arXiv:2603.08273v1 [cs.RO] 9 Mar 2026

# 少即是多：通过表征简约性实现鲁棒的零通信三维追逃

Jialin Ying、Zhihao Li、Zicheng Dong、Guohua Wu和Yihuan Liao*

自动化系，中南大学，长沙，中国

摘要——**杂乱体素环境中的非对称三维追逃在通信延迟、部分可观测性和非完整机动限制下十分困难。尽管许多多智能体强化学习方法依赖更丰富的智能体间耦合或集中式信号，但当通信出现延迟或噪声时，这些依赖可能成为脆弱性来源。在继承的路径引导去中心化追捕框架 [1], 基础上，我们研究一个面向鲁棒性的问题：表征简约性能否改善无通信协作?**

**我们通过以下方式落实这一原则：（i）简约的行动者观测接口，移除团队耦合通道（83维至 → 50维）；（ii）贡献门控信用分配（CGCA），一种面向通信受限协作的局部性感知信用结构。在第5阶段评估中（4名追捕者对1名逃逸者），我们的配置达到** $0.753 \pm 0.091$ **成功率与**$0.223 \pm 0.066$ **碰撞，优于83维完整观测对照（**$0.721 \pm 0.071$，　$0.253 \pm 0.089$）。**该方法在速度/偏航/噪声/延迟压力测试下进一步表现出平缓的性能下降，并在城市峡谷地图上展现出稳健的零样本迁移（密度0.24时成功率约61%）。这些结果支持一条实用设计原则：在通信受限的多机器人追捕中，表征稀疏性可能比更丰富的智能体间耦合更加鲁棒。**

索引词——**多智能体强化学习、追逃、时空解耦、零通信、仿真到真实环境的鲁棒性**

## I. 引言

在杂乱的体素空间中进行三维追逃十分困难，因为通信延迟、部分可观测性和非完整约束会与高度非凸的几何环境相互作用。在本文考虑的4对1场景中，追捕者的机动能力受限（标称速度为$v_x \leq 8\,\mathrm{m/s}$、角速度为 $|\psi| \leq 0.8\,\mathrm{rad/s}$），而逃逸者在第5阶段的速度可达9 m/s，因此过时的估计和延迟的修正可能破坏协同拦截的稳定性。

一种常见的多智能体强化学习策略是通过更丰富的队友条件化观测、显式通信或集中式评论家来增强智能体间的耦合。尽管这些方法在标称仿真中有所帮助，但在存在延迟/噪声时，这类耦合可能会传播过时的队友信念。因此我们提出：在复杂的三维追逃中，去中心化协调是否必然需要更强的跨智能体耦合?

文献 [1] 提出了路径引导的协同搜索与捕获框架（三维A*算法引导 + 方向性前沿分配）。本文不再重新探讨该框架的设计，而是研究一个不同的问题：在杂乱环境下的三维追逃任务中，面对通信拒止、延迟和感知噪声，哪些观测与信用分配结构仍能保持稳健？我们将继承的学习器视为一种参数共享的去中心化PPO/IPPO式演员–评论家（如 [1]中所实现），并将其与集中式评论家的 MAPPO 变体区分开来。

我们的核心洞见是一项以鲁棒性为导向的设计原则：表征简约性能够提升无通信协作。我们的研究揭示了这一情境下两个反直觉的效应：**(i)** 更多团队耦合信息可能降低而非提升鲁棒性；**(ii)** 当稀疏的局部观测与局部性感知信用分配相结合时，无通信协作仍能保持强劲表现。我们的贡献如下：

- **将表征简约性作为通信受限的三维追捕任务中去中心化多智能体强化学习的设计原则**。 我们发现，减少显式的跨智能体观测耦合能够提升系统在延迟和感知噪声条件下的鲁棒性。
- **贡献门控信用分配（CGCA）**。我们提出了一种局部性感知信用结构，无需显式通信信道即可维持协同捕获质量。
- **全面的鲁棒性与泛化能力证据**。 我们提供了基准评估，以及针对速度、偏航角、噪声和延迟的压力测试套件，并在程序生成的城市峡谷地图上进行了零样本迁移。

宏观拓扑规划框架自身沿用自[1]；我们的创新在于在这一继承而来的分散式基础上，面向鲁棒性设计信息机制和信用分配设计。

## II. 相关工作

*1) A.* 三维追逃：经典追踪制导以解析控制器为基础，例如APF式反应势场塑形 [12] 和比例导航（PN） [13]。这些方法轻量且可解释，但本质上是局部的，因此在体素化三维杂乱环境中容易陷入拓扑死锁：控制器可能困于障碍物引起的局部极小值，或陷入振荡式转向—超调循环。拓扑感知规划（例如A*算法）通过揭示同伦层级结构 [11], 缓解了部分死锁，但独立规划在延迟执行和模型失配下仍然脆弱。近期的深度强化学习流程提升了

*通讯作者：Yihuan Liao（邮箱：225189@csu.edu.cn）

适应性，但朴素的端到端形式在高维部分可观测状态空间中常常面临组合爆炸 [4]。 [1] 的路径引导的去中心化框架（该文中命名为PGF-MAPPO）因此促成了一种务实的折中：规划器编码全局几何，而策略则专注于局部的、不确定性感知的残差控制。

