Hierarchical Control Strategy for Cooperative On-Ramp Merging of Connected and Automated Vehicles on Multilane Highways

Rui Peng, Min Yang, Rui Tao, Mingye Zhang, and Renjie Zhang

***Abstract*—In mixed traffic environments, fully utilizing the traffic capacity of the multiple lanes on the inner side of highways and coordinating connected and automated vehicles (CAVs) to complete merging is an extremely challenging task. This article proposes a hierarchical cooperative merging control strategy for CAVs on multilane highways. It comprises two layers. The premerging lane-changing layer aims to guide CAVs from the outer lanes of the mainline to the inner lanes. We establish a CAV lane allocation model and propose a rule-based CAV lane-changing strategy for both free and cooperative lane-changing scenarios. In the merging control layer, we adopt a hierarchical reward optimization multiagent deep deterministic policy gradient (HRO-MADDPG) algorithm to finely optimize the merging trajectories of CAVs. To improve the convergence performance of the merging task algorithm, global and local rewards are set by considering both the overall objectives of the merging control area and the individual objectives of CAVs. A three-lane highway is used as a case study. The results indicate that the proposed strategy effectively promotes the uniform distribution of traffic flow across the mainline lanes, significantly relieving merging pressure between the outermost lane and the ramp. HRO-MADDPG outperforms existing algorithms in terms of convergence speed, efficiency, and safety. The proposed strategy also effectively improves traffic efficiency under different CAV penetration rates and ramp merging rates, providing a reference for future merging management strategies of multilane highways under mixed traffic.**

***Index Terms*—Connected and autonomous vehicles (CAVs) , mixed traffic, multilane traffic, on-ramp merging, reinforcement learning (RL).**

I. Introduction

A. Motivation

With the widespread construction of highways and rapid growth of vehicle ownership, congestion in merging areas has become increasingly severe [1]. Particularly in multilane highway merging scenarios, mainline vehicles often change lanes randomly to avoid merging areas, while ramp vehicles must quickly merge into the outermost lane, resulting in notable conflicts. The development of connected and automated vehicle (CAV) technology provides innovative solutions [2]. By leveraging advanced communication systems, CAVs can interact with nearby vehicles and infrastructure, enhancing the efficiency and safety of merging operations.

Currently, research on CAV merging control mainly focuses on rule-based, optimization, and reinforcement learning (RL) methods [3]. Rule-based or optimization methods typically rely on strict models and complex calculations, which result in insufficient adaptability and real-time performance in dynamic environments. RL has shown significant effectiveness in solving sequential decision problems due to its remarkable self-learning capability [4]. In recent years, it has been widely applied in traffic control, especially in mixed traffic environments consisting of human-driven vehicles (HVs) and CAVs [5].

When applying RL to control the merging of vehicles from mainline multiple lanes and ramp, it also involves the coordinated lane-changing between vehicles in the inner lanes and those in the outermost lane. When traffic density in the outermost lane is high, vehicles in the inner lanes give way to allow some vehicles to move into the outermost lane [6]. While this strategy performs well under low traffic conditions, as traffic volume increases, the exploration space of the model grows exponentially, making it difficult for the learning process to converge, thereby affecting experimental accuracy.

This issue has led us to shift our research focus from the merging of mainline multiple lanes and ramp vehicles to the merging of mainline outermost lane and ramp vehicles. This shift not only effectively reduces the number of agents in RL but also narrows the action exploration space. However, current research in this area mostly fails to fully utilize the overall capacity of multilane highways [7]. The mainline outermost lane often faces excessive traffic pressure from ramp vehicles, while the inner lanes experience relatively lower traffic volumes, resulting in a significant imbalance in lane utilization.

To further alleviate the merging pressure between the mainline outermost lane and ramp vehicles, some researchers have attempted to divide the merging task into two layers. The first layer involves setting up a section upstream of the merging

Received 4 May 2025; revised 27 May 2025; accepted 6 June 2025. Date of publication 10 June 2025; date of current version 8 August 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 52432011 and Grant 524B2153, and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX24_0450. *(Corresponding author: Min Yang.)*

Rui Peng, Min Yang, Mingye Zhang, and Renjie Zhang are with the School of Transportation, Southeast University, Nanjing 211189, China (e-mail: 230228875@seu.edu.cn; yangmin@seu.edu.cn; zhangmingye@seu.edu.cn; 230228877@seu.edu.cn).

Rui Tao is with the School of Civil and Transportation Engineering, Hebei University of Technology, Tianjin 300401, China (e-mail: taorui_95@126.com).

Digital Object Identifier 10.1109/JIOT.2025.3578530

area to guide vehicles in the outermost lane to move into the inner lanes in advance, thereby reducing the traffic pressure in the outermost lane. The second layer focuses on controlling the merging of mainline outermost lane and ramp vehicles [7], [8], [9], [10]. This method has been proven to effectively improve the traffic conditions in the merging area, with rule-based, optimization, and RL algorithms being applied at different layers of the tasks.

Existing hierarchical control strategies are mainly focused on pure CAV environments, where the merging control in the second layer task primarily adopts rule-based or optimization methods. In mixed traffic environments, the uncertainty and dynamic behavior of HVs limit the accuracy and stability of the model. In contrast, RL algorithms allow CAVs to autonomously learn the optimal merging strategy based on real-time traffic conditions. Therefore, RL is particularly suitable for application in the second-layer task in mixed traffic environments. However, current RL algorithms still struggle to balance the overall goals of the merging area with the individual goals of CAVs when handling merging tasks. When faced with the dual-reward problem, the complex interconnection of reward functions often leads to performance fluctuations and instability during the learning process [11]. The algorithm struggles to effectively distinguish and optimize global and local rewards during training, and this issue is particularly pronounced when the reward signal is sparse, often causing the algorithm to overly rely on one type of reward signal [12]. This results in oscillations between the two types of rewards, which affects the effectiveness of optimization [13].

In summary, there are still many unresolved issues regarding multilane highway merging control in mixed traffic environments. This article proposes a hierarchical cooperative merging control strategy for mixed traffic flows. To reduce the overall complexity of the control method, the control process for the premerging lane-changing (PLC) area is relatively simple and regular, which doesn't require complex strategy learning through RL as is needed in merging areas. Therefore, further research is necessary on lane allocation models and lane-changing strategies for CAVs under free and cooperative lane-changing scenarios. For the merging area, this article adopts a novel RL algorithm to address the issue of sacrificing local benefits to improve overall efficiency in merging control. The strategy aims to fully utilize the capacity of the inner lanes of the highway and maximize the traffic efficiency, safety, and learning convergence speed of the multiagent merging task.

B. Literature Review

In recent years, researchers have conducted in-depth studies on highway merging control using rule-based or optimization methods. Xue et al. [14] developed a platoon-based hierarchical algorithm with model predictive control (MPC) to guide vehicles into target merging gaps. Chen and Yang [15] introduced a cooperative strategy based on transferable utility and optimal control to coordinate merging trajectories. Zhang et al. [16] used MPC for trajectory optimization in CAV merging control. A decentralized approach was proposed by Jing et al. [17], where nonlinear MPC tracked optimal trajectories from the upper layer. Additionally, Chen et al. [18] formulated a mixed-integer nonlinear programming model for collaborative merging at highway on-ramp areas, while Ding et al. [19] defined a time window for ramp vehicles entering the mainline to make merging decisions. Hu et al. [20] proposed a critical trajectory point planning model, optimizing the merging task through nonlinear optimization. Rule-based or optimization methods have significant advantages in terms of solution speed, modeling simplicity, and computational efficiency. In the PLC area control discussed in this article, they not only effectively reduce the overall complexity of the hierarchical merging models but also provide valuable insights for designing rule-based lane allocation and lane-changing strategies.

The research on merging control in RL can be categorized into single-agent and multiagent scenarios. In the single-agent scenario, a variety of strategies have been proposed to improve merging efficiency and safety in dense traffic conditions. Brito et al. [21] developed an interaction-aware strategy using deep RL (DRL) to enable safe merging when other vehicles do not yield. To enhance the interpretability of CAVs merging control, Hu et al. [22] combined DRL algorithms with optimization-based methods. Hassani et al. [23] proposed a sample-efficient DRL algorithm with prioritized experience replay to better utilize merging trajectory samples. In the multiagent scenario, research has evolved to address the complexities of multiple vehicles interacting. Chen et al. [6] applied multiagent deep RL (MADRL) to multilane merging. He and Lv [24] introduced an energy-aware cooperative driving strategy based on MADRL for efficient and safe CAV merging. Zhang et al. [25] used the IPPO method with independent learning and parameter sharing to improve ramp merging decision-making. Li et al. [26] applied game theory to model competitive behaviors between ramp and mainline vehicles, developing a Nash double Q-based merging strategy. Li et al. [27] combined communication protocols with the soft actor-critic algorithm to generate optimal, collision-free merging trajectories. Wang et al. [28] proposed an end-to-end merging control framework using spatiotemporal deep Q-networks for multi-CAV decision-making. RL-based methods demonstrate significant advantages in handling complex scenarios, particularly in terms of scalability, adaptability, and dynamic decision-making. Existing research, by setting various rewards and penalties related to efficiency, safety, and driving comfort, provides valuable insights for handling the merging of vehicles from the mainline outermost lane and the ramp in this article.

Based on existing research on highway merging control using rule-based or optimization methods, as well as RL methods, hierarchical control strategies have also received significant attention in vehicle merging studies, particularly in multilane traffic. These strategies typically consist of multiple layers, with each layer focusing on different aspects of the merging process. Liu et al. [7] used RL for lane selection and applied time-energy optimal control for merging in multilane traffic. Chen et al. [29] developed a hierarchical control method for CAV merging, where the tactical layer determines the optimal merging order, and the operational layer uses MPC

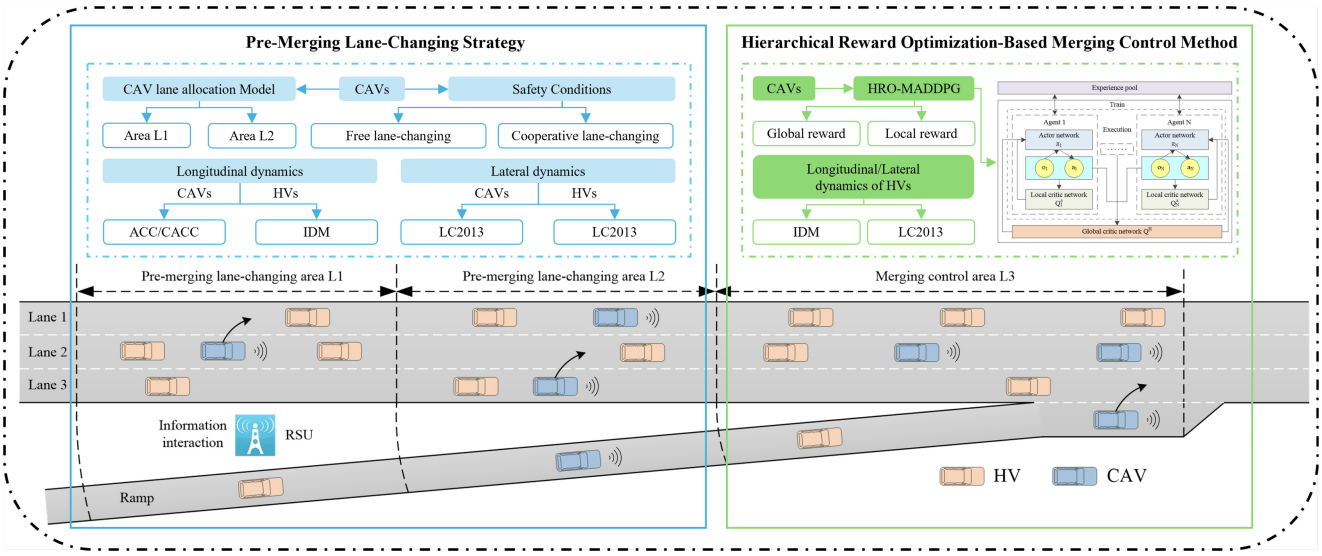


Fig. 1. Hierarchical cooperative merging architecture for multilane highways.

to optimize acceleration. Hou et al. [8] proposed a hierarchical model with an anticipated position search algorithm for determining the expected merging position and a cooperative merging control model to ensure smooth merging. Liu et al. [9] applied proximal policy optimization for optimal lane decision-making and developed an environmentally friendly merging optimization problem to adjust longitudinal speed. Chen and Yang [10] used cooperative game theory for lane allocation based on driving benefits and optimal control to calculate merging trajectories.

