IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 January 2024, accepted 2 March 2024, date of publication 8 March 2024, date of current version 14 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374041

RESEARCH ARTICLE

Research on Multi-Objective Optimization Models for Intersection Crossing of Connected Autonomous Vehicles With Traffic Signals

XIAOBIN NING[1], HAORAN TIAN[1], YONG LIN[2], XUEPING YAO[3], FEI HU[4], AND YUMING YIN[1]

[1]College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310014, China
[2]Zhijiang College of Zhejiang University of Technology, Shaoxing, Zhejiang 310024, China
[3]Hozon New Energy Automobile Company Ltd., Tongxiang, Zhejiang 200062, China
[4]Zhejiang VIE Science and Technology Company Ltd., Zhuji, Zhejiang 311835, China

Corresponding author: Yong Lin (linyong@zjut.edu.cn)

This work was supported in part by the Basic Public Welfare Research Program of Zhejiang Province under Grant LGG22E050019, and in part by the National Natural Science Foundation of China (NSFC) under Grant 51905483.

ABSTRACT This study delves deeply into the traffic light intersection control issue for connected autonomous vehicles (CAVs). First, we employed the vehicle dynamics simulation software, CarSim, to model the vehicle and utilized the intelligent driving simulation software, PreScan, to establish a road environment model. Subsequently, we designed a rule-based traffic light crossing controller (RB-TLCC) for CAVs in MATLAB/Simulink and implemented a co-simulation using CarSim and PreScan. Furthermore, we conducted driver-in-the-loop studies using the Logitech G27 driving simulator kit and compared the experimental results with those from the RB-TLCC. Our findings indicate that RB-TLCC enhances the regenerative braking energy by 23.5%, improves traffic efficiency by 17.9%, and increases driving smoothness by 50.7%. Additionally, based on Markov Chain theory, we proposed a multi-objective optimization model (MO-OM) for CAVs traffic light intersection crossing using the Proximal Policy Optimization algorithm (PPO). A comparison was conducted under two different operating conditions between RB-TLCC and dynamic programming algorithms (DP). The results indicated that the MO-OM proposed in this study exhibited the best comprehensive control performance among the three control methods. While adhering to traffic regulations, it enhances the recuperation of braking energy, efficiency of vehicle passage, and smoothness of travel at traffic light intersections. This study offers effective methodologies for enhancing the performance of CAVs in traffic light intersection control and holds significant reference value for the advancement of future intelligent transportation systems (ITS).

INDEX TERMS Connected autonomous vehicles, intelligent traffic system, multi-objective optimization, optimal speed planning, PPO algorithm, traffic light intersection control.

I. INTRODUCTION

Traffic intersections equipped with traffic lights frequently serve as congestion points in urban transportation because of their role as convergence zones for vehicular flow from different directions. In real-world driving scenarios, drivers often struggle to anticipate upcoming traffic light phases and their precise distance from the stop line. This absence of information complicates the ability of the driver to devise an accurate velocity-adjustment strategy. As vehicles approach an intersection, the clarity of information available to the driver improves; however, decision-making concerning a seamless passage through the intersection remains fraught with uncertainty. This uncertainty often leads to significant speed fluctuations, resulting in frequent stops and starts. This not only exacerbates energy consumption but also increases the risk of traffic accidents [1]. Currently, Vehicle-to-Everything (V2X) technology is advancing

The associate editor coordinating the review of this manuscript and approving it for publication was Hassan Omar.

rapidly. Through V2X, vehicles can obtain real-time traffic information via communication systems between vehicles and roadside infrastructure (V2I) as well as between vehicles themselves (V2V). This enables timely adjustments to vehicle speed, allowing for smoother passage through traffic light intersections [2], [3], [4].

Research on vehicle speed planning directly affects issues such as the energy consumption, traffic efficiency, and ride smoothness during vehicle operation. Currently, there is extensive research on speed planning for Adaptive Cruise Control (ACC) aimed at reducing energy consumption during vehicle operation [5], [6]. Additionally, some studies have established a connection between vehicles and traffic signals based on V2I communication to enable passing through signal intersections without stopping, thereby achieving short travel times and low energy consumption [7], [8], [9]. Scholars have also employed Reinforcement Learning (RL) methods to study speed planning issues, aiming to minimize vehicle waiting time and energy consumption [10], [11].

There is a foundational body of research on energy management strategies for CAVs at traffic light intersections in a V2X environment [12], [13]. Significant progress has been made in studying eco-driving strategies at V2I communication-based signal intersections [14], [15]. Some scholars have focused on multiple objectives in this area, achieving global optimization of traffic efficiency, energy consumption, and passenger comfort [16], [17]. Additionally, research has been conducted to avoid collisions at V2I communication intersections and improve traffic flow [18]. These studies indicate that the current research primarily focuses on using V2X technologies to guide vehicle driving and optimize energy distribution, thereby enhancing fuel economy, traffic efficiency, and safety at signal intersections. However, much of this research focuses on energy management issues in hybrid electric vehicles (HEVs), with a relative scarcity of studies pertaining to purely electric vehicles (EVs).

The current research foundation for ecological driving issues in CAVs has been established [19], [20], [21], [22]. Several scenarios require braking when guiding vehicles through signalized intersections. Some scholars have initiated studies on the braking control issues of CAVs. Existing research on braking control of CAVs mainly focuses on the driving safety of CAVs to prevent collisions between vehicles [23], [24], [25]. Additionally, for EVs equipped with regenerative braking systems (RBS), the deceleration process at traffic light intersections can be optimized using V2X communication technologies. By planning their speed profiles, it is possible to improve the efficiency of braking energy recovery while considering the traffic efficiency, thereby further reducing the overall energy consumption of the vehicle. Regarding related research, Kim et al. [26], [27] introduced an Energy Optimal Deceleration Planning System (EDPS) with the objective of maximizing regenerative braking energy recovery for CAVs. Building on this, they presented a Real-Time Energy Optimal Deceleration Planning System (RT-EDPS). Shang et al. [28] conducted research on regenerative braking control of EVs based on multi-source data, such as image recognition, GPS, and vehicle operating parameters. They designed an adaptive regenerative braking energy recovery strategy based on multi-source information fusion. From the aforementioned research, it is evident that significant progress has been made in the study of regenerative braking for CAVs at traffic light intersections. However, most studies have focused solely on optimizing the efficiency of regenerative braking energy recovery, neglecting other factors, such as vehicle traffic efficiency and driving smoothness at traffic light intersections. Therefore, the author suggests analyzing and studying the braking deceleration process of CAVs at traffic light intersections as a multi-objective optimization model to enhance the multifaceted performance of CAVs in this context.

Reinforcement Learning (RL), an emerging machine learning method, is now widely applied to vehicle motion control issues. Liu et al. [29] proposed a predictive energy management strategy for HEVs based on velocity prediction and RL. Li et al. [30] introduced a hierarchical control architecture composed of a speed planner and an energy management system based on RL to address the challenges of real-time eco-driving control in HEVs. Zhang et al. [31] developed a speed planning algorithm oriented towards energy recovery optimization for autonomous driving vehicles or driver-assist systems at traffic light intersections. Liu et al. [32] proposed an adaptive speed planning method for CAVs based on deep reinforcement learning (DRL) trained with multiple traffic signals, focusing on improving the fuel economy and comfort of CAVs. These studies collectively demonstrate that RL methods are effective in various aspects of vehicle motion control.

In previous research conducted by our research group, Li et al. [33] applied the genetic algorithm (GA) to solve the optimal braking trajectory problem for hybrid energy electric vehicles' (HEEVs) RBS in a V2X communication environment. Building on this, the current study aims to design a multi-objective optimization controller for CAVs passing through traffic light intersections in a V2X setting. This design aims to enhance three key performance aspects: energy recovery from braking, traffic efficiency, and driving smoothness. Additionally, the RL method will be employed to investigate this issue further, in the hope of achieving further improvements in the aforementioned performance.

The following sections of this paper are arranged as follows: In Section II, the single-vehicle traffic problem in the V2X signal area is described, and some typical traffic conditions at the traffic light intersection are presented. In Section III, a rule-based traffic light crossing controller (RB-TLCC) for CAVs is proposed, and the simulation results are compared with real human driver data. In Section IV, a multi-objective optimization model (MO-OM) based on the PPO algorithm is proposed, and the results are compared with those of the RB-TLCC method and dynamic programming

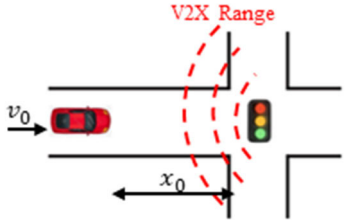

FIGURE 1. Traffic light V2X signal area research scene diagram.

algorithm (DP). Finally, the conclusions of this study are summarized in Section V.

II. PROBLEM DESCRIPTION

Currently, traffic control in the context of CAVs is generally categorized into two types: mixed traffic flow environments and pure CAVs environments. For mixed traffic flow environments that include manually driven vehicles, signal control is often necessary to manage vehicular passages at intersections. By contrast, research on pure CAVs environments typically does not employ signal control. Instead, it relies on the interaction between CAVs to facilitate uninterrupted passage through signalized intersections, which are known as autonomous intersections. However, for the following reasons, the author argues that the use of traffic signal control is also crucial in pure CAVs environments:

1) In complex traffic scenarios, such as high-density areas or four-way intersections, collaboration among pure CAVs might not be sufficient to ensure optimal traffic flow. In such situations, signal control can serve as a supplementary mechanism to coordinate vehicle behavior and reduce congestion and delays.
2) While advancements in autonomous vehicle technology have significantly improved traffic safety, in certain scenarios, such as extreme weather or emergencies, signal control can act as an additional safety layer to manage traffic flow and reduce the risk of accidents.
3) Considering the current state of urban road infrastructure, autonomous intersections are relatively rare, and intersection traffic control relies heavily on traffic lights. With the gradual development of autonomous vehicle technology, our research also considers the transition from existing traffic systems to a purely autonomous vehicle environment. During this transition, signal control can act as a smooth transitional strategy to help gradually adapt to a non-signalized CAVs environment.

