

覃睿, 张晓玲. 基于改进 DQN 算法的多无人机城市火灾协同灭火研究[J]. 安全与环境学报, 2026, 26(6): 2359–2370.

QIN R, ZHANG X L. Cooperative firefighting using multi-UAVs in urban settings based on an improved DQN algorithm[J]. Journal of Safety and Environment, 2026, 26(6): 2359–2370.

基于改进 DQN 算法的多无人机城市火灾协同灭火研究*

覃睿, 张晓玲

(中国民航大学交通科学与工程学院, 天津 300300)

摘要: 针对城市火灾中建筑密集、道路堵塞导致传统消防车救援困难、单无人机灭火能力不足的问题, 提出了一种基于改进深度 Q 网络 (Improved Deep Q -Network, IDQN) 的多无人机协同灭火路径规划算法。首先, 建立基于相同时间间隔的多无人机单目标决策模型; 其次, 通过优先经验回放机制提高算法训练稳定性, 采用 26 向三维动作空间增强机动性, 并构建包含方向引导、步数惩罚、到达奖励和碰撞惩罚的多目标复合奖励函数; 最后, 基于任务优先级, 采用层次化加权法在 MATLAB 平台进行仿真训练, 实现多无人机协同灭火路径规划。仿真结果表明: IDQN 算法多无人机协同灭火方案任务完成时长相比单无人机作业缩短 58.0%, 相比粒子群优化 (Particle Swarm Optimization, PSO) 算法缩短 27.8%, 显著提升了城市火灾救援的响应速度和资源利用率。

关键词: 安全工程; 深度强化学习算法; 多无人机协同; 城市火灾; 协同灭火; 路径规划

中图分类号: X45; X959 **文献标志码:** A **文章编号:** 1009-6094(2026)06-2359-12

DOI: 10.13637/j.issn.1009-6094.2025.1197

0 引言

随着城市化进程的加快, 高楼密集和交通拥堵形成的现代城市典型特点, 使得城市火灾救援难度显著增加^[1], 尤其在建筑密集、道路狭窄且车辆随意停放的老旧城区和商业区, 传统消防车辆常因交通拥堵和消防通道堵塞^[2]而无法及时抵达, 这些人员密集、功能复杂的场所一旦发生火灾, 极易造成重大伤亡和财产损失^[3], 这种救援延误在高层建筑火灾中尤为严重。消防无人机 (Unmanned Aerial Vehicle, UAV) 因其灵活性、快速响应和能够避开障碍物的特点, 逐渐成为高层消防救援的重要工具。通过穿越密集的高楼建筑物搭载干粉灭火器或者配备水基灭火弹执行灭火任务, 有效弥补了人工灭火和高空救援的不足。无人机参战模式逐渐由单无人机转为多无人机协同灭火作战^[4], 从同一消防车停放点以固定时间间隔依次起飞, 沿智能规划的协同路径到达火灾点执行灭火任务并依次返航, 显著提升了灭火效率和作战效能。

复杂环境下的多无人机路径规划已成为当前研究热点, 学者们提出了多种优化方法, 主要包括传统智能优化算法与深度强化学习两大方向。传统方法如粒子群算法^[5]、遗传算法^[6]、人工蜂群算法^[7]和

灰狼算法^[8]等在静态或简单动态环境中表现出良好的寻优能力。2023 年, 郭志明等^[9]创新性地采用 Logistic 函数非线性递减策略改进蝗虫优化算法, 构建了多无人机威胁环境下的协同航迹规划新方法, 在收敛速度和航迹代价优化方面均取得突破性进展, 但在高动态、多约束的城市火灾场景中常面临收敛速度慢和适应性不足的问题。近年来, 深度强化学习因强大的环境感知与决策能力受到广泛关注。2025 年, 王巧等^[10]创新性地提出了基于深度强化学习的价值优先电梯调度算法, 通过优化奖励函数和目标函数, 在高层建筑火灾疏散中实现了电梯与楼梯疏散时间的最优平衡。2025 年, Liu 等^[11]提出了基于 YOLOv8 与动态数据存储器的深度强化学习方法, 为无人机动态避障建立了“视觉-决策-控制”协同优化新体系, 有效引导无人机远离邻近障碍物。针对深度强化学习算法, 学者们为了提高算法在复杂环境中的训练效果, 引入混合优先经验回放^[12], 设计无人机的状态观测空间、动作空间及奖赏函数^[13], 提出分层深度强化学习模型^[14]或引入近端策略优化算法^[15]等。

目前, 针对无人机火灾场景灭火救援有了一定的研究。2020 年, 张小孟等^[7]首次提出基于整数编码改进人工蜂群算法的多无人机山火灭火救援规划

* 收稿日期: 2025-07-10

作者简介: 覃睿, 教授, 硕士, 从事低空经济发展规划、低空空域与低空交通发展规划、新一代通用航空应用创新研究, rtan@cauc.edu.cn。

方法,建立了时变目标价值任务模型,仿真表明,该方法寻优效率显著提升。宋睿等^[16]针对高原森林灭火,提出基于数字高程模型(Digital Elevation Model, DEM)数据构建三维地理空间模型,并改进A*算法有效解决了无人机的路径优化问题。王慧颖等^[17]针对同区域多火点动态救援场景,提出了一种基于时变火情价值的无人机群路径规划方法,使多目标救援总价值提升11.65%。李鸿一等^[2]提出了一种基于随机采样的多无人机高层消防协同搜索规划方法,通过最优搜索点分布、交易规则优化和改进和声搜索算法实现高效路径规划。2023年,为了激励无人机在城市环境火灾中的研究,Perez-Saura等^[18]首次组织了一项创新性无人机城市火灾救援竞赛,通过设计“障碍穿越—目标定位—精准投弹”的完整任务链,开发了基于多模块协同架构的智能任务规划系统,为无人机城市火灾救援研究建立了首个标准化测试平台。

对相关文献进行分析,首先,现有成果多集中于森林或室内火灾,缺乏专门面向城市火灾场景的多无人机协同路径规划方法;其次,已开发的智能任务规划系统大多针对单无人机平台,无法满足多机协同灭火对时效性和资源调配的迫切需求;最后,现有算法的优化设计在飞行时间、避障安全与任务完成等关键目标之间缺乏有效协调,导致算法难以适应城市火灾救援动态复杂的环境要求。本文针对城市环境火灾灭火救援场景,提出一种基于改进深度Q网络(Improved Deep Q-Network, IDQN)的多无人机协同灭火路径规划算法,根据城市高层建筑火灾救援的环境特点、任务需求及消防无人机的性能参数,进行针对性优化:

(1) 建立基于相同时间间隔的多无人机单目标时序协同决策模型,实现多无人机有序抵达火灾点灭火;

