

基于安全强化学习的自主交会对接的制导方法与地面验证

任师琦

中国科学院微小卫星创新研究院

2025 年 11 月 13 日



- ① 课题背景
- ② 研究现状
- ③ 研究内容
- ④ 参考文献

① 课题背景

② 研究现状

③ 研究内容

④ 参考文献

课题背景

- ARVD（自主交会与对接）是空间站在轨服务、深空探测与碎片清除等任务的关键能力，对自主性、鲁棒性与安全性要求严格。
- 传统制导与控制（如 MPC、PID、轨道最优）依赖精确模型，在传感噪声、推力故障或扰动下表现脆弱，可能引致燃料浪费或碰撞风险。
- 强化学习（RL）可在高维、非线性环境中自适应策略，但需引入约束或风险度量以保证安全（Safe RL）。
- 本课题以 Safe RL 为核心，目标是实现从仿真到地面验证的安全自适应制导控制方案，提升在实际任务工况下的鲁棒与可验证性。

- ① 课题背景
- ② 研究现状
- ③ 研究内容
- ④ 参考文献

研究现状

- 国内：研究以高校为主，侧重工程化实现与系统集成（低推力情形、约束处理），多数工作停留在仿真平台，面向安全性的系统性方法与大规模验证不足。
- 国外：着重算法层面的突破（Meta-RL、Safe RL、策略稳定性证明等），代表性工作结合 PPO/SAC/DDPG、LSTM 等提升鲁棒性，但绝大多数仍在复杂模拟或蒙特卡洛评估中验证，实地或半实地验证稀缺。
- 共同瓶颈：现有研究多偏向仿真，缺少可验证、带安全保证的端到端方案与地面到轨道的迁移验证；鲁棒性、可解释性与安全约束整合仍是核心挑战。
- 本课题定位：填补“Safe RL 在 ARVD 场景下从仿真到地面验证”的空白，通过约束化学习与工程化验证推进实用化。

1 课题背景

2 研究现状

3 研究内容

1. 使用 SafeRL 实现 1km 到 20m 的制导和控制
2. 采用端到端的强化学习方法实现 20m 到对接点的 RVD 过程
3. 搭建 10m 级半地面模拟系统验证上述实验

4 参考文献

1 课题背景

2 研究现状

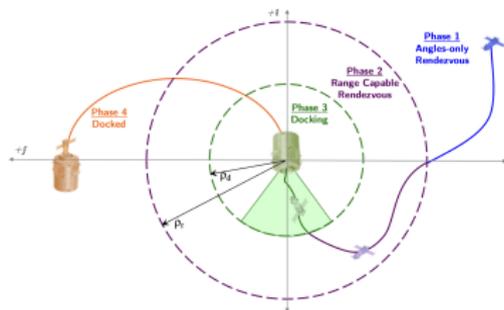
3 研究内容

1. 使用 SafeRL 实现 1km 到 20m 的制导和控制
2. 采用端到端的强化学习方法实现 20m 到对接点的 RVD 过程
3. 搭建 10m 级半地面模拟系统验证上述实验

4 参考文献

工况分析

- 工况划分：我们将交会对接任务划分为远程引导、近程制导、末端接近与对接四个阶段；本研究聚焦于近程制导与末端接近阶段。¹



- 在自主交会对接任务中，包括模糊控制和 MPC 在内的传统控制方法往往依赖精确的系统模型，在面对大型未知干扰时，会出现严重的性能下降。相比之下，RL 具有更高的鲁棒性

¹ Applying Proximal Policy Optimization for Spacecraft Rendezvous and Docking

系统建模

航天器相对运动使用 Clohessy-Wiltshire (CW) 方程描述：

$$\ddot{x} - 3n^2x - 2n\dot{y} = u_x + \delta_x, \quad (1)$$

$$\ddot{y} + 2n\dot{x} = u_y + \delta_y, \quad (2)$$

$$\ddot{z} + n^2z = u_z + \delta_z, \quad (3)$$

其中， $n = \sqrt{\mu/a^3}$ 是轨道角速度， μ 是地心引力常数， a 是半长轴； u_x, u_y, u_z 是控制输入； $\delta_x, \delta_y, \delta_z$ 表示不确定扰动（如噪声）。状态空间表示为：

$$\dot{\mathbf{s}} = \mathbf{A}\mathbf{s} + \mathbf{B}\mathbf{u} + \mathbf{w}, \quad (4)$$

