

文章编号:1671-1637(2026)03-0185-13

# 基于多智能体近端策略优化的低空异构飞行器 实时三维冲突解脱方法

陈运翔<sup>1,2</sup>, 苟明<sup>3</sup>, 张建平<sup>\*1,2</sup>, 芦维宁<sup>4</sup>, 唐凯<sup>3</sup>, 张光远<sup>1,2</sup>

(1. 西南交通大学 交通运输与物流学院, 四川 成都 610031; 2. 低空交通智能管控四川省重点实验室, 四川 成都 610031; 3. 北京信息科技大学 机电工程学院, 北京 100192; 4. 清华大学 北京信息科学与技术国家研究中心, 北京 100084)

**摘要:**针对低空异构飞行器实时三维冲突解脱问题,选取了中大型固定翼飞行器与轻小型多旋翼无人机共享空域运行这一类目前发展迅猛的低空运行场景开展研究;采用集中式训练与分布式执行框架,提出了一种基于多智能体近端策略优化(MAPPO)的解决方法;基于两类飞行器的运行特性确立了固定翼飞机稳定飞行、多旋翼无人机机动避让的实时三维冲突解脱策略,构建了兼顾碰撞避免、任务效率、优先级和平稳性的多维奖励函数;引入了优先级机制以保障固定翼飞机的任务优先性,同时引导多旋翼无人机主动避让。仿真试验表明:选取5、10、20、30架次轻小型多旋翼无人机仿真飞行过程开展基准试验均可实现92%以上任务成功率,计算开销为0.16~0.36 min,平均冲突解脱时间为0.28~1.76 s,飞行冲突占比为0.95%~2.18%,通过优化状态空间、动作空间和奖励函数,该方法在冲突解脱时间上优于现有方法2.25 s,任务成功率上提高2%,为进一步在广域范围开展低空异构飞行器融合运行研究奠定了基础。

**关键词:**低空交通;冲突解脱;多智能体近端策略优化;低空异构飞行器;低空航行系统

中图分类号:U8 文献标志码:A DOI:10.19818/j.cnki.1671-1637.2026.092

## Real-time 3D conflict resolution method for low-altitude heterogeneous aircraft based on multi-agent proximal policy optimization

CHEN Yun-xiang<sup>1,2</sup>, GOU Ming<sup>3</sup>, ZHANG Jian-ping<sup>\*1,2</sup>, LU Wei-ning<sup>4</sup>,  
TANG Kai<sup>3</sup>, ZHANG Guang-yuan<sup>1,2</sup>

(1. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, Sichuan, China; 2. Intelligent Management and Control of Low-altitude Traffic Key Laboratory of Sichuan Province, Chengdu 610031, Sichuan, China; 3. College of Mechanical and Electrical Engineering, Beijing Information Science & Technology University, Beijing 100192, China; 4. Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China)

**Abstract:** In response to the real-time three-dimensional conflict resolution for low-altitude heterogeneous aircraft, a rapidly developing operational scenario was studied, including shared

出版历程:2025-08-31 收稿,2025-10-14 修回,2025-11-27 录用

基金项目:国家重点研发计划(2022YFB4300903);国家自然科学基金民航联合研究项目(U2433217);国家自然科学基金项目(52472332);四川省重大科技专项揭榜挂帅项目(2024ZDZX0044);四川省自然科学基金项目(2025ZNSFSC0394)

作者简介:陈运翔(1992-),男,四川宜宾人,助理研究员,博士,博士后,E-mail:chenyunxiang@swjtu.edu.cn。

\*通信作者:张建平(1976-),男,安徽芜湖人,研究员,工学博士,E-mail:zhangjp@swjtu.edu.cn。

引用格式:陈运翔,苟明,张建平,等.基于多智能体近端策略优化的低空异构飞行器实时三维冲突解脱方法[J].交通运输工程学报,2026,26(3):185-197.

Citation: CHEN Yun-xiang, GOU Ming, ZHANG Jian-ping, et al. Real-time 3D conflict resolution method for low-altitude heterogeneous aircraft based on multi-agent proximal policy optimization[J]. Journal of Traffic and Transportation Engineering, 2026, 26(3): 185-197.

airspace operations between medium-to-large fixed-wing aircraft and light small multi-rotor unmanned aerial vehicles (UAVs). A multi-agent proximal policy optimization (MAPPO)-based method was proposed with a centralized training and decentralized execution framework. Based on the operational characteristics of the two types of aircraft, a real-time three-dimensional conflict resolution strategy was established to allow fixed-wing aircraft to maintain stable flight while multi-rotor UAVs perform avoidance maneuvers. A multi-dimensional reward function was designed, taking into account collision avoidance, mission efficiency, priority, and smoothness. A priority mechanism was introduced to ensure the mission priority of fixed-wing aircraft and encourage proactive avoidance by multi-rotor UAVs. Simulation results show that baseline tests involving 5, 10, 20, and 30 light small multi-rotor UAVs all achieve a mission success rate of over 92%, with computational overhead ranging from 0.16 to 0.36 min, average conflict resolution time between 0.28 and 1.76 s, and flight conflict proportions between 0.95% and 2.18%. Through optimization of the state space, action space, and reward function, the proposed method reduces conflict resolution time by 2.25 s and improves mission success rate by 2% compared to existing methods. A foundation is thus laid for further research on the integrated operation of low-altitude heterogeneous aircraft in wide-area scenarios.

**Keywords:** low-altitude traffic; conflict resolution; multi-agent proximal policy optimization; low-altitude heterogeneous aircraft; low-altitude navigation system

**Publication history:** Received 2025-08-31; Received in revised form 2025-10-14; Accepted 2025-11-27

**Funding:** National Key R&D Program (2022YFB4300903); Civil Aviation Joint Research Fund of National Natural Science Foundation of China (U2433217); National Natural Science Foundation of China (52472332); Sichuan Provincial Major Science and Technology Special Project - Tackling Key Problems Initiative (2024ZDZX0044); Natural Science Foundation of Sichuan Province (2025ZNSFSC0394)

\* **Corresponding author:** ZHANG Jian-ping, research fellow, PhD, Email: zhangjp@swjtu.edu.cn.

## 0 引 言

随着国家低空经济快速发展,低空飞行活动将逐渐呈现大流量、高密度的异构飞行器在共享空域中协同运行的复杂态势,未来甚至出现无人驾驶航空器(无人机)与有人驾驶航空器(有人机)融合运行的高阶运行场景,确保其在动态三维环境下的安全协同已成为空中交通的核心挑战<sup>[1-5]</sup>。欧美等国已前瞻性部署固定翼有人机与多旋翼无人机协同运行技术,在美国陆军的“空中发射效果”项目中,固定翼无人机与多旋翼无人机协同已进入测试阶段,英国的无人机物流演示项目成功验证了固定翼与多旋翼无人机接力的运输方式。中国目前相关技术还不完备。

2023年12月21日,《国家空域基础分类方法》正式发布,首次在真高300 m以下划设G类空域。G类空域是一类对民航公共运输飞行无影响,仅提供飞行信息服务,不提供空中交通管制服务的空域<sup>[4]</sup>。

执行通航任务的中大型固定翼飞行器与轻小型多旋翼无人机在G类空域协同运行场景发展迅猛,两类飞行器在任务需求和优先级上的显著差异,极大增加了共享空域管理的难度。如何在动态环境中实现固定翼飞机与多旋翼无人机的实时三维冲突解脱,确保低空异构飞行器的高效协同与安全运行,已成为空中交通管理亟待解决的技术难题<sup>[6-7]</sup>。

传统的冲突解脱方法主要基于数学规划与规则约束。针对民航运输运行场景,Reich<sup>[8]</sup>提出了基于水平和垂直分离标准的冲突概率分析模型,奠定了有人机之间安全间隔计算方法的基础,然而其静态假设限制了在动态环境中的适用性;Fiorini等<sup>[9]</sup>提出的速度障碍法通过几何航迹预测实现避碰,理论上适用于简单场景,但其计算复杂度随飞行器数量呈指数增长,且难以适应三维动态障碍物环境;Brooker<sup>[10]</sup>的横向冲突风险评估聚焦于水平方向分析,虽在冲突管理中取得广泛应用,却无法有效扩展至多飞行器三维场景;刘洋等<sup>[11]</sup>的探测算法通过雷