(2) 基于灭火任务的优先级,采用层次化加权方法,分别为时间代价、能量消耗代价和飞行距离代价分配权重,从而构建加权路径优化模型;

(3) 设计面向城市密集障碍环境的动态状态表示方法,增强算法对复杂三维空间特征的感知能力;

(4) 设置适用于城市火灾三维场景的状态空间和动作空间,构建包含方向引导、步数惩罚、到达奖励和碰撞惩罚的多目标复合奖励函数,使无人机能自动平衡路径效率与飞行安全。

1 多无人机协同灭火模型

1.1 灭火场景描述

传统城市火灾救援模式主要依赖消防车等地面

设备,但由于交通拥堵、高层建筑遮挡及火场核心区域难以快速接近等限制,救援效率常面临严峻挑战。无人机凭借其快速响应能力、三维空间机动性及不受地面路况制约的优势,逐渐成为城市火灾救援中不可或缺的辅助工具。然而,单架无人机受限于载荷和续航能力,难以独立应对大规模火灾现场需求,因此须构建多无人机协同系统,以提升整体救援效能。本文所述系统由一个集成三架消防无人机的移动平台构成,其系统架构与单架无人机路径如图1所示。以消防车临时停放点作为无人机起降节点,三架无人机按设定的时间间隔依次出发,须在三维空间中实时规避动静态障碍物,以快速、安全地抵达火灾点并执行灭火任务;任务完成后,无人机须沿返程航迹返回起降点,最终实现多无人机系统在灭火任务中的协同作业。因此,该系统主要解决3个问题:(1)多无人机在灭火任务中的协同决策问题;(2)深度强化学习算法在城市火灾场景下的适应性问题;(3)多无人机从起降点到火灾点再返回的路径规划问题。

1.2 目标函数建立

在无人机协同火灾救援任务中,优化目标通常需要考虑任务完成时间、能量消耗及飞行距离等关键因素。为此,构建一个多目标优化模型,旨在通过加权求和的方式平衡不同任务的优先级,模型的变量定义如表1所示。式(1)为任务完成时间代价,式(2)为任务完成能量消耗代价,式(3)为任务完成飞行距离代价。

$$T_{\text{total}} = \max_i \left(t_i^{\text{complete}} + \frac{\|P_{\text{base}} - P_{\text{fire}}\|}{v_i} \right) \quad (1)$$

$$E_{\text{total}} = P_h t_h + P_f t_f \quad (2)$$

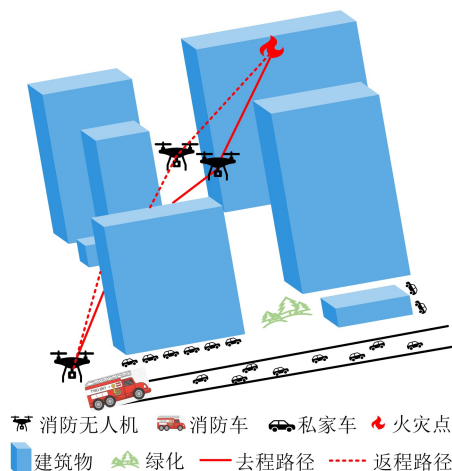


图1 多无人机系统城市火灾协同灭火示意图

Fig.1 Schematic diagram of cooperative firefighting in urban environments using multi-UAV systems

表 1 变量定义表
Table 1 Variable definition table

类型	符号	含义	单位
时间变量	T_{total}	总任务完成时间	s
	t_i^{complete}	第 i 架无人机完成灭火任务的时间	s
	t_{h}	悬停时间	s
	t_{f}	飞行时间	s
	t_i^{start}	系统开始时间	s
	t_i^{depart}	第 i 架无人机起飞时间	s
	t_i^{arrive}	第 i 架无人机到达火灾点时间	s
	$t_{\text{extinguish}}$	单架无人机完成灭火所需时间	s
	Δt	相邻无人机之间的起飞时间间隔	s
	t_{stay}	到达火灾点后停留灭火时间	s
位置和距离 变量	t_{total}	无人机续航时间	min
	$\mathbf{P}_{\text{base}}$	起降地位置坐标	m
	$\mathbf{P}_{\text{fire}}$	火灾点位置坐标	m
	$\mathbf{P}_i(t)$	第 i 架无人机在时刻 t 的三维位置	m
	$\mathbf{O}_k$	第 k 个障碍物的位置坐标	m
	$\mathbf{P}_{\text{l}}$	最终降落点坐标	m
	$\mathbf{P}_{\text{o}}$	起飞点坐标	m
	$\mathbf{P}_t$	无人机 t 时刻的位置坐标 (x_t, y_t, z_t)	m
	$\mathbf{G}_t$	目标点坐标 (x_g, y_g, z_g)	m
	$d^{o,j}$	第 j 个障碍物到无人机的相对距离	m
	$d_{\text{safe}}^{\text{obs}}$	无人机与障碍物的最小安全距离	m
	$d_{\text{safe}}^{\text{uav}}$	无人机间的最小安全距离	m
	D_{max}	初始距离	m
其他变量	$\ \cdot\ $	欧氏距离	m
	v_i	第 i 架无人机的速度	m/s
	v_t	无人机在 t 时刻的速度	m/s
	a_t	无人机在 t 时刻的加速度	m/s ²
	P_{h}	悬停功率	W
	P_{f}	飞行功率	W
	E_{total}	总能量消耗	J
	D_{total}	总飞行距离	m
	$R_{\text{operation}}$	无人机作业半径	m
	Z_t	无人机在 t 时刻的高度	m

$$D_{\text{total}} = \sum_{i=1}^n (\|\mathbf{P}_{\text{fire}} - \mathbf{P}_{\text{base}}\| + \|\mathbf{P}_{\text{base}} - \mathbf{P}_{\text{fire}}\|) \quad (3)$$

将各分项加权求和得到多目标优化函数,即

$$f = \min(w_1 T_{\text{total}} + w_2 E_{\text{total}} + w_3 D_{\text{total}}) \quad (4)$$

式中 w_1 、 w_2 、 w_3 为权重系数,分别取值为 $w_1 = 0.6$ 、 $w_2 = 0.3$ 、 $w_3 = 0.1$,基于任务优先级通过层

次化加权方法确定:时间目标最为关键,赋予主导性权重 0.6,以优先保障快速响应;能量目标关乎安全返航,赋予显著但次级的权重 0.3,以平衡速度与耗能;距离作为优化补充,赋予最低权重 0.1,用于在路径代价相近时细化选择。

