mathematics

Article

Dual-Objective Reinforcement Learning-Based Adaptive Traffic Signal Control for Decarbonization and Efficiency Optimization

Gongquan Zhang [1,2], Fangrong Chang [3,*], Helai Huang [1] and Zilong Zhou [3]

[1] School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[2] Harvard Medical School, Harvard University, Boston, MA 02138, USA
[3] School of Resources and Safety Engineering, Central South University, Changsha 410083, China
* Correspondence: 222023@csu.edu.cn

Abstract: To improve traffic efficiency, adaptive traffic signal control (ATSC) systems have been widely developed. However, few studies have proactively optimized the air environmental issues in the development of ATSC. To fill this research gap, this study proposes an optimized ATSC algorithm to take into consideration both traffic efficiency and decarbonization. The proposed algorithm is developed based on the deep reinforcement learning (DRL) framework with dual goals (DRL-DG) for traffic control system optimization. A novel network structure combining Convolutional Neural Networks and Long Short-Term Memory Networks is designed to map the intersection traffic state to a Q-value, accelerating the learning process. The reward mechanism involves a multi-objective optimization function, employing the entropy weight method to balance the weights among dual goals. Based on a representative intersection in Changsha, Hunan Province, China, a simulated intersection scenario is constructed to train and test the proposed algorithm. The result shows that the ATSC system optimized by the proposed DRL-DG results in a reduction of more than 71% in vehicle waiting time and 46% in carbon emissions compared to traditional traffic signal control systems. It converges faster and achieves a balanced dual-objective optimization compared to the prevailing DRL-based ATSC.

Keywords: adaptive signal control system; intersections; carbon emissions; deep reinforcement learning

MSC: 93C40

Citation: Zhang, G.; Chang, F.; Huang, H.; Zhou, Z. Dual-Objective Reinforcement Learning-Based Adaptive Traffic Signal Control for Decarbonization and Efficiency Optimization. *Mathematics* **2024**, *12*, 2056. https://doi.org/10.3390/math12132056

Academic Editors: Haifeng Song, Min Zhou and Xiaoqing Zeng

Received: 21 May 2024
Revised: 19 June 2024
Accepted: 27 June 2024
Published: 30 June 2024

1. Introduction

The dramatic increase in vehicles on the road has caused serious traffic congestion and environmental pollution issues in urban areas, especially at intersections where vehicle acceleration and deceleration frequently occur [1]. To improve road traffic efficiency, a variety of traffic signal control (TSC) systems have been developed as coordinators for the traffic flows at urban intersections [2]. According to the traffic management policy applied by the authorities, TSC systems are divided into fixed-time signal control (FTSC), actuated/triggered signal control (ASC), and adaptive traffic signal control (ATSC) systems.

The fixed-time control system operates according to a pre-defined signal timing strategy, with a fixed periodic duration, and pre-timed red and green signal phases, regardless of the traffic state [3,4]. Despite its practical importance, the fixed-time control strategy developed based on historical traffic data cannot accommodate variable and unpredictable traffic demands in the real world [5–7]. To this end, an actuated control system was developed. In such a system, the traffic light or its duration time varies with the detected traffic flow at the specific entrance of an intersection according to pre-defined rules [8,9]. Although the actuated control system takes into consideration traffic fluctuations, the traffic flow alone is insufficient in reflecting the actual traffic demands in complex traffic conditions [10].

To relax the limitations of the actuated control system, adaptive traffic control systems have been proposed. In such a system, the real-time traffic state is monitored continuously through several critical parameters, based on which the adaptive control strategies are updated accordingly [11]. The most deployed ATSC systems at urban intersections include the Sydney Coordinated Adaptive Traffic System (SCATS) and the Split Cycle Offset Optimization Technique (SCOOT). The SCATS aims to select the optimal phasing (i.e., cycle times, phase splits, and offsets) scheme for a traffic situation from pre-stored plans according to the vehicle equivalent flow and saturation level calculated from the fluctuations in traffic flow and system capacity. The SCOOT reacts to different traffic states by changing the cycle length, phase splits, and offset in small increments according to vehicle delays and stops calculated from the detected flow [12]. SCATS and SCOOT have proven their great potential in improving traffic efficiency while being human-crafted, given their respective control schemes and incremental designs pre-determined by experts [13]. The experts' knowledge is valuable but may suffer from subjective bias issues.

In recent years, reinforcement learning (RL), particularly data-driven deep reinforcement learning (DRL), has shown excellent application prospects in ATSC [14]. In the ATSC system, RL self-learns the optimal actions through interaction and feedback with the traffic environment instead of manually setting pre-defined rules. One or several intersections are considered an agent. The signal control system of the agent makes a decision after observing the state of the road network, and then learns the optimal signal timing scheme by maximizing the reward of environmental feedback [15]. Mikami and Kakazu [16] first applied RL to TSC optimization, leading to an upsurge in application of RL in TSC systems. However, RL is suitable for models with a discrete state and its direct application to TSC systems increases the computational complexity and requires large storage space [17]. Deep learning (DL) inspired by the working mode of the human brain can effectively process high-dimensional data by transforming low-level features to abstract high-level features, and thus can address the application limitation of RL in traffic signal control systems [18]. By combining the perception capacity of DL with the decision-making capacity of RL, DRL has been widely applied to ATSC [19].

The application of the DRL algorithm to ATSC in most studies focuses on the calculation rate, convergence effect, and application scenarios, in which traffic efficiency is the major goal of TSC optimization [19–22]. Considering the severe air pollution caused by idling times, parking times, and frequent accelerations/decelerations at intersections [23], vehicle emissions are also taken into consideration in the development of FTSC [24,25] and ASC [26], in addition to traffic efficiency. However, these bi-objective traffic control systems are pre-defined optimal timing schemes based on historical traffic data, which cannot be applied to the real-time control of real-world dynamic traffic flow for efficiency and emission optimization. To fill this research gap, this study proposes an optimized algorithm for the development of the ATSC system to take into consideration both vehicle emissions, especially carbon emissions, and traffic efficiency.

The proposed ATSC algorithm utilizes a DRL framework with traffic efficiency and carbon emissions as the dual-objective optimization. The agent in the DRL framework is developed to change the traffic signal phase based on the multiple-reward function related to optimization objectives. More specifically, traffic efficiency and carbon emissions are optimized by reducing the cumulative waiting time (CWT) and carbon dioxide emissions (CDEs) of all vehicles, respectively. The agent self-learns the optimal decision of traffic signal phases by minimizing the CWT and CDE between two adjacent traffic signal phases. To accelerate and balance the agent learning process, we develop a novel neural network comprising Convolutional Neural Networks and Long Short-Term Memory Networks and utilize the entropy weight method to balance the weights among the CWT and CDE. A representative intersection in the real world is simulated for training and testing the proposed algorithm.

2. Literature Review

2.1. TSC System for Decarbonization and Efficiency

Traffic signal control (TSC) systems are the primary means of organizing traffic at intersections, and a reasonable allocation of signal phase durations can improve vehicle passage efficiency. Early studies focused on calculating signal phase durations or setting rules for FTSC and ASC, which are unsuitable for dynamically changing traffic flows [4,8,9]. Hence, ATSC systems were proposed, which dynamically adjust signal timing based on real-time traffic data collected from various sensors [27]. The core principle of ATSC involves real-time analysis of traffic patterns and the optimization of signal phases to improve overall traffic throughput [28,29]. The mainstream approach is dynamic programming models [30,31] that predict traffic patterns based on historical and real-time data to optimize signal timing. Zhao and Ma [32] established an ATSC dynamic planning model to increase traffic volume at intersections under an alternative design. Dynamic programming considers the traffic density, vehicle arrival rates, and intersection geometry in signal timing allocation [33,34]. Although these models can provide optimal solutions for traffic signal planning, they are computationally intensive and resource-demanding, especially for large-scale networks. Therefore, reinforcement learning (RL), particularly deep reinforcement learning (DRL), offers an innovative solution for training agents to manage traffic signals [20]. Agents learn optimal strategies through trial and error to minimize delays and enhance traffic flow stability. Although RL techniques can handle complex nonlinear traffic patterns, they often require long training periods and significant computational power.

Previous studies indicated that TSC systems can alter driving behavior, effectively reducing vehicle carbon emissions and fuel consumption [35–37]. Eco-driving strategies integrated with TSC systems encourage drivers to adopt energy-saving driving habits, such as smooth acceleration and deceleration, maintaining optimal speeds and minimizing idling time [38,39]. By advising drivers to maintain stable speeds and avoid rapid acceleration or braking, TSC systems developed based on eco-driving strategies reduce fuel consumption and emissions. Hao et al. [26] developed a vehicle eco-approach and departure framework controlled by ASC to achieve carbon emissions reduction. Dynamic programming models [40,41] and RL techniques [42,43] are also applied to optimize signal timing specifically for decarbonization. These models can prioritize green waves in high-traffic corridors and adjust signal timings to minimize idling at intersections, both of which reduce fuel consumption. Some studies use multi-agent deep reinforcement learning techniques to coordinate multiple traffic signals, ensuring smooth traffic flow and reducing stop-and-go traffic [44,45]. Additionally, TSC systems are designed to prioritize eco-friendly vehicles, such as electric and hybrid cars, by providing them with longer green phases or giving them priority at intersections to lower overall emissions [46,47].

To achieve the synergistic optimization of decarbonization and efficiency, balancing the demand for efficient traffic flow with the goals of reducing emissions and fuel consumption is necessary. Multi-objective optimization frameworks are employed to tackle this challenge as they can handle conflicting objectives and provide solutions that balance efficiency and decarbonization [48–50]. These multi-objective frameworks often use evolutionary algorithms or other advanced optimization techniques to find Pareto optimal solutions. Lin et al. [51] tackles multi-objective urban traffic light scheduling, minimizing delays and carbon emissions using Q-learning-enhanced algorithms. Furthermore, adaptive TSC systems integrating eco-driving strategies and dynamic programming models are particularly effective in achieving synergistic optimization [52,53]. Integrating eco-driving/carbon emission and DRL allows the system to learn and adapt to real-time traffic conditions. Boukerche et al. [54] used deep reinforcement learning to optimize traffic signals and vehicle speeds, ensuring smooth passage through intersections and reducing delays and emissions. The DRL-based TSC method continuously improves performance by incorporating real-time data from various sources, including vehicle-to-infrastructure (V2I) communication, to enhance its optimization capabilities.

2.2. DRL-Based ATSC

The ATSC system can learn optimal policy through its continuous interactions with the real-time traffic state at intersections by applying different deep reinforcement learning approaches [53,55–57]. In the development of ATSC systems, three crucial parameters are usually used, including state which is the description of the traffic environment, action which is the set of traffic signal phases, and the reward function which serves to measure the changes in traffic efficiency or other relevant traffic indexes caused by the action [58–60]. Existing DRL-based ATSC studies adopted different designs of state, action, and reward. States can be vehicle-based representations, such as discrete traffic state encode (DTSE) including vehicle position and speed information, or feature-based value vector representations, such as vehicle queue length, cumulative delay, and waiting time [61,62]. Action includes selecting a possible green signal phase, keeping the current green signal phase/switching to the next green signal phase in sequence, or changing the signal phase duration [63]. Reward mainly concerns vehicle queue length, delay, etc. [64].

