actuators

MDPI

Article

Multi-Agent Cooperative Control of CAVs in Toll Plaza Diverging Areas: A Target-Path Approach

Siyu Long [1], Lili Zheng [1,*] and Yi Fei [2]

[1] Transportation College, Jilin University, Changchun 130022, China; longsy0323@mails.jlu.edu.cn
[2] School of Software, Henan University of Engineering, Zhengzhou 451191, China; yifei@haue.edu.cn
* Correspondence: lilizheng@jlu.edu.cn; Tel.: +86-13756167376

Abstract

Existing research on cooperative control of connected and autonomous vehicles (CAVs) has primarily focused on structured freeway environments. Most existing approaches adopt lane-based modeling and discrete lane-change actions. These assumptions are unsuitable for toll plaza diverging areas without lane markings, where vehicles move toward multiple tollbooths. The absence of predefined lanes leads to continuous trajectory evolution, dense interactions, and increased safety risk. To address this limitation, this study proposes a multi-agent cooperative control framework based on Multi-Agent Proximal Policy Optimization (MAPPO) under a Centralized Training and Decentralized Execution (CTDE) architecture. The multi-agent formulation captures multi-vehicle interaction in toll plaza diverging areas, while centralized training improves learning stability. A target-path-oriented action space is introduced to replace the discrete lane-change action, enabling flexible tollbooth selection and continuous trajectory generation. The proposed cooperative strategy is trained and evaluated on a simulation platform structured under a Perception-Decision-Action framework, which provides a high-fidelity environment for weak-constraint traffic interactions. Simulation results based on real-world traffic data show that the proposed method improves traffic efficiency and enhances collision avoidance. Furthermore, comparative analyses are conducted to evaluate the model performance under varying traffic environments.

Keywords: toll plaza diverging area; weak-constraint traffic; cooperative vehicle control; MAPPO algorithm

Check for updates

Academic Editor: Xiaozheng Jin

Received: 18 March 2026
Revised: 29 April 2026
Accepted: 29 April 2026
Published: 8 May 2026

Copyright: © 2026 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.

1. Introduction

As a typical class of unmanned ground vehicles (UGVs), connected and autonomous vehicles (CAVs) are increasingly deployed in various transportation scenarios. With their rapid deployment, cooperative control has become a fundamental component of next-generation intelligent transportation systems [1,2]. While substantial progress has been achieved on freeway segments, toll plaza diverging areas remain insufficiently studied [3]. Unlike structured freeway environments, toll plaza diverging areas are more prone to traffic accidents [4], due to their weak lane constraints, high lateral maneuvering freedom, and intensive multi-vehicle weaving interactions [5,6]. The vehicles gradually disperse from upstream approach lanes toward multiple tollbooths without rigid lane guidance, generating continuous lateral movements and complex conflict patterns. This characteristic fundamentally differentiates toll plaza diverging areas from lane-based traffic systems.

Extensive research has investigated cooperative control strategies [7]. However, most existing cooperative control strategies, whether rule-based [8], optimization-driven [9,10], or learning-based [11,12], are developed under explicit lane-based traffic area assumptions. The vehicle control strategies are typically represented as discrete lane-change decisions combined with longitudinal acceleration control [13]. These strategies implicitly constrain conflict modeling within adjacent lanes and assume instantaneous lateral transitions. In rule-based and optimization-based methods, maintaining computational tractability often requires artificial discretization of lanes or trajectories, which constrains the action space [14]. In toll plaza diverging areas, vehicle trajectories evolve in a continuous two-dimensional space where lateral displacement is gradual rather than instantaneous, and conflict regions are spatially overlapping rather than lane-adjacent [15,16]. Under such conditions, the lane-based method fails to reflect the reality of the toll plaza diverging areas. Consequently, although these methods perform well in structured networks, their effectiveness diminishes in weakly constrained diverging regions.

In recent years, Multi-Agent Reinforcement Learning (MARL) has emerged as a powerful method for cooperative vehicle control, enabling agents to learn policies directly through interaction with the environment [17,18]. MARL has shown promising results in freeway merging, variable speed limit control [19,20], and lane-change coordination [21,22], where the state and action spaces remain structured and lane-indexed. Nevertheless, most existing MARL implementations still inherit discrete maneuver abstractions and structured geometric assumptions.

Despite notable progress in vehicle control strategies, a clear gap remains in modeling and coordinating vehicle behaviors in toll plaza diverging areas. Existing methods are largely developed under structured lane assumptions and rely on discrete lane-change or acceleration actions, which cannot capture continuous lateral movements and flexible path selection in weakly constrained environments. In addition, most mainstream traffic simulation platforms inherently adopt one-dimensional, lane-based car-following rules, even when modeling diverging areas. Such modeling assumptions further constrain the representation of vehicle interactions and limit the fidelity of safety evaluation in toll plaza scenarios.

To address this gap, we propose a cooperative control framework based on Multi-Agent Proximal Policy Optimization (MAPPO) under a Centralized Training and Decentralized Execution (CTDE) architecture. The architecture is designed for cooperative decision-making in weak-constraint toll plaza diverging areas and is evaluated in the developed high-fidelity simulation environment. Instead of modeling discrete lane-change actions, a target-path-oriented action space is introduced to represent feasible tollbooth paths under weak lane constraints. The primary contributions of this research are summarized as follows:

(1) A MAPPO-based cooperative control framework is developed for CAVs in mixed traffic flow at toll plaza diverging areas. Different from existing hierarchical or lane-based decision formulations for structured road environments, the proposed method is tailored to cooperative target toll lane selection and maneuvers under weak lane constraints.

(2) By adopting a high-fidelity simulation environment, the state and reward functions are specifically customized for the weakly constrained diverging scenario. The design incorporates path accessibility, queue conditions, surrounding vehicle distribution, and steering-related motion characteristics. This enables the learned policy to better balance traffic efficiency and safety in complex multi-vehicle interactions.

The remainder of this paper is organized as follows. Section 2 presents the methodology. Section 3 describes the simulation-platform development. Section 4 introduces the

multi-agent cooperative decision model. Section 5 outlines data preprocessing and model configuration. Section 6 reports the simulation results and provides analytical discussions. Section 7 concludes the study and highlights future research directions.

2. Methodology

Figure 1 presents the overall methodological framework of this study. The framework consists of two connected parts: a two-dimensional microscopic simulation platform for toll plaza diverging areas under weak lane constraints (left side of Figure 1), and a MAPPO-based multi-agent cooperative control model (right side). The simulation platform provides an interactive environment for policy learning and performance evaluation, while the cooperative control model determines the actions of CAVs within that environment.

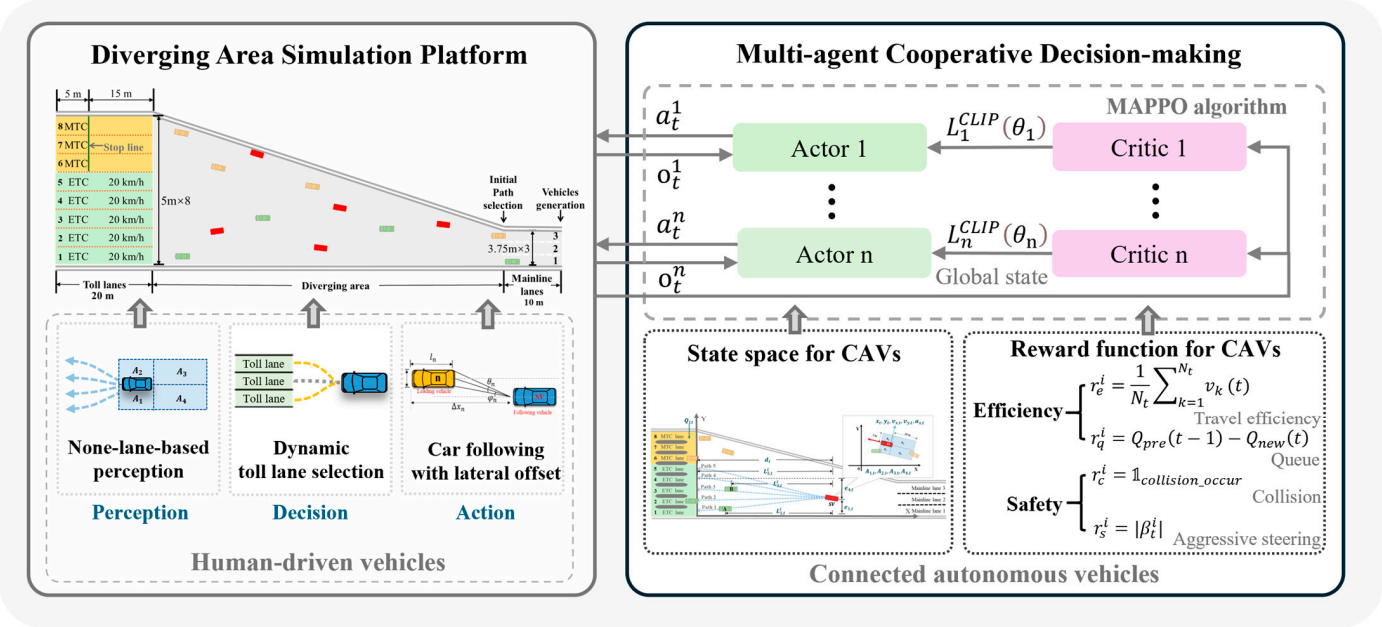


Figure 1. Overall methodology.

