

Received 15 December 2024, accepted 3 February 2025, date of publication 7 February 2025, date of current version 25 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3539685

RESEARCH ARTICLE

Cooperative Control of Intersection Traffic Signals Based on Multi-Agent Reinforcement Learning for Carbon Dioxide Emission Reduction

HYEMIN KIM^{ID}, JINHYUK PARK^{ID}, DONGBEOM KIM^{ID}, AND CHULMIN JUN^{ID}

Department of Geoinformatics, University of Seoul, Dongdaemun-gu, Seoul 02504, Republic of Korea

Corresponding author: Chulmin Jun (cmjun@uos.ac.kr)

This work was supported by the 2023 Research Fund of the University of Seoul.

ABSTRACT Abnormal weather is occurring around the world, including the hottest weather in 174 years of observation records, the largest fire in Europe's observation records, and approximately twice the average annual rainfall recorded in one day. This abnormal climate is highly related to greenhouse gases, and efforts to reduce emissions are required in various fields. This study aims to reduce carbon dioxide emissions in the transportation sector, which accounts for a high proportion of emissions. A multi-agent reinforcement learning technique is used for adaptive traffic signal control, and especially a novel cooperative approach is introduced, when considering neighboring intersections. We consider not only the adjacent intersection's last reward as a Q-function but also its state and action as state. This method has the advantage of considering only vehicles from adjacent intersections that enter an intersection. The proposed method was evaluated on roads in Icheon City, and the results show that it reduces waiting time and carbon dioxide emissions.

INDEX TERMS Traffic signal control, deep reinforcement learning, multi-intersection, intelligent transportation systems, cooperative strategy, carbon dioxide emissions, greenhouse gases.

I. INTRODUCTION

The year 2023 was the hottest on record in 174 years of observations, a year of extremes, with global average near-surface temperatures 1.45 ± 0.12 °C higher than the pre-industrial baseline [1]. Simultaneously, record-breaking heatwaves, wildfires, droughts and floods have wreaked havoc worldwide, upending everyday life for millions and inflicting many billions of dollars in economic losses [2], [3], [4], [5], [6], [7]. If greenhouse gases continue to be emitted at the current levels, the global average temperature is projected to rise by more than 1.5 °C by 2040. Consequently, the frequency of extreme heatwaves is expected to increase 8.6 times, intense rainfall 1.5 times, and droughts twice as much due to climate change [8]. Research results showed that “approximately 75% of extreme abnormal climate events are currently related to climate change caused by carbon emissions” and that “by

2030, the world will face about 560 severe disasters per year, or an average of about 1.5 per day” [9].

Recognizing the gravity of environmental problems, the EU announced plans to reduce net greenhouse gas emissions by more than 55% compared with the 1990 levels by 2030 and achieve carbon-neutral by 2050 [10]. Countries worldwide are focusing on developing carbon reduction plans. Korea has announced Net-Zero scenarios for each sector and is making significant efforts toward carbon neutrality [11]. Korea's greenhouse gas emissions in 2022 were 725.744 Mt CO₂ eq/year, with carbon dioxide accounting for 87.6% [12]. It has the 11th highest carbon dioxide emissions among 210 countries, with emissions from the transportation sector accounting for 107.365 Mt CO₂, or 17% of the total. Traffic congestion has become a daily occurrence in Korea's urban areas, where the number of cars is unusually high in relation to the area, resulting in enormous economic and time losses, as well as serious traffic accidents and air pollution problem. Traffic congestion at signalized intersections results in vehicle idling, where engines remain running without movement,

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose^{ID}.

producing air pollutants at rates up to four times higher than during normal driving [13], [14]. Addressing road congestion could reduce vehicle idling, leading to an estimated annual reduction of 1.96 tons of pollutant emissions [15]. To solve traffic congestion, capacity increases such as road expansion are required; however, capacity increases require a significant amount of time and financial resources [16]. Therefore, the intersection traffic signal control problem (ITSCP) has been emphasized as a means of reducing traffic congestion in urban areas and making efficient use of the limited capacity of roads [17], [18].

Currently, most roads in Republic of Korea use fixed-time signal control. The fixed-signal model is an operating system that repeats preplanned signal patterns over a set period of time [19]. However, this fixed-signal model is limited in its ability to respond flexibly to real-time traffic changes [20], [21]. To address these limitations, recent research has focused on adaptive traffic signal control (ATSC) [22], [23], [24], [25], [26], [27]. Traffic signal control research using multi-agent reinforcement learning (MARL) consists primarily of independent traffic signal control research and cooperative traffic signal control research. Although independent signal control research has been actively conducted, applying independent methods to the real-world traffic networks, which are complex and influenced by adjacent intersections, has limitations in problem solving [28], [29]. Therefore, recent research on traffic signal control has considered adjacent intersections.

Several studies have proven that cooperative signal control considering adjacent intersections have shown good results in solving traffic problems. Most signal control studies aim to reduce vehicle waiting time or vehicle queue length. Reference [30] learned at a 4×4 intersection by accounting for the difference in average waiting time between time t and time $t+1$. The number of vehicles in each lane was input as a state, and the green light direction was selected from four designated directions without maintaining the signal order. Each episode is performed for 36,000 seconds, and the optimizer used is Adam. Reference [31] trained to minimize the standard deviation of queue length at a 2×3 intersection. This ultimately aimed to equalize drivers' individual wait times and increase road user satisfaction. The state used the number of halting in all incoming lanes in percentage format, the number of current intersections in binary format, and the current intersections' number to determine whether to give the green light north-south or east-west directions. Each episode is performed for 1,800 seconds, and the optimizer used is Adam. Reference [32] continued with learning by rewarding changes in the average queue length of vehicles at the intersection between times t and $t+1$. After dividing the lane into lengths of a certain size, the occupancy of each space and the vehicle's speed were stored in a matrix state, and one of the three or four specified phases was selected as an action. Each episode is performed for 4,500 seconds. Most cooperative signal control studies, including the aforementioned research,

consider neighboring intersections through Q-functions and adopt an approach where green signals are directly assigned without maintaining the sequence of signal phases. However, real-world traffic signals operate under constraints regarding the sequence and cycle of signal phases. Modifying phase sequences in conventional systems may influence intersection delay and safety [33]. Additionally, without restrictions on phase sequences, certain phases may be repeated indefinitely, causing vehicles in other lanes to experience prolonged delays [34]. Therefore, in this study, cooperative control of intersection traffic signals based on reinforcement learning is proposed to minimize carbon dioxide emissions, contributing to greenhouse gas reduction. The proposed model enhances performance by improving the cooperation mechanism to share the state of adjacent intersections through not only the Q-function but also the state. Furthermore, an action function reflecting realistic constraints was defined to ensure drivers' comprehension and avoid confusion.

The remainder of this paper is organized into four major sections. Section II explains DQN, one of the most widely used algorithms in adaptive traffic signal control research. Section III describes the construction environment and cooperative method used in this study. Section IV presents the results of an experiment that involved configuring a real intersection environment with the SUMO micro-traffic simulator. Finally, Section V provides concluding remarks.

II. REINFORCEMENT LEARNING

Q-learning is a reinforcement learning algorithm based on the temporal difference, which searches for an optimal policy using an action-value function [35]. Q-learning estimates $q_\pi(s, a)$, which is the value of taking action a based on a certain action strategy π in a given situation s . Q-learning's action-value function update formula is as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (1)$$

where s_t denotes the state at time t , a_t denotes the action taken at time t , r_t denotes rewards received for actions taken at time t , γ denotes the discount factor for long-term/short-term consideration of rewards with values between 0 and 1, and α denotes the learning rate used for Q-learning update, and a' indicates the best action with the largest Q-value at s_{t+1} . Through this update process, the Q-function considers the reward value of any future situation s_t into the future [36]. The Q-learning algorithm stores the Q-value associated with each state-action pair in a table. Therefore, it is also called tabular Q-learning [37]. If the agent continues to update the state-action pairs infinitely according to (1), it can converge to the optimal value [35].

