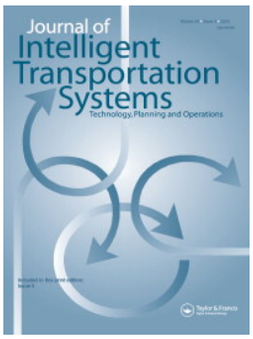


Taylor & Francis
Taylor & Francis Group

Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: 1547-2450 (Print) 1547-2442 (Online) Journal homepage: www.tandfonline.com/journals/gits20

Optimization of hard shoulder running on highways using multi-agent reinforcement learning considering emergency vehicles

Hu Lipeng, Tang Jinjun, Zhe Wang, Zhitao Li, Mingyang Li & Zeng Jie

To cite this article: Hu Lipeng, Tang Jinjun, Zhe Wang, Zhitao Li, Mingyang Li & Zeng Jie (08 Sep 2025): Optimization of hard shoulder running on highways using multi-agent reinforcement learning considering emergency vehicles, Journal of Intelligent Transportation Systems, DOI: 10.1080/15472450.2025.2543823

To link to this article: https://doi.org/10.1080/15472450.2025.2543823

JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS
https://doi.org/10.1080/15472450.2025.2543823

Optimization of hard shoulder running on highways using multi-agent reinforcement learning considering emergency vehicles

Hu Lipeng, Tang Jinjun, Zhe Wang, Zhitao Li, Mingyang Li, and Zeng Jie

Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha, China

ABSTRACT

With increasing travel demand, highway congestion and accidents have become more frequent. As an essential component of intelligent transportation systems (ITS), Hard Shoulder Running (HSR) provides a dynamic strategy to mitigate congestion by temporarily opening shoulder lanes, yet traditional methods often fail to adapt to real-time traffic changes and overlook the shoulder's critical role in ensuring emergency vehicle access. This study proposes a novel HSR optimization framework based on the Long Short-Term Memory (LSTM) and Multi-Agent Deep Deterministic Policy Gradient (MADDPG) reinforcement learning algorithm, integrated with an improved A* algorithm for EV lane clearing. LSTM is used to extract temporal features from traffic data to support intelligent decision-making in MADDPG, enhanced with prioritized experience replay and importance sampling. The EV lane-clearing task is formulated using graph theory, and the improved A* algorithm determines the optimal path for clearing. A simulation case was developed using Simulation of Urban MObility (SUMO) for a section of the Jinan City Ring Highway, China, evaluating four levels of traffic service. Results show the proposed method reduces total travel time by 14.6%, Time Integrated Time-to-Collision (TIT) by 45.2%, and CO2 emissions by 11.9%. Additionally, with EV intervention, braking times are reduced by up to 376.3% and travel time by 18.1% using the improved A* strategy. These findings demonstrate that the integrated LSTM-MADDPG and A* approach effectively enhances highway traffic efficiency, safety, and sustainability under complex real-world conditions.

ARTICLE HISTORY
Received 26 May 2024
Revised 28 July 2025
Accepted 1 August 2025

KEYWORDS
emergency vehicles; hard shoulder running; highway; multi-agent deep deterministic policy gradient; SUMO simulation

1. Introduction

With the rapid development of the economy, the increasing number of vehicles on highways has led to significant issues, including congestion, accidents, and elevated local air pollution. ITS have emerged as an effective solution, leveraging dynamic management technologies to implement HSR—the temporary use of the hard shoulder to improve road capacity and alleviate congestion. As a core service of ITS, HSR dynamically adjusts the operational status of the shoulder lane through real-time traffic monitoring, variable message signs (VMS), and algorithmic optimization methods. Since the 1980s, several European countries, including the United Kingdom, France, and Germany, have been utilizing HSR as a management measure to address highway congestion issues (Guerrieri & Mauro, 2016). The hard shoulder refers to the part of the shoulder adjacent to the main lane, designated for the stopping of vehicles or the passage of special vehicles during emergencies. Considering the special function of the hard shoulder, it is required to provide high-speed access for ambulances, police vehicles, engineering rescue vehicles and other vehicles under emergency conditions. Therefore, the application of HSR needs to take into account special traffic scenarios such as emergency vehicles driving into the regulated area.

Presently, extensive research has been conducted on the impact of HSR strategies on highways and the optimization of these strategies, forming a foundation for subsequent researches. The opening of the hard shoulder significantly affects various aspects of the highway, which can be summarized into three primary categories: efficiency, safety, and vehicle emissions. The implementation of HSR strategies increases the capacity for both local areas and the entire road section, thereby reducing highway congestion and enhancing operational efficiency. The dissipation of congestion also has a significant effect on highway

CONTACT Tang Jinjun jinjuntang@csu.edu.cn Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha, China

safety. Reducing the frequency of occasional accidents, the number of vehicle stops, and vehicle emissions are all benefits of implementing HSR strategies. These conclusions have been sequentially corroborated. Nevertheless, most existing studies predominantly analyze and optimize HSR strategy impacts on highways with a focus on a single objective, such as efficiency or safety. Simultaneously considering the optimization of efficiency, safety, and vehicle emissions represents a challenging yet highly valuable endeavor. Furthermore, a limited number of scholars have explored HSR strategy optimization methods. The optimal HSR strategy is determined by applying the enumeration method and simulation to analyze the implementation effect of different discretized HSR strategies. In addition, Li et al. (2019) employed the genetic algorithm to optimize the HSR strategies, using total travel time as the objective function, aimed at enhancing highway operational efficiency. Methods for optimizing HSR strategies in various scenarios continue to rely on conventional approaches, such as enumeration algorithms and genetic algorithms. Nowadays, real-time changes in freeway traffic flows were more rapid, and the traffic environment has become increasingly complex. Optimizing HSR strategies requires a more detailed consideration of the heterogeneity among various sections of the highway.

Furthermore, current research primarily concentrates on the impact and optimization of HSR strategies, often overlooking the original function of hard shoulders. The most fundamental function of hard shoulders is to serve as a passage for special vehicles during emergencies. EVs are crucial components of emergency systems, requiring prompt and secure responses to various urgent incidents, including ambulances, fire trucks, and rescue vehicles. As the primary means of transportation for emergency response, ensuring their unimpeded passage and reducing response times contributes to an increase in rescue speed. Relevant studies indicate that the "golden hour" for rescuing trauma patients is within the first hour after an accident. The success rate of rescue is closely tied to response speed; the faster emergency vehicles arrive at the accident scene, the higher the likelihood of successful patient rescue. When emergency vehicles reach the accident site within 7 min, the success rate of rescue can reach 94.8% (Newgard et al., 2010). If the arrival time is doubled, the rescue success rate drops to 75.3%. However, the implementation of HSR strategies often results in private vehicles occupying the hard shoulder. When EVs enter the regulated area, they may encounter disruptions from these vehicles, leading to delays and serious damages.

To overcome the limitations of previous studies and address the optimization challenges of HSR strategies—particularly in scenarios where EVs enter the regulated area—this paper introduces a novel L-MADDPG-A* algorithm. The proposed approach is structured within a two-layer problem framework. The primary layer involves optimizing the HSR strategy, while the secondary layer focuses on the EV vehicle lane clearing issue. For the HSR strategy optimization, we present an innovative algorithm that integrates LSTM with MADDPG, effectively leveraging the time-series data extraction capabilities of the LSTM model. Additionally, the reward function within the MADDPG framework is designed to consider three critical performance metrics: safety, efficiency, and emissions. These objectives guide the training process, enabling the agent to learn an optimal control strategy for HSR deployment under dynamic traffic conditions. For the secondary problem of EV lane clearing, we transform the vehicle merging challenge into a graph-based path planning problem. We introduced an A* algorithm that employs heuristic algorithms to determine the optimal clearing solution. Hence, the primary contributions of this paper are as follows: (1) Introduction of a two-tiered model framework for optimizing HSR strategies while considering EVs entering the regulated area. (2) Development of an LSTM-MADDPG algorithm to optimize HSR strategies, considering the temporal feature of traffic data and regional traffic flow heterogeneity. (3) Formulation of the lane clearing problem when EVs enter the regulated area from a graph theory perspective, with the proposal of an improved A* algorithm for lane clearing. (4) Taking into account the heterogeneity of highway traffic flows, construct practical simulation cases for problem-solving and algorithm evaluation across four different highway service level traffic flow scenarios.

The article is structured as follows; The problem formulation and model design are described in Section 2, the methodology is explained in Section 3, the experimental elements and design are described in Section 4, the analysis and discussion of findings are analyzed in Section 5, and the conclusion is summarized in Section 6.

2. Literature review

The following three sections provide a review of existing research outcomes in three main themes: HSR

strategies, EV lane clearing methods, and reinforcement learning algorithms. The HSR strategy research section analyzes the current impact of HSR strategies on highways and the research outcomes of optimization methods. The EV lane clearing methods section reviews the various lane clearing approaches proposed by different scholars. The reinforcement learning algorithm section introduces the current development status of multi-agent reinforcement learning algorithms.

- HSR strategies

To alleviate frequent highway congestion and reduce accidents, implementing Hard Shoulder Running (HSR) strategies has proven to be an effective approach. During periods of heavy traffic, HSR strategies can significantly increase highway capacity, reduce congestion, boost vehicle speeds, and minimize delays (Bhouri et al., 2017; Coffey & Park, 2018; Ghiasi et al., 2018; Li et al., 2017; Newgard et al., 2010; Vadde et al., 2012; Yao et al., 2024). The implementation of HSR strategies can alleviate congestion and improve travel efficiency not only during heavy traffic but also at specific highway bottlenecks, leading to reduction of travel times (Zhang et al., 2022; Zhi et al., 2024). Opening the hard shoulder at the right moment offers a distinct advantage, but it's crucial to note that prolonged open doesn't always improve traffic conditions (Cohen et al., 2010). The effects of implementing the HSR strategy are more noticeable during the time periods with high traffic pressure, particularly on weekends, leading to improvements in both travel time and its reliability (Ma et al., 2016). These findings emphasize to implement different HSR strategies for different traffic scenarios. Furthermore, it is worth noting that, apart from alleviating traffic congestion, the implementation of HSR holds the potential to enhance highway safety to a certain extent (Arora & Kattan, 2023; Chun & Fontaine, 2017; Kononov et al., 2012). The application of HSR strategies on Interstate 2015 (I-66) in northern Virginia resulted in a notable reduction of rear-end collisions by 31% (Waleczek & Geistefeldt, 2021). Aron et al. (2010) study of the implement HSR strategies on a segment of the French freeway (A4-A86) similarly demonstrated a decline in accident occurrence concomitant with reduced road traffic density resulting from the opening of the hard shoulder. Utilizing a Bayesian modeling approach, a comprehensive analysis clarified the impact of broader composite shoulders on highway safety. This analysis indicated a potential decrease of up to 61% in localized accidents and a 31% reduction in fatal crashes (Dutta et al., 2019). These collisions often result from inadvertent causes, and the dynamic opening of the hard shoulder facilitates the rapid evacuation of congested traffic flows, thereby reducing the occurrence of such unintentional accidents (Cohen et al., 2010). In certain cases, by utilizing HSR strategies to alleviate downstream traffic congestion, it is possible to improve the safety conditions upstream. This conclusion stems from Waleczek et al.'s (2021) comprehensive analysis of 13 years' worth of accident data from seven segments of German highways. Improvements in traffic flow conditions and increased highway safety make highway travel smoother, resulting in reduced vehicle occupancy time on the road (Abdel-Aty et al., 2024). During periods of congestion, vehicles frequently experience stop-and-go patterns, resulting in inefficient fuel consumption. As travel times decrease, the cumulative emissions of pollutants from vehicles in localized areas also decrease, subsequently leading to reduced atmospheric pollution. Wilson et al. have discussed the implementation of the hard shoulder strategies to alleviate traffic congestion, concurrently yielding ancillary benefits, including the decrease of carbon emissions and the reduction of environmental pollution (Zeng & Schrock, 2012). Notably, some studies suggest that implementation of HSR strategies exhibits the potential to reduce vehicle fuel consumption and emissions by up to 32% without necessitating road expansion. This result emphasizes the significance of implementing the HSR strategy and shows its economic viability and substantial improvement effects (Wilson, 2009). In the aforementioned study, the improvement effects of HSR strategies on the highway were analyzed from the perspectives of efficiency, safety, and vehicle emissions. Different HSR strategies yield different effects on the highway traffic condition. Consequently, numerous researchers have proposed their own research approaches to address the issue of formulating specific HSR strategies for different traffic flows. Yang et al. (2021) utilized K-Means clustering to categorize traffic conditions, supplemented by factor analysis and TOPSIS method for determining optimal conditions for opening hard shoulder. These initial researches only considered the two fundamental strategies, open and closed, to establish the threshold for opening hard shoulder. Subsequently, some researchers have studied more complex traffic scenarios, utilizing methodologies such as Q-learning (Zhou et al., 2020), IBM CPLEX solver, and others to optimize HSR strategies. However, while the strategy has become more sophisticated, it often

focused on the entire roadway without considering localized traffic flow conditions. This mathematical formalization of the strategy defines the problem as an optimization challenge, making it an invaluable resource for future research endeavors. Li et al. (2019) defined the strategy as the arrangement of the open and closed states of different sub-sections of multiple temporal intervals, and utilizing genetic algorithms to find the optimal HSR strategy with the objective of minimizing travel time. This mathematical formalization of the strategy and the optimize method for HSR strategy provide a valuable reference for future research.

