Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 4419907, 19 pages
https://doi.org/10.1155/2023/4419907

Research Article

Coordinated Variable Speed Limit Control for Consecutive Bottlenecks on Freeways Using Multiagent Reinforcement Learning

Si Zheng[1], **Meng Li**[1], **Zemian Ke**[2] and **Zhibin Li**[1]

[1]*School of Transportation, Southeast University, Nanjing 211189, China*
[2]*Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh 15213, USA*

Correspondence should be addressed to Meng Li; limeng@seu.edu.cn

Received 25 February 2023; Revised 10 June 2023; Accepted 17 June 2023; Published 27 June 2023

Academic Editor: Jinjun Tang

Most of the current variable speed limit (VSL) strategies are designed to alleviate congestion in relatively short freeway segments with a single bottleneck. However, in reality, consecutive bottlenecks can occur simultaneously due to the merging flow from multiple ramps. In such situations, the existing strategies use multiple VSL controllers that operate independently, without considering the traffic flow interactions and speed limit differences. In this research, we introduced a multiagent reinforcement learning-based VSL (MARL-VSL) approach to enhance collaboration among VSL controllers. The MARL-VSL approach employed a centralized training with decentralized execution structure to achieve a joint optimal solution for a series of VSL controllers. The consecutive bottleneck scenarios were simulated in the modified cell transmission model to validate the effectiveness of the proposed strategy. An independent single-agent reinforcement learning-based VSL (ISARL-VSL) and a feedback-based VSL (feedback-VSL) were also applied for comparison. Time-varying heterogeneous traffic flow stemming from the mainline and ramps was loaded into the freeway network. The results demonstrated that the proposed MARL-VSL achieved superior performance compared to the baseline methods. The proposed approach reduced the total time spent by the vehicles by 18.01% and 17.07% in static and dynamic traffic scenarios, respectively. The control actions of the MARL-VSL were more appropriate in maintaining a smooth freeway traffic flow due to its superior collaboration performance. More specifically, the MARL-VSL significantly improved the average driving speed and speed homogeneity across the entire freeway.

1. Introduction

Over the past few decades, traffic congestion has increasingly become a significant problem on freeways [1, 2]. Congestion typically arises in the vicinity of freeway bottlenecks and may propagate in both upstream and downstream directions. With the availability of massive traffic data from motorways, intelligent transportation systems can be implemented practically as traffic control measures to improve the traffic flow [3–5]. The variable speed limit (VSL) control has become a valid method for mitigating jams at freeway bottlenecks and enhancing traffic operations [6, 7].

The previous studies have mainly concentrated on the VSL control methods for shorter freeway segments that have a single bottleneck [8–13]. However, in certain scenarios, multiple bottlenecks may occur due to various reasons, such as consecutive uncontrolled on-ramps, work zones, or lane drops. The previous studies made the assumption that VSL control operations targeting different bottleneck locations could be independently managed without mutual interference. However, in cases where potential bottlenecks are closely located, coordinated VSL strategies that consider the interaction of traffic flow at different bottlenecks should be implemented. By working cooperatively, VSL controllers of road segments can achieve better harmonization effects.

Numerous VSL control methods have been devised specifically to address the issue of multiple bottlenecks [14–18]. For example, Iordanidou et al. introduced a VSL control method for multiple bottlenecks, where they utilized feedback from key bottlenecks to develop a comprehensive

VSL strategy [14]. Lu et al. proposed a feedback control algorithm based on the velocity and extended its application to freeway networks with multiple bottlenecks [15]. The two feedback-based strategies showed efficiency and robustness under practical traffic conditions. However, the current feedback controllers experience time delays between the occurrence of deviations in controlled variables and the implementation of corrective measures, thus affecting their response to continuously changing traffic conditions. To address these limitations, Yu and Fan proposed an online optimization approach to address the issue of the coordinated VSL control [16]. The online optimization approach facilitated precise modeling-based prediction of traffic flow changes. A multiobjective nonlinear integer model was formulated, and a genetic method was developed to optimize the VSL objective function. However, such models may not always be available.

Recently, reinforcement learning (RL) approaches have been explored as potential solutions to overcome the limitations of online optimization and feedback controllers. The RL has proven to be an effective solution for addressing the VSL problem, thanks to its powerful capability in forecasting traffic evolution and implementing proactive control strategies [7, 19–25]. Li et al. reported that VSL based on the Q-learning algorithm yielded superior results compared to the feedback-based algorithms [21]. The proposed Q-learning VSL control strategy demonstrated reductions in the system travel time of 49.34% and 21.84% under different traffic demand scenarios. Later, in this paper [23], a deep reinforcement learning (DRL) method was introduced utilizing an actor-critic architecture to implement lane-specific VSL control on multilane freeways. The results showed that the proposed control approach significantly enhanced both safety and efficiency on the freeway.

Most existing RL-based VSL controllers are primarily designed for shorter freeway segments, necessitating the expansion of their control strategies to enable the speed limit control on longer segments. However, training only one RL agent to generate the joint actions of all VSL controllers may lead to a significant increase in the number of actions and states, which can make the RL convergence intractable. Moreover, training multiple independent single-agent RL models to generate separate VSL controller values could result in each agent receiving misleading reward signals due to unobserved actions taken by their teammates. This leads to a "nonstationary" problem, as each agent's environment is only partially observed [26, 27]. To the best of our knowledge, only a few scholars have utilized a multiagent RL for VSL control. For instance, Kušić et al. proposed a VSL using the multiagent distributed W-learning method. Each individual agent utilized the RL to acquire knowledge of both local and long-range policies, aiming to understand the impact of its actions on the neighboring agents [17, 18]. However, this method was severely constrained by the discretized traffic states and Q-table's storage limit. Later, Wang et al. proposed a cooperative VSL control approach using the distributed Q-learning, in which each agent could observe others' actions and improve the collective reward [28]. However, this algorithm had to assume that each agent optimistically considered that all other agents would act to maximize their reward, which often resulted in poor performance in complex or stochastic environments.

In this study, we developed a centralized training approach with decentralized execution for VSL control in multiple bottlenecks. It is a model-free actor-critic algorithm that uses deep neural networks to learn a policy and a value function [29]. In this strategy, the VSL controllers act as learning agents to cooperate with each other to achieve their individual goals. The centralized critic network helps to address the nonstationarity in a multiple bottlenecks environment caused by the dynamic policy changes of other agents during training. This approach uses a decentralized execution strategy, where each agent uses its local policy to adjust the speed limit values in the environment based on its own observations. Similar structures have been applied in multiple signal control in urban environments with high-dimensional and continuous state space [30–32]. However, unlike signalized control in an urban environment, the effects of control actions propagate quickly along the driving direction due to high speeds and volume on the freeway. Therefore, a careful design of an effective control system and reward functions is necessary to dampen the traffic flow instability and potential traffic oscillations.

The paper is organized as follows: in Section 2, we formulate the VSL control problem in multiple consecutive bottlenecks and introduce the proposed MARL-VSL scheme; in Section 3, we present the simulation model and experiment setups; in Section 4, the performances of different VSL strategies are compared; in Section 5, we test the MARL-VSL scheme in diverse capacity drop scenarios; and in Section 6, we introduce the final conclusions and future directions.

2. Methodology

This section describes the use of MARL as an approach to tackle the issue of cooperation among multiple VSL controllers. Firstly, we introduce the problem of formulating the VSL control within the context of MARL. Next, we propose an MARL-VSL scheme for the consecutive bottlenecks scenario. Finally, we provide an overview of the ISARL-VSL and feedback-VSL methods to be compared with the proposed scheme.

2.1. RL Formulation. A typical VSL controller for a merge bottleneck comprises of the following two primary components: an upstream segment controlled by VSL and a downstream acceleration segment [5, 10]. By modifying the traffic flow entering downstream, the VSL control mechanism can effectively mitigate or eliminate capacity drops at the bottleneck. Considering the VSL controller as an intelligent agent, where each agent corresponds to a specific road segment and is responsible for setting a different speed limit for that segment. The variable speed limit signs are used to display the speed limit. The number of intelligent agents equals the number of road segments, N. Thus, in scenarios featuring continuous bottlenecks, the optimization problem

of VSL can be represented as an N-agent interacting Markov Decision Process (MDP) (see Figure 1). This formulation assumes a set of state, denoted by $s = \{s_1, \ldots, s_N\}$, with one state assigned to each agent in the environment. Additionally, it involves a set of actions, denoted by $a = \{a_1, \ldots, a_N\}$, where each a_i represents the possible actions for the i^{th} agent, and a set of rewards, denoted by $r = \{r_1, \ldots, r_N\}$, with one reward assigned to each agent.

2.1.1. Action. The MARL-VSL and ISARL-VSL control strategies employ VSL to regulate the traffic flow through adjustments of the posted speed limit within the controlled area. Hence, the speed limit is regarded as the action in this context. In practical applications, VSL signs permit the display of only specific speed limit values. Therefore, this study focuses on speed limit values within the action set that are integer multiples of 5 mph. The action set includes speeds of {5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, and 65} mph.

2.1.2. State. At time t, the i^{th} agent can observe the state s_i of the corresponding freeway segment. The traffic data obtained from road measurements can reflect the traffic flow state. In this research study, the state of each agent consisted of six components (d_i^{up}, d_i^{ramp}, k_i^{down}, k_i^{up}, k_i^{ramp}, and SL_i^t). Five traffic measurements were collected for the i^{th} agent to represent the following variables: d_i^{up} represents the traffic demand on the mainline, d_i^{ramp} represents the traffic demand on the ramp, k_i^{down} and k_i^{up} represent the density of different sections of the freeway mainline, and k_i^{ramp} represents the density on the ramp (see Figure 1). Additionally, the i^{th} agent's action SL_i^t at the last time step t was included to prevent significant changes in speed limits between the consecutive time steps.

2.1.3. Reward. Given the intricate spatiotemporal structure of traffic data, it is necessary to decompose the rewards in space and ensure their measurability after the actions are taken. Controlling bottleneck density has been a primary target in various VSL strategies [33–36]. Consistent with prior research, the reward in this research is determined based on the density measured at bottlenecks [19, 21]. As the density approaches the critical density value, the reward increases. The maximum reward is attained when the density matches its critical value (refer to Figure 2). The i^{th} agent's reward r_i is calculated as follows:

$$r_i = \begin{cases} c * d_i, \text{if } d_i < d_c, \\ c * d_c - c * (d_i - d_c), \text{if } d_i \geq d_c, \end{cases} \quad (1)$$

where d_i denoted the downstream density (veh/mile/lane), d_c denoted the critical density (26.75 veh/mile/lane), and the parameter c was set as 0.02. To motivate the agents to rapidly approach the critical density, an additional reward of 0.6 was introduced for values in close proximity to the critical density, specifically, $25.5 \leq d_i \leq 28$. Conversely, −0.6 is applied to the reward value for severe congestion states with $d_i \geq 50$. To ensure safety, it is important to minimize abrupt changes in the speed limits between consecutive time steps, and −0.1 is applied to the reward value when the difference in speed limits exceeds 10 mph. These parameters were meticulously determined through multiple simulation rounds.