达数据提升了冲突检测精度;Hernández-Romero等<sup>[12]</sup>基于天气不确定性的概率方法,结合气象情况优化了冲突过程;管祥民等<sup>[13-14]</sup>提出了使用博弈论以及基于系统理论混合式分析方法,对冲突解脱的安全性进行分析;张启钱等<sup>[15]</sup>提出可同时选择调速度、调航向以及调高度的序列混合整数线性优化模型研究多机冲突解脱问题。针对无人机运行场景,陈运翔等<sup>[16]</sup>提出了一种基于机动避撞策略的无人机安全间隔评估方法;谷志鸣等<sup>[17]</sup>通过模拟测试强化了单机避障的可靠性。然而,上述方法对于处理大流量、高密度的低空异构飞行器三维动态冲突,往往受限于计算效率或环境假设,难以满足实时性需求。例如,朱代武<sup>[18]</sup>指出,几何解脱算法在多飞行器场景下的实时性不足,难以应对动态变化的空域需求。

近年来,强化学习(Reinforcement Learning, RL)因其自适应性在冲突解决中展现潜力。单智能体RL方法在无人机路径规划中取得进展。例如,Pham等<sup>[19]</sup>基于深度Q网络(Deep Q-Network, DQN)优化单架无人机路径;Loquercio等<sup>[20]</sup>通过模仿学习实现复杂环境下的高速避障飞行;Lin等<sup>[21]</sup>通过Q学习优化单架飞行器的避障策略;Jilkov等<sup>[22]</sup>提出多模型方法,结合意图和天气不确定性改进单飞行器解脱精度;Zhao等<sup>[23]</sup>设计了通用单智能体RL框架,适用于简单冲突场景。然而,这些方法缺乏多机协同机制,单智能体策略无法有效协调优先级和任务需求,导致解脱实效受限。

多智能体强化学习(Multi-agent Reinforcement Learning, MARL)因其在动态环境中的协同优化能力成为研究热点。例如,Lai等<sup>[24]</sup>提出的多智能体深度确定性策略梯度(Multi-agent Deep Deterministic Policy Gradient, MADDPG)算法在飞行冲突解决中实现分布式决策,但验证场景局限于二维仿真;Chen等<sup>[25]</sup>提出的三维多机冲突解决框架,通过DQN改进方法实现多智能体协作,验证了其在动态空域的可行性,却没有针对异构飞行器协同运行场景设计差异化奖励机制;Dong等<sup>[26]</sup>设计了基于MARL的三维解脱策略,优化了大规模场景下的协调效率;Brittain等<sup>[27]</sup>引入LSTM网络,增强了冲突检测的长期依赖性。同时,基于图神经网络(Graph Neural Network, GNN)的消息传递<sup>[28]</sup>通过信息共享和动态预测优化了性能。

通过回顾现有研究可以发现,目前针对飞行冲突解脱的研究主要还是在无人机之间或有人机之间,传统方法受限于计算效率和环境假设,单智能体

RL缺乏协同性,MARL虽具潜力但尚未成熟。执行通航任务的中大型固定翼飞行器(可能是有人机,也可能是无人机)与轻小型多旋翼无人机在G类空域协同运行中,中大型固定翼飞行器需优先保障飞行稳定性以确保任务安全完成,而轻小型多旋翼无人机依赖灵活机动能力完成复杂任务。现有研究未充分考虑此类特性差异,导致冲突解脱策略偏离实际。综上,本文针对此类场景提出基于多智能体近端策略优化(Multi-agent Proximal Policy Optimization, MAPPO)的实时三维飞行冲突解决方法,通过集中式训练与分布式执行(Centralized Training with Decentralized Execution, CTDE)框架,设计模块化状态表达、优先级机制和多维奖励函数,实现低空异构飞行器之间的高效协同及冲突解脱。

## 1 实时三维多机冲突解脱问题

### 1.1 问题定义

随着国家空域系统的重构及非管制空域的发布,大流量、高密度的低空异构飞行器在共享空域中协同运行将成为常态,而中大型固定翼飞行器(可能是有人机,也可能是无人机)与轻小型多旋翼无人机在G类空域三维动态环境下同时执行通航任务已经较为普及。然而,由于两类飞行器在任务目标、运动特性和优先级上的显著差异,飞行冲突的发生概率显著增加。本文聚焦于实时三维多飞行器冲突解脱问题,旨在设计一种高效的协同策略,确保两类飞行器在动态空域中的安全运行与任务完成效率。具体而言,该问题涉及以下挑战:一是如何在任务和空域复杂度提升的情况下,快速检测并化解潜在冲突;二是如何平衡中大型固定翼飞机的任务优先性与轻小型多旋翼无人机的机动灵活性,实现全局优化。这一问题的解决对于提升空中交通安全性和资源利用效率具有重要意义。

### 1.2 异构飞行器的运动模型

在三维空域中,固定翼飞机与多旋翼无人机的运行特性存在显著差异,这对冲突解脱策略的设计提出了独特的需求。本文针对两类飞行器的任务目标、物理限制和安全要求,设计了差异化的运动模型,以确保冲突解脱的实时性、安全性和运行效率。在共享空域的仿真环境中,参与任务的飞行器包括 $N_H$ 架中大型固定翼飞机和 $N_D$ 架轻小型多旋翼无人机。

针对中大型固定翼飞机,其通航飞行任务强调航线的稳定性和路径可预测性。因此,其运动模型

被设计为二维平面内的匀速直线运动,其位置在每个时间步的更新依赖于当前速度和航向。为了更精确地描述两类飞行器的行为特性,本文参考Chen<sup>[25]</sup>的方法为两类飞行器设计了各自的动力学模型。中大型固定翼飞机位置更新公式为

$$p_{H,t+1} = p_{H,t} + v_{H,t} \Delta t \quad (1)$$

式中: $p_{H,t}$ 为中大型固定翼飞机 $t$ 时刻的当前位置; $v_{H,t}$ 为中大型固定翼飞机 $t$ 时刻的飞行速度; $\Delta t$ 为时间步长。

由于中大型固定翼飞机的调整动作受到一定的限制,包括速度调整范围以及航向调整范围

$$\Delta v_{H,t} \in [\delta_{\min,v}, \delta_{\max,v}] \quad (2)$$

$$\Delta \theta_{H,t} \in [\delta_{\min,\theta}, \delta_{\max,\theta}] \quad (3)$$

式中: $\Delta v_{H,t}$ 为中大型固定翼飞机的速度调整范围; $\delta_{\min,v}$ 、 $\delta_{\max,v}$ 分别为速度 $v$ 的最小值和最大值; $\Delta \theta_{H,t}$ 为中大型固定翼飞机的航向调整范围; $\delta_{\min,\theta}$ 、 $\delta_{\max,\theta}$ 分别为航向角 $\theta$ 的最小值和最大值。

相比之下,轻小型多旋翼无人机的任务场景更加多样,其设计目标是提供更高的机动性和灵活性,以适应复杂空域中的冲突规避需求。因此,轻小型多旋翼无人机的运动模型被扩展到三维空间,不仅包括平面内的速度和航向调整,还增加了高度维度的变化。轻小型多旋翼无人机的动力学公式为

$$p_{D,t+1} = p_{D,t} + v_{D,t} \Delta t + \Delta z_{D,t} \quad (4)$$

式中: $p_{D,t}$ 为轻小型多旋翼无人机在三维空间中的当前位置; $v_{D,t}$ 为轻小型多旋翼无人机的当前速度; $\Delta z_{D,t}$ 为轻小型多旋翼无人机的高度调整范围。

轻小型多旋翼无人机的速度调整范围、航向调整范围和高度调整范围为

$$\Delta v_{D,t} \in [\delta_{\min,v}, \delta_{\max,v}] \quad (5)$$

$$\Delta \theta_{D,t} \in [\delta_{\min,\theta}, \delta_{\max,\theta}] \quad (6)$$

$$\Delta z_{D,t} \in [\delta_{\min,z}, \delta_{\max,z}] \quad (7)$$

式中: $\Delta v_{D,t}$ 为轻小型多旋翼无人机速度调整范围; $\Delta \theta_{D,t}$ 为轻小型多旋翼无人机航向调整范围。

这种差异化运动模式的设计直接来源于实际空域运行的需要。中大型固定翼飞机的稳定性和有限调整范围反映了其优先执行长距离任务的特点,而轻小型多旋翼无人机的三维机动性则使其成为冲突解脱中的主动调整者。这种分工不仅提升了系统的整体安全性,还优化了空域资源利用效率。

### 1.3 奖励函数设计

在多飞行器协同运行的场景中,首要任务是确

保飞行器之间处于安全间隔,其次需要激励飞行器选择更高效的路径规划,因此设置碰撞惩罚及目标距离奖励是基础性奖励,具有普适性。本文基于中大型固定翼飞机和轻小型多旋翼无人机的任务重要性差异,建立了优先级和平稳性奖励函数机制,面向轻小型多旋翼无人机设计机动灵活的避让策略,以确保中大型固定翼飞机的航迹优先性和安全性。具体考虑如下:首先,中大型固定翼飞机通常执行高优先级的任务,例如紧急救援,其航迹的任何调整可能导致任务延误或安全风险,因此应尽量避免对其航迹的干扰;其次,轻小型多旋翼无人机的任务(如物流运输)通常具有更高的灵活性,能够通过机动调整适应空域变化。因此,在冲突发生时,例如当轻小型多旋翼无人机和中大型固定翼飞机的距离小于安全距离时,通过优先级和平稳性奖励激励轻小型多旋翼无人机调整速度、航向或高度,以确保中大型固定翼飞机的航迹不受影响。这种机制显著降低中大型固定翼飞机的冲突避让频率,符合实际运行需求。