*2*）*B.* 通信受限多智能体强化学习：多智能体强化学习中的显式通信已通过可微消息传递、注意力路由和图交互模块得到广泛研究，包括 CommNet、注意力通信和目标导向的多智能体通信 [5]–[8]。在仿真中，这些通信通道可以提升协作效率；然而，高频空中追捕恰恰是通信质量呈非平稳变化的场景。延迟、数据包丢失和异步时钟会使消息传递转变为过时信念传播，从而可能引发跨智能体错误级联，而非带来协作收益。当控制时域较短且碰撞余量较小时，这种不匹配尤为突出。因此，我们设计零通信智能体并非出于刻意精简，而是将其作为一种鲁棒性先验，明确切断脆弱的耦合路径，并促使策略学习依赖局部几何信息与共享拓扑引导。

*3*）*C.* 协作多智能体强化学习中的信用分配：信用分配仍是协作多智能体强化学习中的核心问题。COMA 等反事实基线，以及 VDN/QMIX 等价值分解方法，通过在训练期间引入集中式结构来降低多智能体非平稳性 [3], [9], [10]。当全局状态可靠时，这些方法效果显著；但在现实的分布式机器人学中，依赖集中式信号可能成为实际瓶颈，因为同步状态聚合成本高昂，甚至无法实现。相比之下，CGCA 仅利用局部距离和接近运动学信号来塑造协作激励。这种轻量的局部机制天然适用于严格的分散式执行，也避免了在本就复杂的三维追捕动力学之上再增加一个高维集中式通道。

## III. 问题表述与环境设置

### *A. Stage-5*环境

世界被离散化为覆盖 $311.0 \times 311.0 \times 110.5$ 米范围的 $52 \times 52 \times 18$ 体素网格（体素尺寸：6.0米）。第5阶段设置包括60米可见范围、8米捕获半径以及70%可见性门限。可通行占用率为0.127，最大自由连通分量覆盖了0.951的自由体素，表明拓扑虽杂乱但全局可导航。

### *B.* 非对称博弈模型

设追捕者 $i \in \{1,2,3,4\}$ 具有位置 $\boldsymbol{p}_t^i \in \mathbb{R}^3$、偏航角 $\psi_t^i$以及机体坐标系控制

$$\boldsymbol{u}_t^i = [v_{x,t}^i, v_{y,t}^i, v_{z,t}^i, \dot{\psi}_t^i]^\top. \tag{1}$$

动力学是

$$\boldsymbol{p}_{t+1}^i = \boldsymbol{p}_t^i + \Delta t\, \mathbf{R}_z(\psi_t^i)\boldsymbol{v}_t^i,\ \psi_{t+1}^i = \psi_t^i + \Delta t\, \dot{\psi}_t^i, \tag{2}$$

表I Stage-5环境与非对称常数。

| Item | 值 |
|---|---|
| 世界/网格 | $311.0 \times 311.0 \times 110.5$ 米， $52 \times 52 \times 18$ 体素 |
| 团队配置 | 4 名追捕者对 1 名逃逸者 |
| 捕获半径（第5阶段） | 8.0 m |
| 可见范围（第5阶段） | 60.0 m |
| 追捕者边界 | $\lvert v_x\rvert \le 8$, $\lvert v_y\rvert \le 4$, $\lvert v_z\rvert \le 3$ 米/秒 |
| 追捕者偏航角上限 | $\lvert\dot{\psi}\rvert \le 0.8$ 弧度/秒（逐步降低至 0.2） |
| 逃逸者速度计划 | 最高可达 9.0 米/秒 |
| 海拔 占据率 | 15.9% / 12.1% / 10.2% |
| 各向异性 | |

带有约束

$$|v_{x,t}^i| \le 8,\ |v_{y,t}^i| \le 4,\ |v_{z,t}^i| \le 3,\ |\dot{\psi}_t^i| \le \bar{\omega}. \tag{3}$$

捕获定义为

$$\min_i \|\boldsymbol{p}_t^i - \boldsymbol{p}_t^e\|_2 \le R_c, \quad R_c = 8\,\mathrm{m}. \tag{4}$$

协作目标为 $J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_t \gamma^t r_t]$，并包含密集的碰撞/安全惩罚和捕获质量激励。

## IV. 方法

### *A.* 路径引导的去中心化框架

继 [1], 之后，每个智能体接收的是拓扑感知三维 A* 引导向量，而非原始路径折线。规划器提供宏观自由空间结构，而循环策略负责局部反应式控制。我们将这一宏观–微观解耦追捕框架保留为继承的基础设施，而非本文的贡献。

为明确方法边界，我们遵循 [1]中的继承实现，将其作为参数共享的去中心化PPO/IPPO式演员–评论家，并单独处理集中式评论家的 MAPPO 变体。

### *B.* 观测空间形式化：通过*83*维到*50*维掩码实现表征简约性

令 $\tilde{\boldsymbol{o}}_i^t \in \mathbb{R}^{83}$ 表示追捕者 $i$ 在时刻 $t$的完整行动者侧观测：

$$\tilde{\boldsymbol{o}}_i^t = \big[\boldsymbol{\ell}_i^t, \Delta\boldsymbol{p}_{ie}^t, \Delta\boldsymbol{v}_{ie}^t, \boldsymbol{v}_i^t, \boldsymbol{a}_i^t, \boldsymbol{\omega}_i^t, \phi(\mathcal{N}_i^t), \boldsymbol{g}_i^t, \boldsymbol{e}_i, m_t, \delta_t, \boldsymbol{s}_i^t, \boldsymbol{c}_i^t\big], \tag{5}$$

