀的检测结果, 但是会增加网络运算的复杂性及训练和推断的时间。因此, 本文选择使用 ResNet50 进行高层语义特征的提取。

在遥感目标检测中, 目标尺度的变化范围较大, 因此, 如何进行多尺度的目标检测是算法设计中需要考虑的问题。特征金字塔网络通过将卷积网络输出的浅层特征与高层特征组合, 生成不同层级的特征图[10]。利用这些特征图, 可以构建特征金字塔, 使得语义特征图同时具备更强的语义抽象能力和更准确的定位能力。通过 stride 为 2 的卷积块可以得到不同大小尺寸的特征图, 本文根据船只目

标尺度的分布，从特征提取器的中间层抽取不同尺度的特征进行融合，最终得到4层特征金字塔（分别由原图下采样1/4、1/8、1/16、1/32），如图1所示。特征金字塔的构建有利于检测出不同尺寸大小的目标，从而提高遥感图像船只目标检测的准确性和效率[11]。

1.3 无锚检测头

如图2所示，相比于水平框和像素级分割，旋转框更能准确描述船只目标的形状与方向。现有的主流算法大多采用基于锚框的检测头，尽管在通用视觉目标检测上能够获得相当的性能，但是存在超参数过多、计算量大、泛化性差等问题。本文针对高分辨率遥感图像船只目标检测这一问题，提出无锚检测器，这一设计可以避免超参数设置的繁琐流程，以及海量预设锚框给计算带来的压力。同时，高分辨率遥感图像船只目标不存在目标重叠和小目标紧密排列的情况，可以从预设多个锚框的范式中跳出而不影响性能。本文设计的旋转框无锚检测头包含4个分支：目标置信度系数预测分支、目标类别预测分支、目标相对中心点的偏移预测分支和基于双曲激活的目标角度预测分支。



(a) 原图 (b) 旋转框标注

(c) 水平框标注 (d) 实例级分割标注

图2 船只目标不同标注方式的对比示意图

Fig. 2 Different annotation methods for ship objects

将船只目标的标注框集合定义为B，其中，$B=(x_c,y_c,w,h,\theta)$。(x_c,y_c)为目标标注框中心点的坐标，w和h分别为标注框的宽度和高度，而θ则为标注框长边与横轴的夹角，取值范围为$\theta\in[-\pi/2,\pi/2)$。

在检测头中，通过共享部分参数可以在降低网络复杂度的同时，保持检测的精度。设计的4个分支共享部分参数，这些共享参数由3个卷积层、批归一化层[12]及非线性激活函数层组成，其中，卷积层均采用3×3的卷积核，共享的卷积层如图1所示。

本文设计的旋转框无锚检测头包含的4个分支，具体如下。

目标类别预测分支。目标类别预测分支的主要任务是预测特征图上像素属于某类目标的概率。由于船只检测属于单类目标检测任务，因此，该分支仅输出一张检测概率图，表示该点是前景或背景。真值标签由正负样本划分确定，若该特征图像素被划分为正样本，则标签为1，否则为0。目标类别预测分支由共享参数的卷积块后接卷积核大小为3×3的卷积层和Sigmoid层构成。Sigmoid层的作用是将卷积输出映射到(0,1)的范围内，这个范围与样本属于目标的概率范围相符。

目标相对中心点的偏移预测分支。偏移预测分支的任务是预测目标检测框4条边距离特征图像素在原图上的坐标，即偏移量。该分支最终输出4个通道的偏移量图，分别对应距左、上、右、下这4边的距离(l,t,r,b)。若点(x_i,y_i)属于$B=(x_c,y_c,w,h,\theta)$，则其监督信息由以下步骤确定。

首先，在图像坐标系中得到(x_i,y_i)相对于标注框中心点(x_c,y_c)的坐标，即

$$x_i^{\mathrm{h}}=x_i-x_c \tag{1}$$

$$y_i^{\mathrm{h}}=y_i-y_c \tag{2}$$

然后，在标注框坐标系中得到(x_i,y_i)相对于标注框中心点(x_c,y_c)的坐标，即

$$\begin{bmatrix}x_i^{\mathrm{r}}\\ y_i^{\mathrm{r}}\end{bmatrix}=\begin{bmatrix}\cos(-\theta) & -\sin(-\theta)\\ \sin(-\theta) & \cos(-\theta)\end{bmatrix}\begin{bmatrix}x_i^{\mathrm{h}}\\ y_i^{\mathrm{h}}\end{bmatrix} \tag{3}$$

最终可以得到(x_i,y_i)距离4条边的距离为

$$\begin{cases}l_i=\dfrac{h}{2}-y_i^{\mathrm{r}}\\ t_i=\dfrac{w}{2}-x_i^{\mathrm{r}}\\ r_i=\dfrac{h}{2}+y_i^{\mathrm{r}}\\ b_i=\dfrac{h}{2}+x_i^{\mathrm{r}}\end{cases} \tag{4}$$

由于不同尺度的特征图共享检测分支参数，为使得网络更好地学习到不同尺度的目标，偏移预测分支在3×3卷积层后利用特征图对应原图的缩放比对结果做缩放处理，使得不同尺度的目标距离预测分支输出值相似。这一做法降低了模型学习难度，提高了模型检测结果。此外，由于距离始终为正值，采用ReLU层保证网络输出结果始终为正。

基于双曲激活的目标角度预测分支。目标角度预测分支的任务是预测特征图上像素点所属的检测框主轴与图像横轴的夹角θ，$\theta\in[-\pi/2,\pi/2)$。夹角的真值标签即为标注框主轴与图像横轴夹角。因为θ范围的限制，直接使用3×3卷积层并不能获得良好的检测效果。因此，本文提出一种新的角度预测分支设计方式，采用双曲非线性激活函数限制角度预测分支预测值的范围，将卷积输出映射

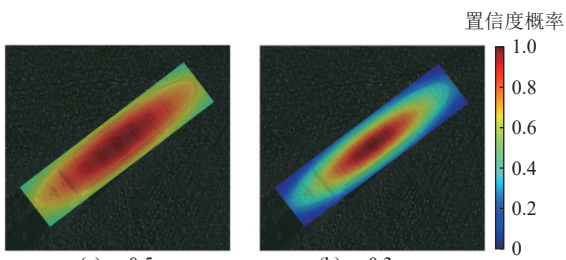
至$[-\pi/2,\pi/2)$, 其激活函数表达式为

$$\sigma(\theta)=\frac{\pi}{2}\tanh\theta \tag{5}$$

目标置信度系数预测分支。目标置信度系数预测分支是输出在后处理过程中与类别预测分支输出置信度相乘的系数。通过赋予低质量样本生成的检测框以小的系数, 降低低质量检测框的置信度, 可有效完成对低质量检测框的筛除。本文设计的置信度监督信息赋予目标中心区域更高的置信度系数, 更符合实际船只的特点, 如:

$$p=\mathrm{e}^{-\frac{1}{2}\left[\left(\frac{x_i-x_c}{w\alpha}\right)^2+\left(\frac{y_i-y_c}{h\alpha}\right)^2\right]} \tag{6}$$

式中: w 和 h 分别为目标包围框的宽和高; α为控制分布的超参数, 图 3 为α分别取 0.5 和 0.3 时可视化结果。由图可知, α可以调整目标检测正样本的数量, α越大, 可用的高响应监督正样本数量越多, 同时前向预测时召回率也会更高, 但是相应的虚警率也会更高。置信度分支的设计包括一个3×3卷积层和 Sigmoid 层。Sigmoid 层将卷积层输出映射到(0,1)。



(a) α=0.5　　(b) α=0.3

图 3　不同控制系数对置信度系数监督图的影响

Fig. 3　Impact of different control coefficients on supervision map

1.4　损失函数

本文提出的网络总体损失函数L由 4 个部分组成, 分别为类别预测损失L_{cls}、偏移预测损失L_{shif}、角度预测损失L_{ang}及置信度系数预测损失L_{conf}。其中, 偏移预测损失、角度预测损失和置信度系数预测损失仅针对被划分为正样本的点进行计算, 而类别预测损失则对正负样本均进行计算。损失函数的具体表达式为

$$L=L_{\mathrm{cls}}+I(L_{\mathrm{shif}}+L_{\mathrm{ang}}+L_{\mathrm{conf}}) \tag{7}$$

式中: I 为指示函数, 当该训练实例为正标注时为 1, 为负标注时为 0。

1.5　标签分配

高质量的训练样本能够提供更精确的监督信息, 从而加速网络的收敛并提高其精度。对于一阶段的目标检测算法而言, 正负样本的标签分配方式对检测结果具有至关重要的影响。传统的基于锚框的检测算法通常利用预设锚框与标注框的交并比(intersection over union, IoU)来划分正负样本, 一般采用硬阈值进行样本的正负判断, 或者采用双阈值来划分出忽略样本。在这类算法中, 选取的划分阈值对结果具有决定性的影响, 需要仔细调节阈值参数以获得良好的检测结果。

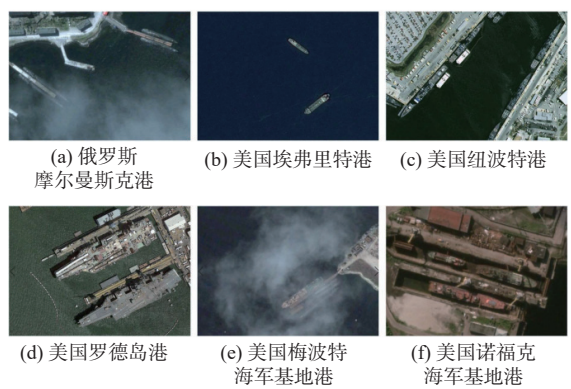
然而, 基于目标中心点进行目标检测的算法及其他基于关键点的目标检测算法仅将目标的几个点作为正样本, 其他点均划分为负样本。这种算法忽视了大量的可用的目标监督信息, 降低了召回率。同时, 由于将大量点划分为负样本, 加剧了正负样本比例不均衡的问题。但若将所有属于目标的点都划分为正样本, 则目标框内不可避免包含的背景信息会生成许多低质量的正样本, 从而降低检测精度。

因此, 本文采用了一种柔性的正负样本划分方式。在训练阶段, 将原本标注框内一定比例范围内的点作为正样本, 同时利用网络的置信度系数分支学习该点属于目标的概率。在推断阶段, 将最终得到的概率与该点生成检测框置信度的乘积作为该候选检测框的最终得分。这种柔性的划分方法充分利用了目标的信息, 缓解了正负样本比例不均衡的问题; 通过置信度系数分支来降低低质量样本生成检测框的得分, 使得这些检测框在后处理中被筛除。

2　实验结果及分析

2.1　数据准备

本文选取了公开的 HRSC2016 数据集作为实验的数据源, 数据集样本如图 4 所示。所有图像均取自 Google Earth, 图像分辨率为 0.4～2 m, 主要覆盖了全球 6 个著名的港口海域, 图像的空间分辨率在 0.4～2 m 之间, 数据包含海上航行和近岸停泊 2 种主要场景, 能够模拟真实的复杂海况和港口环境。此外提供了清晰的舰船细节特征, 使其能够支持细



(a) 俄罗斯摩尔曼斯克港　(b) 美国埃弗里特港　(c) 美国纽波特港

(d) 美国罗德岛港　(e) 美国梅波特海军基地港　(f) 美国诺福克海军基地港

图 4　HRSC2016 数据集样例

Fig. 4　Samples from HRSC2016 dataset

粒度的目标识别。具体的数据集参数及训练测试集的划分情况如表1所示。

表1 数据集信息

Table 1 Dataset statistics

来源	分辨率/m	图像长、宽/像素	训练集样本数	验证集样本数	测试集样本数
Google Earth	0.4～2	300～1 500	1 207	541	1 228

2.2 评价指标和实验细节

为全面评估检测结果，采用MSCOCO[13]数据集上的多种评价指标对本文算法进行评估，包括检测精度(average precision, AP)、AP50、AP75、APs、APl。这些指标分别表示检测精度、在IoU阈值为0.50时的检测精度、在IoU阈值为0.75时的检测精度、在小尺度目标上的平均检测精度及在大尺度目标上的检测精度。

本文中批处理大小设定为8，采用AdamW[14]优化器进行训练，设定的学习率为0.000 1。在训练过程中，采用余弦退火策略对学习率进行衰减，总迭代次数为10 000次。训练图像的输入大小为1 024×1 024。在训练阶段，对图像进行了一系列的几何变换，包括旋转任意角度、随机水平翻转及随机垂直翻转等。此外，为增强网络对光照等因素的鲁棒性，还采用了伽马变换、饱和度变换、对比度变换等色彩变换方式。

2.3 与其他算法对比的结果

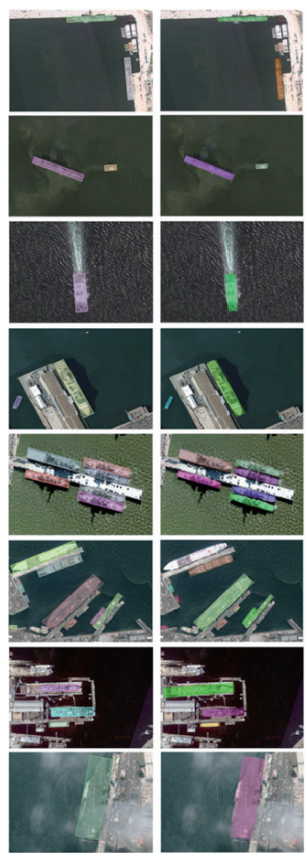
为进一步验证本文算法的有效性，将本文算法与现有的旋转框目标检测算法在HRSC2016测试集上进行了比较。由于其他算法在部分评价指标上的数据缺失，选择在平均精度(mean average precision, mAP)指标下进行比较。如表2所示，其中，尽管本文算法采用了相对简单的特征提取主干网络，但其检测结果仍然优于采用更复杂特征提取网络ResNet101的检测算法。这得益于本文在无锚框架下设计目标检测算法，同时提出了基于分布先验的置信度预测分支、基于双曲激活的角度预测分支和柔性正负样本分配策略，这些针对高分辨率遥感图像船只目标的特殊设计能够使得网络更快收敛，并提高了针对船只的定位精度。与同期最优秀的检测算法基于环形平滑标签的特征金字塔网络(circular smooth label based feature pyramid networks, FPN-CSL)相比，本文算法在平均精度mAP上提高了5.9%，实现了最佳的检测结果(表中加粗部分为最优结果)。

表2 与其他算法的对比

Table 2 Comparison results with other algorithms

算法	主干网络	是否数据增强	mAP/%
R2CNN[15]	ResNet101	否	73.1
RRPN[16]	ResNet101	否	79.1
TOSO[17]	ResNet101	是	79.3
RetinaNet-H[18]	ResNet101	是	82.9
SARD[19]	ResNet101	否	85.4
RoI-Trans[20]	ResNet101	否	86.2
RSDet[21]	ResNet152	否	86.5
Gliding Vertex[22]	ResNet101	否	88.2
R3Det[23]	ResNet152	是	89.3
FPN-CSL[24]	ResNet152	是	89.6
本文算法	ResNet50	是	**95.5**

图5为本文算法的船只目标检测结果的可视化示例。从结果中可以看出，本文算法在面对尺寸差异大、目标角度朝向各异、目标排列紧密等检测难点时，表现出了较好的性能。此外，在存在复杂港口、船只尾部浪花等背景信息的干扰下，以及存在云雾遮挡或图像中存在大量冗余信息的情况下，本文算法也表现出了较好的鲁棒性。



(a) 原图 (b) 本文算法

图5 本文算法检测结果的可视化示例

Fig. 5 Visual examples of detection results of the proposed algorithm

2.4 消融实验

为深入研究本文算法设计的各个模块及改进对最终检测结果的影响，本节设计了相应的消融实验。具体实验包括基于分布先验的置信度预测分支、基于双曲激活的角度预测激活函数、带权重的

回归损失函数及柔性正负标签分配策略对整体检测性能的影响。实验结果如表 3 所示, 针对遥感图像船只目标检测的设计和改进都取得了相应的效果提升(加粗部分为最优结果)。

表 3　消融实验

Table 3　Ablation experiments

算法	AP/%	AP50/%	AP75/%	APs/%	APl/%
本文算法	**70.5**	**95.5**	**87.3**	**59.2**	**73.1**
无分布先验的置信度分支	69.5	95.3	84.0	56.9	71.4
无角度预测激活函数	69.0	94.5	83.4	57.4	70.5
无带权重的回归损失	68.3	94.4	82.3	56.7	70.1
无柔性正负标签分配	67.4	92.5	81.8	55.7	69.6

通过表 3 的实验结果可以得出以下结论: ①置信度系数分支预测的值实质上是标注框中的点属于目标的概率, 其值与该点距标注框中心距离成反比。本文提出了更符合船只特性的置信度系数分布图作为置信度系数分支的真值, 从而获得了更好的检测结果, 在大尺寸船只检测方面效果更为突出; ②本文提出了一种基于双曲正切的角度预测激活函数来控制预测的角度范围, 该方法在所有评价指标下都有增益, 在较为严格的 AP75 指标下表现更为突出; ③本文采用置信度分支的真值标签作为距离预测及角度预测损失函数的系数, 以降低低质量样本对损失函数的贡献, 从而减轻模型对低质量样本的关注度。该设计在对检测框质量要求更高的条件下获得了 5% 的平均精度提升; ④采用柔性正负标签分配策略相较于完全不使用柔性正负标签分配, 在所有评价指标下都取得了较大的增益, 验证了监督信息的分配对网络性能的影响。

3　结　论

本文提出一种无锚旋转框遥感图像船只目标检测算法, 实现了船只目标的旋转框检测。具体结论如下:

1) 本文算法避免了现有基于锚框算法存在的计算量大、超参数多、算法泛化能力差等问题。通过无锚思想, 本文算法一定程度解决了遥感船只检测中存在的目标尺度差异大、目标朝向分布广、目标排列紧密的问题。

2) 本文通过设计基于分布先验的置信度预测分支、基于双曲激活的角度预测分支、带权重的回归损失函数、柔性正负标签分配策略等模块, 提升了自然图像无锚检测器在遥感光学图像船只目标检测任务上的性能, 并通过消融实验验证了各模块的有效性。

参考文献 (References)

[1] 刘俊琦, 李智, 张学阳. 可见光遥感图像海面目标检测技术综述[J]. 计算机科学, 2020, 47(3): 116-123.
LIU J Q, LI Z, ZHANG X Y. Review of maritime target detection in visible bands of optical remote sensing images[J]. Computer Science, 2020, 47(3): 116-123(in Chinese).

[2] CHEN J Q, CHEN K Y, CHEN H, et al. A degraded reconstruction enhancement-based method for tiny ship detection in remote sensing images with a new large-scale dataset[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5625014.

[3] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.

[4] LAW H, DENG J. CornerNet: detecting objects as paired keypoints[C]//Proceedings of the Computer Vision–ECCV 2018. Berlin: Springe, 2018: 765-781.

[5] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-25)[2023-12-28]. https://arxiv.org/abs/1904.07850.

[6] LIU Z K, HU J G, WENG L B, et al. Rotated region based CNN for ship detection[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2017: 900-904.

[7] 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.

[8] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5987-5995.

[9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2023-12-28]. https://arxiv.org/abs/1409.1556.

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

[11] CHEN K Y, LIU C Y, CHEN H, et al. RSPrompter: learning to prompt for remote sensing instance segmentation based on visual foundation model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62: 1-17.

[12] WU Y X, HE K M. Group normalization[EB/OL]. (2018-06-11)[2023-12-28]. https://arxiv.org/abs/1409.1556.https://arxiv.org/abs/1803.08494.

[13] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[EB/OL]. (2015-02-21)[2023-12-29]. https://arxiv.org/abs/1405.0312.

[14] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. (2019-01-04)[2023-12-29]. https://arxiv.org/abs/1711.05101.

[15] JIANG Y Y, ZHU X Y, WANG X B, et al. R2CNN: rotational region CNN for orientation robust scene text detection[EB/OL]. (2017-06-30)[2023-12-29]. https://arxiv.org/abs/1706.09579.

[16] MA J Q, SHAO W Y, YE H, et al. Arbitrary-oriented scene text de-

tection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.

[17] FENG P M, LIN Y T, GUAN J, et al. TOSO: student's-T distribution aided one-stage orientation target detection in remote sensing images[C]//Proceedings of the ICASSP 2020-IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2020: 4057-4061.

[18] LI Z G, YUAN Y, MA D D. One-stage detector from coarse to fine for rotating object of remote sensing[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS. Piscataway: IEEE Press, 2021: 5307-5310.

[19] WANG Y S, ZHANG Y, ZHANG Y, et al. SARD: towards scale-aware rotated object detection in aerial imagery[J]. IEEE Access, 2019, 7: 173855-173865.

[20] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 2844-2853.

[21] QIAN W, YANG X, PENG S L, et al. Learning modulated loss for rotated object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(3): 2458-2466.

[22] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1452-1459.

[23] YANG X, YAN J C, FENG Z M, et al. R3Det: refined single-stage detector with feature refinement for rotating object[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3163-3171.

[24] YANG X, YAN J C. Arbitrary-oriented object detection with circular smooth label[EB/OL]. (2022-03-23)[2023-12-29]. https://arxiv.org/abs/2003.05597v1.

An anchor-free rotated box remote sensing image ship object detection algorithm

LI Lu, CHEN Keyan, LIU Chenyang, SHI Zhenwei*

(School of Astronautics, Beihang University, Beijing 100191, China)

Abstract: Accurate and efficient ship detection plays a crucial role in safeguarding maritime interests and building a maritime powerhouse, with significant practical value. However, anchor boxes are the mainstay of current ship detection algorithms based on optical remote sensing images. These algorithms have limited generalization capabilities, a high computational resource requirement, and a huge number of hyperparameters. Although natural image object detection algorithms have improved these issues by adopting anchor-free methods, they can only achieve horizontal box detection and are unable to handle the unique characteristics of ship targets, such as elongated shapes, varying angles, and tight arrangements. To address these issues, this paper designs an anchor-free, single-stage remote sensing image ship detection algorithm. In particular, based on CenterNet, this research suggests a distribution-prior based confidence coefficient prediction branch to produce higher quality positive samples for the ship detection job. More precise rotation angle representation is achieved by limiting the output space of angles using a hyperbolic activation-based angle prediction branch. The use of a variable positive-negative sample label assignment technique can speed up network convergence by providing dynamic, fine-tuned supervision information. Experiments on the HRSC2016 dataset validate the superiority of the proposed algorithm compared to other advanced algorithms and confirm the effectiveness of each module.

Keywords: remote sensing image; optical remote sensing image; ship object detection; anchor-free detector; rotated box object detection

Received: 2024-01-03; **Accepted**: 2024-04-23; **Published Online**: 2024-05-10 13:46

URL: link.cnki.net/urlid/11.2625.V.20240510.1036.001

Foundation items: National Natural Science Foundation of China (62271018); Teaching Reform Project of Beihang University: Comprehensive Experimental Teaching Reform of Aerospace Information Engineering (2023-15-02-08)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0819

基于SCACGAN的小样本齿轮箱故障诊断

王进花[1,*]，刘秦玮[1]，曹洁[1,2,3]，陈莉[2]

(1. 兰州理工大学 电气工程与信息工程学院，兰州 730050；2. 兰州城市学院 信息工程学院，兰州 730070；
3. 甘肃省制造信息工程研究中心，兰州 730050)

摘 要：针对辅助分类器生成对抗网络(ACGAN)在小样本齿轮箱故障诊断过程中，生成故障样本缺乏多样性，且质量较差，导致诊断准确度不高的问题，提出一种基于自校正辅助分类器生成对抗网络(SCACGAN)的齿轮箱故障诊断方法。在辅助分类器生成对抗网络中引入一个独立的分类器，改善判别器输出错误对生成样本质量造成的不良影响，并对不同齿轮箱样本的健康状况进行分类；采用最小二乘函数，提高模型的生成能力和分类能力，改善训练过程中生成样本质量不高的问题；在生成器中引入自校正卷积神经网络，增强故障特征获取的能力。实验结果表明：在小样本条件下，所提方法能够生成质量较好的故障样本，从而提高了齿轮箱的故障诊断准确度。

关 键 词：齿轮箱；小样本；辅助分类器生成对抗网络；自校正卷积神经网络；故障诊断
中图分类号：TP277; TH133.33
文献标志码：A **文章编号**：1001-5965(2026)03-0713-11

齿轮箱作为关键的机械传动部件，广泛应用于工程机械、航空航天、汽车、风力发电等行业[1]。然而，由于关键部件的磨损和冲击损坏，尤其是断齿故障可能引起机器的不均匀工作，甚至导致停机。为避免不必要的停机和昂贵的维修费用，对齿轮箱进行故障诊断至关重要[2]。

近年来，基于深度学习的故障诊断方法因其有效处理高维数据和避免手动特征提取引起误差的优点而备受关注。这些方法包括卷积神经网络(convolutional neural networks, CNN)[3-4]、自编码器(autoencoder, AE)[5-6]、长短时记忆(long short-term memory, LSTM)网络[7-8]、深度信念网络(deep belief networks, DBN)[9-10]等。然而，这些方法的性能主要受训练标签数据量和质量的制约[11-12]。在实际齿轮箱系统运行中，无法获取足够的齿轮标签故障样本，这将导致监督模型的严重过度拟合问题[13-14]。因此，针对小样本条件下的故障诊断是当前故障诊断领域的一大挑战。

生成对抗网络(generative adversarial networks, GAN)是一种解决无监督小样本故障诊断的经典模型，并取得了不错的成果。Zhou等[15]提出一种采用全局寻优方案的生成对抗网络模型，通过生成器扩充故障数据，使用自动编码器从少量故障样本中提取故障特征。Du等[16]将生成对抗网络与增量学习支持向量机相结合，采用特征选择策略选择敏感特征，用增量学习支持向量机进行故障诊断。Du等[17]引入改进的LSTM网络生成对抗网络，利用时间序列正态大数据训练故障检测模型，并以重构误差作为故障指标。然而，GAN的判别器无法区分不同故障类型，不适用于带有标签的故障样本，因此，引入辅助分类器生成对抗网络解决这一问题[18]，其将标签信息嵌入到生成器的输入噪声中，可以生成较高质量的标签样本，但其损失函数设计不完善，存在梯度消失和模式崩溃的现象，且生成器网络层结

收稿日期：2023-12-18；**录用日期**：2024-03-08；**网络出版时间**：2024-04-30 08:54
网络出版地址：link.cnki.net/urlid/11.2625.V.20240429.1101.001
基金项目：国家自然科学基金(62063020,61763028)；国家重点研发计划(2020YFB1713600)；甘肃省自然科学基金(20JR5RA463)
* **通信作者**. E-mail：wjh0615@lut.edu.cn

引用格式：王进花，刘秦玮，曹洁，等. 基于SCACGAN的小样本齿轮箱故障诊断[J]. 北京航空航天大学学报，2026，52(3)：713-723.
WANG J H，LIU Q W，CAO J，et al. Fault diagnosis of gearbox with small-sample based on SCACGAN[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：713-723 (in Chinese).

构过于单一，获取特征信息的能力较弱，影响生成故障样本的多样性。Sun等[19]提出基于Wasserstein辅助分类器生成对抗网络(Wasserstein auxiliary classifier generative adversarial networks, WAC-GAN)的故障诊断方法，引入Wasserstein距离来改进辅助分类器生成对抗网络(auxiliary classifier generative adversarial networks, ACGAN)的目标函数，并提出信号相似度指标来证明生成样本的真实性。许小伟等[20]提出改进的ACGAN，通过学习原始样本的分布特性，使用Wasserstein距离约束生成数据的重建损失，在生成器中引入循环神经网络以提高生成数据质量。李慧芳等[21]提出一种基于主动生成式过采样与深度堆叠网络的故障诊断方法。该方法使用Wasserstein距离作为新的目标函数，采用基于委员会查询的主动学习算法，设计多样性评价指标，并对ACGAN生成的高信息熵样本进行了二次筛选，同时利用筛选出的样本训练判别器，以引导生成器生成信息量丰富的少数类样本。改进的ACGAN模型在不同数据集和故障模式下表现出令人满意的诊断性能，但其判别器同时执行判别和分类任务，当其中一个环节出现错误时，会影响生成器生成的故障样本质量。

因此，本文提出基于自校正辅助分类器生成对抗网络(self-correcting auxiliary classifier generative adversarial networks, SCACGAN)的齿轮箱故障诊断方法，以ACGAN作为基础模型，引入一个独立的分类器，用来改善判别器输出错误所造成的生成样本质量降低的问题；引入最小二乘函数对其目标函数进行优化，通过博弈对抗机制同时优化生成器和判别器，使模型的泛化能力进一步提升；在生成器中引入自校正卷积模块，增强获取特征信息的能力，增加生成故障样本的多样性。

1 相关理论

1.1 LSGAN模型

GAN在无监督故障诊断领域有很好的表现，模型的损失函数采用了二元交叉熵，具体如下：

$$\min_G \max_D (D,G) = E_{x\sim P_r}(\ln D(x)) + E_{z\sim P_z}(\ln(1-D(G(z)))) \tag{1}$$

式中：x为真实样本；z为随机噪声；D为判别器；G为生成器；$G(z)$为生成样本；E为期望值；P_r为真实数据的分布；P_z为生成器前置的随机分布。

二元交叉熵是衡量2个概率分布之间差异的方法。当真实数据分布P_r和生成数据分布P_g的差异很小时，二元交叉熵的值会减小。当二元交叉熵的值接近0时，生成器的梯度会变得极小，从而使优化生成器参数的过程变得困难。简而言之，当生成的样本被错误地分类为真实数据时，迭代会停滞，导致生成器无法有效学习真实数据分布，从而降低生成样本的质量。

为改善GAN在损失函数设计方面的不足，Mao等[22]提出了最小二乘生成对抗网络(least squares generative adversarial networks, LSGAN)，将原始GAN中的二元交叉熵损失函数替换成最小二乘函数，成功解决了梯度消失的问题。

LSGAN的损失函数为

$$\min_D V_{\mathrm{LSGAN}}(D) = \frac{1}{2}E_{x\sim P_r(x)}((D(x)-b)^2) + \frac{1}{2}E_{z\sim p_z(z)}((D(G(z))-a)^2) \tag{2}$$

$$\min_G V_{\mathrm{LSGAN}}(G) = \frac{1}{2}E_{x\sim P_z(z)}((D(G(z))-c)^2) \tag{3}$$

式中：a、b、c均为常数，表示在训练过程中正例反例的标记。通常情况下，将正例标为1，反例标为0，当参数$a=0$，$b=c=1$时，可以得出适合训练的优化函数。

采用最小二乘函数作为损失函数，可以惩罚来自生成器却又被分类为真的生成数据样本，将其推向决策边界，进而可有效解决梯度消失造成的训练困难问题。虽然LSGAN在GAN的损失函数上进行了有效改进，提升了模型性能，但其网络结构还是无法处理带标签的故障样本，仍然存在不足之处。

1.2 ACGAN模型

GAN作为一种无监督生成模型，能够很好的处理无标签样本数据，但其过于自由的生成模式使得训练过程和结果都不可控。ACGAN是GAN的一个有监督的改进模型，其由生成器G和判别器D组成，并将标签信息嵌入到生成器的随机噪声输入中，其判别器不仅可以区分样本真伪，还可以区分样本类别。ACGAN的模型结构如图1所示。

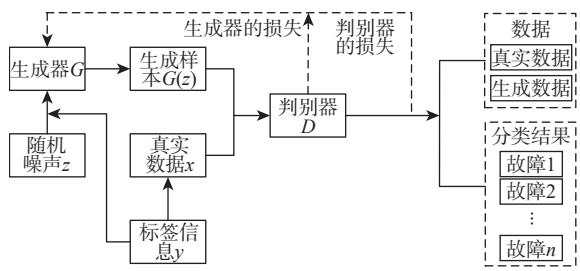


图1 ACGAN结构

Fig. 1 Architecture of ACGAN

ACGAN的损失函数包含以下2部分：

$$L_{\mathrm{S}} = E_{x\sim P_r}(\ln D(x)) + E_{z\sim P_z}(\ln(1-D(G(z,c_{\mathrm{g}})))) \tag{4}$$

$$L_{\mathrm{C}} = E_{x\sim P_r}(\ln P(y=c_{\mathrm{r}}|x)) + E_{z\sim P_z}(\ln P(y=c_{\mathrm{g}}|G(z,c_{\mathrm{g}}))) \tag{5}$$

式中: y为标签信息; L_S为衡量生成样本与真实样本差异的损失函数; L_C为衡量样本类别的损失函数; $G(z,c_g)$为生成的样本; c_r和c_g分别为x和$G(z,c_g)$的标签; $P(y=c_r|x)$和$P(y=c_g|G(z,c_g))$分别为真实样本和生成样本类别标签的条件概率分布。

ACGAN 在 GAN 的网络模型上进行了较大的改进, 可以很好的处理带标签的故障数据, 但是其损失函数仍然采用二元交叉熵设计, 不能很好的衡量生成样本与真实样本之间的差异, 同时网络层结构简单, 获取特征信息的能力不足, 影响生成故障样本的多样性, 而且判别器要同时实现判别和分类的功能, 需要考虑两者的兼容性问题。

1.3　自校正卷积神经网络

ACGAN 的生成器网络层中通常难以获得较大感受野, 无法捕获足够的特征信息, 影响生成故障样本的多样性, 导致诊断准确度降低。

自校正卷积 (self-calibrated convolution, SCConv) 通过在 2 个不同尺度的空间中进行卷积特征转换, 实现了高效地收集每个空间位置的上下文信息。其中, 一个尺度是原始尺度空间, 特征图与输入共享相同的分辨率; 另一个是较小的潜在空间, 经过下采样后分辨率较低。在小的潜在空间中, 将转换后的嵌入视为参考, 用以指导原始特征空间中的特征转换过程, 因为其具有更大的感受野, 有助于捕获整个特征区域。即 SCConv 通道维度被分为 2 个分支: 分支 1 利用下采样来增大特征图的感受野, 而分支 2 用于常规的卷积操作。最后, 合并这 2 个分支的通道信息, 以增强生成器的特征提取和表达能力。

SCConv 结构通过自校正操作, 在每个空间位置周围建立远程空间和通道间的依存关系, 不仅提高模型整体性能, 还不增加额外参数和复杂性。

SCConv 的结构如图 2 所示。

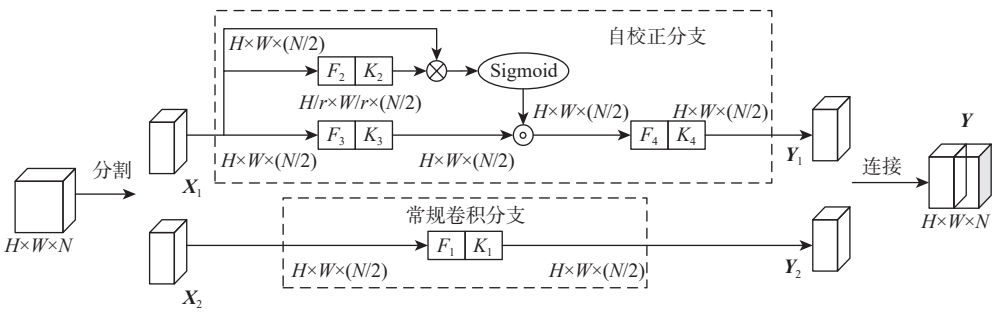


图 2　SCConv 结构

Fig. 2　Architecture of SCConv

首先, 对输入特征图进行拆分, 保持宽、高不变, 通道数减半, 得到 2 个特征图, 如下:

$$\boldsymbol{X}_1,\boldsymbol{X}_2=S(\boldsymbol{X}) \tag{6}$$

式中: $\boldsymbol{X}$为输入特征图; $\boldsymbol{X}_1$和$\boldsymbol{X}_2$为拆分后的 2 个特征图。

首先, 输入$H\times W\times N$的特征图, 其中, H、W和N分别为高度、宽度和通道数, 输出通道也为N的卷积中, 卷积核的维度为$H\times W\times N\times N$; 然后, 将卷积核拆分为$K_1$、$K_2$、$K_3$和$K_4$这 4 个部分, 每个部分的维度为$H\times W\times(N/2)\times(N/2)$, 用以有效地收集每个空间位置的上下文信息; 接着, 执行自校准操作。对特征图$\boldsymbol{X}_1$进行处理, 主要步骤包括平均池化、上采样、Sigmoid 激活和卷积操作, 如式 (7)～式 (10) 所示。

$$\boldsymbol{T}_1=\mathrm{Avgpool}_r(\boldsymbol{X}_1) \tag{7}$$

式中: r为池化处理的下采样率; $\boldsymbol{T}_1$为经过下采样得到的低维特征图; Avgpool(r)进行r倍下采样操作。

$$\boldsymbol{X}_1'=\mathrm{Up}(F_2(\boldsymbol{T}_1))=\mathrm{Up}(\boldsymbol{T}_2\times K_2) \tag{8}$$

式中: Up为双线性上采样操作; K_2为卷积特征变换; F_2为以K_2为核的卷积过程; $\boldsymbol{X}_1'$为上采样后的特征图。

$$\boldsymbol{Y}_1'=F_3(\boldsymbol{X}_1)\cdot\sigma\left(\boldsymbol{X}_1+\boldsymbol{X}_1'\right) \tag{9}$$

式中: σ为 Sigmoid 函数; F_3为以K_3为核的卷积过程; $\boldsymbol{Y}_1'$为经过F_3卷积与 Sigmoid 函数的注意力特征图相融合的输出结果。

$$\boldsymbol{Y}_1=F_4\left(\boldsymbol{Y}_1'\right)=\boldsymbol{Y}_1'K_4 \tag{10}$$

式中: $\boldsymbol{Y}_1$为自校正部分的输出特征; K_4为卷积特征变换。

另一条分支上的特征图$\boldsymbol{X}_2$经过卷积核K_1提取得到特征$\boldsymbol{Y}_2$, 最终将$\boldsymbol{Y}_1$和$\boldsymbol{Y}_2$进行拼接操作, 得到最终的输出特征。与传统卷积相比, 自校正卷积有效地扩大了卷积层的感受野, 增强特征的感知能力, 而且不会引入更多的复杂度。

2　本文方法

针对 LSGAN 和 ACGAN 的不足, 本文提出一种自校正辅助分类器生成对抗网络模型, 在 ACGAN 的网络结构和损失函数上进行优化改进, 改善生成

故障样本的质量和多样性，从而提高齿轮箱故障诊断准确度。

2.1 SCACGAN的模型结构

2.1.1 生成器的结构改进

ACGAN的生成器网络层结构相对简单，对特征信息的学习能力较弱，导致生成的样本缺乏多样性。为此，本文在网络层中引入自校正神经网络，捕捉数据中的多样性和复杂性。改进后的SCACGAN生成器结构如图3所示。

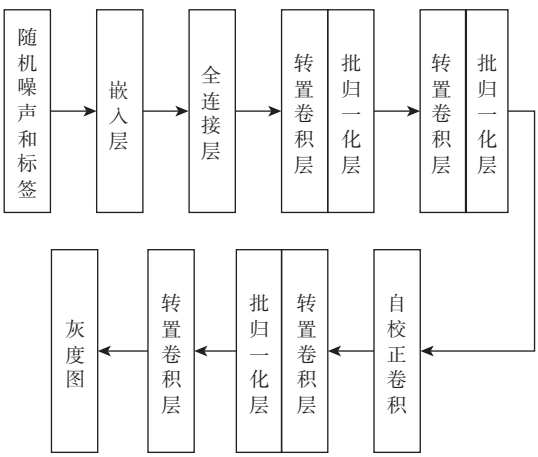


图3 SCACGAN生成器结构

Fig. 3 Architecture of SCACGAN generator

首先，将随机噪声和标签，通过全连接层转换为2×2×512的张量，经由2个步长为2的5×5转置卷积核进行初步特征的提取。

其次，将8×8×64的张量作为SCConv的输入。该张量分成$\boldsymbol{X}_1$和$\boldsymbol{X}_2$两部分，其大小均为8×8×32。将大小为(64,64,5,5)的原卷积核$\boldsymbol{K}$分为4个小卷积核$\boldsymbol{K}_1$、$\boldsymbol{K}_2$、$\boldsymbol{K}_3$和$\boldsymbol{K}_4$，其大小均为(32,32,5,5)，用来收集不同类型的上下文信息，卷积核在分支1中对$\boldsymbol{X}_1$进行自校正操作，得到$\boldsymbol{Y}_1$，而卷积核$\boldsymbol{K}_1$于分支2中对原尺度特征$\boldsymbol{X}_2$进行卷积操作，以保留原始空间信息，得到$\boldsymbol{Y}_2$。将原尺寸空间输出特征$\boldsymbol{Y}_1$和$\boldsymbol{Y}_2$进行拼接，得到与输入张量大小一致的特征$\boldsymbol{Y}$。在分支1的自校正处理中，首先，使用大小为4×4的平均池化层对特征$\boldsymbol{X}_1$进行下采样，得$\boldsymbol{T}_1$，如式(7)所示，再经过卷积核$\boldsymbol{K}_2$提取特征，并进行上采样得$\boldsymbol{X}_1'$，将尺寸恢复到$\boldsymbol{X}_1$大小，如式(8)所示，上采样方式为双线性插值法；后与$\boldsymbol{X}_1$相加，使用Sigmoid激活函数映射计算出校正权重，并与卷积核$\boldsymbol{K}_3$提取特征后相乘进行校正，得$\boldsymbol{Y}_1'$，如式(9)所示，最后，通过$\boldsymbol{K}_4$卷积核操作得到该分支校正后的特征$\boldsymbol{Y}_1$，如式(10)所示。

经过2次转置卷积层，生成尺度为32×32×1的灰度图。在部分转置卷积层后引入批归一化层，对训练参数进行优化，使得模型训练更加稳定和容易。SCACGAN中生成器的具体结构参数如表1所示。

表1 SCACGAN的生成器结构参数

Table 1 Generator structural parameters of SCACGAN

网络层	结构参数	$H\times W\times N$
Dense	2 048	2×2×512
Conv1	(128,5,2)	4×4×128
Conv2	(64,5,2)	8×8×64
SCConv	(32,5)(经过$\boldsymbol{K}_1$) (32,5)(经过$\boldsymbol{K}_2$) (32,5)(经过$\boldsymbol{K}_3$) (32,5)(经过$\boldsymbol{K}_4$)	8×8×64
Conv3	(8,5,2)	16×16×8
Conv4	(1,3,2)	32×32×1

2.1.2 判别器的结构改进

ACGAN的判别器同时进行2种操作，既要判别样本真伪，又要对标签样本进行分类，当其中任何环节出现错误时，就会通过损失函数将错误的信息反馈给生成器，影响生成器生成的故障样本质量。为避免这种情况，在ACGAN模型中引入一个独立分类器，专门处理标签分类任务，从而减轻判别器的负担。本文的SCACGAN判别器和分类器结构如图4所示。

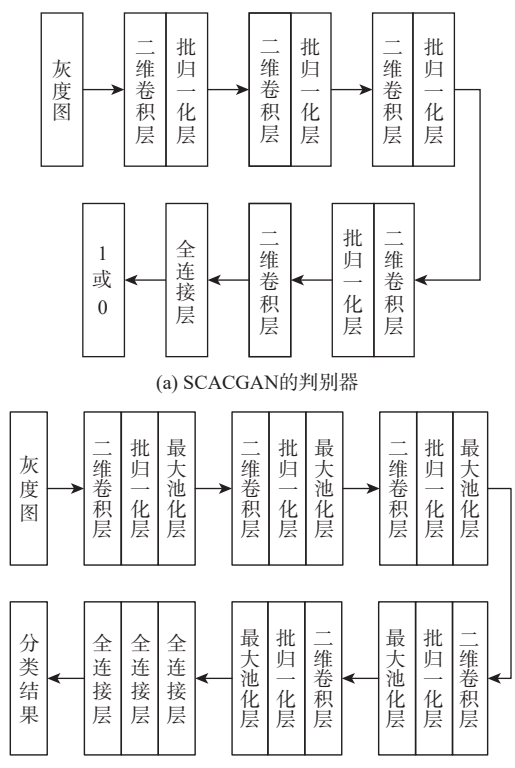


图4 SCACGAN的判别器和分类器

Fig. 4 Discriminator and classifier of SCACGAN

从图4(a)可知，将尺度为32×32×1的灰度图作为判别器的输入，经过5个2维卷积层，可得到2×2×256的张量，通过全连接层进行二元分类。从

图 4(b) 可知, 将尺度为 32×32×1 的灰度图作为分类器的输入, 经过 5 个 2 维卷积层、批归一化层和最大池化层, 得到尺度为 2×2×256 的张量, 最后, 经过 3 个全连接层, 得到分类样本。分类器中的最大池化层卷积核大小为 2×2, 步长为 1, SCACGAN 中判别器和分类器的具体结构参数如表 2 所示。

表 2　SCACGAN 的判别器和分类器结构参数

Table 2　Discriminator and classifier structural parameters of SCACGAN

模型	网络层	结构参数	输出尺寸
判别器	Conv1	(16,5,2)	16×16×16
	Conv2	(32,3,2)	8×8×32
	Conv3	(64,3,2)	4×4×64
	Conv4	(128,3,2)	2×2×128
	Conv5	(256,3,1)	2×2×256
	Dense	1	1
分类器	Conv1	(16,3,2)	16×16×16
	Conv2	(32,3,2)	8×8×32
	Conv3	(64,3,2)	4×4×64
	Conv4	(128,3,2)	2×2×128
	Conv5	(256,3,1)	2×2×256
	Dense	256	256
	Dense	128	128
	Dense	9	9

2.1.3　SCACGAN 的结构

SCACGAN 由生成器、判别器和独立的分类器组成。其中, 生成器以随机噪声z和附加的条件信息y作为输入, 用来生成带标签的故障样本, 判别器以生成样本$G(z)$和附加条件信息y的真实数据x作为输入, 用来判定数据的真伪, 而分类器用来区分标签样本和做最后的故障诊断。为增加对齿轮箱数据的特征提取能力, 分类器的网络层被设计为 CNN, 可以直接从样本数据中提取特征, 并通过 Softmax 层进行准确分类。SCACGAN 结构如图 5 所示。

2.2　SCACGAN 的目标函数

ACGAN 中的损失函数通常采用二元交叉熵设计。当更新生成器时, 若生成的假样本位于靠近决策边界的正确一侧, 会被误判为真实样本, 影响判别器和生成器的训练。通过采用最小二乘函数, 可以对位于决策边界正确一侧的假样本进行惩罚, 并将其拉回决策边界错误一侧, 有助于衡量真假样本的距离, 缓解梯度消失带来的训练困难问题。

本文 SCACGAN 模型, 由于判别器只用来判别真伪, 为更好地衡量生成样本和真实样本的距离,

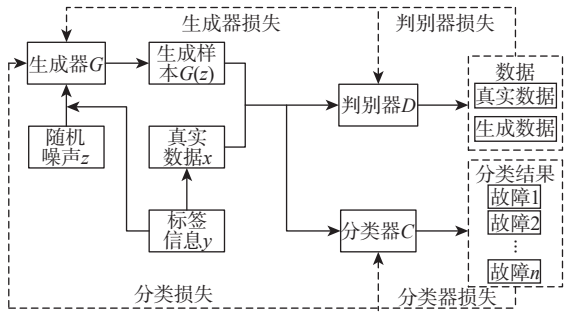


图 5　SCACGAN 结构

Fig. 5　Architectural of SCACGAN

在生成器和判别器中引入最小二乘损失函数, 并结合独立的分类器重新定义所提模型的目标函数, 具体如下:

$$\min_D L'_D = \frac{1}{2}E_{x\sim P_r(x)}((D(x)-1)^2)+\frac{1}{2}E_{z\sim P_z(z)}((D(G(z,c_g)))^2) \tag{11}$$

$$\min_G L'_G = \frac{1}{2}E_{z\sim P_z(z)}\left(\left(D\left(G\left(z,c_g\right)\right)-1\right)^2\right) \tag{12}$$

$$L'_S = \min_D L'_D + \min_G L'_G \tag{13}$$

$$L_C^R = E_{x\sim P_r(x)}(\ln P(y=c_r|x)) \tag{14}$$

$$L_C^G = E_{z\sim P_z(z)}(\ln P(y=c_g|G(z,c_g))) \tag{15}$$

$$L'_C = \lambda_1 L_C^R + \lambda_2 L_C^G \tag{16}$$

$$L = L'_S + L'_C \tag{17}$$

式中: L'_S为衡量生成样本与真实样本差异的损失函数; L'_C为衡量样本类别的损失函数; L为总的损失函数; L'_D和L'_G分别为判别器D和发生器G的损失函数; L_C^R和L_C^G分别为分类器对真实样本和生成样本的损失函数; y为齿轮箱故障样本的类别标签信息; λ_1、λ_2为分类器损失函数的比例因子, 用于调整损失函数中不同部分的重要性, 从而影响模型在训练过程中对真实数据和生成数据的关注程度。为更好地学习真实样本的特征和分布, 需要充分关注分类器对真实样本的损失函数, 将λ_1设置为 1。为平衡模型在真实数据和生成数据上的学习过程, 将λ_2设置为 0.1。

2.3　故障诊断流程图

改进辅助分类器生成对抗网络模型的齿轮箱故障诊断方法训练步骤如下。

1) 数据预处理。使用小波变换将一维齿轮振动信号转化为灰度图, 收集数据并制作新的数据集, 把其按照 7:3 划分为训练集和测试集。

2) 故障样本扩充。用训练集训练构建的 SCACGAN 模型, 学习训练样本的分布特征, 判断其是否达到最大 Epoch, 若不满足性能要求, 则继续

训练网络；若满足性能要求，则终止训练，保存模型参数。

3）故障诊断。SCACGAN生成与真实样本相对应类别的假样本，将其扩充到初始训练集，得到混合训练样本。随后，对分类器诊断模型进行训练，并通过测试集验证诊断效果。故障诊断流程如图6所示。

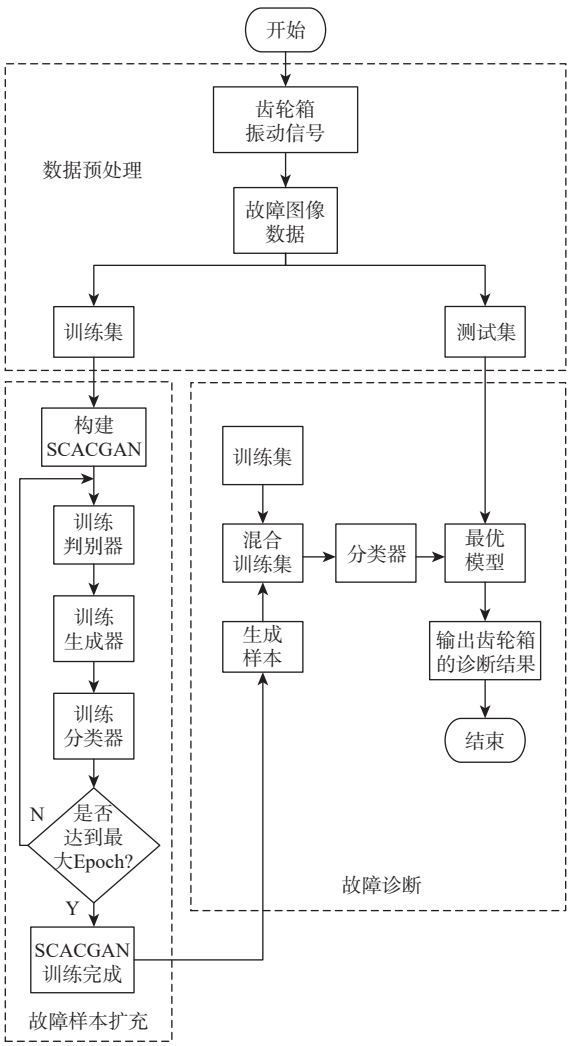


图6 SCACGAN的小样本齿轮箱故障诊断流程图

Fig. 6 Process diagram of SCACGAN for small-sample gearbox fault diagnosis

3 基于SCACGAN的故障诊断实验

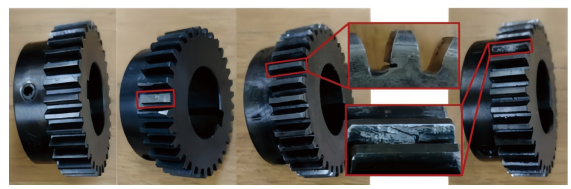
3.1 齿轮箱故障数据的介绍和预处理

使用康涅狄格大学的齿轮振动数据对本文方法进行测试。如图7所示，实验数据来自具有可更换齿轮的基准两级变速箱，其齿轮速度由电机控制，转矩由磁制动器提供，通过改变输入电压进行调节。如图8所示，输入轴上的小齿轮呈现9种不同的齿轮状况。如图8(a)所示，包括健康状况、缺齿、齿根裂纹、剥落。如图8(b)所示，包括5种不同严重程度的削尖，从左到右严重程度逐渐加深。

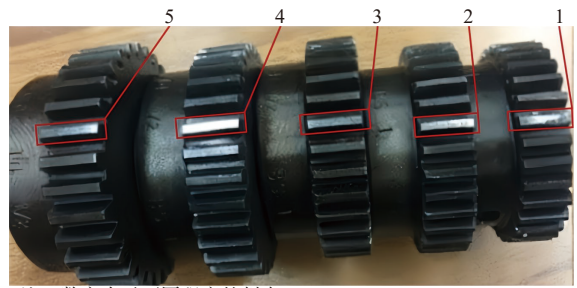


图7 康涅狄格大学的齿轮箱结构

Fig. 7 Gearbox structure at the University of Connecticut



(a) 健康、缺齿、齿根裂纹、剥落



注：数字表示不同程度的削尖

(b) 5种不同程度的削尖

图8 齿轮的9个不同健康状况

Fig. 8 Nine different health conditions of gears

信号通过采样频率为20 kHz。

从每个状态抽取1 024个连续点作为一个样本，每类取50个作为可用样本，用于SCACGAN生成新样本。齿轮样品详情如表3所示。

为从齿轮箱振动信号中获取更加丰富的特征信息，使用小波变换的方法将原始1维振动信号转换为灰度图像，生成图像的像素分辨率为32像素×32像素。9种故障样本的灰度图如图9所示。

3.2 生成样本质量评估

在扩充数据集前，先要确保改进模型能够生成高质量的故障假样本。对于各类故障样本，每种类型对应生成50个假样本。为比较SCACGAN模型、ACGAN模型和WAC-GAN模型生成的样本质量，3个模型采用结构相似性(structural similarity index measure, SSIM)评价指标对生成样本的质量进行定量评价。SSIM旨在从亮度、对比度和结构3个方面衡量2幅图像的相似度，定义如下：

表 3　齿轮样品详情

Table 3　Gear sample details

齿轮的健康状态	分类标签
健康	0
缺齿	1
齿根裂纹	2
剥落	3
碎屑尖端1	4
碎屑尖端2	5
碎屑尖端3	6
碎屑尖端4	7
碎屑尖端5	8

$$f_{\mathrm{SSIM}}(u,v)=l(u,v)c(u,v)s(u,v)=\frac{2\mu_u\mu_v+C_1}{\mu_u^2+\mu_v^2+C_1}\cdot\frac{2\sigma_u\sigma_v+C_2}{\sigma_u^2+\sigma_v^2+C_2}\cdot\frac{\sigma_{uv}+C_3}{\sigma_u\sigma_v+C_3}\tag{18}$$

式中: u和v为 2 个灰度图像样本; μ_u、μ_v、σ_u、σ_v分别为u和v的均值和方差; σ_{uv}为u和v的协方差; $C_1=(0.01V)^2$, $C_2=(0.03V)^2$, $C_3=0.5C_2$为避免分母为 0 的 3 个常数, V为像素值, 取值范围一般为 0～255。f_{SSIM} 值越大表示 2 幅图像之间的相似性越高, 其对比结果如图 10 所示。

由图 10 可知, SCACGAN 模型给出的f_{SSIM}值始终大于 ACGAN 模型和 WAC-GAN 模型的f_{SSIM}值, 这说明 SCACGAN 模型在样本生成过程中能够更好地保留真实样本的细节和结构特征, 导致生成的样本在个体水平上更接近真实样本, 所生成的齿轮箱样本质量更高。

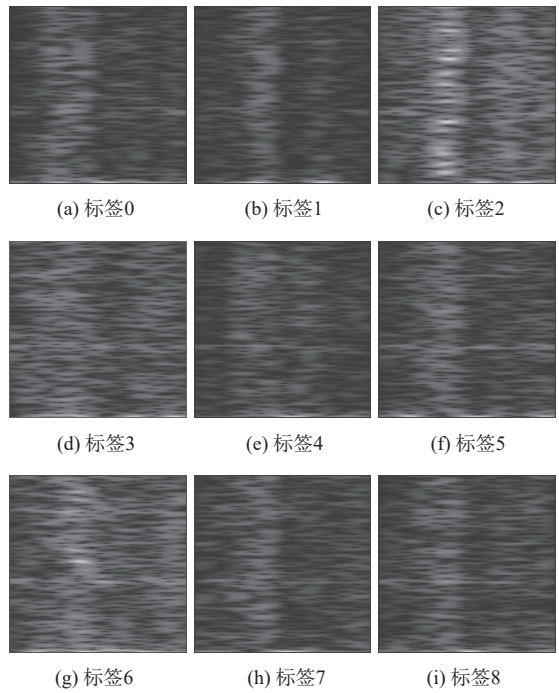


图 9　不同健康状况下的灰度图

Fig. 9　Grayscale images under different health conditions

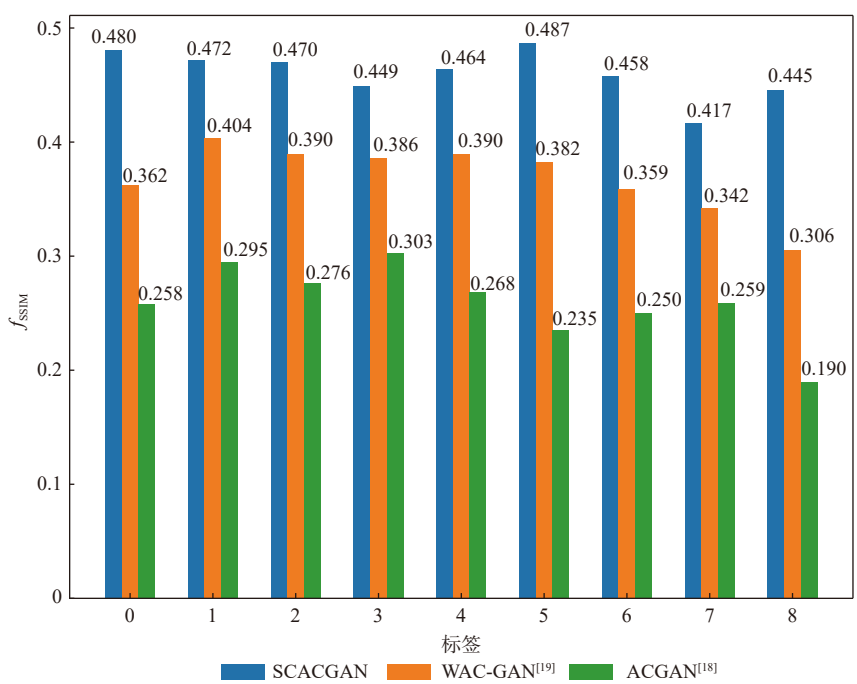
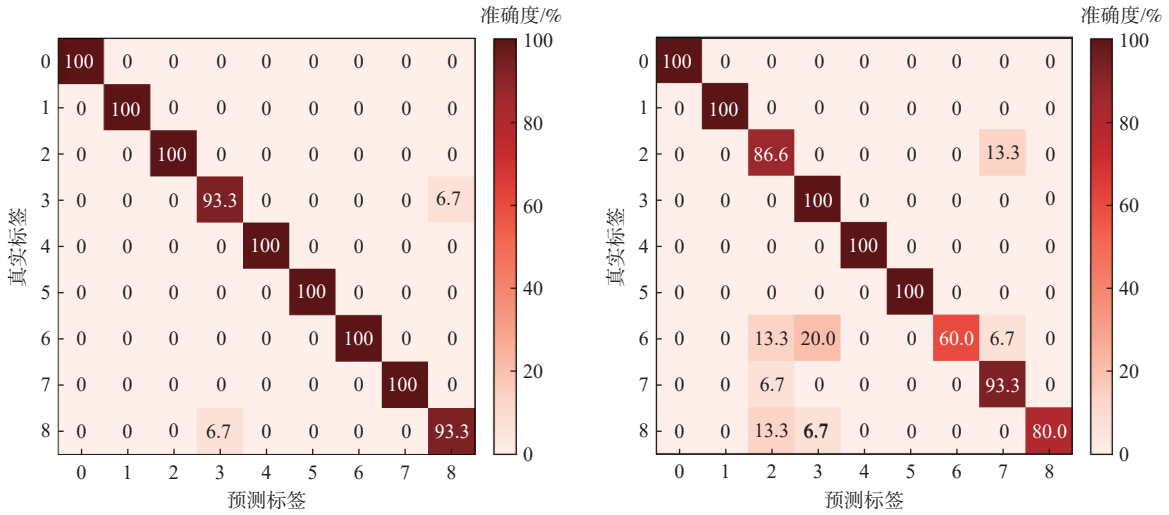


图 10　基于 SSIM 的对比结果

Fig. 10　Contrast results based on SSIM

3.3　相同样本数量下的诊断结果对比

每个状态随机抽取 50 个齿轮样本作为数据集, 训练集用于 SCACGAN 模型训练, 生成每类样本 165 个。将生成的样本扩充进训练集中, 重新训练模型中的分类器, 最后, 用划分好的测试集对齿轮箱故障样本进行诊断, 得到的混淆矩阵如图 11(a)



(a) SCACGAN的混淆矩阵 (b) ACGAN的混淆矩阵[18]

图 11 混淆矩阵图

Fig. 11 Confusion matrix plot

所示。同等条件下训练 ACGAN 模型，并生成同样的样本数量，得到的混淆矩阵如图 11(b) 所示。

由图 11 可知，在测试集每类 15 个的样本条件下，SCACGAN 模型相对于 ACGAN 模型有更好的准确度。

为更近一步的验证本文方法的齿轮箱故障诊断效果，采用聚类的方法，与不同诊断方法进行比较。本文在相同数据集下选择 LSTM 网络模型、CNN 模型、CGAN 模型、ACGAN 模型、WAC-GAN 模型作为对比模型进行实验，所有模型的 batch size 设置为 32，具体参数如表 4 所示。

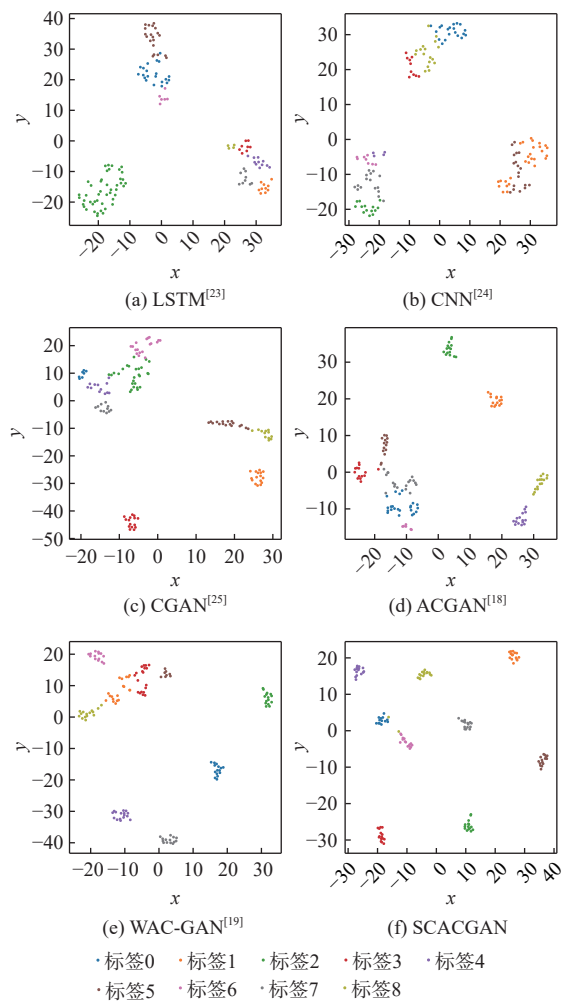
表 4 对比模型参数

Table 4 Comparison model parameters

模型	模块名称	学习率
LSTM[23]		0.01
CNN[24]		0.01
CGAN[25]	生成器	0.000 2
	判别器	0.000 1
ACGAN[18]	生成器	0.000 3
	判别器	0.000 2
WAC-GAN[19]	生成器	0.000 3
	判别器	0.000 2
SCACGAN	生成器	0.000 1
	判别器	0.000 1
	分类器	0.000 01

CGAN 模型、ACGAN 模型、WAC-GAN 模型和 SCACGAN 模型生成的每类样本数均为 165，每种模型的 Epoch 参数设置为 200，此时 SCACGAN 模型训练到最优，聚类效果如图 12 所示。

由图 12 可知，在齿轮箱样本数据为 50 的情况下，SCACGAN 模型相比于其他深度学习模型聚类效果最好，这说明 SCACGAN 模型可以从原始信号中提取高效的故障判别特征，生成质量较高的故障样本，在数据量较低的情况下可以得到一个较好的诊断结果。



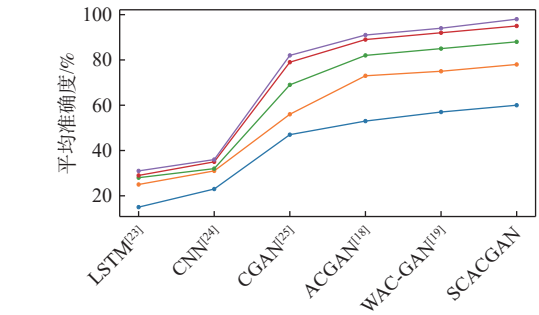
(a) LSTM[23] (b) CNN[24] (c) CGAN[25] (d) ACGAN[18] (e) WAC-GAN[19] (f) SCACGAN

图 12 不同模型的特征散点图

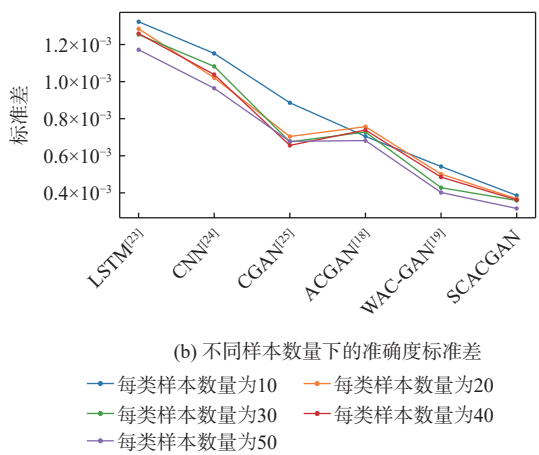
Fig. 12 Scatter plot of features for different models

3.4　不同样本数量下的诊断结果对比

为验证本文 SCACGAN 模型在小样本条件下的性能，每类的样本数量分别取 10、20、30、40 和 50 进行实验。在不同样本数量下，对 LSTM 网络模型、CNN 模型、CGAN 模型、ACGAN 模型、WAC-GAN 模型和 SCACGAN 模型这 6 种深度学习故障诊断模型进行了 10 次实验，对最后的准确度取平均值，并且计算对应的准确度标准差，验证模型的鲁棒性能，CGAN 模型、ACGAN 模型、WAC-GAN 模型和 SCACGAN 模型生成的每类样本数均为 165，每种模型的 Epoch 参数设置为 200，诊断结果如图 13 所示。



(a) 不同样本数量下的平均准确度



(b) 不同样本数量下的准确度标准差

图 13　不同样本数量下的平均准确度和准确度标准差

Fig. 13　Average accuracy and standard deviation under different sample sizes

从图 13(a) 中可以看出，随着样本数量的增多，网络模型得到充分训练，6 种模型的平均准确度都得到了提高，而 SCACGAN 模型在不同的样本条件下，诊断效果最好。

由图 13(b) 可知，SCACGAN 模型的准确度标准差相比其他模型较小，诊断结果较为稳定，拥有更好的鲁棒性。以上结果表明，本文 SCACGAN 模型在小样本条件下，齿轮箱故障诊断方法相比其他几种深度学习故障诊断模型具有明显优势。

4　结　论

本文针对齿轮箱在故障数据不足的情况下，ACGAN 的判别器输出错误会影响生成样本质量，生成器提取特征能力不足，导致故障样本缺乏多样性，损失函数存在设计缺陷，易出现模式崩溃和梯度消失的现象，导致诊断准确度不高的问题，提出自校正辅助分类器生成对抗网络的故障诊断方法。用美国康涅狄格大学的齿轮振动数据集对该方法进行验证，并与其他方法进行了对比，可以得出以下结论。

1）通过引入一个独立的分类器，改进 ACGAN 的判别器，解决判别和分类功能的兼容性问题。

2）在生成器和判别器之间，引入最小二乘损失函数进行改进，结合独立的分类器重新定义目标函数，通过对抗训练提高了模型的生成能力和分类能力，有效改善了训练过程中生成样本质量不高的问题。

3）通过在生成器中引入自校正卷积神经网络，加强特征信息的获取能力，增加生成故障样本的多样性。

在齿轮箱的不同故障样本数据量下，与 LSTM 网络模型、CNN 模型、CGAN 模型、ACGAN 模型、WAC-GAN 模型进行对比实验，本文模型的诊断准确度优于这 5 种模型。说明本文模型在小样本情况下，可以得到较高的诊断准确度，且具有更好的鲁棒性。

参考文献（References）

[1] ZHANG M, LI D C, WANG K S, et al. An adaptive order-band energy ratio method for the fault diagnosis of planetary gearboxes[J]. Mechanical Systems and Signal Processing, 2022, 165: 108336.

[2] XUE S, HOWARD I, WANG C S, et al. Dynamic modelling of the gear system under non-stationary conditions using the iterative convergence of the tooth mesh stiffness[J]. Engineering Failure Analysis, 2022, 131: 105908.

[3] RUAN D W, WANG J, YAN J P, et al. CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis[J]. Advanced Engineering Informatics, 2023, 55: 101877.

[4] JIA L S, CHOW T W S, YUAN Y X. GTFE-Net: a gramian time frequency enhancement CNN for bearing fault diagnosis[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105794.

[5] WANG S D, LIU Z B, JIA Z, et al. Composite fault diagnosis of analog circuit system using chaotic game optimization-assisted deep ELM-AE[J]. Measurement, 2022, 202: 111826.

[6] YANG D G, KARIMI H R, SUN K K. Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples[J]. Neural Networks, 2021, 141: 133-144.

[7] LEI Y X, KARIMI H R, CHEN X F. A novel self-supervised deep LSTM network for industrial temperature prediction in aluminum processes application[J]. Neurocomputing, 2022, 502: 177-185.

[8] ZHANG S Y, QIU T. Semi-supervised LSTM ladder autoencoder

for chemical process fault diagnosis and localization[J]. Chemical Engineering Science, 2022, 251: 117467.

[9] JIN Z Z, HE D Q, WEI Z X. Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN[J]. Engineering Applications of Artificial Intelligence, 2022, 110: 104713.

[10] GAO S Z, XU L T, ZHANG Y M, et al. Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN[J]. ISA Transactions, 2022, 128: 485-502.

[11] LI Y J, WANG Y B, ZHANG Y, et al. Diagnosis of inter-turn short circuit of permanent magnet synchronous motor based on deep learning and small fault samples[J]. Neurocomputing, 2021, 442: 348-358.

[12] MAO W T, FENG W S, LIU Y M, et al. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis[J]. Mechanical Systems and Signal Processing, 2021, 150: 107233.

[13] WANG Z S, XUAN J P. Intelligent fault recognition framework by using deep reinforcement learning with one dimension convolution and improved actor-critic algorithm[J]. Advanced Engineering Informatics, 2021, 49: 101315.

[14] HE Z Y, SHAO H D, DING Z Y, et al. Modified deep autoencoder driven by multisource parameters for fault transfer prognosis of aeroengine[J]. IEEE Transactions on Industrial Electronics, 2021, 69(1): 845-855.

[15] ZHOU F N, YANG S, FUJITA H, et al. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data[J]. Knowledge-Based Systems, 2020, 187: 104837.

[16] DU Z M, CHEN K, CHEN S L, et al. Deep learning GAN-based data generation and fault diagnosis in the data center HVAC system[J]. Energy and Buildings, 2023, 289: 113072.

[17] DU W Y, YANG J P, MENG G L. Fault diagnosis for dynamic system based on the independent latent space reconstruction of generative adversarial network[J]. Journal of Process Control, 2023, 125: 28-40.

[18] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier gans[C]//Proceedings of the International Conference on Machine Learning. Sydney: PMLR, 2017: 2642-2651.

[19] SUN Q, PENG F, YU X H, et al. Data augmentation strategy for power inverter fault diagnosis based on Wasserstein distance and auxiliary classification generative adversarial network[J]. Reliability Engineering & System Safety, 2023, 237: 109360.

[20] 许小伟, 韦道明, 严运兵, 等. 基于改进 ACGAN 的永磁同步电机数据扩张方法[J]. 哈尔滨工业大学学报, 2023, 55(10): 114-121.
XU X W, WEI D M, YAN Y B, et al. Data expansion method of permanent magnet synchronous motor based on improved ACGAN[J]. Journal of Harbin Institute of Technology, 2023, 55(10): 114-121(in Chinese).

[21] 李慧芳, 徐光浩, 黄双喜. 基于主动生成式过采样和深度堆叠网络的轴承故障诊断[J]. 计算机集成制造系统, 2023, 29(1): 146-159.
LI H F, XU G H, HUANG S X. Active generative oversampling and deep stacking network based bearing fault diagnosis approach[J]. Computer Integrated Manufacturing Systems, 2023, 29(1): 146-159 (in Chinese).

[22] MAO X D, LI Q, XIE H R, et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2813-2821.

[23] LEI J H, LIU C, JIANG D X. Fault diagnosis of wind turbine based on long short-term memory networks[J]. Renewable Energy, 2019, 133: 422-432.

[24] LI Y, ZHANG Y, LIU W, et al. A fault pattern and convolutional neural network based single-phase earth fault identification method for distribution network [C]//Proceedings of the IEEE Innovative Smart Grid Technologies-Asia. Piscataway: IEEE press, 2019: 838-843.

[25] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. (2014-12-06)[2023-12-16]. https://arxiv.org/abs/1411.1784.

Fault diagnosis of gearbox with small-sample based on SCACGAN

WANG Jinhua[1, *]，LIU Qinwei[1]，CAO Jie[1, 2, 3]，CHEN Li[2]

(1. School of Electrical Engineering and Information Engineering，Lanzhou University of Technology，Lanzhou 730050，China；
2. School of Information Engineering，Lanzhou City University，Lanzhou 730070，China；
3. Gansu Manufacturing Information Engineering Research Center，Lanzhou 730050，China)

Abstract：A new method for gearbox fault diagnosis based on the self-correcting auxiliary classifier generative adversarial networks (SCACGAN) is suggested in response to the limited diversity and low quality of fault samples produced by the auxiliary classifier generative adversarial networks (ACGAN) during the small-sample gearbox fault diagnosis process, which subsequently results in low diagnostic accuracy. Firstly, an independent classifier is introduced into the auxiliary classifier generative adversarial network to mitigate the adverse impact of discriminator output errors on the quality of generated samples, and to classify the health status of different gearbox samples. Secondly, the problem of low-quality generated samples during the training phase is addressed by using the least squares function to improve the model's generation and classification skills. Lastly, a self-correcting convolutional neural network is integrated into the generator to enhance the capability of fault feature acquisition. Experimental results demonstrate that under small-sample conditions, the proposed approach is capable of generating higher-quality fault samples, thereby improving the accuracy of gearbox fault diagnosis.

Keywords：gearbox；small-sample；auxiliary classifier generative adversarial networks；self-correcting convolutional neural networks；fault diagnosis

Received：2023-12-18；**Accepted**：2024-03-08；**Published Online**：2024-04-30 08：54

URL：link.cnki.net/urlid/11.2625.V.20240429.1101.001

Foundation items：National Natural Science Foundation of China (62063020,61763028); National Key Research and Development Program of China (2020YFB1713600); Natural Science Foundation of Gansu Province (20JR5RA463)

* **Corresponding author**. E-mail：wjh0615@lut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0785

基于信息熵的客机倾斜姿态应急撤离评价

程明，韦柯宇*
(中国民航大学 安全科学与工程学院，天津 300300)

摘 要：为探究与评价客机不同倾斜姿态下人员应急撤离效果，利用信息熵理论建立客机倾斜姿态应急撤离评价模型。构建人员应急撤离评价指标，以真实撤离试验数据进行宽体客机倾斜姿态应急撤离模拟仿真试验，采用熵权-逼近理想解排序(TOPSIS)法完成应急撤离效果的评价和优劣排序。试验数据显示：倾斜姿态对撤离时间影响范围为−7.32%～+3%，俯角与仰角姿态在出口2、出口3位置的人员密度呈现明显差异，速度熵峰值集中发生在撤离初期。分析表明：速度大小熵和出口人员密度为影响倾斜姿态撤离效果的2个重要因素；俯仰5°撤离效果最佳，俯仰−10°撤离效果最差，俯仰倾斜角度为5°与10°呈现出相反的撤离效果。所建模型可实现不同倾斜姿态的评价与排序，为宽体客机设计、制造与运行的安全评估提供参考。

关 键 词：倾斜姿态；信息熵；宽体客机；应急撤离；熵权法；仿真模拟
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2026)03-0724-12

为保证客机突发事件后的人员快速撤离，中国民用航空局(Civil Aviation Administration of China, CAAC)在适航规章中明确要求[1]，对于客座量大于44座的飞机，能够使包括机组成员在内的满载量乘员在90 s内在模拟的应急情况下从机上撤离至地面。AC 25.803-1[2]中规定：考虑到飞机着火的可能性，以便在起落架放下和收起时全部人员能够在紧急着陆时迅速撤离。冲偏出跑道[3]、起落架故障等突发事件易造成机身倾斜姿态，人员在倾斜环境撤离会影响运动速度和平衡性[4]。

对于客机倾斜姿态撤离试验平台方面，在国外，美国联邦航空局(Federal Aviation Administration, FAA)的民用航空医学研究所(Civil Aerospace Medical Institute, CAMI)实验室拥有由C-124运输机机身改造的“FlexSim”客舱试验平台[5]，“FlexSim”可以使用3个连接的升降机来定位，实现交替改变的客舱姿态，可模拟非正常着陆后的客舱倾斜姿态；在国内，中国航空工业集团公司第一飞机设计研究院(以下简称航空工业一飞院)搭建了动态客舱应急撤离试验平台，通过多自由度运动平台调节不同支柱能够实现俯仰、横滚及二者组合的客机倾斜姿态[6]。针对客舱应急撤离问题的研究，学者们大多聚焦在正常姿态下机型的适航验证[7]、客舱布局[8]、应急出口[9]、人员体型变化[10]等影响因素的研究，但真实事故中会造成客舱倾斜姿态下人员应急撤离的情景。针对人员在客机倾斜姿态情景下应急撤离过程研究考虑较少，近几年学者们才开展相应的研究，封文春等[11]基于动态应急撤离模拟舱环境进行了多次不利姿态的个体与群体应急撤离试验，获取了客舱倾斜环境内个体运动与群体撤离的变化规律；Lyu等[12]探究宽体客机火灾情景下机身水平和前倾2种姿态的客舱火情发展态势与人员撤离情况，分析不同座位的安全疏散风险；程明等[13]探究了不同横滚角度下客舱火灾发展规律，得到客

收稿日期：2023-12-01；**录用日期**：2024-04-01；**网络出版时间**：2024-04-15 11:12
网络出版地址：link.cnki.net/urlid/11.2625.V.20240412.1229.001
基金项目：民航安全能力项目(ASSA2022/13)；中国民航大学研究生科研创新项目(2022YJS003)
*通信作者. E-mail：490188084@qq.com

引用格式：程明，韦柯宇. 基于信息熵的客机倾斜姿态应急撤离评价[J]. 北京航空航天大学学报，2026，52(3)：724-735.
CHENG M，WEI K Y. Emergency evacuation assessment of aircraft tilting attitude based on information entropy[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：724-735 (in Chinese).

舱等位置的可用撤离时间; 游乾乾等[14] 构建了倾斜姿态下飞机乘客疏散仿真模型, 研究倾斜姿态对于局部撤离过程的影响。但当前客机倾斜情景下应急撤离研究在倾斜姿态工况设定, 全尺寸撤离模拟及撤离效果评估尚不足, 所以有必要进行客机倾斜姿态情景下人员应急撤离相关的研究, 以往研究大多都是从宏观角度出发, 分析撤离时间、撤离人数变化等。适航规章中[1] 要求的 90 s 撤离时间作为宏观的评判准则, 不能够微观的反映撤离过程中秩序和状态的实时变化。Huang 等[15] 提出速度熵的概念作为检测聚集时人员堵塞情况的评判准则, 利用 Anylogic 的仿真试验进行验证; 魏心泉等[16] 在人员疏散模型中构建了反映运动混乱程度的疏散熵, 通过疏散熵值的变化验证引导者的正向作用。针对客机倾斜姿态应急撤离, 基于信息熵理论量化速度熵可有效反映倾斜作用影响下人员撤离过程中速度和方向的差异性及人员撤离行为特征, 从而使得评估模型和结果更加综合与准确。针对大量撤离试验提取的人员运动数据, 由于缺乏理论基础支撑, 局限于物理层面上的规律总结和分析, 速度熵可较好地处理实时的轨迹数据, 可以实现微观角度对客机应急撤离过程中撤离秩序和效果的量化和评估, 增加试验结果分析的理论深度。

为更好地评价客舱撤离效果, 本文以信息熵理论为基础, 引入速度熵的概念建立客机倾斜姿态应急撤离评价模型; 构建评价指标并提出各自的计算方法, 采用 Anylogic 软件模拟波音 787 宽体客机不同倾斜姿态下人员应急撤离过程, 获取和分析各评价指标仿真数据; 应用熵权-逼近理想解排序(technique for order preference by similarity to an ideal solution, TOPSIS)法对各指标进行评价, 对不同倾斜姿态下人员应急撤离效果进行优劣排序, 以期为宽体客机倾斜姿态下人员应急撤离的安全评估提供参考。

1　理论基础

1.1　信息熵与速度熵

20 世纪 40 年代 Shannon 借鉴了热力学的概念[17], 把信息中排除了冗余后的平均信息量称为“信息熵”; 信息熵可以说是系统有序化程度的一个度量, 系统越是有序, 信息熵就越低: 系统越是混乱, 信息熵就越高。信息熵的数学表达式为

$$H\left(\hat{X}\right)=-\sum_{i=1}^{n}p\left(x_i\right)\log_2 p\left(x_i\right) \tag{1}$$

式中: $H\left(\hat{X}\right)$为信息熵, bit; $p(x_i)$为每个随机变量$\hat{X}$可能取值x_i的概率, 其中, $0\leqslant p(x_i)\leqslant 1$。

针对大量人群聚集拥堵场景, 本文提出了一个新物理熵值的概念“速度熵”: 定义为检测人群拥堵程度的判别准则, 同时也表示人群中个体的运动速度分布和运动方向分布[9]。客舱内部是相对密集、封闭的空间环境。当发生紧急突发情况进行应急撤离时, 受外部环境[18]、人员异质性[19]、人员异常行为[20] 等因素影响会出现拥堵、混乱的撤离秩序和状态。人员实时撤离的速度及方向可以反映人员撤离过程中的移动和行为状态; 故本文引入速度熵的概念, 通过计算整体撤离过程中的速度熵值变化来量化反映客舱内人员撤离过程中秩序与状态。

速度熵E_{s}表示人员移动速度在大小和方向上的分布离散程度, 速度熵包含速度大小熵E_v和速度方向熵E_{d}这 2 部分。速度大小熵E_v和方向熵E_{d}的计算过程如下:

1) 将速度大小范围和速度方向范围分别划分成相同宽度的 P 和 Q 个区域;

2) 用$h_v(i)$表示速度值属于第i个区域的人员数量, 用$h_{\mathrm{d}}(j)$表示速度方向值属于第j个区域的人员数量; 速度值的分布概率通过 $p_v(i)=h_v(i)/S$ 计算, 速度方向值的分布概率通过 $p_{\mathrm{d}}(j)=h_{\mathrm{d}}(j)/S$ 计算, S 为系统内人员总数量, $p_v(i)$和$p_{\mathrm{d}}(j)$的值必须大于 0;

3) 速度大小熵E_v和速度方向熵E_{d}分别为

$$E_v=-\sum_{i=1}^{n}P_v(i)\log_2 p_v(i) \tag{2}$$

$$E_{\mathrm{d}}=-\sum_{j=1}^{n}p_{\mathrm{d}}(j)\log_2 p_{\mathrm{d}}(j) \tag{3}$$

4) 速度熵的计算表达式为

$$E_{\mathrm{s}}=E_vE_{\mathrm{d}} \tag{4}$$

1.2　社会力模型

社会力模型将人员看作满足牛顿运动定律的质点, 将人员考虑为受力粒子, 构建描述人员运动的微分方程[21], 包括 3 个基本力: 自驱力$\boldsymbol{f}_i^{\mathrm{o}}$、人员之间相互作用力$\boldsymbol{f}_{ij}$和与障碍物 w 的作用力$\boldsymbol{f}_{iw}$, 模型方程如下:

$$m_i\frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t}=\boldsymbol{f}_i^{\mathrm{o}}+\sum_{i\neq j}\boldsymbol{f}_{ij}+\sum_{w}\boldsymbol{f}_{iw} \tag{5}$$

式中: m_i为个体i的质量, kg; $\boldsymbol{v}_i$为个体i的移动速度, m/s。

自驱力$\boldsymbol{f}_i^{\mathrm{o}}$为人员在主观意识作用下朝着期望位置移动而施加的力。在运动过程中人员根据周围环境调整速度大小和方向, 以保证最短路径。自驱力表示为

$$\boldsymbol{f}_i^{\mathrm{o}}=\left(v_i^0(t)\boldsymbol{e}_i^0-\boldsymbol{v}_i(t)\right)/\tau_i \tag{6}$$

式中：$\boldsymbol{v}_i(t)$为t时刻人员速度；$v_i^0(t)$为t时刻的期望速度大小；$\boldsymbol{e}_i^0$为t时刻的期望方向；τ_i为反应时长。

在撤离过程中，拥挤情况下人员之间会产生心理排斥与物理力的相互作用力$\boldsymbol{f}_{ij}$，同时人员也会与物理环境中障碍物产生相互排斥力$\boldsymbol{f}_{iw}$，如下：

$$\boldsymbol{f}_{ij}=\left[A_i\exp\left(\frac{r_{ij}-d_{ij}}{B_i}\right)+ku\left(r_{ij}-d_{ij}\right)\right]\boldsymbol{n}_{ij}+ku\left(r_{ij}-d_{ij}\right)\Delta v_{ji}^t\boldsymbol{t}_{ij} \tag{7}$$

$$\boldsymbol{f}_{iw}=\left[A_i\exp\left(\frac{r_i-d_{iw}}{B_i}\right)+ku(r_i-d_{iw})\right]\boldsymbol{n}_{iw}+ku(r_i-d_{iw})(\boldsymbol{v}_i\cdot\boldsymbol{t}_{iw})\boldsymbol{t}_{iw} \tag{8}$$

式中：A_i为力的作用强度；B_i为力的作用范围，m；k为常量系数；u为常量摩擦因数；r_{ij}为2个人员身体不接触的最小距离，m；d_{ij}为人员i与人员j之间的实际距离，m；d_{iw}为人员与障碍物w实际距离，m；$\boldsymbol{n}_{ij}$为人员j指向人员i的方向向量；$\boldsymbol{n}_{iw}$为障碍物指向人员的法向方向向量；$\boldsymbol{t}_{ij}$、$\boldsymbol{t}_{iw}$分别为$\boldsymbol{n}_{ij}$、$\boldsymbol{n}_{iw}$的正交单位向量；Δv_{ji}^t为人员i、j的切向速度差，m/s。

客机倾斜姿态时人员在沿倾斜方向撤离时会接收自身重力分量带来的倾斜力$\boldsymbol{f}_{\mathrm{ax}}^{\mathrm{o}}$，人员在沿着斜坡撤离时需要克服倾斜的影响；倾斜力$\boldsymbol{f}_{\mathrm{ax}}^{\mathrm{o}}$作为倾斜平面中的一个额外力可与自驱力一起进行矢量计算，纳入到基础的社会力模型中，如下：

$$m_i\frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t}=\boldsymbol{f}_i^{\mathrm{o}}+\sum_{i\neq j}\boldsymbol{f}_{ij}+\sum_w\boldsymbol{f}_{iw}+\boldsymbol{f}_{\mathrm{ax}}^{\mathrm{o}} \tag{9}$$

$$\boldsymbol{f}_{\mathrm{ax}}^{\mathrm{o}}=m_i\boldsymbol{g}\sin(\lambda)\boldsymbol{e}_\lambda \tag{10}$$

式中：$\boldsymbol{g}$为重力加速度；λ为客机机身倾斜的角度大小；$\boldsymbol{e}_\lambda$为重力在倾斜平面的分量方向[19]。

2 客机倾斜姿态应急撤离评价模型

2.1 评价指标构建

突发事件发生后，撤离时间为客机应急撤离效率的重要评判准则。客舱内人员撤离秩序会影响撤离效率，通过式(2)～式(4)计算速度熵可量化撤离过程中人员的秩序变化。撤离过程中人员停滞次数直接影响个体的撤离效率。应急出口作为人员撤离的逃生通道，出口位置的人员密度衡量着应急出口的畅通性和安全性。本文建立的客机倾斜姿态应急撤离评价指标如图1所示。

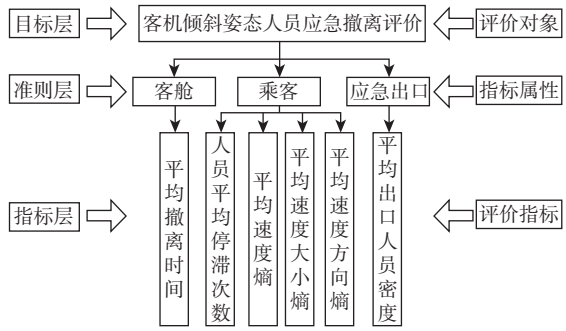


图1 客机倾斜姿态应急撤离评价指标

Fig. 1 Emergency evacuation evaluation index of aircraft tilting attitude

1）平均撤离时间A_1^α指不同倾斜姿态撤离场景客舱人员全部完成撤离所需时间的平均值，撤离时间越短代表撤离效率越好。如下：

$$A_1^\alpha=\sum_{n=1}^{N}T_{\mathrm{time}}/N \tag{11}$$

式中：α为倾斜姿态类型，水平时，$\alpha=0$，俯仰时，$\alpha=\gamma$，横滚时，$\alpha=\theta$；T_{time}为每次仿真试验的撤离时间；N为仿真模拟试验次数，$n=1,2,\cdots,N$。

2）平均出口人员密度A_2^α指一定数量可用应急撤离出口从开始撤离到撤离完成通过一定出口面积区域时间内的人员密度平均值，出口密度越小代表撤离时出口撤离越畅通。如下：

$$A_2^\alpha=\left[\sum_{n=1}^{N_1}\left(\sum_{t=1}^{T_1}P_{\mathrm{density}(n)}\Big/t_{\mathrm{passtime}}\right)\right]\Big/N_{\mathrm{exit}} \tag{12}$$

$$P_{\mathrm{density}(n)}=N_{\mathrm{people}}/S_{\mathrm{area}} \tag{13}$$

式中：$P_{\mathrm{density}(n)}$为出口区域n的人员密度值；N_{people}为区域内人员的数量；S_{area}为设定区域的面积，本文试验将其设定为以应急出口宽度为长的正方形区域；$T_1=t_{\mathrm{passtime}}$为人员通过出口区域$n$的时长，$N_1=N_{\mathrm{exit}}$为可用出口数量，$t=1,2,\cdots,T_1$，$n=1,2,\cdots,N_1$。

3）平均速度熵A_3^α指该撤离场景下人员从开始撤离到撤离完成之间每一时刻速度大小熵与速度方向熵乘积的平均值，熵值越大表明撤离行为越有序。如下：

$$A_3^\alpha=\sum_{t=1}^{T}E_v^tE_{\mathrm{d}}^t/T_{\mathrm{time}} \tag{14}$$

4）平均速度大小熵A_4^α指该撤离场景下人员从开始撤离到撤离完成之间每一时刻速度大小熵的平均值，熵值越小表明人员撤离速度越离散。如下：

$$A_4^\alpha=\frac{\sum_{t=1}^{T}E_v^t}{T_{\mathrm{time}}}=\frac{\sum_{t=1}^{T}\left(-\sum_{i=1}^{n_1}P_v(i)\log_2p_v(i)\right)^t}{T_{\mathrm{time}}} \tag{15}$$

5）平均速度方向熵A_5^α是指该撤离场景下人员从开始撤离到撤离完成之间每一时刻速度方向熵的平均值，熵值越大表明人员撤离方向越统一。如下：

$$A_5^\alpha=\frac{\sum_{t=1}^{T}E_{\mathrm{d}}^t}{T_{\mathrm{time}}}=\frac{\sum_{t=1}^{T}\left(-\sum_{i=1}^{n_2}P_{\mathrm{d}}(i)\log_2p_{\mathrm{d}}(i)\right)^t}{T_{\mathrm{time}}} \tag{16}$$

式中：E_v^t和E_{d}^t分别为t时刻的速度大小熵和速度方

向熵; $T=T_{\text{time}}$为该次仿真的撤离时间, $t=1,2,\cdots,T$。

5) 人员平均停滞次数A_6^{α}指在撤离过程中人员移动停滞总数量与总撤离人数的比值, 比值越大表明撤离过程中堵塞情况越多。如下:

$$A_6^{\alpha}=\text{SUM}\{(x_t-x_{t-1})<\pm D\cap(y_t-y_{t-1})<\pm D\}/N_{\text{headcount}} \tag{17}$$

式中: x_t、y_t为人员在t时刻的x、y位置坐标; D为判定人员停滞的位移偏量值, 本文设D为 0.1 m; $N_{\text{headcount}}$为客舱内总人数。

2.2 基于熵权法的评价指标权重确定

熵权法通过计算评价指标的信息熵值来判断评价指标的离散程度, 进而确定指标权重; 熵权法的具体步骤如下。

1) 数据正向化处理。对于所有指标在综合计算之前把各项指标进行统一化处理, 将负指标通过计算其倒数的形式转变为正指标, 因此, 可以得到由n个评价对象, m个评价指标构成的正向指标集合的评价指标矩阵$\boldsymbol{X}$:

$$\boldsymbol{X}=\begin{bmatrix}X_{11} & \cdots & X_{1m}\\ \vdots & & \vdots\\ X_{n1} & \cdots & X_{nm}\end{bmatrix} \tag{18}$$

2) 矩阵标准化处理。为消除数据间量纲和数量级的差异, 在赋权之前进行数据标准化处理, 使各项指标同质化。同时, 避免当其标准化矩阵中的数值为 0 和 1 时, 所确定熵值均为 0 的情况, 对原标准化矩阵中的值加上某一小数值, 使所确定的熵值发生微小变化进行改进[22]。

将矩阵标准化后记为$\boldsymbol{Z}$, 如下:

$$\boldsymbol{Z}=\begin{bmatrix}Z_{11} & \cdots & Z_{1m}\\ \vdots & & \vdots\\ Z_{n1} & \cdots & Z_{nm}\end{bmatrix} \tag{19}$$

$$Z_{nm}=\frac{X_{nm}-\min(X_{nm})}{\max(X_{nm})-\min(X_{nm})} \tag{20}$$

式中: m 为正指标。

3) 矩阵归一化处理。对标准化后的指标 Z_{mn}进行归一化处理, 计算式为

$$B_{nm}=Z_{nm}\Big/\sum_{n=1}^{N}Z_{nm} \tag{21}$$

4) 计算第m项指标的信息熵值C_m和差异性系数d_m, 其表达式分别为

$$C_m=-\left(\frac{1}{\ln m}\right)\sum_{n=1}^{N}B_{nm}\ln B_{nm} \tag{22}$$

$$d_m=1-C_m \tag{23}$$

d_m的值越大表示指标Z_m在综合评价中影响程度越大。

5) 计算指标Z_m的权重W_m, 计算式为

$$W_m=d_m\Big/\sum_{m=1}^{M}d_m \tag{24}$$

6) 计算评价值R_n, 计算式为

$$R_n=\sum_{m=1}^{M}W_mZ_{nm} \tag{25}$$

式中: R_n为第n个评价对象的评价值。由于Z_{mn}均为正指标, 所以R_n的值越大表示对应的评价对象越优。

2.3 熵权-TOPSIS 评价模型建立

TOPSIS 法是一种多属性的决策分析方法, 核心思想是构建评价对象中的理想解, 即评价指标值中最优值或最差值; 将通过熵权法计算的客观评价指标权重带入计算, 具体过程如下。

1) 构建加权标准评价矩阵。将正向标准化矩阵$\boldsymbol{Z}$与熵权法得到的指标权重相乘, 得到加权评价矩阵$\boldsymbol{U}$如下:

$$\boldsymbol{U}=\begin{bmatrix}U_{11} & \cdots & U_{1m}\\ \vdots & & \vdots\\ U_{n1} & \cdots & U_{nm}\end{bmatrix} \tag{26}$$

2) 确定正理想解U^{+}和负理想解U^{-}。正理想解U^{+}和负理想解U^{-}分别为标准评价矩阵指标中对应的最大值和最小值的集合, 分别表示为

$$U^{+}=\left\{U_1^{+},U_2^{+},\cdots,U_m^{+}\right\} \tag{27}$$

$$U^{-}=\left\{U_1^{-},U_2^{-},\cdots,U_m^{-}\right\} \tag{28}$$

3) 计算各评价对象与理想解的欧氏距离, 计算式分别为

$$D_n^{+}=\sqrt{\sum_{m=1}^{M}\left(U_m^{+}-U_{nm}\right)^2} \tag{29}$$

$$D_n^{-}=\sqrt{\sum_{m=1}^{M}\left(U_m^{-}-U_{nm}\right)^2} \tag{30}$$

4) 计算相对贴近度T_n, 计算式为

$$T_n=D_n^{-}/(D_n^{+}+D_n^{-}) \tag{31}$$

5) 对相对贴近度T_n进行归一化处理, 计算式为

$$P_n=T_n\Big/\sum_{n=1}^{N}T_n \tag{32}$$

依据归一化后的P_n, 可对各评价对象进行排序, P_n的数值越大表明该评价对象越优。

2.4 撤离仿真试验模型搭建

采用仿真建模法建立撤离仿真试验模型, 基于基础社会力模型理论设计的行人库, 搭建人员撤离流程。可以真实模拟不同环境条件下人员应急撤

离过程。仿真试验假设条件如下：①仿真模型中只考虑倾斜姿态对人员撤离速度的影响；②根据适航规章演示准则要求仿真试验中开启50%的应急出口[1]；③仿真试验中人员从最近的可用应急出口进行撤离。具体仿真试验模型参数设定及仿真场景如下。

1）以波音787宽体客机为仿真对象搭建客舱仿真物理模型，采用其公务舱与经济舱混合的客舱布局；将客舱全尺寸布局平面图（CAD图）导入Anylogic软件，在软件中进行客舱座椅、应急出口等设施的绘制。

2）以真实的局部客舱人员应急撤离试验为依据，对于客舱仿真模型中个体速度和尺寸参数进行系数调整校正。人员在客舱环境中正常撤离速度参照文献[23]设定，以航空工业一飞院进行的撤离试验数据为支撑，人员在不同倾斜角度下撤离速度衰减率如图2所示。依据《中国成年人人体尺寸》中的统计数据[24]，同时考虑近年人员体型变化趋势[25]，根据研究中拟合身高与肩宽的线性关系[26]，设定不同类型人员体型参数。人员类型组成按照适航规章中撤离演示准则比例要求设置[1]，人员撤离延迟时间设置为[1.0, 3.2] s[27]。

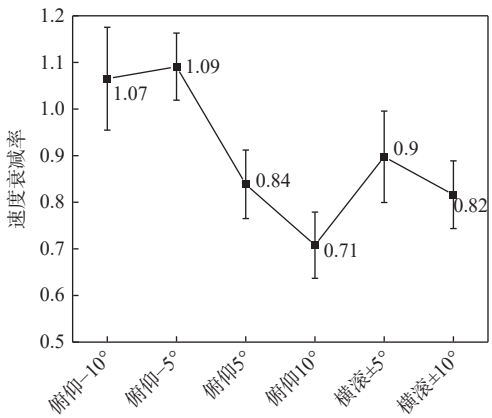


图2 不同倾斜姿态下群体撤离速度衰减率

Fig. 2 Attenuation rate of group withdrawal velocity under oblique posture

3）客机倾斜姿态与应急出口开启示意图如图3和图4所示，参照动态客舱倾斜姿态人员撤离试验[11]进行仿真场景设定如表1所示，每组进行50次随机仿真试验。

2.5 评价方法模型

结合本文第2.1节中构建的客机倾斜姿态应急撤离评价指标，将本文的熵权-TOPSIS法应用到客机不同倾斜姿态下人员应急撤离效率评价过程，详细的评价流程如图5所示。

通过熵权-TOPSIS法对客机不同倾斜姿态应急撤离的相对贴近度计算，根据相对贴近度数值对不同倾斜姿态应急撤离情景由高到低进行排序，表示着应急撤离效果由优到劣的排序结果。

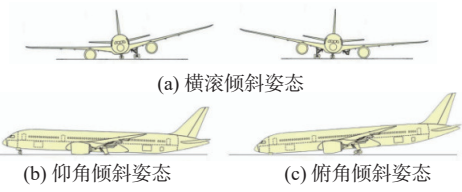


图3 客机倾斜姿态示意图

Fig. 3 Aircraft tilt attitude diagram

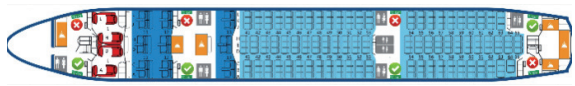


图4 仿真试验应急出口开启示意图

Fig. 4 Simulation test emergency exit opening diagram

表1 客舱倾斜姿态下应急撤离试验场景

Table 1 Emergency evacuation test scenario in cabin tilting attitude

试验序号	撤离场景	出口数量	仿真次数
1	水平0°	4	50
2	俯仰−5°	4	50
3	俯仰5°	4	50
4	俯仰−10°	4	50
5	俯仰10°	4	50
6	横滚±5°	4	50
7	横滚±10°	4	50

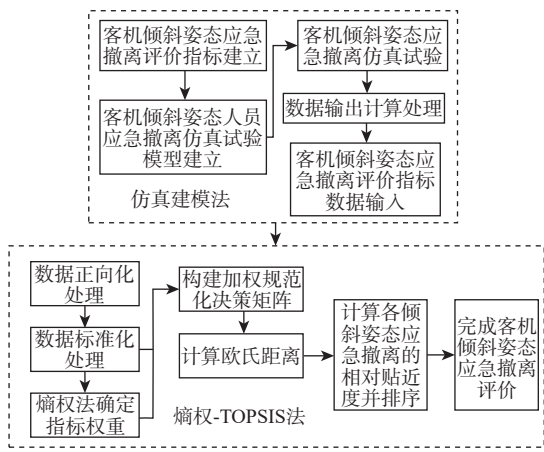


图5 客机倾斜姿态应急撤离评价流程

Fig. 5 Emergency evacuation evaluation process of aircraft tilting attitude

3 结果分析与讨论

3.1 仿真结果

3.1.1 撤离效率

1）撤离时间

不同倾斜姿态下仿真试验的平均撤离时间如

图 6 所示, 以水平姿态平均撤离时间为基准来衡量倾斜姿态对客舱撤离效率的影响如图 7 所示。俯仰−5°撤离时间最短为 83.5 s, 撤离效率最佳为+3%, 横滚±10°撤离时间最长为 92.4 s, 超过规章[1] 中 90 s 撤离时间要求, 撤离效率最差为−7.32%。俯仰±5°与俯仰±10°呈现相反的撤离效果; 相比俯仰姿态, 横滚姿态下对撤离效率都为负向影响, 且两者姿态角度与撤离效率都呈现显著负相关; 同角度下横滚姿态对于撤离效率的影响更显著。

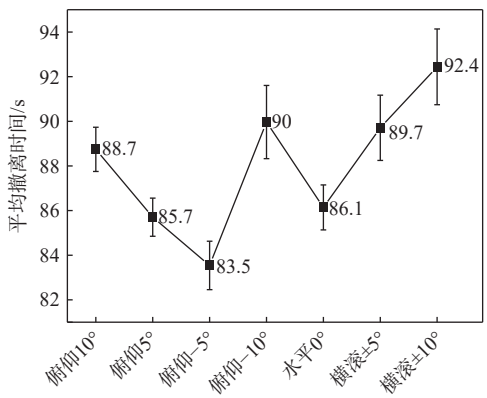


图 6　不同倾斜姿态下平均撤离时间

Fig. 6　Average evacuation time under inclined attitude

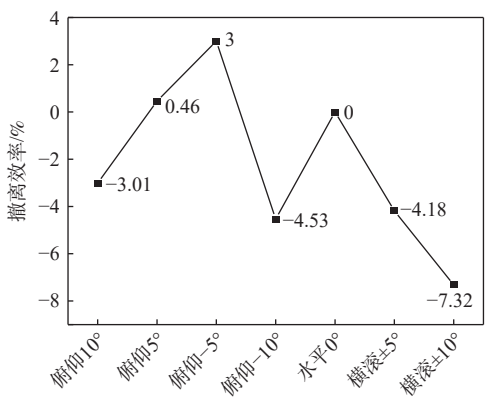


图 7　不同倾斜姿态下人员撤离效率

Fig. 7　Evacuation efficiency of different tilting attitude

2) 人员停滞次数

如图 8 所示, 俯仰 5°停滞次数最少为 957 次, 横滚±10°最多为 1 257 次; 除俯仰 5°, 倾斜角度下平均撤离时间与人员停滞次数的变化趋势相似, 可表明撤离过程中人员停滞次数与撤离时间呈正相关; 俯仰±5°人员停滞次数与撤离时间则呈现显著负相关, 仰角时客舱中后段的人员可借助倾斜力加快撤离速度, 俯角时人员则需克服倾斜力导致移动停滞次数增多; 并且倾斜 10°时的停滞次数明显多于 5°, 反映出倾斜角度越大, 人员在撤离过程中需调整的次数可能就越多, 同时, 人员堵塞的情况越严重。

3.1.2　出口人员密度

应急出口的畅通性直接影响客舱撤离效率及

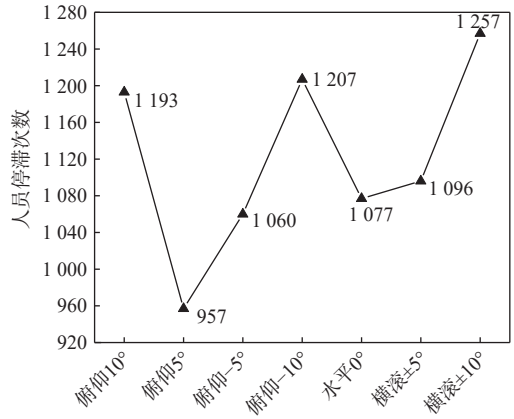


图 8　不同倾斜姿态下人员停滞次数

Fig. 8　Stagnation times of personnel under inclined attitudes

出口位置撤离人员的安全性; 同时, 出口人员密度峰值意味着出口位置的拥堵时刻。

1) 出口畅通性

如图 9 所示, 横滚姿态下出口 2、出口 3 最终的通过时间相比水平姿态有所增长; 横滚±10°时出口 4 最终的通过时间最长, 接近 20 s, 出口 1 的密度峰值出现在撤离前期, 出口 2、出口 3 的密度变化趋势相比水平姿态更加平缓。对比表明: 横滚倾斜姿态对于人员撤离有明显影响, 在撤离前期会加剧出口位置的人员聚集, 并增加出口的撤离通过时间。

俯仰±5°出口 1 的密度变化较水平姿态更平滑, 俯仰姿态下其余出口人员密度在前期的变化也相对平缓, 在撤离初期人员需适应倾斜环境从而导致出口人员通过缓慢。俯角与仰角姿态在出口 2、出口 3 的密度趋势呈现差异性, 俯角姿态下出口 3 为下坡, 出口 2 则为上坡, 俯角姿态下人员可能倾向于借助倾斜力的作用通过下坡出口进行逃生, 导致下坡出口位置的人员聚集。

2) 出口安全性

如表 2 所示, 在出口 1, 水平 0°出口人员密度峰值最高, 俯仰±10°的出口密度峰值明显高于俯仰±5°, 横滚角度下差异则较小; 出口密度峰值时刻发生在不同的撤离时期。在出口 2, 俯角姿态出口平均人员密度明显高于仰角姿态; 俯角姿态下密度峰值均大于其余姿态, 密度峰值时刻都较集中在 20 s。在出口 3, 水平姿态出口平均人员密度均高于倾斜姿态, 仰角姿态下出口密度峰值高, 同时峰值次数多。在出口 4, 水平姿态出口平均人员密度依旧高于其他姿态; 俯仰−5°时出口密度峰值最高, 俯仰−10°时密度峰值时刻最晚, 分析为该姿态下出口 4 为下坡, 部分人员在加大倾斜角度时为确保安全而减缓速度造成的。

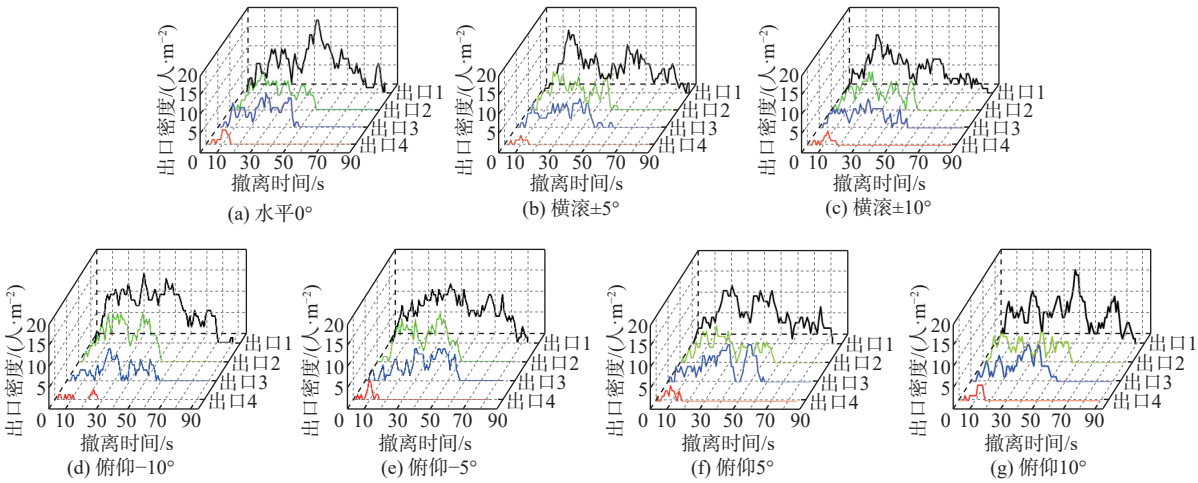


图 9 不同倾斜姿态下应急出口人员密度变化

Fig. 9 Change of emergency exit personnel density under different posture

表 2 不同倾斜姿态下出口人员密度数值统计

Table 2 Numerical statistics of exit personnel density under different posture

出口	平均密度/(人·m⁻²)							密度峰值/(人·m⁻²)						
	水平0°	横滚±5°	横滚±10°	俯仰−10°	俯仰−5°	俯仰5°	俯仰10°	水平0°	横滚±5°	横滚±10°	俯仰−10°	俯仰−5°	俯仰5°	俯仰10°
1	7.619	6.65	6.206	8.161	7.758	6.955	7.278	18.784	16.279	15.027	16.279	13.775	13.775	17.532
2	5.249	4.985	4.867	6.926	5.995	3.962	4.206	10.018	10.018	10.018	11.27	11.27	8.766	7.514
3	4.313	3.757	3.615	2.778	3.85	4.064	4.125	8.766	7.514	7.514	7.514	7.514	8.766	8.766
4	2.049	1.138	1.503	0.601	1.445	1.586	1.83	3.757	2.505	3.757	2.505	5.009	3.757	3.757

出口	峰值时刻/s						
	水平0°	横滚±5°	横滚±10°	俯仰−10°	俯仰−5°	俯仰5°	俯仰10°
1	45、46	17	21、22	34	37	24	51
2	20	15、17	24、26	19、22、25、27、41	38	23	12、37、38
3	29	38	31	27、28	30、31、44、45、47、48	49、50、51	36、37、43
4	10、11、12	10	13	25	10	9	12、13、14、15

3.1.3 速度熵

1）速度熵变化

如图 10 所示，水平和横滚姿态的速度方向熵在撤离前中期平缓变化后发生显著波动，横滚姿态波动程度更严重；速度大小熵在前期发生大幅度波动后趋向平缓，分析为人员刚收到撤离指令后会慌乱撤离，过后撤离速度趋向稳定；水平姿态、横滚±5°、横滚±10°分别在 t=35～55 s、t=65～72 s，t=55～65 s 速度大小熵成“V”形显著变化，是由于出口位置人员堵塞情况造成的。速度熵与速度大小熵的整体波动趋势相似，在撤离前期横滚姿态的速度熵整体水平低于水平姿态，反映出横滚姿态下人员撤离相对有秩序性；到撤离后期横滚±10°速度熵波动显著，可见随着倾斜角度的增加，客舱撤离秩序受影响越明显。

俯仰−5°和−10°的速度方向熵波动变化差异最明显；俯仰±5°速度方向熵较先发生下降趋势，可反映随俯仰角度增加，人员撤离方向秩序受影响越严重。在撤离初期俯仰±10°速度大小熵的波动变化最显著，速度大小熵值水平反映俯仰±5°时人员撤离速度离散程度高；仰角姿态较早出现下降趋势，分析是在后期位于客舱后段的人员受俯角姿态影响造成了撤离速度的波动。速度熵变化趋势较平缓，俯仰±10°的速度熵值水平最低，到撤离后期俯角比仰角姿态波动明显；可表明俯仰±5°的撤离秩序更加混乱，俯角姿态对撤离秩序的影响更显著。

2）速度熵极值

如表 3 所示，水平姿态速度熵峰值最高为 0.786 3 bit，速度熵峰值时刻集中在撤离初期，人员盲目且急于撤离造成了速度及方向的无序性；谷值则出现在撤离后期，倾斜姿态的谷值明显低于水平姿态，横滚±10°速度熵最低为 0.359 6 bit，可见倾斜姿态一定程度上能够稳定客舱撤离秩序。速度大小熵峰值时刻呈现两端性，俯仰10°的峰值最低为 0.718 7 bit；谷值时刻集中在撤离中后期，反映在后期人员撤离速度离散度较低。

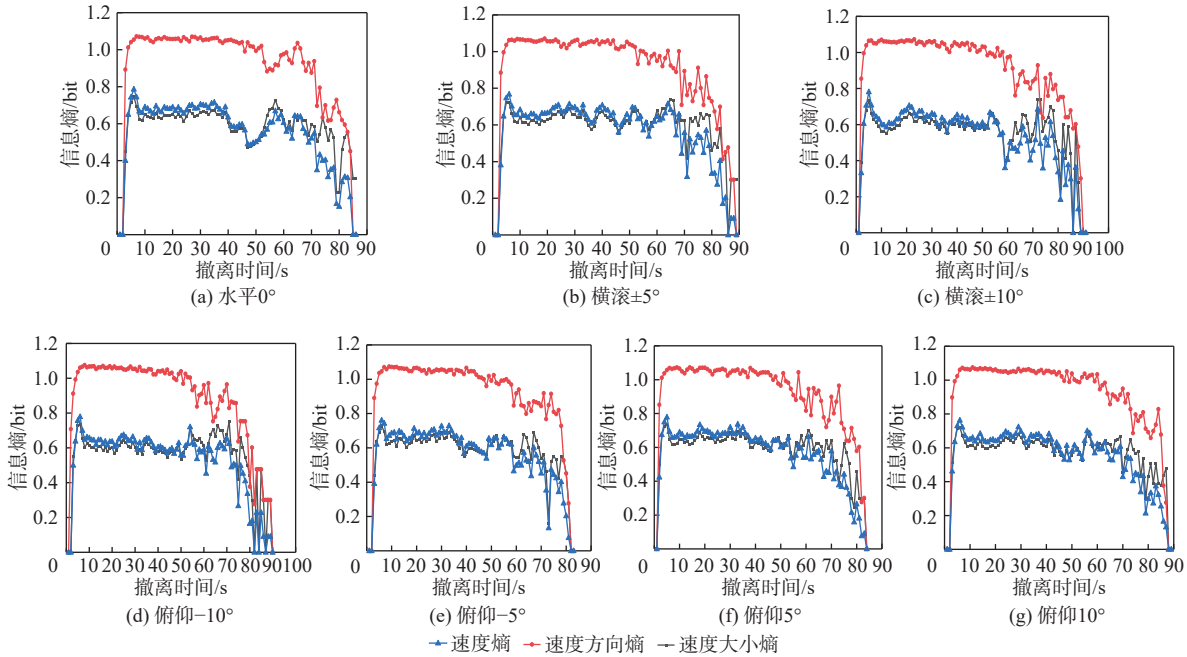


图 10　不同倾斜姿态下信息熵变化

Fig. 10　Information entropy changes under different attitudes

速度方向熵峰值差别不明显, 水平姿态的谷值明显高于倾斜姿态, 可表明在撤离后期倾斜姿态对人员移动方向秩序有正向影响。表 3 中, 括号内为峰/谷值出现的时刻。

表 3　不同倾斜姿态下信息熵极值统计

Table 3　Statistics of extremum information entropy under different attitude

工况	速度熵峰/谷值/bit	速度大小熵峰/谷值/bit	速度方向熵峰/谷值/bit
水平0°	0.786 3(6 s)/0.485 5(49 s)	0.746 2(6 s)/0.471 8(47 s)	1.071 7(7 s)/0.875 6(70 s)
横滚±5°	0.767 3(6 s)/0.443 6(69 s)	0.739 2(65 s)/0.552 6(46 s)	1.070 1(19 s)/0.709 7(70 s)
横滚±10°	0.780 8(5 s)/0.359 6(59 s)	0.740 3(73 s)/0.397 9(59 s)	1.072 2(23 s)/0.761 3(63 s)
俯仰−10°	0.778 7(6 s)/0.453 0(61 s)	0.753 4(71 s)/0.528 3(61 s)	1.075 6(8 s)/0.752 6(64 s)
俯仰−5°	0.761(6 s)/0.436 3(66 s)	0.728 2(6 s)/0.535 0(66 s)	1.073 8(9 s)/0.798 4(64 s)
俯仰5°	0.778 4(5 s)/0.427 0(67 s)	0.736 9(5 s)/0.533 1(55 s)	1.071 9(9 s)/0.719 7(68 s)
俯仰10°	0.762 2(6 s)/0.429 1(70 s)	0.718 7(6 s)/0.452 2(70 s)	1.074 3(11 s)/0.857 6(65 s)

注: 为准确反映撤离过程的秩序性, 熵谷值选取T_{time}=15～70 s区间的数值。

3.2　评价结果

3.2.1　指标权重确定

仿真试验输出的相关数据通过式(11)～式(17)的计算处理得到表 4 的倾斜姿态应急撤离评价指标数据, 以水平、横滚、俯仰 3 种姿态共 7 种倾斜角度工况为评价对象进行评价排序。

根据第 3.2 节中熵权法确定权重的计算步骤对数据进行正向化、标准化处理, 可以得到客机倾斜姿态应急撤离评价规范化决策矩阵; 通过式(22)计算可得各评价指标的信息熵值C_m:

$$C_m = \{0.861\,1, 0.805\,7, 0.869\,4, 079\,46, 0.914\,0, 0.825\,4\}$$

通过式(24)进一步求出各评价指标的客观评价权重W_m:

$$W_m = \{0.170\,9, 0.239\,0, 0.160\,6, 0.252\,7, 0.105\,8, 0.214\,7\}$$

评价指标的信息熵值越小, 则该评价指标数据的离散程度越大, 对评价结果的影响程度就越大, 其权重就越大。平均出口人员密度和平均速度大小熵是影响客舱倾斜姿态应急撤离效果的 2 个重要因素。

出口人员密度的差异性是由出口位置、倾斜环境、通过时间共同造成的, 人员集中在客舱中后段位置导致就近的出口需承担更严峻的撤离压力。同时, 在客舱倾斜姿态的影响下, 各应急出口的通过状态与通过时间也发生着显著变化。由速度熵变化可见, 速度大小熵较大程度上决定着速度熵的波动变化; 倾斜姿态影响着人员撤离的速度与状态, 撤离过程中平均速度大小熵与人员停滞次数呈现负相关, 可更直观的反映出客舱人员的撤离状态。

表4 客机倾斜姿态应急撤离评价指标数据

Table 4 Emergency evacuation evaluation index data of passenger aircraft tilting attitude

序号	工况	平均撤离时间/s	平均出口人员密度/(人·m^{-2})	平均速度熵/bit	平均速度大小熵/bit	平均速度方向熵/bit	人员平均停滞次数
1	水平0°	86.142	4.807 689	0.552 769	0.577 678	0.912 367	4.360 324
2	横滚±5°	89.708	4.132 413	0.566 561	0.591 762	0.909 632	4.437 247
3	横滚±10°	92.440	4.047 715	0.531 763	0.552 871	0.911 086	5.089 069
4	俯仰−10°	89.968	4.616 464	0.524 925	0.553 193	0.888 161	4.886 640
5	俯仰−5°	83.542	4.761 829	0.559 137	0.572 963	0.916 227	4.291 498
6	俯仰5°	85.704	4.141 955	0.566 023	0.586 503	0.922 240	3.874 494
7	俯仰10°	88.744	4.359 759	0.549 797	0.564 031	0.918 651	4.829 960

3.2.2 熵权-TOPSIS法评价

根据TOPSIS法的计算过程，将熵权法计算得到的客观评价指标权重代入计算得到加权规范化决策矩阵，通过式(29)和式(30)计算得到各倾斜姿态距离正、负理想解的欧氏距离；在得到距离测度之后，按照式(31)和式(32)可计算出各倾斜姿态的相对贴近度及其归一化结果，如表5所示。

表5 客机倾斜姿态应急撤离评价计算结果

Table 5 Calculation results of emergency evacuation evaluation of aircraft tilting attitude

工况	与正理想解的欧氏距离	与负理想解的欧氏距离	相对贴近度T_n	归一化相对贴近度P_n	熵权法评价值R_n
水平0°	0.303 73	0.288 33	0.479 970	0.136 87	0.575 63
横滚±5°	0.208 38	0.400 00	0.689 481	0.196 61	0.837 60
横滚±10°	0.452 19	0.250 80	0.386 593	0.110 24	0.336 72
俯仰−10°	0.469 07	0.079 62	0.147 345	0.042 02	0.127 40
俯仰−5°	0.287 75	0.322 20	0.519 149	0.148 04	0.660 21
俯仰5°	0.066 03	0.483 81	0.867 850	0.247 47	1.028 44
俯仰10°	0.350 24	0.218 41	0.416 490	0.118 76	0.497 61

根据式(25)计算得出熵权法的综合评价值如表5所示，由于Z_{nm}均为正指标，所以R_n的值越大代表评价对象越优；采用熵权-TOPSIS法依据相对贴近度的数值完成对客机倾斜姿态应急撤离效果的评价，数值越大表示该姿态应急撤离效果越高效；得到同样的评价结果为：俯仰5°>横滚±5°>俯仰−5°>水平0°>俯仰10°>横滚±10°>俯仰−10°。通过排序结果可知倾斜角度为5°时撤离效果都优于水平姿态，当倾斜角度增加到10°后撤离效果都劣于水平姿态；其中，俯仰5°撤离效果最佳，俯仰−10°撤离效果最差。

根据评价结果，结合实际事故情景进行分析。起落架故障情景：在主落架故障时，仰角姿态小于5°，该情景下撤离效果会有所提高；在前起落架故障时，俯角姿态可能会大于5°，撤离效果会受到一定的影响；当前起落架和单侧主落架故障时，在横滚姿态可能会达到10°，对整体的撤离效果会造成显著影响。冲偏出跑道情景：冲偏出跑道事件通常也是导致起落架损坏从而造成客舱倾斜姿态的，但是对于特殊地形的高原机场可能会由于机场地形等因素造成极端的倾斜角度，通过评价结果可知倾斜角度越大，撤离效果越差，所以在极端角度撤离情况发生时人员撤离效果会更糟糕。

根据上述分析提出相应的对策和建议。在密闭客舱空间内发生倾斜姿态时，乘客可能会因倾斜带来的危机感和不平衡感产生恐慌、不安等负面情绪，所以建议机组人员在开始撤离前通过广播等方式告知乘客撤离情景及相关的注意事项，平复部分乘客的恐慌情绪，降低倾斜姿态带来的恐慌气氛，有利于撤离过程中的引导工作；其次，建议机组人员在客舱倾斜撤离过程中充分发挥引导作用，根据客舱倾斜环境实时合理引导上坡位置的人员进行出口分散撤离，提升下坡应急出口的利用效率，减缓拥挤出口的人员密度同时降低人员堵塞情况，从而达到最佳的撤离效果。

3.2.3 评价结果比较

为进一步验证本文模型的科学性与有效性，同撤离时间和平均速度熵排序结果进行对比，如表6所示。倾斜姿态平均速度熵排序呈现相反顺序，可知撤离秩序与撤离效果呈现负相关，速度熵评价指标具有代表性；对比以撤离时间为评判标准，将出口人员密度、人员停滞次数等指标纳入并结合熵权-TOPSIS评价方法后的评价结果，能更准确的评

表6 评价排序对比结果

Table 6 Evaluation ranking and comparison results

工况	撤离时间排序	平均速度熵排序	熵权-TOPSIS排序
水平0°	3	4	4
横滚±5°	5	7	2
横滚±10°	7	2	6
俯仰−10°	6	1	7
俯仰−5°	2	5	3
俯仰5°	1	6	1
俯仰10°	4	3	5

注：数值1～7表示不同倾斜姿态工况下评估对比排名结果。

判客机不同倾斜姿态下人员应急撤离效果的差异性。

3.2.4　局限性分析

1) 针对客机不同突发情景撤离效果评价, 熵权-TOPSIS 法所用的欧氏距离存在判断失效的风险, 当计算结果相同时无法判断评价对象的优劣, 并且仅能反映评价对象与正、负理想解在欧氏距离上的贴近程度, 侧重于评价对象排序的关系。

2) 以速度熵为评价指标的撤离评价模型不适用于单一出口和试验人数较少的仿真撤离情景, 在单一撤离出口的情景速度方向熵指标的变化差异会不明显; 观察速度熵变化图可见撤离后期剩余人数较少时速度熵值波动显著, 可表明试验人数越少时整体变化受个体影响的程度越显著; 因此, 在上述撤离情景以速度熵指标来分析和评估撤离过程会不理想。

3) 对于真人的撤离模拟试验, 计算速度熵需提取人员的运动轨迹数据, 试验平台则需配备精准的定位设备进行数据收集, 对于搭建的简易试验平台实现会比较困难; 但随着视频图像识别技术的不断提升, 以速度熵值来量化真人模拟试验的撤离秩序将变得更便捷和精确。

4　结　论

本文根据客机倾斜姿态应急撤离效果评价模型, 对客机不同倾斜姿态应急撤离过程进行仿真分析与评价研究, 结论如下:

1) 倾斜姿态对撤离时间影响范围为−7.32%～+3%, 横滚±10°撤离时间最长为 92.4 s; 撤离时间与人员停滞次数呈正相关。

2) 速度熵、出口人员密度等评价指标的仿真结果可较好反映应急撤离过程与效果的差异。

3) 与撤离时间评估准则相比, 建立的客机倾斜姿态应急撤离效果评价模型可实现更精确的综合评价; 评价结果为俯仰 5°>横滚±5°>俯仰−5° >水平 0°>俯仰 10°>横滚±10°>俯仰−10°。

本文只针对客机俯仰、横滚 2 种倾斜姿态下人员应急撤离进行了仿真试验与撤离效果评估研究, 进一步应当考虑客机倾斜环境下人员恐慌情绪、通道障碍物及人员跌倒等影响因素的研究, 同时对于真实事故中出现的组合倾斜姿态下应急撤离情景需进行考虑与研究。

参考文献（References）

[1] 中国民用航空局. 大型飞机公共航空运输承运人运行合格审定规则[EB/OL]. (2024-04-13)[2025-11-19]. https://xxgk.mot.gov.cn/2020/gz/202404/t20240416_4128400.html.
Civil Aviation Administration of China. Rules for the operation qualification examination of large aircraft public air transport carriers[EB/OL]. (2024-04-13)[2025-11-19]. https://xxgk.mot.gov.cn/2020/gz/202404/t20240416_4128400.html(in Chinese).

[2] Federal Aviation Administration . AC 25.803-1 : emergency evacuation demonstrations[EB/OL]. (1989-11-13)[2023-11-22]. https://rosap.ntl.bts.gov/view/dot/71262.

[3] 戴湘龄. 面向机场冲/偏出跑道事件的应急协同决策仿真研究[D]. 南京: 南京航空航天大学, 2017.
DAI X L. Simulation study on emergency collaborative decision-making for airport runway rushing/deviating events[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017(in Chinese).

[4] FANG S M, LIU Z J, WANG X J, et al. Simulation of evacuation in an inclined passenger vessel based on an improved social force model[J]. Safety Science, 2022, 148: 105675.

[5] MCLEAN G A. Access-to-egress: a meta-analysis of the factors that control emergency evacuation through the transport airplane type-III overwing exit: DOT/FAA/AM-01/2 [R]. Washton, D. C. : Federal Aviation Administration, 2001.

[6] 封文春, 李伟. 民用飞机应急撤离影响因素试验研究[J]. 航空科学技术, 2022, 33(3): 11-16.
FENG W C, LI W. Test research on the influencing factors of civil aircraft emergency evacuation[J]. Aeronautical Science & Technology, 2022, 33(3): 11-16(in Chinese).

[7] 张青松, 杨彩红. 飞机客舱应急撤离演示模拟方案设计研究[J]. 安全与环境学报, 2016, 16(3): 154-157.
ZHANG Q S, YANG C H. Simulation design for the aircraft cabin emergency evacuation[J]. Journal of Safety and Environment, 2016, 16(3): 154-157(in Chinese).

[8] 梁文娟, 张建涛. 翼身融合客机应急撤离建模与仿真方法研究[J]. 安全与环境学报, 2023, 23(5): 1717-1724.
LIANG W J, ZHANG J T. Research on modeling and simulation methods for emergency evacuation of blended-wing-body civil aircraft[J]. Journal of Safety and Environment, 2023, 23(5): 1717-1724 (in Chinese).

[9] 田水承, 王启睿, 陈洋, 等. 民机客舱中部应急出口对人员疏散的影响[J]. 科学技术与工程, 2020, 20(20): 8413-8417.
TIAN S C, WANG Q R, CHEN Y, et al. The impact of the central emergency exit of civil aircraft cabin on personnel evacuation[J]. Science Technology and Engineering, 2020, 20(20): 8413-8417(in Chinese).

[10] MELIS D J, SILVA J M, YEUN R, et al. The effect of airline passenger anthropometry on aircraft emergency evacuations[J]. Safety Science, 2020, 128: 104749.

[11] 封文春, 张桐, 李伟. 民用飞机迫降后不利姿态应急撤离实验研究[J]. 北京航空航天大学学报, 2023, 49(7): 1553-1562.
FENG W C, ZHANG T, LI W. Experimental study on adverse attitude emergency evacuation of civil aircraft after crash landing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1553-1562(in Chinese).

[12] LYU W, XING L L, LI J W, et al. Evaluating personnel evacuation risks under fire scenario of Airbus wide-body aircraft: a simulation study[J]. Frontiers in Public Health, 2022, 10: 994031.

[13] 程明, 乔嘉鑫. 横滚姿态下飞机客舱火灾数值模拟研究[J]. 中国

安全生产科学技术, 2024, 20(2): 195-201.
CHENG M, QIAO J X. Numerical simulation of aircraft cabin fire in rolling attitude[J]. Journal of Safety Science and Technology, 2024, 20(2): 195-201(in Chinese).

[14] 游乾乾, 陈琨, 寇志远, 等. 飞机倾斜对乘员应急撤离行为的影响研究[J]. 航空工程进展, 2025, 16(3): 90-98.
YOU Q Q, CHEN K, KOU Z Y, et al. Research on the impact of airplane inclination on passengers emergency evacuation behavior [J]. Advances in Aeronautical Science and Engineering, 2025, 16(3): 90-98(in Chinese).

[15] HUANG L D, CHEN T, WANG Y, et al. Congestion detection of pedestrians using the velocity entropy: a case study of Love Parade 2010 disaster[J]. Physica A: Statistical Mechanics and Its Applications, 2015, 440: 200-209.

[16] 魏心泉, 王坚. 基于熵的火灾场景介观人群疏散模型[J]. 系统工程理论与实践, 2015, 35(10): 2473-2483.
WEI X Q, WANG J. A mesoscopic evacuation model based on multi-agent and entropy with leading behavior under fire conditions[J]. Systems Engineering-Theory & Practice, 2015, 35(10): 2473-2483(in Chinese).

[17] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.

[18] 杜红兵, 张世博, 李弘毅, 等. 火灾场景下民机客舱人员疏散仿真及优化[J]. 中国安全生产科学技术, 2023, 19(2): 187-194.
DU H B, ZHANG S B, LI H Y, et al. Simulation and optimization of personnel evacuation in civil aircraft cabin under fire scenario[J]. Journal of Safety Science and Technology, 2023, 19(2): 187-194(in Chinese).

[19] 封文春, 李伟, 张桐, 等. 基于社会属性的民机复杂群体应急撤离仿真与试验研究[J]. 西北工业大学学报, 2022, 40(4): 853-864.
FENG W C, LI W, ZHANG T, et al. Emergency evacuation simulation and test research of civil aircraft complex groups based on the social attribute[J]. Journal of Northwestern Polytechnical University, 2022, 40(4): 853-864(in Chinese).

[20] SONG C C, SHAO Q, ZHU P, et al. An emergency aircraft evacuation simulation considering passenger overtaking and luggage retrieval[J]. Reliability Engineering & System Safety, 2023, 229: 108851.

[21] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[22] 李军, 张波. 基于 IFAHP-改进熵权法的煤矿综合防尘体系安全评价[J]. 煤炭技术, 2023, 42(9): 195-199.
LI J, ZHANG B. Safety evaluation of coal mine comprehensive dust control system based on IFAHP-improved entropy weight method[J]. Coal Technology, 2023, 42(9): 195-199(in Chinese).

[23] 褚天巍, 朱国庆, 周莹. 基于个体差异的商业综合体人员疏散时间[J]. 消防科学与技术, 2019, 38(5): 626-629.
CHU T W, ZHU G Q, ZHOU Y. Evacuation time of commercial complexes based on individual differences[J]. Fire Science and Technology, 2019, 38(5): 626-629(in Chinese).

[24] 国家标准化管理委员会. 中国成年人人体尺寸: GB/T 10000-2023 [S]. 北京: 国家标准化管理委员会, 2023.
National Standardization Administration. Chinese adult body size: GB/T 10000-2023[S]. Beijing: National Standardization Administration, 2023(in Chinese).

[25] 刘月姣.《中国居民营养与慢性病状况报告 (2020 年)》发布[J]. 中国食物与营养, 2020, 26(12): 2.
LIU Y J. Report on nutrition and chronic diseases of Chinese residents (2020) was published[J]. Food and Nutrition in China, 2020, 26(12): 2(in Chinese).

[26] 李伟. 应用人机工程学研究: 人体数据分析处理及其应用研究 [D]. 上海: 东华大学, 2006.
LI W. Research on applied ergonomics: analysis and processing of human data and its application[D]. Shanghai: Donghua University, 2006(in Chinese).

[27] DE WEESE R, GOWDY V. Human factors associated with the certification of airplane passenger seats: seat belt adjustment and release: DOT/FAA/AM-03/9[R]. Washton, D. C. : Federal Aviation Administration, 2003.

Emergency evacuation assessment of aircraft tilting attitude based on information entropy

CHENG Ming，WEI Keyu*

(College of Safety Science and Engineering，Civil Aviation University of China，Tianjin 300300，China)

Abstract：To explore and evaluate the emergency evacuation effectiveness of passengers under different tilt attitudes of aircraft, an emergency evacuation evaluation model was established using information entropy theory. Evaluation indicators for emergency evacuation were constructed, and wide-body aircraft tilt attitude emergency evacuation simulation experiments were conducted using real evacuation test data. The entropy weight-technique for order preference by similarity to an ideal solution (TOPSIS) method was employed to assess and rank the effectiveness of emergency evacuation. The experimental data showed that the influence of tilt attitude on evacuation time ranged from −7.32% to +3%. With the speed entropy peak focused in the early stage, there were notable disparities in pitch and roll attitudes in the personnel density at exits 2 and 3. Analysis indicated that the speed entropy and exit personnel density were two crucial factors affecting the effectiveness of emergency evacuation in tilted attitudes. The evacuation performance was optimal at a pitch angle of 5°and worst at a pitch angle of −10°, showing opposite evacuation effects at tilt angles of 5° and 10°. As a guide for the safety evaluation of wide-body aircraft design, production, and operation, the established model for aircraft is capable of rating and evaluating various tilt attitudes.

Keywords：inclined attitude; information entropy; wide-body aircraft; emergency evacuation; entropy weight method; simulation

Received：2023-12-01；**Accepted**：2024-04-01；**Published Online**：2024-04-15 11：12

URL：link.cnki.net/urlid/11.2625.V.20240412.1229.001

Foundation items：Civil Aviation Safety Capability Project (ASSA2022/13); Civil Aviation University of China Graduate Science and Technology Innovation Project (2022YJS003)

* **Corresponding author**. E-mail：490188084@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0853

飞机复合材料结构载荷测量的温度修正方法

李俊*，李文龙，高腾龙，李亚南，刘敬礼
(中国飞行试验研究院，西安 710089)

摘 要：针对温度对飞机复合材料结构载荷测量结果的严重影响问题，分析飞机结构载荷测量温度影响机理，建立包含应变电桥相对响应系数和热输出的温度修正模型，并分别给出相对响应系数和热输出与结构温度的通用关系模型。通过复合材料层压板和翼面盒段试验件的温度试验及温度-载荷组合试验对关系模型进行了验证，结果表明：在黏接剂温度特性满足要求的情况下，相对响应系数倒数与无量纲结构温度呈幂函数关系；热输出与结构温度变化量呈截距为0的三次多项式的关系。利用所建模型对某型飞机复合材料翼面结构载荷飞行测量结果进行了修正，修正后的压心相对变化最大值达到30.68%，显著提高了载荷测量的精准度，为飞机飞行载荷验证提供了可靠的数据。

关 键 词：飞机复合材料结构；载荷测量；应变电桥热输出；应变电桥相对响应系数；温度试验；温度-载荷组合试验
中图分类号：V217+.32
文献标志码：A **文章编号**：1001-5965(2026)03-0736-08

在飞机结构载荷飞行测量中，结构温度会随内、外部热源的状态变化而变化，这将导致材料的热胀冷缩，产生热应力，以及弹性模量的变化[1]。在利用应变法对飞机金属结构载荷进行飞行测量中，主要基于金属材料的各向同性，通过设计合适的应变电桥，来减少热胀冷缩对载荷测量的影响[2-5]，但未考虑被测结构热应力和弹性模量的影响，也未考虑应变计及其黏接剂温度特性的影响，因此，温度变化对载荷测量结果的影响问题时有发生，主要表现为载荷测量应变电桥热输出大和响应系数随温度变化大。由于复合材料结构的各向异性和湿热特性，在利用应变法进行复合材料结构载荷测量中，温度变化的影响将更加严重，其中，各向异性将直接影响到应变电桥的热输出；湿热特性将直接影响到应变电桥响应系数。如果不进行分析和修正，可能会严重影响载荷测量的精准度。

国外，针对高速或高超声速飞机，研究了高温对结构载荷飞行测量的影响分析和修正方法，主要通过整机或部件的地面温度试验得到载荷测量温度影响修正量[6-8]，该修正量考虑了结构热应力、应变计及其黏接剂温度特性的影响，即应变电桥热输出的影响，但未考虑结构弹性模量和应变计黏接刚度的影响，即应变电桥响应系数的影响。国内，主要在复合材料翼面结构的正常飞行使用温度范围(−55～70 ℃)内开展相关研究[9-11]，其中，载荷测量应变电桥热输出与结构温度之间的关系模型，既有采用多项式关系模型的用例[9]，也有采用线性关系模型的用例[10]；应变电桥响应系数与结构温度之间的关系模型均采用线性关系模型[10-11]。但这些模型都是基于工程试验得到的，尚未基于载荷测量温度影响机理，给出应变电桥热输出和响应系数与结构温度的通用关系模型，不利于精准分析和修正。

本文针对基于应变法的飞机复合材料结构载荷飞行测量的应变电桥安装、载荷校准试验和飞行试验等关键环节，开展理论分析和试验研究，以形成在−55～70 ℃内，既考虑应变电桥热输出影响，

收稿日期：2024-01-03；录用日期：2024-07-19；网络出版时间：2024-08-29 09:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240828.1355.002
*通信作者. E-mail：cfte_lj@163.com

引用格式：李俊，李文龙，高腾龙，等. 飞机复合材料结构载荷测量的温度修正方法[J]. 北京航空航天大学学报，2026，52(3)：736-743.
LI J，LI W L，GAO T L，et al. Temperature correction method for load measurement of aircraft composite structures[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：736-743 (in Chinese).

又考虑应变电桥响应系数影响的载荷测量温度修正模型和方法。

1　载荷测量温度影响修正方法

基于应变法的飞机结构载荷测量是通过在真实的飞行环境中, 测量安装在结构上的专门用于载荷测量的应变电桥响应, 利用载荷校准地面试验获得的载荷测量模型, 得到飞机的结构载荷[2]。载荷测量温度影响修正主要是将真实飞行温度环境中应变电桥响应修正到载荷校准地面试验温度环境(室温)中, 涉及到载荷测量温度影响修正模型和模型参数(包括应变电桥热输出和相对响应系数)的确定方法。

1.1　载荷测量温度影响修正模型

飞机翼面结构载荷测量模型[2]为

$$\begin{bmatrix} V \\ M \\ T \end{bmatrix} = \begin{bmatrix} \beta_{V_1} & \beta_{V_2} & \cdots & \beta_{V_k} \\ \beta_{M_1} & \beta_{M_2} & \cdots & \beta_{M_k} \\ \beta_{T_1} & \beta_{T_2} & \cdots & \beta_{T_k} \end{bmatrix} \begin{bmatrix} \varepsilon_{\mathrm{L}1} \\ \varepsilon_{\mathrm{L}2} \\ \vdots \\ \varepsilon_{\mathrm{L}k} \end{bmatrix} \tag{1}$$

式中: V、M和T分别为飞机翼面结构载荷测量剖面的剪力、弯矩和扭矩; β_{V_i}、β_{M_i}和β_{T_i} $(i=1,2,\cdots,k)$分别为剪力、弯矩和扭矩测量模型系数, 一般是在参考温度(通常为室温)下通过载荷校准地面试验得到的, 其中, k为参与建模的应变电桥的数量; $\varepsilon_{\mathrm{L}i}(i=1,2,\cdots,k)$为结构载荷在参考温度下引起的应变电桥响应(即载荷应变)。

在飞机结构载荷飞行测量中, 应变电桥响应飞行测量值(实测应变)ε_i为

$$\varepsilon_i = \alpha_{ti}\varepsilon_{\mathrm{L}i} + \varepsilon_{ti} \tag{2}$$

式中: α_{ti}为应变电桥相对响应系数, 其值为试验温度下的应变电桥响应系数相对于参考温度下的比值; ε_{ti}为应变电桥热输出。

由式(2)可以得到实测应变的修正式:

$$\varepsilon_{\mathrm{L}i} = (\varepsilon_i - \varepsilon_{ti})/\alpha_{ti} \tag{3}$$

将式(3)代入式(1)即可得载荷测量温度影响修正模型。下面将重点研究应变电桥热输出ε_{ti}和相对响应系数α_{ti}的确定方法, 其核心是确定ε_{ti}、α_{ti}与结构温度的关系模型。

1.2　载荷测量应变电桥热输出确定方法

根据热输出产生机理[4], 载荷测量应变电桥的第i个电阻应变计的热输出可以表示为

$$\varepsilon_{ti} = \left[\beta_{\mathrm{G}} + k_{\mathrm{G}}\left(\frac{1+k_{\mathrm{St}}H_{\mathrm{G}}}{1-\mu_0 H_{\mathrm{G}}}\right)(\alpha_{\mathrm{S}} - \alpha_{\mathrm{G}})\right]\Delta t \Big/ k_{\mathrm{I}} \tag{4}$$

式中: k_{I}为测试仪器设置的灵敏系数, 通常为 2.0; β_{G}为应变计敏感栅的电阻温度系数; k_{G}为应变计灵敏系数; k_{St}为由温度变化导致的被测结构部位横向应变与轴向应变的比值; H_{G}为应变计横向效应系数; μ_0为在校准应变计的灵敏系数时使用的标准试验件材料的泊松比; α_{G}为应变计敏感栅的线膨胀系数; α_{S}为被测结构部位的线膨胀系数; Δt为相对于参考温度的变化量, 即$\Delta t = t - t_0$, 其中, t为试验温度, t_0为参考温度。

式(4)中, k_{I}、k_{St}、H_{G}和μ_0一般为常量, 不随温度变化。令$\gamma = (1+k_{\mathrm{St}}H_{\mathrm{G}})/(1-\mu_0 H_{\mathrm{G}})$, 则$\gamma$为$k_{\mathrm{G}}$的修正系数。

应变计敏感栅的电阻随着温度上升而增加, 在常规载荷测量温度范围内, 电阻与温度的关系一般可用二次多项式来表达[12], 从而可以得到应变计敏感栅的电阻温度系数β_{G}与温度变化量Δt之间的关系式:

$$\beta_{\mathrm{G}} = \beta_{\mathrm{G0}} + D\Delta t \tag{5}$$

式中: β_{G0}为参考温度t_0时的电阻温度系数; D为斜率。

考虑到应变计的灵敏系数k_{G}与温度t成线性关系[4], 并令k_{G0}为参考温度t_0时的灵敏系数, N为斜率, 可得

$$k_{\mathrm{G}} = k_{\mathrm{G0}} + N\Delta t \tag{6}$$

考虑到线膨胀系数α_{G}和α_{S}与温度t近似成线性关系[13], 并令α_{G0}和α_{S0}分别为参考温度t_0时应变计敏感栅和被测结构部位的线膨胀系数, P和Q为斜率, 则

$$\alpha_{\mathrm{G}} = \alpha_{\mathrm{G0}} + P\Delta t \tag{7}$$

$$\alpha_{\mathrm{S}} = \alpha_{\mathrm{S0}} + Q\Delta t \tag{8}$$

将式(5)～式(8)代入式(4)中, 可得

$$\varepsilon_{ti} = A_{i1}\Delta t + A_{i2}(\Delta t)^2 + A_{i3}(\Delta t)^3 \tag{9}$$

式中: $A_{i1} = [\beta_{\mathrm{G0}} + \gamma k_{\mathrm{G0}}(\alpha_{\mathrm{S0}} - \alpha_{\mathrm{G0}})]/k_{\mathrm{I}}$

$$A_{i2} = \{D + \gamma[k_{\mathrm{G0}}(Q-P) + N(\alpha_{\mathrm{S0}} - \alpha_{\mathrm{G0}})]\}/k_{\mathrm{I}}$$

$$A_{i3} = \gamma N(Q-P)/k_{\mathrm{I}}$$

由式(9)可知, 在考虑应变计和被测结构材料特性的情况下, 载荷测量应变电桥单个电阻应变计热输出ε_{ti}与温度变化量Δt呈截距为 0 的三次多项式的关系。

由载荷测量应变电桥加减特性可以得到应变电桥热输出与温度变化量之间的关系[9], 即

$$\varepsilon_{\mathrm{t}} = B_1\Delta t + B_2(\Delta t)^2 + B_3(\Delta t)^3 \tag{10}$$

式中: 当应变电桥为全桥时, 系数$B_j = A_{1j} - A_{2j} + A_{3j} - A_{4j}$; 当应变电桥为半桥时, 系数$B_j = A_{1j} - A_{2j}$, 其中, A_{ij}为应变电桥第i电阻应变计热输出与温度

变化量之间的三次多项式的j次项系数，$i=1,2,3,4$，$j=1,2,3$。

由式(10)可知，只有当三次多项式系数$B_j=0$时，才能始终确保应变电桥热输出为0，这在工程测量中很难实现，因此，载荷测量应变电桥热输出是客观存在的。在飞机结构载荷测量中，需要利用应变电桥热输出与温度变化量之间的关系对载荷测量应变电桥响应进行分析和修正。

1.3 载荷测量应变电桥相对响应系数确定方法

载荷测量应变电桥响应[2-3]为

$$\mu=\frac{1}{4}k_{\mathrm{G}}U\varepsilon \tag{11}$$

式中：U为应变电桥激励电压；ε为应变电桥各桥臂应变值的代数和。

由于结构应变为载荷F与刚度E之比[14]，即

$$\varepsilon=F/E \tag{12}$$

将式(12)代入式(11)，得到

$$\mu=k_{\mathrm{G}}UF/4E$$

应变电桥响应系数[2-3]为

$$I=\mu/F=k_{\mathrm{G}}U/4E$$

进而得到应变电桥响应系数倒数：

$$1/I=4E/k_{\mathrm{G}}U \tag{13}$$

由于温度对复合材料模量(包括横向模量、纵向模量和剪切模量等)的影响可以用无量纲温度$\overline{T}$的幂函数来近似表达[15]，并考虑到结构形式一旦确定，温度对复合材料结构刚度的影响主要体现在温度对材料模量的影响上，因此，结构刚度E也可以用无量纲温度$\overline{T}$的幂函数来近似表达，即

$$E=A+B\overline{T}^{c} \tag{14}$$

式中：A、B和c分别为幂函数的截距、系数和指数。

将式(14)代入式(13)得到应变电桥响应系数的倒数与无量纲温度$\overline{T}$之间存在幂函数关系，即

$$\frac{1}{I}=\frac{4}{k_{\mathrm{G}}U}\left(A+B\overline{T}^{c}\right) \tag{15}$$

式中：

$$\overline{T}=(T_{\mathrm{g}}-t)/(T_{\mathrm{g}}-t_0) \tag{16}$$

式中：T_{g}为干态下复合材料的玻璃化转变温度。

在参考温度试验时，将$t=t_0$代入式(16)，得到：$\overline{T}=1$。将其代入式(15)，进而可得

$$1/I_0=4(A+B)/k_{\mathrm{G}}U \tag{17}$$

考虑到应变电桥相对响应系数为试验温度下的应变电桥响应系数相对于参考温度下的比值，即$\alpha_{\mathrm{t}}=I/I_0$，由式(15)和式(17)可得

$$\frac{1}{\alpha_{\mathrm{t}}}=\frac{1}{I/I_0}=\frac{A}{A+B}+\frac{B}{A+B}\overline{T}^{c}$$

令$A'=\dfrac{A}{A+B}$，$B'=\dfrac{B}{A+B}$，则

$$1/\alpha_{\mathrm{t}}=A'+B'\overline{T}^{c} \tag{18}$$

由式(18)可知，在考虑被测结构材料特性的情况下，α_{t}的倒数与无量纲温度$\overline{T}$之间存在幂函数关系。

在飞机结构载荷测量中，确定$\alpha_{\mathrm{t}i}$的步骤为：①通过载荷测量部件的温度-载荷组合试验，得到应变电桥在不同温度下对不同加载点的响应系数；②计算各温度下的应变电桥相对响应系数；③计算同一电桥在同一温度下针对不同加载点的相对响应系数的平均值；④利用式(18)，对各应变电桥在给定温度下相对响应系数平均值与相应的结构温度进行拟合分析，得到相对响应系数与温度的拟合关系式；⑤将结构温度实测值代入拟合关系式，得到相应的相对响应系数。

将确定的应变电桥热输出和相对响应系数代入式(3)得到参考温度下的应变电桥响应，并将其代入式(1)，得到真实的飞机结构载荷，具体流程如图1所示。

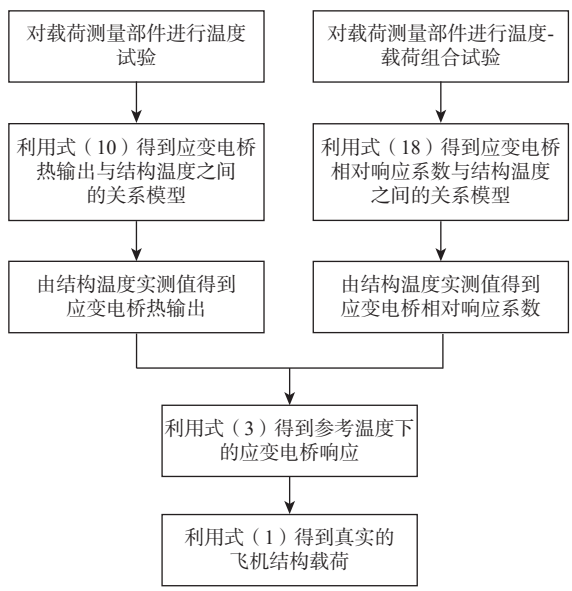


图1 载荷测量温度修正流程图

Fig. 1 Temperature correction flow diagram of load measurement

2 地面验证试验

地面验证试验包括复合材料层压板、翼面盒段的温度试验和翼面盒段的温度-载荷组合试验。其中，温度试验主要用于验证应变电桥热输出与结构温度的关系模型正确性；温度-载荷组合试验主要用于验证应变电桥相对响应系数与结构温度的关

系模型正确性。为验证本文模型的普适性, 试验考虑了不同的铺层形式和加载类型的试验件及不同的应变电桥安装工艺方案。

2.1　试验件

试验件包括复合材料层压板和翼面盒段。试验件材料体系为 T800-180℃ 固化环氧树脂碳纤维复合材料。层压板试验件共 144 件, 涉及 8 种铺层形式和 3 种加载类型, 具体种类及其数量如表 1 所示, 层压板试验件厚度取决于具体的铺层形式, 长度和宽度在满足相关试验标准情况下取较大值, 以便于安装载荷测量应变电桥。翼面盒段试验件与某飞机平尾翼盒结构相似, 尺寸相当。

表 1　层压板试验件种类和数量

Table 1　Type and quantity of laminate coupons

铺层编号	拉伸件数量	压缩件数量	剪切件数量	合计/件
A1	6	6	6	18
A2	6	6	6	18
A3	6	6	6	18
A4	6	6	6	18
A5	6	6	6	18
A6	6	6	6	18
A7	6	6	6	18
A8	6	6	6	18

2.2　应变电桥安装

根据应变计、黏接剂和电桥类型等工艺要素差异形成 6 种应变电桥安装工艺方案, 具体如表 2 所示。其中, 应变计为来自 3 个厂家的电阻应变计, 分别用“H”、“Z”和“M”表示; 黏接剂包括 Z70、HY-914 和 AE-10 这 3 种, 分别用“A”、“B”和“C”表示; 电桥类型包括全桥和半桥 2 种。在翼面盒段试验件上采用工艺方案Ⅰ、Ⅱ、Ⅲ、Ⅴ和Ⅵ安装应变电桥; 在层压板试验件上采用工艺方案Ⅰ、Ⅲ和Ⅳ安装应变电桥。

表 2　应变电桥安装工艺方案

Table 2　Technology scheme for installation of strain-gage bridge

工艺方案	应变计	黏接剂	电桥类型
Ⅰ	H	A	全桥
Ⅱ	H	A	半桥
Ⅲ	H	B	全桥
Ⅳ	Z	A	全桥
Ⅴ	Z	B	全桥
Ⅵ	M	C	半桥

2.3　试验内容及方法

层压板和翼面盒段温度试验包括低温试验(室温～−55 ℃)和高温试验(室温～70 ℃)。翼面盒段温度-载荷组合试验包括分别在−55～70 ℃ 之间的 9 个试验温度下的拉、压加载试验, 具体实施方法: 利用承力台架约束试验件, 将试验件的载荷测量剖面置于温控箱中, 在试验件外段下表面粘贴加载垫板, 采用液压协调加载系统通过加载垫板对试验件施加拉伸和压缩载荷, 如图 2 所示。



图 2　翼面盒段试验件温度-载荷组合试验

Fig. 2　Temperature-load combined test for specimen of wing-box structure

2.4　载荷测量应变电桥热输出分析

复合材料层压板和翼面盒段温度试验共得到了 656 组试验数据。利用式(10)对应变电桥热输出与结构温度变化量进行拟合分析, 结果如表 3 所示。结果表明: ①拟合优度 R^2 大于 0.9 的应变电桥数占 94.05%, R^2 小于 0.9 的原因有 2 种: 一种为应变电桥热输出小, 约占 3.51%, 这是载荷测量所期望的; 另一种为应变电桥热输出异常波动, 约占 2.44%, 这与应变电桥安装工艺质量有关; ②采用工

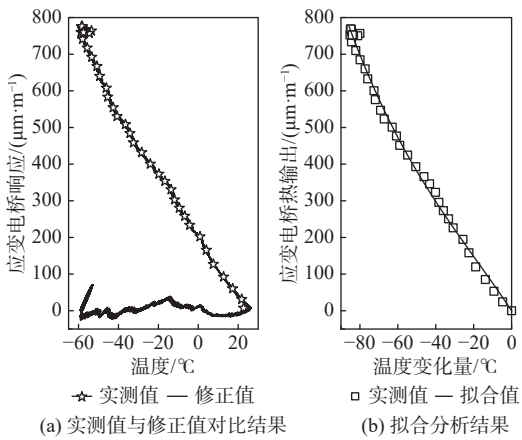
表 3　应变电桥热输出与结构温度变化量的三次多项式拟合结果

Table 3　Results of cubic polynomial fitting between thermal output of strain bridge and structural temperature change

试验件类型	温度范围	工艺方案	电桥数	R^2≥0.9的电桥占比/%	R^2<0.9的电桥占比/%	
					响应小	响应异常
层压板	室温～−50℃	Ⅰ	144	95.83	2.78	1.39
		Ⅳ	143	93.71	4.90	1.40
	室温～65℃	Ⅰ	123	94.31	4.88	0.81
		Ⅳ	123	92.68	3.25	4.07
盒段件	室温～−50℃	Ⅰ	4	100.00		
		Ⅱ	11	100.00		
		Ⅲ	20	100.00		
		Ⅴ	15	100.00		
		Ⅵ	12	83.33	16.67	
	室温～65℃	Ⅰ	4	100.00		
		Ⅱ	11	100.00		
		Ⅲ	19	84.21		15.79
		Ⅴ	15	80.00		20.00
		Ⅵ	12	100.00		

艺方案Ⅲ和Ⅴ安装的应变电桥响应在高温试验中更容易出现异常波动，这主要与 B 型黏接剂玻璃化转变温度低有关。

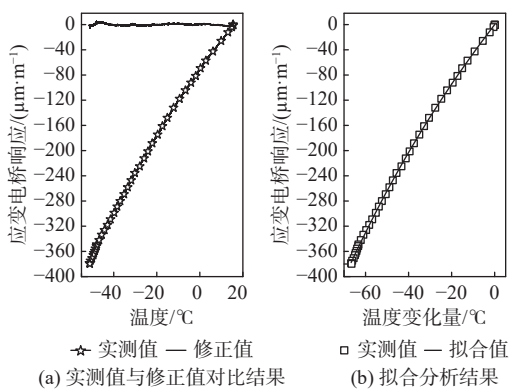
图 3 为层压板载荷测量应变电桥 A6T2HC 热输出拟合分析和修正的典型结果。由图可知，热输出与结构温度变化量的三次多项式拟合效果很好，R^2达到 0.999 07；修正后的应变电桥热输出明显减小，绝大部分达到±50 μm/m 以内。图 4 为翼面盒段试验件载荷测量应变电桥 2HW 热输出拟合分析和修正的典型结果。由图可知，热输出与结构温度变化量的三次多项式拟合效果很好，R^2达到 0.999 96；修正后的应变电桥热输出明显减小，达到±10 μm/m 以内。



(a) 实测值与修正值对比结果　(b) 拟合分析结果

图 3　安装在层压板上的应变电桥热输出拟合分析与修正结果

Fig. 3　Fitting analysis and correction results of thermal output of strain bridge installed on laminate



(a) 实测值与修正值对比结果　(b) 拟合分析结果

图 4　安装在翼面盒段上的应变电桥 2HW 热输出拟合分析与修正结果

Fig. 4　Fitting analysis and correction results of thermal output of strain bridge 2HW installed on wing box

通过复合材料层压板和翼面盒段温度试验表明，在黏接剂温度特性满足要求的情况下，大部分应变电桥热输出较大，并且应变电桥热输出与结构温度变化量呈截距为 0 的三次多项式关系，利用这种关系进行分析和修正，可以明显减小应变电桥热输出。

2.5　应变电桥相对响应系数分析

复合材料翼面盒段温度-载荷组合试验共得到了 972 组试验数据（涉及 27 个应变电桥、9 个温度状态和 4 个加载点）。利用本文 1.3 节给出的方法，得到应变电桥 HJ 相对响应系数倒数$1/\alpha_t$与无量纲结构温度$\overline{T}$（T_g=190 ℃，t_0=18 ℃）的幂函数拟合关系，典型结果如图 5 所示。结果表明：在−55～70 ℃范围内，采用 B 型黏接剂安装的应变电桥拟合效果较差，最大拟合优度R^2仅为 0.587 03，而且相应的响应系数相对于参考温度的变化较大，最大相对变化达到了 12.77%～47.69%，这与 B 型黏接剂玻璃化转变温度低有关，典型结果如图 6 所示；采用 A 型和

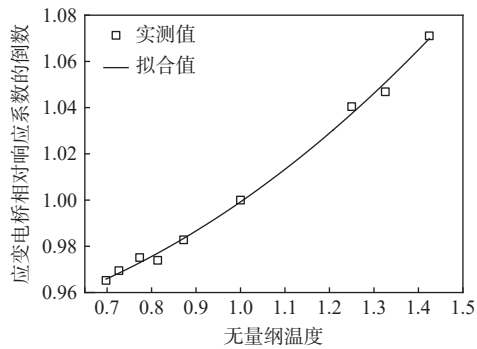


图 5　应变电桥 HJ 相对响应系数的倒数与无量纲温度关系

Fig. 5　Relationship between reciprocal of relative response coefficients of strain bridge HJ and dimensionless temperature

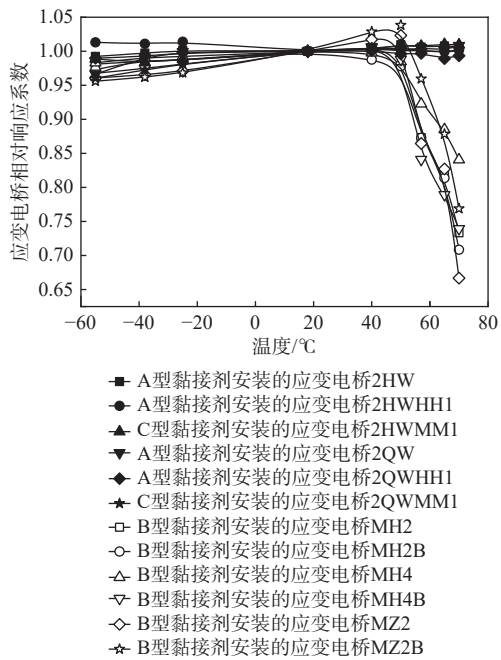


图 6　采用不同黏接剂在翼面盒段上安装的应变电桥相对响应系数随温度变化

Fig. 6　Relative response coefficients of strain bridges installed on wing box with different adhesives varies with temperature

C 型黏接剂安装的应变电桥拟合效果很好, 大部分应变电桥的R^2大于 0.9, 最大R^2达到 0.999 22, 并且相应的响应系数相对变化均较小, 最大相对变化仅为 0.96%～8.84%, R^2小于 0.9 的应变电桥响应系数最大相对变化均小于 1.4%。

基于上述分析, 选取采用温度特性良好的黏接剂(A 或 C)安装的应变电桥, 通过室温下的载荷校准地面试验, 建立载荷测量方程。利用幂函数拟合关系式, 得到−55 ℃ 和 70 ℃ 下的剪力方程中所选应变电桥(其编号分别为 HW、QW、HJ、QJ)的相对响应系数α_{ti}(见表 4), 利用式(3)可得到修正后的应变电桥响应, 将其代入剪力测量方程可得到修正后的实测剪力, 然后, 用−55 ℃ 和 70 ℃ 下的实际施加的剪力进行检验, 得到修正后的检验误差(见表 4)。结果表明, 修正后剪力测量方程的检验误差明显减小。

表 4　剪力方程中应变电桥的相对响应系数及方程检验误差

Table 4　Relative response coefficients of strain bridge in shear equation and check error of equation

温度/℃	应变电桥相对响应系数α_{ti}				检验误差/%	
	HW	QW	HJ	QJ	修正前	修正后
−55	0.984 5	0.959 4	0.934 7	0.942 6	5.00	0.92
70	1.011 4	1.001 9	1.035 4	1.030 1	3.69	0.60

通过复合材料翼面盒段温度-载荷组合试验表明, 在黏接剂温度特性满足要求的情况下, 温度对应变电桥响应系数影响较小, 应变电桥相对响应系数倒数与无量纲结构温度之间呈幂函数关系, 利用这种关系进行分析和修正, 可以提升极限温度(−55 ℃ 和 70 ℃)条件下载荷测量精度。

3　飞行测量载荷修正

复合材料层压板和翼面盒段试验件的温度试验及温度-载荷组合试验结果表明, 在−55～70 ℃ 范围内, 并在应变电桥黏接剂温度特性满足要求的情况下: ①应变电桥相对响应系数对载荷测量影响小, 即$\alpha_{ti}\to 1$, 在飞机结构载荷测量中, 可以通过修正得到更高的载荷测量精度, 但其修正模型系数需要通过真实的载荷测量部件的地面温度-载荷组合试验确定, 目前国内尚不具备相应的试验条件, 无法对其进行修正, 但需要在载荷测量最终结果的不确定度分析中考虑其影响; ②应变电桥热输出对载荷测量影响大, 在飞机结构载荷测量中, 如果不进行修正, 可能会严重影响载荷测量的精准度。通常, 应变电桥热输出修正模型系数需要通过真实载荷测量部件的地面温度试验确定, 但目前所有进行载荷测量的飞机均未开展地面温度试验, 考虑到, 当$\alpha_{ti}\to 1$时, $\varepsilon_i\approx\varepsilon_{Li}+\varepsilon_{ti}$, 在载荷应变不变或变化很小的情况下, 实测应变相对参考温度下的变化量$\Delta\varepsilon_i\approx\varepsilon_{ti}$。因此, 可以通过载荷测量飞行试验数据, 分析确定应变电桥热输出与结构温度变化量之间三次多项式的系数。

飞机结构载荷飞行测量数据修正的步骤: ①分析提取载荷变化小、温度变化大的飞行测量数据; ②分析实测应变相对变化量与结构温度变化量的相关性, 选取相关系数较大的实测数据进行拟合分析, 得到截距为 0 的三次多项式拟合关系式; ③将实测的结构温度变化量代入上述关系式, 可得到实测应变的修正量, 进而可得到修正后的实测应变; ④将其代入相应的载荷方程, 即可得到修正后的实测载荷。

利用上述方法对某型飞机机翼、平尾和垂尾结构载荷飞行测量结果进行了修正。结果表明, 分析提取的应变电桥 YJY2W2 热输出与结构温度变化量之间存在明显的三次多项式关系, 典型结果如图 7 所示; 修正后的结构载荷不再随温度变化, 典型结果如图 8 所示。

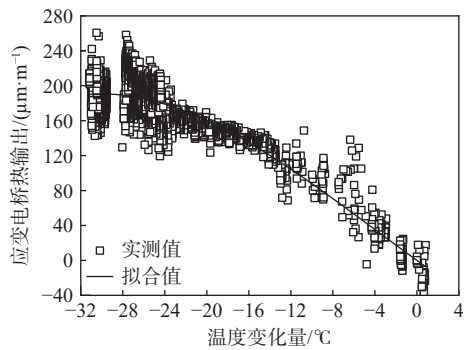


图 7　机翼载荷测量应变电桥 YJY2W2 热输出与结构温度变化量的三次多项式拟合结果

Fig. 7　Results of cubic polynomial fitting between thermal output of strain bridge YJY2W2 and structural temperature change for wing load measurement

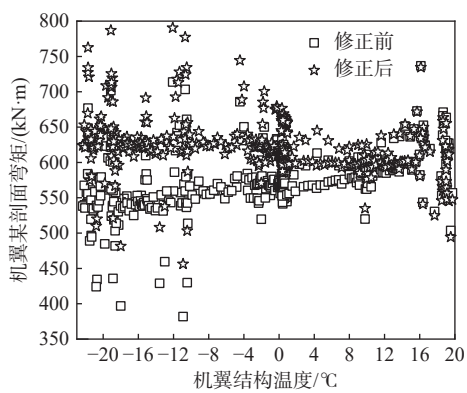


图 8　机翼载荷与结构温度的关系

Fig. 8　Relationship between wing load and structural temperature

表5为某型飞机机翼、平尾和垂尾在典型飞行状态下的实测载荷修正结果。表中，J_1和J_2分别代表在高度h_p=7 620 m，Ma=0.70，进行对称拉起机动过程中，法向过载n_z=0.97、2.13时机翼载荷情况；P_1～P_3分别代表在高度h_p=7 620 m，Ma=0.70，进行对称拉起机动过程中，法向过载n_z=0.97、0.08、2.12时平尾载荷情况；C_1和C_2分别代表在高度h_p=9 144 m，Ma=0.78，进行稳定侧滑机动过程中，侧向过载n_y=0.27、−0.23时垂尾载荷情况；剪力和弯矩为实测剪力V和弯矩M与速压q和机翼面积S乘积的比值，即$V/(qS)$和$M/(qS)$；压心为剖面外载荷压心到测量剖面的展向距离，即M/V；压心相对变化为修正后压心相对于修正前的变化，反映了弯矩、剪力修正程度的综合影响。从表5可以看出：①机翼、平尾和垂尾弯矩和剪力都有不同程度的修正，压心相对变化最大值达到30.68%，对应的载荷情况为P_1；②修正后展向压心更符合该型飞机机翼、平尾和垂尾受载特点，其中，在对称拉起机动中，机翼在n_z=2.13状态下的展向压心明显小于n_z=0.97状态，平尾在n_z=0.97、0.08、2.12状态下的展向压心基本相同，机翼和平尾压心随着载荷的增大而减小；在稳定侧滑机动中，垂尾在n_y=0.27、n_y = −0.23状态下的展向压心基本相同。综上，通过修正显著提高了载荷测量的精准度。

表5 实测载荷及其修正结果

Table 5 Measured loads and their Correction results

载荷情况	剪力	弯矩/m	压心/m	修正后剪力	修正后弯矩/m	修正后压心/m	压心相对变化/%
J_1	0.071 6	0.289 5	4.04	0.075 5	0.330 0	4.37	8.17
J_2	0.168 2	0.672 8	4.00	0.172 1	0.712 3	4.14	3.50
P_1	−0.005 1	−0.009 0	1.76	−0.004 3	−0.009 9	2.30	30.68
P_2	−0.010 9	−0.021 7	1.99	−0.010 1	−0.022 7	2.25	13.07
P_3	0.005 6	0.015 4	2.75	0.006 4	0.014 5	2.27	−17.45
C_1	0.025 1	0.057 6	2.29	0.024 1	0.060 1	2.49	8.73
C_2	−0.019 6	−0.053 3	2.72	−0.020 5	−0.050 8	2.48	−8.82

在飞机结构载荷测量中，针对载荷测量部件未开展地面温度试验的情况，采用基于载荷测量飞行试验数据识别应变电桥热输出修正模型系数的方法，解决了当前载荷测量温度影响修正难题，既有利于提高载荷测量的精准度，又有利于缩短试验周期，降低试验成本。

4 结 论

1）采用包含应变电桥热输出和相对响应系数的载荷测量温度修正模型，在常规的载荷测量温度范围内修正飞行实测载荷，修正结果是可信可靠的，修正方法是合理可行的。

2）在黏接剂温度特性满足要求的情况下，应变电桥热输出与结构温度变化量呈截距为0的三次多项式关系；应变电桥相对响应系数的倒数与无量纲结构温度之间呈幂函数关系。

3）在应变电桥相对响应系数趋近于1的情况下，通过提取飞机结构载荷变化小和结构温度变化大的实测应变，确定应变电桥热输出与结构温度变化量之间的三次多项式关系模型系数的方法是科学有效的。

4）该方法已应用于某型飞机飞行载荷测量试飞，修正后的压心相对变化最大值达到30.68%，显著提高了载荷测量精准度，为飞行载荷验证提供了可靠的数据。

参考文献（References）

[1] 中国飞行试验研究院. 飞机结构飞行试验指南[M]. 北京: 航空工业出版社, 2022: 29-37.
Chinese Flight Test Establishment. Aircraft structural flight test guide[M]. Beijing: Aviation Industry Press, 2022: 29-37(in Chinese).

