arXiv:2601.03301v1 [cs.MA] 6 Jan 2026

PC^{2}P：通过个性化增强通信和人群感知实现多智能体路径规划

Guotao Li[1,2], Shaoyun Xu[1],Yuexing Hao[1,2†], Yang Wang[1],Yuhui Sun[1,2]

摘要——**结合多智能体强化学习（MARL）的分布式多智能体路径规划（MAPF）已成为重要的研究方向。通过智能体间通信，该方法能够在部分可观测环境中实现实时协同决策。然而，由于协同与感知能力不足，现有方法难以适应多样化的环境条件。为应对这些挑战，我们提出了 PC^{2}P，一种基于 Q学习的多智能体强化学习框架构建的新型分布式多智能体路径规划方法。首先，我们引入了一种基于动态图拓扑的个性化增强通信机制，通过选择、生成和聚合三个阶段的操作，确定交互过程中的核心问题，即“谁”与“什么”。与此同时，我们融入局部人群感知，以丰富智能体的启发式观测；通过整合静态空间约束与动态占用变化，进一步增强模型对有效动作的引导能力。为解决极端死锁问题，我们提出了一种基于区域的死锁破解策略，借助专家指导在受限区域内实现高效协同。实验结果表明，与当前最先进的分布式多智能体路径规划方法相比，PC^{2}P 在多种环境中均取得了更优的性能。消融研究进一步验证了各模块对整体性能的有效性。**

I. 引言

多智能体路径规划（MAPF）旨在为多个智能体规划从起始位置到目标位置的无碰撞路径，同时优化总路径长度或行驶时间[1], [2]。鉴于其在智能仓库 [3],自动驾驶车辆[4], 和工业机器人 [5], 等多个领域得到广泛应用，如何提升规划能力已成为当前的研究热点。当前的多智能体路径规划方法通常可分为集中式和分布式两类范式。集中式算法大多基于启发式搜索进行扩展，例如 ODrM* [6] 和 CBS [7]。这些方法严重依赖预先获取的全局信息，并且在动态环境中需要重新规划，因此限制了其可扩展性和实时性能。

分布式算法使用多智能体强化学习（MARL）解决多智能体路径规划问题，仅依赖有限视野（FOV）内的观测，并通过智能体间通信实现实时协作策略。PRIMAL [8] 是首个结合强化学习（RL）与模仿学习（IL）的去中心化多智能体路径规划框架。DHC [9] 提出了一种多智能体通信网络机制，无需专家指导即可增强不同智能体之间的协同。后续研究（如 [10], [11], 和 [12], ）通过优化通信模块，进一步提升了协作效率。同时，已有若干方法专门针对特定规划场景而开发，包括小规模密集环境 [13], [14] 和大规模环境 [15], [16]。然而，当前的分布式多智能体路径规划算法在不同规模、密度和类型的环境中，规划成功率和解的质量表现不稳定；尤其是在大规模、高密度场景中，性能会急剧下降。主要原因在于环境密度增加会导致智能体更加集中于有限区域。在这种场景下，协作能力对无碰撞路径规划起着决定性作用。但现有通信机制在管理不同智能体之间的交互关系和交互内容方面存在不足，削弱了算法在各种环境中的鲁棒性。此外，在密集区域内，典型的智能体观测在多个时间步之间往往重复或相似，因此无法为协作策略提供有效的感知信息。如果相互避让失败，智能体可能频繁停滞或在狭窄空间内徘徊，久而久之可能造成复杂死锁，最终导致任务失败。

本文提出了 PC^{2}P，一种基于协同与感知增强的新型分布式 MARL-MAPF 方法。首先，我们设计了三阶段的个性化增强通信机制“选择-生成-聚合”，动态选择通信智能体，并结合图注意力机制从中获取定制消息。其次，我们引入人群增强局部观测，将静态空间约束与动态占用变化相结合。这增强了智能体对当前及未来环境状态的感知能力，为实时策略输出提供启发式引导。最后，我们提出了一种基于区域的协同死锁破解策略，利用专家指导解决有限区域内的局部极端死锁。实证结果表明，在各种测试环境中，PC^{2}P 始终优于其他分布式算法。

[1]作者就职于中国科学院微电子研究所，地址：中国北京市朝阳区土城西路北三号 {liguotao24, xushaoyun, haoyuexing, wangyang, sunyuhui}@ime.ac.cn[2]中国科学院大学

[†]通讯作者

II. 相关工作

A. 面向 *MAPF* 的多智能体强化学习

近期研究日益探索将多智能体路径规划任务与多智能体强化学习相结合，开发出创新方法

以增强智能体间的协同。PRIMAL [8] 率先将异步优势演员-评论家（A3C）[17] 与模仿学习（IL）相结合，利用专家指导优化策略网络。在此基础上，PRIMAL 2 [18] 将该方法扩展至终身多智能体路径规划任务。此后，大量研究将通信机制引入多智能体强化学习框架，以提升协作能力。DHC [9] 引入启发式信道引导来指导智能体移动，并通过课程学习 [19] 和分布式优先经验回放缓冲区 [20]，以及图卷积[21] 实现通信。DCC [10] 提出了决策因果通信，利用决策因果单元和请求-响应机制来提升通信效率。PICO [13] 从专家示范中推断智能体之间的隐式优先级关系，并构建动态通信拓扑。MAGAT [15] 通过监督学习构建分布式通信网络。SACHA [22]利用基于启发式的注意力机制，促进视野内智能体之间的协作，并采用图卷积实现高效的信息交换。SCRIMP [11] 使用类似Transformer的架构实现可扩展通信模块 [23]。HELSA [16] 利用分层强化学习分解复杂任务，从而在大规模环境中实现高效规划。CRAMP [14] 通过针对环境的奖励塑形增强密集场景中的智能体协作。EPH [12]提出了三种推理技术，实现了与传统算法的有效整合。然而，这些方法在不同环境中的可扩展性仍然不足，尤其是在大规模、高密度场景下。

