

Prior knowledge-assisted reinforcement learning for vehicle platoon control in cut-in scenarios

Junru Yang¹, Sifa Zheng^{1,2,✉}, Chuan Sun¹, Haoran Li¹, Lin Xu¹

 Cite this article: Yang J, Zheng S, Sun C, et al. *J Intel Connect Veh* 2026, 9(3): 9210090. <https://doi.org/10.26599/JICV.2026.9210090>

ABSTRACT: In mixed traffic environments, vehicle platoons are inevitably exposed to cut-in maneuvers performed by human-driven vehicles (HDVs), which introduce substantial uncertainty and degrade platoon control performance. To address this challenge, this study proposed a prior knowledge-assisted reinforcement learning (PKARL) control framework for vehicle platoons. A distributed model predictive control (DMPC)-based optimization controller was first developed to provide structured guidance for deep reinforcement learning (DRL) policy learning. On this basis, the state space, action space, reward function, and training scenarios were systematically designed to enable effective responses to HDV cut-in disturbances. Furthermore, a safety filter was incorporated to enforce stability and safety constraints on the control actions generated by the DRL policy. By integrating prior control knowledge, the proposed framework improved training efficiency and enhanced safety compared with conventional DRL methods. Driver-in-the-loop experimental results showed that, relative to the DMPC approach, the proposed PKARL method reduced the cumulative speed error by 23.0% and the cumulative spacing error by 20.1%. In addition, compared with the DRL method, PKARL achieved further reductions of 10.5% in cumulative speed error and 1.1% in cumulative spacing error. These results demonstrate the effectiveness of the proposed method in mitigating HDV cut-in disturbances and improving platoon tracking performance.

KEYWORDS: vehicle platoon control; mixed traffic; distributed model predictive control (DMPC); deep reinforcement learning (DRL); safety filter

1 Introduction

1.1 Motivation

Vehicle platooning has been widely recognized as an effective approach to improving road traffic safety and operational efficiency while also contributing to energy savings and environmental sustainability. Consequently, it has attracted considerable attention in recent years (Pi et al., 2023; Zhan et al., 2023). Nevertheless, in the foreseeable future, road traffic systems are expected to remain in a mixed traffic state, where conventional human-driven vehicles (HDVs) coexist with connected autonomous vehicles (CAVs) for an extended period (Guo and Jia, 2022; Yang et al., 2024). The inherent uncertainty and irregular driving behaviors of HDVs introduce substantial disturbances to platoon operations, making it difficult for vehicle platoons to maintain stable and compact formations (Stefansson et al., 2019).

Among various interaction behaviors, cut-in and cut-out maneuvers performed by HDVs are inevitable in mixed traffic environments. In particular, a cut-in maneuver refers to an HDV merging into the target lane and occupying the gap directly in front of a following vehicle. Owing to the typically short intervehicle spacing within platoons, cut-in events present a

higher safety risk than cut-out maneuvers (Yang et al., 2023; Zhu et al., 2022). Moreover, such maneuvers can significantly degrade the platoon control performance. To avoid potential collisions, the following vehicle often has to respond with abrupt acceleration or braking, which may compromise platoon stability and lead to a decline in overall traffic efficiency (Xu et al., 2021). Existing platoon control strategies, which are primarily designed for fully connected autonomous driving scenarios, exhibit limited robustness when directly applied to mixed traffic conditions, thereby posing challenges to both safety and stability. Consequently, there is a pressing need to develop advanced platoon control approaches that can effectively mitigate the adverse effects of HDV cut-in behaviors while ensuring safe and stable platoon operation.

1.2 Literature review

Seminal research and real-world experiments regarding vehicle cut-in scenarios have been conducted by the energy intelligent transport systems (ITS) project in Japan (Omae et al., 2014) and the California Partners for Advanced Transportation Technology (PATH) Program in USA (Milanés and Shladover, 2016). These pioneering studies established a fundamental response protocol: If the intruding vehicle is equipped with vehicle-to-vehicle (V2V)

¹ Suzhou Automotive Research Institute, Tsinghua University, Suzhou 215134, China. ² School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China.

✉ Corresponding author. E-mail: zsf@tsinghua.edu.cn

Received: December 25, 2025; Revised: June 11, 2026; Accepted: July 20, 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/>).

Nomenclature

ACC	Adaptive cruise control
CACC	Cooperative adaptive cruise control
CAV	Connected autonomous vehicle
CBF	Control barrier function
CLF	Control Lyapunov function
DDPG	Deep deterministic policy gradient
DMPC	Distributed model predictive control
DRL	Deep reinforcement learning
FH-DDPG-SS	Finite-horizon-DDPG with sweeping through reduced state space using stationary approximation
HDV	Human-driven vehicle
PKARL	Prior knowledge-assisted reinforcement learning
QP	Quadratic program
TD3	Twin-delayed deep deterministic policy gradient

communication, it is designated as the new leader for the trailing members; otherwise, the control logic of the following vehicle degrades from cooperative adaptive cruise control (CACC) to standard adaptive cruise control (ACC). Building on these protocols, [Tu et al. \(2019\)](#) employed PATH models to quantitatively assess the longitudinal safety risks associated with such control degradation in ten-vehicle CACC platoons. More recently, [Lu et al. \(2023\)](#) proposed an intention-driven control approach that employs a finite state machine to alternate between a preventive mode for discretionary lane changes and a yielding mode for mandatory cut-ins, aiming to preserve platoon cohesion while maintaining traffic safety.

Beyond platoon-specific research, extensive studies have addressed cut-in maneuvers for single vehicles equipped with advanced driver assistance systems. [Zhang et al. \(2021\)](#) developed a lateral behavior prediction model leveraging historical driving data to formulate a novel range policy. Similarly, [Ma et al. \(2020\)](#) proposed a hybrid method that integrates fuzzy support vector machines with a sliding window detection technique to evaluate the likelihood of neighboring vehicles executing a merge. [Lyu et al. \(2022\)](#) developed a trajectory forecasting framework for V2V scenarios, employing long short-term memory networks in conjunction with a collision alert system. Furthermore, [Zhang et al. \(2024\)](#) proposed a refined risk assessment metric known as variable time-to-collision, which utilizes elliptical vehicle geometry to accurately estimate collision risks from multiple directions.

It is important to note that control objectives for platoons differ fundamentally from those of single vehicles due to the requirement for string stability. Consequently, distinct controller designs are necessary. [Lu et al. \(2022\)](#) incorporated both lateral and longitudinal control within a queuing network-based cognitive architecture to simulate cut-in maneuvers in vehicle platoons. Model predictive control (MPC) has been widely adopted in this domain; for instance, [Kazemi et al. \(2018\)](#) designed a stochastic MPC that incorporates cut-in probabilities to improve responsiveness to hazardous maneuvers. For distributed systems, [Liu et al. \(2019\)](#) proposed a two-step noniterative distributed MPC (DMPC) strategy for solving optimal control problems across vehicle clusters, while [Basiri et al. \(2020\)](#) developed a nonlinear DMPC approach to manage heterogeneous platoons under unidirectional topologies. Addressing disturbance propagation, [Liu et al. \(2022\)](#) introduced a feedforward variable acceleration-limited control algorithm. Conversely, [An and Talebpour \(2023\)](#) proposed a unique strategy focused on the leading vehicle, which accelerates the proactively

creation of gaps for lane-changing vehicles, thereby minimizing downstream disturbances.

Recently, deep reinforcement learning (DRL) has gained prominence as a data-driven approach for autonomous vehicle decision-making. Unlike traditional methods that depend on physical models, DRL learns optimal strategies through environmental interaction, making it particularly adept at addressing uncertain control issues. For example, [Chen et al. \(2022\)](#) developed a deep Q-learning framework incorporating experience screening to control vehicle speed under uncertain cut-in scenarios. [Selvaraj et al. \(2023\)](#) designed a hierarchical DRL architecture to jointly enhance traffic efficiency, safety, and ride comfort. Additionally, [Li et al. \(2024\)](#) introduced an online learning-based predictive control structure to enhance trajectory tracking and cooperative efficiency for CAVs.

Despite these advancements, significant challenges remain. Most existing methods rely heavily on predicting HDV intentions to trigger preemptive braking. However, due to the stochastic nature of human drivers, prediction errors can lead to unnecessary deceleration or safety hazards. Furthermore, while current DRL approaches circumvent the difficulties of modeling complex human-machine interactions, they are typically model-free and data-hungry. This results in sample inefficiency and protracted training periods, as they fail to leverage system a priori knowledge. More critically, current DRL lacks theoretical safety guarantees; relying solely on reward function penalties does not impose hard constraints on control outputs, leaving the system vulnerable during exploration.

To address the challenge of enhancing platoon safety and tracking performance in HDV cut-in scenarios, it is crucial to improve the DRL method by incorporating prior knowledge. This study proposes a platoon control method termed prior knowledge-assisted reinforcement learning (PKARL), which integrates prior knowledge into both the decision-making process and safety mechanisms of the DRL framework. Specifically, a DMPC controller is employed to guide the DRL algorithm during decision-making, leveraging prior knowledge to improve convergence speed and generalization capability. Additionally, the safety filter, informed by prior knowledge, enforces stability and safety constraints on the DRL outputs. By integrating DMPC guidance and safety filtering into the DRL training process, this method effectively enhances learning efficiency, mitigates unsafe state exploration, and ensures reliable platoon control in HDV cut-in scenarios.

