Modeling and Simulation 建模与仿真, 2025, 14(9), 21-29
Published Online September 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.149581

Hans汉斯

基于驾驶风格识别与协作学习的交通信号强化学习方法

李　琦[1]，周鲁露[2]

[1]上海电科智能系统股份有限公司，上海
[2]上海市公安局交通管理总队，上海

收稿日期：2025年8月4日；录用日期：2025年8月28日；发布日期：2025年9月5日

摘　要

随着城市化进程的加快，城市交通拥堵问题日益严峻，已成为制约城市可持续发展的核心瓶颈。近年来，智能交通信号控制技术迅速发展，尤其是基于多智能体深度强化学习(Multi-Agent Deep Reinforcement Learning, MADRL)的方法，为缓解交通拥堵提供了新思路。然而，现有研究普遍忽视了驾驶员行为的异质性，并在大规模交叉口网络中面临状态空间维度过高与智能体协作效率不足的双重挑战。针对上述问题，本文提出一种融合驾驶风格识别与邻居协作机制的多智能体深度强化学习算法——CDS-DQN (Collaborative Driving Style-aware Deep Q-Network)。该算法设计了一种轻量化驾驶风格识别模块，通过量化车辆行为特征(如激进型、普通型、保守型)，构建“有效占有率”指标作为状态输入，以增强对微观交通特性的感知能力。同时，提出了一种邻居状态共享机制，使各智能体能够获取相邻交叉口的关键信息，实现局部协同感知并缓解多智能体系统中的环境非平稳性问题。基于SUMO (Simulation of Urban Mobility)仿真平台构建的城市路网环境中对该算法进行了实验验证。结果表明，CDS-DQN在平均等待时间、队列长度与通行效率等指标上，均优于传统固定配时控制、独立DQN与主流的MA2C算法，充分展示了其有效性与先进性。

关键词

多智能体深度强化学习，智能交通信号控制，驾驶风格，协作学习

A Reinforcement Learning Approach for Traffic Signal Control Integrating Driving Style Recognition and Collaborative Learning

Qi Li[1], Lulu Zhou[2]

[1]Shanghai SEARI Intelligent System Co., Ltd., Shanghai

文章引用: 李琦, 周鲁露. 基于驾驶风格识别与协作学习的交通信号强化学习方法[J]. 建模与仿真, 2025, 14(9): 21-29.
DOI: 10.12677/mos.2025.149581

[2]Traffic Police Corps of Shanghai Municipal Public Security Bureau, Shanghai

Received: Aug. 4th, 2025; accepted: Aug. 28th, 2025; published: Sep. 5th, 2025

Abstract

With the accelerating process of urbanization, urban traffic congestion has become increasingly severe, emerging as a critical bottleneck for sustainable urban development. In recent years, intelligent traffic signal control technology has undergone rapid development, particularly methods based on Multi-Agent Deep Reinforcement Learning (MADRL), which offer new avenues for alleviating traffic congestion. However, existing research often overlooks the heterogeneity of driver behavior and faces the dual challenges of high-dimensional state spaces and inefficient agent collaboration in large-scale intersection networks. To address these issues, this paper introduces a novel Multi-Agent Deep Reinforcement Learning algorithm, the Collaborative Driving Style-aware Deep Q-Network (CDS-DQN). The proposed algorithm features a lightweight driving style recognition module that quantifies vehicle behavioral characteristics (e.g., aggressive, normal, and conservative) to formulate an "effective occupancy" metric. This metric serves as a state input, enhancing the agent's perceptual capabilities regarding micro-level traffic dynamics. Furthermore, a neighbor state-sharing mechanism is proposed, enabling each agent to access critical information from adjacent intersections. This facilitates local collaborative perception and mitigates the environmental non-stationarity problem inherent in multi-agent systems. The algorithm was experimentally validated in an urban road network environment constructed on the SUMO (Simulation of Urban Mobility) platform. The results demonstrate that CDS-DQN outperforms traditional fixed-time control, independent DQN, and the state-of-the-art MA2C algorithm across key performance indicators, including average waiting time, queue length, and traffic throughput. These findings fully showcase the effectiveness and advanced nature of the proposed algorithm.

