

doi:10.13682/j.issn.2095-6533.2026.05.009

端到端模仿学习的免通信多机器人分布式包围控制算法

钟远湘, 刘明雍, 闫志强, 玄立伟

(西北工业大学 航海学院, 陕西 西安 710072)

摘要: 针对现有多智能体深度强化学习方法依赖大量在线交互探索, 样本效率低, 且训练不稳定, 难以满足实际应用时离线训练完成后可直接部署的需求, 提出一种端到端模仿学习的免通信多机器人分布式包围控制算法。该算法使机器人仅依靠自身对邻近机器人和障碍物的局部感知信息, 通过一个预训练的、基于注意力机制的端到端神经网络, 直接生成运动控制指令, 实现连通、无碰撞的目标包围。神经网络通过监督学习方式, 使用离线数据进行训练, 复现集中式最优决策行为, 避免深度强化学习中的探索成本与收敛不稳定性。仿真结果表明, 所提算法在保持连通性与无碰撞约束的前提下, 在 6~15 个机器人规模的测试场景中, 均能成功完成包围任务, 在多种规模场景与复杂环境下展现出良好的泛化性与鲁棒性。

关键词: 多机器人系统; 包围控制; 行为克隆; 注意力机制; 局部感知; 无通信协同

中图分类号: TP273; TP181

文献标志码: A

文章编号: 2095-6533(2026)05-0092-12

Communication-free distributed multi-robot enclosing control algorithm based on end-to-end imitation learning

ZHONG Yuanxiang, LIU Mingyong, YAN Zhiqiang, XUAN Liwei

(School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Existing multi-agent deep reinforcement learning methods typically rely on extensive online interaction and exploration, suffering from low sample efficiency and training instability, which fail to meet the requirement of direct deployment after offline training in practical applications. To address these challenges, a communication-free distributed encirclement control algorithm based on end-to-end imitation learning is proposed. The algorithm enables robots to directly generate motion control commands using only local perception information of neighboring robots and obstacles through a pre-trained end-to-end neural network with attention mechanisms, thereby achieving connectivity-preserving and collision-free target encirclement. The neural network is trained via supervised learning with offline data to replicate centralized optimal decision-making behavior, which avoids the exploration costs and convergence instability inherent in deep reinforcement learning. Simulation results demonstrate that the proposed algorithm can successfully complete the encirclement task in test scenarios involving 6 to 15 robots, while maintaining connectivity and collision-free constraints. It exhibits good generalization and robustness across various scale scenarios and complex environments.

收稿日期: 2025-10-22

基金项目: 国家自然科学基金项目(52271332, 52071271)

通信作者: 刘明雍, liumingyong@nwpu.edu.cn

引文格式: 钟远湘, 刘明雍, 闫志强, 等. 端到端模仿学习的免通信多机器人分布式包围控制算法[J]. 西安邮电大学学报, 2026, 31(5): 92-103.

ZHONG Yuanxiang, LIU Mingyong, YAN Zhiqiang, et al. Communication-free distributed multi-robot enclosing control algorithm based on end-to-end imitation learning[J]. Journal of Xi'an University of Posts and Telecommunications, 2026, 31(5): 92-103.

Keywords: multi-robot systems; encirclement control; behavior cloning; attention mechanism; local perception; communication-free coordination

多机器人包围控制旨在使多个自主体在目标周围形成稳定且具有约束性的空间分布结构,是群体机器人在军事侦察、海洋探测、灾害救援及安防巡逻等场景中的核心协作能力。例如,在反恐作战中,多架无人机需对高速移动的可疑车辆进行持续跟踪包围以实施精确打击;在深海探测任务中,多台自主水下航行器需在全球定位系统(Global Positioning System, GPS)信号缺失、声学通信严重受限的条件下对目标沉船或海底热泉进行协同围捕式探测。与编队控制或集群跟踪相比,包围控制通常需要同时满足目标动态跟踪、均匀分布、碰撞避免、鲁棒性与实时性等多重要求,且系统结构呈现强耦合、高维度与动态拓扑等特征^[1-4]。在实践中,由于通信带宽有限、传感不完备或存在潜在对抗干扰,机器人往往难以获取全局信息或维持稳定通信,使得在局部感知、通信受限甚至完全无通信条件下实现可信赖的分布式包围控制成为重大挑战^[5]。因此,构建一种在复杂环境下仍具备可扩展性、鲁棒性与实时性的包围控制策略,是该领域长期关注的关键问题^[6-7]。

为应对上述挑战,已有研究提出了多种解决方案。传统方法主要依赖基于图论的控制律、协同优化方法或模型预测控制(Model Predictive Control, MPC)等,并在一定程度上实现了稳定的包围结构^[5]。然而,这类方法往往假设可获得准确系统模型、稳定通信或全局信息,并在高维、强非线性或拓扑动态变化的场景中表现受限。近年来,深度强化学习(Deep Reinforcement Learning, DRL)方法及其多智能体 DRL 方法为解决多机器人系统中的高维状态、复杂动态与部分可观测性问题提供了新的技术路线^[8-12]。典型方法包括价值分解类方法 VDN^[13] (Value-Decomposition Networks)、QMIX^[14] (Monotonic Value Function Factorisation)、多智能体 Actor-Critic 方法 MADDPG^[15] (Multi-Agent Deep Deterministic Policy Gradient)、COMA^[16] (Counterfactual Multi-Agent Policy Gradients),以及基于通信机制的学习方法 CommNet^[17] (Communication Network) 和 TarMAC^[18] (Targeted Multi-Agent Communication)。这些方法已在路径规划、编队控制等任务中展示出良好性能。

然而,将多智能体深度强化学习(Multi-Agent Deep Reinforcement Learning, MADRL)直接应用

于包围控制仍遇到诸多瓶颈。首先是低样本效率问题, MADRL 通常依赖大量在线交互样本,当环境较为复杂、机器人数量较多,或在水下等部署困难的场景中,难以获取足够数量的实物机器人训练数据集,使用纯仿真训练数据则又会引入误差。其次是训练难度大,多智能体目标围捕过程涉及高层任务分配、中层路径规划及底层运动控制等子任务,需要调用一系列规划器进行组合优化,这些复杂的任务目标导致奖励函数设计困难、神经网络难以收敛。再次是在线探索成本高, MADRL 需要多智能体系统不断对未知的环境及状态进行探索,并同步对 Actor、Critic 等神经网络进行迭代,当智能体数量较少时,探索过程能够高效推进。随着智能体数量增加,组合状态空间会指数级增长,导致 MADRL 的同步探索—优化过程难以维系。最后,现有 MADRL 方法普遍需要显式通信、中央状态信息或共享中间特征,而这些假设在通信受限的实际包围场景中往往难以满足。因此,单纯依赖多智能体深度强化学习方法难以全面解决包围控制任务中“多约束、多目标、无通信、高鲁棒性”的综合需求。

相比之下,模仿学习提供了一种更为高效的替代方案。模仿学习通过学习专家演示数据直接获取控制策略,避免了强化学习中的试错过程和奖励函数设计难题^[19]。近期研究表明,模仿学习相比深度强化学习具有显著的样本优势,在连续控制任务中,行为克隆方法相比深度强化学习方法可减少 39~100 倍的样本需求^[20-21]。最重要的是,在多机器人包围控制场景中,模仿学习可以利用已知全局信息的集中式最优控制器^[22],如 MPC 或最优控制算法作为专家,生成大量高质量的演示轨迹,通过行为克隆,将其“蒸馏”至仅依赖局部感知的分布式策略中,在无需通信的前提下实现高效协调控制。这一思路已在多机器人编队控制中得到了验证^[23-24]。此外,多智能体模仿学习的另一个优势在于其能够以 off-line 的方式进行,而无需像强化学习一样,以 on-line 的方式与环境实时交互,从而获取训练数据。该优势使其在水下、战场等部署较为困难的场景中,能够更便捷地使用实物机器人(例如无人机、无人船、水下自主航行器等)采集的数据训练,使其拥有更高的工程适用性。