1.3 Contribution

The primary objective of this study is to develop a robust and safety-critical platoon control paradigm that tightly integrates the reliability of model-based control with the adaptability of data-driven reinforcement learning. In this framework, prior knowledge is mainly introduced from two aspects: The DMPC-based optimization controller provides structured decision guidance for the DRL policy, while the safety filter performs real-time correction of the DRL outputs according to safety and stability constraints. In this way, the proposed method improves the learning efficiency and operational safety of conventional DRL in stochastic HDV cut-in scenarios. To the best of our knowledge, this work represents the first systematic attempt to introduce a PKARL framework for safe and stable vehicle platoon control under mixed traffic conditions involving stochastic HDV cut-in disturbances. The main contributions of this study are summarized as follows:

- 1) We develop a cut-in-focused learning and control

formulation for platoon operation in mixed traffic. The approach is specifically designed for HDV cut-in scenarios by jointly structuring disturbance-aware state variables, control actions, training scenario construction with domain randomization, and cut-in–recovery performance metrics. This integrated design enables the learning-based controller to proactively respond to HDV cut-in disturbances while maintaining tracking performance and safety.

2) We propose a novel PKARL framework that bridges the gap between model-based control and data-driven learning. This architecture introduces a guidance–exploration–constraint mechanism, in which a DMPC-based optimization controller provides structured control guidance to the DRL policy during both training and online execution, thereby improving learning efficiency and online decision quality. Meanwhile, a physics-informed safety filter operates in both phases to correct potentially unsafe actions and enforce stability and safety constraints.

1.4 Organization

The remainder of this paper is organized as follows: Section 2 formulates the problem and presents the overall control architecture. Section 3 develops the PKARL-based platoon controller, including algorithm design, training scenarios, and hyperparameter settings. Section 4 describes the driver-in-the-loop experimental setup and reports the corresponding performance analysis. Section 5 concludes the paper and discusses directions for future research.

2 Problem formulation and control framework

This section delineates the dynamic impact of HDV cut-in and cut-out maneuvers on vehicle platooning and introduces the structural architecture of the proposed PKARL control framework.

2.1 Problem formulation

This study focuses on a representative mixed traffic scenario near highway ramp exits, where an HDV performs sequential lane-change maneuvers through a vehicle platoon. Typically, as a platoon travels in the rightmost lane, its longitudinal occupancy may impede HDVs in adjacent lanes attempting to access the exit ramp. Due to the relatively larger gaps within the platoon compared to dense traffic, an HDV may execute a strategic sequence: initially cutting into the platoon, maintaining a temporary position within the formation for a certain distance, and finally cutting out to reach the exit ramp. This intricate interaction process introduces transient disturbances that challenge the robustness of the platoon control system.

As shown in Fig. 1, a platoon comprising $N + 1$ CAVs travels in the right lane, where the leader is indexed as 0 and the subsequent vehicles are indexed from 1 to N . The platoon leverages V2X communication to acquire the real-time motion states of surrounding HDVs. At time $t = t_0$, an HDV from the adjacent left lane initiates a lateral maneuver, merging into the platoon between CAV- $i-1$ and CAV- i . This intrusion necessitates

that CAV- i immediately modulates its speed to ensure a longitudinal safety margin, thereby triggering a ripple effect of speed and spacing adjustments among all downstream vehicles. At time $t = t_i$, the HDV departs the platoon to the right. This maneuver results in an instantaneous expansion of the intervehicle headway between CAV- $i-1$ and CAV- i . To restore the cohesive formation, CAV- i and its followers must accelerate to close the gap until the platoon converges to a new equilibrium state. Such abrupt transitions in intervehicle spacing not only compromise string stability but also escalate collision risks, posing a significant challenge for maintaining seamless and safe platoon operations in mixed traffic.

2.2 Control framework

The vehicle platoon adopts a predecessor-follower (PF) communication topology, in which each vehicle receives real-time acceleration information from its immediate predecessor. In addition, the relative distance and relative speed are obtained through on-board sensing devices. Based on this information, the control system regulates the longitudinal acceleration of each vehicle to ensure accurate intervehicle spacing regulation and to suppress relative speed fluctuations with respect to the preceding vehicle.

The proposed PKARL framework, illustrated in Fig. 2, integrates model-based optimization with data-driven adaptation. Within this architecture, a DMPC module leverages system dynamics to execute receding horizon optimization, generating a reference control output $u_{i,m}^*(k)$ based on the current state $\mathbf{x}_i(k)$. This output $u_{i,m}^*(k)$ is used during both training and online execution. In both phases, it is fed into the DRL agent as an augmented state input to provide structured control guidance, while during training, it is also incorporated into the reward function to guide policy learning. This mechanism provides structured guidance for DRL policy, effectively narrowing the search space and accelerating convergence in complex cut-in scenarios.

To ensure rigorous operational bounds, the control action $u_{i,d}(k)$ generated by the DRL agent is processed through a safety filter. This filter, grounded in control barrier functions (CBFs) and control Lyapunov functions (CLFs), solves a quadratic programming problem in real time to adjust $u_{i,d}(k)$ into the final control command $u_i^*(k)$. This procedure enforces stability and safety constraints on the control input, ensuring safe platoon operation during both learning and execution.

The PKARL framework adopts a hierarchical, distributed control architecture across platoon vehicles. The proposed algorithm functions as an upper-level decision-making module that generates desired longitudinal acceleration commands. These commands are executed by a lower-level controller, which converts the acceleration references into actuator inputs via an inverse longitudinal dynamics model.

3 Platoon control with PKARL

This section elucidates the proposed PKARL method in detail.

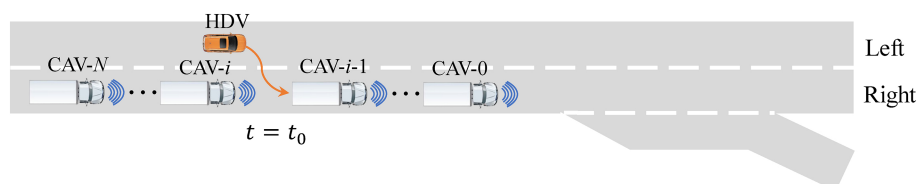


Fig. 1 HDV cut-in and cut-out platoon scenarios.

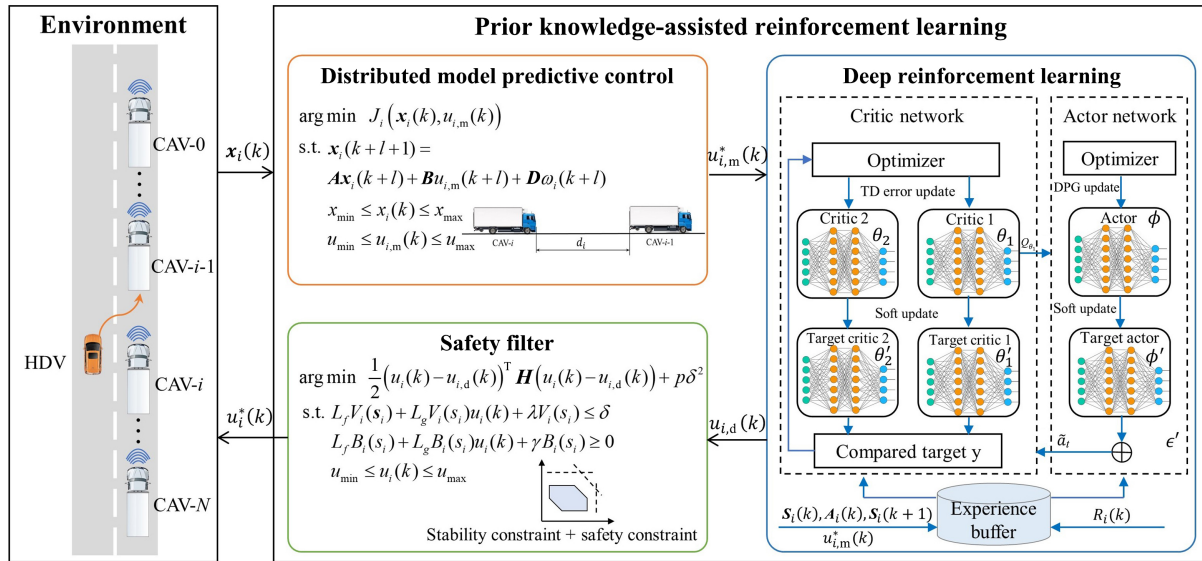


Fig. 2 Framework of the proposed PKARL method.

The PKARL framework consists of three key components: DMPC, DRL, and a safety filter. DMPC utilizes a priori information to generate outputs that guide the DRL decision-making process, providing a structured foundation for learning and improving convergence speed. The DRL component integrates prior knowledge into the entire learning and decision-making pipeline, including the definition of state and action representations, reward function design, and scenario construction, enabling it to effectively handle the uncertainties of HDV cut-in maneuvers. Meanwhile, the safety filter, also informed by established knowledge, optimizes and dynamically adjusts the DRL outputs during real-time control to enforce system safety and stability constraints.

3.1 DMPC

The DMPC module serves as the primary source of prior control knowledge. With a constant time headway policy, the desired intervehicle distance is expressed as

$$d_i^*(t) = t_d v_i(t) + L \quad (1)$$

where $d_i^*(t)$ and $v_i(t)$ represent the desired intervehicle distance and speed for CAV- i , respectively; t_d is the preset constant time headway; and L denotes the safety distance.

The tracking errors for spacing $\Delta d_i(t)$ and speed $\Delta v_i(t)$ are defined as

$$\begin{cases} \Delta d_i(t) = d_i(t) - d_i^*(t) \\ \Delta v_i(t) = v_{i-1}(t) - v_i(t) \end{cases} \quad (2)$$

where $d_i(t)$ is the actual intervehicle distance; and $v_{i-1}(t)$ and $v_i(t)$ represent the speeds of the preceding vehicle and the ego vehicle, respectively.

