

## Article

# Research on Integrated Control Strategy for Highway Merging Bottlenecks Based on Collaborative Multi-Agent Reinforcement Learning

Juan Du <sup>1,2</sup> , Anshuang Yu <sup>1,2</sup>, Hao Zhou <sup>1,2,\*</sup>, Qianli Jiang <sup>1,2</sup> and Xueying Bai <sup>1</sup>

<sup>1</sup> SILC Business School, Shanghai University, Shanghai 201800, China; ritadu@shu.edu.cn (J.D.); yas2022@shu.edu.cn (A.Y.); 17673156854@shu.edu.cn (Q.J.); bxy@shu.edu.cn (X.B.)

<sup>2</sup> SHU-SUCG Research Centre for Building Industrialization, Shanghai University, Shanghai 201800, China

\* Correspondence: hao\_hao333@shu.edu.cn

**Abstract:** The merging behavior of vehicles at entry ramps and the speed differences between ramps and mainline traffic cause merging traffic bottlenecks. Current research, primarily focusing on single traffic control strategies, fails to achieve the desired outcomes. To address this issue, this paper explores an integrated control strategy combining Variable Speed Limits (VSL) and Lane Change Control (LCC) to optimize traffic efficiency in ramp merging areas. For scenarios involving multiple ramp merges, a multi-agent reinforcement learning approach is introduced to optimize control strategies in these areas. An integrated control system based on the Factored Multi-Agent Centralized Policy Gradients (FACMAC) algorithm is developed. By transforming the control framework into a Decentralized Partially Observable Markov Decision Process (Dec-POMDP), state and action spaces for heterogeneous agents are designed. These agents dynamically adjust control strategies and control area lengths based on real-time traffic conditions, adapting to the changing traffic environment. The proposed Factored Multi-Agent Centralized Policy Gradients for Integrated Traffic Control in Dynamic Areas (FM-ITC-Darea) control strategy is simulated and tested on a multi-ramp scenario built on a multi-lane Cell Transmission Model (CTM) simulation platform. Comparisons are made with no control and Factored Multi-Agent Centralized Policy Gradients for Integrated Traffic Control (FM-ITC) strategies, demonstrating the effectiveness of the proposed integrated control strategy in alleviating highway ramp merging bottlenecks.

**Keywords:** variable speed limits; lane change control; multi-agent reinforcement learning; highway merging bottlenecks; integrated control strategy



Academic Editor: Benoit Lung

Received: 20 November 2024

Revised: 28 December 2024

Accepted: 9 January 2025

Published: 16 January 2025

**Citation:** Du, J.; Yu, A.; Zhou, H.; Jiang, Q.; Bai, X. Research on Integrated Control Strategy for Highway Merging Bottlenecks Based on Collaborative Multi-Agent Reinforcement Learning. *Appl. Sci.* **2025**, *15*, 836. <https://doi.org/10.3390/app15020836>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Highways serve as the core of modern transportation networks, significantly impacting China's transportation system and regional economic development. However, the current construction and operation of highway infrastructure fail to effectively meet growing regional traffic demands, leading to frequent congestion and accidents. Consequently, road efficiency decreases, driving safety is compromised, and travel times are extended. Traffic bottlenecks frequently occur at the merging points of mainline and entrance ramp lanes due to lane reductions and vehicle lane changes, severely impacting vehicle speeds. Frequent deceleration, acceleration, or stoppages in these bottleneck areas result in significant delays and congestion [1].

To address the imbalance between highway traffic demand and supply, a comprehensive approach [2] that integrates technological measures, infrastructure improvements, and demand-side optimization is employed. Technological measures, such as Intelligent Transportation Systems (ITS), are used to enhance traffic operational efficiency. Infrastructure improvements, including overall network planning [3], roundabout intersections [4], and ramp optimization designs [5], are implemented to increase capacity. Demand-side optimization, through regional traffic diversion, public travel behavior guidance, and demand substitution, effectively mitigates peak traffic pressure.

Among these, Intelligent Transportation Systems (ITS) are crucial for resolving highway traffic demand conflicts by providing control instructions to vehicles based on traffic flow data patterns. This approach achieves control objectives for specific road segments. For instance, Variable Speed Limit (VSL) control optimizes traffic efficiency by adjusting upstream speed limits to reduce spatiotemporal variations in traffic speed, thereby improving bottlenecks and preventing further deterioration. This method enhances traffic performance, safety, and mobility. However, vehicle weaving caused by lane changes in merging areas, aside from increasingly saturated traffic demand, also contributes significantly to bottleneck issues. Lane Change Control (LCC) calculates congestion levels in different lanes and issues mandatory lane change instructions to vehicles upstream of merging bottlenecks, optimizing lane density distribution and alleviating bottlenecks caused by vehicle weaving. VSL and LCC are complementary and synergistic methods. Implementing LCC requires that the road segment is not overly saturated; otherwise, vehicles may not respond to lane change instructions. VSL can mitigate this issue. Therefore, integrating VSL and LCC effectively addresses ramp merging bottlenecks.

Numerous studies have integrated VSL and LCC to alleviate ramp merging bottlenecks. For example, Zhang et al. [6] found that integrating VSL and LCC stabilizes upstream vehicle density and homogenizes traffic flow, reducing frequent lane changes near bottlenecks. Additionally, LCC provides lane change recommendations for vehicles away from bottlenecks, mitigating capacity drop issues. However, previous research limited control areas to fixed and unchanging ranges, resulting in inflexible control strategies. Furthermore, when multiple entrance ramps are present, local control by different controllers impacts overall traffic due to the upstream and downstream propagation characteristics of traffic flow. Integrated control research on multiple entrance ramps is necessary to alleviate merging bottlenecks in large-scale road networks effectively. This approach mitigates congestion at each ramp and maintains the stability of overall network traffic operations.

Based on the current analysis of research on highway ramp merging, this paper's main contribution lies in the combination of controllers and their integration with reinforcement learning. The specific contributions are as follows:

- To address the difficulty in coordinating traffic operations among multiple ramps in large-scale road networks using integrated control methods, this paper applies collaborative Multi-Agent Reinforcement Learning (MARL) to optimize traffic control strategies for multiple ramp merges. By transforming the integration of VSL and LCC into a Decentralized Partially Observable Markov Decision Process (Dec-POMDP), a ramp merging integrated control strategy framework based on the Factored Multi-Agent Centralized policy gradients (FACMAC) algorithm, termed FM-ITC, is proposed.
- On this basis, the FM-ITC strategy framework is expanded by incorporating the agents' actions. The improved strategy framework, FM-ITC-Darea, evaluates the impact of control area configuration changes on traffic flow optimization, providing a novel solution for traffic optimization in large-scale complex road networks.

- The proposed integrated control strategy is simulated and tested on a multi-ramp scenario built on a multi-lane Cell Transmission Model (CTM) simulation platform, demonstrating its effectiveness in alleviating highway ramp merging bottlenecks.

## 2. Related Works

### 2.1. VSL and LCC Traffic Control Methods

VSL systems are primarily categorized into open-loop and closed-loop control. Initially, open-loop control was favored for its simple structure and ease of maintenance, as it operates based on feedforward characteristics. For instance, Lee et al. [7] proposed a VSL method using a collision prediction model to evaluate highway collision risks, demonstrating that variable speed limits significantly reduce overall collision probability compared to fixed speed limits. Optimal control is a typical open-loop method designed for dynamic systems. Miao et al. [8] proposed an optimal VSL strategy for connected and automated vehicles using optimal control methods, showing significant improvements in traffic efficiency and emission reductions in simulations. However, due to the lack of feedback in open-loop control, systems are highly susceptible to disturbances and exhibit poor stability. Therefore, many researchers have shifted toward closed-loop control. Papamichail et al. [9] employed a proportional–integral feedback regulator for highway traffic control, validated with real road data from Germany, showing substantial improvements in performance metrics.

Model Predictive Control (MPC) methods represent another control category, predicting traffic flow trends to enable preemptive congestion control. Hegyi et al. [10] first applied MPC to VSL, with subsequent research expanding this approach. New MPC types and hierarchical control methods based on discrete LWR models have been developed and validated in microsimulation environments to effectively mitigate highway congestion waves [11]. Han et al. [12] introduced a fast MPC method for VSL coordination, addressing highway congestion waves by considering the linear properties of classical discrete first-order models and congestion wave propagation characteristics.

VSL research in traffic operations focuses on enhancing road traffic efficiency and safety. Studies indicate that VSL can boost traffic efficiency by 10% to 40% [13], especially when optimizing total travel time. However, recent studies on VSL effectiveness present mixed results [14], questioning VSL's impact on traffic efficiency under high-demand conditions. This disparity is attributed to highly disordered and randomly changing traffic flows at congestion bottlenecks.

To address VSL's limitations in traffic efficiency improvement, research has increasingly integrated VSL with other active control methods. Current studies on ramp control often integrate VSL with Ramp Metering (RM) [15,16]. While VSL-RM strategies enhance traffic efficiency and safety more than VSL alone, ramp queue lengths can overflow when mainline traffic is heavy.

With advancements in vehicle-to-everything (V2X) technology and intelligent transportation systems, integrated control of connected vehicles can be achieved through spatial management, combining lane change control. Guo et al. [17] proposed an integrated VSL and LCC method using an MPC framework in heterogeneous multi-lane CTM, optimizing the weighted sum of TTS and TTD in each control cycle. This method maximizes traffic efficiency by determining lane change numbers and speed limits while considering their interactions. Markantonakis et al. [18] applied MTFC as a VSL strategy at lane-drop bottlenecks, treating lane change as a linear quadratic optimal control problem and implementing results every 10 s in an Aimsun microsimulation model.

## 2.2. Application of Reinforcement Learning in VSL and LCC

In traffic control, feedback control and MPC control face challenges, such as frequent strategy adjustments leading to secondary congestion, system oscillation instability, poor convergence in complex traffic environments, and occasional failure to converge [19]. To address these issues, more flexible and intelligent control strategies are introduced. Reinforcement learning (RL) can learn optimal strategies directly from interactions with the environment without pre-modeling the system. This adaptability allows RL to better handle complex and dynamic traffic environments, providing robust responses to system changes and uncertainties [20].

Reinforcement learning, a popular intelligent control method in automatic control, has achieved notable results in previous traffic control studies [21–24]. However, most RL-based traffic control research has focused on small-scale road networks, mainly targeting single-point ramp control. Recently, researchers have turned to more realistic multi-ramp coordinated control. Applying RL to multiple ramp traffic networks presents challenges, as agents typically react to environmental changes individually, without considering coordination among agents (e.g., VSL controllers). This lack of coordination can result in suboptimal overall behavior [25]. To maximize regional traffic efficiency, more studies combine traffic control systems with collaborative Multi-agent Reinforcement Learning (MARL) algorithms to achieve cooperative control among multiple ramps [26].

Wang et al. [27] were the first to transform the VSL problem into a MARL problem, considering a distributed MARL system in a V2I environment. The system comprises multiple ramp control agents that communicate and cooperate to maximize highway traffic flow and safety benefits. To address non-stationarity issues among multiple traffic control agents, Zheng et al. [28] employed the Centralized Training and Decentralized Execution (CTDE) paradigm with the MADDPG algorithm to optimize large-scale VSL control for continuous traffic bottlenecks. Compared to traditional independent agent or feedback-based VSL control strategies, their method effectively improves overall highway network flow and safety by coordinating cooperation among multiple VSL controllers.

Research combining RL with multiple traffic control strategies is relatively sparse, likely due to significant coordination challenges: determining the mutual influence of different control methods, setting control cycles for different controllers, and accurately and efficiently controlling multiple traffic controllers [29]. Schmidt-Dumon et al. [30] proposed a decentralized reinforcement learning method for highway traffic control in South Africa. They simultaneously applied RM and VSL, using decentralized RL to solve control problems online. Their study claims to be the first to apply MARL methods to real traffic control scenarios, particularly in handling concurrent RM and VSL issues. Ko et al. [31] proposed a method combining speed harmonization and lane merging control to improve traffic flow and fuel efficiency by controlling CAVs. CAVs adjust speed based on sensor data, optimizing vehicle-following speeds and prioritizing merging in bottleneck areas to enhance traffic efficiency. In their study, speed harmonization and merging control are treated as a homogeneous group, using a single neural network for training and taking coordinated control actions based on the same network. The results indicate significant benefits in fuel consumption and traffic congestion reduction from the synergy between speed harmonization and merging control. Peng and Xu [32] proposed a control strategy combining VSL and lane change guidance to enhance highway traffic safety and efficiency through distributed deep reinforcement learning. Their study simultaneously employs VSL and LCC to address complex traffic scenarios that could lead to secondary collisions. Simulation experiments demonstrate that the combined controller outperforms single sub-controllers in reducing traffic accidents and increasing road capacity, showcasing the complementary benefits of VSL and LCC.

