Green Energy and Intelligent Transportation 5 (2026) 100309

Contents lists available at ScienceDirect

Green Energy and Intelligent Transportation

journal homepage: www.journals.elsevier.com/green-energy-and-intelligent-transportation

Full length article

Eco-driving framework for hybrid electric vehicles in multi-lane scenarios by using deep reinforcement learning methods

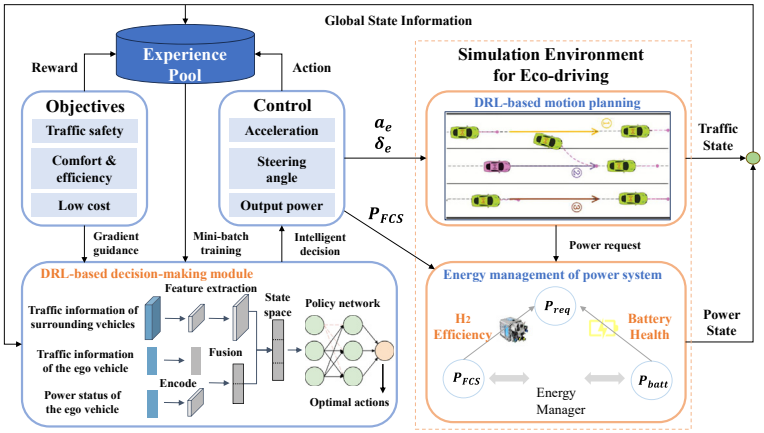
Weiqi Chen [a], Jiankun Peng [a,*], Yuhan Ma [a], Hongwen He [b], Tinghui Ren [a], Chunhai Wang [c]

[a] School of Transportation, Southeast University, Nanjing 211189, China
[b] School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China
[c] Sky-well New Energy Automobile Group Co. Ltd., Nanjing 211100, China

HIGHLIGHTS

- An integrated eco-driving framework is proposed, where motion trajectory planning and energy management are synchronously optimized.
- Traffic spatial information and power operating conditions are fused into the DRL state matrix for decisionmaking.
- The energy conservation mechanism of eco-driving is revealed by analyzing the collaborative optimization process.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Eco-driving strategy
Trajectory planning
Deep reinforcement learning
Hybrid electric vehicles

ABSTRACT

The eco-driving strategy is crucial for hybrid electric vehicles to save energy and reduce emissions. Most studies focused on longitudinal car-following or lane-changing maneuvers, lacking the consideration of continuous lateral dynamics, leading to insufficient optimization of energy-saving. This paper proposes an integrated eco-driving framework for fuel cell hybrid electric vehicles in multi-lane highway scenarios, in which trajectory planning and energy management are synchronously optimized by unified continuous control variables: acceleration, steering angle, and engine power, so as to maximize vehicle energy economy in real traffic environments. The key features of spatial traffic information and vehicular power conditions are extracted and formulated as the decision-making input. Then, the Soft Actor-Critic algorithm is utilized to optimize the eco-driving framework due to its good ability to explore complex strategy spaces for multi-objective optimization tasks. Analyses of the co-optimization process for motion trajectory planning and energy management show that, the proposed eco-driving strategy achieves better transverse-longitudinal comfort and energy economy by sacrificing 14.07% of the average speed, which results in an 87.65% improvement in the State-of-Health performance of the power

This article is part of a special issue entitled: LITS published in Green Energy and Intelligent Transportation.

* Corresponding author.
E-mail addresses: cwq@seu.edu.cn (W. Chen), jkpeng@seu.edu.cn (J. Peng).

https://doi.org/10.1016/j.geits.2025.100309
Received 25 October 2024; Received in revised form 1 January 2025; Accepted 25 February 2025
Available online 29 March 2025

system, and a reduction in the hydrogen consumption and the driving cost by 86.17% and 89.58%, respectively. This project is available at https://github.com/sicilyala/EcoAD.

1. Introduction

The carbon dioxide emission from the transportation industry accounted for 21.32% of the global CO_2 emission in 2022, and the CO_2 emission from the road transportation sector accounted for 75.74% of that from the transportation industry, according to the statistics of the International Energy Agency [1]. Therefore, it is of central importance to advance environmentally friendly vehicles in response to global warming and energy crises. Among currently feasible types of no-emission vehicles, the Fuel Cell Hybrid Electric Vehicle (FCHEV) is a research focus due to their advantages of low-contamination, low-noise, and high efficiency [2].

