

基于数字孪生的智能路口协同自动驾驶：多智能体强化学习方法

Taoyuan Yu*, Kui Wang*, Zongdian Li*, Tao Yu*, Kei Sakaguchi*, and Walid Saad†

*Department of Electrical and Electronic Engineering, Institute of Science Tokyo, Tokyo, TYO 152-8550, Japan

†Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA 22203, USA

Email: {yuty, kuiw, lizd, yutao, sakaguchi}@mobile.ee.titech.ac.jp, walids@vt.edu

摘要—无信号灯交叉口由于复杂的交通流和盲点而存在安全和效率挑战。本文提出了一种基于数字孪生 (DT) 的合作驾驶系统，该系统采用路边单元 (RSU) 为中心的架构，以提高无信号灯交叉口的安全性和效率。该系统利用全面的鸟瞰图 (BEV) 感知来消除盲点，并采用了结合离线预训练和在线微调的混合强化学习 (RL) 框架。具体来说，驾驶策略最初使用行为克隆 (BC) 在真实数据集上通过保守 Q 学习 (CQL) 进行训练，然后使用具有自我注意力机制的多智能体近端策略优化 (MAPPO) 进行微调以处理动态多智能体协调。RSU 通过车辆到基础设施 (V2I) 通信实现实时命令。实验结果表明，所提出的方法在协调多达三辆联网自动驾驶汽车 (CAVs) 时失败率低于 0.03%，显著优于传统方法。此外，该系统的计算扩展性呈次线性增长，推理时间小于 40 毫秒。进一步地，它展示了在各种无信号灯交叉口场景中的鲁棒泛化能力，表明其实用性和准备投入实际应用。

Index Terms—数字孪生，协作驾驶，智能运输系统，生成式人工智能模型，盲点消除。

I. 介绍

交叉口管理仍然是智能交通系统 (ITS) 中的关键瓶颈，这归因于交叉口的复杂性和不确定性 [1]。根据联邦公路管理局 (FHWA) 和国家公路交通安全管理局 (NHTSA)，与交叉口相关的死亡事故占交通事故死亡人数的重要部分，在 2024 年无信号灯交叉口导致的死亡占比达到 68% [2], [3]。盲点和模糊不清的互动规则使得无信号灯交叉口特别危险。为了解决这些问题，数字孪生 (DT) 的概念提供了一个有前景的解决方案，通过创建物理交叉口的实时虚拟复制品，提供了超越单个车辆有限感知能力的全局感知和智能协调 [4], [5]。

涉及自动驾驶车辆 (AVs) 和人工驾驶车辆 (HDVs) 的混合交通场景越来越普遍，从而增加了交通参与者之

间的协调复杂性。车联网 (V2X) 通信技术，包括车对车 (V2V)、车对基础设施 (V2I)、车对行人 (V2P) 和车对网络 (V2N)，可以帮助提高交通安全和效率 [6], [7]。在这些技术中，V2I 通信在 DT 系统中起着核心作用，因为它允许物理车辆与路侧单元 (RSUs) 之间的实时同步，从而支持协作驾驶策略，并将传统的交叉路口基础设施转变为智能控制中心 [8]。

利用 V2I 通信，DT 系统已被应用于具有各种架构的交叉口管理。例如，在 [9] 和 [10] 中，作者开发了基于 RSU 的 DT 系统，通过云计算进行连续交通监控和实时分析。然而，这些方法专注于一般的交通监控和基本感知增强，没有解决交叉口的盲点和遮挡问题。[11] 中的工作解决了交叉口遮挡问题，但 [11] 的解决方案受限于本地车辆感应，无法实现完全消除盲点。尽管 DT 技术有潜力提供全面的环境意识，当前的实施 [9]–[11] 既没有通过全局鸟瞰图 (BEV) 消除盲点，也没有支持遮挡区域中的协同驾驶策略。

为了充分利用全面的 DT 感知，已经在 [12]–[16] 中探索了各种交叉口协调算法。传统方法使用优化和博弈论算法来管理交通流 [12], [13]，但在动态环境中缺乏适应性。多智能体强化学习 (MARL) 已被提出作为在部分可观测情况下实现灵活且可扩展协调的有效解决方案 [14]–[16]。最近的研究进一步通过引入自注意力机制来改进 MARL，以增强跨智能体通信和决策制定。然而，大多数 MARL 模型在整个智能体中采用统一的策略，并未能建模如左转、直行或右转等不同的驾驶意图。此外，这些模型很少在各种车辆密度下进行评估，使其在动态交通条件下的鲁棒性不确定。关键的是，当

