Traffic Signal Timing Optimization at Intersections Using an Improved Q-Learning Algorithm

Yuanshuai LAN[1]*, Chuan LI[2], Min LIAO[1], Ting GUO[1]

[1] Geely University, Chengjian Avenue, Eastern New Area, Chengdu, 317000, China
448916030@qq.com (**Corresponding author*)

[2] Zhongke Ruihai (Dalian) Intelligent Technology Research Institute Co., Ltd., 1 Huoju Road, Ganjingzi District, Liaoning Province, 116085, China
m21z50c71@163.com

Abstract: In order to address the limitations of the conventional traffic signal control methods with regard to adapting to dynamic traffic scenarios and the challenges associated with multi-objective collaborative optimization, this study proposes a Firefly-guided Hybrid Deep Q-Network (FH-DQN) algorithm for optimizing the intersection signal timing. The proposed framework synergistically integrates firefly algorithm-based swarm intelligence with deep reinforcement learning, employing a brightness-driven hierarchical exploration strategy for enhancing the action selection efficiency. A multi-objective dynamic reward function incorporating vehicle delay, queue length, carbon emissions, and traffic throughput is developed in order to achieve a balanced optimization of traffic efficiency and environmental sustainability. The algorithm architecture includes the following innovative components: a dual-network collaborative structure that enhances the vehicle responsiveness to sudden congestion at individual intersections, an adaptive brightness update rule derived from the firefly algorithm-based optimization principles and the multi-objective dynamic reward function which achieves the dynamic adjustment of the intersection traffic signals. The extensive experiments conducted on the Simulation of Urban Mobility (SUMO) platform demonstrate that the FH-DQN algorithm achieves a superior performance in comparison with the fixed-time and conventional DQN approaches in typical cross-intersection scenarios. Specifically, the average queue waiting time for vehicles when applying the proposed method is reduced by 26.09% in comparison with the DQN-based approach. The ablation experiment confirms the individual contributions of the dynamic reward function and global network components to enhancing the overall performance of the FH-DQN algorithm. To sum up, this study provides a novel framework for collaborative traffic signal optimization in complex urban road networks, significantly improving both the adaptive capability and multi-objective optimization performance of intelligent transportation systems. Future work could focus on adapting the algorithm to metropolitan-scale networks and on integrating multimodal traffic data for enhancing the operational robustness of the proposed model.

Keywords: Intersection Traffic Lights Timing, Q-Learning Algorithm, Reinforcement Learning, Dynamic Reward Function, Firefly Algorithm, Regional Cooperative Control.

1. Introduction

In recent years, many scholars have put forward a variety of traffic signal optimization methods, For example, the Convergent Deep Q-network algorithm (C-DKN) is used for solving the divergence problem of the traditional DQN (Deep Q-network) algorithm in reinforcement learning tasks. By modifying the loss function, the network model ensures that the loss will not increase when the target network is updated, thus ensuring the convergence of the model. However, the model cannot completely converge to the optimal solution. The adaptive traffic signal control algorithm of the vehicular ad hoc network (VANET) (Pandit et al., 2013) achieves traffic signal control by transforming the traffic signal control problem into a job scheduling problem. The strategy of Platooning is put forward to enhance the flexibility and response speed of traffic scheduling. However, this method is too idealistic, relying on the VANET equipment on the vehicle, and only for a single intersection, and the real environmental factors are not considered in the simulation. An improved firefly algorithm (FA1->3) proposed by Ghasemi et al. (2022) by introducing three firefly movement modes (approaching the best solution, staying away from the bad solution and balancing randomness) improves the global search ability and can effectively avoid premature convergence. It was verified in six practical engineering optimization problems (spring design, pressure vessel design etc.), however it has not been verified in the field of transportation. From the above-mentioned research, it can be found that although some improvements have been made on DQN and the firefly algorithm, and they are also used for traffic intersections, the existing research mainly focuses on the optimization of a single intersection, which does not include any adjacent intersections (Eom & Kim, 2020) and has not been specifically applied to the research on traffic volume and throughput in traffic intersections. It can be stated that although some improvements were made to the DQN and firefly algorithms, and they are also used for traffic intersections, the existing research mainly focuses on the optimization of single intersections, which has nothing to do with the adjacent intersections, and has not been specifically applied to the

https://doi.org/10.24846/v35i1y202603

research on traffic volume and throughput in traffic intersections.

In order to solve the above problems, a firefly-guided hybrid depth Q-network (FH-DQN) algorithm is proposed. The core innovations are as follows:

1. The integration of swarm intelligence and deep reinforcement learning: The brightness-based attraction mechanism simulated ofby the firefly algorithm is embedded in the action selection process of DQN, and its exploration ability is enhanced through random disturbance. This effectively solves the inherent problem of slow convergence in traditional Q-learning;

2. A dynamic multi-objective return function: By integrating four key traffic parameters, such as vehicle delay, queue length, carbon emission and intersection traffic throughput, an adaptive weight adjustment strategy was designed to dynamically balance the objectives related to efficiency and environmental sustainability;

3. Hierarchical experience playback and a dual-network architecture: A global-local experience pool framework is developed to store intersection coordination data and single intersection status information, respectively. Combined with the dual-network parameter synchronization mechanism, this architecture significantly enhances the robustness of regional road network control.

Based on the above improvements, the paper propose a cross-optimization strategy combining the firefly algorithm and a deep Q-network, which leads to a multi-objective coordinated control and the collaborative optimization of regional traffic.