[2] 郭正旺, 周友明. 结构完整性试飞-飞机载荷与强度试飞[M]. 北京: 航空工业出版社, 2018: 18-51.
GUO Z W, ZHOU Y M. Flight test of structural integrity- flight test of aircraft load and strength[M]. Beijing: Aviation Industry Press, 2018: 18-51(in Chinese).

[3] JECKINS J M, DE ANGELIS V M. A summary of numerous strain-gage load calibrations on aircraft wings and tails in a technology format: NASA TM-4804[R]. Washton, D. C. : National Aeronautics and Space Administration, 1997.

[4] Micro-Measurements. Strain gage thermal output and gage factor variation with temperature: TN-504-1[R]. North Carolina: Micro-Measurements, 2014.

[5] 李俊, 蒋献. 飞机主动式余度作动舵面铰链力矩飞行测量方法[J]. 空军工程大学学报, 2023, 24(3): 43-49.
LI J, JIANG X. A flight measurement method for hinge moment of aircraft control surface with redundant actuation system in active mode[J]. Journal of Air Force Engineering University, 2023, 24(3): 43-49(in Chinese).

[6] FIELDS R A. A study of the accuracy of a flight-heating simulation and its effect on load measurement: NASA TN D-5741[R]. Washton, D. C. : National Aeronautics and Space Administration, 1970.

[7] REARDON L R. Evaluation of a strain-gage load calibration on a low aspect-ratio wing structure at elevented temperature: NASA TP-2921[R]. Washton, D. C. : National Aeronautics and Space Administration, 1989.

[8] JENKINS J M, QUINN R D. A historical perspective of the YF-12A thermal loads and structures program: ASA TM-104317[R]. Washton, D. C. : National Aeronautics and Space Administration, 1996.

[9] 李俊. 飞机结构载荷测量应变电桥热输出修正方法[J]. 空军工程大学学报 (自然科学版), 2022, 23(1): 64-69.

LI J. A thermal output correction method of strain-gage bridge for load measurement of aircraft structure[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(1): 64-69 (in Chinese).

[10] 朱江辉. 复合材料翼面结构载荷测量及温度修正技术研究[J]. 中国测试, 2023, 49(7): 101-111.
ZHU J H. Study on load measurement and temperature correction technology of composite wing structure[J]. China Measurement & Test, 2023, 49(7): 101-111(in Chinese).

[11] 李文龙, 李俊, 刘敬礼, 等. 考虑温度影响的复合材料结构载荷测量方法[J]. 空军工程大学学报, 2024, 25(1): 47-51.
LI W L, LI J, LIU J L, et al. A load measurement method for composite structure considering the effect of temperature[J]. Journal of Air Force Engineering University, 2024, 25(1): 47-51(in Chinese).

[12] 丁冬, 濮兴庭. 利用 Origin 处理测金属电阻温度系数实验数据[J]. 大学物理实验, 2011, 24(6): 59-60.
DING D, PU X T. Processing the experiment data of measuring temperature coefficient of metallic resistance with origin[J]. Physical Experiment of College, 2011, 24(6): 59-60(in Chinese).

[13] 陆集平. 碳-环氧复合材料的热物理和热冲击性能[J]. 宇航材料工艺, 1983, 13(6): 7-10.
LU J P. Thermophysical and thermal shock properties of carbon-epoxy composites[J]. Aerospace Materials & Technology, 1983, 13(6): 7-10(in Chinese).

[14] ALAN B, STUART D, DONALD K. 飞机结构复合材料技术[M]. 2 版. 柴亚南, 丁惠梁, 译. 北京: 航空工业出版社, 2015: 140-156.
ALAN B, STUART D, DONALD K. Composite materials for aircraft structures[M]. 2nd ed. CHAI Y N, DING H L, translated. Beijing: Aviation Industry Press, 2015: 140-156(in Chinese).

[15] 谢伟, 窦鹏鹏, 薛展. 湿热环境下复合材料层合板本构模型研究及其应用[J]. 航空工程进展, 2019, 10(1): 62-72.
XIE W, DOU P P, XUE Z. Research and application of the constitutive model of composite laminates in hydrothermal environments[J]. Advances in Aeronautical Science and Engineering, 2019, 10(1): 62-72(in Chinese).

Temperature correction method for load measurement of aircraft composite structures

LI Jun*, LI Wenlong, GAO Tenglong, LI Yanan, LIU Jingli

(Chinese Flight Test Establishment, Xi'an 710089, China)

Abstract: In view of the serious influence of temperature on load measurement results of aircraft composite structures, the influence mechanism of temperature on load measurement of aircraft structure was analyzed, and a temperature correction model including the relative response coefficient and thermal output of the strain bridge is established, and the general relation models of relative response coefficient and thermal output with structural temperature are given respectively. These relation models are verified by the temperature test and the temperature-load combined test of composite laminate coupons and the wing box test component. The findings indicate that, when the adhesive's temperature characteristics are met, the relationship between the thermal output and the structural temperature change is a cubic polynomial with zero intercept, and the reciprocal of the relative response coefficient is a power function of the dimensionless structure temperature. By using this technique, the flight measurement results of an aircraft's composite wing structure load can be corrected. The maximum relative change of the pressure center following modification is 30.68%, which greatly increases load measurement accuracy and yields trustworthy data for aircraft flight load verification.

Keywords: aircraft composite structures; load measurement; thermal output of strain bridge; relative response coefficient of strain bridge; temperature test; temperature-load combined test

Received: 2024-01-03; **Accepted**: 2024-07-19; **Published Online**: 2024-08-29 09:33

URL: link.cnki.net/urlid/11.2625.V.20240828.1355.002

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0774

反舰弹道导弹杀伤效能的不确定性研究

王志辉[1, 2, *]，相志宁[1]，高萍[3]

(1. 陆军炮兵防空兵学院南京校区，南京 211132；2. 中国人民解放军国防大学 联合作战学院，北京 100091；
3. 装备发展部 装备体系评估中心，北京 100009)

摘 要：在强对抗和不完备信息环境下获取较为准确的杀伤效能是反舰弹道导弹作战运用研究的重难点。基于此，提出迭代减少反舰弹道导弹杀伤效能不确定性的研究方法。建立杀伤效能评估的仿真模型；采用信度分布函数刻画影响因素的不确定性；基于仿真模型计算杀伤效能的信度分布函数；利用信度熵分析每个因素对杀伤效能不确定性的全局影响度。通过针对性减少关键因素的不确定性，高效减少杀伤效能的不确定性。在示例中，杀伤概率不确定范围由0.2～0.6缩小至0.41～0.50。所提方法为作战效能评估提供了一种新的思路。

关 键 词：反舰弹道导弹；不确定性；信度分布函数；信度熵；效能评估
中图分类号：V37; N945
文献标志码：A **文章编号**：1001-5965(2026)03-0744-09

反舰弹道导弹的杀伤效能评估是反舰弹道导弹作战运用研究的重要内容。通常把反舰弹道导弹杀伤敌方舰艇的概率作为杀伤效能的评估指标。在特定作战场景下，给定数量的反舰弹道导弹杀伤某型敌舰的概率(长期频率)是固定的，但由于反舰作战对抗性强，信息不完备，这个固有概率是不确定(未确知)的。这主要有2方面的原因，①影响因素与杀伤效能的映射关系不确定，②影响因素自身取值不确定。当不确定程度过高时，就无法为相关运用与研发工作提供支撑。如何在强对抗、信息不完备环境下获取较为准确的杀伤效能，对相关系统的发展具有重要意义。

未确知性是由于信息不完备所导致的一种认知型不确定性，本文将其称为“不确定性”。不确定性是由于信息不完备导致的认知不确定性。关于不确定条件下的作战效能研究，文献[1-2]用模糊综合评估法分别评估了海军鱼雷系统的作战效能和舰艇在复杂海况下的环境适应性。文献[3]提出将灰色关联分析与模糊综合评判相结合的评估方法，并应用于GNSS欺骗干扰设备的效能评估。文献[4]运用灰色层次分析法并结合径向基函数神经网络进行数据分析，评估有人-无人协同作战效能。文献[5]把直觉模糊集和综合评估模型结合，提出海上联合作战体系在不确定多任务场景下的作战效能的评估方法。文献[6]采用层次分析法(analytic hierarchy process，AHP)和熵值法确定指标权重；然后，运用灰色模糊综合评价法评估模型中各项指标灰色度和模糊性；最后，计算网络防御作战能力优良等级。文献[7]把粗糙集和综合评估模型相结合，提出不确定条件下地面作战目标威胁评估方法。文献[8]通过构建证据网络提出体系作战效能的评估方法。文献[9]提出基于改进的Dempster-Shafer(D-S)证据理论的相控阵雷达作战效能的评估方法。该方法采用Kulsinski差异和模糊熵对传统的证据理论方法进行改进，旨在减少指标数据在融合过程中融合权重不确定性的影响。这些方法适合不确定条件下的方案综合排序，但不适合本文的杀伤效能评估问题，因为无法帮助评估人员概略

收稿日期：2023-11-30；**录用日期**：2024-01-12；**网络出版时间**：2024-02-02 10:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20240201.1029.003
基金项目：国家自然科学基金 (61703412)
* **通信作者**. E-mail：zhwang66@foxmail.com

引用格式：王志辉，相志宁，高萍. 反舰弹道导弹杀伤效能的不确定性研究[J]. 北京航空航天大学学报，2026，52(3)：744-752.
WANG Z H，XIANG Z N，GAO P. Research on uncertainty in kill effectiveness of anti-ship ballistic missiles[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：744-752 (in Chinese).

掌握杀伤目标的概率。

只有充分减少杀伤效能的不确定性, 才能获得更为准确的杀伤效能。目前, 很少有指导评估人员如何减少杀伤效能不确定性的研究。针对这个问题, 本文以反舰弹道导弹杀伤效能为研究背景, 研究如何量化减少杀伤效能的不确定性。首先, 对相关理论进行了阐释; 然后, 分析了反舰弹道导弹的杀伤过程, 找出主要影响因素; 接着, 通过仿真手段建立了给定参数下的反舰弹道导弹杀伤效能评估模型; 在此基础上, 用信度分布函数表征重要变量的不确定性, 计算杀伤效能的信度分布函数曲线; 最后, 利用信度熵找出对杀伤效能不确定性影响大的因素, 针对性减少关键因素的不确定性, 从而迭代减少导弹杀伤效能的不确定性, 直至获取能够支撑决策的评估结果。

1　基本理论

1.1　不确定性的分类

不确定性可以粗略地分为偶然型和认知型[10-12]。偶然型描述系统固有的不确定性, 比如抛一枚硬币出现正反面的概率是固定的。这种不确定性不能被减少。认知型描述由于信息不足而导致的认知上的不确定, 比如对导弹杀伤半径的不确定, 通过做试验就能逐步减少不确定性。这种不确定性可以通过引入更多信息来减少。杀伤效能不确定性属于认知型, 强调在一定论域内不确定, 并非完全未知。

在特定作战场景下, 多枚反舰弹道导弹杀伤敌方舰艇的概率(长期频率)是固定的, 是偶然型不确定。本文通过仿真手段构建了反舰弹道导弹杀伤效能的评估模型, 能够概略评估给定参数下杀伤敌舰的概率。但在实际问题中, 模型参数(如导弹的性能参数、敌方的防御水平)往往不确定, 杀伤概率也就不确定。如何描述不确定变量的不确定性, 本文引入了信度分布函数。

1.2　信度分布函数原理

信度表示主观相信某件事发生的程度, 取值范围为 (0, 1)。形象地理解, 信度取 0.9 代表有 90% 把握, 信度取 0.2 代表有 20% 把握。主观概率也有相近的含义, 但和信度不同的是, 主观概率的运算规则是乘法规则, 而信度通常是取小规则。从理论依据看, 信度比主观概率的容错性更强, 可以更好地处理不确定性[13]。对于一个不确定变量, 如果评估人员所拥有的信息或经验不足, 则估计出的分布函数与其固有频率往往不够接近。这时, 采用信度工具处理将更加稳健可靠[14]。

信度分布函数描述的是不确定变量ξ小于某个值的信度。信度分布函数的相关定义如下:

定义 1[13]　对于任意实数 x, 不确定变量ξ的信度分布函数Φ被定义为

$$\Phi(x) = M\{\xi \leqslant x\} \tag{1}$$

式中: 集合函数M被称为不确定测度[13]。

定义 2[13]　对于x在$0 < \Phi(x) < 1$上, 如果信度分布函数是一个连续的严格递增函数, 那么$\Phi(x)$是正则的, 并且 $\lim\limits_{x\to-\infty}\Phi(x) = 0$, $\lim\limits_{x\to+\infty}\Phi(x) = 1$。

定义 3[13]　假设ξ是具有信度分布函数$\Phi(x)$的一个不确定变量, 那么反函数$\Phi^{-1}(x)$称为ξ的信度分布的逆函数。

定义 4[13]　如果在任意 $B_1, B_2, \cdots, B_n$的 Borel 集上, 有

$$M\left\{\bigcap_{i=1}^{n}(\xi_i \in B_i)\right\} = \bigwedge_{i=1}^{n} M\{\xi_i \in B_i\} \tag{2}$$

则不确定变量$\xi_1, \xi_2, \cdots, \xi_n$是独立的。

信度分布函数的运算具有如下定理:

定理[13]　$\xi_1, \xi_2, \cdots, \xi_n$为独立的不确定变量, $\Phi_1, \Phi_2, \cdots, \Phi_n$分别为其正则信度分布函数, 如果$f(\cdot)$为一个严格单调递增函数, α为信度值, 那么不确定变量$\xi = f(\xi_1, \xi_2, \cdots, \xi_n)$具有信度分布的逆函数表示为

$$\psi^{-1}(\alpha) = f(\Phi_1^{-1}(\alpha), \Phi_2^{-1}(\alpha), \cdots, \Phi_n^{-1}(\alpha)) \tag{3}$$

同理, 经过变换, 如果$f(\cdot)$为一个严格单调递减函数, 则

$$\psi^{-1}(\alpha) = f(\Phi_1^{-1}(1-\alpha), \Phi_2^{-1}(1-\alpha), \cdots, \Phi_n^{-1}(1-\alpha)) \tag{4}$$

1.3　不确定性的度量—信度熵

信度熵是度量某个变量不确定程度的数学工具[15], 定义如下。

定义 5　设ξ为一个不确定的连续变量, 其信度分布函数为Φ_ξ, 对应的信度熵为

$$S_\xi = \int_L^U \left[\Phi_\xi(x)\log_2\left(\frac{1}{\Phi_\xi(x)}\right) + (1-\Phi_\xi(x))\log_2\left(\frac{1}{1-\Phi_\xi(x)}\right)\right]\mathrm{d}x \tag{5}$$

式中: S_ξ为ξ的信度熵; U和L分别为ξ的上限和下限。信度熵可以衡量某个不确定变量的不确定程度。信度熵越大, 表明不确定程度越高; 信度熵越小, 表明不确定程度越低。传统的不确定性度量与解析主要是基于方差, 如 Sobol 方法[16-17], 但方差在使用时很难承载某个变量在其整个区间的分布信息, 而且计算量巨大。本文利用信度熵度量和解析评估要素的不确定性, 比方差更具有优势。

2 问题分析与建模

2.1 反舰弹道导弹原理简介

反舰弹道导弹实质是在弹道导弹的基础上实现了反舰功能。弹道导弹依靠火箭助推升空，火箭脱离后以自由落体轨迹高速飞往目标区域。反舰弹道导弹是在弹道导弹的弹头上加装雷达或红外寻的导引头，实现对舰艇的快速搜索和跟踪，在弹道末段控制导弹飞行。反舰弹道导弹是高超音速导弹，射程可达数千千米。反舰弹道导弹的飞行轨迹整体上看是个抛物线轨迹，可以粗略地分为3个阶段，如图1所示。其中，末段速度最快，耗时最短，拦截也最困难。

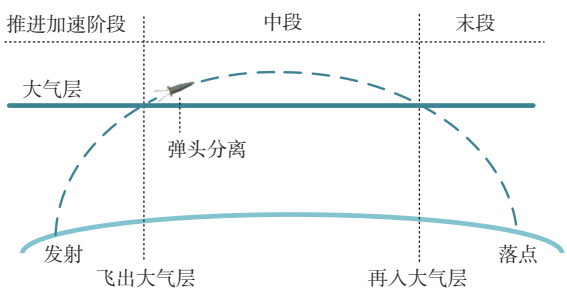


图1 反舰弹道导弹飞行的3个阶段

Fig. 1 Three stages of flight for an anti-ship ballistic missile

普通的弹道导弹由于最后阶段的速度过快难以控制，只能打固定目标。随着空间变轨、末端修正和黑障通信等技术的突破，弹道导弹具备了打击海上移动目标的能力，也就是反舰弹道导弹。反舰弹道导弹先飞出大气层，在达到目标区域后，重返大气层，搜索敌舰并以极高的速度（超过10马赫），从上而下实施灌顶攻击[18]。这种速度如同陨石撞地球，仅靠动能也能重创敌舰的甲板，因此，是远程打击敌人舰艇的利器。

2.2 杀伤效能影响因素分析

结合文献[19]的分析，反舰弹道导弹打击敌人舰艇的作战效能评估（以杀伤敌舰的概率为衡量指标），主要影响因素包括5个不确定变量和1个需要设定的条件变量。

1）目标定位圆概率误差 (circular error probable, CEP)。其指发射导弹时目标实际位置与瞄准位置之间的误差（本文定义目标实际位置与超视距目标瞄准体系提供的位置之间的距离误差小于50%的圆概率误差）。这个误差不仅包含超视距目标瞄准体系对目标的定位误差，还包括识别与定位数据传输至发射单元的时间和导弹飞行时间（从发射到杀伤的时间）所增加的圆概率误差。

2）导弹杀伤半径 R。其指单枚导弹发射前瞄准点与导弹能够发现和定位并机动至目标的最大距离，如图2所示。杀伤（或命中）半径不仅取决于导弹的运动性能，还取决于飞行中的搜索、识别、跟踪和数据链的能力。

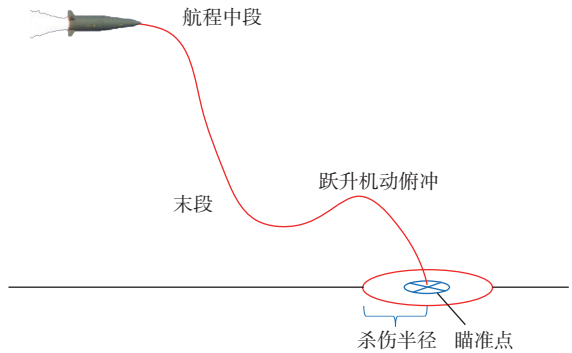


图2 反舰弹道导弹飞行路线及导弹杀伤半径[19]

Fig. 2 Flight path of anti-ship ballistic missiles and missile kill radius[19]

3）导弹杀伤概率 P_k。其指目标在反舰弹道导弹瞄准点杀伤半径范围之内，到达时击中目标的可能性，是性能水平（即速度、可靠性、机动性、传感器）与目标防御水平之间的函数关系，既有运动性也有非运动性因素。

4）导弹的突防概率 P_e。其指反舰弹道导弹躲避敌方防空系统拦截的概率。虽然弹道导弹的速度很高，但也并非拦截不了。据报道，在弹道导弹再次进入大气层之前的中途阶段有机会摧毁来袭的弹道导弹[20]。

5）有效发现目标的概率 P_d。其指超视距目标瞄准体系有效发现、定位敌方舰艇的概率。因为反舰弹道导弹自身不能形成一个完整的杀伤链，需要依托超视距目标瞄准体系提供信息保障，所以需要评估有效发现敌方舰艇的概率，再评估一个完整的“观察-判断-决策-行动”（observe-orient-decide-act, OODA）环下反舰弹道导弹成功杀伤航母的概率。

6）反舰弹道导弹的发射数量 N。该变量在作战中是固定的，在评估工作中根据实际情况设定该变量。

2.3 反舰弹道导弹杀伤效能评估模型

依据2.2节的分析，直接影响杀伤效能不确定性的主要因素包括：发射数量、弹道导弹成功突防概率、目标定位圆概率误差、杀伤概率、杀伤半径、有效发现目标的概率。反舰弹道导弹杀伤效能评估要素的主要关系如图3所示。

反舰弹道导弹杀伤效能评估模型如下：

$$P = f(N, P_d, R, C_{EP}, P_k, P_e) \tag{6}$$

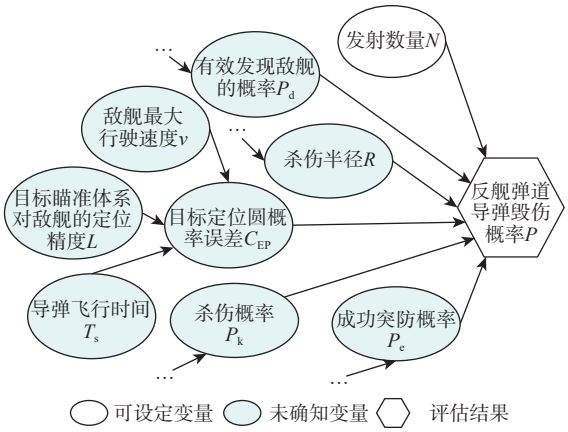


图 3　反舰弹道导弹作战效能影响因素关系

Fig. 3　Relationship of factors influencing kill effectiveness of anti-ship ballistic missiles

式中: P 为 N 枚反舰弹道导弹对敌舰的杀伤概率; $f(\cdot)$表示仿真实验模型。下面给出反舰弹道导弹打击敌舰仿真模型$f(\cdot)$的设计。

文献 [19] 设计了反舰弹道导弹的仿真实验模型, 汇总并分析了弹道导弹在不同参数下对敌舰杀伤概率的仿真计算结果。图 4 为文献 [19] 文献中反舰弹道导弹杀伤半径明显小于目标定位圆概率误差情况下的实验效果, 其中, TLE 表示目标定位偏差。本文对文献 [19] 中的反舰弹道导弹打击舰艇的杀伤模型进行了复现和改进, 加入了有效发现目标概率、突防概率等因素。仿真模型的设计思路如图 5 所示。首先是超视距目标瞄准体系探测目标, 如果发现目标, 则依据目标定位 C_{EP} 计算反舰弹道导弹的瞄准点; 然后, 发射导弹, 依据突防概率判断能否突防; 接着, 计算导弹落点, 并判断目标是否在导弹杀伤半径范围内; 最后, 统计覆盖目标(敌舰)杀伤圆的个数, 设为 t, 按照式 (7) 计算随机一次实验导弹杀伤目标的概率。

$$P_1 = 1-(1-P_k)^t \tag{7}$$

重复足够多次实验, 统计计算反舰弹道导弹对敌舰的杀伤概率, 作为最终反舰弹道导弹杀伤航母的有效概率。如图 6 所示, 10 枚反舰弹道导弹随

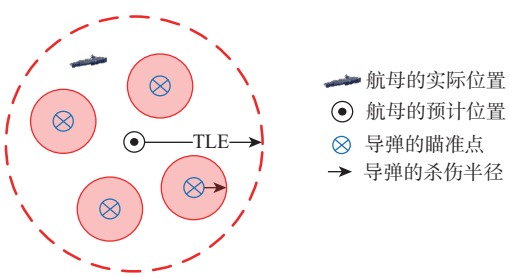


图 4　反舰弹道导弹打击敌舰的示意图[19]

Fig. 4　Illustration of anti-ship ballistic missile striking an enemy vessel[19]

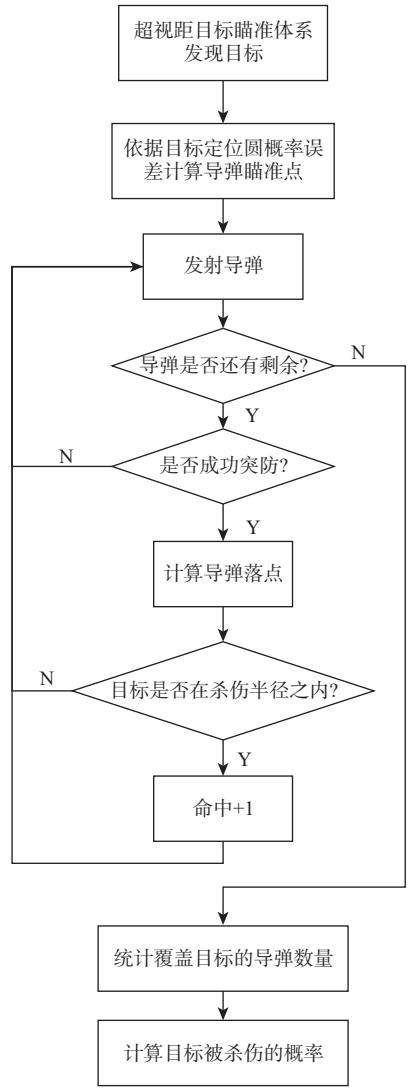


图 5　反舰弹道导弹打击目标仿真模型流程

Fig. 5　Flow of simulation model of anti-ship ballistic missile striking target

机一次打击敌舰随机一次仿真实验的仿真结果。在这次实验中, 各项参数设定为: $N=10$, $R=20$ km, $C_{EP}=40$ km, $P_k=0.6$, $P_e=0.8$, $P_d=0.6$。

由图 6 可知, 发射 10 枚反舰弹道导弹, 有 8 枚突防成功, 1 枚命中目标。该次仿真实验的结果是, 杀伤敌舰的概率为 0.6, 考虑超视距目标瞄准体系对敌舰的有效发现概率, 反舰弹道导弹实际能够杀伤敌舰的概率为 0.36。在上述参数下, 本文进行 5 000 次仿真实验, 得到基于超视距目标瞄准体系的反舰弹道导弹杀伤敌舰的概率为 0.42。这个结果可以作为给定参数下的反舰弹道导弹作战效能的评估结果。

然而, 在实际问题中, R、C_{EP}、P_k、P_e、P_d这些参数往往很难获取到精确数据, 都是不确定的, 这就意味着杀伤效能也是不确定的。下一步对杀伤效能的不确定性进行量化分析。

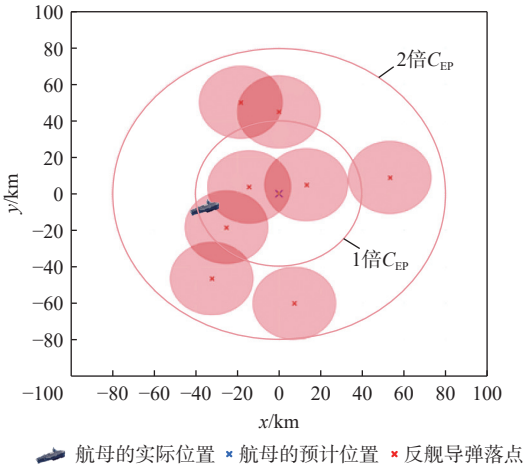


图6 10枚反舰弹道导弹随机打击敌舰的仿真结果

Fig. 6 Simulated results of 10 anti-ship ballistic missiles randomly striking enemy ship

3 不确定性的建模与计算

3.1 参数的信度分布函数

对于杀伤评估模型中不确定变量的不确定性，本文采用信度分布函数进行建模。信度分布函数是在信息不足甚至没有信息时依靠专家的主观判断数据生成的。在生成信度分布函数前，先要对专家进行校准练习[21]，避免专家过于自信或保守；然后，收集专家对该变量取值的判断数据；最后，对离散数据进行分段插值，得到信度分布函数。

在对专家进行校准练习后，以“提问-回答”的交互形式收集R、C_{EP}、P_k、P_e、P_d等参数的评估数据。

例如，对于某型反舰弹道导弹的杀伤半径，由于缺乏准确数据，邀请该型导弹研制方和试验方的专家开会讨论、评估，交互过程如下。

问：请问您认为该导弹的杀伤半径最大有多少km？答：40 km。

问：请问您认为导弹杀伤半径最小有多少km？答：20 km。

问：您认为导弹杀伤半径小于22 km的信度是多少？答：0.1。

问：您认为导弹杀伤半径小于25 km的信度是多少？答：0.22。

问：您认为导弹杀伤半径小于30 km的信度是多少？答：0.5。

…

汇总评估数据。由于每个人知识经验不同，每个人的判断存在差异。可以采用Delphi方法[22]对每个刻度下的评估数据进行收敛、统一，也可以用文献[21]所述的多专家信度集成方法进行集成处理。

设想一个贝塔世界线，红军(A国)反击蓝军(B国)航母战斗群。红军拟发射8枚反舰弹道导弹打击蓝军航母，需要评估8枚反舰弹道导弹对航母的杀伤效能。

先设定条件参数：导弹发射数量8枚，天气状况好(天晴明朗，白日)，敌航母距离海岸线中等距离(1 000～2 000 km)。用C_1和C_2这2个变量分别表示天气状况和敌舰距离，即$N=8$，C_1为好，$C_2=1\,000\sim2\,000$ km。

按照设定的作战概念，邀请多领域的专家进行研讨，给出R、C_{EP}、P_k、P_e、P_d各个参数的信度分布数据，如表1所示。本文主要把信度分为11个刻度进行相应的判断(也可以分成6个或其他数量刻度)。单人对5～9量级不容易混淆，9个刻度以上需要多专家集成，集成方法可参考文献[21]。分别将各个参数小于等于某个值的信度标定为0、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0。例如，C_{EP}小于23 km的信度是0.1，C_{EP}小于26 km的信度是0.3。

表1 R、C_{EP}、P_k、P_e、P_d的信度分布

Table 1 Belief degree distribution of R, C_{EP}, P_k, P_e, P_d

信度	R/km	C_{EP}/km	P_k	P_e	P_d
0	20	20	0	0.09	0
0.1	22	23	0.06	0.79	0.45
0.2	24	26	0.13	0.80	0.48
0.3	26	29	0.22	0.81	0.50
0.4	28	32	0.31	0.82	0.51
0.5	30	35	0.42	0.83	0.53
0.6	32	38	0.54	0.84	0.54
0.7	34	41	0.66	0.85	0.56
0.8	36	44	0.76	0.86	0.57
0.9	38	50	0.88	0.87	0.61
1.0	40	60	1	1	0.88

对表1中的信度分布数据采用三次样条分段插值处理，得到如图7所示的信度分布函数曲线。

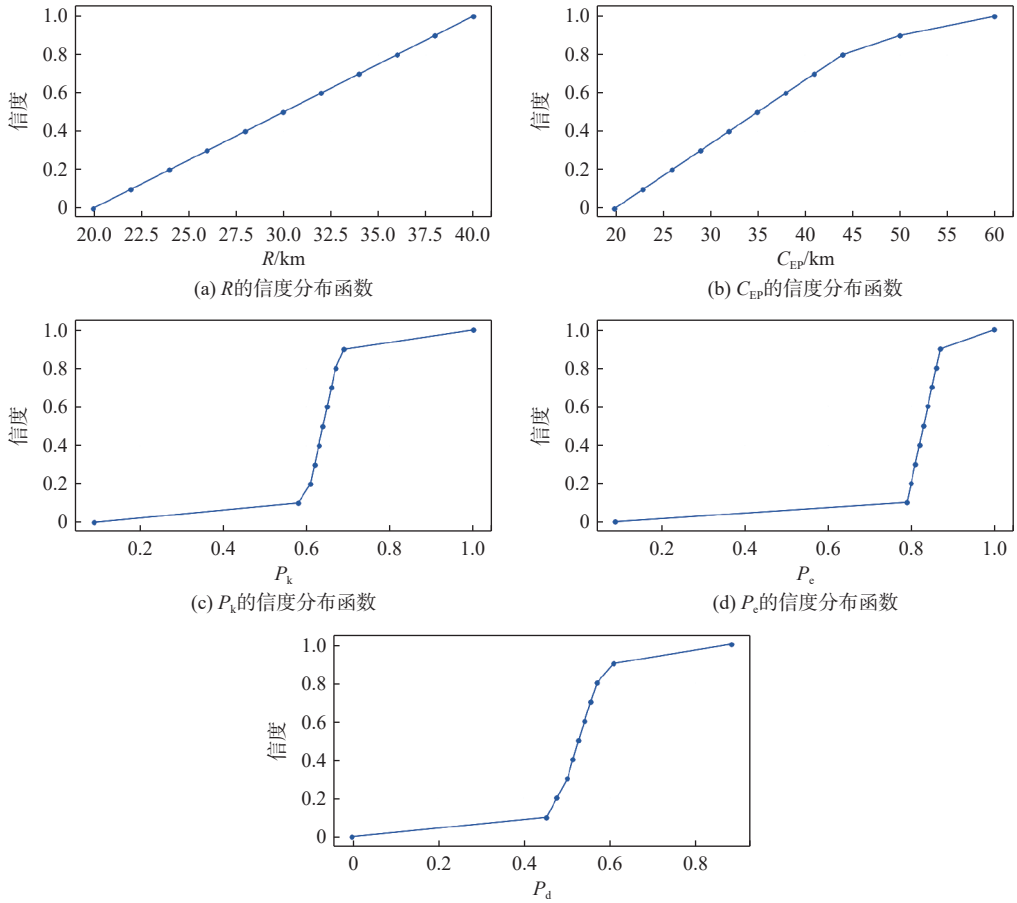
3.2 杀伤效能的信度分布函数

依据定理，不难得出，反舰弹道导弹杀伤效能的信度分布函数的计算式为

$$P=f(N,P_d,R,C_{EP},P_k,P_e)=f(N,\Phi_{P_d}^{-1}(\alpha),\Phi_R^{-1}(\alpha),\Phi_{C_{EP}}^{-1}(1-\alpha),\Phi_{P_k}^{-1}(\alpha),\Phi_{P_e}^{-1}(\alpha)) \tag{8}$$

式中：$f(\cdot)$表示2.3节用仿真手段构造的评估模型。将$N=8$，以及R、C_{EP}、P_k、P_e、P_d各个参数的信度分布函数的逆函数，代入到式(8)中，计算出杀伤效能P的信度分布函数，其函数曲线如图8所示。

由图8可见，反舰弹道导弹有效杀伤航母的概



(a) R的信度分布函数　(b) C_{EP}的信度分布函数

(c) P_k的信度分布函数　(d) P_e的信度分布函数

(e) P_d的信度分布函数

图 7　R、C_{EP}、P_k、P_e、P_d 的信度分布函数

Fig. 7　Belief degree distribution function of R, C_{EP}, P_k, P_e, P_d

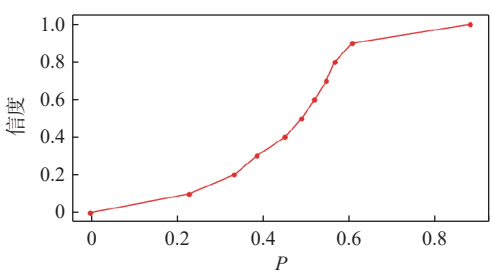


图 8　P 的信度分布函数

Fig. 8　Belief degree distribution function of P

率P在 0.2～0.6 范围内不确定程度很高。显然，这个杀伤效能的评估结果无法支撑决策。依据式 (5) 计算，P的信度熵为 0.44。P的期望值为 0.45。下节讨论如何运用信度熵指导评估活动，迭代渐进减少P的不确定性，从而获取较为准确的 P。

4　迭代减少杀伤效能的不确定性

4.1　不确定性全局贡献率分析

S_P为杀伤效能P的信度熵。对S_P进行分解，分解成各个不确定变量 (R、C_{EP}、P_k、P_e、P_d) 对 S_P的“贡献”。

为推导简便，先令

$$G(\Phi_\xi)=\Phi_\xi\log_2\frac{1}{\Phi_\xi}+(1-\Phi_\xi)\log_2\frac{1}{1-\Phi_\xi}$$

P 的信度熵可分解为

$$S_P=\int_{L_P}^{U_P}G(\Phi(P))\mathrm{d}P=\int_0^1 G(\alpha)(\Phi_P^{-1}(\alpha))'\mathrm{d}\alpha=$$
$$\int_0^1 G(\alpha)\left[\frac{\partial f}{\partial\Phi_{P_d}^{-1}(\alpha)}\Phi_{P_d}^{-1}(\alpha)'+\frac{\partial f}{\partial\Phi_R^{-1}(\alpha)}\Phi_R^{-1}(\alpha)'+\right.$$
$$\frac{\partial f}{\partial\Phi_{C_{EP}}^{-1}(1-\alpha)}\Phi_{C_{EP}}^{-1}(1-\alpha)'+\frac{\partial f}{\partial\Phi_{P_k}^{-1}(\alpha)}\Phi_{P_k}^{-1}(\alpha)'+$$
$$\left.\frac{\partial f}{\partial\Phi_{P_e}^{-1}(\alpha)}\Phi_{P_e}^{-1}(\alpha)'\right]\mathrm{d}\alpha=\int_0^1\frac{\partial f}{\partial\Phi_R^{-1}(\alpha)}G(\alpha)\Phi_R^{-1}(\alpha)'\mathrm{d}\alpha+$$
$$\int_0^1\frac{\partial f}{\partial\Phi_{C_{EP}}^{-1}(1-\alpha)}G(1-\alpha)\Phi_{C_{EP}}^{-1}(1-\alpha)'\mathrm{d}\alpha+\int_0^1\frac{\partial f}{\partial\Phi_{P_k}^{-1}(\alpha)}\cdot$$
$$G(\alpha)\Phi_{P_k}^{-1}(\alpha)'\mathrm{d}\alpha+\int_0^1\frac{\partial f}{\partial\Phi_{P_e}^{-1}(\alpha)}G(\alpha)\Phi_{P_e}^{-1}(\alpha)'\mathrm{d}\alpha+$$
$$\int_0^1\frac{\partial f}{\partial\Phi_{P_d}^{-1}(\alpha)}G(\alpha)\Phi_{P_d}^{-1}(\alpha)'\mathrm{d}\alpha=C_R+C_{C_{EP}}+C_{P_k}+$$
$$C_{P_e}+C_{P_d}\qquad(9)$$

式中：C_R、$C_{C_{EP}}$、C_{P_k}、C_{P_e}、C_{P_d}分别为 R、C_{EP}、P_k、P_e、P_d对P的信度熵的贡献。

经编程计算，R、C_{EP}、P_k、P_e、P_d 对信度熵的贡献分别为 0.083 6, 0.066 0, 0.039 6、0.039 6、0.211 2。归一化后，R、C_{EP}、P_k、P_e、P_d 的贡献率分别为

19%、15%、9%、9%、48%。P_d 贡献率最大，说明超视距目标瞄准体系发现目标的概率的不确定性对反舰弹道导弹有效杀伤航母概率的不确定性影响最大。

4.2 迭代修正杀伤效能评估结果

应当优先减少超视距目标瞄准体系发现目标概率的不确定性。注意在实际问题中，量化的目的就是获得更高的收益，优先减少哪个要素的不确定性要结合投入的成本。

针对超视距目标瞄准体系投入更多的研究，如组织更多的讨论，甚至做更多的试验，摸清超视距目标瞄准体系的实际能力，最终修正了 P_d 的信度分布函数，如图 9 所示。

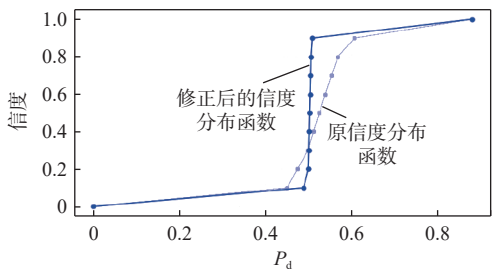


图 9 P_d 修正后的信度分布函数

Fig. 9 Belief degree distribution function of P_d after modification

将 R 重新代入式 (8) 的仿真模型来计算 P，得到修正后的 P 信度分布函数，如图 10 所示。

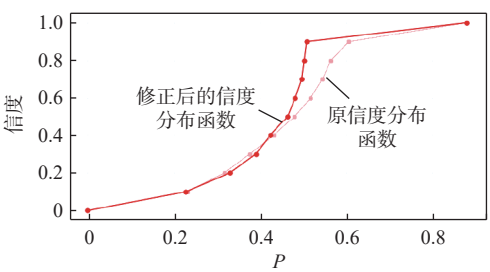


图 10 P 第 1 次修正后的信度分布函数

Fig. 10 Belief degree distribution function of P after the first modification

此时，P 的信度熵由 0.44 降至 0.38，还需要进一步减少。R、C_{EP}、P_k、P_e、P_d 对 P 信度熵的贡献率分别更新为 23%、25%、9%、2%、41%。由于 P_d 的不确定性已经大幅度减少，在实际问题中再减少需要投入巨大成本，只能在属性 R 和 C_{EP} 二者考虑。经研讨分析，减少目标定位圆概率误差 C_{EP} 信度熵的难度和成本相对小些，因此，优先选择属性 C_{EP}。

针对目标定位圆概率误差 C_{EP} 加大研究力度，修正后的 C_{EP} 的信度分布函数，如图 11 所示。

将 C_{EP} 重新代入式 (8) 的仿真模型计算 P，第 2 次修正后的 P 信度分布函数如图 12 所示。

经过 2 次修正，P 的信度熵由 0.44 降至 0.38 再降至 0.29。由图 12 可知，P 在 0.41～0.50 的可能性比较大。经计算，P 此时的期望值为 0.47。如果决策者对杀伤效能的不确定程度满意，则停止减少不确定性的工作；否则，继续迭代减少。

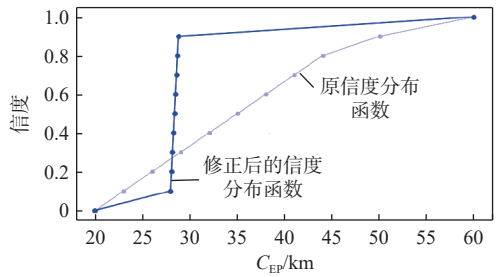


图 11 C_{EP} 修正后的信度分布函数

Fig. 11 Belief degree distribution function of C_{EP} after modification

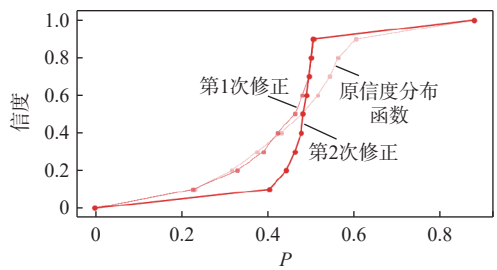


图 12 P 第 2 次修正的信度分布函数

Fig. 12 Bbelief degree distribution function of P after the twice modification

反舰弹道导弹杀伤航母的概率最可能在 0.41～0.50 范围内，这主要受到超视距目标瞄准体系发现目标概率的影响。如果把超视距目标瞄准体系目标概率始终设为 1，即设定在超视距目标体系肯定发现目标的前提下，计算反舰弹道导弹杀伤蓝军航母的概率 P'（即不考虑有效发现航母的概率）。按照相同的步骤调整其信度分布函数如图 13 所示。

P' 为红军 8 枚反舰弹道导弹在有效发现蓝军航母的前提下有效杀伤航母的概率。P' 的信度熵，在 2 次修正下，由 0.55 降至 0.45 再降至 0.31。由图 13 可知，P' 在 0.9～1.0 的信度较大。这个结果意味着，只要超视距目标瞄准体系有效发现目标，红军

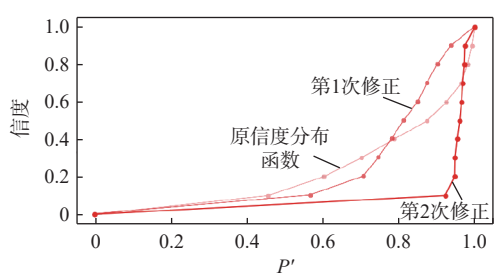


图 13 P' 的 2 次修正的信度分布函数

Fig. 13 Belief degree distribution function of P' after the twice modification

的 8 枚反舰弹道导弹足以杀伤蓝军的航母。

5 结　论

获取较为准确的反舰弹道导弹杀伤效能, 不仅能为设计人员迭代改进反舰弹道导弹提供参考, 而且能为指挥员的作战决策提供支撑。主要结论如下:

1) 使用仿真模型大量模拟, 通过统计手段评估杀伤效能, 结果收敛且可控。但在强对抗和不完备信息环境下, 许多因素不确定, 杀伤效能也就不确定。参数的不确定性用信度分布函数进行表征, 代入仿真模型, 便可计算杀伤效能的信度分布函数, 该信度分布函数可以刻画杀伤效能的不确定性。

2) 使用信度熵能够帮助解析不确定性, 找出关键因素。敏感且不确定的因素对杀伤效能的不确定性全局影响程度最大。

3) 迭代渐进减少反舰弹道导弹杀伤效能不确定性是处理这类问题的复杂性和不确定性的有效手段。

参考文献（References）

[1] XIA Z J, XU L Z, FAN W T. Evaluation of operational effectiveness for TWS based on fuzzy probability[C]//Proceedings of the Sixth International Symposium on Computational Intelligence and Design. Piscataway: IEEE Press, 2013: 424-427.

[2] JIAO J L, REN H L, SUN S Z. Assessment of surface ship environment adaptability in seaways: a fuzzy comprehensive evaluation method[J]. International Journal of Naval Architecture and Ocean Engineering, 2016, 8(4): 418-431.

[3] 田苗苗, 彭进先, 周伦, 等. 基于模糊综合评价法的雷达对抗装备能力评估研究[J]. 现代防御技术, 2025, 53(3): 182-190.
TIAN M M, PENG J X, ZHOU L, et al. Research on the capability evaluation of radar countermeasure equipment based on fuzzy comprehensive model[J]. Modern Defence Technology, 2025, 53(3): 182-190(in Chinese).

[4] 马九方, 杨森, 黄欣鑫. 基于灰色层次分析法的有人/无人协同作战效能评估[J]. 电讯技术, 2023, 63(10): 1625-1630.
MA J F, YANG S, HUANG X X. Effectiveness evaluation of manned-unmanned collaborative combat based on grey analytic hierarchy process[J]. Telecommunication Engineering, 2023, 63(10): 1625-1630(in Chinese).

[5] 孔德鹏, 马溢清, 郑保华, 等. 面向不确定多任务场景的海上联合作战装备体系贡献率评估方法[J]. 系统工程与电子技术, 2022, 44(12): 3775-3782.
KONG D P, MA Y Q, ZHENG B H, et al. Contribution rate assessment method of maritime joint operations equipment system of systems for uncertain multi-mission scenes[J]. Systems Engineering and Electronics, 2022, 44(12): 3775-3782(in Chinese).

[6] 蔡青, 关志军, 赵若言. 基于灰色模糊综合评价法的网络防御作战效能评估[J]. 中国电子科学研究院学报, 2022, 17(10): 991-996.
CAI Q, GUAN Z J, ZHAO R Y. Effectiveness evaluation of cyber defense operation based on grey fuzzy comprehensive evaluation method[J]. Journal of China Academy of Electronics and Information Technology, 2022, 17(10): 991-996(in Chinese).

[7] 孔德鹏, 常天庆, 郝娜, 等. 地面作战目标威胁评估多属性指标处理方法[J]. 自动化学报, 2021, 47(1): 161-172.
KONG D P, CHANG T Q, HAO N, et al. Multi-attribute index processing method of target threat assessment in ground combat[J]. Acta Automatica Sinica, 2021, 47(1): 161-172(in Chinese).

[8] 游雅倩, 姜江, 孙建彬, 等. 基于证据网络的装备体系贡献率评估方法研究[J]. 系统工程与电子技术, 2019, 41(8): 1780-1788.
YOU Y Q, JIANG J, SUN J B, et al. Evidential network-based evaluation method of contribution to weapon system-of-systems[J]. Systems Engineering and Electronics, 2019, 41(8): 1780-1788(in Chinese).

[9] 郑丽莎, 尹东亮, 王旋. 基于改进 D-S 证据理论的相控阵雷达作战效能评估[J]. 系统工程与电子技术, 2024, 46(4): 1330-1336.
ZHENG L S, YIN D L, WANG X. Operational effectiveness evaluation of phased array radar based on improved D-S evidence theory[J]. Systems Engineering and Electronics, 2024, 46(4): 1330-1336(in Chinese).

[10] DER KIUREGHIAN A, DITLEVSEN O. Aleatory or epistemic? does it matter?[J]. Structural Safety, 2009, 31(2): 105-112.

[11] PATÉ-CORNELL M E. What epistemic uncertainty is, and why we need it in risk assessment[J]. Reliability Engineering & System Safety, 2018(180): 117-126.

[12] HOLWERDA J, NEWELL B R. Epistemic and aleatory uncertainty in decisions from experience[C]//Proceedings of the Annual Meeting of the Cognitive Science Society. Vienna: Cognitive Science Society, 2021: 1278-1284.

[13] LIU B D. Uncertainty theory: a branch of mathematics for modeling human uncertainty[M]. Springer: Berlin, 2014.

[14] LIU B D. Why is there a need for uncertainty theory[J]. Journal of Uncertain Systems, 2012, 6(1): 3-10.

[15] WANG Z H, HU J W, XIANG Z N, et al. Belief entropy-based uncertainty analysis[C]//Proceedings of the 3rd International Conference on Computer Science and Management Technology. Piscataway: IEEE Press, 2022: 471-475.

[16] SOBOL I M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates[J]. Mathematics and Computers in Simulation, 2001, 55(1-3): 271-280.

[17] HOMMA T, SALTELLI A. Importance measures in global sensitivity analysis of nonlinear models[J]. Reliability Engineering & System Safety, 1996, 52(1): 1-17.

[18] 隋先辉, 时筱惠, 张小东, 等. 反舰弹道导弹关键技术和作战使用特点研究[J]. 飞航导弹, 2018(5): 29-32.
SUI X H, SHI X H, ZHANG X D, et al. Research on key technologies and operational characteristics of anti-ship ballistic missile[J]. Aerodynamic Missile Journal, 2018(5): 29-32(in Chinese).

[19] HEGINBOTHAM E, NIXON M, HEIM F E, et al. The U. S. -China military scorecard: forces, geography, and the evolving balance of power, 1996—2017[EB/OL]. (2015-03-13)[2023-11-20]. https://www.rand.org/content/dam/rand/pubs/research_reports/RR300/RR392/RAND_RR392.pdf?source=post_page----------.

[20] 易芳, 谢永亮. 美军多招并举应对反舰弹道导弹[J]. 军事文摘,

2018(13): 41-44.
YI F, XIE Y L. The U. S. military responded to anti-ship ballistic missiles with multiple measures[J]. Military Digest, 2018(13): 41-44 (in Chinese).
[21] 胡剑文. 探索性评估论证方法[M]. 北京: 国防工业出版社, 2020.
HU J W. Exploratory decision analysis method [M]. Beijing: National Defense Industry Press, 2020(in Chinese) .
[22] DALKEY N C. Delphi[M]. New York: Routledge, 2018.

Research on uncertainty in kill effectiveness of anti-ship ballistic missiles

WANG Zhihui[1, 2, *], XIANG Zhining[1], GAO Ping[3]

(1. Army Academy of Artillery and Air Defense (Nanjing Campus), Nanjing 211132, China;
2. Joint Warfare College, China People's Liberation Army National Defence University, Beijing 100091, China;
3. Equipment System Evaluation Center, Equipment Development Department, Beijing 100009, China)

Abstract: The focus and challenge of research on the operational application of anti-ship ballistic missiles (ASBMs) is how to achieve a more accurate kill effectiveness in a highly adversarial and information-incomplete environment. A research framework designed to iteratively reduce the uncertainty of ASBM kill effectiveness is proposed. A simulation model for the evaluation of kill efficiency is established. The uncertainty (unknowns) of influencing factors is characterized by belief degree distribution functions. The belief degree distribution function of kill efficiency is computed based on the simulation model. Additionally, belief entropy is used to examine the global impact measure on the indeterminacy of kill efficiency caused by each element. In the example provided, the unknown range of kill probability is reduced from 0.2～0.6 to 0.41～0.5. The proposed method provides a new way of thinking for the evaluation of the operational effectiveness of military equipment.

Keywords: anti-ship ballistic missile; uncertainty; belief degree distribution function; belief degree entropy; effectiveness evaluation

Received: 2023-11-30; **Accepted**: 2024-01-12; **Published Online**: 2024-02-02 10:28
URL: link.cnki.net/urlid/11.2625.V.20240201.1029.003
Foundation item: National Natural Science Foundation of China (61703412)
* **Corresponding author**. E-mail: zhwang66@foxmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0775

双连孔力传感器柔度建模与应变解析

李立建[1,2]，唐守乾[1]，姚建涛[2,3,*]，王迎佳[1]

(1. 华北水利水电大学 机械学院，郑州 450045；2. 燕山大学 河北省并联机器人与机电系统实验室，秦皇岛 066004；
3. 燕山大学 先进锻压成形技术与科学教育部重点实验室，秦皇岛 066004)

摘 要：双连孔力传感器在航空航天和工业计量等领域中应用广泛，但因其结构中存在变截面设计，导致其载荷位移特性和应变载荷关系较难解析描述，进而影响含双连孔结构的多维力传感器设计与性能优化。基于此，基于弹性梁理论，推导得到变截面力敏单元的柔度矩阵和应变载荷关系，进而通过柔度矩阵建模方法构建出双连孔力传感器的整体柔度模型，并以此为桥梁，获得双连孔结构上力敏单元应变片感知应变与传感器测力端作用载荷间的解析映射关系。通过有限元和实验分别对所构建的模型和应变解析关系进行验证。结果表明：期望维解析柔度相对有限元结果误差控制在3%以内，桥路输出电压理论相对于实验结果误差控制在5%以内，说明推导得到的解析公式可用于正确评价双连孔力传感器载荷位移特性和桥路输出应变与测力载荷间的映射，进而为含双连孔结构的多维力传感器优化设计提供可靠的理论技术支持。

关 键 词：力传感器；双连孔结构；柔度建模；应变解析；标定实验
中图分类号：TH73
文献标志码：A **文章编号**：1001-5965(2026)03-0753-10

随着物联网和信息技术的快速发展，智能感知技术迎来新的发展机遇。力传感器作为机械设备或智能机器人与外界环境力交互的关键器件，是其实现精准操作[1]和实时感知外界力觉信息[2]的重要载体。双连孔力传感器作为常见的单维载荷测力单元，因结构紧凑、贴片简易、抗偏载和精度高等优点[3]，在制药、生产制造及食品工业等中小量程称重场合得到广泛应用。

自20世纪70年代，为满足商业电子秤的应用需求，平行梁型传感器结构被提出并大量投入市场应用，后在双连孔结构基础上又演化设计出带辅助梁型、S型和带过载保护型等多种类型的称重传感器[4-5]。刘九卿[6]对平行梁型和S型双连孔力传感器的力学特性、非期望维偏载误差及修正等问题进行分析讨论。双连孔力传感器通过合理减材和贴片组桥设计，使其在测力方向具有较高的测量灵敏度，同时其他维方向上也有较高的抗维间耦合和抗偏载能力，因此，在轮辐式扭矩传感器[7-9]和多维测力传感器[10-12]弹性体设计中双连孔结构备受青睐。Dwarakanath等[13]将双连孔结构作为支链测力单元，提出基于Hexapod机构的并联装配式六维力传感器设计。Wu等[14]利用双连孔结构设计出用于外科手术器械和软组织间力觉交互的小量程解耦型双十字梁结构六维力传感器。Zhang等[15]研制出一款基于双连孔正交串联支链结构和光纤光栅检测原理的十字梁型六维牵引力传感器，用于人机协作时末端执行器和被操纵对象间的接触力感知。Mayetin等[16]通过将多个双连孔力传感器进行组合设计，提出一种用于腕关节康复机器人的三维力传感器。然而，由于双连孔结构中存在变截面设计，导致其力学特性和测量性能较难描述，尤其在多维测力传感器设计中，结构参量涉及较多，传感

收稿日期：2023-11-30；**录用日期**：2024-01-05；**网络出版时间**：2024-01-18 11:22
网络出版地址：link.cnki.net/urlid/11.2625.V.20240117.1645.001
基金项目：国家自然科学基金(52005181,U2037202)；河南省科技攻关项目(252102220107,222102220039)；河南省高等学校重点科研项目计划(24A460016)；河南省高等学校青年骨干教师培养计划项目；华北水利水电大学青年骨干教师培养计划项目
* **通信作者**. E-mail：jtyao@ysu.edu.cn

引用格式：李立建，唐守乾，姚建涛，等. 双连孔力传感器柔度建模与应变解析[J]. 北京航空航天大学学报，2026，52(3)：753-762.
LI L J，TANG S Q，YAO J T，et al. Compliance modeling and strain analysis of double-hole force sensor[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：753-762 (in Chinese).

器性能与其结构参数间呈现明显的非线性耦合关联[17]，在对传感器结构及量程有所限制时，优化设计的初始模型和设计变量往往依据经验或借助有限元分析进行设计和选择，会耗费大量时间和精力，故而双连孔力传感器解析模型的构建将会为含双连孔结构的多维测力传感器性能建模及优化提供有益指导，进而加快多维力传感器设计和研发进程。

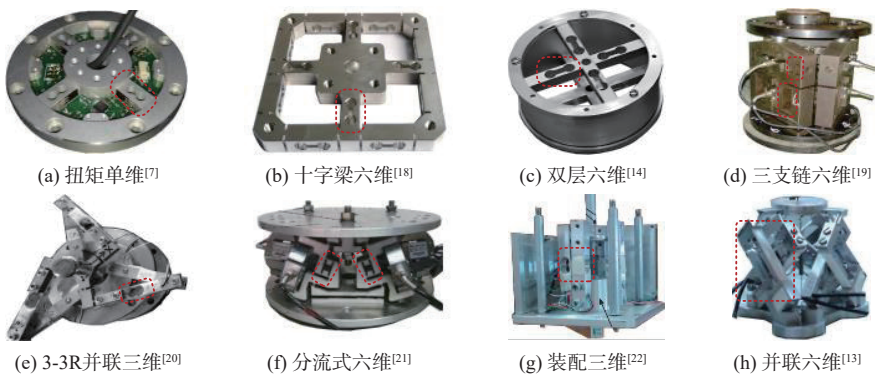
针对双连孔力传感器非规则变截面结构特点，在小变形假设框架下，本文将其视为柔性并联机构，基于弹性梁理论推导得到了其力敏单元柔度矩阵、弹性体整体柔度矩阵和力敏单元应变-载荷解析映射关系，从有限元仿真和实验两方面对本文模型的预测精度进行了验证，同时研究了传感器结构参数和其他维载荷对测量桥路输出应变的影响。

1 双连孔结构的应用与结构特点

1.1 力传感器中的应用

双连孔结构作为铰链四杆机构的典型代表，以其作为主要测力单元的扭矩传感器和多维力传感器在人机协作、碰撞检测和康复机器人等智能感知领域发挥着重要作用，其在扭矩及多维力传感器中的相关典型应用如图 1 所示。其中，图 1(d) 和图 1(f) 为双连孔结构作为弹性运动单元在六维力传感器中的应用，其余分别为双连孔结构作为力敏单元在扭矩和多维力传感器中的应用。多维测力传感器因其涉及参变量较多，且含双连孔结构后其传感器性能与参数间的关系呈明显的非线性关联，这对传感器建模和性能优化带来很大挑战，同时也是现有关于含双连孔结构的多维测力传感器相关研究中，含较少传感器性能解析建模环节的重要原因所在[7-9,16]，而大多采用响应面优化方法对力觉传感器弹性体进行优化设计。

响应面优化方法可有效减低实验次数，构建力觉传感器弹性体各设计变量与性能指标间的多元二次回归响应面模型，进而利用所拟合获得的响应面模型进行传感器性能预测和优化设计，其优势在于针对复杂难解析弹性体结构可快速构建用于性能预测和评估的数学模型，获取传感器性能指标与其结构参数设计域内的关联影响。然而在进行多维力觉传感器优化设计时，由于涉及的参数变量较多，尝试构建不同性能指标的响应面模型势必将耗费大量时间和精力，对设计者和软硬件资源均有较高要求，设计变量对性能的敏感度及设计域选取大多依赖设计者经验，且所构建的响应面模型仅对在设计域内的结构参数有效，当最优解不在设计域或设计域变更时，模型预测精度将受到显著影响。



(a) 扭矩单维[7] (b) 十字梁六维[18] (c) 双层六维[14] (d) 三支链六维[19]

(e) 3-3R并联三维[20] (f) 分流式六维[21] (g) 装配三维[22] (h) 并联六维[13]

图 1 双连孔结构的典型应用

Fig. 1 Typical applications of double-hole structure

与此同时，响应面优化方法对仿真软件的依赖性较强，其响应面拟合所用数据集需要依靠足够的仿真数据做支撑。从另一个层面来讲，基于响应面优化的力觉传感器设计所依赖的响应面模型在功能上可等效于基于弹性梁理论而获得的力学或数学模型，故而响应面模型在传感器弹性体结构复杂难于进行性能建模时经常被使用。因此，通过基于弹性梁理论构建难解析双连孔结构刚度模型和应变-载荷关系，不仅可在全设计域内快速进行参数对传感器性能的敏度和影响分析，而且将所构建的数学模型作为子结构单元融合进多维力传感器整体性能建模当中，进而显著降低含双连孔结构的力觉传感器建模难度，更快捷有效地提升多维力传感器的参数寻优效率。

为进一步探究双连孔结构的结构特征、力学机理和载荷传递机制，以双连孔力传感器为研究对象，综合考虑其结构特点、应变片布置位置和支链并联组合特征，解析建模其结构，必将显著缩短多维测力传感器的研发设计及优化进程，进而为获取高性能多维力觉传感器提供技术支持。

1.2 结构特点

双连孔力传感器为悬臂梁结构，其弹性体结构如图 2 所示，是由固定端、测力端和双连孔结构经串联组合而成，w 表示传感器的宽度，其中，双连孔

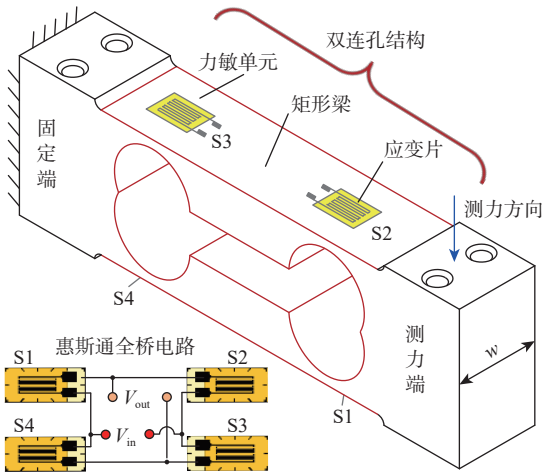


图 2　弹性体结构及惠斯通桥路

Fig. 2　Elastic-body structure and Wheatstone bridge

结构上下双梁采用变截面设计, 这种设计在保留弹性体整体刚度的同时, 兼顾优化传感器精度、输出灵敏度和滞后等性能指标。

双连孔结构为并联式构型, 其上下两变截面梁采用对称设计, 其中, 变截面梁可由 2 个圆弧型力敏单元和矩形梁经串联组合而成。力敏单元从运动机理上可等效为非对称变截面柔性转动副, 其平面侧最小槽口处可布置应变片以感知其上因载荷作用产生的正向线应变。通过运动等效, 传感器弹性体可看作是由 4 个柔性转动副分别经矩形梁连接而形成的平行四杆机构, 故测力端相对于固定端在载荷作用下可沿测力方向上产生微小移动, 而非期望方向则具有较强的抗偏载能力。

双连孔结构共包含 4 个力敏单元, 因此, 平面侧对称粘贴有 4 个应变片 S1～S4, 分别布置在惠斯通全桥电路的 4 个桥臂上, 传感器受测力方向载荷作用时, 应变片 S1 和 S3 产生拉应变, 应变片 S2 和 S4 产生压应变, 该传感器的惠斯通桥路输出应变值和输出灵敏度可表示为

$$\frac{V_{out}}{V_{in}}=K\varepsilon_{out}=\frac{1}{4}K(\varepsilon_{S1}-\varepsilon_{S2}+\varepsilon_{S3}-\varepsilon_{S4}) \tag{1}$$

式中: V_{in}和V_{out}分别为惠斯通桥路的输入和输出电压; K 为应变片灵敏系数; ε_{S1}～ε_{S4} 分别为应变片 S1～S4 处的微应变; ε_{out} 为惠斯通桥路的输出应变。

2　双连孔结构刚度建模

2.1　力敏单元柔度建模

力敏单元属于变截面圆弧型单轴柔性铰链, 而柔性铰链又因其结构紧凑、运动精度高、无摩擦等特点, 在精密微动台、大口径主镜及望远镜并联调整机构等[23-25]领域应用广泛。力敏单元正向视图及结构参数如图 3 所示, 在下侧矩形梁左端面几何中心处建立总体坐标系 $Oxyz$, z 轴由右手定则确

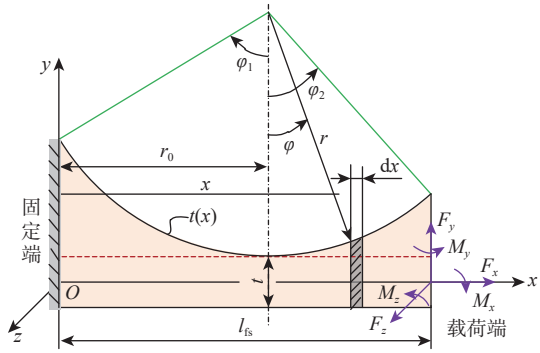


图 3　力敏单元正向视图及结构参数

Fig. 3　View and structural parameters of force-sensitive element

定。力敏单元的总长为 l_{fs}, 最薄处厚度为 t, 圆弧型槽口的半径为 r, r_0 为槽口圆弧圆心到 y 轴的横向距离, $\boldsymbol{F}=[F_x, F_y, F_z, M_x, M_y, M_z]^{\mathrm{T}}$ 为施加于力敏单元自由端处的作用载荷。

在 xOy 平面内, 非对称椭圆弧段的变厚度函数 $t(x)$ 可表示为

$$t(x)=t+r-[r^2-(x-r_0)^2]^{1/2}\qquad 0<x\leqslant l_{fs} \tag{2}$$

令 $x=r_0+r\sin\varphi$, 则 $\mathrm{d}x=r\cos\varphi\mathrm{d}\varphi$ 且 $\varphi\in[-\varphi_1,\varphi_2]$, 其最大离心角 $\varphi_1=\arcsin(r_0/r)$、$\varphi_2=\arcsin[(l_{fs}-r_0)/r]$, 变量 $\xi=t/r$, 由此 $t(x)$ 可进一步转化为

$$t(\varphi)=r(1+\xi-\cos\varphi) \tag{3}$$

基于小变形假设和卡氏第二定理[26-27], 可得在末端受拉/压、弯曲、扭转作用下, 变截面力敏单元空间柔度矩阵 $\boldsymbol{C}_{fs}$ 的所有非零柔度项分别为

$$\begin{cases}C_{fs,11}=w^{-1}N_1/E\\ C_{fs,22}=12w^{-1}N_6/E+\alpha_s w^{-1}N_1/G\\ C_{fs,26}=C_{fs,62}=12w^{-1}N_5/E\\ C_{fs,33}=12w^{-3}N_3/E+\alpha_s w^{-1}N_1/G\\ C_{fs,35}=C_{fs,53}=12w^{-3}N_2/E\\ C_{fs,44}=3.5f(\eta)w^{-3}(N_1+w^2N_4)/G\\ C_{fs,55}=12w^{-3}N_1/E\\ C_{fs,66}=12w^{-1}N_4/E\end{cases} \tag{4}$$

式中: 柔度元素下标中右侧的 2 个数字分别表示元素所在柔度矩阵的行和列; E、$G=E/[2(1+\mu)]$ 和 μ 分别为材料的弹性模量、剪切模量和泊松比; 矩形截面的剪切系数可用 $\alpha_s=(12+11\mu)/[10(1+\mu)]$ 计得; $f(\eta)=(\eta^2+2.609\eta+1)/(1.17\eta^2+2.191\eta+1.17)$[28], 其中, $\eta=t/w$; N_1～N_6 为与力敏单元槽口形状函数有关的定积分, 分别为

$$\begin{cases}N_1=I_1(\varphi)\big|_{-\varphi_1}^{\varphi_2}\\ N_2=-r(\sin\varphi_2 I_1(\varphi)-I_2(\varphi))\big|_{-\varphi_1}^{\varphi_2}\\ N_3=r^2(\sin^2\varphi_2 I_1(\varphi)-2\sin\varphi_2 I_2(\varphi)+I_3(\varphi))\big|_{-\varphi_1}^{\varphi_2}\\ N_4=r^{-2}I_4(\varphi)\big|_{-\varphi_1}^{\varphi_2}\\ N_5=r^{-1}(\sin\varphi_2 I_4(\varphi)-I_5(\varphi))\big|_{-\varphi_1}^{\varphi_2}\\ N_6=(\sin^2\varphi_2 I_4(\varphi)-2\sin\varphi_2 I_5(\varphi)+I_6(\varphi))\big|_{-\varphi_1}^{\varphi_2}\end{cases} \tag{5}$$

其中：关于离心角φ的公式$I_1(\varphi)$～$I_6(\varphi)$分别为

$$I_1(\varphi)=(1+\xi)g(\varphi)-\varphi \tag{6a}$$

$$I_2(\varphi)=(1+\xi)\ln(1+\xi-\cos\varphi)/\xi+\cos\varphi \tag{6b}$$

$$I_3(\varphi)=(\xi^2+2\xi+1/2)\varphi-\xi(\xi+1)(\xi+2)g(\varphi)+(1+\xi+\cos\varphi/2)\sin\varphi \tag{6c}$$

$$I_4(\varphi)=\left\{\tan(\varphi/2)(1+\cos\varphi)^2/(1+\xi-\cos\varphi)^2\cdot\left[\left(2\xi^3+7\xi^2+9\xi+6\right)\tan^2(\varphi/2)+2\xi^3+5\xi^2+5\xi\right]+6(1+\xi)g(\varphi)\right\}/\left[4\xi^2(\xi+2)^2\right] \tag{6d}$$

$$I_5(\varphi)=(1+\xi-2\cos\varphi)/\left[2(1+\xi-\cos\varphi)^2\right]+(1-\xi)/\left(2\xi^2\right) \tag{6e}$$

$$I_6(\varphi)=(1+\cos\varphi)^2\tan(\varphi/2)/\left[4(1+\xi-\cos\varphi)^2\right]\cdot\left[(2\xi+5+1/\xi)\tan^2(\varphi/2)+\left(2\xi^2+3\xi-1\right)/(\xi+2)\right]-(\xi+1)\left(\xi^2+2\xi-1/2\right)g/[\xi(\xi+2)]+\varphi \tag{6f}$$

其中：$g(\varphi)=2\arctan\{[(\xi+2)/\xi)]^{1/2}\tan(\varphi/2)\}/[\xi(\xi+2)]^{1/2}$。

联立式(4)～式(6)，便可解析获得变截面圆弧型力敏单元的柔度矩阵。对于力敏单元经镜像得到的力敏单元结构，其柔度矩阵仅需将式(5)中定积分上下限中的φ_1和φ_2相互调换位置。当然也可利用镜像混合型柔性单元的柔度变换公式，得到镜像两力敏单元间的柔度映射关系为

$$\boldsymbol{C}_{\mathrm{m,fs}}=\boldsymbol{T}\boldsymbol{C}_{\mathrm{fs}}\boldsymbol{T}^{\mathrm{T}} \tag{7}$$

式中：$\boldsymbol{C}_{\mathrm{m,fs}}$为镜像段力敏单元结构的柔度矩阵；$\boldsymbol{T}\in\mathbf{R}^{6\times6}$为镜像混合型两力敏单元间的映射矩阵，其形式为

$$\boldsymbol{T}=\left[\begin{array}{ccc:ccc}1&0&0&0&0&0\\0&-1&0&0&0&l_{\mathrm{fs}}\\0&0&-1&0&-l_{\mathrm{fs}}&0\\\hdashline\multicolumn{3}{c:}{\boldsymbol{O}_{3\times3}}&\multicolumn{3}{c}{\boldsymbol{I}_{3\times3}}\end{array}\right] \tag{8}$$

其中：$\boldsymbol{O}_{3\times3}$和$\boldsymbol{I}_{3\times3}$分别为三阶的零方阵和单位阵。

2.2 双连孔结构所含支链柔度建模

双连孔结构两变截面梁结构对称，均可等效为由力敏单元I、连接梁和力敏单元II首尾连接组合而成的柔性串联支链，简称RTR支链，如图4所示，其中，参考坐标系$\{O_{if}\}$:$O_{if}x_{if}y_{if}z_{if}$建立在力敏单元II右端面处，局部坐标系$\{O_{ij}\}$:$O_{ij}x_{ij}y_{ij}z_{ij}$(i=1,2和j=1,2,3)分别建立在力敏单元I、连接梁和力敏单元II右端面处，x_{i1}和x_{i3}轴共线，$\{O_{i3}\}$和$\{O_{if}\}$系对应坐标轴重合，$t_{\mathrm{b}}=r(1-\cos\varphi_2)+t$和$l_{\mathrm{b}}$分别为连接梁的厚度和长度。

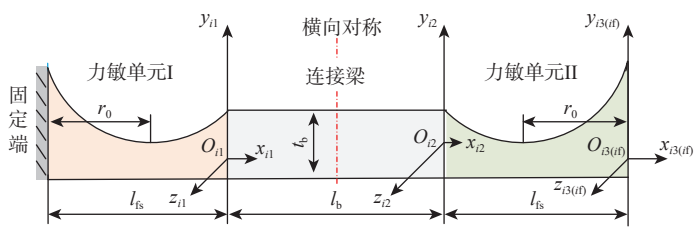


图4 柔性RTR串联支链坐标系建立示意图

Fig. 4 Coordinate frames of flexible RTR serial chain

基于虚功原理和变形叠加原理[29]，柔性RTR串联支链的柔度矩阵在$\{O_{if}\}$系下表示为

$$\boldsymbol{C}_{i,\mathrm{RTR}}=\boldsymbol{J}_{i1}\boldsymbol{C}_{\mathrm{fs}}\boldsymbol{J}_{i1}^{\mathrm{T}}+\boldsymbol{J}_{i2}\boldsymbol{C}_{\mathrm{b}}\boldsymbol{J}_{i2}^{\mathrm{T}}+\boldsymbol{J}_{i3}\boldsymbol{C}_{\mathrm{m,fs}}\boldsymbol{J}_{i3}^{\mathrm{T}} \tag{9}$$

式中：$\boldsymbol{J}_{ij}$为坐标变换矩阵，其形式为

$$\begin{cases}\boldsymbol{J}_{ij}=\begin{bmatrix}\boldsymbol{I}_{3\times3}&-\boldsymbol{S}(\boldsymbol{r}_{ij})\\\boldsymbol{O}_{3\times3}&\boldsymbol{I}_{3\times3}\end{bmatrix}\\\boldsymbol{S}(\boldsymbol{r}_{ij})=\begin{bmatrix}0&-r_{ij,z}&r_{ij,y}\\r_{ij,z}&0&-r_{ij,x}\\-r_{ij,y}&r_{ij,x}&0\end{bmatrix}\end{cases} \tag{10}$$

其中：$\boldsymbol{S}(\cdot)$为反对称算子；$\boldsymbol{r}_{ij}$为$\{O_{if}\}$系坐标原点O_{if}在$\{O_{ij}\}$系下的表示，分别为

$$\begin{cases}\boldsymbol{r}_{i1}=[l_{\mathrm{fs}}+l_{\mathrm{b}}\quad0\quad0]^{\mathrm{T}}\\\boldsymbol{r}_{i2}=[l_{\mathrm{fs}}\quad(t-t_b)/2\quad0]^{\mathrm{T}}\\\boldsymbol{r}_{i3}=[0\quad0\quad0]^{\mathrm{T}}\end{cases} \tag{11}$$

由于连接梁的厚度函数为定值，因而其形状函数$t_{\mathrm{b}}(x)=t_{\mathrm{b}}$，根据式(5)可得到其关于厚度函数的定积分分别为

$$\begin{cases}N_{\mathrm{b},1}=l_{\mathrm{b}}t_{\mathrm{b}}^{-1},N_{\mathrm{b},2}=-l_{\mathrm{b}}^2t_{\mathrm{b}}^{-1}/2,N_{\mathrm{b},3}=l_{\mathrm{b}}^3t_{\mathrm{b}}^{-1}/3\\N_{\mathrm{b},4}=l_{\mathrm{b}}t_{\mathrm{b}}^{-3},N_{\mathrm{b},5}=l_{\mathrm{b}}^2t_{\mathrm{b}}^{-3}/2,N_{\mathrm{b},6}=l_{\mathrm{b}}^3t_{\mathrm{b}}^{-3}/3\end{cases} \tag{12}$$

将式(12)代入式(4)，便可得到连接梁柔度矩阵$\boldsymbol{C}_{\mathrm{b}}$所有非零元素。联立式(4)～式(12)，即可获得双连孔结构的2条柔性RTR串联支链的柔度矩阵。

2.3 双连孔结构整体刚度建模

双连孔结构由2条柔性RTR支链并联组成，其坐标系建立如图5所示。$\{O_{\mathrm{d}}\}$:$O_{\mathrm{d}}x_{\mathrm{d}}y_{\mathrm{d}}z_{\mathrm{d}}$为建立在双连孔结构右端面几何中心处的参考坐标系，其x_{d}和y_{d}轴与两柔性RTR串联支链$\{O_{ij}\}$及$\{O_{if}\}$系的$x_{ij(if)}$和$y_{ij(if)}$轴所在平面共面，各坐标系的z轴可根据右手法则确定，h为双连孔结构的厚度。

基于变形协调条件[29]，并列写出双连孔结构右端面的静力平衡方程，经整理可得双连孔结构在$\{O_{\mathrm{d}}\}$系下表示的整体刚度矩阵为

$$\boldsymbol{K}_{\mathrm{d}}=\boldsymbol{J}_{\mathrm{1f\text{-}d}}^{-\mathrm{T}}\boldsymbol{C}_{1,\mathrm{RTR}}^{-1}\boldsymbol{J}_{\mathrm{1f\text{-}d}}^{-1}+\boldsymbol{J}_{\mathrm{2f\text{-}d}}^{-\mathrm{T}}\boldsymbol{C}_{2,\mathrm{RTR}}^{-1}\boldsymbol{J}_{\mathrm{2f\text{-}d}}^{-1} \tag{13}$$

式中：位移变换矩阵$\boldsymbol{J}_{\mathrm{if\text{-}d}}$为

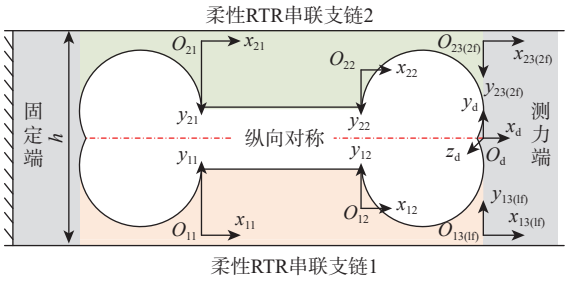


图 5　双连孔结构坐标系建立示意图

Fig. 5　Coordinate frames of double-hole structure

$$\boldsymbol{J}_{i\text{f-d}}=\begin{bmatrix}\boldsymbol{R}_{i\text{f-d}} & \boldsymbol{S}(\boldsymbol{r}_{\text{d-}i\text{f}})\boldsymbol{R}_{i\text{f-d}}\\ \boldsymbol{O}_{3\times3} & \boldsymbol{R}_{i\text{f-d}}\end{bmatrix}\qquad i=1,2 \tag{14}$$

其中: $\boldsymbol{R}_{i\text{f-d}}$ 为 $\{O_{i\text{f}}\}$系相对于$\{O_{\text{d}}\}$系的姿态矩阵; $\boldsymbol{r}_{\text{d-}i\text{f}}$ 为$\{O_{i\text{f}}\}$系坐标原点 $O_{i\text{f}}$ 在$\{O_{\text{d}}\}$系下的表示, 分别为

$$\begin{cases}\boldsymbol{r}_{\text{d-2f}}=[0\quad (h-t)/2\quad 0]^{\text{T}}\\ \boldsymbol{r}_{\text{d-1f}}=-\boldsymbol{r}_{\text{d-2f}}\end{cases} \tag{15}$$

联立式 (9) 和式 (13)～式 (15) 便可解析获得双连孔结构的刚度矩阵。

3　传感器柔度建模与应变解析

3.1　弹性体结构柔度建模

双连孔力传感器由固定端、双连孔结构和测力端串联组成, 坐标系建立及待测载荷如图 6 所示, 参考坐标系 $\{O_{\text{f}}\}$:$O_{\text{f}}x_{\text{f}}y_{\text{f}}z_{\text{f}}$ 建立在测力端待测载荷 $\boldsymbol{F}=[F_x\ F_y\ F_z\ M_x\ M_y\ M_z]^{\text{T}}$ 中心处, 坐标系 $\{O_{\text{d}}\}$的 x_{d} 和 y_{d} 轴与参考坐标系 $\{O_{\text{f}}\}$的 x_{f} 轴和 y_{f} 轴所在平面共面, 且 x_{d} 轴和 x_{f} 轴共线, z 轴可根据右手法则确定。双连孔结构右端到坐标系 $\{O_{\text{f}}\}$的横向距离为 l。

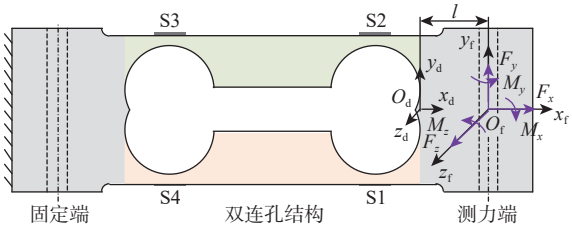


图 6　传感器坐标系建立及待测载荷

Fig. 6　Coordinate frames and measured loads of sensor

将固定端和测力端等效为刚体, 类似式 (9) 可得双连孔力传感器在 $\{O_{\text{f}}\}$系下的柔度矩阵 $\boldsymbol{C}$ 为

$$\begin{cases}\boldsymbol{C}=\boldsymbol{J}_{\text{d-f}}\boldsymbol{K}_{\text{d}}^{-1}\boldsymbol{J}_{\text{d-f}}^{\text{T}}\\ \boldsymbol{J}_{\text{d-f}}=\begin{bmatrix}\boldsymbol{I}_{3\times3} & -\boldsymbol{S}(\boldsymbol{r}_{\text{d-f}})\\ \boldsymbol{O}_{3\times3} & \boldsymbol{I}_{3\times3}\end{bmatrix}\end{cases} \tag{16}$$

式中: $\boldsymbol{r}_{\text{d-f}}=[l,0,0]^{\text{T}}$ 为坐标原点 O_{f} 在$\{O_{\text{d}}\}$下的表示。

3.2　力敏单元应变-载荷关系

圆弧型力敏单元最小槽口处平面侧居中对称位置为敏感区域, 用于粘贴应变片, 如图 7 所示, 其中,

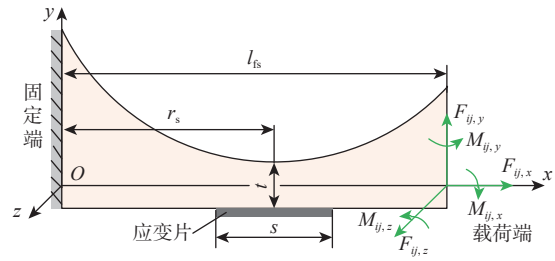


图 7　应变片位置及待测载荷

Fig. 7　Strain-gage location and measured loads

应变片敏感栅的栅丝长度为 s, 应变片编号 k=1～4, 规定表面拉应变为正, 压应变为负。

圆弧转动梁下表面沿轴向布置的应变片所产生的总应变 $\varepsilon_{\text{S}k}$ 与载荷端作用的六维外载荷 $\boldsymbol{F}_{ij}$ 间的关系为

$$\varepsilon_{\text{S}k}=\boldsymbol{C}_{\varepsilon}\boldsymbol{F}_{ij}=[C_{\varepsilon1}\quad C_{\varepsilon2}\quad 0\quad 0\quad 0\quad C_{\varepsilon3}]\boldsymbol{F}_{ij} \tag{17}$$

式中: $\boldsymbol{C}_{\varepsilon}$ 为应变柔度向量, 与力作用的圆弧段有关, $C_{\varepsilon1}$～$C_{\varepsilon3}$ 为应变柔度因子, 利用下式计算得到

$$\begin{cases}C_{\varepsilon1}=\dfrac{1}{Ews}\displaystyle\int_{r_{\text{s}}-s/2}^{r_{\text{s}}+s/2}\frac{\text{d}x}{t(x)}\\ C_{\varepsilon2}=\dfrac{6}{Ews}\displaystyle\int_{r_{\text{s}}-s/2}^{r_{\text{s}}+s/2}\frac{(l_{\text{fs}}-x)\text{d}x}{t^2(x)}\\ C_{\varepsilon3}=\dfrac{6}{Ews}\displaystyle\int_{r_{\text{s}}-s/2}^{r_{\text{s}}+s/2}\frac{\text{d}x}{t^2(x)}\end{cases} \tag{18}$$

将 $x=r_{\text{s}}+r\sin\varphi$、$\text{d}x=r\cos\varphi\text{d}\varphi$、$\varphi_{\text{s}}=\arcsin(s/2r)$ 和式 (3) 代入式 (18) 并经整理可进一步得到

$$\begin{cases}C_{\varepsilon1}=\dfrac{2\lambda_1}{Ews}\\ C_{\varepsilon2}=\dfrac{12(l_{\text{fs}}-r_{\text{s}})\lambda_2}{Ewrs}-\dfrac{12\lambda_3}{Ews}\\ C_{\varepsilon3}=\dfrac{12\lambda_2}{Ewrs}\end{cases} \tag{19}$$

式中: 关于离心角φ的定积分 λ_1～λ_3分别为

$$\begin{cases}\lambda_1=\displaystyle\int_0^{\varphi_{\text{s}}}\cos\varphi(1+\xi-\cos\varphi)^{-1}\text{d}\varphi=I_1(\varphi_{\text{s}})\\ \lambda_2=\displaystyle\int_0^{\varphi_{\text{s}}}\cos\varphi(1+\xi-\cos\varphi)^{-2}\text{d}\varphi=I_7(\varphi_{\text{s}})\\ \lambda_3=\displaystyle\int_0^{\varphi_{\text{s}}}\sin\varphi\cos\varphi(1+\xi-\cos\varphi)^{-2}\text{d}\varphi=I_8(\varphi_{\text{s}})\end{cases} \tag{20}$$

其中: $I_1(\varphi_{\text{s}})$ 可由式 (6) 求得, $I_7(\varphi_{\text{s}})$ 和 $I_8(\varphi_{\text{s}})$ 分别为

$$\begin{cases}I_7(\varphi_{\text{s}})=(2\xi+2)\tan(\varphi_{\text{s}}/2)/\{(\xi^2+2\xi)[\xi\tan^2(\varphi_{\text{s}}/2)+\\ \qquad 2\tan^2(\varphi_{\text{s}}/2)+\xi]\}+g(\varphi_{\text{s}})/(\xi^2+2\xi)\\ I_8(\varphi_{\text{s}})=(\xi+1)/(\cos\varphi_{\text{s}}-\xi-1)-\ln(\xi+1-\cos\varphi_{\text{s}})+\\ \qquad (\xi+1)/\xi+\ln\xi\end{cases} \tag{21}$$

值得注意的是, 对于靠近固定端的两力敏单元 r_{s} 取 r_0, 而对于靠近测力端的两力敏单元 r_{s} 则取 $l_{\text{fs}}-r_0$。

3.3　传感器应变解析

根据图 6 位置关系和力的平移变换, 可得双连孔结构右端面作用载荷与待测载荷间的关系为

$$\boldsymbol{F}_{\mathrm{d}}=\boldsymbol{J}_{\mathrm{d\text{-}f}}^{\mathrm{T}}\boldsymbol{F} \tag{22}$$

依据图 5, 基于测力端的变形协调条件和利用双连孔结构载荷-位移关系, 可进一步获得柔性 RTR 串联支链末端位移与双连孔结构右端面作用载荷间关系为

$$\boldsymbol{X}_{\mathrm{if}}=\boldsymbol{J}_{\mathrm{if\text{-}d}}^{-1}\boldsymbol{X}_{\mathrm{d}}=\boldsymbol{J}_{\mathrm{if\text{-}d}}^{-1}\boldsymbol{K}_{\mathrm{d}}^{-1}\boldsymbol{F}_{\mathrm{d}} \tag{23}$$

基于虚功原理和柔性 RTR 支链载荷-位移关系, 并根据图 4 可得到柔性 RTR 串联支链上两柔性单元作用载荷与其末端位移间的关系为

$$\boldsymbol{F}_{ij}=\boldsymbol{J}_{ij}^{\mathrm{T}}\boldsymbol{F}_{\mathrm{if}}=\boldsymbol{J}_{ij}^{\mathrm{T}}\boldsymbol{C}_{i,\mathrm{RTR}}^{-1}\boldsymbol{X}_{\mathrm{if}}\qquad j=1,3 \tag{24}$$

联立式 (17) 和式 (22)～式 (24), 可最终获得各力敏单元应变片感知应变与六维待测载荷间的解析映射关系为

$$\varepsilon_{sk}=\boldsymbol{C}_{\varepsilon}\boldsymbol{J}_{ij}^{\mathrm{T}}\boldsymbol{C}_{i,\mathrm{RTR}}^{-1}\boldsymbol{J}_{\mathrm{if\text{-}d}}^{-1}\boldsymbol{K}_{\mathrm{d}}^{-1}\boldsymbol{J}_{\mathrm{d\text{-}f}}^{\mathrm{T}}\boldsymbol{F} \tag{25}$$

当确定应变片规格型号后, 结合式 (1) 便可建立双连孔力传感器惠斯通桥路输出灵敏度与待测载荷间的映射关系, 以及非期望载荷间的维间耦合程度。当双连孔结构作为柔性移动模块用于高精密柔性微动平台或作为力敏单元用于多维测力传感器时, 可将其视为类梁单元, 而所构建的柔度模型和应变-载荷关系则可进一步视为该类梁单元的力学属性, 如此便可用于快速构建柔性微动平台的整体刚度模型或多维测力传感器的应变测量模型, 适用性较强, 且可有效降低含双连孔结构的多维测力传感器性能建模的解析难度。

4 双连孔力传感器解析模型验证

选取 4 组不同的双连孔力传感器结构, 其结构参数如表 1 所示, 通过有限元仿真软件和实验分别验证所构建双连孔力传感器的柔度模型和所建立的应变-载荷映射关系的正确性。其中, 弹性体材料选用 2A12-T4 铝合金, 弹性模量 E=73.1 GPa, 泊松比 μ=0.33, 屈服极限为 230 MPa。

表 1 力传感器结构参数

Table 1 Structural parameters of force sensors mm

算例	w	r	t	l_{fs}	r_0	l_{b}	h	l
1	12	6	1	12	6	20	20	7
2	12	7	1.1	14	7	16	20	7
3	14	7	1.2	14	7	16	20	7
4	12	8	1.2	16	8	12	20	7

4.1 有限元仿真验证

双连孔力传感器弹性体结构的有限元网格模型、载荷施加、约束条件和路径定义如图 8 所示, 其中, 路径 A～D 定义在圆弧段外平面最薄处居中对称位置且长度 s=2 mm, 用于记录对应载荷下应变片贴片位置处的平均应变, 进而解算传感器的桥路输出应变仿真值。路径 E'端点 E_1'位于自由端两通孔轴线所确定的平面且相交于该平面截切自由端所形成断面的几何中心, 端点 E_2'位于断面上边线且路径 E'平行于通孔轴线, 仿真时对传感器期望维施加单位力载荷, 记录路径 E'在期望维载荷作用下沿测力方向上的线位移, 即可解算双连孔结构沿测力方向上的线柔度因子仿真值。

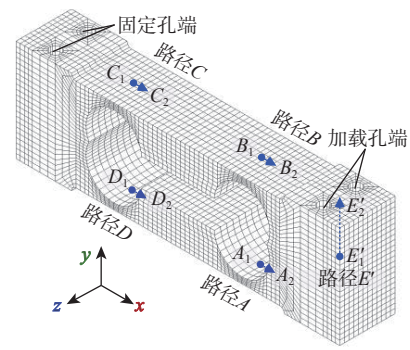


图 8 有限元网格模型及路径定义

Fig. 8 Finite element meshing model and path definitions

利用式 (16) 获得表 1 所列 4 组双连孔力传感器算例的期望维柔度因子解析解和利用有限元软件获得的算例的仿真结果, 两者的对比分析结果如表 2 所示, 其中, 解析解 A_{n} 相对于有限元结果 F_{E} 相对误差可利用 E_{rr}=abs[(A_{n}−F_{E})/F_{E}]×100% 计算得到。从表 2 可看出, 本文构建的柔度模型可精确预测双连孔结构力传感器沿测力方向上的线柔度, 解析解和有限元结果间的相对误差不超过 3%, 解析模型能够满足双连孔力传感器期望维刚度或柔度设计, 非期望维载荷作用下弹性体的变形也可依据式 (16) 进行解算。

表 2 期望维解析柔度与有限元结果对比

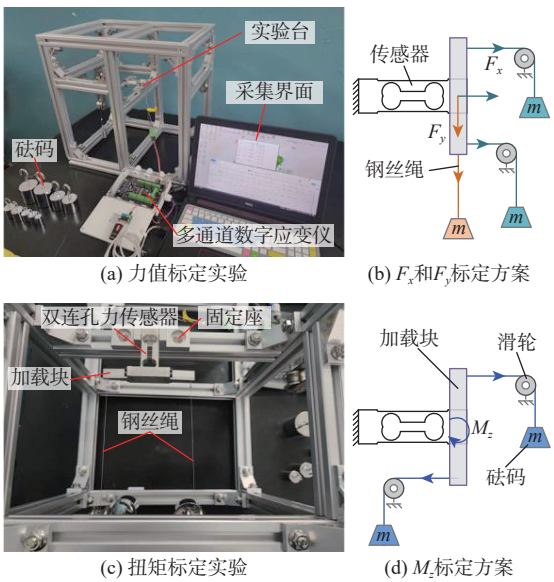
Table 2 Comparison of analytical results relative to finite element results

算例	A_{n}/10^{-6}(m·N^{-1})	F_{E}/10^{-6}(m·N^{-1})	E_{rr}/%
1	13.997	14.198	1.42
2	10.574	10.759	1.72
3	7.291	7.450	2.13
4	8.019	8.200	2.21

4.2 实验验证

为进一步验证双连孔力传感器桥路输出应变与测力载荷间的映射关系和非期望维载荷对期望维测力的耦合影响, 搭建出如图 9 所示的实验装置对其进行实验验证, 实验装置包括实验平台、固定座、双连孔力传感器、加载块、多通道数字应变仪、电脑、滑轮和标准砝码等, 图 9(a) 和图 9(c) 为

用于传感器各维标定的实验装置, 图 9(b) 和图 9(d) 为期望维和非期望维载荷分步施加方式。需要注意的是, 算例 1 被用于实验测量对象, 材料属性和结构参数均与算例 1 保持一致, 传感器在期望维测力方向上的额定量程为 20 N, 多通道数字应变仪的分辨率为 1/20 000 mV。



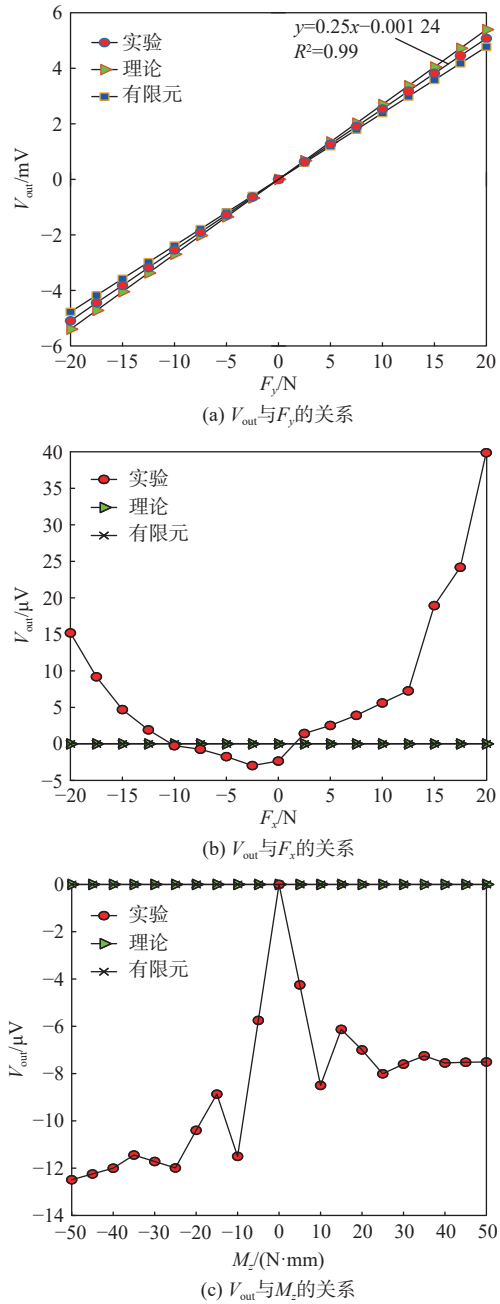
(a) 力值标定实验　(b) F_x和F_y标定方案

(c) 扭矩标定实验　(d) M_z标定方案

图 9　实验装置和加载示意图

Fig. 9　Experimental apparatus and loading schematic diagram

实验时通过螺栓将双连孔力传感器与实验台、固定座和加载块进行紧固连接, 对标定系统进行放置和预热后, 按照检测规范实施标定实验, 钢丝绳由加载块引出, 经滑轮后通过在其末端挂载不同重量的标准砝码, 依次对测力载荷 F_x 和非期望维载荷分量 F_y 和 M_z 在载荷区间内等间隔进行加载和卸载标定, 利用数据采集界面记录数字应变仪采集到的毫伏电压信号, 实验重复 3 次并经处理后, 可得到双连孔力传感器理论和有限元与实测桥路的输出电压信号间的关系如图 10 所示。



(a) V_{out}与F_y的关系

(b) V_{out}与F_x的关系

(c) V_{out}与M_z的关系

图 10　桥路输出电压与作用载荷间的关系

Fig. 10　Relationships between bridge output voltage and applied loads

由图 10 可知, 标定实验数据在测力方向量程范围内具有较高的线性度, 且理论预测桥路电压输出和实测结果具有较好一致性, 相对误差控制在 5% 以内, 施加非期望力和力矩载荷时, 耦合电压信号输出分别控制在 0.04 mV 和 0.014 mV 以内, 耦合电压输出产生的原因可能在于桥路电压波动、环境干扰或贴片工艺等因素, 但耦合电压占满量程测力载荷下电压信号输出小于 1%, 进而从多维标定的角度说明双连孔力传感器具有很强的抗维间耦合能力。

关于双连孔力传感器的精度, 不仅与其弹性体结构设计和参数寻优有关, 还涉及后续材料热处理、贴片工艺、处理电路、变送器分辨率、温度补偿、标准力加载机精度等级、线材线缆等诸多细节和流程, 故而传感器结构建模仅能确保在设计初期使其性能达到最优, 而实际测量精度和稳定性则与关键工艺和技术沉淀存在直接关联, 故传感器的精度仍需通过后期标定等多种手段共同确定。

4.3　耦合分析

为研究非期望维载荷作用下对双连孔力传感器测力维桥路维间耦合的影响程度, 在传感器加载孔端依次施加单维力及力矩, 力和力矩载荷分别为 1 N 和 1 N·mm, 经仿真并结合式 (1) 可分别获取表 1 中 4 个算例的桥路输出应变和最大等效应力如表 3 所示。

表3 输出应变和最大等效应力结果对比

Table 3 Comparison for the output strain and the maximum equivalent stress

算例	ε_{out}						σ_{max}/MPa					
	F_x=1 N	F_y=1 N	F_z=1 N	M_x=1 N·mm	M_y==1 N·mm	M_z=1 N·mm	F_x=1 N	F_y=1 N	F_z=1 N	M_x=1 N·mm	M_y==1 N·mm	M_z=1 N·mm
1	2.3×10^{-6}	−46.2	3.9×10^{-5}	-2.5×10^{-6}	-3.1×10^{-6}	-4.2×10^{-5}	8.4×10^{-2}	4.1	1.6	6.0×10^{-2}	3.5×10^{-2}	7.9×10^{-3}
2	2.5×10^{-7}	−36.1	-3.9×10^{-5}	1.8×10^{-6}	9.6×10^{-7}	-2.8×10^{-5}	7.5×10^{-2}	3.2	1.3	5.1×10^{-2}	3.1×10^{-2}	7.1×10^{-3}
3	6.8×10^{-6}	−26.1	1.2×10^{-4}	2.2×10^{-6}	-4.5×10^{-6}	-4.8×10^{-5}	5.8×10^{-2}	2.3	0.9	3.5×10^{-2}	2.1×10^{-2}	5.5×10^{-3}
4	-3.9×10^{-5}	−28.4	2.9×10^{-5}	7.7×10^{-8}	-9.5×10^{-7}	-1.3×10^{-5}	6.8×10^{-2}	2.57	1.2	4.3×10^{-2}	2.7×10^{-2}	6.4×10^{-3}

对表3结果进行分析及结合式(25)可知，双连孔力传感器桥路仅对测力方向上的力敏感，而在传感器破坏前所能承受其他维的最大载荷分量作用时，桥路耦合应变输出仍较小，这与理论上无维间耦合结果相一致，进一步验证传感器的抗维间耦合能力，故而使其能较好地用于多维力传感器的解耦设计。

4.4 影响分析

根据式(25)可直接获得宽度 w 和弹性模量 E 对传感器桥路输出应变的影响呈负相关，而连接梁长度 l_b 可视为放大系数[6]，与输出应变呈正相关。不失一般性，以表1中的算例1为对象，依次将参数 r、t 和 l_{fs} 选为变量且相对算例1取值做归一化处理，其他参数取值与算例1保持一致，可得到传感器主要结构参数对其桥路输出应变在期望维载荷为单位力载荷条件下的影响趋势如图11所示。

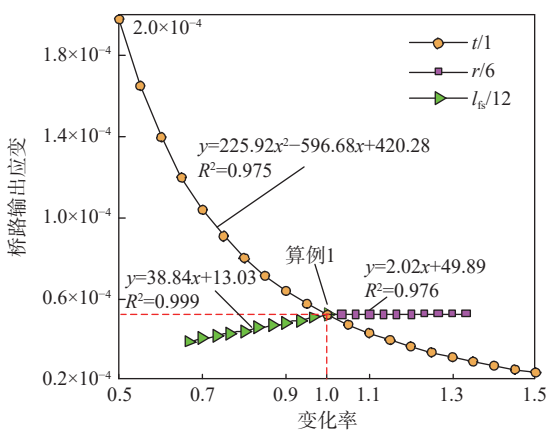


图11 结构参数对输出应变的影响分析

Fig. 11 Effect of structural parameters on the output strain

由图11对应的拟合方程和决定性系数 R^2 可看出，ε_{out} 与 t 呈非线性递减趋势，且 t 越小传感器测量灵敏度越高，但 t 值的选取需同时兼顾考虑传感器的刚度和材料厚度的可加工限制，ε_{out} 与 r 呈线性正相关，与 l_{fs} 呈线性负相关，参数允许范围内 ε_{out} 变化不大，说明力敏单元变形主要发生于其最薄处两端较小距离内。

5 结 论

1）本文基于弹性梁理论和利用矩阵建模法推导得到可解析描述的双连孔力传感器整体柔度模型和应变与载荷间的映射关系，并通过有限元软件和实验验证了解析模型的正确性。

2）本文研究了传感器主要结构参数和其他维载荷对测量桥路输出应变的影响，其中，变截面力敏单元槽口最薄处厚度对传感器桥路输出应变影响较为显著。

3）本文构建的柔度和应变解析模型可为含复杂变截面力敏单元的单维力传感器建模与分析提供技术支持，也可为集成双连孔结构的多维力传感器性能建模及优化设计提供有效手段。

参考文献（References）

[1] 于昌新, 何彦霖, 何超江, 等. 用于微创手术探针的光纤力传感器设计[J]. 光学精密工程, 2022, 30(20): 2421-2429.
YU C X, HE Y L, HE C J, et al. Design of fiber force sensors for surgical probes[J]. Optics and Precision Engineering, 2022, 30(20): 2421-2429(in Chinese).

[2] 胡一达, 李大寨, 杨洋, 等. 显微手术机器人末端微小型传感器集成研究[J]. 北京航空航天大学学报, 2005, 31(7): 794-798.
HU Y D, LI D Z, YANG Y, et al. Integration of microsensor for microsurgery robot's end-effector[J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(7): 794-798(in Chinese).

[3] 王建乐, 裘国华, 李青. 悬臂梁式电子秤全周倾斜称重技术研究[J]. 传感技术学报, 2022, 35(7): 881-887.
WANG J L, QIU G H, LI Q. Research on the whole cycle tilt weighing technology of cantilever beam electronic scale[J]. Chinese Journal of Sensors and Actuators, 2022, 35(7): 881-887(in Chinese).

[4] 宋文敏. 高精确度S型双连孔测力传感器[J]. 自动化仪表, 1983, 4(5): 21-24.
SONG W M. High-precision S-type double-hole force sensor[J]. Process Automation Instrumentation, 1983, 4(5): 21-24(in Chinese).

[5] 钟海见. XFb型过载保护式称重传感器[J]. 自动化仪表, 1990, 11(7): 16-18.
ZHONG H J. XFb overload protection weighing sensor[J]. Process Automation Instrumentation, 1990, 11(7): 16-18(in Chinese).

[6] 刘九卿. 平行梁型称重传感器的力学特性[J]. 衡器, 2009, 38(6): 5-7.
LIU J Q. Mechanics characteristics of parallel beam load cell[J].

Weighing Instrument, 2009, 38(6): 5-7(in Chinese).

[7] LOU Y J, WEI J H, SONG S. Design and optimization of a joint torque sensor for robot collision detection[J]. IEEE Sensors Journal, 2019, 19(16): 6618-6627.

[8] NGUYEN D D, KUO C H. Design and optimization of a joint torque sensor for lightweight robots[J]. IEEE Sensors Journal, 2021, 21(8): 9788-9797.

[9] FU B, CAI G W. Design and calibration of a joint torque sensor for robot compliance control[J]. IEEE Sensors Journal, 2021, 21(19): 21378-21389.

[10] 孙玉香, 曹会彬, 汪正全, 等. 空间多维力/力矩传感器关键技术研究综述[J]. 测控技术, 2020, 39(5): 9-14.
SUN Y X, CAO H B, WANG Z Q, et al. Review on key technologies of space multi-dimensional force/torque sensors[J]. Measurement & Control Technology, 2020, 39(5): 9-14(in Chinese).

[11] SONG A G, FU L Y. Multi-dimensional force sensor for haptic interaction: a review[J]. Virtual Reality & Intelligent Hardware, 2019, 1(2): 121-135.

[12] CAO M Y, LAWS S, BAENA F R Y. Six-axis force/torque sensors for robotics applications: a review[J]. IEEE Sensors Journal, 2021, 21(24): 27238-27251.

[13] DWARAKANATH T A, BHUTANI G. Beam type hexapod structure based six component force-torque sensor[J]. Mechatronics, 2011, 21(8): 1279-1287.

[14] WU B, CAI P. Decoupling analysis of a sliding structure six-axis force/torque sensor[J]. Measurement Science Review, 2013, 13(4): 187-193.

[15] ZHANG Z J, CHEN Y P, ZHANG D L, et al. A six-dimensional traction force sensor used for human-robot collaboration[J]. Mechatronics, 2019, 57: 164-172.

[16] MAYETIN U, KUCUK S. A low cost 3-DOF force sensing unit design for wrist rehabilitation robots[J]. Mechatronics, 2021, 78: 102623.

[17] AKBARI H, KAZEROONI A. Improving the coupling errors of a Maltese cross-beams type six-axis force/moment sensor using numerical shape-optimization technique[J]. Measurement, 2018, 126: 342-355.

[18] 王延深, 闫志彪. 一种柔性机构自解耦六维力传感器: CN108918013A [P]. 2018-11-30.
WANG Y S, YAN Z B. A decoupling six-axis force sensor based on a flexible mechanism: CN108918013A[P]. 2018-11-30(in Chinese).

[19] 路懿, 陈立伟. 刚柔混合三腿六维力传感器测力性能分析[J]. 燕山大学学报, 2014, 38(2): 127-133.
LU Y, CHEN L W. Analysis of measuring performance for a parallel sensor with three rigid-compliment hybrid limb[J]. Journal of Yanshan University, 2014, 38(2): 127-133(in Chinese).

[20] 张晨, 张元开, 王均. 基于串并混联柔顺机构的三维力传感器设计[J]. 仪表技术与传感器, 2019(8): 4-7.
ZHANG C, ZHANG Y K, WANG J. Design of three-dimensional force sensor based on string and hybrid compliant mechanism[J]. Instrument Technique and Sensor, 2019(8): 4-7(in Chinese).

[21] 王永立, 路懿. 分流式柔性铰六维力传感器刚度分析与结构优化[J]. 农业机械学报, 2019, 50(6): 419-426.
WANG Y L, LU Y. Stiffness analysis and structure optimization of shunted sensor with flexure hinge[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(6): 419-426(in Chinese).

[22] SHARIFZADEH M, MASOULEH M T, KALHOR A. Design, construction & calibration of a novel Human-Robot Interaction 3-DOF force sensor[C]//Proceedings of the 3rd RSI International Conference on Robotics and Mechatronics. Piscataway: IEEE Press, 2015: 182-187.

[23] 王保兴, 孟刚, 林苗, 等. 3-PPP 型柔性并联微定位平台的设计与分析[J]. 北京航空航天大学学报, 2020, 46(4): 798-807.
WANG B X, MENG G, LIN M, et al. Design and analysis of a 3-PPP compliant parallel micro-positioning stage[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(4): 798-807(in Chinese).

[24] 于致远, 吴小霞, 王富国. 大口径主镜的六杆硬点定位机构参数优化[J]. 光学精密工程, 2023, 31(2): 200-213.
YU Z Y, WU X X, WANG F G. Optimal design of parameters of the six-hardpoint positioning mechanism for large aperture primary mirror[J]. Optics and Precision Engineering, 2023, 31(2): 200-213 (in Chinese).

[25] 安守和, 汪满新, 李正亮. 两操作模式 2-R(SS)$_2$-R(RR)$_2$(RR)$_2$ 并联机构误差建模及精度设计[J]. 西安交通大学学报, 2020, 54(7): 146-157.
AN S H, WANG M X, LI Z L. Error modeling and accuracy design of a 2-R(SS)$_2$-R(RR)$_2$(RR)$_2$ parallel mechanism with two operation modes[J]. Journal of Xi'an Jiaotong University, 2020, 54(7): 146-157 (in Chinese).

[26] IVANOV I, CORVES B. Stiffness-oriented design of a flexure hinge-based parallel manipulator[J]. Mechanics Based Design of Structures and Machines, 2014, 42(3): 326-342.

[27] 李立建, 姚建涛, 郭飞, 等. 混合型柔性铰链构型设计与柔度建模[J]. 机械工程学报, 2022, 58(21): 78-91.
LI L J, YAO J T, GUO F, et al. Configuration design and compliance modeling of hybrid flexure hinges[J]. Journal of Mechanical Engineering, 2022, 58(21): 78-91(in Chinese).

[28] CHEN G M, HOWELL L L. Two general solutions of torsional compliance for variable rectangular cross-section hinges in compliant mechanisms[J]. Precision Engineering, 2009, 33(3): 268-274.

[29] PHAM H H, CHEN I M. Stiffness modeling of flexure parallel mechanism[J]. Precision Engineering, 2005, 29(4): 467-478.

Compliance modeling and strain analysis of double-hole force sensor

LI Lijian[1, 2], TANG Shouqian[1], YAO Jiantao[2, 3, *], WANG Yingjia[1]

(1. School of Mechanical Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China;
2. Parallel Robot and Mechatronic System Laboratory of Hebei Province, Yanshan University, Qinhuangdao 066004, China;
3. Key Laboratory of Advanced Forging & Stamping Technology and Science of Ministry of Education, Yanshan University, Qinhuangdao 066004, China)

Abstract: Double-hole force sensors are widely used in aerospace, industrial metrology and other fields, but their load-displacement characteristics and strain-load relationship are difficult to be formulated analytically due to the variable cross-section design, which will affect the design and performance optimization of multi-axis force sensors with double-hole structure. This study uses the compliance matrix modeling method to build the overall compliance model of the double-hole force sensor after determining the compliance matrix and strain-load relationship of the force-sensitive element with variable cross-section based on the elastic-beam theory. The analytical relationship between the sensed strain of the strain gauges on the force-sensitive element of the double-hole structure and the applied load acting on the force-measuring end of the sensor is finally obtained with the compliance model. The presented model and strain analytical relationship are validated by the finite element analysis and experiment, respectively. The results show that the relative errors for the expected-dimensional analytical compliances relative to finite element results are within 3%, and the bridge output voltages are within 5% compared with experimental results. These findings demonstrate that the analytical equations that were derived are capable of accurately assessing the load-displacement characteristics of double-hole force sensors as well as the mapping between applied loads and bridge output strains. They can also offer dependable technical assistance for the best possible design of multi-axis force sensors that have double-hole structures.

Keywords: force sensor; double-hole structure; compliance modeling; strain analysis; calibration experiment

Received: 2023-11-30; **Accepted**: 2024-01-05; **Published Online**: 2024-01-18 11:22

URL: link.cnki.net/urlid/11.2625.V.20240117.1645.001

Foundation items: National Natural Science Foundation of China (52005181, U2037202); Science and Technology Project of Henan Province (252102220107,222102220039); Key Research Project Plan for Higher Education Institutions in Henan Province (24A460016); The Training Program for Young Backbone Teachers in Higher Education Institutions of Henan Province; The Training Program for Young Backbone Teachers in North China University of Water Resources and Electric Power

* **Corresponding author**. E-mail: jtyao@ysu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0816

多因素飞行冲突探测中的认知负荷评估方法

杨越*，马博凯，程龙
(中国民航大学 空中交通管理学院，天津 300300)

摘 要：管制员认知负荷的实时监测与评估对空管系统的安全运行具有重要意义，以生理参数反映管制指挥中认知负荷的变化情况，便于及时发现影响管制效能的不良工作状态，从而实现风险管控关口前移。基于雷达管制仿真实验平台搭建不同空域复杂性条件下的冲突探测场景，设计最小间距（4 km、12 km、16 km）×汇聚角度（45°、90°、135°）×速度特性（速度相同、快速优先、慢速优先）的被试内三因素重复测量实验方案；采集眼动和心电数据，基于多因素方差分析研究不同复杂性因素的影响效应及生理指标的变化规律，筛选能够有效表征空域复杂性的特征生理指标；在此基础上，采用随机森林（RF）、支持向量机（SVM）和长短期记忆网络（LSTM）3种机器学习算法对被试的认知负荷进行预测。研究表明，在同类因素的不同水平中，被试的认知负荷分别在最小间距（MD）接近告警间隔和汇聚角度（CA）为锐角的水平下达到最高，在速度特性（SC）为速度相同时达到最低。注视点总数（NrF）、扫视点总数（NrS）、平均扫视持续时间（SD）、平均扫视幅度（SA）、平均扫视峰值速度（SPV）、眨眼频率（BF）、瞳孔直径（PD）、平均呼吸频率（RR）及心率变异性中低频与高频功率的比值（LF/HF）9项生理指标可以作为有效表征不同MD水平下认知负荷的特征指标。基于单模态眼动信号的SVM模型预测准确率为94.69%，高于单模态心电信号预测模型和眼动+心电双模态信号预测模型，删除强相关的特征指标会在一定程度影响模型性能。

关 键 词：冲突探测；认知负荷；眼动指标；心电指标；机器学习；模型性能
中图分类号：V355
文献标志码：A **文章编号**：1001-5965(2026)03-0763-09

飞行冲突的探测和调配是管制员对空指挥的主要任务之一[1]。尽管现有的空管自动化系统能够提供识别和判断飞行冲突的辅助工具，但管制员作为空管指挥的主体，仍需通过实时探测各类飞行态势下的潜在冲突，使其情景意识始终保持在“人-机-环”的闭环框架中[2]。近几年，随着中国空中交通流量逐步恢复，管制员的工作压力也不断增加，由人因引发的不安全事件时有发生。其中，航空器的运行间隔、飞行态势、相对速度等不确定因素都会使管制员在冲突探测过程中的认知负荷产生波动，而长期在高负荷的状态下工作，极易导致人因事故的发生。因此，实时监测管制员的认知负荷水平，有助于及时发现影响空管运行的安全隐患，实现风险管控关口前移。

冲突探测任务中，最小间距（minimum distance, MD）、汇聚角度（convergence angle, CA）及速度特性（speed characteristics, SC）是决定个体认知复杂性的关键因素。最小间距指2架飞机在最接近点的距离，是一个影响认知过程的因素，相比于飞机对之间明显小于或大于告警间隔的情况，当最小间距接近告警间隔时，被试处于边界认知状态，难以准确判断航空器之间是否会产生冲突。汇聚角度指2

收稿日期：2023-12-15；**录用日期**：2024-05-10；**网络出版时间**：2024-05-22 11:02
网络出版地址：link.cnki.net/urlid/11.2625.V.20240521.1318.001
基金项目：天津市教委科研计划 (2023KJ234)
* **通信作者**. E-mail：yueyang0124@126.com

引用格式：杨越，马博凯，程龙. 多因素飞行冲突探测中的认知负荷评估方法[J]. 北京航空航天大学学报，2026，52（3）：763-771.
YANG Y，MA B K，CHENG L. Cognitive load assessment method for multifactorial flight conflict detection[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：763-771（in Chinese）.

架飞机航向之间的夹角，是一个影响视觉信息获取的几何因素。速度特性指2架飞机在保持速度不变的情况下，哪架飞机先通过汇聚点。在冲突探测场景中，研究者发现最小间距、汇聚角度和速度特性会影响被试推断航空器轨迹的精确性：航空器在最小间距接近告警间隔时更加影响对冲突的判断[3]；当航空器的汇聚角度较小时会导致更高的认知复杂性[4]；在航空器以相同速度飞行时，冲突探测的反应时比不同速度飞行时更快[5]。

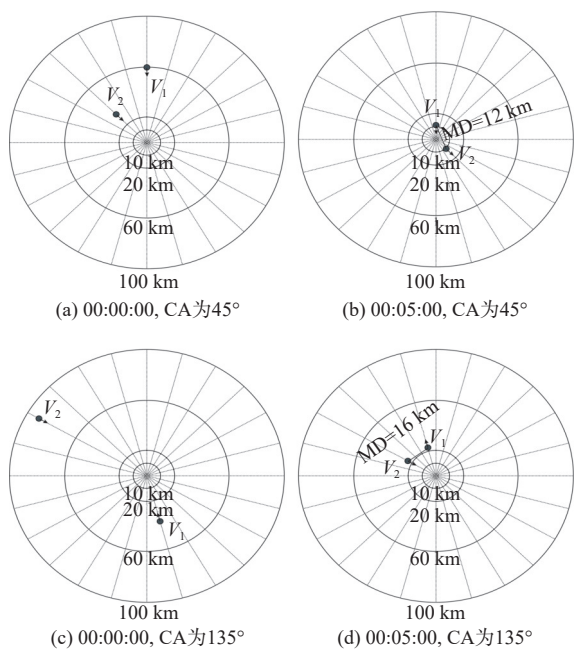
认知负荷（又称脑力负荷）是心理学、人因工效学领域研究的重要问题之一，通常被描述为对某项任务具有一定特征的主观感受[6]。一般来说，较高的任务需求会导致较高的认知负荷，但由于知识、经验及能力的差异，即使面对相同数量或难度的任务时，不同个体的认知负荷水平也有所不同。认知负荷的测量主要包括主观评定、任务绩效和生理测量3种方法[7-10]。与主观评定和任务绩效方法相比，生理测量法以其客观性和高时效性的优点，更加适用于人机交互过程中的认知负荷实时监测。多项研究成果表明，人体生理信号与认知负荷的变化关联紧密，其中，在眼动数据方面：He等[11]发现在模拟飞行任务中，随着认知负荷的增加，注视次数和注视持续时间也会增加；Bodala等[12]发现平均峰值扫视速度会随着认知负荷的增加而增加；考虑到高视觉负荷情况下的短眨眼发生率较高，眨眼频率被认为是评估认知负荷的敏感指标[13]；另外，瞳孔直径也被证明与各种活动的认知负荷呈正相关关系[14]。在心电数据方面：De Rivecourt等[15]在飞行任务实验中发现，心率会随着任务难度增加而显著增加；李鹏杰等[16]发现心率和心率变异性都能明显区分不同的认知负荷水平，但心率变异性的敏感性要优于心率；同时，呼吸频率也被认为是心理任务需求的敏感指标之一[17]。然而，人体生理信号的变化机制极其复杂，不同认知情境下的生理指标灵敏度差异明显，如何筛选针对特定情境的特征生理指标是要解决的关键问题。另外，单一模态的生理指标可能无法全面反映认知负荷的时变特性，基于多模态信息的模型评估方法已成为该领域的研究热点，例如：基于眼动、心电、脑电等多模态信息的反向传播（back propagation, BP）神经网络、支持向量机（support vector machine, SVM）、随机森林（random forest, RF）、K最近邻模型已逐步应用于汽车驾驶员[18-19]、航空飞行员[20-21]、软件程序员[22]的认知负荷评估，而上述方法在民航空管领域的应用却鲜有报道。

基于此，本文针对空中交通管制员的飞行冲突探测任务，研究空域复杂性因素对认知负荷的影响效应及生理指标的变化规律，通过构建机器学习评估模型，系统评价基于多模态生理指标的模型评估能力，为推动认知负荷实时监测技术在空管领域的应用提供支撑。

1 实验方法

1.1 实验方案设计

相关研究表明，影响管制员冲突探测准确率的空域复杂性因素包括航空器的运行间隔[23]、飞行夹角[24]及相对速度[25]等。因此，实验设计中综合考虑MD、CA及SC这3种复杂性因素对被试认知负荷水平的影响效应。基于雷达管制模拟实验平台构建半径为100 km的圆形空域，分别设置距圆心10 km、20 km和60 km的位置参考线及等分空域的24条虚拟航线。在每个实验场景中划设2架航空器的初始位置，使其以相同高度和恒定速度沿着随机的虚拟航线朝向圆心汇聚飞行，并在各场景进行到5 min时到达MD，如图1所示。采用3×3×3的被试内三因素重复测量实验设计方案，每个因素均包括3个水平：MD分别为4 km、12 km和16 km；CA分别为45°、90°和135°；SC分别为速度相同、快速优先和慢速优先。例如，图1(a)和图1 (b)所示的实验场景表示2架航空器在5 min后到达最小间距12 km，汇聚过程中的角度保持45°，且速度快的航空器优先通过汇聚点；图1 (c)和图1 (d)所示的实验场景表示2架航空器在5 min后到达最小间距16 km，汇聚过程中的角度保持135°，且速度慢的航



(a) 00:00:00, CA为45° (b) 00:05:00, CA为45°
(c) 00:00:00, CA为135° (d) 00:05:00, CA为135°

图1 冲突探测实验场景

Fig. 1 Conflict detection experimental scenarios

空器优先通过汇聚点; 航空器速度 V_1 小于 V_2。

实验中, 若 2 架航空器的水平间距小于 10 km 则产生冲突, 人机交互界面显示短期冲突告警(short term conflict alert, STCA)。各场景进行期间, 被试必须尽快判断该场景是否存在冲突(存在冲突按下鼠标左键, 不存在冲突按下鼠标右键; 若已出现 STCA 提示再进行判断则视为无效, 按判断错误统计)。判断完成之后该场景结束, 立即进入下一个实验场景。

本次实验共招募 20 名年龄在 20～23 岁(平均值 M=21.58 岁, 标准差 S_D=1.3)的管制学员作为被试, 其中, 男性 16 名, 女性 4 名, 均为交通运输专业背景的大四学生, 身体健康状况良好, 视力或矫正视力正常。实验前, 被试均进行了为期 4 周的雷达管制模拟机培训并通过考核, 具备基本的冲突探测能力和管制安全意识。每位被试共需完成 27 个场景的测试, 各场景随机出现以消除遗留效应的影响, 整个实验大约持续 40 min。

1.2　测量指标和实验假设

实验中, 以任务绩效指标反映被试的认知负荷水平, 包括冲突探测的正确率(accuracy, ACC)和反应时(response time, RT)。ACC 定义为正确判断冲突的次数与总判断次数的比值; RT 定义为判断冲突用时的平均时长。ACC 值越小, RT 值越大, 表明被试的认知负荷水平越高。根据文献 [3-5], 在实验前针对本文的 3 项复杂性因素作出以下假设: ①在 MD 为 12 km 的场景中被试的认知负荷相比于其他 MD 水平更高; ②当 CA 为 45°时相比于其他 CA 水平的认知负荷更高; ③在航空器速度相同的场景中被试的认知负荷相比于速度不同时更低。

为研究不同认知负荷水平下生理指标的变化规律, 实验中分别采集了被试的眼动和心电数据, 计算并分析了以下 13 项生理指标: 注视点总数(number of fixations, NrF)、平均注视持续时间(mean fixation duration, FD)、扫视点总数(number of saccades, NrS)、平均扫视持续时间(mean saccade duration, SD)、平均扫视幅度(average saccadic amplitude, SA)、平均扫视峰值速度(average saccadic peak velocity, SPV)、眨眼频率(blinking frequency, BF)、瞳孔直径(pupil diameter, PD)、平均呼吸频率(mean respiratory rate, RR)、平均心率(mean heart rate, HR)、全部正常心跳间期标准差(standard deviation of normal to normal, SDNN)、相邻正常心跳间期之差大于 50 ms 的比例(percentage of successive normal-to-normal intervals differing by more than 50 milliseconds, PNN50)及心率变异性中低频与高频功率的比值(power ratio of low frequency to high frequency in heart rate variability, LF/HF)。

1.3　实验设备和数据处理

实验中的眼动数据采用瑞典 Tobii 公司的 Pro Glasses2 眼动仪采集, 基于速度阈值识别(velocity threshold identification, I-VT)算法辨识注视和扫视行为[26], 其中, 注视持续时间的阈值设置为 100 ms, 扫视行为的速度阈值设置为 45(°)/s[27], 采样频率设置为 50 Hz。心电数据采用北京津发公司的 ErgoLAB 第 I 代人因生理记录仪采集, 通过小波去噪、带阻滤波去除肌电干扰和基线漂移等噪声[28], 采样频率设置为 256 Hz。实验中有 2 名被试的生理数据采集异常不计入分析, 其余各项指标的计算结果通过正态性检验和方差齐次性检验后, 采用 SPSS Statistics 24.0 软件进行三因素重复测量的方差分析。当 Mauchly 球形检验显著时, 使用 Greenhouse-Geisser 方法进行校正。为有效控制 I 型错误的发生率, 事后比较采用 Bonferroni 检验, 所有统计检验均使用 0.05 的置信度(p=0.05)。

2　实验结果

2.1　绩效指标测量结果

ACC 的事先对比结果。MD 为 12 km 的 ACC 显著低于 MD 为 4 km 和 MD 为 16 km(p<0.001); CA 为 45°的 ACC 显著低于 CA 为 90°和 CA 为 135°(p=0.001); SC 为速度相同的 ACC 显著高于 SC 为快速优先和慢速优先(p=0.005)。

RT 的事先对比结果。MD 为 12 km 的 RT 显著慢于 MD 为 4 km 和 MD 为 16 km(p<0.001); CA 为 45°的 RT 显著慢于 CA 为 90°和 CA 为 135°(p=0.039); SC 为速度相同的 RT 显著快于 SC 为快速优先和慢速优先(p<0.001)。

任务绩效指标的事先对比结果表明: 在航空器的最小间距接近告警间隔、汇聚角度较小、飞行速度不同的实验场景中, 相比于各因素的其他研究水平, 被试执行冲突探测任务时的认知负荷明显升高, 实验结果符合事先假设。

2.2　生理指标测量结果

2.2.1　眼动指标

NrF 的方差分析结果。MD 的主效应显著(统计量 F(2,34)=4.033, p=0.027), MD 为 12 km 的 NrF 显著高于 MD 为 4 km(p=0.001)和 MD 为 16 km(p<0.001)。CA 和 SC 的主效应不显著。

FD 的方差分析结果。MD、CA 和 SC 的主效应均不显著。

NrS的方差分析结果。MD的主效应显著($F(2,34)=5.925$, $p=0.006$)，MD为12 km的NrS显著高于MD为4 km($p=0.001$)和MD为16 km($p<0.001$)。CA和SC的主效应不显著。

SD的方差分析结果。MD的主效应显著($F(2,34)=37.740$, $p<0.001$)，MD为12 km的SD显著低于MD为4 km($p<0.001$)，MD为16 km的SD显著低于MD为4 km($p<0.001$)；CA的主效应显著($F(2,34)=42.438$, $p<0.001$)，CA为90°的SD显著高于CA为45°($p<0.001$)，CA为135°的SD显著高于CA为45°($p<0.001$)；SC的主效应显著($F(2,34)=7.236$, $p=0.002$)，SC为速度相同的SD显著高于SC为快速优先($p=0.034$)和慢速优先($p=0.002$)。

SA的方差分析结果。MD的主效应显著($F(2,34)=35.532$, $p<0.001$)，MD为12 km的SA显著低于MD为4 km($p<0.001$)，MD为16 km的SA显著低于MD为4 km($p<0.001$)；CA的主效应显著($F(2,34)=57.219$, $p<0.001$)，CA为45°的SA显著低于CA为90°($p<0.001$)，CA为90°的SA显著低于CA为135°($p=0.027$)。SC的主效应不显著。

SPV的方差分析结果。MD的主效应显著($F(2,34)=11.915$, $p<0.001$)，MD为12 km的SPV显著低于MD为4 km($p=0.001$)，MD为16 km的SPV显著低于MD为4 km($p=0.001$)；CA的主效应显著($F(2,34)=23.583$, $p<0.001$)，CA为45°的SPV显著低于CA为90°($p<0.001$)，CA为90°的SPV显著低于CA为135°($p=0.044$)。SC的主效应不显著。

BF的方差分析结果。MD的主效应显著($F(2,34)=13.885$, $p<0.001$)，MD为12 km的BF显著高于MD为4 km($p<0.001$)，MD为16 km的BF显著高于MD为4 km($p<0.001$)；CA的主效应显著($F(2,34)=5.030$, $p=0.012$)，CA为45°的BF显著高于CA为90°($p=0.026$)和CA为135°($p=0.013$)；SC的主效应显著($F(2,34)=7.366$, $p=0.002$)，SC为速度相同的BF显著低于SC为快速优先($p=0.015$)和慢速优先($p=0.002$)。

PD的方差分析结果。MD的主效应显著($F(2,34)=5.352$, $p=0.010$)，MD为12 km的PD显著高于MD为4 km($p=0.009$)。CA和SC的主效应不显著。

2.2.2 心电指标

RR的方差分析结果。MD的主效应显著($F(2,34)=3.987$, $p=0.028$)，MD为4 km的RR显著高于MD为12 km($p=0.024$)。CA和SC的主效应不显著。

HR的方差分析结果。CA的主效应显著($F(2,34)=3.810$, $p=0.032$)，CA为90°的HR显著高于CA为45°($p=0.048$)和CA为135°($p=0.030$)；SC的主效应显著($F(2,34)=8.913$, $p=0.001$)，SC为速度相同的HR显著高于SC为快速优先($p=0.008$)和慢速优先($p=0.002$)。MD的主效应不显著。

SDNN和PNN50的方差分析结果。MD、CA和SC的主效应均不显著。

LF/HF的方差分析结果。MD的主效应显著($F(2,34)=3.830$, $p=0.032$)，MD为12 km的LF/HF显著高于MD为16 km($p=0.028$)。CA和SC的主效应不显著。

各项生理指标的方差分析结果如表1所示。

表1 各项生理指标的显著性检验 p 值

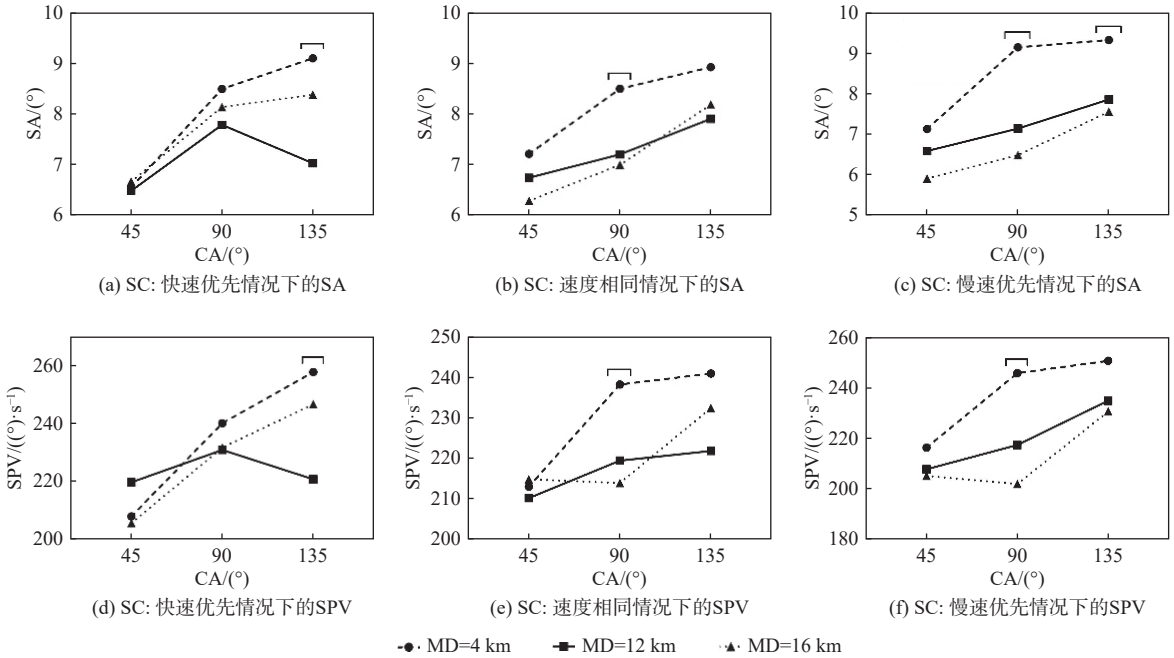
Table 1 Analysis of variance results of physiological indexes

因素	NrF	FD	NrS	SD	SA	SPV	BF	PD	RR	HR	SDNN	PNN50	LF/HF
MD	0.027*	0.114	0.006**	0.000**	0.000**	0.000**	0.000**	0.010*	0.028*	0.102	0.33	0.229	0.032*
CA	0.531	0.213	0.332	0.000**	0.000**	0.000**	0.012*	0.102	0.399	0.032*	0.396	0.292	0.721
SC	0.158	0.445	0.086	0.002**	0.08	0.598	0.002**	0.32	0.45	0.001**	0.397	0.789	0.619

注: ** 表示在置信度$p=0.01$级别(双尾)相关性显著; * 表示在置信度$p=0.05$级别(双尾)相关性显著。

2.2.3 三因素交互作用

MD、CA与SC的三因素交互作用仅在SA和SPV这2项指标中表现出显著性差异，如图2所示(图中，水平线表示在CA中有显著的MD效应)。其中，当SC为快速优先、CA为135°时，MD为12 km的SA显著低于MD为4 km($p<0.001$)和MD为16 km($p=0.006$)，MD为12 km的SPV显著低于MD为4 km($p<0.001$)；当SC为速度相同、CA为90°时，MD为12 km的SA显著低于MD为4 km($p<0.001$)，MD为12 km的SPV显著低于MD为4 km($p<0.001$)；当SC为慢速优先、CA为90°时，MD为12 km的SA显著低于MD为4 km($p<0.001$)，MD为12 km的SPV显著低于MD为4 km($p<0.001$)；当SC为慢速优先、CA为135°时，MD为12 km的SA显著低于MD为4 km($p=0.018$)。



(a) SC: 快速优先情况下的SA　(b) SC: 速度相同情况下的SA　(c) SC: 慢速优先情况下的SA

(d) SC: 快速优先情况下的SPV　(e) SC: 速度相同情况下的SPV　(f) SC: 慢速优先情况下的SPV

图 2　CA×MD×SC 交互作用显著的生理指标分析结果

Fig. 2　Analysis results of physiological indexes with significant effect of CA × MD × SC interactions

3　冲突探测中的认知负荷评估

3.1　特征指标和模型参数的选择

当 MD 为主要因素时, 被试的多项生理指标变化明显, 而其他 2 种因素影响显著的生理指标相对较少。因此, 本文以 MD 作为典型的复杂性因素, 基于该因素影响显著的特征生理指标对被试的认知负荷进行评估, 筛选的特征指标为: NrF、NrS、SD、SA、SPV、BF、PD、RR 和 LF/HF(见表 1)。为避免不同特征指标数据量级相差过大而导致的量纲问题, 将各指标数据进行归一化处理以作为模型的输入参数。为进一步对模型的评估性能进行评价, 选取反向效率得分(inverse efficiency score, IES)作为模型的输出参数[29]。IES 的计算方法为: 被试在 MD 因素下的冲突判断 RT 与 ACC 的比值。

3.2　认知负荷评估方法和结果

为系统比较不同模态的生理指标对认知负荷的评估效果, 本文分别采用 RF、SVM 和长短期记忆网络(long short-term memory network, LSTM)3 种机器学习模型。其中, RF 算法可以有效处理缺失值和异常值并自动选择最优特征, 能够有效降低过拟合的风险; SVM 算法在处理高维数据和非线性问题时表现优异, 具有较好的泛化能力和处理小样本数据的能力; LSTM 算法可以处理复杂关联序列数据中的多变量和多时间阶段的评估问题, 具有良好的时序评估能力。以不同模态的生理指标作为模型的输入参数, 将其定义为: ① 单模态: 眼动数据、心电数据; ② 双模态: 眼动数据+心电数据。各模型中采用的每个特征指标共有 54 组数据(18 名被试×3 种 MD 水平), 设置训练集为 36, 测试集为 18, 选择 10 折交叉验证评估模型性能, 实施流程如图 3 所示, 模型评估结果如表 2 所示, 其中, R^2 为评估结果拟合度, R 为评估结果与真实结果的相关系数。

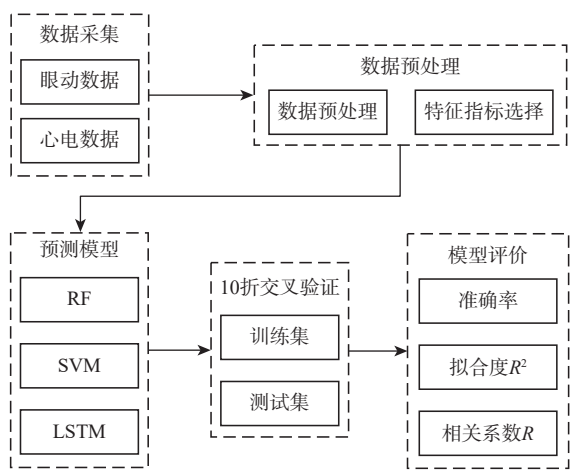


图 3　认知负荷评估方法流程图

Fig. 3　Flowchart of cognitive load assessment method

由表 2 可知, 以单模态眼动特征指标作为输入的模型评估准确率明显高于心电特征指标; SVM 模型在单模态生理数据中的评估效果普遍最佳, 基于眼动特征指标的评估准确率为 94. 69%, 模型的拟合度 R^2 达到 0.91, 评估值与真实值的相关系数 R 达到 0.95。眼动+心电双模态特征指标同样在 SVM 模型中的表现最为优异, 评估准确率为 93.28%, 模型的拟合度 R^2 和相关系数 R 分别达到 0.89 和 0.94, 与单模态心电和眼动特征指标相比, 准确率分

表2 不同模态生理指标的模型评估结果

Table 2 Model assessment results by different modal physiological indexes

指标模态	评估模型	准确率/%	R^2	R
眼动特征指标	RF	91.13	0.78	0.92
	SVM	94.69	0.91	0.95
	LSTM	93.37	0.88	0.93
心电特征指标	RF	74.25	0.029	0.23
	SVM	83.59	0.029	0.38
	LSTM	74.68	0.020	0.17
眼动+心电指标	RF	88.46	0.76	0.91
	SVM	93.28	0.89	0.94
	LSTM	91.58	0.82	0.90

别提高了9.69%和降低了1.41%。

以上结果表明,在单模态生理数据的评估模型中,基于眼动特征指标的RF、SVM和LSTM模型均具有较好的评估性能,其中,以SVM模型表现最佳;在双模态生理数据的评估模型中,基于眼动和心电特征指标的SVM模型的评估准确率最高,与单模态眼动特征指标的评估结果相比,加入心电指标反而降低了模型的评估准确率。

4 实验结果分析

4.1 复杂性因素对认知负荷的影响

本文主要研究了MD、CA和SC这3种因素对冲突探测任务的影响,以绩效指标ACC和RT反映被试在不同因素影响下的认知负荷水平。实验结果表明,当MD更接近告警间隔时,被试判断冲突的正确率更低、反应时更长,认知负荷水平更高。这是因为实验方案中设定航空器的高度相同,水平间距成为决定是否产生冲突的关键因素;在MD接近告警间隔的场景中,被试需要耗费更多的认知资源去判断航空器的间距是否会小于安全间隔,相比于存在冲突的场景(MD为4 km)和不存在冲突的场景(MD为16 km),被试做出判断的反应时更长、正确率更低。在CA较小的实验场景中(CA为45°),相比于较大角度的汇聚飞行(CA为90°、CA为135°),航空器之间的相对接近率更小、间距变化率也更小,被试需要耗费更长时间和更多精力去判断航空器之间的位置关系,认知负荷也会增加。而在同速飞行的实验场景中(SC为速度相同),航空器在相同时间内的位移量相同,相比于变速飞行(SC为快速优先和慢速优先),被试能更准确地推测未来时段内航空器所在的位置,从而使冲突判断的反应时更短、正确率更高。MD、CA和SC都是管制员在探测冲突时需要分析的重要因素,实验中任务绩效指标的事先对比结果完全符合假设,再次验证了航空器的飞行距离、角度和速度对认知复杂性的影响效应,也为管制方案和指挥策略的制定提供了依据。例如,管制员要始终关注航空器的间隔,在可能出现汇聚飞行的态势下,确保提前使用测距工具来精确监控航空器的位置关系;空域管理者在制定空中飞行方案时要特别关注交叉航路与航班流向之间的关系,尽量避免航空器以小角度汇聚的方式互相穿越高度;在指挥高空航路飞行时,管制员要更加注意不同机型航空器的速度区间,避免速度差异性对冲突判断产生的影响等。

4.2 冲突探测中生理指标的变化规律

在不同因素影响下的冲突探测场景中,被试的眼动指标存在明显的变化规律,而心电指标的变化规律不明显。实验结果表明,当MD接近告警间隔时,被试更容易出现判断失误,注视点和扫视点数量较多,扫视持续时间较短,扫视幅度和峰值速度较小,眨眼频率较高,瞳孔直径较大,说明自主神经系统的交感神经活动相对增加,认知负荷水平较高;而在MD小于告警间隔即存在冲突的实验场景中,被试的扫视时间更长,扫视幅度和峰值速度更大,瞳孔直径更小,说明该场景下的认知负荷水平较低。当航空器的汇聚角度较小时,被试的扫视持续时间较短,扫视幅度和峰值速度较小,眨眼频率较高。而相比于速度变化的实验场景,被试在完成同速汇聚的冲突探测任务时,扫视时间更长,眨眼频率更低,但在比较快速优先和慢速优先的影响程度时,未发现生理指标存在显著差异。相比于眼动指标明显的变化规律,各因素对多数心电指标的影响效应并不显著,事后比较发现部分特征指标不满足认知负荷的变化规律。这可能是由于本实验的任务场景中没有添加额外的刺激源,相比于真实的管制场景,当发生威胁安全的飞行冲突时并不会引发被试情绪和心理状态的明显波动,从而导致采集到的心电数据未能有效反映认知负荷的变化特征。因此,实验再次验证了生理信号在表征认知负荷的有效性方面会受到模型类型和任务场景的共同影响[30]。此外,实验结果显示各因素对多数生理指标未表现出明显的交互影响效应,说明本实验设计的复杂性因素对生理数据的主效应更加明显,生理指标在多因素交互作用下的变化规律还需通过更多数据样本和针对性更强的实验设计方案进行研究。

4.3 基于生理指标的认知负荷评估水平

本文还研究了基于不同模态生理数据的模型评估水平。分析发现,以眼动数据作为模型输入

时, 3 种机器学习算法的模型评估准确率均显著高于心电数据的结果。由于缺乏足够的特征指标会限制捕捉与目标变量相关的关键信息, 相比于心电数据, 实验中采集到了更多有效反映认知负荷水平的眼动特征数据, 进一步证实了敏感性指标数量对模型性能的影响[31]。研究发现, 基于双模态生理数据(眼动+心电)的 3 种模型评估准确率均低于单模态眼动数据的评估模型, 这可能源于本实验场景与眼动行为的密切关联性: 眼动数据在冲突探测中提供了更加直接和精确的信息源, 而心电数据虽然反映了被试个体自主神经系统的生理活动水平, 但未能有效反映各种复杂性因素对认知负荷的影响效应, 这相当于增加了额外的干扰因素, 对模型的评估能力产生了负面作用。在实际管制工作中, 除了冲突探测外, 管制员还需完成其他任务, 包括冲突解脱、雷达引导、管制移交及特情处置等。上述针对冲突探测场景的结论并不一定适用于其他任务, 如在特情处置中管制员的生理参数大概率会发生明显波动, 此时采用多模态的生理数据监控认知负荷的效果可能要优于单模态数据。因此, 未来研究中应选取更加符合任务特点的特征生理数据以建立适用于不同工作场景的认知负荷评估模型。研究还发现, 无论是基于单模态还是双模态生理数据的评估模型, SVM 模型的评估准确率较高。这源于本文所采集的生理数据在不同 MD 水平下是非线性变化的, 且数据样本量较少, 当多个生理指标作为模型的输入参数时构成了高维数据集; 而 SVM 算法在处理高维数据和非线性问题时表现优异, 并具有较好的泛化能力和处理小样本数据的能力, 因此模型性能表现较好。

此外, 为减少强相关指标可能对模型评估准确率产生的影响, 本文采用 Pearson 相关分析方法删除具有强相关性的特征指标以降低数据冗余度。当特征指标之间的相关系数绝对值超过 0.6 时, 仅保留其中 1 个指标[32]。表 3 给出了以筛选后的 NrF、SD、PD、RR 和 LF/HF 这 5 个指标作为输入参数的结果, 与表 2 相比, 各类模型的评估准确率均有所降低。由此可知, 虽然删除了可能导致数据冗余的强相关指标, 但实验表明这些指标与认知负荷存在密切的关联性, 且在所研究的变量水平下表现出明显的变化规律, 删除这些指标反而削弱了模型获取关键信息的能力, 进而影响了模型的评估水平。该结果为未来研究提供了重要启示: 在多维特征指标的选择中通常采用删除强相关指标的处理方式以减少数据冗余度, 但在特定情况下, 保留能够明显表征研究对象水平的特征指标会提高模型性能。因此, 在筛选特征指标时要进行更加精细和全面的分析, 综合考虑研究目标的特性和采集数据的可靠性, 保证不遗漏对目标变量有重要影响的敏感特征信息, 从而进一步提高模型的性能表现。

表 3　去除强相关特征指标的模型评估结果

Table 3　Model assessment results after removing strongly correlated feature indexes

指标模态	评估模型	准确率/%	R^2	R
眼动特征指标	RF	90.05	0.74	0.88
	SVM	92.90	0.90	0.93
	LSTM	92.35	0.86	0.93
心电特征指标	RF	74.25	0.029	0.23
	SVM	83.59	0.029	0.38
	LSTM	74.68	0.020	0.17
眼动+心电指标	RF	88.00	0.71	0.87
	SVM	92.39	0.89	0.93
	LSTM	91.15	0.73	0.92

5　结　论

1) 空域复杂性因素对冲突探测中的认知负荷的影响效应显著: 当 MD 接近告警间隔时, 认知负荷水平最高; 随着 CA 减小, 认知负荷逐步增加; 当 SC 为速度相同时, 认知负荷水平最低。以 MD 为主要因素的冲突探测场景中, NrF、NrS、SD、SA、SPV 等 9 项生理指标可作为表征认知负荷的特征指标。

2) 本文多因素冲突探测实验中, 眼动指标相比于心电指标具有更明显的变化规律; 基于单模态眼动数据的模型评估准确率高于单模态心电数据和双模态生理数据, 各模态组合中的 SVM 模型评估准确率高于 RF 模型和 LSTM 模型。

3) 基于多模态生理数据的模型评估技术结合感知信号和生理信号的优势, 为认知负荷的评估提供了更加精细和全面的分析方法, 但在应用前必须明确特征指标的适用条件和对研究目标的影响效应, 进一步完善该技术对推动人机交互领域的发展具有重要意义。

由于条件有限, 本文结论主要基于实验室环境下的小样本数据, 缺乏真实工作场景下的验证和应用, 需要在考虑可行性的前提下将研究成果更多地应用到实际的管制工作场景中。未来还应选取有工作经验的成熟管制员作为被试并深入考虑个体之间的生理差异, 包括不同年龄、性别、体质等因素对认知负荷评估结果的影响, 从而更好地研究空管中人的认知过程。

参考文献（References）

[1] HAO S Q, CHENG S W, ZHANG Y P. A multi-aircraft conflict detection and resolution method for 4-dimensional trajectory-based operation[J]. Chinese Journal of Aeronautics, 2018, 31(7): 1579-1593.

[2] ARICO P, BORGHINI G, DI FLUMERI G, et al. Human factors and neurophysiological metrics in air traffic control: a critical review[J]. IEEE Reviews in Biomedical Engineering, 2017, 10: 250-263.

[3] STANKOVIC S, RAUFASTE É, AVERTY P. Determinants of conflict detection: a model of risk judgments in air traffic control[J]. Human Factors, 2008, 50(1): 121-134.

[4] MARCHITTO M, BENEDETTO S, BACCINO T, et al. Air traffic control: ocular metrics reflect cognitive complexity[J]. International Journal of Industrial Ergonomics, 2016, 54: 120-130.

[5] RANTANEN E M, NUNES A. Hierarchical conflict detection in air traffic control[J]. The International Journal of Aviation Psychology, 2005, 15(4): 339-362.

[6] CHARLES R L, NIXON J. Measuring mental workload using physiological measures: a systematic review[J]. Applied Ergonomics, 2019, 74: 221-232.

[7] FRANK N, LE P, MILLS E, et al. Micromovements and discomfort associated with flight mission with helmet operation tasks with different levels of cognitive workload[J]. International Journal of Industrial Ergonomics, 2023, 95: 103441.

[8] GOH S Y M, LYE S W, SOU K L, et al. Effects of increased cognitive load on field of view in multi-task operations involving surveillance[J]. Human-Intelligent Systems Integration, 2021, 3(4): 277-285.

[9] JOHN A R, SINGH A K, DO T N, et al. Unraveling the physiological correlates of mental workload variations in tracking and collision prediction tasks[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 770-781.

[10] BERNHARDT K A, POLTAVSKI D, PETROS T, et al. The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment[J]. Applied Ergonomics, 2019, 77: 83-91.

[11] HE X L, WANG L J, GAO X H, et al. The eye activity measurement of mental workload based on basic flight task[C]//Proceedings of the IEEE 10th International Conference on Industrial Informatics. Piscataway: IEEE Press, 2012: 502-507.

[12] BODALA I P, KE Y, MIR H, et al. Cognitive workload estimation due to vague visual stimuli using saccadic eye movements[C]//Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway: IEEE Press, 2014: 2993-2996.

[13] MALLICK R, SLAYBACK D, TOURYAN J, et al. The use of eye metrics to index cognitive workload in video games[C]//Proceedings of the IEEE Second Workshop on Eye Tracking and Visualization. Piscataway: IEEE Press, 2016: 60-64.

[14] PRABHAKAR G, MUKHOPADHYAY A, MURTHY L, et al. Cognitive load estimation using ocular parameters in automotive[J]. Transportation Engineering, 2020, 2: 100008.

[15] DE RIVECOURT M, N KUPERUS M, J POST W, et al. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight[J]. Ergonomics, 2008, 51(9): 1295-1319.

[16] 李鹏杰, 姚志, 王萌, 等. 心率变异性在手控交会对接操作脑力负荷评价中的应用[J]. 人类工效学, 2013, 19(3): 1-5.
LI P J, YAO Z, WANG M, et al. The application of heart rate variability in the evaluation of mental workload during manual controlled rendezvous and docking[J]. Chinese Journal of Ergonomics, 2013, 19(3): 1-5(in Chinese).

[17] GRASSMANN M, VLEMINCX E, VON LEUPOLDT A, et al. Individual differences in cardiorespiratory measures of mental workload: an investigation of negative affectivity and cognitive avoidant coping in pilot candidates[J]. Applied Ergonomics, 2017, 59: 274-282.

[18] AYGUN A, LYU B Y, NGUYEN T, et al. Cognitive workload assessment via eye gaze and EEG in an interactive multi-modal driving task[C]//Proceedings of the International Conference on Multimodal Interaction. New York: ACM, 2022: 337-348.

[19] OPPELT M P, FOLTYN A, DEUSCHEL J, et al. ADABase: a multimodal dataset for cognitive load estimation[J]. Sensors, 2022, 23(1): 340.

[20] CHEN J, XUE L, LIU Z C. A pilot workload evaluation method based on EEG data and physiological data[C]//Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing. Piscataway: IEEE Press, 2020: 1-6.

[21] 王煜文, 王盛, 韩明秀, 等. 基于机器学习的飞行员脑力负荷评估研究进展[J]. 载人航天, 2021, 27(6): 789-796.
WANG Y W, WANG S, HAN M X, et al. Research progress of mental workload assessment based on machine learning in pilots[J]. Manned Spaceflight, 2021, 27(6): 789-796(in Chinese).

[22] MEDEIROS J, COUCEIRO R, DUARTE G, et al. Can EEG be adopted as a neuroscience reference for assessing software programmers' cognitive load?[J]. Sensors, 2021, 21(7): 2338.

[23] LOFT S, BOLLAND S, HUMPHREYS M S, et al. A theory and model of conflict detection in air traffic control: incorporating environmental constraints[J]. Journal of Experimental Psychology Applied, 2009, 15(2): 106-124.

[24] MARCHITTO M, DI STASI L L, CAÑAS J J. Ocular movements under taskload manipulations: influence of geometry on saccades in air traffic control simulated tasks[J]. Human Factors and Ergonomics in Manufacturing & Service Industries, 2012, 22(5): 407-419.

[25] NEAL A, KWANTES P J. An evidence accumulation model for conflict detection performance in a simulated air traffic control task[J]. Human Factors, 2009, 51(2): 164-180.

[26] SALVUCCI D D, GOLDBERG J H. Identifying fixations and saccades in eye-tracking protocols[C]//Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00. New York: ACM, 2000: 71-78.

[27] ANDERSSON R, LARSSON L, HOLMQVIST K, et al. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms[J]. Behavior Research Methods, 2017, 49(2): 616-637.

[28] PAN T, WANG H B, SI H Q, et al. Identification of pilots' fatigue status based on electrocardiogram signals[J]. Sensors, 2021, 21(9): 3003.

[29] STATSENKO Y, HABUZA T, GORKOM K N, et al. Applying the

inverse efficiency score to visual-motor task for studying speed-accuracy performance while aging[J]. Frontiers in Aging Neuroscience, 2020, 12: 574401.

[30] DING Y, CAO Y Q, DUFFY V G, et al. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning[J]. Ergonomics, 2020, 63(7): 896-908.

[31] 张奇良, 杨坤华, 曲行达, 等. 基于多模态生理信号的驾驶人脑力负荷评估[J]. 深圳大学学报 (理工版), 2022, 39(3): 278-286.
ZHANG Q L, YANG K H, QU X D, et al. Evaluation of drivers' mental workload based on multi-modal physiological signals[J]. Journal of Shenzhen University (Science and Engineering), 2022, 39(3): 278-286(in Chinese).

[32] 郑玲, 乔旭强, 倪涛, 等. 基于多维信息特征分析的驾驶人认知负荷研究[J]. 中国公路学报, 2021, 34(4): 240-250.
ZHENG L, QIAO X Q, NI T, et al. Driver cognitive loads based on multi-dimensional information feature analysis[J]. China Journal of Highway and Transport, 2021, 34(4): 240-250(in Chinese).

Cognitive load assessment method for multifactorial flight conflict detection

YANG Yue*, MA Bokai, CHENG Long

(College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Real-time monitoring and evaluation of air traffic controllers' cognitive load is of great significance to the operational safety of the air traffic control system. Cognitive load variation in control command, which is reflected by the physiological parameters, facilitates timely discovery of the undesirable working condition that affecting control effectiveness, in order to realize forward movement of the risk control gate. The radar control simulation experimental platform is used to create conflict detection scenarios under various airspace complexity conditions. A repeated within-subjects measurement experimental scheme is designed with three factors: minimum distance (4 km, 12 km, 16 km), convergence angle (45°, 90°, 135°), and speed characteristics (fast speed priority, same speed, slow speed priority). The effect of different complexity factors on the cognitive load and the changing laws of physiological indexes are studied based on multifactorial analysis of variance through collecting the subjects' eye movement and electrocardiogram data, so as to select feature physiological indexes that can effectively express airspace complexity factors. Based on this, the individuals' cognitive load is assessed using three machine learning algorithms: random forest (RF), support vector machine (SVM), and long short-term memory network (LSTM). The results show that among different levels of the same type factor, the cognitive load is highest when the minimum distance (MD) is closed to the warning interval and the convergence angle (CA) is an acute angle, as well as lowest when the speed characteristics (SC) is the same speed respectively. Nine physiological indexes can be used as feature indexes to effectively express cognitive load at different levels of MD, including number of fixations (NrF), number of saccades (NrS), mean saccade duration (SD), average saccade amplitude (SA), average saccade peak velocity (SPV), blinking frequency (BF), pupil diameter (PD), mean respiratory rate (RR) and power ratio of low-frequency to high-frequency in heart rate variability (LF/HF). The assessment accuracy of the SVM model based on a single-modal eye movement signal is 94.69%, which is higher than the single-modal assessment model with electrocardiographic signal and the dual-modal assessment model with eye movement and electrocardiographic signals. Removing strongly correlated feature indexes will affect model performance to some extent.

