Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 20, No. 1, 2026

Efficient State Representation for Multi-Agent Reinforcement Learning in Traffic Signal Control

Mengze Wang*

School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou, China
* **Corresponding author Email:** wangmz0901@163.com

Abstract: With the continuous advancement of urbanization, traffic congestion has emerged as a critical bottleneck limiting the efficiency of urban transportation systems. Conventional traffic signal control strategies, which rely on fixed-time schemes or rule-based adaptive methods, struggle to cope with the highly dynamic and stochastic nature of real-world traffic conditions. In recent years, reinforcement learning (RL) has gained increasing attention in the field of traffic signal control (TSC) due to its ability to autonomously optimize decision-making in dynamic environments. As the foundation of agent decision-making, the representation of environmental states plays a decisive role in control performance. However, most existing studies construct traffic states using only a limited set of representative features, such as queue lengths and signal phase information, which are insufficient to comprehensively capture the complex spatiotemporal dynamics of traffic flows, thereby constraining the learning capability of agents in complex environments. To address these limitations, this paper proposes an Efficient State Representation for Multi-Agent Reinforcement Learning (ESR-MARL) framework. The proposed method incorporates richer traffic information for fine-grained modeling and employs a channel-wise attention mechanism to independently learn and effectively fuse heterogeneous traffic features, enabling the extraction of a more comprehensive and informative traffic state representation. Extensive experiments conducted on both synthetic and real-world traffic datasets demonstrate that ESR-MARL achieves at least a 27.37% improvement in average travel time compared with state-of-the-art baseline methods, thereby validating the effectiveness and superiority of the proposed approach.

Keywords: Intelligent transportation system; Traffic signal control; Multi-agent reinforcement learning; Graph neural network.

1. Introduction

With the accelerating pace of urbanization, traffic congestion has become one of the major challenges constraining urban transportation efficiency and deteriorating travelers' mobility experience [1]. In urban road networks, signalized intersections are widely recognized as the most common bottlenecks, and their control strategies play a pivotal role in determining the overall operational efficiency of the network. Conventional traffic signal control (TSC) approaches predominantly rely on fixed-time, actuated, or rule-based adaptive control mechanisms. Among them, the pre-timed signal control scheme proposed by Webster remains extensively deployed in traffic systems worldwide [2]. However, these methods generally assume relatively stable and predictable traffic flows, making it difficult for them to maintain satisfactory performance in complex and highly dynamic urban traffic environments. When multiple intersections interact and vehicle movements exhibit strong stochasticity, traditional approaches show evident limitations in accurately capturing real-time traffic states.

To address these challenges, reinforcement learning (RL) and its deep variants, i.e., deep reinforcement learning (DRL), have attracted increasing attention in the field of intelligent traffic control due to their capability of adaptive decision-making and automatic feature representation learning [3]. RL-based traffic signal control methods enable agents to iteratively optimize control policies through continuous interaction with the environment, gradually approaching optimal signal timing strategies [4]. The learning performance of such agents is highly dependent on the accuracy and expressiveness of the state definition [5]. Consequently, a growing body of research has focused on traffic state representation. From the perspective of representation form, the most widely adopted approaches can be broadly categorized into discrete traffic state encoding (DTSE) [6] and feature-vector-based representations [7].

Discrete Traffic State Encoding (DTSE). As one of the most commonly used state representation paradigms, DTSE maps continuous or high-dimensional traffic information into a finite set of discrete states to reflect real-world traffic conditions. Existing studies typically encode one or two types of vehicle-level information, such as position, speed, or acceleration, as the state representation [8]. In [9], graph-structured information is further incorporated after encoding to enhance the expressiveness of the traffic state.

Feature-vector-based state representation. Unlike DTSE, feature-vector-based representations do not explicitly describe individual vehicles on lanes. Instead, they aggregate specific traffic attributes across all lanes and represent them as a compact vector using average or cumulative statistics [5]. Commonly used features include the number of vehicles, queue length, average speed, waiting time, and signal phase information [10]. Considerable efforts have been devoted to improving such representations. For instance, Efficient Pressure (EP) [11] reformulates the Max-Pressure concept into a state representation. Advanced Traffic State (ATS) [12] extends EP by incorporating phase demand (PD) to better characterize traffic conditions. Cooperative Max-Pressure (CMP) [13] further propagates intersection pressure to neighboring intersections to obtain more accurate environmental information. Dynamic Cutting Length [14] combines vehicle count and queue length to maintain a compact state while adaptively emphasizing congested lanes. Maximum/Minimum Green Phase Indicators [15] represent phase duration using binary vectors to describe the utilization

of traffic resources. Queue Dynamic State Encoding (QDSE) [16] integrates key lane features derived from queue dynamic models, enhancing the agent's capability to analyze, predict, and respond to imminent congestion.