The microscopic simulation platform based on the Perception-Decision-Action (PDA) framework is adopted from our previous work [6,23]. The platform is specifically designed for toll plaza diverging areas, where vehicles move toward their target toll lanes without lateral constraints, giving rise to intensive weaving interactions. To accurately capture these driving behaviors, the platform integrates three core components: (i) accessible path perception, which maps feasible paths toward tollbooths in continuous space; (ii) dynamic toll lane selection, which determines the routing of vehicles under real-time traffic conditions; and (iii) a car-following model considering lateral offsets. These components enable the platform to reproduce realistic weak-constraint driving behaviors in the diverging area and provide a reliable environment for reinforcement learning training.

Based on this environment, each CAV is modeled as an agent in the MAPPO framework. At time step t, agent i receives its observation o_t^i from the simulation environment and outputs an action a_t^i. Here, a_t^i includes longitudinal acceleration control and target toll lane selection. The reward function is designed from the perspectives of both traffic efficiency and safety to encourage coordinated behaviors among agents. Under the CTDE architecture, the actors generate actions based on individual states, while the critics are trained using the global state to evaluate the joint behavior of all agents, and the policy of each actor is optimized through the clipped objective $L_i^{CLIP}(\theta_i)$ [24]. Through repeated

interaction with the simulation platform, the policy is iteratively updated and then used for cooperative control and performance evaluation in the diverging area.

3. Simulation Platform Establishment

The simulation platform is built on the PDA framework to simulate the driver's cognitive process in the toll plaza diverging areas. It decomposes the complex driving behaviors into three aspects: accessible path perception, dynamic toll lane selection, and a car-following model considering lateral offsets. The main components of this platform are introduced below.

3.1. Accessible Path Perception

As mentioned earlier, although vehicles in the diverging area have a longitudinal target lane, they lack lateral motion constraints and typically drive directly toward the end of their target toll lane's queue. During this process, vehicles no longer rely on lane markings but on paths to each accessible toll lane to guide their perception and make decisions [25,26]. To incorporate this characteristic, this study proposes a path-oriented perception method, with its two components detailed below.

3.1.1. Accessible Diverging Path Generation

Polynomial curves are commonly used to characterize the lane-changing trajectories [27], as their continuous curvature ensures smooth transitions in velocity and acceleration. Therefore, a cubic polynomial function is used to simulate the diverging paths of vehicles, as it provides a good balance between computational efficiency and trajectory fitting performance in the present simulation scenario:

$$f(x) = ax^3 + bx^2 + cx + d, \tag{1}$$

where a, b, and c represent parameters of the curve function, d is a constant term.

As shown in Figure 2, the path-generation mechanism illustrates how multiple accessible paths are constructed for vehicles in the toll plaza diverging area under the mixed ETC-MTC tolling scheme. Since the electronic toll collection (ETC) system has not yet achieved full coverage in China, toll plazas typically operate under a mixed tolling scheme that integrates ETC and manual toll collection (MTC), resulting in multiple feasible approach paths toward different toll booths within the diverging area. The accessible paths shown in Figure 2 are generated using the polynomial curve function defined above. The path parameters are defined by the coordinates of four key points along the vehicle's trajectory during its diverging process, including the vehicle's current and previous positions, and two fixed points on the centerline of the accessible toll lane. When a vehicle enters the diverging area (e.g., ETC vehicle SV 1), the simulation platform will generate multiple accessible paths (Path 1 to Path 5) based on its two trajectory points before entering the diverging area (P1 and P2) and two fixed points on the center-line of each accessible ETC lane (P3 and P4). If a vehicle detects a preceding vehicle (e.g., SV2) on its current path during the diverging process, it will dynamically generate new candidate paths (Path 1′ and Path 2′). These paths are created based on the vehicle's current and previous positions (P1′ and P2′), as well as fixed points on the center lines of other accessible lanes (e.g., P3′ and P4′), effectively simulating the vehicle's dynamic adjustment behavior.

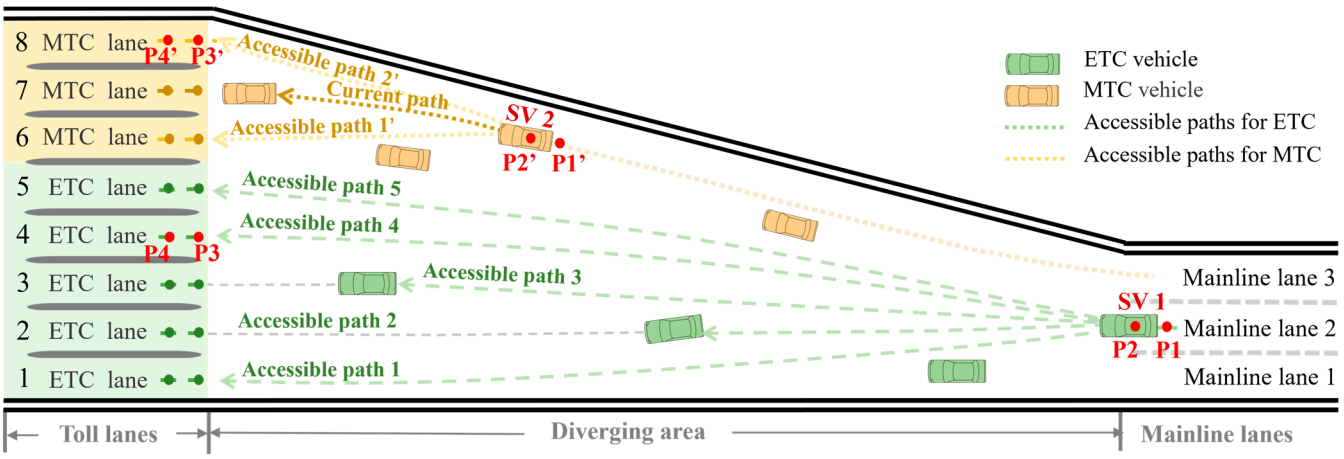


Figure 2. Illustration of accessible path generation.

3.1.2. Perception Based on Path

Vehicle perception information in the diverging area is divided into two categories: vehicle-related information and path-related information. The detailed definitions of all variables are given in Table 1, while Figure 3 provides a schematic illustration of these variables in the diverging area.

Table 1. Variable definition.

Variable		Description
Vehicle-related variables	$x_{i,t}$	Longitudinal position of SV at time step t.
	$y_{i,t}$	Lateral position of SV at time step t.
	$v_{x,i,t}$	The velocity of SV in X direction at time step t.
	$v_{y,i,t}$	The velocity of SV in Y direction at time step t.
	$a_{x,i,t}$	Longitudinal acceleration of SV at time step t.
	T_c	The current toll collection type of SV, 0 for a MTC vehicle, 1 for an ETC vehicle.
	L_c	The initial lane of SV before it enters the diverging area.
	$A^i_{1,t}$	Presence of another vehicle in the left area at time t. (1 = Yes, 0 = No)
	$A^i_{2,t}$	Presence of another vehicle in the right area at time t. (1 = Yes, 0 = No)
	$A^i_{3,t}$	Presence of another vehicle in the right-behind area at time t. (1 = Yes, 0 = No)
	$A^i_{4,t}$	Presence of another vehicle in the left-behind area at time t. (1 = Yes, 0 = No)
Path-related variables	$L^i_{j,t}$	Available longitudinal distance on path j at time t.
	$\beta^i_{j,t}$	Required steering magnitude for selecting path j at time t. (positive: leftward turn, negative: rightward turn)
	$Q^i_{j,t}$	The number of vehicles queued on path j at time t

- Vehicle-related information: It comprises three categories: (i) dynamic kinematic states, including vehicle's longitudinal and lateral position ($x_{i,t}$, $y_{i,t}$), longitudinal and lateral velocity ($v_{x,i,t}$, $v_{y,i,t}$), and longitudinal acceleration ($a_{x,i,t}$); (ii) static attributes, including vehicle's toll collection type (T_c) and initial lane (L_c); and (iii) surrounding vehicle indicators ($A^i_{1,t} - A^i_{4,t}$) for the presence of other vehicles in predefined surrounding zones.
- Path-related information: It includes the longitudinal available moving distance ($L^i_{j,t}$), lateral moving magnitude ($\beta^i_{j,t}$), and the queue length ($Q^i_{j,t}$) for path j at time step t, where j is the toll lane number.

As illustrated in Figure 3, the diverging area consists of five ETC lanes and three MTC lanes. Therefore, the accessible paths to the current vehicle depend on its toll collection

type: for an ETC vehicle, its accessible paths are $j = 1, 2, 3, 4, 5$; if it is an MTC vehicle, $j = 6, 7, 8$. For a given accessible path j, the value of $L^i_{j,t}$ depends on whether there is a preceding vehicle on that path. If a preceding vehicle is present, $L^i_{j,t}$ is equal to the longitudinal distance between the current vehicle and the preceding vehicle. For example, in Figure 3, the preceding vehicles A and B are located on Paths 1 and 3, respectively; therefore, $L^i_{1,t}$ and $L^i_{3,t}$ represent the longitudinal distances from the current vehicle to A and B. By contrast, if no preceding vehicle exists on path j, $L^i_{j,t}$ reduces to d_t, where d_t denotes the longitudinal distance from the current vehicle to the toll lane entrance line. So in Figure 3, since Paths 2, 4, and 5 are free of preceding vehicles, $L^i_{2,t}$, $L^i_{4,t}$, and $L^i_{5,t}$ are all equal to d_t. In addition, $\beta^i_{j,t}$ denotes the required steering magnitude for selecting path j at time t, defined as $\beta^i_{j,t} = e^i_{j,t}/d_t$, where $e^i_{j,t}$ is the lateral distance from the current position of agent i to the merge-in point of path j at time t. Finally, $Q^i_{j,t}$ denotes the number of vehicles queued on path j at time t.

