arXiv:2603.03848v1 [eess.SY] 4 Mar 2026

用于缓解混合交通拥堵的双重交互感知协同控制策略

Zhengxuan Liu, *IEEE* 学生会员，Yuxin Cai,*IEEE* 学生会员，Yijing Wang, XiangkunHe, *IEEE* 高级会员，Chen Lv, *IEEE* 高级会员and Zhiqiang Zuo, *IEEE* 高级会员

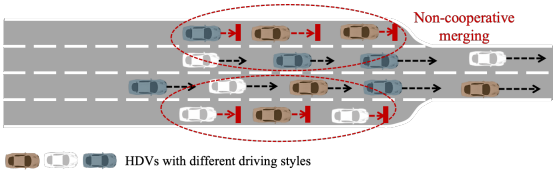
摘要——随着智能交通系统（ITS）的发展，网联自动驾驶车辆（CAV）有望通过协作策略显著缓解交通拥堵，例如在瓶颈区域。然而，在混合交通环境中，人工驾驶车辆（HDV）行为的不确定性和多样性给网联自动驾驶车辆的协作带来了重大挑战。本文提出一种双交互感知协同控制（DIACC）策略，在多智能体强化学习（MARL）框架下增强网联自动驾驶车辆在混合交通瓶颈场景中的局部和全局交互感知。DIACC策略包含三项关键创新：1）分散式交互自适应决策（D-IADM）模块，通过区分 CAV-CAV 协作交互与 CAV-HDV 观测交互，增强行动者的局部交互感知。2）集中式交互增强评论家（C- IEC），通过交互感知价值估计提升评论家的全局交通理解，从而为策略更新提供更准确的指导。3）奖励设计采用 softmin 聚合与温度退火，优先关注混合交通中的交互密集场景。此外，轻量级的基于主动安全的动作优化（PSAR）模块运用基于规则的修正，加快训练收敛。实验结果表明，与基于规则的方法和基准多智能体强化学习模型相比，DIACC 显著提高了交通效率和适应性。

索引词—混合交通、协同控制、交互感知、多智能体强化学习、交互密集型场景。

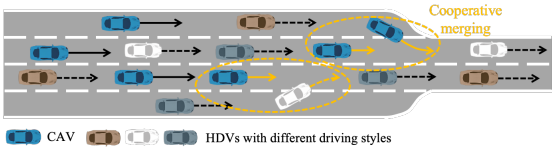
一、引言

WITH 随着车辆保有量快速增长，交通拥堵已成为一项严峻挑战，而网联自动驾驶车辆（CAVs）已展现出提升交通性能的巨大潜力 [1], [2]。然而，在可预见的未来，人工驾驶车辆（HDVs）仍将与网联自动驾驶车辆共存 [3]。人类驾驶员的驾驶风格多样且难以预测，这对网联自动驾驶车辆构成了根本性挑战：它们必须适应不确定的人工驾驶车辆行为，同时协调车辆以实现整体交通优化。

*本研究得到国家自然科学基金委员会资助（项目编号：62173243和61933014）。Zhengxuan Liu、Yijing Wang和Zhiqiang Zuo任职于天津大学电气与信息工程学院天津市智能无人集群技术与系统重点实验室，中国天津300072。（电子邮箱：liuzhengxuan@tju.edu.cn；yjwang@tju.edu.cn；zqzuo@tju.edu.cn）Yuxin Cai和Chen Lv任职于新加坡南洋理工大学机械与航空航天工程学院。（电子邮箱：caiy0039@e.ntu.edu.sg；lyuchen@ntu.edu.sg）Xiangkun He任职于中国电子科技大学深圳先进研究院。（电子邮箱：xiangkun.he@uestc.edu.cn）



(a) 在纯人工驾驶车辆交通中，人类驾驶员通常优先考虑自身的驾驶效率，因此会采取非合作行为，并导致合流区域拥堵。不同颜色表示具有不同驾驶风格的人工驾驶车辆。



(b) 在混合交通中，网联自动驾驶车辆会采取协作合流行为，为相邻车辆变道
创造机会，从而缓解拥堵。蓝色车辆表示网联自动驾驶车辆。

图1：瓶颈场景中纯人工驾驶车辆交通的非合作合流与网联自动驾驶车辆参与的协作合流示意图。

本文聚焦于高密度混合交通下的瓶颈合流场景，其中驾驶风格各异的人工驾驶车辆与协同控制的网联自动驾驶车辆共存（如图 1 所示）。车辆交互是决定瓶颈场景交通性能的关键因素。在完全由人工驾驶车辆组成的交通流中，缺乏协作会导致合流点严重拥堵；而采用协作策略的网联自动驾驶车辆能够提前规划合流路径，并礼让周围车辆，从而提升整体效率。然而，人工驾驶车辆行为的不确定性和多样性大大增加了车辆交互建模的复杂度，给混合交通中的网联自动驾驶车辆协同控制带来了核心挑战。

针对瓶颈合流场景中网联自动驾驶车辆的协同控制，现有解决方案可分为两类。第一类包括基于规则的方法或基于优化的方法 [4]–[8]，第二类则依赖于强化学习（RL） [9]–[12]。集中式优化方法在解决全局交通问题方面具有优势，但其计算成本相对较高 [13]。此外，这些方法往往难以应对复杂的交通环境和车辆交互，因为它们严重依赖交通系统和车辆运动的模型 [5]–[7]。相比之下，基于强化学习的方法通常是无模型的，通过与环境的交互学习有效策略。它们在处理混合交通环境的不确定性和复杂性方面具有优势

。然而，单智能体强化学习算法在为混合交通中的网联自动驾驶车辆设计有效协作策略方面存在局限。它们更适合解决局部控制问题 [9]。多智能体强化学习（MARL）通过交互与信息共享促进多个智能体之间的协同，从而改进协作策略的学习并提高训练效率，提供了一种更为稳健的解决方案。因此，对于本文所讨论的瓶颈场景中网联自动驾驶车辆的协同控制，MARL是一种适当且有效的方案。然而，在分散式多智能体强化学习框架中，网联自动驾驶车辆之间的协作仅限于局部观测，从而限制了全局交通效率的提升。事实上，车辆交互对交通性能的影响是一个全局问题，而现有的MARL方法侧重于局部车辆交互，缺乏对全局交通动态的考虑。

如上所述，车辆交互是影响瓶颈场景中交通性能的关键因素，但现有方法对此仍缺乏充分考虑。为弥补这一不足，我们提出了一种双重交互感知协同控制（DIACC）策略，在多智能体强化学习框架下，从局部和全局视角增强交互感知能力。本文的主要贡献如下：

- 为解决行动者无法区分异构车辆交互（即网联自动驾驶车辆和人工驾驶车辆）的问题，我们提出了一种分散式交互自适应决策（D-IADM）模块。其核心设计采用独立的图注意力网络，分别处理网联自动驾驶车辆之间的协作交互（纳入上一时间步的决策反馈）与网联自动驾驶车辆和人工驾驶车辆之间仅依赖历史轨迹的观测交互。实验结果表明，该模块提升了训练稳定性和交通效率，并减少了安全关键事件。

- 提出了集中式交互增强评论家（C-IEC），使评论家能够在训练过程中捕捉车辆交互对全局交通动态的影响。这可以提供更准确的价值估计，引导 D-IADM 制定兼顾系统级交通动态的协作策略。这一性能优势在高密度场景中尤为显著。

- 本文提出了一种采用 softmin 聚合和温度退火的协作奖励机制。该机制从均匀探索逐渐转向针对困难案例的优化，从而将训练重点集中在具有挑战性的交互场景上。由此带来的训练稳定性和收敛性提升，使 D-IADM 能够在各种交互场景中制定更稳健的协作策略。

本文其余部分的结构安排如下。第二节回顾最新的网联自动驾驶车辆协同控制策略。第三节给出本研究的基础数学描述。第四节介绍 DIACC策略的框架，并详细说明各模块的设计。实验设计及相应分析在第五节中展开，同时讨论研究发现。第六节对本文进行总结。

为便于解释本文设计的协同控制策略，表 I 列出了两类缩写：策略中定义的术语缩写，以及后续实验评估中使用的性能指标缩写。

表 I：缩写。

全称	缩写
集中式训练与分散式执行	CTDE
双交互感知协同控制	DIACC
分散式交互自适应决策	D-IADM
轨迹感知交互编码器	TAIE
基于主动安全的动作优化	PSAR
集中式交互增强评论家	C-IEC
综合交通动态表示	ITDR
带有D-IADM模块的MAPPO模型	MAPPO-IADM
安全关键事件	SCE
等待事件	WE

II. 相关工作

研究人员长期以来一直关注网联自动驾驶车辆的协同控制问题。现有研究结果表明，网联自动驾驶车辆协同控制能够显著提升交通性能，例如交通安全、交通效率和能耗降低。本节总结了最新的网联自动驾驶车辆协同控制策略，包括基于控制理论的方法、基于单智能体强化学习的方法和基于多智能体强化学习的方法。我们分析了这些方法的特点，从而阐明其优势与局限。

A. 基于控制理论的方法

基于控制理论的方法通常依托交通流理论（如宏观和微观交通流模型） [5], [14] 及优化算法，例如模型预测控制（MPC） [8], [15]。在考虑局部或微观层面的优化目标时，重点关注跟驰速度和换道间隙 [4],[16],，并根据这两个因素优化车辆轨迹。而在全局或宏观层面，通常主要关注交通流的平均速度和流量 [5]。在这些研究中，车辆行为由人工设计的驾驶规则或目标函数 [15],[17]引导。集中式优化策略能够为提升交通效率等全局交通问题提供最优解，但计算成本较高 [18], [19]。基于控制理论的方法的另一个重要局限是难以处理复杂的交通环境和车辆交互，因为这类方法通常依赖于对交通环境和车辆行为的建模 [20], [21]。然而，要准确建模复杂的交通演化和车辆交互颇具挑战。

B. 基于单智能体强化学习的方法

基于强化学习的方案在处理环境不确定性和复杂性方面具有优势。它不依赖精确的环境建模，而是通过与环境交互来学习最优决策策略 [10]。强化学习通过设计

奖励函数和训练损失来引导学习过程，以实现预期目标 [11], [22]。在协同控制中，研究人员主要关注两个关键问题：理解交通环境中的车辆交互，以及通过通信增强车辆之间的协作。文献 [3],提出了一种多视图图卷积强化学习（MVGRL）算法，用于处理交通参与者之间的交互及其导致的复杂交通状态。 [23] 提出了参数共享机制和密集通信拓扑，使网联自动驾驶车辆能够应对交通波动并提高驾驶效率。然而，需要注意的是，在处理混合交通中网联自动驾驶车辆之间的协作时，单智能体强化学习方法仍存在一些不足。一方面，单智能体方法更适合解决局部控制问题 [9],，但在全局交通优化方面面临挑战，需要大量交互经验才能学习有效的决策策略。这导致训练成本高昂，也增加了大规模部署的难度。另一方面，这种方法难以处理多车辆之间的协调问题，通常需要额外的规则引导或人工设计的协调交互设置。例如， [24] 通过学习车辆交互来评估驾驶员之间的合作程度，并将这一合作信念状态作为强化学习算法的输入。

C. 基于多智能体强化学习的方法

多智能体强化学习有助于对交通环境中智能体之间的交互进行建模和分析 [25], ，其灵活的协作与共享机制也推动了协同控制研究。目前，多智能体强化学习方法已应用于多种交通优化问题，例如交通信号灯控制 [26], [27],可变限速控制 [9], 多车协同控制 [28], [29],以及车路协同控制 [30], [31]。多智能体强化学习的优势体现在多个方面。首先，它通过多个智能体之间的交互和信息共享促进协作 [10], [11], ，从而能够学习最优协同策略。其次，多智能体强化学习能够应对环境中的非平稳性，尤其是在智能体数量增加时 [32], [33]。第三，多智能体强化学习在收敛性、最优性和稳定性方面表现出优势 [34]。

在交通系统中，多智能体强化学习为实现高度智能化和适应性强的交通管理带来了新的可能。然而，由于多个智能体之间交互的复杂性，用于多车协同控制的多智能体强化学习算法设计仍处于初期阶段。针对多车协作， [35] 采用多智能体深度确定性策略梯度算法，训练主线和匝道上的多辆网联自动驾驶车辆； [11] 则提出多智能体优势演员-评论家算法，训练合流区域内的所有网联自动驾驶车辆。此外， [10] 通过通信协作和参数共享，为车辆设计了协作式换道行为。这些研究主要聚焦于多智能体强化学习算法中个体智能体的设计，但尚未充分探索评论家的引导作用。同时，它们忽略了车辆交互对交通优化的全局影响。

在涉及集中训练、分散执行的多智能体强化学习框架下，本文强调从局部和全局视角综合考量车辆交互的重要影响。基于局部交互感知，我们设计了 D-IADM，以增强网联自动驾驶车辆对行为不可预测且多样化的人工驾驶车辆的适应性。为更好地引导车辆协作，本文提出 C-IEC，借助全局交互感知研究车辆交互对交通效率的影响。通过这种方式，我们显著提升了瓶颈场景下的交通效率，并增强了多智能体强化学习模型对不同场景的适应性。