其中 $\phi(\mathcal{N}_i^t) \in \mathbb{R}^{24}$ 编码前$k$ 名队友状态， $\boldsymbol{s}_i^t \in \mathbb{R}^7$为战术槽位块， $\boldsymbol{c}_i^t \in \mathbb{R}^2$ 包含包围拓扑线索。为在通信拒止下强制表征简约性，我们应用二值掩蔽算子$\mathbf{M} \in \{0,1\}^{50\times 83}$:

$$\boldsymbol{o}_i^t = \mathbf{M}\tilde{\boldsymbol{o}}_i^t = \big[\tilde{\boldsymbol{o}}_{i,1:41}^t, \tilde{\boldsymbol{o}}_{i,66:74}^t\big] \in \mathbb{R}^{50}. \tag{6}$$

等价地，该算子会显式地将团队耦合通道置零，

$$\mathbf{M}\phi(\mathcal{N}_i^t) = \mathbf{0}, \qquad \mathbf{M}\boldsymbol{s}_i^t = \mathbf{0}, \qquad \mathbf{M}\boldsymbol{c}_i^t = \mathbf{0}, \tag{7}$$

从而移除 24 维队友状态和 9 维槽位／包围描述符。该操作并非作为通用特征剪枝引入；它明确移除了团队耦合

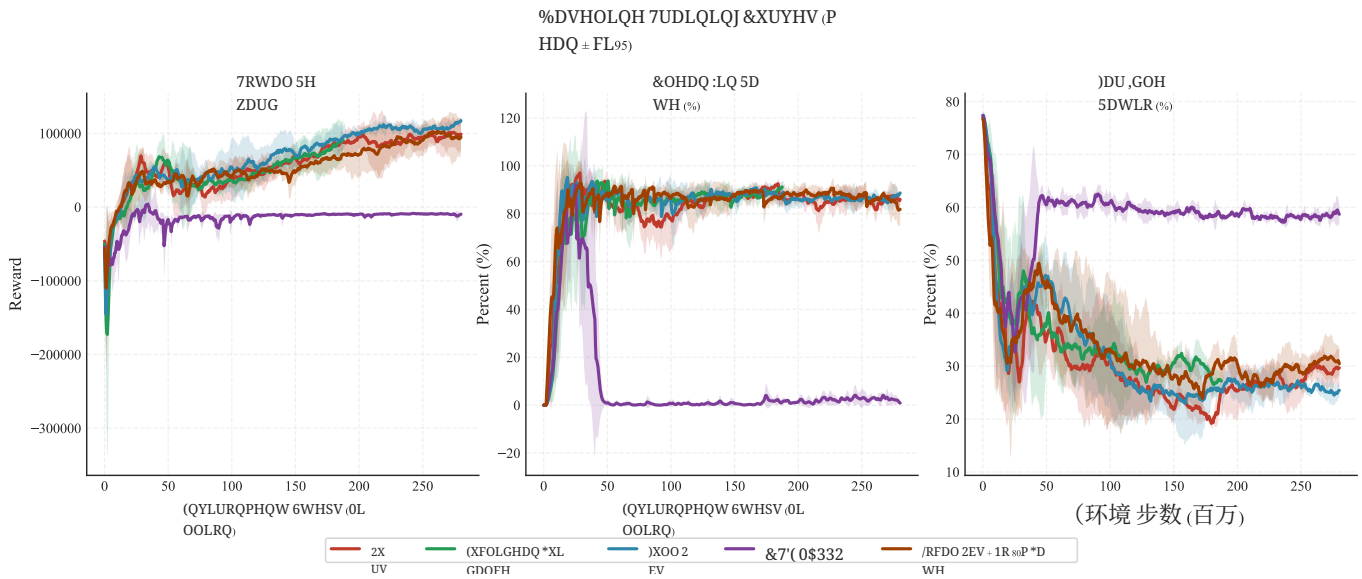


图 1. 第5阶段下的训练奖励曲线。CTDEMAPPO 在可见性门控转换后表现出明显的不稳定性，而 OURS-LITE 收敛至最高的稳定回报区间。

通道，这些通道最容易受到过时的队友估计影响，从而降低对延迟/噪声的敏感性，同时保持动作空间和优化骨干不变。

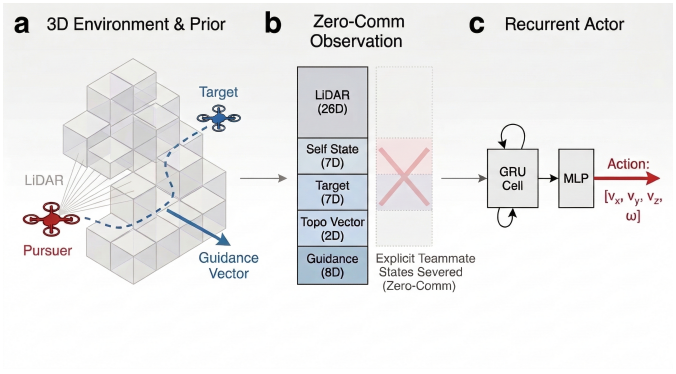


图 2. 继承的路径引导的去中心化流程 [1]。我们的改动位于行动者接口：通过移除队友/槽位/包围通道，将 83 维完整观测掩码为 50 维，以实现零通信鲁棒性。

