自动化学报
ACTA AUTOMATICA SINICA

# 基于多模态特征融合与局部感知推理的多智能体分布式路径协作方法

霍琳 毛剑琳 伞红军 付丽霞 宣志玮 贺志刚

# A Distributed Multi-Agent Path Coordination Method Based on Multimodal Feature Fusion and Local Perception Reasoning

HUO Lin, MAO Jian-Lin, SAN Hong-Jun, FU Li-Xia, XUAN Zhi-Wei, HE Zhi-Gang

在线阅读 View online: https://doi.org/10.16383/j.aas.c250593

---

# 基于多模态特征融合与局部感知推理的多智能体分布式路径协作方法

霍 琳 [1] 毛剑琳 [1, 2] 伞红军 [1] 付丽霞 [2] 宣志玮 [1] 贺志刚 [1]

**摘 要** 在部分可观测且动态变化的环境中, 深度强化学习 (DRL) 为多智能体路径规划提供具备自学习、泛化与动态适应能力的分布式求解途径. 然而, DRL 在该问题中仍存在协调性不足与局部近视两个挑战: 智能体间的隐式交互易引发冲突, 仅依赖局部观测的反应式避障又导致路径冗长与全局目标偏离. 为此, 提出一种基于多模态特征融合与局部感知推理的分布式路径协作方法——LYRA. 该方法在分布式 DRL 框架下构建从感知到决策的学习体系, 使智能体能依托局部观测实现推理与隐式让行. 多模态状态编码器融合局部碰撞线索与全局路径语义, 平衡即时避障与长期导航目标; 奖励引导的路径学习策略保证局部决策与全局任务一致; 群体协作机制通过隐式优先级推理化解局部冲突; 自适应学习率机制动态调节策略更新以提升训练稳定性. 实验结果表明, LYRA 在任务成功率和安全性上较基线方法有大幅提升, 在不同环境复杂度下保持良好泛化性, 为实现高效鲁棒的分布式多智能体路径规划提供了新范式.

**关键词** 多智能体路径规划; 深度强化学习; 局部观测; 多模态; 推理; 协作

**引用格式** 霍琳, 毛剑琳, 伞红军, 付丽霞, 宣志玮, 贺志刚. 基于多模态特征融合与局部感知推理的多智能体分布式路径协作方法. 自动化学报, 2026, **52**(7): 1347−1359

**DOI** 10.16383/j.aas.c250593 **CSTR** 32138.14.j.aas.c250593

# A Distributed Multi-Agent Path Coordination Method Based on Multimodal Feature Fusion and Local Perception Reasoning

HUO Lin[1] MAO Jian-Lin[1, 2] SAN Hong-Jun[1] FU Li-Xia[2] XUAN Zhi-Wei[1] HE Zhi-Gang[1]

**Abstract** In partially observable and dynamically changing environments, deep reinforcement learning (DRL) offers a distributed solution for multi-agent path finding (MAPF) with self-learning capability, generalization, and dynamic adaptability. Nevertheless, DRL-based MAPF still suffers from insufficient coordination and local myopia: Implicit agent interactions easily cause conflicts, while reactive collision avoidance relying solely on local observations often leads to elongated paths and deviations from global objectives. To address these challenges, we propose local yielding and reasoning for agents (LYRA), a distributed path coordination method based on multi-modal feature fusion and local perceptual reasoning. LYRA constructs an end-to-end learning framework from perception to decision-making under distributed DRL, enabling agents to perform reasoning and implicit yielding using only local observations. A multi-modal state encoder integrates local collision cues with global path semantics to balance immediate avoidance and long-term navigation goals. A reward-guided path learning strategy aligns local decisions with global tasks, while a collective coordination mechanism resolves local conflicts through implicit priority reasoning. In addition, an adaptive learning rate mechanism dynamically adjusts strategy updates to improve training stability. Experiments show that LYRA improves task success rate and safety by approximately 35 % and 25 %, respectively, over baselines, while maintaining strong generalization across environments of varying complexity. These results suggest a new paradigm for efficient and robust distributed multi-agent path planning.

**Keywords** multi-agent path finding; deep reinforcement learning; partial observability; multi-modal; inference; coordination

**Citation** Huo Lin, Mao Jian-Lin, San Hong-Jun, Fu Li-Xia, Xuan Zhi-Wei, He Zhi-Gang. A distributed multi-agent path coordination method based on multimodal feature fusion and local perception reasoning. *Acta Automatica Sinica*, 2026, **52**(7): 1347−1359

收稿日期 2025-11-03 录用日期 2026-01-30

Manuscript received November 3, 2025; accepted January 30, 2026

国家自然科学基金 (62263017), 云南省科技人才与平台计划 (202505AT350001), 云南省基础研究计划 (202301AU070059) 资助

Supported by National Natural Science Foundation of China (62263017), Science and Technology Talent and Platform Program of Yunnan Province (202505AT350001), and Basic Research Program Project of Yunnan Province (202301AU070059)

本文责任编委 罗彪

Recommended by Associate Editor LUO Biao

1. 昆明理工大学机电工程学院 昆明 650500 2. 昆明理工大学信息工程与自动化学院 昆明 650500

1. Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500 2. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500

多智能体路径规划 (multi-agent path finding, MAPF) 的目标是为一组智能体规划无碰撞的轨迹, 使其从起始位置到达各自的目标点[1]. 在自动化仓储[2]、地空协同交通管理[3] 等实际场景中, MAPF 系统需要具备实时规划能力, 能够在大规模智能体数量下保持高效运行, 并在部分可观测和动态环境中保持稳定性与可靠性.

传统的多智能体路径规划算法, 如 HCA* (hierarchical cooperative A*)[1], CBS (conflict-based search)[4] 以及 PBS (priority-based search)[5] 等算法, 在理论上可保证路径规划结果的完整性与最优性[6]. 然而, 当智能体密度或障碍物复杂度增加时, 这类方法的计算代价呈指数级增长[7]. 更重要的是, 它们通常依赖全局可观测和集中控制的假设, 这在大规模、去中心化或环境信息不完全的场景下往往难以满足[8].

由于深度强化学习 (deep reinforcement learning, DRL) 具备较好的自学习能力、泛化性以及动态适应性, 研究者开始尝试利用其解决 MAPF 问题. 目前主要形成两类代表性范式: 专家引导学习[9] 和自学习反应式学习[10]. 前者通过模仿集中式求解器生成的专家轨迹来加速早期学习阶段, 但其性能高度依赖示例数据的质量与覆盖范围[11−12]. 相比之下, 后者不依赖外部监督, 而是直接基于局部观测进行学习[13], 并常借助耦合机制或注意力机制[14−15] 来推断邻近智能体的意图, 从而在无显式通信的情况下实现协作[16]. 尽管该范式能够提高智能体的自主性[17], 但在实际应用中仍存在两方面的突出问题.

一方面, 多智能体在动态和部分可观测环境下协调性脆弱[18−19]. 许多方法假设邻居集合固定[20]、可视图静态[21] 或局部规则恒定[22], 这些假设在动态环境下容易失效. 同时, 在局部观测视野的限制下, 智能体缺乏显式通信, 隐式交互易引发冲突和死锁[23]. 另一方面, “局部近视”问题突出. 反应式避障忽视全局语义[24], 常导致路径偏离和效率下降[25].

针对上述两方面的问题, 本文提出一种去中心化 DRL 框架——LYRA (local yielding and reasoning for agents)[1]. LYRA 的核心思想是将局部感知与长时规划线索相结合, 使不同尺度的决策过程自然对齐, 进而在无显式通信的条件下实现稳健的协作. 与现有将局部反应和全局目标分离优化的方法不同, LYRA 将感知、推理与协同相结合, 使短期安全性与长期效率在策略优化过程中同步提升.

本文的主要贡献如下:

1) 为缓解局部近视问题, 设计一种融合局部碰撞线索与全局路径语义的多模态状态编码器, 以弥补观测不完备导致的信息缺失, 促使策略在即时安全与长期目标间做出平衡性推理;

2) 针对多智能体协调性脆弱的问题, 提出群体协作机制, 通过局部可观测标识和动态优先级策略, 使智能体在狭窄通道或高密度区域中能够自发地分配通行顺序, 能有效避免冲突与死锁;

3) 针对局部近视导致智能体绕行的问题, 设计基于奖励的全局引导机制, 将轨迹级先验融入稀疏–稠密奖励设计, 约束局部避障选择向全局目标对齐, 减少路径碎片化与冗长绕行;

4) 为保证大规模训练的稳定性与收敛性, 进一步引入自适应学习率机制, 结合指数衰减与周期调制, 动态调节策略与价值网络的更新幅度, 提升大规模去中心化训练的稳定性与收敛效率.

实验结果表明, LYRA 在多种复杂场景下的表现均优于现有 DRL-MAPF 方法, 成功率提升超过 35%, 碰撞率降低超过 25%, 并在受限空间中表现出更高的通行效率.

本文的结构安排如下: 第 1 节回顾相关研究; 第 2 节介绍问题建模; 第 3 节阐述系统结构; 第 4 节给出实验结果; 第 5 节总结全文并讨论未来工作.

## 1 相关工作

近年来, 基于 DRL 的去中心化 MAPF 研究取得显著进展. 过去十年中, PRIMAL (pathfinding via reinforcement and imitation multi-agent learning) 方法[26] 首次提出将集中式求解器生成的专家轨迹与在线强化学习相结合的混合策略. 这一设计显著加快了早期收敛速度, 并降低了初始阶段的碰撞率. 然而, 该方法对高质量专家示例的依赖较强, 当示例稀疏或目标位置动态变化时, 性能会明显下降[27]. PRIMAL2[28] 在此基础上引入动态目标重分配机制, 使智能体能够持续在线更新, 从而提高对任务分布变化的适应能力. 但其实时重分配的计算开销较高, 并且仍依赖于示例先验, 限制了在大规模非结构化环境中的扩展性[29].