本文设计了适合轻小型多旋翼无人机和中大型固定翼飞机运行的奖励函数,涵盖了飞行器的安全性、任务完成效率、优先级约束和飞行平稳性等关键方面。

针对飞行器 $i$ ,参考MAPPO利用CTDE框架进行策略更新方法,给出综合奖励函数的形式如下<sup>[29]</sup>

$$R_{t,i} = R_{c,i} + R_{g,i} + R_{p,i} + R_{s,i} \quad (8)$$

式中: $R_{t,i}$ 为综合奖励; $R_{c,i}$ 为碰撞惩罚; $R_{g,i}$ 为目标距离奖励; $R_{p,i}$ 为优先级奖励; $R_{s,i}$ 为平稳性奖励。

#### 1.3.1 碰撞惩罚

该奖励项通过负向奖励惩罚飞行器间的过近距离,确保其始终保持安全距离 $d_s$ 。碰撞惩罚 $R_{c,i}$ 定义为

$$R_{c,i} = \begin{cases} -C_c & \|p_{t,i} - p_{t,j}\| < d_s \\ 0 & \text{其他} \end{cases} \quad (9)$$

式中: $C_c$ 为激励系数; $\|p_{t,i} - p_{t,j}\|$ 为飞行器 $i$ 和 $j$ 之间的欧几里得距离,这一约束确保了飞行器在任务执行过程中保持足够的安全间隔<sup>[14]</sup>。

#### 1.3.2 目标距离奖励

该奖励项通过对目标点距离的负向奖励,激励飞行器选择更高效的路径规划。目标距离奖励 $R_{g,i}$ 定义为

$$R_{g,i} = -\|p_{t,i} - g_i\| \quad (10)$$

式中: $p_{t,i}$ 为智能体 $i$ 的当前位置; $g_i$ 为智能体 $i$ 的目标点位置。

### 1.3.3 优先级奖励

优先级奖励是本文的一项关键设计,主要用于体现中大型固定翼飞机的直航优先权,旨在激励轻小型多旋翼无人机主动避让中大型固定翼飞机,同时对轻小型多旋翼无人机的避让行为施加正向激励。具体而言,对于一架轻小型多旋翼无人机,其优先级奖励在每一时间步根据与中大型固定翼飞机的距离变化确定。如果轻小型多旋翼无人机在当前时间步采取动作后,与最近的中大型固定翼飞机的距离增加,则获得正向奖励;如果距离没有增加,则奖励为0。优先级奖励  $R_{p,i}$  定义为

$$R_{p,i} = \begin{cases} C_p \left\| p_{D,t+1} - p_{H,t+1} \right\| > \left\| p_{D,t} - p_{H,t} \right\| \\ 0 & \text{其他} \end{cases} \quad (11)$$

式中:  $C_p > 0$  为正向激励系数。

### 1.3.4 平稳性奖励

为了优化飞行器的运行效率和路径平稳性,本文设计了平稳性奖励,适用于所有飞行器。具体而言,平稳性奖励根据飞行器在速度、航向和高度上的调整幅度计算,其目标是减少剧烈的动作变化。同时通过对轻小型多旋翼无人机和中大型固定翼飞机不同平稳性权重大小的设置,系统会优先要求轻小型多旋翼无人机调整速度、航向或高度,以确保中大型固定翼飞机的航迹不受影响。平稳性奖励为

$$R_{s,i} = w_i \left( -\frac{|\Delta v_{t,i}|^2}{f_v} - \frac{|\Delta \theta_{t,i}|^2}{f_\theta} - \frac{|\Delta z_{t,i}|^2}{f_z} \right) \quad (12)$$

式中:  $w_i$  为平稳性权重;  $\Delta v_{t,i}$ 、 $\Delta \theta_{t,i}$ 、 $\Delta z_{t,i}$  分别为飞行器  $i$  在时间  $t$  的速度、航向和高度的调整量;  $f_v$ 、 $f_\theta$ 、 $f_z$  分别为对应的平滑因子,用于限制剧烈调整。

将中大型固定翼飞机的平稳性权重设置较大,轻小型多旋翼无人机平稳性权重设置较小,使得中大型固定翼飞机飞行状态的改变会对奖励结果带来更大的负反馈。这样在策略生成过程中,为了获得更多的正向奖励结果,策略更倾向于优先通过调整轻小型多旋翼无人机的飞行状态来完成解冲突动作。

## 1.4 问题的优化目标

基于上述奖励函数设计,本文问题的优化目标为最大化所有飞行器的期望累计回报,表示为一个价值函数,即

$$R = \sum_{i=1}^{N_n + N_d} E \left( \sum_{t=0}^T \gamma^t R_{t,i} \right) \quad (13)$$

式中:  $E(\cdot)$  为期望;  $\gamma_t$  为折扣因子;  $T$  为任务周期。

此目标通过平衡安全性和任务效率,指导后续强化学习策略的优化。

## 2 多智能体强化学习冲突解脱方案

### 2.1 MAPPO 算法框架

本文采用MAPPO作为实时三维冲突解脱的核心算法。MAPPO是一种MARL方法,通过CTDE框架,优化轻小型多旋翼无人机与中大型固定翼飞机在动态空域中的协同行为。

MAPPO算法采用CTDE框架,以平衡协同性和实时性。对于每架飞行器(中大型固定翼飞机或轻小型多旋翼无人机),设计了一个独立的策略网络  $\pi_\theta(a_{t,i}|o_{t,i})$ ,其中:  $\theta$  为策略网络的参数;  $a_{t,i}$  为飞行器  $i$  在时刻  $t$  的动作;  $o_{t,i}$  为飞行器  $i$  在时刻  $t$  的局部观测。全局状态  $s_t$  由所有飞行器的关键状态信息构成,包括所有飞行器的位置、速度、航向、与邻近飞行器的相对状态、与目标点的距离。在训练阶段,MAPPO围绕所设计的奖励函数,利用全局状态  $s_t$  集中优化策略网络  $\pi_\theta(a_{t,i}|o_{t,i})$  和价值网络  $V_\phi(s_t)$ 。训练过程从飞行器与仿真环境的交互开始。每架飞行器根据策略网络  $\pi_\theta(a_{t,i}|o_{t,i})$  生成动作  $a_{t,i}$ ,环境返回奖励  $R_{t,i}$  和下一状态  $s_{t+1}$ ,形成经验元组  $\{s_{t,i}, a_{t,i}, R_{t,i}, s_{t+1,i}\}$ 。在执行阶段,MAPPO采用分布式决策,即每架飞行器仅依赖局部观测  $o_{t,i}$  执行动作,而无需获取全局状态  $s_t$ 。执行后,环境更新全局状态,并根据即时状态计算每个智能体的即时奖励。由于每架飞行器无需访问全局状态,因而计算需求显著降低。算法框架如图1所示。

### 2.2 MAPPO 神经网络设计

MAPPO神经网络设计包括策略网络和价值网络,分别负责动作生成和状态价值评估,确保飞行器在冲突解脱中的高效决策。在MAPPO算法中,智能体的策略网络  $\pi_\theta(\cdot)$  和价值网络  $V_\phi(\cdot)$  的结构设计如图2所示。

策略网络的任务是将观测  $o_{t,i}$  映射到动作空间,由3层神经网络构成:输入层接收局部观测  $o_{t,i}$ ; 2个隐藏层各有128个神经元(ReLU激活);输出层生成动作的概率分布  $\pi_\theta(a_{t,i}|o_{t,i})$ 。价值网络  $V_\phi(s_t)$  用于评估当前全局状态的价值,结构同样为3层:输入层接收全局状态  $s_t$ ; 2个隐藏层各有128个神经元(ReLU激活);输出层生成状态价值  $V_\phi(s_t)$ 。由于  $s_t$  包含全局信息,价值网络在所有飞行器间共享参数,通过集中式训练减少参数量并加速收敛,从而显著提升训练效率。