Keywords: conflict detection; cognitive load; eye movement index; electrocardiographic index; machine learning; model performance

Received: 2023-12-15; **Accepted**: 2024-05-10; **Published Online**: 2024-05-22 11:02

URL: link.cnki.net/urlid/11.2625.V.20240521.1318.001

Foundation item: Education Commission Research Program of Tianjin (2023KJ234)

* **Corresponding author**. E-mail: yueyang0124@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0828

矩形高层建筑顺风向脉动风荷载空间相关性

袁家辉[1,2]，陈水福[1,*]，夏俞超[3]，刘奕[1]

(1. 浙江大学 建筑工程学院，杭州 310058；2. 中冶南方城市建设工程技术有限公司，武汉 430063；
3. 浙江理工大学 建筑工程学院，杭州 310018)

摘 要：为探讨建筑深宽比和来流湍流特性对矩形高层建筑顺风向脉动风荷载空间相关性的影响，对深宽比为1/9～9的矩形高层建筑在4种风场中进行了同步测压风洞试验。基于试验结果，分析建筑深宽比、湍流强度和湍流积分尺度对顺风向脉动风荷载竖向相关系数和相干函数的影响；通过非线性最小二乘法，拟合得到了适用于深宽比为1/9～9的矩形高层建筑竖向相关性数学模型。结果表明：顺风向脉动风荷载相关系数同时受到建筑深宽比和测点层高差影响，相关系数随测点层高差增大呈指数衰减；顺风向脉动相干函数随频率增大呈指数衰减，且衰减速率与高差和平均风速的比值大致成正比；对于不同深宽比建筑，湍流积分尺度和湍流强度对顺风向脉动风荷载相关性的影响不同；本文针对不同风场提出的矩形高层建筑顺风向脉动风荷载相关系数和相干函数公式和试验结果吻合良好，可为建筑结构设计及荷载规范修订提供参考。

关 键 词：矩形高层建筑；脉动风荷载；空间相关性；风洞试验；深宽比
中图分类号：TU973.213
文献标志码：A **文章编号**：1001-5965(2026)03-0772-12

空间中不同位置的脉动风荷载一般不会同时达到最大值，空间距离越远，脉动风荷载同时达到最大值的可能性就越低，这种性质被称为脉动风荷载的空间相关性。在时域内一般采用相关系数描述空间两点的相关性，在频域内则采用相干函数来描述。高层建筑风致响应计算中，需要确定不同位置风荷载的空间相关性，即获得脉动风荷载相关系数或相干函数的数学模型。

20世纪60年代，Davenport[1]提出顺风向风致响应的完整计算方法，建议使用顺风向脉动风速相干函数代替脉动风荷载相干函数，并指出脉动风速相干函数与两点平均风速及间距有关，提出指数形式的相干函数经验公式。Shiotani等[2]在试验观测的基础上建议使用与频率无关的相干函数表达式，式中考虑了湍流积分尺度的影响。文献[2]提出的经验公式形式简单，至今仍被中国建筑结构荷载规范所采用[3]。Kareem[4]通过对比顺风向脉动风速和脉动风荷载的功率谱发现，脉动风荷载比脉动风速的相关性更大，但是用脉动风速的相关性代替风荷载的相关性依然能在计算中取得良好的结果。因为假定迎风面和背风面风荷载同步脉动会使计算结果偏大，该假设产生的误差和使用脉动风速相关性代替风荷载相关性产生的误差在一定程度上相互抵消。文献[1-2]提出的相干函数存在如下问题：①脉动风速相干函数恒大于0，这和观测结果是不完全相符的；②当频率为0时，无论两点距离多远，相干函数值恒等于1，这和“空间两点距离远大于湍流积分尺度时，脉动风速不相干”的假设相互

收稿日期：2023-12-21；**录用日期**：2024-01-28；**网络出版时间**：2024-03-19 07:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20240318.0957.001
基金项目：国家自然科学基金(51878607)
* **通信作者**. E-mail：csf@zju.edu.cn

引用格式：*袁家辉，陈水福，夏俞超，等. 矩形高层建筑顺风向脉动风荷载空间相关性[J]. 北京航空航天大学学报，2026，52(3)：772-783. YUAN J H，CHEN S F，XIA Y C，et al. Spatial correlation of along-wind fluctuating wind loads on rectangular high-rise buildings[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：772-783 (in Chinese).*

矛盾。

针对以上问题, Krenk[5] 通过引入频率修正, 提出基于 von Karman 谱的相干函数模型。张建胜等[6] 分别使用 Davenport、Shiotani 和 Krenk 相干函数经验公式进行风致响应计算。结果表明: 使用 Shiotani 相干函数的计算结果相较于使用 Davenport 相干函数的计算结果偏大约 60%, 使用 Krenk 相干函数的计算结果为使用 Davenport 相干函数的计算结果的 75%～90%。欧洲钢结构建设规范 ECCS[7] 给出的脉动风速相干函数在 Shiotani 公式基础上考虑了湍流积分尺度随高度的变化。顾明等[8] 在 Davenport 相干函数的基础上, 通过多个高层建筑同步测压风洞试验结果, 分析了脉动阻力在 B 类和 D 类风场中的竖向相干特性, 给出了竖向相干函数衰减系数的经验公式。黄东梅等[9-11] 对某 492 m 超高层建筑进行同步测压风洞试验, 分析了建筑不同高度处顺风向脉动风荷载的相干函数, 提出对应的经验公式。曾加东等[12] 对深宽比为 2 的矩形高层建筑进行同步测压风洞试验, 在 Krenk 模型的基础上提出了考虑湍流积分尺度影响的顺风向脉动风荷载相干函数模型。Li 等[13-14] 通过高频天平测力风洞试验, 分析了格构式塔架顺风向脉动风荷载的空间相关性, 给出了对应的相干函数经验公式。Zeng 等[15] 对深宽比为 1/4～4 的矩形高层建筑进行同步测压风洞试验, 分析了相关系数和相干函数特性, 提出一种考虑迎风面宽度、模型高宽比和频率修正的相干函数模型。

以往学者所提出的相干函数模型都是根据特定建筑模型或小范围深宽比矩形高层建筑模型的风洞试验数据得到, 无法排除特殊建筑截面形式对相干函数的影响, 对大深宽比的建筑也可能不适用。Shiotani[2]、ECCS[7] 的相干函数模型形式简单, 但本质其实是时域内的相关系数, 不能反映相干函数随频率变化的特性。

为获得适用范围更广泛、实用性更佳的高层建筑顺风向脉动风荷载相关性的数学模型, 给建筑结构荷载规范的修订提供参考。本文探讨了正交风向下深宽比 1/9～9 的矩形截面高层建筑顺风向脉动风荷载的相关系数及相干函数特性, 总结了相关系数和相干函数随建筑深宽比和风场类别的变化规律, 然后, 通过最小二乘法和归纳分析方法获得相关系数及相干函数的数学模型。

1　风洞试验

1.1　风场模拟

本次试验按照工程科学数据库 (engineering sciences data unit, ESDU) 模拟了 4 种风场, 分别为 O1、S1、O2 和 S2。S 类风场的湍流强度大于 O 类, 1 类风场的湍流积分尺度大于 2 类。通过在风洞入口处放置格栅、尖劈和挡板等一系列装置来模拟不同特征的湍流。放置挡板可以产生更大尺寸的湍流, 实现对湍流强度和湍流积分尺度的同步控制[16]。在转盘上散布螺母可以维持建筑模型周边近地面高度处的湍流强度[17]。风速采样频率为 1 250 Hz, 平均风速剖面按照 ESDU-85020[18] 推荐的对数率公式模拟, 理论值计算式为

$$\frac{\overline{U}_z}{\overline{U}_{10}}=\frac{\ln(z/z_0)}{\ln(10/z_0)}+\frac{86.25f_cz}{\overline{U}_{10}} \tag{1}$$

式中: $\overline{U}_{10}$为 10 m 高度处的平均风速; z 为所在高度; z_0 为地表粗糙长度; f_c 为科里奥利力。O1、S1、O2 和 S2 风场在模型顶部的缩尺平均风速分别为 8.86 m/s、8.32 m/s、8.07 m/s 和 8.11 m/s。

湍流强度剖面按照 ESDU-82026[19] 推荐的公式模拟, 理论值计算式为

$$I_{uz}=\frac{7.5\eta u_*[0.538+0.09\ln(z/z_0)]^p}{[1+0.156\ln(u_*/f_cz_0)]\overline{U}_z} \tag{2}$$

式中: η、p 为拟合参数; u_*为摩擦速度。O1、S1、O2 和 S2 风场在模型顶部的湍流强度分别为 0.124、0.187、0.122 和 0.170。

湍流积分尺度按 ESDU-74031[20] 建议的公式模拟:

$$L_u=\overline{U}_z\int_0^{\tau_0}\rho_{uu}(\tau)\mathrm{d}\tau \tag{3}$$

式中: ρ_{uu}为脉动风速自相关系数; τ_0 为自相关系数收敛到 0 时对应的时差; O1、S1、O2 和 S2 风场在模型顶部的缩尺湍流积分尺度分别为 1.328 m、1.545 m、0.504 m 和 0.656 m。

图 1 为 O1、S1、O2 和 S2 风场的平均风速和湍流度剖面, 由图可知, 试验值和理论值吻合良好; 图 1 中同时给出了各风场足尺地表粗糙长度 z_0 和指数率形式的剖面参数 α。

脉动风速谱模拟采用 ESDU-74031[20] 建议的 von-Karman 谱模拟:

$$\begin{cases}\dfrac{fS_{uu}(f)}{\sigma_u^2}=\dfrac{4f_u}{(1+70.8f_u^2)^{5/6}}\\ f_u=\dfrac{fL_u}{\overline{U}_z}\end{cases} \tag{4}$$

式中: f为频率; S_{uu}为脉动风速功率谱密度; σ_u为脉动风速标准差。图 2 给出了本次试验 4 种风场在 z=0.762H 处 (驻点位置) 的脉动风速谱试验值和理论值, H 为模型高度, 可见二者吻合良好。

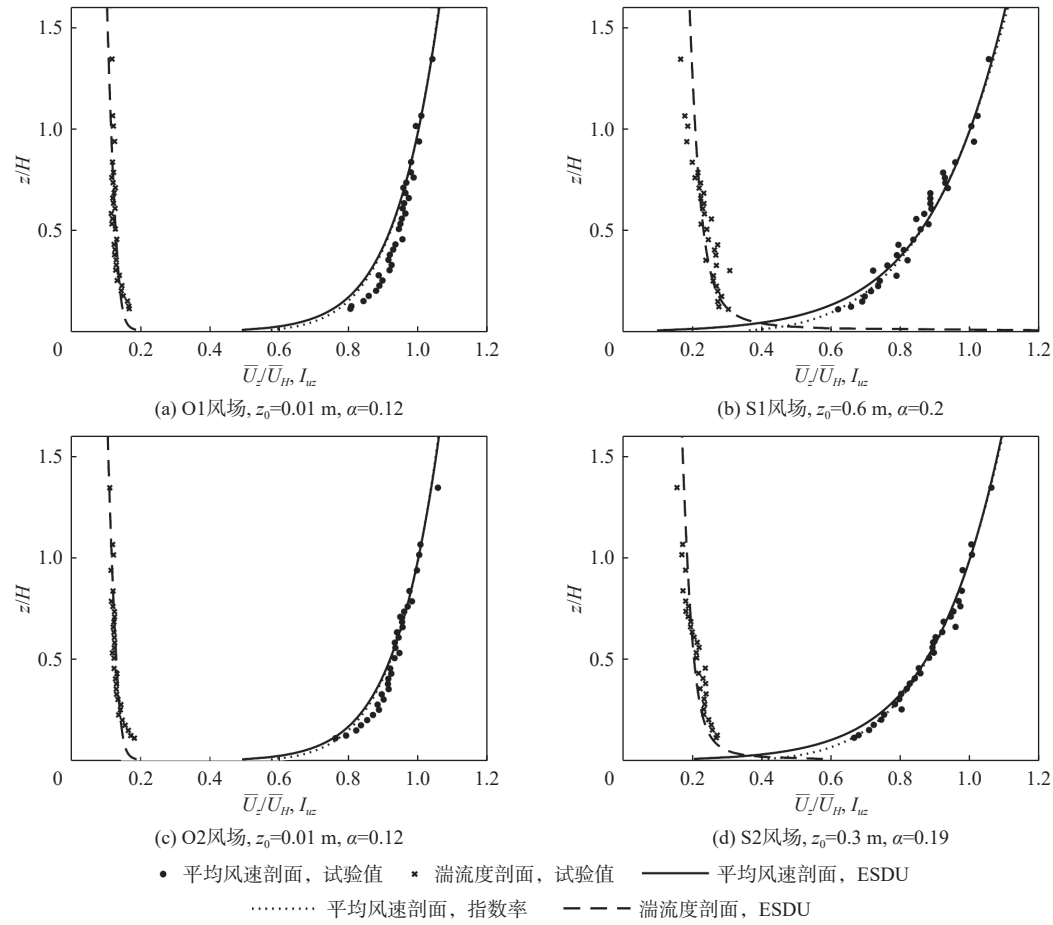


图 1 平均风速和湍流度剖面

Fig. 1 Mean velocity and turbulence intensity profiles

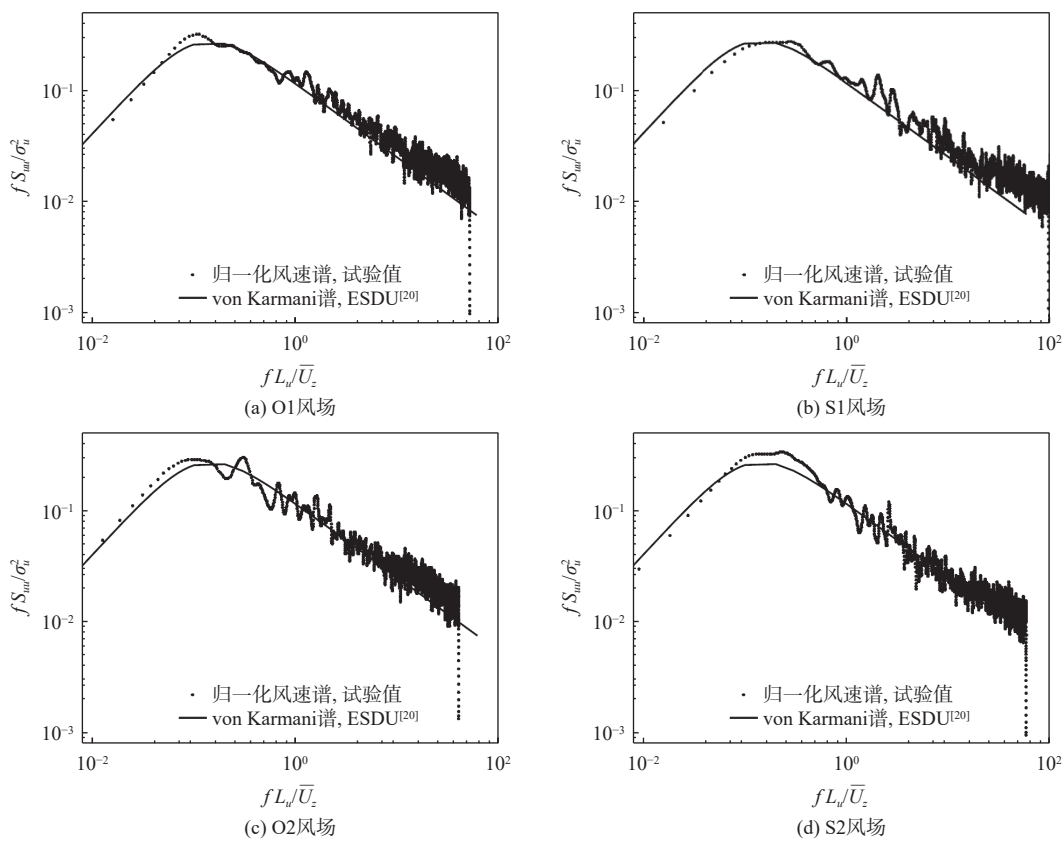


图 2 z=0.762H 高度处脉动风速谱

Fig. 2 Fluctuating velocity spectra at z=0.762H

1.2　试验模型

同步测压试验采用缩尺比为 1:200 的刚性节段模型, 模型高为 0.5 m, 宽为 0.06 m, 长为 0.06～0.54 m。模型共有 12 段, 1～6 段的长宽比为 0.5, 7～12 段的长宽比为 1.0。模型沿高度方向布置了 7 个测点层, 分别位于 0.1*H*、0.3*H*、0.5*H*、0.65*H*、0.8*H*、0.9*H* 和 0.98*H* 高度处, 测点层从低到高分别编号为 1～7, 各测点层的布置方式相同, 模型平面图与测点布置如图 3 所示。定义模型的深宽比为 *D*/*B*, 其中, *B* 为迎风面宽度, *D* 为侧风面深度。短边迎风时深宽比大于等于 1, 长边迎风时深宽比小于等于 1。通过不同的拼接方式可以得到深宽比为 1/9～9 的试验模型, 具体拼接方式和试验模型参数如表 1 所示。当 *D*/*B*=1/9 时, 试验时阻塞效应最大。根据模型尺寸和风洞测试段尺寸计算的阻塞比小于 5%, 本次试验可以忽略阻塞效应[21]。针对每种深宽比模型进行了测压试验, 获得了各测点的风压系数时程, 风压采样频率为 400 Hz。图 4 给出了同步测压试验照片。更详细的试验参数见文献 [21-23]。

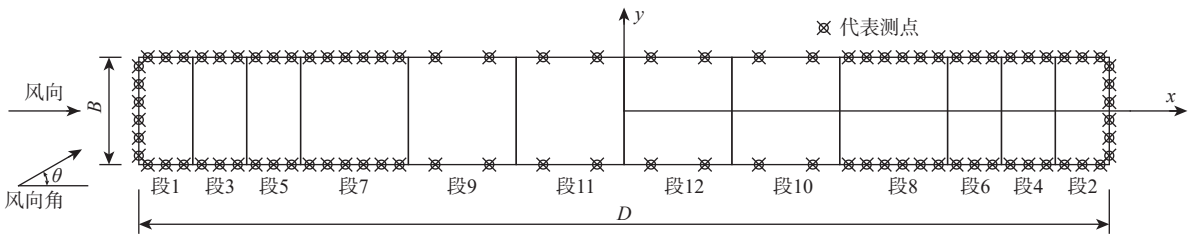


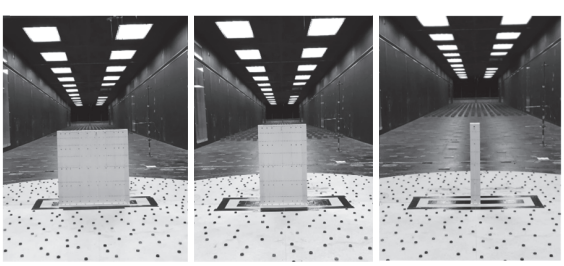
图 3　模型平面图与测点布置

Fig. 3　Plan sketch of model and layout of pressure taps

表 1　试验模型参数

Table 1　Parameters of experiment models

工况	风场类型	深宽比	拼接方式(从左至右)	测压点个数
1	O1	1/9, 9	段1, 段3, 段5, 段7, 段9, 段11, 段12, 段10, 段8, 段6, 段4, 段2	616
2	O1, O2, S1, S2	1/8, 8	段1, 段3, 段5, 段7, 段9, 段11, 段10, 段8, 段6, 段4, 段2	588
3	O1	1/7, 7	段1, 段3, 段5, 段7, 段9, 段10, 段8, 段6, 段4, 段2	560
4	O1	1/6, 6	段1, 段3, 段5, 段7, 段10, 段8, 段6, 段4, 段2	532
5	O1, O2, S1, S2	1/5, 5	段1, 段3, 段5, 段7, 段8, 段6, 段4, 段2	504
6	O1	1/4, 4	段1, 段3, 段5, 段7, 段6, 段4, 段2	420
7	O1, O2, S1, S2	1/3, 3	段1, 段3, 段5, 段6, 段4, 段2	336
8	O1	1/2.5, 2.5	段1, 段3, 段5, 段4, 段2	294
9	O1, O2, S1, S2	1/2, 2	段1, 段3, 段5, 段2	252
10	O1	1/1.5, 1.5	段1, 段4, 段2	210
11	O1, O2, S1, S2	1	段1, 段2	168



(a) *D*/*B*=1/8　(b) *D*/*B*=1/5　(c) *D*/*B*=1

图 4　同步测压试验照片

Fig. 4　Synchronization pressure test photos

2　脉动风荷载空间相关性分析

2.1　相关系数特性

顺风向脉动风荷载相关系数计算式为

$$c_{\mathrm{D}}(z_i,z_j)=\frac{\mathrm{cov}[F_{\mathrm{D}}(z_i),F_{\mathrm{D}}(z_j)]}{\sigma_{F_{\mathrm{D}}}(z_i)\sigma_{F_{\mathrm{D}}}(z_j)} \tag{5}$$

式中: $F_{\mathrm{D}}(z_i)$ 和 $F_{\mathrm{D}}(z_j)$ 分别为 z_i 和 z_j 高度处的测点层顺风向气动力, $\sigma_{F_{\mathrm{D}}}(z_i)$和$\sigma_{F_{\mathrm{D}}}(z_j)$分别为其标准差。

图 5 为 O1 风场下顺风向脉动风荷载相关系

数。由图可知，顺风向脉动风荷载相关系数同时受到建筑深宽比和测点层高差影响。相关系数随测点层高差增大呈指数形式衰减。随建筑深宽比增大，相关系数随测点层高差的衰减速率增大。在 $D/B\leqslant 1$ 时，相关系数衰减速率增速明显；在 $D/B>1$ 时，相关系数衰减速率增速减慢。这可能是因为随着建筑深宽比增大，背风面的脉动风荷载受到侧风面不规则流动的干扰。在 $D/B\leqslant 1$ 时，随深宽比增大，背风面面积持续减小，受侧面不规则流动干扰的脉动风荷载面积比率不断增大，相关系数衰减速率不断增大；在 $D/B>1$ 时，随深宽比增大，侧风面面积持续增大，流动更复杂，但背风面面积保持不变，相关系数衰减速率增速减慢。

为考察来流湍流特性对顺风向脉动风荷载相关系数的影响，定义湍流积分尺度影响系数为 O1 与 O2 风场下顺风向脉动风荷载相关系数的比值，湍流强度影响系数为 S1 与 O1 风场下顺风向脉动风荷载相关系数比值，计算式为

$$\begin{cases} R_{\mathrm{D}}^{L_u}=\dfrac{c_{\mathrm{D}}^{\mathrm{O1}}}{c_{\mathrm{D}}^{\mathrm{O2}}} \\ R_{\mathrm{D}}^{I_u}=\dfrac{c_{\mathrm{D}}^{\mathrm{S1}}}{c_{\mathrm{D}}^{\mathrm{O1}}} \end{cases} \tag{6}$$

图 6 为湍流特性对顺风向脉动风荷载相关系数的影响。由图 6 可知，对于不同深宽比和不同高差的情况下，湍流积分尺度影响系数和湍流强度影响系数都大于 1。由图 6(a) 可知，随测点层高差增大，湍流积分尺度影响系数增大；随建筑深宽比增大，湍流积分尺度影响系数先增大后减小。当 $D/B=5$ 时，湍流积分尺度对顺风向脉动风荷载相关系数的影响最大。由图 6(b) 可知，湍流强度影响系数大致也随测点层高差增大；但湍流强度影响系数与建筑深宽比无明显关系。

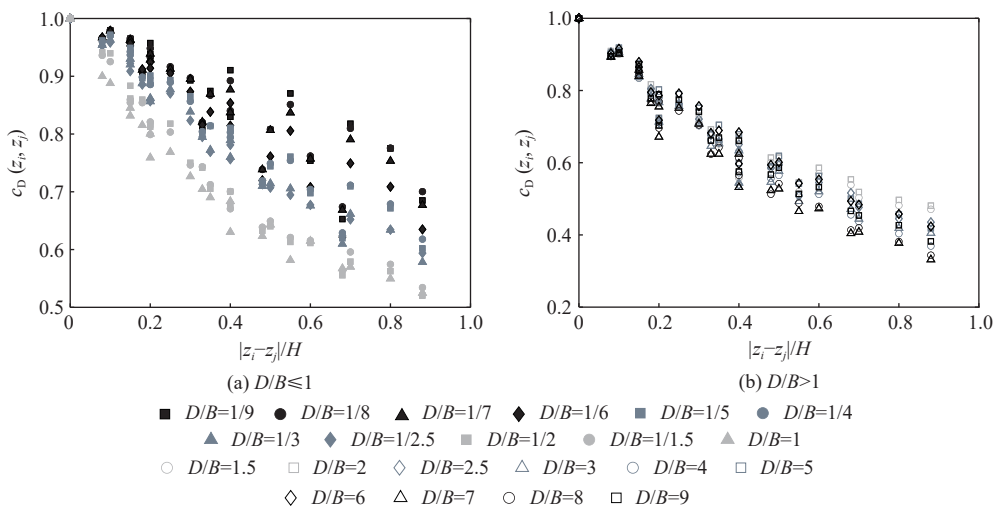


图 5 O1 风场下顺风向脉动风荷载相关系数

Fig. 5 Correlation coefficients of along-wind fluctuating wind loads in O1 wind field

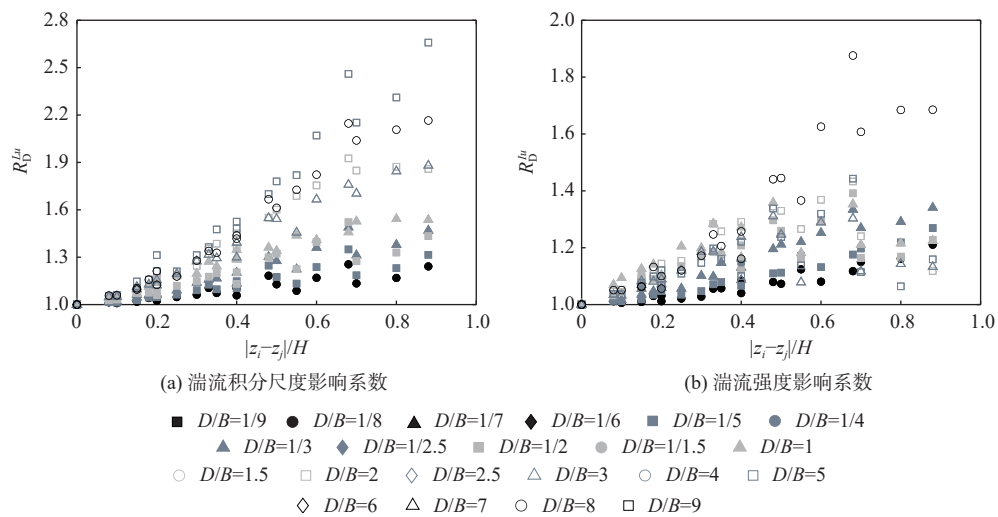


图 6 湍流特性对顺风向脉动风荷载相关系数的影响

Fig. 6 Effect of turbulence characteristics on correlation coefficients of along-wind fluctuating wind loads

2.2　相干函数特性

顺风向脉动风荷载相干函数计算式为

$$\mathrm{coh}_{\mathrm{D}}(z_i,z_j,f)=\frac{\mathrm{Re}[S_{\mathrm{D}}(z_i,z_j,f)]}{\sqrt{S_{\mathrm{D}}(z_i,f)\cdot S_{\mathrm{D}}(z_j,f)}} \tag{7}$$

式中: $\mathrm{Re}[S_{\mathrm{D}}(z_i,z_j,f)]$ 为 z_i 和 z_j 高度处测点层顺风向脉动风荷载互功率谱实部; $S_{\mathrm{D}}(z_i,f)$ 和 $S_{\mathrm{D}}(z_j,f)$ 分别为 z_i 和 z_j 高度处测点层顺风向脉动风荷载自功率谱。

图 7 为 O1 风场下顺风向脉动风荷载相干函数, 图中以 1# vs 4#、2# vs4#、3# vs 4#、4# vs 5#、4# vs 6#和 4# vs 7#测点层的 6 组相干函数为例进行分析, 其中, 1#vs4#表示测点层 1 位置的风荷载与测点层 4 位置的风荷载之间的相干函数, 其他含义类同。6 组相干函数的测点层高差 Δz 分别为: 0.55H、0.35H、0.15H、0.15H、0.25H 和 0.33H; 测点层平均风速 $\overline{U}$ 分别为: 7.83 m/s、8.32 m/s、8.47 m/s、8.62 m/s、8.68 m/s 和 8.70 m/s。由图 7 可知, 顺风向脉动风荷载相干函数存在以下特性: ①所有的相干函数都随频率增大呈指数形式衰减, 频率为 0 时, 相干函数值接近 1; ②当 D/B<1/3 时, 相干函数衰减速率较小, 当 D/B≥1/3 时, 相干函数衰减速率迅速增大; ③对于 D/B≥1/3 的建筑, 测点层高差较大时 (1# vs 4#、2# vs 4#和 4# vs 7#), 相干函数先迅速衰减, 然后, 在 0.1 附近波动, 测点层高差较小时 (3# vs 4#、4# vs 5#), 相干函数持续缓慢衰减; ④相干函数衰减速率大致与 $\Delta z/\overline{U}$ 成正比, 当 $\Delta z/\overline{U}$ 接近时, 相干函数的衰减速率也接近。例如: 3# vs 4#测点层, $\Delta z/\overline{U}=8.85\times10^{-3}$ s; 4# vs 5#测点层, $\Delta z/\overline{U}=8.75\times10^{-3}$ s。

图 8 为不同风场下顺风向脉动风荷载相干函数, 图中以 3# vs 4#测点层的相干函数为例进行分析。由图 8 可知, 不同风场下顺风向脉动风荷载相干函数变化趋势基本一致, 数值上存在差别。当 D/B≤1 时, 来流的湍流特性对相干函数的影响规律不明显; 当 D/B>1 时, O2 风场和 S2 风场的相干函数小于 O1 风场和 S1 风场, 这说明湍

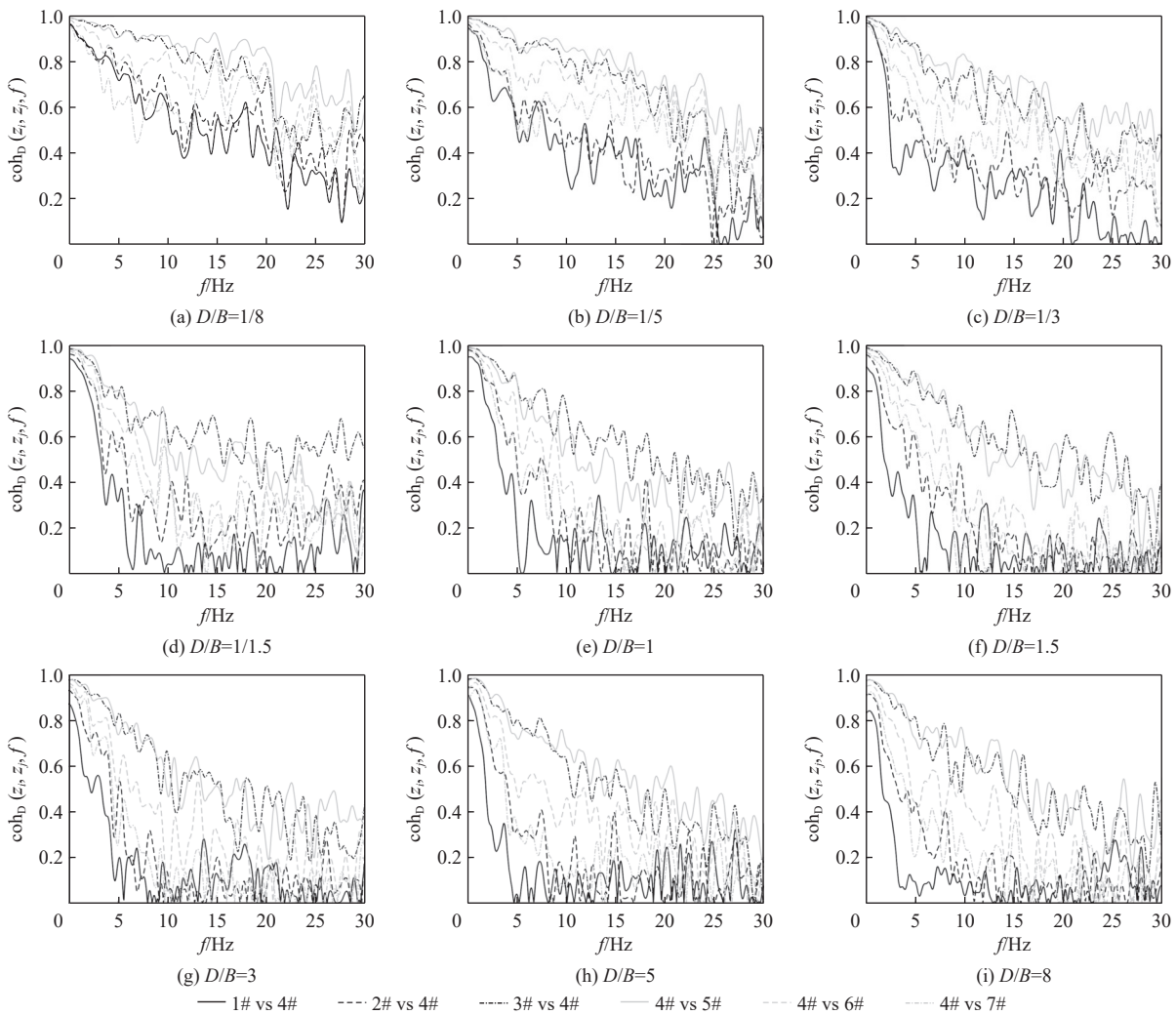


图 7　O1 风场下顺风向脉动风荷载相干函数

Fig. 7　Coherence functions of along-wind fluctuating wind loads in O1 wind field

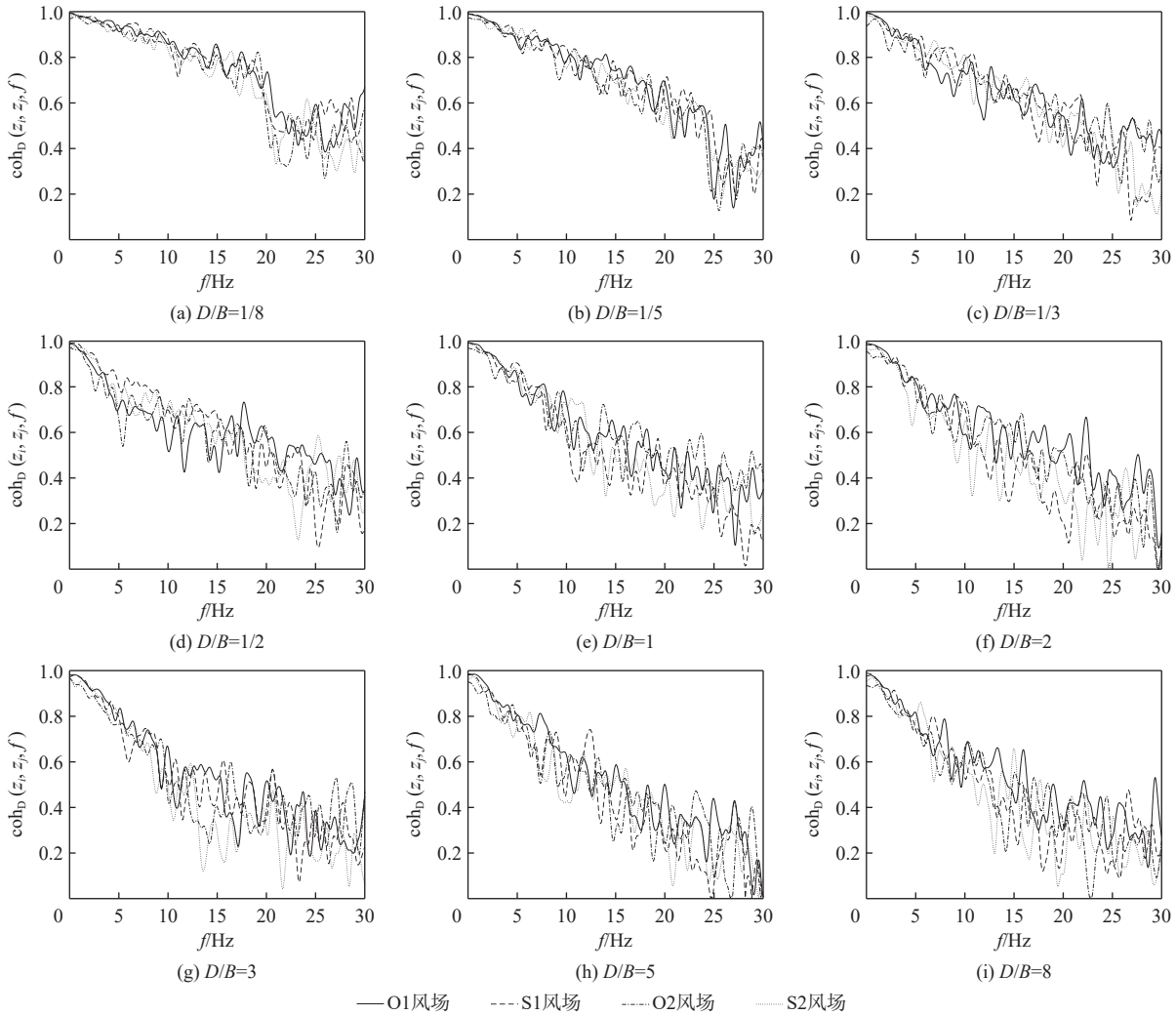


图 8 不同风场下顺风向脉动风荷载相干函数

Fig. 8 Coherence functions of along-wind fluctuating wind loads in different wind fields

流积分尺度会对大深宽比建筑的相干函数产生影响。

3 脉动风荷载相关性数学模型

3.1 相关系数数学模型

为给建筑结构设计及荷载规范修订提供定量参考，以本次风洞试验的数据拟合顺风向脉动风荷载相关系数与相干函数公式。表 2 给出了典型顺风向相干函数经验公式，作为本文拟合公式形式的参照对象。

从 2.1 节的分析可知，顺风向脉动风荷载相关系数和测点层高差、建筑深宽比及风场类别均有关系，且随测点层高差增大呈指数衰减。经过多次试算和比较，本文选取如下形式作为顺风向脉动风荷载相关系数的拟合式：

$$c_{\mathrm{D}}(z_i,z_j)=\exp[-p_1(\Delta z/B)^{p_2}] \tag{8}$$

式中：p_1、p_2 为待拟合参数；$\Delta z=|z_i-z_j|$为测点层高差；B 为迎风面投影宽度。对于不同风场类别下、不同深宽比建筑，拟合式的参数计算式为

$$p_1=-0.043t^3+0.314t^2-0.665t+0.541+(-0.015t^3+0.102t^2-0.202t+0.145)s^{-0.054t^2+0.27t-1.27} \tag{9}$$

$$p_2=\begin{cases}(0.174t^3-1.13t^2+3.03t-2.3)s-0.068t^3+0.502t^2-1.11t+1.77 & s\leqslant 1\\(-0.057t^3+0.407t^2-0.882t+0.629)\ln s+0.09t^3-0.667t^2+1.56t-0.224 & s>1\end{cases} \tag{10}$$

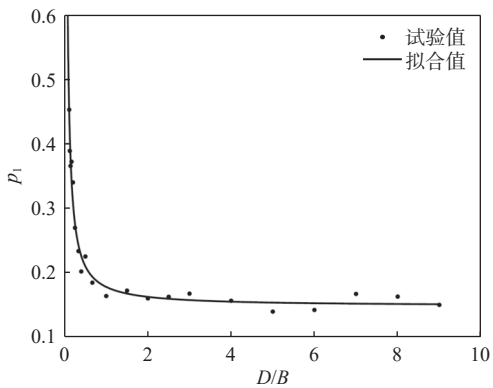
其中：t 为风场类别，对于 O1、S1、O2 和 S2 风场分别取 1、2、3 和 4；s 为建筑深宽比，取值范围为 1/9～9。

图 9 为 O1 风场下式 (8) 中拟合参数与式 (9) 和式 (10) 计算值的比较，由图 9 可知，公式计算值与离散的拟合参数误差很小。为检验式 (9) 和式 (10) 的准确性，图 10 给出了 O1 风场下顺风向脉动风荷载相关系数拟合曲线与试验值比较。由图 10 可知，拟合曲线和试验值吻合良好，说明式 (8)～式 (10) 组成的闭合求解公式可以反映顺风向脉动风荷载相关系数随高差的变化关系。

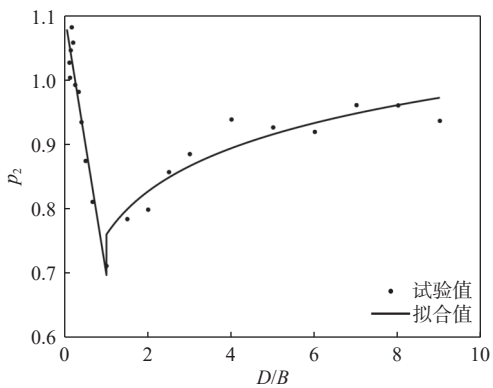
表 2　典型顺风向相干函数经验公式

Table 2　Empirical formula of typical along-wind coherence function

顺风向相干函数公式	参数取法
$\mathrm{coh}(z_i,z_j,f)=\exp\left(-c_z\frac{f\lvert z_i-z_j\rvert}{\overline{U}}\right)$	$c_z=7,\overline{U}=(\overline{U}_i+\overline{U}_j)/2$
$\mathrm{coh}(z_i,z_j)=\exp\left(-\frac{\lvert z_i-z_j\rvert}{L_{uz}}\right)$	$L_{uz}=60$
$\mathrm{coh}(z_i,z_j)=\exp\left(-\frac{\lvert z_i-z_j\rvert}{L_{uz}}\right)$	$L_{uz}=\sqrt{37\bar{z}},\bar{z}=(z_i+z_j)/2$
$\mathrm{coh}(z_i,z_j,f)=\left(1-\frac{c_z f^*\lvert z_i-z_j\rvert}{2\overline{U}}\right)\exp\left(-c_z\frac{f^*\lvert z_i-z_j\rvert}{\overline{U}}\right)$	$f^*=\sqrt{f^2+\left(\frac{\overline{U}}{2\pi L}\right)^2},L=1.34L_u$
$\mathrm{coh}(z_i,z_j,f)=\exp\left(-c_z\frac{f\lvert z_i-z_j\rvert}{\overline{U}}\right)$	$c_z=-2.3\frac{\lvert z_i-z_j\rvert}{\bar{z}}+5.1$
$\mathrm{coh}(z_i,z_j,f)=A\exp\left(-c_z\frac{f\lvert z_i-z_j\rvert}{\overline{U}}\right)$	$A=\left\{1.0+27.4\left[\left(\frac{\lvert z_i-z_j\rvert}{H_r}\right)\Big/\left(\frac{\bar{z}}{H_r}\right)^{1.5}\right]^{1.64}\right\}^{-0.253}$ $c_z=4.22\exp\left(-1.22\lvert z_i-z_j\rvert/H_r\right)+3.57\exp\left(-33.9\lvert z_i-z_j\rvert/H_r\right)$ H_r为梯度风高度
$\mathrm{coh}(z_i,z_j,f)=\left(1-\frac{\eta\beta(s)f^*\lvert z_i-z_j\rvert}{\overline{U}}\right)\exp\left(-\beta(s)\frac{f^*\lvert z_i-z_j\rvert}{\overline{U}}\right)$	$f^*=\sqrt{f^2+\left(\frac{\overline{U}}{2\pi L}\right)^2},L=1.34L_u$ $s=\lvert z_i-z_j\rvert B^{\alpha/2}/H^{1+\alpha/2}$, α为风速剖面指数 $\beta(s)=as^2+bs+c$, $a=-66.26$, $b=27.39$, $c=2.354$, $\eta=1/3$
$\mathrm{coh}(z_i,z_j,f)=\exp\left(-c_1\frac{f\lvert z_i-z_j\rvert}{\overline{U}}+c_2\frac{\lvert z_i-z_j\rvert}{\bar{z}}\right)$	$c_1=1.750,c_2=2.152$
$\mathrm{coh}(k',\delta,s)=[1-\eta\delta\alpha(s)k']\exp[-\alpha(s)\delta k']$	$k'=\sqrt{f^2+\left(\frac{\overline{U}}{2\pi L}\right)^2}\Big/\overline{U},L=[\Gamma(1/3)/\sqrt{\pi}(5/6)]L_{ux}$ $\delta=\lvert z_i-z_j\rvert\bar{z}/B(B/L)^{0.6}$, $\alpha(s)=as^2+bs+c$ a, b, c, η取值和建筑深宽比有关



(a) 参数p_1拟合



(b) 参数p_2拟合

图 9　O1 风场下顺风向脉动风荷载相关系数公式拟合参数与计算值比较

Fig. 9　Comparison between fitting and calculated values of correlation coefficient formula parameters in O1 wind field

3.2　相干函数数学模型

从 2.2 节的结果可知, 顺风向脉动风荷载相干函数和频率、测点层高差、平均风速、建筑深宽比及风场类别均有关系, 且随频率增大呈指数衰减。Davenport 指数模型能反映频率、测点层高差和平均风速的影响, 形式又相对比较简单, 便于实际应用。因此, 本文选取 Davenport 指数模型作为顺风向脉动风荷载相干函数的基础拟合式, 再探究衰减系数 c 与建筑深宽比和风场类别等因素的关系。具体拟合计算式为

$$\mathrm{coh_D}(z_i,z_j,f)=\exp\left(-cf\Delta z/\overline{U}\right) \tag{11}$$

式中: c 为衰减系数; $\overline{U}=(\overline{U}_\mathrm{i}+\overline{U}_\mathrm{j})/2$为测点层平均风速。

本次风洞试验模型竖向共有 7 个测点层, 每种工况下的模型存在 21 个相干函数。考虑 O1、S1、O2 和 S2 这 4 种风场及建筑深宽比的变化, 总共拟合了 1 008 个相干函数。图 11 为 O1 风场下部分顺风向脉动风荷载相干函数拟合结果, 图中, 同时还给出了相干函数对应的测点层编号、测点层高差和平均风速, 方形表示试验时的模型, 箭头方向对应风向。由图 11 可知, 衰减系数变化的 Davenport 指数模型能较好地拟合不同深宽比建筑不同测点层之间的顺风向脉动风荷载相干函数。汇总 O1 风场下所有顺风向脉动风荷载相干函数的拟合结果

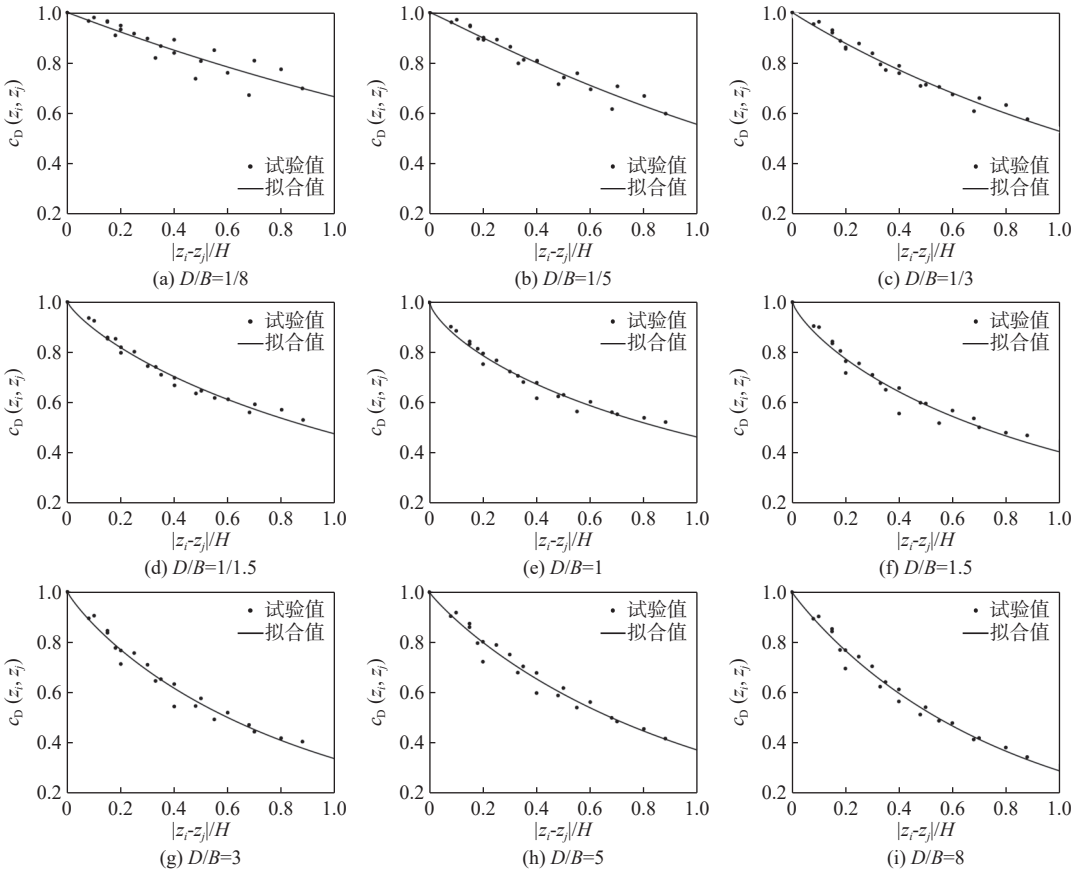


图 10 O1 风场下顺风向脉动风荷载相关系数拟合值与试验值比较

Fig. 10 Comparison of fitting and experimental value of along-wind fluctuating wind loads in O1 wind field

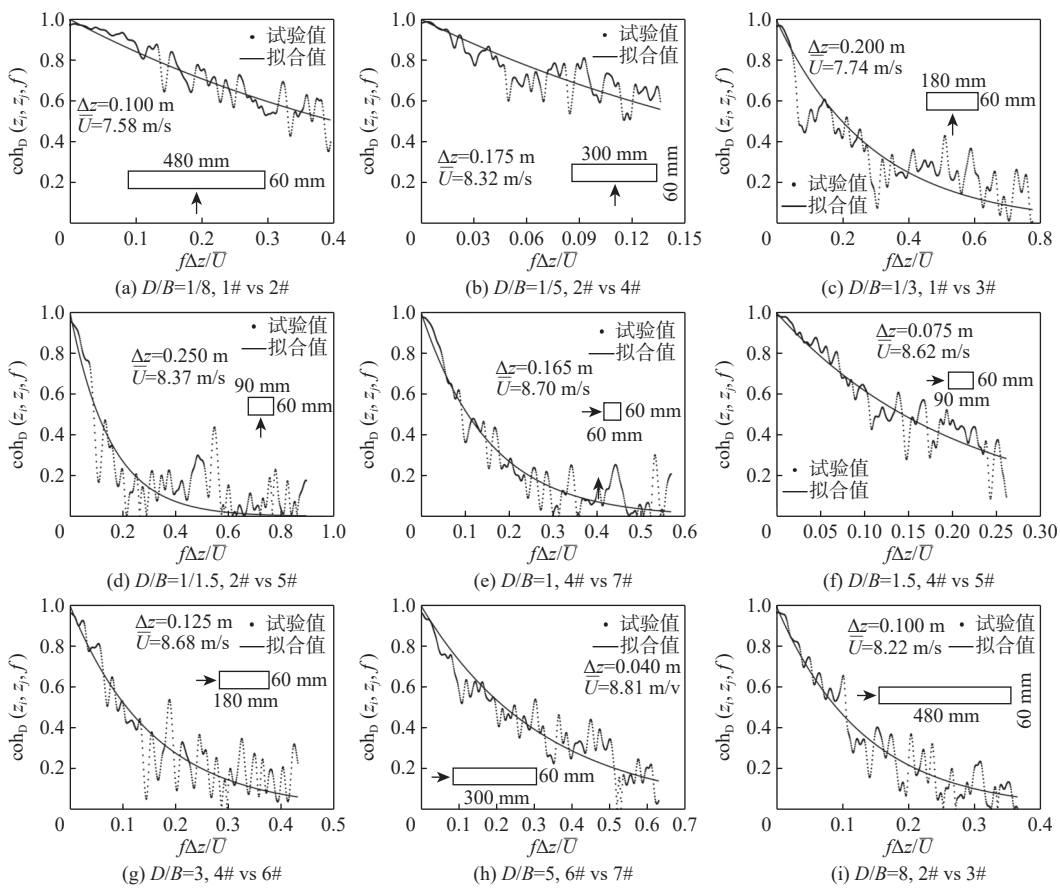


图 11 O1 风场下部分顺风向脉动风荷载相干函数拟合结果

Fig. 11 Some coherence function fitting results of along-wind fluctuating wind loads in O1 wind field

可知, 衰减系数在 0.66～13.77 范围内随建筑深宽比、测点层高差和测点层平均风速变化。最小的衰减系数出现在建筑 D/B=1/9, 测点层 6# vs 7#位置; 最大衰减系数出现在建筑 D/B=9, 测点层 1# vs 4#位置。Davenport 建议的 c=7[1] 和 Simiu 建议的 c=10[24] 都不能反映相干函数衰减系数的变化特性。

图 12 为 O1 风场下顺风向脉动风荷载相干函数衰减系数随建筑深宽比的变化规律。由图 12 可知, 对于不同测点层之间的相干函数, 衰减系数随建筑深宽比的变化规律都是相近的, 衰减系数随建筑深宽比增大, 大致呈对数增长。因此, 将衰减系数按照式 (12) 进行第 2 次拟合:

$$c = p_1 \ln s + p_2 \tag{12}$$

式中: s 为建筑深宽比, 取值范围为 1/9～9。

第 2 次拟合得到不同测点层位置的参数 p_1 和 p_2, 根据分析选用$\Delta z/\bar{z}$为变量来描述 p_1 和 p_2 的变化规律, 图 13 给出了参数 p_1 和 p_2 随$\Delta z/\bar{z}$变化规律, 由图可知, O1 和 S1 风场的衰减系数拟合参数随$\Delta z/\bar{z}$变化规律相近, O2 和 S2 风场的衰减系数拟合参数随$\Delta z/\bar{z}$变化规律相近。这说明, 在湍流积分尺度相近的风场中, 脉动风荷载的相干函数也相近, 因此, 将 O1 和 S1 风场的参数一起拟合, O2 和 S2 风场的参数一起拟合。图 13 为不同风场下衰减系数参数拟合结果, 由图可知, 衰减系数参数拟合结果良好, 拟合得到的计算式为

$$p_1 = (1.15 + 0.66t)\Delta z/\bar{z} + (0.768 - 0.015t) \tag{13}$$

$$p_2 = (1.3 + 0.484t)\ln(\Delta z/\bar{z}) + 6.095 + 1.098t \tag{14}$$

式中: $\Delta z=|z_i-z_j|$为测点层高差, $\bar{z}=(z_i+z_j)/2$ 为测点层平均高度; t 为风场类型参数, O1 和 S1 风场取 0, O2 和 S2 风场取 1。

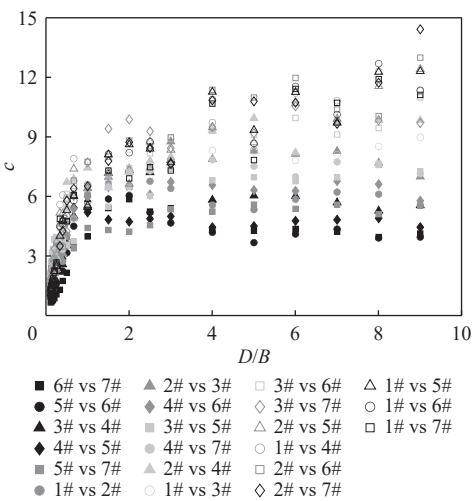


图 12　O1 风场下顺风向脉动风荷载相干函数衰减系数随建筑深宽比的变化规律

Fig. 12　Variation of attenuation coefficient of coherence function with side ratio of building in O1 wind field

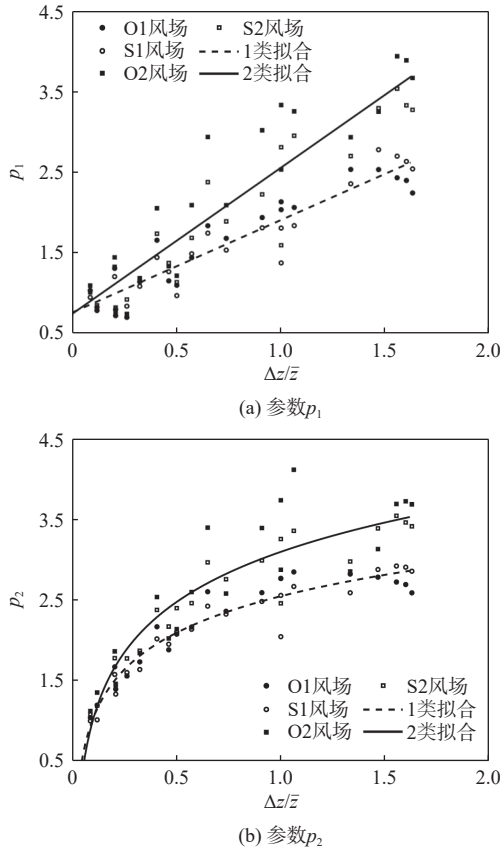


图 13　参数 p_1 和 p_2 随$\Delta z/\bar{z}$变化规律

Fig. 13　Variation law of parameters p_1 and p_2 with $\Delta z/\bar{z}$

4　结　论

本文基于同步测压风洞试验, 研究了 4 种风场下深宽比为 1/9～9 的矩形高层建筑顺风向脉动风荷载的竖相关系数和相干函数特性, 拟合了适用于不同风场、不同深宽比建筑的相关系数和相干函数公式。通过本文的研究, 可以得出以下结论:

1) 顺风向脉动风荷载相关系数同时受到建筑深宽比和测点层高差影响, 相关系数随测点层高差增大呈指数形式衰减; 随建筑深宽比增大, 相关系数衰减速率增大; 在 $D/B\leqslant1$ 时, 相关系数衰减速率增速明显; 在 $D/B>1$ 时, 相关系数衰减速率增速减慢。

2) 增大湍流积分尺度或者湍流强度, 顺风向脉动风荷载相关系数都增大; 测点层高差越大, 湍流积分尺度和湍流强度的影响越明显; 随建筑深宽比增大, 湍流积分尺度的影响先增大后减小。

3) 顺风向脉动相干函数随频率增大呈指数形式衰减, 衰减速率大致与$\Delta z/\overline{U}$成正比; 当 $D/B\geqslant1/3$ 时, 相干函数衰减速率会显著增大; 当 $D/B>1$ 时, 湍流积分尺度增大会导致相干函数增大。

4) 通过最小二乘法拟合得到的相关系数和相干函数系列闭合求解公式和试验值吻合良好, 可为

建筑结构设计及荷载规范修订提供参考。

参考文献（References）

[1] DAVENPORT A G. Gust loading factors[J]. Journal of the Structural Division, 1967, 93(3): 11-34.

[2] SHIOTANI M, AVAI H. Lateral structures of gusts in high winds[C]//Proceedings of the International Conference on the Wind Effect on Buildings and Structures. Cambridge: Cambridge University Press, 1967: 535-555.

[3] 中华人民共和国住房和城乡建设部. 建筑结构荷载规范: GB 50009—2012[S]. 北京: 中国建筑工业出版社, 2012.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Load code for the design of building structures: GB 50009—2012[S]. Beijing: China Architecture & Building Press, 2012(in Chinese).

[4] KAREEM A. Synthesis of fluctuating along wind loads on buildings[J]. Journal of Engineering Mechanics, 1986, 112(1): 121-125.

[5] KRENK S. Wind field coherence and dynamic wind forces[C]// Proceedings of the IUTAM Symposium on Advances in Nonlinear Stochastic Mechanics. Berlin: Springer, 1996: 269-278.

[6] 张建胜, 武岳, 沈世钊. 不同脉动风相干函数对高层建筑风振响应的影响[J]. 振动工程学报, 2009, 22(2): 117-122.
ZHANG J S, WU Y, SHEN S Z. Wind-induced response of high-rise buildings analyzed by different coherence functions of gust[J]. Journal of Vibration Engineering, 2009, 22(2): 117-122(in Chinese).

[7] European Convention for Construction Steelwork. Recommendation for the calculation of wind effects on buildings and structures: ECCS Technical Committee T12[S]. Brussels: European Convention for Construction Steelwork, 1978.

[8] 顾明, 张建国. 高层建筑顺风向脉动荷载相干性研究[J]. 土木工程学报, 2008, 41(11): 18-22.
GU M, ZHANG J G. Coherence analysis of along-wind fluctuating loads on high-rise buildings[J]. China Civil Engineering Journal, 2008, 41(11): 18-22(in Chinese).

[9] 黄东梅, 朱乐东. 超高层建筑层风力空间相关性数学模型: 综合分析法[J]. 土木工程学报, 2009, 42(8): 26-36.
HUANG D M, ZHU L D. Mathematical model of spatial correlation of wind pressure coefficients for super-tall buildings: comprehensive analysis method[J]. China Civil Engineering Journal, 2009, 42(8): 26-36(in Chinese).

[10] 黄东梅, 朱乐东. 超高层建筑层风力空间相干函数研究: 分析归纳法[J]. 土木工程学报, 2010, 43(9): 32-39.
HUANG D M, ZHU L D. Study on spatial correlation functions of wind loads on a super-tall building: analysis and induction method[J]. China Civil Engineering Journal, 2010, 43(9): 32-39(in Chinese).

[11] HUANG D M, ZHU L D, CHEN W, et al. Vertical coherence functions of wind forces and influences on wind-induced responses of a high-rise building with section varying along height[J]. Wind and Structures, 2015, 21(2): 119-158.

[12] 曾加东, 李明水, 李少鹏. 矩形高层建筑顺风向脉动风荷载空间相关性[J]. 哈尔滨工业大学学报, 2017, 49(6): 150-155.
ZENG J D, LI M S, LI S P. Spatial correlation analysis of fluctuating along-wind loads on high-rise buildings with rectangular section[J]. Journal of Harbin Institute of Technology, 2017, 49(6): 150-155(in Chinese).

[13] 邹良浩, 李峰, 梁枢果, 等. 格构式塔架顺风向脉动风荷载空间相关性研究[J]. 湖南大学学报 (自然科学版), 2019, 46(7): 96-103.
ZOU L H, LI F, LIANG S G, et al. Study on spatial correlation of along-wind fluctuating wind load of lattice tower[J]. Journal of Hunan University (Natural Sciences), 2019, 46(7): 96-103(in Chinese).

[14] LI F, ZOU L H, SONG J, et al. Investigation of the spatial coherence function of wind loads on lattice frame structures[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2021, 215: 104675.

[15] ZENG J D, LI M S, LI Z G, et al. Spatial correlation of along-wind fluctuating aerodynamic force acting on large aspect-ratio rectangular prisms[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2022, 224: 104951.

[16] HOLMES J D, OSONPHASOP C. Flow behind two-dimensional barriers on a roughened ground plane, and applications for atmospheric boundary-layer modelling[C]//Proceedings of the 8th Australasian Fluid Mechanics Conference. Newcastle: NSW, 1983.

[17] HO T C E, SURRY D, MORRISH D, et al. The UWO contribution to the NIST aerodynamic database for wind loads on low buildings: Part 1. Archiving format and basic aerodynamic data[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2005, 93(1): 1-30.

[18] Engineering Sciences Data Unit. Characteristics of atmospheric turbulence near the ground. Part II: single point data for strong winds (neutral atmosphere): ESDU-85020[S]. London: Engineering Sciences Data Unit, 1985.

[19] Engineering Sciences Data Unit. Strong winds in the atmospheric boundary layer. Part I: hourly-mean wind speeds: ESDU-82026[S]. London: Engineering Sciences Data Unit, 1982.

[20] Engineering Sciences Data Unit. Characteristics of atmospheric turbulence near the ground. Part II: single point data for strong winds (neutral atmosphere): ESDU-74031[S]. London: Engineering Sciences Data Unit, 1974.

[21] HUNT A. Wind-tunnel measurements of surface pressures on cubic building models at several scales[J]. Journal of Wind Engineering and Industrial Aerodynamics, 1982, 10(2): 137-163.

[22] 袁家辉, 陈水福, 刘奕. 矩形高层建筑气动基底力矩系数研究[J]. 哈尔滨工业大学学报, 2023, 55(9): 54-62.
YUAN J H, CHEN S F, LIU Y. Aerodynamic base moment coefficients of rectangular high-rise buildings[J]. Journal of Harbin Institute of Technology, 2023, 55(9): 54-62(in Chinese).

[23] 刘奕. 板式高层建筑风荷载研究[D]. 杭州: 浙江大学, 2019: 28-45.
LIU Y. Study on wind load of slab high-rise building[D]. Hangzhou: Zhejiang University, 2019: 28-45(in Chinese).

[24] SIMIU E, SCANLAN R H, SACHS P, et al. Wind effects on structures: an introduction to wind engineering and wind forces in engineering[M]. New York: John Wiley & Sons, 1980.

Spatial correlation of along-wind fluctuating wind loads on rectangular high-rise buildings

YUAN Jiahui[1, 2], CHEN Shuifu[1, *], XIA Yuchao[3], LIU Yi[1]

(1. College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China;
2. WISDRI City Construction Engineering & Research Incorporation Ltd., Wuhan 430063, China;
3. School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To investigate the effects of side ratio and turbulent characteristics on the spatial correlation of along-wind fluctuating wind loads on rectangular tall buildings, synchronization pressure wind tunnel tests for rectangular tall buildings with side ratios ranging from 1/9～9 were carried out in four wind fields. Based on the experimental findings, the coherence function and vertical correlation coefficient of along-wind fluctuating wind loads were examined in relation to the effects of side ratio, turbulence intensity, and turbulence integral scale. The mathematical model of vertical correlation of rectangular high-rise buildings with a side ratio of 1/9～9 was obtained by the nonlinear least square method. The results show that the correlation coefficient of along-wind fluctuating wind load is affected by both the side ratio and separation distance, and decreases exponentially with the increase of separation distance. The along-wind fluctuating coherence function decays exponentially with the increase of frequency, and the decay rate of the coherence function is roughly positively related to the ratio of separation distance to average wind speed. The effects of turbulence integral scale and turbulence intensity on the correlation of along-wind fluctuating wind loads are different for buildings with different side ratios. It can serve as a guide for structural design and load code revision since the coherence functions and correlation coefficients of along-wind fluctuating wind loads on rectangular high-rise buildings suggested in this work correspond well with the experimental data.

Keywords: rectangular high-rise building; fluctuating wind load; spatial correlation; wind tunnel test; side ratio

Received: 2023-12-21; **Accepted**: 2024-01-28; **Published Online**: 2024-03-19 07:07
URL: link.cnki.net/urlid/11.2625.V.20240318.0957.001
Foundation item: National Natural Science Foundation of China (51878607)
* **Corresponding author**. E-mail: csf@zju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0840

基于相似性加权的复杂系统费用测算模型

常菊[1,2]，刘晓东[1,*]，何荧[2]

(1. 空军工程大学 装备管理与无人机工程学院，西安 710051；2. 江南机电设计研究所，贵阳 550009)

摘 要：为进一步提高航天器等大型复杂系统的费用测算精度，将大型复杂系统看作一定参数集合的系统分布，定义待测系统与样本的 Jensen-Shannon(JS) 散度、灰色关联度及综合相似度，根据综合相似度计算样本权重大小，构建复杂系统费用测算的加权回归模型。当样本量较少不满足最小二乘法建模条件时，选择综合相似度最大的样本作为基准样本，构建费用的驱动效应矩阵，根据矩阵中参数与费用的 JS 散度大小，按序选取散度较大的参数作为测算模型的自变量。针对样本量大于和不大于参数量 2 种情况实例计算结果。结果表明：基于 JS 散度和灰色关联度融合的相似性加权回归测算模型具有较高的预测精度和稳定性。

关 键 词：复杂系统；JS 散度；灰色关联度；综合相似度；驱动因子；精确度；费用测算
中图分类号：V267
文献标志码：A **文章编号**：1001-5965(2026)03-0784-07

大型复杂系统如航天器等，往往存在结构复杂、研制费用巨额等特点，近年来随着新材料、新工艺、新技术等在复杂系统研制中的不断引入，其研制费用呈现大幅增长的趋势。费用预测既是经费配置决策的理论依据，也是研制及生产单位优化设计方案的重要基础，其预测精度的高低不仅直接影响产品经费的使用效益，也对于解决技术、经济统一视角的研制方案优选问题具有突出的实际意义。

从费用测算的顶层逻辑出发，可将费用测算方法归类为从下到上和从上到下 2 种。从下到上的方法是通过测算系统的所有费用单元(如零部件单价或各工序成本)加和，得到系统总研制费用，主要适用于系统研制中后期；而从上到下的逻辑则相反，主要从系统整体角度(如功能及性能指标)来预测系统的总研制费用，主要适用于项目论证阶段或系统研制初期。传统的工程估算法属于从下到上预测整体费用的方法，参数法和类比法则属于从上到下整体测算费用的方法[1]。基于从下到上的逻辑发展起来的测算方法有基于作业成本测算方法、基于知识工程的测算方法及基于特征的测算方法等。而基于从上到下的逻辑发展起来的费用测算方法主要有：基于神经网络测算方法、基于偏最小二乘回归的测算方法、基于灰色系统理论的测算方法及基于支持向量机的预测方法等。其中，基于神经网络的方法属于“黑箱”建模[2]，其测算精度主要受训练样本数量及与待测系统相似度的影响；基于偏最小二乘回归的预测方法主要为多元线性建模[3]，其测算精度主要受费用驱动因子的选择及样本与待估系统相似度的影响；基于灰色系统理论的测算方法则认为费用与其驱动因子之间的关系存在大量的不确定性，是一种灰色关系[4]，其预测精度也主要受样本与待估系统相似度的影响；基于支持向量机的预测方法也主要适用于样本量少的场合[5]，其预测精度也主要受样本与待估系统相似度的影响。复杂系统费用预测的高精确度一直是该领域研究人员探索的目标，针对复杂系统研制周期长、费用影响因素多、保密因素导致公开信息少等特点，

收稿日期：2023-12-29；录用日期：2024-03-06；网络出版时间：2024-04-10 14:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20240409.0907.001
*通信作者. E-mail：xdliu6609@qq.com

引用格式：常菊，刘晓东，何荧. 基于相似性加权的复杂系统费用测算模型 [J]. 北京航空航天大学学报，2026，52（3）：784-790.
CHANG J，LIU X D，HE Y. Complex equipment cost estimation model based on similarity weight[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：784-790（in Chinese）.

从上到下的系统费用测算方法一直是受关注的重点, 本文方法也属于该逻辑范畴。在小样本条件下, 参考样本与待测系统相似性研究对于费用预测有较为迫切的现实需求, 即与待估系统指标参数越相似的样本系统费用, 理应与待估系统费用越相似[6]。KL(KULLback-Leibler)散度表征 2 种分布之间的相似程度, JS(Jensen-Shannon)散度是在 KL 散度基础上进一步提升, 用于解决 KL 散度不对称问题, 已被用在量化评估[7]、能量变换[8]、故障检测[9]、互联网[10]等领域。系统的参数指标分布是整个系统的定量表现, 某种程度上可将系统看作一定参数集合的系统分布[11], 据此可引入 JS 散度衡量 2 种系统之间的相似程度[12-14]。本文基于 JS 散度和灰色关联度融合, 针对系统研制费用预测, 提出一种复杂系统相似信息优先的费用测算建模方法。

1　复杂系统费用测算模型

1.1　复杂系统综合相似度

设复杂系统样本 $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\cdots$, $\boldsymbol{A}_n$, 每个样本 $\boldsymbol{A}_i(i=1,2,\cdots,n)$有 m 个参数 $y_j(j=1,2,\cdots,m)$, 则 x_{ij} 表示样本 $\boldsymbol{A}_i$关于参数 y_j的指标数据, c_i表示样本 $\boldsymbol{A}_i$的费用, $\boldsymbol{A}_0$表示待测系统, x_{0j}表示 $\boldsymbol{A}_0$关于参数 y_j的指标数据, c_0表示 $\boldsymbol{A}_0$的费用。样本的指标矩阵如下:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} & c_1 \\ x_{21} & x_{22} & \cdots & x_{2m} & c_2 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} & c_n \\ x_{01} & x_{02} & \cdots & x_{0m} & c_0 \end{bmatrix}$$

为解决参数 y_j量纲不一致的问题, 本文采用 min-max 向量化规范化方法, 即

$$x'_{ij}=\frac{x_{ij}-x_{\min,j}}{x_{\max,j}-x_{\min,j}} \tag{1}$$

式中: $x_{\max,j}$、$x_{\min,j}$分别为样本和待估系统关于参数 y_j的最大值、最小值。

计算样本 $\boldsymbol{A}_i$和待测系统 $\boldsymbol{A}_0$关于参数 y_j的比重为

$$p_{ij}=x'_{ij}\Big/\sum_{j=1}^{m}x'_{ij} \tag{2}$$

定义 1　样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的 KL 散度为

$$D_{\mathrm{KL}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)=\sum_{j=1}^{m}p_{0j}\log_2(p_{0j}/p_{ij}) \tag{3}$$

散度 $D_{\mathrm{KL}}(\boldsymbol{A}_i\|\boldsymbol{A}_0)$利用系统特性参数的归一化值来表征系统 $\boldsymbol{A}_i$与待估系统 $\boldsymbol{A}_0$之间的差异程度, KL 散度非负, 且 KL 散度值越接近 0, 表征两者差异性越小, 反之越大。但 KL 散度不具有对称性[15], 即 $D_{\mathrm{KL}}(\boldsymbol{A}_i\|\boldsymbol{A}_0)\neq D_{\mathrm{KL}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)$, 这与事实不太相符。

定义 2　样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的 JS 散度为

$$D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)=\frac{1}{2}\sum_{j=1}^{m}p_{0j}\log_2\left(p_{0j}/2(p_{0j}+p_{ij})\right)+\frac{1}{2}\sum_{j=1}^{m}p_{ij}\log_2\left(p_{ij}/2(p_{0j}+p_{ij})\right) \tag{4}$$

$D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)$是利用 KL 散度演变而得, 显然 $D_{\mathrm{JS}}(\boldsymbol{A}_i\|\boldsymbol{A}_0)=D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)$, 即 JS 散度克服 KL 散度的非对称性问题, 其计算结果为 0～1 之间[15]。同样地, JS 散度值越小, 表明系统 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j越接近, 二者之间的相似程度越高。

定义 3　样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的灰色关联系数为[16]

$$\zeta_{ij}=\frac{\min\limits_{i}\min\limits_{j}\left|x'_{ij}-x'_{0j}\right|+\rho\max\limits_{i}\max\limits_{j}\left|x'_{ij}-\mathrm{x}'_{0j}\right|}{\left|x'_{ij}-x'_{0j}\right|+\rho\max\limits_{i}\max\limits_{j}\left|x'_{ij}-x'_{0j}\right|} \tag{5}$$

式中: $\rho\in[0,1]$为分辨系数。得到样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的关联度为

$$\gamma_{i0}=\frac{1}{m}\sum_{j=1}^{m}\zeta_{ij} \tag{6}$$

灰色关联度 γ_{i0}可以部分表征样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$的相似程度, γ_{i0}值越大, 表明两者关于 y_j相似程度越高, 其相似程度表征情况与 JS 散度相反。

定义 4　样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$的综合相似度为

$$R_{i0}=\lambda[1-D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)]+(1-\lambda)\gamma_{i0} \tag{7}$$

式中: λ为权重系数。

综合相似度表征待测系统 $\boldsymbol{A}_0$与样本 $\boldsymbol{A}_i$的整体相似程度, R_{i0}越大表征二者越相似, 就有理由认为样本 $\boldsymbol{A}_i$的费用能一定程度上较好的反应待测系统 $\boldsymbol{A}_0$的真实费用, 故而可将样本 $\boldsymbol{A}_i$赋予较大权重。

定理　当灰色关联分辨系数 ρ和综合相似度权重系数 λ一定, 样本参数发生仿射变换时, 即 $(\widehat{x}_{1j},\widehat{x}_{2j},\cdots,\widehat{x}_{nj},\widehat{x}_{0j})=k(x_{1j},x_{2j},\cdots,x_{nj},x_{0j})+b$时, 样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$的综合相似度不变。

证明　在计算 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$散度及灰色关联度之前, 需将样本数据采用 min-max 向量化规范化方法进行处理, 当样本参数发生仿射变换时, 规范化处理为

$$\widetilde{x}'_{ij}=\frac{kx_{ij}+b-(kx_{\min,j}+b)}{kx_{\max,j}+b-(kx_{\min,j}+b)}=\frac{x_{ij}-x_{\min,j}}{x_{\max,j}-x_{\min,j}} \tag{8}$$

步骤 1　由式(8)显然得到指标仿射变换后, 规范化处理得到的 x'_{ij}保持不变, 由式(2)可得关于参数 y_j的比重 p_{ij}不变, 由式(3)和式(4)得样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的 JS 散度保持不变, 即 $\widehat{D}_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)=D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)$。

步骤 2　由于规范化处理得到的 x'_{ij}保持不变, 由式(5)得到样本 $\boldsymbol{A}_i$与 $\boldsymbol{A}_0$关于 y_j的灰色关联系数

ζ_{ij}不变，进而由式(6)得到关联度γ_{i0}保持不变，即$\widehat{\gamma}_{i0}=\gamma_{i0}$。

步骤 3 由式(7)可知，当样本指标仿射变换后，$\boldsymbol{A}_i$与$\boldsymbol{A}_0$的综合相似度R_{i0}保持不变，即$\widehat{R}_{i0}=\lambda\left[1-\widehat{D}_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)\right]+(1-\lambda)\widehat{\gamma}_{i0}=\lambda[1-D_{\mathrm{JS}}(\boldsymbol{A}_0\|\boldsymbol{A}_i)]+(1-\lambda)\gamma_{i0}=R_{i0}$。

证毕

定理表明，当灰色关联分辨系数ρ和综合相似度权重系数λ取值一定，样本参数发生仿射变换时，样本之间的综合相似度不会变化。

1.2 测算模型

在1.1节样本指标矩阵中，设复杂系统的费用c_i与y_1，y_2，…，y_m存在线性关系，即$c_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_mx_{im}+\varepsilon_i$，其中，$\varepsilon_i$为误差，则$n$个样本费用$c$与参数$y$的线性关系矩阵表示为

$$\boldsymbol{C}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{9}$$

式中：

$$\boldsymbol{C}=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix};\quad \boldsymbol{X}=\begin{bmatrix}1&x_{11}&x_{12}&\cdots&x_{1m}\\1&x_{21}&x_{22}&\cdots&x_{2m}\\\vdots&\vdots&\vdots&&\vdots\\1&x_{n1}&x_{n2}&\cdots&x_{nm}\end{bmatrix};$$

$$\boldsymbol{\beta}=\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\\\vdots\\\beta_m\end{bmatrix};\quad \boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}。$$

通过测算出未知参数$\hat{\boldsymbol{\beta}}$，则待估系统$\boldsymbol{A}_0$的费用为

$$c_0=\begin{bmatrix}1&x_{01}&x_{02}&\cdots&x_{0m}\end{bmatrix}\hat{\boldsymbol{\beta}} \tag{10}$$

式中：

$$\hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\beta}_0&\hat{\beta}_1&\cdots&\hat{\beta}_m\end{bmatrix}^{\mathrm{T}}$$

1.2.1 样本量n大于系统参数m的情况

当$n>m+1$时，式(9)中未知参数$\boldsymbol{\beta}$可以采用最小二乘法进行估计，根据相似信息优先原理，将参数与待测系统参数综合相似度大的$\boldsymbol{A}_i$赋予较大权重，则

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{C} \tag{11}$$

式中：$\boldsymbol{W}=\begin{bmatrix}\omega_1&0&\cdots&0\\0&\omega_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&\omega_n\end{bmatrix}$。

ω_i为样本$\boldsymbol{A}_i$与$\boldsymbol{A}_0$的综合相似度权重[17]，且

$$\omega_i=R_{i0}\Big/\sum_{i=1}^{n}R_{i0} \tag{12}$$

证明 采用最大似然估计方法来估算未知参数$\boldsymbol{\beta}$，当$\beta_0=\hat{\beta}_0$，$\beta_1=\hat{\beta}_1$，…，$\beta_m=\hat{\beta}_m$时达到最小。

$$\boldsymbol{Q}=\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}=\sum_{i=1}^{n}\omega_i(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_mx_{im}-c_i)^2 \tag{13}$$

求式(13)$\boldsymbol{Q}$分别关于β_0，β_1，…，β_m的偏导数，并令偏导数等于0，得

$$\begin{cases}\dfrac{\partial\boldsymbol{Q}}{\partial\beta_0}=2\displaystyle\sum_{i=1}^{n}\omega_i(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_mx_{im}-c_i)=0\\[2ex]\dfrac{\partial\boldsymbol{Q}}{\partial\beta_j}=2\displaystyle\sum_{i=1}^{n}\omega_i(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_mx_{im}-c_i)x_{ij}=0\end{cases} \tag{14}$$

式中：$j=1,2,\cdots,m$。

简化式(14)得到

$$\begin{cases}\beta_0\displaystyle\sum_{i=1}^{n}\omega_i+\beta_1\sum_{i=1}^{n}\omega_ix_{i1}+\beta_2\sum_{i=1}^{n}\omega_ix_{i2}+\cdots+\beta_m\sum_{i=1}^{n}\omega_ix_{im}=\sum_{i=1}^{n}\omega_ic_i\\\beta_0\displaystyle\sum_{i=1}^{n}\omega_ix_{i1}+\beta_1\sum_{i=1}^{n}\omega_ix_{i1}^2+\beta_2\sum_{i=1}^{n}\omega_ix_{i1}x_{i2}+\cdots+\beta_m\sum_{i=1}^{n}\omega_ix_{i1}x_{im}=\sum_{i=1}^{n}\omega_ix_{i1}c_i\\\vdots\\\beta_0\displaystyle\sum_{i=1}^{n}\omega_ix_{im}+\beta_1\sum_{i=1}^{n}\omega_ix_{i1}x_{im}+\beta_2\sum_{i=1}^{n}\omega_ix_{i2}x_{im}+\cdots+\beta_m\sum_{i=1}^{n}\omega_ix_{im}{}^2=\sum_{i=1}^{n}\omega_ix_{im}c_i\end{cases} \tag{15}$$

式(15)可写成矩阵形式

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{C} \tag{16}$$

由于$\boldsymbol{W}$为n阶对角方阵(ω_i均大于0)，得到$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}$为$(m+1)\times(m+1)$矩阵，根据文献[17]证明得到$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$满秩，不难得出$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}$也满秩，故式(16)两边同时乘以逆矩阵$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}$得

$$\hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\beta}_0&\hat{\beta}_1&\cdots&\hat{\beta}_m\end{bmatrix}^{\mathrm{T}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{C}$$

证毕

1.2.2 样本量n不大于系统参数m的情况

当样本量n不大于系统参数m时，不满足最小二乘法的条件，故不能直接采用1.2.1节的方式计算$\hat{\boldsymbol{\beta}}$。此种小样本条件下，首先，建立费用的驱动效应矩阵，然后，采取一定方法选择对费用驱动较大的参数作为自变量，再利用1.2.1节方法进行多元线性回归。一般地，选择参数个数比样本数量少2为最佳[18]。本文采用JS散度方法选择与费用相似度较高的参数作为自变量，主要步骤如下：

步骤 1　计算样本$\boldsymbol{A}_i$与待测系统$\boldsymbol{A}_0$的综合相似度, 选择相似度最大的$\boldsymbol{A}_k$作为基准, $\boldsymbol{A}_k$的数据记作$x_{k1},x_{k2},\cdots,x_{km},c_k$。

步骤 2　建立费用的驱动效应矩阵:

$$\begin{cases} v_{ij}=\dfrac{|x_{ij}-x_{kj}|}{\sum\limits_{i=1}^{n}|x_{ij}-x_{kj}|} \qquad j=1,2,\cdots,m \\ u_i=\dfrac{|c_i-c_k|}{\sum\limits_{i=1}^{n}|c_i-c_k|} \end{cases} \tag{17}$$

步骤 3　计算$\boldsymbol{v}_j$与$\boldsymbol{u}$的 KL 散度:

$$D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{v}_j)=\sum_{i=1}^{n}u_i\log_2(u_i/v_{ij}) \tag{18}$$

步骤 4　计算$\boldsymbol{v}_j$与$\boldsymbol{u}$的 JS 散度:

$$D_{\mathrm{JS}}(\boldsymbol{u}\|\boldsymbol{v}_j)=\frac{1}{2}\sum_{j=1}^{m}\boldsymbol{v}_j\log_2\left[\boldsymbol{v}_j/2(\boldsymbol{v}_j+\boldsymbol{u})\right]+\frac{1}{2}\sum_{j=1}^{m}\boldsymbol{u}\log_2\left[\boldsymbol{u}/2(\boldsymbol{v}_j+\boldsymbol{u})\right] \tag{19}$$

步骤 5　根据$D_{\mathrm{JS}}(\boldsymbol{u}\|\boldsymbol{v}_j)$大小排序, 依次选取较大的参数作为自变量, 选取参数个数比样本数量少 2 即可(若多个参数 JS 散度相同, 则根据最小二乘条件同时舍弃或保留该多个参数)。

步骤 6　利用 1.2.1 节方法进行多元线性回归得到$\hat{\boldsymbol{\beta}}$。

2　实例分析

1) 例 1, 分析样本量 n 大于系统参数 m 时本文预测方法的精度。选择与文献 [17] 相同的实例, 数据如表 1 所示。为与文献 [17] 形成对比, 样本数据也选取 A-6E～FB-11A,而 F-14A、A-7E 及 F-111E 这 3 种型号作为费用预测对象。表中, x_1,x_2,x_3,x_4 为样本参数, 采用 1.1 节方法分别计算样本与待测系统的综合相似度及其相应权重, 计算结果如表 2 所示, 再利用 1.2 节建模方法分别建立 F-14A、A-7E 及 F-111E 这 3 种型号飞机机载电子设备费用模型为

$$\begin{cases} c_{\mathrm{F-14A}}=-8\,256.87+114.03x_1+2.3x_2-500.73x_3+39.69x_4 \\ c_{\mathrm{A-7E}}=-7\,973.45+109.56x_1+2.36x_2-529.75x_3+40.67x_4 \\ c_{\mathrm{F-111E}}=-7\,934.99+109.32x_1+2.38x_2-557.15x_3+38.45x_4 \end{cases}$$

费用预测统计结果如表 3 所示。

由表 3 和表 4 可知, 基于相似性理论的方法对于 3 种电子设备费用平均预测精度为 8.35%(其他 3 种方法依次为 24.93%、19.12%、11.11%), 预测误差方差为 2.26(其他 3 种方法依次为 552.96、241.85、4.73)。相比其他 3 种预测方法, 本文方法平均精度最高, 且预测精度变化浮动(预测误差方差)最小, 表明基于相似性理论的费用预测方法不仅预测精度高, 且预测结果较稳定, 模型准确度较高。

2) 例 2, 分析样本量 n 不大于系统参数 m 时本文预测方法的精度。选择与文献 [6,17-19] 相同的实例, 数据如表 5 所示。为形成对比, 样本数据也选取 A～H, 预测飞机 I 的费用, 表中 $x_1\sim x_8$ 为样本参数, 分析过程如下:

步骤 1　采用 1.1 节方法分别计算样本 A～H 与待估系统 I 的综合相似度, 依次为: 0.798 9、0.769 5、

表 1　机载电子设备费用样本

Table 1　Sample of airborne electronic equipment costs

飞机型号	首飞时间x_1/a	质量x_2/kg	体积x_3/m³	功率x_4/kW	实际平均费用c/10³美元
A-6E	70	624.25	0.569 1	6.4	1 069
A-7D	68	508.25	0.710 1	10.5	669
F-4D	65	790.41	0.8430	8.2	582
A-10A	72	265.14	0.239 4	3.1	315
E-4E	67	566.14	0.677 3	5.3	662
F-4J	66	1 021.05	0.982 4	19.4	1 329
F-15A	72	717.32	0.833 1	22.5	2 488
F-111A	64	805.40	0.877 4	5.6	1 267
F-111D	68	1 068.72	0.910 2	12.5	2 392
F-111F	71	933.88	1.061 1	8.9	1 577
FB-111A	70	1 136.36	1.343 2	7.9	1965
F-14A	70	998.35	1.062 8	29.4	2 383
A-7E	68	653.76	0.841 3	8.3	828
F-111E	69	987.00	1.105 4	8.9	1 564

表2 3个待预测机载电子设备综合相似度和权重

Table 2 Comprehensive similarity and weight of three predicted airborne electronic devices

飞机型号	相似度			权重		
	与F-14A对比	与A-7E对比	与F-111E对比	与F-14A对比	与A-7E对比	与F-111E对比
A-6E	0.77	0.85	0.80	0.09	0.10	0.09
A-7D	0.76	0.90	0.81	0.09	0.10	0.09
F-4D	0.75	0.88	0.82	0.09	0.10	0.09
A-10A	0.52	0.55	0.53	0.06	0.06	0.06
E-4E	0.73	0.87	0.78	0.09	0.10	0.09
F-4J	0.85	0.75	0.81	0.10	0.08	0.09
F-15A	0.82	0.81	0.74	0.10	0.09	0.08
F-111A	0.70	0.78	0.76	0.08	0.09	0.09
F-111D	0.83	0.86	0.87	0.10	0.10	0.10
F-111F	0.86	0.82	0.92	0.10	0.09	0.11
FB-111A	0.82	0.79	0.88	0.10	0.09	0.10

表3 4种预测方法的结果比较

Table 3 Result comparison of four prediction methods

万元

飞机型号	实际费用	回归预测费用[17]	支持向量机预测[17]	基于熵理论预测[17]	本文方法预测
F-14A	2 383	2 182	2 465	2 695.8	2 558.80
A-7E	828	708	1 114	922.4	911.49
F-111E	1 564	753	1 867	1 701.8	1 682.80

表4 4种预测方法的预测误差比较

Table 4 Comparison of prediction errors among four prediction methods

飞机型号	回归预测误差/%[17]	支持向量机预测误差/%[17]	基于熵理论预测误差/%[17]	本文方法预测误差/%
F-14A	8.43	3.44	13.13	7.37
A-7E	14.49	34.54	11.40	10.08
F-111E	51.85	19.37	8.81	7.60

0.816 8、0.626 5、0.737 5、0.888 4、0.890 5、0.690 7，可见，样本G与待测系统I的相似度最高(0.890 5)，故取样本G为基准样本；

步骤2 采用1.2.2节方法，对参数$x_1 \sim x_8$和费用c预处理后，依次计算各参数与费用的JS散度为：0.036 4、0.102 1、0.028 5、0.182 8、0.031 7、0.205 8、0.099 0、0.102 8，由于散度越小表明二者越相似，且样本数量为8，参数数量为8，故舍弃散度较大的2个参数x_4和x_6，其他参数作为费用驱动因子；

步骤3 采用1.1节方法分别计算新样本与待估系统的综合相似度及其权重，再利用1.2节建模方法建立飞机I费用模型为

$$c_{\mathrm{I}} = 23510.10 - 0.83x_1 + 1467.82x_2 - 4494.25x_3 - 5.35x_5 - 0.22x_7 + 3.96x_8$$

费用预测统计结果如表6所示。

由表6可以得到，基于相似性理论方法预测的飞机I的费用为14 028.3万元，相对误差为1.56%，相较于其他5种方法(26.2%[18]、7.1%[6]、4.0%[19]、1.97%[19]、1.64%[17])，本文方法精度更高。

综合本文中对于样本量n大于和不大于系统参数m这2种情况下实例数据的分析，与其他算法进行比较，本文模型预测精度和稳定性更高，表明，本文基于相似性加权的方法模型较理想。

表5 运输机性能数据与费用

Table 5 Performance data and cost of transport aircraft

飞机型号	最大起飞质量 x_1/kg	机身长 x_2/m	机高 x_3/m	起飞距离 x_4/m	满油航程 x_5/km	最大平飞速度 x_6/(m·s^{-1})	空重 x_7/kg	载油量 x_8/kg	费用 c/万元
A	13 494	23.500	8.43	867	4 262	425.0	6 597	5 683	6 666.70
B	6 849	14.390	4.57	987	3 701	746.0	3 655	2 640	3 524.30
C	9 979	16.900	5.12	1 581	4 679	874.0	5 357	3 350	6 569.90
D	5 670	13.340	4.57	536	3 641	536.0	3 656	1 653	5 586.23
E	63 503	39.750	9.30	1 859	6 764	925.0	33 183	21 273	27 768.80
F	22 000	29.870	6.75	1 200	2 870	907.0	34 360	5 500	17 575.20
G	21 500	27.170	7.65	1 050	2 000	580.0	12 200	5 000	18 137.60
H	70 310	29.790	11.66	1 091	7 876	602.0	36 300	36 300	50 476.00
I	21 000	24.615	7.30	1 300	3 100	819.2	11 700	6 000	14 250.00

表 6　6 种预测方法的比较

Table 6　Comparison of six prediction methods

飞机型号	实际费用/万元	不同预测方法预测费用/万元					
		偏最小二乘回归[18]	相似信息优先原理[6]	逐步多元回归预测[19]	极大熵[19]	基于熵理论[17]	本文预测方法
I	14 250	10 510	15 262	13 680	14 530.27	14 016	14 028.3

3　结　论

针对小样本的复杂系统费用预测, 基于待测系统与样本的 JS 散度和灰色关联度融合, 定义综合相似度度量样本与待测系统的整体相似程度, 据此建立加权回归模型预测复杂系统费用。

经实例分析与对比, 可得到如下结论:

1)本文预测模型相较回归预测、支持向量机及熵理论等方法预测精度较高且稳定;

2)当样本量不大于参数量时, 构造参数与费用 JS 散度选取费用驱动因子是一种有效途径;

3)本文方法作为现有复杂系统费用测算方法的拓展和补充, 具有一定的理论和应用价值。

参考文献（References）

[1] 王骏, 罗鹏程, 周经伦, 等. 军用飞机出厂费用估算方法研究进展综述[J]. 系统工程与电子技术, 2017, 39(9): 2012-2021.
WANG J, LUO P C, ZHOU J L, et al. Survey on military aircraft fly-away cost estimation methods[J]. Systems Engineering and Electronics, 2017, 39(9): 2012-2021(in Chinese).

[2] 金锋, 崔卉, 张玉平, 等. 武器装备研制阶段费用预测方法及仿真[J]. 北京理工大学学报, 2006, 26(11): 991-994.
JIN F, CUI H, ZHANG Y P, et al. Forecast and emulation research on the cost of development for military equipment[J]. Transactions of Beijing Institute of Technology, 2006, 26(11): 991-994(in Chinese).

[3] 王礼沅, 郭基联, 张恒喜. 递阶偏最小二乘回归在飞机研制费用预测中的应用[J]. 航空学报, 2009, 30(8): 1380-1384.
WANG L Y, GUO J L, ZHANG H X. Application of hierarchical partial least squares regression to development cost prediction of aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(8): 1380-1384(in Chinese).

[4] 何鹏翔, 孙胜祥. 基于改进 GM(0, N) 的舰船装备维修费预测模型[J]. 舰船电子工程, 2022, 42(6): 151-154.
HE P X, SUN S X. Ship equipment maintenance cost prediction model based on improved GM(0, N)[J]. Ship Electronic Engineering, 2022, 42(6): 151-154(in Chinese).

[5] 钟诗胜, 付旭云, 胡淑荣. 小样本条件下航空装备费用预测[J]. 哈尔滨工业大学学报, 2011, 43(5): 52-55.
ZHONG S S, FU X Y, HU S R. Aviation equipment cost prediction under small sample size[J]. Journal of Harbin Institute of Technology, 2011, 43(5): 52-55(in Chinese).

[6] 吴利丰, 刘思峰, 方世力, 等. 相似信息优先的复杂装备费用测算模型[J]. 系统工程与电子技术, 2014, 36(10): 2024-2028.
WU L F, LIU S F, FANG S L, et al. Complicated equipment cost estimation model with similarity information priority[J]. Systems Engineering and Electronics, 2014, 36(10): 2024-2028(in Chinese).

[7] 王晨光, 白金鹏, 李婷斑, 等. 基于空战机理的雷达仿真模型可信度评估方法[J]. 系统仿真学报, 2023, 35(10): 2113-2121.
WANG C G, BAI J P, LI T T, et al. Reliability evaluation method of radar simulation model based on air combat mechanism[J]. Journal of System Simulation, 2023, 35(10): 2113-2121(in Chinese).

[8] 毛春翔, 韦鹏, 刘若鹏, 等. 基于行波功率 S 变换能量 JS 散度的直流换流器站间保护方案研究[J]. 电网与清洁能源, 2023, 39(8): 17-26.
MAO C X, WEI P, LIU R P, et al. Research on the inter station protection scheme of the DC converter based on traveling wave power S-transform energy JS divergence[J]. Power System and Clean Energy, 2023, 39(8): 17-26(in Chinese).

[9] 熊亚军, 谢林柏, 彭力. 基于 JS 散度和潜在特征提取的多块 PCA 故障监测[J]. 仪表技术与传感器, 2022(5): 105-110.
XIONG Y J, XIE L B, PENG L. Multiblock PCA fault monitoring based on JS divergence and latent feature extraction[J]. Instrument Technique and Sensor, 2022(5): 105-110(in Chinese).

[10] 程光, 钱德鑫, 郭建伟, 等. 基于散度的网络流概念漂移分类方法[J]. 计算机研究与发展, 2020, 57(12): 2673-2682.
CHENG G, QIAN D X, GUO J W, et al. A classification approach based on divergence for network traffic in presence of concept drift[J]. Journal of Computer Research and Development, 2020, 57(12): 2673-2682(in Chinese).

[11] 明春强. 基于复杂网络的装备体系重要节点识别与链路重构方法研究[D]. 重庆: 重庆邮电大学, 2020.
MING C Q. Research on key node identification and link reconstruction of equipment system based on complex network[D]. Chongqing : Chongqing University of Posts and Telecommunications, 2020(in Chinese).

[12] MENÉNDEZ M L, PARDO J A, PARDO L, et al. Asymptotic approximations for the distributions of the (h., φ.)-divergence goodness-of-fit statistics: application to Renyi's statistic[J]. Kybernetes, 1997, 26(4): 442-452.

[13] SCHIEBER T A, CARPI L, DÍAZ-GUILERA A, et al. Quantification of network structural dissimilarities[J]. Nature Communications, 2017, 8: 13928.

[14] 陈浩宇, 徐涛, 刘闯, 等. 基于高阶信息的网络相似性比较方法[J]. 物理学报, 2024, 73(3): 038901.
CHEN H Y, XU T, LIU C, et al. Network similarity comparison method based on higher-order information[J]. Acta Physica Sinica, 2024, 73(3): 038901(in Chinese).

[15] 胡文, 景玉海. 基于 KL 散度与 JS 散度相似度融合推荐算法[J]. 哈尔滨商业大学学报 (自然科学版), 2020, 36(1): 48-53.
HU W, JING Y H. Recommendation algorithm based on fusion of KL divergence and JS divergence similarity[J]. Journal of Harbin University of Commerce (Natural Sciences Edition), 2020, 36(1):

48-53(in Chinese).

[16] 冯卉, 宋宝军, 张春梅. 基于改进灰色关联法的空袭目标威胁评估方法[J]. 运筹与管理, 2023, 32(9): 1-6.
FENG H, SONG B J, ZHANG C M. Aerial targets threat evaluation based on grey incidence method[J]. Operations Research and Management Science, 2023, 32(9): 1-6(in Chinese).

[17] 魏东涛, 刘晓东, 丁刚, 等. 基于熵理论的复杂装备费用测算模型[J]. 北京航空航天大学学报, 2022, 48(5): 816-823.
WEI D T, LIU X D, DING G, et al. Complex equipment cost estimation model based on entropy theory[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 816-823(in Chinese).

[18] 李寿安, 张恒喜, 童中翔, 等. 偏最小二乘回归在军用飞机价格预测中的应用[J]. 航空学报, 2006, 27(4): 600-604.
LI S A, ZHANG H X, TONG Z X , et al. Application of partial least-squares regression to the acquisition cost estimating of military aircrafts[J]. Acta Aeronautica et Astronautica Sinica, 2006, 27(4): 600-604(in Chinese).

[19] 方世力, 吴利丰, 刘思峰, 等. 复杂装备费用驱动因子识别的效应比重极大熵模型[J]. 系统工程, 2014, 32(10): 149-153.
FANG S L, WU L F, LIU S F, et al. Complex equipment cost drive factors identification model with maximization of the driving effect proportion[J]. Systems Engineering, 2014, 32(10): 149-153(in Chinese).

Complex equipment cost estimation model based on similarity weight

CHANG Ju[1, 2], LIU Xiaodong[1, *], HE Ying[2]

(1. College of Equipment Management and UAV Engineering, Air Force Engineering University, Xi'an 710051, China;
2. Jiangnan Mechanical and Electrical Design Institute, Guiyang 550009, China)

Abstract: In order to further improve the accuracy of cost estimation for large and complex equipment, such as spacecraft and weapon systems, the large and complex equipment is regarded as a system distribution of a certain parameter set. Define the Jensen-Shannon (JS) divergence, grey correlation, and comprehensive similarity between the tested equipment and the equipment samples, and calculate the sample weight based on the similarity to construct a weighted regression model for cost estimation of complex equipment. To create the cost driving impact matrix, the sample with the highest complete similarity is chosen as the benchmark sample when the sample size does not satisfy the requirements of the least squares modeling. Based on the JS divergence between the parameters and the cost in the matrix, the parameters with larger divergence are selected as the independent variables for the prediction model. By comparing two scenarios in which the sample size is larger and smaller than the parameter size, the comparative analysis demonstrates the excellent prediction accuracy and stability of the similarity weighted regression calculation model based on the combination of JS divergence and grey correlation degree.

Keywords: complex equipment; JS divergence; grey correlation; comprehensive similarity; driving factor; accuracy; cost estimation

Received: 2023-12-29; **Accepted**: 2024-03-06; **Published Online**: 2024-04-10 14:24
URL: link.cnki.net/urlid/11.2625.V.20240409.0907.001

* **Corresponding author**. E-mail: xdliu6609@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0514

半倾转构型旋翼飞行器纵向操纵策略研究

田洪源*，鲁森魁，武上景
(中国直升机设计研究所 直升机动力学重点实验室，景德镇 333001)

摘　　要：半倾转构型旋翼飞行器作为一种新构型的垂直起降飞行器，由于其多舵面带来的操作冗余与倾转过渡轨迹相耦合问题，仍没有较好的解决方案。为此，基于最优控制理论，针对半倾转构型旋翼飞行器纵向操纵冗余问题，同步开展舵面分配与动态倾转过渡轨迹优化研究。针对半倾转构型特点建立纵向刚体飞行力学模型，采用操纵速率作为控制量，避免优化过程中出现跳跃不连续的操纵策略，分析纵向舵面分配问题，建立混合操纵方程。将飞行器过渡倾转过程的轨迹优化问题转化为非线性动态最优控制问题，选取合理的优化目标与约束条件，建立最优控制模型。选取不同的舵面分配方程与目标函数，多次开展正向和逆向倾转过渡过程的优化计算，并进行参数影响性分析，同步确定飞行器舵面分配及过渡飞行最优操纵策略与飞行轨迹。采用仿真分析验证了方法的有效性。相比于传统的先基于配平计算确定舵面分配，再通过优化计算确定飞行轨迹的策略制定方式，姿态的稳定性与操纵负荷得到显著的优化。

关　键　词：半倾转构型；最优控制；倾转过渡；操纵策略；轨迹优化
中图分类号：V212.4
文献标志码：A　　**文章编号**：1001-5965(2026)03-0791-10

近年来，电动垂直起降 (electric vertical take-off and landing, eVTOL) 飞行器随着相关技术的进步开始快速发展。通过采用电机分布式驱动，电动垂直起降飞行器的构型相比传统飞行器更加多样，主要包括多旋翼、复合翼、倾转旋翼和半倾转旋翼构型[1-3]。其中，半倾转旋翼构型结合了倾转旋翼和复合翼构型的特点，通常在机翼前方布置倾转旋翼，机翼后方布置非倾转旋翼，其在研发效率、成本控制和噪声优化方面具有一定的优势。在纵向操纵平面，位于机翼前方的前旋翼通过操纵倾转角和旋翼总距来控制旋翼拉力，位于机翼后方的后旋翼通过操纵旋翼转速来控制旋翼拉力。英国 Vertical Aerospace 公司的 VX4 飞行器与美国 Archer Aviation 公司的 Midnight 飞行器均采用该构型设计。

该构型飞行器在有以上优点的同时，也存在特有的问题。一方面，操纵量冗余，面临舵面控制分配问题，且与倾转旋翼带来的过渡飞行策略问题相耦合，需要有针对性地制定操纵策略，以保障飞行安全[4-5]；另一方面，前后旋翼距离重心位置较远，会产生较大的俯仰力矩，同时尾翼上的舵面随着速度的增大也会有较大的俯仰操纵功效，需要合适的操纵策略以保证在全机俯仰力矩平衡的状态下完成倾转过渡飞行[6-7]。机翼上除了控制滚转操纵的副翼，通常还会设计襟翼以增加低速段的机翼升力，也会对纵向操纵策略产生影响。

为了让倾转旋翼飞行器完成倾转过渡转换，许多学者研究了不同的操纵策略设计方案[8-10]。舵面分配负责将飞控律产生的广义力矩指令高效、可靠地分配给多个冗余的执行机构。Wang 等系统总结了航天器和飞行器控制分配方法的最新进展，涵盖优化、智能及容错方法[11]。Liu 等研究了基于强化学习的自适应容错控制分配[12]。Johansen 等研究了

收稿日期：2025-07-22；录用日期：2025-10-17；网络出版时间：2025-10-21 15:23
网络出版地址：link.cnki.net/urlid/11.2625.V.20251021.1309.002
基金项目：直升机动力学全国重点实验室 2022 年基金（61422202205）
*通信作者. E-mail：tianhy@avic.com

引用格式：田洪源，鲁森魁，武上景. 半倾转构型旋翼飞行器纵向操纵策略研究 [J]. 北京航空航天大学学报，2026，52（3）：791-800.
TIAN H Y，LU S K，WU S J. Research on longitudinal manipulation strategy of partial tilting aircraft [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：791-800 （in Chinese）.

使用序列二次规划(SQP)及其改进算法解决过驱动系统控制分配问题[13]。Di Milano等建立了一种线性变参数小扰动模型,在此基础上设计数值优化方法来跟踪预先设定的飞行轨迹,并研究了倾转旋翼机从直升机模式倾转至固定翼模式过程中的倾转走廊曲线和操纵策略[14]。蒋海辉针对倾转四旋翼飞行器,采用混合操纵来解决操纵冗余问题[15]。Kleinhesselink研究了XV-15的飞行特性,并制定了相关的操纵策略[16]。针对倾转旋翼构型飞行器的操纵策略问题,大部分研究均以设计较为成熟的倾转旋翼机为研究对象,而对新构型的半倾转多旋翼飞行器研究较少[17-18]。以上研究对于多操纵舵面的操纵冗余问题,多采用稳态配平的计算方式,通过限制操纵或姿态减少配平量或增加约束方程的方式来解决,或通过控制系统设计跟踪预定的飞行轨迹,无法得到不同飞行任务下的最优操纵策略与飞行轨迹。

目前,许多学者基于最优控制理论,针对飞行轨迹优化开展了大量工作[19-21]。在飞行器飞行过程中,其运动、操纵策略与性能指标均为时间和空间的函数,因此,飞行轨迹优化问题可以归结为一种含有状态和控制约束的非线性动态最优控制问题。陈金鹤等对倾转旋翼机纵向飞行动力学模型进行地面效应影响修正,开展基于最优控制理论的倾转旋翼机跃障飞行仿真与低速回避区研究[22]。严旭飞等以XV-15为研究对象,根据直升机模式与固定翼飞机模式之间的转换特点,利用非线性最优控制理论,以性能指标、路径约束和边界条件的形式构建了倾转过渡最优控制模型,利用最优控制方法研究XV-15倾转旋翼机的最优动态倾转过渡过程,并得到最优操纵策略,使得由时间、姿态角变化及驾驶员工作负荷等组成的性能指标达到最小[23]。以上大部分研究,对于飞行轨迹优化问题,通常以完成了多操纵舵面分配设计的飞行器为研究对象,开展操纵设计与轨迹优化。这些研究在优化过程中,往往事先确定舵面分配,将混合操纵模式认定为已知情形,没有考虑舵面分配与动态倾转过渡轨迹优化的耦合。

本文针对新构型的半倾转多旋翼飞行器,开展关于旋翼的总距、倾转角、转速和翼面舵面偏角的舵面分配问题分析,在混合操纵模式未知的情况下,建立纵向二维刚体飞行动力学模型;针对倾转过渡问题,确定合适的优化目标函数和约束条件,建立最优控制模型;通过多轮优化计算,在确定飞行器舵面分配的同时,得到对应的过渡飞行最优操纵策略与飞行轨迹。

1 飞行动力学建模

1.1 基本飞行动力学模型

本文研究对象为图1所示的半倾转构型多旋翼垂直起降飞行器。其以固定翼飞机为主体,机翼下吊挂数个支臂,支臂前端各安装有一个可变距操纵的固定转速倾转旋翼,支臂后端各安装有一个可变转速操纵的固定总距旋翼。在由多旋翼模式倾转过渡到固定翼模式的过程中,前旋翼随倾转角减小到轴流状态,后旋翼转速逐渐降低到停转。为研究过渡模式下旋翼及机翼和尾翼舵面的操纵策略,建立该构型纵向平面内的飞行动力学模型。

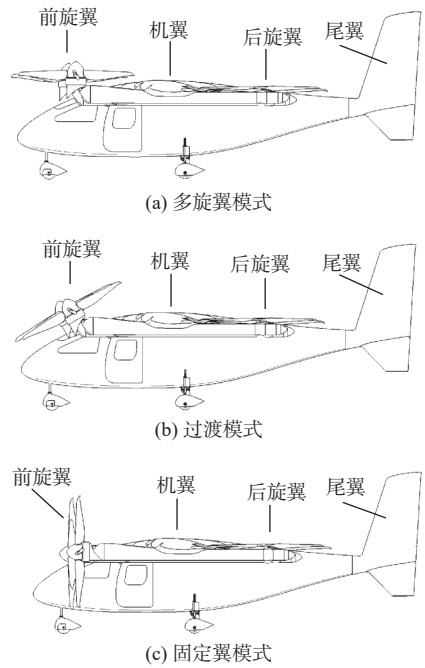


图1 半倾转构型旋翼飞行器飞行模式

Fig. 1 Partial tilting configuration rotor aircraft flight modes

建立各部件的气动力模型[23],其中,前旋翼的气动力$[F_x,F_z,M_y]_{\rm fr}$与旋翼的总距$\theta_{\rm coll}$、倾转角$i_{\rm n}$、前进比λ和入流比μ有关;后旋翼的气动力$[F_x,F_z,M_y]_{\rm br}$与旋翼的转速Ω、前进比λ和入流比μ有关;翼面的气动力$[F_x,F_z,M_y]_{\rm w/t/fus}$与飞行器的速度$u$、迎角$\alpha$、襟翼和尾翼的舵面偏角$\theta_{\rm flap}$和$\theta_{\rm tail}$、不同倾转角$i_{\rm n}$和前飞速度$u$下前旋翼对机翼的气动干扰有关:

$$\begin{cases}[F_x,F_z,M_y]_{\rm fr}=f_{\rm fr}(\theta_{\rm coll},i_{\rm n},\lambda,\mu)\\ [F_x,F_z,M_y]_{\rm br}=f_{\rm br}(\Omega,\lambda,\mu)\\ [F_x,F_z,M_y]_{\rm w/t/fus}=f_{\rm w/t/fus}(\alpha,u,w,\theta_{\rm flap},\theta_{\rm tail},i_{\rm n})\end{cases}\tag{1}$$

基于建立各部件的气动力模型,飞行器的运动方程可表示为

$$\begin{cases}\dot{x}_{\rm b}=f(\boldsymbol{x}_{\rm b},\boldsymbol{u}_{\rm b},t)\\ \boldsymbol{x}_{\rm b}=[u\quad w\quad q\quad \theta]^{\rm T}\\ \boldsymbol{u}_{\rm b}=[\theta_{\rm coll}\quad \Omega\quad \theta_{\rm flap}\quad \theta_{\rm tail}\quad i_{\rm n}]^{\rm T}\end{cases}\tag{2}$$

式中: t 为时间; x_b 为状态量, 包括前飞速度 u、垂向速度 w、俯仰角速度 q 和俯仰角 θ; u_b 为控制量, 包括前旋翼的总距 θ_{coll}、后旋翼的转速 Ω、襟翼和尾翼的舵面偏角 θ_{flap} 和 θ_{tail} 及倾转角 i_n。

通过最优控制方法, 在舵面分配方程未知的情况下, 开展该构型飞行器倾转过渡过程操纵优化计算。考虑到操纵舵面速率的限制, 同时避免在优化计算过程中出现控制量的不连续点或 “bang-bang” 型控制[23], 将原控制量 u 改为状态量, 其一阶导数作为新的控制量。飞行器的运动方程变为

$$\begin{cases} \dot{x}_b = f(\boldsymbol{x}_b, \boldsymbol{u}_b, t) \\ \boldsymbol{x}_b = [u \quad w \quad q \quad \theta \quad \theta_{coll} \quad \Omega \quad \theta_{flap} \quad \theta_{tail} \quad i_n]^T \\ \boldsymbol{u}_b = [\dot{\theta}_{coll} \quad \dot{\Omega} \quad \dot{\theta}_{flap} \quad \dot{\theta}_{tail} \quad \dot{i}_n]^T \end{cases} \tag{3}$$

1.2　舵面分配问题分析

该构型飞行器在纵向平面内存在旋翼的总距 θ_{coll}、转速 Ω、襟翼舵面偏角 θ_{flap}、尾翼舵面偏角 θ_{tail} 和倾转角 i_n 共 5 个操纵舵面。其中, 倾转角在过渡模式下可通过驾驶员或飞控系统独立控制。襟翼主要起到增升的作用, 需要在确定其随前飞速度的操纵规律后, 由飞控系统自动控制。其余 3 个操纵通道均可用于控制全机的力与力矩, 存在操纵冗余, 需确定其与总距杆 δ_{coll} 和纵向杆 δ_{lon} 的舵面分配方程。

半倾转构型多旋翼垂直起降飞行器的舵面分配逻辑与常规倾转旋翼机有一定的相似性[24], 又存在一些差异。其舵面分配方程可表示为

$$\begin{cases} \theta_{coll} = \dfrac{\partial \theta_{coll}}{\partial \delta}\delta_{coll} + \dfrac{\partial \theta_{coll}}{\partial \delta}\delta_{lon}\sin i_n + \theta_{coll0}(i_n) \\ \Omega = \left(\dfrac{\partial \Omega}{\partial \delta}\delta_{coll} + \dfrac{\partial \Omega}{\partial \delta}\delta_{lon}\right)\sin i_n + \Omega_0(i_n) \\ \theta_{tail} = \dfrac{\partial \theta_{tail}}{\partial \delta}\delta_{lon} + \theta_{tail0}(i_n) \\ \theta_{flap} = f_{flap}(u) \end{cases} \tag{4}$$

式中: $\partial\theta_{coll}/\partial\delta$、$\partial\Omega/\partial\delta$ 和 $\partial\theta_{tail}/\partial\delta$ 分别为总距 θ_{coll}、转速 Ω 和尾翼舵面偏角 θ_{tail} 与操纵杆量 δ 的操纵系数。飞行器在不同倾转角下, 其旋翼总距、转速和尾翼舵面的操纵范围差异较大, 需要确定不同倾转角对应的前旋翼总距补偿角 $\theta_{coll0}(i_n)$、后旋翼补偿转速 $\Omega_0(i_n)$、尾翼舵面补偿角 $\theta_{tail0}(i_n)$。前旋翼随倾转角减小, 拉力对全机俯仰力矩减小, 纵向操纵舵效降低; 后旋翼为平衡全机俯仰力矩, 其转速随倾转角减小而降低, 垂向与纵向操纵舵效同时降低。

当舵面分配确定后, 飞行器的运动方程变为

$$\begin{cases} \dot{x}_b = f(\boldsymbol{x}_b, \boldsymbol{u}_b, t) \\ \boldsymbol{x}_b = [u \quad w \quad q \quad \theta \quad \delta_{coll} \quad \delta_{lon} \quad i_n]^T \\ \boldsymbol{u}_b = [\dot{\delta}_{coll} \quad \dot{\delta}_{lon} \quad \dot{i}_n]^T \end{cases} \tag{5}$$

2　最优控制问题分析

为解决半倾转构型多旋翼垂直起降飞行器倾转过渡过程的操纵策略与轨迹优化问题, 使用求解最优控制问题的方法对倾转过渡过程进行分析, 建立对应的最优控制模型。

最优控制问题中需优化的变量通常为优化目标随时间变化的控制量, 即 1.2 节所述飞行器的舵面在整个倾转过程中的操纵量。为求解满足需求的操纵策略, 针对倾转过渡过程, 确定优化模型的目标函数、状态方程、边界条件和路径约束。

2.1　目标函数

飞行器的姿态与操纵的稳定性是保证飞行安全的重要指标, 因此, 将目标函数定义为

$$\min J = \frac{\omega_1}{t_f - t_0}\int_{t_0}^{t_f}\left[\left(\frac{\theta}{\theta_{max}}\right)^2 + \left(\frac{q}{q_{max}}\right)^2\right]\mathrm{d}t + \frac{\omega_2}{t_f - t_0}\int_{t_0}^{t_f}\sum\left[\left(\frac{\Delta}{\Delta_{max}}\right)^2 + \left(\frac{\dot{\Delta}}{\dot{\Delta}_{max}}\right)^2\right]\mathrm{d}t \tag{6}$$

式中: t_0 为初始时刻; t_f 为末端时刻; θ_{max}、q_{max}、Δ_{max}、$\dot{\Delta}_{max}$ 分别为俯仰角、俯仰角速度、控制量、控制量操纵速率的最大值, 以避免不同变量数量级相差过大导致的数值求解困难。ω_1 和 ω_2 为权重系数, 本文将针对不同的权重系数开展分析计算。

2.2　状态方程

优化问题的状态方程为飞行器的运动方程。当舵面分配方程未知时, 状态方程为式 (3); 确定操纵补偿量后, 将舵面分配方程代入飞行器的运动方程, 状态方程为式 (5)。

2.3　边界条件

优化问题的边界条件为飞行器的初始和末端的飞行状态。针对过渡飞行问题, 正向过渡时多旋翼模式悬停为初始状态, 固定翼模式定速前飞为末端状态, 逆向过渡时将初始和末端状态对换。

正向过渡时:

$$\begin{cases} q(t_0) = 0\ (^\circ)/\mathrm{s}, u(t_0) = 0\ \mathrm{m/s}, w(t_0) = 0\ \mathrm{m/s} \\ i_n(t_0) = 90^\circ, \theta_{flap}(t_0) = \theta_{flap}^{max}, q(t_f) = 0\ (^\circ)/\mathrm{s} \\ u(t_f) = V_c, w(t_f) = 0\ \mathrm{m/s}, \Omega(t_f) = 0\ (^\circ)/\mathrm{s} \\ i_n(t_f) = 0^\circ, \theta_{flap}(t_f) = 0^\circ \end{cases} \tag{7}$$

式中: V_c 为飞行器固定翼模式的巡航速度, 为保证初始和末端的飞行状态稳定, 边界的俯仰角速度应限制为 0 (°)/s, 垂向速度应限制为 0 m/s; 固定翼模式为降低气动阻力, 后旋翼停转并锁桨, 限制转速 Ω 为 0 (°)/s; 在悬停及小速度状态下, 为增加翼面升力并降低垂直增重, 襟翼满偏到 θ_{flap}^{max}, 定速巡航时为 0 (°)/s。

同理，逆向过渡时初始和末端状态相反，在此不做赘述。

2.4 路径约束

优化问题的边界条件为飞行器过渡飞行过程中的状态限制，包括

$$\begin{cases} x_{\min} \leqslant x \leqslant x_{\max} \\ u_{\min} \leqslant u \leqslant u_{\max} \\ \dot{x}_{\min} \leqslant \dot{x} \leqslant \dot{x}_{\max} \\ 0 \leqslant P_{\mathrm{r}} \leqslant P_{\max} \end{cases} \tag{8}$$

综合考虑飞行器的典型任务剖面，给出倾转过渡截段的最大时间限制 $t_{\max}$。状态量 x 的限制保证飞行器不出现过大的俯仰角、俯仰角速度和速度。其中，为保证最终得到的舵面分配方程具有一定的普适性，对过渡过程的垂向速度进行严格的限制，定义 $w_{\min}=w_{\max}=0$，使过渡过程中无高度的变化。同时状态量 x 中的操纵量由实际舵面的操纵范围限制；控制量 u 由实际舵面的操纵速率限制；$\dot{x}$ 的限制保证飞行器不出现过大的加速度和角加速度，正向过渡时 $\dot{u}_{\min}=0$，逆向过渡时 $\dot{u}_{\max}=0$，保证前飞速度的单调性；对于分布式多旋翼飞行器，还需给出各旋翼独立的电机功率限制 P_{r}。

3 倾转过渡操纵策略研究

针对半倾转多旋翼垂直起降新构型飞行器，由于其前后旋翼及机翼和尾翼舵面数量较多，且操作方式各不相同，过渡飞行时的最优飞行轨迹是未知的，其冗余舵面的操纵分配方程同样是未知的。分析其舵面分配方程，可知方程中的未知量主要包括2类：一种是不同倾转角下的操纵面补偿量 $\theta_{\mathrm{coll0}}(i_{\mathrm{n}})$、$\Omega_0(i_{\mathrm{n}})$ 和 $\theta_{\mathrm{tail0}}(i_{\mathrm{n}})$，适当的补偿量可以有效降低操纵负荷；另一种是舵面与杆量的操纵系数 $\partial\theta_{\mathrm{coll}}/\partial\delta$、$\partial\Omega/\partial\delta$ 和 $\partial\theta_{\mathrm{tail}}/\partial\delta$，其决定了操纵驾驶杆时同时控制的不同舵面间的分配比例关系。

若基于传统的稳态配平计算方式，参考倾转旋翼机的舵面分配策略[24]，可以多旋翼模式悬停状态和固定翼模式巡航状态的操纵量配平结果为基准，通过线性差值的方式确定不同倾转角下的补偿量；以各个舵面操纵范围比例为基准，确定不同舵面间的分配比例关系。此时

$$\begin{cases} \theta_{\mathrm{coll0}}(i_{\mathrm{n}}) = \dfrac{\theta_{\mathrm{coll}}^{\mathrm{H}}-\theta_{\mathrm{coll}}^{\mathrm{W}}}{i_{\mathrm{n}}^{\mathrm{H}}-i_{\mathrm{n}}^{\mathrm{W}}} i_{\mathrm{n}} + \theta_{\mathrm{coll}}^{\mathrm{W}} \\ \Omega_0(i_{\mathrm{n}}) = \dfrac{\Omega^{\mathrm{H}}-\Omega^{\mathrm{W}}}{i_{\mathrm{n}}^{\mathrm{H}}-i_{\mathrm{n}}^{\mathrm{W}}} i_{\mathrm{n}} + \Omega^{\mathrm{W}} \\ \theta_{\mathrm{tail0}}(i_{\mathrm{n}}) = \dfrac{\theta_{\mathrm{tail}}^{\mathrm{H}}-\theta_{\mathrm{tail}}^{\mathrm{W}}}{i_{\mathrm{n}}^{\mathrm{H}}-i_{\mathrm{n}}^{\mathrm{W}}} i_{\mathrm{n}} + \theta_{\mathrm{tail}}^{\mathrm{W}} \\ \dfrac{\partial\theta_{\mathrm{coll}}}{\partial\delta} : \dfrac{\partial\Omega}{\partial\delta} : \dfrac{\partial\theta_{\mathrm{tail}}}{\partial\delta} = \theta_{\mathrm{coll}}^{\max} : \Omega^{\max} : \theta_{\mathrm{tail}}^{\max} \end{cases} \tag{9}$$

式中：$\theta_{\mathrm{coll}}^{\mathrm{H}}$、$\Omega^{\mathrm{H}}$、$\theta_{\mathrm{tail}}^{\mathrm{H}}$ 和 $i_{\mathrm{n}}^{\mathrm{H}}$ 为多旋翼模式悬停状态的操纵量；$\theta_{\mathrm{coll}}^{\mathrm{W}}$、$\Omega^{\mathrm{W}}$、$\theta_{\mathrm{tail}}^{\mathrm{W}}$ 和 $i_{\mathrm{n}}^{\mathrm{W}}$ 为固定翼模式巡航状态的操纵量；$\theta_{\mathrm{coll}}^{\max}$、$\Omega^{\max}$、$\theta_{\mathrm{tail}}^{\max}$ 为各个舵面操纵范围，通常 $i^{\mathrm{H}}=90°$、$i^{\mathrm{W}}=0°$、$\theta_{\mathrm{tail}}^{\mathrm{H}}=0°$、$\Omega^{\mathrm{W}}=0\,(°)/\mathrm{s}$。

本文基于最优控制理论，通过调整状态方程与目标函数，分步多次计算，得到半倾转构型多旋翼垂直起降飞行器过渡飞行舵面分配策略与最优飞行路径。首先，将各独立舵面的操纵速率作为控制量，优化计算得到舵面分配方程的操纵面补偿量；然后，将操纵杆的操纵速率作为控制量，优化计算得到舵面分配方程的舵面操纵系数；最后，得到完整的舵面分配方程后，优化计算得到最优过渡路径。

3.1 操纵面补偿量求解

当舵面分配方程未知时，假设所有的舵面均可独立操纵，此时优化计算的状态方程为式(3)，目标函数中：

$$\begin{cases} \boldsymbol{\Lambda} = [\theta_{\mathrm{coll}} \quad \Omega \quad \theta_{\mathrm{flap}} \quad \theta_{\mathrm{tail}} \quad i_{\mathrm{n}}]^{\mathrm{T}} \\ \dot{\boldsymbol{\Lambda}} = u_{\mathrm{b}} = [\dot{\theta}_{\mathrm{coll}} \quad \dot{\Omega} \quad \dot{\theta}_{\mathrm{flap}} \quad \dot{\theta}_{\mathrm{tail}} \quad \dot{i}_{\mathrm{n}}]^{\mathrm{T}} \end{cases} \tag{10}$$

ω_1 和 ω_2 为姿态角与操纵量权重系数[25]，由于2.1节中已对姿态角与操纵量进行了无量纲化处理，因此，可以直接取 $\omega_1:\omega_2=1:1$、$\omega_1:\omega_2=2:1$、$\omega_1:\omega_2=1:2$ 开展优化计算。

当飞行器正向过渡时，计算结果如图2所示。在正向过渡的计算结果中，不同 $\omega_1:\omega_2$ 值下的过渡路径的状态量变换趋势一致，ω_1 增大，飞行器俯仰角变化量略有减小，ω_2 增大，舵面操纵量略有减小，不同权重系数下优化计算得到的优化目标的比值如表1所示。其中，优化目标变量的定义同式(6)，下标(1∶1)、(2∶1)和(1∶2)分别表示取 $\omega_1:\omega_2=1:1$、$\omega_1:\omega_2=2:1$、$\omega_1:\omega_2=1:2$。

如图2所示，飞行器的倾转角以较为稳定的操纵速率减小；前飞速度在以较为稳定的加速度增大到倾转角为40°左右时，加速度开始降低；飞行器全程以较小的俯仰角抬头飞行，俯仰角呈先增大后减小再增大的趋势；前旋翼总距维持在悬停状态大小直至倾转角为60°左右，随后开始以较为稳定的操纵速率增大；后旋翼转速维持在悬停状态大小直至倾转角为70°左右，随后开始以较为稳定的操纵速率减小，到倾转角为20°左右，操纵速率逐渐增大；尾翼舵面全程以较小的角度下偏；襟翼舵面偏角全程匀速减小。

在逆向过渡的计算结果中，不同 $\omega_1:\omega_2$ 值对计算结果的影响与正向过渡时类似，不同权重系数下优化计算得到的优化目标的比值如表2所示。

如图3所示，飞行器的倾转角以较为稳定的操纵速率增大到70°左右后，倾转速率开始减小；前飞

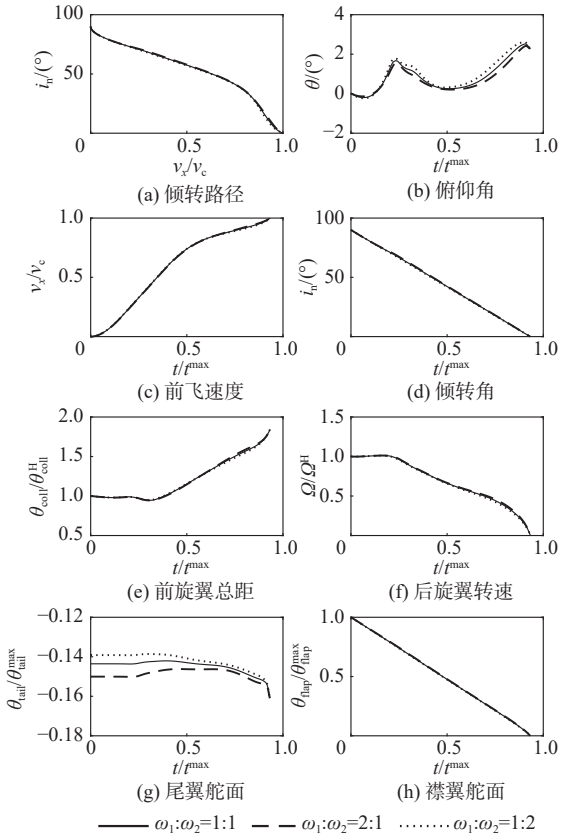


图 2　基于操纵面的正向动态倾转过渡最优解

Fig. 2　Optimal solution for accelerating dynamic conversion based on control surfaces

表 1　基于操纵面的正向过渡不同权重系数优化目标比值

Table 1　Optimizing objective ratios for different weight coefficients for accelerating dynamic conversion based on control surfaces

优化目标变量	优化目标比值
$\int\theta_{(1:1)}^2:\int\theta_{(2:1)}^2:\int\theta_{(1:2)}^2$	1∶0.78∶1.27
$\int q_{(1:1)}^2:\int q_{(2:1)}^2:\int q_{(1:2)}^2$	1∶0.85∶1.16
$\sum\int\Delta_{(1:1)}^2:\sum\int\Delta_{(2:1)}^2:\sum\int\Delta_{(1:2)}^2$	1∶1.01∶0.99
$\sum\int\dot{\Delta}_{(1:1)}^2:\sum\int\dot{\Delta}_{(2:1)}^2:\sum\int\dot{\Delta}_{(1:2)}^2$	1∶1.09∶0.98

表 2　基于操纵面的逆向过渡不同权重系数优化目标比值

Table 2　Different weight coefficients for optimizing objective ratios for decelerating dynamic conversion based on control surfaces

优化目标变量	优化目标比值
$\int\theta_{(1:1)}^2:\int\theta_{(2:1)}^2:\int\theta_{(1:2)}^2$	1∶0.67∶1.44
$\int q_{(1:1)}^2:\int q_{(2:1)}^2:\int q_{(1:2)}^2$	1∶0.61∶1.91
$\sum\int\Delta_{(1:1)}^2:\sum\int\Delta_{(2:1)}^2:\sum\int\Delta_{(1:2)}^2$	1∶1.12∶0.91
$\sum\int\dot{\Delta}_{(1:1)}^2:\sum\int\dot{\Delta}_{(2:1)}^2:\sum\int\dot{\Delta}_{(1:2)}^2$	1∶1.75∶0.64

速度在以较为稳定的加速度减速到倾转角为 70°左右时, 加速度开始增大; 飞行器全程抬头飞行, 俯仰角起始时随倾转角增大而逐渐增大, 在倾转角为 70°左右, 角速度迅速增大, 随后俯仰角开始逐渐减小到 0°; 前旋翼总距起始时随倾转角增大而逐渐减小, 在倾转角为 70°左右开始逐渐增大; 后旋翼随倾转角增大而逐渐增大, 到倾转角为 70°左右到达并稳定在悬停转速左右, 操纵速率呈先增大后减小再增大的趋势; 尾翼下偏角度随倾转角增大而逐渐增大, 到倾转角为 70°左右开始稳定在一个较大的角度; 襟翼舵面偏角随速度增大而增大, 操纵速率先增大后减小。

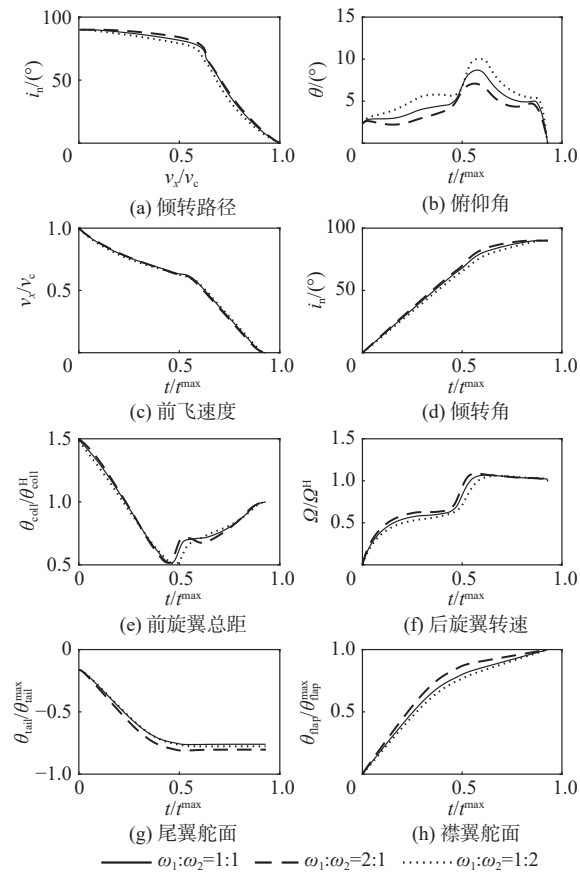


图 3　基于操纵面的逆向动态倾转过渡最优解

Fig. 3　Optimal solution for decelerating dynamic conversion based on control surfaces

基于上述计算结果, 可得到不同目标函数下正向与逆向过渡过程中各操纵量随倾转角和前飞速度的变化规律$\theta_{\mathrm{coll}}(i_{\mathrm{n}})$、$\Omega(i_{\mathrm{n}})$、$\theta_{\mathrm{tail}}(i_{\mathrm{n}})$和$\theta_{\mathrm{flap}}(u)$。$\omega_1$越大, 飞行姿态越稳定, ω_2越大, 飞行操纵负荷越低, 正向与逆向过渡的结果变化趋势相似, 数值有一定的差异。

为保证在确定舵面分配策略后, 飞行器在正向与逆向过渡中都能有稳定的姿态和较低的操纵负荷, 综合考虑选取$\omega_1:\omega_2=1:1$的计算结果, 取正向与逆向计算的中值作为舵面分配方程中的补偿量$\theta_{\mathrm{coll0}}(i_{\mathrm{n}})$、$\Omega_0(i_{\mathrm{n}})$、$\theta_{\mathrm{tail0}}(i_{\mathrm{n}})$和$\theta_{\mathrm{flap0}}(u)$, 计算结果与式 (9) 的结果对比如图 4 所示。

与基于配平计算的结果相比, 基于优化计算得到的舵面补偿量中, 前旋翼总距随倾转角增大先减

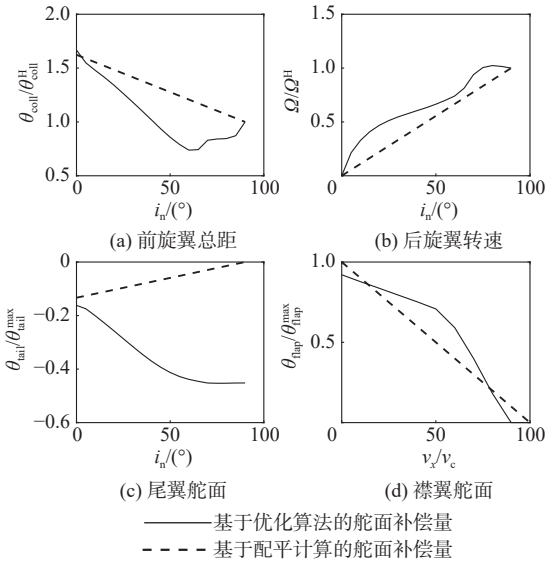


图 4 舵面分配补偿量

Fig. 4 The compensation of the rudder allocation

小后增大，总体比基于配平插值的计算结果更小，而后旋翼转速随倾转角增大而逐渐增大，总体比基于配平插值的计算结果更大，这是由旋翼总距操纵与转速操纵特性的差异导致的，优化结果相比直接插值更加合理；优化计算得到的襟翼与尾翼舵面操纵量更大，充分利用舵面操纵，更有利于改善飞行器过渡飞行中的操纵量和姿态角。

3.2 舵面操纵系数求解

确定操纵补偿量后，将舵面分配方程代入飞行器运动方程，此时，操纵系数为未知量，需要通过优化计算求解，状态方程变为

$$\begin{cases}\dot{\boldsymbol{x}}_{\mathrm{b}}=f(\boldsymbol{x}_{\mathrm{b}},\boldsymbol{u}_{\mathrm{b}},p_{\mathrm{b}},t)\\ \boldsymbol{x}_{\mathrm{b}}=[u\quad w\quad q\quad \theta\quad \delta_{\mathrm{coll}}\quad \delta_{\mathrm{lon}}\quad i_{\mathrm{n}}]^{\mathrm{T}}\\ \boldsymbol{u}_{\mathrm{b}}=[\dot{\delta}_{\mathrm{coll}}\quad \dot{\delta}_{\mathrm{lon}}\quad \dot{i}_{\mathrm{n}}]^{\mathrm{T}}\\ \boldsymbol{p}_{\mathrm{b}}=\left[\dfrac{\theta_{\mathrm{coll}}}{\delta}\quad \dfrac{\Omega}{\delta}\quad \dfrac{\theta_{\mathrm{tail}}}{\delta}\right]^{\mathrm{T}}\end{cases}\tag{11}$$

需要注意的是，由于操纵杆量δ_{coll}、δ_{lon}与操纵系数p_{b}有关，若目标函数中的Δ定义为操纵杆量，为使目标函数最小，优化计算结果中的p_{b}会取限制范围内的最大值，不满足设计需求。因此，控制量定义与 3.1 中一致。但对于操纵速率，应重点关注操纵杆处驾驶员/飞控系统的操纵负荷，因此，

$$\begin{cases}\boldsymbol{\Delta}=[\theta_{\mathrm{coll}}\quad \Omega\quad \theta_{\mathrm{flap}}\quad \theta_{\mathrm{tail}}\quad i_{\mathrm{n}}]^{\mathrm{T}}\\ \dot{\boldsymbol{\Delta}}=[\dot{\delta}_{\mathrm{coll}}\quad \dot{\delta}_{\mathrm{lon}}\quad \dot{i}_{\mathrm{n}}]^{\mathrm{T}}\end{cases}\tag{12}$$

当飞行器正向过渡时，计算结果如表 3 所示。图 5 为正向动态倾转过渡最优解。

在正向过渡的计算结果中，不同$\omega_1:\omega_2$值下的过渡路径的状态量变换趋势一致，ω_1越大，俯仰角变化量越小，操纵杆变化量越大，操纵系数中Ω/δ和$\theta_{\mathrm{tail}}/\delta$的比例越大，不同权重系数下优化计算得到的优化目标的比值如表 4 所示。

表 3 正向过渡舵面操纵分配系数

Table 3 Control distribution coefficients of rudder for accelerating dynamic conversion

$\omega_1:\omega_2$	$\dfrac{\partial\theta_{\mathrm{coll}}}{\partial\delta}:\dfrac{\partial\Omega}{\partial\delta}:\dfrac{\partial\theta_{\mathrm{tail}}}{\partial\delta}$
2:1	$\theta_{\mathrm{coll}}^{\max}:1.94\Omega^{\max}:0.41\theta_{\mathrm{tail}}^{\max}$
1:1	$\theta_{\mathrm{coll}}^{\max}:1.54\Omega^{\max}:0.37\theta_{\mathrm{tail}}^{\max}$
1:2	$\theta_{\mathrm{coll}}^{\max}:1.31\Omega^{\max}:0.33\theta_{\mathrm{tail}}^{\max}$

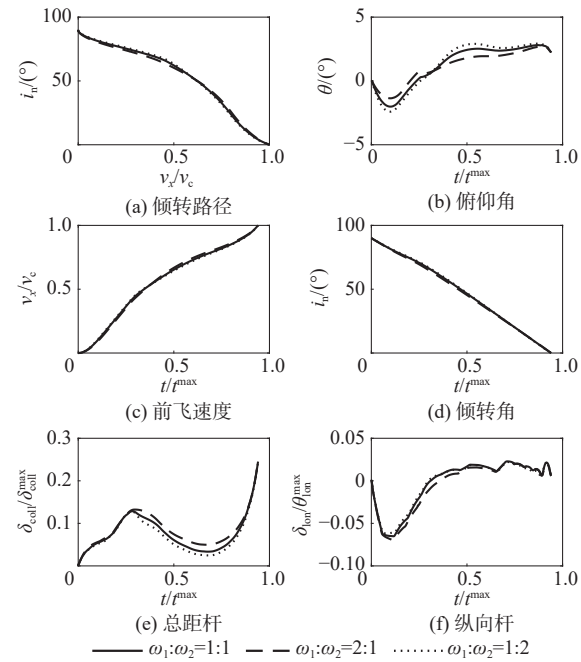


图 5 基于舵面分配的正向动态倾转过渡最优解

Fig. 5 Optimal solution for accelerating dynamic conversion based on control surface allocation

图 5 中，前飞速度与倾转角的变化趋势与 3.1 中的图 2 计算结果一致，俯仰角在 70°前出现较小的低头角度；总距杆量保持为正值，呈先增大后减小再增大的趋势；纵向杆操纵量较小，呈先减小再增大的趋势。

当飞行器逆向过渡时，计算结果如表 5 所示。

在逆向过渡的计算结果中，不同$\omega_1:\omega_2$值对计算结果的影响与正向过渡时类似，不同权重系数下优化计算得到的优化目标的比值如表 6 所示。

表 4 基于舵面分配的正向过渡不同权重系数优化目标比值

Table 4 Optimizing objective ratios for different weight coefficients for accelerating dynamic conversion based on control surface allocation

优化目标变量	优化目标比值
$\int\theta_{(1:1)}^2:\int\theta_{(2:1)}^2:\int\theta_{(1:2)}^2$	1:0.69:1.24
$\int q_{(1:1)}^2:\int q_{(2:1)}^2:\int q_{(1:2)}^2$	1:0.61:1.37
$\sum\int\Delta_{(1:1)}^2:\sum\int\Delta_{(2:1)}^2:\sum\int\Delta_{(1:2)}^2$	1:1.32:0.87
$\sum\int\dot{\Delta}_{(1:1)}^2:\sum\int\dot{\Delta}_{(2:1)}^2:\sum\int\dot{\Delta}_{(1:2)}^2$	1:1.57:0.83

表 5　逆向过渡舵面操纵分配系数

Table 5　Control distribution coefficients of rudder for decelerating dynamic conversion

$\omega_1:\omega_2$	$\frac{\partial\theta_{\text{coll}}}{\partial\delta}:\frac{\partial\Omega}{\partial\delta}:\frac{\partial\theta_{\text{tail}}}{\partial\delta}$
2:1	$\theta_{\text{coll}}^{\max}:0.42\Omega^{\max}:1.55\theta_{\text{tail}}^{\max}$
1:1	$\theta_{\text{coll}}^{\max}:0.45\Omega^{\max}:1.11\theta_{\text{tail}}^{\max}$
1:2	$\theta_{\text{coll}}^{\max}:0.45\Omega^{\max}:0.98\theta_{\text{tail}}^{\max}$

表 6　基于舵面分配的逆向过渡不同权重系数优化目标比值

Table 6　Optimizing objective ratios for different weight coefficients for decelerating dynamic conversion based on control surface allocation

优化目标变量	优化目标比值
$\int\theta_{(1:1)}^2:\int\theta_{(2:1)}^2:\int\theta_{(1:2)}^2$	1:1.01:0.93
$\int q_{(1:1)}^2:\int q_{(2:1)}^2:\int q_{(1:2)}^2$	1:1.17:0.82
$\sum\int\Delta_{(1:1)}^2:\sum\int\Delta_{(2:1)}^2:\sum\int\Delta_{(1:2)}^2$	1:0.98:1.01
$\sum\int\dot{\Delta}_{(1:1)}^2:\sum\int\dot{\Delta}_{(2:1)}^2:\sum\int\dot{\Delta}_{(1:2)}^2$	1:0.97:1.05

如图 6 所示, 前飞速度、倾转角和俯仰角的变化趋势与 3.1 中的图 3 计算结果一致; 总距杆量与纵向杆量均保持为负值, 呈先增大后减小的趋势。

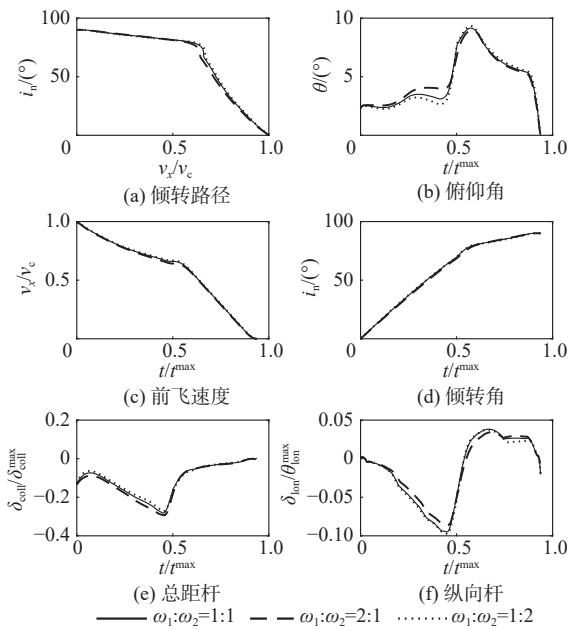


图 6　基于舵面分配的逆向动态倾转过渡最优解

Fig. 6　Optimal solution for decelerating dynamic conversion based on control surface allocation

对比正向过渡与逆向过渡的计算结果, 其舵面操纵系数有一定的差异: 正向过渡中Ω/δ的占比更大, 即后旋翼转速的操纵权限更大; 逆向过渡中$\theta_{\text{tail}}/\delta$的占比更大, 即尾翼舵面的操纵权限更大。为保证在确定舵面分配策略后, 飞行器在正向与逆向过渡中都能有较低的操纵负荷, 取正向与逆向计算的中值作为舵面分配方程中的操纵系数, 结果如表 7 所示。综合考虑飞行姿态和操纵负荷, 以$\omega_1:\omega_2=1:1$

表 7　舵面操纵分配系数

Table 7　Control distribution coefficients of rudder

$\omega_1:\omega_2$	$\frac{\partial\theta_{\text{coll}}}{\partial\delta}:\frac{\partial\Omega}{\partial\delta}:\frac{\partial\theta_{\text{tail}}}{\partial\delta}$
2:1	$\theta_{\text{coll}}^{\max}:0.98\Omega^{\max}:0.96\theta_{\text{tail}}^{\max}$
1:1	$\theta_{\text{coll}}^{\max}:0.88\Omega^{\max}:0.72\theta_{\text{tail}}^{\max}$
1:2	$\theta_{\text{coll}}^{\max}:0.85\Omega^{\max}:0.64\theta_{\text{tail}}^{\max}$

的计算结果作为最终的舵面操纵系数。

3.3　倾转过渡操纵策略确定

基于上述计算结果, 最终确定了飞行器的舵面分配方程, 为验证优化计算的有效性, 对比式 (9) 确定的舵面分配策略, 开展优化计算。此时, 优化计算的状态方程为式 (5), 目标函数中:

$$\begin{cases}\boldsymbol{\Delta}=[\delta_{\text{coll}}\quad\delta_{\text{lon}}\quad i_{\text{n}}]^{\text{T}}\\ \dot{\boldsymbol{\Delta}}=[\dot{\delta}_{\text{coll}}\quad\dot{\delta}_{\text{lon}}\quad\dot{i}_{\text{n}}]^{\text{T}}\end{cases}\tag{13}$$

当飞行器正向过渡时, 计算结果如图 7 所示。

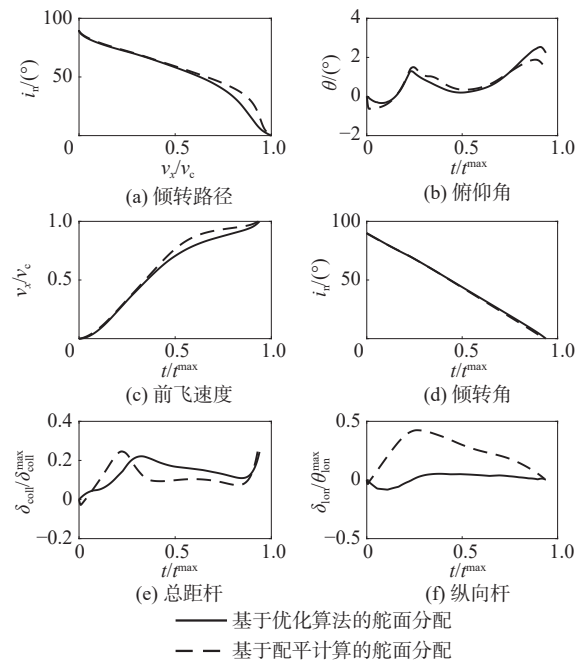


图 7　正向动态倾转过渡最优解

Fig. 7　Optimal solution for accelerating dynamic conversion

在正向过渡的计算结果中, 采用 2 种舵面分配, 速度、倾转角和俯仰角的变化趋势基本一致; 总距杆量均呈先增大后减小的趋势; 使用基于优化算法的舵面分配, 纵向杆先减小后增大, 最后保持在 0 附近, 操纵量较小, 使用基于配平计算的舵面分配, 纵向杆先增大后减小, 操纵量较大。对数据进行定量分析, 使用优化算法计算得到过渡飞行时的总距杆量和纵向杆量最大值分别为使用配平计算的舵面分配计算的 0.95 和 0.19 倍, 俯仰角基本一致, 俯仰角速度、总距操纵速率和纵向操纵速率最大值分别为 0.2、0.79 和 0.72 倍。

当飞行器逆向过渡时, 计算结果如图 8 所示。

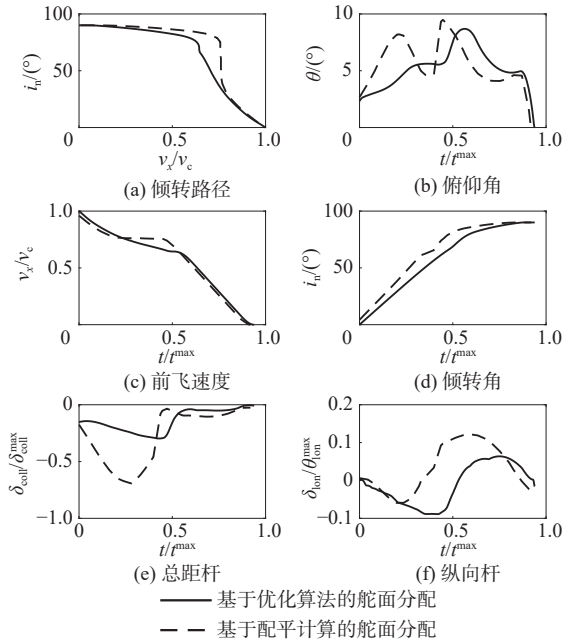


图8 逆向动态倾转过渡最优解

Fig. 8 Optimal solution for decelerating dynamic conversion

在逆向过渡的计算结果中，采用2种舵面分配，速度和倾转角的变化趋势基本一致；俯仰角呈增大-减小-增大-减小的趋势；总距杆先减小后增大，最后保持在0附近；纵向杆呈减小-增大-减小的趋势。使用优化算法计算的总距杆和纵向杆操纵量明显小于基于配平计算的操纵量。对数据进行定量分析，使用优化算法计算得到过渡飞行时的俯仰角、总距杆量和纵向杆量最大值分别为使用配平计算的舵面分配计算的0.92、0.42和0.72倍，俯仰角速度和总距操纵速率最大值分别为0.99和0.39倍，纵向操纵速率基本一致。

综上所述，相比传统的舵面分配策略，使用优化算法计算得到的结果在操纵量和姿态角上均有明显的改善。对正向与逆向过渡飞行的计算结果取均值，俯仰角、总距杆量和纵向杆量最大值平均降低为传统策略的0.96、0.685和0.455倍，俯仰角速度、总距操纵速率和纵向操纵速率最大值平均降低为传统策略的0.6、0.89和0.555倍。

4 结 论

本文以半倾转构型多旋翼垂直起降飞行器为研究对象，基于最优控制理论，开展关于舵面分配与过渡路径优化的操纵策略问题研究。

1）建立飞行器的飞行动力学模型，开展舵面分配问题分析。建立最优控制模型，并基于优化模型开展多次优化计算，逐步确定了飞行器舵面分配方程中的操纵面补偿量和舵面操纵系数，同步确定了最优过渡路径与操纵策略。

2）相比于传统的操纵策略研究方法，使用优化算法得到的过渡阶段俯仰角、总距杆量和纵向杆量最大值分别降低为0.96、0.685和0.455倍，俯仰角速度、总距操纵速率和纵向操纵速率最大值分别降低为0.6、0.89和0.555倍，显著改善了操纵量和姿态角，这对于降低飞行器过渡飞行的操纵负荷、提高飞行安全性具有重要意义。

参考文献（References）

[1] YAN X F, YUAN Y, CHEN R L. Research on pilot control strategy and workload for tilt-rotor aircraft conversion procedure[J]. Aerospace, 2023, 10(9): 742.

[2] LI F Y, XU W Q, SHI Y, et al Multi-body dynamic modeling, simulation and control strategy design of a Y6 tilt rotor UAV[C]//Proceedins of the 2017 2nd International Conference on Advanced Robotics and Mechatronics. Piscataway: IEEE Press, 2017: 373-379.

[3] BAUERSFELD L, SPANNAGL L, DUCARD G J J, et al. MPC flight control for a tilt-rotor VTOL aircraft[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(4): 2395-2409.

[4] ZHOU P, CHEN R L, YU Z M. Investigation of control strategy for quad-tilt-rotor aircraft in helicopter mode[J]. Journal of Aeronautical Dynamics, 2021, 36(10): 2036-2051.

[5] NIETO-HERNÁNDEZ D, MÉNDEZ-BARRIOS C F, ESCARENO J, et al. Non-holonomic flight modeling and control of a tilt-rotor MAV[C]//Proceedins of the 2019 6th International Conference on Control, Decision and Information Technologies. Piscataway: IEEE Press, 2019: 1947-1952.

[6] MA Z Q, SMEUR E J J, DE CROON G C H E. Design and control of a tilt-rotor tail sitter aircraft with pivoting VTOL capability[J]. IEEE Robotics and Automation Letters, 2025, 10(6): 5911-5918.

[7] FANG T R, DU C P, YANG R, et al. Nonlinear model predictive control of tilt-rotor UAV under A dual differential model based on INDI strategy[J]. Unmanned Systems , 2025: 1-14.

[8] 王梓旭, 李攀, 沈俊彪, 等. 加/减速状态下倾转旋翼飞行器动态过渡走廊[J]. 航空学报, 2025, 46(11): 366-385.
WANG Z X, LI P, SHEN J B, et al. Dynamic conversion corridor of tiltrotor aircraft under accelerating and decelerating conditions[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(11): 366-385(in Chinese).

[9] 徐绍峰, 常绍平, 金鑫, 等. 面向倾转旋翼机的TRC响应类型控制律设计[J]. 航空科学技术, 2023, 34(9): 87-93.
XU S F, CHANG S P, JIN X, et al. Design of TRC response type control law for tiltrotor aircraft[J]. Aeronautical Science & Technology, 2023, 34(9): 87-93(in Chinese).

[10] 吴榕. 变转速/变直径倾转旋翼机飞行性能及操纵策略分析[D]. 南京: 南京航空航天大学, 2022.
WU R. Analyses on flight performance and control strategy of variable rotation speed/variable diameter tilt-rotor aircraft[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2022(in Chinese).

[11] WANG Z, BAO W, LI Y, et al. A survey of control allocation methods for spacecraft and aerial vehicles[J]. Acta Astronautica, 2022, 190: 85-103.

[12] LIU X X, YUAN Z K, GAO Z W, et al. Reinforcement learning-based fault-tolerant control for quadrotor UAVs under actuator fault[J]. IEEE Transactions on Industrial Informatics, 2024, 20(12): 13926-13935.

[13] JOHANSEN T A, FOSSEN T I, BERGE S P. Constrained nonlinear control allocation with singularity avoidance using sequential quadratic programming[J]. IEEE Transactions on Control Systems Technology, 2004, 12(1): 211-216.

[14] DI MILANO P, RIGHETTI A, MUSCARELLO V, et al. Linear parameter varying models for the optimization of tiltrotor conversion maneuver[C]//Proceedings of the Vertical Flight Society 73rd Annual Forum. Fort Worth: The Vertical Flight Society, 2017: 1-8.

[15] 蒋海辉. 倾转四旋翼飞行器操纵策略研究[D]. 南京: 南京航空航天大学, 2019.
JIANG H H. Manipulation strategy for quad tilt rotor aircraft[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019(in Chinese).

[16] KLEINHESSELINK K M. Stability and control modeling of tiltrotor aircraft[D]. College Park: University of Maryland, 2007.

[17] LUO L W, XU S, SU C Y, et al. Fixed-time composite observer-based adaptive recursive sliding mode control for tiltrotor UAV attitude tracking[J]. Nonlinear Dynamics, 2025, 113(15): 20031-20056.

[18] LI W, SHI S, CHEN M, et al. Composite anti-disturbance tracking control for the conversion mode of tilt-rotor aircraft via switched fuzzy modeling[J]. International Journal of Control, Automation and Systems, 2025, 23(3): 960-972.

[19] CHEN Q, ZHANG N, XUE L L, et al. Research on mathematical modeling method of tilt-rotor aircraft[C]//Proceedins of the Advances in Guidance, Navigation and Control. Berlin: Springer, 2025: 330-338.

[20] CHEN Z Y, ZHANG Z Q, SHENG S Z, et al. Transition trajectory planning for tilt-rotor aircraft based on SA-IGA algorithm[C]//Proceedins of the Advances in Guidance, Navigation and Control. Berlin: Springer, 2025: 298-308.

[21] GONG H X, HE W, HOU S P, et al. Design and validation of a new tilting rotor VTOL drone: structural optimization, flight dynamics, and PID control[J]. Sensors, 2025, 25(11): 3537-3579.

[22] 陈金鹤, 汪正中, 田洪源. 基于最优控制理论的倾转旋翼机跃障飞行仿真[J]. 西北工业大学学报, 2020, 38(6): 1266-1274.
CHEN J H, WANG Z Z, TIAN H Y. Hurdle-hop simulation of tilt-rotor aircraft based on optimal control theory[J]. Journal of Northwestern Polytechnical University, 2020, 38(6): 1266-1274(in Chinese).

[23] 严旭飞, 陈仁良. 倾转旋翼机动态倾转过渡过程的操纵策略优化[J]. 航空学报, 2017, 38(7): 520865.
YAN X F, CHEN R L. Control strategy optimization of dynamic conversion procedure of tilt-rotor Aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(7): 520865(in Chinese).

[24] FERGUSON S W. A mathematical model for real time flight simulation of a generic tilt rotor aircraft: NASA CR-166536[R]. Washington, D. C. : NASA, 1988.

[25] 严旭飞. 倾转旋翼机倾转过渡操纵策略研究[D]. 南京: 南京航空航天大学, 2019.
YAN X F. Research on control strategy of conversion procedure of tilt-rotor aircraft[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019(in Chinese).

Research on longitudinal manipulation strategy of partial tilting aircraft

TIAN Hongyuan*, LU Senkui, WU Shangjing

(China Helicopter Research and Development Institute, National Key Laboratory of Helicopter Aeromechanics, Jingdezhen 333001, China)

Abstract: The partial tilting aircraft, a novel kind of vertical take-off and landing aircraft, lacks a suitable way to address the operational redundancy brought on by its multiple rudders. Based on the optimal control theory, the problem of longitudinal manipulation of the partial tilting aircraft is carried out, and the distribution of rudder and the dynamic tilt rotation are optimized. The longitudinal rigid body flight mechanics model was established. The first derivative of the control quantity is used as the control quantity to avoid the jumping discontinuity in the optimization process. In this research, a hybrid control equation is established and the longitudinal control surface allocation problem is investigated. The trajectory optimization problem of the transition of the aircraft is transformed into a nonlinear dynamic optimal control problem. This study selects reasonable optimization objectives and constraints to establish an optimal control model. The equation and the target function of the rudder surface are selected. The control surface allocation and optimal maneuvering strategy for transitional flight of the aircraft are determined simultaneously with the flight trajectory. Compared with the traditional approach of first determining control surface allocation based on trim calculations and then determining the flight trajectory through optimization, this method significantly improves attitude stability and reduces control load. The effectiveness of the method has been validated through simulation analysis.

Keywords: partial tilting configuration; optimal control; dynamic conversion; maneuvering strategy; trajectory optimization

Received: 2025-07-22; **Accepted**: 2025-10-17; **Published Online**: 2025-10-21 15:23

URL: link.cnki.net/urlid/11.2625.V.20251021.1309.002

Foundation item: National Key Laboratory of Helicopter Aeromechanica Fund (61422202205)

* **Corresponding author**. E-mail: tianhy@avic.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0781

基于小波变换的压力脉动信号分析方法

覃子宇[1,2]，韩猛[1]，韩啸[1,*]，赵硕[3]，姜宇涵[1]，林宇震[1]

(1. 北京航空航天大学 航空发动机研究院，北京 100191；2. 北京航空航天大学 沈元学院，北京 100191；
3. 中国航发湖南动力机械研究所，株洲 412002)

摘　　要：在航空发动机燃烧室高温高压实验中采集动态压力脉动信号，通过离散小波变换(DWT)和连续小波变换(CWT)分析信号从稳定状态演变为不稳定状态的非平稳特性。结果表明：压力序列存在2种振荡模式的切换，小波分析方法能够有效识别燃烧室压力脉动动态转变过程。在转变时刻后，压力脉动表现为声学模态频率下的极限环热声振荡，在转变时刻前，压力脉动具有频率调制的特性，瞬时振荡频率随时间周期性变化。同时，运用离散小波变换可实现相比傅里叶变换更好的早期预报性能。

关 键 词：热声不稳定性；小波分析；时频表示；频率调制；早期预报
中图分类号：V231.2
文献标志码：A　　**文章编号**：1001-5965(2026)03-0801-08

实时监测、早期预报、主动控制对保障燃气轮机的平稳运行具有重要意义，有助于从源头遏制各类故障模式发展，避免其对燃气轮机产生持续严重的损害。而对于燃气轮机的燃烧室部件，热声不稳定性(也称为燃烧振荡)是实际应用中常见的一种故障类型。这一现象源于释热场、压力场、流场等多物理场耦合作用引发的燃烧室声学模态共振，压力脉动振幅逐渐增大并在非线性作用下趋于饱和[1]。在实际应用环境中，压力脉动是一种相对容易获取的监测信号。已有研究通常使用压力脉动判断系统是否处于热声振荡状态[2-3]，如傅里叶变换频谱图[4]、相空间重构方法[5]、递归图[6]等。同时压力脉动也常用于热声振荡的预报和控制中。从压力信号的演化过程中识别先兆特征并给出预警判断，以使控制系统参与调控。例如，An等[7]基于临界慢化现象中的典型参数，包括滑窗脉动方差和一阶自回归系数，通过计算发现这2个参数在热声振荡发生前具有明显的增长趋势。Delaat等[8]提出自适应滑动相平均控制算法对热声振荡进行主动控制，当压力传感器监测热声振荡发生时，通过高频燃油阀门扰动燃料流并使其与热声振荡相位相互抵消，达到抑制效果。Zhang等[9]设计了应用神经网络的主动控制方法，包含2个神经网络模块，一个用于预测被控对象下一时刻压力脉动值，另一个用于根据观测值和预测值输出控制变量的调节幅度。基于统计学分析的压力脉动处理方法依赖于人工构造或算法自适应优化得到的特征，难以从深层机理层面做出全面的解释。

在燃烧振荡现象中，压力脉动表现为声学模态频率处的窄带振荡，在频域上具有明显的共振峰值。但通过傅里叶变换将信号从时域变换到频域时将丢失时序信息，在频域上不能区分2个时刻的振荡差异，因此，更适合于瞬时频率随时间变化不明显的平稳信号。当系统完全进入热声振荡状态时，振荡频率几乎不变，可使用傅里叶变换。但要深入研究系统从稳定状态演化到不稳定状态的非

收稿日期：2023-11-30；**录用日期**：2024-02-23；**网络出版时间**：2024-04-02 16:12
网络出版地址：link.cnki.net/urlid/11.2625.V.20240401.1203.001
基金项目：国家自然科学基金(52106128)；航空发动机及燃气轮机基础科学中心项目(P2022-A-II-006-001)；中央高校基本科研业务费专项资金

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

引用格式：*覃子宇，韩猛，韩啸，等. 基于小波变换的压力脉动信号分析方法[J]. 北京航空航天大学学报，2026，52(3)：801-808.*
QIN Z Y，HAN M，HAN X，et al. Method for analyzing pressure fluctuation signals based on wavelet transform[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：801-808 (in Chinese).