Although existing research has made progress in the field of traffic control, several issues and challenges remain. First, the synergy between different control methods is difficult to achieve, and determining their interactions and control cycles is a pressing issue. Second, while decentralized reinforcement learning is theoretically feasible, its effectiveness in practical applications requires further empirical validation. Moreover, single neural networks may have limitations in handling the complexity and variability of traffic environments, restricting their potential in practical traffic control scenarios. Finally, despite the theoretical appeal of joint control strategies, their practical application is highly complex, requiring significant advancements in computational capability and execution efficiency.

### 3. Construction of Integrated Control Problems Based on Heterogeneous FACMAC

The multi-agent integrated control framework in this study is based on the FACMAC algorithm, which computes continuous actions using a model-free approach. Continuous actions offer greater flexibility when transitioning into VSL and LCC strategies.

#### 3.1. Collaborative Multi-Agent Reinforcement Learning

In real-world scenarios, larger-scale highways are often influenced by multiple ramps and controllers, a situation that can be depicted using MARL. Unlike single-agent reinforcement learning, multi-agent systems, composed of numerous agents, form complex systems where accurately perceiving the global state of the environment is exceedingly difficult. Instead, agents can only access partial and local information, referred to as observations  $o \in O$ . Consequently, the environment in MARL is influenced not solely by the actions of a single agent but by the interactions among multiple agents and their respective local environments. Additionally, direct or indirect interactions between agents lead to changes in the overall environment, influenced by the behavior strategies of different agents [33]. The rewards for all agents depend on the strategies of other agents. In MARL, each agent aims to learn the optimal Q-value or strategy, but the environment transition and reward function depend on the actions of all agents, making the environment appear unstable from the perspective of any single agent [34].

This situation violates the Markov process assumption that agents have full observability of the environment. In the real world, due to sensor limitations, decisions are made based on limited or noise-interfered observations. In MARL, this is referred to as partial observability, leading to state uncertainty and interaction uncertainty among agents during the learning process. For example, even though traffic controllers on different road sections may share a common goal, limited communication capabilities prevent them from providing accurate traffic information to other agents. Additionally, coordinating respective traffic control signals can lead to mutual influences, making overall traffic complex and unpredictable. For this reason, training in partially observable MARL scenarios is conducted through a distributed partially observable Markov decision process.

Dec-POMDP considers that agents in the environment cannot fully observe the state of the environment and face state uncertainty induced by noise. Dec-POMDP is defined as  $M_{\text{DecP}} = \langle I, S, A, R, T, O, o, \gamma \rangle$ . Here,  $I$  represents a finite set of agents;  $S$  denotes the set of environmental states;  $A$  indicates the set of joint action spaces, which includes each agent's individual actions  $A_i$ ;  $R$  denotes the reward function for all agents; and  $T$  represents the joint state transition probability function that describes the probability of transitioning to a new state given the current state and joint actions. In this paper, this function is modeled using macroscopic traffic simulation to obtain realistic information. The function's probability is regarded as 1,  $O$  represents the set of observations, with  $o$  indicating the joint

observation set, where  $o_i$  represents the local observations of the corresponding agent;  $\gamma$  represents the discount factor reflecting the parameter of reward change over time.

In Dec-POMDP, agents cannot obtain the true state of the entire environment and can only take actions based on their current local observations, without sharing observational information with other agents. Under these conditions, the goal of Dec-POMDP is to achieve effective cooperation among agents, finding a set of optimal strategies to maximize the total expected cumulative return for agents in the environment:

$$Q_{tot}^{\pi}(s_t, a_t) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_0 = a \right], \tag{1}$$

### 3.2. Heterogeneous FACMAC Algorithm

The FACMAC algorithm is a MARL algorithm that follows the CTDE paradigm. In FACMAC, a multi-agent actor–critic framework is employed, where a centralized critic is trained by considering the joint actions of all agents. The joint action value function is optimized by decomposing it across individual agents, utilizing a weighted distribution trained with a mixer network. This distribution more accurately reflects the true traffic environment changes in the global  $Q_{tot}^{\mu}$ . In the actor, decentralized execution of individual actions is achieved through deep deterministic policy gradient learning. The process of the FACMAC algorithm is illustrated in Figure 1.

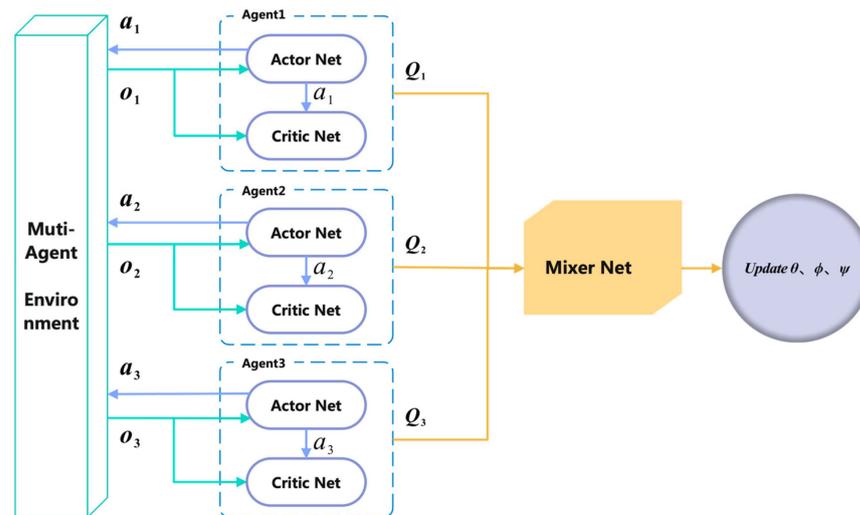


Figure 1. FACMAC algorithm diagram.

Specifically, FACMAC recognizes the challenge of learning in a centralized critic due to the large number of agents or action spaces. Therefore, the global action value function is decomposed into individual agents, which not only reduces training difficulty but also preserves the decentralized execution of the algorithm.

$$Q_{tot}^{\mu}(\tau, a, s; \phi, \psi) = g_{\psi} \left( s, \left\{ Q_i^{\mu_i}(\tau_i, a_i; \phi_i) \right\}_{i=1}^n \right), \tag{2}$$

In Equation (2), the global action value function  $Q_{tot}^{\mu}$  depends on the joint action-observation history  $\tau$ , joint action  $a$ , and global state  $s$ . Here,  $\phi$  represents the parameters of the agent’s policy network. Each agent has its own critic with an action value network  $Q_i^{\mu_i}$ , which is used to evaluate the performance of its policy. The individual action value  $Q_i^{\mu_i}$  can be computed using the nonlinear monotonic function  $g_{\psi}$  as part of  $Q_{tot}^{\mu}$ . The parameters

$\psi$  are computed using a mixer network. The loss function in Equation (3) is used to train  $Q_{tot}^\mu$  centrally:

$$\mathcal{L}(\phi, \psi) = \mathbb{E}_{\mathcal{D}}[(y^{tot} - Q_{tot}^\mu(\tau, a, s; \phi, \psi))^2], \tag{3}$$

where  $y^{tot} = r + \gamma Q_{tot}^\mu(\tau', \mu(\tau'; \theta^-), s'; \phi^-, \psi^-)$  and the parameters  $\theta^-, \phi^-, \psi^-$  represent the target actor, critic, and mixer network parameters, respectively.

To address the issues arising from agents' dependence on other agents' policy estimates, FACMAC centralizes the policy gradient update across the entire joint action space. This centralization corrects strategy errors and inaccurate evaluations during the learning process:

$$\nabla_{\theta} J(\mu) = \mathbb{E}_{\mathcal{D}}[\nabla_{\theta} \mu \nabla_{\mu} Q_{tot}^\mu(\tau, \mu_1(\tau_1), \dots, \mu_n(\tau_n), s)], \tag{4}$$

where all agents use a unified centralized actor policy network. When estimating the global action value function  $Q_{tot}^\mu$ , the current policies of all agents,  $\mu = \mu_1(\tau_1), \dots, \mu_n(\tau_n)$ , are sampled to obtain joint actions, and the overall joint action space is optimized through gradient ascent.

This updating method is particularly important for optimizing the overall efficiency of highway traffic control. For a highway traffic system involving multiple entrance and exit ramps, traffic integration control points are generally set in merging and diverging areas prone to bottlenecks. Due to the continuity of the highway system and the fluidity of vehicles, decisions made in a single traffic integration control area affect not only the traffic flow near the ramp but also the entire system, especially downstream areas. The FACMAC algorithm can optimize across the entire joint action space rather than being limited to a single control point.

However, considering that the FACMAC algorithm primarily addresses control issues in homogeneous cooperative multi-agent environments where all agents share the same policy network, this requires that all agents in the network have identical environmental perceptions and behavior strategies. This means the input and output dimensions of all agents' control processes are fixed. This approach limits the integrated control of the VSL and LCC controllers in single entry ramp merging areas, as it fails to account for differences between various traffic controllers. In this paper's highway environment, the two controllers are heterogeneous agent controllers because the traffic states they need to obtain and their control methods differ. Thus, the FACMAC algorithm cannot be directly applied to train and provide effective control strategies in such a multi-agent environment.

To solve this problem, this paper draws on the concept of classes used in the object-oriented distributed reinforcement learning method by Da Silva et al. [35], constructing heterogeneous agents into different classes through object-oriented principles. Specifically, in this paper, when considering the optimization problem of traffic flow in the highway merging area within a multi-agent system, an effective method is to model the VSL controller and the lane LCC controller in the ramp merging area as independent class objects. The integrated control is then achieved by integrating the state inputs and action outputs of these entities.

In this design, each type of controller is considered an independent agent class AC, with its unique properties and behaviors. The VSL agent class  $AC_{VSL}$  determines speed limits based on traffic density and flow information within the ramp merging area. Its input state dimension is denoted as  $D_s^{vsl}$ . The LCC agent class  $AC_{LCC}$  focuses on the density differences between different lanes in the merging area, with its input state dimension represented as  $D_s^{lcc}$ . By concatenating the state inputs of these two agent types, the overall input dimension becomes  $D_s^{vsl} + D_s^{lcc}$ . Unnecessary state dimensions for different agent types are filled with null values to maintain data consistency. Furthermore, since the control

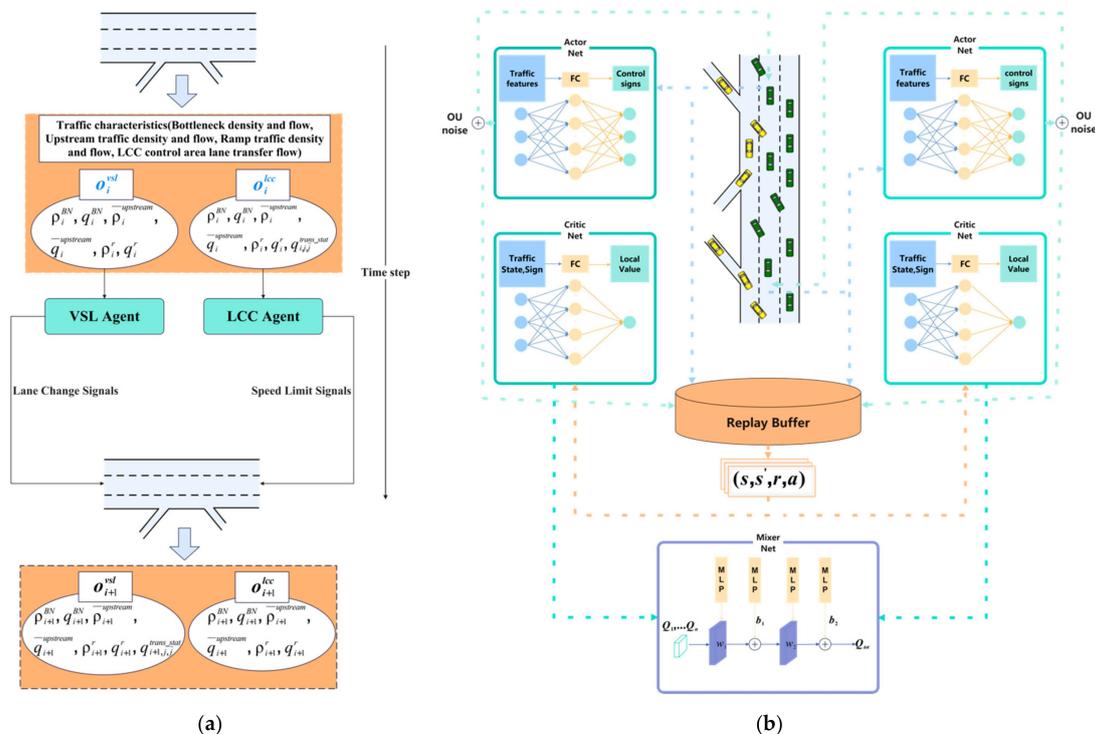
methods of the two agent types vary, the action output dimensions also differ. Thus, the output dimension of the policy network is designed to match the maximum action output dimensions among the agent classes. Agents with fewer action output dimensions extract only the required dimensions for their actions.