Keywords

Multi-Agent Deep Reinforcement Learning, Intelligent Traffic Signal Control, Driving Style, Collaborative Learning

1. 引言

随着全球城市化进程的加速，机动车保有量激增，交通拥堵、能源消耗与环境污染已成为严峻的城市挑战[1]。传统的固定配时控制(Fixed-Time Control) [2]因其无法适应动态变化的交通流，常常导致绿灯时间的浪费和不必要的延误，加剧了路网的拥堵。尽管自适应系统如 SCATS [3]和 SCOOT [4]能够在一定程度上缓解交通压力，但它们高度依赖复杂的工程规则和昂贵的部署成本，在多变的交通环境下优化能力有限。

近年来，以强化学习(Reinforcement Learning, RL) [5]为代表的人工智能技术为交通信号控制问题提供了全新的解决范式。RL 通过智能体(Agent)与环境的交互试错来自动学习最优控制策略。早期研究多集中于单交叉口控制[6]，将每个交叉口视为独立的智能体。然而，这种方式忽略了交叉口间的强关联性——一个路口的拥堵极易传导至相邻路口，因而难以实现网络层面的全局最优。

为此，多智能体强化学习(MARL) [7]被引入该领域，将各交叉口建模为相互协作的智能体。尽管

MARL 展现了巨大潜力，但在大规模实际应用中仍面临两大核心挑战：(1) 维度灾难与环境非平稳性。随着智能体数量增多，联合状态–动作空间呈指数级增长，导致学习效率低下。同时，单个智能体策略的更新会改变其他智能体的学习环境，造成了 MARL 固有的非平稳性问题。(2) 状态表征的局限性。现有方法大多采用车流量、队列长度等宏观交通参数作为状态输入[8]，这些参数有效地反映了交通拥堵的整体状况。然而，较少有研究在强化学习模型中直接考虑微观层面驾驶行为的异质性。如研究[9]所示，不同驾驶风格对车辆的启动延时和跟驰行为有显著影响，进而影响交叉口的实际通行能力。若状态表示中未能包含此类微观信息，可能会限制模型对交通流动态的精细感知能力，从而影响控制策略的性能上限。虽然已有研究尝试对驾驶风格进行识别[9]，但其模型通常较为复杂，难以直接、轻量化地集成到多智能体强化学习的状态空间中用于实时控制。

针对上述挑战，本文提出一种融合驾驶风格感知与协同感知的多智能体深度强化学习算法(CDS-DQN)，我们将 DQN 选为基础框架，主要基于其与问题特性的高度契合及其作为基石算法的稳健性。交通信号控制本质上是一个离散动作空间决策问题(选择下一相位)，这与 DQN 通过价值函数逼近来评估离散动作的方法天然契合。此外，DQN 作为深度强化学习的基石算法，其原理清晰、实现稳定，这使我们能够将研究重点聚焦于验证本文提出的“有效占有率”状态和邻居协作机制这两个核心创新点的有效性，而无需应对更复杂的 Actor-Critic 架构可能引入的额外变量和不确定性。本文的主要贡献可概括为以下两点：

(1) 提出一种融合驾驶风格感知的状态表征方法：通过设计一个简化的驾驶风格分类模块，并在此基础上构建“有效占有率”指标，使智能体能够感知微观驾驶行为差异，从而更精确地评估车道的拥堵程度与通行潜力。

(2) 设计一种高效的邻居协作机制：通过让智能体共享关键状态信息，实现局部范围内的协同感知。这不仅为智能体决策提供了更全面的信息视角，还有效缓解了多智能体环境的非平稳性，提升了学习的效率和稳定性。