Ecological driving (eco-driving) strategy is a kind of technology that maximizes energy efficiency and fuel economy, and minimizes degradation of the State-of-Health (SOH) for Hybrid Electric Vehicles (HEV) [3]. With the advancement of intelligent methods, eco-driving strategies are gradually combining with autonomous driving technologies to expedite the development of an environmentally friendly intelligent transportation industry.

1.1. Eco-driving strategy

The eco-driving strategy consists of two aspects of function, one is the energy management of hybrid power systems, and the other is the vehicular motion planning [4]. As a kind of fundamental technologies of hybrid power systems, the Energy Management Strategy (EMS) aims to improve energy efficiency and fuel economy, and simultaneously reduces health degradation of the Fuel Cell System (FCS) and the power battery pack [4]. Three categories of EMS have been developed [5]: 1) rule--based; 2) optimization-based; 3) learning-based.

The rule-based EMS is easy to implement and doesn't rely on heavy computation [6]. However, it is highly dependent on engineers' practical experience, and the lack of adaptability also leads to its inability to maintain optimal performance in complicated scenarios [7].

The optimization-based EMS depends on numeric optimization algorithms, which has been proved to be effective in complex scenarios and less dependent on intricate intuitions and experiences. Peng et al. [8] utilized the Dynamic Programming (DP) method to search for optimal engine actions in a plug-in hybrid electric vehicle, and reduced diesel consumption by 10.45%. The DP-based EMS needs prior knowledge about driving conditions, and is often regarded as comparison benchmarks due to its global optimal performance. Ou et al. [9] developed an EMS for a fuel cell and power battery hybrid driving system based on Pontryagin's Minimum Principle (PMP). However, DP and PMP are both constrained by heavy computing demands and thus not practical to real-time applications. Guo et al. [10] designed an EMS using Model Predictive Control (MPC) method, and improved fuel economy by 6.48%. But the MPC method requires complex linear approximations of nonlinear powertrain systems, which weakens the control accuracy of time-variant systems [11].

The academic community has seen rapid development of learning-based EMS recently, especially deep reinforcement learning (DRL) based. A series of DRL algorithms, such as Deep Deterministic Policy Gradient (DDPG) [12,13], Twin Delay Deep Deterministic Policy Gradient (TD3) [5,14], and Soft Actor-Critic (SAC) [15], have been employed to develop EMS with continuous control. In our previous work [2], the EMS considering health degradation of both FCS and power battery was proposed based on the improved SAC method, and the overall driving cost was reduced by 15.84%. These works show that DRL-based EMS can reach near-optimal performances and eliminate unexpected model sensitivity caused by model approximations due to the model-free feature, providing a solid foundation for the research on intelligent eco-driving methods.

The operating conditions of hybrid power systems is deeply coupled to vehicle kinematic states through velocity and acceleration, which are influenced by the traffic environment, especially the preceding vehicles. So it is necessary for the eco-driving strategy to take traffic domain variables into consideration to achieve better energy economy. Ye et al. [16] reported an eco-driving strategy applied to urban intersections for approach and departure based on prediction methods, and improved fuel economy by 1.9%. In our previous works [4,17,18], integrated eco-driving strategies in car-following scenarios were proposed by using DRL methods, in which the Adaptive Cruise Control (ACC) and EMS were synchronously controlled in an uniformed optimization framework. The experiment results demonstrates that the eco-driving strategy can achieve better fuel economy when considering traffic conditions. However, these studies only focused on longitudinal movement, neglecting energy-related implications of lateral vehicle dynamics.

Vehicle motion on a level road can be orthogonally decomposed into longitudinal and lateral motion, which are coupled continuous processes and interrelated with energy efficiency through the vehicle powertrain. In order to improve vehicle energy efficiency, it is of crucial necessity to analyze the influence mechanism of vehicle kinematics and hybrid power system energy allocation, and to deeply explore the energy-saving potential of different driving trajectories in multi-lane complex traffic scenarios [19]. Huang et al. [20] developed a collaborative optimization framework of vehicle lateral dynamics and energy management based on a neural network approach to select energy-efficient lanes. Considering the coupled control of longitudinal-lateral motion and energy management, Guo et al. [21] designed a RL-based energy-efficient driving strategy to optimize the power distribution while controlling continuous longitudinal speed and discrete lane-changing maneuvers. Then, Yu et al. [22] proposed a DDPG-based lane-changing model which can simultaneously control the longitudinal and lateral motions with continuous accelerations and yaw accelerations outputted by the RL model. Although their motion planning models with continuous control actions considered energy efficiency, they focused only on the overall power output and ignored the more refined energy flow and distribution processes within the powertrain. To maximize the energy efficiency of hybrid electric vehicles, continuous lateral and longitudinal motion trajectory planning is collaboratively optimized with energy management in this work.