The two controllers in an integrated control group belong to class objects of different intelligences, each having its own critic and actor networks. These controllers differ from other agent classes only in terms of input and output dimensions, while agents of the same class within different integrated control groups share network parameters. This design leverages similar observation spaces and behavior strategies, enabling the sharing of network parameters among homogeneous agents.

This approach effectively addresses the failure of the FACMAC algorithm in heterogeneous intelligent body environments. Although the approach may encounter dimensionality issues due to the presence of multiple agent classes in a heterogeneous environment, it remains applicable for the highway merging bottleneck control problem discussed in this paper. Prior studies on traffic active control strategies have focused on variable speed limit control, ramp control, and lane change control.

### 3.3. FACMAC-Based Combined VSL and LCC Controller

Consider two controllers, VSL and LCC, for each entrance ramp area, and call such an integrated entrance ramp controller involving VSL controllers and LCC controllers an integrated control group. As shown in Figure 2a, the data fusion process integrates key information from the highway environment. The integrated control group monitors real-time traffic speed and density near the entrance ramp. Centralized training simultaneously provides speed limit information and lane change information to alleviate bottlenecks in the merging area. Interaction between the two controllers within the integrated control group and between different integrated control groups impacts the overall system.



**Figure 2.** Architecture diagram of FACMAC-based ramp merge integration control system, where (a) depicts the data fusion process, which integrates crucial information from the surrounding highway environment, and (b) represents the integrated control framework designed for the merging area in this study.

Consider two controllers, the VSL and LCC, for each entrance ramp area. These controllers form an integrated control group that observes traffic speed and density related to the entrance ramp. Centralized training simultaneously provides speed limit and lane change information to alleviate bottlenecks in the merging area. The interaction between the controllers within the integrated control group and across different integrated control groups impacts the overall system performance.

Figure 2b illustrates the merging area integrated control framework designed in this study. Each controller agent has a value network and a policy network, which are trained with data  $(s, s', r, a)$  collected during interactions with the traffic environment and stored in a replay buffer. The output value of each policy network serves as the input to the mixer network. The detailed pseudocode is provided in Algorithm 1.

Each agent's actor network is designed using a three-layer fully connected architecture to map states to actions, with two hidden layers following ReLU activation functions to introduce nonlinearity. The output layer uses the tanh function to constrain the action output to the range  $[-1, 1]$ , matching continuous action requirements. The critic network uses a fully connected layer to evaluate state-action values, combining ReLU-activated hidden layers for processing, and directly outputs the estimated value, supporting the optimization of agent policies.

For the mixer network, two super networks are employed to calculate and generate its parameters. Each super network consists of a linear layer and an absolute value activation function. These two super networks calculate the weights for the first layer neurons  $w_1$  and the hidden layer to output the global  $Q_{tot}$  weights  $w_2$  by receiving the global state  $s$ . Similar network configurations are used to generate biases for the two-layer network, ultimately forming the nonlinear mixer network.

In the policy gradient optimization process of reinforcement learning, the high nonlinearity and dimensionality of the objective function often result in multiple local optima. To address this, we adopt the  $\epsilon$ -greedy strategy to balance exploration and exploitation. By dynamically adjusting the intensity of exploration, this strategy ensures that the agent extensively explores diverse policy spaces during the early stages and gradually converges to the optimal policy in the later stages of training. Additionally, Ornstein–Uhlenbeck (OU) noise is introduced to further enhance exploration diversity. In continuous action spaces, OU noise helps prevent premature convergence to local optima. In traffic system control tasks, it effectively balances exploration and exploitation, avoids abrupt action jumps, and strengthens the diversity and robustness of policy exploration. Its temporal correlation and smoothness not only meet practical control requirements but also significantly improve the algorithm's global optimization performance in complex nonlinear environments.

The following details the key elements of converting the integrated controller into a DEC-POMDP process.

### 3.3.1. Agents

In different merging areas of the highway, the VSL agent class  $AC_{VSL}$  and the LCC agent class  $AC_{LCC}$  are constructed with instance objects  $BN_i^{vsl}$  and  $BN_i^{lcc}$ , where  $i$  represents the location number of the entrance ramp merging area.  $BN_i^{vsl}$  mainly observes and estimates the traffic conditions in the speed control area upstream of the ramp merging bottleneck and adjusts the speed limits accordingly. This adjustment aims to alleviate congestion caused by excessive traffic density in the merging area.  $BN_i^{lcc}$ , on the other hand, adjusts the traffic flow on different lanes within the current ramp merging area to optimize lane density distribution. The integrated control group  $G_i = BN_i^{vsl} + BN_i^{lcc}$  for ramp  $G_i$  is responsible for the integrated control of the entire merging bottleneck, enabling vehicles from both the mainline and the ramp to merge more efficiently.

**Algorithm 1** FACMAC-based combined control of VSL and LCC

---

```

1: Initialize policy networks  $\pi_\theta$  (actor), value networks  $Q_\phi$  (critic), and mixer network  $M$  (centralized  $Q$  computation)
2: Initialize target networks  $\pi_{\theta'}$ ,  $Q_{\phi'}$  and  $M'$  by copying parameters from  $\pi_\theta$ ,  $Q_\phi$  and  $M$ 
3: Initialize replay buffer  $D$  and hyperparameters (learning rates  $\alpha_{actor}$ ,  $\alpha_{critic}$ , discount factor  $\gamma$ , and soft update factor  $\tau$ )
4: while training steps < maximum training steps do
5:   Reset the environment, initialize state  $s_0$ 
6:   Reset exploration noise
7:   for  $t = 1$  to  $T$  do
8:     for each agent  $i = 1$  to  $N$  do
9:       Observe state  $s_t^i$  (local observation and history)
10:      Generate action  $a_t^i = \pi_\theta(s_t^i) + \text{noise}$ 
11:    end for
12:    Execute joint actions  $a_t = \{a_t^1, \dots, a_t^N\}$ , receive reward  $\gamma_t$ , next state  $s_{t+1}$ , and done flag  $d_t$ 
13:    Store  $(s_t, a_t, \gamma_t, s_{t+1}, d_t)$  in replay buffer  $D$ 
14:  end for
15:  if  $D.size \geq batch\_size$  then
16:    Sample a batch of transitions from  $D$ 
17:    Compute target  $Q$  – values using target networks  $\pi_{\theta'}$ ,  $Q_{\phi'}$  and  $M'$ 
18:    Update critic networks  $Q_\phi$  and mixer network  $M$  using Adam optimizer to minimize the TD error
19:    Update actor networks  $\pi_\theta$  using Adam optimizer to maximize the expected return
20:    Perform soft updates for target networks  $\pi_{\theta'}$ ,  $Q_{\phi'}$  and  $M'$ 
21:  end if
22: end while

```

---

### 3.3.2. Joint State Space

In macroscopic traffic flow theory, traffic conditions are usually described by a series of macroscopic variables such as traffic speed  $v$ , density  $\rho$ , and flow  $q$ . According to the fundamental diagram (FD), these variables exhibit significant nonlinear characteristics. This is mainly because for a given traffic flow value, there can be two distinctly different traffic states: one is a stable free-flow state below the critical density  $\rho^{cr}$ , and the other is an unstable flowing state at low speeds above  $\rho^{cr}$ . This phenomenon indicates that simply using the flow  $q$  to represent the traffic conditions perceived by agents can easily lead to perceptual confusion. Therefore, when constructing the Dec-POMDP, both the density  $\rho$  and flow  $q$  are comprehensively considered as macroscopic traffic states.

For an integrated ramp control group  $i$ , there is a local observation  $o_i^{vsl}$  for the VSL agent and a local observation  $o_i^{lcc}$  for the LCC agent. Considering that  $o_i^{vsl}$  and  $o_i^{lcc}$  belong to the same integrated control group, information sharing exists during the training process. Specifically, at ramp  $i$ , the local observation considers the following traffic characteristics:

1. Bottleneck density  $\rho_i^{BN}$  and flow  $q_i^{BN}$ : Typically, when traffic congestion occurs at a highway ramp merging bottleneck, there is significant traffic pressure and high traffic density at the bottleneck location due to inconsistent flow rates caused by vehicles merging into lanes. According to the definition, the bottleneck traffic state is represented by the lane cells at the entrance ramp.
2. Upstream traffic density  $\rho_i^{upstream}$  and flow  $q_i^{upstream}$ : Due to the strong state correlation between upstream and downstream highway traffic, understanding the upstream traffic state is essential for controlling ramp traffic bottlenecks. Considering that too many state variables increase the difficulty of training, traffic information within 1.5 km upstream is used as the current observation, with this information averaged as  $\bar{\rho}_i^{upstream}$  and  $\bar{q}_i^{upstream}$ .

3. Ramp traffic density  $\rho_i^r$  and flow  $q_i^r$ : Similarly, it is crucial to understand the traffic information related to the ramp, which is especially critical for the LCC agent, as it optimizes the lane density distribution based on these differences.
4. LCC control area lane transfer flow  $q_i^{trans\_stat}$ : Here, we define an information coefficient  $q_{i,j,\tilde{j}}^{trans\_stat}$  that reflects the amount of traffic that can be shifted from a lane to a neighbouring lane when a lateral lane change is performed, where  $q_{i,j,\tilde{j}}^{trans\_stat} = \max\{Nveh_{i,j}, l * \rho_{i,j} / T\}$ , and, in most cases,  $q_{i,j,\tilde{j}}^{trans\_stat}$  is determined by the shifted traffic  $Nveh_{i,j}$ .

Based on the above state information, the VSL agent's observation  $o_i^{vsl}$  in the integrated ramp control group  $i$  is defined as  $o_i^{vsl} \in \rho_i^{BN} + q_i^{BN} + \bar{\rho}_i^{upstream} + \bar{q}_i^{upstream} + \rho_i^r + q_i^r$ , which is acquired through sensors such as roadside detectors to capture the current traffic state at ramp  $i$ . The LCC agent's observation  $o_i^{lcc}$  includes lane transfer flow information and is defined as  $o_i^{lcc} \in \rho_i^{BN} + q_i^{BN} + \bar{\rho}_i^{upstream} + \bar{q}_i^{upstream} + \rho_i^r + q_i^r + q_{i,j,\tilde{j}}^{trans\_stat}$ .

### 3.3.3. Joint Action Space

At each time step  $t$ , the actions performed by different agent classes based on the current environment are not consistent. For an integrated ramp control group, the joint action  $u_t = \{a_{1,t}^{vsl}, a_{1,t}^{lcc}, a_{2,t}^{vsl}, a_{2,t}^{lcc}, \dots, a_{ramp_i,t}^{vsl}, a_{ramp_i,t}^{lcc}\}$  represents the set of actions taken by the VSL agent  $a_{ramp_i,t}^{vsl}$  and the LCC agent  $a_{ramp_i,t}^{lcc}$  for each ramp  $ramp_i$ .