2.2. MARL-VSL Strategy. In a cooperative multiagent RL task, multiple agents interact with the environment and with each other. When N independent agents are used to control N VSL signs, the agents may receive misleading reward signals due to the unobserved actions of their teammates, resulting in nonstationary issues [26].

To address this challenge, several MARL algorithms have been developed, with the multiagent deep deterministic policy gradient (MADDPG) method being the most widely used [29, 37]. The MADDPG algorithm uses a structure of centralized training with decentralized execution. Each agent is constructed using an actor-critic architecture. The actor network generates the policy and provides corresponding speed limits according to the real-time traffic state. The critic network evaluates the policy of the agent according to the state values and actions of all agents, i.e., real-time traffic flow states and speed limits for all road sections. During the execution phase, following the completion of training, only local actors are utilized to operate in a decentralized fashion, which is well-suited for cooperative environments. If the actions of other agents are already known, the environment becomes stationary and its transitions become stable. Here, distributed actors generate each control action separately, while a central critic can make accurate evaluations and predictions for all agents' state transitions and reward values. The authentic assessment of the reward signal makes joint optimization feasible, which inspires us to utilize the MADDPG algorithm for VSL controls in the consecutive bottlenecks on the freeway (see Figure 1).

Specifically, the MADDPG algorithm starts with setting up an experience replay buffer $\mathcal{D}$ for each agent i. The replay buffer $\mathcal{D}$ contains the control experiences of all agents, denoted as $(s_1, s_2, \ldots, s_N, SL_1, SL_2, \ldots, SL_N, r_1, r_2, \ldots, r_N)$. Given the states series and the actions series, we can obtain a relatively deterministic reward series $\{r_1, r_2, \ldots, r_N\}$ and the subsequent observations s'. Therefore, MADDPG can guide policy learning by constructing an effective critic network for reward estimation, making it suitable for cooperative multiagent control tasks, such as multiple VSL control.

Agents interact with environments and store the control experiences in replay buffer $\mathcal{D}$ for neural network updating. Let $\mu_\theta = \{\mu_{\theta_1}, \ldots, \mu_{\theta_N}\}$ represent the RL policy for all agents with parameter $\theta = \{\theta_1, \ldots, \theta_N\}$. In order to update the actor network, the gradient of each agent's expected reward is calculated in equation (2) according to [29].

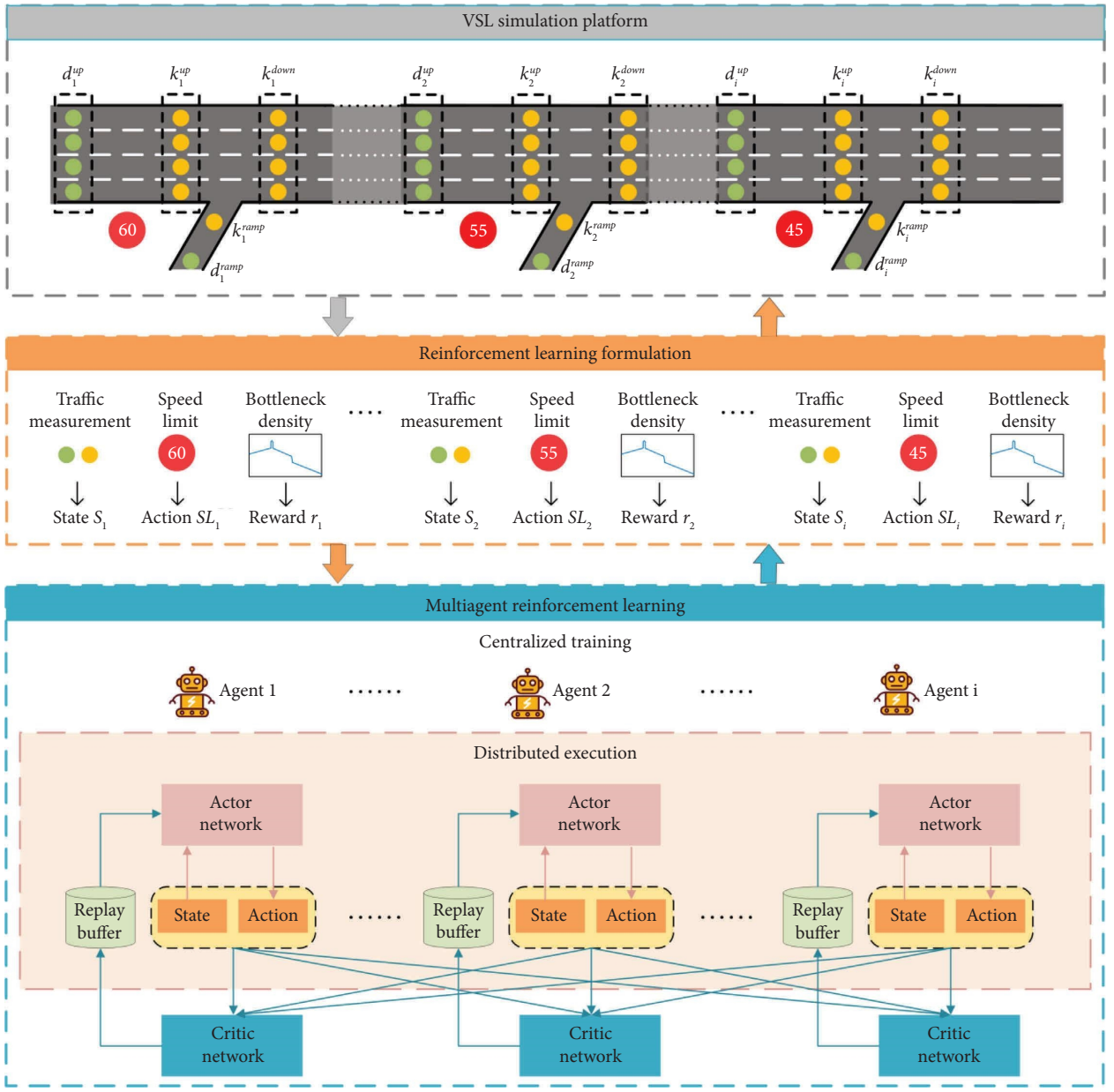

Figure 1: Overall research framework.

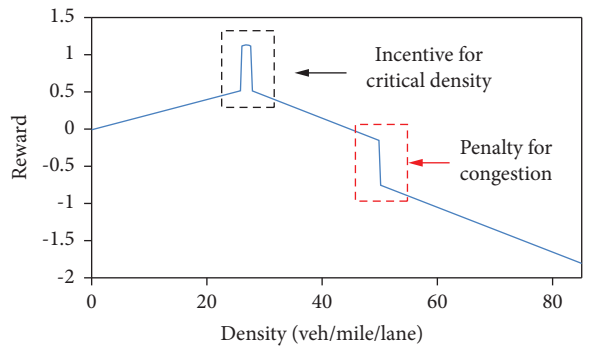

Figure 2: Reward function.

$$\nabla_{\theta_i} J\left(\mu_{\theta_i}\right) = \mathbb{E}_{\mathbf{s},\mathbf{a}\sim\mathscr{D}}\left[\nabla_{\theta_i}\mu_{\theta_i}\left(a_i|s_i\right) \times \nabla_{a_i} Q_i^{\mu}\left(s_1, s_2, \ldots, s_N, SL_1, SL_2, \ldots, SL_N\right)\Big|_{a_i=\mu_{\theta_i}(s_i)}\right], \tag{2}$$

where $Q_i^{\mu}(s_1, s_2, \ldots, s_N, SL_1, SL_2, \ldots, SL_N)$ is the centralized Q function with the inputs of all the agents' actions and the observations.

The critic network is updated by minimizing the loss according to the temporal difference (TD) error as follows:

$$\begin{aligned} \mathscr{L}(\theta_i) &= \mathbb{E}_{\mathbf{s},\mathbf{s}',\mathbf{a},\mathbf{r}}\left[\left(Q_i^{\mu}\left(s_1, s_2, \ldots, s_N, SL_1, SL_2, \ldots, SL_N\right) - y\right)^2\right], \\ y &= r_i + \gamma Q_i^{\mu'}\left(s_1', s_2', \ldots, s_N', SL_1', SL_2', \ldots, SL_N'\right)\Big|_{a_i'=\mu'(s_i')}, \end{aligned} \tag{3}$$

where a_i' is the action of the i^{th} agent at the next time step, γ is the discount factor, and μ' represents the set of target policies.

Within the framework of centralized Q function learning and decentralized policy execution, N agents can cooperatively learn the optimal policy. After sufficient training, the neural networks will converge, enabling each agent to generate the optimal VSL actions with the maximal Q value (see Figure 3). In our experiment, we employed four agents to generate four VSL control actions, denoted as $\{SL_1, SL_2, SL_3, SL_4\}$, on a freeway with four bottlenecks. Upon executing the actions, we acquire a reward series $\{r_1, r_2, r_3, r_4\}$ and new states $\{s_1', s_2', s_3', s_4'\}$.

2.3. VSL Baselines. The MARL-VSL proposed in this paper is based on a density measurement. To ensure that the experimental results are fair and unbiased, any alternative VSL control strategies should also be based on the density measurement.

2.3.1. ISARL-VSL Strategy. This study used the ISARL-VSL strategy as a baseline for comparison because of its simple structure. This strategy employed a deep deterministic policy gradient (DDPG) algorithm [38], and each agent learned an optimal function independently. Multiple agents operated independently in parallel, with each agent attempting to maximize its own reward without any communication or coordination between them. In this case, each agent learned from its own experiences and independently updated its policy or value function to optimize its own reward. Similar to the MARL-VSL approach, the ISARL-VSL agent perceived a one-dimensional vector with six components (d_i^{up}, d_i^{ramp}, k_i^{down}, k_i^{up}, k_i^{ramp}, and SL_i^t), took actions to control the mainline speed, received rewards measured from loop detectors, and updated the network. Notably, the state, actions, and reward settings utilized in the ISARL-VSL strategy were consistent with those of the MARL-VSL algorithm.

2.3.2. Feedback-VSL Strategy. A commonly used VSL strategy in practical engineering is the Feedback-VSL strategy, which shows efficiency and robustness under practical traffic conditions. In the current research studies, the density measurement-based feedback-VSL strategy is implemented locally and lacks cooperative capabilities. Therefore, in scenarios involving multiple bottlenecks, we adopted the typical multiple independent local feedback VSL controllers as a baseline for comparison [33].

The feedback VSL controller was described by the following equation:

$$\begin{aligned} s(k) &= s(k-1) + K_I e_f(k), \\ \widehat{f}(k) &= \widehat{f}(k-1) + \left(K_P' + K_I'\right)e_\rho(k) - K_P' e_\rho(k-1), \end{aligned} \tag{4}$$

where K_I, K_P', and K_I' are the controller parameters, $e_f(k)$ represents control error of the flow, $e_\rho(k)$ represents control error of the density, $s(k)$ represents the speed limit rate calculated by dividing the speed limit value by the free-flow velocity, and $\widehat{f}(k)$ represents the adjusted outflow.

