FGLight: Learning Neighbor-level Information for Traffic Signal Control

Hang Xiao
School of Software, Northwestern Polytechnical University, 710072
Xi'an, China
shawh@mail.nwpu.edu.cn

Huale Li*
School of Software, Northwestern Polytechnical University, 710072
Yangtze River Delta Research Institute of NPU, Taicang, 215400
Xi'an, China
hualeli@nwpu.edu.cn

Shuhan Qi
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
shuhanqi@cs.hitsz.edu.cn

Jiajia Zhang
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
zhangjiajia@hit.edu.cn

DingZhong Cai
School of Software, Northwestern Polytechnical University, 710072
Xi'an, China
caidz@mail.nwpu.edu.cn

ABSTRACT

In recent years, multi-agent reinforcement learning (MARL) methods have increasingly been applied to traffic signal control and have achieved some success. However, most of existing MARL methods often underemphasize the heterogeneity in neighborhood-level information of the same agent. This results in highly sensitive performances and a long learning process. To address this challenge, we propose FGLight, a novel Feudal MARL method for traffic signal control. FGLight leverages Adaptive Graph Attention Networks (AGAT) to dynamically model the interactive relationships between intersections. Through adaptive neighbor selection and weight-based attention mechanisms, AGAT dynamically assigns importance weights to neighbor-level information, thereby improving the accuracy of local policies by more effectively exploiting neighborhood information. Moreover, FGLight introduces a Smooth Hysteretic Deep Q-Network (SHDQN) based on an optimistic assumption mechanism, which enhances the stability of the global policy. We conducted experiments on both synthetic and real-world datasets, and the results demonstrate that, compared to several state-of-the-art MARL methods, FGLight performs better as the complexity of the road network increases, exhibiting faster convergence and greater policy stability.

KEYWORDS

Traffic Signal Control; Multi-agent Reinforcement Learning; Graph Attention Networks

ACM Reference Format:
Hang Xiao, Huale Li*, Shuhan Qi, Jiajia Zhang, and DingZhong Cai. 2025. FGLight: Learning Neighbor-level Information for Traffic Signal Control.

*Huale Li is the corresponding author.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

The ongoing process of urbanization has led to an escalation in urban traffic congestion, presenting a significant challenge for modern cities. In recent years, the rapid advancement of artificial intelligence (AI) technologies has led to the growing trend of utilizing AI to address traffic congestion issues. Traffic signal control plays a critical role in urban traffic management, and optimizing traffic signal control is a key approach to alleviating congestion. Historical data suggests that optimizing traffic signal control systems can improve surface traffic efficiency by 10% to 20% [1]. Traditional traffic signal control methods primarily rely on fixed timing and manual settings, which are based on predefined rules and lack the ability to dynamically adjust signals according to real-time traffic conditions [2]. Given the sharp increase in urban traffic flow, these traditional methods are no longer adequate to meet the current traffic demands.

In recent years, researchers have increasingly applied Deep Reinforcement Learning (DRL) to traffic signal control, achieving notable results that generally surpass traditional control methods [3–6]. Wei, H et al. proposed the CoLight model [7], which achieved efficient cooperation among traffic signals through learning dynamic communication and index-free modeling, significantly improving traffic signal control performance. Chacha Chen et al. proposed the MPLight model [8], which further realized large-scale traffic signal control through a decentralized reinforcement learning paradigm and parameter sharing. Bingyu Xu et al. proposed a novel hierarchical and cooperative RL method, HiLight [9], which adopts a hierarchical structure and introduces multi-critic and adaptive weighting mechanisms to optimize long-term objectives. Ye, Y et al. presented the InitLight model [10], which generates initial models for traffic signal control by combining adversarial inverse reinforcement learning with expert trajectories from single-intersection environments. Zhang, H et al. proposed the MATLight model [11],

which achieves coordinated control of multi-intersection traffic signals through heterogeneous agent mirror learning combined with Transformer architecture.

Despite the progress achieved by DRL-based models [7–11], several critical challenges remain in large-scale traffic signal control. A key objective is to optimize the average travel time of all vehicles across the entire road network. However, coordinating traffic signals across the network to directly optimize this metric is still an open problem that requires further exploration. HiLight [9] was specifically designed to address the challenge of reducing average travel time across the network. It enables each agent to learn its own control strategy, using a multi-critic mechanism to optimize both local and neighborhood average travel times. An adaptive weighting system promotes agent cooperation by balancing local and neighborhood objectives. However, the reliance on coordinating multiple sub-policies limits its ability to achieve global optimization in complex networks, particularly in areas with variable traffic and dense intersections, where low-level optimization alone proves insufficient. These challenges hinder HiLight's scalability and broader application.

To address this challenge, we propose a hierarchical reinforcement learning method based on Adaptive Graph Attention Networks (AGAT), called FGLight. FGLight employs a hierarchical multi-objective optimization reinforcement learning framework to achieve efficient intersection neighborhood communication and cooperation. Specifically, at the lower-level interaction layer, FGLight leverages AGAT to dynamically assign importance weights to neighbor-level information, enhancing the accuracy of local decision-making by further exploring neighborhood information. This approach enables each intersection agent to adaptively select relevant neighboring intersections and selectively share and leverage traffic information based on importance-weighted criteria. At the upper-level decision-making layer, FGLight introduces an optimistic assumption mechanism and proposes a Smooth Hysteretic Deep Q-Network (SHDQN). This innovative approach employs dynamic learning rate adjustments to differentiate between positive and negative temporal difference errors during Q-value updates, thereby increasing the probability of positive updates. Through this design, the upper-level decision-making layer effectively mitigates the impact of environmental uncertainties, enhances the stability of the learning process, and accelerates convergence towards optimal traffic management strategies. Our contributions are summarized as follows:

- We propose a DRL-based traffic signal control method, FGLight, which further explores the importance of neighbor information, focusing on the optimization of multiple traffic signal controls in complex road networks.
- FGLight dynamically balances the importance of local and neighborhood objectives through an adaptive weighting mechanism, enhancing the accuracy of local policies by efficiently utilizing neighborhood information. Furthermore, FGLight introduces a SHDQN based on an hysteretic assumption mechanism, improving the stability of the global policy.
- We conducted extensive experiments on both synthetic traffic grids and real-world traffic networks. The experimental results demonstrate that our proposed FGLight outperforms existing DRL methods and traditional control approaches across all evaluation metrics.

