Transportation Research Part C 148 (2023) 104019

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A deep reinforcement learning based distributed control strategy for connected automated vehicles in mixed traffic platoon

Haotian Shi [a], Danjue Chen [b], Nan Zheng [c], Xin Wang [d], Yang Zhou [e,*], Bin Ran [f]

[a] Department of Civil and Environmental Engineering, University of Wisconsin, Madison, 1208 Engineering Hall 1415 Engineering Drive, Madison, WI 53706, United States
[b] Department of Civil and Environmental Engineering, University of Massachusetts Lowell, Shal Hall 200O, 1 University Ave, Lowell, MA 01854, United States
[c] Department of Civil Engineering, Monash University, 23 College Walk (B60), Clayton Campus, Monash Clayton 3800, United States
[d] Department of Industrial System Engineering, 3258 Mechanical Engineering Building 1513 University Ave, Madison, WI 53706, United States
[e] Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, 199 Spence Street, DLEB 301 A, 3136 TAMU, College Station, TX 77843, United States
[f] Department of Civil and Environmental Engineering, University of Wisconsin, Madison, 1208 Engineering Hall 1415 Engineering Drive, Madison, WI 53706, United States

ARTICLE INFO

Keywords:
Mixed traffic environment
Distributed control
Deep reinforcement learning
Traffic oscillation dampening
Connected automated vehicle

ABSTRACT

This paper proposes an innovative distributed longitudinal control strategy for connected automated vehicles (CAVs) in the mixed traffic environment of CAV and human-driven vehicles (HDVs), incorporating high-dimensional platoon information. For mixed traffic, the traditional CAV control method focuses on microscopic trajectory information, which may not be efficient in handling the HDV stochasticity (e.g., long reaction time; various driving styles) and mixed traffic heterogeneities. Different from traditional methods, our method, for the first time, characterizes consecutive HDVs as a whole (i.e., AHDV) to reduce the HDV stochasticity and utilize its macroscopic features to control the following CAVs. The new control strategy takes advantage of platoon information to anticipate the disturbances and traffic features induced downstream under mixed traffic scenarios and greatly outperforms the traditional methods. In particular, the control algorithm is based on deep reinforcement learning (DRL) to fulfill car-following control efficiency and further address the stochasticity for the aggregated car following behavior by embedding it in the training environment. To better utilize the macroscopic traffic features, a general platoon of mixed traffic is categorized as a CAV-HDVs-CAV pattern and described by corresponding DRL states. The macroscopic traffic flow properties are built upon the Newell car-following model to capture the characteristics of aggregated HDVs' joint behaviors. Simulated experiments are conducted to validate our proposed strategy. The results demonstrate that the proposed control method has outstanding performances in terms of oscillation dampening, eco-driving, and generalization capability.

* Corresponding author.
E-mail addresses: hshi84@wisc.edu (H. Shi), danjue_chen@uml.edu (D. Chen), Nan.Zheng@monash.edu (N. Zheng), xin.wang@wisc.edu (X. Wang), yangzhou295@tamu.edu (Y. Zhou), bran@wisc.edu (B. Ran).

https://doi.org/10.1016/j.trc.2023.104019
Received 25 May 2022; Received in revised form 2 January 2023; Accepted 9 January 2023
Available online 30 January 2023

1. Introduction

Connected and automated vehicles (CAVs), supported by intelligent onboard units and V2X wireless communication, have gained extensive research interest in recent years and gradually occupied the vehicle market (Zou and Qu, 2018). It is envisioned that the CAVs and human-driven vehicles (HDVs) will co-exist in the near future (Zhou et al., 2019a), which will change the pure traffic flow of HDVs to the mixed traffic flow of CAVs and HDVs (Lu and Liu, 2021; Zhang et al., 2021). Despite the changes in the traffic environment, traffic oscillations (i.e., “stop-and-go” traffic), which impair traffic flow stability, traffic safety, vehicle energy efficiency, environmental sustainability, and driving comfort, remain a demanding issue in the congested mixed traffic flow (Shi et al., 2021b). CAVs can drive efficiently using advanced control strategies such as ACC (Takahama and Akasaka, 2018) and CACC technologies (Knorn et al., 2014), which have the potential to stabilize traffic oscillations and optimize the mixed traffic flow performance. Therefore, efficiently controlling CAVs to drive safely and smartly in a mixed traffic environment is important both in academia and industry implementation.

Currently, the mainstream tools for CAV control can be divided into three categories: linear-based controller, model predictive control (MPC) based controller, and deep reinforcement learning (DRL) based controller. The closed-form linear or nonlinear state-feedback CAV controller (e.g., Naus et al., 2010; Morbidi et al., 2013; Li et al., 2019; Wang et al., 2020b) has been favored since it is fast and easy to implement due to its analytical representation. Despite the simplicity, this type of controller is hard to support the constrained optimization framework with multiple explicit objectives and constraints. The MPC-based CAV control (e.g., Gong et al., 2016; Zhou et al., 2017; Zhou et al., 2019b) supported by a flexible framework solves this limitation, which optimizes multiple objectives with constraints in a rolling horizon. However, MPC usually requires the problem to be convex and takes a relatively high computation burden, which makes it hard for real-time implementation with high resolution (e.g., per 0.01 sec). Compared with MPC, DRL-based control (e.g., Wang et al., 2019b; Duan et al., 2020; Qu et al., 2020) is suitable for capturing the stochastic characteristics of a complex system and normally enables great generalization capability through self-learning. Furthermore, the DRL-based controller is fast computing for real-time implementation since the offline training process takes the major computational burden (Görges, 2017; Shi et al., 2021a). However, the equilibrium concept from the control theory, which regulates an explicit gap policy and directly guarantees a stable traffic flow, and the consensus concept (Zhang and Orosz, 2017), which facilitates a system-level control performance in a distributed manner, are seldom incorporated in the DRL control framework. To the authors’ knowledge, only Shi et al. (2021) incorporate the equilibrium concept into DRL-based control to conduct stability analysis regarding string stability (Ploeg et al., 2014) and local stability (Willems, 2013).

Although these CAV longitudinal controllers can provide strong tools to control CAVs efficiently, how to handle mixed traffic still remains a problem due to the heterogeneity and HDVs’ stochastic and uncertain movements (Gong and Du, 2018). The stochastic HDV driving behavior triggers traffic disturbances and amplifies the oscillation amplitude through the vehicular stream, which impairs the mixed traffic flow stability, travel efficiency, and energy (Zheng et al., 2020a). Moreover, the heterogeneous driving behaviors in mixed traffic may create voids and further reduce traffic throughput (Chen et al., 2020). To handle the HDV uncertainty and mixed traffic heterogeneities, approaches of recent studies can be largely divided into two categories: 1). predict the proceeding HDV’s driving behaviors (Gong and Du, 2018; Bang and Ahn, 2019; Wang et al., 2020a; Zhu et al., 2018) and incorporate the prediction in the control strategy (e.g., by MPC); 2). divide mixed traffic into sub-platoons (Shi et al., 2021b; Wang, 2018) for more efficient control and apply a cooperative control strategy for CAVs assuming that the disturbances triggered by HDVs are completely random. For the first type of method, Gong and Du (2018) utilized an online curve matching algorithm to predict the HDV trajectory and developed a cooperative platoon control for the mixed traffic environment. However, this paper only predicts the last HDV of consecutive HDVs in the mixed traffic, which does not fully use the downstream traffic information that can be potentially conveyed by proceeding CAVs. The downstream traffic information can be very helpful in predicting the wave propagation and oncoming traffic scenario. On the other hand, Shi et al. (2021) proposed a DRL-based cooperative CAV longitudinal control strategy for the mixed traffic environment, which divides the mixed platoon into multiple subsystems for centralized cooperative control. However, the method cannot utilize the downstream information of each sub-platoon and model it as random noise in the environment. Moreover, this approach lacks sufficient generalization with the increased centralized CAV sizes. With this concern, a distributed manner may be better for CAV control in mixed traffic. Besides the two primary approaches mentioned above, the connected cruise control (CCC) strategy was developed to handle the various connectivity structures in heterogeneous platoons by assuming an oversimplified nonlinear car following laws for HDVs. Based on that, they designed a control law for CAVs to improve the car following performance and traffic efficiency (e.g., Ge and Orosz, 2014; Orosz, 2016; Zhang and Orosz, 2016). However, these studies did not focus on addressing HDVs’ inherent stochasticity and stabilizing the mixed platoon.

In general, although these approaches (e.g., HDV behavior prediction; sub-platoon) could improve CAV control performances in the mixed traffic environment, the limits still remain as follows. Firstly, it is challenging to effectively incorporate HDVs’ behavior for control due to its inherent stochastic and personalized characteristics, especially for the aggregated (i.e., multiple consecutive) HDVs in the mixed traffic. The joint behaviors of multiple HDVs are hard to model. In the case of platooning, while only the individual behaviors are modeled, the predicting error will be accumulated and propagated over time and space (Lin et al., 2020). Even the microscopic behavior is stochastic and difficult to capture, the aggregated HDV driving behaviors exhibit macroscopic traffic flow properties (e.g., kinematic wave propagating time, density) with typical traffic phenomena (e.g., shock wave propagation), and they can be modeled by for example the fundamental diagram (Meng et al., 2021; Tian et al., 2021). Comparing to the microscopic behaviors, the aggregated driving behaviors show less stochasticity, as indicated by the central limit theorem (Kwak and Kim, 2017). Therefore, this paper aims to attenuate the aggregated HDVs’ stochasticity by incorporating their macroscopic traffic properties into the control framework. Secondly, the mixed traffic environment has various vehicular compositions due to the different combinations of HDVs and CAVs. Approaches such as sub-platoon or centralized cooperative control (Du et al., 2020; Shi et al., 2021b; Wang, 2018; Zheng et al., 2020b) lack adequate flexibility and may suffer from computation burden. Such heterogeneity makes it challenging to develop a generic CAV control approach with a system-level control performance. With this concern, the study aims to build a generic CAV control framework to generalize the varied CAV-HDV topologies and incorporate the macroscopic features.

Based on the two gaps identified above, we propose a novel vehicle following structure, "CAV-AHDV-CAV," as a generic unit for mixed traffic of any vehicle ordering and simultaneously embedded platoon-level features in the distributed CAV control framework. 'AHDV' means the aggregated HDVs between the two CAVs in the structure. This novel structure characterizes the aggregated consecutive HDVs in the mixed traffic as a whole, denoted as the 'AHDV,' whose aggregated HDV car-following behaviors and stochasticity can be further captured by the macroscopic traffic features. Specifically, we propose an estimated time-varying Newell car-following method (Chen et al., 2012), which links the fundamental diagram to the microscopic driving behavior parameters.

Furthermore, DRL is suitable for capturing stochastic characteristics and embedding them in the environment with great generalization capability. Thus, this structure is incorporated into the DRL framework to fulfill car-following control efficiency and further reduce stochasticity based on the following two aspects. First, the ground-truth HDV trajectory data are embedded into the DRL training process, by which we incorporate real HDV stochastic characteristics implicitly. Second, the macroscopic features captured by the 'CAV-AHDV-CAV' structure are weighted and fused into the DRL state and reward function based on the equilibrium concept. In this way, the HDVs' stochasticity is alleviated by regulating CAVs close to the pre-defined equilibrium state. With the proposed 'CAV-AHDV-CAV' structure and the designed DRL framework, the HDV stochasticity is efficiently alleviated for CAV control.

To summarize, this paper utilizes the DRL framework to develop a generic distributed CAV longitudinal control approach for a mixed traffic environment. The contribution can be summarized in terms of methodology and application. From the methodology-wise perspective, the aggregated HDVs' macroscopic traffic flow features are real-time estimated based on the generic 'CAV-AHDV-CAV' structure. The structure is embedded into the DRL control framework by a specially designed DRL state and reward function, which efficiently alleviates the adverse impact of HDVs' stochasticity and optimizes the whole mixed traffic flow. From the application-wise side, a generic strategy for any CAV-HDV topology of a mixed vehicular platoon is developed to stabilize the traffic oscillations efficiently. Specifically, each controlled CAV receives the information from the local downstream vehicles for real-time control. The received information is fused as the DRL state based on the philosophy of equilibrium concept and the consensus concept, which helps develop a robust control policy and gives the base for analyzing car-following control efficiency and vehicular string stability. The DRL reward function is then designed based on the fused DRL state in a quadratic form to efficiently fulfill the car-following control efficiency and improve driving comfort performance.

The paper is organized as follows. Section 2 provides the CAV longitudinal control scheme, including the environment setting, the distributed control scheme, and the state fusion strategy. Section 3 gives the details of DRL-based control model development, in which the basics of the DRL algorithm are discussed in Section 3.1; the policy updating algorithm is given in Section 3.2, and the training procedure is described in Section 3.3. Section 4 analyzes the results of simulated experiments in terms of control performance, driving comfort, and generalization capability. Section 5 concludes the work.

2. Cav control scheme

2.1. Assumptions and environment setting

2.1.1. Assumptions

This paper focuses on the CAV longitudinal control in mixed traffic of CAVs and HDVs. We consider the car following process without lateral movement in an infinite highway segment. The communication between CAVs follows the Federal Communications Commission, allocating a dedicated short-range communication (DSRC) radio with a 5.9-GHz frequency (Du and Dao, 2015). The environment assumptions are given as follows: (i) The CAV can obtain the real-time state information (e.g., speed, position) of its immediate preceding vehicle using onboard sensors. (ii) The CAV can receive its own real-time state information. (iii) The CAV's real-time state information is broadcasted to the upstream CAVs by vehicle-to-vehicle (V2V) communication. (iv) The Signal-Interference-plus-Noise Ratio (SINR) condition dynamically determines the transmission status (fail/success) between any CAV pairs. (v) The communication delay is negligible due to the short communication distance in a road segment. (vi) HDVs have no communication capability. (vii) The lane-changing maneuvers are not considered in the vehicular platoon.

2.1.2. Communications

For the DSRC-based V2V communication environment given in the above assumptions, the information transmission status between CAVs can change dynamically under communication failure due to communication interference or information congestion (Wang et al., 2019a). The communication failure will undermine driving performance. We embed this realistic communication property into the control framework to enhance the robustness and practicality of the CAV controller. The information flow topology (IFT), which indicates the dynamic transmission status of links in the vehicular platoon, is described from the receiver side (i.e., controlled CAV) to illustrate the communication environment. Specifically, the IFT of CAV i at timestep t is defined as $\boldsymbol{\xi}_i^t = [\eta_{i,i-1}^t, \eta_{i,i-2}^t, \cdots, \eta_{i,i-N}^t]$, where $\eta_{i,i-m}^t \in \{0, 1\}$ denotes the information transmission status between the receiver CAV i and the downstream vehicle $i-m$. $\eta_{i,i-m}^t = 1$ indicates a successful transmission, while $\eta_{i,i-m}^t = 0$ can happen either when a communication loss or vehicle $i-m$ is an HDV. In addition, we assume a permanently successful transmission status for the immediate preceding vehicle (i.e., $\eta_{i,i-1}^t \equiv 1$) due to CAV's robust onboard sensors. To better replicate the DSRC-based V2V communication, the SINR communication model (Du and Dao, 2015), which demonstrates great estimation of communication loss on a one-way road segment, is adopted to identify CAVs' IFTs. The SINR model determines the real-time transmission quality $y^t_{i,j}$ between the transmitter CAV j and the receiver CAV i at timestep t, defined as Eq. (1):

$$y^t_{i,j} = \frac{P_j\left(X^t_{ji}\right)^{-\alpha}}{\sum_{k=j+1}^{i-1} P_k\left(X^t_{ki}\right)^{-\alpha} + O}, \tag{1}$$

where P_j denotes the transmission power of vehicle i; X^t_{ji} is the distance between two CAVs; α is the parameter adjusting the signal power decay. $\sum_{k=j+1}^{i-1} P_k(X^t_{ki})^{-\alpha}$ represents the sum of the interference signal power of vehicles between the receiver CAV i and transmitter CAV j. The noise term O $N(\mu,\sigma^2)$ is used to simulate the random noise affecting the communication environment. Based on $y^t_{i,j}$, a threshold value β related to the communication capability (e.g., modulation, code rate) is introduced to determine the real-time transmission status $\eta^t_{i,i-m}$:

$$\eta^t_{i,i-m} = \begin{cases} 1, y^t_{i,i-m} > \beta \\ 0, y^t_{i,i-m} \leq \beta \end{cases} \tag{2}$$

Based on the SINR model, we embed critical communication features of the practical condition in the simulated V2V communication environment, making the simulation more realistic.

2.1.3. Vehicle dynamics

Given the assumptions and communication environment, the vehicle dynamics are modeled by a first-order approximation to capture multiple factors (e.g., gear position, road gradient, air drag force) of the vehicle linearized dynamics (Wang, 2018):

$$a^{t+1}_i = e^{-\frac{\Delta t}{I_{i,L}}} \times a^t_i + \left(1 - e^{-\frac{\Delta t}{I_{i,L}}}\right) \times K_{i,L} u^t_i, \tag{3}$$

where $K_{i,L}$ and $I_{i,L}$ denote the system gain (ratio of the control demand that can be realized) and actuation time lag of CAV i, respectively; u^t_i and a^t_i are the demanded acceleration and realized acceleration. With acceleration a^t_i, the real-time vehicle state is updated according to the kinematic point-mass equations (Zhu et al., 2018):

$$v^{t+1}_i = v^t_i + a^t_i \Delta t \tag{4a}$$

$$\Delta v^{t+1}_{i,i-1} = v^{t+1}_{i-1} - v^{t+1}_i \tag{4b}$$

$$d^{t+1}_{i,i-1} = d^t_{i,i-1} + \frac{\Delta v^t_{i,i-1} + \Delta v^{t+1}_{i,i-1}}{2} \Delta t \tag{4c}$$

where Δt is the update interval; v^t_i is the velocity of CAV i at timestep t; $d^t_{i,i-1}$ denotes the front-bumper distance between CAV i and CAV $i-1$.