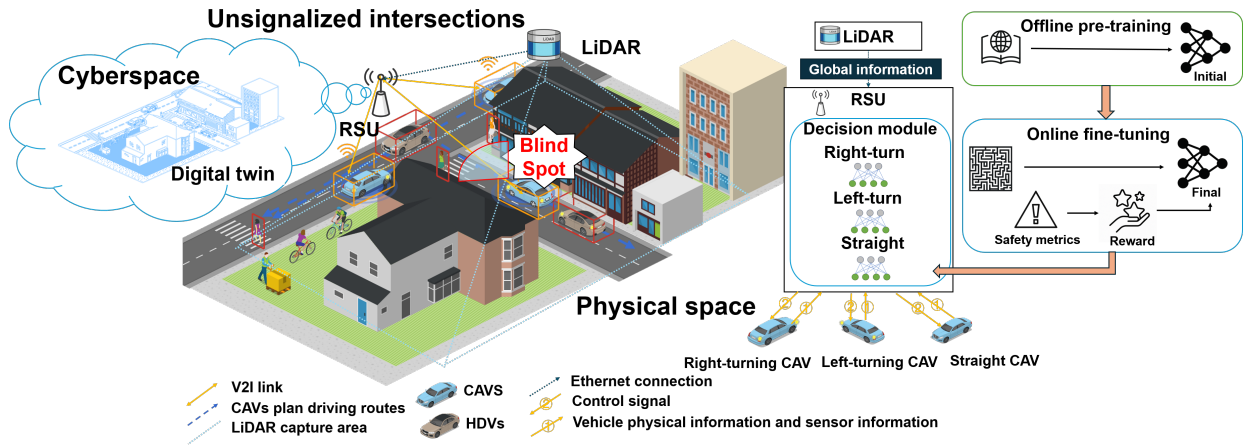


图 1. DT 协作系统的高层架构

前的 MARL 方法未能利用 DT 的全局感知来消除盲点，从而在 ITS 研究中留下了一个空白。

本文的主要贡献在于通过开发一种基于 DT 的无信号灯交叉口合作驾驶系统来解决上述限制。该系统利用安装在 RSU 上的激光雷达构建全面的 BEV 感知，以消除盲点，创建交叉口环境的实时数字副本。我们的方法利用带有角色特定策略网络和自我注意力机制的集中式 MARL 决策模块。此方法允许各种数量车辆实现稳健的合作驾驶。通过结合离线预训练与在线微调的混合学习框架，该系统开发了能够有效部署在真实世界场景中的决策能力。这种设计实现了盲点消除、系统适应性和交通效率方面的显著改进。总结来说，我们的关键贡献包括：

- 我们开发了一个基于 DT 的 MARL 框架，通过 RSU 全局感知消除无信号交叉路口的盲点。
- 我们引入了带有自注意力机制的角色特定策略网络，以实现连接的自动驾驶车辆（CAVs）之间的自适应协调。
- 我们提出了一种混合离线在线强化学习方法以确保策略学习的稳健性和效率。
- 我们进行了广泛的实验，展示了系统在各种场景中的有效性及泛化能力。

本文的其余部分组织如下：第二节介绍了 DT 系统架构。第三节详细描述了提出的算法。第四节讨论了实验结果。第五节总结了论文并概述了未来工作。

II. RSU-CAV 协作系统

我们考虑基于 DT 的协作驾驶系统架构如图 1 所示。该系统在物理交叉口与其在网络空间中的 DT 之间建立了实时同步。安装在 RSU 上的 LiDAR 传感器提供了全面的 BEV 感知，反过来可以帮助消除全球交通监控的盲点。与传统的以车辆为中心的方法（专注于单个车辆）不同，这个基于 DT 的系统为无信号灯交叉路口的多个 CAVs 提供集中式决策支持。RSU 的全局感知克服了单一车辆传感器的局限性，实现了旨在最小化潜在冲突并最大化交叉口吞吐量的合作驾驶策略。