在多智能体包围控制场景中,每个智能体感知到的邻居数量会随着相对位置变化而动态波动,

这使得传统的固定维度神经网络输入难以适应可规模变化的邻居信息。近年来,注意力机制的提出也为解决该问题提供了有力的技术支撑。借助注意力机制,智能体能够聚焦于对运动决策更为关键的数据信息、捕捉环境与状态空间的变化,并结合历史序列信息,对当前状态下的最优运动指令进行准确预测。例如 Shariq 等^[25]利用自注意力模块提取关键邻居特征以改善编队控制性能;Zhang 等^[26]利用安全注意力模块处理时变连接性下的邻居信息聚合,有效提升了多智能体系统的安全性与协调能力。因此,本文拟提出一种端到端模仿学习的免通信多机器人分布式包围控制算法。该算法采用端到端学习方式,直接从激光雷达等局部传感器数据映射到运动控制指令,避免复杂的特征工程。同时采用免通信架构,机器人仅依靠局部感知即可实现分布式协同,无需显式通信基础设施。在训练时,采用行为克隆方法,通过离线学习集中式专家控制器生成的演示轨迹,训练一个仅依赖局部感知的端到端神经网络。该网络输入包括自身状态、邻近机器人相对位置速度及局部障碍物信息,使用注意力机制处理可变数量的邻居特征,直接输出速度控制指令,从而将集中式最优决策能力迁移至分布式执行层。最后,在仿真环境中针对不同机器人数量、障碍物分布及目标运动模式进行测试,评估算法的性能表现,并分析其在规模变化和扰动下的泛化能力。

1 任务描述与模型构建

1.1 多机器人围捕任务描述

多机器人围捕任务场景^[27]:在一个包含静态障碍物的二维平面环境中, N 个同构移动机器人需要协同完成对一个动态目标的包围与追踪。目标以时变速度在环境中移动,试图逃离机器人的包围。机器人的任务目标:在目标周围形成均匀分布的圆形包围阵型^[28],保持期望的包围半径 r_d ,并持续跟踪目标运动,限制其活动范围。

在执行过程中,每个机器人仅依靠局部传感器获取感知半径 r_s 内的邻居机器人、目标及障碍物的相对位置信息,无法获取全局定位信息或与其他机器人进行显式通信。同时,机器人需满足避免相互碰撞、避免障碍物碰撞以及维持编队连通性等安全约束。这种设定模拟了水下搜救、无人机集群侦察及反恐围捕等实际应用场景中的通信受限、环境复杂等典型特征,对算法的分布式决策能力和鲁棒性

提出了严峻挑战。

1.2 机器人运动模型

多机器人目标包围任务存在 N 个被建模为具有非完整约束的单积分器模型的机器人,以及一个待包围目标。设机器人 i 在全局坐标系下的位置为 $P_i=(x_i, y_i) \in \mathbb{R}^2$,其运动学方程为

$$\dot{P}_i = v_i, i=1, 2, \dots, N \quad (1)$$

式中: $v_i=(v_{i,x}, v_{i,y})$ 表示机器人 i 的控制输入,即速度控制指令满足速度约束 $\|v_i\| \leq v_{\max}$, $v_{\max}$ 为机器人的最大速度。

这种单积分器模型被广泛应用于多机器人协同控制研究中,能够有效描述具有全向运动能力的移动机器人,如全向轮机器人、无人机等的运动特性。

1.3 目标运动模型

设目标位置为 $P_T=(x_T, y_T) \in \mathbb{R}^2$,目标以时变速度 v_T 运动,其运动轨迹表达式为

$$\dot{P}_T = v_T \quad (2)$$

式中: v_T 满足 $\|v_T\| \leq v_{T,\max}$, $v_{T,\max}$ 为目标的最大速度,同时满足约束 $v_{T,\max} < v_{\max}$,确保机器人能够追踪目标。

1.4 感知模型

采用局部感知模型,每个机器人仅能获取其感知范围内的信息,无需全局定位系统或机器人间通信,这一设定更符合水下、室内等复杂环境的实际应用场景。

1.4.1 相对位置感知

每个机器人可以在有限的感知范围 L 内相对定位附近的队友和目标,并受到时间延迟 τ 影响^[29]。设 $\mathcal{N}_i(t)=\{i, S_i(t)\}$ 表示机器人 i 在 t 时刻与其邻域集 S_i 的索引并集。机器人 i 直接观察到的相邻机器人 $j \in \mathcal{N}_i$ 的相对位置表示为 $P_i^j=(x_i^j, y_i^j)$,对目标的相对位置观测表示为 $P_i^T(t-\tau)$ 。传感器测量误差服从零均值正态分布,即 $e_{ij}(t) \sim \mathcal{N}(0, \sigma_p^2 \mathbf{I}_2)$,其中: $e_{ij}(t)$ 表示机器人 i 对邻居机器人 j 的相对位置测量误差; σ_p^2 表示位置测量误差标准差; $\mathbf{I}_2$ 为二维单位矩阵。因此,机器人 i 在 t 时刻的邻居位置感知信息集可表示为

$$P_{ij}(t)=\{P_i(t-\tau)-P_i(t)+e_{ij}(t) | j \in \mathcal{N}_i(t)\} \quad (3)$$

1.4.2 障碍物感知

设环境中存在静态障碍物,距离机器人最近的障碍物 $k(k=1, 2, \dots, N)$ 位置为 $P_{k,\text{obs}}=(x_k, y_k)$,则机器人 i 能够检测到感知范围内的障碍物信息集形式为

$$\mathcal{O}_i = \{k \mid \|P_i - P_{k, \text{obs}}\| \leq r_s\} \quad (4)$$

对于每个检测到的障碍物 $k \in \mathcal{O}_i$, 机器人获取相对位置信息表达式为

$$P_i^{k, \text{obs}} = P_i - P_{k, \text{obs}} \quad (5)$$

1.5 约束条件

约束条件主要包括避碰约束和连通性约束。

1) 避碰约束。在集中式训练与分布式执行阶段, 皆使用统一的双层避碰策略: 规划层使用 On-Conflict 算法^[30] 进行无碰撞的协同路径规划、运动层使用最优互惠碰撞避免 (Optimal Reciprocal Collision Avoidance, ORCA) 算法^[31] 实现实时避障。通过该策略, 能够保证集群在协同围捕的过程中不发生碰撞。机器人之间以及机器人与障碍物之间需保持安全距离 d_{safe} , 即任意机器人 i 需分别满足^[32]

$$\|P_i - P_j\| \geq d_{\text{safe}}, \exists j \in \mathcal{N}_i \quad (6)$$

$$\|P_i - P_{k, \text{obs}}\| \geq d_{\text{safe}}, \exists j \in \mathcal{N}_i \quad (7)$$

2) 连通性约束。在集中式训练阶段, 由于专家控制器使用全局信息进行规划, 且目标围捕队型在设计时已保证了全局连通性, 因此不额外引入连通性保持机制。在分布式执行阶段, 为每个机器人引入一个临界阈值 r_c , 要求机器人 i 至少与一个邻居

保持连通, 表示为

$$\|P_i - P_j\| \leq r_c, \exists j \in \mathcal{N}_i \quad (8)$$

式中: r_c 通常设置为感知半径 r_s , 确保至少存在一个可感知的邻居。

当机器人与最近队友的距离大于 r_c 时, 则不再执行神经网络输出的指令, 转为对最近队友的追踪指令, 直至该条件不再成立。该设定保证了机器人的局部连通性, 全局连通性则通过对神经网络对专家算法的模仿隐式维持。此外, 当机器人由于外部扰动、感知误差及通信断连等原因断开与群体的连接时, 则依据对机器人群体最近的感知信息进行追踪与搜索, 直至重新连接。

2 端到端模仿学习的控制算法

端到端模仿学习的免通信多机器人分布式包围控制算法通过神经网络学习从局部感知信息到机器人运动控制的映射关系, 在已知全局信息条件下生成最优包围轨迹, 构建端到端神经网络, 以局部感知信息为输入, 通过行为克隆学习专家策略, 实现多机器人对目标的有效包围。算法原理示意图如图 1 所示。