表 II 面向表征简约性的观测配置缩减。

| 模块 | 83-D | 50-D |
|---|---|---|
| 激光雷达 + 目标 +自身 + 惯性测量单元 | 41 | 41 |
| 队友前$k$ | 24 | 0 |
| 引导 + 标识符 + 模式 + 延迟 | 9 | 9 |
| 槽位 +包围 | 9 | 0 |
| **总计** | 83 | **50** |

### C. 奖励构成与贡献门控信用分配（CGCA）

CGCA 被设计为面向无通信协作的局部性感知信用结构，而非通用的奖励加成。它奖励在拦截中具有几何意义的局部参与，并在无法进行显式消息传递时抑制搭便车均衡。

参数 $\theta$ 对应的协作目标为

$$J(\theta)=\mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T}\gamma^t\frac{1}{N_t}\sum_{i=1}^{N_t}r_i^t\right], \tag{8}$$

其中逐智能体奖励分解为

$$\begin{aligned}r_i^t=\lambda_{\text{dir}}\,g_{\text{dir}}(d_i^t)\,\Delta d_i^t+\lambda_{\text{cap}}\,\rho^t\,\hat{c}_i^t\,\mathbf{1}_{\text{cap}}^t+\lambda_{\text{qual}}\,q_i^t\\-\lambda_{\text{col}}\,\mathbf{1}_{\text{col},i}^t-\lambda_{\text{imp}}\,\kappa_i^t-\lambda_{\text{lazy}}\,\mathbf{1}_{\text{lazy},i}^t,\end{aligned} \tag{9}$$

其中，$\Delta d_i^t=d_i^{t-1}-d_i^t$、$d_i^t=\|\boldsymbol{p}_i^t-\boldsymbol{p}_e^t\|_2$和 $\kappa_i^t$ 表示碰撞冲击强度。

为在零通信条件下稳定协作，CGCA 引入局部性感知方向门控：

$$g_{\text{dir}}(d)=\begin{cases}1, & d\le 40\,\text{m},\\ \frac{80-d}{40}, & 40<d\le 80\,\text{m},\\ 0, & d>80\,\text{m}.\end{cases} \tag{10}$$

捕获份额信用在 60 m 以外采用硬门控，并根据接近行为加权。定义接近速度 $v_{i,\text{ clo}}^t=-\dot{d}_{i_0}^t$。原始贡献为

$$\tilde{c}_i^t=\mathbf{1}[d_i^t\le 60]\Big(\alpha_p e^{-d_i^t/d_0}+\alpha_v[v_{i,\text{clo}}^t]_+ +\alpha_r\mathbf{1}[d_i^t\le R_c]\Big), \tag{11}$$

且归一化份额为

$$\hat{c}_i^t=\frac{\tilde{c}_i^t}{\sum_{j=1}^{N_t}\tilde{c}_j^t+\epsilon}. \tag{12}$$

为抑制搭便车行为，参与率乘以全局捕获奖励：

$$\eta_i^t = \mathbf{1}[d_i^t \le 60]\,\mathbf{1}[v_{i,\text{clo}}^t > 0.5], \qquad \rho^t = \min\left(1, \frac{\sum_i \eta_i^t}{0.5 N_t}\right). \tag{13}$$

因此，如果存活的追捕者中主动逼近目标者不足一半，集体捕获奖励将按比例下调。碰撞抑制被显式编码为

$$\mathbf{1}_{\text{col},i}^t = \mathbf{1}_{\text{obs-hit},i}^t + \mathbf{1}_{\text{team-hit},i}^t + \mathbf{1}_{\text{shield},i}^t, \tag{14}$$

该项在统一项中惩罚障碍物碰撞、智能体间接触以及安全屏蔽触发。

选择40/60/80米阈值，是为了与第5阶段中60米的感知极限以及平台的实际机动范围对齐：40米作为高置信度的局部交互核心，60米匹配可靠可观测范围，80米则抑制无贡献的远场奖励泄漏。

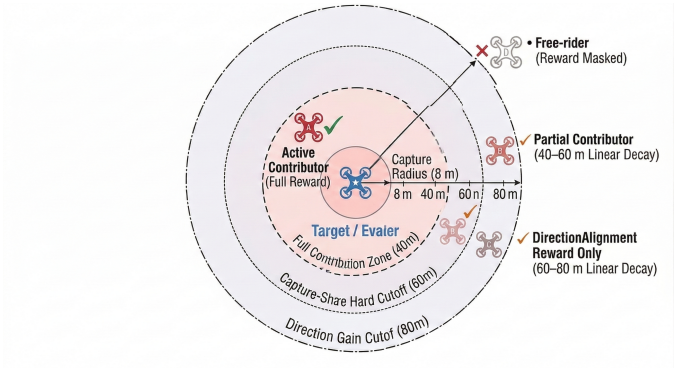


图3. 贡献门控信用分配。方向奖励在40米以内完全生效，在40–80米之间衰减，并在80米以外消失；捕获份额在60米以外被硬门控。这种局部信用分配几何结构抑制了零通信团队追捕中的搭便车均衡。

