Reinforcement Learning for Freeway Lane-Change Regulation via Connected Vehicles

Ke Sun, *Student Member, IEEE*, and Huan Yu, *Senior Member, IEEE*

***Abstract*—Lane change decision-making is a complex task due to intricate vehicle-vehicle and vehicle-infrastructure interactions. Existing algorithms for lane-change control often depend on vehicles with a certain level of autonomy (e.g., autonomous or connected autonomous vehicles), which limits their applicability under low penetration rates of automated vehicles. To address this issue, this study proposes a lane-change regulation framework using multi-agent reinforcement learning (MARL) to improve freeway traffic efficiency via connected vehicles. Specifically, lane-change regulation signals (e.g., 'left change allowed" or 'right change disallowed") are computed at a traffic management center and then broadcast to connected vehicles. Thus, human-driven vehicles remain uncontrolled, while connected vehicles follow the received regulation signals. Compared to vehicle-level maneuver control, this approach reduces communication and positioning requirements, avoids direct trajectory intervention, and can accommodate different levels of connectivity and autonomy. To further bridge the gap between theoretical macroscopic analysis and microscopic simulations, the proposed framework combines a microscopic simulation environment and a macroscopic lane-grid representation. The microscopic simulator is used to generate vehicle-level trajectories and execute the realized traffic dynamics, while the macroscopic framework aggregates microscopic vehicle information into lane-grid states, enabling low-cost control at the grid-agent level. Based on a multi-lane macroscopic traffic model represented by partial differential equations (PDEs), lane changes are interpreted as source-term exchanges between adjacent lanes, which can be regulated by the proposed lane-change regulation signals. Our model is evaluated in varied traffic scenarios and demand conditions using microscopic simulation. Experimental results demonstrate that the method improves overall traffic efficiency with limited additional energy consumption while maintaining driving safety.**

***Index Terms*—Multi-agent reinforcement learning, traffic flow, lane change regulation, macroscopic traffic model.**

I. Introduction

FREEWAY traffic control has relied on infrastructure-based measures such as speed limits and ramp traffic lights. Recently, there has been a growing focus on leveraging connected automated vehicles (CAVs) for traffic flow improvement. The cruising behaviors of the CAVs are controlled to enhance traffic stability [1], [2], [3], [4], safety [5], [6], [7], energy consumption [8], [9], [10], and efficiency [3], [11], [12], [13]. While significant advancements have been achieved in longitudinal vehicle and platoon control, lateral control—particularly in lane change control—remains a challenging task due to the complexity of optimization in a higher-dimensional solution space.

Many studies focus on ramp-merging areas in freeway traffic control, where mandatory lane changes of ramp-merging vehicles can cause substantial disturbances to mainline traffic. Traditionally, macroscopic approaches such as ramp metering have been employed to control the inflow from on-ramps. Classical methods include the feedback control strategy ALINEA [14] and its extensions [15], [16], [17]. More recently, Deep Reinforcement Learning (DRL)-based approaches have been explored to achieve superior performance under model uncertainty and incomplete information [18], [19], [20], [21]. With advances in connectivity and autonomy technologies, CAVs offer an opportunity to control individual vehicles in ramp-merging areas instead of relying solely on aggregated control. A classical method is to map the merging problem to a platoon control framework using virtual mapping [22], [23], [24]. Beyond this mapping framework, optimization-based methods have been proposed to search for optimal solutions [25], [26], [27]. DRL-based methods have also been introduced [28], [29].

Another critical challenge involves managing lane changes on freeways to improve traffic efficiency, which is the primary focus of our work. Studies such as [30] and [31] demonstrate that lane changes significantly impact traffic safety and efficiency. Further research shows that CAVs can enhance these metrics in traffic congestion, particularly with high CAV penetration rates [32], [33]. At the macroscopic level, an implicit strategy for managing lane changes is the implementation of restricted-use lanes, such as High Occupancy Vehicle (HOV) lanes [34], [35]. Related multi-class macroscopic traffic models further show the importance of class-aware and lane-level representations. Classical multi-class extensions of LWR-type and kinematic-wave models describe heterogeneous traffic by assigning different states or behavioral parameters to different vehicle classes [36], [37]. More recent multiclass CTM-based models describe mixed traffic with human-driven vehicles and autonomous or connected automated vehicles, where traffic capacity, wave speed, and lane-level flow dynamics depend on the class composition and penetration rate [38], [39]. To optimize individual lane change decisions and trajectories, researchers have developed collaborative driving strategies. For example, game-theory-based methods, such as the Nash bargaining approach in [40], aim to optimize traffic flow. Optimization-based methods [41], [42] seek to mitigate

Received 5 December 2024; revised 1 July 2025 and 7 June 2026; accepted 5 August 2026. Date of publication 14 August 2026; date of current version 6 October 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 12526214 and Grant 62203131 and in part by Guangzhou Municipal Education Bureau University under Project 2024312102. The Associate Editor for this article was A. Sciarretta. *(Corresponding author: Huan Yu.)*

The authors are with the Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong 511453, China (e-mail: huanyu@hkust-hz.edu.cn).

Digital Object Identifier 10.1109/TITS.2026.3722245

1558-0016 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

congestion near conflict areas by coordinating CAV lane change decisions. Additionally, Multi-Agent Reinforcement Learning (MARL) methods have been proposed to achieve robust generalization performance [43], [44].

However, existing DRL models for lane change control face two significant challenges. First, given the expected low penetration of CAVs in the near future, performance under varying penetration rates must be carefully evaluated. Experimental findings in [3] reveal that while CAVs can enhance traffic stability, substantial improvements are primarily observed at higher penetration rates. Second, the accurate collection of traffic state information, particularly in large-scale environments, is often prohibitively expensive or even infeasible. Many existing approaches assume robust vehicle-to-vehicle or vehicle-to-infrastructure communication. However, as noted in [45], challenges such as communication latency, limited data bandwidth, and privacy concerns can impede the effectiveness of centralized training schemes.

To address these deficiencies, we propose a traffic flow control strategy within a multi-agent reinforcement learning (MARL) framework. This strategy regulates lane-changing behaviors of connected vehicles (CVs) at the macroscopic level to enhance the overall efficiency of mixed freeway traffic. Inspired by infrastructure-based control methods such as ramp metering and restricted-use lanes for freeway management, our model employs an infrastructure-based regulation framework that operates at the decision level. Regulation signals—either permitting or prohibiting maneuvers—are dispatched through a traffic management center to CVs. Consequently, our algorithm follows the paradigm of centralized training with decentralized execution (CTDE). A shared policy is trained in the traffic management center, and each agent selects its own actions independently.

The minimal regulation units in our model are spatially discretized cells of the macroscopic PDE model, referred to as lane grids. Lane-grid-based methods have become increasingly common in recent lane-level variable speed limit (VSL) control. Lane changes often reduce the effectiveness of VSL strategies, especially near traffic bottlenecks. To address this issue, lane grids are preferred over aggregated road segments because they capture inter-lane speed differences more accurately [46]. In [47] and [48], experimental results show that fine-resolution VSL control outperforms traditional road-segment-based approaches. In [49], mandatory lane-change control is combined with lane-grid VSL, demonstrating superior efficiency compared to VSL-only methods.

Our design provides two significant advantages in overcoming the aforementioned challenges. First, our regulation model requires only connectivity functionalities in vehicles, avoiding reliance on high levels of autonomy. Consequently, it regulates a wide range of vehicle types, including human-driven vehicles (HVs), by equipping them with vehicle-to-infrastructure communication technologies. Our proposed approach enables broader compatibility and achieves a higher regulation rate compared to CAV-specific control methods. Second, the macroscopic traffic flow representation substantially reduces data transmission requirements and eliminates the need for precise vehicle positioning, thereby enhancing scalability and practicality in real-world applications.

This paper theoretically addresses in-domain distributed control of PDE systems using a robust MARL framework. Two closely related works are [18], [19], which propose scalable ramp metering control via MARL applied to the traffic PDE models. Several studies have also demonstrated the successful application of MARL to PDE dynamics in diverse domains, including fluid flow control, drag reduction, and turbulence modeling [50], [51], [52], and [53] further provides a theoretical foundation for distributed PDE control using MARL. To address the concern about the inconsistent improvements between theoretical macroscopic analysis and microscopic simulations [54], we adopt a practical evaluation strategy—directly testing in a microscopic simulator.

The main contributions of this work are as follows:

1) We propose a novel fine-resolution, lane-grid-based lane-change regulation approach for freeway traffic control. The infrastructure-based framework is leveraged to enhance traffic efficiency by regulating lane-change maneuvers at a macroscopic level via a CTDE paradigm.
2) In contrast to traditional MARL-based vehicle control methods, the proposed traffic flow control strategy simplifies communication requirements and reduces reliance on precise vehicle positioning, enabling efficient and scalable implementation in large-scale traffic systems.
3) We establish a training framework by utilizing a microscopic simulator to generate microscopic vehicle trajectories and adopting the multi-lane macroscopic traffic PDEs to compute aggregated lane-changing regulation for CVs. The experimental results show improvements in traffic performance.

To the best of our knowledge, this is the first attempt to employ CVs to regulate discretionary lane change at the lane-grid level. Positioned between infrastructure-level control and direct lane-change maneuver control, our regulation design minimizes information collection burdens while managing lane changes at a macroscopic level without introducing additional safety risks.

This paper is organized as follows. Section II details the modeling of our control problem using PDEs and derives the distributed regulation framework using multi-agent reinforcement learning. Section III-A describes the setup of our experimental environments in the SUMO microscopic simulator. Section IV presents extensive numerical results demonstrating the performance of the proposed model. Finally, Section V concludes the paper.

II. Lane Change Regulation Model

In this section, we first describe the freeway traffic dynamics with lane-change using a generic multi-lane macroscopic traffic flow model, and then propose a multi-agent reinforcement learning control formulation to address this problem.