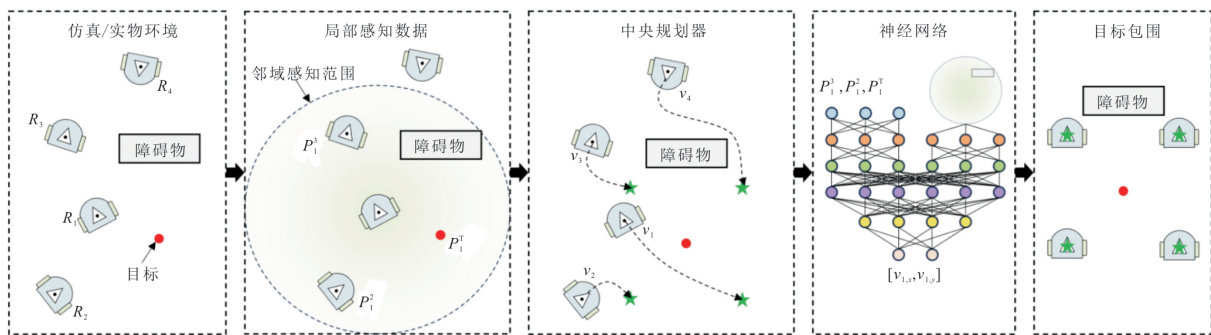


图 1 算法原理示意图

2.1 端到端模仿学习框架

在多机器人包围控制任务中, 端到端架构展现出独特的适用性。传统方法通常将包围控制分解为目标检测、相对定位、轨迹规划和运动控制等多个子任务, 每个子任务都需要精心设计算法和调整参数。然而, 端到端方法可以直接学习局部感知信息, 如相邻机器人位置、障碍物分布及目标方向等运动控制指令的直接映射, 极大简化了系统设计^[33]。

现有多机器人端到端学习方法主要采用图神经网络 (Graph Neural Networks, GNN) 架构处理机器人间通信。例如, Jiang 等^[34] 使用卷积神经网络提取局部特征后通过 GNN 进行单跳信息聚合,

实现了三角编队控制; Li 等^[35] 的去中心化路径规划方法通过 GNN 自动合成局部通信策略。这些方法虽然实现了端到端学习, 但仍需要相邻机器人间的显式信息交换。与上述方法不同, 所提算法采用完全免通信的端到端架构, 机器人仅通过被动感知获取邻居位置信息, 而不进行主动信息交换。为实现免通信下的有效协调, 所提算法设计了双输入融合网络, 将相对几何特征 (目标和邻居位置) 与环境特征 (虚拟激光雷达扫描) 有机结合, 使网络能够在局部信息下隐式学习全局协调策略。

端到端模仿学习框架直接建立从原始感知数据到控制指令的映射关系, 消除中间处理环节。在多机器人系统中, 对于机器人 i , 其控制策略表示为

$$u_i = \pi_{\theta}(\mathbf{Z}_i, \Theta) \quad (9)$$

式中: π_{θ} 表示参数化的深度神经网络, Θ 为网络参数; $\mathbf{Z}_i = (P_i(t), \mathcal{O}_i)$ 为局部观测向量。

训练完成后, 每个机器人装载相同的端到端控制网络 π_{θ} , 仅依靠 $\mathbf{Z}_i$ 进行决策^[36]。该框架具有完全分布式特性, 所有信息均可通过机载传感器和邻域通信获取, 无需全局信息或机器人间显式通信, 满足通信受限环境的应用需求。

2.2 专家算法

模仿学习作为一种高效的端到端学习范式, 通过学习专家演示数据训练控制策略, 避免了强化学习中奖励函数设计困难和大量试错的问题^[37]。在多机器人系统中, 将已知全局信息的集中式最优控制器作为专家, 生成大量演示数据, 然后训练端到端神经网络学习这些专家行为。这种方法将全局最优决策知识有效地迁移到仅依赖局部信息的分布式控制器中, 实现了从集中到分布的转换。

模仿学习的关键在于获取高质量的专家示范数据。专家控制器的设计需同时兼顾包围队形质量、控制策略可解释性与计算实时性 3 个方面。为此, 所提算法采用分层架构: 匈牙利算法用于实现机器人到目标点的最优分配; A^* 与人工势场法用于生成避障路径; 比例反馈控制用于实现轨迹跟踪, 确保生成的示范轨迹同时满足队形约束、最优性和安全性要求这种经典控制方法的组合具有明确的物理意义, 便于神经网络学习其行为模式。

2.2.1 队形规划层

圆形包围编队因其对称性和均匀性成为理想的包围构型。根据目标当前位置 P_T 和系统机器人数量 N , 在目标周围生成与机器人数量相同的 N 个期望包围位置, 实时生成期望包围位置集合^[38]。设定保持距离 r_d 作为包围半径, 各期望位置按圆周均匀分布, 角度间隔 $\Delta\theta = 2\pi/N$ 。第 m 个期望位置计算表达式为

$$P_m^{\text{exp}} = P_T + r_d [\cos(\theta_m), \sin(\theta_m)]^T \quad (10)$$

这种几何化队形设计具有尺度不变性, 当机器人数量变化时, 仅需调整角度间隔即可适应, 无需重新设计控制律。对于大规模系统 ($N > 8$), 专家算法支持多层同心圆编队扩展, 层间距离设为 $r_{\text{layer}} = r_d \sqrt{3}/2$, 确保不同圈层机器人互不干涉。圆形编队的旋转对称性使得各机器人的控制策略具有一致性, 降低了学习难度。

2.2.2 任务分配层

期望位置生成后, 需解决“哪个机器人前往哪

个位置”的指派问题, 以及最小化总移动距离的最优分配问题。因此, 采用匈牙利算法^[39]求解最优匹配问题, 即最小化机器人与其目标点的距离之和, 表示为

$$M^* = \arg \min_M \sum_{i=1}^N \|P_i - P^{\text{exp}}(M[i])\| \quad (11)$$

式中: M 为分配向量; $M[i]$ 表示机器人 i 分配到的目标点编号。

采用匈牙利算法实现一对一的目标点分配, 避免多机器人对同一目标点的重复指派。同时, 在代价函数约束下倾向于为机器人分配距离较近且位于其当前运动方向前方的目标点, 从分配层面降低路径交叉冲突的发生概率, 使各机器人以较小的运动代价到达各自的分配位置。同时, 分配结果实时更新, 当目标移动或机器人数量变化时, 算法自动重新分配, 保证队形动态适应性。机器人包围任务分配示意图如图 2 所示。

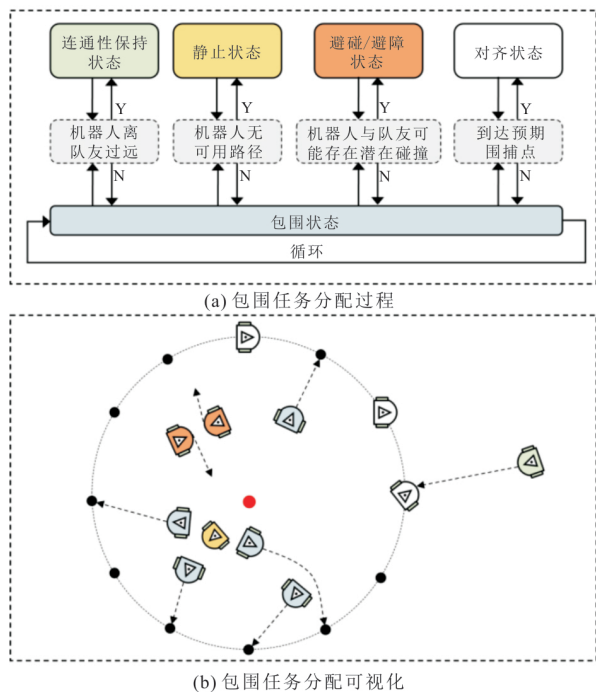


图 2 机器人包围任务分配示意图

2.2.3 轨迹跟踪层

考虑到全向轮机器人平台具有全方向移动能力和零转弯半径的优势, 拥有更高的运动灵活性, 因此采用简化的控制策略实现高效的轨迹跟踪性能^[40]。将人工势场法^[41]的导向思想转化为可执行的速度指令, 引力势场提供从当前位置 P_i 指向期望位置 P_i^{exp} 的吸引力和期望航向角, 为专家算法提供全局导向信息。

