

Infrastructure-Assisted Cooperative Decision Model With Priority Awareness at Unsignalized Intersections

Sifan Wu^{1,2,3}, Xuting Duan^{1,2,3}✉, Hao Zhang^{1,2,3}, Feiyang Zhao^{1,2,3}, Jianshan Zhou^{1,2,3}, Kaige Qu^{1,2,3}, Ling Wang⁴,
Daxin Tian^{1,2,3}✉

Cite this article: <https://doi.org/10.26599/COMMTR.2026.9640054>

ABSTRACT: Unsignalized intersections pose a significant challenge for multi-vehicle cooperative decision-making, where safety and efficiency must be ensured simultaneously under dynamic traffic conditions. To address this challenge, this study proposes a Vehicle-to-Infrastructure Integrated Distributed Agent Decision-Making (V2I-IDADM) framework, which leverages vehicle–infrastructure cooperation to enhance both coordination performance and safety at unsignalized intersections. Built upon a learning paradigm with centralized training and decentralized execution, the proposed V2I-IDADM framework utilizes the global perception of roadside infrastructure to assign passing priorities to connected and autonomous vehicles (CAVs), ensuring safe passage through intersections. Meanwhile, the framework’s global coordination mechanism constructs a unified and structured representation of intersection-level traffic states, enabling scalable, real-time decision-making. Specifically, a priority-based safety decision model is developed by jointly integrating passing-priority constraints with action optimization to promote multi-vehicle cooperation. To enhance training efficiency, a hierarchical weighted sampling strategy is introduced to emphasize high-value episodic experiences and accelerate iterative self-learning. Extensive experiments in pure CAV and mixed-traffic scenarios demonstrate that the proposed framework achieves superior safety and efficiency compared with state-of-the-art methods. Experiments conducted on both a miniature intelligent vehicle platform and a full-scale vehicle further validate the practical feasibility and deployment potential of the proposed framework.

KEYWORDS: Unsignalized intersection, cooperative decision-making, vehicle–infrastructure cooperation, connected and autonomous vehicles (CAVs), iterative self-learning.

1 Introduction

In recent years, autonomous vehicles have demonstrated significant potential in enhancing road safety, reducing traffic accidents, and improving overall traffic efficiency (Huang et al., 2024). With the rapid advancement of intelligent transportation systems (ITS), breakthroughs in autonomous driving technologies have fundamentally transformed traffic control paradigms, data environments, and control methodologies (Liu et al., 2025b). Autonomous vehicles are becoming increasingly capable of handling complex traffic scenarios, playing a critical role in addressing traditional road safety challenges, optimizing traffic flow, and improving passenger efficiency (Makridis et al., 2025). The rapid development of Connected and Autonomous Vehicles (CAVs) and Vehicle-to-Infrastructure (V2I) technologies is accelerating the transformation of traditional urban transportation systems toward greater intelligence and coordination (Yang et al., 2026). By establishing efficient information exchange among vehicles, infrastructure, and the surrounding environment, these technologies provide critical enhancements in perception and decision support for high-level automated driving, effectively mitigating the inherent limitations of vehicle-centric autonomy in global environment perception, scene prediction, and collaborative decision-making (Liu et al., 2024). In particular, roadside units (RSUs), equipped with wide-area sensing coverage, substantial computational resources, and advanced information fusion capabilities, enable autonomous vehicles to overcome the constraints of limited onboard sensing range and unavoidable blind

spots. Through infrastructure-assisted perception, CAVs can obtain more comprehensive and high-precision traffic and environmental information in real time, thereby significantly supporting safer and more reliable cooperative decision-making.

Under this cooperative architecture, CAVs no longer rely solely on onboard sensors for isolated decision-making. Instead, they leverage V2I communication to share critical information with RSUs, such as real-time traffic dynamics, potential conflicts, vehicle trajectory and vehicle priorities, enabling safer, more informed decision-making (George et al., 2025). This advantage is particularly pronounced at unsignalized intersections, which require strong vehicle interaction (el Abidine Kherroubi et al., 2021). By operating at a global scale, roadside infrastructure aggregates and processes traffic data, forming a real-time decision model that identifies potential conflicts and assigns passing priorities to ensure safe intersection passage. As a result, the infrastructure-assisted cooperative decision-making model enhances safety and traffic efficiency, while supporting reasoning for large-scale, data-driven decision-making and learning.

1.1. Related works

Current decision-making approaches for multi-vehicle autonomous driving at unsignalized intersections can be broadly categorized into rule-based, optimization-based, game-theoretic, and learning-based methods. Among these, rule-driven strategies have been extensively studied due to their conceptual simplicity and ease of implementation. Dresner and Stone (2008) were among the first to propose a reservation-based motion planning framework,

¹ State Key Laboratory of Intelligent Transportation Systems, Beihang University, Beijing 100191, China. ² Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beihang University, Beijing 102206, China. ³ School of Transportation Science and Engineering, Beihang University, Beijing 102206, China. ⁴ The Department of Automation, Tsinghua University, Beijing 100084, China.

✉ Corresponding author. E-mail: X. Duan, duanxuting@buaa.edu.cn; D. Tian, dtian@buaa.edu.cn

Received: January 9, 2026; Revised: May 9, 2026; Accepted: September 22, 2026

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>).

which optimizes intersection crossing order according to a First-Come-First-Served (FCFS) principle, thereby improving traffic throughput and reducing conflicts. Building upon this framework, Huang et al. (2012) extended the reservation mechanism by introducing a dynamic hierarchical reservation protocol that jointly coordinates vehicle priorities and velocities, further enhancing overall intersection efficiency. In addition, Lin et al. (2017) proposed a buffer-allocation-based vehicular coordination control method, which employs a three-phase velocity profile to optimize vehicle trajectories at intersections and refines passing rules to improve practical applicability and robustness. Despite the advantages of rule-based approaches in terms of design simplicity and interpretability, they typically rely on fixed logical structures and greedy decision mechanisms, which limit their ability to achieve global optimality under high traffic demand. In dense traffic scenarios, such methods may lead to inefficient utilization of intersection space and even exacerbate congestion risks, thereby constraining their scalability and performance in complex real-world environments.

Within optimization-based approaches, Zohdy and Rakha (2016) developed a nonlinear, constrained trajectory optimization model that minimizes traffic delay and incorporates a car-following formulation to ensure collision avoidance. Pan et al. (2022) introduced a convex optimal control framework for unsignalized intersections, decomposing the multi-vehicle coordination problem into passing-order planning and trajectory optimization, thereby reformulating it into two efficiently solvable second-order cone programs. Yao et al. (2023) proposed a two-stage optimization scheme in which a mixed-integer scheduler first allocates conflict-free arrival times, followed by a linear program that computes fuel-efficient trajectories consistent with the assigned schedule. Pei et al. (2021) further modeled the coordination task as a dynamic-programming-based optimization problem, embedding all feasible passing sequences into a polynomial-size state space to enable efficient search for a globally optimal conflict-free order. Although optimization-based methods provide strong theoretical guarantees and can capture complex interaction structures, their computational burden remains substantial, which may limit real-time applicability in dense or highly dynamic traffic environments.

Game-theoretic decision-making approaches offer a structured modeling paradigm for multi-vehicle coordination at

unsignalized intersections, explicitly characterizing strategic interactions among traffic participants (Jeong, 2023). Li et al. (2020) proposed a game-theoretic framework that incorporates traffic rules and leader–follower relationships, and further integrates yielding behaviors and limited perception ranges into the decision structure to enhance behavioral rationality and interpretability. Chandra and Manocha (2022) developed a multi-agent game-based planning framework grounded in driving style modeling, in which aggressive, normal, and conservative driver behaviors are explicitly represented. Extensive experiments in realistic scenarios demonstrated the effectiveness and stability of the resulting strategies. Building upon these efforts, Yuan et al. (2023) proposed a scalable game-theoretic decision framework that combines receding-horizon optimization with level- k reasoning. By constructing a dynamically updated switching directed graph that retains only interaction-relevant vehicles, their approach significantly reduces computational complexity while maintaining safety guarantees and improving real-time performance. Similarly, Lu et al. (2023) formulated multi-vehicle interactions as a dynamic non-cooperative game, where the solution space is reduced through participant selection and game-mode switching. Their reward design explicitly accounts for safety, efficiency, and heterogeneous driving styles, yielding decision policies that better reflect real-world traffic behaviors. Despite the theoretical advantages of game-theoretic frameworks in modeling multi-vehicle interactions, such methods typically rely on strong assumptions regarding agent rationality, strategy observability, and consistency of driving preferences, which limit their applicability in realistic traffic environments. Moreover, as the number of interacting vehicles increases, the strategy space grows exponentially, leading to rapidly escalating computational complexity and consequently constraining real-time performance and practical deployability in complex traffic scenarios.

Among learning-based decision-making approaches for unsignalized intersections, Deep Reinforcement Learning (DRL) has emerged as a prominent paradigm due to its ability to autonomously extract decision policies from high-dimensional sensory inputs, adapt to dynamic environments, and handle highly nonlinear decision boundaries (Al-Sharman et al., 2023; Huang et al., 2024). Isele et al. (2018) were among the first to apply single-agent reinforcement learning to occluded unsignalized intersection scenarios, demonstrating the effectiveness of RL in addressing

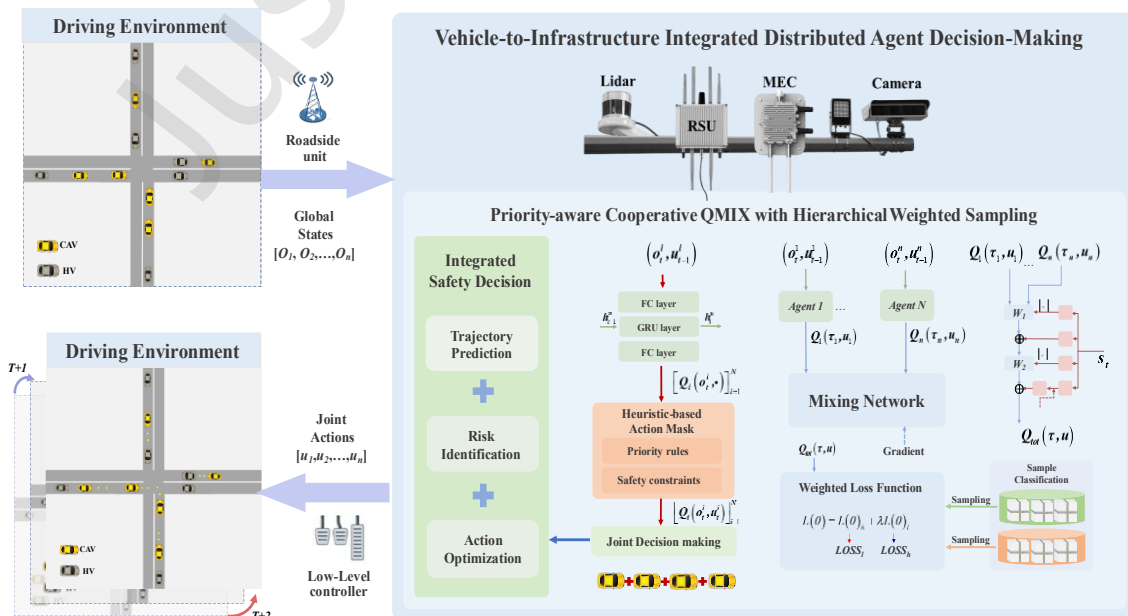


Fig. 1. Overall framework for the proposed model.

occlusion and partial observability challenges. Subsequently, Li et al. (2024) proposed an attention-based reinforcement learning framework that explicitly models interaction dependencies between the ego vehicle and surrounding vehicles, enabling the model to capture the importance of critical traffic elements. In addition, they designed a safety-aware reward function that jointly considers task efficiency and potential risks, achieving a dynamic balance between safety and performance during decision optimization. Building upon this work, Xiao et al. (2024) developed a hybrid attention mechanism tailored for unsignalized intersections. They introduced a state representation suitable for multi-task driving scenarios, allowing the model to maintain stable performance across left-turning, right-turning, and straight-going maneuvers. Despite the strong learning capability and generalization potential demonstrated by single-agent DRL in these studies, significant challenges remain when extending such approaches to multi-agent traffic environments. As the number of interacting vehicles increases, the state space grows exponentially, making it difficult for single-agent methods to capture global traffic structures and interaction patterns. Consequently, deploying single-agent DRL in complex, multi-vehicle traffic scenarios is inherently constrained by dimensionality explosion and limited scalability, which has further motivated the adoption of multi-agent reinforcement learning (MARL) for cooperative driving and unsignalized intersection decision-making.

Consequently, MARL has gradually emerged as a focal research direction. By adopting a centralized training and decentralized execution (CTDE) paradigm, MARL enhances global optimization capability while preserving the autonomy of individual agents (Chen et al., 2024a; Jadoon et al., 2023). In the context of intersection decision-making, MARL methods can effectively handle complex traffic flows and leverage inter-agent cooperation to optimize vehicle-level driving strategies, thereby improving overall traffic efficiency and safety. Huang et al. (2023) improved the initial performance and exploration direction of MARL algorithms by incorporating expert demonstrations and interaction data into the training process. Zhuang et al. (2023) proposed a MARL framework that employs an attention module to extract the most informative features and introduces a weighted reward allocation scheme, assigning greater importance to vehicles approaching the intersection. Chen et al. (2024b) developed a MARL approach based on heterogeneous vehicle attention mechanisms to address cooperative decision-making at intersections. Their method employs a sequential update strategy to enhance adaptability to heterogeneous vehicles. In addition, Liu et al. (2025a) proposed a cooperative decision-making framework for CAVs at unsignalized intersections, in which attention mechanisms are used to model inter-vehicle interactions, select relevant interaction partners, and construct game-theoretic relationships. At the same time, a safety detector is employed to generate admissible actions. Despite these advances, existing MARL approaches still face notable limitations when applied to cooperative decision-making.

1.2. Motivation and Contributions

To address the aforementioned challenges, this study proposes a Vehicle-to-Infrastructure Integrated Distributed Agent Decision-Making (V2I-IDADM) framework for cooperative decision-making at unsignalized intersections, as illustrated in Fig. 1. The framework uses roadside infrastructure for decision support, enabling model-driven cooperative reasoning and large-scale learning for real-time coordination among connected and autonomous vehicles (CAVs). By shifting decision-making from vehicles to infrastructure, it creates a scalable, safety-critical model that optimizes multi-vehicle coordination at intersections. The proposed framework embeds a priority-aware cooperative safety model that integrates heuristic priority rules with explicit safety

constraints. Furthermore, an integrated safety-oriented decision optimization strategy is incorporated into the model to jointly prune and optimize the action space, thereby enhancing decision safety while improving model interpretability and controllability. In addition, a hierarchical weighted sampling strategy is developed to emphasize episodes that contribute more significantly to policy learning, thereby improving training efficiency and focusing the model on high-value experience regions. The main contributions of this study are summarized as follows:

Vehicle-to-Infrastructure Integrated Distributed Agent Decision-Making Framework: This study proposes a Vehicle-to-Infrastructure Integrated Distributed Agent Decision-Making (V2I-IDADM) framework for unsignalized intersections, in which roadside infrastructure plays a central role in aggregating global traffic information and coordinating multi-vehicle behaviors. By shifting global coordination from individual vehicles to the infrastructure, the proposed framework establishes a scalable and safety-critical cooperative decision-making paradigm. This infrastructure-centric organization provides a deployable framework for model-driven cooperative reasoning and large-scale learning at intersections. Building on this foundation, a priority-aware cooperative safety model, informed by globally aggregated traffic data, is embedded in the framework to assign passing priorities and crossing orders to CAVs at intersections. Furthermore, an integrated safety-oriented decision optimization strategy is incorporated into the model to jointly refine and optimize global actions, thereby ensuring safety while improving overall traffic efficiency.