其中 $\mathbf{s} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$ ，矩阵 \mathbf{A} 和 \mathbf{B} 来自 CW 方程线性化， \mathbf{w} 是过程噪声。

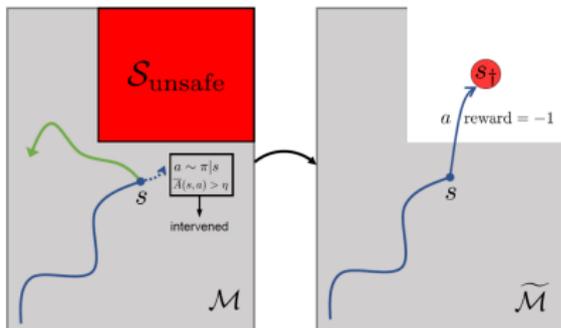
观测模型（考虑测量噪声）：

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{v}, \quad (5)$$

其中 \mathbf{H} 是观测矩阵， $\mathbf{v} \sim \mathcal{N}(0, R)$ 是测量噪声。

Safe RL framework

- Safe RL 的核心是在安全约束下建立目标的最大化期望回报：²



- 安全约束被表征为：

$$C_i(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t c_i(s_t, a_t) \right] \leq d_i, \quad \forall i = 1, \dots, m, \quad (6)$$

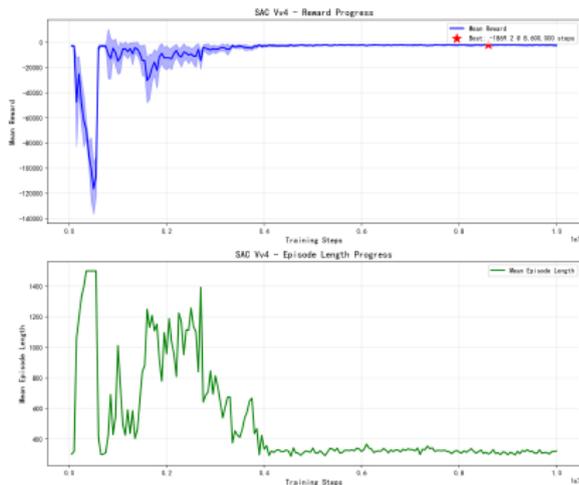
- 其中 $r(s_t, a_t)$ 是奖励函数, $c_i(s_t, a_t)$ 是第 i 个成本函数 (e.g., 碰撞风险、燃料超限), $\gamma \in (0, 1)$ 是折扣因子, d_i 是阈值。

²Safe Reinforcement Learning Using Advantage-Based Intervention

已有的工作

使用 RL 制导控制器实现 1km 到 100m 的制导过程

- 在之前的工作中，我们分别使用 PPO 和 SAC 算法训练了 RL 模型
- 在设定工况中，RL 制导控制器用于将 Chaser 从距离目标 1 公里处准确引导至 100 米以内，且全程应处于 LOS 安全区域

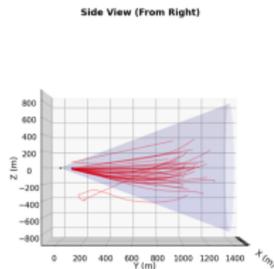
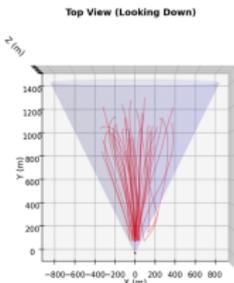
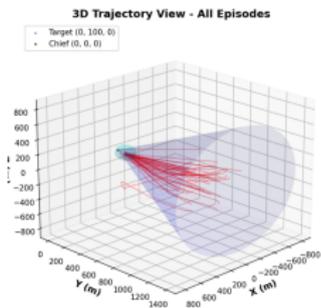