- EV lane clearing methods

The implementation of the HSR strategy yields numerous benefits for highways, but it also introduces certain challenges. Vehicles occupying the hard shoulder during strategy implementation can disrupt approaching EVs which often perform time-sensitive tasks. Such disruptions frequently lead to delays and substantial losses. To address this issue effectively, the clearance of lane for EVs is necessary. This task shares similarities with the process of vehicle merging, where vehicles from the hard shoulder integrate into the primary lanes within a designated segment area (Zhou et al., 2020). In the initial stages of this study, researchers primarily utilized traffic signals to control the merging process. For instance, Li et al. (2014) introduced an ALINEA control algorithm to control the merging process in the context of ramp metering. Advancements in emerging technologies have led to increasing levels of vehicle automation, enabling more efficient and safer merging operations through collaborative vehicle control. The key of effective cooperative control for vehicle merging primarily lies on the determination of the merging order (Cao et al., 2021; Papageorgiou et al., 2008). There exist two approaches for resolving the issue of merging order: the rule-based method and the optimization-based method. Rule-based methodologies ascertain the sequence of vehicle merging by specific rules, which may include the first-in-first-out rule, the arrival time rule, the local gap optimization rule, and others. While these methods are straightforward to implement, they do not ensure the optimality of the merging solution. In contrast, optimization-based methods have the capability to systematically search for the most optimal local or global merging order, resulting in a more precise solution (Ding et al., 2020). Awal et al. (2013) proposed a recursive pruning algorithm with the optimization objective of total travel time to find the optimal sequence for merging a set of vehicles on the main road and ramps. In a similar way, Fukuyama (2020) proposed a decentralized control framework within the merging region and utilized dynamic game theory to optimize vehicle merging trajectories. They applied a zero-suppressed binary decision diagram (ZDD) approach to solve this problem. Mu et al. (2021) formulated the merging problem as a mixed-integer linear programming problem and employed a heuristic algorithm to solve it. Experimental validation of their method confirmed its capability to meet performance expectations within a context of Connected and Autonomous Vehicle (CAV) traffic flow (Spatharis & Blekas, 2024).

- Reinforcement learning algorithms

The existing methods for optimizing HSR strategies are relatively conventional and struggle to adapt to the heterogeneous traffic flow states on highways. Employing a multi-agent reinforcement learning algorithm for regional control can effectively address the aforementioned issues. The multi-agent reinforcement learning algorithm is an extension of single-agent reinforcement learning, accommodating multiple control entities. Early reinforcement learning (RL) mainly focused on tabular and approximate algorithms, suitable for tasks with low-dimensional state and action spaces. To tackle more complex tasks with higher dimensional state or action spaces, methods with robust capabilities for high-dimensional data representation and abstraction are required. Deep neural networks are well-suited for this purpose. Deep learning, with its robust data representation capabilities, has transformed feature engineering from manual processes to network-driven extraction, enhancing efficiency and accuracy while addressing the issue of exploding data dimensions. The combination of deep learning and reinforcement learning is referred to as deep reinforcement learning (Mnih, 2013). Deep Q networks (DQNs), one of the most classic integration methods, have gained substantial attention due to remarkable success in gaming applications, consequently leading to the development of various deep reinforcement learning algorithms (Van Hasselt et al., 2016). Single-agent deep reinforcement learning algorithms such as Double Deep Q-Networks (DDQN), Asynchronous Advantage Actor-Critic (A3C), and Soft Actor-Critic (SAC) have been proposed and applied across diverse domains, encompassing robot control, video processing, text mining, and more

(Balmer et al., 2004; Haarnoja et al., 2018; Hou et al., 2023; Kim et al., 2025; Low et al., 2024; Mnih, 2016). Nevertheless, in scenarios involving multiple control entities in a system, such as UAV formation problems, fleet cooperative navigation, multi-robot path planning, and similar applications, single-agent reinforcement learning algorithms are no longer suitable. Consequently, multi-agent reinforcement learning algorithms have emerged to address these challenges. Multi-agent algorithms, in contrast to single-agent ones, are more intricate. Multiple agents not only interact with the environment but also communicate with each other (Huang et al., 2024). In recent years, the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm proposed by Lowe et al. (2017) has become a research hotspot, widely applied in areas such as energy trading, regional traffic signal coordination, multi-task resource allocation, and drone swarm coordination. Under the MADDPG algorithm framework, addressing the optimization problem of HSR strategies involves segmenting the control region into multiple sub-regions, each treated as an intelligent agent. Each agent independently makes control decisions while simultaneously sharing data with other agents to optimize its policy network. Traditional heuristic algorithms and Q-learning algorithms for control optimization often overlook the distinctiveness of controlled road sections. The MADDPG algorithm defines reward functions for agents, guiding their training through interactions with the environment, ultimately computing the optimal control strategy for each intelligent agent.

Based on the analysis of the existing researches and research problem, this paper proposed a multi-agent reinforcement learning algorithm for optimizing the implementation effect of HSR strategies in different traffic scenarios. Furthermore, we utilized the A* algorithm to address the challenge of merging hard shoulder vehicles into the primary road lane, particularly during EVs enter the regulated area. Ultimately, experimental results conclusively demonstrate the effectiveness of the proposed methodology.

3. Problem formulation and model design

3.1. Statement of the problem

Assume that not all highway segments will implement the HSR strategy, nor will EVs consistently enter segments where the HSR strategy is applied during all time periods. Based on this assumption, as illustrated in Figure 1, the problem is formulated as a two-level structure. The first level focuses on optimizing strategies for specific highway segments where the HSR strategy will be implemented, aiming to enhance overall traffic efficiency, safety, and emissions within the areas affected by the strategy. The second level addresses the process of lane clearing when EVs enter areas subject to the HSR strategy, ensuring that vehicles on the hard shoulder do not obstruct their passage.

At the first level, referred to as the major problem, the main objective is to manage and regulate the use of the hard shoulder based on real-time fluctuations in traffic flow. The solution process for this problem consists of five distinct steps: (1) real-time data collection, (2) data processing and feature extraction, (3) HSR strategy definition and optimization, (4) strategy implementation and data feedback, and (5) strategy evaluation and adjustment. Initially, real-time traffic data is collected using highway detectors, capturing key metrics such as vehicle count, speed, and average travel time. The data is then cleaned and processed to extract relevant features, ensuring the quality of the data for subsequent analysis and optimization. Following this, the HSR strategy is mathematically formulated, and optimization algorithms are employed to search for the optimal strategy. The goal is to enhance traffic flow, passage efficiency, and other traffic indicators on the highway segments where the strategy is applied. The optimization goal is to enhance traffic flow, passage efficiency, and other key performance indicators on the highway segments where the strategy is implemented. The optimized strategy is then implemented by transmitting real-time strategy information to vehicles, with feedback data collected after implementation. Lastly, the effectiveness of the strategy is assessed by analyzing its impact on traffic flow, accident rates, and emissions. Based on these evaluation results, adjustments are made to refine the strategy for the next iteration.

When EVs approach or enter segments where the HSR strategy is active, it is crucial to ensure their smooth passage through the areas affected by the strategy. This concern forms the basis of the second level, referred to as the subproblem, which specifically addresses the lane-clearing process for EVs. This subproblem is resolved through three key steps: (1) emergency vehicle data collection, (2) lane-clearing algorithm design, and (3) implementation and evaluation of clearing measures. First, real-time data on EVs is collected, including their location, speed, and acceleration, to enable accurate tracking. Building on this information, a lane-clearing algorithm is designed to clear vehicles obstructing the EVs' movement,

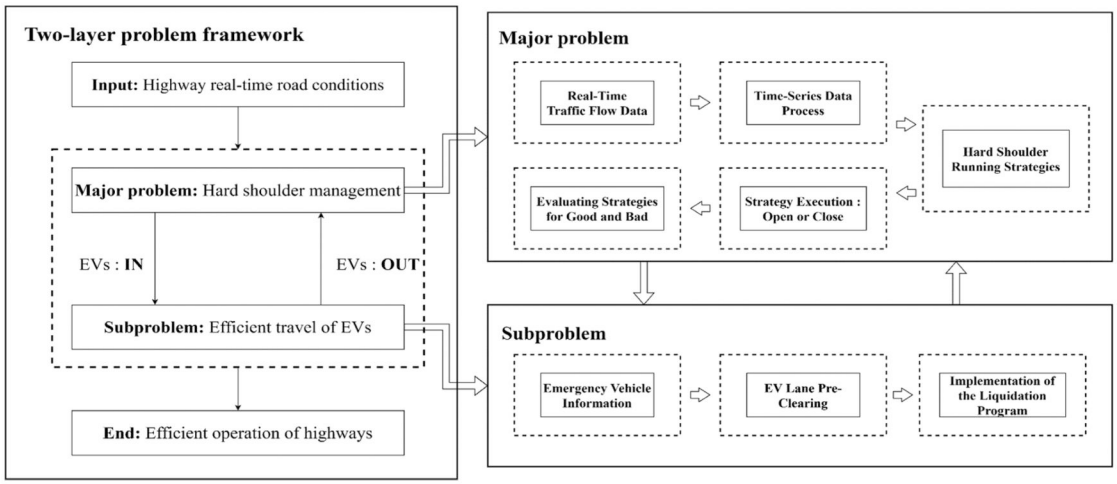


Figure 1. Framework of two-layer problems.

thereby ensuring unobstructed passage. The lane-clearing measures are implemented, followed by an evaluation of their effectiveness. Once the EVs exit the controlled area, the subproblem transitions back to the major problem, resuming the iterative process to maintain efficient road management.

Due to the challenges associated with testing strategies in real-world conditions, a simulation platform has been developed using advanced simulation software. This platform enables data collection and strategy implementation within an interactive environment. It facilitates the simulation of the HSR strategy optimization process under varying traffic conditions, as well as the testing and evaluation of different lane-clearing solutions. In the following sections, we will provide a detailed overview of the strategy optimization framework, the iterative process, and the methods for implementing and evaluating the lane-clearing algorithm.

3.2. Strategic formulation

In the process of solving the major problem, it is necessary to define the HSR strategy with mathematical form. To determine the open state of different road sections at different time intervals, the design of the HSR strategy needs to be considered from the aspects of control distance and time. As depicted in Figure 2, this study employs a temporal and spatial discretization approach to formulate the highway HSR strategies. The total length of the highway segment in the figure is denoted as L, and this segment is subdivided into multiple discrete sub-segments, labeled as $\{l_1, l_2, l_3, ..., l_n\}$. Each of these sub-segments is the smallest control unit subject to the HSR strategy, with two potential states either open or close. When sub-segment l_i is in the open state, it facilitates the passage of vehicles from the primary lane, allowing them to transition into the sub-segment. Conversely, if the sub-segment is closed, it restricts access for vehicles within the primary lane or those traveling straight from the preceding sub-segment. Let each sub-segment l_i correspond to a variable c_i, which is a 0–1 categorical variable. When $c_i = 1$, it means sub-segment l_i is open; when $c_i = 0$, it means sub-segment l_i is closed. Then the mathematical representation of the strategy for the control segment L is: $[c_1, c_2, c_3, ..., c_n]$. Employing a similar approach for temporal discretization, we define the total duration of the control segment as T. Each time interval, denoted as T_i, constitutes a control cycle for the modification of the open and closed status of individual sub-segments. The strategy of each control cycle is expressed as: $CT_i = [c_1 = 0, c_2 = 1, c_3 = 0, ..., c_n = 1]$. The HSR strategy is then composed of the combination of strategies from all cycles: $[CT_1, CT_2, CT_3, ..., CT_N]$.