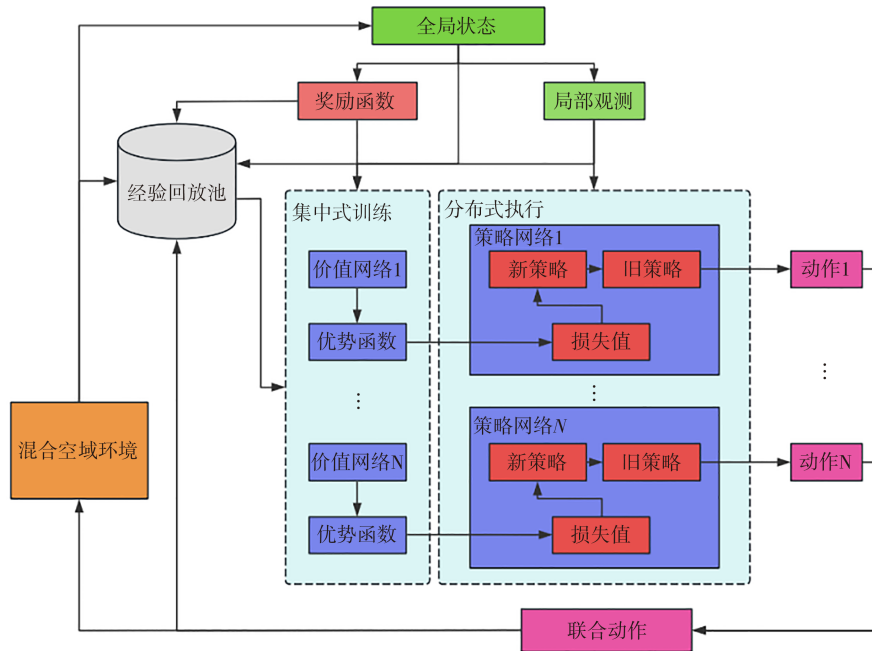


图1 MAPPO算法框架

Fig. 1 Framework of the MAPPO algorithm

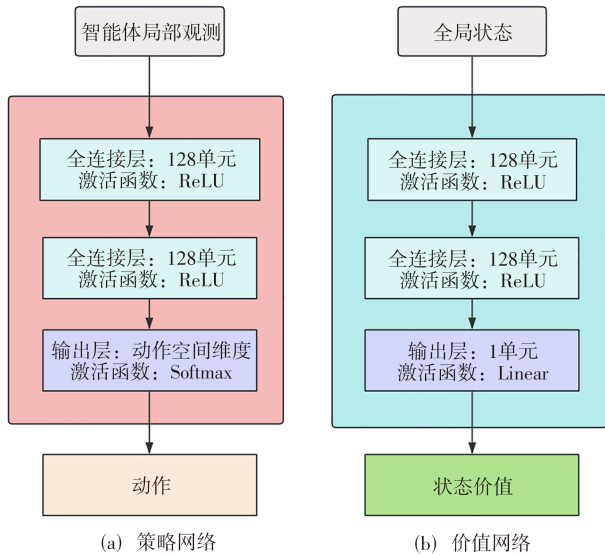


图2 MAPPO算法网络结构

Fig. 2 Network structure of the MAPPO algorithm

策略网络的优化使用剪切概率比目标函数,用以解决无剪切的目标函数策略更新步幅过大时可能会导致策略性能的剧烈下降问题,参考常用处理方法<sup>[30]</sup>,记为

$$L_t^{\text{CLIP}}(\theta) = E \left\{ \min \left\{ r_t(\theta) A_t, \text{CLIP} \left[ r_t(\theta), 1 - \epsilon, 1 + \epsilon \right] A_t \right\} \right\} \quad (14)$$

$$r_t(\theta) = \frac{\pi_\theta(a_{t,i} | o_{t,i})}{\pi_{\theta_0}(a_{t,i} | o_{t,i})}$$

式中:  $L_t^{\text{CLIP}}(\theta)$  为在时间步  $t$  的 CLIP 损失函数;  $A_t$  为

优势函数(通过广义优势估计 GAE 计算),表示动作相对于平均水平的好坏;  $\epsilon$  为一个超参数,将概率比限制为  $[1 - \epsilon, 1 + \epsilon]$  内,确保了新策略不会离旧策略太远,实现了信赖域的效果;  $\pi_{\theta_0}(a_{t,i} | o_{t,i})$  为旧的集中优化策略网络;  $r_t(\theta)$  为新旧策略的概率比。

价值网络用于评估全局状态的价值,以指导策略优化。价值网络的优化目标是最小化均方差损失函数,具体为

$$L_t^{\text{VF}}(\phi) = E \left\{ \left[ V_\phi(s_t) - R_t \right]^2 \right\} \quad (15)$$

式中:  $L_t^{\text{VF}}(\phi)$  为价值函数的损失函数,  $\phi$  为价值网络的参数;  $R_t$  为实际回报。

策略网络和价值网络承担不同任务。策略网络的任务是将局部观测映射到动作空间,输出动作的概率分布(离散动作空间,如加速、减速、左偏航等);因此输出层通常使用 Softmax 激活函数,将神经网络的原始输出转换为概率分布,确保所有动作的概率和为 1。价值网络的任务是评估全局状态的未来累计奖励期望,输出为标量;因此输出层通常采用线性激活,直接输出实数值。策略网络和价值网络的网络结构如图 2 所示。

### 2.3 动作空间

动作空间定义了飞行器在冲突解脱中的调整能力,直接实现第 1.2 节中运动模型的行为表达。基于中大型固定翼飞机和轻小型多旋翼无人机的运行特性,本文设计了离散化的动作空间,涵盖速度、航

向和高度三个维度。动作离散化参考Chen等<sup>[25]</sup>的多级意图划分方法,并根据民航规范和轻小型多旋翼无人机机动需求进行调整。速度意图包括减速、保持速度、加速;航向意图包括左偏航(负调整)、保持航向( $0^\circ$ )、右偏航(正调整);高度意图(仅轻小型多旋翼无人机)包括下降高度、保持高度、爬升高度。

中大型固定翼飞机的动作空间设计旨在保持飞行的稳定性。对于中大型固定翼飞机,其动作空间包含2个维度:速度变化和航向变化。其动作向量 $\mathbf{a}_{H,t}$ 定义为

$$\mathbf{a}_{H,t} = (\Delta v_{H,t}, \Delta \theta_{H,t}) \quad (16)$$

轻小型多旋翼无人机的动作空间设计需要支持其在三维空间中的灵活调整能力。轻小型多旋翼无人机的动作空间则扩展到三个维度,包括速度变化、航向变化和高度变化。其动作向量 $\mathbf{a}_{D,t}$ 定义为

$$\mathbf{a}_{D,t} = (\Delta v_{D,t}, \Delta \theta_{D,t}, \Delta z_{D,t}) \quad (17)$$

这种差异化设计确保中大型固定翼飞机优先保持稳定航迹,而轻小型多旋翼无人机通过三维动作主动规避冲突,优化空域协同效率。

## 2.4 状态空间

### 2.4.1 局部观测

局部观测 $O_t$ 确保飞行器在分布式执行中基于有限信息生成合理动作。参考Chen等<sup>[25]</sup>的扇区划分方法,观测范围以飞行器为中心的检测区域划分为8个 $45^\circ$ 检测扇区,每个扇区记录最近的一架轻小型多旋翼无人机和一架中大型固定翼飞机信息。

如图3所示,右侧的集合是局部观测矩阵,通过在绿色检测区域内观测到的不同机型,多旋翼无人机和中大型固定翼飞机分别用橙色和红色的状态空间;不同机型的状态空间 $O_T$ 定义为

$$O_T = (s_{H,t}, s_{D,t})^T \quad (18)$$

式中: $s_{H,t}$ 和 $s_{D,t}$ 分别为中大型固定翼飞机状态空间和轻小型多旋翼无人机状态空间。

### 2.4.2 中大型固定翼飞机状态空间

中大型固定翼飞机的任务目标较为单一,通常是沿直航路径完成飞行任务。中大型固定翼飞机的状态空间 $s_{H,t}$ 定义为

$$s_{H,t} = (x_{H,t}, y_{H,t}, v_{H,t}, \theta_{H,t}, \Delta s_{H,t}, g_H) \quad (19)$$

式中: $x_{H,t}$ 、 $y_{H,t}$ 分别为中大型固定翼飞机在二维空间中的当前横、纵向位置; $\theta_{H,t}$ 为中大型固定翼飞机的航向; $\Delta s_{H,t}$ 为中大型固定翼飞机与邻近飞行