Sampling strategy: Furthermore, a hierarchical weighted sampling strategy is proposed. By categorizing episodic samples according to their cumulative rewards and applying a weighted loss function during training, the proposed strategy emphasizes informative high-reward experience samples while maintaining sufficient exploration, thereby significantly improving self-learning performance.

Simulation and Real-World Validation: Extensive experiments are conducted under both pure CAV and mixed-traffic environments involving human-driven vehicles (HVs), demonstrating the robustness and effectiveness of the proposed framework. Furthermore, the miniature intelligent vehicle and the full-scale vehicle experiments are developed to validate the practical feasibility and real-world deployment potential of the proposed V2I-IDADM framework.

The remainder of this study is organized as follows. Section 2 formulates the multi-vehicle cooperative decision-making problem at unsignalized intersections. Section 3 presents the overall framework. Section 4 discusses the experimental results and validates the proposed framework in a miniature intelligent vehicle, and full-scale vehicle experiments. Finally, Section 5 concludes the study and discusses directions for future work.

2 Problem Description

2.1. Scenario Description

As shown in Fig. 2, CAVs approach the intersection from four different directions, aiming to achieve safe, efficient, and collaborative passage. To enhance the complexity of the environment, the study randomly introduces HVs into the traffic flow, thereby constructing a mixed traffic flow scenario. Within the proposed V2I-IDADM framework, each CAV is equipped with onboard sensing and communication capabilities to acquire local state information and transmit it to the roadside unit (RSU) via V2I wireless communication (Xu et al., 2022). The roadside infrastructure, consisting of RSUs and sensors, continuously collects and aggregates multi-vehicle traffic information within the intersection area. A mobile edge computing (MEC) server deployed at the roadside performs centralized information fusion

and cooperative decision support during the training phase, enabling efficient coordination among multiple agents.

The primary objective of the V2I-IDADM framework is to optimize cooperative decision-making among CAVs in a complex unsignalized intersection scenario. During policy execution, the roadside MEC server integrates the global traffic information collected from RSUs and roadside sensors and dispatches real-time decision commands to individual CAVs. Leveraging this aggregated global perception, CAVs are guided to perform cooperative scheduling and decision optimization, ensuring collision avoidance and traffic efficiency. The received decision commands are subsequently transformed by onboard controllers into corresponding control actions, thereby realizing coordinated vehicle behavior at the intersection. Accordingly, the proposed scenario is designed to comprehensively evaluate the effectiveness and robustness of the proposed model under the V2I-IDADM framework, demonstrating its capability to enable safe and efficient intersection passage.

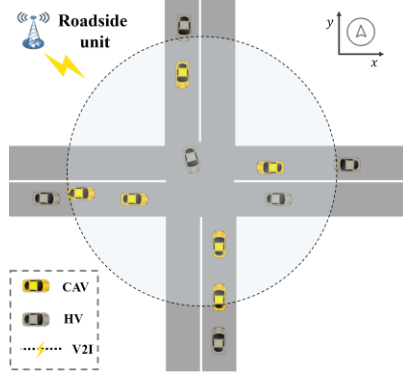


Fig. 2. Scenario description of the unsignalized intersection.

2.2. Vehicle Model

In this study, a bicycle model is employed to represent the vehicle dynamics (Kong et al., 2015), as illustrated in Fig. 3. The bicycle model is commonly used in vehicle dynamics simulations due to its simplicity and ability to effectively capture the essential characteristics of vehicle motion.

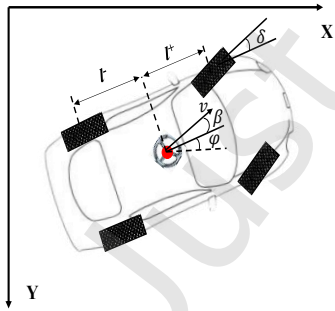


Fig. 3. Bicycle model.

By modeling the bicycle model, the vehicle kinematic model is described as follows:

$$\begin{cases} \dot{x} = v \cos(\varphi + \beta), \\ \dot{y} = v \sin(\varphi + \beta), \\ \dot{v} = a, \\ \dot{\varphi} = \frac{v}{l_r} \tan \delta \cos \beta, \end{cases} \quad (1)$$

where x and y represent the lateral and longitudinal positions of the vehicle, respectively. v represents the velocity, and a represents the acceleration. φ represents the heading angle of the vehicle, and l_r is the distance from the center of mass to the rear axle. β represents the sideslip angle. The relationship between the sideslip

angle β and the front wheel angle δ is given by the following:

$$\beta = \tan^{-1} \left(\frac{l_r}{l_r + l_f} \tan \delta \right), \quad (2)$$

where l_f represents the distance from the vehicle's center of mass to the front axle.

At the vehicle control layer, the lower-level controller is responsible for tracking the desired velocity and steering angle. For longitudinal control, a proportional controller is used to track the vehicle's velocity.

$$a = K_a (v_r - v), \quad (3)$$

where v_r represents the desired velocity of the vehicle, and k_a represents the acceleration proportional control parameter.

$$v_x = -K_x (x_r - x), \quad (4)$$

$$\Delta \varphi_r = \arcsin \left(\frac{v_x}{v} \right), \quad (5)$$

where v_x represents the lateral velocity of the vehicle, x_r represents the target lateral position of the vehicle, K_x represents the proportional control parameter for lateral velocity, $\Delta \varphi_r$ represents the change in heading angle corresponding to the applied lateral velocity command.

Additionally, the lateral angle control of the vehicle is described by the following equation:

$$\varphi_r = \varphi_L + \Delta \varphi_r, \quad (6)$$

$$\dot{\varphi}_r = K_\varphi (\varphi_r - \varphi), \quad (7)$$

$$\delta = \arcsin \left(\frac{1}{2} \frac{l_f}{v} \dot{\varphi}_r \right), \quad (8)$$

where φ_r represents the desired heading angle, used to follow the lane's direction and position. $\dot{\varphi}_r$ represents the desired yaw rate, and K_φ is the proportional parameter for the yaw rate.

2.3. Distributed Partially Observable Markov Decision Process (POMDP)

The cooperative decision-making problem for multiple CAVs at unsignalized intersections is formulated as a V2I-enabled Decentralized Partially Observable Markov Decision Process (Dec-POMDP) (Pey et al., 2025). In this environment, each CAV has inherently limited onboard perception; therefore, it cannot fully observe the global traffic state. To alleviate this limitation, the proposed V2I-IDADM framework exploits roadside sensing and V2I communication: RSUs collect multi-vehicle traffic information, and a MEC server performs real-time information fusion and decision support. Under this paradigm, each CAV receives an infrastructure-assisted observation constructed from the fused global information, while continuing to execute its local actions. The V2I-enabled cooperative decision-making problem is modeled as a Dec-POMDP, defined by the tuple $(I, S, \{O_i\}_{i \in I}, \{U_i\}_{i \in I}, P, r, \gamma)$, where I represents the number of controlled vehicles CAV, S represents the global state space of the intersection maintained by the RSU/MEC through multi-source sensing and fusion, O_i and U_i represent the local observation space and action space for each CAV_{*i*}. At each decision step, each controlled CAV ($i \in I$) executes an action $u_i \in U_i$, forming a joint action $\mathbf{u} = (u_1, u_2, \dots, u_{|I|})$. The environment evolves according to the state transition function $P(s' | s, \mathbf{u})$, which governs the

transition from state s to state s' . In this problem, all CAVs share a reward function $r: S \times U \rightarrow R$, where $U = U_1 \times \dots \times U_{|I|}$, and $\gamma \in [0, 1]$ is the discount factor.

The objective is to learn a cooperative policy that maximizes the expected discounted reward:

$$\max_{\pi} E \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \mathbf{u}_t) \right]. \quad (9)$$

3 Proposed Framework

3.1. QMIX Model

In this framework, QMIX (Rashid, 2018) is adopted as the decision-making algorithm for the MARL framework. QMIX follows the CTDE paradigm of Value Decomposition Network (VDN) (Sunehag et al., 2017), effectively overcoming the difficulty of handling cooperative behaviors in multi-agent systems compared to conventional independent learning methods. For each agent i , given its local observation o_i and action u_i , a local Q-value $Q_i(\tau_i, u_i)$ is first computed. QMIX then employs a nonlinear mixing network to combine all local Q-values into a global Q-value:

$$Q_{tot}(\tau, \mathbf{u}; \theta) = f_{mix}(Q_1(\tau_1, u_1), \dots, Q_n(\tau_n, u_n), s; \theta_{mix}), \quad (10)$$

where τ represents the joint action-observation history of CAVs, f_{mix} is the mixing network parameterized by θ_{mix} . To guarantee that the globally optimal joint action can be obtained by independently maximizing each agent's local Q-value, QMIX imposes a monotonicity constraint on the mixing network:

$$\frac{\partial Q_{tot}(\tau, \mathbf{u})}{\partial Q_i(\tau_i, u_i)} \geq 0, i = 1, \dots, |I|. \quad (11)$$

This constraint preserves monotonicity between local and global Q-values, enabling coordinated optimization toward a shared objective.

Each agent first evaluates its own local Q-value through its individual network, after which all local Q-values are fed into the mixing network to produce the global Q-value $Q_{tot}(\tau, \mathbf{u})$. This architecture enables centralized learning with access to global information during training, while still supporting decentralized execution based solely on local observations at runtime, thus allowing for flexible cooperative decision-making in complex multi-agent environments.

The loss function of QMIX is defined as:

$$L(\theta) = \frac{1}{b} \sum_{m=1}^b \left[\left(y_m^{tot} - Q_{tot}(\tau, \mathbf{u}; \theta) \right)^2 \right], \quad (12)$$

where b represents the mini-batch size, y_m^{tot} is the target Q-value, typically given by:

$$y^{tot} = r + \gamma \max_{\mathbf{u}'} Q_{tot}'(\tau', \mathbf{u}'; \theta^-), \quad (13)$$

where r represents the immediate reward, γ represents the discount factor, τ' represents the joint action-observation history at the next time step, $\mathbf{u}'$ represents the corresponding joint action, s' represents the global state at the next time step, and Q_{tot}' represents the target network parameterized by θ^- .

1) State space: Under the proposed V2I-IDADM framework, the state space is constructed based on globally fused traffic information collected by roadside infrastructure and processed by the MEC server. Specifically, the state space for each CAV includes not only the ego-vehicle's own kinematic states but also the global states of surrounding vehicles within the intersection, as obtained through roadside sensing and V2I communication. The state space encompasses the ego vehicle's absolute position, velocity, and heading angle, as well as the absolute positions, velocities, and heading angles of surrounding vehicles detected by

roadside sensors. Unlike purely vehicle-centric perception, this infrastructure-assisted state representation provides each CAV with a consistent and unified global coordinate frame, thereby enabling coherent state interpretation.

$$O_i = \begin{bmatrix} \text{presence}, & x_i, & y_i, & v_i^x, & v_i^y, & \phi_i \\ \text{presence}, & x_{\text{nearest}1}, & y_{\text{nearest}1}, & v_{\text{nearest}1}^x, & v_{\text{nearest}1}^y, & \phi_{\text{nearest}1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{presence}, & x_{\text{nearest}N}, & y_{\text{nearest}N}, & v_{\text{nearest}N}^x, & v_{\text{nearest}N}^y, & \phi_{\text{nearest}N} \end{bmatrix} \in \mathbb{R}^{N \times 6}, \quad (14)$$

where *presence* represents a binary vehicle indicator, taking the value 1 if the corresponding vehicle is present and 0 otherwise. x_i and y_i represent the lateral and longitudinal positions of CAV_{*i*} in a shared global coordinate maintained by the RSU/MEC, respectively, v_i^x and v_i^y represent the lateral and longitudinal velocity of the CAV_{*i*}; and ϕ_i represent the heading angle of CAV_{*i*}.

Similarly, $x_{\text{nearest}j}$, $y_{\text{nearest}j}$, $v_{\text{nearest}j}^x$, $v_{\text{nearest}j}^y$, and $\phi_{\text{nearest}j}$ represent the corresponding state information of the j -th nearest surrounding vehicle ($j = 1, 2, \dots, N$) as detected by roadside sensors and fused at the MEC. Therefore, the observation vector comprehensively characterizes the local traffic environment perceived by CAV_{*i*}. Accordingly, the observation delivered to CAV_{*i*}, denoted by O_i , is formed by the MEC using the fused global traffic information and transmitted to CAV_{*i*} via V2I communication. This observation vector characterizes the traffic context relevant to CAV_{*i*} under the infrastructure-assisted perception paradigm. The overall observation of the multi-agent system is then given by aggregating the CAV_{*i*} observations $O = [O_1, O_2, \dots, O_{|I|}]$.

2) Action space: In this study, the decision-making process focuses on the longitudinal decision behavior of CAVs, the operational velocity range in our framework is constrained to [3.0, 9.0] m/s (Liu et al., 2025a). For lateral control, a built-in controller, as described in Section 2.2, is employed to achieve accurate trajectory tracking along a predefined path. For longitudinal control, the proposed decision-making model does not directly output acceleration. Instead, the high-level MARL policy generates discrete target-speed commands from the predefined action set of [3, 6, 9] m/s, which are subsequently executed by the low-level controller through an action-smoothing mechanism to ensure stable vehicle motion.

$$U_i = \{\text{decelerate}, \text{idle}, \text{accelerate}\}, \quad (15)$$

where *decelerate* corresponds to reducing the current velocity by one discrete level, *idle* represents maintaining the current velocity, and *accelerate* corresponds to increasing the current velocity by one discrete level, respectively. The selected discrete action is subsequently mapped to a specific target velocity by a low-level longitudinal velocity tracking controller, and then outputs the corresponding acceleration command for vehicle execution (Chen et al., 2024b; Huang et al., 2023; Xiao et al., 2024). The overall action space of the system is formed by aggregating the observations of all CAVs, which can be expressed as

$$U = [U_1, U_2, \dots, U_{|I|}].$$

3) Reward: To encourage safe and efficient intersection crossing, a step-wise reward function is designed.

Collision reward: Safety is of paramount importance in autonomous driving applications. To explicitly enforce safety during the learning process, a collision penalty is imposed whenever a CAV is involved in a collision, this penalty term serves to discourage risky or aggressive behaviors strongly, guiding the policy toward safe and conservative decision-making. To ensure that each CAV can safely reach its target location, this study

introduces a penalty when a vehicle collision occurs to prevent unsafe behavior from happening.

$$r_{i,c} = \begin{cases} -6, & \text{if } CAV_i \text{ collide} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Completion reward: When CAV_i successfully reaches the designated target location, it will be assigned a reward to encourage it to complete the task.

$$r_{i,a} = \begin{cases} 3, & \text{if } CAV_i \text{ arrive} \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Efficiency reward: Throughout the mission, the permissible velocity range through intersections is set at [3 m/s, 9 m/s] according to actual traffic regulations (Liu et al., 2025a). To improve the overall efficiency of the CAV, an efficiency-oriented reward is introduced to encourage the CAV to travel at higher velocities.

$$r_{i,e} = 0.01 * v_i. \quad (18)$$

Finally, the reward function is defined as follows:

$$r = \sum_{i=1}^{|I|} \left(\underbrace{w_c * r_{i,c}}_{\text{collision reward}} + \underbrace{w_a * r_{i,a}}_{\text{completion reward}} + \underbrace{w_e * r_{i,e}}_{\text{efficiency reward}} \right), \quad (19)$$

where r represents the instantaneous reward of CAV_i , and $|I|$ is the total number of CAVs. The coefficients w_c , w_a , and w_e are the corresponding weighting factors. In this study, $w_c = 1$, $w_a = 1$, and $w_e = 1$.