B. 多智能体强化学习中的通信机制

利用多智能体强化学习解决多智能体路径规划问题，核心在于掌握协作，而高效通信对于智能体间的协同至关重要。在部分可观测性下实现有效的信息共享，已成为多智能体强化学习的一项关键研究重点 [24]。然而，传统的广播通信受制于较高的计算开销和低效率。为解决这些局限，研究人员提出了许多创新的通信策略。TarMAC [25]采用注意力机制实现选择性消息传输，并通过多轮通信提升效率。I2C [26] 引入决策因果单元以实现选择性通信，并采用请求-响应机制降低通信开销。MAGIC [27] 将图注意力网络（GAT） [28] 同时用作调度器和聚合器，以优化通信目标选择和信息聚合。AC2C [29] 采用两跳通信机制，使智能体能够获取视野之外的远程信息。CACOM [30] 提出了一个两阶段通信框架，可根据上下文感知消息动态调整通信拓扑，并利用Transformer [23] 高效地进行信息聚合。

III. 问题表述

A. 多智能体路径寻找定义

多智能体路径规划问题旨在图 $G = (V, E)$ 上为 k 个智能体计算无碰撞路径，其中 V 为顶点集，E 为边集。每个智能体都从唯一的起始位置 $s_i \in S = \{s_1, s_2, \ldots, s_k\}$ 移动到指定的目标位置 $g_i \in T = \{g_1, g_2, \ldots, g_k\}$。时间被离散为多个时间步，在每个时间步中，每个智能体可以移动到相邻顶点，也可以停留在当前位置。智能体 i的完整路径记为 $\pi_i = (v_1^i, v_2^i, \ldots, v_T^i)$，必须满足三个条件：(1) $v_1^i = s_i$；(2) $v_T^i = g_i$；(3) 对于每个时间步 t，(v_t^i, v_{t+1}^i) 要么是 E 中的一条有效边，要么是静止动作，其中 $v_t^i = v_{t+1}^i$。为确保无冲突规划，需要施加两类约束：顶点冲突（$v_t^i \neq v_t^j$），用于防止智能体同时占据同一顶点；以及边冲突（$(v_t^i, v_{t+1}^i) \neq (v_{t+1}^j, v_t^j)$），用于避免智能体沿同一条边反向穿行。目标是最小化总时间步数 $\sum_{i=1}^{k} T_i$，其中 T_i 表示智能体 i 到达其目标 g_i的首个时间步。

B. 环境设置

根据多智能体路径规划任务定义，我们将环境建模为二维四邻域网格世界，其中每个单元格要么为空，要么为障碍物。 m智能体最初占据m随机选取的空单元格 m 初始位置和 m 目标位置均从回合开始时的空单元格中随机选取，以确保起点与目标位置之间的连通性，并避免二者重叠。在每个时间步，智能体可以向上、向下、向左、向右移动，或保持静止。在移动过程中，如EPH 中所提出的，通过执行 Q值第二高的动作来解决冲突 [12]。本文考虑的是部分可观测环境，其中每个智能体只能观测其 $l \times l$ 视野范围内的环境信息。奖励函数采用 DHC 提出的形式 [9]，，如表I所示，旨在鼓励智能体快速到达目标，同时避免冲突。

表 I：奖励结构

动作	奖励
移动（上/下/左/右）	-0.075
停留（在目标点上，远离目标点）	0, -0.075
碰撞（障碍物/智能体）	-0.5
完成	3

IV. 方法

PC^{2}P通过Q学习驱动的多智能体强化学习，解决部分可观测的分布式多智能体路径规划问题：提取各个智能体的局部观测信息，通过深度融合个性化信息促进智能体间通信，并最终实时生成用于策略选择的Q值。图 1 展示了PC的整体框架^{2}P。第IV-A 节详细介绍个性化增强通信和Q学习训练，第IV-B 节探讨人群增强局部观测，第IV-C节讨论用于解决极端死锁情形的协作式死锁破解策略。