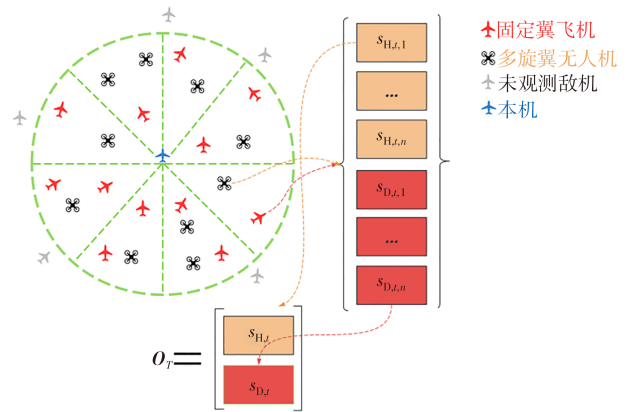


图3 局部观测矩阵和飞行器分类

Fig. 3 Local observation matrix and aircraft classification

器的相对状态; $g_H$ 为中大型固定翼飞机的目标点位置。

### 2.4.3 轻小型多旋翼无人机状态空间

轻小型多旋翼无人机不仅需要完成自身任务,还需主动避让中大型固定翼飞机或其他轻小型多旋翼无人机。轻小型多旋翼无人机的状态向量 $s_{D,t}$ 定义为

$$s_{D,t} = (x_{D,t}, y_{D,t}, z_{D,t}, v_{D,t}, \theta_{D,t}, \Delta z_{D,t}, \Delta s_{D,t}, g_D) \quad (20)$$

式中: $x_{D,t}$ 、 $y_{D,t}$ 、 $z_{D,t}$ 分别为轻小型多旋翼无人机在三维空间中的横、纵、垂位置; $\Delta s_{D,t}$ 为轻小型多旋翼无人机与邻近飞行器的动态交互关系; $g_D$ 为轻小型多旋翼无人机的目标点位置。

## 3 仿真试验

### 3.1 试验环境与参数配置

为验证基于MAPPO的冲突解脱方法的有效性,本文在三维空域仿真环境中开展试验,场景假设为中大型固定翼飞机执行应急救援任务,轻小型多旋翼无人机执行物流运输任务。仿真基于OpenAI Gym框架实现,运行于Linux系统,使用Python 3.8和PyTorch 1.9搭建算法。空域范围设定为 $10 \text{ km} \times 10 \text{ km}$ ,共3个飞行高度层(真高300 m以下,每层间隔30 m)。飞行器初始位置随机分布,轨迹使用Matplotlib的Axes3D模块可视化。试验参数如表1~3所示,所有参数经预试验优化以确保算法稳定性。

### 3.2 基准试验

本文着重讨论低空中大型固定翼飞机与轻小型多旋翼无人机之间的冲突解脱,但考虑到低空轻小型多旋翼无人机的高机动高密度飞行已经是当前涌现的现实场景,分别选取5、10、20、30架次轻小型多

表 1 算法参数

Table 1 Algorithm parameters

参数	取值
折旧因子	0.99
学习率	0.000 04
批量大小	64
剪切函数参数	0.2

表 2 轻小型多旋翼无人机参数

Table 2 Parameters of light small multi-rotor UAV

参数	取值
轻小型多旋翼无人机初始速度/( $\text{m}\cdot\text{s}^{-1}$ )	15
轻小型多旋翼无人机安全间隔/m	250
轻小型多旋翼无人机最大航向调整量/( $^{\circ}$ )	60
轻小型多旋翼无人机速度调整范围/( $\text{m}\cdot\text{s}^{-1}$ )	[12.75, 17.25]
轻小型多旋翼无人机平稳性权重	1

表 3 中大型固定翼飞机参数

Table 3 Parameters of medium-to-large fixed-wing aircraft

参数	取值
中大型固定翼飞机初始速度/( $\text{km}\cdot\text{h}^{-1}$ )	135
中大型固定翼飞机安全间隔/km	2
中大型固定翼飞机最大航向调整量/( $^{\circ}$ )	30
中大型固定翼飞机速度调整范围/( $\text{km}\cdot\text{h}^{-1}$ )	[115, 155]
中大型固定翼飞机平稳性权重	2

旋翼无人机仿真飞行过程开展基准试验(试验 1, 无优先级、平稳性奖励), 每组运行 1 000 次, 记录飞行冲突占比、平均冲突解脱时间、任务成功率和计算开销。此场景下冲突解脱仿真过程切片如图 4 所示, 试验结果如表 4 所示。

由图 4 可以看出: 4 组试验均可实现 92% 以上任务成功率, 计算开销为 0.16~0.36 min, 平均冲突解脱时间为 0.28~1.76 s, 飞行冲突占比为 0.95%~2.18%, 验证了本文所提方法对于单纯低空轻小型多旋翼无人机实现规模化高机动性飞行冲突解脱同样有效, 且计算开销结果显示具备实时性。

### 3.3 奖励机制的有效性分析

在多智能体系统中, 奖励函数的设计对于冲突解脱算法的性能至关重要。本节通过分别移除奖励函数中的优先级、平稳性奖励项, 设计 3 类试验, 以考察上述针对任务差异性的奖励项设置的有效性: ①完整方法(含所有奖励项, 试验 2); ②无优先级奖励(试验 3); ③无平稳性奖励(试验 4)。

为评估方法的扩展性, 本节试验中测试不同飞行器数量下的算法性能表现, 设置 4 种规模, 均按照轻小型多旋翼无人机和中大型固定翼飞机数量配比为 4:1, 分别为 5 架次(4 架次轻小型多旋翼无人机和 1 架次中大型固定翼飞机)、10 架次(8 架次轻小型多

旋翼无人机和 2 架次中大型固定翼飞机)、20 架次(16 架次轻小型多旋翼无人机和 4 架次中大型固定翼飞机)、30 架次(24 架次轻小型多旋翼无人机和 6 架次中大型固定翼飞机), 每组运行 1 000 次, 记录飞行冲突占比、平均冲突解脱时间、任务成功率和计算开销。试验中的冲突解脱仿真过程切片如图 5 所示, 试验结果如表 5~7 所示。可以看出: 随着飞行器数量增加, 飞行冲突占比从 3.23% 上升到 5.38%; 平均冲突解脱时间从 0.50 s 上升到 2.33 s; 任务成功率由 98% 下降到 88%; 计算开销从 0.08 min 上升到 0.33 min; 优先级和平稳性奖励函数是在保证有效解冲突核心目标前提下引入的规则函数, 其设计目的在于更合理地处理真实的解冲突场景, 提升解冲突策略的实用性。通过不同类型不同规模试验结果对比, 证明引入优先级、平稳性奖励函数之后解冲突策略性能依旧保持, 达到设计目标。

试验结果表明: ①随着智能体数量的增加, 系统的平均冲突解脱时间呈现出逐步上升的趋势, 成功率则有所下降, 然而, 尽管冲突解脱时间有所增加, 所提出的 MAPPO 算法依然能够保持较高的任务成功率, 证明了其在复杂多智能体系统中的鲁棒性; ②针对随机生成的 30 架次异构飞行器完成全部冲突解脱解算的计算时间开销在 20 s 以内, 具备实时性。

### 3.4 空域连通性对飞行冲突解脱的影响

在试验 2 基础上添加 3 个圆柱体禁飞区(模拟高楼或雷暴区)重复试验, 考察空域连通性受损对冲突解脱性能的影响。试验过程的直观演示切片如图 6 所示, 试验结果如表 8 所示。可以看出: 随着飞行器数量增加, 飞行冲突占比从 3.45% 上升到 5.41%; 平均冲突解脱时间从 1.30 s 上升到 4.36 s; 任务成功率由 95% 下降到 86%; 计算开销从 0.28 min 上升到 0.53 min。分析表明, 障碍物的引入增加了飞行路径规划难度, 导致飞行冲突增加、解脱时间延长、任务成功率下降、计算开销增加。原因在于, 障碍物限制了飞行器的机动空间。尽管如此, MAPPO 仍维持较高任务成功率, 表明其飞行路径规划算法有效; 且计算开销仍具备一定实时性, 至少可应用于预战术阶段。

### 3.5 与基准方法的性能对比

为进一步验证方法的优越性, 本节将 MAPPO 与 DQN 改进方法(基准方法)进行对比。DQN 采用独立 Q 学习, 每个飞行器有单独的 Q 网络(3 层全连接, 各 64 个神经元, ReLU 激活), 无全局价值