1.2. Vehicle motion trajectory planning

There are five main categories of vehicle motion trajectory planning algorithms [23]: 1)graph search; 2) sampling-based; 3) curve interpolation; 4) numerical optimization-based; and 5) machine-learning-based.

Graph search algorithms abstract the vehicle motion space as grid maps and search for trajectories which satisfy the constraints, mainly including Dijkstra and A* [24]. But the computational complexity grows exponentially with the scale of the motion space and the output trajectories can't be used for control modules directly, which restricts their application in high-dimension spaces and high-speed driving scenarios [25].

Sampling-based algorithms perform random sampling in the vehicle motion space so as to search for feasible trajectories in complex environments, including the Probabilistic Roadmap Method [26] and the Rapidly-exploring Random Tree [27]. Its mathematical principle comes from probabilistic completeness, thus it lacks the ability to solve complex constraints, and the uncertainty caused by probability leads to its inability to guarantee optimal results within the sampling time [28].

Curve interpolation methods utilize prior known way point information to generate paths that satisfy driving constraints such as continuity, smoothness, and obstacle avoidance requirements, mainly including polynomial curves [29], Bézier curves [30]. These methods have a low computational burden and are usually used in combination with other methods to smooth the generated trajectories [31].

Numerical optimization-based approaches model trajectory planning as a constrained multi-objective optimization task, and then solve the optimal trajectory that satisfies constraints by numerical optimization methods [32]. Guo et al. [33] considered the vehicle's lateral speed at the end of the lane-changing process as a middle parameter, thus incorporated vehicle's control sequences into a MPC problem, and then generated obstacle avoidance trajectories. Dixit et al. [34] also proposed a trajectory planning method for automated overtaking by MPC, which combined potential class field functions with reach ability sets to identify safe regions to output feasible lateral and longitudinal motions.

Machine learning-based trajectory planning methods have the advantages of good generalization and fast computation, mainly including two categories: imitation learning and DRL, and usually adopt an end-to-end architecture to process vehicle and environment interaction data [35]. Cai et al. [36] utilized the Robotcar dataset to develop a trajectory planner based on vision and imitation-learning, which learns human driver trajectory styles in the next few seconds to generate collision-free trajectories. The challenges of imitation-based approaches are how to collect enormous and consistent demonstration samples and how to ensure learning efficiency.

DRL methods have been a recent research trend in the field of motion trajectory planning, due to the advantages of no need for predefined training samples, self-learning process through trial-error and exploration in simulated interactive environments, and good compatibility with end-to-end frameworks [37]. Zhang et al. [38] proposed a motion trajectory planning method with stability guarantee for autonomous vehicles based on SAC and the Lyapunov function, in which the linear and angular velocities are optimized in an end-to-end control architecture. The end-to-end framework maximizes system performance by self-optimizing internal components to achieve better performance with a simpler systematic network size [39], providing fresh insights into eco-driving methods.

1.3. Motivation and contribution

From the perspective of methodology, rule-based eco-driving approaches utilize daily driving experiences (e.g., slow acceleration) to design velocity control algorithms. Despite their practicality, their robustness and performance are not satisfactory in dynamic traffic scenarios [40]. Optimization-based eco-driving methods assemble traffic information into hierarchical structures to allocate power flow according to the optimized speed profile; however, the limited information interaction and the hierarchical optimization framework hinder their performance and impede them to realize collaborative multi-objective optimization [41].

In order for the intelligent transportation sector to promote the application of eco-driving strategies, as shown in Fig. 1, this paper designs an eco-driving framework based on the SAC algorithm. Firstly, the traffic information of surrounding vehicles and the operating conditions of the ego vehicle are gather by various sensors. Then, based on the extracted state feature matrix, the SAC decision-making module generates the optimal control sequence, including acceleration, steering angle, and output power of FCS, towards to the optimization objectives. Finally, the simulation environment executes the actions simultaneously in motion planning and energy management modules respectively, and then returns the reward and the next state. To the best of our knowledge, it is the first time to utilize DRL methods to synchronously optimize trajectory planning and EMS for FCHEV. Here are the main contributions:

1) An integrated eco-driving framework in multi-lane highway scenarios is proposed to achieve energy-efficient driving, in which the motion trajectory planning and the energy management for FCHEV is synchronously optimized.
2) Utilizing the CNN-based deep encoder, the state matrix consisting of traffic spatial information and power operating condition is built to instruct the DRL agent better comprehend the controlled environment in the end-to-end paradigm.
3) The energy conservation mechanism of eco-driving strategy is revealed by analyzing the collaborative optimization process of vehicle motion trajectory planning and hybrid power system energy management.