The state vector is defined as $\mathbf{x}_i(t) = [\Delta d_i(t) \ \Delta v_i(t) \ a_i(t)]^T$. By incorporating longitudinal vehicle dynamics and the acceleration of the preceding vehicle, a discrete-time prediction model is obtained using a zero-order hold with sampling period T_s , where k denotes the discrete-time index:

$$\mathbf{x}_i(k+1) = \mathbf{A}\mathbf{x}_i(k) + \mathbf{B}u_{i,m}(k) + \mathbf{D}w_i(k) \quad (3)$$

where the system matrices are given by

$$\mathbf{A} = \begin{bmatrix} 1 & T_s & -t_d T_s \\ 0 & 1 & -T_s \\ 0 & 0 & 1 - \frac{T_s}{\tau_i} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ \frac{T_s}{\tau_i} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 \\ T_s \\ 0 \end{bmatrix} \quad (4)$$

where $u_{i,m}(k)$ and $a_i(k)$ denote the desired and actual accelerations, respectively, while $w_i(k) = a_{i-1}(k)$ represents the disturbance introduced by the acceleration of the preceding vehicle, and τ_i denotes the actuation time lag.

For the scenarios where HDV cut-in occurs within the platoon, the DMPC objective is to minimize tracking errors while ensuring comfort and safety. Therefore, the cost function for the i -th following CAV is defined as

$$J_i(\mathbf{x}_i(k), u_{i,m}(k)) = \sum_{l=0}^{N_p-1} \left(\mathbf{x}_i^T(k+l) Q_x \mathbf{x}_i(k+l) + u_{i,m}^T(k+l) R_u u_{i,m}(k+l) \right) + \mathbf{x}_i^T(k+N_p) P_N \mathbf{x}_i(k+N_p) \quad (5)$$

where N_p denotes the prediction horizon; and Q_x , R_u , and P_N denote the weight matrices of the predicted state, inputs, and the terminal state, respectively. The terminal weight P_N is obtained by solving the discrete-time algebraic Riccati equation (González et al., 2020).

The longitudinal control optimization problem for the i -th following CAV can be described as

$$u_{i,m}^*(k) = \operatorname{argmin} J_i(\mathbf{x}_i(k), u_{i,m}(k)) \quad (6a)$$

$$\text{s.t. } \mathbf{x}_i(k+l+1) = \mathbf{A}\mathbf{x}_i(k+l) + \mathbf{B}u_{i,m}(k+l) + \mathbf{D}w_i(k+l) \quad (6b)$$

$$\mathbf{x}_{\min} \leq \mathbf{x}_i(k+l) \leq \mathbf{x}_{\max} \quad (6c)$$

$$u_{\min} \leq u_{i,m}(k+l) \leq u_{\max} \quad (6d)$$

$$\forall i \in \{1, 2, \dots, N\} \quad (6e)$$

$$\forall l \in \{0, 1, \dots, N_p - 1\} \quad (6f)$$

At each sampling instant, the DMPC optimization problem is solved subject to system dynamics and state and input constraints. The first element of the optimal control sequence $u_{i,m}^*(k)$ is applied as the reference control input and provided to the DRL component. The DMPC parameters are summarized in Table 1.

Remark 1. The sampling period $T_s = 0.05$ s is selected to match the real-time longitudinal control requirement and to capture the transient response caused by HDV cut-in disturbances. The prediction horizon $N_p = 20$ provides a 1 s prediction window, which is consistent with the adopted constant time headway $t_d = 1$ s and is sufficient for short-term spacing regulation. The control bounds $u_{\min} = -3$ m/s² and $u_{\max} = 3$ m/s² are used to reflect feasible acceleration and braking limits while avoiding overly aggressive control actions. The weighting matrices are selected to penalize the spacing error, speed error, acceleration fluctuation, and control effort in a balanced manner.

Although the DMPC model and parameters are predefined, the proposed PKARL method does not simply apply the DMPC output as the final control command. Instead, the DMPC output is used as prior guidance for DRL, while the DRL policy learns adaptive control behavior from randomized cut-in scenarios, and the safety filter further corrects unsafe actions in real time. Therefore, moderate modeling errors or parameter variations in DMPC can be partially compensated by the learning and safety-filtering components, which supports the generalization of the proposed method in similar HDV cut-in scenarios.

3.2 DRL

The twin delayed deep deterministic policy gradient (TD3) algorithm is adopted as the learning-based controller due to its suitability for continuous action spaces and its ability to mitigate overestimation bias in value function approximation (Fujimoto et al., 2018). The effectiveness of the learning process is governed by the meticulous design of the state space, action space, and reward function.

1) State space

In the event of an HDV cutting into the platoon, it is crucial to make informed decisions based on the states of both the ego vehicle and the HDV. Furthermore, to leverage the prior knowledge provided by the DMPC controller, the output $u_{i,m}^*(k)$ is incorporated as one of the elements in the DRL state space.

$$\mathbf{S}_i(k) = [\Delta d_i(k), \Delta v_i(k), v_i(k), a_i(k), a_{i-1}(k), u_{i,m}^*(k)] \quad (7)$$

2) Action space

The action space $\mathbf{A}_i(k)$ is the output $u_{i,d}(k)$ of the DRL, representing the desired acceleration of the vehicle, which will be used as the safety filter input for the i -th following CAV:

$$\mathbf{A}_i(k) = [u_{i,d}(k)] \quad (8)$$

3) Reward function

The reward function is a critical component in evaluating an agent's performance within the environment. It must precisely encapsulate the task's objectives and provide robust feedback to steer the agent toward learning an optimal strategy.

In designing the reward function, several key considerations are

addressed. First, following an HDV cut-in or cut-out event, the ego vehicle must minimize the spacing error $\Delta d_i(k)$ and speed error $\Delta v_i(k)$ relative to the preceding vehicle to achieve platoon control. Second, penalizing the rate of change of acceleration (jerk) and the control output $u_{i,d}(k)$ is essential to ensure passenger comfort and prevent increased fuel consumption due to frequent acceleration and deceleration. Third, to reflect the guidance of DMPC on DRL, it is desirable for $u_{i,d}(k)$ to closely align with $u_{i,m}^*(k)$. Finally, normalization is necessary to account for the differing units of the various variables involved. The reward function is defined as

$$R_i(k) = - \left[\omega_1 \left| \frac{\Delta d_i(k)}{\Delta d_{\max}} \right| + \omega_2 \left| \frac{\Delta v_i(k)}{\Delta v_{\max}} \right| + \omega_3 \left| \frac{\Delta u_{i,d}(k)}{u_{\max}} \right| + \omega_4 \left| \frac{\text{jerk}_i(k)}{2a_{\max}/T_s} \right| + \omega_5 \left| \frac{u_{i,d}(k) - u_{i,m}^*(k)}{u_{\max}} \right| \right] \quad (9)$$

where the jerk is computed as

$$\text{jerk}_i(k) = \frac{a_i(k) - a_i(k-1)}{T_s} \quad (10)$$

where ω_1 – ω_5 denote the weighting coefficients of the reward function; $\Delta d_{\max}$ and $\Delta v_{\max}$ represent the upper bounds of the spacing error and speed error, respectively; $a_{\max}$ and $u_{\max}$ denote the maximum allowable acceleration and control input, respectively; and $|\cdot|$ denotes the absolute value. The parameter values for the reward function are summarized in Table 2.

Remark 2. The terms in Eq. (9) are normalized by their corresponding upper bounds to reduce the influence of different physical units and numerical scales. The weighting coefficients are selected according to the control priorities in HDV cut-in scenarios and are iteratively adjusted through repeated training and simulation tests. Specifically, $\omega_1 = 1$ is assigned to the spacing-error term as the reference weight because maintaining a safe intervehicle distance is the primary control objective. The speed-error weight $\omega_2 = 0.1$ is selected to improve tracking performance and disturbance attenuation. The control-input weight $\omega_3 = 0.1$ and jerk weight $\omega_4 = 0.2$ are used to suppress excessive acceleration and rapid control variations, thereby improving control smoothness and passenger comfort. The DMPC-guidance weight is set to a relatively small value, $\omega_5 = 0.025$, so that the DMPC output provides physically reasonable guidance without dominating the DRL policy or restricting its exploration capability. The final values provide a balance among spacing accuracy, speed tracking, control effort, passenger comfort, and policy adaptation.

3.3 Safety filter

To guarantee real-time safety, a safety filter based on the CLF and CBF is employed (Ames et al., 2019). For the i -th following CAV, its car-following model can be expressed in the form of a nonlinear affine control system:

Table 2 DRL reward function parameters

Parameter	Value	Parameter	Value
ω_1	1	$\Delta d_{\max}$	15 m
ω_2	0.1	$\Delta v_{\max}$	15 m/s
ω_3	0.1	$a_{\max}$	3 m/s ²
ω_4	0.2	$u_{\max}$	3 m/s ²
ω_5	0.025	T_s	0.05 s

Table 1 DMPC parameters

Parameter	Value	Parameter	Value
T_s	0.05 s	R_u	0.5
N_p	20	$u_{\min}$	−3 m/s ²
Q_x	Diag (1, 1, 1)	$u_{\max}$	3 m/s ²

$$\dot{\mathbf{s}}_i(k) = \begin{bmatrix} \dot{p}_i(k) \\ \dot{v}_i(k) \\ \dot{d}_i(k) \end{bmatrix} = \underbrace{\begin{bmatrix} v_i(k) \\ 0 \\ v_{i-1}(k) - v_i(k) \end{bmatrix}}_{f(\mathbf{s}_i)} + \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{g(\mathbf{s}_i)} u_i(k) \quad (11)$$

where $\mathbf{s}_i(k) = [p_i(k) \ v_i(k) \ d_i(k)]^T$ represents the vehicle's position $p_i(k)$, speed $v_i(k)$, and intervehicle spacing $d_i(k)$; $u_i(k) = a_i(k)$ is the control input, i.e., the vehicle's acceleration; $f(\mathbf{s}_i)$ denotes the state-related function; and $g(\mathbf{s}_i)$ denotes the input-related function. Both $f(\mathbf{s}_i)$ and $g(\mathbf{s}_i)$ are local Lipschitz functions.