A. Multi-Lane Macroscopic Traffic Models

While the first-order LWR model [55], [56] fails to capture stop-and-go waves due to its assumption of an equilibrium

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

relationship between vehicle speed and density, second-order traffic models, such as the Aw-Rascle-Zhang (ARZ) model [57], [58], introduce an additional velocity equation to account for traffic waves, acceleration, and deceleration effects. To model lane-change behavior in multi-lane settings, extended models treat each lane separately and couple them through lane-change dynamics.

We consider a highway of length L with $m \in \mathbb{N}$ lanes, where lane 0 denotes the rightmost lane in the direction of travel. Traffic evolves over a time period T. For each lane α, $\rho_\alpha(x,t)$ and $v_\alpha(x,t)$ denote the traffic density and average speed at position x and time t, respectively. The general second-order model incorporating lane-change behavior is:

$$\begin{aligned}
&\frac{\partial \rho_\alpha}{\partial t} + \frac{\partial (\rho_\alpha v_\alpha)}{\partial x} = S_1(\rho, v), \\
&\frac{\partial (\rho_\alpha v_\alpha)}{\partial t} + \frac{\partial (\rho_\alpha v_\alpha^2)}{\partial x} = \rho_\alpha R(\rho_\alpha, v_\alpha) - \frac{\partial p(\rho_\alpha)}{\partial x} + S_2(\rho, v), \quad (1)
\end{aligned}$$

where the relaxation term $R(\rho_\alpha, v_\alpha)$ describes how vehicles adjust to their steady-state speed corresponding to the local density, and the pressure term $p(\rho_\alpha)$ captures the response of the vehicle ensemble to density gradients [59]. The source terms $S_1(\rho, v)$ and $S_2(\rho, v)$ quantify the mass and momentum exchange due to lane changes, respectively. Here, $\rho = (\rho_0, \ldots, \rho_{m-1})$ and $v = (v_0, \ldots, v_{m-1})$ denote the density and speed states of all lanes. Following prior studies [60], [61], the source terms can be explicitly written as:

$$\begin{aligned}
S_1(\rho, v) = &\frac{\rho_{\alpha-1}}{f^{\text{left}}_{\rho,\alpha-1}} a^{\text{left}}_{\alpha-1} + \frac{\rho_{\alpha+1}}{f^{\text{right}}_{\rho,\alpha+1}} a^{\text{right}}_{\alpha+1} \\
&- \frac{\rho_\alpha}{f^{\text{left}}_{\rho,\alpha}} a^{\text{left}}_{\alpha} - \frac{\rho_\alpha}{f^{\text{right}}_{\rho,\alpha}} a^{\text{right}}_{\alpha}, \\
S_2(\rho, v) = &\frac{\rho_{\alpha-1} v_{\alpha-1}}{f^{\text{left}}_{\rho v,\alpha-1}} a^{\text{left}}_{\alpha-1} + \frac{\rho_{\alpha+1} v_{\alpha+1}}{f^{\text{right}}_{\rho v,\alpha+1}} a^{\text{right}}_{\alpha+1} \\
&- \frac{\rho_\alpha v_\alpha}{f^{\text{left}}_{\rho v,\alpha}} a^{\text{left}}_{\alpha} - \frac{\rho_\alpha v_\alpha}{f^{\text{right}}_{\rho v,\alpha}} a^{\text{right}}_{\alpha}, \quad (2)
\end{aligned}$$

where $f^{\text{left, right}}_{\rho,\alpha}(\rho, v)$ and $f^{\text{left, right}}_{\rho v,\alpha}(\rho, v)$ are transition rate functions defined over space and time, representing vehicle mass and momentum exchange from lane α to its adjacent lanes. These functions can be modeled and calibrated based on vehicle interactions and traffic regulations [62], [63].

The actions $a^{\text{left}}_\alpha(x,t), a^{\text{right}}_\alpha(x,t) \in \mathbb{R}$ regulate the magnitude of mass and momentum exchange due to lane changes in the respective directions. The uncontrolled case, where lane changes are not subject to regulation, corresponds to $a^{\text{left}}_\alpha(x,t) \equiv 1$ and $a^{\text{right}}_\alpha(x,t) \equiv 1$ for all (x,t). This formulation allows the uncontrolled scenario to be viewed as a special case of the general model.

When modeling all lanes jointly, we define the global state $y = (\rho_0, v_0, \ldots, \rho_{m-1}, v_{m-1})^T$ and control action $a = (a^{\text{left}}_0, a^{\text{right}}_0, \ldots, a^{\text{left}}_{m-1}, a^{\text{right}}_{m-1})^T$. The system is described using the following partial differential operator:

$$\frac{\partial y}{\partial t} = \mathcal{N}(y, a) = -A(y)\frac{\partial y}{\partial x} + S(y, a), \quad (3)$$

where $A(y)$ governs the characteristic speeds, and $S(y,a)$ represents the aggregated mass and momentum exchange across all lanes.

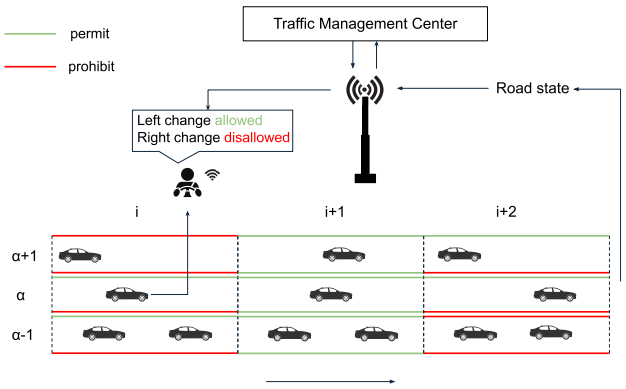


Fig. 1. Model demonstration of lane change regulation signals broadcast to drivers via traffic management center.

B. Grid-Based Lane Change Regulation Strategies

We discretize the system both temporally and spatially. The spatial domain $[0, L]$ is divided into equally spaced intervals, each with grid length Δx, such that a lane is divided into $N_x = L/\Delta x$ grids. The time domain $[0, T]$ is partitioned with time-steps $\Delta T = t_q - t_{q-1}$, where $q = 1, \ldots, N_T$.

Let $\mathcal{V}(t_k) = \{1, 2, \ldots, n_{\text{veh}}(t_k)\}$ denote the set of vehicles at time t_k, where $n_{\text{veh}}(t_k)$ is the number of vehicles in the system. For vehicle $j \in \mathcal{V}(t_k)$, let $\ell^j(t_k)$ denote its lane index, and let $x^j(t_k)$ and $u^j(t_k)$ denote its longitudinal position and velocity, respectively. To incorporate the microscopic data into a PDE model, we need to compute the macroscopic traffic state for each spatial grid. [64] employs a kernel density estimation (KDE) method to estimate density, velocity, and flow functions for the PDE model. Our approach adopts a simpler version of KDE, where each grid is represented by an averaged quantity, rather than requiring detailed positional data, which reduces the need for precise vehicle positioning.

Specifically, the density at the i-th grid in lane α is computed by counting the number of vehicles within the grid at x_i, and the velocity is estimated by averaging their speeds:

$$\begin{aligned}
N_{\alpha_i}(t_k) &= \sum_{j \in \mathcal{V}} \mathbf{1}_{\left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}\right],\ \ell^j = \alpha}(x^j), \\
\rho_{\alpha_i}(t_k) &= \frac{N_{\alpha_i}(t_k)}{\Delta x}, \\
v_{\alpha_i}(t_k) &= \frac{1}{N_{\alpha_i}(t_k)} \sum_{j: x^j \in \left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}\right],\ \ell^j = \alpha} u^j. \quad (4)
\end{aligned}$$

In practice, average speed and density in each grid are acquired through measurements from local detectors, or reports from CVs [18], [65]. The flow is then calculated by multiplying the estimated density and velocity:$q_{\alpha_i}(t_k) = \rho_{\alpha_i}(t_k) \cdot v_{\alpha_i}(t_k)$.

Within our system, we assume the presence of two types of vehicles: HVs, which execute lane changes without intervention from the regulation model, and CVs, which are HVs equipped with connectivity technologies that adhere strictly to regulation signals. Notably, both CVs and HVs follow the same microscopic dynamic models. In the discretized setting of Eq. (2), our model selectively controls lane changes of CVs

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

by permitting or prohibiting transitions from grid α_i to its adjacent lane grids. Specifically, the action for any grid α_i at time t_k is discretized as a two-dimensional vector:

$$a_{\alpha_i}(t_k) = (a_{\alpha_i}^{\text{left}}(t_k), a_{\alpha_i}^{\text{right}}(t_k)) \in \{0,1\}^2. \tag{5}$$

Each component indicates whether lane change requests from grid α_i in a specific direction are permitted at time t_k. If a lane change is disallowed, the corresponding gain or loss term in Eq. (2) is suppressed by multiplying with zero. Otherwise, the source terms behave as in the uncontrolled case, where all actions are set to one. Fig. 1 illustrates how our model functions in a real-world scenario by broadcasting lane change regulations to vehicles through road infrastructure.

In the context of reinforcement learning, the discretized transition function expressed using the partial differential operator in Eq. (3) for each grid column is:

$$\begin{aligned} y_{x=x_i}(t_{k+1}) &= g_{x_i}(y(t_k), a(t_k)) \\ &= y_{x=x_i}(t_k) + \int_{t_k}^{t_{k+1}} \mathcal{N}(y(\tau), a(t_k))d\tau|_{x=x_i}, \end{aligned} \tag{6}$$

where the action $a_{x_i}(t_k)$ is held constant during the time interval (zero-order hold).

Two important properties of traffic PDE models simplify the formulation of the transition function. First, the hyperbolic structure ensures finite wave propagation speeds, so the evolution of a grid over a short time depends only on a localized region of the whole traffic state. Second, the dynamics in Eq. (3) exhibit spatial translation invariance, meaning the system's behavior does not depend on specific spatial locations when border effects are negligible. As a result, the transition function for all grid columns in Eq. (6) can be uniformly approximated by a shared function $\tilde{g}$, enabling parameter sharing across all grid columns.