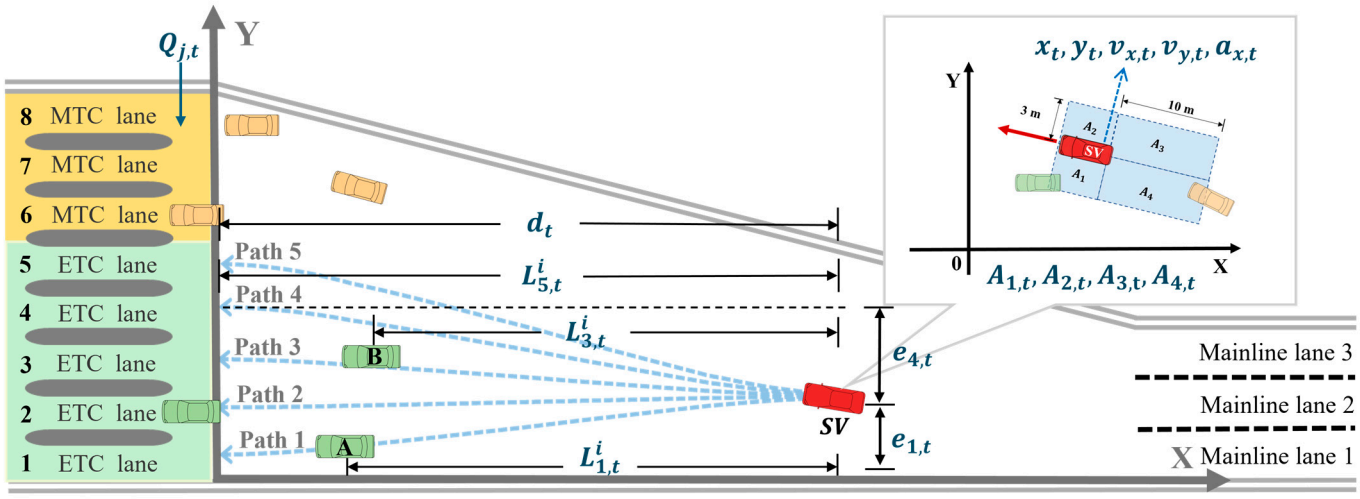


Figure 3. Illustration of vehicle state and path information. Different vehicle types are distinguished by colors: red for CAVs, yellow for human-driven MTC vehicles (MTC HV), and green for human-driven ETC vehicles (ETC HV). SV denotes the subject vehicle. A and B represent preceding vehicles on different paths. The blue dashed lines indicate the generated accessible paths for SV.

3.2. Dynamic Toll Lane Decision

Based on the environmental information provided by the perception layer, the decision layer dynamically selects a target toll lane to simulate the decision adjustment in the human driving process. Upon entering the area, the vehicle first selects an initial target lane and then makes real-time adjustments based on the traffic conditions ahead. This dynamic path selection can be framed as a multi-class classification problem [28,29]. To model this behavior, we use a Multi-layer Perceptron (MLP) neural network [30] that takes vehicle-related and path-related information (as detailed in Table 1) as its input and, in turn, outputs the vehicle's optimal target toll lane at each time step t. The neural network was trained offline using trajectory-based target lane selection samples extracted from the study [6] and achieved over 90% prediction accuracy.

3.3. Car-Following Model Considering Lateral Offsets

On structured road segments with clear lane markings, car-following models often assume that the centers of the preceding and following vehicles are on the same straight line. However, in unstructured scenarios like diverging areas without lane markings, there exists a significant lateral offset between the preceding and following vehicles [31]. In these cases, common car-following models fail to accurately capture their driving behaviors.

Therefore, it is necessary to modify the original model to incorporate the weak-constraint motion features at the diverging area.

The Full Velocity Difference (FVD) model is one of the most widely used car-following models. It simulates the following vehicle's reaction to its predecessor by comprehensively considering factors such as the spacing, the velocity difference, and the driver's response sensitivity. The model can be formulated as:

$$a_{SV}(t) = \alpha\{V[\Delta x_{SV}(t)] - V_{SV}(t)\} + \lambda \Delta V_{SV}(t), \tag{2}$$

where $V[\Delta x_{SV}(t)]$ is the driver's optimal speed function based on the spacing to the preceding vehicle $\Delta x_{SV}(t)$; $V_{SV}(t)$ is the velocity of the following vehicle at time t; $\Delta V_{SV}(t)$ is the velocity difference between the current vehicle and the leading vehicle; α and λ respectively represent the sensitivity coefficient of the driver to the difference between the optimal velocity and the current velocity.

To account for the substantial lateral offsets between vehicles in the diverging area, an improved FVD model is adopted to describe car-following behavior [32]. Figure 4 illustrates the car-following geometry with lateral offset, including the relative position relationship between the leading and following vehicles as well as the associated visual and offset angles. The modified FVD model is given by:

$$a_{SV}(t) = \alpha\{V[\Delta x_{SV}(t)] - V_{SV}(t)\} - \lambda_1 \frac{d\theta(t)}{dt} + \lambda_2 \frac{d\varphi(t)}{dt} \tag{3}$$

where λ_1 and λ_2 are sensitivity coefficients to the rate of change in the visual angle and the lateral offset angle ($\lambda_1 = 40,\ \lambda_2 = 20$). $\theta(t)$ is the visual angle of the preceding vehicle, while $\varphi(t)$ is the lateral offset angle.

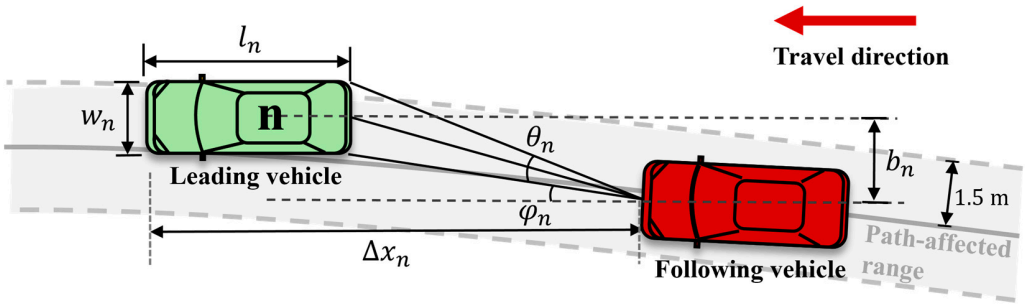


Figure 4. Schematic of the car-following model considering lateral offsets.

The specific calculation methods for these angles are as follows:

$$\theta(t) = \arctan\frac{b_n(t) + w_n/2}{\Delta x_n(t) - l_n} - arctan\frac{b_n(t) - w_n/2}{\Delta x_n(t) - l_n}, \tag{4}$$

$$\varphi(t) = arctan\frac{b_n(t)}{\Delta x_n(t) - l_n}, \tag{5}$$

where the vehicle length l_n and width w_n are set to 5 m and 1.6 m in this study.

The optimal velocity $V[\Delta x_{SV}(t)]$ is calculated as follows:

$$V[\Delta x_{SV}(t)] = V_1 + V_2 \tanh\left(C_1 \frac{b_n(t)}{\tan\varphi_n(t)} + C_2\right), \tag{6}$$

where V_1, V_2, C_1, and C_2 are the parameters for the optimal velocity function. This study adopts the classic parameter set ($V_1 = 6.75,\ V_2 = 7.91,\ C_1 = 0.13,\ C_2 = 1.57$), which was

determined based on empirical traffic data from Stuttgart, Germany, and has been used in subsequent car-following research [33].

Note that $\frac{b_n(t)}{\tan \varphi_n(t)}$ in Equation (6) is geometrically equivalent to the effective longitudinal spacing $\Delta x_n(t) - l_n$ in Equation (5), and thus remains defined when $\varphi_n(t)$ approaches zero.

To ensure continuous car-following behavior even when no preceding vehicle is on the chosen path, the simulation platform places a virtual vehicle at the end of the target toll lane. This allows the current vehicle to adhere to the car-following rules throughout the diverging process.

4. Multi-Agent Cooperative Decision Model

To address the challenge of multi-vehicle cooperative control in the weak-constraint environment of the toll plaza diverging area, each CAV is defined as an agent capable of independent decision-making. By adopting the CTDE architecture, the model enables these agents to train using global information while executing efficient and safe real-time coordination based solely on local observations. The following sections describe the agent's action space, state space, reward function, and the principles of the MAPPO algorithm.

4.1. Action Space

In structured road environments with explicit lane markings, an agent's action space is typically defined by discrete actions such as lane change or lane keeping [34]. However, this definition is inapplicable in the weak-constraint environment of a toll plaza diverging area, which lacks clear lane markings. The actual vehicle movement in this area is characterized by a continuous and smooth merging motion toward the target toll lane, unconstrained by fixed lanes. To accurately represent this weak-constraint driving behavior, this paper redesigns the action space for each agent. The action a_t^i executed by agent i at each simulation time step t comprises two components: longitudinal acceleration control and target lane selection, formulated as follows:

$$a_t^i = [a_{acc}(t), a_{lane}(t)], \tag{7}$$

where the longitudinal acceleration $a_{acc}(t)$ of the agent is constrained to the range $[-4,\ 3]$ $\mathrm{m/s^2}$. $a_{lane}(t)$ represents the target toll lane number of the agent. Accordingly, the value of $a_{lane}(t)$ is drawn from the accessible toll lane set $\mathcal{J}_i$, corresponding to the agent's specific toll collection type:

$$\mathcal{J}_i = \begin{cases} \{1,2,3,4,5\}, & if\ ETC \\ \{6,7,8\}, & if\ MTC \end{cases}. \tag{8}$$

Consequently, the complete action space A_i of each agent i, defined as the set of all possible actions, can be formally expressed as:

$$\mathcal{A}_i = \{(a_{acc}, a_{lane}) \mid a_{acc} \in [-4,3], a_{lane} \in \mathcal{J}_i\}. \tag{9}$$

4.2. State Space

Under the CTDE architecture adopted in this paper, the definition of the state information involves a clear distinction between local observations for each agent and the global state used by the centralized controller for evaluation during training.