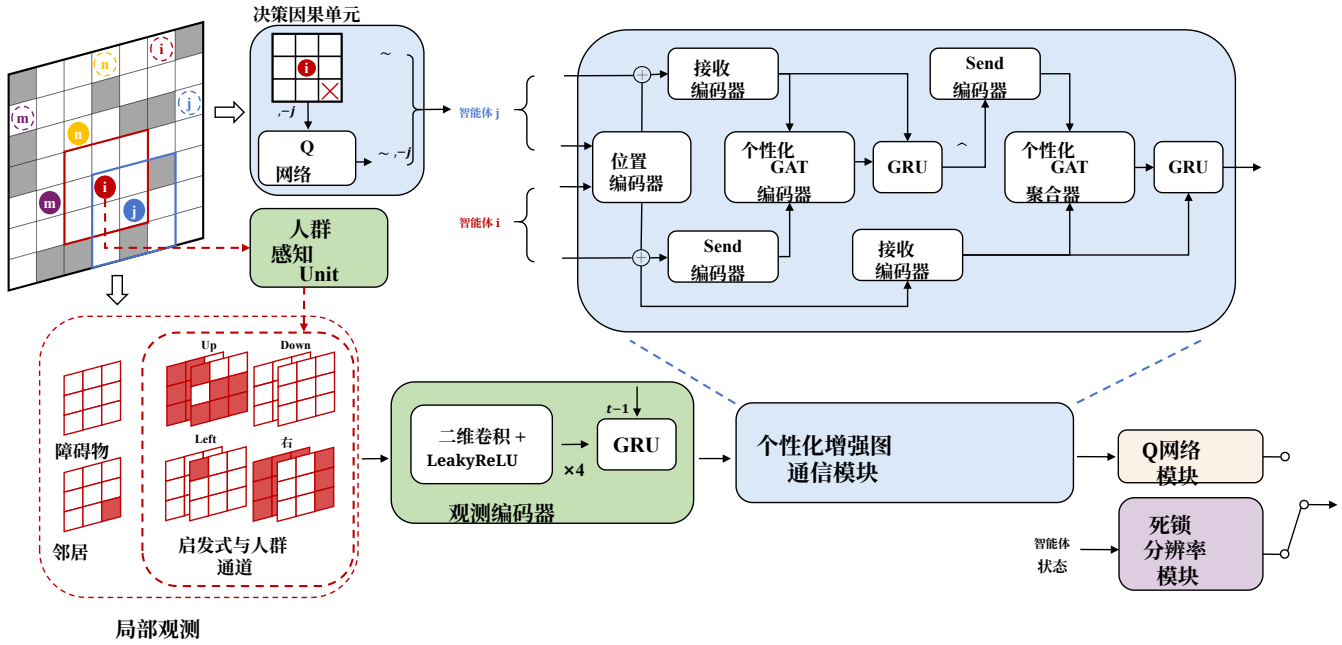


图 1：PC²P 框架由四个组件构成：观测模块（绿色）、通信模块（蓝色）、Q网络模块（粉色）和死锁消解模块（紫色）。智能体收集局部观测（红色虚线框），包括基本信息和人群增强信息，随后由观测编码器进行编码。通信模块选择相关智能体并聚合个性化信息（详见上文），使 Q网络模块能够生成策略。最后，在检测到极端死锁时，PC²P 激活死锁消解模块，输出替代策略。

A. 面向多智能体路径规划的个性化增强通信

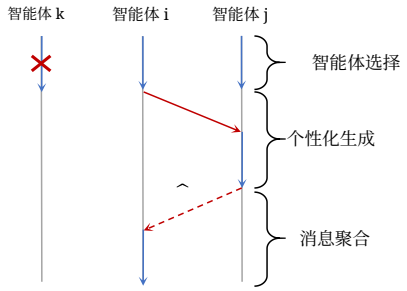


图 2：个性化增强通信机制示意图。（1）智能体 i通过决策因果单元选择智能体 j 进行通信。（2）智能体 i 向智能体 j 发送信息 m_t^i，以生成个性化消息 $\hat{m}_t^{ij}$。（3）智能体 i 将接收到的 $\hat{m}_t^{ij}$ 与自身信息 m_t^i 进行聚合，完成通信过程。

由于智能体状态（例如位置、目标和局部环境）存在差异，智能体的通信需求也大不相同。现有方法通常采用广播或选择性通信机制，以决定与“谁”交互。然而，这些方法忽略了“通信什么”这一问题，即通信内容。我们提出了一种基于图拓扑增强的个性化通信机制。构建基于因果决策的选择性通信，每个智能体根据自身信息发送或接收个性化消息，从而实现定制化通信。该机制包含三个阶段：智能体选择、个性化消息生成和消息聚合，如图2所示。在第一阶段，智能体 i按照 DCC 提出的方法，通过决策因果单元识别其通信伙伴 [10]。通信伙伴集合记为 $\mathcal{C}_i$，定义如下：

$$\mathcal{C}_i = \{j \mid \tilde{a}_t^i \neq \tilde{a}_t^{i,-j}\}_{j \in \mathcal{B}_i}, \tag{1}$$

其中，$\mathcal{B}_i$ 表示局部观测邻域内的智能体，$\tilde{a}_t^i$ 表示智能体 i根据其观测在时刻 t 采取的动作，而 $\tilde{a}^{i,-j}_t$ 表示排除智能体 j 影响时采取的动作。如果发生动作变化，则选择智能体 j 进行通信。在第二阶段，智能体 i 向智能体 j 发送信息 m_t^i，以生成针对智能体 i的个性化消息 $\hat{m}_t^{ij}$。在第三阶段，将个性化消息 $\hat{m}_t^{ij}$ 返回，并与智能体 i的原始信息 m_t^i 进行聚合，从而得到通信后消息 m_t^{ij}。最终，使用 m_t^{ij} 通过 Q网络计算 Q值，并通过 arg max操作确定最终策略a_t^i。

1）个性化增强图通信模块：鉴于智能体间通信关系具有动态特性，我们采用基于图的拓扑，在智能体之间建立实时通信网络。值得注意的是，我们使用图注意力网络（GAT） [28]

来实现消息生成与聚合，并自适应地为随时间变化的通信伙伴分配不同权重，如图1所示。智能体信息由局部观测 m_t 和位置信息p_t组成。以智能体 i 和智能体 j 为例，通过位置编码器将 p_t^i 和 p_t^j 转换为位置向量 l_t^i 和 l_t^j，并分别与局部观测 m_t^i 和 m_t^j 拼接。拼接后的表示分别经过发送编码器 f_s 和接收编码器 f_r 进行预处理。随后，个性化GAT编码器根据观测 m_t^i 和 m_t^j 计算个性化消息$\hat{m}_t^{ij}$。个性化GAT聚合器将智能体 i的信息 m_t^i 与个性化消息$\hat{m}_t^{ij}$ 整合，最终生成通信后消息 m_t^{ij}。上述两个网络组件均采用基于GATv2 [31]的相同结构。为清晰起见，我们以个性化GAT编码器为例进行说明。首先，计算智能体i 与 j之间的注意力系数 e_{ij}：

$$e_{ij} = \mathbf{a}^\top \text{LeakyReLU}\left(\left[f_s(m_t^i) \| f_r(m_t^j)\right]\right), \tag{2}$$

其中，$\mathbf{a}$ 表示线性变换权重向量，$f_s(m_t^i)$表示发送特征向量，$f_r(m_t^j)$表示接收特征向量，而 LeakyReLU 引入非线性特征。随后，将注意力系数归一化为 α_{ij}:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{j\in\mathcal{C}_i}\exp(e_{ij})}. \tag{3}$$