Two further approximations are made in the training algorithm design:

1) We assume identical geometric characteristics, speed limits, and lane discipline across all lanes in each simulated environment. Neglecting border effects (e.g., where border lanes have only one adjacent lane), spatial translation invariance extends to all grids. Consequently, we assign one agent to each grid and train a model with shared parameters across the entire grid system.
2) The dynamics during safety-critical events may evolve with a spatially inhomogeneous partial differential operator, deviating from Eq. (3). However, these anomalies are confined to a small fraction of the overall space-time domain and thus be neglected in experiment.

C. Multi-Agent Reinforcement Learning

We address the control problem using multi-agent reinforcement learning. The overall architecture of the proposed lane change regulation system is illustrated in Fig. 2. Each grid has an agent, and all agents share network parameters. The target system is formulated as a Partially Observable Markov Decision Process (POMDP). Formally, a POMDP can be represented as a 6-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \mathbf{\Omega}, \mathcal{O})$, where $\mathcal{S}$ denotes the state space, $\mathbf{\Omega}$ represents the observation space, $\mathcal{A}$ is the action space, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ describes the state transition distribution, $\mathcal{O}$ is the probabilistic observation model, and $\mathcal{R}$ is the reward function.

1) State space: The entire state space of the highway is defined as $\mathcal{S} = \prod_{\alpha=0}^{m-1} \prod_{i=1}^{N_x} S_{\alpha_i}$, where the product symbol denotes the Cartesian product over all lane-grid state spaces. Here S_{α_i} represents the state of the grid α_i. Features encoded in each state with the grid include:

- ρ: the total vehicle density.
- ρ_c: the CV density, where $\rho_c \in [0, \rho]$. Here, $\rho_c = \rho$ indicates that all vehicles in the grid are connected vehicles and are subject to regulation.
- v: the average longitudinal vehicle velocity.
- v_c: the average longitudinal velocity of CVs.

In Section II-B, we derive the dynamics model in a fully connected environment where all vehicles are subject to regulation. This model can be extended to a partially connected environment where both HVs and CVs coexist. We assume that HVs and CVs follow the same microscopic models, as CVs are essentially HVs equipped with connectivity functionalities. This assumption of driving behavioral homogeneity ensures that the fundamental diagram and the transition rates f_ρ and $f_{\rho v}$ in Eq. (2) depend only on the combined density, velocity, and flow of both CVs and HVs. The evolution of density and velocity for either CV or HV is governed by the same equations, as expressed in Eq. (2), while regulation actions are applied only to CVs.

The observation space for each grid agent $\mathbf{\Omega}_{\alpha_i}$ consists of a localized subset of the global state, defined as $\mathbf{\Omega}_{\alpha_i} = \{S_{\mathcal{I}_\alpha, \mathcal{N}_i}\}$, where $\mathcal{N}_i \subset \{1, \ldots, N_x\}$ represents the set of neighboring grids along the longitudinal direction, and $\mathcal{I}_\alpha \subset \{0, \ldots, m-1\}$ denotes the set of adjacent lanes. In practice, the scope of $\mathcal{N}_i$ and $\mathcal{I}_\alpha$ is determined by the receptive field of the convolutional layers used in the policy and value networks.

2) Action space: As described in Section II-B, the action space $\mathcal{A}_{\alpha_i}$ for the grid α_i is two-dimensional. Only the lane change requests from CVs are regulated by the action output.

$$a_{t,\alpha_i} \sim \pi_\theta(\mathbf{o}_{t,\alpha_i}) \tag{7}$$

It is important to note that lane change requests from individual vehicles already meet the safety criterion defined in the microscopic physical models. This facilitates the design of the reward function for the proposed model π_θ, as safety measures do not need to be explicitly included.

3) Reward function: The reward function evaluates the local performance of each grid agent, focusing primarily on traffic efficiency by assessing congestion levels. A comprehensive review of road traffic congestion measures is provided in [66].

The reward r_1 measures efficiency based on the speed performance index [67]. For a given grid agent α_i with local area $\mathcal{B}(\alpha_i)$, r_1 is defined as:

$$r_1 = \frac{1}{|\mathcal{B}(\alpha_i)|} \sum_{\tilde{\alpha}_k \in \mathcal{B}(\alpha_i)} \frac{v_{\tilde{\alpha}_k}}{v_{max}}. \tag{8}$$

Here, the aggregated speed information of all vehicles in a grid is evaluated, rather than just connected vehicles. This approach is chosen because the objective is to improve overall traffic efficiency through the actions of CVs.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

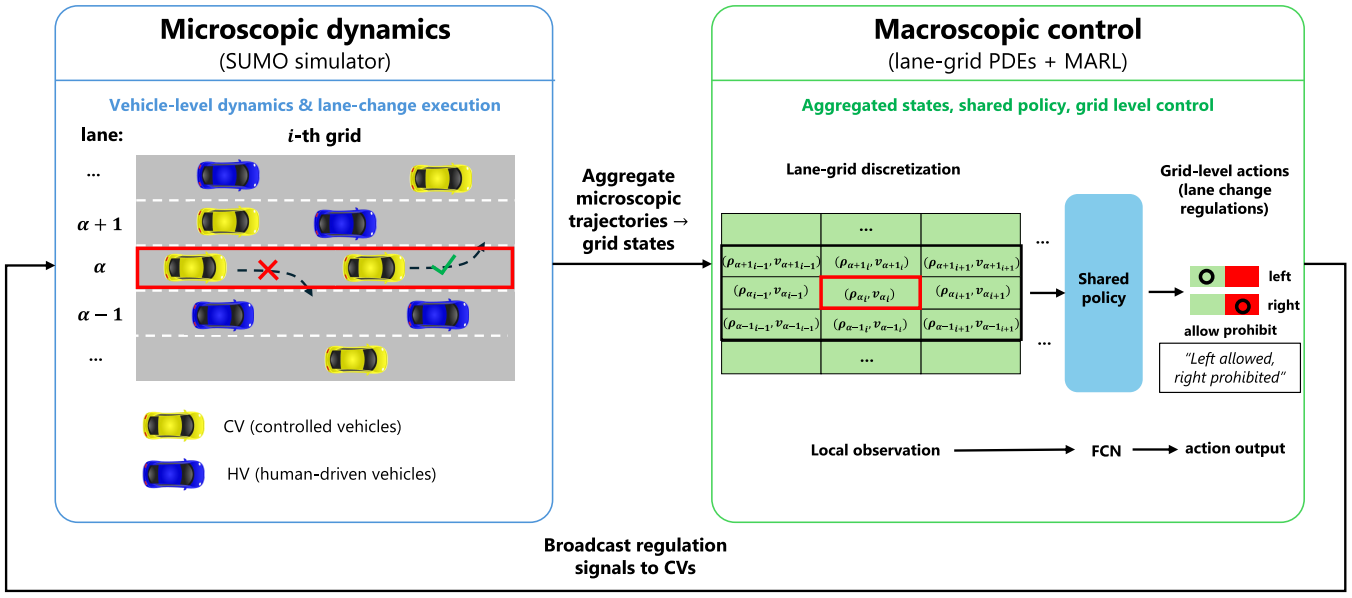


Fig. 2. The architecture of the proposed lane-change regulation model, integrating the microscopic traffic simulator SUMO for vehicle-level dynamics and macroscopic traffic PDEs for aggregated lane-change decisions.

TABLE I
PRIMARY HYPERPARAMETERS USED DURING TRAINING

Parameters	Value
lane grid length (m)	100
discount factor γ	0.95
reward weight of speed ω_1	0.5
reward weight of density ω_2	0.5
road max density (veh/(lane∗m))	0.133
road max speed (m/s)	24.59

We also consider traffic density by defining r_2 based on the volume-to-capacity ratio (V/C):

$$r_2 = \frac{1}{|\mathcal{B}(\alpha_i)|} \sum_{\tilde{\alpha}_k \in \mathcal{B}(\alpha_i)} \left(1 - \frac{\rho_{\tilde{\alpha}_k}}{\rho_{max}}\right). \tag{9}$$

Unlike the discretized Level of Service (LOS) indicator used in [66], we adopt a continuous function with respect to ρ as defined in Eq. (9). This choice ensures training stability, especially in scenarios with high-frequency oscillations, whereas the LOS value can fluctuate significantly over time. In Eq. (8) and (9), v_{max} and ρ_{max} are determined by the simulation settings, as shown in Table I. The $\mathcal{B}(\alpha_i)$ used in all the experiments includes all grids within a rectangular area: $\mathcal{B}(\alpha_i) = \{\tilde{\alpha}_k | \tilde{\alpha} = \alpha - 1, \alpha, \alpha + 1; |x_i - x_k| \leq 2\Delta x\}$. The overall reward function for grid agent α_i is defined as a weighted sum of r_1 and r_2:

$$r = w_1 r_1 + w_2 r_2. \tag{10}$$

III. EXPERIMENTAL SETUP

A. Simulation Environment and Settings

We build a 5-lane highway road in the SUMO microscopic simulator [68]. The road length is 1000 m and lane width is 3.2 m. The speed limit of each lane is set to 24.59 m/s (55 mph). Vehicles arrive from the entrance of the road, and the number of vehicles emitted each second follows

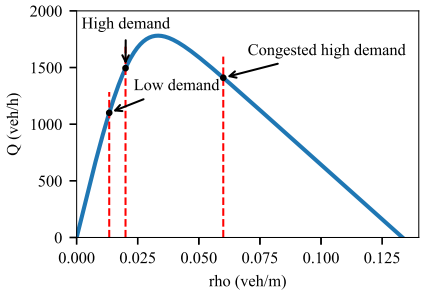


Fig. 3. Demand settings with the fundamental diagram.