平稳过程，以便进行早期预报时，传统的傅里叶变换不再合适。为解决这一问题，可采用短时傅里叶变换（short time fourier transform, STFT），每次取一小段信号进行短时傅里叶变换[10]以同时获得时间频率分布。可以证明在时域上区分不同时刻和频域上区分不同频率这2个目的是矛盾的，不能通过某种时频变换同时具有较高的时间分辨率和频率分辨率。在实际问题中，高频分量更关注时间分辨率，低频分量更关注频率分辨率。短时傅里叶变换不能随着频率变化而自适应调整时频分辨能力，为此提出小波变换[11]、S变换[12]等方法，这些方法在低频时具有更高的频率分辨率，在高频时具有更高的时间分辨率，更好地缓解了时频分辨率的矛盾。对于小波变换，仍有时频分布不够聚焦，进而出现误读等问题。Huang等[13]提出基于经验模态分解的希尔伯特-黄变换，Daubechies等[14]提出同步挤压小波变换，可以使时频分布更为集中。同步挤压技术还可用于同步挤压短时傅里叶变换[15]，同步挤压S变换[16]等。目前，小波分析及其衍生方法已广泛用于地震波研究[17-18]、宏观经济数据分析[19-20]、生理信号分析[21]、故障诊断[22]等各种领域中[23]。

本文将小波变换方法用于分析燃烧振荡中的压力脉动信号。首先，简要介绍离散小波变换和连续小波变换；然后，以高温高压实验中采集的压力脉动序列为研究对象展开讨论，运用小波分析在处理非平稳信号上的优势，深入探讨压力信号的动态演化特征，并为基于压力信号的早期预报打下基础。

1 小波变换简介

1.1 离散小波变换

小波分析的研究对象是平方可积函数，其物理含义为在时域上的能量有限。正交多分辨分析将高分辨率的尺度空间 V_{j+1} 分解为低分辨率的尺度空间 V_j 和反映降低分辨率时所丢失信息的小波空间 W_j，并递归地对尺度空间进行分解。最高分辨率的尺度空间对应为全空间，最低分辨率的尺度空间对应为仅含零元素的子空间。于是平方可积函数空间可表示为一系列小波空间的直和。对于不同分辨率下的尺度空间和小波空间，存在相应的尺度函数 $\phi(t)$ 和小波函数 $\psi(t)$，使得 $\{\phi(2^j t-k)\}_{k\in\mathbf{Z}}$、$\{\psi(2^j t-k)\}_{k\in\mathbf{Z}}$ 分别构成 V_j、W_j 的正交基。进而 $\psi(t)$ 的二进伸缩和整数平移所生成的函数组构成全空间的一组标准正交基。因此，小波分析的主要任务是构造合适的小波函数及由小波表示分析信号的性质。

信号在尺度空间中的分量反映低频平均信息，在小波空间中的分量反映高频细节信息。并且信号在不同分辨率子空间下的投影分量又表现出不同的特性。通过小波基系数，能够研究信号在不同层次上的性质。与由正交三角函数基得到的傅里叶分析类似，正交小波基系数可表示为

$$d_{j,k}=2^{j/2}\int_{-\infty}^{+\infty}f(t)\psi^*(2^j t-k)\mathrm{d}t \tag{1}$$

式中：$f(t)$ 为任意时间信号；$d_{j,k}$ 为小波基系数；j 为尺度伸缩量；k 为平移量。

Mallat[24]发现连续的尺度函数和小波函数分别对应于离散的低通滤波器和高通滤波器。离散小波变换（discrete wavelet transform, DWT）可通过信号与滤波器的卷积下采样运算进行，而无需知道小波函数的具体表达式，这为快速计算离散小波变换提供了一种方法。应用Mallat算法可用如下公式计算小波基系数：

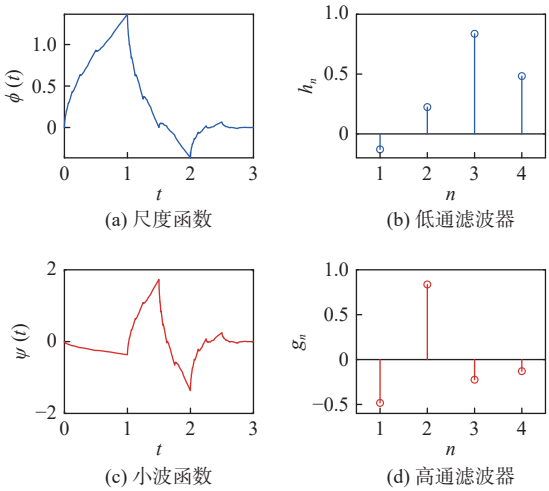
$$\begin{cases}c_{j,k}=\sum\limits_{n=-\infty}^{+\infty}c_{j+1,n}h^*_{n-2k}\\ d_{j,k}=\sum\limits_{n=-\infty}^{+\infty}c_{j+1,n}g^*_{n-2k}\end{cases} \tag{2}$$

式中：$c_{j,k}$ 和 $d_{j,k}$ 分别为不同分辨率尺度空间 V_j 和小波空间 W_j 中基函数的系数；$\{h_n\}$ 和 $\{g_n\}$ 为低通滤波器和高通滤波器；下标 n 为第 n 个元素。当分辨率 J 充分大时，信号 $f(t)$ 在尺度空间投影系数近似为信号的等距均匀采样。

$$c_{J,k}=2^{-J/2}f\left(k/2^J\right)\qquad J\gg 1 \tag{3}$$

式(2)和式(3)给出了从分辨率 J 开始降低分辨率逐层计算分解系数的过程。下一层的尺度基函数系数和小波基函数系数由上一层的尺度基函数系数与滤波器进行卷积，并且以2为步长下采样得到。每进行一次分解，系数的长度减半。对于长度为 2^N 的信号，最多可递归进行 N 次分解。由于实际信号的长度有限，卷积运算时还需在边界处进行合适的延拓，常用的延拓方式有对称延拓、周期延拓等。此外，仅进行卷积操作的离散小波变换称为最大重叠离散小波变换（maximal overlap discrete wavelet transform, MODWT）。MODWT可以使分解前后系数长度保持不变。

在构造尺度函数、小波函数及相应的滤波器方面，Cohen等[25]进行了创造性的工作。其思想是使小波函数具有 p 阶消失矩。由此得到（Daubechies-m）系列小波，m 为消失矩阶数。图1为具有二阶消失矩的db2小波所对应的尺度函数 $\phi(t)$ 及低通滤波器 h_n，小波函数 $\psi(t)$ 及高通滤波器 g_n。尺度函数和小波函数在区间[0,3]之外取值为0，滤波器组的长度为4。



(a) 尺度函数　(b) 低通滤波器

(c) 小波函数　(d) 高通滤波器

图 1　db2 小波对应的基函数和滤波器

Fig. 1　Basis function and filter corresponding to db2 wavelet

滤波器组表达式为

$$\begin{cases} \boldsymbol{h} = \left[\dfrac{1-\sqrt{3}}{4\sqrt{2}}, \dfrac{3-\sqrt{3}}{4\sqrt{2}}, \dfrac{3+\sqrt{3}}{4\sqrt{2}}, \dfrac{1+\sqrt{3}}{4\sqrt{2}}\right] \\ \boldsymbol{g} = \left[-\dfrac{1+\sqrt{3}}{4\sqrt{2}}, \dfrac{3+\sqrt{3}}{4\sqrt{2}}, -\dfrac{3-\sqrt{3}}{4\sqrt{2}}, \dfrac{1-\sqrt{3}}{4\sqrt{2}}\right] \end{cases} \tag{4}$$

1.2　连续小波变换

连续小波变换 (continuous wavelet transform, CWT) 的一个重要应用是非平稳信号的时频表示。时频表示的目的是得到信号在时间-频率平面上的联合分布, 在时频平面上刻画信号的动态演化特性。重点关注如下问题, 某个时间段内存在何种频率分量, 某个频率分量存在于哪个时间段内。一种朴素的想法是短时傅里叶变换, 即对时域信号乘以时域上窄带分布的窗函数后再进行傅里叶变换, 滑动窗口以获得不同时刻的频率分布:

$$T_{\mathrm{STFT}}(f,b) = \int_{-\infty}^{+\infty} f(t) w^*(t-b) \mathrm{e}^{-2\pi i f t} \mathrm{d}t \tag{5}$$

式中: $w(t)$ 为窗函数; 上标*表示复共轭; f 为频率; t 为时间平移量。

时频表示希望同时具有较高的时间分辨率和频率分辨率, 即时频局部化能力较强。但事实上, 时间局部化与频率局部化是矛盾的。若窗函数在时域上分布较窄, 则在频域上分布较宽, 不能同时使时域和频域上的分布宽度都任意小。在时频分析中, 对于低频分量应追求更高的频率分辨率, 对于高频分量应追求更高的时间分辨率。给定窗函数时, 时域宽度和频域宽度也随之确定。因此, 短时傅里叶变换是一种单分辨时频分析, 在时间分辨率与频率分辨率上只能满足一方面的要求, 不能克服时频局部化的矛盾。

相比而言, 连续小波变换可以自适应调整时频局部化特性。式 (6) 中, a 为尺度因子, b 为平移因子, 用以调整对小波函数的伸缩和平移作用。尺度因子较大时, 小波函数被拉伸, 时域宽度较大, 对应频域宽度较小, 并且中心频率较小, 用于分析低频分量; 尺度因子较小时, 小波函数被压缩, 时域宽度较小, 对应频域宽度较大, 并且中心频率较大, 用于分析高频分量。通过调节尺度因子, 可在低频时具有较高的频率分辨率, 高频时具有较高的时间分辨率。因此, 连续小波变换是一种多分辨时频分析, 能够得到更加细致的时间频率联合分布表示:

$$T_{\mathrm{CWT}} = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi^* \left(\frac{t-b}{a}\right) \mathrm{d}t \tag{6}$$

在连续小波变换的基础上, Daubechies 引入同步挤压变换 (synchrosqueezed transform, SST), 在每个时刻对瞬时频率附近的小波系数加和得到同步挤压小波系数, 这一过程类似于在频率轴上同步地逐点挤压使小波系数更为聚集。同步挤压小波变换得到的时频分布能量更集中, 时频关系更为清晰。

值得注意的是, 在数值计算中连续小波变换也要进行离散化处理。离散小波变换与连续小波变换是相互联系的。离散小波变换相邻尺度对应的频率比为 2, 而连续小波变换的尺度划分更为精细, 相邻尺度的频率比小于 2, 一般为 $2^{1/K}$。K 为正整数, 每经过 K 个尺度, 对应的频率为之前的 2 倍。

2　数据处理

2.1　数据获取

本文在如图 2 所示的实验装置上开展高温高压液雾燃烧实验, 以模拟真实航空发动机燃烧室工作特点。实验件为具有中心分级特征的单头部燃烧室。试验件下游安装有长度为 567 mm 的水冷套和临界孔板。经过声学仿真计算, 整体实验系统具有 500 Hz 的理论振荡频率, 实验过程中重点关注热声不稳定性问题。在实验中, 动态压力脉动序列由压力传感器采集, 采集频率为 20 000 Hz。

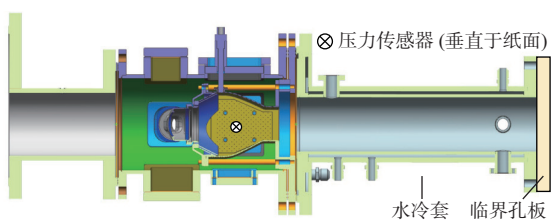


图 2　试验段示意图

Fig. 2　Schematic diagram of experiment section

实验工况条件为, 进口温度为 645 K, 进口压力为 1.14 MPa, 总空气流量为 1.57 kg/s, 预燃级燃油流量为 8.7 kg/h, 主燃级燃油流量为 123.5 kg/h。采集

得到压力脉动序列如图3所示，采集时长为2 s，共有40 000个数据点，p为压力脉动时间序列的大小。图4为压力脉动频谱图，A为压力脉动振幅。压力脉动整体上具有434 Hz的振荡频率，这与稳态振荡的理论频率500 Hz并不完全一致，表明压力序列存在较强的非平稳特征。

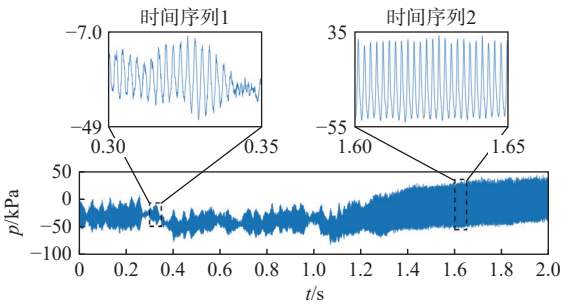


图3 压力脉动序列

Fig. 3 Pressure fluctuation series

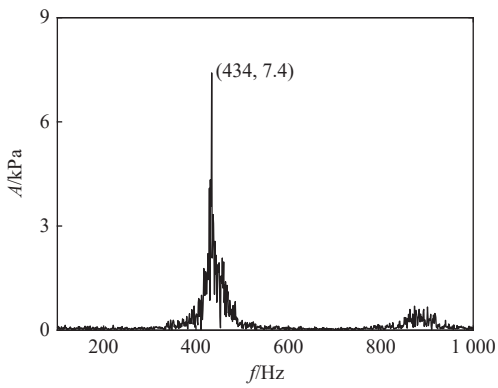


图4 压力脉动频谱

Fig. 4 Pressure fluctuation spectrogram

对压力数据整体进行傅里叶变换将丢失信号的时变特性。截取图3中2个时长为0.05 s的子序列，分别进行傅里叶变换、相空间重构和递归图分析，结果如图5所示，其中，τ为延迟时间，是平移指标。不难发现，2个序列具有截然不同的动力学特征。对于1.60～1.65 s内的时间序列2，具有预期的500 Hz的振荡频率，并且在相空间中形成了封闭的空心极限环轨道，递归图显示了黑白交错的等距条纹。这些行为都表明在这段时间中，系统发生了明显的热声不稳定性现象。而对于0.30～0.35 s内的时间序列1，峰值频率为420 Hz，相空间轨道与递归图较为不规律，表明此时压力脉动演化具有一定程度的随机混沌特性，并且热声不稳定性现象不如时间序列2显著。这2段代表性的压力序列进一步证明所采集的压力脉动数据的非平稳特性，演化过程中存在从相对稳定且带有随机性的振荡到大幅值极限环振荡这2种模式的切换。接下来需要通过对脉动信号的处理准确得到切换时间。

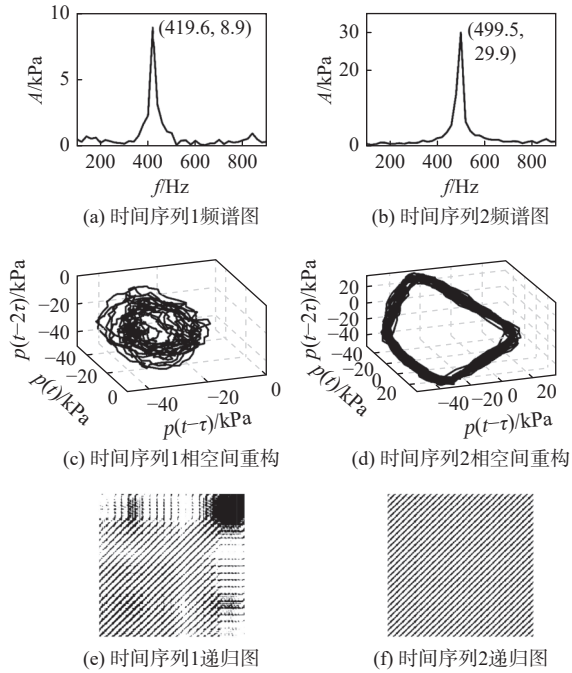


图5 压力脉动子序列动力学分析

Fig. 5 Dynamics analysis of pressure fluctuation subsequences

2.2 离散小波分析

在各种离散小波中，db2小波计算量较小，且具有二阶消失矩，有利于定位信号中的突变点。本文运用db2小波对压力脉动序列进行最大重叠离散小波变换，可得到一系列不同分辨率小波空间中的小波表示。首先，考察压力数据在离散小波变换中的总方差分解，如图6所示。本文中具有40 000个采样点的压力脉动最多可分解15层（$\log_2 40\,000 \approx 15.29$）。最大重叠离散小波分解具有方差守恒的性质，15个分辨率下的小波表示和剩余最低分辨率下尺度空间表示的方差对原始压力信号方差的累计贡献率为1。信号的最高分辨率为奈奎斯特频率10 000 Hz（采样频率的一半），于是第1层小波空间的分辨率对应频率为5 000～10 000 Hz，第2层小波空间的分辨率对应频率为2 500～5 000 Hz，之后逐

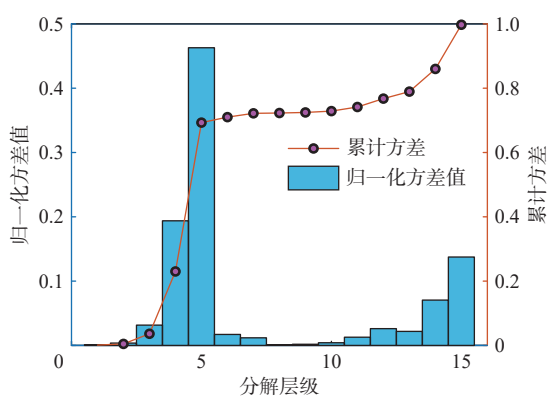


图6 压力脉动的总方差分解

Fig. 6 Total variance decomposition of pressure fluctuation

层按 1/2 递减。分解层数越高, 分辨率越低, 对应低频性质。方差是对信号脉动的度量, 由图 6 可知, 第 5 层分解的方差贡献率最高, 为 46.3%。这表明原始压力数据的脉动特性主要由位于区间 312.5～625 Hz 内的振荡频率体现。此外, 在较低分辨率下的方差贡献率也相对较大, 这对应于信号的低频噪声。

进一步考察信号在小波空间中投影分量的时间演化规律, 如图 7 所示。压力信号在第 3 层小波空间的投影分量的脉动程度具有突增的特点, 脉动方差由 0.72 增大到 6.56; 在第 6 层小波空间的投影分量的脉动程度具有突降的特点, 脉动方差由 0.85 降低到 0.17。且转变时刻在 1.25 s 左右。这一转变对应于前述压力脉动 2 种振荡模式的切换。在 1.25 s 后, 系统进入极限环热声振荡状态。小波分解能够较好地捕捉系统的热声状态从稳定燃烧演化至不稳定振荡过程中的非平稳特性。

如上所述, 对于给定的压力脉动序列, 离散小波分析可以定位热声振荡发生的时刻, 实现故障诊断。而在实际应用中, 更为关心的是通过信号的实时采集和分析, 在热声振荡演化过程的早期阶段尽早给出预警, 便于控制系统进行干预。为探讨小波分析方法的早期预报性能, 对给定大小的滑窗内的

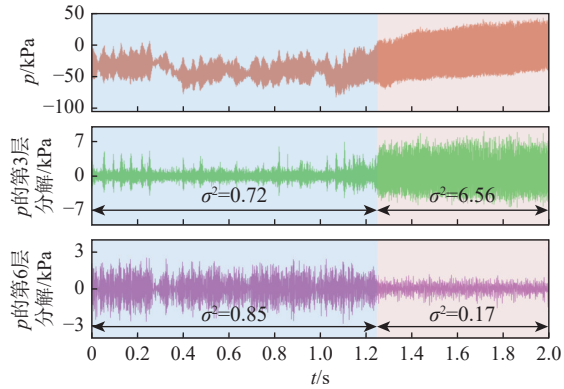
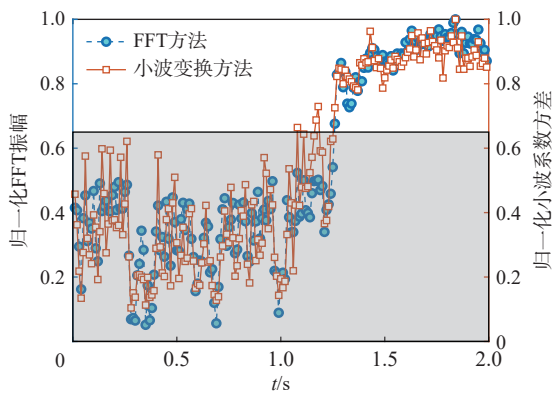


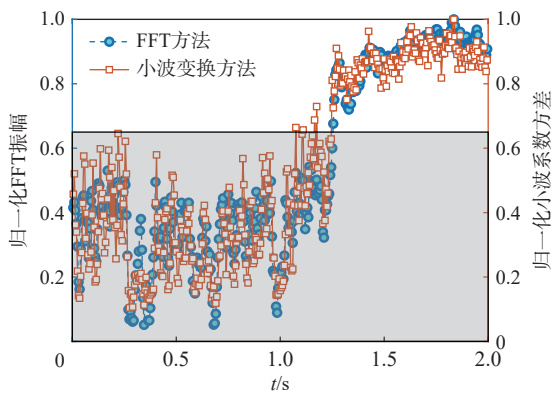
图 7　原始信号和小波分解投影的时间演化规律

Fig. 7　Time evolution law of original signal and wavelet decomposition projection

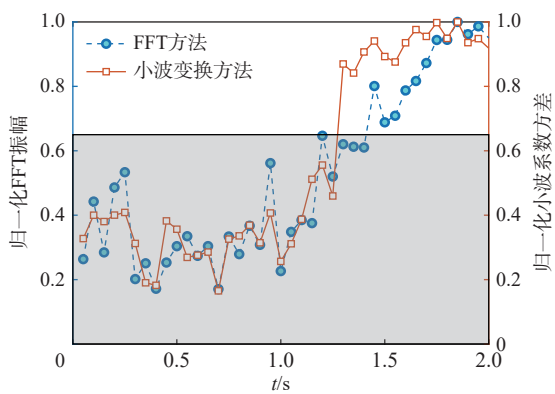
压力脉动序列进行离散小波变换并计算第 3 层小波系数方差。滑窗移动步长为窗口大小时, 相邻滑窗的重合度为 0; 滑窗移动步长为窗口大小的一半时, 相邻滑窗的重合度为 50%。作为对照, 对相同的序列进行傅里叶变换, 得到峰值频率处的振幅, 结果如图 8 所示, 图中, 灰色区为低振幅区。对于不同滑窗步长, 当滑窗大小为 10 ms 时, 小波系数方差和峰值频率振幅随时间的演化曲线在整体上发展趋势一致, 都在 1.2 s 左右处突然增大并达到饱和。而当滑窗大小为 50 ms 时, 小波方法能够更早



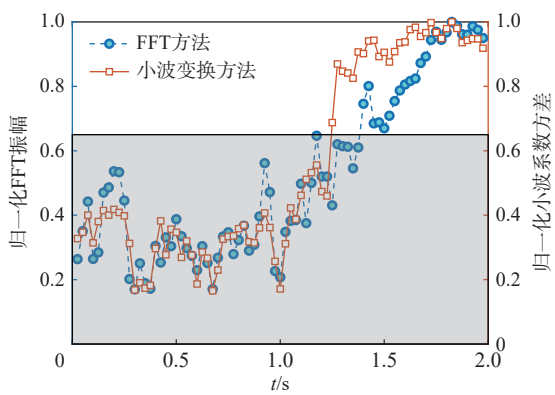
(a) 时间窗口为10 ms, 重叠率为0%



(b) 时间窗口为10 ms, 重叠率为50%



(c) 时间窗口为50 ms, 重叠率为0%



(d) 时间窗口为50 ms, 重叠率为50%

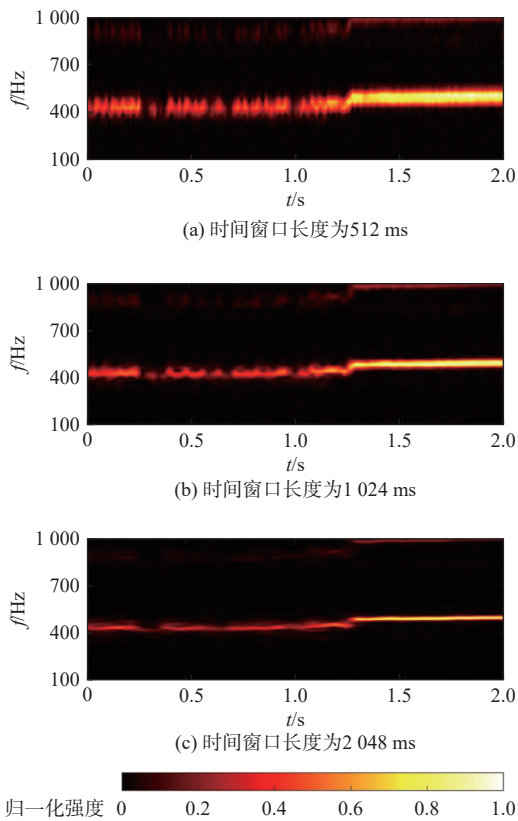
图 8　小波变换和傅里叶变换方法的早期预报性能比较

Fig. 8　Comparison of early prediction performance between wavelet transform and Fourier transform methods

达到饱和，比傅里叶变换方法提前约300 ms，判断热声振荡时延迟时间较短。且在指标数值上突变特性明显，不容易造成误判。因此，小波变换方法的早期预报性能相比傅里叶变换方法更为优异，留有更为充足的时间裕度。

2.3 连续小波分析

压力脉动信号的短时傅里叶变换时频图如图9所示。可以看到，信号具有显著的非平稳特征，也同样反映出2种振荡模式的切换。同时，随着窗函数中信号长度的增加，频率分辨率提高，在频域上局部化能力越强，在时频图上能够得到更窄的频带。但与此同时时间分辨率降低，频率分量幅值沿时间轴分布更为模糊，更难以看出时变规律。短时傅里叶变换难以同时满足时间分辨率和频率分辨率的要求。



(a) 时间窗口长度为512 ms

(b) 时间窗口长度为1 024 ms

(c) 时间窗口长度为2 048 ms

图9 短时傅里叶变换

Fig. 9 Short time Fourier transform

同步挤压小波变换能够较好地缓解时频局部化的矛盾，得到更为清晰的时频分布，如图10所示。可以看出，2种振荡模式的切换时间在1.25 s附近，这与离散小波变换结果一致。在1.25 s后，压力脉动表现为500 Hz的窄带振荡，这与所设计的声学模态频率一致，在这段时间内系统处于热声极限环振荡状态。而在1.25 s之前，同步挤压小波变换时频图提供了离散小波变换难以观察到的信息。压力脉动表现出类似于频率调制(frequency modulation, FM)的性质。在频率调制现象中，瞬时频率具有周期性变化的规律：

$$p' = A\cos(2\pi f_c t + \varphi_c + B\cos(2\pi f_m t + \varphi_m)) \tag{7}$$

式中：A和B分别为压力脉动振幅和信号调频幅度；φ_c和φ_m分别为压力脉动相位和信号调频相位；f_c和f_m分别为压力脉动中心频率和信号调频频率。由图10可知，在热声振荡发生前，压力脉动的瞬时频率以f_c=435 Hz为中心，进行调制频率f_m=25 Hz(周期为40 ms)的振荡。该现象表明燃烧室中压力脉动的声学振荡受到其他相干过程(如涡波等流体动力学过程)的影响。另一方面，频率调制现象可能是热声振荡的某种先兆效应，可进行更深入研究。

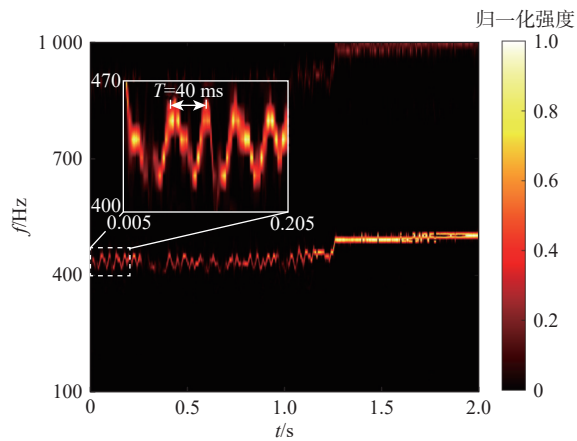


图10 同步挤压小波变换

Fig. 10 Synchro squeezed wavelet transform

对比离散小波变换和连续小波变换，可以发现，连续小波变换使用的离散尺度更多样，能够得到更为丰富细致的信息，但同时结果有一定的冗余性，所需要的计算量较大，可用于信号的详细分析研究。而在故障实时检测等应用情形下，可用离散小波变换快速定位故障。

3 结 论

1）离散小波变换和连续小波变换都表明压力信号在演化过程中存在2种振荡模式的切换。

2）振荡模式的切换时刻在1.25 s附近。在切换时刻前，表现出频率调制的特征；在切换时刻后，表现出热声振荡的特征。

3）基于离散小波变换系数方差的方法可以实现更好的早期预报性能，更加适用于实际应用。

因此，本文工作将有助于更加细致地分析实验所采集的信号的性质。

参考文献（References）

[1] LIEUWEN T C , YANG V. Combustion instabilities in gas turbine engines: operational experience, fundamental mechanisms, and

modeling[M]. Reston: American Institute of Aeronautics and Astronautics, 2005.

[2] WANG X Y, HAN X, SONG H, et al. Multi-bifurcation behaviors of stability regimes in a centrally staged swirl burner[J]. Physics of Fluids, 2021, 33(9): 095121.

[3] WANG X Y, HAN X, WANG J C, et al. Flame stabilization and thermoacoustic instability during operating condition modulations: roles of pilot and main flames[J]. Physics of Fluids, 2022, 34(12): 125102.

[4] COOLEY J W, LEWIS P A W, WELCH P D. The fast Fourier transform and its applications[J]. IEEE Transactions on Education, 2007, 12(1): 27-34.

[5] ABARBANEL H D I, BROWN R, SIDOROWICH J J, et al. The analysis of observed chaotic data in physical systems[J]. Reviews of Modern Physics, 1993, 65(4): 1331-1392.

[6] MARWAN N, ROMANO M C, THIEL M, et al. Recurrence plots for the analysis of complex systems[J]. Physics Reports, 2007, 438(5-6): 237-329.

[7] AN Q, STEINBERG A M, JELLA S, et al. Early warning signs of imminent thermoacoustic oscillations through critical slowing down[J]. Journal of Engineering for Gas Turbines and Power, 2019, 141(5): 054501.

[8] DELAAT J C, KOPASAKIS G, SAUS J R, et al. Active combustion control for a low-emissions aircraft engine combustor prototype: experimental results[J]. Journal of Propulsion and Power, 2013, 29(4): 991-1000.

[9] ZHANG L, SU X Y, ZHOU H, et al. Active control of multiple neural networks for oscillating combustion[J]. AIAA Journal, 2022, 60(6): 3821-3833.

[10] ALLEN J B, RABINER L R. A unified approach to short-time Fourier analysis and synthesis[J]. Proceedings of the IEEE, 1977, 65(11): 1558-1564.

[11] AGUIAR-CONRARIA L F, SOARES M J. The continuous wavelet transform: a primer: NIPE WP 16/2011[R]. Braga: NIPE, 2011.

[12] STOCKWELL R G, MANSINHA L, LOWE R P. Localization of the complex spectrum: the S transform[J]. IEEE Transactions on Signal Processing, 1996, 44(4): 998-1001.

[13] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.

[14] DAUBECHIES I, LU J F, WU H T. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool[J]. Applied and Computational Harmonic Analysis, 2011, 30(2): 243-261.

[15] AUGER F, FLANDRIN P, LIN Y T, et al. Time-frequency reassignment and synchrosqueezing: an overview[J]. IEEE Signal Processing Magazine, 2013, 30(6): 32-41.

[16] HUANG Z L, ZHANG J Z, ZHAO T H, et al. Synchrosqueezing S-transform and its application in seismic spectral decomposition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(2): 817-825.

[17] ZHOU Z Q, ADELI H. Time-frequency signal analysis of earthquake records using Mexican hat wavelets[J]. Computer-Aided Civil and Infrastructure Engineering, 2003, 18(5): 379-389.

[18] 张敏, 孙雷鸣, 欧阳敏, 等. 小波变换在 OBN 地震资料双检合并中的研究与应用[J]. 工程地球物理学报, 2022, 19(6): 877-885.
ZHANG M, SUN L M, OUYANG M, et al. Research and application of wavelet transform in OBN seismic data dual-sensor summation[J]. Chinese Journal of Engineering Geophysics, 2022, 19(6): 877-885(in Chinese).

[19] CROWLEY P M. A guide to wavelets for economists[J]. Journal of Economic Surveys, 2007, 21(2): 207-267.

[20] AGUIAR-CONRARIA L, MARTINS M M F, SOARES M J. The yield curve and the macro-economy across time and frequencies[J]. Journal of Economic Dynamics and Control, 2012, 36(12): 1950-1970.

[21] CUI X, BRYANT D M, REISS A L. NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation[J]. NeuroImage, 2012, 59(3): 2430-2437.

[22] 于广宇. 基于小波分析与 EMD 分解的球磨机齿轮齿面故障诊断研究[D]. 南宁: 广西大学, 2021.
YU G Y. Research on fault diagnosis of gear tooth surface of ball mill based on wavelet analysis and EMD decomposition[D]. Nanning: Guangxi University, 2021(in Chinese).

[23] RHIF M, BEN ABBES A, FARAH I R, et al. Wavelet transform application for/in non-stationary time-series analysis: a review[J]. Applied Sciences, 2019, 9(7): 1345.

[24] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.

[25] COHEN A, DAUBECHIES I, CFEAUVEAU J. Biorthogonal bases of compactly supported wavelets[J]. Communications on Pure and Applied Mathematics, 1992, 45(5): 485-560.

Method for analyzing pressure fluctuation signals based on wavelet transform

QIN Ziyu[1,2], HAN Meng[1], HAN Xiao[1,*], ZHAO Shuo[3], JIANG Yuhan[1], LIN Yuzhen[1]

(1. Research Institute of Aero-Engine, Beihang University, Beijing 100191, China;
2. Shen Yuan Honors College, Beihang University, Beijing 100191, China;
3. China Aerospace Hunan Power Machinery Research Institute, Zhuzhou 412002, China)

Abstract: Dynamic pressure fluctuation signals were acquired in the high-temperature and high-pressure experiments of aero-engine combustors. The non-stationary characteristics of signals transitioning from stable to unstable states are analyzed using the discrete wavelet transform (DWT) and continuous wavelet transform (CWT). The results indicate that the pressure fluctuation exhibits a switching between two oscillation modes. The wavelet analysis method effectively identifies the dynamic transition process of combustor pressure fluctuation. After the transition, the pressure fluctuation manifests as limit cycle thermoacoustic oscillations at acoustic modal frequencies, while before the transition, the pressure fluctuation exhibits frequency modulation characteristics, with instantaneous oscillation frequencies varying periodically over time. Additionally, the discrete wavelet transform achieves superior early warning performance compared to the Fourier transform.

Keywords: thermoacoustic instability; wavelet analysis; time frequency representation; frequency modulation; early warning

Received: 2023-11-30; **Accepted**: 2024-02-23; **Published Online**: 2024-04-02 16:12

URL: link.cnki.net/urlid/11.2625.V.20240401.1203.001

Foundation items: National Natural Science Foundation of China (52106128); Core Program of the Basic Science Center for Aero-Engine and Gas Turbine (P2022-A-II-006-001); The Fundamental Research Funds for the Central Universities

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0787

涵道式冲压发电涡轮部件级性能模型建模和参数设计方法

张华波[1]，郭迎清[1,*]，李桂财[1]，赵万里[1,2]，叶鹏[3]

(1. 西北工业大学 动力与能源学院，西安 710072；2. 中国空间技术研究院西安分院，西安 710000；
3. 中国电子科技集团 29 研究所，成都 610036)

摘　　要：随着机载附件系统电力需求的日益增长，涵道式冲压发电涡轮因其高效、低阻的特性，被视为理想的机载发电方案。针对涵道式冲压发电涡轮布局结构导致的部件耦合性强、系统建模与性能分析难度大等挑战，提出涵道式冲压涡轮部件级模型建模和参数设计方法。采用部件法进行建模，并通过粒子群优化算法实现模型参数设计，自动实现设计点性能精确匹配；同时采用 Newton Raphson 方法进行模型求解，为系统分析提供可靠依据。针对某型带有旁路和增压级的涵道式冲压涡轮系统，进行模型建立和参数设计，验证了所提方法的有效性，并初步揭示了该系统的潜在优势。非设计点仿真结果显示：此系统旁路开启时 26 kW 下工作包线面积拓展为关闭状态的 332.48%，显著增大了有效工作范围，动态阶跃干扰后响应达到 95% 的时间增大至 307.59%，系统的稳定性显著提升。

关 键 词：机载发电；涵道式冲压涡轮；部件级性能模型；参数优化；工作范围；系统稳定性
中图分类号：V231.3
文献标志码：A　　**文章编号**：1001-5965(2026)03-0809-09

随着电子和信息技术不断进步，各种功能的高功率机载电子设备纷纷涌现，而机载外挂吊舱成为这些设备在飞机平台上重要搭载方式之一，越来越受到世界各国的关注与重视[1-3]。然而，随着电子吊舱电力需求不断上升，飞机供电系统逐渐难以满足电子战设备的电力需求，有必要研发一种不依赖飞机电网、自成体系、体积小巧且易于拆卸或换装的自发电电子吊舱[2,4]。吊舱自主发电系统通常采用冲压空气涡轮驱动发电机，该项技术已在欧美国家得到工程化应用。如 EA-18G 装载的 AN/ALO-99 战术干扰吊舱利用吊舱前端的风扇驱动发电机发电，以满足吊舱的电力需求[5-6]。

当前，自发电电子吊舱的发电系统主要分为 2 种：涵道式和非涵道式。非涵道[2,4,7] 直接将冲压空气作用于涡轮，使其驱动发电机产生电能。但是，非涵道式冲压涡轮存在一系列不足之处[2,4,7,8]，如影响设备的前方视线，增大飞机的气动阻力，随着飞行速度增加阻力急剧增加、效率急剧下降，不能利用废气提供系统冷却。相反，涵道式冲压涡轮在亚音速和超音速条件下都能正常工作，是一种非常有前景的技术，具有降低阻力、提高速度和航程、增大功率提取能力、雷达信号较低等诸多优势[2]。由于非涵道式冲压涡轮整体的热力学性能仅通过涡轮本身的性能就能完整反映[9]，涵道式冲压涡轮的热力学性能需要考虑到包含进排气系统及其他所有附件的整体气动热力学平衡，因此，涵道式冲

收稿日期：2023-12-04；录用日期：2024-03-01；网络出版时间：2024-03-18 10:23
网络出版地址：link.cnki.net/urlid/11.2625.V.20240316.1458.003
基金项目：国家科技重大专项 (J2019-V-0003-0094)
* 通信作者. E-mail：yqguo@nwpu.edu.cn

引用格式：张华波，郭迎清，李桂财，等. 涵道式冲压发电涡轮部件级性能模型建模和参数设计方法 [J]. 北京航空航天大学学报，2026，52（3）：809-817. ZHANG H B，GUO Y Q，LI G C，et al. Modeling and parameter design methodology for component-level performance model of ducted ram air generation turbine[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：809-817（in Chinese）.

压涡轮全系统的热力学建模不仅困难得多，也更为必要。

目前国内外对冲压发电涡轮的研究主要集中在涡轮设计方法方面。王建平等[10]采用商用软件 Fluent 数值仿真证明了涵道式涡轮具有更高的效率并进行了实验验证。郭涛[11]基于商用软件 Fluent 研究了不同叶尖间隙下涵道式冲压涡轮性能。傅达旺[8]利用 NUMECA 软件研究了叶尖间隙等对流动损失的影响，给出了合理的间隙值。叶国祥等[4,12]基于 Kriging 模型进行了冲压涡轮叶片几何外形的气动优化。Haid 等[2]研发了一种新型涵道式冲压涡轮基于 ANSYS CFX 仿真和风洞试验得到其直径为 0.635 m，可输出 110 kW 的电力。此外，针对吊舱的环控制冷系统开展了一些相关的研究工作[13]。其中，王超等[14]介绍了双涡轮并行制冷吊舱环控系统，并建立涡轮和换热器模型，仿真和实验检验了系统性能。绳春晨等[15-17]通过 ANSYS CFX 仿真得到涡轮通用特性曲线，建立了同轴三涡轮制冷系统的耦合模型，其与实验值误差不超过 15%。Luo 等[18]提出 3 种制冷发电一体化方案，建立了涡轮和换热器仿真模型，获得了其性能差异。

以上研究主要集中在涡轮设计或部分耦合分析上，缺乏冲压发电涡轮系统的研究分析，限制了总体性能设计和控制策略的研究。本文提出涵道式冲压发电涡轮的部件级性能模型建模和参数设计方法，实现模型参数设计和建模仿真的完整路径。本文针对某型包含增压级、换热器和旁路(可开、关)的涵道式冲压发电涡轮，建立了从大气、进气道、核心部件到排气道的全系统部件级性能模型，进行参数优化设计和性能分析，验证了本文方法的可行性。

1 冲压发电涡轮的部件级建模方法及系统分析

1.1 部件级建模方法

在选定的冲压发电涡轮结构下，通用建模方法流程如图 1 所示，主要包含 3 大部分：①确定各部件的工作原理和计算方法，并分析部件协同的共同工作条件，这是建立模型的基础；②在部件计算方法和共同工作的基础上，基于实际对象或设计性能指标通过优化方法进行模型参数选取；③在以上 2 步的基础上，建立起稳态模型，通过加入动态因素建立起动态模型，文中考虑了转子动态。

其中，参数设计和模型建立是 2 条关键主线，2 部分涉及到的核心问题是相似的，即以共同工作条件为基础的非线性方程求解，但是两者具有一定

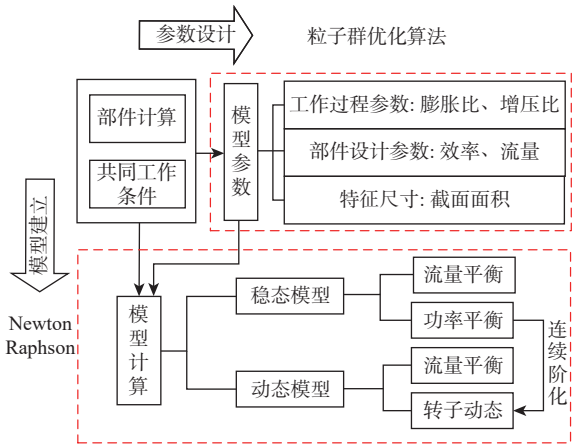


图 1 涵道式冲压涡轮部件级性能模型建立通用方法流程

Fig. 1 General method flow for building component level performance model of ducted ram air turbines

的差异。其中，模型参数设计部分通常在总体设计阶段进行，此时往往需要通过设计性能参数选择部件参数及设计性能参数等。由于涉及变量复杂繁多、可调节范围宽大、系统约束关系相互耦合，采用手工或局部优化算法进行参数选择都十分困难，因此，应该采用全局优化算法如粒子群优化算法[19]来进行参数设计。在模型计算部分通常用于系统分析、控制设计及作为实时验证系统，在这些场合中，通常要求较好的计算精度和速度，因此，适合采用局部收敛的迭代方法如 Newton Raphson 方法[20-21]，具有可控的计算精度和更有限的计算时间。

1.2 冲压发电涡轮系统分析

为兼具大功率发电和高效冷却性能，采用涡轮-发电机-增压级串联形式(turbine-generator-compressor type, TGC)的涵道式冲压发电涡轮系统(后文简称为“冲压发电涡轮系统”)结构，设定工作范围：高度为 0.1～15 km，马赫数为 0.4*Ma* ～0.95*Ma*。其原理如图 2 所示，图中，0～6 及 6'表示特征截面编号，①～⑥表示部件编号，系统包含膨胀涡轮②、发电机③、增压级④、进气道①、排气道⑤及旁路⑥等。此结构在保证散热的同时可以抑制由于发电机换热器在涡轮后的涵道内产生的总压损失对排气道的流通能力的限制，有效地拓展

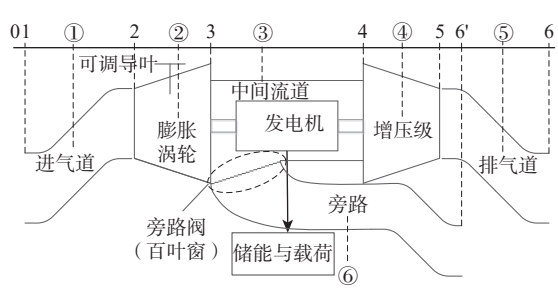


图 2 冲压发电涡轮系统结构原理

Fig. 2 Ram air generation turbine system structural principle

系统的工作范围。

冲压发电涡轮系统，利用冲压空气驱动涡轮旋转输出功率，带动发电机和压气机，通过可调导叶进行调节。其热力循环过程示意图如图 3 所示，图中，比体积为密度的倒数，0 截面自由气流经过大气和进气道冲压后压力增加，增压的气流经涡轮膨胀释放能量，产生轴功率；在主流道内，涡轮出口气流经中间流道进入增压级中提升压力后由排气道排出，最后，在大气中进行等压放热过程。在旁路开启状态下，涡轮后 3 部分气流经过旁路排气道（3～6'）排出。

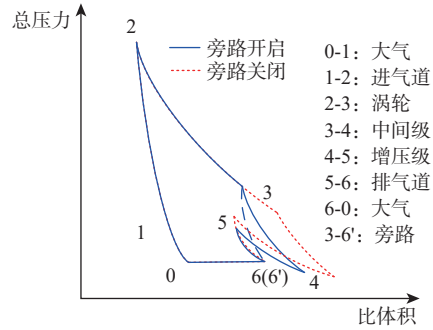


图 3　热力循环过程

Fig. 3　Thermodynamic cycle process

在此热力循环过程中：增压级可以弥补中间流道损失，削弱了涡轮膨胀比的限制，增强涡轮做功能力，同时能防止排气道拥塞；旁路开启时流经增压级的流量减少，使增压级功耗降低，并具有增大其稳定性的作用。

在稳态条件下冲压发电涡轮系统共同工作总体框架如图 4 所示，系统输入为负载和大气条件，通过 4 个迭代参数，实现流量、功率平衡。

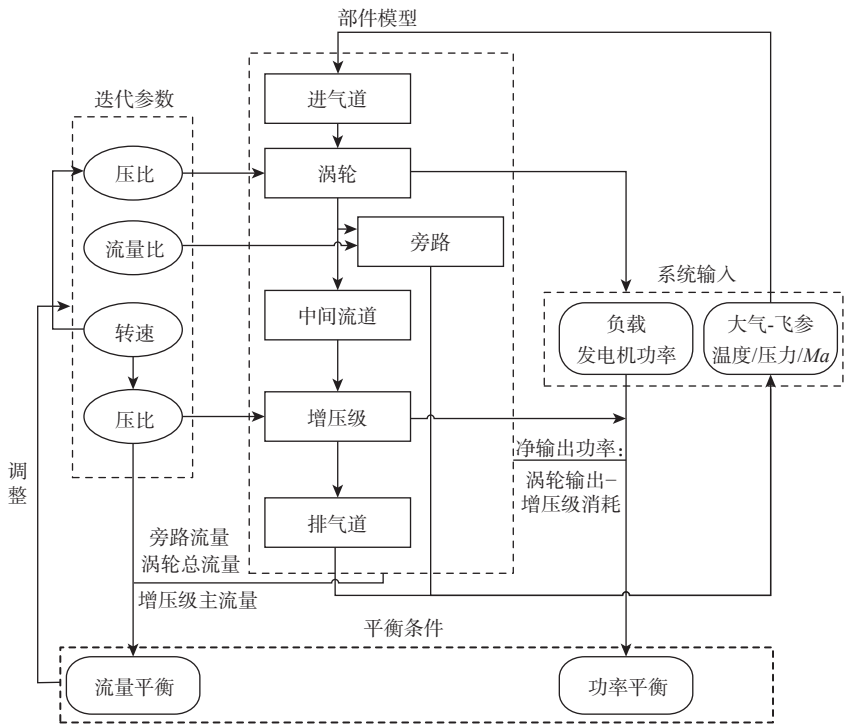


图 4　稳态冲压发电涡轮模型原理

Fig. 4　Steady-state ram air generation turbine model principle

4 个迭代参数包含转子转速 n、涡轮压力比函数 β_{T}、增压级压力比函数 β_{C}、旁路流量比 τ，令 $\boldsymbol{x}=\left[n,\beta_{\mathrm{T}},\beta_{\mathrm{C}},\tau\right]$，根据共同工作条件，可建立如下共同工作方程：

$$W_{\mathrm{b1}}-W_{\mathrm{b2}}=Z_1(\boldsymbol{x},\boldsymbol{\theta})=0 \tag{1}$$

$$W_{\mathrm{m}}-W_{\mathrm{C}}=Z_2(\boldsymbol{x},\boldsymbol{\theta})=0 \tag{2}$$

$$W_{\mathrm{m}}-W_{6}=Z_3(\boldsymbol{x},\boldsymbol{\theta})=0 \tag{3}$$

$$P_{\mathrm{T}}-P_{\mathrm{C}}-P_{\mathrm{e}}=Z_4(\boldsymbol{x},\boldsymbol{\theta})=0 \tag{4}$$

式中：W_{b1}、W_{b2} 为旁路进、出口流量；W_{m} 为主流道流量；W_{C} 为增压级流量；W_6 为排气道流量；P_{T}、P_{C}、P_{e} 为涡轮、增压级、发电机轴功率，$\boldsymbol{\theta}$ 为模型的参数向量。稳态条件下共同工作方程组 $\boldsymbol{F}_1(\boldsymbol{x})=[Z_1(\boldsymbol{x},\boldsymbol{\theta}),Z_2(\boldsymbol{x},\boldsymbol{\theta}),Z_3(\boldsymbol{x},\boldsymbol{\theta}),Z_4(\boldsymbol{x},\boldsymbol{\theta})]^{\mathrm{T}}=\boldsymbol{0}$。

动态条件下，由于转子惯性的存在，导致冲压发电涡轮系统涡轮与增压级和发电机的转子功率不再平衡，因此，动态条件下共同工作方程组 $\boldsymbol{F}_2(\boldsymbol{x})=[Z_1(\boldsymbol{x},\boldsymbol{\theta}),Z_2(\boldsymbol{x},\boldsymbol{\theta}),Z_3(\boldsymbol{x},\boldsymbol{\theta})]^{\mathrm{T}}=\boldsymbol{0}$，此时动态转速按照如下动力学微分方程确定的规律变化：

$$P_{\mathrm{T}}-P_{\mathrm{C}}-P_{\mathrm{e}}=I\left(\frac{\pi}{30}\right)^2 n\frac{\mathrm{d}n}{\mathrm{d}t} \tag{5}$$

式中：I 为转子转动惯量。

1.3 部件计算方案

采用国际标准大气进行大气参数计算。涡轮和增压级基于通用特性曲线建模，在设计点处利用耦合系数缩放进行匹配。根据导叶开度距离加权不同导叶开度下的涡轮特性，得到某一开度下涡轮性能参数压比 π、换算流量 W_{cor}、效率 η：

$$\begin{cases} \pi_{\mathrm{T},\alpha}-1=\dfrac{\sum\limits_{i=1}^{N}\left(\left(\pi_{\mathrm{T},\alpha,i}-1\right)\sum\limits_{j=1,j\neq i}^{N}\left|\alpha-\alpha_j\right|\right)}{\sum\limits_{i=1}^{N}\left|\alpha-\alpha_i\right|} \\ W_{\mathrm{cor,T},\alpha}=\dfrac{\sum\limits_{i=1}^{N}\left(W_{\mathrm{cor,T},\alpha,i}\sum\limits_{j=1,j\neq i}^{N}\left|\alpha-\alpha_j\right|\right)}{\sum\limits_{i=1}^{N}\left|\alpha-\alpha_i\right|} \\ \eta_{\mathrm{T},\alpha}=\dfrac{\sum\limits_{i=1}^{N}\left(\eta_{\mathrm{T},\alpha,i}\sum\limits_{j=1,j\neq i}^{N}\left|\alpha-\alpha_j\right|\right)}{\sum\limits_{i=1}^{N}\left|\alpha-\alpha_i\right|} \end{cases} \tag{6}$$

式中：i 表示对应导叶开度为 α_i 的特性图序号。中间流道、旁路和排气道的总压损失特性表示为

$$\Delta p^{*}=\Delta p_0^{*}(W/W_0)^{q} \tag{7}$$

其中：Δp^{*} 为流过通道的压力损失；W 为流经通道的流量；W_0 为基准流量；q 为损失指数；Δp_0^{*} 为基准状态的压力损失。

2 冲压发电涡轮模型建立

2.1 模型参数设计

按照设计性能要求和共同工作关系进行参数优化，调整模型参数向量 $\boldsymbol{\theta}$ 及设计点迭代参数 $\boldsymbol{x}$，使其在各部件协同工作的条件下并满足设计点性能要求。冲压发电涡轮采用双设计点设计，设计点参数如表 1 所示。

表 1 设计点参数

Table 1 Design point parameters

设计点	旁路状态	导叶开度α/%	转速n/(r·min^{-1})	高度 H/km	马赫数 Ma	标准温差 dT/K	功率 P_e/kW
设计点1	关闭	56.67	18 000	0.5	0.65	23.00	26
设计点2	开启	100.00	18 000	10.0	0.70	−0.15	26

在设计点条件下，确定模型参数和设计点迭代参数，记作 $\boldsymbol{\Theta}=[\boldsymbol{\theta},\boldsymbol{x}]$，需求解以下方程组：

$$\boldsymbol{g}(\boldsymbol{\Theta}=[\boldsymbol{\theta},\boldsymbol{x}])=\begin{bmatrix} g_1(\boldsymbol{\Theta}) \\ g_2(\boldsymbol{\Theta}) \\ g_3(\boldsymbol{\Theta}) \\ g_4(\boldsymbol{\Theta}) \end{bmatrix}=\begin{bmatrix} Z_1(\boldsymbol{x},\boldsymbol{\theta}) \\ Z_2(\boldsymbol{x},\boldsymbol{\theta}) \\ Z_3(\boldsymbol{x},\boldsymbol{\theta}) \\ Z_4(\boldsymbol{x},\boldsymbol{\theta}) \end{bmatrix}=\boldsymbol{0} \tag{8}$$

直接求解该高维方程是十分困难的，因此，将该问题转化为如下优化问题以寻求方程组的损失最小化间接得到方程的解。

$$\begin{cases} \min\limits_{\boldsymbol{\Theta}\in\mathbf{R}^n} & \varphi(\boldsymbol{\Theta})=\omega_1|g_1(\boldsymbol{\Theta})|+\omega_2|g_2(\boldsymbol{\Theta})|+ \\ & \qquad \omega_3|g_3(\boldsymbol{\Theta})|+\omega_4|g_4(\boldsymbol{\Theta})| \\ \text{s.t.} & \boldsymbol{\Theta}\leqslant\boldsymbol{\Theta}_{\max} \\ & \boldsymbol{\Theta}\geqslant\boldsymbol{\Theta}_{\min} \end{cases} \tag{9}$$

式中：归一化权系数 $\omega_1=1/W_{\mathrm{b1}}$，$\omega_2=\omega_3=1/W_{\mathrm{m}}$，$\omega_4=1/P_{\mathrm{e}}$。采用粒子群优化算法进行计算参数优化设计，对于此问题，其实施方式如下：

$$\begin{cases} \boldsymbol{\Theta}_{k+1}^{i}=\boldsymbol{\Theta}_{k}^{i}+v_{k+1}^{i} \\ v_{k+1}^{i}=wv_{k}^{i}+c_1r_1\left(p^{i}-\boldsymbol{\Theta}_{k}^{i}\right)+c_2r_2\left(G-\boldsymbol{\Theta}_{k}^{i}\right) \end{cases} \tag{10}$$

式中：$\boldsymbol{\Theta}_k^i$ 和 v_k^i 分别为第 k 步迭代时粒子 i 的位置和速度；r_1、r_2 为 [0,1] 内的随机因子；w、c_1 和 c_2 分别为惯性权重、认知学习因子和社会学习因子；p^i 表示粒子 i 的最优点；G 表示全局最优点。惯性权重 $w\in[0.1,1.1]$，初始为 1.1，根据历史结果动态调整，学习因子 c_1=c_2=1.49，群大小为 2 000。

参数优化流程如图 5 所示，依次在 2 个设计点进行参数设计。在设计点 2 处，优化模型参数向量 $\boldsymbol{\theta}_1$=[$C_{\pi,\mathrm{T},\alpha_2}$，$C_{W,\mathrm{T},\alpha_2}$，$C_{\eta,\mathrm{T},\alpha_2}$，$C_{\pi,\mathrm{C}}$，$C_{W,\mathrm{C}}$，$C_{\eta,\mathrm{C}}$，$\beta_{\mathrm{T},2}$，$\beta_{\mathrm{C},2}$，$\tau$，$A_z$]，其中，$C$ 分别表示涡轮和增压级压比π、流量 W、效率η的耦合系数，A_z 为排气道出口面积，将设计点 2 得到的参数结果（$C_{\pi,\mathrm{C}}$，$C_{W,\mathrm{C}}$，$C_{\eta,\mathrm{C}}$，A_z）应用于设计点 1 继续优化剩余参数 $\boldsymbol{\theta}_2$=[$C_{\pi,\mathrm{T},\alpha_1}$，$C_{W,\mathrm{T},\alpha_1}$，$C_{\eta,\mathrm{T},\alpha_1}$，$\beta_{\mathrm{T},1}$，$\beta_{\mathrm{C},1}$]。

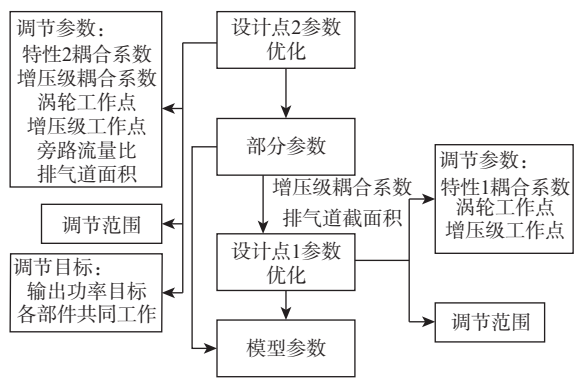


图 5 部件参数优化流程

Fig. 5 Component parameters optimization process

2.2 模型求解器

在非设计点计算时仍然可以通过粒子群优化算法进行，但是该算法收敛速度低，计算时间长，收敛次数难以确定，无法进行实时计算。因此，共同工作方程组 $\boldsymbol{F}(\boldsymbol{x})$（稳态 $\boldsymbol{F}_1(\boldsymbol{x})$ 或动态 $\boldsymbol{F}_2(\boldsymbol{x})$）通过 Newton Raphson 方法进行迭代求解，该方法在平衡点附近可以快速收敛性。Newton Raphson 方法通过迭代使方程残差向量 $\boldsymbol{F}(\boldsymbol{x})$ 的无穷范数在容许范围 $e_{\max}$ 内，具体迭代过程为

$$\begin{cases} \boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\mathrm{sign}(\Delta\boldsymbol{x})\cdot\min\{|\Delta\boldsymbol{x}|,L_{\max}\} \\ \Delta\boldsymbol{x}=-\boldsymbol{J}^{-1}(\boldsymbol{x}_k)\boldsymbol{F}(\boldsymbol{x}_k) \end{cases} \tag{11}$$

式中：$\boldsymbol{x}_k$表示 k 步迭代的$\boldsymbol{x}$值，给定初始猜测 $\boldsymbol{x}_0$；sign 为符号函数；$L_{\max}$为最大步长；$\boldsymbol{J}(\boldsymbol{x})$为雅可比矩阵通过对$\boldsymbol{x}$的微小摄动δ$\boldsymbol{x}$求解；Δ$\boldsymbol{x}$为更新步长。迭代至满足精度要求或最大迭代次数。计算中设置摄动大小δ$\boldsymbol{x}$为 10^{-7}，初始值$\boldsymbol{x}_0$为设计点参数，参数 β_T、β_C、τ的范围均为0～1，最大步长$L_{\max}$为 0.1，迭代误差限$e_{\max}$为 10^{-6}。

2.3　冲压发电涡轮模型建立

涵道式冲压发电的数学模型包含稳态模型和动态模型，其框架如图 6 所示。

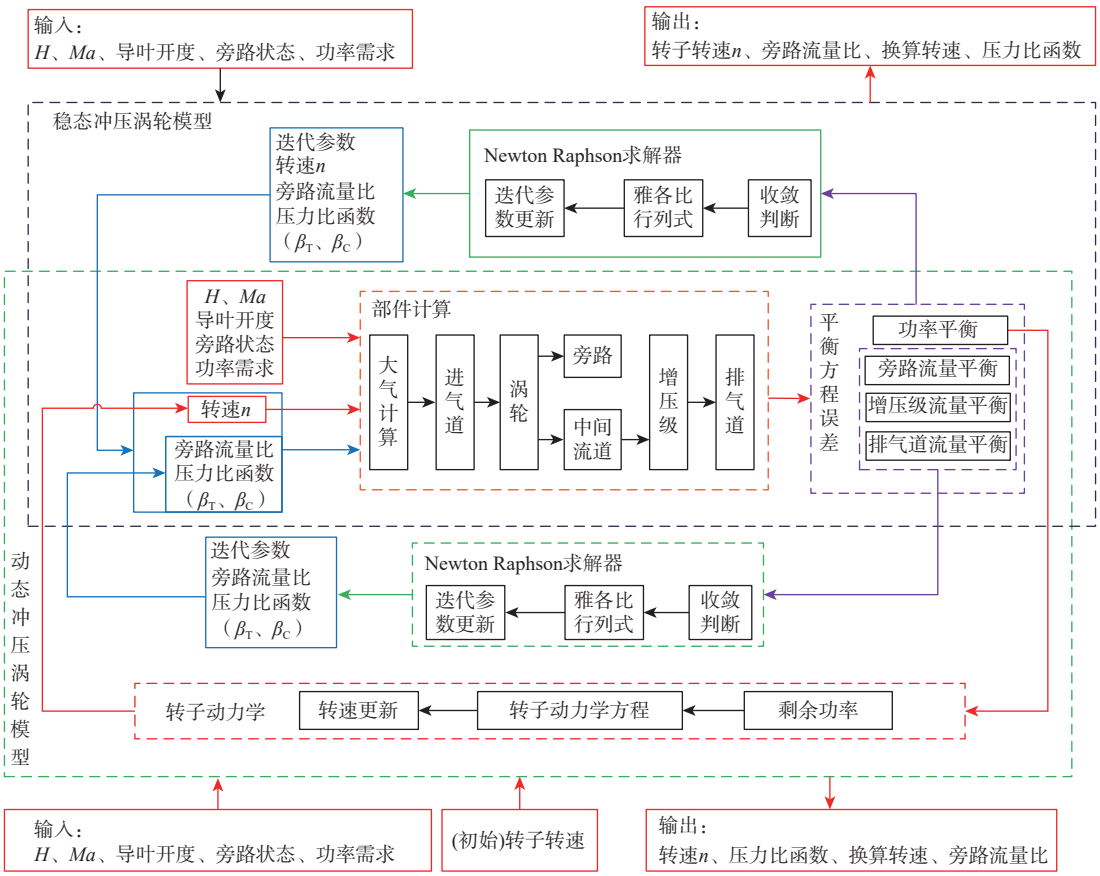


图 6　涵道式冲压涡轮数学模型框架

Fig. 6　Mathematical model framework for ducted ram air generation turbines

稳态模型和动态模型采用相同的部件计算层。在稳态模型中，模型的输入为 H、Ma、导叶开度、旁路状态及发电机功率需求。输出为部件工作状态参数，包括转子转速等迭代参数。在稳态模型迭代层内，部件计算部分根据模型输入及迭代参数进行计算，通过部件计算结果进行流量平衡和功率平衡方程误差计算，利用 Newton Raphson 求解器实现迭代参数更新，驱使系统达到平衡。

在动态模型中，模型的输入还需增加转速初始值。在动态模型迭代层内，部件计算部分根据模型输入、当前时刻转速及迭代参数进行计算，在平衡方程误差计算中不再考虑功率平衡，利用 Newton Raphson 求解器实现迭代参数更新，直到系统达到流量平衡。在迭代层外，由功率偏差驱动转速变化，随着时间变化，系统达到新的平衡或失稳。

3　冲压发电涡轮仿真分析

3.1　冲压涡轮设计点仿真

本节进行涡轮设计点仿真，检验模型设计点性能匹配程度。设计点在未经过耦合系数缩放的涡轮、增压级特性图上位置如图 7 和图 8 所示，图中，等转速线上的数值为换算转速 n_{cor}，等高线上的数值为效率。

由图 7 和图 8 可知，2 个设计点均处于特性图中相对中间位置，保证了冲压涡轮具有较大的工作范围。

设计点计算关键参数结果如表 2 所示，采用基于粒子群优化算法进行模型参数优化设计后，该模型实现了设计点性能的精确匹配，18 000 r/min 条件下得到输出功率 26 kW，当设计性能存在性能冲突时，本文方法能够得到尽可能小的性能偏差。

3.2　冲压涡轮非设计点仿真

3.2.1　稳态仿真

利用稳态模型进行非设计点仿真以了解冲压涡轮的稳态特性。由于发电机的工作状态由负载和转速直接决定，因此，为保持发电机稳定工作，需要维持稳定的转速，按照额定工作转速 18 000 r/min

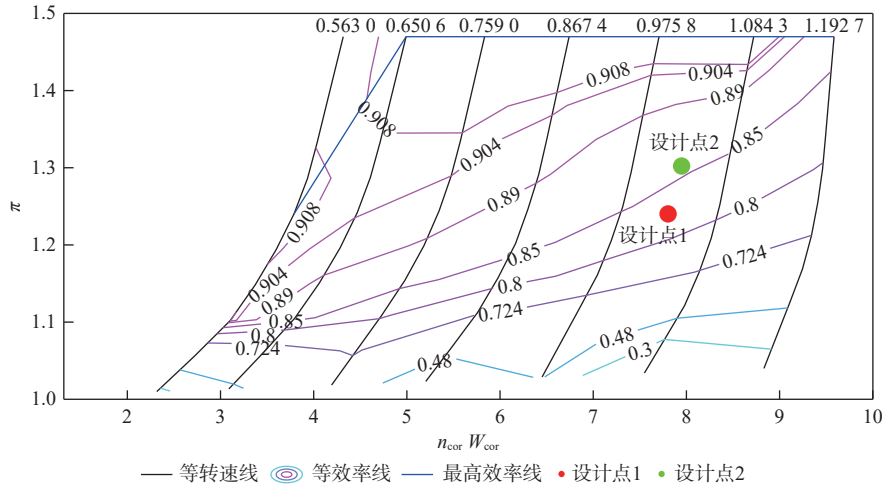


图 7 涡轮设计点 2 特性中设计点位置

Fig. 7 Position of design point in turbine design point 2 characteristics

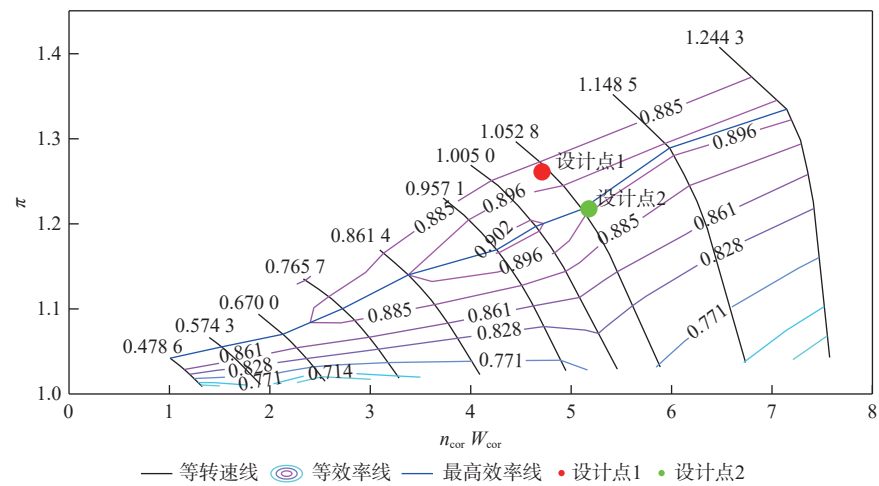


图 8 增压级特性中设计点位置

Fig. 8 Design point position in booster stage characteristics

表 2 设计点参数计算结果

Table 2 Design point parameter calculation results

设计点	旁路比τ	涡轮功P_T/ kW	增压级功P_c/ kW	输出功率P_e/ kW	转速n/(r·min^{-1})
设计点1	0	47.21	21.21	26	18 000
设计点2	0.355 6	33.70	7.70	26	18 000

进行稳态仿真，分别变化额定功率和飞行条件，通过改变导叶开度达到额定转速要求。

1）额定功率变化

飞行条件为设计状态，仿真改变功率需求，通过调整导叶开度使冲压涡轮达到额定功率需求，仿真分别基于旁路处于开启、关闭状态的 2 个设计点进行，其结果如图 9 所示。旁路开启时，改变导叶开度，得到需求功率与导叶开度的关系，可见，导叶开度与功率的关系是强非线性的，对于相同的需求功率导叶开度存在多平衡点。即，随着导叶开度由设计值 100% 的极限值逐渐减小，系统输出功率先增加后减小。旁路关闭功率需求与导叶开度的规律同样是非单调的，其中，上半条曲线，完全覆盖了冲压涡轮可达功率范围。

2）飞行条件变化

仿真中通过改变飞行参数（H、Ma）改变冲压涡轮工作条件，调整导叶开度达到额定工作转速。仿真得到在此额定条件下的稳定工作包线如图 10 所示，2 种状态下的工作包线在飞行高度马赫数的二维平面内呈现条状，其中，旁路开启状态的工作范围较宽，在 H=5 km，Ma=0.7 附近，2 种状态的工作

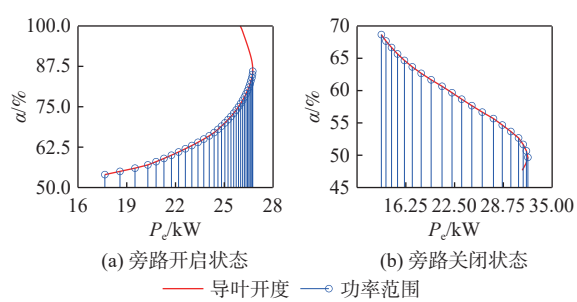


图 9 功率与导叶开度关系

Fig. 9 Relationship between power and guide vane opening

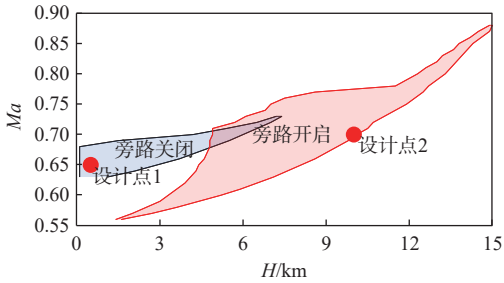


图 10　稳定工作包线

Fig. 10　Stable operating envelope

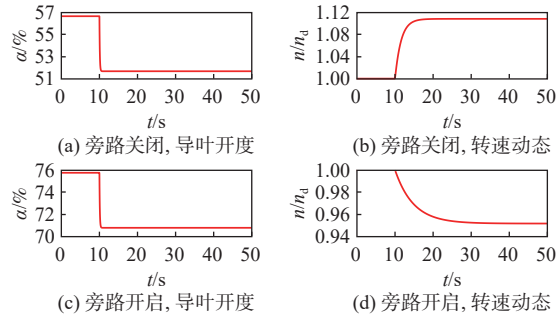


图 11　导叶开度阶跃转子转速动态

Fig. 11　Rotor speed dynamics of guide vane opening step

包线相交。在旁路关闭和开启状态下工作包线的包围面积分别为 0.254 0 km 和 0.844 5 km, 可见旁路存在可以显著拓宽相同额定状态的工作范围为关闭状态的 332.48%, 对于拓宽电子吊舱用电设备的稳定工作范围具有重要意义。

3.2.2　动态仿真

在旁路开启和关闭的 2 种状态下, 通过阶跃输入观察系统的动态特性, 以导叶开度阶跃输入情况分别对 2 个设计工作状态进行仿真, 结果如图 11 所示, 其中, n_d 为设计转速。旁路关闭和开启状态时, 导叶开度阶跃降低 5% 后转子转速分别上升和下降至新的平衡, 达到稳态差值 95% 的时间分别为 4.74 s 和 14.58 s。可见, 旁路存在使得系统动态时间显著增加至 307.59%, 反映出系统内部稳定性增强, 具有更强的抗干扰能力, 有利于稳定地输出功率。

在此动态过程中共同工作点在涡轮和压气机特性图中的运动情况分别如图 12 和图 13 所示, 图

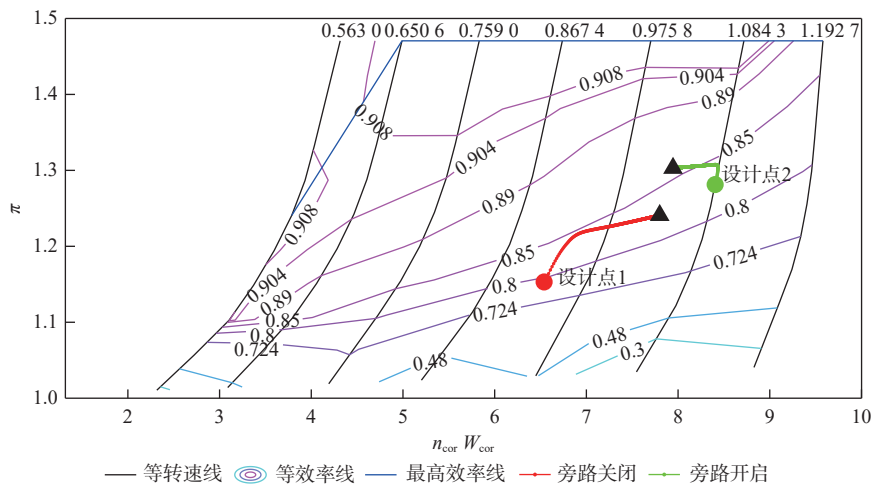


图 12　导叶开度阶跃共同工作点在涡轮特性 2 上的移动

Fig. 12　Movement of common operating point on turbine characteristic 2 when guide vane opening step

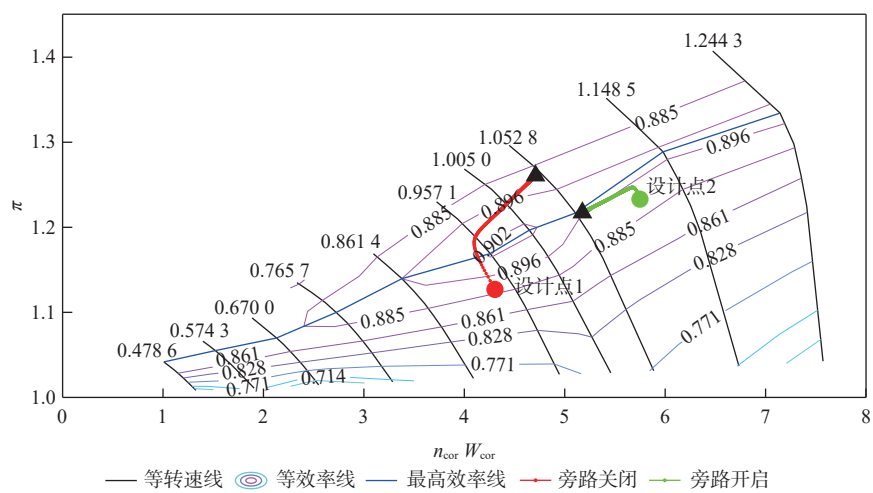


图 13　导叶开度阶跃共同工作点在增压级特性上的移动

Fig. 13　Movement of common working point on compressor characteristic when guide vane opening step

中，等转速线上的数值为换算转速 n_{cor}，等高线上的数值为效率。导叶开度的阶跃变化，导致共同工作点位置的突变，由于涡轮进口参数不变，且转速无法瞬时变化，共同工作点首先沿着等换算转速线移动，之后由于功率不平衡，转速变化，工作点跨越等转速线移动，整个动态过中涡轮处于可工作范围内。

增压级工作点在增压级特性中的移动形式与涡轮的工作点运动过程是相似的。增压级工作点首先近似沿着等转速线跳变，然后，在转速动态下，穿过等转速线移动。

4 结 论

1）本文提出涵道式冲压发电涡轮的建模和参数设计方法，通过对某型涵道式冲压涡轮建模并实现设计点性能的精确匹配，验证了该方法的可行性。

2）稳态仿真结果说明，旁路打开时比旁路关闭时显著拓宽了工作范围，26 kW 下的工作包线面积扩大为原来的 332.48%。

3）动态仿真结果显示，旁路开启时冲压发电涡轮系统稳态时间由 4.74 s 升高至 14.58 s，意味着系统抗干扰能力增强，有利于系统可靠工作。

4）从仿真结果可知，涵道式冲压发电涡轮呈现出多平衡点等非线性的系统特征，针对其的控制系统设计将是具有挑战的工作。

参考文献（References）

[1] 杨康, 刘禹, 郝汀, 等. 机载干扰吊舱总体设计技术的探讨[J]. 现代信息科技, 2020, 4(15): 37-41.
YANG K, LIU Y, HAO T, et al. Discussion on the overall design technology of airborne jamming pod[J]. Modern Information Technology, 2020, 4(15): 37-41(in Chinese).

[2] HAID D, JUSTAK J. Innovative ram air turbine for airborne power generation[C]//Proceedings of the ASME Turbo Expo 2015: Turbine Technical Conference and Exposition. New York: ASME, 2015.

[3] 杨汉华. 飞机自主发电系统的试验研究[D]. 南京: 南京航空航天大学, 2006.
YANG H H. Experimental study on aircraft autonomous power generation system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2006(in Chinese).

[4] 叶国祥, 陆启航, 张大林. 吊舱发电用涵道式冲压涡轮优化设计[J]. 航空兵器, 2016, 23(5): 61-65.
YE G X, LU Q H, ZHANG D L. Optimization design for ducted ram air turbine of pod power[J]. Aero Weaponry, 2016, 23(5): 61-65 (in Chinese).

[5] MAJUMDAR S. Growler, the 'electronic' hornet[J]. Vayu Aerospace and Defence Review, 2021(1): 126-127.

[6] ARUMUGAM A, THOMAS S, MADHUSUDAN M, et al. Utilisation of ram air turbine on a fighter platform for energy extraction failure mode study[J]. Defence Science Journal, 2020, 70(6): 583-589.

[7] 汪涛, 楚武利, 卢家玲, 等. 亚音速涵道式冲压空气涡轮性能数值仿真[J]. 计算机仿真, 2009, 26(10): 34-38.
WANG T, CHU W L, LU J L, et al. Numerical simulation of performance of subsonic ducted ram air turbine[J]. Computer Simulation, 2009, 26(10): 34-38(in Chinese).

[8] 傅达旺. 某型冲压涡轮性能计算及分析[D]. 南京: 南京航空航天大学, 2007.
FU D W. Performance calculation and analysis of a ramjet turbine[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2007(in Chinese).

[9] SAAD M M M, MOHD S B, ZULKAFLI M F. A survey on the use of ram air turbine in aircraft[J]. AIP Conference Proceedings, 2017, 1831(1): 020048.

[10] 王建平, 朱春玲. 冲压涡轮发电系统数值分析及试验研究[C]//第二届中国航空学会青年科技论坛会议论文集. 北京: 中国航空学会, 2006.
WANG J P, ZHU C L. The numerical analysis and experiment study of ram air turbine generator system[C]//Proceedings of the 2nd Chinese Society of Aeronautics and Astronautics Youth Science and Technology Forum. Beijing: Chinese society of aeronautics and astronautics, 2006(in Chinese).

[11] 郭涛. 冲压空气涡轮性能研究[D]. 南京: 南京航空航天大学, 2007.
GUO T. Study on performance of ram air turbine[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2007(in Chinese).

[12] 叶国祥. 吊舱发电用涵道式冲压涡轮优化设计[D]. 南京: 南京航空航天大学, 2016.
YE G X. Optimal design of ducted ramjet turbine for pod power generation[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2016(in Chinese).

[13] LIU S L, MA G Y, XU S X, et al. A review of reverse Brayton air cycle refrigerators[J]. International Journal of Refrigeration, 2023, 150: 200-214.

[14] 王超, 包胜, 王璐璐. TTC 双涡轮并行制冷吊舱环控系统性能研究[J]. 电子机械工程, 2017, 33(1): 48-51.
WANG C, BAO S, WANG L L. Capability analysis of double-turbo parallel refrigeration of turbo-turbo-compressor environment control system of pods[J]. Electro-Mechanical Engineering, 2017, 33(1): 48-51(in Chinese).

[15] 绳春晨, 杨榆, 谢洪涛, 等. 逆升压式吊舱涡轮冷却器变工况特性研究[J]. 真空与低温, 2020, 26(6): 466-470.
SHENG C C, YANG Y, XIE H T, et al. Performance study on variable working condition of reverse-boost pod turbine cooler[J]. Vacuum and Cryogenics, 2020, 26(6): 466-470(in Chinese).

[16] 绳春晨, 杨榆, 谢洪涛, 等. 机载吊舱环控空气制冷系统分析及试验研究[J]. 低温与超导, 2020, 48(1): 80-85.
SHENG C C, YANG Y, XIE H T, et al. Performance analysis and experimental study on air cooling system for airborne pod ECS[J]. Cryogenics & Superconductivity, 2020, 48(1): 80-85(in Chinese).

[17] YANG Y, CHEN S T, SHENG C C, et al. Study on coupling performance of turbo-cooler in aircraft environmental control system[J]. Energy, 2021, 224: 120029.

[18] LUO K, LV B. Study on performance of autonomous power generation and refrigeration system of the jamming pod[C]//Proceedings of the CSAA/IET International Conference on Aircraft Utility Systems. London: IET, 2022: 719-724.

[19] KENNEDY J, EBERHART R. Particle swarm optimization[C]// Proceedings of the ICNN'95 - International Conference on Neural Networks. Piscataway: IEEE Press, 1995: 1942-1948.

[20] GALÁNTAI A. The theory of Newton's method[J]. Journal of Computational and Applied Mathematics, 2000, 124(1): 25-44.

[21] BEN-ISRAEL A. A Newton-Raphson method for the solution of systems of equations[J]. Journal of Mathematical Analysis and Applications, 1966, 15(2): 243-252.

Modeling and parameter design methodology for component-level performance model of ducted ram air generation turbine

ZHANG Huabo[1]，GUO Yingqing[1,*]，LI Guicai[1]，ZHAO Wanli[1,2]，YE Peng[3]

(1. School of Power and Energy，Northwestern Polytechnical University，Xi'an 710072，China；
2. China Academy of Space Technology（Xi'an），Xi'an 710000，China；
3. China Electronic Technology Group Corporation No.29 Research Institute，Chengdu 610036，China)

Abstract：With the growing power demand of airborne systems, the ducted ram air turbine is considered an ideal onboard power-generation solution due to its high efficiency and low aerodynamic drag. However, its ducted configuration leads to strong component coupling, complicating system modeling and performance analysis. To address these challenges, this paper proposes a component-level modeling approach and a parameter design methodology for the ducted ram air turbine. In this method, the model is modeled by a component method, and the model parameters are designed by a particle swarm optimization algorithm to automatically match the performance parameters of design points accurately, and Newton-Raphson method was used to solve the model, which provided a reliable basis for system analysis. The effectiveness of this method is validated by the modeling and parameter design of a type of ducted ram turbine system with bypass and compressor, which also reveals the potential advantages of this system. The simulation results at non-design points demonstrate that, when the bypass is opened, the working envelope area of this ram air generation turbine expands to 332.48% of the closed state at 26 kW. This substantial increase significantly enhances the effective working range. A notable improvement in system stability is also indicated by the time it takes for the response to reach 95% following dynamic step interference, which rises to 307.59%.

Keywords：onboard power-generation; ducted ram air turbine; component-level performance model; parameter optimization; working range; system stability

Received：2023-12-04；**Accepted**：2024-03-01；**Published Online**：2024-03-18 10：23

URL：link.cnki.net/urlid/11.2625.V.20240316.1458.003

Foundation item：National Science and Technology Major Project (J2019-V-0003-0094)

* **Corresponding author**. E-mail：yqguo@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0842

一种定转速涡桨发动机的简化建模方法

杨阳，戴树岭，雷小永*
(北京航空航天大学 虚拟现实技术与系统全国重点实验室，北京 100191)

摘　　要：为实现飞行仿真中涡桨发动机的精确建模，针对定转速控制规律下的涡桨发动机，深入剖析其部件工作特性，并运用简化部件模型对核心机进行一体化建模。为提高动态建模精度，引入系统辨识方法，采用二阶传递函数对模型动态过程进行近似，使加速过程的拟合度达到95.5%。此外，为提高模型的通用性，根据相似理论准则进行归一化处理，实现了模型飞行包线内运行的推广。利用公开的涡桨发动机部分试验和仿真数据，以及AH-20M涡桨发动机总体性能数据，通过调整模型设计点参数，快速进行建模。仿真结果表明：仅需利用有限的总体性能参数数据，即可迅速构建出具有一定稳态精度、动态变化过程合理的发动机模型。仿真结果为飞行仿真中的涡桨发动机建模提供了一种高效且精确的解决方案。

关 键 词：飞行仿真；涡桨发动机；定转速控制；系统辨识；发动机仿真；相似理论
中图分类号：V231; TP391.9
文献标志码：A　　**文章编号**：1001-5965(2026)03-0818-08

航空发动机建模是飞行仿真中非常重要的一部分，既要求模型性能数据足够准确，又要求动态变化过程基本合理。在实时的飞行仿真环境中，还要求模型不能过于复杂以满足实时运行要求。然而，由于航空发动机涉及高温、高压、高转速等极端工况，以及复杂的控制规律和物理过程，使得其建模和仿真成为一项极具挑战性的任务。

为更好地理解和预测航空发动机的性能，研究者们提出多种建模方法，包括部件法、总体法等。其中，部件法具有精度高、物理意义明确等优点，但面临着仿真过程复杂和部件特性数据获取困难的问题，模型很难实时运行。而飞行仿真更注重发动机总体性能的仿真，包括静态参数的精度和动态参数的合理性。为满足实时性和静态精度的要求，飞行仿真中通常采用总体法建模。但这种方法依赖详细的试飞数据，动态变化过程的精度有限且物理意义不明确[1]。

为解决上述问题，本文提出一种简化的部件建模方法，用于飞行仿真中的定转速控制规律类型的涡轮螺旋桨发动机建模。首先，将压气机、涡轮和转子在设计点线性化后进行一体化建模，并利用数值模型对螺旋桨进行仿真。然后，通过系统辨识方法的应用，提高模型的动态精度。之后，运用气动相似理论，将模型的应用范围扩展到包线内的各种飞行情况，以提高模型的通用性。最后，为验证模型的有效性和准确性，选取AH-20M涡轮螺旋桨发动机性能参数来调整模型。通过搭建整体仿真测试环境，对发动机从起动到停车的各个试车阶段进行模拟验证。

1　涡桨发动机简化部件模型

为使发动机模型进行简化，需要解决2个问题：①如何对部件模型进行简化来降低建模复杂度；②简化后的模型在建模精度方面表现如何。定转速控制规律类型的涡桨发动机的转子和主要部件(压气机和涡轮等)正常工作时保持在恒定转速，

收稿日期：2024-01-02；录用日期：2024-03-26；网络出版时间：2024-04-11 18:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20240410.2040.001
*通信作者. E-mail：leixy@buaa.edu.cn

引用格式：杨阳，戴树岭，雷小永. 一种定转速涡桨发动机的简化建模方法[J]. 北京航空航天大学学报，2026，52(3)：818-825.
YANG Y，DAI S L，LEI X Y. A simplified modeling method for a constant speed turboprop engine[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：818-825 (in Chinese).

这意味着建立工作状态模型时只需要考虑设计点的特性, 以此为基础可以使得模型大大简化, 本节对这一简化过程进行热力学推导。

1.1　转速恒定下的部件运行特性

航空发动机部件的特性通常由部件特性图来确定, 如图 1 所示, 以压气机为例, 可以使用换算流量来确定增压比π和工作效率η。

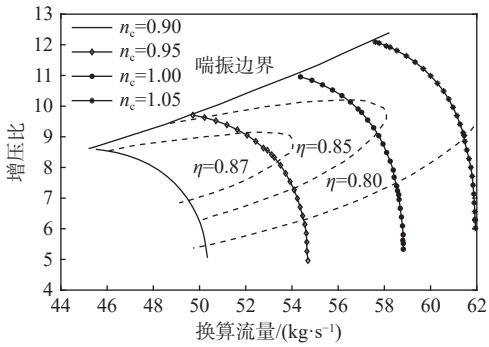


图 1　典型压气机部件特性

Fig. 1　Typical compressor component characteristic

压气机在不同转速和流量下的效率和增压比分别为

$$\eta = f_1(n_c, \dot{m}_c) \tag{1}$$

$$\pi = f_2(n_c, \dot{m}_c) \tag{2}$$

式中: n_c为换算转速, $n_c = n/\sqrt{T^*}$, n为转速, T^*为进口总温; $\dot{m}_c$为换算空气流量, $\dot{m}_c = \dot{m}\sqrt{T^*}/P^*$, $\dot{m}$为流量, P^*为进口总压。

压气机轮缘功计算式为

$$L = \left(\pi^{[k-1/k]} - 1\right) \tag{3}$$

$$N_c = L\dot{m}_c\eta_c \tag{4}$$

式中: L为轮缘功; k为气体常数, 对燃气k=1.3; N_c为压气机功率[2]; η_c为压气机效率。

对于涡轮, 同样有涡轮功率:

$$N_t = L_t \sim \dot{m}_c\eta_t \tag{5}$$

式中: L_t为涡轮轮缘功; η_t为涡轮效率。

因此, 由转速n, 流量$\dot{m}$就能确定η和π, 从而压气机的状态得以确定。

共同工作条件:

$$N_c = N_t - N_s \tag{6}$$

式中: N_s为转子输出功率。

同理, 对于涡轮功, 有

$$L_c\dot{m}_c = L_t\eta_t\eta_c\dot{m}_{gc} - N_s \tag{7}$$

式中: $\dot{m}_{gc}$为换算燃气流量。

结合式 (3), 有轮缘功$L = f(n_c, \dot{m}_c)$, $\dot{m}_g = \dot{m}_c + \dot{m}_f$, $n_c = n_t$。

其中, $\dot{m}_f$为燃油流量, 于是

$$f_c(n, \dot{m}_c) = f_t(n, \dot{m}_c + \dot{m}_f) + N_s \tag{8}$$

当有唯一解时, 也可以写作:

$$n = f(\dot{m}_c, \dot{m}_f, N_s) \tag{9}$$

在设计点一阶展开后, 有

$$n = f'(\dot{m}_{c0})(\dot{m}_c - \dot{m}_{c0}) + f'(\dot{m}_{f0})(\dot{m}_f - \dot{m}_{f0}) + f'(N_{s0})(N_s - N_{s0}) + n_0 + O(2) \tag{10}$$

式中: $O(2)$ 表示 2 阶小量。

因此, 只要知道设计点空气流量和燃油对转速的导数$f'(\dot{m}_{c0})$和$f'(\dot{m}_{f0})$, 即可得到设计点附近的稳态转速。

在标况下, 当转速和马赫数确定时, 空气流量$\dot{m}_{c0}$也是确定的。

一般来说, $\dot{m}_c >> \dot{m}_f$, 因此, 即使$\dot{m}_f$偏离设计点比较远, 仍然可以用式 (10) 近似计算。

综上所述, 仅需要添加在不同马赫数下设计点的特性参数, 在转速偏差不大的情况下, 即可实现对部件特性的简化仿真。

1.2　转子部件一体化仿真

由 1.1 节可知, 可以在定转速情况下, 通过在设计点附近线性化部件特性来简化部件建模过程。具体来说, 将压气机、涡轮的部件特性和转子的动力特性进行线性化处理, 然后, 作为一个整体考虑, 用一个转子、涡轮、压气机的一体化部件来代替。

对于单转子涡桨发动机, 桨叶通过减速器与转子相连。发动机的转速与油门的关系是非线性的。当处于设计点附近定转速工作时, 可以用线性模型代替复杂的非线性关系。由式(10)可知, 在设计点附近, 稳定转速n_{stab}:

$$n_{stab} - n_0 = f'(\dot{m}_{c0})(\dot{m}_c - \dot{m}_{c0}) + f'(\dot{m}_{f0})(\dot{m}_f - \dot{m}_{f0}) - f'(N_{s0})(N_s - N_{s0}) \tag{11}$$

因此, 只要知道设计点的燃油和桨功率及其导数, 在转速偏离设计点不大的情况下, 可以得到比较准确的工作状态的转子模型, 如图 2 所示。

1.3　螺旋桨模型

涡桨发动机的主要动力来源是螺旋桨。螺旋桨的拉力和功率由多种因素决定, 包括温度、压力、马赫数、转速和桨距角等。常用的螺旋桨仿真计算方法包括片条法和涡格法等[3]。

飞行仿真对螺旋桨的建模一般考虑整体性能, 使用无量纲的特性曲线可以很好的模拟螺旋桨性能, 美国国家航空航天局(National Aeronautics and Space Administration, NASA)很早就开展了对各种

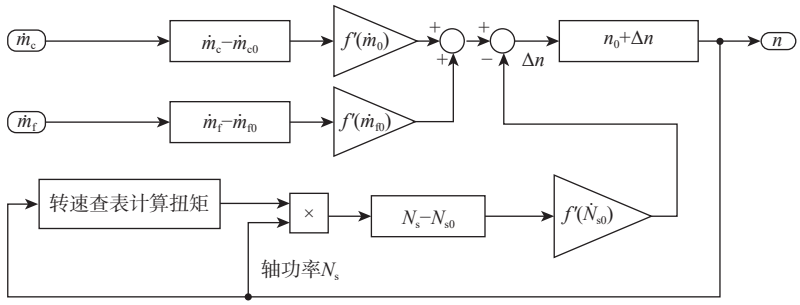


图 2 一体化转子 Simulink 仿真模型

Fig. 2 Integrated rotor Simulink simulation model

螺旋桨性能的试验，取得了丰富的数据[4-7]。变桨距螺旋桨的性能一般可以使用功率特性图来表示，如图 3 所示[8]，其中，J为进动比，C_P为功率系数。

在标况下，螺旋桨的功率系数C_P与进动比J、桨角θ有关[9]。

$$C_P = f(\theta, J) \tag{12}$$

于是，功率P：

$$P = C_P\rho\left(n_p/60\right)^3\left(2R_p\right)^5 \tag{13}$$

式中：ρ为空气密度；R_p为螺旋桨半径；n_p为螺旋桨转速，假设减速器能量转换效率η为定值，且R_p为常数，在标况下，轴功率N_P：

$$N_P = P/\eta = f\left(n_p, \theta\right) \tag{14}$$

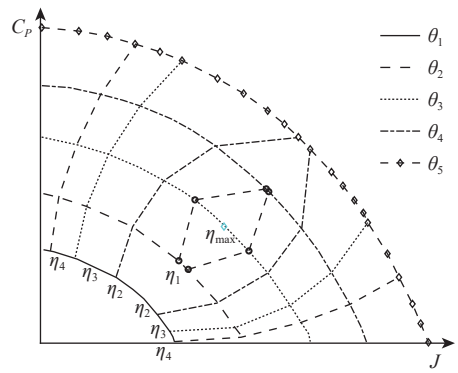


图 3 典型螺旋桨功率特性[8]

Fig. 3 Typical propellers power characteristic[8]

螺旋桨功率模型如图 4 所示，图中，K_1为螺旋桨转速系数，K_2为转换功率系数。

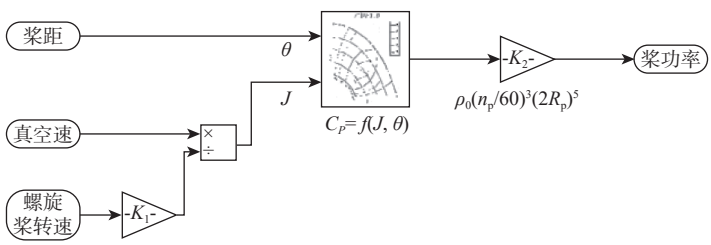


图 4 螺旋桨 Simulink 仿真模型

Fig. 4 Propeller Simulink simulation model

2 动态过程仿真

在系统稳态点模型建立之后，还需要对动态变化过程进行仿真。这主要包括从慢车加速到工作转速及从工作转速减速到慢车 2 种情况。对于飞行模拟机而言，通常只对发动机的过渡时间有要求，而对变化过程的精度要求相对较低。因此，可以使用简单的惯性环节或二阶环节来近似模拟动态过程。

2.1 利用系统辨识对动态变化进行仿真

对于动态过程精度要求比较高的情形，可以使用系统识别方法来建立传递函数。系统辨识是一种通用的测试建模方法，可以通过分析未知系统的输入输出数据，拟合一个与所测系统外特性等价的数学模型。在已知动态过程数据的情况下，参数辨识方法对于获取传递函数非常有效。文献 [3,10] 使用系统辨识方法，成功使用实验数据对涡桨和涡扇发动机进行建模。

本文选取一组测试数据，输入数据为目标转速，输出数据为转速变化。动态过程识别测试参数如图 5 所示。

选取零点为 2、极点为 2 的传递函数进行辨识，得到的传递函数如下：

$$G(s) = \frac{0.992\,3s^2 + 4.512s + 0.643}{s^2 + 5.082s + 0.609\,5} \tag{15}$$

使用 MATLAB 的 pem 函数对仿真模型进行测

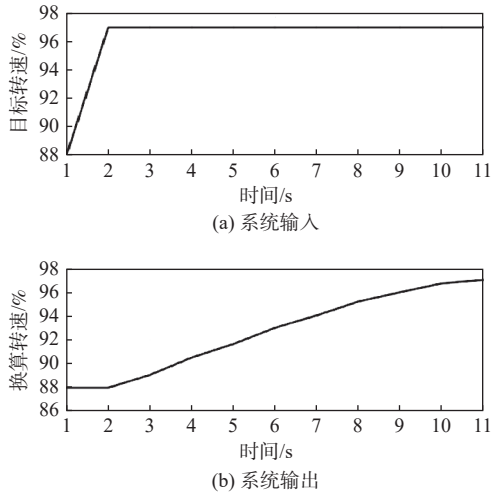


图 5　动态过程识别测试参数

Fig. 5　Dynamic process identification test parameters

试, 数据拟合度为 95.5%。

如图 6 所示, 使用系统辨识的方法可以很好的拟合动态变化过程。然而, 需要指出的是, 使用系统辨识方法得到的黑盒模型, 本文只将其用于动态变化过程的模拟。具体来说, 包括起动过程和从慢车到工作状态的加减速过程。在设计点附近, 由于转子和螺旋桨等主要部件特性偏差较小, 与在飞行仿真中仅模拟整体性能的发动机模型不同, 本文模型对主要部件的特性拥有一定的仿真能力, 因而相对整体模型更具有可解释性。

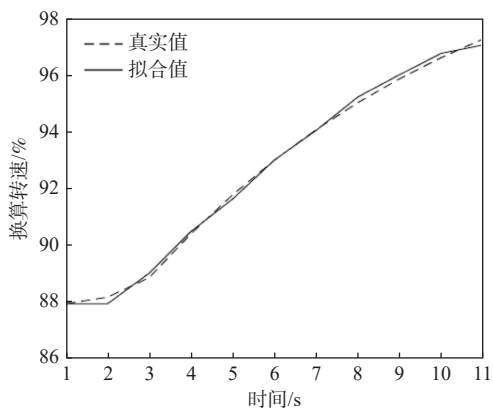


图 6　传递函数模型拟合结果

Fig. 6　Fitting results of transfer function model

2.2　定转速控制规律简易模型

涡桨发动机存在多种控制形式, 如机械控制、全权限电子控制器(full authority digital engine control, FADEC)等[11]。本文以机械液压控制结构为例, 来说明控制系统模型的建立方法。机械液压控制系统主要由发动机控制器和螺旋桨控制系统 2 个独立的系统组成。油门杆用于控制发动机的燃油量, 而螺旋桨调节器则通过调整桨距角来保持转速的恒定。计算流程涉及燃油流量与油门杆角度的关系, 以及目标转速与螺旋桨控制函数的关系[9], 计算流程如下:

$$\dot{m}_{\mathrm{f}} = f(\theta_{\mathrm{TLA}}) \tag{16}$$

式中: θ_{TLA}为油门杆角度。

$$\mathrm{d}\theta_{\mathrm{p}} = \mathrm{PID}\left(n_{\mathrm{tag}}, n\right) \tag{17}$$

式中: $\mathrm{d}\theta_{\mathrm{p}}$为桨距角变化量; n_{tag}为目标转速; PID为PID控制函数。

3　发动机模型的包线内推广

目前的发动机模型可以在标准大气状态下运行, 在缺少包线内详细数据的情况时, 为适应包线内不同高度、速度和大气状态的飞行, 可以利用相似准则进行推广, 虽然会损失部分精度, 但建模简单快速。

3.1　发动机相似准则

相似准则可以将标准大气条件下计算的参数推广到非标大气条件下。对于发动机, 其环境无量纲换算参数为

$$\theta_T = T_0 / T_{\mathrm{std}} \tag{18}$$

$$\delta = P_0 / P_{\mathrm{std}} \tag{19}$$

式中: θ_T为温度折合系数; δ为压力折合系数; T_0为环境温度; P_0为环境压力; T_{std}为标准大气温度; P_{std}为标准大气压力。

通过折合系数将标准大气条件下的仿真发动机参数换算到其他大气条件[12], 发动机相似参数计算式为

$$n = n_{\mathrm{c}} \sqrt{\theta_T} \tag{20}$$

$$\dot{m}_{\mathrm{f}} = \dot{m}_{\mathrm{fc}} \delta \sqrt{\theta_T} \tag{21}$$

$$N_{\mathrm{s}} = N_{\mathrm{sc}} \delta \sqrt{\theta_T} \tag{22}$$

式中: N_{sc}为对应的折合参数。

3.2　螺旋桨相似准则

螺旋桨与发动机相似准则略有不同, 文献 [13] 给出了考虑雷诺数的相似准则。

主要换算系数如下。

长度系数η_{L}:

$$\eta_{\mathrm{L}} = D / D_0 \tag{23}$$

式中: D为螺旋桨直径; D_0为参考螺旋桨直径。

运动黏性系数η_υ:

$$\eta_\upsilon = \upsilon / \upsilon_{\mathrm{std}} \tag{24}$$

式中: υ空气黏度; υ_{std}为参考标准状况下空气黏度。

密度系数η_ρ:

$$\eta_\rho = \rho / \rho_{\mathrm{std}} \tag{25}$$

式中：ρ为空气密度；ρ_{std}为参考标准状况下空气密度。

类似地，螺旋桨转速n、拉力T和轴功率N_P计算式分别为

$$n = n_{\mathrm{c}}\eta_v/\eta_{\mathrm{L}} \tag{26}$$

$$T = T_{\mathrm{c}}\eta_\rho\eta_v^2 \tag{27}$$

$$N_P = N_{\mathrm{c}}\eta_\rho\eta_v^3/\eta_{\mathrm{L}} \tag{28}$$

需要指出的是，使用相似准则推广是一种近似方法，虽然可以加速建模，但当偏离设计点较远时误差也会变大[14]，在拥有更全面的数据后，应该将推广参数补充为包线内参数，由固定值改为插值表。

4 仿真验证

4.1 模型精度验证

文献[15]为某型涡桨的起动过程试验数据，连接起动机和时序的燃油输入，对起动过程进行验证。

供油使用试验数据的插值表得到，空气流量等特性值用设计点参数代替，起动机使用文献[15]中的数据进行建模，起动过程验证模型如图7所示。

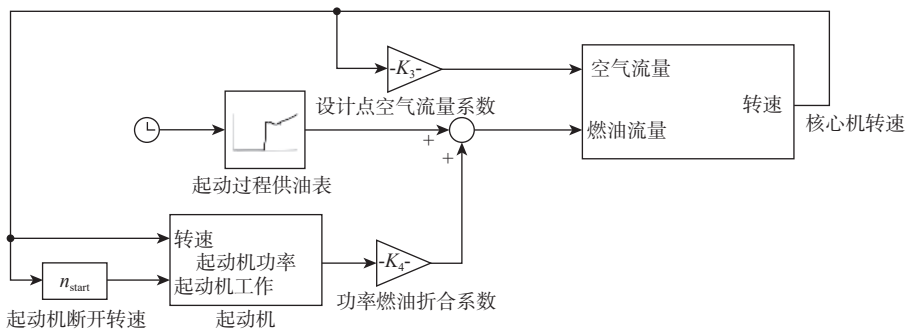


图7 起动过程验证模型

Fig. 7 Model for verifying the starting process

发动机起动过程对比如图8所示，虽然起动时部件特性偏离设计点较远，但起动过程仿真仍具有一定精度。另一方面，部分变化比较大的区域，如起动刚开始时的转速变化，有明显的细节缺失，这表明传递函数动态过程仿真在细节建模方面有所不足。

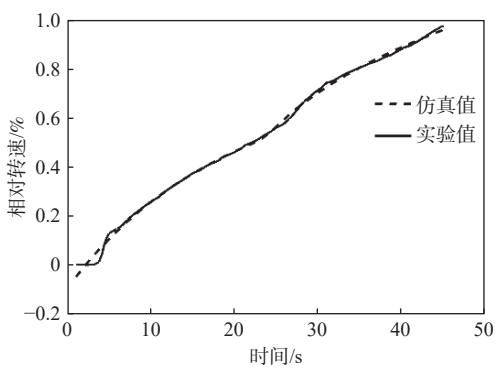


图8 发动机起动过程对比

Fig. 8 Validation model of the start-up process

GASTURB是燃气轮机仿真的一款比较成功的软件，文献[15]基于该软件建立了涡桨发动机的核心机模型，使用该模型的仿真数据与本文建模方法进行动态变化对比。

测试通过输入阶跃的油门杆角度，如图9所示，对比转速变化，如图10所示。结果表明，转速的总体趋势和响应基本符合，但动态过程细节和时间有所差异。

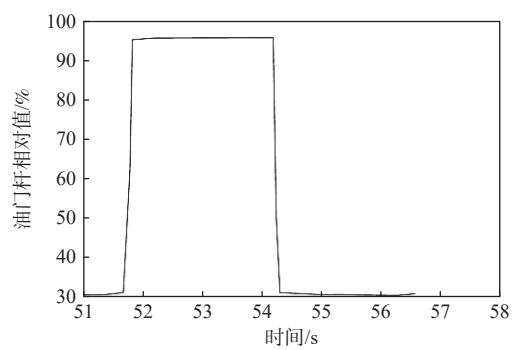


图9 阶跃油门响应测试

Fig. 9 Step throttle response test

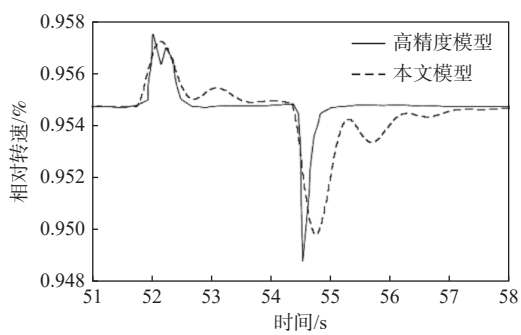


图10 与高精度模型仿真对比的转子转速输出

Fig. 10 Comparison of rotor speed output with high-precision model simulation

4.2　仿真试车验证

为验证模型是否满足飞行仿真功能要求, 选取 AH-20M 发动机参数进行调参[16], 使用试车条目要求来对模型进行功能性测试。选取参数如表 1 所示。

使用 MATLAB/Simulink 搭建了实时仿真环境, 如图 11 所示, 并额外搭建了简单外围控制逻辑, 用来测试模型的实时性能和各类工作状态下的响应。

测试工况包含飞行仿真中涡桨发动机典型的工作状态, 由于缺乏详细的数据, 测试结果分为性能验证和功能验证 2 个部分[15]。

表 1　选取的 AH-20M 设计点参数

Table 1　Selected design point parameters of AH-20M

参数	设计值
*慢车转速/%	81
*设计转速/%	91
起飞耗油率/(kg·kWh^{-1})	0.375 5
起飞功率/kW	3 169
空气流量/(kg·s^{-1})	20
巡航功率/kW	2 013
*起动时间/s	55
*加速时间/s	20
*停车时间/s	75

注: 带“*”的数据来自类似型号推测的数据。

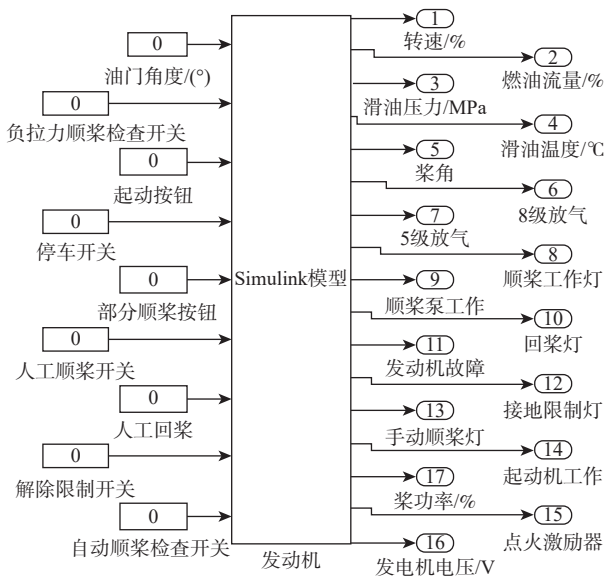


图 11　Simulink 集成仿真环境

Fig. 11　Simulink integrated simulation environment

性能指标是对模型精度进行考察, 包括稳态性能验证、动态性能验证、起动性能验证等。本文选取稳态转速、动态变化时间为验证依据。

功能验证主要指发动机正常运行的各类功能性响应, 对于涡桨发动机, 主要验证加减速和螺旋桨顺桨、回桨变化响应。

涡桨发动机的试验(检验试车)具有复杂的程序[17-18], 本文根据试车程序表, 结合仿真需要简化后制定测试工况。表 2 给出了测试工况和测试结果, 图 12 给出了主要参数变化。

表 2　测试工况和测试结果

Table 2　Test conditions and test results

阶段	测试工况	性能测试结果	功能测试结果
起动	测试从停车到慢车的动态过程	①起动时间约为51 s; ②转速稳定在81%	在起动机工作和供油时有明显增速过程
加速	从慢车到设计转速的加速过程	①加速时间约为19 s; ②设计转速为91%	转速平滑增加
顺桨	螺旋桨顺桨响应	①桨角迅速增大至90°; ②桨功率迅速下降	桨角迅速增大
回桨	螺旋桨回桨响应	①桨角逐渐恢复; ②桨功率随桨角恢复	桨角逐渐恢复
停车	发动机从慢车到停车的动态过程	停车时间约为75 s	指数型减小

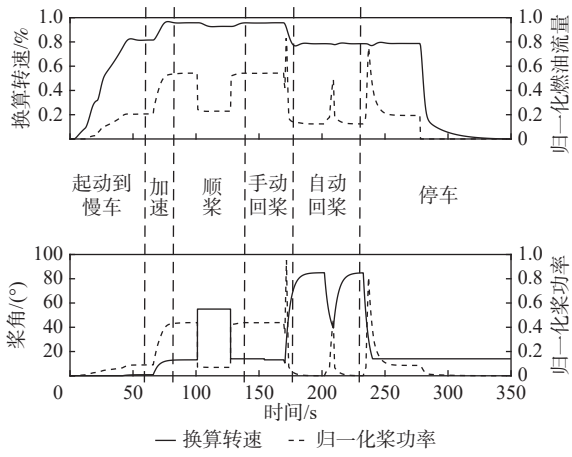


图 12 典型工况下主要参数变化

Fig. 12 Variations of main parameters under typical operating conditions

5 结 论

1）一体化的部件简化建模方法可以成功应用于涡桨发动机建模，仿真模型在设计点具有较高的可信度。

2）使用系统辨识的方法得到的二阶传递函数，可以建立动态变化过程模型，对加速过程拟和度达到95.5%。在典型工况下进行测试，起动、停车等过程变化也符合正常发动机规律。

3）通过对比试验数据和精细模型仿真结果，使用传递函数加线性化模型的方法可以得到比较好的稳态点精度和基本正确的动态响应。但在动态变化的细节方面有所不足。

4）通过使用AH-20M发动机的公开参数来建立模型，并进行仿真试车测试。表明该模型具有良好的通用性和适用性，能够满足飞行仿真中对定转速涡桨发动机性能、功能的建模要求。

综上所述，本文模型在稳态计算和动态仿真方面均表现出良好的准确性和可靠性。通过进一步完善和优化模型参数和控制策略，可以提高仿真的逼真度和应用范围，为涡桨发动机的研究和发展提供更有价值的工具和参考。

参考文献（References）

[1] 熊海国. 飞行模拟器发动机系统建模与面向对象的仿真研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
XIONG H G. Research on modeling and object-oriented simulation of flight simulator engine system[D]. Harbin: Harbin Institute of Technology, 2010(in Chinese).

[2] 彭泽琰. 航空燃气轮机原理[M]. 北京：国防工业出版社, 2008: 97-107.
PENG Z Y. Principles of aero gas turbine [M]. Beijing: National Defense Industry Press, 2008: 97-107(in Chinese).

[3] 时培燕, 毛宁, 杨恒辉, 等. 涡桨发动机螺旋桨建模与控制系统设计研究[J]. 计算机测量与控制, 2016, 24(5): 103-105.
SHI P Y, MAO N, YANG H H, et al. Research on modeling and control system designing of turboprop engine propeller[J]. Computer Measurement & Control, 2016, 24(5): 103-105(in Chinese).

[4] DURAND W F. Experimental research on air propellers: NACA-TR-14 [R]. Washington, D. C. : NASA, 1918.

[5] DURAND W F, LESLEY E P. Experimental research on air propellers II: NACA-TR-30[R]. Washington, D. C. : NASA, 1920.

[6] DURAND W F, LESLEY E P. Experimental research on air propellers III: NACA-TR-64[R]. Washington, D. C. : NASA, 1920.

[7] DURAND W F, LESLEY E P. Experimental research on air propellers IV: NACA-TR-109[R]. Washington, D. C. : NASA, 1921.

[8] 史永运, 钟易成, 邓君湘, 等. 涡轮螺旋桨动力飞机桨发匹配性能仿真研究[J]. 机械制造与自动化, 2019, 48(4): 116-120.
SHI Y Y, ZHONG Y C, DENG J X, et al. Research on prop-engine cooperation performance simulation of propeller powered aircraft[J]. Machine Building & Automation, 2019, 48(4): 116-120 (in Chinese).

[9] 张孟伟. 单杆操控的变桨距微型涡桨发动机控制系统设计与验证[D]. 南京: 南京航空航天大学, 2020.
ZHANG M W. Design and verification of control system of variable pitch micro turboprop engine controlled by single rod[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020 (in Chinese).

[10] 郭迎清, 王海泉. 涡扇发动机模型辨识及其控制器设计[J]. 现代制造工程, 2006(9): 73-74.
GUO Y Q, WANG H Q. System identification of a turbofan engine and design of its controller[J]. Modern Manufacturing Engineering, 2006(9): 73-74(in Chinese).

[11] 姚华. 航空发动机全权限数字电子控制系统[M]. 北京: 航空工业出版社, 2014.
YAO H. Full authority digital electronic control system for aeroengine [M]. Beijing: Aviation Industry Press, 2014(in Chinese).

[12] 廉筱纯, 吴虎. 航空发动机原理[M]. 西安: 西北工业大学出版社, 2005.
LIAN X C, WU H. Aeroengine principle[M]. Xi'an: Northwestern Polytechnical University Press, 2005(in Chinese).

[13] 刘沛清, 马利川, 段中喆, 等. 平流层飞艇螺旋桨相似准则分析与验证[J]. 北京航空航天大学学报, 2012, 38(7): 957-961.
LIU P Q, MA L C, DUAN Z Z, et al. Study and verification on similarity theory for propellers of stratospheric airships[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(7): 957-961(in Chinese).

[14] 缪辉, 臧朝平, 王晓伟, 等. 航空发动机高压转子试验模型的动力学相似设计[J]. 推进技术, 2021, 42(10): 2340-2348.
MIAO H, ZANG C P, WANG X W, et al. Dynamic similarity design for experimental model of aero engine high-pressure rotor[J]. Journal of Propulsion Technology, 2021, 42(10): 2340-2348(in Chinese).

[15] 田超. 航空涡桨发动机建模与控制的仿真研究[D]. 南京: 南京航空航天大学, 2010.
TIAN C. Simulation study on modeling and control of aircraft turboprop engines[D]. Nanjing : Nanjing University of Aeronautics

and Astronautics , 2010(in Chinese).
[16] 方昌德. 世界航空发动机手册[M]. 北京: 航空工业出版社, 1996.
FANG C D. The world aircraft engines handbook[M]. Beijing: Aviation Industry Press, 1996(in Chinese).
[17] 高亚奎, 朱江, 林皓, 等. 飞行仿真技术[M]. 上海: 上海交通大学出版社, 2015.
GAO Y K, ZHU J, LIN H, et al. Flight simulation technology[M]. Shanghai: Shanghai Jiao Tong University Press, 2015(in Chinese).
[18] 黄庆南. 航空发动机设计手册. 第 6 册, 涡轮[M]. 北京: 航空工业出版社, 2001.
HUANG Q N. Aerospace engine design handbook. Volume 6 , turbines [M]. Beijing: Aviation Industry Press, 2001(in Chinese).

A simplified modeling method for a constant speed turboprop engine

YANG Yang，DAI Shuling，LEI Xiaoyong*

(National Key Laboratory of Virtual Reality Technology and System，Beihang University，Beijing 100191，China)

Abstract：To achieve precise modeling of turboprop engines in flight simulation, this study provides an in-depth analysis of the component operating characteristics of the turboprop engine under the constant speed control law. Based on these characteristics, we integrated the core engine using a simplified component model. A 95.5% match for the acceleration process is obtained by approximating the model's dynamic processes using second-order transfer functions, which are introduced as a means of improving the accuracy of dynamic modeling. In addition, to enhance model generality, normalization processing is conducted based on similarity theory principles, enabling extrapolation within the flight envelope. In the validation phase, we carried out a simulation and comparative validation of the model using publicly available data on the amount of overall performance of the AH-20M turboprop engine, as well as some experimental and simulation data. The results showed that only a limited overall performance parameter data was needed to quickly build an engine model with certain steady-state accuracy and reasonable dynamic changes. The model is an effective and precise way to model turboprop engines in flight simulation, and it satisfies the performance and functionality criteria of flight simulation.

Keywords：flight simulation；turboprop engine；constant speed control；system identification；engine simulation；similarity theory

Received：2024-01-02；**Accepted**：2024-03-26；**Published Online**：2024-04-11 18：28
URL：link.cnki.net/urlid/11.2625.V.20240410.2040.001
* **Corresponding author**. E-mail：leixy@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0771

非均匀加热下超临界燃油在并联双管中质量流量分配特性

董昊[1]，赵超凡[2,*]，朱剑琴[3,4]，程泽源[3]

(1. 长光卫星技术股份有限公司，长春 130033；2. 北京机电工程总体设计部，北京 100854；3. 北京航空航天大学 能源与动力工程学院 航空发动机气动热力国家级重点实验室，北京 100191；4. 天目山实验室，杭州 310023)

摘 要：为研究线性非均匀加热下超临界燃油在并联双管中的质量流量分配特性，通过瞬态模拟超临界燃油RP-3流经并联双管的流动过程，探究质量流量分配特性的影响因素及其作用机理；通过稳态模拟平均热流密度在2.7～3.9 MW/m^2、热流增长率在0.025～0.225范围内的质量流量分配特性，研究不同非均匀加热条件对质量流量分配特性的影响。结果表明：拟临界状态、裂解反应、固体导热、分流阻力差是影响质量流量分配特性的主要因素；拟临界状态正反馈、裂解反应正反馈会增大双管间的质量流量偏差；固体导热负反馈、分流阻力差负反馈会减小双管间的质量流量偏差。裂解反应正反馈被激活的必要条件是总加热功率可使全部冷却工质达到裂解温度。保持平均热流密度3.7 MW/m^{2}不变，热流增长率增大了8倍，质量流量相对标准差增大了11%。非均匀加热条件主要起质量流量分配偏差的诱导作用，对最终质量流量分配特性的影响较小。

关 键 词：超临界燃油；并联管道；非均匀加热；质量流量分配；机理
中图分类号：V231$^+$1
文献标志码：A **文章编号**：1001-5965(2026)03-0826-11

超声速飞行器能够稳定地在超声速状态下巡航，离不开可靠高效的发动机热防护系统。对于巡航速度达到6倍声速的超燃冲压发动机，其燃烧室中燃气的总温达到2 800 K以上[1-2]，而材料无法承受如此高的温度；高超声速下来流气体的滞止温度也较高，无法满足发动机内高温部件的冷却需求。因此，高超声速飞行器的热防护系统需要解决缺少冷源、提高冷却效率两大难题。再生冷却技术作为新型高效的冷却技术应运而生[3]，其以燃油作为冷源，综合利用燃油的物理热沉及化学热沉，将高温部件的能量转换为燃油的内能及化学能，并在之后的燃烧中重新释放。再生冷却系统一般采用并排密布的平板冷却通道，当冷却平板处于非均匀加热状态时，不同管道间的质量流量会出现不均匀分配，导致部分管道冷却工质质量流量无法满足冷却需求。

对于并排管道质量流量分配问题，目前的数学研究方法有3种：经验关系式法[4]、管道分段法[5-6]和数值模拟法。经验关系式法是通过建立各个管道的质量流量与流动阻力的经验关系式，将质量流量、流动阻力分别与电流、电阻类比，基于电路的基本规律，得到各分支管道的质量流量情况；管道分段法是将总管道及分支管道按各自流向分成若干段，在每段上建立相应的控制方程，并联立求解，得到各管道的质量流量分配情况；数值模拟法是使用三维数值模拟软件来进行仿真，得到各管道质量流量分配情况。对于超临界燃油，其质量流量与流动阻力的经验关系式适用范围小，拟临界状态导致其经验关系式易出现间断点而不连续，不满足经验关系式法所要求的连续性要求；管道分段法会忽略

收稿日期：2023-11-30；录用日期：2024-04-05；网络出版时间：2024-05-20 19:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20240518.1422.001
基金项目：国家自然科学基金(52122604,52306064)；天目山实验室项目(TK-2023-C-009)
*通信作者. E-mail：zhaochaofan@buaa.edu.cn

引用格式：董昊，赵超凡，朱剑琴，等. 非均匀加热下超临界燃油在并联双管中质量流量分配特性[J]. 北京航空航天大学学报，2026，52(3)：826-836. DONG H，ZHAO C F，ZHU J Q，et al. Mass flow distribution characteristics of supercritical fuel oil in parallel tubes under non-uniform heating conditions[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：826-836 (in Chinese).

转角等结构导致的流动分离的影响, 在再生冷却通道中, 转角相对通道尺寸较大, 流体流速较高, 转角处的阻力影响不可被忽略。因此, 针对再生冷却通道的质量流量分配问题, 当前研究以数值模拟法为主。

针对超临界燃油, 部分学者结合其特殊的物性及化学性质, 开展了不同温度范围内的质量流量分配机理研究。Jiang 等[7-10] 基于文献 [11] 的实验数据, 通过线性变化的热流密度分布来近似真实发动机中的热环境, 数值模拟了 5 条并联管在热流不均匀分布情况下的质量流量分配情况, 分析了质量流量分配特性对各物性变化的敏感度, 并分别通过设置入口温度为拟临界温度和裂解温度, 研究了拟临界状态及裂解反应对质量流量的影响, 结果表明, 质量流量分配对密度的敏感性最强, 且拟临界温度引起的物性变化及裂解反应都会扩大质量流量的差异。此外, 部分学者开展了初始条件对质量流量分配影响的研究。Chen 等[12] 通过数值仿真研究了再生冷却通道在不均匀加热情况下的质量流量分配情况及管道温度分布, 并证明了初始条件的微小质量流量差异会在加热后产生明显的质量流量差。除初始条件的影响外, 也有部分学者研究了分汇流结构对质量流量分配特性的影响。Ma 等[13] 针对并联蛇形管排的质量流量分配特性进行了数值研究, 发现管道入口及出口处的分汇流装置导致的流动损失对质量流量分配的影响不可忽略。Zhang 等[14] 通过数值模拟研究的分汇流结构参数对质量流量分配的影响也支持该结论, 其发现较大的分流结构能有效降低流体惯性力对质量流量分配的影响。针对质量流量不均匀分配的现象, 学者开展了管排结构的优化设计。Jing 等[15] 对含有联通结构的并联双管通道进行了数值模拟研究, 结果表明, 对于 10% 偏差及 20% 偏差比例的不均匀加热工况, 均可以在 200 mm 处布置联通结构, 以改善质量流量分配的差异。Chen 等[16] 通过数值模拟含有节流孔结构并联管道的质量流量分配情况, 使超临界燃油在节流孔处达到声速, 降低了下游因素对质量流量分配的影响。

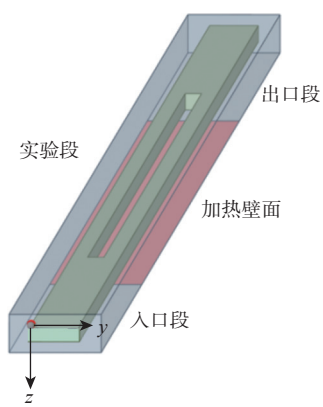
综上, 目前超临界燃油的质量流量分配特性研究主要包含特定温度范围内导致质量流量不均匀分配的正反馈因素机理研究及不同结构参数对质量流量分配特性的影响。而再生冷却平板中燃料的质量流量分配特性受拟临界状态和裂解反应的共同影响, 其共同作用下对质量流量分配特性的影响因素探究仍不全面, 联立考虑拟临界状态和裂解反应的一体化质量流量分配机理研究仍有不足, 非均匀加热边界的变化对质量流量分配特性的影响仍不明确。因此, 本文聚焦于超临界燃油的质量流量分配特性问题, 基于再生冷却平板的典型通道结构——并联双管结构, 开展线性非均匀加热条件下的质量流量分配特性研究, 探究综合考虑拟临界状态、裂解反应影响下的质量流量分配影响因素及其作用机理, 得到非均匀加热工况变化对质量流量分配偏差的影响, 为复杂热环境下的再生冷却平板质量流量分配特性研究及其设计提供理论支撑。

1　数值计算方法

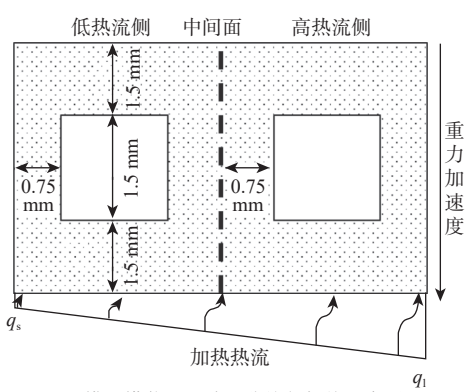
1.1　物理模型

图 1 为本文采用的并联冷却管道物理模型。其为总长 1 100 mm 水平放置的并排双管结构, 分别设置 250 mm 的入口段和出口段。入口段与实验段的交界处为分流位置, 出口段与实验段的交界处为汇流位置。分支管为实验段的 2 个子管道, 其为边长 1.5 mm 的方形管, 2 分支管的间距为 1.5 mm, 管道侧壁厚为 0.75 mm, 上下壁厚为 1.5 mm。实验段的加热面在底面, 采用如图 1(b) 所示的非均匀加热方式, 热流密度沿 y 正方向从低热流侧热流密度 q_s 线性增长到高热流侧热流密度 q_l, 定义热流增长率 θ 为

$$\theta = \frac{q_l}{q_s} - 1 \tag{1}$$



(a) 几何模型



(b) 模型横截面尺寸及非均匀加热示意图

图 1　并联冷却管道物理模型

Fig. 1　Physical model of parallel cooling channels

参考超燃冲压发动机工况范围[10]，热流增长率变化范围为0.025～0.225，平均热流密度q_{ave}的变化范围为2.7～3.9 MW/m^2。所有算例在初始状态下2个管道质量流量相同，壁面采用无滑移边界条件，流固耦合界面采用热耦合边界条件，考虑到真实发动机的隔热布局[10]，除加热面外的其余壁面为绝热条件，固体材料选用304不锈钢。表1为边界条件汇总。

表1 边界条件

Table 1 Boundary conditions

工况组	平均热流密度/(MW·m^{-2})	热流增长率	冷却工质入口速度/(m·s^{-1})	冷却工质入口温度/K	稳态/瞬态
A	3.7	0.1	1.8	300	瞬态
B	3.7	0.025～0.225	1.8	300	稳态
C	2.7～3.9	0.1	1.8	300	稳态

1.2 控制方程

采用商业流体计算软件进行数值模拟。超临界燃油在管排中的流动涉及组分扩散方程、连续方程、动量方程、能量方程，因此，流体区域的控制方程如下。

组分扩散方程：

$$\frac{\partial(\rho Y_m)}{\partial t}+\frac{\partial(\rho u_i Y_m)}{\partial x_i}=\frac{\partial}{\partial x_i}\left(\rho D_m\frac{\partial Y_m}{\partial x_i}\right)+R_m \tag{2}$$

式中：ρ为流体密度；u_i为流体速度分量；x_i为坐标系的分量；下标i=x, y, z分别表示x、y、z这3个方向；t为时间；D_m为组分m的扩散系数；Y_m为组分m的质量分数；R_m为组分m的质量源。

连续方程：

$$\frac{\partial\rho}{\partial t}+\frac{\partial(\rho u_i)}{\partial x_i}=0 \tag{3}$$

动量方程：

$$\frac{\partial(\rho u_i)}{\partial t}+\frac{\partial(\rho u_i u_j)}{\partial x_j}=-\frac{\partial p}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\eta\frac{\partial u_i}{\partial x_j}-\rho\overline{u_i'u_j'}\right)+\rho g_i \tag{4}$$

式中：p为压力；g_i为加速度的分量；$\eta\dfrac{\partial u_i}{\partial x_j}$表征黏性应力项；$\rho\overline{u_i'u_j'}$为Reynolds湍流应力。

能量方程：

$$\frac{\partial(\rho h)}{\partial t}+\frac{\partial(\rho u_i h)}{\partial x_i}=\frac{\partial}{\partial x_i}\left(\frac{\lambda}{c_p}\cdot\frac{\partial h}{\partial x_i}+\frac{\mu_t}{Pr_t}\cdot\frac{\partial h}{\partial x_i}\right)+S \tag{5}$$

式中：h为比焓；λ为导热系数；c_p为比热容；μ_t为湍流黏性系数；Pr_t为湍流普朗特数；S为能量源，表示裂解反应引起的控制体内的能量变化。

根据文献[17-18]，k-ω SST湍流模型能较准确地预测超临界流体流动换热过程。因此，采用k-ω SST湍流模型进行计算，其控制方程如下。

湍动能方程：

$$\frac{\partial(\rho k)}{\partial t}+\frac{\partial(\rho k)}{\partial x_i}+\frac{\partial(\rho u_i k)}{\partial x_i}=\frac{\partial}{\partial x_i}\left[(\mu+\mu_t/\sigma_k)\frac{\partial k}{\partial x_i}\right]+G_k-Y_k+S_k+G_b \tag{6}$$

比耗散率方程：

$$\frac{\partial(\rho\omega)}{\partial t}+\frac{\partial(\rho\omega)}{\partial x_i}+\frac{\partial(\rho u_i\omega)}{\partial x_i}=\frac{\partial}{\partial x_i}\left[(\mu+\mu_t/\sigma_\omega)\frac{\partial\omega}{\partial x_i}\right]+G_\omega-Y_\omega+G_{\omega b}+S_\omega \tag{7}$$

式中：k为湍动能；ω为比耗散率；μ为层流黏性系数；G_k和G_ω分别为湍动能k和比耗散率ω的生成量；Y_k和Y_ω分别为湍动能k和比耗散率ω的消散量；σ_k和σ_ω分别为湍动能k和比耗散率ω的湍流普朗特数；S_k和S_ω分别为湍动能k和比耗散率ω的源项；G_b和$G_{\omega b}$为表明浮升力对湍流方程影响的参数。

能量方程在固体区域退化为固体导热方程：

$$\frac{\partial(\rho h)}{\partial t}=k_s\frac{\partial^2 T}{\partial x_i^2} \tag{8}$$

式中：k_s为固体导热系数；T为温度。

瞬态计算采用压力隐式算子分离(pressure implicit with splitting of operators, PISO)算法，稳态计算采用压力耦合方程组的半隐式算法(semi-implicit method for pressure linked equations, SIMPLE)，压力、组分质量分数、速度、能量采用二阶离散格式，湍动能、比耗散率采用一阶离散格式。计算中，监视质量流量和出口温度的变化，当各控制方程及监视点的残差降至10^{-6}时，认为达到收敛。计算网格在流体域与固体域耦合面共节点，实现温度与热流密度数据的准确传递。

1.3 计算模型

RP-3作为吸热型碳氢燃料，被广泛应用在再生冷却技术中[19-25]。选用RP-3作为冷却工质，研究超临界燃油RP-3在并联双管中的质量流量分配特性。

1.3.1 裂解模型

采用Jiang等[20]通过超临界实验获得的24步RP-3裂解反应动力学模型(包含1个主反应和23步二次裂解反应)。主反应为

$$\begin{aligned}\text{RP-3}\rightarrow{}&0.108\ 6H_2+0.477\ 3CH_4+0.558\ 6C_2H_4+\\&0.39C_2H_6+0.41C_3H_6+0.200\ 1C_3H_8+\\&0.224\ 6C_4H_8+0.035\ 3C_4H_{10}+0.031C_4H_6+\\&0.720\ 1C_{5+}+0.27CC_{5+}+0.022\ 2C_nH_{2n-6}\end{aligned} \tag{9}$$

式中：C_{5+}为C_5～C_{11}的烯烃和少量烷烃；CC_{5+}包含环烯烃和少量环烷烃；C_nH_{2n-6}为不包含有苯、甲苯、乙苯、苯乙烯的芳香烃混合物。剩余23个二次反应模型可参考文献[20]。

1.3.2　物性模型

采用广义对应态法则计算 RP-3 及其裂解产物的密度、黏度、导热系数和定压比热容。广义对应态法则是根据已知参考物质的物性信息来预测其他物质物性的方法。使用丙烷作为参考物质, 其密度通过 32 参数修正的贝内迪克特-韦德-鲁宾(modified Benedict Webb-Rubin, MBWR)方程[26-27]计算得到, 定压比热容由偏差函数法计算得到。物性的计算结果与标准数据库[28]进行对比, 最大误差小于 5%。物性详细求解方法可参见文献 [29]。C_{5+}、CC_{5+}、C_nH_{2n-6}、RP-3 的替代模型如表 2 所示。

表 2　混合物物性的替代模型

Table 2　Proxy models for mixture physical properties

混合物	替代模型
C_{5+}[30]	40%的戊烯(C_5H_{10})和60%的己烷(C_6H_{14})
CC_{5+}[30]	50%的环己烷(C_6H_{12})和50%的甲基环己烷(C_7H_{14})
C_nH_{2n-6}[30]	100% 三甲苯(C_9H_{12})
RP-3[31]	19.1% 正癸烷、36.5% 正十二烷、14.5%甲基环己烷和29.9%正丁基苯

通过上述方法可得到 RP-3 及其裂解产物的物性数据库, 在数值计算中调用该数据库可得到相应条件下的物性数据。

1.4　网格无关解验证

采用结构化网格, 对壁面附近的网格进行加密, 保证第 1 层网格高度 y^+<1。网格采用线性增长的方式, 增长率为 1.1。通过改变分支管道径向及流向节点数来调节网格数量。以本文关注的物理量——质量流量偏差作为监视信息, 定义质量流量相对标准差 R:

$$R=\left|\frac{m_1-(m_1+m_2)/2}{(m_1+m_2)/2}\right| \tag{10}$$

式中: m_1、m_2分别为 2 个管道的质量流量。

图 2 为并联双管系统的网格无关解分布。可知, 在网格数达到 1.77×10^6 时, 质量流量相对标准差达到无关解标准, 本文将基于该网格开展研究。

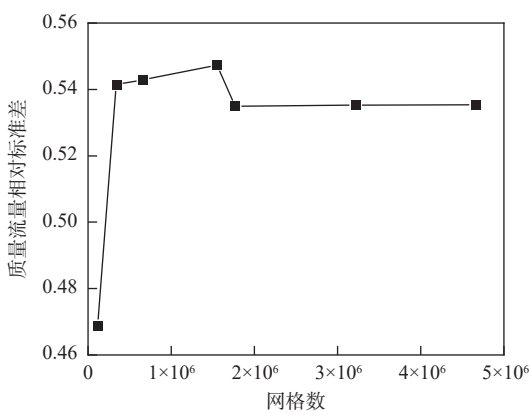


图 2　网格无关解验证

Fig. 2　Mesh independence verification

1.5　数值方法验证

基于 Li 等[31]的超临界航空煤油的流动换热实验数据开展数值方法验证。实验模型为边长为 1.8 mm 的水平方管, 管长为 450 mm。超临界航空煤油以 1 g/s 的速度注入管道中, 入口温度为 673.15 K, 管道出口压力为 3.5 MPa。定义截面平均流体温度 T_f及截面平均裂解度Z为

$$\begin{cases} T_f=\int_A \rho u c_p T\mathrm{d}A \Big/ \int_A \rho u c_p \mathrm{d}A \\ Z=\left(1-\int_A \rho u Y_{\mathrm{fuel}}\mathrm{d}A \Big/ \int_A \rho u \mathrm{d}A\right)\times 100\% \end{cases} \tag{11}$$

式中: dA为管道某横截面上的积分微元面; Y_{fuel}为该微元面上燃油的质量分数; u为该微元面上流体流速。

将沿程壁温、出口处的平均流体温度、出口处的截面平均裂解度与实验值进行比较, 如图 3 所示。可见, 计算结果与实验结果相吻合, 最大误差在 3.2% 以内, 满足本文基于工程设计的精度要求。

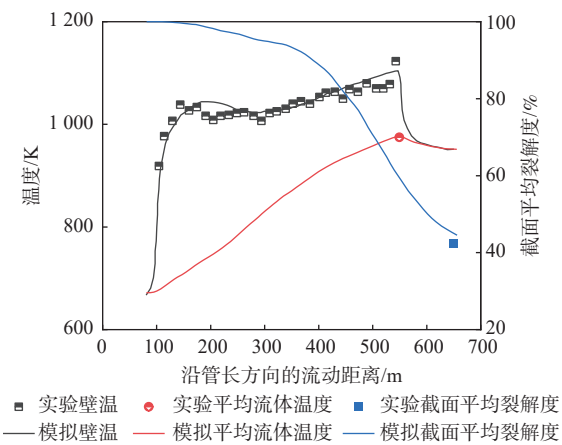


图 3　数值结果与实验结果对比

Fig. 3　Comparison of calculated results with experimental data

2　结果与分析

2.1　非均匀加热工况下的质量流量分配机理

图 4 为 2 个管道之间的质量流量分配随时间的变化。可以看出, 在热流不均匀分布的工况下, 2 个管道之间的质量流量出现了较为显著的差异; 当时间推进到 15 s 时, 2 个管道的质量流量分配达到最终稳定状态, 高热流侧管道流入 23.2% 的冷却工质, 低热流侧管道流入 76.8% 的冷却工质。定义高热流侧管道从初始条件 (t=0 s) 到最终稳定状态 (t>15 s) 的质量流量变化为 Δm, 所需时间为 Δt。在约$0.51\Delta t$的时间完成了$0.9\Delta m$的质量流量的差异化分配过程。2 个管道质量流量从初始的等质量流量

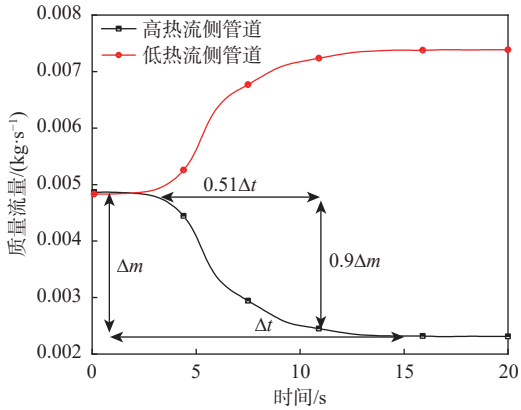


图4 2个管道质量流量分配随时间变化

Fig. 4 Variation of flow distribution with time in two tubes

状态经过3 s的时间缓慢变化后，在3 s<t<6.1 s的时间内，2个管道的质量流量差异迅速变大，且增大的速度也在增加，说明在该时间段，管排系统中存在引起质量流量不均匀分配的正反馈机制，且该机制主导了此阶段的质量流量过程。而在6.1 s<t<15 s的时间段，质量流量差异依然增大，但增大的速度越来越小，说明系统中存在抑制质量流量差异进一步扩大的因素，并随着质量流量差异的增大而逐渐占据主导因素，最终使管排系统达到新的平衡状态。

高热流侧管道流入了较少的冷却工质，因此，2个管道的吸热能力会出现较大的差异。图5为分支管对流换热功率随时间的变化，其是2个管道中流体吸热的能力随时间的变化，表征了2个分支管道内对流换热能力随时间的变化。与质量流量分配相似，在最终平衡状态时，流入23.2%质量流量的高热流侧管道吸收了36.33%的热量。

图5中的小图是高热流侧管道的热功率与低热流侧管道的热功率之比，虚线表示若高热流侧管道与低热流侧管道为绝热连接，即中间面为绝热，则2个管道之间的热功率比为1.05。可以发现，在

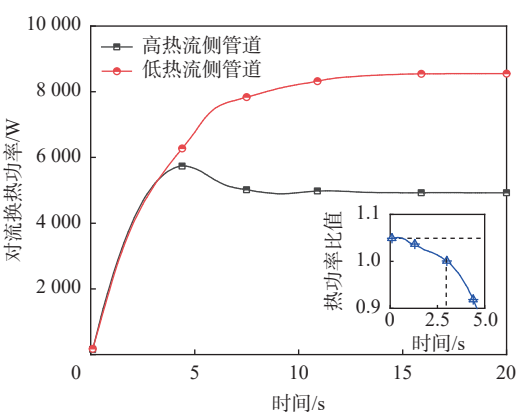


图5 分支管对流换热功率随时间变化

Fig. 5 Variation of convection heat transfer power of branch tubes with time

前2.5 s的时间内，高热流侧管道的热功率大于低热流侧，但随着时间的推进，高热流侧管道吸收热量的占比持续降低。

图6为最终稳定后x=0.55 m处的横截面上的导热热流及温度分布。高热流侧结构温度高于低热流侧结构，因此，在横截面上存在明显的沿y负方向的导热热流。由于高热流侧冷却工质无法满足冷却需求，靠近加热面的结构上，导热热流一方面向冷却通道传递，另一方面由高热流侧的管道流向低热流侧的管道。而固体导热使高热流侧加热面吸收的热量被转移到低热流侧管道，起到抑制高热流侧管道温度过高的作用，防止进一步增大质量流量分配的差异。

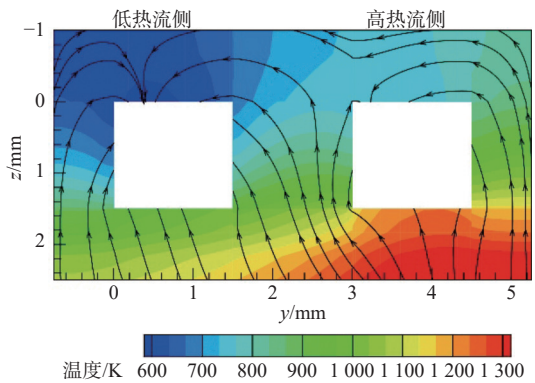


图6 x=0.55 m横截面处固体导热热流分布

Fig. 6 Solid conduction heat flux distribution at the x=0.55 m cross-section

图7表示整个并联管排结构中，通过中间截面的热功率及其在总加热功率中的占比。可以发现，在0～3 s的时间内，通过中间截面的热功率缓慢增加，而在3 s后，流经中间截面的热功率迅速上升，并最终在14.9 s时达到稳定，转移的热功率约占总

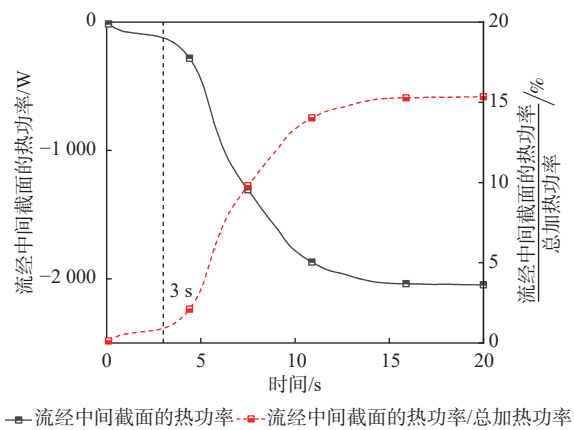


图7 通过中间截面的热功率及其在总加热功率中的占比随时间的变化

Fig. 7 Variation of thermal power passing through the middle section and its proportion in the total heating power with time

加热功率的 15.23%, 热功率达到 2 028 W, 由固体导热转移的热功率在热功率的重新分配过程中起到重要作用。

压力是驱动并联管排系统质量流量分配的主要因素。为分析压力对质量流量分配的影响, 取并联管排 2 个管道的轴线及其延长线为特征线, 图 8 为流动稳定后特征线上的压力分布。可以看出, 低热流侧管道在入口和出口处均出现了明显的压力波动, 而高热流侧只有出口处出现了明显的压力波动。2 个分支管道的出口处压力波动是由于并排管道在汇合处流通面积增大, 流动减速, 压力增加导致的, 而入口处的压力波动是由流道突变引起的。

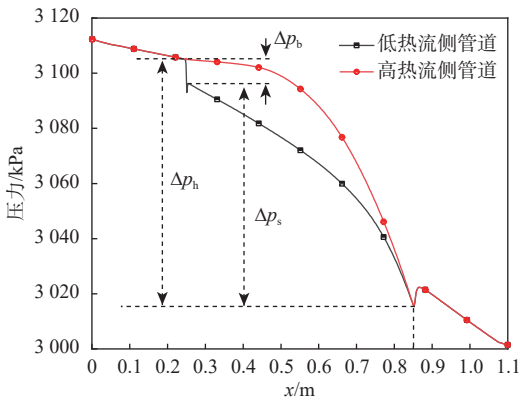


图 8　2 个管道轴线延长线上的压力分布

Fig. 8　Pressure distribution along the axes extension line of the two tubes

由图 8 中的压力分布曲线可得

$$\Delta p_b = \Delta p_h - \Delta p_s \tag{12}$$

式中: Δp_b为 2 个管道内流体在经过分流装置时需要的驱动力之差; Δp_h为高热流侧流体的流动驱动力; Δp_s为低热流侧流体的流动驱动力。此时, 管道质量流量分配已经稳定, 流体驱动力克服流动阻力。因此, Δp_b和Δp_h、Δp_s分别表示分流阻力差和高、低热流侧管道的流体流动阻力。

图 9 为并排管道在入口处的压力分布及流线分布。可以看出, 流体的流动方向在分流口处由于结构的突变会产生多次改变。流入低热流侧通道的流体产生了明显的流动分离区, 并引起了低热流侧管道压力在分流后的骤降。结合图 8, 分流阻力差使流入低热流侧管道的流体需要先克服由于分流装置引起的更多的流动损失, 提高了流体流入低热流侧管道的阻力, 一定程度上降低了不均匀分配的程度, 分流阻力差起到抑制质量流量分配差异增大的作用。

为分析质量流量分配过程中的压力变化, 图 10 展示了 2 个管道流动阻力及分流阻力差随时间的

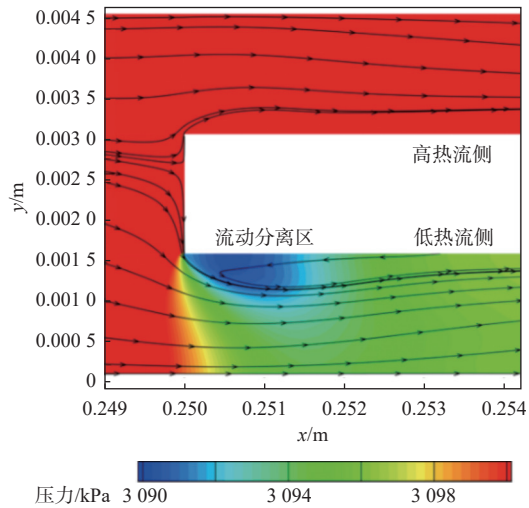


图 9　分流区纵截面处的流线分布及压力分布

Fig. 9　Streamline distribution and pressure distribution of the longitudinal section at diversion zone

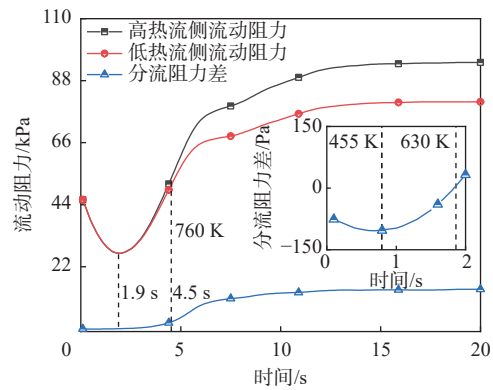


图 10　2 个管道流动阻力及分流阻力差随时间的变化

Fig. 10　Variation of flow resistance in two tube and the difference in flow resistance with time

变化关系。可以发现, 0～1.9 s 的时间内, 高热流侧管道及低热流侧管道的流动阻力均出现了显著的下降, 之后又迅速提升。Zhu 等[32] 研究发现, RP-3 的流动阻力与黏度成正比, 与密度成反比。结合图 11, 0～1.9 s 内流动阻力下降是由于超临界燃油处于较低温度时, 随着温度的升高, 流体的黏度迅速下降, 密度变化相对较小, 在此阶段, 温度升高导致流动阻力下降。由于高热流侧流体在起始阶段加热功率略大, 管内温度略高于低热流侧管道, 因此, 其阻力降的更低, 使流体更多流入高热流侧管道, 导致分流阻力差在起始阶段表现为负值, 即高热流侧的流动阻力大于低热流侧的流动阻力。

0.8 s 时, 分流阻力差由减小变为增大, 出口温度对应为 455 K。结合图 11 可知, 此时密度开始加速下降, 而黏度变化较小, 由于密度降低引起的流体加速使雷诺数增加, 流动阻力增加, 即随着温度的升高, 流动阻力增大。1.9 s 后, 随着管道流体温

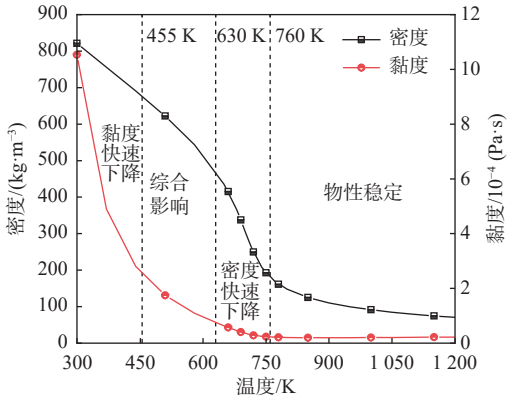


图 11 物性随温度的变化

Fig. 11 Variation of physical properties with temperature

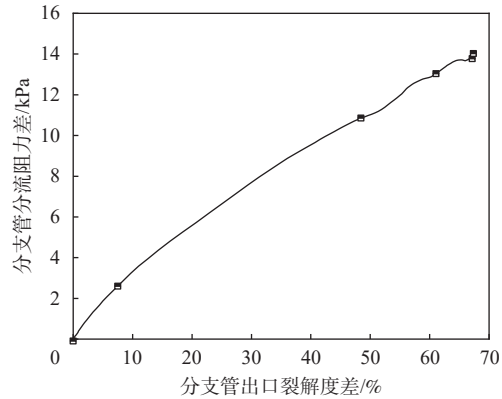


图 13 分支管出口裂解度差与分支管分流阻力差的关系

Fig. 13 The relationships of the conversion difference and the flow resistance difference of branch tubes

度逐渐升高，2 个管道之间开始出现稍大的质量流量差。4.5 s 后，2 个管道之间的分流阻力差迅速拉大，此时对应的出口温度达到 760 K 以上，达到了超临界燃油的裂解温度。这说明碳氢燃料高温下的裂解反应会引起流动阻力的剧烈变化，并导致质量流量不均匀分配。

为分析裂解对质量流量分配的影响，图 12 展示了 2 个管道出口处平均裂解度与时间的变化关系。在最终稳定状态时，2 个管道流体的裂解度差异巨大，低热流侧管道出口处的裂解度只有 3%，而高热流侧管道的裂解度却高达 68.5%。2 个管道在 0～4.5 s 的时间内，平均裂解度均上升，高热流侧管道的平均裂解度增长速度略快；而 4.5 s 后，低热流侧管道出口处平均裂解度开始下降，这是由于更多的流体流入了低热流侧管道，使低热流侧管道流体的热功率不足以继续提高冷却工质的裂解水平。

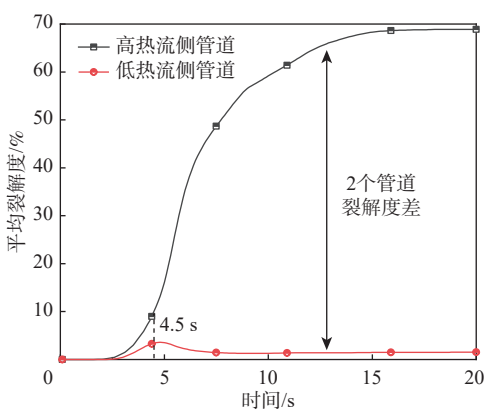


图 12 2 个管道出口处平均裂解度随时间的变化

Fig. 12 Variation of conversion with time at two tubes outlet

图 13 为瞬态模拟中 2 个管道出口处裂解度差与分流阻力差之间的变化关系。可以发现，两者具有很强的正相关性。随着裂解度差的增加，分流阻力差迅速增大，说明裂解是导致 2 个管道出现压力差的重要因素。

为进一步分析裂解对质量流量分配的影响机理，使用 6 个算例，采用工况组 A 的边界条件，分别控制各个算例的物性因素及裂解反应进行仿真。其中，工况 A1 为正常物性工况，考虑裂解反应及裂解反应产生的混合物对各物性的影响；工况 A2～A5 为控制物性的工况，考虑裂解反应，但分别忽略了裂解产物对燃料定压比热容、导热系数、黏度、密度的影响，即分别采用 RP-3 的物性来替代发生裂解后燃油的物性；工况 A6 为忽略裂解反应的工况，即只对 RP-3 进行相同边界条件下的流动换热模拟。表 3 为物性计算方式汇总，其中，RP-3 表示使用 RP-3 的对应物性，忽略裂解产物对冷却工质对应物性的影响（工况 A2～A5）；混合物表示考虑裂解产物对冷却工质的对应物性的影响。

表 3 工况组 A 仿真的物性计算方式

Table 3 Physical property calculation method of group A simulation

工况	裂解	定压比热容	导热系数	黏度	密度
A1	√	混合物	混合物	混合物	混合物
A2	√	RP-3	混合物	混合物	混合物
A3	√	混合物	RP-3	混合物	混合物
A4	√	混合物	混合物	RP-3	混合物
A5	√	混合物	混合物	混合物	RP-3
A6	×	RP-3	RP-3	RP-3	RP-3

图 14 为 6 个控制工况下的质量流量相对标准差的分布。可以发现，相对于实际物性（工况 A1），忽略裂解产物对定压比热容变化的工况（工况 A2）产生更恶化的质量流量偏差，而忽略导热系数（工况 A3）及黏度（工况 A4）变化的工况，质量流量分布情况与真实环境相差较小。忽略密度变化（工况 A5）的工况及忽略裂解反应（工况 A6）的工况，质量流量分配的差异远远小于其他工况，该现象说明密度是引起质量流量分配差异的主要原因，而裂解引起的定压比热容的变化则会抑制质量流量的不均

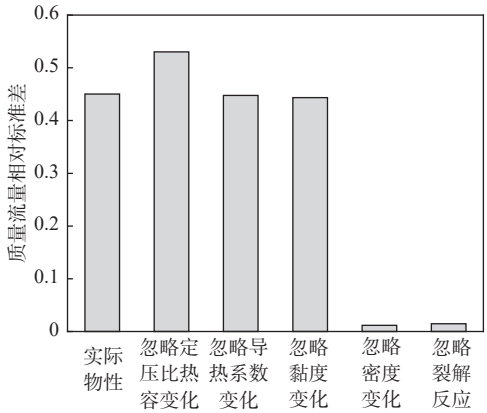


图 14　工况 A1～A6 下质量流量相对标准差的分布

Fig. 14　Relative standard deviation of mass flow rate under condition A1−A6

匀分配。

基于上述分析, 可得到超临界燃油在并联双管中的质量流量分配机理(图 15)。在均衡质量流量的初始条件下, 非均匀热流的边界条件会导致 2 个管道之间出现温度差, 此时, 高热流侧的管道流体温度高。随着时间的推进, 高热流侧的流体温度逐渐上升到拟临界温度附近, 此时, 冷却工质的密度随温度上升会出现较大程度的下降, 导致高热流侧管道的流体密度下降速度大于低热流侧流体, 使高热流侧流体流动阻力的增加速度大于低热流侧。该情况下, 流体偏向于流向低热流侧的管道, 并产生小程度的质量流量分配差异。而高热流侧管道的质量流量减少使高热流侧的冷却工质不足, 温度进一步上升, 高热流侧管道内的超临界燃油先达到裂解温度, 发生化学裂解。而裂解反应产生大量的 H_2、CH_4 等小分子化合物, 使高热流侧的管道密度进一步下降, 流动阻力增加, 从而扩大质量流量分配的差异。而在上述过程中, 固体导热起到将高热流侧管道的热量转移到低热流侧管道的作用, 约束了高热流侧管道温度的上升。而大量流体在流入低热流侧管道的过程中, 其流动阻力远大于流入高热流侧管道的流动阻力, 抑制流体过多流入低热流侧管道。综上, 并联双管系统中, 质量流量分配差异中存在 2 个正反馈因素: 拟临界状态正反馈及裂解反应正反馈。同时, 固体导热及分流阻力差也作为负反馈因素来抑制质量流量分配差异的增大。

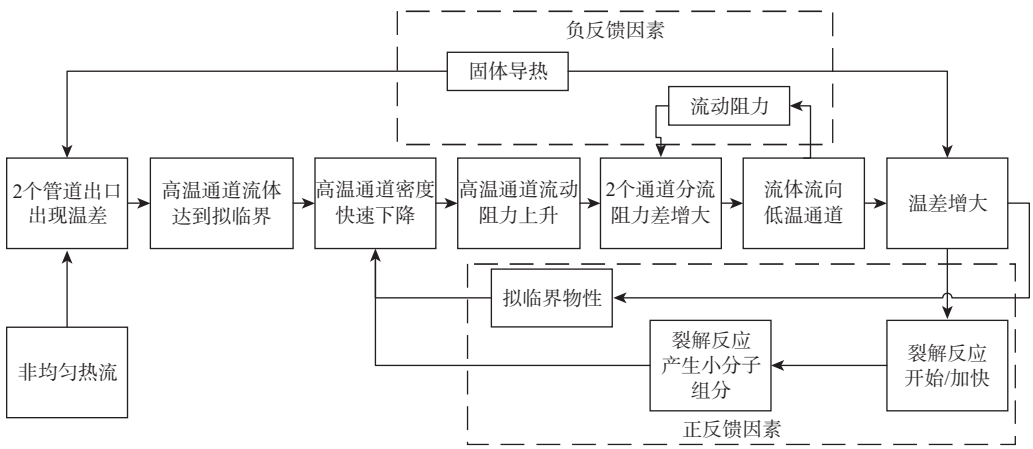


图 15　超临界燃油在并联双管中的质量流量分配机理

Fig. 15　Flow distribution mechanism of supercritical fuel oil in parallel tubes

2.2　非均匀加热边界条件对质量流量分配特性的影响

采用表 1 中工况组 B、C 的边界条件对并联双管模型进行稳态数值模拟, 研究平均热流密度、热流增长率对质量流量分配特性的影响。

图 16 为 2 个管道质量流量与平均热流密度的变化关系。可见, 质量流量分配的差异随着平均热流密度的增加而增大。当平均热流密度小于 3.5 MW/m^2 时, 其增加对质量流量分配特性的影响相对平稳; 当平均热流密度大于 3.5 MW/m^2 时, 其微小变化均会明显影响质量流量分配特性。

结合 2.1 节的分析, 说明虽然具有明显质量流量分配偏差的工况会出现大质量流量分支管内的

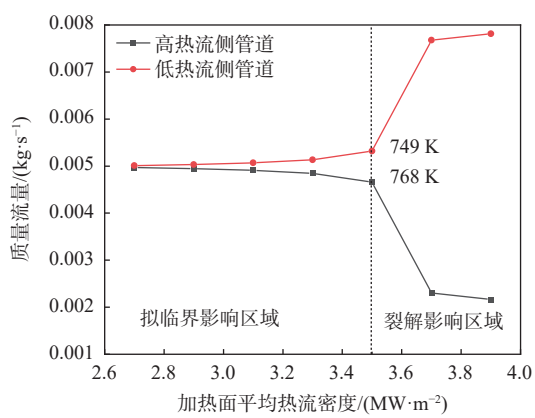


图 16　分支管质量流量随平均热流密度的变化

Fig. 16　Variation of mass flow distribution of branch tubes with average heat flux

超临界燃油裂解度较低甚至不裂解的特性，但裂解反应正反馈被激活的必要条件是加热的总功率能够使全部的冷却工质达到裂解温度以上。定义q_{py}为冷却工质从入口状态到裂解状态所需要吸收的热量，则冷却通道的总加热功率q_{total}应大于q_{py}。因此，在较小加热热流密度工况下（总加热功率不足以使全部的冷却工质达到裂解温度），不会出现小质量流量管道内流体裂解度较高、而大质量流量管道流体温度上升较小的质量流量分配特性，即在较小加热热流密度的工况下，影响质量流量分配的正反馈因素只有拟临界物性正反馈。

采用平均热流密度为3.7 MW/m^{2}的加热工况进一步研究热流增长率对质量流量分配特性的影响。图17为分支管质量流量与热流增长率的关系。可见，2个管道的质量流量差随着热流增长率的增加而略微增大，热流增长率从0.025增大到0.225，质量流量相对标准差增大了11%，变化较为平缓。

图18为2个管道的对流换热功率占总加热功率之比随热流增长率的变化关系。对流换热功率之比是进入高热流侧管道流体的热功率与进入低热流侧管道流体的热功率之比。随着热流增长率的增加，高热流侧管道的加热功率占比也缓慢增大，但高热流侧管道加热功率增加的趋势逐渐降低，说明固体导热在一定程度上缓解了热流增长率增大带来的影响，这有助于降低不均匀加热程度较高工况下高温管道的温度，进一步缓解质量流量分配差异的情况。

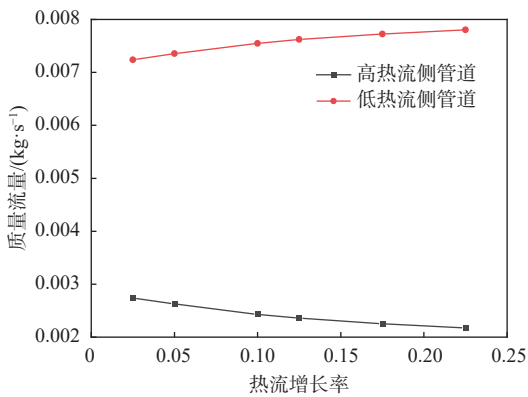


图17 分支管质量流量随热流增长率的变化

Fig. 17 Variation of mass flow rate with heat flux growth rate of branch tubes

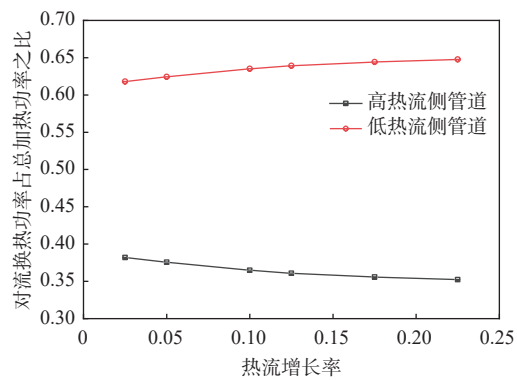


图18 分支管对流换热功率占总加热功率之比随热流增长率的变化

Fig. 18 Variation of the ratio between convection heat transfer power and total heating power with heat flux growth rate of branch tubes

图19为通过中间截面的导热功率及加热面上的热功率差随热流增长率的变化关系。加热面上的热功率差是加热面分别流入高热流侧结构及低热流侧结构的热功率的差值，其与热流增长率呈线性关系。随热流增长率从0.025增加至0.225，由加热面进入的热功率之差从83.15 W增加到677.56 W，而对应的通过中间截面的热功率从1 669.14 W增加到2 385.42 W，远大于非均匀加热边界条件所带来的热功率偏差，这说明非均匀加热对质量流量重新分配结果及热功率再分配的影响相对较小，非均匀加热主要起诱导质量流量出现差异的作用。

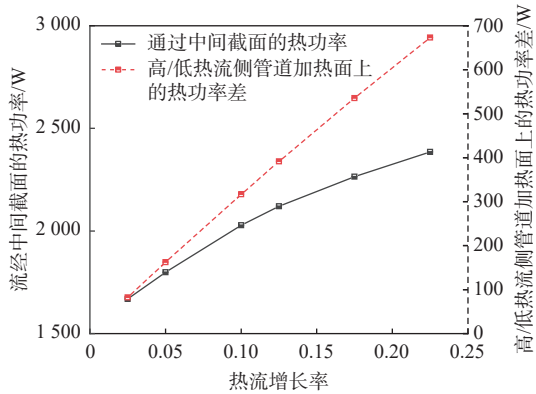


图19 通过中间截面的热功率及加热面的热功率差随热流增长率的变化

Fig. 19 Variation of thermal power passing through the middle section and thermal power difference at heating surface with heat flux growth rate

非均匀的热流在质量流量分配中主要起前期诱导质量流量发生初始分配差异的作用，使某个管道先于另一管道进入拟临界温度范围。而热流增长率对于质量流量分配的正反馈因素、负反馈因素的影响较小，而后两者是影响稳定后质量流量分配情况的主要影响因素，因此，会出现如图17中所示质量流量分配的变化随热流增长率的变化相对较小的现象。

3 结 论

本文采用并联双排管道模型，通过数值研究超

临界燃油在非均匀加热下的质量流量分配机理, 并探究热流增长率和平均热流密度对质量流量分配的影响, 得到以下结论:

1) 质量流量分配特性的影响因素有拟临界物性、裂解反应、固体导热、分流阻力差。其中, 正反馈因素有拟临界物性正反馈、裂解反应正反馈; 负反馈因素有固体导热负反馈、分流阻力差负反馈。

2) 非均匀加热工况下, 超临界燃油流经并联双管系统会产生明显的质量流量分配偏差。其机理为: 高热流侧的管道温度上升较快, 在拟临界物性及裂解反应影响下, 高热流侧管道流动阻力增加大于低热流侧管道, 使流体倾向于流入低热流侧管道; 而高热流侧管道温度较高, 引起向低热流侧管道的导热, 抑制 2 个管道温差增大, 抑制流体过多流入低热流侧管道; 而流体较多地流入低热流侧管道会增大低热流侧管道中分流结构附加的流动阻力, 抑制流体流入低热流侧管道。

3) 激活裂解反应正反馈的必要条件是总加热功率可使全部冷却工质被加热到裂解温度以上, 即 $q_{\text{total}} > q_{\text{py}}$。低于该加热功率的并联双管系统质量流量分配特性仅受拟临界物性正反馈的影响。

4) 保持平均热流密度不变, 质量流量分配对热流增长率变化的敏感性较低, 热流增长率增大 8 倍, 质量流量相对标准差增大 11%。热流增长率在质量流量分配中主要起诱导因素的作用, 对最终的质量流量分配特性影响较小。

参考文献（References）

[1] 张丽静, 刘东升, 于存贵, 等. 高超声速飞行器[J]. 航空兵器, 2010, 17(2): 13-16.
ZHANG L J, LIU D S, YU C G, et al. Hypersonic vehicle[J]. Aero Weaponry, 2010, 17(2): 13-16(in Chinese).

[2] 王振国, 梁剑寒, 丁猛, 等. 高超声速飞行器动力系统研究进展[J]. 力学进展, 2009, 39(6): 716-739.
WANG Z G, LIANG J H, DING M, et al. A review on hypersonic airbreathing propulsion system[J]. Advances in Mechanics, 2009, 39(6): 716-739(in Chinese).

[3] 刘世俭, 刘兴洲. 超燃冲压发动机可贮存碳氢燃料再生主动冷却换热过程分析[J]. 飞航导弹, 2009(3): 48-52.
LIU S J, LIU X Z. Analysis of regenerative active cooling heat transfer process of storable hydrocarbon fuel in scramjet[J]. Winged Missiles Journal, 2009(3): 48-52(in Chinese).

[4] LIU J L, LI H X, LEI X L, et al. An improved model on flow distributions of supercritical pressure water in parallel heated pipes[J]. Applied Thermal Engineering, 2018, 130: 793-803.

[5] BAJURA R A, JONES E H J. Flow distribution manifolds[J]. Journal of Fluids Engineering, 1976, 98(4): 654-665.

[6] CAMILLERI R, HOWEY D A, MCCULLOCH M D. Predicting the flow distribution in compact parallel flow heat exchangers[J]. Applied Thermal Engineering, 2015, 90: 551-558.

[7] JIANG Y G, XU Y X, ZHANG S L, et al. Parametric study on the distribution of flow rate and heat sink utilization in cooling channels of advanced aero-engines[J]. Energy, 2017, 138: 1056-1068.

[8] JIANG Y G, ZHANG S L, FENG Y, et al. A control method for flow rate distribution of cracked hydrocarbon fuel in parallel channels[J]. Applied Thermal Engineering, 2016, 105: 531-536.

[9] JIANG Y G, QIN J, CHETEHOUNA K, et al. Parametric study on the hydrocarbon fuel flow rate distribution and cooling effect in non-uniformly heated parallel cooling channels[J]. International Journal of Heat and Mass Transfer, 2018, 126: 267-276.

[10] 姜俞光. 热裂解型碳氢燃料并联通道流量分配及偏差抑制研究[D]. 哈尔滨: 哈尔滨工业大学, 2018.
JIANG Y G. Investigation of endothermic hydrocarbon fuel flow distribution characteristics and flow deviation control[D]. Harbin: Harbin Institute of Technology, 2018(in Chinese).

[11] ZHANG C, YAO Z L, QIN J, et al. Experimental study on measurement and calculation of heat flux in supersonic combustor of scramjet[J]. Journal of Thermal Science, 2015, 24(3): 254-259.

[12] CHEN Y, LEI Z L, ZHANG T H, et al. Flow distribution of hydrocarbon fuel in parallel minichannels heat exchanger[J]. AIChE Journal, 2018, 64(7): 2781-2791.

[13] MA T, ZHANG P, SHI H N, et al. Prediction of flow maldistribution in printed circuit heat exchanger[J]. International Journal of Heat and Mass Transfer, 2020, 152: 119560.

[14] ZHANG W Q, LI A G, GAO R, et al. Effects of geometric structures on flow uniformity and pressure drop in dividing manifold systems with parallel pipe arrays[J]. International Journal of Heat and Mass Transfer, 2018, 127: 870-881.

[15] JING T T, HE G Q, QIN F, et al. An innovative self-adaptive method for improving heat sink utilization efficiency of hydrocarbon fuel in regenerative thermal protection system of combined cycle engine[J]. Energy Conversion and Management, 2018, 178: 369-382.

[16] CHEN Y, LIU B, LEI Z L, et al. A control method for flow distribution in fuel-cooled plate based on choked flow effect[J]. Applied Thermal Engineering, 2018, 142: 127-137.

[17] TAO Z, CHENG Z Y, ZHU J Q, et al. Effect of turbulence models on predicting convective heat transfer to hydrocarbon fuel at supercritical pressure[J]. Chinese Journal of Aeronautics, 2016, 29(5): 1247-1261.

[18] LI Y, SUN F, XIE G N, et al. Numerical analysis of supercritical n-decane upward flow and heat transfer characteristics in the buffer layer of a vertical tube[J]. Numerical Heat Transfer, Part A: Applications, 2020, 77(3): 247-265.

[19] ZHONG F Q, FAN X J, YU G, et al. Thermal cracking of aviation kerosene for scramjet applications[J]. Science in China Series E: Technological Sciences, 2009, 52(9): 2644-2652.

[20] JIANG R P, LIU G Z, ZHANG X W. Thermal cracking of hydrocarbon aviation fuels in regenerative cooling microchannels[J]. Energy & Fuels, 2013, 27(5): 2563-2577.

[21] 阮波, 孟华. 裂解吸热反应对乙烷超临界传热的影响[J]. 工程热物理学报, 2012, 33(1): 121-124.
RUAN B, MENG H. Effects of endothermic cracking on supercritical heat transfer of ethane[J]. Journal of Engineering Thermophysics, 2012, 33(1): 121-124(in Chinese).

[22] 阮波, 孟华. 碳氢燃料裂解吸热反应及超临界传热现象数值模型

的构建与验证[J]. 航空学报, 2011, 32(12): 2220-2226.
RUAN B, MENG H. Numerical model development and validation for hydrocarbon fuel supercritical heat transfer with endothermic pyrolysis[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(12): 2220-2226(in Chinese).

[23] 徐可可. 航空煤油 RP-3 超临界压力湍流传热和裂解吸热现象的数值模拟研究[D]. 杭州: 浙江大学, 2017.
XU K K. Numerical studies of turbulent heat transfer and endothermic pyrolysis of aviation kerosene RP-3 at supercritical pressures[D]. Hangzhou: Zhejiang University, 2017(in Chinese) .

[24] RUAN B, MENG H, YANG V. Simplification of pyrolytic reaction mechanism and turbulent heat transfer of n-decane at supercritical pressures[J]. International Journal of Heat and Mass Transfer, 2014, 69: 455-463.

[25] XU K K, MENG H. Modeling and simulation of supercritical-pressure turbulent heat transfer of aviation kerosene with detailed pyrolytic chemical reactions[J]. Energy & Fuels, 2015, 29(7): 4137-4149.

[26] YOUNGLOVE B A, ELY J F. Thermophysical properties of fluids. II. Methane, ethane, propane, isobutane, and normal butane[J]. Journal of Physical and Chemical Reference Data, 1987, 16(4): 577-798.

[27] HUBER M L. Properties of fluids, the correlation, prediction and estimation[M]. Cambridge: Cambridge University Press, 1996.

[28] HUBER M L. NIST thermophysical properties of hydrocarbon mixtures database (SUPERSTRAPP)–Version 3.2[M]. Gaithersburg: National Institute of Standards and Technology, 2007.

[29] 程泽源, 朱剑琴. 低裂解度正癸烷物性快速计算方法[J]. 推进技术, 2016, 37(8): 1586-1593.
CHENG Z Y, ZHU J Q. Fast calculation method on physical properties in mild cracking of decane[J]. Journal of Propulsion Technology, 2016, 37(8): 1586-1593(in Chinese).

[30] 程泽源, 朱剑琴, 金钊. 吸热型碳氢燃料 RP-3 替代模型研究[J]. 航空动力学报, 2016, 31(2): 391-398.
CHENG Z Y, ZHU J Q, JIN Z. Study on surrogate model of endothermic hydrocarbon fuel RP-3[J]. Journal of Aerospace Power, 2016, 31(2): 391-398(in Chinese).

[31] LI Z Z, LI Y, ZHANG X W, et al. Coupling of pyrolysis and heat transfer of supercritical hydrocarbon fuel in rectangular minichannels[J]. Chemical Engineering Science, 2022, 247: 116924.

[32] ZHU K, XU G Q, TAO Z, et al. Flow frictional resistance characteristics of kerosene RP-3 in horizontal circular tube at supercritical pressure[J]. Experimental Thermal and Fluid Science, 2013, 44: 245-252.

Mass flow distribution characteristics of supercritical fuel oil in parallel tubes under non-uniform heating conditions

DONG Hao[1], ZHAO Chaofan[2,*], ZHU Jianqin[3,4], CHENG Zeyuan[3]

(1. Chang Guang Satellite Technology Co., Ltd., Changchun 130033, China;
2. Beijing System Design Institute of Electro Mechanic Engineering, Beijing 100854, China;
3. National Key Laboratory of Science and Technology on Aero-Engine Aero-Thermodynamics, School of Energy and Power Engineering, Beihang University, Beijing 100191, China;
4. Tianmushan Laboratory, Hangzhou 310023, China)

Abstract: To study the flow distribution characteristics in parallel tubes under leaner non-uniform heating, a transient simulation of supercritical fuel oil RP-3 flowing through parallel tubes was conducted in this paper, and the influence factors and their mechanisms were studied. The influence of different uniform heating conditions on flow distribution characteristics was investigated by steady-state simulations with average heat flux in the range of 2.7−3.9 MW/m^2 and heat flux growth rate in the range of 0.025−0.225. The results show that the influence factors include pseudo critical state, pyrolysis, heat conduction, and shunting resistance difference. The positive feedback of pseudo-critical state and pyrolysis would increase the difference in flow distribution between the two tubes. The flow distribution discrepancy will be lessened by the negative feedback of the shunting resistance differential and heat conduction. When the average heat flux remained constant at 3.7 MW/m^2, the growth rate of the heat flux increased eight times, which resulted in an 11% rise in the relative standard deviation of mass flow. Non-uniform heating conditions mainly guide the flow distribution deviation and have a slight influence on the final flow distribution characteristics.

Keywords: supercritical fuel oil; parallel pipes; non-uniform heating; flow distribution; mechanism

Received: 2023-11-30; **Accepted**: 2024-04-05; **Published Online**: 2024-05-20 19:03
URL: link.cnki.net/urlid/11.2625.V.20240518.1422.001
Foundation items: National Natural Science Foundation of China (52122604,52306064); Project of Tianmushan Laboratory (TK-2023-C-009)
* **Corresponding author**. E-mail: zhaochaofan@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0827

基于2D-FFT与2D-CFAR的调频引信快速高精度信号处理方法

张安琪[1]，曹荣刚[1,2,*]，周宇[1]，李佳武[1]，曹宇曦[1]，于永滨[1]

(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 063611)

摘 要：针对调频引信在实际应用中谐波定距法定距精度不高的问题，根据快速傅里叶变换(FFT)及恒虚警率(CFAR)理论，设计以二维快速傅里叶变换(2D-FFT)方法与频域二维恒虚警率(2D-CFAR)自适应检测方法为核心的差频信号处理方法。该方法将一维差频信号转换为二维矩阵，通过2D-FFT变换得到二维距离-速度频谱，并将静态杂波滤除，抑制噪声对调频引信的影响，对滤波后的二维频谱使用2D-CFAR提取目标的距离和速度。通过仿真验证所提方法性能，结果表明：所提方法在干信比15 dB的情况下，定距相对误差为0.033，测速相对误差为0.004，能够在10.26 μs内准确地输出起爆信号。所提方法能够满足调频引信与地面目标在高速交会情况下目标距离和速度的准确获取和实时性，并提高调频引信的抗扫频干扰能力。

关 键 词：调频引信；二维快速傅里叶变换；二维恒虚警率；现场可编程门阵列；信号处理

中图分类号：TJ43^{+}4.1

文献标志码：A **文章编号**：1001-5965(2026)03-0837-09

毫米波调频引信是一种基于毫米波技术实现精准定距和目标定位的装置，相比于其他体制的引信，其优点十分明显，在各国都备受重视，研究机构投入巨大力量进行研究和开发[1]。与激光调频引信相比，毫米波调频引信可以在更加复杂的环境下稳定工作，对目标的识别更灵活；与微波调频引信相比，毫米波调频引信具有更高的精度和分辨率，能够在复杂环境下提供更为准确的定距数据。此外，毫米波调频引信在一些特殊应用场景下具有明显优势，如精确火力打击、隐蔽作战等[2]。

毫米波调频引信发射信号的频率随时间变化，当发射信号与接收信号进行混频后，得到的差频信号蕴含了弹目之间的距离信息和速度信息。针对差频信号中的距离信息，国内外学者做了大量研究，目前主要分为3类：①谐波定距法。例如，文献[3]利用差频信号的各次谐波分量与其本振信号进行混频，通过检测多普勒检波信号中的峰值信息即可获取当前的弹目距离。虽然简单且实现便捷，但仍存在一些无法避免的缺点，如对本振频率稳定性的要求较高、对弹目交会的速度有一定限制、易于受干扰信号干扰等。文献[4]对谐波定距法进行改进，提出多谐波联合识别的定距方法，与传统的单谐波定距法相比，提高了调频引信的抗干扰能力，但该方法需要对多次谐波进行比较，运算量大大增加，实时性难以满足需求。②连续测距方法。例如，文献[5]采用魏格纳-维尔分布(Wigner-Ville distribution, WVD)最大值法在频域中估计出毫米波调频引信差频信号的瞬时频率，从而对弹目距离信息进行连续估计，但该方法在信噪比较低时，瞬时频率的估计值和实际值误差很大，导致定距精度很低。文献[6]研究了基于分数阶域瞬时初始频率估计和基于分数阶相关的2种定距方法，这2种方法

收稿日期：2023-12-20；录用日期：2024-03-01；网络出版时间：2024-04-03 19:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20240403.1147.007
*通信作者. E-mail：crgeea@bit.edu.cn

引用格式：张安琪，曹荣刚，周宇，等. 基于2D-FFT与2D-CFAR的调频引信快速高精度信号处理方法[J]. 北京航空航天大学学报，2026，52(3)：837-845. ZHANG A Q，CAO R G，ZHOU Y，et al. Fast and high precision signal processing method for frequency modulation fuze based on 2D-FFT and 2D-CFAR[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：837-845 (in Chinese).

