基于深度学习的航天器自主指挥控制策略研究

赵唯 柴朔 张杰[通讯作者]

北京宇航系统工程研究所，北京市，100076；

摘要：随着航天任务复杂度的持续提升与深空探测的不断推进，航天器自主指挥控制技术已成为航天领域的前沿研究方向。本文提出一种融合深度强化学习与模型预测控制的航天器自主指挥控制策略，通过构建多模态感知-决策-执行一体化控制框架，实现复杂空间环境下的自主任务规划、动态避障及故障容错控制。研究首先剖析航天器控制面临的核心挑战，包括环境不确定性、实时性约束及星载资源限制；继而设计基于多模态特征金字塔网络的感知架构、鲁棒对抗强化学习决策模型与显式模型预测控制的混合框架；最终通过 STK 仿真平台构建闭环验证环境。实验结果表明，所提策略在轨道机动与故障容错场景下的控制精度较传统方法提升 20%以上，燃料消耗降低 15%，为深空探测任务的工程实现提供了理论支撑与技术路径。

关键词：航天器控制；强化学习；多模态融合

DOI：10.69979/3060-8767.25.05.038

引言

航天器自主指挥控制技术是实现深空探测、在轨服务等复杂航天任务的核心使能技术。传统控制方法依赖地面预编程指令序列与实时遥测遥控，在面对地月通信延迟（约 2.5s 往返时延）、非球形引力场建模误差（如小行星引力场误差可达 15%）及执行器突发故障时，难以满足毫秒级实时决策与高可靠性控制需求。近年来，深度学习技术凭借其强大的非线性映射能力与数据驱动优化特性，为航天器控制提供了全新解决方案：卷积神经网络可自动提取多源异构传感器数据的层次化特征表示，强化学习通过与环境的交互实现控制策略的动态优化，而模型预测控制则为多约束动态优化问题提供了高效求解框架。

针对航天器控制中环境不确定性（如空间碎片分布动态变化、行星表面地形非结构化）、实时性要求（交会对接需≤5ms 决策延迟）与资源约束（星载 CPU 算力≤10GFLOPS，能源供应≤50W）三大核心挑战，本文提出一种混合智能控制架构。该架构通过多模态感知网络融合视觉、惯性与雷达数据，构建包含空间几何、光谱特性与运动状态的统一环境表征；利用鲁棒对抗强化学习算法提升系统对故障与干扰的容错能力；结合显式模型预测控制（EMPC）实现控制输入的实时约束优化，确保指令的物理可执行性。

1 相关工作

1.1 航天器自主控制技术研究进展

航天器自主控制技术经历了从有限状态机到智能决策系统的演进。早期 DS-1 探测器采用基于规则的预编程控制，仅能处理有限工况。随着模型预测控制（MPC）的发展，其滚动时域优化能力在卫星编队重构中实现 12%的燃料节省，但依赖精确模型的缺陷在强不确定性环境中导致控制偏差可达标称值 30%[1]。

1.2 深度学习在航天器控制中的应用现状

深度学习在航天器控制中展现多维度优势：

环境感知：CNN 视觉导航对月面≥20cm 坑洞识别准确率达 97.3%，高光谱迁移学习实现火星土壤 89.6%分类精度[2-3]。

自主决策：西北工业大学团队提出的多智能体强化学习算法，使轨道追逃博弈目标捕获时间缩短 25%[4]。

故障容错：天津大学研发的鲁棒对抗强化学习框架，将姿态控制故障稳定时间缩短 40%[5]。

1.3 当前技术瓶颈与研究缺口

现有技术存在四大挑战：

模型泛化不足：空间辐射导致传感器噪声使特征提取精度下降 20%-30%[6]；

实时性瓶颈：ResNet-50 单步推理耗时 25ms，远超交会对接任务 5ms 的阈值要求[7]；

决策不可解释：深度强化学习的黑箱特性难以满足载人航天功能安全标准[8]；

数据融合低效：传统融合方法导致障碍物定位误差增加 18%[9]。

2 基于深度学习的航天器自主控制策略设计

2.1 系统总体架构

本文提出的三层分布式控制架构如图 1 所示，通过标准化接口实现模块交互：

感知层：集成高光谱相机、激光雷达与星敏感器，

经多模态特征金字塔网络（MFPN）处理后，输出 512 维环境表征向量。

决策层：鲁棒对抗强化学习（RA-RL）生成初始控制指令，结合显式模型预测控制（EMPC）完成约束校验与修正。

执行层：驱动推进系统与姿态机构，通过 IMU 和星敏感器形成控制闭环。

注：架构包含感知层多模态融合、决策层混合控制与执行层闭环反馈，箭头标注数据流与控制信号流向

2.2 多模态感知网络设计

2.2.1 跨模态特征融合

设计三分支编码器处理视觉（CNN）、点云（PointNet）与惯性数据（LSTM），通过自注意力机制计算模态权重矩阵\(\mathbf{W}_m\in\mathbb{R}^{3\times1}\)，实现动态加权融合：\(\mathbf{F}_{\text{fusion}}=\sum_{m=1}^3\mathbf{W}_m\odot\mathbf{F}_m\)