采用比例反馈控制律^[42]定义位置误差 $\Delta P_i = P_i^{\text{exp}} - P_i$ 和航向误差 $\Delta\theta_i = \theta_i^{\text{exp}} - \theta_i$, 其中 θ_i^{exp} 为期望

航向角,由目标航向角 θ_{obs} 确定,机器人 i 的控制律表达式为

$$\begin{cases} v_i = K_v \cdot \Delta P_i \\ \omega_i = K_\theta \cdot \Delta \theta_i \end{cases} \quad (12)$$

式中: ω_i 为机器人 i 的角速度; K_v 和 K_θ 分别为速度和角速度反馈增益。

在复杂障碍物环境中,纯人工势场法容易陷入局部极小点或在狭窄通道处产生振荡导致机器人卡死^[43]。为此,采用 A* 算法作为全局路径规划器为机器人提供从初始位置到期望包围位置的无碰撞路径。A* 算法路径规划示意图如图 3 所示,该混合架构结合了 A* 算法的全局最优性和比例控制的实时响应能力,在保证无碰撞的同时实现高效轨迹跟踪。在规划路径时,A* 算法还会使用线性观测器,对队友的位置进行修正与补偿,并基于此优化生成的路径。

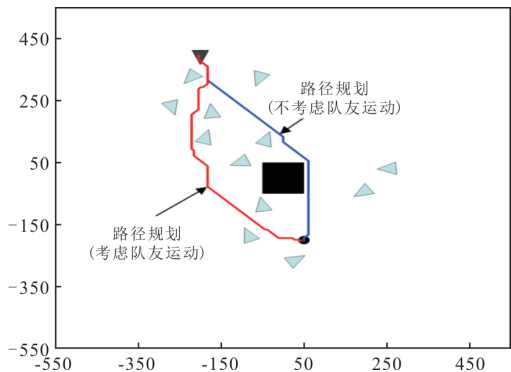


图 3 A* 算法路径规划示意图

专家算法的优势在于能够利用全局信息,如所有机器人位置、目标轨迹、完整地图等进行集中式优化决策。通过匈牙利算法进行最优目标点分配,A* 算法规划无碰撞全局路径,人工势场法提供实时避障和吸引导向,这些算法的协同工作保证了专家轨迹的高质量,为端到端神经网络学习提供丰富、高质量的专家示范样本。

2.3 神经网络

采用深度神经网络(Deep Neural Network,DNN)作为端到端控制器,直接建立从机器人局部观测量到速度控制指令的映射关系。不同于传统模块化方法将感知、规划及控制分离处理,端到端架构通过单一网络实现从原始传感器数据到执行指令的直接映射,避免了多模块串联导致的误差累积问题。

神经网络采用双输入架构实现从局部感知到运动控制的端到端映射。网络包含两路输入:局部几何特征集合 $\mathcal{N}_i$,其前 2 个元素为目标点相对于自身的位置偏差,后续元素为邻居机器人相对于自身

的位置信息;环境特征集合 $\mathcal{O}_i$ 为基于激光雷达点云数据生成的障碍物分布信息。两路输入经神经网络处理后输出二维运动控制执行指令 u_i ,指导机器人的速度与航向。

2.3.1 双输入融合架构

考虑到多机器人包围控制任务需要同时处理邻居协同信息和环境障碍信息,设计双输入融合网络架构。该架构将几何特征(机器人相对位置)和环境特征分别编码后融合,充分利用不同模态信息的互补性。双输入融合架构接收输入信息分为局部几何特征向量和环境特征向量两类。

1)局部几何特征向量 $\mathbf{X}_{\text{fea}} \in \mathcal{N}_i(t)$ 包含目标相对位置和感知范围内所有邻居机器人的相对位置信息,采用拉平方式组织。机器人 i 的局部几何特征向量表达式为

$$\mathbf{X}_{i,\text{fea}} =$$

$$[\Delta x_{i,T}, \Delta y_{i,T}, \Delta x_{i,1}, \Delta y_{i,1}, \dots, \Delta x_{i,j}, \Delta y_{i,j}]^T \quad (13)$$

式中: $\Delta x_{i,T}, \Delta y_{i,T}$ 为目标相对机器人 i 的位置偏移量; $\Delta x_{i,j}, \Delta y_{i,j}$ 表示邻居机器人 j 相对于机器人 i 的位置偏移量。

这种相对坐标表示使网络具有平移不变性,即控制策略不依赖于全局坐标系的选择。前两维度始终保留给目标信息,确保目标跟踪的优先级;后续维度根据实际感知到的邻居数量动态填充,未感知到邻居时对应位置填零。局部几何特征向量的最大维度根据可能感知的最大邻居数量进行预设。相比直接使用全局坐标,相对位置编码能够使网络更好地学习机器人间的协同模式。

2)环境特征向量 $\mathbf{X}_{\text{map}} \in \mathcal{O}_i(t)$ 为感知周围障碍物分布,采用虚拟激光雷达扫描机制生成环境特征向量。通过在机器人周围生成预定义的采样点集合,测量每个方向上的障碍物距离,形成类似激光雷达的环境感知数据。这种表示方法类似于真实激光雷达传感器的工作原理,使网络能够感知复杂几何形状的障碍物分布。相比直接使用原始地图图像,激光雷达特征具有更紧凑的表示和更好的旋转不变性^[44]。

局部几何特征向量 $\mathbf{X}_{\text{fea}}$ 与环境特征向量 $\mathbf{X}_{\text{map}}$ 分别使用多层感知机(全连接层+ReLU 激活层)进行数据预处理,分别生成尺寸为 $[1, 32]$ 的张量;之后,将两项张量合并,生成尺寸为 $[1, 64]$ 的张量。最后,再使用一个多层感知机,生成尺寸为 $[1, 2]$ 的张量,将该张量作为神经网络的输出。

2.3.2 网络输出设计

网络输出为二维速度控制指令,直接作用于

全向轮机器人的运动执行,即

$$u_i = (v_{i,x}, v_{i,y}) = \pi_{\Theta}(X_{i, \text{fea}}, X_{i, \text{map}}) \quad (14)$$

相比输出转向角和速度标量的方式,直接输出笛卡尔速度分量更适合全向运动平台。作为全向轮机器人的运动执行指令,该网络输出设计遵循端到端学习的核心理念,即从原始感知数据到最终控制指令的直接映射,省去了传统方法中轨迹规划、路径跟踪等中间处理环节。神经网络通过学习专家演示数据,隐式地将全局最优决策知识编码到局部感知与控制指令之间的映射关系中。

2.3.3 训练方法

神经网络的训练方法采用监督学习框架下的模仿学习方式。模仿学习通过观察专家演示来学习控制策略,相比强化学习具有样本效率高、训练稳定、易于引入先验知识等优点^[45]。在多机器人系统中,模仿学习能够有效地将全局最优控制器的知识迁移到分布式控制器中^[46-47]。

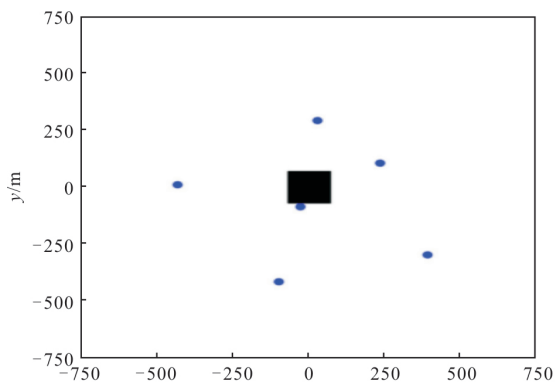
模仿学习的目标^[48]是训练一个仅依赖局部观测的神经网络策略,使其能够在不进行显式通信的条件下复现专家行为。这种从集中式到分布式、从全局到局部的知识迁移是本研究的核心创新。通过学习专家在大量不同场景下的决策模式,神经网络能够隐式地学会协同策略,而无需显式的通信协议设计。