Despite these advances, regardless of the specific representation paradigm, most existing studies rely on only a limited number of representative traffic features, with only a few considering additional factors such as pedestrians or emergency vehicles [15]. Such sparse traffic information fails to provide a holistic view of dynamic traffic conditions, making it difficult for agents to learn effective control policies in rapidly changing environments. To overcome this limitation, we propose an Efficient State Representation for Multi-Agent Reinforcement Learning (ESR-MARL) framework, which leverages richer traffic information for fine-grained modeling. Specifically, ESR-MARL independently learns different types of traffic features and employs a channel-wise attention mechanism to effectively fuse heterogeneous information, thereby constructing a more comprehensive and informative traffic state representation.

The main contributions of this paper are summarized as follows:

1)We propose an efficient state representation framework, termed ESR-MARL, which incorporates richer traffic information for fine-grained modeling and applies a channel-wise attention mechanism to independently learn and effectively fuse heterogeneous traffic features, resulting in a more comprehensive traffic state representation for multi-agent reinforcement learning.

2)Extensive experiments on both synthetic and real-world urban traffic datasets demonstrate that the proposed ESR-MARL method achieves at least a 27.37% reduction in average travel time compared with state-of-the-art baseline approaches, validating its effectiveness and superiority.

2. Preliminaries

2.1. Traffic Signal Control Modeling

In this study, the multi-intersection traffic signal control problem is formulated as a Markov Decision Process (MDP), defined by a tuple $(\mathcal{S},\mathcal{O},\mathcal{A},\mathcal{P},r,\pi,\gamma)$, whose components are described as follows:

State space $\mathcal{S}$ and observation space $\mathcal{O}$: At each time step t, every intersection can observe its local state $o_i^t \in \mathcal{O}$ from the global system state $s_i^t \in \mathcal{S}$.

Action space $\mathcal{A}$: Each agent corresponds to a signalized intersection, and its action $a_i^t \in \mathcal{A}$ represents the currently selected signal phase. A typical four-phase signal control scheme is adopted, where the RL agent selects only one phase at each decision step. The minimum execution duration of each action is set to τ (i.e., the green signal duration), and a 5-second all-red interval is enforced between consecutive phases to ensure safe vehicle clearance at intersections.

State transition probability $\mathcal{P}$: $\mathcal{P}(s_{t+1} \mid s_t, a_t)$ denotes the probability that the environment transitions to the next state s_{t+1} after agent i executes action a_i^t in state s_t.

Reward function r: The reward r_i^t measures the feedback received by agent i from the environment after taking action a_i^t. In this work, the sum of queue lengths of all incoming lanes at an intersection is adopted as the reward signal to reflect the level of traffic congestion. The reward for agent i at time step t is computed as:

$$r_i^t = -\mathcal{Q}_i^t = -\sum_{nl} q^t. \tag{1}$$

Policy and discount factor (π,γ): Each agent learns an optimal policy π to maximize its expected cumulative discounted return:

$$G_i^t = \sum_{t=\tau}^{T} \gamma^{t-\tau} r_i^t, \tag{2}$$

where the discount factor $\gamma \in [0,1]$ balances short-term and long-term rewards.

2.2. State Definition

To describe traffic conditions more accurately, multiple categories of traffic information are jointly utilized to define the intersection state. Figure 1 illustrates the schematic diagram of a signalized intersection.

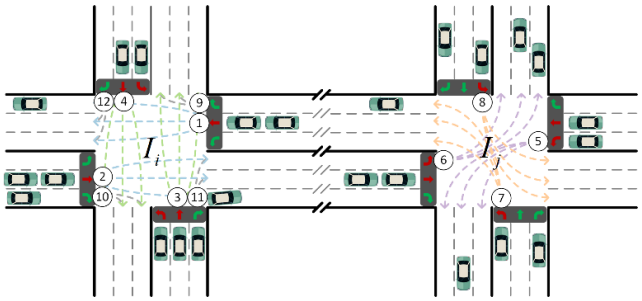


Fig. 1 Intersection diagram

2.2.1. Signal Phase

The signal phase [5], representing a set of movement permissions for vehicles, serves as an intermediate variable linking traffic conditions and control actions. In this study, the current signal phase is encoded as part of the state using one-hot encoding. Let the phase vector be denoted as a one-hot representation over the phase set, as summarized in Table 1.