III. 问题表述

A. 协同决策问题建模

在本研究中，我们将混合交通环境下瓶颈合流场景中网联自动驾驶车辆的协同决策过程建模为一个多智能体强化学习问题，该问题由分散式部分可观测马尔可夫决策过程（Dec-POMDP） [36]表征。多智能体强化学习模型可表示为一个元组（$\mathcal{N}$， $\mathcal{S}$， $\boldsymbol{\mathcal{A}}$， $\boldsymbol{\mathcal{O}}$ R， P， γ），其中：

- $\mathcal{N} = \{1, \ldots, n\}$ 表示混合交通环境中的网联自动驾驶车辆集合。
- $\mathcal{S}$ 表示环境的全局状态空间。
- $\boldsymbol{\mathcal{A}} = \prod_{i=1}^{n} \mathcal{A}^i$ 表示所有网联自动驾驶车辆的联合动作空间。
- $\boldsymbol{\mathcal{O}} = \prod_{i=1}^{n} \mathcal{O}^i$ 表示所有网联自动驾驶车辆的联合观测空间。
- $R: \boldsymbol{\mathcal{O}} \times \boldsymbol{\mathcal{A}} \to \mathbb{R}$ 表示联合奖励函数。
- $P(s_{t+1}|s_t, a_t^1, \ldots, a_t^n)$: $\mathcal{S} \times \boldsymbol{\mathcal{A}} \times \mathcal{S} \to [0, 1]$ 定义状态转移概率函数。
- $\gamma \in [0, 1)$为折扣因子，用于权衡即时奖励与未来奖励的重要性。

在此设定中，每辆网联自动驾驶车辆 i 在任意时间步 t 观测其局部状态 $o_t^i \in \mathcal{O}^i$ ，并根据由 θ参数化的策略$\pi_\theta^i(\cdot|o_t^i)$选择动作 a_t^i 。

所有网联自动驾驶车辆的共同目标是优化团队策略 π_θ^*，以最大化在 L个时间步范围内的期望累积奖励，即，

$$J(\pi_\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{L} \gamma_t R_t\right]. \tag{1}$$

B. 多智能体演员-评论家框架

在本小节中，我们采用基于集中训练、分散执行（CTDE）的多智能体强化学习（MARL），其包含两个关键部分：集中式评论家网络和 n 分散式演员网络。前者通过获取所有智能体的全局信息来辅助训练阶段，从而缓解环境的非平稳性并提高学习效率。同时，每个分散式演员网络都力求最大化自身的个体奖励，并受益于评论家提供的全局洞见。一个广为人知的代表方法是多智能体近端策略优化（MAPPO） [37], ，它将近端策略优化（PPO） [38] 扩展至多智能体场景。MAPPO 能够适应多个交互智能体带来的复杂性，因此特别适用于制定分散式控制策略，使每个智能体都能独立地与环境及其他智能体交互。

在 MAPPO 中，每个智能体的演员网络 π_θ^i 将其局部观测 o_t^i 映射为离散动作空间 $\mathcal{A}^i$上的分类分布，旨在优化其策略 π_θ^i（$a_t^i|o_t^i$）以最大化期望奖励。具体而言，演员网络专注于优化裁剪代理目标 [38],

$$L^{CLIP}(\theta) = \mathbb{E}\left[\min\left(r_t^i(\theta)\hat{A}_t, \text{clip}(r_t^i(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \tag{2}$$

其中，$r_t^i(\theta) = \frac{\pi_\theta(a_t^i|o_t^i)}{\pi_{\theta_{\text{old}}}(a_t^i|o_t^i)}$ 表示新旧策略之间的概率比。由超参数 ϵ 控制的裁剪机制可防止策略偏离旧策略过远，从而确保更新的稳定性。在此目标函数中，优势估计器 A_t 根据广义优势估计器（GAE）计算，该方法在准确估计优势时实现偏差-方差权衡。联合优势估计的计算方式如下:

$$\hat{A}_t^{GAE}(\gamma,\lambda) = \sum_{l=0}^{T-t}(\gamma\lambda)^l\delta_{t+l}, \tag{3}$$

其中，λ 是 GAE 参数，T 是回合长度。δ_t 是时序差分误差，其形式为

$$\delta_t = R_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t), \tag{4}$$

其中，R_t 是即时奖励，V_ϕ 是集中式评论家的价值函数（详见下文）。时序差分误差量化实际回报与评论家估计值之间的差异，该差异通过 GAE 累积，以计算用于策略更新的优势。

除了裁剪目标函数外，还在演员损失中加入熵项，以防止过早收敛并促进充分探索:

$$L(\theta) = L^{CLIP}(\theta) + \sigma\mathbb{E}\left[H(\pi_\theta(a_t^i|o_t^i))\right] \tag{5}$$

其中，H 表示策略的熵，σ 为熵系数。引入熵奖励有助于探索，并避免对特定场景特征过拟合，从而促进动作多样性，并增强在相似状态下的泛化能力。

尽管每个行动者都基于各自的观测独立优化其局部策略，集中式评论家会利用全局状态信息来指导训练过程。集中式评论家记为 V_ϕ，执行以下映射：$\mathcal{S} \to \mathbb{R}$。它提供全局价值函数 $V_\phi(s_t)$，用于评估当前状态 s_t，并指导每个智能体的策略更新。为稳定评论家的更新，参数化为 ϕ 的集中式评论家也通过裁剪后的值预测进行训练:

$$V^{CLIP}(\phi) = \text{clip}(V_\phi(s_t), V_{\phi_{old}}(s_t)-\epsilon, V_{\phi_{old}}(s_t)+\epsilon) \tag{6}$$

价值函数损失的形式为

$$L_V(\phi) = \mathbb{E}\left[\max\left((V_\phi(s_t)-\hat{R}_t)^2, (V^{CLIP}(\phi)-\hat{R}_t)^2\right)\right] \tag{7}$$

其中，R_t 表示折扣未来累计奖励 $\sum_{t'=t}^{T}\gamma_{t'}R(s_{t'}, a_{t'}, s_{t'+1})$，即预期的未来回报。

总而言之，我们采用 MAPPO 框架来学习去中心化策略（π_θ^1，π_θ^2，...，π_θ^n），其中每个智能体的独立策略更新均根据目标函数（2）进行裁剪。随后，采用集中式评论家估计价值函数，为各智能体提供共享信息，以稳定学习并改善协同。然而，将原始多智能体近端策略优化应用于混合交通协同控制时，暴露出若干局限性，这促使我们在下一节提出 DIACC策略:

- 行动者不会显式建模智能体之间的交互，这在混合交通中尤具挑战性，因为CAV-CAV协作交互和CAV-HDV观测交互具有根本不同的特征。为此，我们提出D-IADM模块，专注于对混合交通环境中网联自动驾驶车辆的局部交互关系进行建模。

- 尽管集中式评论家能够获取全局信息，但缺乏充分利用这些信息的有效机制，因此难以获得准确的价值估计。为此，我们提出 C-IEC，通过利用全局车辆交互与交通动态之间的关系映射，提供更准确的价值估计，以指导策略更新。

- 标准奖励设计未能充分强调智能体面临复杂协同挑战的高交互场景，因此可能忽略瓶颈汇入中的关键情况。为解决这一问题，我们提出一种采用 softmin 聚合和温度退火的协作奖励机制，引导训练关注困难的交互场景，并增强学习鲁棒性。

IV. 面向网联自动驾驶车辆的双交互感知协同控制策略

基于 MAPPO 框架，我们提出面向网联自动驾驶车辆的双交互感知协同控制（DIACC）策略，以提升瓶颈场景下协同控制的有效性。本节详细介绍该策略的设计与实现。

gy.

A. 框架

所提出的协同控制模型结合了多智能体行动者-评论家框架，其中包括若干去中心化交互自适应决策（D-IADM）模块和一个集中式交互增强评论家（C-IEC），如图 2 所示。在训练过程中，每辆网联自动驾驶车辆从环境中获取局部观测，并通过 D-IADM 输出控制动作。D-IADM 首先通过轨迹感知交互编码器（TAIE）获取自车与周围车辆之间的交互特征。随后，它将上一次决策输出 a_{t-1}^i 与门控循环单元（GRU） [39] 及决策层结合，生成可学习的动作输出 $\hat{a}_t^i$。GRU 捕捉不同时间步观测特征之间的时间依赖，从而增强决策的连续性。输出 $\hat{a}_t^i$ 随后被送入 PSAR 模块；该模块是一个轻量级动作优化层，用于在变道和跟车操作期间监测车间距和碰撞时间（TTC）。

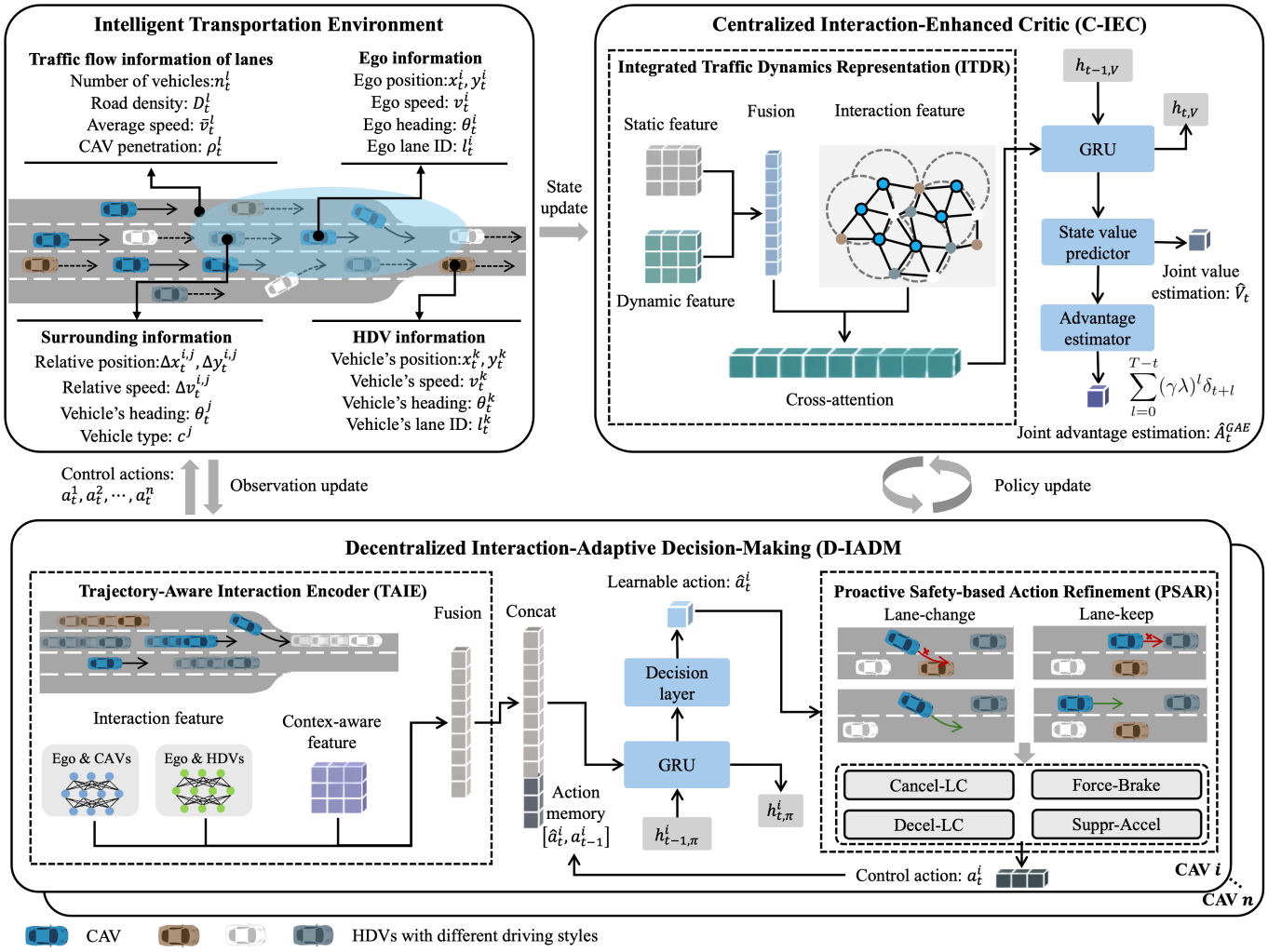


图 2：DIACC策略框架。

当车间距或碰撞时间（TTC）低于安全阈值时，PSAR 会对 $\hat{a}_t^i$ 进行微调，以降低潜在的碰撞风险。C-IEC 模块接收来自环境的所有状态信息，并输出全局状态价值函数，以评估整体状态质量并指导 D-IADM 中的参数更新。为了更好地理解交通演变并引导网联自动驾驶车辆的运动，我们使用综合交通动态表示（ITDR）模块来捕捉交通动态的关键特征以及车辆之间的交互。

下面将详细说明状态空间、动作空间和奖励函数。

1）状态空间

车辆状态包括位置（x，y）、速度 v、航向角 θ、车道索引 l 和车辆类型 c。此外，还考虑了车道特定信息，包括车辆数 n^l、密度 D^l、平均速度v^l 以及网联自动驾驶车辆渗透率 ρ^l，其中 l 表示瓶颈场景中各车道的索引。此外，高精度地图提供了详细的道路结构信息。这些数据构成了集中式评论家可从环境中获取的状态信息。