The simulation and modeling testing employed the following parameters: $K_I = 0.0007$ mile·lane/veh, $K_I' = 5.0$ mile·lane/hour, and $K_P' = 50.0$ mile·lane/hour. The speed limit was rounded to the nearest 5 mph. More details can be found in reference [33].

3. Simulation Design

3.1. Development of Simulation Platform. A macroscopic simulation platform was developed in this research, utilizing the cell transmission model (CTM). Our research primarily examined the macroscopic characteristics of freeways without delving into the microscopic driving behaviors of individual vehicles [39–41]. Additionally, our experiments involved extensive iterative computations, and decisions regarding VSL control were made using the loop detector data. The computational cost of the microscopic simulation platform is much higher compared to that of CTM. Therefore, CTM is deemed more suitable for our research purposes.

In CTM, a road is depicted as a sequence of contiguous cells. Every cell corresponds to the distance covered by a vehicle traveling at 65 mph within a single time step. A fundamental diagram is utilized to approximate the traffic conditions in each cell. The initial parameters of CTM have been fine-tuned to accurately simulate the effects of capacity drops and VSL strategies.

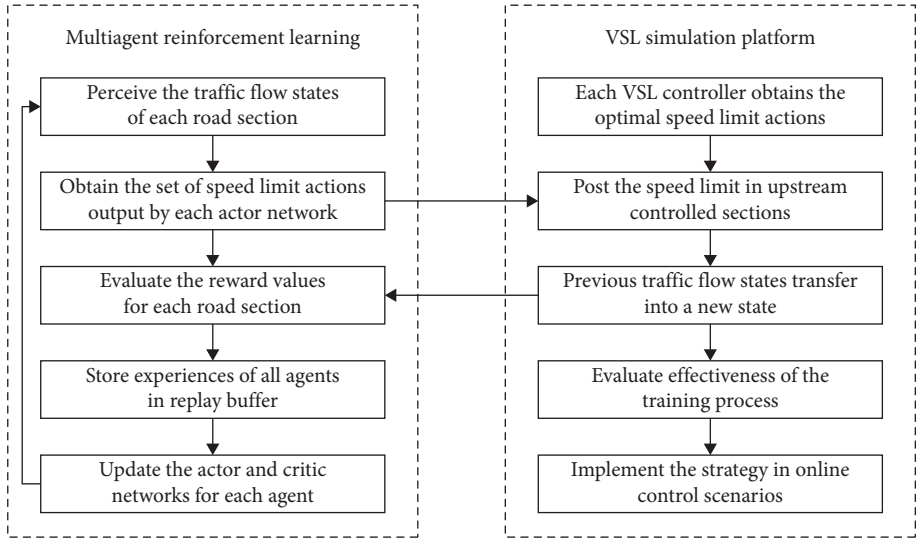

Figure 3: Flowchart of the MARL-VSL strategy.

The fundamental diagram utilized in this research is illustrated in Figure 4(a). The left side represents the outflow rate $\sigma_i(k)$ of the cell i, and the right side represents the inflow rate $\delta_i(k)$ of the cell i. The expressions for the outflow rate and inflow rate are given by the following equations:

$$\begin{aligned} \sigma_i(k) &= \min\{\min\{V_{SL}(k), V_F\} \cdot d_i(k) \cdot n_i, Q_{\text{VSL}}\}, \\ \delta_i(k) &= \min\{w_i \cdot (d_{i,\text{jam}} - d_i(k)) \cdot n_i, Q_{\text{VSL}}\}, \end{aligned} \tag{5}$$

where k represents the time step, $d_i(k)$ represents the vehicle density, n_i represents the number of freeway lanes, Q_{VSL} represents the maximum outflow under the VSL control, w_i represents the kinematic wave speed, V_{SL} represents the values of the VSL control, V_F represents the free-flow speed, and $d_{i,\text{jam}}$ represents the congestion density.

The phenomenon of capacity drop occurs when congestion happens [2, 42]. A reversed lambda-shaped fundamental diagram was used to describe the variation in the bottleneck cell flow (refer to Figure 4(b)). The outflow rate of the bottleneck cell with a capacity drop is defined by the following equation:

$$\sigma_i(k) = \begin{cases} V_F \cdot d_i(k) \cdot n_i, & \text{if } d_i(k) \le d_C, \\ Q_D, & \text{if } d_i(k) > d_C, \end{cases} \tag{6}$$

where $Q_C = V_F \cdot d_C \cdot n_i$ represents the bottleneck capacity before the capacity drop (veh/hour), Q_D represents the maximum discharge flow after the capacity drop (veh/hour), and d_C represents the critical density.

3.2. Experiment Setup. This study focused on a 4.2-mile segment of freeway, which comprises four merging on-ramps (see Figure 5). The segment was divided into four regulated sections, and the speed limit for each section was determined by a controller. There were four controllers in total. The study solely considered eastbound traffic, with the four single-lane on-ramps positioned at intervals of 0.7, 1.7, 2.7, and 3.7 miles from the beginning of the segment. The mainline featured four lanes, and the segment was divided into 42 cells, each with a length of 0.1 mile. The MARL-VSL and ISARL-VSL strategies were integrated with the CTM, and the speed limit was determined at each control period by the agents. Then, the traffic data from the simulated measurements were sent to the agents for training and decision-making. After undergoing the training process, the MARL-VSL strategy was implemented in online control scenarios. Two distinct traffic demand scenarios were employed for evaluation purposes, namely, a static traffic scenario and a dynamic traffic scenario. We primarily evaluated simulation results based on the following metrics: traffic flow operations (including volume, density, and speed) and total time spent (TTS).

The training quality directly affected the performance of the algorithm. During training, the MARL-VSL and ISARL-VSL strategies interact with the modified CTM. The training process is divided into distinct episodes [43]. Each episode comprises 122 control cycles. Different cycle lengths ranging from 30 seconds to 10 minutes were tested, revealing a higher likelihood of convergence for the MARL-VSL and ISARL-VSL control policies when the control cycle was set to 30 seconds. A dynamic traffic scenario was used to ensure that the agents could gain sufficient experience under different traffic conditions. The entire training process lasted approximately 326 minutes, and after the training, the time to compute the optimal VSL at a time step was less than 0.03 seconds.

The parameters were carefully determined in our research based on the preliminary testing and recommendations from previous studies [21, 29, 44–47]. The parameters are shown in Table 1. The discount factor is a parameter that specifies the relative weight assigned to current and past rewards in a reinforcement learning context. The learning rate parameter regulates the speed at which Q-values are updated in the learning process. The batch size determines

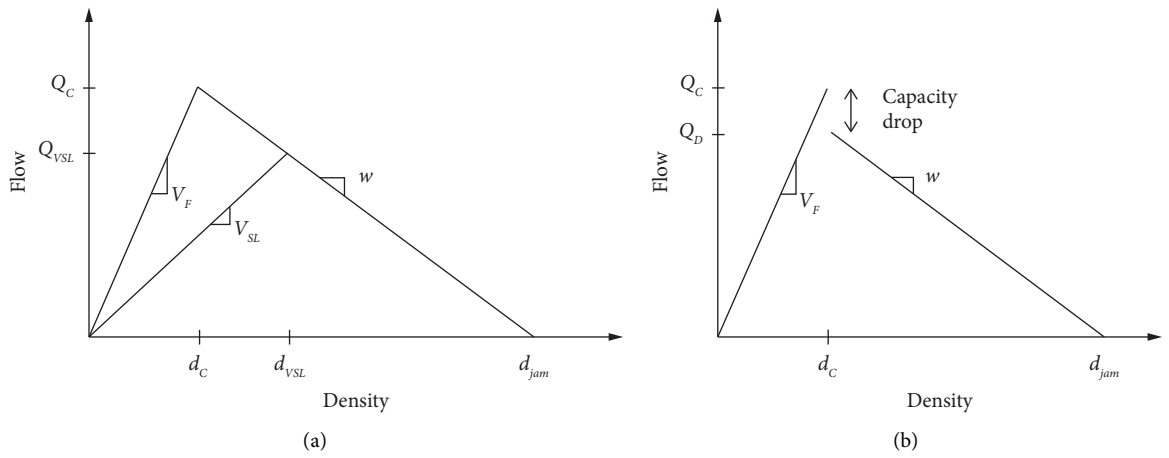

Figure 4: (a) The fundamental diagram of the CTM. (b) The fundamental diagram depicting capacity drop.

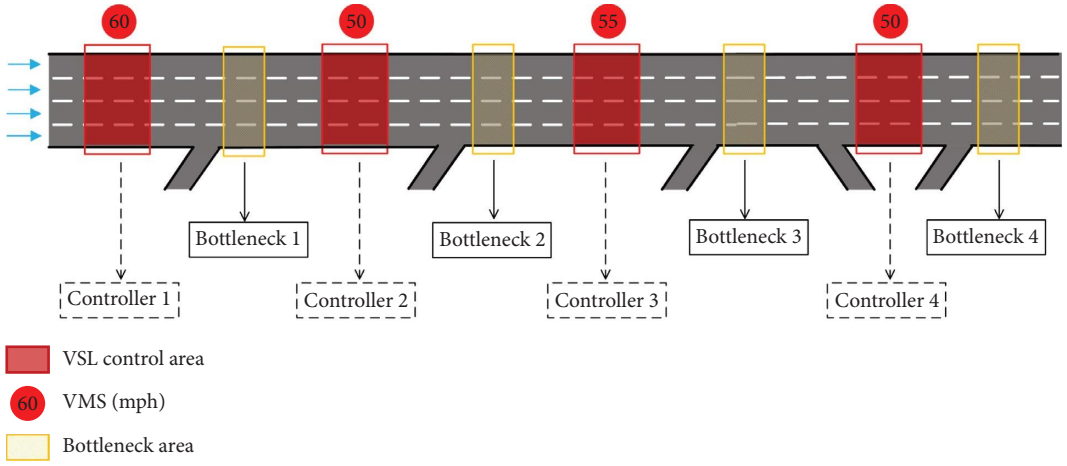

Figure 5: A freeway stretch with consecutive bottlenecks.

Table 1: Parameter values of the experiment.

Parameters	Value
Free-flow speed	65 mph
Critical density	26.75 veh/mile/lane
Capacity of the freeway mainline	1750 veh/hour/lane
Magnitude of the capacity drop	7.6%
Cycle of VSL control	30 s
Discount factor	0.95
Size of the replay buffer	1000000
Batch size	1024
Learning rate	0.001
Frequency of updating network	1 time/2 episodes

the number of state-action pairs used to update the neural network parameters in each iteration of the learning process.

To evaluate the effectiveness of the training, we plotted the reward and the TTS curves. The reduction of freeway congestion can lead to a significant decrease in TTS [48].

