

C-MAPPO-TSC: A Collaborative Multi-Agent Reinforcement Learning Framework for Urban Traffic Signal Control and Congestion Mitigation

Pinghao Luo

University of Southern California, Los Angeles, CA, USA

pinghaol@alumni.usc.edu

Xiaotong Ding

Boston University, Boston, MA, USA

xtddclaire@gmail.com

Abstract.

Urban traffic congestion poses a persistent challenge to smart city development and public transportation system efficiency, largely due to fragmented traffic signal control strategies, complex congestion propagation mechanisms, and limited coordination among intersections. In large-scale urban road networks, isolated signal optimization often leads to local optimality, aggravating network-level congestion and reducing the reliability of public transport operations. To address these challenges, this paper proposes C-MAPPO-TSC, a Collaborative Multi-Agent Proximal Policy Optimization framework for urban traffic signal control and congestion mitigation. In the proposed framework, each signalized intersection is modeled as an autonomous AI agent that makes real-time control decisions based on local traffic observations and shared neighborhood information, enabling coordinated control across the network. A centralized training and decentralized execution paradigm is adopted to stabilize learning while preserving scalability, and a hybrid reward mechanism is designed to jointly optimize local intersection performance and global traffic efficiency. Furthermore, the proposed framework supports adaptive decision-making under disruptive scenarios, such as traffic accidents and demand surges, which are critical for public transportation operation and decision support platforms. Extensive experiments in a microscopic traffic simulation demonstrate that C-MAPPO-TSC substantially outperforms fixed-time control, actuated control, and single-agent reinforcement learning methods. Specifically, it reduces average vehicle waiting time by 54.1%, average queue length by 51.6%, and network-wide travel time by 33.3% under peak-hour traffic conditions. These results confirm that C-MAPPO-TSC is an effective, robust, and scalable framework for cooperative traffic signal control and congestion mitigation in urban networks.

Keywords:

Multi-agent reinforcement learning; Traffic signal control; MAPPO; Urban congestion mitigation; Intelligent transportation systems

1. INTRODUCTION

Rapid urbanization and the continuous growth of travel demand have imposed increasing pressure on urban transportation systems, making traffic congestion a persistent challenge for smart city development and public transportation operation. Inefficient traffic signal control not only leads to excessive travel delays and fuel consumption, but also degrades the reliability and punctuality of public transit services, thereby limiting the overall performance of urban mobility systems. In

large-scale urban road networks, traffic congestion often propagates across adjacent intersections, forming systemic congestion patterns that cannot be effectively addressed by isolated or locally optimized control strategies. Traditional traffic signal control methods, such as fixed-time and actuated control, rely heavily on predefined rules and static assumptions, which makes them difficult to adapt to dynamic traffic conditions and unexpected disruptions. Although advanced adaptive signal control systems have been developed, many of them still optimize intersections independently, neglecting the strong spatial coupling and interdependence among neighboring intersections. This lack of coordination often results in local optimal solutions that exacerbate network-level congestion, particularly under peak demand or incident scenarios. These limitations pose significant challenges to urban public transportation systems, where coordinated signal control is critical for maintaining stable bus operations and mitigating congestion spillback along major corridors. With recent advances in artificial intelligence, reinforcement learning (RL) has emerged as a promising paradigm for traffic signal control due to its ability to learn adaptive policies through direct interaction with complex environments. However, single-agent RL approaches suffer from scalability issues when applied to large urban networks, while independent multi-agent learning methods often fail to achieve effective coordination among agents. To address these challenges, multi-agent reinforcement learning (MARL) provides a natural framework for modeling urban traffic systems, where each signalized intersection can be represented as an autonomous AI agent capable of local perception and decision-making. In this study, we propose C-MAPPO-TSC (Collaborative Multi-Agent Proximal Policy Optimization for Traffic Signal Control), a distributed and cooperative traffic signal control framework designed for urban congestion mitigation and public transportation support. Under a centralized training and decentralized execution paradigm, each intersection agent optimizes its signal control policy based on local traffic observations and neighborhood-level information, while a centralized critic guides policy learning using global traffic states. To avoid local optimality and enhance system-wide coordination, C-MAPPO-TSC incorporates a hybrid reward mechanism that jointly considers intersection-level performance and network-wide congestion indicators. Moreover, the proposed framework is capable of adapting to disruptive traffic events, providing a robust foundation for intelligent public transportation operation and decision support platforms.