- Local observation space: During the decentralized execution phase, each agent i solely perceives the environment information through its own sensors, then forms a local observation o_t^i. This allows any differences in their traffic performance to be attributed solely to their respective control strategies or human behavioral models.

Specifically, o_t^i includes the vehicle's ego state ($x_{i,t}$, $y_{i,t}$, $v_{x,i,t}$, $v_{y,i,t}$,$a_{x,i,t}$), surrounding vehicles information ($A_{1,t}^i$–$A_{4,t}^i$), and path-related information ($Q_{j,t}^i$, $L_{j,t}^i$, $\beta_{j,t}^i$).

Therefore, the local observation can be formalized as:

$$o_t^i = \left[x_{i,t}, y_{i,t}, v_{x,i,t}, v_{y,i,t}, a_{x,i,t}, \left\{A_{k,t}^i\right\}_{k=1}^4, \left\{Q_{j,t}^i, L_{j,t}^i, \beta_{j,t}^i\right\}_{j\in\mathcal{J}_i}\right]^\top. \tag{10}$$

- Global state space: During the centralized training phase, the critic network takes the global state information as input to accurately estimate the expected joint return of the agents, enabling the learning of cooperative policies. Consequently, the global state is defined as:

$$s_t = \left[o_t^1, o_t^2, \ldots, o_t^N\right], \tag{11}$$

where N is the total number of agents in the diverging area at the current time step t.

4.3. Reward Function

Conventional lateral decision-making models typically calculate rewards solely based on the state upon lane-change completion, thereby overlooking the vehicle's behavior throughout the motion. To address this limitation, the reward function we propose provides agent i with immediate feedback at each time step. It jointly optimizes for both traffic efficiency and safety, as detailed below.

4.3.1. Traffic Efficiency Reward

To promote coordination among agents in the diverging area and improve overall traffic efficiency, we introduce the average speed of all agents within the area at each time step t as a shared reward, which is then distributed to each agent:

$$r_e^i = \frac{1}{N_t}\sum_{k=1}^{N_t} v_k(t), \tag{12}$$

where N_t denotes the total number of agents in diverging area at time t, and $v_k(t)$ represents the speed of CAV_k at time t. If $N_t = 0$, the reward r_e^i is set to 0. This reward encourages cooperative behaviors by reflecting the overall traffic efficiency at each time step.

Furthermore, to balance traffic load across lanes, we introduce a queue-balancing reward r_q^i that encourages agents to select toll lanes with shorter queues, formulated as follows:

$$r_q^i = Q_{t-1} - Q_t, \tag{13}$$

where Q_{t-1} and Q_t denote the queue lengths on the target toll lane selected at the previous and current time step, respectively. If the currently selected lane has a shorter queue than the previous one, the agent receives a positive reward; otherwise, it receives a penalty.

4.3.2. Traffic Safety Reward

To guarantee vehicle safety during the diverging process, we incorporate two penalty mechanisms. First, a collision penalty is designed to train agents to avoid conflicts with surrounding vehicles. When an agent's lateral maneuver results in a collision, the penalty is triggered:

$$r_c^i = \begin{cases} 1, & if\ a\ collision\ occurs \\ 0, & otherwise \end{cases}. \tag{14}$$

Also, a steering penalty is designed to penalize aggressive steering actions:

$$r_s^i = \left|\beta_t^i\right|. \tag{15}$$

Here, β_t^i denotes the steering magnitude corresponding to the target toll lane selected by agent i at time step t. A larger value of r_s^i indicates more aggressive steering, thus leads to a heavier penalty.

The final weighted reward received by agent i at time t is formulated as:

$$r_t^i = \alpha_1 r_e^i + \alpha_2 r_q^i - \alpha_3 r_c^i - \alpha_4 r_s^i \tag{16}$$

where the weighting coefficients are set as $\alpha_1 = 0.1$, $\alpha_2 = 5$, $\alpha_3 = 20$, $\alpha_4 = 10$. The weights were determined through extensive testing to optimize the performance of the cooperative control strategy.

4.4. MAPPO Training Framework

MAPPO is adopted in this study because it provides better stability and convergence in the dynamic and highly interactive toll-plaza diverging environment than off-policy alternatives such as MADDPG or QMIX.

MAPPO is extended from the single-agent Proximal Policy Optimization (PPO) algorithm. When adapted to multi-agent scenarios, MAPPO assigns an independent actor and an independent critic network to each agent. Each agent's actor selects actions based on local observation o_t^i, with its clipped objective function formulated as follows:

$$L_i^{CLIP}(\theta^i) = \hat{\mathbb{E}}_t\left[\min\left(\rho_t^i(\theta^i)\hat{A}_t^i, clip\left(\rho_t^i(\theta^i), 1-\epsilon, 1+\epsilon\right)\hat{A}_t^i\right)\right], \tag{17}$$

where $\rho_t^i(\theta^i) = \frac{\pi_{\theta^i}^i(a_t^i|o_t^i)}{\pi_{\theta^i\,old}^i(a_t^i|o_t^i)}$ denotes the probability ratio between the new and old policies of agent i. $\hat{A}_t^i$ is the advantage function estimated by the critic network, and ϵ is the clipping parameter (typically set between 0.1 and 0.3). The $clip$ function restricts $\rho_t^i(\theta^i)$ to the interval $[1-\epsilon, 1+\epsilon]$, preventing excessive policy updates.

To estimate the advantage function $\hat{A}_t^i$ accurately, the critic network first receives the global state s_t during centralized training, outputs $V_\varphi^i(s_t)$, thus using generalized advantage estimation (GAE):

$$\hat{A}_t^i = \delta_t^i + (\gamma\lambda)\delta_{t+1}^i + \cdots + (\gamma\lambda)^{T-t+1}\delta_{T+1}^i, \tag{18}$$

$$\delta_t^i = r_t^i + \gamma V_\varphi^i(s_{t+1}) - V_\varphi^i(s_t), \tag{19}$$

where δ_t^i is the temporal difference (TD) error. γ is the discount factor and λ is the GAE parameter.

To ensure accurate value estimation, the critic network is trained to minimize the mean squared error between the predicted value and the discounted cumulative return R_t^i:

$$L_i^{critic}(\varphi^i) = \frac{1}{2}E\left[\left(V_\varphi^i(s_t) - R_t^i\right)^2\right], \tag{20}$$

where $R_t^i = \sum_{k=0}^{T-t}\gamma^k r_{t+k}^i$ represents the actual return obtained by agent i from time t to the end of the episode.

Furthermore, to prevent premature convergence to sub-optimal policies, a mixed entropy term $\mathcal{H}_{mix}$ is introduced. The overall objective function for the actor network is thus formulated as:

$$L_i^{actor}(\theta^i) = L_i^{CLIP}(\theta^i) + \varepsilon\mathcal{H}_{mix}, \tag{21}$$

with $\mathcal{H}_{mix} = H_{lane} + H_{acc}$, where H_{lane} sums over discrete lanes and H_{acc} integrates over $[-4,\ 3]$, and ε is the entropy coefficient.

Finally, the network parameters are iteratively updated using the Adam optimizer with learning rate η_{actor} and η_{critic}:

$$\theta^i \leftarrow \theta^i + \eta_{actor} \nabla_{\theta^i} L^i_{actor}(\theta^i), \tag{22}$$

$$\varphi^i \leftarrow \varphi^i - \eta_{critic} \nabla_{\varphi^i} L^i_{critic}(\varphi^i), \tag{23}$$

5. Simulation Experiments

5.1. Data Collection and Processing

The data for this study were collected at the Changsha West Toll Plaza on the G55 Changsha-Zhangjiajie Freeway (east to west), a major traffic node in the western part of Changsha. An aerial view of the study area is shown in Figure 5. The upstream section of the study area consists of three mainline lanes, each with a width of 3.75 m. It is followed by a diverging area approximately 145 m in length, which leads to a toll plaza with eight lanes. Five of these are ETC lanes on the left, and the remaining three are MTC lanes on the right. Each toll lane is 5 m wide. The vehicle trajectory data were captured through vertical aerial filming from an Unmanned Aerial Vehicle (UAV). The footage was recorded in May 2021. A total of 55 min of video was recorded in 4K resolution at 30 fps. After excluding segments with no traffic or heavy congestion to ensure data quality, approximately 25 min of continuous video footage were kept for analysis. Based on 10-min statistical intervals, the traffic flow during the observation period ranged from 1578 to 2004 vehicles per hour.