The hierarchical cooperative merging strategy proposed in this article integrates findings from various fields, not only filling existing research gaps but also providing an innovative approach to alleviating traffic congestion in merging areas.

C. Contribution of this Article

The main contributions of this article are as follows.

1) To enhance traffic efficiency and safety in the merging areas of multilane highways under mixed traffic flow, this article proposes a hierarchical cooperative merging control strategy for CAVs, which divides the merging area into two PLC sections and one merging control section. It can coordinate the lane distribution of mainline vehicles and finely optimize the merging trajectories of CAVs.
2) This article proposes a lane allocation model for mainline CAVs, determines the number of CAVs on the outer lanes that need to be guided to the inner lanes in different PLC sections, and establishes a rule-based CAV lane-changing strategy for both free and cooperative lane-changing scenarios. It addresses the problem of uneven distribution of traffic flow on multilane highways and the underutilization of the inner lanes.
3) This article adopts the HRO-MADDPG algorithm to address the problem of traditional single reward settings struggling to effectively distinguish the impact of agents on local traffic conditions in highway merging scenarios. By considering both the overall objectives of the merging control area and the individual objectives of CAVs to set global and local rewards, it enhances the convergence speed and accuracy of model training.

D. Organization of this Article

The rest of this article is organized as follows. In the second section, we introduce the methods used in this study. In the third section, we establish a traffic scenario and describe the experimental details. In the fourth section, we give the simulation results and evaluate the proposed strategy. In the fifth section, we conclude this article.

II. Methodology

This article proposes a cooperative merging control method for CAVs on multilane highways, which includes a PLC strategy and a merging trajectory optimization model. A three-lane highway serves as the case study. Specifically, the PLC strategy is applied in PLC areas L1 and L2 for CAVs. In the merging control area L3, a trajectory optimization model is employed for CAVs in the ramp and the outermost lane of the mainline. As shown in Fig. 1.

A. Premerging Lane-Changing Strategy for Mainline CAVs

For a three-lane highway, if traffic density is uneven across lanes and the inner lane has less traffic, it is possible to guide some vehicles from the outer lanes to the inner lane to reduce the traffic density in the outermost lane. It can also effectively improve the merging efficiency and safety of vehicles.

Considering the complexity of simultaneously coordinating CAVs from lane 2 to lane 1, and from lane 3 to lane 2, it may cause traffic flow disruptions, we propose a hierarchical coordination method. The PLC area is divided into two sections. The area L1 aims to guide some CAVs from lane 2 to lane 1, while the area L2 aims to guide some CAVs from

lane 3 to lane 2. To simplify the research scenario, we have made the following assumptions.

1) CAVs are equipped with advanced sensors and controllers to obtain information about surrounding vehicles via V2V and V2I communications.
2) Communication delays are not considered, and CAVs can receive control instructions from roadside units.
3) CAVs entering areas L1, L2, and L3 must strictly follow lane-changing and speed-changing instructions.
4) CAVs in the PLC area can only change lanes once and cannot change across multiple lanes.

1) CAV Lane Allocation Model: It is necessary to determine the number of vehicles that should be allocated to each lane. For area L1, the number of CAVs that need to change from lane 2 to lane 1 is as follows:

$$N_{2\to1}^{\mathrm{CAV}} = \begin{cases} 0, & \zeta \le 0 \\ \zeta, & 0 < \zeta \le N_2^{\mathrm{CAV}} \\ N_2^{\mathrm{CAV}}, & \zeta > N_2^{\mathrm{CAV}} \end{cases} \tag{1}$$

$$\zeta = \mathrm{ceil}\left(\frac{1}{3}(N_1 + N_2 + (N_3 + N_r))\right) - N_1 \tag{2}$$

where $N_{2\to1}^{\mathrm{CAV}}$ is the number of CAVs in lane 2 that should change to lane 1 within area L1. N_1, N_2, and N_3 represent the number of vehicles in lanes 1, 2, and 3 of area L1, respectively. N_r is the number of ramp vehicles within the projection distance from area L1 to the ramp. N_2^{CAV} is the number of CAVs in lane 2 of area L1. ζ represents the CAV lane allocation model for area L1. The ceil$(\cdot)$ function is used to round up to the nearest whole number.

Similarly, for area L2, we can determine the number of CAVs that need to change from lane 3 to lane 2.

$$N_{3\to2}^{\mathrm{CAV}} = \begin{cases} 0, & \zeta' \le 0 \\ \zeta', & 0 < \zeta' \le N_{3'}^{\mathrm{CAV}} \\ N_{3'}^{\mathrm{CAV}}, & \zeta' > N_{3'}^{\mathrm{CAV}} \end{cases} \tag{3}$$

$$\zeta' = \mathrm{ceil}\left(\frac{1}{2}\left(N_2' + \left(N_3' + N_r'\right)\right)\right) - N_2' \tag{4}$$

where $N_{3\to2}^{\mathrm{CAV}}$ is the number of CAVs in lane 3 that should change to lane 2 within area L2. N_2' and N_3' represent the number of vehicles in lanes 2 and 3 of area L2, respectively. N_r' is the number of ramp vehicles within the projection distance from area L2 to the ramp. $N_{3'}^{\mathrm{CAV}}$ is the number of CAVs in lane 3 of area L2. ζ' represents the CAV lane allocation model for area L2.

2) Safety Conditions for Premerging Lane-Changing: After calculating the number of CAVs that need to be guided in each lane, it is necessary to determine which CAVs meet the conditions for lane-changing, including scenarios of free lane-changing and cooperative lane-changing.

a) Free lane-changing scenario: As shown in Fig. 2, in a free lane-changing scenario, vehicle i decides to change to an adjacent lane, with vehicle $j-1$ and vehicle j being the downstream and upstream vehicles in the target lane, respectively. At any moment, vehicle i will not attempt to follow vehicle $j-1$ with a time gap less than the desired time gap. The safe distance gap G_i is defined as the maximum value between the following distance maintained by vehicle i through the desired time gap and the minimum standstill safe gap [19]. The minimum safe space headway H_i is as follows:

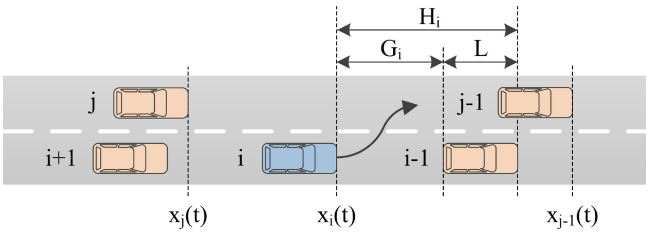


Fig. 2. CAV lane-changing diagram.

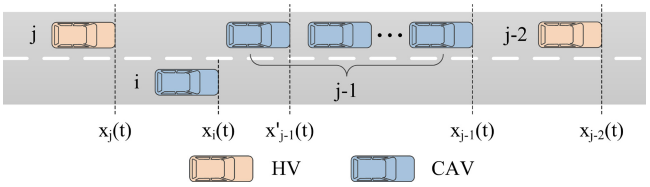


Fig. 3. Leading vehicle in the target lane is a CAV or a CAV platoon and the following vehicle is an HV.

$$G_i = \max\{v_i t_0, G_0\} \tag{5}$$

$$H_i = G_i + L \tag{6}$$

where v_i is the speed, t_0 is the desired time gap, G_0 is the minimum standstill safe gap, L is the length of the vehicle.

To ensure the necessary gap ahead, CAVs that need to change lanes will adjust their speeds to align with following vehicles in the target lane, while also ensuring a safe distance from the preceding vehicle in the target lane [30]. At time t, the positions of vehicle i relative to vehicles $j-1$ and j must satisfy the following conditions:

$$x_{j-1}(t) - x_i(t) \ge H_i = \max\{v_i t_0, G_0\} + L \tag{7}$$

$$x_i(t) - x_j(t) \ge H_j = \max\left\{v_j t_0, G_0\right\} + L. \tag{8}$$

After the system issues lane-changing instructions to the CAVs in the PLC area, the CAVs that meet the conditions begin to change lanes. When the number of required CAVs is insufficient, the CAVs in the target lane will assist these CAVs in completing the maneuver.

b) Cooperative lane-changing scenario: In the target lane, both individual CAVs and CAV platoons are treated as a single cooperative vehicle. When the leading vehicle in the target lane is a CAV or a CAV platoon, and the following vehicle is an HV, as shown in Fig. 3. The leading vehicle $j-1$ serves as the cooperative vehicle with $j-2$ positioned directly in front of it. The longitudinal position of vehicle $j-1$ at time t is $x_{j\text{-}1}(t)$. When the cooperative vehicle is a platoon, $x'_{j\text{-}1}(t)$ is the longitudinal position of the last vehicle in the platoon. When the cooperative vehicle is a single vehicle, $x'_{j\text{-}1}(t) = x_{j\text{-}1}(t)$. The longitudinal positions of vehicles $j-2$ and j at time t are $x_{j\text{-}2}(t)$ and $x_j(t)$, respectively.

In Fig. 3, the cooperative vehicle $j-1$ assesses the distance to the vehicle $j-2$ ahead and accelerates to meet the spatial conditions for lane-changing. At time t, the positions of vehicle

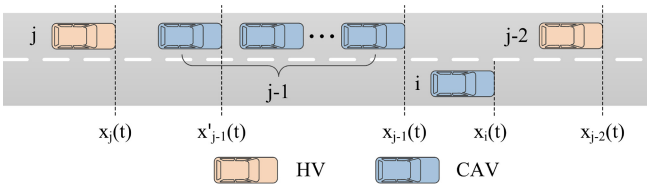


Fig. 4. Leading vehicle in the target lane is an HV and the following vehicle is a CAV or a CAV platoon.