2.2. Distributed control scheme

Based on the above environment setting, this section provides a generic distributed control framework to regulate CAVs'

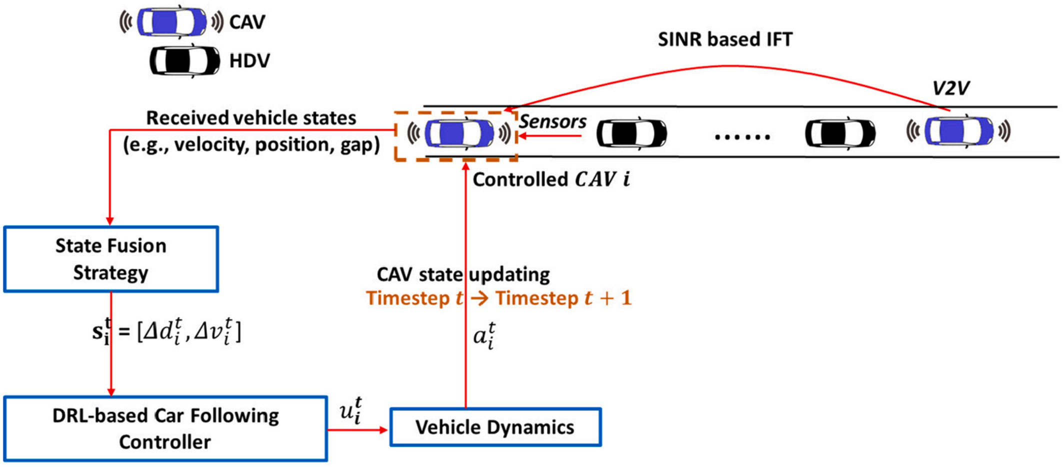

Fig. 1. The distributed control scheme for the vehicular platoon.

longitudinal movements in a mixed traffic environment, as presented in Fig. 1. The communication topology setting assumed by the SINR model in the framework is a common V2V communication topology, which is widely utilized in the CACC control (Wang et al., 2020c). For this topology, each controlled CAV (i.e., CAV i) broadcasts its state information (e.g., velocity, position) to the upstream vehicles within a certain communication range and simultaneously communicates with the multiple downstream vehicles within the communication range (i.e., at most K downstream vehicles) at each timestep for real-time longitudinal control. For each timestep, the received information from the downstream vehicles is fused as a weighted DRL state s_i^t, which will be explained with details later. After the fusion process, the DRL-based controller generates the real-time demanded acceleration u_i^t, and u_i^t is then implemented based on the above vehicle dynamics, regulating CAV i's longitudinal movements.

Within the above framework, the fused DRL state s_i^t is notably designed to better utilize the downstream vehicles' information. The communication range for state fusion, defined as the 'local downstream environment,' is restricted to cover at most K downstream vehicles. The communication quality within the range should be basically stable (i.e., rarely fails), and the kinematic traffic waves (Whitham, 1955) can be quickly propagated to the controlled CAV i. Despite the limited range, the diversified downstream CAV-HDV topologies make developing a generic distributed controller challenging. To this end, we describe any local mixed downstream environment as the generic CAV-HDVs-CAV pattern, which consists of a nearest downstream CAV followed by a single or multiple HDVs, as presented in Fig. 2(a). In this heterogeneous local environment, the traffic oscillation amplitude usually grows upstream through the consecutive HDVs between CAV i and CAV $i-m$ (Zhou et al., 2019a), which hinders CAV i from driving smoothly. To alleviate this issue, we firstly fuse the nearest downstream CAV (i.e., CAV $i-m$)' s state information to 'actively' anticipate its relatively smooth and stable driving behavior for more efficient control. Furthermore, directly modeling or predicting each HDV's stochastic behavior between CAV i and CAV $i-m$ is very challenging. To this end, we characterize the consecutive HDVs as a whole 'large' HDV (i.e., AHDV) and utilize its macroscopic traffic features to attenuate stochasticity, thus enhancing CAV i' s driving behavior. As presented in Fig. 2(b), we neglect each HDV's microscopic driving behavior between the preceding HDV (i.e., HDV $i-1$) and CAV $i-m$ and define this 'CAV-HDVs-CAV' pattern as a novel car-following structure 'CAV i → AHDV → CAV $i-m$. ' In this way, CAV i receives the real-time state information of its preceding HDV $i-1$ and the nearest downstream CAV $i-m$ to generate the fused DRL state s_i^t for the DRL-based control. It should be noted that if CAV $i-m$ is out of the communication range (i.e., $m > K$), CAV i only receives the information of HDV $i-1$ for control. The proposed distributed control scheme is downgraded to the 'decentralized control', which will be explained in Section 4.2 with details.

2.3. State fusion formulation

A generic state fusion strategy is designed based on the equilibrium concept to regulate each CAV close to the pre-defined equilibrium state and meanwhile effectively stabilize traffic oscillations. The equilibrium concept from the modern control theory defines the equilibrium state (equilibrium point) for a dynamical system, which represents a state where the system can stabilize after being affected by external disturbances or forces (Absil and Kurdyka, 2006). A system will remain at the (stable) equilibrium state once reached, given the perturbation and inputs are small enough.

In longitudinal car-following control, the equilibrium state represents the desired ideal vehicle state (i.e., equilibrium spacing and speed) during driving for each car following pair, which avoids arbitrary variation of the inter-vehicle spacing for control. Incorporating the equilibrium concept in DRL provides the exploration direction in DRL training to help develop a robust control policy and

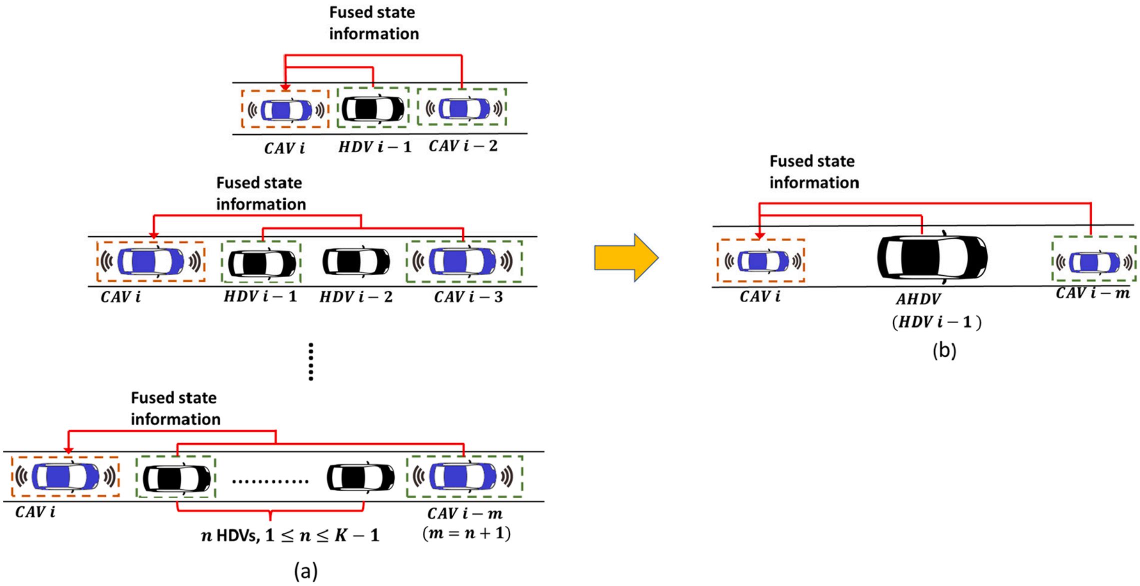

Fig. 2. The mixed local downstream environment: (a) characterized as 'CAV-HDVs-CAV' pattern; (b) "three-vehicle" car-following structure.

gives the base for analyzing vehicle string stability and car following control efficiency. Based on the concept, this subsection derives the DRL state s^t_i as the weighted deviation from the equilibrium spacing Δd^t_i and the weighted deviation from the equilibrium speed Δv^t_i regarding its downstream vehicles HDV $i-1$ and CAV $i-m$. Four parameters are predefined to fuse the DRL state, including the equilibrium spacing $d^{*t}_{i,i-1}$ and equilibrium speed $v^{*t}_{i,i-1}$ regarding HDV $i-1$; the equilibrium spacing $d^{*t}_{i,i-m}$ and equilibrium speed $v^{*t}_{i,i-m}$ regarding CAV $i-m$. The following content describes the derivations of the DRL state.

2.3.1. *Local equilibrium*

The derivation of the DRL state starts with the local equilibrium state. The local equilibrium state for each CAV car following pair follows the constant time gap (CTG) policy from the Society of Automotive Engineer Standard (SAE), which regulates the CAV to set the same speed as its preceding vehicle and maintain the preset equilibrium spacing, defined as below:

$$d^{*t}_{i,i-1} = v^t_i \tau^*_i + l_i, \tag{5a}$$

$$v^{*t}_{i,i-1} = v^t_{i-1}, \tag{5b}$$

where v^t_i denotes CAV i's real-time velocity at timestep t; τ^*_i and l_i are the constant time gap and the standstill spacing between CAV i and vehicle $i-1$, respectively. The two equations above define the local equilibrium for a car following pair.

2.3.2. *Multi-agent equilibrium*

Furthermore, to consider the impact of multiple vehicles in the local downstream environment, Eq. (5a) and Eq. (5b) are expanded to a distributed multi-agent version, whose equilibrium spacing $d^{*t}_{i,i-m}$ and speed $v^{*t}_{i,i-m}$ between CAV i and any downstream vehicle $i-m$ is specified as Eq. (6a) and Eq. (6b):

$$d^{*t}_{i,i-m} = v^t_i T^*_{i,i-m} + L_{i,i-m}, \tag{6a}$$

$$v^{*t}_{i,i-m} = v^t_{i-m}. \tag{6b}$$

$T^*_{i,i-m}$ and $L_{i,i-m}$ denote the equilibrium time gap and standstill spacing between CAV i and CAV $i-m$. The two terms regulate CAV i's desired microscopic driving behavior considering its downstream vehicles, which will be explained with mathematical details later. To facilitate the system-optimal consensus of the vehicular platoon, we measure and embed the actual spacing and speed deviations from the equilibrium between CAV i and CAV $i-m$ into the DRL framework, which is specified as:

$$\Delta d^t_{i,i-m} = d^t_{i,i-m} - d^{*t}_{i,i-m}, \tag{7a}$$

$$\Delta v^t_{i,i-m} = v^t_{i-m} - v^t_i, \tag{7b}$$

Based on Eq. (7), the equilibrium deviations for multiple downstream vehicles can be determined for the state fusion.

2.3.3. *Estimation from Newell's Car-following model*

The remaining problem lies in calculating the equilibrium spacing $d^{*t}_{i,i-m}$ in Eq. (6a), which needs to specify the corresponding time gap $T^*_{i,i-m}$ and the standstill spacing $L_{i,i-m}$. Specifically, for the defined 'CAV-AHDV-CAV' structure, the equilibrium spacing regarding HDV $i-1$ (i.e., $d^{*t}_{i,i-1}$) and CAV $i-m$ (i.e., $d^{*t}_{i,i-m}$) needs to be configured. Regarding HDV $i-1$, $d^{*t}_{i,i-1}$ is directly given in Eq. (5). Based on that, the deviation from equilibrium spacing is specified as $\Delta d^t_{i,i-1} = d^t_{i,i-1} - d^{*t}_{i,i-1}$. Regarding CAV $i-m$, the equilibrium time gap $T^*_{i,i-m}$, standstill spacing $L_{i,i-m}$, and equilibrium spacing $d^{*t}_{i,i-m}$ are defined as follows based on the three-vehicle following scheme 'CAV i → HDV $i-1$ → CAV $i-m$':

$$T^*_{i,i-m} = \tau^*_i + T^{*t}_{i-1,i-m}, \tag{8a}$$

$$L_{i,i-m} = l_i + L^t_{i-1,i-m} \tag{8b}$$

$$d^{*t}_{i,i-m} = v_i(\tau^*_i + T^{*t}_{i-1,i-m}) + \left(l_i + L^t_{i-1,i-m}\right) \tag{8c}$$

where $T^{*t}_{i-1,i-m}$ and $L^t_{i-1,i-m}$ represent the time-varying time gap and spacing between HDV $i-1$ and CAV $i-m$, respectively. Considering the aggregated HDVs in-between, $T^*_{i-1,i-m}$ can be denoted as:

$$T^{*t}_{i-1,i-m} = \sum_{j=1}^{m-1} \tau^{*t}_{i-j}, \tag{9}$$

where τ^{*t}_{i-j} denotes the time gap between HDV $i-j$ and its preceding vehicle. Since HDV has inherent stochastic nature with great diversities, τ^{*t}_{i-j} is time-varying and follows varied distributions for different HDVs. Moreover, τ^{*t}_{i-j} is unmeasurable due to the lack of communication capability of HDVs, making it challenging to determine $T^{*t}_{i-1,i-m}$. Compared with a single HDV's microscopic behavior, the aggregated HDV driving behaviors exhibit macroscopic traffic flow properties, which show less stochasticity. Thus, rather than measuring τ^{*t}_{i-j} individually, the aggregated HDV driving characteristics can be better captured by the macroscopic traffic features to address stochasticity, as indicated by the philosophy of central limit theorem (CLT) (Kwak and Kim, 2017). Though m may not be sufficiently large to apply CLT, the aggregation treatment of multiple HDVs brought promises to capture HDVs features. Moreover, we leave the remaining uncertainties by embedding the field-measured HDV trajectories in the DRL training process.

Precisely, the two time-varying terms $T^{*t}_{i-1,i-m}$ and $L^{t}_{i-1,i-m}$ need to be real-time estimated in the state fusion process to capture the macroscopic features. The schematic diagram for the real-time estimation is presented in Fig. 3(c). Newell's car following model (Newell, 2002), which bridges the fundamental diagram and microscopic driving behavior features, and meanwhile efficiently models the kinematic oscillation waves (Richards, 2013), is adopted after modification, by allowing the two time-varying terms $T^{*t}_{i-1,i-m}$ and $L^{t}_{i-1,i-m}$ to be time-variant and real-time estimated. Rather than directly modeling the microscopic driving behaviors of any CAV or HDV, the time-varying version of Newell's car-following model describes the aggregated HDVs' (AHDV's) driving behavior to capture

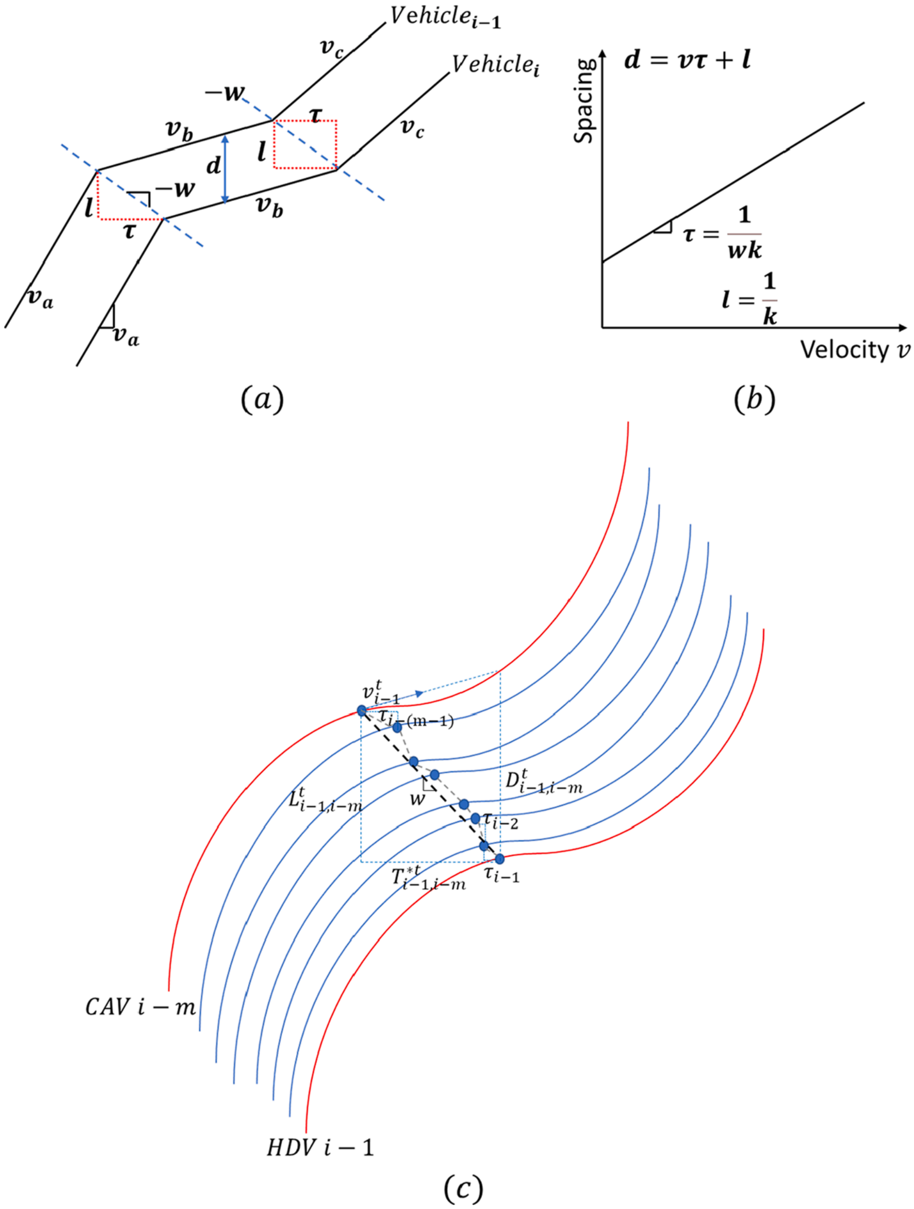

Fig. 3. Model schematic diagram: (a) Newell's car following model; (b) Speed-spacing relationship; (c) Real-time estimation diagram of the time-varying time-gap $T^{*t}_{i-1,i-m}$ and spacing..$L^{t}_{i-1,i-m}$

its exhibited macroscopic traffic features in real-time.