均不依靠谐波获取距离信息，与WVD最大值法相比，具有更高的定距精度，能够在低信噪比情况下精确定距，但该方法需要的运算点数较大，计算量要求高。③多普勒包络提取方法。文献[7]提出向量机方法，对连续波多普勒引信检波输出信号的频谱进行训练和识别，从而实现极高的目标检测率和干扰检测率，但该方法需要大量样本进行向量机学习训练，计算量大、复杂度高，不适用于引信平台。文献[8]对宽带线性调频信号进行了一维距离像研究，多普勒频移使一维距离像产生平移，无法获得精确一维距离像，不能实现距离速度联合估计。

因此，针对调频引信高定距精度、高实时性的需求，设计一种基于二维快速傅里叶变换(two-dimensional fast Fourier transform, 2D-FFT)和二维恒虚警率(two-dimensional constant false alarm rate, 2D-CFAR)的距离和速度联合估计方法。该方法将一维差频信号转换为二维矩阵，通过2D-FFT变换得到二维距离-速度频谱，并将杂波滤除，抑制噪声对调频引信的影响，对滤波后的二维频谱使用2D-CFAR提取目标的距离和速度，实现了在复杂杂波环境下对弹目的距离和速度信息的准确提取和识别，并提高毫米波调频引信的抗扫频干扰能力。

1 调频引信理论模型

图1为毫米波调频引信的理论框图。若调制信号为锯齿波，调制频偏为F_m，调制周期为T_m，调制频率为f_m，载波中心频率为f_c，发射信号、回波信号及差频信号的时频特性如图2所示。

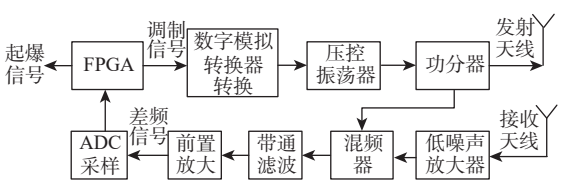


图1 毫米波调频引信理论框图

Fig. 1 Theoretical block diagram of millimeter wave frequency modulation fuze

为了便于分析，将坐标轴x轴零点设置在发射信号与回波信号中心连线处，并将一个周期的信号区分为区间1和区间2，分别对应规则区和不规则区。图2中，$f_T(t)$为发射信号频率，$f_R(t)$为回波信号频率，$f_i(t)$为差频信号频率，τ为回波信号相对于发射信号的时间延迟，记为时延。

时延τ表达式为

$$\tau=[2(R-vt)]/c \tag{1}$$

式中：R为弹目交会距离；v为弹目交会速度；t为时间；c为光速。

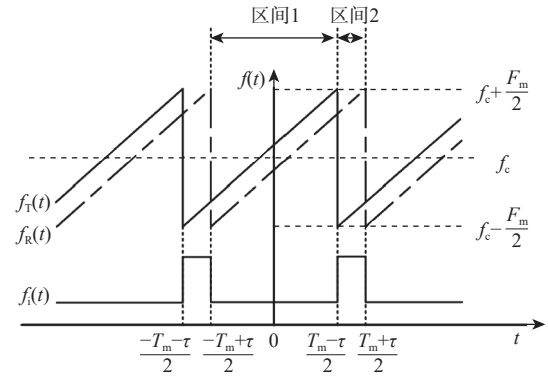


图2 锯齿波调频定距引信信号时频特性曲线

Fig. 2 Time frequency characteristic curve of sawtooth wave frequency modulation fixed distance fuse signal

差频信号表达式[9-10]为

$$s_i(t)=\begin{cases}\cos(2\pi f_c\tau)\cos(2\pi F_m f_m\tau t)-\\ \quad\sin(2\pi f_c\tau)\sin(2\pi F_m f_m\tau t) & \text{区间1}\\ \cos[2\pi F_m f_m(T_m-\tau)t-\\ \quad\pi F_m f_m(T_m-\tau)]\cos(2\pi f_c\tau)+\\ \quad\sin[2\pi F_m f_m(T_m-\tau)t-\\ \quad\pi F_m f_m(T_m-\tau)]\cos(2\pi f_c\tau) & \text{区间2}\end{cases} \tag{2}$$

由式(2)可知，单周期的差频信号由区间1和区间2共同组成。而由于在毫米波调频引信系统的实际应用中，区间1占据了整个周期中很大的比例($T_m\gg\tau$)，因此，在信号分析过程中，主要研究区间1的信号，不规则区是由时延τ引起的。

区间1的多普勒信号频率f_d和差频信号频率f_i表达式分别为

$$f_d=(2f_cv)/c \tag{3}$$

$$f_i=\frac{2F_m}{cT_m}R-f_d \tag{4}$$

当引信以速度v接近地面目标时，将产生多普勒效应。如式(4)所示，弹目相对运动时，差频信号频率与相对静止时相比，减小一个多普勒频率[11]，且差频信号频率中包含弹目距离信息，多普勒信号频率中包含弹目速度信息。因此，可以通过提取多普勒信号频率，并对其做FFT，获取毫米波调频引信弹目交会时的速度，从而实现调频引信的精确定距，使引信的执行级能够准确地输出起爆信号。

2 差频信号处理方法

差频信号处理方法流程如图3所示。可知，差频信号经模数转换器(analog-to-digital converter, ADC)采样后得到一维差频信号。需对一维差频信号进行数据预处理即数据重排，将一维差频信号根据发射信号的采样点数与周期数排列成二维矩阵信号，如图4所示。

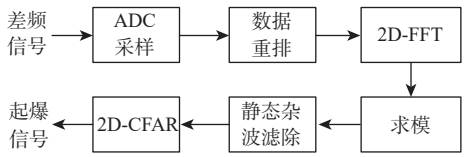


图 3　差频信号处理方法流程框图

Fig. 3　Flow chart of differential frequency signal processing method

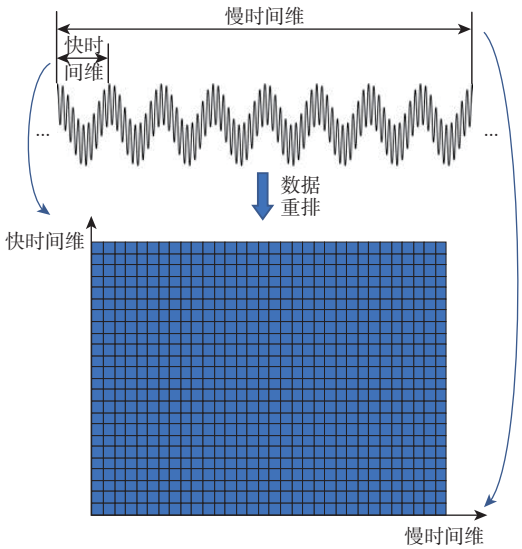


图 4　数据重排流程

Fig. 4　Flow chart of data rearrangement

在慢时间维上对采集到的回波数据进行一次 FFT 变换, 称作距离维 FFT, 可以得到距离信息。通过将差频信号在距离维度上从时域变换为频域, 可以从频谱图中提取不同距离上的目标反射能量分布。对距离维 FFT 后的差频信号矩阵进行静态杂波滤除处理, 去除静态杂波和背景杂波, 提高目标信号的显著性。静态杂波滤除可以消除目标和背景的相位和幅度差异, 并保留目标的相对运动速度信息。在快时间维上进行一次 FFT 变换, 称作速度维 FFT, 可以得到速度信息, 并构建出二维距离多普勒矩阵 (range Doppler matrix, RDM) 并对其求模。通过 2D-CFAR 检测方法, 在 RDM 中搜索目标信号, 并判断其是否符合预设的虚警概率要求, 识别出目标距离速度信息。

2.1　2D-FFT 处理

2D-FFT 是一维 FFT 的推广, 在图像处理中被广泛应用[12-13], 其可将图像从空间域转换到频率域, 通过对频率域图像进行操作, 可以实现很多图像处理任务, 如图像增强、去噪、边缘检测、纹理分析等。同样, 2D-FFT 也可以将二维时间信号转化为二维 RDM。

调制周期 T_{m} 是微秒级, 弹目速度是百米每秒级, 因此, 可以假设弹目距离在一个调制周期内不变, 在相邻调制周期内变化。则在第 n 个调制周期内弹目距离 $R_n = R_0 - v(n-1)T_{\mathrm{m}}$, 其中, R_0 为弹目初始距离。第 n 个调制周期的差频信号经采样和幅度归一化后, 用复信号表示为

$$s(m,n) = \exp\left(\mathrm{j}2\pi\left(\frac{2F_{\mathrm{m}}R_n}{T_{\mathrm{m}}c}\cdot\frac{T_{\mathrm{m}}}{M}m + \frac{2R_n}{\lambda}\right)\right) \tag{5}$$

式中: m 为距离维的采样点数索引; n 为速度维的调制周期数目索引; M 为一个调制周期内的采样点数; λ 为发射信号的波长。

将 N 个差频信号作为列向量, 得到 $M\times N$ 的矩阵, 对该二维矩阵先做行方向上的离散傅里叶变换, 再做列方向上的离散傅里叶变化, 便可得到 RDM:

$$\begin{aligned} F(p,q) = &\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} s(m,n)\exp\left(-\mathrm{j}2\pi\frac{m}{M}p\right)\exp\left(-\mathrm{j}2\pi\frac{n}{N}q\right) = \\ &\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\exp\left(\mathrm{j}2\pi\left((f_{\mathrm{r}}T_{\mathrm{m}}-p)\frac{m}{M} - \right.\right. \\ &\left.\left.\left(f_{\mathrm{d}}T_{\mathrm{m}}+\frac{q}{N}\right)n + \frac{2R_0}{\lambda}\right)\right) \end{aligned} \tag{6}$$

式中: $f_{\mathrm{r}} = 2F_{\mathrm{m}}R_n/(T_{\mathrm{m}}c)$; p 为距离维方向的频率索引; q 为速度维方向的频率索引。

2D-FFT 处理流程如图 5 所示。可以看出, 过程中, 2D-FFT 将时域信号转换为频域信号, 并通过频谱分析获得距离差频频率和多普勒频率, 且距离差频频率蕴含弹目距离信息, 多普勒频率蕴含当前导弹的速度。因此, 通过对差频信号做 2D-FFT 能够实现精准起爆的功能。

2.2　静态杂波滤除

本文中的静态杂波为速度维上的直流分量, 即速度为零的通道。实际应用中, 调频引信天线面临的难题为直流均衡问题, 其会对系统的动态范围产生影响。在引信起爆时, 不关心距离接近零或速度为零的目标, 可以将其抑制掉, 减少不必要的干扰, 便于后续的信号处理。直流分量的产生原因如下。

差频信号除受地面目标的影响, 还与器件的电路有关。当高频回波信号和发射信号作用于混频器时, 由于混频器是非线性元件, 混频后的信号包含直流分量、基波、和频、差频等, 经过带通滤波, 基波与和频信号被滤除。由于器件的直接耦合, 使各级的直流通路相连, 从而使静态工作点相互影响, 导致系统的零点漂移, 因此, 仍有直流分量存在于差频信号中。此外, 在电路设计与应用中, 温度变化是不可避免的因素, 电路元器件对温度变化的高敏感性会导致电路元器件的静态工作点发生温度漂移[14-15]。

本文采用的直流分量去除法为相量均值相消

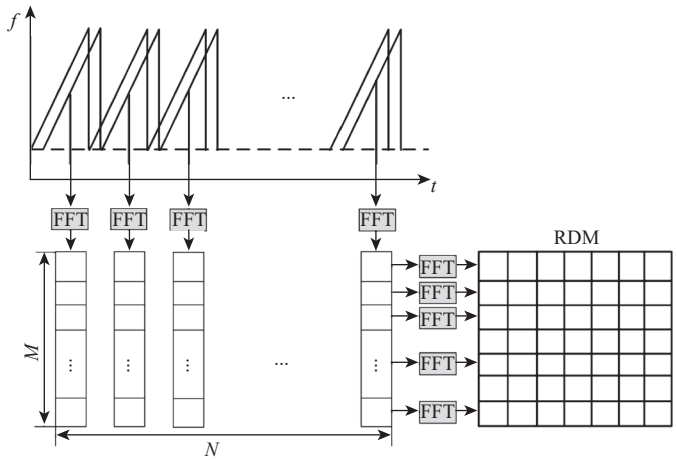


图5 2D-FFT距离和速度提取方法

Fig. 5 2D-FFT distance velocity extraction method

法，其实现原理为：静止目标到毫米波调频引信天线的距离不变，每束回波信号上静止目标的时延也不变。首先，对一帧回波信号进行累加求和，将累加的结果除以一帧的回波信号数目，得到参考回波信号。然后，每周期回波信号与参考回波信号相减，得到目标回波信号。其中，参考回波信号的表达式为

$$C[m]=\frac{1}{N}\sum_{n=1}^{N}X(m,n) \tag{7}$$

式中：$X(m,n)$为数据重排后的二维时域信号经过距离维FFT处理后的信号。

相量均值相消法的表达式为

$$Y(m,n)=X(m,n)-C[m] \tag{8}$$

2.3 2D-CFAR检测

对距离维和速度维的差频信号进行直流分量去除后，需要检测RDM矩阵中的目标信息。本文提出2D-CFAR进行处理，图6为目标检测方案。



图6 2D-CFAR检测的信号处理步骤

Fig. 6 Signal processing steps for 2D-CFAR detection

引信接收机的平均噪声电平通常是未知的，需要先对RDM幅值平方后的矩阵进行噪声估计，噪声估计后的电平称为阈值H。

$$a=L(P_{\mathrm{f}}^{-1/L}-1) \tag{9}$$

$$H=a\left(\frac{1}{L}\sum_{i=1}^{L}S_i\right) \tag{10}$$

式中：a为比例因子；L为参考单元数目，本方法中的差频信号通过ADC采样得到，每个采样点所得到的信号值即为一个单元；S_i为第i个参考单元信号强度；P_{f}为虚警概率。

目标的检出率与虚警概率和信噪比均成正相关。而对地调频引信作用距离近，回波信号功率大，且差频信号经过静态杂波滤除后噪声也会减小，即差频信号有较高的信噪比。在较高信噪比下，本文虚警概率较小，为恒定值10^{-5}。

信号检测部分通过待检测单元的信号值S与阈值H进行比较，从而实现目标判别。即信号值S大于或等于阈值H，判断为目标，否则，判断为噪声。

2D-CFAR检测是一种基于2个维度做CFAR检测的目标检测方法，其可以同时处理待检测矩阵的行列2个维度[16]。以本文介绍的RDM矩阵为例，通过2D-CFAR寻找目标信号，如图7所示。

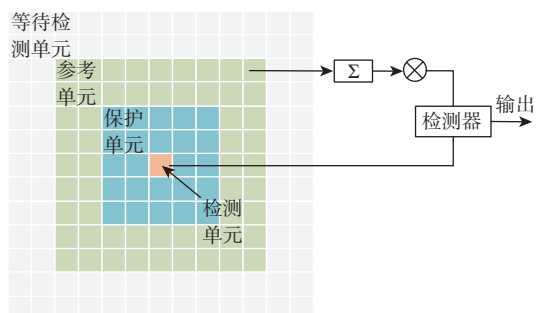


图7 2D-CFAR检测方法

Fig. 7 2D-CFAR detection method

该方法的实施过程如下：首先，需确定检测单元周围的保护单元，保护单元的作用是防止目标能量泄漏到参考单元；然后，计算保护单元周围参考单元的所有信号电平和，将电平和与比例因数相乘后取平均得到阈值；最后，将检测单元的电平值与该阈值进行比较，若检测单元的电平值大于阈值，则判定为目标信号，并将检测单元分配为1，否则，判定为噪声信号，并将检测单元分配为0。

3　差频信号处理方法仿真

3.1　调频引信系统参数

毫米波调频引信系统发射信号的载波中心频率为 60 GHz, 调制频偏为 450 MHz, 调制周期为 10 μs, 弹目交会速度为 100 m/s, 系统定距范围为 9～12 m。由第 1 节可知, 差频信号频率范围为 2.7～3.6 MHz, 多普勒频率为 40 kHz。具体参数如表 1 所示。

表 1　毫米波调频引信的系统参数

Table 1　System parameters of millimeter wave frequency modulation fuse

载波中心频率 f_c/GHz	调制频偏 F_m/MHz	调制周期 T_m/μs	弹目交会距离 R/m	弹目交会速度 v/(m·s^{-1})	FIR带通滤波器通带范围/MHz
60	450	10	9～12	100	2.7～3.6

3.2　MATLAB 仿真结果

搭建 MATLAB 和现场可编程门阵列 (field-programmable gate array, FPGA) 联合仿真平台。通过 MATLAB 构建 3.1 节的毫米波调频引信系统, 并得到差频信号作为仿真数据, 调频引信发射信号、回波信号和差频信号如图 8 所示。分别通过 MATLAB 和 FPGA 对差频信号进行仿真对比, 以验证差频信号处理方法的正确性及 FPGA 实现方法的准确性和实时性。

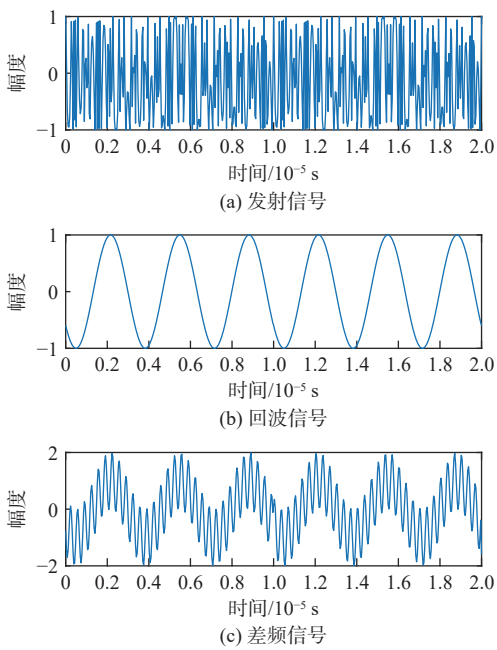


图 8　调频引信发射信号、回波信号和差频信号模型

Fig. 8　The model of transmitting signal, receiving signal, and differential frequency signal for frequency modulation fuze

为验证本文设计的静态杂波滤除方法的有效性, 仿真设置弹目交会距离为 10 m, 且在 1 m 处设置一个静止杂波。由图 9 和图 10 对比可以看出, 1 m 处的静止杂波经静态杂波滤除方法处理后能够被有效消除, 并留下需要识别的运动目标。

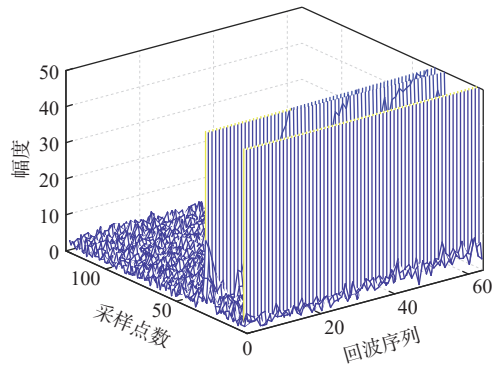


图 9　未经静态杂波滤除的 1D-FFT 处理结果

Fig. 9　1D-FFT processing results without static clutter filtering

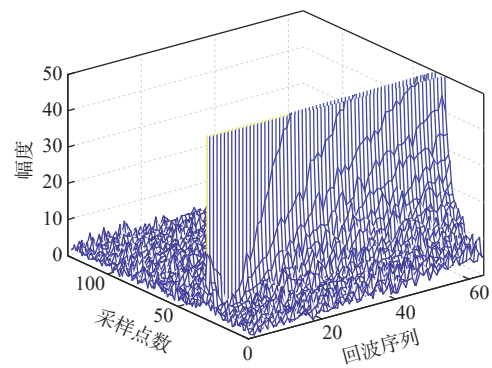


图 10　静态杂波滤除后的 1D-FFT 处理结果

Fig. 10　1D-FFT processing results after static clutter filtering

由图 11 和图 12 可以看出, 2D-FFT 能有效地得到运动目标的距离和速度, 2D-CFAR 能有效检测出运动目标。在 2D-CFAR 检测后的矩阵中查找为 1 的值可以确定目标的位置。经 2D-CFAR 检测出为 1 的值在 RDM 中对应的坐标为(51, 30), 51 为目标在速度维对应的坐标, 30 为目标在距离维对应的坐标, 且速度维的 FFT 点数为 512, 距离维的 FFT 点数为 256。

由第 2 节可以推出该方法距离分辨率ΔR和距离 R 表达式分别为

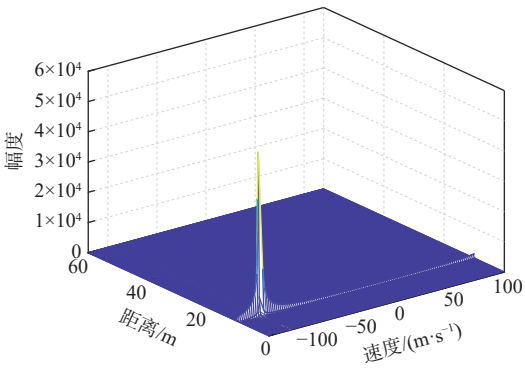
$$\Delta R = \frac{c}{2F_m} \tag{11}$$

$$R = \frac{cTf_i}{2F_m} = \frac{cT}{2F_m}\cdot\frac{F_s}{M}p = \frac{c}{2F_m}\cdot\frac{TF_s}{M}p = \Delta R\cdot p \tag{12}$$

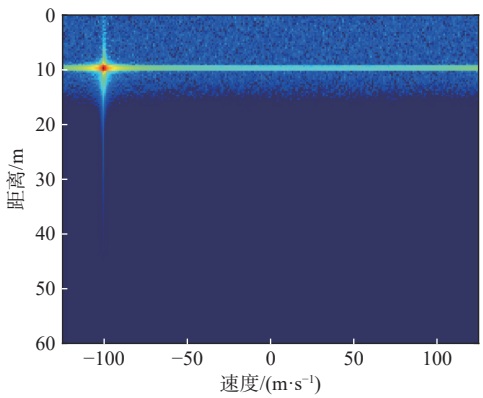
式中: T为采样周期; F_s为采样频率。

由式 (11) 可以计算出距离分辨率ΔR为 0.333 3 m。由于 MATLAB 矩阵中的起始点从 1 开始计算, 实际距离为$R=(n-1)\Delta R$, 即仿真计算出的实际距离为 9.666 7 m, 与设定的目标距离 10 m 相比, 在误差允许范围内, 且定距相对误差为 0.033。

同样, 速度分辨率Δv和速度 v 表达式分别为



(a) 2D-FFT处理后幅度



(b) 2D-FFT处理后距离-速度图

图 11 2D-FFT 处理结果

Fig. 11 2D-FFT processing result

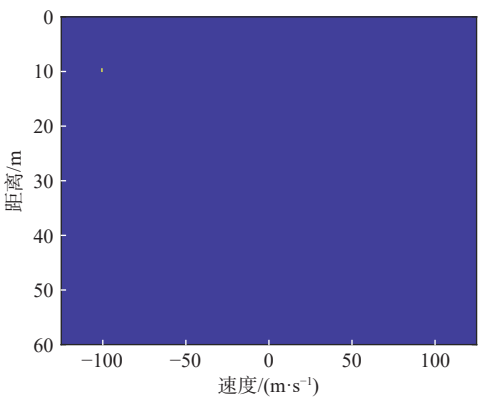


图 12 2D-CFAR 检测处理结果

Fig. 12 2D-CFAR detection processing results

$$\Delta v = \frac{\lambda}{2NT_{\mathrm{m}}} \tag{13}$$

$$v = \frac{\lambda\Delta\phi}{4\pi T_{\mathrm{m}}} = \frac{\lambda}{4\pi T_{\mathrm{m}}} \cdot \frac{2\pi}{N} q = \frac{\lambda}{2NT_{\mathrm{m}}} q = \Delta V \cdot q \tag{14}$$

式中: $\Delta\phi$为调制周期内的多普勒相位变化。

由式 (13) 可以算出速度分辨率ΔV为 0.488 3 m/s。由于速度有正负性, 且当弹目距离接近时, 速度为负值, 零速度通道所对应的速度维点数为 256, 故速度为 0.488 3×(51−256+1)=−99.613 2 m/s。可以看出, 与设定的弹目交会速度 100 m/s 相比, 该方法能够在误差允许范围内检测出弹目交会过程中的相对速度, 测速相对误差为 0.004。

研究表明, 扫频干扰对调频引信的干扰效果较为显著[17], 选择扫频干扰来验证本文提出的差频信号处理方法的有效性。通常, 扫频干扰信号的调制周期远大于调频引信信号调制周期, 使目标回波的差频信号在距离维上受扫频干扰污染较严重。扫频干扰的参数选择如表 2 所示, 讨论 2 个不同干信比下本文方法对目标的检测是否有效, 仿真结果如图 13 所示。

表 2 扫频干扰参数

Table 2 Sweep frequency interference parameters

载波中心频率/GHz	扫频干扰带宽/MHz	干扰调制周期/μs	干信比/dB
60	600	500	15, 20

由图 11 和图 13 对比可以看出, 在干信比为 15 dB 的条件下, 目标回波差频信号在距离维上的底噪被显著抬升, 但本文提出的差频信号处理方法仍能将目标检测出来。当干信比提高到 20 dB 时, 干扰信号完全将目标产生的峰值淹没, 此时, 方法失效。经过仿真发现, 在扫频干扰带宽为 600 MHz、干扰调制周期为 500 μs 的条件下, 当干信比不大于 15 dB 时, 本文提出的差频信号处理方法有效。

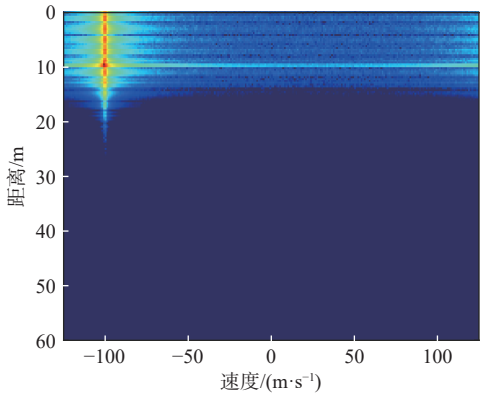
3.3 FPGA 仿真结果

为实现该差频信号处理方法的高实时性需求, 使用 FPGA 对该方法进行硬件加速。将差频信号处理方法通过 Verilog 语言实现, 使用图 8 中的差频信号作为输入, 输入到 FPGA 中, 使用 Modelsim 和 FPGA 联合仿真得到的仿真结果如图 14 所示。可以看出, Verilog 的仿真结果与 MATLAB 一致, 经 2D-CFAR 检测出为 1 的值在 RDM 矩阵中对应的坐标为(51, 30)。图 14(b) 中高亮信号线为仿真输入的差频信号, 图 14(c) 中高亮信号线为 2D-CFAR 检测的输出值。

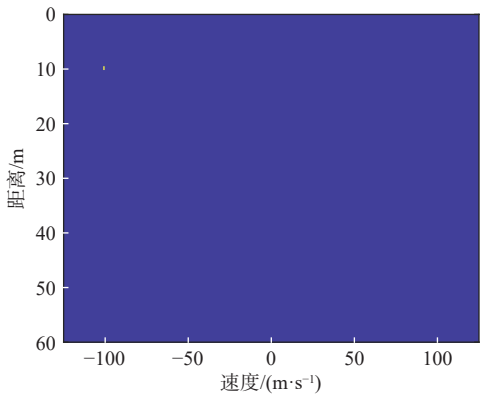
为提高本系统的实时性, 2D-CFAR 参考单元的选取不宜过大, 参考单元为 5×5 矩阵, 且该检测模块工作时钟为 200 MHz。从差频信号的输入到第 1 个检测的值被输出, 共需要经过 2 053(512×4+5) 个工作时钟周期的延迟。在该检测模块的工作时钟下, 2 053 个工作时钟延迟仅为 10.26 μs, 且之后的 2D-CFAR 检测结果可流水线式输出。故该差频信号处理方法使用硬件加速, 在 FPAG 环境下能够满足毫米波调频引信高速飞行时定距的高实时性要求。

3.4 系统资源消耗分析

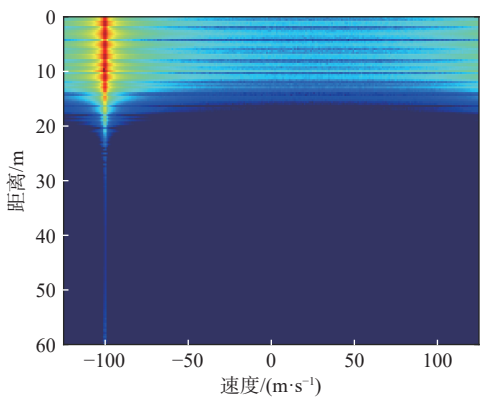
对于 FPGA 上的方法, 芯片的逻辑资源消耗无法忽视。本系统使用的硬件开发平台为 Xilinx 的 FPGA, 其工艺为 Airtix-7、型号为 XC7a200t、封装为



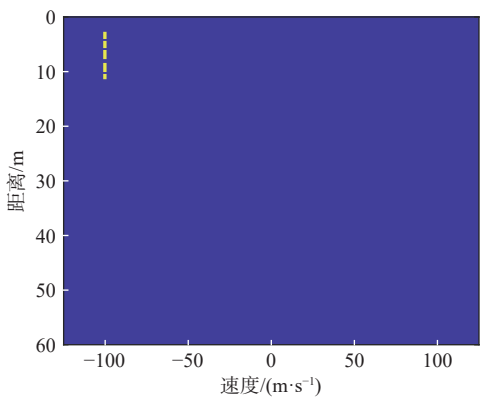
(a) 干信比为15 dB时2D-FFT处理后的距离-速度图



(b) 干信比为15 dB时2D-CFAR检测处理结果



(c) 干信比为20 dB时2D-FFT处理后的距离-速度图



(d) 干信比为20 dB时2D-CFAR检测处理结果

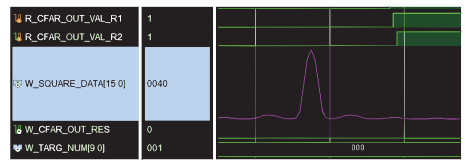
图 13　不同干信比下差频信号处理方法处理结果

Fig. 13　Processing results of differential frequency signals processing method under different jamming-to-signal ratios

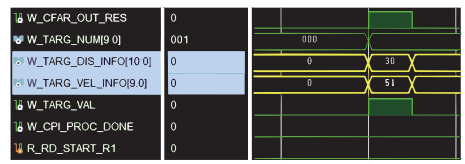
fbg484、速度等级为中等。FPGA 布局布线工具为 Vivado, 版本号为 2 019.1。该方法综合后的资源占



(a) 仿真结果的总体时序



(b) 2D-CFAR的待检测波形



(c) 运动目标在RDM中的坐标

图 14　差频信号处理方法在 FPGA 上运行的仿真结果

Fig. 14　Simulation results of differential frequency signal processing method running on the FPGA

用情况如表 3 所示。

由表 3 可以看出, 在 2D-FFT 点数为 256、512, 2D-CFAR 参考单元为 5×5 矩阵时, Xilinx Xca200t 的 Aritix 芯片资源消耗率很小, 因此, 在此条件下使用 FPGA 平台进行高速测距非常适合。

表 3　FPGA 的片上资源消耗情况

Table 3　The on-chip resource consumption of FPGA

资源	已使用	可用	利用率/%
LUT	13 604	134 600	10.11
LUTRAM	2 040	46 200	4.42
FF	19 270	269 200	7.16
BRAM	68	365	18.63
DSP	14	740	1.89
IO	8	285	2.81

4　结　论

本文提出并仿真验证了基于 2D-FFT 和 2D-CFAR 的调频引信信号处理方法, 得出如下结论:

1) 通过理论分析和仿真计算, 验证了本文提出的毫米波调频引信差频信号处理方法可以实现定距测速功能。

2) 在弹目交会距离为 10 m、速度为 100 m/s 时, 毫米波调频引信的定距相对误差为 0.033, 测速相对误差为 0.004, 均在误差允许范围内, 提高了调频引信系统的定距测速精度。

3) 为了满足毫米波调频引信的高实时性需求,

将差频信号处理方法通过Verilog语言实现，并进行FPGA和Modelsim联合仿真。在2D-FFT点数为256、512，且2D-CFAR的参考单元为5×5的矩阵时，毫米波调频引信定距测速的延时为10.26 μs，能够满足工程需求。

4）为了验证本方法抗干扰的能力，对目标回波信号加上扫频干扰。经仿真验证，在扫频干扰带宽为600 MHz、干扰调制周期为500 μs的条件下，当干信比不大于15 dB时，本文提出的差频信号处理方法有效。

参考文献（References）

[1] 张合, 戴可人. 先进引信技术的发展与展望[J]. 科学通报, 2023, 68(25): 3293-3310.
ZHANG H, DAI K R. Development and prospect of advanced fuze technology[J]. Chinese Science Bulletin, 2023, 68(25): 3293-3310 (in Chinese).

[2] 崔占忠, 宋世和, 徐立新. 近炸引信原理[M]. 3版. 北京: 北京理工大学出版社, 2009: 19-20.
CUI Z Z, SONG S H, XU L X. Principle of proximity fuze[M]. 3rd ed. Beijing: Beijing Institute of Technology Press, 2009: 19-20(in Chinese).

[3] 李鑫. 谐波比较式调频多普勒引信信号处理技术[D]. 南京: 南京理工大学, 2010: 15-17.
LI X. Signal processing technology of harmonic comparison FM Doppler fuze[D]. Nanjing: Nanjing University of Technology, 2010: 15-17(in Chinese).

[4] 梁远, 黄岩, 张玉波, 等. 基于多谐波联合识别的调频引信定距方法[J]. 探测与控制学报, 2022, 44(3): 22-26.
LIANG Y, HUANG Y, ZHANG Y B, et al. FM fuze ranging method based on multiple harmonics associated recognition[J]. Journal of Detection & Control, 2022, 44(3): 22-26(in Chinese).

[5] 袁汉钦, 张向龙. 基于WVD的调频引信瞬时频率估计算法[J]. 海军航空工程学院学报, 2011, 26(3): 301-304.
YUAN H Q, ZHANG X L. Instantaneous frequency estimation of FM radio fuse based on WVD[J]. Journal of Naval Aeronautical and Astronautical University, 2011, 26(3): 301-304(in Chinese).

[6] 岳凯, 郝新红, 栗苹, 等. 基于分数阶傅里叶变换的线性调频引信定距方法[J]. 兵工学报, 2015, 36(5): 801-808.
YUE K, HAO X H, LI P, et al. Research on ranging method for linear frequency modulation radio fuze based on fractional Fourier transform[J]. Acta Armamentarii, 2015, 36(5): 801-808(in Chinese).

[7] 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).

[8] 林慧君, 刘景萍, 张帆, 等. 基于线性调频脉冲压缩引信一维距离像的研究[J]. 微波学报, 2019, 35(S1): 239-242.
LIN H J, LIU J P, ZHANG F, et al. Research on one-dimensional distance image based on chirped pulse compression fuze[J]. Journal of Microwaves, 2019, 35(S1): 239-242(in Chinese).

[9] 孔志杰, 郝新红, 栗苹, 等. 调频引信谐波时序检测抗干扰方法及实现[J]. 北京航空航天大学学报, 2018, 44(3): 549-555.
KONG Z J, HAO X H, LI P, et al. Harmonic timing sequence detection anti-jamming method and its implementation for FM fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(3): 549-555(in Chinese).

[10] 赵惠昌. 无线电引信设计原理与方法[M]. 北京: 国防工业出版社, 2012.
ZHAO H C. Fundamentals and methodology of radio fuze[M]. Beijing: National Defense Industry Press, 2012(in Chinese).

[11] 刘志伟. 毫米波定距射频接收及发射电路设计[D]. 南京: 南京理工大学, 2020: 9-10.
LIU Z W. Design of millimeter wave fixed distance radio frequency reception and transmission circuit[D]. Nanjing: Nanjing University of Technology, 2020: 9-10(in Chinese).

[12] 杜晨阳, 沈俊逸, 刘扬扬, 等. 基于时频分析与相关系数算法的目标探测与距离估计方法研究[J]. 空天防御, 2020, 3(3): 71-77.
DU C Y, SHEN J Y, LIU Y Y, et al. Target detection and range estimation based on time-frequency analysis and correlation coefficient calculation[J]. Air & Space Defense, 2020, 3(3): 71-77(in Chinese).

[13] KIM B, JIN Y, CHOI Y, et al. Low-complexity super-resolution detection for range-vital Doppler estimation FMCW radar[J]. Journal of Electromagnetic Engineering and Science, 2021, 21(3): 236-245.

[14] 畅言, 蔡兴雨, 罗利强, 等. 基于时空相关性的雷达杂波滤除方法[J]. 火控雷达技术, 2020, 49(3): 47-50.
CHANG Y, CAI X Y, LUO L Q, et al. A radar clutter filtering method based on space-time correlation[J]. Fire Control Radar Technology, 2020, 49(3): 47-50(in Chinese).

[15] 杨国成, 余慧敏. 基于N次峰值捕捉的超宽带雷达生命体征检测[J]. 电子测量与仪器学报, 2020, 34(11): 204-210.
YANG G C, YU H M. Vital sign detection of ultra-wide band radar based on N peaks capture[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(11): 204-210(in Chinese).

[16] KHALDI F, SOLTANI F, BAADECHE M. Fuzzy CFAR detectors for MIMO radars in homogeneous and non-homogeneous Pareto clutter[J]. Journal of Communications Technology and Electronics, 2021, 66(1): 62-69.

[17] 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).

Fast and high precision signal processing method for frequency modulation fuze based on 2D-FFT and 2D-CFAR

ZHANG Anqi[1]，CAO Ronggang[1, 2, *]，ZHOU Yu[1]，LI Jiawu[1]，CAO Yuxi[1]，YU Yongbin[1]

(1. Science of Technology on Electromechanical Dynamic Control Laboratory，Beijing Institute of Technology，Beijing 100081，China；
2. Tangshan Research Institute，Beijing Institute of Technology，Tangshan 063611，China)

Abstract：Aiming at the problem of low distance accuracy of the harmonic distance determination method in the practical application of frequency modulation fuze, a differential frequency signal processing method is designed based on fast Fourier transform (FFT) and constant false alarm rate (CFAR) theory, with a two-dimensional fast Fourier transform (2D-FFT) method and frequency domain two-dimensional constant false alarm rate (2D-CFAR) adaptive detection method as the core. To reduce the impact of noise on the frequency modulation fusion, this approach transforms a one-dimensional differential frequency signal into a two-dimensional matrix, applies 2D-FFT transformation to produce a two-dimensional distance velocity frequency domain, and removes static clutter. The filtered two-dimensional frequency domain is used to extract the distance and velocity of the target using 2D-CFAR. The performance of the system is verified by simulation. According to the research findings, at a jamming-to-signal ratio of 15 dB, the relative error of the velocity measurement is 0.004 and the relative error of the differential frequency signal processing system is 0.033. The system has high real-time performance and can accurately output the initiation signal within 10.26 μs. The proposed method can obtain the target distance and speed accurately and in real time when the frequency modulation fuze and the ground target intersect at high speed, and improve the anti-sweep jamming ability of the frequency modulation fuze.

Keywords：frequency modulation fuze; two-dimensional fast Fourier transform; two-dimensional constant false alarm rate; field-programmable gate array; signal processing

Received：2023-12-20；**Accepted**：2024-03-01；**Published Online**：2024-04-03 19：00
URL：link.cnki.net/urlid/11.2625.V.20240403.1147.007

* **Corresponding author**. E-mail：crgeea@bit.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0848

强夯作用下夯沉量实用计算方法

许国一，冯锦艳，OCHIENG J. OBONGO，于卓奇，童朝霞*
(北京航空航天大学 交通科学与工程学院，北京 100191)

摘 要：强夯夯沉量的准确预测对其设计施工至关重要。我国规范中地基土静荷载沉降计算多采用分层总和法，该方法简便实用，且工程经验丰富。强夯冲击荷载下的竖向动应力分布与静荷载下按弹性理论计算的土中附加应力分布类似。在此基础上，借鉴分层总和法思路，提出一种计算强夯夯沉量的拟静力实用计算方法。通过便携式落锤弯沉仪试验和贯入仪试验，对计算方法中涉及的动荷载特性参数及动回弹模量与压缩模量的关系进行了初步探讨，建议了平均夯锤接触动应力的计算方法。为验证所提夯沉量实用计算方法的有效性，将3个工程实例的实测夯沉量与计算夯沉量进行对比分析，结果表明：所提方法能较好地反映地基土体在不同能级下夯击的沉降特性，对于低、中、高3种不同能级的强夯工程，计算夯沉量与实测夯沉量的相对误差分别为2.28%、2.27%和9.89%。所提强夯夯沉量实用计算方法公式简单，使用方便，可为冲击荷载作用下地基土的沉降计算提供新思路。

关 键 词：强夯；夯沉量；拟静力；分层总和法；实用计算方法
中图分类号：TU433
文献标志码：A **文章编号**：1001-5965(2026)03-0846-07

强夯法目前广泛用于处理碎石土、砂土、低饱和度的粉土、黏性土、湿陷性黄土等，已被证明是一种经济、高效的地基加固方法[1]。其中，强夯接触力和夯后地基沉降一直是工程界关心的重要问题，直接影响强夯加固效果。国内外学者对强夯接触力和夯后地基沉降进行了广泛而深入的研究[1]。钱家欢[2]通过改进Scott等[3]的理论，提出加卸载弹性模型，可计算强夯作用下的冲击应力和夯沉量。孔令伟等[4]在考虑夯锤自重的基础上，结合夯锤的刚体运动方程和成层弹性地基空间轴对称动力问题的传递矩阵法，导出了强夯的边界接触应力和沉降在时域上的解析解。刘汉龙等[5]将强夯冲击荷载形式简化为标准正弦函数，推导了一种计算强夯作用下土体夯沉量的方法。强夯加固地基是一个复杂的夯锤与土体动力相互作用过程。也有不少学者从工程应用的角度出发，绕开复杂的动力相互作用过程，按照拟静力的思路提出各具特色的强夯夯沉量计算方法。例如，白冰[6]将强夯冲击荷载作用过程分解为多级不同时间发生的等幅均布荷载，采用拟静力方法预估强夯荷载作用瞬间饱和土因孔压消散引起的固结变形。罗嗣海等[7]从功能原理出发，建立了一种无黏性土强夯加固程度定量估算的拟静力分析方法。

与已有学者拟静力思路不同，本文基于国内外强夯的已有研究成果，通过分析强夯夯锤接触动应力和土中动应力的特点，借鉴静荷载作用下地基沉降量计算的分层总和法思路，提出强夯夯沉量的一种实用计算方法，并用3个工程实例对其有效性进行初步验证。

1 强夯作用下夯沉量计算

1.1 强夯夯锤接触动应力和土中动应力的特点

强夯作用下，夯锤与土体接触动应力的发展过

收稿日期：2024-01-02；**录用日期**：2024-03-15；**网络出版时间**：2024-04-04 07:48
网络出版地址：link.cnki.net/urlid/11.2625.V.20240402.1923.003
基金项目：国家重点研发计划(2023YFB2603605)
*通信作者. E-mail：tongzx@buaa.edu.cn

引用格式：许国一，冯锦艳，OCHIENG J. OBONGO，等. 强夯作用下夯沉量实用计算方法[J]. 北京航空航天大学学报，2026，52(3)：846-852. XU G Y，FENG J Y，OCHIENG J. OBONGO，et al. Practical calculation method for settlement under dynamic compaction[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：846-852 (in Chinese).

程及在土中产生的动应力特点是正确分析和计算强夯夯沉量的首要条件。图 1 给出了采用不同方法获得的强夯夯锤接触动应力时程曲线, 曲线图改编自白冰[6] 对不同夯锤接触动应力时程曲线的对比分析, 主要用于分析冲击荷载作用波形的特点, 故未设置纵轴刻度。其中, 曲线 1 和曲线 2 是裘以惠等[8] 现场实测的湿陷性黄土强夯夯锤接触动应力时程曲线, 曲线 1 对应的土层比较干硬, 含水量较低, 应力波为一个尖峰, 曲线 2 对应的土层含水量较大, 应力波也为一个尖峰, 但衰减较慢, 应力波作用时间约为 0.1 s; 曲线 3 为 Thilakasiri 等[9] 根据室内模拟试验测得的夯锤接触动应力时程曲线; 曲线 4 为钱家欢等[10] 根据室内动力固结试验仪测定的动应力时程曲线; 曲线 5 为孔令伟等[4] 通过计算公式模拟的结果; 曲线 6 为蒋鹏等[11] 根据大变形数值分析计算的强夯夯锤接触动应力时程曲线。分析图 1 可知, 强夯作用下夯锤接触动应力时程曲线呈典型的单脉冲尖峰模式, 没有第 2 峰出现, 可大致分为动应力快速增加到峰值的加载和由峰值降为零的卸载 2 个阶段, 加荷持续时间小于卸荷持续时间。

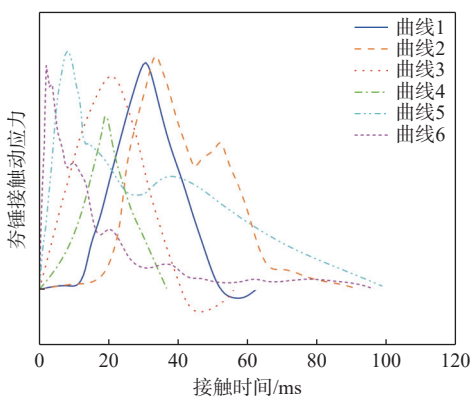


图 1　夯锤接触动应力时程曲线

Fig. 1　Hammer contact dynamic stress-time history curves

强夯作用下的夯沉量与夯锤接触动应力在土体中产生的竖向动应力大小与分布密切相关。Chow 等[12] 通过对比测得的夯锤中心线竖向动应力分布及静荷载作用下弹性理论解发现, 强夯作用下竖向动应力分布与静荷载作用下弹性理论计算的土中附加应力分布类似。这为本文借鉴静荷载作用下土体沉降计算的分层总和法思路, 提出强夯夯沉量实用计算方法提供了坚实基础。

1.2　强夯夯沉量拟静力实用计算方法

我国《建筑地基基础设计规范》(GB 50007—2011)[13] 中, 静荷载作用下地基沉降量的计算采用各向同性均质线性变形体理论的分层总和法, 该方法简单实用, 并积累了丰富的经验。地基最终沉降量的计算如下:

$$s=\psi_s\sum_{i=1}^{n}\frac{p_0}{E_{si}}(z_i\bar{\alpha}_i-z_{i-1}\bar{\alpha}_{i-1}) \tag{1}$$

式中: s为地基最终变形量; ψ_s为沉降计算经验系数; n 为地基变形计算深度内的土层数; p_0为准永久组合作用下基础底面附加压力; E_{si}为基础底面下第 i 层土的压缩模量; z_i、z_{i-1}分别为基础底面至第 i 层、i–1 层土底面的距离; $\bar{\alpha}_i$、$\bar{\alpha}_{i-1}$分别为基础底面计算点至第 i 层、i–1 层土底面范围内的平均附加应力系数。

基于强夯作用下竖向动应力分布与静荷载作用下弹性理论解类似及静荷载作用下地基沉降量计算公式(1)的形式, 并考虑: ①夯锤与土体的接触动应力 p_d为随时间变化的单脉冲模式, 采用将动荷载拟静力的思路, 并考虑 p_d的动荷载特性, 引入平均接触动应力 $\overline{p}_d$及动荷载特性参数 λ; ②土的动回弹模量 E_{vd}是反映地基土动荷载作用下抗变形能力的主要力学指标, 故强夯作用下夯沉量的计算采用土的动回弹模量 E_{vd}替代压缩模量 E_s。基于以上考虑, 提出强夯作用下夯沉量的拟静力实用计算公式:

$$s=\sum_{i=1}^{n}\frac{\overline{p}_d}{\lambda E_{vdi}}(z_i\bar{\alpha}_i-z_{i-1}\bar{\alpha}_{i-1}) \tag{2}$$

强夯冲击荷载持续时间短, 仅为几十毫秒, 且动应力时程曲线为单脉冲尖峰模式, 考虑强夯对有效加固深度范围内的土体有较好的密实作用, 因此, 基于实用考虑, 式(2)强夯夯沉量计算深度采用《建筑地基处理技术规范》(JGJ 79—2012)[14] 中的有效加固深度, 有效加固深度以外土体的变形忽略不计。

1.3　实用计算方法参数分析

强夯夯沉量的拟静力实用计算公式(2)涉及 3 个新变量: 平均接触动应力 $\overline{p}_d$、动荷载特性参数 λ及土的动回弹模量 E_{vd}。

1.3.1　平均接触动应力近似计算

在强夯夯锤冲击过程中, 夯锤与土体的接触动应力随时间变化, 分布也不均匀。从实用考虑, 假定夯锤为刚体, 锤底接触动应力均匀分布, 且夯锤与土体接触为光滑接触, 利用冲量守恒, 有

$$(mg-\overline{p}_dA)T=m\Delta v\Rightarrow\overline{p}_d=\frac{mg-m\Delta v/T}{A} \tag{3}$$

式中: m为夯锤质量; g 为重力加速度; Δv为夯锤冲击土体过程中的速度改变量; T为冲击荷载历时; A为夯锤锤底面积。用式(3)计算夯锤平均接触动应力 $\overline{p}_d$时, 夯锤为自由落体运动, 夯锤冲击土体后

回弹量很小，可忽略不计，方便利用自由落体速度公式计算Δv。式(3)的另一个关键参数是冲击荷载历时T。在图1所示的强夯夯锤接触动应力时程曲线中，加载段均为快速增加的上升段，而卸载段却形式各不同。考虑到卸载段的时间比加载段的时间长，而时间短的加载段对最终的沉降有重要影响，为获得有效的强夯冲击荷载作用历时，可利用冲量守恒定律，将多种不同形式的夯锤接触动应力时程曲线归一化为与原时程曲线中峰值动应力p_{dmax}相等的半正弦函数形式，如图2所示，获得冲击荷载等效历时$\bar{t}$，即

$$\bar{t}=\frac{\pi\int_0^T p(t)\mathrm{d}t}{2p_{\mathrm{dmax}}} \tag{4}$$

利用式(3)计算$\bar{p}_{\mathrm{d}}$时，参数T用式(4)计算的$\bar{t}$代入进行计算。若接触动应力时程曲线无法获得，初步估算可采用吴忠怀[15]和罗嗣海[16]等建议的40 ms进行计算。

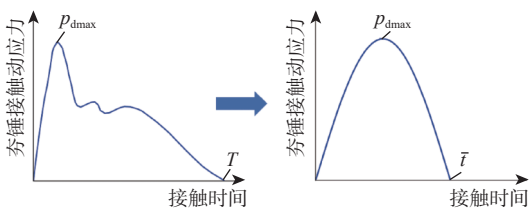


图2 荷载半正弦形式简化

Fig. 2 Simplified half-sine load profile

1.3.2 动荷载特性参数

便携式落锤弯沉仪(portable falling weight deflectometer, PFWD)可以测量冲击荷载作用下场地沉降和动回弹模量，试验操作简单，数据可信度高。为探究动荷载特性参数λ的大小，采用PFWD对北京航空航天大学校内一处较均匀的黏土场地进行落锤弯沉试验。选取6个试验点，每个试验点进行3次平行试验，测得的夯沉量s和动回弹模量E_{vd}如表1所示。

同时，PFWD技术参数有：落锤质量m为10 kg，落锤下落高度为700 mm，回弹高度为400 mm，夯锤接触动应力时程曲线为半正弦形式，加载历时即冲击荷载等效历时$\bar{t}$为18 ms。将上述参数代入式(3)，可计算夯锤平均接触动应力$\bar{p}_{\mathrm{d}}$=52.5 kPa。PFWD试验地基土的影响深度z为0.5 m，将表1中测得的6组平均夯沉量s、土的动回弹模量E_{vd}、夯锤平均接触动应力$\bar{p}_{\mathrm{d}}$及平均附加应力系数$\bar{\alpha}$代入式(2)，可计算出动荷载特性参数λ非常稳定，在0.548～0.550之间，可取$\lambda=0.55$。

1.3.3 动回弹模量与压缩模量关系

利用贯入仪进行贯入试验。图3为采用的

表1 便携式落锤弯沉仪试验数据

Table 1 Test data of PFWD

实验编号	夯沉量 s_1/mm	夯沉量 s_2/mm	夯沉量 s_3/mm	平均夯沉量 s/mm	动回弹模量 E_{vd}/MPa
1	1.027	1.006	0.996	1.010	22.2
2	0.892	0.866	0.861	0.873	25.7
3	1.010	0.989	1.006	1.002	22.4
4	1.002	1.003	0.975	0.993	22.6
5	1.019	0.990	0.964	0.991	22.7
6	0.941	0.911	0.893	0.915	24.5

WG-VI型智能贯入仪，可根据贯入阻力确定地基土的承载力特征值和压缩模量。同样选取6个试验点、对每个试验点进行3次平行贯入试验。测得的贯入阻力P、地基土的承载力特征值f_{ak}和压缩模量E_{s}如表2所示。图4给出了表1中的动回弹模量E_{vd}与表2中的压缩模量E_{s}的关系曲线。可知，动回弹模量E_{vd}与压缩模量E_{s}之间关系较简单，$E_{\mathrm{vd}}\approx 4.5E_{\mathrm{s}}$。值得指出的是，上述得到的$E_{\mathrm{vd}}\approx 4.5E_{\mathrm{s}}$只是针对本文试验场地，土的动回弹模量$E_{\mathrm{vd}}$和压缩模量$E_{\mathrm{s}}$之间的关系与土的类别、土的状态、应力水平等诸多因素有关，有待进一步深入研究。



图3 WG-VI型智能贯入仪

Fig. 3 Intelligent penetration apparatus WG-VI

表2 贯入仪试验数据

Table 2 Test data of penetration

实验编号	贯入阻力 P_1/N	贯入阻力 P_2/N	贯入阻力 P_3/N	平均贯入阻力P/N	承载力特征值f_{ak}/kPa	压缩模量 E_{s}/MPa
1	33.1	29.8	32.0	31.63	289.1	5.74
2	29.2	25.9	28.1	27.73	256.6	5.19
3	25.3	25.8	31.2	27.43	253.9	5.17
4	28.6	25.4	25.4	26.47	245.7	5.02
5	33.1	31.8	34.8	33.20	301.9	5.83
6	30.8	32.1	30.0	30.97	283.7	5.65

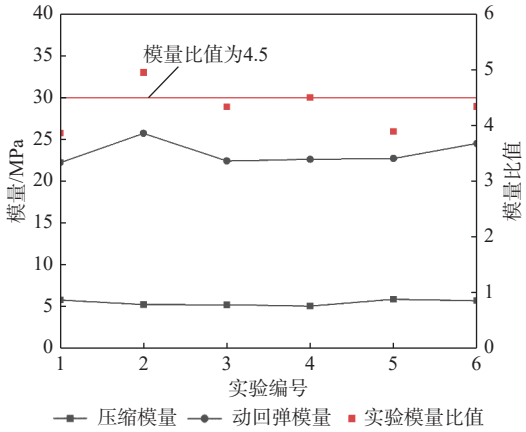


图 4　动回弹模量E_{vd}与压缩模量E_s的关系曲线

Fig. 4　Relationship curves between dynamic rebound modulus E_{vd} and compression modulus E_s

2　工程实例验证

为校核本文提出的强夯夯沉量实用计算方法的有效性, 用 3 个不同夯击能的强夯工程实例进行验证, 并与其他文献方法进行对比。

2.1　秦皇岛煤码头堆场工程

秦皇岛煤码头堆场工程[17] 现场为人工回填细砂, 重度为 19.1 kN/m^3, 变形模量 E_0=6 377 kPa, 锤重 98 kN, 落距为 13 m, 锤底面积为 4.0 m^2, 平均夯击能为 325 kN·m/m^2。夯击后的压缩模量与夯击次数 N 的关系经验公式为 $E_s = E_0N^{0.516}$。取泊松比 μ 为 0.33, 根据 $E_0 = \left[1-\left(2\mu^2/1-\mu\right)\right]E_s$[18], 可算得每次夯击后的压缩模量$E_s$。

根据文献 [4] 对该工程采用数值方法计算各夯击次数对应的夯锤接触动应力时程关系曲线, 如图 5 所示, 利用式(4)可得图 6 所示等效半正弦形式的冲击时程曲线。可知, 除第 1 击的冲击荷载等效历时$\bar{t}$接近 50 ms 外, 其余各击的冲击荷载等效历时$\bar{t}$均在 40 ms 附近, 这与吴忠怀[15] 和罗嗣海[16] 等建议强夯计算时冲击荷载历时采用 40 ms 较为一

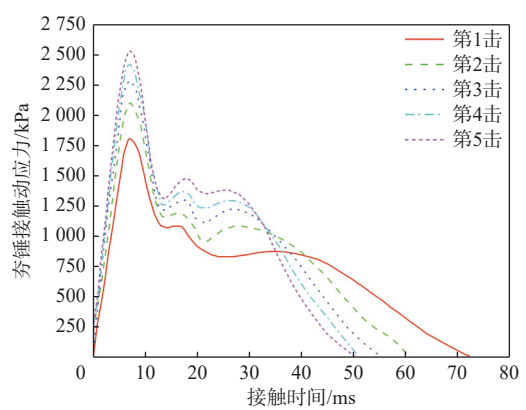


图 5　夯锤接触动应力时程关系曲线[4]

Fig. 5　Hammer contact dynamic stress-time relationship curves[4]

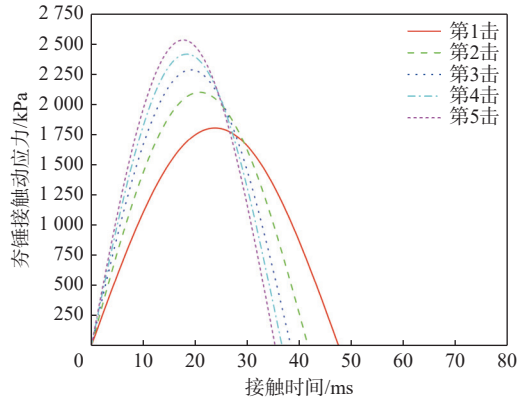


图 6　等效半正弦形式夯锤接触动应力时程关系曲线

Fig. 6　Equivalent half-sine hammer contact dynamic stress-time relationship curves

致[14]。将$\bar{t}$代入式(3), 可计算出每一击的平均接触动应力$\bar{p}_d$。根据《建筑地基处理技术规范》(JGJ 79—2012)[14], 可查得该场地相应夯击能的有效加固深度 z=6 m。

用式(2)计算出各次夯击下夯坑的夯沉量, 如表 3 所示, 表中同时列出了文献 [4] 及边界元法(boundary element method, BEM)[17] 的数值模拟结果。可见, 本文计算的 5 次夯击夯沉量和实测值及 BEM 法[17] 计算的夯沉量较接近, 均小于文献 [4] 的数值计算结果。

表 3　秦皇岛工程计算结果比较

Table 3　Calculation result comparison of Qinhuangdao engineering

夯击次数	压缩模量/MPa	夯沉量/cm		
		文献[4]计算	BEM法[17]计算	本文计算
1	9.4	12.6	11.0	9.4
2	13.4	10.5	8.0	7.5
3	16.6	9.2	7.0	6.5
4	19.2	8.5	6.0	5.9
5	21.6	8.0	5.0	5.4

注: 文献[4]、BEM法、本文计算及实测5次夯击合计夯沉量分别为 48.8、37.0、34.7、37 cm。

2.2　西藏某机场强夯工程

西藏某机场强夯工程[19] 现场地表以粉细砂和植物土为主, 场地浅层分布有 0～2 m 不等的松散粉土地层, 粉土层压缩模量为 5.5 MPa。以 Q1 区第 1 遍点夯为例, 夯击能级为 2 000 kN·m, 夯锤底面积为 3.79 m^2, 夯锤质量为 17.8 t, 夯锤落距约为 11.2 m。夯击后的压缩模量与夯击次数关系采用经验公式 $E = E_0N^{0.516}$。

根据《建筑地基处理技术规范》(JGJ 79—2012)[14], 可查得该场地相应夯击能的有效加固深度 z=5 m, 强夯夯击次数应满足最后 2 击平均夯沉量不大于 50 mm 的要求。利用式(2)计算出各次夯击下夯坑的夯沉量。图 7 为各次夯击下计算夯沉量与实

测值的对比。可知，本文计算的达到收锤标准的夯沉量为114.90 mm，现场实测夯沉量为112.34 mm，两者相对误差仅为2.28%。

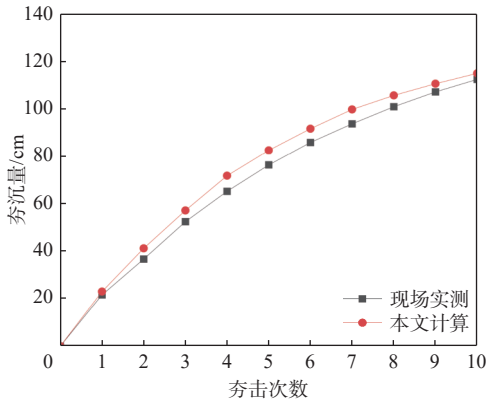


图7 西藏某机场强夯工程计算结果比较

Fig. 7 Comparison of engineering calculation results for an airport dynamic compaction project in Tibet

2.3 某碎石土地基高能级强夯工程

某碎石土地基高能级强夯工程[20]现场地表以碎石土为主，根据Liu等[21]给出的离散元模型宏微观力学参数转换公式，计算得到碎石土层压缩模量为18 MPa。以Ⅰ、Ⅱ区夯击为例，Ⅰ区夯击能级为6 000 kN·m，夯锤底面积为4.91 m^2，夯锤质量为47.94 t，夯锤落距约为13.0 m。Ⅱ区夯击能级为10 000 kN·m，夯锤底面积为4.91 m^2，夯锤质量为61.98 t，夯锤落距约为16.5 m。夯击后的压缩模量与夯击次数的关系经验公式为$E_s = E_0 N^{0.516}$。

根据《建筑地基处理技术规范》(JGJ 79—2012)[14]，可查得夯击能级为6 000 kN·m的有效加固深度z=9 m，强夯夯击次数应满足最后2击平均夯沉量不大于100 mm；夯击能级为10 000 kN·m的有效加固深度z=9.5 m，强夯夯击次数应满足最后2击平均夯沉量不大于200 mm。利用式(2)计算出各次夯击下夯坑的夯沉量。图8和图9分别展示了Ⅰ区和Ⅱ区各次夯击下计算夯沉量与实测值的对比。可知，在高能级工况下，本文计算方法仍然适用，达到收锤标准计算的夯沉量与现场实测结果相近，其中，6 000 kN·m夯击能的相对误差为2.27%，10 000 kN·m夯击能的相对误差为9.89%。

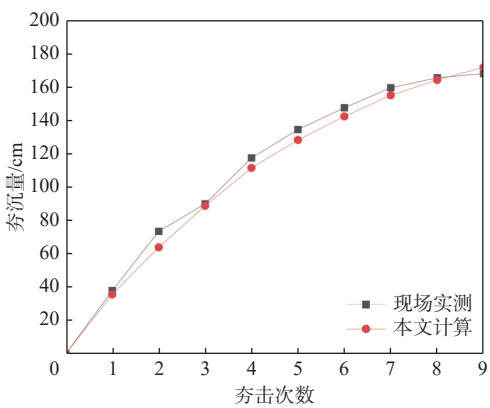


图8 6 000 kN·m工况计算结果对比

Fig. 8 Comparison of calculation results under the 6 000 kN·m operating condition

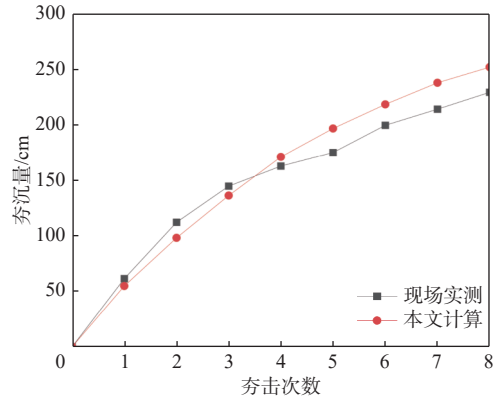


图9 10 000 kN·m工况计算结果对比

Fig. 9 Comparison of calculation results under the 10 000 kN·m operating condition

3 结 论

1）本文在分析夯锤接触动应力和土中动应力特点的基础上，借鉴静荷载作用下沉降量计算的分层总和法思路，提出强夯夯沉量拟静力实用计算方法。

2）对实用计算方法中涉及的平均接触动应力$\overline{p}_d$、动荷载特性参数λ及土的动回弹模量E_{vd}与压缩模量E_s的近似关系进行初步探讨。

3）对3个不同能级的工程实例进行数值和试验结果的对比，发现在夯击能为2000，6 000，10 000 kN·m工况下，相对误差分别为2.28%、2.27%、9.89%，初步验证了本文提出的强夯夯沉量拟静力实用计算方法的有效性。

参考文献（References）

[1] 吴帅峰，蔡红，严俊，等. 强夯沉降模型及参数特征分析研究[J]. 振动与冲击，2021，40(7)：112-118.
WU S F，CAI H，YAN J，et al. Dynamic compaction settlement model and parameter characteristic analysis[J]. Journal of Vibration and Shock，2021，40(7)：112-118(in Chinese).

[2] 钱家欢. 土工原理与计算[M]. 2版. 北京：中国水利水电出版社，1996：251-257.
QIAN J H. Geotechnical principle and calculation[M]. 2nd ed. Beijing：China Water & Power Press，1996：251-257(in Chinese).

[3] SCOTT R A，PEARCE R W. Soil compaction by impact[J]. Géotechnique，1975，25(1)：19-30.

[4] 孔令伟，袁建新. 强夯的边界接触应力与沉降特性研究[J]. 岩土工程学报，1998，20(2)：86-92.
KONG L W，YUAN J X. Study on surface contact stress and settle-

ment properties during dynamic consolidation[J]. Chinese Journal of Geotechnical Engineering, 1998, 20(2): 86-92(in Chinese).

[5] 刘汉龙, 高有斌, 曹建建, 等. 强夯作用下接触应力与土体竖向位移计算[J]. 岩土工程学报, 2009, 31(10): 1493-1497.
LIU H L, GAO Y B, CAO J J, et al. Calculation of contact stress and soil vertical displacement under dynamic compaction[J]. Chinese Journal of Geotechnical Engineering, 2009, 31(10): 1493-1497(in Chinese).

[6] 白冰. 强夯荷载作用瞬间饱和土层固结变形计算[J]. 岩土力学, 2003, 24(1): 57-60.
BAI B. Calculation of consolidation deformation of saturated stratum under dynamic compaction loading[J]. Rock and Soil Mechanics, 2003, 24(1): 57-60(in Chinese).

[7] 罗嗣海, 龚晓南. 无黏性土强夯加固效果定量估算的拟静力分析法[J]. 岩土工程学报, 2008, 30(4): 480-486.
LUO S H, GONG X N. Quasi-static analysis for quantitative estimation of improvement effect of cohesionless soil treated by dynamic compaction[J]. Chinese Journal of Geotechnical Engineering, 2008, 30(4): 480-486(in Chinese).

[8] 裘以惠, 郭玉玲. 强夯法加固地基的土体动应力量测[J]. 太原工学院学报, 1984, 15(1): 45-52.
QIU Y H, GUO Y L. The in-situ measurement of dynamic stress in soil mass during heavy teping by dynamic consolidation method in ground improvement[J]. Journal of Taiyuan University of Technology, 1984, 15(1): 45-52(in Chinese).

[9] THILAKASIRI H S, GUNARATNE M, MULLINS G, et al. Investigation of impact stresses induced in laboratory dynamic compaction of soft soils[J]. International Journal for Numerical and Analytical Methods in Geomechanics, 1996, 20(10): 753-767.

[10] 钱家欢, 钱学德, 赵维炳, 等. 动力固结的理论与实践[J]. 岩土工程学报, 1986, 8(6): 1-17.
QIAN J H, QIAN X D, ZHAO W B, et al. Theory and practice of dynamic conslidation[J]. Chinese Journal of Geotechnical Engineering, 1986, 8(6): 1-17(in Chinese).

[11] 蒋鹏, 李荣强, 孔德坊. 强夯大变形冲击碰撞数值分析[J]. 岩土工程学报, 2000, 22(2): 222-226.
JIANG P, LI R Q, KONG D F. Numerical analysis of large deformation impact and collision properties during dynamic compaction[J]. Chinese Journal of Geotechnical Engineering, 2000, 22(2): 222-226 (in Chinese).

[12] CHOW Y K, YONG D M, YONG K Y, et al. Dynamic compaction of loose granular soils: effect of print spacing[J]. Journal of Geotechnical Engineering, 1994, 120(7): 1115-1133.

[13] 中华人民共和国住房和城乡建设部. 建筑地基基础设计规范: GB 50007—2011[S]. 北京: 中国计划出版社, 2012.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Code for design of building foundation: GB 50007—2011[S]. Beijing: China Planning Press, 2012(in Chinese).

[14] 中华人民共和国住房和城乡建设部. 建筑地基处理技术规范: JGJ 79—2012[S]. 北京: 中国建筑工业出版社, 2012: 35.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Technical code for ground treatment of buildings: JGJ 79—2012[S]. Beijing: China Architecture & Building Press, 2012: 35(in Chinese).

[15] 吴忠怀, 吴武胜, 龚晓南. 强夯置换深度估算的拟静力法[J]. 华东地质学院学报, 2001, 24(4): 306-308.
WU Z H, WU W S, GONG X N. Quasi-static model for estimating replacement depth of dynamic replacement[J]. Journal of East China Geological Institute, 2001, 24(4): 306-308(in Chinese).

[16] 罗嗣海. 软弱地基强夯与强夯置换加固效果计算[D]. 杭州: 浙江大学, 1999.
LUO S H. Calculation of flabby ground reinforced by dynamic compaction[D]. Hangzhou: Zhejiang University, 1999(in Chinese).

[17] 帅方生. 边界元法在强夯法中的应用[D]. 南京: 河海大学, 1985: 3.
SHUAI F S. The application of boundary element method to dynamic compaction[D]. Nanjing: Hohai University, 1985: 3(in Chinese).

[18] 钱家欢. 土力学[M]. 南京: 河海大学出版社, 1988.
QIAN J H. Soil mechanics[M]. Nanjing: Hohai University Press, 1988(in Chinese).

[19] 张强, 卓拥. 垫层强夯法加固高水位粉土地基现场试验研究[J]. 民航学报, 2022, 6(6): 44-49.
ZHANG Q, ZHUO Y. Field test study on strengthening high water level silt foundation with cushion dynamic compaction method[J]. Journal of Civil Aviation, 2022, 6(6): 44-49(in Chinese).

[20] 陈南, 尚晓一, 吴彪, 等. 基于可破碎颗粒的高填方碎石土地基强夯颗粒流数值分析[J]. 宁波大学学报 (理工版), 2022, 35(4): 58-64.
CHEN N, SHANG X Y, WU B, et al. Numerical analysis of dynamic compaction particle flow of high fill gravel foundation based on broken particles[J]. Journal of Ningbo University (Natural Science & Engineering Edition), 2022, 35(4): 58-64(in Chinese).

[21] LIU C, XU Q, SHI B, et al. Mechanical properties and energy conversion of 3D close-packed lattice model for brittle rocks[J]. Computers & Geosciences, 2017, 103: 12-20.

Practical calculation method for settlement under dynamic compaction

XU Guoyi，FENG Jinyan，OCHIENG J. OBONGO，YU Zhuoqi，TONG Zhaoxia*

(School of Transportation Science and Engineering，Beihang University，Beijing 100191，China)

Abstract：Accurately predicting the compaction settlement under the impact load of dynamic compaction has an important impact on the design and construction of dynamic compaction. The calculation of subgrade soil settlement under static load generally adopts the layered summation method in China. The layered summation method is simple and practical with rich engineering experience. According to the analysis of the dynamic stress in the soil-hammer contact, the additional stress distribution in the soil determined by the elastic theory under static load is comparable to the vertical dynamic stress distribution under the impact load of dynamic compaction. Based on the layered summation method, a practical calculation method for the compaction settlement of dynamic compaction is proposed. By using a portable drop hammer deflection test and a penetrometer test, a preliminary discussion was conducted on the dynamic load parameters involved in the calculation method, as well as the relationship between the dynamic rebound modulus and compression modulus. Meanwhile, a calculation method for the average contact dynamic stress of the hammer was provided. Lastly, a comparison between the measured and estimated settlement of three engineering examples was carried out in order to confirm the efficacy of the suggested practical calculation approach for compaction settlement. The results show that this method can effectively capture the settlement characteristics of subgrade soil under different energy levels of compaction. For dynamic compaction projects with low, medium, and high energy levels, the errors between calculated settlement and measured settlement are 2.28%, 2.27%, and 9.89%, respectively. This practical calculation method is simple and convenient to use, which provides a new approach for calculating the settlement of subgrade soil under impact loads.

Keywords：dynamic compaction；settlement；quasi-static force；layered summation method；practical calculation method

Received：2024-01-02；**Accepted**：2024-03-15；**Published Online**：2024-04-04 07：48

URL：link.cnki.net/urlid/11.2625.V.20240402.1923.003

Foundation item：National Key Research and Development Program of China (2023YFB2603605)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0802

双置信度下多特征自适应融合的红外弱小目标实时跟踪

陈家俊[1,2]，李响[2,*]，宋延嵩[2]，董小娜[1,2]

(1. 长春理工大学 光电工程学院，长春 130022；2. 长春理工大学 空间光电技术研究所，长春 130022)

摘 要：针对在红外弱小目标跟踪过程中出现的地面背景杂乱、相似目标干扰及目标微弱等一系列难题，为提高算法在复杂地面背景下跟踪的鲁棒性和实时性，利用卡尔曼滤波预测目标的初始位置，将初始位置设置为感兴趣区域(ROI)的中心。提取ROI区域中目标的灰度(GRAY)特征、梯度(HOG)特征和局部二值(LBP)特征，分别计算滤波响应图，依据3个响应结果的平均峰值相关能量(APCE)和相邻2帧的响应一致性(CFR)获得融合权重，采用自适应加权融合的方式将3个特征的响应结果进行融合，从而估计目标的最佳位置。对目标模型进行更新，并将目标位置作为卡尔曼滤波的量度。针对不同场景的红外地面背景图像序列的实验结果表明：平均距离精度(DP)为0.782，平均重叠精度(OP)为0.731，实时跟踪速度为94.7帧/s，所提算法能够有效提升复杂环境下跟踪的准确性和鲁棒性，整体性能优于对比算法。

关 键 词：红外弱小目标跟踪；核相关滤波；多特征自适应融合；卡尔曼滤波；相邻帧一致性

中图分类号：TP391.4

文献标志码：A **文章编号**：1001-5965(2026)03-0853-11

红外检测与跟踪技术具有灵敏度高、抗电磁干扰能力强、动态范围大和实时性好等优势，被广泛运用在导弹制导、交通监控、医疗诊断、气象分析等领域中[1-2]。在红外制导和红外反导的军事场景中，场景复杂多变，红外目标只占图像中的少量像素，图像细节不明显，大部分纹理信息都以点状出现，导致图像对比度低，经常缺少细节纹理信息，在复杂地面背景情况下容易出现背景杂波干扰、相似目标干扰及目标微弱等一系列难题，这些因素都要求更加高效、鲁棒性和实时性好的红外弱小目标跟踪算法[3-6]。

近年来，判别式跟踪算法在实时目标跟踪领域应用广泛，通过将时域计算转化为频域计算，可以获得显著的跟踪速度，在许多数据集上都取得了良好表现[7]。针对原有最小误差平方和相关滤波(minimum output sum of squared error, MOSSE)[8]和核相关滤波(kernelized correlation filters, KCF)[9]算法中存在负样本带来的大量边界效应[10]的问题，Fan利用高斯过程回归视图构建了一种空间图正则化分层子任务多核互相关跟踪器[11]，Tang等在此基础上提出快速核化相关滤波器，解决了部分边界效应[12]。

上述算法仅能解决简单背景下的跟踪问题，遇到复杂背景时，需要构建时空正则化模型[13]或判别模型和生成模型[14]的混合模型。Peng等提出在滤波器上集成稀疏性约束以减少生成模型中干扰特征的响应[15]。Li等提出一种基于贝叶斯推理的预测方法来处理目标空间分布的时间变化[16]。Tai等通过在子空间内用重构样本替换训练样本来检索目标的信息重建被污染的训练样本[17]。该类算法虽然可以在复杂背景中取得良好的效果，但计算过

收稿日期：2023-12-09；**录用日期**：2024-01-26；**网络出版时间**：2024-03-12 09:16
网络出版地址：link.cnki.net/urlid/11.2625.V.20240311.1451.003
基金项目：国家重点研发计划(2022YFB3902505)；吉林省自然科学基金重大项目(20230301002GX)
*通信作者. E-mail：lixiang@cust.edu.cn

引用格式：陈家俊，李响，宋延嵩，等. 双置信度下多特征自适应融合的红外弱小目标实时跟踪[J]. 北京航空航天大学学报，2026，52(3)：853-863. CHEN J J，LI X，SONG Y S，et al. Real-time tracking of infrared dim-small target with multi-feature adaptive fusion under double confidence[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：853-863 (in Chinese).

程较复杂，实时性较差。因此，如何平衡复杂环境下鲁棒性和实时性的要求仍然是一个难题。

红外弱小目标通常仅有2像素×2像素～5像素×5像素，深度特征表达不明显、目标不清晰的情况下，只能通过表征信息来确认目标状态，而传统特征在红外弱小目标的跟踪中，局部刻画能力和几何不变性有更优越的效果，跟踪过程中，单一特征无法表达目标足够的信息，需要采取梯度(gradient, HOG)特征、灰度(grayscale, GRAY)特征及局部二值(local binary, LBP)特征进行补充。而在复杂地面背景中，存在背景杂乱、相似目标干扰、目标微弱等一系列问题，采用双置信度的方式来衡量不同特征的响应程度，极大抑制了背景干扰和相似目标干扰的问题，提高了目标的特征表达能力。为解决鲁棒性和实时性的平衡，加入目标预定位模块，使用卡尔曼滤波构造感兴趣区域(region of interest, ROI)，将目标锁定在小范围区域中进行跟踪，避免了冗余的搜索区别，在确保跟踪精度的同时，实现了90帧/s的跟踪速度。

针对上述问题，本文在KCF算法[9]的基础上，提出一种核相关滤波算法，基于双置信度来确认权值，将多特征自适应融合，采用卡尔曼滤波优化算法效率，减少不必要的背景干扰。首先，利用卡尔曼滤波对ROI区域进行预测，提高了跟踪效率；其次，提取ROI区域中的灰度特征、梯度特征和局部二值特征分别训练3个核相关滤波器，通过平均峰值相关能量(average peak correlation energy, APCE)和相邻2帧的响应一致性(consistent frame response, CFR)确定不同特征的权重占比；再次，根据权重占比对3个滤波器的结果进行自适应加权融合，从而确定目标位置；最后，通过目标跟踪的置信度来更新跟踪模型，以适应变化的环境条件，并将当前帧所估计的目标位置作为卡尔曼滤波的量度。

1 核相关滤波

核相关滤波目标跟踪算法主要包含训练、检测和更新3个阶段。在视频序列的第1帧，通过矩形框选定跟踪目标，收集矩形框中跟踪目标的需要特征。为确保有效的样本集，需要采用循环矩阵对目标样本进行密集采样，构建不同位置的样本集。通过对样本集及其对应标签的训练，可以得到判别目标相应强度的分类器。因此，在之后的每一帧中，只需要将视频序列中的图像特征输入分类器就可以找到最大响应位置，从而判别目标所在位置，但由于跟踪过程中会出现背景干扰，需要针对每一帧的跟踪结果对分类器和样本集进行更新[18]。

1.1 训练阶段

训练阶段要求核相关滤波算法在初始跟踪过程中提取训练样本，得到分类器函数$f(\boldsymbol{x})=\langle\boldsymbol{\omega},\varphi(\boldsymbol{x})\rangle$，需要在跟踪过程中损失最小，其中，$\boldsymbol{x}$为训练样本，$\boldsymbol{\omega}$为分类器的待求参数，$\varphi(\boldsymbol{x})$为原低维空间中由训练样本构成的非线性映射函数，使子样本空间中的$\boldsymbol{\omega}$的求解问题转化为对偶空间$\boldsymbol{\alpha}$的求解问题。通过岭回归的方法求解$\boldsymbol{\omega}$，建立最小化目标函数模型：

$$Y=\min_{\omega}\left(\sum_{i=1}^{n}(f(\boldsymbol{x}_i)-\boldsymbol{y}_i)^2+\lambda\|\boldsymbol{\omega}\|^2\right) \tag{1}$$

式中：$\boldsymbol{x}_i$为样本数据；$\boldsymbol{y}_i$为回归目标；n为训练样本数；λ为正则项系数，防止分类器过拟合。

将式(1)求偏导可得

$$\boldsymbol{\omega}=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+\lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{2}$$

式中：$\boldsymbol{X}$为由基样本$\boldsymbol{x}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]^{\mathrm{T}}$构造的$n\times n$循环矩阵；$\boldsymbol{I}$为单位矩阵；$\boldsymbol{y}$为每个训练样本$\boldsymbol{x}_i$所对应的样本标签$\boldsymbol{y}_i$的集合。

利用循环矩阵$\boldsymbol{X}$在傅里叶变换中对角化的性质可以很好地进行特征训练，得到特征标签，但在实际跟踪过程中，在图像数据中提取的图像特征往往是非线性问题，因此，需要引入非线性函数核函数，使新的样本$\boldsymbol{x}$在高维特征空间也是线性可分的，从而得到如下分类器：

$$\boldsymbol{\omega}=\sum\boldsymbol{\alpha}_i\varphi(\boldsymbol{x}_i) \tag{3}$$

式中：$\boldsymbol{\alpha}_i$为对偶空间的系数。

可通过一阶求导得到分类器参数$\hat{\boldsymbol{\alpha}}$和对应标签$\hat{\boldsymbol{y}}$的闭式解：

$$\hat{\boldsymbol{\alpha}}=\left(\frac{1}{\hat{\kappa}^{xx}+\lambda}\right)^{*}\odot\hat{\boldsymbol{y}} \tag{4}$$

式中：$\hat{\kappa}^{xx}$为高斯核函数$K=\langle\varphi(\boldsymbol{x}),\varphi(\boldsymbol{x})\rangle$的离散傅里叶变换；“*”表示复共轭；“⊙”表示逐元素相乘。

1.2 检测阶段

在检测过程中，用已经训练完成的分类器对新一帧图像$\boldsymbol{z}$进行滤波计算，得到响应图，检测过程中输出相应方程为

$$\hat{f}(\boldsymbol{z})=\hat{\kappa}^{xz}\odot\hat{\boldsymbol{\alpha}} \tag{5}$$

式中：$\hat{\kappa}^{xz}$为高斯核函数$K=\langle\varphi(\boldsymbol{x}),\varphi(\boldsymbol{z})\rangle$的离散傅里叶变换，将$\hat{f}(\boldsymbol{z})$转换回空间域，最大的响应位置即为预测目标的位置。

1.3 更新阶段

对跟踪目标完成检测后，随着背景发生变化，跟踪目标也有可能发生变化。因此，需要对目标样本$\hat{\boldsymbol{x}}$和分类器参数$\hat{\boldsymbol{\alpha}}$进行更新，更新模型为

$$\hat{\boldsymbol{\alpha}}_t=(1-\eta)\hat{\boldsymbol{\alpha}}_{t-1}+\eta\hat{\boldsymbol{\alpha}} \tag{6}$$

$$\hat{\boldsymbol{x}}_t = (1-\eta)\hat{\boldsymbol{x}}_{t-1} + \eta\hat{\boldsymbol{x}} \tag{7}$$

式中: η为学习因子; t为帧数; $\hat{\boldsymbol{\alpha}}_{t-1}$和$\hat{\boldsymbol{x}}_{t-1}$分别为上一帧分类器的模板参数和目标外观模型; $\hat{\boldsymbol{\alpha}}_t$和$\hat{\boldsymbol{x}}_t$分别为当前帧分类器模板参数和目标外观模型。

2　目标跟踪框架

基于 KCF 算法[9] 的框架进行改进, 通过卡尔曼滤波预测 ROI 区域, 极大减少了全局搜索时间, 提高了跟踪效率, 提取目标的灰度、梯度及局部二值特征, 通过自适应融合的策略, 根据不同特征的响应情况得到最佳的融合结果, 完成对目标位置的定位。根据当前的跟踪状态, 实时对分类器和目标模板进行自适应更新, 减少了过程中可能出现的模型漂移, 再将定位的目标位置作为更新卡尔曼滤波的量度。

2.1　特征选择

在相关滤波跟踪的过程中, 需要考虑特征的选用及特征的表达方式。在核相关滤波框架中, 单一特征点的方式在红外图像中会带来干扰, 因此, 需要提取图像的全局特征。在提取全局特征时, 如果只提取单一特征, 面对复杂的背景变换, 特征的有效性可能会发生变化, 导致跟踪丢失。因此, 针对上述问题, 结合红外图像的特点, 选用灰度特征、梯度特征和局部二值特征。

灰度特征作为最简单的一维特征, 虽然在描述能力上有所欠缺, 但运行速度快, 可以将其他特征串联起来。梯度特征对目标轮廓和表面纹理有良好的描述能力, 具有一定的平移、旋转不变性, 针对目标在背景干扰较小的情况下有较好的表现, 但当背景出现很多噪点干扰或目标运动速度过快时, 具有旋转不变性和灰度不变性的局部二值特征可以发挥很好的作用, 有效滤除高频噪声, 并处理高速运动带来的运动模糊问题。因此, 本文将灰度特征、梯度特征、局部二值特征这 3 种特征进行融合, 提升目标跟踪的性能。

2.2　特征融合策略

在目标跟踪的过程中, 由于背景复杂多变, 目标不断受到外界的干扰, 尤其是目标被遮挡或背景中出现相似目标时。因此, 需要引入多种关于目标跟踪的置信度指标。在相关滤波中, 常引入峰值旁瓣比作为置信度指标, 但本文主要引入 APCE[19] 和相邻 2 帧目标的相关滤波 CFR, 能得到更好的融合效果。APCE[20] 定义为

$$P_{\mathrm{APCE}} = \frac{|F_{\max} - F_{\min}|^2}{\mathrm{mean}\left\{\sum_{m,n}\left(F_{m,n} - F_{\min}\right)^2\right\}} \tag{8}$$

式中: $F_{\max}$、$F_{\min}$和$F_{m,n}$分别为目标融合响应图中的最大值、最小值和第m行第n列的元素; mean(·) 表示做最大值和最小值的极差运算。

式 (8) 体现了整体跟踪过程中的干扰影响。当出现干扰时, 响应图出现多峰或震荡情况, P_{APCE} 值变小, 跟踪结果可靠性较低; 反之, 如果 P_{APCE} 值越高, 跟踪结果更可靠。但仅使用 P_{APCE} 值作为唯一的置信度指标无法应对更复杂的跟踪情况, 当目标出现不规则机动时, 会导致更多的冗杂背景信息进入置信度指标。因此, 针对这种情况, 引入第 2 个置信度指标, 即相邻 2 帧目标的相关滤波 CFR, 定义为

$$P_{\mathrm{CFR}} = \|\hat{g}_t\left(\boldsymbol{x}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{p}}\right) - \hat{g}_{t-1}\left(\boldsymbol{x}_{\mathrm{p}} + \Delta\boldsymbol{x}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{p}} + \Delta\boldsymbol{y}_{\mathrm{p}}\right)\|_2^2 \tag{9}$$

式中: $\hat{g}_t\left(\boldsymbol{x}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{p}}\right)$和$\hat{g}_{t-1}\left(\boldsymbol{x}_{\mathrm{p}} + \Delta\boldsymbol{x}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{p}} + \Delta\boldsymbol{y}_{\mathrm{p}}\right)$分别为目标在第$t$帧和第$t-1$帧的滤波响应; $\Delta\boldsymbol{x}_{\mathrm{p}}$和$\Delta\boldsymbol{y}_{\mathrm{p}}$为相邻 2 帧目标的位置变化。

在实时红外目标跟踪过程中, 无论目标做怎样的运动轨迹, 由于 2 帧之间时间较短, 往往只有 20 ms, 目标的特征变化及背景信息往往都是连续的, 体现在式(9)中, 即为相邻 2 帧之间的相关性越高, P_{CFR} 值越低, 而相邻 2 帧之间相关性越低, P_{CFR} 值明显升高。可以借此来判断整个跟踪过程的置信度。

可以将 2 个置信度指标 P_{APCE} 和 P_{CFR} 进行权重计算, 作为综合评价函数来衡量跟踪过程的可靠性, 其定义为

$$F(P_{\mathrm{APCE}}, P_{\mathrm{CFR}}) = (1-\rho)P_{\mathrm{APCE}} + \rho\frac{1}{P_{\mathrm{CFR}} + \varepsilon} \tag{10}$$

式中: ρ为 2 个置信度之间的权重调节系数, 取值范围为[0,1]; ε为 0.01, 用以避免分母为 0。

提取 ROI 区域的灰度特征、梯度特征和局部二值特征, 针对不同特征分别训练出 3 个不同的核相关滤波器。当得到核相关滤波响应时, 使用式(10)构造的评价函数计算不同特征之间的置信度, 计算式为

$$\begin{cases} f_{\mathrm{Gray}} = (1-\rho)P_{\mathrm{APCE.Gray}} + \rho\dfrac{1}{P_{\mathrm{CFR.Gray}} + \varepsilon} \\ f_{\mathrm{Hog}} = (1-\rho)P_{\mathrm{APCE.Hog}} + \rho\dfrac{1}{P_{\mathrm{CFR.Hog}} + \varepsilon} \\ f_{\mathrm{LBP}} = (1-\rho)P_{\mathrm{APCE.LBP}} + \rho\dfrac{1}{P_{\mathrm{CFR.LBP}} + \varepsilon} \end{cases} \tag{11}$$

式中: 下标 Gray、Hog、LBP 分别表示灰度、梯度、局部二值; f_{Gray}、f_{Hog}和f_{LBP}分别为 3 个特征滤波响应的置信度。

由于目标运动及场景变化, 不同特征之间所表现的效果也不同。因此, 通过评价函数所得的不同特征之间的置信度将 3 个特征的相关滤波响应融合为最终的相关滤波响应结果, 融合后的滤波响应

结果为

$$\hat{f}_{\text{final}}=\frac{f_{\text{Gray}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}\hat{f}_{\text{Gray}}+\frac{f_{\text{Hog}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}\hat{f}_{\text{Hog}}+\frac{f_{\text{LBP}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}\hat{f}_{\text{LBP}} \tag{12}$$

式中：$\hat{f}_{\text{Gray}}$、$\hat{f}_{\text{Hog}}$和$\hat{f}_{\text{LBP}}$分别为灰度特征、梯度特征和局部二值特征的核相关滤波响应；$\frac{f_{\text{Gray}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}$、$\frac{f_{\text{Hog}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}$、$\frac{f_{\text{LBP}}}{f_{\text{Gray}}+f_{\text{Hog}}+f_{\text{LBP}}}$分别为各个特征的权重因子。融合方式如图1所示。

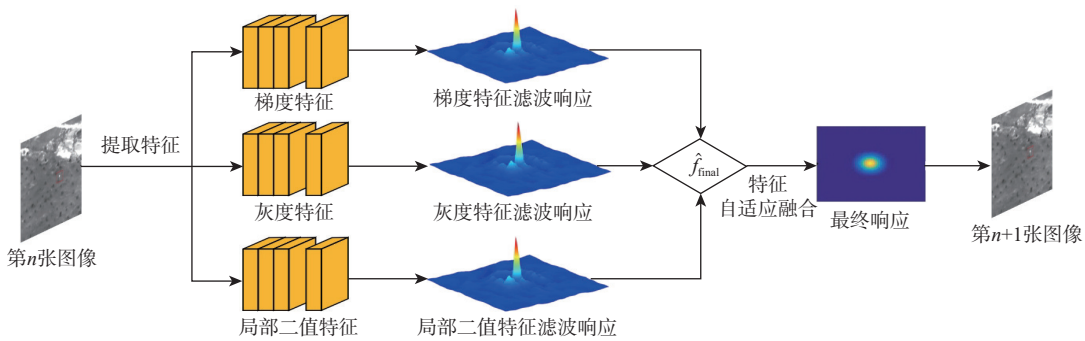


图1 特征自适应融合示意图

Fig. 1 Feature adaptive fusion diagram

2.3 更新策略

在跟踪过程中，由式(6)和式(7)可知，目标样本和分类器都会不断更新，在实际场景中，背景杂乱和目标遮挡等情况下，如果采用原更新方式，会引入不可靠的跟踪结果到模型中，随着模型误差的的累积，会导致跟踪失败，因此，需要根据置信度结果来决定目标样本和分类器是否更新。根据式(12)的特征融合策略，将加权融合后的相关滤波响应的APCE值和CFR值作为目标跟踪的置信度评价标准，经过3个不同特征加权融合后的置信度为

$$f_{\text{final}}=(1-\rho)P_{\text{APCE.final}}+\rho\frac{1}{P_{\text{CFR.final}}+\varepsilon} \tag{13}$$

式中：最终APCE值$P_{\text{APCE.final}}$和最终CFR值$P_{\text{CFR.final}}$的计算方法如式(8)和式(9)所示。

当f_{final}值较小时，说明目标跟踪的置信度较低，此时，目标发生了遮挡或模糊，不对跟踪模型进行更新；反之，如果f_{final}值较大，目标的跟踪置信度高，可以对跟踪模型进行更新。因此，如何判断f_{final}值的大小，需要引入置信度阈值f_{th}来判断跟踪模型是否更新，如果f_{final}大于设定的阈值f_{th}，则用当前帧信息对跟踪模型进行更新，否则不更新。由式(6)和式(7)可知，跟踪模型中训练样本$\boldsymbol{x}$和对位空间$\boldsymbol{\alpha}$都需要进行更新。由于本文有3个特征的跟踪模型，关于灰度特征、梯度特征、局部二值特征的跟踪模型都需要进行更新，3个特征追踪模型的分类器和目标样本具体的更新公式分别为

$$\begin{cases}\hat{\boldsymbol{\alpha}}_{\text{Gray}_t}=(1-\eta)\hat{\boldsymbol{\alpha}}_{\text{Gray}_t-1}+\eta\hat{\boldsymbol{\alpha}}_{\text{Gray}}\\ \hat{\boldsymbol{\alpha}}_{\text{Hog}_t}=(1-\eta)\hat{\boldsymbol{\alpha}}_{\text{Hog}_t-1}+\eta\hat{\boldsymbol{\alpha}}_{\text{Hog}}\\ \hat{\boldsymbol{\alpha}}_{\text{LBP}_t}=(1-\eta)\hat{\boldsymbol{\alpha}}_{\text{LBP}_t-1}+\eta\hat{\boldsymbol{\alpha}}_{\text{LBP}}\end{cases} \tag{14}$$

$$\begin{cases}\hat{\boldsymbol{x}}_{\text{Gray}_t}=(1-\eta)\hat{\boldsymbol{x}}_{\text{Gray}_t-1}+\eta\hat{\boldsymbol{x}}_{\text{Gray}}\\ \hat{\boldsymbol{x}}_{\text{Hog}_t}=(1-\eta)\hat{\boldsymbol{x}}_{\text{Hog}_t-1}+\eta\hat{\boldsymbol{x}}_{\text{Hog}}\\ \hat{\boldsymbol{x}}_{\text{LBP}_t}=(1-\eta)\hat{\boldsymbol{x}}_{\text{LBP}_t-1}+\eta\hat{\boldsymbol{x}}_{\text{LBP}}\end{cases} \tag{15}$$

式中：$\hat{\boldsymbol{\alpha}}_{\text{Gray}_t}$、$\hat{\boldsymbol{\alpha}}_{\text{Hog}_t}$和$\hat{\boldsymbol{\alpha}}_{\text{LBP}_t}$分别为3个特征核相关滤波器在当前帧的分类器参数；$\hat{\boldsymbol{\alpha}}_{\text{Gray}_t-1}$、$\hat{\boldsymbol{\alpha}}_{\text{Hog}_t-1}$和$\hat{\boldsymbol{\alpha}}_{\text{LBP}_t-1}$分别为3个特征核相关滤波器在前一帧的分类器参数；$\hat{\boldsymbol{x}}_{\text{Gray}_t}$、$\hat{\boldsymbol{x}}_{\text{Hog}_t}$和$\hat{\boldsymbol{x}}_{\text{LBP}_t}$分别为当前帧3个特征核相关滤波器的目标外观模型；$\hat{\boldsymbol{x}}_{\text{Gray}_t-1}$、$\hat{\boldsymbol{x}}_{\text{Hog}_t-1}$和$\hat{\boldsymbol{x}}_{\text{LBP}_t-1}$分别为前一帧3个特征核相关滤波器的目标外观模型。

通过该方式可以实现在确保不丢失前一帧关键信息的同时，将当前帧的有效信息更新到跟踪模型中。

2.4 目标预定位

针对红外图像的第1帧，将卡尔曼滤波[21]的初始目标状态向量设置为$\boldsymbol{X}_0=[\boldsymbol{x}_0,\boldsymbol{0},\boldsymbol{y}_0,\boldsymbol{0}]$，其中，$[\boldsymbol{x}_0,\boldsymbol{y}_0]$和$[\boldsymbol{0},\boldsymbol{0}]$分别为第1帧目标的初始位置和速度。通过目标在$t$帧时的状态方程和观测方程可以得到目标在$t$帧的预测位置，分别为

$$\boldsymbol{X}_t=\boldsymbol{A}\boldsymbol{X}_{t-1}+\boldsymbol{W}_t \tag{16}$$

$$\boldsymbol{Z}_t=\boldsymbol{H}_t\boldsymbol{X}_t+\boldsymbol{V}_t \tag{17}$$

式中：$\boldsymbol{X}_t$和$\boldsymbol{Z}_t$分别为目标在第t帧状态下的状态向量和观测向量；$\boldsymbol{A}$为状态矩阵；$\boldsymbol{H}_t$为观测矩阵；$\boldsymbol{W}_t$和$\boldsymbol{V}_t$分别表示均值为零的状态误差和均值误差。

根据式(16)和式(17)可以得到目标的预测位置$(\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)$，以目标的预测位置为中心设置ROI区域，ROI为矩形区域，两侧长度为目标的6倍，可以达到跟踪性能与计算时间的平衡。根据式(12)可以得到ROI区域中滤波响应的峰值目标为$(\boldsymbol{x}^*,\boldsymbol{y}^*)$，结

合目标的预测位置, 可以得到目标的估计位置为

$$\begin{cases}\hat{\boldsymbol{x}}_t=\dot{\boldsymbol{x}}_t+\boldsymbol{x}^*-\boldsymbol{MI}/2\\ \hat{\boldsymbol{y}}_t=\dot{\boldsymbol{y}}_t+\boldsymbol{y}^*-\boldsymbol{NI}/2\end{cases}\tag{18}$$

式中: M 和 N 分别为 ROI 区域的宽度和高度。

2.5　算法流程

在跟踪的第 1 帧过程中, 给定目标位置, 设置卡尔曼滤波的初始目标状态, 收集目标的 3 个特征信息, 构建 3 个特征的核相关滤波器, 得到分类器参数 $\hat{\boldsymbol{\alpha}}_{\text{Gray}}$、$\hat{\boldsymbol{\alpha}}_{\text{Hog}}$ 和 $\hat{\boldsymbol{\alpha}}_{\text{LBP}}$ 及目标外观参数 $\hat{\boldsymbol{x}}_{\text{Gray}}$、$\hat{\boldsymbol{x}}_{\text{Hog}}$ 和 $\hat{\boldsymbol{x}}_{\text{LBP}}$。利用卡尔曼滤波预测当前帧中的初始目标位置, 以目标的预测位置设置 ROI 区域; 提取 ROI 区域的 3 个特征信息, 计算其特征响应。根据式(11)计算 3 个特征的置信度 f_{Gray}、f_{Hog} 和 f_{LBP}, 根据式(12)分配特征权重得到特征融合后的相关滤波响应 $\hat{f}_{\text{final}}$, 得到滤波相应峰值位置 $(\boldsymbol{x}^*,\boldsymbol{y}^*)$。在更新阶段设定阈值 f_{th}, 根据式(13)得到特征融合后相关滤波响应的置信度 f_{final}, 通过计算置信度 f_{final} 和阈值 f_{th}, 当置信度大于设定阈值时, 按照式(14)和式(15)对特征模型进行更新, 若置信度小于设定阈值, 不进行更新, 并将式(18)得到的目标估计位置作为更新卡尔曼滤波的量度。具体流程如图 2 所示。

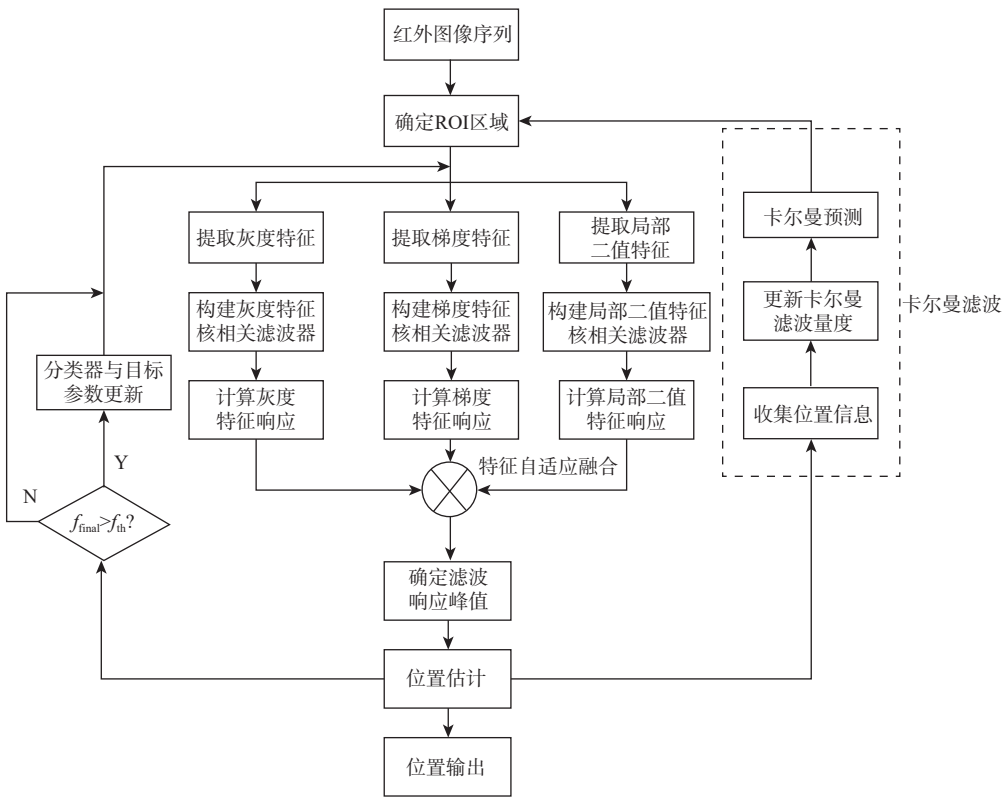


图 2　本文算法流程

Fig. 2　Flow chart of the proposed algorithm

3　实验结果与分析

3.1　测试数据集及参数设置

测试数据集使用红外序列图像中小飞机目标检测跟踪数据集[22]。该数据集中出现了复杂地面背景干扰、背景噪声杂乱、相似目标干扰、目标微弱、目标做不规则运动等一系列挑战。

实验硬件环境条件为 Intel Core i5-12500H, 主频为 2.40 GHz, 内存为 16 GB; 开发软件为 MATLAB-R2021a。本文算法的具体参数设置如下: 目标跟踪的置信度阈值 $f_{\text{th}}=6$, 学习因子 $\eta=0.02$, 置信度权重调节系数 $\rho=0.5$, 状态矩阵和观测矩阵分别为 $\boldsymbol{A}=\begin{bmatrix}1&1&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&1\end{bmatrix}$, $\boldsymbol{H}=\begin{bmatrix}1&0&0&0\\0&1&0&0\end{bmatrix}$, 其余参数设置与原 KCF 算法[9] 相同。

采用 3 种标准评价指标, 分别为重叠精度(overlap precision, OP)、距离精度(distance precision, DP)及跟踪速度。DP 定义为跟踪框的中心位置与目标实际中心位置之间的欧氏距离小于 20 像素的帧数占总帧数的比例。OP 定义为跟踪区域与实际区域的重叠率大于 0.5 的帧数占比。而跟踪速度用每秒平均帧数表征。

3.2　消融实验

为验证本文提出的将灰度特征、梯度特征和局

部二值特征进行自适应加权融合的模块性能及卡尔曼滤波预定位的模块性能，在红外序列图像中小飞机目标检测跟踪数据集[22]中进行消融实验测试。为验证每个改进点对算法的提升效果，设计消融实验，结果如表1所示。

通过对比表1中组别1、3、5、7可以看出，采用不同特征进行融合，跟踪效率有显著差异，其中，影响最大的为图像的局部二值特征，影响较小的为图像的梯度特征，且随着选用特征数量的增加，目标跟踪过程中的跟踪速度降低，由此可知，增加特征的选取数量，可以有效提升算法的跟踪效率。通过对比组别1～6可以看出，本文提出的多特征自适应融合模块在针对采用2个特征进行融合的情况下，跟踪的DP和OP都有很好的提升效果，但由于融合特征的方式更加复杂，导致跟踪过程的跟踪速度显著降低；对比组别7和组别9在采用3个特征的情况下，多特征自适应融合模块跟踪的DP提升了0.093，OP提升了0.092，即在特征数量提取越多的情况下，自适应加权模式带来的提升效果更加明显。通过对比组别9和组别10的实验结果发现，由于采用预定位的方式减少了大量不必要的搜索区域，显著提升了跟踪速度，且对跟踪的鲁棒性未造成较大影响。

表1 消融实验结果

Table 1 Ablation experimental results

组别	特征选择	多特征自适应融合模块	卡尔曼滤波预定位模块	距离精度	重叠精度	跟踪速度/(帧·s^{-1})
1	GRAY+HOG			0.611	0.569	135.1
2	GRAY+HOG	√		0.647	0.581	51.9
3	HOG+LBP			0.667	0.606	107.6
4	HOG+LBP	√		0.689	0.625	44.5
5	GRAY+LBP			0.683	0.639	124.2
6	GRAY+LBP	√		0.711	0.657	48.3
7	GRAY+HOG+LBP			0.702	0.654	89.7
8	GRAY+HOG+LBP		√	0.693	0.647	128.4
9	GRAY+HOG+LBP	√		0.795	0.746	32.2
10(本文)	GRAY+HOG+LBP	√	√	0.782	0.731	94.7

综上所述，通过消融实验可以看出，本文提出的多特征自适应融合模块可以有效地提升目标跟踪的鲁棒性，卡尔曼滤波预定位模块可以有效地提高跟踪算法的跟踪速度。

3.3 目标跟踪算法对比定量分析

为验证本文算法的有效性，将本文算法与当前使用较多的7种目标跟踪算法进行对比分析，7种目标跟踪算法分别为通过连续卷积再进行特征降维的高效卷积算子(efficient convolution operators, ECO)算法，加入时间正则项、使用乘子交替方向法(alternating direction method of multipliers, ADMM)迭代求解的时空正则化相关滤波(spatio-temporal regularized correlation filters, STRCF)算法，使用裁剪算子优化样本空间的样本裁剪空间滤波(background-aware correlation filters, BACF)算法，利用正则矩阵在频域稀疏性进行迭代收敛的空间正则化判别相关滤波(spatially regularized discriminative correlation filters, SRDCF)算法，利用循环位移获得训练样本、再使用正则项防止过度拟合的KCF算法，将一维特征扩充到多维、融合多特征的多特征融合(discriminative scale space tracker, DSST)算法，采用多尺度图像金字塔以适应尺度变化的多尺度金字塔自适应(scale adaptive with multiple features tracker, SAMF)算法。

在红外序列图像中小飞机目标检测跟踪数据集[22]中选取6种具有代表性的场景进行测试，数据集特征描述如表2所示。

为更好地描述上述算法在不同场景的表现，给出DP曲线和OP曲线，如图3和图4所示。DP和OP的详细结果如表3和表4所示。

距离阈值为20像素时，从图3和表3可以看

表2 数据集特征描述

Table 2 Data set feature description

场景图像序列	帧数	目标大小/(像素×像素)	目标平均速度/(像素·s^{-1})	场景描述
场景1	399	3×3	0.47	地面背景干扰较小
场景2	399	4×4	0.38	地面背景干扰较小，目标做不规则俯仰运动
场景3	399	5×5	0.40	地面背景干扰较小，相似目标干扰
场景4	399	5×5	1.91	地面背景干扰较大，目标做快速移动
场景5	749	2×2	0.49	地面背景干扰较大，目标微弱，相似目标干扰
场景6	499	3×3	0.42	地面背景干扰较大，背景噪声杂乱

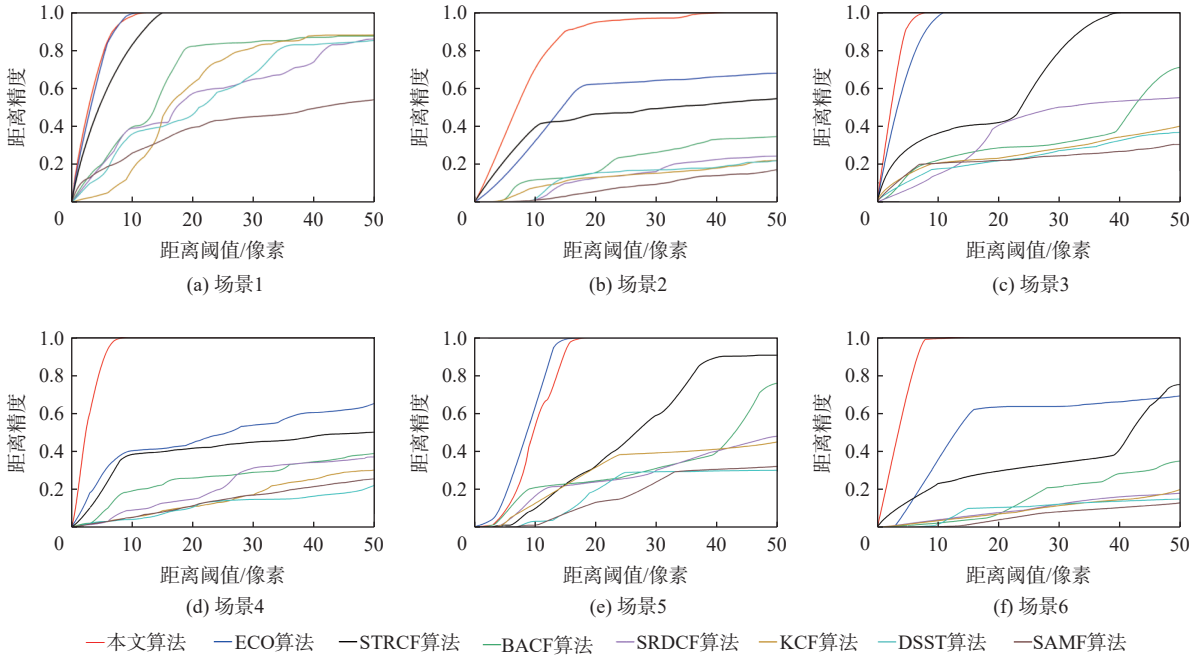


图 3　8 种对比算法在不同场景下的距离精度曲线

Fig. 3　Distance precision curves of 8 comparison algorithms in different scenarios

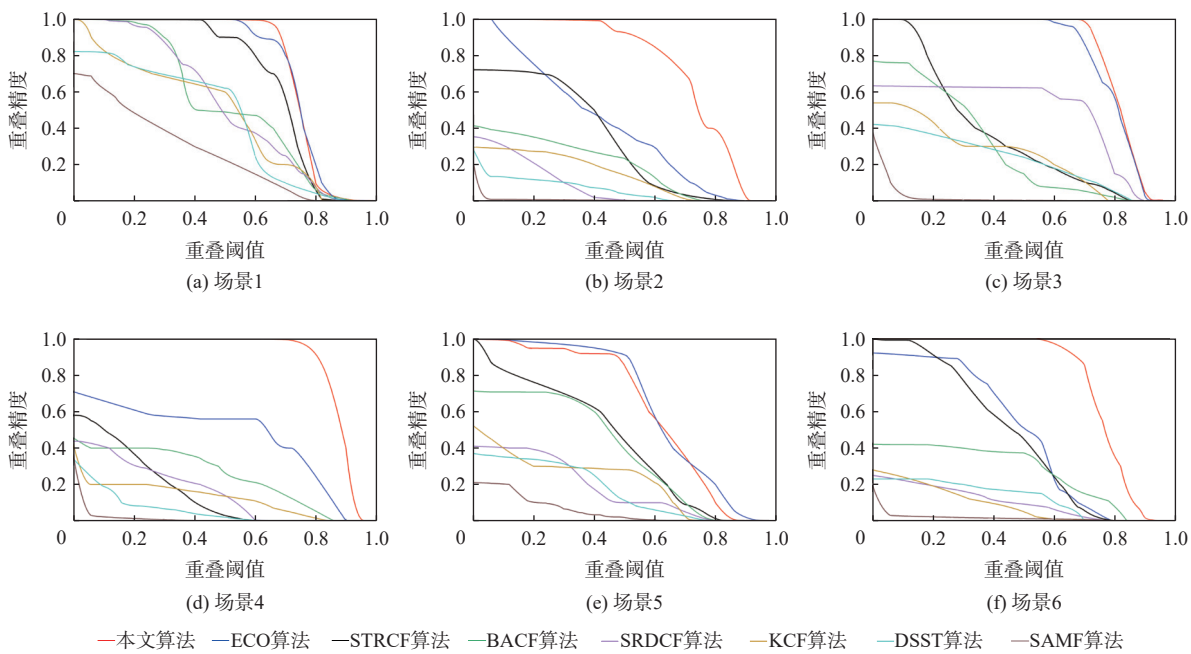


图 4　8 种对比算法在不同场景下的重叠精度曲线

Fig. 4　Overlap precision curves of 8 contrast algorithms in different scenarios

到, 本文算法在 DP 的表现上, 除在场景 4 为次优, 其余 5 个场景中均为最优表现。

重叠阈值为 0.5 时, 从图 4 和表 4 可以看到, 本文算法在 OP 的表现上, 除在场景 4 为次优, 其余 5 个场景中均为最优表现。

从表 5 可以看到, 跟踪速度大于 90 帧/s 的高速跟踪算法有本文算法、KCF 算法和 DSST 算法, 但 KCF 算法和 DSST 算法在跟踪结果上却不尽人意。虽然采用深度学习的 ECO 算法跟踪结果较好, 但无法满足 30 帧/s 的实时跟踪需求。

3.4　目标跟踪算法对比定性分析

为更加直观地评估各算法的跟踪性能, 对 8 种算法在 6 个场景下进行定性分析, 结果如图 5 所示。针对表 2 中各个场景的特性, 场景 1～场景 3 都为背景干扰较小的情况, 场景 4～场景 6 都为背景干扰较大的情况。在场景 1 中, 由于背景干扰较小, 大部分跟踪算法都能在前 300 帧内跟定目标, 而在后续的跟踪过程中, 表现最优异的为本文算法

表3 8种对比算法在不同场景下的距离精度

Table 3 Distance precision of 8 comparison algorithms in different scenarios

算法	场景1	场景2	场景3	场景4	场景5	场景6
本文	**1.000**	**1.000**	**1.000**	0.984	**0.921**	**1.000**
ECO	**1.000**	0.427	**1.000**	**1.000**	0.613	0.628
STRCF	**1.000**	0.406	0.409	0.329	0.425	0.294
BACF	0.817	0.245	0.263	0.217	0.131	0.086
SRDCF	0.553	0.136	0.395	0.213	0.105	0.087
KCF	0.627	0.076	0.214	0.324	0.109	0.080
DSST	0.413	0.069	0.206	0.188	0.130	0.094
SAMF	0.368	0.072	0.204	0.124	0.036	0.021

注: 加粗数字表示最优值。

表4 8种对比算法在不同场景下的重叠精度

Table 4 Overlap precision of 8 comparison algorithms in different scenarios

算法	场景1	场景2	场景3	场景4	场景5	场景6
本文	**0.997**	**1.000**	**1.000**	0.864	**0.901**	**1.000**
ECO	0.995	0.581	**1.000**	**0.916**	0.409	0.512
STRCF	0.892	0.023	0.294	0.423	0.308	0.477
BACF	0.503	0.259	0.168	0.397	0.256	0.382
SRDCF	0.506	0.182	0.617	0.308	0.101	0.113
KCF	0.616	0.174	0.289	0.105	0.164	0.026
DSST	0.627	0.019	0.254	0.113	0.029	0.184
SAMF	0.218	0.001	0.003	0.027	0.157	0.018

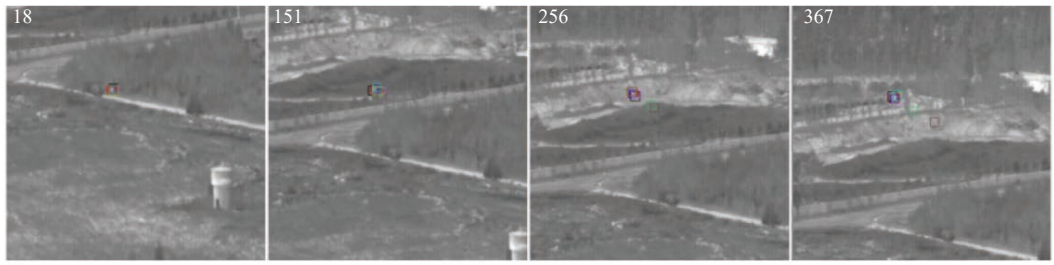
注: 加粗数字表示最优值。

表5 8种对比算法的跟踪速度

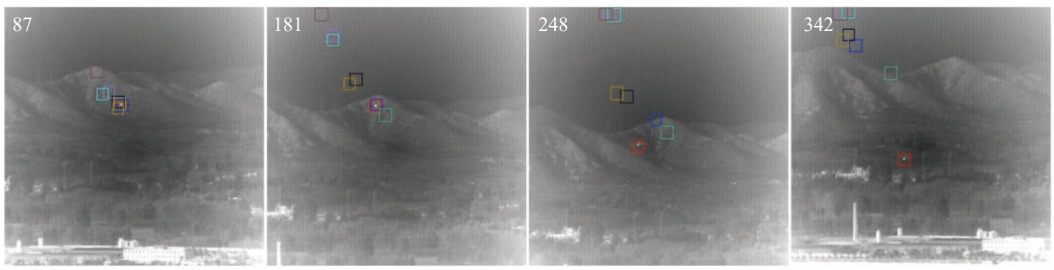
Table 5 Tracking speed of 8 comparison algorithms

算法	跟踪速度/(帧·s^{-1})
本文	94.7
ECO	18.6
STRCF	45.8
BACF	33.2
SRDCF	22.4
KCF	316.2
DSST	289.4
SAMF	46.9

和ECO算法。而在场景2中,在背景干扰较小的前提下,目标做不规则的俯仰运动,目标在前181帧都在做上升运动,可以看到,此时,本文算法和ECO算法都可以对目标进行跟踪,但在181帧后,目标迅速进入俯冲运动,由于本文算法加入了前后帧的连贯性置信度判别,此时可以捕捉到目标运动方向的改变,本文算法能够应对地面背景下目标做运动改变的情况,而其他算法会在目标发生运动时无法继续跟踪目标。场景3中出现了相似目标干扰,在背景干扰较小的情况下,BACF算法和STRCF算法都因为相似目标的干扰丢失目标,而KCF算法和DSST算法因为跟踪性能的原因而丢失了目标。场景4为背景干扰较大且目标进行高速运动的情况,在115帧,目标进入大片遮挡区域,只有本文算法和ECO算法能够对目标继续进行稳定跟踪,但在后续跟踪过程中,本文算法略逊色于ECO算法,而其余算法都发生了偏离,导致后续跟踪不准确或失败。场景5为出现相似目标干扰的情况,存在背景干扰较大且目标微弱的问题,第284帧时,由于目标过于微弱,ECO算法开始丢失目标,而其他算法已经无法继续跟踪目标,其中,BACF算法和STRCF算法还是因为在相似目标的干扰下出现了跟踪失误,而其他算法在复杂的背景情况下,早已无法对目标进行跟踪。场景6为大量的背景噪声影响,第65帧时,除ECO算法和STRCF算法,都因为噪声的原因跟丢了目标,随后,出现更多噪点干扰,导致最终只有本文算法能够继续跟踪目标。



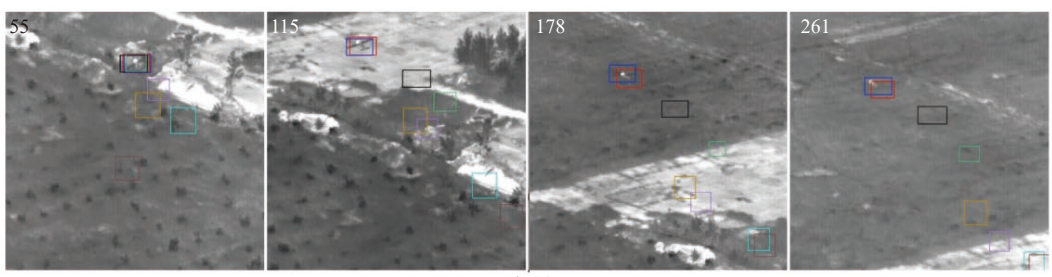
(a) 场景1



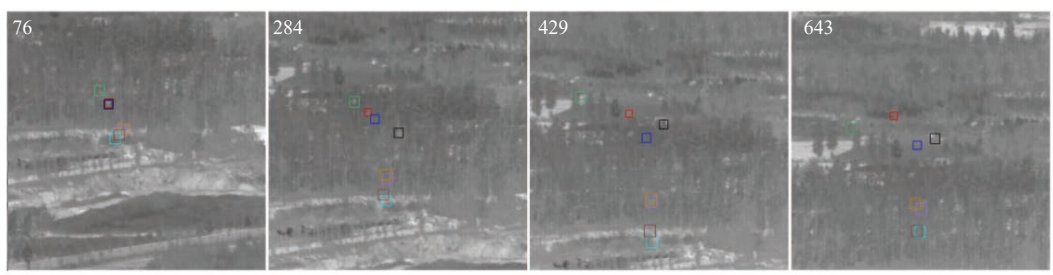
(b) 场景2



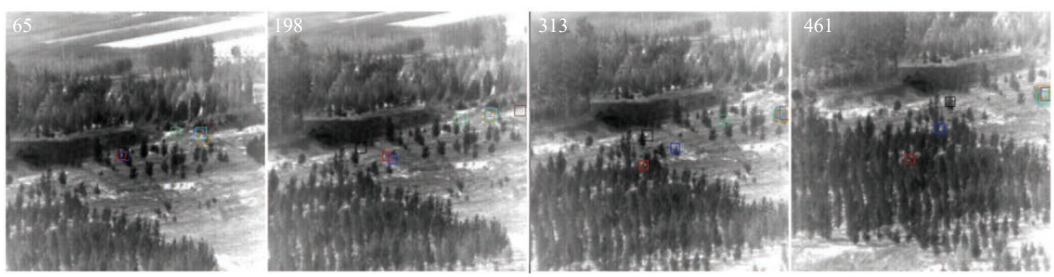
(c) 场景3



(d) 场景4



(e) 场景5



(f) 场景6

图 5　8 种对比算法在不同场景的跟踪结果对比

Fig. 5　Comparison of tracking results of 8 comparison algorithms in different scenarios

4 结 论

1）本文算法可以实现较为优异的跟踪精度和跟踪成功率，当距离阈值为15像素时，平均DP为0.782；当重叠阈值为0.5时，平均OP为0.731。

2）本文算法能够在达到较高的跟踪精度和跟踪成功率的同时，拥有较高的跟踪速度，可以达到94.7帧/s，相比于深度学习算法，能够完成对目标的实时跟踪。

3）本文提出的由不同置信度确定权值的特征自适应融合方式能够有效提高复杂地面背景下目标跟踪的跟踪精度和跟踪成功率，且可以应对不同场景的跟踪问题。

4）本文算法在分类器和目标参数更新的方式上采用双置信度阈值判别的方式，能够有效减少冗余信息影响目标跟踪进程的情况。

5）本文算法加入了卡尔曼滤波构建ROI区域来减少不必要的搜索区域，大大提高了算法效率，完成了对目标的实时跟踪。

参考文献（References）

[1] 房胜男, 谷小婧, 顾幸生. 基于自适应响应融合的相关滤波红外目标跟踪[J]. 红外与激光工程, 2019, 48(6): 479-486.
FANG S N, GU X J, GU X S. Infrared target tracking with correlation filter based on adaptive fusion of responses[J]. Infrared and Laser Engineering, 2019, 48(6): 479-486(in Chinese).

[2] YANG X, LI S Y, YU J T, et al. GF-KCF: aerial infrared target tracking algorithm based on kernel correlation filters under complex interference environment[J]. Infrared Physics & Technology, 2021, 119: 103958.

[3] WANG X T, ZHANG K, ZHANG X M, et al. Aerial infrared object tracking via an improved long-term correlation filter with optical flow estimation and SURF matching[J]. Infrared Physics & Technology, 2021, 116: 103790.

[4] ZHOU H, ZHANG Z Y, CHEN Z H, et al. Infrared dim small target tracking based on inter-frame consistency under complex background[C]//Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Computer Engineering. Piscataway: IEEE Press, 2022: 728-735.

[5] 常敏, 沈凯, 张学典, 等. 自适应特征融合和模型更新的相关滤波跟踪[J]. 光学学报, 2019, 39(9): 0915001.
CHANG M, SHEN K, ZHANG X D, et al. Adaptive feature fusion and correlation filtering tracking for model updating[J]. Acta Optica Sinica, 2019, 39(9): 0915001(in Chinese).

[6] WU S J, ZHANG K, LI S Y, et al. Joint feature embedding learning and correlation filters for aircraft tracking with infrared imagery[J]. Neurocomputing, 2021, 450: 104-118.

[7] 卢明瑞, 韩超, 鲁帆, 等. 复杂场景下的自适应相关滤波跟踪算法[J]. 激光与光电子学进展, 2022, 59(24): 174-183.
LU M R, HAN C, LU F, et al. Adaptive correlation filtering tracking algorithm in complex scene[J]. Laser & Optoelectronics Progress, 2022, 59(24): 174-183(in Chinese).

[8] 姜珊, 张超, 韩成, 等. 基于相关滤波的目标重检测跟踪[J]. 红外与激光工程, 2021, 50(2): 224-235.
JIANG S, ZHANG C, HAN C, et al. Target detection and tracking based on correlation filtering[J]. Infrared and Laser Engineering, 2021, 50(2): 224-235(in Chinese).

[9] SHU Q Z, LAI H C, WANG L J, et al. Multi-feature fusion target re-location tracking based on correlation filters[J]. IEEE Access, 2021, 9: 28954-28964.

[10] ZHA Y F, QIU Z L, SUN J X, et al. Learning spatial-channel regularization jointly with correlation filter for visual tracking[J]. Neurocomputing, 2021, 453: 839-852.

[11] FAN B J. Spatial graph regularized multi-kernel subtask cross-correlation tracker[C]//Proceedings of the 2021 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2021: 14248-14253.

[12] TANG M, ZHENG L Y, YU B, et al. Fast kernelized correlation filter without boundary effect[C]//Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2021: 2998-3007.

[13] YUAN D, SHU X, HE Z Y. TRBACF: learning temporal regularized correlation filters for high performance online visual object tracking[J]. Journal of Visual Communication and Image Representation, 2020, 72: 102882.

[14] CHEN J, LIN Y M, HUANG D T, et al. Robust tracking algorithm for infrared target via correlation filter and particle filter[J]. Infrared Physics & Technology, 2020, 111: 103516.

[15] PENG Z, LU X J. Learning region sparse constraint correlation filter for tracking[J]. Signal Processing: Image Communication, 2021, 90: 116042.

[16] LI Z Z, CHEN C, LIU D P, et al. Infrared small target tracking algorithm based on temporal-spatial structure sparse Bayesian estimation[J]. Infrared Physics & Technology, 2020, 105: 103160.

[17] TAI Y, TAN Y H, XIONG S Z, et al. Subspace reconstruction based correlation filter for object tracking[J]. Computer Vision and Image Understanding, 2021, 212: 103272.

[18] PAVLOVIĆ M, BANJAC Z, KOVAČEVIĆ B. Object tracking in SWIR imaging based on both correlation and robust Kalman filters [J]. IEEE Access, 2023, 11: 63834-63851.

[19] LIU P Q, DU J, HE S Y, et al. A real-time target tracking algorithm based on improved kernel correlation filter[C]//Proceedings of the 2021 5th International Conference on Imaging, Signal Processing and Communications. Piscataway: IEEE Press, 2021: 5-9.

[20] FENG M Z, SONG K C, WANG Y Y, et al. Learning discriminative update adaptive spatial-temporal regularized correlation filter for RGB-T tracking[J]. Journal of Visual Communication and Image Representation, 2020, 72: 102881.

[21] SHI G Q, ZHANG B Y, GENG X T, et al. Improvement of kernel correlation filtering algorithm based on Kalman filter[C]//Proceeding of the 2022 IEEE 17th International Conference on Control & Automation. Piscataway: IEEE Press, 2022: 394-399.

[22] 回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集[J]. 中国科学数据, 2020, 5(3): 286-297.

HUI B W, SONG Z Y, FAN H Q, et al. A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background[J]. China Scientific Data, 2020, 5(3): 286-297(in Chinese).

Real-time tracking of infrared dim-small target with multi-feature adaptive fusion under double confidence

CHEN Jiajun[1, 2]，LI Xiang[2, *]，SONG Yansong[2]，DONG Xiaona[1, 2]

(1. School of Opto-electronic Engineering，Changchun University of Science and Technology，Changchun 130022，China；
2. Institute of Space Photoelectric Technology，Changchun University of Science and Technology，Changchun 130022，China)

Abstract：In order to improve the robustness and real-time performance of the algorithm, a series of problems, such as ground background clutter, similar target interference and weak targets appear in the process of infrared dim small target tracking. Kalman filter is first used to predict the initial position of the target, and the initial position is set as the center of the region of interest (ROI). Next, determine the filter response graph by extracting the target's local binary (LBP) features, gradient (HOG) features, and grayscale (GRAY) features within the ROI region. Fusion weights are obtained according to the average peak correlation energy (APCE) of the three response results and the consistent frame response (CFR) of the adjacent frames, and the response results of the three features are fused by adaptive weighted fusion. To estimate the optimal location of the target. Finally, the target model is updated and the target position is taken as the measure of the Kalman filter. According to experimental data, the average distance precision (DP), overlap precision (OP), and real-time tracking speed for infrared ground background image sequences in various scenarios are 0.782, 0.731, and 94.7 frames per second, respectively. The algorithm can effectively improve the accuracy and robustness of tracking in complex environments.

Keywords：infrared dim-small target tracking；kernel correlation filtering；multi-feature adaptive fusion；Kalman filter；consistent frame of the adjacent frames

Received：2023-12-09；**Accepted**：2024-01-26；**Published Online**：2024-03-12 09：16

URL：link.cnki.net/urlid/11.2625.V.20240311.1451.003

Foundation items：National Key Research and Development Program of China (2022YFB3902505)；Key Project of Natural Science Foundation of Jilin Province of China (20230301002GX)

* **Corresponding author**. E-mail：lixiang@cust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0813

超声速盘收口盘缝带伞不同收口比下气动性能

代雨柔[1,2]，李健[1,2,*]，荣伟[1,2]，薛晓鹏[3]，黄明星[1,2]，方康[1,2]

(1. 北京空间机电研究所，北京 100094；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 100094；
3. 中南大学 自动化学院，长沙 410083)

摘　　要：火星着陆探测器的质量提升要求降落伞增大面积以保障稳降性能，同时，需控制开伞力以避免探测器结构负担过大质量。超声速盘缝带伞盘收口技术可同时满足上述2点需求。基于此，以火星环境为背景，采用流固耦合仿真分析，得到收口比与阻力系数比、投影面积比、收口绳载荷的函数关系；采用计算流体力学仿真分析方法，研究盘缝带伞的稳定性特征。研究表明，超声速盘收口盘缝带伞能够实现阻力性能的受控变化，提供满足工程应用的稳定性能。研究结果可为超声速盘收口盘缝带伞的工程设计提供参考。

关 键 词：盘缝带伞；收口技术；盘收口；阻力特征；稳定性
中图分类号：V445.2+3
文献标志码：A　　**文章编号**：1001-5965(2026)03-0864-10

盘缝带伞是一种密实织物降落伞，具有优良的阻力和稳定性能，常用于火星着陆探测等超声速低密度减速领域，目前成功软着陆火星的探测器使用的均是盘缝带伞[1-8]。

随着深空探测技术的进步，对未来先进盘缝带伞的需求主要有2方面：①面积更大，能够满足大型探测器的减速需求；②开伞力更小，使探测器不用付出过大的结构质量负担。但2个需求相互矛盾，现有的盘缝带伞技术不能同时满足。对超声速盘缝带伞进行收口是解决上述矛盾的技术途径之一，目的是主动控制降落伞的展开面积，通过渐次展开的方式实现开伞力的大幅降低。

盘缝带伞由1个带有中心通气孔的“盘”、1个圆环型开“缝”和1个圆筒型的“带”组成。盘缝带伞有2种常用的收口方式：盘收口和带收口，即将收口绳安装在盘或带的边缘，将下缘伞衣收紧，通过减小盘缝带伞投影面积，达到降低气动阻力的目的。代雨柔等[9]对盘收口和带收口盘缝带伞在超声速情况下的气动特性进行研究，结果表明，在相同收口直径比(以下简称“收口比”)下，盘收口能够提供更高的阻力和更好的稳定性。因此，本文重点针对盘收口盘缝带伞，进一步研究其阻力性能和稳定性能。

超声速条件下，盘缝带伞会出现剧烈的“呼吸与振荡”现象，开伞力增大，充气形状在充满与欠充满(甚至闭合)之间快速振荡，伞衣在空间位置上摆动剧烈，导致盘缝带伞阻力性能降低，稳定性变差。超声速减速是盘收口盘缝带伞的重要应用场景之一，但国内外对盘收口盘缝带伞的超声速气动特性研究很少。现阶段，国内外仅做过少量的风洞试验、高空开伞试验，还未得到工程应用[10-11]。现有的研究基础不足以满足未来的工程需求。

本文采用流固耦合和计算流体力学方法，对不同收口比的超声速盘收口盘缝带伞气动性能和稳

收稿日期：2023-12-13；**录用日期**：2024-03-27；**网络出版时间**：2024-04-03 18:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20240403.1145.006
基金项目：国家自然科学基金(12072377)；湖南省自然科学基金(2022JJ30678)
*通信作者. E-mail：lijian_bbmouth@163.com

*引用格式：代雨柔，李健，荣伟，等. 超声速盘收口盘缝带伞不同收口比下气动性能[J]. 北京航空航天大学学报，2026，52(3)：864-873.
DAI Y R，LI J，RONG W，et al. Aerodynamic characteristics of supersonic mid-gore reefing disk-gap-band parachute under different reefing ratios[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：864-873 (in Chinese).*

定性进行研究，得到投影面积比和阻力系数比与收口比的关系及收口绳载荷与盘缝带伞总阻力的关系，并分析不同攻角下盘缝带伞的稳定性。研究结果可为超声速盘收口盘缝带伞的气动设计和结构设计提供参考和支持。

1　研究对象

1.1　盘收口盘缝带伞

充满状态的盘缝带伞如图 1 所示。盘收口盘缝带伞是在盘的下缘，用收口绳将盘下缘伞衣收紧，减小投影面积来降低阻力。收口绳通过收口环等装置实现结构固定，使用收口绳切割器（火工品）切断收口绳，实现解除收口。盘收口盘缝带伞充气形状、收口绳及切割器的安装位置、安装形式示意图如图 2 所示[11-12]。

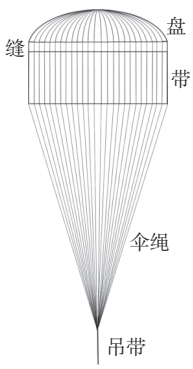


图 1　盘缝带伞

Fig. 1　Disk-gap-band parachute

1.2　盘缝带伞参数设置

使用有已公开试验数据的美国“海盗号”火星探测器用盘缝带伞的 4% 缩比伞为研究对象，具体尺寸如图 3 和表 1 所示[13]。本文主要研究内容是不同收口比对盘缝带伞气动性能的影响，故只对盘收口盘缝带伞进行研究，不考虑前体（被减速物体）的影响。

1.3　盘缝带伞收口参数设置

使用收口比作为收口程度的量化定义，即以收口绳长度为周长的圆直径与伞衣名义直径的比。在马赫数为 1.8～2.6 下，对收口比为 10%～40% 的盘收口盘缝带伞进行仿真分析，具体研究工况如表 2 所示。

超声速情况下，盘缝带伞的流场复杂，伞衣振动剧烈，收口比过小会导致盘缝带伞充气形状不稳定，因此，以收口比 10% 作为下限。马赫数为 1.8 时，在收口比为 50% 的盘收口情况下，收口阻力系数非常接近全展开情况的阻力系数，收口绳没有明显约束作用[9]，因此，选择收口比 40% 为上限。

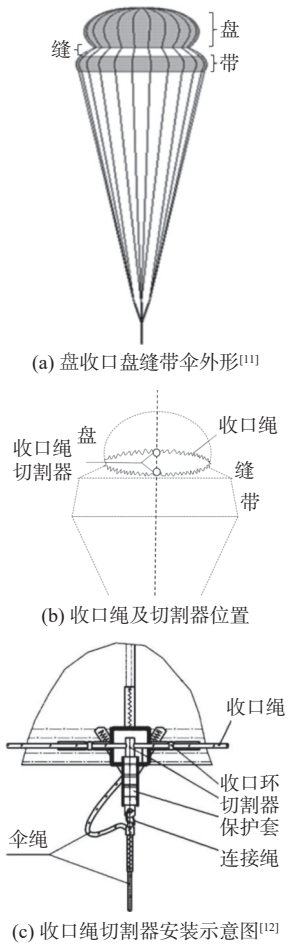


图 2　盘收口盘缝带伞[11-12]

Fig. 2　Disk-gap-band parachute with mid-gore reefing[11-12]

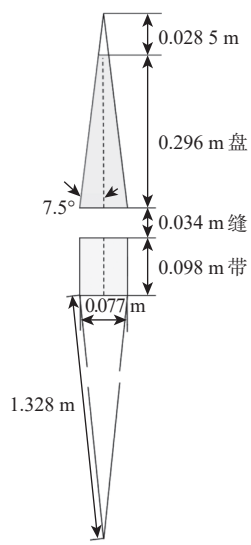


图 3　盘缝带伞结构[13]

Fig. 3　Disk-gap-band parachute structure[13]

表 1　盘缝带伞结构尺寸[13]

Table 1　Structural size of the disk-gap-band parachute[13]

名义直径/m	顶孔直径/m	盘高/m	缝高/m	带高/m	单幅带宽/m	伞绳长度/m
0.813	0.057	0.296	0.034	0.098	0.077	1.328

表 2 研究算例

Table 2 Research examples

马赫数	收口比/%
1.8	10, 15, 25, 30, 35, 40
2.2	10, 15, 25, 30, 35, 40
2.6	10, 15, 25, 30, 35, 40

2 研究方法及数值模型

使用流固耦合方法，对不同马赫数、收口比下，盘缝带伞的动态开伞过程进行模拟，得到稳定后盘缝带伞的稳定外形、集束点受力、伞衣网格节点位置和收口绳单元受力，计算盘缝带伞的阻力系数、投影面积和收口绳载荷，最终分析不同收口比下盘缝带伞的阻力系数比、投影面积比和收口绳载荷与盘缝带伞总阻力的比值。

同时，使用计算流体力学方法对稳态情况下，不同收口比的盘缝带伞外形进行分析，得到不同攻角下的轴向力和法向力，计算得到轴向力系数、力矩系数和力矩系数导数，根据结果分析盘缝带伞在流场中的稳定性。

2.1 研究方法

应用 LS-DYNA 软件中的 DUALCE/SE 求解器模拟盘缝带伞收口状态的动态开伞至稳定过程。该方法流场域使用 DUALCE/SE 方法进行计算，盘缝带伞柔性结构域使用有限元方法，使用浸入边界法实现流场与结构之间的流固耦合。

浸入边界法是一种简单、高效、稳定、可用来处理大变形流固耦合问题的方法。其中，流体计算网格固定，固体结构网格随着固体结构一起运动，流体网格和固体结构网格完全独立。在流体网格中，接触界面的位置由固体结构模块提供的位移信息决定[14]。

2.1.1 DUALCE/SE 方法

DUALCE/SE 方法 (dual-meshCE/SE, 简称为双网格 CE/SE 方法) 是对 CE/SE(时-空守恒元解元) 方法的小范围改进。CE/SE 方法物理概念清晰，计算精度高，格式构造简单。传统 CE/SE 方法求解器的流体变量都在网格中心点上，而 DUALCE/SE 方法采用双网格交替在流体单元中心和流体单元节点上求解流体状态变量[14]。在相同网格数下，DUALCE/SE 方法更精确。以一维非定常纳维-斯托克斯 (Navier-Stokes, N-S) 方程为例的求解步骤[15-19]如下：

1) 式 (1) 为一维情况下的 N-S 方程，将时间作为额外的空间维度，构建二维欧拉空间域 E_2，如图 4[16] 中①所示，x、t 为 E_2 的 2 个坐标：

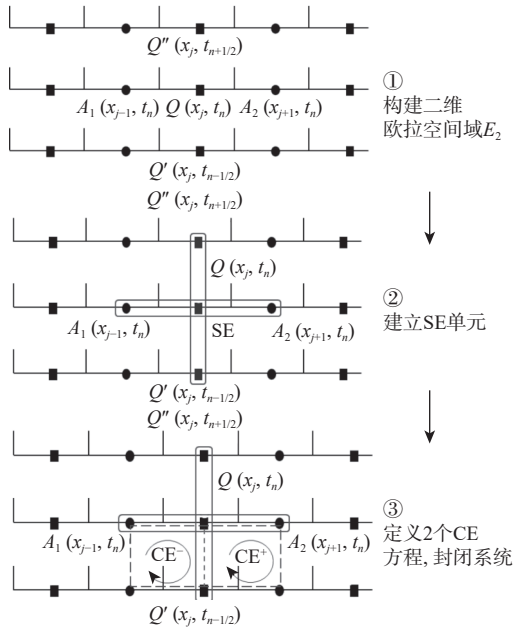


图 4 DUAL CE/SE 方法一维求解步骤[16]

Fig. 4 DUAL CE/SE one-dimensional solution procedure[16]

$$\frac{\partial \boldsymbol{u}_m}{\partial t}+\frac{\partial \boldsymbol{f}_m}{\partial x}-\frac{\partial^2 \boldsymbol{g}_m}{\partial x^2}=\boldsymbol{0} \qquad m=1,2,3 \tag{1}$$

式中：$\boldsymbol{u}_m$为速度；$\boldsymbol{f}_m$为体积力；$\boldsymbol{g}_m$为重力加速度。定义 $\boldsymbol{h}_m=\left(\boldsymbol{f}_m-\frac{\partial \boldsymbol{g}_m}{\partial x},\boldsymbol{u}_m\right)$，$m$=1,2,3 为时间-空间中的质量、动量、能量的流密度向量。

2) 利用高斯散度定理，将时间和空间完全统一，可将微分方程写为如下形式：

$$\oint_{S(V)} \boldsymbol{h}_m \mathrm{d}s=\boldsymbol{0} \qquad m=1,2,3 \tag{2}$$

式中：$S(V)$ 为 E_2 中任意区域 V 的边界。

3) 建立 SE 单元，如图 4 中②所示，沿 SE 域对流体变量进行泰勒级数展开，令

$$\boldsymbol{u}_m^*(x,t;j,n)=(\boldsymbol{u}_m)_j^n+(\boldsymbol{u}_{mx})_j^n(x-x_j)+(\boldsymbol{u}_{mt})_j^n(t-t^n) \tag{3}$$

$$\boldsymbol{f}_m^*(x,t;j,n)=(\boldsymbol{f}_m)_j^n+(\boldsymbol{f}_{mx})_j^n(x-x_j)+(\boldsymbol{f}_{mt})_j^n(t-t^n) \tag{4}$$

$$\boldsymbol{g}_m^*(x,t;j,n)=(\boldsymbol{g}_m)_j^n+(\boldsymbol{g}_{mx})_j^n(x-x_j)+(\boldsymbol{g}_{mt})_j^n(t-t^n) \tag{5}$$

同理可得

$$\boldsymbol{h}_m^*(x,t;j,n)=\left(\boldsymbol{f}_m^*(x,t;j,n)-\frac{\partial \boldsymbol{g}_m^*(x,t;j,n)}{\partial x},\boldsymbol{u}_m^*(x,t;j,n)\right) \tag{6}$$

式中：j 和 n 分别为对应空间和时间方向的编号；下标 mx 和 mt 为偏导数离散标识。

4) 将式 (2) 用式 (7) 逼近，并将式 (6) 代入式 (2)，得到式 (8)。

$$F(j,n)=\oint_{S(v_{\mathrm{CE}}\pm(j,n))} \boldsymbol{h}_m^* \mathrm{d}s=0 \qquad \forall (j,n)\in \Omega \tag{7}$$

$$(\boldsymbol{u}_{mj})_j^n=-(\boldsymbol{f}_{mx})_j^n \tag{8}$$

式中：$F(j, n)$ 为离散节点 (j, n) 的残余量；v_{CE} 为 2 个

守恒元 CE^-和 CE^+; 下标 mj 表示变量在空间节点 j 处的离散值。

此时, 在每个网格点的变量只有 $\boldsymbol{u}_m$ 和 $\boldsymbol{u}_{mx}$, 在 2 个守恒元 CE^-和 CE^+中, 如图 4 中③所示, 联立式 (7) 和式 (8), 得出计算结果。

2.1.2　计算流体力学方法

求出盘缝带伞的动态结果后, 使用计算流体力学方法, 将盘缝带伞的稳定外形固定在流场中, 对不同攻角下的盘收口盘缝带伞进行分析。在该方法中, 使用三维可压缩理想气体 N-S 方程并利用 HLLC 格式计算无黏通量。采用有限体积法进行空间离散, 双时间步隐式推进方法为时间格式。设置入口和出口为远场来流, 伞衣壁面设置为无滑移绝热壁面。有研究表明, 对于刚性盘缝带伞, 采用层流进行计算的结果与试验结果吻合较好[20], 因此, 本文选用层流计算。

2.1.3　基本假设

对仿真模型作如下假设: ①盘缝带伞初始状态为轴对称结构; ②不考虑前体影响; ③忽略伞绳气动力; ④不考虑盘缝带伞织物透气性。

2.2　盘缝带伞及流场域模型

2.2.1　流固耦合仿真模型

伞衣部分有限元使用四边形膜单元, 选用 LS-DYNA 材料库中的织物材料(FABRIC)关键字对膜单元定义材料属性, 织物材料视为各向同性材料; 伞绳、纬向加强带和收口绳有限元使用索单元。具体材料属性如表 3 所示。

表 3　材料属性

Table 3　Material property

结构名称	密度/(kg·m^{-3})	弹性模量/Pa	泊松比
伞衣	538	4.3×10^8	0.14
纬向加强带	1 194	1.5×10^{10}	
伞绳	468	1.5×10^{10}	
收口绳	800	5×10^9	

数值模拟中, 主要研究伞衣部分的动态变化, 忽略伞绳气动力, 且设置伞绳与流场不发生耦合, 因此, 流场域部分主要覆盖伞衣部分。作者在前期研究中, 已对相同大小的盘缝带伞和流场域进行了网格准确性验证, 发现当流场域网格大小为 0.02 m 时, 计算结果已收敛, 流场可以精确地捕捉盘缝带伞顶孔的高速射流和伞后的非定常尾流[9]。本文选择单位长度为 0.015 m 的网格, 流场域沿盘缝带伞的轴向、法向、侧向的长度分别为 2 m、1.6 m、1.6 m, 约 150 万网格, 具体流场域和盘缝带伞模型及两者位置关系如图 5 所示。

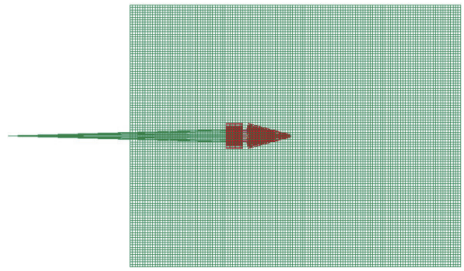
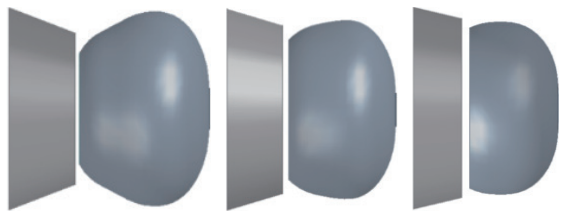


图 5　流场域及盘缝带伞模型

Fig. 5　Fluid and disk-gap-band parachute model

2.2.2　计算流体力学仿真模型

在计算流体力学计算中, 取 2.3.1 节得到的盘缝带伞稳态下的平均气动外形, 做成如图 6 所示的刚性不透气外壳, 并对该模型进行三维结构化网格建模, 壁面处采用加密处理, 最终形成约 200 万的流场域网格。



(a) 15%收口比　(b) 25%收口比　(c) 35%收口比

图 6　盘收口盘缝带伞平均气动外形

Fig. 6　Average aerodynamic shape of mid-gore reefing disk-gap-band parachute

2.3　仿真工况

以火星低密度环境为应用背景, 大气模型取自火星科学实验室[21]和火星全球勘探者得到的火星大气参数拟合公式[22], 计算距火星地面 8 500 m 高空的大气环境条件, 具体参数如表 4 所示。

表 4　仿真使用来流条件

Table 4　Freestream condition for simulation

马赫数	来流速度/(m·s^{-1})	静压/Pa	密度/(kg·m^{-3})	温度/K
1.8	429.13	330.56	0.007 44	231.28
2.2	524.50	330.56	0.007 44	231.28
2.6	619.85	330.56	0.007 44	231.28

对于火星探测, 主要要求盘缝带伞在 16～4 km 内, 具备开伞速度为马赫数 1.5～2.3、动压为 250～850 Pa 的宽域适应能力[8]。为更好地覆盖上述条件, 马赫数范围为 1.8～2.6, 对马赫数为 1.8、2.2、2.6 时不同收口比情况进行研究。

3　仿真模型验证

作者在前期研究中, 使用相同仿真模型进行数值模拟, 并从阻力系数、伞衣充满过程的外形变化、投影等方面与已公开数据进行对比, 证明了该

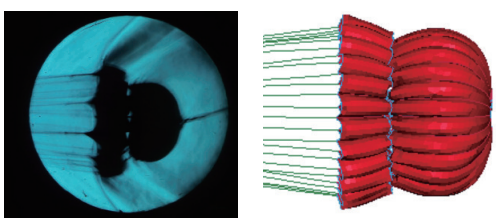
数值模拟方法有效，计算结果可靠[9]。

北京空间机电研究所曾进行过盘收口盘缝带伞的风洞试验。以该风洞试验为对标工况，对收口比为25%的盘收口盘缝带伞在不同马赫数下进行数值模拟。从盘缝带伞的阻力系数、稳定充气外形2方面进行对比。如表5所示，对风洞试验所得的阻力系数和数值模拟结果进行对比，发现阻力系数的误差均在10%以下，数值模拟所得结果较为准确。图7(a)和图7(b)分别为风洞试验和数值模拟得到的盘收口盘缝带伞稳定时的侧视图。数值模拟结果显示，盘缝带伞的伞盘部分收束于收口绳处，呈饱满的半圆形，伞带部分向外扩张，与风洞试验结果一致。且由于收口绳的限制，伞盘底部向内收缩，伞带部分向外扩张，导致侧视图中的盘与带的间隙极小，而该特点在数值模拟的结果中可以清晰地观察到。因此，从稳定外形对比，数值模拟结果与风洞试验结果一致。

表5 阻力系数对比

Table 5 Comparison of drag coefficients

马赫数	阻力系数		误差/%
	试验结果	数值模拟结果	
2.0	0.213 44	0.194 64	8.81
2.5	0.214 84	0.196 62	8.48



(a) 风洞试验外形 (b) 数值模拟外形

图7 盘收口盘缝带伞外形对比

Fig. 7 Shape comparison of mid-gore reefing disk-gap-band parachute

4 盘缝带伞气动性能

4.1 投影面积比随收口比的变化

投影面积比是描述盘缝带伞充气形状的重要参数，定义为在相同环境下收口降落伞展开后垂直于轴向最大的投影面积与不收口降落伞完全展开后垂直于轴向的最大投影面积的比值。理论上，投影面积比与收口比呈二次方关系。

图8为30%收口比下盘缝带伞在马赫数2.2开伞时投影面积随时间的变化曲线。可见，盘缝带伞在开伞时达到了最大的投影面积，在经历4次较大的震荡后，投影面积趋于稳定。对比不同工况的投影面积，取0.48～0.68 s稳定阶段的平均投影面积。

计算得到不同马赫数、不同收口比下的投影面积比，拟合得到投影面积比G与收口比X的函数关系如式(9)所示，关系曲线如图9所示。

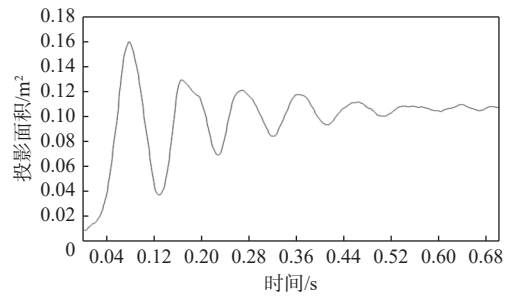


图8 投影面积与时间的关系

Fig. 8 Relation between projected area and time

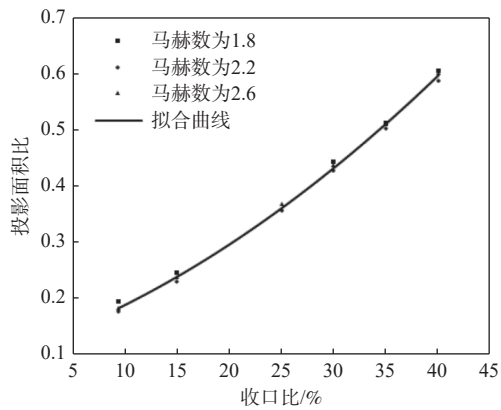


图9 投影面积比与收口比的关系

Fig. 9 Relation between projected area ratio and reefing ratio

$$G = 0.107\,88 + 0.652\,72X + 1.424\,02X^2 \tag{9}$$

数值模拟结果表明，不同马赫数下，投影面积比与收口比呈二次方关系；投影面积比与收口比的关系主要受收口比变化的影响，基本与马赫数无关。

4.2 阻力系数比随收口比的变化

建立准确的阻力系数比随收口比的变化曲线是工程师进行收口绳结构设计的前提，本节使用流固耦合仿真方法对该函数关系进行研究。

盘缝带伞的阻力系数定义为

$$C_D = \frac{D}{qS_0} \tag{10}$$

式中：C_D为盘缝带伞阻力系数；D为盘缝带伞阻力；q为盘缝带伞动压；S_0为盘缝带伞名义面积。

定义阻力系数比为伞衣收口状态的阻力系数与完全张满状态的阻力系数之比。根据数值模拟结果拟合得到阻力系数比K与收口比X的函数关系为

$$K = 0.076\,64 + 0.677\,25X + 1.118\,77X^2 \tag{11}$$

风洞试验获取了不同马赫数、不同收口比下的阻力系数试验结果, 将数值模拟得到的拟合曲线与风洞试验结果进行对比, 如图 10 所示。可见, 仿真结果与风洞试验结果基本一致。分析结果表明, 不同马赫数下阻力系数比与收口比呈二次方关系; 阻力系数比与收口比的关系主要受收口比变化的影响, 基本与马赫数无关; 由式 (11) 推算, 在收口比为 55% 时, 阻力系数比可达到 0.8 左右, 基本可认为 55% 收口比是有工程意义的盘收口盘缝带伞的收口比上限。

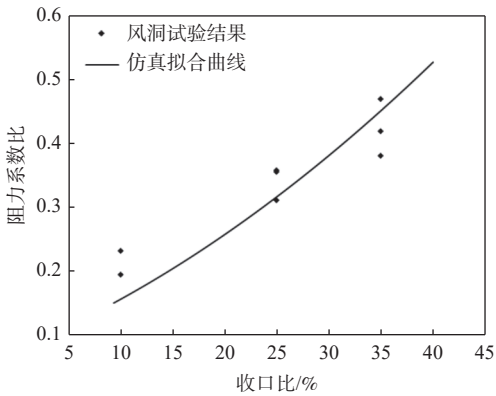


图 10　阻力系数比与收口比的关系

Fig. 10　Relation between drag coefficients ratio and reefing ratio

4.3　收口绳载荷和盘缝带伞总阻力的关系

准确预估收口绳载荷是工程师选择收口绳材料的前提, 本节使用流固耦合仿真方法进行研究。

对收口绳进行受力分析, 如图 11 所示, 其中, F_D 为盘缝带伞的总阻力, F_r 为收口绳的载荷。此时, 以伞盘和伞绳的交界点为研究对象, 如图 12(a) 所示。其中, F_R 为收口绳在水平面上的合力, F_{cd} 为伞盘部分径向带上的载荷, F_{ml} 为中间伞绳的张力载荷。以伞带部分受力为研究对象, 如图 12(b) 所示, 其中, F_b 为伞带部分所受合力, F_{dl} 为每根伞绳所受分力。

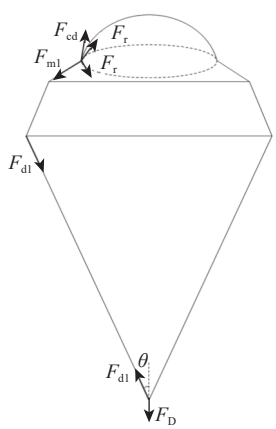


图 11　盘收口盘缝带伞受力图

Fig. 11　Mid-gore reefing disk-gap-band parachute force diagram

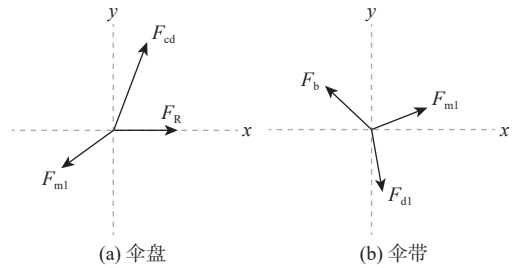


图 12　局部受力图

Fig. 12　Local stress diagram

在盘缝带伞的各类试验中, 由于很难在收口绳安装测力传感器, 因此, 收口绳载荷很难直接测量, 但盘缝带伞所受到的总阻力往往可以由拉力传感器直接测量。如果能够建立收口绳载荷与总阻力的关系式, 可以在工程上根据阻力实测值方便地估算收口绳载荷。采用流固耦合数值模拟方法, 得到不同收口比下, 马赫数为 2.2 和 2.6 的载荷 H 与收口比 X 的关系式 (12), 拟合曲线如图 13 所示。

$$H = 0.366\,6 - 0.477\,84X \tag{12}$$

结果表明, 收口绳载荷与盘缝带伞总阻力比值随收口比的增大呈线性减小趋势, 且基本与马赫数无关。该研究结果可以为收口绳设计材料的选择提供依据。

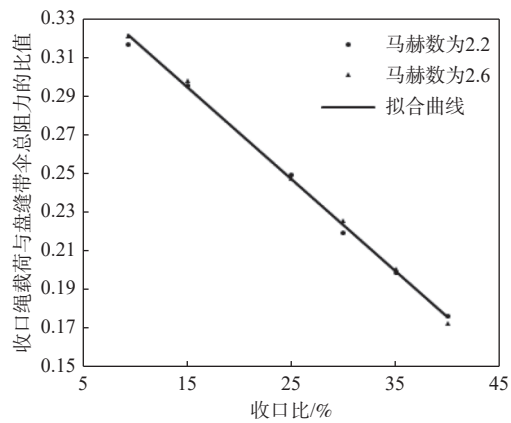


图 13　载荷比与收口比的关系

Fig. 13　Relation between load ratio and reefing ratio

4.4　不同收口比稳定性分析

稳定性是降落伞的重要特性之一, 某些应用场景下, 对降落伞的稳定性有特殊要求。为探究收口状态下盘缝带伞的稳定性, 对马赫数为 2.2 时收口比为 15%、25%、35% 盘收口盘缝带伞在攻角为 0°、3°、6°、9°、12°情况下使用计算流体力学方法进行分析。

在仿真结果中, 定义垂直于盘缝带伞几何轴线的力为法向力 N, 法向力系数为 C_N; 沿盘缝带伞几何轴线的力为轴向力 T, 轴向力系数为 C_T; 盘缝带伞所受力矩为 M, 其中, 力臂为 L, 力矩系数为 C_M,

力矩作用点为伞绳集束点。法向力、轴向力、力矩系数的计算公式如式(13)～式(15)所示。盘缝带伞的阻力为轴向力和法向力的合力,具体法向力、轴向力、力矩系数方向如图14所示。

$$C_N = \frac{N}{qS_0} \tag{13}$$

$$C_T = \frac{T}{qS_0} \tag{14}$$

$$C_M = \frac{M}{qS_0L} \tag{15}$$

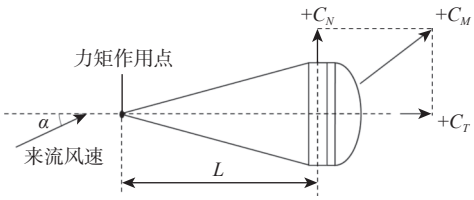


图14 盘缝带伞力矩方向示意图

Fig. 14 Diagram of moment direction of disk-gap-band parachute

取计算流体力学计算稳定后0.04～0.1 s的平均数,得到不同收口比下的轴向力系数、力矩系数和力矩系数导数,这些气动参数与攻角的关系如图15～图17所示。

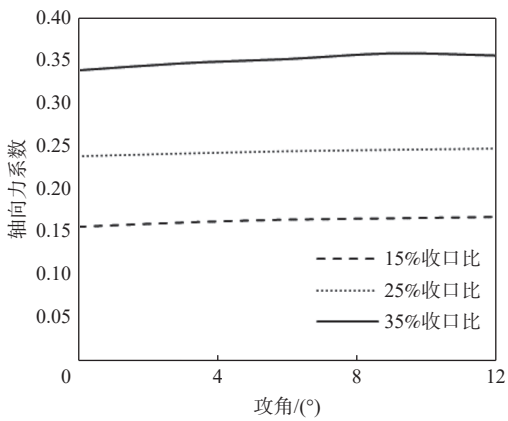


图15 轴向力系数与攻角的关系

Fig. 15 Relation between axial force coefficient and angle of attack

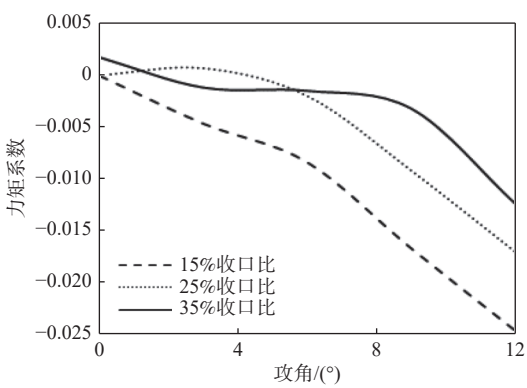


图16 力矩系数与攻角的关系

Fig. 16 Relation between moment coefficient and angle of attack

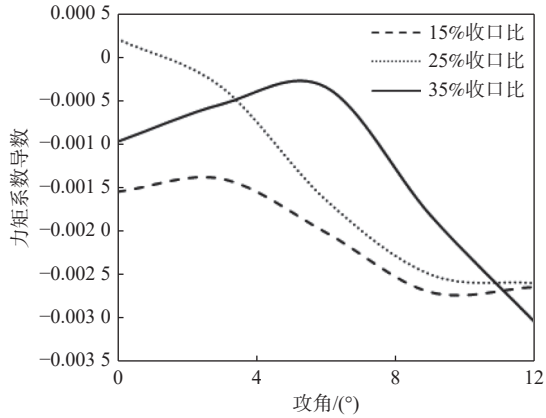


图17 力矩系数导数与攻角的关系

Fig. 17 Relation between the derivative of the moment coefficient and the angle of attack

由图15看出,在0°～12°范围内,随攻角的增大,轴向力系数并没有明显变化。

由图16和图17看出,盘收口盘缝带伞的配平攻角在6°左右,即0°～6°时,降落伞的力矩系数接近0,在小攻角时,收口盘缝带的回复力矩很小,不能有效抑制盘缝带伞的摆动。在攻角为6°～12°时,盘收口盘缝带伞的力矩系数负向增加迅速,说明大攻角时,降落伞能够提供较大的回复力矩,有效抑制攻角的增加。在来流扰动下,盘收口盘缝带伞的摆角在6°左右振荡。还可以看出,收口比越小,盘缝带伞的稳定性越好。

为了解释小攻角情况下,力矩系数的变化特点,对比计算流体力学方法测定的收口比为35%的情况下,攻角为3°和12°的力矩系数与时间的关系,如图18所示。可以看到,在小攻角情况下,盘缝带伞的力矩不稳定,力矩系数震动剧烈。而在大攻角的情况下,初始状态过后,力矩系数较稳定。在其他收口比下,该情况依然明显。选取收口比为35%的情况下,攻角为3°和12°时0.055～0.07 s的速度云图及其流场的流线图进行分析,如图19所示。

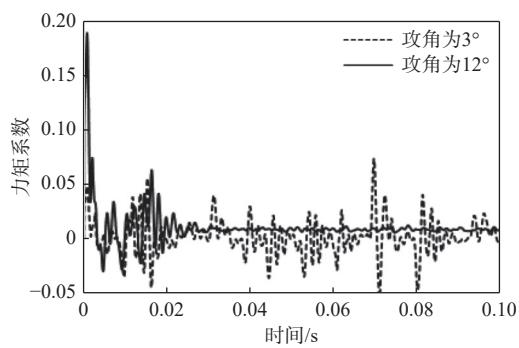
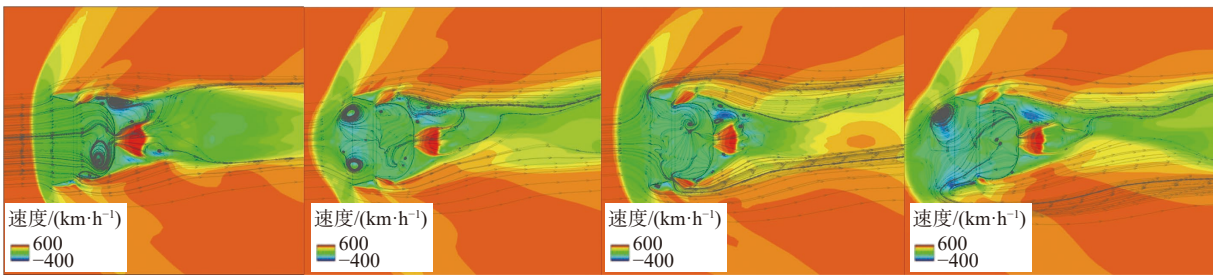


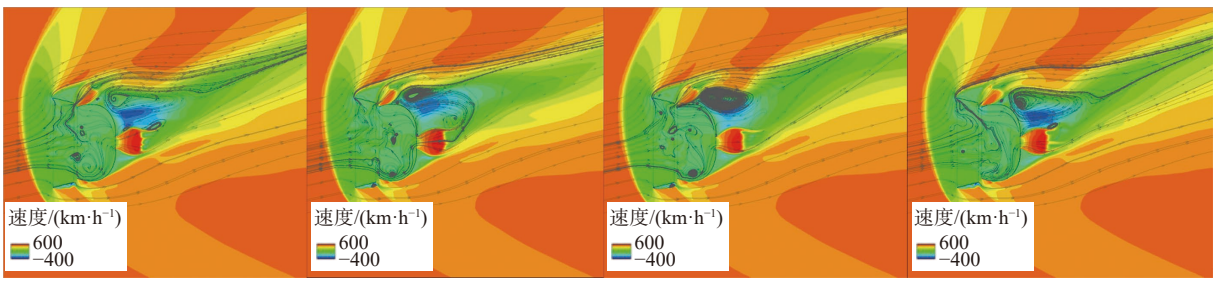
图18 力矩系数随时间变化曲线

Fig. 18 Curve of moment coefficient with time

由图19可以观察到,在小攻角情况下,伞衣对高速气流起阻滞作用,更多气体涌向伞缝处,伞内



(a) 攻角为3°



(b) 攻角为12°

图 19　速度云图

Fig. 19　Velocity cloud map

的流线呈纵向和横向交替变化, 导致伞内壁附近有涡产生, 且涡量较大。涡的产生和消散都会导致伞内横向力分布不均匀, 导致盘缝带伞的轴向力始终处于不稳定的状态。另外, 盘缝带伞背风面尾流部分始终进行从 1 个附着涡分裂成 2 个涡, 再逐渐组合为 1 个涡的周期性运动。这也导致流场不稳定, 侧向力系数波动较大。但在大攻角情况下, 流场始终处于较稳定的状态, 盘缝带伞内涡远离伞内壁, 涡量小且位置分散, 背风面的上游部分形成一个稳定的回流区, 且随着时间的推进, 回流区始终较为稳定。小攻角法向力系数的震荡情况说明, 小攻角处于不稳定的状态, 可能导致伞衣充气形状震荡, 且伞系快速摆动。

5　结　论

本文基于流固耦合方法, 得到了不同马赫数和不同收口比情况下, 投影面积比、阻力系数比、收口绳载荷与收口比的关系。对稳态外形进行分析, 得到不同攻角下的稳定性结果。研究结果可指导工程设计, 也为类似问题提供了分析思路。结论如下:

1) 通过对盘缝带伞进行盘收口, 可以实现盘缝带伞阻力、稳定性等性能调整, 具有较强的工程实用价值。

2) 投影面积比和阻力系数比与收口比呈二次方关系, 与马赫数基本无关。

3) 收口绳载荷和盘缝带伞总阻力的比值与收口比呈线性减小的关系, 与马赫数基本无关。

4) 盘收口盘缝带伞的配平攻角基本在 6°左右, 一般地, 收口比越小, 稳定性越好。

参考文献（References）

[1] 《降落伞技术导论》编写组. 降落伞技术导论[M]. 北京: 国防工业出版社, 1977: 113-117.
Introduction to Parachute Technology Editorial Group. Introduction to parachute technology[M]. Beijing: National Defense Industry Press, 1977: 113-117(in Chinese).

[2] WITKOWSKI A, BRUNO R. Mars exploration rover parachute decelerator system program overview[C]//Proceedings of the 17th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar. Reston: AIAA, 2003: 2100.

[3] ADAMS D S, WITKOWSKI A, KANDIS M. Phoenix Mars scout parachute flight behavior and observations[C]//Proceedings of the 2011 Aerospace Conference. Piscataway: IEEE Press, 2011: 1-8.

[4] CRUZ J R, WAY D W, SHIDNER J D, et al. Reconstruction of the Mars science laboratory parachute performance[J]. Journal of Spacecraft and Rockets, 2014, 51(4): 1185-1196.

[5] STEINBERG S Y, SIEMERS P M, SLAYMAN R G. Development of the Viking parachute configuration by wind-tunnel investigation[J]. Journal of Spacecraft and Rockets, 1974, 11(2): 101-107.

[6] 李健, 房冠辉, 吕智慧, 等. 天问一号火星探测器伞系减速分系统设计与验证[J]. 中国科学 (技术科学), 2022, 52(2): 264-277.
LI J, FANG G H, LYU Z H, et al. Design and verification of parachute deceleration subsystem of Tianwen-1 Mars probe[J]. Scientia Sinica (Technologica), 2022, 52(2): 264-277(in Chinese).

[7] MUPPIDI S, O'FARRELL C, VAN NORMAN J W, et al. ASPIRE aerodynamic models and flight performance[C]//Proceedings of the AIAA Aviation 2019 Forum. Reston: AIAA, 2019: 3376.