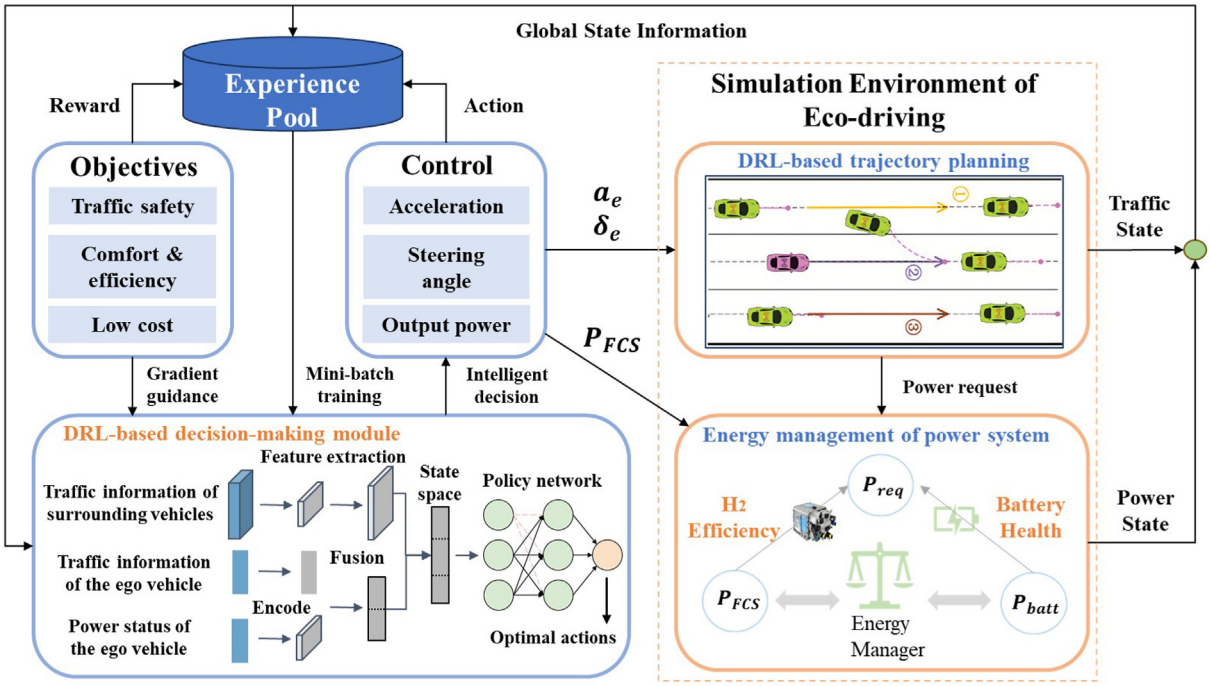


Fig. 1. The framework of the proposed eco-driving strategy.

4) The proposed eco-driving strategy shows superior performance to existing DRL methods, and demonstrates good robustness in dynamic environment with different traffic flow density.

The remainder is organized as follows. Section 2 elaborates on the driving scenario, vehicle dynamics and energy-flow models. Section 3 formulates the DRL-based eco-driving strategy. Section 4 discusses the simulation results, and Section 5 concludes this paper.

2. Model description

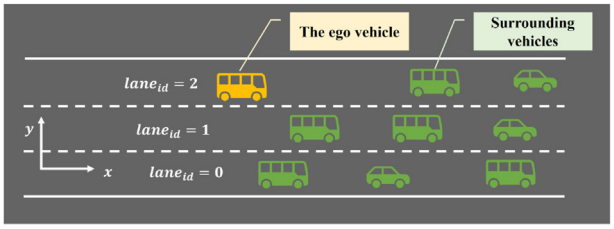
2.1. Driving scenario and vehicle kinematics

The Highway-env [42], an open-source simulation platform for automated vehicles, is employed in this paper to construct the straight highway driving scenario as shown in Fig. 2(a). The acceleration and steering angle of the ego vehicle are controlled by the proposed DRL-based strategy, and the longitudinal and lateral maneuvers of surrounding vehicles are controlled by the IDM (Intelligent Driver Model) and the MOBIL (Minimize Overall Braking Induced by Lane Change) respectively, which are built into the Highway-env.