为帮助网联自动驾驶车辆更好地理解其与周围环境的交互特征，我们为每辆网联自动驾驶车辆 i设计了局部观测状态，具体如下：

- **自身状态信息**：位置坐标（x_t^i，y_t^i）、速度 v_t^i、航向角 θ_t^i 和车道索引 l_t^i。
- **相邻车辆状态信息**：该信息由集合 $\mathcal{M}_s^i$ 定义，其中包含同车道和相邻车道中距离最近的六辆车。每辆相邻车辆 $j \in \mathcal{M}_s^i$ 相对于网联自动驾驶车辆 i 的状态由相对位置（$\Delta x^{i,j}_t$，$\Delta y^{i,j}_t$）和相对速度 $\Delta v^{i,j}_t$，以及相邻车辆的航向角 θ_t^j 和车辆类型 c^j 描述。
- **相邻车道交通流统计**：网联自动驾驶车辆通过共享车道统计信息了解周围交通状态，包括车辆数 n^l、密度 D^l、平均速度 v^l和网联自动驾驶车辆渗透率 ρ^l。
- **静态道路结构信息**：合流点前后路段的坐标和车道数用作静态道路结构信息。

2）动作空间

每辆网联自动驾驶车辆 i 具有离散动作空间 $\mathcal{A}_i$，涵盖横向控制与纵向控制。具体而言，$\mathcal{A}_i$

包含五种决策行为：保持当前车道和速度（记为“保持”）、向左或向右变道、加速或减速。因此，每辆网联自动驾驶车辆的动作集正式定义为 $\mathcal{A}_i = \{0:$ 保持,1: 向左,2: 向右, 3: 加速,4: 减速$\}$。在每个时间步，网联自动驾驶车辆 i 通过其策略 π_θ^i选择一个动作 $a_t^i \in \mathcal{A}_i$，以最大化其期望累积奖励。

3）奖励函数

总奖励由三个部分组成：基于softmin聚合的局部自车奖励 $\bar{r}_e$、全局奖励 r_g和完成奖励 r_{done}:

$$R_t = w_e \bar{r}_e + w_g r_g + r_{done}, \tag{8}$$

其中，w_e 和 w_g 为固定权重系数。奖励r_{done} 也被纳入其中，以激励网联自动驾驶车辆驶过瓶颈区域，从而提高通行量。

在混合交通瓶颈场景中，网联自动驾驶车辆面临不同的交互复杂度：有些车辆处于动态简单的稀疏区域，而另一些则需要在车辆交互频繁的密集合流区域中行驶，这些构成了交互密集情形。为使学习过程侧重于这些情形，对局部自车奖励采用基于softmin的聚合:

$$\bar{r}_e = \sum_{i=1}^{n} w^i r_e^i, \tag{9}$$

其中，softmin 权重的计算方式如下:

$$w^i = \frac{\exp(-r_e^i/\tau)}{\sum_{j=1}^{n} \exp(-r_e^j/\tau)}. \tag{10}$$

权重 $w^i \propto \exp(-r_e^i/\tau)$ 为奖励较低的智能体赋予更高的重要性，从而将学习重点引向表现不佳的智能体，以及瓶颈处最具挑战性的协作场景。温度参数 τ 控制关注强度：较大的 τ 会产生近乎均匀的权重，以进行广泛探索；较小的τ 则会将学习集中在表现最差的智能体上。为在整个训练过程中兼顾这两种模式，采用线性退火计划:

$$\tau(t) = \begin{cases} \tau_{init} - \frac{(\tau_{init}-\tau_{final})\cdot t}{T_{anneal}} & \text{if } t \le T_{anneal}, \\ \tau_{final} & \text{if } t > T_{anneal}, \end{cases} \tag{11}$$

其中，τ_{init} 和τ_{final} 分别为初始和最终温度值，T_{anneal} 为退火时长。这种受课程学习启发的策略 [40]能够在训练早期实现广泛探索，并在后期阶段专注于困难交互的优化，从而防止策略被更简单环境中的智能体主导。自车奖励 r_e^i 由多个子奖励组成部分构成，即

$$\begin{aligned} r_e^i &= w_{e,v} r_{e,v}^i + w_{e,w} r_{e,w}^i + w_{e,c} r_{e,c}^i + w_{e,tp} r_{e,tp}^i, \\ r_{e,v}^i &= -\frac{|v^i - v_{max}|}{v_{max}} + p_{e,v}, \\ r_{e,w}^i &= \sum_{j \in \mathcal{M}_w^i} -\frac{d_{th,w} - \sqrt{\Delta x^{i,j^2} + \Delta y^{i,j^2}}}{d_{th,w} - d_{th,c}}, \\ r_{e,c}^i &= \sum_{j \in \mathcal{M}_c^i} \left[-\frac{d_{th,c} - \sqrt{\Delta x^{i,j^2} + \Delta y^{i,j^2}}}{d_{th,c}} + p_{e,c} \right], \\ r_{e,tp}^i &= \text{sigmoid}(v^i - v_{th}) - 1. \end{aligned} \tag{12}$$

其中，$w_{e,v}$、$w_{e,w}$、$w_{e,c}$和$w_{e,tp}$ 分别是速度、警告距离、碰撞惩罚和时间惩罚的加权参数。参数 v_{max} 表示规定的最高速度。集合 $\mathcal{M}_w^i$ 和 $\mathcal{M}_c^i$ 分别表示处于警告范围和实际碰撞范围内的车辆，其距离阈值分别为$d_{th,w}$ 和 $d_{th,c}$。参数$p_{e,v}$和$p_{e,c}$ 用于调整速度奖励和碰撞惩罚。时间惩罚$r_{e,tp}^i$ 会对速度低于阈值 v_{th} 的车辆进行惩罚，以防止懒惰行为。

全局奖励 r_g 反映了所有网联自动驾驶车辆在平均速度方面的整体表现:

$$r_g = -\frac{\left|\frac{1}{n}\sum_{i=1}^{n} v^i - v_{max}\right|}{v_{max}} + p_{g,v}. \tag{13}$$

其中$p_{g,v}$ 是用于调整全局速度奖励的参数。

*B. D-IADM*模块

D-IADM模块通过区分 CAV-CAV 协作交互与 CAV-HDV 观测交互，增强行动者的局部交互感知。该模块包含两个关键组件：1）轨迹感知交互编码器（TAIE），通过独立的注意力网络对这两种交互类型进行建模；2）基于主动安全的动作优化（PSAR）模块，通过基于规则的修正减少高风险动作。动作记忆机制会将行动者生成的动作和经 PSAR 优化的动作纳入下一时间步的观测中，从而为行动者提供其意图决策与执行决策的完整记录。

1) 轨迹感知交互编码器（*TAIE*）

TAIE模块的目的是处理每辆网联自动驾驶车辆的局部观测，并提取决策所需的关键交互特征，如图3所示。每辆网联自动驾驶车辆的局部观测状态包括其自身状态信息、相邻车辆运动信息、车道统计信息以及静态道路结构信息。相邻车辆集合记为 $\mathcal{M}_s^i$，包含同车道及相邻车道中最近的六辆车。考虑到人工驾驶车辆与网联自动驾驶车辆之间的行为差异，我们将相邻车辆划分为人工驾驶车辆集合$\mathcal{M}_{s,H}^i$ 和网联自动驾驶车辆集合$\mathcal{M}_{s,C}^i$。该分类增强了模型捕捉网联自动驾驶车辆与人工驾驶车辆之间交互的能力，并降低了偏差。

为说明交互建模过程，我们以提取人工驾驶车辆交互特征为例。为此，构建网联自动驾驶车辆 i 与周围人类驾驶车辆之间的交互图，将每辆网联自动驾驶车辆与指定范围内相邻的人工驾驶车辆连接起来。人工驾驶车辆交互图表示为 $\mathcal{G}_{s,H}^i = (\mathcal{V}_{s,H}^i, \mathcal{E}_{s,H}^i)$，其中 $\mathcal{V}_{s,H}^i = i \cup \mathcal{M}_{s,H}^i$ 表示节点，$\mathcal{E}_{s,H}^i$ 表示连接这些节点的边。请注意，这些边仅连接网联自动驾驶车辆 i 与其相邻的人工驾驶车辆，人工驾驶车辆之间彼此不连接。将 $m_t^i = (x_t^i, y_t^i, v_t^i, \theta_t^i)$ 记为车辆 i的状态，其中 x_t^i、y_t^i、v_t^i和 θ_t^i 分别表示

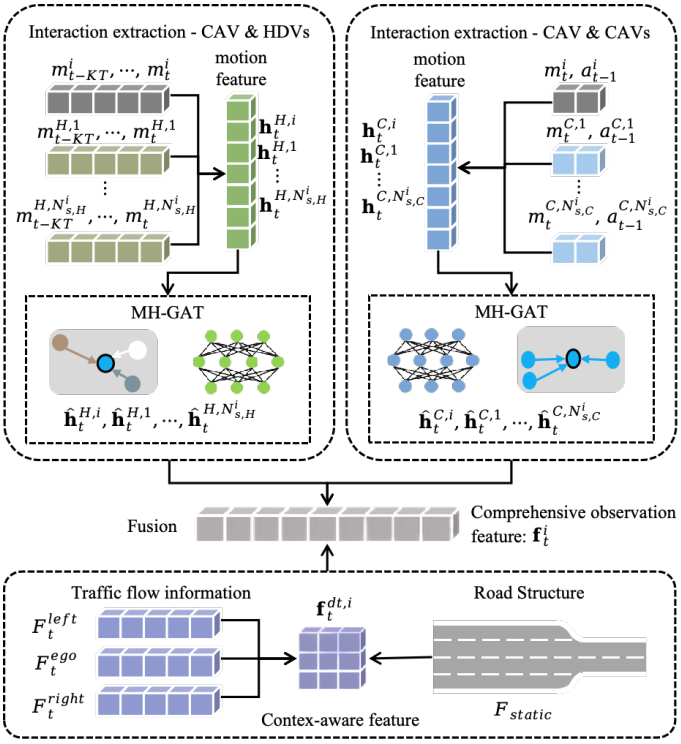


图3：TAIE的网络框架。

位置、速度和航向角。网联自动驾驶车辆—人工驾驶车辆层的输入由网联自动驾驶车辆 i 及其 $N_{s,H}^i$ 周围人工驾驶车辆的当前状态和历史状态构成。历史状态经拼接后嵌入节点特征$\mathbf{h}_t^{H,j} \in \mathbb{R}^{S_H}$，然后输入多头图注意力（MH-GAT）[41]以提取交互特征。

网联自动驾驶车辆 i 的增强节点特征$\hat{\mathbf{h}}_t^{H,i}$,通过对所有 B 个注意力头加权聚合相邻特征得到:

$$\hat{\mathbf{h}}_t^{H,i} = \sigma\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j\in\mathcal{M}_{s,H}^i}\alpha_{ij}^b W^b \mathbf{h}_t^{H,j}\right), \tag{14}$$

其中 α_{ij} 为注意力权重，用于量化周围人工驾驶车辆 j 对网联自动驾驶车辆 i的相对重要性，计算方式为:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\hat{\mathbf{a}}[W\mathbf{h}_t^{H,i} \| W\mathbf{h}_t^{H,j}]))}{\sum_{k\in\mathcal{M}_{s,H}^i}\exp(\text{LeakyReLU}(\hat{\mathbf{a}}[W\mathbf{h}_t^{H,i} \| W\mathbf{h}_t^{H,k}]))} \tag{15}$$

其中 W 为权重矩阵，$\hat{\mathbf{a}}$ 为单层前馈网络，其后接LeakyReLU非线性与softmax归一化。该聚合生成了网联自动驾驶车辆 i周围人工驾驶车辆交互动态的丰富表示。

以类似的过程生成 CAV 交互图。与人工驾驶车辆图不同，CAV 图还纳入了每辆 CAV 的先前决策信息 a_{t-1}^i，从而得到 CAV交互特征 $\mathbf{f}_t^{C,i}$。这种设计体现了关键差异：CAV-CAV 交互涉及共享决策信息，而联网自动驾驶车辆与人工驾驶车辆之间的交互则依赖观测到的轨迹模式。此外，局部交通上下文信息（车道统计信息 $\mathbf{F}_t^{ego}$、$\mathbf{F}_t^{left}$、$\mathbf{F}_t^{right}$ 以及静态道路结构 $\mathbf{F}_{static}$）经多层感知机处理后，得到上下文嵌入 $\mathbf{f}_t^{dt,i}$。最后，人工驾驶车辆交互嵌入、CAV 交互嵌入和局部交通上下文嵌入在基于多层感知机的融合层中进行组合，形成时间步 t 的综合观测嵌入特征 $\mathbf{f}_t^i$。

2) 基于主动安全的动作优化（*PSAR*）

算法 1 基于主动安全的动作优化 (PSAR)