2 RELATED WORK

In the field of traffic signal control, traditional methods rely on manually designed signal rules [3, 4, 6], which employ predefined signal cycles and are broadly applicable to stable traffic flow control. However, these approaches primarily focus on steady traffic states and lack the ability to dynamically adjust based on real-time traffic conditions. To address these limitations, adaptive traffic control systems [12–14] have been developed, which predefine a range of traffic signal parameters (e.g., cycle lengths, phase allocations, and offsets) and adjust these in response to real-time traffic volumes. Despite these advances, designing effective traffic signal rules remains a challenging task. In response to these challenges, Varaiya et al. proposed the max-pressure [5] algorithm, which balances queue lengths between adjacent intersections by minimizing the pressure at each intersection, assuming infinite downstream link capacity. However, these assumptions often do not hold in real-world environments, resulting in suboptimal performance in practice.

In recent years, many researchers in traffic management have begun exploring reinforcement learning (RL) methods for traffic signal control, which bypass the need for predefined rules and assumptions. RL-based methods [15, 16] have shown superior performance compared to traditional approaches. Some researchers have treated each intersection as an independent agent, employing Independent Reinforcement Learning (IRL) methods [17–21], enabling each agent to adjust signal phases and cycles based on locally observed traffic conditions. For example, Wei et al. introduced the IntelliLight [17], which adjusts signal phases in real-time. Zhang et al. developed the FRAP [22], addressing phase competition through symmetry modeling. Liang et al. proposed a DRL method [16], utilizing convolutional neural networks to map states to rewards. These methods are scalable to multi-intersection scenarios. However, a major limitation of IRL approaches is their failure to account for interactions between adjacent intersections within the road network. Agents make decisions based solely on local information, without a global perspective, hindering the optimization of global objectives. To address this, some studies [23, 24] have proposed using observations from all intersections as inputs to a centralized model, which makes decisions for each intersection. However, this centralized approach suffers from the curse of dimensionality as the number of intersections increases, leading to exponential growth in the joint action space.

To enable large-scale coordinated traffic signal control, researchers have increasingly turned to multi-agent reinforcement learning (MARL) methods. Chu et al. introduced the MA2C algorithm [25], designed to address the scalability issues inherent in centralized RL approaches. Building on MA2C, Jinming et al. proposed FMA2C [26], integrating hierarchical reinforcement learning with MA2C to achieve both global coordination and scalability. Xu et al. introduced HiLight [9], a hierarchical and cooperative RL method utilizing a layered structure with multiple critics and adaptive weighting to optimize long-term objectives. Chen et al. proposed MPLight [8], advancing large-scale traffic signal control through decentralized

RL and parameter sharing. Wei et al. developed CoLight [7], a MARL model that enhances agent cooperation by learning dynamic communication patterns and index-free modeling, improving signal control performance. Despite the progress of MARL methods such as CoLight and PressLight [27], these models often lack effective global coordination mechanisms. They fail to fully leverage dynamic information from neighboring intersections, leading to suboptimal global performance. Additionally, their limited communication mechanisms inhibit the ability to exploit the temporal and spatial characteristics of traffic flow, hindering optimal control across the entire road network.

3 METHOD

In this section, we introduce the architecture of proposed FGLight in the paper. FGLight is an innovative end-to-end Federated Multi-Agent Reinforcement Learning (FMARL) framework specifically developed to optimize large-scale traffic signal control. As shown in figure 1, FGLight is composed of three essential components: the feature extraction layer, lower-level interaction layer, and upper-level decision-making layer. The feature extraction layer processes real-time traffic data such as vehicle counts and queue lengths, providing crucial information for decision-making. The upper-level decision-making layer leverages AGAT, enabling adaptive and dynamic communication between intersections. AGAT's adaptive neighbor selection mechanism allows each intersection to dynamically choose relevant neighboring nodes as traffic conditions change. This is complemented by a weight-based attention mechanism that assigns importance weights to neighbor-level information. At the lower-level, FGLight employs an Hysteretic Update Module (HUM) along with a SHDQN to ensure stability and accuracy in policy updates. This two-tiered system balances local decision accuracy with global traffic optimization, ensuring effective control across complex road networks.

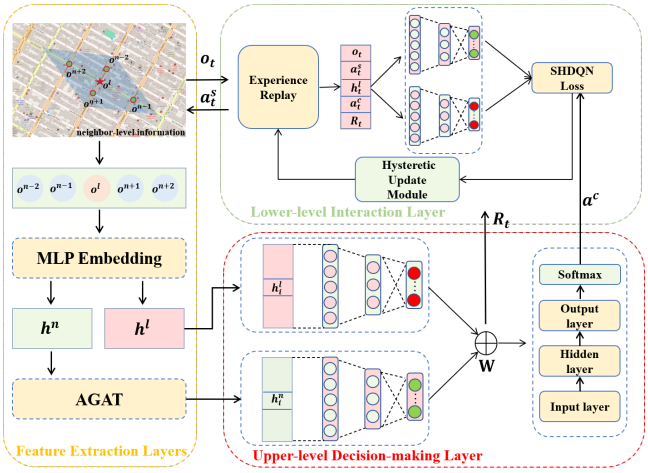


Figure 1: The framework of FGLight. The yellow, red, and green boxes represent the feature extraction layers, upper-level decision-making layer, and lower-level interaction layer, respectively.