[8] SONNEVELDT B S, CLARK I G, O'FARRELL C. Summary of the advanced supersonic parachute inflation research experiments (ASPIRE) sounding rocket tests with a disk-gap-band parachute [C]//Proceedings of the AIAA Aviation 2019 Forum. Reston: AIAA, 2019: 3482.

[9] 代雨柔, 李健, 薛晓鹏, 等. 超声速下盘缝带伞不同收口方式的气动特性[J]. 航空学报, 2024, 45(7): 68-80.
DAI Y R, LI J, XUE X P, et al. Aerodynamic characteristics of supersonic disk-gap-band parachute with different reefing ways[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(7): 68-80(in Chinese).

[10] PREISSER J S, GROW R B. High-altitude flight test of a reefed 12.2-meter-diameter disk-gap-band parachute with deployment at a mach number of 2.58[M]. Washington, D.C.: National Aeronautics and Space Administration, 1971: 20-25.

[11] WITKOWSKI A, KANDIS M. Reefing the Mars science laboratory parachute[C]//Proceedings of the 2010 IEEE Aerospace Conference. Piscataway: IEEE Press, 2010: 1-6.

[12] 林斌, 吴卓. 降落伞收口技术[J]. 航天返回与遥感, 2012, 33(4): 1-12.
LIN B, WU Z. Reefing technology of parachute[J]. Aerospace Return and Remote Sensing, 2012, 33(4): 1-12(in Chinese).

[13] WITKOWSKI A, KANDIS M, REUTER J, et al. Design of subscale parachute models for MSL supersonic wind tunnel testing[C]//Proceedings of the 20th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar. Reston: AIAA, 2009: 2999.

[14] 辛春亮, 朱星宇, 王凯, 等. LS-DYNA 有限元建模、分析和优化设计[M]. 北京: 清华大学出版社, 2022.
XIN C L, ZHU X Y, WANG K, et al. Finite element modeling analysis and optimization design of LS-DYNA[M]. Beijing: Tsinghua University Press, 2022(in Chinese).

[15] 刘凯欣, 王景焘, 王刚, 等. 时-空守恒元解元 (CE/SE) 方法综述[J]. 力学进展, 2011, 41(4): 447-461.
LIU K X, WANG J T, WANG G, et al. A review on the CE/SE method[J]. Advances in Mechanics, 2011, 41(4): 447-461(in Chinese).

[16] CHANG S C. The method of space-time conservation element and solution element: a new approach for solving the Navier-Stokes and Euler equations[J]. Journal of Computational Physics, 1995, 119(2): 295-324.

[17] CHANG S C, WANG X Y, CHOW C Y. The space-time conservation element and solution element method: a new high-resolution and genuinely multidimensional paradigm for solving conservation laws[J]. Journal of Computational Physics, 1999, 156(1): 89-136.

[18] 刘海涛, 徐建中. 求解 Euler 方程的空间-时间守恒格式[J]. 工程热物理学报, 1997, 18(3): 294-299.
LIU H T, XU J Z. A space-time conservation scheme for solving the two-dimensional Euler equations[J]. Journal of Engineering Thermophysics, 1997, 18(3): 294-299(in Chinese).

[19] 张增产, 沈孟育. 改进的时空守恒元和解元方法[J]. 清华大学学报 (自然科学版), 1997, 37(8): 65-68.
ZHANG Z C, SHEN M Y. Improved scheme of space time conservation element and solution element[J]. Journal of Tsinghua University (Science and Technology), 1997, 37(8): 65-68(in Chinese).

[20] 徐欣, 贾贺, 陈雅倩, 等. 织物透气性对火星用降落伞气动特性影响机理[J]. 航空学报, 2022, 43(12): 301-318.
XU X, JIA H, CHEN Y Q, et al. Influence mechanism of fabric permeability of canopy on aerodynamic performance of Mars parachute [J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(12): 301-318 (in Chinese).

[21] 夏元清. 火星探测器进入、下降与着陆过程的导航、制导与控制——“恐怖”七分钟[M]. 北京: 科学出版社, 2017: 34-38.
XIA Y Q. Navigation, guidance and control of Mars rover entry, descent and landing—seven minutes of terror[M]. Beijing: Science Press, 2017: 34-38(in Chinese).

[22] HALL N. Mars atmosphere model[EB/OL]. (2021-05-13) [2021-07-27]. https://www.grc.nasa.gov/www/k-12/airplane/atmosmrm.html.

Aerodynamic characteristics of supersonic mid-gore reefing disk-gap-band parachute under different reefing ratios

DAI Yurou[1, 2], LI Jian[1, 2, *], RONG Wei[1, 2], XUE Xiaopeng[3], HUANG Mingxing[1, 2], FANG Kang[1, 2]

(1. Beijing Institute of Space Mechanics and Electricity, Beijing 100094, China;
2. Space Entry Deceleration and Landing Technology Laboratory, China Aerospace Science and Technology Corporation Limited, Beijing 100094, China;
3. School of Automation, Central South University, Changsha 410083, China)

Abstract: The increasing the mass of Mars landing probes requires enlarging the parachute area to ensure stable deceleration performance while controlling the parachute opening force to avoid excessive structural weight penalties on the probe. One of the effective ways to meet the above needs is the reefing disk-gap-band parachute. In this article, the functional link between the reefing ratio and the resistance coefficient ratio, projection area ratio, and reefing rope load is obtained using fluid-structure coupling simulation analysis based on the Mars environment. The stability characteristics of the reefing disk-gap-band parachute were studied by computational fluid dynamics simulation. The research of this paper shows that the disk-gap-band parachute can achieve the controlled change of resistance performance, and can provide stable performance to meet the engineering application.

Keywords: disk-gap-band parachute; reefing technology; mid-gore reefing; drag characteristics; stability

Received: 2023-12-13; **Accepted**: 2024-03-27; **Published Online**: 2024-04-03 18:43
URL: link.cnki.net/urlid/11.2625.V.20240403.1145.006
Foundation items: National Natural Science Foundation of China (12072377); Hunan Provincial Natural Science Foundation of China (2022JJ30678)
* **Corresponding author**. E-mail: lijian_bbmouth@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0814

丙烯工质平板环路热管的稳态建模与特性分析

贾志超[1]，毕涵礼[1]，彭正[2]，李国广[1]，吴琪[1]，张红星[1,*]，苗建印[1]

(1. 北京空间飞行器总体设计部 航天器热控全国重点实验室，北京 100094；
2. 北京航空航天大学 航空科学与工程学院，北京 100191)

摘　　要：丙烯工质平板环路热管(LHP)具有耐低温、质量轻等优势，是解决深空探测任务热控难题的关键技术。丙烯工质平板环路热管的工作性能和特性亟需掌握。基于此，建立丙烯工质新型平板环路热管的稳态模型，基于模型分析其最大传热能力和流阻特性，指出提升传热性能的主要途径。改进储液器容积和工质充装质量的计算方法，提出一种带有副储液器的环路热管设计方案，可在拓宽工作温区的同时减小质量和体积。

关 键 词：丙烯；环路热管；平板蒸发器；建模分析；空间热控
中图分类号：V444.3^{+}6
文献标志码：A　　**文章编号**：1001-5965(2026)03-0874-11

环路热管(loop heat pipe, LHP)是一种依靠毛细力驱动的高效两相传热装置，可解决航天器大功率、远距离、逆重力热传输等诸多热控难题，已广泛应用于各国航天器。

随着我国空间技术的发展，对月球极区和太阳系边际的科学探测将相继开展。此类深空探测任务面临能源紧缺、辐射器长期低温(≤−100 ℃)、质量要求苛刻等热控难题，热控装置须具备良好的低温适应性，降低对航天器供电和遥测资源的需求。由于丙烯的冰点足够低(−185 ℃)，采用丙烯工质的环路热管无需补偿加热即可实现长期低温存储和运行，有望解决上述热控难题。

此前，俄罗斯和美国的多个地球轨道航天器应用丙烯环路热管为载荷散热或控温[1-4]，实现了节约能源和降低冻结风险的目的。在深空探测领域，俄罗斯的Mars 96[5]采用丙烯环路热管为推进系统散热；欧洲提出的欧洲火星车[6]和美国提出的国际月球网络锚节点[7]均计划使用丙烯环路热管作为热开关解决设备昼间散热、夜间保温的问题。

在建模分析方面，Pauken[8]和Mishkinis[6]等分别建立了一维稳态模型，分析了丙烯环路热管的地面运行特性。闫凯芬[9]建立了基于实验关联式的稳态模型，并利用可视化手段研究丙烯环路热管在低温区的特性。

近年来，新型平板蒸发器结构的环路热管技术逐渐成为研究热点。平板蒸发器可直接与热源贴合，无需鞍座，使其具有热阻低、质量轻、体积小等优点，更适合航天器应用。Maydanik等[10]研制了可为电子设备高效散热的平板环路热管，使用水工质可实现900 W传热能力。Yang等[11]研制了纵向补液平板环路热管，研究其传热性能和启动特性。Zhang等[12]研制了用于航天器热控的平板环路热管，并完成了搭载飞行试验。平板环路热管已展现出优秀的传热性能，且技术趋于成熟。因此，基于平板环路热管技术发展丙烯工质环路热管更有利于解决深空探测任务中的热控难题。

本文建立丙烯工质平板环路热管的稳态模型，可准确预测稳态工作温度，结合实验结果修正丙烯

收稿日期：2023-12-14；录用日期：2024-03-08；网络出版时间：2024-03-18 21:50
网络出版地址：link.cnki.net/urlid/11.2625.V.20240316.1501.004
基金项目：国家自然科学基金(U21B2084)
*通信作者. E-mail：redlincoco@hotmail.com

引用格式：贾志超，毕涵礼，彭正，等. 丙烯工质平板环路热管的稳态建模与特性分析[J]. 北京航空航天大学学报，2026，52(3)：874-884. JIA Z C，BI H L，PENG Z，et al. Steady state modeling and characteristic analysis of propylene flat loop heat pipes[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：874-884 (in Chinese).

与镍粉毛细芯的接触角，分析丙烯平板环路热管的全温区最大传热量、流阻特性及热沉温度对工作温度的影响，指出提高最大传热量的具体优化方法。此外，改进储液器容积和工质充装质量的计算方法，提出一种带有副储液器的环路热管设计方案，在拓宽工作温区的同时，可减小质量和体积，为宽温区环路热管设计提供参考。

1　平板环路热管构成及工作原理

平板环路热管一般由蒸发器、蒸气/液体管路、冷凝器和储液器组成，如图 1 所示。平板环路热管的最主要特征是具有扁平的蒸发器：蒸发器壳体、毛细芯、吸液芯等均为平板结构。大多数热控对象都以平面散热，传统的圆柱蒸发器需要在壳体与热源间设置鞍座实现热连接。鞍座本身具有一定热阻，同时具有较大的质量和体积。平板蒸发器可直接与热源贴合，降低了热阻和质量。

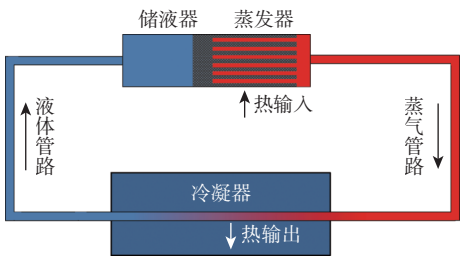


图 1　平板环路热管的结构

Fig. 1　The construction of the flat loop heat pipe

平板环路热管的工作原理与圆柱环路热管类似，仅储液器与蒸发器间的流动过程存在区别。平板蒸发器一般不具有液体引管，由液体管路回流的工质直接进入储液器，储液器对蒸发器的供液则靠吸液芯和毛细芯的抽吸作用实现。工质的相变过程及在蒸发器以外管路内的流动过程与圆柱环路热管相同。如图 1 所示，液体工质在蒸发器受热变为蒸气，蒸气沿蒸气管路流动进入冷凝器放热冷凝为液体，并沿液体管路流回储液器及蒸发器，环路热管即利用此循环实现传热。

2　平板环路热管建模

建立各部件的一维稳态数学模型，以此为基础建立平板环路热管的整体稳态模型。部件的数学模型包括蒸发器、储液器、冷凝器、蒸气/液体管路。蒸气/液体管路内的计算按单相流处理，管路与环境的换热按自然对流考虑。

2.1　部件数学模型

2.1.1　蒸发器与储液器

横向供液型平板蒸发器结构及流场如图 2 所示。储液器内壁及毛细芯（主芯）表面铺有不锈钢丝网，作为吸液芯（副芯）实现储液器至毛细芯的水力连接。工质在蒸发器与储液器内的传热和流动过程较为复杂，建模基于如下假设：①储液器内工质处于同一温度；②蒸发器对储液器的漏热为交界面处换热；③吸液芯和毛细芯内均充满液体工质，蒸气槽内充满蒸气；④毛细芯内工质自上表面至蒸气槽的流动是一维的，且流速均匀，吸液芯及蒸气槽道内工质也是一维流动，质量流量沿流动方向均匀变化；⑤蒸发器受热面受到均匀热流的加热，液态工质在蒸气槽表面蒸发，工质蒸发后为饱和态。

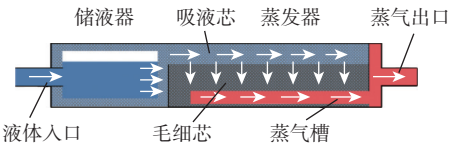


图 2　平板蒸发器结构及流场示意图

Fig. 2　The construction and flow field diagram of the flat evaporator

将毛细芯和壳体对吸液芯及储液器的漏热视为将储液器液体入口的过冷态工质加热至饱和态所需的热流。因此，蒸发器的能量平衡方程为

$$Q_{\text{in}} = Q_{\text{cond}} + Q_{\text{hl}} + Q_{\text{hw}} + Q_{\text{ev-am}} \tag{1}$$

式中：Q_{in} 为施加到蒸发器上的热负载；Q_{cond} 为冷凝带走的热量；Q_{hl} 为蒸发器对储液器的漏热；Q_{hw} 为用于加热储液器来流的工质；$Q_{\text{ev-am}}$ 为蒸发器与环境的换热。

Q_{hl} 可分为 2 部分：①热流将储液器入口的液体加热至储液器温度；②热流经壳体向环境漏热。由此，可将式 (1) 改写为式 (2)。模型由式 (2) 计算工质的质量流量。

$$Q_{\text{in}} = \dot{m}\left[h_{\text{fg}} + c_{\text{p,l}}\left(T_{\text{sat}} - T_{\text{cc,in}}\right)\right] + Q_{\text{ev-am}} + Q_{\text{cc-am}} \tag{2}$$

式中：$\dot{m}$ 为工质的质量流量；T_{sat} 和 $T_{\text{cc,in}}$ 分别为工作温度（工质的饱和温度）和储液器入口工质温度；h_{fg} 和 $c_{\text{p,l}}$ 分别为工质在饱和温度下的气液相变潜热和在特征温度为 $\left(T_{\text{sat}} + T_{\text{cc,in}}\right)/2$下的液体定压比热容；$Q_{\text{cc-am}}$ 为储液器与环境的换热。

毛细芯和吸液芯均为多孔结构，用达西定律计算流动压降：

$$\Delta P_{\text{w}} = \frac{\dot{m}\mu_{\text{l}} L}{\rho_{\text{l}} A_{\text{w}} K} \tag{3}$$

式中：ΔP_{w} 为毛细芯内液体工质的流动压降；μ_{l} 为液体工质的动力黏性系数；ρ_{l} 为液体工质的密度；L 为流动距离（主芯厚度或副芯长度）；A_{w} 为流通截面积（主、副芯截面积）；K 为渗透率。根据假设，副芯的计算长度为实际值的 1/2。

蒸气槽道内的压降计算与蒸气管路一致，蒸气的质量流量被 n 个蒸气槽平均分配，质量流量沿长度方向向出口处线性增加，蒸气槽计算长度为实际值的 1/2。蒸气槽至蒸气管路入口的压降按突扩或突缩导致局部损失计算。

通过计算平板环路热管蒸发器和储液器 2 个腔室界面处的换热来计算该漏热。漏热的换热形式可以分为相变换热、热对流、热传导等。在重力环境下，蒸发器与储液器处于水平姿态时，储液器内工质的气液分布呈现上层为蒸气、下层为液体的状态。液体工质受毛细作用进入吸液芯，一部分液体会在吸液芯上受热蒸发，蒸气回到储液器腔内，在温度较低处放热冷凝；另一部分液体则沿副芯流动，最终进入主芯，这些液体与吸液芯存在对流/导热等换热形式。此外，蒸发器壳体对储液器壳体的漏热也将传递给储液器内的工质。在稳态下，式 (4) 始终成立。

$$\dot{m}c_{\mathrm{p,l}}\left(T_{\mathrm{cc}}-T_{\mathrm{cc,in}}\right)=h_{\mathrm{ev\text{-}cc}}A_{\mathrm{hl}}\left(T_{\mathrm{sat}}-T_{\mathrm{cc}}\right)-Q_{\mathrm{cc\text{-}am}} \tag{4}$$

式中：$h_{\mathrm{ev\text{-}cc}}$ 为蒸发器对储液器漏热的换热系数，其值决定了储液器的温度；T_{cc} 为储液器工质温度；A_{hl} 为蒸发器与储液器间的换热面积。

根据丁汀等[13]的实验结论和 Zhou 等[14]的理论分析，蒸发器对储液器的漏热具有如下特性：在可变热导区内，漏热以相变形式为主，$h_{\mathrm{ev\text{-}cc}}$ 较大，并随流量增大而降低；在固定热导区内，储液器充满液体，漏热以液体对流和壳体导热为主，$h_{\mathrm{ev\text{-}cc}}$ 较小。

根据 $h_{\mathrm{ev\text{-}cc}}$ 的特性，提出一种简便的拟合方法，以储液器内气体和液体工质的体积分数对蒸发器和储液器间的相变换热和对流/导热换热分别进行加权，得到 $h_{\mathrm{ev\text{-}cc}}$ 的表达式为

$$h_{\mathrm{ev\text{-}cc}}=C_1\alpha_{\mathrm{cc}}+(1-\alpha_{\mathrm{cc}})C_2 \tag{5}$$

式中：C_1 和 C_2 分别为两相漏热和液体对流/壳体导热的等效换热系数；α_{cc} 为储液器内气体工质的体积占比，其计算方法为

$$\alpha_{\mathrm{cc}}=\frac{m^{*}-m}{\left(\rho_{\mathrm{l,cc}}-\rho_{\mathrm{v,cc}}\right)V_{\mathrm{cc}}} \tag{6}$$

式中：m*和 m 分别为模型计算得到的工质充装质量和实际充装质量；$\rho_{\mathrm{l,cc}}$ 和 $\rho_{\mathrm{v,cc}}$ 分别为储液器内液体和气体工质的密度，取储液器内饱和温度下的值；V_{cc} 为储液器容积。

2.1.2 冷凝器

冷凝器的计算分为过热区、两相区、过冷区 3 部分。仅当工作温度低于环境温度时，冷凝器存在过热区，过热蒸气进入冷凝器，先放热变为饱和蒸气；在两相区内，工质为饱和态，饱和蒸气随着冷凝放热逐渐转变为饱和液；在过冷区内，工质已完全冷凝为液态，并进一步放热降温变为过冷液。

1）两相区。两相区内的能量平衡方程为

$$-Q_{\mathrm{cond}}\mathrm{d}x=(UA/L)^{\mathrm{tp}}_{\mathrm{cond\text{-}sink}}\left(T_{\mathrm{sat}}-T_{\mathrm{sink}}\right)\mathrm{d}l \tag{7}$$

式中：dl 和 dx 分别为两相区微元长度和该长度内的干度变化量；T_{sink} 为热沉温度；$(UA/L)^{\mathrm{tp}}_{\mathrm{cond\text{-}sink}}$ 为两相区内冷凝器和热沉在单位长度上的热导，计算式为

$$(UA/L)^{\mathrm{tp}}_{\mathrm{cond\text{-}sink}}=1\Big/\left(\frac{1}{h_{\mathrm{tp}}\pi D_{\mathrm{cond}}}+\frac{1}{(UA/L)_{\mathrm{p\text{-}sink}}}\right) \tag{8}$$

式中：h_{tp} 为冷凝换热系数，由 Cavallini 分ΔT模型[15]计算；D_{cond} 为冷凝器管路内径；$(UA/L)_{\mathrm{p\text{-}sink}}$ 为冷凝器管壁至热沉在单位长度上的热导。

将式 (8) 代入式 (7)，并将式 (7) 对干度做积分可得两相区长度 L_{tp}：

$$L_{\mathrm{tp}}=\frac{Q_{\mathrm{cond}}}{T_{\mathrm{sat}}-T_{\mathrm{sink}}}\int_0^1\frac{1}{(UA/L)^{\mathrm{tp}}_{\mathrm{cond\text{-}sink}}}\mathrm{d}x \tag{9}$$

两相区内的流阻采用 Muller-Steinhagen-Heck 全液相摩擦压降倍率模型[16]计算，全液相摩擦压降倍率ϕ^2_{lo}定义为

$$\phi^2_{\mathrm{lo}}=\left(\frac{\mathrm{d}P}{\mathrm{d}L}\right)_{\mathrm{tp}}\Big/\left(\frac{\mathrm{d}P}{\mathrm{d}L}\right)_{\mathrm{lo}} \tag{10}$$

式中：$(\mathrm{d}P/\mathrm{d}L)_{\mathrm{tp}}$和$(\mathrm{d}P/\mathrm{d}L)_{\mathrm{lo}}$分别为两相压降梯度和全液相压降梯度。

压降的计算与两相区长度的计算方法类似，对式 (10) 进行积分，可以得到两相区压降 ΔP_{tp}：

$$\Delta P_{\mathrm{tp}}=\frac{Q_{\mathrm{cond}}}{\left(T_{\mathrm{sat}}-T_{\mathrm{sink}}\right)}\left(\frac{\mathrm{d}P}{\mathrm{d}L}\right)_{\mathrm{lo}}\int_0^1\frac{\phi^2_{\mathrm{lo}}}{(UA/L)^{\mathrm{tp}}_{\mathrm{cond\text{-}sink}}}\mathrm{d}x \tag{11}$$

2）过热区和过冷区。过热区和过冷区的计算过程十分相似，过热区终点和过冷区起点均为饱和态，2 区内的对流换热系数和压降计算均可按光滑圆管内的单相流计算。

过热区的计算须求出过热区长度、过热区终点工质温度和压力，可联立能量平衡方程和压降计算方程求解。过热区内的能量平衡方程为

$$\dot{m}c_{\mathrm{p,v}}\left(T_{\mathrm{cond,in}}-T_{\mathrm{sat}}\right)=(UA/L)^{\mathrm{v}}_{\mathrm{cond\text{-}sink}}L_{\mathrm{sup}}\Delta T_{\mathrm{sup}} \tag{12}$$

式中：$c_{\mathrm{p,v}}$ 为饱和温度对应的气体工质定压比热容；$T_{\mathrm{cond,in}}$ 和 L_{sup} 分别为冷凝器入口工质温度和过热区长度；$(UA/L)^{\mathrm{v}}_{\mathrm{cond-sink}}$为过热区内冷凝器和热沉在单位长度上的热导；$\Delta T_{\mathrm{sup}}$ 为平均温差，取冷凝器入口温度与热沉温度之差。

过冷区长度 L_{sub} 可由冷凝器管路长度减去过热区和两相区长度得到，区域内的计算需求出冷凝器出口的工质温度和压力。过冷区内的能量平衡方程为

$$\dot{m}c_{p,l}\left(T_{sat}-T_{cond,out}\right)=(UA/L)^{l}_{cond\text{-}sink}L_{sub}\Delta T_{sub} \quad (13)$$

式中: $T_{cond,out}$ 为冷凝器出口工质温度; $(UA/L)^{l}_{cond-sink}$ 为过冷区内冷凝器和热沉在单位长度上的热导; ΔT_{sub} 为对数平均温差, 计算如下:

$$\Delta T_{sub}=\frac{(T_{sat}-T_{sink})-\left(T_{cond,out}-T_{sink}\right)}{\ln\left((T_{sat}-T_{sink})/\left(T_{cond,out}-T_{sink}\right)\right)} \quad (14)$$

2.2　模型求解

平板环路热管的工作区域可分为可变热导区和固定热导区。张红星[17]、柏立战等[18]阐述了不同工作区域内的运行特性。

在可变热导区内, 模型求解收敛的条件是储液器能量平衡方程 (式 (4))。储液器内存在温度梯度。因此, 储液器温度计算值向饱和温度和储液器入口温度的平均值 $T_{cc,av}$ 收敛, 两者差值满足误差要求时求解完毕。

计算蒸发器对储液器漏热时, 储液器内气体工质的体积占比 α_{cc} 可按式(6)计算。

在固定热导区内, 求解收敛的条件是工质质量守恒, 即工质质量 m^*与实际充装质量 m 之差应满足误差要求。

模型的计算流程如图 3 所示。首先, 判断平板环路热管所处的工作区域, 再根据不同区域的收敛条件求解, 迭代计算基于部件模型完成, 并以各部件进出口为结点, 保存工质温度和压力的数据。

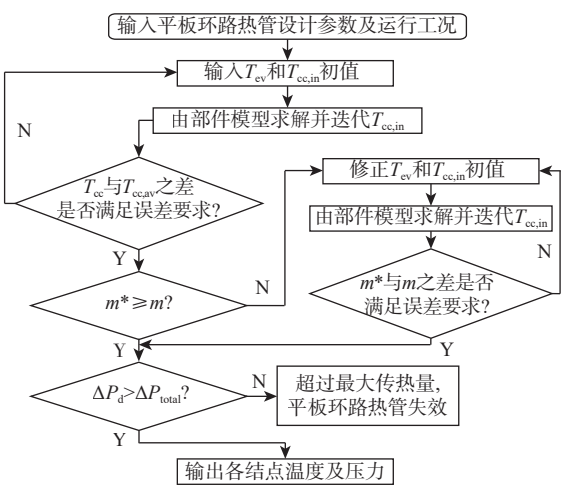


图 3　模型计算流程

Fig. 3　Flow chart for model calculations

根据不同工作区域内平板环路热管气液分布的特点, 储液器是否被液体工质充满可作为环路热管工作区域的判断依据。可先假定平板环路热管处于可变热导区, 储液器达到能量平衡, 计算此状态下的工质质量 m^*, 并与实际充装质量 m 对比, 即可判断环路热管所处的工作区域。若 $m^*>m$, 说明高估了储液器工质质量, 储液器未被液体充满, 平板环路热管处于可变热导区; 若 $m^*<m$, 则说明储液器已被充满, 且两相区长度大于实际值, 平板环路热管处于固定热导区。

判断该热负载是否超过平板环路热管的最大传热量, 若超过则判定环路热管失效, 未超过则输出各结点的温度和压力。

2.2.1　工质质量计算

对于稳态运行中的平板环路热管, 蒸气槽道、蒸气管路、冷凝器过热区内均为饱和蒸气; 冷凝器两相区为气液两相态; 主芯、副芯、冷凝器过冷区和液体管路内充满液体。此外, 假设储液器内充满液体工质, 以使工质质量计算满足模型求解的需要。

因此, 工质质量 m^*可按式 (15) 计算, 式中, 工质物性参数可按各部件对应位置的饱和态或平均温度取值。使用 Xu-Fang 模型[19]计算两相区空泡率 α_v, 进而得到两相区工质质量。

$$\begin{aligned}m^*=&\left(V_{vg}+V_{vl}+V_{sup}+\alpha_v V_{tp}\right)\rho_{v,sat}+\\&\left[V_{pw}+V_{sw}+(1-\alpha_v)V_{tp}\right]\rho_{l,sat}+\\&V_{sub}\rho_{l,sub}+V_{ll}\rho_{l,ll}+V_{cc}\rho_{l,cc}\end{aligned} \quad (15)$$

式中: V_{vg}、V_{vl}、V_{sup} 和 V_{tp} 分别为蒸气槽、气体管路、过热段和两相区的容积; α_v 为两相区含气率; $\rho_{v,sat}$ 为饱和温度下气体工质的密度; V_{pw} 和 V_{sw} 分别为主芯和副芯内孔隙的容积; $\rho_{l,sat}$ 为饱和温度下液体工质的密度; V_{sub}、V_{ll} 和 V_{cc} 分别为过冷区、液体管路和储液器的容积; $\rho_{l,sub}$ 为特征温度取饱和温度和热沉温度平均值时过冷区液体工质的密度; $\rho_{l,ll}$ 和 $\rho_{l,cc}$ 分别为液体管路和储液器内液体工质的密度。

2.2.2　最大传热量计算

平板环路热管的最大传热量根据毛细驱动力和流动压降的平衡计算, 如式 (16) 所示。当环路总压降 ΔP_{total} 与最大毛细力 ΔP_c 相等时, 平板环路热管的传热量达到极限。

$$\Delta P_c=\Delta P_{ev}+\Delta P_{vl}+\Delta P_{cond}+\Delta P_{ll} \quad (16)$$

式中: ΔP_{ev}、ΔP_{vl}、ΔP_{cond} 和 ΔP_{ll} 分别为蒸发器、气体管路、冷凝器和液体管路内的流动压降。

毛细力由 3 个因素共同决定, 计算方式如式 (17) 所示, 其中, 毛细芯孔径可由实验测得, 表面张力可取工作温度对应的值, 而接触角难以通过实验方式直接得到。根据闫凯芬[9]的推荐, 丙烯和镍粉毛细芯的接触角可取 75°, 而根据刘成等[20]的结论, 接触角可取 60°。

$$\Delta P_c=\frac{4\sigma\cos\theta}{d_{pw}} \quad (17)$$

式中: σ 为液体工质在饱和温度下的表面张力; θ 为

液体工质与主芯的接触角；d_{pw} 为主芯最大孔径。

此外，低温工况中，最大毛细力与流动压降的关系不是判断平板环路热管传热能力是否到达极限的唯一根据。丙烯和氨等工质的饱和蒸气压随温度降低，在低温下可能低于最大毛细力，并替代后者成为限制传热能力的因素。

因此，以最大驱动力 ΔP_d 与环路总压降 ΔP_{total} 的平衡关系确定平板环路热管的最大传热量，ΔP_d 取值为最大毛细力和饱和蒸气压中较大者。

3 模型实验验证

3.1 实验装置

实验所用丙烯工质平板环路热管的主要参数如表 1 所示。

表 1 丙烯工质平板环路热管主要参数

Table 1 Main parameters of the propylene flat loop heat pipe

毛细芯最大孔径/μm	毛细芯渗透率/m^2	毛细芯厚度/mm	毛细芯面积/mm^2	蒸气槽数量	蒸气槽直径/mm	蒸气槽长度/mm	副芯长度/mm	副芯截面积/mm^2	副芯渗透率/m^2	储液器容积/mL	蒸气管路尺寸/mm	液体管路尺寸/mm	冷凝器管路尺寸/mm	工质充装质量/g
1.35	8.3×10^{-15}	4	2 300	32	1	50	70	150	3.4×10^{-11}	15	ϕ3×1 750	ϕ2×2 000	ϕ2×3 200	21.7

平板蒸发器与储液器为横向供液型，蒸气槽道设置在毛细芯与壳体受热面之间，集气腔设置在蒸气槽道与蒸气管路接口之间。冷凝器管路钎焊在预留凹槽的铜板上，铜板与水冷板间紧密贴合并涂抹导热脂。实验使用陶瓷加热片作为模拟热源，由程控电源供电施加热负载；水冷板作为热沉，由水冷机控温。实验使用 T 型热电偶测温，由安捷伦 34970A 收集温度数据。实验平板环路热管及实验台设备如图 4 所示。平板环路热管示意图及测温点位置如图 5 所示。其中，T119 位于储液器入口处，测量值视为储液器入口温度 $T_{cc,in}$；T115 位于储液器上表面正中位置，测量值视为储液器温度 T_{cc}；T113 位于蒸气管路入口处，测量值视为工作温度 T_{ev}。

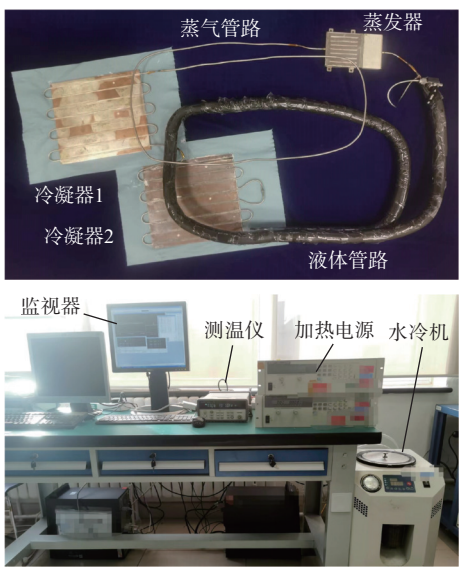


图 4 实验平板环路热管及实验设备

Fig. 4 The experimental flat loop heat pipe and equipment

3.2 模型结果与实验结果对比

平板环路热管在实验中的工作温度曲线如图 6 所示。实验中，热沉温度为 18.0 ℃，环境温度为 23.5 ℃。平板环路热管以 20.5 W 热负载启动，之后依次增大热负载至 26.5, 33.4, 41.1, 49.6, 59.4, 66.6, 70.8, 73.3, 76.2, 78.5 W。启动后，蒸发器与储液器几乎同时升温，且温度相近，随后冷凝器两相区增长（T108、T110、T101 依次升温），平板环路热管运行在可变热导区。热负载超过 34.4 W 后，平板环路热管进入固定热导区，蒸发器与储液器温差逐渐变大。热负载为 78.5 W 时，工作温度显著升高，最终出现温度飙升现象，表明该热负载已超过最大传热量。

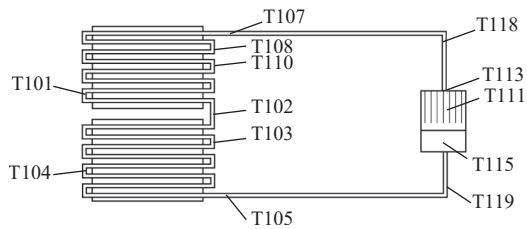


图 5 平板环路热管结构及测温点布局示意图

Fig. 5 Schematic drawing of the flat loop heat pipe and temperature measuring points layout

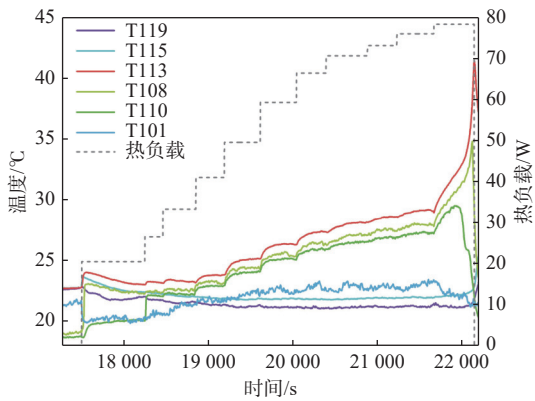


图 6 不同热负载下平板环路热管工作温度曲线

Fig. 6 The operating temperature curves of the flat loop heat pipe in different heat loads

实验发现，平板环路热管处于可变热导区时，储液器温度低于饱和温度，十分接近饱和温度和储

液器入口温度的平均值，说明储液器内工质处在两相态时存在温度梯度。

将不同热负载下 T_{sat}、T_{cc}、$T_{cc,in}$ 的实验数据代入式(4)可以推导 $h_{ev\text{-}cc}$ 的值。用式(5)对其进行拟合，发现 C_1 取 1 000 W(m^2·K^{-1})，C_2 取 121 W(m^2·K^{-1})时，拟合曲线与实验推导所得曲线相近，如图 7 所示。

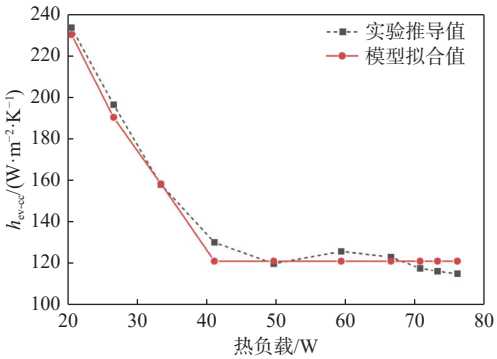


图 7　$h_{ev\text{-}cc}$ 的实验推导值和模型拟合值

Fig. 7　Experimental derivation values and simulation fitting values of $h_{ev\text{-}cc}$

蒸发器和储液器稳态工作温度的模型预测值与实验值的对比如图 8 所示。模型预测结果与实验结果最大差值不超过 2 ℃，吻合度良好，变化趋势一致。热负载较大时，模型对储液器温度的预测值偏低，这是因为低估了回流液的温度在固定热导区内，蒸发器对储液器的漏热占热负载的比例很小，因此，储液器温度预测值的少量偏差对工作温度的预测影响并不大。

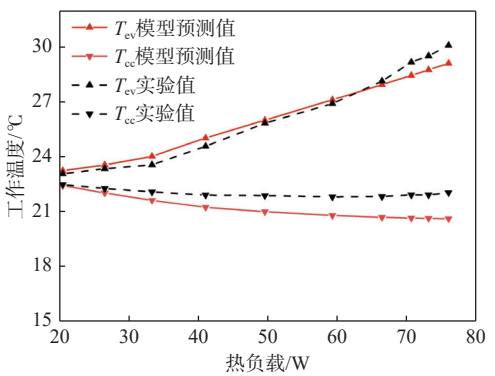


图 8　蒸发器及储液器稳态工作温度实验值和模型预测值的对比

Fig. 8　Comparison of the temperature of evaporator and compensation chamber between experiment and simulation

4　运行特性分析

4.1　最大传热量

根据 2.2.2 节的分析，对最大传热量的准确预测要求掌握丙烯与毛细芯的接触角。基于毛细力和总压降的平衡关系，可由最大传热量下的总压降反推接触角。实验中，平板环路热管飙升前可在 29.7 ℃、76.2 W 热负载下维持稳定，而施加 78.5 W 热负载后，工作温度飙升，因此，其在该工作温度下的传热极限为 76.2～78.5 W。计算不同接触角下平板环路热管的最大传热量，如图 9 所示。可见，反推得到的丙烯工质与镍粉毛细芯的接触角约为 52°～54°，小于文献 [10,21] 的推荐值。

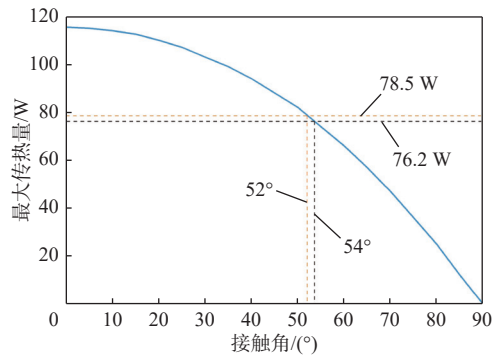


图 9　不同接触角拟合的最大传热量

Fig. 9　The maximum heat transfer capabilities fitted with different contact angles

丙烯和氨的主要热物性参数如表 2 所示。与航天器常用的氨工质平板环路热管相比，丙烯工质平板环路热管的传热能力较差，丙烯和氨工质平板环路热管的品质因数曲线(Dunbar 参数曲线)和基于本文模型计算的最大传热量曲线分别如图 10 和图 11 所示。其中，丙烯工质的接触角取 53°，氨工质取 0°。

Dunbar 参数曲线与本文计算的最大传热量曲线具有一致的变化趋势，表明在−50～60 ℃ 温区内，丙烯工质平板环路热管的最大传热量远小于氨工质平板环路热管；丙烯工质平板环路热管的传热能力随温度变化较小，氨工质平板环路热管的传热能力随温度变化更加显著；在低温区，氨工质平板环路热管传热能力降低，两者差距缩小。

此外，本文计算结果显示，氨工质平板环路热

表 2　丙烯和氨的主要热物性参数

Table 2　Main thermophysical parameters of propylene and ammonia

材料	冰点/℃	沸点/℃	临界点	气液相变潜热(0 ℃)/(kJ·kg^{-1})	表面张力(0 ℃)/(mN·m^{-1})	饱和液密度(0 ℃)/(kg·m^{-3})
丙烯	−185	−47.7	91.9 ℃、4.62 MPa	377.59	10.178	546.13
氨	−77.7	−33.5	132.5 ℃、11.40 MPa	1 262.25	26.295	638.57

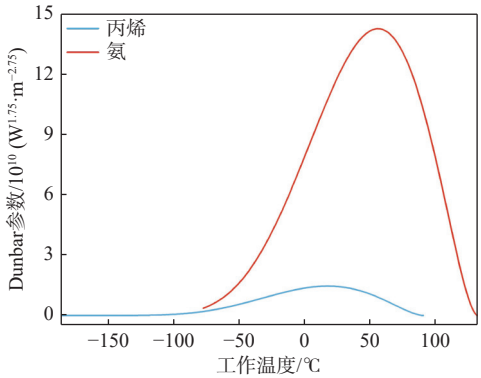


图 10 丙烯和氨工质平板环路热管的 Dunbar 参数曲线

Fig. 10 Dunbar parameter curves of propylene and ammonia flat loop heat pipe

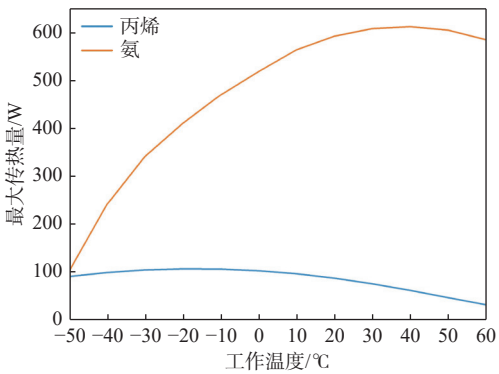


图 11 丙烯和氨工质平板环路热管的最大传热量

Fig. 11 The maximum heat transfer capabilities of propylene and ammonia flat loop heat pipe

管最大传热量曲线在−34 ℃附近的斜率发生了变化。这是因为驱动力与总压降的平衡关系发生了变化。对于氨工质平板环路热管，常温区最大驱动力为毛细力；而在−34 ℃以下，氨工质平板环路热管的饱和蒸气压小于最大毛细力，成为最大驱动力的限制因素，总压降应小于或等于饱和蒸气压。对于丙烯工质平板环路热管，在−50～60 ℃温区内，毛细力始终小于饱和蒸气压，因此，最大传热量受毛细力限制。

4.2 流阻分布

基于模型对平板环路热管的压降计算，给出不同热负载下各部件的压降，如图 12 所示。可见，蒸发器压降是总压降的主要组成部分，占比约为81%，管路内的压降占比相对较小，蒸气管路、冷凝器管路、液体管路分别约为 4%、13%、2%。

蒸发器压降占比大的原因有 2 个：①毛细芯渗透率低，根据式(2)可知，渗透率与压降成反比，毛细芯渗透率低会直接增大蒸发器压降；②丙烯工质产生的毛细力过低，使工质质量流量难以提高，而小质量流量下液体在多孔介质内的压降显著大于蒸气在管路内的压降。

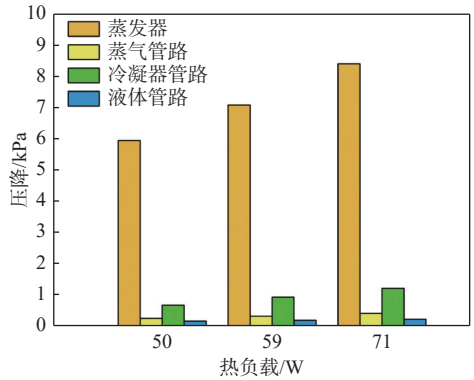


图 12 丙烯平板环路热管各部件压降

Fig. 12 Pressure drops of propylene flat loop heat pipe components

因此，未来丙烯平板环路热管传热性能提升的合理途径是优化蒸发器性能，包括减小毛细芯材料与丙烯的接触角、缩小微孔孔径、提高渗透率、增大面积与厚度比等。

4.3 热沉温度对工作温度的影响

不同热沉温度下工作温度随热负载的变化如图 13 所示，模型的参数设置与实验状态相同。可见，热沉温度降低会使稳态工作温度降低，且平板环路热管进入固定热导区的临界热负载增大。

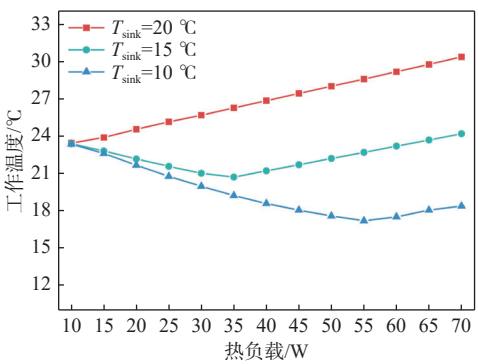


图 13 不同热沉温度下工作温度随热负载的变化

Fig. 13 Variation of operating temperature with thermal loads at different heat sink temperatures

由于丙烯工质平板环路热管的最大传热量较小，在热沉温度低时，临界热负载将接近最大传热量。为了降低热阻并提高运行稳定性，对平板环路热管的设计一般要求固定热导区的热负载范围足够大。因此，储液器容积和工质充装质量的匹配是实现设计目标的关键。

5 丙烯工质平板环路热管的宽温区设计

丙烯工质平板环路热管的主要应用需求是深空探测航天器的热控制。在深空探测任务中，航天器将面对比地球轨道更复杂且极端的热环境，同

时，航天器可为热控系统提供的能源、质量等预算极少。因此，平板环路热管需具备良好的低温适应性及宽温区运行能力。

宽温区设计的目的是保证平板环路热管的低温启动性能和多种工况下的运行性能。这要求储液器容积和工质充装质量的设计应满足以下要求：①低温存储时储液器内有充足的液体；②对于不同温度的运行工况，冷凝器两相区长度均处于合理范围内。

用于深空探测任务的丙烯工质平板环路热管预期工作温度范围为−30～30 ℃，最低存储温度为−150 ℃。存储温度与运行温度间的差值越大，液体工质的密度变化越大，所需储液器的容积就越大。冷凝器有效换热面积（两相区长度）对平板环路热管的传热性能有显著影响，其最大长度与工质密度变化相关。在不同的运行工况中，冷凝器有效换热面积（两相区长度）受液体工质体积影响：在高温工况中，液体工质密度小体积膨胀，将两相区压缩；而在低温工况中，液体工质密度增大、体积缩小使两相区增长。显然，储液器容积越大，其对两相区长度的影响也越大。

因此，低温存储与宽温区运行在储液器容积设计上存在矛盾。提出储液器容积和工质充装质量的改进计算方法及带有副储液器的设计方案以解决该矛盾。

5.1　储液器容积和工质充装质量的计算方法

柏立战等[18]提出了储液器体积和工质充装质量的计算方法，对工作在常温区的平板环路热管合理有效的。然而，丙烯工质平板环路热管具有低温存储和低温运行的要求，需对此算法作出 2 点改进：①根据航天器舱内和舱外温度分别计算低温存储时蒸发器（含储液器）和外部管路的质量；②在低温运行工况中，工质应足够多以避免两相流冲出冷凝器引发温度振荡等问题[21]。改进后的计算方法如下：

$$m \geqslant \rho_{\mathrm{l,ev,st}}\left(V_{\mathrm{pw}}+V_{\mathrm{sw}}+V_{\mathrm{vg}}+\beta V_{\mathrm{cc}}\right)+\rho_{\mathrm{l,st}}\left(V_{\mathrm{vl}}+V_{\mathrm{cond}}+V_{\mathrm{ll}}\right)+\rho_{\mathrm{v,cc,st}}\left(1-\beta\right)V_{\mathrm{cc}} \tag{18}$$

$$m \leqslant \rho_{\mathrm{l,ev,hot}}\left(V_{\mathrm{pw}}+V_{\mathrm{sw}}\right)+\rho_{\mathrm{v,hot}}\left(V_{\mathrm{vg}}+V_{\mathrm{vl}}+\varphi_1\alpha_{\mathrm{v}}V_{\mathrm{cond}}\right)+\rho_{\mathrm{l,hot}}\left[\left(1-\varphi_1\alpha_{\mathrm{v}}\right)V_{\mathrm{cond}}+V_{\mathrm{ll}}+V_{\mathrm{cc}}\right] \tag{19}$$

$$m \geqslant \rho_{\mathrm{l,ev,cold}}\left(V_{\mathrm{pw}}+V_{\mathrm{sw}}\right)+\rho_{\mathrm{v,cold}}\left(V_{\mathrm{vg}}+V_{\mathrm{vl}}+\varphi_2\alpha_{\mathrm{v}}V_{\mathrm{cond}}\right)+\rho_{\mathrm{l,cold}}\left[\left(1-\varphi_2\alpha_{\mathrm{v}}\right)V_{\mathrm{cond}}+V_{\mathrm{ll}}+V_{\mathrm{cc}}\right] \tag{20}$$

式中：$\rho_{\mathrm{l,ev,st}}$、$\rho_{\mathrm{l,st}}$、$\rho_{\mathrm{v,cc,st}}$ 分别为存储状态下蒸发器内液体、外部管路内液体、储液器内气体的密度；β 为储液器内的含液率；$\rho_{\mathrm{l,ev,hot}}$、$\rho_{\mathrm{v,hot}}$、$\rho_{\mathrm{l,hot}}$ 分别为高温运行工况下蒸发器内液体、环路内饱和蒸气、外部管路和储液器内液体的密度；$\rho_{\mathrm{l,ev,cold}}$、$\rho_{\mathrm{v,cold}}$、$\rho_{\mathrm{l,cold}}$ 分别为低温运行工况下蒸发器内液体、环路内饱和蒸气、外部管路和储液器内液体的密度；V_{cond} 为冷凝器管路容积；$\varphi_1=30\%$ 和 $\varphi_2=90\%$ 分别为高温和低温运行工况下两相区长度占冷凝器总长的比值。

基于实验环路热管的设计参数，设定存储状态下外回路温度为−150 ℃，蒸发器和储液器温度为−40 ℃，储液器含液率为 30%；高温运行状态下，蒸发器及气相管路部分（含冷凝器两相区）为 30 ℃，液体管路部分及储液器为 20 ℃，两相区含气率为 80%。式（18）和式（19）联立可求得最小储液器容积为 64.4 mL，最小工质充装质量为 47.4 g。实际上，若不对低温存储方程做改进，用柏立战等[18]方法计算的结果正是该最小值。

将式（18）～式（20）联立求解，可以考察储液器容积能否满足低温运行的要求，计算结果如图 14 所示。可见，解的范围收缩为狭长三角形，这是因为式（18）和式（19）所假设的气液分布相似，使 2 条曲线近乎平行。新的解确定了储液器容积的上限，同时，给出了储液器容积对应的工质充装质量。可见，本文的平板环路热管可以适应的工作温区较窄，在低温工况为 20 ℃ 时有解，而低温工况设为 10 ℃ 时无解。

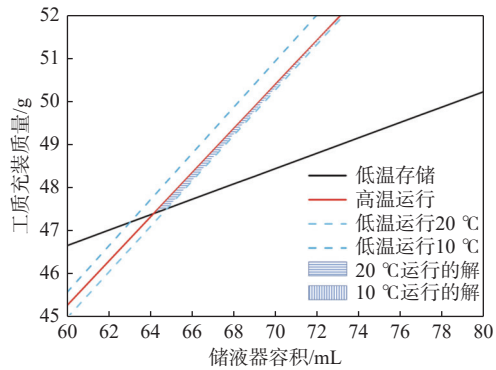


图 14　不同低温工况下解的范围

Fig. 14　Region of solutions under different low temperature conditions

在高、低温工况间，冷凝器两相区最大长度的变化是由工质密度变化引起的。理论上，扩大冷凝器管路截面积或减小储液器容积可削弱工质密度变化对两相区长度的影响。按我国航天器环路热管常用的最大管径（内径 4 mm）计算，储液器容积为 99.8 mL、工质充装质量为 74.9 g，可将低温运行温度拓宽至−3.8 ℃。而若要将工作温度拓宽至−30 ℃，理论上至少需要将冷凝器管径扩大至 8.8 mm，

此时，储液器体积为281.5 mL，工质充装质量为216.4 g，环路热管的体积和质量将随之剧增。因此，仅扩大冷凝器管路内径的效果并不理想，需在减小储液器容积的方向做额外改进。

5.2 带副储液器的平板环路热管设计方案

为了解决低温存储和宽温区运行对储液器容积设计的矛盾，并控制储液器的体积和环路热管的整体质量，提出一种设计方案，如图15所示。在环路外设置一个置于舱外的副储液器，通过支路接入蒸气管路或液体管路，并在副储液器出口设置控制阀，副储液器及控制阀仅有少量能耗，航天器负担很小。平板环路热管进入存储状态时，加热副储液器，使工质蒸发并过热，补偿工质受压力推动进入环路热管后关闭控制阀、停止加热。补偿工质进入液体管路和主储液器，可在管路内充满液体时保证主储液器工质充足；在运行状态下，控制阀打开，控制副储液器温度低于环路热管内工质的饱和温度，工质将受压力推动回到副储液器内并将其充满。当主储液器出现过冷度，且冷凝器两相区长度足够大后，关闭控制阀副储液器内的工质将不再参与环路中的工质循环。除对副储液器控温时外，其始终处于舱外平衡温度(假设为−150 ℃)。

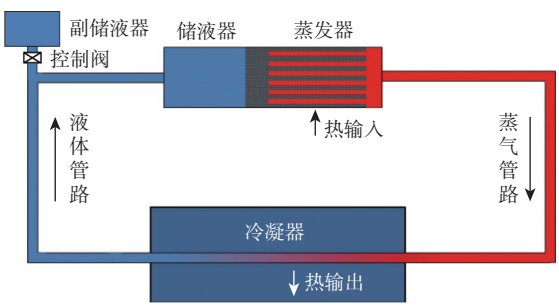


图15 带副储液器平板环路热管的结构

Fig. 15 The construction of the flat loop heat pipe with a secondary compensation chamber

该方案使主储液器容积和工质充装质量的计算方法产生了变化，式(18)改为

$$m \geqslant \rho_{\mathrm{l,ev,st}}\left(V_{\mathrm{pw}}+V_{\mathrm{sw}}+V_{\mathrm{vg}}+\beta V_{\mathrm{cc}}\right)+\rho_{\mathrm{l,st}}\left(V_{\mathrm{vl}}+V_{\mathrm{cond}}+V_{\mathrm{ll}}\right)+\rho_{\mathrm{v,ev,st}}\left(1-\beta\right)V_{\mathrm{cc}}-\left(\rho_{\mathrm{l,st}}-\rho_{\mathrm{v,ev,st}}\right)V_{\mathrm{scc}} \tag{21}$$

式中：V_{scc}为副储液器容积；$\rho_{\mathrm{v,ev,st}}$为蒸发器中饱和态的蒸气密度。

联立式(19)～式(21)，并将不等号改为等号，可解得主储液器容积的最大值和副储液器容积最小值，解得的工质质量为运行工况中参与循环的质量。总工质充装质量可由式(22)计算得到：

$$m=\rho_{\mathrm{l,ev,st}}\left(V_{\mathrm{pw}}+V_{\mathrm{sw}}+V_{\mathrm{vg}}+\beta V_{\mathrm{cc}}\right)+\rho_{\mathrm{l,st}}\left(V_{\mathrm{vl}}+V_{\mathrm{cond}}+V_{\mathrm{ll}}\right)+\rho_{\mathrm{v,ev,st}}\left[\left(1-\beta\right)V_{\mathrm{cc}}+V_{\mathrm{scc}}\right] \tag{22}$$

对于本文的实验件，采用4 mm内径冷凝器管路时，按此方案解得的主储液器容积最大值为49.7 mL，副储液器容积最小值为23.2 mL，总工质充装质量为66.1 g。与常规设计相比，冷凝器管路内径缩小了55%，主储液器容积减小了82%，总工质充装质量减小了70%。增大副储液器的容积还可进一步减小主储液器的容积，并减弱其对两相区长度的影响。维持不等式组的不等号不变，则可得到与图14类似的解，即不同副储液器容积下，对应的主储液器的容积设计范围和工质充装质量范围。

可见，带副储液器的平板环路热管设计方案可以解决低温存储和宽温区运行需求带来的矛盾。采用带副储液器设计的环路热管不仅可拓宽工作温区，同时，还具有质量轻、体积小的优点。然而，补偿工质的“进、出”过程对环路热管稳态及动态运行特性的影响有待研究。此外，根据不同蒸发器的结构，主储液器容积都应有最小值限制，以确保在任何状态下都能稳定地对毛细芯供液，此约束条件有待确定。

6 结 论

本文建立一种理论模型，实现了对丙烯工质平板环路热管稳态工作温度的准确预测，并通过实验验证了模型的正确性。具体结论如下：

1) 所建模型可在常温温区准确预测丙烯工质平板环路热管的蒸发器和储液器稳态工作温度，变化趋势吻合，最大误差小于2 ℃，对丙烯工质平板环路热管的设计和参数优化有一定参考意义。

2) 经模型分析和实验修正，丙烯与镍粉毛细芯的接触角接近52°～54°，是影响毛细力的重要因素，后续应开展研究掌握丙烯与镍粉及其他常用毛细芯材料的浸润特性。

3) 丙烯工质的表面张力小，且蒸发器内压降在总压降中占比约为81%。因此，提高蒸发器驱动力、降低压降是优化丙烯平板环路热管传热性能的主要途径。

4) 由于毛细力的差异，丙烯工质平板环路热管的最大传热量在全温区内低于氨工质平板环路热管。丙烯工质平板环路热管在低温区传热能力较好，而氨工质平板环路热管在常温区传热能力最佳。

5) 改进了储液器容积和工质充装量的计算方法，明确了不同工况下储液器容积的设计范围。带副储液器的平板环路热管设计方案可有效拓宽丙烯工质平板环路热管的工作温区，同时具备质量轻、体积小的优点。

参考文献（References）

[1] GROB E W. System accommodation of propylene loop heat pipes for the geoscience laser altimeter system (GLAS) instrument[J]. SAE Technical Paper Series, 2001, 1: 2263.

[2] GONCHAROV K A, NIKITKIN M N, GOLOVIN O A, et al. Loop heat pipes in thermal control systems for "OBZOR" spacecraft[J]. SAE Technical Paper Series, 1995, 1: 951555.

[3] CHOI M. Thermal design to meet stringent temperature gradient/stability requirements of SWIFT BAT detectors[C]//Proceedings of the 35th Intersociety Energy Conversion Engineering Conference and Exhibit. Reston: AIAA, 2000: 2905.

[4] RODRIGUEZ J I. Thermal design of the tropospheric emission spectrometer instrument[J]. SAE Technical Paper Series, 2000, 1: 2274.

[5] KOZMINE D, GONCHAROV K, NIKITKIN M, et al. Loop heat pipes for space mission Mars 96[J]. SAE Technical Paper Series, 1996, 1: 961602.

[6] MISHKINIS D, GREGORI C, ROMERA F, et al. Development of propylene LHP for European Mars Rover applications[J]. Heat Pipe Science and Technology, 2010(1): 19-46.

[7] WALKER K L, HARTENSTINE J, TARAU C, et al. Ammonia and propylene loop heat pipes with thermal control valves-thermal/vacuum and freeze/thaw testing[C]//Proceedings of the 43rd International Conference on Environmental Systems. Reston: AIAA, 2013: 3434.

[8] PAUKEN M, RODRIGUEZ J I. Performance characterization and model verification of a loop heat pipe[J]. SAE Technical Paper Series, 2000, 1: 2317.

[9] 闫凯芬. 丙烯环路热管 150-290 K 温区冷凝流型及换热特性研究[D]. 北京: 中国科学院大学, 2021: 34-40.
YAN K F. Condensation flow pattern and heat transfer characteristics of propylene loop heat pipes in 150-290 K[D]. Beijing: University of Chinese Academy of Sciences, 2021: 34-40(in Chinese).

[10] MAYDANIK Y, CHERNYSHEVA M, VERSHININ S. High-capacity loop heat pipe with flat evaporator for efficient cooling systems[J]. Journal of Thermophysics and Heat Transfer, 2020, 34(3): 465-475.

[11] YANG Z H, ZHANG Y F, BAI L Z, et al. Experimental study on the thermal performance of an ammonia loop heat pipe using a rectangular evaporator with longitudinal replenishment[J]. Applied Thermal Engineering, 2022, 207: 118199.

[12] ZHANG H X, LI G G, CHEN L, et al. Development of flat-plate loop heat pipes for spacecraft thermal control[J]. Microgravity Science and Technology, 2019, 31(4): 435-443.

[13] 丁汀, 张红星, 苗建印, 等. 环路热管启动性能实验与分析[J]. 航天器工程, 2006, 15(1): 25-30.
DING T, ZHANG H X, MIAO J Y, et al. Experimental study and analysis of the startup performance of LHP[J]. Spacecraft Engineering, 2006, 15(1): 25-30(in Chinese).

[14] ZHOU L, QU Z G, CHEN G, et al. One-dimensional numerical study for loop heat pipe with two-phase heat leak model[J]. International Journal of Thermal Sciences, 2019, 137: 467-481.

[15] CAVALLINI A, DEL COL D, DORETTI L, et al. Condensation in horizontal smooth tubes: a new heat transfer model for heat exchanger design[J]. Heat Transfer Engineering, 2006, 27(8): 31-38.

[16] MÜLLER-STEINHAGEN H, HECK K. A simple friction pressure drop correlation for two-phase flow in pipes[J]. Chemical Engineering and Processing: Process Intensification, 1986, 20(6): 297-308.

[17] 张红星. 环路热管 2 相传热技术的理论和实验研究[D]. 北京: 北京航空航天大学, 2006: 55-65.
ZHANG H X. Theoretical and experimental investigation of loop heat pipes[D]. Beijing: Beihang University, 2006: 55-65(in Chinese).

[18] 柏立战, 林贵平, 张红星, 等. 环路热管稳态建模及运行特性分析[J]. 北京航空航天大学学报, 2006, 32(8): 894-898.
BAI L Z, LIN G P, ZHANG H X, et al. Steady state modeling of loop heat pipes and operating characteristics analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, 2006, 32(8): 894-898(in Chinese).

[19] XU Y, FANG X D. Correlations of void fraction for two-phase refrigerant flow in pipes[J]. Applied Thermal Engineering, 2014, 64(1-2): 242-251.

[20] 刘成, 谢荣建, 吴亦农, 等. 工质对 180～230 K 低温环路热管工作特性的影响[J]. 工程科学与技术, 2020, 52(6): 190-198.
LIU C, XIE R J, WU Y N, et al. Effects of working fluids on operating characteristics of the cryogenic loop heat pipe in 180～230 K[J]. Advanced Engineering Sciences, 2020, 52(6): 190-198(in Chinese).

[21] WANG N H, JOSEPH B, CHENG L. Design and experimental study of thermal control system for AMS cryocoolers[J]. Chinese Science Bulletin, 2013, 58(10): 1200-1204.

Steady state modeling and characteristic analysis of propylene flat loop heat pipes

JIA Zhichao[1], BI Hanli[1], PENG Zheng[2], LI Guoguang[1], WU Qi[1], ZHANG Hongxing[1,*], MIAO Jianyin[1]

(1. National Key Laboratory of Spacecraft Thermal Control, Beijing Institute of Spacecraft System Engineering, Beijing 100094, China;
2. School of Aeronautical Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: The flat loop heat pipe (LHP) using propylene as the working fluid has advantages such as low-temperature adaptability and light weight, making it the vital technology for solving the thermal control problems of deep space exploration missions. Urgent demands are raised to investigate the heat transfer performance and characteristics of propylene flat LHPs. This paper established a steady-state model, which can accurately predict the operating temperature of a propylene flat LHP. The maximum heat transfer capability and flow resistance characteristics of propylene LHP were analyzed. The calculation method was improved, which is used for computing the volume of the compensation chamber and the mass of working fluid. A design of an LHP with a secondary compensation chamber is put forward. The propylene LHP's operating temperature range is expanded by the design, which also reduces its weight and volume.

Keywords: propylene; loop heat pipes; flat evaporator; modeling analysis; space thermal control

Received: 2023-12-14; **Accepted**: 2024-03-08; **Published Online**: 2024-03-18 21:50

URL: link.cnki.net/urlid/11.2625.V.20240316.1501.004

Foundation item: National Natural Science Foundation of China (U21B2084)

* **Corresponding author**. E-mail: redlincoco@hotmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0826

生物航煤燃烧低压模化数值模拟

何彦潼[1]，邓甜[1]，王伟[2,*]

(1. 中国民航大学 中欧航空工程师学院，天津 300300；2. 中国民航大学 适航学院，天津 300300)

摘 要：全球能源需求增长和环境污染加剧背景下，生物航煤作为清洁、低碳、可持续燃料，引领未来燃料发展。航空发动机燃烧室的原型试验受到实际条件限制，低压模化试验可简化工况参数，有效解决实际问题并满足工程需求。为此，以生物航煤为燃料，采用3种模化准则，对燃烧室进行低压模化数值模拟研究，并与原设计状态对比。结果表明：等容积流率模化准则（Q准则）预测的流场分布与原设计状态最接近，且轴向和切向温度场分布优于等燃烧效率模化准则（K准则）和L准则。低压模化条件下，与原设计状态相比，燃烧效率降低2%～3%。特别是当压力指数过高时，L准则模化后的燃油流量过小，不适用于航空发动机的模化试验。出口截面温度分布系数的预测结果显示，K准则的预测值与原设计值相比，存在负偏差，Q准则预测结果的平均误差最小。综合来看，Q准则为最优模化准则，其计算结果可为航空发动机燃烧室的低压模化试验研究提供参考。

关 键 词：航空发动机；低压模化；数值模拟；Q准则；生物航煤
中图分类号：V231.21
文献标志码：A **文章编号**：1001-5965(2026)03-0885-10

随着全球能源需求的持续攀升，传统化石燃料引发的环境污染和资源枯竭等问题日益凸显。作为碳排放的主要来源之一，航空业正面临着巨大的环保压力。因此，具有清洁、低碳、可持续特点的生物航煤成为未来燃料发展的重要方向。航空业的迅猛发展对发动机的性能要求也日益苛刻。发动机正朝着高压比、高初温、大流量、大功率及更优热力性能的方向发展[1]。燃烧室必须具备更高的出口温度和增压比以实现航空发动机的性能目标。因此，深入研究生物航煤作为燃料对发动机燃烧室燃烧特性的影响，显得尤为重要。试验方法和数值模拟构成解决现代工程问题的两大主流方法。在燃烧室的研究和设计过程中，试验方法得到了广泛应用，主要包括原型试验方法和物理模型试验方法。前者是指在实际燃烧室上进行全压测试，以获得与实际情况最接近的数据；而后者则在简化(缩小工况)的参数条件下使用原型尺寸的燃烧室模型进行试验，通常被称为低压试验[2]。采用原型试验方法得到的结果虽然真实可靠，但在设计和改型阶段进行大量原型试验，耗费巨大的时间和能源成本。因此，为了有效解决实际问题，并满足工程上快速分析的需求，科研人员一直在努力探索和研究燃烧室的物理模型试验方法。对燃烧室进行低压模化数值模拟计算，可以在短时间内以更低的成本获得与低压试验同样准确的结果，从而为燃烧室的设计和优化提供更有力的支持。

燃烧室的工作过程涉及多个物理化学反应，包括气体掺混、燃烧、传热等多个过程，且之间存在强耦合关系[3]。仅依赖理论公式进行全尺寸燃烧室的数值模拟，资源消耗巨大且周期长。因此，在进行发动机燃烧室的模拟仿真时，考虑多种复杂因素和边界条件，运用相似理论能够在很大程度上简化

收稿日期：2023-12-20；**录用日期**：2024-04-02；**网络出版时间**：2024-04-10 11:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20240408.1837.004
基金项目：天津市教委科研计划 (2020KJ036)
* **通信作者**. E-mail：weiwangcauc@126.com

引用格式：何彦潼，邓甜，王伟. 生物航煤燃烧低压模化数值模拟 [J]. 北京航空航天大学学报，2026，52（3）：885-894.
HE Y T，DENG T，WANG W. Numerical simulation of low-pressure modeling of bio-jet coal combustion [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：885-894（in Chinese）.

复杂的工程问题,并节省计算资源。相似三定理包含相似正定理、相似逆定理及π定理[3],其为研究各过程中的相似性提供核心标准,确保在低压模化试验中能够真实地描述原型设备中的工作过程。其中,相似正定理阐明2种相似现象的相似判据必须相等,在进行低压模型试验时,需要保证模型与实际设备在某些重要参数上的一致性。例如,燃烧室的尺寸、形状、材料等都需要与原型设备保持一致,以保证试验结果的可靠性。相似逆定理表明,在物理方程经无量纲化后,其各项即为相似判据,故在进行低压模型试验时,需要根据物理方程进行无量纲化,确定相似判据,并确保模型与实际设备在这些相似判据上具有一致性。π定理是指2种物理现象中的基本物理量所组成的无量纲数群(相似准则)必须相等,即在进行低压模型试验时,需要确定适当的相似准则,并保证模型与实际设备在这些准则上一致。总之,相似理论的3个定理为低压模型试验提供了理论基础。通过合理应用这些定理,可以确保研究过程与实际设备的相似性,从而提高试验的准确性和可靠性。相似理论的应用不仅在燃烧室工作过程中起到重要作用,也广泛应用于其他领域的模型试验研究中。

自20世纪50年代开始,许多国内外学者已经对燃烧的相似性问题进行了深入的探讨和研究。Steward[4]基于Damkohler[5]理论,针对均匀混气的燃烧相似问题,提出9个无量纲相似准则参数,分析得出判断燃烧过程相似的2个关键相似准则数,即燃料在燃烧室中停留时间与反应时间的比值和雷诺数。王宏基[6]在研究空气喷气发动机燃烧室模化问题的过程中,发现等压模化规律能够有效确保带有蒸发器(或均匀混气)的燃烧室的燃烧相似现象。Childs[7]和Greenhough[8]等基于火焰传播理论、燃烧速率理论和化学反应动力学理论,提出不同的模化准则模型,以评估燃烧效率。陈丹之[9]通过理论分析,得到涡轮喷气发动机和燃气轮机燃烧室的模化准则,并强调模化准则并不唯一,需根据具体条件确定关键的模化参数。郭伯伟[10]认为任何燃烧装置的试验研究都应以模化理论为指导,使用正确的相似准则,以使试验结果具有指导意义。在此基础上,焦树建[11]采用等容积流率模化准则(*Q*准则)在原型燃烧室上进行低压模拟试验,研究表明,模拟得到的试验结果与原设计参数存在一定偏差,但总体上两者的相似性依然很高。汪凤山等[12]进行了微型燃气轮机燃油燃烧室的模化试验研究,结果表明,经过模化后,燃烧室的燃烧特性满足指标要求。赖哲渊等[13]对柴油机喷雾燃烧相似性展开研究,理论证明了在单值条件相同的情况下燃烧过程具有相似性。高玉闪等[14]采用可视化的燃烧室,对全流量补燃循环气气燃烧相似性开展缩尺试验研究,结果表明,随着喷嘴流量的增大,喷嘴出口的火焰结构均具有一定的相似性。杨强等[15]采用*Q*准则分析某燃气轮机低压模化试验对燃烧室试验结果的影响,发现模化结果的速度场变化较小,出口温度分布较好,火焰筒壁温偏低,火焰长度因压力和质量流量的降低而偏短。尽管前人已经针对燃烧室的低压模化研究提出许多模化准则,但以上模型均用于燃气轮机燃烧室,未在航空发动机燃烧室上进行验证。且大多数研究仅针对其中某种模化准则进行流场、温度场及燃烧效率等工作特性的比较,对不同模化准则下燃烧室的性能特点仍缺乏系统性研究。此前的研究多数采用传统航煤作为燃料,针对生物航煤在燃烧室中的数值模拟研究较为有限。

本文主要针对某型航空发动机燃烧室,以生物航煤为燃料,开展低压模化试验,结合精细的多步燃烧反应机理,使用*Q*准则、等燃烧效率模化准则(*K*准则)及*L*准则对发动机燃烧室进行数值模拟研究,通过与原设计状态条件下燃烧室的工作特性进行对比分析,探究适用于生物燃料航空发动机燃烧室物理模型试验方法的最优模化准则。

1 数值模拟方法

1.1 模化准则

1.1.1 *Q*准则

等容积流率模化准则通常简称为*Q*准则,是我国目前采用最多的模拟试验方法[11]。为保证得到较为准确的模拟结果,该准则要求:①在几何尺寸相同的燃烧室上进行试验;②使用同种燃料和氧化剂,同时保证两者初温相同;③模化后,进入燃烧室的空气容积流率与原设计状态相同,即$q_{va,M}=q_{va,R}$;④保证模化压力在0.3~0.5 MPa范围内[16]。确定模化空气压力后,模化空气质量流量$q_{ma,M}$的表达式为

$$q_{ma,M}=\frac{P_{in,M}}{P_{in,R}}q_{ma,R} \tag{1}$$

式中:P_{in}为燃烧室前空气压力,MPa;q_{ma}为空气质量流量,kg/s;下标R、M分别表示原型和模型。

王威[17]和田学清[18]等指出,模化压力在0.3~0.5 MPa范围内能够得到较精确的出口温度分布和流场分布的模拟结果。然而,燃烧效率及污染物排放的模化结果准确性在模化压力小于0.2 MPa时显著下降,这主要是因为低压环境下的燃烧过程和高

压环境相比, 有显著不同。另外, 如果使用过高的模化压力, 会削减低压模化试验的意义, 因为过高的压力可能会导致模化结果失去实际应用价值。因此, 对于模化压力的选择, 需要综合考虑多方面因素, 以达到最佳的模化效果。

1.1.2　*K* 准则

等燃烧效率模化准则即 *K* 准则, 是一种针对联管型燃烧室的模化准则, 通过对大量试验数据分析, 得出模化准则数 *K*[19]。*K* 值基于当量比相同的条件下得出, 能够衡量燃烧效率的模化结果与设计工况的一致性。相同尺寸的燃烧室, 只需通过改变进口空气质量流量、压力和温度, 使模型的准则数 *K* 值与原设计状态的 *K* 值相同, 即可满足两者之间燃烧效率的相似性。*K* 的表达式为

$$K=\frac{q_{\mathrm{ma}}}{P_{\mathrm{in}}^{1.15}T_{\mathrm{in}}l^{3}} \tag{2}$$

式中: l为燃烧室的定形尺寸, m; T_{in}为进口空气温度, K。

1.1.3　*L* 准则

Greenhough 等[8] 基于反应动力学理论, 并结合试验, 在各种燃烧室的设计压力、温度和空气质量流量范围内获得大量数据。当压力指数值为 1.75 时, 式(3)获得最佳关联性, 从而得到 *L* 准则。该准则认为燃烧效率为综合数群 *L* 的单值函数, *L* 的表达式为

$$L=\frac{q_{\mathrm{ma}}}{P_{\mathrm{in}}^{1.75}Fl\exp\left(\dfrac{T_{\mathrm{in}}}{b}\right)} \tag{3}$$

式中: F为燃烧室的参考面积, m^2; b为进口温度因子。

在原设计状态和低压模化状态下, 模型的流场结构存在一定差异, 这主要是由于在理论推导阶段, *K* 准则并没有充分考虑确保流动相似性的相应准则。这种差异可能会导致燃烧组织发生改变, 从而对燃烧室的性能产生一定程度的影响。为了更准确地评估这种影响, 在低压模化试验中, 需要对 *K* 准则和 *L* 准则进行更深入的研究和探讨。

1.1.4　压力指数

为了方便计算模化后的空气质量流量, 引入压力指数n, 得到公式如下[20]:

$$q_{\mathrm{ma,M}}=\left(\frac{P_{\mathrm{in,M}}}{P_{\mathrm{in,R}}}\right)^{n}q_{\mathrm{ma,R}} \tag{4}$$

Q 准则、*K* 准则和 *L* 准则对应的压力指数n分别为 1、1.15、1.75[16]。通过式(4)可以计算出模化压力与模化空气质量流量之间的关系。

1.2　几何模型及网格

针对文献 [21] 中单头部燃烧室建立几何模型, 如图 1 所示。该单头部燃烧室为全环燃烧室的 1/18, 包括双径向旋流器、文丘里管、火焰筒、主燃孔、掺混孔和气膜冷却孔等。

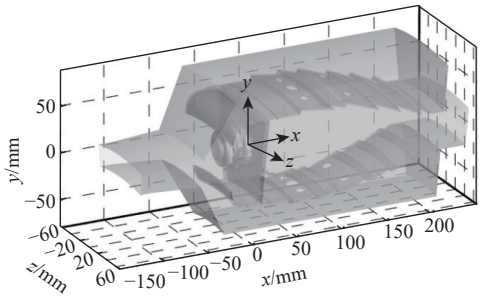


图 1　燃烧室几何结构示意图

Fig. 1　Schematic diagram of combustion chamber geometry

为避免指定火焰筒各进气孔的空气质量流量及人为计算给定旋流器和冷却气膜的进口条件而带来误差, 对某环形燃烧室的内部和外部通道进行模拟分析, 该燃烧室的计算范围从前置扩压器的喉道开始, 一直延伸到火焰筒出口。为了兼顾计算的精度和效率, 将数值模拟的区域进一步细分为 3 个子区域: ①旋流器和文丘里管区域。采用加密后的小尺寸结构化网格能够较好地捕捉空气与燃料的掺混及油滴的雾化过程。②火焰筒内部。在该区域, 空气和燃料产生剧烈的燃烧反应, 通过使用结构化网格优化计算速率和准确性。③进气道、燃烧室出口及火焰筒壁面到燃烧室内壁之间的区域。该区域对燃烧过程的影响相对较小, 主要涉及流体动力过程和冷、热气体掺混过程, 故选用较大尺寸的结构化网格来缩短计算时间。图 2 为整个计算域网格。

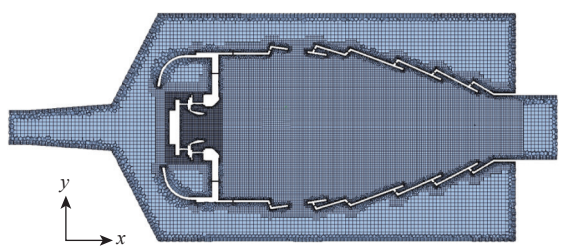


图 2　计算网格

Fig. 2　Computational grid

1.3　计算方法及边界条件

航空发动机燃烧室中的燃烧过程是一个复杂且动态的过程。在该过程中, 各项物理参数(如温度、压力、速度、流场等)都是随时间变化的。为了更好地研究这些物理量的变化规律, 采用商用软件 ANSYS Fluent 进行非稳态燃烧模拟, 时间步长设置为 $1\times e^{-5}$ s, 选取燃烧稳定后 t=50 ms 时的模拟结果进行分析。分别使用 *Q* 准则、*K* 准则和 *L* 准则计算进口参数。表 1 详细展示了不同模化准则下燃烧

室的进口参数。在进行燃烧仿真过程中，将流域内的燃油液滴视为离散相，忽略颗粒碰撞，只考虑燃油液滴与空气之间的相互作用，故选用离散相模型（discrete phase model, DPM）计算。湍流模型使用可实现 k-ε 模型，燃烧模型选择非预混概率密度函数（probability density function, PDF）燃烧模型。为准确捕捉燃烧动态行为，需要更加详细准确的燃料燃烧反应模拟。航空煤油是C1～C16大分子碳氢复杂烃类混合物，包含链烷烃、环烷烃及芳香烃等成百上千种组分。在研究航空煤油燃烧特性时，鉴于航空煤油成分的复杂性，无法针对某种航空煤油直接构建其化学反应动力学模型。故通常选取一种或几种具有代表性的成分作为航空煤油替代燃料的组分，进而构建其详细替代燃料的化学反应机理。杨飞[22]通过试验测得生物航煤与传统航煤有相似的化学分子结构，且生物燃料通过加氢加工从植物油和动物脂肪中提取，具有黏度高、密度高、热值高等优秀的物理特性，是理想的航空替代燃料。因此，选择73% 正十二烷、14.7% 的1,3,5-三甲基环己烷、12.3% 正丙基苯3种组分作为RP-3航空煤油的代表成分，选取44组分数，78反应数的RP-3骨架简化机理[23]模拟航空煤油复杂的燃烧动力学，同时，通过改变RP-3传统航煤的理化性质实现生物航煤的燃烧模拟。计算采用的生物航煤基本性质如表2所示[22]。

表1 原设计状态和不同模化准则下燃烧室的进口参数

Table 1 Inlet parameters of the combustion chamber in the design state and under different modeling criteria

状态	空气质量流量/($kg\cdot s^{-1}$)	空气温度/K	空气压力/MPa	燃油质量流量/10^{-3}($kg\cdot s^{-1}$)
原设计状态	1.3	800	2.5	17.95
Q准则	0.26	800	0.5	3.59
K准则	0.204	800	0.5	2.82
L准则	0.078	800	0.5	1.08

表2 生物航煤的基本性质[22]

Table 2 Basic properties of bio-jet coal[22]

密度/($kg\cdot m^{-3}$)	黏度/($mm^2\cdot s^{-1}$)	表面张力/($N\cdot m^{-1}$)	沸点/K	蒸发温度/K	气化潜热/($kJ\cdot kg^{-1}$)	热值/($MJ\cdot kg^{-1}$)
808	2.11	0.023 6	477.95	311	252.6	44.4

2 结果与分析

2.1 网格无关性验证及模型验证

使用Fluent Meshing软件对单头部燃烧室计算域划分多面体-正六面体混合类型网格，其中，大部分计算域为正六面体网格，而边界和结构复杂的区域为多面体网格。设立4组对照网格，网格数分别为160万、209万、330万、492万。对以上4组网格做常压冷态流场及热态燃烧场计算验证，具体的参数设置如表3和表4所示。验证计算的出口类型为outflow，选取燃烧室 x=30 mm 截面处中心轴线（z=0, x=30 mm）的轴向速度进行对比，如图3所示。由于燃烧室壁面摩擦力对流体的拖拽作用，形成速度梯度和边界层，对比其他网格算例可以发现：越粗糙的网格越难捕捉靠近燃烧室壁面边界层发展，对壁面处进行加密的330万网格算例和492万网格算例则可以捕捉上述现象。选取燃烧室 z=0 截面处中心轴线的温度进行对比，如图4所示，330万网格算例与492万网格算例所对应的温度分布差异很小。为保证计算准确度和计算效率，选用网格数量为330万的网格。网格质量生成良好、扭曲率大于0.85的网格数目控制在1.5% 以内。

为了验证模型的正确性，使用330万网格做常压冷态流场计算验证（详细参数设置见表3），并与文献[21]进行对比。由于常压冷态流场计算未涉及复杂的燃烧反应，采用稳态问题的隐式求解器模拟仿真。图5和图6分别为 x=10 mm 和 x=50 mm 截面的轴向速度对比情况。可以看出，使用可实现 k-ε 湍流模型可以较准确地对单头部燃烧室冷态流场进行模拟，330万网格算例轴向速度预测平均误

表3 冷态条件计算模型

Table 3 Calculation model for cold conditions

模型算法				射流参数				边界条件(进口)			
求解器	时间	湍流模型	近壁处理	进口类型	粒子类型	速度/($m\cdot s^{-1}$)	温度/K	总压/kPa	总温/K	湍流强度/%	水利直径/mm
隐式求解器	稳态	可实现k-ε模型	标准壁面函数	圆锥	液滴	30	300	110	300	5	34.56

表4 热态条件计算模型

Table 4 Calculation model for hot conditions

模型算法				射流参数				边界条件(进口)			
求解器	燃烧模型	湍流模型	近壁处理	进口类型	粒子类型	速度/($m\cdot s^{-1}$)	温度/K	总压/MPa	总温/K	湍流强度/%	水利直径/mm
非稳态问题的隐式求解器	非预混PDF燃烧模型	可实现k-ε模型	标准壁面函数	圆锥	液滴	30	400	2.5	800	5	34.56

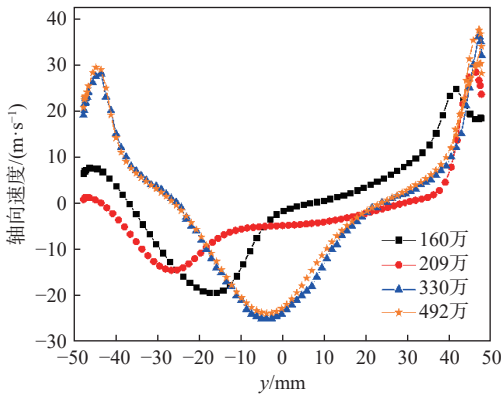


图 3　不同网格数下轴向速度对比

Fig. 3　Comparison of axial velocity at different mesh numbers

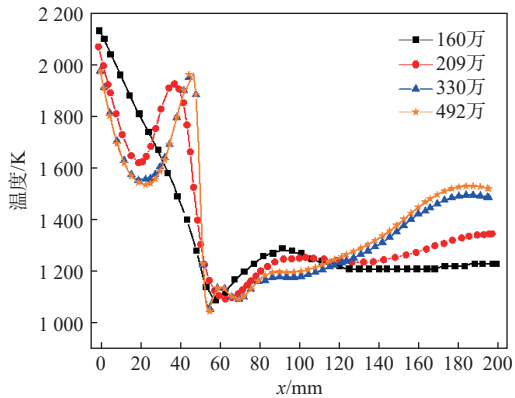


图 4　不同网格数下温度对比

Fig. 4　Comparison of temperature at different mesh numbers

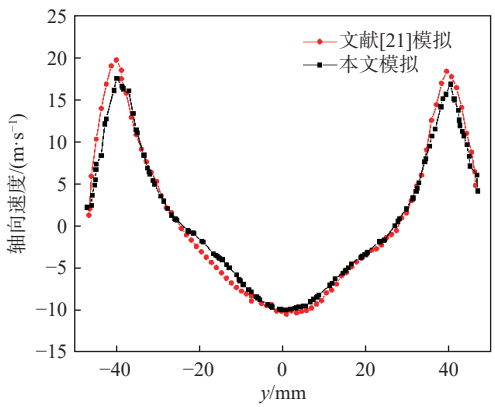


图 5　x=10 mm 截面处轴向速度模拟结果对比

Fig. 5　Comparison of the simulation results of axial velocity at x=10 mm cross-section

差为 4.69%。误差最大处为y=±40 mm 处, 为 7.32%, 验证了模型在冷态流场模拟方面的精确性及一致性。

燃烧室热态温度场的分布不仅与冷态的纯流场有关, 还与燃油的浓度场密切相关。为进一步验证模型的正确性, 对热态条件下的温度场模拟结果(详细参数设置见表 4)进行对比分析。由于主燃孔前侧回流区温度对燃烧过程模拟结果准确性影响较大, 取主燃孔前侧中心轴线上的点进行模拟温度结果对比, 如图 7 所示, 模拟结果平均误差仅为 3.97%。因此, 本文的数值模拟研究以该模型为基础。

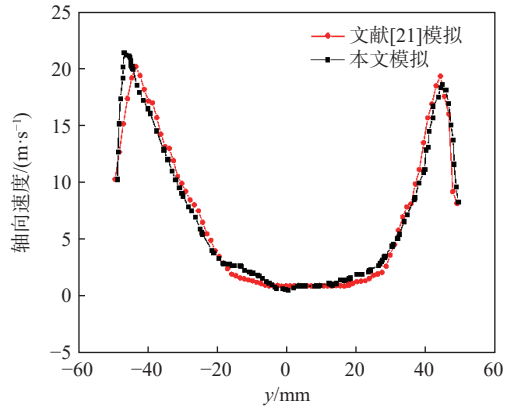


图 6　x=50 mm 截面处轴向速度模拟结果对比

Fig. 6　Comparison of the simulation results of axial velocity at x=50 mm cross-section

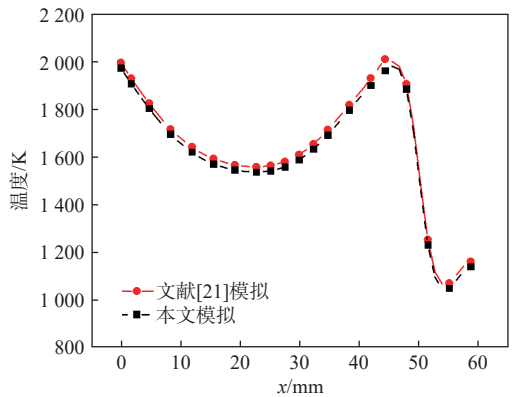


图 7　中心轴线上温度模拟结果对比

Fig. 7　Comparison of temperature simulation results on the central axis

2.2　流场分布

图 8 呈现了 3 种模化状态的流场结构, 并与原设计状态进行比较。可见, Q 准则模型的模化结果与原设计状态最为贴近, 主燃孔前后侧回流区的分布与面积基本一致, 说明该模型能够较好地反映燃烧室流动的真实情况。而 K 准则和 L 准则的模拟结果与原设计状态存在一定差异, 这是由于在 K 准则的理论推导阶段, 并未考虑保证流动相似的相应准则。在低压模拟条件下, 观察到燃烧室头部靠近轴线附近形成了明显的中心回流区。该现象说明, 当燃料喷射位置偏离火焰面时, 会引起局部气流紊乱, 从而造成燃烧室内部温度升高。此外, 在燃烧室头部壁面附近也出现了一些“涡”, 这些“涡”并不会随燃烧室工作压力或燃油质量流量的变化而消失。这表明, 当燃烧过程中产生大量气体时, 会导致燃料与空气之间存在强烈的相互作用。燃烧

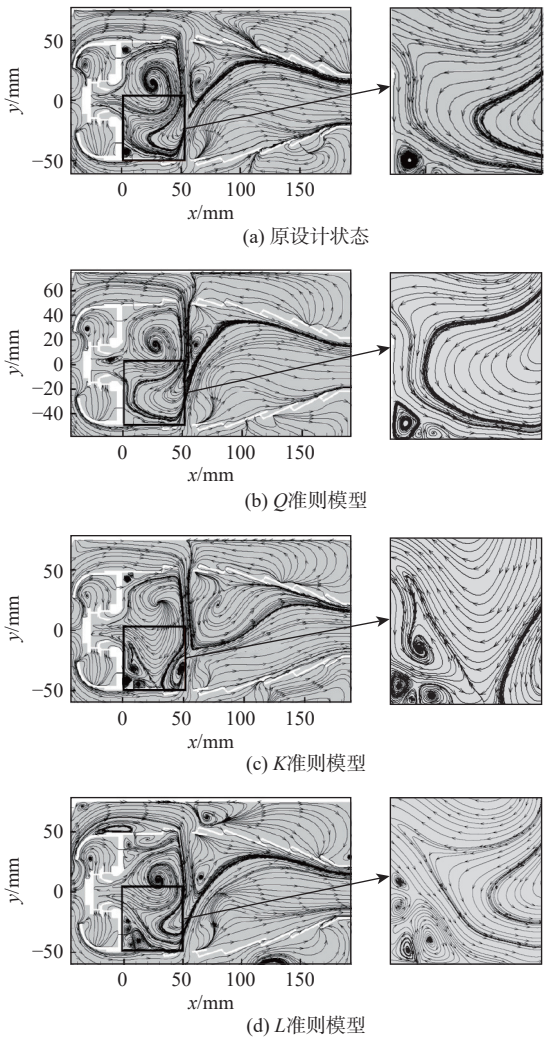


图8 原设计状态和不同模化状态的流场结构

Fig. 8 Flow field structure of original design state and different modeling states

室头部的特殊构造是其形成的主要因素。因此，对于燃烧室的设计和优化，需要充分考虑这些因素，以实现更接近理论推导的流动状态。

2.3 温度分布

在工程领域中，原设计状态的温度场非常重要，模化后的温度场结果越接近原设计状态，模化效果越好，越有利于保证工程的稳定性和可靠性。对比轴向原设计状态和不同模化状态的温度情况，如图9所示。可见，在头部旋流器出口的高温区前段，*K*准则和*Q*准则的差异显著，与原设计状态相比，模化后的出口温度明显更高，且主燃孔后侧的区域高温区的分布和面积与原设计状态存在一定的差异。使用*Q*准则模化得到的温度场与原设计状态更为吻合，但在主燃孔的前侧高温区，局部温度和平均温度相对于原设计状态略有升高。因此，需要关注局部高温和平均温度过高的现象，以确保燃烧室的稳定性和可靠性。

对比切向原设计状态和不同模化状态的温度

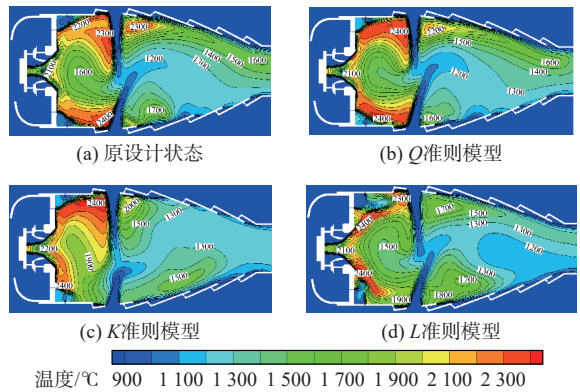


图9 原设计状态和不同模化状态的温度场分布

Fig. 9 Temperature field distribution of original design state and different modeling states

情况，沿轴向分别取主燃孔前后*x*=30 mm和*x*=70 mm处截面，温度分布对比如图10和图11所示。从图10可见，原设计状态在主燃孔前侧轴向切面*x*=30 mm处的高温区分布呈环状，局部最高温达2 500 K。*Q*准则模化得到的切向温度场分布与原设计状态最为接近，但局部最高温2 500 K的面积略小于原设计状态。*K*准则模化得到的高温区呈条状分布，并未形成环状，与原设计状态存在一定差距。*L*准则模化得到的高温区部分呈环状分布，且高温区的面积远小于原设计状态。

从图11可见，原设计状态在主燃孔后侧轴向切面*x*=70 mm处的高温区分布呈对称点状，分别在

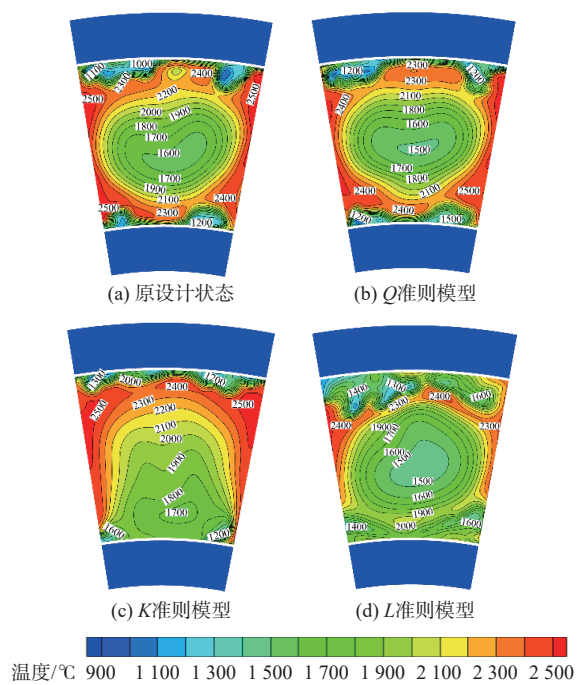


图10 原设计状态和不同模化状态轴向切面*x*=30 mm处的温度场分布对比

Fig. 10 Comparison of temperature field distribution at axial section *x*=30 mm between original design state and different modeling states



图 11　原设计状态和不同模化状态轴向切面 x=70 mm 处的温度场分布对比

Fig. 11　Comparison of temperature field distribution at axial section x=70 mm between original design state and different modeling states

燃烧室的上下靠近壁面处及中心轴线的上侧形成高温区。Q 准则模化得到的切向温度场分布与原设计状态最为接近, 但高温区面积略小于原设计状态, 低温区面积略大于原设计状态。K 准则模化得到的燃烧室高温区只存在于燃烧室上侧, 燃烧室下侧靠近壁面的高温区消失。L 准则模化的温度分布与原设计状态基本一致, 但总体模化温度较原设计状态稍低。

2.4　出口截面温度分布系数

燃烧室出口径向温度分布系数(radial temperature distribution factor, RTDF)α_{RTDF}及周向温度分布系数(overall temperature distribution factor, OTDF)β_{OTDF}是评价出口温度品质的重要数据标准, 其指数也是衡量导向叶片寿命的重要参数, 计算公式如下[1]:

$$\alpha_{\mathrm{RTDF}}=\frac{\left(\overline{T}_{\mathrm{t4r}}-T_{\mathrm{t4ave}}\right)_{\max}}{\left(T_{\mathrm{t4ave}}-T_{\mathrm{t3ave}}\right)} \tag{5}$$

$$\beta_{\mathrm{OTDF}}=\frac{T_{\mathrm{t4max}}-T_{\mathrm{t4ave}}}{T_{\mathrm{t4ave}}-T_{\mathrm{t3ave}}} \tag{6}$$

式中: T_{t4ave}为燃烧室出口截面平均温度; $\overline{T}_{\mathrm{t4r}}$为出口截面某点半径径向平均温度; T_{t3ave}为进口平均温度; T_{t4max}为出口截面最高温度; 下标 max 表示最大值。

图 12 展示了在 3 种模拟条件及原设计状态下, 燃烧室出口截面上的温度分布情况。可以明显看出, 高温区域主要集中于出口截面的左侧。在 Q 准则的模化条件下, 燃烧室出口截面的中心温度比原设计状态下更高; 而 K 准则和 L 准则下的温度比原设计状态下更低。在原设计状态和 L 准则下, 高温区域呈现带状分布; 而在 Q 准则和 K 准则下, 高温区域呈现中心对称分布。涡轮受到的影响主要集中在涡轮叶片的中部, 这与高温区域的分布形态关联性较弱。而叶片强度最高的位置也恰好在中部。故可以得出结论: 高温区域越小, 最高温度越低, 对叶片的影响越小, 该设计被认为更优。

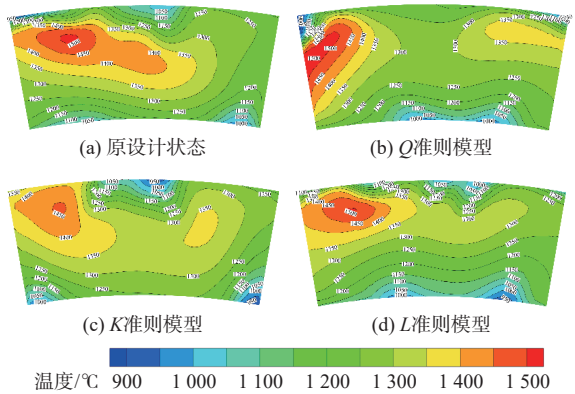


图 12　原设计状态和不同模化状态出口截面温度分布对比

Fig. 12　Comparison of temperature distribution of exit section between original design state and different modeling states

为研究不同模化条件及原设计状态下燃烧室出口温度随高度的变化规律, 在燃烧室出口截面上取左 1/4 处直线, 在该直线上等距离取 100 个点。由图 13 可见, 燃烧室出口温度沿高度的变化规律基本一致。靠近截面两端的模化温度高于原设计状态, 而中间的模化温度低于原设计状态。

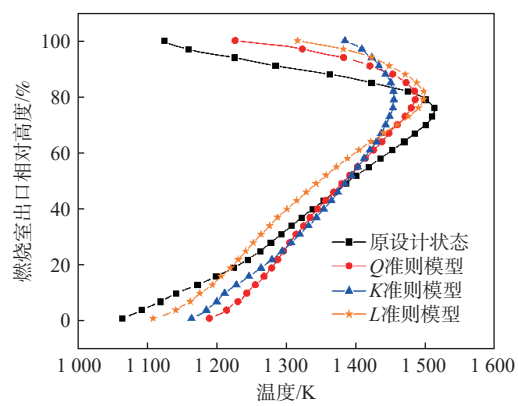


图 13　燃烧室出口温度沿高度的变化

Fig. 13　Combustion chamber outlet temperature change with height

在 3 种不同模拟条件及原设计状态下, 燃烧室出口温度分布指标的具体统计数据如表 5 所示。可以明显看出, 在 3 种模拟条件下, 燃烧室出口截面的平均温度都低于原设计状态。其中, 最大误差仅为 2.48%, 焦树建[24]通过大量试验数据得出, 压

表5 原设计状态和不同模化准则下燃烧室的出口参数

Table 5 Outlet parameters of combustion chamber under original design state and different modeling criteria

状态/准则	出口截面平均温度/K	出口截面最高温度/K	α_{RTDF}	β_{OTDF}
原设计状态	1 289	1 515	0.14	0.46
*Q*准则	1 279	1 577	0.16	0.62
*K*准则	1 288	1 456	0.10	0.34
*L*准则	1 257	1 506	0.21	0.55

力的变化会对化学反应平衡产生影响[25-26]，使在低压模化条件下的燃烧室出口截面平均温度较低。*Q*准则模拟的出口截面最大温度超出预期，而*K*准则和*Q*准则模拟的结果低于原设计状态。OTDF描述了燃烧室出口周向温度的分布情况，其值越小，说明周向温度分布越均匀，燃烧越充分。RTDF反映了燃烧室出口径向温度的变化规律，径向温差越大，RTDF越大。在实际应用中，希望RTDF的值尽可能小，以减小燃烧室内部的温度梯度，降低热应力，从而延长叶片的使用寿命。*K*准则预测的出口截面温度分布系数与原设计状态相比，RTDF降低28.6%，OTDF降低26.1%，即预测值与原设计值相比存在负偏差。计算*Q*准则、*K*准则、*L*准则出口截面温度分布系数的平均误差，分别为24.6%、27.4%、34.8%，可知，*Q*准则的预测值与原设计值最接近。

2.5 燃烧效率

使用热焓法计算燃烧效率（结果见表6）：

$$\eta=\frac{G_{\mathrm{g}}C_{\mathrm{pg}}(T_{\mathrm{g}}^{*}-T_{0})-G_{\mathrm{a}}C_{\mathrm{pa}}(T_{\mathrm{a}}^{*}-T_{0})-G_{\mathrm{f}}C_{\mathrm{pf}}(T_{\mathrm{f}}^{*}-T_{0})}{G_{\mathrm{f}}Q_{\mathrm{H}}^{\mathrm{f}}} \tag{7}$$

式中：G_{g}为燃料和空气质量流量之和；T_{g}^{*}为燃烧室出口截面平均温度；T_{a}^{*}为进口空气平均温度；T_{f}^{*}为进口燃料平均温度；G_{a}为空气的质量流量；G_{f}为燃料的质量流量；$Q_{\mathrm{H}}^{\mathrm{f}}$为参考温度$T_0$（293 K）下燃料的低发热值；$C_{\mathrm{pg}}$为燃料和空气的比定压热容；$C_{\mathrm{pa}}$为空气的比定压热容；$C_{\mathrm{pf}}$为燃料的比定压热容。

表6 原设计状态和不同模化准则下的燃烧效率

Table 6 Combustion efficiency under original design state and different modeling criteria

状态/准则	燃烧效率
原设计状态	0.976 8
*Q*准则	0.955 3
*K*准则	0.958 9
*L*准则	0.947 8

计算燃烧室的燃烧效率时，在低压模化条件下，燃烧室的燃烧效率相较于原设计状态降低2%～3%。这表明，在特定条件下，压力对燃烧效率有显著影响。然而，进口空气温度是影响燃烧效率的重要因素。当进口空气温度保持一致时，随着进口空气压力的升高，燃烧室的效率也会逐渐提高。这一发现进一步强调了压力和温度在燃烧过程中的重要作用。因此，对于燃烧室的设计和优化，需要综合考虑压力、温度及其对燃烧效率的影响，以实现更高的燃烧效率和更环保的燃烧过程。

3 结 论

1）本文获得了以生物航煤为燃料的燃烧室在3种模化状态下的流场分布、轴向和切向温度场分布。对比分析发现，*Q*准则模型预测的燃烧室流场分布及燃烧场分布与原设计状态最为接近，其模化结果优于*K*准则和*L*准则。

2）在预测出口截面温度分布时，*Q*准则模化条件下燃烧室出口截面中心的温度比原设计状态下高，*K*准则和*L*准则下比原设计状态下低。因此，使用*K*准则和*L*准则模化过程中，要提高出口截面温度的设定，以保障飞机发动机的安全性能；*K*准则预测的出口截面温度分布系数与原设计状态相比，存在负偏差，*Q*准则预测结果的平均误差最小。

3）在低压模化条件下，由于压力的变化会对化学反应平衡和燃烧过程产生影响，燃烧效率相较于原设计状态降低2%～3%。

4）由于*L*准则的压力系数过高，模化后的燃油质量流量过小，使其在燃烧的过程中失去非定常捕捉的流动特征，故不在航空发动机的模化试验中使用该准则。

参考文献（References）

[1] 刘威, 王成军, 郑顺, 等. 旋流器结构参数对燃烧室出口温度分布的影响[J]. 上海航天 (中英文), 2020,37(4): 79-87.
LIU W, WANG C J, ZHENG S, et al. Effects of structure parameters of swirlers on the outlet temperature distribution of combustor [J]. Aerospace Shanghai (Chinese& English), 2020, 37(4): 79-87(in Chinese).

[2] 何念, 陈明敏, 高贤智, 等. 低压模化对燃气轮机燃烧室工作特性影响的数值分析[J]. 航空发动机, 2017, 43(2): 10-16.
HE N, CHEN M M, GAO X Z, et al. Numerical analysis on influence of low pressure modeling on operation characteristics of gas turbine combustor[J]. Aeroengine, 2017, 43(2): 10-16(in Chinese).

[3] 任广旭, 霍志成, 刘秀娟. 基于相似理论的航空发动机燃烧室工作过程试验模化研究[J]. 工程与试验, 2010, 50(2): 37-38.
REN G X, HUO Z C, LIU X J. Test model research on the combustor working process of aeronautical engine based on similarity theory[J]. Engineering and Test, 2010, 50(2): 37-38(in Chinese).

[4] STEWARD D G. Scaling of gas turbine combustion systems[M]. London: Butterworths, 1956, 5: 35-43.

[5] DAMKOHLER G. The influence of flow, diffusion and heat transfer on the performance of reaction furnaces. I. General considerations of the transfer of chemical processes from small to large size equipment[J]. Zt. Elektrochemie, 1936, 42: 846-862.

[6] 王宏基. 空气喷气发动机中燃烧室的模化问题[J]. 西北工业大学学报, 1964(2): 39-51.
WANG H J. Modulation of the combustion chamber in an air-jet engine[J]. Journal of Northwestern Polytechnical University, 1964(2): 39-51(in Chinese).

[7] CHILDS J H, GRAVES C C. Correlation of turbine-engine combustion efficiency with theoretical equations[C]//Symposium (International) on Combustion. Pittsburgh: The Combustion Institute, 1957, 6(1): 869-878.

[8] GREENHOUGH V W, LEFEBVRE A H. Some applications of combustion theory to gas turbine development[C]//Symposium (International) on Combustion. Pittsburgh: The Combustion Institute, 1957, 6(1): 858-869.

[9] 陈丹之. 燃烧系统的相似理论与模化法则[J]. 西安交通大学学报, 1963(4): 71-86.
CHEN D Z. The similarity theory and scaling procedure of combustion systems[J]. Journal of Xi'an Jiaotong University, 1963(4): 71-86(in Chinese).

[10] 郭伯伟. 关于燃烧装置模化的准则[J]. 工业炉, 1986(2): 2-9.
GUO B W. Modeling criterion about combustion equipment[J]. Industrial Furnace, 1986(2): 2-9(in Chinese).

[11] 焦树建. 原型燃烧室的低压模拟试验[J]. 燃气轮机技术, 1995, 8(4): 36-43.
JIAO S J. Low pressure simulation test of prototype combustion chamber[J]. Gas Turbine Technology, 1995, 8(4): 36-43(in Chinese).

[12] 汪凤山, 孔文俊, 王宝瑞, 等. 微型燃气轮机燃油燃烧室燃烧特性的模化试验研究[J]. 热能动力工程, 2010, 25(1): 8-11.
WANG F S, KONG W J, WANG B R, et al. Experimental study of the modeling of the combustion characteristics of a micro gas turbine oil-fired combustor[J]. Journal of Engineering for Thermal Energy and Power, 2010, 25(1): 8-11(in Chinese).

[13] 赖哲渊, 李铁, 周昕毅, 等. 柴油机喷雾燃烧相似性理论基础研究[J]. 内燃机工程, 2018, 39(3): 1-7.
LAI Z Y, LI T, ZHOU X Y, et al. Fundamental theory of diesel spray and combustion similarity[J]. Chinese Internal Combustion Engine Engineering, 2018, 39(3): 1-7(in Chinese).

[14] 高玉闪, 刘小勇, 金平. 全流量补燃循环气气燃烧相似性缩尺试验研究[J]. 推进技术, 2019, 40(7): 1554-1559.
GAO Y S, LIU X Y, JIN P. Reduced-scale test of gas-gas combustion similarity in full-flow stage combustion engine[J]. Journal of Propulsion Technology, 2019, 40(7): 1554-1559(in Chinese).

[15] 杨强, 丛敏, 姜雪梅, 等. 燃烧室低压模化试验的数值分析与试验对比[J]. 热能动力工程, 2020, 35(4): 46-51.
YANG Q, CONG M, JIANG X M, et al. Numerical analysis and comparison of combustion chamber low pressure modeling test[J]. Journal of Engineering for Thermal Energy and Power, 2020, 35(4): 46-51(in Chinese).

[16] 沈思远, 宗超, 季晨振, 等. 微小型燃气轮机燃烧室模化方法适用性的数值研究[J]. 热能动力工程, 2021, 36(8): 56-64.
SHEN S Y, ZONG C, JI C Z, et al. Numerical study on applicability of modeling methods for micro gas turbine combustion chamber[J]. Journal of Engineering for Thermal Energy and Power, 2021, 36(8): 56-64(in Chinese).

[17] 王威, 吴艳, 张国强, 等. 中心值班火焰对燃烧室性能影响的试验研究[J]. 热能动力工程, 2020, 35(3): 51-57.
WANG W, WU Y, ZHANG G Q, et al. Experiment study of the influence of the central pilot flame combustion on combustor characteristics[J]. Journal of Engineering for Thermal Energy and Power, 2020, 35(3): 51-57(in Chinese).

[18] 田学清. 燃气轮机燃烧室模化准则的数值模拟研究[D]. 北京: 中国科学院工程热物理研究所, 2004.
TIAN X Q. Numerical simulation study of modeling criterion for gas turbine combustion chambers[D]. Beijing: Institute of Engineering Thermophysics, Chinese Academy of Sciences, 2004(in Chinese).

[19] 赵钦煊. 相似原理在燃烧室研制中的应用[J]. 航空制造工程, 1997(5): 18-20.
ZHAO Q X. Application of similarity principle in combustion chamber development[J]. Aviation Engineering & Maintenance, 1997(5): 18-20(in Chinese).

[20] 朱杰文, 刘江, 翁培奋, 等. 燃气轮机燃烧室模化试验方法研究[J]. 上海电力学院学报, 2019, 35(2): 127-131.
ZHU J W, LIU J, WENG P F, et al. Study on the modeling test of gas turbine combustion chamber[J]. Journal of Shanghai University of Electric Power, 2019, 35(2): 127-131(in Chinese).

[21] 张雪琴, 黄勇. 某旋流杯环形燃烧室三维湍流流动的数值模拟[C]//中国航空学会第六届动力年会. 北京: 航空工业出版社, 2006: 242-248.
ZHANG X Q, HUANG Y. Numerical simulation of three-dimensional turbulent flow in a circular combustion chamber of a swirl cup[C]//Proceedings of the 6th Annual Meeting of the Chinese Society of Aeronautics and Astronautics. Beijing: Aviation Industry Press, 2006: 242-248(in Chinese).

[22] 杨飞. 某型生物航空煤油的基础特性研究[D]. 南京: 南京航空航天大学, 2020: 35.
YANG F. Study on the basic characteristics of bio-aviation kerosene[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020: 35(in Chinese).

[23] 王秧, 王静波, 李象远. RP-3 航空燃料中低温燃烧机理构建及动力学模拟[J]. 化学研究与应用, 2018, 30(6): 946-952.
WANG Y, WANG J B, LI X Y. Low and intermediate-temperature mechanism and simulation for the combustion of RP-3 kerosene[J]. Chemical Research and Application, 2018, 30(6): 946-952(in Chinese).

[24] 焦树建. 燃气轮机燃烧室[M]. 北京: 机械工业出版社, 1981: 301-318.
JIAO S J. Gas turbine combustor[M]. Beijing: China Machine Press, 1981: 301-318(in Chinese).

[25] CUI Y F, XU G, YU B, et al. The effects of pressure on gas turbine combustor performance: an investigation *via* numerical simulation [C]//Proceedings of the ASME Turbo Expo 2006: Power for Land, Sea, and Air. New York: ASME, 2008: 533-540.

[26] LEFEBVRE A H. Gas turbine combustion[M]. Washington, D. C. : Hemisphere Pub. Corp. , 1983: 311-379.

Numerical simulation of low-pressure modeling of bio-jet coal combustion

HE Yantong[1]，DENG Tian[1]，WANG Wei[2,*]

(1. Sino-European Institute of Aviation Engineering，Civil Aviation University of China，Tianjin 300300，China；
2. School of Airworthiness，Civil Aviation University of China，Tianjin 300300，China)

Abstract：In the context of global energy demand growth and environmental pollution intensification, bio-jet coal as a clean, low-carbon, sustainable fuel, leading the future fuel development. The prototype test of an aero-engine combustion chamber is limited by practical conditions, while the low-pressure modeling test simplifies working parameters, effectively solves practical problems and meets engineering requirements. Three modeling criteria are used in this article's low-pressure modeling numerical simulation of the combustor, which uses bio-jet as fuel. The results are compared to those computed in the original design condition. The results show that the flow field distribution predicted by the equal volume flow rate criterion (Q criterion) is the closest to the original design state, and the axial and tangential temperature field distribution is better than the equal combustion efficiency criterion (K criterion) and L criterion. Under the condition of low-pressure modeling, the combustion efficiency is reduced by 2% to 3% compared to the original design state. Especially when the pressure index is too high, the fuel flow after the L criterion is too small, and it is not suitable for aero-engine modeling test. The prediction results of the temperature distribution coefficient of the exit section show that the predicted value of the K criterion has a negative deviation compared with the original design value. The Q criterion has the smallest average error in predicting results. In conclusion, the Q criterion is the best modeling criterion, and the results of its computations can be used as a guide for the low-pressure modeling test of the combustion chamber of an aero engine.

Keywords：aero-engine；low-pressure modeling；numerical simulation；Q criterion；bio-jet coal

Received：2023-12-20；**Accepted**：2024-04-02；**Published Online**：2024-04-10 11：07

URL：link.cnki.net/urlid/11.2625.V.20240408.1837.004

Foundation item：Scientific Research Project of Tianjin Municipal Education Commission (2020KJ036)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0834

红外遥感弱小船舶目标检测算法

谢友晨，徐天阳*，汤张泳，吴小俊
(江南大学 人工智能与计算机学院，无锡 214122)

摘 要：针对红外遥感弱小目标检测面临的挑战，提出一种面向弱小船舶目标的检测算法(YOLO-WIT)。针对弱小目标对位置偏移低容忍的难点，优化削减检测头，降低模型体积的同时限制锚框的最大偏移量，并设计了一种复合距离相似度度量作为损失函数降低回归分支敏感度，从而提升定位精度；针对图像背景高亮而目标灰暗的情况，设计了联接型信息扩展模块(DECO)保留微弱信号，增强对微弱特征的感知；为区分与船舶目标灰度幅值、形状相似的干扰物，利用Sobel算子求解浅层特征图的一阶导数，以边缘特征指导模型进行判断，并采用时空注意力机制进行特征强化。在数据集NUDT-SIRST-Sea上实验验证：与基线模型相比，所提算法参数量减少31.6%，在指标E_{mAP50}上提升9%，$E_{\mathrm{mAP50\text{-}95}}$提升4.9%；与主流检测算法相比，所提算法所需资源更少，模型体积仅为9.2×10^{6}，对弱小船舶目标的检测效果更加显著。

关 键 词：弱小目标检测；红外遥感；YOLOv5；海面船舶检测；深度学习
中图分类号：TP391.4; V243.5
文献标志码：A **文章编号**：1001-5965(2026)03-0895-13

随着空间遥感技术的快速发展，遥感图像广泛应用于各个检测识别领域[1]。其中，海面船舶检测在民生、军事2方面都有重要价值。民生方面，中国海外贸易繁荣、海洋资源丰富，航行船舶数量与种类与日俱增，海上交通事故常有发生[2]，海面船舶检测对事故前预警、事故后搜救都有关键作用。而在军事上，境外船舶以海域航行自由为幌子进行海权侵犯、在争议岛礁巡游的情况时有发生[3]，提升领海监测能力是维护海洋安全的重要保障。

现有遥感船舶检测主要依靠可见光、合成孔径雷达(synthetic aperture radar, SAR)、高光谱和红外成像技术[4]。可见光具有传感范围广、纹理特征丰富、分辨率高的优势，但在光线反射条件差的情况下难以识别目标。SAR可以实现全天候和长距离探测，但在海面探测过程中容易受到电波、岛屿、港口[5]和大气噪声影响。不同于可见光成像，红外遥感成像烟雾穿透性强、不受光照条件限制；红外成像被动接受辐射，比SAR具有更强的隐蔽性[6]。因此，基于红外成像的遥感检测方法在复杂海况中具备通用性。

其中，弱小目标是制约红外成像系统检测性能的关键因素[7]。自2012年深度学习[8]浪潮席卷计算机视觉领域，越来越多先进检测算法给弱小目标检测带来新思路[9]。在通用检测算法层面，YOLO[10]是现今主流的单阶段检测器，具有检测速度快、模型轻量等特点。而二阶段检测算法则以区域卷积神经网络(faster region-based convolutional network, Fater R-CNN)[11]这一系列为主，在多尺度检测、可解释性上具备优势。另外还有基于Transformer的检测算法检测Transformer(detection Transformer, DETR)[12]系列，此类算法注重图像全局关系建模，对局部细节的感知能力不足，因此，在小目标检测领域中表现较差。

针对可见光图像中的弱小目标，已有诸多策略从不同角度强化检测性能：路径聚合网络(path aggregation network, PAnet)[13]受到特征金字塔(feature

收稿日期：2023-12-25；录用日期：2024-05-03；网络出版时间：2024-06-04 14:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20240604.1045.001
基金项目：国家自然科学基金(62106089,62020106012,62332008)
*通信作者. E-mail：tianyang.xu@jiangnan.edu.cn

引用格式：谢友晨，徐天阳，汤张泳，等. 红外遥感弱小船舶目标检测算法[J]. 北京航空航天大学学报，2026，52(3)：895-907. XIE Y C，XU T Y，TANG Z Y，et al. Detecting small ship targets via infrared remote sensing[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：895-907 (in Chinese).

pyramid network, FPN)[14]浅层与深层特征进行交互的设计启发,利用双向路径丰富了特征层次,用浅层精准的位置信息增强深层高级语义特征;文献[15]设计了一种跨层注意力模块,以获得更强的小目标响应;文献[16]采用提出的点距离(dot distance, DotD)(定义为2个包围盒中心点之间的归一化欧氏距离)来代替常用的交并比(intersection over union, IoU)。

近年来,也出现了面向红外图像的弱小目标检测算法:文献[17]设计了一个非对称的上下文模块来对高层和低层语义特征进行调制;文献[18]提出了一种大小和极性(亮或暗)未知的低对比度红外目标检测方法,采用背景减法和对数直方图变换对目标进行增强;文献[19]设计了一种基于空间、频率注意力的解码器,用以提取空间上下文和频域上下文提高特征表示能力;文献[20]提出了一种基于梯度相关滤波和梯度对比度测量的方法,用以提高复杂背景下对目标的检测能力。

与可见光图像和一般的红外图像相比,红外遥感船舶检测图像拍摄距离更远[21],船舶目标在图像中尺寸占比更小,主流深度检测算法使用IoU作为定位损失函数,对尺度极小的目标位置偏移容忍度很低,难以实现精确定位。同时红外遥感目标具有低对比度、低信噪比[22]的特点,形状和纹理特征不明显,易被背景杂波的噪声淹没而发生漏检[23]。

此外,海面充满岛礁、海浪、积云等与船舶目标亮度幅值相似的干扰物,容易导致错检;再者,复杂的场景也给检测带来很大挑战[24],陆地区域较多的海岸、河道、港口场景,而陆地建筑在红外遥感图像中往往是温度较高的高亮区域,灰暗的船舶目标在提取语义的过程中易被模型忽略。因红外遥感船舶检测的以上特点,基于传统方法的船舶检测抗干扰能力弱、泛化性不足,而现有基于深度学习的检测算法则针对性不强,均难以实现对红外遥感船舶目标实现准确的识别、精准的定位。

针对上述难点,本文以YOLOv5S为基线,设计了一种面向红外遥感弱小船舶目标的检测算法(you only look once for weak infrared remote sensing target, YOLO-WIT),并通过实验进行验证,探索弱小目标检测的本质,本文主要创新点如下:

1)针对弱小目标对边界框定位偏移低容忍的特性,基于YOLOv5S设计了一个适用于弱小目标的模型框架,并引入复合距离相似度度量提升定位精度。

2)提出了联接型信息扩展模块(dense concat information expansion convolution block, DECO)存储、强化弱小目标特征,以减少因背景区域信号过于显著而导致空间降采样过程中弱小目标信息的损失。

3)在特征拼接(concat)过程中采用Grad模块求解一阶导数提取边缘信息,通过待检测目标与海面之间的像素梯度判断其是否为船舶目标,并采用空间、通道注意力机制进行特征增强。

1 数据处理与分析

由于现有红外弱小目标检测数据集很少,本文将红外卫星遥感船舶目标分割数据集NUDT-SIRST-Sea[25]进行处理,生成用于检测任务的标注,并将其作为主要验证数据。NUDT-SIRST-Sea包含5 687张切片图像(patch,分辨率大部分为1 024像素×1 024像素,另一部分为1 024像素×740像素),17 598个船舶分割实例。其拍摄范围包含海面、海岸、河道、陆地等多个场景,在样本的多样性和丰富度上都能较好体现红外弱小船舶目标的特点。如图1(a)~图1(d)所示,船舶目标尺寸极小,比其他红外小目标分割数据集[17,26-27]中目标占背景的比值都要小,检测难度更大。

1.1 数据处理

1.1.1 数据集划分

数据集NUDT-SIRST-Sea由48张低轨道卫星上传感器捕获的红外图像处理而来,每幅图像覆盖面积约1×10^{10} m^2,像素为10 000像素×10 000像素。为保证检测的连贯性,以整张图像为单位进行划分:前35张作为训练集,总共4 235张patch、11 193个实例;其余用于验证和测试,为平衡两者的实例数目符合1∶2的比例,选定第35~43张作为测试集,总共968张patch、3 921个实例;第44~48张作为验证集,总共605个patch、1 815个实例。

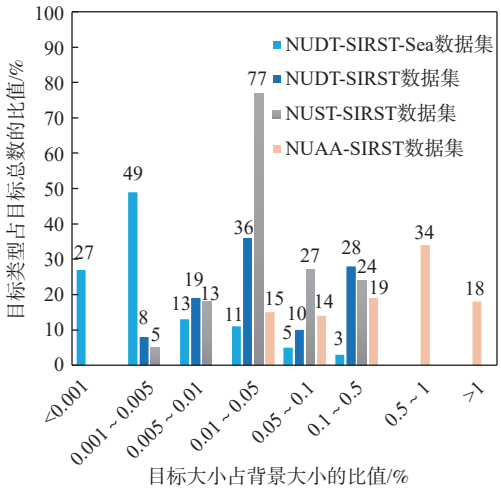
1.1.2 标注生成

NUDT-SIRST-Sea数据集原本面向于分割任务(segmentation),其标签为实例分割像素级手工标注,因此,在进行检测任务之前需利用原先的标注生成能够被检测算法使用的格式(标签的第1列为类别,后4列依次是:归一化后的中心点x、y坐标,归一化后的目标框宽、高)。标注转换流程:①从掩码提取轮廓;②求解轮廓外接矩形;③获得外接矩形的宽、高和中心点位置;④将矩形的中心点坐标和宽高除以图片的尺寸进行归一化。

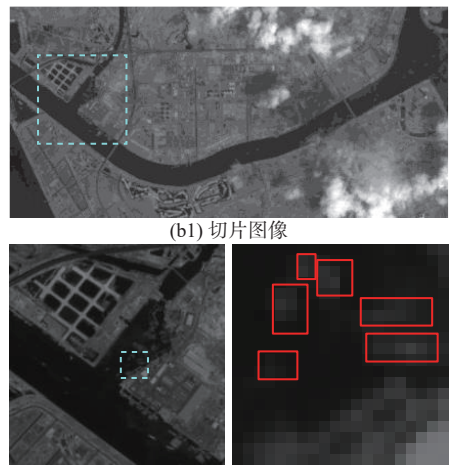
1.2 挑战分析

1.2.1 目标尺寸更小

如图1(b)所示,本文中的船舶目标尺寸极小,平均大小29个像素,最小的目标仅含2个像素,比一般意义上(COCO数据集中32像素×32像素[6,28])的小目标更加微小。而对于微小的物体,像素少的

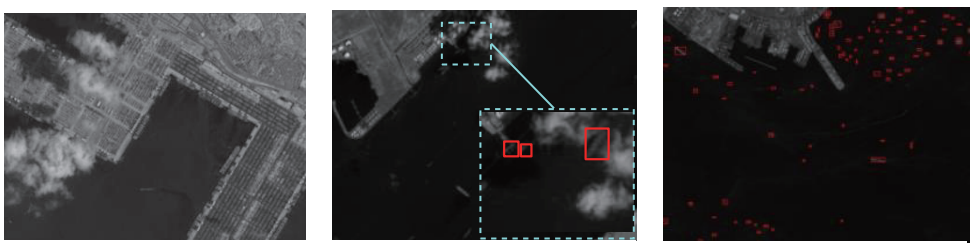


(a) 船舶目标占图像背景比例的分布



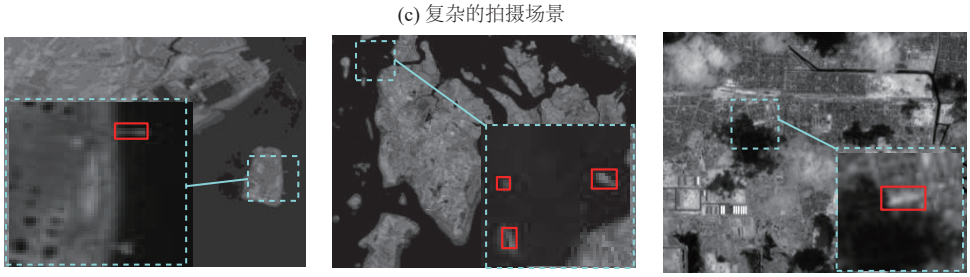
(b1) 切片图像　(b2) 放大5倍　(b3) 放大10倍

(b) 切片图像放大5倍和10倍下的船舶目标



(c1) 停泊港口　(c2) 云层遮挡　(c3) 密集场景

(c) 复杂的拍摄场景



(d1) 近岸设施　(d2) 岛屿礁石　(d3) 陆上建筑

(d) 多种类型可疑相似目标

图 1　数据集的总体描述和挑战分析

Fig. 1　Overall description and challenge analysis of weak and small target datasets

特性会增加定位的难度。

定位作为检测的主要任务之一, 在大多数检测范式中被表述为一个回归问题, 设计定位分支输出包围框偏移量[10-11,29-31] E_{IoU} 度量来评估精度, E_{IoU} 的定义为

$$E_{\mathrm{IoU}}=\frac{T\cap P}{T\cup P} \tag{1}$$

式中: T 为标注框; P 为预测框。E_{IoU} 具体指的是标签框、预测框相交部分与两者并集之间的比值。

如图 2 所示, 可见 E_{IoU} 对不同尺度物体的敏感度差异很大。与大型目标 (岛屿)、中型目标 (礁石) 相比, 小型目标 (船舶) 的预测框位置稍有偏差 (对角线方向 1 个像素, 见图中 B), E_{IoU} 就会显著下降 (从 1 下降到 0.286)。当更大的偏移发生时 (如 4 像素, 见图中 C), 小物体的 E_{IoU} 直接下降到 0, 偏移超过此值时, E_{IoU} 将无法衡量预测框位置的好坏 (E_{IoU} 均为 0)。与大目标相比, 弱小目标对边界框的位置

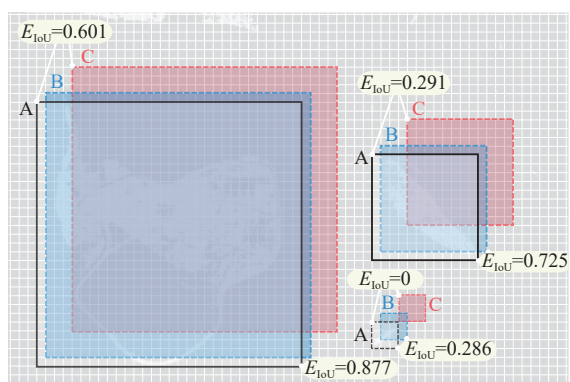


图 2　不同尺度目标 E_{IoU} 对位置偏移的容忍程度

Fig. 2　Tolerance of different scale target IoU to locate offset

扰动容忍度较低[9], 如何提升对弱小目标的定位精度是一个难题。

1.2.2　特征灰暗

本文数据集的拍摄场景复杂多变, 船舶目标与

背景像素的亮度对比并不明显，在图像中难以被观测到。如图 1(c) 所示，在背景噪声和显著区域的干扰下，船舶目标的特征微弱。

如在港口场景，受太阳光照和人类活动影响，陆地建筑比海面的温度更高，因此，在红外图像中呈现出较高的灰度值，而船舶目标在图像中则往往较为黯淡。如何在复杂和高亮的背景中，关注到灰暗微弱的船舶目标，并对其实现检测是一个挑战。

1.2.3 *干扰物多样*

红外遥感图像中除了船舶目标之外，还存在多种疑似目标。如图 1(d) 所示，靠岸的设施、礁石、岸上的建筑等物体在尺寸、形状上都与船舶目标极其相似，因此，如何避免将干扰物错检为船舶目标是一个重要任务。

2 方法介绍

2.1 YOLO-WIT 算法框架

本文以 YOLOv5S 为基线进行设计，改进后的模型结构如图 3 所示，图中 CBS、C3、SPPE 分别为对应的模块缩写，如 CBS 是 Conv2+BatchNorm+SiLU 模块。首先，在预测端(prediction)设计一种复合距离相似度度量作为损失函数，并将基线的 3 个检测头削减到 1 个，以提升模型对弱小目标的定位精度；同时，中间层(neck)也进行了针对性的简化，并设计了 Grad 模块区分干扰目标与船舶目标；在主干网络(backbone)中设计了 DECO 模块保留微弱信号，并采用卷积块注意力模块 (convolutional block attention module, CBAM)时空注意力机制进行特征增强。

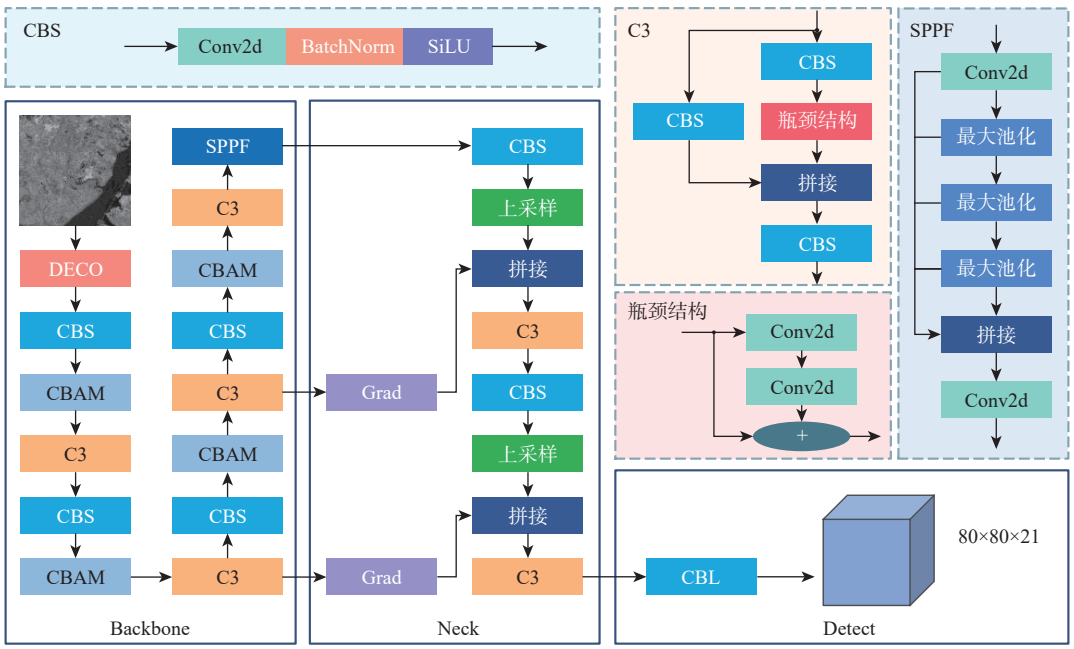


图 3 YOLO-WIT 算法框架总览图

Fig. 3 Overview of the YOLO-WIT algorithm framework

由于 YOLOv5 的定位损失 IoU 在弱小目标场景中难以准确判断预测框的好坏，因此，设计复合距离相似度度量作为损失函数，降低回归分支的敏感度。YOLOv5 最终输出 3 个不同空间维度的特征图 (80 像素×80 像素，40 像素×40 像素，20 像素×20 像素)，并采用预设锚框 (初始位置固定) 加偏移 (学习得到) 的方式进行定位。而数据集 NUDT-SIRST-Sea 中的船舶属于弱小目标，在小尺度特征图中易产生较大的定位偏移。因为当特征图的空间尺度越小，每个格子对应图像中的相应区域就越大，锚框偏移量的上限就越高。因此，本文仅输出一个 80 像素×80 像素的检测头用以计算 Loss，从而限制锚框偏移带来的定位误差上限。

在 Neck 中，为区分干扰目标与船舶目标，设计了 Gard 模块提取边缘信息，在进行特征拼接之前，利用 Sobel 算子求解浅层特征的一阶导数，随后再与深层特征融合，进而指导模型进行判断；在 Backbone 中，为减少因过早进行空间层面的下采样而导致的微弱信号，设计了 DECO 以空间低损的方式提取特征，并采用 CBAM 从通道和空间 2 个层面进行强化，从而进一步提升特征的表示能力。

经过改进的算法被命名为 YOLO-WIT，其整体描述如下所示。

算法 1 YOLO-WIT 的训练与测试算法。

输入 红外图像训练集 $G=\{\boldsymbol{x}(n),\boldsymbol{y}(n)\}_{n=0}^{N}$，测试 $D=\{z(m)\}_{m=0}^{M}$，模型初始权重 $\boldsymbol{W}_0$。

输出　训练模型权重$\boldsymbol{W}$, 检测结果$\boldsymbol{B}_{\text{pred}}$。

定义　损失函数记为f_{Loss},学习率为γ,最大训练轮次为E, 训练和测试的批量大小均为b, μ为动量系数, τ为阻尼系数, f_{ceil}为向上取整函数。

$\boldsymbol{W} \leftarrow \boldsymbol{W}_0$//对模型赋初始值

$B_1 \leftarrow f_{\text{ceil}}\left(\dfrac{N}{b}\right)$//计算训练批量数目

$B_2 \leftarrow f_{\text{ceil}}\left(\dfrac{M}{b}\right)$//计算测试批量数目

for $t = 1$ to E do: //训练

　for $i = 0$ to B_1 do: //对Sobel中的样本进行随机重排序后执行

　　if $(i+1)b < N$:

　　$\boldsymbol{X} \leftarrow \{\boldsymbol{x}(n)\}_{n=i\times b}^{(i+1)b}$, $\boldsymbol{Y} \leftarrow \{\boldsymbol{y}(n)\}_{n=i\times b}^{(i+1)b}$

　　else:

　　$\boldsymbol{X} \leftarrow \{\boldsymbol{x}(n)\}_{n=i\times b}^{N}$, $\boldsymbol{Y} \leftarrow \{\boldsymbol{y}(n)\}_{n=i\times b}^{N}$

　　$g_i \leftarrow \nabla f_{\text{Loss}}(\boldsymbol{W}(\boldsymbol{X}), \boldsymbol{Y})$//计算梯度

　　if $i > 0$:

　　$\boldsymbol{g}_i \leftarrow \mu \boldsymbol{g}_{i-1} + (1-\tau)\boldsymbol{g}_i$

　　else:

　　$\boldsymbol{g}_i \leftarrow \boldsymbol{g}_i$

　　$\boldsymbol{W} \leftarrow \boldsymbol{W} + \gamma \boldsymbol{g}_i$

　for $j = 0$ to B_2 do: //测试

　　if $(j+1)b < M$:

　　$\boldsymbol{Z} \leftarrow \{z(m)\}_{m=j\times b}^{(j+1)b}$

　　else:

　　$\boldsymbol{Z} \leftarrow \{z(m)\}_{m=j\times b}^{M}$

　　$\boldsymbol{B}_{\text{box}} \leftarrow \boldsymbol{W}(\boldsymbol{Z})$//产生预测框

　　$\boldsymbol{B}_{\text{pred}} \leftarrow f_{\text{plot}}(\boldsymbol{B}_{\text{box}}, \boldsymbol{Z})$//在图像中绘制检测框

在训练过程, 采用 Pytorch 框架中的动量随机梯度下降对模型进行优化。具体地, 每次迭代随机选取一个批次的训练数据, 计算该批次中模型的梯度, 并与上一批次的梯度进行加权, 随后利用历史加权后的梯度对模型进行更新。所有训练轮次完成之后, 输出得到的模型参数作为后续测试的模型权重, 以实现对测试图像中弱小船舶目标的检测。

2.2　DECO 模块

本文大部分船舶目标在图像中暗淡微弱, 主流算法检测过程中易发生漏检。基于卷积神经网络的检测算法图像输入后, 通常采用步长为 2 的卷积提取特征, 尽管卷积之后语义信息得到加强, 但会使特征图的尺寸减半, 空间位置信息缺失[32], 船舶目标由于特征信号微弱, 更易被当作非主要信息而受到降采样的影响。当监督信号通过反向传播抵达卷积层时, 信息损失已然发生, 大量灰暗微弱的船舶目标特征被网络忽略, 从而导致漏检。

针对以上特点, 受文献 [33] 启发, 本文设计了 DECO, 将原始输入图像以空间信息低损的方式提取成特征, 保留弱小船舶目标的信号, 以提升模型对微弱特征的感知能力, 模块结构如图 4 所示。

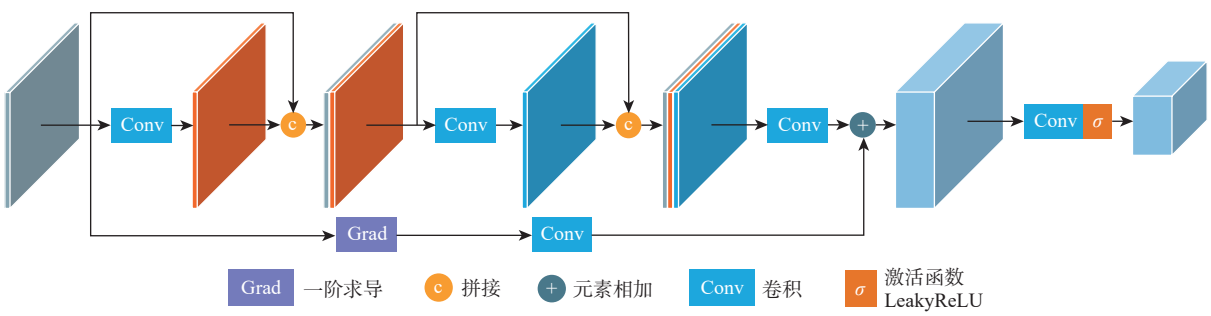


图 4　DECO 模块结构图

Fig. 4　Structure diagram of DECO module

在 DECO 模块主干部分, 采用 2 个步长为 1 的卷积层提取特征, 减少图像降采样的空间信息损失, 并通过拼接整合提取到的信息, 起交互、存储的作用。具体, 对于批量大小为 1、通道数为 3、尺寸为 640 像素×640 像素的图像输入$\boldsymbol{X}^{3\times640\times640}$, 采用 3 像素×3 像素的卷积核进行步长为 1 的卷积, 并将其输出与输入拼接在一起得到特征图$\boldsymbol{P}^{6\times640\times640}$, 如下:

$$\boldsymbol{P}^{6\times640\times640} = F_{\text{cat}}[f_{\text{conv}}^{3\times3,1}(\boldsymbol{X}^{3\times640\times640}), \boldsymbol{X}^{3\times640\times640}] \tag{2}$$

式中: $f_{\text{conv}}^{3\times3,1}$表示卷积核大小 3 像素×3 像素、步长为 1 的卷积函数; F_{cat}表示拼接函数。

随后将特征图$\boldsymbol{P}^{6\times640\times640}$作为输入, 重复上述过程, 并对其输出采用 1×1 的卷积核进行卷积得到特征图$\boldsymbol{P}_1^{32\times640\times640}$, 如下:

$$\boldsymbol{P}_1^{32\times640\times640} = f_{\text{conv}}^{1\times1,1}(F_{\text{cat}}[f_{\text{conv}}^{3\times3,1}(\boldsymbol{P}^{6\times640\times640}), \boldsymbol{P}^{6\times640\times640}]) \tag{3}$$

为避免模型进行反向传播过程中出现梯度消失的情况, 模块设计了跳跃连接部分: 图像输入之后对其一阶求导提取边缘信息, 然后, 利用卷积层将特征图的通道数扩充到 32, 如下:

$$\boldsymbol{P}_2^{32\times640\times640} = f_{\text{conv}}^{1\times1,1}(F_{\text{grad}}[\boldsymbol{X}]) \tag{4}$$

式中: F_{grad}表示求解图像一阶导数的函数。

在对图像进行空间信息低损的特征提取, 保留

足够的微弱信息之后，对特征图压缩以便于学习更高级的语义信息：将主干部分输出 $\boldsymbol{P}_1^{32\times640\times640}$ 和跳跃连接部分的输出 $\boldsymbol{P}_2^{32\times640\times640}$ 元素相加，进行步长为2的卷积把特征图宽高减半，然后，采用LeakyReLU激活函数实现非线性映射，增强模型拟合数据分布的能力，最终得到形状为32像素×320像素×320像素的特征图。具体为

$$\boldsymbol{S}^{32\times320\times320}=f_{\mathrm{relu}}(f_{\mathrm{conv}}^{3\times3,2}(\boldsymbol{P}_1^{32\times640\times640}+\boldsymbol{P}_2^{32\times640\times640})) \tag{5}$$

式中：f_{relu} 表示LeakyReLU激活函数。

2.3 特征增强

为对干扰物与待检测的船舶目标进行区分，本文采用了特征增强策略。由1.1.3节所述，图像数据集中存在积云、岛屿、近岸建筑等与船舶目标的尺寸、形状相似，但与海面之间的像素幅值对比存在差异的干扰物。

为利用待检测目标与海面之间的边缘信息，设计了Grad模块求解一阶导数以获得边缘信息进行特征增强。采用Sobel算子对特征图求解一阶导数，其中，对$\boldsymbol{x}$方向进行求导的Sobel($\boldsymbol{x}$)算子如下：

$$\mathrm{Sobel}(\boldsymbol{x})=\begin{bmatrix}1&0&-1\\2&0&-2\\1&0&-1\end{bmatrix} \tag{6}$$

对Sobel($\boldsymbol{x}$)算子转置即为对$\boldsymbol{y}$方向求导的算子 $\mathrm{Sobel}(\boldsymbol{y})=(\mathrm{Sobel}(\boldsymbol{x}))^{\mathrm{T}}$。对于输入特征图$\boldsymbol{P}$，将其2个方向的一阶导数幅值相加，可近似得到输入的一阶导数：

$$F_{\mathrm{sobelxy}}(\boldsymbol{P})\approx F_{\mathrm{abs}}[F_{\mathrm{sobelx}}(\boldsymbol{P})]+F_{\mathrm{abs}}[F_{\mathrm{sobely}}(\boldsymbol{P})] \tag{7}$$

式中：F_{abs} 为取绝对值函数；F_{sobelx} 为卷积核为算子Sobel($\boldsymbol{x}$)的卷积函数，F_{sobely} 为卷积核为Sobel$\boldsymbol{y}$算子的卷积函数。

与现有算法不同，在本文的拼接操作中，首先，利用Grad模块进行增强：对被拼接的浅层特征求解一阶导数，并将求得的一阶导数与原浅层特征图相加，然后，再与待拼接的深层特征进行拼接融合。由于船舶目标和干扰物与海面的像素梯度存在差异，而一阶导数中包含特征的幅值对比度信息，因此，可以指导模型分辨干扰物与真正的船舶。此外，引入浅层特征的一阶导数可以增强模型的定位能力：网络经过多层卷积之后，空间位置关系逐渐模糊，浅层特征的一阶导数含有丰富的位置信息，可以提升弱小目标的定位精度。

随后采用CBAM[34]模块从空间和通道2个层面对特征作进一步增强，模块结构如图5所示。实验表明：把CBAM模块放置于Backbone的卷积模块CBS前，之后能够提升模型的检测性能(见3.3节)，CBS模块由一个2D卷积层(Conv2D)、批量归一化(batch normalization, BN)层及SiLU激活函数组成，当SiLU激活函数替换为Leaky_ReLU激活函数模块简写为CBL。CBAM注意力模块在空间层面上选择关键的位置信息进行增强，以提高模型的局部感知能力；而在通道层面则有选择性地对重要模式的语义信息进行增强。

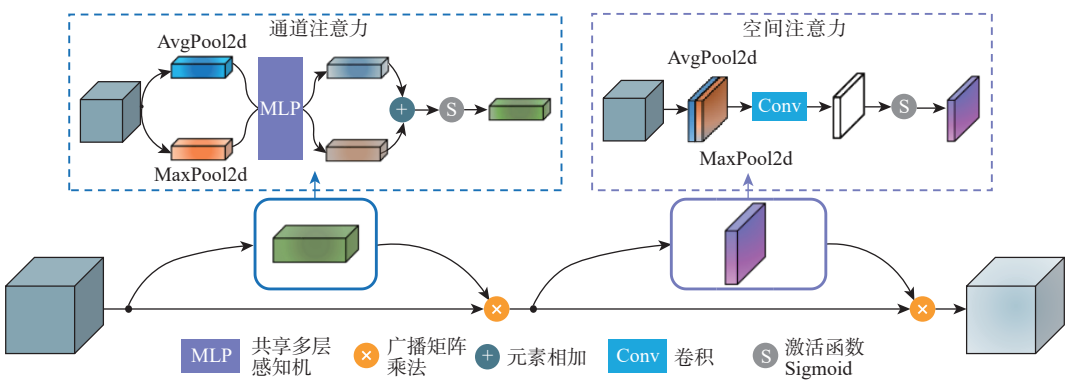


图5 CBAM模块结构

Fig. 5 Structure of CBAM module

2.4 复合距离相似度度量

基线算法YOLOv5S定位损失计算采用的是IoU，IoU是预测框和真实框交集与两者并集的比值，当预测框和真实框无交集时，IoU恒等于0，此时IoU无法衡量预测结果的好坏。又因本文的检测目标属于小目标，其预测框和真实框都很小。如图2所示，小目标的预测框位置稍有偏移IoU就会大幅减小，也容易出现预测框IoU为0的情况，故而IoU无法准确衡量弱小船舶的定位精度。为解决此难题，本文设计了一个中心点位置扰动敏感度低且预测框与真实框无须重叠也能判断其好坏的度量方式。

受归一化瓦瑟斯坦距离(normalized Gaussian Wasserstein distance, NWD)[35]的启发，本文提出一种适用于弱小目标的复合距离相似度度量作为损失函数，用以改善IoU度量小目标位置偏移容忍度低[9]

的弊端, 具有更好的抗扰动能力。

首先, 引入中心点欧氏距离度量:

$$D^2_{\text{center}} = (x_1 - x_2)^2 + (y_1 - y_2)^2 \tag{8}$$

式中: x_1、y_1分别为预测框中心点x、y方向的坐标; x_2、y_2分别为真实框中心点x、y坐标。

引入预测框与真实框之间的宽高 (广义上的) 距离度量:

$$D^2_{wh} = (w_1 - w_2)^2 + (h_1 - h_2)^2 \tag{9}$$

式中: h_1、w_1分别为预测框的高度、宽度; h_2、w_2分别为真实框的高度、宽度。

为均衡不同尺度的目标, 引入凸对角线的平方作为距离度量的量纲统一除数项:

$$\begin{aligned} D^2_{\text{con}} = {} & (F_{\max}[b_1x_2, b_2x_2] - F_{\min}[b_1x_1, b_2x_1])^2 + \\ & (F_{\max}[b_1y_2, b_2y_2] - F_{\min}[b_1y_1, b_2y_1])^2 \end{aligned} \tag{10}$$

式中: $F_{\max}$表示取最大值函数; $F_{\min}$表示取最小值函数; (b_1x_1, b_1y_1)、(b_1x_2, b_1y_2)分别为真实框左上角、右下角的横纵坐标; (b_2x_1, b_2y_1)、(b_2x_2, b_2y_2)分别为预测框左上角、右下角的横纵坐标。

因此, 引入量纲统一项, 并综合考量欧氏距离和宽高距离之后的复合距离度量计算式为

$$R_{\text{com}} = \frac{D^2_{\text{center}} + D^2_{\text{wh}}}{2D^2_{\text{con}}} \tag{11}$$

最后, 将复合距离度量R_{com}利用e的指数进行归一化, 即为值域在 (0, 1] 的相似度度量:

$$r_{\text{com}} = \mathrm{e}^{-K \times R_{\text{com}}} \tag{12}$$

式中: K 为常量归一化系数。

该r_{com}相似度度量对预测框的定位具有更好的距离偏置一致性, 即使与真实框没有重合也能对预测的好坏进行判断。当归一化系数 K 取 4 时, 各相似度度量响应差异如图 6 所示, 由图可以发现, 采用r_{com}作为定位损失函数后, 预测框定位对位置偏移敏感度大幅降低。相比于单一的欧氏距离度量, 引入凸对角线的平方进行平衡之后, 可以维持相同距离偏移下对小目标框更高权重的优势。此外, r_{com}将预测框的宽高纳入衡量因素的一部分, 更加全面的对预测框进行评价。

引入复合距离相似度度量增强对弱小目标的定位能力之后, 还需考虑检测过程中不同任务损失之间的关系。YOLOv5 将检测任务定义为回归任务, 其损失函数由定位损失、置信度损失、分类损失 3 部分构成:

$$L_{\text{oss}} = l_{\text{box}} + l_{\text{obj}} + l_{\text{cls}} \tag{13}$$

式中: l_{box}为定位损失; l_{obj}为置信度损失; l_{cls}为分类

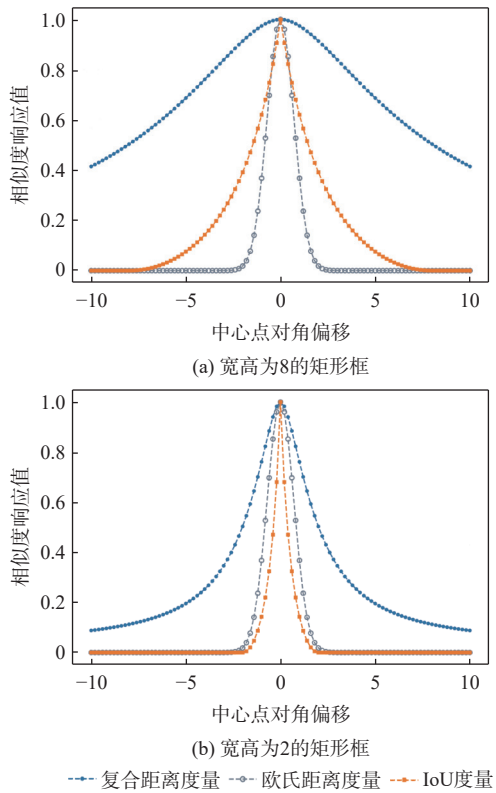


图 6　中心点位置对角偏移的相似度响应曲线

Fig. 6　Similarity response curve of center point position diagonal offset

损失。为平衡各个任务L_{oss}之间的关系, 本文根据文献 [36] 的思想, 设计了一种柔性的加权方式:

$$L_{\text{oss}} = \frac{l_{\text{box}}}{\sigma_1^2} + \frac{l_{\text{obj}}}{\sigma_2^2} + \frac{l_{\text{cls}}}{\sigma_3^2} + \ln(\sigma_1\sigma_2\sigma_3) \tag{14}$$

式中: σ_1、σ_2、σ_3为引入的加权平衡系数, 在训练中被初始化为 1。

针对多个L_{oss}的量纲不统一或反向传播中梯度数量级相差悬殊的情况, 引入加权平衡系数能够平衡各个L_{oss}之间的关系, 从而使得训练更加平稳。其中, $\ln(\sigma_1\sigma_2\sigma_3)$项起正则项的作用, 对加权平衡系数进行约束; 否则, 其加权系数将趋于无穷大而L_{oss}始终为于 0, 导致监督信号无法传导。

3　实验分析

3.1　实验条件

训练平台为 Ubuntu 20.04.4 LTS, GPU 为 NVIDIA GeForce RTX 3090。本文使用的数据集为 1.1 节中转换成检测任务标签之后的 NUDT-SIRST-Sea, 图片输入尺寸固定为 640 像素×640 像素, CUDA 版本为 11.7, PyTorch 版本为 1.13.0, 优化器为 SGD。

3.2　评价指标

本文采用平均精度均值E_{mAP50}、$E_{\text{mAP50-95}}$, 查准率$E_{\text{precision}}$、召回率E_{recall}、参数量$N_{\text{parameters}}$和浮点运算速度指标对模型性能进行综合评价, 其中, E_{mAP50}为

E_{IoU} 阈值为 0.5 时的平均精度均值，$E_{\mathrm{mAP50\text{-}95}}$ 为 E_{IoU} 阈值从 0.5～0.95 时的平均精度均值如图 7 所示，$E_{\mathrm{precision}}$ 是所有预测为正样本的结果中，预测正确的比率：

$$E_{\mathrm{precision}}=\frac{A}{A+B} \tag{15}$$

而 E_{recall} 是所有正样本中，预测正确的比率：

$$E_{\mathrm{recall}}=\frac{A}{A+C} \tag{16}$$

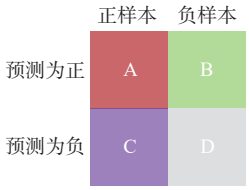


图 7 指标示意图

Fig. 7 Indicator diagram

3.3 消融实验

基线算法的具体版本为 YOLOv5-master 中的 YOLOv5S，为保证公平，所有的消融实验的批量大小均为 32，线程数为 4，最大训练 Epoch 为 1 200。其余参数均为 YOLOv5-master 默认值，其中：随机种子为 0，训练容忍度为 100。

为直观地展示本文算法对检测性能的影响，以基线算法 YOLOv5s 作对比进行消融实验。首先，进行探究的是 YOLOv5S 进行检测头优化设计产生的影响，优化设计后的模型被命名为 YOLOv5-re，其只有一个空间尺度为 80 像素×80 像素的检测头，并且移除了 Neck、Backbone 的一部分，采用复合距离相似度度量作为定位损失。

如表 1 所示：对 YOLOv5S 优化设计限制锚框偏移，并引入复合距离相似度度量的改进，能提升模型对弱小船舶目标的检测性能和定位精度。YOLOv5S-re 相比于基线网络 YOLOv5S 的 $N_{\mathrm{parameters}}$ 减少 31.8%，浮点运算速度减少 18.6% 的条件下 E_{mAP50} 提升 5.7%，$E_{\mathrm{mAP50\text{-}95}}$ 提升 2.8%。

表 1 消融实验结果

Table 1 Ablation experiment results

算法	$E_{\mathrm{precision}}$/%	E_{recall}/%	E_{mAP50}/%	$E_{\mathrm{mAP50\text{-}95}}$/%	$N_{\mathrm{parameters}}$	浮点运算速度/10^9 s^{-1}
YOLOv5S	65.4	33.6	38.7	16.4	7.02×10^6	15.8
YOLOv5S-re	66.3	**38.4**	44.4	19.2	**4.78**$\times10^6$	**12.9**
Yolov5s-re+Crad	70.1	35.9	45.4	19.9	4.79×10^6	13.0
Yolov5s-re+CBAM	66.9	38.3	43.7	18.7	4.79×10^6	12.9
YOLOv5S-re+GradCB	71.7	36.2	45.9	19.9	4.80×10^6	13.0
YOLOv5S-re+DECO	67.6	37.8	44.7	19.6	4.79×10^6	14.6
YOLO-WIT(本文算法)	**75.1**	35.3	**47.7**	**21.3**	4.80×10^6	14.7

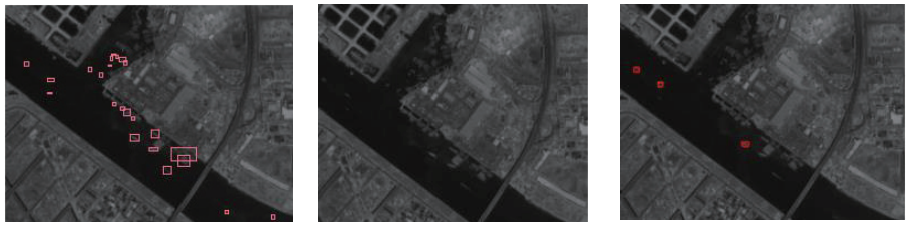
注：粗体表示最优值。

随后将设计的模块和策略依次加入到重构后的 YOLOv5S-re 中，验证其有效性。首先，在 Backbone 中加入 CBAM 特征增强模块，实验结果表明 (见表 1 中 Yolov5s-re+CBAM 一栏)，利用 CBAM 注意力本身并不能提升算法性能；而在利用 Grad 模块求解浅层特征图的一阶导数之后 (Yolov5s-re+Grad)，再加入 CBAM 模块对边缘信息进行增强 (YOLOv5S-re+GradCB)，便能有更好的检测效果，相比于 YOLOv5S-re 在指标 E_{mAP50} 上提升 1.5%，$E_{\mathrm{mAP50\text{-}95}}$ 上提升 0.7%。实验证明直接采用时空注意力机制并不能提升对弱小船舶目标的检测性能，但若有 Grad 模块提取边缘信息加以指导利用，则能关注到弱小目标并进行有效的特征增强。

维持以上改动，再将 YOLOv5S-re 模型的第一个 CBS 替换成 DECO 模块，就是本文算法 YOLOv5-WIT，如图 8(a) 所示，该算法能有效地减少因目标过于微弱导致的漏检，图 8(b) 则展示错检的情况也相应减少。实验结果表明 YOLO-WIT 比 YOLOv5S 的 E_{mAP50} 提升了 9%，$E_{\mathrm{mAP50\text{-}95}}$ 提升 4.9%，同时 $N_{\mathrm{parameters}}$ 减少了 31.6%，浮点运算速度减少 7.0%。

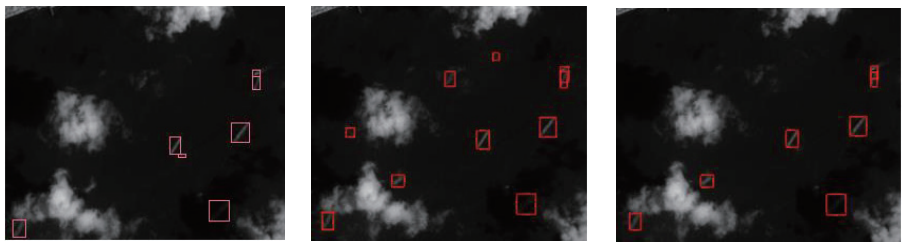
图 9 展示了引入 DECO 模块前后，模型 backbone 部分产生的热力图对比。由热力图可知 YOLOv5S-re+GradCB 模型对更加显著的陆地区域产生较大的响应，而加入 DECO 模块之后的 YOLO-WIT 模型注意力则集中在海面或船舶本身，尽管海面区域在红外遥感图像中呈现像素灰暗、特征微弱的特点。热力图差异说明 DECO 能降低模型在提取语义过程中空间信息的损失，在 Backbone 中便将注意力集中于船舶所在的更重要但灰暗的海面区域，提升了对微弱特征的感知能力。

此外，本文还对 DECO 模块进行了消融探究，如表 2 所示，表中列举了 DECO 模块设计过程中尝试的 4 种结构相应的实验结果：主干部分不采用任何联接、采用 DenseNet 结构进行信息互联、跳跃连接部分不计算梯度直接卷积得到输出、空间降采样时再进行通道扩充。



(a1) 挑战：目标过于微弱　(a2) 未检测出船舶目标　(a3) 增加3个检出目标

(a) 因目标过于微弱而导致的漏检

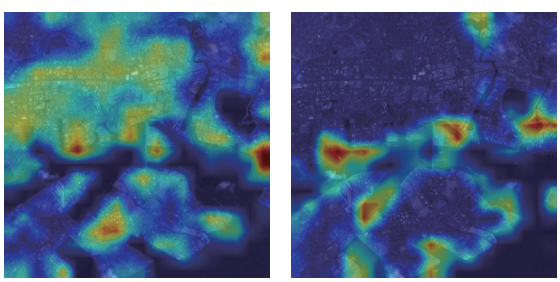


(b1) 挑战：云层干扰
真实标签　(b2) 4片积云误检为船舶
YOLOv5S　(b3) 误检减少到1个
YOLO-WIT

(b) 积云导致的错检

图 8　基线、改进算法预测与真实标签结果对比

Fig. 8　Comparison of baseline and improved algorithm prediction and real label results



(a) YOLOv5S-re+GradCB上的结果　(b) YOLO-WIT上的结果

图 9　Backbone 产生的热力图对比

Fig. 9　Comparison of heat maps generated by Backbone

消融结果如表 2 所示, 实验证明了本文设计的 DECO 模块结构具备合理性和必要性。首先, 采用拼接的方式联接模块主干部分的卷积层, 对于浅层信息的交互、扩展、存储更加有效; 而在跳跃连接部分计算一阶导数能更充分的利用图像的初始位置信息; 在使用卷积对图像进行特征提取的过程中, 同步扩充通道有助于进行模态的识别。

综上, 在定量分析层面, YOLO-WIT 比基线算法在指标 E_{mAP50} 和 $E_{\mathrm{mAP50\text{-}95}}$ 上分别提升 9% 和 4.9%, 同时减少了 31.6% 参数量和 7.0% 浮点运算速度。而在定性分析层面, YOLO-WIT 能更加有效地平衡对微弱信息和显著信息的感知关注度, 从而减少因船舶目标的特征过于微弱而导致的漏检; 并且, YOLO-WIT 利用时空注意力机制增强后的边缘信息, 提升了对干扰物的辨别能力, 减少了将干扰物检测为船舶目标的情况。

表 2　模块消融结果

Table 2　Module ablation results

算法或改动	$E_{\mathrm{precision}}$/%	E_{recall}/%	E_{mAP50}/%	$E_{\mathrm{mAP50\text{-}95}}$/%	$N_{\mathrm{parameters}}$	浮点运算速度/10^9 s^{-1}
DECO模块	**75.1**	35.3	**47.7**	**21.3**	4.80×10^6	14.7
主干部分不采用任何方式联接	71.8	36.8	46.6	20.7	4.80×10^6	14.5
主干采用DenseNet结构联接	69.2	36.9	44.8	19.7	4.80×10^6	14.5
取消一阶导数计算	71.4	**37.1**	46.6	20.3	4.80×10^6	14.7
空间降采样时通道扩充	71.0	35.4	45.7	19.9	$\mathbf{4.79}\times10^6$	**13.1**

注: 粗体表示最优值。

3.4　对比实验

通过对比实验验证算法的性能优势, 因现有红外小目标检测方法[17-20]大多采用分割算法的范式, 无法通过统一的指标进行对比, 故选取已公开代码的检测算法 (YOLOv3[29]、YOLOv7[37]、YOLOv8、YOLOv9[38]、SSD[39]、Faster-RCNN[11]、RetinaNet[30]、DETR[12]、RT-DETR[40]) 作对比, 其中, YOLOv7、YOLOv3 和 YOLOv7-Tiny 在训练 1 200 个 Epoch 后尚未完全收敛, 因此, 最大训练 Epoch 设置为 2 000, 其余实验最大训练 Epoch 均为 1 200。而 DETR 因其在训练 1 012 个 epoch 之后 mAP 始终为 0, 因此, 其训练被终止。

DETR采用Transformer结构，将图片特征分割成片段作为序列后应用注意力机制，优势是具有很强的全局关系映射学习能力，可以打破距离对信息关联的限制。而正因如此，DETR缺乏局部信息归纳能力，难以对此尺寸极小目标进行检测。

实验对比结果如表3所示，YOLO-WIT在评价指标 E_{mAP50} 和 $E_{\mathrm{mAP50\text{-}95}}$ 上均领先于其他对比算法，并且 $N_{\mathrm{parameters}}$ 仅为 4.8×10^6，浮点运算速度仅为 $1.470\times10^{10}\ \mathrm{s}^{-1}$，模型体积为 9.2×2^{20}，远比YOLOv8s之外的模型所需计算资源更少。

表3 对比实验结果

Table 3 Comparative experimental results

算法	主干网络	E_{mAP50}/%	$E_{\mathrm{mAP50\text{-}95}}$/%	$N_{\mathrm{parameters}}$	浮点运算速度/$10^9\ \mathrm{s}^{-1}$	模型体积
YOLOv5S	CSPDarkNet	38.7	16.4	7.010×10^6	15.80	13.4×2^{20}
YOLO-WIT(本文算法)	CSPDarkNet	**47.7**	**21.3**	4.800×10^6	14.70	9.2×2^{20}
YOLOv3	DarkNet-53	24.7	10.3	6.150×10^7	154.60	114.9×2^{20}
YOLOv7		22.7	9.7	3.649×10^7	103.20	69.7×2^{20}
YOLOv7-Tiny		20.0	8.09	6.010×10^6	13.00	45.0×2^{20}
YOLOv8n		23.9	10.8	**3.010**$\times10^6$	**8.10**	**5.9**$\times2^{20}$
YOLOv8s		25.1	11.5	1.110×10^7	28.40	21.5×2^{20}
YOLOv9-c		23.0	10.8	5.070×10^7	236.60	98.0×2^{20}
SSD	SSDVGG	30.0	9.6	2.388×10^7	137.30	182.2×2^{20}
Faster-RCNN	ResNet-50	2.1	1.3	4.113×10^7	91.00	307.6×2^{20}
RetinaNet	ResNet-50	8.5	2.5	3.613×10^7	81.90	276.8×2^{20}
RT-DETR	ResNet-18	22.5	8.7	2.010×10^7	60.00	307.3×2^{20}
DETR	ResNet-50	0	0	4.128×10^7	37.10	474.1×2^{20}

注：粗体表示最优值。

综上所述，本文算法在5类指标的表现均优于其他检测算法，能够有效减少因船舶特征过于灰暗微弱导致的漏检、因干扰物导致的错检，提升定位精度。本文模型在减少参数量、计算量和模型体积的条件下，检测性能依然具备优势，证明YOLO-WIT算法的有效性与先进性。

为进一步验证本文算法对船舶以外的红外弱小目标检测是否有效，分别标注已公开的分割数据集NUAA-SIRST[17]、NUDT-SIRST[27]、IRSTD-1k[28]目标框生成相应的检测数据集，进行补充对比实验，从而探究本文算法在其他红外弱小目标检测场景下的鲁棒性和泛化性。实验结果如表4所示，补充实验表明本文的改进策略对于其他红外弱小目标同样有检测性能上的提升，本文算法具备一定的泛化性。

表4 泛化性验证实验结果

Table 4 Generalizability verification experimental results

算法	$E_{\mathrm{mAP_{50}}}$/%			$E_{\mathrm{mAP_{50-95}}}$/%			$N_{\mathrm{parameters}}$	浮点运算速度/$10^9\ \mathrm{s}^{-1}$	模型体积
	NUAA-SIRST	NUDT-SIRST	IRSTD-1k	NUAA-SIRST	NUDT-SIRST	IRSTD-1k			
YOLOv5S	79.8	96.2	71.9	28.2	74.1	30.1	7.01×10^6	15.80	13.4×2^{20}
YOLO-WIT(本文算法)	**87.3**	98.3	**72.6**	**32.4**	76.4	**32.5**	4.80×10^6	14.70	9.2×2^{20}
YOLOv8n	77.1	97.3	71.3	30.2	75.6	31.2	**3.01**$\times10^6$	**8.10**	**5.9**$\times2^{20}$
YOLOv8S	83.0	97.8	72.3	31.4	75.9	31.8	1.11×10^7	28.40	21.5×2^{20}
YOLOv9-c	73.8	**98.7**	70.2	27.2	**76.5**	31.9	5.07×10^7	236.60	98.0×2^{20}
RT-DETR	72.3	97.3	71.1	31.7	71.7	31.1	2.01×10^7	60.00	307.3×2^{20}

注：粗体表示最优值。

在泛化性验证实验中，与表3的结果相比，本文算法在检测指标上的提升更小，并且在数据集NUDT-SIRST上YOLOv9-c的检测性能已经超过了本文算法。说明本文YOLO-WIT算法在红外弱小目标检测任务上具备一定的泛化性，但依然受制于任务面向的特定场景，算法的适用范围存在边界。

4 结 论

本文设计了一种面向红外遥感弱小船舶目标的检测算法，针对弱小船舶目标尺寸过小定位困

难、特征微弱、干扰物多 3 个挑战, 提出了针对性的模块和策略。

1) 对模型检测头进行了优化设计, 限制了锚框的位置偏移量上限, 减少定位误差; 并设计了一种复合距离相似度度量作为定位损失函数, 降低弱小目标对位置偏移的敏感度从而提升定位精度。

2) 在网络的 Backbone 设计了 DECO 模块, 以空间信息低损的方式将图像提取为特征, 增强对微弱特征的感知能力。

3) 设计了 Grad 模块, 在拼接之前利用浅层特征图的一阶导数引入边缘信息, 并采用 CBAM 进行增强, 提升了对干扰物的辨别能力。

4) 经实验验证, YOLO-WIT 作为浮点运算速度仅为 1.47×10^{10}、模型体积为 9.2×2^{20} 的红外遥感检测算法, 在弱小船舶目标检测任务中具备良好的性能, 比体积更大、浮点运算更多的其他算法在检测指标上表现更优越。此外, 泛化性验证实验表明本文的改进策略对于其他红外弱小目标检测同样有效。

参考文献 (References)

[1] LIU J, HE Z Q, CHEN Z L, et al. Tiny and dim infrared target detection based on weighted local contrast[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(11): 1780-1784.