TABLE II
PARAMETER SETTING FOR IDM MODEL

Symbol	Description	Value
L	Vehicle length	5 m
v_0	Desired speed	24.59 m/s
T	Time gap	1.4 s
s_0	Minimum gap	2.5 m
δ	Acceleration exponent	4
a	Acceleration	0.73 m/s^2
b	Comfortable deceleration	1.67 m/s^2

TABLE III
TRAFFIC DEMAND SETTINGS IN SIMULATION SCENARIOS

	Inflow (veh/(lane*h))	Density (veh/(lane*m))	Velocity (m/s)
Low demand	1100	0.013	22.93
High demand	1495	0.02	20.76
Congested High demand	1410	0.06	6.53

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

a binomial distribution. The intelligent driver model (IDM) [69] is selected as the car-following model for all simulated vehicles. The corresponding parameter settings are shown in Table II. In addition, we set the lateral resolution to 0.8m and choose the SL2015 model [70] as the lane-changing model. Table III shows the traffic demands in the simulation scenarios. The relation of inflow and density is decided by the fundamental diagram derived from IDM car-following model, as illustrated in Fig. 3.

Given a demand setting, we test our model in different scenarios. In addition to a stable inflow scenario, typical safety-critical events are included to evaluate the model's performance under transient and hazardous traffic conditions. The scenarios are described as follows.

1) *Stable Flow*: In a stable flow scenario, all lanes share a constant inflow rate selected from Table III. All vehicles depart from the entry position of the road.
2) *Lane Degrade*: In a lane degrade scenario, one or more lanes can experience reduced capacity due to narrow lane width, dangerous road conditions, or reduced visibility range. This can be simulated by assigning an increased time gap parameter in the IDM model to vehicles in the affected lane [69]. Built upon a *Stable Flow* scenario, we create lane degrade events with random lane, location range, and time range. The new time gap is set in the range [4.0, 10.0] seconds. Lane change actions are prohibited for all vehicles traveling in hazardous lane grids, regardless of the decisions made by the driver or our model.
3) *Vehicle Stop*: In a vehicle stop scenario, a lane block event is created by an accidental vehicle. We simulate this event in this way: at a given trigger time, a vehicle will decelerate relatively aggressively until the speed reaches zero. Then the vehicle remains stopped for a random period of time. During the recovery stage, the accidental vehicle and its blocked followers will resume their route.

Scenario identification is essential for selecting the appropriate policy for deployment. In practice, real-time road conditions can be inferred from official sources (e.g., transportation agencies) and crowdsourced GPS data. Commercial navigation apps like Google Maps use such data to detect incidents and congestion [71], which can also support timely policy switching in our system.

B. Training

We train our policy using the Double Deep Q-Network (Double DQN) algorithm, which is well-suited for tasks with discrete action spaces. Double DQN addresses the overestimation bias commonly found in traditional Deep Q-Networks (DQN) by decoupling action selection from value estimation [72]. This improves training stability and results in more accurate value function estimates. The full training procedure is detailed in Algorithm 1. To evaluate the effectiveness of our method, we compare it against a benchmark policy trained using the Proximal Policy Optimization (PPO) algorithm, a widely used on-policy reinforcement learning method known for its sample efficiency and robustness.

Algorithm 1 The Pseudo-Code for Training the Regulation Policy Based on Double DQN

Initialize: Q-network Q_θ, target network Q_{θ^-}, and replay buffer $\mathcal{D}$.
For step=1,..., M
  Reset the environment state s_0.
  While not *done*
    For each agent
      Select action with exploration probability ϵ
        $a_{t,\alpha_i} = \arg\max_a Q_\theta(\mathbf{s}_{t,\alpha_i}, a)$
    End
    Execute a_t in SUMO and observe the aggregated lane-grid state s_{t+1}.
    Store global transition in replay buffer $\mathcal{D}$
      $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathbf{s}_t, a_t, r_t, \mathbf{s}_{t+1})$
    Sample $B = \{(\mathbf{s}_k, a_k, r_k, \mathbf{s}_{k+1})\}$ from $\mathcal{D}$
    For each agent
      Calculate agent reward r_{j,α_i} by using Eq. (10).
      Compute target Q value:
      $y_{j,\alpha_i} = r_{j,\alpha_i} +$
        $\gamma Q_{\theta^-}(\mathbf{s}_{j+1,\alpha_i}, \arg\max_a Q_\theta(\mathbf{s}_{j+1,\alpha_i}, a))$
    End
    Reshape batch into $\tilde{B}$ with size of $|B| \times num_{agents}$.
    Perform gradient descent on loss
      $\mathcal{L}(\theta) = \mathbb{E}_{\tilde{B}}[(y_j - Q_\theta(\mathbf{s}_j, a_j))^2]$
    If step mod T_{update} == 0
      Update target network $Q_{\theta^-} \leftarrow Q_\theta$.
    End
    step$\leftarrow$ step+1
  End
End

For each scenario and demand condition, we train a separate policy network which employs a fully convolutional neural network (FCN) with ReLU activation functions and a discount factor of 0.95. A training environment is created using the SUMO. Simulation parameters include a unified simulation step $\delta = 0.1s$, a reward step $\Delta t_r = 1.0s$, and an environmental step size $\Delta t_{env} = 4.0s$ for consistency across environments. The reward for each step is the average reward calculated over the interval Δt_{env}, with samples taken at regular intervals of Δt_r. The lane grid length Δx_{env} is set to 100 m, as summarized in Table I. At the end of each episode, the training environment is reset by generating the traffic demand using a random seed. In scenarios involving safety-critical events, additional parameters such as the location range and affected lanes are also randomized to improve generalization.

Training is conducted over 250,000 steps, with performance metrics calculated at the end of each episode. The experimental environment is implemented using SUMO, the Traffic Control Interface (TraCI), and OpenAI Gym APIs. All experiments are conducted on a computer equipped with an Nvidia GeForce RTX 4090 GPU and 64 GB RAM. The software environment uses Python 3.9 and PyTorch 1.13.1.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

C. Evaluation

The primary goal of the proposed model is to optimize the efficiency of the entire freeway section. At the same time, we need to monitor whether our model causes negative effects on driving safety and energy cost. Therefore, traffic efficiency, safety, and energy cost are the performance metrics of interest. In this context, we use average travel speed of all vehicles for evaluating traffic efficiency. Energy cost is measured by CO_2 emission. We adopt the *HBEFA4/PC_petrol_Euro-6d* emission model for all simulated vehicles. Since our model does not intervene directly on vehicle-level maneuvers, vehicle-level collision avoidance is handled by the microscopic simulator, and no collisions are observed in our evaluations. Besides the collision rate indicator, time-to-collision (TTC) based measures are also commonly used [33], [73]. When TTC is less than a threshold, a collision risk is indicated. We use two TTC-based indicators: average TTC and time exposed TTC (TET) [74]. Compared to average TTC, time exposed TTC is normalized by travel time of each vehicle. Definitions of the two safety indicators are given in Eq. (11) and Eq. (12). The critical threshold TTC^* is selected according to the traffic condition. For the low and high demand scenarios, we use $TTC^* = 1.5$s, following common practice in surrogate safety analysis [75], [76]. For the congested high demand scenario, we use $TTC^* = 5.0$s as a more conservative threshold to capture potential conflicts earlier under low-speed and small-gap traffic conditions [74].

$$\overline{\text{TTC}} = \frac{1}{N_{veh}} \sum_{i=1}^{N_{veh}} \sum_{t=1}^{T_i} \mathbf{1}_i(TTC_i(t) < TTC^*) \cdot \delta, \tag{11}$$

$$\text{TET} = \frac{1}{N_{veh}} \sum_{i=1}^{N_{veh}} \sum_{t=1}^{T_i} \frac{1}{T_i} \cdot \mathbf{1}_i(TTC_i(t) < TTC^*). \tag{12}$$

We evaluate models trained in each environment at varying checkpoints. The baseline model is defined to follow the SUMO IDM car-following model and SL2015 lane-changing model for all vehicles. Since each traffic environment in SUMO is initialized with an empty road, all data collected during the warm-up phase is excluded when computing metrics in both the training and evaluation stages.

IV. Numerical Results

This section evaluates the performance of the lane-change regulation model in different traffic scenarios and under different penetration rates of CVs. We first evaluate the model's performance under a 100% CV penetration rate, and then examine how it varies in mixed-traffic scenarios with partial CV penetration in Section IV-D.

A. Performance of the Lane Change Regulation Model

To minimize the dependency on precise parameter initialization and enhance training stability, the states are normalized to the range $[0, 1]$ by dividing each dimension by its respective maximum value. These maximum values of vehicle speed and density are determined by the fundamental diagram. The primary hyperparameters of our algorithm are listed in Table I. Fig. 4 illustrates the average accumulated reward across agents per episode during the training phase. Under free-flow demand conditions (low and high), a vehicle stop event substantially reduces reward performance. In contrast, compared to congested high demand conditions, a significant improvement trend is observed under free-flow demand conditions across all scenarios. As shown in Fig. 4, this improvement is primarily attributed to the enhancement in the speed performance index reward term associated with vehicle travel speeds. This indicates that the policy seeks to optimize traffic speed without significantly changing the equilibrium state of the traffic, which is governed by the spatial distribution of traffic density.

Table IV summarizes the results of our model and the baselines across all evaluation environments. Comparative metrics against the baselines are presented in Table V, with arrows indicating the optimization direction for each indicator. Our model outperforms the baseline in terms of traffic efficiency under low and high demand conditions, while demonstrating comparable performance under congested high demand conditions. As shown in Table VI, the average number of lane changes per vehicle in the *Stable Flow* scenario decreases sharply under congested high demand conditions compared to low demand conditions. Specifically, it decreases by 87.1% for the baseline and 89.5% for our model. Moreover, our regulation model suppresses lane-changing behavior compared to the baseline. For example, in the *Stable Flow* scenario, the lane change frequency decreases by 7.5%, 39.0%, and 25.0% under low, high, and congested high demand conditions, respectively. This indicates a link between the optimization strategy of our model and the suppression of lane change frequency. Consequently, it can be inferred that low lane change frequency in congested conditions restricts the optimization potential of our model, leading to its comparable performance under congested high demand conditions.