为降低对专家数据的依赖, 后续研究开始探索基于成对耦合的机制. DHC (distributed, heuristic, communication) 方法[30] 在图神经网络结构上引入启发式通道筛选, 通过剪除远距离或弱相关的邻居节点来提高在中等密度场景下的计算效率. 然而, 该人工设计的筛选策略在拥挤环境中易过度剪枝, 在稀疏环境中又可能筛选不足, 从而导致协调误差[31]. DCC (decision causal communication) 方法[32] 用可学习的门控函数取代静态启发式规则, 使智能体能够根据场景动态调整邻居影响权重. 该机制在不同密度场景中表现出更强的泛化性, 但双门控结构在极端密度变化下仍可能误判关键邻居, 限制了系统的鲁棒性[33].

[1] 项目演示视频见: https://github.com/HuoLinL/LYRA.

在此基础上, MAPPER (multi-agent path planning with evolutionary reinforcement learning) 方法[34] 进一步提出在训练阶段为每对智能体学习亲和函数, 以引导联合动作趋向更安全的配置, 从而提高协作质量. 然而, 成对亲和计算的平方级复杂度限制了其在大规模智能体场景中的可扩展性. SCRIMP (scalable communication for reinforcement- and imitation-learning-based multi-agent pathfinding) 方法[16] 通过预测邻居的下一步动作并将其作为输入, 避免显式的成对计算, 实现了基于预测的协同规划. 然而, 在非平稳人群中, 预测误差容易累积, 导致在动态环境下的可靠性下降[35].

尽管成对耦合机制在处理局部交互方面表现出色, 但它们通常缺乏对长距离依赖和长时目标的建模能力[36]. 因此, 研究者开始将 Transformer 结构引入 MAPF. TIRL (Transformer-based imitative reinforcement learning) 方法[37] 采用视觉 Transformer 骨干网络, 通过多头自注意力机制同时聚合邻近智能体、远处障碍物与队友信息, 从而在密集及宽域场景中显著提升了协调能力. 然而, 该方法的计算与存储开销较高[38], 并且在稀疏奖励条件下训练稳定性较差, 导致大规模训练收敛缓慢[39].

总体来看, 现有方法虽然在性能与泛化方面取得进展[40], 但仍存在两个关键问题亟待解决: 1) 大多数方法在观测质量变化或局部密度波动时适应性较差[41−42], 在动态群体中易出现协调不稳与行为漂移[43]; 2) 多数方法难以将局部避障行为与长期任务目标有机结合[44], 导致在长时间或复杂任务中的路径碎片化与效率下降.

为此, 本文提出 LYRA 框架, 该框架融合了多模态感知、自适应邻居推理与轨迹引导决策. 这种设计使智能体能够在响应局部交互的同时保持全局任务一致性, 从而在无通信约束下实现可扩展的协调控制, 适用于大规模动态环境.

## 2　问题描述

从智能体个体自主决策的角度出发, 在无通信依赖的条件下, 每个智能体必须独立完成环境感知、冲突处理与路径选择. 然而, 由于个体在竞争时空资源时往往表现出贪婪性决策特征, 系统整体容易陷入协作困境, 表现为协调成功率低、避障安全性差以及路径规划效率下降. 在局部感知受限的条件下, 多智能体分布式路径协作的决策过程高度复杂 (如图 1 所示), 不完整的环境观测、局部路径的高度耦合以及视野外动态变化共同增加了决策不确定性. 这些问题使智能体在感知、推理与学习层面都面临显著困难, 也成为实现高效、安全分布式路径规划的核心瓶颈.

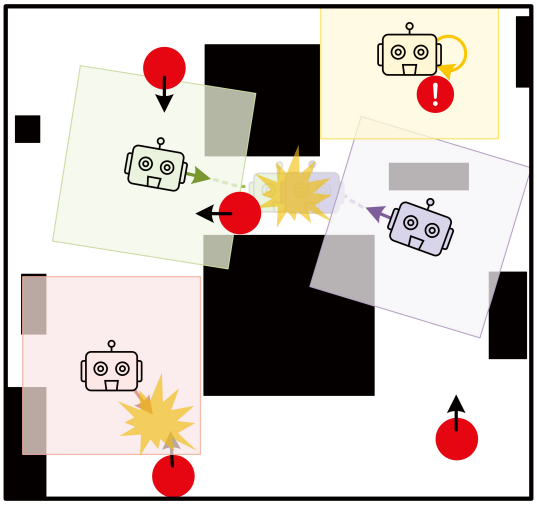


图 1　问题描述

Fig. 1　Problem description

为解决上述问题, 本文将该任务建模为部分可观测动态时空博弈 (partially observable dynamic spatio-temporal game, PODSTG). 在该博弈框架中, 智能体需基于有限的局部观测与全局路径先验, 推断出协同策略, 从而在动态且部分可观测的环境中完成导航任务.

本文将问题建模分解为两个关键部分: 时空交互建模与观测空间设计. 二者共同支撑在不确定条件下的可扩展、去中心化策略学习.

### 2.1　基于 PODSTG 的时空交互建模

在 PODSTG 框架下, 每个智能体 $i \in \boldsymbol{N}$ 在动态且部分可观测的环境中进行导航, 其决策仅依赖于局部观测信息, 但同时隐式地考虑了静态障碍、动态实体以及其他智能体的影响, $\boldsymbol{N} = \{1, 2, \cdots, N\}$ 为智能体集合. 这些因素可统一建模为时空势场:

$$U(x,\ y,\ t) = U_{\text{static}}(x,\ y) + \sum_{dy \in \mathcal{G}} U^{dy}_{\text{dynamic}}(x,\ y,\ t) + \sum_{j \neq i} U^{j}_{\text{agent}}(x,\ y,\ t) \tag{1}$$

其中, $x$ 与 $y$ 表示二维离散栅格地图中的空间位置坐标, $t$ 表示当前决策时刻的时间步. $U_{\text{static}}$ 表示静态占据网格 $\mathcal{G} \in \{-1,\ 1\}^{H \times W}$, 其中, 1 代表障碍物, $-1$ 代表可通行区域; $U^{dy}_{\text{dynamic}}$ 表示由动态障碍物引起的势场, 其位置 $\boldsymbol{p}^{dy}_t$ 按离散时间半马尔可夫过程 (semi-Markov process) 演化[45], 以保证系统的下一状态仅依赖当前状态, 即

$$\boldsymbol{p}^{dy}_{t+1} = \boldsymbol{p}^{dy}_t + \boldsymbol{\Delta p}^{dy},\ \boldsymbol{\Delta p}^{dy} \sim \mathcal{U}\{\uparrow,\ \downarrow,\ \leftarrow,\ \rightarrow,\ \circ\} \tag{2}$$

式中, $\boldsymbol{\Delta p}^{dy}$ 表示 $\boldsymbol{p}^{dy}$ 的变化量; $\mathcal{U}\{\uparrow,\ \downarrow,\ \leftarrow,\ \rightarrow,\ \circ\}$ 表示离散时间半马尔可夫的演化过程, 其中参数 $\uparrow$、

↓、←、→、∘ 分别代表移动方式为上、下、左、右、停止. 同时, $U^{j}_{\text{agent}}$ 则刻画智能体间的相互作用, 其运动状态 $\boldsymbol{s}_{t,\,i}=(x_{t,\,i},\ y_{t,\,i},\ \boldsymbol{v}_{t,\,i})$ 满足以下约束:

$$\begin{cases} x_{t+1,\,i}=x_{t,\,i}+v_{t,\,x}\cdot\Delta t \\ y_{t+1,\,i}=y_{t,\,i}+v_{t,\,y}\cdot\Delta t \\ \|\boldsymbol{v}_{t,\,i}\|\le\|\boldsymbol{v}_{\max}\|=1.0 \end{cases} \tag{3}$$

复合势场的梯度 $\nabla U(x,\ y,\ t)$ 可直接用于引导局部动作选择, 既能有效避免碰撞, 又能在拥挤或动态区域实现平滑的路径调整.

### 2.2 基于拓扑编码的局部–全局感知解耦

为补偿纯局部感知的局限性, 每个智能体被赋予由全局路径规划算法生成的高层轨迹先验. 全局路径 $Path_i=[path_{1,\,i},\ \cdots,\ path_{Q,\,i}]$ 由 A*算法[46] 在仅考虑静态障碍的条件下生成, 并编码为拓扑热力图 $HF_{\text{path},\,i}\in\{-1,\ 0,\ \cdots,\ Q\}^{H\times W}$, 其定义如下:

$$HF_{\text{path},\,i}(x,\ y)=\begin{cases} q, & (x,\ y)=path_{q,\,i},\quad 1\le q\le Q \\ -1, & \text{其他} \end{cases} \tag{4}$$

该编码提供对长期导航意图的空间表征, 使局部动作选择能够与全局路径引导保持一致, 同时在动态变化环境中仍具备自适应调整能力.

### 2.3 部分可观测条件下的观测空间设计

在每个时间步 $t$, 智能体 $i$ 的观测应能够以紧凑的形式同时反映空间布局与运动语义 (如图 2 所示).

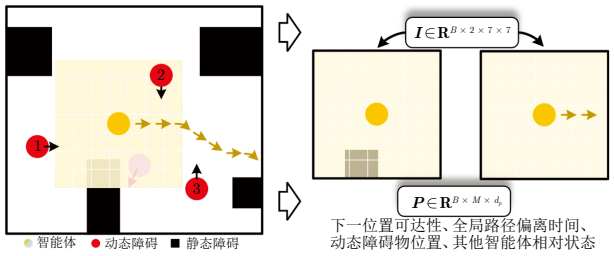



图 2 多模态观测

Fig. 2 Multimodal observation

具体而言, 智能体 $i$ 在时刻 $t$ 的观测可表示为

$$\boldsymbol{o}_{t,\,i}=[\boldsymbol{I}_{t,\,i},\ \boldsymbol{P}_{t,\,i}] \tag{5}$$

其中, $\boldsymbol{I}_{t,\,i}\in\mathbf{R}^{B\times 2\times 7\times 7}$ 为双通道图像张量, $B$ 表示批大小 (batch size). 第一通道表示以智能体为中心、视野范围为 $7\times 7$ 的局部静态障碍分布; 第二通道编码由 A* 算法生成的全局路径方向信息, 将长期导航意图嵌入空间表示中. 向量 $\boldsymbol{P}_{t,\,i}\in\mathbf{R}^{B\times M\times d_p}$ 用于描述非视觉特征, 其中 $M$ 表示在当前观测窗口内的最大邻居数量, $d_p$ 表示动态障碍物位置以及其他智能体的相对状态. 该向量记录了异质的语义与时间线索, 包括可通行单元的二值可达性标志、智能体偏离引导路径后的累计时间 $t_{i,\,\text{off}}$, 以及周围动态障碍与邻居智能体的相对状态信息.