3.3. Modeling the problem

Figure 3 provides a schematic representation of the major problem and sub-problem. To enable the transmission of the HSR strategy information, a crucial assumption is made, wherein a signal bar is positioned at each sub-segment l_i. These signal bars have dual-color indicators: a red light signifies that the segment is closed, while a green light indicates that it is accessible for entry and exit. These signal bars are designed to facilitate bidirectional communication with connected vehicles, transmitting state information of segment. Connected vehicles are equipped with sensors to perceive the positions of surrounding vehicles, as

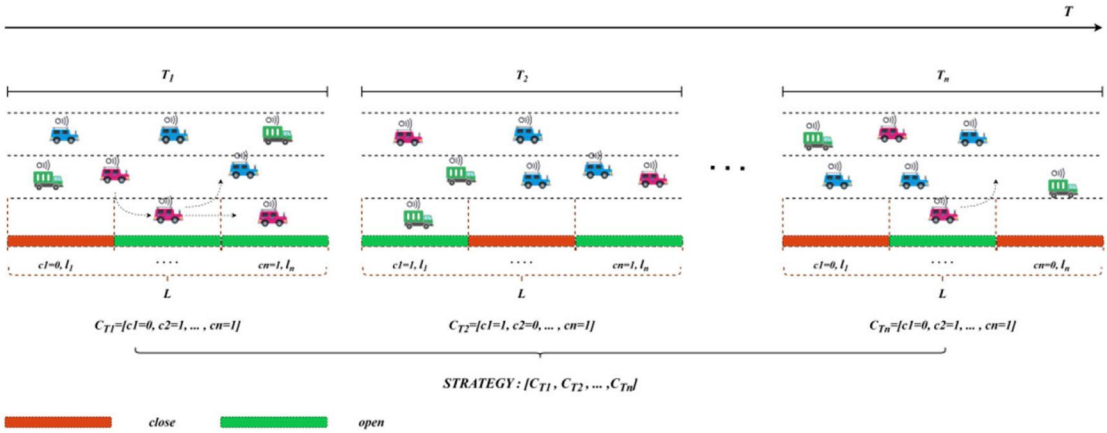


Figure 2. Strategy of HSR operation.

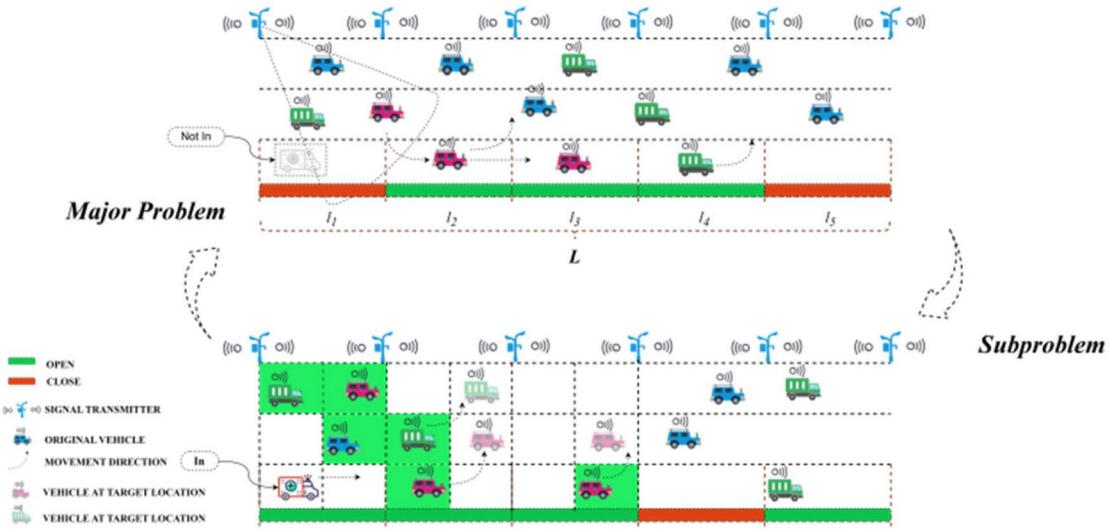


Figure 3. Framework for problem switching.

well as to gather data like their speeds, accelerations, the status of hard shoulder segment, and the next open decision time. Moreover, the highway lanes also can collect different information by equipping various detection devices, including the number of vehicles within each roadway, waiting times for vehicles, average speeds of roadways, average travel times, roadway occupancy rates, and other data (Azimjonov & Özmen, 2021).

In the major problem, the EVs did not enter the regulated road section. As illustrated in Figure 3, within a control temporal interval denoted as T_i, the sub-segment opening strategy is defined as follows: $[c_1 = 0, c_2 = 1, c_3 = 1, c_4 = 1, c_5 = 0]$. During this temporal interval, the signal indicators for sub-segments l_2, l_3, and l_4 illuminate in green, transmitting instantaneous signals to the vehicles driving on the lanes. A connected vehicle in the main lane receives this signal, or the driver observes the color of the signal bars, allowing for a lane change to access these three sub-segments. For sub-segments l_1 and l_5, the signal bar displays as red, permitting only EVs to access them. Once an EV enters, the situation transitions from the major problem to the sub-problem. In this context, ensuring unimpeded EV travel is paramount, which necessitates the clearance of vehicles on the hard shoulder. As depicted in Figure 4, utilizing a grid-based approach, all lane vehicles are divided into cells. Each vehicle corresponds to a cell, forming an $M \times N$ vehicle arrangement matrix. Vehicle behaviors, including lane changes and straight-line driving, are transformed into movement between cells. The movement of each vehicle to an adjacent cell leads to a new arrangement. The hard shoulder lane corresponds to one of the columns in the matrix, and the task at hand involves the transfer of all vehicles from this column to other cells. As in the example in Figure 4, the black dashed box indicates the column corresponding to the hard shoulder lane. The clearing task is completed by moving vehicles No. 1 and No. 2 to the left

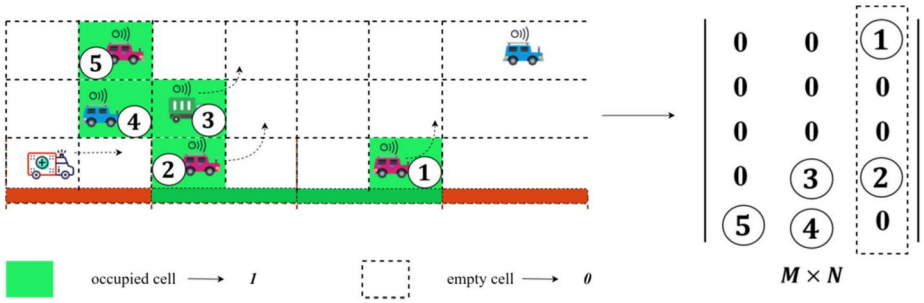


Figure 4. Lane clearing process in HSR.

two columns. In this scenario, an issue arises wherein vehicle No. 2 is obstructed by the presence of vehicle No. 3. Consequently, it becomes imperative to design an effective algorithm with the purpose of determine the optimal resolution for this clearing problem.

4. Methodology

4.1. Multi-agent reinforcement learning

Reinforcement learning (RL) is an important branch of machine learning. It is based on a trial-and-error learning mechanism, where an optimal policy is trained by agents interacting with its environment to maximize rewards. Deep reinforcement learning (DRL) combines reinforcement learning with deep learning as a new approach to solve complex control problems. When multiple control objects are involved in the environment, multi-intelligence deep reinforcement learning emerges. Specifically, multi-agent reinforcement learning refers to a method where multiple agents, in the context of sequential decision-making problems, interact with the environment and learn to achieve an optimal equilibrium state for the task under predefined reward rules. In general, we model the problem of multi-agent reinforcement learning as a partially observable Markov decision problem (Busoniu et al., 2008). The decision-making process, involving N agents, is described by the tuple $(N, S_t, A_t, \mathrm{P}, S_{t+1}, R_t, \gamma)$. Here, $S_t = \{\mathrm{O}_1 \times \mathrm{O}_2 \times \cdots \times \mathrm{O}_N\}$ represents the environmental information data observed by all agents at time t, denoting the current state. $A_t = \{\mathrm{a}_1 \times \mathrm{a}_2 \times \cdots \times \mathrm{a}_N\}$ denotes the joint action taken by all the agents under their existing policy, where each action a_i belongs to the action space ($\mathrm{a}_i \in \Omega$). The state transition function, denoted as P, describes the probability of transitioning from the current state to the next state. S_{t+1} represents the state at the next time step. $R_t = \{\mathrm{r}_1, \mathrm{r}_2, ..., \mathrm{r}_N\}$ are the rewards received by each agent during the current state transition. The discount factor, $\gamma \in [0, 1]$, is used to calculate the expected cumulative reward. At time t, each agent $\mathbb{N}_i$ selects an action a_i based on their individual policy function π_i and forms a joint action A_t which is subsequently executed. Then, the state of the environment is transformed to S_{t+1}, and all agents receive corresponding rewards R_t. Typically, at a certain moment T, the interaction between agents and the environment concludes when the tasks of all agents are completed or the reward criteria are met. The primary objective of multi-agent reinforcement learning is to determine a set of optimal policies, $\{\pi_1, \pi_2, ..., \pi_N\}$, that maximize the expected cumulative discounted reward of all agent: $maxE\left[\sum_{t=0}^{T} \gamma^t R_t\right]$.

4.2. MADDPG algorithm

In a multi-agent system, different relationships exist between agents, often determined by the designed objectives. Depending on the specific optimization objectives, relationships are broadly classified into three primary categories: cooperation, competition, and mixed tasks. Cooperation is particularly applicable in practical production scenarios. Notably, the Multi-Agent Deep-Deterministic Policy Gradient Algorithm (MADDPG) has emerged as an efficient multi-agent reinforcement learning algorithm for 'cooperative-competitive' mixed tasks. As one of the most popular multi-agent reinforcement learning algorithms, it has been widely applied across various domains (Gronauer & Diepold, 2022). This algorithm extends from the single-agent reinforcement learning algorithm, DDPG (Deep Deterministic Policy Gradient), employing the well-known Actor-Critic (AC) framework. The DDPG algorithm can deal with the task of continuous action space, which mainly consists of two critical networks, an Actor network, and a Critic network. The Actor network is responsible for receiving state information and outputting corresponding actions, representing the agent's policy. The Critic network models the state value Q-function to evaluate policies and perform parameter updates. Both the Actor and Critic networks encompass two

sub-networks: one designated as the 'current network' and the other as the 'target network', with identical network architectures. The MADDPG algorithm utilizes the foundation of the DDPG algorithm to advance it and extend its applicability to multi-agent systems. It operates within a framework of centralized training and decentralized execution. Specifically, it conducts centralized training by training a global critic network Q_{gobal} with inputs from all agents to update the policy networks π_i for each agent. The current Actor network, corresponding to the agent's policy, receives the environmental data observed by the agent, calculates the distribution of the agent's actions, and selects the actions through the '$\varepsilon - greedy$' methodology. Subsequently, agents execute these actions and receive environmental rewards, while the Critic network scores actions based on both actions and environmental states. The Actor network updates the network parameters based on the environmental rewards and scores, enabling the agents to select actions that result in higher cumulative rewards. This iterative process continues until parameter updates are complete, yielding the optimal policy. Assuming ϑ_i represent the current Actor network π_i parameters and the Critic network Q as θ_i. Then the set of all Actor networks is $\Pi = \{\pi_1(s,\vartheta_1), \pi_2(s,\vartheta_2)\cdots\pi_{N-1}(s,\vartheta_{N-1}), \pi_N(s,\vartheta_N)\}$ for N agents. where $s = \{o_1 \times o_2 \times \cdots \times o_N\}$ denotes the environmental data acquired by all agents. The target Actor network, which shares the same structure as the current Actor network, denoted as $\Pi' = \{\pi'_1(s,\vartheta'_1), \pi'_2(s,\vartheta'_2)\cdots\pi'_{N-1}(s,\vartheta'_{N-1}), \pi'_N(s,\vartheta'_N)\}$. The Critic network of all agents can be represented as: $Q = \{Q^1(s,a,\theta_1), Q^2(s,a,\theta_2)\cdots Q^{N-1}(s,a,\theta_{N-1}), Q^N(s,a,\theta_N)\}$. The target Critic network is: $Q' = \{Q_1'(s,a,\theta_1'), Q_2'(s,a,\theta_2')\cdots Q_{N-1}'(s,a,\theta_{N-1}'), Q_N'(s,a,\theta_N')\}$, where $a = \{a_1 \times a_2 \times \cdots \times a_N\}$ denotes the joint action of agents. The update formula for the Critic network is as follows:

$$\mathcal{LOSS}(\theta_i) = \mathbb{E}_{s,a,r,s'}\left[Q_i^{\pi}(s,a,\theta_i) - y)^2\right], y = r_i + \gamma Q_i^{\pi'}(s',a',\theta_i)|_{a'_k=\pi'_k(o_k)} \quad (1)$$

Where γ is the discount factor, r_i is the reward of the agent, and π' denotes the target Critic network. According to the definition of cumulative discount reward, for agent i: $\mathbb{J}(\vartheta_i) = \mathbb{E}[R_i]$. Its gradient can be written as:

$$\nabla_{\vartheta_i} J(\pi_{\vartheta_i}) = E_{(s,a)\sim D}[\nabla_{\vartheta_i}\pi_i(a_i|s_i)\nabla_{a_i}Q_i^{\pi}(s,a_1,a_2\cdots a_N)|_{a_i=\pi_{\vartheta_i}(s_i)}] \quad (2)$$

Where D represents an empirical data replay pool and is used for network training by sampling data. It is composed of tuples $(s_t, a_1^t, a_2^t, ..., a_{N-1}^t, a_N^t, r_t^1, r_t^2, ..., r_t^N, s_{t+1})$. For the target Actor network and the target Critic network the network parameters are updated by using the soft update mechanism:

$$\vartheta^{Q_i'} = (1-\partial)\vartheta^{Q_i} + \partial\vartheta^{Q_i'} \quad (3)$$

$$\theta^{\pi_i'} = (1-\partial)\theta^{\pi_i} + \partial\theta^{\pi_i'} \quad (4)$$

Where $\vartheta^{Q_i'}, \theta^{\pi_i'}$ are the parameters of the target Actor, target Critic network respectively. ∂ is the update step size, which is used to control the speed of updating the network parameters.