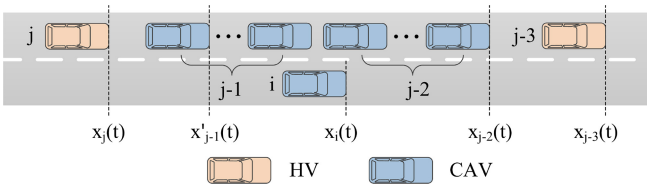


Fig. 5. Both the leading and following vehicles in the target lane are CAVs or CAV platoons.

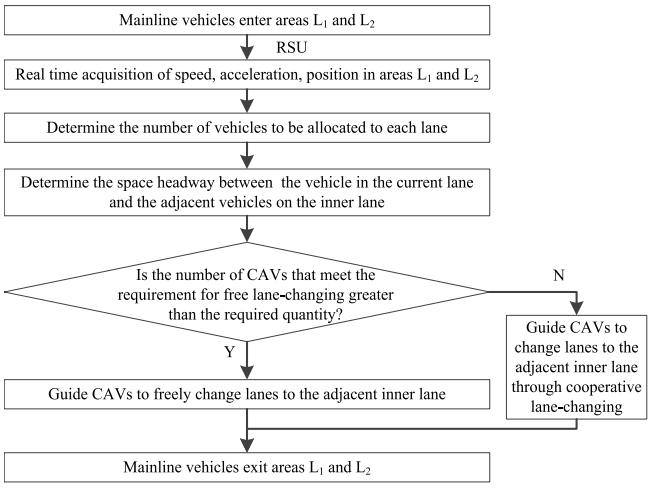


Fig. 6. PLC process for mainline CAVs.

i relative to vehicles $j-1$ and j, and the position of vehicle $j-2$ relative to vehicle $j-1$, must meet the following conditions:

$$x_{j\text{-}2}(t) - x_{j\text{-}1}(t) \geq H_{j\text{-}1} = \max\left\{v_{j\text{-}1}t_0, G_0\right\} + L \tag{9}$$

$$x_i(t) - x_j(t) \geq H_j = \max\left\{v_j t_0, G_0\right\} + L \tag{10}$$

$$x'_{j\text{-}1}(t) - x_i(t) \geq H_i = \max\{v_i t_0, G_0\} + L. \tag{11}$$

When the leading vehicle in the target lane is an HV, and the following vehicle is a CAV or a CAV platoon, as shown in Fig. 4. The following vehicle $j-1$ in the target lane acts as the cooperative vehicle, with vehicle j positioned behind $j-1$. The longitudinal position of vehicle $j-1$ at time t is $x_{j\text{-}1}(t)$. When the cooperative vehicle is a CAV platoon, $x'_{j\text{-}1}(t)$ is the longitudinal position of the last vehicle in the CAV platoon. When the cooperative vehicle is a single vehicle, $x'_{j\text{-}1}(t) = x_{j\text{-}1}(t)$. The longitudinal positions of vehicles $j-2$ and j at time t are $x_{j\text{-}2}(t)$ and $x_j(t)$, respectively.

In Fig. 4, the cooperative vehicle $j-1$ evaluates the spacing to the following vehicle j and decelerates to meet the spatial requirements necessary for lane-changing. At time t, the positions of vehicle i relative to vehicles $j-2$ and $j-1$, and the position of vehicle $j-1$ relative to vehicle j, must meet the following conditions:

$$x_{j\text{-}2}(t) - x_i(t) \geq H_i = \max\{v_i t_0, G_0\} + L \tag{12}$$

$$x_i(t) - x_{j\text{-}1}(t) \geq H_{j\text{-}1} = \max\left\{v_{j\text{-}1}t_0, G_0\right\} + L \tag{13}$$

$$x'_{j\text{-}1}(t) - x_j(t) \geq H_j = \max\left\{v_j t_0, G_0\right\} + L \tag{14}$$

When both the preceding and following vehicles in the target lane are CAVs or CAV platoons, as shown in Fig. 5. The CAV platoon is split, with the vehicle positioned before vehicle i acting as the cooperative vehicle $j-2$, and the vehicle positioned behind vehicle i as the cooperative vehicle $j-1$. The longitudinal positions of vehicle $j-2$ and $j-1$ at time t are $x_{j\text{-}2}(t)$ and $x_{j\text{-}1}(t)$, respectively. When the cooperative vehicle $j-2$ is a CAV platoon, $x'_{j\text{-}2}(t)$ is the longitudinal position of the last vehicle in the CAV platoon. When the cooperative vehicle $j-1$ is a CAV platoon, $x'_{j\text{-}1}(t)$ is the longitudinal position of the last vehicle in the CAV platoon. When the cooperative vehicle $j-2$ is a single vehicle, $x'_{j\text{-}2}(t) = x_{j\text{-}2}(t)$. When the cooperative vehicle $j-1$ is a single vehicle, $x'_{j\text{-}1}(t) = x_{j\text{-}1}(t)$. The longitudinal positions of vehicles $j-3$ and j at time t are $x_{j\text{-}3}(t)$ and $x_j(t)$, respectively.

In Fig. 5, cooperative vehicle $j-2$ assesses the distance to the preceding vehicle $j-3$ and accelerates to satisfy the spatial conditions for lane-changing. Cooperative vehicle $j-1$ evaluates the distance to the following vehicle j and decelerates to meet the spatial requirements for changing lanes. At time t, the positions of vehicle i relative to vehicles $j-2$ and $j-1$, the position of vehicle $j-3$ relative to vehicle $j-2$, and the position of vehicle $j-1$ relative to vehicle j, must meet the following conditions:

$$x_{j\text{-}3}(t) - x_{j\text{-}2}(t) \geq H_{j\text{-}2} = \max\left\{v_{j\text{-}2}t_0, G_0\right\} + L \tag{15}$$

$$x'_{j\text{-}2}(t) - x_i(t) \geq H_i = \max\{v_i t_0, G_0\} + L \tag{16}$$

$$x_i(t) - x_{j\text{-}1}(t) \geq H_{j\text{-}1} = \max\left\{v_{j\text{-}1}t_0, G_0\right\} + L \tag{17}$$

$$x'_{j\text{-}1}(t) - x_j(t) \geq H_j = \max\left\{v_j t_0, G_0\right\} + L. \tag{18}$$

In summary, the control process for the mainline CAVs in the PLC area is shown in Fig. 6.

B. Hierarchical Reward Optimization-Based CAVs Merging Control Method

1) Markov Modeling of CAVs Merging Decisions: Due to the dynamic and partially observable nature of traffic flow and vehicle behavior, CAV decision-making systems must respond with high precision and timeliness. In this article, we formulate the vehicle merging behavior as a partially observable Markov decision process (POMDP) [6]. Each CAV is modeled as an independent agent that makes collaborative decisions [31]. The system is represented as a multiagent deep RL process, defined as a tuple $(S, A_1, \ldots, A_N, R_1, \ldots, R_N, P, \gamma)$. Where N is the number of agents, S is the state set, A_i is the action set for agent i, R_i is the reward, P is the state transition function, and γ is the discount factor. At time t, each agent gets its own state $o_i^t \in S$ and executes joint actions $A^t = (a_1^t, \ldots, a_N^t)$ according to the policy $\pi_i(a_i|o_i)$. The rewards for each agent are as follows:

$$R_i^t = E\left[R_i^{t+1} \middle| S^t = o_i, A_i^t = a_i, H\right] \tag{19}$$

where $H = \prod_{i \in N} \pi_i(a_i|o_i)$ is the joint policy of all agents.

The action-value function of each agent depends on the joint actions, with the formula as follows:

$$Q_i(o, a) = E_i\Big[R_i^{t+1} + \gamma Q_i(S_{t+1}, A_{t+1})\Big]. \tag{20}$$

2) MADDPG Algorithm: The multiagent deep deterministic policy gradient (MADDPG) is a DRL algorithm for multiagent environments. Each agent consists of a critic network, which requires global information, and an actor network, which only needs local observations [32]. This architecture allows MADDPG to search for the optimal joint policy H through centralized training and decentralized execution. During centralized training, the critic network uses both its own and other agents' observations and actions to train the Q-value function. The calculation formula for the action-value function is as follows:

$$Q = Q\big(s^t, a_1, a_2, \ldots, a_N\big) \tag{21}$$

where s^t is the current state of the environment, and a_i is the actions of each agent.

The policy parameters for each agent are denoted by $\theta = \{\theta_1, \theta_2, , \ldots, \theta_N\}$. The set of deterministic policies for all agents is represented as $\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$. This article employs the policy gradient method to update the actor network. For the ith agent, the policy gradient formula for its deterministic policy is as follows:

$$\nabla \mathcal{J}(\theta_i) = E_{s,a\sim D}\Big[\nabla_{\theta_i}\pi_i(a_i|o_i)\nabla_{a_i}Q_i^{\pi}\ (s, a_1, \ldots, a_N)|_{a_i=\pi_i(o_i)}\Big] \tag{22}$$

where $\mathcal{J}(\theta_i)$ is the cumulative expected return for the ith agent, and $s = (o_1, o_2, \ldots, o_N)$ is the state vector of the environment. $Q_i^{\pi}(s, a_1, \ldots, a_N)$ is the action-value function calculated during centralized training, which takes as input the actions and state information of all agents and outputs the action value for each agent. The experience replay pool $D = \{s, s', a_1, \ldots, a_N, r_1, \ldots, r_N\}$ is used to store previous training data. $s' = (o_1', o_2', \ldots, o_N')$ is the new state vector of the environment obtained after all agents have executed their actions. $a_i = \pi_i(o_i)$ is the action policy for agent i.

After randomly sampling data from the experience replay pool, the stored data s' is input into the actor target network to obtain the actions a' at the next step. Then, s' and a' are input into the critic target network to obtain the target action value for the next step, and the current target action value is calculated as follows:

$$y = r_i + \gamma Q_i^{\pi'}\ \big(s', a_1', \ldots, a_N'\big)\big|_{a_j'=\pi_j'\left(o_j'\right)} \tag{23}$$

where $\pi' = \{\pi_1', \pi_2', \ldots, \pi_N'\}$ is the set of policies with delayed updated parameters θ_i'. r_i is the immediate reward for the ith agent.