2.2.2. Lane Vehicle Count

The lane vehicle count [7] refers to the number of vehicles on each incoming lane at an intersection and constitutes the most fundamental descriptor of traffic conditions. It is defined as:

$$\mathcal{NV} = [nv_1, nv_2, \cdots, nv_{nl}], \tag{3}$$

where $nv \in [0, nv_{\max}]$, $nv_{\max}$ denotes the maximum vehicle capacity of lane, and nl represents the total number of lanes at the intersection.

Table 1. Set of signal phase

Signal phase	One-hot code							
North-South straight	0	0	0	0	0	1	0	1
North-South left	0	0	0	0	1	0	1	0
East-West straight	0	1	0	1	0	0	0	0
East-West left	1	0	1	0	0	0	0	0

2.2.3. Lane Vehicle Density

Lane vehicle density [15] refers to the vehicle density on each incoming lane at an intersection and directly reflects the occupancy level of the intersection. It is defined as:

$$\mathcal{DV} = [\rho_1, \rho_2, \cdots, \rho_{nl}], \tag{4}$$

where $\rho \in [0,1]$ is calculated as:

$$\rho = \frac{nv \cdot \left(L_v + Gap_{\min}\right)}{L_l}, \tag{5}$$

in which L_v denotes the vehicle length, L_l denotes the length of lane l, $Gap_{\min}$ denotes the minimum spacing between vehicles.

2.2.4. Average System Delay

Average system delay [17] refers to the average ratio between the difference of the actual vehicle speed and the maximum allowable speed on the lane and the maximum allowable speed, which reflects the overall operational efficiency of the intersection. It is defined as:

$$\mathcal{SD} = [f_1, f_2, \cdots, f_{nl}], \tag{6}$$

Where $f \in [0,1]$ is calculated as:

$$f = 1 - \frac{\sum_i^{nv} V_i}{nv \cdot V_{\max}}, \tag{7}$$

in which $V_i \in [0, V_{\max}]$ denotes the actual vehicle speed on lane l, and $V_{\max}$ denotes the maximum allowable vehicle speed on lane l.

2.2.5. Lane Queue Length

Lane queue length [18] refers to the queue length of waiting vehicles on each incoming lane at an intersection. Vehicles with speeds lower than 0.1 m/s are regarded as waiting vehicles. The lane queue length is defined as:

$$\mathcal{Q} = [q_1, q_2, \cdots, q_{nl}]. \tag{8}$$

2.2.6. Efficient Pressure

Efficient pressure [11] is defined as the difference between the queue lengths of incoming and outgoing lanes at an intersection and is used to measure the potential efficiency improvement brought by releasing a specific signal phase. It is defined as:

$$EP = [ep_1, ep_2, \cdots, ep_{nl}], \tag{9}$$

where ep is calculated as:

$$ep = \sum_{i \in \mathcal{E}} q_i - \sum_{i \in \mathcal{X}} q_i, \tag{10}$$

in which $\mathcal{E}$ and $\mathcal{X}$ denote the sets of incoming and outgoing lanes, respectively.

2.2.7. Phase Demand

Phase demand [12] refers to the number of vehicles on each incoming lane that are able to leave the lane within one unit of time t, which is used to quantify the demand of the current signal phase. It is defined as:

$$PD = [pd_1, pd_2, \cdots, pd_{nl}]. \tag{11}$$

2.2.8. Maximum/Minimum Phase Indicators

The minimum/maximum green phase indicators [15] are incorporated into the state representation to balance the trade-off between excessively short green phases, which may aggravate congestion and increase accident risk during peak periods, and excessively long green phases, which reduce resource utilization efficiency. The indicator δ is represented as a binary vector. Specifically, $\delta = 1$ if the phase duration is shorter than the minimum green time plus yellow time ; otherwise, $\delta = 0$. Similarly, $\delta = 1$ if the phase duration exceeds the maximum green time plus yellow time; otherwise, $\delta = 0$.

2.3. Optimization Objective

An urban road network consists of multiple road segments and signalized intersections, where traffic signals regulate vehicle movements. Based on real-time traffic states, the control algorithm selects appropriate signal phases for each intersection. In this study, every phase transition is determined by the control algorithm, with the objective of minimizing the Average Travel Time (ATT) over the entire road network. ATT represents the average duration required for vehicles to travel from their origins to destinations and is a key indicator of commuting efficiency. It is calculated as:

$$\mathrm{ATT} = \frac{1}{NV} \sum_{i=1}^{NV} t_{i,out} - t_{i,in}, \tag{12}$$

where $t_{i,in}$ and $t_{i,out}$ denote the entry and exit times of vehicle i in the road network, respectively.

