

Joint control of traffic signal phase sequence and timing: a deep reinforcement learning method

Zhanbo Sun¹, Xinben Jia¹, Yiming Cai², Ang Ji^{1*} , Xia Lin³, Lin Liu³, Wenjun Wang³ and Yuexuan Tu³

¹ School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China

² SWJTU-Leeds Joint School, Southwest Jiaotong University, Chengdu 611756, China

³ Tencent Technology (Chengdu) Co., Ltd., Chengdu 610094, China

* Corresponding author, E-mail: ang.ji@swjtu.edu.cn

Abstract

Recent advances in artificial intelligence have opened up new possibilities for optimizing traffic operations. In this study, a novel deep reinforcement learning-based traffic signal control strategy is proposed to jointly optimize phase sequence and signal timing, allowing for more efficient and flexible signal control. A comparison between the proposed approach and two traditional methods, Webster and MaxPressure, is conducted to highlight the advantages of AI-empowered signal control. Different techniques in state representations and action selections are explored to enhance the performance of the DRL-based agent. Results from simulation experiments indicate that the 3DQN framework with prioritized experience replay outperforms other methods by at least a 7.56% queue length reduction. Additionally, the combined state representation of macroscopic feature-based and microscopic cell-based information presents a valuable enhancement for model performance. The ablation experiments demonstrate that considering microscopic information only leads to a 2.44% increase in queue length compared to the proposed method with combined micro-level and macro-level information. However, depending only on the macroscopic feature-based representation, it fails to converge during the training session. Furthermore, the proposed joint control method reduces queue length by 6.37% compared to the phase switching control method, while the single-agent model that optimizes the phase duration performs even worse. Hopefully, this study can offer references for future research in deep reinforcement learning-based traffic signal control schemes and reveal their potential to cope with more dynamic and complex scenarios.

Keywords: Traffic signal control; Deep reinforcement learning; Micro and macro state representation; Joint actions

Citation: Sun Z, Jia X, Cai Y, Ji A, Lin X, et al. 2025. Joint control of traffic signal phase sequence and timing: a deep reinforcement learning method. *Digital Transportation and Safety* 4(2): 118–126 <https://doi.org/10.48130/dts-0025-0008>

Introduction

Advances in automation and communication technology have reshaped intelligent transportation systems (ITS), revealing the potential to further enhance traffic control and operations. The emergence of artificial intelligence (AI) provides a novel insight to manage ITS with minimal human intervention. Traffic signal control (TSC) at intersections, as one of the crucial problems in urban traffic systems, is expected to combine with advanced AI technology to alleviate traffic congestion^[1–3]. Reinforcement learning (RL) or deep reinforcement learning (DRL) with neural networks has been deemed promising in adaptive traffic control, which could hopefully reduce total intersection delay.

Back to the inception of traffic signals, fixed-time control has been widely deployed in the real world. It operates by repeating a preset pattern made according to historical data. Since traffic demand is unpredictable and fluctuates in time^[4,5], fixed-time control may limit its adaptability to dynamic traffic situations. Actuated control was proposed to address this limitation^[6], which is usually applied for isolated intersections. By employing loop detectors at intersections, real-time traffic data can be collected and utilized to adjust signal control strategies. It changes signal strategies based on a set of predefined, static parameters such as unit extension time, and minimum and maximum green time^[7,8]. For instance, if the number of vehicles in a specific approach exceeds a predefined threshold, the signal will prioritize that approach to release more vehicles. Although actuated control demonstrates the ability to respond to varying traffic conditions, it faces challenges in maintaining effectiveness under highly saturated traffic volumes. In response to this

challenge, adaptive signal control systems such as SCATS^[9], OPAC^[10], SCOOT^[11], RHODES^[1], PROLYN^[12], and MOTION^[13] have been proposed. Built upon strategies designed by experts, adaptive signal control can integrate additional traffic-related knowledge and operate with a greater intelligence level compared to pure actuated control that relies solely on collected data. By utilizing upstream detector data to estimate incoming traffic flow, the adaptive control method seeks an optimal strategy based on objectives such as traffic delay and emission reduction^[14–16].

Conventional adaptive control grapples with many significant challenges^[17]. A noteworthy limitation stems from its reliance on predefined rules (e.g., restrictions on the number of vehicles/lane occupancy and fixed thresholds of signal duration), which may impact the control performance, particularly in dynamic real-world scenarios. Furthermore, during demand surge periods, it often struggles to dynamically adjust signal timings, resulting in prolonged delays in densely populated urban areas with excessive traffic volumes.

Recent decades have witnessed an increasing amount of learning-based applications in TSC optimization. In the early stage, fuzzy logic, neural networks (NNs), and RL are mostly adopted for intelligent signal control^[18,19]. Due to the curse of dimensionality, as the information collected from various sources (such as fixed-location sensors and probe vehicles) continues to increase, conventional methods are difficult to cope with such high-dimensional inputs on a large scale^[20,21]. Therefore, DRL-based TSC algorithms have been developed to handle such complexity^[22]. Numerous studies have integrated DRL methods to tackle traffic-related problems including

TSC, showcasing its effectiveness in managing complex traffic scenarios^[23,24]. By leveraging DRL, the agent autonomously learns intricate strategies directly from data in a model-free manner, enabling more flexible and adaptive signal control strategies^[25,26]. Consequently, its performance relies primarily on collected data, rather than simulation settings or model assumptions, ensuring model robustness.

In general, DRL agents are trained to control traffic signals by selecting appropriate phases in a cycle (i.e., phase allocation)^[27,28], or keeping/switching the current phase (i.e., phase split)^[29–31]. In response to dynamic and/or unpredictable conditions at an intersection, agents may also adjust their control strategies to minimize overall waiting time or queue length from the system perspective.

In the optimization problem of DRL-based signal control, there are two main tasks: signal timing and phase sequencing optimization. In specific, a signal timing optimization strategy should determine the duration of each phase given a fixed phase sequence. For example, the phase duration can be adjusted in a phase-by-phase manner based on real-time traffic demands^[32]. Although the strategy allows for some adaptability, it has certain limitations, particularly during peak hours or high-demand traffic flow scenarios. Moreover, the non-uniform distribution of vehicle arrivals in different directions further complicates its practical application, making it difficult to achieve consistent and effective performance.

The phase-switching strategy, on the other hand, provides potential solutions to overcoming the drawbacks mentioned above^[33]. The method introduces a dynamic control mechanism for signals by selecting phases in a flexible sequence. Additionally, if a selected phase aligns with the previous one, its duration will be extended, leading to smooth traffic flow under stable traffic conditions^[34–36]. The strategy is implemented by switching phases at fixed time intervals, thereby enhancing the flexibility of signal control. Then, it can be adjusted to accommodate dynamic traffic patterns during different periods throughout the day. Some recent relevant papers are recommended for interested readers^[37–39].