3.1 PROBLEM DEFINITION

The traffic signal control problem is formulated as a Markov Decision Process (MDP) [28], where each agent is tasked with controlling a single intersection within the traffic network. Each agent operates under partial observability, meaning it can only observe its own state and those of its local neighborhood. The objective is to derive a policy that selects the optimal control phase for each intersection based on these partially observable states, thereby minimizing the average travel time across the entire traffic network. In our problem framework, we assume that there are n intersections in the system, and the number of agents is the same as the number of intersections, which is defined as n. The Markov game is defined by the following components:

- State Space: The state space $S = \{s_1, s_2, ..., s_m\}$ represents the set of all possible states of the entire traffic network, including current traffic signal phases and the number of vehicles on the roads, where each state represents the current phase combination and the number of vehicles on the lanes of the road network globally.
- Observation Space: Due to partial observability, each agent's observation space O includes only the states of its own intersection and those within its local neighborhood.
- Action Space: The action space $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ comprises the possible actions available to each agent, which in this case are the different traffic signal phases. Agents select the next phase based on the observed states. Possible phases include red, yellow, and green lights, in which each phase is encoded as one-hot encoding, and four signals at each intersection, which are connected to form a 12-dimensional vector.
- Reward Function: The reward function $R_s = E[R_{t+1}|S_t = s]$ defines the feedback each agent receives after taking an action. Upon executing an action, the system receives a reward signal derived from the reward function.
- State Transition Function: The state transition function $P = P[S_{t+1} = s'|S_t = s]$ describes the probability distribution of the traffic system transitioning from one state to another after an action is performed.
- Policy: The policy determines how agents select actions based on their current observations. Each agent follows a policy to take actions that aim to maximize long-term cumulative rewards, with reinforcement learning being used to discover the optimal policy.

3.2 The Architecture of FGLight

In large-scale traffic signal control, FGLight adopts a distributed approach by assigning agents to individual intersections within a hierarchical feudal architecture. As illustrated in Figure 1, this structure consists of two key components: an upper-level decision-making layer and a lower-level interaction layer. The upper-level layer, featuring an AGAT module, dynamically adjusts relevant intersection sets and assigns importance weights to neighbor-level information based on changing traffic conditions. It operates on a macro time scale, selecting sub-policies every T steps to optimize long-term goals. In contrast, the lower-level layer functions on a micro time scale, implementing the HUM for learning stability and

executing specific control actions at each step within the T-step interval. This dual-layer design creates a synergistic approach to traffic management, with the upper-level layer overseeing broad, long-term strategies through AGAT's adaptive spatial awareness, while the lower-level layer ensures real-time responsiveness to dynamic traffic conditions.

The subsequent sections will provide a detailed examination of the four core components of the FGLight framework: the upper-level decision-making layer, the AGAT module, the lower-level interaction layer, and the HUM module, offering detailed insights into how they contribute to achieving efficient and effective traffic signal control.

3.2.1 The Upper-level Decision-making Layer. The action of the upper-level decision-making layer, denoted as a^c, involves selecting and implementing a sub-policy for T consecutive time steps. This layer receives a reward r_l, defined as the negative local travel time. However, the primary objective is to minimize the average travel time across the entire road network, not just locally. Optimizing local travel times in isolation does not necessarily lead to global optimization due to potential conflicts between individual upper-level decision-making layer strategies. Such conflicts may result in suboptimal performance with respect to the network-wide average travel time.

To address this issue, the upper-level decision-making layer incorporates an additional reward metric: the neighborhood travel time r_n. For a given intersection, r_n is defined as a statistical measure of local travel times for all intersections within its neighborhood, including itself. By optimizing r_n, the upper-level decision-making layer accounts for the strategies of neighboring agents, thereby enhancing performance over a broader area. This approach facilitates inter-agent coordination and promotes strategy alignment, minimizing conflicts. Consequently, it contributes to the optimization of the overall average travel time across the entire road network.

To jointly optimize both local and neighborhood travel times, the upper-level decision-making layer employs an actor-critic reinforcement learning approach incorporating an AGAT mechanism. This layer utilizes two value networks: $V_l(o;\phi_l)$ and $V_n(\hat{o};\phi_n)$, which estimate the value functions for local and neighborhood travel times, respectively, under policy $\pi(a_c \mid o;\phi_\pi)$. These value functions and the policy are parameterized by ϕ_l, ϕ_n, and ϕ_π, respectively. Notably, the input $\hat{o}$ to the neighborhood value network V_n differs from the input o to the local value network V_l and policy π. The input $\hat{o}$ is constructed by concatenating observations from multiple intersections within the neighborhood. This actor-critic approach effectively integrates gradient signals from both V_l and V_n, facilitating the computation of the policy gradient for $\pi(a_c \mid o;\phi_\pi)$. The policy gradient is calculated as follows:

$$\nabla\phi^\pi = \mathbb{E}[\log \pi(a^c|o;\phi^\pi)(\delta^l + w\delta^n)] \tag{1}$$

$$\delta^l = r^l + \gamma V^l(Embed(o');\phi^l) - V^l(Embed(o);\phi^l) \tag{2}$$

$$\delta^n = r^n + \gamma V^n(AGAT(Embed(\widehat{o'}));\phi^n) - V^n(AGAT(Embed(\widehat{o}));\phi^n) \tag{3}$$

where δ^l denotes the advantage function for the local travel time, and δ^n represents the advantage function for the neighborhood travel time. The weight coefficient w is employed to balance the contributions of these two advantage functions in the overall policy optimization process.