1.3 约束条件

多无人机在城市火灾协同灭火任务中需考虑多方面的影响。式(5)约束所有无人机必须按预定时间间隔依次出发,式(6)约束无人机在飞行过程中必须避开所有障碍物,式(7)为无人机间的动态防撞约束,式(8)约束所有无人机最终必须返回起点,式(9)表示无人机到达火灾点和完成灭火任务的时间约束。

$$t_i^{\text{start}} + (i-1)\Delta T \leq t_i^{\text{depart}} \leq t_i^{\text{start}} + i\Delta T \quad (5)$$

$$\|P_i(t) - O_k\| \geq d_{\text{safe}}^{\text{obs}} \quad \forall t, \forall k \quad (6)$$

$$\|P_i(t) - P_j(t)\| \geq d_{\text{safe}}^{\text{uav}} \quad \forall t, \forall i \neq j \quad (7)$$

$$P_1 = P_o \quad (8)$$

$$t_i^{\text{complete}} = t_i^{\text{arrive}} + t_{\text{extinguish}} \quad (9)$$

1.4 时序协同灭火模型

针对城市火灾场景中的多无人机协同需求,采用固定时间间隔的递进式起飞方案。

$$t_i^{\text{depart}} = t_{i-1}^{\text{depart}} + \Delta t \quad \Delta t = 40 \text{ s}, i \in \{2, 3\} \quad (10)$$

$$t_{\text{slay}} = 30 \text{ s} \quad (11)$$

2 改进深度强化学习算法构建

2.1 深度强化学习算法

强化学习的核心理念是智能体主动与环境进行交互,在探索中通过行为动作获得奖励,模型如图2所示。

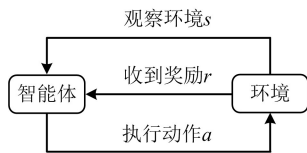


图2 强化学习模型

Fig.2 Reinforcement learning model

Q-learning 是一种基于值函数的无模型学习方法,通过学习状态动作对获得所对应的评价函数 $Q(s, a)$,即当前状态执行动作后获得的奖励反馈。Q-learning 的更新公式为

$$\begin{cases} Q(s, a) + \alpha(r + D) \rightarrow Q'(s, a) \\ D = \gamma \max_{a'} Q(s', a') - Q(s, a) \end{cases} \quad (12)$$

式中 s 表示当前状态; a 为当前动作; s' 表示下一个状态; a' 为下一个状态的动作; γ 为学习率; α 为折扣因子; r 表示即时奖励; $Q'(s, a)$ 表示更新后的 Q 值。

深度强化学习通过结合强化学习的决策优化能力与深度神经网络的高维表征能力,实现了复杂环境下的智能策略学习。其中, DQN 算法采用经验回放机制提高数据利用率并稳定训练过程,同时引入目标网络技术降低 Q 估计波动,从而有效解决了传统 Q-learning 在高维状态空间中的局限性问题。

2.2 改进深度强化学习算法

2.2.1 优先经验回放

经验回放是在训练过程中存储智能体与环境交互产生的经验样本,然后从存储的经验中随机抽取样本进行训练。优先经验回放是对传统 DQN 算法中均匀采样经验回放方法的改进,它根据每条经验的重要性赋予不同的采样概率,其中重要性通过时序差分(Temporal Difference, TD)误差来衡量。优先经验回放优先选择那些对训练更有帮助的经验进行学习,从而提高训练稳定性,并适应复杂任务的需求。每条经验的 TD 误差用于衡量其重要性,误差越大,表示该经验对网络更新的贡献越大。TD 误差的计算公式为

$$d_t = |r + \gamma \max_{a'} Q(s', a', \theta_t) - Q(s, a, \theta)| \quad (13)$$

式中 d_t 为 TD 误差; θ 为当前网络参数; θ_t 为目标网络参数; $Q(s, a, \theta)$ 为预测值; $r + \gamma \max_{a'} Q(s', a', \theta_t)$ 为目标值。

每条经验的采样概率与其 TD 误差成正比,且为了抵消优先采样带来的偏差,引入重要性采样权重,即

$$P(i) = \frac{p_i^\sigma}{\sum_i p_i^\sigma} \quad (14)$$

$$w_i = \left[\frac{1}{N} \times \frac{1}{P(i)} \right]^\beta \quad (15)$$

式中 p_i 为第 i 条经验的优先级; σ 为控制优先程度的超参数; N 为经验池容量,即经验池中最多可存储的经验样本总数; β 为控制权重调整的超参数; w_i 为第 i 条经验的重要性采样权重,是一个标量。

2.2.2 状态及动作空间设计

状态空间是描述无人机在完成所有可能状态的集合,用来表示环境中的无人机状态和相关信息。状态空间帮助无人机感知自身位置、速度及加速度,避开障碍物并接近目标点,状态空间可以定

义为

$$S_t = \{P_t, G_t, Z_t, t_{\text{total}}, d^{0,0}, \dots, d^{0,k-1}\} \quad (16)$$

$$x_{t+1} = x_t + v_t \Delta t + \frac{1}{2} a_t (\Delta t)^2 \quad (17)$$

$$R_{\text{operation}} = \|(x_{t+1}, y_{t+1})\|_2 \quad (18)$$

式中 P_t 表示无人机在 t 时刻的当前位置为三维坐标 (x_t, y_t, z_t) ; G_t 表示目标点坐标 (x_g, y_g, z_g) ; Z_t 为无人机在 t 时刻的高度 $Z_t \in [0, 200]$ m; t_{total} 为无人机的续航时间, $t_{\text{total}} \leq 30$ min; $d^{0,j}$ 表示第 j 个障碍物到无人机在 (x, y, z) 方向上的相对距离 $(x^{0,j}, y^{0,j}, z^{0,j})$, 其中 $j = 0, \dots, k-1$, k 为障碍物总数; v_t 为无人机在 t 时刻的速度, $v_t \in [0, 8.33]$ m/s; a_t 为无人机在 t 时刻的加速度, $a_t \in [-3, 3]$ m/s²; $R_{\text{operation}}$ 为无人机作业半径, $R_{\text{operation}} \leq 5\ 000$ m。

无人机基于当前策略评估环境状态,从 26 个三维移动方向中选择预期回报最高的动作执行,通过策略优化最大化长期奖励。如图 3 所示, O 点为无人机当前位置,动作空间分为三层:上层 9 个向上移动方向,中层 8 个水平移动方向,下层 9 个向下移动方向。单步步长为 3 单位,支持垂直、水平及斜向组合移动,实现灵活三维机动。