对于特定邻居 $j\in\boldsymbol{N}$, 其相对状态定义为

$$f^{(j)}_{\text{nei}}=[\boldsymbol{\Delta p}_{ij},\ ID_{j,\,\text{mod}},\ t_{j,\,\text{off}}] \tag{6}$$

其中, $\boldsymbol{\Delta p}_{ij}$ 表示智能体 $i$ 到邻居 $j$ 在局部坐标系下的相对位置向量. 该向量既作为交互推理的空间特征输入, 也用于第 3.2 节中的避碰检测, 其模长可直接用于判断是否违反了距离约束. $ID_{j,\,\text{mod}}$ 表示在每个训练回合开始时为智能体随机分配的索引. 该索引在取模约束后用于限制嵌入空间范围, 避免模型过度拟合到固定身份信息. 在第 3.2 节中, 该索引还被群体协作 (group collaborative consensus, GCC) 机制用于优先级比较. 对于在运行过程中新进入观测视野、但未参与初始索引分配的智能体, 则基于局部可观测量实时计算临时优先级得分, 具体方法将在第 3.2 节给出. 变量 $t_{j,\,\text{off}}$ 表示邻居偏离其全局路径的时间长度.

组合特征 $[\boldsymbol{I}_{t,\,i},\ \boldsymbol{P}_{t,\,i}]$ 同时保留局部避碰所需的空间上下文与长期协作推理所需的语义结构信息. 该特征将在第 3.1 节作为多模态状态编码器的直接输入, 用于统一感知与决策过程.

### 2.4 任务目标

在去中心化 MAPF 任务中, 优化目标不仅是获得高累积奖励, 还需确保所有智能体能够安全且成功地到达各自目标点, 并最大限度减少碰撞. 该目标可表述为在强化学习框架下最大化任务成功概率, 并在任务完成的条件下进一步最大化期望折扣回报[47], 即

$$\max_{\pi_\theta}\ \mathrm{E}_{\tau\sim\pi_\theta}\left[\frac{1}{N}\sum_{t=0}^{T}\sum_{i=0}^{N-1}\gamma_t R_{t,\,i}\ \middle|\ \max_{\pi_\theta}\ P(SR)\right] \tag{7}$$

其中, $\pi_\theta$ 为参数化策略, $\tau$ 表示状态与动作序列, $\gamma\in(0,\ 1]$ 为折扣因子, $R_{t,\,i}$ 为智能体 $i$ 在时刻 $t$ 的即时奖励, $P(SR)$ 表示所有智能体在给定时间范围内成功到达目标的概率. 该优化形式明确体现了任务完成的优先性: 任务成功是首要目标, 而奖励最大化则在任务完成的条件下进行次级优化.

## 3 基于多模态特征融合与局部感知推理的多智能体分布式路径协作方法

LYRA 采用分布式 DRL-MAPF 框架, 其核心思想是从状态感知到学习流程进行协同设计, 使智能体在部分可观测条件下无需显式通信, 而是通过

对局部环境变化与同伴行为的持续感知, 在策略执行过程中自发形成让行与顺序通行等协作行为, 从而在大规模动态环境中保持稳定的协调性能. 框架的总体结构如图 3 所示.

LYRA 将局部观测编码为结构化的空间–语义状态表示, 统一刻画静态障碍、邻近智能体动态及轨迹层级的引导信息, 并将其输入至局部可观测的优先级机制, 以在受限区域内有序化解瓶颈与死锁, 实现局部避碰与全局路径一致性的协调. 此外, 通过将全局路径先验嵌入奖励信号, 该框架在去中心化条件下仍能维持一致的交互规则与稳定的策略收敛, 从而为多智能体的可扩展协作提供支撑.

### 3.1　多模态状态编码器

为预测并适应多智能体交互模式的动态变化, LYRA 设计了一个多模态状态编码器 (multi-modal state encoder, MMSE), 用于融合局部空间布局与邻近智能体及障碍物的运动趋势, 使智能体能够在仅依赖局部观测的条件下, 对潜在交互关系与冲突态势进行持续判断, 并据此调整当前决策. 该编码器能够在部分可观测条件下为智能体提供结构化的情境表征, 从而在保持局部避碰安全性的同时兼顾长期导航一致性.

MMSE 由两个关键组件组成: 多尺度特征融合模块 (multi-scale feature fusion, MSFF) 和动态协作注意力模块 (dynamic collaborative attention, DCA). 前者用于联合建模短期避碰信息与全局路径线索, 后者通过自适应注意力机制聚合邻居的关键行为特征, 为后续的协作决策提供支持.

#### 3.1.1　多尺度特征融合模块

有效的 MAPF 感知需要在局部避碰与全局路径感知之间取得平衡. 过度强调局部风险容易导致冗长绕行, 而过度依赖目标方向则可能引发冒险行为或碰撞. MSFF 通过多尺度感知结构对这两类信息进行联合建模, 在统一的嵌入空间内同时编码局部安全性与全局一致性.

在每个时间步 $t$, 智能体 $i$ 的观测 $[\boldsymbol{I}_{t,i},\ \boldsymbol{P}_{t,i}]$ 的定义与第 2.3 节一致. 如图 4 所示, 图像张量 $\boldsymbol{I}_{t,i}$ 被分为两个尺度路径处理: 高分辨率网络保持原始 $7\times 7$ 尺度, 输出 $\boldsymbol{H}_{\mathrm{HR}}\in\mathbf{R}^{B\times 2\times 7\times 7}$, 用于检测狭窄通道或拥挤交叉口中的碰撞风险; 低分辨率路径将输入下采样至 $\boldsymbol{H}_{\mathrm{LR}}\in\mathbf{R}^{B\times 2\times 2\times 2}$, 提取粗粒度的导航方向信息, 帮助智能体在避障过程中跟随全局指引路径. 高分辨率网络卷积操作 $\mathrm{Conv}_{k3,s1,p1}$ 中, 卷积核 $k$ 取值为 3, 步长 $s$ 取值为 1, 填充 $p$ 取值为 1; 低分辨率网络的卷积操作 $\mathrm{Conv}_{k3,s2,p1}$ 中,卷积核 $k$ 取值为 3, 步长 $s$ 取值为 2, 填充 $p$ 取值为 1.

在高分辨率分支中, 通过中心聚焦操作提取以智能体为中心的 $3\times 3$ 区域 $\boldsymbol{H}_{\mathrm{HC}}\in\mathbf{R}^{B\times 2\times 3\times 3}$, 用于捕获即时行动最关键的局部区域. 随后将 $\boldsymbol{H}_{\mathrm{HC}}$ 展平并输入多层感知机 ($\mathrm{MLP}_1$, 线性层 + ReLU), 生成具备碰撞敏感性的特征嵌入. 与此同时, 非视觉特征向量 $\boldsymbol{P}_{t,i}$ 通过另一条 $\mathrm{MLP}_1$ 分支处理, 以提取邻近动态元素的语义运动线索. 两者输出拼接后经由 $\mathrm{MLP}_2$ (线性层 + ReLU + 线性层) 融合, 得到多尺度观测嵌入 $\boldsymbol{f}_{\mathrm{self}}\in\mathbf{R}^{B\times 128}$.

该结构避免将局部避碰信号与全局导航信息割裂建模, 使即时反应始终受到长期路径约束, 从而缓解纯反应式策略带来的路径偏离问题.

#### 3.1.2　动态协作注意力模块

尽管 MSFF 为单个智能体提供了丰富的局部表征, 但在多智能体场景中, 有效决策仍需结合邻居的运动趋势进行推理. 在狭窄通道或密集区域中, 若未能充分建模邻居的行为, 常会出现振荡、无效等待或偏离全局路径的绕行. 图 5 所示的动态协作