Determining the appropriate intervals, however, presents a significant challenge when conducting the phase-switching strategy^[40]. The short phase-switching interval will result in frequent phase changes, limiting the number of vehicles that can pass through the intersection at once, leading to deteriorated intersection capacity. Additionally, it also increases the required computational resources during training and operation for strategy updating. Conversely, excessive intervals bring prolonged green duration, posing difficulties in promptly responding to real-time dynamic traffic. This may also hinder the agent's learning process, leading to sub-optimal solutions.

In summary, many existing reinforcement learning-based signal control models may result in ineffective phase allocation. For example, some strategies may not adapt to peak traffic flow scenarios in specific directions, resulting in extended periods of green light with no vehicles passing through the intersection. Meanwhile, phase-switching strategy may create excessive yellow light time due to inappropriate phase-switching intervals, both of which hinder the intersection traffic efficiency. To address this issue, a joint approach can be employed: the phase-switching agent can control the signal phases to improve the efficiency of vehicle movements, while the phase duration optimization agent can determine the duration of each phase and reduce the proportion of yellow light time.

Focusing on the TSC of isolated intersections, which can potentially be similarly extended to multiple intersections, we propose a DRL-based control strategy that jointly optimizes phase sequence and duration. The contribution of this study includes:

- We train two DRL-based agents to jointly optimize phase sequence and timing for intersection control, which aims to address

the limitations of control strategies with fixed sequences or durations.

- We stitch image-like cell information (such as vehicle positions and speeds) and feature-based information (such as queue length and phase duration) as state representation. The question of whether the training of DRL-based agents benefits from the knowledge of microscopic and macroscopic scales is investigated.

- We conduct ablation tests to identify joint action settings within the proposed framework, compared with methods that focus on phase-switching and phase duration only, which could hopefully provide guidelines for future DRL-based TSC studies.

Materials and methods

A review of Deep Q-Networks

Deep Q-Network (DQN) is commonly used in the field of intelligent TSC because of its straightforward implementation and tuning process. Additionally, DQN and its variants are off-policy methods that allow more exploration and generate discrete actions that are practical and suitable for TSC problems, adopted by most existing studies. Upon learning a policy π , the RL agent observes the state s_i , takes the action a_i , and finally obtains the value Q at time i . The action-value function for this so-called Q-learning is given as:

$$Q^*(s, a) = \max_{\pi} \mathbb{E}[R_t | s_t = s, a_t = a, \pi] \quad (1)$$

In this procedure, the agent learns the optimal policy π^* by Q-values — expected future rewards. However, $Q^*(s, a)$ is often difficult to obtain by conventional methods due to high-dimensional input (curse of dimensionality)^[41]. The recent integration of the neural network (NN) makes it possible to interpret an approximator $Q(s, a; \theta)$, which is called the DQN structure. Using the Bellman equation^[42], the NN parameters are updated by stochastic gradient descent, and the loss or temporal difference (TD) error function is finally minimized.

$$\mathcal{L} = \mathbb{E}[(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))^2] \quad (2)$$

Many DQN extensions have emerged that can boost the agents' performance by 2–6 or even more in some tasks (e.g., Atari games). Therefore, we posit that those DQN variants will also find their value in solving TSC problems.

It is found that Eqn (2) uses the same neural network for 'greedy' action selection and action evaluation, leading to overestimation bias in the training process. To overcome this problem, the idea of double DQN (DDQN) is proposed by separating the two procedures^[43]. The loss function of DDQN becomes:

$$\mathcal{L} = \mathbb{E}[(R_t + \gamma Q_{\text{target}}(s_{t+1}, \arg\max_a Q_{\text{main}}(s_{t+1}, a)) - Q_{\text{main}}(s_t, a_t))^2] \quad (3)$$

Considering sample selection, the agent's experiences are typically archived within a dataset. During the training phase, samples are randomly selected from this dataset to disrupt the inter-sample correlations, a process known as experience replay^[44]. However, in this way, rare samples that are valuable to the goals can hardly be selected compared to some high-frequency redundancy. To address this imbalance, a prioritized experience replay mechanism was introduced, which allocates higher priority to samples exhibiting substantial absolute temporal difference (TD) errors^[45]:

$$p_i = |r_i + \gamma \max_a Q(s_i, a) - Q(s_{i-1}, a_{i-1})| \quad (4)$$

where, p_i denotes the priority of sample i , updated by the loss after the neural network's forward pass.

Based on the advantages of DDQN, a novel network architecture named 'dueling network' (3DQN) was presented in a study by Wang et al.^[46], in which the state-value function $V(s)$ and the advantage function $A(s, a)$ are estimated by two collateral networks with

different sets of hyper-parameters. By aggregating the collateral networks, the Q-value function can be approximated by the following formula:

$$Q(s, a | \theta, \theta') = V(s | \theta) + (A(s, a | \theta') - \frac{1}{|\Delta_\pi|} \sum_{a'} A(s, a' | \theta')) \quad (5)$$

in which θ and θ' are two different sets of hyper-parameters and Δ_π is the discrete action space of policy π .

Some other extensions have contributed to addressing different issues during the training of DQNs, such as Distributional RL, Noisy Nets, and others^[47,48]. An algorithm called 'Rainbow' aims to integrate comprehensively with all the aforementioned improvements to train the agents^[49]. Due to limited space, these variants will not be introduced here. To date, DQN and similar algorithms remain one of the most widely used approaches in addressing TSC problems^[50–52]. In this study, we deploy the framework of 3DQN to train the signal controller and examine the impacts of different training settings on its performance.

Methodology

The problem of traffic signal control can be modeled as a Markov Decision Process (MDP) due to its sequential characteristics. In this process, an agent interacts with traffic scenarios to capture feature data, makes decisions according to the dynamics of the environment, and collects evaluation data as feedback to improve the agent's strategies (as presented in Fig. 1). The key point of dealing with TSC problems lies in the design of the state, action, and reward of the DRL, which are presented in the following subsections.

State representation

The state space should contain information that can accurately depict the environment. For TSC, an isolated intersection is the basic

unit for collecting information that includes vehicle, road, and signal in the agent's learning process, which is divided into cells, as shown in Fig. 2. We discretize the real-time position, speed, and acceleration into three matrices. When a vehicle is located in a certain cell, the corresponding value in the position matrix P is set to 1. The corresponding value of the speed matrix V and the acceleration matrix A is set to contain all vehicles' instant speeds and acceleration rates. To ensure the accuracy of vehicle information, the cell size is slightly larger than the vehicle size. For road information, the number of vehicles N , average queue length $\bar{L}$, average waiting time $\bar{W}$, and average speed $\bar{V}$ corresponds to intersection traffic conditions. By collecting these states in different directions, the agent can quickly determine the critical directions/lanes at the intersection using macroscopic information and evaluate the temporal and spatial relationships of vehicles at the intersection using microscopic information. This enables the agent to effectively identify and locate congestion points, allowing for proactive responses and reasonable adjustments to the intersection signal control strategy. The road information can be described as follows:

$$RI = \{N_{ns}, N_{ew}, \bar{L}_{ns}, \bar{L}_{ew}, \bar{W}_{ns}, \bar{W}_{ew}, \bar{V}_{ns}, \bar{V}_{ew}\} \quad (6)$$

In addition, we consider two stages of signal information for agents, which includes the last phase ϕ_t (for the agent that selects the next phase) or the current phase ϕ_{t+1} (for the agent that optimizes the timing) in the state space. Therefore, the state space for TSC incorporates three types of information, namely $s_t = \{P, V, A, RI, \phi\}$.