4.3. LSTM algorithm

Long Short-Term Memory (LSTM) is a type of recurrent neural network that evolved from the traditional Recurrent Neural Network (RNN) (Hochreiter, 1997). LSTM effectively addresses the limitations of traditional RNNs, enabling the processing of long time sequences, feature extraction, and resolving long-term dependencies. As shown in Figure 5, its network structure is built on a chain of repeating neural units, with each unit having a simple architecture, including an input layer, a hidden layer, and an output layer (Xu et al., 2022). In the hidden layer, it replaces the basic RNN units with memory cells that consist of three crucial gates: the input gate, forget gate, and output gate, along with a memory state. Each gate comprises activation functions and element-wise multiplication functions. These gates are responsible for retaining useful information from the current sequence, selectively forgetting previous data, and transferring data information. The memory state is responsible for retaining data states, selectively transmitting information, and organizing all memory

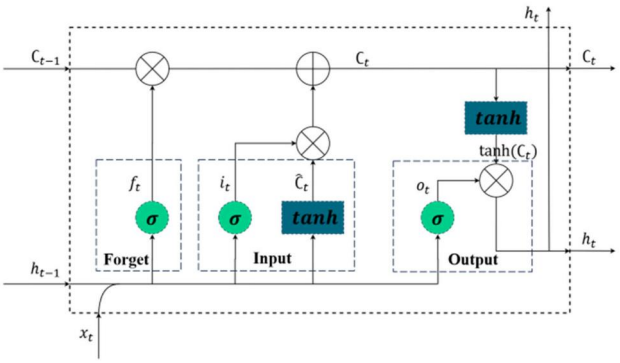


Figure 5. Structure of long short-term memory (LSTM).

cells. As illustrated in Figure 5, f_t, i_t, and o_t are the output signals from the three gates, while C_t represents the current memory state and C_{t-1} signifies the state information that needs to be retained from the previous time step. Notably, x_t corresponds to the input data at the current time step, h_{t-1} refers to the output information from the hidden layer at the previous time step, and h_t denotes the hidden layer output information (Yan et al., 2021). The formula for its calculation is as follows:

$$\text{Forget}: f_t = \sigma\left(W_f \cdot [x_t, h_{t-1}] + b_f\right) \tag{5}$$

$$\text{Input}: i_t = \sigma(W_i \cdot [x_t, h_{t-1}] + b_i) \tag{6}$$

$$\text{Output}: o_t = \sigma(W_o \cdot [x_t, h_{t-1}] + b_o) \tag{7}$$

$$\text{Candidate Data}: \hat{C}_t = \tanh(W_C \cdot [x_t, h_{t-1}] + b_C) \tag{8}$$

$$\text{Record State}: C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t \tag{9}$$

$$\text{Hidden Layer}: h_t = o_t \otimes \tanh(C_t) \tag{10}$$

In the equation, σ denotes the sigmoid function; W, b denote the weight matrix and offset, respectively; tanh denotes the hyperbolic tangent function; and $\otimes$ is an elemental product operator.

4.4. Optimization algorithm: LSTM-MADDPG algorithm

Highway traffic flow data exhibit characteristics of substantial dimensions and extended time series. The process of feature extraction from traffic flow sequence data serves as a fundamental procedure in data processing. In this study, traffic flow data collected through detector is used to characterize the traffic state of a highway segment over a temporal interval, which serves as the basis for formulating HSR strategies. Notably, LSTM networks possess robust capabilities in processing sequential data, effectively extracting features from extended time sequences. As depicted in Figure 6, we incorporate LSTM networks into the MADDPG algorithm as a module for feature extraction from environmental observation data. This combination significantly improves the handling of long-term observation data from various highway segments. Conventional MADDPG algorithms typically lack a temporal dimension in their input data which just provide one-dimensional or multi-dimensional representations of environmental information. Since the traffic flow data have a temporal dimension characteristic, extracting the features of the data through LSTM network can enhance the effectiveness of the MADDPG algorithm. Furthermore, to enhance the effectiveness of the

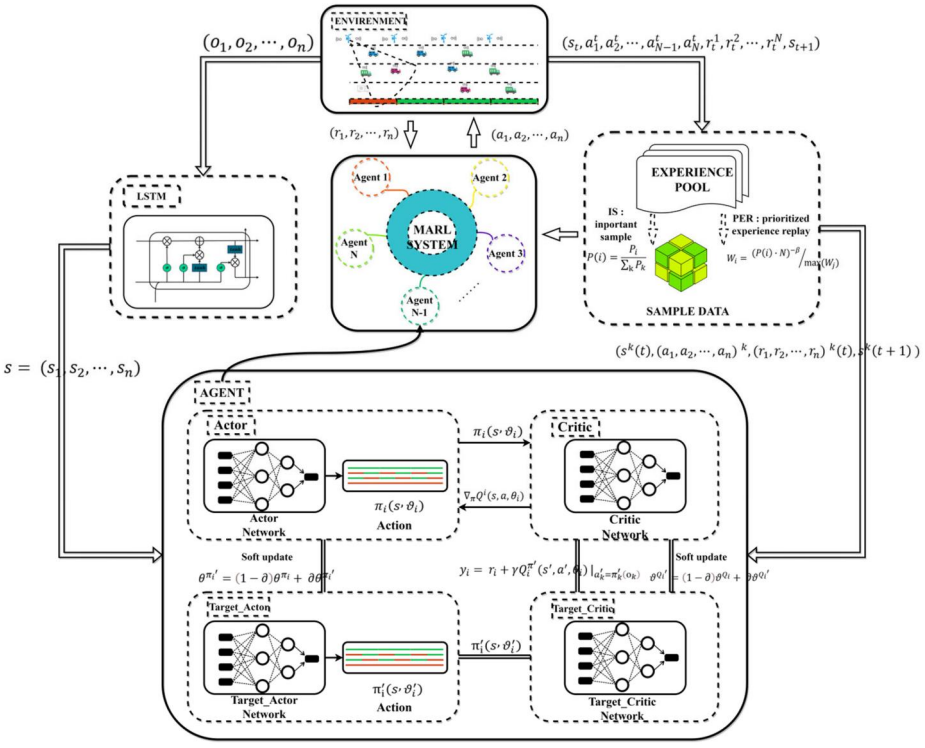


Figure 6. Architectural framework of the LSTM-MADDPG algorithm.

training process for the LSTM-MADDPG algorithm, a series of measures is instituted.

- Prioritized experience replay (PER) (Schaul et al., 2015)

within the training process for conventional reinforcement learning algorithms, the utilization of experience replay is a common practice. To enhance the efficiency of learning from training data in the MADDPG algorithm, a prioritized experience replay approach is adopted. This methodology, in conjunction with the loss function derived from the Critic network, computes a prioritization basis denoted as $\mathcal{L}_i$ for each data sample. The prioritization basis $\mathcal{L}_i$ is calculated according to the following formula: $\mathcal{L}_i = |Q_i^{\pi}(s, a, \theta_i) - y_i|$.The priority $P(i)$ of data sample i is given by the following two equations:

$$P_i = (|\mathcal{L}_i| + \mathcal{T})^{\alpha} \tag{11}$$

$$P(i) = \frac{P_i}{\sum_k P_k} \tag{12}$$

where α is a hyperparameter between 0 and 1 used to control the preference of the empirical priority in the sampling process. $\mathcal{T}$ is a parameter greater than 0.

- Important sample (IS)

In the prioritized experience replay algorithm, individual samples are assigned unique priorities which determine the sampling frequency for each sample. Samples with higher priorities are selected more frequently, while samples with lower priorities have a minimal or zero sampling frequency. Such an approach alters the original distribution of training samples and can lead to unexpected convergence of network parameters, resulting in bias. To address this bias issue, we employ the technique of importance sampling. For each sample's priority, we calculate its importance sampling weight, denoted as W_i. Through priority-based sampling, we collect a set of experience samples along with their corresponding sample weights, W_i. Subsequently, based on these sample weights and the network's loss function, we calculate the sample losses and perform a weighted summation. Specifically, we multiply the loss value computed for each sample by its corresponding weight and then sum these values, updating the network parameter. The formula for this weighting is shown as follows:

$$\mathrm{W_i} = {(\mathrm{P(i)\cdot N})^{-\beta}} \Big/ {\max(\mathrm{W_j})} \tag{13}$$

- Credit Assignment (CA)

In a multi-agent system, evaluating each agent's contribution to the overall performance through credit assignment helps prevent the occurrence of lazy agents. The MADDPG algorithm designs independent reward function for each agent to mitigate bias in assessing agent contributions. Additionally, each agent maintains cooperative relationships with other agents through a global Critic network.

4.5. Improved A* clearing algorithm

The A* algorithm is a heuristic search algorithm widely applied in graph search problems like path planning (Hart et al., 1968; Seet et al., 2004). It combines the greedy strategy from heuristic methods with the breadth-first search algorithm from shortest path search algorithms. Heuristic methods in which past empirical information is rationally utilized to improve the speed of the algorithm, often can only provide approximate optimal solutions. On the other hand, breadth-first search is a precise algorithm that disregards computation time, theoretically capable of finding the shortest path for a given path planning problem. The A* algorithm combine these two algorithms with their strengths. Its core lies in the design of two essential functions: one that estimates the cost of the best path from the current search node to the target node and another that calculates the total cost of the paths searched so far. The algorithm evaluates nodes in the graph by computing the sum of these two functions for each node, thereby determining the direction of each search node and finding the shortest cost path.

As illustrated in Figure 7, for the purpose of EV lane clearing, we discretize the area spatially where vehicles travel. Every vehicle is mapped to a grid cell in this spatial discretization system. Vehicle movement is thus described as transitions between these grid cells. Initially, the vehicle's position is represented as a 0–1 matrix, which serves as the initial node. After completing EV lane clearing, the arrangement matrix of vehicle positions becomes the target node. Vehicle position matrices are altered by moving left or right within the grid cells, with each arrangement matrix representing a node. Different movements of various vehicles from the initial node generate different nodes, interconnected to form a graph. Multiple paths emerge between the

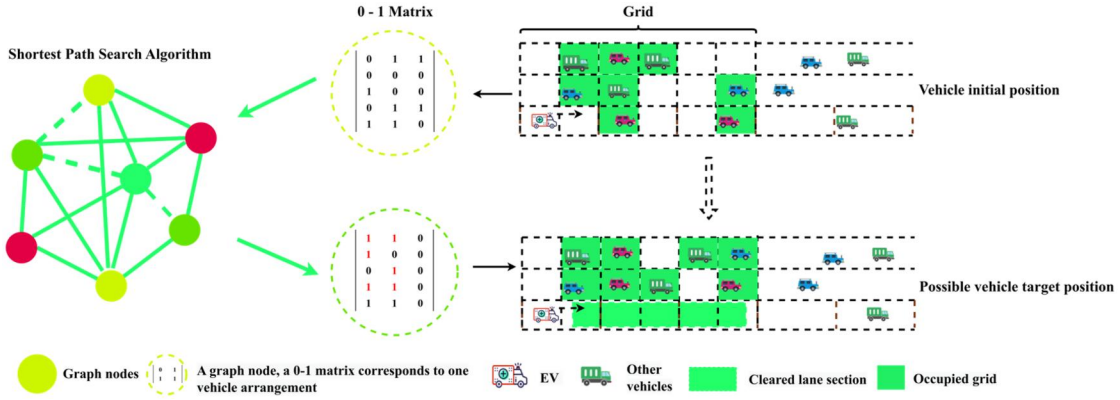


Figure 7. Design of the a* algorithm framework.

Table 1. The improved a* algorithm.