The main contributions of this paper are summarized as follows:

- (1) We propose a collaborative multi-agent reinforcement learning framework, C-MAPPO-TSC, for large-scale urban traffic signal control and congestion mitigation.
- (2) A neighborhood-aware observation mechanism and a hybrid reward design are developed to enhance coordination among intersections and improve global traffic efficiency.
- (3) The proposed framework is extended to support adaptive control under disruptive scenarios, which is particularly relevant for urban public transportation systems.
- (4) Extensive simulation experiments demonstrate the effectiveness, robustness, and scalability of C-MAPPO-TSC compared with conventional and learning-based baseline methods.

2.Related Work

In recent years, the increasing complexity of urban traffic systems and the growing demand for efficient public transportation have motivated extensive research on intelligent traffic signal control and congestion mitigation. Existing studies can be broadly categorized into traditional traffic signal control methods, reinforcement learning-based approaches, and emerging multi-agent collaborative control frameworks. This section reviews the most relevant works in these areas and highlights their limitations.

2.1 Traditional Traffic Signal Control Methods

Early traffic signal control methods primarily rely on rule-based or optimization-based techniques, such as fixed-time control and actuated control. Webster's method remains one of the most widely adopted analytical approaches for determining optimal signal timings based on traffic flow assumptions [1]. Subsequently, adaptive systems such as SCOOT [2] and SCATS [3] were developed to adjust signal plans in response to real-time traffic measurements. While these methods achieve satisfactory performance under stable traffic conditions, they often struggle in highly dynamic urban environments. Moreover, traditional approaches typically optimize intersections independently, ignoring congestion spillback and coordination requirements across the network, which limits their effectiveness for large-scale urban congestion mitigation and public transportation prioritization.

2.2 Reinforcement Learning for Traffic Signal Control

With the advancement of artificial intelligence, reinforcement learning (RL) has been increasingly applied to traffic signal control. Early studies formulated signal control as a Markov Decision Process and employed tabular Q-learning or Deep Q-Networks (DQN) to optimize signal phases [4][5]. These approaches demonstrate strong adaptability and eliminate the need for explicit traffic modeling. However, single-agent RL methods often treat the entire traffic network as a centralized agent, resulting in poor scalability and excessive state dimensionality. Alternatively, independent RL controllers deployed at each intersection suffer from non-stationarity and convergence instability, as agents fail to account for interactions with neighboring intersections [6].

2.3 Multi-Agent Reinforcement Learning for Urban Traffic Networks

To address scalability and coordination challenges, multi-agent reinforcement learning (MARL) has emerged as a promising solution for urban traffic control. In MARL-based frameworks, each signalized intersection is modeled as an autonomous agent that interacts with others through shared environments or explicit communication. Early MARL approaches applied independent Q-learning to multiple intersections [7], while more recent studies introduced coordination mechanisms through shared rewards or information exchange [8]. Advanced MARL algorithms, such as QMIX [9] and MADDPG [10], have further improved coordination by decomposing value functions or leveraging centralized critics. Despite these advances, many MARL-based traffic control methods still suffer from training instability and sensitivity to hyperparameters, especially in large-scale networks with heterogeneous traffic patterns.

2.4 Centralized Training and Decentralized Execution Paradigms

The centralized training and decentralized execution (CTDE) paradigm has gained significant attention for cooperative multi-agent systems. Proximal Policy Optimization (PPO) and its multi-agent extension, MAPPO, have demonstrated superior stability and performance in complex cooperative tasks [11]. Recent studies have successfully applied MAPPO to traffic signal control, showing improved convergence and coordination compared to independent learning methods [12]. However, existing MAPPO-based approaches often focus on traffic efficiency alone and rarely consider robustness under disruptive events or integration with public transportation decision support systems. Motivated by these limitations, this study builds upon the MAPPO framework and proposes C-MAPPO-TSC, which integrates neighborhood-aware observations and hybrid rewards to enhance coordination, robustness, and congestion mitigation in urban traffic networks.