In a freeway system, traffic variables are discretely represented with a time index, k, and a time interval, η. The TTS within the time range K can be calculated according to the following equation [48]:

$$\text{TTS} = \eta \sum_{k=1}^{K} N(k), \tag{7}$$

where $N(k)$ is the total number of vehicles in the freeway system at time k. The total reward received by all agents at each episode is shown in Figure 6(a), while the rewards received by each individual agent at each episode are shown in Figures 6(b) and 6(c). After the fluctuation in the first 2,860 episodes, the total reward of MARL agents increased significantly and reached a maximum around the 3,592nd episode, reflecting that the MARL agents learned an optimal policy to maximize the reward through appropriate training. We estimated the performance of the MARL-VSL controllers by analyzing the TTS of each episode (see

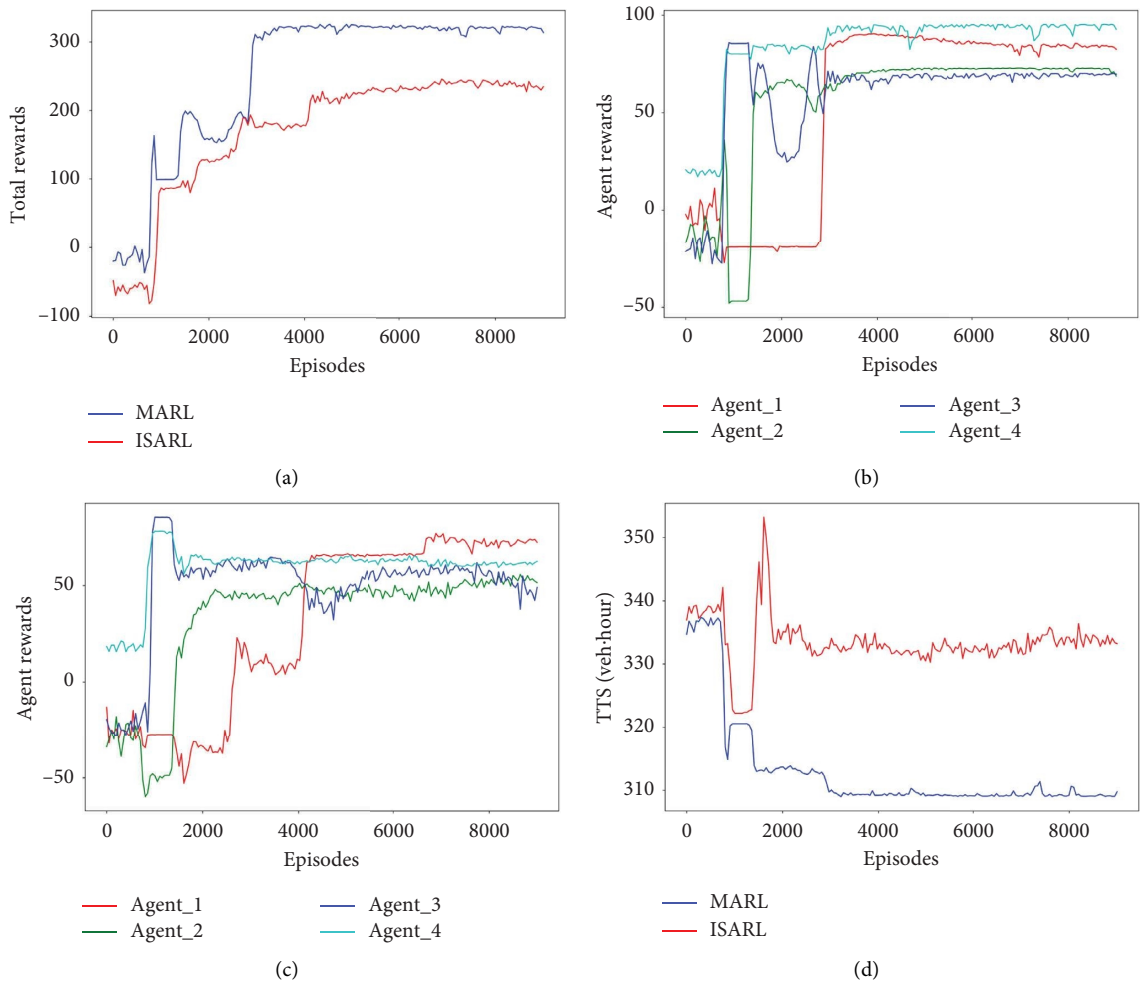

Figure 6: (a) Total reward during training. (b) Reward of each individual agent during MARL-VSL training. (c) Reward of each individual agent during ISARL-VSL training. (d) TTS during training.

Figure 6(d)). The TTS was relatively high in the first 2,860 episodes and decreased significantly later, indicating the effectiveness of the MARL-VSL strategy. However, the ISARL-VSL did not converge during training, and its total reward was substantially lower than that of MARL. Furthermore, the TTS of the ISARL-VSL was significantly higher than that of the MARL-VSL, indicating that the ISARL-VSL was unable to obtain the global optimum.

4. Evaluation of VSL Strategies

After the training process described in the previous section, the learned MARL-VSL strategy was evaluated in two different traffic demand scenarios, namely, a static traffic scenario (see Figures 7(a) and 7(c)–7(f)) and a dynamic traffic scenario (Figures 7(b) and 7(g)–7(j)). The static traffic scenario had a duration of 1 hour, with a 15-minute warm-up period. The purpose of this scenario was to confirm the stability of the VSL control method when operating in different traffic regimes, particularly in cases where traffic remains stable with low deviations in demand. In the dynamic scenario, traffic demand on both the mainline and ramp underwent significant variations over time. As with the training process, the simulation duration for both scenarios was set to be 1 hour, and the VSL control cycle was set to be 30 seconds.

4.1. Effect of VSL Strategies on Traffic Operation. The mainline traffic flow operations were reflected in the evolution of traffic density (see Figure 8). For comparison, we also considered three base cases as follows: feedback-VSL, ISARL-VSL, and no control (No-VSL). In the static traffic scenario, obvious congestion occurred at the bottleneck areas 1 and 3. The MARL-VSL strategy held vehicles near the VSL control zone, ensuring that the density near the downstream bottleneck approached the critical density (approximately 26.75 veh/mile/lane) to optimize the outflow. With the implementation of ISARL-VSL, the congestion at bottleneck area 1 was alleviated, but a congestion problem emerged in bottleneck area 2. This was attributed to an inadequate control action taken by Controller 2.

In the dynamic traffic scenario, the MARL-VSL approach still achieved a good control effect. Implementation of ISARL-VSL helped to ease the congestion in

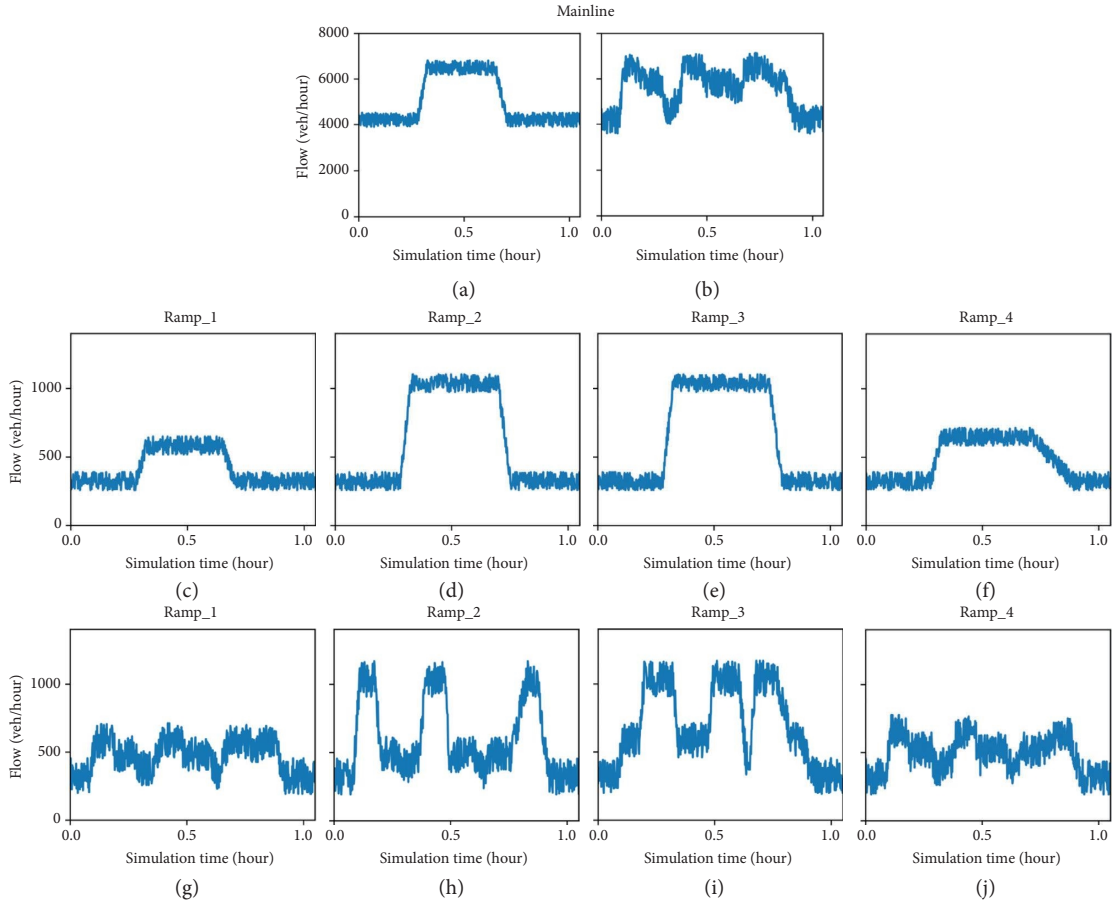

FIGURE 7: (a, c–f) The traffic demand for the mainline and ramps in the static traffic scenario. (b, g–j) The traffic demand for the mainline and ramps in the dynamic traffic scenario.

bottleneck areas 1 and 2, although its efficacy in bottleneck area 3 was suboptimal. The feedback-VSL strategy partially alleviated congestion, but it also led to the occurrence of high-density sections. It is noteworthy that, while Feedback Controller 3 was able to alleviate congestion in bottleneck area 3, a new congestion wave emerged in the speed-limited zone 3 and propagated upstream, ultimately increasing the density of bottleneck area 2 (see the black box in Figure 8(d)). This was attributed to the feedback controller's failure to incorporate information from other upstream controllers when generating control actions.

Overall, under the control of the MARL-VSL, the congestion of the entire road network had been eliminated. The ISARL-VSL and feedback-VSL methods were able to learn effective strategies to some degree, but many vehicles remained trapped in high-density regions (over 50 veh/mile/lane), accompanied by traffic oscillations propagating along the freeway. The differences in control effectiveness arose from MARL-VSL's ability to learn precise actions and collaborate to mitigate congestion effects between multiple bottlenecks.

To explore the fluctuations in traffic flow characteristics at four freeway bottlenecks, we presented the changes in three primary elements, influenced by the different VSL strategies (see Figures 9 and 10).