Improved A* Algorithm
1. Define τ as the unit cost of each vehicular movement
2. The initial node, denoted as S, and the terminal state, represented as Terminate
3. Design the function $G(n)$: represent the cost of the path from the initial node S to node n
4. Design the admissible heuristic function $H(n)$: estimate the cost of reaching the target state Terminate from node n
5. Design the node evaluation function $F(n)$: serves to assess the overall cost of a given node, $F(n) = G(n) + H(n)$
6. Initialize the open list (O_{list}) and add the initial node S to the list. Initialize the mark list (M_{list})
7. Calculate the F function values for all nodes in the open list and select the node with the minimum value, denoted as O
8. If the node O belongs to the target state Terminate, the process concludes. If it does not, node O is added to M_{list}, and its adjacent nodes are placed into O_{list}, and the process repeats at step 7.
9. After the termination of the algorithm, the trajectory of the vehicle will be determined by a path backtracking process.

initial and target nodes, and the shortest path search provides the optimal lane clearing solution. Consequently, the original problem of lane clearing is transformed into a graph shortest path search problem.

This study employs the improved A* algorithm as the solution methodology for finding the shortest path. To reduce the computational complexity of the search process, it is assumed that only one vehicle moves at a time in each step. Since the existence of multiple potential arrangements satisfying the requisite conditions in the context of EV lane clearing, the A* algorithm is enhanced by redefining the termination criteria. Instead of terminating at a specific target node as before, the new termination condition essentially comprises a set of many nodes. In the EV lane clearing problem, the termination state is defined as the condition where all vehicles on the EV lane within a specified area have been completely cleared (Li et al., 2022). Thus, the improved A* algorithm, when combined with the original algorithm steps, is shown in Table 1.

5. Experimental elements and design

5.1. Design of experimental factors

5.1.1. Agent

In accordance with Figure 8, this study combines consecutive adjacent road segments into a unified entity, which is designated as an agent ($AGENT_i$). Assuming that each agent controls two consecutive road segments, denoted as l_1 and l_2, managing their open and closed states while collecting data information. The entire set of road segments is divided into N agents, with each agent controlling an equal number of segments.

5.1.2. Action space

Each agent controls a combination of consecutive road segments, and its action space consists of the various combinations of open and closed states for these segments. As Figure 8, each agent controls two sub-segments (such as: l_1, l_2), then the action space will be: $\{(0,0),(0,1),(1,0),(1,1)\}$, and the corresponding control variables take the values of: $\{(c_i = 0,\ c_{i+1} = 0), (c_i = 0, c_{i+1} = 1),\ (c_i = 1, c_{i+1} = 0),\ (c_i = 1, c_{i+1} = 1)\}$. Since each agent controls the same number of sub-segments, each agent has the same action space. The action space for each agent changes with the number of road segments it controls.

5.1.3. State

Agents gather information from the environment to determine their control strategies. Information observed by agents consists of two parts: autonomous agent-specific observations and global observations for the entire regulated segment. The autonomous

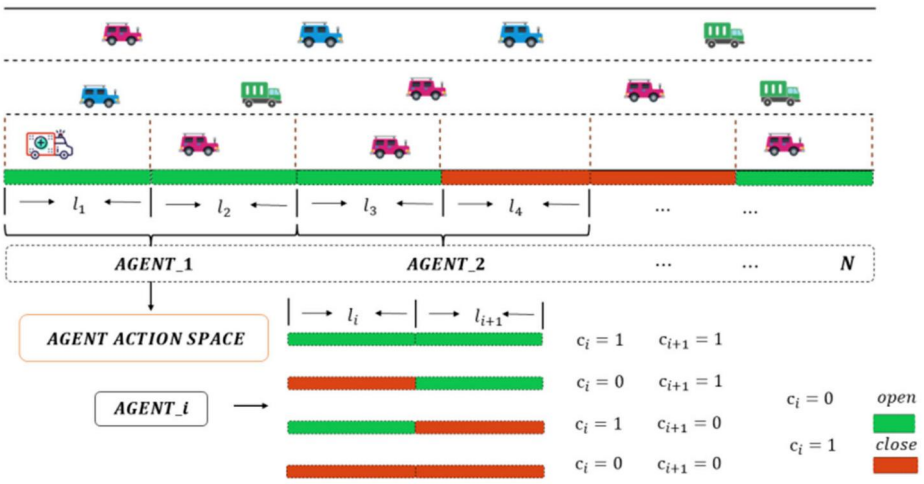


Figure 8. Design of individual components in agents.

observation information for each agent is denoted as $s_{agent\_i} = [c, w, v, tt, occ]$, where c represents the total number of vehicles traveling on the roadway, w indicates the waiting time of the roadway, v represents the average speed of vehicles on the managed roadway, tt denotes the average travel time of vehicles on the roadway, and occ indicates the occupancy rate of the roadway. The global observation information is represented as $s_{environment} = [C, W, V]$, where C denotes the total number of vehicles in the global network, W is the global waiting time, and V represents global traffic data. Each agent's observation data is represented as $S_i = [s_{agent_i}, s_{environment}]$, and the overall network state information can be expressed using the following formula: $S = [S_1, S_2, ..., S_{N-1}, S_N]$.

5.1.4. Reward function

To improve highway traffic operation through the implementation of HSR, this study designs reward functions from three perspectives: efficiency, safety, and emissions. From an efficiency standpoint, the total travel time of vehicles is considered as the efficiency reward function.

$$r^i_{eff} = \log_{10}(c^i_{total\_travel\_time}) \tag{14}$$

$$c^i_{total\_travel\_time} = \sum_{t=0}^{T} c^i_t \tag{15}$$

where $c^i_{total\_travel\_time}$ denotes the total travel time of the i-th agent during a period of T.

From a security perspective, the safety reward function is designed by incorporating two safety indicators: Time Exposed Time-to-Collision (TET) and Time Integrated Time-to-Collision (TIT). TET represents the total time vehicles spend in a risky driving state, where vehicles are considered in a risky state when their Time-to-Collision (TTC) falls below a threshold, typically around 2 s (Li et al., 2022). On the other hand, TIT represents the integral of the vehicle's collision-time curve, a metric utilized for the comprehensive assessment of vehicular safety. The indicator of safety combing TIT and TET is shown as follows:

$$r^i_{safe} = \alpha * \log_{10}(tit_i) + \beta * \log_{10}(tet_i) \tag{16}$$

To reduce vehicle emissions on the highway, an emission reward function is designed based on the emissions of three gases: carbon monoxide (CO), carbon dioxide (CO_2), and nitrogen oxide (NO). Herein, c^i_{CO} represents the CO emissions of the i-th agent, $c^i_{CO_2}$ denotes the CO_2 emissions produced by the i-th agent, and c^i_{NO} is the NO emissions generated by the i-th agent.

$$r^i_{em} = \kappa * \log_{10}(c^i_{CO}) + \mu * \log_{10}(c^i_{COO}) + \gamma * \log_{10}(c^i_{NO}) \tag{17}$$

The reward function r for the total number of all agents is designed as follows:

$$r = \sum_{i=1}^{N} \rho * r^i_{eff} + \sigma * r^i_{safe} + \tau * r^i_{em} \tag{18}$$

5.2. Design of experimental simulation

To validate the effectiveness of the proposed method, SUMO simulation software is employed for case analysis. As illustrated in Figure 9, a simulation segment model is established for a 6.2 km highway section between "Xingcun Interchange" and "Ganggou Hub Interchange" in Jinan city, Shandong province. This research segment is divided into 16 sub-sections, and the length of each sub-segment ranges from 300 to

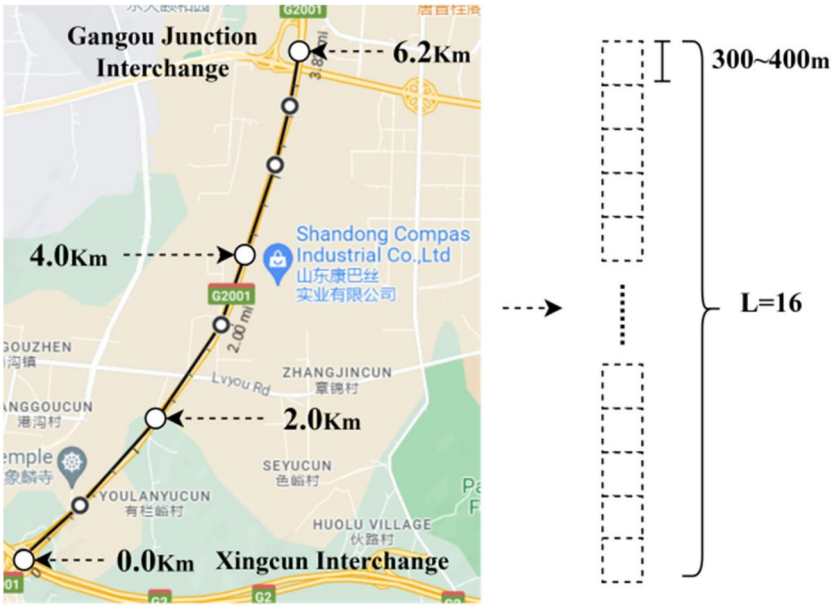


Figure 9. Simulation section.

400 meters. The width of the hard shoulder is 3.5 meters, while the main lane road width is 3.75 meters.

To investigate the adaptability of the proposed methodology under different traffic flow conditions, traffic flow simulation data is categorized into four levels based on the service levels of the highway. Highway service levels are generally categorized into four types, corresponding to free flow, light congestion, heavy congestion, and severe congestion. Under different service levels, the proposed method is applied to optimize the HSR strategy to achieve maximum control benefits.

In accordance with the standards in highway service level criteria of China, the traffic flow data for these four service levels are defined as follows: 1400 vehicles per hour, 3000 vehicles per hour, 3750 vehicles per hour, and 4300 vehicles per hour. Furthermore, different vehicle types are also considered in SUMO simulation using actual traffic flow data collected in Jinan City. It contains three vehicle types: private, delivery, and trucks, which correspond to small, medium, and large vehicles, respectively. The ratio of these three categories of vehicles account for 75%, 12.5%, and 12.5%. Considering driver physiological response characteristics, the Widemann 99 and LC2013 models are selected to describe vehicle following and lane-changing behavior in SUMO simulations.

6. Analysis and discussion of findings

6.1. Results of training

Figure 10 illustrates the convergence results of the reward function in response to the training of the L-MADDPG algorithm under four service levels. In each level, the L-MADDPG algorithm, the MADDPG algorithm, no open (ZEROS), and all open (ALL) were tested as four distinct HSR control strategies. In Figure 10(a), the purple curve represents the convergence results achieved by the L-MADDPG algorithm, which exhibits higher reward convergence compared to the ZEROS strategy. For the second and third service levels, the L-MADDPG algorithm demonstrates higher convergence rewards compared to the MADDPG baseline algorithm and ZEROS, indicating that L-MADDPG becomes more effective in highway performance after congestion occurs. However, at the fourth service level, the difference in rewards among the four strategies become less pronounced. This finding suggests that the influence of adjusting the HSR strategy on highway traffic flow is constrained during periods of severe congestion. From the four graphs in Figure 10, it is evident that the strategy computed using the L-MADDPG algorithm consistently yield higher reward values compared to the ZEROS strategy.

6.2. Comparative analysis of various metrics

To further assess the impact of various algorithms on highway operation, this study conducted comparing tests of the L-MADDPG, MADDPG, ZEROS, and ALL methods across three aspects: efficiency, safety, and emissions. Regarding efficiency, the Total Travel Time was selected as a comparative metric. For safety evaluation, a composite safety indicator consisting of