3.2. Priority-aware cooperative safety model

Due to limited perception range and field of view, autonomous vehicles cannot reliably obtain global environmental information in complex traffic scenarios. Although V2V communication can partially mitigate these limitations through information sharing, it assumes that all vehicles are communication-enabled, which remains difficult to achieve at scale. Therefore, the priority-aware cooperative safety model leverages global traffic information acquired by roadside infrastructure to improve coordinated decision-making and safety among CAVs at unsignalized intersections.

In this study, we first define the priority list H for all CAVs at the intersection, where f_i represents the priority level of CAV_i . The priority list H dictates the order in which vehicles are allowed to pass through the intersection, with higher-priority vehicles being granted passage first. The priority of the vehicles is defined as follows:

$$H = [CAV_1, CAV_2, \dots, CAV_{|I|}], \quad (20)$$

$$s.t. \ f_1 > f_2 > \dots > f_{|I|}. \quad (21)$$

Then, we define the priority of each CAV based on the order in which it enters the intersection. If the priority of CAV_i is higher than that of another CAV_j , then CAV_i is granted precedence to pass through the intersection. Once a CAV enters the intersection, it is assigned a position in the priority list H .

$$H = H \cup [CAV_i]. \quad (22)$$

Before all CAVs enter the intersection, each CAV may accelerate according to its priority. Specifically, if CAV_i has the highest priority and meets the safety requirements, it is allowed to accelerate, as defined by the following condition:

$$u_i = \{accelerate\}, \quad (23)$$

$$s.t. \ f_i = \max(f_1, f_2, \dots, f_{|I|}). \quad (24)$$

If CAV_i does not satisfy the priority, it must either maintain its current velocity or decelerate until the conditions are met.

$$u_i = \{decelerate, idle\}. \quad (25)$$

To ensure traffic safety, two longitudinal safety constraints, Time-Headway (THW) and Time-to-Collision (TTC), are enforced exclusively outside the intersection area. Specifically, these constraints apply only when a vehicle travels along the approach or exit lanes, to maintain a safe distance from the preceding vehicle:

$$d_i = d_{front} - d_{ego} \geq \Delta d_{min}, \quad (26)$$

$$\Delta d_{min} = v_{ego} THW_{min}, \quad (27)$$

$$TTC_i = \frac{d_i}{v_{ego} - v_i} \geq TTC_{min}, \quad (28)$$

where d_i is the distance between CAV_i and the vehicle ahead, and v_{ego} is the velocity of CAV_i , v_i is the velocity of the vehicle ahead. According to (Mo et al., 2021), the safe THW_{min} is set to 1.5 s, while the safe TTC_{min} is set to 1.0 s.

Similarly, if CAV_i fails to satisfy the safety constraints, it is required to decelerate until the constraints are fulfilled:

$$u_i = \{decelerate\}. \quad (29)$$

Once a high-priority CAV has completed its passage and exited the intersection, it is promptly removed from the priority list H to ensure orderly passage of subsequent vehicles and maintain the effectiveness of cooperative decision-making. Given that a minimum vehicle velocity threshold of 3 m/s is imposed in this study, the priority assignment strategy is activated when a CAV reaches a predefined distance d from the intersection, thereby effectively preventing potential deadlock situations at the intersection.

When vehicles enter the intersection, global traffic information collected by RSUs is used to optimize CAV actions and enhance safety. Since jointly optimizing all CAVs and HVs incurs considerable computational overhead, an RAT-based screening mechanism is introduced to identify interaction-relevant vehicles and reduce the optimization scope, as detailed below:

$$RAT_i = \left| \frac{P_{i,j}^{conflict} - P_i}{V_i} - \frac{P_{i,j}^{conflict} - P_j}{V_j} \right|, \quad (30)$$

$$f_{inter} = \begin{cases} 1, & RAT_i \leq RAT_{safe} \\ 0, & \text{otherwise} \end{cases}, \quad (31)$$

where $P_{i,j}^{conflict}$, P_i and P_j represent the positions of the conflict point between vehicles i and j , vehicle i , and vehicle j , respectively, while V_i and V_j represent the corresponding velocities of vehicles i and j . For each pair of potentially conflicting predefined paths, the corresponding principal geometric conflict location is determined according to the associated origin-destination (OD) path combination. When RAT_i lies within a predefined safety threshold RAT_{safe} , the corresponding CAV is identified as interaction-relevant and incorporated into the interaction list H_{inter} . Only CAVs in H_{inter} are subsequently selected as optimization targets, thereby focusing computational resources on safety-critical interactions and significantly reducing the overall complexity of the optimization. It should be emphasized that RAT serves here as a lightweight temporal indicator for interaction screening rather than as a standalone geometric collision-detection criterion.

After identifying the interaction-relevant vehicles, their processing order is further determined according to the current traffic conditions. Therefore, the Time to Conflict (TTR) is used to determine the priority of CAVs:

$$TTR_i = \frac{P_{i,j}^{conflict} - P_{CAV_i}}{V_{CAV_i}}, \quad (32)$$

where P_{CAV_i} and V_{CAV_i} represent the position and velocity of the CAV_i , respectively.

A smaller TTR indicates that the vehicle is expected to reach the corresponding conflict location earlier; therefore, it is assigned a higher priority in the sequential action-optimization procedure.

Following the priority ordering, the model sequentially predicts the trajectories of the controlled CAVs and their associated conflict vehicles over a finite prediction horizon of T time steps. At each prediction step, the position, velocity, and heading of each vehicle are dynamically updated according to its motion model. For CAVs, the vehicle-state transition model associated with the MDP is used to propagate their future positions, velocities, and headings under the evaluated candidate actions:

$$P_k^{CAV_i} = \left\{ \begin{bmatrix} x_{k+1}^{CAV_i} \\ y_{k+1}^{CAV_i} \end{bmatrix}, \begin{bmatrix} x_{k+2}^{CAV_i} \\ y_{k+2}^{CAV_i} \end{bmatrix}, \dots, \begin{bmatrix} x_{k+T}^{CAV_i} \\ y_{k+T}^{CAV_i} \end{bmatrix} \right\}, \quad (33)$$

$$P_k^{CAV_j} = \left\{ \begin{bmatrix} x_{k+1}^{CAV_j} \\ y_{k+1}^{CAV_j} \end{bmatrix}, \begin{bmatrix} x_{k+2}^{CAV_j} \\ y_{k+2}^{CAV_j} \end{bmatrix}, \dots, \begin{bmatrix} x_{k+T}^{CAV_j} \\ y_{k+T}^{CAV_j} \end{bmatrix} \right\}, \quad (34)$$

where $P_k^{CAV_i}$ and $P_k^{CAV_j}$ represent the global positions of CAV_i and CAV_j at time step k , respectively, and T represents the length of the prediction horizon.

For HVs, an Intelligent Driver Model (IDM) (Kesting et al., 2010) is adopted to generate trajectory predictions over a T -step prediction horizon (Guo et al., 2024).

$$P_k^{HV} = \left\{ \begin{bmatrix} x_{k+1}^{HV} \\ y_{k+1}^{HV} \end{bmatrix}, \begin{bmatrix} x_{k+2}^{HV} \\ y_{k+2}^{HV} \end{bmatrix}, \dots, \begin{bmatrix} x_{k+T}^{HV} \\ y_{k+T}^{HV} \end{bmatrix} \right\}, \quad (35)$$

where P_k^{HV} represents the global position of the HV at time k .

At each prediction step, the model performs a trajectory-based geometric collision assessment between the controlled CAV and its conflict vehicles. Each vehicle is represented by an oriented rectangular footprint constructed from its predicted position, heading, length, and width. A potential collision is identified when the rectangular footprints of two vehicles intersect at any prediction step.

During action optimization, QMIX adopts a discrete action space to reduce the computational burden of multi-vehicle decision-making. If a footprint intersection is predicted, all available actions are evaluated over the prediction horizon, and the action maximizing the minimum temporal safety margin is selected as follows:

$$u^* = \arg \max_{u_i \in U_i} (\min_{k \in T} RAT_{i,k}). \quad (36)$$

Each conflicting OD-path pair is assigned a predefined conflict point for RAT/TTR-based screening and priority ordering. For multiple conflict points, the minimum temporal margin is adopted. RAT measures temporal separation, whereas trajectory-based geometric collision detection determines the final safety assessment. The overall procedure is summarized in Algorithm 1.

Algorithm 1 Priority-aware cooperative safety model

```

1:  $f_i$  denotes the priority of  $CAV_i$ ,  $f_{i,enter}$  denotes  $CAV_i$  enters the intersection, and  $f_{i,out}$  denotes  $CAV_i$  exits the intersection.
2: Priority Initialization
3: for each  $CAV_i$  do
4:   Compute priority  $f_i$  based on the entry order at an intersection
5: end for
6: Sort the CAVs in descending priority to obtain the sequence:  $H = [CAV_1, \dots, CAV_N]$ 
7: Priority-Based Passing Permission
8: for each  $CAV_i$  do
9:   if  $f_i$  is the highest priority then
10:     Allowed action  $U_i = \{accelerate\}$ 
11:   else
12:     Allowed action  $U_i = \{idle, decelerate\}$ 
13:   end if

```

```

14: if  $CAV_i$  exits the intersection:
15:    $H \leftarrow H \setminus \{CAV_i\}$ 
16: end if
17: end for
18: Safety Constraint Evaluation
19: if  $f_{i,enter}$  or  $f_{i,out}$  then
20:   for each  $CAV_i$  in  $H$  do
21:     Compute THW $i$  and TTC $i$ 
22:     if THW $i$   $\leq$  THW $min$  and TTC $i$   $\leq$  TTC $min$  then
23:       Allowed action  $U_i = \{decelerate\}$ 
24:     end if
25:   end for
26: end if
27: Integrated Safe Decision
28: if  $f_{i,enter}$  and  $f_{i,out}$  then
29:   for each  $CAV_i$  do
30:     if  $CAV_i$  enters the intersection then
31:       Compute RAT $i$  using Eq. (30)
32:       if RAT $i$   $<$  RAT $safe$  then
33:         add  $CAV_i$  to  $H_{inter}$ 
34:       end if
35:     end if
36:   end for
37:   for each  $CAV_i$  in  $H_{inter}$  do
38:     Compute TTR $i$  using Eq. (32)
39:   end for
40:   Sort the CAVs in ascending order of TTR $i$ 
41:   for each  $CAV_i$  in ascending order of TTR $i$  do
42:     for prediction step  $k = 1$  to  $T$  do
43:       predict  $CAV_i$  and  $CAV_j$  trajectories using Eq. (33) and (34)
44:       predict HV trajectories using Eq. (35)
45:       Construct oriented footprints  $CAV_i$  and  $Vehicle_j$  using predicted positions, headings, lengths, and widths
46:       if  $CAV_i$  intersects that of a conflict  $Vehicle_j$  then
47:         Mark  $u_i$  as unsafe
48:       for each candidate action  $u_i^c$  in  $U_i$  do
49:         Predict  $CAV_i$  under  $u_i^c$ 
50:         Compute minimum temporal safety margin  $u_i^c$ 
51:       end for
52:       Solve Eq. (36) and get  $u^*$ 
53:       Replace action with  $u^*$  and update the predicted trajectory
54:     break
55:   end if
56: end for
57: end for
58: end if

```

3.3. Hierarchical weighted sampling strategy

At the first level, episodes are classified according to their cumulative reward. To this end, two replay buffers are maintained: a high-reward buffer D_h and a low-reward buffer D_l . A temporary cache D_c is used to store transitions within the current episode. During the episode, all transitions are first stored in D_c . After the episode terminates at time step T , its cumulative reward is computed as:

$$R_t = \sum_{\tau=0}^{T-1} r_\tau. \quad (37)$$

Given the current maximum cumulative reward R_{max} observed so far and a classification threshold k , the episode is assigned to the corresponding replay buffer according to:

$$\begin{cases} \text{store } R_t \text{ in } D_h, & \text{if } R_t > k * R_{max} \\ \text{store } R_t \text{ in } D_l, & \text{otherwise} \end{cases} \quad (38)$$

After classification, the maximum cumulative reward is updated online as:

$$R_{max} \leftarrow \max(R_{max}, R_t), \quad (39)$$

where R_{max} ensures that the high-reward buffer D_h always contains the most rewarding episodes observed during training. This hierarchical episode partitioning provides a basis for the subsequent weighted sampling strategy.

After the hierarchical episode classification, the replay memory is partitioned into a high-reward buffer D_h and a low-reward buffer D_l . Previous studies have shown that high-reward episodes play a dominant role in accelerating model improvement, while low-reward episodes still provide helpful information for exploration (Guan et al., 2022; Wu et al., 2024). To balance the contributions of these two types of experiences, a weighted dual-buffer loss is designed in this study.

During each update step, two mini-batches are independently sampled: $b_h \subset D_h$ from the high-reward buffer and $b_l \subset D_l$ from the low-reward buffer, where b_h and b_l denote the batch sizes drawn from D_h and D_l , respectively.

The loss function with high-reward episodes is defined as:

$$L_h(\theta) = \frac{1}{b_h} \sum_{m_1=1}^{b_h} (y^{tot} - Q_{tot}(s, \mathbf{u}; \theta))^2. \quad (40)$$

While the loss function for low-reward episodes is given by:

$$L_l(\theta) = \frac{1}{b_l} \sum_{m_2=1}^{b_l} (y^{tot} - Q_{tot}(s, \mathbf{u}; \theta))^2. \quad (41)$$

To emphasize the positive effect of high-reward trajectories and to prevent low-reward trajectories from dominating the optimization process, the overall training objective is formulated as a weighted combination of the two losses:

$$L(\theta) = L_h(\theta) + \lambda L_l(\theta), \quad (42)$$

where $\lambda \in (0,1)$ is a weighting coefficient that controls the relative influence of low-reward episodes. This dual-buffer and weighted-loss design encourages the policy to gradually shift toward the high-return region of the episode space, while maintaining sufficient diversity from low-reward experiences to preserve exploration and training stability.

4 Experiment and Results

4.1. Experiment setting

This study utilizes the highway_env simulation (Leurent, 2018) to model a multi-agent environment at an intersection, setting up two experimental scenarios: the first scenario includes only CAVs. In contrast, the second scenario incorporates HVs alongside the CAVs. The maximum simulation time for each

scenario is set to 100 s, with evaluations conducted every 200 training episodes. The simulation time step is 0.1 s, and the decision execution time is 0.2 s. The computer configuration features an Intel Core i7-11700F processor and an NVIDIA GeForce RTX 3060 Ti graphics card. The operating system is Ubuntu 18.04, running on 16GB of memory. The programming environment uses Python 3.8, and the deep learning framework is PyTorch 1.10.0. The hyperparameters adopted for QMIX are summarized as follows. The learning rate is set to 0.0005, and the discount factor is 0.99. The batch size is fixed at 32, and the target network is updated every 200 episodes. An ϵ -greedy exploration strategy is employed, with the exploration rate initialized to 1.0 and gradually decayed to a final value of 0.05. For the hierarchical sampling scheme, the two sub-mini-batch sizes, m_1 and m_2 , are both set to 16. The weighting coefficient λ is fixed at 0.3, and the episodic sample classification threshold k , is set to 0.3. The detailed architecture parameters of the model are shown in Table 1.