Tabular Q-learning shows good performance in problems where small-scale discrete states and actions are defined; however, generalization problems are difficult to apply in real-world problems where large-scale continuous states and

actions are defined with the rapid increase in computation time [38], [39]. To address this problem, Q-learning with neural networks has been proposed, and a deep Q-learning network (DQN) algorithm was developed [40], [41]. A DQN is a popular reinforcement learning algorithm, and numerous studies have applied it to solve adaptive traffic signal control problems [42], [43], [44]. In a DQN, instead of individually estimating the Q-value of each state–action pair, a deep neural network is used as a function approximator that maps states to Q-values. These functional approximations allow the use of larger continuous-state spaces [45]. Equation (2) expresses the loss function of DQN as follows:

$$MSE(\theta_i) = \frac{1}{m} \sum_{t=1}^m (r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta_i^*) - Q(s_t, a_t; \theta_i)) \quad (2)$$

where m denotes the batch size, s_t denotes the state at time t , a_t denotes the action taken at time t , r_t denotes rewards received for actions taken at time t , γ denotes the discount factor for long-term/short-term consideration of reward value, and θ denotes the parameter used when estimating the Q-function. Deep Q Networks, which are Q-functions with deep learning, are optimized to minimize the loss function [46].

After DQN, extended methods such as double DQN, dueling DQN, and prioritized experience replay were introduced. The Double DQN algorithms propose to select the action on the basis of the Online Q Network but to use the values of the target state-action value corresponding to this particular state-action from the Target Q Network. This algorithm can reduce the bias problems that occur in the expected reward predictions of traditional Q-functions [47], [48]. Dueling DQN's algorithm divides the neural network into value and advantage networks. This algorithm achieves faster learning speeds because the advantage function focuses only on action values [49], [50]. In this study, an extended DQN, which combines a double DQN, dueling DQN, and prioritized experience replay, was applied to traffic signal control.

III. METHODOLOGY

A. STATE

In reinforcement learning, an agent recognizes the state based on the information it observes in the environment, which becomes the basis for the agent to select its actions [51]. TSC problems largely use two state–space representations: in the first case, the state representation is vector-based [52], [53], [54], and in the second case, it is a snapshot representation [55], [56]. In this study, we used a vector-based state expression based on the information that can be collected, which has become more diverse owing to the development of sensing and V2X technologies. Specifically, this study used the number of vehicles and the average speed, which are variables commonly used to understand the situation at an intersection, as well as whether the signal is on and the elapsed time of the current green signal, which can provide information regarding traffic lights.

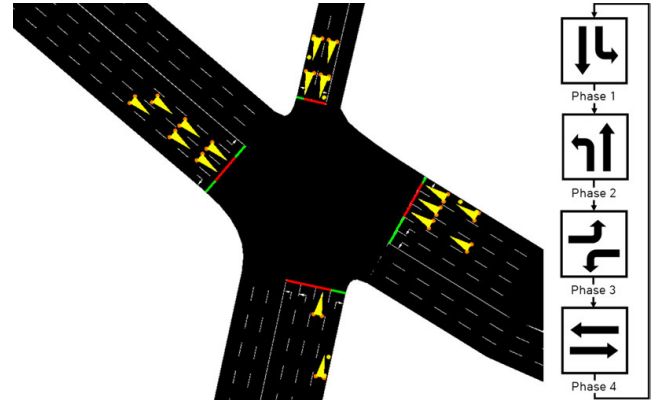


FIGURE 1. Example of intersection state and intersection signal phase.

Thus, the state expression of the intersection is $S_t = \{P_t, d_t, N_t, V_t\}$. P_t indicates whether each phase is green or not in binary form. Green light has a value of 1, red light and yellow light have a value of 0. As shown in Fig. 1, green light is on in Phase 3 and red light is on in the rest; therefore, it has a value of $P_t = [0, 0, 1, 0]$. Subsequently, d_t indicates the elapsed time of the currently turned on phase. As shown in Fig. 1, because 10 s have passed since Phase 3 turned green, $d_t = 10$. Third, N_t is the number of vehicles affected by each signal. We used the number of vehicles counted by signal, not the number of vehicles counted by lane. In reinforcement learning, the state should encapsulate adequate information necessary for decision-making [57]. However, an increase in the complexity or size of the state space can lead to slower learning speeds and higher memory requirements, posing challenges to system efficiency [58]. This method can produce better results than lane-based aggregation methods by reducing unnecessary dimensions while retaining important traffic information. There are four vehicles in Phase 1, one vehicle in Phase 2, all vehicles have left and zero vehicles in Phase 3, and 11 vehicles in Phase 4. Therefore, $N_t = [4, 1, 0, 11]$. Right turns were not considered because they were not controlled by signals. Finally, V_t is the average speed of the vehicles aggregated from N_t . Therefore, the state shown in Fig. 1 at t s can be expressed as $S_t = \{0, 0, 1, 0, 10, 4, 1, 0, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$.

B. ACTION

Because reinforcement learning agents control signals through actions, action design is important for creating a model applicable to real-world scenarios [59]. Existing reinforcement learning-based TSC studies largely use two action expressions. First, to select the most appropriate display among the set of signal patterns, that is, to directly select the signal to provide (dynamic phase selection type), and second, to decide whether to move on to the next signal or maintain the current signal while maintaining the specified phase sequence (binary action selection type). Whereas both approaches have been actively studied in research considering independent intersections, cooperative signal control

research is dominated by dynamic phase selection studies. Although a method that does not maintain the phase sequence can achieve better results because it provides the most appropriate signal based on the intersection situation, it can confuse drivers familiar with the signal patterns of existing intersections, thereby increasing the possibility of an accident [34]. Therefore, in this study, the phase sequence was maintained, and based on the state, the phase was maintained ($A_t = 0$) or changed to the next phase ($A_t = 1$). In other words, the set of possible action operations was $A_t = \{0, 1\}$. Additionally, the maximum green constraint was used to prevent certain vehicles from waiting endlessly by continuously providing green signals to lanes with many vehicles, and a minimum green constraint was applied to protect the pedestrian signal and prevent too frequent changes in the phase (signal flickering). Therefore, if the minimum green time of Phase 3 in Fig. 1 is set to 15 s, the action of changing to the next signal ($A_t = 1$) cannot be selected at this point, when only 10 s has been maintained.

C. REWARD

A reward is the value an agent receives when choosing an action; therefore, it is important to define the reward appropriately [60]. In reinforcement learning, the agent aims to maximize long-term accumulated rewards through continuous interaction with the environment; therefore, learning results may vary based on how the reward value is defined [61], [62]. The goal of this study was to reduce carbon dioxide emissions from vehicles. However, when we attempted to minimize carbon dioxide emissions, we discovered that many cars ended up sitting on the roads. That is because a vehicle's carbon dioxide emissions are related to its acceleration, and the agent decided that stopping more vehicles on the road by not giving them a green signal would be a way to further reduce emissions than if they were running. However, keeping many cars on the road to reduce carbon dioxide emissions does not solve this problem and is not the desired outcome. Therefore, the main goal of this study was to reduce carbon dioxide emissions from vehicles without unrealistically impeding traffic at intersections. According to [63] and [64], vehicle carbon dioxide emissions were affected by acceleration and deceleration; therefore, emissions on roads that included signalized intersections appeared to be greater than those on roads that did not. In addition, vehicle emissions were particularly high around intersections where more vehicles stopped. We defined the reward as minimizing the carbon dioxide emissions from waiting vehicles. In other words, our reward was $R_t = -\sum E_t$.