When energy cost is considered, a clear trade-off emerges between travel efficiency and energy cost in our model's performance. If the trade-off ratio between these two metrics is set to 1:1, our model performs better under high demand conditions compared to low demand conditions. For example, in the *Lane Degrade* scenario under low demand conditions, our model achieves a 4.4% improvement in average speed at the cost of a 3.2% increase in CO_2 emissions. In contrast, under high demand conditions in the same scenario, these values are 4.5% and 1.9%, respectively. Considering the performance uplift margin, it is crucial to assess whether our model consistently achieves these gains. The environment is constructed to ensure fairness in comparison for our model and the baseline by maintaining the same vehicle route demand, safety-critical event settings, and random seeds in each evaluation round.

Fig. 5 shows the distribution of performance uplift across episodes in all environments. In low and high demand conditions, where our model outperforms the baseline, performance uplift is consistent across episodes. The *Stable Flow* scenario exhibits less fluctuation than the *Lane Degrade* and *Vehicle Stop* scenarios, where fluctuations are primarily attributed

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

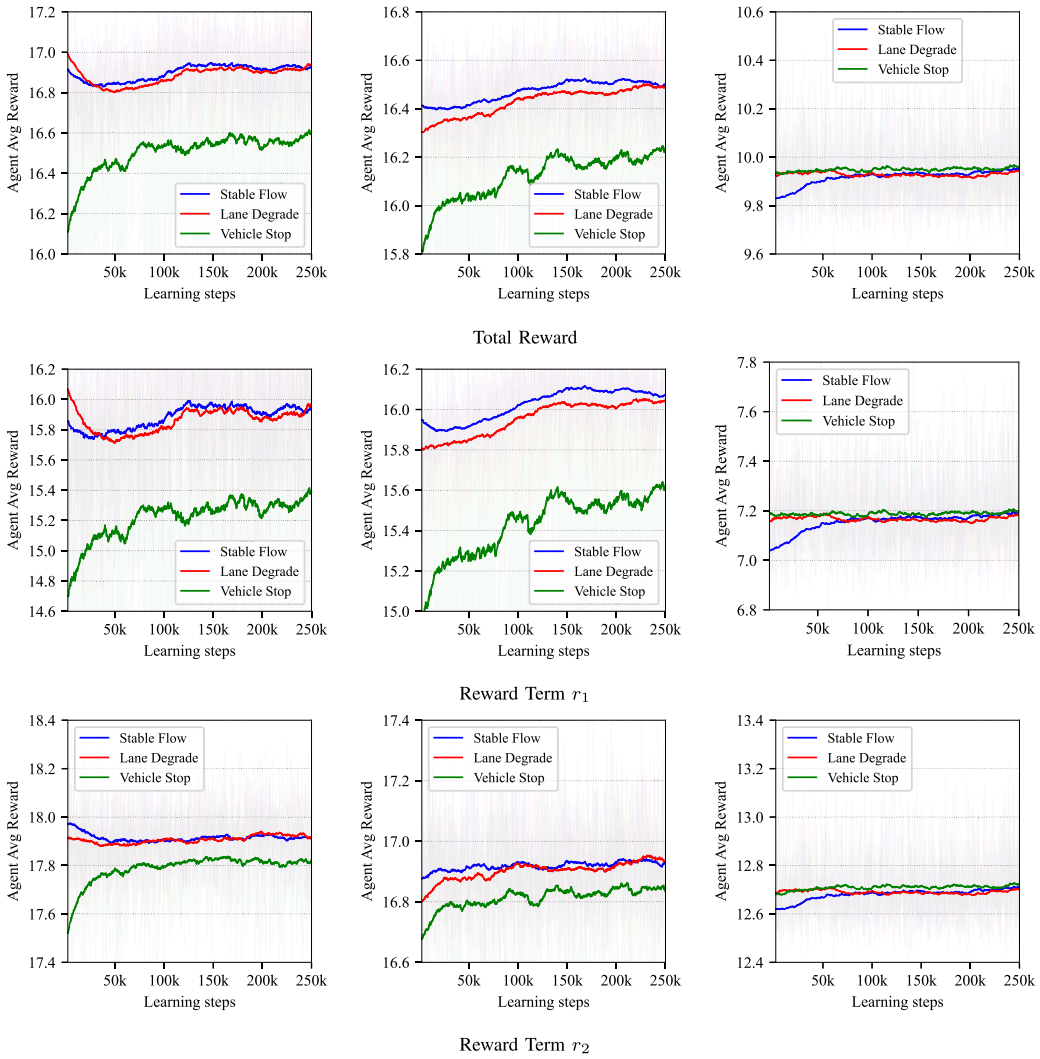


Fig. 4. Average agent reward per episode during the training phase under varying environments. From left to right, the columns represent low, high, and congested high demand settings. The top row shows the total reward, while the second and third rows display the rewards for terms r_1 and r_2, respectively. Solid lines represent an exponential moving average for smoothing.

TABLE IV

PERFORMANCE METRICS (MEAN AND STANDARD DEVIATION) ACROSS SCENARIOS UNDER DIFFERENT DEMAND CONDITIONS

Metric	Demand	Stable Flow		Lane Degrade		Vehicle Stop	
		Baseline	Our model	Baseline	Our model	Baseline	Our model
Average speed (m/s)	Low	21.78 ± 0.12	22.38 ± 0.10	20.69 ± 0.94	21.60 ± 0.62	20.08 ± 1.30	21.11 ± 0.83
	High	19.29 ± 0.18	19.78 ± 0.17	18.46 ± 0.88	19.28 ± 0.52	18.07 ± 0.87	18.93 ± 0.61
	Congested High	8.54 ± 0.12	8.56 ± 0.15	8.87 ± 0.36	8.83 ± 0.33	8.84 ± 0.42	8.9 ± 0.47
CO_2 emission (g)	Low	122.68 ± 0.89	126.69 ± 0.70	126.12 ± 3.53	130.11 ± 3.13	132.76 ± 9.69	134.14 ± 6.58
	High	119.61 ± 1.06	122.73 ± 1.08	125.71 ± 5.77	128.14 ± 4.59	133.23 ± 11.27	132.89 ± 7.32
	Congested High	256.5 ± 3.24	256.25 ± 3.83	259.56 ± 11.85	259.93 ± 11.83	259.24 ± 7.83	258.87 ± 8.19
$\overline{\text{TTC}}$ (s)	Low	0.00 ± 0.00	0.00 ± 0.00	0.04 ± 0.07	0.03 ± 0.07	0.05 ± 0.10	0.01 ± 0.02
	High	0.04 ± 0.03	0.00 ± 0.01	0.16 ± 0.19	0.02 ± 0.03	0.13 ± 0.14	0.04 ± 0.08
	Congested High	3.01 ± 0.28	2.98 ± 0.29	2.86 ± 0.46	2.88 ± 0.46	2.7 ± 0.58	2.67 ± 0.62
TET (%)	Low	0.00 ± 0.01	0.00 ± 0.00	0.05 ± 0.08	0.04 ± 0.09	0.07 ± 0.12	0.01 ± 0.02
	High	0.08 ± 0.05	0.01 ± 0.01	0.22 ± 0.20	0.03 ± 0.04	0.19 ± 0.16	0.06 ± 0.09
	Congested High	2.37 ± 0.23	2.34 ± 0.23	2.15 ± 0.29	2.18 ± 0.28	2.07 ± 0.46	2.05 ± 0.49

to safety-critical events. This observation aligns with the higher fluctuation in training rewards under scenarios involving safety-critical events, as shown in Fig. 4.

Regarding safety metrics $\overline{\text{TTC}}$ and TET, the values under low and high demand conditions are close to zero for both the baseline and our model under the TTC^* = 1.5s threshold.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

TABLE V

COMPARISON OF PERFORMANCE METRICS BETWEEN THE PROPOSED MODEL AND BASELINES ACROSS SCENARIOS UNDER VARYING DEMAND CONDITIONS

Metric	Stable Flow			Lane Degrade			Vehicle Stop		
	Low	High	Congested High	Low	High	Congested High	Low	High	Congested High
Average Speed (↑)	2.8	2.6	0.3	**4.4**	**4.5**	-0.5	**5.1**	**4.7**	0.5
CO_2 Emission (↓)	3.3	2.6	-0.1	3.2	1.9	0.1	1.0	-0.3	-0.1
$\overline{\text{TTC}}$ (↓)	-	-	-1.2	-	-	1.0	-	-	-1.2
TET (↓)	-	-	-1.2	-	-	1.2	-	-	-1.2

The symbol "-" indicates that the corresponding values are close to zero.

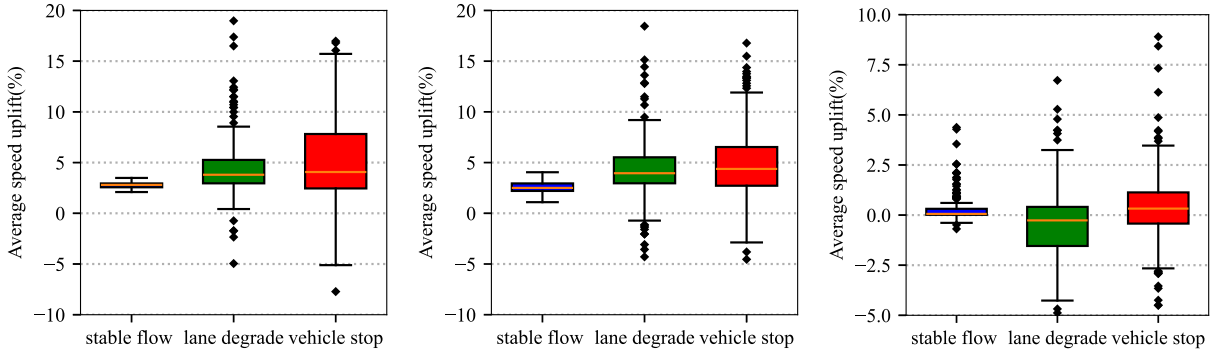


Fig. 5. Boxplots of performance uplift across episodes under varied demand conditions. Each environment is evaluated over 200 episodes. From left to right, the plots represent low, high, and congested high demand conditions. The colored boxes represent the interquartile range (IQR) containing 50% of the data near the median. Outliers are shown as black diamonds.