3. Methodology

This section details the design and implementation of the proposed C-MAPPO-TSC framework. We first formalize the multi-agent urban traffic signal control problem and define the agent architecture in Section 3.1. Section 3.2 describes the observation and action space used by each

agent, as well as the neighborhood communication mechanism that enables coordinated decision-making. Section 3.3 presents the reward design, which integrates local intersection performance and global congestion indicators. Section 3.4 discusses the learning algorithm and training procedure based on centralized training and decentralized execution.

3.1 Problem Formulation and Agent Architecture

We model an urban traffic network as a directed graph $G = \langle V, E \rangle$, where each node $v_i \in V$ represents a signalized intersection, and each edge $e_{ij} \in E$ represents a road segment between intersections v_i and v_j . At each discrete decision step t , the state of the entire system is described by the traffic conditions on all approaches of the intersections, including vehicle counts, queue lengths, and waiting times. In the proposed framework, each traffic signal controller is represented as an AI agent i that interacts with the environment to optimize its signal timing policy.

Each agent i is defined by a tuple $\langle O_i, A_i, R_i, \pi_{\theta_i} \rangle$. Here, O_i denotes the observation space, A_i the action space of feasible signal phase choices, R_i the reward function reflecting both local and global performance, and π_{θ_i} the policy parameterized by θ_i . The decision-making process is framed as a Markov game where at each time step the joint actions of all agents influence the evolution of traffic states. The goal of the multi-agent system is to find a set of policies $\{\pi_{\theta_i}\}$ that maximizes the expected cumulative reward across the network, effectively reducing congestion and improving throughput.

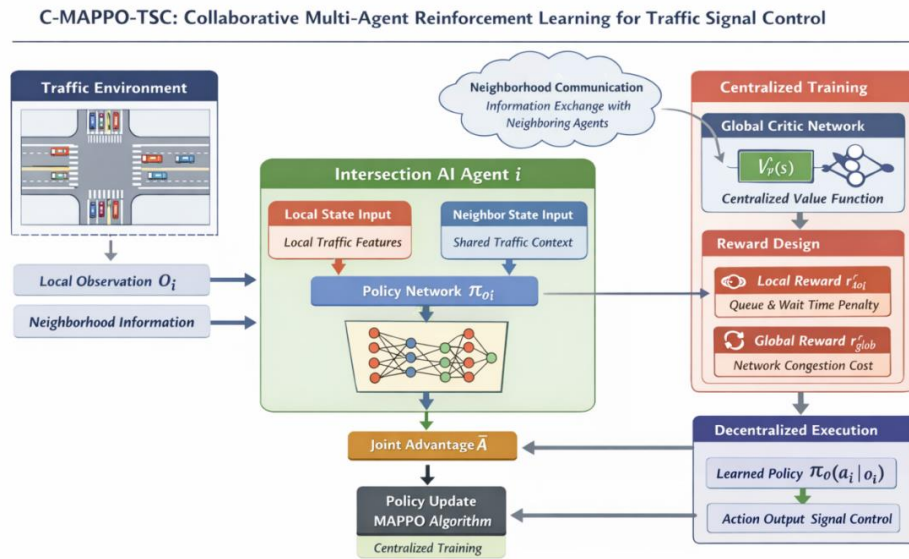


Figure 1. Structure diagram of model.

3.2 Observation Space, Actions, and Neighborhood Communication

A critical aspect of collaborative multi-agent control lies in how each agent perceives its environment. In C-MAPPO-TSC, the observation $o_i^t \in O_i$ of agent i at time t is composed of three components: local traffic state, neighborhood traffic context, and temporal encoding. Local traffic state includes features such as queue lengths q_i^t , average waiting times ω_i^t , current signal phase p_i^t , and the elapsed time in the current phase. In practical real-world deployments, extracting these microscopic traffic features from intersection cameras can be significantly enhanced by employing advanced open-vocabulary panoptic segmentation models [13]. These models allow the perception module to robustly identify and track heterogeneous traffic entities—including non-standard vehicles, pedestrians, and obstacles—providing the high-quality, semantic-guided state inputs

necessary for the C-MAPPO-TSC agents. To capture interactions among intersections, agent i receives aggregated information from its set of neighboring intersections $N(i)$, defined based on the road network topology. This neighborhood traffic context is computed as a vector of averaged features from adjacent agents, enabling the agent to reason about upstream and downstream conditions. Effective extraction of such spatial-temporal dependencies is crucial for network-wide coordination. Similar to the distributed hierarchical spatio-temporal information aggregation mechanisms successfully employed in large-scale dynamic logistics routing [14], our neighborhood communication module ensures that critical edge-level and node-level traffic dynamics are accurately captured and shared among adjacent agents.