## V. 实验与评估

### A. 设置与基线

评估遵循所提供的第5阶段协议（主要基准测试：3 个随机种子，每个随机种子 500 个回合；鲁棒性测试套件：每种设置、每个随机种子 200 个回合，其中延迟测试使用 4–5 个评估随机种子）。基线包括 FULL OBS、EUCLIDEAN、LOCAL-NO-GATE、CTDE MAPPO 和 APF+PN。

比较采用有意拆分因素的方式。OURS-LITE、FULL OBS 和 LOCAL-NO-GATE 属于同一继承而来的参数共享分散式近端策略优化系列；OURS-LITE 与 FULL OBS 的对比隔离了信息结构效应（简约的 50 维与团队耦合的 83 维），OURS-LITE 与 LOCAL-NO-GATE 的对比则在相同的 50 维设置下隔离了 CGCA 的作用。CTDE MAPPO 则单独作为集中式评论家基线，用于对比该任务中的集中式耦合与分散式执行。

### B. 课程成熟与可见性门控

为阐明搜索能力的来源，我们采用分阶段课程，并明确实施可见性门控。训练期间，仅当已探索空间比例超过阶段特定阈值 $\tau$ 或达到困难回合超时（后期阶段设置为$>$ 1500步）时，才会有条件地开放目标可观测性。随着课程成熟，阈值 $\tau$ 逐步提高：阶段1不设门控，而后续盲搜索阶段则从45%（阶段3）收紧至60%（阶段4）和70%（第5阶段）。

这一机制造成了结构性瓶颈：智能体必须先内化三维探索，之后才能获得可靠的跟踪信号。换言之，可观测性是通过覆盖率赢得的，而非预先赋予的；这能防止早期的偶然发现主导学习。关键在于，这一门控仅在训练期间生效，作为促使战术涌现的手段；评估期间则会将其完全移除，届时策略将在没有人为可见性约束的情况下接受测试。

### C. 训练动态与样本效率

图1表明，OURS-LITE比完整观测和CTDE MAPPO更早进入高回报区间，且振荡幅度更低。一个关键原因是表征级压缩：移除33维延迟敏感的团队耦合特征会缩小演员输入流形，并降低延迟观测下的干扰协方差。在PPO中，策略梯度估计器

$$\hat{\boldsymbol{g}}_t = \nabla_\theta \log \pi_\theta(a_i^t \mid \boldsymbol{o}_i^t)\, \hat{A}_i^t \tag{15}$$

可以从策略输入方差的角度加以解释：更少的延迟敏感通道会降低干扰协方差，这与我们课程中经验上更稳定的策略更新以及更高的样本效率相一致。相比之下，CTDE MAPPO在该设定中表现出较不稳定的价值学习动态：338维中心化评论家可能增加近似负担，并在可见性门控与杂波诱导的模式切换生效后，潜在地放大对分布偏移的敏感性。

重要的是，图1中间面板（净胜率）表明，CTDE MAPPO并非从一开始就表现较弱：在较简单的课程阶段（阶段1–2）中，它在进入阶段3之前的连续200万步区间内一直保持在90%以上。不稳定性主要出现在阶段3转换处，此时引入了严重的三维非凸遮挡和可见性门控。这些观察表明，当高维价值近似与突发的可见性诱导的分布偏移同时发生时，集中式评论家可能会变得脆弱。

**为什么 CTDEMAPPO 在此处可能变得不稳定？** 中心化评论家接收 338 维输入，这可能会加剧其在高度非凸三维杂波中的近似难度。同时，可见性门控的阶段转换会导致观测分布偏移突然发生，可能进一步放大评论家与目标之间的方差和梯度抖动。综合证据表明，尽管采用了中心化条件化，价值学习仍可能不稳定，策略改进也较弱。

### D. 主要基准测试与消融研究

在去中心化参数共享PPO系列中，OURS-LITE 与 FULLOBS 的对比隔离了信息结构的影响：移除显式的队友耦合通道

表 III 主要阶段 5 基准测试（3 个随机种子上的平均值±标准差，每个随机种子 500 个回合）。OURS-LITE、完整观测和 LOCAL-NO-GATE 是继承参数共享的去中心化 PPO 系列变体。

| 方法 | 成功率 | 无碰撞 | 碰撞 | 平均步数 |
|---|---|---|---|---|
| OURS-LITE | **0.753±0.091** | **0.752±0.091** | **0.223±0.066** | 683.6±269.4 |
| 完整观测 | 0.721±0.071 | 0.721±0.072 | 0.253±0.089 | **655.4±106.7** |
| 欧几里得 | 0.586±0.120 | 0.583±0.121 | 0.353±0.092 | 887.9±237.7 |
| LOCAL-NO-GATE | 0.569±0.090 | 0.568±0.089 | 0.389±0.106 | 669.1±223.2 |
| CTDE MAPPO | 0.006±0.004 | 0.006±0.004 | 0.159±0.111 | 2806.8±139.1 |
| 人工势场+比例导航 | 0.125±0.007 | 0.125±0.007 | 0.875±0.007 | 124.9±2.0 |