Based on these considerations, this study investigated the passage scenarios of CAVs at traffic light intersections. As depicted in Figure 1, the CAV enters the traffic light V2X communication zone at an initial speed of v_0. At this juncture, the distance from the CAV to the stop line of the traffic light intersection was x_0. By utilizing the V2X signal transmission, the CAV acquires the current traffic light phase information

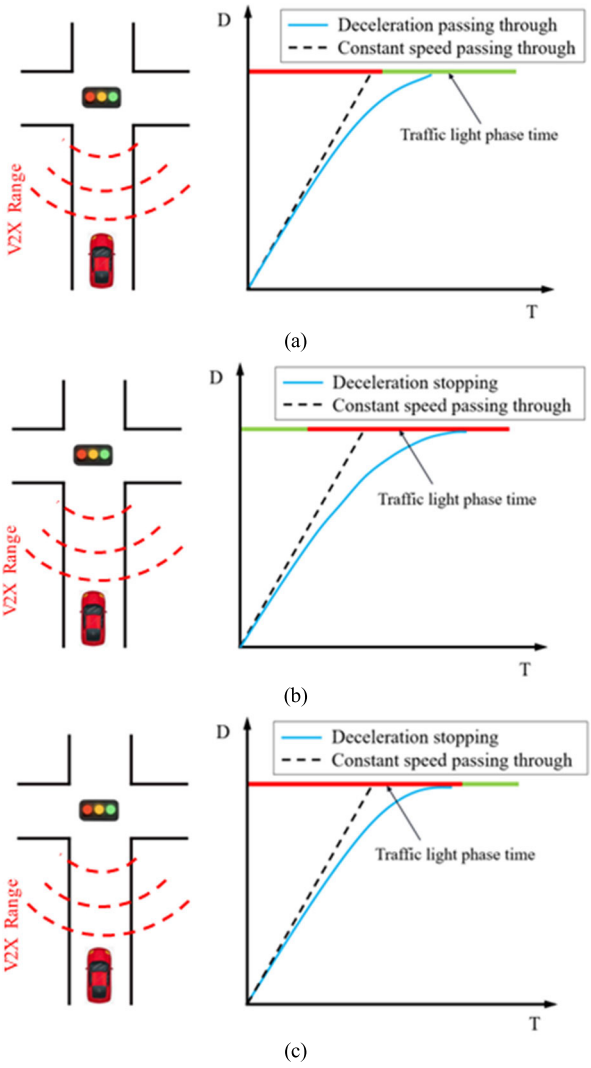

FIGURE 2. (a) Traffic light V2X signal area deceleration passage scenario; (b) (c) Traffic light V2X signal area decelerating and stopping scenario.

and subsequently strategizes its traveling speed based on the procured phase details.

In accordance with real-world driving conditions and Chinese traffic regulations, vehicles must pass the stop line at traffic light intersections during the green light phase. During the yellow and red light phases, vehicles are prohibited from crossing the stop line. Given the multitude of external interferences in the traffic environment, to make our research more targeted, this study primarily focuses on the optimal speed planning for CAVs passing through single signal intersections. Based on practical considerations, the following constraints and assumptions are made:

1) The influence of other vehicles and pedestrians on the subject vehicle is not considered.
2) The primary focus is on vehicles moving straight ahead, without considering lane-changing scenarios.

3) The yellow light phase is simplified to be treated as part of the red light phase, both of which indicate that vehicles are forbidden from crossing the stop line.

Given the changes in the traffic light phases and the vehicle's distance to the stop line, the speed variations of a vehicle from its entry to exit within the V2X signal zone can be segmented into four scenarios: uniform speed passage, accelerating passage, decelerating passage, and decelerating to a halt. In alignment with the focal point of this study, a schematic representation was employed to analyze the vehicle's behaviors during the decelerating passage and decelerating to a stop within the V2X signal zone. In Figure 2, the vertical axis represents the distance covered by the vehicle, and the horizontal axis represents the travel time of the vehicle. The red and green line segments correspond to the durations of the red and green light phases, respectively, during the travel process.

A. DECELERATION PASSAGE SCENARIO

In this scenario, as the CAV enters the V2X signal zone, the current traffic light phase is red, with a relatively short remaining duration. The vehicle cannot maintain its current speed for a uniform passage and must decelerate in advance. This ensures that the vehicle can pass the stop line once the green light phase commences, avoiding the need to halt and wait in front of it. The schematic representation of the decelerating passage scenario is shown in Figure 2 (a).

B. DECELERATING AND STOPPING SCENARIO

Case 1: As shown in Figure 2 (b), the current traffic light phase is green when the CAV enters the V2X signal zone. However, the remaining duration of the green light is insufficient for the vehicle to cross the stop line before the end of this phase. Consequently, the vehicle must decelerate and halt before the stop line, awaiting the next green phase.

Case 2: As illustrated in Figure 2 (c), the current traffic light phase is red when the CAV enters the V2X signal zone. Given the prolonged remaining duration of the red light, the vehicle cannot maintain its current speed and cross the stop line. Consequently, the vehicle must decelerate and stop before the stop line while waiting for the onset of the green phase.

III. RULE-BASED TRAFFIC LIGHT CROSSING CONTROLLER

Upon a comprehensive analysis of vehicular traffic within the V2X signal zones at traffic lights revealed that vehicles tend to undergo frequent braking actions. When CAVs enter the V2X signal zone at traffic lights, both the phase of the traffic light and remaining time are random. Within the constraints imposed by traffic lights and vehicular motion, the planning of the deceleration curve significantly influences the amount of braking energy that can be recuperated, efficiency of traffic, and smoothness of the ride.

TABLE 1. Vehicle dynamic parameters.

Component	Unit	Quantity
Gear ratio (i)	/	4.1
Transmission efficiency (η_T)	/	0.9
Wheel radius (R)	m	0.287
Air drag coefficient (C_D)	/	0.28
Windward area (A)	m^2	1.60
Mass (m)	kg	1260

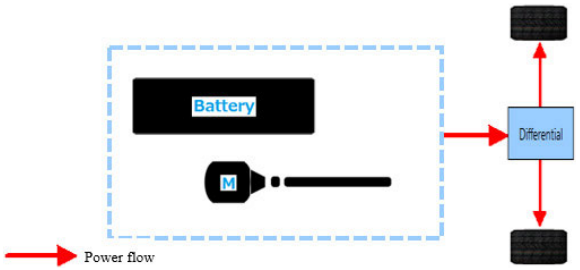

FIGURE 3. Schematic of the EV's power system.

To address the aforementioned issues, this section presents the design of a multi-objective optimization controller for CAVs passing through traffic light intersections.

The optimization objectives are as follows:

1) Maximize the amount of energy recuperation;
2) Enhance the traffic efficiency;
3) Optimize the driving smoothness based on the previous two objectives.

A. CARSIM VEHICLE MODEL

The primary driving resistances of an automobile consist of rolling resistance, aerodynamic drag, grade resistance, and acceleration resistance. The dynamic equation for an electric vehicle (EV) is shown in Equation (1):

$$\frac{T_m i \eta_T}{R} = Gf\cos\alpha + G\sin\alpha + \frac{C_D A}{21.15}V^2 + \delta ma \quad (1)$$

where T_m is the motor drive torque, which is positive during propulsion and negative during braking. i denotes the overall transmission ratio. η_T denotes the mechanical efficiency of the transmission system. R denotes the radius of the wheel. $G = mg$ represents the weight of the EV, where m is the vehicle mass and g is the gravitational acceleration. f denotes the rolling resistance coefficient. α is the road gradient angle. C_D denotes the aerodynamic drag coefficient. A is the frontal area facing wind. V represents the vehicle speed. δ is a coefficient used to convert the rotational inertia of the vehicle. a represents the vehicle acceleration.

According to the above formula, the relevant vehicle dynamic parameters are determined by comparison with a B-class car currently available in the Chinese market, as shown in Table 1.

The subject of this study is a pure electric vehicle with a power system comprising a battery and motor. The power

output from the motor is transmitted to the wheels via a differential. Figure 3 shows a schematic of the EV's power system.

In Equation (2), the motor input torque T behaves as a first-order lag element of the required input torque T_{cmd}. This is constrained by the lookup function f_T, which determines the maximum torque of the motor. The maximum torque of the motor is expressed as a function of motor speed ω.

$$\dot{T} = \frac{sign\left[min\left\{f_T\left(\omega\right), \left|T_{cmd}\right|\right\}, T_{cmd}\right] - T}{t_C} \tag{2}$$

wherein, f_T is the lookup function for the motor's maximum torque, ω is the motor speed, T_{cmd} is the demanded motor input torque, T is the actual motor input torque, and t_C is the time constant.

The motor efficiency E is represented as a lookup function f_E of the motor output torque T and motor speed ω, as shown in Equation (3).

$$E = f_E\left(T, \omega\right) \tag{3}$$

The motor power P is calculated based on the motor output torque T, motor speed ω, and motor efficiency E, and is expressed as

$$\begin{cases} P = T\dfrac{\omega}{E}\left(P \geq 0, discharge\right) \\ P = T\omega E\left(P < 0, charge\right) \end{cases} \tag{4}$$

The required battery power P_{bttry} is represented as:

$$P_{bttry} = P_{motor} \tag{5}$$

wherein, P_{motor} is the demanded power of the motor.

In Equation (6), the open-circuit voltage V_{OC} of the battery is represented as a lookup function f_{VOC} of the battery temperature *temp* and the state of charge SoC.

$$V_{OC} = N_{bttry} \cdot f_{VOC}\left(temp, SoC\right) \tag{6}$$

wherein, N_{bttry} is the number of battery packs.

The battery current A_{bttry} is represented as:

$$A_{bttry} = \frac{V_{OC} - \sqrt{V_{OC}^2 - 4P_{limit} \cdot R}}{R} \tag{7}$$

wherein, P_{limit} is the power limit for the battery discharge and R is the internal resistance of the battery.