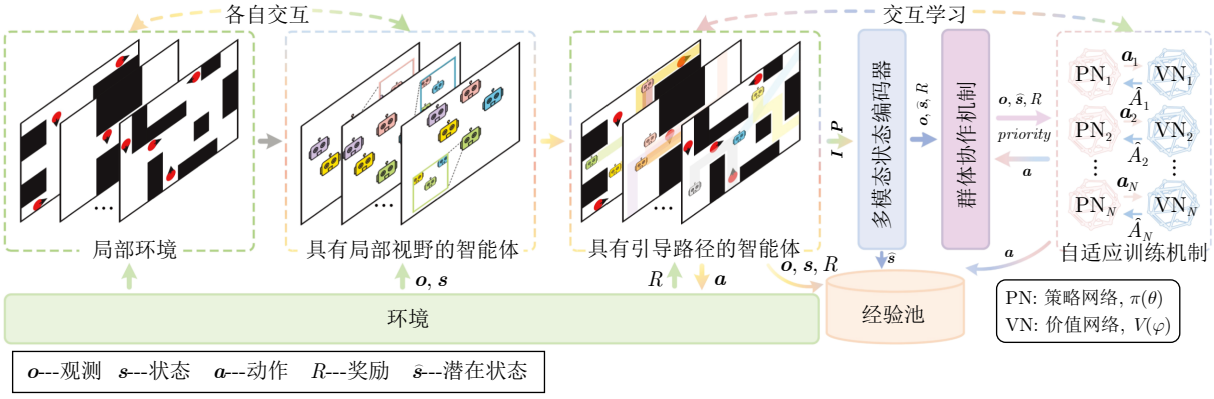



图 3　LYRA 框架总体结构示意图

Fig. 3　Schematic diagram of the overall framework of LYRA

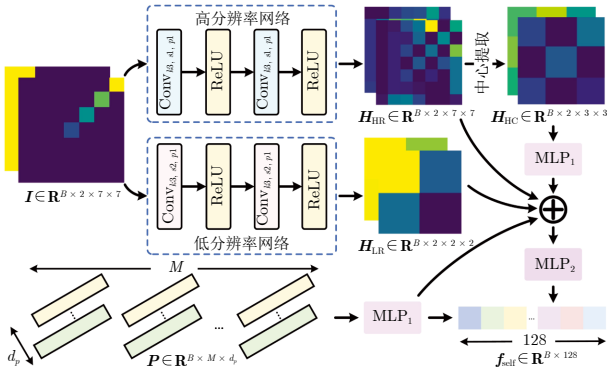



图 4 MSFF 模块的结构示意图

Fig. 4 Schematic diagram of the MSFF module structure

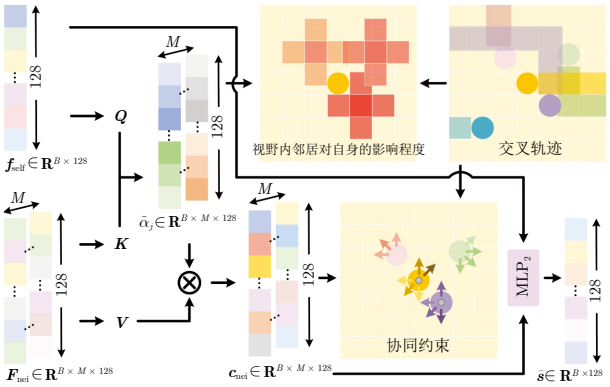



图 5 DCA 模块的结构示意图

Fig. 5 Schematic diagram of the DCA module structure

注意力模块 (DCA) 通过聚合邻居信息, 增强在不同密度条件下的交互建模能力.

在时间步 $t$, 智能体 $i$ 从 $7\times7$ 感知窗口中识别可观测邻居集合 $\mathcal{M}_i$, 并对每个邻居的观测输入相同的 MSFF 结构, 得到邻居特征张量 $\boldsymbol{F}_{\text{nei}}\in\mathbf{R}^{B\times M\times128}$. 当可见邻居数少于 $M$ 时, 采用二值掩码 $\boldsymbol{Mask}\in\{0,\,1\}^M$ 标记无效槽位, 确保聚合时不存在虚假邻居的干扰.

智能体自身的 MSFF 输出 $\boldsymbol{f}_{\text{self}}\in\mathbf{R}^{B\times128}$ 用于查询向量 $\boldsymbol{Q}$, 而每个邻居特征经线性映射生成键 $\boldsymbol{K}$ 与值 $\boldsymbol{V}$, 二者均为 $\mathbf{R}^{B\times128}$. 对于邻居 $j$, 注意力权重计算如下:

$$\alpha_j=\frac{\boldsymbol{Q}\boldsymbol{K}_j^{\mathrm{T}}}{\sqrt{128}} \tag{8}$$

$$\tilde{\alpha}_j=\frac{\exp(\alpha_j)\boldsymbol{Mask}_j}{\sum\limits_{m=1}^{M}\exp(\alpha_m)\boldsymbol{Mask}_m+\epsilon} \tag{9}$$

其中, $\epsilon=10^{-8}$ 用于防止在无邻居情况下出现除零错误. 该机制使与当前路径交叉或运动方向相近的邻居获得更高权重, 从而重点建模潜在的瓶颈与冲突.

最终, 对值向量加权求和得到上下文特征 $\boldsymbol{c}_{\text{nei}}\in\mathbf{R}^{B\times M\times128}$, 再与 $\boldsymbol{f}_{\text{self}}$ 拼接, 经 $\text{MLP}_2$ 映射后生成最终的潜在状态嵌入 $\hat{\boldsymbol{s}}\in\mathbf{R}^{B\times128}$. 该嵌入不仅保留 MSFF 的多尺度空间特征, 还融合了动态拥堵与协作意图信号, 从而在决策阶段显式地偏向冲突化解与路径一致性. DCA 使邻居推理从被动的轨迹感知转变为主动的、路径感知的协作过程, 是 LYRA 实现可扩展分布式 DRL-MAPF 的关键模块.

### 3.2 群体协作机制

当多个智能体在狭窄通道或交叉区域汇聚时, 仅依靠被动避让往往不足以避免拥堵. 若缺乏优先级机制, 局部振荡容易扩散为全局死锁. LYRA 的群体协作机制 (GCC) 通过在每个实验开始时为智能体分配随机化的局部可观测标识符 $ID$ 打破对称性, 实现有序让行与去中心化协调, 而无需显式通信.

在策略执行阶段, GCC 机制使用相同的标识符来确定通行优先级与动态避让规则, 如图 6 所示. 具体而言, 当多个智能体同时接近受限区域时, 它们会比较各自的 $ID_{\text{mod}}$ 值: 标识值较小的智能体拥有更高的通行优先级, 而其他智能体则根据结果调整速度以防止阻塞. 速度更新规则定义为

$$\|\boldsymbol{v}_{t+1,\,i}\|=\begin{cases}0, & \exists j,\ ID_{j,\,\text{mod}}<ID_{i,\,\text{mod}}\\ \|\boldsymbol{v}_{t,\,i}\|, & \text{否则}\end{cases} \tag{10}$$

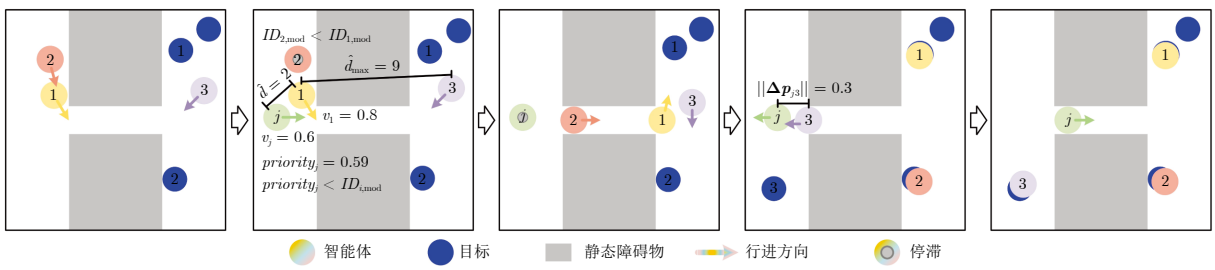



图 6 GCC 机制示意图

Fig. 6 Schematic diagram of the GCC mechanism

对于在运行过程中新进入 $7\times7$ 观测范围但尚未分配标识符的智能体, GCC 基于局部可观测信息计算临时优先级评分, 即

$$priority_{t,\,j}=0.5\cdot\left(1-\frac{\hat{d}_{t,\,j}}{\hat{d}_{\max}}\right)+0.5\cdot\left(1-\frac{\|\boldsymbol{v}_{t,\,j}\|}{\|\boldsymbol{v}_{\max}\|}\right)\tag{11}$$

其中, $\hat{d}_{t,\,j}$ 为智能体 $i$ 与新观测到智能体 $j$ 之间的曼哈顿距离, $\|\boldsymbol{v}_{t,\,j}\|$ 为局部观测得到的速度模值, $\hat{d}_{\max}$ 与 $\|\boldsymbol{v}_{\max}\|$ 分别为当前邻居集合中的最大距离与最大速度. 新检测到的智能体按 $priority_{t,\,j}$ 的降序插入局部优先级列表, 若优先级相同, 则根据其相对网格位置打破平局.

当高优先级智能体通过受限区域后, 低优先级智能体可通过局部观测检测其离开并恢复正常速度, 从而有效减少死锁传播. 此外, 由于轨迹偏离等动态因素可能导致优先级临时反转, GCC 增加了一个动态排斥项. 每个智能体持续监测与邻居的欧氏距离 $\|\boldsymbol{\Delta p}_{ij}\|$, 若该距离低于阈值 $\xi$ (实验中设为 0.3), 则触发横向偏移或速度调整以避免碰撞:

$$\|\boldsymbol{\Delta p}_{ij}\|=\|\boldsymbol{p}_{t,\,i}-\boldsymbol{p}_{t,\,j}\|_2\tag{12}$$

这种混合设计显著降低了局部拥堵与碰撞概率, 为后续路径学习与策略优化提供了稳定的协作基础.

### 3.3　全局引导的路径学习策略

即便具备稳健的局部避碰机制, 在部分可观测环境中, 智能体仍可能偏离长期目标路径. 为保持局部机动与长期任务目标的一致性, LYRA 设计基于奖励驱动的全局引导策略. 该机制将轨迹级先验嵌入策略学习过程, 并通过提供密集、方向性反馈, 引导智能体在适应动态环境变化的同时, 保持向目标区域的有效推进, 从而减少长期绕行与重复拥堵.

具体而言, 本研究采用包含稀疏项 $R_{t,\,i,\,s}$ 与稠密项 $R_{t,\,i,\,d}$ 相结合的层级奖励结构, 即

$$R_{t,\,i}=R_{t,\,i,\,s}+R_{t,\,i,\,d}\tag{13}$$

其中, 稀疏奖励 $R_{t,\,i,\,s}$ 反映关键任务事件, 通过二值指示函数定义为

$$R_{t,\,i,\,s}=I_{\mathrm{g}}\cdot100-\left(1-I_{\mathrm{g}}\right)\cdot\left(200\cdot I_{\mathrm{c}}+200\cdot I_{\mathrm{pd}}\right)\tag{14}$$

其中, $I_{\mathrm{g}},\ I_{\mathrm{c}},\ I_{\mathrm{pd}}\in\{0,\ 1\}$. 当智能体在时间步 $t$ 到达目标位置时, $I_{\mathrm{g}}=1$; 当发生与静态障碍、动态障碍或其他智能体的碰撞时, $I_{\mathrm{c}}=1$; 当智能体偏离全局路径的持续时间超过容差 $l_{\mathrm{dev}}$ (实验中设为 20 步) 时, $I_{\mathrm{pd}}=1$. 该设计保证任务完成优先, 避免同一时间步内出现奖励与惩罚并存, 同时能够允许在未达成目标时因碰撞与偏离共同累计惩罚. 稀疏奖励机制如图 7 所示.