4.2. Training in Simulation Environments

To empirically validate the effectiveness of the proposed model, experiments are conducted under two different scenarios, and the episode reward and success rate of QMIX, Priority-Aware Cooperative Safety QMIX (PCSQMIX), Hierarchical Weighted Sampling (QMIX-HWS), and Priority-Aware Cooperative Safety QMIX with Hierarchical Weighted Sampling (PCSQMIX-HWS) are shown for different scenarios. Additionally, the proposed model is also compared with the Proximal Policy Optimization (PPO) (Schulman et al., 2017) and the VDN (Sunehag et al., 2017), as illustrated in Figs. 4 and 5.

As shown in Fig. 4(a), in Scenario 1, PCSQMIX-HWS achieves the highest episode reward, converging to approximately 40, which clearly demonstrates that the joint integration of the priority-aware cooperative safety model and the hierarchical weighted sampling strategy substantially enhances model performance. The PCSQMIX model exhibits the second-best performance among all compared methods, indicating that the priority-aware cooperative safety model plays a crucial role in improving the vanilla QMIX framework. In contrast, the PPO and VDN models perform the worst during training, which can be attributed to their inherent limitations. Specifically, VDN aggregates local Q-values through a simple summation, which is insufficient to capture the complex inter-agent interactions and dependencies, thereby limiting its ability to approximate the joint action-value function accurately. On the other hand, PPO, which strictly follows individual reward maximization, is prone to inducing goal conflicts among agents in scenarios requiring tight cooperation, ultimately leading to degraded overall performance.

Fig. 4(b) compares the training success rates of PPO, VDN, QMIX, PCSQMIX, QMIX-HWS, and PCSQMIX-HWS in Scenario 1. The results show that both PCSQMIX and QMIX-HWS achieve higher success rates than the vanilla QMIX, with PCSQMIX exhibiting the most pronounced improvement.

Table 1 Model hyperparameter setting.

Parameter Name	Parameter Description	Parameter Value
Local Network Layer 1	Number of neurons in the first local network layer	[State input dimension, 64]
Local Network GRU Layer	Number of neurons in the GRU layer of the local network	[64,64]
Local Network Layer 2	Number of neurons in the second local network layer	[64, action output dimension]
Global Network W_1	Number of neurons in the bias term of the first global network	[Joint state input dimension, 4*32]
Global Network b_1	Number of neurons in the second global network layer	[Joint state input dimension, 32]
Global Network W_2	Number of neurons in the second global network layer	[Joint state input dimension, 32]
Global Network b_2	Number of neurons in the bias term of the second global network	[Joint state input dimension, 32]

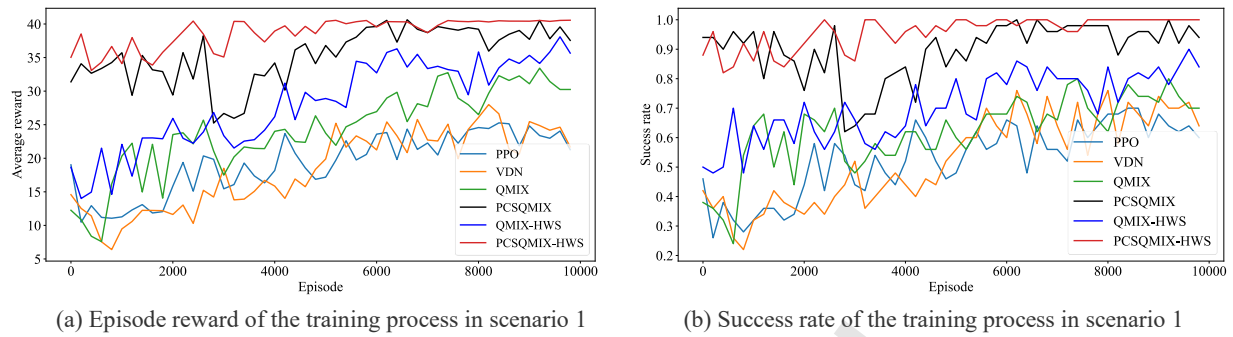


Fig. 4. Episode reward and success rate of the training process in scenario 1.

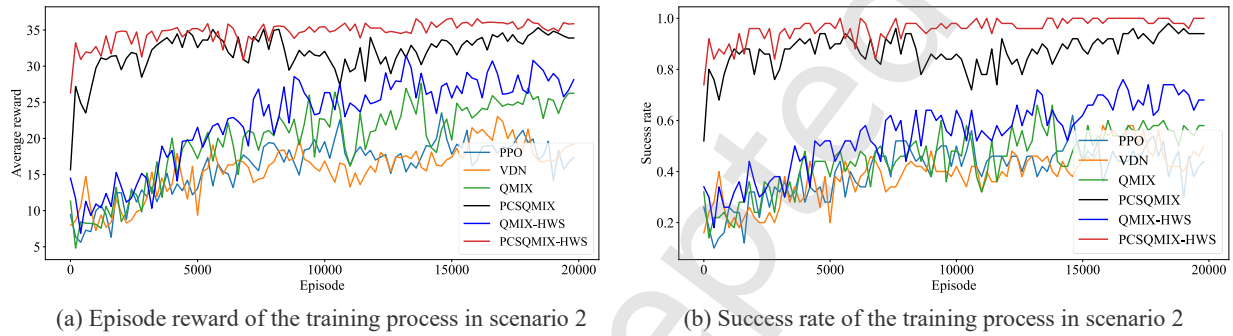


Fig. 5. Episode reward and success rate of the training process in scenario 2.

This result further confirms the effectiveness of the priority-aware cooperative safety model. Although the success rate of QMIX-HWS is slightly lower than that of PCSQMIX-HWS and PCSQMIX, it still significantly outperforms vanilla QMIX, indicating that the hierarchical weighted sampling strategy also contributes to performance enhancement. This advantage primarily stems from its ability to assign greater importance to high-value samples, thereby enabling the model to learn more informative experiences. Ultimately, by jointly leveraging the weighted sampling strategy, PCSQMIX-HWS achieves the highest success rate among all compared methods, reaching 100% after convergence. This result further demonstrates that the proposed model is capable of learning a more effective policy.

Similar trends are also observed in scenario 2. As shown in Fig. 5, with the further increase in environmental complexity, both the episode rewards and success rates of all models exhibit varying degrees of degradation. Nevertheless, PCSQMIX-HWS consistently outperforms all competing methods, achieving a final episode reward of approximately 37 and a success rate of 98%, which remains the highest among all models. In contrast, the PPO model demonstrates the weakest performance in scenario 2 with a final success rate of only 46%. This result further highlights the inherent limitations of single-agent models, which focus solely on individual reward maximization and fail to explicitly model inter-agent cooperation, thereby leading to significant performance degradation. Meanwhile, PCSQMIX and QMIX-HWS exhibit a certain degree of performance degradation, with final episode rewards of approximately 32 and 27, and success rates of 94% and 72%, respectively. To further verify the effect of the safety constraints, we conduct additional experiments in scenario 2. As shown in Fig. 6, PCSQMIX-HWS without safety constraints exhibits inferior performance to PCSQMIX-HWS in terms of both average reward and success rate. Specifically, its final average reward is only 34, and its success rate reaches only 90%. These results indicate that the safety constraints not only enhance safety-related performance but also contribute to more stable policy

learning and more reliable task execution.

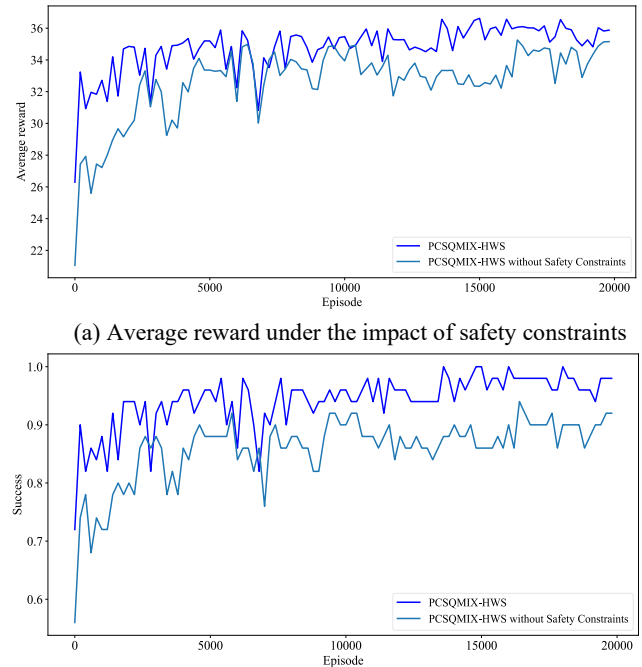


Fig. 6. Impact of safety constraints on model performance.

In addition, we conduct experiments with different weighting coefficients in scenario 2. As illustrated in Fig. 7, when $\lambda=0.5, 0.7$, or 0.9 , the final performance of the model is closer to that of the QMIX model. By comparison, the best overall performance is achieved when $\lambda=0.3$. However, further decreasing λ below 0.3 leads to performance degradation. These findings suggest that assigning a relatively smaller weight can effectively reduce their adverse influence during training and thus improve the overall model performance. At the same time, overly reducing weight may

result in insufficient effective training data, ultimately causing performance deterioration.

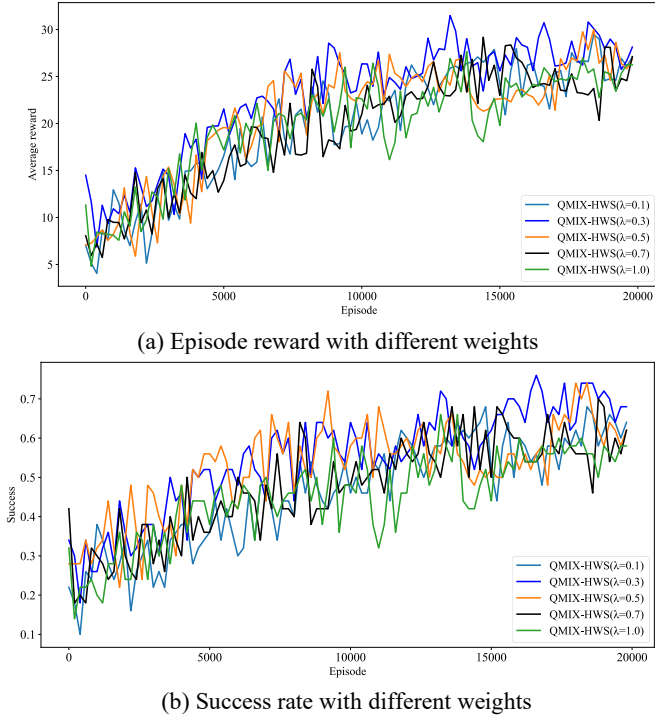
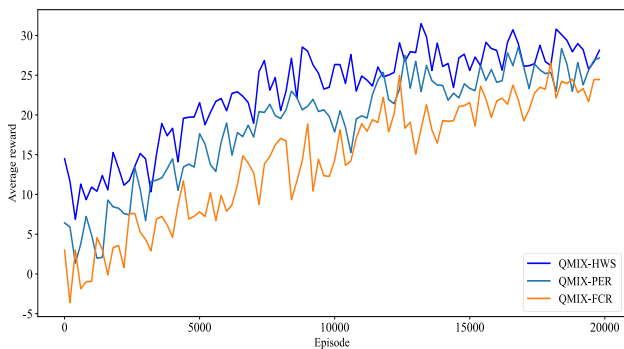
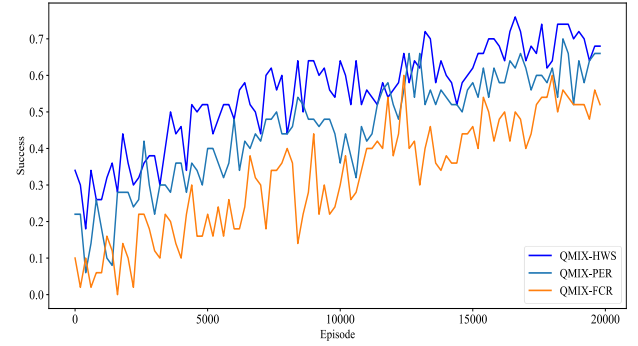


Fig. 7. Episodes rewards and success rates with different weights in scenario 2.

Fig. 8 compares the training performance of different experience replay sampling strategies in scenario 2, including QMIX with Prioritized Experience Replay (QMIX-PER), QMIX with Failure-case Replay (QMIX-FCR), and the proposed QMIX-HWS. To isolate the effect of the replay mechanism, all three methods are implemented with the same vanilla QMIX and identical training settings, except for the replay sampling strategy. The results show that QMIX-HWS achieves the highest episode reward and success rate, demonstrating the effectiveness of the proposed sampling mechanism in improving cooperative decision-making performance. In contrast, QMIX-FCR performs the worst. Although failure samples, such as collision and near-collision cases, contain valuable safety-boundary information, excessive oversampling of such low-reward samples may bias the training distribution toward failure interactions and weaken the learning of successful cooperative strategies. QMIX-PER improves the performance relative to QMIX-FCR, but its final reward and success rate are still lower than those of QMIX-HWS. These results indicate that QMIX-HWS provides a more effective balance between high- and low-reward samples, reducing overreliance on low-reward episodes and consequently achieving better training performance.



(a) Episode reward with different experience replay strategies



(b) Success rate with different experience replay strategies
Fig. 8. Episode reward and success rate with different experience replay strategies in scenario 2.

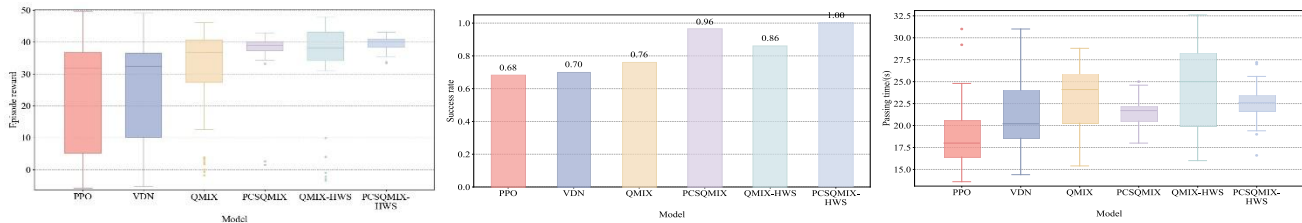
4.3. Testing in Simulation Environments

After the training phase, 50 testing episodes that are not encountered during training are selected for evaluation. Fig. 9 and Table 2 present the episode rewards, success rates, and passing times of different models in Scenario 1. As shown in Figs. 9(a) and (b), the episode rewards obtained by PPO, VDN, QMIX, PCSQMIX, QMIX-HWS, and PCSQMIX-HWS are 24.95, 26.24, 30.22, 37.31, 34.20, and 39.48, respectively. PCSQMIX-HWS achieves the highest success rate of 100%, followed by PCSQMIX at 96%, QMIX-HWS at 86%, QMIX at 76%, VDN at 70%, and PPO at 68%. This indicates that incorporating a priority-aware cooperative safety model and hierarchical sampling strategies can achieve higher cumulative rewards and ensure safe driving performance. Regarding passage efficiency, Fig. 9(c) shows the average passing time results. The average passing times for PCSQMIX-HWS, PCSQMIX, QMIX-HWS, and QMIX are 22.58 s, 21.50 s, 24.14 s, and 23.08 s, respectively, reflecting that PCSQMIX-HWS achieves competitive passage efficiency while maximizing security. In summary, through the synergistic integration of the priority-aware collaborative security model and the hierarchical weighted sampling strategy, PCSQMIX-HWS exhibits the best overall performance, achieving effective collaborative decision-making in Scenario 1. In contrast, the single-agent PPO model performs the worst because its optimization process only focuses on individual rewards, lacking effective collaboration, leading to overly aggressive individual behavior, thus compromising overall security and operational efficiency.