2.4 性能指标

理想情况下,机器人 i 应位于以目标为中心、半径为 r_d 的圆周上,且相邻机器人均匀分布。为了定量分析包围过程,使用指标 $D_{\text{en}}(t)$ 表示 t 时刻机器人与其包围点之间距离之和,反映整体队形误差, $D_{\text{max}}(t)$ 表示 t 时刻机器人与其包围点之间的最大值,反映当前队形中误差最大的机器人状态。 $D_{\text{en}}(t)$ 和 $D_{\text{max}}(t)$ 的表达式分别为

$$D_{\text{en}}(t) = \sum_{i=1}^N \|P_i(t) - P_i^{\text{exp}}(t)\|_2 \quad (15)$$

$$D_{\text{max}}(t) = \max_{i \in \{1, \dots, N\}} \|P_i(t) - P_i^{\text{exp}}\|_2 \quad (16)$$



(a) $t=0$ s

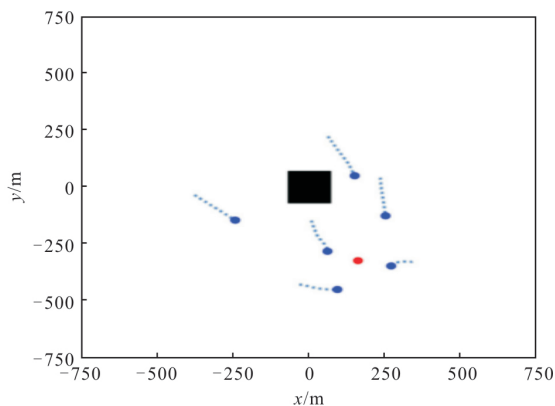
式中:当 $D_{\text{en}}(t)$ 值与 $D_{\text{max}}(t)$ 值较小时,表示机器人编队更接近期望的包围构型。

3 实验结果与分析

在正式仿真实验前,首先对专家控制器中任务分配模块的计算效率进行测试,并完成端到端神经网络的离线训练。任务分配效率测试平台配置为Intel Core i7 处理器和16 GB RAM。针对50个机器人的任务分配问题,采用匈牙利算法独立重复计算1 000次,平均单次计算时间约为0.20 ms,表明该模块能够在毫秒级完成任务分配。神经网络训练阶段采用Adam优化器,最大训练轮数设置为50,批大小设置为256,初始学习率设置为0.001,并在每轮训练前随机打乱样本。完成上述计算效率测试和网络训练后,进一步开展不同机器人规模、目标运动状态及障碍物环境下的仿真实验。

为验证端到端模仿学习算法在多机器人包围控制任务中的有效性和适应性,在多种典型场景下进行仿真实验,全面测试所提算法在不同环境复杂度和机器人规模下的性能表现。在这些仿真测试中,设计环境大小为 $1\,500 \times 1\,500 \text{ m}^2$ 、机器人的速度为5 m/s、目标速度为2.5 m/s、机器人的感知距离 $r_s=350 \text{ m}$ 、第一层围捕半径 $r_d=130 \text{ m}$ 、后续多层间的距离为86 m、最大围捕时间300 s, K_v 和 K_θ 均设为2。

选取6个机器人,机器人静态围捕仿真过程如图4所示。可以观察到,机器人从随机初始位置出发,逐渐向目标区域运动,并在100 s后开始形成包围构型;至150 s时,机器人已形成稳定的包围队形,完成围捕任务。该场景验证了端到端神经网络策略在小规模团队下能够有效克隆学习专家算法的包围行为,实现了从随机初始位置到稳定包围队形的自主转换。



(b) $t=50$ s

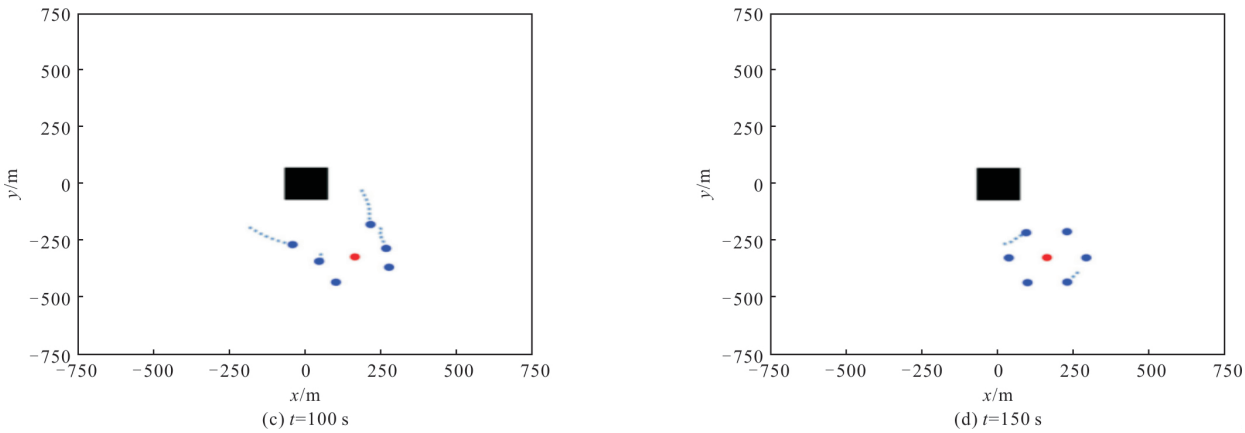


图 4 6 个机器人静态围捕仿真过程

6 个机器人静态围捕任务的性能指标如图 5 所示。由图 5 可知,围捕开始后,机器人总距离之和 D_{en} 和与最大队形误差 D_{max} 均呈现快速下降趋势,并在约 135 s 时同时降至 0,之后基本保持稳定。这表明机器人能够在较短时间内完成对目标的围捕,并逐步调整至理想的包围构型,验证了所提算法在静态目标围捕任务中的有效性。为了进一步验证所提算法在多机器人包围控制任务中的适应性、稳定性与可扩展性,在基础实验上引入更高难度的测试场景,并分别开展 6 个机器人动态围捕实验与 15 个机器人静态围捕仿真实验,结果分别如图 6 和图 7 所示。其中,动态围捕场景旨在考察算法在目标持续移动、队形不断调整情况下的实时规划能力与编队保持性能;而静态围捕场景则主要用于评估算