The action of the VSL agent for each ramp  $a_{r_i,t}^{vsl}$  is represented by  $l + 1$  continuous variables (where  $l$  is the number of lanes) as

$$a_{ramp_i,t}^{vsl} = \{v_1, v_2, \dots, v_l, pos_{ramp_i,vsl}\}, \tag{5}$$

where  $pos_{ramp_i,vsl}$  in Equation (5) represents the position of the VSL dynamic control area, serving as a unified decision variable for the control lengths of all lanes before the ramp. To maintain the consistency of continuous decision variables, all variables are set within the range of  $[-1, 1]$ . Considering the efficiency and safety of highway traffic, VSLs are generally used to issue advisory speeds to vehicles in different lanes. The speed limit is defined as follows:

1. To prevent the speed limit from being too low and causing vehicles to fail to move smoothly, the speed limit should not be lower than the minimum highway speed,  $Vsl \geq V_{min}$ .
2. To ensure the effectiveness of speed limit control, the speed limit should not exceed the free flow speed,  $Vsl \leq V_{free}$ .
3. To maintain traffic flow smoothness and continuity, the speed limit change between consecutive control periods should be within a certain range, i.e.,  $|a_{ramp_i,t}^{vsl} - a_{ramp_i,t-1}^{vsl}| \leq m$ . This avoids drastic changes in traffic speed due to speed limit control, thereby reducing congestion and accident risks.
4. The speed limit should be a multiple of 5, i.e., the speed limit value should be adjusted to the nearest multiple of 5.

Based on the above constraints, the lane speed limit can be calculated using Equation (6):

$$V_{ramp_i,j,t}^{sl} = clip[a_{ramp_i,j,t}^{vsl} * m, v_{min}, v_{max}], \tag{6}$$

The  $pos_{ramp_i,vsl}$  in Equation (5) is converted to the corresponding length range in CTM by Equation (7):

$$pos_{ramp_i,vsl} = (pos_{ramp_i,vsl} + 1)/2 * (pos_{max}^{vsl} - pos_{min}^{vsl}) + 1, \tag{7}$$

where  $pos_{max}^{vsl}$  and  $pos_{min}^{vsl}$  represent the farthest and nearest cell positions in the VSL control area from the merging zone, respectively. The difference  $pos_{max}^{vsl} - pos_{min}^{vsl}$  ensures that the VSL dynamic control area remains within a limited range. Finally, the length of  $pos_{ramp_i,vsl}$  is increased by 1 to ensure that the minimum control range of the dynamic speed limit area is one cell length. Figure 3 illustrates the variation of dynamic control areas on a single entry ramp. It is worth noting that the range of dynamic control areas for VSL and LCC at different entry ramps is influenced by their respective strategies.

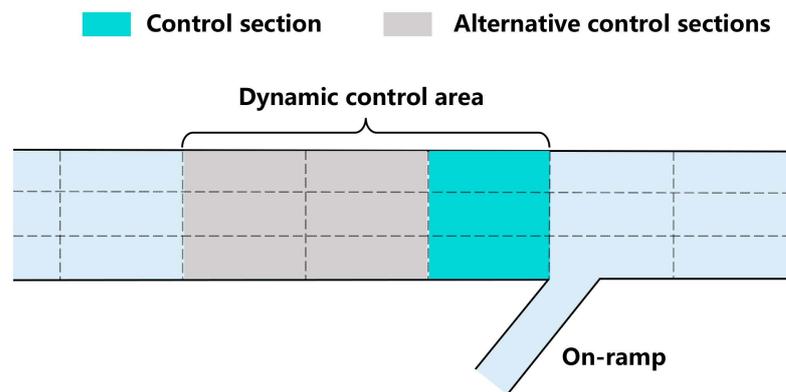


Figure 3. Dynamic control zones in integrated control strategies.

For the LCC agent action  $a_{ramp_i,t}^{lcc}$ , considering the multi-lane CTM model where the transferable vehicle number  $q_{i,j,\tilde{j}}^{trans\_stat}$  is primarily determined by the density difference between two adjacent lane cells, we use the lane transfer coefficient  $Ts_{ramp_i,j \rightarrow \tilde{j}}$ . This controls the proportion of vehicle flow transfer between lanes based on density rather than judging from a high-density lane to a low-density lane. This approach effectively manages the flow distribution between lanes. Therefore, the LCC agent action  $a_{ramp_i,t}^{lcc}$  at each ramp is expressed as Equation (8):

$$a_{ramp_i,t}^{lcc} = \{ Ts_{ramp_i,1 \rightarrow 2}, Ts_{ramp_i,2 \rightarrow 1}, Ts_{ramp_i,2 \rightarrow 3}, Ts_{ramp_i,3 \rightarrow 2}, pos_{ramp_i,lcc} \}, \tag{8}$$

For a three-lane highway, within the LCC control area, vehicles on the inner and outer lanes can only transfer to one adjacent lane, while the middle lane has the ability to transfer vehicles to both sides. Therefore,  $a_{ramp_i,t}^{lcc}$  for the control section contains four continuous variables representing the number of transferable vehicles between lanes  $Ts_{ramp_i,j \rightarrow \tilde{j}}$ . The LCC agent’s control range variable  $pos_{ramp_i,lcc}$  is also determined. Considering the constraints of the transferable vehicle number  $q_{i,j,\tilde{j}}^{trans\_stat}$  during lane changes between adjacent lanes, the LCC agent action  $Ts_{ramp_i,j \rightarrow \tilde{j}}$  is converted into lateral flow through Equations (9) and (10):

$$Ts_{ramp_i,j \rightarrow \tilde{j}} = (Ts_{ramp_i,j \rightarrow \tilde{j}} + 1)/2, \tag{9}$$

$$\hat{f}_{i,j \rightarrow \tilde{j}} = q_i^{trans\_stat} * Ts_{ramp_i,j \rightarrow \tilde{j}} \tag{10}$$

Since the output variables of the strategy network in this study are set in the range of  $[-1, 1]$ , and considering that the lane transfer coefficient  $Ts_{ramp_i,j \rightarrow \tilde{j}}$  cannot be negative, it

is scaled to  $[0, 1]$  using Equation (9). When  $Ts_{ramp_{i,j} \rightarrow j}^{\sim} = 0$ , it means the transfer power to adjacent lanes is completely closed, and  $Ts_{ramp_{i,j} \rightarrow j}^{\sim} = 1$  means optimizing the traffic efficiency for lane transfer. Thus, Equation (10) directly controls the lateral flow of transferable vehicles between adjacent lanes.

### 3.3.4. Reward

In Dec-POMDP problems, cooperative multi-agent systems typically share a global reward function. This means that all agents aim to maximize the overall performance of the entire system when making decisions, rather than just their individual performance. This emphasizes the importance of teamwork and collective goals.

In the multi-agent merging bottleneck integrated control model discussed in this paper, local integrated control groups optimize traffic flow efficiency and alleviate high lane density at specific ramps through VSL and LCC control. Considering the upstream and downstream interactions in the global traffic integrated control system, the integrated control agents at different ramps influence each other's traffic conditions, creating a chain reaction. The propagation of traffic flow can impact the broader highway network globally.

Previous studies aimed at mitigating merging traffic bottlenecks often focused on reducing the total travel spend (TTS) within the entire road network system, which includes the total travel time within the study area and the queuing time upon entry [27,36]. This approach is somewhat effective in simplified single-agent scenarios or with single traffic control measures. However, using TTS in multi-agent integrated control can lead to unnecessary confusion among agents.

Considering the distinction between VSL and LCC control, as well as the relationship between the global and local levels, this paper proposes a mixed reward function. Specifically, based on the objective of integrated control, we aim to maximize the traffic efficiency  $BN_{effect}$  in the merging bottleneck area. Although the traffic flow  $F_{BN}$  is a direct indicator of traffic efficiency, the nonlinear relationship between flow and density in the fundamental diagram better represents road traffic conditions. The FD reveals that as road density approaches the critical density  $\rho_{cr}$ , the traffic flow increases, but it sharply decreases when exceeding  $\rho_{cr}$ . Therefore, maintaining road density near the critical density  $BN_{effect}$  is crucial for improving traffic efficiency. For outer lanes, the density  $\rho_r$  of the ramp lane is also considered due to the impact of merging traffic. The reward function is defined by Equations (11)–(13):

$$r_{i,j}^{BN} = \begin{cases} \frac{\rho_{i,j}^2}{\rho_{cr}^2}, & \rho_{i,j} \leq \rho_{cr} \\ \max\left\{0, \frac{\rho_{cr}^2 - (\rho_{i,j} - \rho_{cr})^2}{\rho_{cr}^2}\right\}, & \rho_{i,j} > \rho_{cr} \end{cases}, \quad (11)$$

$$r_{i,out}^{BN} = \begin{cases} \frac{\rho_{i,out}^2}{\rho_{wish}^2}, & \rho_{i,out} \leq \rho_{wish} \\ \max\left\{0, \frac{\rho_{wish}^2 - (\rho_{i,out} - \rho_{wish})^2}{\rho_{wish}^2}\right\}, & \rho_{i,out} > \rho_{wish} \end{cases}, \quad (12)$$

$$r_{i,j}^{upstream} = -\frac{\rho_{i,j}^2}{\rho_{cr}^2}, \quad (13)$$

At time step  $t$ , the system receives an overall reward  $r_t$ , which is determined by the density of the merging bottleneck area and the upstream region. In Equation (11), the inner lane reward  $r_{i,j}^{BN}$  of the merging bottleneck area is designed to optimize traffic flow when the density  $\rho_{i,j}$  is below the critical density  $\rho_{cr}$ , ensuring maximum traffic throughput. We also aim to avoid a rapid decrease in capacity when  $\rho_{i,j}$  exceeds  $\rho_{cr}$ .

Similar to Equation (11), the density  $\rho_{i,out}$  of the outer lane in the merging bottleneck area is significantly influenced by the merging traffic at the entry ramp. Therefore, the reward  $r_{i,out}^{BN}$  for the outer lane in the merging bottleneck area is defined in Equation (12). The desired lane density  $\rho_{wish} = \rho^{cr} - \rho_r$ , where  $\rho_r$  is the ramp lane density, indicates that the outer lane should transfer more excess traffic to the inner lanes via the LCC to obtain a higher reward.

For Equation (13), during traffic control, especially speed limit control, there is often a tendency to lower the speed limit on upstream control sections to alleviate traffic pressure at merging points. While this approach mitigates traffic pressure at the bottleneck, it can also cause overall system congestion, reducing overall traffic efficiency. Thus,  $r_{i,j}^{upstream}$  represents the reward for the upstream section affected negatively by the VSL.

The global reward function  $r_t$ , in conjunction with Equations (11)–(13), considers both the overall traffic efficiency and the mitigation of local traffic pressure. The parameter  $c$  reflects the desired improvement in overall traffic capacity when control measures are implemented to alleviate traffic congestion at specific ramp bottlenecks:

$$r_t = \sum_i^{BN} \sum_{j \neq out}^l r_{i,j}^{BN} + \sum_i^{BN} r_{i,out}^{BN} + c \cdot \sum_i^{upstream} \sum_j^l r_{i,j}^{upstream}, \tag{14}$$

#### 4. Multi-Lane CTM Simulation Model

This study performs simulation experiments on the proposed integrated control based on the multi-lane CTM model. In CTM, the road is divided into several equal-length cells along the direction of traffic flow. In the multi-lane CTM model, the highway consists of  $n$  segments ( $i = 1, 2, \dots, n$ ) of length  $l$ . Each segment  $i$  is subdivided into  $m$  lanes ( $j = 1, 2, \dots, m$ ). A lane cell can be indexed as  $(i, j)$ . In this paper, the initial segment is defined as  $i = 1$ , and the outermost lane is defined as  $j = 1$ . As shown in Figure 4, the traffic density of a three-lane highway model is updated using the density conservation formula in Equation (15):

$$\rho_{i,j}(k+1) = \rho_{i,j}(k) + \frac{T}{l} [q_{i-1,j}(k) + r_{i,j}(k) - q_{i,j}(k) + f_{i,j+1,j}(k) + f_{i,j-1,j}(k) - f_{i,j,j-1}(k) - f_{i,j,j+1}(k)] \tag{15}$$

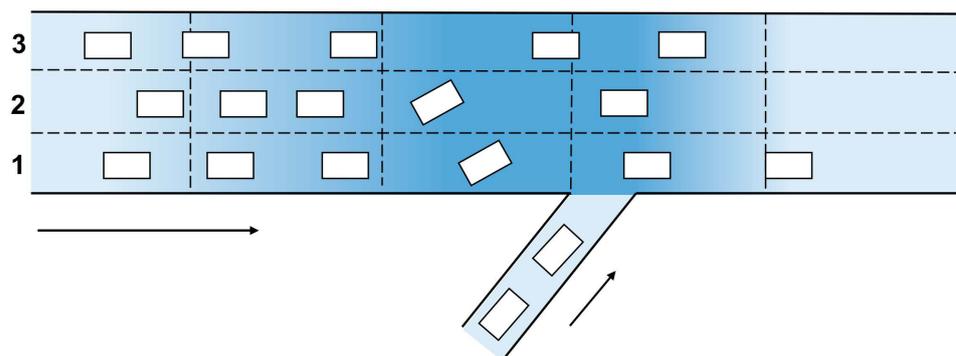
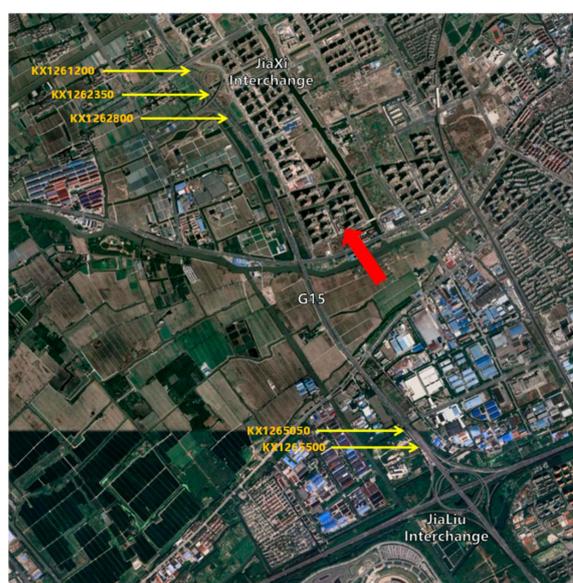


Figure 4. Freeway multi-lane traffic flow model.