Similarly, Fig. 10 and Table 2 show the episode rewards, success rates, and passing times of different models in scenario 2. In terms of episode reward performance, the episode rewards achieved by PPO, VDN, QMIX, PCSQMIX, QMIX-HWS, PCSQMIX-HWS are 18.16, 20.30, 27.20, 36.34, 30.11, and 38.79, respectively. Although the episode reward of PCSQMIX-HWS decreases slightly compared to Scenario 1, it still reaches 38.79, remaining the highest among all compared models. In terms of success rate, PCSQMIX-HWS achieves a 98% success rate, indicating more stable performance under complex traffic conditions. Regarding traffic efficiency, the average passing time of PCSQMIX-HWS is 24.22 s, only 0.22 s longer than PCSQMIX, while the success rate improves by 4%. This demonstrates that although PCSQMIX-HWS slightly sacrifices some efficiency, it improves safety.

Subsequently, the proposed algorithm is compared with the IDM (Kesting et al., 2010), the rule-based FCFS, the DRL-based HAG-TRPO (Chen et al., 2024b), and the DRL-based MA-GA-DDPG (Liu et al., 2025a) with the results summarized in Figs. 11, 12 and Table 3. As shown in Fig. 11 and Table 3, although the IDM algorithm achieves a relatively high average velocity of 8.66 m/s, its success rate in Scenario 1 is limited to 38%, indicating that while

it optimizes velocity, it fails to ensure overall safety, resulting in a

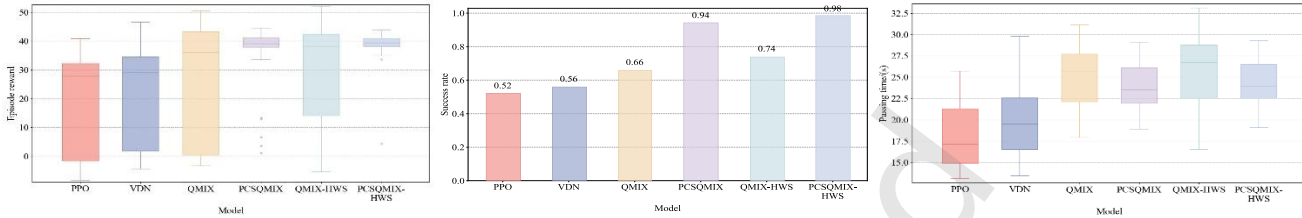


(a) Episode reward in scenario 1

(b) Success rate in scenario 1

(c) Passing time in scenario 1

Fig. 9. Testing the performance of different models in scenario 1.

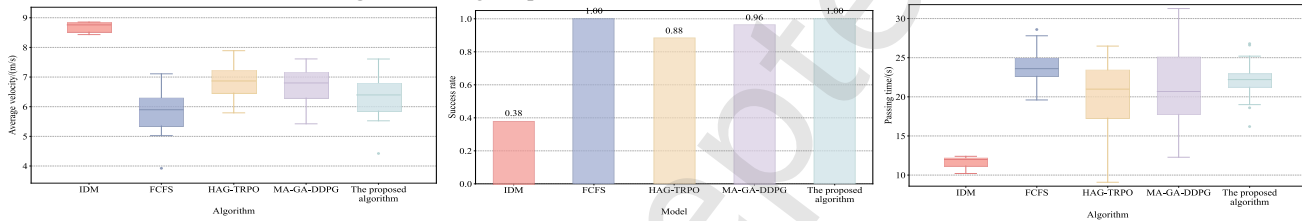


(a) Episode reward in scenario 2

(b) Success rate in scenario 2

(c) Passing time in scenario 2

Fig. 10. Testing the performance for different models in scenario 2.

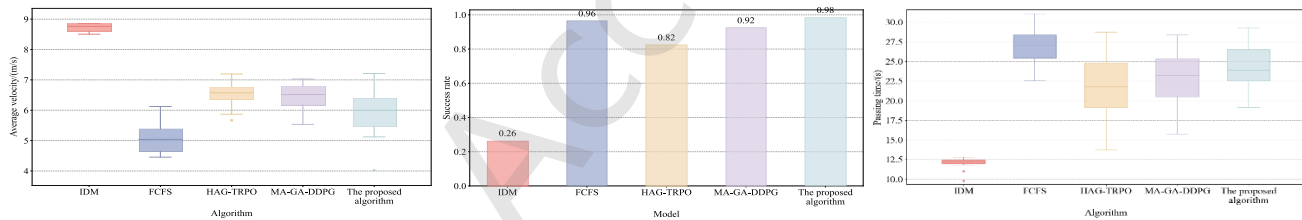


(a) Average velocity in scenario 1

(b) Success rate in scenario 1

(c) Passing time in scenario 1

Fig. 11. Testing performance for different algorithms in scenario 1.



(a) Average velocity in scenario 2

(b) Success rate in scenario 2

(c) Passing time in scenario 2

Fig. 12. Testing performance for different algorithms in scenario 2.

Table 2 Comparison of model performance in different scenarios.

Model	Scenario 1			Scenario 2		
	Average episodic reward	Success rates	Average passing time (s)	Average episodic reward	Success rates	Average passing time (s)
PPO	24.95	68%	18.97	18.16	52%	18.00
VDN	26.24	70%	21.00	20.30	56%	20.03
QMIX	30.22	76%	23.08	27.20	66%	24.97
PCSQMIX	37.31	96%	21.50	36.34	94%	24.00
QMIX-HWS	34.20	86%	24.14	30.11	74%	25.59
PCSQMIX-HWS	39.48	100%	22.58	38.79	98%	24.22

higher incidence of conflict. In contrast, both the proposed algorithm and the FCFS algorithm achieve a 100% success rate in Scenario 1. Notably, the proposed algorithm consistently outperforms the FCFS algorithm, improving the average velocity by 0.49 m/s and reducing the overall passing time by 1.29 s. These results demonstrate that the proposed algorithm not only maintains a high level of safety but also significantly enhances traffic efficiency, thereby improving the overall effectiveness of cooperative vehicle coordination. Compared with DRL-based baselines, although HAG-TRPO and MA-GA-DDPG achieve

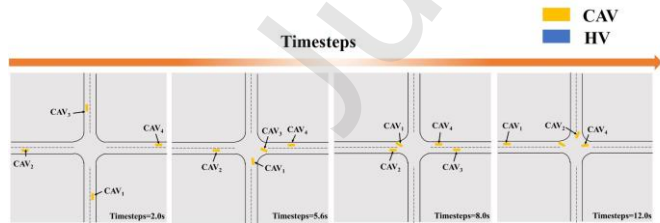
relatively high average velocities, their success rates are limited to 88% and 96%, respectively, indicating that safety cannot be fully guaranteed due to the lack of effective coordination. In contrast, the proposed algorithm achieves a no-collision operation while maintaining a competitive average velocity, highlighting its superior performance.

Similarly, as illustrated in Fig. 12 and Table 3, in scenario 2, the performance degradation is more significant with increasing task difficulty; the success rate of the IDM model is only 26%. In contrast, the proposed algorithm and the FCFS algorithm

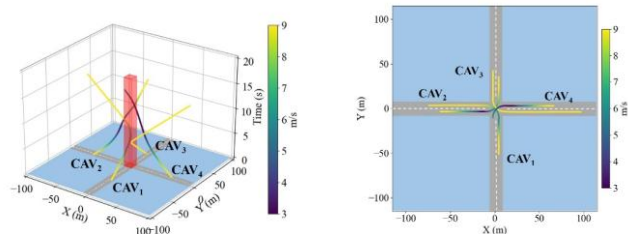
consistently maintain high success rates of 98% and 96%, respectively, demonstrating their ability to handle complex interaction scenarios more effectively. Compared to the FCFS algorithm, the proposed algorithm not only improves the average velocity by 0.90 m/s but also reduces the total passage time by 2.55 s. Furthermore, compared to the DRL-based benchmark algorithms, although HAG-TRPO and MA-GA-DDPG achieve relatively high average speeds of 6.52 m/s and 6.44 m/s, respectively, their success rates drop to 82% and 92% with increasing scenario complexity, showing a significant decline in safety performance. In contrast, although the proposed algorithm has a slightly lower average speed of 5.94 m/s, it consistently maintains a high success rate, thus achieving a good balance between efficiency and safety.

To further evaluate the performance of the proposed model in different scenarios, representative cases in scenario 1 and scenario 2 are illustrated in Figs. 13 and 14, respectively. In Scenario 1, all four CAVs initially drive at their maximum velocities. After approximately 1 s, CAV₂ and CAV₄ decelerate, followed by CAV₁, which reduces its velocity while approaching the intersection to avoid a potential conflict with others. Benefiting from the assigned priority, CAV₃ is granted right-of-way and enters the intersection first, passing the conflict area at the maximum velocity. After CAV₃ leaves the conflict area, CAV₁ accelerates and passes the intersection. Subsequently, CAV₂ enters and passes through the intersection without incident, while the remaining CAV₄ adjusts its velocity accordingly. Overall, within the proposed framework, each CAV can dynamically adjust its motion state and traffic sequence, thereby enabling all vehicles to pass through the intersection in an orderly and collision-free.

In scenario 2, CAV₁, CAV₂, CAV₃, and CAV₄ all initially travel at a speed of 9 m/s. When CAV₃ approaches the intersection at approximately 4 s, the RSU detects a potential collision between CAV₃ and HV and immediately issues a deceleration instruction. CAV₃ yields to HV before it passes, and accelerates to complete the crossing after HV has passed. After CAV₃ passes the intersection, CAV₁ begins to accelerate and proceeds toward the intersection. At time step 9 s, when CAV₂ attempts to enter the intersection, another potential conflict with CAV₁ is identified by the RSUs, triggering a deceleration command; once CAV₁ clears the intersection, CAV₂ accelerates and passes safely. Subsequently, CAV₄ passes the intersection in sequence. This case study demonstrates that the proposed V2I-IDADM framework effectively coordinates CAV behaviors in the presence of HVs through priority-based passage and infrastructure-assisted safety interventions, thereby ensuring orderly operations and safety.

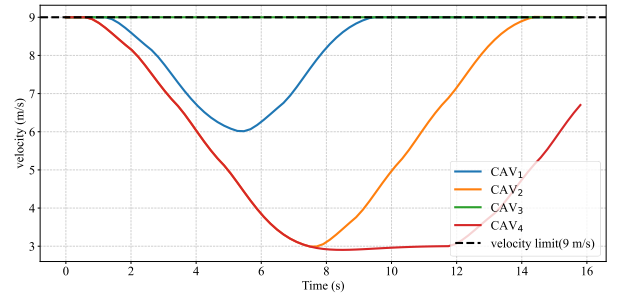


(a) Snapshots of critical moments in scenario 1



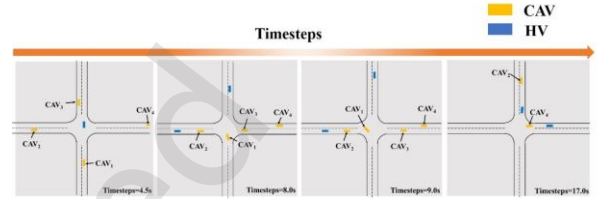
(b) The 3D movement paths in scenario 1

(c) Planar trajectory in scenario 1

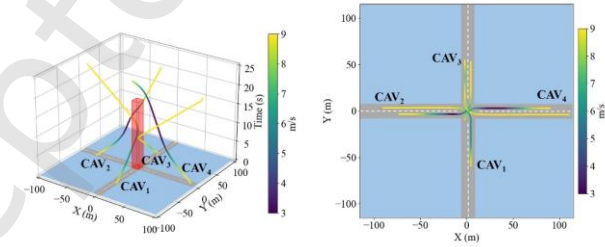


(d) Actual operating velocity in scenario 1.

Fig. 13. Test results of a typical case in scenario 1.

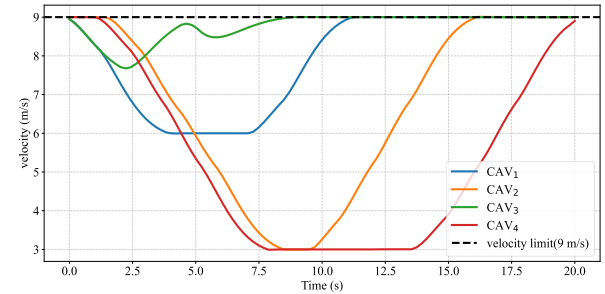


(a) Snapshots of critical moments in scenario 2



(b) The 3D movement paths in scenario 2

(c) Planar trajectory in scenario 2



(d) Actual operating velocity in scenario 2.

Fig. 14. Test results of a typical case in scenario 2.

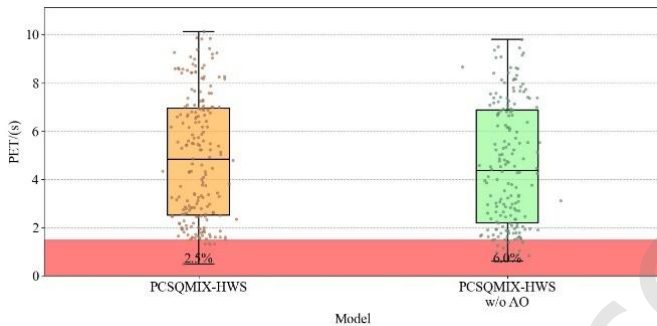
4.4. Sensitivity and Robustness Analysis

To evaluate the safety effectiveness of the proposed action optimization mechanism, we compare PCSQMIX-HWS with PCSQMIX-HWS without action optimization, denoted as PCSQMIX-HWS w/o AO. In addition to the collision rate and success rate, the post-encroachment time (PET) metric is introduced to quantify temporal safety at conflict points, where PET values below 1.5 s are regarded as PET violations (Lin et al., 2025). The results are presented in Fig. 15 and Table 4. As shown in Table 4, incorporating the action optimization mechanism improves the success rate from 86% to 98%, corresponding to a 12% increase, while reducing the collision rate from 14% to 2%. Moreover, PCSQMIX-HWS achieves a higher mean PET of 4.88 s, compared with 4.52 s for PCSQMIX-HWS w/o AO, indicating that the proposed mechanism provides a larger temporal safety margin during conflict resolution. The PET distribution in Fig. 15 further shows that PCSQMIX-HWS exhibits the lowest PET violation rate, with only 2.5% of cases falling below the 1.5 s

Table 3 Comparison of different algorithms.