The remainder of this paper is structured as follows. Section 2 reviews the three technical pillars of this work – the fundamentals of Deep Q-Networks and their challenges in traffic control, the core mechanisms of the Firefly Algorithm, and how vehicle-detection technologies support traffic-state perception - thus providing the theoretical background for subsequent innovations. Section 3 presents theproposed Firefly-guided Hybrid Deep Q-Network (FH-DQN) algorithm. Specifically, it details a brightness-based hierarchical exploration strategy inspired by fireflies, a dynamic multi-objective reward function, a layered experience-replay mechanism, and a cooperative dual-network architecture. Further on, Section 4 describes the SUMO-based simulations in which the FH-DQN algorithm is compared against traditional fixed-timing plans, the baseline DQN, and several algorithmic variants based on both performance and ablation studies, demonstrating its superiority in reducing vehicle delay, queue length and carbon emissions while improving traffic throughput. Finally, Section 5 summarizes the work, highlighting its main innovations and experimental conclusions, and it outlines possible future research directions.

2. Related Technologies

2.1 Deep Q-network

2.1.1 Basic Theory and Core Mechanism

The Deep Q-Network (DQN)- an algorithm integrating deep learning with reinforcement learning, addresses Markov Decision Process (MDP) problems in discrete action spaces to obtain a data-driven scheduling framework (Liu et al., 2024). By modeling the flexible job scheduling problemas a Markov decision process, the rewards are centered around maximizing the overall completion time, delays, and energy consumption, enabling the agent to output process priorities at a millisecond level. This verifies the efficiency and scalability of deep reinforcement learning in real-time production scheduling. If the concept of 'state-action value' from Q-Learning is embedded into the ant colony optimization framework, a new method for automated driving path planning is obtained (Zhao et al., 2024). This algorithm allows ants to consider not only pheromones and heuristic distances when choosing their next step, but it also incorporates Q-values as a third decision component, enabling ants to gradually learn to avoid congested and high-risk areas. Additionally, a dynamic evaporation coefficient is introduced to prevent good paths from being forgotten too early.

By employing a deep neural network for approximating the Q-value function, DQN effectively handles high-dimensional state spaces, thereby overcoming the limitations of conventional Q-learning in practical applications. This integration of deep neural networks with Q-learning fundamentally resolves the curse of dimensionality inherent to traditional reinforcement learning when dealing with high-dimensional state spaces (Wu et al., 2019). The core operational mechanisms comprise:

1. Experience Replay: Stores the agent-environment interaction experiences in a memory buffer and utilizes random sampling to break the temporal correlations among sequential data, thereby effectively reducing the variance of training updates (Downs, 2004);

2. Target Network: An independent target network is introduced to calculate Q-value targets, mitigating the overestimation bias caused by bootstrap approximation. The parameter synchronization interval C (typically configured as including 103 steps) enhances the learning stability through delayed network updates;

3. Reward function: DQN updates the Q-value function through the Bellman equation of Q-learning. Reward is the key parameter for calculating the target Q-value, but the reward function depends on artificial design, which leads to a low exploration efficiency, a significantly prolonged training time and even to the inability to converge.

While DQN demonstrates a superior performance in gaming domains (Nguyen et al., 2020), its implementation in dynamic traffic control scenarios encounters three fundamental challenges:

1. Suboptimal exploration efficiency: The conventional ε-greedy strategy (with a fixed $\varepsilon = 0.1$) exhibits a slow convergence when handling burst congestion scenarios;

2. Multi-objective optimization conflicts: Simplified reward structures (e.g. pure delay minimization) tend to produce suboptimal control policies due to competing optimization goals;

3. Deficient coordination mechanisms: Isolated intersection-level optimization frequently results in local optimal but globally suboptimal control scenarios across road networks.

2.1.2 Related Research

The DynamicLight (Zhang et al., 2022) network involves a two-stage control framework, which achieves the flexible optimization of traffic signal timing through a two-stage decision-making mechanism (phase selection and duration adjustment). In the design of the reward mechanism, combining Queue Length and traffic Pressure, DQN is driven by a negative reward signal to optimize the phase duration, so as to control traffic signals. Based on this network, this paper proposes a dynamic multi-objective reward mechanism in order to achieve the dynamic adjustment of traffic signals. In the field of robotics, the DQN algorithm is used for achieving the cooperative control of transport robots. It has significant cross-domain transfer capabilities in handling large-scale multi-agent cooperation, real-time conflict resolution and distributed resource optimization (Li & Cheng, 2025), and provides a reliable methodological support for traffic signal timing and vehicle cooperative control at urban intersections.

2.2 Swarm Intelligence Algorithm

The Firefly Algorithm (FA)fundamentally operates by simulating the bioluminescence-based attraction mechanism observed in firefly swarms. In this metaheuristic optimization framework, each artificial firefly represents a candidate solution, with its luminous intensity corresponding to the objective function value. A higher luminance indicates a superior solution quality (Sankalap & Singh, 2013). Fireflies exhibit phototaxis towards the neighboring individuals with a greater luminance, enabling position updates based on local optimum information.

When searching for the optimal layer arrangement of composite laminated plates in a discrete-continuous hybrid design space, the Firefly Algorithm can be used as an optimization strategy, allowing it to bear higher loads even after buckling (Koide et al., 2024). To that, this algorithm framework, when combined with K-means clustering, has been successfully applied to the precise segmentation of brain tumors in the field of medical imaging (Capor Hrosik et al., 2019), which proves its excellent optimization capabilities and strong robustness in handling high-dimensional, non-convex clustering problems. For solving the Capacitated Vehicle Routing Problem (CVRP), a Cooperative Firefly Algorithm (CFA) is proposed, which first uses ‘double-chain encoding + greedy correction’ for discretization and then introduces a collaborative mechanism for information exchange and sharing, thereby balancing exploration and convergence. These all indicate the wide application of the Firefly Algorithm (Altabeeb et al., 2021). However, during the optimization process, the following core rules will be satisfied:

1. All the fireflies interact indiscriminately within the population, where any individual exerts an attraction force on all others;

2. The attraction force of fireflies is directly proportional to their brightness. The fireflies with a low brightness will move to

the individuals with a high brightness. If the brightness of two fireflies is equal, the fireflies will move randomly;

3. Individual brightness li is directly mapped to the optimization problem's objective function f(xi), such that $Ii \propto f(xi)$, establishing an explicit fitness-intensity relationship.