[2] 卢宇韦. 海面红外弱小目标检测算法研究[D]. 大连: 大连海事大学, 2021.
LU Y W. Research on detection algorithm of infrared dim and small targets on the sea surface[D]. Dalian: Dalian Maritime University, 2021(in Chinese).

[3] 吴科君. 基于深度学习的海面船舶目标检测[D]. 哈尔滨: 哈尔滨工程大学, 2018.
WU K J. Ship target detection on sea surface based on deep learning[D]. Harbin: Harbin Engineering University, 2018(in Chinese).

[4] LI L Y, JIANG L Y, ZHANG J W, et al. A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds[J]. Remote Sensing, 2022, 14(7): 1534.

[5] CHEN P, ZHOU H, YING L, et al. A novel deep learning network with deformable convolution and attention mechanisms for complex scenes ship detection in SAR images[J]. Remote Sensing, 2023, 15(10): 2589.

[6] KOU R K, WANG C P, PENG Z M, et al. Infrared small target segmentation networks: a survey[J]. Pattern Recognition, 2023, 143: 109788.

[7] GUAN X W, ZHANG L D, HUANG S Q, et al. Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy[J]. Remote Sensing, 2020, 12(9): 1520.

[8] KRIZHEVSKY A , SUTSKEVER I , HINTON G .ImageNet classification with deep convolutional neural networks[C]//NIPS.Curran Associates Inc.. [S.l.:s.n.], 2012.

[9] CHENG G, YUAN X, YAO X W, et al. Towards large-scale small object detection: survey and benchmarks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.

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

[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] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//Computer Vision-ECCV 2020. Berlin: Springer, 2020: 213-229.

[13] 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.

[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] LI Y Y, HUANG Q, PEI X, et al. Cross-layer attention network for small object detection in remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14: 2148-2161.

[16] XU C, WANG J W, YANG W, et al. Dot distance for tiny object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2021: 1192-1201.

[17] DAI Y M, WU Y Q, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2021: 949-958.

[18] DAI Y M, WU Y Q, ZHOU F, et al. Attentional local contrast networks for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9813-9824.

[19] ZHANG M J, ZHANG R, ZHANG J, et al. Dim2Clear network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5001714.

[20] ZHANG X Y, RU J Y, WU C D. Infrared small target detection based on gradient correlation filtering and contrast measurement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5603012.

[21] LIU S, CHEN P F, Woźniak M. Image enhancement-based detection with small infrared targets[J]. Remote Sensing, 2022, 14(13): 3232.

[22] XIAO S, MA Y, FAN F, et al. Tracking small targets in infrared image sequences under complex environmental conditions[J]. Infrared Physics & Technology, 2020, 104: 103102.

[23] 詹令明, 李翠芸, 姬红兵. 基于显著图的红外弱小目标动态规划检测前跟踪算法[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1061-1066.
ZHAN L M, LI C Y, JI H B. Dynamic programming track-before-detect algorithm based on saliency map for infrared dim and small target[J]. Journal of Computer-Aided Design & Computer Graphics,

2019, 31(7): 1061-1066(in Chinese).

[24] 黄康, 毛峡, 梁晓庚. 一种新的红外背景抑制滤波算法[J]. 航空学报, 2010, 31(6): 1239-1244.
HUANG K, MAO X, LIANG X G. A novel background suppression algorithm for infrared images[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(6): 1239-1244(in Chinese).

[25] WU T H, LI B Y, LUO Y H, et al. MTU-Net: multilevel TransUNet for space-based infrared tiny ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5601015.

[26] WANG H, ZHOU L P, WANG L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 8509-8518.

[27] LI B Y, XIAO C, WANG L G, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2022, 32: 1745-1758.

[28] ZHANG M J, ZHANG R, YANG Y X, et al. ISNet: Shape matters for infrared small target detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 877-886.

[29] REDMON J, FARHADI A . YOLOv3: an incremental Improvement[EB/OL].(2018-04-08)[2022-07-22].https://arxiv.org/pdf/1804.02767.

[30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2999-3007.

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

[32] 张瑞鑫, 黎宁, 张夏夏, 等. 基于优化 CenterNet 的低空无人机检测方法[J]. 北京航空航天大学学报, 2022, 48(11): 2335-2344.
ZHANG R X, LI N, ZHANG X X, et al. Low-altitude UAV detection method based on optimized CenterNet[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2335-2344(in Chinese).

[33] 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: 2261-2269.

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

[35] WANG J , XU C , YANG W , et al. A normalized Gaussian wasserstein distance for tiny object detection[EB/OL]. (2022-06-22)[2022-07-22]. https://arxiv.org/pdf/2110.13389.

[36] CIPOLLA R, GAL Y, KENDALL A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7482-7491.

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

[38] WANG C Y, YEH I H, MARK LIAO H Y. Yolov9: learning what you want to learn using programmable gradient information[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2024: 1-21.

[39] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Berlin: Springer, 2016: 21-37.

[40] ZHAO Y, LV W, XU S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 16965-16974.

Detecting small ship targets via infrared remote sensing

XIE Youchen，XU Tianyang*，TANG Zhangyong，WU Xiaojun

(School of Artificial Intelligence and Computer Science，Jiangnan University，Wuxi 214122，China)

Abstract: A detection technique called YOLO for weak infrared remote sensing ship target (YOLO-WIT) is presented to overcome the difficulties in detecting weak and small ship targets using infrared remote sensing. Firstly, due to the difficulty of low tolerance against positional offset for dim and small targets, the detection head is optimized to reduce model size while limiting the maximum offset of the anchor. Additionally, a composite distance similarity measure is designed to decrease location sensitivity in the regression branch, thereby enhancing regression accuracy. Secondly, in view of the situation where the infrared image background is bright and the target is dark, a dense concat information expansion convolution block (DECO) is designed to retain weak signals and enhance the perception of weak features. A spatio-temporal attention mechanism is employed for feature enhancement, and the Sobel operator is utilized to solve the first-order derivative of the shallow feature maps in order to direct the model to make judgments with edge features in order to differentiate interference objects with similar shapes and grayscale amplitudes. The experimental results on the NUDT-SIRST-Sea dataset demonstrate that: YOLO-WIT reduces parameters by 31.6% compared to the baseline model, increases mAP_{50} by 9%, and raises $mAP_{50\text{-}95}$ by 4.9%. In comparison to mainstream detection algorithms, YOLO-WIT demands fewer resources with a model size of just 9.2×10^6. Its detection performance on dim and small ship targets is notably superior to other methods.

Keywords: dim and small target detection; infrared remote sensing; YOLOv5; sea surface ship detection; deep learning

Received：2023-12-25；**Accepted**：2024-05-03；**Published Online**：2024-06-04 14：15

URL：link.cnki.net/urlid/11.2625.V.20240604.1045.001

Foundation items：National Natural Science Foundation of China (62106089,62020106012,62332008)

* **Corresponding author**. E-mail：tianyang.xu@jiangnan.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0838

基于近似计算技术的逻辑电路HDL级面积优化

于宗源，王伦耀*，储著飞，夏银水
(宁波大学 信息科学与工程学院，宁波 315211)

摘 要：针对基于近似计算技术优化算法在硬件描述语言（HDL）级电路面积优化中不能有效利用结果质量（QoR）约束所提供的优化空间问题，提出了一种包括内部信号位宽优化操作、算术运算符替换操作、近似算术电路调用及近似操作选择机制等算术电路近似优化方法，在QoR约束下实现了电路的面积优化。所提方法使用C语言编程，电路面积用EDA工具Design Compiler结合工艺库进行估算。实验结果显示：与精确优化结果相比，电路面积平均优化55.2%；与现有近似优化方法相比，平均面积优化可进一步获得24.9%提升。

关 键 词：近似计算；面积优化；HDL级优化；Verilog硬件描述语言；逻辑综合
中图分类号：TP391.41
文献标志码：A **文章编号**：1001-5965(2026)03-0908-09

在传统逻辑优化中，电路功能在逻辑优化前后等价是首要满足的条件。但在某些领域，如音频信号处理、模式识别、数据挖掘、神经网络优化等，由于待处理对象具有容错特性，使得一定程度上的计算错误并不影响实际的应用。现实应用中存在的容错特性促使人们重新对电路逻辑优化前后的“等价”含义进行思考，进而提出了在逻辑优化中使用近似计算[1]技术的设想。基于近似计算的近似逻辑综合(approximate logic synthesis, ALS)其基本思想是通过人为引入错误输出，在满足结果质量(quality-of-result, QoR)的基础上，通过降低电路计算精度，实现电路结构简化，从而获得面积、功耗及延迟等性能的提升[2]。

近似逻辑综合可以在电路功能描述的不同层级，使用不同方法实现。如逻辑网表修改、布尔重写及基于硬件描述语言(hardware description language, HDL)的逻辑设计过程中引入近似计算技术等途径实现[3]。其中，在HDL层级进行近似逻辑综合又称为近似高层级逻辑综合(approximate high-level synthesis, AHLS)。文献[4]把电路功能描述转换为有向无环图，并且根据2个指标即电路节点的重要性和活跃度进行节点修剪。文献[5]通过计算构成电路各节点的重要性及分析删除某节点所导致的误差，决定是否对部分节点删减，实现误差约束下的面积优化。文献[6]通过获取电路关键路径的切割集，用常量节点来替换的方法实现电路性能优化。文献[7]将近似逻辑综合技术与深度学习相结合，提出的算法可以在映射阶段更改或删除门来进一步优化电路的功耗。文献[8]使用了布尔矩阵分解的方法重写逻辑，把表示电路输出真值表的矩阵分解为2个子矩阵，通过综合得到新的压缩后的电路，压缩后的电路输出比原始电路少，从而达到了面积和功耗的降低。文献[9]利用雷德密勒(Reed-Muller, RM)逻辑“异或”运算的特点实现了固定极性下RM逻辑电路的面积优化，拓展了近似优化技术在RM函数近似优化过程中的应用。上述基于逻辑网表修改或布尔重写的近似优化的优势在于可以对电路节点逐个或多个进行替换，或对逻辑表达式的

收稿日期：2023-12-29；**录用日期**：2024-02-25；**网络出版时间**：2024-03-11 10:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20240308.1611.001
基金项目：国家自然科学基金(U23A20351,U22A2013)；宁波市重点研发计划(2023Z233)
* **通信作者**. E-mail：wanglunyao@nbu.edu.cn

引用格式：于宗源，王伦耀，储著飞，等. 基于近似计算技术的逻辑电路HDL级面积优化[J]. 北京航空航天大学学报，2026，52(3)：908-916. YU Z Y，WANG L Y，CHU Z F，et al. Area optimization of logic circuits in HDL-level using approximate computing[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：908-916 (in Chinese).

乘积项进行修改, 使得对电路近似设计的控制更加精细, 不过这些方法可能会受限于电路的结构, 从而无法获得好的优化结果[10]。高层级逻辑优化是把行为级的语言解析为语法树然后进行优化。因此,同属于高层级逻辑综合的 AHLS, 其某个优化步骤往往能达到网表转换和布尔重写方法中多步骤才能达到的优化效果。另外, 因 AHLS 是从 HDL 层面进行近似优化, 使得它可以与商业 EDA 工具实现充分结合, 将近似优化技术融入到现有的集成电路设计流程中, 具有明显的实用价值。

本文主要讨论用 AHLS 实现算术逻辑电路的面积优化。在算术电路中, 电路的面积与待运算信号的宽度成正相关, 因此, 在满足 QoR 约束下, 对信号的位宽进行缩减有利于电路面积的优化。同时考虑到逻辑运算电路比算术电路简单, 因此, 在近似电路实现中, 用逻辑运算来代替算术运算也可以实现电路面积优化[11-12]; 此外, 在逻辑综合过程中, 用面积更小的近似单元电路来替换原来的正确电路, 也可以实现电路面积的优化。为此, 在对算术逻辑电路用 AHLS 实现面积优化时, 本文引入了内部信号的位宽优化、运算符替换和近似单元库调用等 3 种方法。同时, 为了加快算法的运行效率, 提出了近似操作选择机制。该机制通过评估每次近似操作带来的面积优化效果及其负面影响, 如电路输出的错误偏差的大小、经近似操作后得到的 HDL 代码是否可以综合等情况, 来指导近似操作的选择, 进而实现 QoR 约束情况下的电路面积的优化。

1　QoR 的常用判别指标

基于近似计算的优化其前提必须满足给定的约束条件即 QoR。但电路因其功能不同, 对应的 QoR 评判标准也不同。

在算术电路中常用的 QoR 评判指标有误差距离, 以及与误差距离紧密相关的平均误差距离、相对误差距离、平均相对误差距离和均方误差等指标[13]。

若 R_c 和 R_e 分别表示正确值和近似值, 则错误距离 d_{ED} 可以表示为

$$d_{ED} = |R_c - R_e| \tag{1}$$

平均误差距离 m_{ED} 可以表示为

$$m_{ED} = \frac{1}{\alpha}\sum_{i=0}^{\alpha-1} d_{EDi} \tag{2}$$

相对误差距离 r_{ED} 可以表示为

$$r_{ED} = d_{ED}/R_c \tag{3}$$

平均相对误差距离 m_{RED} 为 α 个 r_{ED} 的平均值:

$$m_{RED} = \frac{1}{\alpha}\sum_{i=0}^{\alpha-1} r_{EDi} \tag{4}$$

均方误差 m_{SE} 为 α 个 r_{ED} 的平方的平均值:

$$m_{SE} = \frac{1}{\alpha}\sum_{i=0}^{\alpha-1} (r_{EDi})^2 \tag{5}$$

2　HDL 级近似优化操作

2.1　内部信号位宽优化操作

Verilog 是常用的数字电路逻辑功能硬件描述语言, 其基本结构如图 1 所示。图 1 的 Verilog 描述中包含了 4 个部分, 即电路的输入/输出端口声明、信号类型声明、电路功能描述及包括与信号传输、建立、保持等相关的时间定义。

由于电路输入/输出端口直接与其他电路相连, 输入/输出端口的位宽变化会对整个系统带来不利影响, 因此, 本文信号位宽优化操作仅针对内部信号。通过重新定义图 1 中某些内部信号的位宽, 或令内部信号的某些位为常数, 则可以实现内部信号线的减少, 达到减少电路面积的目的。

信号位宽优化分为高位优化和低位优化 2 种。其中, 高位优化通过删除高位信号实现, 而低位优化则是通过将低位信号置为常数实现。

对于一个位宽为 W 的内部信号 v, 本文用 $P_B(v,w,r)$ 来表示对信号 v 的位宽优化操作。其中, w 表示待优化的位宽, $w \leqslant W$。$P_B(v,w,r)$ 的功能可表示为

$$P_B(v,w,r)=\begin{cases}\{v[W-1:w],N_c\} & r=N_c\\ v[W-w-1:0] & r=i_{DC}\end{cases} \tag{6}$$

式中: N_c 为一个位宽为 w 的二进制常数; 符号 "{}" 为信号位的拼接, 即 $\{v[W-1:w],N_c\}$ 表示新生成的信号的第 w 到 W−1 位来自原始信号 v, 而低 w 位为

```
module module_name (port_list);
  // 端口声明;
input [31:0] a, b;
input c;
output [31:0] o;
output o2;
...
  // 信号类型声明;
...
reg [31:0] o1;
wire signed[31:0] m1, m2;
wire signed[15:0] n1, n2;
...
  // 电路功能描述;
    ...
    m=m1+m2;
    n=n1×n2;
    ...
  // 时间定义;
    ...
endmodule
```

图 1　电路功能 Verilog 建模的基本结构

Fig. 1　Basic structure of circuit function Verilog modeling

常数N_c；i_{DC}为无关项。从式(6)可得，通过赋予r不同值，可以实现高位删除或低位的常数化。

以图1中信号m_1为例，$P_B(m_1,3,3'b001)$，表示将信号m_1的低3位信号被置为二进制常数001，即$m_1[2:0]=3'b001$。$P_B(m_1,0,3)=3'b001$操作等效如下Verilog代码描述：

$$P_B(m_1,0,3'b001) \Leftrightarrow \quad \text{assign} \quad m_1[2:0]=3'b001$$

而$P_B(m_1,2,i_{DC})$则表示m_1高2位被省略。其操作可以等效为如下的Verilog代码描述：

$$P_B(m_1,2,i_{DC}) \Leftrightarrow \quad \text{wire signed}[30:0]m_1$$

在运算电路中，因低位错误而引入的相对误差往往比较小，因此，将低位输出置为常数也是一些近似运算电路在电路级设计中常用的方法[6]。高位删除的方法在Verilog代码描述上比较容易实现，但是由于删除的是内部信号的高位，引入的误差会比较大，因此，高位删除适合那些在多数情况下高位输出为0的运算。这样即使因高位删除导致误差比较大，但因误差引入的次数不多，对误差的平均值影响会比较小。

2.2 运算符替换

在数字电路中，逻辑运算比算术运算更容易实现，且电路面积代价更小。因此，在满足QoR约束下，用逻辑运算代替算术运算是AHLS中实现面积优化的有效途径。

对于加、减或乘运算的近似替换操作，本文用$P_R(e,i,g)$表示。其中，e表示一个只包含由2个位宽为W的变量构成的加、减或乘运算表达式，i表示运算替换的终止位，$i \leqslant W$，g为逻辑门，$g \in \{\text{AND},\text{OR},\text{XOR}\}$。$P_R(e,i,g)$的功能可以描述为

$$P_R(e,i,g)=\{e[W-1:i],b_t(e[i-1:0],g)\} \tag{7}$$

式中：$b_t(e[i-1:0],g)$表示e从0到$i-1$位的运算用逻辑位操作g来替代；符号“{}”的作用同式(6)。从式(7)可得，$P_R(e,i,g)$的作用就是将原来的加、减和乘运算用2部分来替代，高位部分保持原来的运算不变，而低位部分则用逻辑位操作来代替。

以图1中32位加法运算$m=m_1+m_2$为例，操作$P_R(m,4,\text{OR})$表示m_1和m_2的低4位求和运算用逻辑位“或”替换。$P_R(m,4,\text{OR})$操作可等效为如下的Verilog代码描述：

$$P_R(m,4,\text{OR}) \Leftrightarrow \begin{aligned} M_1 &= m_1[31:4]+m_2[31:4] \\ M_2 &= m_1[3:0]|m_2[3:0] \\ m &= \{M_1,M_2\} \end{aligned}$$

式(7)中，当$i=W$时表示用逻辑位运算完全代替算术运算，但这种替代带来误差往往会超过QoR约束，尤其是用逻辑位“与”替换乘运算，会引入很大误差，导致无法满足QoR约束。为实现满足QoR约束下的面积优化，一个折中的解决办法就是逻辑综合过程中，引入近似运算电路单元库[14]，用结构相对简单的近似运算电路来代替原设计中的精确运算电路，实现QoR约束下的面积优化。

引入的近似运算电路单元库中包含多种不同精度的近似运算电路可供选择。近似电路调用过程主要包含2个步骤：①判定被替换的子电路与近似电路单元的输入/输出位宽是否一致或接近。如被替换的子电路是8位乘法器，则首先寻找近似电路库中是否有8位近似乘法器，如果没有找到就向临近的较少位数寻找，假设找到了一个6位乘法器，就将这个子电路的高的6位用6位近似乘法器替换，较低的2位将会直接置为常数0。②近似运算电路单元精度的选择。一般而言，精度越高，电路越复杂，面积也越大。在精度选择方面，优化算法值初始阶段是随机的，但是随着迭代次数的增加，在近似操作选择机制引导下，优化算法会选择满足QoR约束且面积较小的电路单元。

本文用$P_C(e,c)$实现近似运算电路单元调用。其中，e表示一个只包含由2个位宽为W的变量构成的加、减或乘运算表达式，c为被调用的近似运算电路单元。

以图1中表达式$n=n_1\times n_2$为例，假设该乘法运算通过调用近似运算电路单元库中16位近似乘法器实现，且该电路名称为Mult_ap16×16，则$P_C(n,\text{Mult_ap16}\times 16)$操作可等效为如下的Verilog代码描述：

$$P_C(n,\text{Mult_ap16}\times 16) \Leftrightarrow \quad \text{Mult_ap16}\times 16 \quad U(n,n_1,n_2)$$

2.3 近似操作选择策略

在算术逻辑电路的Verilog描述中，内部信号和运算符数量都比较大，这使得每次面积优化都是由多个近似操作构成的一个操作组合来实现。考虑到每种近似操作均有操作符和参数2部分组成，且参数的变化会引起操作结果的变化。因此，可用于面积优化的近似操作组合数量极为庞大。若采用穷举方法，会消耗大量的时间，须引入相应的近似操作选择策略，提高算法效率。

待优化算术电路的Verilog描述经一个近似操作组合作用后，得到的Verilog文件在编译和综合过程中会出现如下4种情况：

case1：有语法错误，无法编译；

case2：可以编译，但代码无法综合生成实际电路；

case3：面积优化与QoR控制不理想；

case4：符合QoR约束，并获得面积优化。

上面 4 种情况中, 导致 case1 和 case2 的近似操作组合需要尽量避免, 而导致 case4 的近似操作组合则需要保留。为尽快搜索到符合 case4 的近似操作组合, 本文通过分类赋分的方法对不同近似操作及其对应的参数进行赋分并累计。在不断迭代中, 通过对近似操作和参数对应的积分情况分析, 选取符合 QoR 约束的近似操作组合, 实现电路面积优化。其基本原理如下。

假设$\mathcal{F}$为待近似优化逻辑电路的 Verilog 描述, κ为$\mathcal{F}$中可以进行近似操作的对象, 如内部变量, 算术运算表达式等。用$\boldsymbol{O}_i \times \boldsymbol{P}_j$表示一个近似操作组合, $\boldsymbol{O}_i$表示第 i 个近似操作符的组合, $\boldsymbol{P}_j$表示与$\boldsymbol{O}_i$对应的第 j 种参数的集合。假设近似操作组合$\boldsymbol{O}_i \times \boldsymbol{P}_j$因包含了某个近似操作$o(p)$, $o \in \boldsymbol{O}_i$, $p \in \boldsymbol{P}_j$, 使得近似操作后得到的 Verilog 文件无法编译, 即对应 case1 情况。由于 case1 为无效操作, 但占用了优化时间, 因此, 要提高效率就应该尽可能避免使用$o(p)$。由于每个近似操作组合都由多个不同近似操作组成, 无法通过一次的运算结果来确定到底是哪个或哪些近似操作导致出现 case1。而通过分类赋分并结合多次迭代分析可以将导致 case1 的近似操作给区分出来。现规定对导致 case1 的近似操作赋分最低, case4 赋分最高, 其他情况次之。当近似操作组合$\boldsymbol{O}_i \times \boldsymbol{P}_j$的结果为 case1 时, 所有构成$\boldsymbol{O}_i \times \boldsymbol{P}_j$的近似操作都赋予相同的最低分, 并对每个近似操作得到的赋分累计。经过多次迭代以后, 那些导致 case1 的近似操作的积分会比较低; 反之, 那些导致 case4 的近似操作的积分会比较高。最后, 利用积分值作为主要参数引导算法搜索到满足 QoR 约束并能实现面积优化的近似操作组合。

式 (8) 为某个近似操作$o(p)$在不同 case 下的赋分情况。

$$m_{\mathrm{K}}(o(p)) = \begin{cases} c_1 & \text{case1} \\ c_2 & \text{case2} \\ c_3 & \text{case3} \\ c_4 \times I_{\mathrm{P}} & \text{case4} \end{cases} \tag{8}$$

式中: I_{P}为面积优化系数, 可以表示为$I_{\mathrm{P}} = s_{i-1}/s_i$, 其中, s_{i-1}和s_i分别表示第$(i-1)$次和第 i 次近似优化后电路的面积, I_{P}越大, 表示面积优化提升效果就越好, 其对应的赋分值也越高。由于 case1 和 case2 中的情况需要尽量避免, 为减小此类近似操作的发生, 令 c_1 和 c_2 的取值小于 0。m_{K}的初始值为 0。

$\mathcal{F}$中的κ被选中进行近似操作的概率$p_{\mathrm{bt}}(\kappa)$表示为

$$p_{\mathrm{bt}}(\kappa) = \begin{cases} \alpha^{-\sum m_{\mathrm{K}}(o(p))} & \sum m_{\mathrm{K}}(o(p)) < 0 \\ 1 & \sum m_{\mathrm{K}}(o(p)) \geqslant 0 \end{cases} \tag{9}$$

式中: α为收敛系数, $\alpha < 1$。由式 (9) 可得, $\sum m_{\mathrm{K}}$如果大于 0, 则对κ的近似操作进行的概率为 100%。$p_{\mathrm{bt}}(\kappa)$的初始值为 1。

对于每个κ, 一般有多种近似操作可选择。如假设κ为加法运算, 则κ的近似操作包括逻辑位“或”替换, 不同精度的近似加法器电路替换等多种选择。假设对κ存在 L 种可能的近似操作, 则选择第 l 种近似操作的概率$R_l(\kappa)$可以表示为

$$R_l(\kappa) = \frac{m_{\mathrm{K}l}}{\sum m_{\mathrm{K}}} \tag{10}$$

式中: $m_{\mathrm{K}l}$为对κ的第 l 种近似操作的m_{K}值; $\sum m_{\mathrm{K}}$为关于κ的所有近似操作的m_{K}值总和。利用式 (10) 采用轮盘赌的方式决定对κ采用何种近似操作。

3　电路面积 HDL 级近似优化

在对电路的 Verilog 描述进行近似优化前, 先要确认 Verilog 描述中, 哪些代码片段可进行近似优化操作。假设$\mathcal{F}$为待优化的 Verilog 的代码, $s[i]$ 为$\mathcal{F}$的某一代码片段, S_{et}为 $s[i]$ 的集合, 则$\mathcal{F}$中可近似优化的代码片段的搜索用算法 1 实现。

算法 1　$\mathcal{F}$中可近似优化的代码片段的搜索。

输入　待优化的 Verilog 的代码$\mathcal{F}$。

输出　待位宽优化对象S_{et}; 待算术运算优化对象O_{set}。

```
1: WHILE(s[i]! =′ endmodule′)
2: IF( (F_C(s[i])! = input &&
        F_C(s[i])! = output) &&
        (F_C(s[i]) == wire ||
        F_C(s[i]) == reg  ||
        F_C(s[i]) == wire singed ||
        F_C(s[i]) == reg singed))
3: store  s[i] in  S_et;
4: ELSE IF (F_C(s[i]) ==′ +′  ||
   F_C(s[i]) ==′ −′  ||
   F_C(s[i]) ==′ ∗′   )
5: store  s[i] in  O_set;
```

算法 1 中, $F_{\mathrm{C}}(s[i])$表示对代码片段 $s[i]$ 的描述进行检查。考虑到输入、输出信号不进行位宽优化, 因此, 在步骤 2 中将$F_{\mathrm{C}}(s[i])$为“input”和“output”的信号排除在外, 而将 $F_{\mathrm{C}}(s[i])$为“wire”、“wire signed”、“reg”和“reg signed”等 4 种类型的内部变量作为可近似优化的对象, 并存储在S_{et}中。同理, 因只对加、减和乘 3 种算术运算进行优化, 因此, 步骤 4 中仅将包含上述 3 种运算的代码片段存储在O_{set}中, 用于算术运算近似优化的对象。由于“endmodule”用来表示 Verilog 代码的结束, 因此,

步骤1用来判断是否遍历完整个Verilog文件。

3.1 电路内部信号位宽优化算法

由位宽优化操作$P_{\mathrm{B}}(v,w,r)$定义可知，内部信号v的高位或低位被优化的信号位数量w的取值只要满足$w\leqslant W$就可以了。但考虑到舍弃的信号位越多，带来的误差将越大，因此，在实际近似优化中，本文将w取值限于$w\leqslant 1$。由于$w=0$表示不执行位宽优化操作，因此，实行的$P_{\mathrm{B}}(v,w,r)$操作实际上只有3种情况，即

1）省略高位，即$P_{\mathrm{B}}(v,1,i_{\mathrm{DC}})$；

2）最低位值置“1”，即$P_{\mathrm{B}}(v,1,1'b1)$；

3）最低位值置“0”，即$P_{\mathrm{B}}(v,1,1'b0)$；

而最终采用哪种情况可以由式(10)来确定。位宽优化的算法伪代码见算法2。算法2中$S_{\mathrm{et_M}}$用于存储S_{et}中各个内部信号的m_{K}值。一般每个信号有3个m_{K}值，分别对应上面3种情况。输出为$S_{\mathrm{et_ap}}$。

算法2 位宽优化算法。

输入 待位宽优化对象S_{et}，待位宽优化对象的m_{K}值$S_{\mathrm{et_M}}$。

输出 位宽优化后的对象$S_{\mathrm{et_ap}}$。

1: copy S_{et} to $S_{\mathrm{et_ap}}$

2: FOR($i=0,i<N(S_{\mathrm{et_ap}}),i=i+1$)

3: {/*$s[i]\in S_{\mathrm{et_ap}}$ */

4: $Y=$do_ap($S_{\mathrm{et_M}},s[i]$);

5: IF(Y) {

6: $r=$comp($S_{\mathrm{et_M}},s[i]$);

7: $P_{\mathrm{B}}(s[i],1,r)$;

8: replace $s[i]$ in $S_{\mathrm{et_ap}}$ with a new one.};

9: ELSE

10: keep $s[i]$ in $S_{\mathrm{et_ap}}$ unchanged};

算法2中，$N(S_{\mathrm{et_ap}})$表示集合$S_{\mathrm{et_ap}}$包含的元素的个数。子程序do_ap($S_{\mathrm{et_M}}$,$s[i]$)利用$S_{\mathrm{et_M}}$寄存的m_{K}值和式(9)确定是否对$s[i]$对应的内部信号进行位宽优化操作，若需要操作，则返回值为“1”，否则为“0”。一旦确定需要位宽优化操作，则由子程序comp($S_{\mathrm{et_M}}$,i)利用式(10)来计算$P_{\mathrm{B}}(s[i],1,r)$中参数r的值。

3.2 算术运算符替换算法

算术运算符替换近似操作包括$P_{\mathrm{R}}(e,i,g)$和$P_{\mathrm{C}}(e,c)$。其中，$P_{\mathrm{R}}(e,i,g)$只对加、减2个运算进行近似操作；为加快优化算法速度，将运算替换的终止位i的值限定为W和$\lfloor W/2\rfloor$二者之一。其中，W为参与运算的变量的位宽。当$i=W$表示加或减运算完全用逻辑位运算替代，而$i=\lfloor W/2\rfloor$则表示加或减运算的低位部分用逻辑位运算代替，而高位部分运算保持不变。$P_{\mathrm{C}}(e,c)$表示表达式e中的加、减或乘运算是通过调用近似算术单元电路来实现的。算术运算符替换的算法伪代码见算法3。算法3中$O_{\mathrm{set_M}}$用于存储O_{set}中各个内部信号的m_{K}值。输出为$O_{\mathrm{set_ap}}$。

算法3 算术运算符替换的算法。

输入 待算术优化对象O_{set}，待算术优化对象的m_{K}值$O_{\mathrm{set_M}}$。

输出 算术优化后的对象$O_{\mathrm{set_ap}}$。

1: copy O_{set} to $O_{\mathrm{set_ap}}$

2: FOR($i=0,i<N(O_{\mathrm{set_ap}}),i=i+1$)

3: {/*$s[i]\in O_{\mathrm{set_ap}}$ */

4: $Y=$do_ap($O_{\mathrm{set_M}},s[i]$);

5: IF(Y) {

6: IF($s[i]$ =='addition' ‖ $s[i]$ =='subtraction')

7: {$s_{\mathrm{t}}=$operator_sel($O_{\mathrm{set_M}},s[i]$);

8: CASE(s_{t}){

$1: P_{\mathrm{R}}(s[i],W,g)$;

$2: P_{\mathrm{R}}(s[i],\left\lfloor\frac{W}{2}\right\rfloor,g)$;

$3: P_{\mathrm{C}}(s[i],c_j)$;

9: replace $s[i]$ in $O_{\mathrm{set_ap}}$ with a new one.}

10: ELSE

11: {$c_k=$Ap_gate($O_{\mathrm{set_M}},s[i]$);

12: $P_{\mathrm{C}}(s[i],c_k)$;

13: replace $s[i]$ in $O_{\mathrm{set_ap}}$ with a new one.}

14: ELSE keep $s[i]$ in $O_{\mathrm{set_ap}}$ unchanged;}

算法3中，$N(O_{\mathrm{set_ap}})$和do_ap($O_{\mathrm{set_M}}$,$s[i]$)意义同算法2所示算法。子程序operator_sel($O_{\mathrm{set_M}}$,$s[i]$)用来确定对加法和减法的近似操作。假设$s[i]$为A+B或A−B，且A、B位宽为W；则当operator_sel($O_{\mathrm{set_M}}$,$s[i]$)返回值为1时，A+B用位“或”$A|B$代替，A−B用位“异或”$A\hat{}B$代替；当返回值为2时，A+B用$\left\{A\left[W-1:\left\lfloor\frac{W}{2}\right\rfloor\right]+B\left[W-1:\left\lfloor\frac{W}{2}\right\rfloor\right],A\left[\left\lfloor\frac{W}{2}\right\rfloor-1:0\right]|B\left[\left\lfloor\frac{W}{2}\right\rfloor-1:0\right]\right\}$来代替，而$A$−$B$用$\left\{A\left[W-1:\left\lfloor\frac{W}{2}\right\rfloor\right]-B\left[W-1:\left\lfloor\frac{W}{2}\right\rfloor\right],A\left[\left\lfloor\frac{W}{2}\right\rfloor-1:0\right]\hat{}B\left[\left\lfloor\frac{W}{2}\right\rfloor-1:0\right]\right\}$来代替。式中符号“{}”作用同式(6)。当返回值为3时，进行近似单元电路替换，用近似运算单元c_j代替“加”或“减”运算。每次调用都会根据优化结果赋予一个m_{K}值，根据多次迭代后m_{K}值的累计值，并结合选择概率，选择不同的近似单元电路替代加、减运算，或者用“或”、“异或”操作来替代全部或部分的加、减运算。

对于乘法运算，在确定需要进行近似替换后，由子程序Ap_gate($O_{\mathrm{set_M}}$,$s[i]$)来确定选用哪个近似乘法单元来替代乘法运算，子程序返回的值就是用

来替换的近似乘法器单元电路的编号 c_k。

表 1 给出了本文调用的近似运算电路库中的部分有符号乘法器的数据对比, mul8s_1KV8 为一个精确乘法器。在近似电路库中表征错误参数有平均绝对误差 m_{AE}、最差绝对误差 w_{CE} 和 m_{RED}。可见在各项错误参数更高的情况下, 乘法器的功耗和面积获得了进一步的优化, 如在多个近似乘法模块都满足电路 QoR 的前提下, 算法就会趋向于采用功耗、面积更小的乘法模块进行替换。

表 1　近似电路库中部分乘法器

Table 1　Multipliers in library of approximate circuits

电路名称	m_{AE}/%	w_{CE}/%	m_{RED}/%	功耗	面积
mul8s_1KV8	0	0	0	0.425	729.8
mul8s_1L1G	0.52	2.66	27.44	0.126	284.9
mul8s_1KR3	3.08	12.30	135.77	0.052	172.2

3.3　电路面积近似优化算法

设 $\mathcal{F}$ 为待优化的算术电路的 Verilog 描述, $\mathcal{F}_{ap}$ 为近似优化都得到的 Verilog 描述。则本文基于近似计算技术的 HDL 级电路面积近化算法的伪代码如算法 4 所示。S_{et_M} 与 O_{set_M} 的含义如前述。在步骤 1 中, 对 S_{et_M} 与 O_{set_M} 中实现初始化。在步骤 2 中找到所有的可以进行位宽优化和算术运算符替换的 Verilog 描述片段, 并分别记为 S_{et_org} 和 O_{set_org}。在步骤 4 和步骤 5 中完成在 S_{et_org} 中随机挑选 m 个随机的内部变量进行位宽优化算法。在步骤 6 和步骤 7 完成在 O_{set_org} 中随机挑选 n 个算术运算表达式, 进行算术运算符替换。

在步骤 8 和步骤 9 中, 用近似操作后得到的内部变量和算术表达式替换原来的内部变量和算术表达式, 并寄存在 S_{et_ap} 与 O_{set_ap} 中。并在步骤 10 中, 用 S_{et_ap} 与 O_{set_ap} 寄存的变量和算术表达式形式替代原 Verilog 文件中对应的描述, 得到近似操作后的 Verilog 描述 $\mathcal{F}'$。在步骤 11 中, 对 $\mathcal{F}'$ 进行综合和结果的分析, 给出此次近似操作的赋分, 并对积分进行更新。步骤 12 为停止循环判断。在本文中停止循环的条件为: 迭代 30 次或连续 3 次迭代优化改进小于 1%。

步骤 13 中找到优化效果最好的 Verilog 描述作为近似优化的最终结果。

算法 4　HDL 级面积近似优化化算法。

输入　待优化的算术电路的 Verilog 描述 $\mathcal{F}$。

输出　近似优化都得到的 Verilog 描述 $\mathcal{F}_{ap}$。

1: initialize S_{et_M} and O_{set_M};

2: search_Sections_for_ap($\mathcal{F}$), get S_{et_org} and O_{set_org};

3: DO{

4: select m elements in S_{et_org} randomly and form S_{et};

5: exclute Bit_op(S_{et}, S_{et_M}, S_{et_ap});

6: select n elements in O_{set_org} randomly and form O_{set};

7: exclute Op_replace(O_{set}, O_{set_M}, O_{set_ap});

8: replace the variables in $\mathcal{F}$ with the counterparts in S_{et_ap};

9: replace the arithmatic expressions in $\mathcal{F}$ with the counterparts in O_{set_ap};

10: get the approximate verilog fileafter the above replacements, named $\mathcal{F}'$;

11: synthesize $\mathcal{F}'$, analyze the results and update S_{et_M} and O_{set_M};}

12: WHILE(the stop condition unmatched);

13: select the best one as $\mathcal{F}_{ap}$;

4　测试电路

参与近似优化测试的为如下 10 个电路, 这些电路的逻辑功能均用 Verilog HDL 描述。

1) 支持向量机(support vector machine, SVM)。SVM 是机器学习中一个常用的应用。感知器接受一个输入数据, 用向量 $\boldsymbol{X}$ 表示, 并预测 $\boldsymbol{X}$ 的类别 (例如, −1 或+1), 感知器本质上定义了一个超平面, 将训练数据分成 2 个类。本文的感知器测试数据使用二次函数来映射输入空间。输入数据集由 1 000 个随机生成的二维点组成。分类结果与真实值进行比较, 因此, 衡量 SVM 的 QoR 为与原始 SVM 输出不同的百分比。

2) 有限长单位冲激响应(finite impulse response, FIR)滤波器。采用 25 分频 FIR 滤波器接收信号, 并与低通滤波器系数阵列进行卷积, 计算滤波输出。该设计近似优化后的质量评估通过信号振幅上的 m_{SE} 来描述。使用函数生成随机信号作为输入, 分别得到近似电路和原始电路的输出, QoR 为信号振幅的均方误差。

3) FFT 电路。对于 FFT 电路, 参与测试的为一个 8 点 FFT。输入随机生成, 最终通过计算近似电路与原电路的输出振幅谱的 m_{SE} 作为近似电路的 QoR。

4) Multi_4 为 4 位乘法器。

5) Multi_8_3 为 8×3 乘法器即 2 个输入分别为 8 位和 3 位。

6) Multi_16 为 16 位乘法器。

上述乘法器的 QoR 为平均相对误差距离 m_{RED}。

7) 加法器电路 Add_8。Add_8 是一个 8 位全加

器，输入为2个8位信号，输出为9位信号，QoR为平均相对误差距离 m_{RED}。

8）方差（Variance）电路。输入为随机10组8位二进制数据，先计算10组数据的平均值，再计算其方差。以 m_{RED} 作为QoR评判指标。

9）无限冲击响应滤波器（infinite impulse response, IIR）。参加测试的IIR滤波器为巴特沃斯型低通滤波器，截止频率为1 000 Hz，通带衰减1 dB，阻带衰减15 dB。该设计的近似优化后的QoR是通过计算原始电路和近似电路输出的 m_{RED} 来评估。

10）阶乘（Factorial）电路。阶乘电路的输入是4位二进制数，本文测试了4位二进制数的所有输入组合，用近似电路与原始电路输出的 m_{RED} 作为近似电路的QoR。

5 实验结果与分析

本文方法由C语言编程实现，并结合近似高层级逻辑综合工具abacus，得到近似电路的Verilog描述。电路的面积由synopsys公司Design Compiler结合45 nm工艺库得到。电路输出在近似优化前后的差异由开源模拟器verilator[15]仿真获得。式(8)和式(9)中的 α、c_1、c_2、c_3、c_4 在本节分别被设置为0.95、−8、−5、−4、25。

在QoR计算中，测试输入的使用数量将直接影响QoR计算的准确性。一般而言，测试输入越多，QoR计算就越准确，但是算法的效率就越低。以表2中的FIR电路为例，输入数据位宽为27位，理论上输入组合有227种。显然，用所有可能的输入组合来测试电路近似优化前后的输出误差在效率上不可行。为此需要在计算精度与测试输入数量之间做出某种折中，使得在误差可接受的范围内，减少测试输入的数量，进而提高算法效率。

表2给出了10个测试电路经过近似优化后，测试输入数与误差之间的关系。从表2可知，随着测试输入组合数的增加，输出正确率的逐渐趋向在一个稳定值上下小范围波动。当测试输入超过100组数据时，给出的测试电路中大部分电路输出的QoR波动范围不超过1%，当测试数据大于500组数据时，全部测试电路输出QoR波动范围都不超过1%，且输出的波动范围更小。为此，从提高算法速度和保证一定估算精度前提下，本文以500组输入得到的输出结果作为每次近似迭代后，得到的近似电路输出QoR。由于Factorial电路输入组合数最多只有15种，故表2中Factorial电路的值为输入15组的结果。

表3给出了本文方法的实验结果。其中，“精确优化面积”表示没有引入近似计算技术时优化后电路的面积。在QoR阈值设定中，前3个采用文献[16]提供的阈值，后7个电路统一设置为80%。从实验结果可见，在满足约束情况下，待测的10个电路的面积均获得了比较大的优化。相比于原始电路，面积最大优化了87.9%，最小有33.1%的改进。在SVM电路近似优化上，本文方法与文献[16]的效果比较接近，而其余7个电路上，本文方法要远好于文献[16]。其主要原因是SVM电路主要逻辑功能是对输入数据实现分类，而分类功能对计算的精度要求不高，因此，SVM电路中许多算术运算都可以用逻辑位运算来替换，从而实现面积的极大优化。但SVM电路中还包含了控制逻辑部分，这部分电路如引入近似优化，往往会引起很大的输出误差，其优化结果一般不会被保留。对于SVM电路中的算术运算部分，由于本文与文献[16]都采用了类似的优化策略，使得优化效果比较接近。对于电路Multi_8_3是直接由8位乘3位的乘法电路，因为其中一个输入位宽仅为3位，一旦对其进行位宽缩减等优化，其误差会很容易突破QoR约束。因此，经过多次迭代后，本文方法最终采用和文献[16]相同的优化策略，导致二者优化效果一样。

在表3其他电路近似优化方面，文献[16]采用单一的逻辑位运算替代方法。该方法在QoR控制方面不够精细，往往因单次的近似操作引入较大误差而突破总的QoR约束，导致优化空间没有充分利用。相比于文献[16]，本文引入算术运算符替换操作不仅实现用逻辑位运算替代算术运算，而且通过算术运算高位保持精确运算，低位采用逻辑位运算替代等操作实现QoR约束下的面积优化尽可能最大化。在单元库映射阶段，通过调用近似运算电路库等多种方法实现计算精度与面积优化的折中。

表2 测试激励的数量对QoR的影响

Table 2 Impact of the number of testing vectors on QoRs

电路名称	QoR/%				
	输入组合数为100	输入组合数为500	输入组合数为1 000	输入组合数为5 000	波动范围
SVM	74.69	75.35	74.56	74.55	0.80
FIR	94.77	94.35	94.34	94.31	0.42
FFT	91.55	92.25	91.92	91.88	0.70
Multi_4	84.77	84.74	85.19	84.61	0.45
Multi_8_3	95.11	94.98	95.14	95.10	0.16
Multi_16	83.01	81.98	81.95	81.77	1.03
Add_8	94.39	94.34	92.71	93.10	0.63
Variance	81.53	81.77	81.65	80.96	0.81
IIR	87.81	87.97	88.10	87.93	0.16
Factorial	80.79				0

表 3　本文方法与其他方法的比较结果

Table 3　Comparison results of the proposed method and other methods

电路名称	电路的输入/输出数量	精确优化面积	设定的 QoR值/%	本文面积优化率/%	文献[16]面积优化率/%	本文功耗优化率/%	文献[16]功耗优化率/%	本文延迟优化率/%	文献[16]延迟优化率/%
SVM	68/2	43 858	76.3	78.4	72.2	72.0	70.8	0	0
FIR	27/15	47 823	91.5	60.5	37.7	31.9	36.7	0	0
FFT	258/256	21 423	92.0	45.5	20.5	38.0	48.9	0	0
Multi_4	8/8	220	80.0	79.4	33.5*	83.8	34.5*	85.9	20.3*
Multi_8_3	11/11	205	80.0	36.5	36.5*	41.7	41.7*	33.3	33.3*
Multi_16	32/32	2 949	80.0	87.9	34.0*	94.1	40.1*	64.2	19.5*
Add_8	16/9	69	80.0	39.9	12.8*	55.4	14.7*	58.0	14.5*
Variance	80/16	11 369	80.0	34.2	13.2*	51.4	25.9*	4.5	0*
IIR	34/32	49 696	80.0	56.1	23.1*	63.3	5.4*	14.1	0.1*
Factorial	4/32	11 260	80.0	33.1	19.1*	32.2	16.0*	28.2	17.0*

注: “*” 表示该实验结果利用文献[16]提供的方法获得。

由于存在多种近似替代方法, 可以较好控制单次近似优化操作带来的 QoR 的增量, 使得优化方法能充分利用 QoR 约束提供的优化空间, 实现电路的面积优化。这也是表 3 中本文方法要比文献 [16] 的方法优化效果更好的原因。相比文献 [16] 的方法, 本文方法使平均面积优化进一步提升了 24.9%, 并且在平均功耗和平均延迟上, 也分别提高了 22.9% 和 18.3%。本文方法相比于精确优化, 在面积、功耗和延时上分别平均提升了 55.2%、56.4% 和 28.8%。

表 3 中, 相比于精确优化方法, 本文方法在面积、功耗和延时上都有提升。其原因是表 3 中的测试电路均包含了大量的加法、乘法等算术运算。对于以算术运算为主且允许有容错输出的电路适合用近似计算技术实现电路性能的提升。但若待优化电路的输出容错很小, 甚至没有, 如控制逻辑电路, 则很难用近似计算技术对其性能优化。

6　结　论

本文提出一种在 QoR 约束下逻辑电路在 HDL 级的面积优化方法。其中 HDL 级电路的描述采用 Verilog 语言。本文优化方法主要包括: ①搜索并确定可以优化的内部变量和算术运算。在满足 QoR 约束下, 通过对搜索到的内部变量低位部分置常数或删除部分高位等内部变量位宽优化操作实现内部信号线数量的减少, 实现电路面积的减少。②算术运算的近似替代。通过用逻辑运算替代部分算术运算, 以及近似运算电路单元库调用等近似优化操作进一步实现电路面积的减少。③提出了基于优化结果的近似操作分类赋分和基于积分结果的近似操作选择方法。该方法通过减少无效近似优化操作的调用次数, 提高了优化方法的效率。在引入本文近似优化方法后, 在满足 QoR 约束下, 不但电路的面积得到了优化, 延时和功耗也得到了进一步改善。

参考文献（References）

[1] LIU W Q, LOMBARDI F, SHULTE M. A retrospective and prospective view of approximate computing[J]. Proceedings of the IEEE, 2020, 108(3): 394-399.

[2] SCARABOTTOLO I, ANSALONI G, CONSTANTINIDES G A, et al. Approximate logic synthesis: a survey[J]. Proceedings of the IEEE, 2020, 108(12): 2195-2213.

[3] MITTAL S. A survey of techniques for approximate computing[J]. ACM Computing Surveys, 2016, 48(4): 1-33.

[4] SCHLACHTER J, CAMUS V, PALEM K V, et al. Design and applications of approximate circuits by gate-level pruning[J]. IEEE Transactions on Very Large Scale Integration Systems, 2017, 25(5): 1694-1702.

[5] SCARABOTTOLO I, ANSALONI G, POZZI L. Circuit carving: a methodology for the design of approximate hardware[C]// Proceedings of the Design, Automation & Test in Europe Conference & Exhibition. Piscataway: IEEE Press, 2018: 545-550.

[6] CHANDRASEKHARAN A, SOEKEN M, GROßE D, et al. Approximation-aware rewriting of AIGs for error tolerant applications [C]//Proceedings of the 35th International Conference on Computer-Aided Design. New York: ACM, 2016: 1-8.

[7] PASANDI G, PETERSON M, HERRERA M, et al. Deep-PowerX: a deep learning-based framework for low-power approximate logic synthesis[C]//Proceedings of the Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design. New York: ACM, 2020: 73-78.

[8] MA J X, HASHEMI S, REDA S. Approximate logic synthesis using Boolean matrix factorization[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2022, 41(1): 15-28.

[9] 王伦耀, 夏银水, 储著飞. 基于近似计算技术的 FPRM 电路面积优化[J]. 电子学报, 2019, 47(9): 1868-1874.
WANG L Y, XIA Y S, CHU Z F. Area optimization of FPRM circuits using approximate computing[J]. Acta Electronica Sinica,

2019, 47(9): 1868-1874(in Chinese).

[10] NEPAL K, LI Y T, BAHAR R I, et al. ABACUS: a technique for automated behavioral synthesis of approximate computing circuits[C]//Proceedings of the Design, Automation & Test in Europe Conference & Exhibition. Piscataway: IEEE Press, 2014: 1-6.

[11] PRIYADHARSHNI M, KUMARAVEL S. Low power and area efficient error tolerant adder for image processing application[J]. International Journal of Circuit Theory and Applications, 2020, 48(5): 696-708.

[12] PRIYADHARSHNI M, KUMARAVEL S. Design of imprecise multipliers by using approximate technique for error resilient applications[J]. Journal of Circuits, Systems and Computers, 2021, 30(7): 2150114.

[13] JIANG H L, LIU C, LIU L B, et al. A review, classification, and comparative evaluation of approximate arithmetic circuits[J]. ACM Journal on Emerging Technologies in Computing Systems, 2017, 13(4): 1-34.

[14] MRAZEK V, HRBACEK R, VASICEK Z, et al. EvoApprox8b: library of approximate adders and multipliers for circuit design and benchmarking of approximation methods[C]//Proceedings of the Design, Automation & Test in Europe Conference & Exhibition. Piscataway: IEEE Press, 2017: 258-261.

[15] SNYDER W. Verilator, the fastest free Verilog HDL simulator [EB/OL]. (2023-05-05)[2023-12-20]. http://www.veripool.org/wiki/verilator.

[16] NEPAL K, HASHEMI S, TANN H, et al. Automated high-level generation of low-power approximate computing circuits[J]. IEEE Transactions on Emerging Topics in Computing, 2019, 7(1): 18-30.

Area optimization of logic circuits in HDL-level using approximate computing

YU Zongyuan, WANG Lunyao*, CHU Zhufei, XIA Yinshui

(Factual of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: Given that the current area optimization at the hardware description language (HDL) level using approximate computing techniques is unable to effectively utilize the optimization space provided by quality-of-result (QoR) constraints, an approximate optimization algorithm for arithmetic circuits is proposed. This algorithm includes the selection mechanism of the proposed approximate operations, internal signal bit-width reduction, arithmetic operator replacement, and approximate arithmetic cell circuits calling. The proposed algorithm is programmed in C and the circuit area is estimated by Design Compiler. The experimental results show that under the constraint of QoR, compared with the non-approximate optimization result, the average area saving is 55.2%. The suggested approach can further increase the average area save by 24.9% when compared to the reported approximate strategy.

Keywords: approximate computing; area optimization; HDL-level optimization; Verilog HDL; logic synthesis

Received: 2023-12-29; **Accepted**: 2024-02-25; **Published Online**: 2024-03-11 10:07

URL: link.cnki.net/urlid/11.2625.V.20240308.1611.001

Foundation items: National Natural Science Foundation of China (U23A20351,U22A2013); Major Research Plan of Ningbo (2023Z233)

* **Corresponding author**. E-mail: wanglunyao@nbu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0844

SMA驱动多状态变后缘弯度机翼的设计、分析与实验

王子昂[1]，卢志荣[1]，李鸿濠[1]，周文雅[2]，王晓明[1,*]

(1. 广州大学 机械与电气工程学院，广州 510006；2. 大连理工大学 力学与航空航天学院，大连 116024)

摘 要：变弯度作为变形机翼最典型的一种变形形式，在未来智能飞行器设计领域具有广阔应用前景。针对多状态、连续变弯度机翼的设计需求，提出一种结合模块化刚柔耦合结构、交叉簧片式铰链、多级形状记忆合金(SMA)驱动器的新型变弯度机翼设计方案。针对该方案，结合交叉簧片式铰链等效刚度计算方法，推导单级、多级模块单元的等效刚度结构力学模型，并通过有限元分析完成对比验证；搭建准静态流固耦合分析模型，通过SMA完全相变驱动方式验证所提变弯度机翼方案的多状态调节能力、多工况适应性，并给出各变形状态对气动系数的影响规律；研制原理样机并搭建测试平台，完成基于前馈开环控制策略的多状态变形控制效果，验证了所提多状态变弯度机翼设计的工程可行性。

关 键 词：变弯度机翼；交叉簧片；多状态；形状记忆合金；实验验证

中图分类号：V221+.3; TB553

文献标志码：A **文章编号**：1001-5965(2026)03-0917-09

当前航空技术的快速发展促使飞行器设计向着智能化、多用途化方向发展。变形机翼能根据飞行任务和环境主动改变自身形状，提升复杂多变环境下的气动特性和飞行性能，扩大飞行包线，是未来智能飞行器的典型特征[1]。

机翼变弯度是目前研究最为广泛和发展相对成熟的一种变体形式[2]，国内外研究机构对变弯度机翼技术开展了一系列研究。Shi等[3]提出一种基于多旋转肋的变后缘弯度方案，实现了连续平滑变形。Li等[4]通过集中柔性铰链和齿轮连杆机构的配合来实现机翼变弯度。Zhao等[5]提出一种双肋片结构，既避免了机翼变弯度过程的应力集中问题，又提升了承载能力。然而，传统机械结构设计方案存在部件数多、重量大、维护性差等问题，难以满足未来简洁轻量化设计需求。因此，轻质、高效的柔性结构被越来越多地应用于变弯度机翼设计中。Companile等[6]提出一种“肋带”型设计方案，采用分布式柔性结构替代传统机械结构。张盛等[7]针对“基板”方案存在的问题，提出多变形控制点偏心梁方案。Gong等[8]将新型三维零泊松比蜂窝结构应用在变弯度机翼中。王志刚等[9]通过拓扑优化方法设计了柔性机翼后缘。赵立杰等[10]在机翼结构设计中，使用多相材料拓扑优化技术。徐钧恒等[11]提出以交叉簧片式铰链为柔性变形单元的设计方案。

柔性结构为机翼基体设计带来了多种潜在的优势，但大部分依然采用传统的机械驱动方式，如电机、连杆等，存在磨损、润滑不足、结构复杂等问题。近些年，融合智能材料的柔性变形机翼设计方案成为新趋势，其中，形状记忆合金(shape memory alloy, SMA)因其变形量大、可靠性高、易维护等优势受到了越来越多的关注。Braga等[12]使用SMA

收稿日期：2024-01-02；**录用日期**：2024-03-15；**网络出版时间**：2024-04-03 15:59
网络出版地址：link.cnki.net/urlid/11.2625.V.20240403.1040.003
基金项目：国家自然科学基金(12102096)；广东省自然科学基金(2022A1515011885)
*通信作者. E-mail：wangxm@gzhu.edu.cn

引用格式：王子昂，卢志荣，李鸿濠，等. SMA驱动多状态变后缘弯度机翼的设计、分析与实验[J]. 北京航空航天大学学报，2026，52(3)：917-925. WANG Z A，LU Z R，LI H H，et al. Design, analysis, and experimentation of SMA-driven multi-state variable camber wing[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：917-925 (in Chinese).

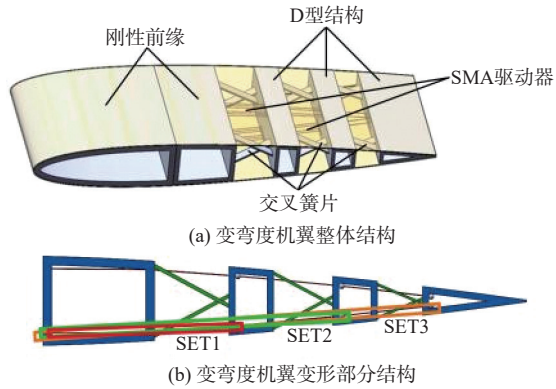
弹簧驱动器实现了机翼变形及最终偏转角度的控制。Lee等[13]将新型SMA驱动器嵌入柔性机翼内，展示了其在变形机翼领域的广阔应用前景。Silva等[14]验证了SMA在减低机翼颤振等方面的有效性。然而，SMA驱动器在加热到相变温度并完全相变后，通常只能达到一种固定状态。SMA材料特性难以准确建模预测，且温度无法精确控制，难以在实际应用中实现精确的不完全相变状态控制。因此，采用单个SMA驱动器通常难以满足不同飞行状态下的多状态机翼变形控制需求。

综上所述，传统机械结构复杂笨重，不利于轻量化，柔性结构是未来的发展方向，但现阶段难以满足大的承载要求；在驱动方式上，智能驱动材料替代传统驱动方式已成为趋势，但由于智能材料自身特性的制约，其在变形精度和能实现的变形形式之间较难以找到平衡点。为了解决上述2大问题，本文提出一种由模块化刚柔耦合结构、交叉簧片式铰链、多级SMA驱动器等构成的新型多状态连续变弯度机翼设计方案。刚柔耦合结构相比于传统机械结构和柔性结构，其结构较为简洁轻量，且承载能力较强。多级SMA驱动器以丝状SMA为基本组成单元，以SMA完全相变后状态为基本特征，通过各级SMA驱动器之间的配合，实现变弯度机翼在较高精度下的多状态变形。

1 变弯度机翼结构设计

以连续变后缘弯度为设计目标，选取NACA0012翼型作为初始翼型，翼型弦长为600 mm，后缘可变形段长度为弦长的50%。

新型多状态连续变弯度机翼设计方案主要由3部分组成，分别为刚性前缘、多级SMA驱动器和模块化刚柔耦合后缘结构，如图1所示。模块化刚柔耦合后缘结构（下文简称后缘结构）的变形精度决定了变弯度机翼的最终变形形态，在本文设计中，后缘结构由3个子模块单元组成，每个子模块单元由交叉簧片式铰链（下文简称铰链）和D型结构构成。SMA驱动器的材料特性导致难以对其相变过程中的状态进行精确控制，从而不易满足机翼对不同弯度调节的需求，为此，本文采用多级SMA驱动方式。与常规单级SMA驱动方式相比，其主要具有以下优势：①利用SMA完全相变特性，提高变形精度，降低控制难度；②通过各级SMA驱动器之间的配合，实现变弯度机翼多种状态的偏转，增强机翼的承载和环境适应性。另外，采用的丝状SMA驱动器具有加热散热快、安装维护方便等独特优势。



(a) 变弯度机翼整体结构

(b) 变弯度机翼变形部分结构

图1 变弯度机翼结构设计

Fig. 1 Variable camber wing structure design

多级SMA驱动器位于后缘结构的上、下翼面，控制机翼向上偏转和向下偏转，后缘结构和多级SMA驱动器共同构成变弯度机翼的变形部分。根据特定飞行控制需求，加热不同侧、不同级的SMA驱动器，可驱动对应位置的铰链弯曲变形，从而实现机翼整体连续、光滑的变形效果。这种拮抗型设计方案还可以通过加热另一侧SMA驱动器实现快速变形恢复[15-16]，降低SMA冷却时间慢导致的滞后控制问题。此外，同时加热两侧SMA驱动器可达到提升结构整体弯曲刚度和气动承载能力的效果，在一定程度上缓解变弯度机翼变形和承载间的矛盾。考虑到篇幅问题，本文将聚焦于利用单侧、不同级SMA驱动器实现变弯度机翼的多状态变形及控制，为了便于区分，将单侧各级SMA驱动器分别命名为SET1、SET2和SET3，如图1(b)所示。通过各级SMA驱动器间的配合，机翼能够实现包括初始状态在内的8种变形状态。

2 变弯度机翼后缘结构力学模型

分析所提新型变弯度机翼的变形能力，关键在于建立SMA相变产生的力/力矩对结构变形的影响规律，因此，需研究后缘结构的刚度特性。由于采用了交叉簧片式铰链，难以直接获得其刚度系数，因此，研究中借鉴文献[17]中铰链等效刚度的计算方法，对后缘结构力学模型进行推导。

2.1 SMA驱动特性模型

根据文献[18]所提的SMA本构模型和参数，在一维模型中，SMA应力、应变和温度的关系式如下：

$$\sigma-\sigma_0=E\varepsilon-E_0\varepsilon_0+\Lambda\xi_S-\Lambda_0\xi_{S0}+\omega(T-T_0) \tag{1}$$

式中：σ为应力；E为SMA弹性模量；ε为应变；Λ为相变张量；ξ_S为应力诱发马氏体体积分数；ω为热膨胀系数；T为温度；σ_0、E_0、ε_0、Λ_0、ξ_{S0}、T_0分别为初始状态的应力、弹性模量、应变、相变张量、孪晶马氏体体积分数及温度。其中，弹性模量E、相变

张量Λ与总马氏体体积分数ξ相关, 关系式如下:

$$E = E_{\mathrm{A}} - (E_{\mathrm{M}} - E_{\mathrm{A}})\xi \tag{2}$$

$$\Lambda = -\varepsilon_{\max}\left[E_{\mathrm{A}} - (E_{\mathrm{M}} - E_{\mathrm{A}})\xi\right] \tag{3}$$

式中: E_{A}为奥氏体弹性模量; E_{M}为马氏体弹性模量; $\xi = \xi_{\mathrm{S}} + \xi_{\mathrm{T}}$, ξ_{T}为温度诱发马氏体体积分数; $\varepsilon_{\max}$为最大残余应变。

为简化后续分析和建模工作, 使用等效驱动力来表示 SMA 相变过程中的力/力矩。为获得 SMA 等效驱动力与温度的关系, 对 SMA 升温阶段的相变过程进行研究, 结果如图 2 所示。

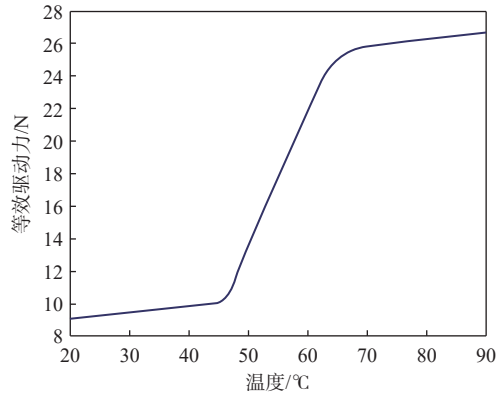
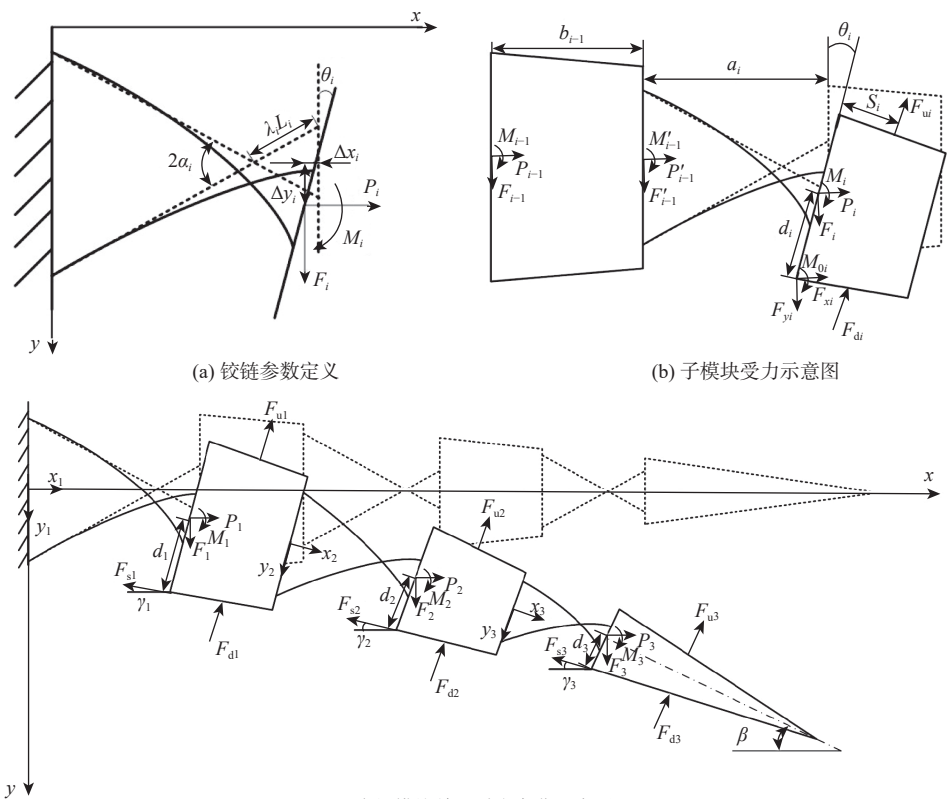


图 2　SMA 升温段温度与力的关系

Fig. 2　The relationship between temperature and force in SMA during heating

2.2　后缘结构力学模型的建立

机翼后缘结构及其子模块单元的受力见图 3。为了得到后缘结构的力学模型, 需要先建立单级模块单元的等效刚度计算方法, 再找出子模块单元间的受力平衡关系, 最后, 得到多级模块单元的等效刚度计算方法。



(a) 铰链参数定义　　(b) 子模块受力示意图

(c) 多级模块单元受力弯曲示意图

图 3　后缘结构及其子模块单元受力示意图

Fig. 3　The force diagram depicting the trailing edge structure and its sub-module units

1) 对于第i个子模块单元, 借鉴文献 [17] 的建模方法, 消去方程的非线性, 得到铰链的等效刚度计算公式为

$$\boldsymbol{\theta}_i^* = \frac{15\cos\boldsymbol{\alpha}_i(\lambda_i \boldsymbol{f}_i \cos\boldsymbol{\alpha}_i + \boldsymbol{m}_i)}{(18\lambda_i^2 - 18\lambda_i + 15\lambda_i\cos\boldsymbol{\alpha}_i^2 + 2)\boldsymbol{p}_i + 120\cos\boldsymbol{\alpha}_i(3\lambda_i^2 - 3\lambda_i + 1)} \tag{4}$$

Δx和Δy为偏转后, 铰链上载荷作用点在x、y轴方向的位移量可表示为

$$\Delta x_i = \lambda_i L_i \cos\boldsymbol{\alpha}_i(1 - \cos\theta_i) \tag{5}$$

$$\Delta y_i = \lambda_i L_i \cos\boldsymbol{\alpha}_i \sin\theta_i \tag{6}$$

式中: $\boldsymbol{m}_i$、$\boldsymbol{f}_i$、$\boldsymbol{p}_i$为无量纲参数, $\boldsymbol{m}_i = \dfrac{\boldsymbol{M}_i L_i}{E' I_i}$, $\boldsymbol{f}_i = \dfrac{\boldsymbol{F}_i L_i^2}{E' I_i}$, $\boldsymbol{p}_i = \dfrac{\boldsymbol{P}_i L_i^2}{E' I_i}$, E'和I_i分别为该子模块单元中铰链单一柔性梁的材料弹性模量和惯性矩; θ_i为该子模

块单元相对于上一个模块单元的偏转角度，且有$\sum_{i=1}^{3}\theta_i=\beta$；$\boldsymbol{M}_i$、$\boldsymbol{F}_i$、$\boldsymbol{P}_i$分别为第$i$个子模块单元中铰链所受的力矩、切向力和法向力；$\alpha_i$、$L_i$、$\lambda_i$均为铰链的几何参数，分别为第$i$个子模块单元中铰链柔性梁夹角的一半、单一柔性梁的长度和柔性梁交叉点到自由端的长度与柔性梁长度的比值，如图3(a)所示。

2）根据静力平衡关系，可以得到第i个子模块单元铰链两侧存在以下关系：

$$\begin{bmatrix}1&0&0\\0&1&0\\(a_i+\Delta x_i)&-\Delta y_i&1\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_i\\\boldsymbol{P}_i\\\boldsymbol{M}_i\end{bmatrix}=\begin{bmatrix}\boldsymbol{F}'_{i-1}\\\boldsymbol{P}'_{i-1}\\\boldsymbol{M}'_{i-1}\end{bmatrix}\tag{7}$$

考虑到气动载荷对第i个子模块单元的作用及D型结构对力矩的影响，可得到作用于第$i-1$个子模块单元铰链上的合力为

$$\begin{bmatrix}1&0&0\\0&1&0\\b_{i-1}\cos\delta_{i-1}+a_i+\Delta x_i&-b_{i-1}\sin\delta_{i-1}-\Delta y_i&1\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{F}_i\\\boldsymbol{P}_i\\\boldsymbol{M}_i\end{bmatrix}+\begin{bmatrix}1&0&0\\0&1&0\\b_{i-1}\cos\delta_{i-1}&-b_{i-1}\sin\delta_{i-1}&1\end{bmatrix}\cdot\begin{bmatrix}-(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\cos\theta_i\\(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\sin\theta_i\\-(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\boldsymbol{S}_i\end{bmatrix}=\begin{bmatrix}\boldsymbol{F}_{i-1}\\\boldsymbol{P}_{i-1}\\\boldsymbol{M}_{i-1}\end{bmatrix}\tag{8}$$

式中：$\boldsymbol{F}'_{i-1}$、$\boldsymbol{P}'_{i-1}$、$\boldsymbol{M}'_{i-1}$为第i个子模块单元对第$i-1$个子模块单元的作用力；a_i表示铰链的长度；$\boldsymbol{F}_{ui}$、$\boldsymbol{F}_{di}$分别为机翼上、下表面受到的气动载荷；S_i为气动载荷与作用点之间的距离；b_{i-1}和δ_{i-1}分别为第$i-1$个子模块单元中D型结构的长度和子模块单元在未受到作用力前的初始偏转角度；$\boldsymbol{F}_{i-1}$、$\boldsymbol{P}_{i-1}$、$\boldsymbol{M}_{i-1}$为传递到第$i-1$个子模块单元中铰链上的作用力。如图3(b)所示。

3）在第i个子模块单元中，将SMA完全相变后的等效驱动力用$\boldsymbol{F}_{si}$表示，为了方便后续计算，将作用力的作用点移至铰链的中心线位置，其对铰链产生的作用力可表示为

$$\begin{bmatrix}1&0&0\\0&1&0\\-d_i\sin\delta_i&-d_i\cos\delta_i&1\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{si}\sin(\gamma_i-\delta_i)\\-\boldsymbol{F}_{si}\cos(\gamma_i-\delta_i)\\0\end{bmatrix}=\begin{bmatrix}\boldsymbol{F}_i\\\boldsymbol{P}_i\\\boldsymbol{M}_i\end{bmatrix}\tag{9}$$

式中：d_i为SMA作用点与中心线位置点的距离；γ_i为初始状态下等效驱动力$\boldsymbol{F}_{si}$与x轴方向的夹角；δ_i为未受等效驱动力作用前的初始偏转角度。

综上所述，后缘结构多级模块单元间力学关系可以表示为

$$\boldsymbol{B}_i\boldsymbol{F}_i+\boldsymbol{C}_i=\boldsymbol{F}_{i-1}\tag{10}$$

$$\begin{cases}\boldsymbol{B}_i=\begin{bmatrix}1&0&0\\0&1&0\\b_{i-1}\cos\delta_{i-1}+a_i+\Delta x_i&-b_{i-1}\sin\delta_{i-1}-\Delta y_i&1\end{bmatrix}\\\boldsymbol{C}_i=\begin{bmatrix}1&0&0\\0&1&0\\b_{i-1}\cos\delta_{i-1}&-b_{i-1}\sin\delta_{i-1}&1\end{bmatrix}\begin{bmatrix}-(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\cos\theta_i\\(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\sin\theta_i\\-(\boldsymbol{F}_{ui}+\boldsymbol{F}_{di})\boldsymbol{S}_i\end{bmatrix}\\\boldsymbol{F}_i=\begin{bmatrix}\boldsymbol{F}_i&\boldsymbol{P}_i&\boldsymbol{M}_i\end{bmatrix}^{\mathrm{T}}\\\boldsymbol{F}_{i-1}=\begin{bmatrix}\boldsymbol{F}_{i-1}&\boldsymbol{P}_{i-1}&\boldsymbol{M}_{i-1}\end{bmatrix}^{\mathrm{T}}\end{cases}\tag{11}$$

联立式(4)和式(10)，即可求解后缘结构多级模块单元等效刚度。

根据上述给出的单级、多级模块单元等效刚度计算公式，只需要知道等效驱动力$\boldsymbol{F}_{si}$的作用位置及大小，就可以求解出机翼后缘结构的变形状态。

2.3 后缘结构力学模型的验证

为了验证后缘结构力学模型的准确性，针对所建立的力学模型，开展基于ABAQUS软件的有限元仿真验证。机翼使用高性能尼龙材料制造，材料弹性模量为4.3×10^3 MPa。

以图2中得到的SMA完全相变后的等效驱动力作为输入载荷，以SET1、SET2和SET3分别驱动的3种驱动状态作为计算算例，对其基于后缘结构力学模型得到的理论结果和基于有限元模型得到的仿真结果进行分析对比。在ABAQUS中，将机翼后缘左侧固定，针对不同算例在后缘对应载荷作用点处施加载荷，SET1、SET2和SET3产生的等效驱动力分别对应载荷作用点1、2、3，如图4所示。以各子模块单元中铰链理论结果偏转角度θ_i与仿真结果偏转角度θ'_i的相对误差作为评价指标，可表示为

$$e=\frac{(\theta_i-\theta'_i)}{\theta_i}\times100\%\tag{12}$$

力学模型与有限元模型的结果对比如图5所示。在SET1驱动下，第1级铰链的相对误差为0.41%，后缘结构偏转角度的相对误差为0.82%；在

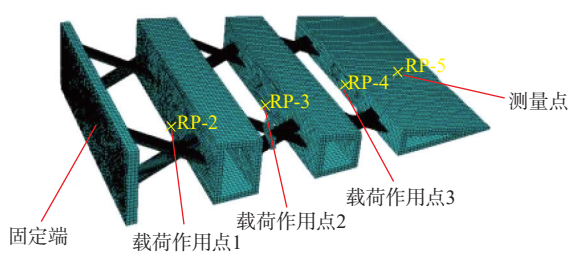


图4 后缘结构有限元网格及载荷作用位置示意图

Fig. 4 Finite element mesh and load application position diagram of the trailing edge structure

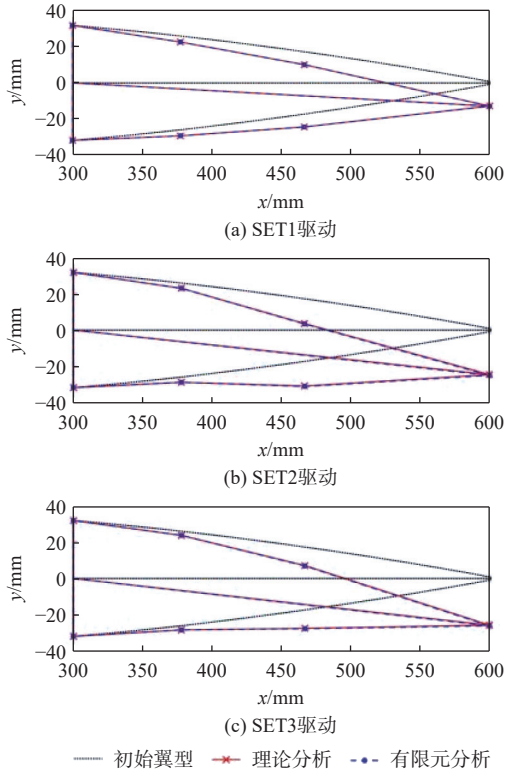


图 5　理论模型与有限元模型分析结果对比

Fig. 5　Comparison of analysis results between theoretical model and finite element model

SET2 驱动下, 第 1 级铰链和第 2 级铰链的相对误差分别是 1.77% 和 4.08%, 后缘结构偏转角度的相对误差为 1.49%; 在 SET3 驱动下, 第 1 级、第 2 级和第 3 级铰链的相对误差分别是 0.24%、3.49% 和 0.14%, 后缘结构偏转角度的相对误差为 1.02%。上述对比结果表明, 所采用的等效刚度模型具有较高的计算精度。

3　机构与气动弹性分析

3.1　机翼变形能力分析

机翼的变形能力取决于各子模块单元中铰链变形量的叠加, 而铰链的弯曲变形又由 SMA 驱动器控制, 因此, 各级 SMA 驱动器的驱动状态会直接影响机翼的变形效果。通过各级 SMA 驱动器的单独或协同驱动, 机翼可以产生不同的变弯度效果, 因而可实现多种状态变形。为了便于展示, 同时避免重复分析, 后续不考虑初始状态和 SET1、SET3 协同驱动状态, 以表 1 中的 6 种变形状态作为分析对象进行展示, 包括单组加热、双组加热、三组加热等情形。

为了直观地展示多级 SMA 驱动器对机翼变形的作用效果, 基于推导的后缘结构力学模型, 对状态 1、状态 4 和状态 6 进行数值分析, 得到各状态下机翼的变形效果, 如图 6 所示。

表 1　机翼多状态变形

Table 1　Multiple-state deformation of the wing

加热位置	变形状态
SET1	状态1
SET2	状态2
SET3	状态3
SET1、SET2	状态4
SET2、SET3	状态5
SET1、SET2、SET3	状态6

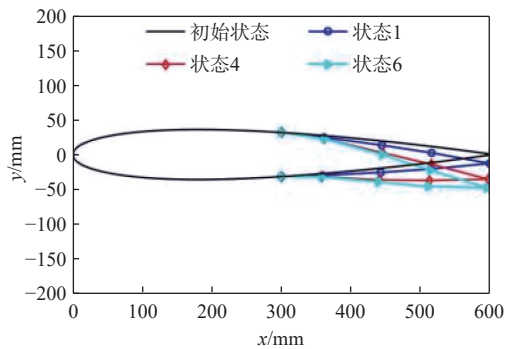


图 6　后缘结构多状态变形效果示意图

Fig. 6　Schematic diagram of multi-state deformation effect of the trailing edge structure

3.2　准静态气动弹性分析

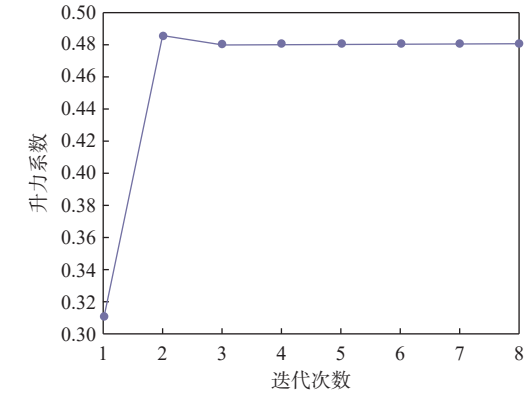
在变弯度机翼的设计过程中, 有必要分析机翼在实际工况下的飞行性能及与气动载荷之间的相互作用关系。因此, 以表 1 所示的机翼 6 种变形状态为分析算例, 对所设计变弯度机翼在等效驱动力作用下的结构变形与气动载荷双向耦合下的气动特性进行分析。

为了获取机翼在双向耦合作用下的气动特性变化情况, 建立以 XFOIL 为气动特性参数求解器, 同时, 结合 MATLAB 搭建了机翼流固耦合分析平台, 整个求解过程主要在 MATLAB 中实现, MATLAB 自主调用 XFOIL 软件进行气动载荷求解。变形后的机翼结构会导致翼面气动载荷发生改变, 附加的气动载荷又会反过来影响机翼的结构变形, 经过多次迭代运算后, 翼面气动载荷变化量和机翼变形量会逐渐收敛到平衡态。所设计变形机翼在表 2 的计算工况下, 机翼变形状态 3 的升、阻力系数随流固耦合迭代步数的变化曲线如图 7 所示, 机翼各状态气动系数变化情况如图 8 所示。结果表明, 该设计方案具备多状态调节能力, 进一步验证了机翼多状态变形方案的有益性和可行性。

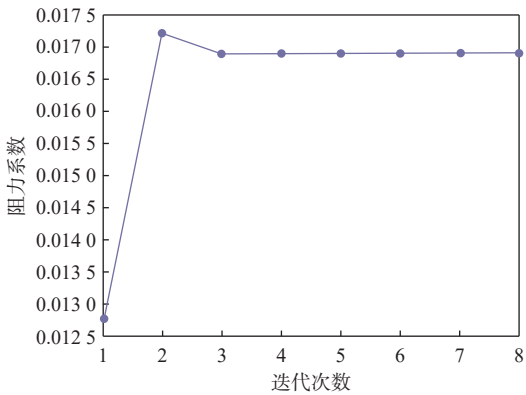
表 2　计算工况

Table 2　Calculate operating conditions

Ma	*Re*	*α*/(°)
0.3	10^5	4



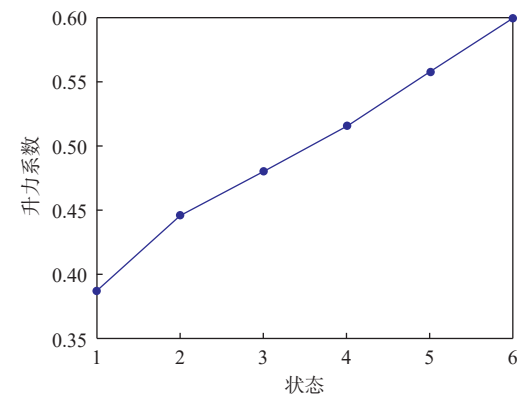
(a) 升力系数变化情况



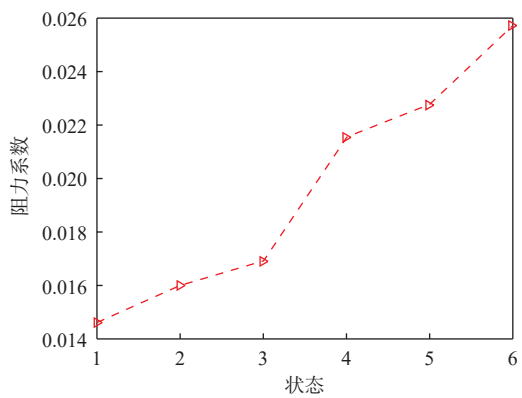
(b) 阻力系数变化情况

图 7 机翼变形状态 3 的气动系数变化

Fig. 7 The aerodynamic coefficient variation in wing state 3



(a) 各状态升力系数变化



(b) 各状态阻力系数变化

图 8 机翼各状态气动系数变化

Fig. 8 The aerodynamic coefficient variations across various wing states

4 变弯度机翼样机实验

4.1 机翼多状态变形能力验证

为了验证该变弯度机翼设计方案的实际变形能力及多状态变形效果，根据前述机翼结构模型完成样机的制造及安装，并搭建实验平台如图 9 所示。该实验平台用于开展变弯度机翼的静态变形实验，通过计算机终端控制可编程电源输出恒定电流对预紧后的 SMA 驱动器进行加热，使其完全相变，进而驱动机翼后缘结构变形，最后，通过激光位移传感器收集测量点的位移变化数据。SMA 预紧装置除了控制 SMA 初始预紧力外，还能实时监控相变过程中的等效力变化。



图 9 变弯度机翼样机实验平台

Fig. 9 Variable camber wing prototype experimental platform

基于该实验平台，对表 1 所示的机翼 6 种变形状态分别进行实验，各变形状态下测量点位移变化数据如图 10 所示。数据显示，多级 SMA 驱动器能够使机翼产生差异化的变弯度效果，满足其对多种弯度调节的需求。

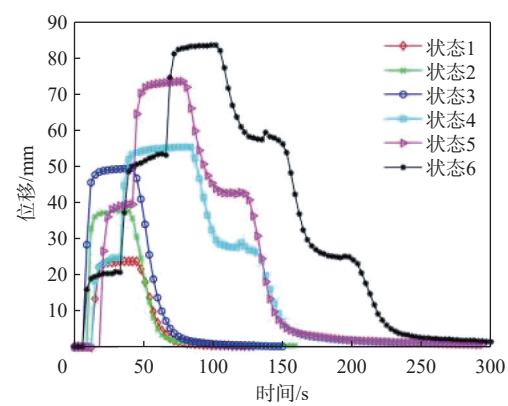


图 10 测量点位移变化曲线

Fig. 10 The displacement change curve of the measuring point

机翼后缘结构在各变形状态下的实验与仿真效果对比如图 11 所示, 两者变形趋势基本吻合, 说明该方案的设计、分析是合理的。在此基础上, 将各变形状态测量点最终位移量与基于有限元模型的仿真数据进行对比, 如图 12 所示, 并计算各变形状态下实验与仿真数据相对误差, 如表 3 所示。造成误差的主要原因包括: ①SMA 驱动器的预紧力设定不够精确; ②机翼结构的加工制造误差; ③仿真分析时施加的等效驱动力与实验中的实际驱动力略有差别。在后续研究过程中, 可以通过优化 SMA 预紧装置和提高仿真分析精度等方式减小结果误差。

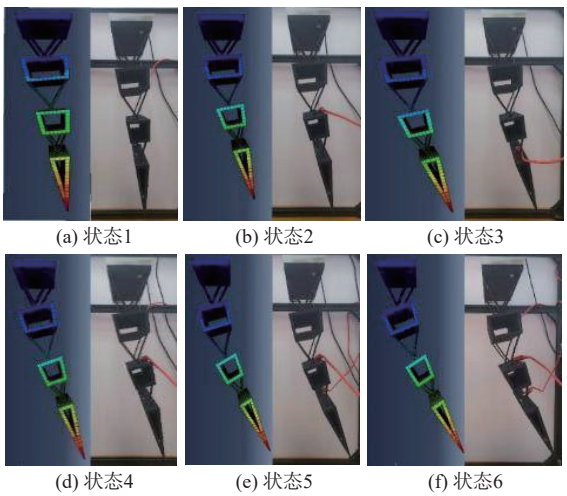


图 11　机翼结构实验与仿真变形趋势

Fig. 11　Experimental and simulation deformation trend in wing structure

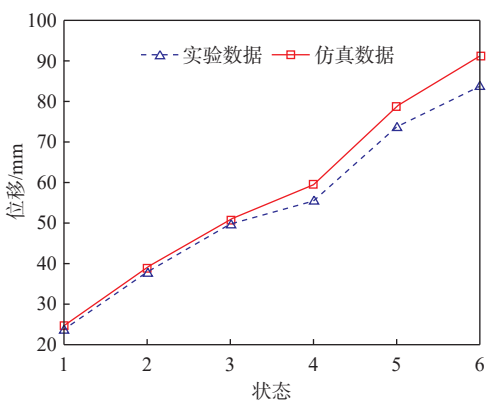


图 12　实验数据与仿真数据对比

Fig. 12　Comparison of experimental data and simulation data

当机翼变形至状态 6 时, 偏转角度最大, 如图 13 所示。该变弯度机翼向下偏转测量点最大位移可达到 83.6 mm, 综合考虑测量点与翼尖的位置关系, 可计算出等效偏转角为 17.8°。同理可得机翼各变形状态的等效偏转角, 如表 4 所示。

4.2　机翼多状态变形开环控制

为了探索该型机翼在实际应用中的变形适应能力, 在上述分析的基础上, 设计机翼多状态变形开环控制策略。机翼的多状态变形开环控制是在充分发挥 SMA 相变驱动能力的前提下, 进而提高机翼变形精度, 扩展机翼适用范围。机翼的多状态变形与各级 SMA 驱动器的驱动状态密切相关, 因而将多级 SMA 驱动器的驱动状态作为控制变量。

在实际情况下, 变形机翼需要根据不同任务的需求, 控制机翼变形至最优外形, 从而提高飞行器的气动效率。因此, 机翼多状态变形控制思想即: 根据不同工况条件下的最优气动特性, 同时结合多状态控制策略, 得到机翼最优变形状态, 进而确定需要参与驱动的 SMA 驱动器的位置, 最后驱动机翼变形, 如图 14 所示。控制策略主要基于前期实验得到的数据库及目标就近原则和能量最低原则。目标就近原则指数据库中机翼某个变形状态与目标状态最接近, 就将其作为最优变形状态; 能量最低原则指当数据库中有 2 个变形状态与目标状态较为接近时, 选择加热所需能量最少的变形状态作为最优变形状态。由于实验中未考虑气动力, 因此, 评价接近程度的指标为后缘结构上测量点(如图 4 所示)的位移量。

表 3　实验数据与仿真数据相对误差

Table 3　Relative error between experimental and simulation data

变形状态	相对误差/%
状态1	3.88
状态2	3.09
状态3	2.59
状态4	7.36
状态5	7.08
状态6	8.96

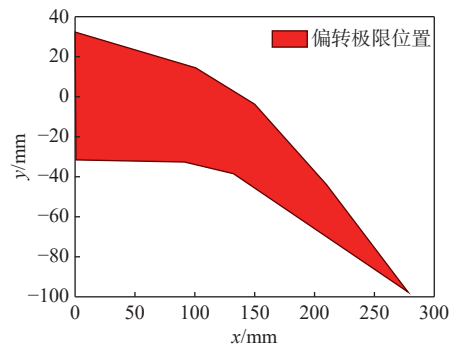


图 13　机翼最大偏转位置

Fig. 13　The maximum deflection position of the wing

表 4　机翼各变形状态的等效偏转角

Table 4　The equivalent deflection angles of the wing in various deformation states

变形状态	等效偏转角/(°)
状态1	5.0
状态2	8.0
状态3	10.5
状态4	11.8
状态5	15.6
状态6	17.8

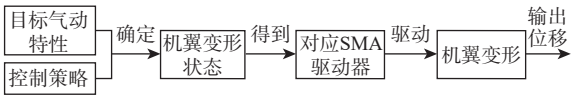


图 14 机翼多状态变形控制策略

Fig. 14 Multi-state wing deformation control strategy for aircraft wings

基于上述控制思想，开展面向目标变形状态的机翼多状态变形控制实验。将目标状态测量点位移数据输入控制电脑中，结合控制策略与数据库进行比对分析，得到机翼最优变形状态，通过对应的SMA驱动器驱动机翼结构变形，记录测量点的位移变化数据。参考变形机翼在单组加热、双组加热和三组加热等驱动情况下的驱动能力，选定开环控制的目标位移，并将其分别命名为目标位移1、目标位移2和目标位移3。将3组目标位移作为控制目标分别输入控制电脑，可得机翼多状态变形控制效果，如图15所示。结果表明，机翼多状态变形开环控制能够实现对不同目标状态的良好追踪。

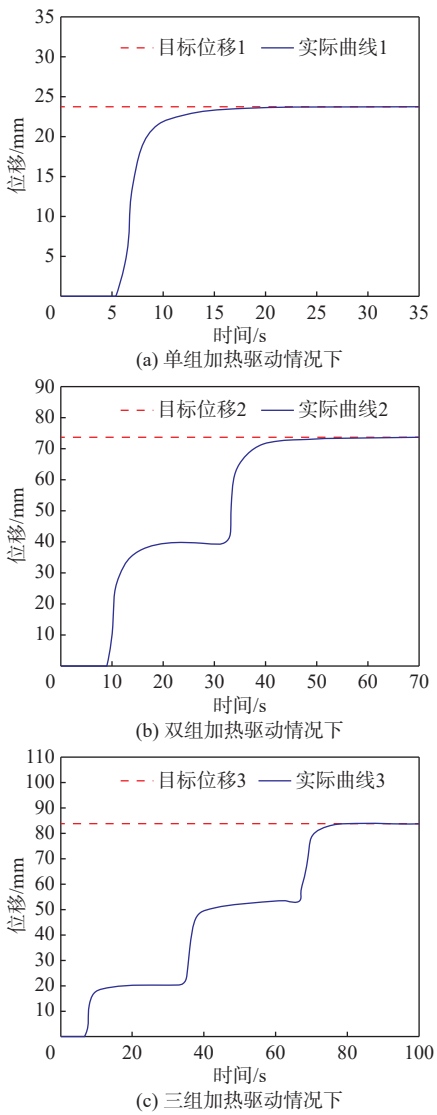


图 15 对目标位移的控制效果

Fig. 15 The control effect of target displacement

5 结 论

1）设计了一种基于SMA驱动的新型多状态连续变弯度机翼，具备较高精度下的多状态变形能力。

2）建立后缘结构单级、多级模块单元等效刚度力学模型，并进行了验证。

3）开展面向机翼多状态变形的气弹分析，验证了该设计方案在设定工况下的多状态调节能力。

4）搭建实验平台，验证了机翼的多状态变形能力及控制效果，具备工程可行性。

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] 王彬文, 杨宇, 钱战森, 等. 机翼变弯度技术研究进展[J]. 航空学报, 2022, 43(1): 024943.
WANG B W, YANG Y, QIAN Z S, et al. Technical development of variable camber wing: review[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(1): 024943(in Chinese).

[3] SHI X T, YANG Y, WANG Z G, et al. Design and shape monitoring of a morphing wing trailing edge[J]. Aerospace, 2023, 10(2): 127.

[4] LI Y Z, GE W J, ZHOU J, et al. Design and experiment of concentrated flexibility-based variable camber morphing wing[J]. Chinese Journal of Aeronautics, 2022, 35(5): 455-469.

[5] ZHAO A M, ZOU H, JIN H C, et al. Structural design and verification of an innovative whole adaptive variable camber wing[J]. Aerospace Science and Technology, 2019, 89: 11-18.

[6] CAMPANILE L F, SACHAU D. The belt-rib concept: a structronic approach to variable camber[J]. Journal of Intelligent Material Systems and Structures, 2000, 11(3): 215-224.

[7] 张盛, 杨宇, 王志刚, 等. 变弯度机翼后缘偏心梁设计与验证[J]. 航空学报, 2022, 43(6): 525892.
ZHANG S, YANG Y, WANG Z G, et al. Design and validation of eccentric beam for variable camber trailing edge[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(6): 525892(in Chinese).

[8] GONG X B, REN C W, SUN J, et al. 3D zero Poisson's ratio honeycomb structure for morphing wing applications[J]. Biomimetics, 2022, 7(4): 198.

[9] 王志刚, 杨宇, 段世慧. 基于参数化分析的柔性后缘优化设计[J]. 航空学报, 2017, 38(增刊 1): 721562.
WANG Z G, YANG Y, DUAN S H. Optimal design of flexible trailing edge based on parametric analysis[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(Sup 1): 721562(in Chinese).

[10] 赵立杰, 邸超, 李凯, 等. 多相材料机翼前缘柔性机构拓扑优化设计[J]. 机械设计, 2019, 36(1): 86-90.
ZHAO L J, DI C, LI K, et al. Topology optimization for the compli-

ant wing's leading edge based on the multi-phase materials[J]. Journal of Machine Design, 2019, 36(1): 86-90(in Chinese).

[11] 徐钧恒, 杨晓钧, 李兵. 基于交叉簧片式铰链的变弯度机翼机构设计[J]. 浙江大学学报 (工学版), 2022, 56(3): 444-451.
XU J H, YANG X J, LI B. Design of wing mechanism with variable camber based on cross-spring flexural pivots[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(3): 444-451(in Chinese).

[12] BRAGA R, RODRIGUES P F, CORDEIRO H, et al. The study of new NiTi actuators to reinforce the wing movement of aircraft systems[J]. Materials, 2022, 15(14): 4787.

[13] LEE S Y, LEE G Y. Largely deformable torsional soft morphing actuator created by twisted shape memory alloy wire and its application to a soft morphing wing[J]. Scientific Reports, 2023, 13(1): 17629.

[14] SILVA G C, SILVESTRE F, DONADON M V. A nonlinear aerothermoelastic model for slender composite beam-like wings with embedded shape memory alloys[J]. Composite Structures, 2022, 287: 115367.

[15] PICCIRILLO V, GÓES L C S, BALTHAZAR J M, et al. Deflection control of an aeroelastic system utilizing an antagonistic shape memory alloy actuator[J]. Meccanica, 2018, 53(4-5): 727-745.

[16] CHEN J J, SHEN X, LI J F. Zero Poisson's ratio flexible skin for potential two-dimensional wing morphing[J]. Aerospace Science and Technology, 2015, 45: 228-241.

[17] ZHAO H Z, BI S S. Accuracy characteristics of the generalized cross-spring pivot[J]. Mechanism and Machine Theory, 2010, 45(10): 1434-1448.

[18] BRINSON L C. One-dimensional constitutive behavior of shape memory alloys: thermomechanical derivation with non-constant material functions and redefined martensite internal variable[J]. Journal of Intelligent Material Systems and Structures, 1993, 4(2): 229-242.

Design, analysis, and experimentation of SMA-driven multi-state variable camber wing

WANG Ziang[1], LU Zhirong[1], LI Honghao[1], ZHOU Wenya[2], WANG Xiaoming[1,*]

(1. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China;
2. School of Mechanics and Aerospace Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: The most common type of morphing wing in the field of intelligent aircraft design for the future is variable camber, which offers a wealth of opportunities. This paper addresses the design requirements of a multi-state, continuously variable camber wing and proposes a novel design scheme integrating a modular rigid-flexible coupling structure, cross-leaf hinge, and multi-stage shape memory alloy (SMA) actuators. According to the scheme, the equivalent stiffness structural mechanics model of single-stage and multi-stage modular elements is derived by combining the equivalent stiffness calculation method of the cross-leaf hinge, and the comparison and verification are completed by finite element analysis. Furthermore, a quasi-static fluid-structure coupling analysis model was established, verifying the proposed variable camber wing scheme's multi-state adjustment capability and adaptability under various conditions using SMA's full phase transition driving, and providing the impact patterns of each deformation state on aerodynamic coefficients. Finally, a conceptual prototype was developed, and a testing platform was constructed. A feedforward open-loop control technique based on multi-state deformation control was successfully implemented, confirming the engineering feasibility of the suggested multi-state variable camber wing design.

Keywords: variable camber wing; cross-leaf hinge; multi-state; shape memory alloy; experimental verification

Received: 2024-01-02; **Accepted**: 2024-03-15; **Published Online**: 2024-04-03 15:59

URL: link.cnki.net/urlid/11.2625.V.20240403.1040.003

Foundation items: National Natural Science Foundation of China (12102096); Natural Science Foundation of Guangdong Province (2022A1515011885)

* **Corresponding author**. E-mail: wangxm@gzhu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0847

MRAM芯片单粒子功能中断机制研究

刘羽寒[1,2]，马英起[1,2,*]，李世行[3]，王菡滨[3]，王碧[3]

(1. 中国科学院国家空间科学中心，北京 100190；2. 中国科学院大学，北京 100049；
3. 北京航空航天大学 集成电路学院，北京 100191)

摘　　要：随着工艺尺寸的微缩，单粒子效应成为航天器故障的重要因素之一，单粒子功能中断(SEFI)在先进工艺节点集成电路中日益突出。磁性随机存储器(MRAM)作为一种高速读写的新型存储器件，存储单元抗辐射性能优异，但其外围电路对单粒子效应更为敏感。利用脉冲激光分析SEFI对芯片的功能影响是一种高效方法。采用脉冲激光聚焦微束定位其敏感区域位于MRAM外围读电路中的控制模块，读0循环下SEFI的等效LET阈值为25 (MeV·cm^2)/mg，发生SEFI时数据呈现百万级别的翻转错误，其翻转比特数随辐照的持续而不断上升，电流持续上升超过3 mA。分析发现，发生翻转时电流在控制端口存在正反馈效应，对MRAM的读电路通过电阻等效存储单元来建立仿真模型，得到了控制模块、灵敏放大和内存单元之间形成环路反馈结论。同时，利用电路级仿真软件对MRAM敏感区域控制模块进行单粒子故障注入，发现，控制模块同时控制的灵敏放大器(SA)单元数目越多，SA发生输出错误的时间越快。当超过3个SA被同时控制时，越易受辐照影响而发生功能中断。基于上述研究，建议从MRAM外围电路结构上抑制环路反馈，或将控制模块与SA之间采用异步处理进行抗辐射加固设计。

关 键 词：单粒子效应；单粒子翻转；单粒子功能中断；脉冲激光；磁性随机存储器
中图分类号：TN406; TN43; V241.07
文献标志码：A　　**文章编号**：1001-5965(2026)03-0926-09

磁性随机存储器(magnetic random access memory, MRAM)作为新型存储器的一种，通过磁阻状态切换模式来进行数据存储，提高了其抗辐射能力，被视为未来空间飞行任务普遍采用的理想器件[1-2]。但其外围电路依旧采用对空间辐射敏感的传统体硅工艺，在辐射环境下会发生数据读取错误，严重时会导致整个存储器无法正常工作。本文针对MRAM中产生的单粒子功能中断(single-event function interrupt, SEFI)进行研究，分析其特征规律，并探究SEFI的产生机制。

当前，国内外对SEFI的研究多是通过地面辐射手段得到关于SEFI的表现形式及敏感性分析。Colorado Springs公司在MRAM的重离子试验中发现2类错误持续时间不同的SEFI，并发现SEFI仅在读功能下发生[3]；加州理工学院同样在一款MRAM的重离子测试中观察到多种SEFI类型，范围从一百万到几百万比特错误[4]。有些错误可以通过设备重置清除，而有些错误只能通过重新写入内存来清除。法国EADS机构在对动态随机存储器(dynamic random access memory，DRAM)的单粒子测试中，定位出SEFI存在于逻辑区域内的寄存器；同时发现，重复与非重复(随机)模式会影响SEFI的发生概率[5]。北京微电子技术研究所对一款Space wire通讯路由器进行的重离子试验中，发

收稿日期：2024-01-02；**录用日期**：2024-04-17；**网络出版时间**：2024-05-06 10:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20240430.1403.001
基金项目：国家自然科学基金(U2241280，62104015)；中国科学院青年创新促进会优秀会员人才基金(Y2022057)
*通信作者. E-mail：myq@nssc.ac.cn

引用格式：刘羽寒，马英起，李世行，等. MRAM芯片单粒子功能中断机制研究[J]. 北京航空航天大学学报，2026，52(3)：926-934.
LIU Y H，MA Y Q，LI S X，et al. Research on single-event function interruption mechanism of MRAM chips[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：926-934 (in Chinese).

现随重离子线型能量转移(linear energy transfer, LET)值增大, SEFI 的发生次数增多, 且寄存器对 SEFI 更为敏感[6]。湘潭大学在铁电随机存储器(ferroelectric random access memory, FRAM)的辐照研究中, 根据束流停止后错误的表现形式将 SEFI 分为 2 类, 并通过对电路可控硅结构的分析认为是由寄存器发生微锁定导致[7]。国防科技大学通过对 SOC 芯片上的静态随机存储器(static random access memory, SRAM)及中央处理器内部的寄存器文件进行脉冲激光扫描, 发现 SEFI 发生于程序运行控制模块, 且 SRAM 对此更为敏感[8]。

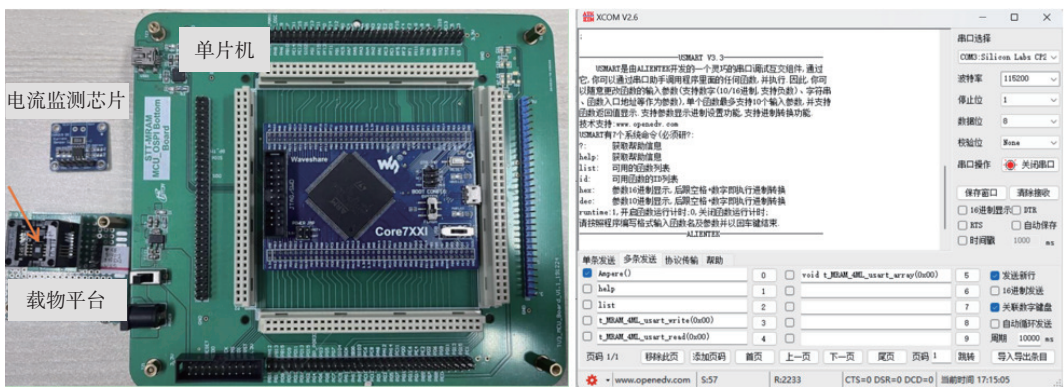
研究发现, SEFI 广泛存在于存储器件中(如 DRAM、FRAM、SRAM、MRAM 等), 可采用多种辐照模拟源来触发 SEFI, 但针对 SEFI 的综合特征及敏感区域电路的仿真及演化机制的分析相对缺乏。本文选择一款商用 MRAM 存储器作为试验对象, 将 MRAM 的 SEFI 产生特征及演化规律作为主要研究点, 采用脉冲激光装置对其进行辐照, 定位其敏感区域, 通过对敏感区域进行等效电路模拟仿真来研究 MRAM 的 SEFI 机制, 发现 MRAM 的控制模块、灵敏放大器(sense amplifier, SA)和内存单元在电路结构上构成环路, 同步控制的 SA 数目在持续辐照下对其工作状态产生影响, 为后续电路系统级抗 SEFI 防护提供试验与仿真支撑。

1　单粒子功能中断试验

1.1　试验对象及测试

选用一款商用 MRAM, 类型为 STT-MRAM, 其主要由存储单元和外围电路构成。该 MRAM 芯片的内存容量为 256 kB, 工作电流为 2～3 mA, 工作电压为 3.3 V, 对芯片塑封的背部进行开封装。

如图 1(a) 中箭头所示位置即为 MRAM 芯片, 该芯片依托于载物平台, 金属引脚与芯片一一对应。整套装置可以通过电脑上安装的测试系统程序代码向芯片存储位写入数据, 如图 1(b) 所示, 并通过电流监测芯片实时读取电流。



(a) Socket装置、单片机和电流监测芯片实物　　(b) SEU、SEFI测试系统

图 1　试验器件装置及测试软件

Fig. 1　Experimental device and test software

测试系统可采用 3 种方案来监测数据和电流: 静态读写、静态写/动态读及动态写读。考虑到需要实时监测辐照过程中的电流和错误比特以方便定位敏感区域, 本文试验采用静态写/动态读的方式。通过中国科学院国家空间科学中心空间天气效应中心自主研发的脉冲激光试验装置平台对器件进行辐照(等效 LET 值的范围为 0.1～120 (MeV·cm^2)/mg, 且连续可调), 激光的脉冲宽度为 25 ps, 波长为 1 064 nm, 器件背部入射的激光等效 LET 值换算公式如下:

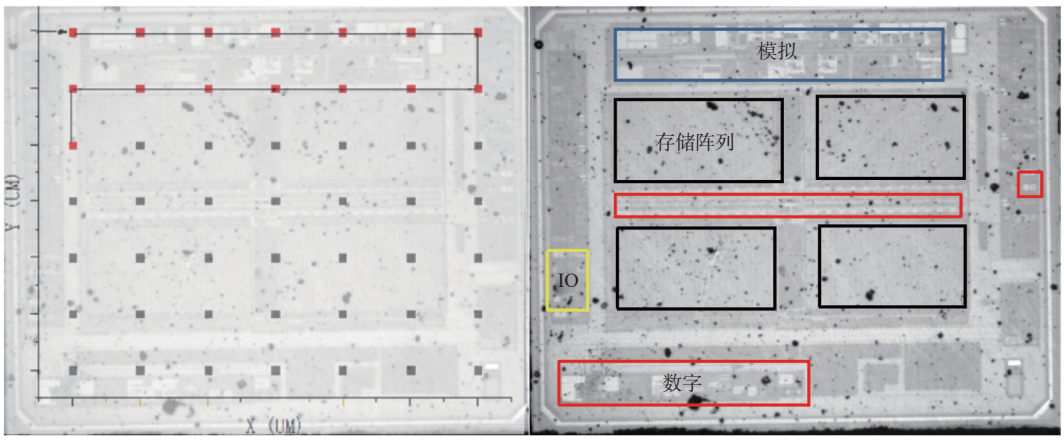
$$E_{\mathrm{LET}} = 0.082E_{\mathrm{eff}} \qquad (1)$$

式中: E_{eff}为诱发单粒子效应的有效激光能量, 其与硅衬底对激光的反射系数及吸收系数等参数相关[9]。

打开单粒子翻转 (single-event up, SEU)、SEFI 测试系统的串口与芯片相连, 开启电源, 设置工作参数。在试验前进行单次的全局写和循环读操作, 观测读取功能是否正常, 如果正常, 则开启激光对芯片进行辐照。在不同等效 LET 值下对芯片进行全局扫描和分区域扫描, 扫描方式如图 2 所示。

1) 故障注入 MRAM 芯片。激光辐照至芯片衬底有源区, 调整激光控制器, 使激光能量从 200 pJ 逐步增加, 根据式 (1) 对试验中的激光能量进行 LET 值换算。以 5 μm的扫描步距对全芯片进行扫描, 读取周期设置为 10 s。

对比 MRAM 的拓扑图, 可以将其分为 4 个模块——存储阵列、输入/输出(input/output, IO)模块、数字模块及模拟模块, 如图 2(b) 所示。对芯片进行分区域扫描, 利用激光注入芯片内部, 从芯片左上角每隔 2 μm注入一个脉冲激光, 读取周期设置为 6 s。



(a) 全局扫描 (b) 分区域扫描

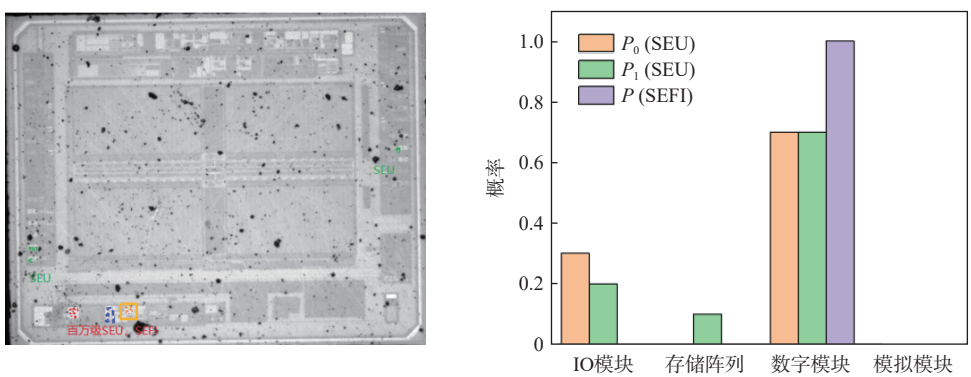
图 2 芯片背部激光红外成像扫描示意图

Fig. 2 Schematic diagram of laser infrared imaging scanning on the rear surface of a chip

2）SEFI的特征表现。MRAM试验前写入数据，开启激光，对芯片进行辐照，由于芯片背部开封装，激光可从芯片背部穿过硅衬底到达有源区，在辐照过程中实时读取芯片的工作电流和read模式下芯片的错误比特数。观测SEFI产生的特征现象、SEU数值变化和电流变化及在不同区域产生的单粒子效应现象。

1.2 试验结果

聚焦完成的脉冲激光对芯片进行分区域的逐点辐照，同时，上位机动态读取存储的数据（读取周期设置为10 s，激光扫描的速度为5 000 μm/s，因此，在每次进行读取时，激光都已辐照至新的地方，避免了辐照结果的重复性）。获得各种单粒子效应(single-event effect, SEE)的物理位置，对应到红外成像上的敏感区域，如图3(a)所示，SEFI的敏感区域为橙色框处电路，对比其电路功能模块判定，SEFI的敏感区域为控制电路。各个模块产生的单粒子效应概率如图3(b)所示：P_0(SEU)为存储数据为0时读出的翻转概率，P_1(SEU)为存储数据为1时读出的翻转概率。可以看出，SEFI只在数字模块发生，而SEU存在于IO模块和数字模块。在不同模块下发生的SEU比特数和特征都不一致，在数字模块的SEU是百万级的，在IO模块发生的SEU数目较少且大多数时候可以自恢复。



(a) 辐照敏感区域 (b) 各模块单粒子效应敏感区域

图 3 MRAM单粒子效应敏感区域

Fig. 3 SEE-sensitive areas in MRAM

在以上参数的激光辐照下，MRAM产生的各种单粒子效应如表1所示：激光辐照芯片时，上位机系统读取到翻转错误，有些翻转错误在下一个读取周期到来时就会消失，而有些翻转错误则持续存在，只有当激光关闭后才会消失。因此，根据其特征和恢复操作，将辐照中产生的SEE分为几种，如表1所示。根据读取的错误比特数，分为SEU和百万级SEU。动态回读时，随着激光能量的上升、LET值的增大，芯片在发生多位翻转后进而功能中断。

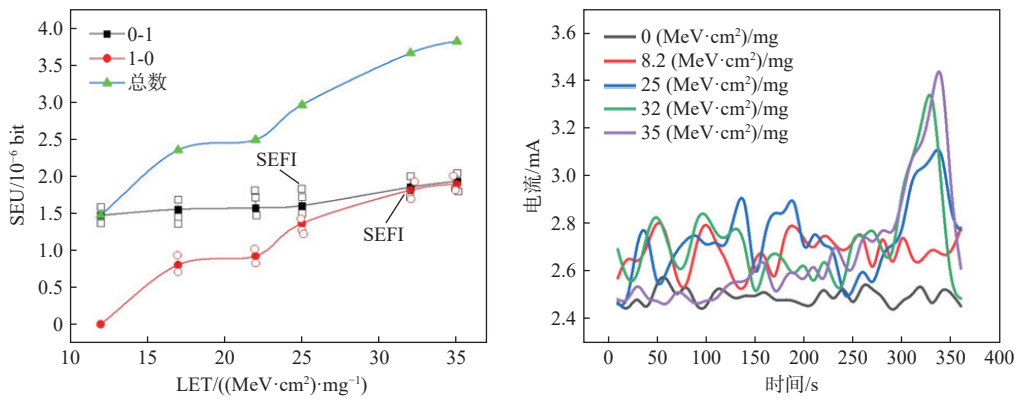
百万级翻转错误的平均数随着LET的变化如图4(a)所示：总的来看，LET值越大，翻转的数目越多；读0时，激光等效LET值为25 (MeV·cm^2)/mg，MRAM发生了SEFI，其翻转错误数随LET值的增长幅度较小，而读1时，SEFI等效LET阈值为

表 1　辐照中芯片产生的单粒子效应

Table 1　The single-event effect produced by the chip in irradiation

效应	现象	恢复操作
SEU	在下一个读取命令到来时, 消失	自恢复
	在下一个读取命令到来时, 会增大	自恢复
	在下一个读取命令到来时不会消失, 不随着辐照时间的增加而变化	关闭激光或直接重写
百万级SEU	数据百万级翻转, 在下一个读循环到来时持续增大	关闭激光或直接重写
SEFI	读取功能失效, 电流增大至超过3 mA, 下一个读取命令到来时消失	自恢复
	读取功能失效, 无法正常读取数据, 电流增大至超过3 mA, 激光停止后消失	关闭激光

35 (MeV·cm²)/mg, 其翻转数随着 LET 值的增大而显著增加。在发生 SEFI 的 LET 阈值之上(读 0)对其进行 2 组高能量值的辐照试验(LET 值分别为 32 (MeV·cm²)/mg、35 (MeV·cm²)/mg), 并在辐照过程中动态记录工作电流。工作电流变化如图 4(b)所示, 发生 SEFI 的扫描时刻为 280～310 s, 与辐照下未发生 SEFI 的电流及无辐照时的电流对比, 此时电流急剧增大, 且超过正常工作电流的范围。



(a) 翻转数随LET变化曲线　　(b) 读0时不同LET值下的工作电流

图 4　不同 LET 值下的翻转数和工作电流

Fig. 4　Number of flips and operating current under different LET values

对于 MRAM 中发生的 SEFI, 可以发现, 其读 0 时对 SEFI 更为敏感, 等效 LET 阈值为 25 (MeV·cm²)/mg。SEFI 的敏感区域主要位于参与读取过程的寄存器控制模块, 发生 SEFI 时数据先呈现百万级错误, 随着辐照的持续, 当下一次读取周期到来时, 错误比特数增加, 且读电流增大至超过 3 mA时发生 SEFI, 读功能中断无法读取数据, 可以通过关闭激光的方式恢复读功能。

试验中发生 SEFI 时的翻转位数和电流值如表 2 所示(读 0 状态下)。可以明显地观测到以下特征(此处区域在无激光辐照时的正常工作电流均值 $I_0 \approx 2.506\,6$ mA): ① 在此处读取命令成功运行后, 读取出成百上万位的数据翻转, 这种翻转错误数在激光的持续辐照下不断增加; ② 翻转电流略有增大, 并随着辐照的持续而不断增大; ③ 当翻转电流超过 3 mA时, 则在下一次读取命令运行时显示“FALL”而功能中断。

将上述过程视作翻转电流和读取命令控制端口的“反馈”过程, 如图 5 所示。发生多位翻转(multiple bit upset, MBU)时的电流做“反馈”给读取命令端口而产生中断, 利用反馈方程对此进行演算:

表 2　SEFI 时的翻转错误数和电流值（读 0 状态下）

Table 2　Number of upset errors and current value in SEFI (read 0)

等效LET/((MeV·cm²)·mg⁻¹)	翻转错误数/bit	翻转电流/mA
25	1 597 273	2.706 3
	1 599 136	2.745 1
	1 613 194	3.017 4
		3.026 4
		3.073
		3.162 2
32	1 786 082	2.968 2
	1 987 861	3.065 2
		3.210 7
		3.453 2
35	1 877 612	2.981 9
	1 991 041	3.103 3
		3.304 9
	FALL	3.571 9

注: 正常工作电流为2.57、2.425和2.541 4 mA, 均值约为2.506 6 mA。

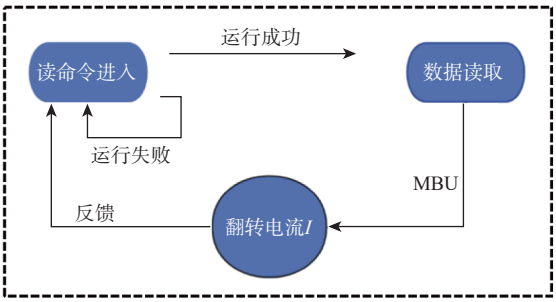


图 5 反馈电路示意图

Fig. 5 Schematic diagram of feedback circuit

$$GI_0 = I_{SEU} \tag{2}$$

$$G(I_0 - \beta I_{SEU}) = I_{SEU} \tag{3}$$

$$GI_0 = I_{SEU}(1 + \beta G) \tag{4}$$

式中：G为反馈增长因子；I_{SEU}为输出发生翻转时的电流；$\bar{I}_{SEU}$为翻转电流平均值。

根据表 2 的数据可以得出：$G_{25} = I_{SEU}/I_0 = \bar{I}_{SEU}/\bar{I}_0 = 2.823/2.506\,6 \approx 1.126$；$G_{30} = I_{SEU}/I_0 = \bar{I}_{SEU}/\bar{I}_0 = 3.016\,7/2.506\,6 \approx 1.204$；$G_{32} = I_{SEU}/I_0 = \bar{I}_{SEU}/\bar{I}_0 = 3.042\,6/2.506\,6 \approx 1.214$。

根据式 (2)～式 (4)，推导出反馈因子（也称反馈深度）$A = 1 + \beta G < 1$，可以发现，翻转产生的电流“正反馈放大”到读命令控制端口，使端口的净输入量增大，电路工作变得不稳定。

2 单粒子功能中断敏感区域电路级仿真

在激光辅助试验中，激光光斑最小可为1.3 μm，可以更详细地定位 SEFI 发生时的物理位置，对比拓扑图中的区域可以判定，SEFI 的敏感位置为控制模块。该模块输出信号来控制灵敏放大器的工作模式，而灵敏放大器主要是读取内存单元的数据。此时控制模块受辐照影响而导致读取数据发生翻转错误，且电流急剧上升超过最大工作电流时，功能中断。为了分析控制模块在辐照下导致数据翻转最终丧失读取功能的机制，对 MRAM 中 SEFI 的敏感区域——控制模块参与的读电路整体进行分析。

2.1 读电路单元模型的构建和功能验证

如图 6 所示，为一个读取电路单元，其主要构成为：控制模块、SA 及磁性隧道结（magnetic tunnel junctions, MTJ）。控制模块设计为一个简单的与门电路，由 CE 和 SE 这 2 个输入信号组成，输出信号作为读电路的使能端，控制读电路的工作状态；M2、M3、M4、M6 作为互锁结构的反相器，对内存单元 MTJ 进行读取输出。CE 和 SE 信号为高电平时，输出给 SA 使能端，使之进入正常工作状态，对内存数据进行读取。

其中，MTJ 的阻值作为数据输入，在相同读取电流下，MTJ1 处于高或低阻态（MTJ2 取值相反），存储的数据为“1”或“0”[10-16]，而正常工作下对应的输出电压则是高电平或低电平。利用 HSPICE 对读模式下的通用电路进行仿真，分析控制模块受辐照时导致读命令功能中断的机制。由于 HSPICE 系统中采用 0.18 μm工艺，在代码写入中晶体管有相应工艺参数设置，同时，仿真中 MTJ 通过电阻等效写入，采用北京航天航空大学在集成 MRAM 的设计中给出的典型 MTJ 器件对应阻值，将高阻态的电

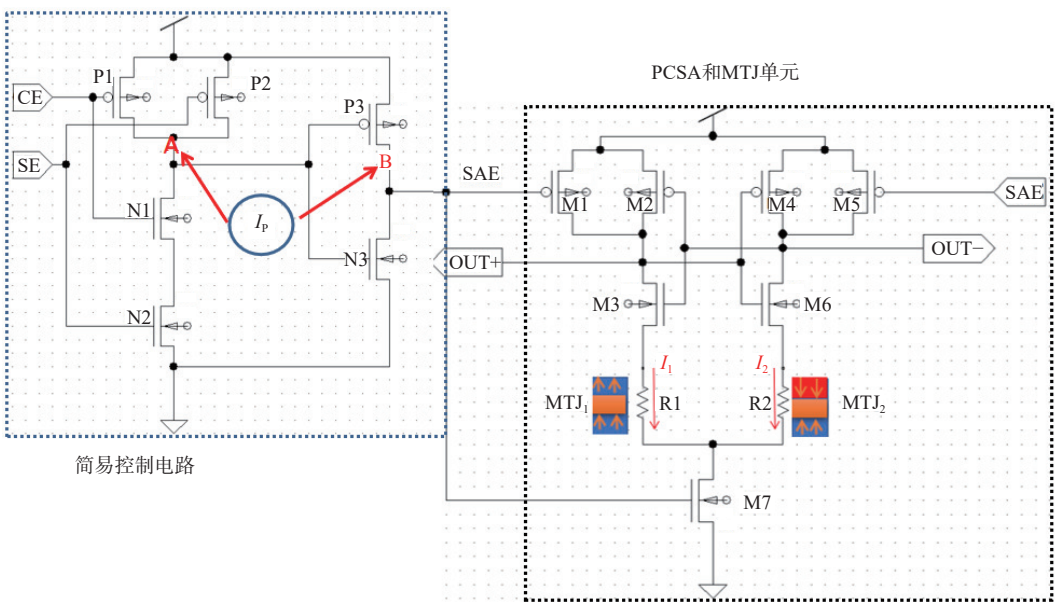


图 6 读电路示意图

Fig. 6 Schematic diagram of the reading circuit

阻设置为 9 kΩ, 低阻态下的等效电阻值为 3 kΩ[11]。对其进行等效电路的可行性验证。验证表明, 当 MTJ1 等效电阻为 9 kΩ时, 输出端为 1, 当 MTJ1 等效电阻为 3 kΩ时, 输出为 0, 满足 MRAM 读取机制。

2.2　仿真的故障注入

试验中激光辐照到控制模块, 读出数据发生多位翻转, 且电流不断上升, 导致功能中断。在 HSPICE 的仿真中, 通常采用对目标电路敏感节点注入单粒子效应电流源, 以此来获得电路的单粒子效应特征[11,17-18]。目前国际上常用的电流源模型为 Messenger 提出的双指数电流源模型[19-21]。对寄存器控制模块的敏感点引入双指数脉冲电流源模型I_P:

$$I_P = I_0(e^{-t/\tau_\alpha} - e^{-t/\tau_\beta})^* \tag{5}$$

$$I_0 = \frac{Q_{coll}}{\tau_\alpha - \tau_\beta} \tag{6}$$

$$Q_{coll} = 10.8LS \tag{7}$$

式中: I_0为脉冲电流的幅度; τ_α和τ_β分别为上升时间参数和下降时间参数, 其取值与半导体工艺相关; Q_{coll}为在衬底收集的电荷; L 为收集深度, 其中, 1 (MeV·cm^2)/mg的典型电荷收集深度为 2 μm[22-24]; S 为线性能量转移 LET, MeV。此时, τ_α和τ_β取值分别为 203 ns、200 ns[13]。将 SEFI 的 LET 阈值代入式 (5)、式 (6) 和式 (7), 得到电流源模型参数。

SEE 是由于高能粒子与器件相互作用, 产生电子-空穴对并进行电荷收集而导致的各种错误[8]。当 CE=1、SE=1 时, 读命令成功运行, PCSA 开始读取工作。此时, 图 6 控制电路中的 P1、P2 和 N3 处于截止状态, 体硅结构中漏极/衬底结处于偏压状态[25], 此节点在高能粒子的入射下更容易发生电荷收集[26], 因此, P1、P2 和 N3 的漏极对单粒子效应更为敏感。将图 6 中的 A、B 作为敏感节点注入瞬态脉冲I_P, 观测到 PCSA 读取的数据与正常状态下相比发生翻转。

为了分析 SEFI 的产生机制, 对读工作模式下的敏感区域控制模块与 SA 及内存单元之间的相互作用进行分析, 如图 7 所示。在该 MRAM 中, 上位机写入读命令, 使所有 SA 单元进入数据读取状态, 即控制模块同时输出多个 SA 单元使能命令。电路设计中, 单个 SA 对应读取多个内存单元, 当一个 SA 发生输出错误时, 对应读取的多个内存数据发生翻转[16]。而每个内存单元产生的读电流 I_1、I_2 又输入使能端口的 NMOS 器件, 因此, 电流在电路中构成环路, 反馈给信号输入端。

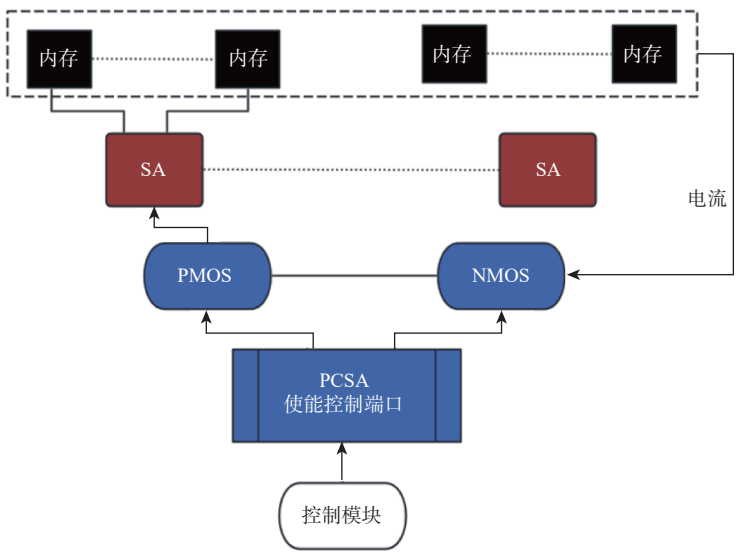


图 7　MRAM 中控制模块、SA 以及内存单元的相互作用

Fig. 7　Schematic diagram of the overall simplification of the MRAM read circuit

2.3　仿真结果

从试验结果可以发现, 读 0 时的 SEFI 阈值小于读 1 状态下的 SEFI 阈值, 读 0 状态下对 SEFI 更为敏感。因此, 对读 0 状态下的读电路进行分析。

选取控制模块 A 点注入 SEE 电流源, 如图 8 所示, 当控制模块与 1 个 SA 和与 2 个 SA 相连时的输出电压均发生翻转, 由 0 转变为 1.8 V; 当控制模块只控制 1 个 SA 单元时, 在 630 ps 发生翻转, 当其同时控制 2 个 SA 单元时, 输出发生翻转的时间为 590 ps, 比 1 个 SA 提前 40 ps 发生输出错误。同时, 随着控制的 SA 数目越多, 对应读取的内存单元越多, 受瞬态脉冲影响的翻转错误数不断增加。

控制模块的输出电压 SAE 如图 9 所示: V(SAE) 设置为 1.8 V, 假设 1.5～1.8 V的范围为高电平, 0～0.3 V的范围为低电平, 中间范围为亚稳态。随着控制的 SA 数目越多, 控制模块的输出电压 SAE 由高

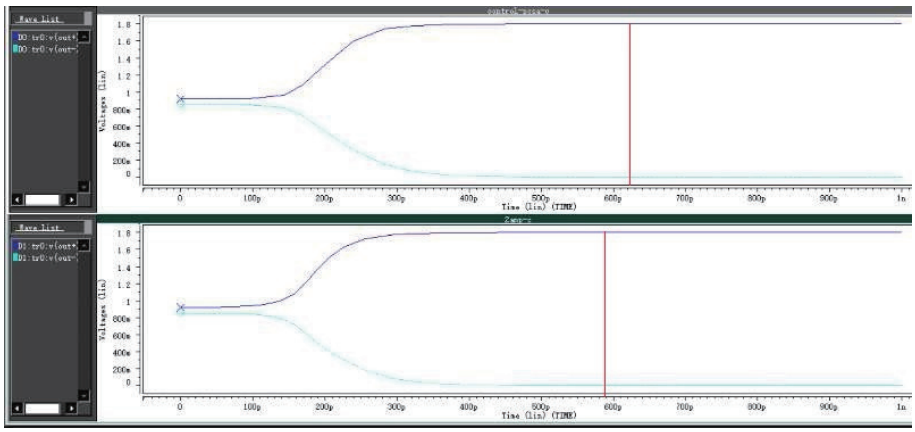


图 8 脉冲注入控制模块,连接不同数目 SA 时的输出电压

Fig. 8 Pulse injection control module with output voltage when connecting different numbers of SAs

电平不断降低,当超过 3 个 SA 被控制模块中的单粒子瞬态电流影响时,其输出在仿真周期中将持续处于低电平状态。SAE 电平信号决定 SA 的工作状态,控制模块输出的 SAE 信号处于低电平的时间越长,SA 越会由感应放大的工作状态变为预充电状态,无法读取内存数据。

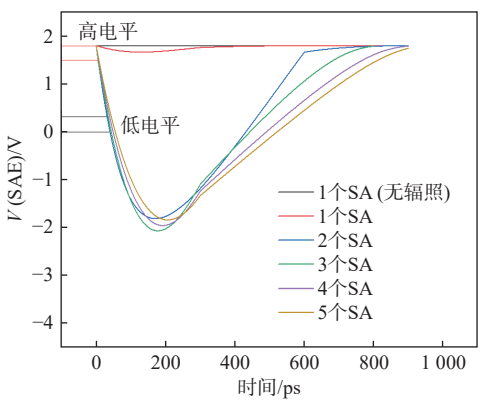


图 9 脉冲注入控制模块,连接不同数目 SA 时的使能电压 SAE

Fig. 9 Pulse injection control module with enabling voltage SAE when connected to different number of SAs

3 结 论

1)通过对商用 STT-MRAM 开展试验研究发现,利用不同激光有效能量脉冲对其进行辐照,MRAM 动态读 0 循环,当等效 LET 阈值为 25 (MeV·cm^2)/mg 时发生了 SEFI 效应,伴随着百万级的翻转错误,且随着激光等效 LET 值的增加,翻转错误数显著增大,产生的电流上升直至超过 3 mA时,读命令失效。发现此时电流反馈深度小于 1,处于正反馈状态,导致读命令的控制端口工作变得不稳定。

2)利用激光定位到发生 SEFI 的敏感区域位于外围读电路中的控制模块,对 MRAM 的外围读电路整体进行等效仿真分析,控制模块输出 SAE 信号作为灵敏放大器的使能端,当 SAE 信号为高电平时,控制灵敏放大器对内存单元 MTJ 的数据进行读取。此时,向控制模块敏感节点注入 SEE 电流源,发现随着控制模块注入的瞬态脉冲电流流经更多的 SA 后,其控制模块输出的电压 V(SAE) 下降幅度增大,且持续时间增长,影响 SA 的工作状态,无法读取内存数据。

3)通过仿真对读电路结构进行分析,发现从控制模块到 SA 再到内存数据,三者之间在电路结构上形成环路,瞬态电流在其中传播反馈到控制模块,影响读电路整体的工作模式,造成读功能中断。

基于上述工作给出 SEFI 防护建议:从电路结构上抑制“环路反馈”或将控制模块与 SA 之间采用异步处理进行抗辐射加固。同时,在试验中发现,MRAM 在读 1 状态下会在存储单元发生 SEU,且随着 LET 值增大,翻转数目显著增加,对于这 2 点的分析还需要后续对整个 MRAM 电路结构进行分析。本文的研究结果可为 MRAM 电路级抗 SEFI 技术的实现提供参考。

参考文献(References)

[1] 戈勇, 高一, 梅博, 等. 磁阻式随机存储器 (MRAM) 重离子单粒子效应试验研究[J]. 航天器环境工程, 2018, 35(6): 561-567.
GE Y, GAO Y, MEI B, et al. Experimental study of single event effect of heavy ions on magneto-resistive random access memory (MRAM) in space applications[J]. Spacecraft Environment Engineering, 2018, 35(6): 561-567(in Chinese).

[2] 杨茂森, 周昕杰, 陈瑶. MRAM 空间粒子辐射效应关键技术研究[J]. 电子与封装, 2021, 21(4): 62-70.
YANG M S, ZHOU X J, CHEN Y. Study on the key technologies of space particles particle radiation effects of MRAM[J]. Electronics & Packaging, 2021, 21(4): 62-70(in Chinese).

[3] HAFER C, VON THUN M, MUNDIE M, et al. SEU, SET, and SEFI test results of a hardened 16Mbit MRAM device[C]// Proceedings of the 2012 IEEE Radiation Effects Data Workshop. Piscataway: IEEE Press, 2012: 1-4.

[4] VARTANIAN S, YANG-SCHARLOTTA J, ALLEN G R, et al. Single event effects characterization of the ST-DDR4 spin-transfer torque magnetoresistive random access memory (STT-MRAM)[C]// Proceedings of the 2021 IEEE Nuclear and Space Radiation Effects Conference. Piscataway: IEEE Press, 2021: 1-7.

[5] BOUGEROL A, MILLER F, GUIBBAUD N, et al. Experimental demonstration of pattern influence on DRAM SEU and SEFI radiation sensitivities[J]. IEEE Transactions on Nuclear Science, 2011, 58(3): 1032-1039.

[6] 董攀, 陈莉明, 范隆, 等. Spacewire 通讯路由器单粒子功能中断测试系统研究[J]. 微电子学与计算机, 2014, 31(4): 173-176.
DONG P, CHEN L M, FAN L, et al. Research on the SEFI TEST system of spacewire communication router[J]. Microelectronics & Computer, 2014, 31(4): 173-176(in Chinese).

[7] 琚安安, 郭红霞, 张凤祁, 等. 铁电存储器中高能质子引发的单粒子功能中断效应实验研究[J]. 物理学报, 2018, 67(23): 248-254.
JU A A, GUO H X, ZHANG F Q, et al. Experimental study about single event functional interrupt of ferroelectric random access memory induced by 30–90 MeV proton[J]. Acta Physica Sinica, 2018, 67(23): 248-254(in Chinese).

[8] 胡春媚, 陈书明, 吴振宇, 等. 利用脉冲激光的片上系统芯片单粒子效应试验[J]. 国防科技大学学报, 2017, 39(2): 134-139.
HU C M, CHEN S M, WU Z Y, et al. Single event effect experiment on SoC using pulsed laser[J]. Journal of National University of Defense Technology, 2017, 39(2): 134-139(in Chinese).

[9] 上官士鹏. 脉冲激光模拟试验数字器件单粒子效应的机理与方法研究[D]. 中国科学院研究生院 (空间科学与应用研究中心), 2011.
SHANGGUAN S P. Study on the mechanism and method of single event effect of digital devices by pulsed laser simulation test[D]. Graduate School of Chinese Academy of Sciences (Space Science and Application Research Center), 2011(in Chinese).

[10] KIRKPATRICK S. Modeling diffusion and collection of charge from ionizing radiation in silicon devices[J]. IEEE Transactions on Electron Devices, 1979, 26(11): 1742-1753.

[11] SUN H L, ZHAO X P, HAN R K, et al. Compositionally modulated perpendicular magnetic anisotropy in tetragonal Mn_xAl films[J]. Applied Physics Letters, 2024, 124(10): 102404.

[12] 戴培芳. 基于 STT-MRAM 的存内计算逻辑设计及抗辐照特性的研究[D]. 无锡: 江南大学, 2022.
DAI P F. Research on in-memory calculation logic design and anti-irradiation characteristics based on STT-MRAM[D]. Wuxi: Jiangnan University, 2022(in Chinese).

[13] TAUSCH J, SLEETER D, RADAELLI D, et al. Neutron induced micro SEL events in COTS SRAM devices[C]// Proceedings of the 2007 IEEE Radiation Effects Data Workshop. Piscataway: IEEE Press, 2007, 168-188.

[14] 杨榕, 杜卓宏, 王乾元, 等. 国产先进工艺 SoC 器件空间单粒子效应试验研究[J]. 航天器环境工程, 2022, 39(2): 193-199.
YANG R, DU Z H, WANG Q Y, et al. Experimental study on space single event effect of domestic advanced SoC devices[J]. Spacecraft Environmental Engineering, 2022, 39(2): 193-199(in Chinese).

[15] 刘淼, 牛英山. 抗单粒子功能中断的加固技术研究[J]. 微处理机, 2020, 41(1): 1-5.
LIU M, NIU Y S. Research on reinforcement technology against single event functional interruption[J]. Microprocessor, 2020, 41(1): 1-5(in Chinese).

[16] Parkin S S P, Kaiser C, Panchula A, et al. Giant tunerlling magnetoresistance at room temperature with MgO(100) tunnel barriers[J]. Nature Materials, 2004, 3(12): 862-867.

[17] 贺泽. 纳米存储器件和组合逻辑器件的单粒子效应研究[D]. 兰州: 中国科学院大学 (中国科学院近代物理研究所), 2022.
HE Z. Single-event effect studies of nano-memory devices and combinatorial logic devices[D]. Lanzhou: University of Chinese Academy of Sciences (Institute of Modern Physics, Chinese Academy of Sciences), 2022(in Chinese).

[18] 陶元, 甯德雄. 现代 VLSI 器件基础[M]. 黄如, 王润声, 黎明, 等译. 第二版. 北京: 电子工业出版社, 2020: 45-46.
TAO Y, YONG D X. Fundamentals of modern VLSI devices[M] Translated by HUANG R, WANG R S, LI M, et al. 2nd Ed. Beijing: Publishing House of Electronics Industry, 2020: 45-46 (in Chinese).

[19] 江新帅, 罗尹虹, 赵雯, 等. 阱接触对 28 nm SRAM 单粒子多位翻转的影响[J]. 物理学报, 2023, 72(3): 229-235.
JIANG X S, LUO Y H, ZHAO W, et al. Effect of well contact on single event multi-bit flipping of 28 nm SRAM[J]. Acta Physica Sinica, 2023, 72(3): 229-235(in Chinese).

[20] 秦可勉, 彭海波, 茆亚南, 等. 激光诱发单粒子翻转的物理过程模拟[J]. 现代应用物理, 2023, 14(2): 161-166.
QING K M, PENG H B, MAO Y N, et al. Simulation of the physical process of laser-induced single event flipping[J]. Modern Applied Physics, 2023, 14(2): 161-166(in Chinese).

[21] Rzeszut P, Chęciński J, Brzozowski I, et al. Multi-state MRAM cells for hardware neuromorphic computing[J]. Sci Rep, 2022, 12: 71-78.

[22] Yoon I, Raychowdhury A. Test challenges in embedded STT-MRAM arrays[C]// Proceedings of the 2017 18th International Symposium on Quality Electronic Design, Santa Clara: [s.n.], 2017: 35-38.

[23] Prieto-Alfonso H, Del Peral L, Casolino M, et al. Radiation hardness assurance for the JEM-EUSO space mission[J]. Reliability Engineering & System Safety, 2015, 133: 137-145.

[24] 林珊珊. 比较器和放大器 SET 仿真分析研究[D]. 西安: 西安电子科技大学, 2018.
LIN S S. SET simulation analysis of comparator and amplifier [D]. Xi'an: Xidian University, 2018(in Chinese).

[25] 王定洪. 标准单元库单粒子效应敏感性分析[D]. 湘潭: 湘潭大学, 2021.
WANG D H. Sensitivity analysis of single particle effect of standard cell library[D]. Xiangtan: Xiangtan University, 2021(in Chinese).

[26] DODD P E, SHANEYFELT M R, FELIX J A, et al. Production and propagation of single-event transients in high-speed digital logic ICs[J]. IEEE Transactions on Nuclear Science, 2004, 51(6): 3278-3284.

Research on single-event function interruption mechanism of MRAM chips

LIU Yuhan[1, 2], MA Yingqi[1, 2, *], LI Shixing[3], WANG Hanbin[3], WANG Bi[3]

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. School of Integrated Circuits Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: The single-event functional interruption (SEFI) has grown in importance in advanced process node integrated circuits, and the single-event effect has become one of the key causes of spacecraft failure as process sizes have shrunk. As a new type of memory device with high-speed reading and writing, the magnetic random access memory (MRAM) has excellent radiation resistance of the memory cell, but its peripheral circuits are more sensitive to single-event effects. Pulsed laser analysis of the functional effect of SEFI on chips is an efficient method. The pulsed laser focusing micro-beam is used to locate the control module whose sensitive area is located in the peripheral reading circuit of MRAM, and the equivalent LET threshold of SEFI under the reading 0 cycle is 25 $(\mathrm{MeV}\cdot\mathrm{cm}^2)/\mathrm{mg}$, and the data presents a million-level flip error when SEFI occurs, and the number of flip bits continues to rise with the continuation of irradiation, and the current continues to rise by more than 3 mA. The analysis shows that there is a positive feedback effect of the current on the control port when the flip occurs, and the simulation model of the MRAM reading circuit is established through the resistor-equivalent memory unit, and the loop feedback conclusion between the control module, the sensitive amplification, and the memory unit is obtained. Simultaneously, single-event faults are injected into the MRAM sensitive area control module using circuit-level simulation software. It is discovered that the more sense amplifier (SA) units that are simultaneously controlled by the control module, the faster the SA's output error time. It is also discovered that the more SA that are simultaneously controlled, the more vulnerable they are to irradiation and function interruption. Based on the above research, it is proposed to suppress the "loop feedback" from the MRAM peripheral circuit structure or use asynchronous processing between the control module and the SA for the anti-radiation reinforcement design.

Keywords: single-event effects; single-event upset; single event function interruption; pulsed laser; MRAM

Received: 2024-01-02; **Accepted**: 2024-04-17; **Published Online**: 2024-05-06 10:39

URL: link.cnki.net/urlid/11.2625.V.20240430.1403.001

Foundation items: National Natural Science Foundation of China (U2241280, 62104015); Outstanding Member Talent Fund of Youth Innovation Promotion Association of Chinese Academy of Sciences (Y2022057)

* **Corresponding author**. E-mail: myq@nssc.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0791

融合邻域信息的民间文学知识图谱补全

高方正[1,2]，王笳辉[1,2]，段亮[1,2,*]，岳昆[1,2]，李忠斌[1,2]

(1. 云南大学 信息学院，昆明 650500；2. 云南大学 云南省智能系统与计算重点实验室，昆明 650500)

摘 要：作为民间文学的知识库，民间文学知识图谱(KG)在文本理解和知识服务等方面具有良好的应用前景，受到广泛关注。民间文学作品语言简洁，且常用短句，仅阐述民间文学人、事物间的部分关系，且KG通常难以包含领域内的所有知识，需对其进行补全。针对现有知识图谱补全(KGC)方法未能充分利用民间文学实体的不同邻域信息和关系内部的细粒度语义等问题，提出融合邻域信息的民间文学KG补全模型FolkKGC。设计关系感知注意力门控机制，融合实体邻域信息，获得民间文学实体的有效表示；设计基于交叉邻域相似度的关系学习器，对关系的邻域信息进行融合，构建细粒度的民间文学关系表示；对关系表示进行打分，完成知识图谱补全。从民间文学文本构造KG数据集，并对所提方法进行对比测试和消融实验。FolkKGC模型的MRR和Hits@*n*指标均高于现有模型，验证了其有效性。

关 键 词：民间文学知识图谱；知识图谱补全；小样本学习；邻域信息；门控循环单元
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2026)03-0935-10

近年来，随着大数据和人工智能技术的快速发展，数据密集型和知识密集型范式下的计算社会科学研究日益受到学界和业界的广泛关注。从海量数据中发现领域知识并进行计算分析，已成为相关领域研究的基本范式[1]。民间文学是民众在生活文化中传承、传播、共享的口头传统和语辞艺术，反映了人类各群体的基础性文化观念[2]。知识图谱(knowledge graph, KG)是由<头实体，关系，尾实体>形式三元组构成的大规模知识库[3]，能从海量数据中挖掘、组织和管理知识；民间文学知识图谱(folklore KG, FolkKG)一般通过对民间文学文本进行知识抽取而构建，是研究和传承民族历史文化知识的文学领域知识库。FolkKG从技术层面拓展了对民间文学文本的阐释和解读方式，可为问答系统、情感分析和文本挖掘等下游任务提供技术支撑[4]。

FolkKG来源于民间文学中古典史诗的活态文本和传承人口述[2]，大量的新知识会随着民间文学调查研究的推进而产生，其包含的知识也并不完善。然而，先前的FolkKG修缮工作很大程度上依赖于专家手工修补编制的案例数据，效率低且人工成本高。此外，FolkKG中存在大量样本数少于20条的关系三元组，甚至部分不常见关系三元组少于5条，小样本特征明显，因此，需要借助小样本场景下的知识图谱补全(KG completion, KGC)技术[4]对其进行完善，从而发掘民间文学作品中缺失的知识及情节之间的逻辑关系，补全人物在故事中隐含的社会关系和身份信息，帮助人们理解民间文学的深层内涵。

KGC旨在预测和补全KG三元组中缺失的实体或关系，通用的KGC模型包括翻译模型和张量分解模型[5]。翻译模型将实体和关系表示为连续向量空间中的低维向量，通过计算头尾实体间的距离为三元组打分，如TransE[6]及其扩展模型TransH[7]、TransR[8]和RotatE[9]等。张量分解模型将KGC定义

收稿日期：2023-12-04；**录用日期**：2024-04-19；**网络出版时间**：2024-05-07 14:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20240506.1326.001
基金项目：云南省基础研究计划面上项目(202301AT070193)；云南省教育厅科研基金项目(2022Y010)
*通信作者. E-mail：duanl@ynu.edu.cn

引用格式：高方正，王笳辉，段亮，等. 融合邻域信息的民间文学知识图谱补全[J]. 北京航空航天大学学报，2026，52(3)：935-944.
GAO F Z，WANG J H，DUAN L，et al. Folklore knowledge graph completion by fusing neighborhood information[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：935-944 (in Chinese).

为三阶二进制张量补全，基于张量与其分解来计算三元组得分，如RESCAL[10]及其扩展模型DistMult[11]和ComplEx[12]。在领域KGC中，文献[13]提出一种面向无对齐实体场景的多语言KGC模型，通过预训练语言模型生成伪对齐实体，以实现无对齐实体场景下的知识迁移；文献[14]提出一种基于时序KG的电力设备缺陷预测模型，通过注意力机制对多模态信息进行融合，利用关系感知的图神经网络得到实体和关系的时序表示，从而对电力设备进行缺陷预测。以上方法均需要大量的训练数据，但现实中的KG大都存在长尾关系，且需要动态更新，针对领域KGC任务，这一特点尤为突出[15]。

为了在关系样本较少的情况下，有效完成KGC，小样本知识图谱补全（few-shot KGC，FKGC）模型的研究日益受到学界关注，主要包括基于度量学习和基于元学习2类方法。基于度量学习的模型中，GMatching[16]使用邻域编码器和匹配处理器聚合实体的邻域特征，并比较查询集与参考集之间的相似性；FAAN[17]提出自适应注意力的邻域编码器和Transformer[18]关系编码器，能捕捉实体关系动态信息，但忽略了三元组间的实体交互。基于元学习的模型中，MetaR[19]将元学习应用于FKGC，针对没有背景知识的FKGC任务，可利用元信息快速更新模型参数，但忽略了不同参考集对关系表示的影响；GANA[20]使用门控网络和图注意力机制过滤邻域中的噪声，并设计用于处理复杂关系的打分函数，但对于多对多关系的处理效果仍欠佳；RSCL[21]使用三元组的图上下文来学习全局和局部的关系表示，能捕捉无直接邻域的三元组的相关信息，但不能有效减少长尾实体对关系表示的影响；HiRe[22]通过分层关系学习器以细化关系在实体、三元组和上下文3个层次上的表示，能有效利用三元组的交互信息，但很大程度上增加了算法复杂度；目前效果最佳的方法是NP-FKGC[23]，其利用神经过程和归一化流计算关系的不同概率分布，设计了基于路径感知注意力的匹配处理器，但不能有效减少噪声邻域的不良影响。

上述FKGC模型在通用数据集上取得了良好的效果，但针对FolkKG的补全任务仍存在以下2方面的问题：①民间文学实体拥有故事语义和属性语义2种不同的邻域语义信息，无法通过直接聚合实体邻域特征而生成准确的实体表示。②若不对同类民间文学关系内部存在的细粒度语义差距加以区分，则难以生成准确的关系表示。对此，本文提出一种融合邻域信息的民间文学知识图谱补全（FolkKG completion，FolkKGC）模型，通过聚合文学实体和文学关系的邻域特征来优化两者的表示，并通过自适应注意力机制为特定查询实体对打分，进而实现对FolkKG的补全。

针对问题①，本文以门控循环单元（gated recurrent unit，GRU）[24]这一由重置门和更新门组成、能较好捕捉时间序列中依赖关系的门控循环神经网络为基础，旨在有效表示民间文学实体所拥有的故事语义和属性语义2种不同的邻域信息，构造基于关系感知注意力和GRU的实体邻域聚合模块来更新实体表示。设计关系感知注意力函数，通过民间文学实体与其邻域特征间的运算来度量邻域特征的重要性，从而生成实体的故事邻域特征和属性邻域特征表示。进一步重构GRU的门控机制，使其通过特征信息而非状态信息进行更新，利用故事邻域特征和属性邻域特征生成门控信号，再自适应地控制2种特征的融合，并更新民间文学实体表示。

针对问题②，本文以通过跟踪数据中的关系来学习上下文并学习其含义的神经网络Transformer为基础，充分利用其自注意力机制可捕捉KG中的依赖关系的优势[15]，旨在有效刻画特定民间文学关系内部存在的细粒度语义差距，并学习关系的表示，构造基于交叉邻域相似度和Transformer的评分预测模块。首先，针对民间文学关系同侧实体相似度高的特点，设计交叉邻域相似度函数，将头/尾实体向量与尾/头实体邻域向量间的余弦相似度作为学习关系邻域特征的权重，优化参考集民间文学实体对的表示。然后，给出Transformer评分模块生成参考集和查询集实体对的语义表示，通过对比两者的相关程度，生成针对具体查询集实体对的细粒度关系表示，并为其评分。最后，根据评分结果补全FolkKG。

本文从民间文学文本中抽取到测试数据集，对所提方法进行性能对比、消融实验和参数影响测试。实验结果表明，本文提出的FolkKGC模型的MRR和Hits@n指标都高于对比模型，且其性能在参考集容量增大时仍保持稳定，验证了本文模型的有效性。

1 民间文学知识图谱补全

1.1 相关定义

定义1 给定一个民间文学三元组$<h,r,t>$ $(h,t\in E,r\in R)$，其中，E为民间文学实体集，R为民间文学关系集，r为头实体h和尾实体t之间的关系。FolkKG由大量民间文学三元组构成。

FolkKG可按关系类型分为故事知识图谱（story

KG, SKG)和属性知识图谱(attribute KG, AKG):

$$G_s=\{<h_s,r_s,t_s>|h_s,t_s\in E,r_s\in R_s\} \quad (1)$$

$$G_a=\{<h_a,r_a,t_a>|h_a,t_a\in E,r_a\in R_a\} \quad (2)$$

式中: 下标 s 和 a 分别表示故事和属性; R_s为故事关系集, 包含召相勐和哈依迭古等民族英雄的血缘、阵营和战争等故事关系; R_a为属性关系集, 包含属性关系, 如埋直积立别、英麻板和低雅蜡分别属于工具、山峰和妖怪实体类型。图 1 给出一个简单的 FolkKG 示例: SKG 记录了召相勐和嫡西里总布之间的夫妻关系、出生地、比喻修辞等故事信息, 而 AKG 记录了相关实体的属性信息。

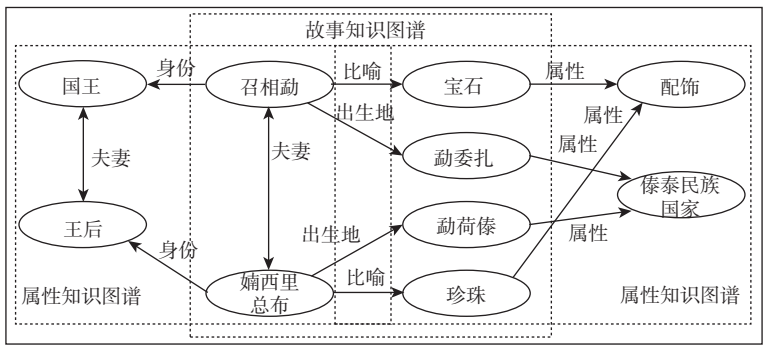


图 1　故事知识图谱和属性知识图谱示例

Fig. 1　Example of SKG and AKG

定义 2　给定 G 中的民间文学关系 $r\in R$及其参考集 $S_r=\{(h_s^1,t_s^1),(h_s^2,t_s^2),\cdots,(h_s^K,t_s^K)|h_s,t_s\in E\}$, 小样本 FolkKGC 旨在补全三元组 $<h,r,?>$中缺失的尾实体 t, 即评判 r的查询集 $Q_r=\{(h_q,t_q/t_q^-)|_q,t_q,t_q^-\in E\}$中候选尾实体的正确性, 其中, (h_s,t_s)为 S_r实体对, K 为 S_r中实体对的数量, (h_q,t_q)和(h_q,t_q^-)分别为 Q_r中的正负样本实体对。

每个民间文学关系 r对应一个补全任务 $T_r=\{S_r,Q_r\}$, 如表 1 所示。本文研究不同 K 值(K=1, 3, 5)的 FolkKGC 方法。

1.2　模型框架

本文提出的 FolkKGC 模型框架如图 2 所示。各层描述如下:

1) 预嵌入层(pre-embedding layer, PEL): 基于 TransH 的评分函数将 G 中的民间文学实体和关系表示为 d 维向量。

表 1　民间文学知识图谱补全任务示例

Table 1　Example of FolkKGC tasks

关系	集合	头实体	尾实体
r_1演奏	S_{r_1}	俸改	仙笛
		杰罗	口琴?
	Q_{r_1}	荷花仙女	十二弦琴?
		⋮	⋮
r_2坐骑	S_{r_2}	十头王	飞车
		嫡西里总布	白象?
	Q_{r_2}	撒赛歇	龙马?
		⋮	⋮

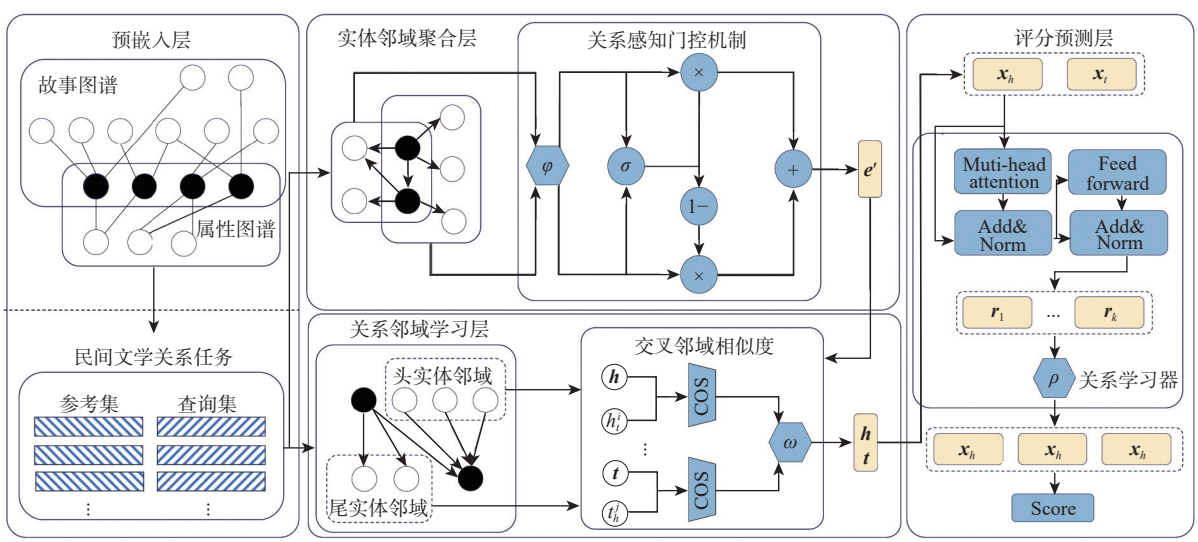


图 2　FolkKGC 模型框架

Fig. 2　Framework of FolkKGC

$$\begin{cases} \boldsymbol{h}_\perp = \boldsymbol{h}' - \boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{h}'\boldsymbol{w}_r \\ \boldsymbol{t}_\perp = \boldsymbol{t}' - \boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{t}'\boldsymbol{w}_r \end{cases} \tag{3}$$

$$S(h,t) = \|\boldsymbol{h}_\perp + \boldsymbol{r} - \boldsymbol{t}_\perp\|_{L1/L2} \tag{4}$$

式中：$\boldsymbol{h}'$和$\boldsymbol{t}'$为实体h和t的特征向量；$\boldsymbol{w}_r$为$\boldsymbol{r}$超平面的法向量，$\|\boldsymbol{w}_r\|_2 = 1$。实体对向量$(h, t)$映射为$\boldsymbol{r}$超平面上的投影。

2）实体邻域聚合（entity neighbor aggregation, ENA）层：融合实体邻域特征，生成民间文学实体的增强表示。

3）关系邻域学习（relation neighbor learning, RNL）层：学习关系的邻域特征，生成S_r实体对的增强表示。

4）评分预测层（score prediction layer, SPL）：生成针对具体Q_r实体对的民间文学关系表示，进而对实体进行打分，并完成补全。

1.3 民间文学实体表示

经PEL将G中所有实体和关系表示为向量后，由ENA层聚合民间文学实体邻域信息，以获得高质量的实体表示。定义距离T_r中民间文学实体e一跳的关系和实体集合为e的邻域，e在SKG中的邻域为N_s，表示其故事信息，在AKG中的邻域为N_a，表示其属性信息，具体描述如下：

$$N_\mathrm{s} = \{(r_\mathrm{s}, e_\mathrm{s}) | < e, r_\mathrm{s}, e_\mathrm{s} > \in G_\mathrm{s}, e_\mathrm{s} \in E\} \tag{5}$$

$$N_\mathrm{a} = \{(r_\mathrm{a}, e_\mathrm{a}) | < e, r_\mathrm{a}, e_\mathrm{a} > \in G_\mathrm{a}, e_\mathrm{a} \in E\} \tag{6}$$

ENA层在区分2个不同邻域信息的基础上对e进行特征聚合。首先，设计如下关系感知注意力φ度量邻域实体对e的重要性，使用tanh激活函数将(e, r_s)映射到故事邻域实体e_s上，得到重要性权重φ。φ越大，e_s与$\boldsymbol{e}'$关系越紧密，对$\boldsymbol{e}'$的特征更新越重要：

$$\varphi'(r_\mathrm{s}, e_\mathrm{s}) = \boldsymbol{e}_{\mathrm{s}\perp}^{\mathrm{T}} \tanh(\boldsymbol{e}_\perp + \boldsymbol{r}_\mathrm{s}) \tag{7}$$

$$\varphi(r_\mathrm{s}, e_\mathrm{s}) = \mathrm{soft}\max(\varphi'(r_\mathrm{s}, e_\mathrm{s})) \tag{8}$$

式中：$\boldsymbol{e}_{\mathrm{s}\perp}$和$\boldsymbol{e}_\perp$分别为经式（3）得到的$e_\mathrm{s}$和$e$在邻域关系$r_\mathrm{s}$超平面上的投影；$(\cdot)^{\mathrm{T}}$表示向量转置运算。

进一步基于φ聚合e的邻域特征生成故事邻域特征表示$\boldsymbol{e}_{N_\mathrm{s}}$：

$$\boldsymbol{e}_{N_\mathrm{s}} = \sum_{(r_\mathrm{s}, e_\mathrm{s}) \in N_\mathrm{s}} \varphi(r_\mathrm{s}, e_\mathrm{s}) \boldsymbol{e}_\mathrm{s} \tag{9}$$

使用相同的方法获得e的属性邻域特征表示$\boldsymbol{e}_{N_\mathrm{a}}$。针对$e$的2种邻域特征表示，本节扩展GRU更新门的计算公式，学习控制$\boldsymbol{e}_{N_\mathrm{s}}$和$\boldsymbol{e}_{N_\mathrm{a}}$两者融合的门控信号$\boldsymbol{g}$，再根据$\boldsymbol{g}$自适应地将2种特征信息融合为总邻域特征表示$\boldsymbol{e}_\mathrm{n}$：

$$\boldsymbol{g} = \sigma(\boldsymbol{W}_{N_\mathrm{s}}\boldsymbol{e}_{N_\mathrm{s}} + \boldsymbol{W}_{N_\mathrm{a}}\boldsymbol{e}_{N_\mathrm{a}}) \tag{10}$$

$$\boldsymbol{e}_\mathrm{n} = \boldsymbol{g} \odot \boldsymbol{e}_{N_\mathrm{s}} + (\boldsymbol{I} - \boldsymbol{g}) \odot \boldsymbol{e}_{N_\mathrm{a}} \tag{11}$$

式中：$\boldsymbol{W}_{N_\mathrm{s}}$和$\boldsymbol{W}_{N_\mathrm{a}}(\boldsymbol{W}_{N_\mathrm{s}}, \boldsymbol{W}_{N_\mathrm{a}} \in \mathbf{R}^{d\times d})$为模型可学习的权重矩阵；$\boldsymbol{g} \in \mathbf{R}^{d\times 1}$；$\boldsymbol{I}$为$d\times 1$维全1向量；$\odot$为向量对应维度分量间的乘法运算；$\sigma$为sigmoid激活函数。

由于民间文学实体e的邻域大小不固定，不能用$\boldsymbol{e}_\mathrm{n}$完全代替其初始嵌入表示$\boldsymbol{e}_\mathrm{o}$，故定义可学习的参数$\boldsymbol{\alpha}$来平衡$\boldsymbol{e}_\mathrm{o}$和$\boldsymbol{e}_\mathrm{n}$，得到如下融合邻域特征的实体表示：

$$\boldsymbol{e}' = \boldsymbol{\alpha} \odot \boldsymbol{e}_\mathrm{n} + (\boldsymbol{I} - \boldsymbol{\alpha}) \odot \boldsymbol{e}_\mathrm{o} \qquad \boldsymbol{\alpha} \in \sigma(\boldsymbol{W}_\mathrm{n}\boldsymbol{e}_\mathrm{n}) \tag{12}$$

式中：$\boldsymbol{W}_\mathrm{n}(\boldsymbol{W}_\mathrm{n} \in \mathbf{R}^{d\times d})$为模型可学习的权重矩阵，$\boldsymbol{\alpha} \in \mathbf{R}^{d\times 1}$。

1.4 民间文学关系表示

基于ENA层获得的T_r中所有民间文学实体的表示，由RNL层和SPL联合完成学习民间文学关系表示的任务。

T_r中民间文学三元组在一侧（头/尾）实体确定时，另一侧（尾/头）实体的相似度越高，对应关系的相关性也越强，且民间文学三元组也普遍存在头/尾实体邻域较小的情况，故RNL层在学习关系的邻域特征时，需综合考虑头尾2个邻域。在图3的示例中，古嫡提拉和召相勐均为皇室成员，而皇室成员常被比喻为宝石和明珠等各种配饰；兹兹朴乌、只都城和甘洛只孜均为彝区地名，都是英雄哈依迭古去过的地方。

民间文学关系r的头实体邻域N_h为S_r实体对(h, t)的相邻三元组中指向尾实体t的实体和关系：

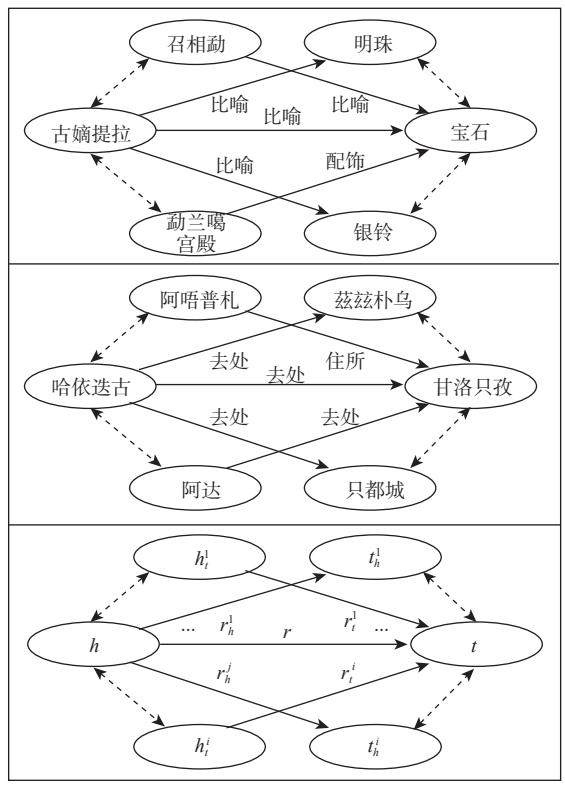


图3 民间文学关系邻域示例

Fig. 3 Example of FolkKG's relation neighborhood

$$N_h=\{(h_t^1,r_t^1),(h_t^2,r_t^2),\cdots,(h_t^M,r_t^M)|<h_t,r_t,t>\in G,M=|N_h|\} \tag{13}$$

r 的尾实体邻域 N_t 为头实体 h 指出的关系和实体:

$$N_t=\{(r_h^1,t_h^1),(r_h^2,t_h^2),\cdots,(r_h^J,t_h^J)|<h,r_h,t_h>\in G,J=|N_t|\} \tag{14}$$

式中: M 和 J 分别为 N_h 和 N_t 的大小。

在 RNL 层中设计度量关系邻域重要性的交叉邻域相似度函数 ω, 将 N_h 中的 h_t^i 与 h 间的余弦相似度作为聚合 (h_t^i,r_t^i) 特征的权重, ω 越大, 邻域关系 r_t^i 与 r 的相关性越强:

$$\omega'(h_t^i)=\cos(\boldsymbol{h}_t^i,\boldsymbol{h}') \tag{15}$$

$$\omega(h_t^i)=\mathrm{soft}\max(\omega'(h_t^i)) \tag{16}$$

式中: h_t^i 为 N_h 中第 $i(i\leqslant I)$ 个实体; $\cos(\boldsymbol{a},\boldsymbol{b})$ 为 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 这 2 个向量间的余弦相似度。

基于 ω 生成 h 的 N_h 邻域特征表示 $\boldsymbol{h}_{N_h}$:

$$\boldsymbol{h}_{N_h}=\sum_{j=1}^{J}\omega(h_t^i)\boldsymbol{h}_t^i \tag{17}$$

进一步定义可学习的参数 $\boldsymbol{\beta}$, 以平衡 $\boldsymbol{h}'$ 和 $\boldsymbol{h}_{N_h}$, 从而得到 h 的邻域特征增强表示 $\boldsymbol{x}_h$:

$$\boldsymbol{x}_h=\boldsymbol{\beta}\odot\boldsymbol{h}_{N_h}+(\boldsymbol{I}-\boldsymbol{\beta})\odot\boldsymbol{h}_{\mathrm{o}},\boldsymbol{\beta}\in\sigma(\boldsymbol{W}_r\boldsymbol{h}_{N_h}) \tag{18}$$

式中: $\boldsymbol{W}_r(\boldsymbol{W}_r\in\mathbf{R}^{d\times d})$ 为模型可学习的权重矩阵; $\boldsymbol{\beta}\in\mathbf{R}^{d\times 1}$; $\boldsymbol{h}_{\mathrm{o}}$ 为 h 经过 ENA 层更新后的实体表示。

同理可得 t 的增强表示 $\boldsymbol{x}_t$。在 RNL 层处理完所有 S_r 上民间文学实体对后, 令 $\boldsymbol{X}=(\boldsymbol{x}_h,\boldsymbol{x}_t)$ 为 S_r 上第 $k(k\leqslant 5)$ 个实体对在 SPL 上的初始输入。将 $\boldsymbol{x}_h$ 和 $\boldsymbol{x}_t$ 分别加上各自的位置嵌入向量作为 SPL 的实际输入:

$$\boldsymbol{X}_h^0=\boldsymbol{x}_h+\boldsymbol{x}_h^{\mathrm{p}} \tag{19}$$

式中: $\boldsymbol{x}_h^{\mathrm{p}}$ 为 $\boldsymbol{x}_h$ 的正弦位置嵌入。同理可得 $\boldsymbol{X}_t^0$。

SPL 使用 L 层的 Transformer 编码器来学习实体对代表的关系表示, Transformer 编码器由多头注意力模块和前馈神经网络组成。SPL 层通过多头注意力模块感知 $\boldsymbol{X}$ 保存的位置信息, 从而学习 $\boldsymbol{X}$ 对应实体对的上下文语义特征:

$$\boldsymbol{X}_i^l=\mathrm{Transformer}(\boldsymbol{X}_i^{l-1}) \tag{20}$$

式中: Transformer$(\cdot)$ 表示 Transformer 编码操作; $\boldsymbol{X}_i^l$ 为实体对序列 $\boldsymbol{X}$ 经过第 l 层后的隐藏状态。取 $\boldsymbol{r}_k=\boldsymbol{X}_i^l$ 作为 S_r 实体对 (h_k,t_k) 对应的最终民间文学关系表示, $\boldsymbol{r}_k$ 可反映实体对隐含的关系语义。

SPL 进一步将关系表示 $\boldsymbol{r}_k$ 迁移到 Q_r 上, 将 Q_r 实体对 (h_q,t_q) 输入 Transformer 编码器, 得到其对应的关系表示 $\boldsymbol{r}_q$, 设计 $\boldsymbol{r}_q$ 与 $\boldsymbol{r}_k$ 间的关系相似度函数 ρ 来衡量两者在语义上的相关性, 并将其作为权重计算特定于 Q_r 实体对的民间文学关系表示 $\boldsymbol{R}_q$:

$$\rho'(\boldsymbol{r}_k)=\boldsymbol{r}_q\boldsymbol{r}_k \tag{21}$$

$$\rho(\boldsymbol{r}_k)=\mathrm{soft}\max(\rho'(\boldsymbol{r}_k)) \tag{22}$$

$$\boldsymbol{R}_q=\sum_{\boldsymbol{r}_k\in S_r}\rho(\boldsymbol{r}_k)\times\boldsymbol{r}_k \tag{23}$$

最后在 SPL 层中, 使用 $\boldsymbol{R}_q$ 替换 TransH 的评分函数中的 r, 为 Q_r 候选实体对打分:

$$S=\left\|(\boldsymbol{h}_q-\boldsymbol{h}_{q\perp})+\boldsymbol{R}_q-(\boldsymbol{t}_q-\boldsymbol{t}_{q\perp})\right\|_{L1/L2} \tag{24}$$

1.5　模型训练

FolkKGC 遵循元学习 FKGC 模型的训练设置, 采用基于批量抽样的元训练过程[20] 来优化损失函数和模型参数。在 T_r 中, 先通过 S_r 上的合页损失函数(hinge loss)来更新民间文学关系表示, 再使用 Q_r 上损失函数来优化模型效果, 并更新参数(见式(25)), 具体训练方法如下:

输入: 故事图谱 G_{s}, 属性图谱 G_{a}, 训练集 D_{train}。

输出: 更新后的模型参数 ϕ。

1. while not done do
2. 在 D_{train} 中随机采样得到 $T_r=\{S_r,Q_r\}$;
3. 使用式 (7)～式 (12) 生成 T_r 中实体的增强表示 $\boldsymbol{e}'$;
4. 使用式 (15)～式 (18) 优化 S_r 上的实体对表示 $\boldsymbol{X}$;
5. 使用式 (19)～式 (20) 生成 S_r 上的关系表示 $\boldsymbol{r}_k$;
6. 计算 S_r 的损失函数 $L(S_r)$;
7. 使用 $L(S_r)$ 的梯度信息更新 $\boldsymbol{r}_k$;
8. 使用式 (21)～式 (23) 生成 Q_r 上的关系 $\boldsymbol{r}_q$;
9. 使用式 (25) 计算 Q_r 的损失函数 $L(Q_r)$;
10. 使用 $L(Q_r)$ 更新模型参数 ϕ;
11. end while

$$L=\sum_r\sum_{(\boldsymbol{h}_q,\boldsymbol{t}_q)\in Q_r}\sum_{(\boldsymbol{h}_q,\boldsymbol{t}_q^-)\in Q_r^-}[\gamma+S_q(\boldsymbol{h}_q,\boldsymbol{t}_q^-)-S_q(\boldsymbol{h}_q,\boldsymbol{t}_q)]^+ \tag{25}$$

式中: $[x]^+=\max(0,x)$ 为标准的合页损失函数; γ 为用来分隔正负查询集三元组的边际值。

2　实　验

2.1　实验设置

本节在从民间文学本文抽取到的 FolkKG 数据集上进行实验测试, 先按照 FKGC 的原则[15] 划分民间文学关系任务 T_r: 将三元组按照关系类型分组, 分组后按(7:2):1 的比例分为训练阶段任务集(训练集和验证集)及测试阶段任务集(测试集); 再从每个关系组中随机选取 K 个实体对(特定关系下的头尾实体)作为参考集 S_r, 剩余实体对作为查询集

Q_r的正样本，将正样本的尾实体随机替换为同类型其他实体来构造Q_r的负样本；最后，将T_r数据处理为FKGC数据集的规范格式。实验数据集的统计信息如表2所示。

表2 民间文学知识图谱数据集的统计信息

Table 2 Statistics of the FolkKG dataset

实体	关系	三元组	关系任务	训练集	验证集	测试集
1 907	48	6 106	40	28	8	4

选取GMatching[16]和FAAN[17]2个基于度量学习的FKGC方法及MetaR[19]、GANA[20]、RSCL[21]、HiRe[22]和NP-FKGC[23]5个基于元学习的FKGC方法作为对比模型。对比方法按照各自模型的最佳参数进行设置，在1-shot、3-shot和5-shot的场景下，分别进行3次实验并取各场景实验结果的平均值作为最终的补全结果。本文模型的嵌入维度d为95，Transformer编码层数为3，多头数目为4，实体邻域聚合层的最大邻域数为20。模型使用Adam优化器、基于Pytorch实现，初始学习率l_r和边际值分别设置为5×10^{-5}和5，将验证集上的MRR作为早停策略的标准。

准确率评价指标采用平均倒数排名MRR和Hits@n，对于Q_r上的正负实体对，MRR表示正确实体对排名倒数的平均值，Hits@n表示评分排在前n名的正确实体对的比例，MRR和Hits@n越大，说明实体排名越符合事实，补全结果越准确，计算公式如下：

$$\mathrm{MRR}=\frac{1}{|T|}\sum_{i=1}^{|T|}\frac{1}{r_i} \tag{26}$$

$$\mathrm{Hits@}n=\frac{1}{|T|}\Pi(r_i\leqslant n) \tag{27}$$

式中：$|T|$为候选尾实体的数量；r_i为Q_r实体对(h,t_i)评分的正序排名；$\Pi(\cdot)$为指示函数，条件为真时，函数值为1，反之，为0。

2.2 对比实验

本节比较了多个FKGC模型在K-shot(K=1, 3, 5)场景的表现，如表3所示。可以看出：

表3 对比实验结果

Table 3 Results of comparative experiments

模型类别	模型名称	MRR			Hits@10			Hits@5			Hits@1		
		1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
度量学习FKGC模型	GMatching[16]	0.193	—	—	0.350	—	—	0.272	—	—	0.116	—	—
	FAAN[17]	0.227	0.302	0.333	0.411	0.463	0.511	0.319	0.388	0.424	0.137	0.218	0.245
元学习FKGC模型	MetaR[19]	0.207	—	0.241	0.406	—	0.458	0.313	—	0.355	0.101	—	0.141
	GANA[20]	0.331	0.354	0.382	0.571	0.577	0.653	0.448	0.464	0.536	0.206	0.237	0.256
	RSCL[21]	0.338	—	0.395	0.516	—	0.532	0.412	—	0.490	0.215	—	0.271
	HiRe[22]	0.341	—	0.435	0.538	—	0.627	0.453	—	0.608	0.219	—	0.311
	NP-FKGC[23]	0.343	0.360	0.418	0.557	0.569	0.593	0.460	0.476	0.574	0.225	0.244	0.383
	FolkKGC	**0.358**	**0.392**	**0.463**	**0.595**	**0.614**	**0.694**	**0.498**	**0.506**	**0.646**	**0.240**	**0.272**	**0.394**

注：加粗数字表示最优结果；下划线数字表示次优结果；“—”表示该模型不支持当前K-shot场景。

1）FolkKGC与对比模型的最优结果相比，5-shot上的MRR提升2.8%，Hits@10提升4.1%，Hits@5提升3.8%，Hits@1提升1.1%；3-shot上的MRR提升3.2%，Hits@10提升3.7%，Hits@5提升3%，Hits@1提升2.8%；1-shot上的MRR提升1.5%，Hits@10提升2.4%，Hits@5提升3.8%，Hits@1提升1.5%。以上结果表明，FolkKGC在所有评价指标上均优于对比模型且达到最优。

2）FolkKGC优于GMatching和MetaR，说明融合关系邻域信息生成关系表示的方法优于不使用邻域信息；FolkKGC优于FAAN和GANA，说明对邻域信息进行分类处理后再学习优于直接学习；FolkKGC优于RSCL和HiRe，说明FolkKGC的邻域信息学习方法更为合理；FolkKGC优于NP-FKGC，说明FolkKG的关系模式比通用数据集简单，且本文生成民间文学关系表示的方法比NP-FKGC融合路径信息和关系分布的方法更有效。

上述结果表明，本文方法能针对民间文学文本的特点，有效利用邻域信息来挖掘实体对的隐含语义信息，使FolkKGC取得最优性能。

2.3 消融实验

为了验证FolkKGC中各模块的有效性，本文对各模块进行修改，并进行如下消融实验，实验结果如表4所示：

1）为了验证ENA层的有效性，移除ENA层得到w/o ENA，直接使用民间文学实体的预嵌入进行后续计算。GAT ENA使用图注意力(graph attention network, GAT)[25]代替ENA层。FolkKGC在1/3/5-shot上的MRR比w/o ENA高10.7%、11.5%和12.4%，验证了ENA层优化民间文学实体表示的有效

表 4　消融实验结果

Table 4　Results of ablation experiments

模型名称	模块			MRR			Hits@10			Hits@5			Hits@1		
	1	2	3	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
FolkKGC	√	√	√	**0.358**	**0.392**	**0.463**	**0.595**	**0.614**	**0.694**	**0.498**	**0.506**	**0.646**	**0.240**	**0.272**	**0.394**
w/o ENA	×	√	√	0.251	0.277	0.339	0.490	0.512	0.584	0.369	0.387	0.481	0.139	0.174	0.203
GAT ENA	≒	√	√	0.306	0.315	0.380	0.545	0.579	0.657	0.428	0.446	0.546	0.182	0.194	0.260
w/o RNL	√	×	√	0.222	0.273	0.394	0.474	0.495	0.612	0.328	0.370	0.501	0.125	0.131	0.229
GMatc SPL	√	√	≒	0.296	0.293	0.356	0.502	0.535	0.573	0.410	0.421	0.528	0.155	0.181	0.244
GANA SPL	√	√	≒	0.344	0.352	0.431	0.573	0.596	0.662	0.460	0.489	0.593	0.212	0.224	0.324
w/o ρ	√	√	≒	0.358	0.356	0.412	0.595	0.592	0.641	0.498	0.473	0.553	0.240	0.213	0.291

注: 加粗数字表示最优结果; 下划线数字表示次优结果; “—” 表示该模型不支持当前K-shot场景。

性; FolkKGC 在 1/3/5-shot 上的 MRR 比 GAT ENA 高 5.2%、7.7% 和 8.3%, 表明本文的实体邻域聚合方法优于 GAT, 且随着 K 的增加, 优势更显著。

2) 为了验证 RNL 层的有效性, 移除 RNL 层得到 w/o RNL, 直接使用经过 ENA 层的S_r实体对进行后续计算。FolkKGC 的 1/3/5-shot 补全结果分别比 w/o RNL 高 13.6%、11.9% 和 6.9%, 验证了 RNL 层优化S_r实体对表示的有效性。

3) 为了验证 SPL 的有效性, GMatc SPL 和 GANA SPL 分别使用 GMatching 和 GANA 模型的关系学习层代替 SPL, 分别代表基于度量学习和基于元学习的 FKGC 模型关系学习层的基准水平。移除 ρ 注意力函数得到 w/o ρ, 直接将 Transformer 学习到的S_r关系表示取平均作为Q_r上的关系表示。FolkKGC 在 1/3/5-shot 上的 MRR 分别比 GMatc SPL 高 6.2%、9.9% 和 10.7%, 验证了 SPL 关系表示方法的有效性。FolkKGC 在 1/3/5-shot 上的 MRR 比 GANA SPL 分别高 1.4%、4% 和 3.2%, 表明本文的关系表示方法比 GANA 更有效。FolkKGC 在 3/5-shot 上分别比 w/o ρ 高 3.6% 和 5.2%, 验证了基于相似度函数生成民间文学关系的有效性, 且随着 K 值的增大, 优势更显著。

2.4　预嵌入模型影响

在 FolkKG 数据集上对 4 个常用预嵌入模型 TransE[6]、TransH[7]、DistMult[11] 和 ComplEx[12] 做测试, 结果如表 5 所示, TransH 在 3-shot 和 5-shot 上效果均优于其他模型, 说明 TransH 是最适合本文 FolkKGC 任务的预嵌入模型。

表 5　预嵌入模型实验结果

Table 5　Results of pre-embedding models experiments

模型类别	模型名称	MRR			Hits@10			Hits@5			Hits@1		
		1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
预嵌入模型	TransE[6]	0.068	0.102	0.107	0.172	0.209	0.228	0.084	0.129	0.135	0.028	0.050	0.056
	TransH[7]	0.117	**0.175**	**0.229**	0.201	**0.368**	**0.372**	0.141	**0.267**	**0.285**	0.084	**0.103**	**0.115**
	DistMult[11]	0.113	0.139	0.150	0.208	0.216	0.236	0.133	0.172	0.190	0.073	0.101	0.099
	ComplEx[12]	**0.124**	0.168	0.187	**0.231**	0.274	0.292	**0.184**	0.211	0.212	**0.092**	0.133	0.105

2.5　参数影响

为确定 FolkKGC 实体和关系的最佳嵌入维度 d, 在民间文学数据集上测试 5-shot 补全的准确率, 结果如图 4 所示, FolkKGC 补全的准确率随着 d 的增大而上升, 当 d 为 95 时, 各项评价指标达到最大, d 大于 95 时, 各项指标不再提升, 反而整体呈下降趋势, d 为 95 时的 MRR 比 d 为 90 和 100 时分别高 6.4% 和 2%, Hits@n 分别高 7.3% 和 8.5%。因此, 选取 95 作为最佳嵌入维度。

为测试S_r容量 K 对模型性能的影响, 对 FolkKGC 和支持不同 K-shot 场景的 FKGC 对比模型进行 K-shot(K=1, 3, 5, 7, 9)补全结果对比, 如图 5 所示。可

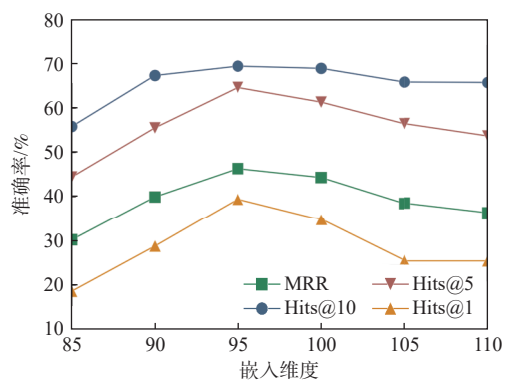


图 4　嵌入维度影响

Fig. 4　Impacts of embedding dimensions

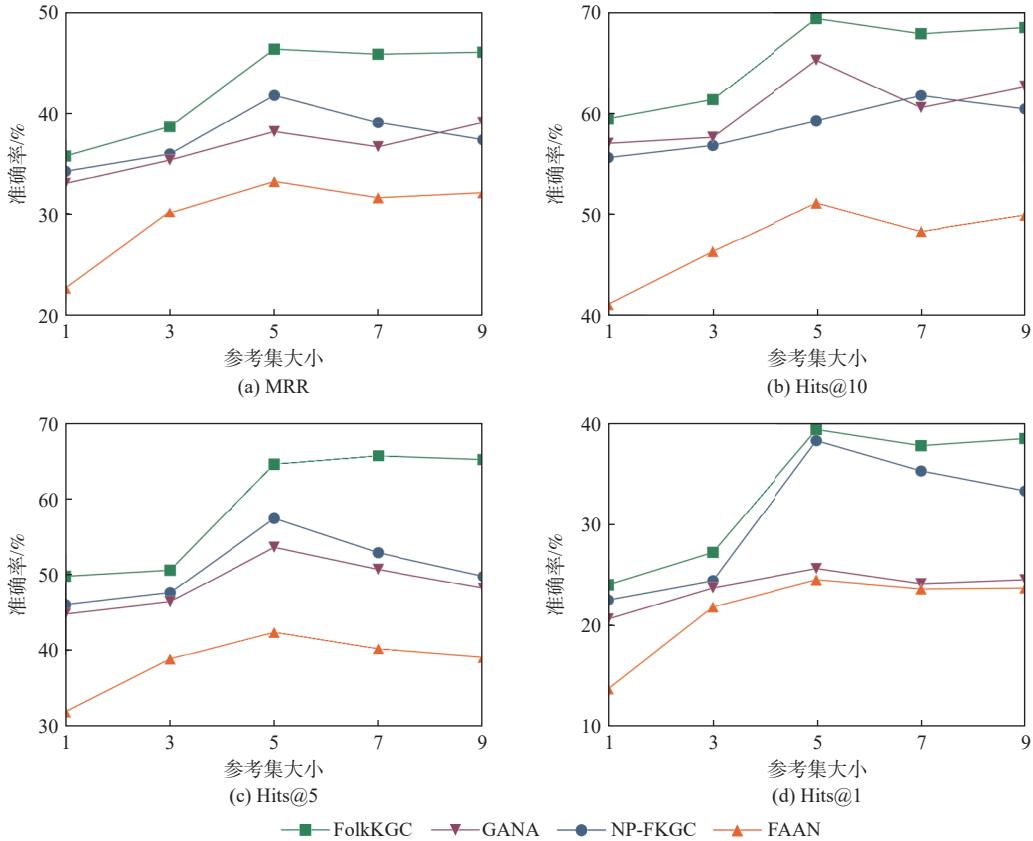


图5 参考集容量影响

Fig. 5 Impacts of support set's capacity

以看出：

1）随着 K 值从 1 增加至 9，FolkKGC 与对比模型的 MRR 和 Hits@n 均保持总体上升趋势，说明参考集实体对的增加可提供更多的关系信息，提升了补全的准确率。

2）K 值大于 5 时，FolkKGC 结果的准确率保持相对稳定，而对比模型出现不同程度的波动，说明当 K 较大时，FolkKG 中同类关系内部的语义差距在 S_r 实体对之间相互影响，FolkKGC 通过相似度函数保证了 Q_r 实体对关系表示的准确性。

3）FolkKGC 的MRR、Hits@10、Hits@5 和Hits@1，比其他模型的最优结果平均提升了 4.5%、4.4%、8.4% 和 2.6%，说明 FolkKGC 在 FolkKGC 任务中取得了最优性能。

3 结 论

本文针对民间文学知识图谱补全问题，结合民间文学数据的特点，提出了融合领域信息的民间文学知识图谱补全模型 FolkKGC，实验结果表明：

1）本文模型实现了较为优异的知识图谱补全能力，最高可达 46.3% 的 MRR 和 69.4% 的 Hits@10。与其他模型的最优结果相比，FolkKGC 的 MRR 和 Hits@10 分别提升了 2.5% 和 3.4%。

2）本文通过融合关系领域信息生成关系表示的方法优于不使用领域信息，同时 FolkKGC 先对领域信息分类处理再进行特征学习的方式也比直接学习更为有效，使得 FolkKGC 取得最优性能。

3）本文方法有效克服了民间文学领域知识库补全高度依赖专家知识、人工成本高等困难，也弥补了通用补全方法无法生成高质量实体关系表示的不足。 为使本文方法能够高效处理大规模的民间文学实体和关系，仍需要进一步优化 FolkKGC 模型的补全效率，并在更大规模的 KG 上进行测试。

参考文献（References）

[1] EDELMANN A, WOLFF T, MONTAGNE D, et al. Computational social science and sociology[J]. Annual Review of Sociology, 2020, 46(1): 61-81.