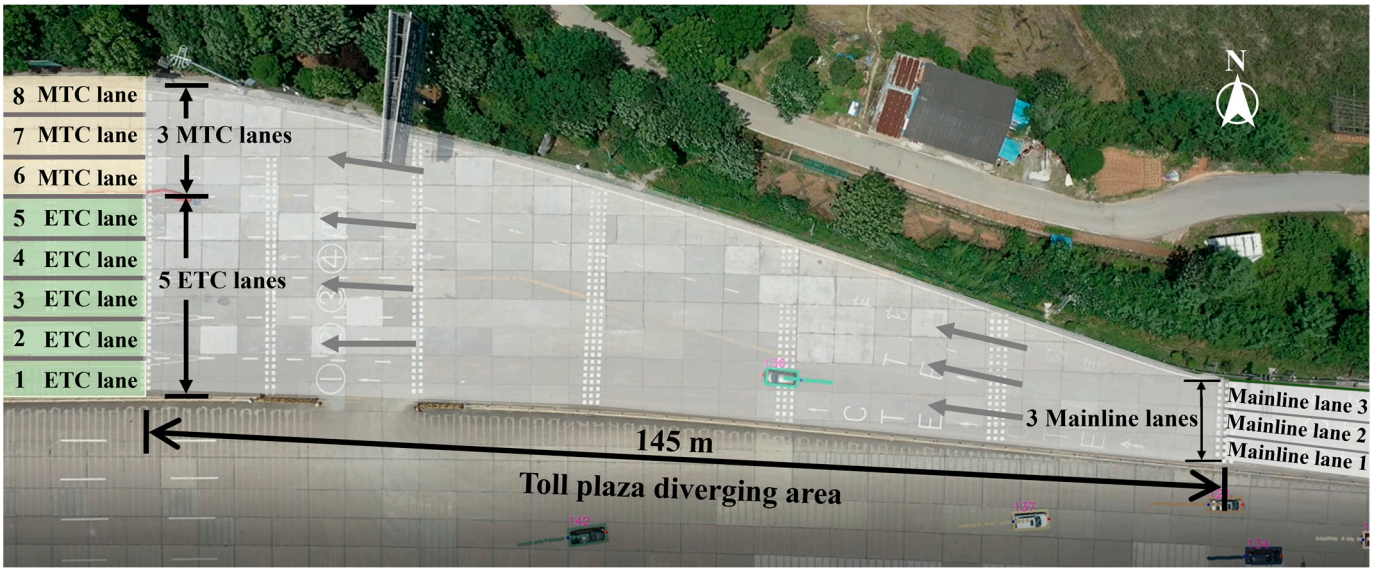


Figure 5. Aerial view of the diverging area at Changsha West Toll Station.

The vehicle trajectories were extracted from the video data using the Automated Roadway Conflicts Identify System (ARCIS) [35], developed by the University of Central Florida, Orlando, FL, USA. A final dataset of 692 complete vehicle trajectories was obtained through this method, comprising 628 passenger cars (439 ETC, 189 MTC) and 64 large vehicles (trucks and buses). Given their significant differences in acceleration performance, turning radii, and lane-changing behaviors compared to passenger cars, large vehicles were excluded from direct modeling and analysis. However, to ensure a realistic traffic environment, the trajectories of large vehicles were preserved and incorporated as background traffic flow within the simulation environment.

The number of vehicles entering the diverging area from each mainline lane and the throughput of each toll lane are presented in Table 2. The data reveal two primary trends. First, in terms of entry lane selection, ETC vehicles tended to enter from the middle mainline lanes, while MTC vehicles primarily opted for the outer lanes. Second, regarding toll lane selection, drivers showed a clear preference for the lane with the shortest lateral distance. These behavioral tendencies led to higher traffic volumes for ETC lanes 1–3 and MTC lanes 1–2 compared to other lanes of their respective types.

Table 2. Vehicle counts by entry lanes and toll lanes.

	Lane ID	**1**		**2**		**3**		**Total**	
Mainline lane	Toll type	**ETC**	**MTC**	**ETC**	**MTC**	**ETC**	**MTC**	**ETC**	**MTC**
	Vehicle counts	115	29	202	54	122	106	439	189
	Lane ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Toll lane	Toll type	**ETC**	**ETC**	**ETC**	**ETC**	**ETC**	**MTC**	**MTC**	**MTC**
	Vehicle counts	165	128	94	42	10	94	69	26

5.2. Model Setup

5.2.1. Simulation Platform Setup

A scenario was constructed in the simulation platform based on the actual layout of the Changsha West Toll Plaza diverging area, as shown in Figure 6. The figure presents the actual geometric layout of the simulated scenario, including the upstream mainline lanes, the widening diverging area, and the downstream toll lanes. The simulated highway mainline consists of a 10-m section with three 3.75-m-wide lanes (numbered 1–3). This section transitions into a gradually widening, diverging area and ultimately connects to a toll plaza with eight 5-m-wide toll lanes. The traffic volume is set to 1500 vehicles per hour. In this environment, both ETC and MTC vehicles are generated at the beginning of the mainline section, with initial conditions consistent with the measured data shown in Table 2: the lane-entry proportion of ETC vehicles is 1:2:1, with speeds following a normal distribution $N(13.7, 3^2)$ m/s; while the lane-entry proportion of MTC vehicles is 1:2:4, with speeds following $N(12, 3^2)$ m/s. The terminal rules for vehicles are defined as follows: ETC vehicles must pass through toll lanes at a speed not exceeding 20 km/h, whereas MTC vehicles stop for 20 s after traveling 15 m into the toll lane to simulate manual payment processing. To evaluate the system efficiency, the metric "diverging time" is defined as a vehicle's total travel time within the diverging area, excluding any dwell time within the toll lanes. The simulation process is executed on a self-developed Python platform, comprising three modules: a visualization interface, a simulation engine, and a data logging module. The accuracy of this platform has been validated in previous work [6,23].

5.2.2. MAPPO Algorithm Configuration

To train a cooperative driving policy for CAVs in the diverging area, our study integrated the MAPPO algorithm into the simulation platform. Both the actor and critic networks are MLPs, each consisting of one input layer, two hidden layers, and one output layer. The key hyper-parameters are specified in Table 3. During policy training, the CAV penetration rate was set to 50%. At the beginning of each episode, a corresponding proportion of the simulated human-driven vehicles was randomly designated as CAVs. The simulation step length was set to 0.1 s, each episode contained 10,000 time steps, and the model was trained for 500 episodes. The platform assumes no communication delay. The entire training procedure was conducted on a computer equipped with a 2.30 GHz Intel Core i7-12700H CPU, 32.0 GB RAM, and an NVIDIA GeForce RTX 3070 Ti GPU. The

program was implemented in Python 3.8, and the neural network was developed using TensorFlow 2.6.0.

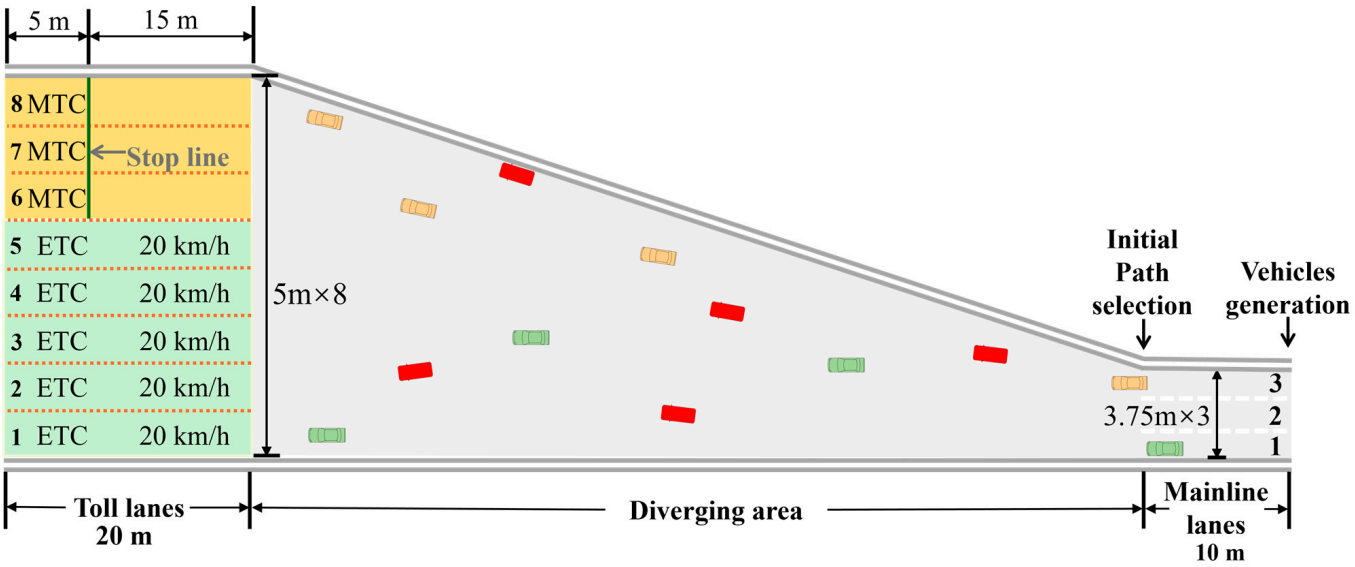


Figure 6. Layout of the simulation platform. Different vehicle types are distinguished by colors: red for CAVs, yellow for MTC HV, and green for ETC HV.

Table 3. Simulation parameters for MAPPO.

Parameters	Values	Parameters	Values
Number of hidden layers	2	Actor learning rate η_{actor}	0.001
Number of units per layer	256	Critic learning rate η_{critic}	0.001
Entropy coefficient ε	0.1	Batch size B	128
Discount coefficient γ	0.98	Buffer size M	20,000
Clip coefficient ϵ	0.2		

As shown in Figure 7, the average reward of each CAV during the training process of the MAPPO model is presented. In the initial training phase, the average reward exhibits significant volatility, suggesting that CAVs may encounter issues such as collisions, abrupt maneuvers, or inefficient queuing in the early learning process. As the number of iterations increases, the reward value gradually rises and converges after approximately 350 episodes, indicating the continuous optimization and stabilization of the policy network. Furthermore, to verify the effectiveness of the multi-agent cooperative strategy, a comparative analysis was conducted against the single-agent PPO algorithm under identical parameters and environmental conditions. While the MAPPO model achieves stable convergence after 350 episodes, the PPO algorithm's reward continues to exhibit marked fluctuations during the same period. The persistent instability observed in the PPO results highlights underlying issues with collision avoidance and effective lane selection, whereas MAPPO successfully ensures its stable performance.

