Alexandria Engineering Journal 117 (2025) 99–115

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original article

Simulation optimization of highway hard shoulder running based on multi-agent deep deterministic policy gradient algorithm

Lipeng Hu [a], Jinjun Tang [a,*], Guoqing Zou [b], Zhitao Li [a], Jie Zeng [a], Mingyang Li [a]

[a] Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha, China
[b] Hunan Pingyi Expressway Construction Development Company Limited, Yueyang, China

ARTICLE INFO

Keywords:
Highway
Hard Shoulder Running
MADDPG
Spatial-temporal constrains
SUMO

ABSTRACT

To alleviate traffic congestion and reduce vehicle emissions, the use of hard shoulder running (HSR) has emerged as a sustainable and cost-effective active traffic management technology. However, optimizing the utilization of HSR remains a critical challenge for improving highway traffic congestion. To tackle this issue, the Multi-Agent Deep Deterministic Policy Gradient with spatio-temporal constraints (STC-MADDPG) algorithm based on multi-agent reinforcement learning is proposed in this paper. To verify the effectiveness of the proposed algorithm, the present study utilizes a Simulation of Urban Mobility (SUMO) platform to construct a simulation environment. The optimal HSR strategy is then determined for four different service levels of highways. Additionally, the granularity of control is adjusted by varying the number of agents, allowing for a comprehensive analysis and evaluation of the varying effectiveness of different control levels across different service levels. Through in-depth investigation into the two strategies under the fourth service level, it is discovered that fewer sections each agent controls yields better results when congestion becomes more severe. The experimental results clearly demonstrate the superiority of the optimized strategy for HSR using the STC-MADDPG algorithm, compared to the "no open" strategy. Specifically, the maximum reductions achieved in terms of total vehicle travel time, Time Integrated Time-to-collision, CO emissions, CO_2 emissions, and NO_x emissions are 37.4 %, 34.1 %, 28.0 %, 17.1 %, and 27.2 % respectively. This comprehensive evaluation of the algorithm's effectiveness covers three key aspects: driving efficiency, driving safety, and environmental protection. The findings conclusively demonstrate the positive impact of the proposed algorithm on all three fronts.

1. Introduction

Highway hard shoulder running (HSR) is an active traffic management measure aimed at temporarily utilizing the hard shoulder, which can effectively reduce congestion and enhance traffic flow by increasing road capacity [1]. By allowing vehicles, or specific types of vehicles, to use the hard shoulder as an additional lane, the pressure on the main road can be alleviated, particularly during peak hours. The implementation of this measure typically relies on real-time traffic flow conditions, and the decision to open the hard shoulder is made by traffic management operations [2].

The implementation of HSR can be categorized into three main types: BOS (bus-on-shoulder), static, and dynamic HSR [3]. Among these, dynamic HSR stands out as it allows the temporary utilization of shoulders as additional lanes based on real-time traffic conditions, without being restricted to specific time periods or vehicle types [4]. This approach has gained widespread adoption worldwide, with over 700 miles of dynamic HSR implemented in Europe alone [5]. The impact of different HSR strategies on highway operation systems has been the subject of numerous studies conducted by researchers. These studies aim to understand and analyze the effects and benefits of various strategies, shedding light on their implications for optimizing overall highway operations.

Several studies have investigated the impact of HSR strategies on highway operational efficiency, emissions, and safety, among other factors. These studies have revealed that the use of HSR strategies can reduce delays, total consumption time, emissions, and road congestion, increase average speeds and throughput, and improve highway accident management. However, the current literature lacks proper analysis of the differences between strategies and how to optimize their impact across various indicators. Moreover, there is a lack of problem-specific analysis for different traffic flow scenarios, and the environmental

* Corresponding author.
E-mail address: jinjuntang@csu.edu.cn (J. Tang).

https://doi.org/10.1016/j.aej.2024.12.110
Received 22 August 2023; Received in revised form 4 November 2024; Accepted 30 December 2024
Available online 11 January 2025

setting conditions are often unspecified. While some studies have explored reasonable and effective HSR strategy opening to improve the impact of strategies on highway efficiency, many such studies use traditional heuristic algorithms such as genetic algorithms to optimize travel time while ignoring advanced optimization techniques such as dynamic control methods. In summary, current research on HSR strategies has several limitations. Firstly, existing studies lack sufficient depth of analysis on the impact of different strategies on highway efficiency, safety, and other factors, and application effects of various HSR strategies are vaguely compared. Secondly, there is limited research on how to optimize HSR strategies to improve highway operational efficiency or traffic safety. Using traditional heuristic algorithms in the optimization making its inadequate ability for adapting and optimizing dynamic HSR control methods. Lastly, there is inadequate research on optimizing HSR strategies under different traffic flow conditions.

In order to address the shortcomings of the above study and to optimally solve the HSR control strategy problem under different traffic flow states, this paper proposes the Multi-Agent Deep Deterministic Policy Gradient with spatio-temporal constraints (STC-MADDPG) algorithm. Initially, the study mathematically models the HSR control optimization problem and introduces a spatio-temporal constraint for the open strategy. Subsequently, this paper integrates the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm and spatio-temporal constraints to formulate the STC-MADDPG algorithm. Furthermore, the HSR control strategy is optimized by applying varying numbers of agent control methods at four levels of service on the highway. The optimized strategy exhibits benefit in terms of efficiency, safety, and environmental impact. Lastly, an extended study examines four levels of service, comparing the effectiveness of two control granularity strategies utilizing four and eight agents during severe congestion. Through the aforementioned research components, this paper provides a mathematical description of the HSR strategy optimization problem. It presents a feasible and effective method for optimizing solutions across different service levels. Moreover, the analysis compares and evaluates the impact of different control granularity strategies in the context of severe congestion, offering valuable insights for engineering applications.

The structure of this paper is described, Section 2 presents the problem to be optimised and the mathematical formulation of the problem. Section 3 elaborates on the proposed methodology. Section 4 describes the setting of the various elements of the experimental environment. Section 5 analyses the experimental results. Section 6 summarizes the results and shortcomings of the research.

2. Literature review

The HSR was introduced early in several European countries as a traffic management measure. In Germany, the utilization of hard shoulders, either temporarily or permanently, has been studied since 2000, and has significant implications for road safety, traffic flow management, and highway operation [6]. By using the hard shoulder as an additional travel lane, it has been observed that highway congestion can be reduced, and traffic flow efficiency can be increased [7–20]. Specifically, the implementation of the dynamic HSR strategy has been found to effectively reduce departure delays [21], significantly reduce travel times during peak commuting hours [19], increase average speeds, and enhance vehicle throughput [22]. Additionally, the dynamic HSR has a crucial role to play in reducing vehicle emissions and environmental pollution. For example, Truck-Specific HSR (T-HSR) strategies have been proposed and implemented to reduce vehicle CO_2 and NO_X emissions [23]. HSR measures have also been employed theoretically as well as practically to decrease fuel consumption and emissions by up to 41–44 % [20]. There exist numerous studies on the impact of implementing dynamic HSR strategies on road safety [6–9,19,21–34]. According to these studies, the implementation of dynamic HSR has proven to effectively reduce congestion and other causes of accidents, thereby decreasing collision rates and enhancing safety on urban weaving sections, reducing the impact of episodic accidents [27,31]. In conclusion, reasonable and effective HSR strategies can alleviate highway congestion, improve operational efficiency, reduce accidents, and enhance highway safety.

The numerous benefits of HSR strategies for all aspects of highways have led to extensive research from scholars. Kellermann et al. proposed five ways to adapt to changing traffic conditions by opening the hard shoulder appropriately. It was discovered that mobile HSR strategies improve traffic quality, reduce congestion, do not compromise safety, are cost-effective to maintain, and incur lower costs than adding a third lane [9]. The substantial benefits of HSR strategies relative to the investment have prompted national researchers to consider their implementation. In the United Kingdom (UK), Chase et al. outlined the advantages and limitations of HSR, emphasizing that with the right information, targeted driver education, and training, implementation of HSR strategies could be significantly successful [31]. In Germany, Geistefeldt et al. conducted a study on the temporary use of hard shoulders on the A3 and A5 highways in the Black Forest region to analyze the effects on highway traffic flow and road safety. The study revealed that HSR could increase capacity by 20–25 % without compromising road safety [10]. In France, Aron et al. evaluated and analyzed two partially implemented hard shoulder strategies and found that HSR could either increase or decrease the number of accidents [25]. Furthermore, Lemke et al. pointed out that in 2010, the pilot distance on the hard shoulder in Germany was already 200 km [32]. The extensive implementation of HSR strategies and its impact on highway safety have been studied by numerous researchers. According to Kononov et al., HSR was effective in reducing collision rates by decreasing traffic volume per lane, with safety benefits outweighing adverse effects [26]. Additionally, in the works of [33], Zeng et al. improved the safety of hard shoulders using a joint empirical Bayesian approach. In the works of [30], Maurice et al. carried out a safety assessment of the implementation of dynamic HSR measures on braided sections of urban highway in French, assessed the impact of HSR strategies on highway safety.

Effectively implementing the HSR strategy requires identifying the appropriate trigger conditions and controlling critical elements such as control time, control space, and other parameters. Scholars have proposed different solutions to optimize the strategy. Li et al. investigated an optimal framework for managing highway traffic congestion with three control modes: ramp metering, variable speed limit, and hard shoulder running. They took the problem as an integer linear programming task, with the objective of reducing the total highway delay, and used the IBM CPLEX to solve it [18]. Li et al. proposed a genetic algorithm with time windows to optimize the hard shoulder control strategy, which reduced the total travel time of the road network by 30.61 % [35]. Zhou et al. proposed a reinforcement learning algorithm, namely the Q-learning algorithm, to optimally coordinate Variable Speed Limit (VSL) and HSR control strategies. Experimental results showed that the proposed method reduced travel time by up to 27 % [36]. Hussein et al. introduced a hybrid operating system for HSR that consists of three modules, namely a data manager, technology engine, and transportation management center [37]. By using road sensors to collect traffic data and applying hyperbolic fuzzy affiliation functions, it was possible to analyze real-time traffic flow states and decide on activating hard shoulder control. The transportation management center executed these decisions. Fan et al. employed the K-means clustering algorithm to categorize traffic states into three clusters and used factor analysis and TOPSIS methods to identify the ideal conditions for activating hard shoulders [38]. Arora et al. considered the dynamic combination of VSL and HSR to design an integrated control strategy based on model predictive control. This strategy increased the average speed and vehicle throughput by 21.09 % and 33.44 %, respectively, in experimental studies carried out on the Deerfoot Trail section in Calgary, Alberta [22]. Although there is little research on HSR strategy optimization methods, characteristic methods such as genetic

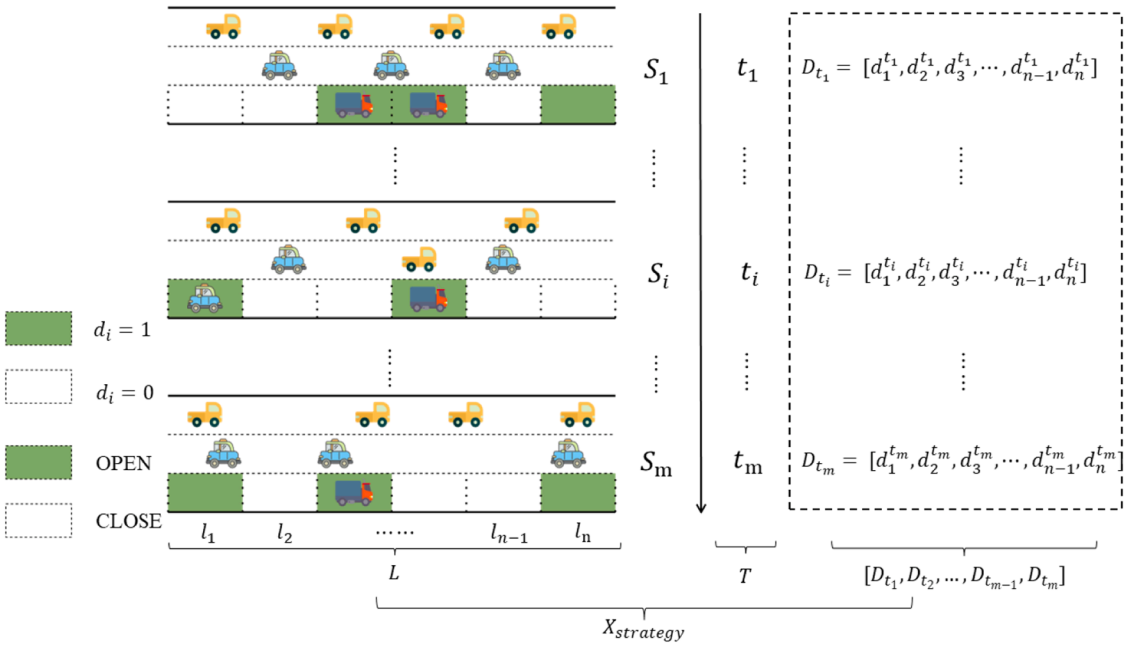


Fig. 1. Strategy description for HSR.