D. COOPERATION OF ADJACENT INTERSECTION

Traffic flow at an intersection is affected not only by the traffic conditions of the intersection itself but also by the traffic conditions of adjacent intersections. The interconnection of signals can reduce the greenhouse gases emissions of vehicles [65]. Therefore, cooperative signal control that

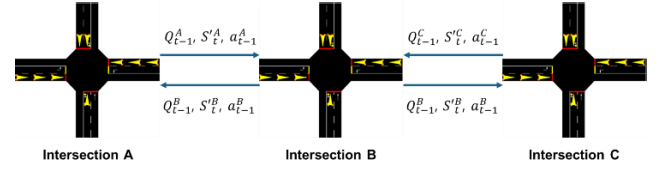


FIGURE 2. Cooperative algorithm process with adjacent intersections.

considers neighbors can solve traffic congestion and carbon dioxide emission problems more effectively [66], [67]. In this study, the cooperation mechanism was improved by incorporating state-based integration into the existing Q-function-based approach. In the proposed approach, as shown in Figure 2, the agent's action selection is influenced not only by its own state, action, and reward values but also by the state, action, and reward values of adjacent intersections. Therefore, ATSC-based systems can equalize the traffic flow between adjacent intersections while improving the overall performance of road networks. In conclusion, we update the Q-function of each agent by considering the previous reward values of adjacent intersections as follows:

$$Q_{t+1}^i(s_t^i, a_t^i) \leftarrow Q_t^i(s_t^i, a_t^i; \theta_i) + \alpha(t)[r_t + \gamma \max_{a'} Q_t^i(s_{t+1}^i, a'; \theta_i^*) - Q_t^i(s_t^i, a_t^i; \theta_i)] + \frac{1}{N_{adj}} \sum_{j \in N_{adj}} r_{t-1}^j \quad (3)$$

where N_{adj} denotes the number of adjacent intersections, θ_i and θ_j denote the parameters of evaluation network of intersections i and j , respectively. It also monitors not only the conditions of the current intersection but also the states and actions of adjacent intersections. In this study, the state of each agent considering the adjacent intersections is defined as follows:

$$S_t^i = \{S_t^i, S_t^{i,j}, a_{t-1}^j\} \quad (4)$$

where the intersection state (S_t^i) uses $S_t = \{P_t, d_t, N_t, V_t\}$ described in 3.A state; however, when considering the state of adjacent intersections, it is a slightly modified state ($S_t^{i,j}$). The difference between S and S' is $S'_t = \{P_t, d_t, N'_t, V'_t\}$, and when receiving the state (S') of a adjacent intersection, the number of vehicles entering an intersection only considers information. For example, to the right of the intersection in Fig. 1, there is an intersection shown in Fig. 3.

At the intersection shown in Fig. 3, lanes 1 and 3 are right-turn lanes, lanes 2 and 8 are left-turn lanes, and lanes 4, 5, 6, and 7 are straight lanes. Among these, the only lanes heading toward the intersection in Fig. 1 are lanes 1, 4, and 5. Lane 1 was excluded from the analysis because it was a right turn and not controlled by traffic lights. Therefore, when the intersection in Fig. 3 is considered for its own signal, it corresponds to $N'_t = [2, 0, 14]$ and $V'_t = [0, 0, 1.3]$, but when shared with the intersection in Fig. 1, it corresponds to $N'_t = [0, 0, 10]$ and $V'_t = [0, 0, 0]$. In other words, the state of the intersection in Fig. 1 that is received

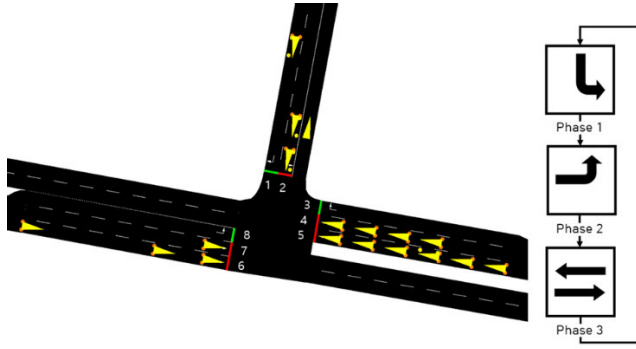


FIGURE 3. Example of adjacent intersection state and intersection signal phase.

by the intersection in Fig. 3 is $S'_t = \{P_t, d_t, N'_t, V'_t\} = \{0, 1, 0, 5, 0, 0, 10, 0, 0, 0\}$.

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL ENVIRONMENTS

This study conducted experiments using Simulation of Urban Mobility (SUMO), an open-source simulator widely used in traffic signal research. References [68] and [69]. Using SUMO, we can calculate the movements of individual vehicles and implement dynamic traffic signal control. As shown in Fig. 4, the real-time traffic situation implemented in SUMO was transmitted to the Python-based reinforcement learning model, which determined an action (signal control) through the received state, which was transmitted back to SUMO to change the traffic flow.

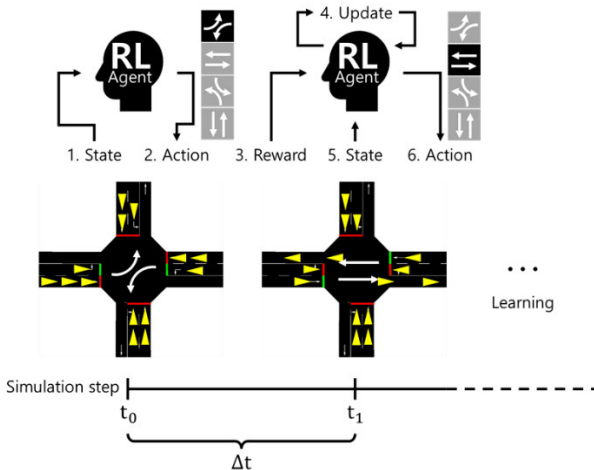


FIGURE 4. Reinforcement learning-based signal optimization model learning process using SUMO.

B. SCENARIOS

The experiment was conducted at six partially contiguous intersections on Gyeongchung-daero in Icheon-si, Gyeonggi-do, Republic of Korea, as shown in Fig. 5(a), and consisted of three- or four-way intersections, as shown in Fig. 5(b). This road passes through the downtown area of Icheon-si,

Gyeonggi-do, where commercial and residential areas are concentrated, and signals of different cycles are used during off-peak and peak times. Therefore, in this study, we divided the simulation into two cases based on the off-peak hours of 15:00-16:00 and the peak hours of 18:00-19:00. Table 1 shows the traffic volume during off-peak hours and Table 2 shows the traffic volume during peak hours. The study conducted experiments based on observed real-world traffic when evaluating the trained model. If we look at the traffic volume, Gyeongchung-daero, which runs from intersections 1 to 6, is mainstream.

Although the numbers vary by time of day, at both the off-peak and peak hours, between 60% and more than 90% of vehicles at each intersection pass on the signal to go straight in the mainstream.