DRL-based ATSC algorithms are always trained and implemented by a single agent of value-based or policy-gradient-based methods [65]. Popular DRL algorithms used for traffic-efficiency-based ATSC systems include the Deep Q Network (DQN) based on the value function, Deep Deterministic Policy Gradient (DDPG), Advantage Actor-Critic (A2C), and Asynchronous Advantage Actor-Critic (A3C) based on the policy gradient [66]. For example, Arel et al. [18] applied DRL in ATSC using a neural network to fit the Q-value, in which the Q-value may be overestimated by the DQN. To solve the overestimation problem, Van Hasselt et al. [67] proposed the Double Deep Q Network to decouple action selection and value function estimation. Given that uniform sampling reduces the learning effect, a priority experience replay method which gives priority to important operations and a Dueling Deep Q Network that attaches an extra Q-value to the selection of each action was proposed to accelerate the learning effect [68,69]. To enhance the stability of the model, a Double Dueling Deep Q Network was proposed by using the Convolutional Neural Network (CNN) as the Q-value function approximator and target network, respectively [49,70]. In terms of policy-gradient DRL in ATSC algorithms, the A2C method and DDPG, which uses the nonlinear function to approximate the Q-value function, were applied to improve the efficiency and stability of the model [71–73]. In addition, the DRL model for ATSC is optimized by considering vehicle heterogeneity, improving the model input, considering the key points of the road network, and so on [74].

In summary, traditional TSC systems can only achieve emission reductions based on specific scenarios and optimization algorithms, limiting their adaptability and practicality. ATSC systems based on eco-driving, dynamic programming, and DRL show good performance in decarbonization, but most still primarily focus on efficiency optimization, treating decarbonization as a constraint or secondary strategy rather than proactively linking it to signal timing. To fill this gap, this study proposes an optimized DRL algorithm that targets carbon emissions and efficiency as primary optimization objectives. A multi-objective reward function is designed to accelerate model convergence, advancing the development of an ATSC system that optimizes decarbonization and efficiency.

3. Deep Reinforcement Learning-Based ATSC Algorithm

This section presents the research problem statement of this paper and describes the DRL-based ATSC algorithm for improving traffic efficiency and reducing carbon emissions.

3.1. Research Problem Statement

In this study, the research design is to reduce carbon emissions and improve traffic efficiency by designing a DRL-based ATSC system for optimizing dual goals (DRL-DG). In this application of DRL to the TSC problem, the signal control unit is regarded as an agent that takes actions by observing the state in the intersection environment. Its action a is defined as the appropriate traffic signal phase. Its reward r is defined as the multi-objective combination of the traffic efficiency index and carbon emission index. Through a self-learning process, the agent makes optimal decisions of traffic signal phases to achieve the goals.

In the environment, a signalized intersection is designed to connect with four access roads. Each road is a dual carriageway allowing vehicles to travel in both directions. For each direction, three types of lanes are designed. Along the direction of vehicles approaching the intersection, the inside lane is used for left-turning vehicles; the middle lane is used for straight-ahead vehicles; and the outside lane is used for right-turning or straight-ahead vehicles. The specific width, quantity, and rules of lanes at the intersection are designed based on travel demand.

The traffic signal phase defines the releasing and waiting time for traffic flow in different directions, consisting of red, yellow, and green signal policies to ensure the orderly movement of vehicles at intersections. In our problem, four traffic signal phases are designed for the movement of vehicles, as shown in Figure 1. Phase 1 sets the green signal for the middle and outside lanes; phase 2 sets the green signal for the inside lane in the east–west direction; Phase 3 sets the green signal for the middle and outside lanes; and Phase 4 sets the green signal for the inside lane in the north–south direction. According to the real-world rules of TSC, the yellow signal timing is set as four seconds.

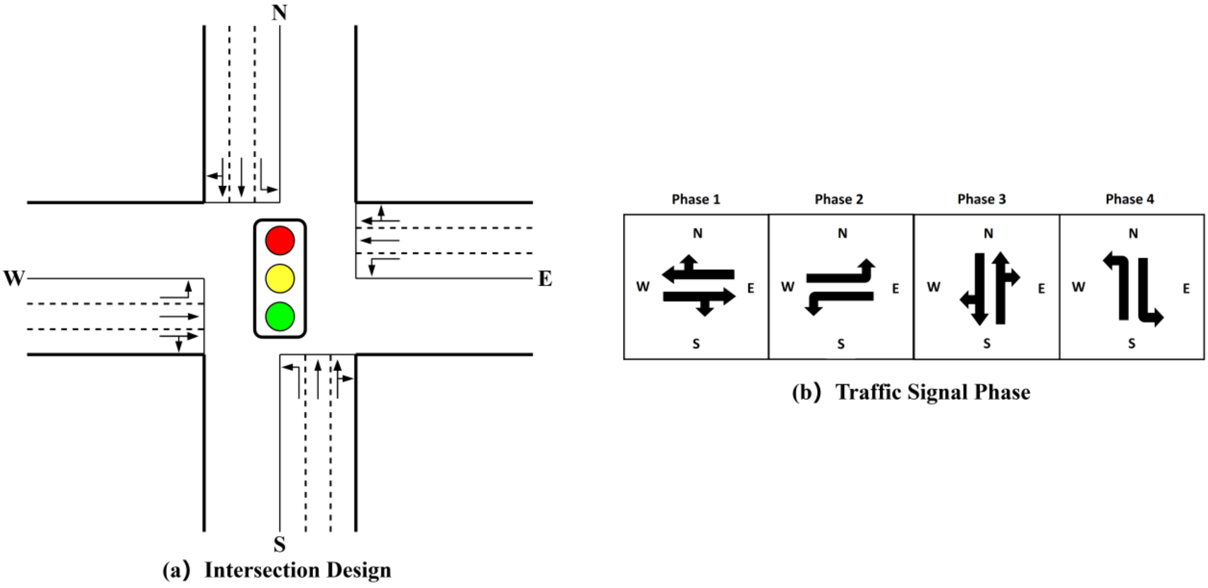


Figure 1. Definition of the intersection and four traffic signal phases.

3.2. Deep Reinforcement Learning

By the RL approach, agents learn the optimal policy to achieve definitive goals through continuous interactions with the environment. The Markov decision process is a theoretical framework to achieve goals through interactive learning, which can explain well the basic process of RL [14]. As TSC is the process of discrete action selection, value-based RL is appropriate for the current application. Specifically, the state is expressed as the characteristic matrix or vector of the traffic environment. The action is shown as the discrete selection vector while the reward is presented as a scalar value related to the traffic data. RL learns strategies/policies to maximize returns or achieve specific goals via continuous interaction with the environment. The real-time state s of the environment is first input to the agent for taking the corresponding action a according to its current knowledge of policy π. Then, the agent obtains feedback reward R (or punishment) from the environment, and accumulates long-term goals based on the reward. Under the action a_t, the state s_t transits to the state s'_{t+1} with a probability of p_a. In the learning process, policy is constantly updated to maximize the expected value of the long-term reward (action-value function) until the expected value stabilizes in the optimal policy π_* (term: "Converge"). The action-value function is defined as

$$Q_\pi(s,a) \doteq E_\pi[G_t | s_t = s, a_t = a] = E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | s_t = s, a_t = a] \tag{1}$$

where $Q_\pi(s,a)$ denotes the expected return of adopted policy π after taking action a at state s, E_π is the expected value of the adopted policy π, G_t is the cumulative discounted future reward, s_t is the state at the time step t, a_t is the action taken at the time step t, k is an incremental value from 0 to positive infinity, γ is the discount factor, and R_{t+k+1} is the reward at the time step $t+k+1$.

According to the Bellman equation, the action-value function decomposes as [75]:

$$\begin{aligned} Q_*(s,a) &= E\left[R_{t+1} + \gamma \max_{a'} Q_*(s_{t+1}, a') \,\middle|\, s_t = s, a_t = a\right] \\ &= \sum_{s',r} p(s', r | s, a)\left[r + \gamma \max_{a'} Q_*(s', a')\right] \end{aligned} \tag{2}$$

where $Q_*(s,a)$ denotes the optimal expected return of the optimal adopted policy π after action a is taken at state s, E is the expected value, R_{t+1} is the reward at the time step $t+1$, s_{t+1} is the state at the time step $t+1$, a' is the action taken at the time step $t+1$, and $Q_*(s',a')$ is the optimal expected return of the optimal adopted policy π after action a' is taken at state s_{t+1}. In addition, s' is the state at the time step $t+1$, r is the reward after action a is taken at state s, and p is the probability of the state transition. The optimal policy by iterating the optimal action value function continuously is solved:

$$\pi_* = arg \max_{a \in A} Q_*(s,a) \tag{3}$$

where π_* is the optimal policy and A is the set of actions.

DRL is the combination of RL and DL, which is one of the advanced learning frameworks in the current control system. Deepmind [19] proposed the Deep Q Network (DQN) in 2013. The DQN uses the experience playback to renew the neural network of the Q-value calculation instead of the tabular form and stores the samples (s,a,r,s') from the interaction in the memory of experience. Then, small batches of samples are uniformly sampled from the memory of experience. The depth neural network is trained by the random gradient descent method to approximate the Q-value. A strong correlation in samples can be interrupted by random sampling, which stabilizes the convergence.

$$\pi_* = arg \max_{a \in A} Q_*\left(s, a; w^\theta\right) \tag{4}$$

where w^θ is the parameter of the neural network.

$$y(s,a) = r + \gamma \max_{a'} Q_*\left(s', a'; w^\theta\right) \tag{5}$$

However, the DQN tends to overestimate Q-values. Therefore, this study employs the Double DQN (DDQN) framework to design the agent, whose current target action-value function is defined as

$$y(s,a) = r + \gamma Q(s', arg \max_{a'} Q\left(s', a'; w^\theta\right); w^t) \tag{6}$$

where w_t represent the parameters of the target network.

3.3. Framework

Based on the DRL's architecture, the conceptual framework of the DRL-DG approach consists of the environment and the agent is composed of a self-learning algorithm and a TSC component as shown in Figure 2. The agent applies the DDQN algorithm and receives a reward related to optimization goals after executing actions affecting the environment. The TSC system takes actions to adjust traffic signal phases to smooth traffic flow.

Traffic environment information is collected and transformed to the state s_t at the t time step as the input of the agent in DRL-DG. Based on s_t, an action a_t is selected for the agent through an ε-greedy policy. According to the action a_t, the TSC system remains in the current traffic signal phase or switches to another traffic signal phase to change vehicular movements on specific lanes. After taking action a_t, the traffic environment changes to the state s_{t+1} at the next time step $t+1$. The reward r_t of the state–action pair (s_t, a_t) is calculated according to the definition of reward functions. Next, the reward r_t and the state s_{t+1} are returned from the environment, forming (s_t, a_t, r_t, s_{t+1}) together with state–action pair (s_t, a_t), stored as the agent's experience in the memory pool. The state s_{t+1} is used as the agent's input at the next time step $t+1$. All procedures involving the input and feedback mechanism between the agent and environment are iterative. Finally, the agent learns the optimal traffic signal phases and updates the DDQN model from the memory pool by the experience replay method.

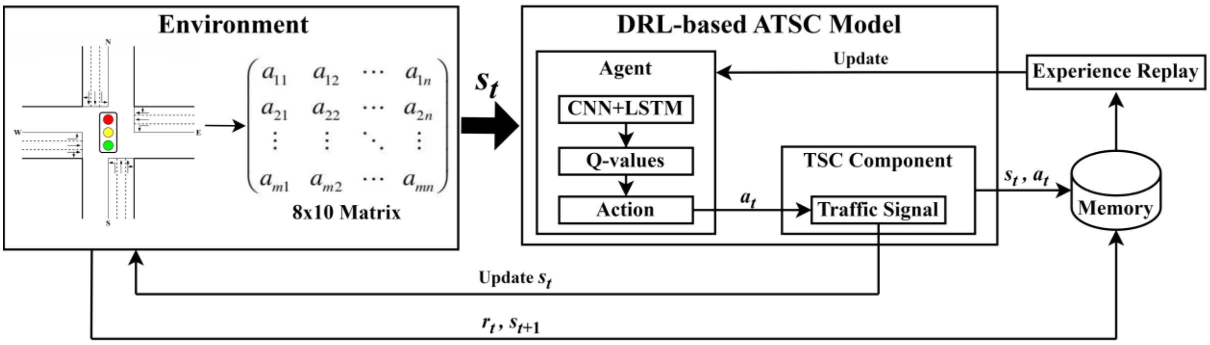


Figure 2. The conceptual framework of DRL-DG.