algorithms, integer programming methods, and Q-learning methods have been proposed. However, all these methods are appropriate for static control methods. They pose encoding solutions that result in dimensional explosion, especially for genetic algorithms when the path length is long and the control time is uncertain. Q-learning is a reinforcement learning algorithm that is inefficient and unsuitable for solving problems with large action state spaces. Advanced models such as multi-agent deep reinforcement learning algorithms are efficient in solving control and management problems. The problem is divided into more subproblems, each of which is solved separately by one agent. Roadway-based traffic control, such as ramp meter, variable speed limit, hard shoulder running, traffic signal control, etc., can achieve better results by applying the MARL compared to traditional methods. Chen et al. used a multi-agent reinforcement learning algorithm to control the hybrid vehicles for ramp meter, which resulted in an increase in the average speed of the vehicle fusion and the conflict rate compared to the traditional algorithms [39]. Fang et al. used a multi-agent proximal optimization algorithm for variable speed limit control, which resulted in a more stable traffic flow on the highway and reduced the waiting time of the entire network by 15.8 % [40]. Based on the multi-agent reinforcement learning algorithm, the W-Learning algorithm was used to control the congested zone section and dynamically adjust the speed limit, which led to an 18 % improvement in the traffic parameters [41]. In addition, Máté et al. optimized traffic signal control through a multi-agent deep Q-learning approach reduced fuel consumption by 11 % and average travel time by 13 % for regional traffic [42]. MARL, a data-driven approach based on the application of data to traffic control systems, effectively utilizes the monitored road information data for different scenarios where the model is more effective. This paper designs the HSR strategy optimization as a sequential decision problem, exploring advanced multi-agent reinforcement learning algorithms as a new approach to HSR strategy control.

This paper proposes a novel approach adapting the optimized dynamic HSR control strategy to address the issues of dimensional explosion and inefficiency that existing methods pose. Starting with the mathematical model of the HSR strategy control problem, this study considers multiple traffic flow scenarios and implements the multi-agent reinforcement learning method to optimize the HSR strategy. The effectiveness of the proposed method is evaluated in terms of efficiency, safety, and environmental impact [43,44].

3. Problem

3.1. Problem description

HSR control is an active traffic management measure that involves the timely opening of the hard shoulder when the highway experiences traffic overload or congestion due to accidents or incidents. Effective control of the opening time and distance of the hard shoulder is crucial for its optimal utilization. To address this, this study proposes segmenting the entire highway section and dividing it into n sub-section sets, denoted as $L = \{l_1, l_2, l_3, \cdots, l_{n-1}, l_n\}$, where each section is defined as the minimum control unit. Moreover, the control time is discretized into segments $T = (t_1, t_2, t_3, \cdots, t_{m-1}, t_m)$, with the smallest time slice, t_i, representing the unit of open time. The sequence $Decision_{squence} = (\{\{L_{sub1}\}, t_1), (\{L_{sub2}\}, t_2), \cdots, (\{L_{subm}\}, t_m))$ represents the mathematical expression of the HSR strategy (shown in Fig. 1). In this expression, $(\{L_{sub1}\}, t_1)$ indicates the opening of all road sections in L_{sub1}, a subset of L, for a duration of t_1, followed by the closure of all open sections at the end of t_1, awaiting the next decision. The strategy decision time, being very short, is excluded from the control time and, thus, disregarded. The subset $\{L_{subi}\}$ is empty if there is no need to open any section during a particular time slice t_i. Therefore, the hard shoulder control strategy problem can be transformed into optimizing the $Decision_{squence}$. When the control time is fixed, T is determined, and m represents the number of open time slices, indicating the number of decisions to be made. In the case of fixed control time, the section length is also determined, resulting in static control, wherein the hard shoulder is open during specified time periods and closed during other times. Conversely, when the control time is uncertain, T becomes a variable, and the number of open time slices, m, also becomes a variable, leading to dynamic control. Since static control is typically applied to characteristic events or specific time intervals, this study focuses on solving the dynamic HSR control problem with variable time periods.

3.2. Mathematical model of the problem

The subsequent section will discuss the mathematical formulation of dynamic control strategies. As depicted in Fig. 1, let each road segment, denoted as $L = \{l_1, l_2, l_3, \cdots, l_{n-1}, l_n\}$, correspond to decision variables $D = \{d_1, d_2, d_3, \cdots, d_{n-1}, d_n\}$. Here, road segment l_i

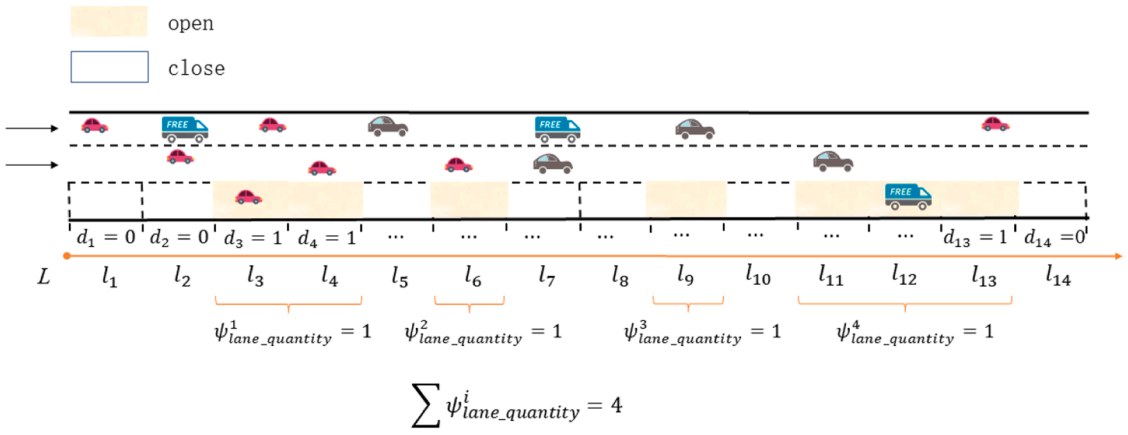


Fig. 2. Constraint diagram.

corresponds to decision variable d_i, which is a binary variable (0 or 1). When $d_i = 0$, it means that section l_i is in the closed state during time slice t_i. Conversely, when $d_i = 1$, it indicates that section l_i is open during time slice t_i. Let $\{L_{subi}\}$ represent the set of all roadway decision variables d_j with a value of 1. Within time slice t_i, the open state of the section can be represented by the tuple $(\{L_{subi}\}, t_i)$. Consequently, for the time series $(t_1, t_2, t_3, \cdots, t_{m-1}, t_m)$, the set of open road sections $((\{L_{sub1}\}, t_1), (\{L_{sub2}\}, t_2), \cdots, (\{L_{subm}\}, t_m))$ is determined, which represents the open status of road sections throughout the control time period. The current roadway open strategy is defined as the combination $D_{t_i} = [d^{t_i}_1, d^{t_i}_2, d^{t_i}_3, \cdots, d^{t_i}_{n-1}, d^{t_i}_n]$ comprising all roadway decision variables within time slice t_i, which is a column vector. The variable corresponding to the strategy considering the entire section length L and control time T is denoted as $X_{strategy} = [D_{t_1}, D_{t_2}, \ldots, D_{t_{m-1}}, D_{t_m}]$, forming an $n \times m$ matrix of binary variables (0 or 1). During time slice t_i, the hard shoulder is opened using the combination of D_{t_i}, and the highway operates in state S_i. At time slice t_{i+1}, the highway's operational state transitions to S_{i+1}. The operational state of the road segment during time slice t_{i+1} is solely influenced by the preceding time slice t_i, giving the problem has Markovian property.

We assume that various indicators of highway efficiency, safety, and environmental impact are characterized by a function Y. The application of a specific hard shoulder control strategy, denoted as $X'_{strategy}$, affects each indicator of the highway, with the corresponding function values being denoted as Y'. Thus, there exists a relationship between function Y and strategy $X_{strategy}$. However, it is not feasible to explicitly resolve the expression between Y and $X_{strategy}$ using function analysis. Instead, the corresponding value of Y can be determined by simulating the highway's operational state for different $X_{strategy}$ configurations. To find the optimal value of Y corresponding to $X_{strategy}$, various optimization algorithms such as genetic algorithms, particle swarm algorithms, reinforcement learning algorithms, and others can be employed. In cases where the control time T is uncertain, and the value of m in the strategy variable matrix $X_{strategy} = [D_{t_1}, D_{t_2}, \ldots, D_{t_{m-1}}, D_{t_m}]$ is not fixed, traditional heuristic algorithm optimization methods are not applicable. By modeling the problem as a sequential decision process and leveraging a reinforcement learning algorithm, the problem can be solved even with a variable value of m. The specific modeling approach is elaborated in the subsequent sections. Given that the aforementioned problem represents a sequential decision problem with Markovian properties, it is feasible to employ a reinforcement learning algorithm for its resolution [45].

3.3. Problem constraints

In consideration of practical requirements, the control strategy should account for both time and space constraints to be effective. Time constraints involve the count of jump changes, which is the sum of changes in all roadway decision variables d_i between two consecutive time slices $[t_i \sim t_{i+1}]$. A jump change occurs when the control variable d_i transitions from 0 to 1 or from 1 to 0 within the time window. Eq. (1) illustrates this relationship, where $d_i(t_j)$ represents the switching state of the i-th section at time slice t_j, and $d_i(t_{j+1})$ represents the switching state at time slice t_{j+1}. The parameter Φ_{time} denotes the total number of jump changes, reflecting the overall transformation of the decision variable D within two adjacent time slices $[t_i \sim t_{i+1}]$. Frequent opening and closing of the hard shoulder can increase lane change occurrences, consequently elevating driving risks. To ensure highway safety, it is essential to establish reasonable time constraints and maintain the number of jump changes within an acceptable range. Therefore, careful consideration is given to time constraints, balancing the need for efficient traffic flow while prioritizing driving safety.