对于每个注意力头，我们通过线性映射 f_p，使用归一化系数 α_{ij} 计算个性化信息 $\hat{m}_t^{ij}$：

$$\hat{m}_t^{ij} = f_p\left(\text{concat}\left(\sum_{j\in\mathcal{C}_i}\alpha_{ij}^h f_r(m_t^j), \forall h\in\mathcal{H}\right)\right), \tag{4}$$

其中，$\mathcal{H}$ 表示注意力头集合，α_{ij}^h 表示第 h个注意力头对应的归一化注意力系数。最后，个性化GAT编码器通过门控循环单元（GRU）将个性化通信信息 $\hat{m}_t^{ij}$ 传递给智能体 i。

2）Q网络：我们采用双重决斗深度Q网络（D3QN）进行模型训练，该网络结合了决斗DQN [32]和双重DQN [33]的优势。这种方法能有效稳定训练过程，并缓解Q值高估问题。智能体 i 在时刻 t的动作价值函数定义如下：

$$Q_{s,a}^i = V_s(m_t^{ij}) + \left[A(m_t^{ij})_a - \frac{1}{|\mathcal{A}|}\sum_{a'}A(m_t^{ij})_{a'}\right], \tag{5}$$

其中，$\mathcal{A}$表示动作空间，m_t^{ij} 表示智能体 i的通信后消息。在获得智能体 i的Q值后，我们采用时序差分（TD）误差作为模型训练的损失函数。使用均方误差（MSE）计算多步 TD 误差：

$$\mathcal{L}(\theta) = \text{MSE}\left(R_t^i - Q_{s_t,a_t}^i(\theta)\right), \tag{6}$$

其中，$R_t^i = r_t^i + \gamma r_{t+1}^i + \cdots + \gamma^n Q_{s_{t+n},\ a_{t+n}}^i(\bar{\theta})$表示多步期望回报，$r_t^i$表示智能体 i 在时间 t获得的奖励。θ 表示目标网络参数，而 θ 对应当前网络参数。

B. 人群增强局部观测

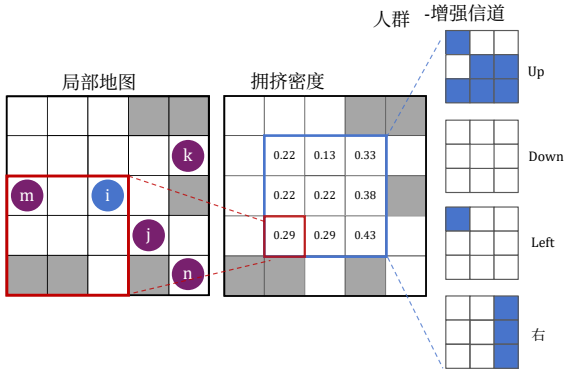


图 3：智能体 i 获取人群增强通道的示意图。为便于说明，图中视野（蓝色边界框）和人群感知区域（红色边界框）均以面积为 3×3 的区域为例。拥挤密度根据 i 周围 3×3 区域内的智能体和障碍物数量计算。若相应动作能降低 i 的综合拥堵程度，则将每个方向通道（上、下、左、右）标记为 1；否则标记为 0。

我们的方法中，智能体的局部观测由两个部分组成。一部分是基本信息，包括障碍物位置、智能体位置和基于距离的启发式信息。另一部分是人群感知信息，用于动态评估局部区域内当前和未来的拥堵情况。如图 3 所示，每个智能体通过评估其局部区域内的拥堵程度，计算四个人群增强通道。具体而言，每个位置的拥堵程度由对周围5 × 5 区域的综合评估确定，该区域由静态拥堵 ρ_{static} 和动态拥堵 $\rho_{dynamic}$构成。静态拥堵 $\rho_{static}(x,y)$用于量化由障碍物分布造成的空间约束，其度量值计算如下：

$$\rho_{static}(x,y) = \frac{O(x,y)}{N(x,y)}, \tag{7}$$

其中，$O(x,\ y)$表示位置（$x,\ y$）处局部区域内障碍物的数量，$N(x,\ y)$表示位置（$x,\ y$）处局部区域内网格单元的总数。动态拥堵 $\rho_{dynamic}(x,\ y,\ t)$用于衡量智能体移动造成的空间占用，其度量方式计算如下：

$$\rho_{dynamic}(x,y,t) = \frac{A(x,y,t)}{N(x,y) - A(x,y,t) - O(x,y)}, \tag{8}$$

其中，$A(x,y,t)$ 表示时间 t 时位置（x,y）周围局部区域内的智能体数量。综合

拥堵程度 $\rho(x,y,t)$ 通过对静态拥堵和动态拥堵进行加权求和得到，如下所示：

$$\rho(x,y,t) = \alpha \cdot \rho_{\text{static}}(x,y) + (1-\alpha) \cdot \rho_{\text{dynamic}}(x,y,t), \quad (9)$$

其中，α 用于平衡静态拥堵与动态拥堵的相对影响。通过整合局部人群感知信息，智能体不仅能够监测当前区域的可通行性，还能预测未来潜在活动区域的空间状态，从而增强整体环境感知能力。

在时刻 t，智能体 i 接收一个维度为 $H \times W \times 10$ 的二值观测矩阵o_t^i，其中 $H \times W$ 表示视野的尺寸。六个通道编码障碍物位置、智能体位置和基于距离的启发式信息。四个通道表示人群增强特征。观测编码器包含四个卷积层和一个GRU模块。观测 o_t^i 首先经过卷积层预处理，然后与上一个时间步的通信信息 m_{t-1}^{ij}相结合，最后输入GRU模块以生成当前特征 m_t^i。