Here,  $k$  represents the discrete time step;  $T$  is the aggregation period;  $q_{i,j}(k)$  and  $r_{i,j}(k)$  denote the flow rates of the mainline cell and the on-ramp cell, respectively; and  $f(k)$  represents the lateral flow between lane cells. To adapt the multi-lane CTM for the purposes of this study, corresponding improvements have been made. For more detailed information, interested readers can refer to [37].

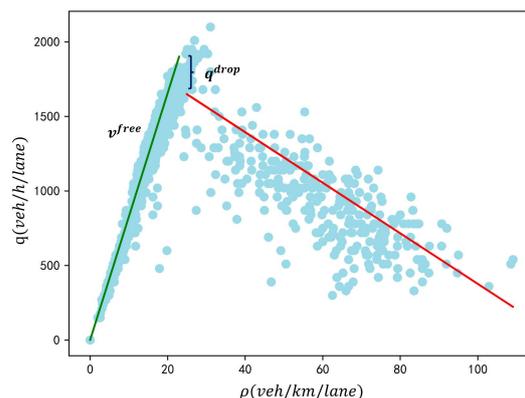
To facilitate the effective integration of the multi-lane CTM with the heterogeneous FACMAC algorithm, a `ctm_sim` class is developed in Python 3.9.13. This class encom-

passes modules for basic parameter configuration and initialization, simulation execution, traffic demand generation, traffic state retrieval, and reward calculation. Among these, model initialization is critical to the accuracy and reliability of the simulation. Therefore, prior to each simulation, the scenario must be confirmed, and model parameters must be appropriately configured. In this study, the downward direction of the Jiali section on the G15 Shenhai Expressway in Shanghai is selected as the research focus (indicated by the red arrow). A satellite image of the road is presented in Figure 5. The research scope includes two interchanges: Jiaxi Interchange and Jiali Interchange. The Jiali Interchange contains an entry ramp, while the Jiaxi Interchange includes both an entry ramp and an exit ramp. The two entry ramps are approximately 4.3 km in length. The mainline of the expressway primarily consists of three lanes, whereas the ramps typically feature one lane. Additionally, five pavement detectors are installed at specific stake points (highlighted in yellow) along the study segment.



**Figure 5.** Satellite picture of G15 Jiali section.

The construction and calibration of the multi-lane CTM are determined by two key factors: the road's geometric characteristics and the model's traffic parameters. For this study, the basic model parameters are set as follows: road length  $l = 0.5\text{km}$ , number of lanes  $m = 3$ , and simulation time step  $k$ , aggregated over an interval of  $T = 15\text{s}$ . The calibration of the traffic parameters is conducted using a flow-density scatter plot at the bottleneck upstream of the entry ramp (KX1262350) (Figure 6). Initially, the critical density  $\rho^{cr}$  is determined by observing that it is smaller than a specific density threshold  $\rho^a$ . Sample points satisfying  $\rho^a$  are identified, sorted in descending order of flow, and the top 3% of these points are selected. The mean density of the selected points is defined as the critical density  $\rho^{cr} = 21 \text{ veh/km/lane}$ , while their mean flow is defined as the road capacity  $Q^{max} = 1950 \text{ veh/lane}$ . Subsequently, all sample points are divided into two sets, a left set and a right set, based on  $\rho^{cr}$ . Linear regression using the least squares method is applied to both sets. The left set's regression line represents the free-flow speed  $v^{free}$  while the right set's regression line represents the congested flow density  $\rho^{cr} = 21 \text{ veh/km/lane}$ . Finally,  $\rho^{jam}$  is used to calculate the capacity drop coefficient  $q^{drop} = 12\%$ .



**Figure 6.** Traffic flow density scatter diagram.

## 5. Results and Discussion

To evaluate the effectiveness of the proposed FM-ITC-Darea strategy, benchmark schemes were established, including no control, the SPSC-LCC integrated control strategy based on the simple proportional speed controller (SPSC) algorithm, and the FM-ITC strategy using multi-agent reinforcement learning for static control areas. These benchmarks illustrate the advantages of the FM-ITC-Darea strategy with dynamic control areas. Below, the basic experimental setup is introduced, and the training process of FM-ITC-Darea is analyzed.

### 5.1. Algorithm Training Settings

In the experimental traffic flow settings, considering that highway traffic flow originates from the upstream mainline and on-ramps, the mainline traffic demand generally ranges from 3000 to 7000 vehicles per hour (veh/h), while the ramp demand generally ranges from 500 to 1200 veh/h. To simulate real-world traffic conditions, the simulation environment is set to a medium flow scenario, with a total traffic demand of 5100 veh/h.

Considering that real-world traffic demand typically undergoes a stochastic variation process, random noise is defined as fluctuations around the traffic demand that do not exceed 10% of the maximum traffic demand. During the first 100 time steps of the simulation, the mainline traffic demand continuously increases, eventually stabilizing around 1700 veh/h/lane. The demands at both on-ramps are set to be the same, fluctuating around 800 veh/h/lane.

The experiment sets the maximum training epochs for the RL algorithm to 1000. The algorithm terminates training when it reaches the maximum number of training rounds or a convergence condition. During each training epoch, the simulation steps in the CTM simulation environment include an environment warm-up period and a formal training period. At the beginning of each training session, the CTM initializes the simulation environment and warms up the traffic operation environment within 20 simulation steps. Then, over the following 600 simulation steps, agents interact with the stable CTM model environment to obtain observational information. In the traffic control system, the control cycle is set to 3 min, with the integrated controller updating the speed limits and lane change transition coefficients for each lane every 3 min.

Additionally, by comparing the effects of different experiments, the parameters related to OU noise were selected as  $\theta = 0.15$ ,  $\mu = 0$ ,  $\sigma = 0.4$ .

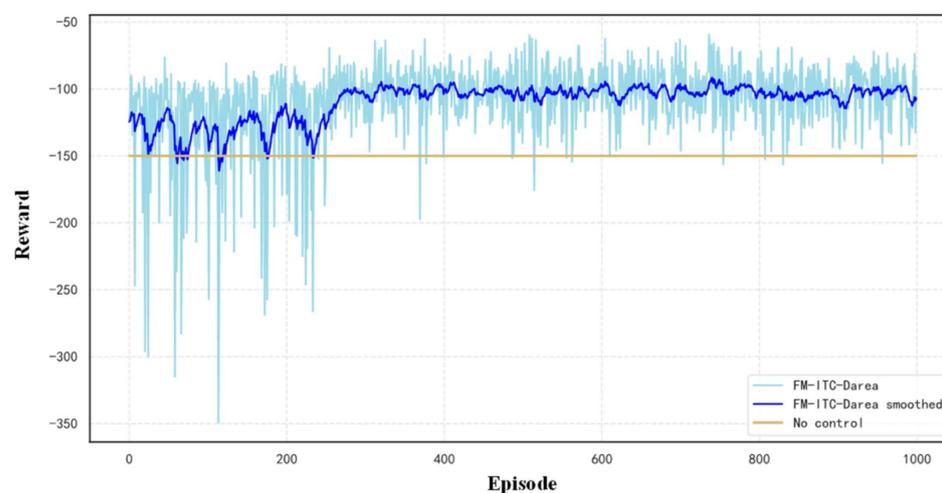
In Table 1, the settings for the algorithm-related parameters are provided:

**Table 1.** Parameter list of FACMAC-based ramp merge integration control methods.

Parameter	Value
Number of agents	4
Training epochs	1000
Testing epochs	10
Simulation steps	620
Effective simulation duration	7200 s
Experience replay buffer capacity	4000
Network training batch size	64
Discount factor	0.99
Actor learning rate	0.0003
Critic learning rate	0.003
Soft update parameter	0.01

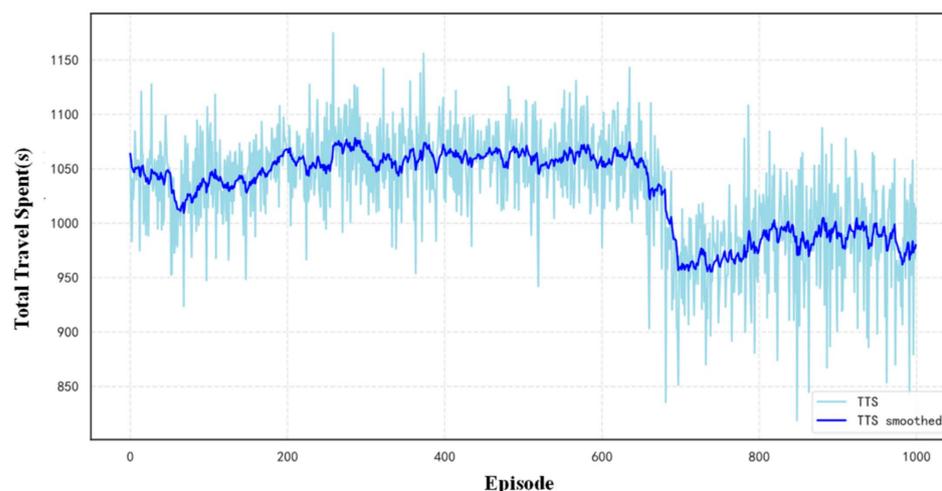
### 5.2. Training Process

Below, the training process of the FM-ITC-Darea algorithm is analyzed, and its convergence is evaluated. The simulation experiments, running 1000 training epochs on a computer with an AMD Ryzen 5600X six-core 3.7 GHz processor and an NVIDIA 1660S graphics card, take approximately 20 to 30 min. The training process is illustrated in Figure 7.

**Figure 7.** Reward curves for FM-ITC-Darea training processes.

For comparison, this paper also considers the scenario without control in the current simulation environment, where the reward value is calculated to be  $-150$ , as indicated by the yellow line in Figure 7. From the training of the agent in the FM-ITC-Darea strategy, it can be observed that during the first 200 training episodes, the agent continuously learns from the environment. Due to the randomness of the environment, there are significant oscillations during this period. After sufficient exploration, the agent shows an upward trend after 200 episodes and converges around  $-100$ . The final reward of the strategy shows an improvement compared to the no-control scenario.

To further demonstrate the effectiveness of the strategy, Figure 8 shows the variation of TTS during the training process. Compared to the initial stages of training, the FM-ITC-Darea strategy, after training, reduced the TTS by approximately 9.0%, indicating that the strategy can, overall, improve traffic efficiency.