For the platooning system, one of the control objectives is to maintain the same speed as the preceding vehicle. The Lyapunov function is designed as

$$V_i(\mathbf{s}_i) = (v_i(k) - v_{i-1}(k))^2 \quad (12)$$

where $v_{i-1}(k)$ is the speed of the preceding vehicle. According to Ames et al. (2017), for any $\lambda > 0$, the following inequality constraint holds:

$$\inf_{u_i \in U_i} [L_f V_i(\mathbf{s}_i) + L_g V_i(\mathbf{s}_i) u_i(k)] + \lambda V_i(\mathbf{s}_i) \leq 0 \quad (13)$$

The satisfaction of Eq. (13) ensures that $V_i(\mathbf{s}_i)$ decreases exponentially, thereby guaranteeing exponential convergence of the relative-speed error. Here, $L_f V_i(\mathbf{s}_i) = \nabla V_i(\mathbf{s}_i) \cdot f(\mathbf{s}_i)$ and $L_g V_i(\mathbf{s}_i) = \nabla V_i(\mathbf{s}_i) \cdot g(\mathbf{s}_i)$. We obtain:

$$\nabla V_i(\mathbf{s}_i) = [0 \ 2(v_i(k) - v_{i-1}(k)) \ 0] \quad (14)$$

The CBF ensures safety by maintaining the system state within a safe invariant set using a Lyapunov-like function. To avoid collisions and ensure safe intervehicle distances, the CBF is designed based on the braking distance model:

$$B_i(\mathbf{s}_i) = d_i(k) - \frac{1}{2} \cdot \frac{(v_{i-1}(k) - v_i(k))^2}{u_{\min}} - L \quad (15)$$

where L is the safety distance and $u_{\min}$ represents the vehicle's lower bound on acceleration. Then,

$$\nabla B_i(\mathbf{s}_i) = \begin{bmatrix} 0 & -\frac{v_i(k) - v_{i-1}(k)}{u_{\min}} & 0 \end{bmatrix} \quad (16)$$

We have

$$\sup_{u_i \in U_i} [L_f B_i(\mathbf{s}_i) + L_g B_i(\mathbf{s}_i) u_i(k)] + \gamma B_i(\mathbf{s}_i) \geq 0 \quad (17)$$

where $L_f B_i(\mathbf{s}_i) = \nabla B_i(\mathbf{s}_i) \cdot f(\mathbf{s}_i)$ and $L_g B_i(\mathbf{s}_i) = \nabla B_i(\mathbf{s}_i) \cdot g(\mathbf{s}_i)$, with a constant $\gamma > 0$.

By integrating CLF and CBF constraints, the safety filter is formulated as a quadratic program (QP) problem:

$$u_i^*(k) = \operatorname{argmin} \frac{1}{2} (u_i(k) - u_{i,d}(k))^T \mathbf{H} (u_i(k) - u_{i,d}(k)) + p \delta^2 \quad (18a)$$

$$\text{s.t. } L_f V_i(\mathbf{s}_i) + L_g V_i(\mathbf{s}_i) u_i + \lambda V_i(\mathbf{s}_i) \leq \delta \quad (18b)$$

$$L_f B_i(\mathbf{s}_i) + L_g B_i(\mathbf{s}_i) u_i + \gamma B_i(\mathbf{s}_i) \geq 0 \quad (18c)$$

$$u_{\min} \leq u_i(k) \leq u_{\max} \quad (18d)$$

$$\forall i \in \{1, 2, \dots, N\} \quad (18e)$$

where $u_{i,d}(k)$ represents the output of the DRL algorithm, and $u_i^*(k)$ denotes the output of the safety filter, specifically the desired acceleration applied to the controlled vehicle. $\mathbf{H}$ is the positive definite matrix; $p > 0$ is the penalization factor; and δ is the slack variable.

The safety-filtering mechanism can be interpreted as a minimum-deviation projection of the TD3 output onto the feasible control set defined by the CLF, CBF, and actuator constraints. Specifically, $u_{i,d}(k)$ is the nominal acceleration generated by the TD3 policy, whereas $u_i^*(k)$ is the acceleration command finally applied to the controlled vehicle. If $u_{i,d}(k)$ satisfies all constraints in Eq. (18), the safety filter leaves the TD3 output unchanged, i.e., $u_i^*(k) = u_{i,d}(k)$. Otherwise, the QP modifies the nominal action by selecting the closest feasible acceleration command. The CBF constraint is imposed as a hard constraint to maintain the longitudinal safe set, while the slack variable δ is introduced only into the CLF constraint. Therefore, speed-tracking performance can be temporarily relaxed when it conflicts with longitudinal safety. Provided that $B_i(\mathbf{s}_i) \geq 0$ and the QP remains feasible, the CBF condition ensures the forward invariance of the safe set under the adopted longitudinal vehicle model.

The real-time feasibility of QP-based safety filters constructed with CLF/CBF constraints has been supported in autonomous driving and learning-based control applications. For example, CLF-CBF-QP has demonstrated real-time feasibility in safe motion control with moving obstacles (Desai and Ghaffari, 2022). In addition, CBF-QP has been used to correct unsafe actions generated by reinforcement learning controllers and can be solved efficiently at real-time control frequencies (Hou et al., 2025).

In the proposed distributed framework, the safety filter is solved independently for each following CAV, and the QP problem only involves the longitudinal control input $u_i(k)$ and the slack variable δ . Therefore, the dimension of each local QP remains unchanged as the platoon size increases, while the overall computational load increases approximately linearly with the number of following vehicles. This low-dimensional formulation makes it suitable for real-time platoon control.

Remark 3. The terms $v_{i-1}(k)$ and $a_{i-1}(k)$ denote the speed and acceleration of the preceding vehicle, respectively. Under normal platooning conditions, the predecessor is the $(i-1)$ -th CAV. When an HDV executes a cut-in maneuver, the HDV is treated as the preceding vehicle for the affected CAV.

3.4 Training scenario and hyperparameter settings

MATLAB/Simulink serves as the platform for algorithm training, with the TD3 algorithm implemented using the reinforcement learning toolbox. During each control period, the quadratic programming problem of DMPC and the safety filter is solved using MATLAB's quadprog function. Initially, the leading vehicle and ego vehicle scenario is examined. Upon achieving convergence and stability in algorithm training, the learned strategies are extended to multivehicle cooperative driving scenarios. In this context, the vehicle model is treated as a first-order inertial system as described in Eq. (3). The ego vehicle, a CAV, is controlled based on the PKARL algorithm. The ego vehicle's tracking target switches from the leading vehicle to the HDV when the HDV crosses the lane line.

The training scenario is illustrated in Fig. 3, where the time headway t_d and the safety distance L are set to 1 s and 5 m, respectively. The ego vehicle is initialized at a longitudinal position

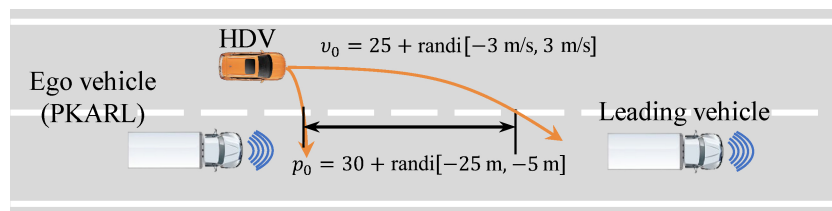


Fig. 3 Training scenario setup.

of 0 m with an initial speed of 25 m/s. To emulate HDV cut-in maneuvers occurring at different positions and speeds ahead of the ego vehicle, the initial position of the HDV is defined as a nominal value with random perturbation, i.e., $p_0 = 30 + \text{randi}[-25, -5]$. Considering that longitudinal speed variations during cut-in maneuvers are typically limited to within 10% of the nominal speed, the variation range for the initial speed is set narrower than that for the position. Accordingly, the initial speed of the HDV is $v_0 = 25 + \text{randi}[-3 \text{ m/s}, 3 \text{ m/s}]$.

With these settings, the initial HDV position varies from 5 to 25 m ahead of the ego vehicle, and the initial HDV speed varies from 22 to 28 m/s. These ranges generate cut-in cases with different relative distances and relative speeds, covering both close-

distance and moderate-distance disturbances. Therefore, the agent is not trained under a fixed cut-in condition but learns to respond to diverse HDV behaviors, which improves the robustness and generalization capability of the learned policy. The selected ranges are consistent with typical highway cut-in conditions and are reasonable for training scenario construction.