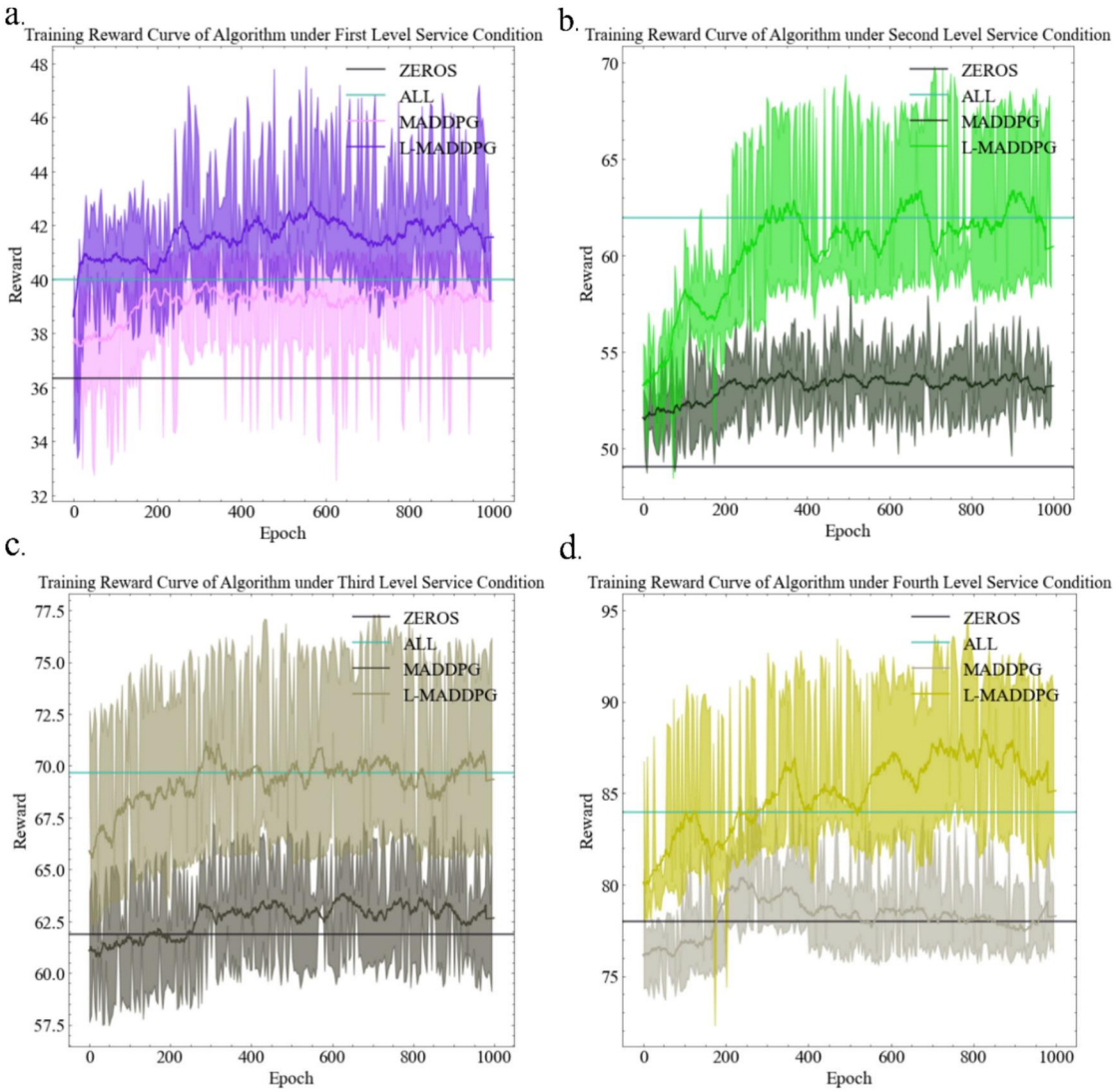


Figure 10. Training results of the algorithms.

TIT and TET was utilized to assess the safety level of highway over a specified period (Wu et al., 2020). In term of emissions, three major gases—CO, CO_2, and NO—were selected as crucial measures to assess the level of environmental emissions on the highway. Under the four service level conditions, each strategy was tested ten times, and the averages were taken as the final results to enhance the validity of the test outcomes.

6.2.1. Efficiency metrics

Figure 11 presents test results for total travel time across the four methods. Total travel time is a significant measure that reflects the duration during which highway lanes are occupied by vehicles, with shorter times indicating higher highway operational efficiency. It is evident that under all four service level conditions, the L-MADDPG algorithm consistently outperforms the other three strategies, highlighting the significant effectiveness of the proposed method in enhancing highway efficiency. Specifically, at the both second and thrid service level, the L-MADDPG algorithm achieves a reduction in total travel time of highway vehicle by 11.4 and 7.6 h, respectively, in comparison to the ZEROS strategy. From the first to the fourth service level, as vehicle density increases and congestion intensifies, the optimization space for vehicle travel time decreases.

6.2.2. Safety metrics

Figure 12 presents results of two safety indicators. The TIT value calculates the cumulative time during which vehicles on the highway are in a hazardous state, effectively reflecting the safety of highway traffic. From the figure, it is evident that as the service level improves, the cumulative duration of vehicles in

hazardous driving situations significantly increases with the higher volume on highway. This observation suggests that an increase in vehicle density and the time vehicles spend in hazardous driving conditions may lead to a higher frequency of accidents. Comparatively, during the implementation of the HSR strategy demonstrates, a substantial reduction in TIT in contrast to the ZEROS strategy can be observed, resulting in an obvious improvement of safety. Notably, at service level 4, the HSR strategy devised by the L-MADDPG algorithm leads to a significant 26.1% reduction in the TIT for all vehicles. While TIT primarily focuses on the duration of vehicles in hazardous driving conditions, the TET value reflects the frequency of highway vehicles below a certain TTC threshold. The comparison of the two highway safety indicators, TIT and TET, reveals that the L-MADDPG algorithm effectively reduces the traffic risk at different service levels, ensuring smooth traffic operations.

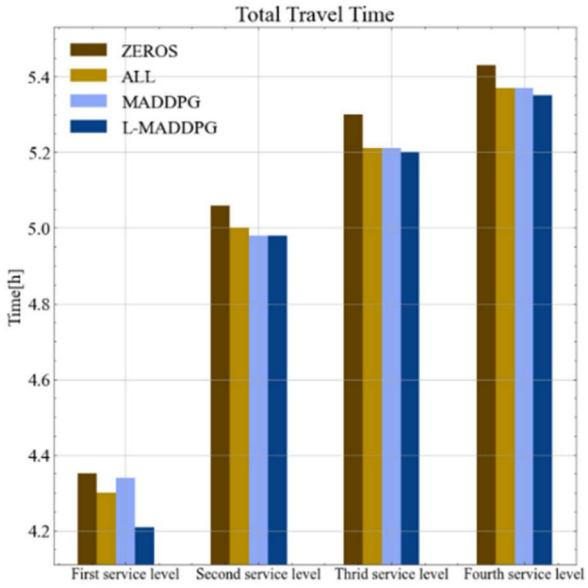
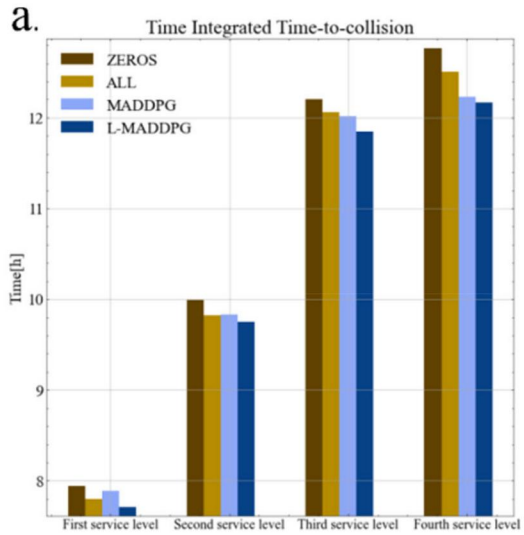


Figure 11. Comparison of efficiency across various algorithms.



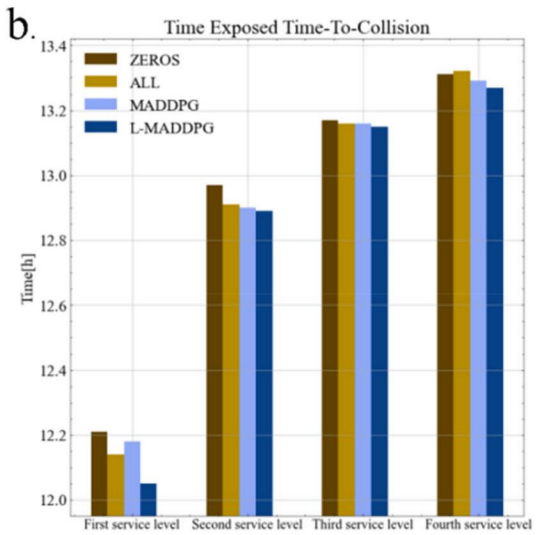


Figure 12. Comparison of safety across various algorithms.

6.2.3. Emission metrics

Figure 13 presents a comparison of the emissions of three gases: CO, CO_2, and NO. CO and NO are harmful gases, and their excessive emissions exacerbate environmental pollution, posing significant health risks. CO_2, as a greenhouse gas, contributes to rising atmospheric carbon dioxide levels, leading to phenomena like the greenhouse effect and global climate change. As traffic volumes increases, emissions from vehicles also grow. The implementation of HSR strategy on highways facilitates faster vehicle exit, resulting in reducing vehicular emissions. The application of the L-MADDPG algorithm to determine HSR strategies for different service levels has different impacts on vehicle emissions. At the first service level, compared to the ZEROS strategy, the implementation of the HSR strategy through L-MADDPG algorithm results in a significant reduction of CO emissions by 3625.35 Kg, CO_2 emissions by 461787.71 Kg, and NO emissions by 2159.68 Kg, effectively reducing pollution. At service level 2, the reduction in CO emissions is even more obvious, with a decrease of 599107.08 Kg. Notably, at the fourth service level, the NO emissions of L-MADDPG algorithm are lowered by 2440.63 Kg compared with the ZEROS strategy, demonstrating effectiveness of the L-MADDPG algorithm. Experimental findings demonstrate that optimizing hard shoulder strategies through the

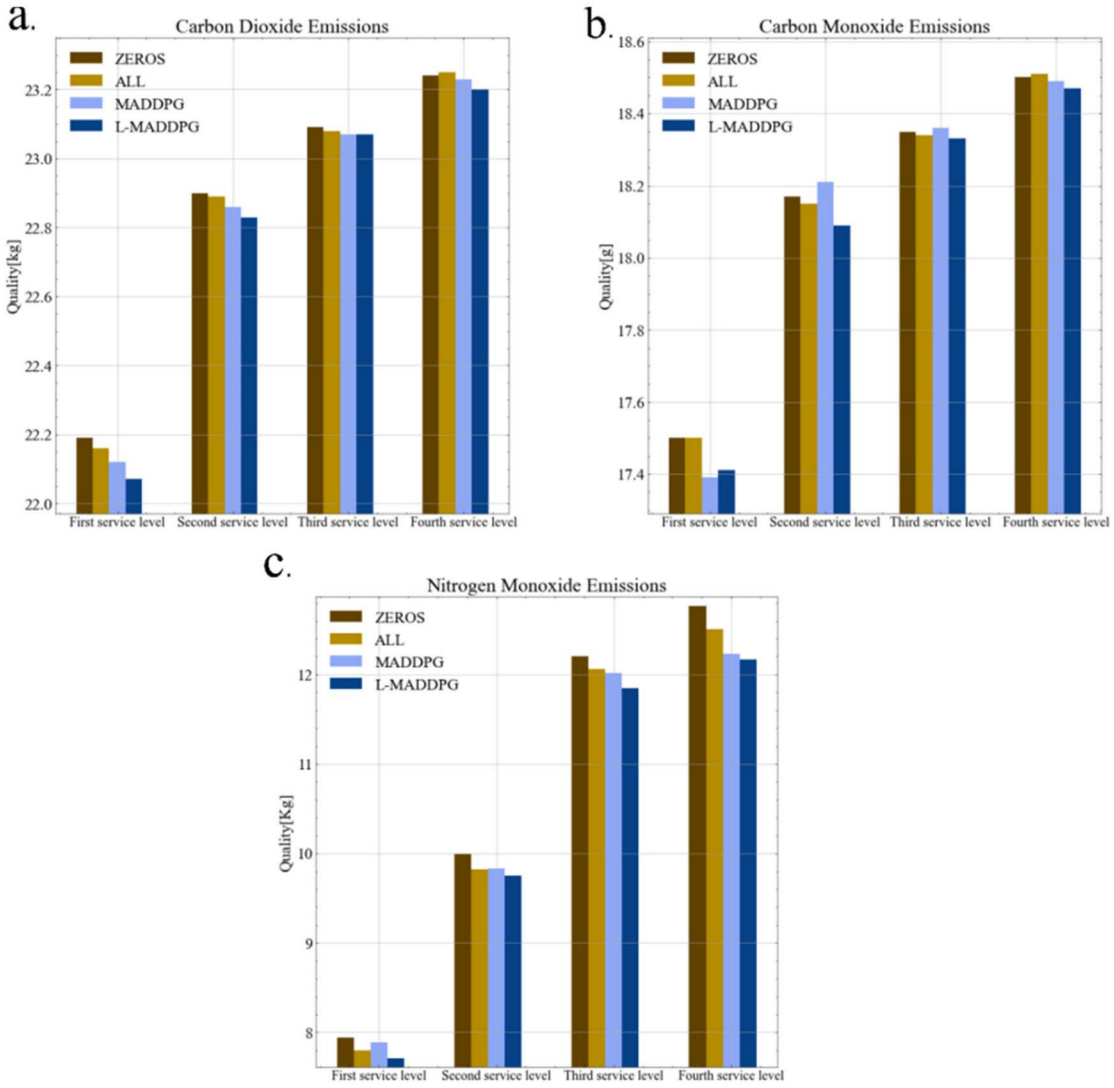


Figure 13. Comparison of emission across various algorithms.

L-MADDPG algorithm leads to a obvious reduction in emissions of three types of pollutants.