虽然稀疏奖励 $R_{t,\,i,\,s}$ 能显式强化安全性与任务完成, 但它无法细化连续时间步间的行为调节. 为此, LYRA 增加稠密奖励项 $R_{t,\,i,\,d}$ 以引导逐步决策, 即

$$R_{t,\,i,\,d}=0.1\cdot I_{\mathrm{path}}-0.1\cdot t\tag{15}$$

其中, 当智能体在当前时间步沿全局路径前进时, $I_{\mathrm{path}}=1$, 否则为 0; $t$ 为截至当前时刻行进的时间. 前者鼓励持续朝目标方向前进, 后者惩罚冗余绕行与无效等待, 从而在高密度场景下维持路径通畅与系统吞吐率. 在 $R_{t,\,i,\,d}$ 中, 本文并未对智能体施加严格的几何路径一致性约束, 而是通过指标 $I_{\mathrm{path}}$ 描述其在连续决策过程中是否保持朝目标区域推进的整体趋势. 该设计允许智能体在局部避让或交互过程中产生必要的短时偏离, 同时仍可在后续决策中逐步回归全局引导路径, 从而在去中心化条件下兼顾局部机动灵活性与长期导航一致性.

通过将轨迹先验直接嵌入奖励结构, 该策略有效缓解了去中心化 DRL-MAPF 中因过度依赖即时避障而产生的局部近视问题, 使智能体能够在局部避碰与全局导航之间实现动态平衡.

### 3.4　自适应学习率机制

在无中心控制的动态环境中训练多智能体协同行动, 极易导致学习过程不稳定. 若采用固定或单调递减的学习率, 往往会进一步恶化这一问题: 前者可能使策略陷入局部安全但效率低下的模式, 后

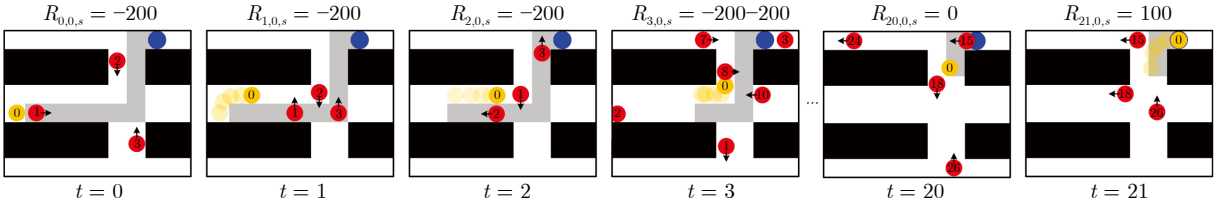



图 7　稀疏奖励机制示意图 (黑、红、灰、蓝、橙色分别代表静态障碍、动态障碍、全局引导路径、目标点与智能体)

Fig. 7　Schematic diagram of the sparse reward mechanism (the black, red, grey, blue, and orange cells represent static obstacles, dynamic obstacles, global guidance, goals, and agents, respectively)

者则易引发振荡, 阻碍收敛. 为此, LYRA 进一步引入自适应学习率 (adaptive learning rate, ALR) 机制, 通过动态调整参数更新幅度, 在训练过程中平衡稳定性与响应性.

受周期性学习率方法[48] 的启发, 本文提出一种针对 MAPF 训练优化的自适应学习率机制. 该机制依据训练阶段与收敛信号, 分别调整策略网络与价值网络的学习率, 使早期探索保持足够多样性, 而后期更新趋于稳定.

具体地, ALR 将指数衰减与周期性余弦调制相结合:

$$lr_{\pi}^{(k)} = lr_{0,\,\pi} \cdot \left(\frac{lr_{K,\,\pi}}{lr_{0,\,\pi}}\right)^{\frac{k}{K}} \cdot \left(1 + A\cos\left(\frac{2\pi}{C}k\right)\right) \tag{16}$$

$$lr_{v}^{(k)} = lr_{0,\,v} \cdot \left(\frac{lr_{K,\,v}}{lr_{0,\,v}}\right)^{\frac{k}{K}} \cdot \left(1 + A\cos\left(\frac{2\pi}{C}k\right)\right) \tag{17}$$

其中, $lr_{0,\,*}$ 和 $lr_{K,\,*}$ 分别为初始学习率和最终学习率, $K$ 为训练总轮数, $A$ 为调制幅度系数, $C$ 为调制周期. 在 MAPF 场景中, 这种周期性调制使策略能够周期性地重新探索不同的协作模式, 有助于避免智能体在狭窄通道中出现长期死锁. 此外, 为保持策略探索的多样性, ALR 机制还引入指数衰减的熵正则项:

$$\beta^{(k)} = \beta_0 \mathrm{e}^{-\lambda k} \tag{18}$$

其中, $\beta_0$ 为初始熵权重, $\lambda$ 为衰减速率. 该设计在训练早期保持动作的多样性 (对高密度 MAPF 场景中发现非平凡避障策略尤为重要), 而在系统形成稳定流动后逐渐降低随机性, 从而促进收敛.

通过显式地将学习率与熵权衰减设计关联至 MAPF 的可扩展性与稳定性挑战, ALR 机制确保在严重部分可观测与高拥挤条件下, 去中心化策略仍能稳定可靠地收敛.

## 4 实验结果

### 4.1 实验设置

为全面验证所提方法在不同环境与规模下的有效性与泛化能力, 本文选取四类具有代表性的测试地图 (如图 8 所示). 四种测试地图分别为: 规则、随机、空白与狭隘. 前三种地图尺寸为 $48 \times 48$, 来自公开数据集[49], 其障碍密度分别为 0.35、0.30 和 0.00. 狭隘地图尺寸为 $9 \times 9$, 障碍密度为 0.25, 并引入双向动态障碍以模拟瓶颈拥堵场景. 除狭隘地图外, 其余环境均包含 30 个动态障碍 (运动方式见第 2.1 节), 用于进一步评估算法的避障性能与鲁棒性.

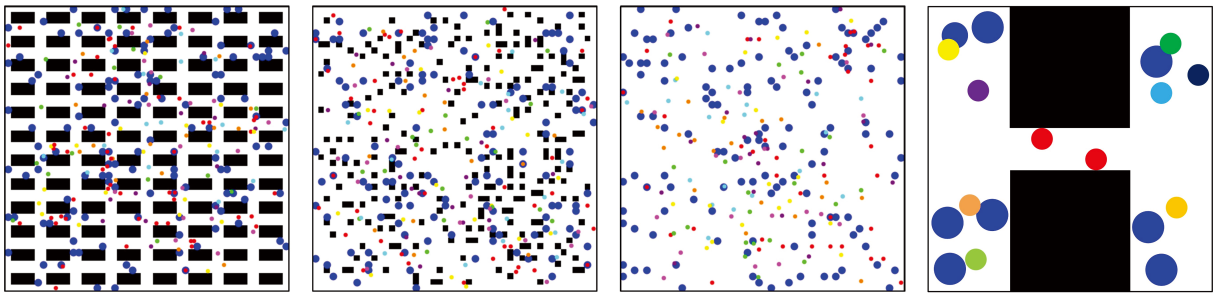
所有实验均在配备 NVIDIA GeForce RTX 3090 GPU、CUDA 11.8 与 PyTorch 1.11 的服务器上进行. LYRA 的主要超参数如表 1 所示.

此外, 为确保全面评估, 本文同时采用路径性能指标与行为合规性指标. 路径性能指标用于评估智能体导航的质量与效率, 包括成功率 ($SR$)、路径代价 ($C_{\mathrm{act}}$) 和平均完成时间 ($T_{\mathrm{avg}}$). 行为合规性指标用于评估智能体是否安全行驶并遵守环境约束, 包括碰撞率 ($CR$) 和越界概率 ($OOBP$).

### 4.2 实验结果

本文将 LYRA 与六种先进的分布式 DRL-MAPF 方法进行比较, 包括 MAPPER[34]、PRIMAL2[28]、DCC[32]、TIRL[37]、SCRIMP[20] 和 SIGMA[50]. 路径性能指标结果如表 2 和表 3 所示, 行为合规性指标结果如表 4 和表 5 所示. 图 9 展示了各方法的成功率变化趋势.