The actor and critic networks of the TD3 algorithm are shown in Fig. 4. Both networks adopt fully connected structures. The actor network comprises four hidden layers with 48, 48, 48, and 1 neuron. The input layer dimension matches the number of state variables (6 neurons). Rectified linear unit (ReLU) activation functions are applied to all hidden layers, while the output layer uses a hyperbolic tangent (Tanh) activation to map actions to

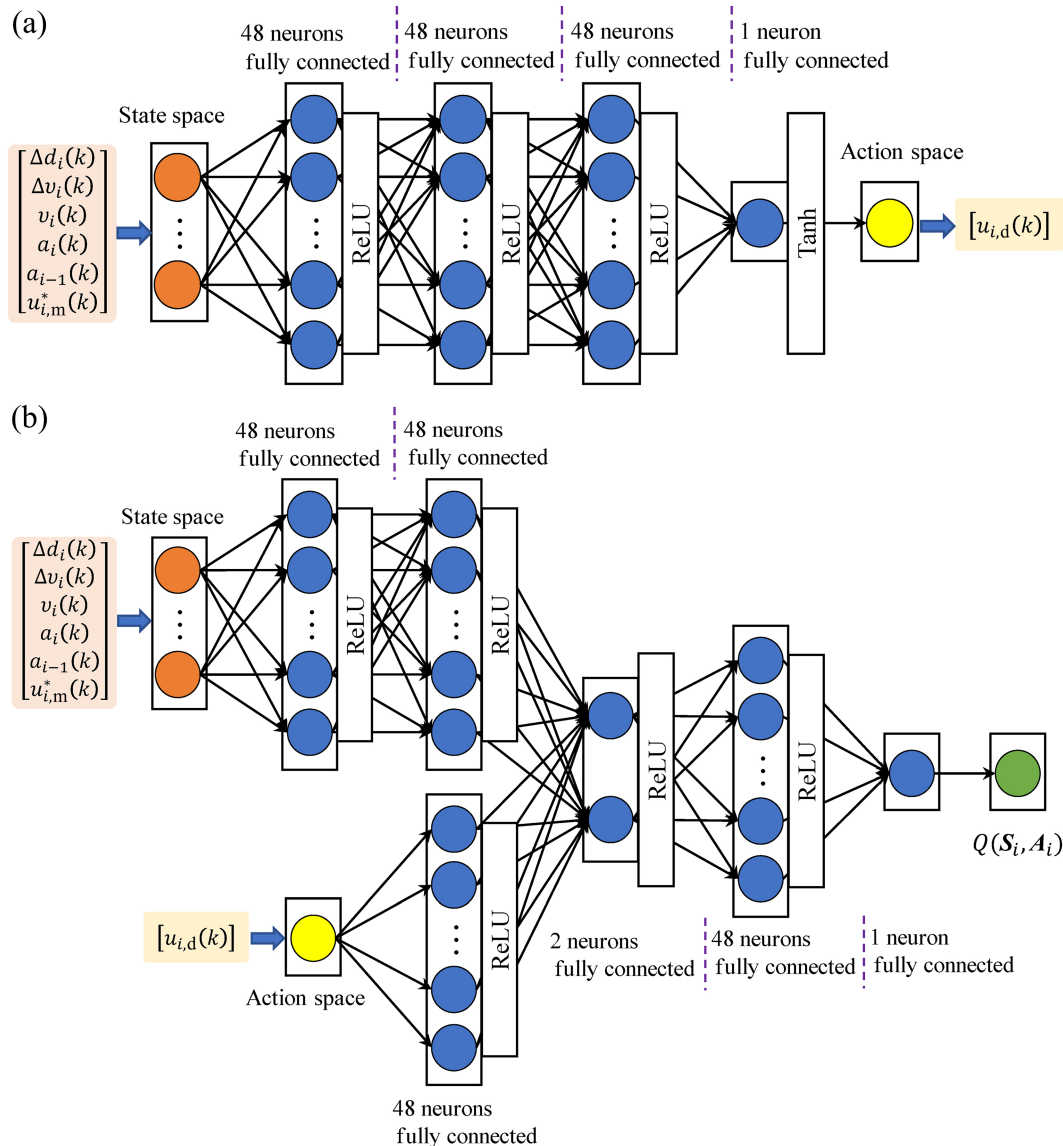


Fig. 4 TD3 algorithm network structure. (a) Actor network; (b) critic network.

$[-1, 1]$, subsequently scaled to the feasible control range $[-3 \text{ m/s}^2, 3 \text{ m/s}^2]$.

The critic network incorporates a state path and an action path, each employing ReLU activation functions. The state path comprises two hidden layers, each comprising 48 neurons, while the action path has a single hidden layer with 48 neurons. These paths converge at an addition layer with 2 neurons, which sums the outputs of the state and action paths. This combined output is then processed through a ReLU activation function before being passed to the subsequent network layer. Ultimately, the critic network outputs the $Q(\mathbf{S}_i, \mathbf{A}_i)$ value for state-action pairs, necessitating a hidden layer with 1 neuron.

Fig. 5 presents the cumulative reward curves for the PKARL and TD3 algorithms. Both approaches converge within 300 episodes, validating the reward design and parameter selection. The PKARL algorithm stabilizes around a cumulative reward of -300 by the 35th episode, exhibiting minor fluctuations due to variations in initial vehicle states. In contrast, the TD3 algorithm achieves convergence only after approximately 200 episodes and attains a lower cumulative reward.

These results indicate that the PKARL method not only accelerates convergence but also improves performance relative to conventional DRL approaches. The performance gain arises from the integration of a safety filter during training, which constrains the system to operate within a probabilistically safe set. The synergy between DMPC guidance and safety-constrained exploration significantly enhances training efficiency. The operational framework for platoon control using the PKARL method is comprehensively outlined in Algorithm 1.

4 Driver-in-the-loop experiment

To assess the efficacy of the PKARL method in managing scenarios involving HDV cut-ins and cut-outs within a platoon, we conducted driver-in-the-loop experiments. It should be emphasized that a uniform decision-making and control strategy was adopted across all vehicles in the platoon. Upon detection of an HDV cut-in by the onboard target recognition system, all impacted CAVs within the platoon reacted using the PKARL control strategy. Following an HDV cut-out, the PKARL strategy remains active until the platoon returns to a steady state, after which the controller switches back to the conventional DMPC method.

4.1 Experimental setup

The driver-in-the-loop experimental platform, illustrated in Fig. 6, was equipped with a Logitech G29 steering wheel, an accelerator and brake pedal assembly, a gear lever, a display, a computer, and other ancillary devices. The simulation environment and vehicle

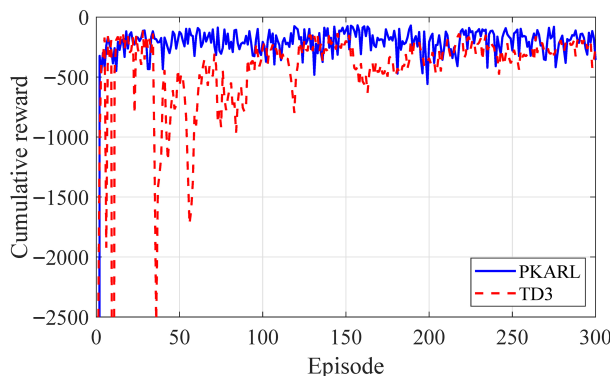


Fig. 5 Cumulative reward curves during training.

Algorithm 1 PKARL-Based platoon control strategy

Input: States of the preceding vehicle $i-1$: position p_{i-1} , speed v_{i-1} , and acceleration a_{i-1} ; states of the ego vehicle i : position p_i , speed v_i and acceleration a_i

Output: Optimal longitudinal acceleration command of the ego vehicle $u_i^*(k)$

- 1: Initialize actor network π_ϕ and twin critic networks $Q_{\theta_1}, Q_{\theta_2}$, with random parameters ϕ, θ_1, θ_2
- 2: Initialize target networks $\theta'_1 \leftarrow \theta_1, \theta'_2 \leftarrow \theta_2, \phi' \leftarrow \phi$
- 3: Initialize the position, speed and acceleration of all platoon vehicles and the HDV
- 4: **For** episode $m = 1$ to M **do**
- 5: **For** $k = 1$ to T **do**
- 6: Solve DMPC problem Eq. (6) to obtain nominal control $u_{i,m}^*(k)$
- 7: Construct state vector $\mathbf{S}_i(k)$
- 8: Select action $\mathbf{A}_i(k) \sim \pi_\phi(\mathbf{S}_i) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma)$
- 9: Compute safety-filtered control input $u_i^*(k)$ via Eq. (18)
- 10: Apply $u_i^*(k)$ to the vehicle system, observe the reward $R_i(k)$, and obtain the next state $\mathbf{S}'_i(k)$
- 11: Store $(\mathbf{S}_i, \mathbf{A}_i, R_i, \mathbf{S}'_i)$ in replay buffer $\mathcal{B}$
- 12: Compute TD target
 $\tilde{\mathbf{A}}_i(k) \leftarrow \pi_{\phi'}(\mathbf{S}'_i(k)) + \varepsilon, \varepsilon \sim \text{clip}(\mathcal{N}(0, \bar{\sigma}), -c, c)$
 $y \leftarrow R + \gamma \min_{i=1,2} Q_{\theta'_i}(\mathbf{S}'_i(k), \tilde{\mathbf{A}}_i(k))$
- 13: Update critics by minimizing Bellman error
- 14: **If** $k \bmod d = 0$:
- 15: Update actor using deterministic policy gradient
- 16: Soft-update target networks:
 $\theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i, \phi' \leftarrow \tau \phi + (1 - \tau) \phi'$
- 17: **End if**
- 18: **End for**
- 19: **End for**

dynamics were modeled on the high-fidelity TruckSim platform, which processes real-time input from the driver, such as the steering angle, throttle, and brake signals, to simulate vehicle dynamics. The decision-making and control algorithms for the platoon were developed and implemented using matrix laboratory (MATLAB)/Simulink.