法在大规模机器人数量下的空间协同能力与收敛效率。

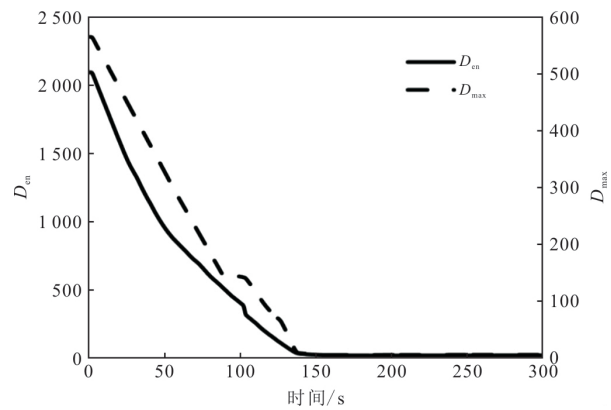


图 5 6 个机器人静态围捕性能指标

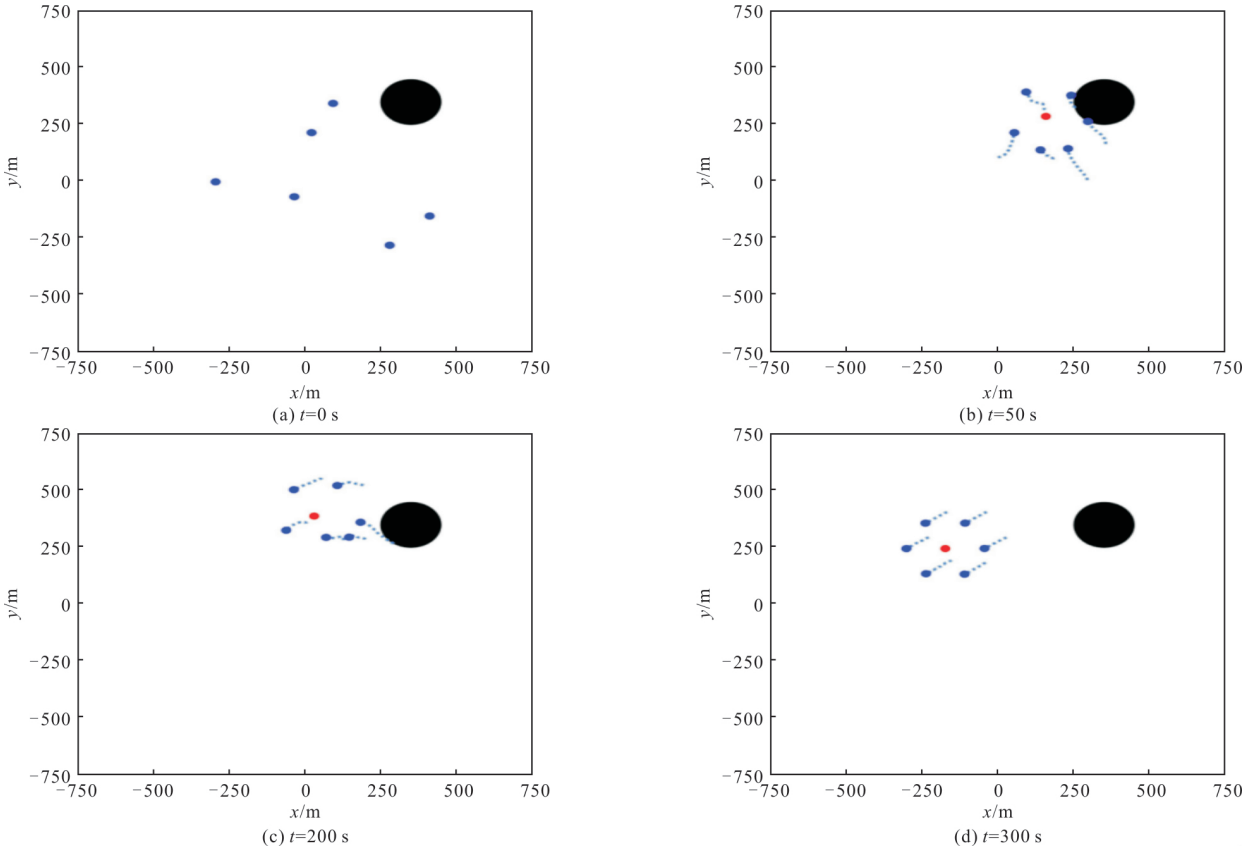


图 6 6 个机器人动态围捕仿真结果

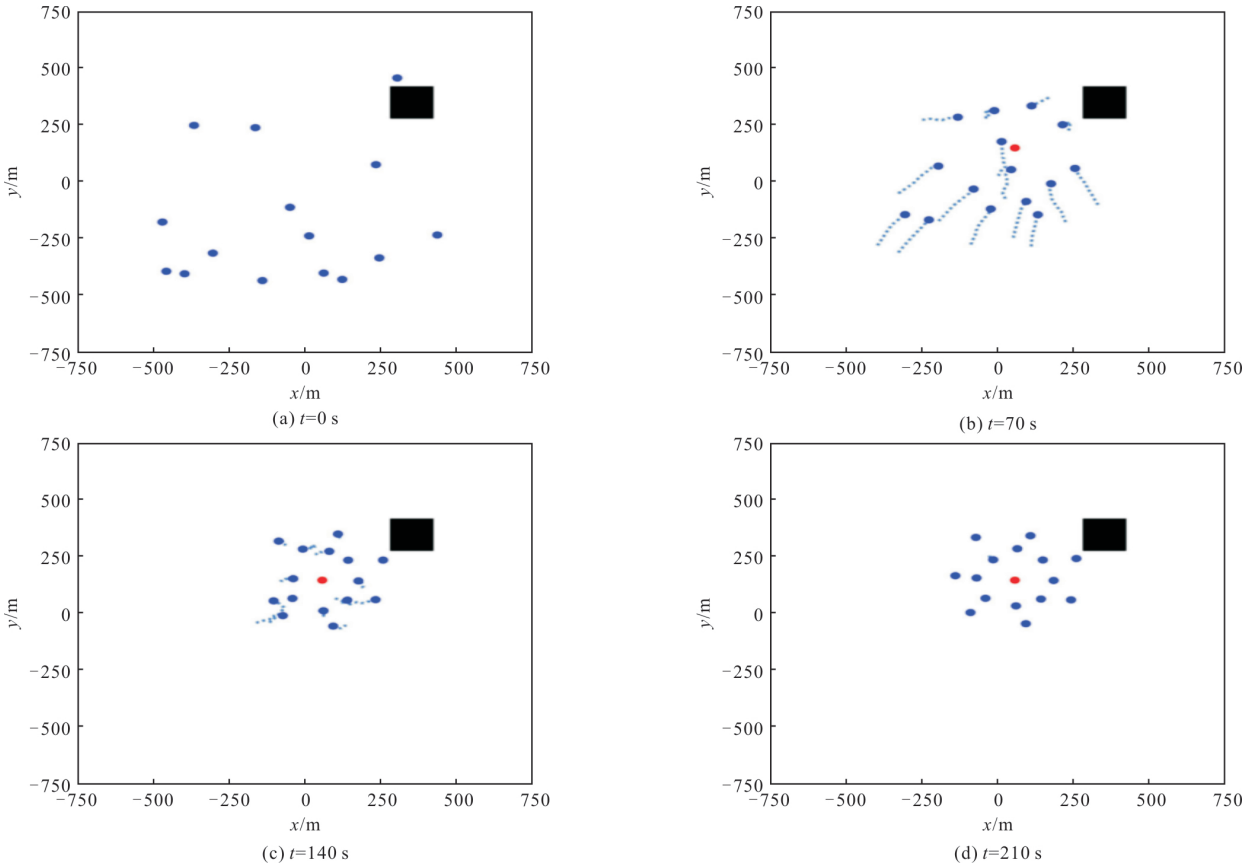


图 7 15 个机器人静态围捕仿真结果

由图 6 和图 7 可知,通过在不同包围场景环境下的训练,所提算法展现出优异的泛化能力。对于环境的泛化能力:训练好的控制器在面对全新的环境、全新的障碍物分布和队友分布时,能够基于积累的经验,自适应地实现避碰、避障等运动决策;对于多机器人系统规模的泛化能力:当机器人数量从 6 个增加到 15 个时,使用规模为 6,15 的多机器人系统的仿真数据训练的神经网络控制器,可以无障碍迁移至规模介于二者之间的多机器人群体,而不需要(或只需要少量)重新训练。仿真结果验证了基于端到端模仿学习网络架构对可变团队规模的强泛化能力。