TABLE VI

AVERAGE NUMBER OF LANE CHANGES PER VEHICLE PER EVALUATION EPISODE UNDER VARIOUS DEMAND CONDITIONS

Scenario	Models	Demand		
		Low	High	Congested High
Stable Flow	Baseline	0.093±0.012	0.059±0.012	0.012±0.012
	Ours	0.086±0.011	0.036±0.008	0.009±0.009
Lane Degrade	Baseline	0.149±0.062	0.142±0.076	0.049±0.030
	Ours	0.106±0.043	0.057±0.021	0.043±0.027
Vehicle Stop	Baseline	0.143±0.058	0.135±0.061	0.061±0.036
	Ours	0.104±0.025	0.076±0.038	0.050±0.032

TABLE VII

CROSS-EVALUATION RESULTS OF THE MODEL TRAINED IN THE STABLE FLOW SCENARIO APPLIED TO OTHER SCENARIOS

Metric	Lane Degrade		Vehicle Stop	
	Low	High	Low	High
Average Speed (↑)	4.4/4.2*	4.5/2.9	5.1/4.1	4.7/2.7
CO_2 Emission (↓)	3.2/3.0	1.9/2.2	1.0/1.8	-0.3/1.0

*The value on the left is from Table V, while the number on the right is from model cross validation.

Under congested high demand conditions, where $\text{TTC}^* = 5.0$s is used, the changes remain small, within approximately ±1.2% across all scenarios. Although no safety-related term is included in the reward function, the model does not introduce evident safety degradation in the evaluated scenarios. This indicates that the proposed regulation framework can improve traffic efficiency while maintaining comparable safety performance in the microscopic simulation.

B. Policy Behavior Analysis

Here, we analyze how our model optimizes traffic efficiency through lane-change regulation. The SL2015 lane-change model used in our experiments closely aligns with driver lane-change behaviors in real-world freeways, as described in [70]. Specifically, the real-world regulation "Keep to the Right" is represented in SL2015 by the "Keep Right" lane-change intention, where vehicles are obligated to clear the overtaking lane. Additionally, lane changes based on the "Speed Gain" intention are accurately modeled in SL2015, capturing expected speed gains exceeding an empirical threshold. By default, decisions to speed up by changing to the right lane require more deliberation compared to leftward lane changes. To evaluate the policy behavior, we visualize the distribution of actions taken by our model in the *Stable Flow* scenario under low and high demand conditions in Fig. 6. We compare the action distribution at training checkpoints 100k and 200k for each evaluation environment and highlight two observations:

1) The policy remains relatively permissive on the middle lanes. The rate of allowing at least one lane-change direction, denoted by 'left | right' stays above 80% under low demand and above 40% under high demand. This

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII
COMPARISON OF PERFORMANCE METRICS (AVERAGE SPEED / CO_2 EMISSIONS) BETWEEN DOUBLE DQN AND PPO

Algorithms (state, checkpoints)	Stable Flow		Lane Degrade		Vehicle Stop	
	Low	High	Low	High	Low	High
PPO (converged, $150k$)	0.0/0.0	3.0/2.5	4.3/3.2	0.3/-1.1	0.0/0.0	4.3/-0.4
DDQN (not converged, $150k$)	2.1/2.7	1.2/1.7	2.5/2.3	3.6/1.9	4.8/1.4	1.3/-1.8
DDQN (converged, $200k$)	2.8/3.3	2.6/2.6	4.4/3.2	4.5/1.9	5.1/1.0	4.7/-0.3

All values are percentage changes compared with the uncontrolled baseline in each environment.

indicates that the model does not simply prohibit all lane changes, but learns different regulation patterns under different demand conditions.

2) The learned action distribution shows clear demand-dependent behavior. Under high demand, the policy rarely selects the "right only" action. In addition, on the middle lanes, the rate of allowing at least one lane-change direction gradually decreases from the right-side middle lanes to the left-side middle lanes under high demand, while it remains close to 100% under low demand. This pattern is consistent with the sparse traffic state under low demand and the stronger need for lane-change regulation under high demand.

Through this analysis, we further evaluate the performance of the policy trained in the *Stable Flow* scenario when applied to other scenarios. As shown in Table VII, cross-validation performance deteriorates under high demand conditions while remaining relatively close under low demand conditions. Under low demand, the average speed improvement changes from 4.4% to 4.2% in the *Lane Degrade* scenario and from 5.1% to 4.1% in the *Vehicle Stop* scenario. Under high demand, the deterioration is more evident. For example, the uplift in average speed decreases by 35.6%, from 4.5% to 2.9%, in the *Lane Degrade* scenario. This performance discrepancy can be attributed to the vehicle density differences between the two demand conditions. The sparse vehicle distribution under low demand conditions mitigates the impact of safety-critical events, whereas out-of-distribution congested traffic states induced by safety-critical events under high demand conditions are the primary cause of poor performance.

C. Performance Comparison With PPO

Previous work [77] investigates the performance of the off-policy Deterministic Policy Gradient (DDPG) and the on-policy PPO algorithm in an autonomous driving task with a continuous action space. The authors conclude that while PPO achieves competitive performance within a short training period, the off-policy DDPG algorithm performs better after convergence. In our setting, although the action space is discrete, PPO remains a suitable baseline for performance comparison. We evaluate both PPO and Double DQN at the same training checkpoint where PPO demonstrates convergence, and also include the performance of the converged Double DQN model from Section IV-A. The results are summarized in Table VIII.

PPO exhibits faster convergence and better sample efficiency at the evaluated $150k$ checkpoint. At this intermediate training stage, PPO performs competitively in several non-congested cases, such as the *Stable Flow* scenario under high demand and the *Lane Degrade* scenario under low demand. However, its performance is less stable across scenarios. For example, in the *Stable Flow* and *Vehicle Stop* scenarios under low demand, PPO converges to a dummy policy where all lane changes are allowed, resulting in no improvement over the uncontrolled baseline. In addition, PPO achieves only a 0.3% speed improvement in the *Lane Degrade* scenario under high demand.

Compared with PPO and the intermediate Double DQN checkpoint, the converged Double DQN checkpoint provides more consistent improvements across the evaluated scenarios. Given the relatively small discrete action space in our lane-change regulation problem, Double DQN proves to be an effective choice for training the proposed policy, offering stable performance with an acceptable training cost.

D. Performance Under Varying Penetration Rate of CVs

In previous experiments, all vehicles were subject to regulation (i.e., 100% of vehicles were CVs). Here, we investigate how the performance of our model changes under different CV penetration rates. For each specified regulation rate, vehicles are randomly assigned as CVs in proportion to that rate, and the model is trained and evaluated accordingly. We select the *Stable Flow* scenario under low demand conditions to examine the effect of regulation rate on model performance. Fig. 7 presents the training rewards and evaluation metrics under varying CV penetration rates.

As the regulation rate decreases from 100% to 10%, the average agent reward per episode during training also declines. Evaluation results from converged checkpoints show that both average travel speed and CO_2 emissions decrease as the regulation rate drops. In Section IV-A, we observed that the efficiency gain in speed closely balances the cost increase in emissions for the *Stable Flow* scenario under low demand. This conclusion remains valid for partial regulation, as long as the CV penetration rate exceeds a certain threshold.

When the regulation rate falls below 30%, the model's performance becomes comparable to the uncontrolled baseline. This suggests that a minimum level of regulation is necessary to activate the benefits of our framework. Once this requirement is met, our model continues to function effectively even when full lane change regulation is not accessible.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

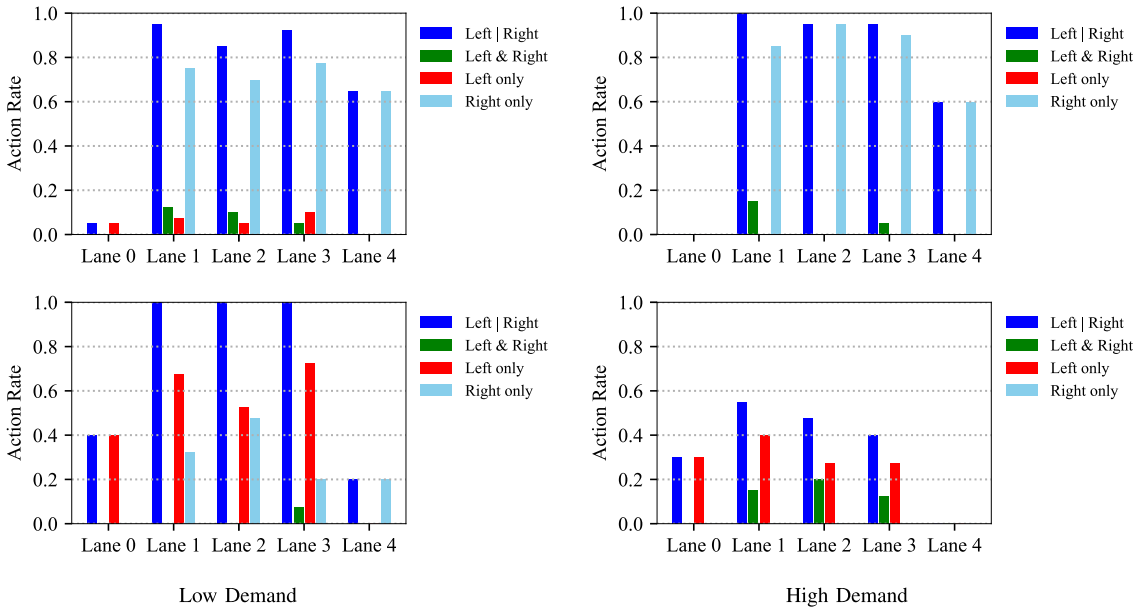


Fig. 6. Distribution of actions across different training checkpoints for the *Stable Flow* scenario under low and high demand conditions. The training steps for each checkpoint, from top to bottom, are 100k and 200k steps. The lane index starts from 0, corresponding to the rightmost lane. Actions are collected every 1000 steps during evaluation for all lane grids along each lane.