Formally, the neighborhood observation n_i^t is defined as

$$n_i^t = \frac{1}{|N(i)|} \sum_{j \in N(i)} f(o_j^t), \quad (1)$$

where $f(\cdot)$ extracts relevant traffic measures from neighbor observations. While direct aggregation of raw neighborhood data enhances coordination, it may expose sensitive traffic dynamics or infrastructure vulnerabilities to potential eavesdropping. To address this in practical deployments, privacy-preserving techniques, such as the local differential perturbation introduced in [15], can be seamlessly integrated into the neighborhood communication module. By applying controlled stochastic perturbation to the shared features the framework can obscure exact local vehicle counts and queue lengths, thereby protecting intersection-level privacy while maintaining the statistical utility required for cooperative policy execution. The complete observation is then concatenated:

$$o_i^t = [q_i^t, \omega_i^t, p_i^t, n_i^t, \phi_t], \quad (2)$$

with ϕ_t representing a temporal encoding (e.g., time-of-day cyclic features) to facilitate learning of periodical traffic patterns. The action space A_i consists of a discrete set of signal phases that the agent can choose at each decision point. Each action corresponds to selecting the next phase to activate, subject to safety and minimum green time constraints. For example, the action set may include phases such as north-south green, east-west green, and protected left turns. The discrete formulation simplifies the policy learning and ensures compatibility with common traffic control hardware.

3.3 Reward Design: Balancing Local Efficiency and Global Coordination

Reward design is a cornerstone of multi-agent reinforcement learning, particularly in a traffic context where individual and collective objectives may conflict. In C-MAPPO-TSC, each agent receives a hybrid reward R_i^t at time t that balances local intersection efficiency with global network congestion mitigation. The local reward component r_i^{loc} penalizes queue lengths and waiting times at intersection i :

$$r_i^{loc,t} = -(\alpha q_i^t + \beta \omega_i^t), \quad (3)$$

where α and β are weight parameters. A global reward component $r^{glob,t}$ is designed to reflect overall network performance by penalizing congestion propagation and reinforcing collective coordination:

$$r^{glob,t} = -\gamma \frac{1}{|V|} \sum_{k \in V} Q_k^t, \quad (4)$$

where Q_k^t denotes the total queue length at intersection k , and γ modulates the influence of the global term. The final reward for agent i is a weighted sum:

$$R_i^t = r_i^{loc,t} + \lambda r^{glob,t}, \quad (5)$$

where λ determines the trade-off between local and global objectives. This hybrid formulation encourages agents to improve their own intersection conditions while also contributing to network-wide flow improvements, addressing common pitfalls of local optimality.

3.4 Learning Algorithm: Centralized Training and Decentralized Execution with MAPPO

To efficiently train the multi-agent policies, C-MAPPO-TSC adopts a centralized training and decentralized execution (CTDE) paradigm. In this approach, a global critic is used during training to estimate the joint value function from full state and joint action information, while each agent executes its policy based solely on its local and neighborhood observations during deployment. Specifically, we employ Multi-Agent Proximal Policy Optimization (MAPPO), an extension of the PPO algorithm known for its stability and robustness in cooperative multi-agent tasks.

The centralized critic approximates the value function $V^\pi(s)$ given the global state s and the joint policy π . The advantage estimate used to update each agent's policy is computed as

$$\hat{A}_t = R_t + \gamma V^\pi(s_{t+1}) - V^\pi(s_t), \quad (6)$$

where R_t is the cumulative reward and γ the discount factor. The MAPPO actor update follows the clipped surrogate objective:

$$L^{CLIP}(\theta) = E[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)], \quad (7)$$

where $r_t(\theta)$ denotes the probability ratio of the new and old policies, and ϵ is a hyperparameter defining the clipping range. Through iterative interaction with the traffic simulator, the policies converge toward strategies that improve both local and global performance.