6 个机器人动态围捕任务的性能指标和 15 个机器人静态围捕性能指标分别如图 8 和图 9 所示。

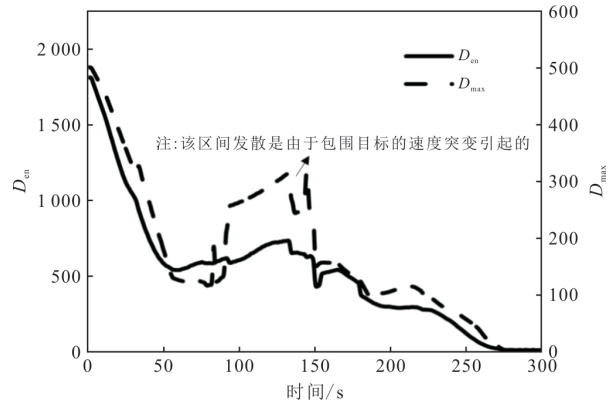


图 8 6 机器人动态围捕性能指标

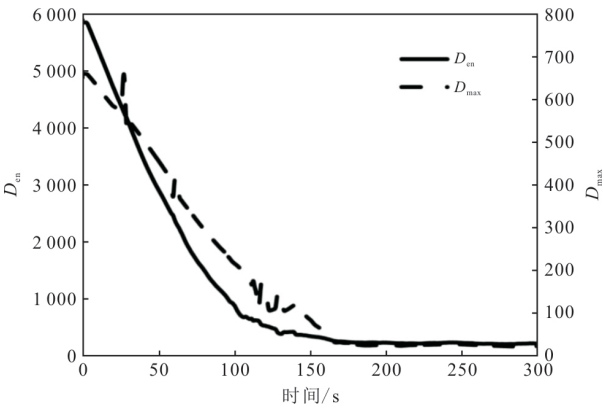


图 9 15 个机器人静态围捕性能指标

由图 8 和图 9 可观察得出,随着任务难度的加大,系统仍能保持一个良好的包围构型和较快的收敛时间,特别是静态围捕场景下,机器人数量的增加对系统的收敛速度和包围构型影响较小,展现出了良好的适应性和可扩展性。

为验证所提算法在机器人包围任务中的性能,分别使用基于模仿学习控制器的所提算法和基于密度交互的光滑粒子流体动力学(Smoothed Particle Hydrodynamics, SPH)算法^[48]在两种不同初始配置场景下开展 15 个机器人包围任务实验,结果如表 1 所示,表中结果为两种初始配置下测试结果的均值±标准差。

表 1 两种算法的机器人包围任务实验结果

算法	平均收敛步数/步	最终队形误差
所提算法	210.5 ± 10.6	80.0 ± 21.9
SPH	308.5 ± 37.5	94.3 ± 13.2

由表 1 可以看出,所提算法的平均收敛时间为 210.5 ± 10.6 步,SPH 算法为 308.5 ± 37.5 步,相较 SPH 算法,所提算法的收敛速度提升约 31.8%。在队形精度方面,所提算法平均最终误差为 80.0 ± 21.9 ,略优于 SPH 算法的 94.3 ± 13.2 。两种算法在测试中均未发生碰撞,安全性相当,结果证明了所提算法在收敛性上的优势。

4 结语

端到端模仿学习的免通信多机器人分布式包围控制算法通过构建仅依赖局部感知的分布式控制框架,显著降低了对通信系统的依赖。同时,端到端神经网络采用注意力机制处理可变数量的邻居信息,实现了从观测到控制的直接映射。模仿学习方法通过克隆集中式专家算法的离线训练方式,成功模仿了全局最优控制行为,有效避免了深度强化学习中的样本低效和训练不稳定问题。仿真结果表明,所提算法能够在 6~15 个机器人场景中完成静态和动态目标包围,实现了端到端学习架构、模仿学习方法、免通信协调和多机器人包围控制 4 个关键特征的结合,为通信受限环境下的多机器人包围任务提供了新的解决方案。

所提算法的性能依赖于专家数据的质量与分布覆盖范围,在实际部署时需考虑仿真-实物差距的影响,适用于邻居数量动态变化、通信受限且对实时性要求较高的分布式多机器人协同场景。未来工作将进一步研究该算法在真实多机器人平台中的部署问题,探索在多目标协同、异构机器人系统及三维空间任务中的扩展应用,并研究在线增量学习机制,以提升算法在实际复杂场景中的可行性与泛化能力。

参 考 文 献

[1] Tanner H G, Jadbabaie A, Pappas G J. Stable flocking of mobile agents, part I: Fixed topology[C]// 42nd IEEE International Conference on Decision and Control. Maui: IEEE, 2003: 2010-2015.

[2] Navarro I, Matia F. A survey of collective movement of mobile robots[J]. International Journal of Advanced Robotic Systems, 2013, 10: 73.

[3] 胡子谅. 无人船协同目标包围控制方法[J]. 计算机与

网络, 2024, 50(6): 549-554.

Hu Ziliang. Unmanned surface vehicle cooperative target encirclement control method[J]. Computer & Network, 2024, 50(6): 549-554.

[4] 贾永楠, 李擎. 多机器人编队控制研究进展[J]. 工程科学学报, 2018, 40(8): 893-900.

Jia Yongnan, Li Qing. Research development of multi-robot formation control[J]. Chinese Journal of Engineering, 2018, 40(8): 893-900.

[5] Cortés J, Egerstedt M. Coordinated control of multi-robot systems: A survey[J]. SICE Journal of Control, Measurement, and System Integration, 2017, 10(6): 495-503.

[6] 张辉, 郑晓娟. 速度未知的无人机群包围及碰撞避免控制[J]. 兵器装备工程学报, 2022, 43(10): 72-76.

Zhang Hui, Zheng Xiaojuan. UAVs fencing and collision avoided control with unknown velocity[J]. Journal of Ordnance Equipment Engineering, 2022, 43(10): 72-76.

[7] 路张昕, 关永强, 许世阳. 基于零和博弈的多智能体二分包围控制研究[J]. 系统科学与数学, 2026, 46(5): 1413-1430.

Lu Zhangxin, Guan Yongqiang, Xu Shiyang. Research on multi-agent bipartite containment control based on zero-sum game[J]. Journal of Systems Science and Mathematical Sciences, 2026, 46(5): 1413-1430.

[8] Dong Z, Jia S, Zhang C, et al. Deep manifold learning of symmetric positive definite matrices with application to face recognition[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 4009-4015.

[9] Nguyen T T, Nguyen N D, Nahavandi S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications[J]. IEEE Transactions on Cybernetics, 2020, 50(9): 3826-3839.

[10] Gronauer S, Diepold K. Multi-agent deep reinforcement learning: A survey[J]. Artificial Intelligence Review, 2022, 55(2): 895-943.

[11] 陈佳盼, 郑敏华. 基于深度强化学习的机器人操作行为研究综述[J]. 机器人, 2022, 44(2): 236-256.

Chen Jiapan, Zheng Minhua. A survey of robot manipulation behavior research based on deep reinforcement learning[J]. Robot, 2022, 44(2): 236-256.

[12] 曾毓凌, 郝宇清, 于颖, 等. 基于深度强化学习的多无人车系统编队控制[J]. 力学学报, 2024, 56(2): 460-471.

Zeng Yuling, Hao Yuqing, Yu Ying, et al. Formation control for multi-unmanned vehicles via deep reinforcement learning[J]. Chinese Journal of Theoretical and Applied Mechanics, 2024, 56(2): 460-471.