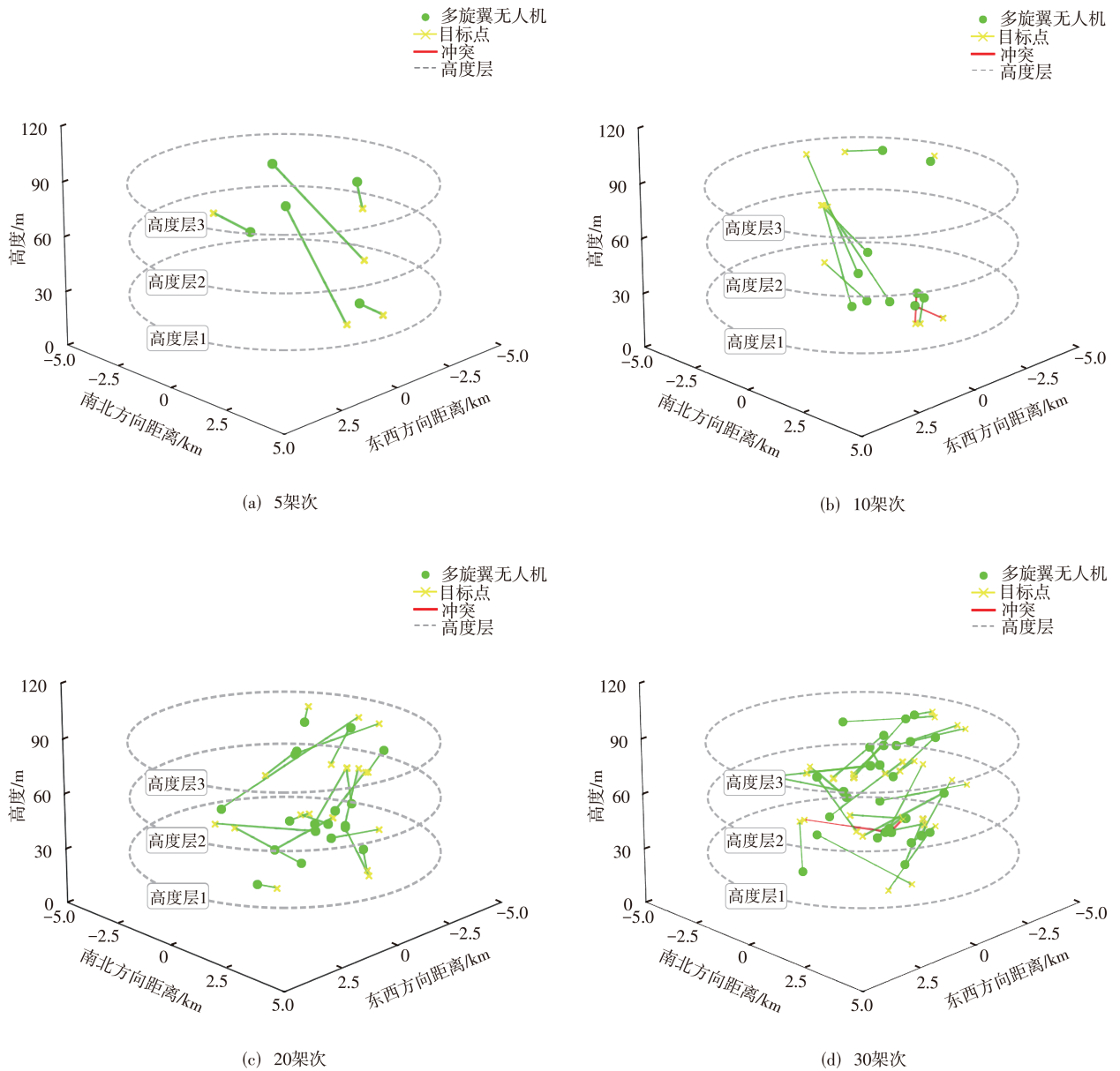


图 4 轻小型多旋翼无人机飞行冲突解脱仿真矢量图(试验 1)

Fig. 4 Simulation vector diagrams of flight conflict resolution for light small multi-rotor UAVs (experiment 1)

表 4 轻小型多旋翼无人机飞行冲突解脱仿真试验结果(试验 1)

Table 4 Simulation results of conflict resolution for light small multi-rotor UAVs (experiment 1)

飞行器数量/架次	飞行冲突占比/%	平均冲突解脱时间/s	任务成功率/%	计算时间/min
5	0.95	0.28	95	0.16
10	1.03	0.33	94	0.19
20	1.44	0.89	92	0.26
30	2.18	1.79	92	0.36

共享。试验设置飞行器数量为 30 架次(24 架次轻小型多旋翼无人机和 6 架次中大型固定翼飞机),运行 1 000 次,记录解脱时间和任务成功率,结果如表 9 所示。结果显示:MAPPO 的解脱时间和任务成功率分别为 2.14 s 和 90%;DQN 改进方法的解

脱时间和任务成功率分别为 4.39 s 和 88%;MAPPO 在冲突解脱时间上显著高于 DQN 改进方法,在任务成功率上略高于 DQN 改进方法。原因在于,MAPPO 的策略网络通过集中式训练捕捉全局依赖,而 DQN 改进方法的独立训练忽略了飞行器间的协同性,导致冲突解决效率降低。

## 4 结 语

(1)本文提出了一种适用于通用实时三维轻小型多旋翼无人机与中大型固定翼飞机冲突解脱场景的多智能体强化学习框架,为解决复杂动态环境中的轻小型多旋翼无人机与中大型固定翼飞机冲突问题提供了高效的解决方案;通过依次开展单纯轻小

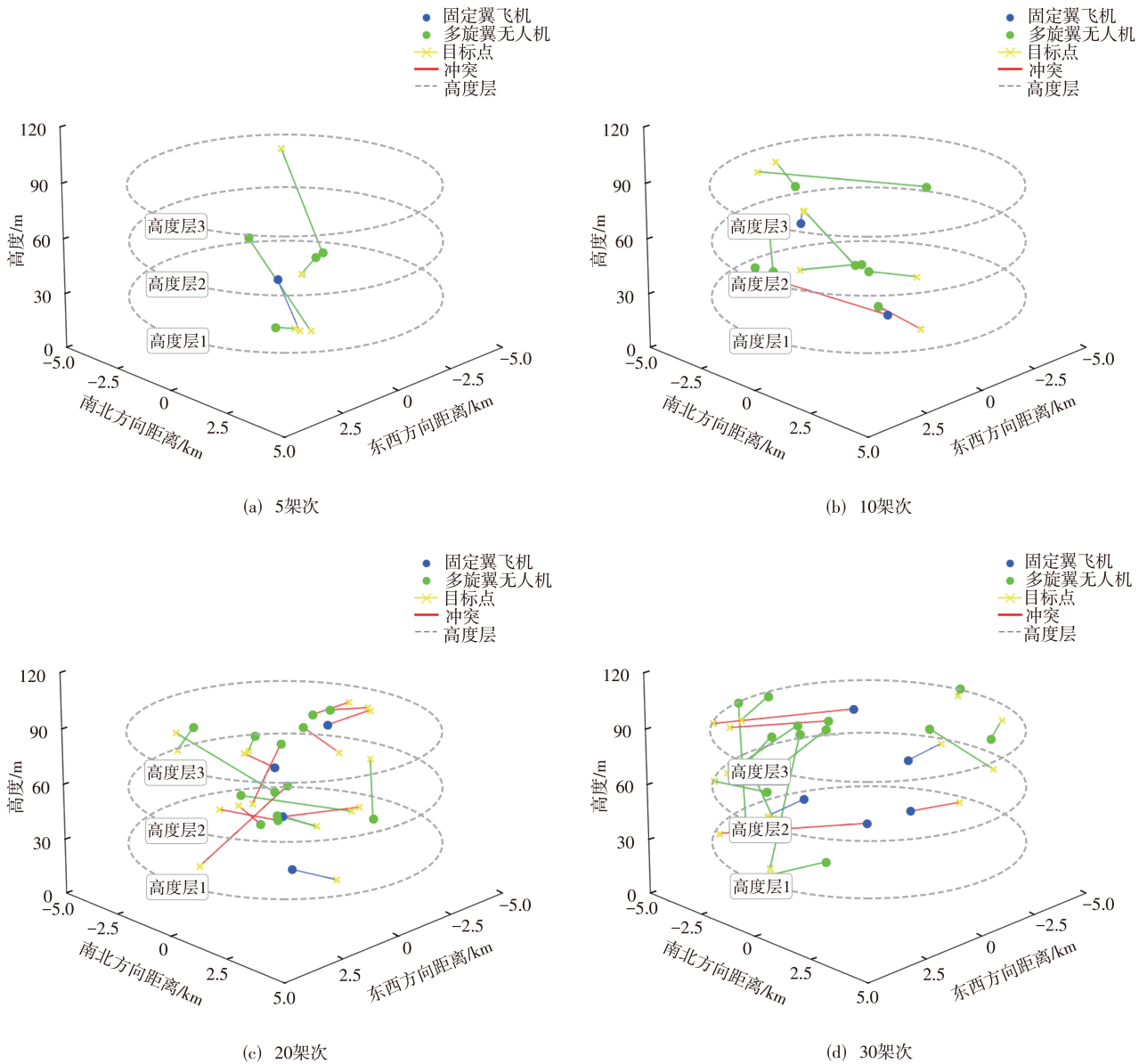


图5 完整奖励下异构飞行器冲突解脱仿真矢量图(试验2)