本文通过将微观驾驶行为特征与宏观交叉口协作机制相结合，为多智能体交通信号控制提供了新的视角，旨在提升城市交通网络的整体运行效率。

2. 研究背景

2.1. 深度强化学习在交通信号控制中的应用

深度强化学习(DRL)通过深度神经网络(DNN)来近似值函数或策略函数，成功解决了强化学习在处理高维状态空间时的难题。深度 Q 网络(DQN) [10]是其中的里程碑算法，它通过经验回放(Experience Replay)和目标网络(Target Network)两大技术，打破了数据间的相关性并稳定了训练目标，从而稳定了 Q-learning 的训练过程。许多研究将其应用于交通信号控制，并取得了良好效果。后续的改进算法如 Double DQN [11]、Dueling DQN [12]等也通过解决 Q 值高估、优化网络结构等方式进一步提升了算法性能。

2.2. 多智能体协作机制

在多交叉口场景下，智能体间的协作至关重要。现有的协作方式大致可分为：共享奖励、共享状态及动作、共享值函数等。例如，Chu 等人[13]提出的 MA2C 算法通过“指纹”信息来捕捉邻居的策略变化。Haddad 等人[14]则通过让智能体直接接收邻居的状态、动作和奖励来实现协作。然而，完全共享所有信息会带来巨大的通信开销和计算负担。因此，如何设计一种高效且轻量化的协作机制是当前研究的热点。

2.3. 驾驶风格识别

驾驶风格是影响交通流稳定性和安全性的关键因素。研究表明，激进型驾驶员通常伴随着更短的跟

车时距和更剧烈的加减速行为，而保守型驾驶员则相反[9]。传统的交通工程理论通过引入“客车当量”来处理重型车等异质车辆，但很少有研究将普通客车驾驶员的风格差异纳入信号控制模型。文献[9]中提出了一种基于跟驰模型(IDM)参数标定的方法来识别驾驶风格，但该方法数据需求量大，计算复杂，不适用于实时控制。本文将对其进行简化，使其更适用于 DRL 框架。

3. 模型设计

本文提出的 CDS-DQN 模型框架如图 1 所示。每个交叉口的智能体首先处理自身和邻居交叉口的观测数据，生成联合状态，然后输入到 DQN 网络中，输出动作选择的 Q 值，最后根据奖励来优化网络。

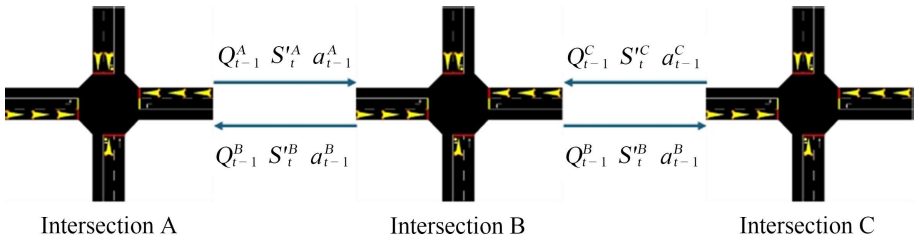


Figure 1. The collaborative framework of the CDS-DQN model
图 1. CDS-DQN 模型协作框架示意图

3.1. 状态空间定义

为了让智能体全面感知交通环境，我们设计了包含宏观和微观信息的联合状态。

(1) 基础状态

对于单个交叉口 k，其基础状态 s_k 由两部分组成：

1) 队列长度(Queue Length) $Q_k = \{q_1, q_2, \cdots, q_m\}$：各进口道的车辆排队长度，直接反映了当前的拥堵程度。

2) 当前相位(Current Phase) P_k：当前绿灯相位，告知智能体当前的路权分配情况。

(2) 驾驶风格感知的有效占有率(Effective Occupancy)

传统的状态表示忽略了驾驶员行为。例如，一条排满了“保守型”驾驶员的车道，其通行效率远低于排满了“激进型”驾驶员的车道。为了量化这一差异，我们提出“有效占有率” O_{lane}^{*}。