在所有环境下, LYRA 在各项配置中均取得最高的成功率 ($SR$), 在高密度场景中尤为显著. 例如, 在三种地图 (规则、随机、空白) 的 128 个智能体任务中, LYRA 的 $SR$ 达到 87.21%, 而 PRIMAL2 和



(a) 规则 (a) Regular　(b) 随机 (b) Random　(c) 空白 (c) Free　(d) 狭隘 (d) Narrow

图 8 MAPF 实验环境示意图 (黑色、红色、蓝色及其他颜色分别代表静态障碍、动态障碍、目标点与智能体)

Fig. 8 Schematic diagram of the MAPF experimental environment (the black, red, blue, and other color cells represent the static obstacle, the dynamic obstacle, the goals and the agents, respectively)

表 1　超参数设置

Table 1　Hyperparameter settings

| 超参数 | 取值 |
|---|---|
| 折扣因子 $\gamma$ | 0.95 |
| 策略网络初始学习率 $lr_{0,\,\pi}$ | $1\times10^{-5}$ |
| 价值网络初始学习率 $lr_{0,\,v}$ | $5\times10^{-5}$ |
| 衰减速率 $\lambda$ | 0.10 |
| 调制幅度系数 $A$ | 0.10 |
| 经验缓冲区分段长度 $L$ | 20 |
| 批大小 $B$ | 512 |
| 观测视野 $\boldsymbol{FOV}$ | $7\times7$ |
| 优化器类型 | Adam |
| 隐藏层维度 | 128 |
| 智能体 $ID$ 更新频率 | 每轮更新 |
| 裁剪系数 $\varepsilon$ | 0.15 |
| 策略网络最终学习率 $lr_{K,\,\pi}$ | $3\times10^{-7}$ |
| 价值网络最终学习率 $lr_{K,\,v}$ | $1\times10^{-7}$ |
| 初始熵权重 $\beta_0$ | $5\times10^{-3}$ |
| 周期系数 $C$ | 10.0 |
| 最小批量 $minibatch$ | 1 |
| 训练轮数 $K$ | 600 |
| 最大可见邻居数 $M$ | 8 |
| Adam 参数 ($\rho_1$, $\rho_2$, $\rho_3$) | (0.9, 0.999, $1\times10^{-8}$) |
| 激活函数 | ReLU |
| 动态障碍更新频率 | 每轮更新 |

DCC 分别仅为 49.77% 和 23.59%. 虽然 DCC 在小规模场景中由于其激进的局部规划策略可生成较短路径, 但在高密度环境中常导致拥堵与死锁, 其急剧下降的 $SR$ 与高碰撞率 ($CR$) 充分反映了这一点. PRIMAL2 依赖模仿学习, 在简单布局下表现良好, 但在缺乏显式拥堵处理机制时难以维持稳定协调. SCRIMP 通过消息传递增强环境感知, 但通信开销与响应延迟在瓶颈区域造成振荡与绕行. 同时, SIGMA 以局部安全约束为主的决策方式在缺乏显式冲突消解机制时更易产生减速与绕行, 从而在密集交互场景中成功率受限. 相比之下, LYRA 的基于共识的局部让行机制在密集交互条件下能够维持平稳通行, 从而在高负载场景中提升整体完成率.

在路径效率方面 ($C_{\text{act}}$ 与 $T_{\text{avg}}$), LYRA 与最优基线方法保持竞争力. 尽管 DCC 在小规模测试中获得最短路径, 但其高频碰撞在动态环境中抵消了该优势. LYRA 略高的 $T_{\text{avg}}$ 主要源于在密集交互中的主动礼让行为, 这种策略以有限的时间开销换取安全性与任务完成稳定性. 此外, SIGMA 在高密度场景中同样采取较为保守的通行策略, 其路径效率随规模增加而有所下降, 进一步体现出不同方法在安全性与效率之间的权衡差异.

在行为合规性指标 (表 4) 方面, LYRA 在所有场景下均实现了最低的碰撞率 ($CR$) 和接近零的越界概率 ($OOBP$). 在 128 个智能体条件下, LYRA 的 $CR$ 仅为 12.11%, 相比 MAPPER 的 38.84% 和 PRIMAL2 的 49.15% 有大幅改善. 这表明多模态感知与群体协作机制的结合能够有效抑制高风险行为, 使智能体始终保持在导航约束范围内.

在狭隘地图 (表 3 和表 5) 中, 算法在严苛空间约束下的差异进一步放大. 大多数基线方法的 $SR$ 随智能体数量增加而显著下降, 例如 DCC 从 2 个智能体时的 86.00% 下降至 8 个智能体时的 53.75%. 而 LYRA 始终保持 $SR>99\%$, 且无任何碰撞或越界事件, 表明其礼让机制在单通道场景中能够有效

表 2　三种地图 (规则、随机、空白) 中的路径质量对比

Table 2　Comparison of path quality in three maps (regular, random, and free)

| 智能体数量 | 指标 | MAPPER | PRIMAL2 | DCC | TIRL | SCRIMP | SIGMA | LYRA |
|---|---|---|---|---|---|---|---|---|
| 16 | $SR$ | 0.8704 | 0.8353 | 0.6333 | 0.8052 | 0.8479 | 0.9722 | **0.9831** |
| | $C_{\text{act}}$ | 32.8703 | 32.7627 | **29.5958** | 33.8917 | 33.4555 | 35.7576 | 30.0737 |
| | $T_{\text{avg}}$ (s) | 32.8947 | 32.7668 | **29.7577** | 34.0307 | 33.8156 | 36.6788 | 36.9454 |
| 32 | $SR$ | 0.8080 | 0.7794 | 0.5483 | 0.7640 | 0.8128 | 0.8826 | **0.9727** |
| | $C_{\text{act}}$ | 32.8446 | 32.8876 | **27.7264** | 34.6483 | 35.5310 | 32.5537 | 30.2058 |
| | $T_{\text{avg}}$ (s) | 32.8822 | 32.8954 | **27.9363** | 34.8102 | 36.1857 | 33.9869 | 36.6436 |
| 64 | $SR$ | 0.7158 | 0.6693 | 0.4090 | 0.6761 | 0.7494 | 0.7432 | **0.9238** |
| | $C_{\text{act}}$ | 33.7400 | 30.2338 | **25.2589** | 35.8568 | 39.2429 | 32.9997 | 30.2854 |
| | $T_{\text{avg}}$ (s) | 33.8197 | 30.2379 | **25.6223** | 36.0231 | 40.5643 | 35.6018 | 35.6982 |
| 128 | $SR$ | 0.5530 | 0.4977 | 0.2359 | 0.5336 | 0.6380 | 0.5747 | **0.8721** |
| | $C_{\text{act}}$ | 34.5973 | 29.1432 | **21.0439** | 36.8009 | 44.7254 | 32.1888 | 30.5756 |
| | $T_{\text{avg}}$ (s) | 34.7071 | 29.1453 | **21.5344** | 36.9884 | 47.0501 | 37.2716 | 35.1631 |

注: 加粗字体表示对应指标的最优结果.

表 3 狭隘地图中的路径质量对比
Table 3 Comparison of path quality in narrow map

| 智能体数量 | 指标 | MAPPER | PRIMAL2 | DCC | TIRL | SCRIMP | SIGMA | LYRA |
|---|---|---|---|---|---|---|---|---|
| 2 | $SR$ | 0.9430 | 0.9930 | 0.8600 | **0.9980** | 0.9930 | 0.9972 | 0.9936 |
| | $C_{\text{act}}$ | 7.2322 | **7.0081** | 8.4350 | 7.8367 | 9.1934 | 9.9995 | 10.4306 |
| | $T_{\text{avg}}$ (s) | 7.2492 | **7.0101** | 8.4950 | 7.8537 | 9.5438 | 10.0199 | 11.1547 |
| 4 | $SR$ | 0.8520 | 0.9640 | 0.5962 | 0.9480 | 0.9710 | **0.9923** | 0.9902 |
| | $C_{\text{act}}$ | 7.4988 | **7.3734** | 12.0792 | 8.2521 | 12.2307 | 11.9899 | 10.3881 |
| | $T_{\text{avg}}$ (s) | 7.5481 | **7.3786** | 12.2738 | 8.2880 | 13.1452 | 12.0885 | 11.4061 |
| 6 | $SR$ | 0.8054 | 0.9421 | 0.6117 | 0.9301 | 0.9361 | 0.9745 | **0.9976** |
| | $C_{\text{act}}$ | 7.5539 | 7.5339 | **5.8033** | 8.6406 | 14.1738 | 16.9993 | 10.4422 |
| | $T_{\text{avg}}$ (s) | 7.5539 | 7.5339 | **6.0068** | 8.6749 | 15.5160 | 17.2584 | 11.8271 |
| 8 | $SR$ | 0.7330 | 0.9140 | 0.5375 | 0.8680 | 0.8770 | 0.8803 | **0.9988** |
| | $C_{\text{act}}$ | 7.7203 | 8.0230 | **5.0262** | 9.2419 | 15.3615 | 17.7298 | 10.5779 |
| | $T_{\text{avg}}$ (s) | 7.8199 | 8.0295 | **5.1764** | 9.3007 | 17.0764 | 17.8249 | 12.3714 |

表 4 三种地图 (规则、随机、空白) 中的行为合规性对比 (%)
Table 4 Comparison of behavioral compliance in three maps (regular, random, and free) (%)

| 智能体数量 | 指标 | MAPPER | PRIMAL2 | DCC | TIRL | SCRIMP | SIGMA | LYRA |
|---|---|---|---|---|---|---|---|---|
| 16 | $CR$ | 10.61 | 14.35 | 36.64 | 19.34 | 15.15 | 2.78 | **1.39** |
| | $OOBP$ | 2.35 | 2.12 | 0.03 | 0.13 | 0.07 | **0.00** | 0.10 |
| 32 | $CR$ | 16.02 | 20.46 | 45.16 | 23.31 | 18.65 | 11.74 | **2.54** |
| | $OOBP$ | 3.19 | 1.60 | 0.01 | 0.29 | 0.07 | **0.00** | **0.00** |
| 64 | $CR$ | 24.48 | 31.60 | 59.10 | 32.29 | 24.90 | 25.68 | **7.23** |
| | $OOBP$ | 3.94 | 1.47 | 0.01 | 0.10 | 0.16 | **0.00** | 0.20 |
| 128 | $CR$ | 38.84 | 49.15 | 76.42 | 46.52 | 35.91 | 42.53 | **12.11** |
| | $OOBP$ | 5.86 | 1.08 | **0.00** | 0.13 | 0.29 | **0.00** | 0.10 |

表 5 狭隘地图中的行为合规性对比 (%)
Table 5 Comparison of behavioral compliance in narrow map (%)

| 智能体数量 | 指标 | MAPPER | PRIMAL2 | DCC | TIRL | SCRIMP | SIGMA | LYRA |
|---|---|---|---|---|---|---|---|---|
| 2 | $CR$ | 3.00 | 0.60 | 14.00 | 0.20 | 0.30 | 0.28 | **0.00** |
| | $OOBP$ | 2.70 | 0.10 | 2.50 | **0.00** | 0.40 | **0.00** | **0.00** |
| 4 | $CR$ | 11.60 | 2.90 | 41.38 | 5.20 | 2.40 | 0.77 | **0.00** |
| | $OOBP$ | 3.20 | 0.70 | 8.00 | **0.00** | 0.50 | **0.00** | **0.00** |
| 6 | $CR$ | 16.07 | 4.59 | 38.83 | 6.99 | 5.79 | 2.55 | **0.00** |
| | $OOBP$ | 3.39 | 1.20 | **0.00** | **0.00** | 0.60 | **0.00** | **0.00** |
| 8 | $CR$ | 22.30 | 5.80 | 46.25 | 13.20 | 11.60 | 11.97 | **0.00** |
| | $OOBP$ | 4.40 | 2.80 | **0.00** | **0.00** | 0.70 | **0.00** | **0.00** |