其中\(\mathbf{F}_m\)为各模态特征，\(\odot\)表示逐元素乘积。该机制可根据环境动态调整模态权重，提升复杂光照与遮挡条件下的感知鲁棒性。

2.2.2 时序校准与轻量化优化

引入双向 LSTM 处理 100ms 内的特征序列，通过状态转移矩阵\(\mathbf{U}\)抑制传感器噪声；采用通道剪枝与深度可分离卷积，将模型参数量降至 2.3MB，同时保持地形识别准确率≥98.5%。轻量化设计满足星载算力限制，单帧推理时间控制在 2ms 以内。

2.3 鲁棒对抗强化学习框架

2.3.1 双网络架构设计

Actor 网络：基于 Transformer 架构，输出 6 维连续控制量（3 维推力矢量+3 维角速率指令），通过位置编码处理状态序列的时序依赖。

Critic 网络：近似值函数评估策略价值，损失函数定义为时序差分误差平方和：\(\mathcal{L}_{\text{Critic}}=\mathbb{E}_{s\sim\rho^\pi}\left[\left(V(s|\theta)-\hat{V}(s)\right)^2\right]\)

其中\(\rho^\pi\)为策略\(\pi\)诱导的状态分布，\(\hat{V}(s)\)为目标值函数。

2.3.2 对抗训练机制

构建对抗网络生成执行器故障扰动\(\delta\)（如±20%推力偏差或 15° 矢量偏转），通过添加扰动生成对抗样本\(s'=s+\delta\)，迫使控制网络在扰动环境下保持控制精度。对抗损失函数定义为：\(\mathcal{L}_{\text{Adversary}}=\mathbb{E}_{s\sim\rho^\pi,\delta\sim\mathcal{N}(0,\Sigma)}\left[\|\mathbf{u}(s)-\mathbf{u}(s')\|_2^2\right]\)

通过对抗训练，策略对推力失效、传感器漂移等故障的容错能力显著提升。

2.3.3 奖励函数设计

采用多目标加权奖励函数，综合姿态精度（权重 0.6）、燃料消耗（权重 0.3）与约束违反（权重 0.1）：\(R=0.6\cdot\exp(-\|\Delta\theta\|_2^2)-0.3\cdot\|\Delta\mathbf{u}\|_1-0.1\cdot\sum_{i=1}^n\max(0,c_i)\)

其中\(\Delta\theta\)为姿态误差，\(\Delta\mathbf{u}\)为控制量变化，\(c_i\)为约束违反指标（如推力饱和、角速率超限）。

2.4 显式模型预测控制集成

2.4.1 离线-在线优化策略

采用多参数规划（MP-SQP）求解 20 步预测时域的最优控制问题，生成状态分区\(\{\mathcal{X}_k\}\)与线性控制律\(\mathbf{u}_k=\mathbf{K}_k\mathbf{x}+\mathbf{f}_k\)，存储为查找表。在线阶段通过状态定位快速匹配分区，将计算时间从传统 MPC 的 8ms 降至 0.3ms，满足实时性要求。

2.4.2 安全切换逻辑设计

设置双阈值切换策略保障系统安全性：当约束违反度\(c>0.1\)或推进剂余量\(m<0.2m_{\text{max}}\)时，决策层优先启用 EMPC 进行硬约束优化；正常工况下由 RA-RL 主导控制，实现灵活性与安全性的平衡。