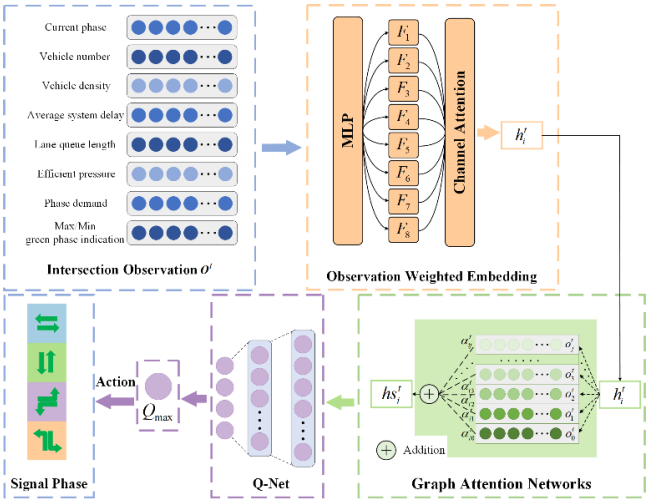


Fig. 2 Model Framework

3. Method

The proposed algorithm consists of three main components: traffic information weighted embedding, multi-intersection cooperative learning, and action-value estimation, as illustrated in Fig. 2. First, each agent obtains the observed state information of intersections from the environment and independently learns different types of traffic features through a channel attention mechanism, which enables effective feature fusion and yields a more comprehensive traffic state representation. Then, a graph neural network (GNN) is employed to capture the spatial correlations of the road network, facilitating information interaction and cooperative learning among multiple intersections. Finally, an action-value estimation module evaluates the potential benefits of different signal phases to generate the optimal traffic signal control policy. The details of each module are described as follows.

3.1. Traffic Information Weighted Embedding

At time step t, the raw observed state of intersection i is first fed into a multi-layer perceptron (MLP) with ReLU activation functions for feature extraction, generating the initial embedding representation of the node:

$$f_i^t = \mathrm{ReLU}\left(o_i^t W_e + b_e\right), \tag{13}$$

where f_i^t is used to describe the traffic state features of intersection i, o_i^t denotes the observation vector of intersection i, and W and b represent the learnable weight matrix and bias term, respectively.

Next, each type of traffic feature is treated as an independent channel. Average pooling and max pooling operations are applied to aggregate the state information, producing two different context descriptors, denoted as F_{avg}^c and F_{max}^c, which correspond to the average-pooled and max-pooled features, respectively. These two descriptors are then forwarded into a shared network composed of multi-layer perceptrons to generate the channel attention [19], as formulated as:

$$M_c = \sigma\left(\begin{array}{l}\mathrm{MLP}\left(\mathrm{AvgPool}\left(f_i^t\right)\right)\\ \quad +\mathrm{MLP}\left(\mathrm{MaxPool}\left(f_i^t\right)\right)\end{array}\right) \tag{14}$$
$$= \sigma\left(\mathrm{ReLU}\left(F_{avg}^c W_0\right)W_1 + \mathrm{ReLU}\left(F_{max}^c W_0\right)W_1\right),$$

where c is the number of channels, and the ReLU activation is applied after the first fully connected layer. Subsequently, the traffic state features are weighted and fused using the learned channel attention:

$$h_i^t = \mathrm{channelavg}\left(M_c \odot f_i^t\right), \tag{15}$$

where h_i^t denotes the weighted traffic state features of intersection i, and $\mathrm{channelavg}$ represents the summation and averaging operation along the channel dimension.

3.2. Graph Attention Mechanism

In this study, intersections are modeled as nodes in a graph, and road segments are modeled as edges. Since intersections are connected by bidirectional lanes, each edge can be represented as a pair of directed connections. Accordingly, a traffic network graph $G=(V,E)$ is constructed, where V denotes the set of nodes and E denotes the set of directed edges. The adjacency matrix is defined as $A \in \mathbb{R}^{N\times N}$, where N is the number of intersections. The neighborhood set of intersection i, including itself, is denoted as N_i. Each node contains features such as vehicle counts and signal phases. The global state space can be expressed as S, with its feature representation denoted as H.