Fig. 3(a) and Fig. 3(b) demonstrates the principle of Newell's car-following model. From the microscopic perspective, Newell's car-following model gives a linear speed-spacing relationship in congested traffic flow for the following vehicle i, which assumes the follower reproduces the preceding leader's trajectory with a time–space displacement (τ, l):

$$d_i = v\tau + l \tag{10}$$

where v is the vehicle speed; d_i is the spacing; τ represents the time shift for vehicle i to match its leader's speed; l denotes the displacement of the speed change point. Moreover, from a macro perspective, the Newell's car following model models the kinematic wave with a triangular fundamental diagram, in which the parameters τ and l represent macroscopic traffic features to describe traffic wave speed w and jam density k:

$$w = \frac{1}{\tau k}, \tag{11a}$$

$$k = \frac{1}{l} = \frac{1}{w\tau}, \tag{11b}$$

where τ denotes the wave propagating time between two consecutive vehicles; l indicates the jam spacing. Based on Eq. (10) and Eq. (11), τ and l are two key terms for modeling the car-following behavior and simultaneously capturing the macroscopic features. Furthermore, for the vehicle following structure 'CAV-AHDV-CAV' described above, the time gap $T^{*t}_{i-1,i-m}$ and spacing $L^t_{i-1,i-m}$ can be interpreted as Newell's parameters τ and l in the car following pair 'HDV $i-1 \rightarrow$ CAV $i-m$', which anticipates the relatively smooth behavior of CAV $i-m$ and incorporates macroscopic features of the aggregated HDVs. Specifically, Eq. (12) is proposed to real-time estimate $T^{*t}_{i-1,i-m}$ and $L^t_{i-1,i-m}$ based on the integration form of Eq. (10) and Eq. (11):

$$T^{*t}_{i-1,i-m} = \frac{D^t_{i-1,i-m}}{w + v^t_{i-1}}, \tag{12a}$$

$$L^t_{i-1,i-m} = wT^{*t}_{i-1,i-m}, \tag{12b}$$

where v^t_{i-1} is the speed of HDV $i-1$ at timestep t; $D^t_{i-1,i-m}$ is the actual spacing between HDV $i-1$ and CAV $i-m$; w denotes the average kinematic wave speed, which is a pre-calibrated value determined by the road infrastructure's features and configuration. Since w plays an important role in the method, applications of our methods should regularly measure and update the w value. There are methods available for w measurement, including direct measurement w using wavelet transform (Zheng et al., 2011; Zheng and Washington, 2012), or indirect estimation by first estimating the fundamental diagram to derive w per Li et al., (2022). We set w to 16 km/h due to the generalized settings in studies using Next Generation Simulation (NGSIM) data collected on eastbound I-80 with an on-ramp at Powell Street (e.g., Laval and Leclercq, 2010; Duret et al., 2011; Chen et al., 2012).

In addition, it should be noted that $T^{*t}_{i-1,i-m}$ and $L^t_{i-1,i-m}$ are not derived from the steady-state spacing defined in Chen et al. (2012). We use the actual spacing $D^t_{i-1,i-m}$ to approximate the steady-state spacing for real-time estimation, which better anticipates the actual disturbances from downstream (e.g., sudden change in spacing) and thus achieves adaptive control performances. Moreover, the estimation method for $T^{*t}_{i-1,i-m}$ and $L^t_{i-1,i-m}$ are only suitable for heavily congested traffic conditions, where the traffic oscillations are continuously propagated upstream.

2.3.4. DRL state fusion

Based on the estimated time-varying gap $T^{*t}_{i-1,i-m}$ and spacing $L^t_{i-1,i-m}$ in Eq. (12), the equilibrium spacing $d^{*t}_{i,i-m}$ between CAV i and CAV $i-m$ in Eq. (8c) can be real-time determined. Thus, the deviation between actual spacing and equilibrium spacing is defined as $\Delta d^t_{i,i-m} = d^t_{i,i-m} - d^{*t}_{i,i-m}$. To better regulate the CAVs close to the equilibrium and reduce the DRL state dimension for greater training performance, the DRL state $s^t_i = [\Delta d^t_i, \Delta v^t_i]$ is generated by fusing the weighted equilibrium deviations of $HDVi-1$ and $CAVi-m$:

$$\Delta d^t_i = \frac{q_{i-1}\eta^t_{i,i-1}\Delta d^t_{i,i-1} + q_{i-m}\eta^t_{i,i-m}\Delta d^t_{i,i-m}}{q_{i-1}\eta^t_{i,i-1} + q_{i-m}\eta^t_{i,i-m}} \tag{13a}$$

$$\Delta v^t_i = \frac{q_{i-1}\eta^t_{i,i-1}\Delta v^t_{i,i-1} + q_{i-m}\eta^t_{i,i-m}\Delta v^t_{i,i-m}}{q_{i-1}\eta^t_{i,i-1} + q_{i-m}\eta^t_{i,i-m}} \tag{13b}$$

where weights q_{i-1} and q_{i-m} represent the information importance for HDV $i-1$ and CAV $i-m$, respectively. The coefficient for the two components is computed based on the function $q_{i-j} = \begin{cases} \frac{1}{2}, j = 1 \\ \frac{1}{2}, j = m \end{cases}$. The sum of all the weights should be equal to 1 without loss of generality. Since the information of both the two components is critical, we make it decay with 1/2 order to give equal weight to the two components based on the above function. Despite the equal weight settings in this study, the weights can be adjusted to balance the impact of the two components. The impact of both components can be summarized as follows. The preceding HDV $i-1$'s information is necessary due to safety concerns. The information $\Delta d^t_{i,i-1}$ and $\Delta v^t_{i,i-1}$ from HDV $i-1$ should always be incorporated into the fused DRL state to enhance safety. In particular, it regulates the local equilibrium spacing deviation $\Delta d^t_{i,i-1} \rightarrow 0$ and relative speed $\Delta v^t_{i,i-1} \rightarrow 0$, which significantly lowers the driving risks as manifested by safety surrogate measures such as time-to-collision (TTC) (Jiménez et al., 2013). On the other hand, the information from CAV $i-m$ is the key part of the state fusion process, which anticipates the relatively smooth driving behaviors from CAV $i-m$ and alleviates the HDVs' stochasticity to facilitate control performances.

For a better understanding, the above-proposed estimation method and state fusion process can be interpreted in this way. The controlled CAV i adapts its car-following strategy according to the state of its actual immediate preceding vehicle HDV $i-1$ and the state of the fictive vehicle (i.e., AHDV) in front of the controlled CAV. To refine the effective position of the fictive vehicle, the Newell-based methodology is introduced to estimate the traffic state and propagate the current state of the leader CAV $i-m$ through a set of HDV vehicles. Specifically, the time-varying terms $T^{*t}_{i-1,i-m}$ and $L^t_{i-1,i-m}$, which are real-time estimated from Newell's model, determine the position of the fictive vehicle. Adopting the interpretation makes the introduced method similar to Multi-Anticipative ACC car-following rules (Lin et al., 2012; Wang et al., 2014a, 2014b). Rather than defining a fictive vehicle like in MA-ACC rules only based on perception and communication sensors, which provide information regarding the immediate leader HDV $i-1$ and the CAV leader CAV $i-m$, the introduced method refines the approach by making use of the Newell's car-following model to capture the macroscopic traffic features between the two leaders.

From the equilibrium concept perspective, the control design maintains the CAV in the pre-defined equilibrium state, considering its preceding HDV and the nearest downstream CAV. The equilibrium state with the preceding HDV (i.e., $d^{*t}_{i,i-1}$, v^t_{i-1}) ensures local stability (Willems, 2013), representing the capability to remain in a car-following pair of equilibrium under disturbances. The equilibrium state with the nearest downstream CAV (i.e., $d^{*t}_{i,i-m}, v^t_{i-m}$) incorporates the relatively stable driving motion of the downstream CAV and the macroscopic traffic features of aggregated HDVs, which further enhances the car-following performances. Moreover, this control design is generic since it is suitable for diversified compositions of the mixed local downstream environment.

2.3.5. Extension to the full CAV environment

It should be noted that the local communication range can be a pure connected automated environment, in which consecutive downstream vehicles are CAVs (i.e., CAV-CAVs patten), as presented in Fig. 4. The above generic state fusion approach can also be applied to this full CAV condition, whose control design is to achieve a platoon-level consensus by fusing received information from all the aggregated CAVs (i.e., m CAVs ,$1 \leq m \leq K$) as the DRL state s^i_t for control (Shi et al., 2022). Specifically, the fusion process follows the same formulations from Eq. (5) to Eq. (7) to calculate the equilibrium speed deviation $\Delta v^t_{i,i-m}$ and spacing deviation $\Delta d^t_{i,i-m}$ between $CAVi$ and $CAVi-m$. Due to the full CAV environment, the equilibrium time gap $T^*_{i,i-m}$ and equilibrium standstill spacing $L_{i,i-m}$ in Eq. (6) can be directly defined in a multi-agent version, which means $T^*_{i,i-m} = m\tau^*_i$; $L_{i,i-m} = ml_i$. Similar to Eq. (13), the fused DRL state $s^t_i = [\Delta d^t_i,\ \Delta v^t_i]$ incorporates the weighted equilibrium deviations for the m aggregated CAVs to anticipate the disturbances induced downstream and thus achieve great system-level consensus:

$$\Delta d^t_i = \frac{q_{i-1}\eta^t_{i,i-1}\Delta d^t_{i,i-1} + q_{i-2}\eta^t_{i,i-2}\Delta d^t_{i,i-2} + \cdots + q_{i-k}\eta^t_{i,i-m}\Delta d^t_{i,i-m}}{q_{i-1}\eta^t_{i,i-1} + q_{i-2}\eta^t_{i,i-2} + \cdots q_{i-m}\eta^t_{i,i-m}}, \tag{14a}$$

$$\Delta v^t_i = \frac{q_{i-1}\eta^t_{i,i-1}\Delta v^t_{i,i-1} + q_{i-2}\eta^t_{i,i-2}\Delta v^t_{i,i-2} + \cdots + q_{i-k}\eta^t_{i,i-m}\Delta v^t_{i,i-m}}{q_{i-1}\eta^t_{i,i-1} + w_{i-2}\eta^t_{i,i-2} + \cdots q_{i-k}\eta^t_{i,i-m}}, \tag{14b}$$

where the coefficient $q_{i-j} = \begin{cases} \frac{1}{2^j}, 1 \leq j \leq m-1 \\ \frac{1}{2^{j-1}}, j = m \end{cases}$ represents that the closer CAV is assigned with greater power on the control decision.

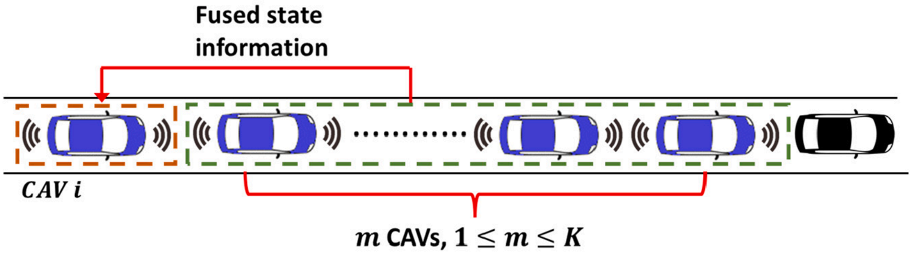

Fig. 4. Scenario extended to the full CAV downstream environment.

3. Development of Drl-Based controller

Based on the defined control scheme in Section 2, this section develops the DRL-based controller. We discuss the detailed DRL scheme (Section 3.1), the adopted DRL algorithm (Section 3.2), and the training process (Section 3.3). The simulation experiments, including training and evaluation, are performed via Python. TensorFlow package is used to build the DRL algorithm. Pyomo package is applied to develop the MPC-based controller for performance comparison in the experimental part (Section 4).

3.1. DRL scheme and formulation

The basis of DRL is Markov Decision Process (MDP), in which the DRL agent (i.e., CAV controller) and environment (given in Section 2) interact with each other based on four basic elements: state, action, policy, and reward ($\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{\pi}, \boldsymbol{r}$). As discussed in the previous section, state $\boldsymbol{s}$ represents the fused state information, which contains the weighted deviations of spacing Δd_i^t and speed Δv_i^t, denoted as $\boldsymbol{s}_i^t = [\Delta d_i^t, \Delta v_i^t]$. The DRL agent receives $\boldsymbol{s}_i^t$ at each timestep and outputs the action $\boldsymbol{a}$ (i.e., the control signal u_i^t) based on the policy $\boldsymbol{\pi}$ to regulate CAV i's longitudinal movements. As an implicit function that assigns the action probability for each state, policy $\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s})$ needs to be updated to achieve optimal control performance through the training process.

The reward $\boldsymbol{r}$ determines the control objectives. In our design, the objectives of the car following control efficiency, which aims to maintain CAV in the pre-defined equilibrium state, and driving comfort, which pursues a smoother driving behavior with greater eco-driving performance, are incorporated in the DRL framework. In particular, the cost of the car following control efficiency c_i^t is defined as the quadratic form of deviation from the equilibrium state, which is a common control design in modern control theories such as Linear Quadratic Regulator (LQR) and MPC. This design facilitates stability analysis, as manifested by numerous control papers (e.g., Fisher and Bhattacharya, 2009; Zhou et al., 2019b). Specifically, the quadratic cost c_i^t is defined as:

$$c_i^t = (\boldsymbol{s}_i^t)^T Q_i \boldsymbol{s}_i^t, \tag{15}$$

where $Q_i = \begin{bmatrix} \alpha_{1,i} & \\ & \alpha_{2,i} \end{bmatrix}$ is a positive definite diagonal coefficient matrix with weights $\alpha_{1,i} > 0$ and $\alpha_{2,i} > 0$.

Further, the driving comfort cost g_i^t suggested by Wang et al., (2014a) is defined in Eq. (16), which alleviates the acceleration to improve the eco-driving performance and empirical string stability (i.e., acceleration energy) (Shi et al., 2021b; Feng et al., 2019):

$$g_i^t = \alpha_{3,i}\left(a_i^t\right)^2, \tag{16}$$

where $\alpha_{3,i}$ denotes the weight for driving comfort. It should be noted that both the two costs regarding the car following control efficiency cost c_i^t and driving comfort cost g_i^t are unitless. The weight's unit is the reciprocal of its valuable unit. Thus, each weight of $\alpha_{1,i}$, $\alpha_{2,i}$, and $\alpha_{3,i}$ offsets the units of its variables, making the whole cost unitless.

Combining the two control objectives above, the running cost e_i^t of CAV i at timestep t is defined as the sum of the car following efficiency cost c_i^t and driving comfort cost g_i^t:

$$e_i^t = c_i^t + g_i^t. \tag{17}$$

Since the quadratic running cost e_i^t is similar to the cost function in the constrained optimization framework, and the training environment is similar to the state space as Zhou et al. (2019b), the coefficients setting ($\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$) is set to be same as Zhou et al. (2019b) to improve the string stability performance further. Though it is prohibitive to conduct mathematical string stability as Zhou et al. (2019b), due to the intrinsic complexity of DRL, we envision the setting coefficients in the same fashion as is helpful to enhance the string stability by the similarities mentioned above.

Based on the above systematic cost design, we convert the running cost e_i^t as the immediate reward r_i^t using the exponential function, as shown in Equation (18), which calculates the reward value as feedback for the control action at each timestep. The exponential equation serves the following two purposes. First, the reward value needs to be maximized in the DRL framework, whose optimization direction is opposite to the cost function of optimal control. Using the exponential function changes the optimization direction from minimization to maximization. Second, the above exponential function plays a normalization function, normalizing the immediate reward r_i^t within the boundary [0, 1] and further enhancing the training performance.

$$r_i^t = \exp\left(-e_i^t\right) \tag{18}$$

The above exponential function also normalizes the immediate reward r_i^t within the boundary [0, 1] to enhance the training performance. With the reward value, an infinite-horizon optimal control problem is formulated for maximizing the discounted cumulative rewards to find the optimal control policy π^*:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \sum_{m=0}^{\infty} \Upsilon^m r_i^{t+m}, \tag{19}$$

where Υ is the discount factor .

3.2. Policy update algorithm

The DRL solves the optimization problem in Eq. (20) by continuously updating policy π in training. The choice of DRL algorithm for the CAV control is based on the following aspects: (i) the action space (discrete/continuous); and (ii) the algorithm performance. The algorithm should support continuous action space for the instance of the microscopic CAV control with great sampling efficiency and converging performances. The Distributed Proximal Policy Optimization (DPPO) algorithm (Heess, 2017), one of the Actor-Critic DRL algorithms, is adopted for policy updating in training. The Actor-Critic DRL algorithm combines the merits of Policy-Based RL and Value-Based RL algorithms, which performs faster than traditional RL algorithms and supports continuous action space in the training process. Based on its merits, the Actor-Critic framework is widely used in the most popular reinforcement learning algorithms, such as the A3C algorithm (Mnih et al., 2016), DDPG algorithm (Lillicrap et al., 2016), and PPO (DPPO) algorithm (Heess, 2017). In this paper, we adopted the DPPO algorithm to update policy due to its great balance between sampling efficiency, implementation simplicity, and converging performance (Schulman et al., 2017). Compared with traditional policy gradient RL algorithms, the DPPO algorithm makes policy gradient less sensitive to a large step and improves the convergence of policy updates by clipping the divergence of the strategy update. Besides, the DPPO algorithm updates the policy of the global agent in parallel through multiple parallel agents, which further improves training efficiency.