3.2.2 The Lower-level Interaction Layer. The lower-level interaction action a^s primarily offers two options: maintaining the current phase or transitioning to the next phase in the subsequent time step. However, research by Chen et al. [8] at Pennsylvania State University suggests that a^s is not confined to this binary choice; it can directly select a specific phase for the next time step. The rewards for the lower-level interaction action are defined as the negative values of the total waiting time r^w, the total delay r^d, and the total queue length r^q across all lanes. These rewards are defined as follows: the waiting time is the total time vehicles spend waiting (with speeds lower than 0.1 meters per second); the delay is the difference between the maximum speed and the current speed of the lane, divided by the maximum speed; and the queue length is the total number of waiting vehicles on incoming lanes.

In the lower-level interaction layer, SHDQN is implemented to model the agent-environment interaction process. This method employs differential learning rates to minimize the loss function, thereby enhancing decision-making efficacy. The loss function is defined as:

$$\mathcal{L}_\theta^{SHDQN} = \sum_{k=1}^{B} \delta_k^2 \tag{4}$$

$$\delta = r + \gamma \max_{a'} Q\left(AGAT(o'), a'; \theta'\right) - Q(AGAT(o), a; \theta) \tag{5}$$

where δ denotes the temporal difference (TD) error, which quantifies the discrepancy between the predicted Q-value and the updated Q-value based on the subsequent observation. B represents the batch size, indicating the number of samples utilized in each network update. θ' denotes the parameters of the target network, which is periodically synchronized with the main network parameters θ. o' represents the observation at the subsequent time step.

3.2.3 Adaptive Graph Attention Networks for Cooperation. In a MARL environment, the judicious utilization of heterogeneity in neighborhood-level information for individual agents emerges as a critical factor in optimizing inter-agent communication processes. To address this issue, FGLight employs an AGAT to learn the relative importance of neighboring intersections for communication, generating decision importance weight information for the agents' neighborhood. Based on this mechanism, agents learn to model the influence of their neighborhood and take actions according to the weighted observations within the neighborhood, considering the varying importance of different neighbors and adapting to evolving traffic patterns.

Given the raw data from local observations, comprising the vehicle count on each lane and the current traffic signal phase, we initially embed these k-dimensional data into an m-dimensional latent space. This embedding is achieved through a multilayer perceptron (MLP):

$$h_i = Embed(o_i^t) = \sigma(o_i W_e + b_e) \tag{6}$$

where $o_i^t \in \mathbb{R}^k$ denotes the observation vector of intersection i at time step t; k represents the dimensionality of the feature space; W_e and b_e are the learnable weight matrix and bias vector, respectively; $\sigma(\cdot)$ is the ReLU activation function; and $h_i \in \mathbb{R}^m$ represents the

resulting embedded representation, capturing the current traffic state at intersection i.

To implement the dynamic neighbor selection mechanism, the AGAT utilizes a LSTM-based approach that adaptively determines agent interactions. This mechanism encodes each agent's local observation into a feature vector, which is then processed by a LSTM to extract time-series features and historical information. For each agent pair (i, j), their LSTM hidden states are concatenated and fed into a MLP, outputting a binary value that determines the existence of an interaction relationship. This can be summarized by the formula:

$$W_{i,j} = f^g(\text{concat}(h_i, h_j)) \tag{7}$$

where $W_{i,j}$ is the binary interaction indicator, h_i and h_j are LSTM hidden states of agents i and j, f^g is the MLP function, and concat is the concatenation operation.

To quantify the relevance of observational information from neighboring agents in the decision-making process of the target agent, we first extract representation vectors from the observation embedding layer for both the target and neighboring agents. The importance weight of agent j to the decision-making of agent i, denoted as e_{ij}, is then computed using the following attention mechanism:

$$e_{ij} = W_{i,j} \cdot (h_i W_t) \cdot (h_j W_s)^T \tag{8}$$

where $W_t, W_s \in \mathbb{R}^{m \times n}$ represent the observation embedding parameters for the target agent and the neighboring agents, respectively.

To derive generalized attention values that quantify the relative importance of neighboring agents to the target agent, we further normalize the interaction scores between the target agent i and its neighboring agents. This normalization is achieved through the following formula:

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp\left(\frac{e_{ij}}{\tau}\right)}{\sum_{j \in \mathcal{N}_i} \exp\left(\frac{e_{ij}}{\tau}\right)} \tag{9}$$

where τ is the temperature factor, N_i represents the set of agents within the neighborhood of the target agent, and $\mid \mathcal{N}_i \mid$ is the number of neighboring agents in the target agent's vicinity.

It is crucial to note that the set N_i includes the target agent i itself, enabling the agent to assess the relative importance of its own traffic conditions in the decision-making process. This self-attention mechanism allows the agent to balance the significance of its local state against the information from neighboring agents. The generalized attention score α_{ij} offers several advantages. It is adaptable to diverse road network topologies, including intersections with varying numbers of branches. Furthermore, it employs a relaxed neighborhood concept, allowing for the inclusion of non-adjacent agents in N_i under certain conditions, extending beyond immediate neighbors. This approach facilitates comprehensive information integration, potentially incorporating a wider range of agents and considering a more expansive and nuanced view of the traffic situation.

To encapsulate the aggregated influence of neighboring agents on the target agent, we compute a weighted sum of the hidden states of multiple neighboring agents, with the weights determined by their respective importance. This aggregation can be formalized as follows:

$$hs_i = \sigma(W_q \cdot \sum_{j \in \mathcal{N}_i} \alpha_{ij}(h_j W_c) + b_q) \tag{10}$$

where $W_c \in \mathbb{R}^{m \times c}$ denotes the weight matrix for embedding the source intersection, and W_q and b_q are trainable parameters. The weighted sum of the neighborhood representation $hs_i \in \mathbb{R}^c$ aggregates critical information from the surrounding environment, facilitating the execution of efficient signal strategies. For agent i, the cooperative information hs_i encapsulates the importance-based relationships with its neighboring agents.