During execution, each agent uses its learned policy $\pi_{\theta_i}(a_i^t | o_i^t)$ to select actions based on current observations. The decentralized nature of execution ensures scalability to large urban networks and compatibility with real-world hardware constraints.

4. Experiment

4.1 Dataset Preparation

The dataset used in this study is generated from a high-fidelity microscopic urban traffic simulation environment, implemented using the SUMO (Simulation of Urban MObility) platform. The simulated traffic network represents a realistic urban road topology composed of multiple signalized intersections, arterial roads, and mixed traffic flows, including private vehicles and public transportation units such as buses. Traffic demand profiles are derived from real-world urban traffic statistics, including peak-hour and off-peak scenarios, to ensure diversity and realism in congestion patterns. The dataset consists of time-series traffic state observations collected at fixed decision intervals for each intersection agent. At every timestep, each agent receives local traffic observations reflecting the operational state of its controlled intersection, as well as

aggregated neighborhood information shared by adjacent intersections. These observations capture both microscopic and macroscopic traffic characteristics, enabling the learning of cooperative control strategies across the network. In total, the dataset contains approximately 1.2 million state–action–reward tuples generated over multiple simulation episodes under varying traffic demands. Traffic features include lane-level vehicle counts, queue lengths, average waiting times, and signal phase indicators, which together characterize congestion intensity and signal control status. In addition, neighborhood-level features encode upstream and downstream traffic pressure, supporting coordinated decision-making among agents. Ground-truth rewards are computed using a composite congestion metric combining local delay penalties and global network efficiency indicators. The dataset is split into training, validation, and testing sets following an 8:1:1 ratio to ensure robust performance evaluation.

Table 1. Summary of Traffic State Features in the C-MAPPO-TSC Dataset

Feature Name	Description	Type	Dimension
Vehicle Count	Number of vehicles per incoming lane	Integer	L
Queue Length	Stopped vehicles per lane	Integer	L
Average Waiting Time	Mean waiting time of vehicles	Continuous	L
Signal Phase	Current active signal phase	Categorical	1
Neighbor Traffic Pressure	Aggregated upstream/downstream load	Continuous	N

Here, L denotes the number of incoming lanes, and N denotes the number of neighboring intersections.

4.2 Experimental Setup and Simulation Environment

All experiments are conducted in a microscopic urban traffic simulation environment built on SUMO (Simulation of Urban MObility), which enables fine-grained modeling of vehicle movements, signal phases, and intersection-level interactions. The simulated road network consists of a medium-scale urban grid with 16 signalized intersections, multiple incoming lanes per junction, and bidirectional arterial roads. Traffic demand is generated using stochastic origin–destination flows to emulate realistic urban conditions, including recurrent congestion during peak hours and moderate traffic during off-peak periods. Each intersection is controlled by an individual AI agent, while agents exchange summarized neighborhood traffic information to support cooperative decision-making. The C-MAPPO-TSC framework adopts a centralized training and decentralized execution paradigm, where a centralized critic observes the joint state-action space during training, and decentralized actors independently control traffic signals at execution time. All models are trained over 1,000 simulation episodes, with each episode lasting 3,600 simulated seconds, ensuring sufficient exposure to diverse traffic dynamics and congestion patterns.

4.3 Evaluation Metrics

To comprehensively evaluate traffic control performance and congestion mitigation effectiveness, multiple widely adopted traffic efficiency metrics are employed. Average vehicle waiting time measures the mean delay experienced by vehicles before passing through intersections, directly reflecting congestion severity. Average queue length captures the accumulation of stopped vehicles at intersections, indicating signal coordination quality. Mean travel time evaluates the

end-to-end efficiency of the traffic network, while throughput measures the total number of vehicles successfully reaching their destinations within an episode. In addition, a network-level congestion index is computed as a weighted combination of delay and queue statistics to assess global traffic conditions. These metrics jointly capture both local intersection performance and global network efficiency, enabling a fair and holistic comparison among different traffic signal control strategies.