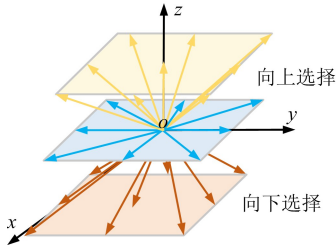


图 3 无人机动作方向图

Fig. 3 Diagram of UAV action directions

2.2.3 奖励函数设计

针对多无人机灭火任务的特点,本算法设计了多目标复合奖励函数,通过以下改进解决关键问题。

(1) 方向奖励。

通过设置动态方向奖励以激励无人机缩短与火灾点的欧氏距离,解决稀疏奖励问题,如式(19)和(20)所示。

$$r_d = 10(\|p - g\|_2 - \|p' - g\|_2) \quad (19)$$

$$\|p - g\|_2 = \sqrt{(x_i - x_g)^2 + (y_i - y_g)^2 + (z_i - z_g)^2} \quad (20)$$

式中 r_d 表示方向奖励值,为一个标量; p 表示无人机当前状态的位置坐标, m; p' 表示无人机下一状态的位置坐标, m; g 为目标点位置坐标, m; $\|p - g\|_2$

表示当前状态到火灾点的欧氏距离, m; $\|p' - g\|_2$ 表示下一状态到火灾点的欧氏距离, m。

(2) 步数惩罚。

步数惩罚是一种负向激励机制,旨在优化无人机的路径规划效率,避免无效探索或冗余移动。该惩罚项对每一步行动施加固定的小额负奖励,促使无人机在训练过程中优先选择更短的路径,如式(21)所示。

$$r_s = -0.05 \times \left(1 - \frac{\|p_i - g\|_2}{D_{\text{max}}}\right) \quad (21)$$

式中 r_s 表示步数惩罚奖励值,为一个标量; p_i 表示无人机在第 i 步时的位置坐标, m; D_{max} 表示最大初始距离, m。

(3) 到达奖励。

采用渐进式奖励机制,在接近目标时提供梯度信号,最终抵达时给予大额奖励,如式(22)所示。

$$r_g = \begin{cases} 1000 & \|p_t - g\|_1 = 0 \\ 0 & \text{其他} \end{cases} \quad (22)$$

式中 r_g 表示到达奖励值,为一个标量; p_t 表示无人机在 t 时刻的位置坐标, m; $\|p_t - g\|_1$ 表示无人机与火灾点的 L_1 范数距离(即曼哈顿距离), m。

(4) 碰撞惩罚。

根据无人机与障碍物的实时距离,动态调整惩罚强度,惩罚呈指数增长,如式(23)所示。

$$r_c = \begin{cases} -100 \exp\left(-\frac{d^2}{2\sigma^2}\right) & d \leq 3 \text{ m} \\ 0 & \text{其他} \end{cases} \quad (23)$$

式中 r_c 表示碰撞惩罚奖励值,为一个标量; $d = \|p_t - g\|_2$ 为到最近障碍物的欧氏距离; $\sigma = 2$ m 为控制惩罚衰减速率。

总奖励函数 r 由各奖励函数值线性相加得到,即

$$r = r_d + r_s + r_g + r_c \quad (24)$$

2.2.4 算法流程图

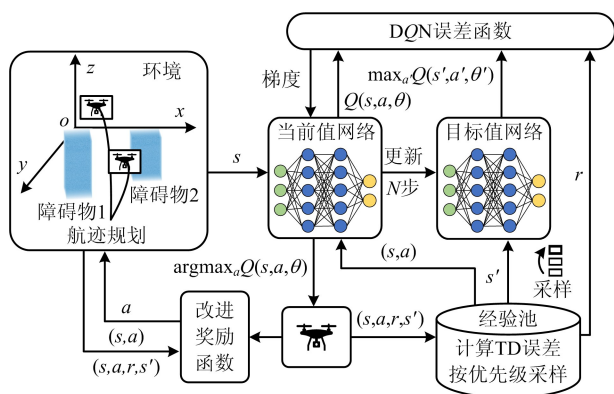
IDQN 算法流程图如图 4 所示。

3 试验仿真与结果分析

3.1 仿真试验参数设置

试验平台搭载 R7-6800H 8 核处理器、16 GB 内存和 RTX 3050 显卡,操作系统为 Windows 11,仿真软件使用 MATLAB R2024b。试验相关超参数值如表 2 所示。

选用 TH-MH-30 多旋翼无人机作为消防平台,该机型具备卓越的飞行稳定性、载重能力和抗风



图中 s 表示无人机当前状态, 包含位置、速度、加速度等本体信息, 以及与障碍物、火灾点间的距离等环境信息; a 表示无人机当前动作; r 表示无人机执行当前动作后获得的奖励值; θ 表示当前网络参数。

图4 IDQN 算法流程图

Fig. 4 Flow diagram of the IDQN algorithm

表2 算法相关参数

Table 2 Algorithm related parameters

参数符号	参数名	参数值
α	学习率	10^{-3}
γ	折扣因子	0.99
σ	超参数	0.8
ε_0	初始探索率	0.9
ε_d	探索衰减率	10^{-3}
δ	时间惩罚系数	10
μ	惩罚增速	1.2
W_1	隐藏层1 神经元数量	256
W_2	隐藏层2 神经元数量	128
N	经验回放池容量	10^6

性能,配合高性能电池与先进飞行控制系统,可满足消防任务需求。其基本参数如下:最大负载能力为 30 kg,最大续航时间为 30 min,最大巡航速度为 30 km/h,最大作业高度为 200 m,机身长 1 150 mm、宽 1 150 mm、高 615 mm,作业半径为 5 km,旋翼数量为 8。

3.2 仿真环境建模

以南京市雨花台区 6 号楼火灾事故为案例,在仿真试验中,设置了一个长 200 m、宽 200 m、高 150 m 的三维区域,楼层建筑物为 8 幢,试验部署 3 架无人机。基于失火小区三维地图数据,通过 MATLAB 构建障碍物模型,如图 5 所示。建筑物长方体的顶点坐标可以通过一个统一的公式表示,假设长方体

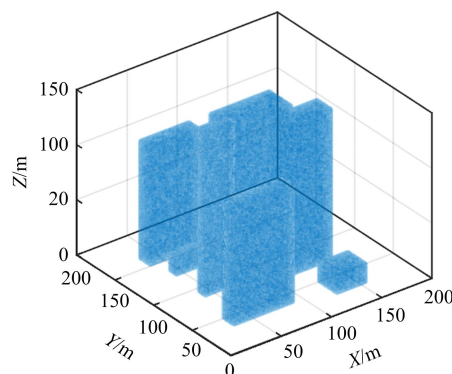
的起始点为 (x, y, z) , 长度为 l , 宽度为 w , 高度为 h , 则其他顶点坐标可表示为

$$V_i = (x + l, y + w, z + h) \quad (25)$$

各建筑物的参数信息如表 3 所示。



(a) 小区附近三维地图



(b) 小区附近MATLAB仿真图

图5 小区三维环境与仿真图

Fig. 5 Three-dimensional environment and simulation diagram of the residential area