To simultaneously focus on neighborhood information in different representation subspaces across various locations, this study extends the single-head attention mechanism in the AGAT to multi-head attention. Specifically, K different linear projections (i.e., multiple sets of trainable parameters W_c, W_t, W_s) are employed to perform the attention function in parallel. This attention function encompasses observation interaction, attention distribution computation, and neighborhood cooperation. The diverse representations of neighborhood conditions generated by these K attention heads are subsequently aggregated and averaged into a final representation hm_i:

$$hm_i = \sigma(W_q \cdot (\frac{1}{H} \sum_{h=1}^{h=H} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^h (h_j W_c^h)\Big) + b_q) \tag{11}$$

3.2.4 Hysteretic Update Mechanisms. In the lower-level interaction layer, a Hysteretic Update Mechanism is proposed and implemented through a Smooth Hysteretic Deep Q-Network (SHDQN) method. This novel approach employs differential learning rates during the Q-value update process to distinguish between updates arising from positive and negative TD errors. This adaptive learning rate mechanism enhances the stability and performance of the policy by facilitating more nuanced adjustments based on the direction and magnitude of the TD error.

Specifically, the SHDQN employs two distinct learning rates in the Q-value update formula: one for positive TD errors (i.e., when updating the Q-value to increase it) and another for negative TD errors (i.e., when updating the Q-value to decrease it). The core principle of this approach is to stabilize the learning process by attenuating the response to negative TD errors, thereby promoting exploration. The update formula for the Smoothed Hysteretic Deep Q-Network can be expressed as follows:

$$L_\theta^{SHDQN} = \sum_{k=1}^{B} \bar{\delta}_k^2 \tag{12}$$

$$\bar{\delta_k} = \delta_k(\sigma(\delta_k) + 0.5) \tag{13}$$

where $\bar{\delta_k}$ denotes the weighted temporal difference error, σ denotes the sigmoid function, and δ_k denotes the temporal difference error.

Traditional Q-value update methods often lead to unstable learning processes and premature exploitation. To address these issues, this study proposes the Smoothed Hysteretic mechanism, which effectively balances exploration and exploitation by differentiating the impact of positive and negative TD errors. This approach mitigates the problem of insufficient exploration caused by excessive penalization of negative experiences, thereby enhancing the algorithm's performance in complex traffic environments.

4 EXPERIMENTS

In this section, we outline the experimental setting, including datasets, baselines, evaluation metrics, and describes various comparative experiments designed to validate the effectiveness of the proposed FGLight method. Additionally, ablation studies are conducted by removing the AGAT mechanism and the HUM module from FGLight to demonstrate the contribution of each component to the improvement of traffic signal control effectiveness.

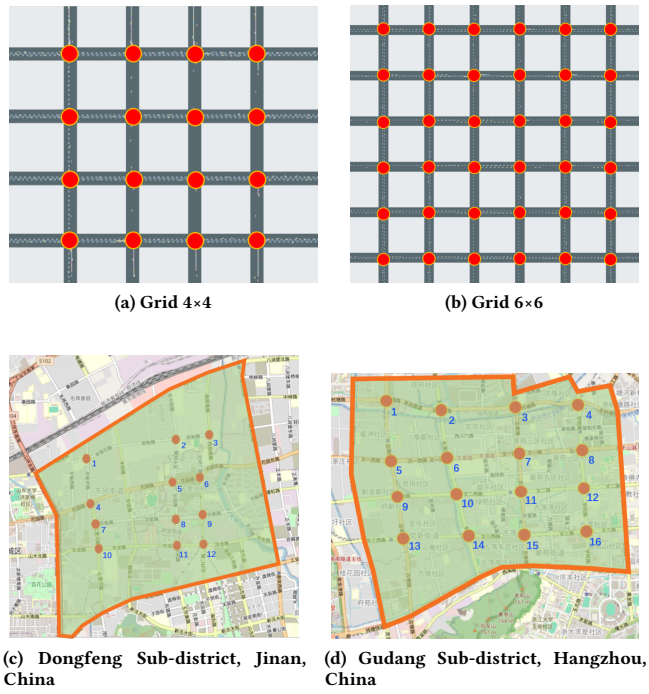
4.1 Experiment Settings

For performance evaluation, we utilized CityFlow [29], a popular open-source traffic simulator that supports large-scale traffic signal control, as our experimental simulation environment. CityFlow is widely adopted in traffic signal control research due to its efficient simulation speed and flexible configuration capabilities. Within this simulation environment, various traffic flow scenarios were designed and configured, including both synthetic scenarios and real-world traffic patterns, to assess the performance of the FGLight method under diverse traffic conditions.

4.1.1 Datasets. Our experiments utilized two types of datasets [30] obtained through different methods: synthetic datasets and real-world datasets. The synthetic dataset includes the Grid 4×4 dataset and the Grid 6×6 dataset. The real-world datasets comprise the Jinan 3×4 dataset, and the Hangzhou 4×4 dataset. Each vehicle data entry includes a timestamp, the vehicle's starting position, and its destination within the environment. These data were collected in real-time through roadside sensors and surveillance systems, ensuring accuracy and reliability. The datasets encompass both straight and turning vehicle flows, with all lanes designed for two-way traffic. As a result, the data encapsulate intricate and dynamic bidirectional traffic information, offering a thorough assessment of the algorithm's performance in complex traffic scenarios.