Then, the critic network uses the mean squared error to update its loss function, with the formula:

$$L(\theta_i) = E_{s,a,r,s'}\Big[\big(Q_i^{\pi}(s, a_1, \ldots, a_N) - y\big)^2\Big]. \tag{24}$$

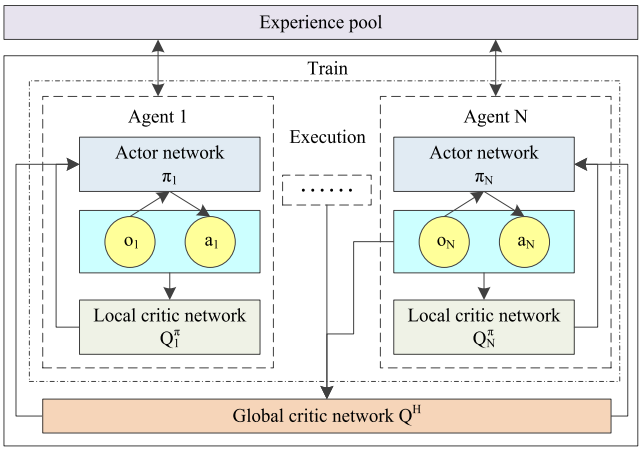


Fig. 7. HRO-MADDPG structure.

3) HRO-MADDPG Algorithm: The hierarchical reward optimization MADDPG (HRO-MADDPG) algorithm is an extension of the traditional MADDPG algorithm. MADDPG employs a global centralized critic network for centralized training, whereas HRO-MADDPG improves the network architecture. In addition to the global critic network, it constructs a local critic network for each agent, as shown in Fig. 7. The global critic network aims to maximize the global reward, while the local critic network focuses on maximizing local rewards. During centralized training of multiagent systems, each agent can simultaneously maximize both global and local returns. On one hand, it prevents the generation of suboptimal and unstable solutions, and on the other hand, it also prevents the strategy of the agent with the highest return value, caused by the centralized critic network of the MADDPG algorithm, from playing a dominant role in the group. This further ensures that each agent can effectively make optimal decisions.

The architecture of HRO-MADDPG combines the MADDPG algorithm with the DDPG algorithm, maintaining the characteristics of centralized training and decentralized execution. During the training phase, the algorithm requires additional state and action information from other agents. In the execution phase, agents only need to generate policy actions based on their own states. For agent i, the policy gradient is calculated as follows:

$$\begin{aligned}\nabla \mathcal{J}(\theta_i) = {} & E_{s,a\sim D}\Big[\nabla_{\theta_i}\pi_i(a_i|o_i)\nabla_{a_i}Q_{\psi}^{\pi}(s, a_1, \ldots, a_N)\Big] \\ & + E_{o_i,a_i\sim D}\Big[\nabla_{\theta_i}\pi_i(a_i|o_i)\nabla_{a_i}Q_{\varphi_i}^{\pi}(o_i, a_i)\Big]\end{aligned} \tag{25}$$

where Q_{ψ}^{π} is the global critic network parameterized by ψ, used to evaluate the joint policy of all agents. $Q_{\varphi_i}^{\pi}$ is the local critic network parameterized by φ_i, used to evaluate the local policy of agent i.

Since HRO-MADDPG has two types of critic networks, there are two corresponding loss functions. The loss function of the global critic network Q_{ψ} is as follows:

$$L(\psi) = E_{s,a,r,s'}\left[\left(Q_{\psi}^{\pi}(s, a_1, a_2, \ldots, a_N) - y_G\right)^2\right] \tag{26}$$

$$y_G = r_G + \gamma Q_{\psi'}^{\pi'}\ \big(s', a_1', a_2', \ldots, a_N'\big)\big|_{a_i'=\pi_i'(o_i')} \tag{27}$$

Algorithm 1 Training Process of Global Critic Network

Initialize main global critic network Q_ψ and target global critic network $Q_{\psi'}$
for episode = 1 to D **do**
  for t= 1 to T **do**
    Get initial state $s = (o_1, o_2, \ldots, o_N)$
    Choose action $a_i = \pi_i(o_i)$ for each agent i
    Execute action $a = (a_1, a_2, \ldots, a_N)$
    Receive global reward r_G and local reward r_L
    Store $\left(s, a, r_L, r_G, s'\right)$ into replay buffer D
/* Train global critic */
    Randomly sample a minibatch of S samples $\left(s^j, a^j, r_G^j, s'^j\right)$ from D
    Set $y_G^j = r_G^j + \gamma Q_{\psi'}^{\pi'}\left(s'^j, a'_1, a'_2, \ldots, a'_N\right)\Big|_{a'_i=\pi'_i\left(o_i'^j\right)}$
    Update global critic Q_ψ by minimizing $\frac{1}{S}\sum_j\left(Q_\psi^\pi\left(s^j, a_1^j, a_2^j, \ldots, a_N^j\right) - y_G^j\right)^2$
    Update target global critic $Q_{\psi'}$: $\psi'_i \leftarrow \tau\psi_i + (1-\tau)\psi'_i$
  end for
end for

where r_G is the global reward, and $Q_{\psi'}$ is the target global critic network parameterized by ψ'.

The loss function of the local critic network Q_{φ_i} is as follows:

$$L(\varphi_i) = E_{o,a,r,o'}\left[\left(Q_i^L(o_i, a_i) - y_L^i\right)^2\right] \tag{28}$$

$$y_L = r_L^i + \gamma Q_{\varphi_i'}^{\pi'}\left(o_i', a_i'\right)\Big|_{a_i'=\pi_i'(o_i')} \tag{29}$$

where r_L^i is the local reward of agent i, and $Q_{\varphi_i'}$ is the target local critic network parameterized by φ_i'.

The algorithm for training the global critic network is shown in Algorithm 1.

During the training process, in addition to updating the global critic network, it is also necessary to train the local critic and actor networks for each agent to ensure that each agent can adjust its policy in real time based on global feedback. The algorithm for training the local Critic network is shown in Algorithm 2.

4) CAVs Merging Control Strategy: This article designs multilevel reward functions based on the specific merging tasks of CAVs and then adopts the HRO-MADDPG algorithm to train multiple agents.

a) Action space: The action space of agent i is defined as a set of high-level control decisions, namely, $a_i = (a_l, a_r, a_c, a_a, a_d)$, where a_l and a_r are the lateral controls for left and right turns, respectively. a_c, a_a and a_d are the longitudinal controls for cruising, acceleration, and deceleration, respectively. Through the chosen high-level decisions, the lower-level controllers generate the corresponding steering and throttle control signals. The overall action space of the system is composed of the joint actions of all CAVs, namely, $a = a_1 \times a_1 \times \cdots \times a_N$ [6].

b) State space: CAVs can obtain their state information in real-time through onboard sensors, V2X communication, and RSU technology. In real-world scenarios, due to the lack of real-time communication and advanced sensors in HVs, their state is more dependent on prediction models. However, considering that the focus of this article is on vehicle merging control, the acquisition of HVs' states has been simplified. In the simulation, the real-time state of HVs is directly obtained through SUMO's TraCI package.

Algorithm 2 Training Process of Local Critic and Actor Networks

Initialize critic network Q_{φ_i} and target critic network $Q_{\varphi_i'}$ of each agent
Initialize actor network π_i and target actor network π_i' of each agent
for episode = 1 to D **do**
  for t= 1 to T **do**
/* Train local critics and update actor network */
    for agent i = 1 to N **do**
      Randomly sample a minibatch of S samples $\left(s^j, a^j, r_L^j, s'^j\right)$ from D
      Set $y_L^j = r_L^j + \gamma Q_{\varphi_i'}^{\pi'}\left(o'^j, \pi_i'\left(o_i'^j\right)\right)$
      Update local critic $Q_{\varphi_i}^\pi$ by minimizing $(1/S)\sum_j\left(Q_{\varphi_i}^\pi\left(o^j, a_i^j\right) - y_L^j\right)^2$
      Update actor network: $\theta_i \leftarrow \theta_i + \frac{1}{S}\sum_j \nabla_{\theta_i}\pi_i\left(a_i^j\middle|o_i^j\right)\nabla_{a_i}Q_\psi^\pi\left(s^j, a_1^j, \ldots, a_N^j\right) + \nabla_{\theta_i}\pi_i\left(a_i^j\middle|o_i^j\right)\nabla_{a_i}Q_{\varphi_i}^\pi\left(o_i^j, a_i^j\right)$
    end for
    Update target network parameters for each agent: $\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i'\varphi_i' \leftarrow \tau\varphi_i + (1-\tau)\varphi_i'$
  end for
end for

We define the state space of agent i as $o_i = (v_{\text{ev}}, p_{\text{ev}}, v_{\text{nv}}, p_{\text{nv}})$, where v_{ev} and p_{ev} are the speed and position of the ego vehicle. v_{nv} and p_{nv} are the speed and position of nearby vehicles (within a distance of no more than 150 m) [33]. The state of the environment is $s = (o_1, o_2, \ldots, o_N)$.

c) Reward function: The reward functions are designed based on the individual goals of each agent and the task objectives of the multiagent system. For the cooperative merging strategy of multiple CAVs, the combination of global reward function and local reward function will promote the decision network to update to optimal parameters faster.

The global reward R_G considers the overall objectives of the merging control area, aiming to maximize the traffic efficiency and safety of all CAVs and HVs.

The efficiency reward r_G^e is defined as [25]

$$r_G^e = \frac{v_t - v_{\min}}{v_{\max} - v_{\min}} \tag{30}$$

where v_t is the vehicle speed. Considering the maximum speed limit in the research scenario and existing studies, the minimum speed $v_{\min}$ and maximum speed $v_{\max}$ are set to 10 m/s and 33.33 m/s, respectively.

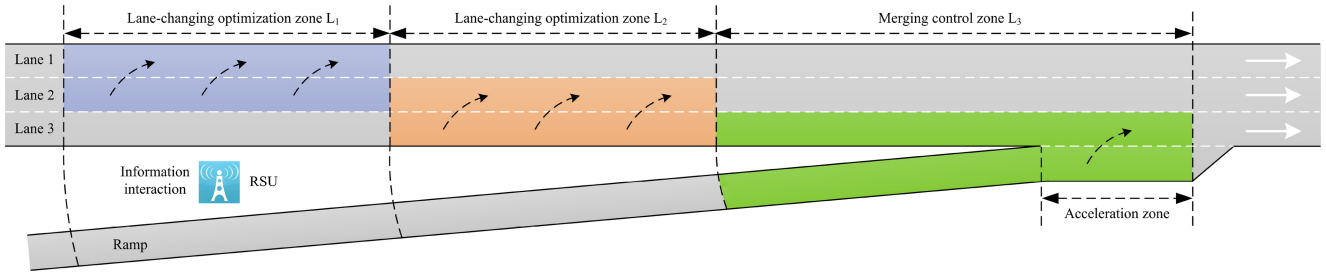


Fig. 8. Three-lane highway traffic scenario.

In terms of safety, the headway distance reward r_G^h is defined as [6]

$$r_G^h = \log \frac{d_{\text{headway}}}{t_h v_t} \tag{31}$$

where d_{headway} is the headway distance, and t_h is the predefined headway threshold. In this study, t_h is set to 1.2 s. Therefore, when the headway distance is less than t_h, the ego vehicle is penalized, and it receives a reward only when the headway distance exceeds t_h.