The firefly algorithm is modified, and the priority sampling for fireflies is adopted in order to simulate the three previously-mentioned firefly movement modes, so as to accelerate model convergence, quickly select the right action and avoid indiscriminate attraction during exploration.

2.3 Vehicle Inspection

GHAFNet (Global-context Hierarchical Attention Fusion Network) (Li, Qu & Wang, 2024) is a deep learning network focused on traffic object detection. It is a global-context feature fusion network in combination with a hierarchical mixed-attention module" / "It involves a global-context feature fusion network in combination with a hierarchical mixed-attention module, where the global context feature fusion network is responsible for constructing global contextual semantic information and the hierarchical mixed-attention module refines feature expression through a novel fusion strategy, ultimately generating a feature map that contains multi-dimensional weight information related to spatial and channel attributes. This allows it to adaptively emphasize the key features while suppressing irrelevant background noise, significantly enhancing the sensitivity and accuracy of detecting small-sized vehicles, and enabling the precise localization and bounding box regression for vehicle targets in traffic images. Although this method is used for vehicle detection, it can also be applied in the traffic signal domain to focus on traffic flow data at single intersections and their adjoining intersections.

Based on this local and global method, this paper proposes a dual-network collaborative architecture allowing the combined use of local and global information in the field of traffic signals.

3. Methods

In order to solve the intersection traffic signal timing problem of the traffic signal control system, this study proposes a novel adaptive exploration strategy inspired by the firefly algorithm. The related flow chart is shown in Figure 1, and its core operating mechanism is as follows.

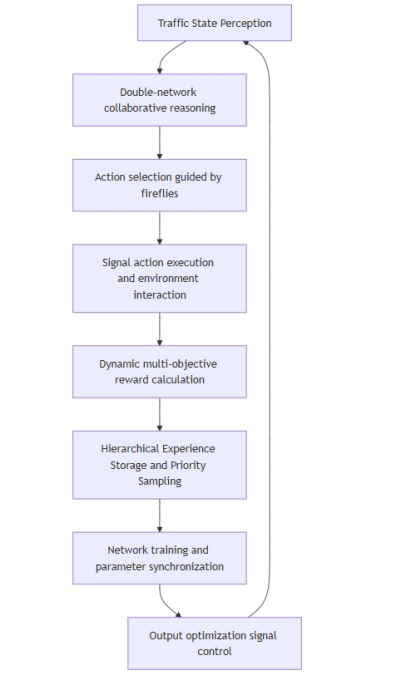

Figure 1. Algorithm implementation process

3.1 Hierarchical Exploration Strategy Based on the Firefly Algorithm

During peak hours when intersection traffic flow exhibits surge patterns ($Q_p > Q_{th}$), conventional ε-greedy strategies frequently converge to local optima. (e.g. excessively prolonging the green light phases when approaching arterial roads).The proposed brightness-based attraction adaptive exploration mechanism addresses this through the following:

3.1.1 Phase-adaptive Luminance Adjustment

High-flow phase prioritization: When the queue length (Q_p) in stage (p) exceeds the threshold (Q_{th}), the luminance value is enhanced logarithmically (Qiu et al., 2019):

$$B'_{\alpha\_t} = B_{\alpha_t} \times [1 + \log(\frac{Q_p}{Q_{th}})] \quad (1)$$

where:

- Q_p: Real-time queue length (in meters) of phase p;
- Q_{th}: Predefined threshold (e.g., $Q_{th} = 30$ meters);
- $\log(\frac{Q_p}{Q_{th}})$: Natural logarithm function to smooth the enhancement magnitude;

- α_t: Candidate action (green light phase extension);
- B_{α_t}: Brightness value before adjustment;
- B'_{α_t}: Adjusted brightness value.

Physical Meaning: Dynamically amplifying the luminance of high-flow phases prioritizes exploration resources, accelerating traffic congestion mitigation.

Burst congestion response: If abrupt queue changes ($\Delta Q>30\%$) are detected in the adjacent intersections (Zhang et al., 2020):

$$\beta'_0 = \beta_0 \times 1.2 (\beta_0^{max} = 2.4) \quad (2)$$

where:

ΔQ: Queue length variation rate (%) in the adjacent intersections;

β_0: Baseline attraction coefficient, initialized as $\beta_0 = 1.0$, with an upper bound $\beta_0^{max} = 2.4$.

Physical Meaning: Enhancing β_0 facilitates a rapid escape from suboptimal policies for handling burst congestion.

3.1.2 Luminance-driven Action Selection

The luminance value of the candidate action α_t is determined by the Q-value prediction and stochastic perturbation:

$$A = Q(s_t, \alpha_t) \quad (3)$$

$$Q = \alpha^{(k)} \cdot \epsilon_t (\epsilon_t \sim N(0, (\alpha^{(k)})^2)) \quad (4)$$

$$B(\alpha_t) = A * Q \quad (5)$$

where:

α_t: Candidate actions;

s_t: Current action;

$Q(s_t, \alpha_t)$: Action value predicted by the local Q-network;

$\alpha^{(k)}$: Perturbation intensity at iteration k, which can be expressed as:

$$\alpha^{(k)} = \alpha^{(0)} \times 0.95^{\frac{k}{100}} (\alpha^{(0)} = 0.3) \quad (6)$$

ϵ_t: Gaussian-distributed noise with variance controlled by $\alpha^{(k)}$.

Physical Meaning: The luminance integrates the Q-value exploitation and noise-driven exploration, with perturbation intensity decaying in order to balance exploration and convergence stability.