需要： 执行者动作 $\hat{a}_t^i \in \{0,1,2,3,4\}$；自车的状态以及周围车辆；
距离阈值 d_{LC} ,min $\le d_{\text{safe}} < d_{\text{warn}} < d_{\text{att}}$; TTC阈值 $\tau_{\text{safe}}<\tau_{\text{warn}}<\tau_{\text{att}}$;
风险$(d_0, \tau_0) \equiv |d^{i,k}_{t,long}| \le d_0 \wedge 0 < \text{TTC}_t^{i,j} \le \tau_0$，(用于当前情境中的相关车辆对)；
确保： 精细化动作 a^i $_t$
$a_t^i \leftarrow \hat{a}_t^i$
如果 $a_t^i \in \{1,2\}$ **则**
  $b_{lc} \leftarrow 0$;
  如果 目标车道前车 j 存在 **则**
    如果 $d^{i,j}_{t,long} \le d_{\text{LC, min}}$ **或** 风险$(d_{\text{safe}}, \tau_{\text{safe}})$ **则**
      $a_t^i \leftarrow 0$ ▷ 取消变道
    否则，如果 风险$(d_{\text{safe}}, \tau_{\text{att}})$ **或** 风险$(d_{\text{att}}, \tau_{\text{safe}})$ **则**
      $b_{lc} \leftarrow$ 最小值$(|\Delta v^{i,j}_{t,long}|, b_{\max})$ ▷ 减速变道
    结束如果
  结束如果
  如果 目标车道后方车辆 k 存在 **且** $a_t^i \neq 0$ **则**
    如果 $|d^{i,k}_{t,long}| \le d_{\text{LC, min}}$ **或** 风险$(d_{\text{safe}}, \tau_{\text{safe}})$ **则**
      $a_t^i \leftarrow 0$ ▷ 取消变道
    否则如果 [风险$(d_{\text{safe}}, \tau_{\text{att}})$ **或** 风险$(d_{\text{att}}, \tau_{\text{safe}})$] **且** $b_{lc} \neq 0$ **则**
      $a_t^i \leftarrow 0$ ▷ 取消变道
    结束如果
  结束如果
结束如果
如果 $a_t^i \in \{0,3,4\}$ **则**
  如果前方车辆 j 存在 **则**
    如果 $d^{i,j}_{t,long} \le d_{\text{warn}}$ **且** 碰撞时间$t_t^{i,j} > 0$，**那么**
      $a_t^i \leftarrow 4$; ▷ 强制制动
    否则如果 $d^{i,j}_{t,long} \le d_{\text{safe}}$ **或**$(d^{i,j}_{t,long} \le d_{\text{warn}}$ **且** $a_t^i=3)$ **则**
      $a_t^i \leftarrow 0$ ▷ 抑制加速
    否则如果风险$(d_{\text{att}}, \tau_{\text{warn}})$ **则**
      $a_t^i \leftarrow 4$; ▷ 强制制动
    否则如果风险$(d_{\text{att}}, \tau_{\text{att}})$ **且** $a_t^i=3$ **则**
      $a_t^i \leftarrow 0$ ▷ 抑制加速
    结束条件判断
  结束条件判断
结束条件判断
返回 a_t^i

PSAR 是一个基于规则的动作优化层，将行动者提议的动作 $\hat{a}_t^i \in \mathcal{A}_i$ 作为输入，并输出精细化动作 a_t^i。与相关周围车辆的相对纵向间距 $d^{i,j}_{t,long}$ 以及碰撞时间（TTC）$\text{TTC}_t^{i,j}$，均作为风险指标。

$$d_{t,long}^{i,j} = (x_t^j - x_t^i - L_{veh})\cos\theta_t^i - \frac{W_{veh}^i}{2}|\sin\theta_t^i|$$

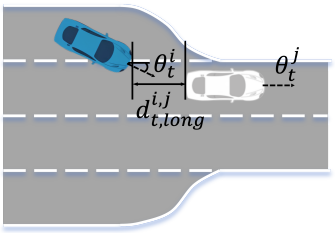


图4：碰撞时间的改善。

算法 2 面向网联自动驾驶车辆的去中心化交互自适应决策（D-IADM）

1：初始化：局部观测 $o_0^1, \ldots, o_0^n$，**done** $=$**False 2：for** $t = 1$ **to** L **do 3：更新观测** $o_t^1, \ldots, o_t^n$ **4：for** $i = 1$ **to** n **do 5：使用 MH-GAT 获取人工驾驶车辆交互特征** $\mathbf{f}_t^{H\,i}$ **6：使用 MH-GAT 获取 CAV交互特征** $\mathbf{f}_t^{C\,i}$ **7：使用 MLP 获取上下文感知交通特征** $\mathbf{f}_t^{dt\,i}$ **8：使用 MLP 融合特征，得到综合观测** $\mathbf{f}_t^i$ **9：将** $\mathbf{f}_t^i$ **与上一个动作** a_{t-1}^i **拼接 10：使用 GRU 和决策层生成动作** $\hat{a}_t^i$ **11：if** $\hat{a}_t^i$ **违反碰撞时间/间距安全规则 then 12：应用启发式修正，得到精细化动作** a_t^i **（PSAR） 13：else 14：保持动作** $a_t^i = \hat{a}_t^i$ **15：end if 16：end for 17：执行动作** $a_t^1, \ldots, a_t^n$ **并接收奖励** $r_t^1, \ldots, r_t^n$ **18：if 发生碰撞或所有车辆驶离道路 then 19：done** $=$**True 20：break 21：end if 22：end for**

$$\mathrm{TTC}_t^{i,j} = \frac{d_{t,long}^{i,j}}{\Delta v_{t,long}^{i,j}}, \quad j \in \mathcal{M}_s^i$$

$$\Delta v_{t,long}^{i,j} = v_t^i |\cos\theta_t^i| - v_t^j |\cos\theta_t^j|$$

其中，W_{veh}^i 表示网联自动驾驶车辆 i的宽度，$\Delta v^i_{t^j,long}$表示网联自动驾驶车辆 i与其周围车辆 $j \in \mathcal{M}_s^i$之间的纵向速度差，同时考虑车辆的航向角。

PSAR采用四种动作细化方式：

- **取消变道**：取消计划中的换道动作；
- **减速变道**：在减速的同时保持换道意图；
- **强制制动**：用减速覆盖纵向指令；
- **抑制加速**：抑制加速指令。

完整的触发条件和降速公式见算法1。行动者生成的动作 $\hat{a}_t^i$ 和经 PSAR 精细化的动作 a_t^i 均作为动作记忆纳入下一时间步的观测中。保留这两种动作后，行动者便能感知其预期决策如何被安全模块修改，并在整个训练过程中跟踪计划行为与实际执行行为之间的差异。

总之，混合交通中网联自动驾驶车辆的 D-IADM 流程如算法 *2*所示。

C. 集中式交互增强评论家（*C-IEC*）

C-IEC模块旨在增强评论家的全局交通理解能力。C-IEC明确建模车辆交互如何影响交通动态，从而实现更准确的价值估计，为策略更新提供更清晰的指导。其核心观点是，通过理解交互与交通演化之间的关系，评论家能够更好地评估局部协作动作是否有助于全局交通改善。C-IEC模块通过其核心组件综合交通动态表示（ITDR）模块实现这一目标；该模块构建全局车辆交互图，并采用交叉注意力机制捕捉交互与交通动态之间的因果关系。随后，ITDR的输出经门控循环单元处理，以捕捉时间依赖，最后通过全连接层输出价值函数预测结果，为行动者更新提供指导。

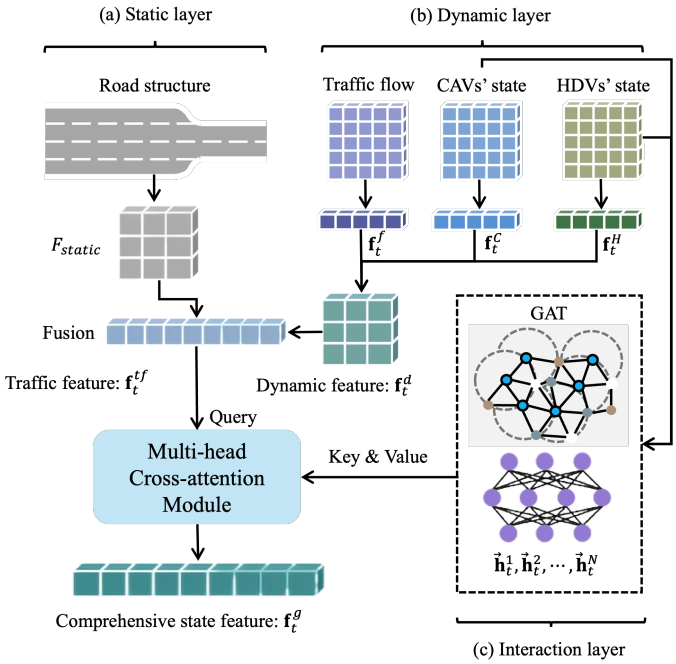


图5：ITDR的网络框架。

ITDR模块通过处理三个不同的信息层，分析车辆交互与交通演化之间的关系：(a) 静态道路结构$\mathbf{F}_{static}$，(b) 动态交通信息（全局车道统计、网联自动驾驶车辆和人工驾驶车辆的状态），以及 (c) 车辆交互信息。前两个信息层分别通过专用的多层感知机处理，并拼接形成交通特征$\mathbf{f}_t^{tf}$。

对于第三个信息层，车辆交互信息通过图注意力网络处理，以提取全局交互特征。构建一个全局车辆交互图，记为 $\mathcal{G}_{global}=(\mathcal{V},\mathcal{E})$，其中 $\mathcal{V}$ 表示所有车辆的节点，$\mathcal{E}$ 表示连接这些节点的边。此处仅连接相邻车辆，即 $\mathcal{E}=\{(i,j)|i\in\mathcal{V},j\in\mathcal{M}_s^i\}$。每个车辆节点的状态 $m_t^i=(x_t^i,y_t^i,v_t^i,\theta_t^i)$被嵌入节点特征$\mathbf{h}_t^i\in\mathbb{R}^{S_N}$中，再输入图注意力网络以提取全局交互特征 $\overrightarrow{\mathbf{h}}_t^i\in\mathbb{R}^{S_N}$。节点 i 到节点 j 的注意力权重 α_{ij} 计算如下：

$$\alpha_{ij}=\frac{\exp(\text{LeakyReLU}(\overrightarrow{\mathbf{a}}[W\mathbf{h}_t^i\|W\mathbf{h}_t^j]))}{\sum_{k\in\mathcal{M}_s^i}\exp(\text{LeakyReLU}(\overrightarrow{\mathbf{a}}[W\mathbf{h}_t^i\|W\mathbf{h}_t^k]))} \tag{16}$$

其中，W 是权重矩阵，$\overrightarrow{\mathbf{a}}$ 是一个单层前馈神经网络，后接LeakyReLU非线性函数和softmax归一化。节点 i 在时间 t 的增强节点特征 $\overrightarrow{\mathbf{h}}_t^i$ 是所有邻居特征的加权和：

$$\overrightarrow{\mathbf{h}}_t^i=\sigma\left(\sum_{j\in\mathcal{M}_s^i}\alpha_{ij}W\mathbf{h}_t^j\right), \tag{17}$$

其中，$\overrightarrow{\mathbf{h}}_t=[\overrightarrow{\mathbf{h}}_t^i]_{i\in\mathcal{V}}$ 表示所有车辆 $i\in\mathcal{V}$的增强节点特征堆叠矩阵。为捕捉车辆交互对交通演化的影响，采用多头交叉注意力机制，以交通特征$\mathbf{f}_t^{tf}$作为查询，以交互特征 $\overrightarrow{\mathbf{h}}_t$作为键和值。每个头 h的注意力输出为：

$$\text{Attention}(Q_h,K_h,V_h)=\text{softmax}\left(\frac{Q_hK_h^T}{\sqrt{d_k}}\right)V_h, \tag{18}$$

这里，d_k 是键维度，Q_h、K_h和 V_h 通过学习得到的投影获得：

$$Q_h=\mathbf{f}_t^{tf}W_Q^h,\quad K_h=\overrightarrow{\mathbf{h}}_tW_K^h,\quad V_h=\overrightarrow{\mathbf{h}}_tW_V^h, \tag{19}$$

其中，W_Q^h、W_K^h和 W_V^h是头h的投影矩阵。所有 H 个头的输出拼接后再进行线性投影：

$$\text{MultiHead}(Q,K,V)=W_O\left[\text{head}_1;\text{head}_2;\ldots;\text{head}_H\right], \tag{20}$$

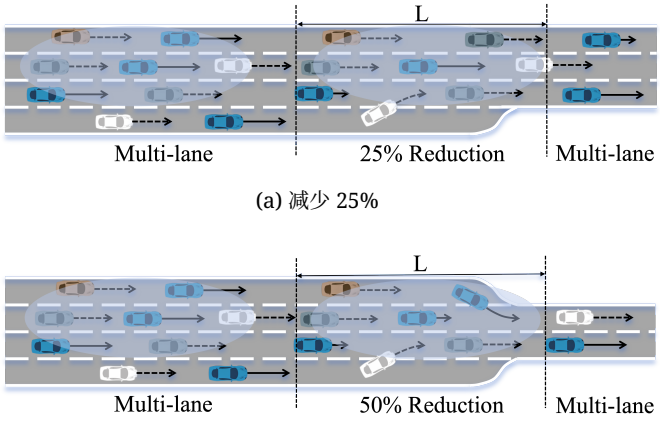
其中，$\text{head}_h=\text{Attention}(Q_h,K_h,V_h)$，$W_O$ 为输出投影矩阵。