C. RESULTS

The proposed model was trained for over 150 iterations. The termination condition for each episode was the passage of approximately 8,500 vehicles. Batch size $|B|$ was 32, and the replay memory size was 1000. The optimizer's update rule was Adam, and the learning rate was 0.0001. To allow the agent to change its role from exploration to exploitation, epsilon decreased throughout the training process from a starting value of 0.9 to an ending value of 0.03. The parameters used in the simulations are enumerated in Table 3. These values were experimentally determined.

Fig. 6 shows the learning process at each traffic light. These plots were obtained by running 150 episodes. We can observe that as the number of episodes increased, the value of the reward earned by each traffic light increased and gradually converged. Therefore, we can conclude that the trained model was able to predict the outcome appropriately.

In this experiment, to test the proposed model, the off peak/peak-time fixed signals currently applied on the road, a cooperative model that consider adjacent signals only by the Q-function and a model without constraints on signal sequence as well as minimum and maximum green times were used as comparison models. Considering the convergence speed and stability of the reinforcement learning algorithm, appropriate information should be selected to define the states [34], [70]. Cooperative approaches that consider adjacent intersections through states, such as the proposed model, have the disadvantage of complicating the state representation. However, advances in DQNs allow us to consider more complex states than ever before, and this state cooperation allows us to consider only the flows coming into our intersection from neighbors. Therefore, to evaluate the method proposed in this study, a model that does not consider neighbors in the state but only considers neighbors in the Q-function was attached as a comparison model 1. The Q-function update equation for comparison model that considers neighbors was the same as that of the proposed model. Comparison model 2 was attached to evaluate the effects of signal constraints. This model adopts a dynamic phase selection approach, which directly determines signal allocation

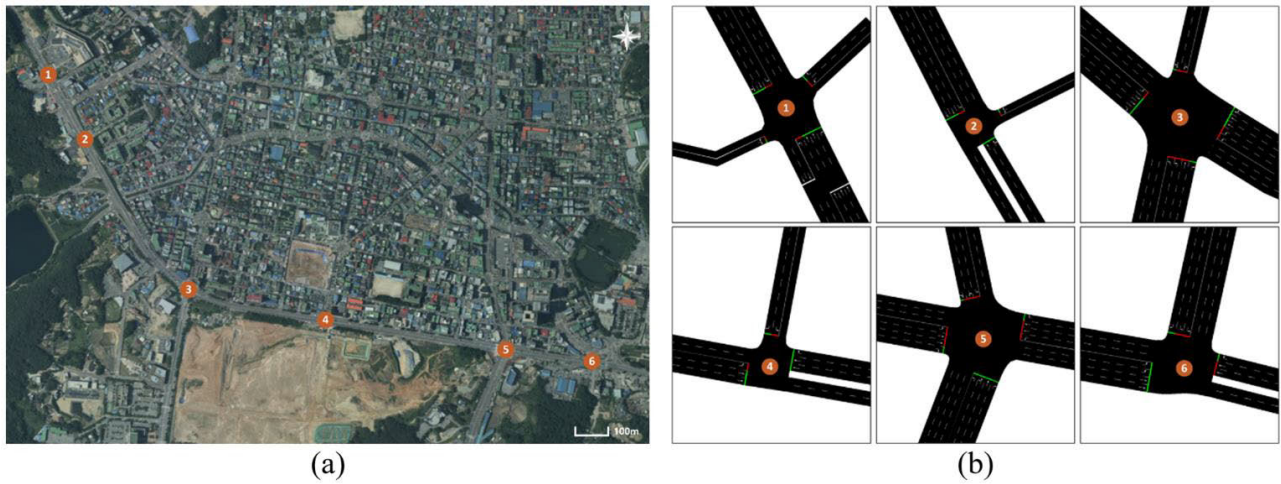


FIGURE 5. Reinforcement learning-based signal optimization model learning process using SUMO.

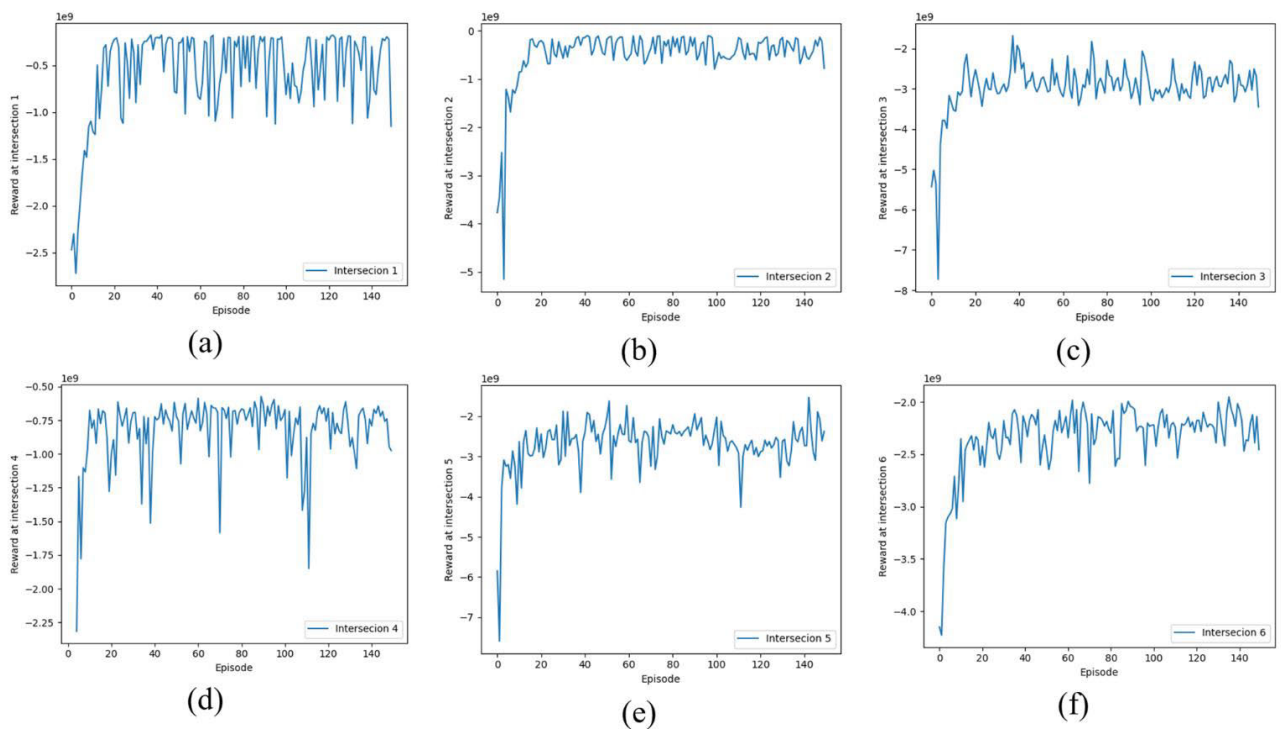


FIGURE 6. The learning process at each intersection.

without preserving the predefined sequence. Additionally, the constraints on minimum and maximum phase durations were removed. Whereas it is a commonly adopted approach in cooperative traffic signal control research and is expected to achieve higher performance, it has limitations in real-world applications. We used cumulative waiting time and carbon dioxide emissions as metrics to evaluate the performance of the proposed algorithm.

1) OFF-PEAK

Fig. 7 shows a comparison of the cumulative waiting times during off-peak hours. Compared to the fix and comparison

1 models, the proposed model showed the best results for all traffic lights, except for the second. In particular, we can see that the proposed model shows noticeably better results compared to fixed signals. The proposed approach reduced the waiting time by approximately 54% on average (250,779 in total) for the fixed method and approximately 18% on average (123,190 in total) for the comparison models. This means that existing fixed signals are not suitable for controlling the flow of dynamic vehicles.