The Distributed Proximal Policy Optimization (DPPO) algorithm (Heess, 2017) is adopted for policy updating due to its great balance between sampling efficiency, implementation simplicity, and converging performance. The DPPO algorithm is a typical Actor-Critic DRL algorithm, with objective $L^{CLIP}(\theta)$ updating in the actor network and critic loss $L_c(\Phi)$ updating in the critic network. The overall actor-network framework with network structures is presented in Fig. 5. The detailed hyperparameters settings are demonstrated in Table 1. The number of neurons for the actor network (200) and critic network (100) is tuned by experiences to achieve the desired performances without causing underfitting and overfitting issues. Since the actor network learns a more complex policy function that maps the DRL state to a probability distribution over all actions, thus setting with more neurons in this paper (Grondman et al., 2012).

3.2.1. Actor network

The actor network determines the policy π with parameter θ. It receives the DRL state s_i^t as the input and outputs a probability distribution over actions. The control signal u_i^t is then sampled from the distribution. For the network structure, there is one hidden layer with 200 neurons, and the ReLu function is adopted as the activation function for the output. The actor network is updated by maximizing the objective function $L^{CLIP}(\theta)$:

$$L^{CLIP}(\theta) = \widehat{E}_t[\min(p_t(\theta)\widehat{A}_t, clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)\widehat{A}_t] \tag{20}$$

where $p_t(\theta) = \frac{\pi_\theta(a_t|\mathbf{s}_t)}{\pi_{\theta_{old}}(a_t|\mathbf{s}_t)}$ identifies the probability ratio of the new policy $\pi_\theta(a_t|\mathbf{s}_t)$ and old policy $\pi_{old}(a_t|\mathbf{s}_t)$. The clipping function $clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)$ function restricts $p_t(\theta)$ between $1-\varepsilon$ and $1+\varepsilon$ to limit the update range of new policy, making the policy gradient less sensitive to the step size and improving the convergence. Ɛ is the clipping parameter. $\widehat{A}_t$ is the estimated advantage at state s_i^t,

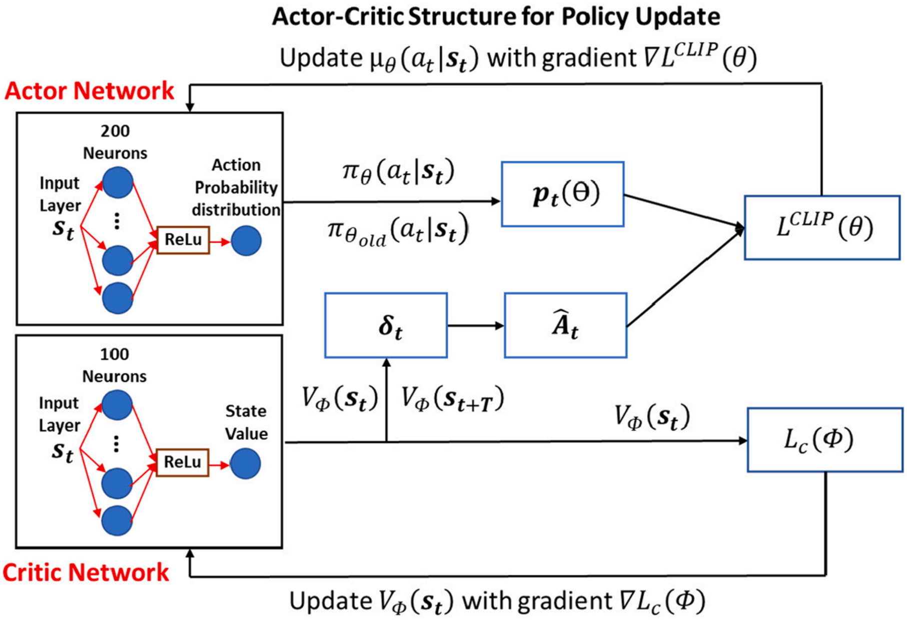
Actor-Critic Structure for Policy Update
Update $\mu_\theta(a_t|\boldsymbol{s_t})$ with gradient $\nabla L^{CLIP}(\theta)$
Actor Network
200 Neurons
Input Layer $\boldsymbol{s_t}$
ReLu
Action Probability distribution
$\pi_\theta(a_t|\boldsymbol{s_t})$
$\pi_{\theta_{old}}(a_t|\boldsymbol{s_t})$
$\boldsymbol{p_t}(\Theta)$
$L^{CLIP}(\theta)$
$\boldsymbol{\delta_t}$
$\boldsymbol{\widehat{A}_t}$
100 Neurons
Input Layer $\boldsymbol{s_t}$
ReLu
State Value
$V_\Phi(\boldsymbol{s_t})$
$V_\Phi(\boldsymbol{s_{t+T}})$
$V_\Phi(\boldsymbol{s_t})$
$L_c(\Phi)$
Critic Network
Update $V_\Phi(\boldsymbol{s_t})$ with gradient $\nabla L_c(\Phi)$

Fig. 5. The actor-critic structure of the policy iteration algorithm.

Table 1
Hyper parameters of the DPPO algorithm.

Hyperparameter	Value
clipping value ε	0.2
minibatch T	256
discount factor Y	0.99
Hidden layer of actor	1
Hidden layer of critic	1
actor network neurons	200
critic network neurons	100
parallel worker numbers	4
actor learning rate	0.00001
critic learning rate	0.00001

which is provided from the critic network:

$$\widehat{A}_t = R_t - V_\Phi\left(s_i^t\right), \tag{21}$$

where $V_\Phi\left(s_i^t\right)$ is the value estimated from the critic network; R_t denotes the discounted sum of rewards in T steps at state s_i^t:

$$R_t = \sum_{m=0}^{T-1}\gamma^m r_i^{t+m} + \gamma^T V_\Phi\left(s_i^{t+T}\right), \tag{22}$$

where T is the minibatch size; r_i^{t+m} is the immediate reward defined in Eq. (18); γ is the discount factor. Therefore, the parameter θ of the actor network is updated based on the gradient of $L^{CLIP}(\theta)$ with learning rate α_θ:

$$\theta = \theta - \alpha_\theta \nabla L^{CLIP}(\theta). \tag{23}$$

3.2.2. Critic network

On the other hand, the critic network with parameter Φ evaluates the decision u_i^t output by the actor network. The critic network receives the DRL state s_i^t as the input and outputs the estimated state value $V_\Phi\left(s_i^t\right)$. For the network structure, there is one hidden layer with 100 neurons, and the ReLu function is used as the activation function for the output. The critic network is updated by minimizing the critic loss function $L_c(\Phi)$:

$$L_c(\Phi) = \widehat{E}_t\left(V_\Phi\left(s_i^t\right) - R_t\right)^2 \tag{24}$$

where Temporal Differences (TD) error δ_t is denoted as $\delta_t = V_\Phi\left(s_i^t\right) - R_t$ in the loss function. The TD error δ_t estimates the advantage value $\widehat{A}_t$ in actor since $\delta_t = -\widehat{A}_t$. Thus, the parameter Φ is iteratively optimized based on the gradient $\nabla L_c(\Phi)$ with learning rate α_Φ: $\Phi = \Phi - \alpha_\Phi \nabla L_c(\Phi)$.

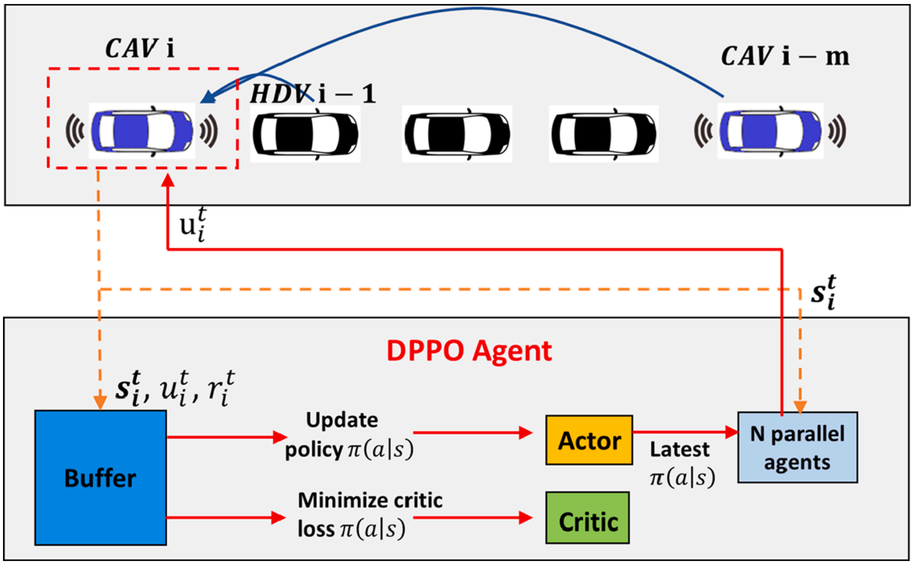

Fig. 6. The schematic diagram of training framework.

3.3. Training procedure and results

Based on the proposed DRL scheme (given in Section 3.1) and the adopted policy updating algorithm (given in Section 3.2), this section describes the detailed training procedure in which the DPPO agent continuously interacts with the simulation environment. The DPPO agent consists of one global agent for the actor-critic network updating and multiple parallel agents interacting with their independent simulation environments for data collecting (e.g., state, action, reward), which further improves the sampling efficiency and training speed. The detailed training process is demonstrated in Fig. 6. At timestep t, each parallel agent receives the CAV i's state s_i^t of its corresponding simulation environment and outputs the control signal u_i^t to update the longitudinal movement based on the current policy $\pi(a|s)$. Concurrently, the collecting data, including the calculated reward r_i^t, state s_i^t, and action u_i^t, is sent to the memory buffer for storage. The update of the policy and actor-critic network is triggered after a certain batch of data is stored in the memory buffer.

Regarding the training environment settings, ten sets of five-vehicle ground-truth vehicular platoon trajectories with a time length of 334 timesteps from NGSIM datasets are embedded for the initial configuration. For each training episode, one of the ten platoon trajectories is randomly sampled and assigned for the trajectories of the leading vehicle CAV $i-m$ and the following aggregated HDVs. Fig. 7 below shows details of one platoon trajectory. The platoon trajectories that incorporate typical phases of acceleration, deceleration, and uniform speed are embedded in the training process for the DRL control model to better capture stochastic driving characteristics. It should be noted that the proposed "CAV-AHDV-CAV" structure is a generic unit in the mixed traffic flow, whose leading CAV's driving behavior could be varied and even stochastic due to the impact of the downstream traffic. Thus, rather than setting a deterministic leading CAV, using NGSM data to represent the stochastic leading vehicle behavior should be rational.

Specifically, the number of HDVs $n(1 \leq n \leq K-1)$ between the two CAVs is randomly sampled in the simulation environment at each training episode to enhance the generalized capability for different topologies. Based on sampled topology, the trajectories of the leading CAV $i-m$ and following HDVs are assigned from the above NGSIM platoon data. Taking an example of the topology 'CAV-three HDVs-CAV', the NGSIM leader trajectory and three follower trajectories are assigned to CAV $i-m$ and three HDVs. For the controlled CAV i, it starts with the initial equilibrium state defined in Section 2 and is then controlled by the DRL learning agent without the loss of generality.

For the scenario where the local downstream environment is pure connected and automated (i.e., full CAVs), only the trajectory of the leading vehicle is from the NGSIM dataset, and the corresponding DRL-based models control the other downstream CAVs.

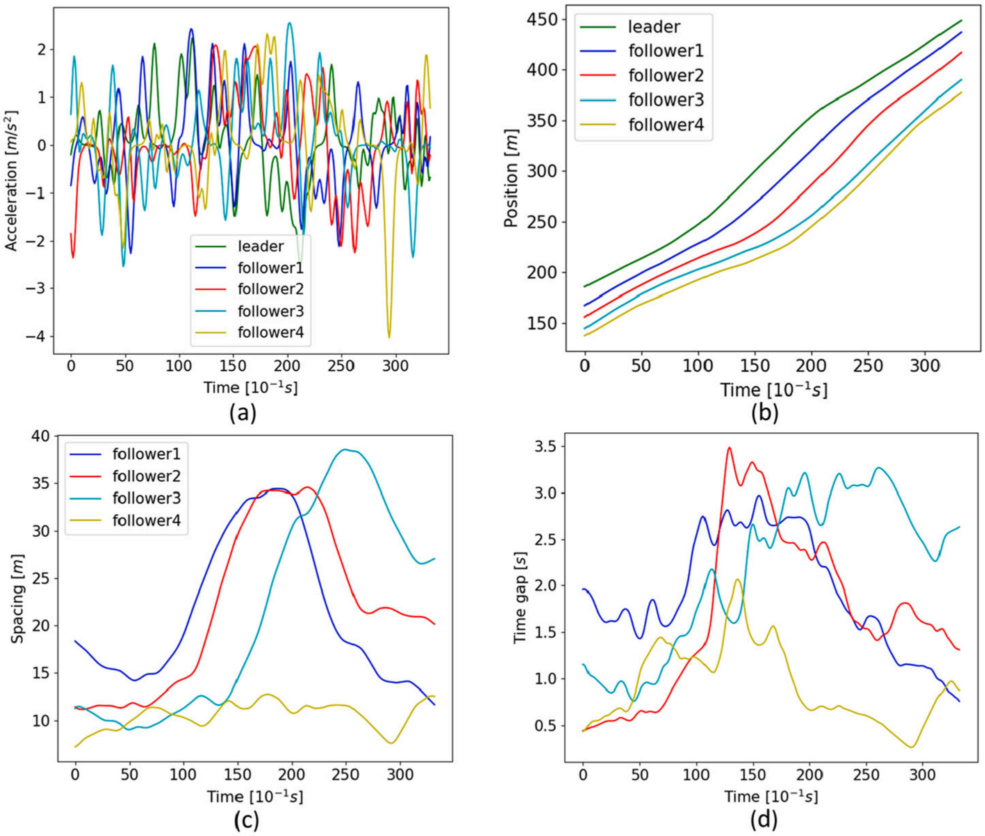

Fig. 7. Trajectories of the vehicular platoon in the training process regarding (a) acceleration; (b) position; (c) spacing; and (d) time gap (g_i^t/v_i^t).

Training results are represented by the moving reward trajectory (Qu et al., 2020) demonstrated in Fig. 8. For the mixed (heterogeneous) traffic environment (Fig. 8(a)), the training platoon trajectories and number of HDVs are randomly sampled, which leads to a varied mixed environment. Thus, the rewards fluctuate in the converging area. On the other hand, the full CAV downstream environment leads to more stable reward trajectories, as presented in Fig. 8(b). In general, the rewards for both cases monotonically increase until convergence, indicating good converging performance.

4. Simulation experiments

4.1. Experiment settings

4.1.1. Experiment overview

After developing the DRL-based control models, we conduct several numerical experiments to evaluate the control approach using NGSIM datasets of I-80 in California. To remove the noises and handle the missing data, we reconstruct the datasets using a low-pass filter proposed by (Punzo et al., 2011) and (Montanino and Punzo, 2015). The trajectories in Lane 2 from 4:00 pm to 4:15 pm are selected for experiments and analysis due to the frequent traffic congestions and oscillations and the fewer lane-changing maneuvers. The default experimental setting is shown in Table 2.

The simulation experiments can be divided into three parts: (i) model performance analysis; (ii) application of the proposed model in a long vehicular platoon with different CAV penetration rates; (iii) generalization capability validation. Based on these experiments, the effectiveness, robustness, and generalization of the proposed control strategy are analyzed comprehensively. For the simulated platoons in these experiments, the leader's trajectory is picked from NGSIM datasets or the customized trajectory profile to reproduce traffic disturbances. The initial states of followers start with the pre-defined equilibrium states or start with the ground-truth NGSIM data. With the leader profile and followers' initial states, the vehicular platoon trajectory can be simulated based on the proposed model or other compared methods.

4.1.2. Evaluation metrics

There are several performance indicators for quantitatively evaluating the control performance: cumulative dampening ratio $d_{p,i}$, local stability measured by $\Delta d_{i,i-1}^t$ and Δv_{i-1}^t, driving comfort g_i^t (given in Eq. (16)), and average velocity $\overline{v}_i$. The cumulative dampening ratio $d_{p,i}$ quantifies the empirical string stability, an important property that measures the capability of the CAV controller in dampening traffic oscillations. The traffic oscillation magnitude is reduced or remains the same as it goes through a string stable CAV. Specifically, the l_2-norm acceleration dampening ratio $d_{p,i}$ (Ploeg et al., 2014) is specified as:

$$d_{p,i} = \frac{\|a_i^t\|_2}{\|a_0^t\|_2} = \frac{\left(\sum_{t=0}^{N} |a_i^t|^2\right)^{\frac{1}{2}}}{\left(\sum_{t=0}^{N} |a_0^t|^2\right)^{\frac{1}{2}}}, \tag{25}$$

where N denotes the time length; i is the vehicle index. Index 0 represents the leader of the whole vehicular platoon. The smaller dampening ratio $d_{p,i}$ indicates that the disturbances are dampened to a greater extent, leading to a more string stable driving behavior. The local stability is another important property of CAV longitudinal control, denoting a vehicle's ability to remain in the equilibrium state with its immediate preceding vehicle (Willems, 2013). The deviations from equilibrium spacing $\Delta d_{i,i-1}^t$ and equilibrium speed Δv_{i-1}^t regarding vehicle $i-1$ are the indicators for local stability. Great local stability with low equilibrium deviations indirectly guarantees driving safety since it leads to large time-to-collision (TTC) (Minderhoud and Bovy, 2001). The average velocity $\overline{v}_i$ refers to

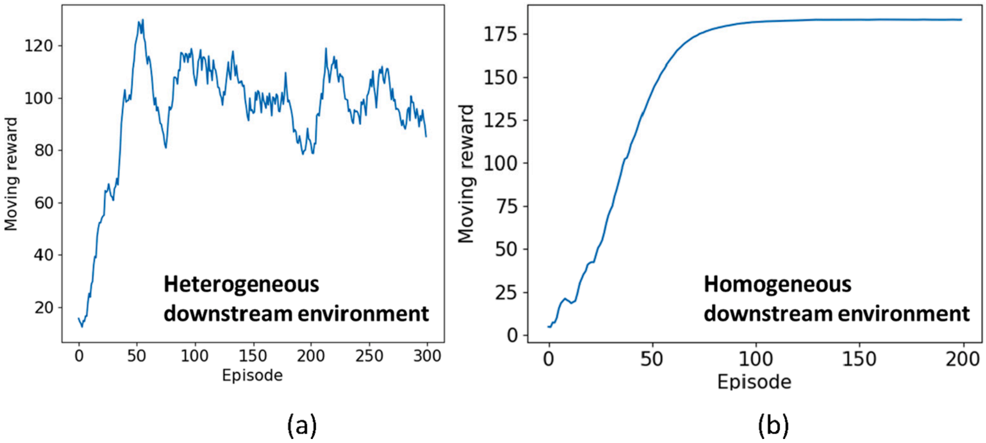

Fig. 8. The moving reward trajectory results for mixed traffic environment (a) and homogeneous traffic environment (b).