4.1.2 Baselines. To evaluate the effectiveness of the proposed FGLight, the experiments compare two types of methods: traditional traffic signal control algorithms and multi-agent reinforcement learning algorithms. The details are as follows:

- Fixed-Time [3]: This method sets a fixed timing plan for traffic signals based on historical traffic flow data. The green light duration for each phase and the sequence of signal phases are predetermined and do not adapt in response to real-time traffic conditions. This approach is characterized by its simplicity of implementation and is suitable for environments with stable traffic flow.
- SOTL [14]: This method determines the timing of signal changes based on local traffic data, such as the number of vehicles and waiting time. Specific thresholds are set, and when the number of waiting vehicles or the waiting time on one side of the intersection reaches these preset values, the traffic signal automatically switches to allow vehicles to pass.
- MaxPressure [5]: This method optimizes traffic signal control by calculating the pressure at intersections in real-time, which is defined as the difference between the number of vehicles entering and exiting the lanes. The goal is to maximize overall network traffic flow and reduce congestion. During each signal cycle, the MaxPressure method dynamically adjusts the green light duration based on the pressure from each direction, prioritizing the signal phases that can most effectively reduce overall pressure.
- PressLight [27]: A reinforcement learning-based traffic signal control strategy that uses pressure as the reward signal. The core idea is to dynamically optimize traffic signals by training a deep neural network.
- CoLight [7]: A traffic signal control strategy based on Graph Neural Networks (GNN) [31]. It aims to optimize signal control by modeling the interactions between intersections in an urban traffic network.
- HiLight [9]:A hierarchical reinforcement learning-based traffic signal control strategy. This method decomposes the traffic signal control problem into multiple layers, with each layer responsible for different decision dimensions, such as policy selection and timing optimization.
- InitLight [10]: A pre-training method for traffic signal control using Adversarial Inverse Reinforcement Learning. It pre-trains a single agent model on multiple single-intersection scenarios, learning a reward function that generalizes well.



(a) Grid 4×4 (b) Grid 6×6

(c) Dongfeng Sub-district, Jinan, China (d) Gudang Sub-district, Hangzhou, China

Figure 2: (a) and (b) are Synthetic Map with 16 and 36 intersections. (c) and (d) are Real-world Maps with 12 and 16 intersections. The green areas on the maps are the ones we use. The intersections within the red circles will be used in the experiment.

4.1.3 Evaluation Metrics. Consistent with existing research [9], we use widely applied evaluation metrics in the field of traffic signal control to measure the performance of different traffic signal control

Table 1: The comparison results of average travel time, including the mean and standard deviation (in parentheses). Best results in boldface, and the second best results underlined.

Method	*Grid4×4*	*Grid6×6*	*Jinan*	*Hangzhou*
Fixedtime	1165.69 (0.00)	1086.45 (0.00)	1054.85 (0.00)	1149.88 (0.00)
SOTL	1304.03 (0.00)	1003.74 (0.00)	1075.90 (0.00)	1110.66 (0.00)
MaxPressure	1066.43 (0.00)	563.27 (0.00)	917.56 (0.00)	864.53 (0.00)
PressLight	300.62 (54.57)	287.11 (1.79)	325.34 (7.94)	345.91 (1.26)
CoLight	328.57 (1.26)	294.97 (0.95)	368.23 (13.43)	337.46 (1.38)
HiLight	267.61 (4.28)	389.69 (17.82)	296.74 (2.75)	343.68 (2.69)
InitLight	270.92 (30.64)	**242.62 (4.05)**	301.37 (3.17)	336.19 (2.86)
FGLight	**259.73 (0.73)**	251.91 (2.08)	**290.50 (1.13)**	**330.86 (2.11)**

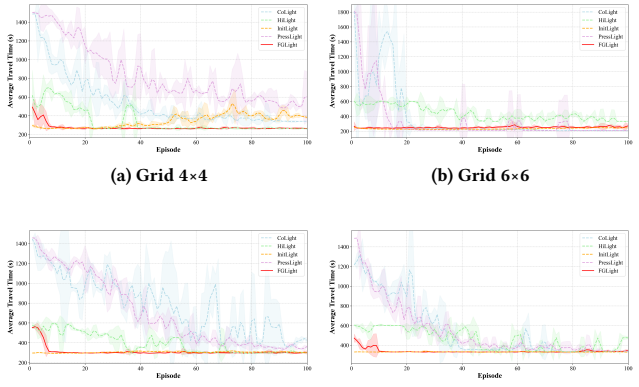
Table 2: The comparison results of Throughput, including the mean and standard deviation (in parentheses). Best results in boldface, and the second best results underlined.

Method	*Grid4×4*	*Grid6×6*	*Jinan*	*Hangzhou*
Fixedtime	6090 (0.0)	3390 (0.0)	4465 (0.0)	2359 (0.0)
SOTL	3831 (0.0)	3476 (0.0)	4631 (0.0)	2267 (0.0)
MaxPressure	6897 (0.0)	3576 (0.0)	5381 (0.0)	2695 (0.0)
PressLight	10971 (84.7)	4343 (1.3)	5517 (18.9)	2733 (4.4)
CoLight	10529 (4.5)	4330 (0.4)	5673 (54.6)	2736 (1.7)
HiLight	10958 (32.0)	2729 (21.0)	5531 (7.8)	2689 (10.8)
InitLight	9810 (52.9)	4345 (17.1)	**5788 (11.7)**	2618 (8.8)
FGLight	**11054 (4.0)**	**4403 (5.0)**	5764 (3.5)	**2747 (0.7)**

methods. These metrics are crucial for reflecting the efficiency of the traffic system. The two primary evaluation metrics employed are the average travel time of all vehicles and the throughput of the road network. The average travel time refers to the duration required for vehicles to travel from their origin to their destination, reflecting the impact of traffic signal control on the efficiency of vehicle movement. The throughput of the road network represents the total number of vehicles that pass through the traffic network within a given period, serving as a measure of the traffic signal control system's capacity to optimize overall traffic flow.

4.2 Comparison Experiment Results

In this section, FGLight is compared with state-of-the-art baselines across various traffic datasets. To ensure comparability, identical experimental settings are maintained across all methods. For reinforcement learning algorithms, each experiment is replicated five times on every dataset to assess algorithmic stability and performance consistency.