通过这一机制，ITDR 捕捉车辆交互如何影响交通演化。与 TAIE 的局部视角不同，ITDR 为评论家提供了车辆交互和交通动态的全局视角，从而能够更准确地进行价值估计，为策略更新提供指导。

V. 实验

A. 实验设置

仿真环境基于 SUMO[42] 和 TransSimHub [43]，开发，后者是一个用于多模态感知与决策的统一仿真平台。本研究基于 HARL 代码库 [44]，开发，实现代码仓库地址为 https://github.com/lzxphhh/DIACC.git。



(a) 减少 25%

(b) 减少 50%

图6：瓶颈设置，其中（a）车道缩减25%，（b）车道缩减50%

如图6所示，道路网络由容量降低 25% 的路段、容量降低 50% 的路段以及多车道路段组成，形成一条长 1.3 km 、最高限速为 25 m/s 的长距离路线。为进行泛化测试，另设计了一张 1.3 km 的测试地图，其中容量缩减路段的位置与训练地图不同。人工驾驶车辆采用智能驾驶员模型（IDM） [45]进行纵向控制，并采用 LC2013模型 [46]进行换道操作。

实验采用系统化训练和零样本测试框架，以评估模型在多种交通场景下的性能。训练期间，场景中包含25辆车，其中有10辆网联自动驾驶车辆（渗透率为0.4）和15辆遵循表II所规定的D1驾驶风格分布的人工驾驶车辆。三组参数分别代表激进、正常和谨慎的驾驶行为：激进驾驶的跟车距离较短、变道更频繁，而谨慎驾驶的跟车距离较长、变道较少。正常驾驶采用默认的SUMO模型参数。

表II：人工驾驶车辆驾驶风格分布。

Name	激进	正常	谨慎
D1	0.2	0.6	0.2
D2	0.2	0.4	0.4
D3	0.4	0.4	0.2

在零样本泛化测试中，我们在测试地图上评估训练好的模型，测试地图中的减速路段位置各不相同，且测试条件有所变化：网联自动驾驶车辆渗透率为 0.2 和 0.3，车辆数量为 20 、 25 、 30 和 40，人工驾驶车辆驾驶风格分布为 D2 和 D3。这些测试用于评估模型对未见过的道路配置和交通组成的适应性。

1) 基线模型

对比实验涵盖基于多智能体近端策略优化的基线配置和消融配置：

- 纯人工驾驶车辆场景：所有车辆均为人工驾驶车辆，采用IDM+LC2013模型，并启用SUMO的安全模式。

- SUMO协同场景：根据渗透率引入网联自动驾驶车辆，采用协同式自适应巡航控制模型，并进行高度协同变道（lcCooperative=1）。
- MAPPO基线场景：网联自动驾驶车辆由原始多智能体近端策略优化模型控制。
- 我们的模型：包含三种基于多智能体近端策略优化的变体：– DIACC（不含PSAR）：不带PSAR动作精炼模块的DIACC模型，用于消融分析。–MAPPO-IADM：仅加入D-IADM模块的MAPPO模型。– DIACC：同时加入D-IADM和C-IEC模块并采用PSAR的MAPPO模型（我们的主要提出模型）。

在实验中，我们启用了SUMO模拟器针对人工驾驶车辆的安全模式，以防止人工驾驶车辆之间发生碰撞，否则这会使评估结果产生偏差。

2）训练参数

我们在包含三个人工驾驶车辆的密集混合交通瓶颈场景中，使用三个随机种子将各比较模型训练了300万步。训练基础设施采用分布式框架，在CPU上的20个并行环境中收集轨迹数据，并使用单个GPU专门进行策略学习。训练期间，从回放缓冲区中随机采样大小为32的转移数据批次，并将其输入学习器。行动者网络和评论家网络均使用Adam优化器进行优化，学习率为 5×10^{-4}，折扣因子为0.99。

B. 结果与分析

1) 训练性能

图7(a)–(c)比较了基线模型和消融模型的训练动态，包括多智能体近端策略优化、MAPPO-IADM、DIACC（无PSAR）和DIACC。如图所示，在所有对比方法中，DIACC在收敛时实现了最低的碰撞率和最高的奖励值，验证了完整框架的有效性。缺少C-IEC的MAPPO-IADM在训练早期能够更快地降低碰撞率，但DIACC在后期超过了它。这一行为表明，C-IEC通过在ITDR中提供交互感知全局表征，实现更准确的价值估计；在基本交互模式建立后，它能够引导策略精炼，实现更优的协同。与DIACC（无PSAR）相比，DIACC还实现了更快且更稳定的碰撞率下降，这是因为PSAR通过基于TTC和间距的规则过滤高风险探索性动作，减少了训练早期造成不稳定的不安全转移。在所有变体中，原始多智能体近端策略优化表现最差，无法处理混合交通中固有的复杂交互动态。

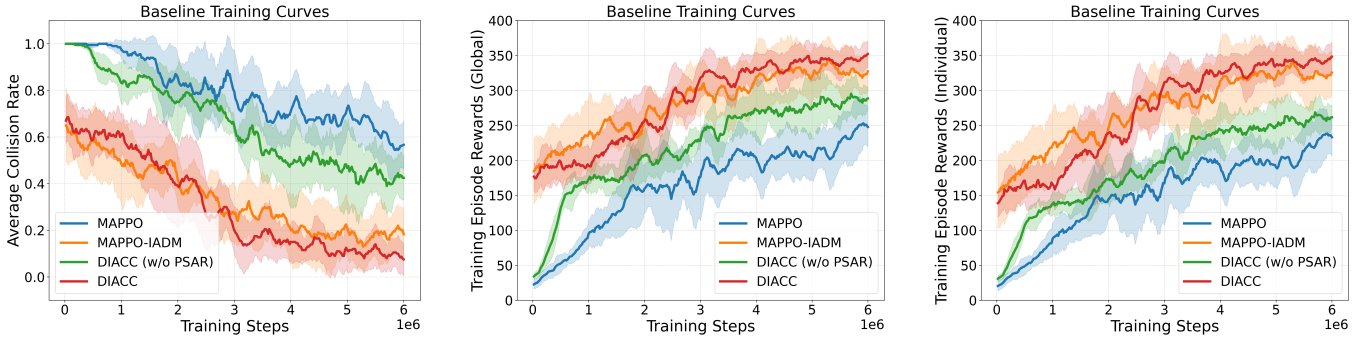
图 7(d)–(f) 通过将退火策略（如式 (11) 所定义，在训练前半程内采用 $\tau_{init} = 2 \rightarrow \tau_{final} = 0.05$ ）与固定的 τ 设置进行比较，考察温度参数 τ 的影响。退火策略在收敛时实现了最低的碰撞率和最高的奖励值，表明受课程学习启发的调度策略能够有效平衡广泛探索与定向优化。固定的 $\tau = 0.05$ 会导致较高的训练方差（置信区间的阴影范围较宽）且碰撞率收敛情况较差，因为在基本协同尚未建立之前，策略就被引导去处理困难案例。固定 $\tau = 2$或 $\tau = 10$时，智能体间的 softmin 权重会变得越来越均匀（当 $\tau \to \infty$时， $w^i \to 1/n$ ），从而减少对交互密集型场景的关注；随着 τ 增加，训练性能会单调下降。

图 7(g)–(i)考察了式 (8) 中自车奖励权重w_e 的影响，该权重用于平衡自车的局部目标与全局协同奖励。在所有测试设置中，$w_e = 0.6$ 的表现最佳，在收敛时实现了最低的碰撞率和最高的奖励值。自车奖励权重 w_e 较小（$w_e = 0$ 至 0.4）时，自车奖励不足以区分单个智能体的性能，因此策略无法有选择地引导学习侧重于交互密集的智能体；并且，随着 w_e 减小，训练性能会下降。自车奖励权重 w_e 较大（$w_e = 0.8$或 $w_e = 1.0$）时，这两种设置的表现均不如$w_e = 0.6$，因为全局奖励r_g 的权重降低，导致每辆网联自动驾驶车辆过度优化其局部目标，削弱协同合作，并使策略趋向次优解。

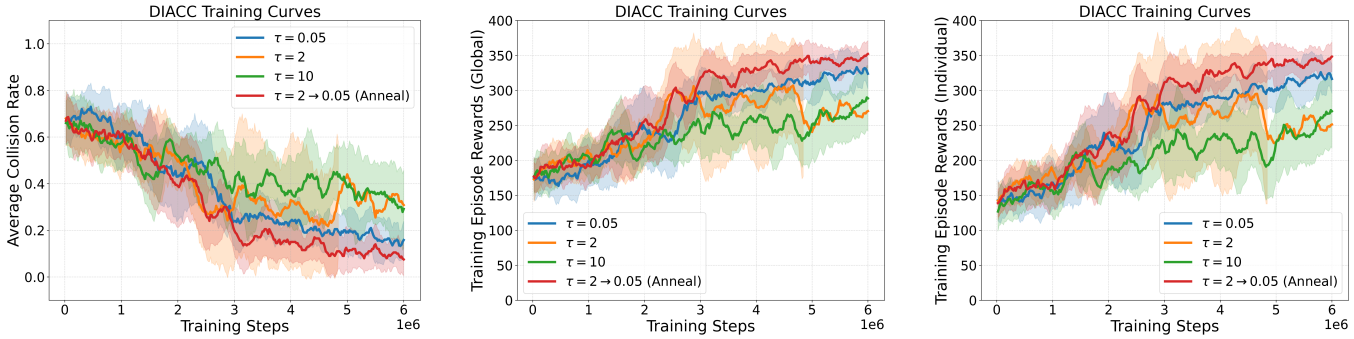
2) 测试性能

表 III 和表 IV 全面展示了在容量降低 25% 和 50% 的瓶颈场景下的性能评估，其中 N 表示通过瓶颈的车辆总数。评估了三项指标：(1) 速度（米/秒），通过测量平均速度和标准差反映交通效率；(2) WE 的 p 值，即车辆速度低于 3 米/秒时发生等待事件的概率；(3) p（SCE），即发生安全关键事件的概率，包括近距离违规（距离低于最低安全阈值）、碰撞时间低于临界值、碰撞和紧急制动 [47]。

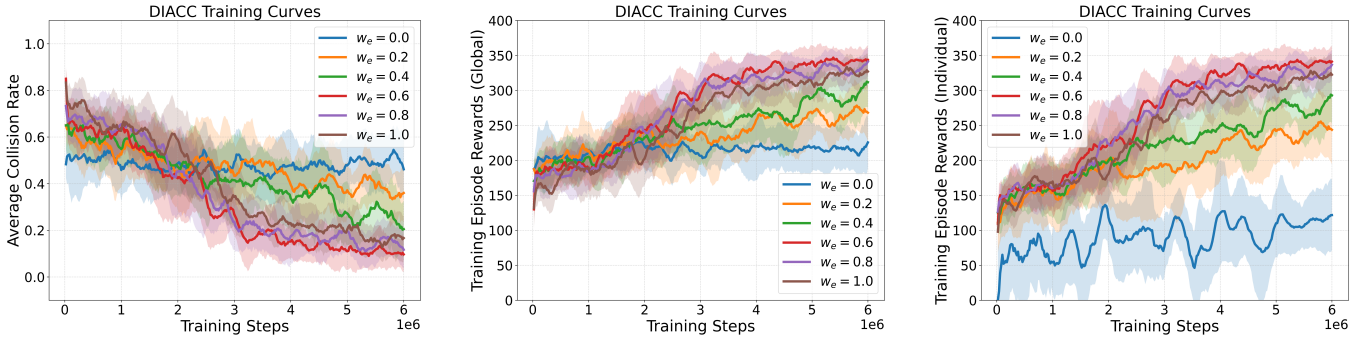
在表 III 和表 IV 中，DIACC 在效率（高速度和低 WE 的 p 值）与安全性（低 p（SCE））两方面均持续取得最优或接近最优的结果，证实了完整架构的有效性。值得注意的是，在表 III 的全部五个零样本场景中，DIACC 将 p（SCE）降至 0%，表明完整框架能够在多样化的驾驶风格分布下有效消除安全关键事件。MAPPO-IADM 采用相同的行动者架构，但省略了 C-IEC，在表 III 中取得了具有竞争力的效率表现，甚至在某些场景中超过 DIACC（例如，N=25，D1：20.77 对 19.04 m/s，N=20，D1：21.36 对 19.26 m/s）。这表明仅 D-IADM 行动者设计就足以在简单设置中保持高效率。然而，MAPPO-IADM 在表 III 中并不能持续实现零 p（SCE），并且在要求更高的 50% 缩减场景中显著下降，尤其是在表 IV 的 N=30 和 N=40 设置中，而 DIACC 在效率和安全性方面均展现出明显优势。这些对比结果印证了 C-IEC 的作用：它在行动者训练期间提供具备交互感知的全局价值估计，从而促成协同策略的学习，而这些策略是行动者模块单独在高密度、容量受限场景中无法实现的。