已有的工作

使用 RL 制导控制器实现 1km 到 100m 的制导过程

- 这套模型在模拟环境中得以验证
- 在多个独立评估集上拥有稳定的表现，成功率达到了实验指标

SAC V7.5 - 3D Spacecraft Docking Trajectories (Best Model @ 12.55M Steps)



1 课题背景

2 研究现状

3 研究内容

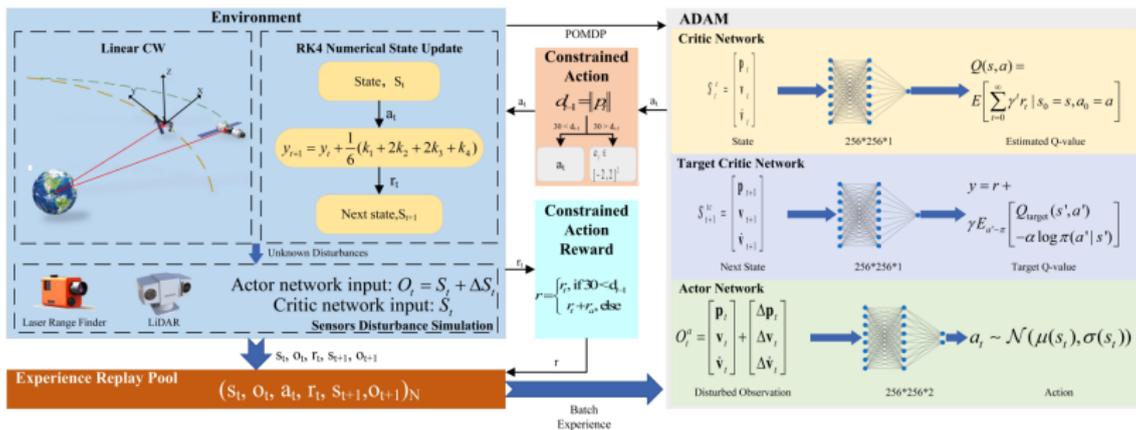
1. 使用 SafeRL 实现 1km 到 20m 的制导和控制
2. 采用端到端的强化学习方法实现 20m 到对接点的 RVD 过程
3. 搭建 10m 级半地面模拟系统验证上述实验

4 参考文献

end to end GNC framework

由 RL 直接接管 GNC 全流程

- 使用单一的深度神经网络，实现端到端决策。
- 输入：原始传感器数据 (相机图像 + LiDAR 点云)。
- 输出：直接的航天器控制指令 (如推力、力矩)。
- 绕过传统的状态估计与 GNC 分离设计，对传感器噪声鲁棒性更强。³



³Space Noncooperative Object Active Tracking With Deep Reinforcement Learning

1 课题背景

2 研究现状

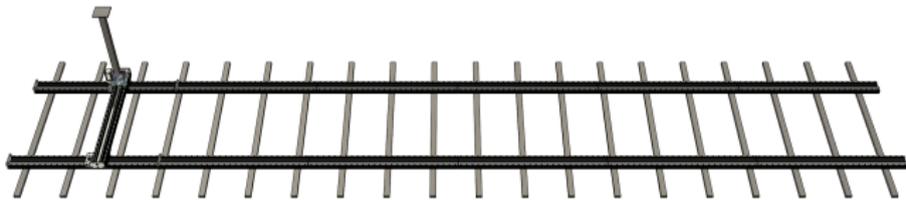
3 研究内容

1. 使用 SafeRL 实现 1km 到 20m 的制导和控制
2. 采用端到端的强化学习方法实现 20m 到对接点的 RVD 过程
3. 搭建 10m 级半地面模拟系统验证上述实验