Graph attention can dynamically learn the spatial dependency relationships among nodes [7]. Its core idea is to compute the importance weights between nodes through a trainable weight matrix W. The attention coefficient from node j to node i is defined as:

$$e_{ij} = \left(h_i^t W_q\right)\cdot\left(h_j^t W_k\right)^T. \tag{16}$$

To reflect the relative importance of neighboring nodes to the current node, the attention coefficients are normalized using the Softmax function:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in N_i}\exp(e_{ik})}. \tag{17}$$

Finally, the aggregated representation of intersection i is obtained by the weighted summation of the features of its neighboring nodes:

$$hs_i^t = \mathrm{ReLU}\left(\sum_{j\in N_i}\alpha_{ij}h_i^t W_v\right), \tag{18}$$

where W_v is a learnable weight matrix, and hs_i^t denotes the spatial representation after fusing neighborhood information. Through this mechanism, the model can adaptively capture the interactions among different intersections, yielding more representative spatial features to support subsequent traffic signal control decisions.

3.3. Action-Value Estimation

In this module, fully connected layers are used to learn the fused spatiotemporal hidden states and to predict the action values (Q-values) of each intersection under different signal phases. The optimal signal action is selected according to the maximum Q-value. During training, an ε-greedy policy is adopted to balance exploration and exploitation, enabling the agent to gradually approach the optimal policy:

$$Q_i^t = hs_i^t W_p + b_p, \tag{19}$$

where Q_i^t denotes the action-value vector of intersection i at time t for all signal phases, p represents the dimension of the action space, i.e., the number of selectable signal phases, and W_p and b_p are the learnable weight matrix and bias term, respectively. During decision-making, a random action is selected with probability ε, while the action with the maximum Q-value is chosen with probability $1-\varepsilon$.

During the training phase, the policy parameters are updated by minimizing the loss function, with the optimization objective defined as:

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i}^{N}\left(Q\left(o_i^t,a_i^t\right)-\widetilde{Q}\left(o_i^t,a_i^t,\theta\right)\right)^2, \tag{20}$$

where T denotes the total number of training time steps, and θ represents the set of all trainable parameters.

4. Experiments

4.1. Experimental Settings

All experiments are conducted on the open-source traffic simulation platform CityFlow [20], which has been widely adopted for evaluating reinforcement learning–based traffic signal control methods. The main hyperparameter settings used during model training are summarized in Table 2. To ensure a fair comparison, all methods are evaluated under identical experimental configurations.

4.2. Datasets

Four traffic datasets are employed in the experiments, including two synthetic datasets and two real-world datasets. The real-world traffic data are collected from Jinan and Hangzhou. Detailed information about all datasets is provided in Table 3.

Synthetic datasets: Both synthetic datasets are constructed on grid road networks with a size of 4×4. Each intersection follows a four-leg configuration, with three lanes per direction.

The lane length and width are set to 300 m and 3 m, respectively. Vehicles enter and leave the network randomly from boundary roads. Based on statistical analysis of real-world traffic data, the turning ratios at intersections are set to 10% for left turns, 60% for straight movements, and 30% for right turns, making the synthetic traffic flows more consistent with realistic distributions.

Jinan dataset: The road network consists of 3×4 intersections, each with a four-leg structure. The network includes two east–west lanes of 400 m and two north–south lanes of 800 m at each intersection.

Hangzhou dataset: The road network scale is 4×4, and each intersection also follows a four-leg configuration. The east–west roads are 800 m long, while the north–south roads are 600 m long.

Table 2. Experimental parameters

Hyperparameter	Value
Discount Factor γ	0.8
Learning rate	0.001
Phase duration Δt	10s
Number of training episodes	200
Simulation length	3600s
Training steps per episode	100
Replay buffer size	10000
Sample size	1000
Batch size	20
Target network update interval C	5 episodes
$\varepsilon_{\max}$	0.8
$\varepsilon_{\min}$	0.2
decay factor λ	0.95

Table 3. Statistics of traffic dataset

Dataset	Arrival rate(vehicles/300s)				Total traffic volume
	Mean	Std	Max	Min	
Grid-1	201.95	24.54	238	156	12117
Grid-2	213.47	24.65	235	163	12808
Jinan	104.92	19.79	136	50	6295
Hangzhou	108.97	19.87	134	75	6538

4.3. Evaluation Metrics

To evaluate the performance of different multi-intersection traffic signal control methods, Average Travel Time (ATT), defined in Eq. (12), is adopted as the primary evaluation metric. ATT measures the average time required for vehicles to travel from their origins to destinations and provides a direct reflection of overall traffic efficiency.