Consequently, the global reward R_G is defined as

$$R_G = \omega_e r_G^e + \omega_h r_G^h \tag{32}$$

where ω_e and ω_h are the weights of r_G^e and r_G^h, respectively. $\omega_e = 0.4$ and $\omega_h = 0.6$.

The local reward R_L primarily focuses on objectives of each CAV on the outermost mainline lane and ramp, encouraging CAVs to exit the merging area quickly while preventing collisions and not exceeding boundaries.

The speed reward r_L^s is designed to encourage CAVs to drive at higher speeds, while penalizing unreasonable speeds and excessive speeding. Within the range of 0 m/s to the maximum speed limit $v_{\text{Lane x}}^{\text{limit}}$ of the lane, the reward r_L^s is set to monotonically increase from 0 to 1. Once it exceeds this range, the reward is set to –1 to penalize speeding or unreasonable speeds. The r_L^s is defined as [34]

$$r_L^s = \begin{cases} \frac{v_t^{\text{CAV}}}{v_{\text{Lane x}}^{\text{limit}}}, & 0 \le v_t^{\text{CAV}} \le v_{\text{Lane x}}^{\text{limit}} \\ -1, & \text{otherwise} \end{cases} \tag{33}$$

where v_t^{CAV} is the speed of the CAV at time t.

The collision penalty r_L^c is applied to ensure safe driving. The r_L^c is defined as

$$r_L^c = \begin{cases} -1, & \text{collision} \\ 0, & \text{otherwise} \end{cases}. \tag{34}$$

The merging cost r_L^m aims to penalize the waiting time of CAVs on the acceleration lane to avoid deadlock. The r_L^m is defined as [25]

$$r_L^m = -\exp\left(\frac{-(x-L)^2}{10L}\right) \tag{35}$$

where x is the distance traveled by the CAV on the acceleration lane, and L is the length of the acceleration lane. When the CAV approaches the end of the acceleration lane, the penalty increases.

To reduce frequent lane changes by CAVs, the lane-changing penalty function r_L^l is defined as [25]

$$r_L^l = \begin{cases} -1, & \text{change lane} \\ 0, & \text{otherwise} \end{cases}. \tag{36}$$

Thus, the local reward R_L is defined as

$$R_L = \omega_s r_L^s + \omega_c r_L^c + \omega_m r_L^m + \omega_l r_L^l \tag{37}$$

where ω_s, ω_c, ω_m and ω_l are the weights of r_L^s, r_L^c, r_L^m and r_L^l, respectively. $\omega_s = 0.3$, $\omega_c = 0.4$, $\omega_m = 0.2$, and $\omega_l = 0.1$.

III. Modeling and Experiments

A. Traffic Scenario

This article takes a three-lane highway as an example. The speed limit of the two inner lanes of the mainline is 33.33 m/s, the speed limit of the outermost lane is 27.78 m/s, and the speed limit of the ramp is 22.22 m/s. In this study, we number the lanes along the highway from left to right as lane 1 to lane 3. The lengths of the PLC areas L1 and L2 are taken as 300 m, respectively. The length of the merging control area L3 is taken as 550 m, with the acceleration lane being 250 m. Area L1 aims to guide some CAVs from lane 2 to lane 1. Area L2 aims to guide some CAVs from lane 3 to lane 2. Area L3 aims to control the CAVs in the outermost lane of the mainline and ramps to coordinate merging. As shown in Fig. 8.

In current research, SUMO has become a widely used traffic simulation platform [27]. It offers numerous vehicle following and lane-changing models that meet various research needs. Moreover, as an open-source software, SUMO provides greater flexibility for further development [35]. Thus, we use SUMO to simulate and validate the proposed strategy. As shown in Fig. 9.

B. Vehicle Dynamics

In this study, the HRO-MADDPG algorithm is adopted for the driving policy of the CAVs in the merging control area. For CAVs in the PLC areas, this article employs the CACC platoon control strategy, where CAVs following HVs are controlled by the adaptive cruise control (ACC) model, while those following CAVs are controlled by the cooperative ACC (CACC) model [36], [37]. The intelligent driver model (IDM) is used to simulate the longitudinal dynamics of all HVs [38]. The default LC2013 model in SUMO is employed to simulate the lateral dynamics of all vehicles [39], [40]. The LC2013 model in SUMO uses a four-tiered motivation

Fig. 9. SUMO simulation platform.

TABLE I
SUMMARY OF MAIN PARAMETERS FOR HVS AND CAVS IN SIMULATION

Parameters	HVs	CAVs
acceleration (m/s^2)	2.9	2.9
deceleration (m/s^2)	7.5	7.5
minGap (m)	2.5	1.0
tau (s)	1.5	0.5
sigma	0.5	0.0
impatience	0.5	0.0

Note: minGap is the minimum following distance; tau is the desired time headway; sigma is the driver's degree of imperfect driving, range [0,1]; impatience is the driver's patience, range [0,1].

hierarchy to guide vehicle behavior during each simulation step. It makes lane-changing decisions based on the vehicle's planned route, current and historical traffic conditions, and adjusts the vehicle's speed to facilitate successful maneuvers. Motivations are categorized into strategic, cooperative, tactical, and obligatory changes, reflecting the driver's decisions based on long-term routing, coordination with other vehicles, immediate traffic conditions, and compliance with road rules or obstacles. These factors work together to enable vehicles to make rational lane-changing decisions in complex traffic environments. The basic parameter settings are shown in Table I [38].

C. Experiments Details

In the training environment, two typical vehicle streams stemming from the mainline and the ramp are 1200 and 800 passenger car unit per hour per lane (pcu/h/lane), respectively [27]. Moreover, to evaluate the effectiveness of the proposed strategy under different traffic densities, tests are conducted based on the highway capacity manual (HCM) at 800, 1200, and 1600 pcu/h/lane on the mainline. In the SUMO simulation, the vehicle input for the mainline is set to three times the traffic volume of each lane. To ensure a certain level of randomness in the distribution of vehicles across the main lanes, we have used the departLane="random" configuration, which better reflects real-world traffic conditions. During these tests, the ramp merging rates remain constant, with a CAV penetration rate of 80%.

Our experiments are conducted in a highway merging scenario using FLOW, which provides a convenient interface between SUMO and DRL algorithms. All algorithms are implemented on the computer (RAM: 64 GB, processor: Intel Core i9-14900K, GPU: NVIDIA GeForce RTX 4090D). Regarding the main training parameters for the experiments, the discount factor γ is 0.99, the batch size is 64, the actor network learning rate is 0.001, the critic network learning rate is 0.002, the target network soft update rate τ is 0.01, MaxEpisode is 10000. The local critic network for each agent consists of a three-layer fully connected neural network, which includes an input layer, a hidden layer, and an output layer. The hidden layer is equipped with 64 neurons featuring the ReLU activation function. The global critic network is a four-layer fully connected neural network with two hidden layers. The actor network is a four-layer fully connected neural network, which is the same as the global critic network. To minimize the impact of simulation randomness, each simulation is repeated ten times with different random seeds. The results are averaged to eliminate randomness and enhance the reliability of the outcomes.

D. Comparison Baselines

In this section, we introduce some existing MADRL algorithms in the field of highway vehicle merging. In addition to the MADDPG baseline model mentioned earlier, we also introduce three other baseline models here. MAPPO is the multiagent extension of PPO [41]. By limiting the magnitude of policy updates, this algorithm effectively avoids instability caused by overly large policy updates. MAA2C is the multiagent version of A2C [42]. This algorithm enables efficient learning of agents in multiagent environments through expected updates while also having the ability to adapt to complex environmental changes [6]. MASAC is a multiagent extension of SAC that adds an entropy term to encourage exploration. It is suited for high-dimensional continuous action spaces and can handle both cooperative and partially competitive environments [43].

To evaluate the performance of the HRO-MADDPG algorithm, we analyze the average episode reward of each model. These baseline algorithms typically focus on optimizing global rewards without separately considering local rewards. To ensure fairness in comparison, although the HRO-MADDPG model includes both global and local rewards, we only compare its global reward with the rewards of the other models. This ensures that the effectiveness of different algorithms in optimizing the overall performance of the merging area can be evaluated under the same standards. Specifically, we conduct a comparative analysis of the rewards for all models in 10 000 episodes. The reward results are averaged every 100 episodes

TABLE II
EXPERIMENTAL GROUP OF GLOBAL REWARD WEIGHT

Group	Weight	
	r_G^e	r_G^h
1	0.5	0.5
2	0.4	0.6
3	0.3	0.7
4	0.2	0.8

TABLE III
EXPERIMENTAL GROUP OF LOCAL REWARD WEIGHT

Group	Weight			
	r_L^s	r_L^c	r_L^m	r_L^l
1	0.4	0.4	0.1	0.1
2	0.3	0.5	0.1	0.1
3	0.2	0.6	0.1	0.1
4	0.3	0.4	0.2	0.1
5	0.2	0.5	0.2	0.1

to smooth the reward curve and reduce random fluctuations during the training process.

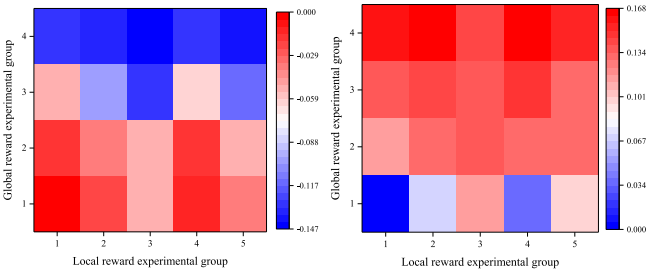
IV. SIMULATION RESULTS

A. Sensitivity Analysis

This section aims to validate the rationality of the reward function weights. The weight settings are based on experience and consider the task requirements. According to existing literature, safety is considered the top priority in merging control, thus it is given a higher weight. When designing the global and local reward weights, we ensure that the weight for safety-related rewards is no less than that for efficiency-related rewards. However, placing too much emphasis on safety may result in overly conservative vehicle behavior, which could reduce efficiency. Therefore, we avoid setting a large disparity between the two weights. The priorities for avoiding deadlock and lane-change penalties are lower, with the lane-change penalty weight set to the minimum to prevent vehicles from deliberately refusing to change lanes.

We designed four experimental groups for global reward weights and five experimental groups for local reward weights, resulting in a total of 20 experimental combinations, with the combination of global reward experimental group 1 and local reward experimental group 1 serving as the baseline. As shown in Tables II and III. We studied the changes in average speed and average Time to Collision (TTC) for different combinations compared to the baseline, as shown in Fig. 10.

Fig. 10(a) shows that when selecting global reward group 1 and group 2, the traffic efficiency slightly decreases compared to the baseline, which places the highest emphasis on global efficiency, while the traffic efficiency in global reward groups 3 and 4 significantly drops. Fig. 10(b) indicates that when selecting global reward group 1, safety is much lower than in other groups due to the lower weight of the safety reward. As the safety reward weight in the global reward increases, the safety performance of global reward groups 2, 3, and 4 significantly improves. Building on the excellent traffic efficiency of global reward groups 1 and 2 shown in Fig. 10(a), and combining with Fig. 10(b), we find that global reward group 2 can achieve a good balance between both efficiency and safety. For global reward group 2, by observing the impact of local rewards shown in Fig. 10(a) and (b), we find that the combination used in this article (global reward group 2 and local reward group 4) exhibits excellent traffic efficiency while maintaining safety similar to other groups, demonstrating high overall performance.