Methods	Scenario 1			Scenario 2		
	Average velocity (m/s)	Success rates	Average passing time (s)	Average velocity (m/s)	Success rates	Average passing time (s)
IDM	8.66	38%	11.70	8.73	26%	12.06
FCFS	5.84	100%	23.87	5.04	96%	26.87
HAG-TRPO	6.81	88%	20.14	6.52	82%	22.04
MA-GA-DDPG	6.69	96%	20.61	6.44	92%	22.67
The proposed algorithm	6.33	100%	22.58	5.94	98%	24.22

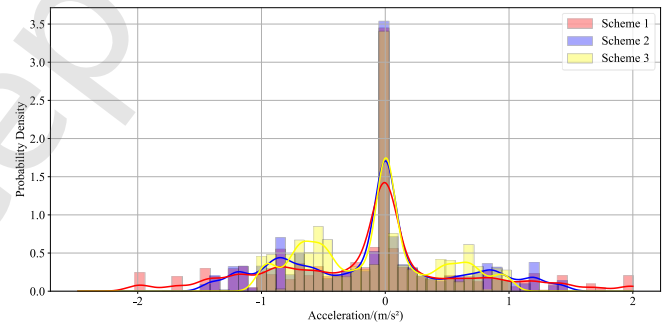
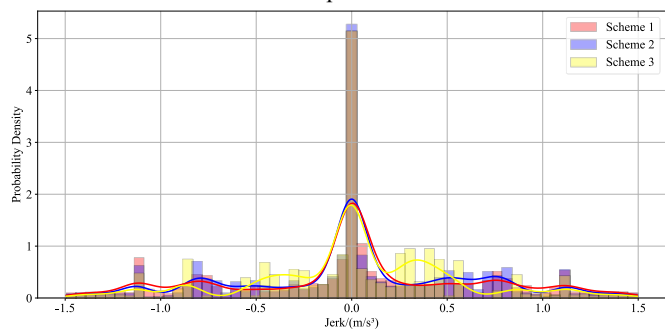
threshold, whereas PCSQMIX-HWS w/o AO results in a considerably higher PET violation rate of 6.0%. These results demonstrate that the proposed action optimization mechanism can effectively mitigate conflict risks at unsignalized intersections and enhance the overall safety of cooperative decision-making.

**Fig. 15.** Effect of action optimization on PET distribution in scenario 2.**Table 4** Ablation comparison of performance in scenario 2.

Model	Success rates	Collision rates	Mean PET (s)
PCSQMIX-HWS	98%	2%	4.88
PCSQMIX-HWS w/o AO	86%	14%	4.52

To further examine the influence of action design, we compare three speed discretization schemes: Scheme 1 with three discrete speed values [3, 6, 9] m/s, Scheme 2 with five values [3, 4.5, 6, 7.5, 9] m/s, and Scheme 3 with seven values [3, 4, 5, 6, 7, 8, 9] m/s. As shown in Table 5, Scheme 1 achieves the highest success rate of 98% and the lowest collision rate of 2%. Moreover, no emergency braking events, defined as acceleration lower than -9m/s^2 (el Abidine Kherroubi et al., 2021), are observed under any of the three schemes. Although a finer speed discretization slightly reduces the average passing time, it also leads to an increase in the collision rate. This result indicates that a denser speed action scheme can improve driving efficiency, but may also increase decision-making complexity and introduce greater instability in action selection. Furthermore, as shown in Figs. 16, 17 and Table 6, the acceleration and jerk distributions indicate that the mean absolute acceleration values of Scheme 1, Scheme 2, and Scheme 3 are 0.56 m/s^2 , 0.48 m/s^2 , and 0.39 m/s^2 , respectively, with corresponding RMS acceleration values of 0.79 m/s^2 , 0.66 m/s^2 , and 0.52 m/s^2 . Similarly, the mean absolute jerk values are 0.46 m/s^3 , 0.43 m/s^3 , and 0.37 m/s^3 , while the corresponding RMS jerk values are 0.64 m/s^3 , 0.61 m/s^3 , and 0.53 m/s^3 . These results suggest that all three schemes maintain acceptable longitudinal smoothness without introducing excessive acceleration or jerk fluctuations.

Overall, although Scheme 1 exhibits slightly higher acceleration and jerk than Schemes 2 and 3, it achieves the highest success rate and lowest collision rate while maintaining acceptable longitudinal smoothness. In contrast, finer discretization provides only marginal improvements in efficiency and smoothness but results in higher collision rates. Therefore, Scheme 1 offers a more favorable balance among safety, efficiency, and ride comfort, supporting its use in the proposed framework.

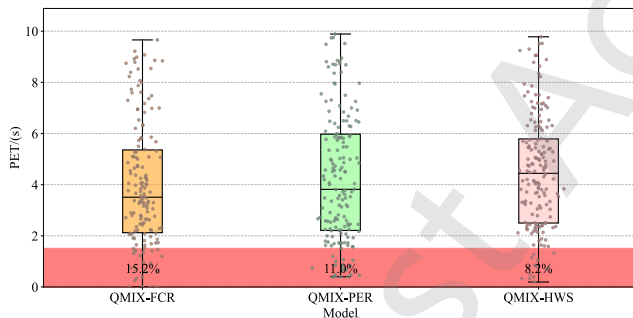
**Fig. 16.** Acceleration probability density distributions under different action spaces in scenario 2.**Fig. 17.** Jerk probability density distributions under different action spaces in scenario 2.**Table 5** Comparison of different speed values in scenario 2.

Scheme	Success rates	Collision rates	Average passing time (s)	Number of emergency braking
Scheme 1	98%	2%	24.22	0
Scheme 2	94%	6%	24.08	0
Scheme 3	92%	8%	23.89	0

Table 6 Comparison of accelerations and jerks in scenario 2.

Scheme	Mean absolute acceleration (m/s ²)	RMS acceleration (m/s ²)	Mean absolute jerk (m/s ³)	RMS jerk (m/s ³)
Scheme 1	0.56	0.79	0.46	0.64
Scheme 2	0.48	0.66	0.43	0.61
Scheme 3	0.39	0.52	0.37	0.53

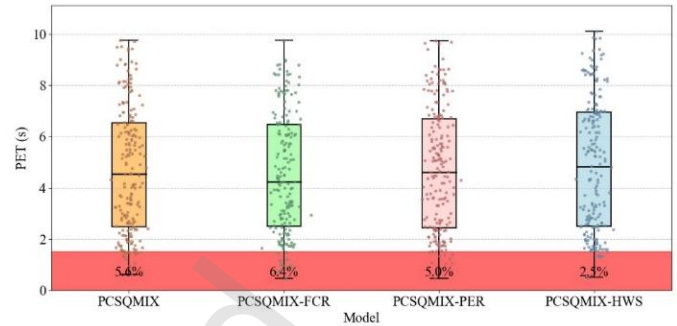
Fig. 18 and Table 7 present the performance comparison of three experience sampling strategies. The results show that QMIX-FCR exhibits the weakest safety performance, with a success rate of only 62%, a collision rate as high as 38%, the lowest average PET among the three methods at 3.97 s, and an average passing time of 24.75 s. In comparison, QMIX-PER improves the overall performance to some extent, reducing the collision rate to 30%, increasing the average PET to 4.23 s, and yielding an average passing time of 25.40 s. Furthermore, the proposed QMIX-HWS achieves the best safety performance among the three strategies, with the lowest collision rate of 26% and the highest average PET of 4.41 s, while its average passing time is 25.59 s. In addition, the PET distribution further confirms the safety advantage of QMIX-HWS. Specifically, QMIX-FCR has the highest PET violation rate, reaching 15.2%, whereas QMIX-PER reduces to 11.0%. By contrast, QMIX-HWS achieves the lowest PET violation rate of 8.2%. These results indicate that the proposed HWS strategy improves the utilization of high-reward episodes while retaining sufficient safety-critical information from collision- and near-miss-related experiences, thereby enhancing the safety of multi-agent cooperative decision-making.

**Fig. 18.** PET distribution under different experience replay strategies in scenario 2.**Table 7** Comparison of different experience replay strategies in scenario 2.

Model	Success rates	Collision rates	Average passing time(s)	Mean PET (s)
QMIX-HWS	74%	26%	25.59	4.41
QMIX-PER	70%	30%	25.40	4.23
QMIX-FCR	62%	38%	24.75	3.97

To further evaluate the effectiveness of HWS, Fig. 19 and Table 8 present the performance of PCSQMIX integrated with different experience replay strategies. Consistent with the comparisons conducted using the QMIX, PCSQMIX-HWS achieves the best overall performance among all evaluated variants, attaining a success rate of 98% and a mean PET of 4.88 s. In contrast, PCSQMIX-FCR exhibits the poorest performance, with a success rate of only 90% and a mean PET of 4.58 s. Although PCSQMIX-PER achieves the same success rate as the standard

PCSQMIX, it improves both the PET violation rate and the mean PET relative to PCSQMIX. Taken together, the consistent results obtained using both the QMIX and PCSQMIX backbones demonstrate that HWS provides an additional and independently measurable safety improvement and achieves the best overall performance among the compared experience replay strategies.

**Fig. 19.** PET distribution under different experience replay strategies in scenario 2.**Table 8** Comparison of different experience replay strategies in scenario 2.

Model	Success rates	Collision rates	Average passing time(s)	Mean PET (s)
PCSQMIX	94%	6%	24.00	4.67
PCSQMIX-FCR	90%	10%	23.43	4.58
PCSQMIX-PER	94%	6%	23.95	4.70
PCSQMIX-HWS	98%	2%	24.22	4.88

Furthermore, we evaluate the safety performance of the proposed PCSQMIX-HWS under various disturbance conditions and compare it with that of PCSQMIX. As shown in Table 9, under adverse conditions such as perceptual noise, communication latency, abnormal HV behavior, communication interruption, and boundary state initialization, the performance of both models degrades to varying degrees. Nevertheless, in all test scenarios, PCSQMIX-HWS consistently outperforms PCSQMIX. Specifically, PCSQMIX-HWS maintains a success rate of 90%–96% and a collision rate of 4%–10%, while PCSQMIX has a lower success rate of 84%–94% and a higher collision rate of 6%–16%. Across the different disturbance conditions, PCSQMIX-HWS reduces the collision rate by 2–6% compared with PCSQMIX. Moreover, it consistently achieves higher mean PET values and lower PET violation rates, indicating a stronger capability to mitigate potential conflicts and preserve temporal safety margins under disturbed operating conditions. Among all tested cases, boundary-state initialization with PET < 1.5 s results in the most pronounced performance degradation, confirming that low-PET interactions pose the greatest challenge to policy safety due to the limited time available for conflict resolution. Even under this highly critical condition, PCSQMIX-HWS achieves a success rate of 90% and a collision rate of 10%, compared with 84% and 16% for PCSQMIX. These results demonstrate that the proposed HWS strategy improves the robustness of cooperative decision-making and enhances safety performance under diverse disturbances and safety-critical interaction conditions.

To further evaluate the generalization capability of the proposed model under more realistic HV behaviors, we introduce a Gap Acceptance Model (Boon et al., 2019) into the unsignalized intersection scenario. Specifically, the model is employed to characterize HVs' gap-acceptance behavior, yielding, waiting, and

crossing decisions in the intersection, while the IDM is retained to describe longitudinal car-following behavior when vehicles are

Table 9 Comparison of safety performance under different conditions in scenario 2.

Cases	Models	Success rate	Collision rate	Mean PET (s)	PET violation rate
Perception Noise (Gaussian noise 10%)	PCSQMIX-HWS	96%	4%	4.81	3.1%
	PCSQMIX	92%	8%	4.66	4.4%
Communication delay (300ms)	PCSQMIX-HWS	96%	4%	4.72	3.0%
	PCSQMIX	94%	6%	4.65	5.3%
Boundary initialization (PET<1.5s)	PCSQMIX-HWS	90%	10%	4.59	5.9%
	PCSQMIX	84%	16%	4.50	7.2%
Abnormal HV behavior (Random acceleration changes)	PCSQMIX-HWS	96%	4%	4.79	4.1%
	PCSQMIX	92%	8%	4.67	4.9%
Communication interrupt (Interruption possibility 5%)	PCSQMIX-HWS	96%	4%	4.73	4.0%
	PCSQMIX	90%	10%	4.62	5.6%

sufficiently far from the intersection. This hybrid IDM–Gap Acceptance formulation allows the HVs to exhibit more diverse and interaction-dependent behaviors near the conflict zone, thereby providing a more realistic test environment for assessing the robustness of the proposed cooperative decision-making framework.

As shown in Fig. 20 and Table 10, replacing the IDM-only HV model with the IDM–Gap Acceptance model does not reduce the success rate, which remains at 98%. Meanwhile, the mean PET increases to 5.09 s, the PET violation rate slightly decreases to approximately 2.0%, and the average passing time is 24.80 s. These results indicate that the proposed model maintains stable safety performance despite the increased uncertainty and behavioral heterogeneity introduced by the more realistic HV model. Compared with the IDM-only model, the IDM–Gap Acceptance model more explicitly represents yielding, waiting, gap acceptance, and crossing decisions at unsignalized intersections, and therefore introduces more complex interaction patterns between CAVs and HVs. Overall, these results provide further evidence of the generalization capability of the proposed framework under more realistic mixed-traffic interaction conditions.

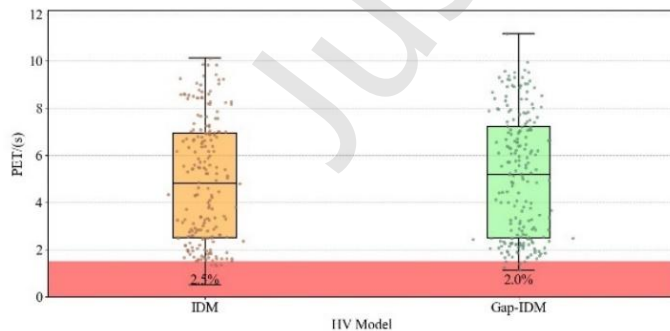


Fig. 20. PET distributions under different HV models in scenario 2.

Table 10 Performance comparison under different HV behavior models in scenario 2.

HV Model	Success rates	Collision rates	Average passing time(s)	Mean PET (s)
IDM	98%	2%	24.22	4.88
IDM+Gap Acceptance	98%	2%	24.80	5.09

Furthermore, to evaluate the real-time performance and computational scalability of the proposed framework under different traffic scales, runtime tests are conducted in scenario 2 with 4, 12, and 20 CAVs. The results are presented in Fig. 21 and Table 11. With 4 CAVs, the average decision time is 6.81 ms, with a maximum of 61.46 ms. When the number of CAVs increases to 12, the average and maximum decision times increase to 8.70 and 86.13 ms, respectively. With 20 CAVs, the corresponding values further increase to 10.73 and 112.98 ms. These results indicate that the increasing number of inter-vehicle interactions and potential conflicts introduces additional computational overhead as the traffic scale grows, resulting in increases in both the average and maximum decision times. Nevertheless, even under the 20-CAV setting, the maximum decision time remains below the 200-ms decision cycle. This demonstrates that the proposed framework maintains satisfactory computational scalability and satisfies the real-time decision-making requirement within the evaluated traffic scales.

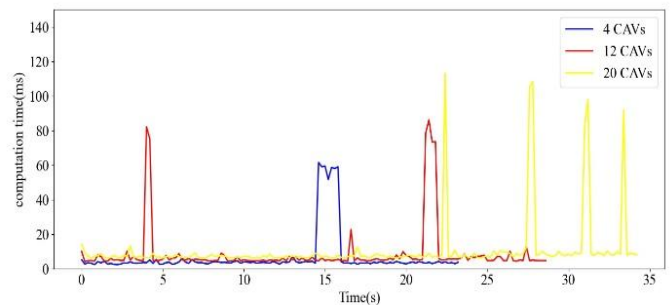


Fig. 21. Computation time with different numbers of CAVs.