All experimental results are obtained from five independent runs with different random seeds and are reported as the mean value with standard deviation. For all tables, the reported ATT values correspond to the average results over the last 10 testing episodes, ensuring evaluation stability and statistical reliability.

4.4. Baseline Methods

The proposed method is compared against a range of representative baseline approaches, including both traditional traffic signal control methods and reinforcement learning–based methods.

Traditional methods:

Fixed Time [21]: This method employs predefined cycle lengths and phase durations with added random offsets. It is commonly used in scenarios with relatively stable traffic flows.

Max Pressure [22]: This approach selects green phases based on the pressure, defined as the difference between upstream and downstream queue lengths at intersections. It aims to minimize the total traffic pressure at the network level and is regarded as one of the most effective traditional signal control methods.

Reinforcement learning–based methods:

One Model [23]: In this method, agents adopt the same state and reward design as conventional single-intersection RL approaches and only consider traffic conditions on roads directly connected to the controlled intersection. All agents share a single policy network instead of maintaining separate parameters for each intersection.

Press Light [24]: An advanced traffic signal control algorithm that utilizes traffic pressure (queue length differences) as the reward signal to optimize multi-intersection signal control at the network level.

Co Light [7]: This method employs a Graph Attention Network (GAT) to adaptively aggregate information from neighboring intersections, thereby improving queue length performance in multi-intersection traffic signal control.

Table 4. Performance comparison of all methods on synthetic and real-world datasets in terms of Average Travel Time (in seconds). Values before the symbol "±" indicate the mean, and values after represent the standard deviation.

Methods	Grid-1	Grid-2	Jinan	Hangzhou	Average
Fixedtime	923.54	913.61	814.09	806.36	864.40
MaxPressure	767.92	803.55	394.68	583.69	637.46
OneModel	559.14±164.84	540.49±75.04	291.54±17.03	515.40±17.03	476.64
PressLight	583.57±31.40	617.03±25.14	333.32±28.56	619.07±23.53	538.25
CoLight	383.12±61.30	532.86±49.98	282.96±2.49	528.19±36.17	431.78
Ours	300.70±22.65	283.54±7.28	245.92±2.55	424.21±17.91	313.59

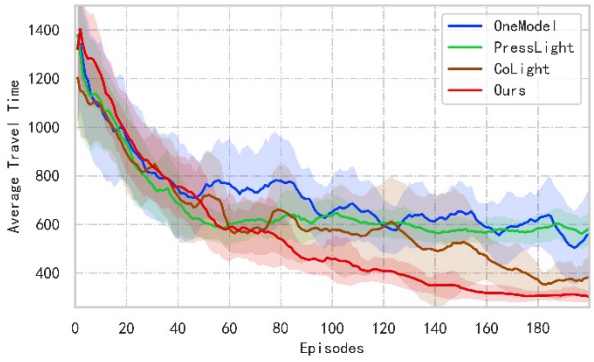
4.5. Overall Performance

The proposed method is compared with all baseline approaches, and the results are summarized in Table 4. As can be observed, the proposed method consistently outperforms all baselines across all datasets, achieving a performance improvement of at least 27.37%. Notably, the performance gains are particularly significant on the two synthetic datasets with heavier traffic demand. Compared with the best-performing baseline method, the proposed approach reduces the average travel time by 21.51% and 46.78%, respectively. On the real-world datasets with relatively lower traffic volumes, the average travel time is reduced by 13.09% and 19.69%, respectively. These results indicate that as traffic demand increases and congestion pressure intensifies, the requirement for richer and more expressive traffic

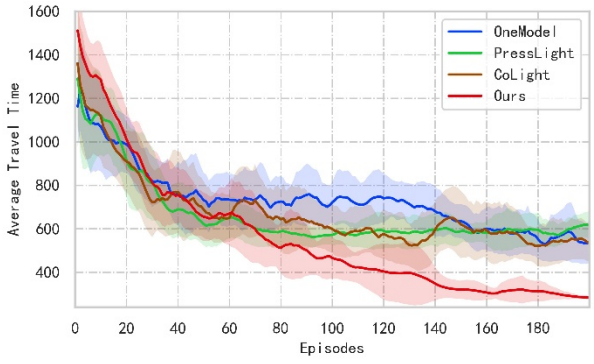
information becomes increasingly critical. By incorporating diverse traffic features and performing fine-grained modeling, ESR-MARL effectively enhances agent decision-making capability, leading to substantial reductions in average travel time.