The battery output voltage V_{bttry} is represented as:

$$V_{bttry} = V_{OC} - A_{bttry} \cdot R \tag{8}$$

The battery discharge capacity C_{btr_Ds} is determined through the integration of the current, which is expressed as

$$C_{btr\_Ds} = \frac{\int A_{bttry} dt}{3600} \tag{9}$$

The state of charge SoC of the battery is represented as:

$$SoC = \frac{C_{max} - C_{btr\_Ds}}{C_{max}} \tag{10}$$

TABLE 2. Motor and battery model parameters.

Component	Unit	Quantity
Maximum output torque	Nm	305
Maximum speed	RPM	6300
Motor peak power	kW	165
Number of battery packs	/	40
Initial SoC	/	0.95
Minimum battery voltage	V	6
Maximum battery voltage	V	9
Maximum battery capacity	Ah	60

wherein, C_{max} is the maximum capacity of the battery, represented as a lookup function f_{bttry_cap} of the battery temperature:

$$C_{max} = f_{bttry\_cap}\left(temp\right) \tag{11}$$

Table 2 lists the model parameters of the motor and battery obtained after the parameter matching of the B-class vehicle.

B. PRESCAN ROAD ENVIRONMENT MODEL

To make the simulation scenario more in line with the real road environment, Liuhe Road in Liuxia Street, Xihu District, Hangzhou, Zhejiang Province was chosen as the road model for the simulation. Liuhe Road is an urban road that runs from the southwest to the northeast. The road is located on the Xiaohe Mountain Higher Education Park in Hangzhou, where there are several universities and multiple residential communities. The road has a total length of 5.5 km, with three lanes in each direction, and a maximum speed limit of 50 km/h. Information on Liuhe Road from OpenStreetMap is shown in Figure 4.

There are 12 traffic lights on the driving route, denoted as TL1-TL12. The longest and shortest distances between two consecutive traffic lights are 697 and 220 m, respectively, with an average distance of 458 m. The 12 traffic lights divide the road into 13 segments, denoted as Seg1–Seg13. The length of each segment and the corresponding traffic light phase information are listed in Table 3.

To study the vehicle traffic control strategy within the V2X signal zone at traffic signal intersections, V2X sensor models are added to both the CAV and traffic light. The signal transmission range of the V2X sensor is 200 m. When the CAV enters the V2X signal zone, it receives real-time phase information from the current traffic light. The road environment model constructed using PreScan is illustrated in Figure 5.

C. CONTROLLER MODEL

At the start of the simulation, the 12 traffic lights on the road simultaneously begin their red-green cycle from the green phase, following the phase information defined in Table 3. The CAV accelerates from the starting point of the road at a rate of 1 m/s^2 until it reaches 50 km/h and then maintains this speed. Every time the CAV enters the V2X signal zone

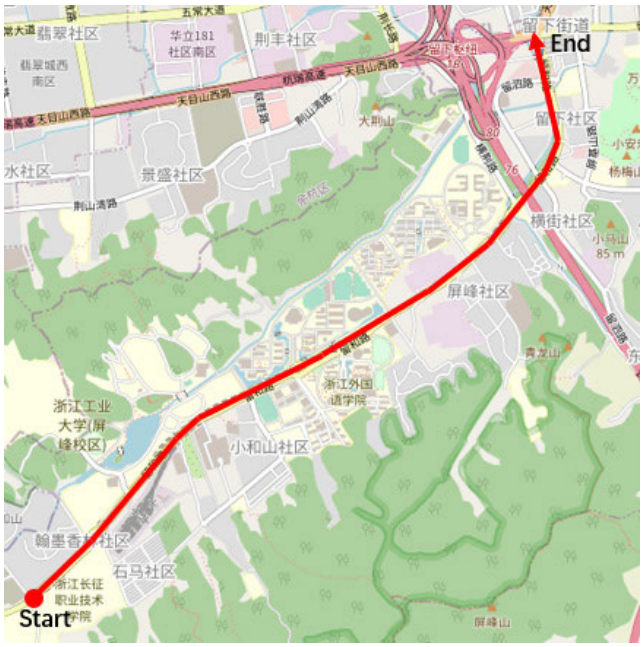

FIGURE 4. Information of Liuhe Road under OpenStreetMap.

TABLE 3. Road length and corresponding traffic light phase information.

Segment number	Road length [m]	Traffic light number	Green light time [s]	Red light time [s]	Time for one cycle [s]
Seg1	538	-	-	-	-
Seg2	417	TL1	50	60	110
Seg3	697	TL2	54	57	111
Seg4	363	TL3	70	40	110
Seg5	323	TL4	43	67	110
Seg6	400	TL5	65	45	110
Seg7	377	TL6	34	22	56
Seg8	660	TL7	56	53	109
Seg9	220	TL8	57	51	108
Seg10	497	TL9	66	42	108
Seg11	465	TL10	42	68	110
Seg12	245	TL11	47	63	110
Seg13	300	TL12	70	38	108

of a traffic light, the RB-TLCC receives the current traffic light phase information transmitted from the V2X traffic light transmitter, as shown in Equation (12):

$$t_{totalphase\_i} = t_{greenphase\_i} + t_{redphase\_i} \tag{12}$$

wherein, $t_{totalphase_i}$is the total phase duration of the i^{th} traffic light, $t_{greenphase\_i}$ is the green light phase duration of the i^{th} traffic light, and $t_{redphase\_i}$ is the red light phase duration of the i^{th} traffic light.

Subsequently, the RB-TLCC calculates the minimum green light duration t_{min_i} during which the CAV can pass the current traffic light without decelerating, as shown in Equation (13):

$$t_{\min\_i} = \frac{d_{range}}{v_{0\_i}} \tag{13}$$

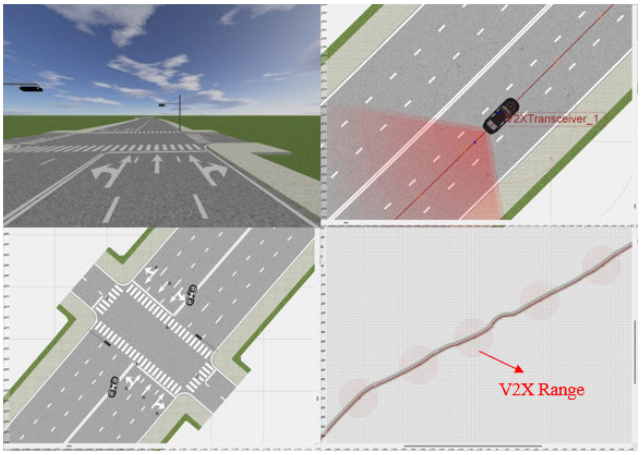

FIGURE 5. The road environment model constructed in PreScan.

wherein, $d_{range} = 200$ m is the set effective transmission range of the V2X sensor and v_{0_i} is the initial speed of the CAV when it reaches the V2X signal transmission range of the i^{th} traffic light.

If the CAV continues to travel at speed v_{0_i}, then when t_{0_i} satisfies Equation (14):

$$t_{0\_i} \in [0\ ,\ t_{greenphase\_i} - t_{min\_i}) \times \cup \left(t_{totalphase\_i} - t_{min\_i}\ ,\ t_{totalphase\_i}\right] \tag{14}$$

The CAV can pass through the traffic signal intersection before the current green light ends. Wherein, t_{0_i} is the current time-count value when the CAV reaches the V2X signal transmission range of the i^{th} traffic light. In the design process of RB-TLCC, we allocate a corresponding cycle count time based on the phase of each traffic light. Each traffic light transmits its current time-count value (i.e., t_{0_i}) to the CAV via V2X signals, thereby achieving the transition and cycle counting of traffic lights.

Otherwise, the CAV cannot pass through the traffic signal intersection before the current green light ends, and must decelerate by braking. The time range $t_{dec\_range\_i}$ required for deceleration satisfies Equation (15).

$$\begin{cases} t_{min\_dec\_range\_i} = t_{greenphase\_i} - t_{min\_i} \\ t_{max\_dec\_range\_i} = t_{totalphase\_i} - t_{min\_i} \\ t_{dec\_range\_i} = \left[t_{min\_dec\_range\_i}, t_{max\_dec\_range\_i}\right] \end{cases} \tag{15}$$

wherein, $t_{min\_dec\_range\_i}$ is the minimum value of the required deceleration time range and $t_{max\_dec\_range\_i}$ is the maximum value of the required deceleration time range.

When the CAV meets the deceleration criteria, the desired speed v_{dec} in the RB-TLCC is determined using Equation (16) - (20):

$$d_{range} = \frac{v + v_{0\_i}}{2} t_{dec\_i} \tag{16}$$

$$t_{dec\_i} = \frac{2d_{range}}{v_{0\_i}} \tag{17}$$

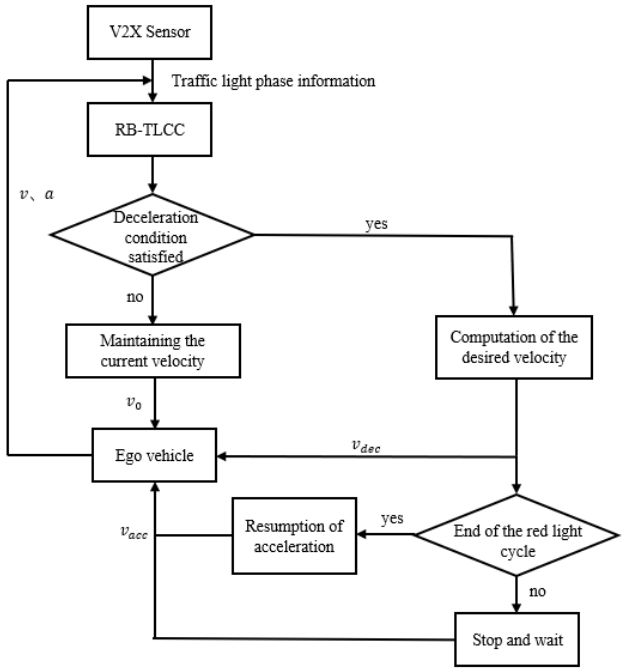

FIGURE 6. Flowchart of the RB-TLCC control logic.

$$a_{dec\_i} = \frac{\Delta v}{\Delta t} = -\frac{v_{0\_i}^2}{2d_{range}} \tag{18}$$

$$v_{dec} = v_{0\_i} + \int_0^{t_{dec\_i}} a_{dec\_i} dt \tag{19}$$

$$v_{dec} = v_{0\_i} + \int_0^{\frac{2d_{range}}{v_{0\_i}}} \left(-\frac{v_{0\_i}^2}{2d_{range}}\right) dt \tag{20}$$

wherein, $v = 0$ is the final speed of the CAV, t_{dec_i} is the time required for the CAV to decelerate to zero within the deceleration range of the i^{th} traffic light, and a_{dec_i} is the desired deceleration of the CAV.