In these experiments, the HDV was manually operated by the same driver across all trials, while the CAVs were autonomously controlled by the algorithms. This setting was adopted to reduce the influence of interdriver variability and to provide comparable cut-in and cut-out disturbances for evaluating different platoon control methods. During each trial, operational inputs from the driver and state data from both the HDV and the platoon CAVs were recorded for subsequent analysis.

The platoon is required to maintain a constant cruising speed of 22 m/s . The experiments are designed to evaluate the disturbance rejection capability of the platoon in response to HDV cut-in and cut-out maneuvers. With the time headway $t_d = 1 \text{ s}$ and the minimum safety distance $L = 5 \text{ m}$, the desired intervehicle spacing is set to 27 m . To account for the stochastic nature of real-world traffic conditions and to assess system robustness, the initial speed of the HDV is randomly selected within the range of $21\text{--}23 \text{ m/s}$ across different trials.

As shown in Fig. 7, during the experiment, the HDV initially cuts into the platoon from the left lane, subsequently cuts out, and

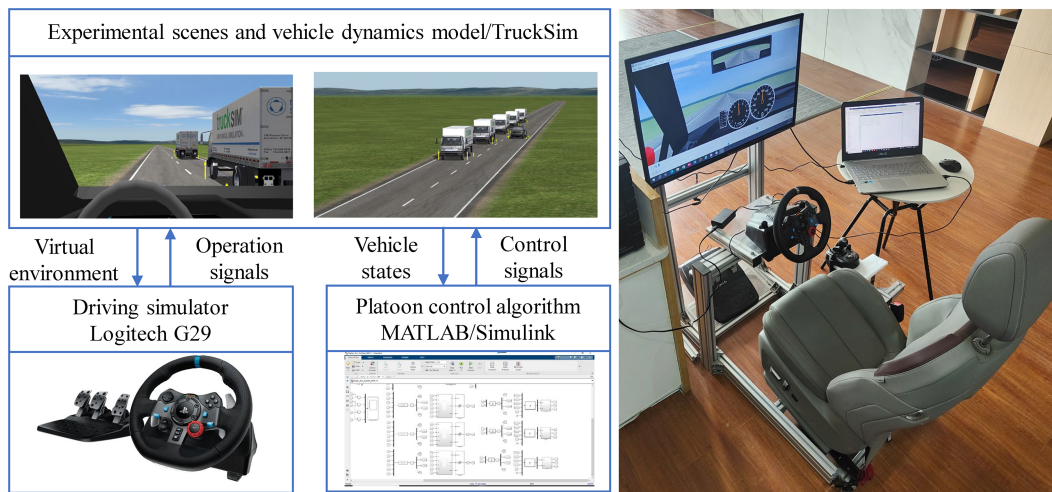


Fig. 6 Driver-in-the-loop experimental platform.

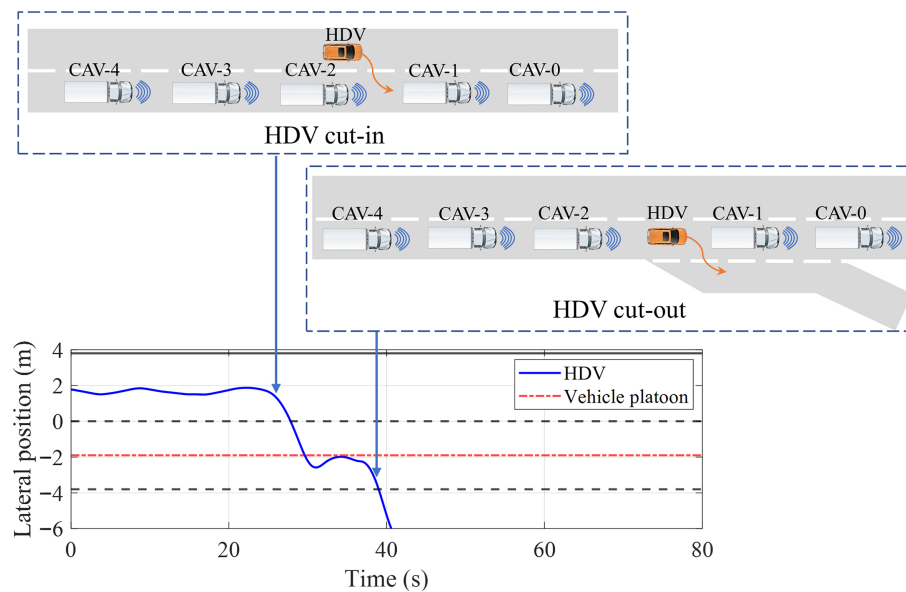


Fig. 7 Vehicle trajectories in HDV cut-in and cut-out scenarios.

eventually exits the highway. Due to the inherent limitations of human drivers in controlling vehicles with the precision of automated systems, small lateral oscillations occur during the driving process. In contrast, the algorithm-controlled platoon vehicles maintain precise lane centering.

4.2 Results and analysis

Fig. 8 illustrates the variations in speed and intervehicle distance for the platoon vehicles during HDV cut-in and cut-out scenarios. At a simulation time of 27.9 s, the HDV crosses the lane line and

merges into the platoon behind CAV-1. At this juncture, the following target of CAV-2 shifts from CAV-1 to the HDV, resulting in a sudden reduction in the intervehicle distance of CAV-2 from 27 to 10.36 m. To maintain a safe distance, CAV-2 applies braking and decelerates to a minimum speed of 18.63 m/s before beginning to accelerate again. Due to the inherent instability and randomness of human driver behavior, the speed profile of the HDV exhibits minor fluctuations around the desired speed of 22 m/s.

At simulation time $t = 39.4$ s, the HDV crosses the lane line

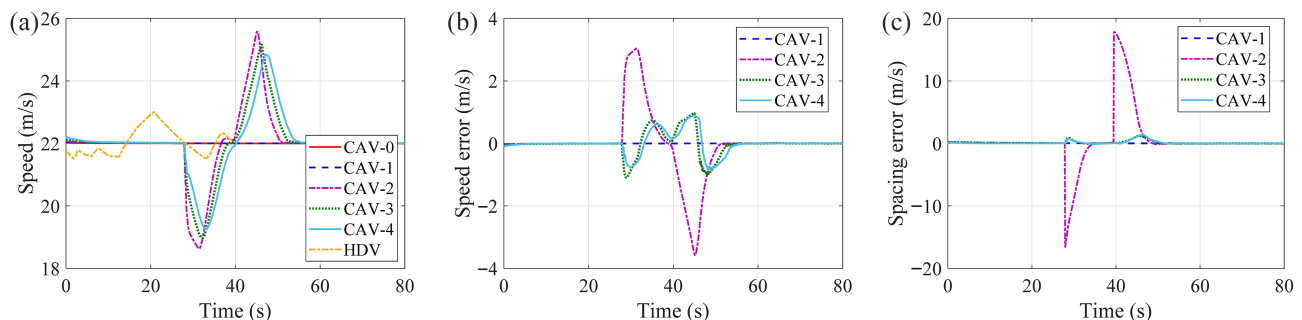


Fig. 8 PKARL control performance. (a) Speed; (b) speed error; (c) spacing error.

again and exits the platoon. Consequently, the following target of CAV-2 reverts from the HDV back to CAV-1. The intervehicle distance of CAV-2 rapidly increases from 27 to 44.9 m. To reduce this distance, CAV-2 accelerates and then decelerates to reach an equilibrium state, attaining a maximum speed of 25.58 m/s. Throughout the experiment, CAV-2, CAV-3, and CAV-4 exhibit prompt responses to changes in the state of the leading vehicle, with both speed error and spacing error showing a decreasing trend. This indicates that the platoon can effectively mitigate the effects of HDV cut-in and cut-out disturbances under the PKARL method.

Further analysis compares the performance of the PKARL method with DMPC (Ozkan and Ma, 2021; Zhou et al., 2019), Gaussian-process (GP)-MPC (Wang et al., 2024), and FH-deep deterministic policy gradient (DDPG)-SS (Liu et al., 2023) methods. Fig. 9 presents the performance comparison of the following vehicle, CAV-2, which experiences the largest HDV disturbance, under the four different control strategies. All four methods mitigate the HDV cut-in and cut-out disturbances and restore the equilibrium state of the platoon vehicles. The impact of the cut-in is more pronounced compared to the cut-out across all four methods.

At a simulation time of 27.9 s, when the HDV changes lanes and cuts in, the PKARL method exhibits a maximum speed error of -3.58 m/s, while the DMPC, GP-MPC, and FH-DDPG-SS methods show larger speed fluctuations. Compared with DMPC, the GP-MPC method reduces the peak tracking error and improves disturbance rejection performance by incorporating uncertainty-aware prediction, whereas FH-DDPG-SS further benefits from learning-based adaptation. Subsequently, the platoon returns to a stable state, with the DMPC and FH-DDPG-SS methods displaying spacing error overshoots of 6.65 and 3.22 m, respectively, whereas PKARL shows no overshoot. The GP-MPC method achieves a smaller spacing deviation than DMPC but still

exhibits larger transient fluctuations than PKARL. The proposed PKARL method demonstrates a more stable control effect compared to the other methods.

Following Yan et al. (2022), the cumulative speed error and cumulative spacing error are adopted as performance evaluation metrics. Smaller cumulative error values indicate improved tracking performance. Moreover, a progressive reduction in the cumulative speed and spacing errors of the following vehicles implies that tracking errors are not amplified during backward propagation, thereby indicating an empirical tendency of disturbance attenuation along the platoon. The effectiveness of the proposed control method is evaluated by comparing the cumulative speed error and cumulative spacing error of the three affected following vehicles after the HDV cut-in event.