3.4. Agent Design

3.4.1. State

In this study, based on the discrete traffic state encoding (DTSE), non-uniform quantization and one-hot encoding are used to design the state vector as the state representation. The intersection used for simulation is an isolated cruciform intersection with an eight-lane dual carriageway whose length is 500 m in four directions, respectively [76]. For each direction, the inside lane is designed for left-turning vehicles; two middle lanes are designed for straight-ahead vehicles; and the outside lane is designed for right-turning or straight-ahead vehicles. Lanes are divided into cells according to a certain length proportion. Taking the west approach entrance of the intersection as an example, the cell design is illustrated in Figure 3. The three lanes on the right are divided as a whole, while the left turn lane on the left is divided separately. Ten cells are obtained for the west entrance direction. A total of 80 cells are set for the lanes in four directions of an intersection. Whether a car is present in cells represents the state. The state value of the cell is one if a vehicle exists; otherwise, it is zero.

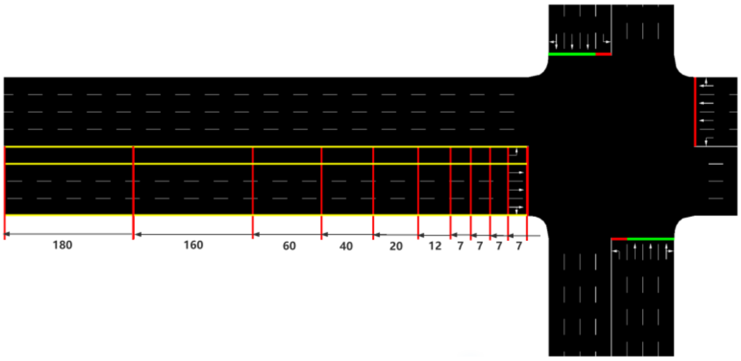


Figure 3. Schematic diagram of cells designed for west entrance at intersection (state presentation).

As for the design of each cell, it aims to reflect the distribution of vehicles along the road. As shown in Figure 3, the cell nearest to the intersection is 7 m long, which can accommodate only one vehicle. Considering the relatively low traffic density in the roadway sections far from the intersection, the cell farthest from the intersection is 180 m long. Compared with the method of using a real-time image or lane uniform division to represent the state, the proposed division method can reflect the actual nonuniform traffic density along the road, reduce the data dimension, and shorten the calculation time [77]. Using the presence of vehicles in each cell as the state can simplify traffic information, give samples specific labels of the environmental features, facilitate the feature extraction in the model, and thus increase the stability of convergence.

3.4.2. Action

The agent selects appropriate actions to divert traffic flow based on the traffic state. The action in this study is defined as the selection of a possible traffic signal phase. The action set is $A = \{EWG, EWLG, NSG, NSLG\}$ representing the east–west straight movement and right turn, the east–west left turn, the north–south straight movement and right turn, and the north–south left turn, respectively. The minimum duration of each green traffic signal phase is set to 10 s [63]. Meanwhile, a 4 s yellow signal is set during the switching between green and red signals for intersection safety. At each signal phase decision, if the agent selects the same phase, the green light for that phase is extended by 10 s. Otherwise, a 4 s yellow light is executed before switching to the next phase. In the DRL-DG system, after a phase has been selected consecutively six times, it will trigger the enforcement of other phases. Each green phase duration ranges from 10 to 60 s.

3.4.3. Reward

At a certain moment, the agent selects an action according to the observed state. Once the action is executed, the feedback, i.e., reward, is obtained for evaluating the performance of the action. The reward function is a key factor in ensuring the convergence of DRL and the achievement of optimization goals. The dual-objective reward function is defined by the reward functions of traffic efficiency R_{TE} and carbon emissions R_{CE}.

$$R^{(t)} = W_{TE}R_{TE}^{(t)} + W_{CE}R_{CE}^{(t)} \quad (7)$$

where W_{TE} and W_{CE} are the weights of traffic efficiency and carbon emissions set in the dual-objective reward function, respectively.

The weight values in the reward function influence the model's convergence. Compared to the expert scoring method, analytic hierarchy process, or simple linear weighting, the entropy weight method calculates weights based on data distribution, reducing subjective biases and providing a more data-driven and adaptable solution. The entropy weight method is used to adjust the weights based on the reward values in the DRL-based ATSC system [49,78]. Given that the entropy method is sensitive to data distribution and has initial subjective weighting issues, data normalization and the dynamic adjustment of weights based on real-time traffic data and reward values are implemented, ensuring stable and reliable weighting results.

$$y_i = \frac{x_i - min\,\{x_i\}}{max\{x_i\} - \min\{x_i\}} \tag{8}$$

where $min\,\{x_i\}$ and $max\{x_i\}$ represent the maximum and minimum value of the i reward.

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}, 0 \leq P_{ij} \leq 1 \tag{9}$$

where x_{ij} is the reward value at action i calculated by the reward function j.

$$Y = \{P_{ij}\}_{m\times n} \tag{10}$$

where Y is the standardized matrix.

$$H_j = -\frac{1}{\ln m}\sum_{i=1}^{m}(P_{ij} \times \ln P_{ij}) \tag{11}$$

where H_j is the entropy value of the reward function j.

$$g_j = 1 - H_j \tag{12}$$

where g_j is the coefficient of variation in the reward function j.

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j} \tag{13}$$

where w_j is the weight value of the reward function j, i.e., the value of W_{TE} and W_{CE}.

In terms of traffic efficiency, minimizing travel delays is the primary goal. Previous studies have proved that the waiting time of vehicles at the intersection can be used as an indicator of travel delay [58,61,64].

CWT denotes the cumulative or total waiting time of all vehicles stopping and waiting at the lane before crossing the intersection. A longer waiting time indicates longer delays. The difference in CWT between two adjacent execution time steps refers to the reward function indicating traffic efficiency:

$$R_{TE}^{(t)} = -\left(CWT_{(t+1)} - CWT_{(t)}\right) \tag{14}$$

where $CWT_{(t)}$ and $CWT_{(t+1)}$ denote the cumulative waiting time at step t and $t+1$, respectively.

In terms of carbon emissions, its major source is carbon dioxide emissions. Thus, the second goal is minimizing carbon dioxide emissions (CDEs). The difference in CDE in two adjacent executing actions refers to the reward function indicating carbon emission reductions:

$$R_{CE}^{(t)} = -\left(PE_{(t+1)} - PE_{(t)}\right) \tag{15}$$

where $PE_{(t)}$ and $PE_{(t+1)}$ denote the cumulative carbon dioxide emissions of step t and $t+1$, respectively.

Carbon dioxide emissions are acquired by the pollutant emission model of SUMO [79], which defines the emission quantity (g/h) as a function of the vehicular current engine power using typical emission curves over power (CEPs). The total carbon dioxide emissions PE are defined as

$$PE = (P_{Roll} + P_{Air} + P_{Accel} + P_{Grad})/\eta_{gearbox} \tag{16}$$

where

$$P_{Roll} = (m_{vehicle} + m_{load}) \times g \times (Fr_0 + Fr_1 v + Fr_2 v^4) \times v \tag{17}$$

$$P_{Air} = (c_d \times A \times \frac{\rho}{2})v^3 \tag{18}$$

$$P_{Accel} = (m_{vehicle} + m_{rot} + m_{load})av \tag{19}$$

$$P_{Grad} = (m_{vehicle} + m_{load}) \times Gradient \times 0.01 \times v \tag{20}$$

with the following definitions:

$\eta_{gearbox}$	driver train loss (set to 0.95);
$m_{vehicle}$, m_{load}	vehicular masses and load masses, respectively;
g	gravitational constant ($6.673 \times 10^{-11}\mathrm{m}^3/(\mathrm{kg} \times \mathrm{s}^2)$);
Fr_0, Fr_1, Fr_2	coefficients of resistance/friction;
v	the instantaneous velocity of vehicles;
c_d	the vehicular coefficient from drag;
A	cross-sectional area (m^2);
ρ	air mass density (~1.225 $\mathrm{kg/m}^3$);
m_{rot}	rotating mass.

3.4.4. DQN Model

The optimal policy is learned according to the optimization goal through the DQN model based on the traffic state. The traditional DQN model is a complete neural network with a full connection layer. Figure 4 presents the designed novel neural network for the DQN model in this study, which is a neural network linked by convolutional, long short-term, and fully connected layers. The target network utilizes the same neural network.

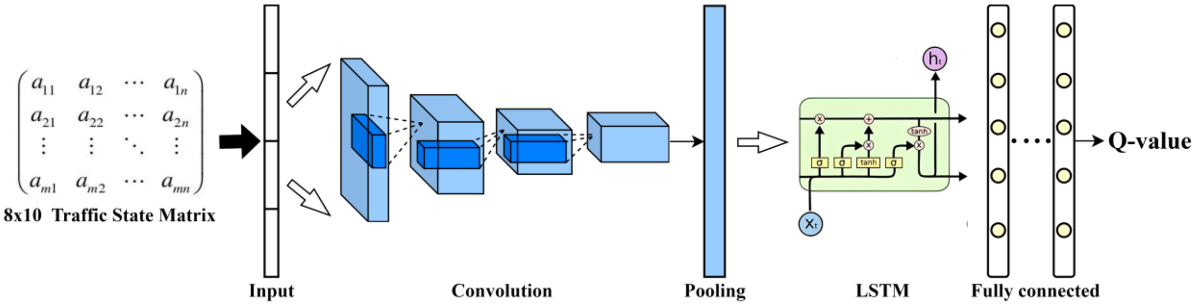

$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$
8x10 Traffic State Matrix
Input
Convolution
Pooling
h_t
tanh
σ
σ
tanh
σ
X_t
LSTM
Fully connected
Q-value

Figure 4. Structure design of the DQN model.

The given state represented by 80 cells, i.e., the 0–1 matrix with size 8×10, is calibrated to the Q-value of the action through convolutional and fully connected layers. Based on the size of the 0–1 state matrix, two convolution layers with 100 and 10 kernels are set up [80], whose filter size is 1×3 and stride is 2, to create labels with a lane characteristic. The final convolution layer's output is flattened via a pooling layer as the state vector to fully connected layers. The LSTM includes two layers with 80 units and a 0.2 dropout rate. It is noted that the number of fully connected layers is 5, whose width is 400. Using the Adam optimizer, the learning rate is 0.001, the batch size is 100, and the training iteration is 800 times per round, using the mean square error as the loss function. The Q-value indicates the reward value. Thus, the optimal selection is the action which has the highest Q-value. The agent's experience at every time step is stored in the memory pool. The DQN is trained by the experience replay method to update the weight parameter of the neural network.

3.5. Overall Algorithm

Algorithm 1 presents the pseudo-code of the proposed algorithm. The time step, agent, TSC component, and memory pool are initialized first (Line 1), after which the interface between the algorithm and simulation platform is established (Line 2). Based on the ε-greedy policy, the optimal action is obtained for the current traffic state (Lines 3–7). The optimal action is then activated for the TSC component, causing a traffic signal phase variation, as illustrated in lines 8–10. Following the action, the next state and reward returned from the environment are stored in the memory pool together with the current state and action (Lines 11–13). By sampling from the memory pool, DQN is trained (Lines 14–17). The algorithm is implemented through the DL framework Keras using the Python programming language (Version 3.8.12).