Fig. 5 Simulation vector diagrams of flight conflict resolution for heterogeneous aircraft under comprehensive reward (experiment 2)

表5 完整奖励下异构飞行器冲突解脱仿真结果(试验2)

Table 5 Simulation results of conflict resolution for heterogeneous aircraft under comprehensive reward (experiment 2)

飞行器数量/架次	飞行冲突占比/%	平均冲突解脱时间/s	任务成功率/%	计算开销/min
5	3.23	0.50	98	0.08
10	4.35	1.67	92	0.22
20	5.28	1.55	90	0.24
30	5.38	2.33	88	0.33

表6 无优先级奖励下异构飞行器冲突解脱仿真结果(试验3)

Table 6 Simulation results of conflict resolution for heterogeneous aircraft under non-prioritized reward (experiment 3)

飞行器数量/架次	飞行冲突占比/%	平均冲突解脱时间/s	任务成功率/%	计算开销/min
5	4.13	1.05	98	0.13
10	4.37	2.28	91	0.16
20	5.93	2.41	90	0.28
30	6.28	3.05	86	0.31

型多旋翼无人机间冲突解脱基准试验、异构飞行器间完整奖励下、无优先级奖励下、无平稳性奖励下冲突解脱仿真试验,以及在空域连通性受损环境下的异构飞行器冲突解脱仿真试验,证明所提反映任务差异性的优先级、平稳性奖励项在多智能体冲突解

脱中均具有有效性。

(2)采用集中式训练与分布式执行(CTDE)框架,构建了基于多智能体近端策略优化(MAPPO)的低空异构飞行器实时三维冲突解脱方法;为评估方法的扩展性,试验中测试了不同飞行器数量下的

表 7 无平稳性奖励下异构飞行器冲突解脱  
仿真结果(试验 4)

Table 7 Simulation results of conflict resolution for heterogeneous aircraft under non-stationary rewards (experiment 4)

飞行器数量/架次	飞行冲突占比/%	平均冲突解脱时间/s	任务成功率/%	计算开销/min
5	3.09	0.37	98	0.10
10	3.68	1.42	98	0.18
20	4.95	2.16	90	0.23
30	4.62	2.50	88	0.30

算法性能表现。结果表明:所提出的 MAPPO 算法针对随机生成的 30 架次异构飞行器规模,能够保持较高的任务成功率,在复杂多智能体系统中具有鲁棒性;完成全部冲突解脱解算的计算时间开销在

表 8 空域连通性受损下的异构飞行器冲突解脱仿真结果(试验 5)

Table 8 Simulation results of conflict resolution for heterogeneous aircraft in degraded airspace connectivity (experiment 5)

飞行器数量/架次	飞行冲突占比/%	平均冲突解脱时间/s	任务成功率/%	计算开销/min
5	3.45	1.30	95	0.28
10	5.15	1.87	93	0.31
20	5.18	3.00	88	0.43
30	5.41	4.36	86	0.53

1 min 以内,具有一定实时性。

(3) 本研究所提方法是在相对理想的情况下开展试验,在冲突环境的条件设定上还有进一步改进的空间;在未来研究中可立足轻小型多旋翼无人机

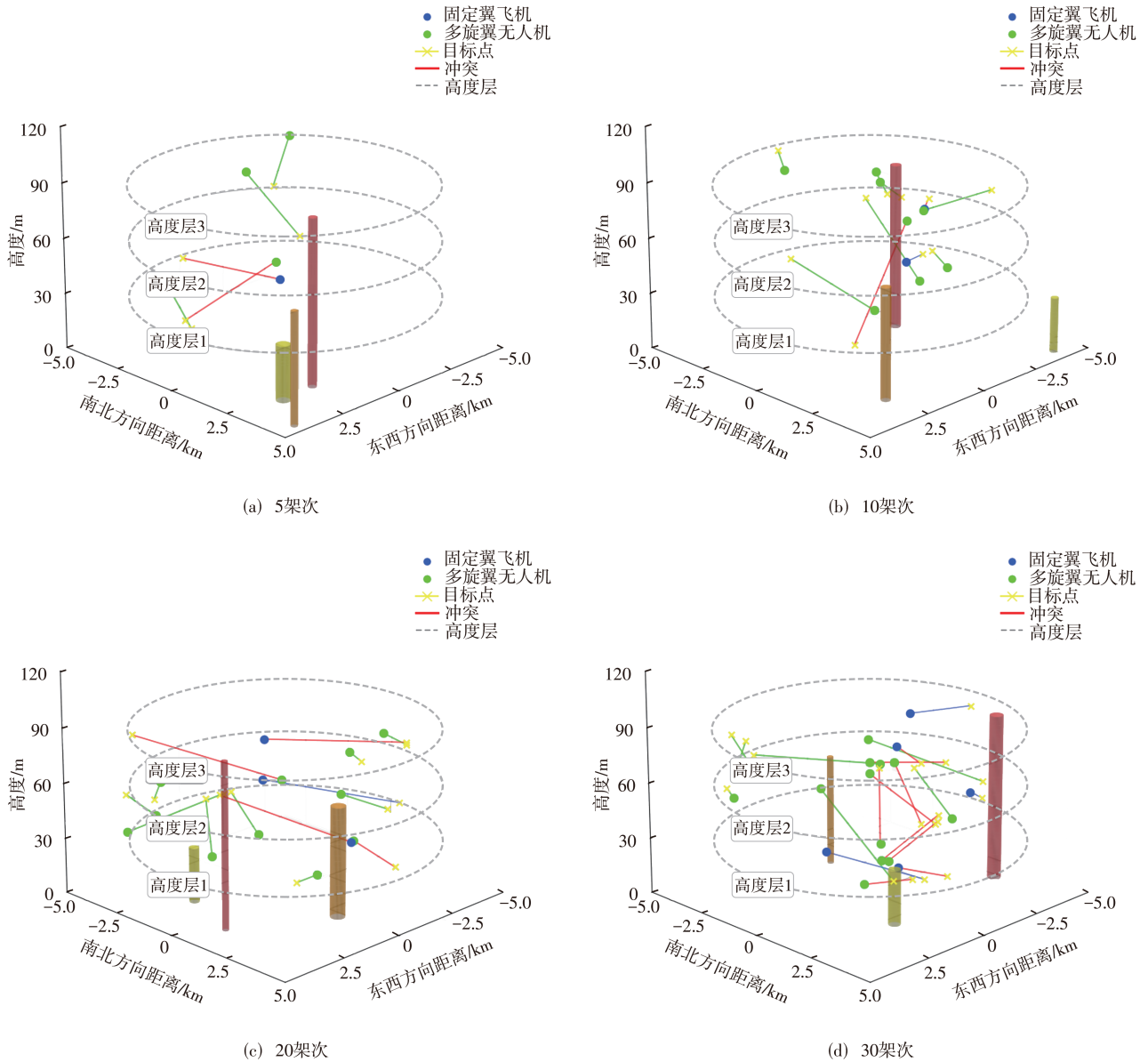


图 6 空域连通性受损的异构飞行器冲突解脱仿真矢量图(试验 5)

Fig. 6 Simulation vector diagrams of flight conflict resolution for heterogeneous aircraft in degraded airspace connectivity (experiment 5)

表9 MAPPO与DQN改进方法的冲突解脱性能对比

Table 9 Performance comparison of MAPPO and advanced DQN method

方法	平均冲突解脱时间/s	任务成功率/%
MAPPO	2.14	90
DQN改进方法	4.39	88

与中大型固定翼飞机更有针对性的场景飞行任务特点进行条件设定。对于所提方法而言,设计奖励函数十分关键,而平衡多维奖励项的影响是一个重要研究问题。在后续研究中将专题研究更加细致的量化方法,以更优地平衡各奖励项之间的影响。