为了有效管理交叉口的复杂性和不确定性，RSU 采用通过两阶段学习方法开发的决策策略。鉴于交通环境的动态和部分可观察特性，强化学习（RL）为在不确定性下的顺序决策提供了一个框架，该框架被建模为一个部分可观测马尔可夫决策过程（POMDP）。训练过程首先从真实世界的交通数据集上的离线 RL 开始，建立基础驾驶策略，然后在线上模拟环境中通过在线 RL 增强适应性和鲁棒性。这种混合范式确保了生成的策略能够在处理多样化的交通场景的同时保持安全约束。部署后，RSU 在其 DT 系统中利用这些训练好的策略进行实时决策，减少 CAVs 上的计算需求同时保证低延迟响应 [17]。

为了解决无信号交叉路口的盲点和有限的车载感知问题，我们引入了一个基于 DT 的合作系统。当 CAVs 接近交叉路口时，它们通过 V2I 通信同时在 DT 中表示。DT 保持物理车辆及其数字对应物之间的实时同步，使 RSU 能够根据完整的交通状态信息做出决策。使用

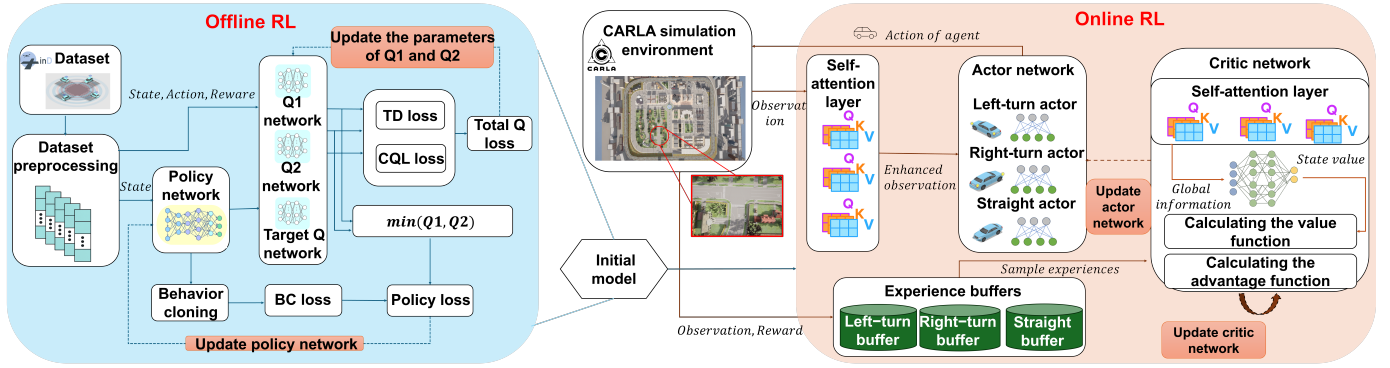


图 2. 离线-在线混合强化学习算法框架设计

实时数据，RSU 确定每辆车在 DT 中的驾驶角色。随后，RSU 利用集中式决策模块中预先加载的角色基础策略网络来计算控制信号。这些信号通过 V2I 通信实时传输到相应的 CAVs。同时，DT 持续监控交通状况，包括所有交通参与者的状态和预测移动、交通流的平滑性以及异常情况。物理空间与网络空间之间的这种同步为决策网络提供了必要的实时输入，并促进性能评估。

III. 混合强化学习框架

如图 2 所示，我们提出了一种基于数字孪生的两阶段学习框架，用于开发无信号控制交叉口的合作驾驶策略。该方法首先利用收集的数据集进行离线预训练，采用离线强化学习来获取基础驾驶技能和交通先验知识。随后，在 CARLA 模拟器 [18] 中进行在线微调，使代理能够适应动态环境。这种混合方法结合了离线强化学习的安全性和在线强化学习的适应性，确保训练模型能够在 RSU 的数字孪生系统中实现实时决策。本节详细介绍了离线预训练和在线微调的方法。

A. 观测空间

在每个时间步长 t ，状态空间 $\mathbf{s}(t)$ 包含了 RSU 监控到的所有交通参与者。RSU 使用全局信息为每个 CAV 构建个体观测向量 $\mathbf{o}(t)$ ，捕捉部分可观测且可能有噪声的表示：

$$\mathbf{o}(t) = [\mathbf{o}_{\text{core}}, \mathbf{o}_{\text{veh}}, \mathbf{o}_{\text{ped}}, \mathbf{o}_{\text{role}}, \mathbf{o}_{\text{ctx}}], \quad (1)$$