Incorporating the principles of connected components from mathematical graph theory, the dispersion level of the open state of the hard shoulder is characterized by the count of lane components. Each time slice t_i, counts the spatially adjacent and simultaneously open hard shoulder subsegments as a single lane component. As illustrated in Fig. 2, the entire lane is represented by L, while l_i represents the subsections under control. $d_i = 0$ means the subsegment is closed. $d_i = 1$ means the subsegment is open. The $\psi_{lane\_quantity} = 1$ indicates the existence of a continuous open subsegment forming a lane component. At a given time slice t_i, the sum of all components within the entire lane corresponds to the total count of lane components, denoted as $\Psi_{lane\_quantity}$. In the provided case of Fig. 2, the total number of lane components $\Psi_{lane\_quantity}$ is 4. Eq. (2) represents the constraint on the number of lane components.

$$\sum_{i=1}^{L} \sum_{j=1}^{T} \left(d_i(t_{j+1}) - d_i(t_j)\right)^2 \le \Phi^*_{time} \tag{1}$$

$$\sum \Psi_{lane\_quantity} \le \Psi^*_{lane\_quantity} \tag{2}$$

4. Methodology

4.1. Multi-agent reinforcement learning

Reinforcement learning is a method of autonomous learning through the interaction of an agent with its environment, developed from a combination of statistics, psychology, and other disciplines [46]. The agent perceives the external state of environment and selects actions accordingly. It then evaluates the rewards associated with the impact of those actions and adjusts its strategy for subsequent action selection based on the reward outcomes. In this manner, the agent strives to perform optimally across various environmental states, aiming to

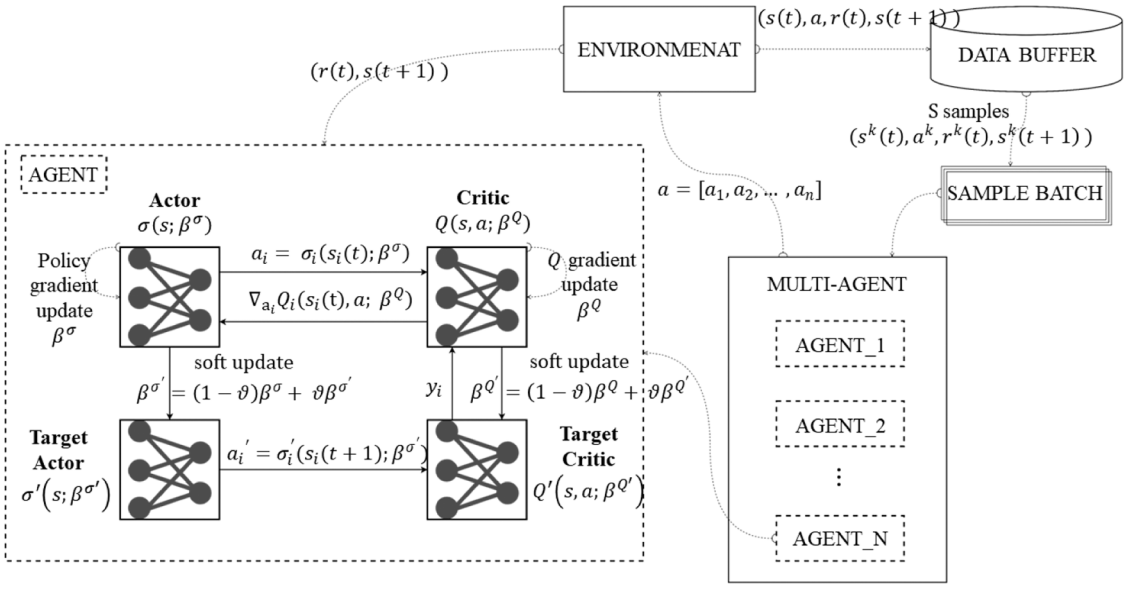


Fig. 3. Framework of MADDPG algorithm.

achieve the most favorable outcomes.

Deep learning has emerged as a prominent avenue for advancing machine learning techniques, offering remarkable capabilities in characterizing high-dimensional data structures across various domains such as images, videos, and texts [47]. One challenging task for an agent is to perceive the real-world environment and derive meaningful representations from high-dimensional inputs. It is necessary to conduct sophisticated approaches to solve this complex problem. Utilizing deep neural networks in conjunction with reinforcement learning methods can improve the effectiveness of the latter. This approach is referred to deep reinforcement learning, and it takes issues like data representation in reinforcement learning. Deep reinforcement learning has demonstrated significant progress of diverse domains [48–50], particularly in fields like robotics and gaming. Volodymyr et al. proposed a fusion of deep neural networks with Q-learning methods in reinforcement learning, leading to significant breakthroughs by replacing human agents with deep Q-networks. Extensive testing was conducted on 49 games of Atari 2600, achieving performance comparable to professional human players and elevating deep reinforcement learning as a major research focus [51]. In complex systems involving multiple control entities, the problem often needs to be modeled as a multi-agent system [52–56]. In such systems, multiple agents control objects and make behavioral decisions while interacting within a shared environment. Each agent learns from its environment, constructs its behavioral framework, and pursues specific objectives. Examples of multi-agent systems encompass multiplayer games, multi-drone formations, multi-robot collaboration, and collaborative driving of multiple vehicles [57–59]. The application of deep reinforcement learning methods to multi-agent systems, tackling diverse problems, is known as multi-agent reinforcement learning.

Problems addressed by multi-agent reinforcement learning algorithms frequently exhibit Markovian properties, characterized by a tuple representation $(S, A^{\{1,\cdots,N\}}, P, S', R^{\{1,\cdots,N\}}, N, \tau)$. Here, N denotes the number of agents, S denotes the observed data of all agents $S = (o_1, o_2, \cdots, o_N)$, $A^{\{1,\cdots,N\}}$ denotes the joint action of all agents $A^{\{1,\cdots,N\}} = (A^1 \times A^2 \times \cdots \times A^N)$, P represents the state transfer probability, indicating the probability distribution of transitioning from the current state S to the next state S' upon taking joint action A. $R^{\{1,\cdots,N\}}$ represents the reward set for all agents, and the objective of the multi-agent reinforcement learning algorithm is to maximize the cumulative rewards of all agents. This objective can be expressed as $J_{MDR} = \max E\left[\sum_{i=1}^{i=N} R_i\right] = \max E\left[\sum_{i=1}^{i=N} \sum_{t=0}^{T} \tau^t r_i^t\right]$, where $\tau \epsilon [0, 1)$ is the discount factor, which focuses the agent on immediate rewards while attenuating the impact of future rewards [60].

4.2. MADDPG method

Multi-agent reinforcement learning methods can be classified into three types: cooperative, competitive, and a mixture of both, depending on the setting between the agents [61]. The MADDPG algorithm is a hybrid cooperative-competitive multi-agent reinforcement learning algorithm built upon the Actor-Critic network framework, which is an extension of the single-agent reinforcement learning algorithm Deep Deterministic Policy Gradient (DDPG) [62]. To ensure training stability, a single agent is equipped with a dual Actor-Critic framework, wherein each agent has two Actor networks and two Critic networks. The first set represents the current network, while the second set denotes the target network [63]. The Actor network receives information about the state of the external environment sensed by the agent and gives the corresponding action distribution. The Critic network scores and evaluates the executed actions based on the current state information. The Actor network updates the network parameters based on the Critic network scoring evaluation, and continuously updates the iterative policy to find the optimal policy. The important feature of the MADDPG algorithm is that it introduces global observation data and combinations of actions of all other agents as training data when training a single agents Critic network, but receives only local information as input when executing the actions. The framework of the MADDPG algorithm is shown below in Fig. 3:

Specifically, the environment is set to contain N agents and the set of strategies characterized by the Actor network is $\sigma = \{\sigma_1, \sigma_2, \sigma_3 \cdots \sigma_{N-1}, \sigma_N\}$, the set of policy parameters is: $\beta = \{\beta_1, \beta_2, \beta_3 \cdots \beta_{N-1}, \beta_N\}$, then the i-th agent expects the gradient of the reward $J(\sigma_i) = E[R_i]$:

$$\nabla_{\beta_i} J(\sigma_{\beta_i}) = E_{s,a\sim D}[\nabla_{\beta_i} \sigma_i(a_i|s_i) \nabla_{a_i} Q_i^\sigma(s, a_1, a_2 \cdots a_N)|_{a_i = \sigma_{\beta_i}(s_i)}] \quad (3)$$

where s denotes the observed data for all agents, $s = (o_1, o_2, \cdots o_{N-1}, o_N)$, a denotes all intelligent body action sequences, $a = (a_1, a_2, \cdots a_{N-1}, a_N)$. D denotes the empirical playback pool, which records all the agent sample data for training the neural network, which consists of the tuples $(s_t, a_t^1, \cdots, a_t^N, r_t^1, \cdots, r_t^N, s_{t+1})$. $Q_i^\sigma(s, a_1 a_2 \cdots a_N)$ denotes the i-th agent action state Q-value function, the input is the action of all agent $(a_1 a_2 \cdots a_N)$ and the

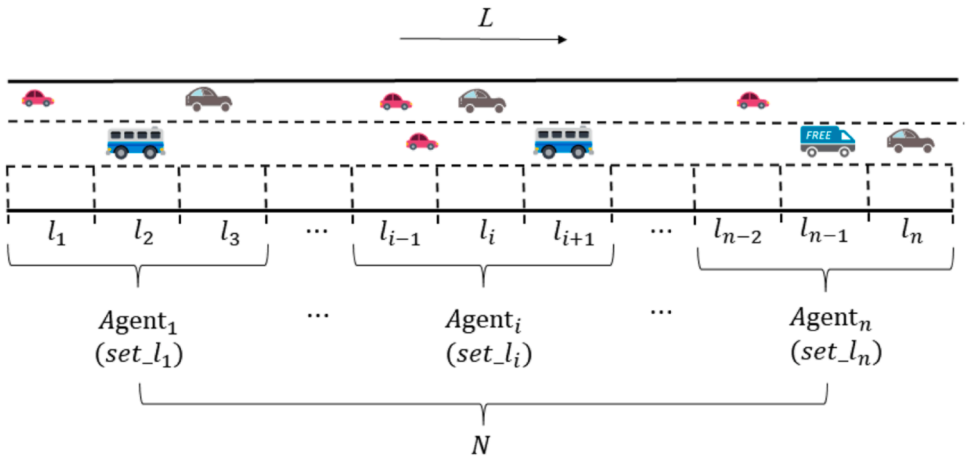


Fig. 4. Agent control section diagram.

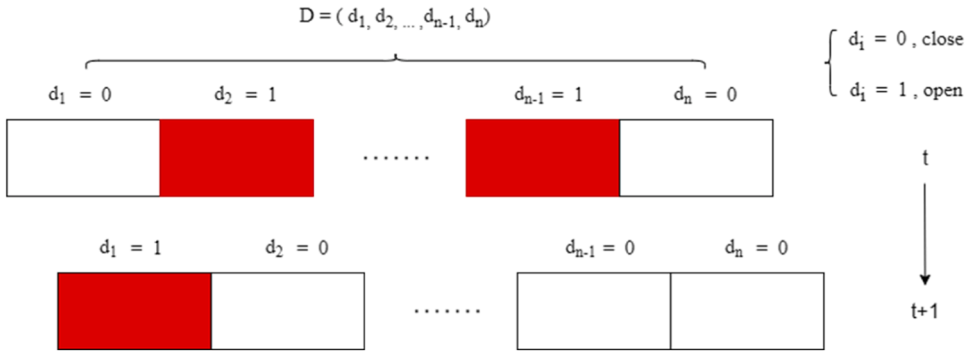


Fig. 5. Action process.

agent observation data s. The output is the action state Q-value, which is the evaluation of the goodness of taking action a in the current state s. The network parameter update formula for the state value function Q_i^{σ} is as follows

$$LOSS(\beta_i) = E_{s_t,a_t,r_t,s_{t+1}}\left[\left(Q_i^{\sigma}(s_t,a_t) - y\right)^2\right] \tag{4}$$

$$y = r_i + \tau Q_i^{\sigma'}(s_{t+1},a_{t+1})\big|_{a_{t+1}^j=\beta_j'(o_j)} \tag{5}$$

where $\beta' = \{\beta_1', \beta_2', \cdots, \beta_N'\}$ denotes the parameters of the target Actor network strategy $\sigma' = \{\sigma_1', \sigma_2', \sigma_3', \cdots, \sigma_{N-1}', \sigma_N'\}$. The target Actor network update method uses a soft update method according to the following equation:

$$\beta_i' = \vartheta * \beta_i + (1-\vartheta) * \beta_i' \tag{6}$$

4.3. STC-MADDPG algorithm

The algorithm interacts with the simulation module and real-time data transmission channel to facilitate the training and testing process. Chapter 5 will provide a detailed description of this process. Specifically, the algorithm retrieves real-time data from the simulation module through a feedback mechanism, generates control strategies, and executes them in the simulation environment while updating network parameters to optimize these strategies. This process concludes when the predefined reward function converges or a specified termination condition is met. At a certain time step during the algorithm's training, it acquires real-time input data from the simulation environment, including environmental and agent states. Each agent processes the input data, generating corresponding actions (open strategy) based on its action network and scoring the generated actions through its evaluation network. By combining the actions of all agents, a joint action (joint open strategy) is formed and evaluated through a spatiotemporal constraint module. Joint strategies that meet the spatiotemporal constraints are fed back to the simulation platform as open strategies, which are then executed to collect reward data and state information for the next time step, thereby constructing a training data pool. Based on this data pool, the algorithm uses the target network to randomly sample data, updating both the action network and the evaluation network. After a certain number of training steps, the parameters of the target network are adjusted through a soft update process. This iterative process continues until the algorithm's training is completed.