(a) (b)

Fig. 10. Sensitivity analysis of different reward function weight combinations. (a) Average speed. (b) Average TTC.

B. Learning Performance

To study the training process of the HRO-MADDPG algorithm, we compare the average reward of the proposed model with the baseline models. As shown in Fig. 11.

Fig. 11 shows that the rewards of all algorithms increase as training progresses. HRO-MADDPG converges around 7000 episodes, achieving the highest reward and the fastest convergence. Compared to other algorithms, HRO-MADDPG can reach the optimal state more quickly in the same training scenario. Its rapid convergence not only improves training efficiency but also reduces the consumption of computational resources, highlighting its advantages in terms of time cost.

MAPPO and MASAC converge faster than HRO-MADDPG before 3000 episodes, as HRO-MADDPG requires each agent to explore both how to coordinate actions in the global environment and make optimal decisions in the local environment, making early training more complex. In contrast, MAPPO focuses on optimizing global information, making the early learning task relatively simpler. On the other hand, MASAC accelerates exploration and learning through entropy maximization, but this also leads to higher reward fluctuations compared to MAPPO. As training progresses, the dual optimization mechanism of HRO-MADDPG begins to show its effectiveness. The local critic network gradually helps agents learn local policies more effectively, combining them with the optimization of global rewards, ultimately leading to more stable policy execution.

In comparison, the traditional MADDPG performs worse than HRO-MADDPG in both reward and convergence speed. MADDPG relies on a centralized critic for global evaluation and lacks effective local feedback, limiting its performance in multiagent tasks. The MAA2C algorithm shows the greatest instability, with its strategy update mechanism more prone to interference between agents, leading to unstable policies.

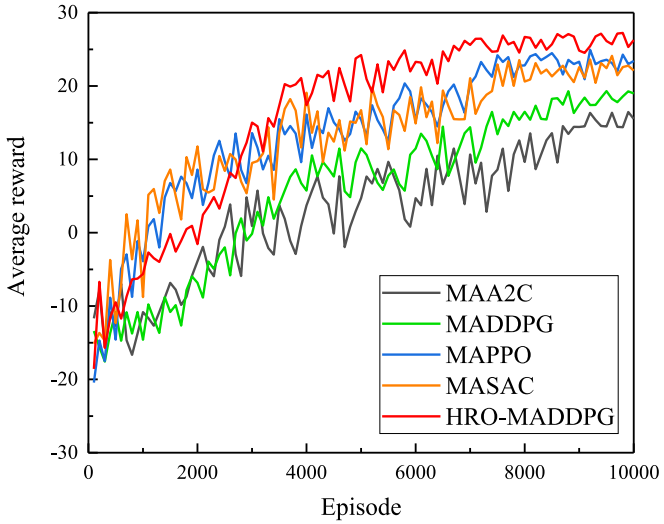


Fig. 11. Average reward every 100 episodes.

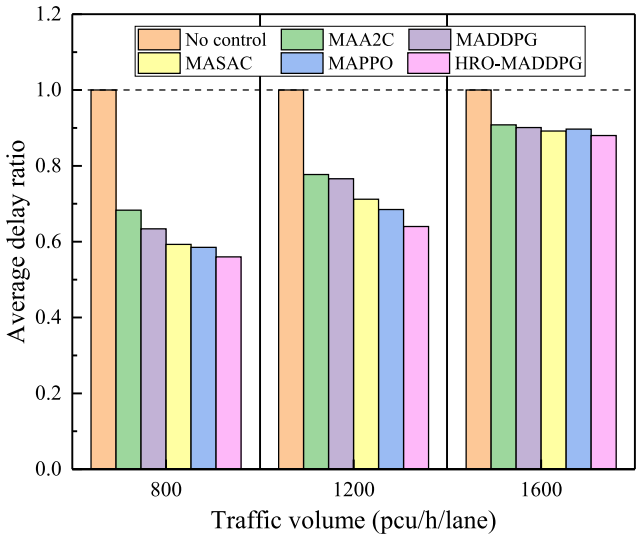


Fig. 12. Average delay ratio.

C. Traffic Efficiency

This section studies the average delay of different algorithms under various traffic volume levels, with the PLC strategy employed throughout. The no control scenario is used as the baseline, as shown in Fig. 12.

Fig. 12 illustrates that, in comparison to medium and high-density scenarios, all algorithms demonstrate greater effectiveness in reducing average delay under low-density conditions. This is largely attributed to the greater spacing between vehicles, which facilitates the implementation of the PLC strategy. As density gradually increases, the average delay ratio under medium-density is higher than under low-density. Under high-density conditions, severe traffic congestion restricts the available space for lane changes, thereby reducing the effectiveness of the PLC strategy. The improvement in this scenario is further diminished.

HRO-MADDPG consistently achieves the lowest average delay across various traffic volumes compared to other state-of-the-art algorithms. Notably, under medium-density conditions, it shows a significantly higher improvement in delay reduction than the other algorithms. However, under high-density conditions, the improvement is more similar to that of the other algorithms.

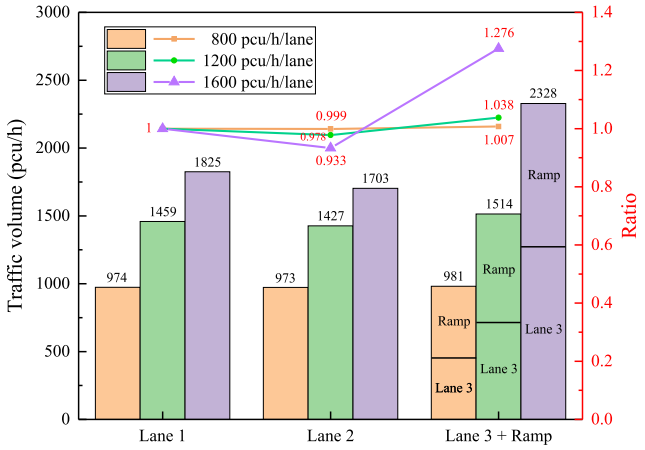


Fig. 13. Unevenness of traffic flow on the mainline.

D. Unevenness of Traffic Flow

This article studies the impact of the PLC strategy on the unevenness of traffic flow on the mainline under different traffic volume levels. We analyze the cross sectional traffic volume between area L2 and area L3. The traffic flow of Lane 1 is used as the baseline. As shown in Fig. 13.

Fig. 13 shows that the PLC strategy effectively alleviates the problem of increased density in the outermost lane of the mainline due to merging under varying traffic volume levels. Specifically, when the mainline traffic volume is 800 pcu/h/lane and 1200 pcu/h/lane, respectively, the sum of the cross sectional traffic volume of lane 3 and the ramp is close to the traffic volumes of the innermost lane 1 and the middle lane 2. This indicates that the PLC strategy can ensure a uniform distribution of traffic flow on the mainline downstream of the merging area, while also reducing the computational burden of using deep RL for trajectory control.

When the mainline traffic volume reaches 1600 pcu/h/lane, although the PLC strategy can still promote a uniform distribution of mainline traffic flow, its improvement is not as significant as in the low and medium-density scenarios. At this point, the combined cross sectional traffic volume of lane 3 and the ramp is approximately 1.276 times that of the innermost lane 1. This indicates that an increase in traffic volume will significantly reduce the gap between mainline vehicles, preventing some CAVs with lane-change intentions from successfully shifting to the inner lanes.

E. Congested Patterns

This section compares the spatiotemporal trajectories of traffic flows in the outermost lane of the mainline and the ramp within the merging control area at medium traffic density, as shown in Figs. 14 and 15.

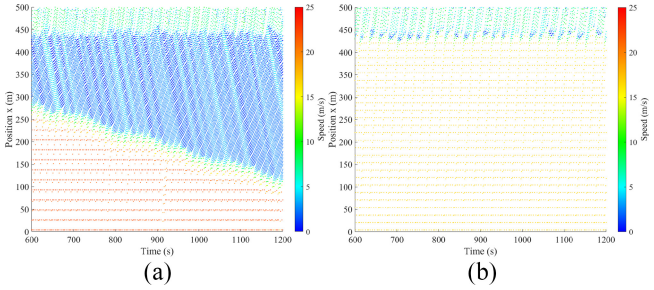


Fig. 14. No control—Spatiotemporal trajectory. (a) Lane 3. (b) Ramp.

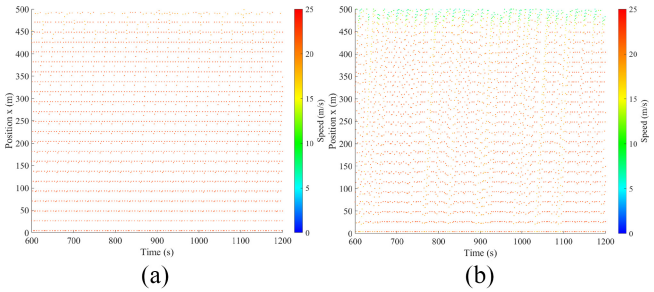


Fig. 15. Proposed strategy—Spatiotemporal trajectory. (a) Lane 3. (b) Ramp.

Fig. 14 shows the significant congestion in the outermost lane of the mainline and the ramp within the highway merging area under conditions without control strategies. Vehicles in the outermost lane experience extensive stopping and starting due to frequent merging actions and spatial limitations, severely hindering the smooth flow of traffic. Additionally, the speed on the entrance ramp drops significantly below 10 m/s for about 70 m, indicating substantial difficulties for ramp vehicles attempting to merge onto the mainline, which further reduces traffic flow efficiency.

Fig. 15 demonstrates that proposed strategy significantly improves the spatiotemporal trajectories of traffic flow. The PLC strategy optimizes vehicle lane positions before merging, effectively reducing congestion in the merging area. Vehicles in the outermost lane of the mainline can pass through the merging area more orderly. The speed of traffic on the ramp has increased, and although there is a slight reduction in speed over a 20 m section of the entrance ramp compared to upstream speeds, the overall speed remains between 10 m/s and 20 m/s. This indicates that the HRO-MADDPG algorithm promotes vehicles to respond more flexibly.

F. Generalization Performance

This section studies the average speed growth rate of HRO-MADDPG compared to MAPPO and the no-control algorithm under different CAV penetration rates and ramp inflow rates, with the PLC strategy employed throughout. The color bar on the right side of both heatmaps represents the mapping from values to colors. The redder the color, the higher the growth rate, while the bluer the color, the lower the growth rate. As shown in Fig. 16.