其中 $\mathbf{o}_{\text{core}}$ 包括自车速度、全局位置、航向角和交叉口占用情况； $\mathbf{o}_{\text{veh}}$ 包括附近车辆的相对位置和速度； $\mathbf{o}_{\text{ped}}$ 代表行人检测、距离和角度； $\mathbf{o}_{\text{role}}$ 编码代理的驾驶角色； $\mathbf{o}_{\text{ctx}}$ 包含场景标识符。

B. 动作空间

我们定义了一个统一的二维连续动作空间 $\mathcal{A}$ ，其结构如下：

$$\mathbf{a}(t) = [\mathbf{a}_{\text{acc}}, \mathbf{a}_{\text{steer}}] \in \mathbb{R}^2 \quad (2)$$

其中， $\mathbf{a}_{\text{acc}}$ 是纵向加速度， $\mathbf{a}_{\text{steer}}$ 是转向角速度。在离线预训练期间，动作是从连续的状态转换中估算出来的，因为真实的控制输入不可用。在线微调过程中，动作直接由策略网络预测。

C. 奖励函数

为了实现协同驾驶，我们设计了一个结构化的奖励函数 $\mathcal{R}_{\text{online}}(\mathbf{s}(t), \mathbf{a}(t), \mathbf{s}(t+1))$ 将高层次目标转化为实时反馈。总体奖励 $r(t)$ 定义为：

$$r(t) = \sum w_k r_k(\mathbf{s}(t), \mathbf{a}(t), \mathbf{s}(t+1)), \quad (3)$$

其中 r_i 代表个体奖励成分， w_i 捕获相应的权重。奖励项包括：

$$r_i \in \{r_{\text{safety}}, r_{\text{eff}}, r_{\text{comfort}}, r_{\text{task}}, r_{\text{yield}}, r_{\text{coop}}, r_{\text{penalty}}\} \quad (4)$$

其中 r_{safety} 基于最小时间至碰撞（TTC）等指标惩罚危险行为； r_{eff} 鼓励与交通流相兼容的速度； r_{comfort} 惩罚大幅度的加速度变化； r_{task} 奖励合作达成导航目标的代理； r_{yield} 和 r_{coop} 奖励遵守交通规则和合作行为； r_{penalty} 严重惩罚碰撞或超时。每一项都按其对应的权重 w_k 进行缩放，其中 w_{safety} 和 w_{penalty} 通常被赋予较大的值，因为它们具有关键的重要性。

D. 离线预训练：网络和算法

离线预训练的主要目标是为在线微调提供高质量的初始化。模型针对每个驾驶角色独立训练，使用的是基于车辆意图划分的 InD 数据集 [19] 的子集。

对于每个子集，我们在演员-评论家框架中采用结合保守 Q 学习 (CQL) 和行为克隆 (BC) [20], [21] 的离线 RL 算法。评论家使用双 Q 网络 $Q_{\theta_{i,1}}, Q_{\theta_{i,2}}$ 及其目标网络来稳定学习并减少高估，优化如下：

$$L_Q(\theta_{i,j}) = \mathbb{E}_{(\mathbf{o}, \mathbf{a}, r, \mathbf{o}') \sim \mathcal{D}_{\text{role}=i}} \left[\frac{1}{2} (Q_{\theta_{i,j}}(\mathbf{o}, \mathbf{a}) - y)^2 \right] + \alpha_{\text{CQL}} L_{\text{CQL_reg}}(\theta_{i,j}) \quad (5)$$

这里， $y = r + \gamma(1 - d) \min_j Q_{\theta'_{i,j}}(\mathbf{o}', \pi_{\phi_i}(\mathbf{o}'))$ 是时间差分 (TD) 目标。

策略网络 π_{ϕ_i} 最小化 BC 损失并最大化保守的 Q 值：

$$L_\pi(\phi_i) = \mathbb{E}_{\mathbf{o} \sim \mathcal{D}_{\text{role}=i}} \left[- \min_{j=1,2} Q_{\theta_{i,j}}(\mathbf{o}, \pi_{\phi_i}(\mathbf{o})) \right] + \lambda_{\text{BC}} \mathbb{E}_{(\mathbf{o}, \mathbf{a}) \sim \mathcal{D}_{\text{role}=i}} [\|\pi_{\phi_i}(\mathbf{o}) - \mathbf{a}\|^2] \quad (6)$$