Fig. 8 presents a comparison of carbon dioxide emissions during off-peak hours. Compared to the fix and comparison 1 models, the proposed model showed the best results for

TABLE 1. Off-peak time traffic volume at each intersection (The numbers next to the arrows indicate traffic volume).

Directions to Chungju ↔ Direction to Seoul	
Division	Traffic volume by direction
1	
2	
3	
4	
5	
6	

TABLE 2. Peak time traffic volume at each intersection (The numbers next to the arrows indicate traffic volume).

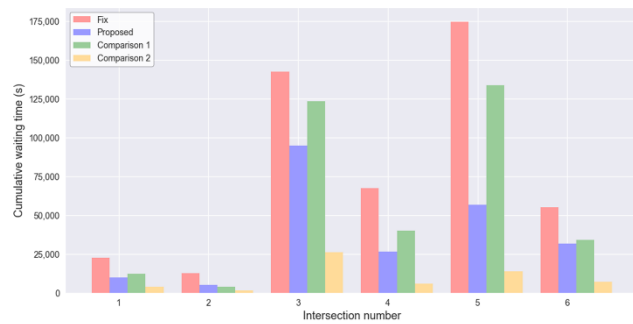
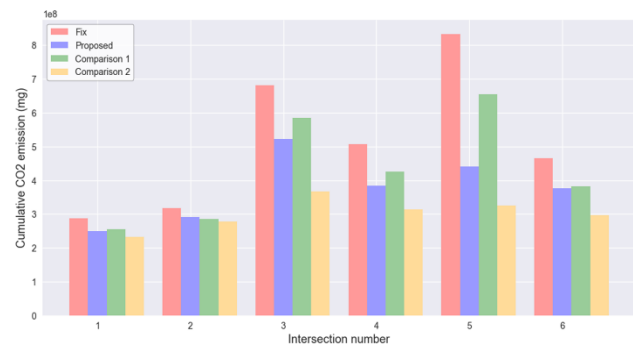
Directions to Chungju ↔ Direction to Seoul	
Division	Traffic volume by direction
1	
2	
3	
4	
5	
6	

all traffic lights, except for the second. At traffic light 2, the proposed model recorded slightly higher emissions than the comparison model 1, but at the rest of the traffic lights, it performed better. The proposed model reduced carbon dioxide

emissions by an average of approximately 23% (829,369,598 in total) compared to the fixed method and an average of approximately 9% (325,579,461 in total) compared with the comparison model.

TABLE 3. Reinforcement learning-based signal optimization model learning process using SUMO.

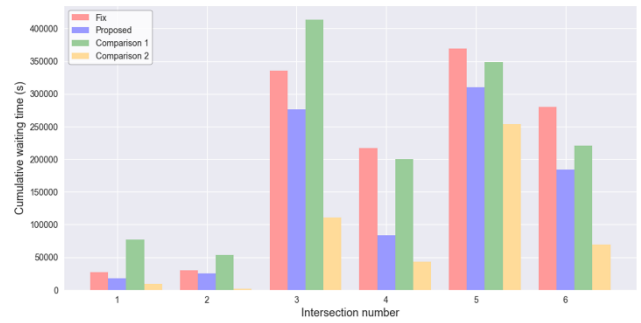
Parameters	Values
Batch size($ B $)	32
Replay memory size	1000
Discount factor(γ)	0.9
Learning rate(α)	0.0001
Episode	150
Starting epsilon	0.9
Ending epsilon	0.03
Epsilon decay	20000

**FIGURE 7.** Compare cumulative waiting times at each intersection at off-peak times.**FIGURE 8.** Compare carbon dioxide emissions at each intersection at off-peak times.

In both cumulative waiting times and carbon dioxide emissions, Comparison 2 outperformed the proposed model due to the removal of constraints, which allowed immediate signal allocation to roads with higher traffic volumes. However, the random signal sequences may confuse drivers, increasing the risk of accidents and potentially causing infinite delays for certain lanes.

2) PEAK

Fig. 9 shows a comparison of the cumulative waiting times during peak hours. Compared to the fix and comparison 1 models, the proposed model showed the best results for all traffic lights. The proposed approach reduced the waiting time by approximately 30% on average (360,627 in total)

**FIGURE 9.** Compare cumulative waiting times at each intersection at peak times.

for the fixed method and approximately 41% on average (416,632 in total) for the comparison models.

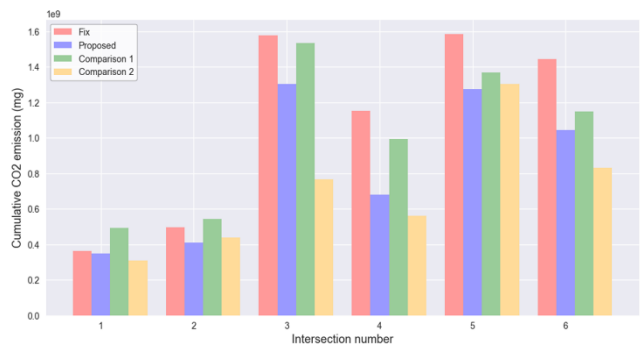
**FIGURE 10.** Compare carbon dioxide emissions at each intersection at peak times.

Fig. 10 presents a comparison of carbon dioxide emissions during peak hours. Compared to the fix and comparison 1 models, the proposed model yielded the best results for all the traffic lights. The proposed model reduced carbon dioxide emissions by an average of approximately 21% (1,560,811,796 in total) compared to the fixed method and an average of approximately 19% (1,023,801,766 in total) compared with the comparison model. Whereas Comparison model 2 demonstrates better performance than the proposed model in terms of cumulative waiting times and carbon dioxide emissions. Nonetheless, its random signal sequence is impractical for real-world application. Therefore, we can see that the proposed model provides better signal control than the fixed signal and comparison 1 models in both off-peak and peak hours.

Fig. 11 shows the average speeds of all the vehicles in the fixed and proposed models during peak hours. Since the velocity comparison graph shows similar shapes in off-peak and peak hours, peak hours are used as a representative example. As shown in Fig. 11, the average speed of the proposed model was similar to that of the fixed model; however, the speed variation was relatively small. Because carbon dioxide emissions are highly related to the acceleration and deceleration of vehicles, it is expected that the proposed model, with relatively less variation in vehicle speed, will perform well in terms of carbon dioxide emissions.

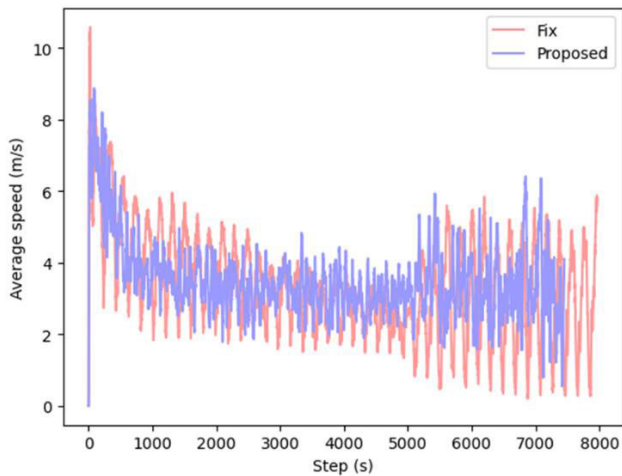


FIGURE 11. Change in average speed of vehicles over simulation time in fixed and proposed models.