The previous analysis demonstrated that the lack of VSL implementation led to reduced speed and increased density in bottleneck areas caused by traffic congestion. Capacity reduction occurred shortly after congestion onset and persisted consistently throughout the entire duration. In the static traffic scenario, both the feedback-VSL and the ISARL-VSL attempted to mitigate the effects of traffic congestion but still led to a noticeable capacity drop (see the yellow boxes in Figure 9). On the contrary, there were only slight speed fluctuations with the MARL-VSL in four bottlenecks. Moreover, average speeds at all bottlenecks remained consistently above 50 mph, indicating smoother driving behaviors with reduced unnecessary acceleration and deceleration and resulting in fewer traffic oscillations. Consequently, our proposed strategy led to more stable density and flow curves.

It can be observed that the dynamic traffic scenario resulted in greater speed fluctuations (see Figure 10). Nevertheless, the MARL-VSL strategy was able to maintain relatively high speeds (above 40 mph). In the dynamic traffic scenario, our proposed strategy was able to swiftly facilitate traffic recovery to a low-density state. However, the other control strategies exhibited significant fluctuations in flow curves, revealing the presence of interference between bottlenecks.

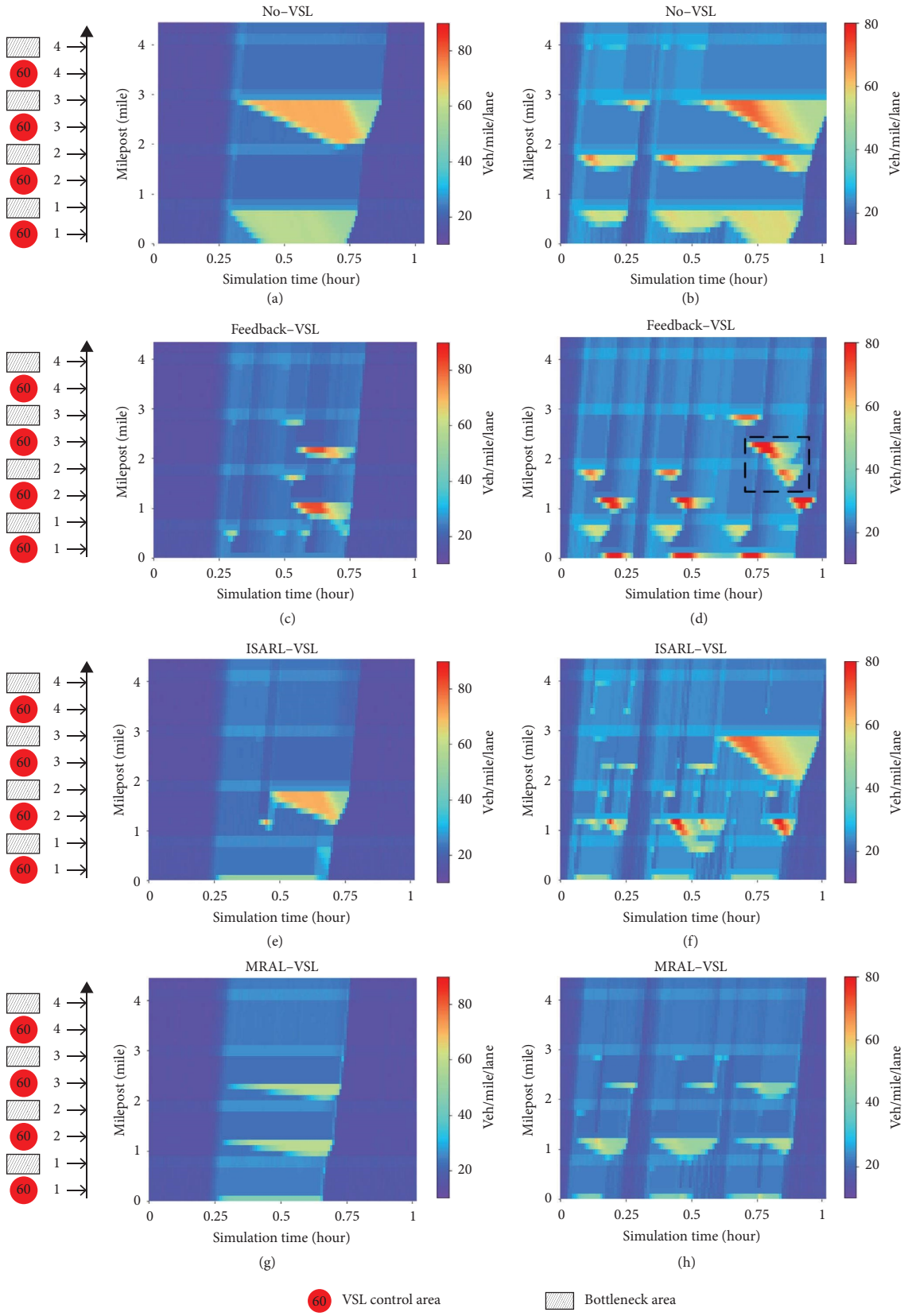

FIGURE 8: (a, c, e, g) The evolution of traffic density in the static traffic scenario. (b, d, f, h) The evolution of traffic density in the dynamic traffic scenario.

In both traffic scenarios, our proposed MARL-VSL strategy resulted in significantly higher average speeds than other control strategies, while also exhibiting smaller standard deviation of speeds (see Table 2). Especially under dynamic traffic demand, the effect of the MARL-VSL was obviously better than that of the other two control strategies. Specifically, the MARL-VSL reduced the speed standard deviation by 45.03%, whereas the other two control strategies only reduced it by 22.85% (ISARL-VSL) and 17.03% (feedback-VSL). Possibly due to the lack of coordination among controllers, the congestion waves formed by different bottlenecks would propagate upstream and exacerbate traffic flow disturbances. The results demonstrated that the MARL-VSL could effectively improve the average driving speed and the speed homogeneity of the entire freeway, thereby improving traffic efficiency.

4.2. Effect of VSL Strategies on TTS. In this section, we calculated vehicles' TTS on the freeway with different VSL strategies. The simulations were subjected to a repetition of ten trials, and subsequently, the outcomes were subjected to a process of averaging prior to their presentation (see Figures 11(a) and 11(b)). In the static traffic scenario, the TTS value was the same for all cases in the initial stage when traffic congestion had not formed. After 0.3 hour, four curves diverged. The cases with no control reached the largest TTS (over 4 veh·hour) at around 0.7 hour. The feedback-VSL and the ISARL-VSL reduced TTS from 0.4 hour to 0.8 hour. However, the TTS value with the two control strategies was always larger than our proposed strategy over the simulation period. Hence, the MARL-VSL outperformed the existing methods, consistent with the previous results obtained by analyzing the traffic density evolution (see Figures 8–10). The same results were shown for the dynamic traffic scenario (see Figure 11(b)). We noticed that the feedback-VSL control presented a higher TTS value than the No-VSL around 0.3 hour. The reason might be that the feedback-VSL generated incompatible speed limit values, resulting in worse traffic operations.

We calculated the statistical results of TTS (see Table 3). As anticipated, the MARL-VSL strategy yielded the lowest TTS value (235.19 veh·hour) in the static traffic scenario, resulting in an 18.01% reduction in TTS. In comparison, the feedback-VSL and ISARL-VSL strategies reduced the TTS by 8.02% and 13.82%, respectively. In the dynamic traffic scenario, the MARL-VSL strategy achieved a 17.07% reduction in TTS, while the feedback-VSL and ISARL-VSL strategies reduced TTS by 6.08% and 8.30%, respectively. It is noteworthy that the MARL-VSL strategy demonstrated a significant reduction in TTS compared to the other VSL strategies in the dynamic traffic scenario. Hence, our proposed control strategy was more effective in a more complex flow-demand scenario. This was because the traffic flow interference phenomenon between different bottlenecks was more evident in a complex environment. However, our proposed strategy could consider the traffic flow information of the entire road section and take cooperative control actions, which further reduced the interference effect of traffic flow.

4.3. Discussion on Differences between VSL Actions. To explore the reasons why the MARL-VSL outperformed the other two compared strategies, the speed limits generated by different strategies were plotted (see Figures 12 and 13). The ISARL-VSL and feedback-VSL strategies were used for comparison. When congestion was forming, the MARL-VSL reduced the speed limits quickly to different extents at different bottlenecks to prevent capacity drops. It tried to keep the speed limit values at a fixed value with less fluctuation to provide stable speed limit signals for drivers.

In the static traffic scenario, looking at the red offsets in Figure 12, it was evident that the MARL-VSL strategy started to reduce the speed limit by 30 mph at 0.3 hour for Controller 1, Controller 2, and Controller 3. Subsequently, the three controllers maintained a stable speed limit. However, controllers of the other VSL strategies did not cooperate well, leading to speed limit fluctuations (see red boxes in Figure 12).

In the dynamic traffic scenario, the MARL-VSL strategy again demonstrated effective coordination among controllers. Specifically, Controller 1 reduced the speed to 35 mph at the 5^{th} minute, Controller 2 reduced the speed to 25 mph at the 7^{th} minute, and Controller 3 reduced the speed to 25 mph at the 11^{th} minute (see the red offsets in Figure 13). The MARL-VSL strategy enabled the controllers to limit the speed synchronously, making traffic speed evolution smoother. Conversely, other control strategies experienced difficulties in finding a suitable control action for Controller 3, resulting in a suboptimal control effect on the road (see red boxes in Figure 13). The MARL-VSL strategy showcased strong coordination, as each controller considered the speed limit values of other controllers when taking actions. Overall, well-trained MARL-VSL controllers cooperated effectively in generating optimal VSL values for mainline traffic.

5. Testing of MARL-VSL Strategy in Diverse Capacity Drop Scenarios

The extent of capacity drop at merged bottlenecks can vary across different time periods or dates [2]. A greater degree of capacity drop results in higher vehicle densities and lower discharge flow rates. Therefore, evaluating the effectiveness of the MARL-VSL control approach under various capacity drop scenarios is valuable. In our research, we set the capacity drop levels at 5%, 10%, 12.5%, and 15%, respectively (refer to Table 4). The fundamental diagram in the CTM was adjusted based on the studies of the authors mentioned in references [8, 49].