其中， α_{CQL} 和 λ_{BC} 表示控制 CQL 正则化和 BC 模仿强度的超参数。

特定角色的演员网络 $\pi_{\phi_{\text{role}}}$ 和批评家网络 $Q_{\theta_{\text{role}}}$ 实现为多层感知器 (MLPs)。在这一阶段省略了自注意力机制以确保稳健的训练稳定性。生成的预训练权重在在线微调过程中被重复使用，以提高性能并加速适应。

E. 在线微调：网络和算法

在线微调采用多智能体近端策略优化 (MAPPO) [22]，结合角色特定网络 ($\pi_{\phi_{\text{left}}}, \pi_{\phi_{\text{straight}}}, \pi_{\phi_{\text{right}}}$) 与共享的批评家网络 V_ψ 。

为了捕捉动态交互，我们通过多头自注意力 (MHSA) 增强了演员和评论家网络。MHSA 允许模型同时关注来自不同表示子空间的不同位置的信息。缩放点积注意力定义为：

$$\text{A}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V \quad (7)$$

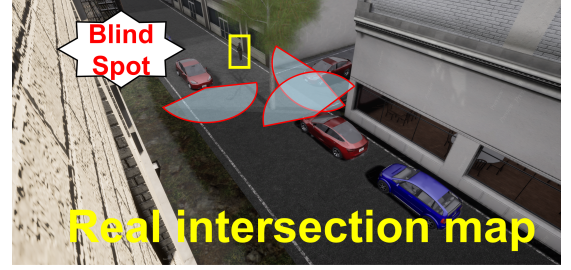
其中 Q 、 K 和 V 分别表示查询、键和值矩阵。MHSA 并行计算多个注意力头，并将它们的输出连接起来形成最终嵌入，捕捉观察特征之间的依赖关系。

在线学习通过一个互动-学习循环进行。智能体生成轨迹：

$$\tau = \{(\mathbf{o}_t, \mathbf{a}_t, r_{t+1}, V_\psi(\mathbf{o}_t), \log \pi_{\phi_{\text{role}}}(\mathbf{a}_t | \mathbf{o}_t))\}_{t=0}^T \quad (8)$$



(a)



(b)

图 3. 实验场景和泛化场景设置 (a) CARLA 示例地图, (b) 真实交叉路口地图



图 4. 离线预训练结果

优势估计 $\hat{A}_t^{\text{GAE}}$ 和返回值 $\hat{R}_t$ 使用广义优势估计 (GAE) 进行计算，基于从评论者价值计算出的时间差分 (TD) 误差 δ_t ：

$$\delta_t = r_{t+1} + \gamma V_\psi(\mathbf{o}_{t+1}) - V_\psi(\mathbf{o}_t) \quad (9)$$

优先经验回放 (PER) 根据与绝对 TD 误差成比例的优先级采样转换，并使用重要性采样 (IS) 权重来校正采样偏差：

$$w_t = \left(\frac{1}{B \cdot P(t)} \right)^\beta \quad (10)$$

其中 B 是回放缓冲区大小， β 控制 IS 校正强度。

每个特定角色的参与者 $\pi_{\phi_{\text{role}}}$ 都是使用以下加权目标进行训练的，该目标包括 PPO 截断代理损失和一个

熵奖励 $S[\cdot]$:

$$L^{\text{CLIP}+\text{S}}(\phi_{\text{role}}) = \mathbb{E}_{t \sim \text{PER}} \left[w_t \left(-L_t^{\text{CLIP}}(\phi_{\text{role}}) - c_2 \cdot S[\pi_{\phi_{\text{role}}}](\mathbf{o}_t) \right) \right] \quad (11)$$

PPO 替代损失函数 L_t^{CLIP} 定义为:

$$L_t^{\text{CLIP}} = \min \left(r_t \hat{A}_t, \text{clip}(r_t, 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \quad (12)$$