Table 2
Default parameters for the experimental setting.

Parameters	Value
number of local downstream vehicles K	5
update interval Δt	0.1 s
vehicle length l_v	4.6 m
standstill spacing l_i	6.4 m
constant time gap τ_i^*	1 s
SINR threshold β	0.01
control demand ratio $K_{i,L}$	1
actuation time lag $I_{i,L}$	0.1
noise term $N(\mu, \sigma^2)$	$N(0, 0.1)$
$\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$	$1\left(\frac{1}{m^2}\right), 0.5\left(\frac{s^2}{m^2}\right), 0.5\left(\frac{s^4}{m^2}\right)$
$[a_{i,min}, a_{i,max}]$	[-4 m/s^2, 4 m/s^2]
free flow speed v_f	33.3 m/s (120 km/h)
wave speed w	4.4 m $/s$ (16 km/h)

the mean velocity of all timesteps ($\overline{v}_i = \frac{\sum_{t=0}^{N} v_i^t}{N}$).

4.1.3. HDV modeling method

Furthermore, the precise modeling of HDV driving behavior in the mixed traffic simulation contributes to a more realistic simulation environment and convincing results. This paper uses a calibrated Intelligent Driver Model (IDM) (Kesting and Treiber, 2008; Treiber et al., 2000), which can be representative of HDV's string instability property, to model the HDV behaviors in the experiments. The IDM parameters are calibrated by (Kesting and Treiber, 2008) using ground-truth datasets of HDV behaviors. The calibrated datasets show complex situations of daily city traffic with several accelerations, decelerations, or standstill periods, which is quite similar to the adopted NGSIM datasets for experiments. Therefore, the calibrated IDM model can be applied in the experiments of this paper. The calibrated parameters are presented in Table 3.

4.2. Control performance evaluation

For the first part of the experiments, this section analyzes the performance of the proposed distributed control strategy. The proposed distributed control performance is analyzed in the mixed local environment compared with the following state-of-art CAV controllers as comparisons:

- **Decentralized DRL-based controller.** The decentralized controller, also developed by the DRL, downgrades the CAV to the autonomous vehicle (AV) that can only receive its immediate preceding vehicle's information through onboard sensors. The decentralized DRL control model is developed using the same methodology (i.e., same reward design and training data) given in Section 3. The only difference lies in that the decentralized DRL state is defined as the local equilibrium deviations regarding the immediate preceding vehicle (i.e., $s_i^t = [\Delta d_{i,i-1}^t, \Delta v_{i,i-1}^t]$).
- **Linear-based CACC controller.** The compared linear-based CACC controller (Zhou et al., 2019a) is based on the constant time gap (CTG) policy, which has been proved to have excellent traffic oscillation dampening performances and guaranteed string stability performance.
- **MPC-based CACC controller.** The compared MPC-based CACC controller (Wang et al., 2016) has explicit constraints of velocity and acceleration to meet restrictions of the vehicle kinematics. The cost function is designed to achieve control efficiency and driving comfort criteria. Similarly, the CTG policy is incorporated into the control model to enhance the empirical string stability performances.

The simulated mixed platoon follows a topology $T_p = \{1', 0, 1, 0, 0, 1, 0, 0, 0, 1\}$, where 1' represents the leading CAV of the platoon with its trajectory from the NGSIM dataset; 0 denotes the simulated HDV follower; 1 denotes the simulated CAV follower. Each fol-

Table 3
Calibrated Parameters of IDM.

Variable	Parameter	Values
V_0	Desired velocity	33.3 m/s
T	Safe time headway	1.12 s
a	Maximum acceleration	1.23 m/s^2
b	Comfortable Deceleration	3.2 m/s^2
sigma	Acceleration exponent	4
S_0	Minimum distance	2.3 m

lower starts with the pre-defined equilibrium state. This topology provides the typically mixed traffic local environment, in which driving behaviors of CAV 2, CAV 5, and CAV 9 are determined by the proposed distributed control approach. The mixed platoons generated from the decentralized control and linear-based CACC strategy follow the same topology.

Fig. 9 presents the position, velocity, and acceleration trajectory results under DRL-based distributed control (Fig. 9(a)) and DRL-based decentralized control (Fig. 9(b)). The leading CAV's trajectory (black trajectory) shows frequent acceleration-deceleration waves and a short standstill period. For the mixed platoon under the decentralized control strategy (Fig. 9(b)), the HDV tends to amplify the traffic oscillations due to the long reaction time, aggravating the traffic jam. Compared with HDVs, the decentralized CAVs are more responsive to their leaders with smaller spacing, showing efficient car-following behaviors. However, the decentralized CAVs only slightly dampen the traffic disturbances since they are separated and distributed in the mixed platoon. The decentralized CAV makes the control decision only based on its preceding HDV, whose driving behavior is negatively affected by the propagated traffic disturbances. Thus, the decentralized CAV is hard to diminish the disturbances in this mixed traffic scenario.

On the other hand, the distributed CAVs, as presented in Fig. 9(a), also demonstrate responsive driving behaviors with smaller spacings compared with the HDVs. Moreover, the distributed CAV can dampen the traffic oscillation significantly, showing great string stability. The reason is that the downstream CAV's driving state and macroscopic traffic flow property of the aggregated HDVs are

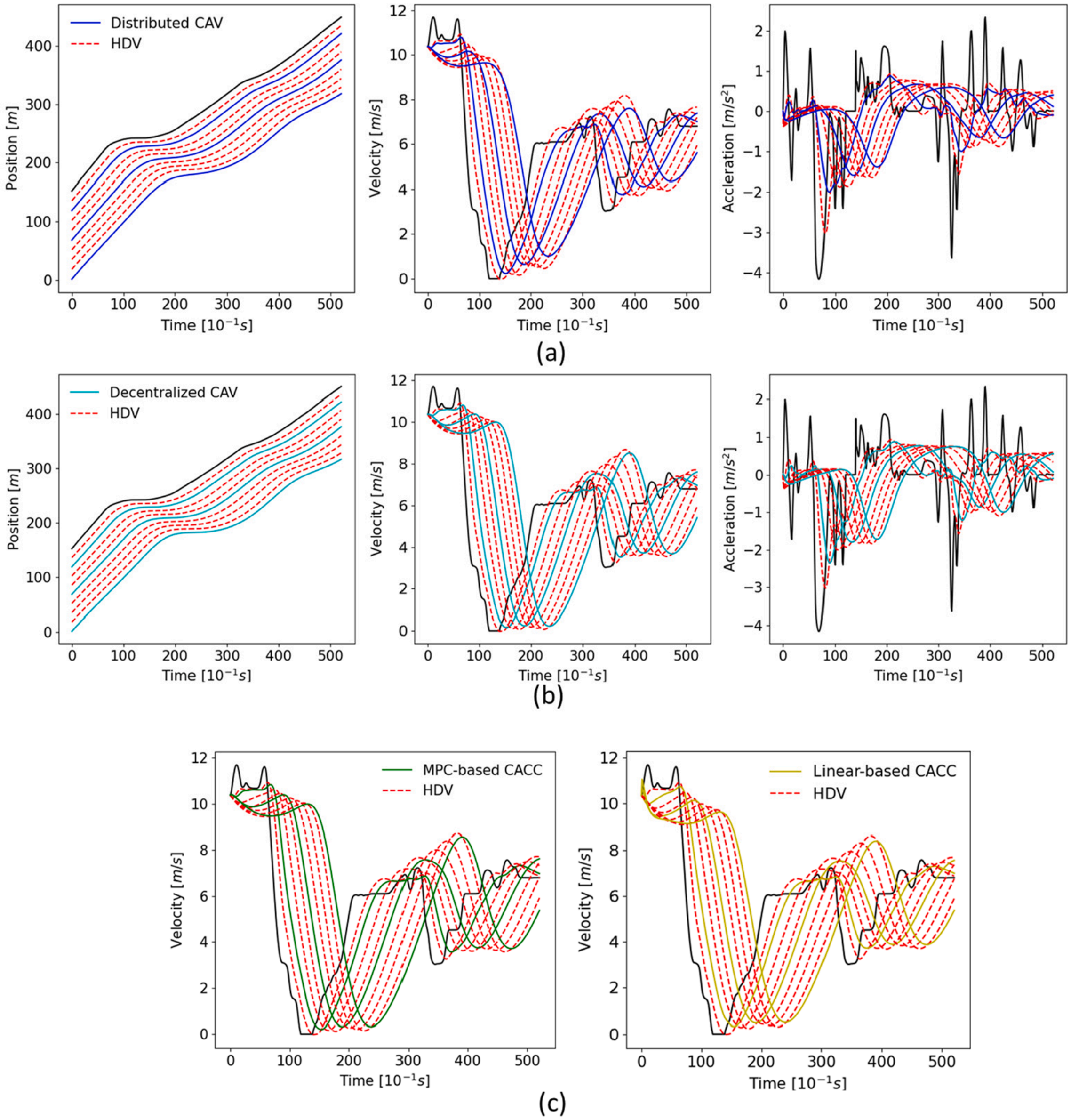

Fig. 9. The position, velocity and acceleration trajectory results comparison: (a) distributed control; (b) decentralized control; (c) MPC-based CACC and Linear-based CACC velocity trajectories.

conveyed into the DRL control framework, which enhances the car following performance and better optimizes the entire mixed traffic flow. Fig. 9(c) gives the velocity portfolio of the linear-based CACC controller and MPC-based CACC controller. Compared with these two approaches, the distributed DRL-based control can alleviate the propagated oscillations to a more significant extent. The performance of these approaches can be differentiated around the inflection point of the acceleration-deceleration process (e.g., timestep 280, timestep 320). The underlying reason is that the DRL can better capture leading HDV characteristics and stochasticity with the proposed ‘CAV-AHDV-CAV’ structure and ground truth training dataset. Whereas Zhou et al. (2019a) focused on the frequency predominant range, Wang et al. (2016) focused on the formation and propagation of moving jams, which may lose some nuanced characteristics of leading HDV behaviors.

The quantified performance indicators of the nine followers from the distributed control, decentralized control, and CACC strategy are shown in Fig. 10, respectively, in which we focus on each vehicle’s average performance. In general, the mixed platoon under distributed control framework greatly outperforms the decentralized control-based mixed platoon in terms of string stability and driving comfort. The performance of the linear-based CACC strategy is more akin to the proposed distributed DRL-based approach, while it scarifies certain performances in velocity. Particularly, the three distributed CAVs (CAV 2, CAV 5, and CAV 9) differentiate their performances from other strategies, in which the most upstream CAV 9 has the greatest advantage. This indicates that the more HDVs between the controlled CAV and the downstream CAV, the distributed control can better dampen traffic oscillations and have higher advantages than other strategies. Specifically, compared with the decentralized CAV 9, the distributed CAV 9 can reduce a 20.23% dampening ratio, 36.38% driving comfort cost, and increase by 0.52% average velocity. Regarding the local stability, Fig. 11 demonstrates the trajectories of equilibrium spacing $\Delta d_{i,i-1}^t$ and equilibrium speed $\Delta v_{i,i-1}^t$ of vehicle $i-1$. The equilibrium deviations of the CAVs are within a relatively small range (i.e., $-2.5\ m/s$ to $0.8\ m/s$ for Δv_{i-1}^t; $-1.6m$ to $2.4m$ for $\Delta d_{i,i-1}^t$), which indicates that local stability is achieved empirically.

4.2.1. Evaluation under the extreme scenario

Furthermore, we conduct an experiment under an extreme traffic scenario to validate the robustness of the proposed controller. The vehicular platoon topology has the same topology as the previous experiment, while the leading vehicle profile is customized with one rapid deceleration-stall-acceleration cycle ($-2.4m/s^2 \rightarrow 0ft/s^2 \rightarrow 1.5m/s^2$) traffic oscillation. Fig. 12(a) gives the position, velocity, and acceleration for the proposed DRL-based controller, and Fig. 12 (b) shows the velocity for the other three compared approaches. Similarly, the aggregated HDVs amplify traffic oscillations while the distributed CAVs greatly dampen traffic oscillations with stability-wise performances. The quantified results of indicators are presented in Fig. 13, which shows the proposed DRL-based control has manifest advantages over other approaches regarding string stability and driving comfort.

4.2.2. Evaluation under communication failure

Although we assume the communication quality within the communication range should be stable (i.e., communication rarely fails), the communication loss could happen at a certain period when the communication distance is relatively far (e.g., four aggre-

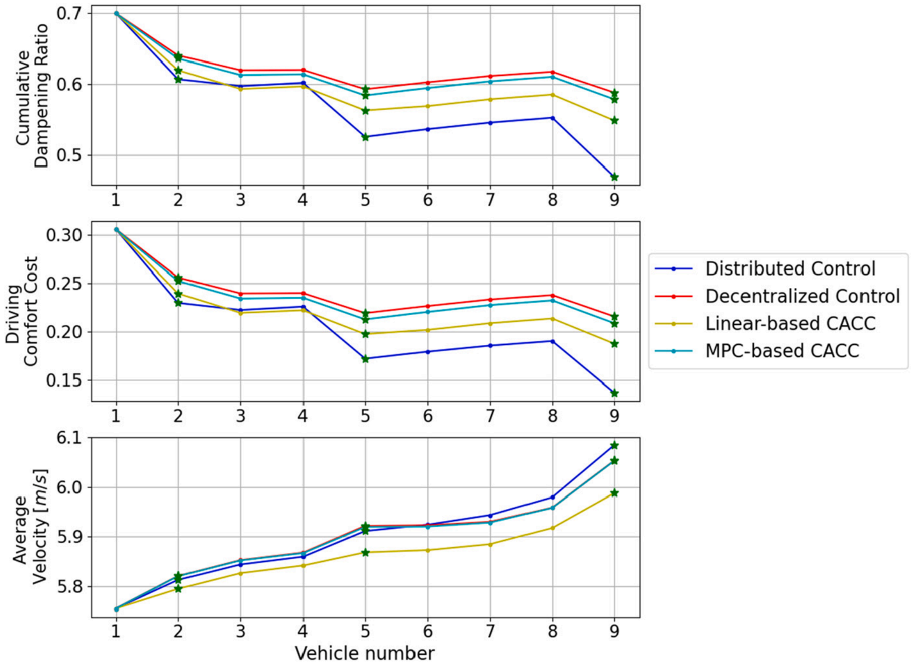

Fig. 10. The detailed performance indicators of the distributed control, decentralized control, linear-based CACC, and MPC-based CACC.

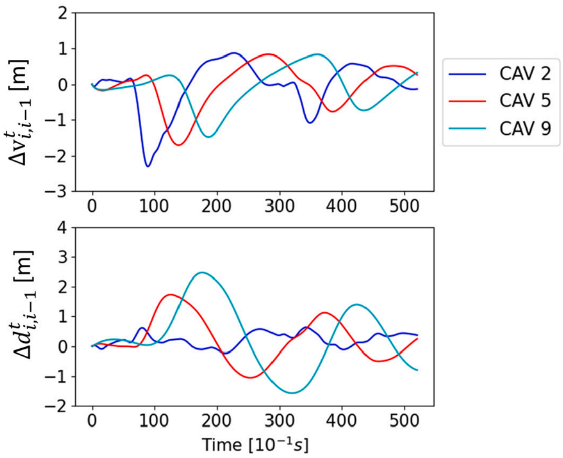

Fig. 11. The spacing equilibrium deviation $\Delta d^t_{i,i-1}$ and speed equilibrium deviation $\Delta v^t_{i,i-1}$ trajectories (equilibrium deviation with the preceding vehicle).