6.3. Emergency vehicle performance metrics

Implementation and optimization through the HSR strategy yields improvement for highway efficiency. Nevertheless, the hard shoulder, serving as a pathway to ensure smooth EVs travel on the highway, should consider its original function. Focusing on this, this study introduces an improved A* algorithm for clearing the hard shoulder lane to ensure access for EVs entering the regulated area. As in Figure 14, five strategies, ZEROS, ALL, ALL-A* algorithm, L-MADDPG algorithm and L-MADDPG-A* algorithm, are tested under four service level introduced before. Algorithm effectiveness is evaluated through the analysis of two key metrics related to EV performance: vehicle travel time and the braking times of EVs. In Figure 14(a), a comparison of EV vehicle travel times revealed a consistent reduction across all four service level conditions when using the A* algorithm for lane clearing. Notably, at service level 4, the L-MADDPG-A* algorithm reduced EV travel time by 18.1% compared to the ALL strategy. Both stopping and slowing down of vehicles significantly affect normal vehicle operation, and this study statistically compared the total stops or slowdowns of EVs during the experiment. In Figure 14(b), it was observed that the lane clearing through the improved A* algorithm results in a marked reduction in braking times of EVs. At the third service level, the L-MADDPG-A* algorithm reduced the braking times by 451.6%, thus greatly enhancing the smoothness of EVs operation. It is important to note that each optimal lane clearing solution incurs a certain clearing cost using the improved A* algorithm to clear lane. A larger clearance cost implies a greater number of vehicles that

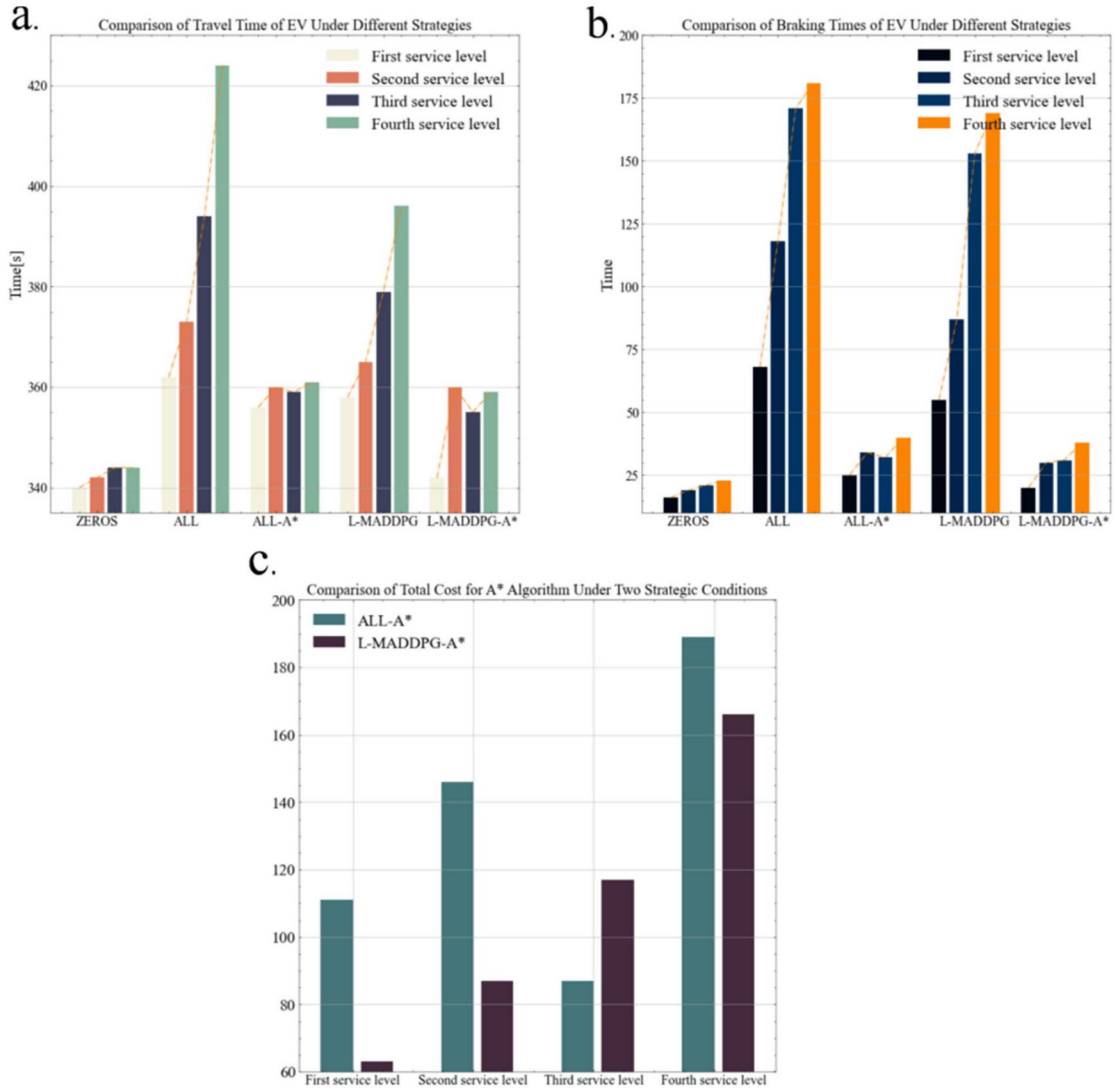


Figure 14. Comparison of EV performance across various algorithms.

need to be cleared. A comparison was made between the ALL-A* algorithm and the L-MADDPG-A* algorithm in terms of the total lane clearing cost for the entire EV travel process under four service level. The results show that as traffic volume increases and more vehicles transition from the main road to the hard shoulder lane, lane clearing becomes more challenging. Figure 14(c) illustrates that the ALL-A* algorithm incurs a higher total cost in all four service level conditions compared to the L-MADDPG algorithm, implying that the L-MADDPG-A* algorithm needs the clearance of fewer vehicles. This further demonstrates that the optimal HSR strategy obtained through the L-MADDPG-A* algorithm can reduce the likelihood of EV vehicles being disrupted by vehicles on the main road. In summary, the comparison of multiple indicators reveals the necessity of using the A* algorithm for hard shoulder lane clearing, which has significant implications for ensuring the smooth travel of EV vehicles.

6.4. Comparison result on the entire road segment

To investigate how the traffic changes on different lanes after EVs enter the regulated area, we selected lane average speeds for analysis. As shown in Figure 15, the speed variation of two main lanes and one hard shoulder lane was displayed under four service level. In Figure 15(a), at lower service levels with lower traffic volume, the various strategies resulted in similar average vehicle speeds. However, with increasing traffic volume, the lane average speed decreased, particularly under the ZEROS strategy. On the other hand, the ALL strategy enables smoother and faster vehicle travel, as the hard shoulder helps alleviate part of traffic burden. By optimizing the strategy through

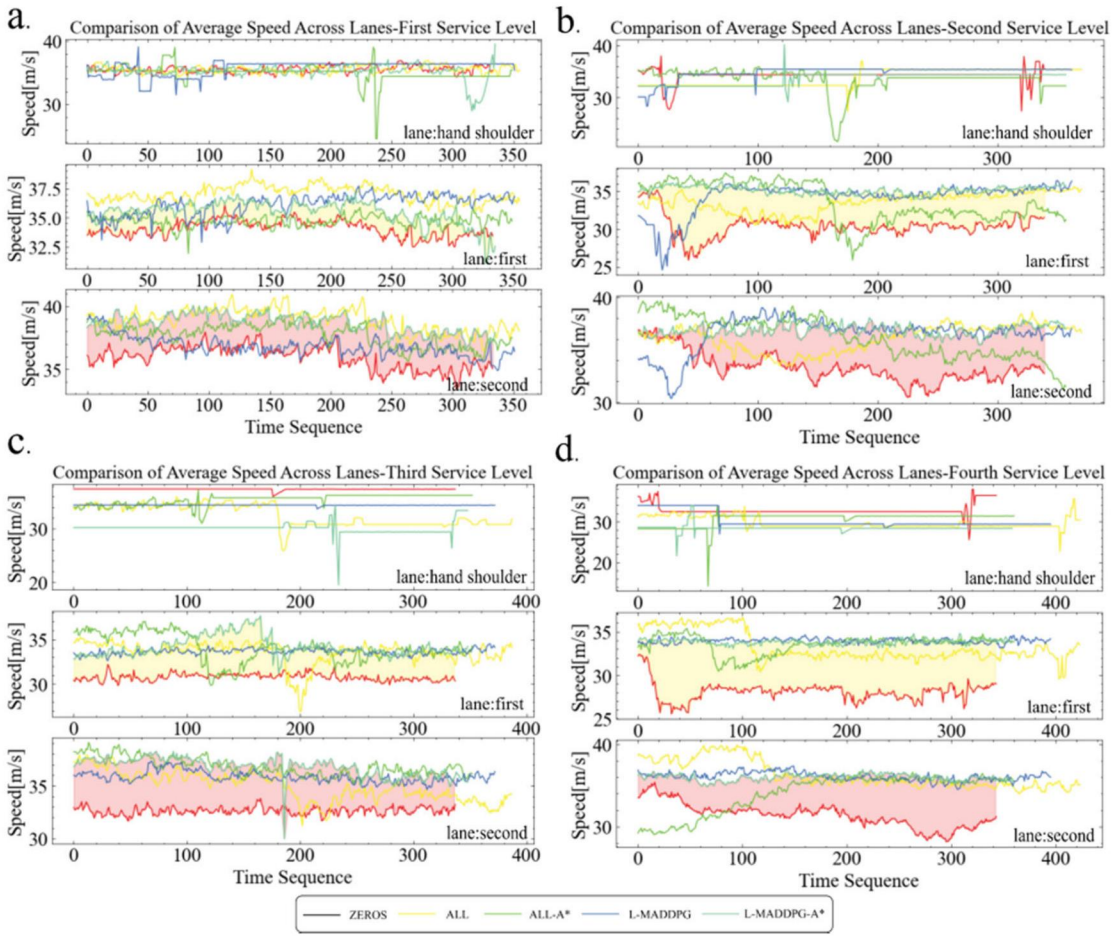


Figure 15. Comparison of average speed on different lanes.

the L-MADDPG-A* algorithm to open the hard shoulder, main road vehicles could enter the hard shoulder in the appropriate areas, reducing the number of vehicles on the main road. This led to higher lane average speeds compared to the ZEROS strategy. Notably, the yellow and red areas in Figure 15 represent the difference in average speed between the ZEROS strategy and the L-MADDPG-A* algorithm's optimized strategy. It is found that under the four service level conditions, the utilization of the L-MADDPG-A* algorithm has demonstrated an enhancement in the average road speed and a reduction in the time vehicles occupy the highway. This indicates that optimizing HSR strategies through the L-MADDPG-A* algorithm effectively enhances highway operational efficiency while ensures smooth traffic flow.

7. Conclusion

HSR strategies have been widely implemented in many countries, with numerous positive impacts on highway traffic operations. However, addressing how HSR strategies can positively affect highway efficiency, safety, and vehicle emissions while ensuring smooth passage for EVs remains a challenge. Based on the mathematical model of HSR strategy, we proposed a MADDPG algorithm combined with LSTM data feature extraction to optimize HSR strategies. To ensure unimpeded EV travel, an improved A* algorithm was employed to find the optimal lane clearing solution. The effectiveness of the proposed methods was validated through experiment using the SUMO simulation software. The experimental findings revealed the superior performance of the L-MADDPG algorithm in terms of training reward convergence across all four service levels when compared to other strategies. Moreover, an evaluation encompassing total travel time, TIT metrics, CO emissions and others demonstrated the effectiveness of the L-MADDPG algorithm. The application of the improved A* algorithm for lane clearing yields significant reductions in travel time of EVs, as well as braking times. These findings underscore the effectiveness of opening hard shoulders for highway traffic conditions across different service levels. It is worth mentioning that optimizing HSR

strategies requires consideration of EV entry into regulated area, and lane clearing through the algorithm can effectively guarantee the smooth passage of EVs.

Although this study has made progress in optimizing the HSR strategy, several limitations remain. First, the study does not account for external factors such as extreme weather conditions or construction activities, which could significantly impact traffic flow and, consequently, the effectiveness of the HSR strategy. Future research could incorporate additional external variables, such as weather variations and construction influences, to enhance the model's adaptability across diverse scenarios. Second, this study primarily focuses on the interaction between the HSR strategy and EVs, without adequately exploring the combined application of the HSR strategy with other active traffic management measures, such as variable speed limits and ramp metering. Future studies could investigate the integrated optimization of the HSR strategy alongside these measures to address complex and dynamic traffic scenarios, thereby improving the overall efficiency and safety of traffic management systems. Furthermore, future research could consider scenarios where all vehicles are intelligent, connected, and autonomous, capable of real-time, latency-free communication with one another. Under such conditions, the hard shoulder could be treated as a regular lane. From the perspective of autonomous vehicle control and lane-changing behavior, the smooth passage of emergency vehicles could be ensured through precise vehicle control mechanisms. In this scenario, the core challenge of optimizing the hard shoulder strategy would shift from traditional traffic management to the control of autonomous vehicles. This shift could further enhance the adaptability and dynamic adjustment capabilities of intelligent transportation systems. Such research would provide new perspectives and solutions for the integration of traffic management and autonomous driving technologies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by National Natural Science Foundation of China; The Key R&D Program of Hunan Province; Science Research Foundation of Hunan Provincial Department of Education.