According to the findings presented in Table 4, the MARL-VSL strategy demonstrated TTS values in different traffic scenarios. In the static traffic scenario, TTS values of the MARL-VSL approach ranged from 235.57 to 237.25 veh·hour when the capacity drops were set at 5%, 10%, and 12.5%. However, a noticeable increase in TTS was observed with a larger capacity drop of 15%, reaching 261.17 veh·hour. Similarly, in the dynamic traffic scenario, TTS values of the MARL-VSL approach remained relatively stable within a range of 264.42 to 267.74 veh·hour as the

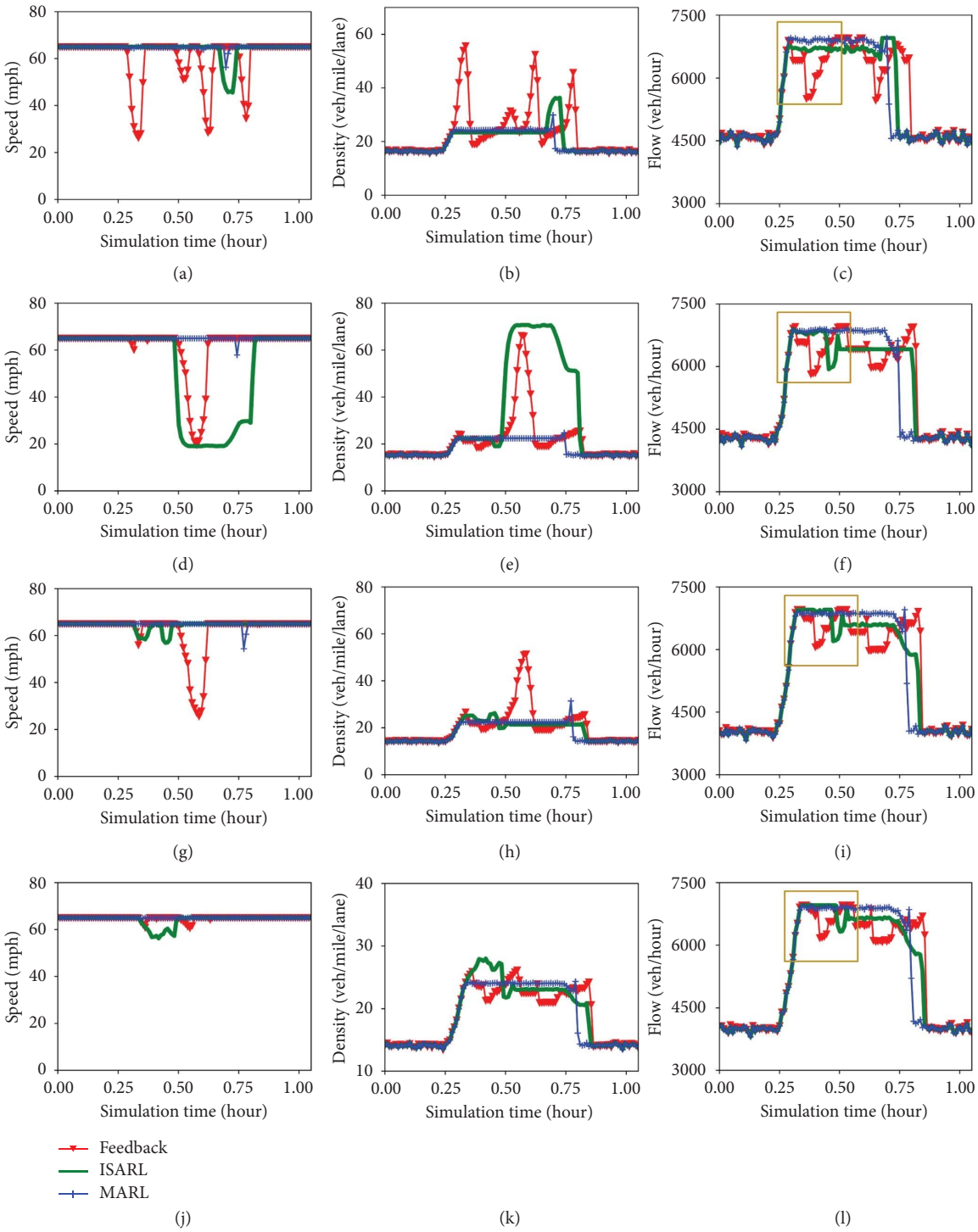

Figure 9: Traffic operations at bottlenecks 1–4 in the static traffic scenario. (a–c) Bottleneck 1, (d–f) bottleneck 2, (g–i) bottleneck 3, and (j–l) bottleneck 4.

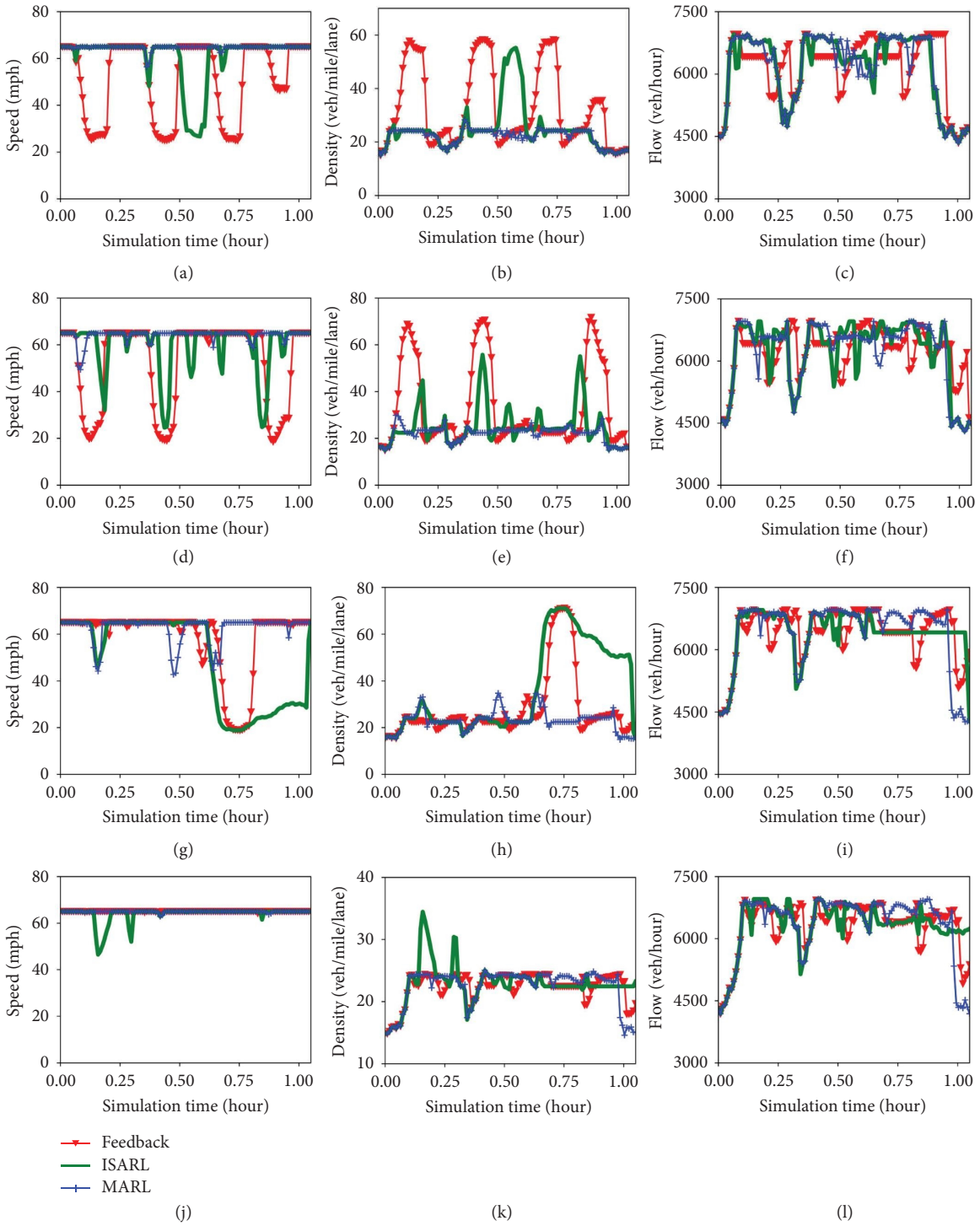

Figure 10: Traffic operations at bottlenecks 1–4 in the dynamic traffic scenario. (a–c) Bottleneck 1, (d–f) bottleneck 2, (g–i) bottleneck 3, and (j–l) bottleneck 4.

Table 2: Average speed and the speed standard deviation with different control strategies.

Scenario	Control strategy	Average speed (mile/hour)	Improvement (%)	Speed standard deviation (mile/hour)	Reduction (%)
Static traffic scenario	No-VSL control	51.77	None	8.70	None
	Feedback	57.31	+10.70	6.44	−25.97
	ISARL	57.95	+11.93	5.97	−31.37
	MARL (ours)	**59.71**	**+15.33**	**5.49**	**−36.89**
Dynamic traffic scenario	No-VSL control	50.43	None	11.86	None
	Feedback	54.31	+7.69	9.84	−17.03
	ISARL	55.11	+9.28	9.15	−22.85
	MARL (ours)	**59.31**	**+17.61**	**6.52**	**−45.03**

The bold numbers represent the results of our proposed MARL algorithm. The bold formatting is solely for emphasis.

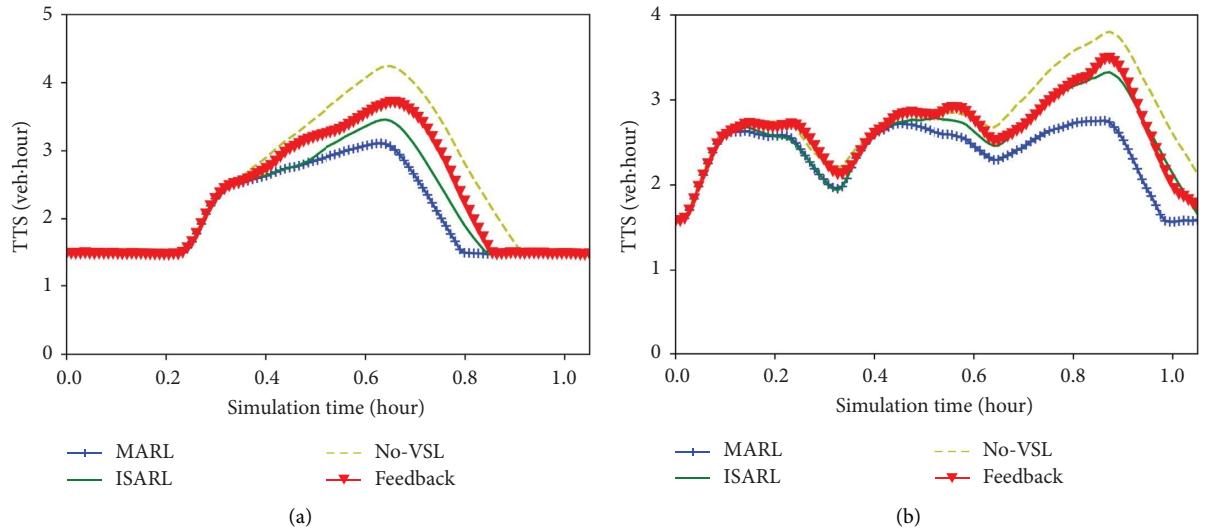

Figure 11: (a) Total time spent with different control conditions in the static traffic scenario; (b) total time spent with different control conditions in the dynamic traffic scenario.

Table 3: Total time spent with different control strategies.