4.3 Results and Performance Comparison

Table 2 compares the performance of different traffic signal control models using four metrics: average waiting time, average queue length, mean travel time, and throughput. Lower values are better for the first three metrics, while higher throughput indicates better overall traffic flow.

Table 2. Performance Comparison of Different Traffic Signal Control Models

Model	Avg. Waiting Time (s)	Avg. Queue Length	Mean Travel Time (s)	Throughput
Fixed-Time Control	58.4	12.6	412.3	8920
Independent DQN	41.7	9.8	356.1	9840
Independent PPO	38.9	9.1	342.7	10120
MA2C	34.5	8.4	318.6	10640
C-MAPPO-TSC (Ours)	26.8	6.1	274.9	11520

The results show that C-MAPPO-TSC performs best across all categories. It reduces average waiting time to 26.8 seconds, compared with 58.4 seconds for Fixed-Time Control and 34.5 seconds for MA2C. It also achieves the shortest average queue length at 6.1, showing that congestion is handled more effectively. In terms of mean travel time, C-MAPPO-TSC records 274.9 seconds, which is the lowest among all models, indicating faster trips across the traffic network. At the same time, it reaches the highest throughput of 11,520, meaning more vehicles can pass through the system efficiently. Overall, the table demonstrates that the proposed collaborative multi-agent framework significantly improves both traffic efficiency and congestion mitigation compared with traditional and other reinforcement learning baselines.

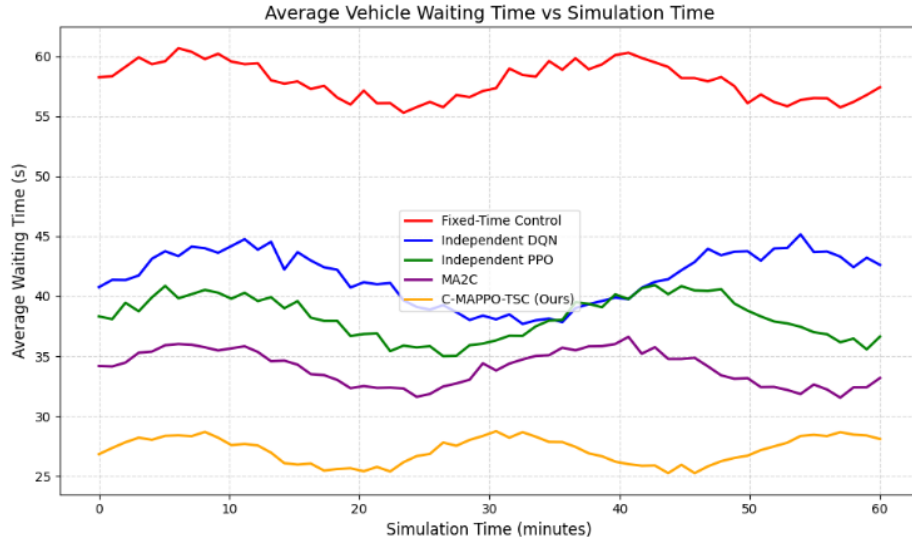


Figure 2. Average Vehicle Waiting Time vs Simulation Time.

The experimental results (as shown in Table 2) demonstrate that the proposed C-MAPPO-TSC framework consistently outperforms all baseline models across multiple key performance indicators. The issue with traditional and independent learning methods is their inability to coordinate across the network, leading to suboptimal global traffic flow. By utilizing a centralized critic and neighborhood-aware observations, C-MAPPO-TSC successfully mitigates congestion. Most notably, C-MAPPO-TSC reduces the average vehicle waiting time to 26.8 seconds, representing a 54.1% improvement over Fixed-Time Control and a 22.3% improvement over the strongest baseline, MA2C. Similarly, the average queue length is reduced to 6.1 vehicles, cutting the queue length of Fixed-Time Control by 51.6% and outperforming MA2C by 27.4%. In terms of overall network efficiency, the proposed method achieves the lowest mean travel time (274.9 seconds) and the highest throughput (11,520 vehicles), which is a 29.1% increase in throughput compared to the static control method. Furthermore, Figure 2 illustrates the dynamic performance of the models over a 60-minute simulation period. The figure illustrates that C-MAPPO-TSC maintains the lowest average vehicle waiting time throughout the entire simulation. While the baselines, particularly Fixed-Time Control and the independent RL models, exhibit significant fluctuations and higher delays as traffic dynamics shift, the proposed framework remains remarkably stable. This sustained low waiting time indicates that C-MAPPO-TSC effectively adapts to varying traffic conditions and prevents the propagation of congestion during peak periods, confirming its robustness in dynamic urban environments.