Actions

In this study, we design separate actions for two different agents, namely, phase switching, and phase duration. The phase-switching

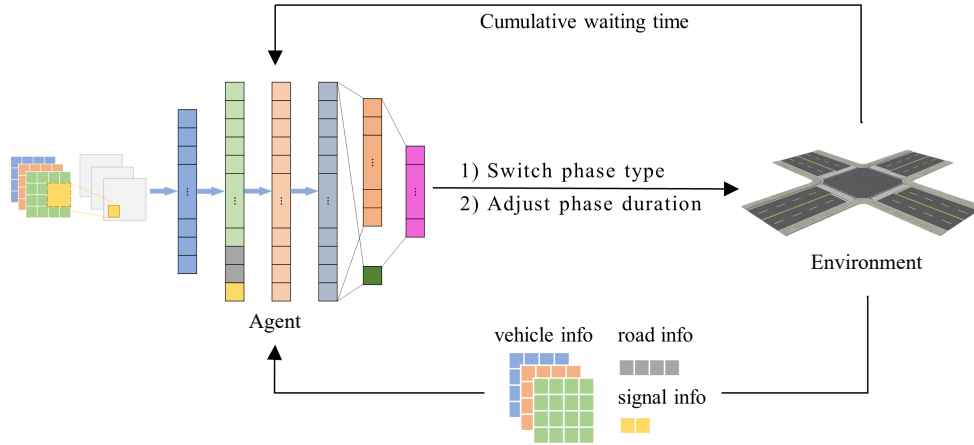


Fig. 1 The schematic diagram of the reinforcement learning process.

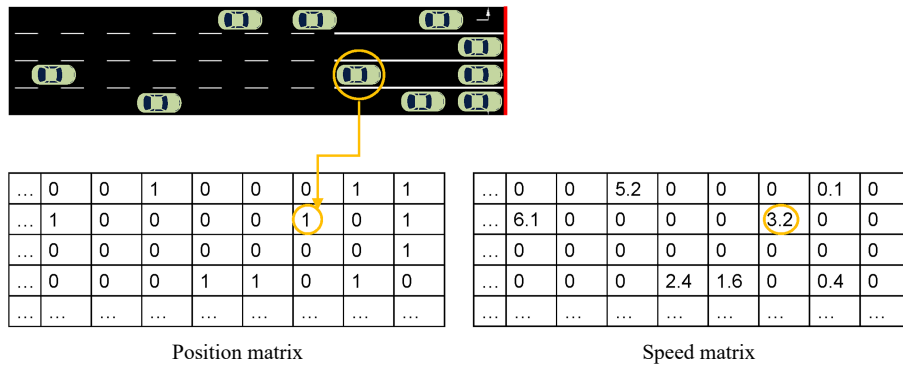


Fig. 2 Vehicle information extraction at one entrance of the intersection.

agent selects which phase should be switched to at the next interval. Thus, the action space A_1 consists of all available phases ϕ . The phase duration agent receives action a_1 from A_1 and determines the duration of a_1 . The action space A_2 can then be defined as a discrete set of integers to represent the phase duration in seconds because the signal timing values such as cycle length, green time, and red time are all implemented in integers in real practice.

$$\begin{aligned} A_1 &= \{\phi_1, \phi_2, \dots, \phi_p\} \\ A_2 &= \{0, 1, \dots, T\} \end{aligned} \quad (7)$$

To address the problems (insufficient or excessive green signal duration, unexpected phase switching, etc.) of RL-based TSC, several widely-adopted rules are implemented to restrict arbitrary action selection:

(1) Minimum green time: Is employed in TSC to align with drivers' psychological expectations. Consequently, only when the minimum green time has been reached, the controller can perform a phase-switching.

(2) Default phase: During the learning stage, RL-based TSC would randomly select a phase to activate in the absence of vehicles at the intersection. This random selection can have adverse effects on both traffic efficiency and safety. To address this, a default phase representing the primary movements of the major approach is established to prevent arbitrary phase changes. Specifically, the default phase is set when there is no vehicle within the intersection range, with transitions following the phase switching sequence of traditional signal timing, such as switching from north-south straight-through to north-south left-turn.

(3) Maximum green time: Setting a maximum green time helps to ensure fair allocation of right-of-way and prevent excessive delay for directions with low traffic volume. If the duration of any phase exceeds the maximum green time, the signal will be instantly switched to other phases.

Reward function

The reward decides whether an agent can achieve the objectives as expected. Most of the related studies exploited waiting time^[19,43,53], queue length^[54–56], or vehicle throughput^[57–59] as components of the reward function. While lots of evaluation metrics, including queue length, average speed, and throughput can train agents to achieve the learning goal to some extent, they may encounter problems of unclear reward signals in the training process. In this study, we consider the total waiting time as the reward component, referred to in the studies by Mannion et al.^[19], Van Hasselt et al.^[43], and Nishi et al.^[53], which can be expressed as:

$$W_t = \sum_{i=1}^n w_{t(i,t)} \quad (8)$$

where, W_t is the total waiting time of all vehicles in the scenario currently, $w_{t(i,t)}$ is the waiting time of vehicle i at the time step t , and n is the total number of vehicles at this time. The reward is defined as follows:

$$R_t = W_{t+1} - W_t \quad (9)$$

where, R_t is the reward at the time step t . When $R_t < 0$, it indicates that the cumulative waiting time of the vehicles in the road network is reduced.

Results and discussion

To validate the performance of the proposed joint control model, experiments were conducted using the Simulation of Urban Mobility (SUMO) platform. Firstly, we provide a brief introduction to the simulation settings. Then, we conduct experiments and compare the results with the widely-adopted Webster and MaxPressure methods. Finally, ablation experiments are performed using different representations of states and actions to figure out how these elements affect the experiment results.

Simulation settings

3DQN was chosen to learn and make decisions for the joint control of two agents, and ReLU was used as the activation function for all networks. Parameter settings are demonstrated in Table 1^[22,29]. The ϵ -greedy was added to ensure that the method can explore all executable actions even near the end of the training process to alleviate the Exploration-Exploitation dilemma^[60]. In this study, the exploration rate ϵ decreased from 1 to 0 as the training episode progressed.