Algorithm 1: DRL-DG Algorithm Flow

```
1:  Initialize: Evaluation DQN, Target DQN, TSC component, memory pool
2:  Establish simulation interface (TRACI)
3:  for episode = 1 to total episode do
4:     Initialize: road network environment, import traffic flow data
5:     for time step = 1 to maximum time do
6:        Agent observes the current environment $s_t$
7:  Choose $a_t$ based on $\varepsilon$-greedy policy
8:  Import $a_t$ to TSC component
9:  TSC component changes traffic signal phase
10: Output $s_{t+1}$ and calculates reward $r_t$
11: Store $(s_t, a_t, r_t, s_{t+1})$ in memory pool
12:    end for
13:    Extract samples from the memory pool to train the network
14:    Based on $Q_*(s, a)$ to calculate optimization goals
15:    Update the parameters of Evaluation/Target DQN using the mean square error
    loss function
16: end for
```

Several variables act as obstacles in achieving convergence, including a high variability in traffic flow, complex traffic scenarios, the reward design, and algorithm parameters. The DRL-DG method improves convergence by employing a sophisticated reward function that uses the entropy weight method to balance traffic efficiency and carbon emissions. Additionally, a neural network structure combining CNN and LSTM is used to handle complex traffic scenarios and improve agent-learning efficiency. To enhance convergence stability, the experience replay mechanism and target networks are utilized. Algorithm parameters are carefully tuned through extensive experimentation.

4. Case Validation

Based on a representative signalized intersection in the Changsha urban road network, Simulation of Urban Mobility (SUMO) software is adopted to build the simulated intersection scenario for training and testing the proposed algorithm. In the simulation, the algorithm collects traffic information and controls traffic signal phases by the Traci interface coded directly in Python. The agent in DRL-DG is trained under a random traffic flow generated by a Weibull distribution. The performance of the proposed DRL-DG is evaluated at the simulated intersection with the real-world traffic flow data recorded by photography and compared with that of three classic traffic signal control algorithms.

4.1. Scenario

The experimental scenario refers to the intersection of Kaiyuan East Road and Huangxing Avenue in Changsha City, Hunan Province, China. The intersection is a typical cruciform signalized intersection in China, which connects four 500 m long dual carriageways with four lanes each way in Figure 5. For each entrance direction, there is an inside lane for left-turning vehicles, two middle lanes for straight-ahead vehicles, and an outside lane for right-turning or straight-ahead vehicles. The vehicles on the outside lane are permitted to turn right during the red signal phase without conflicts in the intersection area.

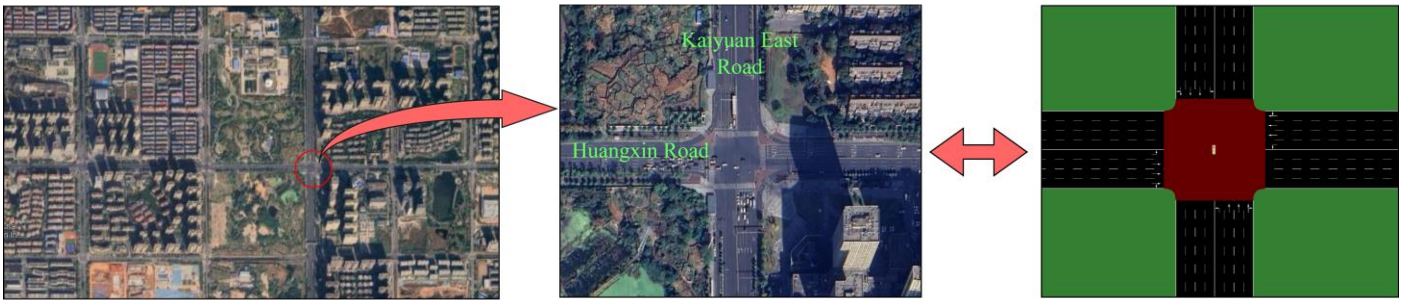


Figure 5. Real-world intersection and scenario.

The four directions of the real-world intersection are all business areas with a balanced traffic flow distribution. The real-world traffic flow data were collected from 7:30 a.m. to 8:30 a.m., which is part of the peak hours in Changsha City on Thursday 23 June, 2022. A total of 979 vehicles were observed during such a period. The number of vehicles increases significantly during the peak hours, especially in the middle lanes which account for about 70% of the total number of vehicles as presented in Figure 6, causing traffic congestion. In such a case, the traffic flow approximately obeys a Weibull distribution, which is thus used to simulate the flow distributions during peak hours.

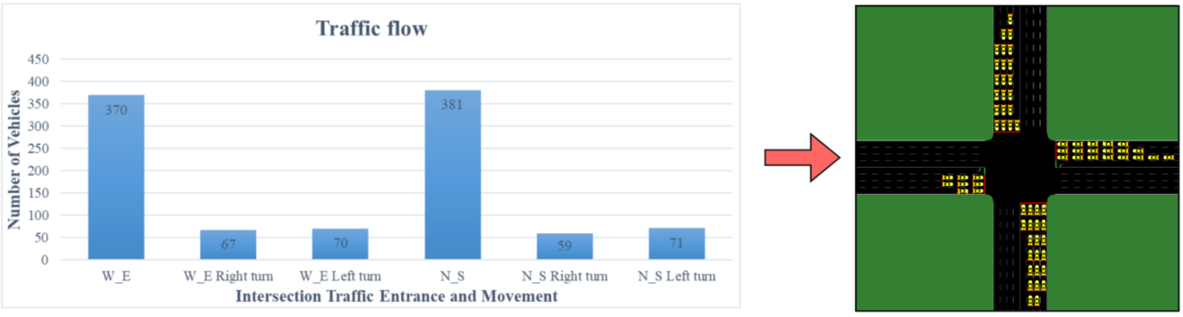


Figure 6. Real-world traffic flow.

4.2. Simulation and Algorithm Setting

To simulate the traffic flow during the peak hours at the real-world intersection, the probability density of a Weibull distribution is assumed for vehicles in the traffic flow.

$$f(x;\lambda,a)=\begin{cases}\frac{a}{\lambda}\left(\frac{x}{\lambda}\right)^{a-1}e^{-\left(\frac{x}{\lambda}\right)^{a}} & x\geq 0\\ 0 & x<0\end{cases} \tag{21}$$

where λ is the scale parameter, set as 1; a is the shape parameter, set as 2.

As for the vehicle movement from any approaching entrance of the intersection, the probability of vehicles turning left, traveling straight, or turning right is 12.5%, 75%, and 12.5%, respectively. In the simulation, the car-following behavior obeys the Krauss car-following model. The vehicle is $5\ \mathrm{m}$ long with the minimum distance between the adjacent vehicle of $2.5\ \mathrm{m}$. The maximum velocity of vehicles is $35\ \mathrm{km/h}$, setting $1\ \mathrm{m/s^2}$ as the maximum acceleration, $4.5\ \mathrm{m/s^2}$ as the maximum deceleration, and 0.5 ($sigma$) as the driver defect.

Table 1 shows the detailed setting for the simulation and algorithm. The parameters presented are for the agent (action number, duration of signal phases) and algorithm (episode, step, learning rate, batch size, memory, etc.).

Table 1. Algorithm setting.

Parameter	**Value**	**Note**
Number of actions N_a	4	Number of phases
Minimum green time g_{min}	10 s	
Yellow time t_y	4 s	
Default phase p_0	EWG	Initial traffic signal phase
Episode	400	Number of trainings
Step	3600	Step length of one training
Weight of CWT W_{CWT}	0.5	
Weight of CDE W_{CDE}	0.5	
Batch size B	100	
Learning rate L_r	0.001	
Epoch E	800	
Starting ε	0.99	
Ending ε	0.01	To avoid local optimal solution
Minimum memory pool size M_{min}	600	To obtain all samples
Maximum memory pool size M_{max}	60,000	To remove the oldest element
Discount factor γ	0.8	
Leaky ReLU β	0.01	
Length of training step t_{train}	1 s	

4.3. Comparisons among Different TSC Systems

To evaluate the performance of the proposed algorithm, the DRL-DG-based adaptive traffic control system is compared with traditional traffic control systems, including FTSC and ASC, and the DRL-SG-based ATSC system, which applies an advanced DRL algorithm for efficiency optimization. The comparisons are conducted in terms of VWT, VQL, CDE, ADF, VFC, and NGE [13].

(1) **Fixed-time signal control (FTSC)**: FTSC predefines a set of timing schemes by the classic Webster timing method and is widely used for real-world traffic intersections. The duration set for phase 1, phase 2, phase 3, and phase 4 is 60, 40, 60, and 40 s, respectively. Between two adjacent phases, a four-second yellow signal is set.

(2) **Actuated signal control (ASC)**: ASC adjusts the traffic signal phase and the duration time based on the queue length and traffic flow. Once the queue length of the lane during the red signal phase reaches the threshold which is set as 70 m, the signal for this lane turns green. In case many vehicles are still in the lane during the green signal, the duration of the green signal will be extended up to 60 s [81].

(3) **DRL-based ATSC Optimizing Single Goal (DRL-SG)**: DRL-SG applies the DRL algorithm framework into ATSC to optimize traffic efficiency that receives the most attention. Similarly, the DQN model used for efficiency optimization is a conventional, long short-term, and fully connected neural network. The reward refers to the difference in the vehicular waiting time at two adjacent time steps.

The signal phase duration and rules of DRL-DG are different to FTSC and ASC. FTSC utilizes the Webster method, a historical traffic data-driven approach that calculates predetermined and fixed green phase durations. The ASC method incorporates a queue length detection mechanism based on FTSC, allowing for early phase changes and dynamic signal phase adjustments based on real-time traffic flow. In contrast, the proposed DRL-DG approach entails developing a phase-selecting agent capable of dynamically adjusting green phase durations in response to real-time traffic conditions [11,61]. The minimum green phase duration is set to 10 s to account for driver reaction times and ensure an optimal optimization effect [5,20].

4.4. Evaluation Metrics

The primary optimization objectives of DRL-DG include vehicle waiting time and vehicle driving delay. In addition, the vehicle queue length is calculated for traffic efficiency evaluation. While vehicle fuel consumption, carbon dioxide emissions, and toxic gas emissions are estimated for vehicle emissions. The explanations for all metrics are given below.

(1) **Vehicle waiting time (VWT)**: This refers to the cumulative waiting time of vehicles stopping at the intersection in each time stamp (5 min). A lower VWT indicates a shorter time that vehicles are stopped at the intersection, contributing to higher traffic efficiency.

(2) **Vehicle queue length (VQL)**: This refers to the cumulative quantity of vehicles stopping at the intersection entrance in each time stamp (5 min). A lower VQL implies more vehicles crossing the intersection, reducing the possibility of congestion.

(3) **Carbon Dioxide Emissions (CDEs)**: These refer to the total carbon dioxide emissions of vehicles in each time stamp (5 min). CDEs are used as the main index to evaluate vehicle carbon emissions. Lower CDEs denote lower carbon emissions.

(4) **Acceleration–deceleration frequency (ADF)**: This refers to the total frequency of vehicle accelerations or deceleration in each time stamp (5 min). A lower ADF reveals lower extra carbon dioxide emissions (Vasconcelos et al., 2014).

(5) **Vehicle fuel consumption (VFC)**: This refers to the cumulative fuel consumed during driving in each time stamp (5 min). A lower VFC denotes a higher energy efficiency.

(6) **Noxious gas emissions (NGEs)**: These are the total emissions of carbon monoxide (CO) and nitrogen oxides (NOx) emitted by vehicles in each time stamp (5 min). Lower NGEs denote less toxic air pollutants. The CO and NOx emissions are estimated using SUMO's pollutant emission model, which calculates emissions based on the vehicle's current engine power and typical emission curves [79].

$$E_{CO} = P \cdot EF_{CO} \tag{22}$$

$$E_{NOx} = P \cdot EF_{NOx} \tag{23}$$

$$P = \left(m \cdot a + m \cdot g \cdot C_r + \frac{1}{2} \cdot p \cdot A \cdot C_d \cdot v^2\right) \cdot v \tag{24}$$

where P is the engine power in kilowatts (kW). EF_{CO} and EF_{NOx} are the emission factors for CO and NOx (grams/kWh). v is the vehicle speed (m/s). m is the vehicle mass (kg). a is the vehicle acceleration in (m/s^2). g is the gravitational acceleration, typically 9.81 m/s^2. C_r is the rolling resistance coefficient. p is the air density (kg/m^3), typically 1.225 kg/m^3 at sea level and 15 °C. A is the vehicle frontal area (m^2). C_d is the air resistance coefficient.