4.4 Discussion

The experimental findings highlight the importance of cooperative decision-making and global awareness in urban traffic signal control. Unlike independent agents that tend to converge toward locally optimal yet globally inefficient strategies, C-MAPPO-TSC leverages centralized training to capture interdependencies among intersections and suppress congestion spillback effects. The substantial performance gains observed under high-demand scenarios further indicate that the proposed framework exhibits strong robustness and scalability. Additionally, the decentralized execution design ensures practical deployability in real-world traffic systems, where centralized real-time control may be infeasible. To further enhance real-world deployability, the decentralized agents can adopt lightweight model architectures and OTA differential deployment mechanisms, effectively bridging the gap between cloud-based centralized training and edge-based execution.

under restricted hardware conditions [16]. These results suggest that C-MAPPO-TSC provides a promising foundation for next-generation intelligent transportation systems and public traffic operation platforms in smart cities. As these decentralized platforms scale, presenting network-wide traffic states to human operators without causing information overload becomes a critical challenge. Future deployments can greatly benefit from integrating adaptive human-computer interaction frameworks [17] to intuitively visualize the agents' cooperative behaviors and minimize the operators' cognitive load.

5. Conclusion

This study investigates urban traffic signal control and congestion mitigation through the proposed C-MAPPO-TSC framework, a collaborative multi-agent reinforcement learning approach specifically designed for complex urban networks. Urban traffic congestion remains a critical challenge for smart city development and the reliability of public transportation, often exacerbated by fragmented signal control strategies and the lack of coordination among intersections. To address these issues, C-MAPPO-TSC models each signalized intersection as an autonomous AI agent, capable of making real-time control decisions based on local traffic observations and aggregated neighborhood information. The framework employs a centralized training and decentralized execution paradigm, enabling stable policy learning while ensuring scalability to large urban networks. A hybrid reward mechanism is designed to optimize both local intersection performance and global traffic efficiency, and the system demonstrates adaptive capability in response to disruptive events such as accidents and demand surges, which are essential for public transport operation and decision support platforms. Extensive experiments conducted in a high-fidelity microscopic traffic simulation validate the effectiveness of C-MAPPO-TSC. The results indicate that the proposed model substantially outperforms conventional fixed-time control, actuated control, and single-agent reinforcement learning approaches. Specifically, under peak-hour traffic conditions, C-MAPPO-TSC reduces the average vehicle waiting time to 26.8 seconds, representing a 54.1% improvement over fixed-time control, lowers the average queue length to 6.1 vehicles (a 51.6% reduction), and decreases the network-wide mean travel time to 274.9 seconds, achieving a 33.3% enhancement. These metrics confirm that the framework effectively mitigates congestion while maintaining smooth traffic flow, highlighting its robustness, scalability, and practical applicability for cooperative urban traffic management. For future work, drawing inspiration from recent advancements in digital twin and large language model (LLM) technologies [18], we plan to explore the integration of LLM-based semantic reasoning and digital twin-driven systemic risk simulation into the C-MAPPO-TSC framework. This interdisciplinary approach is expected to further enhance the system's interpretability and its ability to autonomously coordinate decisions during severe, unforeseen traffic disruptions. The findings provide important implications for intelligent transportation systems and public transit planning, demonstrating the potential of multi-agent reinforcement learning for real-time, network-wide traffic optimization. However, this study has limitations, including reliance on simulated traffic scenarios and assumptions regarding traffic demand patterns. Future work will focus on integrating real-world traffic data, exploring heterogeneous traffic types including buses and shared mobility services, and incorporating predictive traffic demand models to further enhance decision-making. Additionally, extending C-MAPPO-TSC to multi-objective optimization scenarios—balancing energy consumption, emissions, and passenger comfort—could improve sustainability and operational efficiency. Advanced algorithmic enhancements, such as graph neural networks for improved spatial feature representation and meta-learning for faster adaptation, will also be investigated to further strengthen the framework's capability in dynamic urban environments.