The experiment scenario is illustrated in Fig. 3. We assumed a four-legged intersection with four lanes in each direction, including two through lanes, a through and right-turn lane, and a left-turn lane. The lane length was set to be 750 m. Vehicle arrival conforms to the Weibull distribution. To test the performance of each method under different traffic conditions, we first examined three traffic volumes to simulate low, medium, and high traffic conditions, corresponding to the traffic flow of 1,000, 1,500, and 2,000 veh/h for the junction, respectively. As for specific traffic volume in different directions, we simulated the routing of vehicles in a stochastic manner to validate the model's robustness. Traffic flow directions are categorized into straight-through and turning cases. Straight-through situations include four scenarios (e.g., north to south), while turning situations consist of eight scenarios (including left and right turns, e.g., north to east). The generation positions of vehicles are uniformly distributed across all entrances, with a 75% probability of going straight and a 25% probability of turning. Among turning movements, the probabilities of left turns and right turns are equal. All experiments were implemented using the Python API provided by SUMO. The neural networks were trained by Pytorch.

Table 1. Parameter settings of the training process.

Parameters	Description	Value
total_episodes	Total number of training episodes where agents interact with the environment and update strategies	2,000
max_steps	Maximum steps (s) in one episode	3,600
iterations	Number of batches extracted during training	100
batch_size	Number of data in one batch	256
memory_size_min	Minimum memory size	512
memory_size_max	Maximum memory size	20,480
learning_rate	Step size in the optimization process	0.001
gamma	Discount factor	0.75

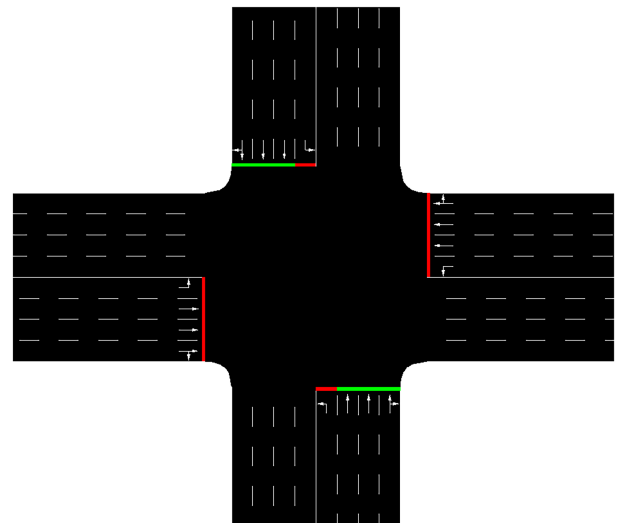


Fig. 3 The simulated intersection scenario in all experiments.

Table 2. Comparative results with benchmark models.

	Average queue length (m)			Average travel time (s)			Average vehicle speed (m/s)		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Webster	22.15	46.35	94.85	139.61	147.95	161.59	11.90	11.23	10.30
Max pressure	18.95	38.25	81.90	136.24	143.35	154.76	12.03	11.44	10.45
Joint control	17.65	35.80	77.95	135.74	141.34	153.98	12.11	11.46	10.71

Values in bold indicate the optimal results across different models.

Comparisons with benchmark methods

We compared the joint control model with the benchmark methods listed below. To ensure the performance of the methods, all of them are carefully tuned in the experiments. In the study, we chose queue length, vehicle speed, and travel time as evaluation metrics because they are more intuitive to reflect traffic mobility, congestion, and throughput efficiency, and allowing for a comprehensive assessment of the overall performance of the traffic system and the effectiveness of the signal control strategy.

- Webster: It operates on a fixed cycle, using road and traffic flow data from intersections to determine the duration of each phase within the cycle. In addition, phase switching is carried out in a fixed sequence^[61].

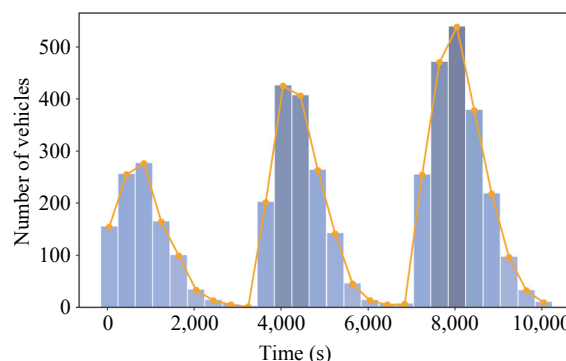
- MaxPressure: It selects a phase for switching based on the current traffic situation at the intersection at every fixed time interval. The selection criteria are to prioritize the phase with the highest pressure. The phase duration of the signal timing scheme is fixed while the sequence is dynamic^[62].

To comprehensively understand the experimental results, we exploit average queue length, average travel time, and average vehicle speed to evaluate the performance. In existing studies, these metrics are often used to reflect the actual capacity of an intersection, which is calculated by the average length of all queuing vehicles, the average travel time of all vehicles spent, and the average speed of all vehicles in the system. The comparative results of all metrics are given in Table 2. By analyzing these results, we have the following findings:

(1) Among the traditional methods, the Webster method performs the worst according to the three metrics, mainly due to its limited expert experience guidance and inflexibility. MaxPressure, known as a sufficient signal timing control method, exhibits satisfactory outcomes in terms of average queue length, travel time, and vehicle speed. The proposed joint control model outperforms all other methods and is the most effective in improving intersection conditions.

(2) The analysis under varying traffic volumes reveals that the Webster method faces great challenges when applied to scenarios with high traffic flow, as its performance drastically declines with the increasing demand. In contrast, both the MaxPressure and joint control models can exhibit stable adaptability to different traffic conditions. In particular, the joint control model outperforms others in low traffic flow, achieving a 6.86% reduction in the average queue length compared to MaxPressure.

(3) The selection of phase switching interval significantly impacts the performance of MaxPressure, so determining the most appropriate interval in advance is challenging. Enumerating all possible intervals is computationally inefficient and time-consuming. In contrast, the joint control model does not require to pre-determine the interval, making it more flexible in dynamic traffic conditions. Further simulation experiments with dynamic traffic volumes ranging from 1,000 to 2,000 vehicles per hour (as shown in Fig. 4) were conducted, and the results are summarized in Table 3. The joint control model outperforms MaxPressure by achieving a 7.56% reduction in average queue length.

**Fig. 4** Settings of dynamic traffic flow.**Table 3.** The results of tests under the temporal dynamic traffic flow.

	JointControl	MaxPressure	Improvement
Average queue length (m)	48.50	52.47	-7.56%
Average travel time (s)	146.48	147.87	-0.94%
Average vehicle speed (m/s)	11.46	11.19	+2.41%

Ablation tests

We conducted two ablation experiments about the proposed method to explore the effects of state representation and action selection. The settings are presented as follows:

State representation is a crucial aspect of the joint control model, where both microscopic vehicle information and macroscopic road information are employed as its state. To investigate how the information influences strategy selection and the model's performance, we conducted experiments using micro vehicle information and macro road information as separate state representations.

Action selection has been explored in some existing studies, but typically, only one single type of action has been applied to control signal settings. Few of them have delved into the impact of action selection on the model performance, considering the type and dimension of the action. In this study, we introduce a novel approach involving two dimensions of different types of actions, simultaneously controlling phase and timing. By comparing phase control only, timing control only, and the combination of those two, we aim to gain valuable insights into the impacts of different action selections on overall performance.