首先，我们根据车辆在一段时间内的平均速度和加速度标准差，通过一个简化的分类器将其划分为三类：激进型(Aggressive)、普通型(Normal)和保守型(Conservative)。然后为每种类型分配一个权重 w_{style}。该权重的设置参考文献[9]，以普通型(Normal)驾驶员为基准，其行为代表了标准的交通流特性，因此其权重设为 $w_{norm} = 1.0$，激进型(Aggressive)驾驶员因其更短的启动延时和跟车时距能有效提高车道通行能力，故为其分配一个小于 1 的权重 $w_{agg} = 0.8$，保守型(Conservative)驾驶员则会降低通行效率，因此分配一个大于 1 的权重 $w_{con} = 1.2$。本文所采用的具体数值是在上述原则指导下，通过初步的超参数扫描和经验调优确定的，旨在找到一组能够在不过度影响模型稳定性的前提下，最有效地表征不同驾驶风格对交通流影响的数值，从而在我们的仿真环境中取得了最佳性能。权重小于 1 表示通行能力强，大于 1 表示通行能力弱。

车道有效占有率 O_{lane}^{*} 定义为：

$$O_{lane}^{*} = O_{lane} \times \frac{\sum_{i=1}^{N_{lane}} w_{style,i}}{N_{lane}} \tag{1}$$

其中，O^*_{lane} 是车道的传统占有率(车辆总长度/车道长度)，N_{lane} 是车道上的车辆数，$w_{style,i}$ 是第 i 辆车的风格权重。有效占有率综合了车辆数量和其驾驶特性，能更准确地反映车道的拥堵程度和通行潜力。

(3) 联合状态

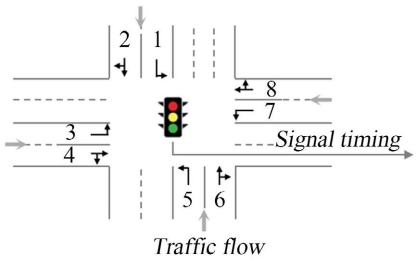
智能体 k 的最终输入状态 $\hat{S}_k$ 由其自身状态和来自邻居的状态拼接而成：

$$\hat{S}_k = \text{Concat}\left(S_k, O^*_k, \overline{O}^*_{neighbors}\right) \tag{2}$$

其中，O^*_k 是交叉口 k 所有车道的有效占有率，$\overline{O}^*_{neighbors}$ 是其所有邻居交叉口传递过来的“有效占有率”的平均值或拼接向量。这种设计使智能体既能关注自身细节，又能了解周边宏观态势，从而做出更优的协同决策。

3.2. 动作空间

我们将每个交叉口的信号控制简化为选择下一个绿灯相位。如图 2 所示，对于一个标准的十字交叉口，动作空间 A 通常包含四个主要相位：东西向直行、东西向左转、南北向直行、南北向左转。智能体在每个决策时刻选择其中一个动作执行。为了保证安全，相位切换之间会设置固定的黄灯时间。



action	phase	lanes in green
NSLg		1 5
NSg		2 6
EWLg		3 7
EWg		4 8

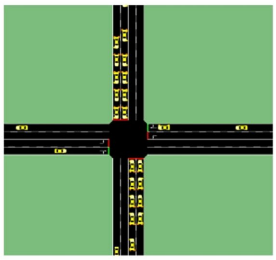


Figure 2. Illustration of intersection phases and the action space
图 2. 交叉口相位与动作空间示意图

3.3. 奖励函数

奖励函数的设计目标是引导智能体学习到减少拥堵的策略。我们采用一个简单而直观的奖励函数，即交叉口所有进口道车辆的平均等待时间的负值。

$$R_t = -\frac{\sum_{i=1}^{m} W_{t,i}}{N_{total}} \tag{3}$$

其中，$W_{t,i}$ 是第 i 条进口道上所有车辆在 t 时刻的总等待时间，N_{total} 是交叉口范围内的车辆总数。该奖励函数的目标是让智能体最大化这个值，也就是最小化平均等待时间。

3.4. 算法框架

本文提出的 CDS-DQN 算法基于经典的 DQN 框架。智能体将 t 时刻的联合状态 $\hat{S}_k$ 输入 Q 网络。得

到每个动作的 Q 值 $Q\left(\hat{S}_t,a;\theta\right)$。在训练阶段，智能体采用 $\epsilon-greedy$ 策略进行探索。智能体执行动作 a_t 后，环境转移到新状态 $\hat{S}_{t+1}$ 并返回奖励 R_t。我们将经验元组 $\left(\hat{S}_t,a_t,R_t,\hat{S}_{t+1}\right)$ 存入经验回放池。