3.1.3 Attraction Dynamics

Low-luminance actions iteratively converge toward high-luminance candidates denoted by α_{best}:

$$\alpha_{t+1} = \alpha_t + \beta_0 e^{-\gamma r^2} (\alpha_{best} - \alpha_t) + \eta \epsilon_t \quad (7)$$

where:

$\gamma = 0.1$: Distance decay factor controlling the attenuation attraction rate with normalized distance r;

r: Normalized distance between the current actionand the optimal action α_{best};

$\eta_t = 0.1 \times 0.99t$: Annealed step-size factor to reduce perturbation magnitude progressively;

$\epsilon_t \sim U(-0.1, 0.1)$: Uniformly distributed noise to prevent premature convergence.

Physical Meaning:

The update rule comprises three components:

Directional Optimization: Movement toward the current optimal action α_{best};

Distance Attenuation: Attraction weakens with distance ($e^{-\gamma r^2}$);

Stochastic Perturbation: Noise injection enhances exploration diversity.

The above algorithm can be visualized as shown in Figure 2.

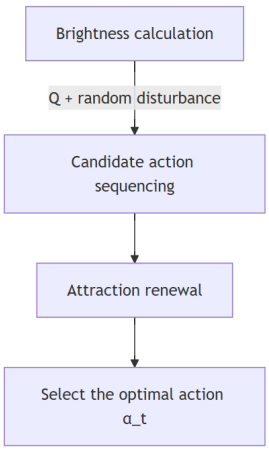

Figure 2. Action selection guided by the firefly algorithm

3.2 Dynamic Multi-Objective Reward Function

According to the real-time state of the road traffic and various complex environments, the adaptive reward mechanism is proposed, and the signal control strategy is dynamically adjusted according to the real-time traffic conditions. Through a real-time adjustment, the rapid response of different vehicles at different time moments can be met. The process is shown in Figure 3.

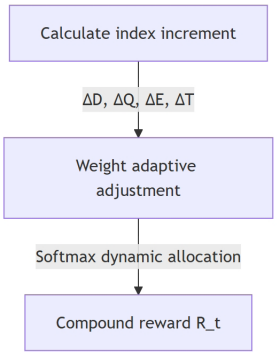

Figure 3. Dynamic multi-objective reward calculation

During peak hours, the optimization of queue lengths should be prioritized to ensure an efficient traffic clearance on main roads; during non-peak hours, idling emissions from stationary vehicles should be reduced to a minimum through adaptive signal maintenance strategies.

A composite reward function including four traffic indicators was designed, and its weight was dynamically adjusted according to the traffic state, with s_{t+1} representing the current moment, and representing the next moment:

$$R_t = \sum_{i=1}^{4} w_i(t) \cdot \Delta M_i(s_t, s_{t+1}) \quad (8)$$

The Core Metrics are as follows:

The reduction in the delay time (seconds), expressed as:

$$\Delta D = D(s_t) - D(s_{t+1}) \quad (9)$$

The change in queue length (meters), expressed as:

$$\Delta Q = Q(s_t) - Q(s_{t+1}) \quad (10)$$

The reduction in carbon emissions (grams/vehicle), expressed as:

$$\Delta E = E(s_t) - E(s_{t+1}) \quad (11)$$

The improvement in traffic throughput (vehicles/cycle), expressed as:

$$\Delta T = T(s_t) - T(s_{t+1}) \quad (12)$$

Weight adaptive mechanism:

The Softmax function is used to dynamically allocate weights:

$$w_i(t) = \frac{\exp(f_0(S_i(t)))}{\sum_{j=1}^{4} \exp(f_0(S_j(t)))} \quad (13)$$

where $S_i(t)$ represents the real-time importance of index I.

3.3 Layered Experience Playback and Firefly Priority Sampling

Based on the design concept for a multi-objective reinforcement learning constructor aimed at travel agency problems (clearly integrating local and global information) (Gobbi et al., 2024), this paper proposes a dual buffer hierarchical storage for empirical data:

3.3.1 Design of the Dual Buffer Hierarchical Storage for Empirical Data

Global experience pool (B_G):

- Captures the regional traffic wave propagation patterns through coordinated signal cycles. (Example: When congestion emerges at upstream intersection A, the downstream intersection B proactively stores the coordination experiences with 3 signal cycles in advance);
- Stores the cross-intersection coordination tuples $(S_{global}, \alpha, s', \gamma)$ with the spatial-temporal correlation weights;
- Identifies the traffic flow correlation matrices $C \in R^{N \times N}$ across N intersections.

Local experience pool (B_L):

- Enhances the emergency response capability through priority tagging. The experiences are flagged as high-priority if any phase exhibits queue length surges ($\Delta Q > 15$ vehicles) for 2 consecutive cycles;
- Maintains single-intersection state-action pairs $(S_{local}, \alpha, s', \gamma)$;
- Optimizes phase transition sequences via adaptive reward shaping.

3.3.2 Firefly Algorithm-base Prioritization of Experience Sampling

The Probability sampling for experience i is determined by its luminance value $B(\alpha_i)$:

$$P_{(i)} = \frac{\exp(f_0(B(\alpha_i)/\tau))}{\sum_j \exp(f_0(B(\alpha_j)/\tau))} \quad (14)$$

where $\tau \in [0.5,2.0]$ serves as the temperature parameter governing the exploration-exploitation balance during training iterations. $B(\alpha_i)$ is the brightness of action α_i.

After double-layer buffer storage, the global and local intersection information is captured, and the two pieces of information are processed by the firefly algorithm for priority experience sampling, and the priority processing method is obtained by deduction. The process is depicted in Figure 4.