4.5. Results and Discussion

FTSC, ASC, DRL-SG, and DRL-DG systems are implemented in the simulation based on the real-world traffic flow. The cumulative, average, and real-time evaluation metrics are obtained and compared for all traffic control systems.

Figure 7 illustrates that the DRL-DG's convergence speed is faster than that of the DRL-SG. The novel network architecture and the design of the multi-objective optimization function have accelerated the agent's learning process.

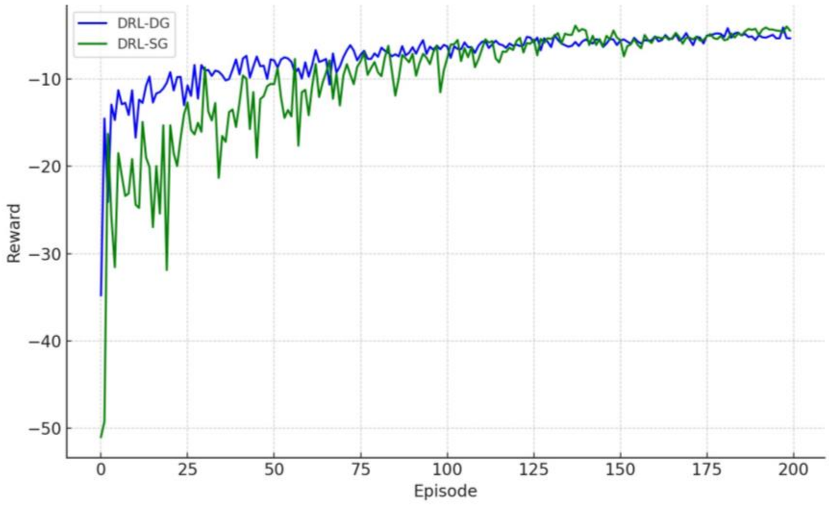


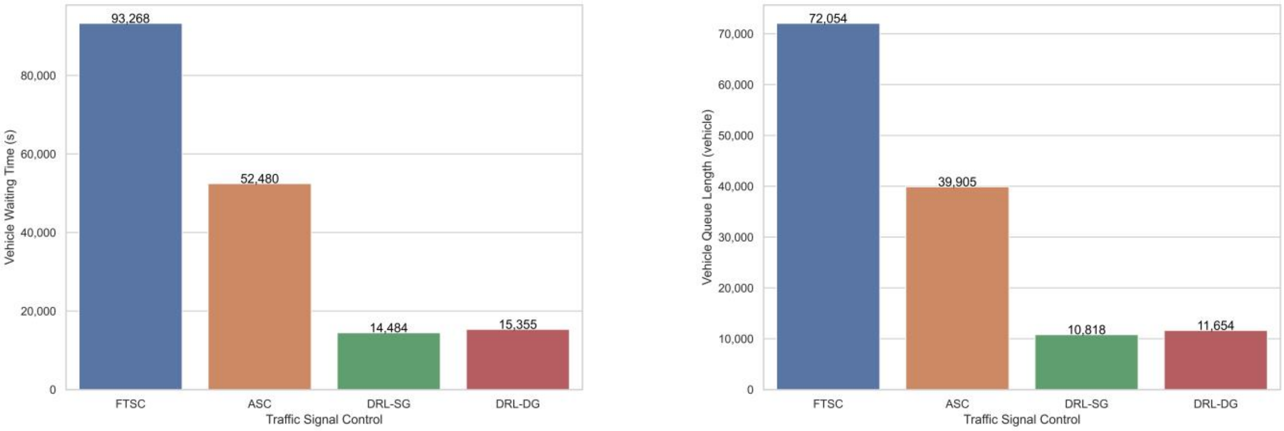
Figure 7. Training process.

4.5.1. Overall Analysis

VWT and VQL are indicators for the evaluation of traffic efficiency at the intersection. VWT indicates the time lost caused by vehicle stopping and waiting for the red signal. VQL is the number of vehicles stopping at the intersection due to a red signal. The cumulative waiting time and queue length for different traffic control systems are shown in Figure 8a,b, respectively. The average value of VWT and VQL is provided in the first two columns of Table 2. Compared to FTSC and ASC, VWT is reduced by 83.54% and 70.74%, respectively, for a DRL-DG-based ATSC. In terms of VQL, its value obtained for a DRL-DG-based ATSC is reduced by 83.83% and 70.79%, respectively. When compared with DRL-SG, DRL-DG has the approximate performance on VWT (15.68 vs. 14.79) and VQL (3.23 vs. 3.01).

In terms of CDE shown in Figure 9a, the vehicle carbon dioxide emissions of the DRL-DG-based ATSC system are reduced by 69.71%, 52.71%, and 41.96%, respectively, compared with FTSC, ASC, and DRL-SG systems. ADF is the acceleration or deceleration frequency/rate, an important component related to carbon emissions [82,83]. The result of cumulative ADF is provided in Figure 9b. From the figure, the ADF value of DRL-DG is smallest, indicating that the traffic flow is stable. The average value of CDE and ADF is provided in the third and fourth columns of Table 2, indicating similar comparison results to the accumulative results.

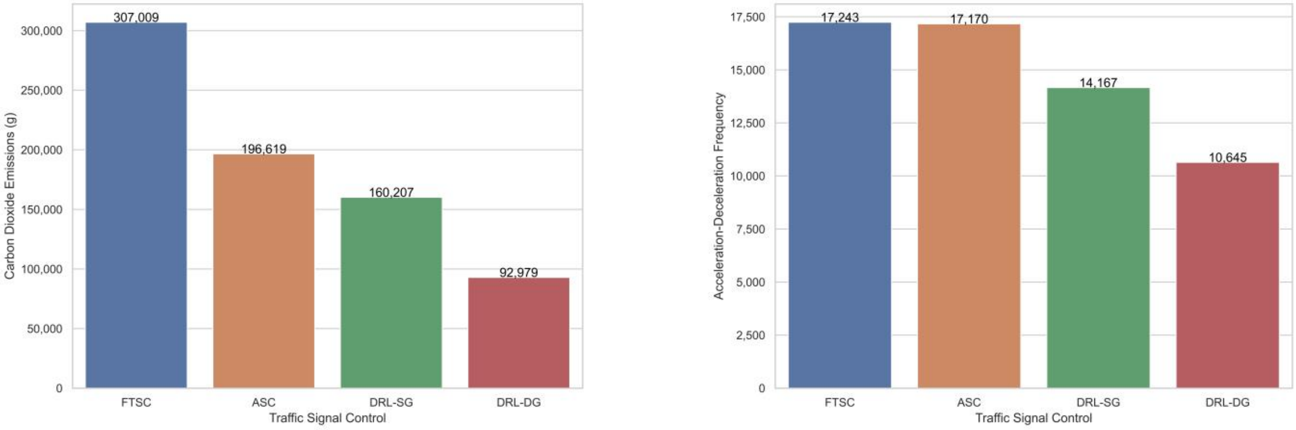
In addition to the primary objective indicators, VFC and NGE are also calculated and used for evaluating the performance of different TSC systems from the perspectives of economic benefits and toxic air pollution [84]. Their cumulative results are given in Figure 10a,b, respectively. Their average values are given in the last two columns of Table 2. Both the figure and table showed that the VFC of DRL-DG is reduced by 69.71%, 52.71%, and 41.96%, respectively, compared to FTSC, ASC, and DRL-SG. Similarly, the NGE of DRL-DG is reduced by 84.85%, 73.14%, and 24.53%, respectively. Therefore, the proposed DRL-DG algorithm can significantly improve fuel economy and reduce toxic air pollution.



(**a**) Vehicle waiting time

(**b**) Vehicle queue length

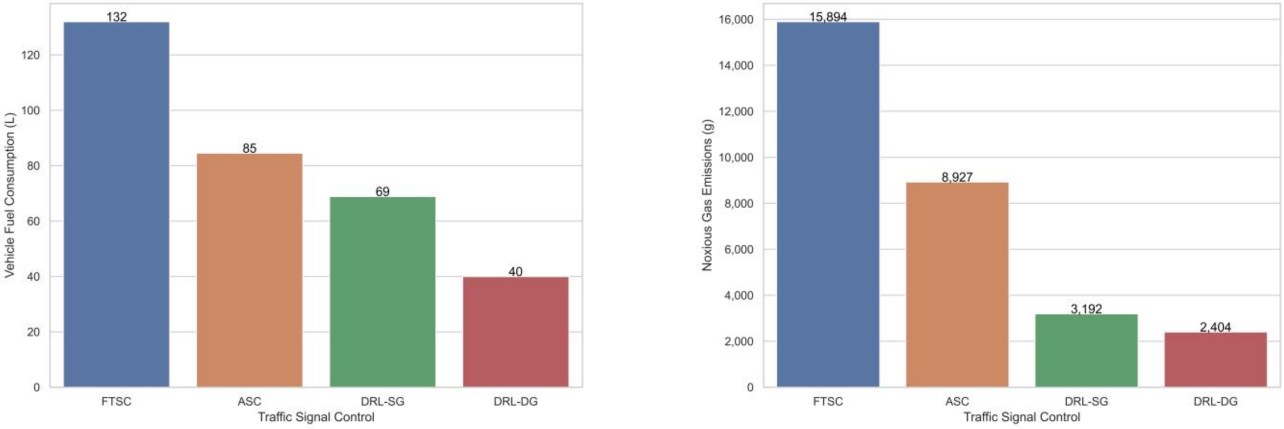
Figure 8. Cumulative performance of TSC systems regarding traffic efficiency.



(**a**) Carbon dioxide emissions

(**b**) Acceleration–deceleration frequency

Figure 9. Cumulative performance of TSC systems regarding carbon dioxide emissions.



(**a**) Vehicle fuel consumption

(**b**) Noxious gas emissions

Figure 10. Cumulative performance of TSC systems regarding secondary index.

Table 2. Average performance of traffic signal control methods with evaluation metrics.

TSC	Average VWT (s/Vehicle)	Average VQL (Vehicle/s)	Average CDE (g/Vehicle)	Average CDE (Rate)	Average VFC (mL/Vehicle)	Average NGE (g/Vehicle)

FTSC	95.27	20.02	313.59	17.61	134.81	16.24
ASC	53.61	11.08	200.84	17.54	86.33	9.16
DRL-SG	14.79	3.01	163.64	14.47	70.35	3.26
DRL-DG	15.68	3.23	94.97	10.87	40.83	2.46

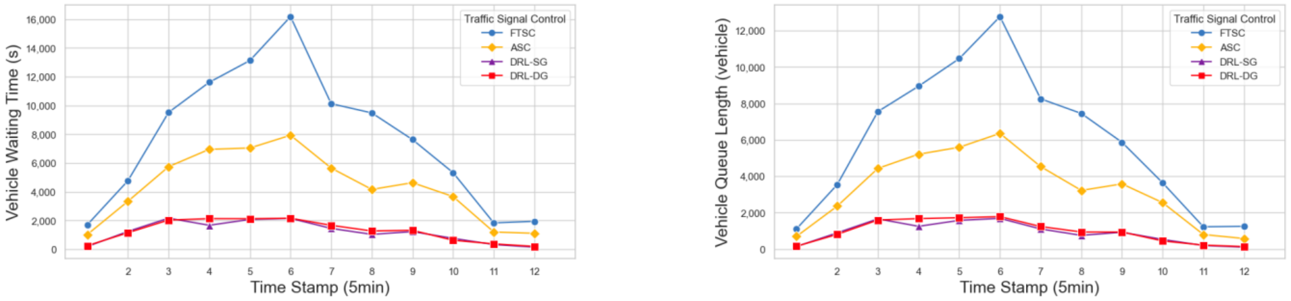
4.5.2. Comparative Analysis in Simulation

According to the real-world traffic flow at intersections, the traffic volume increases dramatically from 0 to 30 min and decreases gradually from 30 to 60 min. Based on the simulation, the real-time performance of DRL-DG on the evaluation metrics collected in five-minute intervals is compared with that of FTSC, ASC, and DRL-SG.