(a) 基线 – 碰撞率 (b) 基线 – 全局奖励 (c) 基线 – 个体奖励



(d) τ 选择 – 碰撞率 (e) τ 选择 – 全局奖励 (f) τ 选择 – 个体奖励



(g) w_e 选择 – 碰撞率 (h) w_e 选择 – 全局奖励 (i) w_e 选择 – 个体奖励

图 7：综合训练曲线比较。(a)–(c)：基线与消融对比。(d)–(f)：softmin温度 τ 选择。(g)–(i)：自车奖励权重 w_e 选择。

3) 协作分析

图8 展示了所有方法在不同交通规模和容量削减场景下的速度分布，从分布视角对表 III 和表 IV 中的汇总指标形成补充。在 $N = 30$ 和 $N = 40$ 的箱线图与核密度估计（KDE）图中（图8(a)–(b)），DIACC 在所有方法中取得了最高的中位速度，并且在正常多车道和 25% 容量降低条件下观察到最集中的四分位距。这表明 DIACC 使车辆能够保持持续较高且稳定的速度，其低速振荡少于任何基线。各道路路段的 KDE 图（图8(c)–(h)）进一步证实了这一规律，表明随着容量降低从 25% 增至 50%，DIACC 的速度分布仍保持最大的

在高速区域的比例最高，在低速区域的比例最低。这种分布优势体现了 DIACC 的双层交互设计：D-IADM 使每个行动者能够主动做出适应交互的决策，而 C-IEC 则提供准确的全局价值估计，二者共同维持协同交通流并抑制速度振荡，且这一效果会随着瓶颈严重程度的增加而持续发挥。

y 增加。

C. 讨论

车辆交互从根本上影响着单个车辆的车辆运动以及混合交通环境中的整体交通演化。DIACC策略在MAPPO 框架内通过三个相互补充的组成部分应对这一挑战：1）D-IADM 通过区分 CAV-CAV 协作交互与 CAV-HDV 观测交互，增强行动者的局部交互感知；2）C-IEC 通过交互感知价值估计强化评论家的全局交通理解，

表 III：容量减少 25% 的瓶颈场景中的性能比较。

场景	指标	SUMO（纯人工驾驶车辆）	SUMO（混合交通）	MAPPO-IADM	DIACC
			训练场景		
N=25, $\rho = 0.4$, D1	速度（米/秒）	16.08 ± 7.94	17.05 ± 7.24 (+6.0%)	**20.77 ± 2.88 (+29.2%)**	19.04 ± 6.19 (+18.4%)
	WE 的 *p* 值	11.20 ± 4.66%	8.80 ± 7.76% (-21.4%)	**0.00 ± 0.00% (-100.0%)**	4.00 ± 0.00% (-64.3%)
	p（*SCE*）	7.20 ± 3.92%	6.40 ± 9.33% (-11.1%)	**0.00 ± 0.00% (-100.0%)**	**0.00 ± 0.00% (-100.0%)**
			零样本测试场景		
N=25, $\rho = 0.4$, D2	速度（米/秒）	14.44 ± 8.74	14.54 ± 9.07 (+0.7%)	**20.13 ± 4.62 (+39.4%)**	19.05 ± 6.61 (+31.9%)
	WE 的 *p* 值	18.40 ± 4.08%	11.20 ± 3.92%（-39.1%）	**4.00 ± 0.00%（-78.3%）**	**4.00 ± 0.00%（-78.3%）**
	p（*SCE*）	6.40 ± 3.20%	5.60 ± 6.97%（-12.5%）	**0.00 ± 0.00%（-100.0%）**	**0.00 ± 0.00%（-100.0%）**
N=25, $\rho = 0.4$, D3	速度（米/秒）	16.48 ± 7.90	16.12 ± 8.73 (-2.2%)	18.86 ± 6.91 (+14.4%)	**20.47 ± 3.64 (+24.2%)**
	WE 的 *p* 值	11.20 ± 5.31%	10.61 ± 5.29% (-5.3%)	12.00 ± 0.00%（+7.1%）	**0.00 ± 0.00% (-100.0%)**
	p（*SCE*）	5.60 ± 6.97%	12.42 ± 12.31%（+121.7%）	8.00 ± 0.00%（+42.9%）	**0.00 ± 0.00% (-100.0%)**
N=20, $\rho = 0.4$, D1	速度（米/秒）	16.26 ± 8.04	17.54 ± 8.58 (+7.9%)	**21.36 ± 2.61 (+31.4%)**	19.26 ± 6.60 (+18.5%)
	WE 的 *p* 值	15.00 ± 5.48%	9.00 ± 2.00% (-40.0%)	**0.00 ± 0.00% (-100.0%)**	5.00 ± 0.00% (-66.7%)
	p（*SCE*）	4.00 ± 8.00%	3.00 ± 6.00% (-25.0%)	**0.00 ± 0.00% (-100.0%)**	**0.00 ± 0.00% (-100.0%)**
N=30, $\rho = 0.3$, D1	速度（米/秒）	15.24 ± 8.26	16.77 ± 7.42 (+10.1%)	**18.65 ± 5.19 (+22.4%)**	18.48 ± 5.20 (+21.3%)
	WE 的 *p* 值	12.00 ± 4.99%	7.33 ± 3.89% (-38.9%)	**3.33 ± 0.00% (-72.2%)**	**3.33 ± 0.00% (-72.2%)**
	p（*SCE*）	8.00 ± 7.77%	2.00 ± 4.00% (-75.0%)	1.33 ± 2.67% (-83.3%)	**0.00 ± 0.00% (-100.0%)**
N=40, $\rho = 0.2$, D1	速度（米/秒）	15.03 ± 8.11	14.52 ± 8.81 (-3.4%)	15.63 ± 7.16 (+4.0%)	**17.24 ± 7.70 (+14.7%)**
	WE 的 *p* 值	10.00 ± 3.16%	13.00 ± 4.58%（+30.0%）	**5.00 ± 0.00% (-50.0%)**	7.50 ± 0.00%（-25.0%）
	p（*SCE*）	8.00 ± 2.92%	15.00 ± 8.51% (+87.5%)	**0.00 ± 0.00% (-100.0%)**	**0.00 ± 0.00% (-100.0%)**

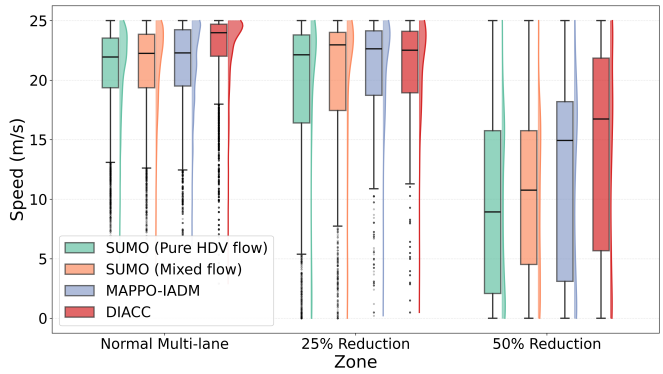
表 IV：容量降低 50% 的瓶颈场景中的性能比较。

场景	指标	SUMO（纯人工驾驶车辆）	SUMO（混合交通）	MAPPO-IADM	DIACC
			训练场景		
N=25, $\rho = 0.4$, D1	速度（米/秒）	10.00 ± 7.30	14.10 ± 6.52 (+41.0%)	17.47 ± 6.65 (+74.8%)	**17.60 ± 7.77 (+76.0%)**
	WE 的 *p* 值	25.60 ± 9.83%	13.23 ± 7.97%（-48.3%）	**6.00 ± 2.00%（-76.6%）**	6.80 ± 5.38%（-73.4%）
	p（*SCE*）	16.40 ± 7.68%	7.60 ± 8.48% (-53.7%)	6.40 ± 7.84% (-61.0%)	**0.00 ± 0.00% (-100.0%)**
			零样本测试场景		
N=25, $\rho = 0.4$, D2	速度（米/秒）	10.55 ± 7.71	13.21 ± 6.67 (+25.2%)	16.47 ± 7.96 (+56.1%)	**16.59 ± 8.48 (+57.2%)**
	WE 的 *p* 值	23.60 ± 9.37%	10.90 ± 6.96%（-53.8%）	12.00 ± 0.00%（-49.2%）	**10.00 ± 6.00%（-57.6%）**
	p（*SCE*）	8.80 ± 6.65%	**3.67 ± 5.26%（-58.3%）**	8.00 ± 0.00%（-9.1%）	4.00 ± 4.00% (-54.5%)
N=25, $\rho = 0.4$, D3	速度（米/秒）	12.65 ± 7.80	14.98 ± 7.47 (+18.5%)	15.32 ± 8.05 (+21.1%)	**15.90 ± 7.78 (+25.7%)**
	WE 的 *p* 值	16.40 ± 5.20%	10.03 ± 8.41%（-38.8%）	12.00 ± 4.00%（-26.8%）	**10.00 ± 2.00%（-39.0%）**
	p（*SCE*）	9.20 ± 7.60%	**3.20 ± 6.40%（-65.2%）**	14.00 ± 2.00% (+52.2%)	8.00 ± 0.00% (-13.0%)
N=20, $\rho = 0.4$, D1	速度（米/秒）	11.99 ± 7.64	14.91 ± 6.30 (+24.3%)	**17.98 ± 5.78 (+50.0%)**	17.62 ± 7.61 (+46.9%)
	WE 的 *p* 值	19.00 ± 8.31%	10.00 ± 4.47% (-47.4%)	**5.00 ± 0.00% (-73.7%)**	7.50 ± 2.50% (-60.5%)
	p（*SCE*）	13.00 ± 8.12%	7.50 ± 6.80% (-42.3%)	**0.00 ± 0.00% (-100.0%)**	**0.00 ± 0.00% (-100.0%)**
N=30, $\rho = 0.3$, D1	速度（米/秒）	9.25 ± 7.36	10.51 ± 6.79 (+13.6%)	11.97 ± 8.10 (+29.4%)	**14.37 ± 8.61 (+55.3%)**
	WE 的 *p* 值	27.67 ± 10.96%	23.67 ± 7.67%（-14.5%）	16.00 ± 3.27%（-42.2%）	**13.33 ± 3.33%（-51.8%）**
	p（*SCE*）	**20.00 ± 6.50%**	20.33 ± 8.62%（+1.7%）	29.00 ± 7.75%（+45.0%）	20.00 ± 10.00%（+0.0%）
N=40, $\rho = 0.2$, D1	速度（米/秒）	8.40 ± 6.93	10.49 ± 7.47 (+24.8%)	10.84 ± 7.97 (+29.0%)	**12.68 ± 7.93 (+51.0%)**
	WE 的 *p* 值	30.75 ± 9.94%	24.25 ± 12.70% (-21.1%)	22.50 ± 10.00%（-26.8%）	**12.25 ± 3.61%（-60.2%）**
	p（*SCE*）	19.00 ± 7.60%	16.00 ± 12.31%（-15.8%）	23.75 ± 11.25%（+25.0%）	**8.75 ± 4.37% (-53.9%)**

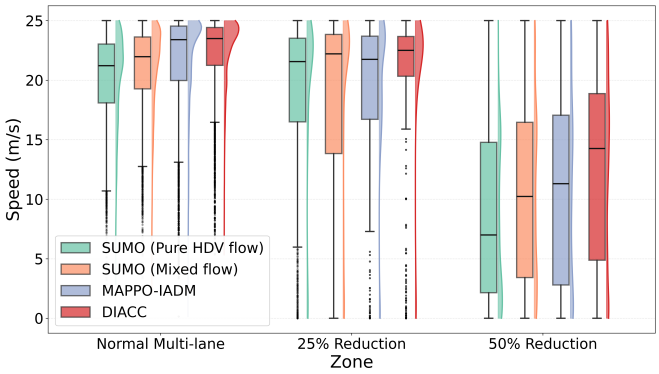
3）协作奖励机制通过采用温度退火的 softmin 聚合，使学习聚焦于交互密集场景。实验结果验证了这三项贡献的有效性。

D-IADM模块体现了其在增强行动者决策的局部交互感知方面的价值。与原始多智能体近端策略优化基线相比，MAPPO-IADM实现了更稳定的训练收敛，且碰撞率更低，如图7(a)–(c)所示。在测试场景中，MAPPO-IADM降低了等待事件 p(WEs)和安全关键事件 p(SCEs)的发生概率，同时在各种交通状况下提高了平均速度（表 III 和表 IV）。对CAV-CAV与CAV-HDV交互的差异化建模使每辆网联自动驾驶车辆能够更好地理解不同的周围车辆的行为模式，从而在复杂交通环境中做出更具适应性的决策。

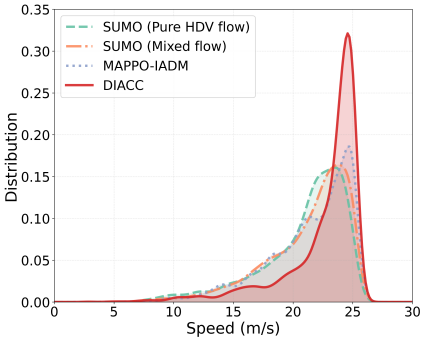
C-IEC模块验证了其在增强评论家的全局交通理解方面的作用，从而提供更有效的策略指导。完整的DIACC模型优于MAPPO-IADM，尤其是在高密度场景（$N = 30$ 和 $N = 40$）中；随着交通复杂度增加，这一优势更加明显。在零样本测试条件下，DIACC展现出比MAPPO-IADM更强的泛化能力，表明交互感知价值估计能为策略学习提供更稳健的指导。C-IEC中的ITDR模块能够捕捉车辆交互对交通动态的影响，使评论家能够提供更准确的价值估计，引导行动者更新以制定有效的协作策略。