表3 建筑物参数

Table 3 Building parameters

建筑物	起始点坐标/m	长度/m	宽度/m	高度/m
1 号楼	(30, 35, 0)	60	15	108
2 号楼	(40, 85, 0)	30	10	150
3 号楼	(105, 70, 0)	50	20	138
4 号楼	(42, 125, 0)	25	10	18
5 号楼	(35, 150, 0)	50	15	108
6 号楼	(105, 150, 0)	60	15	108
7 号楼	(155, 135, 0)	10	25	108
幼儿园	(125, 25, 0)	30	25	18

3.3 仿真结果分析

基于 IDQN 算法分别对 3 架无人机路径规划进

行了 300 次训练,1 号无人机训练奖励值如图 6 所示,2 号无人机训练奖励值如图 7 所示,3 号无人机训练奖励值如图 8 所示。图 6~8 中浅色折线表示每次训练的即时奖励值,深色折线代表平均奖励值,点划线为 Q 值,表示无人机在特定状态下执行某动作后能获得的预期累计折扣奖励。通过对 3 架无人机训练奖励值图的分析可以看出,该协同灭火算法展现出良好的收敛特性和学习效果:(1)3 架无人机的平均奖励值均随训练轮数增加呈稳定上升趋势,最终维持在较高水平,表明算法能有效学习到最优策略;(2)即时奖励值的波动幅度逐渐减小,表明训练过程趋于稳定;(3) Q 值曲线整体呈稳定增长态势,反映了无人机对长期累积奖励的预测能力在逐步提升。图 8 中箭头所指处为第 1 000 轮的 Q 值 (160.748 6),显示出模型在训练过程中的收敛状态。

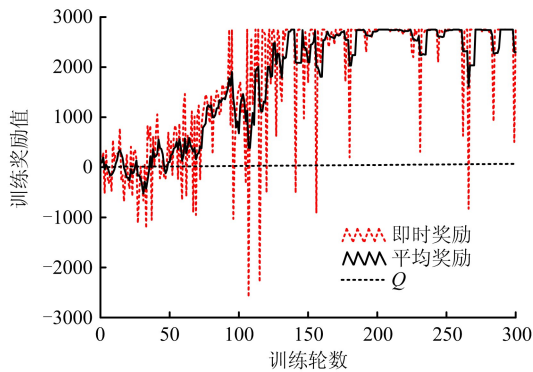


图 6 1 号无人机训练奖励值图

Fig. 6 Diagram of training reward value for UAV No. 1

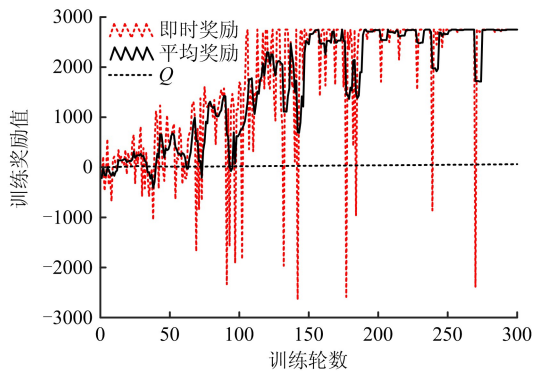


图 7 2 号无人机训练奖励值图

Fig. 7 Diagram of training reward value for UAV No. 2

仿真选定起始坐标 (24, 20, 1) 和终点坐标 (140, 140, 90), 基于 MATLAB 软件对 IDQN 算法进行仿真运行, 并选取粒子群优化 (Particle Swarm

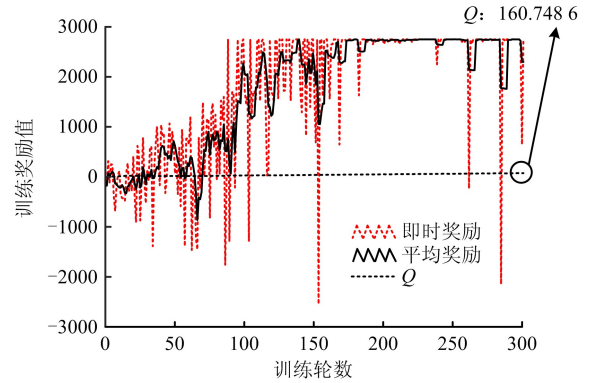


图 8 3 号无人机训练奖励值图

Fig. 8 Diagram of training reward value for UAV No. 3

Optimization, PSO) 算法作为对比算法进行分析。IDQN 算法多无人机路径规划结果如图 9 所示, PSO 算法多无人机路径规划结果如图 10 所示。

基于 IDQN 算法, 3 架无人机协同灭火任务的关键时序参数与路径性能优化结果如表 4 所示。仿真结果表明, 多无人机协同灭火方案展现出显著的时空协同优势。在时间效率方面, 3 架无人机采用梯队式调度策略, 间隔 40 s 依次出发, 并行任务执行将总任务耗时控制在 156 s。较单机完成所有任务点的 371.4 s 缩短了 58.0%, 时间效益提升显著, 展现出良好的多无人机协同灭火路径规划效果。该协同系统实现了多维性能平衡:(1)3 架无人机严格按照 40 s 时间间隔分时起飞, 且各无人机保持一致的 30 s 核心灭火作业时长, 确保灭火效果的稳定性;(2)通过智能路径规划, 后续无人机较首架分别缩短航程 9.0% 和 12.5%, 验证了协同算法的学习优化能力;(3)相比单无人机作业, 多无人机协同灭火方案通过路径规划优化, 将任务完成时间缩短 58.0%, 显著提升了城市火灾救援的响应速度和资源利用率。

基于 PSO 算法, 3 架无人机协同灭火任务的关键时序参数与路径性能优化结果如表 5 所示。IDQN 算法相比 PSO 算法在任务执行时间上平均缩短约 27.8%, 且飞行路径长度整体更优, 总飞行距离减少约 12.5%, 显著提升了多无人机协同作业的效率和经济性。

多无人机性能雷达图如图 11 所示, 从五个维度对 3 架无人机的性能进行了综合评估。图 11 中采用 0~1 的刻度, 间隔为 0.2, 清晰量化了各指标差异, 多边形轮廓直观揭示了机型间的性能平衡与专