Fig. 16 demonstrates that the HRO-MADDPG exhibits strong generalization performance under varying traffic conditions. Compared to the MAPPO and the no-control algorithm,

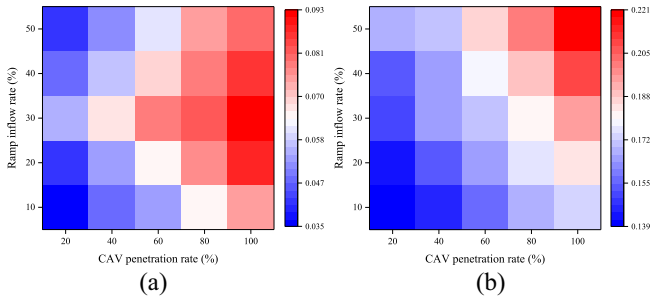


Fig. 16. Impact of different CAV penetration rates and ramp inflow rates. (a) Relative to MAPPO. (b) Relative to none.

it improves the average speed of the merging area to varying degrees. The color bar on the right side of Fig. 16(a) shows that HRO-MADDPG increases the average speed by 3.5%–9.3% over MAPPO, with average speed rising as the CAV penetration rate increases. However, as the ramp inflow rate rises, the average speed first increases, then decreases, indicating that excessive ramp inflow beyond a certain threshold can cause congestion and vehicle interference at the merge point. Even highly automated vehicles must slow down in dense traffic to ensure safe merging. While HRO-MADDPG optimizes traffic flow, it requires further adjustments or integration with other traffic management strategies under extreme conditions to maintain efficiency and safety. Fig. 16(b) shows that compared to the no-control algorithm, HRO-MADDPG increases average speed by 13.9%–22.1%. As both CAV penetration and ramp inflow increase, the average speed rises.

In addition, this article also studies the impact of different car-following and lane-changing models of HVs on the generalization of the proposed algorithm. We additionally adopt the car-following model Krauss and the lane-changing model SL2015, using the combination of the car-following model IDM and the lane-changing model LC2013 as the baseline to analyze the change rates in average speed and average TTC for different model combinations relative to the baseline. Among them, the Krauss model adjusts acceleration based on the safety distance and relative speed between vehicles. The SL2015 model can be used for sublane simulation, with an additional behavioral layer responsible for maintaining safe lateral gaps, as shown in Fig. 17.

Fig. 17 shows that under different car-following and lane-changing models, the proposed algorithm still ensures the merging efficiency and safety. The change rates in average speed and average TTC for different model combinations relative to the baseline used in this article are less than 2%. This shows that even in real-world environments with differences from the simulation models, the proposed algorithm makes appropriate action choices in response to different vehicle behaviors.

G. Traffic Safety Evaluation

To study the impact of the HRO-MADDPG algorithm on the safety of the merging area, this article analyzes the TTC, a crucial metric in traffic safety research [44]. A negative TTC value indicates that the following vehicle is slower than

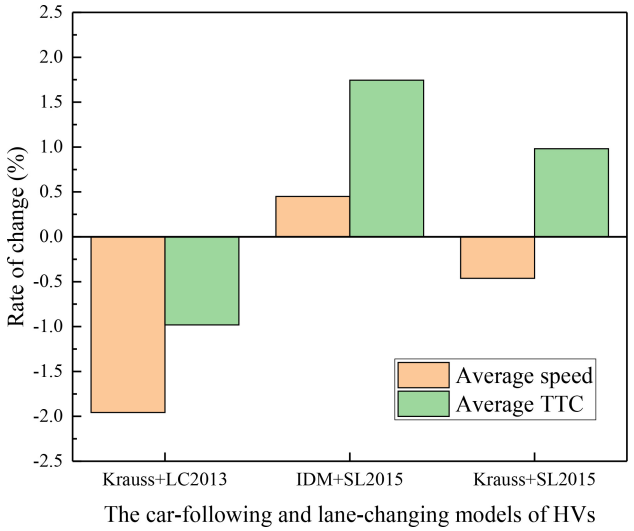


Fig. 17. Impact of different car-following and lane-changing models of HVs.

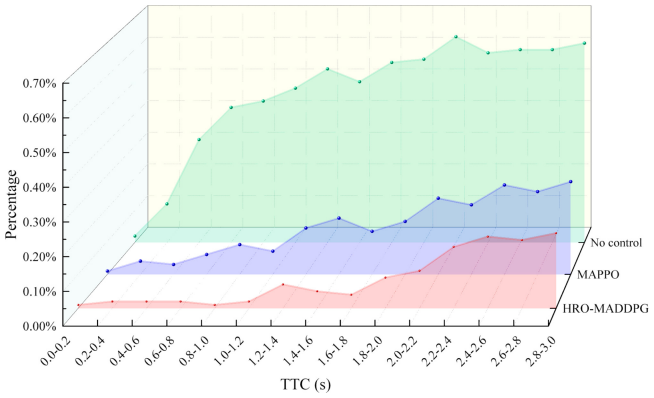


Fig. 18. TTC distribution.

the one ahead, whereas a higher TTC value suggests a lower probability of collision. Therefore, this article focuses on analyzing TTC values within a 0–3 s range. We select the MAPPO algorithm and a scenario without control algorithms as baseline comparisons. The PLC strategy is employed throughout this section. As shown in Fig. 18.

Fig. 18 shows that within the 0–3 s TTC range, the TTC distribution proportion increases as TTC increases across three scenarios. In the scenario without a control algorithm, the TTC distribution proportion is about 7.02% within the 0–3 s range. The proportion sharply rises after exceeding 0.2 s and stabilizes after surpassing 2 s. This sharp increase is due to a lack of effective intervention and adjustment between vehicles, leading to frequent emergency braking. In the scenario using the MAPPO algorithm, the TTC distribution proportion is approximately 2.14% within the 0–3 s range. As the TTC exceeds 1.2 s, the proportion gradually increases. With the application of the HRO-MADDPG algorithm, the TTC distribution proportion in the 0–3 s range is about 1.27%. This proportion gradually rises after the TTC surpasses 1.6 s. The HRO-MADDPG algorithm further optimizes vehicle interaction strategies, significantly reducing disturbances and conflicts between vehicles. Across different TTC distribution ranges, the frequency distribution of TTC using the HRO-MADDPG algorithm is much lower than that of the no-control scenario and significantly lower compared to the MAPPO algorithm. This indicates that the HRO-MADDPG algorithm significantly reduces the risk of collisions between vehicles in merging areas.

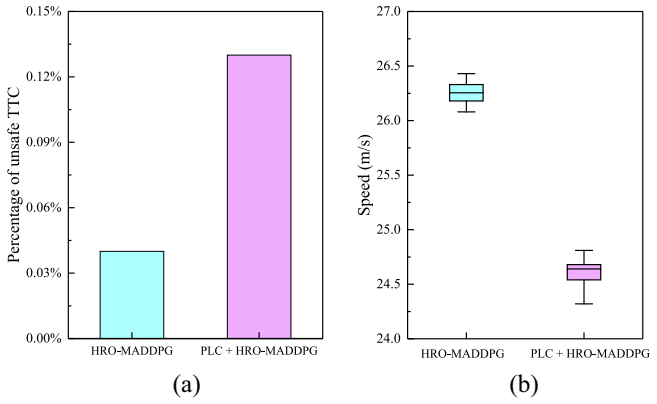


Fig. 19. Impact of the PLC Strategy on PLC areas. (a) Percentage of unsafe TTC. (b) Speed.

This article also analyzes the impact of the PLC strategy on PLC areas, focusing on two indicators: 1) the percentage of unsafe TTC and 2) speed fluctuations. A TTC value less than two seconds is considered unsafe [28], as shown in Fig. 19.

Fig. 19(a) shows that the PLC strategy leads to an increase in the percentage of unsafe TTC in PLC areas. This is primarily due to some vehicles from the outer lanes being guided into the inner lanes. However, despite this, the percentage of unsafe TTC in the merging area (As shown in Fig. 18) is approximately 0.4%, and the percentage of unsafe TTC in PLC areas remains considerably lower than that in the merging area. Fig. 19(b) shows that the PLC strategy results in an overall reduction in speed of about 6% in PLC areas, along with a slight increase in speed fluctuations. Therefore, it can be concluded that the PLC strategy sacrifices a small decrease in the efficiency of PLC areas in exchange for a significant improvement in the efficiency of the merging area.

H. Driving Comfort

This section studies the driving comfort on the highway and its bottleneck segment, considering it as a reflection of traffic flow stability. The International Organization for Standardization 2631-1 introduces comfort index (CI) by calculating and classifying RMS values to clearly evaluate the vibration severity of vehicles [45]. The lower the CI value, the lower the vibration, and the higher the driving comfort. To analyze the CI values more intuitively, we also conducted three supplementary experiments (SE) for comparison with similar studies in the literature [38]. As shown in Table IV. The calculation of CI indicators is expressed by

$$CI = \left(\frac{1}{n}\sum_{i=0}^{n} a_i^2\right)^{0.5} \tag{38}$$

TABLE IV
CI COMPARISON

	This paper	SE 1	SE 2	SE 3	Reference [38]
Mainline (pcu/h/lane)	1200	1200	1200	1500	1500
Ramp (pcu/h/lane)	800	300	300	300	300
CAV penetration rate	80%	80%	10%	10%	10%
Highway CI (m/s^2)	0.43	0.41	0.52	0.56	0.64
Bottleneck CI (m/s^2)	1.38	1.27	1.66	1.89	2.28

where a_i is the ith acceleration obtained from the trajectory data, and n is the number of accelerations in the whole simulation.

Table IV shows that under similar traffic conditions, the proposed strategy can effectively ensure driving comfort in highway merging areas. The driving comfort increases with the penetration rate of CAVs and decreases with the increase in traffic volume.

V. CONCLUSION

This article proposes a hierarchical cooperative merging control strategy for the merging problem of multilane highway in mixed traffic environments. This strategy includes a PLC model for mainline CAVs, and a merging control method based on HRO-MADDPG. The results demonstrate that the proposed strategy effectively mitigates the problem of increased density in the mainline outermost lane due to merging and enhances the uniformity of traffic flow distribution across multiple lanes. HRO-MADDPG outperforms existing algorithms in terms of convergence speed, traffic efficiency, and safety. Additionally, it significantly improves traffic efficiency under varying CAV penetration rates and ramp merging rates. Moreover, the strategy ensures good driving comfort.

Building on the previous research, we believe that the HRO-MADDPG algorithm used in this article still has room for further improvement. By incorporating the actual traffic rules of highway merging areas in mixed traffic environments, we plan to introduce a CAV priority assignment model and use the trajectory information of HVs within a certain prediction range to provide safety supervision for CAV decision-making. This improvement will not only reduce the collision rate but also expedite the training process of the algorithm.

On the other hand, accurate trajectory prediction and effective information transmission are crucial for the successful implementation of the proposed strategy. Both directly impact the safety of real-world merging control. They not only enable CAVs to accurately assess lane-changing conditions and respond swiftly to traffic dynamics, but also encourage CAVs to take necessary evasive actions in advance. Deviations in trajectory prediction and fluctuations in communication quality can lead to unnecessary braking or incorrect lane changes. Future research will focus on improving trajectory prediction accuracy and wireless communication reliability. We will study the effects of different prediction models, processing times, communication ranges, and system delays on vehicle control to further strengthen the proposed strategy.