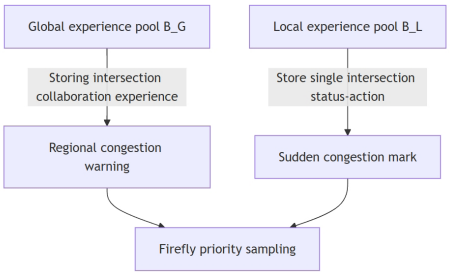

Figure 4. Dual buffer hierarchical storage for empirical data

3.4 Dual Network Collaboration Architecture

Through local and global networks, different information related to intersections is processed. The local network processes vehicle information for single intersections, while the global network processes vehicle information related to the adjacent intersections, driving local traffic from the global perspective (Eom & Kim, 2020). The two networks adjust and constrain each other, achieving a dynamic release of vehicles, that is the rapid passage of vehicles at intersections, as it can be seen in Figure 5.

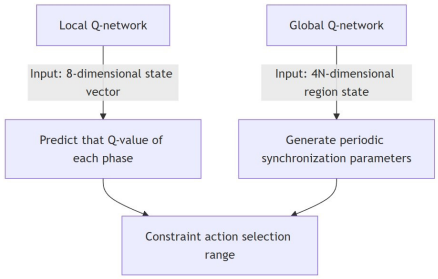

Figure 5. Double-network collaborative reasoning

3.4.1 Local Q-network (Q_L)

It processes real-time intersection states at a frequency of 10Hz.

Input: 8-dimensional feature vector (per-lane queue length $\in [0,1]^4$, delay $\in R^4$).

Architecture: 3-layer MLP with hidden dimensions 64→64→4, ReLU activation, and batch normalization.

3.4.2 Global Q-network (Q_G)

It coordinates the signal phase offsets across the adjacent intersections.

Input: 4N-dimensional regional state vector (N adjacent intersections × 4 features: traffic throughput, queue gradient, delay ratio, and phase status)

Output: Phase synchronization parameters $\phi \in [-0.5T, 0.5T]^N$ ($T = 120s$ base cycle).

3.4.3 Parameter Synchronization Mechanism

It implements periodic hard synchronization with an exponentially decaying interval:

$$\theta_G \leftarrow \theta_L \quad (15)$$

where θ represents the network parameters, maintaining policy stability through delayed target updates.

In comparison with the isolation control methods, it has achieved the following:

- A real-time response to local congestion patterns;
- The global propagation of traffic wave mitigation strategies;
- The balanced between exploration and exploitation through dual-network parameter decoupling.

4. Experiment

Simulation of Urban Mobility (SUMO) (Krajzewicz et al., 2002) is a microscopic, continuous traffic simulation software capable of accurately modeling urban traffic scenarios. It is equipped with a graphical user interface (GUI) supporting various input grid formats and configurable road network designs. Through the TraCI (Traffic Control Interface) module integrated in SUMO, the real-time interaction with the simulation platform can be achieved. This enables the traffic signal timing system to acquire

dynamic traffic status data and optimize the traffic flow management effectively.

4.1 Experimental Setup

The experimental setup includes the following:

- Fixed Time control: The phase is fixed at a 30-second rotation without dynamic adjustment;
- Benchmark DQN: Standard model, with a state space of 8 dimensions (queue length and delay), without hierarchical experience replay and firefly algorithm-based exploration;
- FH-DQN: This method integrates firefly algorithm-based exploration, a layered experience replay, and a dynamic reward function;
- FH-DQN FinedReward: Fixed-weight reward function for controlling the length of vehicle queues;
- FH-DQN NoGlobal: The global Q network was removed (only the local Q network was kept). This method is for controlling the road conditions of the entire area and its surroundings.

Taking the cross-shaped intersections illustrated in Figures 6 and 7 as case studies, the traffic signal control problem for all intersections within the target area is reformulated based on phase-specific vehicle movement patterns. This problem is further abstracted into a temporal decision-making framework where the right-of-way allocation for specific phases at each intersection must be dynamically determined across discrete time intervals. The abstracted problem is subsequently digitized to enable its algorithmic resolution and optimization.

Figure 6 shows the combination of turning lanes and straight lanes into a single lane. When the traffic light is red, the queue time for vehicles increases, leading to road congestion, which then causes a chain reaction on the adjacent roads, further exacerbating the congestion. Figure 7 depicts the separation of turning lanes and straight lanes as two distinct lanes, where each type of lane accommodates certain vehicles. This significantly reduces the congestion on every visible road, with no interference among them. It also reduces the congestion on the surrounding roads, thereby allowing for a more reasonable adjustment of traffic flow across the entire road network. At the red traffic light at the intersection, traffic signal timing will affect the congestion of the road. Therefore, by designing them separately, and making independent decisions for each state, the aim is to minimize the generalized multi-phase intersection to reduce vehicle delays. (Wu et al., 2020; Mohajerpoor et al., 2022).

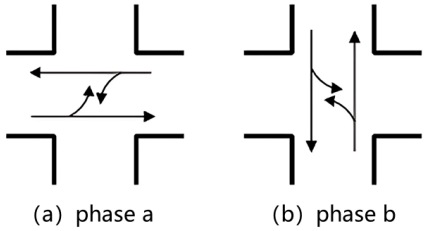

Figure 6. Cross Intersection 1

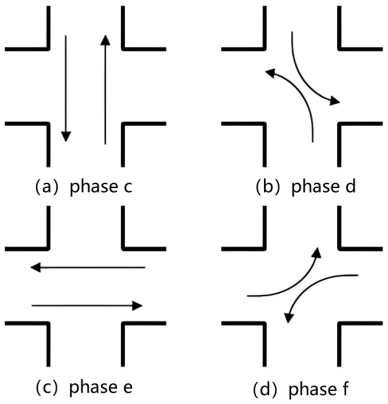

Figure 7. Cross Intersection 2

4.2 Comparative Experiments

Figure 8 depicts the training progression, where the horizontal axis represents the training epochs and the vertical axis represents the average queue length, with the lower values indicating an enhanced control efficacy.