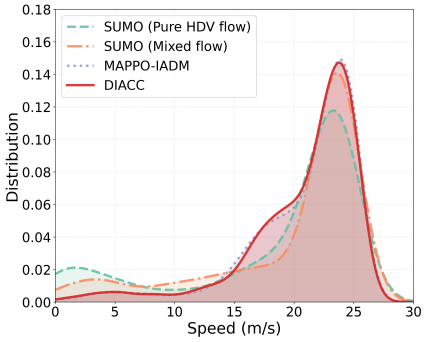
（a） $N = 30$：速度分布



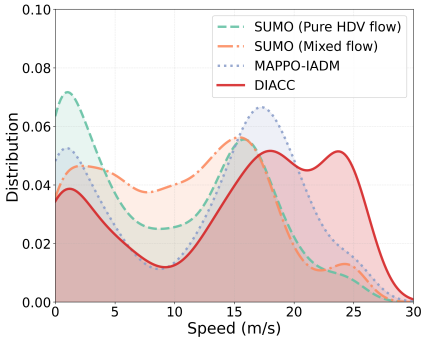
（b） $N = 40$：速度分布



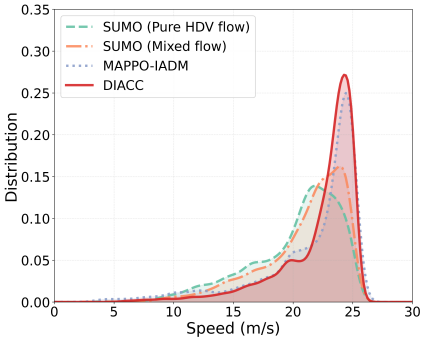
（c） $N = 30$，正常多车道路段



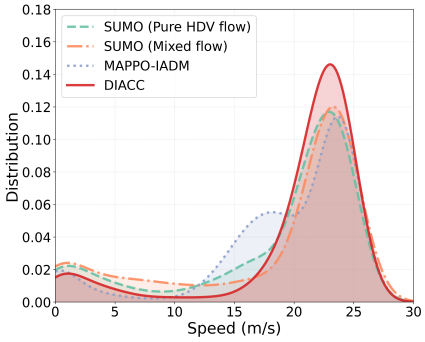
（d） $N = 30$，容量降低25%



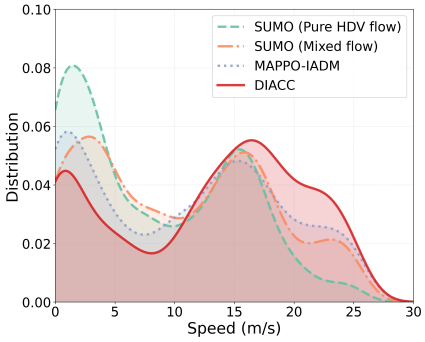
（e） $N = 30$，容量降低50%



(f) $N = 40$，正常多车道路段



(g) $N = 40$，容量降低25%



(h) $N = 40$，容量降低50%

图8：不同混合交通规模和容量削减场景下的速度分布比较。(a)–(b) 整体速度分布的箱线图和核密度估计。(c)–(h) 不同容量削减场景下的速度分布核密度估计。

奖励设计机制对于在交互密集场景中进行学习至关重要。温度参数分析（图7(d)–(f)）表明，固定的 τ 值会导致收敛不稳定（ τ较小）或无法区分场景难度（ τ较大）；而从$\tau = 2$退火至 $\tau = 0.05$ 的策略则能实现稳定训练，并将学习重点放在具有挑战性的交互上。自车奖励权重分析（图7(g)–(i)）表明， $w_e = 0.6$ 至 0。8能够在局部与全局奖励信号之间实现最佳平衡。这些发现证实，s oftmin聚合能有效引导学习注意力聚焦于面临复杂交互场景的智能体，从而确保在各种交通状况下全面改进策略。

然而，仍有相当大的改进空间

本研究模型的适应性和鲁棒性。在后续工作中，我们计划：

- 开发轻量级轨迹预测模块，以提高安全评估计算的准确性。
- 进一步探索泛化能力更强的基于预测的多智能体强化学习模型，以更好地适应规模更大的交通系统。
- 将双交互感知训练框架扩展至多样化交通场景，并建立统一的跨场景协同控制框架。

VI. 结论

混合交通环境中网联自动驾驶车辆的协同控制面临着分散式决策与全局交通优化之间的根本矛盾，而人工驾驶车辆不可预测的行为使这一问题更加复杂

。本文通过 DIACC策略应对这些挑战，该策略为 MAPPO 框架增添了双重交互感知能力。

所提出的方法包含三项相辅相成的技术贡献。首先，D-IADM模块通过分别对网联自动驾驶车辆间的协作交互和网联自动驾驶车辆与人工驾驶车辆间的自适应交互进行建模，使每辆网联自动驾驶车辆都能感知并应对周围车辆各自不同的行为特征。其次，C-IEC模块使评论家能够更深入地理解车辆交互如何影响交通演化，从而生成信息量更丰富的价值估计，为训练过程中的行动者策略更新提供指导。第三，奖励设计机制确保学习过程优先关注协同最具挑战性的场景，避免较简单交通区域中的智能体主导训练过程。

实验验证了这些贡献的有效性，结果表明，在各种交通状况下，安全关键事件减少的同时，通行效率也得到提升。此外，DIACC在零样本场景中仍保持这一优势，展现出交互感知学习所赋予的泛化能力。未来工作将着重拓展理论分析，并研究该框架在规模更大的混合交通系统和更多样化交通场景中的可扩展性。

VII. 附录

参考文献

[1] A. Katriniok、B. Rosarius 和 P. M¨ah¨onen，《交叉路口网联自动驾驶车辆的完全分布式模型预测控制：理论与车辆实验》，*IEEE*智能交通系统汇刊，第23卷，第10期，第18288–18300页，2022年。[2] X. He、J. Wu、Z. Huang、Z. Hu、J. Wang、A. Sangiovanni-Vincentelli 和 C. Lv，《面向安全自动驾驶的恐惧神经启发式强化学习》，*IEEE Transactions on Pattern Analysis and Machine Intelli- gence*，第46卷，第1期，第267–279页，2024年。[3] D. Xu、P. Liu、H. Li、H. Guo、Z. Xie 和 Q. Xuan，《面向高速公路混合交通中网联自动驾驶车辆协同控制的多视图图卷积网络强化学习》，*IEEE*智能车辆汇刊，第9卷，第1期，第2588–2599页，2024年。[4] D. Chen、A. Srivastava 和 S. Ahn，《利用网联自动驾驶汽车技术控制混合交通中高速公路合流瓶颈处的变道》，《交通研究*C*辑：新兴技术》，第123卷，第102950页，2021年。[5] M. ˇCi ˇc´c、X. Xiong、L. Jin 和 K. H. Johansson，《协调车辆编队以缓解高速公路瓶颈拥堵并提升通行能力》，*IEEE*智能交通系统汇刊，第23卷，第7期，第8959–8971页，2022年。[6] Y. Li、B. Pan、Z. Chen 和 L. Xing，《开发动态速度控制系统以降低高速公路瓶颈附近混合交通流的碰撞风险》，*IEEE*智能交通系统汇刊，2023年。[7] Y. Du、M. A. Makridis、C. M. Tamp`ere、A. Kouvelas 和 W. Shang-Guan，《在网联自动驾驶车辆参与的高速公路瓶颈处采用移动执行器进行自适应控制》，《交通研究*C*辑：新兴技术》，第156卷，第104319页，2023年。[8] H. Ding、L. Zhang、J. Chen、X. Zheng、H. Pan 和 W. Zhang，《混合车辆环境中瓶颈区域上游多个路段内网联自动驾驶车辆的基于MPC的动态速度控制》，《物理学*A*辑：统计力学及其应用》，第613卷，第128542页，2023年。[9] C. Wang、Y. Xu、J. Zhang 和 B. Ran，《基于深度强化学习的高速公路周期性瓶颈综合交通控制》，*IEEE*智能交通系统汇刊，第23卷，第9期，第15522–15535页，2022年。[10] M. Li、Z. Li、S. Wang 和 S. Zheng，《利用基于通信的软演员-评论家算法增强繁忙交通中的车辆合流控制协作》，IEEE智能交通系统汇刊，第24卷，第6期，第6491–6506页，2023年。[11] D. Chen、M. R. Hajidavalloo、Z. Li、K. Chen、Y. Wang、L. Jiang 和 Y. Wang，《面向混合交通中高速公路入口匝道合流的深度多智能体强化学习》，IEEE智能交通系统汇刊，第24卷，第11期，第11623–11638页，2023年。[12] L. Li、W. Zhao、C. Wang、A. Fotouhi 和 X. Liu，《用于混合交通中交互式合流策略的基于纳什双Q的多智能体深度强化学习》，Expert Systems with Applications，第237卷，121458，2024年，第。[13] X. Hu 和 J. Sun，《多车道高速公路合流区域网联自动驾驶车辆的轨迹优化》，《交通研究C辑：新兴技术》，第101卷，第111–125页，2019年。[14] J. Ying 和 Y. Feng，《混合交通状况下基础设施辅助协同驾驶与交叉口管理》，《交通研究C辑：新兴技术》，第158卷，第104443页，2024年。[15] S. Huang、D. Sun 和 M. Zhao，《交通信号灯附近采用T-CPS的混合车辆群基于分布式模型预测控制的分层协同控制》，IEEE智能交通系统汇刊，2024年。[16] K. Hou、F. Zheng、X. Liu 和 G. Guo，《多车道高速公路混合交通的协同入口匝道合流控制模型》，IEEE智能交通系统汇刊，第24卷，第10期，第10774–10790页，2023年。[17] J. Shi、K. Li、C. Chen、W. Kong 和 Y. Luo，《基于最优终态相图和灵活高速公路合流点的混合交通协同合流策略》，IEEE智能交通系统汇刊，第24卷，第10期，第11185–11197页，2023年。[18] S. Jing、F. Hui、X. Zhao、J. Rios-Torres 和 A. J. Khattak，《网联自动驾驶车辆最优合流序列与入口匝道合流控制的合作博弈方法》，IEEE智能交通系统汇刊，第20卷，第11期，第4234–4244页，2019年。[19] J. Ding、L. Li、H. Peng 和 Y. Zhang，《面向网联自动驾驶车辆的基于规则的协同合流策略》，IEEE智能交通系统汇刊，第21卷，第8期，第3436–3446页，2020年。[20] L. Li、C. Qian、J. Gan、D. Zhang、X. Qu、F. Xiao 和 B. Ran，《DCoMA：面向混合交通状况下匝道车辆的动态协同合流辅助策略》，《交通研究C辑：新兴技术》，第165卷，第104700页，2024年。[21] K. Sun、S. Gong、Y. Zhou、Z. Chen、X. Zhao 和 X. Wu，《针对单车变道场景、采用平滑跟踪目标切换以缓解交通振荡的多车协同控制方案》，《交通研究C辑：新兴技术》，第159卷，第104487页，2024年。[22] Y. Zhang、W. Macke、J. Cui、S. Hornstein、D. Urieli 和 P. Stone，《学习用于缓解交通拥堵的鲁棒多智能体驾驶策略》，Neural Computing and Applications，第1–14页，2023年。[23] M. Li、Z. Cao 和 Z. Li，《采用密集通信强化学习的增强型混合车辆编队控制，以缓解交通振荡》，IEEE物联网期刊，2024年。[24] M. Bouton、A. Nakhaei、K. Fujimura 和 M. J. Kochenderfer，《面向繁忙交通合流的协作感知强化学习》，载于2019年IEEE智能交通系统会议（ITSC），第3441–3447页，IEEE，2019年。[25] A. Irshayyid、J. Chen 和 G. Xiong，《基于强化学习的高速公路自动驾驶车辆控制综述》，Green Energy and Intelligent Transportation，第3卷，第4期，第100156页，2024年。[26] T. Chu、J. Wang、L. Codec`a 和 Z. Li，《面向大规模交通信号控制的多智能体深度强化学习》，IEEE智能交通系统汇刊，第21卷，第3期，第1086–1095页，2020年。[27] M. Wang、Y. Chen、Y. Kan、C. Xu、M. Lepech、M.-O. Pun 和 X. Xiong，《变化干预频率下采用集中式评论家和分布式演员的交通信号周期控制》，IEEE智能交通系统汇刊，第25卷，20104，2024年12，20085期，第–页。[28] C. Spatharis 和 K. Blekas，《面向无信号交叉口交通区域自动驾驶的多智能体强化学习》，Journal of Intelligent Transportation Systems，第28卷，第1期，第103–119页，2024年。[29] Y. Hu、A. Nakhaei、M. Tomizuka 和 K. Fujimura，《合流场景下采用自适应策略的交互感知决策》，载于2019年IEEE/RSJ智能机器人与系统国际会议（IROS），第151–158页，IEEE，2019年。[30] J. Yang、J. Zhang 和 H. Wang，《通过多智能体深度强化学习实现软件定义物联网中的城市交通控制