Following the aforementioned deceleration control logic of RB-TLCC, when the CAV reaches the stop line of the traffic light, its speed is precisely reduced to zero. At this point, the CAV waits for the next green light phase to appear, and then accelerates at 1 m/s^2 up to 50 km/h and maintains this speed. If the next green light phase appears while the CAV continues to decelerate, that is, before the CAV reaches the stop line of the traffic light, the vehicle will begin accelerating from its current speed of 1 m/s^2 up to 50 km/h and then maintain this speed.

The flowchart of the RB-TLCC control logic is shown in Figure 6.

The RB-TLCC model designed in MATLAB/Simulink, as depicted in Figure 7, is presented herein.

D. DRIVER-IN-THE-LOOP SIMULATION

Additionally, a driver-in-the-loop simulation (DILS) environment was established using the Logitech G27 driving simulation kit to allow human drivers to operate on the virtual road constructed in PreScan. The DILS equipment used in the laboratory experiment is illustrated in Figure 8.

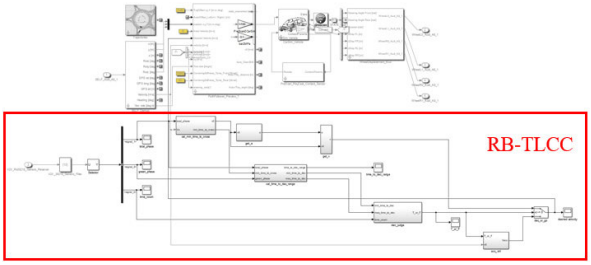

FIGURE 7. The RB-TLCC model designed in MATLAB/Simulink.

FIGURE 8. Logitech G27 driving simulation kit established in the laboratory.

Ten human drivers operated the Logitech G27 driving simulation kit in a simulated environment using the constructed road model. A complete journey is one in which the given requirements are met from start to finish. Each participant completed one full journey, resulting in the collection of ten sets of human driving data.

The driving simulation requirements are as follows:

1) The maximum driving speed must not exceed 50 km/h;
2) Drivers must not cross solid lines while driving;
3) Drivers must adhere to traffic light regulations.

E. RESULTS AND DISCUSSION

Figure 9 shows the comparison of the SoC for 10 human drivers and the RB-TLCC after completing a single journey. When the initial SoC for both human drivers and RB-TLCC is set to 0.95, the maximum and minimum SoC values for human drivers after completing a journey are 0.921 and 0.908, corresponding to SoC consumptions of 0.029 and 0.042, respectively. The average SoC for human drivers after a journey is 0.916, with a corresponding SoC consumption of 0.034 (the results are shown in Table 4). In contrast, RB-TLCC achieves an SoC value of 0.924 after a journey, with a corresponding SoC consumption of 0.026. The RB-TLCC demonstrates the lowest energy consumption among all curves in the graph. Compared with human drivers,

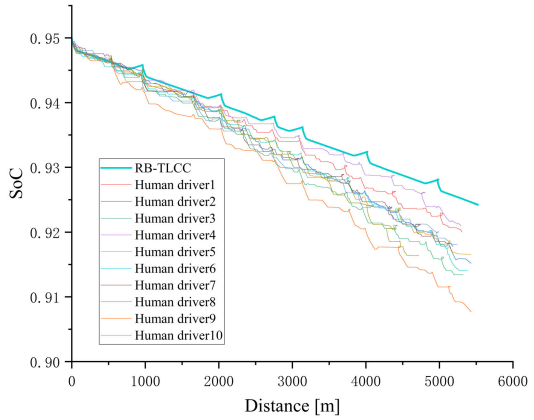

FIGURE 9. SoC comparison between human drivers and RB-TLCC after completing a full journey.

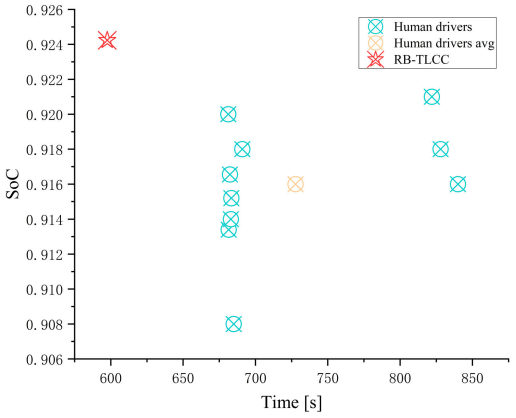

FIGURE 11. Time-SoC scatter plot after completing a full journey between human drivers and RB-TLCC.

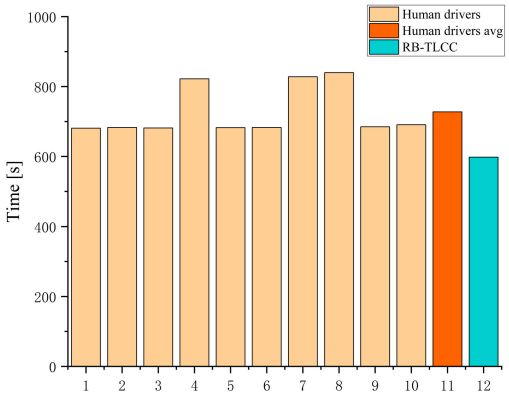

FIGURE 10. Travel time comparison between human drivers and RB-TLCC after completing a full journey.

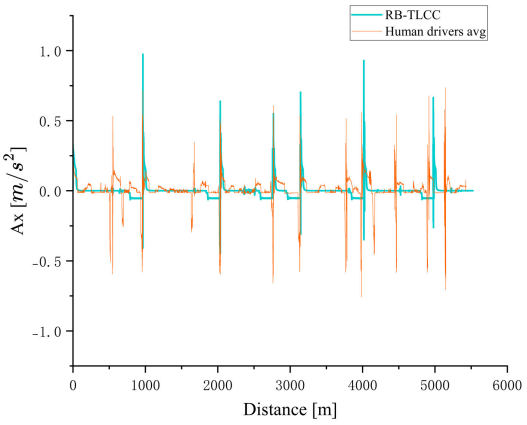

FIGURE 12. Acceleration comparison between human drivers average and RB-TLCC after completing a full journey.

RB-TLCC-controlled CAVs can save approximately 23.5% in energy consumption after completing a journey, thereby enhancing the energy efficiency of CAVs. Additionally, when considering the entire journey, the SoC curves of some human drivers decrease more slowly than those of RB-TLCC at the beginning of the trip, which could be attributed to these drivers' smooth driving habits. However, as the journey progressed, the gap in SoC values between them and the RB-TLCC widened. This indicates that the RB-TLCC is capable of steadily controlling power consumption, with its energy-saving effects becoming more pronounced over longer distances.

Figure 10 compares the time required to complete a single journey between the 10 human drivers and the RB-TLCC. From the graph, it is evident that most human drivers took approximately 700 s to complete a journey. The average time taken by the 10 human drivers for a single journey is 728 s. In contrast, RB-TLCC completes a journey in 598 s, representing a 17.9% reduction in travel time compared with the average time for human drivers. This indicates that the RB-TLCC-controlled CAVs exhibited higher road traffic efficiency.

TABLE 4. Objective optimization parameters comparison between human drivers and RB-TLCC after a full journey.

	Human drivers avg	RB-TLCC	Δ	Δ%
SoC consumption	0.034	0.026	0.008	23.5%
Travel time [s]	728	598	130	17.9%
σ_a [m/s^2]	0.140	0.069	0.071	50.7%

Combining the contents of Figure 9 and 10, Figure 11 is a scatter plot showing the relationship between the travel time and SoC of human drivers and the RB-TLCC at the end of the complete journey. In the figure, the farther the x-axis, the shorter the travel time, and the higher the y-axis, the less power consumed. It can be intuitively observed from the figure that RB-TLCC has the smallest X values and the largest Y values among all points.

Figure 12 compares the acceleration over the entire journey of human drivers average with that of the RB-TLCC. The figure reveals that, compared to human drivers, the overall acceleration curve of the RB-TLCC has fewer fluctuations. During steady driving, the acceleration of the RB-TLCC

remains almost unchanged, resulting in a smoother curve. Particularly in the deceleration phase, RB-TLCC's deceleration values are smaller than those of human drivers, staying within -0.5 m/s^2, which indicates a more stable deceleration process. In the acceleration phase, the RB-TLCC aims to reach the highest driving speed (50 km/h) as quickly as possible to save transit time, resulting in higher acceleration values in some instances compared to human drivers. Nevertheless, the maximum acceleration did not exceed 1 m/s^2.

Standard deviation (σ) is used to measure the extent to which a set of data deviates from its average. A smaller σ value indicates less deviation from the average, signifying more stable data. In Table 4, the average standard deviation of acceleration (σ_a) for human drivers after completing a journey is 0.140 m/s^2, while RB-TLCC achieves a σ_a of 0.069 m/s^2. Compared to human drivers, the use of RB-TLCC results in a 50.7% reduction in σ_a after completing a journey. The data results also show that the RB-TLCC-controlled CAV has less acceleration fluctuations and better ride comfort.