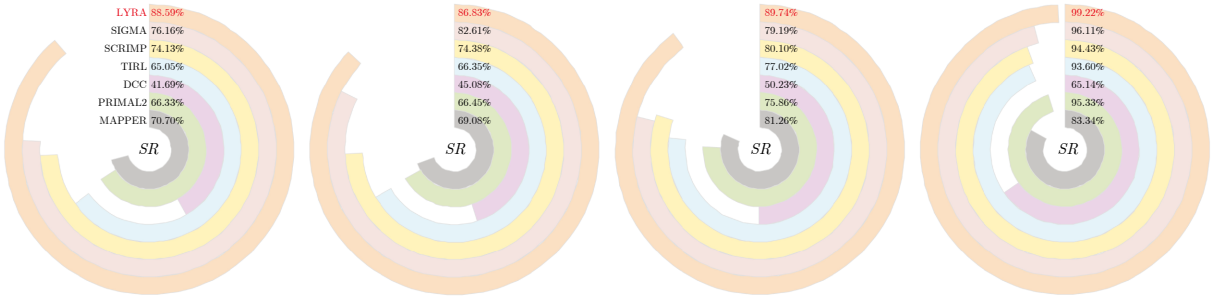
避免阻塞. 虽然 LYRA 的 $C_{\text{act}}$ 与 $T_{\text{avg}}$ 略高于基准规划器, 但这些差异源于有意的协同控制, 而非路径规划效率的下降.

实验结果表明, LYRA 通过多模态状态表示、基于奖励的路径学习及共识驱动的协调机制, 在不同环境密度条件下实现了稳定性、可扩展且安全的多智能体协作表现. 与现有分布式 DRL-MAPF 方法相比, LYRA 在任务完成率与行为规范性两方面均表现出一致优势.

### 4.3 实物实验

为进一步验证 LYRA[2] 在实际部署中的实用性和稳健性, 本文使用索尼 Toio 机器人作为移动代理进行物理多机器人实验 (见图 10). 每个 Toio 单元都是一个紧凑的差速驱动平台 (尺寸为 31.8 mm × 31.8 mm × 35.1 mm), 配备内置惯性传感器、通过官方 Toio 游戏垫实现的光学运动追踪以及低功耗

[2] 实物演示视频见: https://github.com/HuoLinL/LYRA.

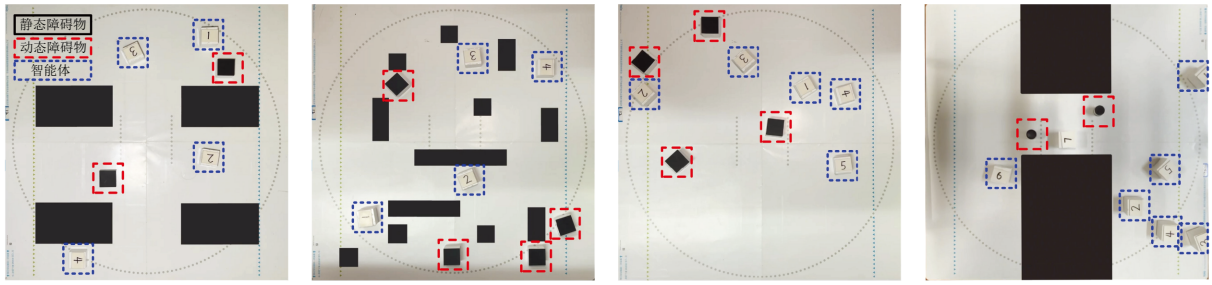



(a) 规则　(b) 随机　(c) 空白　(d) 狭隘

(a) Regular　(b) Random　(c) Free　(d) Narrow

图 9　不同 MAPF 环境的算法成功率对比

Fig. 9　Comparison of success rates of algorithms in different MAPF environments




(a) 规则　(b) 随机　(c) 空白　(d) 狭隘

(a) Regular　(b) Random　(c) Free　(d) Narrow

图 10　实物实验地图

Fig. 10　Physical experiment map

蓝牙 (Bluetooth low energy, BLE) 通信功能.

集中式主机实时运行经过 LYRA 训练的去中心化策略. 主机通过 BLE 以 10 Hz 的频率向每个机器人发送运动命令, 而机器人则自主执行这些命令, 无需智能体间通信. 每个机器人仅接收其本地观测值, 该观测值源自其机载位置以及检测到的附近障碍物和智能体的相对位置. 在狭隘地图场景中, LYRA 基于共识的让行行为自然出现, 允许车辆顺利通过单车道走廊, 而无需明确的优先规则, 由此证明了其在现实世界中的可扩展性和超越模拟的稳健性.

## 5　结束语

本文提出一种面向部分可观测与动态环境的基于局部让行与推理的多智能体路径规划深度强化学习框架. 该框架针对两个挑战展开: 1) 可见性与密度变化下的协作不稳定性; 2) 仅依赖局部避障导致的路径偏移问题. LYRA 将多模态状态编码、群体协作共识与基于奖励的路径学习有机结合, 构建一个统一的决策流程, 使局部动作同时服务于即时避障与长程导航目标. 此外, 自适应学习率机制为大规模训练提供了稳定性与可扩展性支持.

实验结果表明, LYRA 在路径性能与行为合规性方面均显著优于代表性基线方法, 其性能提升在不同环境密度与结构下保持一致, 验证了方法的普适性与稳健性. 这些结果表明, 将结构化感知与基于共识的协调相结合, 为在复杂动态场景中实现更稳健、高效、可扩展的去中心化 DRL-MAPF 提供了有力的实践基础.

未来工作将进一步拓展该框架, 特别是在具备动态目标与连续控制任务的场景中, 深度融合通信感知与规划机制, 以实现更高层次的智能协同控制.

## 参考文献

1　Zhang Kai-Xiang, Mao Jian-Lin, Xiang Feng-Hong, Xuan Zhi-Wei. B-IHCA*, a bargaining game based multi-agent path finding algorithm. *Acta Automatica Sinica*, 2023, **49**(7): 1483−1497
(张凯翔, 毛剑琳, 向凤红, 宣志玮. 基于讨价还价博弈机制的 B-IHCA* 多机器人路径规划算法. 自动化学报, 2023, **49**(7): 1483−1497)

2　Li Jian-Qiang, Cai Jun-Chuang, Sun Tao, Zhu Qing-Ling, Lin Qiu-Zhen. Multitask-based assisted evolutionary algorithm for vehicle routing problems incomplex logistics distribution scenarios. *Acta Automatica Sinica*, 2024, **50**(3): 544−559
(李坚强, 蔡俊创, 孙涛, 朱庆灵, 林秋镇. 面向复杂物流配送场景的车辆路径规划多任务辅助进化算法. 自动化学报, 2024, **50**(3): 544−559)

3　Hao Zhao-Tie, Guo Bin, Zhao Kai-Xing, Wu Lei, Ding Ya-San, Li Zhe-Tao, et al. From rule-driven to collective intelligence emergence: A review of research on multi-robot air-ground collaboration. *Acta Automatica Sinica*, 2024, **50**(10): 1877−1905

(郝肇铁, 郭斌, 赵凯星, 吴磊, 丁亚三, 李哲涛, 等. 从规则驱动到群智涌现: 多机器人空地协同研究综述. 自动化学报, 2024, **50**(10): 1877−1905)

4 Xuan Zhi-Wei, Mao Jian-Lin, Zhang Kai-Xiang. Low-expansion A* algorithm based on CBS framework for complex map. *Acta Electronica Sinica*, 2022, **50**(8): 1943−1950
(宣志玮, 毛剑琳, 张凯翔. CBS 框架下面向复杂地图的低拓展度 A* 算法. 电子学报, 2022, **50**(8): 1943−1950)

5 Qian Cheng-Ze, Mao Jian-Lin, Li Rui-Qi, Zhou Wen-Na, Gong De-Zheng, Zhang Jin-Bao. Improved PBS multi-robot path planning algorithm based on conflict cost Bayesian weighting. *Acta Electronica Sinica*, 2025, **53**(7): 2358−2371
(钱诚泽, 毛剑琳, 李睿祺, 周雯娜, 龚德正, 张进宝. 基于冲突代价 Bayesian 权重的改进 PBS 多智能体路径规划算法. 电子学报, 2025, **53**(7): 2358−2371)

6 Veerapaneni R, Saleem M S, Li J, Likhachev M. Windowed MAPF with completeness guarantees. In: Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI, 2025. 23323−23332

7 Zhang Shu-Fan, Mao Jian-Lin, Zhang Kai-Xiang, Li Rui-Qi, Li Da-Yan, Wang Ni-Ya. Survey on robust multi-robot path planning under uncertainty. *Control and Decision*, 2024, **39**(12): 3873−3888
(张书凡, 毛剑琳, 张凯翔, 李睿祺, 李大炎, 王妮娅. 面向不确定性的多机器人路径鲁棒规划研究综述. 控制与决策, 2024, **39**(12): 3873−3888)

8 Zhu K, Zhang T. Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, 2021, **26**(5): 674−691

9 Wang B, Liu Z, Li Q, Prorok A. Mobile robot path planning in dynamic environments through globally guided reinforcement learning. *IEEE Robotics and Automation Letters*, 2020, **5**(4): 6932−6939

10 Huo L, Mao J, San H, Li R, Zhang S. Deep reinforcement learning of group consciousness for multi-robot pathfinding. *Engineering Applications of Artificial Intelligence*, 2025, **155**: Article No. 110978

11 Andreychuk A, Yakovlev K, Panov A, Skrynnik A. MAPF-GPT: Imitation learning for multi-agent pathfinding at scale. In: Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI, 2025. 23126−23134

12 Jin W, Du H, Zhao B, Tian X, Shi B, Yang G. A comprehensive survey on multi-agent cooperative decision-making: Scenarios, approaches, challenges and perspectives. arXiv preprint arXiv: 2503.13415, 2025.