Combining the spatio-temporal constraints and the above learning process, the STC-MADDPG algorithm is structured as follows:

Algorithm 1. STC-MADDPG algorithm for shoulder running with N agents

Initialize the parameters of the critic network $Q(s,a;\beta^Q)$, the target critic network $Q'(s,a;\beta^{Q'})$, actor network $\sigma(s;\beta^\sigma)$ and target actor network $\sigma'(s;\beta^{\sigma'})$

For episode from 1 to N **do**

  Reset the simulation environment

  For $t=1$ to N_T **do**

    For each agent$_i$, observe the environment state $o_{environment}$ and the state of agent$_i$: o_i, obtain the state of agent$_i$: $s_i(t)=[o_{environment},\ o_i]$, then the current state $s(t)=[s_1(t),s_2(t),\dots,s_N(t)]$

    For each agent i, select the action $a_i=\sigma_i(s_i(t);\beta^\sigma)+\mathcal{N}_a$, the joint action $a=[a_1,a_2,\dots,a_n]$

    Check the joint action a meets the constraint conditions, if not satisfying, repeatedly select the joint action until it is met

    Perform a simulation and compute the reward $r(t)$ for each agent, observe the next state $s(t+1)=[s_1(t+1),s_2(t+1),\dots,s_N(t+1)]$

    Store the transition $(s(t),a,r(t),s(t+1))$ to data set D

    For agent $i=1$ to N **do**

      Randomly sample a minibatch of S samples $(s^k(t),a^k,r^k(t),s^k(t+1))$ from the data set D

      Set $\begin{cases} Q_i^{target(k)}(t)=r_i^{(k)}(t)+\gamma Q_i^{target(k)}\left(s^{(k)}(t+1),a^{s^{(k)}(t+1)};\beta^{Q'}\right) \\ a^{s^{(k)}(t+1)}=[a_1',a_2',\dots,a_n'],\text{where } a_j'=\sigma_j'(s_j^k(t+1);\beta^{\sigma'}) \end{cases}$

      Update the parameters of the critic network by minimizing loss function:

$$\begin{cases} LOSS_i(\beta^Q)=1/N_s\sum_k(Q_i^{target(k)}(t)-Q_i(s^{(k)}(t),a^{(k)};\beta^Q))^2 \\ \beta^Q=\beta^Q-\eta_Q\nabla_{\beta^Q}LOSS_i(\beta^Q) \end{cases}$$

      Update the parameters of the actor network by using the sampled policy gradient:

$$\nabla J_i(\beta^\sigma)\approx 1/N_s\sum_k\nabla_{a_i}Q_i(s^{(k)}(t),a;\beta^Q)\nabla_{\beta_i^\sigma}\sigma_i(s_i^{(k)}(t);\beta^\sigma)$$

    End for

  End for

  Soft update the target network parameters for each agent every $\mathbb{N}$ training times:

$$\begin{cases} \beta^{Q'}=(1-\vartheta)\beta^Q+\vartheta\beta^{Q'} \\ \beta^{\sigma'}=(1-\vartheta)\beta^\sigma+\vartheta\beta^{\sigma'} \end{cases}$$

End for

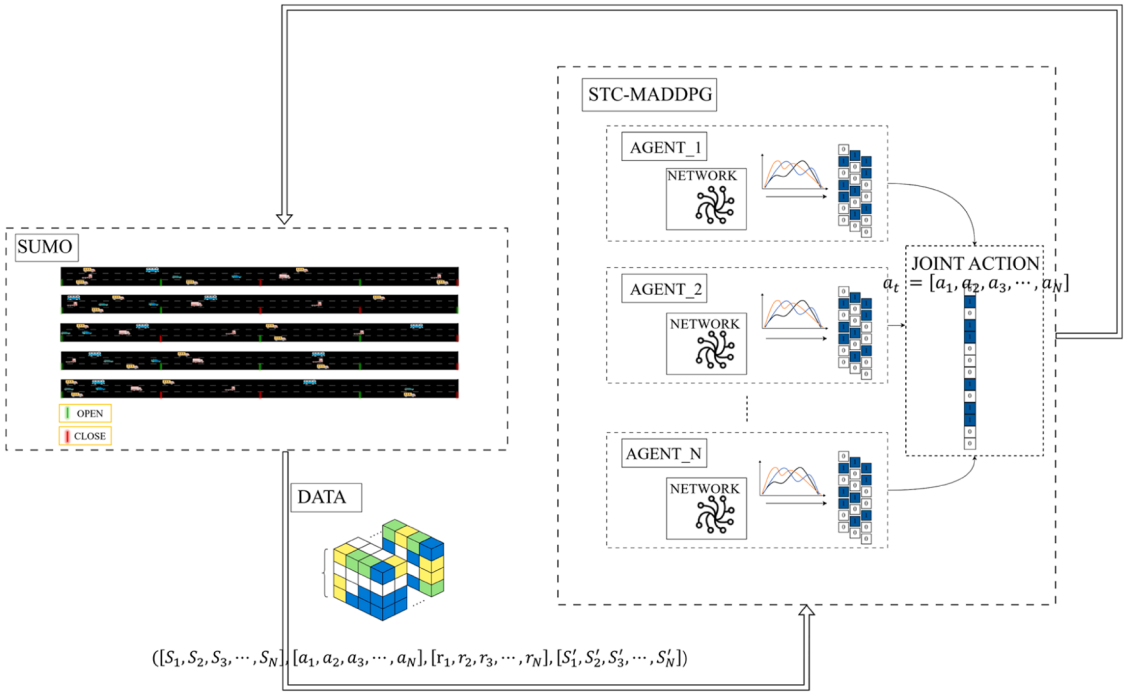


Fig. 6. Sumo simulation framework.

5. Experiment and environment

The following section describes the experimental environment, the experimental environment elements, and the simulation environment. Section 5.1 describes the reinforcement learning agent, the observation space, the action space, and the reward function in the experiment. Section 5.2 describes the experimental sections, the simulation framework, and the experimental data.

5.1. Experimental environment elements

5.1.1. Agent

This paper proposes a methodology where contiguous road sections are combined into a sub-section set, denoted as set_l_i, and treated as individual agents $Agent_i$. The actions of the agent are recorded as a combination of open and closed states of each section in the section set, and thus control the state of the section. By dividing the whole road section into N subsets of road sections, the open and closed states of the whole road section can be controlled by N agents, shown in Fig. 4.

5.1.2. Observe

The observation space S_i of the agent is divided into two parts. One part consists of environmental observation data, $o_{environment} = [cars_{total}, wait\_time_{total}, volume_{total}]$. Among them, $cars_{total}$ indicates the total number of vehicles in the network, which is used to describe the current traffic pressure on the road network; $wait\_time_{total}$ indicates the total waiting time of vehicles on the road network, which is used to describe the current traffic efficiency of the road network; $volume_{total}$ indicates the total volume on the road network, which is used to describe the current traffic flow condition of the road network. A part of the independent observation data from the intelligent body $o_i = [cars_i, wait\_time_i, average\_velocity_i, travel\_time_i, occupancy_i]$, where $cars_i$ denotes the number of all vehicles observed on the road section controlled by the i-th agent, $wait\_time_i$ denotes the waiting time of all cars on the road section controlled by the i-th agent, $average\_velocity_i$ denotes the average speed of vehicles on the i-th agent-controlled road section, $travel\_time_i$ denotes the average travel time on the i-th agent-controlled road section, $occupancy_i$ denotes the average occupancy rate of the i-th agent-controlled road section. The observation data of the i-th agent is $S_i = [o_{environment}, o_i]$, that is, the state information indicating the current state of the road network is $S = [S_1, S_2, S_3, \cdots, S_N]$.

5.1.3. Action

Each agent governs the switching state of a contiguous set of multiple road segments, denoted as set_l_i. Each road segment, denoted as l_i, is regulated by binary decision variables (d_i) to determine control activation. The agent's action space encompasses all possible permutations and combinations of binary control variables within the set of road segments, shown in Fig. 5.

5.1.4. Reward

The objective of this study is to minimize individual vehicle travel time within the network and enhance the efficiency of highway network operations, consequently analyze other impacting factors. In pursuit of these objectives, we consider the combination of both vehicle travel time and traffic flow on the network. The following equation represents the formulated reward function:

$$r_t^i = \frac{1}{\log_{10}(cars_i / volume_i)}, \tag{7}$$

$$r_t = \sum_{i=1}^{n} \varphi_i r_t^i \tag{8}$$

where $cars_i$ denotes the total number of vehicles on the network of the i-th agent control section, $volume_i$ denotes the traffic volume of the i-th agent control section, r_t^i denotes the reward of the i-th agent at the t-th time slice, and r_t denotes the reward of all agents at the t-th time slice.

5.2. Simulation environment

5.2.1. Road Segment introduction

This paper utilizes the urban microsimulation environment, SUMO, to conduct simulations and validate the feasibility and effectiveness of the proposed method. The road section has a total length, denoted as L, of 6.5 km, with each sub-section, represented by l_i, approximately

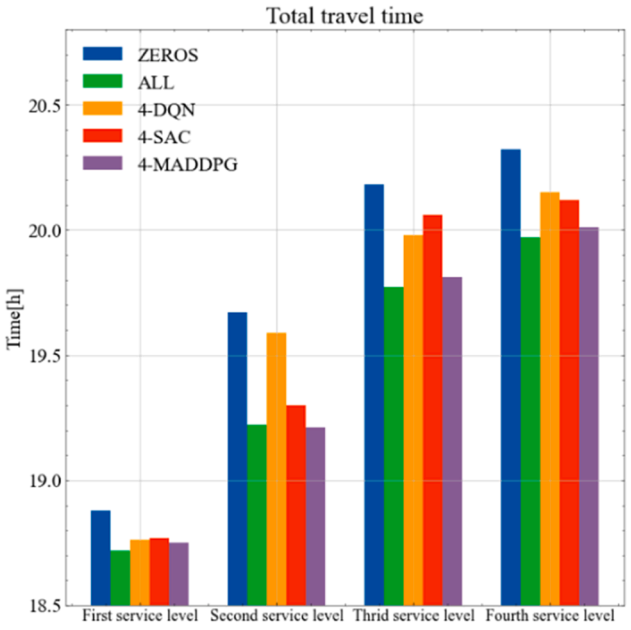
Fig. 7. Training results.

ranging from 300 to 400 m. Consequently, the road section can be divided into 16 sub-sections. This is a section of highway in Jinan, Shandong Province, China, between Ganggou Interchange and Lixingcheng Interchange, which was used as a section for simulation experiments. To achieve optimal results, the time constraint threshold is experimentally adjusted, and Φ^*_{time} is taken as 8. In addition, this paper constrains the number of lane components $\Psi^*_{lane\_quantity}$ in the experiment to the interval [0,7]. It is important to note that the maximum number of lane components can not exceed 7, while the minimum can be 0. Spatial constraints control the degree of dispersion of open hard shoulder subsegments within the same time slice.

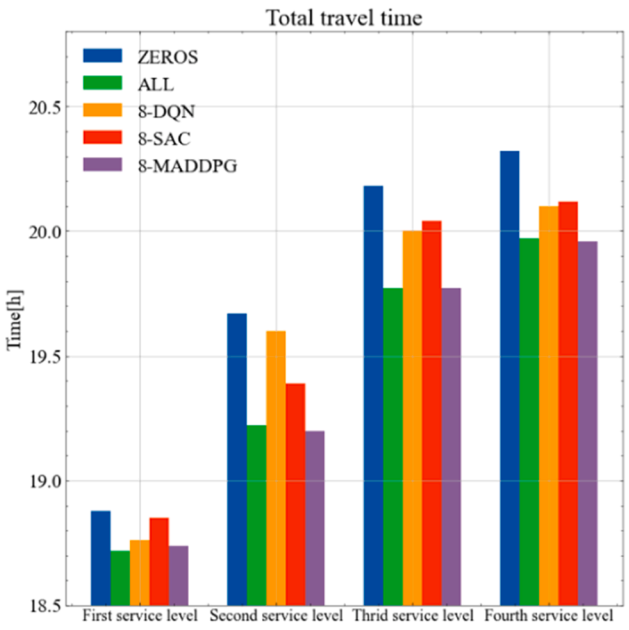
Fig. 6 illustrates the structure used for conducting simulation experiments. The SUMO simulation environment receives joint actions from the individual agent combinations of the STC-MADDPG algorithm, controlling the opening and closing of the hard shoulder through the Traci interface in the Python environment. In real-time, the environment generates traffic flow data, including status and reward data, which is stored in a data buffer pool. At the conclusion of the simulation, each agent updates network parameters by sampling the stored data. After several iterations of this process, the STC-MADDPG algorithm rewards convergence, indicating the end of the experimental training.