3 实验验证与结果分析

3.1 仿真环境搭建

基于 STK12.0 构建三类典型场景：

近地交会场景：500km 圆轨道，配置 8 台变推力发动机，碎片检测范围 5km，模拟空间碎片动态避障需求。

月球软着陆场景：10km 动力下降段，采用 SGM100h 重力场模型，地形分辨率 1m，包含月坑、斜坡等复杂地貌。

故障容错场景：模拟单侧发动机 30%推力衰减与 15° 矢量偏差，姿态角速率限制 5° /s，测试执行器故障

下的恢复能力。

3.2 性能评估指标

精度指标：姿态角均方根误差（RMS）、轨道位置误差（RPE）；

效率指标：总冲量消耗、决策延迟；

鲁棒性指标：故障恢复时间、约束违反次数。

3.3 对比实验结果

3.3.1 近地轨道交会性能

表 1 对比结果显示，本文方法在位置精度、燃料效率与实时性上均显著优于对比方案。与传统 MPC 相比，位置误差降低 63.2%，燃料消耗减少 18.4%，决策延迟仅 3.5ms，满足交会对接任务的严格实时性要求。

表 1 近地轨道交会性能对比

控制方法	位置误差(m)	燃料消耗(N·s)	决策延迟(ms)
传统 MPC	0.87	1250	8.6
深度强化学习	0.51	1180	4.2
本文方法	0.32	1020	3.5

3.3.2 月球软着陆效果

在复杂地形条件下，本文方法实现着陆点精确控制，横向偏差仅 15.2 米，较预设轨迹制导方法提升 65.2%。在±10° 坡度扰动下，着陆成功率达 92%，较基线模型提高 17.9%，验证了多模态感知网络对非结构化环境的适应性。

3.3.3 故障容错能力

当遭遇单侧发动机失效故障时，本文策略在 2.1 秒内完成姿态恢复，稳态误差控制在 0.5° RMS 以内，较传统 PID 控制提升 74.7%。同时，燃料消耗减少 25%，表明对抗训练有效优化了冗余执行器的协同控制策略。

3.4 消融实验分析

通过模块化功能移除实验验证系统组件效能：多模态融合机制：取消跨模态注意力单元后，地形辨识精度降低 3.4 个百分点，着陆坐标偏移量扩大至 48%，体现多源信息整合对场景建模的决定性作用。对抗训练单元：中止对抗性训练机制后，应急响应耗时增加 128.6%，推进剂消耗率上升 18%，表明动态对抗优化对系统稳定性的强化效果。EMPC 控制器：禁用模型预测控制模块引发操作约束突破率提升 300%，伴随推进系统超负荷警报，证实预测算法对机械限阈的防护价值。

4 结论与展望

本文提出一种多模态感知、鲁棒对抗强化学习与显式模型预测控制融合的航天器自主控制策略，经 STK 仿真验证其精度、效率与鲁棒性优势。实验显示该策略在典型场景控制性能较传统方法提升 20%以上。未来研究方向包括：决策可解释性：融合梯度归因与决策树技术，实现策略形式化验证，满足载人航天安全需求；星载硬件适配：研发模型量化与抗辐射 ASIC 芯片，降低计算功耗至≤10W；分布式协同控制：拓展多智能体强化学习在星群协同中的应用。深度学习与航天控制融合推动航天器从"地面遥控"向"自主智能"转变，混合智能控制框架将为火星采样、小行星防御等任务提供可靠技术支撑。

参考文献

[1]李明，王华。航天器模型预测控制研究进展[J].航天学报，2020,41(3):289-298.

[2]ChenX,LiY.CNN-basedVisualNavigationforLunarLanding[J].ActaAstronautica,2021,182:456-465.

[3]张伟，刘畅。高光谱遥感在行星表面物质分类中的应用[J].遥感学报，2019,23(5):890-898.

[4]WangZ,LiuH.Multi-AgentReinforcementLearningforOrbitalPursuit-EvasionGames[J].JournalofGuidance,Control,andDynamics,2022,45(6):1123-1135.

[5]TianJ,SunF.RobustAdversarialReinforcementLearningforSpacecraftAttitudeControl[J].IEEETransactionsonAerospaceandElectronicSystems,2023,59(2):987-1001.

[6]吴敏，陈刚。空间辐射环境下传感器噪声抑制技术研究[J].宇航学报，2021,42(7):901-908.

[7]HeK,ZhangX.DeepResidualLearningforImageRecognition[C].CVPR,2016:770-778.

[8]ISO26262.RoadVehicles-FunctionalSafety[S].2018.

[9]刘洋，王磊。多传感器融合定位误差分析[J].仪器仪表学报，2020,41(4):156-163.

作者简介：赵唯，1991.09，女，汉族，河南新乡人，硕士研究生，工程师，研究方向：指挥控制、健康管理、运维。

柴朔，1992.04，男，汉族，山东潍坊人，硕士研究生，工程师，研究方向：指挥控制。

通讯作者简介:张杰，1993，男，汉族，山西晋中人，最高学历：硕研，工程师，研究方向:健康管理、运维、指挥控制。