In summary, the proposed RB-TLCC outperformed human drivers in terms of SoC consumption, road traffic efficiency, and driving smoothness. Therefore, the application of RB-TLCC to the problem of traffic light control at intersections for CAVs is highly effective.

IV. MULTI-OBJECTIVE OPTIMIZATION MODEL BASED ON PPO ALGORITHM

In Section III, it was observed that RB-TLCC exhibited superior control performance compared to human drivers. However, owing to RB-TLCC's reliance on rule-based deceleration control strategies, constrained by the mutual dependencies between displacement and velocity, it inevitably encounters situations where it must come to a stop at traffic intersections, lacking adaptive adjustment capabilities. If CAVs completely stop at traffic light intersections, it will, to some extent, affect the efficiency of traffic flow. Second, frequent starting and stopping can affect passenger comfort, detracting from the smooth operation of CAVs. Finally, excessive idling during stops can lead to unnecessary consumption of electric energy, which is not conducive to the efficient use of energy. To address this limitation, this section proposes a multi-objective optimization model (MO-OM) for the intersection crossing of CAVs based on the Proximal Policy Optimization algorithm (PPO). The PPO algorithm, a reinforcement learning approach, is employed to model the expected velocity profile of CAVs to further enhance braking energy recovery, traffic efficiency, and driving smoothness.

A. PROXIMAL POLICY OPTIMIZATION ALGORITHM

Reinforcement Learning (RL), a widely adopted machine learning algorithm, offers advantages such as data-driven decision-making, model independence, self-learning, and direct policy determination [34]. The problem under investigation in this study pertains to multi-objective optimization, applications with varying uncertainties, and decision-making in online planning scenarios, making it well-suited for RL methodologies.

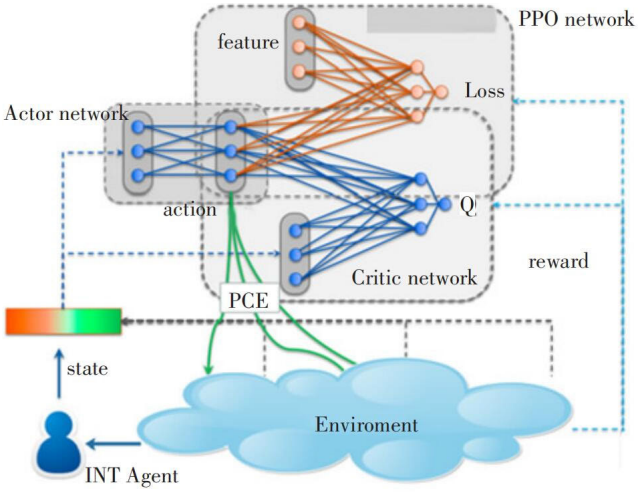

FIGURE 13. PPO algorithm neural network structure diagram.

The PPO algorithm is a policy-based RL algorithm that uses two neural networks. By feeding the current state of the agent into these networks, it generates the corresponding actions and rewards. These actions are then used to update the agent's state. The neural network weight parameters are updated through gradient ascent, optimizing the action choices to maximize the cumulative reward as defined by the objective function encompassing rewards and actions [35].

The neural network structure for the PPO algorithm is depicted in Figure 13.

The pseudocode for the PPO algorithm is as follows:

Algorithm 1 PPO-Clip

1: Input: initial policy parameters θ_0, initial value function parameters ϕ_0
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
4: Compute rewards-to-go $\hat{R}_i$.
5: Compute advantage estimates $\hat{A}_i$ (using any method of advantage estimation) based on the current value function V_{ϕ_K}.
6: Update the policy by maximizing the PPO-Clip objective:

$$\theta_{k+1} = arg\max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{i=0}^{T} min\left(\frac{\pi_\theta(a_i|s_i)}{\pi_{\theta_k}((a_i|s_i))} A^{\pi_{\theta_k}} \times (s_i, a_i), g\left(\epsilon, A^{\pi_{\theta_k}}(s_i, a_i)\right)\right)$$

typically via stochastic gradient ascent with Adam.
7: Fit value function by regression on mean-squared error:

$$\phi_{k+1} = arg\min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{i=0}^{T} \left(V_\phi(s_i) - \hat{R}_i\right)^2,$$

typically via some gradient descent algorithm.
8: **end for**

TABLE 5. PPO algorithm training parameter settings.

Parameters	Value
Timestep	10000
Learning rate	0.0003
Discount factor (γ)	0.99
Gae_lambda (λ)	0.95
Entropy coefficient	0
Value function coefficient	0.5
Max_grad_norm	0.5
Batch size	64
N_epochs	10

In the implementation of the PPO algorithm used in this study, several key sensitive parameters significantly influenced the performance of the algorithm. The discount factor (γ) serves as a crucial parameter for determining the present value of future rewards, with its magnitude dictating the algorithm's balance between immediate and long-term rewards. A higher γ value emphasizes long-term returns, whereas a lower value shifts focus towards short-term gains. The learning rate is a vital determinant of the magnitude of policy updates, where a higher learning rate can accelerate the learning process but may lead to training instability. Conversely, a lower rate, though more stable, slows down the learning pace. The lambda (λ) parameter in Generalized Advantage Estimation (GAE) plays a pivotal role in balancing the bias and variance in the estimation of the advantage function. Variations in λ can significantly alter the performance of the algorithm, with higher λ values potentially leading to increased variance and lower values leading to higher bias. The entropy coefficient is employed to incentivize exploratory actions by providing additional rewards for choosing actions with higher uncertainty. A higher entropy coefficient aids in enhancing the exploratory capability of the algorithm, preventing premature convergence to local optima but may introduce instability in the learning process.

After debugging the main sensitive parameters, the final PPO algorithm related training parameter settings are listed in Table 5.

B. MULTI-OBJECTIVE OPTIMIZATION MODEL

In reinforcement learning, the interaction between the agent and the environment is modeled using a Discrete Markov Decision Process (MDP). Based on the MDP and research scenario described in Section II, the constructed model for the PPO algorithm, vehicle model, and definitions of state (S), action (A), and reward (R) are as follows:

At the beginning of each training round, the current traffic signal phase is randomly selected from the 12 available signal phases listed in Table 3. The initial time, t_{cur_0}, is randomly generated in the range of 0 to 111 s, where 111 s represents the longest cycle time, as shown in Table 3.

The current traffic signal time count, t_{rg_i}, is defined as:

$$t_{rg\_i} = \begin{cases} t_{cur\_0} mod t_{cycle\_i} & \left(t_{cur\_0} > t_{cycle\_i}\right) \\ t_{cur\_0} & \left(t_{cur\_0} \le t_{cycle\_i}\right) \end{cases} \tag{21}$$

where t_{cur_0} is the randomly generated initial time and t_{cycle_i} is the cycle time of the current traffic signal.

The remaining green light time, $t_{g\_remain\_i}$, and red light time, $t_{r\_remain\_i}$, are defined as

$$\begin{cases} t_{g\_remain\_i} = t_{g\_phase\_i} - t_{rg\_i} \\ t_{r\_remain\_i} = 0 \end{cases} \quad \left(t_{rg\_i} < t_{g\_phase\_i}\right) \tag{22}$$

$$\begin{cases} t_{g\_remain\_i} = 0 \\ t_{r\_remain\_i} = t_{cycle\_i} - t_{rg\_i} \end{cases} \quad \left(t_{rg\_i} \ge t_{g\_phase\_i}\right) \tag{23}$$

where $t_{g\_phase\_i}$ represents the green phase time of the current traffic signal.

At the start of each training round, the vehicle enters the V2X signal zone at a speed of 50 km/h within a 200-meter range of the traffic signal. The initial state of the vehicle is defined as

$$\begin{cases} x_0 = 200m \\ v_0 = 13.9\ m/s \\ a_0 = 0m/s^2 \end{cases} \tag{24}$$

where x_0 is the initial vehicle position, v_0 is the initial velocity, and a_0 is the initial acceleration.

The vehicle's state space is defined as:

$$S_t = [x_t v_t] \tag{25}$$

where x_t represents the distance traveled by the vehicle at the current time and v_t is the current velocity of the vehicle.

The vehicle's action space is defined as:

$$A_t = a_t \tag{26}$$

where a_t represents the current acceleration of the vehicle. The action space is discretized into 40 dimensions, ranging from -2 m/s^2 to 2 m/s^2, with intervals of 0.1 m/s^2, resulting in a total of 40 discrete actions.

The model operates in discrete 1-second iterations, and based on the vehicle's velocity and iteration time, it calculates the distance the vehicle moves in each iteration. The vehicle's travel distance, traffic signal status, and duration are updated according to the rules of the environment. The velocity for each iteration is determined using Equation (27):

$$v_t = v_{t-1} + a_t \tag{27}$$

The travel distance for each iteration is calculated by Equation (28):

$$x_t = \frac{v_t + v_{t-1}}{2} \tag{28}$$

The cumulative travel distance for the current time is obtained by accumulating the travel distances for each iteration, as shown in Equation (29).

$$d_t = \sum_{t=0}^{n} x_t \tag{29}$$

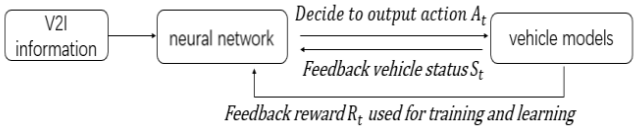

FIGURE 14. Overall block diagram of PPO algorithm.

If the vehicle exits the 200-meter V2X signal zone after the last iteration, it is controlled to stop at the intersection line, and the current training round ends. If the vehicle runs a red light, exceeds the speed limit, or reverses, the current round is immediately terminated and the next training round begins.