5.2.2. *Data introduction*

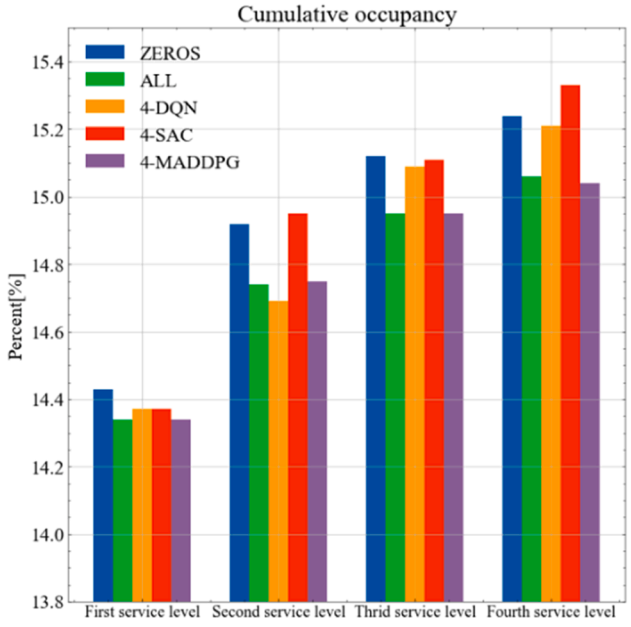
According to the rules for evaluating the service level of Chinese highways, traffic flow conditions on the highway divided into four levels to represent service quality [64], then the corresponding flow data under different service level will be employed as simulation inputs. The total traffic volumes for the four service levels are calculated based on the maximum service traffic volume under different lane speed limits. These volumes are determined to be 1400 (veh/h), 3000 (veh/h), 3750 (veh/h), and 4300 (veh/h) for the respective service levels. To create a more realistic representation of the vehicle composition in the highway traffic flow, the SUMO simulation data includes three types of vehicles: Trailers, Trucks, and Private cars. Trailers, representing large-sized vehicles, including big trucks and other vehicles, have a length exceeding 10 m. Trucks, representing medium-sized vehicles, have a length of



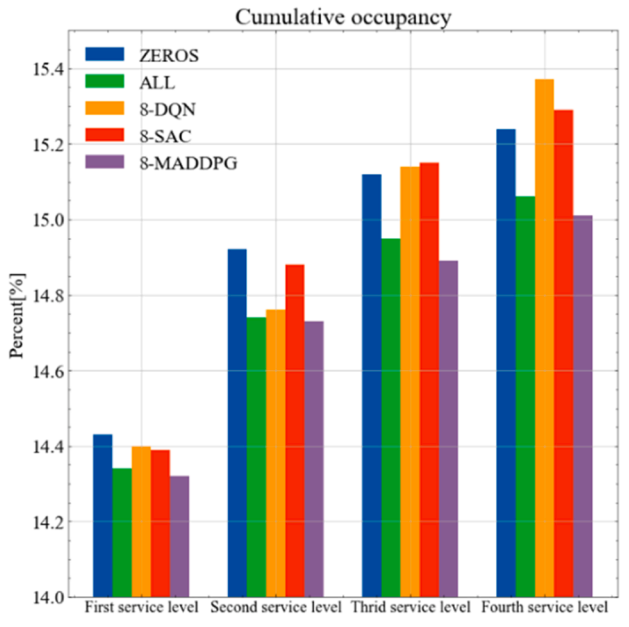
(a) total travel time results for 4 agents



(b) total travel time results for 8 agents



(c) cumulative occupancy results for 4 agents



(d) cumulative occupancy results for 8 agents

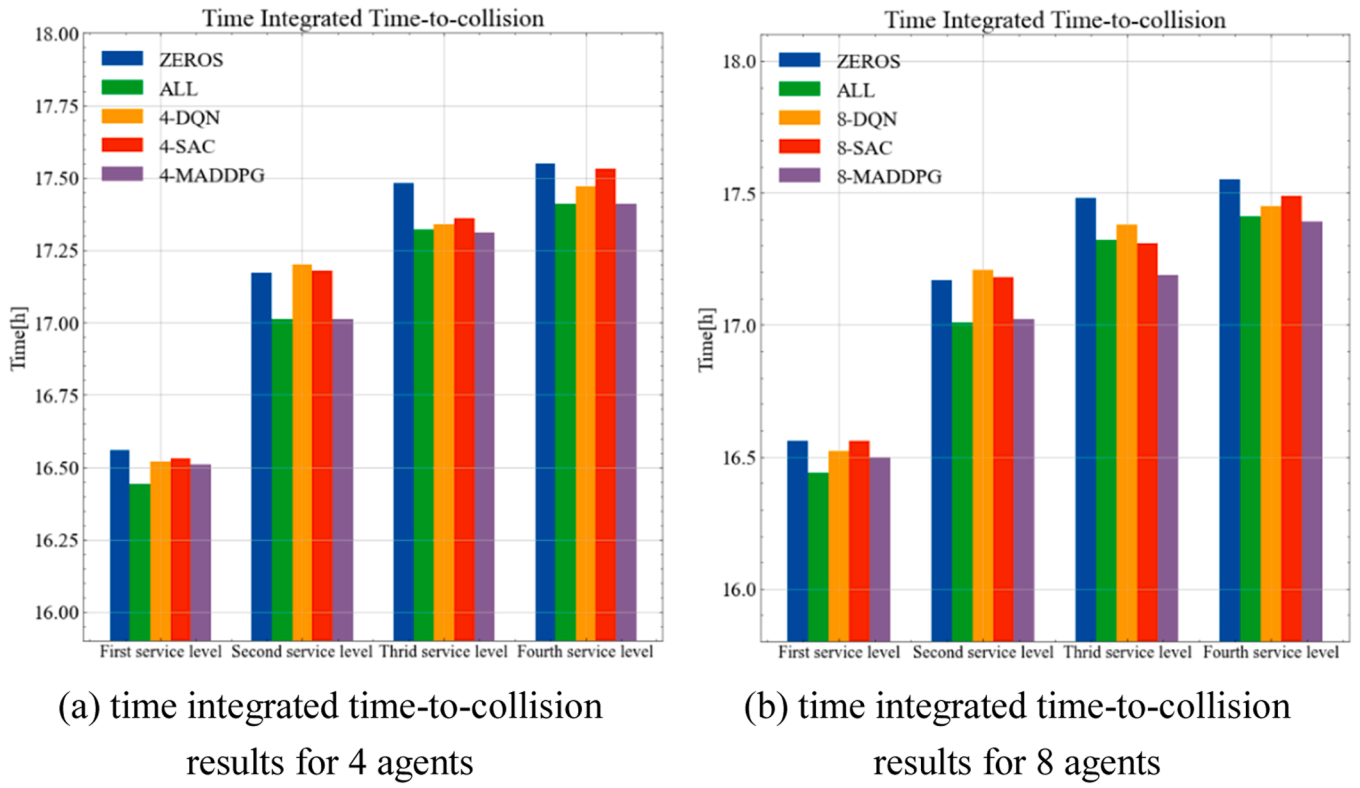
Fig. 8. Efficiency analysis under different service levels.

about 7–10 m. Private cars, representing small vehicles, have a length of approximately 5–6 m. The proportions of these vehicle types are set to 15 %, 15 %, and 70 % respectively, based on the data distribution of different vehicle types observed in the highway in Jinan City, Shandong Province.

6. Analysis of results and discussion

In this section, we assess the efficacy of the proposed STC-MADDPG algorithm through simulation experiments and evaluate its ability to address the HSR control problem on highways. With the widespread adoption of highway traffic flow sensing technology, vehicles can readily acquire real-time traffic flow information on highway sections, thereby facilitating the dynamic opening of the hard shoulder. In this study, we employ SUMO simulation to replicate functions such as roadway information collection, information communication, and information exchange among vehicles. Each $Agent_i$ governs a sub-section set, set_l_i, and determines whether to activate the hard shoulder in this particular sub-section set. To explore the impact of the number of subsections on the control effectiveness of different agents, we consider two scenarios: 2 and 4 controlled subsections per agent. If an agent controls 2 subsections, the total number of agents is 8; if an agent controls 4 subsections, the total number of agents is 4. Training tests are conducted separately for each agent count.

Given the varying traffic flow conditions on highways, the requirements and approaches for hard shoulder activation can differ



(a) time integrated time-to-collision results for 4 agents

(b) time integrated time-to-collision results for 8 agents

Fig. 9. Safety analysis under different service levels.

significantly. This paper categorizes the highway traffic flow status based on service levels, assigns different flow magnitudes to different traffic flow scenarios, and employs the STC-MADDPG algorithm to investigate the optimization problem of dynamic hard shoulder running. This section comprises three parts: training results, evaluation of individual indices, and an extended study.

6.1. Training results

Highway service level characterizes the quality of traffic service provided to drivers on the highway. Highway service levels are categorized as Class I, Class II, Class III, and Class IV. Different levels correspond to highway traffic conditions evaluated by traffic speed and traffic volume. The optimal solution is computed for four classes, each further divided into two cases involving four and eight agents. The reward convergence plots of the training process using the STC-MADDPG algorithm are depicted in Fig. 7, where a, b, c, and d correspond to the first, second, third, and fourth service levels, respectively. In this study, the control algorithms deep Q learning (DQN) [65] and soft actor-critic (SAC) [66] are employed alongside the STC-MADDPG algorithm to validate its effectiveness. As observed in the figure, the blue curve representing the STC-MADDPG algorithm consistently achieves the highest reward across all four service levels. When comparing the four service levels, it becomes evident that the algorithm exhibits slower convergence and greater reward variability as the service level gradually increases. This observation suggests that in practical control scenarios, as traffic flow intensifies, the control effectiveness may become less stable once the highway traffic density reaches a certain threshold. Notably, at service level 4, the utilization of 8 agents for control yields slightly higher rewards compared to other control approaches, implying that a more subtle control strategy may be more effective in high-traffic scenarios.

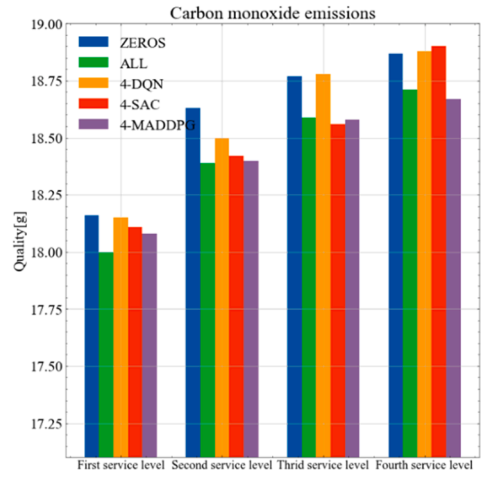
6.2. Evaluation of learning performance

To assess and validate the performance of the approach, this study employs two fundamental control groups: "zeros," representing the scenario where none are open, and "all," signifying the case where all are open. Additionally, the "DQN" and "SAC" schemes serve as the optimized control groups to verify the effectiveness of the STC-MADDPG algorithm. The proposed method is evaluated and validated based on efficiency, safety, and emissions across four service level conditions. These three aspects are compared and analyzed using six indicators: total travel time, cumulative occupancy, TIT (Time Integrated Time-to-collision) [67], CO, CO_2, and NO. To accommodate the large values of these indicators, logarithms are taken to obtain relative values, as illustrated in the following figures.

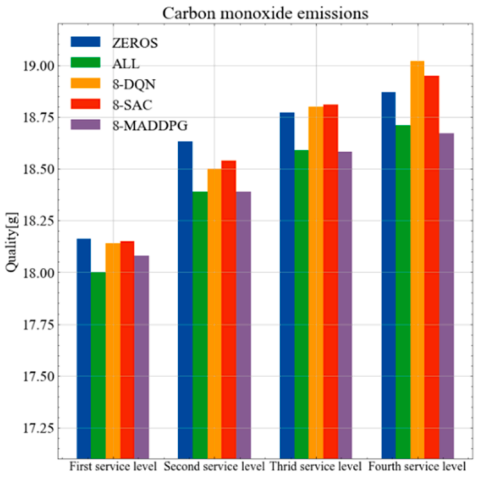
6.2.1. Efficiency

The Fig. 8 presents a comparative analysis of total travel time and cumulative occupancy for different strategies across four service levels. It is evident that the STC-MADDPG algorithm control strategy effectively reduces the total travel time for both vehicle categories compared to the "no open" strategy across all service levels. The "8-MADDPG algorithm" achieved reductions in total travel time of 14.5 %, 37.5 %, 33.5 %, and 30.4 % for the four service levels, respectively, when compared to the "no open" strategy.