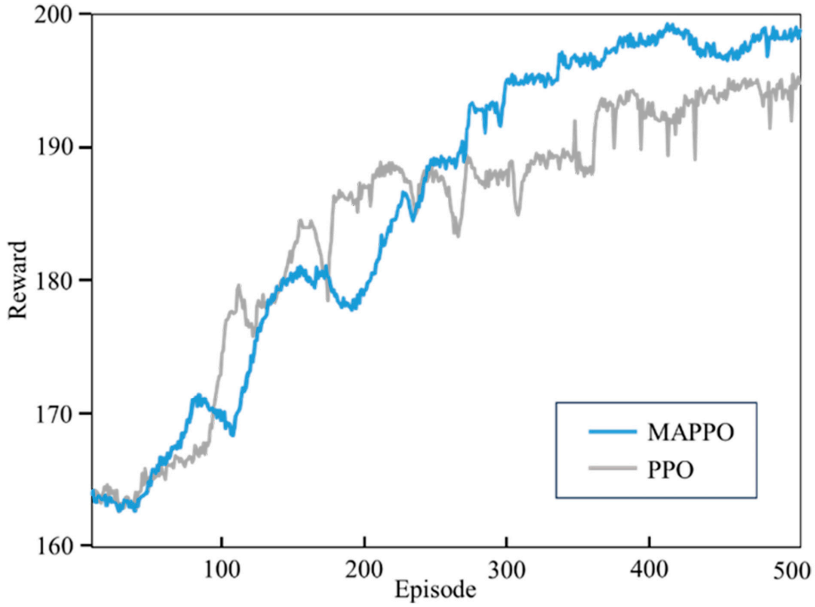


Figure 7. Convergence curve of the average reward during training.

6. Simulation Results and Analysis

6.1. Benchmark Implementation

To examine the superior performance of MARL for cooperative optimization in diverging areas, our study compares traffic performance under three methods: MAPPO, PPO, and a no-control baseline. Each scenario was simulated independently for one hour under identical conditions. To ensure a fair comparison, the PPO model was implemented using the same network framework and core training settings as MAPPO control. In the no-control baseline, HVs still followed the MLP-based toll-lane decision model embedded in the simulation platform without introducing any MARL-based cooperative control strategy. Efficiency is measured by the average vehicle speed, while safety is assessed through the distribution of traffic conflicts.

6.2. Performance Evaluation

Figure 8 presents the average diverging speeds of CAVs, human-driven ETC vehicles (ETC HVs), and MTC vehicles (MTC HVs) under three different scenarios: MAPPO control, PPO control, and the baseline (no control), along with the corresponding variability. The results show that under the control of both the MAPPO and PPO algorithms, the overall diverging speeds improve compared to the no-control baseline. Notably, CAVs under MAPPO control achieve the highest average diverging speed, significantly outperforming the HVs. This demonstrates the effectiveness of this multi-agent strategy in improving efficiency. Furthermore, compared to the PPO algorithm, the MAPPO exhibits a lower standard deviation in speeds, indicating a more stable traffic operation. This stability is attributed to the cooperative mechanism in the CTDE architecture, which enables agents to leverage global state information and thereby mitigate local competition. In contrast, the PPO algorithm, relying solely on local observations for independent decision-making, results in limited speed improvements and larger velocity variance among agents. Additionally, it was observed that in the presence of CAVs, the average speed of human-driven vehicles also slightly increased compared to the no-control scenario. This suggests that the proposed strategy enhances the CAVs' efficiency without compromising the performance of HVs.

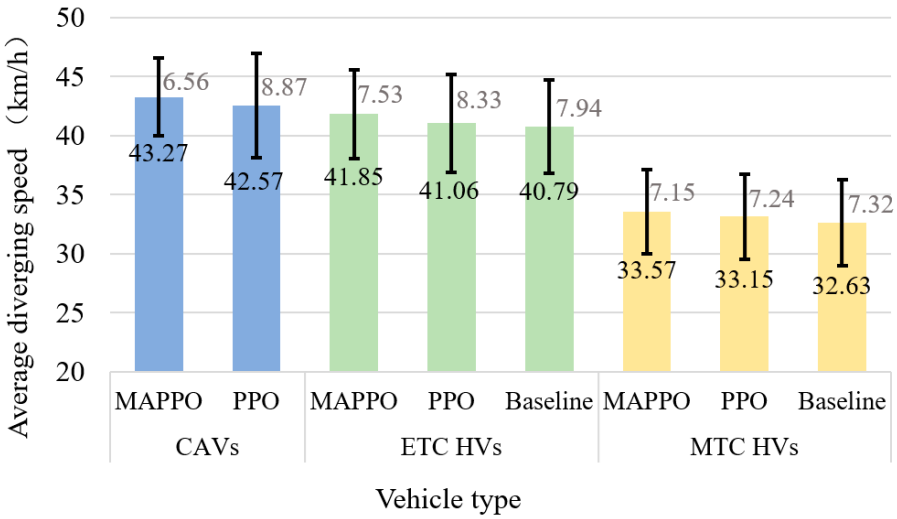


Figure 8. Average diverging speeds of different vehicle types.

Figure 9 compares the distribution of traffic conflicts in the diverging area under the three scenarios. Since conflicts in such areas primarily arise from multi-directional weaving, the Extended Time-to-Collision (ETTC) is adopted as a surrogate safety measure for multi-angle conflicts [15]. Specifically, ETTC is calculated based on vehicle trajectories at 0.1-s simulation time steps. An ETTC value below 2 s is defined as a severe conflict.

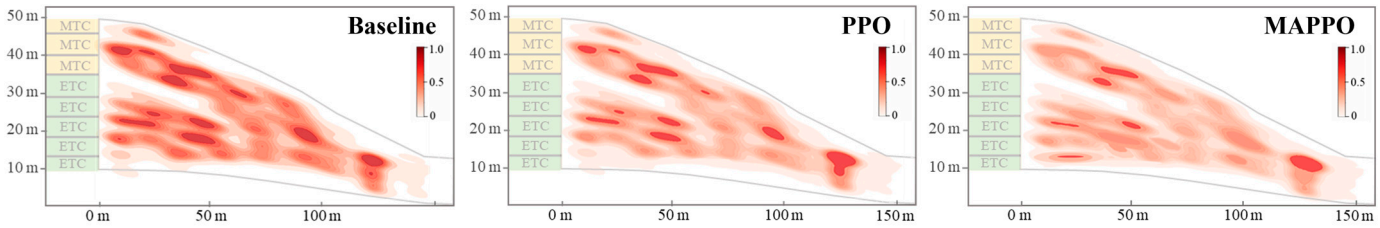


Figure 9. Heat-maps of traffic conflict distribution under different scenarios.

A two-dimensional kernel density estimation is then applied to generate conflict heat-maps, where darker colors indicate higher spatial densities of severe conflicts, as shown in Figure 9. The results show that both PPO and MAPPO significantly reduce the overall density and clustering of severe conflicts, with the conflict hot-spots shifting downstream. MAPPO achieves a greater reduction than PPO, demonstrating the safety advantage of multi-agent cooperative control. However, a local conflict hot-spot persists near the entrance of the MTC lanes. This is likely associated with the stop-and-go behavior of MTC vehicles, which must stop to pay tolls and cannot be fully eliminated.

6.3. Comparative Analysis

To validate the applicability and robustness of the proposed cooperative control strategy under varying traffic demands and geometric conditions, two sets of comparative experiments were designed:

- Traffic volume sensitivity test: The length of the diverging area was fixed at 140 m, while traffic volumes were set to 1500, 1750, and 2000 veh/h.
- Geometric sensitivity test: With the traffic volume fixed at 1500 veh/h, the lengths of the diverging area were set to 120, 140, 160, and 180 m.

In all experiments, other parameters such as the proportion of ETC and MTC vehicles, the CAV penetration rate, and the initial speed distribution remained constant. Each

simulation was independently executed for 1 h. The evaluation metrics include the average diverging speed and the total number of traffic conflicts in the diverging area.

Figure 10 compares the average diverging speeds and their distributions under the MAPPO control and no-control scenarios across different traffic volumes. The results demonstrate that the MAPPO strategy consistently increases the overall speed and narrows the speed distribution at all flow levels. The most significant improvement occurs at 1500 veh/h. As traffic volume increases, the average speed decreases under 1750 and 2000 veh/h, because vehicle interactions become denser and the space available for coordination becomes more limited, which weakens the improvement effect of the proposed strategy. However, under MAPPO, the proportion of low-speed vehicles is still reduced, and the speed distribution remains relatively compact. This indicates that the proposed cooperative control strategy effectively mitigates speed variance among vehicles under high-volume conditions, thereby ensuring traffic flow stability in the diverging area.

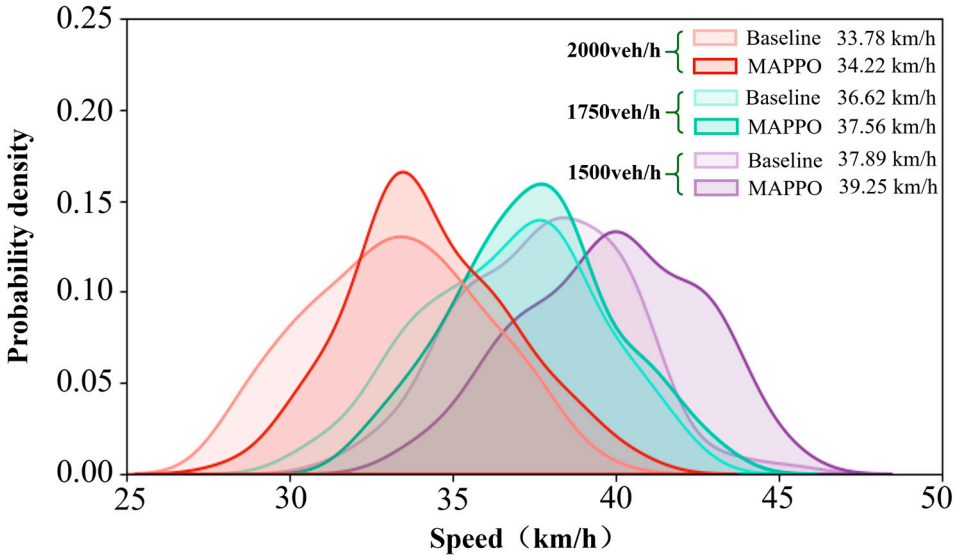


Figure 10. Vehicle speed distributions under different traffic volumes.