The real-time performance of different TSC systems on traffic efficiency in each five-minute interval is shown in Figure 11. The overall trend revealed by Figure 11a shows that the fluctuations in the optimized objective of VWT for all traffic control systems are positively related to those of traffic volume. A similar trend could be observed for VQL in Figure 11b. From 0 to 30 min, the vehicle waiting time and queue length increase, while from 30 to 60 min, the vehicle waiting time and queue length reduce. However, VWT and VQL are always lowest for DRL-SG and DRL-DG, followed by ASC and FTSC, indicating that DRL-based ATSC can significantly improve traffic efficiency compared to traditional traffic control systems. DRL-SG results in a slightly shorter waiting time and queue length than DRL-DG with an insignificant difference. This result can be explained by the sole optimization objective of DRL-SG which is traffic efficiency, while in DRL-DG, the efficiency may compromise the objective of carbon emissions, resulting in a lower efficiency.

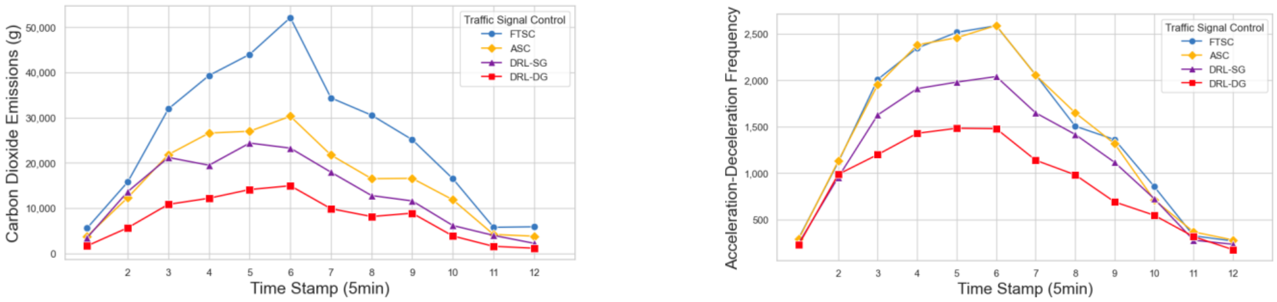
The real-time performance of various TSC systems on carbon emissions in each five-minute interval is presented in Figure 12a. From the figure**,** the CDE values of DRL-DG are always lowest, followed by ASC, FTSC, and DRL-SG, indicating its advantage of carbon emission reductions. The ADF of different traffic control systems is shown in Figure 12b. The FTSC and ASC result in the highest vehicle acceleration or deceleration, followed by DRL-SG. Similar to the comparative result of carbon dioxide emissions, DRL-DG results in the lowest acceleration/deceleration frequency [85,86]. The DRL-DG compels vehicles to avoid inessential acceleration/deceleration and move stably. Acceleration or deceleration rate has been demonstrated to contribute to significant carbon emissions [87]. If there is a speed profile, the explanations can be more convincing.

The real-time performance of traffic control systems on fuel economy and toxic air pollution is illustrated in Figure 13a and Figure 13b, respectively. Given that carbon emissions and fuel consumption are directly related, the images of CDE and VFC are found to be similar. The VFC value of DRL-DG is lowest among the four traffic control systems, indicating its advantage in the improvement in fuel economy. The reduction in fuel consumption can be explained by the decreased VWT and acceleration/deceleration rate [88]. As for NGE in Figure 13b, the toxic pollutant emissions at the intersection controlled by DRL-DG are lowest, indicating the lowest toxic pollutants among all four traffic systems. Based on Equations (22)–(24), the emission of noxious gas is related to the vehicle engine power [89]. The lowest NGE can be partially explained by the reduced vehicle waiting time (as shown in Figure 11a) and acceleration/deceleration rate (as shown in Figure 12b) [90].



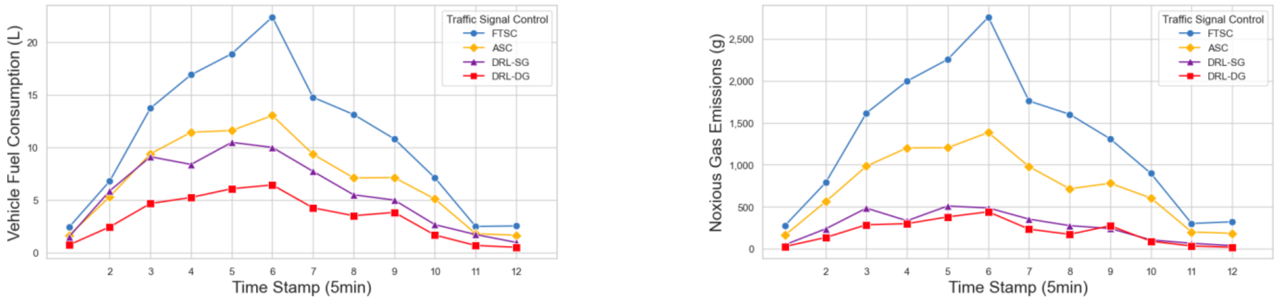
(**a**) Vehicle waiting time (**b**) Vehicle queue length

Figure 11. Real-time performance of TSC methods regarding traffic efficiency.



(**a**) Carbon dioxide emissions (**b**) Acceleration–deceleration frequency

Figure 12. Real-time performance of TSC methods regarding carbon dioxide emissions.



(**a**) Vehicle fuel consumption (**b**) Noxious gas emissions

Figure 13. Real-time performance of TSC methods regarding other air pollution index.

In summary, traffic control systems developed by DRL-based methods, especially DRL-DG, perform significantly better than traditional traffic control systems consisting of fixed and actuated traffic control systems, in terms of traffic efficiency, carbon dioxide emissions, fuel consumption, and toxic gas emissions. The overwhelming advantage of the DRL-DG-based traffic control systems is embodied in high-traffic volume situations. Considering this advantage and the expensive relevant equipment, a DRL-DG-based traffic control system could be applied for intersections with heavy traffic. In the practical application, the weights can be adjusted according to government policies or the demand of practice.

4.5.3. Opening the "Black Box" in DRL-DG

The relationship between the DRL-DG system and vehicle waiting time is fundamental to understanding its impact on traffic efficiency. The reasonable allocation of signal phase durations directly influences driving behavior and improves traffic flow. The DRL-DG system employs deep reinforcement learning to dynamically adjust traffic signal timings based on real-time traffic conditions. Traditional FTSC and ASC systems fail to adapt to fluctuating traffic patterns, resulting in extended vehicle waiting times and increased congestion. In contrast, the DRL-DG system continuously learns and modifies signal phases to optimize traffic flow. Microsimulation results have shown that the DRL-DG system reduces vehicle waiting times by 83.54% compared to FTSC and 70.74% compared to ASC. It intelligently selects signal phases and adjusts green phase durations based on current traffic data, minimizing idle times and enhancing overall traffic flow at intersections. These improvements highlight the system's ability to streamline traffic movement and reduce congestion through real-time optimization.

The impact of the DRL-DG system on carbon emissions is closely linked to its ability to reduce vehicle idle times and optimize acceleration and deceleration patterns. Vehicles emit higher levels of pollutants, such as CO_2, CO, and NOx, during idling and frequent stop-and-go movements. The DRL-DG system effectively lowers emissions by minimizing these periods through optimized signal timings. Specifically, the system adjusts signal phases dynamically, ensuring that vehicles spend less time idling at red lights and experience fewer abrupt stops and starts. This leads to a smoother traffic flow with reduced acceleration and deceleration cycles, significantly reducing carbon emissions. The macroscopic simulation results confirm this point (see Figure 12), showing that the DRL-DG system has the lowest vehicle acceleration/deceleration frequency. In fact, the DRL-DG system influences several critical factors directly impacting emissions: it reduces idle times, ensures smoother transitions through intersections, and adapts to real-time traffic conditions to prevent congestion. These adjustments result in a substantial reduction in overall fuel consumption and emissions. In simulations, the DRL-DG system achieved a 69.71% reduction in CO_2 emissions compared to FTSC and a 52.71% reduction compared to ASC. Further, CO and NOx emissions were significantly reduced, proving the environmental benefits of the DRL-DG system and its potential to improve sustainable urban mobility by lowering harmful emissions.

Several factors also influence vehicle carbon emissions at intersections, including intersection design, vehicle types, and traffic volume. The design of an intersection, such as the number of lanes, presence of dedicated turning lanes, and overall layout, can significantly affect traffic flow and emissions. Well-designed intersections that minimize vehicle idling and facilitate smooth traffic flow can reduce emissions. Additionally, the types of vehicles and their respective emission rates impact overall emissions. Eco-friendly cars produce fewer emissions than conventional vehicles. Traffic volume is another critical factor: high traffic volumes often lead to increased idling times and more frequent stop-and-go movements, contributing to higher emissions. Traffic signal optimization aims to prevent high traffic volumes, thereby reducing carbon emissions. Therefore, effective policy measures are essential to addressing these factors and reducing vehicle emissions at intersections. Implementing congestion pricing can decrease traffic volume during peak hours, reducing emissions. Incentives for eco-driving behaviors and driver education can promote energy-efficient driving practices. Investing in smart infrastructure, such as adaptive traffic signal control systems and real-time traffic monitoring, can enhance traffic flow and reduce emissions.

5. Conclusions

To improve traffic efficiency and reduce carbon emissions at intersections, this study proposes a deep reinforcement learning-based dual-objective optimization algorithm for the adaptive traffic signal control system. The objectives of this study are achieved by reducing vehicle waiting time and carbon dioxide emissions through the proposed DRL-DG-based ATSC traffic control systems. In addition, the performance of the proposed system in reducing vehicle fuel consumption and toxic gas emissions is also evaluated.

Based on the video data collected from an isolated intersection in Changsha City, China, the intersection and traffic flow are simulated through SUMO. Based on the simulated intersection, the proposed DRL-DG algorithm is trained and tested with an equal priority set for vehicle waiting time and carbon dioxide emissions. For comparison purposes, fixed-time signal control (FTSC), actuated signal control (ASC), and DRL-based ATSC that optimizes only traffic efficiency are also trained and tested. In terms of traffic efficiency, the results show that DRL-DG and -SG methods perform similarly on traffic efficiency without significance. But DRL-DG performs much better than FTSC and ASC with a reduction of more than 71% in vehicle waiting time. Regarding carbon dioxide emissions, the DRL-DG method performs best with a reduction of more than 46%. The traffic control system developed based on the proposed DRL-DG also shows its advantage in the reduction in vehicle fuel consumption and toxic gas emissions. For all evaluation metrics, the performance of the proposed algorithm is especially outstanding in high-traffic-flow situations.

The proposed DRL-DG-based traffic control systems are suitable for intersections with heavy traffic, considering their overwhelming advantage in high-traffic-flow situations and the limited funds available for system development. By revising the weights of the two objectives, the algorithms can adjust to government policies and practical demands on the trade-off of traffic efficiency and carbon emissions.

This study is not without limitations. In terms of objectives, road safety, especially traffic conflicts, which is another important issue of traffic in intersections, is not taken into consideration. In addition, the DRL-DG in the ATSC system faces challenges such as requiring extensive high-quality data, hyperparameter tuning, system complexity, lengthy training times, and ensuring robustness under diverse conditions. Future research will address these issues, aiming to develop more efficient, scalable, and practical DRL-DG-optimized ATSC systems for diverse urban environments.