The training curves of all the above settings are demonstrated in Fig. 5, and the results of the ablation experiments are given in Table 4. The results of this study highlight the importance of both state representation and action selection on the performance of the joint control model. Regarding state representation, we observed that the model utilizing micro information outperforms the one using macro information. This suggests that microscopic information provides more detailed and specific descriptions of traffic scenarios, enabling the model to effectively leverage and incorporate. In addition, macro information offers a broader perspective on the overall traffic conditions, enhancing decision-making accuracy as a supplementary factor. However, relying solely on macro information results in challenges in gathering effective traffic data,

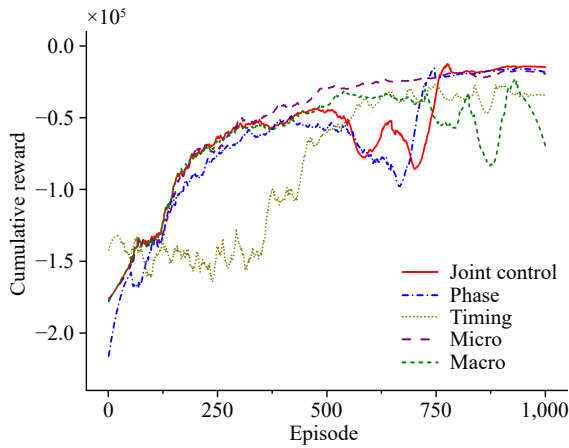


Fig. 5 The training curves of ablation experiments.

leading to sub-optimal decisions and inferior performance. As a result, achieving a balanced integration of both micro and macro information proves critical for optimizing the joint control model's performance.

As for action selection, we found that the phase switching control outperforms the strategy with fixed sequence and can better accommodate the spatio-temporal imbalances in traffic flow. Furthermore,

Table 4. The simulation results of ablation experiments.

	Average queue length (m)	Average travel time (s)	Average vehicle speed (m/s)
Micro info only	79.90	154.50	10.42
Macro info only	215.80	206.62	8.96
Phase control – single agent	83.25	155.97	10.45
Timing control – single agent	135.58	180.44	9.92
Joint control	77.95	153.98	10.71

Values in bold indicate the optimal results across different models.

the joint control model performs more effectively in reducing traffic congestion. Its superiority can be attributed to the model's capability to jointly optimize the phase sequence and the switching interval based on real-time traffic conditions. In contrast, only adjusting the phase duration is insufficient to alleviate congestion caused by excessive vehicle volumes. Thus, the holistic approach of the joint model to control both phase sequencing and duration is vital in optimizing traffic flow and minimizing congestion.

Symmetry test with imbalanced demand in different directions

To further investigate how the proposed joint control method operates under dynamic traffic conditions, Figure 6 provides

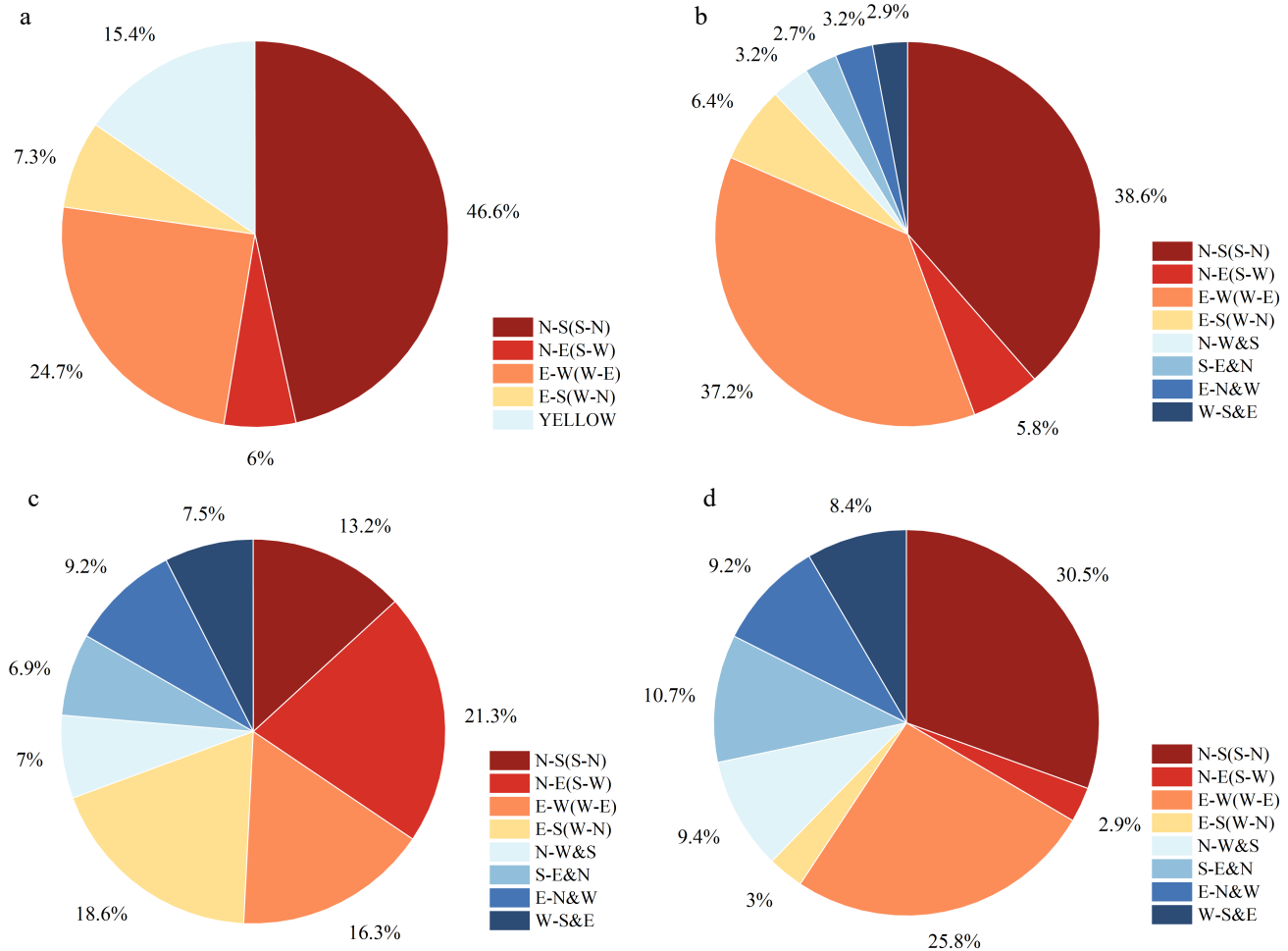


Fig. 6 The ratios of phase duration, the number of vehicles, and cumulative queue length with two different strategies. (a) The proportion of duration of all phases with the asymmetric strategy (the proposed model), (b) The proportion of vehicles in all lanes with the asymmetric strategy (the proposed model), (c) The proportion of queue length in all lanes with the asymmetric strategy (the proposed model), (d) The proportion of queue length in all lanes with the symmetric strategy.