According to the results presented in Fig. 10, the cumulative error of CAV-4 is larger than that of CAV-3 under the DMPC method, indicating that the disturbance is amplified during backward propagation. In contrast, the cumulative speed error and cumulative spacing error generally decrease along the platoon under the FH-DDPG-SS and PKARL methods, demonstrating their capability to attenuate HDV-induced disturbances. Although a slight increase can still be observed between CAV-3 and CAV-4 under the GP-MPC method, its cumulative errors remain significantly lower than those of DMPC, indicating an improved disturbance rejection capability. Among all methods, PKARL achieves the lowest average cumulative speed error and cumulative spacing error, demonstrating its superior ability to suppress disturbance propagation.

The quantitative comparison results are summarized in Table 3, where the DMPC method is selected as the baseline. Percentage changes are reported in parentheses, with negative values indicating reductions and positive values indicating increases. The results show that, on average, GP-MPC, FH-DDPG-SS, and PKARL reduce the cumulative speed and spacing errors relative to

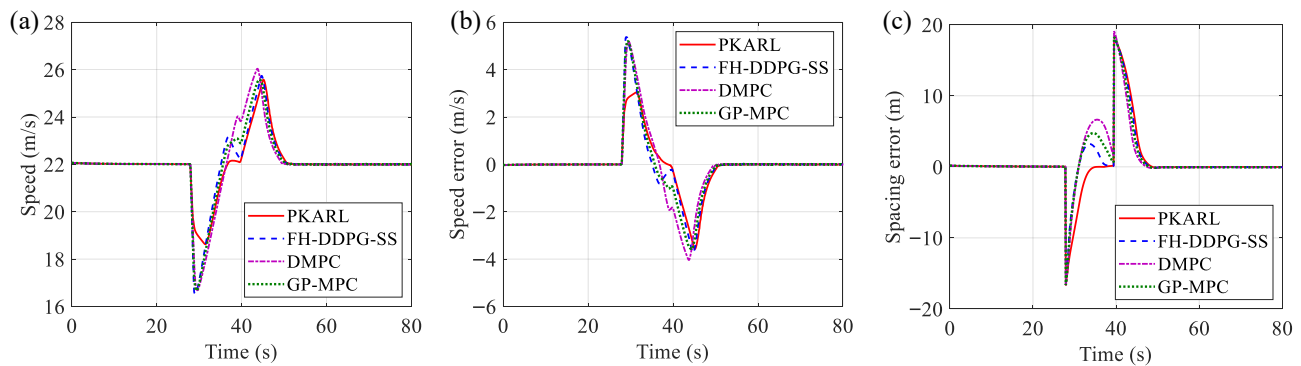


Fig. 9 Control performance comparison of CAV-2. (a) Speed; (b) speed error; (c) spacing error.

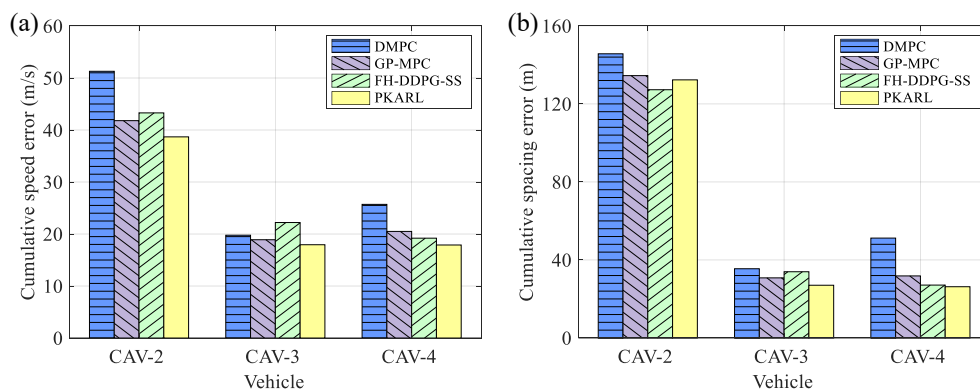


Fig. 10 Cumulative error comparison for the following CAVs. (a) Cumulative speed error; (b) cumulative spacing error.

Table 3 Performance comparison of different control methods

Vehicle number	$\sum_{t=0}^{80} \Delta v_i(t) $ (m/s)				$\sum_{t=0}^{80} \Delta d_i(t) $ (m)			
	DMPC	GP-MPC	FH-DDPG-SS	PKARL	DMPC	GP-MPC	FH-DDPG-SS	PKARL
CAV-2	51.29	41.82 (−18.5%)	43.29 (−15.6%)	38.68 (−24.6%)	145.70	134.52 (−7.7%)	127.27 (−12.6%)	132.33 (−9.2%)
CAV-3	19.77	18.91 (−4.4%)	22.21 (+12.3%)	17.94 (−9.3%)	35.48	30.85 (−13.0%)	33.94 (−4.3%)	27.02 (−23.8%)
CAV-4	25.70	20.55 (−20.0%)	19.20 (−25.3%)	17.89 (−30.4%)	51.20	31.81 (−37.9%)	27.10 (−47.1%)	26.26 (−48.7%)
Average	32.25	27.09 (−16.0%)	28.23 (−12.5%)	24.84 (−23.0%)	77.46	65.73 (−15.1%)	62.77 (−19.0%)	61.87 (−20.1%)

the DMPC baseline, although FH-DDPG-SS produces a larger cumulative speed error for CAV-3. Overall, these results demonstrate the effectiveness of uncertainty-aware prediction and learning-based adaptation in improving platoon disturbance rejection performance. Specifically, the GP-MPC method achieves average reductions of 16.0% in cumulative speed error and 15.1% in cumulative spacing error, demonstrating the effectiveness of incorporating predictive uncertainty into the platoon control framework. Relative to the DMPC baseline, the FH-DDPG-SS method achieves average reductions of 12.5% in cumulative speed error and 19.0% in cumulative spacing error. In comparison, the PKARL method yields larger average reductions of 23.0% and 20.1%, respectively.

Compared with GP-MPC, the proposed PKARL method further improves the reduction rates of cumulative speed error and cumulative spacing error by 7.0% and 5.0%, respectively. Furthermore, compared with the pure DRL method FH-DDPG-SS, the PKARL approach further improves the corresponding reduction rates by 10.5% and 1.1%, respectively, demonstrating its superior capability in mitigating the impact of HDV-induced disturbances.

Overall, the four methods exhibit a clear performance hierarchy, with DMPC providing the baseline model-based control performance, GP-MPC improving disturbance handling through uncertainty-aware prediction, FH-DDPG-SS benefiting from data-driven adaptation, and PKARL achieving the best overall performance by integrating model-based prior knowledge, reinforcement learning, and safety-constrained control.

5 Conclusions and future work

This study proposes a novel platoon control method, termed PKARL, to address the uncertainties associated with HDV cut-in maneuvers and to minimize disturbances within vehicle platoon systems. The method leverages a priori knowledge to enhance the performance of the DRL algorithm, improving convergence speed and generalization while ensuring system safety and stability. DMPC is first employed to provide structured guidance and a priori knowledge during the learning process. The DRL framework integrates this prior knowledge through the design of state representations, action formulations, reward functions, and training scenarios, enabling the learning of efficient control policies for uncertain HDV cut-in events. A safety filter based on CLF and CBF is then applied to optimize the DRL outputs in real time, guaranteeing compliance with safety and stability requirements. Driver-in-the-loop experimental results demonstrate that the proposed PKARL method effectively mitigates HDV cut-in and cut-out disturbances, enhances platoon tracking performance, and reduces cumulative speed and spacing errors along the platoon, indicating improved platoon stability.

Future research will consider imperfect communication, including communication delays and packet loss, to improve the robustness of the proposed method in more stochastic

environments. More complex scenarios, such as multiple HDV cut-ins, continuous lane-changing disturbances, and denser mixed traffic, will also be investigated to further verify the robustness and scalability of the PKARL framework in real traffic environments. In addition, the proposed framework has the potential to be extended to other cooperative autonomous systems, such as unmanned aerial vehicle formations and multirobot systems.

Author contributions

Junru Yang: Conceptualization, Methodology, Writing – original draft. **Sifa Zheng:** Funding acquisition, Resources, Supervision. **Chuan Sun:** Formal analysis, Validation, Writing – review & editing. **Haoran Li:** Project administration, Investigation, Visualization. **Lin Xu:** Data curation, Software.

Replication and data sharing

The data collected within this research can be made accessible upon request via email to the first author or the corresponding author.