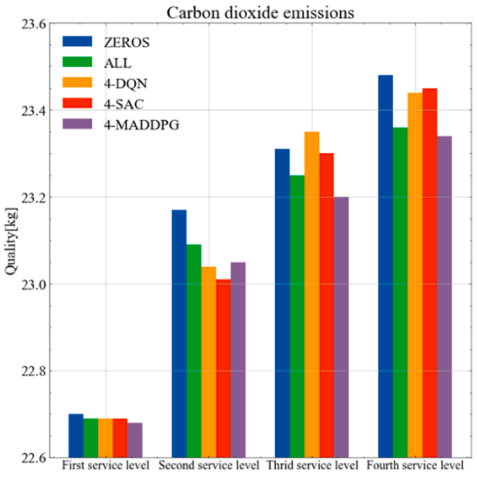
At service level 1, the impact of opening or closing the hard shoulder on improving vehicle driving efficiency is relatively minimal since the highway experiences smooth traffic flow. However, under service level 2 and 3 conditions, where traffic volume is higher, opening the hard shoulder increases capacity and enhances vehicle movement. Notably, the "4-MADDPG" and "8-MADDPG" algorithms yield superior outcomes compared to the "4-DQN," "8-DQN," "4-SAC," and "8-SAC" algorithms, resulting in greater reductions in total travel time. From Section 6.1 we conclude that the "4-MADDPG" and "8-MADDPG" converge to higher reward values compared to "4-DQN," "8-DQN," "4-SAC," and "8-SAC" algorithm for the same number of training rounds. This indicates that



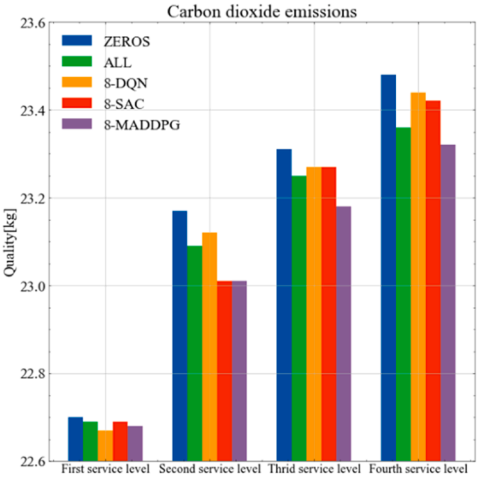
(a) carbon monoxide emissions results for 4 agents



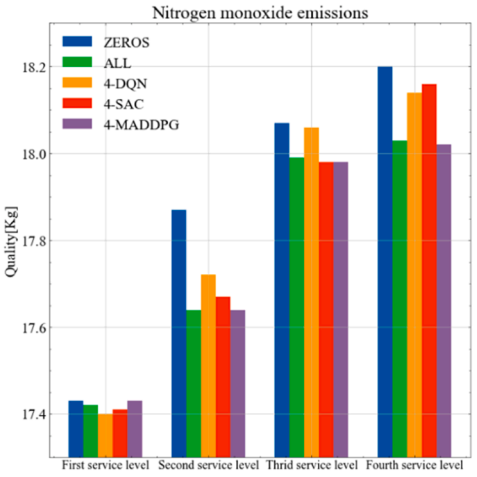
(b) carbon monoxide emissions results for 8 agents



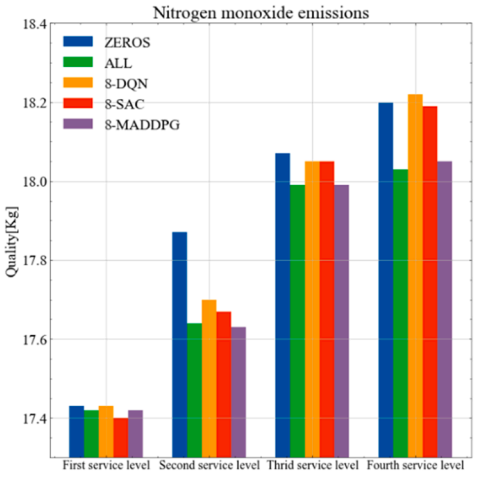
(c) carbon dioxide emissions results for 4 agents



(d) carbon dioxide emissions results for 8 agents



(e) nitrogen monoxide emissions results for 4 agents



(f) nitrogen monoxide emissions results for 8 agents

Fig. 10. Emission analysis under different service levels.

Fig. 11. Test section.

the strategy of STC-MADDPG enables vehicles to leave the control area in a faster time. This will result in all vehicles spending less time on the control road network, which in turn will allow the corresponding strategy of STC-MADDPG to have less total travel time.

The MADDPG algorithm assumes that each agent has its own independent critic network and actor network, and assumes that each agent has its own independent reward function, so that the MADDPG algorithm can solve the multi-agent problem in the collaborative, competitive, and hybrid environments at the same time. Fusing the time and space constraints, the STC-MADDPG algorithm is proposed, which learns the optimal policy to give the optimal action using only local information. In contrast, DQN, SAC algorithms all need to obtain complete global information data in order to generate the corresponding control policies. In addition, when the control section becomes larger, the action space dimension of DQN and SAC algorithms will increase exponentially, and then DQN and SAC algorithms cannot be used. The STC-MADDPG algorithm segments multiple control sections, which can not only control the local area effectively, but also cooperate and win-win situation among multiple agents, which makes the overall effect better.

In the case of the fourth service level, where the traffic flow-to-capacity ratio reaches 0.88–1.0, congested sections become more prevalent. The "4-MADDPG algorithm" and "8-MADDPG algorithm" control strategies achieve reductions in total travel time by 26.8 % and 30.4 %, respectively. Cumulative occupancy rate calculates the sum of highway occupancy rates over a specified time period, providing insights into the overall operational balance. The cumulative occupancy index further demonstrates that the STC-MADDPG algorithm strategy improves highway traffic conditions compared to the "no open" and "all open" strategies.

The aforementioned comparison highlights the effectiveness of optimized HSR control in enhancing highway traffic efficiency. However, the control effects of different agent control methods vary significantly across the four service levels, with the 8-agent method exhibiting

Table 1
Traffic flow expansion under service level 4.

Service level	Ratio	Trailers (veh/h)	Trucks (veh/h)	Private cars (veh/h)	Total flow (veh/h)
4	0.8	516	516	2408	3440
4	1.0	645	645	3010	4300
4	1.2	774	774	3612	5160

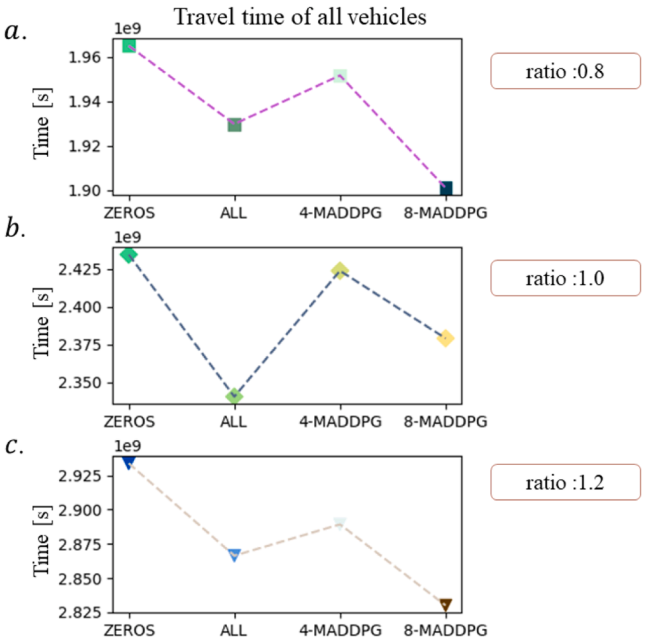


Fig. 12. Total travel time under different traffic flow expansion ratios.

superior performance.

6.2.2. *Safety*

The Time Integrated Time-to-Collision (TIT) index is an indicator that encompasses the impact of risky driving behavior and is calculated by integrating the duration a vehicle spends below the TTC threshold. In this study, the evaluation of the impact of the STC-MADDPG algorithm for highway HSR control on highway safety is approached from the perspective of the TIT index. The Fig. 9 illustrates the comparison results of the four-agent and eight-agent algorithms.

At service level 1, there is minimal variation between the different control strategies. This can be attributed to the smooth movement of vehicles on the highway, resulting in a higher probability of vehicles maintaining a safe distance from each other and consequently higher road safety levels. At the service level 2, both variations of the STC-MADDPG algorithm outperform the "no open," "4-DQN," "8-DQN," "4-SAC," and "8-SAC" algorithms in searching for superior policies. The control strategy output by STC-MADDPG allows vehicles to stay shorter and spend less time in a hazardous state when passing through the control area. As a result, the vehicle TIT value is lower compared to the DQN, SAC algorithm output control strategy. And the effect of the trained model will be different under different service levels.

At the service level 3, adopting the STC-MADDPG algorithm's open strategy led to reductions in TIT values of 18.7 % and 31.4 % for the four-agent and eight-agent scenarios, respectively, compared to the "no open" strategy. This improvement was highest across the four service levels. Notably, as the service level progresses from one to four, the optimization effect on TIT values initially increases and then decreases, with reductions of 4.1 % and 5.8 % transforming to 18.7 % and 31.4 %, and finally keeping at 15.4 % and 17.3 %, respectively. This observation suggests that enhancing highway safety can be more challenging under

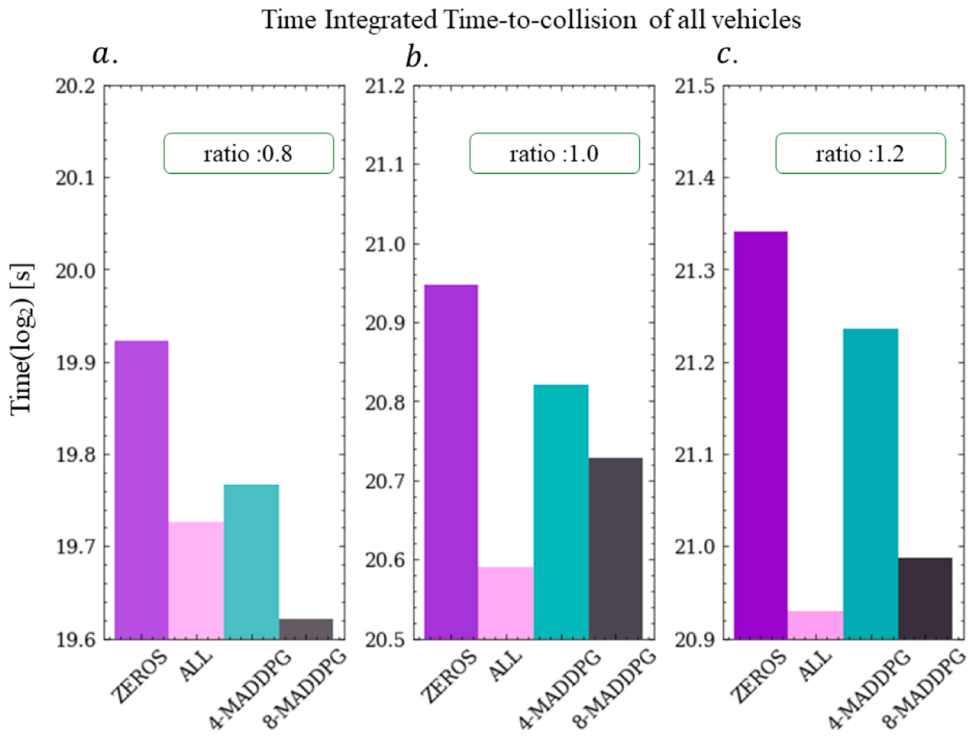


Fig. 13. TIT under different traffic flow expansion ratios.