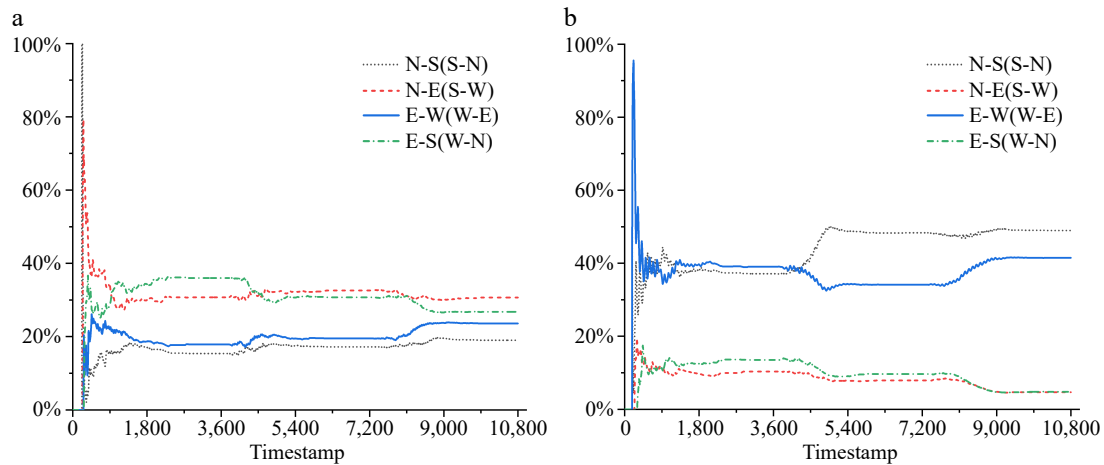


Fig. 7 The time-varying queue length proportion under asymmetric and symmetric strategies. (a) With the asymmetric strategy (the proposed model), (b) With the symmetric strategy.

insights into the ratio of each phase, the number of vehicles, and the cumulative queue length in each lane. Note that 'vehicles in each lane' and 'cumulative queue length in each lane' are categorized by phases. The traffic flow settings were as shown in Table 3. Figure 6b illustrates a well-balanced traffic flow, for example, the number of vehicles driving from N(S) to S(N) is similar to vehicles driving from E(W) to W(E). From Fig. 6a, we find that the duration of phase N-S(S-N) accounts for almost half of the entire process and E-W(W-E) only occupies about a quarter. Hence, the proposed model has learned an asymmetric strategy to evacuate queues. Typically, the ratio of each phase corresponds to the number of vehicles in each lane, as more vehicles in the lanes necessitate more time to clear them, for example, in directions N-E(S-W) and E-S(W-N). However, the cases of N-S(S-N) and E-W(W-E) do not follow this rule, which could be due to variations in vehicle arrival frequency in different directions. In general, the denser arrival of vehicles leads to a shorter duration of time required for them to clear, as they can arrive and leave within a more concentrated period. Examining the proportion of cumulative queue length in each lane in Fig. 6c, it becomes evident that the N-E(S-W) direction exhibits the highest ratio. Subsequently, E-S(W-N), E-W(W-E), and N-S(S-N) account for 18.6%, 16.3%, and 13.2%, respectively. These results show that the proposed asymmetric strategy needs to sacrifice some traffic flow of the left-turn lanes to optimize the overall performance. For comparison purposes, we define a symmetric strategy that allocates phase duration proportionally based on the traffic flow. That is, the strategy will assign more ratios of phases to directions with greater numbers of vehicles in each lane, as shown in Fig. 6d. With the symmetric strategy, the signal timing is set to 20 s for each phase.

Figure 7a demonstrates that, with the proposed asymmetric strategy, the cumulative queue length distribution has been balanced in all under-controlled directions throughout the whole process, while the distribution of vehicles' passing in each lane is asymmetrical. For instance, the cumulative queue length in E-W(W-E) is 3.1% greater than that in N-S(S-N) with 21.9% lower phase duration. However, with the symmetric strategy, the queue length in different directions has a drastic imbalance, which can be found in Fig. 7b.

In conclusion, this study demonstrates the crucial role of state representation and action selection in the joint control model's performance. The findings emphasize the significance of balancing micro and macro information and highlight the advantages of dynamic phase sequence control in traffic management. These insights contribute to enhancing the efficiency of traffic signal control systems, ultimately leading to traffic flow improvements and congestion mitigation in urban areas.

Conclusions

This study introduced a novel deep reinforcement learning-based joint control model for signal optimization at isolated intersections. The model employed vehicle information, road information, and signal data as inputs, allowing a comprehensive understanding of the dynamics of intersections. To effectively smooth traffic flow, this study designed two intelligent agents that work in tandem to control signal phase switching and its duration in a cooperative manner.

Through simulation experiments, we thoroughly evaluated the performance of the proposed approach in various traffic scenarios. Under fixed traffic flow conditions, the joint control model demonstrated a slight but notable advantage over the baseline methods. In dynamic traffic flow scenarios, the algorithm obtains states and rewards from the scenario and makes decisions, generating samples from various traffic flows. By learning from different samples, the algorithm can determine the optimal signal control strategy under various traffic flow conditions. By continuously collecting traffic data, the algorithm can refine the model, enabling it to effectively adapt to dynamic traffic flow. The real strength of the model came to the forefront when traffic flow dynamically changed, as it exhibited significant benefits over other methods.

Delving into the ablation experiments, we unveiled the pivotal role played by microscopic vehicle information in the signal control process. The specific data related to individual vehicles significantly influenced the model's performance, enabling a more precise and effective traffic control strategy. Moreover, we found that incorporating dynamic phase sequence control further boosted the model's performance, particularly in alleviating traffic congestion. The flexibility and adaptability provided by dynamic phase sequence adjustments can better improve traffic conditions compared to other methods.

In summary, the proposed joint control model is a promising solution for traffic management, harnessing the power of microscopic vehicle data and dynamic phase sequence control. By coordinating signal switching and duration, the model effectively improves intersection conditions.