4.6. Convergence Analysis

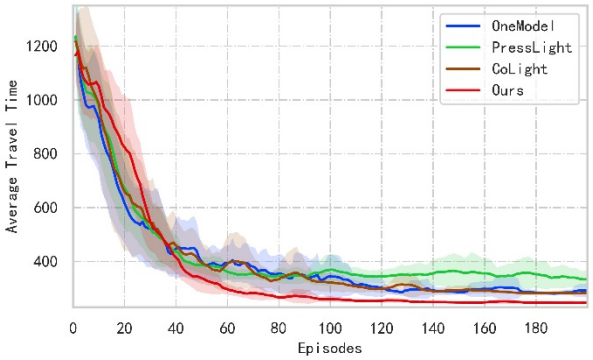
The learning curves of the proposed method and several baseline models are compared in Fig. 3. Traditional traffic signal control methods are not included, as they do not require a training process. It can be observed that the proposed method achieves both faster convergence and superior final performance compared with the other learning-based approaches. Specifically, Press Light, which adopts a decentralized DQN architecture, tends to suffer from training instability and struggles to reach optimal performance even after prolonged training. In contrast, ESR-MARL leverages richer traffic information and applies a channel attention mechanism to independently learn and effectively fuse heterogeneous traffic features, resulting in a more comprehensive traffic state representation. This design significantly improves both stability and performance in multi-intersection cooperative control scenarios, enabling ESR-MARL to outperform advanced methods such as CoLight.



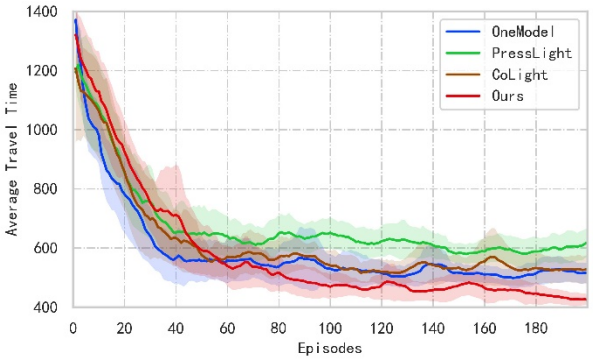
(a) Grid-1



(b) Grid-2



(c) Jinan



(d) Hangzhou

Figure 3. Training curves. The curves are smoothed using a moving average with a window size of 10. Solid lines represent the average performance, and the shaded areas indicate the standard deviation.

5. Conclusion

This paper addresses the limitations of traditional traffic signal control methods in dynamic traffic environments, as well as the insufficient state representations adopted by existing reinforcement learning–based approaches. To this end, we propose an efficient state representation framework for multi-agent reinforcement learning, termed ESR-MARL. The key innovation of the proposed method lies in constructing a fine-grained traffic state by incorporating richer and more informative multi-dimensional traffic features, including queue length, vehicle density, system delay, and network-level pressure. Moreover, a channel attention mechanism is introduced to dynamically evaluate and fuse the importance of different feature channels, thereby yielding a more comprehensive and discriminative representation of the traffic state. Extensive experiments conducted on both synthetic and real-world traffic datasets demonstrate that ESR-MARL significantly enhances the performance of multi-agent cooperative control. In terms of the critical metric of average travel time, the proposed method outperforms state-of-the-art baseline approaches by at least 27.37%, validating its effectiveness and robustness. Beyond providing a high-performance solution for traffic signal control, the attention-based feature fusion framework proposed in this study also offers valuable insights for state modeling in complex dynamic systems.

Future work will focus on incorporating additional real-time traffic characteristics, such as vehicle speed distributions and phase-switching balance, and further exploring the interplay between traffic information and multi-agent coordination mechanisms. These efforts aim to construct

more accurate state representations and further improve traffic signal control strategies.

References

[1] P. W. Shaikh, M. R. Shah, A. A. Shaikh, and S. A. Mahar, “A review on swarm intelligence and evolutionary algorithms for solving the traffic signal control problem,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 48–63, 2022.

[2] M. E. M. Ali, A. Durdu, S. A. Çeltek, and M. A. Özdemir, “An adaptive method for traffic signal control based on fuzzy logic with Webster and modified Webster formula using SUMO traffic simulator,” IEEE Access, vol. 9, pp. 102985–102997, 2021.

[3] Z. Yu, N. Nianwen, Y. Zheng, Y. Lv, F. Liu, and Y. Zhou, “Review of intelligent traffic signal control strategies driven by deep reinforcement learning,” Computer Science, vol. 50, no. 4, pp. 159–171, 2023.