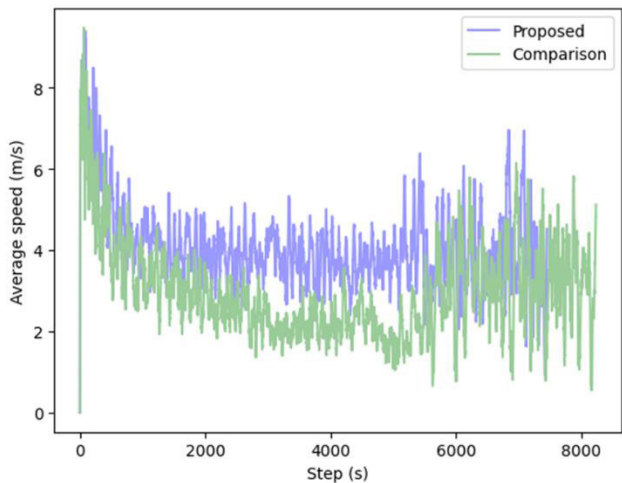


FIGURE 12. Change in average speed of vehicles over simulation time in comparison and proposed models.

Fig. 12 shows the average speeds of all the vehicles in the comparison 1 and proposed models during peak hours. The proposed model has a higher average speed of vehicles than the comparison model, and the episode in the proposed model ended earlier than that in the comparison model. The evaluation in this study is conducted based on the number of vehicles observed at real-world intersections. Therefore, unlike previous studies that define a fixed duration as the termination condition for an episode, this study uses the departure of a specified number of vehicles as the termination condition. Each simulation concludes when the predetermined number of vehicles has exited the roadway. Therefore, the time required for an episode to run may vary between the models. The comparison model would have taken longer for a given number of vehicles to pass through the road than the proposed model, and it would have been expected that the carbon dioxide emissions from the vehicles would have accumulated during that time. Therefore, it can be expected that the proposed model, which took a relatively short time

for the vehicle to exit the road, will perform well in terms of carbon dioxide emissions.

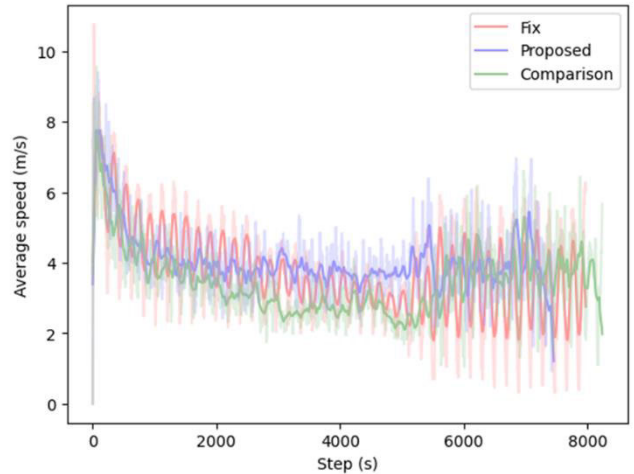


FIGURE 13. Smoothing representation of the change in average speed of vehicles over simulation time in three models.

Fig. 13 shows the smoothed average speed of the overall vehicles for the three models. Whereas the factors that are expected to have contributed significantly to the performance difference with the proposed model are different, we can see that the common thread is that the episodes in the proposed model ended earlier than both models. In fact, when we compared the travel time of vehicles by route, we identified that while not all routes had less travel time, the proposed model allowed people to reach their destination on average approximately 42 s faster than the fixed model and approximately 48 s faster than the comparison model. This is not a small difference, considering that the distance from intersections 1 to 6 was approximately 5 minutes (when traffic was not blocked). In other words, from the perspective of vehicles, people reached their destinations faster in the proposed model than in the comparison model.

V. CONCLUSION

In this study, we propose a multi-intersection signal control model that uses a novel cooperative approach to reduce traffic congestion and carbon dioxide emissions. In the proposed model, we use adjacent intersections to improve overall performance by allowing agents to share their states, actions, and rewards. At each intersection, an agent's action is determined by considering not only the state of its own intersection but also the states and actions of its neighbors. In addition, the estimated Q-value considers the last reward received from the neighbors. Experiments on six contiguous roads in Icheon City show that our method outperforms fixed-signal and comparison model 1 in terms of cumulative waiting time and carbon dioxide emission metrics. The amount of emission saved by the proposed method is expected to significantly reduce a huge amount of carbon dioxide emissions when accumulated over a month or year.