可提升成功率（0.753 vs. 0.721）并减少碰撞（0.223 vs. 0.253）。这支持了这一假设：在存在延迟/噪声的追捕中，更丰富的耦合并不总是更稳健。

**消融分析（OURS-LITE 与 LOCAL-NO- GATE）。**在相同的 50 维零通信设置下移除 CGCA，会导致成功率下降 18.4 个百分点（$0.753 \to 0.569$），碰撞率上升 16.6 个百分点（$0.223 \to 0.389$）。这表明 CGCA 是一种必要机制，而非仅仅用于美化的正则化项。

### *E.* 运动学压力测试

图 4 展示了速度扫描（7.0–10.0 m/s）的结果，其中 OURS-LITE 的成功率从 0.907 平稳下降至 0.790。欧几里得方法则呈现出典型的运动学陷阱：在杂乱环境中，直线追逐会延迟可行的转向决策，导致反复出现超调与恢复，并拉长回合时长。

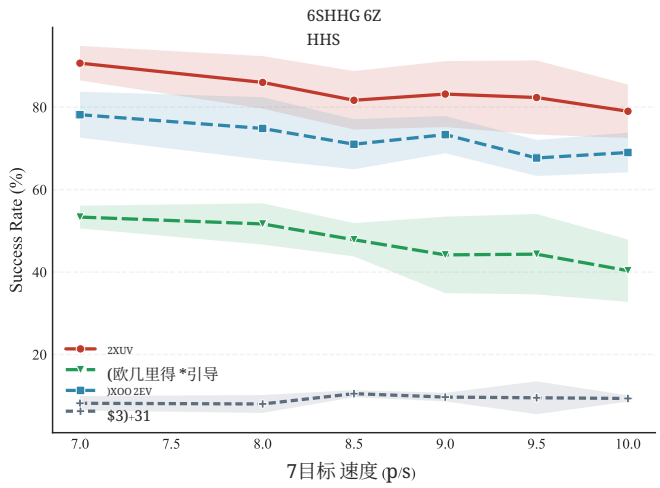


图 4. 速度扫描。随着逃逸者速度增加，OURS-LITE 在成功率与碰撞之间保持了最佳权衡，而欧几里得方法则持续进行步数较多、效率低下的追逐。

图 5 报告了偏航角速度上限限制（$0.8 \to 0.2$ rad/s）下的结果。在严格的转向限制下，性能下降反映了算法与物理特性之间的不匹配：当可行弧线空间缩小时，反应式修正往往为时已晚。