The reward function of the model is defined in Equations (30) - (33):

$$R_1 = SOC_t - SOC_{t-1} \tag{30}$$

$$R_2 = \frac{d_t}{200} \tag{31}$$

$$R_3 = \begin{cases} \left| \dfrac{1}{a_t - a_{t-1}} \right| (a_t - a_{t-1} \neq 0) \\ 1\ (a_t - a_{t-1} = 0) \end{cases} \tag{32}$$

$$R = \beta_1 R_1 + \beta_2 R_2 + \beta_3 R_3 \tag{33}$$

where $\beta_1 = 50$, $\beta_2 = 30$, and $\beta_3 = 20$ are the weight coefficients for the three reward values. SOC_t and SOC_{t-1} represent the current and previous state of charge. R_1 represents the reward for braking energy recovery to enhance the energy efficiency. R_2 represents the reward for the distance from the vehicle to the intersection line to improve the traffic efficiency. R_3 represents the reward for ride comfort to enhance the smoothness of vehicle motion. These three rewards correspond to the three previously defined optimization objectives.

The three rewards have different physical meanings and significantly different numerical ranges. To standardize the measurements and improve the convergence speed and training effectiveness of the neural network, each of these three rewards is normalized separately according to Equation (34):

$$x' = \frac{x - min\,(x)}{max\,(x) - min\,(x)} \tag{34}$$

Additionally, d_t, v_t, and a_t must satisfy the following constraints:

$$\begin{cases} 0 \leq d_t \leq 200\ m \\ 0 \leq v_t \leq 13.9\ m/s \\ -2m/s^2 \leq a_t \leq 2\ m/s^2 \end{cases} \tag{35}$$

The overall block diagram for PPO algorithm is depicted in Figure 14.

C. RESULTS AND DISCUSSION

Training was carried out on a Lenovo laptop powered by an AMD Ryzen 7 5800H processor with 8 cores and 16 threads, operating at a base frequency of 3.2GHz. The GPU utilized was the NVIDIA GeForce RTX 3050 with an operational RAM of 16GB.

The performance evaluation of the PPO generally depends on multiple metrics to ensure that the trained model excels in tasks. The subsequent analysis assesses the model's post-training performance using three indicators: average reward, steps per epoch, and loss.

In RL, rewards evaluate an agent's behavior in the environment and drive the agent to learn a superior strategy throughout training. Figure 15 (a) illustrates the average reward per epoch during training. It is evident that post-initiation of training, the reward gradually ascends. By 25 million epochs, the average reward began to converge, stabilizing at approximately 250. This indicates the agent's discovery of a fitting strategy through environmental interactions. Figure 15 (b) shows the average steps per epoch during training. In the initial phase, the average steps were minimal, suggesting that the vehicles were terminated before successfully reaching the intersection owing to constraint violations. As the training advanced to approximately 20 million epochs, the average steps began to converge, consistently hovering around 30. This suggests that the agent learned an appropriate strategy to navigate the vehicle to the stop line. Loss in PPO primarily comprises Policy loss and Value loss, with the total loss being their weighted sum. The objective of the agent is to minimize this cumulative loss. Figure 15 (c) presents this loss throughout the training, highlighting a decline post-initiation. By 15 million epochs, the loss begins to stabilize near zero.

Upon a CAV entries into the traffic light V2X signal zone, owing to the randomness of the current time and traffic light phase, the agent needs to determine the appropriate action strategy to ensure that the vehicle seamlessly crosses the intersection without halting. Referencing the scenarios in Section II, two typical conditions were selected to validate the training outcomes of the model. The PPO model proposed in the current section, along with the RB-TLCC model discussed in the previous section, is compared with dynamic programming algorithm (DP) models [20] in terms of control effectiveness during deceleration at a single traffic light.

Figure 16 presents the time-distance, time-speed, and time-acceleration curves under three different control methods for a scenario where the traffic light transitions from green to red during deceleration. The green and red segments in the figure represent the green and red light phases, respectively, with the green phase lasting 70 s and red phase lasting 38 s.

In this scenario, when CAVs entered the V2X signal zone of the traffic light, only 8 s remained in the green phase, indicating a short remaining green time. CAVs are unable to pass through the intersection at a constant speed (50 km/h) before the end of the current green phase and must decelerate in advance. This scenario corresponds to the deceleration and stopping situation in Case 1 in Section II.

During deceleration, RB-TLCC performs uniform deceleration at approximately -0.5 m/s^2, reaching the traffic light intersection and decelerating to 0 m/s approximately 29 s into the run, while still in the red phase, requiring a stop and wait for the green phase. After a 16-second wait, the red light

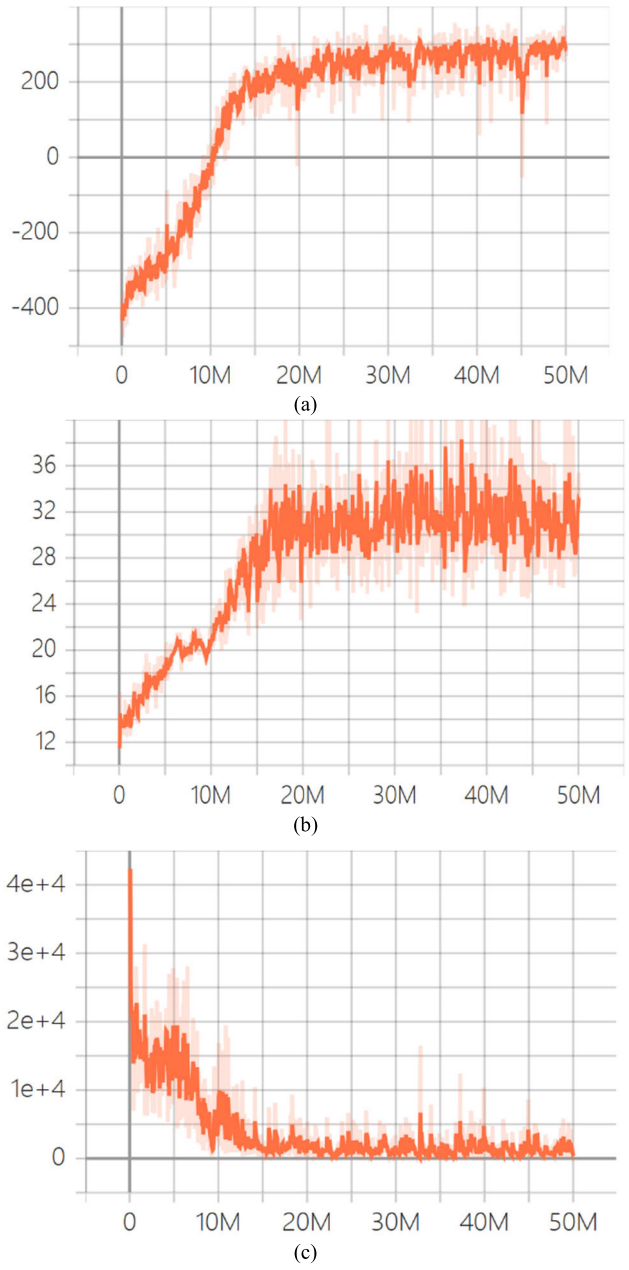

FIGURE 15. (a) The average reward per epoch during training; (b) The average steps per epoch during training; (c) Change in loss during training.

turns green, and the CAV resumes its journey, accelerating at 1 m/s^2. This is not conducive to improving traffic efficiency, and idle stopping also increases the unnecessary power consumption. Moreover, frequent starts and stops are not smooth for the vehicle's driving.

The PPO model proposed in this section initially decelerates the vehicle at approximately -1 m/s^2 for approximately 10 s, then maintains a lower constant speed (3 m/s), and passes through the intersection 3 s after the red light turns green (i.e., at approximately 49 s of running time). There is no stop throughout the journey, which ensures traffic efficiency while extending the time for brake energy recovery and maintaining smooth driving.

Similar to the PPO model, in the initial deceleration phase, the DP algorithm tends to use a larger deceleration rate for braking and then maintains a lower speed to pass through the traffic signal without stopping. However, the initial deceleration rate of the DP algorithm is higher (-1.5 m/s^2), which is less friendly to driving smoothness. Additionally, as shown in the figure, owing to the lower speed of the DP algorithm in the later stages, it fails to pass through the signal light promptly after the red light turns green, leading to suboptimal traffic efficiency.

Figure 17 displays the time-distance, time-speed, and time-acceleration curves under the three control methods for a scenario in which the traffic light transitions from red to green during deceleration. In this scenario, the green phase lasted for 65 s and the red phase lasted for 45 s.

When CAVs enter the V2X signal zone of the traffic light in this scenario, 20 s remain in the red phase, indicating a longer remaining red time. The CAVs cannot maintain their current speed and must decelerate in advance. This scenario corresponds to the deceleration and stopping situation in Case 2, as described in Section II.

During deceleration, the RB-TLCC performs uniform deceleration at approximately -0.5 m/s^2, and after 20 s of uniform deceleration, the red light turns green. The vehicle then starts to accelerate at 1 m/s^2 and leaves the traffic light intersection at a running time of approximately 23 s. This approach improves traffic efficiency; however, early acceleration also leads to additional power consumption.

The PPO model initially decelerates the vehicle at approximately 1 m/s^2 for approximately 8 s, then maintains a lower constant speed (6 m/s), and passes through the intersection 7 s after the red light turns green (i.e., at approximately 27 s of running time). There is no stop throughout the journey, which ensures traffic efficiency while extending the time for brake energy recovery and maintaining smooth driving.

Similar to the PPO model, in the initial deceleration phase, the DP algorithm tends to use a larger deceleration rate for braking and then maintains a lower speed (6 m/s) to pass through the traffic signal without stopping. The overall speed curve of the DP algorithm is very close to that of the PPO model. However, the initial deceleration rate of the DP algorithm is higher (-1.5 m/s^2), which is less friendly to driving smoothness.

In the aforementioned deceleration scenarios, the agent, based on the randomly assigned traffic light phase, chose suitable action strategies for early vehicle deceleration, achieving seamless intersection crossings and thereby enhancing traffic efficiency. Furthermore, the maximum deceleration during this process was only -1 m/s^2, ensuring vehicular smoothness, which is attributable to the multi-objective optimization reward function design of the model.