C. 协作式死锁消解

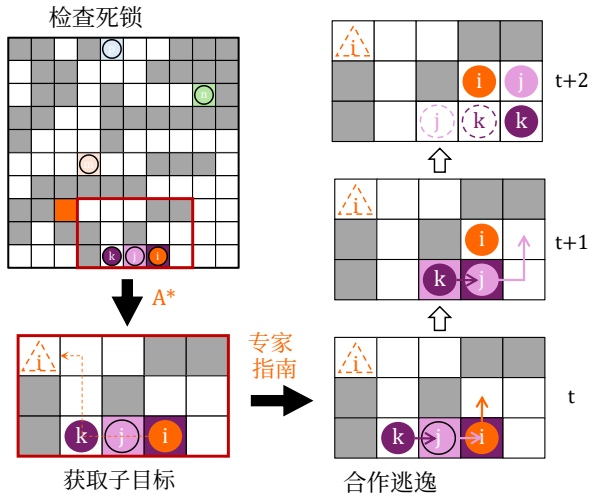


图 4：典型死锁消解示例示意图。首先，将智能体 j，i，k识别为一个死锁智能体组。接下来，构建一个局部地图，并使用 A* 为智能体 i生成一个子目标。在专家指导下，智能体协作解决死锁。最后，在时间步 t+2，强化学习规划器继续执行任务。

死锁是多智能体路径规划中不可避免的挑战，在密集环境中尤为如此，因为狭窄空间会困住智能体，导致任务失败。现有方法采用随机动作或单智能体 A* 算法来解决问题，但它们着眼于个体层面的解决方案，忽视了根本原因：协作能力不足，阻碍了智能体在复杂局部环境中进行有效协同。因此，我们提出了一种基于区域、由专家引导的协作式死锁消解策略。图 4 展示了一个典型示例：

死锁消解过程。我们首先分别识别死锁智能体集合 $A_{\text{deadlocked}}$ 和死锁智能体组 $\mathcal{D}$，以明确执行场景。死锁智能体集合 $A_{\text{deadlocked}}$ 由活跃智能体（尚未到达目标位置）组成，这些智能体在一段时间内保持静止或徘徊。形式化定义如下：

$$A_{\text{deadlocked}} = \left\{ i \in A_{\text{active}} \;\middle|\; v_i^{t-1} = v_i^{t-3} \wedge v_i^{t-2} = v_i^{t-4} \right\}, \quad (10)$$

其中，A_{active} 表示活跃智能体的集合，v_i^{t-} 表示智能体 i^k 在时间 $t-k$的位置。死锁智能体组的集合 $\mathcal{D}$ 定义如下：

$$\begin{aligned} \mathcal{D} = \{ \{i, \mathcal{B}_i\} \mid i \in A_{\text{deadlocked}}, \\ \forall j \in \mathcal{B}_i,\ j \notin A_{\text{active}} \vee j \in A_{\text{deadlocked}} \}, \end{aligned} \quad (11)$$

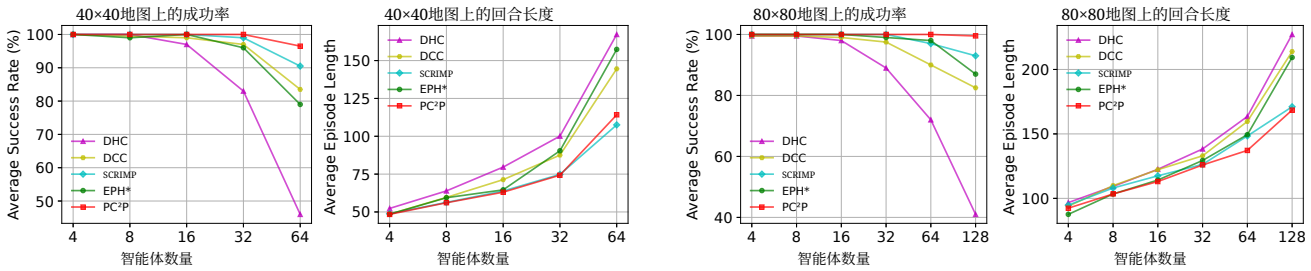
其中 $\mathcal{B}_i$ 表示智能体i的邻近智能体集合。在每个死锁智能体组中，所有智能体要么陷入死锁，要么处于非活跃状态，因为它们无法移动会导致复杂死锁的出现。针对每个死锁智能体组，构建一张最小化局部地图。如果智能体的目标位于该地图之外，则使用 A* 设置一个子目标，使局部移动与全局路径保持一致。此时定义一个局部规划任务，并使用 ODrM*[6] 来解决局部地图中的死锁。该策略能有效帮助分布式算法在视野有限的情况下解决复杂死锁问题。图4 展示了智能体 j、i、k在专家指导下打破死锁的具体过程。在时间步$t+2$，死锁成功解除，智能体得以按照多智能体强化学习方法继续执行后续任务。