### *F.* 仿真到现实的鲁棒性

噪声和延迟鲁棒性如图 6 和图 7 所示。在延迟条件下，FULLOBS 的性能下降幅度比 OURS-LITE 更大，这与状态去同步的情况一致：

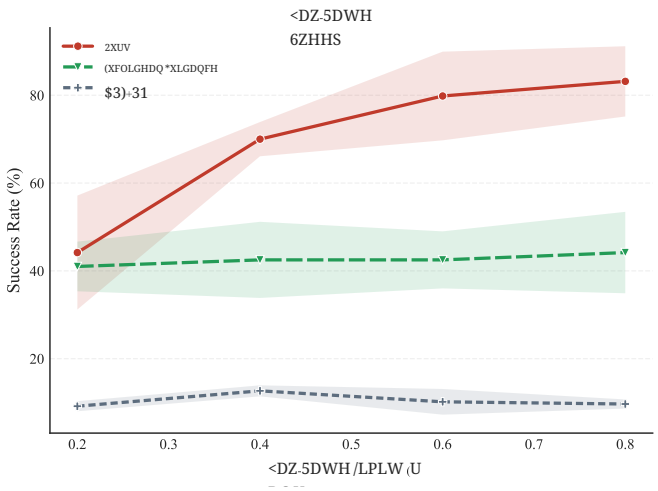


图 5. 偏航角速度扫描。所提出的配置在中等范围内保持了稳健的裕度，并在机动能力减弱时呈现出平稳的性能退化。

紧密耦合的队友通道可能会传播过时的估计值，并可能放大控制误差级联。

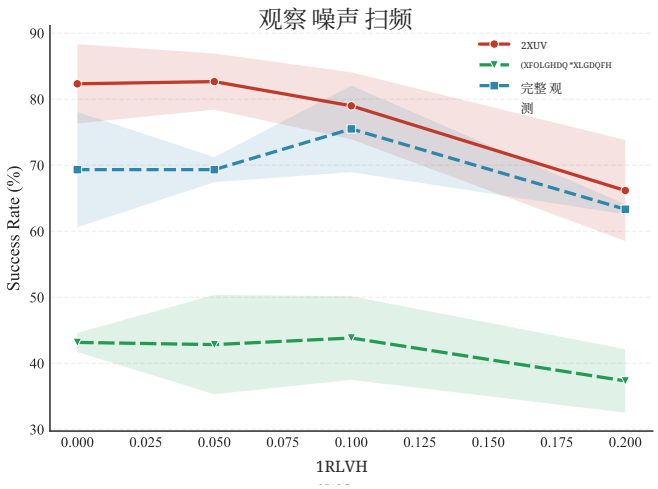


图 6. 观测噪声扫描（$\sigma \in [0, 0.20]$）。在所有测试的噪声水平下，OURS-LITE 的表现始终最佳。

### *G.* 定性分析：跨层垂直扫描

图 8 揭示了一种分为三个阶段的机动模式，而非单一的反应式追逐。初始搜索：在共享的三维A*先验引导下，智能体向与前沿一致的通道分散，同时保持安全间距。高度分层：目标出现的可能性集中后，一名追捕者升至占据率较低的上层，其余智能体则占据中低层，利用测得的各向异性（15.9%/12.1%/10.2%）减少目标的垂直逃逸选项。拓扑围困：随后，队形利用障碍物形态——尤其是 U 形墙体和角落凹处——作为虚拟队友，逐步收紧跨层包围圈，直至最终捕获。

这种分阶段行为解释了为何无通信协作依然有效：智能体不交换显式消息，但共享拓扑先验和 CGCA 所产生的局部

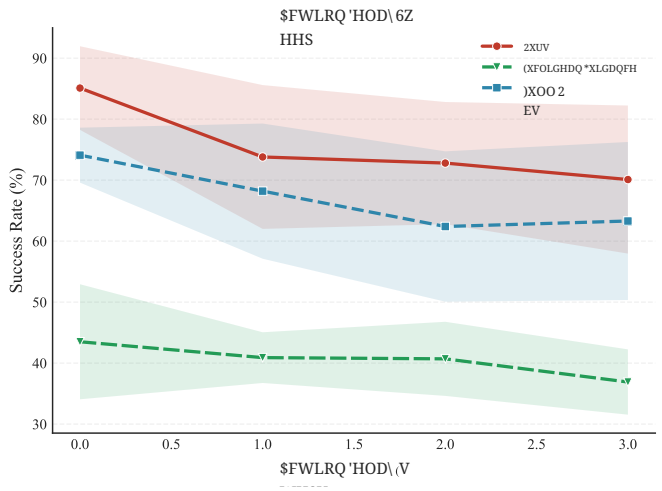


图 7. 动作延迟扫描（0–3 步）。在时延扰动下，OURS-LITE 保持了最佳的绝对鲁棒性表现。

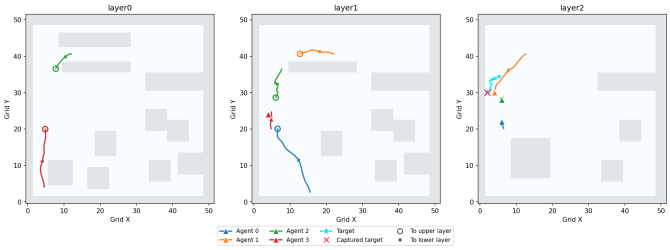


图 8. 三维Z轴战术切片。团队执行跨层垂直扫描：上层压力限制垂直逃逸，同时下层智能体封锁水平走廊，从而形成无需通信的拓扑包围。

激励会随时间同步角色分配。重要的是，包围轨迹在几何上与地图的分层自由空间结构相一致，这表明这是真正的 3D 战术组织，而非偶然的高度抖动。

## VI. 程序化生成城市峡谷中的零样本泛化

为检验地图过拟合，我们在五张程序化生成的 *UrbanCanyon*地图上评估 OURS-LITE 的零样本表现，这些地图的障碍物密度为 0.08–0.24。生成器采用连通的 2.5D 各向异性街道走廊布局（而非随机体素噪声），从而生成具有物理意义的机动走廊，并实现受控的拓扑偏移。

如图 9 所示，性能随密度增加而平缓下降：在最困难的地图（0.24）上，成功率仍约为 61%，碰撞率保持在 40% 以下，平均步数仍处于 350–450 的区间。这些趋势与结构迁移（垂直各向异性条件下的跨层遏制）相符，而非固定地图记忆。

## VII. 讨论与局限

实证结果支持一种以鲁棒性为导向的设计原则：在通信受限的多机器人追捕中，减少跨智能体耦合有时反而能实现更可靠的协同。在我们的设定中，83维团队耦合特征的表现不及50维简约配置；而无需显式通信，局部性感知 CGCA也能维持协作质量。这一结果与这样的观点一致：在延迟/噪声条件下，陈旧的耦合通道可能引入有害方差，而稀疏局部信号则仍然可控。

仍有两项局限。首先，偏航限制扫描目前未包含完整观测，因此低偏航角下的跨方法比较尚不完整。其次，当前噪声测试套件的扰动上限为 $\sigma \leq 0.20$；若要在极端感知损坏情况下得出更可靠的结论，还需要考察更强的扰动范围。

## VIII. 结论

本文基于继承自[1], 的路径引导去中心化框架，研究了一个不同于框架设计的问题：通信受限的三维追逃中，以鲁棒性为导向的信息与信用结构。研究结果支持两项相互关联的发现：（i）更丰富的团队耦合观测并非总是更好，表征简约性能够提升鲁棒性；（ii）通过 CGCA 恰当构建局部信用分配后，无通信协作仍能保持有效。在基准测试、压力测试以及零样本城市峡谷迁移评估中，证据表明这是一项实用的设计原则，而非普遍适用的规则：在通信受限的多机器人追捕中，减少显式的跨智能体耦合有时能够实现更鲁棒的协作。