训练时，从回放池中随机抽取一批样本，通过最小化损失函数来更新 Q 网络参数 θ。损失函数定义为：

$$L\left(\theta\right)=\mathbb{E}\left[\left(y_t-Q\left(\hat{S}_t,a_t;\theta\right)\right)^2\right] \tag{4}$$

其中，目标值 $y_t=R_t+\gamma\max_{a'}Q\left(\hat{S}_{t+1},a';\theta^-\right)$，$\theta^-$ 是目标网络的参数，它会定期从主 Q 网络复制，以保证训练的稳定性。

4. 仿真实验

4.1. 实验设置

我们在开源交通仿真软件 SUMO 中搭建了一个 4 × 4 的网格状交通路网，包含 16 个信号控制交叉口，如图 3 所示。道路为双向四车道，路段长度为 400 米，如图 2 所示。车辆的生成遵循随机韦布尔分布，以模拟真实交通流的到达模式。在仿真过程中，我们设置了三种驾驶风格的车辆：激进型(20%)、普通型(60%)、保守型(20%)，并为其分别设置了符合其行为特征的跟驰模型参数每个决策步长为 5 秒，相位切换间的黄灯时间固定为 3 秒。

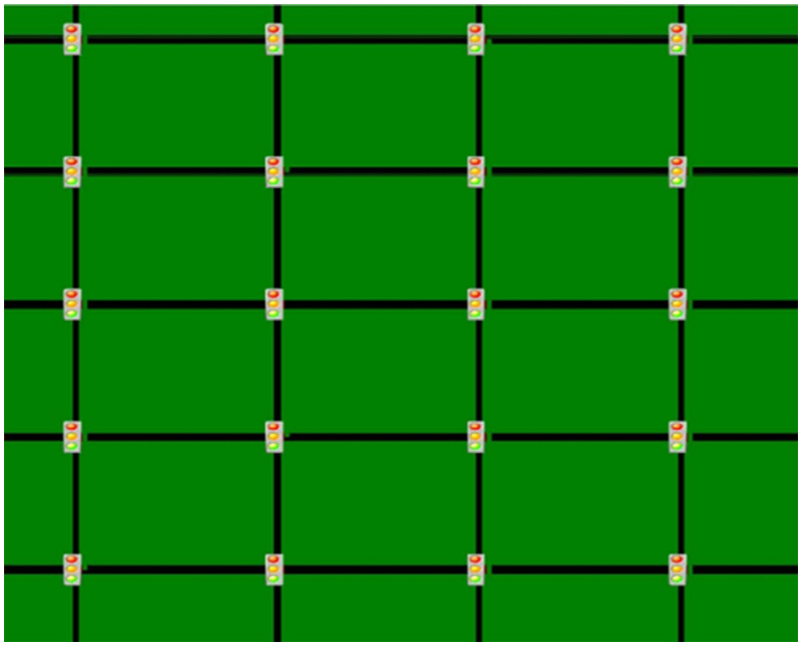


Figure 3. The simulated traffic network environment
图 3. 仿真交通网络环境

4.2. 实验结果与分析

为验证本文算法的性能，我们选取了以下几种方法作为对比。固定配时(Fixed-Time, FT)：一种经典的控制方法，周期和信比固定；独立 Q 学习(Independent DQN, IDQN)：每个智能体独立 DQN 算法，不进行任何协作；MA2C：一种基于 Actor-Critic 框架的先进多智能体强化学习算法，通过邻居指纹进行协作；(CDS-DQN)：融合了驾驶风格与邻居协作的 DQN 算法。