References

Abdel-Aty, M., Hasan, T., & Anik, B. T. H. (2024). An advanced real-time crash prediction framework for combined hard shoulder running and variable speed limits system using transformer. *Scientific Reports*, *14*(1), 26403. https://doi.org/10.1038/s41598-024-75350-z

Aron, M., Cohen, S., & Seidowsky, R. (2010). *Two French Hard-Shoulder Running operations: Some comments on effectiveness and safety* [Paper presentation]. 13th International IEEE Conference on Intelligent Transportation Systems (pp. 230–236). IEEE.

Arora, K., & Kattan, L. (2023). Operational and safety impacts of integrated variable speed limit with dynamic hard shoulder running. *Journal of Intelligent Transportation Systems*, *27*(6), 769–798. https://doi.org/10.1080/15472450.2022.2078664

Awal, T., Kulik, L., & Ramamohanrao, K. (2013). *Optimal traffic merging strategy for communication-and sensor-enabled vehicles* [Paper presentation]. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) (pp. 1468–1474). IEEE.

Azimjonov, J., & Özmen, A. (2021). A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways. *Advanced Engineering Informatics*, *50*, 101393. https://doi.org/10.1016/j.aei.2021.101393

Balmer, M., Nagel, K., & Raney, B. (2004). *Large-scale multi-agent simulations for transportation applications* [Paper presentation]. Intelligent Transportation Systems (Vol. 8, pp. 205–221). Taylor & Francis Group.

Bhouri, N., Aron, M., Lebacque, J., & Haj-Salem, H. (2017). Effectiveness of travel time reliability indicators in the light of the assessment of dynamic managed lane strategy. *Journal of Intelligent Transportation Systems*, *21*(6), 492–506. https://doi.org/10.1080/15472450.2017.1327815

Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(2), 156–172. https://doi.org/10.1109/TSMCC.2007.913919

Cao, D., Wu, J., Wu, J., Kulcsár, B., & Qu, X. (2021). A platoon regulation algorithm to improve the traffic performance of highway work zones. *Computer-Aided Civil and Infrastructure Engineering*, *36*(7), 941–956. https://doi.org/10.1111/mice.12691

Chun, P., & Fontaine, M. D. (2017). Evaluation of operational effects of I-66 active traffic management system. *Transportation Research Record: Journal of the Transportation Research Board*, *2616*(1), 91–103. https://doi.org/10.3141/2616-10

Coffey, S., & Park, S. (2018). Operational evaluation of part-time shoulder use for Interstate 476 in the State of Pennsylvania. *Advances in Civil Engineering*, *2018*(1), 1724646. https://doi.org/10.1155/2018/1724646

Cohen, S., Aron, M., & Seidowsky, R. (2010, July 2010). *Assessment of a dynamic managed lanes operation* [Paper presentation]. 12th World Conference on Transport Research, Lisbon.

Ding, J., Li, L., Peng, H., & Zhang, Y. (2020). A rule-based cooperative merging strategy for connected and automated vehicles. *IEEE Transactions on Intelligent*

Transportation Systems, *21*(8), 3436–3446. https://doi.org/10.1109/TITS.2019.2928969

Dutta, N., Boateng, R. A., & Fontaine, M. D. (2019). Safety and operational effects of the interstate 66 active traffic management system. *Journal of Transportation Engineering, Part A: Systems*, *145*(3), 04018089. https://doi.org/10.1061/JTEPBS.0000220

Fukuyama, S. (2020). Dynamic game-based approach for optimizing merging vehicle trajectories using time-expanded decision diagram. *Transportation Research Part C: Emerging Technologies*, *120*, 102766. https://doi.org/10.1016/j.trc.2020.102766

Ghiasi, A., Hale, D., Bared, J., Kondyli, A., & Ma, J. (2018). *A dynamic signal control approach for integrated ramp and mainline metering* [Paper presentation]. 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 2892–2898). IEEE.

Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: A survey. *Artificial Intelligence Review*, *55*(2), 895–943. https://doi.org/10.1007/s10462-021-09996-w

Guerrieri, M., & Mauro, R. (2016). Capacity and safety analysis of hard-shoulder running (HSR). A motorway case study. *Transportation Research Part A: Policy and Practice*, *92*, 162–183. https://doi.org/10.1016/j.tra.2016.08.003

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). *Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor* [Paper presentation]. International Conference on Machine Learning (pp. 1861–1870). PMLR.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107. https://doi.org/10.1109/TSSC.1968.300136

Hochreiter, S. (1997). *Long short-term memory*. Neural Computation MIT-Press. https://doi.org/10.1162/neco.1997.9.8.1735

Hou, X., Gan, M., Zhang, J., Zhao, S., & Ji, Y. (2023). Secondary crash mitigation controller after rear-end collisions using reinforcement learning. *Advanced Engineering Informatics*, *58*, 102176. https://doi.org/10.1016/j.aei.2023.102176

Huang, Y., Zhou, C., Cui, K., & Lu, X. (2024). A multi-agent reinforcement learning framework for optimizing financial trading strategies based on timesnet. *Expert Systems with Applications*, *237*, 121502. https://doi.org/10.1016/j.eswa.2023.121502

Kim, Y., Kang, K., Park, N., Park, J., & Oh, C. (2025). Reinforcement learning approach to develop variable speed limit strategy using vehicle data and simulations. *Journal of Intelligent Transportation Systems*, *29*(3), 251–268. https://doi.org/10.1080/15472450.2024.2312808

Kononov, J., Hersey, S., Reeves, D., & Allery, B. K. (2012). Relationship between freeway flow parameters and safety and its implications for hard shoulder running. *Transportation Research Record: Journal of the Transportation Research Board*, *2280*(1), 10–17. https://doi.org/10.3141/2280-02

Li, R., Ye, Z., Li, B., & Zhan, X. (2017). Simulation of hard shoulder running combined with queue warning during traffic accident with ctm model. *IET Intelligent Transport Systems*, *11*(9), 553–560. https://doi.org/10.1049/iet-its.2016.0345

Li, R., Zhen, Y., & Li, B. (2019). Optimal control and simulation of hard shoulder running on highways. *Journal of System Simulation*, *30*(3), 1036–1045. https://doi.org/10.16182/j.issn1004731x.joss.201803034

Li, Y., Chow, A. H., & Cassel, D. L. (2014). Optimal control of motorways by ramp metering, variable speed limits, and hard-shoulder running. *Transportation Research Record: Journal of the Transportation Research Board*, *2470*(1), 122–130. https://doi.org/10.3141/2470–13

Li, Y., Pan, B., Xing, L., Yang, M., & Dai, J. (2022). Developing dynamic speed limit strategies for mixed traffic flow to reduce collision risks at freeway bottlenecks. *Accident; Analysis and Prevention*, *175*, 106781. https://doi.org/10.1016/j.aap.2022.106781

Low, V. J. M., Khoo, H. L., & Khoo, W. C. (2024). Robust dynamic real-time control strategies for high-frequency bus service: A multi-agent reinforcement learning framework. *Journal of Intelligent Transportation Systems*. Advance online publication. https://doi.org/10.1080/15472450.2024.2425293

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., P., Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. ArXiv, abs/1706.02275. https://doi.org/10.48550/arxiv.1706.02275.

Ma, J., Hu, J., Hale, D. K., & Bared, J. (2016). Dynamic hard shoulder running for traffic incident management. *Transportation Research Record: Journal of the Transportation Research Board*, *2554*(1), 120–128. https://doi.org/10.3141/2554-13

Mnih, V. (2013). *Playing atari with deep reinforcement learning*. arxiv preprint arxiv:1312.5602.

Mnih, V. (2016). *Asynchronous methods for deep reinforcement learning*. arxiv preprint arxiv:1602.01783.

Mu, C., Du, L., & Zhao, X. (2021). Event triggered rolling horizon based systematical trajectory planning for merging platoons at mainline-ramp intersection. *Transportation Research Part C: Emerging Technologies*, *125*, 103006. https://doi.org/10.1016/j.trc.2021.103006

Newgard, C. D., Schmicker, R. H., Hedges, J. R., Trickett, J. P., Davis, D. P., Bulger, E. M., Aufderheide, T. P., Minei, J. P., Hata, J. S., Gubler, K. D., Brown, T. B., Yelle, J.-D., Bardarson, B., & Nichol, G, Resuscitation Outcomes Consortium Investigators. (2010). Emergency medical services intervals and survival in trauma: Assessment of the "golden hour" in a North American prospective cohort. *Annals of Emergency Medicine*, *55*(3), 235–246.e4. https://doi.org/10.1016/j.annemergmed.2009.07.024

Papageorgiou, M., Papamichail, I., Spiliopoulou, A. D., & Lentzakis, A. F. (2008). Real-time merging traffic control with applications to toll plaza and work zone management. *Transportation Research Part C: Emerging Technologies*, *16*(5), 535–553. https://doi.org/10.1016/j.trc.2007.11.002

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). *Prioritized experience replay*. arXiv preprint arXiv:1511.05952,

Seet, B. C., Liu, G., Lee, B. S., Foh, C. H., Wong, K. J., & Lee, K. K. (2004, May 9–14). *A-STAR: A mobile ad hoc*

routing strategy for metropolis vehicular communications [Paper presentation]. Networking 2004: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Third International IFIP-TC6 Networking Conference Athens, Greece. Proceedings 3 (pp. 989–999). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24693-0_81

Spatharis, C., & Blekas, K. (2024). Multiagent reinforcement learning for autonomous driving in traffic zones with unsignalized intersections. *Journal of Intelligent Transportation Systems*, *28*(1), 103–119. https://doi.org/10.1080/15472450.2022.2109416

Vadde, R., Sun, D., Sai, J. O., Faruqi, M. A., & Leelani, P. T. (2012). A simulation study of using active traffic management strategies on congested freeways. *Journal of Modern Transportation*, *20*(3), 178–184. https://doi.org/10.3969/j.issn.2095087X.2012.03.008

Van Hasselt, H., Guez, A., & Silver, D. (2016). *Deep reinforcement learning with double q-learning* [Paper presentation]. Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30). https://doi.org/10.1609/aaai.v30i1.10295

Waleczek, H., & Geistefeldt, J. (2021). Long-term safety analysis of hard shoulder running on freeways in Germany. *Transportation Research Record: Journal of the Transportation Research Board*, *2675*(8), 345–354. https://doi.org/10.1177/0361198121997836

Wilson, M. (2009). Hard shoulder running eases motorway traffic jams. *Highways*, *v78*(n1), p12–3. http://worldcat.org/oclc/3831968

Wu, J., Kulcsár, B., Ahn, S., & Qu, X. (2020). Emergency vehicle lane pre-clearing: From microscopic cooperation to routing decision making. *Transportation Research Part B: Methodological*, *141*, 223–239. https://doi.org/10.1016/j.trb.2020.09.011

Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., & Wang, S. (2022). Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *Journal of Hydrology*, *608*, 127553. https://doi.org/10.1016/j.jhydrol.2022.127553

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, *169*, 114513. https://doi.org/10.1016/j.eswa.2020.114513

Yang, F., Wang, F., Ding, F., Tan, H., & Ran, B. (2021). Identify optimal traffic condition and speed limit for hard shoulder running strategy. *Sustainability*, *13*(4), 1822. https://doi.org/10.3390/su13041822

Yao, J., Qian, Y., Feng, Z., Zhang, J., Zhang, H., Chen, T., & Meng, S. (2024). Hidden markov model-based dynamic hard shoulders running strategy in hybrid network environments. *Applied Sciences*, *14*(8), 3145. https://doi.org/10.3390/app14083145

Zeng, H., & Schrock, S. D. (2012). Estimation of safety effectiveness of composite shoulders on rural two-lane highways. *Transportation Research Record: Journal of the Transportation Research Board*, *2279*(1), 99–107. https://doi.org/10.3141/2279-12

Zhang, Z., Rui, P., & Qiao, J. (2022). *Analysis of Traffic Characteristics of Dynamic Open Hard Shoulder Road Based on SUMO* [Paper presentation]. 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI) (pp. 100–104). IEEE.

Zhi, Y., Zhang, Z., Zhou, W., Hou, D., & Zhang, J. (2024). Evaluation of mixed traffic flow efficiency and safety on hard-shoulder-running freeways. *Applied Sciences*, *14*(23), 11137. https://doi.org/10.3390/app142311137

Zhou, W., Yang, M., Lee, M., & Zhang, L. (2020). Q-learning-based coordinated variable speed limit and hard shoulder running control strategy to reduce travel time at freeway corridor. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(11), 915–925. https://doi.org/10.1177/0361198120949875