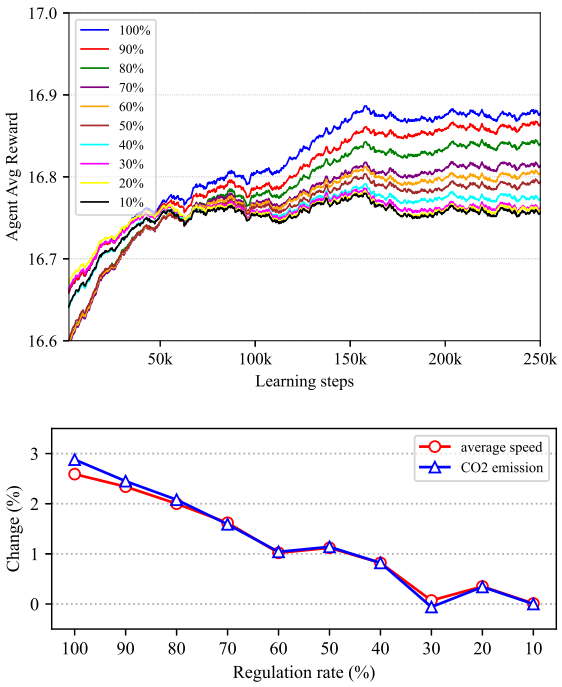


Fig. 7. Training reward and evaluation performance under varying regulation rates. Top: average training reward per episode. Bottom: evaluation results in terms of travel speed and energy cost.

V. Conclusion

This paper presents a novel lane change regulation model designed to enhance overall traffic efficiency. The approach introduces the concept of regulating individual connected vehicle lane change decisions through a macroscopic traffic control framework. A multi-agent reinforcement learning algorithm is proposed, wherein the action space determines whether to permit or prohibit lane change actions within each lane grid. Simulation environments are constructed across various scenarios and demand conditions to comprehensively evaluate the performance of the proposed model.

In scenarios with 100% CV penetration, experimental results demonstrate that our model achieves comparable or superior performance relative to the baseline when balancing traffic efficiency and energy cost. Performance analysis reveals that the model learns demand-dependent and lane-dependent regulation patterns rather than simply prohibiting all lane changes. Additionally, experiments with varying regulation rates indicate that the model continues to positively influence traffic with the control rate above a certain threshold.

Three promising directions for future research are identified. First, the current model assumes that all connected vehicles fully comply with regulation signals. However, individual objectives may conflict with the global goal, potentially leading to regulation violations. To address this, future work can incorporate a high-fidelity simulator that models human compliance behavior. Since cross-environment experiments reveal performance degradation, exploring methods to train a unified, generalized policy adaptable to varied demand conditions and scenarios deserves investigation. Finally, real-world deployment requires bridging the simulation-to-reality gap. Future work should generalize to a wider range of real-world demand conditions and scenarios.

References

[1] S. Cui, B. Seibold, R. Stern, and D. B. Work, "Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations," in *Proc. IEEE Intell. Vehicles Symp.* (IV), Jun. 2017, pp. 1336–1341.

[2] R. E. Stern et al., "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 205–221, Apr. 2018.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

[3] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 143–163, Oct. 2016.

[4] J. Wang, Y. Zheng, K. Li, and Q. Xu, "DeeP-LCC: Data-EnablEd predictive leading cruise control in mixed traffic flow," *IEEE Trans. Control Syst. Technol.*, vol. 31, no. 6, pp. 2760–2776, Nov. 2023.

[5] C. Zhao, H. Yu, and T. G. Molnar, "Safety-critical traffic control by connected automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 154, Sep. 2023, Art. no. 104230.

[6] C. Zhao, T. G. Molnar, and H. Yu, "Safety-critical stabilization of mixed traffic by pairs of CAVs," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2024, pp. 743–748.

[7] Y. Chen, G. Orosz, and T. G. Molnar, "Safety-critical connected cruise control: Leveraging connectivity for safe and efficient longitudinal control of automated vehicles," in *Proc. IEEE 27th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2024, pp. 110–115.

[8] J. Rios-Torres and A. A. Malikopoulos, "Impact of partial penetrations of connected and automated vehicles on fuel consumption and traffic flow," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 4, pp. 453–462, Dec. 2018.

[9] J. I. Ge, S. S. Avedisov, C. R. He, W. B. Qin, M. Sadeghpour, and G. Orosz, "Experimental validation of connected automated vehicle design among human-driven vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 91, pp. 335–352, Jun. 2018.

[10] A. Vahidi and A. Sciarretta, "Energy saving potentials of connected and automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 822–843, Oct. 2018.

[11] M. Cicic, X. Xiong, L. Jin, and K. H. Johansson, "Coordinating vehicle platoons for highway bottleneck decongestion and throughput improvement," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8959–8971, Jul. 2022.

[12] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative adaptive cruise control in real traffic situations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 296–305, Feb. 2014.

[13] N. Hyldmar, Y. He, and A. Prorok, "A fleet of miniature cars for experiments in cooperative driving," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 3238–3244.

[14] M. Papageorgiou, H. Hadj-Salem, and F. Middelham, "ALINEA local ramp metering: Summary of field results," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1603, no. 1, pp. 90–98, Jan. 1997.

[15] M. Zhang, T. Kim, X. Nie, W. Jin, L. Chu, and W. Recker, "Evaluation of on-ramp control algorithms," Inst. Transp. Stud., Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB-ITS-PRR-2001-36, 2001.

[16] E. Smaragdis, M. Papageorgiou, and E. Kosmatopoulos, "A flow-maximizing adaptive local ramp metering strategy," *Transp. Res. B, Methodol.*, vol. 38, no. 3, pp. 251–270, Mar. 2004.

[17] M. Papageorgiou, J.-M. Blosseville, and H. Haj-Salem, "Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in Paris: Part II: Coordinated on-ramp metering," *Transp. Res. A, Gen.*, vol. 24, no. 5, pp. 361–370, Sep. 1990.

[18] F. Belletti, D. Haziza, G. Gomes, and A. M. Bayen, "Expert level control of ramp metering based on multi-task deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1198–1207, Apr. 2018.

[19] H. Yu, S. Park, A. Bayen, S. Moura, and M. Krstic, "Reinforcement learning versus PDE backstepping and PI control for congested freeway traffic," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 4, pp. 1595–1611, Jul. 2022.

[20] A. Fares and W. Gomaa, "Freeway ramp-metering control based on reinforcement learning," in *Proc. 11th IEEE Int. Conf. Control Autom. (ICCA)*, Jun. 2014, pp. 1226–1231.

[21] F. Deng, J. Jin, Y. Shen, and Y. Du, "Advanced self-improving ramp metering algorithm based on multi-agent deep reinforcement learning," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3804–3809.

[22] V. Milanes, J. Godoy, J. Villagra, and J. Perez, "Automated on-ramp merging system for congested traffic situations," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 500–508, Jun. 2011.

[23] T. Chen, M. Wang, S. Gong, Y. Zhou, and B. Ran, "Connected and automated vehicle distributed control for on-ramp merging scenario: A virtual rotation approach," *Transp. Res. C, Emerg. Technol.*, vol. 133, Dec. 2021, Art. no. 103451.

[24] X. Liao et al., "Cooperative ramp merging design and field implementation: A digital twin approach based on vehicle-to-cloud communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4490–4500, May 2022.

[25] W. Cao, M. Mukai, T. Kawabe, H. Nishira, and N. Fujiki, "Cooperative vehicle path generation during merging using model predictive control with real-time optimization," *Control Eng. Pract.*, vol. 34, pp. 98–105, Jan. 2015.

[26] Y. Zhou, M. E. Cholette, A. Bhaskar, and E. Chung, "Optimal vehicle trajectory planning with control constraints and recursive implementation for automated on-ramp merging," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3409–3420, Sep. 2019.

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

[28] J. Guo, S. Cheng, and Y. Liu, "Merging and diverging impact on mixed traffic of regular and autonomous vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1639–1649, Mar. 2021.

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

[30] T. L. Pan, W. H. K. Lam, A. Sumalee, and R. X. Zhong, "Modeling the impacts of mandatory and discretionary lane-changing maneuvers," *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 403–424, Jul. 2016.

[31] R. Xu, L. Huang, S. Yin, D. Yao, H. Zhang, and L. Peng, "An analysis on traffic flow characteristics and lane changing behaviors in Beijing urban expressway bottlenecks," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 822–827.

[32] Z. Wang, X. Shi, and X. Li, "Review of lane-changing maneuvers of connected and automated vehicles: Models, algorithms and traffic impact analyses," *J. Indian Inst. Sci.*, vol. 99, no. 4, pp. 589–599, Dec. 2019.

[33] F. V. Monteiro and P. Ioannou, "Safe autonomous lane changes and impact on traffic flow in a connected vehicle environment," *Transp. Res. C, Emerg. Technol.*, vol. 151, Jun. 2023, Art. no. 104138.

[34] M. Menendez and C. F. Daganzo, "Effects of HOV lanes on freeway bottlenecks," *Transp. Res. B, Methodol.*, vol. 41, no. 8, pp. 809–822, Oct. 2007.

[35] A. Naseri, F. Estelaji, A. Samani, R. Omidifar, R. Zahedi, and H. Yousefi, "Simulating the performance of HOV lanes for optimal urban traffic management," *Transp. Res. Interdiscipl. Perspect.*, vol. 23, Jan. 2024, Art. no. 101010.