### 参 考 文 献:

#### Reference :

- [ 1 ] 中国民用航空局. 关于促进民用无人驾驶航空发展的指导意见(征求意见稿)[EB/OL]. (2019-05-14)[2024-06-06]. [http://www.caac.gov.cn/HDJL/YJZJ/201905/t20190514\\_196175.html](http://www.caac.gov.cn/HDJL/YJZJ/201905/t20190514_196175.html).  
Civil Aviation Administration of China. Guiding opinions on promoting the development of civil unmanned aircraft (Draft for Public Comment)[EB/OL]. (2019-05-14)[2024-06-06]. [http://www.caac.gov.cn/HDJL/YJZJ/201905/t20190514\\_196175.html](http://www.caac.gov.cn/HDJL/YJZJ/201905/t20190514_196175.html).
- [ 2 ] PONS-PRATS J, ŽIVOJINOVIĆ T, KULJANIN J. On the understanding of the current status of urban air mobility development and its future prospects: Commuting in a flying vehicle as a new paradigm [J]. *Transportation Research Part E: Logistics and Transportation Review*, 2022, 166: 102868.
- [ 3 ] GARROW L A, GERMAN B J, LEONARD C E. Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 132: 103377.
- [ 4 ] 张洪海, 夷 珈, 李 珊, 等. 低空空域容量评估研究综述[J]. *交通运输工程学报*, 2023, 23(6): 78-93.  
ZHANG Hong-hai, YI Jia, LI Shan, et al. Review on research of low-altitude airspace capacity evaluation[J]. *Journal of Traffic and Transportation Engineering*, 2023, 23(6): 78-93.
- [ 5 ] 汤新民, 顾俊伟, 张 康, 等. 无人驾驶航空器自主探测与避让技术研究综述[J]. *交通运输工程学报*, 2026, 26(3): 1-24.  
TANG Xin-min, GU Jun-wei, ZHANG Kang, et al. Research review on the autonomous detect and avoid technologies for unmanned aerial vehicles [J]. *Journal of Traffic and Transportation Engineering*, 2026, 26(3): 1-24.
- [ 6 ] 中国民用航空局. 关于发布《国家空域基础分类方法》的通知[EB/OL]. (2023-12-21)[2026-02-27]. [http://www.caac.gov.cn/XXGK/XXGK/TZTG/202312/t20231221\\_222397.html](http://www.caac.gov.cn/XXGK/XXGK/TZTG/202312/t20231221_222397.html).  
Civil Aviation Administration of China. Notice on *Issuing the Basic Classification Method of National Airspace*. (2023-12-21)[2026-02-27]. [http://www.caac.gov.cn/XXGK/XXGK/TZTG/202312/t20231221\\_222397.html](http://www.caac.gov.cn/XXGK/XXGK/TZTG/202312/t20231221_222397.html).
- [ 7 ] 李诚龙, 屈文秋, 李彦冬, 等. 面向eVTOL航空器的城市空中运输交通管理综述[J]. *交通运输工程学报*, 2020, 20(4): 35-54.  
LI Cheng-long, QU Wen-qiu, LI Yan-dong, et al. Overview of traffic management of urban air mobility (UAM) with eVTOL aircraft [J]. *Journal of Traffic and Transportation Engineering*, 2020, 20(4): 35-54.
- [ 8 ] REICH P G. Analysis of long-range air traffic systems: Separation standards: I [J]. *Journal of Navigation*, 1966, 19(1): 88-98.
- [ 9 ] FIORINI P, SHILLER Z. Motion planning in dynamic environments using velocity obstacles [J]. *The International Journal of Robotics Research*, 1998, 17(7): 760-772.
- [ 10 ] BROOKER P. Lateral collision risk in air traffic track systems: A 'post-Reich' event model [J]. *Journal of Navigation*, 2003, 56(3): 399-409.
- [ 11 ] 刘 洋, 向锦武, 罗漳平, 等. 低空自由飞行短期冲突探测算法[J]. *北京航空航天大学学报*, 2017, 43(9): 1873-1881.  
LIU Yang, XIANG Jin-wu, LUO Zhang-ping, et al. Short-term conflict detection algorithm for free flight in low-altitude airspace [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2017, 43(9): 1873-1881.
- [ 12 ] HERNÁNDEZ-ROMERO E, VALENZUELA A, RIVAS D. Probabilistic multi-aircraft conflict detection and resolution considering wind forecast uncertainty [J]. *Aerospace Science and Technology*, 2020, 105: 105973.
- [ 13 ] 管祥民, 吕人力. 基于满意博弈论的复杂低空飞行冲突解脱方法[J]. *航空学报*, 2017, 38(增1): 120-128.  
GUAN Xiang-min, LYU Ren-li. Aircraft conflict resolution method based on satisfying game theory [J]. *Acta Aeronautica et Astronautica Sinica*, 2017, 38(S1): 120-128.
- [ 14 ] 张宏宏, 甘旭升, 孙静娟, 等. 基于STPA-TOPAZ的低空无人机冲突解脱安全性分析[J]. *航空学报*, 2022, 43(7): 255-267.  
ZHANG Hong-hong, GAN Xu-sheng, SUN Jing-juan, et al. Analysis of low altitude UAV conflict resolution safety based on STPA-TOPAZ [J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(7): 255-267.
- [ 15 ] 张启钱, 王中叶, 张洪海, 等. 基于SMILO-VTAC模型的复杂低空多机冲突解脱方法[J]. *交通运输工程学报*, 2019, 19(6): 125-136.  
ZHANG Qi-qian, WANG Zhong-ye, ZHANG Hong-hai, et al. SMILO-VTAC model based multi-aircraft conflict resolution method in complex low-altitude airspace [J]. *Journal of Traffic and Transportation Engineering*, 2019, 19(6): 125-136.
- [ 16 ] 陈运翔, 张建平, 王致远, 等. 基于机动避撞策略的低空多旋翼无人机安全间隔计算模型[J]. *航空学报*, 2025, 46(11): 349-365.  
CHEN Yun-xiang, ZHANG Jian-ping, WANG Zhi-yuan, et al. Safety separation calculation model for multi-rotor drones in low-altitude airspace based on avoidance strategy [J]. *Acta Aeronautica et Astronautica Sinica*, 2025, 46(11): 349-365.

- [17] 谷志鸣,高文明,魏潇龙,等. 基于TOPAZ的无人机冲突解脱安全评估技术[J]. 安全与环境学报,2016,16(5):51-56.  
GU Zhi-ming, GAO Wen-ming, WEI Xiao-long, et al. Safety assessment technology of UAV conflict resolution based on the TOPAZ method [J]. Journal of Safety and Environment, 2016, 16(5): 51-56.
- [18] 朱代武. 低空空域飞行冲突避让算法[J]. 交通运输工程学报, 2005,5(3):73-76.  
ZHU Dai-wu. Calculational methods of avoiding flight conflict in low altitude airspace[J]. Journal of Traffic and Transportation Engineering, 2005, 5(3): 73-76.
- [19] PHAM H, BALASOORIYAN P, YILMAZ Y, et al. Conflict resolution for unmanned aerial vehicles using deep reinforcement learning [J]. Journal of Intelligent & Robotic Systems, 2022, 95(3): 629-644.
- [20] LOQUERCIO A, MAQUEDA A I, DEL-BLANCO C R, et al. DroNet: Learning to fly by driving[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 1088-1095.
- [21] LIN C E, LAI Y H. UAV path prediction for CD&R to manned aircraft in a confined airspace for cooperative mission [J]. International Journal of Aerospace Engineering, 2018, 2018: 8759836.
- [22] JILKOV V P, LEDET J H, LI X R. Multiple model method for aircraft conflict detection and resolution in intent and weather uncertainty[J]. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(2): 1004-1020.
- [23] ZHAO X, LIU Y. Generalised single-agent reinforcement learning for multi-aircraft conflict resolution [J]. Aerospace Science and Technology, 2021, 112: 106649.
- [24] LAI Z, ZHENG Z, QIU S, et al. Multi-agent deep deterministic policy gradient for air traffic conflict resolution [J]. Aerospace Science and Technology, 2021, 115: 106797.
- [25] CHEN Y T, XU Y, YANG L, et al. General real-time three-dimensional multi-aircraft conflict resolution method using multi-agent reinforcement learning [J]. Transportation Research Part C: Emerging Technologies, 2023, 157: 104367.
- [26] DONG S, LI W, LIU S, et al. Deep reinforcement learning for multi-agent conflict resolution in 3D airspace [J]. Aerospace Science and Technology, 2021, 110: 106412.
- [27] BRITTAİN M, WEI P. Long short-term memory network for aircraft conflict detection and resolution [J]. Journal of Guidance, Control, and Dynamics, 2021, 44(2): 330-342.
- [28] DALMAU R, ALLARD E. Air traffic control using message passing neural networks and multi-agent reinforcement learning [C]// SIDs. 10th SESAR Innovation Days. Brussels: SESAR, 2020: 158-167.
- [29] YU C, VELU A, VINITSKY E, et al. The surprising effectiveness of PPO in cooperative multi-agent reinforcement games[C]// NeurIPS. 36th Conference on Neural Information Processing Systems, San Diego: NeurIPS, 2022: 24611-24624.
- [30] WACHI A, SHEN X, SUI Y. A survey of constraint formulations in safe reinforcement learning[C]// IJCAI. Proceedings of the 33rd International Joint Conference on Artificial Intelligence. California: IJCAI, 2024: 8262-8271.