13 Anastassacos N, Hailes S, Musolesi M. Partner selection for the emergence of cooperation in multi-agent systems using reinforcement learning. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 7047−7054

14 Yu C, Velu A, Vinitsky E, Gao J, Wang Y, Bayen A, et al. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 2022, **35**: 24611−24624

15 Li Q, Lin W, Liu Z, Prorok A. Message-aware graph attention networks for large-scale multi-robot path planning. *IEEE Robotics and Automation Letters*, 2021, **6**(3): 5533−5540

16 Yao S, Chen G, Pan L, Ma J, Ji J, Chen X. Multi-robot collision avoidance with map-based deep reinforcement learning. In: Proceedings of the 32nd IEEE International Conference on Tools with Artificial Intelligence. Baltimore, USA: IEEE, 2020. 532−539

17 Alkazzi J M, Okumura K. A comprehensive review on leveraging machine learning for multi-agent path finding. *IEEE Access*, 2024, **12**: 57390−57409

18 Yao Z, Wang W. Layeredmapf: A decomposition of MAPF instance to reduce solving costs. arXiv preprint arXiv: 2404.12773, 2024.

19 Omidshafiei S, Pazis J, Amato C, How P J, Vian J. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR. org, 2017. 2681−2690

20 Wang Y, Xiang B, Huang S, Sartoretti G. Scrimp: Scalable communication for reinforcement-and imitation-learning-based multi-agent pathfinding. In: Proceedings of the International Conference on Intelligent Robots and Systems. Detroit, USA: IEEE, 2023. 9301−9308

21 Okumura K. LaCAM: Search-based algorithm for quick multi-agent pathfinding. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI, 2023. 11655−11662

22 Yang Y, Fan M, He C, Wang J, Huang H, Sartoretti G. Attentionbased priority learning for limited time multi-agent path finding. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems. Auckland, New Zealand: IFAAMAS, 2024. 1993−2001

23 Shi Q, Liu M, Zhang S, Lan X. Reinforcement learning for multi-agent path finding in large-scale warehouses via distributed policy evolution. *IEEE Robotics and Automation Letters*, 2025, **10**(8): 7843−7850

24 Jiang H, Zhang Y, Veerapaneni R, Li J. Scaling lifelong multi-agent path finding to more realistic settings: Research challenges and opportunities. In: Proceedings of the 17th International Symposium on Combinatorial Search. Kananaskis, Canada: AAAI, 2024. 234−242

25 Wu Z, Yu C, Chen C, Hao J, Zhuo H H. Plan to predict: Learning an uncertainty-foreseeing model for model-based reinforcement learning. *Advances in Neural Information Processing Systems*, 2022, **35**: 15849−15861

26 Sartoretti G, Kerr J, Shi Y, Wagner G, Kumar T S, Koenig S, et al. PRIMAL: Pathfinding via reinforcement and imitation multiagent learning. *IEEE Robotics and Automation Letters*, 2019, **4**(3): 2378−2385

27 Wang J, Meng M Q H. Real-time decision making and path planning for robotic autonomous luggage trolley collection at airports. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, **52**(4): 2174−2183

28 Damani M, Luo Z, Wenzel E, Sartoretti G. PRIMAL$_2$: Pathfinding via reinforcement and imitation multi-agent learning-lifelong. *IEEE Robotics and Automation Letters*, 2021, **6**(2): 2666−2673

29 Abreu N. Efficient deep learning for multi agent pathfinding. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual Event: AAAI, 2022. 13122−13123

30 Ma Z, Luo Y, Ma H. Distributed heuristic multi-agent path finding with communication. In: Proceedings of the IEEE International Conference on Robotics and Automation. Xi′an, China: IEEE, 2021. 8699−8705

31 Yan Z, Wu C. Neural neighborhood search for multi-agent path finding. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: IEEE, 2024.

32 Ma Z, Luo Y, Pan J. Learning selective communication for multiagent path finding. *IEEE Robotics and Automation Letters*, 2021, **7**(2): 1455−1462

33 Lin Q, Ma H. Sacha: Soft actor-critic with heuristic-based attention for partially observable multi-agent path finding. *IEEE Robotics and Automation Letters*, 2023, **8**(8): 5100−5107

34 Liu Z, Chen B, Zhou H, Koushik G, Hebert M, Zhao D. MAPPER: Multi-agent path planning with evolutionary reinforcement learning in mixed dynamic environments. In: Proceedings of the International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE, 2020. 11748−11754

35 He C, Duhan T, Tulsyan P, Kim P, Sartoretti G. Social behavior as a key to learning-based multi-agent pathfinding dilemmas. *Artificial Intelligence*, 2025, **348**: Article No. 104397

36 He C, Yang T, Duhan T, Wang Y, Sartoretti G. ALPHA: Attentionbased long-horizon pathfinding in highly-structured areas. In: Proceedings of the International Conference on Robotics and Automation. Yokohama, Japan: IEEE, 2024. 14576−14582

37 Chen L, Wang Y, Miao Z, Mo Y, Feng M, Zhou Z, et al. Transformer-based imitative reinforcement learning for multirobot path planning. *IEEE Transactions on Industrial Informatics*, 2023, **19**(10): 10233−10243

38 Gao J, Li Y, Li X, Yan K, Lin K, Wu X. A review of graph-based multi-agent pathfinding solvers: From classical to beyond classical. *Knowledge-Based Systems*, 2024, **283**: Article No. 111121

39 Hu S, Shen L, Zhang Y, Chen Y, Tao D. On transforming rein-

forcement learning with transformers: The development trajectory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, **46**(12): 8580−8599

40 Shen B, Chen Z, Cheema M A, Harabor D D, Stuckey P J. Tracking progress in multi-agent path finding. arXiv preprint arXiv: 2305.08446, 2023.

41 Pham P, Bera A. Optimizing crowd-aware multi-agent path finding through local communication with graph neural networks. arXiv preprint arXiv: 2309.10275, 2023.

42 Chen Z, Harabor D, Li J, Stuckey P J. Traffic flow optimisation for lifelong multi-agent path finding. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2024. 20674−20682

43 Tang H, Berto F, Park J. Ensembling prioritized hybrid policies for multi-agent pathfinding. In: Proceedings of the International Conference on Intelligent Robots and Systems. Abu Dhabi, United Arab Emirates: IEEE, 2024. 8047−8054

44 Phan T, Phan T, Koenig S. Generative curricula for multi-agent path finding via unsupervised and reinforcement learning. *Journal of Artificial Intelligence Research*, 2025, **82**: 2471−2534

45 Korolyuk V S, Brodi S, Turbin A. Semi-Markov processes and their applications. *Journal of Soviet Mathematics*, 1975, **4**(3): 244−280

46 Hart P E, Nilsson N J, Raphael B. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 1968, **4**(2): 100−107

47 Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. arXiv preprint arXiv: 1312.5602, 2013.

48 Smith L N. Cyclical learning rates for training neural networks. In: Proceedings of the Winter Conference on Applications of Computer Vision. Santa Rosa, USA: IEEE, 2017. 464−472

49 Stern R, Sturtevant N, Felner A, Koenig S, Ma H, Walker T, et al. Multi-agent pathfinding: Definitions, variants, and benchmarks. In: Proceedings of the 12th Conference on International Symposium on Combinatorial Search. Napa, USA: AAAI, 2019. 151−158

50 Liao S, Xia W, Cao Y, Dai W, He C, Wu W, et al. SIGMA: Sheaf-informed geometric multiagent pathfinding. In: Proceedings of the International Conference on Robotics and Automation. Atlanta, USA: IEEE, 2025. 1−7

**霍 琳** 昆明理工大学机电工程学院博士研究生. 主要研究方向为深度强化学习, 多机器人路径规划.
E-mail: huolin1223@foxmail.com
(**HUO Lin** Ph.D. candidate at the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. Her research interests include deep reinforcement learning and multi-robot path finding.)

**毛剑琳** 昆明理工大学信息工程与自动化学院教授. 主要研究方向为多机器人路径规划, 智能优化算法. 本文通信作者.
E-mail: jlmao@kust.edu.cn
(**MAO Jian-Lin** Professor at the Faculty of Information Engineering and Automation, Kunming University of Science and Technology. Her research interests include multi-robot path finding and intelligent optimization algorithms. Corresponding author of this paper.)

**伞红军** 昆明理工大学机电工程学院副教授. 主要研究方向为机械结构设计, 多机器人路径规划.
E-mail: sanhjun@163.com
(**SAN Hong-Jun** Associate professor at the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His research interests include mechanical structure design and multi-robot path finding.)

**付丽霞** 昆明理工大学信息工程与自动化学院副教授. 主要研究方向为集群智能控制优化算法.
E-mail: 12309049@kust.edu.cn
(**FU Li-Xia** Associate professor at the Faculty of Information Engineering and Automation, Kunming University of Science and Technology. Her main research interest is cluster intelligent control optimization algorithm.)

**宣志玮** 昆明理工大学机电工程学院博士研究生. 主要研究方向为深度强化学习, 多机器人路径规划.
E-mail: zhiweixuan@stu.kust.edu.cn
(**XUAN Zhi-Wei** Ph.D. candidate at the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His research interests include deep reinforcement learning and multi-agent path finding.)

**贺志刚** 昆明理工大学机电工程学院博士研究生. 主要研究方向为分布式多机器人路径规划.
E-mail: zhiganghe@stu.kust.edu.cn
(**HE Zhi-Gang** Ph.D. candidate at the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His main research interest is distributed multi-robot path finding.)