表V：符号说明（按类别组织）。

符号	定义
索引与集合	
i, j	车辆索引（例如，用于网联自动驾驶车辆/人工驾驶车辆）
t	时间步索引
l	车道索引
n	网联自动驾驶车辆数量（$\|\mathcal{N}\|$）
$\mathcal{N}$	混合交通中的网联自动驾驶车辆集合
$\mathcal{S}$	全局状态空间
$\mathcal{A}$	联合动作空间
$\mathcal{O}$	联合观测空间
$\mathcal{M}_s^i$	车辆 i的相邻车辆
$\mathcal{M}_{s,H}^i, \mathcal{M}_{s,C}^i$	警告范围/碰撞范围内的车辆
N^i	相邻人工驾驶车辆/网联自动驾驶车辆
$\mathcal{G}^{s,H/C}$	CAV i周围人工驾驶车辆或网联自动驾驶车辆的数量
$\mathcal{V}_{s,H/C}^i, \mathcal{E}_{s,H/C}^i$	CAV i的HDV/CAV交互图
$\mathcal{G}_{global}$	HDV/CAV交互图中的节点/边
$\mathcal{V}, \mathcal{E}$	全局车辆交互图
	全局交互图中的节点/边
变量（状态、动作、观测等）	
s_t, a_t^i, o_t^i	全局状态；车辆 i 在时间 t的动作和观测
$\tilde{o}_{t+1}^i$	包含 $\hat{a}_t^i$ 和 a_t^i的增广观测
$m_t^i = (x_t^i, y_t^i, v_t^i, \theta_t^i)$	车辆 i 的状态（位置、速度、航向等）
$(\Delta x_t^{i,j}, \Delta y_t^{i,j})$	车辆 j 相对于车辆 i 的相对位置
$\Delta v_t^{i,j}$	车辆 j 相对于车辆 i 的相对速度
c^j	车辆类型 j (网联自动驾驶车辆/人工驾驶车辆)
$n^l, D^l, \bar{v}^l, \rho^l$	车道 l中的车辆数、密度、平均速度和网联自动驾驶汽车渗透率
r_e^i, r_g, r_{done}	自车奖励、全局奖励、完成奖励
$r_{e,v}^i, r_{e,w}^i, r_{e,c}^i, r_{e,tp}^i$	自车奖励分量（速度、警告、碰撞、时间）
w^i	智能体 i的Softmin权重
$\bar{r}_e$	Softmin聚合自车奖励
$\mathbf{F}_t^{ego}, \mathbf{F}_t^{left}, \mathbf{F}_t^{right}, \mathbf{F}_{static}$	车道统计和静态道路结构信息
$碰撞时间_t^{i,j}$, $d_{t,long}^{i,j}$, $\Delta v_{long}^{i,t}$	碰撞时间、最小纵向头尾距离和纵向相对速度
α_{ij}	节点 i 与节点 j之间的注意力/重要性系数
$\hat{a}_t^i$	行动者指令及其横向/纵向分量
$\mathbf{h}_{t,H,i}^{a_{lat}}, \mathbf{f}_{t,H,i}^{a_{long}}, \mathbf{h}_t^{H,i}$	HDV节点特征 ，交互特征 ，和增强节点特征
$\mathbf{f}_t^{C,i}$	CAV交互特征
$\mathbf{f}_t^{t,i}, \mathbf{f}_t^i, \mathbf{f}_t^d$	局部上下文、融合观测、全局动态特征
$\mathbf{h}_t^i, \overrightarrow{\mathbf{h}}_t^i$	全局图节点特征和增强节点特征
Q_h, K_h, V_h	注意力机制中的查询、键和值
$\hat{\mathbf{a}}, \hat{\mathbf{a}}$	单层前馈网络（用于注意力或节点更新）
参数（常数、超参数、函数）	
R	奖励函数
$P(s_{t+1}\|s_t, a_t^1, \ldots, a_t^n)$	状态转移概率
γ	折扣因子
π_θ	参数为 θ的策略
$J(\pi_\theta)$	期望累积奖励
V_ϕ	集中式评论家
L	时间范围长度（回合步数）
L^{CLIP}, L_V, V^{CLIP}	近端策略优化的裁剪目标和价值预测
$r_t^i(\theta)$	概率比（新策略与旧策略）
ϵ	裁剪超参数（近端策略优化）
$\hat{A}_t, \delta_t$	优势估计器与时序差分误差
λ	广义优势（广义优势估计）参数
T	广义优势估计中的回合长度
w_e, w_g	$\bar{r}_e$ 和 r_g的权重
$w_{e,v}, w_{e,w}, w_{e,c}, w_{e,tp}$	自车奖励分量的权重
τ	softmin 聚合的温度参数
$\tau_{init}, \tau_{final}$	退火的初始和最终温度
T_{anneal}	退火周期（总训练步数的一半）
v_{max}	规定的最高速度
$d_{th,w}, d_{th,c}$	距离阈值（警告 / 碰撞）
v_{th}	惩罚调节的速度阈值
$p_{e,v}, p_{e,c}, p_{g,v}$	奖励偏置项
L_{veh}, W_{veh}^i	车辆长度与宽度
S_H, S_N	隐藏层大小（例如，用于图注意力网络或多层感知机）
$W, W_Q, W_K, W_V, W_O, d_k$	注意力投影矩阵和键维度
B, H	注意力头数量（图注意力网络 / 交叉注意力）
$\sigma, \sigma(\cdot)$	熵权和激活函数

方法”，《IEEE智能交通系统汇刊》，第22卷，第6期，第3742–3754页，2020年。[31] H. Huang、Z. Hu、M. Li、Z. Lu和X. Wen，“利用一种新颖的多智能体深度强化学习方法协同优化交通信号灯与车辆速度”，《IEEE车辆技术汇刊》，2024年。[32] ´A. Madridano、A. Al-Kaff、D. Mart´ın和A. De La Escalera，“多机器人系统的轨迹规划：方法与应用”，Expert Systems with Applications，第173卷，114660页，2021年。[33] R. Bautista-Montesano、R. Galluzzi、K. Ruan、Y. Fu和X. Di，“无信号交叉口的自主导航：一种结合强化学习与模型预测控制的方法”，《交通研究C辑：新兴技术》，第139卷，103662页，2022年。[34] C. Yu、X. Wang、X. Xu、M. Zhang、H. Ge、J. Ren、L. Sun、B. Chen和G. Tan，“基于动态协同图的高速公路自动驾驶分布式多智能体协同学习”，《IEEE智能交通系统汇刊》，第21卷，第2期，第735–748页，2020年。[35] S. K. S. Nakka、B. Chalaki和A. A. Malikopoulos，“用于网联自动驾驶汽车汇入道路场景的多智能体深度强化学习协同框架”，载于2022年美国控制会议（ACC），第3297–3302页，IEEE，2022年。[36] F. A. Oliehoek、C. Amato等，《分散式部分可观测马尔可夫决策过程简明导论》，第1卷。施普林格，2016年。[37] C. Yu、A. Velu、E. Vinitsky、J. Gao、Y. Wang、A. Bayen和Y. Wu，“近端策略优化在协作式多智能体博弈中的惊人效果”，《神经信息处理系统进展》，第35卷，第24611–24624页，2022年。[38] J. Schulman、F. Wolski、P. Dhariwal、A. Radford和O. Klimov，“近端策略优化算法”，arXiv预印本 arXiv:1707 06347，2017年。。[39] K. Cho、B. Van Merri¨enboer、C. Gulcehre、D. Bahdanau、F. Bougares、H. Schwenk和Y. Bengio，“用于统计机器翻译的循环神经网络编码器—解码器短语表示学习”，arXiv预印本 arXiv:1406.1078，2014年。[40] Y. Bengio、J. Louradour、R. Collobert和J. Weston，“课程学习”，《第26届年度国际机器学习会议论文集》，第41–48页，2009年。[41] P. Veli˘ckovi´c、G. Cucurull、A. Casanova、A. Romero、P. Lio和Y. Bengio，“图注意力网络”，载于《国际学习表征会议》，2018年。[42] P. A. Lopez、M. Behrisch、L. Bieker-Walz、J. Erdmann、Y.-P. Fl¨otter¨od、R. Hilbrich、L. L¨ucken、J. Rummel、P. Wagner和E. Wießner，“使用SUMO进行微观交通仿真”，载于2018年第21届国际智能交通系统会议（ITSC），第2575–2582页，IEEE，2018年。[43] M. Wang、Y. Chen、Y. Cai、A. Pang、Y. Xie、Z. Ma、C. Xu、K. Jiang、D. Wang、L. Roullet、C. S. Chen、Z. Cui、Y. Kan、M. Lepech和M.-O. Pun，“TranSimHub：用于多模态感知与决策的统一空地仿真平台”，arXiv预印本 arXiv:2510.15365，2025年。[44] J. G. Kuba、R. Chen、M. Wen、Y. Wen、F. Sun、J. Wang和Y. Yang，“多智能体强化学习中的信赖域策略优化”，arXiv预印本 arXiv:2109.11251，2021年。[45] M. Treiber和A. Kesting，“交通流动力学”，《交通流动力学：数据、模型与仿真》，施普林格出版社，柏林—海德堡，第983–1000页，2013年。[46] J. Erdmann，“SUMO变道模型”，载于《利用开放数据建模移动性：第2届SUMO会议论文集》，2014年，德国柏林，2014年5月，第105–123页，施普林格，2015年。[47] W. Zhou、X. Wang、Y. Glaser、X. Wu和X. Xu，“基于自然驾驶研究数据中的安全关键型跟驰事件开发改进的自动预防制动系统”，《事故分析与预防》，第178卷，106834页，2022年。

请在此处
放置照片

Zhengxuan Liu 分别于2016年和2019年在中国天津的天津大学获得工学学士学位和硕士学位。2019年至2021年，她在中国汽车技术研究中心（天津）汽车工程研究院有限公司担任EEA工程师，专注于主动安全领域。她目前正在电气与信息工程学院攻读控制科学与工程博士学位。她目前的研究兴趣包括混合交通系统的协同控制与安全问题。

地点
照片
HERE

Yuxin Cai 于2022年获得新加坡南洋理工大学（NTU）工学学士学位。她目前是新加坡NTU机械与航空航天工程学院的博士生。她的主要研究兴趣包括强化学习、多机器人协作和基础模型。

地点
照片
HERE

Yijing Wang 于2000年和2004年分别获得燕山大学控制理论与控制工程硕士学位和北京大学控制理论博士学位。2004年，她加入天津大学电气与信息工程学院，现任正教授。她的研究兴趣包括切换/混杂系统的分析与控制以及鲁棒控制。

地点
照片
HERE

Xiangkun He（IEEE高级会员）现任电子科技大学UESTC100青年教授。此前，他曾任新加坡南洋理工大学研究员，并于2019年至2021年担任华为诺亚方舟实验室高级研究科学家。他于2019年获得清华大学车辆与运载学院博士学位。他已发表60余篇论文，并持有8项授权专利。他的研究兴趣包括强化学习、可信人工智能、自动驾驶车辆和机器人学。他曾获得多项奖项与荣誉，其中包括2019年清华大学优秀博士学位论文奖、2020年IEEE ICMA最佳论文入围奖、2021年华为重大技术突破奖、2022年CVCI最佳论文亚军奖，以及2022年阿里巴巴全球未来汽车挑战赛智能算法决赛亚军。他担任50多种知名期刊和会议的审稿人。

地点
照片
HERE

Chen Lv（IEEE高级会员）于2016年获得中国清华大学汽车工程系博士学位。2014年至2015年，他在加州大学伯克利分校电气工程与计算机科学系担任联合博士研究员。他现任新加坡南洋科技大学副教授。他的研究重点是先进车辆和人机系统，已发表论文100余篇，并在中国获得12项授权专利。他曾获多项奖项和荣誉，包括2012年英国机械工程师学会高度赞誉论文奖、2014年日本NSK优秀机械工程论文奖、2016年清华大学优秀博士论文奖、2018年IEEE智能车辆研讨会/特别分会最佳论文奖、2020年《汽车创新》最佳论文奖、2021年CVPR Waymo开放数据集挑战赛获胜者称号，以及2022年《Machines》青年研究者奖。他担任《IEEE智能交通系统汇刊》《IEEE车辆技术汇刊》和《IEEE智能车辆汇刊》的副编辑；并担任《IEEE智能交通系统杂志》《IEEE/ASME机电一体化汇刊》和《应用能源》的客座编辑。

地点
照片
HERE

Zhiqiang Zuo（IEEE高级会员）于2001年获得燕山大学控制理论与控制工程硕士学位，并于2004年获得中国北京大学控制理论博士学位。2004年，他加入天津大学电气与信息工程学院，现任正教授。2008年至2010年，他在香港城市大学数学系担任研究员。2013年至2014年，他在加州大学河滨分校担任访问学者。他的研究兴趣包括非线性控制、鲁棒控制、多智能体系统和信息物理系统。

Zuo博士是《富兰克林研究所期刊》（爱思唯尔）的副编辑。