Scenario	Control strategy	Total time spent (veh·hour)	Reduction (%)
Static traffic scenario	No-VSL control	286.89	None
	Feedback	263.87	−8.02
	ISARL	247.23	−13.82
	MARL (ours)	**235.19**	**−18.01**
Dynamic traffic scenario	No-VSL control	319.97	None
	Feedback	300.50	−6.08
	ISARL	293.39	−8.30
	MARL (ours)	**265.33**	**−17.07**

The bold text and numbers represent the results of our proposed MARL algorithm. The bold formatting is solely for emphasis.

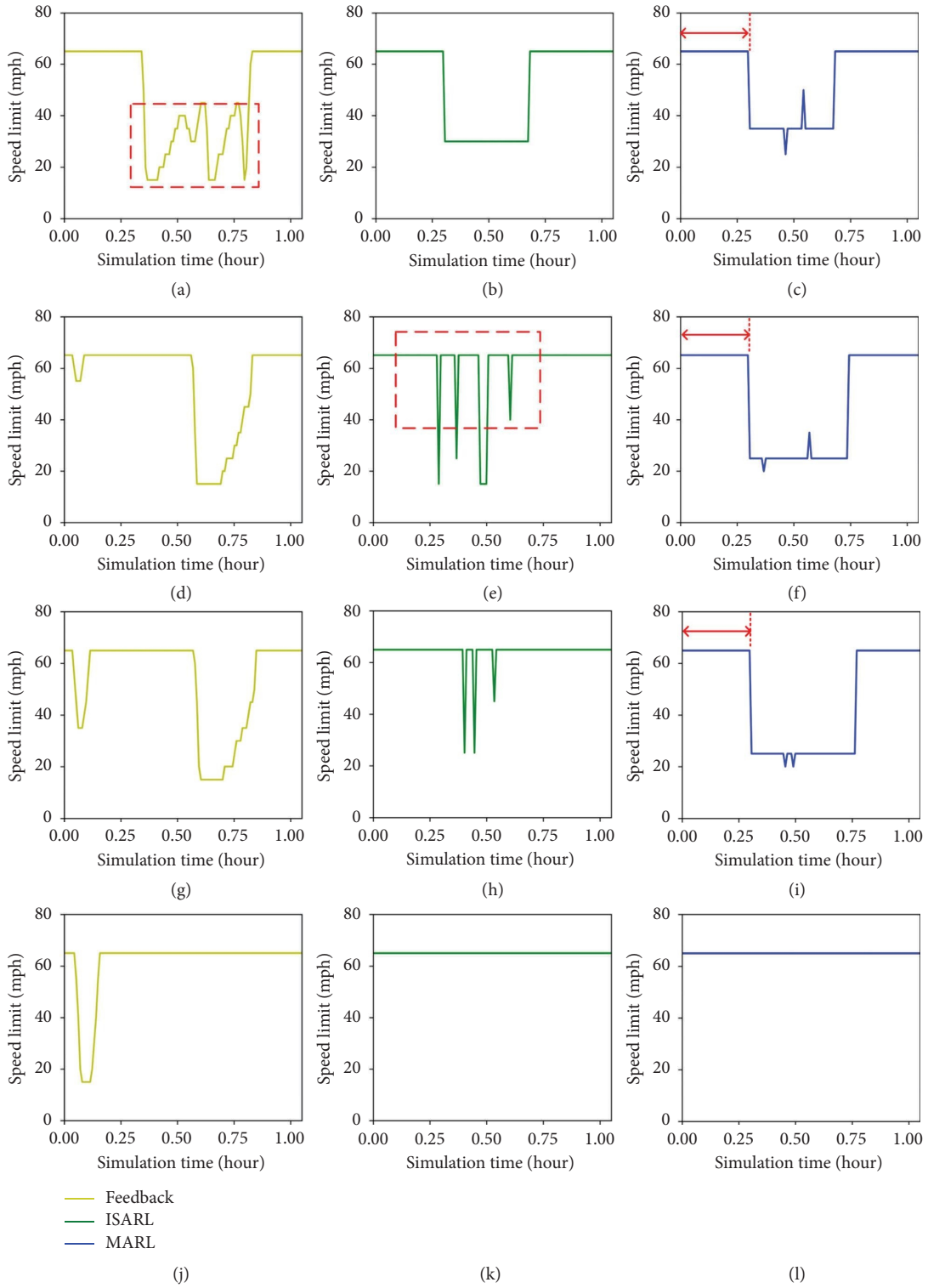

Figure 12: Comparison of speed limit with different VSL approaches in the static traffic scenario. (a–c) Bottleneck 1, (d–f) bottleneck 2, (g–i) bottleneck 3, and (j–l) bottleneck 4.

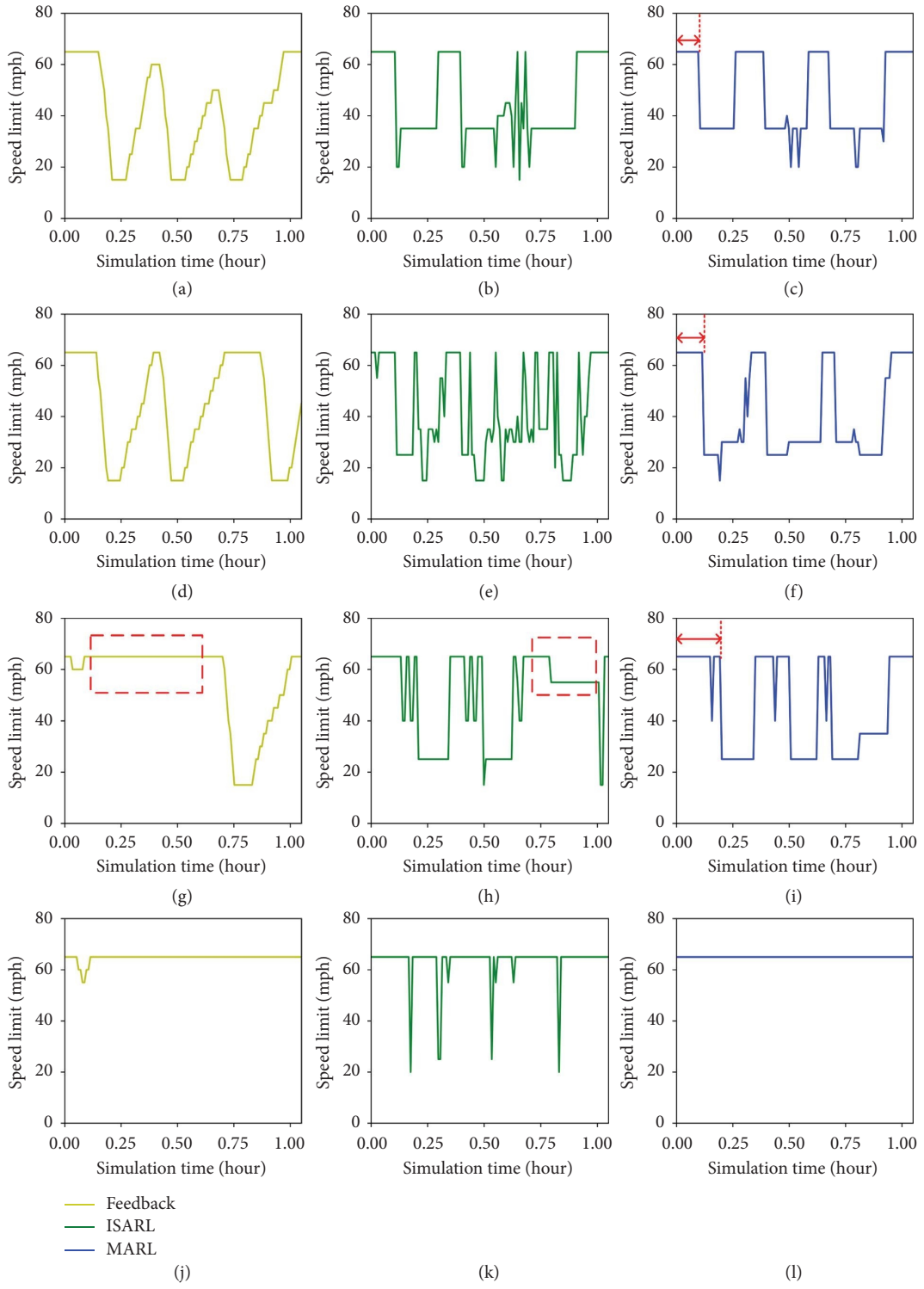

Figure 13: Comparison of speed limit with different VSL approaches in the dynamic traffic scenario. (a–c) Bottleneck 1, (d–f) bottleneck 2, (g–i) bottleneck 3, and (j–l) bottleneck 4.

Table 4: Total time spent in diverse capacity drop scenarios.

Scenario	Capacity drop (%)	TTS of No-VSL (veh·hour)	TTS of MARL (veh·hour)
Static traffic scenario	5.0	275.93	236.38
	10.0	311.78	237.25
	12.5	337.08	235.57
	15	362.34	261.17

Table 4: Continued.

Scenario	Capacity drop (%)	TTS of No-VSL (veh·hour)	TTS of MARL (veh·hour)
Dynamic traffic scenario	5.0	305.06	265.61
	10.0	337.93	267.74
	12.5	364.15	264.42
	15	407.28	314.49

capacity drop varied from 5.00% to 12.50%. Notably, a significant increase in TTS was observed with a capacity drop of 15%, reaching a value of 314.49 veh·hour. These results indicate that the MARL-VSL approach can yield satisfactory performance in most capacity drop scenarios, both in static and dynamic traffic conditions.

6. Conclusions and Recommendations

This study proposed the MARL-VSL approach to reduce the total time spent in scenarios with multiple consecutive bottlenecks. Specifically, the proposed strategy employed a centralized training with a decentralized execution structure to achieve a joint optimal solution for a series of VSL controllers. The MARL-VSL was trained in a predetermined dynamic traffic scenario. Two typical testing scenarios, namely, static and dynamic traffic scenarios, were utilized to evaluate the proposed method. We assessed the effectiveness of both ISARL-VSL and feedback-VSL control strategies for the comparative analysis. The speed limit values of each controller were analyzed, and traffic flow, density, and speed at the four bottlenecks were compared under different control strategies.

The simulation results provide evidence for the effectiveness of the proposed strategies in alleviating congestion at consecutive bottlenecks. The MARL-VSL strategy produced an 18.01% decrease in TTS in the static traffic scenario, surpassing the 8.02% and 13.82% reductions achieved by the feedback-VSL and ISARL-VSL strategies, respectively. In the dynamic traffic scenario, the MARL-VSL strategy demonstrated superior performance by achieving a reduction of 17.07% in TTS, as opposed to the 6.08% and 8.30% reductions observed in the feedback-VSL and ISARL-VSL strategies, respectively. Notably, the MARL-VSL strategy significantly outperformed other approaches in reducing TTS in the dynamic traffic scenario. This was attributed to the more pronounced traffic flow interference phenomenon between different bottlenecks in complex environments. However, our proposed strategy considered traffic flow information for the entire road section and employed a precise and coordinated control scheme that further reduced the interference effect of traffic flow. Additionally, the results showed that the MARL-VSL significantly improved the average driving speed and speed homogeneity of the entire freeway, thereby improving traffic efficiency. Furthermore, compared to other baseline methods, the control actions of the MARL-VSL were more appropriate in maintaining a smooth freeway traffic flow. The findings suggested that the MARL-VSL could effectively improve collaboration among the VSL controllers by considering each other's speed limit values when taking actions. The MARL-VSL also exhibited a degree of robustness by producing satisfactory outcomes under various capacity drop scenarios.