Author Contributions: Conceptualization, G.Z. and F.C.; methodology, G.Z.; software, G.Z. and Z.Z.; validation, G.Z., F.C., and Z.Z.; formal analysis, G.Z. and H.H.; investigation, G.Z. and H.H.; resources, G.Z.; data curation, G.Z. and Z.Z.; writing—original draft preparation, G.Z.; writing—review and editing, G.Z., F.C., H.H., and Z.Z.; visualization, G.Z. and Z.Z.; supervision, F.C. and H.H.; project administration, F.C.; funding acquisition, F.C. and H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant number 2023YFB2504704) and the Natural Science Foundation in Hunan Province (grant number S2023JJQNJJ1969)

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

1. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 383–398. https://doi.org/10.1109/tits.2018.2815678.
2. Zhao, Y.; Tian, Z. *Applicability of Adaptive Traffic Control Systems in Nevada's Urban Areas*. No. 092-09-803. Nevada Department of Transportation: Carson City, NV, USA, 2011.
3. Federal Highway Administration. *Traffic Signal Timing Manual*. Technical Report FHWA-HOP-08-024, U.S. Department of Transportation: Washington, DC, USA, 2008.
4. Muralidharan, A.; Pedarsani, R.; Varaiya, P. Analysis of fixed-time control. *Transp. Res. Part B Methodol.* **2015**, *73*, 81–90.
5. Celtek, S.A.; Durdu, A.; Ali, M.E.M. Evaluating Action Durations for Adaptive Traffic Signal Control Based On Deep Q-Learning. *Int. J. Intell. Transp. Syst. Res.* **2021**, *19*, 557–571. https://doi.org/10.1007/s13177-021-00262-5.
6. Roess, R.P.; Prassas, E.S.; Mcshane, W.R. *Traffic Engineering*; Pearson/Prentice Hall: Hoboken, NJ, USA, 2014.
7. Zhou, P.; Fang, Z.; Dong, H.; Liu, J.; Pan, S. Data analysis with multi-objective optimization algorithm: A study in smart traffic signal system. In Proceedings of the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), London, UK, 7–9 June 2017; pp. 307–310.
8. Cesme, B.; Furth, P.G. Self-organizing traffic signals using secondary extension and dynamic coordination. *Transp. Res. Part C Emerg. Technol.* **2014**, *48*, 1–15. https://doi.org/10.1016/j.trc.2014.08.006.
9. Wang, X.B.; Yin, K.; Liu, H. Vehicle actuated signal performance under general traffic at an isolated intersection. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 582–598. https://doi.org/10.1016/j.trc.2018.08.002.
10. Eom, M.; Kim, B.-I. The traffic signal control problem for intersections: A review. *Eur. Transp. Res. Rev.* **2020**, *12*, 50. https://doi.org/10.1186/s12544-020-00440-8.
11. Wang, Y.; Yang, X.; Liang, H.; Liu, Y. A Review of the Self-Adaptive Traffic Signal Control System Based on Future Traffic Environment. *J. Adv. Transp.* **2018**, *2018*, 1096123. https://doi.org/10.1155/2018/1096123.
12. Stevanovic, A.; Kergaye, C.; Martin, P.T. Scoot and scats: A closer look into their operations. In Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington DC, USA, 11–15 January *2009*.
13. Zhao, D.; Dai, Y.; Zhang, Z. Computational Intelligence in Urban Traffic Signal Control: A Survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2011**, *42*, 485–494. https://doi.org/10.1109/tsmcc.2011.2161577.
14. Wei, H.; Zheng, G.; Gayah, V.; Li, Z. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. *ACM SIGKDD Explor. Newsl.* **2021**, *22*, 12–18.

15. Balaji, P.; German, X.; Srinivasan, D. Urban traffic signal control using reinforcement learning agents. *IET Intell. Transp. Syst.* **2010**, *4*, 177–188. https://doi.org/10.1049/iet-its.2009.0096.
16. Mikami, S.; Kakazu, Y. Genetic reinforcement learning for cooperative traffic signal control. First IEEE Conference on Evolutionary Computation. In Proceedings of the IEEE World Congress on Computational Intelligence, Orlando, FL, USA, 27–29 June 1994; pp. 223-228.
17. Dai, Y.; Hu, J.; Zhao, D.; Zhu, F. Neural network based online traffic signal controller design with reinforcement training. In Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems—(ITSC 2011), Washington, DC, USA, 5–7 October 2011; pp. 1045–1050.
18. Arel, I.; Liu, C.; Urbanik, T.; Kohls, A. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intell. Transp. Syst.* **2010**, *4*, 128. https://doi.org/10.1049/iet-its.2009.0070.
19. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 11–32. https://doi.org/10.1109/tits.2020.3008612.
20. Cel
21. , M.; Naik, A.; Goodman, L.; Crebo, J.; Abrar, T.; Abad, Z.S.H.; Bazzan, A.L.; Far, B. Reinforcement learning in urban network traffic signal control: A systematic literature review. *Expert Syst. Appl.* **2022**, *199*, 116830. https://doi.org/10.1016/j.eswa.2022.116830.
22. Gregurić, M.; Vujić, M.; Alexopoulos, C.; Miletić, M. Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data. *Appl. Sci.* **2020**, *10*, 4011. https://doi.org/10.3390/app10114011.
23. Liu, S.; Wu, G.; Barth, M. A Complete State Transition-Based Traffic Signal Control Using Deep Reinforcement Learning. In Proceedings of the 2022 IEEE Conference on Technologies for Sustainability (SusTech), Corona, CA, USA, 21–23 April 2022; pp. 100–107.
24. Vasconcelos, L.; Silva, A.B.; Seco, Á.M.; Fernandes, P.; Coelho, M.C. Turboroundabouts: Multicriterion assessment of intersection capacity, safety, and emissions. *Transp. Res. Rec.* **2014**, *2402*, 28–37.
25. Yao, R.; Wang, X.; Xu, H.; Lian, L. Emission factor calibration and signal timing optimisation for isolated intersections. *IET Intell. Transp. Syst.* **2018**, *12*, 158–167. https://doi.org/10.1049/iet-its.2016.0332.
26. Yao, R.; Sun, L.; Long, M. VSP-based emission factor calibration and signal timing optimisation for arterial streets. *IET Intell. Transp. Syst.* **2019**, *13*, 228–241. https://doi.org/10.1049/iet-its.2018.5066.
27. Hao, P.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Eco-Approach and Departure (EAD) Application for Actuated Signals in Real-World Traffic. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 30–40. https://doi.org/10.1109/TITS.2018.2794509.
28. Shepelev, V.; Glushkov, A.; Fadina, O.; Gritsenko, A. Comparative Evaluation of Road Vehicle Emissions at Urban Intersections with Detailed Traffic Dynamics. *Mathematics* **2022**, *10*, 1887. https://doi.org/10.3390/math10111887.
29. Shepelev, V.; Glushkov, A.; Slobodin, I.; Balfaqih, M. Studying the Relationship between the Traffic Flow Structure, the Traffic Capacity of Intersections, and Vehicle-Related Emissions. *Mathematics* **2023**, *11*, 3591. https://doi.org/10.3390/math11163591.
30. Jovanović, A.; Kukić, K.; Stevanović, A.; Teodorović, D. Restricted crossing U-turn traffic control by interval Type-2 fuzzy logic. *Expert Syst. Appl.* **2023**, *211*, 118613. https://doi.org/10.1016/j.eswa.2022.118613.
31. Zheng, L.; Li, X. Simulation-based optimization method for arterial signal control considering traffic safety and efficiency under uncertainties. *Comput. Civ. Infrastruct. Eng.* **2023**, *38*, 640–659. https://doi.org/10.1111/mice.12876.
32. Tsitsokas, D.; Kouvelas, A.; Geroliminis, N. Two-layer adaptive signal control framework for large-scale dynamically-congested networks: Combining efficient Max Pressure with Perimeter Control. *Transp. Res. Part C Emerg. Technol.* **2023**, *152*, 104128. https://doi.org/10.1016/j.trc.2023.104128.
33. Zhao, J.; Ma, W. An Alternative Design for the Intersections with Limited Traffic Lanes and Queuing Space. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1473–1483. https://doi.org/10.1109/tits.2020.2971353.
34. Li, J.; Peng, L.; Xu, S.; Li, Z. Distributed edge signal control for cooperating pre-planned connected automated vehicle path and signal timing at edge computing-enabled intersections. *Expert Syst. Appl.* **2024**, *241*, 122570. https://doi.org/10.1016/j.eswa.2023.122570.
35. Li, J.; Yu, C.; Shen, Z.; Su, Z.; Ma, W. A survey on urban traffic control under mixed traffic environment with connected automated vehicles. *Transp. Res. Part C Emerg. Technol.* **2023**, *154*, 104258. https://doi.org/10.1016/j.trc.2023.104258.
36. McKenney, D.; White, T. Distributed and adaptive traffic signal control within a realistic traffic simulation. *Eng. Appl. Artif. Intell.* **2013**, *26*, 574–583. https://doi.org/10.1016/j.engappai.2012.04.008.
37. Tan, W.; Li, Z.C.; Tan, Z.J. Modeling the effects of speed limit, acceleration, and deceleration on overall delay and traffic emission at a signalized intersection. *J. Transp. Eng. Part A-Syst.* **2017**, *143*, 04017063.
38. Shi, X.; Zhang, J.; Jiang, X.; Chen, J.; Hao, W.; Wang, B. Learning eco-driving strategies from human driving trajectories. *Phys. A Stat. Mech. Its Appl.* **2024**, *633*, 129353. https://doi.org/10.1016/j.physa.2023.129353.
39. Rabinowitz, A.I.; Ang, C.C.; Mahmoud, Y.H.; Araghi, F.M.; Meyer, R.T.; Kolmanovsky, I.; Asher, Z.D.; Bradley, T.H. Real-Time Implementation Comparison of Urban Eco-Driving Controls. *IEEE Trans. Control. Syst. Technol.* **2023**, *32*, 143–157. https://doi.org/10.1109/tcst.2023.3304910.
40. Li, Y.; Yang, Y.; Lin, X.; Hu, X. Traffic Information-Based Hierarchical Control Strategies for Eco-Driving of Plug-In Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2023**, *73*, 3206–3217. https://doi.org/10.1109/tvt.2023.3326989.