Figure 11 presents the impact of diverging area length on the average speed. A slight upward trend in average speed is observed as the length increases, with the MAPPO strategy exhibiting a more pronounced increase. Specifically, at the 120 m length, the difference in average speed between the two scenarios is minimal; however, MAPPO yields a more concentrated distribution. At medium lengths (140 m and 160 m), MAPPO achieves a notable improvement in the average diverging speed. When the length extends to 180 m, although the absolute difference in average speed diminishes, the speed distribution under MAPPO exhibits lower dispersion. This corroborates the advantages of the proposed strategy in smoothing traffic flow fluctuations.

To further evaluate the safety performance of the MAPPO strategy, this study compared the numbers of traffic conflicts with ETTC values in the ranges of $(0, 1]$ and $(1, 2]$ between the MAPPO strategy and the baseline methods across all scenarios, with the results shown in Figure 12.

As shown in Figure 12a, the number of conflicts increases as the traffic volume rises. However, MAPPO consistently yields fewer conflicts than the baseline. Notably, for severe conflicts in the (0,1] s range, MAPPO achieves reductions of 16.3%, 15.9%, and 6.0% at volumes of 1500, 1750, and 2000 veh/h, respectively. This indicates that while the safety improvement effect of MAPPO is reduced under high-volume conditions, it remains evident.

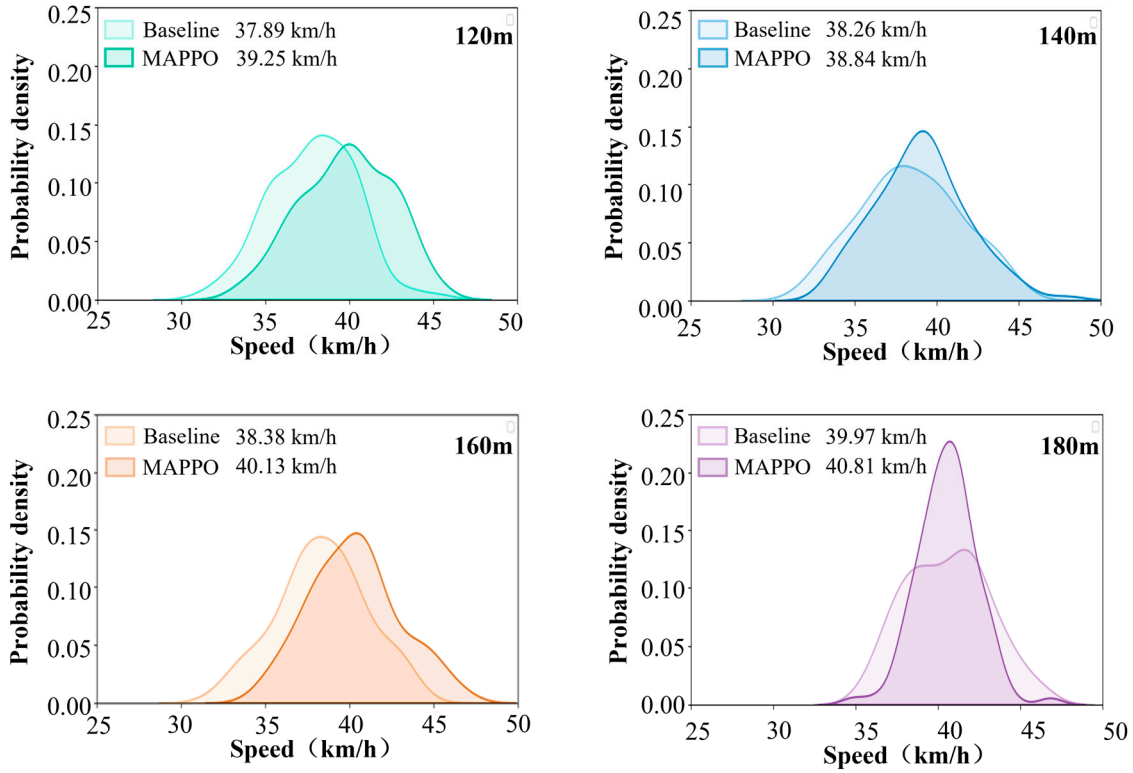


Figure 11. Distribution of average vehicle speeds under different diverging area lengths.

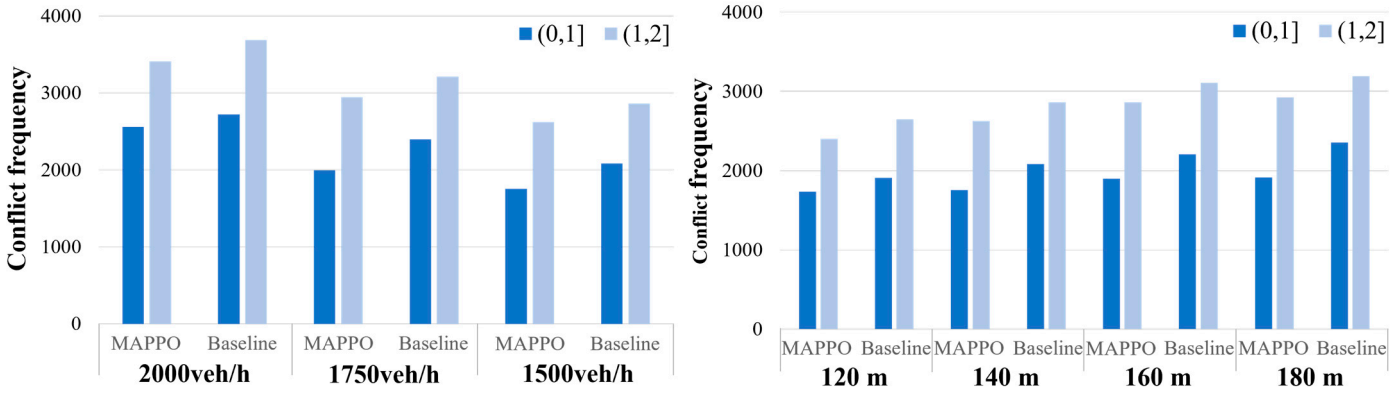


Figure 12. Statistical comparison of traffic conflicts: (**a**) Conflict statistics under different traffic volumes; (**b**) conflict statistics under different lengths.

Figure 12b examines the effect of diverging area length. The total number of conflicts increases slightly as the length increases. In contrast, the relative safety improvement under MAPPO becomes more pronounced. For diverging area lengths of 120, 140, 160, and 180 m, the severe conflicts in the (0,1] s range are reduced by 8.9%, 14.7%, 15.0%, and 18.8%, respectively. These results indicate that additional spatial capacity enhances the effectiveness of cooperative control among CAVs and strengthens the safety benefits of the proposed control strategy.

7. Conclusions

This study aims to develop a cooperative control strategy for CAVs in mixed traffic flow at toll plaza diverging areas to improve both traffic efficiency and operational safety. A two-dimensional microscopic simulation platform under the PDA framework was adopted

as the underlying environment to reproduce the weak-constraint driving behaviors in the diverging area to provide a high-fidelity setting for strategy training and evaluation. Building on this foundation, a MAPPO-based cooperative control method was proposed to enable multi-vehicle coordination.

The results demonstrate clear advantages of the proposed framework in both efficiency and safety. Under MAPPO control, CAVs achieve the highest average diverging speed with minimal fluctuations, indicating smoother and more stable operations. The cooperative mechanism also improves the overall traffic environment, leading to efficiency gains for human-driven vehicles. In addition, conflict analysis based on ETTC shows that the proposed strategy effectively reduces severe conflicts, highlighting the safety benefits of multi-agent coordination in weakly constrained diverging areas. Comparative experiments under different traffic volumes and diverging lengths further confirm the applicability of the proposed strategy across varying traffic demand and geometric conditions. Higher traffic volumes tend to weaken the optimization effect of the cooperative strategy, whereas a longer diverging area can further enhance the coordination effect among vehicles.

This study mainly focuses on the cooperative decision-making of CAVs in weakly constrained toll plaza diverging areas. Future work will further incorporate vehicle dynamics to improve the control performance of the proposed strategy. To enhance generalization, we will examine the influence of different CAV penetration rates on cooperative control under a broader range of scenarios. In addition, more realistic traffic constraints, such as communication limitations, obstacles, and partial lane closures, will be considered to further extend the applicability of the framework and assess its potential for real-world deployment.