There are still some limitations of this study, mainly the exclusive focus on traditional four-way four-phase intersections. To enhance the model's applicability, future studies can explore expanding the scope of scenarios to include other intersection/vehicle types and incorporate pedestrian waiting time, ensuring a more comprehensive and effective approach. The heterogeneity of driving styles

among different drivers is another important factor affecting the effectiveness of signal control. It influences aspects such as following distances, vehicle speeds, and lane-changing behaviors during driving, which change the state representation of the environment and, consequently, impact the agent's ability to determine the optimal signal control strategy. In addition, we need to extend the framework to multi-intersection control. Future research will focus on how to effectively utilize and coordinate the state and reward information of each intersection, such as state representation improvements and comprehensive reward designs. Meanwhile, it is also essential to explore the differences in algorithm performance under different road structures, vehicle arrival patterns, and partially observable conditions to ensure its adaptability to real-world scenarios. Our approach proves more stable and effective than timed signal control in various simulated scenarios. However, applying the algorithm in the real-world requires consideration of various factors, such as bus priority, fairness in traffic flow allocation under mixed conditions, and the ability to handle unexpected events. In future experiments, we will consider incorporating different types of vehicles and pedestrian flows into scenarios, as well as introducing certain unexpected events to test the algorithms' generalization capability. Providing a global-optimal control strategy for multiple agents to solve the failure caused by the dramatic increase in action-space dimension for multi-intersection scenarios is another possible extension for follow-up research. Also, some measures should be implemented to alleviate drivers' confusion and psychological discomfort caused by unexpected phase switching. By addressing these improvements, the joint control model can potentially make significant contributions to intelligent transportation systems, smoothing traffic flow, and minimizing congestion in urban traffic environments.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Sun Z, Jia X, Ji A, Lin X; analysis and interpretation of results: Jia X, Cai Y, Ji A, Liu L, Tu Y; draft manuscript preparation: Jia X, Cai Y, Ji A, Wang W. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

The work is supported by the Tencent-SWJTU Joint Laboratory of Intelligent Transportation (Grant No. R113623H01015), the National Natural Science Foundation of China (Grant Nos 52072316 and 52302418), the Sichuan Science and Technology Program (Grant No. 2024NSFSC0942), the Fundamental Research Funds for the Central Universities (Grant No. 2682023CX047), and the Postdoctoral International Exchange Program (Grant No. YJ20220311).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 15 October 2024; Revised 9 January 2025; Accepted 15 February 2025; Published online 27 June 2025

References

1. Mirchandani P, Head L. 2001. A real-time traffic signal control system: architecture, algorithms, and analysis. *Transportation Research Part C: Emerging Technologies* 9:415–32
2. Rasheed F, Yau KA, Noor RM, Wu C, Low YC. 2020. Deep reinforcement learning for traffic signal control: a review. *IEEE Access* 8:208016–44
3. Qin Z, Ji A, Sun Z, Wu G, Hao P, et al. 2024. Game theoretic application to intersection management: a literature review. *IEEE Transactions on Intelligent Vehicles* 00:1–19
4. Maslekar N, Mouzna J, Boussedjra M, Labiod H. 2013. CATS: an adaptive traffic signal system based on car-to-car communication. *Journal of Network and Computer Applications* 36:1308–15
5. Hoogendoorn S, Knoop V. 2013. Traffic flow theory and modelling. In *The transport system and transport policy: An introduction*, eds. van Wee B, Annema JA, Banister D, Pudāne B. Cheltenham, UK: Edward Elgar Publishing. pp. 125–59
6. Miller A. 1963. A computer control system for traffic networks. *Proceedings of the International Symposium on the Theory of Traffic Flow and Transportation, London, UK, 1963*. London, UK: Organisation for Economic Co-operation and Development. <https://trid.trb.org/View/612653>
7. Zheng X, Recker W, Chu L. 2010. Optimization of control parameters for adaptive traffic-actuated signal control. *Journal of Intelligent Transportation Systems* 14:95–108
8. Zheng X, Chu L. 2008. Optimal parameter settings for adaptive traffic-actuated signal control. *2008 11th International IEEE Conference on Intelligent Transportation Systems*. October 12–15, 2008, Beijing, China. USA: IEEE. pp. 105–10. doi: [10.1109/ITSC.2008.4732676](https://doi.org/10.1109/ITSC.2008.4732676)
9. Sims AG, Dobinson KW. 1980. The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits. *IEEE Transactions on Vehicular Technology* 29:130–37
10. Gartner N. 1983. OPAC: A demand-responsive strategy for traffic signal control. *Transportation Research Record, No. 906*. pp. 75–81
11. Bing B, Carter A. 1995. SCOOT: The world's foremost adaptive TRAFFIC control system. *Traffic Technology International '95*. Surrey, UK: UK and International Press. <https://trid.trb.org/View/415757>
12. Henry JJ, Farges JL, Tuffal J. 1984. The PROLYN real time traffic algorithm. In *Control in Transportation Systems. Proceedings of the 4th IFAC/IFIP/IFORS Conference, Baden-Baden, Federal Republic of Germany, 20–22 April 1983*. Germany: Elsevier. pp. 305–10. doi: [10.1016/B978-0-08-029365-3.50048-1](https://doi.org/10.1016/B978-0-08-029365-3.50048-1)
13. Brilon W, Wietholt T. 2013. Experiences with adaptive signal control in Germany. *Transportation Research Record: Journal of the Transportation Research Board* 2356:9–16
14. Lertworawanich P, Unhasut P. 2021. A CO emission-based adaptive signal control for isolated intersections. *Journal of the Air & Waste Management Association* 71:564–85
15. Mondal MA, Rehena Z. 2022. Priority-based adaptive traffic signal control system for smart cities. *SN Computer Science* 3:417
16. Lee WH, Wang HC. 2022. A person-based adaptive traffic signal control method with cooperative transit signal priority. *Journal of Advanced Transportation* 2022:2205292
17. Jing P, Huang H, Chen L. 2017. An adaptive traffic signal control in a connected vehicle environment: a systematic review. *Information* 8:101
18. Liu Z. 2007. A survey of intelligence methods in urban traffic signal control. *International Journal of Computer Science and Network Security* 7(7):105–12
19. Mannion P, Duggan J, Howley E. 2016. An experimental review of reinforcement learning algorithms for adaptive traffic signal control. In *Autonomic Road Transport Support Systems*, eds. McCluskey T, Kotsialos A, Müller J, Klügl F, Rana O, et al. Cham: Springer International Publishing. pp. 47–66. doi: [10.1007/978-3-319-25808-9_4](https://doi.org/10.1007/978-3-319-25808-9_4)
20. La P, Bhatnagar S. 2011. Reinforcement learning with function approximation for traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 12:412–21
21. Mohamad Alizadeh Shabestary S, Abdulhai B. 2022. Adaptive traffic signal control with deep reinforcement learning and high dimensional sensory inputs: case study and comprehensive sensitivity analyses. *IEEE Transactions on Intelligent Transportation Systems* 23:20021–35