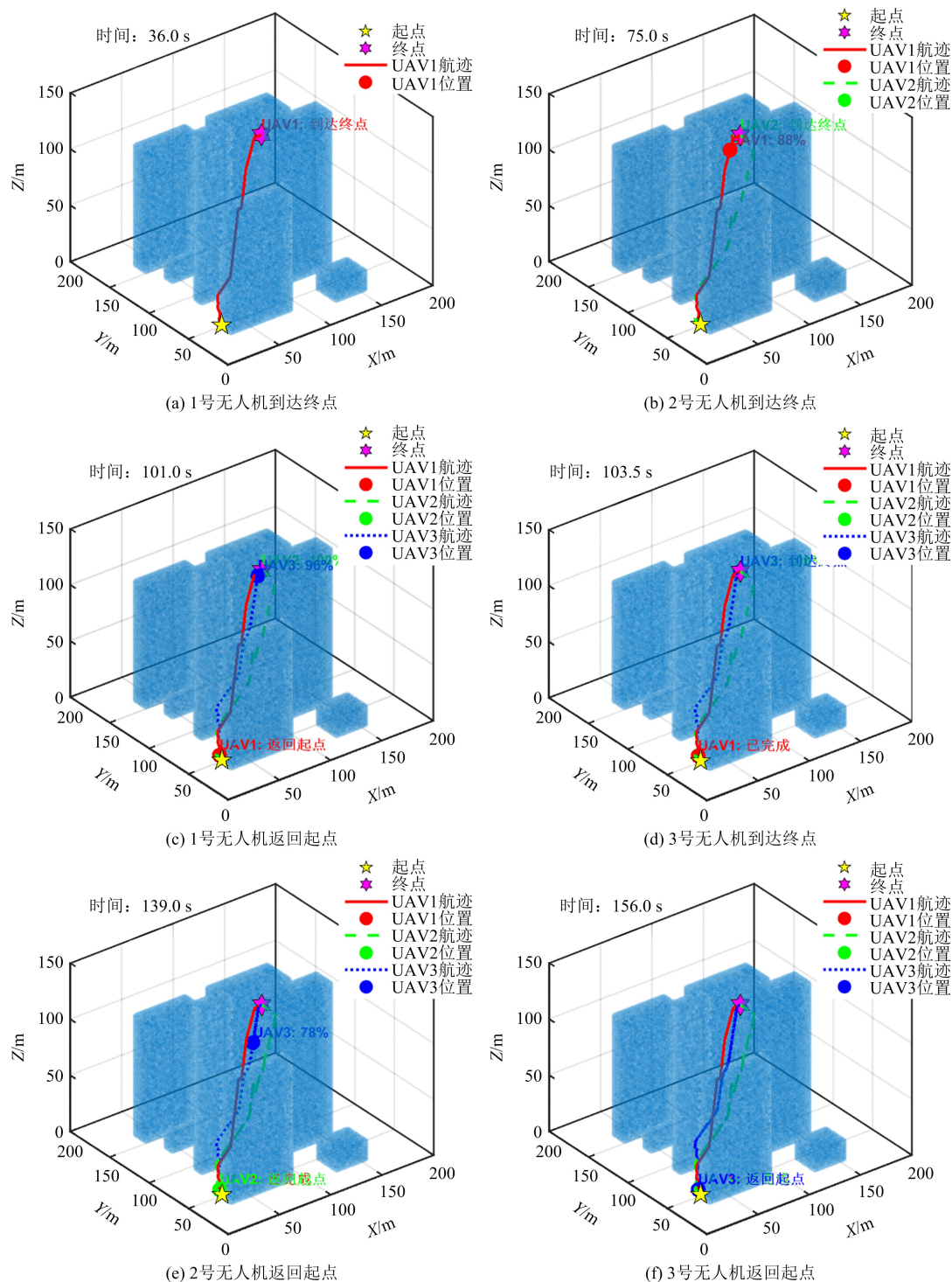


图9 IDQN算法多无人机路径规划结果图

Fig. 9 Path planning result diagram of multi-UAV using the IDQN algorithm

项优势。在路径长度方面,无人机3表现最优,得分接近0.8;在任务时间上,无人机3所用时间最短;无人机1的燃料消耗效率最高;协同效率最佳的是无人机3;在避障能力上无人机3表现最优。3架无人机共同构成了功能完整、专项优势突出的多机协作格局。

2366

4 结论

本文通过建立多无人机单目标时序协同决策模型,并基于任务0优先级采用层次化加权方法融合时间、能量与距离代价,构建了面向城市火灾救援的加权路径优化模型。设置了一个长200 m、宽200