## 参考文献

[1] J. Ying、Z. Li、Z. Dong、G. Wu 和 Y. Liao，“通过路径引导的多智能体近端策略优化和方向性前沿分配，在杂乱环境中实现可泛化的协同搜索与捕获”，*arXiv* 预印本 *arXiv:2512.09410*，2025。[2] R. Lowe 等，“混合协作—竞争环境中的多智能体演员—评论家”，载于神经信息处理系统进展，2017。[3] J. Foerster、G. Farquhar、T. Afouras、N. Nardelli 和 S. Whiteson，“反事实多智能体策略梯度”，载于*AAAI*人工智能会议，2018。[4] C. Yu 等，“PPO 在合作多智能体游戏中的惊人效果”，载于*NeurIPS*数据集与基准，2022。[5] S. Sukhbaatar、A. Szlam 和 R. Fergus，“通过反向传播学习多智能体通信”，载于神经信息处理系统进展，2016。[6] J. Jiang 和 Z. Lu，“学习用于多智能体合作的注意力通信”，载于神经信息处理系统进展，2018。[7] A. Das、T. Gervet、J. Romoff、D. Bouchard、L. V. Belanger 和 J. Pineau，“TarMAC：目标导向的多智能体通信”，载于国际机器学习会议，2019。[8] Y. Jiang、Y. Guo 和 S. Li，“用于多智能体合作的图卷积强化学习”，载于国际学习表征会议研讨会，2020。[9] P. Sunehag 等，“合作式多智能体学习的价值分解网络”，载于国际自主智能体与多智能体系统会议，2018。[10] T. Rashid 等，“QMIX：深度多智能体强化学习的单调价值函数分解”，载于国际机器学习会议，2018。[11] P. E. Hart、N. J. Nilsson 和 B. Raphael，“启发式确定最小代价路径的形式化基础”，*IEEE*系统科学与控制论汇刊，第4卷，第2期，第100–107页，1968。[12] O. Khatib，“机械臂与移动机器人实时避障”，国际机器人研究杂志，第5卷，第1期，第90–98页，1986。

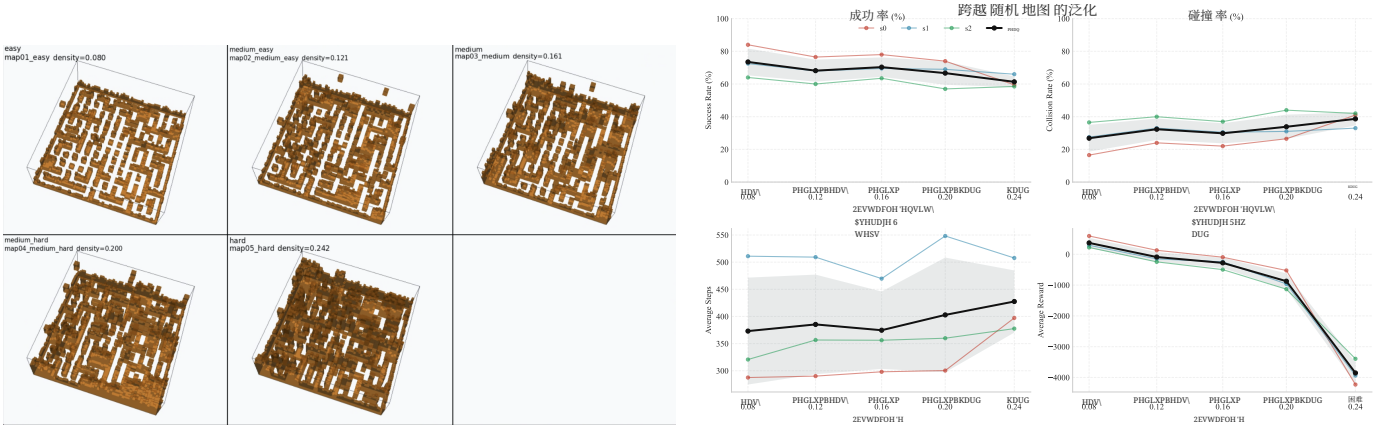


图 9. 在五张未见过的 UrbanCanyon 地图上进行零样本迁移（密度 0.08–0.24）。**左：**从容易到困难密度的地图示例。**右：**未经微调时，按随机种子统计及平均值计算的成功率、碰撞、平均步数和奖励。随着密度增加，OURS-LITE 的性能平稳下降；在密度为 0.24 时，成功率仍接近 61%，碰撞率低于 40%，这与结构迁移而非固定地图记忆的表现一致。

[13] P. Zarchan，战术与战略导弹制导，第6版。美国弗吉尼亚州雷斯顿：美国航空航天学会，2012。[14] F. Sadeghi 和 S. Levine，“CAD$^2$RL：无需真实图像即可实现真实的单图像飞行”，载于机器人学：科学与系统，2017。[15] L. Busoniu、R. Babuska 和 B. De Schutter，“多智能体强化学习综合综述”，*IEEE*《系统、人类与控制论汇刊》*C*辑，第38卷，第2期，第156–172页，2008。