REFERENCES

- [1] *State of the Global Climate 2023*, World Meteorological Org., Geneva, Switzerland, Mar. 2024.
- [2] A. Voiland. (2023). *Tracking Canada's Extreme 2023 Fire Season*. NASA Earth Observatory, Washington, DC, USA. [Online]. Available: <https://earthobservatory.nasa.gov/images/151985/tracking-canadas-extreme-2023-fire-season>
- [3] N. Yousif. (2023). *Canada's 2023 Wildfires Emitted More Carbon Than Most Countries*. BBC, London, U.K. [Online]. Available: <https://news.skocplant.com/plant-tomorrow/13463/>
- [4] J. P. Ferreyra. (2023). *Canada's Blazing Inferno: 2023 Wildfires Swept Through an Area Larger Than Florida*. ELPAIS. [Online]. Available: <https://english.elpais.com/climate/2023-12-08/canadas-blazing-inferno-2023-wildfires-swept-through-an-area-larger-than-florida.html>
- [5] N. Lopez. (2023). *2023: A Year of Intense Global Wildfire Activity*. ECMWF. [Online]. Available: <https://atmosphere.copernicus.eu/2023-year-intense-global-wildfire-activity>
- [6] E. Cassidy. *A Deluge in Greece*. NASA Earth Observatory. [Online]. Available: <https://earthobservatory.nasa.gov/images/151807/a-deluge-in-greece>
- [7] R. Davies. *Greece—'Unimaginable Amounts of Water' As Floods and Rain Continue*. FloodList. [Online]. Available: <https://floodlist.com/europe/greece-floods-september-2023>
- [8] *2023 Climate Change Report*, Meteorological Admin., Daejeon, South Korea, Mar. 2024.
- [9] *GAR Special Report 2023: Mapping Resilience for the Sustainable Development Goals*, UNDRR, Geneva, Switzerland, Jun. 19, 2024.
- [10] *The European Green Deal*, Eur. Commission, Brussels, Belgium, Nov. 2019.
- [11] H. Inchang et al., "Long-term strategy and sectoral approaches of Seoul for achieving carbon neutrality by 2050," Seoul Inst., Seoul, South Korea, Tech. Rep., Jun. 2020.
- [12] M. Crippa et al., "GHG emissions of all world countries 2023," Publications Office Eur. Union, Luxembourg, Europe, Tech. Rep., 2023, doi: [10.2760/953322](https://doi.org/10.2760/953322).
- [13] J. Jin and J. Jin, "A study on the effect of traffic congestion on particulate matter concentration in Seoul : Big data approach," *J. Korea Planning Assoc.*, vol. 56, no. 1, pp. 121–136, Feb. 2021, doi: [10.17208/jkpa.2021.02.56.1.121](https://doi.org/10.17208/jkpa.2021.02.56.1.121).
- [14] *White Paper on the Survey on the Status of Idling Vehicles and the Legislation of Prohibition of Idling Vehicles Project*, Green Transp., Seoul, South Korea, 2001.
- [15] D. L. Eom and S. W. Park, "Effectiveness of idling restriction in signalized intersections," *J. Transp. Res.*, vol. 20, no. 2, pp. 153–161, Jun. 2013, doi: [10.34143/jtr.20.2.153](https://doi.org/10.34143/jtr.20.2.153).
- [16] A. University, E. Han, S. Park, H. Jeong, C. Lee, and I. Yun, "The development of an algorithm for the optimal signal control for isolated intersections under V2X communication environment," *J. Korea Inst. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 90–101, Dec. 2016.
- [17] T. A. Haddad, D. Hedjazi, and S. Aouag, "A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105019, doi: [10.1016/j.engappai.2022.105019](https://doi.org/10.1016/j.engappai.2022.105019).
- [18] D. H. Kim and O. R. Jeong, "A study on cooperative traffic signal control at multi-intersection," *J. IKEEE*, vol. 23, no. 4, pp. 1381–1386, Dec. 2019.
- [19] J. W. Gu, D. G. Kim, M. H. Lee, and C. M. Jun, "Intersection signal model based on reinforcement learning to minimize waiting time," *J. Korean Soc. Geospatial Inf. Sci.*, vol. 28, no. 4, pp. 59–67, Dec. 2020, doi: [10.7319/kogsis.2020.28.4.059](https://doi.org/10.7319/kogsis.2020.28.4.059).
- [20] Y. Jaehong and J. Yookang, "Simulation of traffic signal control with adaptive priority order through object extraction in images," *J. Korea Multimedia Soc.*, vol. 11, no. 8, pp. 1051–1058, Aug. 2008.
- [21] W.-K. Hong and W.-S. Shim, "Traffic signal control scheme for traffic detection system based on wireless sensor network," *J. Inst. Control, Robot. Syst.*, vol. 18, no. 8, pp. 719–724, Aug. 2012, doi: [10.5302/J.ICROS.2012.18.8.719](https://doi.org/10.5302/J.ICROS.2012.18.8.719).
- [22] J. Gao, Y. Shen, J. Liu, M. Ito, and N. Shiratori, "Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network," 2017, *arXiv:1705.02755*.
- [23] Y. Feng, K. L. Head, S. Khoshmaghani, and M. Zamanipour, "A real-time adaptive signal control in a connected vehicle environment," *Transp. Res. C, Emerg. Technol.*, vol. 55, pp. 460–473, Jun. 2015, doi: [10.1016/j.trc.2015.01.007](https://doi.org/10.1016/j.trc.2015.01.007).
- [24] M. Guo, P. Wang, C.-Y. Chan, and S. Askary, "A reinforcement learning approach for intelligent traffic signal control at urban intersections," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 4242–4247, doi: [10.1109/ITSC.2019.8917268](https://doi.org/10.1109/ITSC.2019.8917268).
- [25] F.-X. Devailly, D. Larocque, and L. Charlin, "IG-RL: Inductive graph reinforcement learning for massive-scale traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7496–7507, Jul. 2022, doi: [10.1109/TITS.2021.3070835](https://doi.org/10.1109/TITS.2021.3070835).
- [26] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020, doi: [10.1109/TITS.2019.2901791](https://doi.org/10.1109/TITS.2019.2901791).
- [27] W. Genders and S. Razavi, "Using a deep reinforcement learning agent for traffic signal control," 2016, *arXiv:1611.01142*.
- [28] A. Boukerche, D. Zhong, and P. Sun, "A novel reinforcement learning-based cooperative traffic signal system through max-pressure control," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1187–1198, Feb. 2022.
- [29] M. Abdoos, "A cooperative multiagent system for traffic signal control using game theory and reinforcement learning," *IEEE Intell. Transp. Syst. Mag.*, vol. 13, no. 4, pp. 6–16, May 2021, doi: [10.1109/MITS.2020.2990189](https://doi.org/10.1109/MITS.2020.2990189).
- [30] D. Kim and O. Jeong, "Cooperative traffic signal control with traffic flow prediction in multi-intersection," *Sensors*, vol. 20, no. 1, p. 137, Dec. 2019, doi: [10.3390/s20010137](https://doi.org/10.3390/s20010137).
- [31] M. Kolat, B. Kővári, T. Bércsi, and S. Aradi, "Multi-agent reinforcement learning for traffic signal control: A cooperative approach," *Sustainability*, vol. 15, no. 4, p. 3479, Feb. 2023, doi: [10.3390/su15043479](https://doi.org/10.3390/su15043479).
- [32] H. Ge, Y. Song, C. Wu, J. Ren, and G. Tan, "Cooperative deep Q-learning with Q-value transfer for multi-intersection signal control," *IEEE Access*, vol. 7, pp. 40797–40809, 2019, doi: [10.1109/ACCESS.2019.2907618](https://doi.org/10.1109/ACCESS.2019.2907618).
- [33] J. H. Park and J. S. Huh, "Particle swarm optimization and SUMO based multi-intersection traffic signal optimization," in *Proc. KIIT Conf.*, Jeju, South Korea, Nov. 2023, pp. 347–350.
- [34] M. Pi, H. Lee, and M. Chung, "Reinforcement learning-based traffic signal control under real-world constraints," *J. KIISE*, vol. 48, no. 8, pp. 871–877, Aug. 2021.
- [35] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, May 1992.
- [36] V. Kosana, K. Teeparthi, S. Madasthu, and S. Kumar, "A novel reinforced online model selection using Q-learning technique for wind speed prediction," *Sustain. Energy Technol. Assessments*, vol. 49, Feb. 2022, Art. no. 101780, doi: [10.1016/j.seta.2021.101780](https://doi.org/10.1016/j.seta.2021.101780).
- [37] A. Applebaum, C. Dennler, P. Dwyer, M. Moskowitz, H. Nguyen, N. Nichols, N. Park, P. Rachwalski, F. Rau, A. Webster, and M. Wolk, "Bridging automated to autonomous cyber defense: Foundational analysis of tabular Q-learning," in *Proc. 15th ACM Workshop Artif. Intell. Secur.*, Nov. 2022, pp. 149–159, doi: [10.1145/3560830.3563732](https://doi.org/10.1145/3560830.3563732).
- [38] E. Van der Pol and F. A. Oliehoek, "Coordinated deep reinforcement learners for traffic light control," in *Proc. Learn., Inference Control Multi-Agent Syst.*, vol. 8, Barcelona, Spain, 2016, pp. 21–38.
- [39] H. Wang, M. Emmerich, and A. Plaat, "Assessing the potential of classical Q-learning in general game playing," in *Proc. Benelux Conf. Artif. Intell.*, Nov. 2018, pp. 138–150.
- [40] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.
- [41] J. Clifton and E. Laber, "Q-learning: Theory and applications," *Annu. Rev. Statist. Appl.*, vol. 7, no. 1, pp. 279–301, Mar. 2020, doi: [10.1146/annurev-statistics-031219-041220](https://doi.org/10.1146/annurev-statistics-031219-041220).
- [42] S. Mirbakhsh and M. Azizi, "Adaptive traffic signal safety and efficiency improvement by multi objective deep reinforcement learning approach," 2024, *arXiv:2408.00814*.
- [43] G. Yaobang, "Improving traffic safety and efficiency by adaptive signal control systems based on deep reinforcement learning," Ph.D. dissertation, Dept. Engineering and Computer Science, Univ. Central Florida, Orlando, FL, USA, 2020.
- [44] S. M. A. Shabestary and B. Abdulhai, "Deep learning vs. discrete reinforcement learning for adaptive traffic signal control," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, Nov. 2018, pp. 286–293, doi: [10.1109/ITSC.2018.8569549](https://doi.org/10.1109/ITSC.2018.8569549).
- [45] S. J. Yong, H. G. Park, Y. H. You, and I. Y. Moon, "Q-learning policy and reward design for efficient path selection," *J. Adv. Navigat. Technol.*, vol. 26, no. 2, pp. 72–77, Apr. 2022, doi: [10.12673/jant.2022.26.2.72](https://doi.org/10.12673/jant.2022.26.2.72).