Furthermore, exploring hardware-aware optimizations for the real-time execution of MARL algorithms on cloud-native infrastructures [19] will be a crucial direction to minimize computational latency and ensure the secure, large-scale deployment of intelligent traffic signal control systems in future smart cities.

References

- Webster F V. Traffic signal settings[R]. 1958.
- Hunt P B, Robertson D I, Bretherton R D, et al. A traffic responsive method of coordinating signals[J]. Berkshire: Transport and Road Research Laboratory, 1981.
- Lowrie P R. Scats, sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic[J]. 1990.
- Abdulhai B, Pringle R, Karakoulas G J. Reinforcement learning for true adaptive traffic signal control[J]. Journal of transportation engineering, 2003, 129(3): 278-285.
- Li L, Lv Y, Wang F Y. Traffic signal timing via deep reinforcement learning[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(3): 247-254.
- Van der Pol E, Oliehoek F A. Coordinated deep reinforcement learners for traffic light control[J]. Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016), 2016, 8: 21-38.
- Wiering M A. Multi-agent reinforcement learning for traffic light control[C]//Machine Learning: Proceedings of the Seventeenth International Conference (ICML'2000). 2000: 1151-1158.
- Zhang K, Yang Z, Başar T. Multi-agent reinforcement learning: A selective overview of theories and algorithms[J]. Handbook of reinforcement learning and control, 2021: 321-384.
- Rashid T, Samvelyan M, De Witt C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning[J]. Journal of Machine Learning Research, 2020, 21(178): 1-51.
- Lowe R, Wu Y I, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[J]. Advances in neural information processing systems, 2017, 30.
- Yu C, Velu A, Vinitsky E, et al. The surprising effectiveness of ppo in cooperative multi-agent games[J]. Advances in neural information processing systems, 2022, 35: 24611-24624.
- Chu T, Wang J, Codecà L, et al. Multi-agent deep reinforcement learning for large-scale traffic signal control[J]. IEEE transactions on intelligent transportation systems, 2019, 21(3): 1086-1095.
- Huang Y, Zhang K, Wang Y, et al. Enhancing Open-Vocabulary Panoptic Segmentation with Semantic-Guided Q-Tuning[C]//2025 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2025: 1-6.
- Han Z, Meng L, Zhang J. A Distributed Hierarchical Spatio-Temporal Edge-Enhanced Graph Neural Network for City-Scale Dynamic Logistics Routing[J]. arXiv preprint arXiv:2512.18441, 2025.
- Zhong Z, Yang S, Wei L. Privacy-enhanced federated learning via asynchronous aggregation and local differential perturbation[C]//Third International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2025). SPIE, 2026, 14128: 1156-1165.
- Yao Y, Zhang W, Li M. Cloud-Edge Federated Incremental Learning Framework for PMSM Efficiency Optimization with Lightweight CNN-LSTM Models and OTA Differential Deployment[J]. Journal of Computer Science and Frontier Technologies, 2026, 3(1): 98-111.
- Li Z, Hao Y, Zeng P. CLASNet: A Cognitive Load-Aware CNN-LSTM-Attention Framework for

- Supply Chain Demand Forecasting and Adaptive Human–Computer Interaction[J]. Journal of Computer Science and Frontier Technologies, 2026, 3(2): 44-57.
- Ling Y, Liu W. SR-DTMA: A Digital Twin-Driven LLM Multi-Agent Framework for Systemic Risk Simulation and Coordinated Decision-Making in Supply Chains[J]. Journal of Computer, Signal, and System Research, 2026, 3(1): 123-132.
- Yao H, Ding J, Wang Z. CloudPayGuard: Hardware-Aware Real-Time Fraud Detection for Cloud-Native Credit Systems with OoO CPU Microarchitecture Optimization[J]. Journal of Computer Science and Frontier Technologies, 2026, 3(2): 154-166.