[13] Wei Q L, Li Y G, Zhang J, et al. VGN: Value de-

- composition with graph attention networks for multi-agent reinforcement learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(1): 182-195.
- [14] Rashid T, Samvelyan M, Schroeder C, et al. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning[C]//35th International Conference on Machine Learning. [S. l.]: PMLR, 2018: 4295-4304.
- [15] Du J B, Kong Z W, Sun A J, et al. MADDPG-based joint service placement and task offloading in MEC empowered air-ground integrated networks[J]. *IEEE Internet of Things Journal*, 2024, 11(6): 10600-10615.
- [16] Foerster J, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 2974-2982.
- [17] Sukhbaatar S, Szlam A, Fergus R. Learning multi-agent communication with backpropagation[C]//30th International Conference on Neural Information Processing Systems. [S. l.]: ACM, 2016: 2252-2260.
- [18] Das A, Gervet T, Romoff J, et al. TarMAC: Targeted multi-agent communication[C]//36th International Conference on Machine Learning. [S. l.]: PMLR, 2019: 1538-1546.
- [19] Oroojlooy A, Hajinezhad D. A review of cooperative multi-agent deep reinforcement learning[J]. *Applied Intelligence*, 2023, 53(11): 13677-13722.
- [20] Kober J, Bagnell J A, Peters J. Reinforcement learning in robotics: A survey[J]. *The International Journal of Robotics Research*, 2013, 32(11): 1238-1274.
- [21] Zare M, Kebria P M, Khosravi A, et al. A survey of imitation learning: Algorithms, recent developments, and challenges[J]. *IEEE Transactions on Cybernetics*, 2024, 54(12): 7173-7186.
- [22] Holecek M, Hoskovec A, Baudis P, et al. Table understanding in structured documents[C]//2019 International Conference on Document Analysis and Recognition Workshops. Sydney: IEEE, 2019: 158-164.
- [23] Wang R, Lyu M, Zhang J. A multi-robot collaborative exploration method based on deep reinforcement learning and knowledge distillation[J]. *Mathematics*, 2025, 13(1): 173.
- [24] 范永青, 李嘉睦. 基于 RBF 神经网络的不确定性非线性二阶 MASs 编队控制[J]. *西安邮电大学学报*, 2025, 30(5): 109-118.
- Fan Yongqing, Li Jiamu. Formation control of uncertain nonlinear second-order MASs based on RBF neural network[J]. *Journal of Xi'an University of Posts and Telecommunications*, 2025, 30(5): 109-118.
- [25] Shariq I, Fei S. Actor-attention-critic for multi-agent reinforcement learning[C]//36th International Conference on Machine Learning. Long Beach: ICML, 2019: 2961-2970.
- [26] Zhang T L, Liu Z, Pu Z Q, et al. Multi-target encirclement with collision avoidance via deep reinforcement learning using relational graphs[C]//2022 International Conference on Robotics and Automation. Philadelphia: IEEE, 2022: 8794-8800.
- [27] Lo H W, Pai C J, Deveci M. A multi-objective model for integrated supplier order allocation and supply chain network transportation planning decision-making[J]. *Information Sciences*, 2025, 689: 121487.
- [28] 赵桥, 李博, 项融融. 多动态目标的多机器人协同环航控制[J]. *计算机测量与控制*, 2023, 31(1): 71-78.
- Zhao Qiao, Li Bo, Xiang Rongrong. Cooperative circumnavigation control for robots and multiple dynamic targets[J]. *Computer Measurement & Control*, 2023, 31(1): 71-78.
- [29] Liu F, Yuan S H, Meng W, et al. Multiple noncooperative targets encirclement by relative distance-based positioning and neural antisynchronization control[J]. *IEEE Transactions on Industrial Electronics*, 2024, 71(2): 1675-1685.
- [30] Sharon G, Stern R, Felner A, et al. Conflict-based search for optimal multi-agent pathfinding[J]. *Artificial Intelligence*, 2015, 219: 40-66.
- [31] Van den Berg J, Guy S J, Lin M, et al. Reciprocal n-body collision avoidance[C]//Robotics Research: The 14th International Symposium, ISRR 2009. Berlin: Springer, 2011: 3-19.
- [32] Zhang F, Yang G H, Dimirovski G M. Optimal encirclement control along arbitrary orbits for vehicles considering safety constraints: A safe learning approach[J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(10): 6672-6686.
- [33] Ji M, Egerstedt M. Distributed coordination control of multiagent systems while preserving connectedness[J]. *IEEE Transactions on Robotics*, 2007, 23(4): 693-703.
- [34] Jiang C, Huang X C, Guo Y. End-to-end decentralized formation control using a graph neural network-based learning method[J]. *Frontiers in Robotics and AI*, 2023, 10: 1285412.
- [35] Li Q B, Gama F, Ribeiro A, et al. Graph neural networks for decentralized multi-robot path planning[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas: IEEE,

2020; 11785-11792.

[36] Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[C]//31st International Conference on Neural Information Processing Systems. Long Beach:ACM, 2017; 6382-6393.

[37] Oroojlooy A, Hajinezhad D. A review of cooperative multi-agent deep reinforcement learning[J]. Applied Intelligence, 2023, 53(11): 13677-13722.

[38] 玄立伟,刘明雍,何国元,等. 信息素扰动的多目标蚁群优化动态路径规划算法[J]. 西安邮电大学学报, 2026, 31(3): 110-118.

Xuan Liwei, Liu Mingyong, He Guoyuan, et al. Dynamic path planning algorithm of multi-objective ant colony optimization with pheromone disturbance[J]. Journal of Xi'an University of Posts and Telecommunications, 2026, 31(3): 110-118.

[39] 柳毅, 佟明安. 匈牙利算法在多目标分配中的应用[J]. 火力与指挥控制, 2002, 27(4): 34-37.

Liu Yi, Tong Ming'an. An application of Hungarian algorithm to the multi-target assignment[J]. Fire Control & Command Control, 2002, 27(4): 34-37.

[40] Kamel M, Burri M, Siegwart R. Linear vs nonlinear MPC for trajectory tracking applied to rotary wing micro aerial vehicles[C]//20th World Congress of the International Federation of Automatic Control (IFAC). Toulouse: IFAC-PapersOnLine, 2017, 50(1): 3463-3469.

[41] Ge S S, Cui Y J. Dynamic motion planning for mobile robots using potential field method[J]. Autonomous Robots, 2002, 13(3): 207-222.

[42] 寇立伟, 项基. 基于输出反馈线性化的多移动机器人目标包围控制[J]. 自动化学报, 2022, 48(5): 1285-1291.

Kou Liwei, Xiang Ji. Target fencing control of multiple mobile robots using output feedback linearization[J]. Acta Automatica Sinica, 2022, 48(5): 1285-1291.

[43] 李珣, 何晶晶, 赵征凡, 等. 基于一致性改进模型的多机器人编队运动规划方法[J]. 西安工程大学学报, 2021, 35(3): 44-53.

Li Xun, He Jingjing, Zhao Zhengfan, et al. Multi-robot formation motion planning method based on improved consensus model[J]. Journal of Xi'an Polytechnic University, 2021, 35(3): 44-53.

[44] Tai L, Paolo G, Liu M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vancouver: IEEE, 2017; 31-36.

[45] Hussein A, Gaber M M, Elyan E, et al. Imitation learning: A survey of learning methods[J]. ACM Computing Surveys, 2018, 50(2): 1-35.

[46] Ross S, Gordon G J, Bagnell J A. A reduction of imitation learning and structured prediction to No-regret online learning[C]//Fourteenth International Conference on Artificial Intelligence and Statistics. [S. l.]: PMLR, 2011: 627-635.

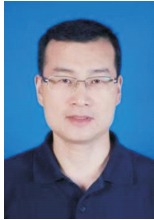
[47] Sartoretti G, Kerr J, Shi Y F, et al. PRIMAL: Path-finding via reinforcement and imitation multi-agent learning[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2378-2385.

[48] Lei X K, Zhang S, Xiang Y L, et al. Self-organized multi-target trapping of swarm robots with density-based interaction[J]. Complex & Intelligent Systems, 2023, 9(5): 5135-5155.

[作者简介]



钟远湘(2001—),男,湖南怀化人,西北工业大学硕士研究生,主要研究方向为多机器人控制、人工智能等。E-mail: zhongyx@mail.nwpu.edu.cn



刘明雍(1971—),男,陕西西安人,博士,西北工业大学教授、博士生导师。主要研究方向为自主导航与控制、蜂拥式机器人集群控制、自主无人水面船等。E-mail: liumingyong@nwpu.edu.cn



闫志强(1996—),男,山西太原人,西北工业大学博士研究生,主要研究方向为计算机视觉、多机器人协同控制等。E-mail: yanzhiqiang@mail.nwpu.edu.cn



玄立伟(1989—),男,吉林德惠人,西北工业大学博士研究生,主要研究方向为群机器人智能控制。E-mail: St0rmer111@163.com