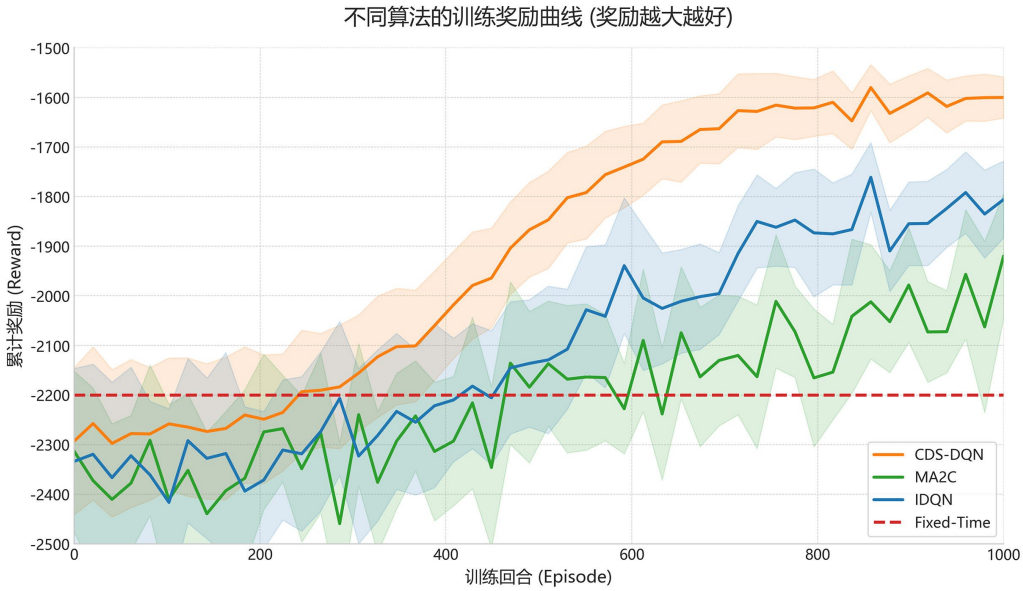


Figure 4. Comparison of cumulative reward curves for different algorithms during the training process
图 4. 同算法在训练过程中的累计奖励曲线对比

图 4 展示了本文提出的 CDS-DQN 算法与其他三种对比算法在学习过程中的效率和性能。可以看出，CDS-DQN 的奖励曲线从训练初期就显著高于其他算法，并以最快的速度收敛到最高的稳定水平。这表明，得益于对驾驶风格的精准感知和高效的邻居协作机制，CDS-DQN 能够更快速地学习到有效的控制策略。相比之下，MA2C 的学习速度次之，而 IDQN 由于缺乏协作，学习过程最慢且性能最差。Fixed-Time 作为非学习方法，其奖励值保持不变且处于最低水平。

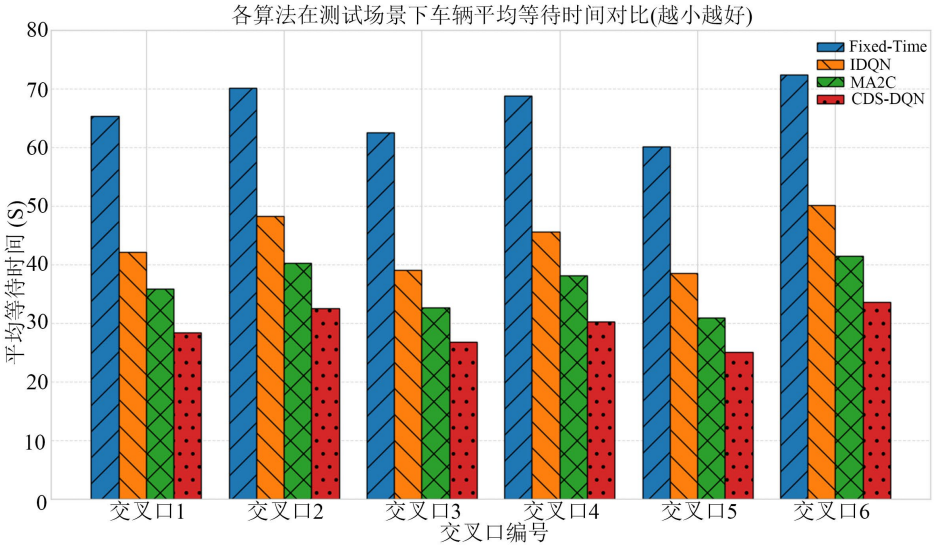


Figure 5. Performance comparison of average vehicle waiting time for different algorithms in the multi-intersection test network
图 5. 各算法在多交叉口测试网络中车辆平均等待时间的性能对比