The PPO-trained model significantly improved energy recuperation, evident through a significant SoC increase during the initial higher deceleration phase. The subsequent

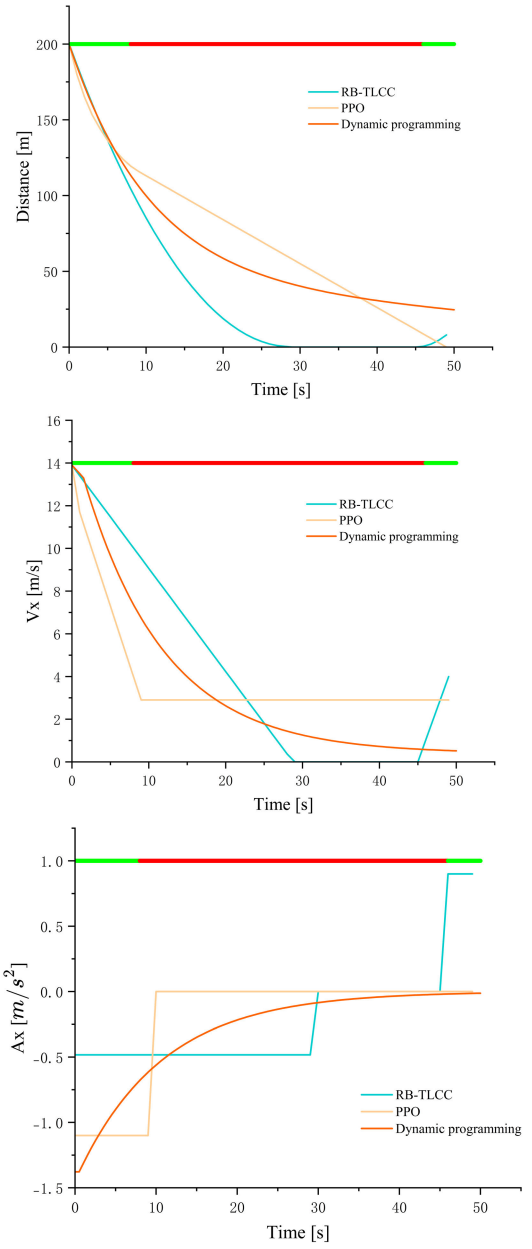

FIGURE 16. The displacement, velocity, acceleration and corresponding traffic light phase of the three control methods under decelerating traffic condition from green to red.

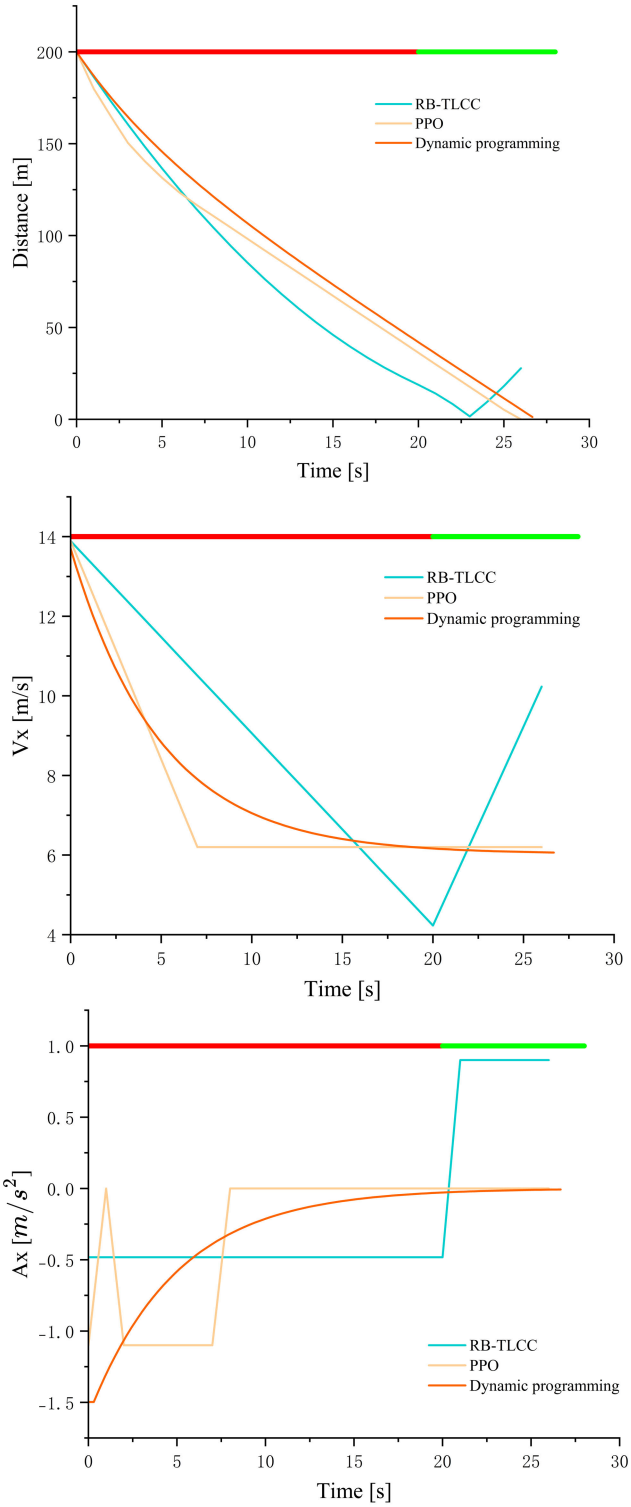

FIGURE 17. The displacement, velocity, acceleration and corresponding traffic light phase of the three control methods under decelerating traffic condition from red to green.

reduced deceleration compensated for the shorter distance traveled owing to the initial deceleration, ensuring intersection passage and extending the brake energy recovery duration.

The speed profiles of the three control methods under the two aforementioned deceleration traffic scenarios were inputted into the co-simulation model proposed in Section III. Under the same traffic light phase conditions, the braking energy recuperation, traffic efficiency, and driving smoothness were compared. The conclusions are presented in Tables 6 and 7, respectively.

In the given two scenarios of single traffic light deceleration, RB-TLCC achieves the shortest transit time. However, the transit time of the PPO model is very close to that of

TABLE 6. Objective optimization parameters comparison between three methods (Green to Red).

	RB-TLCC	PPO	DP
SoC consumption	0.0003	0.0004	0.0003
Travel time [s]	46	49	57
σ_a [m/s^2]	0.569	0.444	0.438

TABLE 7. Objective optimization parameters comparison between three methods (Red to Green).

	RB-TLCC	PPO	DP
SoC consumption	0.0002	0.0003	0.0002
Travel time [s]	23	26	27
σ_a [m/s^2]	0.541	0.491	0.482

RB-TLCC. Specifically, in Scenario 1, the PPO model outperformed DP by 14.0%, and in Scenario 2, it surpassed DP by 3.7%. In terms of energy recuperation, the PPO model exhibited the best performance, surpassing both the RB-TLCC and DP by 33.3% in Scenario 1 and by 50% in Scenario 2. Regarding driving smoothness, DP, as a globally optimized algorithm, achieved the best results. Nevertheless, the standard deviation of acceleration for the PPO model is very close to that of DP, and the PPO model outperformed RB-TLCC by 22.0% in Scenario 1 and by 9.2% in Scenario 2. In summary, the PPO model proposed in this section exhibits the best comprehensive control effect among the three control methods.

The proposed PPO model adeptly selects suitable action strategies for speed profile planning across the various scenarios outlined in Section II. While adhering to traffic regulations, the model improves brake energy recuperation, traffic efficiency, and driving smoothness at intersections, affirming the efficacy of the proposed V2X autonomous driving vehicle multi-objective optimization model based on the PPO algorithm.

V. CONCLUSION

This study addresses the control issues related to vehicular traffic at traffic signal intersections for CAVs with the following findings:

1) A rule-based traffic light control controller (RB-TLCC) for CAVs was designed in MATLAB/Simulink and co-simulated using CarSim and PreScan. Compared with human drivers, the proposed RB-TLCC demonstrated a 23.5% improvement in brake energy recovery, a 17.9% increase in traffic efficiency, and a 50.7% enhancement in ride comfort.
2) Based on Markov Chain theory, a multi-objective optimization model (MO-OM) for traffic light intersection control of CAVs using the PPO algorithm was proposed. In the given two scenarios, compared with RB-TLCC and the Dynamic Programming algorithm (DP), the PPO model proposed in this paper exhibits the best overall control performance among the three control methods. This enhances the recuperation of braking energy, traffic efficiency, and driving smoothness of vehicles at traffic light intersections.
3) For vehicles needing to decelerate at traffic light intersections, the ideal speed profile is as follows: when far from the intersection and at relatively high speeds, vehicles should decelerate aggressively using substantial braking torque, facilitating maximum energy recuperation. Subsequently, vehicles should adopt a gradual and smooth braking approach that extends the regenerative braking duration to ensure efficient energy recovery throughout the braking phase. Simultaneously, ensure smooth running and traffic efficiency of the vehicle during the deceleration process.
4) This study has certain limitations that future research might consider as reference points. Initially, the study focused solely on deceleration scenarios of CAVs at traffic light intersections. Future studies could consider broadening the scope of the scenarios examined, such as the acceleration behavior of CAVs at traffic light intersections. Moreover, this study was limited to simulations with human driver simulators and hardware-in-the-loop experiments without conducting real-vehicle experiments. Subsequent research can further validate the effectiveness of the control strategies proposed herein through real-vehicle testing. Finally, the application of V2X technology in this study was confined to the communication between CAVs and traffic intersections. Future studies could explore communication environments, such as V2V, to achieve higher levels of autonomous driving.