Table 11 Decision Time with different numbers of CAVs.

Number of CAVs	Average decision time (ms)	Max decision time (ms)
4	6.81	61.46
12	8.70	86.13
20	10.73	112.98

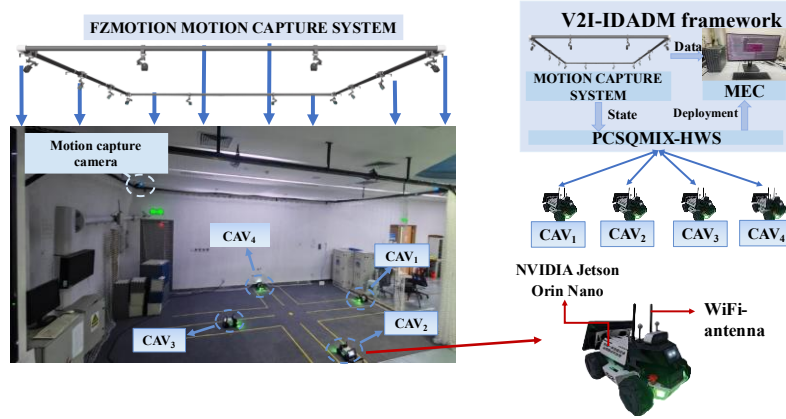


Fig. 22. Development of the miniature intelligent vehicles testing platform.

4.5. Miniature intelligent vehicles testing

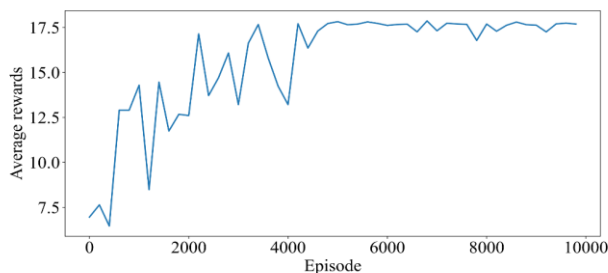
To validate the effectiveness of the proposed method, a miniature intelligent vehicle testing environment is implemented. Four miniature intelligent vehicles are placed at the four directions of the intersection in the simulated experiment. A FZMotion Motion Capture System is positioned above the test area to capture the vehicles' lateral position, longitudinal position, lateral velocity, longitudinal velocity, and heading angle, providing the necessary state information for the vehicles. Each vehicle is equipped with an Nvidia Jetson Orin Nano embedded computer and a Wi-Fi module. To realize the proposed V2I-IDADM framework, a Motion Capture System is employed to emulate roadside sensors, enabling the acquisition of global traffic information at the intersection. All four vehicles are connected within the same local area network (LAN), where each CAV periodically uploads its local state information to a MEC server. The MEC server hosts a PCSQMIX-HWS model. The optimized decision commands are then distributed to each miniature intelligent vehicle through MEC, where the onboard controllers execute them. The overall experimental platform is illustrated in Fig. 22.

A simulation environment is first developed to reproduce the miniature intelligent vehicle scenario, where miniature intelligent vehicles are modeled as scaled representations of real-world vehicles. The kinematics follow the Ackermann steering model, and the road and vehicle parameters are summarized in Table 12.

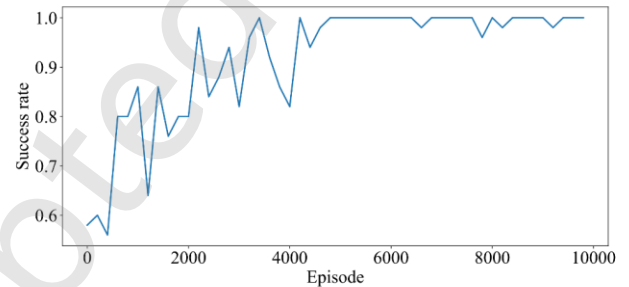
Table 12 Miniature intelligent vehicles platform experimental parameters.

Parameter Name	Unit	Value
Road Width	m	0.5
External Dimensions	mm	322*220*251
Wheelbase	mm	200
Velocity Values	m/s	[0.1,0.2,0.3]
Control Frequency	Hz	20
Decision Frequency	Hz	5

After the training process converges, the evolution of the episode rewards and success rate is presented in Fig. 23. As illustrated, the episode reward gradually stabilizes at approximately 18, and the success rate ultimately reaches 100%, indicating robust and stable learning.



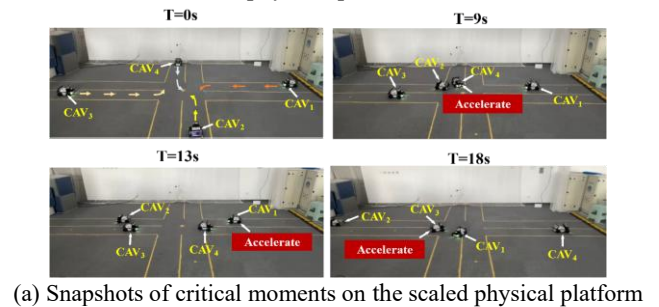
(a) Episode reward in the simulation environment



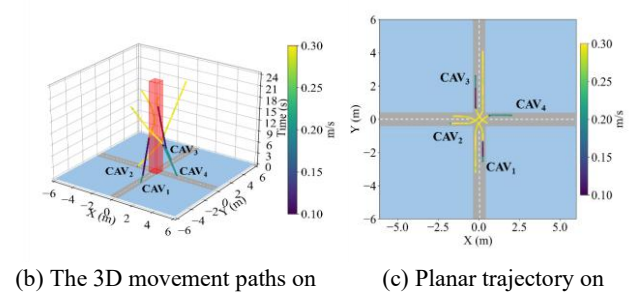
(b) Success rate in the simulation environment

Fig. 23. Training results in the simulation environment.

Finally, all ten experiments conducted on the scaled physical platform are completed successfully, achieving a 100% success rate. Fig. 24 illustrates a representative experimental scenario using the proposed PCSQMIX-HWS model. In this scenario, the initial velocity of all vehicles is set to 0.2 m/s. CAV₂ adopts an acceleration maneuver at the beginning of the experiment and maintains a stable velocity of approximately 0.3 m/s while passing the intersection. CAV₄ initially passes at 0.2 m/s and decelerates at around 8.0 s in response to the presence of CAV₂, resuming acceleration only after CAV₂ has cleared the intersection. Meanwhile, CAV₁ and CAV₃ maintain an initial velocity of 0.2 m/s, subsequently execute deceleration maneuvers to yield, and safely cross the intersection in sequence after CAV₂ and CAV₄ pass. These results demonstrate the application potential of the proposed framework on the scaled physical platform.



(a) Snapshots of critical moments on the scaled physical platform



(b) The 3D movement paths on

(c) Planar trajectory on

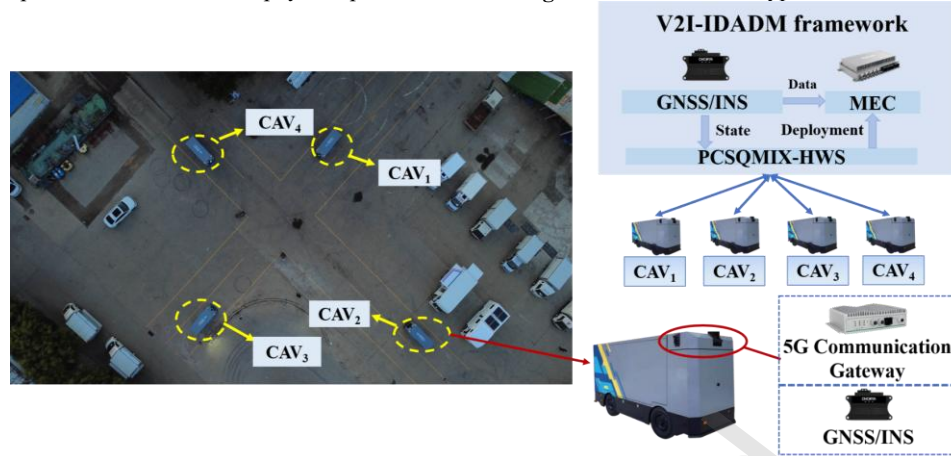
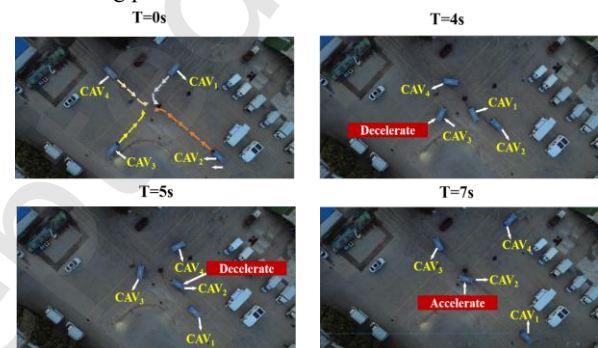


Fig. 25. Development of the real-vehicle testing platform.

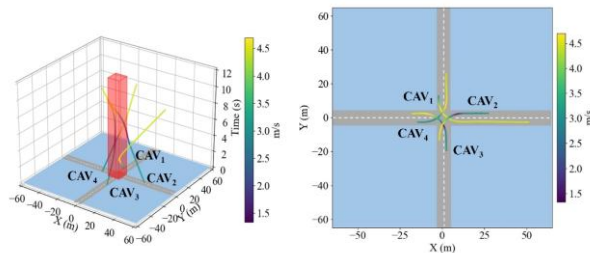
4.6. Real-Vehicle testing

Although the miniature intelligent vehicle tests demonstrate the application potential of the proposed framework, full-scale vehicle experiments are further conducted to validate the practical feasibility of the V2I-IDADM framework. To this end, a real-world experiment is designed in a four-vehicle cooperative unsignalized intersection scenario. In this setup, each CAV is equipped with a Global Navigation Satellite System/Inertial Navigation System (GNSS/INS) integrated navigation unit and a 5G communication gateway, while the RSU is equipped with a 5G communication gateway for V2I communication. The position, velocity, and heading information of each CAV are obtained from the GNSS/INS and transmitted to the MEC server through the 5G communication gateway. The MEC server then computes the cooperative decision-making results using the PCSQMIX-HWS model and sends the corresponding decisions back to the CAVs. The PCSQMIX-HWS model is deployed on a GEACX2 edge computing controller equipped with two NVIDIA Jetson AGX Orin 64-GB modules, providing a total AI computing capability of up to 550 TOPS. As shown in Fig. 25, due to limitations in the experimental environment, the vehicle's motion space is defined as $[1.5, 3.0, 4.5]$ m/s. The frequency of the decision module is defined as 5 Hz, while the frequency of the underlying control module is defined as 50 Hz.

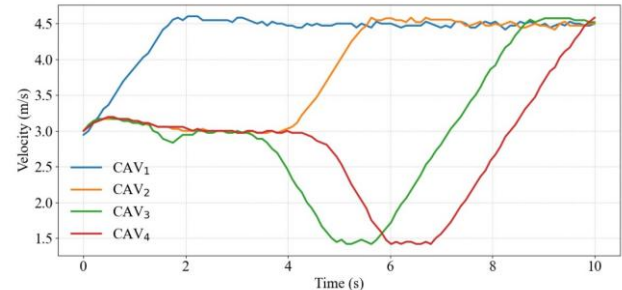
Fig. 26 shows a typical experimental scenario of the proposed PCSQMIX-HWS model. In this scenario, the initial velocity of all vehicles is set to 3.0 m/s. CAV₁ performs an acceleration maneuver at the beginning of the experiment and maintains a stable velocity of 4.5 m/s while passing the intersection. CAV₄ initially passes at 3.0 m/s and accelerates through the intersection after CAV₁ has passed the conflict area. Subsequently, CAV₃ maintains its initial velocity during the early stage. At approximately 4.0 s, after identifying a potential conflict with CAV₄, CAV₃ decelerates and yields until CAV₄ passes the intersection, after which it accelerates and passes through the intersection. Finally, CAV₂ maintains a velocity of 3.0 m/s before decelerating to yield to CAV₃ at the intersection. Once CAV₃ has passed the intersection, CAV₂ accelerates and passes through the intersection. The results demonstrate that the proposed PCSQMIX-HWS model can generate safe and coordinated passing decisions for multiple CAVs at an unsignalized intersection.



(a) Snapshots of critical moments on the real-vehicle platform



(b) The 3D movement paths on the real-vehicle platform (c) Planar trajectory on the real-vehicle platform



(d) Actual operating velocity

Fig. 26. Test results of a typical case on the real-vehicle platform.

To further validate the repeatability of the proposed framework under real-vehicle experimental conditions, five repeated trials are conducted. As shown in Table 13, PCSQMIX-HWS successfully completes the cooperative passing task in all five trials without any collisions, achieving an overall success rate of 100% and an average passing time of 14.12 s. As shown in Fig. 27, all observed PET values remain above the safety threshold of 1.5 s. The mean and minimum PET values are 6.90 and 2.30 s, respectively, resulting in a PET violation rate of 0%. Moreover, to reduce the amplification of measurement noise during numerical

differentiation, the measured velocity signals are lightly smoothed using a Savitzky–Golay filter before calculating acceleration and jerk. The resulting probability density distributions of acceleration and jerk, presented in Figs. 28 and 29, are both strongly concentrated around zero, indicating that vehicle operation is dominated by small acceleration and jerk values. These results further demonstrate that the proposed framework enables safe, smooth, and repeatable multi-vehicle cooperative passing across the repeated real-vehicle experiments.

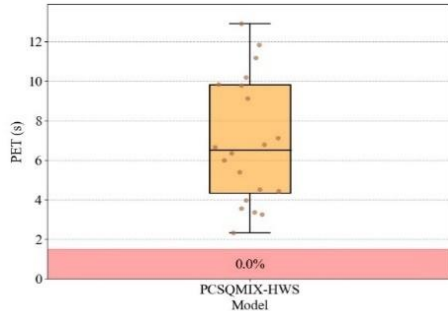


Fig. 27. PET distribution on the real-vehicle platform.

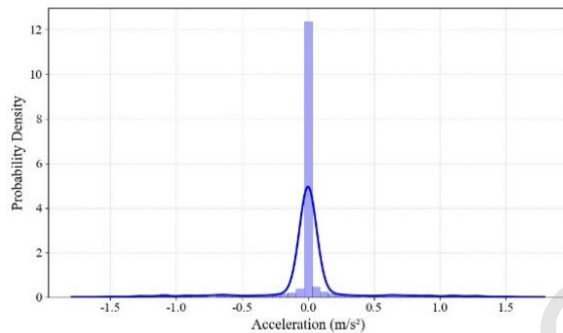


Fig. 28. Acceleration probability density distributions on the real-vehicle platform.

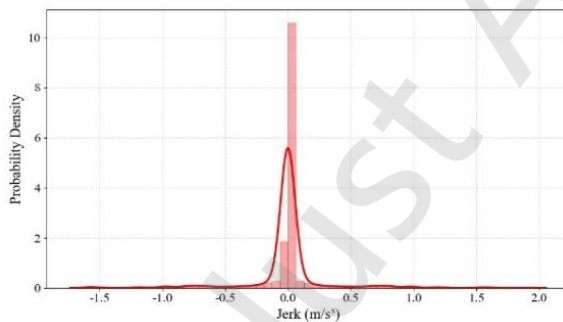


Fig. 29. Jerk probability density distributions on the real-vehicle platform.

Table 13 Experimental Results on the real-vehicle platform.