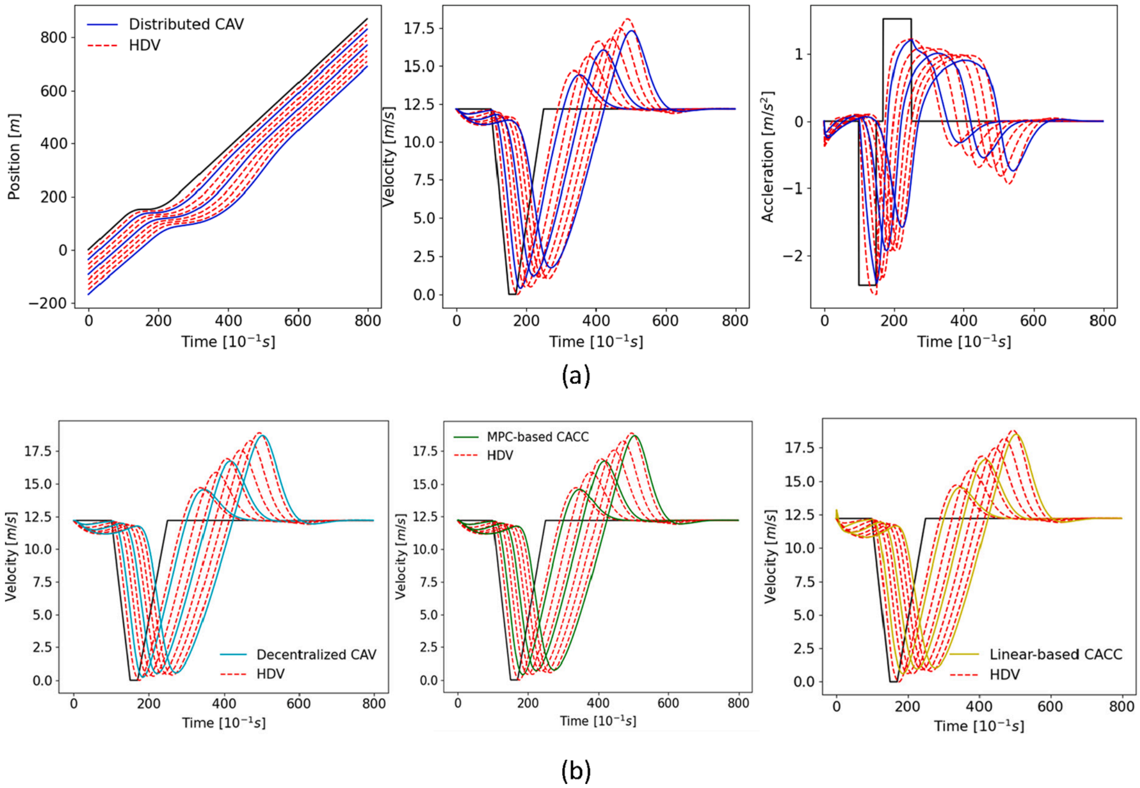

Fig. 12. The position, velocity and acceleration trajectory results under the extreme scenario: (a) Distributed DRL-based controller; (b) Velocity trajectories of decentralized CAV, MPC-based CACC, and linear-based CACC.

gated HDVs in between CAV i and CAV $i-m$). If communication failures between CAV i and CAV $i-m$ happen, the IFT status $\eta^t_{i,i-m}$ (in Equation 14) frequently switches between 0 and 1, which makes the DRL state s^t_i fluctuate**.** This will lead to high-jerk accelerations since the DRL policy directly maps the DRL state to the control action, as presented in Fig. 14(a). Considering the issue, we adopted the 'dynamic information fusion mechanism' proposed by (Shi et al., 2022) to reduce the adverse impact caused by communication losses. The 'dynamic information fusion mechanism' adjusted the IFT status $\eta^t_{i,i-m}$ during communication failures to smooth the acceleration signal, alleviating the high-jerk DRL control issue.

The experiments are conducted to evaluate the control performances under communication failure, as presented in Fig. 14. The

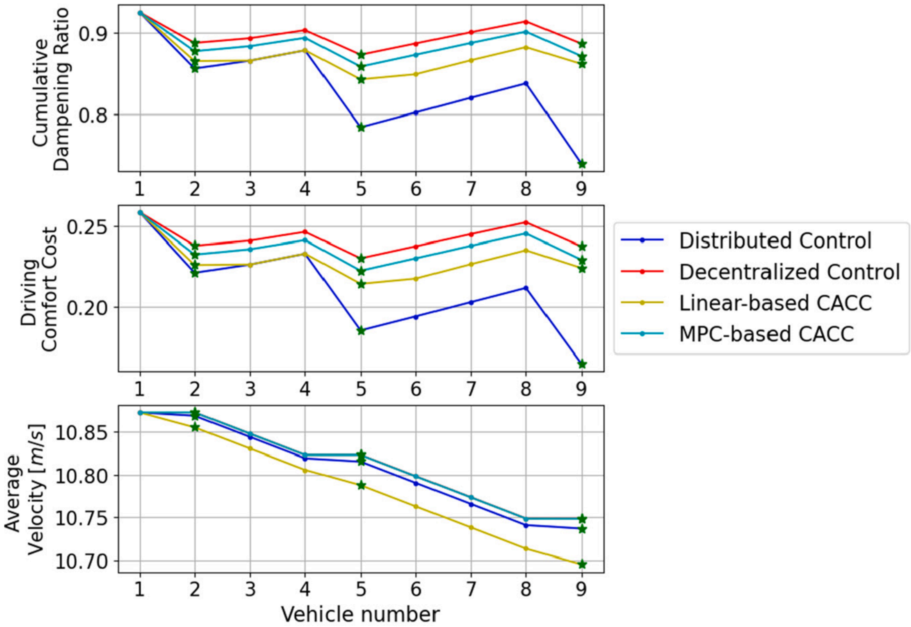

Fig. 13. The detailed performance indicators of the distributed control, decentralized control, linear-based CACC, and MPC-based CACC under the extreme scenarios.

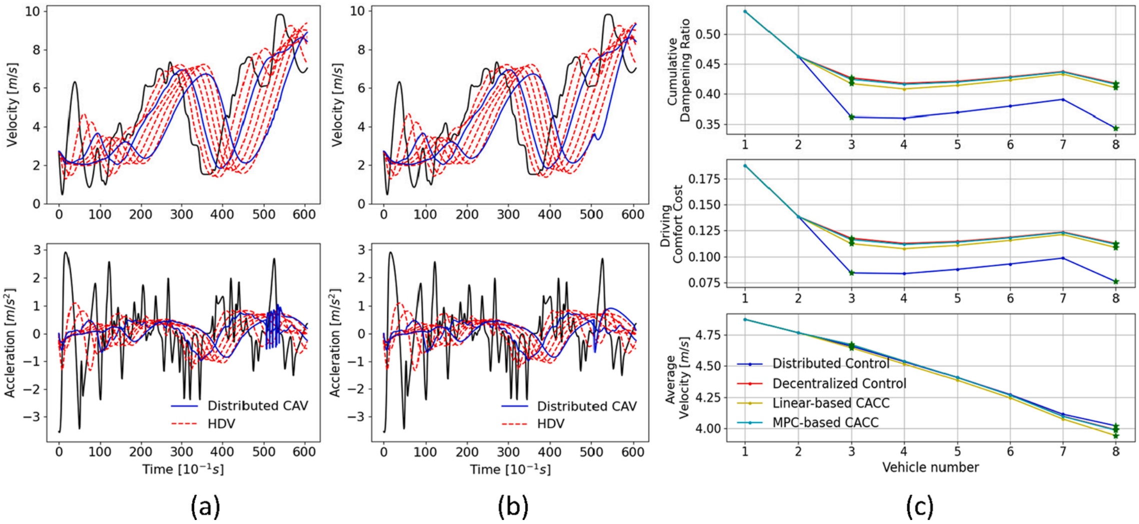

Fig. 14. The performance comparison under communication failure: (a) trajectory under communication failure (b) trajectory under communication failure with adjusted IFT by 'dynamic information fusion mechanism'; (c) performance indicator comparison under communication failure.

communication failure happens during 500 to 550 timesteps, where the IFT status $\eta^{t}_{i,i-m}$ between the receiver CAV 9 and transmitter CAV 5 switches frequently (Fig. 14(a)). With the adopted dynamic information fusion mechanism, the acceleration trajectory suddenly changes when communication failure happens and then performs smoothly without high jerks (Fig/ 12b). The quantified results for indicators are presented in Fig. 14(c). Similar to the previous experiments, the proposed DRL-based control outperforms other approaches in oscillation dampening and driving comfort performances.

4.2.3. Evaluation with different IDM model parameter settings

The adopted IDM model only reproduces one HDV driving pattern. To evaluate the CAV controller considering different HDV driving patterns, we further conducted an experiment using the IDM model with three sets of parameters calibrated based on the NGSIM datasets (Jiang et al., 2023). The trajectory and indicator results are illustrated in Fig. 15. The performances using different IDM models show a similar tendency of control performances, illustrating that the distributed CAV can markedly dampen traffic

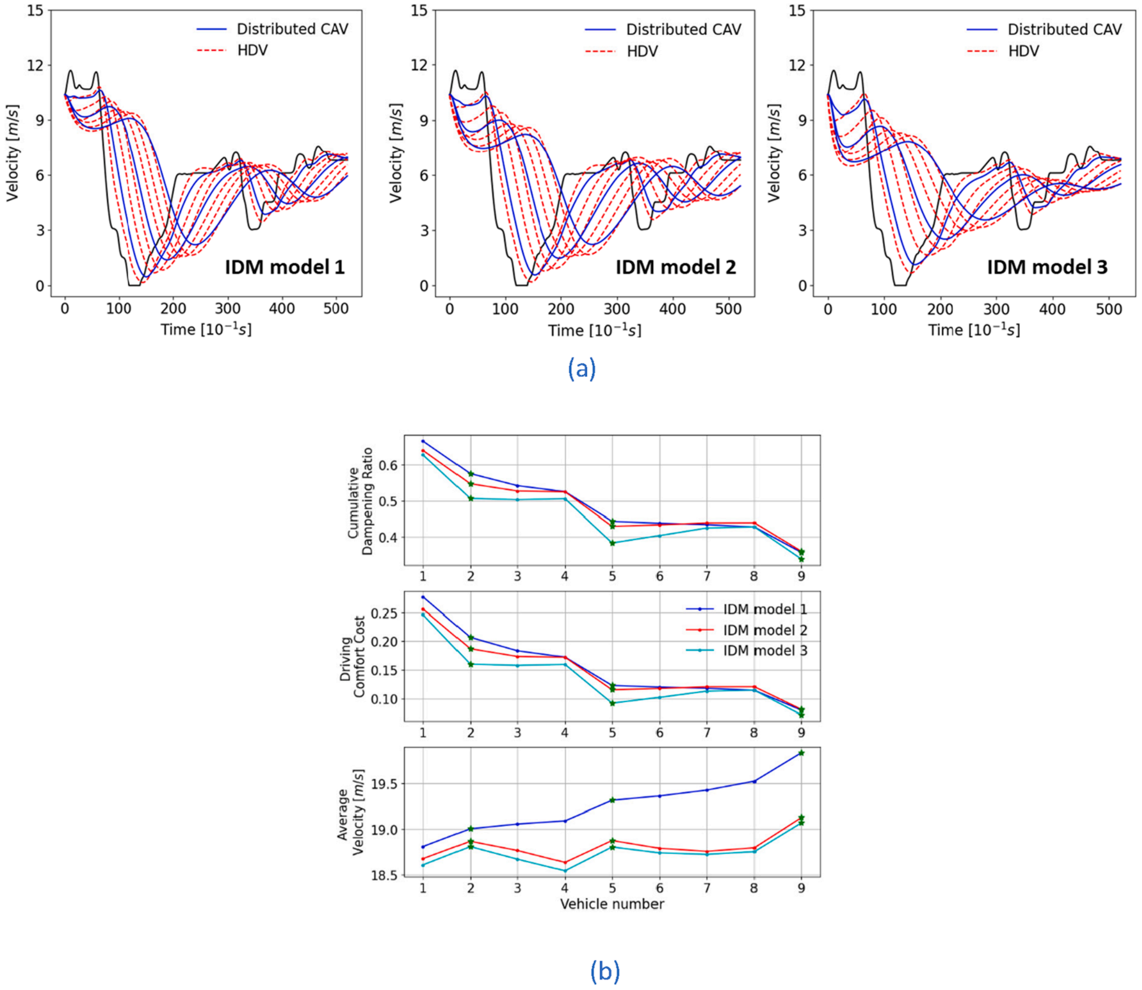

Fig. 15. Performance evaluation results using different calibrated IDM models: (a) trajectory results; (b) indicator results.

disturbances in the mixed traffic flow regarding different HDV driving patterns.

4.2.4. Evaluation using ground-truth AV trajectory

The previous experiments use ground-truth NGSIM data (i.e., HDV trajectory) as the leading vehicle trajectory for evaluation.

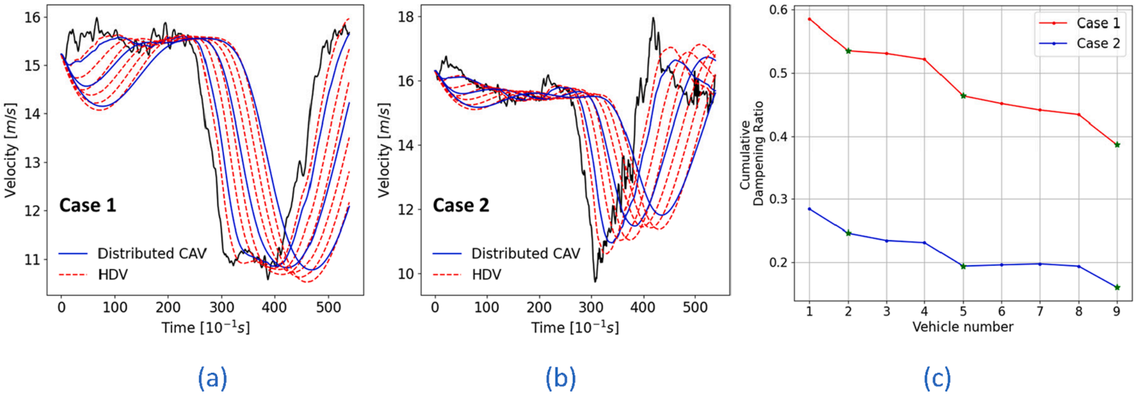

Fig. 16. Two cases of AV ground-truth trajectories as the platoon leading trajectory for control performance evaluation: (a) case 1; (b) case 2; (c) indicator performance.

Furthermore, we evaluate our model using two ground-truth AV trajectories adopted from (Li et al., 2022) as the platoon leader trajectory, with results presented in Fig. 16 below. As suggested by the results, each CAV in the platoon can greatly stabilize the propagated traffic oscillations, suggesting similar control performances as the previous experiments.

4.3. Mixed platoon with different penetration rates

To further visually demonstrate the dampening performance of the proposed control strategy, we utilized the strategy to control CAVs in a 50-follower mixed platoon with different penetration rates (0%, 20%, 40%, 60%, 80%, 100%), where the CAVs are randomly sampled and distributed in the mixed traffic. The platoon leader experienced two typical deceleration-acceleration maneuvers. The followers start with the pre-defined equilibrium states. For results, Fig. 17 illustrates the mixed platoon's velocity and acceleration heat map, and the platoon trajectories under three typical penetration rates (0%, 60%, 100%) are demonstrated in Fig. 18. Generally, the traffic oscillations are dampened gradually with the increasing CAV penetration rate. When the followers are all HDVs (i.e., %0 CAV), the disturbances are propagated and amplified towards the end, seriously impairing the entire traffic flow. Furthermore, the oscillations are undermined to a large extent when the CAV penetration rate reaches 60%, where the upstream vehicles are much less affected by the dampened disturbances. Finally, the disturbances are quickly dissipated in the 100% CAV penetration rate, and the downstream CAVs can promptly recover from the disturbances, showing the controller's strong robustness and resilience. Therefore, the distributed CAV control strategy can effectively stabilize traffic oscillations and significantly improve the entire traffic flow.

4.4. Generalization capability validation

After evaluating the model performance, the generalization capability is validated in this section.

4.4.1. Statistical validation in mixed traffic

First, we use 150 NGSIM ground-truth trajectories excluded from training set, which is with a time length of over 50 s, to validate the statistical robustness of the proposed model's control performances. The experiment is configured with a 15-follower mixed platoon with different penetration rates (0%, 20%, 40%, 60%, 80%, 100%), where CAVs are randomly distributed in the mixed traffic. Each ground-truth trajectory of the 150 NGSIM datasets is assigned as the platoon leader's trajectory for each penetration rate, which means there are 150 simulated vehicular platoons for each penetration rate. The followers start with the initial equilibrium state and are then simulated by the corresponding control models (IDM for HDVs and DRL model for CAVs).

For each simulated platoon in the experiment, we first average the performance indicators of the 15 followers to represent the performance of the whole platoon. Then for each penetration rate, the mean indicator performance value over the 150 platoons is calculated, representing the generalized performance of the penetration rate. The results are demonstrated in Fig. 19, which illustrates that the traffic flow performance in terms of travel efficiency, string stability, and driving comfort is improved monotonically with the increasing CAV penetration rate. Specifically, compared with the HDV platoon, the platoon with a 100% CAV penetration rate reduces a 38.54% dampening ratio, 55.74% driving comfort cost, and increases 5.16% travel efficiency, respectively. These generalized results further validate the generalization capability of the proposed control strategy. To focus on the detailed performance of each vehicle in the platoon, we directly average the indicator performance for each vehicle over the 150 platoons under different CAV penetration rates, as illustrated in Fig. 20. As the CAV penetration rate increases, the traffic disturbances are dampened to a greater extent through the platoon, which optimizes the entire mixed traffic flow.

4.4.2. Cases for irregular initial condition

The initial condition (e.g., initial velocity, acceleration, spacing) of a vehicular platoon has a great impact on the CAV controller (Li et al., 2016; Gao et al., 2019). Irregular initial conditions normally impair the control performances. To further validate the generalization capability, different NGSIM datasets are assigned for both the leader trajectory and the follower's initial states. The vehicular platoon has a topology {0, 1, 0, 0, 0, 1}, where '0′ represents the HDV, and '1′ represents the CAV. The results are presented in Fig. 21. Like the previous experiments, the DRL-based distributed CAV has responsive driving behaviors with great oscillation dampening performances even under the various initial conditions. With the equilibrium-based control philosophy, the CAV can quickly recover to the equilibrium state from the large initial spacing (Fig. 21 (c), (e)) or small initial spacing (Fig. 21 (b), (d)) and maintain close to the equilibrium, which stabilizes the traffic flow and alleviates the adverse impact brought by HDVs' stochasticity. The results validate the great robustness and resilience of the proposed controller.