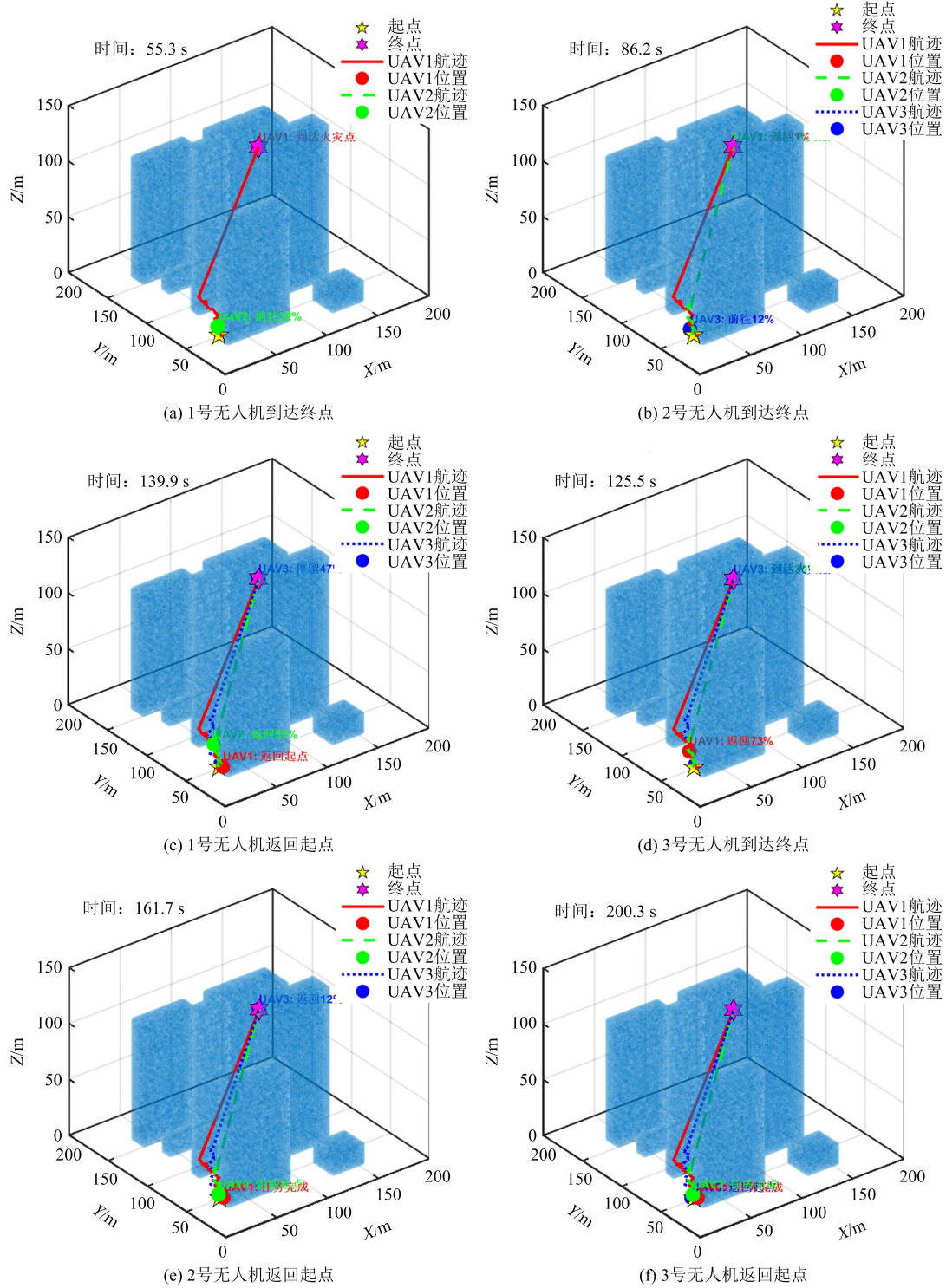


图 10 PSO 算法多无人机路径规划结果图

Fig. 10 Path planning result diagram of multi-UAV using the PSO algorithm

m、高 150 m 的三维区域,使用本文模型与 PSO 算法进行试验,并对多无人机性能雷达图进行分析,得出以下结论。

(1) 与单无人机完成任务耗时 371.4 s 相比,三架无人机基于 IDQN 算法通过相同时间间隔的时序

协同决策模型和加权路径规划模型,将总任务时间缩短至 156 s,耗时缩短 58.0%,且后续无人机航程分别减少 9.0% 和 12.5%,显著提升了时间效率与资源利用率。

(2) 与基于 PSO 算法的协同灭火方案相比,IDQN

表4 IDQN 算法多无人机路径规划结果

Table 4 Path planning results of multi-UAV using the IDQN algorithm

无人机	起飞时间/s	到达时间/s	返航时间/s	返回时间/s	飞行距离/m
1	0.0	36.0	66.0	101.0	443.07
2	40.0	75.0	105.0	139.0	403.15
3	80.0	103.5	133.5	156.0	387.75

表5 PSO 算法多无人机路径规划结果

Table 5 Path planning results of multi-UAV using the PSO algorithm

无人机	起飞时间/s	到达时间/s	返航时间/s	返回时间/s	飞行距离/m
1	0.0	55.3	85.3	139.9	496.10
2	40.0	86.2	116.2	161.7	466.31
3	80.0	125.5	155.5	200.3	448.01

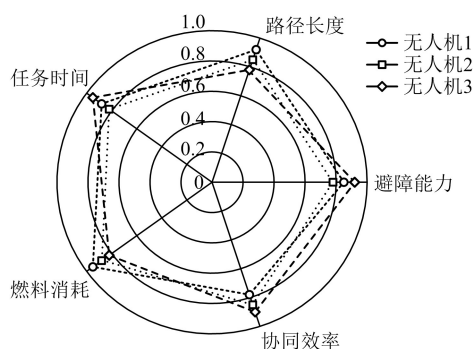


图11 多无人机性能对比雷达图

Fig. 11 Radar diagram of multi-UAV performance comparison

算法在任务执行时间上平均缩短约 27.8%, 总飞行距离减少约 12.5%, 显著提升了多无人机协同作业的时效性与经济性。

(3) 在多无人机性能雷达图评估中, 无人机 3 在路径长度、任务时间、避障能力和协同效率上全面领先, 而无人机 1 燃料消耗效率最高, 3 架无人机共同构成了功能完整、专项优势突出的多机协作格局。

参考文献 (References):

- [1] 刘辉, 程振超, 王丹. 城市消防韧性评价指标体系研究[J]. 灾害学, 2023, 38(2): 25-30.
LIU H, CHENG Z C, WANG D. Study on evaluation index system of urban fire resilience [J]. Journal of Catastrophology, 2023, 38(2): 25-30.

- [2] 李鸿一, 陈锦涛, 任鸿儒, 等. 基于随机采样的高层消防无人机协同搜索规划[J]. 中国科学: 信息科学, 2022, 52(9): 1610-1626.

LI H Y, CHEN J T, REN H R, et al. Random-sampling-based multi-UAV cooperative search planning for high-rise firefighting[J]. Scientia Sinica: Informationis, 2022, 52(9): 1610-1626.

- [3] 张无敌, 陈一洲, 李琪, 等. 城市大型公共建筑火灾风险因素影响程度及可能性分析[J]. 安全与环境学报, 2021, 21(4): 1434-1439.

ZHANG W D, CHEN Y Z, LI Q, et al. Analysis of degree of influence and risk possibility of fire risk factors in large-scale public buildings in cities [J]. Journal of Safety and Environment, 2021, 21(4): 1434-1439.

- [4] BESADA J A, CAMPANA I, BERGESIO L, et al. Drone flight planning for safe urban operations: UTM requirements and tools [C]//2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), March 11-15, 2019, Kyoto, JPN. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2019: 924-930.

- [5] 赵玉花, 石永康, 万晓燕. 基于粒子群优化的多无人机区域覆盖航迹规划[J]. 农机化研究, 2024, 46(6): 63-67.

ZHAO Y H, SHI Y K, WAN X Y. Path planning of multi-UAVs area coverage based on particle swarm optimization [J]. Journal of Agricultural Mechanization Research, 2024, 46(6): 63-67.