其中 ϵ 是 PPO 的剪辑超参数, 而 r_t 表示当前策略与旧策略之间的概率比率。

IV. 实验与分析

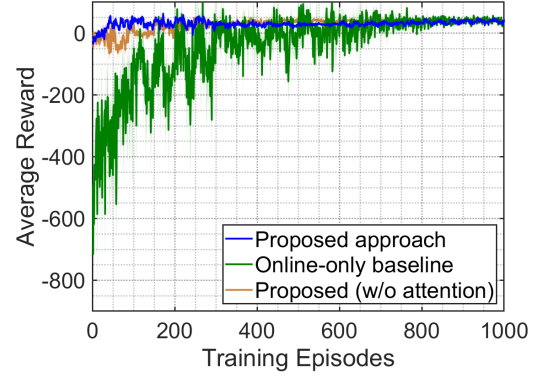
实验在同步模式下使用带有 Unreal Engine 的 CARLA 模拟器进行。主要测试场景是 Town03 中的一个无信号灯交叉口, 如图 3 所示。在每个试验中, 我们的系统控制 1 到 3 辆 CAV (红色), 而背景车辆 (蓝色) 由 CARLA 的交通管理器控制。添加行人以模拟真实的城区环境。为了评估泛化能力, 我们在基于东京理科大学校园的真实交叉口地图上部署模型。RSU 通过 BEV 感知维持全局状态, 并使用微调后的决策模型计算控制命令, 这些命令通过模拟的 V2I 通信发送给 CAVs。

A. 基线和评估指标

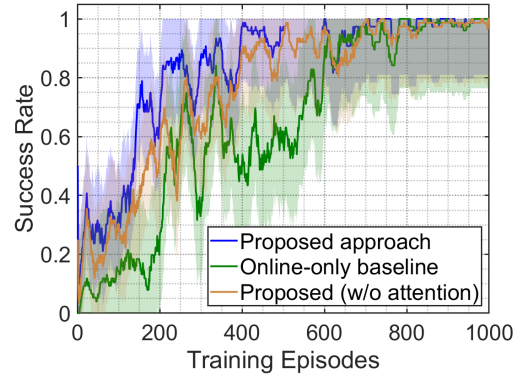
为了评估每个组件的贡献, 我们将我们的模型与几个基线进行了比较。首先, 考虑了两个消融变体: (1) 直接训练的仅在线 MAPPO 基线; (2) 一个带有离线预训练但不带自注意力或角色特定策略的变体。两者都与完整模型共享相同的架构和超参数, 分别隔离了离线预训练和自注意力的影响。其次, 我们包括了 Autoware Universe [23], 这是一个基于规则的自动驾驶堆栈, 配置为控制单辆车辆。所有方法均根据收敛速度、失败率和平均旅行时间进行了评估。

B. 离线预训练结果

离线预训练阶段旨在从 InD 数据集中提取驾驶先验, 以初始化模型进行在线微调。图 4 显示了训练过程中的 Q1/Q2 损失和奖励改进。稳步收敛的损失表明状态-动作值学习稳定, 而奖励指标则稳定在约 112%, 超过了 100% 的基础线。这证实通过将 CQL 与 BC 结合所学到的策略不仅能模仿, 还能超越数据集行为平均表现, 为在线阶段提供了强大的初始化。



(a)



(b)

图 5. 不同方法的训练性能比较, (a) 奖励, (b) 成功率。

表 I
性能比较总结

方法 / 场景	失败率 (%)	平均时间 (秒)
Ours (1 Agent, Town03)	0.01	5.52
Ours (2 Agent, Town03)	0.03	5.49
Ours (3 Agent, Town03)	0.02	5.25
Autoware (1 Agent, Town03)	5.31	5.77
Ours (3 Agent, Real Map)	0.02	5.15

C. 在线训练结果

图 5 展示了我们提出的模型及其两种消融变体的训练收敛情况。完整模型始终优于所有基线模型。它在大约 250 个回合内达到稳定性能, 而仅在线基线需要超过 800 个回合才能收敛。没有自注意力和角色特定策略的消融变体尽管受益于离线预训练, 在约 500 个回合后才收敛。这一比较表明这两个组件对于实现最佳性能至关重要。离线预训练加速了学习并提高了初始性能, 而自注意力和角色特定策略进一步增强了多智能体协调的有效性并维持其效果。这些结果证实我们的混合方法结