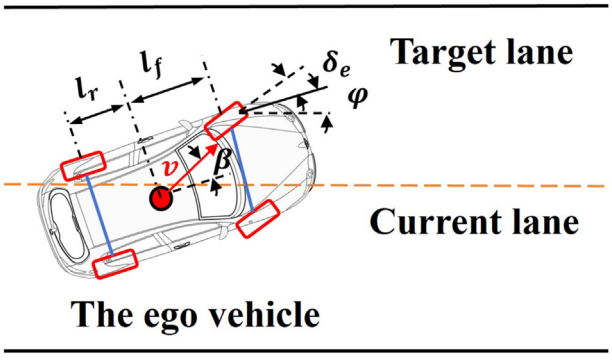
As shown in Fig. 2(b), the Bicycle kinematics model [43] is employed to predict and control the kinematic process of the ego vehicle, which takes steering angle and acceleration as inputs and then predicts kinematic states such as position, velocity and acceleration. It is formulated as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\varphi} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \sin(\beta+\varphi) \\ 0 & 0 & 0 & \cos(\beta+\varphi) \\ 0 & 0 & 0 & (\sin\beta)/l_r \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \varphi \\ v \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} a_e \tag{1}$$

where x and y are the longitudinal and lateral positions respectively, l_f and l_r are the distances between the vehicle mass center and the front and rear axles respectively. a_e and v are the acceleration and speed of the ego vehicle respectively, and β is the side-slip angle of the mass center, which is related to the steering angle δ_e as follows:



(a)



(b)

Fig. 2. The driving scenario and kinematics model. (a) The highway driving scenario. (b) The Bicycle kinematics model.

$$\beta = \tan^{-1}\frac{l_r \tan\delta_e}{l_r + l_f} \tag{2}$$

2.2. Vehicle dynamics

The main configurations of the FCHEV and the power system structure are shown in Table 1 and Fig. 3 respectively. The total tractive force is calculated as follows:

$$F_t = mgf\cos\theta + mgsin\theta + \frac{AC_D v_e^2}{21.15} + \Phi m a_e \tag{3}$$

where θ is the road slope, Φ is the rotational inertia conversion factor, g is the gravity acceleration. According to Fig. 3, the total power required by the motor is formulated as follows:

$$P_{req} = P_{DC/DC} + P_{bat} \tag{4}$$

where $P_{DC/DC}$ and P_{bat} are output powers of the fuel cell system and power battery pack respectively.

2.3. Fuel cell system

According to the physical principles of proton exchange membrane fuel cell system (FCS), the hydrogen consumption rate is dependent on the output power of FCS:

$$\dot{m} = \frac{P_{FCS}}{\eta_{FCS} \cdot L_\nu} \tag{5}$$

where L_ν denotes the chemical energy density of hydrogen, 120 MJ/kg, and η_{FCS} is the working efficiency of FCS. The fitted mathematical relationships between the output power P_{FCS} and hydrogen consumption rate $\dot{m}$ and efficiency η_{FCS} are respectively illustrated in Fig. 4(a).

The DC/DC converter adjusts the output voltage level to align with power battery pack, whose efficiency curve is calculated by numerical fitting [2], thus we have the output power of DC/DC converter $P_{DC/DC}$ as follows:

$$P_{DC/DC} = \eta_{DC/DC} \cdot P_{FCS} \tag{6}$$

The FCS actually degrades during driving mainly due to four categories of operating conditions [2]: 1) start–stop cycle; 2) low-power load; 3) high-power load; 4) load-changing cycle. Thus we have the total health degradation:

$$SOH_{FCS} = 1 - \sum_{t=0}^{n} \left[d_{ss}(t) + d_{low}(t) + d_{high}(t) + d_{cha}(t) \right] \tag{7}$$

Table 1
Parameters of the FCHEV.

Section	Description	Symbol	Value
Vehicle	mass/kg	m	8,400
	front window area/m^2	A	6.56
	Wheel radius/m	r_w	0.467
	Coefficient of air resistance	C_D	0.55
	Coefficient of rolling resistance	f	0.012
	final drive ratio	R_{fd}	6.2
Motor	Rated/peak power/kW	P_m^r, P_m^p	100, 200
	Rated/peak torque/(N·m)	T_m^r, T_m^p	1,200, 2,400
	Rated/peak speed/rpm	W_m^r, W_m^p	800, 3,000
FCS	Rated power/kW	P_{FCS}^r	60
DC/DC converter	Rated power/kW	$P_{DC/DC}^r$	60
	Efficiency	$\eta_{DC/DC}$	[0.90, 0.95]
Power battery	Capacity/kWh	Q_0	108
	Rated voltage/V	V_{OC}^T	633