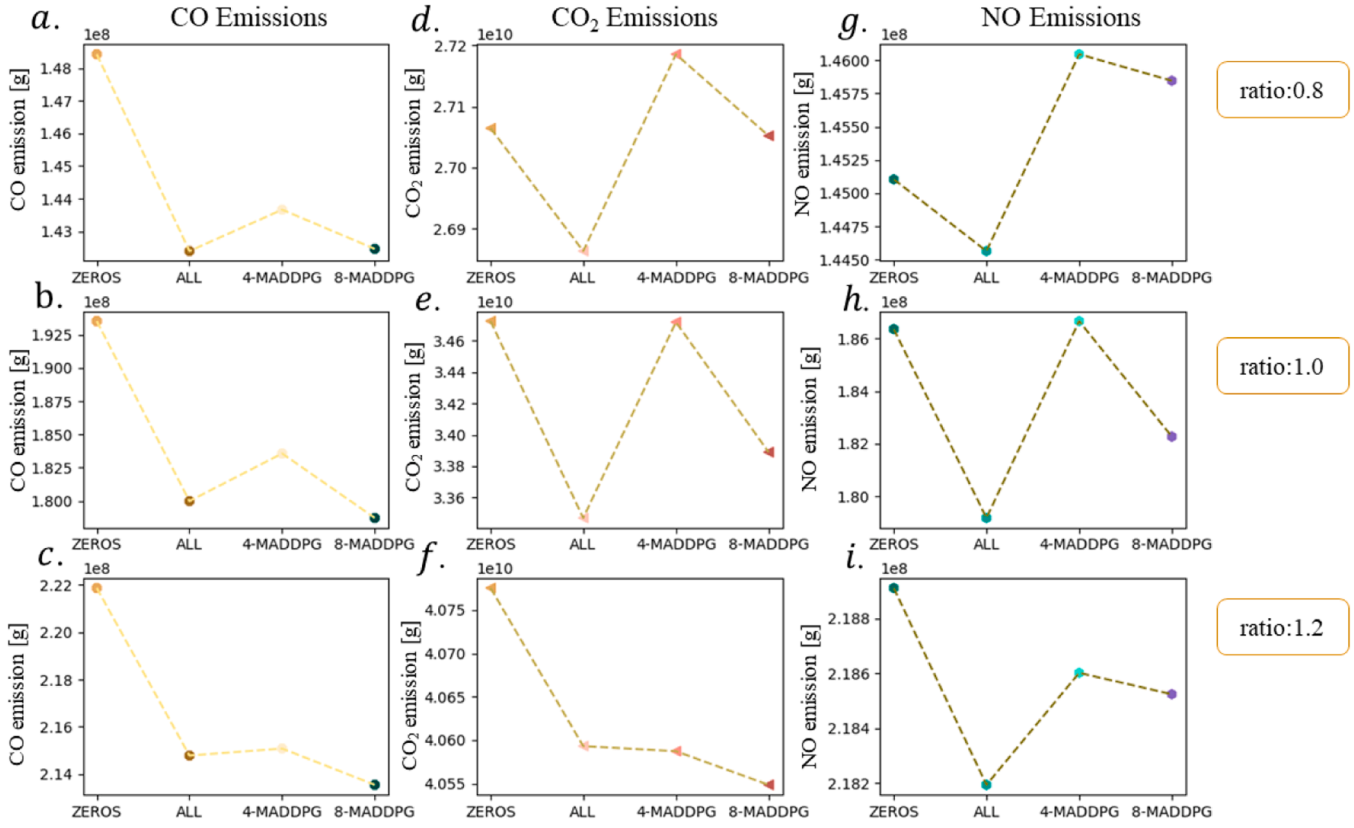


Fig. 14. Exhaust emission index display.

both lower and higher traffic conditions.

6.2.3. *Emissions*

Vehicle exhaust a significant contributor to atmospheric pollutants, encompassing various gases such as carbon dioxide, carbon monoxide, nitrous oxide, and nitrogen dioxide. Carbon dioxide, in particular, acts as a greenhouse gas, leading to global warming and influencing regional climate change. Nitric oxide rapidly oxidizes in the atmosphere to form

nitrogen dioxide, a prominent pollutant that impacts air quality. The Fig. 10 illustrates the successful reduction of highway vehicle emissions through the implementation of the STC-MADDPG algorithm for HSR control.

At the service level 1, consistent with the results for the efficiency and security metrics, the results do not vary significantly across strategies. The STC-MADDPG algorithm optimized for HSR control demonstrates significant reductions in CO_2 emissions of 12.4 % and 16.4 %, respectively, compared to the no-open strategy at service level 2. Remarkably, this optimization yields the most substantial benefits among the four service levels. Likewise, for CO and NO_x emissions, reductions occur across all four service levels. Notably, at service level 4, the optimized control strategy employing either the 4-agent or 8-agent mode results in reductions of carbon monoxide, carbon dioxide, and nitrous oxide emissions by 21.7 %, 14.3 %, and 19.9 %, respectively. The reduction in vehicle exhaust emissions is consistent across both modes, highlighting the efficacy and feasibility of utilizing the STC-MADDPG algorithm for optimized HSR control from an emissions standpoint.

6.3. Further discussion

At the service level 4, we observed that training with 8 agents resulted in faster convergence and higher rewards compared to training with 4 agents. As the total length of the controlled sections remains constant for both the 4- and 8-agent approaches, controlling more subsections per agent is required in the 4-agent approach. Thus, as the number of controls per agent is smaller in the 8-agent approach, it is more effective for higher traffic volume. To validate these findings, this study expanded the length of the control section and conducted comparative experiments by increasing the maximum traffic values by 0.8, 1.0, and 1.2 times for the service level 4. The selected section is illustrated in Jinan city, a length of 16.3 km (shown in Fig. 11), and the traffic flow is presented in Table 1. Three different ratios: 0.8, 1.0, and 1.2 times the total flow at service level 4 (4300 veh/h) were employed as traffic flow data for the simulation.

Under the aforementioned three flow conditions, a comparative analysis was conducted among four strategies: "no open," "full open," "4-MADDPG Algorithm", and "8-MADDPG Algorithm." The obtained results provide comparative insights into the effectiveness of these strategies in terms of efficiency, safety, and emissions. As illustrated in Fig. 12, the results show the comparison of total travel time for the three traffic flow expansion rates. Notably, the "8-MADDPG algorithm" outperforms the "no open" strategy and other approaches. The total travel time reduction achieved by the "8-MADDPG algorithm" compared to the "no open" strategy is 17,795.68 hours, 15,397.3 hours, and 28,806.15 hours across three cases.

The algorithmic search strategy effectively enhances highway efficiency by reducing vehicle travel time. Moreover, when comparing the "8-MADDPG algorithm" with 8 agents to the "4-MADDPG algorithm" with 4 agents, the former demonstrates superior control effectiveness, resulting in a reduction in total travel time of 14,076.32 hours, 12,415.51 hours, and 16,430.69 hours for the three traffic conditions, respectively. These findings indicate that increasing control granularity by adjusting the number of agents under high traffic conditions can enhance efficiency of traffic flow on highway.

The Fig. 13 illustrates the comparison of TIT values for three different traffic ratios. From a safety perspective , the impact of different strategies was evaluated. The "8-MADDPG algorithm" demonstrates significant reductions in the TIT value compared to the "no open" strategy shown in the figure, with reductions of 23.3 %, 16.3 %, and 27.8 % observed under the three traffic conditions, respectively. TIT represents the cumulative time that vehicles spend in critical situations where collision risk is high while traveling on the highway. A decrease in the TIT value indicates reduced time spent in critical conditions, thus contributing to improved highway safety. Notably, at expansion rate of 1.2, the "8-MADDPG algorithm" achieved a reduction of 108.5 hours compared to the "4-MADDPG algorithm." This reduction signifies a substantial enhancement in driving safety, as it represents a decrease in the time during which vehicles are exposed to potential collision risks. The "8-MADDPG algorithm" consistently outperformed the "4-MADDPG algorithm" across all three traffic conditions, demonstrating its superior effectiveness in enhancing highway safety. These findings emphasize the importance of meticulous control measures, particularly in scenarios characterized by high traffic volumes.

The Fig. 14 presents the comparison of CO, CO_2, and NO emissions for all vehicles across three traffic flow expansion ratios. The analysis of CO emissions in the Fig. 14 (a, b, c) reveals the superior effectiveness of the "8-MADDPG algorithm" strategy, resulting in significantly lower CO emissions compared to the other three strategies. It is worth noting that CO_2, as a primary greenhouse gas, plays a important role in climate change, and the reduction of CO_2 emissions can contribute to mitigating the greenhouse effect and alleviating the adverse impacts of environmental degradation. Examining the Fig. 14 (d, e, f), which focuses on CO_2 emissions, the "8-MADDPG algorithm" strategy demonstrates its effect in curbing CO_2 emissions when compared to the baseline "no open" strategy. Specifically, it achieves reductions of 12,351.8 kg, 834,678 kg, and 226,667.5 kg under the respective flow conditions. Furthermore, the "8-MADDPG algorithm" strategy proves effective in minimizing NO emissions, with a substantial reduction of 4097.0 kg at a flow rate of 1.0 in comparison to the "no open" strategy. These reductions provide considerable significance. When assessing the cumulative emissions of the three gases, the "8-MADDPG" strategy yields substantial reductions of 134,729.5 kg, 837,776.2 kg, and 40,344.2 kg when compared to the "4-MADDPG" strategy across all three flow rates. These findings underscore the superior efficacy of the "8-MADDPG algorithm" in mitigating highway vehicle emissions and addressing environmental pollution.

7. Conclusion

Highway congestion causes queuing delays, higher pollution emissions, and significant disruption to highway operations. This paper examines how to improve traffic efficiency and reduce pollution emissions by optimizing the control strategy of the hard shoulder based on the multi-agent reinforcement learning algorithm. Firstly, we propose the multi-agent reinforcement learning algorithm STC-MADDPG based on hard shoulder spatio-temporal constraints. Secondly, using the SUMO simulation environment, we verify the effectiveness of the proposed method on a 6.4 km long section. Experimental results demonstrate the capacity of the proposed method to improve the efficiency of highway traffic, reduce the danger of traffic, and simultaneously reduce vehicle exhaust emissions. Under the four service level conditions, the proposed method reduced total vehicle travel time by up to 37.5 %, TIT values by up to 34.1 %, and emissions of carbon monoxide, carbon dioxide, and nitric oxide by up to 28.0 %, 17.1 %, and 27.2 %, respectively. The comparison of the three aspects of efficiency, safety, and environmental pollution shows that the effective control of the hard shoulder control problem by the STC-MADDPG algorithm can improve highway traffic conditions and provide manifold benefits. Finally, an analysis of the results of the extended study with 4 and 8 agents at longer control sections at service level 4 suggests that a smaller number of control sections per agent is more effective during times of heavy congestion.

The current research on hard shoulder operation strategies in this paper has limitations in its impact on highway traffic flow. Combining various active traffic control methods, such as variable speed limits, ramp metering, and queue warnings, may lead to better results. Secondly, the traffic environment data used in this research does not take into account special factors such as weather and holidays, which will be considered in future studies. To address these limitations, future in-depth studies could investigate the combination of other active traffic control techniques with hard shoulder running to further improve traffic

flow simulation environments and explore alternative indicators for highway evaluation.

CRediT authorship contribution statement

Lipeng Hu: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jinjun Tang:** Writing – review & editing, Methodology, Conceptualization. **Guoqing Zou:** Data curation. **Zhitao Li:** Writing – review & editing, Data curation. **Jie Zeng:** Data curation. **Mingyang Li:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded in part by the National Natural Science Foundation of China (No. 52172310), Science Research Foundation of Hunan Provincial Department of Education (No. 22B0010).