REFERENCES

[1] Z. El Abidine Kherroubi, S. Aknine, and R. Bacha, "Novel decision-making strategy for connected and autonomous vehicles in highway on-ramp merging," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12490–12502, Aug. 2022.

[2] O. Nassef, L. Sequeira, E. Salam, and T. Mahmoodi, "Building a lane merge coordination for connected vehicles using deep reinforcement learning," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2540–2557, Feb. 2021.

[3] H. Dong, H. Zhang, F. Ding, H. Tan, and J. Peng, "Battery-aware cooperative merging strategy of connected electric vehicles based on reinforcement learning with hindsight experience replay," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 3, pp. 3725–3741, Sep. 2022.

[4] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[5] Z. Zheng et al., "Capacity of vehicular networks in mixed traffic with CAVs and human-driven vehicles," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 17852–17865, May 2024.

[6] D. Chen et al., "Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11623–11638, Nov. 2023.

[7] J. Liu, W. Zhao, and C. Xu, "An efficient on-ramp merging strategy for connected and automated vehicles in multi-lane traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5056–5067, Jun. 2022.

[8] K. Hou, F. Zheng, X. Liu, and G. Guo, "Cooperative on-ramp merging control model for mixed traffic on multi-lane freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10774–10790, Oct. 2023.

[9] J. Liu, W. Zhao, C. Wang, C. Xu, L. Li, and Q. Chen, "Eco-friendly on-ramp merging strategy for connected and automated vehicles in heterogeneous traffic," *IEEE Trans. Veh. Technol.*, vol. 72, no. 11, pp. 13888–13900, Nov. 2023.

[10] R. Chen and Z. Yang, "Cooperative ramp merging strategy at multi-lane area for automated vehicles," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 14326–14340, Oct. 2024.

[11] A. Wong, T. Bäck, A. V. Kononova, and A. Plaat, "Deep multiagent reinforcement learning: Challenges and directions," *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 5023–5056, 2023.

[12] J. Chen et al., "Global-and-local attention-based reinforcement learning for cooperative behaviour control of multiple UAVs," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4194–4206, Mar. 2024.

[13] H. U. Sheikh and L. Bölöni, "Multi-agent reinforcement learning for problems with combined individual and team reward," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2020, pp. 1–8.

[14] Y. Xue, C. Ding, B. Yu, and W. Wang, "A platoon-based hierarchical merging control for on-ramp vehicles under connected environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21821–21832, Nov. 2022.

[15] R. Chen and Z. Yang, "A cooperative merging strategy for connected and automated vehicles based on game theory with transferable utility," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19213–19223, Oct. 2022.

[16] C. Zhang, J. Qi, Y. He, M. A. Shafique, and J. Zhao, "Collision-free merging control via trajectory optimization for connected and autonomous vehicles," *Transport. Res. Rec.*, vol. 2678, no. 8, pp. 1077–1087, 2024.

[17] S. Jing, F. Hui, X. Zhao, J. Rios-Torres, and A. J. Khattak, "Integrated longitudinal and lateral hierarchical control of cooperative merging of connected and automated vehicles at on-ramps," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24248–24262, Dec. 2022.

[18] J. Chen, Y. Zhou, and E. Chung, "An integrated approach to optimal merging sequence generation and trajectory planning of connected automated vehicles for freeway on-ramp merging sections," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 2, pp. 1897–1912, Feb. 2024.

[19] H. Ding, Y. Di, X. Zheng, H. Bai, and W. Zhang, "Automated cooperative control of multilane freeway merging areas in connected and autonomous vehicle environments," *Transportmetrica B, Transp. Dyn.*, vol. 9, no. 1, pp. 437–455, 2021.

[20] Y. Hu, C. Yu, Z. Su, W. Ma, Z. Chen, and J. Hou, "Critical trajectory point planning for connected and autonomous vehicles on freeway on-ramps under mixed traffic environment," *IEEE Trans. Veh. Technol.*, vol. 73, no. 12, pp. 18156–18172, Dec. 2024.

[21] B. Brito, A. Agarwal, and J. Alonso-Mora, "Learning interaction-aware guidance for trajectory optimization in dense traffic scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18808–18821, Oct. 2022.

[22] B. Hu, L. Jiang, S. Zhang, and Q. Wang, "An explainable and robust motion planning and control approach for autonomous vehicle on-ramping merging task using deep reinforcement learning," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 3, pp. 6488–6496, Sep. 2024.

[23] H. Hassani, S. Nikan, and A. Shami, "Traffic navigation via reinforcement learning with episodic-guided prioritized experience replay," *Eng. Appl. Artif. Intell.*, vol. 137, Nov. 2024, Art. no. 109147.

[24] X. He and C. Lv, "Toward intelligent connected E-mobility: Energy-aware cooperative driving with deep multiagent reinforcement learning," *IEEE Veh. Technol. Mag.*, vol. 18, no. 3, pp. 101–109, Sep. 2023.

[25] X. Zhang, L. Wu, H. Liu, Y. Wang, H. Li, and B. Xu, "High-speed ramp merging behavior decision for autonomous vehicles based on multiagent reinforcement learning," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 22664–22672, Dec. 2023.

[26] L. Li, W. Zhao, C. Wang, A. Fotouhi, and X. Liu, "Nash double Q-based multi-agent deep reinforcement learning for interactive merging strategy in mixed traffic," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121458.

[27] M. Li, Z. Li, S. Wang, and S. Zheng, "Enhancing cooperation of vehicle merging control in heavy traffic using communication-based soft actor-critic algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6491–6506, Jun. 2023.

[28] S. Wang, H. Fujii, and S. Yoshimura, "Generating merging strategies for connected autonomous vehicles based on spatiotemporal information extraction module and deep reinforcement learning," *Physica A Stat. Mechan. Appl.*, vol. 607, Dec. 2022, Art. no. 128172.

[29] N. Chen, B. van Arem, T. Alkim, and M. Wang, "A hierarchical model-based optimization control approach for cooperative merging by connected automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7712–7725, Dec. 2021.

[30] H. Yeo, A. Skabardonis, J. Halkias, J. Colyar, and V. Alexiadis, "Oversaturated freeway flow algorithm for use in next generation simulation," *Transport. Res. Record*, vol. 2088, no. 1, pp. 68–79, 2008.

[31] L. Liu, X. Li, Y. Li, J. Li, and Z. Liu, "Reinforcement-learning-based multi-lane cooperative control for on-ramp merging in mixed-autonomy traffic," *IEEE Internet Things J.*, , vol. 11, no. 24, pp. 39809–39819, Dec. 2024.

[32] Z. Wang, Y. Xue, L. Liu, H. Zhang, C. Qu, and C. Fang, "Multi-agent DRL-controlled connected and automated vehicles in mixed traffic with time delays," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 11, pp. 17676–17688, Nov. 2024.

[33] C. Yu et al., "Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 735–748, Feb. 2020.

[34] S. Wang, Z. Wang, R. Jiang, R. Yan, and L. Du, "Trajectory jerking suppression for mixed traffic flow at a signalized intersection: A trajectory prediction based deep reinforcement learning method," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18989–19000, Oct. 2022.

[35] T. Wang, J. Cao, and A. Hussain, "Adaptive traffic signal control for large-scale scenario with cooperative group-based multi-agent reinforcement learning," *Transp. Res. C, Emerg. Technol.*, vol. 125, Apr. 2021, Art. no. 103046.

[36] L. Xiao, M. Wang, W. Schakel, and B. van Arem, "Unravelling effects of cooperative adaptive cruise control deactivation on traffic flow characteristics at merging bottlenecks," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 380–397, Nov. 2018.

[37] L. Xiao, M. Wang, and B. Van Arem, "Realistic car-following models for microscopic simulation of adaptive and cooperative adaptive cruise control vehicles," *Transp. Res. Rec.*, vol. 2623, no. 1, pp. 1–9, 2017.

[38] L. Zhu, L. Lu, X. Wang, C. Jiang, and N. Ye, "Operational characteristics of mixed-autonomy traffic flow on the freeway with on-and off-ramps and weaving sections: An RL-based approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13512–13525, Aug. 2022.

[39] S. Chen, J. Dong, P. Ha, Y. Li, and S. Labi, "Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 7, pp. 838–857, 2021.

[40] Y. Wu, H. Tan, L. Qin, and B. Ran, "Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102649.

[41] C. Yu et al., "The surprising effectiveness of PPO in cooperative multi-agent games," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 24611–24624.

[42] K. Lin, R. Zhao, Z. Xu, and J. Zhou, "Efficient large-scale fleet management via multi-agent deep reinforcement learning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 1774–1783.

[43] Z. Yang, Z. Zheng, J. Kim, and H. Rakha, "Eco-cooperative adaptive cruise control for platoons in mixed traffic using single-agent and multi-agent reinforcement learning," *Transp. Res. D, Transp. Environ.*, vol. 142, May 2025, Art. no. 104658.

[44] P. Chen, H. Ni, L. Wang, G. Yu, and J. Sun, "Safety performance evaluation of freeway merging areas under autonomous vehicles environment using a co-simulation platform," *Accid. Anal. Prevent.*, vol. 199, May 2024, Art. no. 107530.

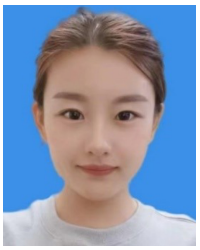
[45] G. Paddan and M. Griffin, "Evaluation of whole-body vibration in vehicles," *J. Sound Vib.*, vol. 253, no. 1, pp. 195–213, 2002.

Rui Peng received the M.E. degree in transportation engineering from Hebei University of Technology, Tianjin, China, in 2022. He is currently pursuing the Ph.D. degree with the School of Transportation, Southeast University, Nanjing, China.

His current research interests include dynamic traffic control and intelligent vehicles.

Min Yang received the Ph.D. degree in transportation planning and management from the School of Transportation, Southeast University, Nanjing, China, in 2007.

He is currently a Professor with Southeast University. His research interests include intelligent transportation systems, traffic flow, and travel demand forecasting.



Rui Tao received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2016. She is currently pursuing the Ph.D. degree with the School of Civil and Transportation Engineering, Hebei University of Technology, Tianjin, China.

Her research interests include traffic safety analysis and intelligent transportation systems.

Mingye Zhang received the M.E. degree in transportation engineering from Jilin University, Changchun, China, in 2021. He is currently pursuing the Ph.D. degree with the School of Transportation, Southeast University, Nanjing, China.

His current research interests include traffic analysis and control.

Renjie Zhang received the M.E. degree in transportation engineering from Nanjing University of Science and Technology, Nanjing, China, in 2022. He is currently pursuing the Ph.D. degree with the School of Transportation, Southeast University, Nanjing, China.

His current research interests include optimization and control of intelligent transportation systems.