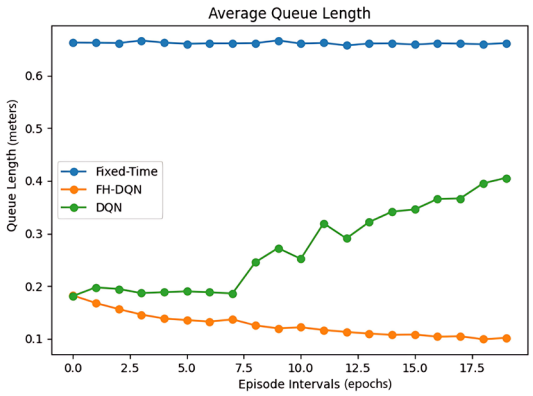

Figure 8. Vehicle Queue Length Results

In this context, the Fixed-Time model exhibits negligible variation in its horizontal trajectory, maintaining consistently elevated queue lengths. This pattern suggests the strategy's inability to adapt to dynamic traffic conditions.

The conventional Deep Q-Network (DQN) algorithm shows an initial decline followed by mid-training escalation before plateauing, as evidenced by its unstable performance in comparison with the FH-DQN algorithm in the later training stages.

The FH-DQN algorithm achieves a progressive optimization, decreasing from the initial higher values to stabilized levels around 0.1. This learning-driven adaptation significantly improves signal control through a continuous queue length reduction.

The comparative analysis demonstrates the FH-DQN algorithm's superior performance, achieving a progressive queue optimization and maintaining the lowest average queue length (0.1) during the later training phases. These results confirm its effectiveness in minimizing vehicular queues at intersections, establishing it as the most effective control methodology among the three evaluated strategies.

Figure 9 illustrates the comparative performance of three traffic signal control strategies (the Fixed-Time, FH-DQN, and DQN models) across two critical metrics: the average pollutant emissions and average traffic throughput.

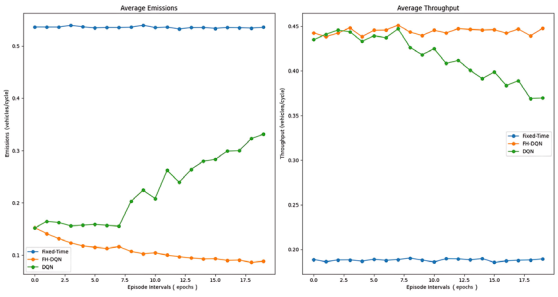

Figure 9. Vehicle throughput and carbon emissions (mean)

In this context, the Fixed-Time strategy exhibits persistently elevated carbon emission levels ($\mu = 0.6 \pm 0.15$ kg/veh) coupled with a suboptimal throughput ($\varphi = 0.2 \pm 1.7$ veh/min), indicating a limited adaptability to dynamic traffic demand fluctuations.

Further on, while the DQN-based approach achieves a good initial performance (Phase I: $t < 7.5$), it demonstrates a progressive decrease in control efficacy during sustained operation (Phase II: $t \geq 10$).

The FH-DQN algorithm manifests a characteristic dual convergence pattern: the carbon emission metrics follow a monotonic decreasing trajectory with asymptotic stabilization (Δemission = 0.05%), while the traffic throughput parameters display a complementary monotonic growth converging to steady-state values (Δthroughput = 0.4%). This bifurcated convergence behavior confirms the algorithm's capability to optimize traffic signal phasing through dynamic right-of-way allocation.

These quantitative comparisons conclusively establish the superiority of the FH-DQN algorithm in multi-objective signal control optimization, achieving a statistically significant pollutant emission reduction and a traffic throughput enhancement relative to the baseline strategies.

Using the TraCI interface module provided in SUMO to interact with the simulation platform online, the traffic signal timing system can obtain real-time traffic state information. By the traffic scene simulation, the normal traffic flow and peak traffic flow are set. The specific settings are as follows: at intersections, the switching time for traffic signal lights is fixed at 30 seconds; 10-20 vehicles are generated every minute, and the vehicles are distributed in a positive way to simulate the traffic conditions during off-peak hours and peak hours on weekdays. In order to ensure traffic safety, the maximum speed of the vehicle is set at 13.89 m/s and the maximum acceleration is 2m/s^2.

According to the set values, three algorithms, Fixed-Time, DQN and FH-DQN, are verified. Based on the 30-minute waiting time, the average waiting time for the vehicle passing through the intersection and the average vehicle queue length are recorded, and Table 1 is obtained.

Table 1. Average waiting time and queue length at intersections

Algorithm	Average waiting time (vehicles/cycle)	Average queue length (meters)
Fixed-Time	22.50	27.29
DQN	20.18	22.18
FH-DQN	18.06	20.17

The experimental results from Table 1 demonstrate that the FH-DQN model achieved a 10.50% reduction in the average waiting time in comparison with the DQN algorithm and a 19.73% reduction versus the Fixed-Time algorithm. In terms of traffic flow optimization, the average queue waiting time decreased by 26.09% and 9.06% respectively in comparison with the values obtained by the DQN and fixed-time algorithms. These quantitative improvements confirm the superiority of the FH-DQN algorithm with regard to the intersection traffic flow optimization, effectively decreasing the vehicle waiting times ($p < 0.01$) and enhancing the traffic throughput efficiency.