References

[1] J. Geistefeldt, Hard shoulder running, in: R. Vickerman (Ed.), International Encyclopedia of Transportation, Elsevier, 2021, pp. 41–44.
[2] R. Li, Z. Ye, B. Li, Simulation of hard shoulder running combined with queue warning during traffic accident with CTM model, IEEE Trans. Intell. Transp. Syst. 11 (9) (2017) 553–560.
[3] P. Metaxatos, P. Thakuriah, Planning for bus-on-shoulders operations in northeastern Illinois: a survey of stakeholders, Transp. Res. Rec. 2111 (1) (2009) 10–17.
[4] Use of Freeway Shoulders for Travel — Guide for Planning, Evaluating, and Designing Part-time Shoulder Use as A Traffic Management Strategy, ⟨https://rosap.ntl.bts.gov/view/dot/42328/dot_42328_DS1.pdf⟩, 2016.
[5] S. Coffey, S. Park, Impact of part-time shoulder use on safety through the highway safety manual, International Conference on Transportation and Development 2018: Connected and Autonomous Vehicles and Transportation Safety, 2018, 180-187.
[6] M. Rohloff, Re-use of herd-shoulders at federal motorways, 2nd International Symposium on Highway Geometric Design, 2000, p. 254-266, Mainz, Germany.
[7] J. Geistefeldt, Operational reliability of hard shoulder running on freeways, Transp. Res. Rec.: J. Transp. Res. Board (2024).
[8] Jenior, P., Bonneson, J.A., Zhao, L., Kittelson, W., Donnell, E.T., & Gayah, V.V. Safety Performance of Part-time Shoulder Use on Freeways, Volume 1: Informational Guide and Safety Evaluation Guidelines, 2021.
[9] G. Kellermann, Experience of using the hard shoulder to improve traffic flows, Traffic Eng. Control 41 (2000) 10.
[10] J. Geistefeldt, Operational experience with temporary hard shoulder running in Germany, Transp. Res. Rec. 2278 (1) (2012).
[11] B. Sultan, R. Meekums, J. Ogawa, S. Self, P. Unwin, M42 aCtive Traffic Management Pilot – Initial Results from Hard Shoulder Running Under 60mph Speed Limit.
[12] M. Wilson, Hard shoulder running eases motorway traffic jams, Highways 78 (1) (2009) 12–13.
[13] S. Cohen, M. Aron, R. Seidowsky, Assessment of a dynamic managed lanes operation, 12 th WCTR, 2010, Lisbon.
[14] Z. Deng, Z. Luo, N. Hockaday, A. Farid, A. PandeEvaluation of left shoulder as part-time travel lane design alternatives and transportation management center staff training module development, 2023.
[15] J. Yao, Y. Qian, Z. Feng, J. Zhang, H. Zhang, T. Chen, S. Meng, Hidden markov model-based dynamic hard shoulders running strategy in hybrid network environments, Appl. Sci. (2024).
[16] W. Lu, Z. Yi, Y. Gu, Y. Rui, B. Ran, TD3LVSL: a lane-level variable speed limit approach based on twin delayed deep deterministic policy gradient in a connected automated vehicle environment, Transp. Res. Part C Emerg. Technol. 153 (2023) 104221.
[17] J. Geistefeldt, Operational experience with temporary hard shoulder running in Germany, Transp. Res. Rec. 2278 (1) (2012) 67–73.
[18] Y. Li, A.H.F. Chow, D.L. Cassel, Optimal control of motorways by ramp metering, variable speed limits, and hard-shoulder running, Transp. Res. Rec. 2470 (1) (2014) 122–130.
[19] S. Coffey, S. Park, Operational evaluation of part-time shoulder use for interstate 476 in the state of Pennsylvania, Adv. Civ. Eng. 2018 (2018) 1724646.
[20] S.G. Farrag, F. Outay, A. Yasar, M.Y. El-Hansali. Evaluating Active Traffic Management (ATM) Strategies under Non-recurring Congestion: Simulation-based with Benefit Cost Analysis Case Study. (2020).
[21] Z. Zhigang, P. Rui, Q. Jiangang, Analysis of traffic characteristics of dynamic open hard shoulder road based on SUMO, 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), Shijiazhuang, China, 2022, 100-104.
[22] K. Arora, L. Kattan, Operational and safety impacts of integrated variable speed limit with dynamic hard shoulder running, J. Intell. Transp. Syst. (2022).
[23] D. Li, J. Lasenby, Mitigating urban motorway congestion and emissions via active traffic management, Res. Transp. Bus. 48 (2023) 100789.
[24] T. Hasan, M.A. Abdel-Aty, Short-term safety performance functions by random parameters negative binomial-lindley model for part-time shoulder use, Accid. Anal. Prev. 199 (2024) 107498.
[25] M. Aron, S. Cohen, R. Seidowsky, Two French hard-shoulder running operations: some comments on effectiveness and safety, 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 2010, 230-236.
[26] J. Kononov, S. Hersey, D. Reeves, B.K. Allery, Relationship between freeway flow parameters and safety and its implications for hard shoulder running, Transp. Res. Rec. 2280 (1) (2012) 10–17.
[27] J. Ma, J. Hu, D.K. Hale, J. Bared, Dynamic hard shoulder running for traffic incident management, Transp. Res. Rec. 2554 (1) (2016) 120–128.
[28] J. Choi, R. Tay, S. Kim, S. Jeong, J. Kim, T.-Y. Heo, Safety effects of freeway hard shoulder running, Appl. Sci. 9 (2019) 3614.
[29] H. Waleczek, J. Geistefeldt, Long-Term safety analysis of hard shoulder running on freeways in Germany, Transp. Res. Rec. 2675 (8) (2021) 345–354.
[30] M. Aron, R. Seidowsky, S. Cohen, Safety impact of using the hard shoulder during congested traffic. The case of a managed lane operation on a French urban motorway, Transp. Res C. -Emer. 28 (2013) 168–180.
[31] P. Chasea, E. Avineri, Maximizing motorway capacity through hard shoulder running: UK perspective, Open Transp. J. 2 (1) (2008) 7–18.
[32] K. Lemke, Hard Shoulder Running as a short-term measure to reduce congestion, 4th International Symposium on Highway Geometric Design, 2010, Valencia, Spain.
[33] H. Zeng, S.D. Schrock, Estimation of safety effectiveness of composite shoulders on rural two-lane highways, Transp. Res. Rec. 2279 (1) (2012) 99–107.
[34] R. Sharma, M.O. Faruk, A. El-Urfali, Operational and safety impact analysis of implementing emergency shoulder use (ESU) for hurricane evacuation, Transp. Res. Rec. 2674 (2020) 282–293.
[35] R. Li, Z. Ye, B. Li, Optimal control and simulation of hard shoulder running on highways, J. Syst. Simul. 30 (3) (2018) 1036–1045.
[36] W. Zhou, M. Yang, M. Lee, L. Zhang, Q-learning-based coordinated variable speed limit and hard shoulder running control strategy to reduce travel time at freeway corridor, Transp. Res. Rec. 2674 (11) (2020) 915–925.
[37] F.F. Hussein, B. Naik, G.A. Süer, Development of hybrid hard shoulder running operation system for active traffic management, Int. Conf. Transp. Dev. (2020).
[38] F. Yang, F. Wang, F. Ding, H. Tan, B. Ran, Identify optimal traffic condition and speed limit for hard shoulder running strategy, Sustainability 13 (2021) 1822.
[39] D. Chen, Mohammad R. Hajidavalloo, L. Zhaojian, et al., Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic, IEEE Trans. Intell. Transp. Syst. 24 (11) (2023) 11623–11638.
[40] X. Fang, T. Péter, T. Tettamanti, Variable speed limit control for the motorway–urban merging bottlenecks using multi-agent reinforcement learning, Sustainability 15 (14) (2023) 11464.
[41] K. Kušić, I. Dusparic, M. Guériau, M. Gregurić and E. Ivanjko Extended variable speed limit control using multi-agent reinforcement learning, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece (2020)1-8.
[42] M. Kolat, B. Kővári, T. Bécsi, et al., Multi-agent reinforcement learning for traffic signal control: a cooperative approach, Sustainability 15 (2023) 3479.
[43] L. Xiaoming, Y. Dong, D.M. Frangopol, Sustainable life-cycle maintenance policymaking for network-level deteriorating bridges with a convolutional autoencoder–structured reinforcement learning agent, J. Bridge Eng. 29 (9) (2023) 04023063.
[44] L. Lai, Y. Dong, C.P. Andriotis, A. Wang, X. Lei, Synergetic-informed deep reinforcement learning for sustainable management of transportation networks with large action spaces, Autom. Constr. 160 (2024) 105302.
[45] C. Ying, A.H.F. Chow, H.T.M. Nguyen, K. Chin, Multi-agent deep reinforcement learning for adaptive coordinated metro service operations with flexible train composition, Transp. Res. Part B: Methodol. 161 (2022) 36–59.
[46] A.E. Sallab, M. Abdou, E. Perot, S. Yogamani, Deep reinforcement learning framework for Autonomous Driving, Electron. Imaging 29 (19) (2017) 70–79.
[47] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[48] Darius Drungilas, Mindaugas Kurmis, Audrius Senulis, Zydrunas Lukosius, Arunas Andziulis, Jolanta Januteniene, Marijonas Bogdevicius, Valdas Jankunas, Miroslav Voznak, Deep reinforcement learning based optimization of automated guided vehicle time and energy consumption in a container terminal, Alex. Eng. J. 67 (2023) 397–407.
[49] Mohammed H. Alabdullah, Mohammad A. Abido, Microgrid energy management using deep Q-network reinforcement learning, Alexandria Engineering Journal, 61 (11)(202) 9069-9078.
[50] Xiaoqing Huang, Dongliang Zhang, XiaoSong Zhang, Energy management of intelligent building based on deep reinforced learning, Alex. Eng. J. 60 (1) (2021) 1509–1517.
[51] V. Mnih, K. Kavukcuoglu, D. Silver, Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.
[52] T.Thi Nguyen, N.D. Nguyen, S. Nahavandi, Deep reinforcement learning for multi-agent systems: a review of challenges, solutions and applications, IEEE Trans. Cybern. 50 (9) (2018) 3826–3839.

[53] F. Martinez-Gil, M. Lozano, F. Fernández, Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models, Simul. Model. Pract. Theory 74 (2017) 117–133.
[54] M. Kejriwal, S. Thomas, A multi-agent simulator for generating novelty in monopoly, Simul. Model. Pract. Theory 112 (2021) 102364.
[55] F. Martinez-Gil, M. Lozano, F. Fernández, MARL-Ped: a multi-agent reinforcement learning based framework to simulate pedestrian groups, Simul. Model. Pract. Theory 47 (2014) 259–275.
[56] Jinghui Wang, Wei Lv, Yajuan Jiang, Shuangshuang Qin, Jiawei Li, A multi-agent based cellular automata model for intersection traffic control simulation, Phys. A Stat. Mech. Appl. 584 (2021) 126356.
[57] B. Lucian, B. Robert, D.S. BartA comprehensive survey of multiagent reinforcement learning, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38(2); 156-1722008.
[58] Ke Zhang, Xi Lin, Meng Li, Graph attention reinforcement learning with flexible matching policies for multi-depot vehicle routing problems, Phys. A Stat. Mech. Appl. 611 (2023) 128451.
[59] Jun-Hao Qian, Yi-Xin Zhao, Wei Huang, Model improvement and scheduling optimization for multi-vehicle charging planning in IoV, Phys. A Stat. Mech. Appl. 621 (2023) 128826.
[60] S. Gronauer, K. Diepold, Multi-agent deep reinforcement learning: a survey, Artif. Intell. Rev. 55 (2022) 895–943.
[61] K. Zhang, Z. Yang, T. Başar, Multi-agent reinforcement learning: a selective overview of theories and algorithms, in: K.G. Vamvoudakis, Y. Wan, F.L. Lewis, D. Cansever (Eds.), Handbook of Reinforcement Learning and Control. Studies in Systems, Decision and Control, Springer, Cham, 2021.
[62] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, Adv. Neural Inf. Process. Syst. 30 (NIPS 2017) (2017) 30.
[63] I. Shariq, F. Sha, Actor-attention-critic for multi-agent reinforcement learning, Proceedings of the 36th International Conference on Machine Learning 97(2019) 2961-2970.
[64] Ministry of Transport of the People's Republic of China: JTG B01-2003, Technical Standard of Highway Engineering[S].
[65] V. Mnih, K. Kavukcuoglu, D. Silver, et al., Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.
[66] T. Haarnoja, A. Zhou, P. Abbeel, et al., Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//International conference on machine learning, PMLR (2018) 1861–1870.
[67] Y. Zheng, B. Ran, X. Qu, Cooperative lane changing strategies to improve traffic operation and safety nearby freeway off-ramps in a connected and automated vehicles environment, IEEE Trans. Intell. Transp. Syst. 21 (11) (2020) 4605–4614.