从图 5 可以看出，在所有被抽样测试的交叉口中，本文提出的 CDS-DQN 算法(红色柱)对应的车辆平均等待时间均为最低。这直观地证明了其在疏导交通、减少拥堵方面的卓越性能。

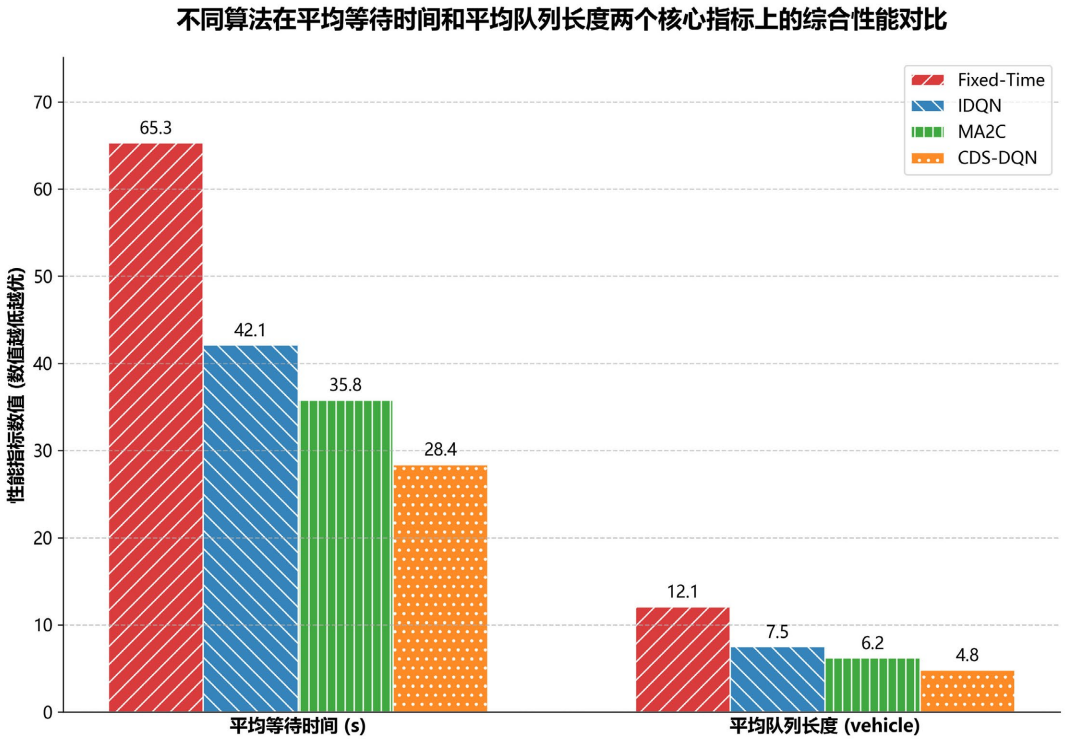


Figure 6. Comprehensive performance comparison of different algorithms on the two core metrics: average waiting time and average queue length
图 6. 不同算法在平均等待时间和平均队列长度两个核心指标上的综合性能对比

图 6 进一步从平均等待时间和平均队列长度两个核心指标上对各算法进行了综合比较。CDS-DQN 在这两项指标上均取得了最优结果，平均等待时间相比次优的 MA2C 算法降低了约 19.8%，相比 IDQN 降低了约 32.5%；平均队列长度相比 MA2C 降低了约 22.6%。这一结果有力地证明，通过融合微观驾驶行为与宏观协同感知，本文算法能够实现对交通流更精细、更高效地调控，其性能全面优于传统的固定配时、无协作的 IDQN 以及主流的 MA2C 算法。