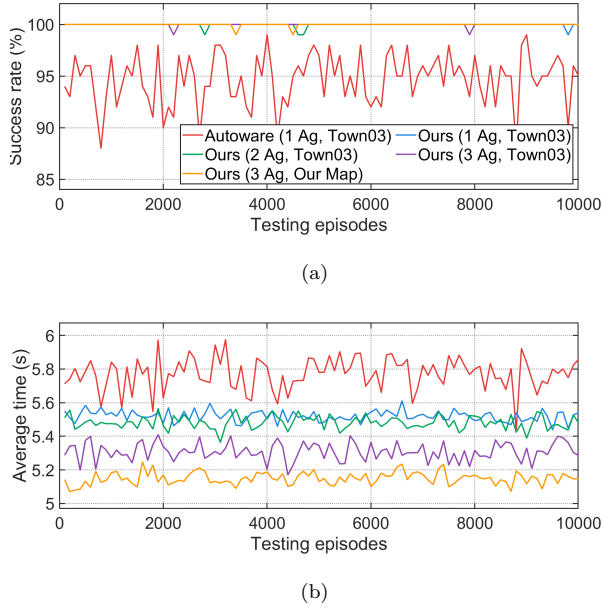


图 6. 最终模型性能评估结果 (a) 通过测试阶段的成功率, (b) 通过测试阶段的平均旅行时间

合了离线 RL 的安全性和复杂多智能体在交叉口协作所需适应性的需求。

D. 性能评估与泛化分析

我们通过在 Town03 交叉路口和实际交叉路口地图上进行 10,000 次性能测试评估了该模型, 并将其与基线进行了比较。关键绩效指标汇总于图 6 和表 I 中, 其中失败率表示以碰撞或超时结束的会话的百分比。我们的模型在 Town03 的所有测试场景中都表现出高安全性和可靠性。当控制单个车辆时, 它实现了 0.01% 的失败率, 超过了 Autoware 基准 5.31% 的失败率。值得注意的是, 随着协调复杂性的增加, 我们的系统没有显示出明显的性能下降。具体来说, 在两车场景中的失败率为 0.03%, 在三车场景中的失败率为 0.02%。BEV 视角和自注意力机制的结合为这种稳健性做出了贡献, 展示了我们模型在处理复杂的多智能体协作任务方面的有效性。

此外, 我们的模型在交通效率方面的性能优势也得到了体现。单辆车场景下的平均行驶时间为 5.52 秒, 而 Autoware 为 5.77 秒。随着受控车辆数量的增加, 平均行驶时间略有下降, 表明多智能体有效协调, 建立了高效的协同驾驶策略, 实际上提高了交叉路口的通行能力。

为了泛化, 训练于 Town03 的三车模型被部署到真实的交叉路口地图上。它在这个新环境中实现了 2% 的失败率和 5.15 秒的平均旅行时间。这表明 BEV 有效消除了单个车辆盲点的影响。这一结果验证了我们模型出色的泛化能力, 并为该方法的实际应用提供了坚实的基础。

为了进一步验证系统的可部署性, 我们评估了其在不同协调场景下的计算性能。实验在 NVIDIA RTX 4070 Ti GPU 上以 10 Hz 的控制频率进行。单辆车的平均推理时间为 23.7 毫秒, 两辆车为 31.4 毫秒, 三辆车为 38.2 毫秒, 所有测试中的最大推理时间为 42.6 毫秒。这种亚线性扩展确认了高效的多智能体处理。即使在最坏的情况下, 推理时间仍然远低于 100 毫秒的控制间隔, 为 V2I 通信和安全检查留出了足够的余地, 验证了系统的实时可部署性。

V. 结论与未来工作

在本文中, 我们提出了一种基于 DT 的无信号交叉口以 RSU 为中心架构的合作驾驶系统。该系统利用 BEV 感知消除盲点, 并采用混合强化学习算法实现鲁棒的多智能体合作驾驶策略。我们开发了特定角色的政策并在多种场景下验证了系统, 实现了 0.03% 的故障率和最多三个 CAVs 的亚 40 毫秒推理时间。未来的主要工作包括概念验证 (PoC) 实验以全面验证该系统的实际性能。