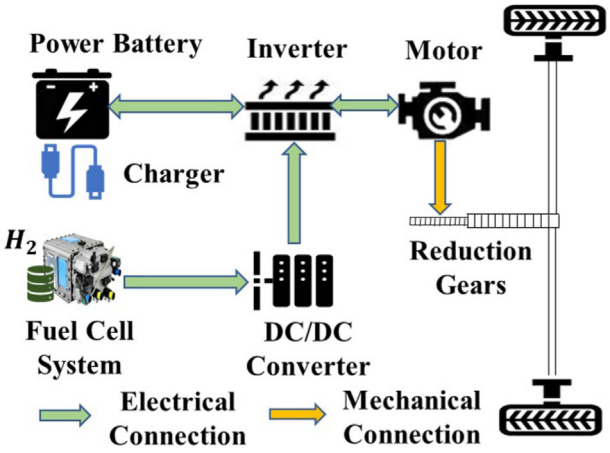
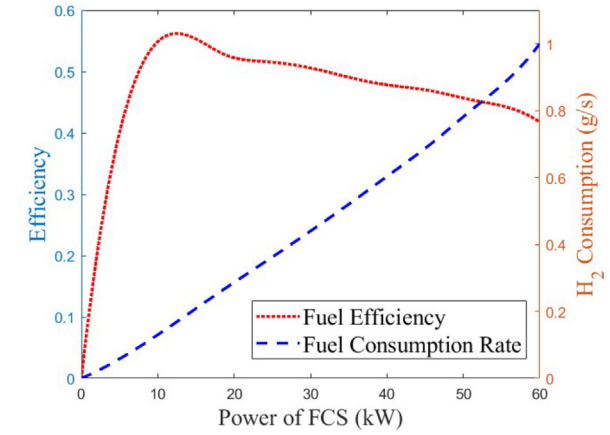
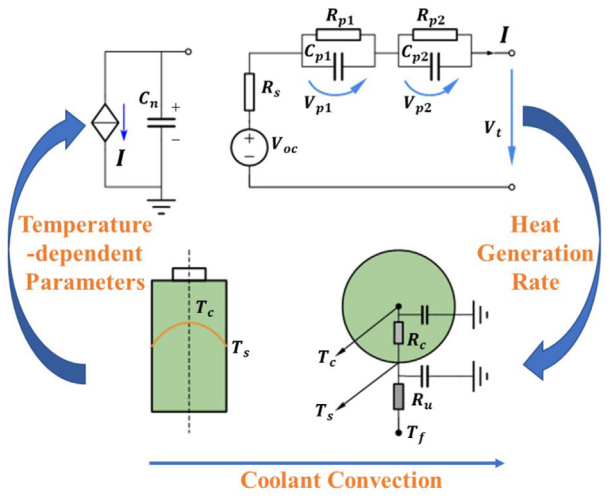


Fig. 3. The structure of power system.

where SOH_{FCS} is the State-of-Health (SOH) of FCS, and n is the number of discrete time steps t. While $d_{ss}(t)$, $d_{low}(t)$, $d_{high}(t)$, and $d_{cha}(t)$ are the four kinds of health degradation respectively.



(a)



(b)

Fig. 4. Models of the power system. (a) Hydrogen consumption rate and FCS efficiency. (b) Coupled electric-thermal model.

2.4. Power battery model

According to Fig. 4(b), the power battery pack is established as a coupled electric-thermal-aging model consisting of three parts: a second-order RC electric model, a two-state thermal model, and an energy-throughput aging model. The electric part is formulated as follows [2]:

$$\begin{cases} \dfrac{\mathrm{dSOC}(t)}{\mathrm{d}t} = \dfrac{I(t)}{3,600C_n} \\ \dfrac{\mathrm{d}V_{p1}(t)}{\mathrm{d}t} = -\dfrac{V_{\mathrm{p1}}(t)}{R_{\mathrm{p1}}(t)C_{\mathrm{p1}}(t)} + \dfrac{I(t)}{C_{\mathrm{p1}}(t)} \\ \dfrac{\mathrm{d}V_{p2}(t)}{\mathrm{d}t} = -\dfrac{V_{\mathrm{p2}}(t)}{R_{\mathrm{p2}}(t)C_{\mathrm{p2}}(t)} + \dfrac{I(t)}{C_{\mathrm{p2}}(t)} \\ V_{\mathrm{t}}(t) = V_{\mathrm{oc}}(\mathrm{SOC}) + V_{\mathrm{p1}}(t) + V_{\mathrm{p2}}(t) + R_{\mathrm{s}}I(t) \end{cases} \tag{8}$$