4.4.3. Generalization capability comparison with other approaches

Finally, multiple ground-truth trajectories from NGSIM datasets are used to further statistically validate the advantage of the proposed DRL-based controller over other compared control methods (given in Section 4.2). The experiments comprise two parts, including (i) equilibrium initial condition and (ii) random initial condition. For each vehicular platoon in the experiments, the first part follows the same experiment configuration in Section 4.2, and the second part follows the same experiment configuration of 'Cases for Random Initial Condition' in Section 4.4. Specifically for each vehicular platoon, a superiority percentage P is defined to quantify the advantage of the proposed DRL-based method over other control approaches:

$$P = \frac{PI_o - PI_d}{PI_o} * 100\% \tag{26}$$

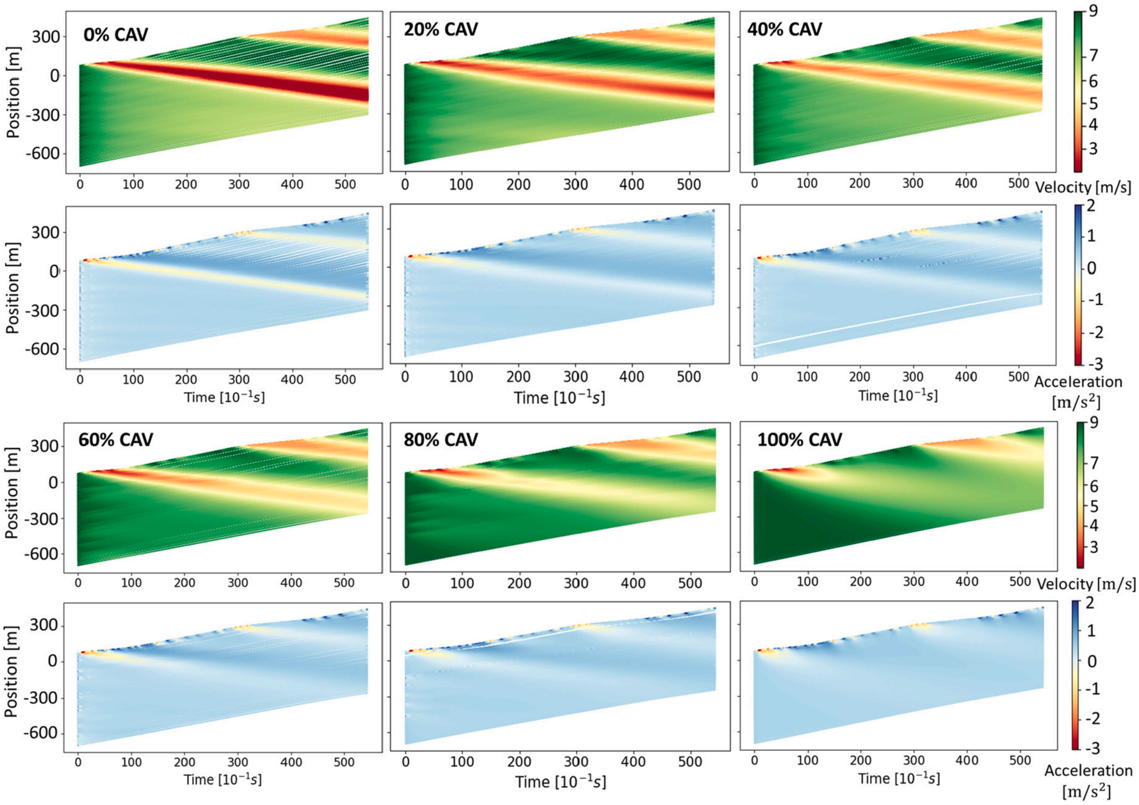

Fig. 17. Velocity and acceleration heat map of mixed platoon with different penetration rates.

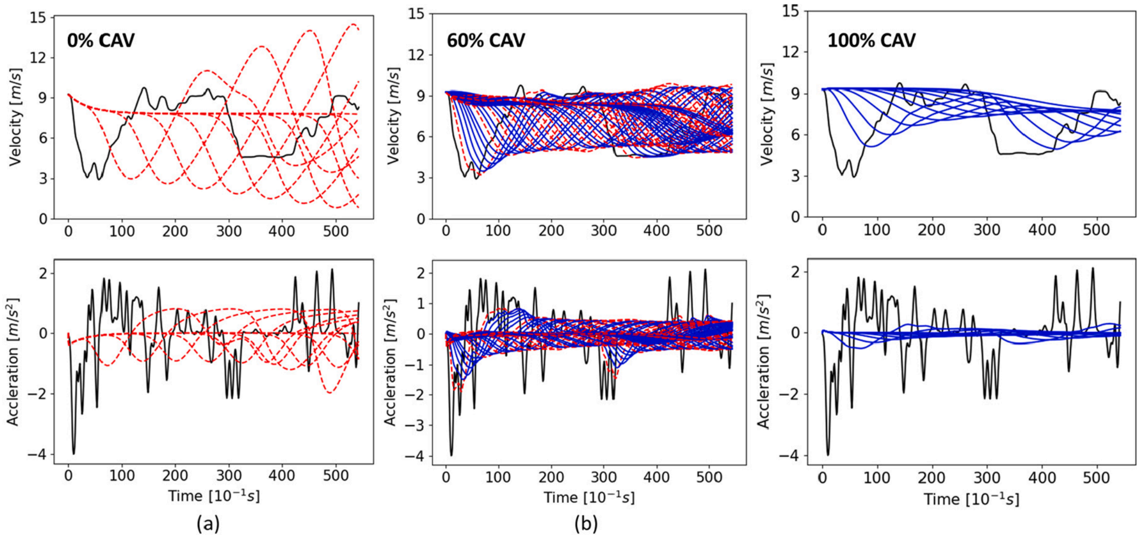

Fig. 18. Trajectories of velocity and acceleration under 0% (Fig. 18 (a)), 60% (Fig. 18 (b)), and 100% (Fig. 18 (c)) penetration rates. (For illustration, only *vehicle*5, *vehicle*10, *vehicle*15⋯*vehicle*50 are plotted for 0% and 100% penetration rates).

where PI_o and PI_d represent the CAV's performance indicator value of the compared control approaches (PI_o) and the proposed DRL-based control approach (PI_d), respectively. Then, the average superiority percentage $\widetilde{P}$ over multiple vehicular platoons is calculated as the final result.

For the first experiment, whose vehicular topology is $T_p = \{1,0,1,0,0,1,0,0,0,1\}$, we focus on the performance of CAV 2, CAV 5, and CAV 9 in the platoon. The average superiority percentage $\widetilde{P}$ is calculated over the 150 vehicular platoons with different NGSIM

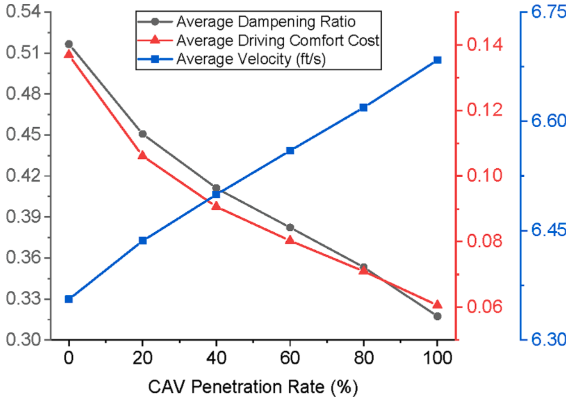

Fig. 19. The generalized statistical results of the mixed platoon with different penetration rates.

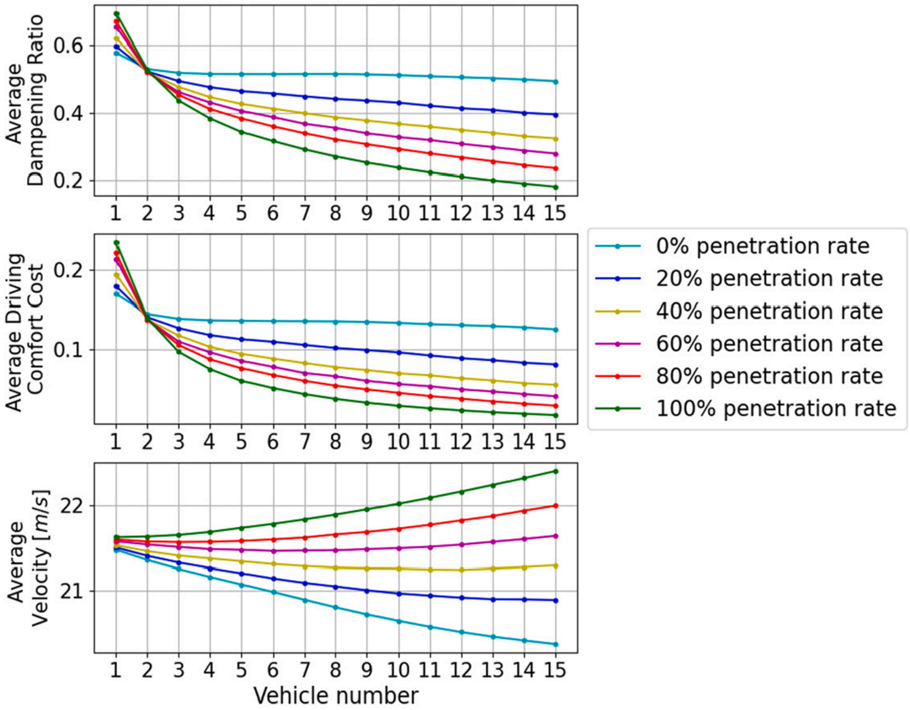

Fig. 20. The generalized indicator results of each vehicle in the mixed platoon under different penetration rates.

leading trajectories (i.e., same NGSIM data in Section 4.4). The results are presented in Fig. 22. As can be found in the Figure, the proposed control method markedly outperforms other control approaches regarding oscillation dampening and driving comfort. Moreover, the more HDVs between the controlled CAV and the immediate CAV downstream, the proposed DRL-based control shows higher advantages than other approaches. With more HDVs and a longer distance between the two CAVs, the proposed control can better capture the stochastic characteristics of the aggregated HDVs joint driving behaviors and stabilize traffic disturbances to a greater extent.

For the second experiment, whose vehicular topology is $T_p = \{0, 1, 0, 0, 0, 1\}$, we focus on the performance of CAV 5, which is the last CAV in the platoon. The average superiority percentage $\widetilde{P}$ is calculated over thirty vehicular platoons with different NGSIM leading trajectories over 500 timesteps and irregular initial conditions for followers. The average superiority percentage $\widetilde{P}$ is shown in Fig. 23. The MPC-based controller does not perform well in this case since the optimized control policy is more sensitive to the initial state. A large initial spacing may lead to a relatively aggressive control policy, which increases acceleration energy. Similarly, the proposed control method performs better in every aspect than other compared approaches for the cases with irregular initialized conditions.

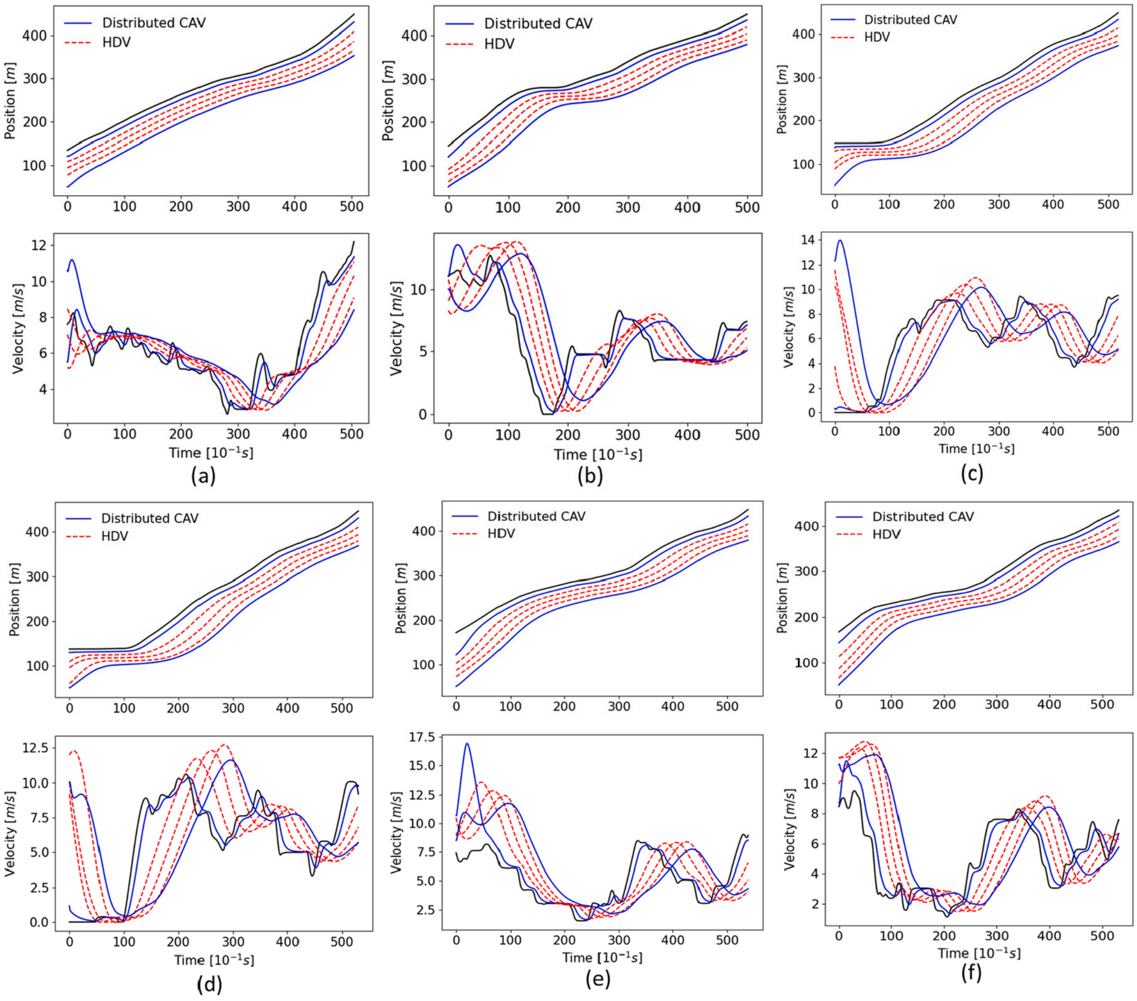

Fig. 21. The generalized mixed platoon trajectories with initial states from ground-truth NGSIM data.

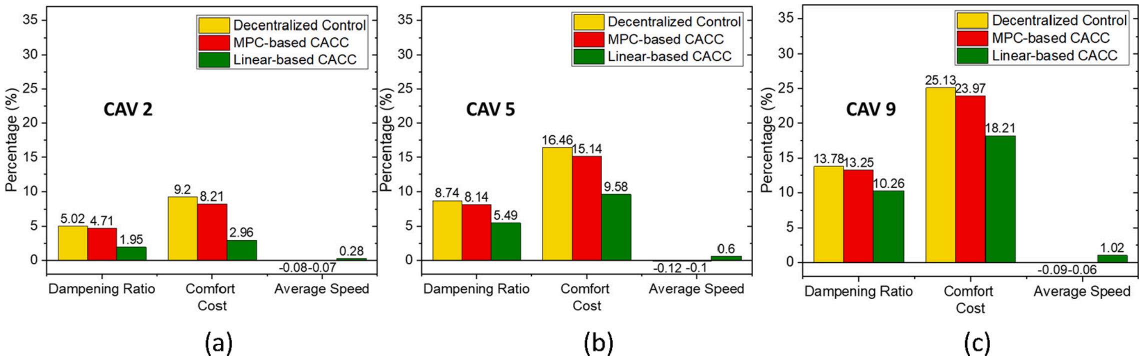

Fig. 22. Generalization capability comparisons for different CAVs in the mixed platoon (CAV 2 with one downstream HDVs (a); CAV 5 with two downstream HDVs (b); CAV 9 with three downstream CAVs (c); 0 line represents the performance of the proposed approach).

5. Conclusion

This research proposes a DRL-based distributed CAV longitudinal control strategy for mixed traffic of CAVs and HDVs. In this generic distributed control framework, each CAV receives the fused real-time information of vehicles in the local downstream environment for longitudinal control. To generalize the diversified downstream topologies, any mixed local downstream environment is

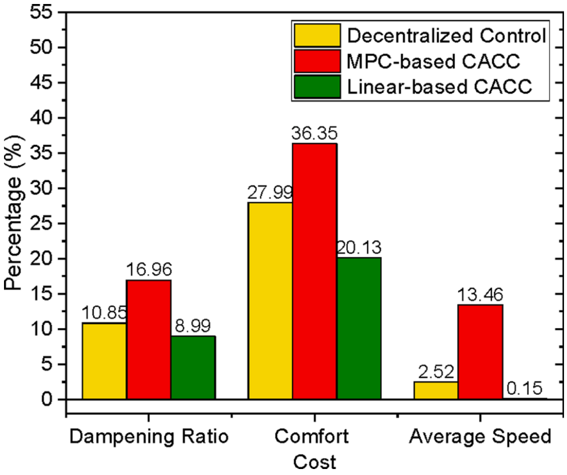

Fig. 23. Generalization capability comparisons with irregular initial states (i.e., CAV 5 in the mixed platoon; 0 line represents the performance of the proposed approach).

categorized as the CAV-HDVs-CAV pattern, which consists of a nearest downstream CAV followed by aggregated HDVs. For this local heterogeneous environment, we construct a novel vehicle-following structure ‘CAV-AHDV-CAV’ based on Newell’s car-following model to capture the macroscopic traffic properties of the aggregated HDVs and embed them into the control framework. This approach efficiently attenuates the HDVs’ stochasticity and enhances the car-following performances. With the philosophy, a novel DRL state fusion strategy based on the equilibrium concept is proposed to regulate each CAV close to the pre-defined equilibrium state and greatly stabilize traffic oscillations. For model development, NGSIM datasets are embedded in training to better incorporate the preceding vehicles’ stochastic characteristics into control. The DPPO algorithm is adopted to enhance the convergence of control policy updated in the training process.

A series of simulated experiments are conducted with NGSIM datasets. The proposed strategy’s control performance is evaluated regarding empirical string stability, travel efficiency, and driving comfort. Numerical results indicate that the proposed distributed control strategy can significantly dampen the traffic oscillation and outperform the decentralized strategy and linear-based CACC strategy in every aspect. Then, the dampening performance of the proposed control strategy is intuitively demonstrated in a 50-follower mixed platoon with different penetration rates, showing its strong robustness and resilience. Finally, the generalization capability of the proposed strategy is validated.