4.3 Ablation Experiment

The left image of Figure 10 shows the comparative performance of the FH-DQN-NoGlobal, FH-DQN-FixedReward and FH-DQN algorithms, highlighting the influence of the global network and of the fixed reward mechanism on the baseline model. The ordinate represents the queue length (the lower the length, the better). By comparing the FH-DQN-NoGlobal algorithm with the FH-DQN algorithm, it can be seen that both of them are effective in traffic optimization in the initial stage, but as the number of the training epochs increases, the optimization efficiency of FH-DQN improves continuously, while for FH-DQN-NoGlobal it begins to decrease. This is because with the increase in the number of epochs, the road conditions at the analysed intersection become more and more complicated, also affecting the adjacent intersections, a problem which the global network can solve. This network can learn the distance between the current intersection and the adjacent intersections, by controlling all the vehicles at these intersections, the traffic signals at the current intersection can be optimized in advance to reduce the congestion at the current intersection, thus reducing the length the vehicle queue at the intersection. The advanced layout of the Global network also influences the real-time change of the traffic signal at the current intersection, that is, the dynamic reward has an effect which achieves the dynamic adjustment of the intersection traffic lights through real-time adjustment, instead of always being fixed for 30 seconds. That is, FH-DQN-FixedReward has a certain decline in the initial stage when using the fixed reward mechanism, but then it gradually increases. However, it can be seen that the fixed reward mechanism can not effectively reduce the queue length. The right image of Figure 10 shows the vehicle throughput at intersections obtained by the three algorithms. The higher the numerical value, the higher the algorithm`s performance. From Figure 10, it can be seen that FH-DQN has a good effect, and its curve change is consistent with the curve change law in drawing. The lower the queue length, the smaller the vehicles throughput at the current intersection, and the faster the vehicles pass through the intersection, and the intersection can accommodate more vehicles. These two experiments show that the dynamic reward function and the dual network can effectively improve the traffic flow at intersections. The FH-DQN algorithm is the best in optimizing the traffic signal control, which can significantly reduce the queue length and improve traffic throughput, making it the most effective traffic signal control method among the three employed strategies.

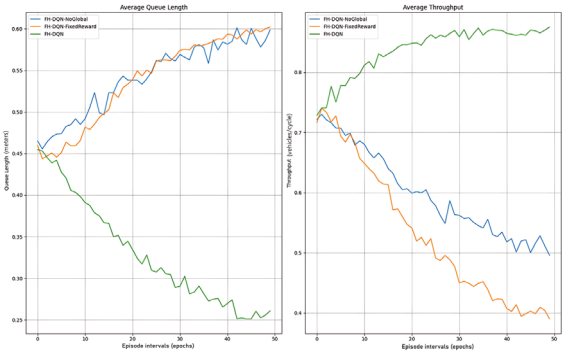

Figure 10. Results of the ablation experiment

5. Conclusion

The FH-DQN algorithm proposed in this paper significantly improves the convergence speed for traffic signal control by combining the firefly algorithm and DQN algorithm with the dynamic reward adjustment mechanism and hierarchical experience playback which can better adapt to a dynamic traffic demand, improve traffic efficiency and reduce congestion and pollutant emissions.

In this context, a brightness-driven hierarchical exploration strategy is proposed, and the brightness-based attraction mechanism of the firefly algorithm is embedded in the DQN-based action selection process. The priority of candidate actions is dynamically adjusted by the brightness value and the exploration efficiency is improved by updating the model parameters.

Further on, a dynamic composite reward function was designed and a weight adaptive mechanism based on Softmax was introduced in order to reach the dynamic balance between traffic efficiency and environmental protection objectives.

Then, a global-local double experience pool was built for storing cross-intersection collaborative information, and the ability to capture regional traffic flow association patterns was enhanced. The ablation experiment involving the global network and the dynamic composite function showed that these two modules could effectively speed up the traffic flow at intersections and reduce the waiting time for vehicles.

In comparison with the traditional model, the FH-DQN model reduced the average waiting time by 10.50%. In terms of traffic flow optimization, the average queue waiting time decreased by 26.09% in comparison with the value obtained by the DQN algorithm.

Future research could explore more complex traffic network topologies and introduce other swarm intelligence algorithms (Pham et al., 2021) to further enhance robustness:

- Complex Road Network Expansion: Exploring the collaborative control capability of FH-DQN in larger urban road networks (e.g. multi-level intersections and roundabouts). The topological relationships between road networks will be modeled using Graph Neural Networks (GNNs (Zhou et al., 2018)) to enhance the algorithm's global awareness of the dynamic traffic flows;
- Multi-Agent Collaborative Optimization: Introducing the Multi-Agent Reinforcement Learning (MARL) framework to investigate the communication and cooperation mechanisms among intersection agents (Goel et al., 2025). For instance, the graph-based reinforcement learning methods can optimize the distributed decision-making process for regional traffic signal timing by modeling the topological connections between agents, thereby addressing the communication bottlenecks of the traditional single-agent methods in road network coordination;
- Multimodal Data Fusion: Integrating multi-source traffic data collected by onboard sensors, cameras, and unmanned aerial vehicles (UAVs) to construct a high-precision traffic state representation model. This integration aims to further improve the algorithm's robustness and real-time performance;
- Cross-Domain Algorithm Fusion: Combining meta-heuristic optimization and PID control technologies for designing a hierarchical control architecture. This architecture will enable the seamless integration of traffic signal timing strategies at both the macro-optimization and micro-execution levels.

The above directions will promote the evolution of intelligent transportation systems toward a "perception-decision-execution" integrated framework, providing more efficient and environmentally friendly solutions for the urban traffic management.

REFERENCES

Altabeeb, A., Mohsen, A. M., Abualigah, L. et al. (2021) Solving capacitated vehicle routing problem using cooperative firefly algorithm. *Applied Soft Computing*. 108, Art. ID 107403. https://doi.org/10.1016/j.asoc.2021.107403.

Arora, S. & Singh, S. (2013) The Firefly Optimization Algorithm: Convergence Analysis and Parameter Selection. *International Journal of Computer Applications*. 69(3), 48-52. https://doi.org/10.5120/11826-7528.