**Figure 8.** TTS convergence curves for FM-ITC-Darea training processes.

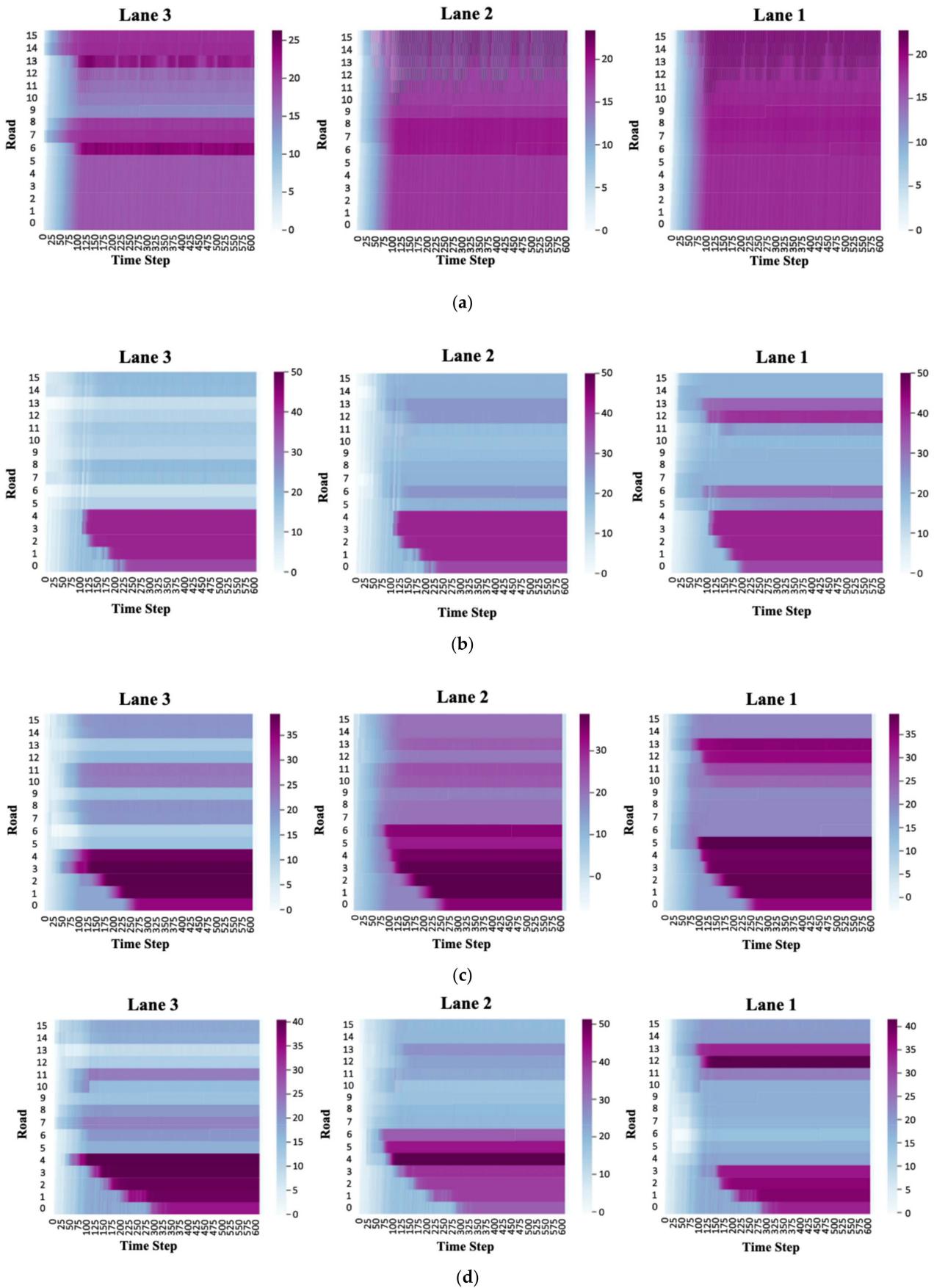
### 5.3. Comparison of Traffic Conditions in the Merge Bottleneck Area

In the study, the primary cause of road congestion is the emergence of bottlenecks at the on-ramp merge points, which subsequently spread to upstream areas. Therefore, a comparative analysis of the traffic conditions for the overall road and the bottleneck areas was conducted.

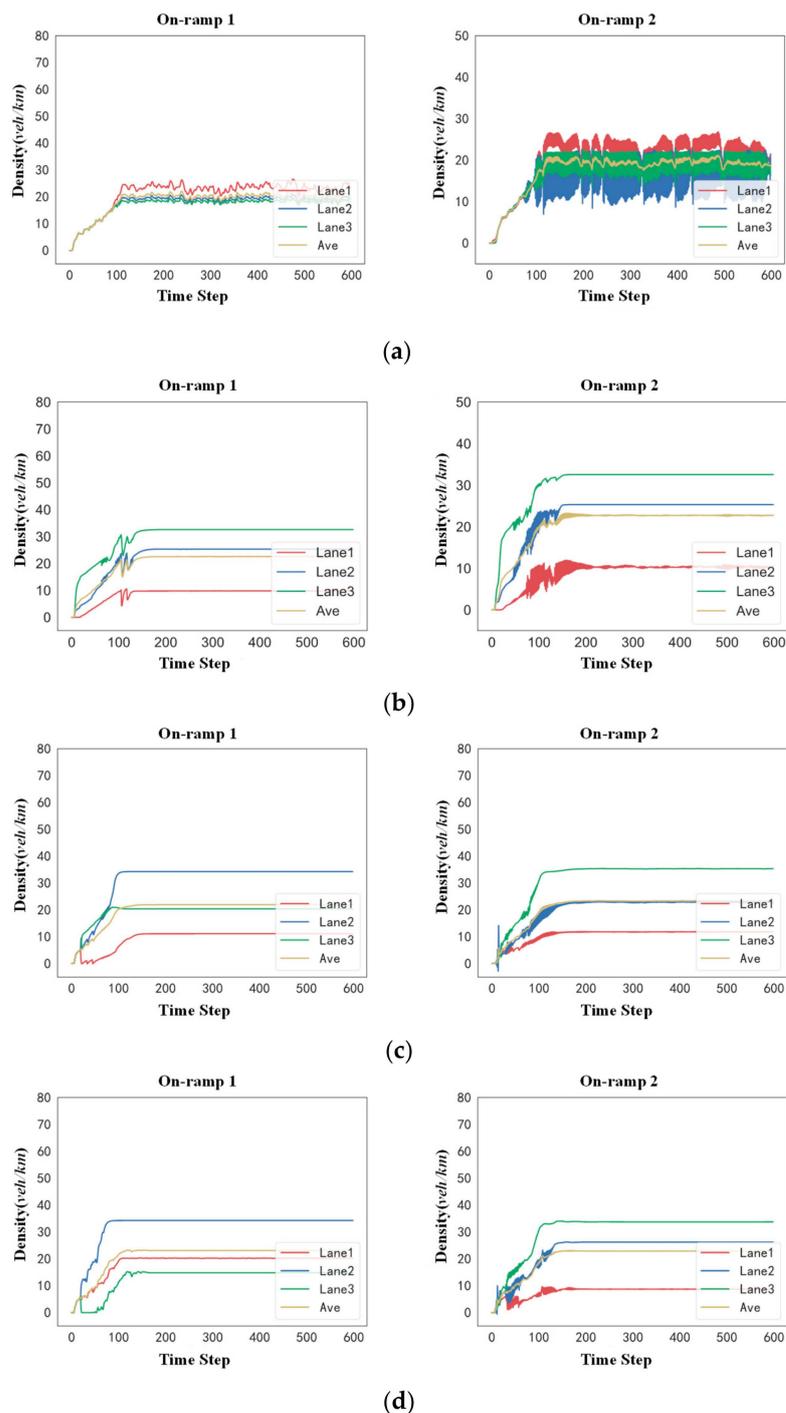
Figure 9 shows the spatiotemporal density variation of the overall mainline traffic, where Lane 1 represents the lane adjacent to the on-ramp. It can be observed that in the no-control scenario, the convergence of traffic from the on-ramp leads to congestion before the on-ramp, which then spreads to the upstream road sections and adjacent lanes. Under the SPSC-LCC strategy, the congestion in the control area before on-ramps is significantly alleviated; however, some congestion still occurs in the upstream area.

In contrast, under the FM-ITC and FM-ITC-Darea control strategies, the integrated control method imposes speed limits on the upstream traffic flow and directs vehicles within the control area to move preferentially toward the inner lanes. This reduces the density of the outer lanes and increases the utilization of the inner two lanes. Overall, the spatiotemporal impact of traffic congestion caused by the bottleneck spread is significantly reduced.

Figures 10 and 11 provide a more detailed analysis of the bottleneck traffic conditions at two on-ramps (i.e., cells 7 and 14). Due to the convergence of traffic from the on-ramps, the density of the outer lanes (i.e., Lane 1) at both on-ramps exceeds the critical density in the no control scenario, affecting the downstream traffic flow. The traffic fluctuations are more pronounced at cell 14 due to the presence of upstream on-ramps. By observing the simulation results of the three control strategies, it is evident that integrated control tends to eliminate traffic bottlenecks in the outer lanes. Although this increases the traffic load on the inner lanes (particularly Lane 2) by directing lane changes, the local objectives are achieved. This approach mitigates the negative impact of merging traffic to some extent and stabilizes the fluctuating traffic flow at cell 14. Given these observations, it can be seen that the SPSC-LCC strictly adheres to its logical settings, aiming to regulate each lane within the control range to the desired density and flow state. However, this setting sometimes has adverse effects on roads outside the control range (e.g., causing new traffic congestion).

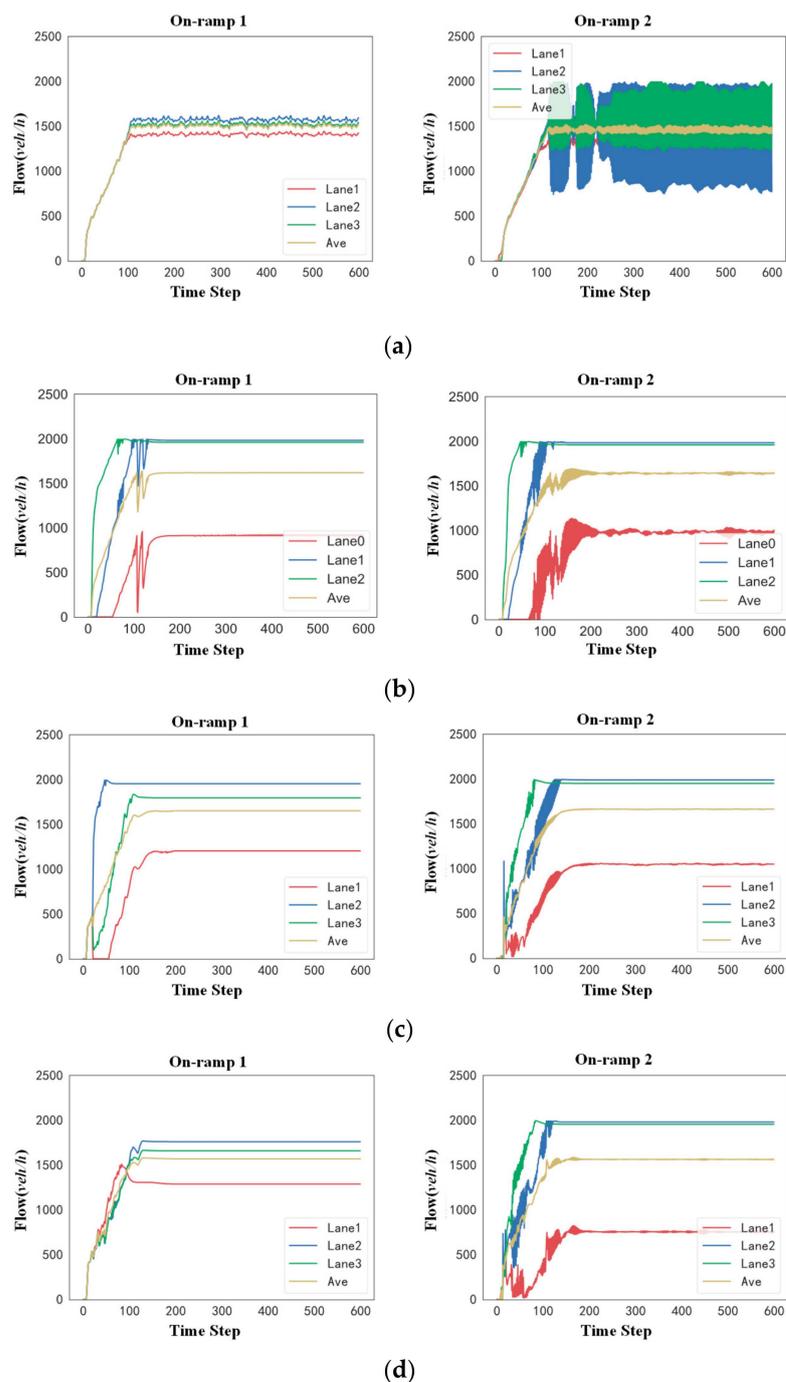


**Figure 9.** Overall spatial-temporal density of each lane of the mainline under different strategies, where (a) denotes the no control scenario, (b) denotes the SPSC-LCC strategy scenario, (c) denotes the FM-ITC strategy scenario, and (d) denotes the FM-ITC-Darea strategy scenario.