This study still has several limitations. The first point lies in that the proposed control method focuses on heavily congested traffic conditions, while it is not suitable for free flow conditions. Besides, this paper does not consider the communication delay, which may lead to an over-optimistic performance. Moreover, the paper only considers the longitudinal car-following movement, which is relatively limited for applications in more complex scenarios (e.g., lane-changing movement). Some future work can be conducted based on the research. For instance, the vehicle dynamics can be built more complex considering the internal vehicle components (e.g., pedal, steering wheel, brake). Moreover, lateral movement can be incorporated into the control framework to reproduce more complex traffic scenarios, such as lane-changing, merging, or diverging behaviors.

CRediT authorship contribution statement

Haotian Shi: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Danjue Chen:** Methodology, Writing – review & editing. **Nan Zheng:** Writing – review & editing. **Xin Wang:** Methodology, Writing – review & editing. **Yang Zhou:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Bin Ran:** Conceptualization, Writing – review & editing.

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 paper.

Acknowledgement

This work was supported by the Wisconsin Traffic Operations and Safety (TOPS) Laboratory.

References

Absil, P.A., Kurdyka, K., 2006. On the stable equilibrium points of gradient systems. Syst. Control Lett. 55 (7), 573–577. https://doi.org/10.1016/j.sysconle.2006.01.002.

Bang, S., Ahn, S., 2019. Mixed traffic of connected and autonomous vehicles and human-driven vehicles: traffic evolution and control using spring-mass-damper system. Transp. Res. Rec. 2673 (7), 504–515. https://doi.org/10.1177/0361198119847618.

Chen, D., Laval, J., Zheng, Z., Ahn, S., 2012. A behavioral car-following model that captures traffic oscillations. Transp. Res. B Methodol. 46 (6), 744–761. https://doi.org/10.1016/j.trb.2012.01.009.

Chen, D., Srivastava, A., Ahn, S., & Li, T. (2020). Traffic dynamics under speed disturbance in mixed traffic with automated and non-automated vehicles. Transport. Res. Part C: Emerg. Technol., 113(April 2019), 293–313. https://doi.org/10.1016/j.trc.2019.03.017.

Du, L., Dao, H., 2015. Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream. IEEE Trans. Intell. Transp. Syst. 16 (1), 66–80. https://doi.org/10.1109/TITS.2014.2326331.

Du, R., Chen, S., Li, Y., Dong, J., Ha, P. Y. J., & Labi, S. (2020). A Cooperative Control Framework for CAV Lane Change in a Mixed Traffic Environment. 0–1. http://arxiv.org/abs/2010.05439.

Duan, J., Li, S.E., Guan, Y., Sun, Q., Cheng, B., 2020. Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data. IET Intel. Transport Syst. 14 (5), 297–305. https://doi.org/10.1049/iet-its.2019.0317.

Duret, A., Ahn, S., Buisson, C., 2011. Passing rates to measure relaxation and impact of lane-changing in congestion. Computer-Aid. Civ. Inf. Eng. 26 (4), 285–297. https://doi.org/10.1111/j.1467-8667.2010.00675.x.

Feng, S., Zhang, Y., Li, S.E., Cao, Z., Liu, H.X., Li, L., 2019. String stability for vehicular platoon control: Definitions and analysis methods. Annu. Rev. Control. 47, 81–97. https://doi.org/10.1016/j.arcontrol.2019.03.001.

Fisher, J., Bhattacharya, R., 2009. Linear quadratic regulation of systems with stochastic parameter uncertainties. Automatica 45 (12), 2831–2841. https://doi.org/10.1016/j.automatica.2009.10.001.

Gao, S., Dong, H., Song, H., Zhou, M., 2019. On state feedback control and Lyapunov analysis of car-following model. Physica A 534, 122320. https://doi.org/10.1016/j.physa.2019.122320.

Ge, J.I., Orosz, G., 2014. Dynamics of connected vehicle systems with delayed acceleration feedback. Transport. Res. Part C: Emerg. Technol. 46, 46–64. https://doi.org/10.1016/j.trc.2014.04.014.

Gong, S., Du, L., 2018. Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles. Transp. Res. B Methodol. 116, 25–61. https://doi.org/10.1016/j.trb.2018.07.005.

Gong, S., Shen, J., Du, L., 2016. Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. Transp. Res. B Methodol. 94, 314–334. https://doi.org/10.1016/j.trb.2016.09.016.

Görges, D., 2017. Relations between Model Predictive Control and Reinforcement Learning. IFAC-PapersOnLine 50 (1), 4920–4928. https://doi.org/10.1016/j.ifacol.2017.08.747.

Grondman, I., Busoniu, L., Lopes, G.A.D., Babuška, R., 2012. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 42 (6), 1291–1307. https://doi.org/10.1109/TSMCC.2012.2218595.

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, M., & Silver, D. (2017). Emergence of Locomotion Behaviours in Rich Environments. http://arxiv.org/abs/1707.02286.

Jiang, J., Zhou, Y., Wang, X., & Ahn., S. A Generic Stochastic Hybrid Car-following Model Based on Approximate Bayesian Computation. Presented at 102ed Transportation Research Board (TRB) Annual Meeting. TRBAM-23-02657. Washington, D.C., 2023.

Jiménez, F., Naranjo, J.E., García, F., 2013. An Improved Method to Calculate the Time-to-Collision of Two Vehicles. Int. J. Intell. Transp. Syst. Res. 11 (1), 34–42. https://doi.org/10.1007/s13177-012-0054-4.

Kesting, A., Treiber, M., 2008. Calibrating car-following models by using trajectory data methodological study. Transp. Res. Rec. 2088, 148–156. https://doi.org/10.3141/2088-16.

Knorn, S., Donaire, A., Agüero, J.C., Middleton, R.H., 2014. Passivity-based control for multi-vehicle systems subject to string constraints. Automatica 50 (12), 3224–3230. https://doi.org/10.1016/j.automatica.2014.10.038.

Kwak, S.G., Kim, J. H. (2017). cornerstone of modern statistics.

Laval, J.A., Leclercq, L., 2010. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 368 (1928), 4519–4541. https://doi.org/10.1098/rsta.2010.0138.

Li, S.E., Qin, X., Zheng, Y., Wang, J., Li, K., Zhang, H., 2019. Distributed Platoon Control under Topologies with Complex Eigenvalues: Stability Analysis and Controller Synthesis. IEEE Trans. Control Syst. Technol. 27 (1), 206–220. https://doi.org/10.1109/TCST.2017.2768041.

Li, T., Chen, D., Zhou, H., Xie, Y., & Laval, J. (2022). Fundamental diagrams of commercial adaptive cruise control: Worldwide experimental evidence. Transport. Res. Part C: Emerg. Technol., 134(October 2021), 103458. https://doi.org/10.1016/j.trc.2021.103458.

Li, Y., Li, K., Zheng, T., Hu, X., Feng, H., Li, Y., 2016. Evaluating the performance of vehicular platoon control under different network topologies of initial states. Physica A 450, 359–368. https://doi.org/10.1016/j.physa.2016.01.006.

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 2016. Continuous control with deep reinforcement learning. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings.

Lin, F., Fardad, M., Jovanović, M.R., 2012. Optimal control of vehicular formations with nearest neighbor interactions. IEEE Trans. Autom. Control 57 (9), 2203–2218. https://doi.org/10.1109/TAC.2011.2181790.

Lin, Y., Wang, P., Zhou, Y., Ding, F., Wang, C., Tan, H., 2020. Platoon Trajectories Generation: A Unidirectional Interconnected LSTM-Based Car-Following Model. IEEE Trans. Intell. Transp. Syst. 1–11 https://doi.org/10.1109/TITS.2020.3031282.

Lu, C., Liu, C., 2021. Ecological control strategy for cooperative autonomous vehicle in mixed traffic considering linear stability. J. Intell. Connected Vehicles 4 (3), 115–124. https://doi.org/10.1108/jicv-08-2021-0012.

Meng, D., Song, G., Wu, Y., Zhai, Z., Yu, L., Zhang, J., 2021. Modification of Newell's car-following model incorporating multidimensional stochastic parameters for emission estimation. Transp. Res. Part D: Transp. Environ. 91 (January), 1–20. https://doi.org/10.1016/j.trd.2020.102692.

Minderhoud, M.M., Bovy, P.H.L., 2001. Extended time-to-collision measures for road traffic safety assessment. Accid. Anal. Prev. 33 (1), 89–97. https://doi.org/10.1016/S0001-4575(00)00019-1.

Mnih, V., Badia, A.P., Mirza, L., Graves, A., Harley, T., Lillicrap, T.P., Silver, D., Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. 33rd International Conference on Machine Learning, ICML 2016, 4, 2850–2869.

Montanino, M., Punzo, V., 2015. Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. Transp. Res. B Methodol. 80, 82–106. https://doi.org/10.1016/j.trb.2015.06.010.

Morbidi, F., Colaneri, P., Stanger, T., 2013. Decentralized optimal control of a car platoon with guaranteed string stability. 2013 European Control Conference ECC 2013, 3494–3499. https://doi.org/10.23919/ecc.2013.6669336.

Naus, G.J.L., Vugts, R.P.A., Ploeg, J., Van De Molengraft, M.J.G., Steinbuch, M., 2010. String-stable CACC design and experimental validation: A frequency-domain approach. IEEE Trans. Veh. Technol. 59 (9), 4268–4279. https://doi.org/10.1109/TVT.2010.2076320.

Newell, G.F., 2002. A simplified car-following theory: a lower order model. Transp. Res. B Methodol. 36 (3), 195–205.

Orosz, G., 2016. Connected cruise control: modelling, delay effects, and nonlinear behaviour. Veh. Syst. Dyn. 54 (8), 1147–1176. https://doi.org/10.1080/00423114.2016.1193209.

Ploeg, J., Van De Wouw, N., Nijmeijer, H., 2014. Lp string stability of cascaded systems: Application to vehicle platooning. IEEE Trans. Control Syst. Technol. 22 (2), 786–793. https://doi.org/10.1109/TCST.2013.2258346.

Punzo, V., Borzacchiello, M.T., Ciuffo, B., 2011. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. Transport. Res. Part C: Emerg. Technol. 19 (6), 1243–1262. https://doi.org/10.1016/j.trc.2010.12.007.

Qu, X., Yu, Y., Zhou, M., Lin, C.T., Wang, X., 2020. Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. Appl. Energy 257 (September 2019), 114030. https://doi.org/10.1016/j.apenergy.2019.114030.

Richards, P.I. (2013). Shock Waves on the Highway Author (s): Paul I . Richards Published by : INFORMS Stable URL : http://www.jstor.org/stable/167515 . SHOCK WAVES ON THE HIGHWAY *. 4(1), 42–51.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal Policy Optimization Algorithms. 1–12. http://arxiv.org/abs/1707.06347.

Shi, H., Nie, Q., Fu, S., Wang, X., Zhou, Y., Ran, B., 2021a. A distributed deep reinforcement learning–based integrated dynamic bus control system in a connected environment. Comput. Aided Civ. Inf. Eng. 1–17 https://doi.org/10.1111/mice.12803.

Shi, H., Zhou, Y., Wang, X., Fu, S., Gong, S., Ran, B., 2022. A deep reinforcement learning-based distributed connected automated vehicle control under communication failure. Comput. Aided Civ. Inf. Eng. 1–19 https://doi.org/10.1111/mice.12825.

Shi, H., Zhou, Y., Wu, K., Wang, X., Lin, Y., Ran, B., 2021b. Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment. Transport. Res. Part C: Emerg. Technol. 133 (August), 103421 https://doi.org/10.1016/j.trc.2021.103421.

Takahama, T., Akasaka, D., 2018. Model Predictive Control Approach to Design Practical Adaptive Cruise Control for traffic jam. Int. J. Automotive Eng. 9 (3), 99–104. https://doi.org/10.20485/jsaeijae.9.3_99.

Tian, J., Zhu, C., Chen, D., Jiang, R., Wang, G., Gao, Z., 2021. Car following behavioral stochasticity analysis and modeling: Perspective from wave travel time. Transp. Res. B Methodol. 143, 160–176. https://doi.org/10.1016/j.trb.2020.11.008.

Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat Interdiscip. Topics 62 (2), 1805–1824. https://doi.org/10.1103/PhysRevE.62.1805.

Wang, C., Gong, S., Zhou, A., Li, T., & Peeta, S. (2019). Cooperative adaptive cruise control for connected autonomous vehicles by factoring communication-related constraints ☆. Transportation Research Part C, April, 1–22. https://doi.org/10.1016/j.trc.2019.04.010.

Wang, J., Zheng, Y., Xu, Q., Wang, J., Li, K., 2020a. Controllability Analysis and Optimal Control of Mixed Traffic Flow With Human-Driven and Autonomous Vehicles. IEEE Trans. Intell. Transp. Syst. 1–15 https://doi.org/10.1109/tits.2020.3002965.

Wang, M., 2018. Infrastructure assisted adaptive driving to stabilise heterogeneous vehicle strings. Transport. Res. Part C: Emerg. Technol. 91, 276–295. https://doi.org/10.1016/j.trc.2018.04.010.

Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014a. Rolling horizon control framework for driver assistance systems. Part I: Mathematical formulation and non-cooperative systems. Transport. Res. Part C: Emerg. Technol. 40, 271–289. https://doi.org/10.1016/j.trc.2013.11.023.

Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014b. Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. Transport. Res. Part C: Emerg. Technol. 40, 290–311. https://doi.org/10.1016/j.trc.2013.11.024.

Wang, M., Daamen, W., Hoogendoorn, S.P., Van Arem, B., 2016. Cooperative Car-Following Control: Distributed Algorithm and Impact on Moving Jam Features. IEEE Trans. Intell. Transp. Syst. 17 (5), 1459–1471. https://doi.org/10.1109/TITS.2015.2505674.

Wang, Y., Hou, S., Wang, X., 2019b. Crossing Traffic Avoidance of Automated Vehicle Through Bird-View Control, a Reinforcement Learning Approach. SSRN Electron. J. https://doi.org/10.2139/ssrn.3495727.

Wang, Y.u., Li, X., Tian, J., Jiang, R., 2020b. Stability analysis of stochastic linear car-following models. Transp. Sci. 54 (1), 274–297. https://doi.org/10.1287/trsc.2019.0932.

Wang, Z., Bian, Y., Shladover, S.E., Wu, G., Li, S.E., Barth, M.J., 2020c. A Survey on Cooperative Longitudinal Motion Control of Multiple Connected and Automated Vehicles. IEEE Intell. Transp. Syst. Mag. 12 (1), 4–24. https://doi.org/10.1109/MITS.2019.2953562.

Whitham, G.B., 1955. On kinematic waves I. Flood movement in long rivers. Proc. R. Soc. Lond. A 229 (1178), 281–316. https://doi.org/10.1098/rspa.1955.0088.

Willems, Polderman, 2013. Introduction to Mathematical Systems Theory: A Behavioral Approach. Springer Science & Business Media.

Zhang, L., Orosz, G., 2016. Motif-Based Design for Connected Vehicle Systems in Presence of Heterogeneous Connectivity Structures and Time Delays. IEEE Trans. Intell. Transp. Syst. 17 (6), 1638–1651. https://doi.org/10.1109/TITS.2015.2509782.

Zhang, L., Orosz, G., 2017. Consensus and disturbance attenuation in multi-agent chains with nonlinear control and time delays. Int. J. Robust Nonlinear Control 27 (5), 781–803. https://doi.org/10.1002/rnc.3600.

Zhang, Z., & Yang, X. (Terry). (2021). Analysis of highway performance under mixed connected and regular vehicle environment. J. Intell. Connected Vehicles, 4(2), 68–79. https://doi.org/10.1108/jicv-10-2020-0011.

Zheng, F., Liu, C., Liu, X., Jabari, S.E., Lu, L., 2020a. Analyzing the impact of automated vehicles on uncertainty and stability of the mixed traffic flow. Transport. Res. Part C: Emerg. Technol. 112 (January), 203–219. https://doi.org/10.1016/j.trc.2020.01.017.

Zheng, Y., Zhang, Y., Ran, B., Xu, Y., Qu, X., 2020b. Cooperative control strategies to stabilise the freeway mixed traffic stability and improve traffic throughput in an intelligent roadside system environment. IET Intel. Transport Syst. 14 (9), 1108–1115. https://doi.org/10.1049/iet-its.2019.0577.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011. Freeway traffic oscillations: Microscopic analysis of formations and propagations using Wavelet Transform. Transp. Res. B Methodol. 45 ‘(9), 1378–1388. https://doi.org/10.1016/j.trb.2011.05.012.

Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. Transport. Res. Part C: Emerg. Technol. 25, 18–33. https://doi.org/10.1016/j.trc.2012.03.006.

Zhou, Y., Ahn, S., Chitturi, M., Noyce, D.A., 2017. Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty. Transport. Res. Part C: Emerg. Technol. 83, 61–76. https://doi.org/10.1016/j.trc.2017.07.011.

Zhou, Y., Ahn, S., Wang, M., & Hoogendoorn, S. (2019). Stabilizing mixed vehicular platoons with connected automated vehicles: An H-infinity approach. Transport. Res. Part B: Methodol., xxxx. https://doi.org/10.1016/j.trb.2019.06.005.

Zhou, Y., Wang, M., Ahn, S., 2019b. Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability. Transp. Res. B Methodol. 128, 69–86. https://doi.org/10.1016/j.trb.2019.07.001.

Zhu, M., Wang, X., Wang, Y., 2018. Human-like autonomous car-following model with deep reinforcement learning. Transport. Res. Part C: Emerg. Technol. 97 (October), 348–368. https://doi.org/10.1016/j.trc.2018.10.024.

Zou, Y., Qu, X., 2018. On the impact of connected automated vehicles in freeway work zones: a cooperative cellular automata model based approach. J. Intelligent Connected Vehicles 1 (1), 1–14. https://doi.org/10.1108/jicv-11-2017-0001.