22. Liang X, Du X, Wang G, Han Z. 2019. A deep reinforcement learning network for traffic light cycle control. *IEEE Transactions on Vehicular Technology* 68:1243–53
23. Ge H, Song Y, Wu C, Ren J, Tan G. 2019. Cooperative deep Q-learning with Q-value transfer for multi-intersection signal control. *IEEE Access* 7:40797–809
24. Haddad TA, Hedjazi D, Aouag S. 2022. A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control. *Engineering Applications of Artificial Intelligence* 114:105019
25. Chu T, Wang J, Codecà L, Li Z. 2020. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 21:1086–95
26. Bouktif S, Cheniki A, Ouni A. 2021. Traffic signal control using hybrid action space deep reinforcement learning. *Sensors* 21:2302
27. Du Y, Shang Guan W, Rong D, Chai L. 2019. RA-TSC: learning adaptive traffic signal control strategy via deep reinforcement learning. 2019 *IEEE Intelligent Transportation Systems Conference (ITSC)*, October 27–30, 2019, Auckland, New Zealand. USA: IEEE. pp. 3275–80. doi: [10.1109/itsc.2019.8916967](https://doi.org/10.1109/itsc.2019.8916967)
28. Kumar N, Mittal S, Garg V, Kumar N. 2022. Deep reinforcement learning-based traffic light scheduling framework for SDN-enabled smart transportation system. *IEEE Transactions on Intelligent Transportation Systems* 23:2411–21
29. Li L, Lv Y, Wang FY. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3:247–54
30. Kumar N, Rahman SS, Dhakad N. 2021. Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system. *IEEE Transactions on Intelligent Transportation Systems* 22:4919–28
31. Ma D, Zhou B, Song X, Dai H. 2022. A deep reinforcement learning approach to traffic signal control with temporal traffic pattern mining. *IEEE Transactions on Intelligent Transportation Systems* 23:11789–800
32. Aslani M, Mesgari MS, Wiering M. 2017. Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events. *Transportation Research Part C: Emerging Technologies* 85:732–52
33. Wei H, Zheng G, Gayah V, Li Z. 2019. A survey on traffic signal control methods. *arXiv Preprint*
34. El-Tantawy S, Abdulhai B. 2010. An agent-based learning towards decentralized and coordinated traffic signal control. 13th *International IEEE Conference on Intelligent Transportation Systems*, September 19–22, 2010, Funchal, Portugal. USA: IEEE. pp. 665–70. doi: [10.1109/ITSC.2010.5625066](https://doi.org/10.1109/ITSC.2010.5625066)
35. Khamis MA, Goma W. 2012. Enhanced multiagent multi-objective reinforcement learning for urban traffic light control. 2012 11th *International Conference on Machine Learning and Applications*, December 12–15, 2012, Boca Raton, FL, USA. USA: IEEE. pp. 586–91. doi: [10.1109/ICMLA.2012.108](https://doi.org/10.1109/ICMLA.2012.108)
36. Mousavi SS, Schukat M, Howley E. 2017. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems* 11:417–23
37. Zhao J, Yao T, Zhang C, Shafique MA. 2024. Signal control for overflow prevention at intersections using partial connected vehicle data. *Transportmetrica A: Transport Science* 1–31
38. Ma C, Yu C, Zhang C, Yang X. 2023. Signal timing at an isolated intersection under mixed traffic environment with self-organizing connected and automated vehicles. *Computer-Aided Civil and Infrastructure Engineering* 38:1955–72
39. Yao T, Zhang C, Zhao J, Gupta A, Mondal S. 2023. Adaptive signal control for overflow prevention at isolated intersections based on fuzzy control. *Transportation Research Record: Journal of the Transportation Research Board* 2677:1387–401
40. Noaen M, Naik A, Goodman L, Crebo J, Abrar T, et al. 2022. Reinforcement learning in urban network traffic signal control: a systematic literature review. *Expert Systems with Applications* 199:116830
41. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. 2017. Deep reinforcement learning: a brief survey. *IEEE Signal Processing Magazine* 34:26–38
42. Bellman R. 1952. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America* 38:716–19
43. Van Hasselt H, Guez A, Silver D. 2016. Deep reinforcement learning with double Q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 30:2094–100
44. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518:529–33
45. Schaul T, Quan J, Antonoglou I, Silver D. 2015. Prioritized experience replay. *arXiv Preprint*
46. Wang Z, Schaul T, Hessel M, Hasselt H, Lanctot M, et al. 2016. Dueling network architectures for deep reinforcement learning. *International Conference on Machine Learning*, New York, USA, 2016. New York, USA: PMLR. pp. 1995–2003. <https://proceedings.mlr.press/v48/wangf16.pdf>
47. Bellemare MG, Dabney W, Munos R. 2017. A distributional perspective on reinforcement learning. *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017. PMLR. pp. 449–58. <https://proceedings.mlr.press/v70/bellemare17a/bellemare17a.pdf>
48. Fortunato M, Azar MG, Piot B, Menick J, Osband I, et al. 2017. Noisy networks for exploration. *arXiv Preprint*
49. Hessel M, Modayil J, Van Hasselt H, Schaul T, Ostrovski G, et al. 2018. Rainbow: combining improvements in deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 32(1):3215–222
50. Gu J, Fang Y, Sheng Z, Wen P. 2020. Double deep Q-network with a dual-agent for traffic signal control. *Applied Sciences* 10:1622
51. Park S, Han E, Park S, Jeong H, Yun I. 2021. Deep Q-network-based traffic signal control models. *PLoS One* 16:e0256405
52. Ducrocq R, Farhi N. 2023. Deep reinforcement Q-learning for intelligent traffic signal control with partial detection. *International Journal of Intelligent Transportation Systems Research* 21:192–206
53. Nishi T, Otaki K, Hayakawa K, Yoshimura T. 2018. Traffic signal control based on reinforcement learning with graph convolutional neural nets. 2018 21st *International Conference on Intelligent Transportation Systems (ITSC)*, November 4–7, 2018, Maui, HI, USA. USA: IEEE. pp. 877–83. doi: [10.1109/ITSC.2018.8569301](https://doi.org/10.1109/ITSC.2018.8569301)
54. Zang X, Yao H, Zheng G, Xu N, Xu K, et al. 2020. MetaLight: value-based meta-reinforcement learning for traffic signal control. *Proceedings of the AAAI Conference on Artificial Intelligence* 34:1153–60
55. Steingrover M, Schouten R, Peelen S, Nijhuis E, Bakker B. 2005. Reinforcement learning of traffic light controllers adapting to traffic congestion. *Proceedings of the Seventeenth Belgium-Netherlands Conference on Artificial Intelligence*, Brussels, Belgium, October 17–18, 2005.
56. Gokulan BP, Srinivasan D. 2010. Distributed geometric fuzzy multiagent urban traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 11:714–27
57. Salkham A. 2010. Decentralized optimization of fluctuating urban traffic using reinforcement learning. PhD thesis. Trinity College Dublin, UK
58. Xu LH, Xia XH, Luo Q. 2013. The study of reinforcement learning for traffic self-adaptive control under multiagent Markov game environment. *Mathematical Problems in Engineering* 2013:962869
59. Salkham A, Cunningham R, Garg A, Cahill V. 2008. A collaborative reinforcement learning approach to urban traffic control optimization. 2008 *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, December 9–12, 2008, Sydney, NSW, Australia. USA: IEEE. pp. 560–66. doi: [10.1109/WIIAT.2008.88](https://doi.org/10.1109/WIIAT.2008.88)
60. Kuleshov V, Precup D. 2014. Algorithms for multi-armed bandit problems. *arXiv Preprint*
61. Webster FV. 1958. *Traffic signal settings*. Road Research Technical Paper, No. 39. London, UK: Department of Scientific and Industrial Research.
62. Varaiya P. 2013. Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies* 36:177–95



Copyright: © 2025 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.