**Figure 10.** Changes in mainline density in ramp merging areas under different integrated control strategies, where (a) denotes the no-control scenario, (b) denotes the SPSC-LCC strategy scenario, (c) denotes the FM-ITC strategy scenario, and (d) denotes the FM-ITC-Darea strategy scenario.

Compared to the SPSC-LCC strategy, the FM-ITC and FM-ITC-Darea control strategies have a more flexible perception of road traffic conditions, enhancing the effectiveness of control commands and reducing traffic fluctuations caused by control commands lagging behind actual traffic conditions. FM-ITC-Darea exhibits similar effects to FM-ITC in controlling local traffic, but has the added advantage of determining the control area. This means it is not limited to achieving control objectives for the outer lanes but is more inclined to optimize traffic conditions within the entire effective bottleneck range of the on-ramps.



**Figure 11.** Changes of mainline flow in ramp merging areas under different integrated control strategies, where (a) denotes the no-control scenario, (b) denotes the SPSC-LCC strategy scenario, (c) denotes the FM-ITC strategy scenario, and (d) denotes the FM-ITC-Darea strategy scenario.

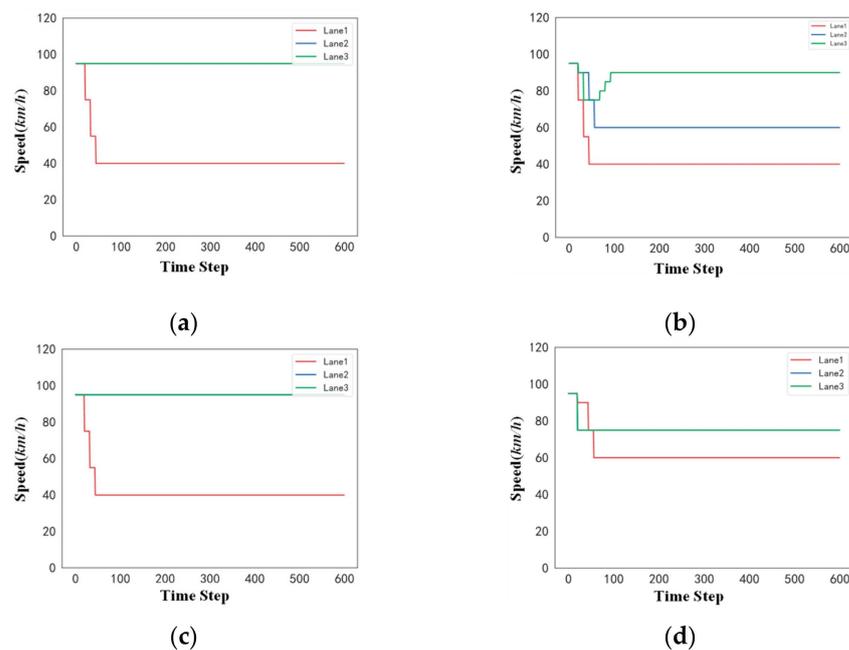
Table 2 presents the overall evaluation metrics under different control strategies. According to the simulation results, compared to the no-control scenario, the SPSC-LCC strategy is less effective in alleviating congestion. This is because SPSC-LCC only monitors flow in the local on-ramp merge areas and controls traffic at the merge bottlenecks only after congestion occurs, resulting in poorer performance on overall metrics. On the other hand, the FM-ITC and FM-ITC-Darea control strategies show improvements in both TTS and total vehicle turnover (TVT) metrics. Compared to integrated control in a static control area, the FM-ITC-Darea strategy, by dynamically adjusting the control area, can more effectively reduce the overall traffic travel time.

**Table 2.** Comparison of traffic performance indicators for different control strategies.

Control Strategy	Simulation Results		Improvement	
	TTS (s)	TVT (veh)	TTS (%)	TVT (%)
No control	1070	70,862	-	-
SPSC-LCC	1098	71,768	-2.6	1.3
FM-ITC	990	72,493	7.5	2.3
FM-ITC-Darea	960	72,759	10.3	2.7

5.4. Comparison of Integrated Controller Actions

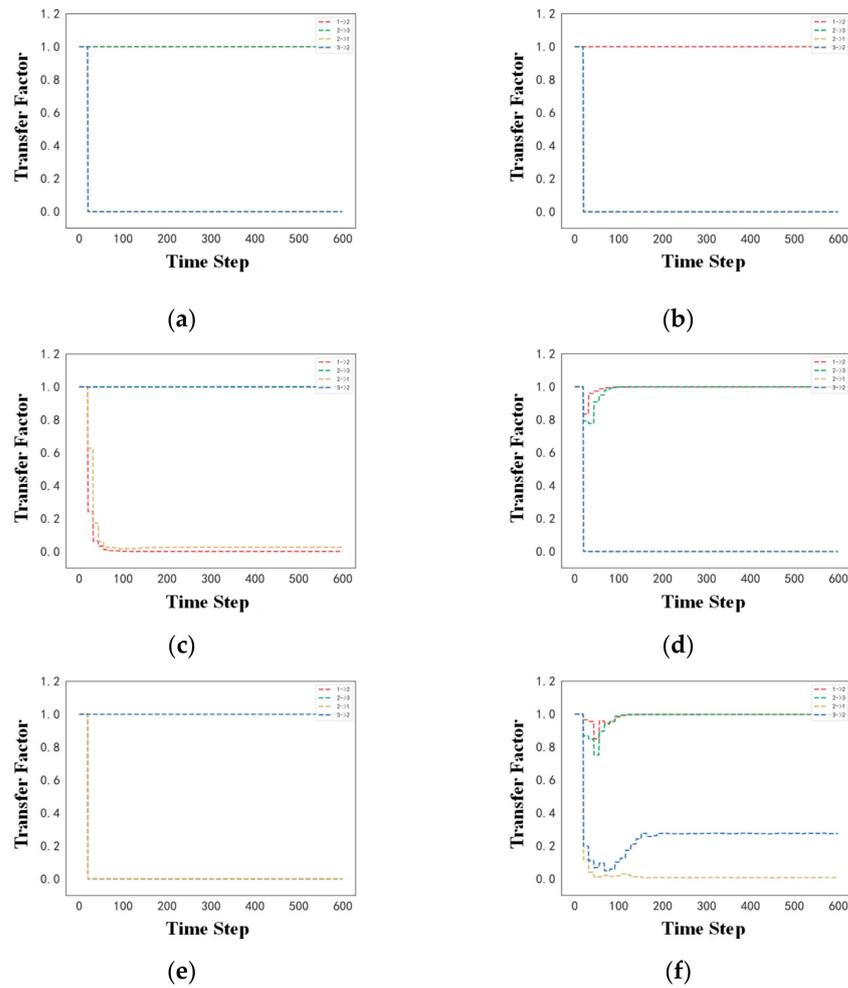
To better compare the differences in strategy selection between the FM-ITC and FM-ITC-Darea algorithm frameworks, the changes in speed limit values, lane-changing coefficients, and control region lengths under two strategies are, respectively, shown in Figure 12. It can be observed that both control strategies tend to impose lower speed limits on the outer lane (Lane 1) to reduce the bottleneck caused by merging. These speed limits are often set at the minimum speed limit, indicating that under higher traffic demand conditions, the FM-ITC with a fixed control region may not be able to achieve the most effective control (i.e., the traffic speed within the speed limit area can even be lower than the speed limit value). In response to different traffic scenarios, the FM-ITC-Darea can avoid this issue by dynamically adjusting the length of the control region. It is evident that an integrated control system using a dynamic control region can adopt more lenient speed limits, thereby reducing the interference of speed limits on overall traffic.



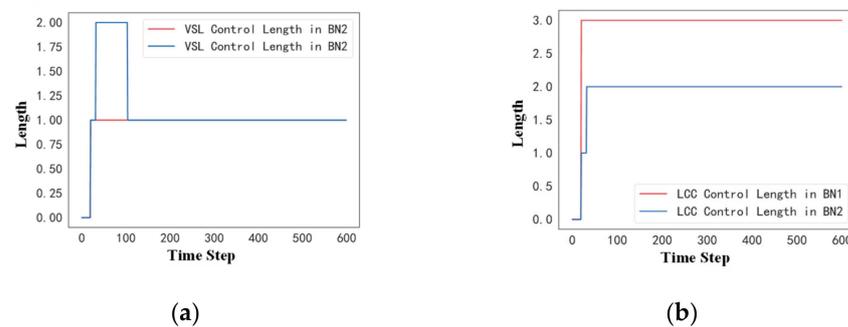
**Figure 12.** Variation of speed limit values for different integrated control strategies, where (a) and (b) correspond to FM-ITC speed limit changes at entrance ramps 1 and 2, respectively, and (c) and (d) correspond to FM-ITC-Darea speed limit changes at entrance ramps 1 and 2, respectively.

Figure 13 shows the lane transfer coefficients of the cells before the two on-ramps under the FM-ITC and FM-ITC-Darea strategies. It can be observed that before Ramp 1, both strategies tend to close the possibility of flow from the inner lanes to the outer lanes. By adjusting the flow from the outer lanes to the inner lanes, they alleviate the traffic conflicts between the outer lanes and the on-ramps. Before Ramp 2, the LCC agent in the FM-ITC-Darea strategy allows some flow from Lane 3 to Lane 2. This is because the LCC

agent’s control range expands under dynamic control areas (see Figure 14b), preventing more severe bottlenecks in Lane 3.



**Figure 13.** Variation of lane transfer factor for different integrated control strategies, where (a–d) denote the lane transfer coefficients of the two entrance ramp merge pre-cells under the FM-ITC strategy, and (e,f) denote the lane transfer coefficients of the two entrance ramp merge pre-cells under the FM-ITC-Darea strategy.



**Figure 14.** Variation of dynamic control area length for different integrated control strategies. where, (a) represents the variation in the control length of VSL agents at different bottlenecks, and (b) represents the variation in the control length of LCC agents at different bottlenecks.

By observing the changes in the dynamic control range actions of the VSL (a) and LCC (b) agents in Figure 14, it can be seen that under the current demand level, the integrated control strategy often prioritizes the adjustment of vehicle transfers and control

ranges of the outer lanes. This not only alleviates the formation of traffic bottlenecks in the outer lanes but also, to some extent, avoids the VSL speed limits from affecting the overall traffic efficiency.

In our experiments, both the VSL agent and the LCC agent in the integrated control strategy framework have the ability to dynamically adjust their control ranges. For the VSL agent, this ability helps avoid the failure of speed limits due to excessive congestion within a limited control range. For the LCC agent, it allows prioritizing the extension of the control range length to better optimize the congested traffic flow in the ramp merge areas. The experimental results support this hypothesis: under medium traffic demand, the speed limits set by the VSL agent generally do not reach the minimum speed limit of 40 km/h, thereby avoiding secondary congestion. When traffic demand changes, dynamically configuring the control area range enables the algorithm to flexibly decide whether to activate alternative control areas to optimize speed limits or lane transfer coefficients. This flexible control strategy is difficult to achieve with traditional traffic control strategies like those primarily based on SPSC.

However, deciding whether to adjust a larger control area or set more extreme control commands presents a challenge for the agents. When the number of training iterations is insufficient, or the training falls into a suboptimal solution, the agents in FM-ITC-Darea might issue less effective control commands compared to the static control area strategy, FM-ITC. Additionally, although using MARL-based integrated control strategies in multi-ramp merging scenarios can effectively handle randomly changing traffic conditions, it sometimes struggles to balance the optimization of local ramp areas with overall traffic flow. Determining the control priorities of the agents can also be challenging, potentially affecting the convergence of training. To address these shortcomings, future research could further explore the trade-offs between integrated control strategies for overall and local traffic optimization and the feasibility of implementing dynamic control areas.