REFERENCES

[1] R. K. Kamalanathsharma, H. A. Rakha, and H. Yang, "Networkwide impacts of vehicle ecospeed control in the vicinity of traffic signalized intersections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2503, no. 1, pp. 91–99, Jan. 2015.
[2] S. Bartoletti, B. M. Masini, V. Martinez, I. Sarris, and A. Bazzi, "Impact of the generation interval on the performance of sidelink C-V2X autonomous mode," *IEEE Access*, vol. 9, pp. 35121–35135, 2021.
[3] Z. Song, K. Song, and T. Zhang, "State-of-the-art and development trends of energy management strategies for intelligent and connected new energy vehicles: A review," SAE, Tongji Univ., Tech. Paper 2019-01-1216, 2019, doi: 10.4271/2019-01-1216.
[4] Z. Wang, G. Wu, and M. J. Barth, "Cooperative eco-driving at signalized intersections in a partially connected and automated vehicle environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 2029–2038, May 2020.
[5] M. Wang, H. Yu, G. Dong, and M. Huang, "Dual-mode adaptive cruise control strategy based on model predictive control and neural network for pure electric vehicles," in *Proc. 5th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Liverpool, U.K., Jul. 2019, pp. 1220–1225.
[6] S. K. Chada, A. Purbai, D. Görges, A. Ebert, and R. Teutsch, "Ecological adaptive cruise control for urban environments using SPaT information," in *Proc. IEEE Vehicle Power Propuls. Conf. (VPPC)*, Gijon, Spain, Nov. 2020, pp. 1–6.
[7] G. De Nunzio, C. Canudas de Wit, P. Moulin, and D. Di Domenico, "Eco-driving in urban traffic networks using traffic signal information," in *Proc. 52nd IEEE Conf. Decis. Control*, Florence, Italy, Dec. 2013, pp. 892–898.

[8] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, "Optimal eco-driving control of connected and autonomous vehicles through signalized intersections," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3759–3773, May 2020.
[9] X. Meng and C. G. Cassandras, "Eco-driving of autonomous vehicles for nonstop crossing of signalized intersections," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 320–331, Jan. 2022.
[10] E. Wang, F. H. Memar, S. Korzelius, A. W. Sadek, and C. Qiao, "A reinforcement learning approach to CAV and intersection control for energy efficiency," in *Proc. 5th Int. Conf. Connected Auto. Driving (MetroCAD)*, Detroit, MI, USA, Apr. 2022, pp. 81–88.
[11] M. Zhu, Y. Wang, Z. Pu, J. Hu, X. Wang, and R. Ke, "Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102662.
[12] Y. Junkai, S. Jinju, L. Xunyi, and Y. Kangjian, "Speed planning and energy optimal control of hybrid electric vehicles based on Internet of Vehicles," *IFAC-PapersOnLine*, vol. 54, no. 10, pp. 169–175, 2021.
[13] B. HomChaudhuri, R. Lin, and P. Pisu, "Hierarchical control strategies for energy management of connected hybrid electric vehicles in urban roads," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 70–86, Jan. 2016.
[14] Z. Bai, P. Hao, W. ShangGuan, B. Cai, and M. J. Barth, "Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15850–15863, Sep. 2022.
[15] R. Zhang and E. J. Yao, "Eco-driving at signalized intersections for electric vehicles," *IET Intell. Transp. Syst.*, vol. 9, no. 5, pp. 488–497, Jun. 2015.
[16] X. Tang, Z. Duan, X. Hu, H. Pu, D. Cao, and X. Lin, "Improving ride comfort and fuel economy of connected hybrid electric vehicles based on traffic signals and real road information," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3101–3112, Apr. 2021.
[17] H. Kui, H. Wang, H. Li, and C. Tian, "An eco-speed optimization for multiple signalized intersections based on V2I," in *Proc. 8th Int. Conf. Green Intell. Transp. Syst.*, Changchun, China, 2019, pp. 309–319.
[18] V. Milanés, J. Villagra, J. Godoy, J. Simo, J. Pérez, and E. Onieva, "An intelligent V2I-based traffic management system," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 49–58, Mar. 2012.
[19] W. Dib, W. Dib, A. Chasse, P. Moulin, A. Sciarretta, and G. Corde, "Optimal energy management for an electric vehicle in eco-driving applications," *Control Eng. Pract.*, vol. 29, pp. 299–307, Aug. 2014.
[20] M. Miyatake, M. Kuriyama, and Y. Takeda, "Theoretical study on eco-driving technique for an electric vehicle considering traffic signals," in *Proc. IEEE 9th Int. Conf. Power Electron. Drive Syst.*, Singapore, Dec. 2011, pp. 733–738.
[21] F. Zhang, J. Xi, and R. Langari, "Real-time energy management strategy based on velocity forecasts using V2V and V2I communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 416–430, Feb. 2017.
[22] N. Wan, A. Vahidi, and A. Luckow, "Optimal speed advisory for connected vehicles in arterial roads and the impact on mixed traffic," *Transp. Res. C, Emerg. Technol.*, vol. 69, pp. 548–563, Aug. 2016.
[23] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, "A decision-making strategy for vehicle autonomous braking in emergency via deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 5876–5888, Jun. 2020.
[24] B. Chen, L. Liao, F. Zou, and Y. Zheng, "Rule-based graded braking for unsignalized intersection collision avoidance via vehicle-to-vehicle communication," in *Proc. 13th Int. Conf. Genetic Evol. Comput.*, Qingdao, China, 2019, pp. 134–142.
[25] W. Yang, B. Wan, and X. Qu, "A forward collision warning system using driving intention recognition of the front vehicle and V2V communication," *IEEE Access*, vol. 8, pp. 11268–11278, 2020.
[26] D. Kim, J. S. Eo, and K. K. Kim, "Parameterized energy-optimal regenerative braking strategy for connected and autonomous electrified vehicles: A real-time dynamic programming approach," *IEEE Access*, vol. 9, pp. 103167–103183, 2021.
[27] D. Kim, J. S. Eo, and K. K. Kim, "Service-oriented real-time energy-optimal regenerative braking strategy for connected and autonomous electrified vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11098–11115, Aug. 2022.
[28] Y. Shang, C. Ma, K. Yang, and D. Tan, "Regenerative braking control strategy based on multi-source information fusion under environment perception," *Int. J. Automot. Technol.*, vol. 23, no. 3, pp. 805–815, Jun. 2022.
[29] T. Liu, X. Hu, S. E. Li, and D. Cao, "Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 4, pp. 1497–1507, Aug. 2017.
[30] Z. Li, W. Zhuang, G. Yin, F. Ju, Q. Wang, and H. Ding, "Learning-based eco-driving strategy design for connected power-split hybrid electric vehicles at signalized corridors," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Aachen, Germany, Jun. 2022, pp. 1226–1233.
[31] Y. Zhang, H. Xie, and K. Song, "An optimal vehicle speed planning algorithm for regenerative braking at traffic lights intersections based on reinforcement learning," in *Proc. 4th CAA Int. Conf. Veh. Control Intell. (CVCI)*, Hangzhou, China, Dec. 2020, pp. 193–198.
[32] B. Liu, C. Sun, B. Wang, and F. Sun, "Adaptive speed planning of connected and automated vehicles using multi-light trained deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 3533–3546, Apr. 2022.
[33] N. Li, J. Yang, J. Jiang, F. Hong, Y. Liu, and X. Ning, "Study on speed planning of signalized intersections with autonomous vehicles considering regenerative braking," *Processes*, vol. 10, no. 7, p. 1414, Jul. 2022.
[34] R. Nian, J. Liu, and B. Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106886.
[35] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

XIAOBIN NING was born in Taiyuan, Shanxi, China, in 1965. He received the B.S. and M.S. degrees in mechanical engineering from Taiyuan University of Science and Technology, in 1991, and the Ph.D. degree in mechanical engineering from the University of Science and Technology Beijing, in 2004.

From 1991 to 2000, he was a Lecturer with Taiyuan University of Science and Technology. Since 2004, he has been an Assistant Professor with the College of Mechanical Engineering, Zhejiang University of Technology. He is the author of more than 30 articles and more than 20 invention patents. His research interests include regenerative braking system control method for automobiles and structural safety of large construction machinery.

HAORAN TIAN was born in Yongkang, Zhejiang, China, in 1999. He received the B.S. degree in vehicle engineering from Zhejiang University of Technology, in 2021, where he is currently pursuing the M.S. degree in mechanical engineering with the College of Mechanical Engineering.

He holds two invention patents. His research interests include intelligent driving and regenerative braking for electric vehicles.

YONG LIN was born in Xinyang, Henan, China, in 1980. He received the Ph.D. degree in vehicle engineering from Zhejiang University, in 2009.

He is currently an Associate Professor and the Deputy Dean of the School of Mechanical Engineering, Zhijiang College of Zhejiang University of Technology. His research interests include mechatronics technology and intelligent testing technology.

XUEPING YAO was born in 1987. He received the B.S. degree in vehicle engineering from Zhejiang University of Technology.

From 2011 to 2013, he was with Zhejiang Youngman Passenger Car Group Company Ltd. From 2013 to 2017, he was with Zotye Automobile Company Ltd. Since 2017, he has been with Hozon New Energy Automobile Company Ltd. He is currently the Director of the Electric Chassis Department, Automotive Engineering Research Institute, Hozon New Energy Automobile Company Ltd. His research interests include chassis wire control and intelligence.

FEI HU was born in 1987. He received the M.S. degree in vehicle engineering from Tongji University.

He is currently the Deputy Director of Zhejiang VIE Science and Technology Company Ltd. His research interest includes automotive electronics.

YUMING YIN received the M.S. degree in vehicle engineering from the University of Science and Technology Beijing, in 2013, and the Ph.D. degree in mechanical engineering from Concordia University, Canada, in 2017.

He is currently an Associate Professor with the School of Mechanical Engineering, Zhejiang University of Technology, and a Visiting Research Fellow with the School of Vehicle and Mobility, Tsinghua University. He is the author of about 40 peer-reviewed journal/conference papers, and the PI/co-PI of national and state projects. His research interests include ground vehicle system dynamics, marginal emergence control, and model-data mixed reinforcement learning.