Author Contributions: Methodology, Y.F. and L.Z.; Conceptualization, Y.F.; writing—original draft preparation, S.L.; writing—review and editing, S.L. and L.Z.; funding acquisition, L.Z.; Software, Y.F. and S.L.; Validation, Y.F.; Supervision, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAVs	Connected and autonomous vehicles
MAPPO	Multi-agent proximal policy optimization
CTDE	Centralized training and decentralized execution
MARL	Multi-agent reinforcement learning
PDA	Perception-Decision-Action
ETC	Electronic toll collection
MTC	Manual toll collection
FVD	Full velocity difference
MADDPG	Multi-agent deep deterministic policy gradient
QMIX	Monotonic mixing network
PPO	Proximal policy optimization
GAE	Generalized advantage estimation

TD	Temporal difference
UAV	Unmanned aerial vehicle
MLP	Multilayer perceptron
HVs	Human-driven vehicles
ETTC	Extended time-to-collision

References

1. Talebpour, A.; Mahmassani, H.S. Influence of Connected and Autonomous Vehicles on Traffic Flow Stability and Throughput. *Transp. Res. Part C Emerg. Technol.* **2016**, *71*, 143–163. [CrossRef]
2. Rahman, M.M.; Thill, J.-C. Impacts of Connected and Autonomous Vehicles on Urban Transportation and Environment: A Comprehensive Review. *Sustain. Cities Soc.* **2023**, *96*, 104649. [CrossRef]
3. Liu, W.; Hua, M.; Deng, Z.; Meng, Z.; Huang, Y.; Hu, C.; Song, S.; Gao, L.; Liu, C.; Shuai, B.; et al. A Systematic Survey of Control Techniques and Applications in Connected and Automated Vehicles. *IEEE Internet Things J.* **2023**, *10*, 21892–21916. [CrossRef]
4. Abdelwahab, H.T.; Abdel-Aty, M.A. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. *Transp. Res. Rec. J. Transp. Res. Board* **2002**, *1784*, 115–125. [CrossRef]
5. Saad, M.; Abdel-Aty, M.; Lee, J. Analysis of Driving Behavior at Expressway Toll Plazas. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *61*, 163–177. [CrossRef]
6. Fei, Y.; Long, K.; Xing, L.; Pei, X.; Li, X.; Yao, L. Safety Performance Analysis of Toll Plaza Diverging Area Based on an Improved Simulation Platform for Weak-Constraint Driving Behaviors. *Accid. Anal. Prev.* **2025**, *220*, 108177. [CrossRef] [PubMed]
7. Shladover, S.E.; Nowakowski, C.; Lu, X.-Y.; Ferlis, R. Cooperative Adaptive Cruise Control: Definitions and Operating Concepts. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2489*, 145–152. [CrossRef]
8. Lukose, E.; Levin, M.W.; Boyles, S.D. Incorporating Insights from Signal Optimization into Reservation-Based Intersection Controls. *J. Intell. Transp. Syst.* **2019**, *23*, 250–264. [CrossRef]
9. Kamal, M.A.S.; Imura, J.; Hayakawa, T.; Ohata, A.; Aihara, K. A Vehicle-Intersection Coordination Scheme for Smooth Flows of Traffic Without Using Traffic Lights. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1136–1147. [CrossRef]
10. Wu, Y.; Chen, H.; Zhu, F. DCL-AIM: Decentralized Coordination Learning of Autonomous Intersection Management for Connected and Automated Vehicles. *Transp. Res. Part C Emerg. Technol.* **2019**, *103*, 246–260. [CrossRef]
11. Boukerche, A.; Zhong, D.; Sun, P. A Novel Reinforcement Learning-Based Cooperative Traffic Signal System Through Max-Pressure Control. *IEEE Trans. Veh. Technol.* **2022**, *71*, 1187–1198. [CrossRef]
12. Zhou, M.; Yu, Y.; Qu, X. Development of an Efficient Driving Strategy for Connected and Automated Vehicles at Signalized Intersections: A Reinforcement Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 433–443. [CrossRef]
13. Zhang, J.; Chang, C.; Zeng, X.; Li, L. Multi-Agent DRL-Based Lane Change with Right-of-Way Collaboration Awareness. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 854–869. [CrossRef]
14. Mirheli, A.; Tajalli, M.; Hajibabai, L.; Hajbabaie, A. A Consensus-Based Distributed Trajectory Control in a Signal-Free Intersection. *Transp. Res. Part C Emerg. Technol.* **2019**, *100*, 161–176. [CrossRef]
15. Xing, L.; He, J.; Abdel-Aty, M.; Cai, Q.; Li, Y.; Zheng, O. Examining Traffic Conflicts of Upstream Toll Plaza Area Using Vehicles' Trajectory Data. *Accid. Anal. Prev.* **2019**, *125*, 174–187. [CrossRef] [PubMed]
16. Xing, L.; He, J.; Li, Y.; Wu, Y.; Yuan, J.; Gu, X. Comparison of Different Models for Evaluating Vehicle Collision Risks at Upstream Diverging Area of Toll Plaza. *Accid. Anal. Prev.* **2020**, *135*, 105343. [CrossRef] [PubMed]
17. Aoki, S.; Higuchi, T.; Altintas, O. Cooperative Perception with Deep Reinforcement Learning for Connected Vehicles. In *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*; IEEE: New York, NY, USA, 2020; pp. 328–334.
18. Waga, A.; Benhlima, S.; Bekri, A.; Abdouni, J.; Saber, F.Z. A Survey on Autonomous Navigation for Mobile Robots: From Traditional Techniques to Deep Learning and Large Language Models. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *37*, 198. [CrossRef]
19. Gregurić, M.; Kušić, K.; Ivanjko, E. Impact of Deep Reinforcement Learning on Variable Speed Limit Strategies in Connected Vehicles Environments. *Eng. Appl. Artif. Intell.* **2022**, *112*, 104850. [CrossRef]
20. Jin, J.; Huang, H.; Li, Y.; Dong, Y.; Zhang, G.; Chen, J. Variable Speed Limit Control Strategy for Freeway Tunnels Based on a Multi-Objective Deep Reinforcement Learning Framework with Safety Perception. *Expert Syst. Appl.* **2025**, *267*, 126277. [CrossRef]
21. Li, G.; Qiu, Y.; Yang, Y.; Li, Z.; Li, S.; Chu, W.; Green, P.; Li, S.E. Lane Change Strategies for Autonomous Vehicles: A Deep Reinforcement Learning Approach Based on Transformer. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2197–2211. [CrossRef]
22. Zhang, S.; Zhuang, W.; Li, B.; Li, K.; Xia, T.; Hu, B. Integration of Planning and Deep Reinforcement Learning in Speed and Lane Change Decision-Making for Highway Autonomous Driving. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 521–535. [CrossRef]
23. Fei, Y.; Xing, L.; Yao, L.; Yang, Z.; Zhang, Y. Deep Reinforcement Learning for Decision Making of Autonomous Vehicle in Non-Lane-Based Traffic Environments. *PLoS ONE* **2025**, *20*, e0320578. [CrossRef]

24. Zhang, J.; Zhang, Y.; Zhang, X.S.; Zang, Y.; Cheng, J. Intrinsic Action Tendency Consistency for Cooperative Multi-Agent Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2024; Volume 38, pp. 17600–17608. [CrossRef]
25. Xing, L.; Zou, D.; Fei, Y.; Long, K.; Wang, J. Safety Evaluation of Toll Plaza Diverging Area Considering Different Vehicles' Toll Collection Types. *Appl. Sci.* **2023**, *13*, 9005. [CrossRef]
26. Bai, R.; Xu, R.; Rui, T.; Liu, J.; Lee, H.L.; Oung, Q.W.; Tian, Z.; Yuan, F. Safe and Efficient Lane-Changing for Autonomous Vehicles: An Improved Double Quintic Polynomial Approach with Time-to-Collision Evaluation. *J. King Saud Univ. Comput. Inf. Sci.* **2026**, *38*, 36. [CrossRef]
27. Li, Y.; Li, L.; Ni, D. Dynamic Trajectory Planning for Automated Lane Changing Using the Quintic Polynomial Curve. *J. Adv. Transp.* **2023**, *2023*, 6926304. [CrossRef]
28. Kumar, P.; Perrollaz, M.; Lefevre, S.; Laugier, C. Learning-Based Approach for Online Lane Change Intention Prediction. In *Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV)*; IEEE: New York, NY, USA, 2013; pp. 797–802.
29. Shi, Q.; Zhang, H. An Improved Learning-Based LSTM Approach for Lane Change Intention Prediction Subject to Imbalanced Data. *Transp. Res. Part C Emerg. Technol.* **2021**, *133*, 103414. [CrossRef]
30. Peng, J.; Guo, Y.; Fu, R.; Yuan, W.; Wang, C. Multi-Parameter Prediction of Drivers' Lane-Changing Behaviour with Neural Network Model. *Appl. Ergon.* **2015**, *50*, 207–217. [CrossRef]
31. Song, X.-M.; Jin, S.; Wang, D.-H.; Cao, J.-H. Vehicle-Following Model Considering Lateral Offset. *J. Jilin Univ. (Eng. Technol. Ed.)* **2011**, *41*, 333–337.
32. Qi, W.; Ma, S.; Fu, C. An Improved Car-Following Model Considering the Influence of Multiple Preceding Vehicles in the Same and Two Adjacent Lanes. *Phys. A Stat. Mech. Its Appl.* **2023**, *632*, 129356. [CrossRef]
33. Helbing, D.; Tilch, B. Generalized Force Model of Traffic Dynamics. *Phys. Rev. E* **1998**, *58*, 133–138. [CrossRef]
34. Hoel, C.-J.; Wolff, K.; Laine, L. Automated Speed and Lane Change Decision Making Using Deep Reinforcement Learning. In *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*; IEEE: New York, NY, USA, 2018; pp. 2148–2155.
35. Zheng, O.; Abdel-Aty, M.; Wu, Y. UCF-SST Automated Roadway Conflicts Identify System (ARCIS). Available online: https://github.com/fatemehjdi/A-R-C-I-S (accessed on 15 March 2026).

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.