- [6] 李章萍, 徐鑫. 改进遗传算法应用于地震场景下无人机路径规划研究[J]. 安全与环境学报, 2025, 25(1): 237–249.
- LI Z P, XU X. Research on drone path planning in earthquake scenarios using an improved genetic algorithm [J]. Journal of Safety and Environment, 2025, 25(1): 237–249.
- [7] 张小孟, 胡永江, 李文广, 等. 基于改进人工蜂群算法的多无人机灭火任务规划[J]. 中国惯性技术学报, 2020, 28(4): 528–536.
- ZHANG X M, HU Y J, LI W G, et al. Multi-UAV fire fighting mission planning based on improved artificial bee colony algorithm [J]. Journal of Chinese Inertial Technology, 2020, 28(4): 528–536.
- [8] 许子良, 胡涛, 刘凯越, 等. 基于多策略改进灰狼算法的多无人机三维路径规划[J]. 航空兵器, 2024, 31(6): 86–93.
- XU Z L, HU T, LIU K Y, et al. Multi-drone 3D path planning based on multi-strategy improved grey wolf algorithm[J]. Aero Weaponry, 2024, 31(6): 86–93.
- [9] 郭志明, 娄文忠, 李涛, 等. 基于改进蝗虫优化算法考虑任务威胁的多无人机协同航迹规划[J]. 兵工学报, 2023, 44(增刊2): 52–60.
- GUO Z M, LOU W Z, LI T, et al. Collaborative route planning of multiple unmanned aerial vehicles considering task threats based on improved grasshopper optimization algorithm[J]. Acta Armamentarii, 2023, 44(Suppl. 2): 52–60.
- [10] 王巧, 宋丹丹, 何姗姗, 等. 高层建筑电梯辅助人员疏散调度策略研究[J]. 安全与环境学报, 2025, 25(6): 2320–2329.
- WANG Q, SONG D D, HE S S, et al. Study of dispatching strategies in the occupants evacuation process of elevator-assisted high-rise buildings [J]. Journal of Safety and Environment, 2025, 25(6): 2320–2329.
- [11] LIU J D, LUO W, ZHANG G Q, et al. Unmanned aerial vehicle path planning in complex dynamic environments based on deep reinforcement learning[J]. Machines, 2025, 13(2): 162.
- [12] 张莉涓, 喻春妮, 高攀, 等. 面向多无人机辅助数据采集的深度强化学习协同航迹规划算法[J]. 中国科学: 信息科学, 2025, 55(2): 426–443.
- ZHANG L J, YU C N, GAO P, et al. Reinforcement learning-based cooperative trajectory planning for UAV enabled data collection [J]. Scientia Sinica: Informationis, 2025, 55(2): 426–443.
- [13] 司鹏搏, 吴兵, 杨睿哲, 等. 基于多智能体深度强化学习的无人机路径规划[J]. 北京工业大学学报, 2023, 49(4): 449–458.
- SI P B, WU B, YANG R Z, et al. UAV path planning based on multi-agent deep reinforcement learning [J]. Journal of Beijing University of Technology, 2023, 49(4): 449–458.
- [14] 吕超, 李慕宸, 欧家骏. 基于分层深度强化学习的无人机混合路径规划[J/OL]. 北京航空航天大学学报: 1–13 [2025–07–09]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0550>.
- LÜ C, LI M C, OU J J. UAV hybrid path planning based on hierarchical deep reinforcement learning [J/OL]. Journal of Beijing University of Aeronautics and Astronautics: 1–13 [2025–07–09]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0550>.
- [15] 江泰民, 谭泰, 李辉, 等. 基于分层深度强化学习的六自由度固定翼无人机路径跟踪方法[J/OL]. 计算机工程: 1–14 [2025–07–09]. <https://doi.org/10.19678/j.issn.1000-3428.0070197>.
- JIANG T M, TAN T, LI H, et al. Path following of 6–DOF fixed-wing UAV based on hierarchical deep reinforcement learning [J/OL]. Computer Engineering: 1–14 [2025–07–09]. <https://doi.org/10.19678/j.issn.1000-3428.0070197>.
- [16] 宋睿, 朱培, 张剑高, 等. 面向高原森林火灾的灭火无人机路径规划[J]. 科学技术与工程, 2024, 24(34): 14863–14870.
- SONG R, ZHU P, ZHANG J G, et al. Path planning of fire fighting UAV for plateau forest fires [J]. Science Technology and Engineering, 2024, 24(34): 14863–14870.
- [17] 王慧颖, 张志强, 杨传杰, 等. 改进 ABC 算法无人机群多目标灭火救援任务规划[J]. 消防科学与技术, 2023, 42(6): 838–843.
- WANG H Y, ZHANG Z Q, YANG C J, et al. Multi-target fire rescue mission planning of UAV swarm based on improved ABC algorithm [J]. Fire Science and Technology, 2023, 42(6): 838–843.
- [18] PEREZ-SAURA D, FERNANDEZ-CORTIZAS M, PEREZ-SEGUI R, et al. Urban firefighting drones: precise throwing from UAV [J]. Journal of Intelligent & Robotic Systems, 2023, 108(4): 66.

Cooperative firefighting using multi-UAVs in urban settings based on an improved DQN algorithm

QIN Rui, ZHANG Xiaoling

(College of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: To enhance the efficiency of urban fire rescue operations, this paper presents a multi-UAV cooperative firefighting path planning algorithm based on an Improved Deep Q -Network (IDQN). First, we establish a sequential collaborative decision-making model with equal time intervals for multi-UAV single-target operations. This model ensures that multiple UAVs arrive at the fire scene in an orderly and efficient manner to execute fire extinguishing tasks, thereby minimizing task conflicts and resource waste. Second, the algorithm incorporates a prioritized experience replay mechanism to improve training stability. Additionally, a dynamic state representation method is developed to navigate dense urban obstacle environments effectively. A 26-direction three-dimensional action space is implemented to enhance the UAVs' maneuverability and obstacle avoidance capabilities in three-dimensional environments, thereby improving their perception of complex spatial features. A multi-objective composite reward function—comprising directional guidance, step penalties, arrival rewards, and collision penalties—is constructed, allowing the UAVs to autonomously balance path length and flight safety during operations, significantly boosting mission reliability. Finally, given the priority requirements of urban firefighting tasks, a hierarchical weighting method is employed to assign differentiated weights to time cost, energy consumption cost, and flight distance cost, leading to the development of a multi-objective weighted path optimization model. Simulation training conducted on the MATLAB platform facilitates efficient cooperative path planning for multiple UAVs. The simulation results indicate that when three UAVs take off at 40-second intervals to perform firefighting tasks, the total mission time is maintained within 156 seconds—a substantial reduction of 58.0% compared to the 371.4 seconds required by a single UAV. Additionally, when compared to the Particle Swarm Optimization (PSO) algorithm, the proposed IDQN algorithm achieves a further reduction in average task execution time by 27.8% and a decrease in total flight distance by 12.5%. These findings demonstrate significant improvements in both time efficiency and path economy, thereby enhancing the overall effectiveness and practicality of the multi-UAV cooperative system. In the evaluation of multi-UAV performance radar chart, UAV 3 is leading in path length, mission time, obstacle avoidance ability and collaborative efficiency. Although UAV 1 consumes the most fuel, the three UAVs together constitute a multi-aircraft cooperation pattern with complete functions and outstanding special advantages.

Key words: safety engineering; deep reinforcement learning algorithm; multi-UAV cooperation; urban fires; collaborative firefighting; path planning