4 参考文献

地面模拟系统的组成

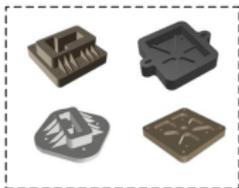
- RVD 过程的最后 10m 包含大量测量内容，为了对其进行高拟实模拟，我们搭建了这套仿真平台
- 通过特别设计的快装夹具，这套系统可以搭载多种传感器载荷，对输入端的传感器进行高保真工况验证
- 地面模拟系统可以实现闭环的动力学再现，被用于运行并验证我们的端到端强化学习模型



10m级多自由度半物理仿真平台



二级负载平台



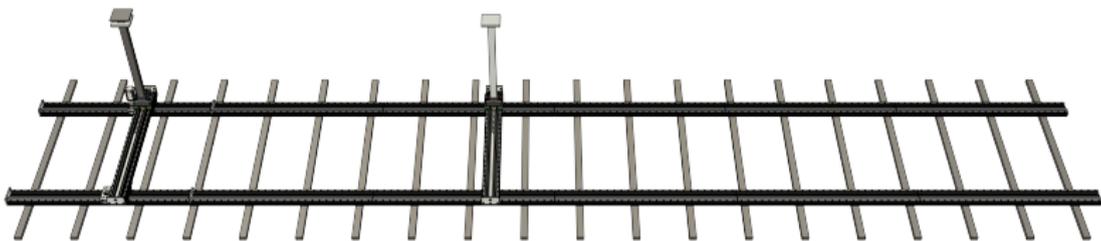
快装夹具



传感器载荷

地面模拟系统的后续规划

- 在我们的后续规划中，包括日光模拟器和消光系统在内的多套设施会被部署以进行高保真的对接环境模拟
- 一套额外的轨道系统会被添加以更好的模拟 RVD 过程中 Target 的行为



- ① 课题背景
- ② 研究现状
- ③ 研究内容
- ④ 参考文献

- [CSL25] S. Chapin, K. Stewart, and R. Leontie.
Reinforcement learning control of free-flyer for spacecraft assembly: Apiary experiment.
Technical report, U.S. Naval Research Laboratory, Washington, DC, USA, Sep 2025.

- [DHL+24] K. Dunlap, N. Hamilton, Z. Lippay, M. Shubert, S. Phillips, and K. L. Hobbs.
Demonstrating reinforcement learning and run time assurance for spacecraft inspection using unmanned aerial vehicles.
arXiv preprint arXiv:2405.06770, May 2024.
[cs.RO].

[EHOGOM25] M. El Hariry, A. Orsula, M. Geist, and M. Olivares Mendez.

RI-avist: Reinforcement learning for autonomous visual inspection of space targets.

arXiv preprint arXiv:2510.22699, Oct 2025.

[cs.RO].

[Fer23] G. Fereoli.

Meta-reinforcement learning for spacecraft proximity operations guidance and control in cislunar space.

Master's thesis, Politecnico di Milano, Milan, Italy, 2023.

[FSZF22] L. Federici, A. Scorsoglio, A. Zavoli, and R. Furfaro.

Meta-reinforcement learning for adaptive spacecraft guidance during finite-thrust rendezvous missions.

Acta Astronautica, 201:129–141, December 2022.

[HU21]

K. Hovell and S. Ulrich.

Deep reinforcement learning for spacecraft proximity operations guidance.

Journal of Spacecraft and Rockets, 58(2):254–266, Mar-Apr 2021.

[SKSA23]

K. P. Sharma, I. Kumar, P. P. Singh, and K. Anbazhagan.

Advancing spacecraft rendezvous and docking through safety reinforcement learning and ubiquitous learning principles.

Comput. Human Behav., 2023.

[SZS+25]

S. Shao, D. Zhou, G. Sun, W. Ma, and R. Deng.

Asymmetric deep reinforcement learning-based spacecraft approaching maneuver under unknown disturbance.

Aerospace, 12(3):170, 2025.

Thanks!