V. 实验

A. 实验设置

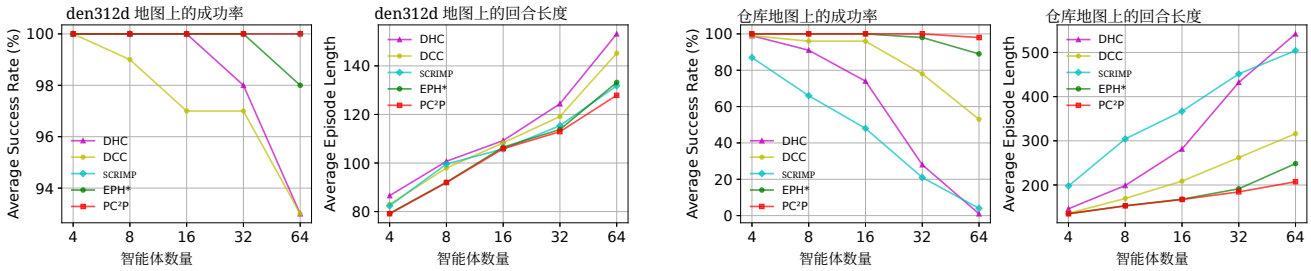
1）训练：PC²P采用基于Q学习的多智能体强化学习方法进行训练，并通过分布式优先经验回放缓冲区提升训练效率。Q网络使用折扣因子为0.99的两步时序差分（TD）误差进行更新。通过课程学习策略逐步提高任务难度。初始环境设置为一个 10×10 网格，包含一个智能体，障碍物按范围为0至0.5、峰值为0.33的三角分布生成。当成功率超过0.9时，地图尺寸逐渐增大至40×40，智能体数量增加至16。每个智能体的视野大小为 9×9，人群感知区域为 5×5。每回合的最大步数设为256。为提高探索效率，32个执行器并行运行，每个执行器每训练1,000步更新一次网络。实验在一台配备NVIDIA 4090 GPU和AMD EPYC 9654（96核）处理器的服务器上进行。总训练步数为30k，训练时长约为24小时。

2）测试：本实验采用与 [9]，[10]，[12] 相同的测试地图来评估 PC²P 的性能，包括随机地图和结构化地图。对于随机地图，我们在 40×40 和 80×80 地图上进行测试，障碍物密度为 0.3



(a) 40×40地图上的结果　　(b) 80×80地图上的结果

图 5：随机地图上的成功率和回合长度实验结果。



(a) den312d 地图上的结果　　(b) 仓库地图上的结果

图 6：结构化地图上的成功率和回合长度实验结果。

最大步长分别为 256 和 386。结构化地图 den312d（65×81）和仓库（161×63）来自 Moving AI 实验室数据集 [1],，最大步长分别为 256 和 512。为了测试模型在更大规模、高密度环境中的性能，120×120 和 240×240 随机地图也用于测试，障碍物密度均为 0.3，最大步长分别为 512 和 1,024。每种随机地图包含 100 个实例，每种结构化地图包含 300 个实例。评估指标包括成功率（SR），表示已完成任务的比例；以及回合长度（EL），反映解决方案的质量。我们选取了四种最先进的分布式多智能体路径规划模型作为对比基线算法：DHC[9], DCC [10], SCRIMP [11], 和 EPH [12], 。评估使用其公开的源代码和预训练模型权重。EPH 集成了 22 种不同的规划器[1]，并采用并行计算为每项任务选取最优结果。由于计算成本高、耗时长，它不适用于实时多智能体路径规划任务。因此，我们采用开源代码中提供的默认规划器进行测试，并将其称为 EPH*。

B. 对比结果

1) 随机地图：如图 5所示，PC²P 在 40×40 和 80×80 地图上均取得了最高的成功率。尤其是当智能体数量增加到 128（尺寸为 80）时，PC²P 仍能将成功率稳定保持在接近 100%，而其他模型的成功率则显著下降。

[1]可访问 https://github.com/ai4co/eph-mapf

此外，PC²P 在两种地图上的平均步数几乎都是最低的。例外情况是智能体数量为 64（尺寸为 40）时，SCRIMP 的表现更好。这可能归因于其有效的冲突解决策略，该策略缩短了发生冲突的智能体的路径长度，进而降低了整体平均步数；然而，在这种情况下，SCRIMP 的成功率比 PC²P 低约 5%。

2) 结构化地图：图6 展示了 PC²P 在结构化地图上的性能，在两种测试地图上均取得了最优结果。尤其是在高度模拟真实环境的仓库地图中，PC²P 即使面对 64 个智能体，成功率仍达到 98%，而大多数其他模型的性能则显著下降。这凸显了该方法在不同环境中处理规划任务的有效性。

3) 大规模地图：表II展示了 PC²P 与其他模型在 120×120和 240×240随机地图上的实验结果。PC²P 在所有大规模地图上均取得了最高的成功率和最低的平均步数。随着智能体数量增加，对协作的需求以及死锁的发生率显著上升，导致其他模型的性能急剧下降。相比之下，PC²P 的成功率始终保持在 90% 以上。这展现了 PC²P 在大规模、高密度环境中的可扩展性和协调能力。

C. 消融研究

表III 评估了 PC²P 中各模块对模型性能的贡献，其中 PGC、CA 和 CDB 分别代表个性化图增强通信模块、

表II：**大规模地图上的成功率和回合长度实验结果。**

方法	*120*大小的地图，密度为*0.3*						*240*大小的地图，密度为*0.3*					
	$n = 128$		$n = 256$		$n = 384$		$n = 256$		$n = 512$		$n = 768$	
	SR↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓
DHC	53	368.1	13	487.9	1	511.8	44	778.9	8	983.2	1	1019.9
DCC	87	260.4	79	329.6	52	436.3	88	495.9	61	696.4	57	749.7
SCRIMP	35	416.1	0	–	0	–	0	–	0	–	0	–
EPH*	93	240.3	82	320.3	29	470.4	94	451.2	66	659.0	29	879.7
PC^{2}P	**100**	**210.0**	**99**	**252.1**	**91**	**338.6**	**98**	**423.3**	**96**	**477.1**	**97**	**510.0**

表III：**不同模块在成功率和回合长度上的消融研究结果。**

方法	*80* 格地图，密度为 *0.3*						*120* 格地图，密度为 *0.3*					
	$n = 128$		$n = 192$		$n = 256$		$n = 256$		$n = 384$		$n = 512$	
	SR↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓	*SR*↑	*EL*↓
不使用 PGC +DHC	80	219.8	50	310.7	11	373.9	64	386.6	20	480.8	2	508.5
不使用 PGC +DCC	95.5	184.5	81	245.0	65	305.3	95	268.1	66	389.3	21	485.1
不使用 PGC +SCRIMP	96	179.9	85	241.8	51	326.2	94	273.6	58	407.5	20	489.1
不使用拥堵感知	94.5	176.3	82	232.6	50	319.9	89	282.2	67	383.7	18	489.9
不使用 CDB	96.5	176.9	90	231.9	67	301.1	95	274.1	82	354.3	42	460.5
PC^{2}P	**99.5**	**168.5**	**94**	**215.1**	**77**	**281.6**	**99**	**252.1**	**91**	**338.6**	**56**	**440.9**