Capor Hrosik, R. Tuba, E., Dolicanin, E. et al. (2019) Brain Image Segmentation Based on Firefly Algorithm Combined with K-means Clustering, *Studies in Informatics and Control*. 28(2), 167-176. https://doi.org/10.24846/v28i2y201905.

Downs, A. (2004) Still Stuck in Traffic: Coping With Peak-Hour Traffic Congestion. Washington, USA, Brookings Institution Press.

Eom, M. & Kim, B.-I. (2020) The traffic signal control problem for intersections: a review. *European Transport Research Review*. 12(1), Art. ID 50. https://doi.org/10.1186/s12544-020-00440-8.

Ghasemi, M., Mohammadi, S. K., Zare, M. et al. (2022) A new firefly algorithm with improved global exploration and convergence with application to engineering optimization. *Decision Analytics Journal*. 5, Art. ID 100125. https://doi.org/10.1016/j.dajour.2022.100125.

Gobbi, H. U., dos Santos, D. G. & Bazzan, A. L. C. (2023). Comparing reinforcement learning algorithms

for a trip building task: A multi-objective approach using non-local information. *Computer Science and Information Systems*. 21(1), 291-308. https://doi.org/10.2298/CSIS221210072G.

Goel, H., Omama, M., Chalaki, B. et al. (2025). R3DM: Enabling Role Discovery and Diversity Through Dynamics Models in Multi-agent Reinforcement Learning. To be published in *ICML 2025 Conference*. [Preprint] https://doi.org/10.48550/arXiv.2505.24265 [Accessed: 18th June 2025].

Koide, R. M., Herrera, P. H., Luersen, M. A. et al. (2024). Post-Buckling Optimisation of Composite Structures Using a Firefly Algorithm. *International Journal of Simulation Modelling,* 23(1), 5-16. https://doi.org/10.2507/IJSIMM23-1-654.

Krajzewicz D, Hertkorn G., Wagner P. et al. (2002) SUMO (Simulation of Urban MObility) - an open-source traffic simulation. In: *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002)*, *28-30 September 2002, Sharjah, UAE*. Amsterdam, Netherlands, SCS Europe. pp. 183-187.

Li, W. & Cheng, C. (2025) Intelligent Robot Cooperative Control Based on Distributed DQN Algorithm, *Studies in Informatics and Control*. 34(1), 57-73, 2025. https://doi.org/10.24846/v34i1y20250.

Liu, A. Y, Yue, D. Z., Chen, J. L. et al. (2024) Deep Learning for Intelligent Production Scheduling Optimization. *International Journal of Simulation Modelling.* 23(1). 172-183. https://doi.org/10.2507/IJSIMM23-1-CO4.

Mohajerpoor, R, Cai, C. & Ramezani, M. (2022) Optimal Traffic Signal Control of Isolated Oversaturated Intersections Using Predicted Demand. *IEEE Transactions on Intelligent Transportation Systems*. 24(1), 815-826. https://doi.org/10.1109/TITS.2022.3209606.

Nguyen, T. T., Nguyen, N. D. & Nahavandi, S. (2020) Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Transactions on Cybernetics*. 50(9), 3826-3839. https://doi.org/10.1109/TCYB.2020.2977374.

Pandit, K., Ghosal, D., Zhang, H. M. et al. (2013) Adaptive Traffic Signal Control With Vehicular Ad hoc Networks in*IEEE Transactions on Vehicular Technology*. 62(4), 1459-1471, https://doi.org/10.1109/TVT.2013.2241460.

Pham, Q.-V., Nguyen, D. C., Mirjalili, S. et al. (2021) Swarm intelligence for next-generation networks: Recent advances and applications. *Journal of Network and Computer Applications*. 191, Art. ID 103141. https://doi.org/10.1016/j.jnca.2021.103141.

Qiu, X., Liu, L., Chen, W. et al. (2019). Online Deep Reinforcement Learning for Computation Offloading in Blockchain-Empowered Mobile Edge Computing. *IEEE Transactions on Vehicular Technology*. 68(8), 8050-8062. https://doi.org/10.1109/TVT.2019.2924015.

Wu, C., Ju, B., Wu, Y et al. (2019) UAV Autonomous Target Search Based on Deep Reinforcement Learning in Complex Disaster Scene. *IEEE Access*. 7. 117227-117245, https://doi.org/10.1109/ACCESS.2019.2933002.

Wu, T., Zhou, P., Liu, K. et al. (2020) Multi-Agent Deep Reinforcement Learning for Urban Traffic Light Control in Vehicular Networks. *IEEE Transactions on Vehicular Technology*. 69(8), 8243-8256. https://doi.org/10.1109/TVT.2020.2997896.

Zhang, Z., Zou, Y., Zhang, X. et al. (2020) Green Light Optimal Speed Advisory System Designed for Electric Vehicles Considering Queuing Effect and Driver's Speed Tracking Error. *IEEE Access*, 8, 208796-208808, https://doi.org/10.1109/ACCESS.2020.3037105.

Zhang, L., Wu, Q., Shen, J. et al. (2022) DynamicLight: Dynamically Tuning Traffic Signal Duration with DRL. To be published in *CoRR 2022*. [Preprint] https://doi.org/10.48550/arXiv.2211.01025 [Accessed: 12nd May 2023].

Zhao, L., Li, F., Sun, D. et al. (2024) An improved ant colony algorithm based on Q-Learning for route planning of autonomous vehicle. *International Journal of Computers Communications & Control*. 19(3). https://doi.org/10.15837/ijccc.2024.3.5382.

Zhou, J., Cui, G., Hu, S. et al. (2018) Graph neural networks: A review of methods and applications. *AI Open*. 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001