To advance the current research, future studies could explore various topics. At present, we only study the cooperative VSL control under the consecutive bottleneck scenarios due to merging flow from multiple on-ramps. The VSL control strategies in scenarios with numerous bottlenecks caused by curvature sections and traffic incidents are ongoing. Our research has mainly focused on the efficiency effect of the MARL-VSL in a multibottleneck control system. However, the safety impact of different control methods remains an important research area that requires further investigation. In addition, although our proposed MARL-VSL approach has demonstrated its effectiveness in improving traffic operation in consecutive bottleneck scenarios, more sophisticated multiagent reinforcement learning algorithms, such as Multi-Agent Variational Exploration (MAVEN) [50] and QTRAN [51], a factorization method for MARL, should undergo further testing.

Data Availability

The data supporting this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Si Zheng, Meng Li, and Zhibin Li conceived and designed the study. Si Zheng and Zemian Ke developed the algorithms and conducted the simulations. Si Zheng, Meng Li, and Zemian Ke analyzed and interpreted the results. Si Zheng and Meng Li drafted the manuscript. All authors have reviewed the results and approved the final version of the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 52272331 and 52232012) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province, China (grant no. KYCX22_0281).

References

[1] A. Srivastava and N. Geroliminis, "Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model," *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 161–177, 2013.

[2] K. Chung, J. Rudjanakanoknad, and M. J. Cassidy, "Relation between traffic density and capacity drop at three freeway

bottlenecks," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 82–95, 2007.

[3] S. Siri, C. Pasquale, S. Sacone, and A. Ferrara, "Freeway traffic control: a survey," *Automatica*, vol. 130, Article ID 109655, 2021.

[4] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 29–40, 2015.

[5] R. Ke, C. Liu, H. Yang, W. Sun, and Y. Wang, "Real-time traffic and road surveillance with parallel edge intelligence," *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 693–696, 2022.

[6] B. Khondaker and L. Kattan, "Variable speed limit: an overview," *Transportation Letters*, vol. 7, no. 5, pp. 264–278, 2015.

[7] K. Kušić, E. Ivanjko, M. Gregurić, and M. Miletić, "An overview of reinforcement learning methods for variable speed limit control," *Applied Sciences*, vol. 10, no. 14, p. 4917, 2020.

[8] Z. Li, P. Liu, W. Wang, and C. Xu, "Development of a control strategy of variable speed limits to reduce rear-end collision risks near freeway recurrent bottlenecks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 866–877, 2014.

[9] C. Lee, B. Hellinga, and F. Saccomanno, "Evaluation of variable speed limits to improve traffic safety," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 3, pp. 213–228, 2006.

[10] P. Allaby, B. Hellinga, and M. Bullock, "Variable speed limits: safety and operational impacts of a candidate control strategy for freeway applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 4, pp. 671–680, 2007.

[11] Z. Li, Y. Li, P. Liu, W. Wang, and C. Xu, "Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers," *Accident Analysis & Prevention*, vol. 72, pp. 134–145, 2014.

[12] L. Zhao, Z. Li, Z. Ke, and M. Li, "Fuzzy self-adaptive proportional–integral–derivative control strategy for ramp metering at distance downstream bottlenecks," *IET Intelligent Transport Systems*, vol. 14, no. 4, pp. 250–256, 2020.

[13] L. Zhao, Z. Li, Z. Ke, and M. Li, "Distant downstream bottlenecks in local ramp metering: comparison of fuzzy self-adaptive PID controller and PI-alinea," in *Proceedings of the ICTE 2019 Sixth International Conference on Transportation Engineering*, pp. 2532–2542, Chengdu, China, July 2019.

[14] G.-R. Iordanidou, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Feedback-based mainstream traffic flow control for multiple bottlenecks on motorways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 1–12, 2014.

[15] X.-Y. Lu, S. E. Shladover, I. Jawad, R. Jagannathan, and T. Phillips, "Novel algorithm for variable speed limits and advisories for a freeway corridor with multiple bottlenecks," *Transportation Research Record*, vol. 2489, no. 1, pp. 86–96, 2015.

[16] M. Yu and W. D. Fan, "Optimal variable speed limit control in connected autonomous vehicle environment for relieving freeway congestion," *Journal of Transportation Engineering, Part A: Systems*, vol. 145, no. 4, Article ID 04019007, 2019.

[17] K. Kušić, I. Dusparic, M. Guériau, M. Gregurić, and E. Ivanjko, "Extended variable speed limit control using multi-agent reinforcement learning," in *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8, Rhodes, Greece, September 2020.

[18] K. Kušić, E. Ivanjko, F. Vrbanić, M. Gregurić, and I. Dusparic, "Spatial-temporal traffic flow control on motorways using distributed multi-agent reinforcement learning," *Mathematics*, vol. 9, no. 23, p. 3081, 2021.

[19] Z. Ke, Z. Li, Z. Cao, and P. Liu, "Enhancing transferability of deep reinforcement learning-based variable speed limit control using transfer learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4684–4695, 2021.

[20] K. Kušić, E. Ivanjko, and M. Gregurić, "A comparison of different state representations for reinforcement learning based variable speed limit control," in *Proceedings of the 2018 26th Mediterranean Conference on Control and Automation (MED)*, pp. 1–6, IEEE, Zadar, Croatia, June 2018.

[21] Z. Li, P. Liu, C. Xu, H. Duan, and W. Wang, "Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3204–3217, Nov, 2017.

[22] Z. Li, C. Xu, Y. Guo, P. Liu, and Z. Pu, "Reinforcement learning-based variable speed limits control to reduce crash risks near traffic oscillations on freeways," *IEEE Intell. Transport. Syst. Mag.,*vol. 13, no. 4, pp. 64–70, 2021.

[23] Y. Wu, H. Tan, and B. Ran, "Differential variable speed limits control for freeway recurrent bottlenecks via deep reinforcement learning," 2018, https://arxiv.org/abs/1810.10952.

[24] Y. Han, A. Hegyi, L. Zhang, Z. He, E. Chung, and P. Liu, "A new reinforcement learning-based variable speed limit control approach to improve traffic efficiency against freeway jam waves," *Transportation Research Part C: Emerging Technologies*, vol. 144, p. 2022, Article ID 103900, 2022.

[25] H. F. Yang, J. Cai, C. Liu, R. Ke, and Y. Wang, "Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning," *Transportation Research Part C: Emerging Technologies*, vol. 148, Article ID 103982, 2023.

[26] M. Li, Z. Cao, and Z. Li, "A reinforcement learning-based vehicle platoon control strategy for reducing energy consumption in traffic oscillations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 12, pp. 5309–5322, 2021.

[27] P. Sunehag, "Value-decomposition networks for cooperative multi-agent learning," 2017, https://arxiv.org/abs/1706.05296.

[28] C. Wang, J. Zhang, L. Xu, L. Li, and B. Ran, "A new solution for freeway congestion: cooperative speed limit control using distributed reinforcement learning," *IEEE Access*, vol. 7, pp. 41947–41957, 2019.

[29] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[30] S. Li, "Multi-agent deep deterministic policy gradient for traffic signal control on urban road network," in *Proceedings of the 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 896–900, Dalian, China, August 2020.

[31] T. Chu, J. Wang, L. Codeca, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1086–1095, 2020.

[32] L. Li, W. Li, X. Chen, J. Wang, and Q. Peng, "Learning based multi-channel access in vehicular networks with illegal operators and users," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 4097–4102, 2023.

[33] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Local feedback-based mainstream traffic flow control on motorways using variable speed limits," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1261–1276, 2011.

[34] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Comparison of local feedback controllers for the mainstream traffic flow on freeways using variable speed limits," *Journal of Intelligent Transportation Systems*, vol. 17, no. 4, pp. 268–281, 2013.

[35] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Integrated feedback ramp metering and mainstream traffic flow control on motorways using variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 209–221, 2014.

[36] K. Ma, H. Wang, Z. Zuo, Y. Hou, X. Li, and R. Jiang, "String stability of automated vehicles based on experimental analysis of feedback delay and parasitic lag," *Transportation Research Part C: Emerging Technologies*, vol. 145, Article ID 103927, 2022.

[37] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 156–172, 2008.

[38] T. P. Lillicrap, "Continuous control with deep reinforcement learning," 2019, https://arxiv.org/abs/1509.02971.

[39] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[40] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Piecewise-linearized cell transmission model and parameter calibration methodology," *Transportation Research Record*, vol. 1965, no. 1, pp. 183–191, 2006.

[41] M. Mauch and M. J. Cassidy, "Freeway traffic oscillations: observations and predictions," in *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pp. 653–673, Emerald Group Publishing Limited, Adelaide, Australia, July 2002.

[42] K. Ma, H. Wang, and T. Ruan, "Analysis of road capacity and pollutant emissions: impacts of Connected and automated vehicle platoons on traffic flow," *Physica A: Statistical Mechanics and Its Applications*, vol. 583, Article ID 126301, 2021.

[43] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Application of reinforcement learning with continuous state space to ramp metering in real-world conditions," in *Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems*, pp. 1590–1595, Anchorage, AK, USA, September 2012.

[44] D. Silver, J. Schrittwieser, K. Simonyan et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[45] D. Silver, A. Huang, C. J. Maddison et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[46] W. Zheng, Q. Guo, H. Yang, P. Wang, and Z. Wang, "Delayed propagation transformer: a universal computation engine towards practical control in cyber-physical systems," in *Advances in Neural Information Processing Systems*, pp 12141–12153, Curran Associates, Inc, Red Hook, NY, USA, 2021.

[47] M. Yang, Z. Li, Z. Ke, and M. Li, "A deep reinforcement learning-based ramp metering control framework for improving traffic operation at freeway weaving sections," in *Proceeding of the Transportation Research Board 98th Annual Meeting*, Washington DC, USA, January 2019.

[48] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: an overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–281, Dec. 2002.

[49] M. D. Hadiuzzaman and T. Z. Qiu, "Cell transmission model based variable speed limit control for freeways," *Canadian Journal of Civil Engineering*, vol. 40, no. 1, pp. 46–56, 2013.

[50] A. Mahajan, T. Rashid, M. Samvelyan, and S. Whiteson, "MAVEN: multi-agent variational exploration," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc, Red Hook, NY, USA, 2019.

[51] K. Son, D. Kim, W. J. Kang, D. E. Hostallero, and Y. Yi, "QTRAN: learning to factorize with transformation for cooperative multi-agent reinforcement learning," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 5887–5896, PMLR, Long Beach, California, USA, May 2019.