人群增强局部观测模块和协作式死锁破解模块。实验在尺寸为 80×80 和 120×120、障碍物密度为 0.3 的随机地图上进行。由于智能体数量较少时成功率始终达到 100%，因此将智能体数量从 128 逐步增加到 512，以准确评估各模块的影响。

1）个性化图增强通信模块：为评估 PGC 模块的有效性，我们将其替换为 DHC [9],DCC [10],和 SCRIMP [11]的通信模块。由于 EPH [12] 采用与 SCRIMP 相同的通信模块，因此未在实验中单独纳入。我们对上述所有模型重新训练并进行测试。如表III所示，PGC 模块优于所有替代方案，而且随着智能体数量增加，其优势更加显著。值得注意的是，在 512 个智能体（规模为 120）的设置下，PC^{2}P 的成功率比表现最佳的替代通信模块高出 35%。我们可以看到，随着场景密度增加，对协作能力的要求也随之提高。PGC 模块会动态调整智能体间关系，并根据个体需求生成个性化信息，从而实现更高效的协作通信与信息融合。

2）人群增强的局部观测：移除 CA 模块后的实验结果如表 III 所示。随着智能体密度增加，CA 模块的作用愈发显著。具体而言，在 256 个智能体（尺寸为 80）和 512 个智能体（尺寸为 120 ）的环境中，成功率分别下降了 27% 和 38%，平均步数分别增加了 38.3 和 49。这些显著变化表明，随着环境密度增加，局部拥堵信息对于有效路径规划变得愈发关键。基于启发式的通道引导智能体朝目的地移动，而人群增强的通道则可能促使它们避开拥堵区域。二者结合为实时策略优化提供了有价值的方向信息，

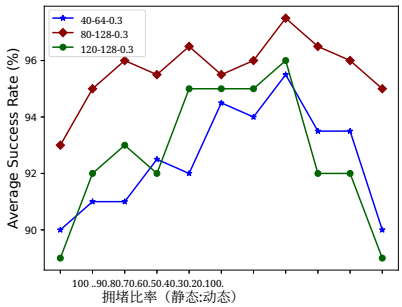


图 7：拥堵比率对成功率的影响。

息以进行实时策略优化。

为了进一步探究静态拥堵 ρ_{static} 与动态拥堵 $\rho_{dynamic}$比例的影响，我们评估了不同的比例，如图7所示。当静态拥堵系数与动态拥堵系数之比设为0.7 : 0.3时，三种随机地图上的成功率均达到最佳水平，显著优于仅使用其中任一参数时的成功率。

3）协作式死锁破解模块：移除 CDB 模块后的实验结果见表III。在高密度环境中，CDB 模块的影响显著增大。在 256 个智能体（地图尺寸为 80）和 512 个智能体（地图尺寸为 120）的环境中，成功率分别提高了 10% 和 14%。这项改进可归因于密度更高的环境中复杂死锁出现的频率增加。CDB 模块通过智能体局部协调实现有效避让，从而解决这一问题。

值得注意的是，在所有消融实验中，移除 CDB 的模型表现仅次于 PC^{2}P。这是因为 PGC 和 CA 模块已经在一定程度上增强了死锁消解能力。只有遇到极其复杂的死锁问题时，才需要借助传统算法进行小规模

局部协同。这些场景往往是阻碍整体任务完成的关键瓶颈。因此，CDB模块对于提升模型在不同环境中的成功率和泛化能力至关重要。

六、结论

本文提出了一种新型分布式实时规划器 PC^{2}P，用于解决多智能体路径规划问题。该方法采用个性化图增强通信模块，利用图注意力机制动态捕捉智能体间关系，并自适应地分配个性化消息。此外，我们引入了整合静态与动态拥堵状态的人群感知信息，以增强智能体在复杂场景中的态势感知能力。我们设计了一种协作式死锁破解策略，在专家指导下应对严重的死锁情况。实验结果表明，PC^{2}P 在不同规模、密度和类型的测试环境中均展现出显著优势，尤其是在高密度环境下表现出更强的协作能力和可扩展性。未来的工作将着重解决独立Q学习（IQL）框架固有的不稳定性，并探索整合集中训练与分散执行（CTDE）框架，以减少多智能体路径规划任务中的冲突和死锁。