Model	Success rates	Collision rates	Average passing time(s)	Mean PET (s)	Min PET (s)
PCSQMIX-HWS	100%	0%	14.12	6.9	2.3

To provide a more comprehensive evaluation of the timing performance of the complete decision loop, the key computation and communication metrics are further measured. Specifically, the model inference time is 5.71 ms, the average V2I transmission delay is 12.18 ms, the average round-trip time (RTT) is 30.33 ms, the latency jitter is 3.45 ms, and the packet loss rate is 0. The average end-to-end latency from vehicle-state input to the output of the corresponding decision command is 58.04 ms, with a maximum of 96.84 ms. Both values remain substantially below the 200 ms decision cycle. These results demonstrate that, under the

evaluated full-scale experimental conditions, the proposed cooperative decision-making framework completes an entire state-input-cooperative-decision-command-output loop within each decision cycle, thereby satisfying the real-time decision-making requirement.

5 Conclusion

This study proposes a V2I-IDADM framework for cooperative decision-making at unsignalized intersections, in which roadside infrastructure leverages global perception and computation to enable more reliable and coordinated decisions. Within this infrastructure-centric framework, a cooperative decision-making model is developed under the CTDE paradigm. Heuristic priority rules, explicit safety constraints, and safety-oriented action optimization are jointly incorporated to improve safety and operational efficiency. In addition, a hierarchical weighted sampling strategy classifies episodic samples according to their training significance and assigns category-specific weights through a weighted loss function, thereby improving iterative learning efficiency. Experimental results demonstrate the effectiveness of the proposed framework in both pure-CAV and mixed-traffic scenarios involving CAVs and HVs. Its practical feasibility and deployment potential are further validated using miniature intelligent vehicle and full-scale experimental platforms. Future research will focus on three main directions. First, we will explore integrating multimodal perception models into roadside infrastructure to support semantic reasoning capabilities in complex traffic scenarios. Second, we will deploy the proposed framework in larger-scale scenarios for testing to further validate its real-time performance and scalability in large-scale collaborative decision-making. Third, future work will extend the proposed framework to continuous action spaces to improve its applicability in real-world scenarios.

Author contributions

Sifan Wu: Writing – original draft, Software, Methodology, Conceptualization. **Xuting Duan:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Hao Zhang:** Writing – review & editing, Investigation. **Feiyang Zhao:** Visualization, Investigation, Formal analysis, Conceptualization. **Jianshan Zhou:** Writing – review & editing, Validation, Investigation. **Kaige Qu:** Writing – review & editing, Conceptualization. **Ling Wang:** Investigation. **Daxin Tian:** Writing – review & editing, Supervision, Funding acquisition.

Replication and data sharing

The data and codes used in this study are available at <https://doi.org/10.26599/ETSD.2026.9190084>.

Acknowledgements

This work was supported in part by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM123, the National Natural Science Foundation of China under Grants T2588101, 62432002, 62173012, and U22A2046, the Fundamental Research Funds for the Central Universities (Beihang Ganwei Action Plan Key Program) under Grant JK2024-19, and the Beijing Municipal Science and Technology Program under Grant Z251100003925004.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

References

- Al-Sharman, M., Dempster, R., Daoud, M.A., Nasr, M., Rayside, D., Melek, W., 2023. Self-learned autonomous driving at unsignalized intersections: A hierarchical reinforced learning approach for feasible decision-making. *IEEE Trans. Intell. Transp. Syst.* 24, 12345–12356.
- Boon, M.A., Mandjes, M., 2019. Generalized gap acceptance models for unsignalized intersections. *Math. Methods Oper. Res.*, 89, 385–409.
- Chandra, R., Manocha, D., 2022. Gameplan: Game-theoretic multi-agent planning with human drivers at intersections, roundabouts, and merging. *IEEE Robot. Autom. Lett.* 7, 2676–2683.
- Chen, C., Yang, H., Zhai, C., Chen, X.M., Mo, D., 2024a. Competitive pricing for ride-sourcing platforms with MARL. *Transp. Res. Part C Emerg. Technol.* 165, 104697.
- Chen, K., Li, B., Zhang, R., Cheng, X., 2024b. Autonomous Intersection Management with Heterogeneous Vehicles: A Multi-Agent Reinforcement Learning Approach. In: 2024 IEEE Intelligent Vehicles Symposium (IV), pp. 2255–2260.
- Dresner, K., Stone, P., 2008. A multiagent approach to autonomous intersection management. *J. Artif. Intell. Res.* 31, 591–656.
- el abidine Kherroubi, Z., Aknine, S., Bacha, R., 2021. Novel decision-making strategy for connected and autonomous vehicles in highway on-ramp merging. *IEEE Trans. Intell. Transp. Syst.* 23, 12490–12502.
- George, R., Clancy, J., Brophy, T., Sistu, G., O'Grady, W., Chandra, S., et al. Infrastructure assisted autonomous driving: Research, challenges, and opportunities. *IEEE Open J. Veh. Technol.* 2025, 6, 662–716.
- Guan, J., Chen, G., Huang, J., Li, Z., Xiong, L., Hou, J., et al., 2022. A discrete soft actor-critic decision-making strategy with sample filter for freeway autonomous driving. *IEEE Trans. Veh. Technol.* 72, 2593–2598.
- Guo, Y., Liu, J., Yu, R., Hang, P., Sun, J., 2024. MAPPO-PIS: A multi-agent proximal policy optimization method with prior intent sharing for CAVs' cooperative decision-making. In: European Conference on Computer Vision, pp. 244–263.
- Huang, C., Zhao, J., Zhou, H., Zhang, H., Zhang, X., Ye, C., 2023. Multi-agent decision-making at unsignalized intersections with reinforcement learning from demonstrations. In: 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1–6.
- Huang, S., Sadek, A.W., Zhao, Y., 2012. Assessing the mobility and environmental benefits of reservation-based intelligent intersections using an integrated simulator. *IEEE Trans. Intell. Transp. Syst.* 13, 1201–1214.
- Huang, Z., Sheng, Z., Ma, C., Chen, S., 2024. Human as AI mentor: Enhanced human-in-the-loop reinforcement learning for safe and efficient autonomous driving. *Commun. Transp. Res.* 4, 100127.
- Isele, D., Rahimi, R., Cosgun, A., Subramanian, K., Fujimura, K., 2018. Navigating occluded intersections with autonomous vehicles using deep reinforcement learning. In: 2018 IEEE international conference on robotics and automation (ICRA), pp. 2034–2039.
- Jadoon, M.A., Pastore, A., Navarro, M., Valcarce, A., 2023. Learning random access schemes for massive machine-type communication with MARL. *IEEE Trans. Mach. Learn. Commun. Netw.* 2, 95–109.
- Jeong, Y., 2023. Probabilistic game theory and stochastic model predictive control-based decision making and motion planning in uncontrolled intersections for autonomous driving. *IEEE Trans. Veh. Technol.* 72, 15254–15267.
- Kesting, A., Treiber, M., Helbing, D., 2010. Enhanced intelligent driver model to assess the impact of driving strategies on traffic capacity. *Philos. Trans. R. Soc. A* 368, 4585–4605.
- Kong, J., Pfeiffer, M., Schildbach, G., Borrelli, F., 2015. Kinematic and dynamic vehicle models for autonomous driving control design. In: 2015 IEEE intelligent vehicles symposium (IV), pp. 1094–1099.
- Leurent, E., 2018. An environment for autonomous driving decision-making.
- Li, N., Yao, Y., Kolmanovsky, I., Atkins, E., Girard, A.R., 2020. Game-theoretic modeling of multi-vehicle interactions at uncontrolled intersections. *IEEE Trans. Intell. Transp. Syst.* 23, 1428–1442.
- Li, S., Peng, K., Hui, F., Li, Z., Wei, C., Wang, W., 2024. A decision-making approach for complex unsignalized intersection by deep reinforcement learning. *IEEE Trans. Veh. Technol.* 73, 16134–16147.
- Lin, P., Liu, J., Jin, P.J., Ran, B., 2017. Autonomous vehicle-intersection coordination method in a connected vehicle environment. *IEEE Intell. Transp. Syst. Mag.* 9, 37–47.
- Lin, Z., Lan, J., Anagnostopoulos, C., Tian, Z., Flynn, D., 2025. Safety-critical multi-agent mcts for mixed traffic coordination at unsignalized intersections. *IEEE Trans. Intell. Transp. Syst.* 26, 18992–19006.
- Liu, J., Hang, P., Na, X., Huang, C., Sun, J., 2025a. Cooperative decision-making for cavs at unsignalized intersections: A marl approach with attention and hierarchical game priors. *IEEE Trans. Intell. Transp. Syst.* 26, 443–456.
- Liu, Z., Li, C., Wang, Y., Yang, N., Fan, X., Ma, J., et al., 2024. Multi-scale temporal fusion transformer for incomplete vehicle trajectory prediction. *IEEE Trans. Intell. Veh.* DOI: 10.1109/TIV.2024.3509461.
- Liu, Z., Sun, Y., Wang, Y., Yang, N., Li, S.E., Zhao, X., 2025b. Beyond conventional vision: RGB-event fusion for robust object detection in dynamic traffic scenarios. *Commun. Transp. Res.* 5, 100202.
- Lu, X., Zhao, H., Li, C., Gao, B., Chen, H., 2023. A game-theoretic approach on conflict resolution of autonomous vehicles at unsignalized intersections. *IEEE Trans. Intell. Transp. Syst.* 24, 12535–12548.
- Makridis, M.A., El-Baklish, S.K., Kouvelas, A., Laval, J.A., 2025. The fundamental diagram of autonomous vehicles: Traffic state estimation and evidence from vehicle trajectories. *Commun. Transp. Res.* 5, 100212.
- Mo, S., Pei, X., Wu, C., 2021. Safe reinforcement learning for autonomous vehicle using monte carlo tree search. *IEEE Trans. Intell. Transp. Syst.* 23, 6766–6773.
- Pan, X., Chen, B., Timotheou, S., Evangelou, S.A., 2022. A convex optimal control framework for autonomous vehicle intersection crossing. *IEEE Trans. Intell. Transp. Syst.* 24, 163–177.
- Pei, H., Zhang, Y., Zhang, Y., Feng, S., 2021. Optimal cooperative driving at signal-free intersections with polynomial-time complexity. *IEEE Trans. Intell. Transp. Syst.* 23, 12908–12920.
- Pey, J., Samarakoon, S.B.P., Muthugala, M.V.J., Elara, M.R., 2025. A Decentralized Partially Observable Markov Decision Process for complete coverage onboard multiple shape changing reconfigurable robots. *Expert Syst. Appl.* 271, 126565.
- Rashid, T., Samvelyan, M., Schroeder, C., Farquhar, G., Foerster, J., Whiteson, S., 2018. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Proceedings of the 35th International Conference on Machine Learning*, 4295–4304.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

- Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W.M., Zambaldi, V., Jaderberg, M., et al., 2017. Value-decomposition networks for cooperative multi-agent learning. arXiv preprint arXiv:1706.05296.
- Wu, S., Tian, D., Duan, X., Zhou, J., Zhao, D., Cao, D., 2024. Continuous decision-making in lane changing and overtaking maneuvers for unmanned vehicles: A risk-aware reinforcement learning approach with task decomposition. *IEEE Trans. Intell. Veh.* 9, 4657–4674.
- Xiao, W., Yang, Y., Mu, X., Xie, Y., Tang, X., Cao, D., et al., 2024. Decision-making for autonomous vehicles in random task scenarios at unsignalized intersection using deep reinforcement learning. *IEEE Trans. Veh. Technol.* 73, 7812–7825.
- Xu, Q., Li, K., Wang, J., Yuan, Q., Yang, Y., Chu, W., 2022. The status, challenges, and trends: An interpretation of technology roadmap of intelligent and connected vehicles in China (2020). *J. Intell. Connect. Veh.* 5, 1–7.
- Xu, R., Xiang, H., Han, X., Xia, X., Meng, Z., Chen, C.-J., et al., 2023. The openca open-source ecosystem for cooperative driving automation research. *IEEE Trans. Intell. Veh.* 8, 2698–2711.
- Yang, Y., Yan, R., Miao, J., Zhao, Z., Wang, K., Jiang, K., et al. Vehicle-Infrastructure Cooperative General Object Detection through Feature Flow and Differentiable Pose-based Spatial Alignment. *J. Intell. Connect. Veh.* 2026, <https://doi.org/10.26599/JICV.2026.9210091>.
- Yao, Z., Jiang, H., Jiang, Y., Ran, B., 2023. A two-stage optimization method for schedule and trajectory of CAVs at an isolated autonomous intersection. *IEEE Trans. Intell. Transp. Syst.* 24, 3263–3281.
- Yuan, M., Shan, J., Schofield, H., 2023. Scalable game-theoretic decision-making for self-driving cars at unsignalized intersections. *IEEE Trans. Ind. Electron.* 71, 5920–5930.
- Zhuang, H., Lei, C., Chen, Y., Tan, X., 2023. Cooperative decision-making for mixed traffic at an unsignalized intersection based on multi-agent reinforcement learning. *Appl. Sci.* 13, 5018.
- Zohdy, I.H., Rakha, H.A., 2016. Intersection management via vehicle connectivity: The intersection cooperative adaptive cruise control system concept. *J. Intell. Transp. Syst.* 20, 17–32.

Author biography



intelligent transportation systems.

Sifan Wu received the M.S. degree in computer science from Beijing Union University, Beijing, China, in 2020. He is currently working towards the Ph.D. degree with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include autonomous driving, reinforcement learning, deep learning, artificial intelligence, and



systems, connected and autonomous vehicles.

Xuting Duan received the Ph.D. degree in traffic information engineering and control from Beihang University, Beijing, China. He is currently a Professor with the School of Transportation Science and Engineering, Beihang University. His research interests are focused on the theories of V2X communications and collaborative intelligence and their systematic engineering applications in three-dimensional intelligent transportation



unmanned systems.

Hao Zhang received the M.S. degree in software engineering from Beijing Union University, Beijing, China, in 2025. He is currently pursuing the Ph.D. degree with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include autonomous driving decision-making, multi-agent reinforcement learning, and intelligent



Feiyang Zhao received a B.S. degree in Engineering from Jilin University in 2024. He is pursuing a M.S. degree at the School of Transportation Science and Engineering, Beihang University. His current research interests include multi-agent collaboration, SLAM, and autonomous driving.



Jianshan Zhou received the Ph.D. degree in Traffic Information Engineering and Control from Beihang University, Beijing, China. He is currently an assistant professor at the School of Transportation Science and Engineering, Beihang University. His current research interests include wireless communication, artificial intelligence systems, and intelligent transportation systems.



currently a professor at the School of Transportation Science and Engineering, Beihang University. Her research interests include network slicing, edge intelligence, machine learning for wireless networks, connected autonomous vehicles, and digital twin-assisted network automation.

Kaige Qu received the M.S. degree in integrated circuits engineering and electrical engineering from Tsinghua University, Beijing, China, and KU Leuven, Leuven, Belgium, in 2016, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2021. She is



scheduling.

Ling Wang received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Tsinghua University, Beijing, China, in 1995 and 1999, respectively. Since 1999, he has been affiliated with the Department of Automation at Tsinghua University, where he was promoted to Full Professor in 2008. His research interests include intelligent optimization and production



Daxin Tian (IEEE Fellow) received the Ph.D. degree in computer application technology from Jilin University, Changchun, China, in 2007. He is currently a Professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His research interests include mobile computing, intelligent transportation systems, vehicular ad-hoc networks, and swarm intelligence.

Just Accepted