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2023YFB4302600), Natural Science Foundation of Jiangsu Province (No. BK20231197), Science and Technology Program of Suzhou (No. SYG2025116), Hubei Science and Technology Project (Nos. 2025CDB015 and 2024BAB087), and Huanggang Technical Breakthrough General Project (No. JGYB20250014).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

- Ames, A. D., Coogan, S., Egerstedt, M., Notomista, G., Sreenath, K., Tabuada, P., 2019. Control barrier functions: Theory and applications. In: 2019 18th European Control Conference (ECC), 3420–3431.
- Ames, A. D., Xu, X., Grizzle, J. W., Tabuada, P., 2017. Control barrier function based quadratic programs for safety critical systems. *IEEE Trans Autom Control*, **62**, 3861–3876.
- An, G., Talebpour, A., 2023. Vehicle platooning for merge coordination in a connected driving environment: A hybrid ACC-DMPC approach. *IEEE Trans Intel Transp Syst*, **24**, 5239–5248.
- Basiri, M. H., Ghogh, B., Azad, N. L., Fischmeister, S., Karray, F., Crowley, M., 2020. Distributed nonlinear model predictive control and metric learning for heterogeneous vehicle platooning with cut-in/cut-out maneuvers. In: 2020 59th IEEE Conference on Decision and Control (CDC), 2849–2856.
- Chen, Q., Zhao, W., Li, L., Wang, C., Chen, F., 2022. ES-DQN: A learning method for vehicle intelligent speed control strategy under uncertain cut-In scenario. *IEEE Trans Veh Technol*, **71**, 2472–2484.
- Desai, M., Ghaffari, A., 2022. CLF-CBF based quadratic programs for safe motion control of nonholonomic mobile robots in presence of mov-

- ing obstacles. In: 2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 16–21.
- Fujimoto, S., van Hoof, H., Meger, D., 2018. Addressing Function Approximation Error in Actor-Critic Methods. <https://doi.org/10.48550/arXiv.1802.09477>
- González, E., Sanchis, J., García-Nieto, S., Salcedo, J., 2020. A comparative study of stochastic model predictive controllers. *Electronics*, **9**, 2078.
- Guo, L., Jia, Y., 2022. Anticipative and predictive control of automated vehicles in communication-constrained connected mixed traffic. *IEEE Trans Intel Transp Syst*, **23**, 7206–7219.
- Hou, Z., Liu, W., Knoll, A., 2025. Safe reinforcement learning for autonomous driving by using disturbance-observer-based control barrier functions. *IEEE Trans Intel Veh*, **10**, 3782–3791.
- Kazemi, H., Mahjoub, H. N., Tahmasbi-Sarvestani, A., Fallah, Y. P., 2018. A learning-based stochastic MPC design for cooperative adaptive cruise control to handle interfering vehicles. *IEEE Trans Intel Veh*, **3**, 266–275.
- Li, Q., Zhang, P., Yao, H., Chen, Z., Li, X., 2024. Online learning-based model predictive trajectory control for connected and autonomous vehicles: Modeling and physical tests. *J Intel Connect Veh*, **7**, 86–96.
- Liu, P., Kurt, A., Ozguner, U., 2019. Distributed model predictive control for cooperative and flexible vehicle platooning. *IEEE Trans Control Syst Technol*, **27**, 1115–1128.
- Liu, T., Lei, L., Zheng, K., Zhang, K., 2023. Autonomous platoon control with integrated deep reinforcement learning and dynamic programming. *IEEE Internet Things J*, **10**, 5476–5489.
- Liu, Y., Wang, W., Zhao, D., 2022. Modified cooperative adaptive cruise control strategy for optimizing cooperative driving in freeway weaving segments. *Transp Res Rec J Transp Res Board*, **2676**, 214–226.
- Lu, Y., Huang, L., Yao, J., Su, R., 2023. Intention prediction-based control for vehicle platoon to handle driver cut-In. *IEEE Trans Intel Transp Syst*, **24**, 5489–5501.
- Lu, Y., Wang, B., Huang, L., Zhao, N., Su, R., 2022. Modeling of driver cut-in behavior towards a platoon. *IEEE Trans Intel Transp Syst*, **23**, 24636–24648.
- Lyu, N., Wen, J., Duan, Z., Wu, C., 2022. Vehicle trajectory prediction and cut-In collision warning model in a connected vehicle environment. *IEEE Trans Intel Transp Syst*, **23**, 966–981.
- Ma, Z., Huo, Q., Yang, X., Zhao, X., 2020. Safety cruise control of connected vehicles using radar and vehicle-to-vehicle communication. *IEEE Syst J*, **14**, 4602–4613.
- Milánés, V., Shladover, S. E., 2016. Handling cut-In vehicles in strings of cooperative adaptive cruise control vehicles. *J Intel Transp Syst*, **20**, 178–191.
- Omae, M., Fukuda, R., Ogitsu, T., Chiang, W. P., 2014. Control procedures and exchanged information for cooperative adaptive cruise control of heavy-duty vehicles using broadcast inter-vehicle communication. *Int J Intel Transp Syst Res*, **12**, 84–97.
- Ozkan, M. F., Ma, Y., 2021. Fuel-economical distributed model predictive control for heavy-duty truck platoon. In: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 1919–1926.
- Pi, D., Xue, P., Wang, W., Xie, B., Wang, H., Wang, X., et al., 2023. Automotive platoon energy-saving: A review. *Renew Sustain Energy Rev*, **179**, 113268.
- Selvaraj, D. C., Hegde, S., Amati, N., Deflorio, F., Chiasserini, C. F., 2023. An ML-aided reinforcement learning approach for challenging vehicle maneuvers. *IEEE Trans Intel Veh*, **8**, 1686–1698.
- Stefansson, E., Fisac, J. F., Sadigh, D., Sastry, S. S., Johansson, K. H., 2019. Human-robot interaction for truck platooning using hierarchical dynamic games. In: 2019 18th European Control Conference (ECC), 3165–3172.
- Tu, Y., Wang, W., Li, Y., Xu, C., Xu, T., Li, X., 2019. Longitudinal safety impacts of cooperative adaptive cruise control vehicle's degradation. *J Saf Res*, **69**, 177–192.
- Wang, J., Pant, Y. V., Jiang, Z., 2024. Learning-based modeling of human-autonomous vehicle interaction for improved safety in mixed-vehicle platooning control. *Transp Res Part C Emerg Technol*, **162**, 104600.
- Xu, X., Wang, X., Wu, X., Hassanin, O., Chai, C., 2021. Calibration and evaluation of the Responsibility-Sensitive Safety model of autonomous car-following maneuvers using naturalistic driving study data. *Transp Res Part C Emerg Technol*, **123**, 102988.
- Yan, R., Jiang, R., Jia, B., Huang, J., Yang, D., 2022. Hybrid car-following strategy based on deep deterministic policy gradient and cooperative adaptive cruise control. *IEEE Trans Autom Sci Eng*, **19**, 2816–2824.
- Yang, J., Chu, D., Yin, J., Pi, D., Wang, J., Lu, L., 2024. Distributed model predictive control for heterogeneous platoon with leading human-driven vehicle acceleration prediction. *IEEE Trans Intel Transp Syst*, **25**, 3944–3959.
- Yang, L., Yuan, J., Zhao, X., Fang, S., He, Z., Zhan, J., et al., 2023. SceGAN: A method for generating autonomous vehicle cut-In scenarios on highways based on deep learning. *J Intel Connect Veh*, **6**, 264–274.
- Zhan, J., Ma, Z., Zhang, L., 2023. Data-driven modeling and distributed predictive control of mixed vehicle platoons. *IEEE Trans Intel Veh*, **8**, 572–582.
- Zhang, T., Zou, Y., Zhang, X., Guo, N., Wang, W., 2021. Data-driven based cruise control of connected and automated vehicles under cyber-physical system framework. *IEEE Trans Intel Transp Syst*, **22**, 6307–6319.
- Zhang, Y., Lin, Y., Qin, Y., Dong, M., Gao, L., Hashemi, E., 2024. A new adaptive cruise control considering crash avoidance for intelligent vehicle. *IEEE Trans Ind Electron*, **71**, 688–696.
- Zhou, Y., Wang, M., Ahn, S., 2019. Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability. *Transp Res Part B Methodol*, **128**, 69–86.
- Zhu, X., Hu, W., Deng, Z., Zhang, J., Hu, F., Zhou, R., et al., 2022. Interaction-aware cut-In trajectory prediction and risk assessment in mixed traffic. *IEEE/CAA J Autom Sin*, **9**, 1752–1762.



control.

Junru Yang received the Ph.D. degree in transportation engineering from Wuhan University of Technology, Wuhan, China, in 2024. He was a Joint Ph.D. Student with the Polytechnic University of Milan, Italy, in 2021. He is currently a Postdoctoral Researcher with Suzhou Automotive Research Institute, Tsinghua University, Suzhou, China. His research interests include decision making in automated driving, vehicle dynamics and



dynamics and control.

Sifa Zheng received the B.E. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1993 and 1997, respectively. He is currently a Professor with the School of Vehicle and Mobility and the State Key Laboratory of Automotive Safety and Energy, Tsinghua University. He is also the Deputy Director of Suzhou Automotive Research Institute, Tsinghua University. His research interests include autonomous driving and vehicle



Chuan Sun received the B.E. degree in vehicle engineering from Hainan University, Haikou, China, in 2011, the M.E. degree in vehicle engineering from Dalian Jiaotong University, Dalian, China, in 2013, and the Ph.D. degree in vehicle engineering from Wuhan University of Technology, Wuhan, China, in 2017. He visited Delft University of Technology, The Netherlands, as a Joint Ph.D. Student, in 2015. He is currently a Senior Engineer with Suzhou Automotive Research Institute, Tsinghua University, Suzhou, China. His research interests include autonomous and connected vehicles.



Haoran Li received the B.E. degree in vehicle engineering from Wuhan University of Science and Technology, Wuhan, China, in 2011, the M.E. degree in vehicle engineering from Wuhan University of Technology, Wuhan, China, in 2014, and the Ph.D. degree in vehicle engineering from Wuhan University of Technology, Wuhan, China, in 2021. He is currently a Senior Engineer with Suzhou Automotive Research Institute, Tsinghua University, Suzhou, China, and a Lecturer with the School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, China. His research interests include autonomous and connected vehicles.



Lin Xu received the M.S. degree in traffic information engineering and control from Shanghai Maritime University, Shanghai, China, in 2022. He is currently an Engineer with Suzhou Automotive Research Institute, Tsinghua University, Suzhou, China. His research interests include the research and development of automated driving algorithms.