5. 结论

针对现有智能交通信号控制方法中普遍存在的状态表征不充分和多智能体协作效率低下的问题，本文提出了一种融合驾驶风格感知与协同感知的多智能体深度强化学习算法(CDS-DQN)。该算法通过引入创新的“有效占有率”状态特征，使模型能够感知微观层面的驾驶行为异质性，从而更精准地评估交通状态。同时，通过设计一种轻量级的邻居状态共享机制，实现了智能体间的有效协作，缓解了环境非平稳性问题。在 SUMO 平台上进行的大规模仿真实验结果表明，本文提出的 CDS-DQN 算法在平均等待时间、队列长度等关键性能指标上均显著优于传统的固定配时方法、无协作的 IDQN 以及主流的 MA2C 算法，验证了其在提升城市交通网络运行效率方面的有效性和先进性。

未来的工作可以从以下几个方面展开：1) 研究更精细的驾驶风格模型，并探索其对交通流的深层影响；2) 将该框架扩展到更复杂的网络拓扑和更真实的交通需求模式中；3) 考虑将算法与车联网技术结合，利用车辆主动上报的信息，实现更高效地控制。

参考文献

[1] Bilbao-Ubillos, J. (2008) The Costs of Urban Congestion: Estimation of Welfare Losses Arising from Congestion on

Cross-Town Link Roads. *Transportation Research Part A*: *Policy and Practice*, **42**, 1098-1108. https://doi.org/10.1016/j.tra.2008.03.015

[2] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., *et al*. (2017) Mastering the Game of Go without Human Knowledge. *Nature*, **550**, 354-359. https://doi.org/10.1038/nature24270

[3] Sims, A.G. and Dobinson, K.W. (1980) The Sydney Coordinated Adaptive Traffic (SCAT) System Philosophy and Benefits. *IEEE Transactions on Vehicular Technology*, **29**, 130-137. https://doi.org/10.1109/T-VT.1980.23833

[4] Hunt, P.B., Robertson, D.I., Bretherton, R.D., *et al*. (1981) SCOOT—A Traffic Responsive Method of Coordinating Signals.

[5] Kaelbling, L.P., Littman, M.L. and Moore, A.W. (1996) Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, **4**, 237-285. https://doi.org/10.1613/jair.301

[6] Li, L., Lv, Y. and Wang, F. (2016) Traffic Signal Timing via Deep Reinforcement Learning. *IEEE*/*CAA Journal of Automatica Sinica*, **3**, 247-254. https://doi.org/10.1109/jas.2016.7508798

[7] Buşoniu, L., Babuška, R. and De Schutter, B. (2010) Multi-Agent Reinforcement Learning: An Overview. In: Srinivasan, D. and Jain, L.C., Eds., *Innovations in Multi-Agent Systems and Applications*—1, Springer, 183-221. https://doi.org/10.1007/978-3-642-14435-6_7

[8] Wei, H., Zheng, G., Gayah, V. and Li, Z. (2021) Recent Advances in Reinforcement Learning for Traffic Signal Control: A Survey of Models and Evaluation. *ACM SIGKDD Explorations Newsletter*, **22**, 12-18. https://doi.org/10.1145/3447556.3447565

[9] Rong, J., Mao, K. and Ma, J. (2011) Effects of Individual Differences on Driving Behavior and Traffic Flow Characteristics. *Transportation Research Record*: *Journal of the Transportation Research Board*, **2248**, 1-9.

[10] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., *et al*. (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[11] Van Hasselt, H., Guez, A. and Silver, D. (2016) Deep Reinforcement Learning with Double Q-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**, 2094-2100. https://doi.org/10.1609/aaai.v30i1.10295

[12] Wang, Z., Schaul, T., Hessel, M., *et al*. (2016) Dueling Network Architectures for Deep Reinforcement Learning. *International Conference on Machine Learning*. *PMLR*, New York, 20-22 June 2016, 1995-2003.

[13] Chu, T., Wang, J., Codecà, L. and Li, Z. (2020) Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 1086-1095. https://doi.org/10.1109/tits.2019.2901791

[14] Haddad, T.A., Hedjazi, D. and Aouag, S. (2022) A Deep Reinforcement Learning-Based Cooperative Approach for Multi-Intersection Traffic Signal Control. *Engineering Applications of Artificial Intelligence*, **114**, Article ID: 105019. https://doi.org/10.1016/j.engappai.2022.105019