[36] G. C. K. Wong and S. C. Wong, "A multi-class traffic flow model—An extension of LWR model with heterogeneous drivers," *Transp. Res. A, Policy Pract.*, vol. 36, no. 9, pp. 827–841, Nov. 2002.

[37] S. Logghe and L. H. Immers, "Multi-class kinematic wave theory of traffic flow," *Transp. Res. Part B: Methodol.*, vol. 42, no. 6, pp. 523–541, Jul. 2008.

[38] M. W. Levin and S. D. Boyles, "A multiclass cell transmission model for shared human and autonomous vehicle roads," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 103–116, Jan. 2016.

[39] T. Pan, W. H. K. Lam, A. Sumalee, and R. Zhong, "Multiclass multilane model for freeway traffic mixed with connected automated vehicles and regular human-piloted vehicles," *Transportmetrica A, Transp. Sci.*, vol. 17, no. 1, pp. 5–33, Jan. 2021.

[40] D. Lin, L. Li, and S. E. Jabari, "Pay to change lanes: A cooperative lane-changing strategy for connected/automated driving," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 550–564, Aug. 2019.

[41] M. Tajalli, R. Niroumand, and A. Hajbabaie, "Distributed cooperative trajectory and lane changing optimization of connected automated vehicles: Freeway segments with lane drop," *Transp. Res. C, Emerg. Technol.*, vol. 143, Oct. 2022, Art. no. 103761.

[42] H. Xu, Y. Zhang, C. G. Cassandras, L. Li, and S. Feng, "A bi-level cooperative driving strategy allowing lane changes," *Transp. Res. C, Emerg. Technol.*, vol. 120, Nov. 2020, Art. no. 102773.

[43] S. Bhalla, S. G. Subramanian, and M. Crowley, "Deep multi agent reinforcement learning for autonomous driving," in *Proc. Can. Conf. Artif. Intell. Cham*, Switzerland: Springer, 2020, pp. 67–78.

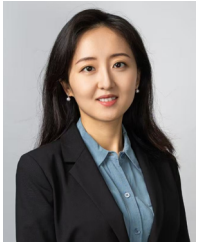
[44] D. Chen et al., "Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11623–11638, Nov. 2023.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.

[45] M. Hua et al., "Multi-agent reinforcement learning for connected and automated vehicles control: Recent advancements and future prospects," 2023, *arXiv:2312.11084*.
[46] M. Carey, C. Balijepalli, and D. Watling, "Extending the cell transmission model to multiple lanes and lane-changing," *Netw. Spatial Econ.*, vol. 15, no. 3, pp. 507–535, Sep. 2015.
[47] W. Lu, Z. Yi, Y. Gu, Y. Rui, and B. Ran, "TD3LVSL: A lane-level variable speed limit approach based on twin delayed deep deterministic policy gradient in a connected automated vehicle environment," *Transp. Res. C, Emerg. Technol.*, vol. 153, Aug. 2023, Art. no. 104221.
[48] W. Lu, Z. Yi, B. Liang, Y. Rui, and B. Ran, "Improving traffic operation of bottleneck in a connected and automated vehicles environment: An integrated lane-level control method," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 11675–11688, Sep. 2024.
[49] Y. Guo, H. Xu, Y. Zhang, and D. Yao, "Integrated variable speed limits and lane-changing control for freeway lane-drop bottlenecks," *IEEE Access*, vol. 8, pp. 54710–54721, 2020.
[50] C. Vignon, J. Rabault, J. Vasanth, F. Alcántara-Ávila, M. Mortensen, and R. Vinuesa, "Effective control of two-dimensional Rayleigh–Bénard convection: Invariant multi-agent reinforcement learning is all you need," *Phys. Fluids*, vol. 35, no. 6, Jun. 2023, Art. no. 065146.
[51] T. Sonoda, Z. Liu, T. Itoh, and Y. Hasegawa, "Reinforcement learning of control strategies for reducing skin friction drag in a fully developed turbulent channel flow," *J. Fluid Mech.*, vol. 960, p. 30, Apr. 2023.
[52] L. Guastoni, J. Rabault, P. Schlatter, H. Azizpour, and R. Vinuesa, "Deep reinforcement learning for turbulent drag reduction in channel flows," *Eur. Phys. J. E*, vol. 46, no. 4, p. 27, Apr. 2023.
[53] S. Peitz, J. Stenner, V. Chidananda, O. Wallscheid, S. L. Brunton, and K. Taira, "Distributed control of partial differential equations using convolutional reinforcement learning," *Phys. D, Nonlinear Phenomena*, vol. 461, May 2024, Art. no. 134096.
[54] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1812–1823, Jul. 2017.
[55] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. London. A. Math. Phys. Sci.*, vol. 229, no. 1178, pp. 317–345, May 1955.
[56] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, Feb. 1956.
[57] A. Aw and M. Rascle, "Resurrection of 'second order' models of traffic flow," *SIAM J. Appl. Math.*, vol. 60, no. 3, pp. 916–938, 2000.
[58] H. M. Zhang, "A non-equilibrium traffic model devoid of gas-like behavior," *Transp. Res. B, Methodol.*, vol. 36, no. 3, pp. 275–290, Mar. 2002.
[59] M. Treiber and A. Kesting, "*Traffic flow dynamics*," in *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Germany: Springer-Verlag, 2013, pp. 983–1000.
[60] J. Song and S. Karni, "A second order traffic flow model with lane changing," *J. Sci. Comput.*, vol. 81, no. 3, pp. 1429–1445, Dec. 2019.
[61] H. Yu and M. Krstic, "Output feedback control of two-lane traffic congestion," *Automatica*, vol. 125, Mar. 2021, Art. no. 109379.
[62] V. Shvetsov and D. Helbing, "Macroscopic dynamics of multilane traffic," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 59, no. 6, pp. 6328–6339, Jun. 1999.
[63] K. N. Porfyri, A. I. Delis, I. K. Nikolos, and M. Papageorgiou, "Calibration and validation of a macroscopic multi-lane traffic flow model using a differential evolution algorithm," in *Proc. Transp. Res. Board 96th Annu. Meeting*, Washington, DC, USA, 2017, pp. 1–17.
[64] S. Fan and B. Seibold, "Data-fitted first-order traffic models and their second-order generalizations: Comparison by trajectory and sensor data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2391, no. 1, pp. 32–43, Jan. 2013.
[65] N. Bekiaris-Liberis and A. I. Delis, "PDE-based feedback control of freeway traffic flow via time-gap manipulation of ACC-equipped vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 1, pp. 461–469, Jan. 2021.
[66] T. Afrin and N. Yodo, "A survey of road traffic congestion measures towards a sustainable and resilient transportation system," *Sustainability*, vol. 12, no. 11, p. 4660, Jun. 2020.
[67] F. He, X. Yan, Y. Liu, and L. Ma, "A traffic congestion assessment method for urban road networks based on speed performance index," *Proc. Eng.*, vol. 137, pp. 425–433, Jan. 2016.
[68] P. A. Lopez et al., "Microscopic traffic simulation using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2575–2582.
[69] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.
[70] J. Erdmann, "SUMO's lane-changing model," in *Proc. 2nd SUMO Conf. Modeling Mobility Open Data*. Berlin, Germany: Springer, 2015, pp. 105–123.
[71] E. D'Andrea and F. Marcelloni, "Detection of traffic congestion and incidents from GPS trace analysis," *Expert Syst. Appl.*, vol. 73, pp. 43–56, May 2017.
[72] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 1–7.
[73] H. Behbahani, N. Nadimi, and S. S. Naseralavi, "New time-based surrogate safety measure to assess crash risk in car-following scenarios," *Transp. Lett.*, vol. 7, no. 4, pp. 229–238, Sep. 2015.
[74] M. M. Minderhoud and P. H. L. Bovy, "Extended time-to-collision measures for road traffic safety assessment," *Accident Anal. Prevention*, vol. 33, no. 1, pp. 89–97, Jan. 2001.
[75] N. Virdi, H. Grzybowska, S. T. Waller, and V. Dixit, "A safety assessment of mixed fleets with connected and autonomous vehicles using the surrogate safety assessment module," *Accident Anal. Prevention*, vol. 131, pp. 95–111, Oct. 2019.
[76] D. Gettman, L. Pu, T. Sayed, and S. G. Shelby, "Surrogate safety assessment model and validation," U.S. Dept. Transp., Federal Highway Admin., Georgetown, VA, USA, Tech. Rep. FHWA-HRT-08-051, 2008, pp. 1–324.
[77] S. Siboo, A. Bhattacharyya, R. Naveen Raj, and S. H. Ashwin, "An empirical study of DDPG and PPO-based reinforcement learning algorithms for autonomous driving," *IEEE Access*, vol. 11, pp. 125094–125108, 2023.



Ke Sun (Student Member, IEEE) received the B.E. degree in automation from Tsinghua University, China, in 2018, and the M.S. degree in computer science from the National University of Singapore, Singapore, in 2019. He is currently pursuing the Ph.D. degree with the Department of Civil and Environmental Engineering, Rice University. His doctoral research focuses on optimization, machine learning, and game theory, with applications to transportation systems. From 2024 to 2025, he was a Research Assistant with the Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou).



Huan Yu (Senior Member, IEEE) received the B.Eng. degree in aerospace engineering from Northwestern Polytechnical University, China, and the M.Sc. and Ph.D. degrees in aerospace engineering from the Department of Mechanical and Aerospace Engineering, University of California, San Diego. She was a Visiting Scholar with the University of California, Berkeley and Massachusetts Institute of Technology. She is currently an Assistant Professor with the Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou). Her research interests include control theory, optimization, and machine learning methodologies and their applications in intelligent transportation systems.

Authorized licensed use limited to: The Hong Kong University of Science and Technology (Guangzhou). Downloaded on October 21,2026 at 02:35:22 UTC from IEEE Xplore. Restrictions apply.