- [46] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: [10.1038/nature14236](https://doi.org/10.1038/nature14236).
- [47] H. Hasselt, "Double Q-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1–9.
- [48] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, Mar. 2016, vol. 30, no. 1, pp. 1–7, doi: [10.1609/aaai.v30i1.10295](https://doi.org/10.1609/aaai.v30i1.10295).
- [49] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, New York, NY, USA, Jun. 2016, pp. 1995–2003.
- [50] M. Sewak, "Deep Q network (DQN), double DQN, and dueling DQN," in *Deep Reinforcement Learning*. Singapore: Springer, 2019, pp. 95–108.
- [51] S. S. Mousavi, M. Schukat, and E. Howley, "Traffic light control using deep policy-gradient and value-function-based reinforcement learning," *IET Intell. Transp. Syst.*, vol. 11, no. 7, pp. 417–423, Sep. 2017, doi: [10.1049/iet-its.2017.0153](https://doi.org/10.1049/iet-its.2017.0153).
- [52] X. Liang, X. Du, G. Wang, and Z. Han, "Deep reinforcement learning for traffic light control in vehicular networks," 2018, *arXiv:1803.11115*.
- [53] L. Li, Y. Lv, and F.-Y. Wang, "Traffic signal timing via deep reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 3, no. 3, pp. 247–254, Jul. 2016, doi: [10.1109/JAS.2016.7508798](https://doi.org/10.1109/JAS.2016.7508798).
- [54] S. Jung, S. Lee, and J. Kim, "The real-time signal control system using reinforcement learning considering priority signaling for emergency vehicle," *J. Korean Soc. Transp.*, vol. 39, no. 3, pp. 329–344, Jun. 2021.
- [55] K.-F. Chu, A. Y. S. Lam, and V. O. K. Li, "Traffic signal control using end-to-end off-policy deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7184–7195, Jul. 2022, doi: [10.1109/TITS.2021.3067057](https://doi.org/10.1109/TITS.2021.3067057).
- [56] D. Ma, B. Zhou, X. Song, and H. Dai, "A deep reinforcement learning approach to traffic signal control with temporal traffic pattern mining," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11789–11800, Aug. 2022, doi: [10.1109/TITS.2021.3107258](https://doi.org/10.1109/TITS.2021.3107258).
- [57] W. T. Scherer, S. Adams, and P. A. Beling, "On the practical art of state definitions for Markov decision process construction," *IEEE Access*, vol. 6, pp. 21115–21128, 2018, doi: [10.1109/ACCESS.2018.2819940](https://doi.org/10.1109/ACCESS.2018.2819940).
- [58] K. Fujita and H. Matsuo, "Multiagent reinforcement learning with the partly high-dimensional state space," *Syst. Comput. Jpn.*, vol. 37, no. 9, pp. 22–31, Jun. 2006, doi: [10.1002/scj.20526](https://doi.org/10.1002/scj.20526).
- [59] B. Seunggho and K. Seunghyun, "Reinforcement learning-based ATCS research trends," *KICS*, vol. 35, no. 12, pp. 3–7, Dec. 2018.
- [60] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1243–1253, Feb. 2019, doi: [10.1109/TVT.2018.2890726](https://doi.org/10.1109/TVT.2018.2890726).
- [61] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *Robotica*, vol. 17, no. 2, pp. 229–235, Mar. 1999, doi: [10.1017/S0263574799271172](https://doi.org/10.1017/S0263574799271172).
- [62] J. Sangchul, "The real-time traffic signal control system using reinforcement learning in V2I environment," M.S. thesis, Dept. Traffic Engineering, Univ. Seoul, Seoul, South Korea, 2019.
- [63] M. C. Coelho, T. L. Farias, and N. M. Rouphail, "Impact of speed control traffic signals on pollutant emissions," *Transp. Res. D, Transp. Environ.*, vol. 10, no. 4, pp. 323–340, Jul. 2005, doi: [10.1016/j.trd.2005.04.005](https://doi.org/10.1016/j.trd.2005.04.005).
- [64] S. L. Hallmark, I. Fomunung, R. Guensler, and W. Bachman, "Assessing impacts of improved signal timing as a transportation control measure using an activity-specific modeling approach," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1738, no. 1, pp. 49–55, Jan. 2000, doi: [10.3141/1738-06](https://doi.org/10.3141/1738-06).
- [65] L. Yunsok and O. Heungun, "Comparison of vehicle carbon emissions in expressway and national highway," *Int. J. Highw. Eng.*, vol. 13, no. 3, pp. 177–184, 2011.
- [66] K. L. Soon, J. M.-Y. Lim, and R. Parthiban, "Coordinated traffic light control in cooperative green vehicle routing for pheromone-based multi-agent systems," *Appl. Soft Comput.*, vol. 81, Aug. 2019, Art. no. 105486, doi: [10.1016/j.asoc.2019.105486](https://doi.org/10.1016/j.asoc.2019.105486).
- [67] L.-W. Chen and C.-C. Chang, "Cooperative traffic control with green wave coordination for multiple intersections based on the Internet of Vehicles," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1321–1335, Jul. 2017, doi: [10.1109/TSMC.2016.2586500](https://doi.org/10.1109/TSMC.2016.2586500).
- [68] G. Li, G. Song, and L. Wen, "Intelligent traffic light system for high priority vehicles," in *Proc. China Conf. Wireless Sensor Netw.*, Oct. 2019, pp. 212–223.
- [69] W. Miao, L. Li, and Z. Wang, "A survey on deep reinforcement learning for traffic signal control," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 1092–1097, doi: [10.1109/CCDC52312.2021.9601529](https://doi.org/10.1109/CCDC52312.2021.9601529).
- [70] G. Zheng, X. Zang, N. Xu, H. Wei, Z. Yu, V. Gayah, K. Xu, and Z. Li, "Diagnosing reinforcement learning for traffic signal control," 2019, *arXiv:1905.04716*.



HYEMIN KIM received the B.S. degree in geoinformatics from the University of Seoul, South Korea, in 2023, where she is currently pursuing the M.S. degree in geoinformatics. Her research interests include spatial databases and spatial analysis, traffic simulation, and reinforcement learning applications in geospatial analysis.



JINHYUK PARK received the B.S. degree in geoinformatics from the University of Seoul, South Korea, in 2024, where he is currently pursuing the M.S. degree in geoinformatics. His research interests include spatial databases and geographic information systems, pedestrian simulation, and reinforcement learning.



DONGBEOM KIM received the B.S. degree in geography from Kongju National University, Gongju, South Korea, in 2021. He is currently pursuing the M.S. degree in geoinformatics with the University of Seoul, South Korea. From May 2021 to December 2022, he was a Developer with Naega System Company, specializing in spatial databases, web programming, and simulations. His research interests include geospatial database management and simulation modeling.



CHULMIN JUN received the Ph.D. degree in urban and regional planning from Texas A&M University, in 1997. He is a Professor of the Department of Geoinformatics, the University of Seoul, South Korea, where he also leads the GeoDB Laboratory. He teaches courses, such as object-oriented programming, databases, spatial databases, and spatial analysis. His research interests include spatial big data analysis, traffic simulation, and pedestrian simulation. In his recent research, he has applied AI techniques to public transport optimization and traffic signal optimization and developed a fire evacuation simulator incorporating sensors, such as people-counting sensors, fire sensors, and CCTV.

...