## 6. Conclusions

This paper improves the FACMAC algorithm by converting the VSL and LCC integrated control strategies into a distributed partially observable Markov decision process. Heterogeneous VSL and LCC agents were designed, and a multi-ramp merging integrated control framework based on FACMAC was proposed. The study considers the introduction of the dynamic control area concept under the premise of a vehicular network environment. By granting VSL and LCC agents the ability to configure spatially, control issues such as speed limit failures are avoided, enhancing the adaptability of control strategies to different traffic environments. Testing the proposed integrated control strategy in a multi-ramp scenario demonstrated its effectiveness in solving merging bottleneck problems and verified the significant effect of dynamic control areas in alleviating traffic pressure, proving the potential and effectiveness of the improved integrated control scheme. To further demonstrate the practicality of the proposed control framework, Figure A1 in Appendix A presents a schematic diagram detailing the framework's process and implementation plan. This diagram encompasses the entire workflow, ranging from real-world data acquisition and simulation environment construction to agent training and testing, as well as the validation of control strategies.

In conclusion, this paper not only provides an effective intelligent traffic integrated control solution for the highway ramp merging bottleneck problem but also introduces the innovative concepts of MARL algorithms and dynamic traffic control areas into the control system. This offers valuable theoretical and practical guidance for the design of future intelligent traffic systems and traffic flow management.

**Author Contributions:** Conceptualization, J.D.; methodology, J.D. and Q.J.; validation, Q.J. and H.Z.; formal analysis, J.D. and Q.J.; investigation, A.Y.; resources, J.D.; data curation, Q.J.; writing—original draft preparation, A.Y., H.Z. and X.B.; writing—review and editing, A.Y., H.Z. and X.B.; visualization, A.Y.; supervision, J.D.; project administration, J.D.; funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is funded by the National Natural Science Foundation of China (grant number 72471133).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

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

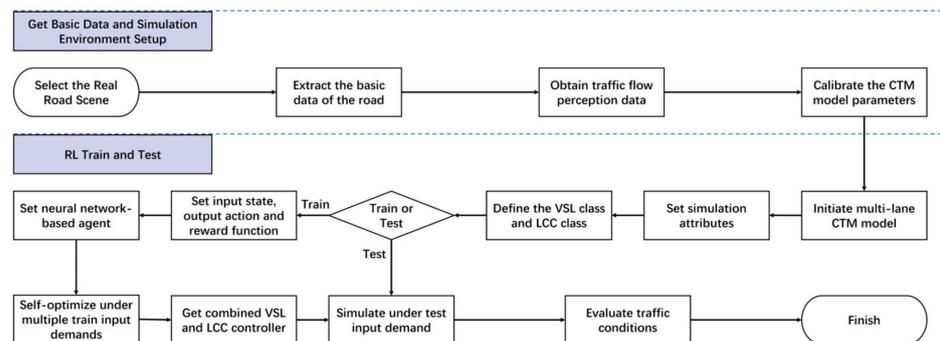
**Conflicts of Interest:** The authors declare no conflicts of interest.

### Abbreviations

The following are the abbreviations used in the manuscript.

ITS	Intelligent Transportation Systems.
VSL	Variable Speed Limits.
LCC	Lane Change Control.
Dec-POMDP	Decentralized Partially Observable Markov Decision Process.
CTM	Cell Transmission Model.
MARL	Multi-Agent Reinforcement Learning.
FACMAC	Factored Multi-Agent Centralized Policy Gradients.
FM-ITC	Factored Multi-Agent Centralized Policy Gradients for Integrated Traffic Control.
FM-ITC-Darea	Factored Multi-Agent Centralized Policy Gradients for Integrated Traffic Control in Dynamic Areas.
MPC	Model Predictive Control.
RM	Ramp Metering.
VSL-RM	Variable Speed Limits with Ramp Metering.
V2X	Vehicle-to-Everything.
RL	Reinforcement Learning.
CTDE	Centralized Training and Decentralized Execution.
MADDPG	Multi-Agent Deep Deterministic Policy Gradient.
OU	Ornstein–Uhlenbeck.
TTS	Total Travel Spend.
SPSC	Simple Proportional Speed Controller.
TVT	Total Vehicle Turnover.

### Appendix A



**Figure A1.** Flowchart of the proposed integrated control framework for application in real-world scenarios.

## References

1. Oh, S.; Yeo, H. Estimation of Capacity Drop in Highway Merging Sections. *Transp. Res. Rec.* **2012**, *2286*, 111–121. [[CrossRef](#)]
2. Zhang, J.; Wang, F.-Y.; Wang, K.; Lin, W.-H.; Xu, X.; Chen, C. Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [[CrossRef](#)]
3. Farahani, R.Z.; Miandoabchi, E.; Szeto, W.Y.; Rashidi, H. A review of urban transportation network design problems. *Eur. J. Oper. Res.* **2013**, *229*, 281–302. [[CrossRef](#)]
4. Gkyrtis, K.; Kokkalis, A. An Overview of the Efficiency of Roundabouts: Design Aspects and Contribution toward Safer Vehicle Movement. *Vehicles* **2024**, *6*, 433–449. [[CrossRef](#)]
5. Papageorgiou, M.; Kotsialos, A. Freeway ramp metering: An overview. *IEEE Trans. Intell. Transp. Syst.* **2002**, *3*, 271–281. [[CrossRef](#)]
6. Zhang, Y.; Ioannou, P.A. Combined Variable Speed Limit and Lane Change Control for Highway Traffic. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1812–1823. [[CrossRef](#)]
7. Lee, C.; Hellinga, B.; Saccomanno, F. Assessing Safety Benefits of Variable Speed Limits. *Transp. Res. Rec.* **2004**, *1897*, 183–190. [[CrossRef](#)]
8. Yu, M.; Fan, W. “David” Optimal Variable Speed Limit Control in Connected Autonomous Vehicle Environment for Relieving Freeway Congestion. *J. Transp. Eng. Part Syst.* **2019**, *145*, 04019007. [[CrossRef](#)]
9. Papamichail, I.; Papageorgiou, M.; Stamatakis, I. Feedback Traffic Control at Highway Work Zones using Variable Speed Limits. *IFAC-Pap.* **2018**, *51*, 329–336. [[CrossRef](#)]
10. Hegyi, A.; Schutter, B.D.; Hellendoorn, J. Optimal coordination of variable speed limits to suppress shock waves. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 102–112. [[CrossRef](#)]
11. Han, Y.; Wang, M.; He, Z.; Li, Z.; Wang, H.; Liu, P. A linear Lagrangian model predictive controller of macro- and micro- variable speed limits to eliminate freeway jam waves. *Transp. Res. Part C Emerg. Technol.* **2021**, *128*, 103121. [[CrossRef](#)]
12. Han, Y.; Hegyi, A.; Yuan, Y.; Hoogendoorn, S.; Papageorgiou, M.; Roncoli, C. Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 405–420. [[CrossRef](#)]
13. Carlson, R.C.; Papamichail, I.; Papageorgiou, M.; Messmer, A. Optimal mainstream traffic flow control of large-scale motorway networks. *Transp. Res. Part C Emerg. Technol.* **2010**, *18*, 193–212. [[CrossRef](#)]
14. Soriguera, F.; Martínez, I.; Sala, M.; Menéndez, M. Effects of low speed limits on freeway traffic flow. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 257–274. [[CrossRef](#)]
15. Ma, M.; Yang, Q.; Liang, S.; Li, Z. Integrated Variable Speed Limits Control and Ramp Metering for Bottleneck Regions on Freeway. *Math. Probl. Eng.* **2015**, *2015*, e313089. [[CrossRef](#)]
16. Fang, J.; Hadiuzzaman, M.; Yin, E.; Qiu, T.Z. DynaTAM: An online algorithm for performing simultaneously optimized proactive traffic control for freeways. *Can. J. Civ. Eng.* **2014**, *41*, 315–322. [[CrossRef](#)]
17. Guo, Y.; Xu, H.; Zhang, Y.; Yao, D. Integrated Variable Speed Limits and Lane-Changing Control for Freeway Lane-Drop Bottlenecks. *IEEE Access* **2020**, *8*, 54710–54721. [[CrossRef](#)]
18. Markantonakis, V.; Skoufoulas, D.I.; Papamichail, I.; Papageorgiou, M. Integrated Traffic Control for Freeways using Variable Speed Limits and Lane Change Control Actions. *Transp. Res. Rec.* **2019**, *2673*, 602–613. [[CrossRef](#)]
19. Kattan, L.; Khondaker, B.; Derushkina, O.; Poosarla, E. A Probe-Based Variable Speed Limit System. *J. Intell. Transp. Syst.* **2015**, *19*, 339–354. [[CrossRef](#)]
20. Kušić, K.; Ivanjko, E.; Gregurić, M.; Miletić, M. An Overview of Reinforcement Learning Methods for Variable Speed Limit Control. *Appl. Sci.* **2020**, *10*, 4917. [[CrossRef](#)]
21. Zhu, F.; Ukkusuri, S.V. Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *41*, 30–47. [[CrossRef](#)]
22. Genders, W.; Razavi, S. Using a Deep Reinforcement Learning Agent for Traffic Signal Control. *arXiv* **2016**. [[CrossRef](#)]
23. Zhou, Y.; Ozbay, K.; Kachroo, P.; Zuo, F. Ramp Metering for a Distant Downstream Bottleneck Using Reinforcement Learning with Value Function Approximation. *J. Adv. Transp.* **2020**, *2020*, e8813467. [[CrossRef](#)]
24. Belletti, F.; Haziza, D.; Gomes, G.; Bayen, A.M. Expert Level Control of Ramp Metering Based on Multi-Task Deep Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1198–1207. [[CrossRef](#)]
25. Bazzan, A.L.C. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. *Auton. Agents Multi-Agent Syst.* **2009**, *18*, 342–375. [[CrossRef](#)]
26. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11–32. [[CrossRef](#)]
27. Wang, C.; Zhang, J.; Xu, L.; Li, L.; Ran, B. A New Solution for Freeway Congestion: Cooperative Speed Limit Control Using Distributed Reinforcement Learning. *IEEE Access* **2019**, *7*, 41947–41957. [[CrossRef](#)]
28. Zheng, S.; Li, M.; Ke, Z.; Li, Z. Coordinated Variable Speed Limit Control for Consecutive Bottlenecks on Freeways Using Multiagent Reinforcement Learning. *J. Adv. Transp.* **2023**, *2023*, e4419907. [[CrossRef](#)]

29. Wang, C.; Xu, Y.; Zhang, J.; Ran, B. Integrated Traffic Control for Freeway Recurrent Bottleneck Based on Deep Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15522–15535. [[CrossRef](#)]
30. Schmidt-Dumont, T.; van Vuuren, J.H. A case for the adoption of decentralised reinforcement learning for the control of traffic flow on South African highways. *J. S. Afr. Inst. Civ. Eng.* **2019**, *61*, 7–19. [[CrossRef](#)]
31. Ko, B.; Ryu, S.; Park, B.B.; Son, S.H. Speed harmonisation and merge control using connected automated vehicles on a highway lane closure: A reinforcement learning approach. *IET Intell. Transp. Syst.* **2020**, *14*, 947–957. [[CrossRef](#)]
32. Peng, C.; Xu, C. Combined variable speed limit and lane change guidance for secondary crash prevention using distributed deep reinforcement learning. *J. Transp. Saf. Secur.* **2022**, *14*, 2166–2191. [[CrossRef](#)]
33. Busoniu, L.; Babuska, R.; De Schutter, B. A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2008**, *38*, 156–172. [[CrossRef](#)]
34. Hernandez-Leal, P.; Kartal, B.; Taylor, M.E. A survey and critique of multiagent deep reinforcement learning. *Auton. Agents Multi-Agent Syst.* **2019**, *33*, 750–797. [[CrossRef](#)]
35. Da Silva, F.L.; Glatt, R.; Costa, A.H.R. MOO-MDP: An Object-Oriented Representation for Cooperative Multiagent Reinforcement Learning. *IEEE Trans. Cybern.* **2019**, *49*, 567–579. [[CrossRef](#)] [[PubMed](#)]
36. Ke, Z.; Li, Z.; Cao, Z.; Liu, P. Enhancing Transferability of Deep Reinforcement Learning-Based Variable Speed Limit Control Using Transfer Learning. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4684–4695. [[CrossRef](#)]
37. Jiang, Q. Improving Traffic Efficiency at Merge Bottleneck: A Combined Variable Speed Limit and Logic Rule-Based Lane Change Control Approach. In Proceedings of the 7th International Conference on Traffic Engineering and Transportation System (ICTETS 2023); Dalian, China, 22–24 September 2023; Volume 13064. [[CrossRef](#)]

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