参考文献

- [1] K. Chu, A. Lam and V. Li, "Traffic Signal Control Using End-to-End Off-Policy Deep Reinforcement Learning," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7184-7195, 2022.
- [2] U.S. Department of Transportation. [Online]. Available: <https://highways.dot.gov/sites/fhwa.dot.gov/files/2024-08>
- [3] National Highway Traffic Safety Administration. [Online]. Available: <https://www.nhtsa.gov/press-releases/nhtsa-2023-traffic-fatalities-estimate-april-2024>
- [4] K. Wang et al., "Smart Mobility Digital Twin Based Automated Vehicle Navigation System: A Proof of Concept," in *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 3, pp. 4348-4361, 2024.
- [5] O. Hashash, C. Chaccour, W. Saad, T. Yu, K. Sakaguchi and M. Debbah, "The Seven Worlds and Experiences of the Wireless Metaverse: Challenges and Opportunities," in *IEEE Communications Magazine*, vol. 63, no. 2, pp. 120-127, 2025.

- [6] K. Wang, C. She, Z. Li, T. Yu, Y. Li, and K. Sakaguchi, "Roadside Units Assisted Localized Automated Vehicle Maneuvering: An Offline Reinforcement Learning Approach," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1709–1715, 2024.
- [7] Z. Li, K. Wang, T. Yu, and K. Sakaguchi, "Het-SDVN: SDN-Based Radio Resource Management of Heterogeneous V2X for Cooperative Perception," *IEEE Access*, vol. 11, pp. 76255–76268, 2023.
- [8] D. Suo, B. Mo, J. Zhao, and S. E. Sarma, "Proof of Travel for Trust-Based Data Validation in V2I Communication," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9565–9584, 2023.
- [9] Y. Cui, H. Xu, J. Wu, Y. Sun and J. Zhao, "Automatic Vehicle Tracking With Roadside LiDAR Data for the Connected-Vehicles System," *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 44–51, 2019.
- [10] L. Wang et al., "Multi-Modal 3D Object Detection in Autonomous Driving: A Survey and Taxonomy," in *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 7, pp. 3781–3798, 2023.
- [11] K. Moller, R. Trauth, and J. Betz, "Overcoming Blind Spots: Occlusion Considerations for Improved Autonomous Driving Safety," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, pp. 819–826, 2024.
- [12] Y. Zhu, Z. He and G. Li, "A bi-Hierarchical Game-Theoretic Approach for Network-Wide Traffic Signal Control Using Trip-Based Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15408–15419, 2022.
- [13] M. Gallo, "Combined Optimisation of Traffic Light Control Parameters and Autonomous Vehicle Routes," *Smart Cities*, no. 3, pp. 1060–1088, 2024.
- [14] Y. Shi, H. Dong, C. He, Y. Chen and Z. Song, "Mixed Vehicle Platoon Forming: A Multiagent Reinforcement Learning Approach," *IEEE Internet of Things Journal*, vol. 12, no. 11, pp. 16886–16898, 2025.
- [15] S. Iqbal and F. Sha, "Actor-Attention-Critic for Multi-Agent Reinforcement Learning," *arXiv preprint arXiv:1810.02912*, 2019.
- [16] R. Younas, H. M. Raza Ur Rehman, I. Lee, B. W. On, S. Yi and G. S. Choi, "SA-MARL: Novel Self-Attention-Based Multi-Agent Reinforcement Learning With Stochastic Gradient Descent," in *IEEE Access*, vol. 13, pp. 35674–35687, 2025.
- [17] K. Wang, T. Yu, Z. Li, K. Sakaguchi, O. Hashash, and W. Saad, "Digital Twins for Autonomous Driving: A Comprehensive Implementation and Demonstration," in *2024 International Conference on Information Net-working (ICOIN)*, pp. 452–457, 2024.
- [18] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, pp. 1–16, 2017.
- [19] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater and L. Eckstein, "The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections," in *2020 IEEE Intelligent Vehicles Symposium(IV)*, p. 1929–1934, 2020.
- [20] A. Kumar, A. Zhou, G. Tucker, and S. Levine, "Conservative Q-Learning for Offline Reinforcement Learning," *arXiv preprint arXiv:2006.04779*, 2020.
- [21] D. A. Pomerleau, "ALVINN: An Autonomous Land Vehicle in a Neural Network," in *Advances in Neural Information Processing Systems*, 1988.
- [22] C. Yu, A. Velu, E. Vinitzky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games," *arXiv preprint arXiv:2103.01955*, 2022.
- [23] Autoware. [Online]. Available: <https://autoware.org/>