41. Dong, S.; Harzer, J.; Frey, J.; Meng, X.; Liu, Q.; Gao, B.; Diehl, M.; Chen, H. Cooperative Eco-Driving Control of Connected Multi-Vehicles With Spatio-Temporal Constraints. *IEEE Trans. Intell. Veh.* **2023**, *9*, 1733–1743. https://doi.org/10.1109/tiv.2023.3282490.
42. Zhang, Z.; Ding, H.; Guo, K.; Zhang, N. An Eco-driving Control Strategy for Connected Electric Vehicles at Intersections Based on Preceding Vehicle Speed Prediction. *IEEE Trans. Transp. Electrif.* **2024**, *PP*, 1–1. https://doi.org/10.1109/tte.2024.3410278.
43. Boukerche, A.; Zhong, D.; Sun, P. FECO: An Efficient Deep Reinforcement Learning-Based Fuel-Economic Traffic Signal Control Scheme. *IEEE Trans. Sustain. Comput.* **2021**, *7*, 144–156. https://doi.org/10.1109/tsusc.2021.3138926.
44. Ding, H.; Zhuang, W.; Dong, H.; Yin, G.; Liu, S.; Bai, S. Eco-Driving Strategy Design of Connected Vehicle among Multiple Signalized Intersections Using Constraint-enforced Reinforcement Learning. *IEEE Trans. Transp. Electrif.* **2024**, *PP*, 1–1. https://doi.org/10.1109/tte.2024.3396122.
45. Wang, Q.; Ju, F.; Wang, H.; Qian, Y.; Zhu, M.; Zhuang, W.; Wang, L. Multi-agent reinforcement learning for ecological car-following control in mixed traffic. *IEEE Trans. Transp. Electrification* **2024**, *PP*, 1–1. https://doi.org/10.1109/tte.2024.3383091.
46. Feng, J.; Lin, K.; Shi, T.; Wu, Y.; Wang, Y.; Zhang, H.; Tan, H. Cooperative traffic optimization with multi-agent reinforcement learning and evolutionary strategy: Bridging the gap between micro and macro traffic control. *Phys. A Stat. Mech. Its Appl.* **2024,** 647, 129734. https://doi.org/10.1016/j.physa.2024.129734.
47. Krishankumar, R.; Pamucar, D.; Deveci, M.; Ravichandran, K.S. Prioritization of zero-carbon measures for sustainable urban mobility using integrated double hierarchy decision framework and EDAS approach. *Sci. Total. Environ.* **2021**, *797*, 149068. https://doi.org/10.1016/j.scitotenv.2021.149068.
48. Liu, J.; Wang, C.; Zhao, W. An eco-driving strategy for autonomous electric vehicles crossing continuous speed-limit signalized intersections. *Energy* **2024**, *294*, 130829. https://doi.org/10.1016/j.energy.2024.130829.
49. Zhang, X.; Fan, X.; Yu, S.; Shan, A.; Fan, S.; Xiao, Y.; Dang, F. Intersection Signal Timing Optimization: A Multi-Objective Evolutionary Algorithm. *Sustainability* **2022**, *14*, 1506. https://doi.org/10.3390/su14031506.
50. Zhang, G.; Chang, F.; Jin, J.; Yang, F.; Huang, H. Multi-objective deep reinforcement learning approach for adaptive traffic signal control system with concurrent optimization of safety, efficiency, and decarbonization at intersections. *Accid. Anal. Prev.* **2024**, *199*, 107451. https://doi.org/10.1016/j.aap.2023.107451.
51. Salem, S.; Leonhardt, A. Optimizing Traffic Adaptive Signal Control: A Multi-Objective Simulation-Based Approach for Enhanced Transportation Efficiency. In Proceedings of the 10th International Conference on Vehicle Technology and Intelligent Transport Systems – VEHITS, Angers, France, 2-4 May 2024; pp. 344-351. https://doi.org/10.5220/0012682100003702.
52. Lin, Z.; Gao, K.; Wu, N.; Suganthan, P.N. Problem-Specific Knowledge Based Multi-Objective Meta-Heuristics Combined Q-Learning for Scheduling Urban Traffic Lights With Carbon Emissions. *IEEE Trans. Intell. Transp. Syst.* **2024**, *PP*, 1–12. https://doi.org/10.1109/tits.2024.3397077.
53. Deshpande, S.R.; Jung, D.; Bauer, L.; Canova, M. Integrated Approximate Dynamic Programming and Equivalent Consumption Minimization Strategy for Eco-Driving in a Connected and Automated Vehicle. *IEEE Trans. Veh. Technol.* **2021**, *70*, 11204–11215. https://doi.org/10.1109/tvt.2021.3102505.
54. Wan, C.; Shan, X.; Hao, P.; Wu, G. Multi-objective coordinated control strategy for mixed traffic with partially connected and automated vehicles in urban corridors. *Phys. A Stat. Mech. Its Appl.* **2024**, *635*, 129485. https://doi.org/10.1016/j.physa.2023.129485.
55. Boukerche, A., Zhong, D., & Sun, P. (2021). Feco: An efficient deep reinforcement learning-based fuel-economic traffic signal control scheme. *IEEE Trans. Sustain. Comput.* **2021**, 7(1), 144-156.
56. Jamil, A.R.M.; Ganguly, K.K.; Nower, N. Adaptive traffic signal control system using composite reward architecture based deep reinforcement learning. *IET Intell. Transp. Syst.* **2020**, *14*, 2030–2041. https://doi.org/10.1049/iet-its.2020.0443.
57. Liu, C.; Sheng, Z.; Chen, S.; Shi, H.; Ran, B. Longitudinal control of connected and automated vehicles among signalized intersections in mixed traffic flow with deep reinforcement learning approach. *Phys. A Stat. Mech. Its Appl.* **2023**, *629*, 129189. https://doi.org/10.1016/j.physa.2023.129189.
58. Hua, C.; Fan, W.D. Safety-oriented dynamic speed harmonization of mixed traffic flow in nonrecurrent congestion. *Phys. A Stat. Mech. Its Appl.* **2024**, *634*, 129439. https://doi.org/10.1016/j.physa.2023.129439.
59. Jamil, A.R.M.; Nower, N. A Comprehensive Analysis of Reward Function for Adaptive Traffic Signal Control. *Knowl. Eng. Data Sci.* **2021**, *4*, 85–96. https://doi.org/10.17977/um018v4i22021p85-96.
60. Ahmed, A.A.; Malebary, S.J.; Ali, W.; Barukab, O.M. Smart Traffic Shaping Based on Distributed Reinforcement Learning for Multimedia Streaming over 5G-VANET Communication Technology. *Mathematics* **2023**, *11*, 700. https://doi.org/10.3390/math11030700.
61. Agafonov, A.; Yumaganov, A.; Myasnikov, V. Cooperative Control for Signalized Intersections in Intelligent Connected Vehicle Environments. *Mathematics* **2023**, *11*, 1540. https://doi.org/10.3390/math11061540.
62. Genders, W.; Razavi, S. Evaluating reinforcement learning state representations for adaptive traffic signal control. *Procedia Comput. Sci.* **2018**, *130*, 26–33. https://doi.org/10.1016/j.procs.2018.04.008.
63. Dong, L.; Xie, X.; Lu, J.; Feng, L.; Zhang, L. OAS Deep Q-Learning-Based Fast and Smooth Control Method for Traffic Signal Transition in Urban Arterial Tidal Lanes. *Sensors* **2024**, *24*, 1845. https://doi.org/10.3390/s24061845.

64. Aslani, M.; Mesgari, M.S.; Wiering, M. Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events. *Transp. Res. Part C Emerg. Technol.* **2017**, *85*, 732–752. https://doi.org/10.1016/j.trc.2017.09.020.
65. Touhbi, S.; Babram, M.A.; Nguyen-Huu, T.; Marilleau, N.; Hbid, M.L.; Cambier, C.; Stinckwich, S. Adaptive Traffic Signal Control : Exploring Reward Definition For Reinforcement Learning. *Procedia Comput. Sci.* **2017**, *109*, 513–520. https://doi.org/10.1016/j.procs.2017.05.327.
66. Li, D.; Wu, J.; Xu, M.; Wang, Z.; Hu, K. Adaptive Traffic Signal Control Model on Intersections Based on Deep Reinforcement Learning. *J. Adv. Transp.* **2020**, *2020*, 6505893. https://doi.org/10.1155/2020/6505893.
67. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. https://doi.org/10.1109/msp.2017.2743240.
68. Van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016. https://doi.org/10.1609/aaai.v30i1.10295.
69. Schaul, T.; Quan, J.; Antonoglou, I.; Silver, D. Prioritized experience replay. *arXiv* **2015**, arXiv:1511.05952.
70. Wang, Z.; Schaul, T.; Hessel, M.; Hasselt, H.; Lanctot, M.; Freitas, N. Dueling network architectures for deep reinforcement learning. *Int. Conf. Mach. Learn.* PMLR, 2016; pp. 1995–2003. https://proceedings.mlr.press/v48/wangf16.html.
71. Liang, X.; Du, X.; Wang, G.; Han, Z. A Deep Reinforcement Learning Network for Traffic Light Cycle Control. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1243–1253. https://doi.org/10.1109/tvt.2018.2890726.
72. Chu, T.; Wang, J.; Codeca, L.; Li, Z. Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1086–1095. https://doi.org/10.1109/TITS.2019.2901791.
73. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
74. Pang, H.; Gao, W. Deep Deterministic Policy Gradient for Traffic Signal Control of Single Intersection. In Proceedings of the 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 5861–5866.
75. Fujimoto, S.; Hoof, H.; Meger, D. Addressing function approximation error in actor-critic methods. In *International Conference on Machine Learning*; PMLR: 2018; pp. 1587–1596. https://proceedings.mlr.press/v80/fujimoto18a.html.
76. Ding, Z.; Huang, Y.; Yuan, H.; Dong, H. Introduction to reinforcement learning. In *Deep Reinforcement Learning*; Springer: Singapore, 2020; pp. 47–123.
77. Zeng, J.; Hu, J.; Zhang, Y. Training Reinforcement Learning Agent for Traffic Signal Control under Different Traffic Conditions. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference—ITSC, Auckland, New Zealand, 27–30 October 2019.
78. Jin, J.; Li, Y.; Huang, H.; Dong, Y.; Liu, P. A variable speed limit control approach for freeway tunnels based on the model-based reinforcement learning framework with safety perception. *Accid. Anal. Prev.* **2024**, *201*, 107570. https://doi.org/10.1016/j.aap.2024.107570.
79. Krajzewicz, D.; Behrisch, M.; Wagner, P.; Luz, R.; Krumnow, M. Second generation of pollutant emission models for SUMO. In *Modeling Mobility with Open Data*; Springer, Cham, Switzerland, 2015; pp. 203–221.
80. Jin, J.; Huang, H.; Yuan, C.; Li, Y.; Zou, G.; Xue, H. Real-time crash risk prediction in freeway tunnels considering features interaction and unobserved heterogeneity: A two-stage deep learning modeling framework. *Anal. Methods Accid. Res.* **2023**, *40*, 100306. https://doi.org/10.1016/j.amar.2023.100306.
81. Wu, Y.; Ho, C. The development of taiwan arterial traffic-adaptive signal control system and its field test: A taiwan experience. *J. Adv. Transp.* **2009**, *43*, 455–480. https://doi.org/10.1002/atr.5670430404.
82. Tsang, K.; Hung, W.; Cheung, C. Emissions and fuel consumption of a Euro 4 car operating along different routes in Hong Kong. *Transp. Res. Part D Transp. Environ.* **2011**, *16*, 415–422. https://doi.org/10.1016/j.trd.2011.02.004.
83. Choudhary, A.; Gokhale, S. Urban real-world driving traffic emissions during interruption and congestion. *Transp. Res. Part D Transp. Environ.* **2016**, *43*, 59–70. https://doi.org/10.1016/j.trd.2015.12.006.
84. Zhou, X.; Tanvir, S.; Lei, H.; Taylor, J.; Liu, B.; Rouphail, N.M.; Frey, H.C. Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies. *Transp. Res. Part D Transp. Environ.* **2015**, *37*, 123–136. https://doi.org/10.1016/j.trd.2015.04.013.
85. Clarke, P.; Muneer, T.; Cullinane, K. Cutting vehicle emissions with regenerative braking. *Transp. Res. Part D Transp. Environ.* **2010**, *15*, 160–167.
86. Gallus, J.; Kirchner, U.; Vogt, R.; Benter, T. Impact of driving style and road grade on gaseous exhaust emissions of passenger vehicles measured by a Portable Emission Measurement System (PEMS). *Transp. Res. Part D Transp. Environ.* **2017**, *52*, 215–226. https://doi.org/10.1016/j.trd.2017.03.011.
87. Ye, Q.; Chen, X.; Liao, R.; Yu, L. Development and evaluation of a vehicle platoon guidance strategy at signalized intersections considering fuel savings. *Transp. Res. Part D Transp. Environ.* **2020**, *77*, 120–131. https://doi.org/10.1016/j.trd.2019.10.020.
88. Pandian, S.; Gokhale, S.; Ghoshal, A.K. Evaluating effects of traffic and vehicle characteristics on vehicular emissions near traffic intersections. *Transp. Res. Part D Transp. Environ.* **2009**, *14*, 180–196. https://doi.org/10.1016/j.trd.2008.12.001.

89. Boryaev, A.; Malygin, I.; Marusin, A. Areas of focus in ensuring the environmental safety of motor transport. *Transp. Res. Procedia* **2020**, *50*, 68–76. https://doi.org/10.1016/j.trpro.2020.10.009.
90. Grote, M.; Williams, I.; Preston, J.; Kemp, S. A practical model for predicting road traffic carbon dioxide emissions using Inductive Loop Detector data. *Transp. Res. Part D Transp. Environ.* **2018**, *63*, 809–825. https://doi.org/10.1016/j.trd.2018.06.026.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.