[4] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[5] A. Haydari and Y. Yilmaz, “Deep reinforcement learning for intelligent transportation systems: A survey,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 11–32, 2022.

[6] J. J. A. Calvo and I. Dusparic, “Heterogeneous multi-agent deep reinforcement learning for traffic lights control,” in Proc. 26th Irish Conf. Artificial Intelligence and Cognitive Science (AICS), 2018, pp. 1–12.

[7] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, and Z. Li, “CoLight: Learning network-level cooperation for traffic signal control,” in Proc. 28th ACM Int. Conf. Information and Knowledge Management (CIKM), 2019, pp. 1913–1922.

[8] D. Garg, M. Chli, and G. Vogiatzis, “Deep reinforcement learning for autonomous traffic light control,” in Proc. 3rd IEEE Int. Conf. Intelligent Transportation Engineering (ICITE), Sep. 2018, pp. 214–218.

[9] Z. Zhao, K. Wang, Y. Wang, and X. Liang, “Enhancing traffic signal control with composite deep intelligence,” Expert Systems with Applications, vol. 244, Art. no. 123020, 2024.

[10] H. Wei, G. Zheng, H. Yao, and Z. Li, “IntelliLight: A reinforcement learning approach for intelligent traffic light control,” in Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2018, pp. 2496–2505.

[11] Q. Wu, L. Zhang, J. Shen, L. Lü, B. Du, and J. Wu, “Efficient pressure: Improving efficiency for signalized intersections,” arXiv preprint arXiv:2112.02336, 2021.

[12] L. Zhang, Q. Wu, J. Shen, L. Lü, B. Du, and J. Wu, “Expression might be enough: Representing pressure and demand for reinforcement learning based traffic signal control,” in Proc. Int. Conf. Machine Learning (ICML), 2022, pp. 26645–26654.

[13] L. Li, R. Li, Y. Peng, C. Huang, and J. Yuan, “Cooperative max-pressure enhanced traffic signal control,” in Proc. 31st ACM Int. Conf. Information and Knowledge Management (CIKM), 2022, pp. 4173–4177.

[14] Y. Sun, K. Lin, and A. Kashif, “KeyLight: Intelligent traffic signal control method based on improved graph neural network,” IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 2861–2871, 2024.

[15] K. Yang, Z. Wang, X. Meng, L. Li, Y. Shi, Y. Yu, and Z. Yao, “Store-and-forward with graph attention: Enhanced multi-agent reinforcement learning for emergency-responsive traffic signal control,” Engineering Applications of Artificial Intelligence, vol. 159, Art. no. 111602, 2025.

[16] Y. Zhang, H. Goel, P. Li, M. Damani, S. Chinchali, and G. Sartoretti, “CoordLight: Learning decentralized coordination for network-wide traffic signal control,” IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 6, pp. 8034–8049, 2025.

[17] S. Mousavi, M. Schukat, and E. Howley, “Traffic light control using deep policy-gradient and value-function-based reinforcement learning,” IET Intelligent Transport Systems, vol. 11, no. 7, pp. 417–423, 2017.

[18] G. Zheng, Y. Xiong, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, and Z. Li, “Learning phase competition for traffic signal control,” in Proc. 28th ACM Int. Conf. Information and Knowledge Management (CIKM), 2019, pp. 1963–1972.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Lecture Notes in Computer Science, vol. 11211, 2018, pp. 3–19.

[20] H. Zhang, S. Feng, C. Chen, W. Zhang, Y. Zhu, Z. Li, and Z. Wang, “CityFlow: A multi-agent reinforcement learning environment for large scale city traffic scenario,” in Proc. The World Wide Web Conf. (WWW), 2019, pp. 3620–3624.

[21] J. Koonce and L. Rodegerdts, Traffic Signal Timing Manual, Tech. Rep. FHWA-HOP-08-024, Federal Highway Administration, Washington, DC, USA, 2008.

[22] P. Varaiya, “The max-pressure controller for arbitrary networks of signalized intersections,” in Advances in Dynamic Network Modeling in Complex Transportation Systems. Springer, 2013, pp. 27–66.

[23] T. Chu, J. Wang, L. Codecà, and Z. Li, “Multi-agent deep reinforcement learning for large-scale traffic signal control,” IEEE Transactions on Intelligent Transportation Systems, 2019.

[24] C. Chen, W. Zhang, Y. Zhu, G. Zheng, and Z. Li, “Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control,” in Proc. AAAI Conf. Artificial Intelligence, vol. 34, 2020, pp. 3414–3421.