(a) Grid 4×4 (b) Grid 6×6

(c) Dongfeng Sub-district,Jinan, China (d) Gudang Sub-district,Hangzhou, China

Figure 3: The convergence of CoLight, HiLight , InitLight , PressLight and FGLight.

4.2.1 Overall Analysis. Tables 1 and 2 present the performance comparison results of FGLight against seven other baseline methods in terms of average travel time and throughput metrics, respectively. The experimental results demonstrate that on synthetic datasets, FGLight achieves an average improvement of 74.74% in average travel time and 61.84% in throughput compared to three traditional methods (Fixedtime, SOTL, and MaxPressure). On real-world datasets, FGLight exhibits an average improvement of 69.82% in average travel time relative and 15.94% in throughput to the three traditional methods. The performance difference observed may be attributed to the nature of the compared methods. Traditional traffic control approaches typically rely on pre-configured schemes and fixed assumptions, potentially limiting their adaptability to real-time traffic variations. In contrast, FGLight's design enables dynamic adjustments to current traffic conditions, which may contribute to its observed performance in the tested scenarios.

We further compared FGLight with four advanced MARL-based TSC methods (PressLight, CoLight, HiLight, and InitLight). As evident from Table 1, FGLight demonstrates notable performance differences compared to these methods. On synthetic datasets, FGLight achieves an average improvement of 14.03% over the other four methods. Moreover, on real-world datasets, FGLight outperforms CoLight by an average of 11.95%, PressLight by 7.43%, HiLight by 2.98%, and InitLight by 2.54%, which show the effectiveness of proposed FGLight.

It is noteworthy that FGLight performs better in both average travel time and throughput metrics for the Grid4×4 and Hangzhou scenarios, compared with comparison methods. In the Jinan scenario, FGLight achieves the lowest average travel time average travel time, albeit with slightly lower throughput than InitLight, indicating its focus on overall efficiency and fairness in traffic flow management. For the Grid6×6 scenario, while FGLight's average travel time is marginally higher than InitLight's, it achieves comparable performance without extensive pre-training, suggesting improved learning efficiency and sample efficiency. These results indicate FGLight's robustness across diverse traffic scenarios and its potential for efficient real-world implementation.

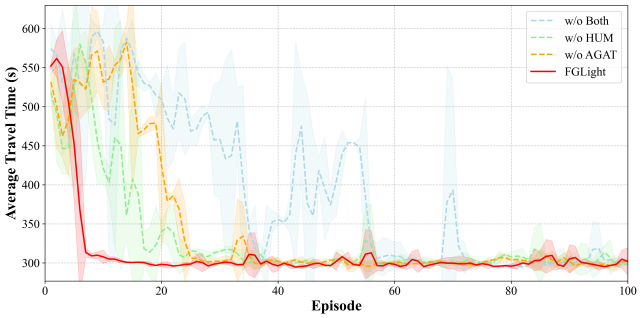


Figure 4: The convergence of ablating AGAT, HUM, and both modules in grid4x4

Table 3: The comparison results of ablating studies in grid4x4

Method	*average travel time*	*convergence round*
w/o Both	347.67	79
w/o AGAT	328.52	41
w/o HUM	325.94	27
FGLight	**306.86**	**12**

4.2.2 Convergence Analysis. Figure 3 illustrates the convergence rates of FGLight compared to PressLight, CoLight, HiLight, and InitLight during the training process, using average vehicle travel time as the evaluation metric. The results indicate that while InitLight exhibits a convergence trend similar to FGLight, FGLight achieves comparable performance without pre-training. FGLight reaches the target objective in 12 iterations after the first iteration, which is 67 iterations fewer than the next fastest method. At the conclusion of training, FGLight achieves an average travel time of 290.5s, representing a 3.61% reduction compared to the second-best performing method in this experiment. These observations suggest that the FGLight model offer advantages in learning decision-making strategies, potentially leading to reduced overall average travel time while maintaining rapid convergence.

4.3 Ablation Studies

To investigate the impact of each module on the overall performance of FGLight, we conduct ablation Studies, including three configurations: (1) without (w/o) the Adaptive Graph Attention Networks (AGAT) module and the Hysteretic Update Mechanism (HUM) module both, (2) w/o the AGAT module only, and (3) w/o the HUM module only.

Table 3 and Figure 4 illustrate the impact of individual components on FGLight's performance. The results indicate that without the HUM, there is a 7.06% increase in average travel time, while without the AGAT, there is a 241.67% extension in convergence time. The configuration without both AGAT and HUM showed the largest performance difference, with average travel time increasing by 13.30% and convergence time extending by 558.33% compared to the full FGLight model. The observed performance differences when these components are removed suggest their potential importance in the model's functionality. These ablation study results comprehensively demonstrate that both AGAT and HUM play indispensable roles in FGLight.

5 CONCLUSION

In this paper, we have introduced FGLight, a hierarchical reinforcement learning algorithm leveraging AGAT to address the challenge of multi-intersection traffic signal control in complex road networks. FGLight employs an adaptive weighting mechanism for neighbor communication through AGAT, ensuring that agents effectively prioritize relevant information from adjacent intersections. Additionally, the integration of a hysteretic update mechanism enhances the stability and accuracy of policy updates, facilitating both short-term and long-term optimization objectives. The experimental results demonstrate that FGLight significantly outperforms existing methods across various traffic performance metrics, including average travel time and throughput, while also achieving faster convergence. These results validate the efficacy and superiority of FGLight in managing complex and dynamic traffic networks, highlighting its potential for real-world deployment.

In the future, we aim to develop an adaptive mechanism for optimizing the selection of neighboring agents in the AGAT, improving traffic signal control performance through more precise communication. Additionally, we plan to introduce an intermediate decision-making layer to strengthen the algorithm's hierarchy, enhancing its scalability and adaptability to larger and more dynamic traffic networks.