参考文献

[1] R. Stern, N. Sturtevant, A. Felner, S. Koenig, H. Ma, T. Walker, J. Li, D. Atzmon, L. Cohen, T. Kumar等人，“多智能体路径规划：定义、变体与基准”，载于国际组合搜索研讨会论文集，第10卷，第1期，2019年，第151–158页。[2] J. Chung、J. Fayyad、Y. A. Younes和H. Najjaran，“团队导航学习：多智能体路径规划深度强化学习技术综述”，《人工智能评论》，第57卷，第2期，第41页，2024年。[3] P. R. Wurman、R. D’Andrea和M. Mountz，“在仓库中协调数百辆协作型自动驾驶车辆”，《AI杂志》，第29卷，第1期，第9–9页，2008年。[4] J. Li、E. Lin、H. L. Vu、S. Koenig等人，“基于优先级搜索的自动驾驶车辆交叉路口协调”，载于*AAAI*人工智能会议论文集，第37卷，第10期，2023年，第11 578–11 585页。[5] F. Ho、A. Goncalves、A. Salta、M. Cavazza、R. Geraldes和H. Prendinger，“无人机交通管理中的多智能体路径规划：机器人赛道”，2019年。[6] C. Ferner、G. Wagner和H. Choset，“ODrM*：低维搜索空间中的最优多机器人路径规划”，载于*2013*年*IEEE*国际机器人与自动化会议。IEEE，2013年，第3854–3859页。[7] G. Sharon、R. Stern、A. Felner和N. R. Sturtevant，“用于最优多智能体路径规划的基于冲突的搜索”，《人工智能》，第219卷，第40–66页，2015年。[8] G. Sartoretti、J. Kerr、Y. Shi、G. Wagner、T. S. Kumar、S. Koenig和H. Choset，“PRIMAL：通过强化与模仿多智能体学习进行路径规划”，IEEE《机器人与自动化快报》，第4卷，第3期，第2378–2385页，2019年。[9] Z. Ma、Y. Luo和H. Ma，“通过通信实现的分布式启发式多智能体路径规划”，载于*2021*年*IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，2021年，第8699–8705页。[10] Z. Ma、Y. Luo和J. Pan，“面向多智能体路径规划的选择性通信学习”，IEEE《机器人与自动化快报》，第7卷，第2期，第1455–1462页，2021年。[11] Y. Wang、B. Xiang、S. Huang和G. Sartoretti，“SCRIMP：适用于基于强化与模仿学习的多智能体路径规划的可扩展通信”，载于*2023*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，2023年，第9301–9308页。

[12] H. Tang、F. Berto和J. Park，“面向多智能体路径规划的优先级混合策略集成”，arXiv预印本 arXiv:2403.07559，2024年。[13] W. Li、H. Chen、B. Jin、W. Tan、H. Zha和X. Wang，“结合优先级通信学习的多智能体路径规划”，2022年。 [在线]。网址：https://arxiv.org/abs/2202 03634 。P. Pham和A. Bera，“通过[14]图神经网络局部通信优化人群感知多智能体路径规划”，载于2024年IEEE/RSJ智能机器人与系统国际会议（IROS），2024年，第8913–8913页。[15] Q. Li、W. Lin、Z. Liu和A. Prorok，“面向大规模多机器人路径规划的消息感知图注意力网络”，IEEE《机器人与自动化快报》，第6卷，第3期，第5533–5540页，2021年。[16] Z. Song、R. Zhang和X. Cheng，“HELSA：利用时空抽象进行大规模多智能体路径规划的分层强化学习”，载于2023年IEEE/RSJ智能机器人与系统国际会议（IROS），2023年，第7318–7325页。[17] V. Mnih，“深度强化学习的异步方法”，arXiv预印本 arXiv:1602.01783，2016年。[18] J. Li、A. Tinka、S. Kiesel、J. W. Durham、T. S. Kumar和S. Koenig，“大型仓库中的终身多智能体路径规划”，载于AAAI人工智能会议论文集，第35卷，第13期，2021年，第11 272–11 281页。[19] Y. Bengio、J. Louradour、R. Collobert和J. Weston，“课程学习”，载于第26届国际机器学习年会论文集，2009年，第41–48页。[20] T. Schaul、J. Quan、I. Antonoglou和D. Silver，“优先经验回放”，arXiv预印本 arXiv:1511.05952，2015年。[21] T. N. Kipf和M. Welling，“基于图卷积网络的半监督分类”，CoRR，第abs/1609.02907卷，2016年。

[ ] 1609 02907 在线 。网址：[22]http://arxiv.org/abs/ 。Q. Lin和H. Ma，“SACHA：结合基于启发式的注意力的软演员-评论家，用于部分可观测多智能体路径规划”，IEEE《机器人与自动化快报》，第8卷，第8期，第5100–5107页，2023年。[23] A. Vaswani，“注意力就是一切”，《神经信息处理系统进展》，2017年。[24] C. Zhu、M. Dastani和S. Wang，“带通信的多智能体深度强化学习综述”，2024年。 [在线]。网址：https://arxiv.org/abs/2203 08975 。A. Das、T. Gervet、J.[25]Romoff、D. Batra、D. Parikh、M. Rabbat和J. Pineau，“TarMAC：目标导向的多智能体通信”，载于第36届国际机器学习会议论文集，2019年，第1538–1546页。[26] Z. Ding、T. Huang和Z. Lu，“面向多智能体协作的个体推断通信学习”，2021年。 [在线]。网址：https://arxiv.org/abs/2006 06455 。Y. Niu、R. R. Paleja和[27]M. C. Gombolay，“多智能体图注意力通信与团队协作”，载于AAMAS，第21卷，2021年，第20届。[28] P. Veli ̆ckovi ́c、G. Cucurull、A. Casanova、A. Romero、P. Li`o和Y. Bengio，“图注意力网络”，2018年。[在线]。网址：https://arxiv.org/abs/1710 10903 。X. Wang、X. Li、J. Shao和J. Zhang，“AC2C：面向多智能体强化学习的[29]自适应控制两跳通信”，2023 [ ] 2302 12515 。在线 。网址：[30]https://arxiv.org/abs/ 。X. Li和J. Zhang，“面向多智能体强化学习的上下文感知通信”，arXiv预印本 arXiv:2312.15600，2023年。[31] S. Brody、U. Alon和E. Yahav，“图注意力网络的注意力程度如何？”，载于国际学习表征会议，2022年。 [在线]。网址：https://openreview.net/ forum?id=F72ximsx7C1[32] Z. Wang、T. Schaul、M. Hessel、H. Hasselt、M. Lanctot和N. Freitas，“深度强化学习的决斗网络架构”，载于国际机器学习会议。PMLR，2016年，第1995–2003页。[33] H. Van Hasselt、A. Guez和D. Silver，“基于双重Q学习的深度强化学习”，载于AAAI人工智能会议论文集，第30卷，第1期，2016年。