[2] 左玉堂. 中国民间文学大系·史诗·云南卷 (1)[M]. 北京: 中国文联出版社, 2022: 3-7.
ZUO Y T. Treasury of Chinese folk literature: Collection of epics. Yunnan volume (1)[M]. Beijing: China Federation of Literary and Art Publishing House, 2022: 3-7(in Chinese).

[3] JI S X, PAN S R, CAMBRIA E, et al. A survey on knowledge graphs: representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514.

[4] 官赛萍, 靳小龙, 贾岩涛, 等. 面向知识图谱的知识推理研究进展

[J]. 软件学报, 2018, 29(10): 2966-2994.
GUAN S P, JIN X L, JIA Y T, et al. Knowledge reasoning over knowledge graph: a survey[J]. Journal of Software, 2018, 29(10): 2966-2994(in Chinese).

[5] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 53(2): 247-261.
LIU Z Y, SUN M S, LIN Y K, et al. Knowledge representation learning: a review[J]. Journal of Computer Research and Development, 2016, 53(2): 247-261(in Chinese).

[6] ANTOINE B, NICOLAS U, ALBERT G, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 27th Conference on Neural Information Processing Systems. Stateline: NeurIPS, 2013: 2787-2795.

[7] WANG Z, ZHANG J W, FENG J L, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Reston: AIAA, 2014, 28(1): 1112-1119.

[8] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[J]. Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin: AAAI, 2015, 29(1): 2181-2187.

[9] SUN Z Q, DENG Z H, NIE J Y, et al. Rotate: knowledge graph embedding by relational rotation in complex space[C]//Proceedings of the 7th International Conference on Learning Representations. New Orleans: Ithaca, 2019: 1-18.

[10] NICKEL M, TRESP V, KRIEGEL H. A three-way model for collective learning on multi-relational data[C]//Proceedings of the 28th International Conference on Machine Learning. Williamstown: ACM, 2011: 809-816.

[11] YANG B S, YIH W T, HE X D, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//International Conference on Learning Representations. San Diego: ICLR, 2015.

[12] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33rd International Conference on Machine Learning. New York: ACM, 2016: 2071-2080.

[13] 唐榕氚, 徐秋程, 汤闻易, 等. 无对齐实体场景的多语言知识图谱补全[J]. 北京航空航天大学学报, 2026, 52(1): 252-259.
TANG R C, XU Q C, TANG W Y, et al. Multilingual knowledge graph completion without aligned entity pairs[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(1): 252-259(in Chinese).

[14] 纪鑫, 武同心, 杨智伟, 等. 基于时序知识图谱的电力设备缺陷预测[J]. 北京航空航天大学学报, 2024, 50(10): 3131-3138.
JI X, WU T X, YANG Z W, et al. Power equipment defect prediction based on temporary knowledge graph[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(10): 3131-3138(in Chinese).

[15] MA H, DAISY W. A survey on few-shot knowledge graph completion with structural and commonsense knowledge[J]. (2023-01-03)[2023-09-07]. https://doi.org/10.48550/arXiv.2301.01172.

[16] XIONG W H, YU M, CHANG S Y, et al. One-shot relational learning for knowledge graphs[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2018: 1980-1990.

[17] SHENG J W, GUO S, CHEN Z Y, et al. Adaptive attentional network for few-shot knowledge graph completion[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2020: 1681-1691.

[18] VASWANI A, SHAZEER N, PARMER N, et al. Attention is all you need[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach: NeurIPS, 2017: 6000-6010.

[19] CHEN M Y, ZHANG W, ZHANG W, et al. Meta relational learning for few-shot link prediction in knowledge graphs[C]//Proceedings of the 2019 Conference on Empirical Methods Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2019: 4216-4225.

[20] NIU G L, LI Y, TANG C G, et al. Relational learning with gated and attentive neighbor aggregator for few-shot knowledge graph completion[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 213-222.

[21] LI Y, YU K, ZHANG Y, et al. Learning relation-specific representations for few-shot knowledge graph completion[J]. (2023-03-22)[2023-09-20]. https://doi.org/10.48550/arXiv.2203.11639.

[22] WU H, YIN J, RAJARATNAM B, et al. Hierarchical relational learning for few-shot knowledge graph completion[C]//Proceedings of the 11th International Conference on Learning Representations. Virtual: Ithaca, 2022: 1-10.

[23] LUO L H, LI Y F, HAFFARI G, et al. Normalizing flow-based neural process for few-shot knowledge graph completion[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2023: 900-910.

[24] CHUANG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[C]//Proceedings of the Conference on Neural Information Processing Systems. Montréal: NeurIPS, 2014.

[25] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of the 6th International Conference on Learning Representations. Ithaca: ICLR, 2018: 1-12.

Folklore knowledge graph completion by fusing neighborhood information

GAO Fangzheng[1, 2], WANG Jiahui[1, 2], DUAN Liang[1, 2, *], YUE Kun[1, 2], LI Zhongbin[1, 2]

(1. School of Information Science and Engineering, Yunnan University, Kunming 650500, China;
2. Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming 650500, China)

Abstract: Folklore knowledge graph (KG), the domain knowledge base, has received a lot of interest and has the potential to be widely employed in context understanding and knowledge service scenarios. However, folk literature works often use short sentences and just elaborate on partial relations between characters and things in detail. The KG does not fully cover all the real domain knowledge, and it is necessary to complete. Existing KG completion (KGC) methods are hard to differentiate various neighborhood information and semantic gaps within relations. This paper proposes the folklore KG completion model FolkKGC by fusing the neighborhood information. First, the relation-aware gate attention mechanism is designed for neighborhood information fusion to effectively represent the folklore representation of entities. The relevant fine-grained folklore representations are then produced by fusing the neighborhood information of relations using a relation learner based on cross-neighborhood similarity. Comparison and ablation experiments are conducted on the datasets extracted from folklore texts. The results show that our method outperforms the state-of-the-art models on MRR and Hits@*n*, which verifies the effectiveness of our proposed model.

Keywords: folklore knowledge graph; knowledge graph completion; few-shot learning; neighborhood information; gated recurrent unit

Received: 2023-12-04; **Accepted**: 2024-04-19; **Published Online**: 2024-05-07 14:05

URL: link.cnki.net/urlid/11.2625.V.20240506.1326.001

Foundation items: Basic Research Project of Yunnan Province (202301AT070193); Research Foundation of Yunnan Education Department (2022Y010)

* **Corresponding author**. E-mail: duanl@ynu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0837

城市环境改进点云配准的 LiDAR 辅助 GNSS/IMU 定位算法

黄明明，孙蕊*
(南京航空航天大学 民航学院，南京 211106)

摘　　要：自动驾驶技术飞速发展，对城市环境定位精度的需求不断提升，但由于城市环境的复杂性，单纯的全球卫星导航系统 (GNSS) 和惯性测量单元 (IMU) 组合无法满足该需求。激光雷达 (LiDAR) 可实时感知周围环境，近年来成本不断降低，由于 GNSS 和 LiDAR 良好的互补特性，GNSS/IMU/LiDAR 组合被广泛地研究。针对目前 GNSS/IMV/LiDAR 组合定位精度无法满足需求的问题，提出一种改进点云配准的 LiDAR 辅助 GNSS/IMU 定位方法。其中，算法构建了 LiDAR 配准时新增误差项和局部点云来实现 GNSS 质量控制，设计双差特征量综合判断卫星信号，实现不同信号质量的 GNSS 信号定权。实测实验表明，在城市环境中，所提算法对比传统 GNSS/IMU 紧组合算法，水平位置精度提高 49.33%，三维位置精度提高 48.31%，对比传统 GNSS/IMU/LiDAR 组合算法，水平、三维位置精度分别提高 40.33%、37.60%。

关　键　词：GNSS/IMU/LiDAR 组合导航；改进点云配准；局部点云地图；双差特征量
中图分类号：V221^{+}.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2026)03-0945-10

汽车行业发展至今，各大车企都纷纷聚焦于无人车技术。无人驾驶技术可用于无人巡逻车、无人快递运输车，解放人力，降低运营成本。准确可靠的导航定位信息是实现无人车运行的必要条件。全球卫星导航系统(global navigation satellite system, GNSS)和惯性测量单元(intertial measurement unit, IMU)组合作为普遍使用的导航定位技术，导航精度在城市环境中受到 GNSS 的严重制约。其中，GNSS 的实时动态测量技术(real time kinematic, RTK)在卫星信号良好时，组合导航系统可以达到厘米级定位精度，但在复杂的城市环境运行区域，卫星信号会被高层建筑遮挡，观测值中包含非直射信号和多路径误差，导致最终定位精度恶化到几十甚至上百米。

近年来，LiDAR 的即时定位与地图构建(simultaneous localization and mapping, SLAM)技术飞速发展。LiDAR 通过发射激光点，获得物体反射回来的信号，形成点云数据，进而求取定位结果，可以全天候使用。其对丰富特征获取的需求，使其适合在城市峡谷环境中运行。GNSS 和激光雷达(light detection and ranging, LiDAR)有很强的互补性，因此，为了提高组合导航的精度和可靠性，将 GNSS/IMU/LiDAR 融合具有明显优势。

当前，LiDAR 与其他传感器的组合已被深入研究，如与惯性传感器的组合[1-2]、惯性导航系统 (inertial navigation system, INS) 和相机的组合[3-4]、GPS 的组合[5-6]、INS 和 GPS 的组合[7-8]。在大多数 INS/LiDAR 组合工作中，LiDAR 与 INS 松散耦合，其中，

收稿日期：2023-12-29；**录用日期**：2024-04-05；**网络出版时间**：2024-10-21 09:40
网络出版地址：link.cnki.net/urlid/11.2625.V.20241018.1419.001
基金项目：国家自然科学基金 (42222401,42174025,41974033)；工信部专项科研项目 (TC220A04A-79)；江苏省“六大人才高峰”项目 (KTHY-014)；江苏省自然科学基金 (BK20211569)；中央高校基本科研业务费专项资金 (xcxjh20230720)
* **通信作者**. E-mail：rui.sun@nuaa.edu.cn

引用格式：*黄明明，孙蕊. 城市环境改进点云配准的 LiDAR 辅助 GNSS/IMU 定位算法 [J]. 北京航空航天大学学报，2026，52（3）：945-954.*
HUANG M M，SUN R. LiDAR aided GNSS/IMU positioning algorithm based on improved point cloud registration in urban environment [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：945-954 （in Chinese）.

激光雷达和INS的姿态估计分别通过卡尔曼滤波(Kalman filtering, KF)或部分滤波(partial filtering, PF)融合。而在文献[9-10]中,实现了LiDAR和INS的紧密集成。其中,在LiDAR导航算法中,扫描匹配是LiDAR获取精准相对定位的重要步骤,扫描匹配质量好坏直接决定最终的定位结果。

扫描匹配方法可分为基于单个点、特征、点云整体数学属性3类。基于特征的扫描匹配通过提取2帧点云间的相同线段、平面、车道标记或曲率函数等,能反映环境中的几何特征,拟合线有助于解决点对点问题。基于点的扫描匹配原理简单,通过比较2帧点云中特征点的属性,如曲率等,可直接处理原始点云中的点。当前对点的扫描匹配方法的研究主要有:Garulli等使用二维激光测距仪提供的距离扫描来更新环境地图,通过基于线特征的环境表示,构建不断增加新特征点的特征线,通过计算线之间的距离得到残差[11]。Yokozuka等利用校正后的点云数据构建全景局部地图,利用该地图将当前帧的点云数据与之进行配准,以获得更精确的位置信息。用改进迭代最近点算法(iterative closest point, ICP)算法计算帧间变换关系,采用对称相对熵散度,缓解LiDAR测程中的轨迹误差积累[12]。雷道竖等将室内周边环境面积表征为极坐标系下测量角度和距离的函数形式,以此描述LiDAR在不同位置上的距离测量分布。考虑到室内环境中广泛存在的线段特性,通过线段检测的方式获取扫描数据中的若干线段特征[13]。廖自威等结合室内结构化环境特点,提出一种完备端点定义与提取方法,将直线段特征与完备端点进行关联,优化几何特征扫描匹配过程,采用基于直线拟合认知的姿态角加权几何平均求解方法确保姿态角收敛[14]。谢勃等提出基于LiDAR点云的建筑物激光扫描重构方法,将LiDAR获取的点云数据做预处理,其中包括数据配准和坐标转换,提取出点云的特征点,根据该特征点获取建筑物的基本轮廓[15]。Chen等从三维激光扫描中提取特征(例如线条和角点),用于确定位置和航向[16]。现有研究集中于特征点和特征线的选取,对扫描匹配的距离误差建模较为粗糙,不够细致。

在LiDAR与GNSS融合的过程中,用LiDAR建立局部地图可以辅助GNSS信号接收类型的判定,有效提高定位精度,目前已经被很多团队所研究。Wen等使用LiDAR获得车辆周围双层巴士的点云图,采用欧几里得聚类算法和分类方法计算车辆相对于巴士的位置,剔除被巴士遮挡的卫星信号,在车站周围做了静态实验,没有进行动态实验[17]。Hsu等通过使用3D建筑模型、射线跟踪模拟,并结合接收机位置,根据伪距量测值检测非视距(non-line-of-sight reception, NLOS)信号,发现NLOS信号与高度角高度相关。进一步使用高度角和接收机与建筑物之间的距离来反映NLOS情况。需要提前准备好3D建筑模型,但模型精度较差时,定位结构受影响较大[18]。Chen等研究了一种利用移动激光扫描点云数据生成GNSS掩模的解决方案。通过去除噪声点,确定衰减卫星信号,提取最高的阻塞点,并计算GNSS阴影掩模的角度。实验结果表明,点云数据可用来提取GNSS阴影掩模,相比传统3D模型,能更准确描述GNSS信号覆盖边界。但只能提取LiDAR探测到的建筑物部分,无法缓解建筑物高层带来的影响[19]。Wen在车辆上搭载3D LiDAR传感器,利用点云生成围绕车辆的滑动窗口地图。利用一种快速搜索方法,在实时滑动窗口地图上检测NLOS接收情况,将修正后的伪距测量值和正常的伪距测量值与INS紧密集成,采用因子图进行优化。该方法通过生成滑动窗口地图,当街道宽度为10 m时,探测的高度角能达到76°。但当街道宽度增加时,探测的高度角显著变小[20]。现有的LiDAR辅助GNSS信号接收类型判定的方法存在局部点云地图建图不全的问题,会导致信号类型误判。

综上所述,目前关于GNSS/IMU/LiDAR组合导航的研究仍有一些问题有待解决。在大多数研究中,对LiDAR配准中的误差建模考虑不够全面。此外,目前,LiDAR辅助GNSS判断信号质量只单独使用LiDAR点云,在无法建模高层建筑影响时,性能下降。本文提出一种基于改进点云配准的LiDAR辅助GNSS/IMU组合导航定位模型。主要贡献如下:

1)提出一种在LiDAR配准过程中,自适应地根据距离大小实现线特征残差修正的方法,从而减小2帧点云之间存在微小真实特征不匹配而导致的旋转和平移误差。

2)提出一种基于局部地图和构建的判断量组合的GNSS伪距测量值评估模型,以减小建筑物高层LiDAR建模不充分引起的GNSS卫星信号类型误判的影响,进而提高组合导航定位精度。

1 算法框架

所提出的LiDAR辅助的基于紧组合扩展卡尔曼滤波器(extended kalman filter, EKF)方案的GNSS/IMU融合算法如图1所示。

该算法包括3个步骤:①LiDAR根据获得的点

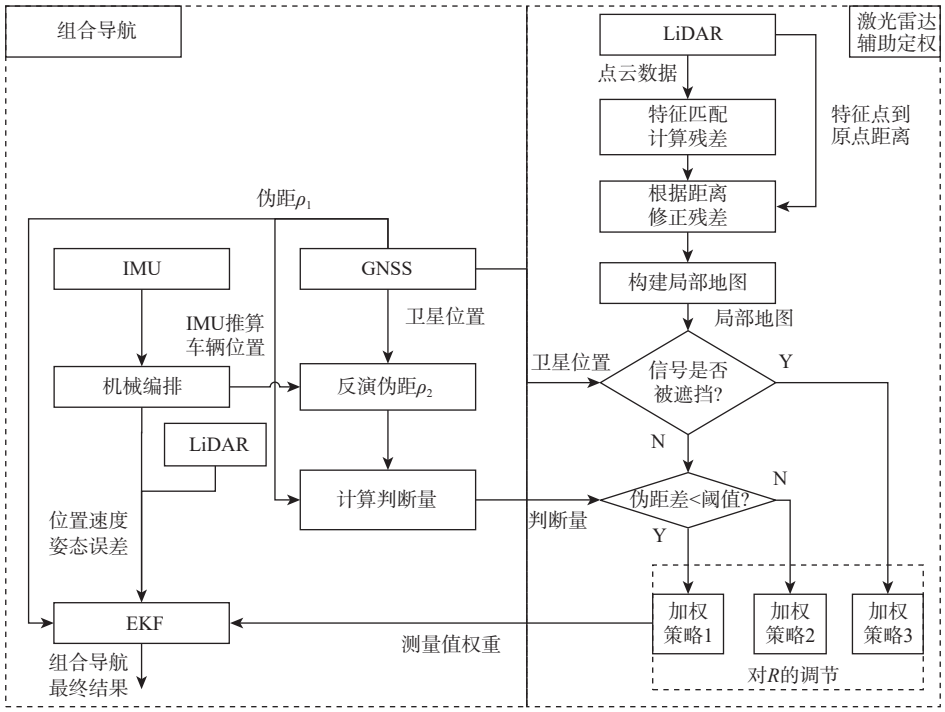


图 1　算法框图

Fig. 1　Scheme of the proposed algorithm

云数据, 通过特征点和特征线、特征面的匹配, 获得匹配对, 使用与距离相关的残差改正量, 修改特征对之间的残差。LiDAR 使用滑动窗口帧构建局部地图。②使用 IMU 一步预测的载体位置和已知卫星位置反演伪距, 与真实测量值作差, 求出伪距差, 结合是否被局部点云地图遮挡和伪距差的值, 卫星观测值被确定为不同的状态: 误差超过阈值的 LOS 信号、误差小于阈值的 LOS 信号、NLOS 信号, 从而赋予当前卫星的观测值不同的权重。如果局部地图显示未遮挡, 伪距差却很大, 认定是由于局部地图构造不全引起的误判, 是 NLOS 信号, 采取加权策略 2。③IMU 量测值和 GNSS 伪距、伪距率进行拓展 KF 紧组合, 输出最终导航信息。

2　LiDAR 局部地图构建

LiDAR 原始点云经过按线数分割, 计算曲率, 删除异常点, 按曲率大小筛选特征点后进行帧间匹配。通过激光雷达第 t_k 帧和第 t_{k+1} 帧的边缘点特征和平面点特征之间的残差和对 2 帧点云进行配准。

边缘特征点到线的距离计算公式为

$$d_{\mathrm{e}}=\frac{\left|\left(\boldsymbol{p}_{(k+1,i)}-\boldsymbol{p}_{(k,a)}\right)\left(\boldsymbol{p}_{(k+1,i)}-\boldsymbol{p}_{(k,b)}\right)\right|}{\left|\boldsymbol{p}_{(k,a)}-\boldsymbol{p}_{(k,b)}\right|} \tag{1}$$

平面特征点到面的距离计算公式为

$$d_{\mathrm{plane}}=\frac{\left|\left(\boldsymbol{p}_{(k+1,i)}-\boldsymbol{p}_{(k,a)}\right)\left(\left(\boldsymbol{p}_{(k,a)}-\boldsymbol{p}_{(k,b)}\right)\left(\boldsymbol{p}_{(k,a)}-\boldsymbol{p}_{(k,c)}\right)\right)\right|}{\left|\left(\boldsymbol{p}_{(k,a)}-\boldsymbol{p}_{(k,b)}\right)\left(\boldsymbol{p}_{(k,a)}-\boldsymbol{p}_{(k,c)}\right)\right|} \tag{2}$$

式中: $\boldsymbol{p}_{(k+1,i)}$ 为 t_k 时刻在 t_{k+1} 时刻对应的第 i 个特征点; $\boldsymbol{p}_{(k,a)}$、$\boldsymbol{p}_{(k,b)}$、$\boldsymbol{p}_{(k,c)}$ 为 t_{k+1} 时刻的特征点在 t_k 时刻对应的特征点中距离最近的 3 个点。

进行激光雷达点云匹配的前提假设是 2 帧点云的边缘特征点在现实世界中表示了同一条边, 实际上由于水平分辨率的存在, 从 2 帧中提取到的边缘特征点事实上有可能不属于同一边缘线, 而是相差了一个小于水平分辨率的距离, 在其各自的点云帧中, 其都是曲率最大的边缘点, 示意图如图 2 所示。

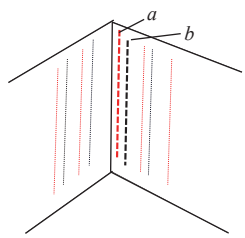


图 2　2 帧特征线不完全重合示意图

Fig. 2　Schematic diagram showing incomplete overlap of two feature lines

这种误差与匹配对距离 LiDAR 坐标原点的距离呈线性相关。使用特征点的位置计算距离, 当 LiDAR 点云帧之间的匹配对距离 LiDAR 坐标原点的距离为 d_{is} 时, 线特征对之间的残差修正为

$$f=d_{\mathrm{e}}+kd_{\mathrm{is}}\sin\theta\sin\varphi \tag{3}$$

式中: k 为常数, 取值范围为 0～1, 为 2 帧点云扫描线之间的微小偏移, 为一个经验值; d_{is} 为特征点到

原点的距离；θ为水平分辨率；φ为特征线与XOY平面的夹角。

使用滑动窗口的方式构建局部地图，不需要构建全局的点云地图，因为判断卫星信号是否被遮挡只需要当前环境的点云，与相隔过远的点云无关，减少了计算量。采用激光雷达测距与测绘（lidar odometry and mapping, LOAM）的建图方法，具体见文献[15]。点云的坐标系为LiDAR原点的坐标系，为了减少计算量，将卫星位置同样转换到LiDAR坐标系中。只使用当前时刻t和前$t-\Delta t$时刻的点云构建局部地图。Δt的大小与车辆的运动速度有关：

$$\Delta tv = 2\sqrt{(L\cos\beta)^2 - a^2} \tag{4}$$

式中：L为LiDAR的量程；a为车辆到建筑物的距离；β为LiDAR的最大高度角。

示意图如图3所示。

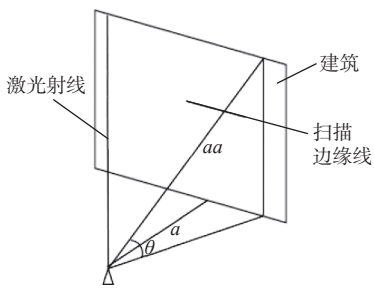


图3 LiDAR射线扫描墙壁示意图

Fig. 3 Diagram of LiDAR ray scan wall

3 局部地图和伪距差共同确定权重

GNSS/IMU松组合需要4颗卫星才能工作，GNSS/IMU紧组合使用的是GNSS原始的观测量、伪距和伪距率，可以在卫星数不足4颗的条件下工作，更符合城市峡谷环境。本文采用GNSS/IMU紧组合方式。

GNSS接收机接收卫星伪距测量值，第i颗卫星在地球坐标系中的位置是(x_s^i, y_s^i, z_s^i)，惯导给出的地球坐标系中的载体位置为$(x_{\text{imu}}, y_{\text{imu}}, z_{\text{imu}})$，反演得到的载体到第$i$颗卫星的伪距为

$$\rho_{\text{imu}}^i = \sqrt{\left(x_{\text{imu}} - x_s^i\right)^2 + \left(y_{\text{imu}} - y_s^i\right)^2 + \left(z_{\text{imu}} - z_s^i\right)^2} \quad i = 1,2,3,\cdots \tag{5}$$

设载体在地球坐标系中的真实位置为(x,y,z)，将式(5)在(x,y,z)处使用泰勒级数展开，IMU数据反演得到的伪距量测值ρ_{imu}^i为

$$\rho_{\text{imu}}^i = \rho_i + e_{i1}\delta x + e_{i2}\delta y + e_{i3}\delta z \tag{6}$$

式中：$\rho_i = \left[\left(x - x_s^i\right)^2 + \left(y - y_s^i\right)^2 + \left(z - z_s^i\right)^2\right]^{\frac{1}{2}}$，为真实伪距。

$$\begin{cases} e_{i1} = \dfrac{\partial \rho_{\text{imu}}^i}{\partial x} = \dfrac{x - x_s^i}{\rho_i} \\ e_{i2} = \dfrac{\partial \rho_{\text{imu}}^i}{\partial y} = \dfrac{y - y_s^i}{\rho_i} \\ e_{i3} = \dfrac{\partial \rho_{\text{imu}}^i}{\partial z} = \dfrac{z - z_s^i}{\rho_i} \end{cases} \tag{7}$$

GNSS接收机测量得到的与第i颗卫星之间的伪距可以表示为

$$\rho_{\text{g}}^i = \rho_i - cV_{t_{\text{R}}} + cV_{t_s}^i - v_i + \delta\rho_{\text{mul}}^i \tag{8}$$

式中：ρ_i为真实伪距；$V_{t_{\text{R}}}$为接收机钟差，$V_{t_s}^i$为卫星钟差，可根据卫星导航电文中给出的钟参数求得；v_i为大气延迟误差，由对流层延迟误差、电离层延迟误差等组成，经过模型修正后可大大减小，$\delta\rho_{\text{mul}}^i$为多路径效应及非视距接收引起的误差，为正值。

将2种方式得到的伪距相减，可得

$$\rho_{\text{g}}^i - \rho_{\text{imu}}^i - cV_{t_s}^i + v_i = -(e_{i1}\delta x + e_{i2}\delta y + e_{i3}\delta z) - cV_{t_{\text{R}}} + \delta\rho_{\text{mul}}^i \tag{9}$$

式中：ρ_{g}^i为GNSS接收机直接量测值，ρ_{imu}^i为IMU反演的伪距。

卫星钟差、大气延迟误差均已知。设T为判断量，则卫星i,j的判断量T可表示为

$$T_i = \rho_{\text{g}}^i - \rho_{\text{imu}}^i - cV_{t_s}^i + v_i = -(e_{i1}\delta x + e_{i2}\delta y + e_{i3}\delta z) - cV_{t_{\text{R}}} + \delta\rho_{\text{mul}}^i \tag{10}$$

$$T_j = \rho_{\text{g}}^j - \rho_{\text{imu}}^j - cV_{t_s}^j + v_j = -\left(e_{j1}\delta x + e_{j2}\delta y + e_{j3}\delta z\right) - cV_{t_{\text{R}}} + \delta\rho_{\text{mul}}^j \tag{11}$$

如果IMU推算的起始位置准确，则在0.1 s的卫星信号间隔内，IMU推算得到的位置的累计误差不大，位置较为准确，反演的伪距ρ_{imu}接近真实车辆卫星之间的距离，$-(e_{i1}\delta x + e_{i2}\delta y + e_{i3}\delta z)$可忽略。同一时刻的伪距观测值、接收机的钟差相等，用T_j减去T_i得

$$T_j - T_i = \delta\rho_{\text{mul}}^j - \delta\rho_{\text{mul}}^i \tag{12}$$

而同一时刻的GNSS信号，对于未被城市峡谷高楼遮挡的直射信号，由非视距接收和多路径效应引起的误差较小，如果卫星信号被高楼遮挡，形成非直射信号，则由非视距接收和多路径效应引起的误差较大。当$T_j - T_i > 0$时，$\delta\rho_{\text{mul}}^j - \delta\rho_{\text{mul}}^i > 0$，说明卫星信号受到城市峡谷的影响比$i$和$j$卫星信号更严重，卫星信号的质量更差，在计算位置时，权重应更低。

利用局部点云地图、IMU推算的载体位置和卫星位置，可以判断出当前时刻卫星是否被局部地图遮挡，判断方法详见文献[5]，被局部地图遮挡记为1，未被点云遮挡记为0，结果为

$$w_{dy}=\{w_{dy1},w_{dy2},\cdots,w_{dyn}\}\quad w_{dyj}=0,1\quad j=1,2,\cdots,n \tag{13}$$

在扩展 KF 中, R表示测量值的随机误差, 某一颗卫星的测量值对应的R越小, 表示这颗卫星的测量值越可靠。T越小, 说明在当前时刻观测到的所有卫星中, 对应的卫星受城市峡谷环境影响带来的误差更小, 观测量越可靠, 对应的R应越小。

设置对应的权重, 同一时刻的所有卫星的判断量序列为$\{T_1\quad T_2\quad\cdots\quad T_n\}$, 其中, 最小值为$T_{\min}$, 最大值为$T_{\max}$。设$T_{\max}$对应的卫星权重为$w_0$, 通过实验确定, 则其他卫星对应的权重为

$$w_n=\frac{T_n-T_{\min}}{T_{\max}-T_{\min}}\cdot w_0 \tag{14}$$

卫星高度角和信噪比在一定程度上能反应伪距信号的质量。将信噪比和卫星高度角计入考虑, 引用文献 [20] 提出的加权模型。加权函数为

$$W(\varepsilon,S)=\frac{1}{\sin^2\varepsilon}\left(10^{-\frac{(S-T)}{b}}\left(\left(\frac{A}{10^{-\frac{(F-T)}{b}}}-1\right)\cdot\frac{(S-T)}{F-T}+1\right)\right) \tag{15}$$

式中: ε为高度角; S为信噪比; 参数T为信噪比阈值; 参数b、A、F通过实验确定。

结合点云地图和伪距差, 最终的综合判断权值策略为

$$W=\begin{cases}W(\varepsilon,\ S) & w_{dy}=1\\ w_nW(\varepsilon,\ S) & w_{dy}=0\end{cases} \tag{16}$$

最终的量测值噪声矩阵为

$$\boldsymbol{R}=R_0\boldsymbol{W} \tag{17}$$

式中: R_0为初始量测噪声值; $\boldsymbol{W}$是由 16 式构造的对角线矩阵。IMU 的测量结果随时间发散, LiDAR 可以辅助 IMU 获得精度更高的相对定位结果, 从而与 GNSS 绝对定位结果融合, 达到更好的定位效果, 本文采用迭代误差状态 KF 的 LiDAR/IMU 紧组合, 文献 [21] 给出了状态量与观测量的转换矩阵。边缘点特征观测方程对状态的转换矩阵为

$$\boldsymbol{H}_{l_k}^{\mathrm{e}}=\frac{\left[\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_a^{l_k}\right)\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_b^{l_k}\right)\right]^{\mathrm{T}}\left(\boldsymbol{p}_b^{l_k}-\boldsymbol{p}_a^{l_k}\right)_{\times}}{\left|\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_a^{l_k}\right)\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_b^{l_k}\right)\right|\left|\boldsymbol{p}_a^{l_k}-\boldsymbol{p}_b^{l_k}\right|}\cdot\frac{\partial\hat{\boldsymbol{p}}_i^{l_k}}{\partial\delta\boldsymbol{x}} \tag{18}$$

平面点特征观测方程对状态的转换矩阵为

$$\boldsymbol{H}_{l_k}^{\mathrm{p}}=\frac{\left[\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_a^{l_k}\right)\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_b^{l_k}\right)\right]^{\mathrm{T}}}{\left|\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_a^{l_k}\right)\left(\hat{\boldsymbol{p}}_i^{l_k}-\boldsymbol{p}_b^{l_k}\right)\right|}\cdot\frac{\partial\hat{\boldsymbol{p}}_i^{l_k}}{\partial\delta\boldsymbol{x}} \tag{19}$$

式中: $\hat{\boldsymbol{p}}_i^{l_k}$为$t_k$时刻第$i$个点$\boldsymbol{p}_i^{l_k}$在$t_{k+1}$时刻对应的点; $\boldsymbol{p}_a^{l_k}$和$\boldsymbol{p}_b^{l_k}$为$\boldsymbol{p}_i^{l_k}$最接近的 2 个点; $\delta\boldsymbol{x}$为误差状态向量。LiDAR/IMU 相对定位系统输出的位置姿态变化与 GNSS 观测量进行紧组合拓展 KF 输出最终定位结果。拓展 KF 状态更新公式为

$$\boldsymbol{K}_{k+1}=\boldsymbol{P}_{k+1,k}\boldsymbol{H}_{k+1}^{\mathrm{T}}\left(\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1,k}\boldsymbol{H}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1}\right)^{-1} \tag{20}$$

$$\boldsymbol{P}_{k+1}=(\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})\boldsymbol{P}_{k+1,k}(\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})^{\mathrm{T}}+\boldsymbol{K}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{K}_{k+1}^{\mathrm{T}} \tag{21}$$

$$\hat{\boldsymbol{X}}_{k+1}=\hat{\boldsymbol{X}}_{k+1,k}+\boldsymbol{K}_{k+1}\left(\boldsymbol{Z}_{k+1}-\boldsymbol{H}_{k+1}\hat{\boldsymbol{X}}_{k+1,k}\right) \tag{22}$$

4　实验结果与分析

为了验证本文算法在城市峡谷环境中的有效性, 在南京南站附近的城区进行车载实验。GNSS 和 IMU 设备为: 采样率为 10 Hz、型号为北斗星通 C200-AT 的 GNSS 接收机; 采样率为 125 Hz、型号为 STIM-300 的微机电系统(micro-electro-mechanical systems, MEMS)IMU, 输出比力和角速度, 其中, 加速度计的零偏为 0.05 mg, 随机游走为 0.06 $(\mathrm{m\cdot s^{-1}})/\sqrt{h}$; 陀螺仪的零偏为 0.5(°)/h, 随机游走为 0.15(°)/$\sqrt{h}$; 采用 Velodyne-16 线激光雷达采集点云数据, 采样频率为 10 Hz, 垂直分辨率为 2°, 水平分辨率为 0.2°。惯性/卫星组合导航设备的数据利用 Inertial Explore(IE)后处理计算得到, 采样率为 250 Hz, 可达 cm 级精度。通过实验确定的参数如表 1 所示。

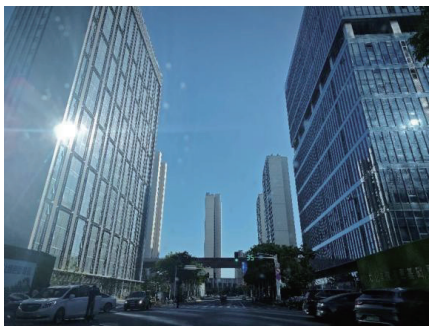
表 1　使用的参数

Table 1　Parameter values used

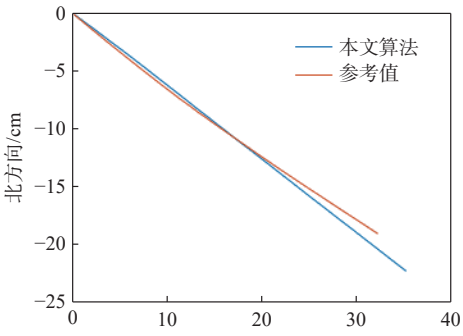
w_0	T	b	A	F
0.6	10	30	32	10

取总距离为 37 m 的直线路段 LiDAR 点云进行对比实验, 实验环境与轨迹对比如图 4 所示, 其中, 轨迹图的参考坐标系为地心地固坐标系。实验结果统计分析如表 2 所示, 参考坐标系为 LiDAR 初始坐标系, 改进的残差修正算法与参考轨迹相比, 总距离误差从 5.27 m 降低到 4.41 m, 误差提高了 16.30%。

实验场景 1 是典型的城市峡谷环境, 接收到的卫星信号较少, 卫星信号会被建筑物遮挡。图 5 展示了数据 1 实验过程中接收到的 GPS 和 BDS 卫星数。传统的 GNSS/IMU 紧组合的量测值平权, 误差较大的伪距测量值严重减弱定位精度。通过局部地图和判断量综合判断卫星信号类型, 对误差较大的测量值降权, 能减弱其带来的影响。局部地图判断卫星类型示意图如图 6 所示。通过判断载体位置和卫星位置的连线是否穿过点云地图, 判断卫星信号的类型。穿过点云地图的卫星信号视为被遮



(a) 实验环境



(b) 轨迹对比

图 4 实验环境与轨迹对比

Fig. 4 Experimental environment and trajectory comparison

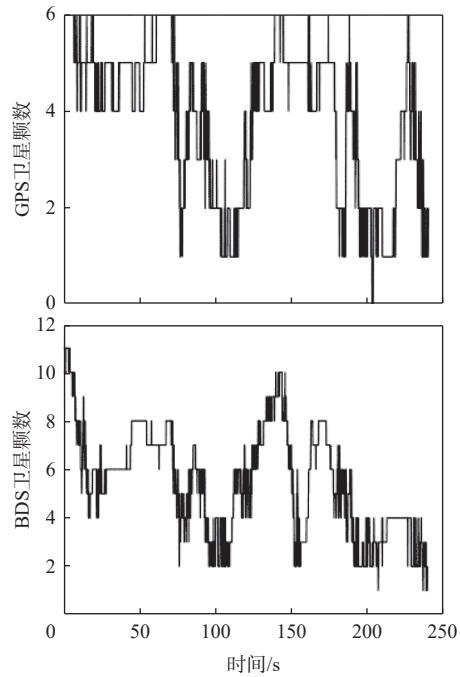


图 5 场景 1 卫星数量

Fig. 5 Number of satellites of scenario 1

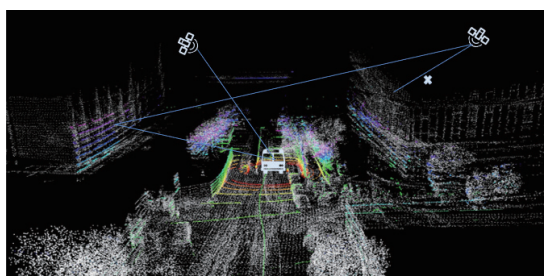


图 6 局部地图判断卫星类型示意图

Fig. 6 Local map with determining satellite type

表 2 改进匹配残差算法与原算法的运动估计误差对比

Table 2 Comparison of motion estimation errors between the improved residual matching algorithm and the original algorithm

算法	东向位移/m	北向位移/m	天向位移/m	总距离/m	误差/m
参考轨迹	−0.480 4	36.816 1	0.025 8	36.97	
原始匹配残差算法	−0.110 2	42.244 4	0.034 2	42.24	5.27
改进匹配残差算法	−0.106 5	41.382 8	0.040 5	41.38	4.41

挡，进行降权。没有穿过局部地图的卫星信号，进一步进行伪距差分析。

本文算法为 LiDAR 辅助的 GNSS/IMU 组合导航算法，使用局部地图辅助 GNSS 卫星信号进行质量判断，与基于伪距伪距率的 GNSS/IMU 紧组合算法和 IMU 与 LiDAR 相对定位结果先进行组合，再与卫星信号结合的传统 GNSS/IMU/LiDAR 算法进行对比。性能比较如表 3、表 4 所示，位置误差、速度误差比较如图 7、图 8 所示。结果表明，本文算法优于传统 GNSS/IMU 算法和传统 GNSS/IMU/LiDAR 算法。本文算法在水平定位和三维定位上的均方根误差分别为 5.94 m 和 11.00 m，东北天方向上分别为 4.20 m、4.21 m 和 9.26 m。在水平位置和三维位置上，其精度分别比 GNSS/IMU 算法提高了 55.29% 和 56.37%，比传统 GNSS/IMU/LiDAR 算法提高了 38.76% 和 44.28%。该算法在水平速度和三维速度上的精度分别达到 1.48 m/s 和 1.54 m/s，比 GNSS/IMU 算法分别提高了 30.28% 和 29.56%，比传统 GNSS/IMU/LiDAR 算法提高了 22.18% 和 21.54%。结合卫星数量可以看到，误差对比图中误

表 3 场景 1 位置误差

Table 3 Scenario 1 analysis of position error

算法	北方向位置误差/m	东方向位置误差/m	地方向位置误差/m	水平位置误差/m	三维位置误差/m
GNSS/IMU紧组合(算法1)	8.31	10.39	21.44	13.31	25.23
传统GNSS/IMU/LiDAR(算法2)	6.67	7.06	17.20	9.72	19.76
本文算法	4.20	4.21	9.26	5.94	11.00

表 4　场景 1 速度误差

Table 4　Scenario 1 analysis of velocity error

算法	北方向速度误差/(m·s^{-1})	东方向速度误差/(m·s^{-1})	地方向速度误差/(m·s^{-1})	水平速度误差/(m·s^{-1})	三维速度误差/(m·s^{-1})
GNSS/IMU紧组合(算法1)	1.64	1.35	0.54	2.12	2.19
传统GNSS/IMU/LiDAR(算法2)	1.44	1.24	0.50	1.91	1.97
本文算法	1.06	1.03	0.43	1.48	1.54

差较大区域的卫星数较少, 卫星被遮挡严重, 本文算法有效地进行了加权, 使误差大幅减小。

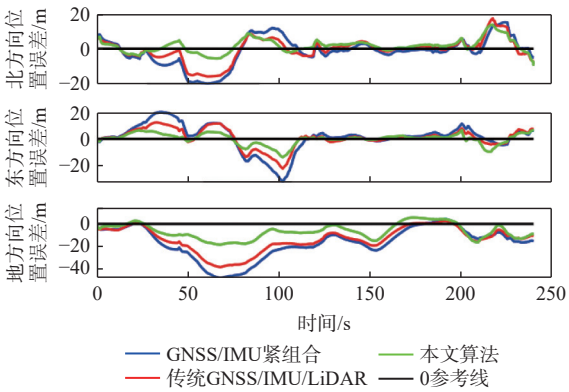


图 7　位置误差对比

Fig. 7　Position error comparison

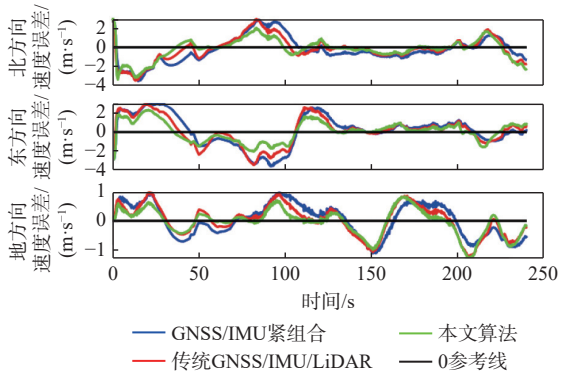


图 8　速度误差对比

Fig. 8　Velocity error comparison

为测试卫星数量较多场景下, 本文算法相对对比算法的性能提升, 采用实验数据 2, 场景 2 的卫星数量如图 9 所示, 与场景 1 的卫星数量对比直方图如图 10 所示。能明显看出场景 2 比场景 1 的卫星数更多, 卫星观测条件更优良。算法测试的性能比较结果如表 5、表 6 所示, 位置误差、速度误差比较如图 11、图 12 所示。在此组数据条件下, 本文算法在水平定位和三维定位上的均方根误差分别为 4.25 m 和 7.11 m, 东、北、地方向上分别为 3.53 m、2.70 m 和 5.69 m。在水平位置和三维位置上, 其精度分别比 GNSS/IMU 算法提高了 49.33% 和 48.31%, 比传统 GNSS/IMU/LiDAR 算法提高了 40.33% 和 37.60%。该算法在水平速度和三维速度上的精度分别达到 0.86 m/s 和 1.01 m/s, 比 GNSS/IMU 算法

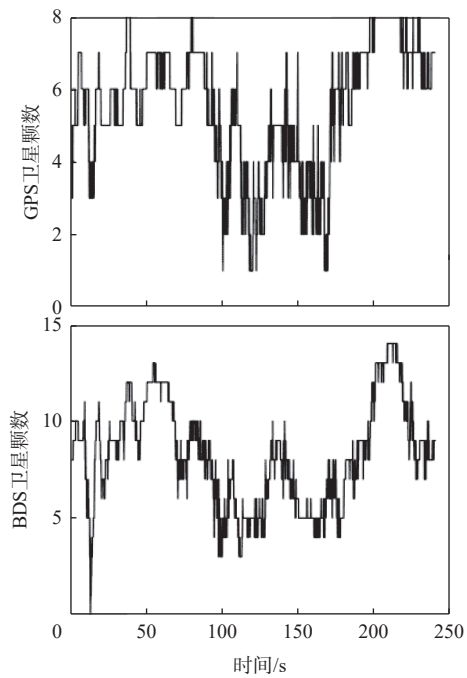


图 9　场景 2 卫星数量

Fig. 9　Number of satellites of scenario 2

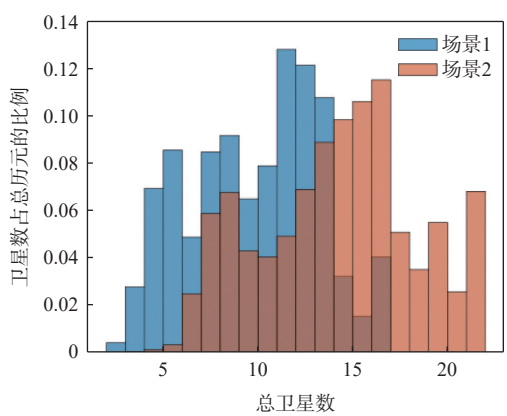


图 10　场景 1 和场景 2 卫星数分布对比直方图

Fig. 10　Comparison of histograms of satellite number distribution in scenario 1 and scenario 2

分别提高了 29.72% 和 25.83%。实测实验表明, 本文算法在卫星观测条件较好时, 也能达到一定的提

图 11 中地方向的误差值较大且波动严重, 这是由于在城市环境中周围建筑的反射和直射导致的多路径效应, 从定位结果的 3 个方向的影响来看, 对地方向的影响最大。地方向的误差大小与载体周围环境高度关联。通过对受到多路径效应影响严

表5 场景2位置误差

Table 5 Scenario 2 analysis of position error

算法	北方向位置误差/m	东方向位置误差/m	地方向位置误差/m	水平位置误差/m	三维位置误差/m
GNSS/IMU紧组合(算法1)	5.56	6.21	10.89	8.39	13.75
传统GNSS/IMU/LiDAR(算法2)	4.98	5.10	8.89	7.13	11.39
本文算法	3.53	2.70	5.69	4.25	7.11

表6 场景2速度误差

Table 6 Scenario 2 analysis of velocity error

算法	北方向速度误差/(m·s^{-1})	东方向速度误差/(m·s^{-1})	地方向速度误差/(m·s^{-1})	水平速度误差/(m·s^{-1})	三维速度误差/(m·s^{-1})
GNSS/IMU紧组合(算法1)	1.02	0.68	0.60	1.22	1.36
传统GNSS/IMU/LiDAR(算法2)	0.84	0.61	0.58	1.04	1.20
本文算法	0.63	0.58	0.53	0.86	1.01

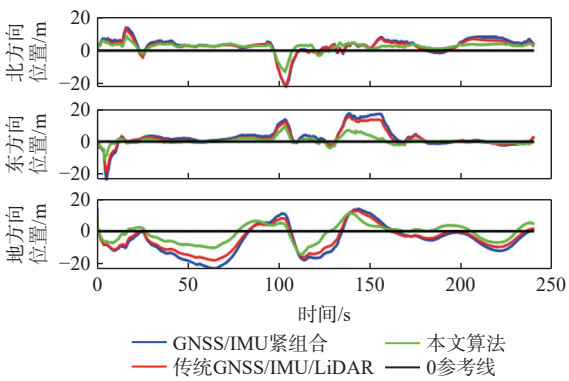


图11 场景2位置误差对比

Fig. 11 Position error comparison in scenario 2

重的伪距信号进行降权，削弱误差严重的信号对定位结果地方向的影响，有效减小了地方向的误差，由对比算法的10.89 m提升到5.69 m。通过对比参考算法结果和本文算法结果，可以看出，绝大部分时间，本文算法的位置速度误差都更小，可以为车辆在城市环境下的安全运行提供支持。

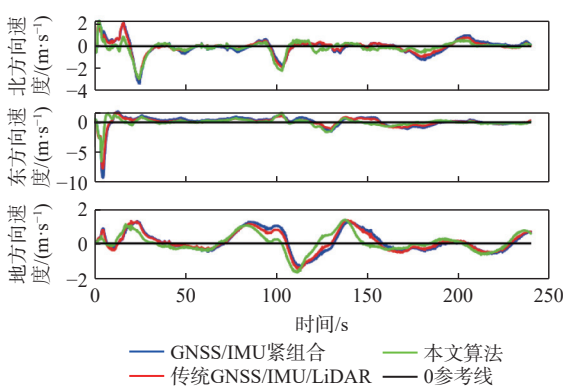


图12 场景2速度误差对比

Fig. 12 Velocity error comparison in scenario 2

5 结 论

1）本文提出的根据距离修正匹配对残差的方法进一步对匹配误差进行建模，提高了LiDAR建图效果，改进扫描匹配算法的总距离误差相比传统算法提高16%。

2）本文提出的使用局部地图和构建的判断量综合确定卫星信号权重的方法一定程度上减弱了点云地图建模不全导致的卫星信号类型误判的影响。

3）本文算法在城市环境实地测试中具备有效性和优越性，选取2种场景进行实验，在卫星质量较差的场景1中，相较传统GNSS/IMU紧组合三维定位精度提高了56.37%，相比传统GNSS/IMU/LiDAR算法三维定位精度提高44.28%，在卫星质量较好的场景2中，相较GNSS/IMU紧组合三维定位精度提高了48.31%，相比传统GNSS/IMU/LiDAR算法三维定位精度提高37.60%。

未来通过继续挖掘LiDAR、GNSS、IMU之间的协调性，可进一步提高算法的性能。

参考文献（References）

[1] KOHLBRECHER, S, VON STRYK O, MEYER J, et al. A flexible and scalable SLAM system with full 3D motion estimation[C]//Proceedings of the 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics . Piscataway: IEEE Press, 2011: 155-160.

[2] BRY A, BACHRACH A, ROY N. State estimation for aggressive flight in GPS-denied environments using onboard sensing[C]// Proceedings of the 2012 IEEE International Conference on Robotics and Automation . Piscataway: IEEE Press, 2012: 1-8.

[3] SHEN S J, MICHAEL N, KUMAR V. Autonomous multi-floor indoor navigation with a computationally constrained MAVC]// Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2011: 20-25.

[4] FALLON M F, JOHANNSSON H, BROOKSHIRE J, et al. Sensor fusion for flexible human-portable building-scale mapping[C]//Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2012: 4405-4412.

[5] Joerger M, Pervan B. Range-domain integration of GPS and laser scanner measurements for outdoor navigation[C]//In Proceedings of

the ION GNSS 19th International Technical Meeting. Fort Worth: [s.n.], 2006: 1115-1123.

[6] HENTSCHEL M, WULF O, WAGNER B. A GPS and laser-based localization for urban and non-urban outdoor environments[C]//Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2008: 149-154.

[7] SOLOVIEV A. Tight coupling of GPS, laser scanner, and inertial measurements for navigation in urban environments[C]//Proceedings of the 2008 IEEE/ION Position, Location and Navigation Symposium. Piscataway: IEEE Press, 2008: 511-525.

[8] JABBOUR M, BONNIFAIT P. Backing up GPS in urban areas using a scanning laser[C]// Proceedings of the 2008 IEEE/ION Position, Location and Navigation Symposium. Piscataway: IEEE Press, 2008: 505-510.

[9] SOLOVIEV A, BATES D, VAN G F. Tight coupling of laser scanner and inertial measurements for a fully autonomous relative navigation solution[J]. Navigation, 2007, 54: 189-205.

[10] LIU S F, ATIA M M, KARAMAT T B, et al. A LiDAR-aided indoor navigation system for UGVs[J]. Journal of Navigation, 2015, 68(2): 253-273.

[11] GARULLI A, GIANNITRAPANI A, ROSSI A , et al. Mobile robot SLAM for line-based environment representation[C]//Proceedings of the 44th IEEE Conference on Decision and Control. Piscataway: IEEE Press, 2005: 2041-2046.

[12] YOKOZUKA M, KOIDE K, OISHI S , et al. LiTAMIN2: ultra light LiDAR-based SLAM using geometric approximation applied with KL-divergence[C]// Proceedings of the 2021 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2021: 11619-11625.

[13] 雷道竖, 刘海波. 一种极坐标系下的激光雷达扫描匹配 SLAM 方法[J]. 中国电子科学研究院学报, 2019, 14(6): 563-567.
LEI D S, LIU H B. A SLAM method based on LiDAR scan matching in the polar coordinates[J]. Journal of China Academy of Electronics and Information Technology, 2019, 14(6): 563-567(in Chinese).

[14] 廖自威, 李荣冰, 雷廷万, 等. 基于几何特征关联的室内扫描匹配 SLAM 方法[J]. 导航与控制, 2016, 15(3): 26-32.
LIAO Z W, LI R B, LEI T W, et al. Indoor scan-matching SLAM method based on geometric features association[J]. Navigation and Control, 2016, 15(3): 26-32(in Chinese).

[15] 谢勃, 张燕. 基于 LiDAR 点云的建筑物激光扫描重构方法仿真[J]. 计算机仿真, 2021, 38(8): 435-439.
XIE B, ZHANG Y. Simulation of LiDAR point cloud based laser scanning reconstruction method for buildings[J]. Computer Simulation, 2021, 38(8): 435-439(in Chinese).

[16] CHEN C X, PEI L, XU C Q, et al. Trajectory optimization of LiDAR SLAM based on local pose graph[C]//China Satellite Navigation Conference 2019 Proceedings. Berlin: Springer, 2019: 360-370.

[17] WEN W S, ZHANG G H, HSU L T . Exclusion of GNSS NLOS receptions caused by dynamic objects in heavy traffic urban scenarios using real-time 3D point cloud: an approach without 3D maps[C]//Proceedings of the 2018 IEEE/ION Position, Location and Navigation Symposium. Piscataway: IEEE Press, 2018: 158-165.

[18] HSU L T, GU Y L, KAMIJO S. 3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation[J]. GPS Solutions, 2016, 20(3): 413-428.

[19] CHEN Y W, ZHU L L, TANG J, et al. Feasibility study of using mobile laser scanning point cloud data for GNSS line of sight analysis[J]. Mobile Information Systems, 2017, 2017: 5407605.

[20] WEN W S. 3D LiDAR aided GNSS and its tightly coupled integration with INS via factor graph optimization[C]// Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation. Vancouver: Institute of Navigation, 2020: 1649-1672.

[21] 文刚, 周仿荣, 李涛, 等. LINS-GNSS: 滤波与优化耦合的 GNSS/INS/LiDAR 巡检机器人定位方法[J]. 南京信息工程大学学报: 自然科学版, 2023, 15(1): 85-93.
WEN G, ZHOU F R, LI T, et al. LINS-GNSS: filter and optimization coupled GNSS/INS/LiDAR positioning method for inspection robot localization[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2023, 15(1): 85-93(in Chinese).

LiDAR aided GNSS/IMU positioning algorithm based on improved point cloud registration in urban environment

HUANG Mingming，SUN Rui*

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

Abstract：Due to the complexity of urban environments, simple global navigation satellite system/inertial measurement unit (GNSS/IMU) cannot meet the positioning demand in urban environments. Light detection and ranging(LiDAR) can perceive the surrounding environment in real-time, and its cost has been continuously decreasing in recent years. Due to the good complementary characteristics, GNSS/IMU/LiDAR has been widely studied. An improved point cloud registration LiDAR assisted GNSS/IMU integrated navigation method is proposed. GNSS quality control is done by the algorithm by adding error terms in LiDAR registration, creating local point clouds, developing double difference feature quantities to evaluate satellite signals thoroughly, and achieving GNSS signal weighting for various signal qualities. Experimental results show that in urban environments, the proposed algorithm achieves a 49.33% improvement in horizontal position accuracy and a 48.31% improvement in 3D position accuracy compared to traditional GNSS/IMU algorithms. Compared to traditional GNSS/IMU/LiDAR algorithms, the horizontal 3D position accuracy has been improved by 40.33% and 37.60%, respectively.

Keywords：GNSS/IMU/LiDAR integrated navigation；improved point cloud registration；local point cloud map；double difference feature quantity

Received：2023-12-29；**Accepted**：2024-04-05；**Published Online**：2024-10-21 09：40

URL：link.cnki.net/urlid/11.2625.V.20241018.1419.001

Foundation items：National Natural Science Foundation of China (42222401,42174025,41974033); Ministry of Industry and Information Technology Special Project of China (TC220A04A-79); Jiangsu Provincial Six Talent Peaks Project (KTHY-014); Natural Science Foundation of Jiangsu Province (BK20211569); the Fundamental Research Funds for the Central Universities (xcxjh20230720)

* **Corresponding author**. E-mail：rui.sun@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0839

小直径C形金属封严环多道次液压成形技术

刘春雷[1]，孟宝[1,*]，朱宇[2]，闫彬宇[1]，万敏[1]

(1. 北京航空航天大学 机械工程及自动化学院，北京 100191；2. 中国航空发动机研究院 仿真技术研究中心，北京 101304)

摘　　要：液压成形是航空发动机复杂薄壁构件精密制造的有效方式。针对某航空发动机薄壁高温合金C形封严环构件尺寸微小的特点，提出2道次液压成形工艺方法。对多道次成形过程进行应力应变分析，建立多道次液压成形过程的有限元分析模型。基于数值模拟和工艺实验，研究合模间距和液压加载路径等工艺参数对封严环成形质量的影响规律，探究截面几何特征丢失、贴模不佳、壁厚过度减薄等失效形式，并优化工艺参数。结果表明：2道次液压成形工艺可实现薄壁C形金属封严环的精确成形，采用优化工艺参数，合模间距1.0 mm，第1道次液室压力140 MPa，第2道次液室压力180 MPa，可制成贴模度为93.9%、减薄率为10.5%以内、壁厚均匀性为85.5%的高质量C形封严环构件。

关　键　词：液压成形；多道次；C形封严环；加载路径；缺陷控制

中图分类号：V261.7; TG394

文献标志码：A　　**文章编号**：1001-5965(2026)03-0955-09

金属封严环是一类具有新型密封形式的轴向自紧式、静密封结构[1]，常见有W形、U形、C形、Ω形等多种截面形状。封严环用于实现航空发动机燃气涡轮、动力涡轮端面密封和新一代飞机气道封严[2]，服役于高温高压环境，通常采用高温合金超薄带材制造。常用材料包括GH4169、GH738等时效硬化镍基变形高温合金，具有高温下屈服强度高、抗拉强度高和变形抗力大等特征。封严环构件一般具有壁厚超薄(0.08～0.3 mm)、变形大、精度高(微米级)等特点[3]。目前，针对封严环的核心设计及加工技术只有少数发达国家掌握，国内尚缺乏成熟、系统化的设计、制造体系和相关数据库[4]。金属封严环的成形工艺主要分为多道次滚压成形和液压成形[5]。在滚压成形研究方面，肖大志等[6]研究了直径为1 600 mm的封严环滚压成形技术，验证了采用滚压成形方法加工薄壁环形件的可行性。郭正华等[7]研究了直径为220 mm的U形封严环滚压成形有限元建模仿真技术，并对有限元模型进行了实验验证，结果表明，滚压成形模拟结果与实验结果十分吻合。李留柱等[8]建立了直径为235.6 mm的W截面封严环多道次滚压成形三维弹塑性有限元模型，研究了工艺参数对封严环壁厚变化的影响规律，并进行了实验研究，得出了合理的工艺参数：驱动辊转速为2 r/s，进给速度为0.2 mm/s，摩擦因数为0.1。赵刚要等[9]对直径为204.55 mm的高温合金薄壁W形封严环多道次滚压成形进行了系统的分析，研究了摩擦对W形封严环多道次滚压成形不均匀变形的影响规律。

液压成形是利用液体作为传力介质或模具，使工件成形的一种塑性加工技术。相较于传统的金属加工技术，液压成形技术具有成形极限高、表面质量好、尺寸精度高等优势。随着航空航天、大型设备等关键构件轻质化整体成形制造的趋势，液压成形技术作为一种先进加工工艺，具有广泛的应用

收稿日期：2023-12-29；**录用日期**：2024-02-08；**网络出版时间**：2024-03-11 20:08
网络出版地址：link.cnki.net/urlid/11.2625.V.20240311.1522.005
基金项目：国家科技重大专项 (J2019-VII-0014-0154)
*通信作者. E-mail：mengbao@buaa.edu.cn

引用格式：刘春雷，孟宝，朱宇，等. 小直径C形金属封严环多道次液压成形技术[J]. 北京航空航天大学学报，2026，52(3)：955-963.
LIU C L，MENG B，ZHU Y，et al. Multi-steps hydroforming technology of small diameter C-shaped metallic seal ring[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(3)：955-963 (in Chinese).

和长久的发展潜力[10-12]。液压成形是另一种常见的封严环柔性成形技术。Yan 等[13] 研究了直径为 35.6 mm 的小直径多波 W 形封严环液压成形，提出了多道次液压成形技术，通过工艺试验验证了提出的多道次液压成形能够准确控制封严环成形过程中的材料流动。Wang 等[14] 针对直径为 230 mm 的 W 形封严环，研究了基于控制材料流动的成形策略，并进行了工艺实验，验证了基于接触压力调整的成形方法的可行性。Wang 等[15] 研究了直径为 100 mm 的 U 形封严环黏介质准体积成形，成形件的最大壁厚减薄率小于 8%。Lang 等[16] 针对直径为 822 mm、具有微小特征尺寸的 W 形环形件，研究了轴向进给与充液胀形相结合的充液成形技术成形此类零件的可行性。朱宇等[17] 研究了直径为 212 mm 的 W 形封严环动模外压成形技术，实现 W 形封严环的整体精确成形。

可见，目前滚压成形和液压成形的封严环直径尺寸一般在 30 mm 以上，对于直径 15 mm 以下封严环的成形技术研究较少。在微介观尺度下，由于界面摩擦增大、微型模具加工制造困难等因素，小直径金属封严环成形过程可能不同于大尺寸封严环。本文针对小直径高温合金 C 形封严环成形过程缺陷控制与参数优化进行了研究，提出多道次液压成形工艺路线。基于理论分析、数值模拟和工艺实验，研究不同工艺参数对成形质量的影响规律，实现了利用 200 MPa 以内液室压力成形出 C 形封严环，拓展了液压成形在小直径复杂薄壁构件制造中的应用。

1 构件工艺分析及成形方案

高温合金 C 形封严环为薄壁复杂截面环形件，其形状和基本尺寸如图 1 所示。封严环材料为镍基变形高温合金 GH4169，其力学性能基本参数如表 1 所示。该零件内径约 10 mm，不便于装夹，截面深度较大(成形深度与直径之比大于 0.6)，变形区圆角半径较小。壁厚超薄，仅为 0.2 mm。

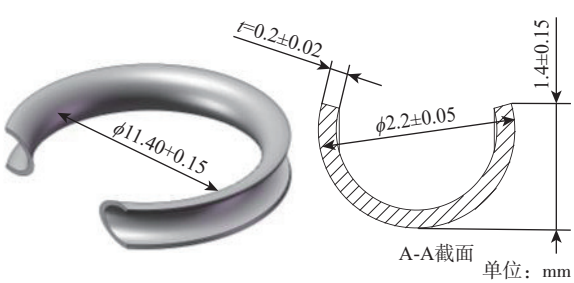


图 1 小直径薄壁 C 形封严环基本尺寸

Fig. 1 Basic dimensions of small diameter thin-wall C-shaped seal ring

表 1 GH4169 材料力学性能参数

Table 1 Material mechanical property parameters of GH4169

抗拉强度σ_b/MPa	屈服强度σ_s/MPa	弹性模量E/GPa	泊松比	密度/ (kg·m^{-3})
965	550	204	0.3	8.24×10^3

在零件的成形过程中，若直接将高压液体作用于坯料成形所需截面形状，则需要加载远超 200 MPa 的超高压液体。该方案存在密封要求高、模具强度要求高及超高压密封难等问题。基于零件的结构特征和材料特性，结合液压成形工艺，提出了小直径 C 形封严环 2 道次液压成形新方法，工艺原理如图 2 所示。该工艺方法的成形步骤包括第 1 道次胀形、合模和第 2 道次胀形。工艺原理如下：第 1 道次胀形阶段，在上、下动模之间放置环形毛坯，设置开模间距且保持不变，充入高压液体，对毛坯进行第 1 道次胀形，使毛坯产生塑性鼓胀，如图 2(b)所示。在合模阶段，两侧动模同步加载，坯料流入型腔内部并逐渐贴模，直至上下动模闭合接触，如图 2(c)所示。在第 2 道次胀形阶段，保持一定的合模力以限制动模移动，增大液室压力，使坯料完全贴模，如图 2(d)所示。

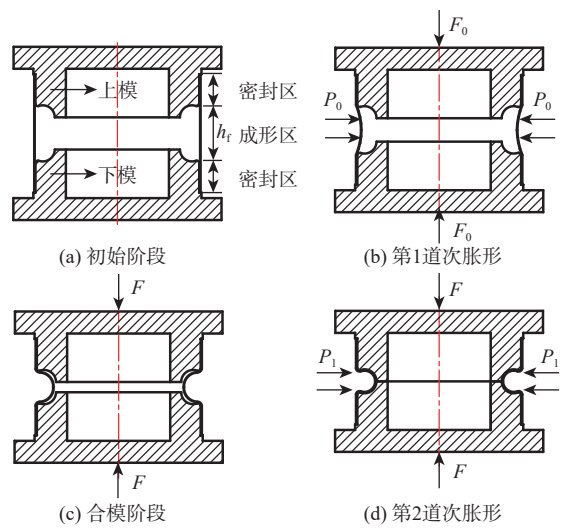
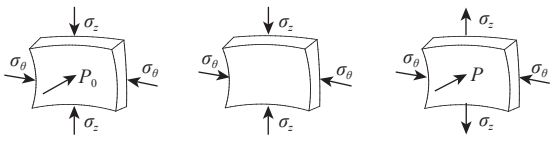


图 2 两道次液压成形工艺原理

Fig. 2 Principle of two-pass hydroforming process

其中，上、下动模初始位置所夹持区域为毛坯的成形区 h_f，成形区主要受液体径向压力，产生塑性变形以成形微小截面特征。环形坯料两端与上下模接触的区域为密封区，其作用是保持型腔内部高压密封。图中，F_0 为锁模力，P_0 为第 1 道次胀形压力，P_1 为第 2 道次胀形压力。

为分析坯料在不同成形阶段的应力应变状态，从环形坯料中切取微小单元，如图 3 所示。在第 1 道次胀形阶段，环形毛坯受到液压径向作用力



(a) 第1道次胀形　(b) 合模　(c) 第2道次胀形

图 3　不同成形阶段应力状态

Fig. 3　Stress state at different forming stages

P_0 作用向内鼓胀, 材料受轴向拉应力 σ_z 和环向压应力 σ_θ 作用。在合模胀形阶段, 材料受到动模轴向作用力逐渐填充型腔, 且成形区坯料内径收缩, 因此, 材料微元轴向和环向皆受压应力。在第 2 道次胀形阶段, 轴向载荷无法作用于型腔内部, 整形压力作用于环形坯料, 使其内径继续收缩, 因此, 材料微元受轴向拉应力和环向压应力。

对成形过程进行应变分析, 厚向应变增量 d_{ε_t} 反映了材料的流动, 根据塑性变形增量理论[18], d_{ε_t} 可表示为

$$d_{\varepsilon_t}=-\frac{d_{\varepsilon_i}}{2\sigma_i}(\sigma_\theta+\sigma_z) \tag{1}$$

式中: d_{ε_i} 和 σ_i 分别为等效应变增量和等效应力。

即环形坯料液压成形过程中材料流动主要受轴向应力和环向应力控制。

当 $\sigma_\theta+\sigma_z=0$ 时, $d_{\varepsilon_t}=0$, 坯料厚向不发生变化, 即轴向拉应力可促使材料向轴向流动, 壁厚不发生变化, 为理想成形状态。

当 $\sigma_\theta+\sigma_z>0$ 时, $d_{\varepsilon_t}<0$, 坯料厚向缩小, 即轴向拉应力促使材料向轴向流动的趋势大于环向压应力对材料流动的补充, 壁厚减薄。

当 $\sigma_\theta+\sigma_z<0$ 时, $d_{\varepsilon_t}>0$, 坯料厚向增大, 若在第 1 道次胀形和第 2 道次胀形阶段, 即轴向拉应力促使材料向轴向流动的趋势小于环向压应力促使材料向中间聚集的趋势, 壁厚增大, 材料易发生聚集, 导致起皱; 若发生在合模胀形阶段, 则轴向和环向都受压, 易发生失稳起皱。

2　数值模拟与实验方法

基于有限元仿真软件 Dynaform 进行 C 形封严环 2 道次液压成形的数值模拟, 包括建立几何和装配模型、划分网格、设置接触和边界条件、设计工艺参数等。

2.1　仿真建模

建立如图 4 所示 2 道次液压成形有限元模型, 模型主要由上动模、下动模和环形坯料组成, 上、动模设置为解析刚体, 环坯设置为变形体。坯料选用适用于薄壁薄板成形的 3 参数 Barlat 材料模型, 选用 4 节点 Be-lytschko-Tsay 壳单元进行网格划分, 应用自适应网格划分。各接触面的接触方式设置为 Surface-to-surface-contact(面面接触), 采用罚函数接触方式; 根据摩擦磨损实验结果, 摩擦系数设置为 0.1。

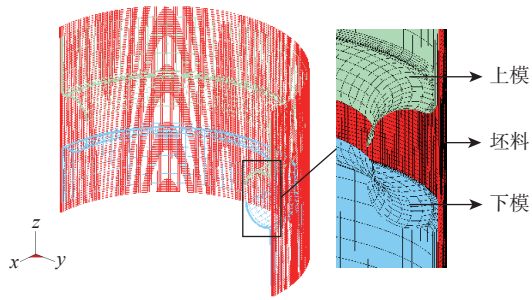


图 4　2 道次液压成形有限元模型

Fig. 4　Finite element model of two-steps hydroforming

2.2　实验设计

合模间距 h_f 决定了轴向送料距离和第 2 道次胀形所需材料。h_f 过大时, 材料环向压应力的作用大于轴向拉应力, 易发生起皱。相反地, h_f 过小时, 第 1 道次胀形所需压力过大。考虑 2 道次液压的工艺特点和材料流动过程中的壁厚变化, 设计环坯合模间距分别为 0.5 mm、1.0 mm 和 1.5 mm。

由外压圆桶理论, 计算出环形坯料发生径向预变形所需要的最小压力为 8 MPa, 因此, 第 1 道次胀形压力 P_0 须大于 8 MPa。在第 2 道次胀形阶段, 胀形压力越大, 贴模精度越高, 最大胀形压力 P_1 设置为 180 MPa。依据理论计算和实际工程经验, 选用系列不同于第 1 道次胀形压力, 设计如图 5 所示的 5 种不同液压加载路径, 曲线 A～F 的第 1 道次胀形压力 P_0 分别为 60 MPa、80 MPa、100 MPa、120 MPa、140 MPa 和 150 MPa。

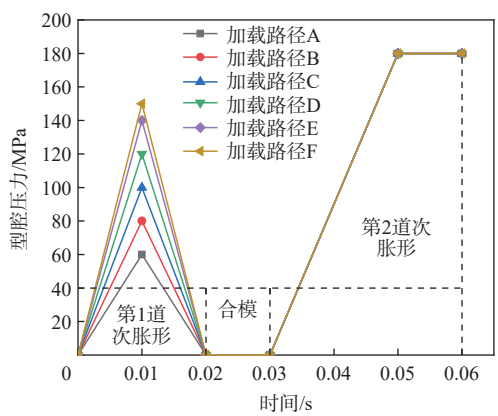


图 5　型腔压力加载曲线

Fig. 5　Loading paths of cavity pressure

按照 2 道次液压成形工艺特点, 结合 C 形封严环小直径、薄壁的特点, 设计多层可分离空心环状模具结构, 关键部件分瓣设计。对 O 形、Y 形等小直径橡胶密封圈进行密封效果对比, 选用 Y 形密封

圈，保障高压密封要求。利用3D液压成形设备提供所需的180 MPa液压，设备和模具如图6所示。



图6 成形设备和模具

Fig. 6 Forming equipment and dies

3 工艺参数确定

工艺参数对2道次液压成形有重要影响，本文采取解析法初步确定了第1道次胀形压力的选取原则，同时分析了模具圆角对成形结果的影响。

3.1 第1道次胀形压力计算

第1道次的作用是增大变形区，为第2道次胀形提供充足的材料储备。依据等体积原理，在理想状态下，合模及第2次胀形只是使材料在型腔内部流动，型腔外坯料不产生壁厚变化。因此，合适的第1道次胀形压力 P_0 应满足：第1道次胀形后的弓形曲面的母线长度等于标准零件母线长度。

当第1道次胀形压力 P_0 较小，坯料变形程度较小时，假定坯料第1次胀形后沿轴截面剖开（母线）为抛物线（见图7），可用抛物线函数表示。其中，标准零件C形张口距离为 $2s$，第1道次胀形深度为 h_1，标准零件母线弧长为 S。型腔部分的坯料中性层抛物线函数为

$$y=a\left(x+s+\frac{h_\mathrm{f}}{2}\right)\left(x-s-\frac{h_\mathrm{f}}{2}\right) \tag{2}$$

式中：a 为抛物线参数；x、y 为自变量和函数。

第1道次胀形后弓形曲面母线弧长为

$$S_1=\int_{-s-\frac{h_\mathrm{f}}{2}}^{s+\frac{h_\mathrm{f}}{2}}\sqrt{1+(y')^2}\,\mathrm{d}x \tag{3}$$

式中：S_1 为第1道次胀形后弓形曲面母线弧长；s 为C形张口距离的一半。

根据第1道次胀形后弓形曲面母线弧长等于标准零件母线长度，则有

$$S_1=S \tag{4}$$

则抛物线参数为

$$a\approx-\frac{\sqrt{S^2-\left(s+\frac{h_\mathrm{f}}{2}\right)^2}}{2\left(s+\frac{h_\mathrm{f}}{2}\right)^2} \tag{5}$$

第1道次胀形深度为

$$h_1\approx-\sqrt{S^2-\left(s+\frac{h_\mathrm{f}}{2}\right)^2}\Big/2 \tag{6}$$

当第1道次胀形压力 P_0 较大、坯料变形程度较大时，假定坯料第一次胀形后沿轴截面剖开（母线）为半圆，可用圆函数表示。此时，第1道次胀形后弓形曲面母线弧长为

$$S_1=\pi h_1 \tag{7}$$

根据第1次胀形后弓形曲面母线弧长等于标准零件母线弧长，即式(3)，则第1道次胀形深度为

$$h_1=S/\pi \tag{8}$$

P_0 的选取原则为：第1道次胀形深度 h_1 为式(6)和式(8)的理论计算值。

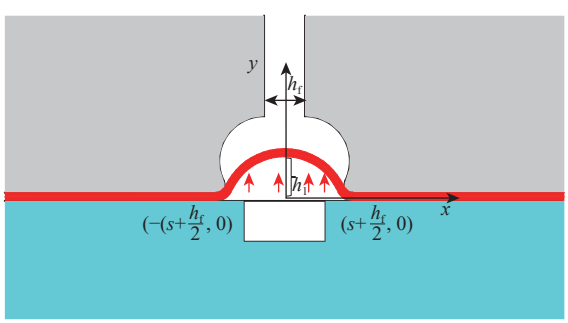


图7 第一道次胀形参数

Fig. 7 First pass hydroforming parameters

3.2 模具圆角确定

C形封严环2道次液压成形过程中，模具圆角半径 r 对坯料的流动有重要影响[19]。成形后，r 仍对回弹有影响。通过对不同的模具圆角半径下C形封严环纯胀形过程的有限元仿真，对成形后封严环的贴模间隙 d（成形结束时工件轴截面胀形顶点到模具型腔顶点的距离）、最大回弹量、最大壁厚减薄率进行对比分析，以得到模具圆角半径 r 对C形封严环成形的影响规律及最优模具圆角尺寸。

模拟时采用纯胀形工艺，最大液压为180 MPa，依据工程经验，r 通常取在0.2 mm附近，为探究 r 对成形的影响，r 取值范围为0.05 mm、0.1 mm、0.15 mm、0.2 mm、0.25 mm、0.3 mm、0.4 mm和0.5 mm。贴模间隙 d、最大回弹量及最大壁厚减薄率随 r 变化如图8～图10所示。d 代表成形的难易程度。从图8可以看出，d 随着 r 的增大先迅速减小趋近于不变然后增加，最大回弹量随着 r 的增大而增大。这是由于在一定范围内，r 增大，便于材料流动，越容易成形；当 r 在最优值附近，成形结果趋近于一致；当 r 超过最优值，模具圆角增大也会使变形区变小、成形深度变大，越难以成形。图9为最大回弹量随 r 的变化，最大回弹量随 r 的增大而增大。这是由于 r 越大，成形件直径越大，回弹量也就越大。图10为最大壁厚减薄率随 r 的变化，可见，最大减薄率集中在10%左右，仅在 r 为0.1 mm

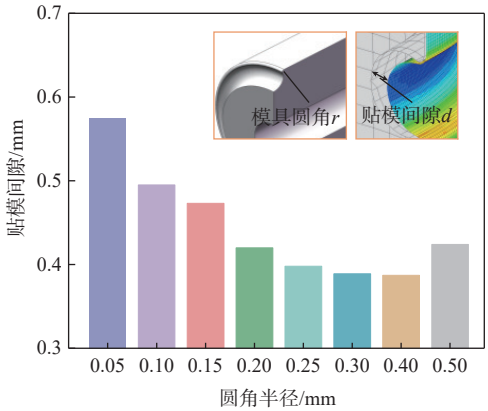


图 8　贴模间隙随模具圆角半径变化曲线

Fig. 8　Curve of the attachment gap between the formed blank and the die with transition fillet radius of die

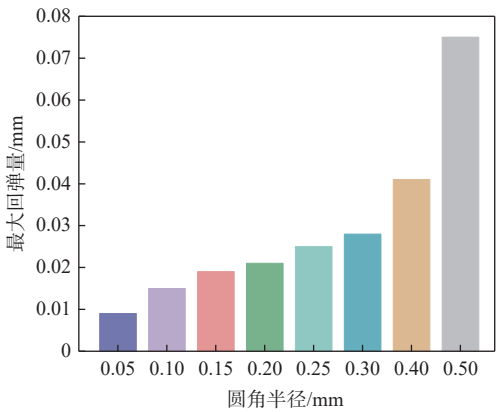


图 9　最大回弹量随模具圆角半径变化曲线

Fig. 9　Curve of maximum springback with transition fillet radius of die

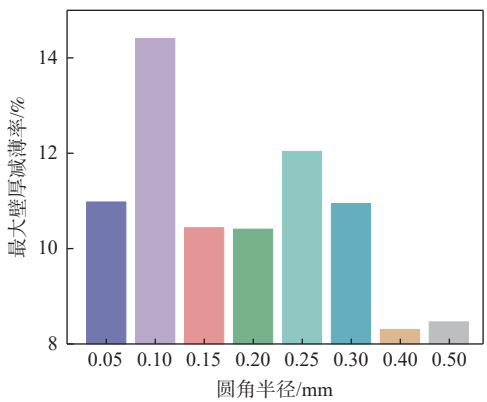


图 10　最大壁厚减薄率随模具圆角半径变化曲线

Fig. 10　Curve of maximum wall thickness thinning rate with transition fillet radius of die

和 0.25 mm 时, 最大减薄率较大。模具圆角的选取要同时满足尽可能小的贴模间隙 d、最大回弹量和最大壁厚减薄率, 可见, 当 r=0.2 mm 时, 同时满足了较小的贴模间隙、较小的最大壁厚减薄率及较小的回弹量, 综合考虑 3 个因素, 选取模具圆角 r 值为 0.2 mm。

4　C 形封严环成形过程分析

2 道次液压成形工艺特点决定了 C 形封严环成形控制的关键是抑制成形过程中的环向失稳并消除起皱。在模具尺寸确定的前提下, 通过数值模拟和工艺试验, 研究合模间距和液压加载路径等工艺参数对成形结果的影响, 进而分析不同成形阶段的变形原理和缺陷控制方法。

4.1　成形过程壁厚变化分析

选取图 5 中加载路径 E, 依据式(1)的塑性变形增量理论, 结合数值模拟结果, 分析不同成形阶段的壁厚变化及缺陷形成原因。

第 1 道次胀形阶段, 毛坯在液压作用下向成形区中部变形, 坯料受环向压应力和轴向拉应力, 环向压应力使材料向型面中间聚集的趋势大于轴向拉应力使材料向轴向流动的趋势, 胀形顶点处壁厚稍微增加。在变形初期, 变形区弓形曲面母线长度增加较少, 轴向拉应力小, 环向压应力大, 材料只能向胀形顶点处聚集, 因而此时坯料壁厚减薄不大。随着胀形力的增加, 弓形曲面母线长度变长, 轴向拉应力大于环向压应力, 顶点处壁厚开始减薄, 直至第 1 次胀形结束。

合模阶段, 毛坯受上下动模的作用力而贴靠型腔, 坯料在轴向和环向都受压应力。由于合模阶段未施加液压胀形力, 型腔内部弓形曲面材料只能向顶点处聚集, 顶点处壁厚在整个合模阶段逐渐增加。若第 1 次胀形形成的弓形曲面较大, 则此时易产生起皱。

第道 2 次胀形阶段, 毛坯在液体高压下贴靠型腔内壁, 坯料受环向压应力和轴向拉应力。此时, 当胀形力较小时, 弓形曲面变形较小, 环向压应力等于轴向拉应力, 顶点处材料不发生流动, 壁厚保持不变。当胀形力较大时, 弓形曲面贴靠型腔, 环形压应力小于轴向拉应力, 壁厚略微较薄。图 11 为成形件顶点处壁厚变化曲线。

4.2　合模间距对成形结果的影响

毛坯合模间距 h_f 影响轴向送料距离和第 2 道次胀形所需材料。分别设置合模间距为 0.5 mm, 1.0 mm 和 1.5 mm, 选择图 5 中加载路径 D 进行数值模拟和工艺试验, 成形结果如图 12 所示。

可见, 不同 h_f 条件下, 成形件壁厚最大值均在圆弧顶点处, 这与传统内高压胀形不同, 其原因为在 C 形封严环外压成形过程中, 坯料向型腔内流动, 环向受压, 轴向压应力使材料向型面顶点流动以补充型面变形所需的材料, 故型面顶点处壁厚减薄不明显甚至不减薄。

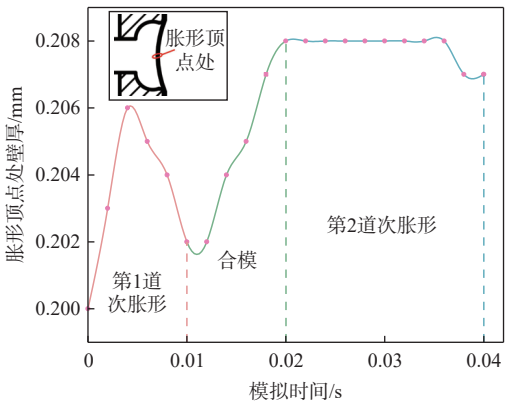
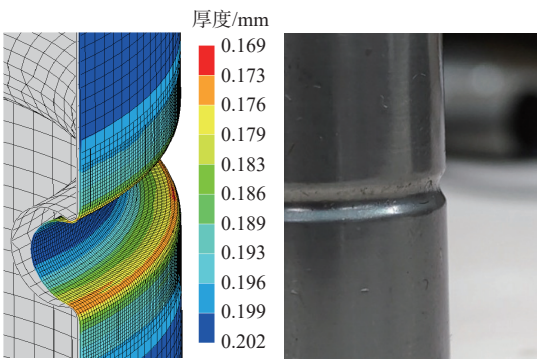
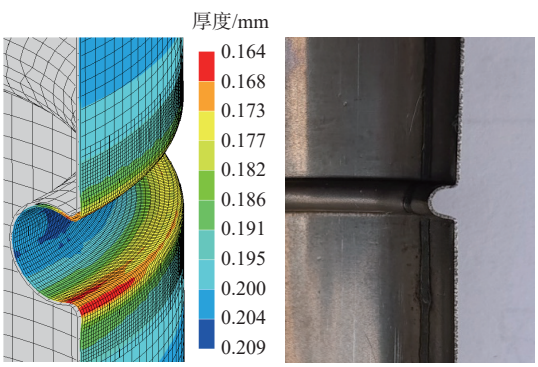


图 11 胀形顶点处壁厚随模拟时间变化曲线

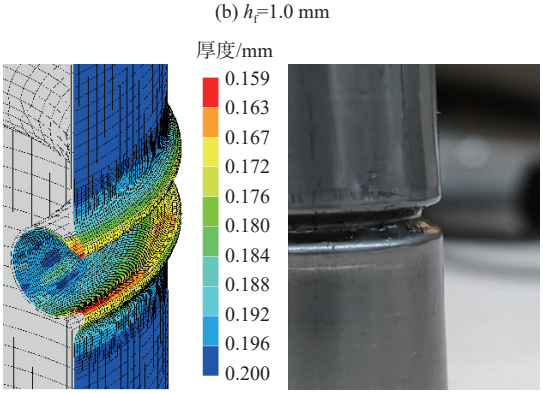
Fig. 11 Curve of wall thickness at bulging vertex changing with simulation time



(a) h_f=0.5 mm



(b) h_f=1.0 mm



(c) h_f=1.5 mm

图 12 不同 h_f 条件下的模拟与实验结果对比

Fig. 12 Comparison of results between simulations and experiments with different h_f

h_f 偏小(0.5 mm),变形区面积较小,毛坯成形区储料过少。在第 1 道次胀形压力下,型面处坯料变形不充分,提供给第 2 道次胀形的材料不足,导致合模后试件在顶点处偏离模具型面较大,第 2 道次胀形后型腔处坯料仍然无法贴靠模具型面。此时,顶点处壁厚略微增加,圆角处壁厚减薄。成形试件内径偏大,截面为凹弧较小的弓形曲面,截面几何特征丢失严重,图 12(a)为模拟与实验结果对比,其中,顶点距离理论型面处偏差为 2.5 mm。

h_f 偏大(1.5 mm),变形区面积较大,毛坯成形区储料过多。在第 1 道次胀形压力下,型面处变形坯料较多,顶点处由于严重材料堆积而产生起皱;同时,第 1 道次胀形后形成弓形曲面的母线长度超过最终零件的母线长度,合模后局部材料在圆角处堆积,形成翘曲,截面几何特征产生冗余。图 12(c)为模拟与实验结果对比,顶点处产生起皱。

当采用适当的 h_f(1.0 mm)时,第 1 道次胀形后形成弓形曲面母线长度接近最终零件母线长度,较好满足了第 2 次胀形所需材料,具备合适的材料储备,几乎不产生截面几何特征丢失,同时,顶点处材料稍有堆积但尚未起皱,增加了顶点处壁厚,具有较好的贴模性和最终成形质量。图 12(b)为模拟与实验结果对比。

4.3 第 1 道次胀形压力对成形结果的影响

由 3.1 节分析可知,第 1 道次胀形的目的是为第 2 道次胀形提供材料储备。理论上,第 1 道次胀形后的弓形曲面的母线长度等于成形零件母线长度。采用模具圆角半径 r=0.2 mm,合模间距 h_f=1.0 mm 进行 C 形封严环 2 道次液压成形数值模拟,对成形后的第 1 道次成形深度、最终成形深度及截面壁厚分布进行对比分析,得到了第 1 道次胀形压力对第 1 道次成形深度、最终成形深度和截面壁厚分布的影响规律。

第 1 道次成形深度随第 1 道次胀形压力 P_0 的变化如图 13 所示。可以看出,第 1 道次成形深度随 P_0 增大而增大。第 1 道次成形深度与弓形曲面母线长度有关,弓形曲面的母线长度越大,第 1 道次成形深度越大。由式(5)可知,h_1 即为第 1 道次成形深度,代入工艺参数,第 1 道次成形深度约为 1.5 mm。相应的,较优的第 1 道次胀形压力 P_0 为 140 MPa 左右。

最终成形深度随第 1 道次胀形压力 P_0 的变化如图 14 所示,图中也给出了 P_0=60 MPa, P_0=100 MPa 和 P_0=140 MPa 时的成形结果。可以看出,最终成形深度随着 P_0 的增大而增加。P_0 为 140 MPa 和 150 MPa 时,两者第 1 道次成形深度差异较大,最终

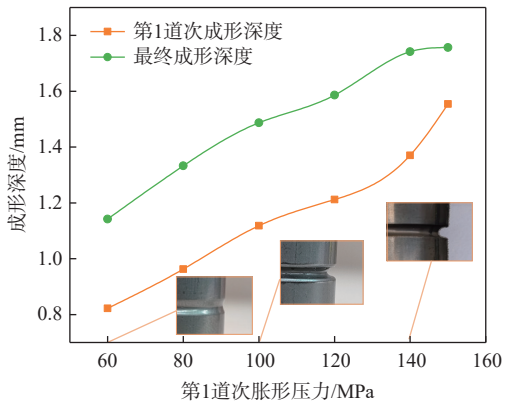


图 13　成形深度随第 1 道次胀形压力变化曲线

Fig. 13　Curve of forming depth with first pass hydroforming pressure

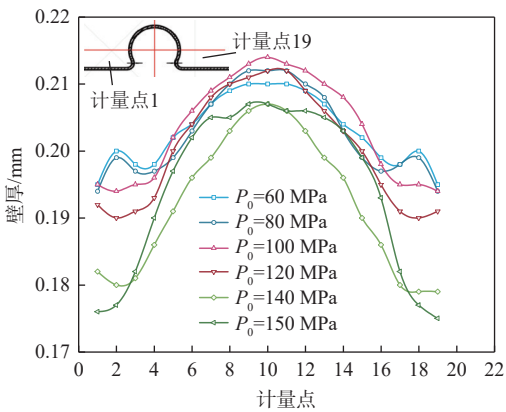


图 14　不同第 1 道次胀形压力下成形零件壁厚分布

Fig. 14　Thickness distribution of forming parts under different first pass hydroforming pressures

成形深度却接近。这是由于 P_0=140 MPa 时, 第 1 道次成形后弓形曲面母线长度已经满足要求, 接近最终零件母线长度。在 P_0=150 MPa 时, 第 1 道次成形后弓形曲面母线长度大于最终成形后零件母线长度, 在第 2 道次胀形时, 部分材料在顶点处堆积或在圆角处形成翘曲。因此, 最优第 1 道次胀形压力 P_0 在 140 MPa 附近。

截面几何特征与最终成形深度有关。第 1 道次胀形压力越大, 最终成形深度越大, 截面几何特征丢失越少。由图 13 可看出, P_0=60 MPa 时, 截面为凹弧较小的弓形, 几何特征丢失严重; P_0=100 MPa 时, 截面接近圆弧, 几何特征略微丢失; P_0=140 MPa 时, 截面为直径 2.2 mm 的圆弧, 几乎不产生截面几何特征丢失。

选取 C 形封严环截面上对称的 19 个点, 测量成形后 C 形封严环截面壁厚 t。不同第 1 道次胀形压力 P_0 下成形后零件截面壁厚分布曲线如图 14 所示。可以看出, 顶点处壁厚最大, 越靠近两边, 壁厚越薄。这是由于 C 形封严环在成形过程中, 型腔内部材料受环向压应力, 因此, 材料聚集、壁厚增加。当 P_0=60 MPa 时, 坯料变形程度较小, 型面中间壁厚增加较小, 两边壁厚减薄较小; 且由于第 1 道次成形后弓形曲面母线长度较短, 在合模时, 大量材料向顶点处聚集, 因此, 顶点处壁厚相对两边较大。随着 P_0 增大(80 MPa), 坯料变形程度变大, 型面中间壁厚增加减小, 两边壁厚减薄增大; 且由于第 1 道次成形后弓形曲面母线长度增长, 在合模时, 材料聚集空间变大, 引起的壁厚增加减小。当 P_0=100 MPa 时, 壁厚最大。随着 P_0 的进一步增大(120 MPa、140 MPa), 坯料变形程度持续变大, 此时型面中间处壁厚增大持续减小, 型面两边壁厚减小持续增大; 且由于第 1 道次成形后弓形曲面母线长度持续增大, 合模时材料聚集引起的壁厚逐渐减小, 因此, 整体壁厚减小。当 P_0=150 MPa, 由于第 1 道次成形后弓形曲面母线长度大于成形后零件母线长度, 在合模时, 材料向型面中间聚集, 型面中间整体壁厚增加, 型面两端壁厚较薄。

由上述壁厚分析可知, 第 1 道次胀形压力对成形后零件壁厚分布有重要影响。采用壁厚均匀性指标 u_t 分析第 1 道次胀形压力对成形质量的影响, 其定义如下:

$$u_t = \left(1 - \frac{(t_{\max} - t_{\min})}{t_{\text{agv}}}\right) \times 100\% \qquad (9)$$

式中: $t_{\max}$ 为最大壁厚值; $t_{\min}$ 为最小壁厚值; t_{agv} 为壁厚平均值。

分析不同第 1 道次胀形压力 P_0 下成形件的壁厚均匀性 u_t, 综合考虑中间壁厚和两端壁厚分布情况, 选取 P_0 为 140 MPa, 此时壁厚均匀性为 85.5%。

4.4　第 2 道次胀形压力对成形结果的影响

第 2 道次胀形的目的是对第 1 道次胀形后的型腔坯料进行整形, 使其完全贴模, 达到理论截面几何特征。因此, 第 2 道次胀形压力 P_1 应大于第 1 道次胀形压力 P_0。

在 r=0.2 mm、h_f=1.0 mm、P_0=140 MPa 条件下, 分别选用 P_1 为 160 MPa、180 MPa 和 200 MPa 进行数值模拟和工艺试验, 探究 P_1 对成形质量的影响。

第 2 道次胀形压力对最终成形深度和最大壁厚减薄率的影响如图 15 所示。可以看出, P_1=180 MPa 和 P_1=200 MPa 时的最终成形深度接近且远大于 P_1=160 MPa 时的最终成形深度, 最大壁厚减薄率随 P_1 增大而略微增大, 都在 10% 附近。可见, 第 2 道次胀形对型腔内部材料产生整形作用, 使其更好贴模。当 P_1=160 MPa 时, 最大成形深度较小, 贴模度较小, 截面特征丢失较为严重; 当 P_1=180 MPa 和 P_1=200 MPa 时, 最大成形深度相当, 贴膜度好,

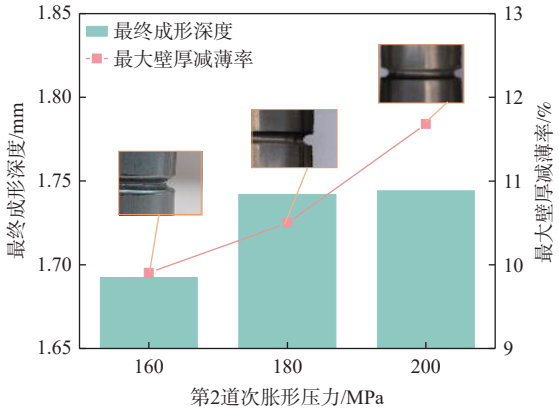


图 15 最终成形深度和最大壁厚减薄率随第二道次胀形压力变化曲线

Fig. 15 Curve of final forming depth and maximum wall thickness thinning rate with second pass hydroforming pressure

截面几何特征几乎不产生丢失，且 P_1=180 MPa 时，最大壁厚减薄率较小。这是因为 P_1=180 MPa 时第 2 道次胀形压力已满足整形作用，产生了较好的贴膜性，此时 2 道次胀形结束后成形件贴模度（最终成形深度/C 形封严环顶点高度）为 93.9%。综合考虑贴模度和壁厚减薄情况，选取 P_1=180 MPa。

对采用优化工艺参数（r=0.2 mm, h_f =1.0 mm, P_0=140 MPa, P_1=180 MPa）成形的 C 形封严环试件（如图 16 所示）检测其成形性能，如表 2 所示。

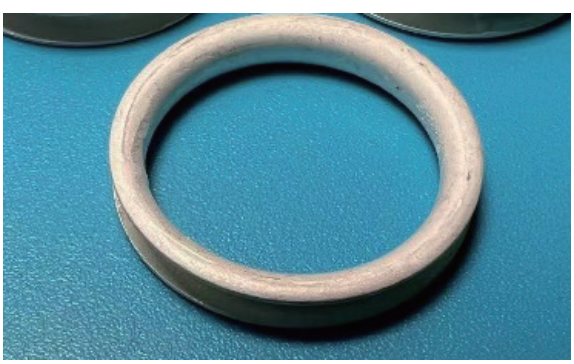


图 16 成形试件

Fig. 16 Forming part

表 2 成形试件性能检测结果

Table 2 Performance test results of forming parts

试件编号	开口面直径/mm	圆度/mm	端面平行度/mm	晶粒度/级	表面粗糙度/μm
C#1	11.504	0.073	0.046	10.5	0.64
C#2	11.526	0.069	0.064	10.5	0.64
C#3	11.503	0.087	0.063	10.5	0.64

5 结 论

1）在小直径 C 形封严环多道次液压成形过程中，通过选择合适的工艺参数，使第 1 道次成形后弓形曲面母线长度接近成形后零件母线长度，可提高成形后零件的贴模度，选择模具圆角半径为 0.2 mm 可降低成形难度，减小回弹量。

2）合模间距和第 1 道次胀形压力是影响 C 形环 2 道次液压成形的主要因素，第 1 道次胀形后型面起皱是主要的失效形式。通过合理控制合模间距和第 1 道次胀形压力，可抑制起皱产生。

3）采用合模间距为 1.0 mm，第 1 道次胀形压力为 140 MPa，第 2 道次胀形压力为 180 MPa 时，成形件贴模度为 93.9%，壁厚最大减薄率为 10.5%，壁厚均匀性为 85.5%。

4）小直径 C 形封严环多道次液压成形实验表明，多道次液压成形工艺方法能够利用小于 200 MPa 液室压力（P_0=140 MPa、P_1=180 MPa）实现小直径 C 形封严环的精确成形。

参考文献（References）

[1] 孟宝, 贺炜林, 万敏. 航空发动机金属封严环成形回弹预测与控制[J]. 锻压技术, 2021, 46(9): 145-153.
MENG B, HE W L, WAN M. Prediction and control of springback for aeroengine metallic sealing ring forming[J]. Forging & Stamping Technology, 2021, 46(9): 145-153(in Chinese).

[2] YANG H, ZHAN M, LIU Y L, et al. Some advanced plastic processing technologies and their numerical simulation[J]. Journal of Materials Processing Technology, 2004, 151(1-3): 63-69.

[3] ZHANG X, LI H W, ZHAN M. Mechanism for the macro and micro behaviors of the Ni-based superalloy during electrically-assisted tension: local Joule heating effect[J]. Journal of Alloys and Compounds, 2018, 742: 480-489.

[4] 王云, 徐江锋. 航空发动机金属波纹封严环性能分析[J]. 航空工程进展, 2012, 3(3): 379-382.
WANG Y, XU J F. Performance analysis of corrugated metal sealing ring used in aviation engine[J]. Advances in Aeronautical Science and Engineering, 2012, 3(3): 379-382(in Chinese).

[5] 朱宇, 张贺刚, 董红瑞, 等. 航空发动机钣金件先进成形技术[J]. 锻压技术, 2023, 48(5): 67-78.
ZHU Y, ZHANG H G, DONG H R, et al. Advanced forming technology on sheet metal parts in aero-engines[J]. Forging & Stamping Technology, 2023, 48(5): 67-78(in Chinese).

[6] 肖大志, 樊泽兴, 杨成林. 薄壁环形零件滚压成形研究[J]. 材料科学与工艺, 2006, 14(1): 75-77.
XIAO D Z, FAN Z X, YANG C L. Research on roll forming of annular thin-walled parts[J]. Materials Science and Technology, 2006, 14(1): 75-77(in Chinese).

[7] 郭正华, 应帅, 赵刚要, 等. 高温合金 U 形环滚压成形有限元建模关键技术研究[J]. 装备制造技术, 2013(6): 1-2.
GUO Z H, YING S, ZHAO G Y, et al. Study on the key technology of rolling finite element modeling forming super alloy U ring[J]. Equipment Manufacturing Technology, 2013(6): 1-2(in Chinese).

[8] 李留柱, 李智军, 李宏伟, 等. 高温合金薄壁 W 截面密封环滚压成形壁厚变化研究[J]. 精密成形工程, 2019, 11(5): 43-49.
LI L Z, LI Z J, LI H W, et al. Wall thickness variation of a superalloy thin-walled W-section seal ring during roll forming[J]. Journal of Netshape Forming Engineering, 2019, 11(5): 43-49(in Chinese).

[9] 赵刚要, 张冉阳, 郭正华, 等. 摩擦对高温合金复杂截面薄壁圆环

多道次滚压成形不均匀变形作用[J]. 稀有金属材料与工程, 2018, 47(12): 3761-3767.
ZHAO G Y, ZHANG R Y, GUO Z H, et al. Effect of friction on nonuniform deformation behavior of thin-walled superalloy ring with complex section in multi-pass rolling[J]. Rare Metal Materials and Engineering, 2018, 47(12): 3761-3767(in Chinese).

[10] 苑世剑, 刘伟, 徐永超. 板材液压成形技术与装备新进展[J]. 机械工程学报, 2015, 51(8): 20-28.
YUAN S J, LIU W, XU Y C. New development on technology and equipment of sheet hydroforming[J]. Journal of Mechanical Engineering, 2015, 51(8): 20-28(in Chinese).

[11] 林俊峰, 苑世剑. 管件液压成形技术应用进展[J]. 机械工人, 2006(3): 26-29.
LIN J F, YUAN S J. Application progress of hydroforming technology for pipe fittings[J]. Machinist Metal Forming, 2006(3): 26-29(in Chinese).

[12] 程东明, 苑世剑, 刘钢. 异型截面轻体构件内高压成形技术[J]. 航空制造技术, 2007, 50(增刊 1): 478-483.
CHENG D M, YUAN S J, LIU G. Internal high pressure forming of lightweight components with irregular cross sections[J]. Aeronautical Manufacturing Technology, 2007, 50(Sup 1): 478-483(in Chinese).

[13] YAN B Y, MENG B. Wall thickness control in multi-stage hydroforming of multiwave seal ring with small diameter[C]//Proceedings of the IOP Conference Series: Materials Science and Engineering. Bristol: IOP, 2022, 1270(1): 012067.

[14] XIANG N, WANG Z J, YI J, et al. Controlling of material flow in the quasi-bulk forming of thin-walled corrugated rings through optimization of contact pressure[J]. The International Journal of Advanced Manufacturing Technology, 2017, 91(5): 2077-2088.

[15] WANG Z J, XIANG N, YI J, et al. Forming thin-walled circular rings with corrugated meridians via quasi-bulk deformation of metal blank and viscous medium[J]. Journal of Materials Processing Technology, 2016, 236: 35-47.

[16] LANG L H, WANG Y M, DUAN X M, et al. Research on multi-step hydroforming of complicated thin wall circular part with small section[J]. Duanya Jishu-Forging & Stamping Technology, 2011, 36(2): 34-38.

[17] 朱宇, 万敏. 航空发动机薄壁 W 形封严环动模外压成形[J]. 航空学报, 2015, 36(7): 2457-2467.
ZHU Y, WAN M. External pressure forming of thin walled W-shaped sealing rings in aircraft engines using movable dies[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(7): 2457-2467(in Chinese).

[18] 梁炳文, 胡世光. 板料成形塑性理论[M]. 北京: 机械工业出版社, 1987: 97-112.
LIANG B W, HU S G. Plastic theory of sheet metal forming[M]. Beijing: China Machine Press, 1987: 97-112(in Chinese).

[19] 李晓冬, 徐雪峰, 华如雨, 等. 5052 铝合金 T 型三通管内高压成形规律研究[J]. 塑性工程学报, 2021, 28(3): 41-49.
LI X D, XU X F, HUA R Y, et al. Research on internal high pressure forming laws of 5052 aluminum alloy T-shape tube[J]. Journal of Plasticity Engineering, 2021, 28(3): 41-49(in Chinese).

Multi-steps hydroforming technology of small diameter C-shaped metallic seal ring

LIU Chunlei[1], MENG Bao[1,*], ZHU Yu[2], YAN Binyu[1], WAN Min[1]

(1. School of Mechanical Engineering & Automation, Beihang University, Beijing 100191, China;
2. Simulation Technology Research Center, China Aero Engine Research Institute, Beijing 101304, China)

Abstract: Hydroforming is an effective way for precision manufacturing of complex thin-walled components of aeroengines. According to the micro size characteristics of thin-walled superalloy C-shaped seal ring components of an aero-engine, a two-steps hydroforming process was proposed. The stress-strain analysis of multi-steps hydroforming process was conducted, and the finite element analysis model of multi-step hydroforming process was established. The impact of process variables, such as the height of the blank forming and the hydraulic loading path, on the forming quality of the seal ring was investigated using numerical simulation and process experimentation. Failure modes, such as the loss of section geometric characteristics, inadequate die attaching, and excessive wall thickness thinning, were also investigated. The process parameters were optimized. The results show that the two-step hydroforming process can achieve accurate forming of a thin-walled C-shaped metal seal ring. By using the optimized process parameters: height of blank forming of 1.0mm, first pass cavity pressure of 140 MPa, second pass cavity pressure of 180 MPa, the high-quality C-shaped seal ring with the degree of blank molding of 93.9%, thinning rate of 10.5% and wall thickness uniformity of 85.5% can be made.

Keywords: hydroforming; multi-steps; C-shaped seal ring; loading path; defect control

Received: 2023-12-29; **Accepted**: 2024-02-08; **Published Online**: 2024-03-11 20:08
URL: link.cnki.net/urlid/11.2625.V.20240311.1522.005
Foundation item: National Science and Technology Major Project (J2019-VII-0014-0154)
* **Corresponding author**. E-mail: mengbao@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0846

基于 FastSAM 辅助表示增强的自监督单目深度估计

张冬冬[1,*]，王春平[2]，付强[2]

(1. 国防科技大学 电子对抗学院，合肥 230037；2. 陆军工程大学石家庄校区，石家庄 050003)

摘 要：针对自监督单目深度估计方法对薄结构区域和边界区域深度估计效果不佳的问题，提出一种基于 FastSAM 辅助表示增强的自监督单目深度估计的方法。首先，引入 FastSAM，在不引入额外监督的情况下为深度网络提供丰富的语义信息。其次，提出语义指导模块 (SGM)，以探索语义特征和深度特征之间的相关性，并增强全局特征表示。另外，设计 (EGM)，以引导网络更加关注局部特征，从而改善边界深度估计性能。大量实验表明，所提方法优于最先进的方法，尤其是在薄结构区域和边界区域的深度估计上具有明显优势。

关 键 词：深度估计；自监督；FastSAM；语义信息；特征表示

中图分类号：TP753

文献标志码：A **文章编号**：1001-5965(2026)03-0964-09

深度估计是计算机视觉领域中的一项基本任务。高质量的深度信息可以为多个领域提供有价值的线索，包括自动驾驶[1]、机器人感知[2]和增强现实[3]等。感知深度通常需要专用传感器(如激光雷达、结构光传感器)。然而，部署此类传感器需要高昂的成本，并且获得的真实深度数据往往是稀疏且有缺陷的，这限制了深度估计方法的实际应用。单目深度估计无需额外感知即可估算给定二维图像中每个像素的深度值，打破了上述限制。单目深度估计是一个高度不适定问题，因为在同一个二维场景中可以投射出无限多的三维场景。因此，从单目图像中估算准确的深度是一个极具挑战性的问题。

由于深度估计的重要性和成本效益，许多研究[4-7]旨在提升单目深度估计的准确性。近年来，随着深度学习的发展成熟，卷积神经网络(convolutional neural networks, CNNs)被成功应用于深度估计领域。自 Eigen 等[8]利用深度学习完成单目深度估计任务开始，各种基于深度学习的完全监督方法被不断提出，如任务转换方法[9-10]、跨域监督方法[11-12]和基于 Transformer 的方法[13-15]。上述方法都需要真实深度进行监督学习。然而，在实际应用中，收集和构建具有地面实况的数据集是一项艰巨的任务。作为一种替代方法，自监督单目深度估计方法利用未标记的立体图像或单目视频进行自我监督，通过最小化光度损失来优化深度。尽管自监督单目深度估计方法的性能可与监督方法相媲美，但仍无法准确地对具有挑战性的图像区域(如薄结构区域或边界区域)进行深度预测。为了克服上述缺陷，Choi 等[16]引入了多任务学习方案，将语义感知纳入深度特征的表示中。Jung 等[17]提出了利用跨域信息(尤其是场景语义)改进自监督单目深度估计的新思路，并将工作的重点放在将隐含语义信息融入几何表示增强中。Li 等[18]提出了一种语义引导的深度表示增强方法，该方法通过利用丰富的上下文信息来促进局部和全局的深度特征表示。虽然上述方法通过利用语义信息来增强深度特征表示能力，改善了现有单目深度估计方法所存在的缺陷，但引入了额外的任务，即语义分割任务，在一定程度上增加了网络学习的难度。另外，语义分割任务需要对地面实况进行监督，这使自监督单目深度

收稿日期：2024-01-03；录用日期：2024-03-22；网络出版时间：2024-04-07 17:54
网络出版地址：link.cnki.net/urlid/11.2625.V.20240406.1021.001
*通信作者. E-mail：1418748495@qq.com

引用格式：张冬冬，王春平，付强. 基于 FastSAM 辅助表示增强的自监督单目深度估计 [J]. 北京航空航天大学学报，2026，52（3）：964-972.
ZHANG D D，WANG C P，FU Q. FastSAM-assisted representation enhancement for self-supervised monocular depth estimation [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（3）：964-972 （in Chinese）.

估计不再是真正意义上的自监督。

最近, Kirillov 等[19] 提出了第 1 个通用图像分割的基础模型 SAM, 并引起了计算机视觉界的极大关注。该模型在 1 100 万张图像和超过 10 亿个分割掩码的广泛数据集上进行训练, 在各种图像类别的“少镜头”甚至是“零镜头”学习方面具有巨大潜力。与语义分割模型相比, SAM 具有 3 个独特的属性: ①可以将提示(点、边界框等)作为输入, 分割出所需的目标区域; ②可以分割任何图像, 且可以适应任何场景; ③可以准确地对像素间的语义关系进行建模, 得到详尽的语义分割结果。然而, 尽管 SAM 具有广泛的普适性和强大的分割能力, 并在多种下游分割任务中取得了令人惊讶的结果, 但其实际应用仍然具有限制性。最突出的问题是 SAM 模型需要大量的计算资源, 阻碍了 SAM 的实际部署, 特别是在实时任务中的应用。针对这一问题, Zhao 等[20] 基于 YOLOv8-seg 提出了 FastSAM 模型, 该模型在单个 NVIDIA RTX 3090 上的运行速度比 SAM 快 50 倍。FastSAM 不仅大幅降低了计算和资源需求, 而且实现了与 SAM 相当的性能。此外, FastSAM 同 SAM 一样具有广泛的普适性, 可以用于多个下游分割任务。

受 SAM 和 FastSAM 的优缺点启发, 本文将 FastSAM 作为提供语义信息的源泉, 以期通过引入 FastSAM 来提高对具有挑战性图像区域的深度估计性能, 并消除因引入分割任务所带来的弊端。由于 FastSAM 具有强大的特征表示和实例分割能力, 无需对其进行额外的监督训练即可得到输入图像的全景实例分割结果。为有效利用 FastSAM 所提供的语义信息, 提升网络对薄结构区域和边界区域的深度估计能力, 本文设计了语义指导模块和边界引导模块。语义指导模块用于提高全局深度表示, 该模块不是简单地将语义特征与深度特征进行融合, 而是通过挖掘语义特征和深度特征的相关性来辅助增强深度特征表示。边界引导模块用于提高网络对边缘深度的感知能力, 该模块可以明确显示建模实例边缘信息, 引导深度网络更加关注边界区域来提高局部特征的表示能力, 进而提升模型对边界深度的估计能力。

1　自监督单目深度估计

自监督单目深度估计能够在没有地面实况的情况下, 预测单个 RGB 图像的深度图。具体来说, 给定输入图像三元组(I_{t-1}, I_t, I_{t+1}), 其中, I_t充当目标图像, $I_{t'}(t' \in \{t-1, t+1\})$是源图像。在训练期间, 目标图像$I_t$被送入深度网络中以预测其对应的深度$\boldsymbol{D}_t$, 而相邻图像对$(I_t, I_{t'})$被送入姿态网络以获得目标图像$I_t$与源图像$I_{t'}$之间的相对姿态$T_{t\to t'}$。源图像$I_{t'}$、深度$\boldsymbol{D}_t$和相对姿态$T_{t\to t'}$被用来估计合成的目标图像$\hat{I}_t$。自监督单目深度估计损失的计算公式为

$$L(I_t, \hat{I}_t) = \mu L_{\mathrm{p}}(I_t, \hat{I}_t) + \lambda L_{\mathrm{s}}(\boldsymbol{D}_t) \tag{1}$$

式中: μ充当静态掩膜, 当重投影的光度误差小于原光度误差时, μ为 1, 反之为 0; L_{p}为光度损失; L_{s}为深度平滑损失; λ为控制平滑度相对强度的因子。根据文献 [7], 光度损失可通过式(2)进行计算:

$$L_p(I_t, \hat{I}_t) = \alpha \frac{1 - \mathrm{SSIM}(I_t, \hat{I}_t)}{2} + (1-\alpha)\left\| I_t - \hat{I}_t \right\| \tag{2}$$

式中: SSIM为结构相似度[21]; α为加权因子, 通常设为 0.85。深度平滑损失[22] 用于确保深度估计中的边缘感知平滑, 计算公式如下:

$$L_{\mathrm{s}}(\boldsymbol{D}_t) = |\nabla_x \boldsymbol{D}_t| \mathrm{e}^{-|\nabla_x I_t|} + \left|\nabla_y \boldsymbol{D}_t\right| \mathrm{e}^{-|\nabla_y I_t|} \tag{3}$$

式中: ∇_x和∇_y分别表示 x 和 y 轴上的梯度运算。

2　本文方法

本节主要介绍基于 FastSAM 辅助表示增强的自监督单目深度估计网络架构, 该网络主要由 2 部分组成, 即 FastSAM 和深度估计网络, 其整体架构如图 1 所示。深度估计网络遵循 U-Net 结构[23], 主要由编码器、跳跃连接、解码器组成。编码器用于从图像中提取特征, 解码器用于恢复空间分辨率, 而跳跃连接则是为了促进信息的流动, 避免细节信息的丢失。为提高网络对薄结构区域和边界区域的深度估计性能, 本文在解码过程中引入语义指导模块(semantic guidance module, SGM)和边界引导模块(edge guidance module, EGM), 以充分利用 FastSAM 所提供的语义信息。语义指导模块可以嵌入多个解码过程中, 通过将语义信息转换为深度网络的低层特征来改进深度特征表示, 进而增强全局深度特征表示。边界引导模块通过实例边界信息引导网络关注局部特征, 以促进边界感知和深度推理。

2.1　语义信息

单目深度估计的目标是从单个 RGB 图像预测深度图, 而无需地面实况。多个研究工作[16-18] 已经证明, 语义信息对单目深度估计任务是有益的, 因为语义信息能够提供丰富的先验知识, 并有助于理解空间上下文。FastSAM 是一个以 YOLOv8-seg 为基础的快速实例分割模型, 可以对“零镜头”样本进行全景实例分割, 并能获得高质量的分割结果。FastSAM 通过主干网络和特征金字塔从图像中提取特征, 并生成多种尺度的特征, 这些特征包含了

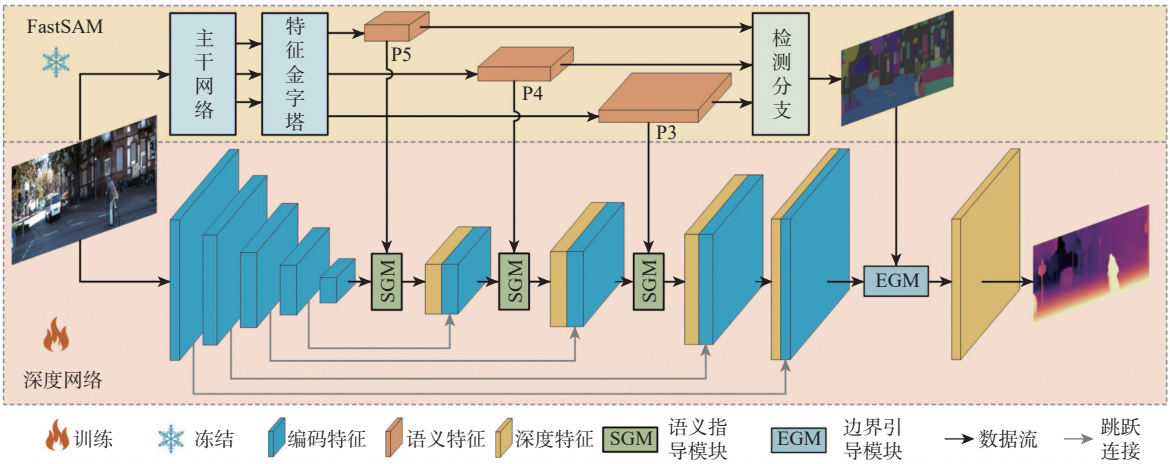


图1 所提方法的整体框架

Fig. 1 Overall framework of the proposed methodology

丰富的语义信息。因此，本文使用FastSAM来提供语义信息，并期望通过利用这些语义信息来应对深度估计任务中具有挑战性的图像区域所带来的挑战。与先前工作[16-18,24-25]相比，本文方法不需要为了获得语义信息而引入额外任务，这在一定程度上降低了网络学习的难度。另外，由于本文方法在训练过程中不需要对分割的地面实况进行监督，因此，更加实用。

2.2 语义指导模块

FastSAM可以为深度估计网络提供丰富的语义信息，增强网络对场景结构的理解，以准确地获得三维空间中的整体结构和相对深度信息，进而提升深度估计的性能。然而，简单的融合操作忽略了不同类型特征之间的语义差距，缺少对特征之间关系的探索。因此，本文提出SGM，该模块探索和利用深度特征和语义特征之间的相关性，通过语义信息增强全局特征表示，以获得更加准确的深度估计。SGM具体结构如图2所示。

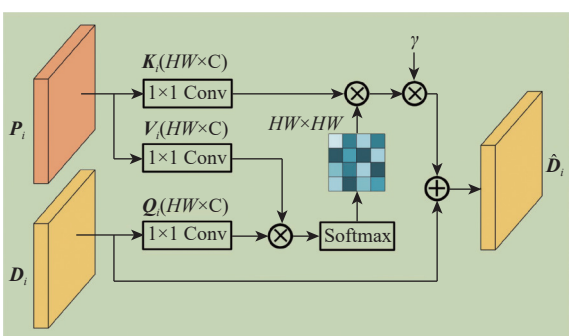


图2 语义指导模块

Fig. 2 Semantic guidance module

语义指导模块以语义特征图$\boldsymbol{P}_i\in\mathbf{R}^{H\times W\times C}$和深度特征图$\boldsymbol{D}_i\in\mathbf{R}^{H\times W\times C}$为输入，通过一系列操作后输出深度增强特征图$\hat{\boldsymbol{D}}_i\in\mathbf{R}^{H\times W\times C}$，其中，$C$表示通道数，$i\in\{5,4,3\}$。首先，使用不同的1×1卷积将$\boldsymbol{P}_i$转换为关键特征$\boldsymbol{K}_i\in\mathbf{R}^{HW\times C}$和值特征$\boldsymbol{V}_i\in\mathbf{R}^{HW\times C}$，并将$\boldsymbol{D}_i$转换为查询特征$\boldsymbol{Q}_i\in\mathbf{R}^{HW\times C}$。然后，在值特征$\boldsymbol{V}_i$和查询特征$\boldsymbol{Q}_i$的转置之间执行矩阵乘法，并通过Softmax函数获得相关性矩阵$\boldsymbol{A}_i\in\mathbf{R}^{HW\times HW}$。$\boldsymbol{A}_i$可以计算如下：

$$\boldsymbol{A}_i=\mathrm{softmax}(\boldsymbol{V}_i\times\boldsymbol{Q}_i^{\mathrm{T}}) \tag{4}$$

为传递像素间的语义相关性，在相关性矩阵$\boldsymbol{A}_i$和关键特征$\boldsymbol{K}_i$之间进行矩阵乘法运算。通过以下方式将语义信息从$\boldsymbol{P}_i$传递给$\boldsymbol{D}_i$：

$$\hat{\boldsymbol{D}}_i=\gamma(\boldsymbol{A}_i\times\boldsymbol{K}_i)+\boldsymbol{D}_i \tag{5}$$

式中：γ为可学习的权重因子。如图1所示，将语义指导模块嵌入3个解码层(层5、4、3)中，实现了对多层深度特征的增强。

2.3 边界引导模块

在深度估计任务中，边界区域比普通区域更值得关注。一方面，边界区域包含丰富的几何结构信息，有助于提高局部特征的表示能力；另一方面，如果网络清楚地知道描述对象边界特征的类别和位置，则可以很容易地得到准确的深度估计结果。为了使网络更好地捕捉边界区域，并提高特征的表示能力，提出EGM，其可以引导网络更加关注局部特征，具体结构如图3所示。

基于FastSAM输出的全景实例分割结果，利用

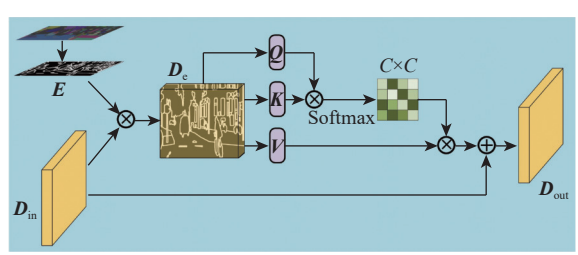


图3 边界引导模块

Fig. 3 Edge guidance module

Sobel 算子得到边界图 $\boldsymbol{E}\in\mathbf{R}^{H\times W\times 1}$。通过逐像素相乘融合边界图 $\boldsymbol{E}$ 和深度特征图 $\boldsymbol{D}_{\text{in}}\in\mathbf{R}^{H\times W\times C}$, 以获得面向边界的特征映射 $\boldsymbol{D}_{\text{e}}\in\mathbf{R}^{H\times W\times C}$。借助自注意力机制对边界区域的特征进行建模, 以鼓励网络关注局部区域特征。对于给定的融合特征, 自注意力模块利用不同的卷积计算得到查询特征、关键特征和值特征 $\{\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\}\in\mathbf{R}^{C\times HW}$, 并通过式 (6) 所示过程得到具有值特征向量的边缘增强特征图 $\boldsymbol{E}'$。融合边缘增强的特征图 $\boldsymbol{E}'$ 和 $\boldsymbol{D}_{\text{in}}$ 以获得输出特征图 $\boldsymbol{D}_{\text{out}}$。

$$\begin{cases}\boldsymbol{E}'=\text{softmax}(\boldsymbol{Q}\times\boldsymbol{K}^{\text{T}})\times\boldsymbol{V}\\ \boldsymbol{D}_{\text{out}}=\boldsymbol{E}'+\boldsymbol{D}_{\text{in}}\end{cases}\tag{6}$$

根据文献 [7], 使用光度损失和深度平滑损失对网络进行自监督训练。因此, 类似式(1), 整体损失函数如下:

$$L=\mu L_{\text{p}}+\lambda L_{\text{s}}\tag{7}$$

在本文中, 训练时 λ 设为 0.001。

3　实　验

3.1　实验设置

使用 KITTI 数据集 [26] 进行网络的训练和评估。为保证实验的公平性, 采用 Eigen 等 [8] 的分割方法, 并使用 Zhou 等 [27] 的预处理方法去除静态帧, 最终得到 39 810 幅训练图像、4 424 幅验证图像及 697 幅评估图像。

使用带有 Imagenet 预训练权重的 ResNet-50 作为骨干网络。所有实验均在配置为 E5-2630v4 CPU、RTX 3 090 GPU(24 GB video memory)、64 GB RMB 的图像工作站上进行, 以深度学习框架 Pytorch 为编译工具完成。网络的输入图像大小为 192 像素×640 像素, 共训练 20 轮, 批量大小是 12, 初始学习率为 1.5×10^{-4}, 在训练到第 15 轮时, 学习率下降 10 倍。

3.2　评价指标

为了定量评估模型性能, 根据之前的研究 [7-8,16], 本文采用以下 5 个常用的评价指标:

1)相对绝对误差 Abs Rel: $\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{|d'_i-d_i|}{d_i}$

2)相对平方误差 Sq Rel: $\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{\|d'_i-d_i\|^2}{d_i}$

3)均方根误差 RMS: $\sqrt{\dfrac{1}{N}\sum_{i=1}^{N}\|d'_i-d_i\|^2}$

4)均方根对数误差 RMS log:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\log d'_i-\log d_i\right\|^2}$$

5)阈值精度 δ_k:

$$\max\left(\frac{d'_i}{d_i},\frac{d_i}{d'_i}\right)=\delta_k<1.25^k\qquad k=1,2,3$$

式中: N 为测试集中像素的总数; d'_i 和 d_i 分别为像素 i 的估计深度值和地面实况深度值。

3.3　实验结果

3.3.1　对比实验

为了验证本文方法的有效性, 在 KITTI 数据集上将本文方法与最先进的自监督模型进行比较。为保证比较的公平性, 所有模型均在图像大小为 192 像素×640 像素下进行测试, 定量结果如表 1 所示。可以看出, 本文方法在 Sq Rel、Abs Rel、RMS log、k 取不同值时的阈值精度 δ_1、δ_2 和 δ_3 指标上取得了最高性能。另外, 本文方法在 RMS 指标上也取得了第 2 的好成绩, 且与最好方法相比差距很小。值得注意的是, 本文方法在评价指标 δ_1 上的得

表 1　不同方法在 KITTI 数据集上的定量检测结果

Table 1　Quantitative detection results of different methods on the KITTI dataset

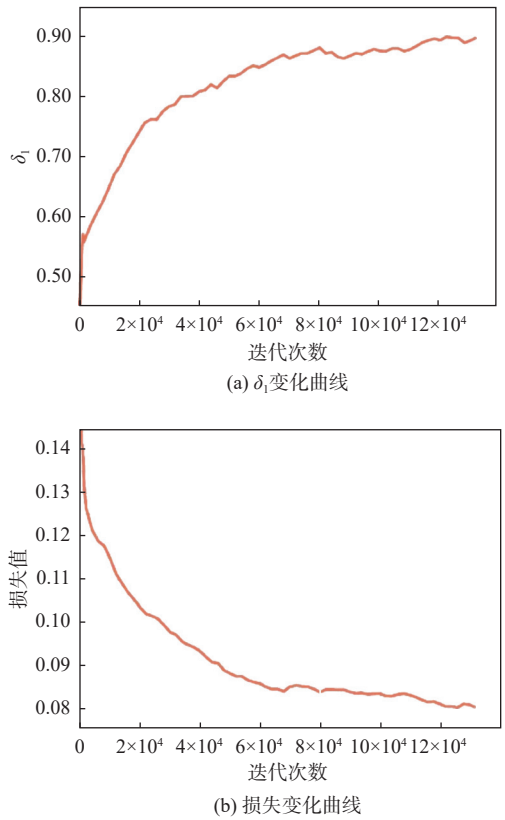
方法	监督方式	Abs Rel↓	Sq Rel↓	RMS↓	RMS log↓	δ_1↑	δ_2↑	δ_3↑
Monodepth2[7]	M	0.115	0.903	4.863	0.193	0.877	0.959	0.981
SAFENet[16]	M+Se	0.112	0.788	4.582	0.187	0.878	0.963	0.983
SGDepth[25]	M+Se	0.113	0.835	4.693	0.191	0.879	0.961	0.981
SGRLDepth[28]	M	0.102	0.698	**4.381**	**0.178**	0.896	0.964	**0.984**
WaveletMonodepth[29]	S	0.106	0.824	4.824	0.205	0.870	0.949	0.975
FSREDepth[17]	M+Se	0.105	0.722	4.547	0.182	0.886	0.964	**0.984**
HRDepth[22]	M	0.109	0.792	4.632	0.185	0.884	0.962	0.983
DepthHints[30]	S	0.102	0.762	4.602	0.189	0.880	0.960	0.981
VADepth[31]	M	0.109	0.785	4.624	0.190	0.875	0.960	0.982
DIFFNet[2]	M	0.102	0.764	4.483	0.180	0.896	**0.965**	0.983
本文	M	**0.101**	**0.688**	4.383	**0.178**	**0.900**	**0.965**	**0.984**

注: "↓" 表示数值越低越好, "↑" 表示数值越高越好, M表示单目自监督, S表示立体对自监督, Se表示语义监督。最优结果用粗体表示, 次优结果添加了下划线。

分最高，这表明，相较于其他模型，本文方法所预测的深度更加详细和准确。总的来说，本文方法在KITTI数据集上取得了最优异的结果，且在所有评价指标上都具有较强的竞争力。

为了直观地展示本文网络的训练过程，利用可视化工具对δ_1和损失函数值的变化进行了可视化，变化曲线如图4所示。可以看出，本文网络在前10轮中收敛较快，在后10轮中逐步收敛到稳定，并且δ_1的值稳定在0.90左右。综上所述，本文网络经过20轮的训练后可以稳定收敛，并取得最佳性能。

为了更加直观地展示本文方法的优越性，将其与Monodepth2、FSREDepth及DIFFNet进行定性比较，定性结果如图5所示。在图5的第1行，本文方法能够准确地估计较小物体的深度。在第2行和第3行，对路标和电线杆等具有较薄结构的对象，本文方法能够进行清晰的深度估计，而其他模型无法准确估计。在第4行中，本文方法精确还原了人体的深度轮廓，对人体各部位（如头部、腿部）的深度也能很好地估计出来。其他模型无法完整检测出人体的所有部位，且检测出部位的深度值不均匀。在最后一行，本文模型正确区分了同一语义对象的不同实例，并成功地估计出了连续对象（汽车）各自的深度。而其他模型在面对连续对象时可能会造成混乱，将多个对象误判为同一对象。综上所述，定性结果表明，与其他方法相比，本文模型能够获得更准确的深度预测，特别是在对象的边界和细节上。



(a) δ_1变化曲线

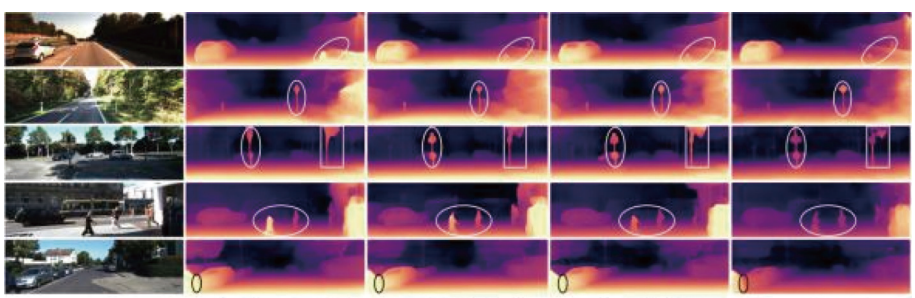
(b) 损失变化曲线

图4 训练过程可视化

Fig. 4 Visualization of training process

3.3.2 消融实验

为了进一步验证本文模块的有效性，通过消融实验分析各模块对模型性能提升的贡献。消融实验的定量结果如表2所示。可以看出，与基线模型相比，增加语义指导模块或边界引导模块的模型在几乎所有评价指标上都实现了稳步提升。同时拥有语义指导模块和边界引导模块的模型性能达到了最佳，这表明2个模块可以共存，并能协同改善网络性能。



(a) 输入图像 (b) Monodepth2 (c) FSREDepth (d) DIFFNet (e) 本文方法

图5 KITTI数据集上可视化结果

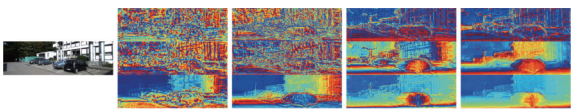
Fig. 5 Visualization results on KITTI dataset

表2 消融实验结果

Table 2 Results of ablation experiments

模型	Abs Rel↓	Sq Rel↓	RMS↓	RMS log↓	δ_1↑	δ_2↑	δ_3↑
Baseline	0.109	0.816	4.605	0.186	0.878	0.962	0.983
Baseline+语义指导模块	0.102	0.709	4.436	0.180	0.893	0.964	**0.984**
Baseline+边界引导模块	0.105	0.743	4.501	0.183	0.886	0.963	0.983
本文	**0.101**	**0.688**	**4.383**	**0.178**	**0.900**	**0.965**	**0.984**

为直观地展示 SGM 和 EGM 所发挥的效用, 对 2 个模块的输入和输出分别进行了可视化, 如图 6 所示。图 6(a)是输入图像, 图 6(b)和图 6(c)分别是语义指导模块的输入深度特征图 $\boldsymbol{D}_3$ 和输出深度特征图 $\hat{\boldsymbol{D}}_3$, 图 6(d)和图 6(e)分别是边界引导模块的输入特征图 $\boldsymbol{D}_{\text{in}}$ 和输出特征图 $\boldsymbol{D}_{\text{out}}$。通过对比图 6(b)和图 6(c)可以发现, 经过 SGM 的处理, 特征图包含更多的细粒度特征, 表征能力得到增强。同时, 由于受 FastSAM 的影响, 对象间的区分更加明确。通过对比图 6(d)和图 6(e)不难发现, 输出特征图中对象的边界显著性得到增强, 丰富了局部细节信息。



(a) 输入图像　(b) $\boldsymbol{D}_3$　(c) $\hat{\boldsymbol{D}}_3$　(d) $\boldsymbol{D}_{\text{in}}$　(e) $\boldsymbol{D}_{\text{out}}$

图 6　特征图可视化示例

Fig. 6　Example of feature map visualization

3.3.3　扩展实验

1) 语义指导模块嵌入方案评估。由于 FastSAM 能够产生 3 种不同尺度的语义特征, 在网络构建时可以利用其中的 1 个或多个来对深度网络的解码过程施加影响。因此, SGM 有多种嵌入方案。以嵌入数量为划分依据, 对不同嵌入数量的方案进行评估。表 3 展示了不同嵌入方案的定量结果, 数字 0、1、2、3 表示所使用 SGM 的数量, 值得注意的是, 嵌入数量不为 0 的情况下, 均以 P5 为基础进行组合, 当嵌入 2 个语义指导模块时, 所使用的语义层输入可以是 P5+P4 的组合或 P5+P3 的组合。随着嵌入数量的增加, SGM 可以充分利用不同模态的信息, 增强深度特征的表示能力。为进一步说明不同嵌入方案的差异性, 将各方案的深度估计结果进行了可视化, 如图 7 所示。通过比较图 7(b)～图 7(f)可以直观地看出, 使用 SGM 的数量越多, 深度估计结果中对象的轮廓越清晰, 边界也越精确。综上所述, 基于本文所使用的嵌入方案, 能够更加充分地利用语义信息, 并获得更为准确的深度预测结果。

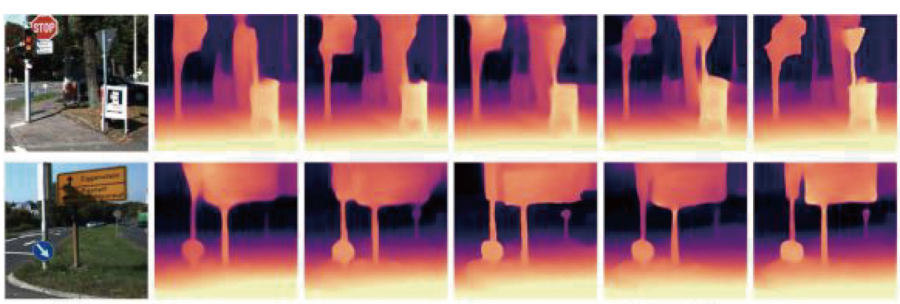
2) 失败案例。尽管本文方法在 KITTI 数据集上取得了极具竞争力的表现, 但仍存在一些失败案例。图 8 展示了本文方法的 2 种典型失败案例: ①本文方法无法预测远处对象的深度, 主要是因为远处对象所包含的特征较少, 在特征编码的过程中, 这部分特征极易丢失, 且未能在解码过程中进行很好的恢复。②本文方法误将密集排列的对象视为一个整体, 把多个对象预测成同一深度, 究其原因是网络对复杂排列对象的区分能力不足, 模糊的层次结构引起了网络的混乱。通过观察图 8(c)不难发现, FastSAM 无法识别出远处对象及密集排列对象, 影响深度网络对场景结构和局部细节的感知, 导致网络无法很好地预测这 2 类对象的深度。未来, 将主要聚焦于 2 个方面: ①探索通过 FastSAM 来丰富深度网络细粒度特征的方法, 以提升网络的鲁棒性; ②构建更加合理的联合机制, 以增加 FastSAM 的效益, 并降低其带来的负面影响。

3) 模型泛化能力评估。为了评估本文模型的泛化能力, 在 Make3D 数据集[32] 上对其进行测试。

表 3　语义指导模块不同嵌入方案的定量结果

Table 3　Quantitative results of different embedding schemes for the semantic guidance module

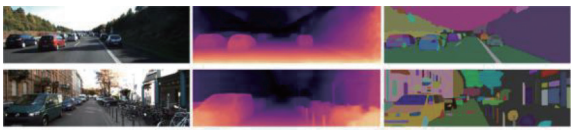
数量	Abs Rel↓	Sq Rel↓	RMS↓	RMS log↓	δ_1↑	δ_2↑	δ_3↑
0	0.109	0.816	4.605	0.186	0.878	0.962	0.983
1(P5)	0.106	0.783	4.545	0.184	0.884	0.963	0.983
2(P5+P4)	0.104	0.761	4.507	0.182	0.887	**0.964**	**0.984**
2(P5+P3)	0.103	0.743	4.474	0.182	0.889	0.963	**0.984**
3(P5+P4+P3)	**0.102**	**0.709**	**4.436**	**0.180**	**0.893**	**0.964**	**0.984**



(a) 输入　(b) 0　(c) 1(P5)　(d) 2(P5+P4)　(e) 2(P5+P3)　(f) 3(P5+P4+P3)

图 7　不同嵌入方案的可视化结果

Fig. 7　Visualization results of different embedding schemes



(a) 输入图像 (b) 输出结果 (c) FastSAM分割结果

图 8 失败案例

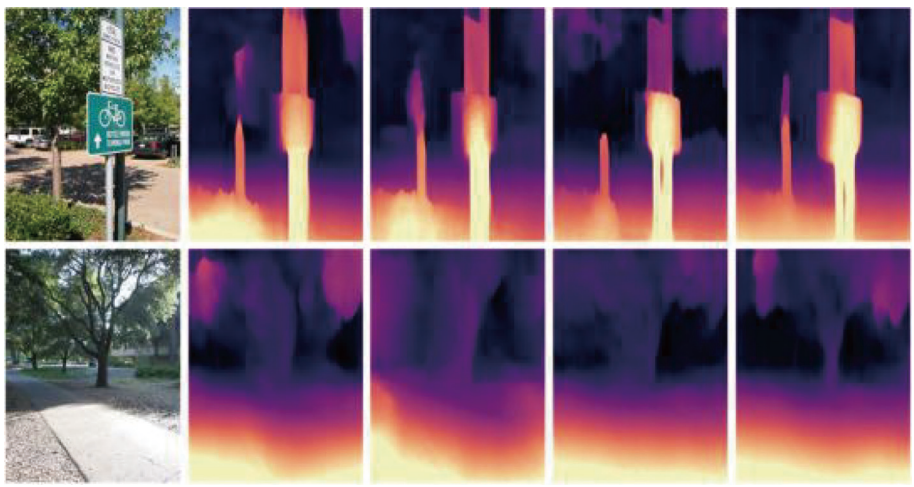
Fig. 8 Failure cases

该数据集包含 134 幅室外场景测试图像。在这些图像上直接加载并推断在 KITTI 数据集上训练的模型。如表 4 所示，与其他先进的自监督方法相比，本文方法获得了更加优越的结果。可视化结果如图 9 所示，本文模型即使在非均匀光照条件下，也能够较为准确地估计出薄结构物体和低纹理区域的深度。在图 9 第 1 行中，与其他模型相比，本文模型对指示牌、树干及灯杆的深度预测结果更加完整，深度值更加均匀，通过颜色差异更加准确地描述了物体的空间分布。在图 9 第 2 行中，本文模型能够清晰地识别出背景中淡化的绿色植物，并很好地与背景进行了区分。

表 4 不同方法在 Make3D 数据集上的定量检测结果

Table 4 Quantitative detection results of different methods on the Make3D dataset

方法	Abs Rel↓	Sq Rel↓	RMS↓	MS logR↓
Monodepth2[7]	0.322	3.589	7.417	0.163
FSREDepth[17]	0.296	2.802	7.003	0.138
DIFFNet[2]	0.287	2.984	6.808	0.133
本文	**0.284**	**2.655**	**6.503**	**0.128**



(a) 输入图像 (b) Monodepth2 (c) FSREDepth (d) DIFFNet (e) 本文方法

图 9 Make3D 数据集上的可视化结果

Fig. 9 Visualization results on the Make3D dataset

4 结 论

1）本文方法在 KITTI 数据集上取得了优异的性能，例如 Abs Rel 为 0.101，Sq Rel 为 0.688，RMS log 为 0.178，δ_1 精度达到 90.0%，δ_2 精度达到 96.5%，δ_3 精度达到 98.4%，在 Sq Rel、Abs Rel、RMS log、δ_1、δ_2 及 δ_3 指标上均优于现有最先进方法。

2）本文提出的语义指导模块和边界引导模块能够协同改善网络性能，消融实验表明，相比基线模型（Abs Rel 为 0.109，δ_1 为 87.8%），同时引入两个模块后 Abs Rel 降至 0.101，δ_1 提升至 90.0%，Sq Rel 从 0.816 降至 0.688。

3）本文方法采用 FastSAM 提供语义信息而无需额外监督，是一种真正意义上的自监督方法，语义指导模块嵌入数量从 1 个增加至 3 个（P5+P4+P3 组合）时，Abs Rel 从 0.106 逐步降至 0.102，δ_1 从 88.4% 提升至 89.3%。

4）本文方法在 Make3D 数据集上展现出良好的泛化能力，Abs Rel 为 0.284，Sq Rel 为 2.655，RMS 为 6.503，均优于 Monodepth2、FSREDepth 和 DIFFNet 等方法。

5）本文方法对远处对象和密集排列对象的深度预测仍存在不足，主要原因是 FastSAM 无法有效识别这两类对象，进而影响深度网络对场景结构和局部细节的感知。

为使本文提出的方法能够处理更具挑战性的场景，未来工作需要探索通过 FastSAM 丰富深度网络细粒度特征的方法，并构建更加合理的联合机制以增加 FastSAM 的效益。

参考文献（References）

[1] 柴国强, 薄祥仕, 刘海军, 等. 基于不确定性单目图像自监督场景深度估计[J]. 北京航空航天大学学报, 2024, 50(12): 3780-3787.
CHAI G Q, BO X S, LIU H J, et al. Self-supervised monocular depth estimation based on uncertainty[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(12): 3780-3787.

[2] ZHOU H, GREENWOOD D, TAYLOR S. Self-supervised monocular depth estimation with internal feature fusion[EB/OL]. (2023-11-20)[2021-10-18]. https://doi.org/10.48550/arXiv.2110.09482.

[3] 张冬冬, 王春平, 付强. 基于人眼视觉机制的伪装目标检测网络[J]. 北京航空航天大学学报, 2025, 51(7): 2553-2561.
ZHANG D D, WANG C P, FU Q. Camouflaged object detection network based on human visual mechanisms[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2553-2561 (in Chinese).

[4] LEE J H, HAN M, KO D W, et al. From big to small: Multi-scale local planar guidance for monocular depth estimation[EB/OL]. (2023-11-21)[2019-07-24]. https://doi.org/10.48550/arXiv.1907.10326.

[5] PATIL V, SAKARIDIS C, LINIGER A, et al. P3Depth: monocular depth estimation with a piecewise planarity prior[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 1600-1611.

[6] LIU C, KUMAR S, GU S, et al. Va-depthnet: A variational approach to single image depth prediction[EB/OL]. (2023-11-22)[2023-02-13]. https://doi.org/10.48550/arXiv.2302.06556.

[7] GODARD C, MAC AODHA O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 3827-3837.

[8] EIGEN D, PUHRSCH C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[C]//Proceedings of Advances in Neural Information Processing Systems. Vancouver: NeurIPS, 2014.

[9] FAROOQ BHAT S, ALHASHIM I, WONKA P. AdaBins: depth estimation using adaptive bins[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 4008-4017.

[10] BHAT S F, ALHASHIM I, WONKA P. Localbins: improving depth estimation by learning local distributions[C]//Computer Vision–ECCV 2022. Berlin: Springer, 2022: 480-496.

[11] GUO X Y, LI H S, YI S, et al. Learning monocular depth by distilling cross-domain stereo networks[C]//Computer Vision–ECCV 2018. Berlin: Springer, 2018: 506-523.

[12] ZHAO S S, FU H, GONG M M, et al. Geometry-aware symmetric domain adaptation for monocular depth estimation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 9780-9790.

[13] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 12159-12168.

[14] KIM D, KA W, AHN P, et al. Global-local path networks for monocular depth estimation with vertical cutdepth[EB/OL]. (2023-11-30)[2022-01-19]. https://doi.org/10.48550/arXiv.2201.07436.

[15] ZHANG N, NEX F, VOSSELMAN G, et al. Lite-mono: a lightweight CNN and transformer architecture for self-supervised monocular depth estimation[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 18537-18546.

[16] CHOI J, JUNG D, LEE D, et al. SAFENet: self-supervised monocular depth estimation with semantic-aware feature extraction [EB/OL]. (2023-12-01)[2020-08-06]. https://doi.org/10.48550/arXiv.2010.02893.

[17] JUNG H, PARK E, YOO S. Fine-grained semantics-aware representation enhancement for self-supervised monocular depth estimation[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 12622-12632.

[18] LI R, MAO Q, WANG P, et al. Semantic-guided representation enhancement for self-supervised monocular trained depth estimation [EB/OL]. (2023-12-03)[2020-12-15]. https://doi.org/10.48550/arXiv.2012.08048.

[19] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [EB/OL]. (2023-12-04)[2024-04-05]. https://doi.org/10.48550/arXiv.2304.02643.

[20] ZHAO X, DING W C, AN Y Q, et al. Fast segment anything [EB/OL]. (2023-12-05)[2023-06-21]. https://doi.org/10.48550/arXiv.2306.12156.

[21] PENG R, WANG R, LAI Y, et al. Excavating the potential capacity of self-supervised monocular depth estimation[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 15540-15549.

[22] LYU X Y, LIU L, WANG M M, et al. HR-Depth: high resolution self-supervised monocular depth estimation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(3): 2294-2301.

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

[24] NEKRASOV V, DHARMASIRI T, SPEK A, et al. Real-time joint semantic segmentation and depth estimation using asymmetric annotations[C]//Proceedings of the 2019 International Conference on Robotics and Automation. Piscataway: IEEE Press, 2019: 7101-7107.

[25] KLINGNER M, TERMÖHlen J A, MIKOLAJCZYK J, et al. Self-supervised monocular depth estimation: solving the dynamic object problem by semantic guidance[C]//Computer Vision–ECCV 2020. Berlin: Springer, 2020: 582-600.

[26] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 3354-3361.

[27] ZHOU T H, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6612-6619.

[28] GUIZILINI V, HOU R, LI J, et al. Semantically-guided representation learning for self-supervised monocular depth[EB/OL]. (2023-12-15)[2020-02-27]. https://doi.org/10.48550/arXiv.2002.12319.

[29] RAMAMONJISOA M, FIRMAN M, WATSON J, et al. Single image depth prediction with wavelet decomposition[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pat-

tern Recognition. Piscataway: IEEE Press, 2021: 11084-11093.

[30] WATSON J, FIRMAN M, BROSTOW G, et al. Self-supervised monocular depth hints[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 2162-2171.

[31] XIANG J, WANG Y, AN L F, et al. Visual attention-based self-supervised absolute depth estimation using geometric priors in autonomous driving[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 11998-12005.

[32] SAXENA A, SUN M, NG A Y. Make3D: learning 3D scene structure from a single still image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824-840.

FastSAM-assisted representation enhancement for self-supervised monocular depth estimation

ZHANG Dongdong[1,*], WANG Chunping[2], FU Qiang[2]

(1. College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China;
2. Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China)

Abstract: In order to address the issue of unsatisfactory performance in self-supervised monocular depth estimation methods when applied to thin structured regions and boundary regions, this paper proposes a method for self-supervised monocular depth estimation based on FastSAM-assisted representation enhancement. Firstly, without the need for extra supervision, FastSAM is presented to supply the depth network with rich semantic information. Secondly, a semantic guidance module (SGM) is proposed to explore the correlation between semantic features and depth features, and to enhance the global feature representation. Furthermore, to enhance the performance of boundary depth estimation, a edge guiding module (EGM) is built to direct the network to focus more on local features. Extensive experiments show that the proposed method outperforms the state-of-the-art methods, especially in depth estimation of thin-structured regions and boundary regions.

Keywords: depth estimation; self-supervised; FastSAM; semantic information; feature representation

Received: 2024-01-03; **Accepted**: 2024-03-22; **Published Online**: 2024-04-07 17:54

URL: link.cnki.net/urlid/11.2625.V.20240406.1021.001

* **Corresponding author**. E-mail: 1418748495@qq.com

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部

办公地点:北京航空航天大学办公楼西配楼412~416房间

电　　话:(010)82316698,82316981,82315499,82317448

E-mail: jbuaa@ buaa. edu. cn

http://bhxb. buaa. edu. cn

http://www. buaa. edu. cn

IISTJA 工信科技期刊集群 集群成员期刊
2023 中国精品科技期刊 OUTSTANDING SCI JOURNALS OF CHINA
2024中国高校杰出科技期刊

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

编辑部主任：张　嵘

编　　辑：孙　芳　万玉敏　王　茜　卞欢欢　李亚泰

北京航空航天大学学报
Beijing Hangkong Hangtian Daxue Xuebao
（原《北京航空学院学报》）
（月刊　1956 年创刊）
第52卷 第3期　2026年3月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.52 No.3　March 2026

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100191
地　　址　北京市海淀区学院路37号
印　　刷　北京科信印刷有限公司
发　　行　《北京航空航天大学学报》编辑部
发行范围　国内外发行
联系电话　(010) 82316698 82316981
82315499 82317448
电子信箱　jbuaa@ buaa.edu.cn

Administrated by Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100191, P. R. China)
Chief Editor ZHAO Qinping
Edited and Published by Editorial Board of JBUAA
Printed by Beijing Kexin Printing Co., Ltd.
Distributed by Editorial Board of JBUAA
Telephone (010) 82316698 82316981
82315499 82317448
E-mail jbuaa@buaa.edu.cn
http: //bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V

ISSN 1001-5965
03>

微信公众号　超星电子书

国内定价：50.00元／期