ACKNOWLEDGMENTS

This research was funded by National Natural Science Foundation of China (No.62406251), Basic Research Programs of Taicang, 2022 (TC2022JC14), Natural Science Foundation of Guang-dong (No.2024A1515030024), Shenzhen Foundational Research Funding Under Grant (No.20220818102414030).

REFERENCES

[1] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2019. A survey on traffic signal control methods. *arXiv preprint arXiv:1904.08117* (2019).
[2] Kok-Lim Alvin Yau, Junaid Qadir, Hooi Ling Khoo, Mee Hong Ling, and Peter Komisarczuk. 2017. A survey on reinforcement learning models and algorithms for traffic signal control. *ACM Computing Surveys (CSUR)* 50, 3 (2017), 1–38.
[3] Peter Koonce et al. 2008. *Traffic signal timing manual.* Technical Report. United States. Federal Highway Administration.
[4] John Little, Mark Kelson, and Nathan Gartner. 1981. MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. *Transportation Research Record Journal of the Transportation Research Board* 795 (12 1981), 40–46.
[5] Pravin Varaiya. 2013. The max-pressure controller for arbitrary networks of signalized intersections. In *Advances in dynamic network modeling in complex transportation systems.* Springer, 27–66.
[6] Roger P Roess, Elena S Prassas, and William R McShane. 2004. *Traffic engineering.* Pearson/Prentice Hall.
[7] Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. 2019. Colight: Learning network-level cooperation for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management.* 1913–1922.
[8] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. 2020. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3414–3421.
[9] Bingyu Xu, Yaowei Wang, Zhaozhi Wang, Huizhu Jia, and Zongqing Lu. 2021. Hierarchically and cooperatively learning traffic signal control. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 669–677.
[10] Yutong Ye, Yingbo Zhou, Jiepin Ding, Ting Wang, Mingsong Chen, and Xiang Lian. 2023. InitLight: initial model generation for traffic signal control using adversarial inverse reinforcement learning. In *IJCAI.*

[11] Haipeng Zhang, Zhiwen Wang, and Na Li. 2024. MATLight: Traffic Signal Coordinated Control Algorithm based on Heterogeneous-Agent Mirror Learning with Transformer. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. 2582–2584.

[12] PB Hunt, DI Robertson, RD Bretherton, and M Cr Royle. 1982. The SCOOT on-line traffic signal optimisation technique. *Traffic Engineering & Control* 23, 4 (1982).

[13] PR Lowrie. 1990. Scats, sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic. (1990).

[14] Seung-Bae Cools, Carlos Gershenson, and Bart D'Hooghe. 2013. Self-organizing traffic lights: A realistic simulation. *Advances in applied self-organizing systems* (2013), 45–55.

[15] Li Li, Yisheng Lv, and Fei-Yue Wang. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3, 3 (2016), 247–254.

[16] Xiaoyuan Liang, Xunsheng Du, Guiling Wang, and Zhu Han. 2018. Deep reinforcement learning for traffic light control in vehicular networks. *arXiv preprint arXiv:1803.11115* (2018).

[17] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2496–2505.

[18] Marco Wiering, Jelle Van Veenen, Jilles Vreeken, and Arne Koopman. 2004. Intelligent traffic light control. *Institute of Information and Computing Sciences. Utrecht University* (2004).

[19] Marco A Wiering et al. 2000. Multi-agent reinforcement learning for traffic light control. In *Machine Learning: Proceedings of the Seventeenth International Conference (ICML'2000)*. 1151–1158.

[20] Tianshu Chu, Shuhui Qu, and Jie Wang. 2016. Large-scale multi-agent reinforcement learning using image-based state representation. In *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 7592–7597.

[21] HM Abdul Aziz, Feng Zhu, and Satish V Ukkusuri. 2018. Learning-based traffic signal control algorithms with neighborhood information sharing: An application for sustainable mobility. *Journal of Intelligent Transportation Systems* 22, 1 (2018), 40–52.

[22] Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, Hua Wei, Huichu Zhang, Yong Li, Kai Xu, and Zhenhui Li. 2019. Learning phase competition for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1963–1972.

[23] LA Prashanth and Shalabh Bhatnagar. 2010. Reinforcement learning with function approximation for traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (2010), 412–421.

[24] Xiaonan Klingbeil, Marius Wegener, Haibo Zhou, Florian Herrmann, and Jakob Andert. 2023. Centralized model-predictive cooperative and adaptive cruise control of automated vehicle platoons in urban traffic environments. *IET Intelligent Transport Systems* 17, 11 (2023), 2154–2170.

[25] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. 2019. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE transactions on intelligent transportation systems* 21, 3 (2019), 1086–1095.

[26] Jinming Ma and Feng Wu. 2020. Feudal multi-agent deep reinforcement learning for traffic signal control. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems (AAMAS)*. 816–824.

[27] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. 2019. Presslight: Learning max pressure control to coordinate traffic signals in arterial network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 1290–1298.

[28] Frans A Oliehoek, Christopher Amato, et al. 2016. *A concise introduction to decentralized POMDPs*. Vol. 1. Springer.

[29] Huichu Zhang, Siyuan Feng, Chang Liu, Yaoyao Ding, Yichen Zhu, Zihan Zhou, Weinan Zhang, Yong Yu, Haiming Jin, and Zhenhui Li. 2019. Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario. In *The world wide web conference*. 3620–3624.

[30] Hao Mei, Xiaoliang Lei, Longchao Da, Bin Shi, and Hua Wei. 2023. Libsignal: an open library for traffic signal control. *Machine Learning* (2023), 1–37.

[31] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE transactions on neural networks* 20, 1 (2008), 61–80.