where SOC is the State-of-Charge of the power battery pack. $I(t)$ and $V_t(t)$ are the load current and terminal voltage at time step t, V_{p1} and V_{p2} are the polarization voltages featured by the capacitance C_{p1}, C_{p2} and resistance R_{p1}, R_{p2}, respectively. The thermal part is constrained by the thermal energy conservation principle:

$$\begin{cases} C_{\mathrm{c}}\dfrac{\mathrm{d}T_{\mathrm{c}}(t)}{\mathrm{d}t} = \dfrac{T_{\mathrm{s}}(t) - T_{\mathrm{c}}(t)}{R_{\mathrm{c}}} + H(t) \\ C_{\mathrm{s}}\dfrac{\mathrm{d}T_{\mathrm{s}}(t)}{\mathrm{d}t} = \dfrac{T_{\mathrm{c}}(t) - T_{\mathrm{s}}(t)}{R_{\mathrm{c}}} + \dfrac{T_{\mathrm{f}}(t) - T_{\mathrm{s}}(t)}{R_{\mathrm{u}}} \\ T_{\mathrm{a}}(t) = \dfrac{T_{\mathrm{c}}(t) + T_{\mathrm{s}}(t)}{2} \end{cases} \tag{9}$$

where T_s, T_c, T_a, T_f are the Celsius temperatures of battery surface, core, internal average and ambient respectively. R_c, R_u are the thermal resistances inside the battery and on the battery surface respectively. C_c, C_u are the equivalent thermal capacitance inside the battery and on the battery surface respectively. The heat generation rate $H(t)$ is calculated by:

$$H(t) = I(t)\left[V_{\mathrm{p1}}(t) + V_{\mathrm{p2}}(t) + R_{\mathrm{s}}(t)I(t)\right] + I(t)[T_{\mathrm{a}}(t) + 273]E_{\mathrm{n}}(\mathrm{SOC}, t) \tag{10}$$

where E_n is the entropy change of electrochemical reactions. The aging part utilizes the energy-throughput model to express battery health degradation mathematically, supposing that the battery can support a certain amount of accumulated charging flow until scrapped [44]. The SOH of power battery pack is calculated by:

$$\Delta\mathrm{SOH}_{\mathrm{bat}}(t) = -\frac{|I(t)|\Delta t}{2N(c, T_{\mathrm{a}})C_{\mathrm{n}}} \tag{11}$$

where Δt denotes the current duration. $N(c, T_a)$ is the equivalent number of cycles until the power battery is scrapped. Considering impacts of C-rate (c) and temperature T_a, the percentage of capacity loss ΔC_n is:

$$\Delta C_n = B(c) \cdot \exp\left(-\frac{E_{\mathrm{a}}(c)}{RT_{\mathrm{a}}}\right) \cdot Ah^z \tag{12}$$

where $B(c)$ is the pre-exponential factor which is fitted by experiment data [4]. R denotes the ideal gas constant, 8.314 J/(mol·K), $z = 0.55$ is a power-law factor, Ah is the charging flow throughput, and E_a is the activation energy (J/mol):

$$E_{\mathrm{a}}(c) = 31,700 - 370.3c \tag{13}$$

The power battery pack reaches its end of life when C_n declines by 20%, hence, Ah and N can be calculated as follows according to Eq. (12):

$$\begin{cases} Ah(c, T_{\mathrm{a}}) = \left[20 \Big/ \left(B(c) \cdot \exp\left(\dfrac{-E_{\mathrm{a}}(c)}{RT_{\mathrm{a}}}\right)\right)\right]^{1/z} \\ N(c, T_{\mathrm{a}}) = 3,600Ah(c, T_{\mathrm{a}})/C_{\mathrm{n}} \end{cases} \tag{14}$$

Finally, given current I, temperature T_a, and other operating conditions, the degradation of SOH_{bat} is calculated by Eq. (11).

3. DRL-based eco-driving strategy

3.1. Soft Actor-Critic algorithm

The SAC algorithm is designed to maximize reward signals and policy entropy simultaneously in response to former RL methods' insufficient exploration of strategy spaces [45]. SAC constructs a critic network to approximate the state-action value function, denoted as Q and parameterized by θ, which is derived by soft Bellman iteration [45]:

$$Q(s_t, a_t) = r_t + \gamma E_{s_{t+1}, a_{t+1}}[Q(s_{t+1}, a_{t+1}) - \alpha \log(\pi(a_{t+1}|s_{t+1}))] \tag{15}$$

where s_t, a_t, r_t are state, action, reward signal at time step t respectively. γ is the discounting factor, and E denotes mathematical expectation. α is the weighting coefficient to accommodate the relative importance between policy entropy and rewards. π denotes the policy network parameterized by ϕ. The critic network is learned by minimizing the soft Bellman error:

$$J_Q(\theta) = E_{(s_t, a_t, r_t, s_{t+1}) \sim M} \frac{1}{2}[Q(s_t, a_t) - [r_t + \gamma[Q'(s_{t+1}, a_{t+1}) - \alpha \log[\pi(a_{t+1}|s_{t+1})]]]]^2 \tag{16}$$

where M is the experience replay pool, and (s_t, a_t, r_t, s_{t+1}) are mini-batches sampled randomly. The target critic network Q' with parameter θ' is established for stable and fast training, whose parameters are softly copied from θ adjusted by a step factor τ:

$$\theta' \leftarrow (1 - \tau)\theta' + \tau\theta \tag{17}$$

Using the information projection defined by the Kullback-Leibler divergence, the policy function $\pi(a_t|s_t)$ is improved by:

$$\pi = \arg\min D_{KL}\left[\pi_\phi(\bullet|s_t) \middle\| \frac{\exp(Q(s_t, \bullet)/\alpha)}{Z(s_t)}\right] \tag{18}$$

where $D_{KL}(\bullet)$ is the KL divergence. $Z(s_t)$ is the logarithm partition function used to normalize the state distribution. Then, policy function π_ϕ can be learned by minimizing the expected KL divergence:

$$J_\pi(\phi) = E_{s_t \sim M}[E_{a_t \sim \pi}[\alpha \log[\pi(a_t|s_t)] - Q(s_t, a_t)]] \tag{19}$$

The weighting coefficient α is adjusted automatically, whose gradient is computed by:

$$J(\alpha) = E_{a_t \sim \pi_\phi}[-\alpha \log \pi_\phi(a_t|s_t) - \alpha \overline{H}] \tag{20}$$

where $\overline{H}$ is the target entropy, defined by the opposite number of action dimension.

3.2. State and action spaces

It is presumed that the motion information of surrounding vehicles within a certain range can be acquired by the ego vehicle through its sensing system, from which key features about driving decisions can be extracted. A V*F state matrix is defined in this work, where V is the number of observed vehicles including the ego vehicle, and F is the number of state feature dimensions. The state matrix of vehicle i consisting of kinematic information is then defined as follows:

$$s_i^{AD} = [p_i, x_i, y_i, v_i^x, v_i^y, \cos\varphi_i, \sin\varphi_i], i \in [1, V] \tag{21}$$

where $i = 1$ indicates the ego vehicle and $i \in [2, V]$ indicates surrounding vehicles. p_i indicates whether the ego vehicle can observe the ith vehicle while $p_1 \equiv 1$. (x_i, y_i), (v_i^x, v_i^y), $(\cos\varphi_i, \sin\varphi_i)$ are position, speed, and heading angle, respectively.

According to the previous work [2], the state matrix of energy management is defined as follows:

$$s^{EMS} = [SOC, SOH_{bat}, SOH_{FCS}, P_{req}, P_{FCS}] \tag{22}$$

After feature extraction and normalization, the integrated state space $\boldsymbol{s}$ consisting of both kinematic information s_i^{AD} and dynamic information s^{EMS} is formulated as follows:

$$\boldsymbol{s} = [s_1^{AD}, \cdots, s_i^{AD}, \cdots, s_V^{AD}, s^{EMS}], i \in [1, V] \tag{23}$$

The proposed eco-driving strategy controls motion trajectory planning and energy management simultaneously, so the action space is defined as:

$$\boldsymbol{a} = [a_e, \delta_e, P_{FCS}] \quad \text{s.t.} \begin{cases} a_e \in [-2, 2]\ \mathrm{m/s^2} \\ \delta_e \in [-0.785, 4, 7, 854]\ \mathrm{rad} \\ P_{FCS} \in [0, 60]\ \mathrm{kW} \end{cases} \tag{24}$$

3.3. Reward function

There are two major optimization objectives of the proposed eco-driving strategy, including energy management and motion trajectory planning. The integrated reward function which has the same mathematical meaning as the cost function, is defined as follows:

$$r(t) = -\omega \cdot r_{EMS}(t) + r_{AD}(t) \tag{25}$$

where ω is the weight coefficient, adjusting the relative importance between r_{EMS} and r_{AD}. The reward about energy management is defined as follows:

$$r_{EMS}(t) = \rho_1 \dot{m}(t) + \rho_2 \Delta SOH_{FCS}(t) + \rho_3 \Delta SOH_{bat}(t) + \rho_4 |SOC(t) - SOC_{ref}| \tag{26}$$

where ρ_1, ρ_2, ρ_3 are the market prices of hydrogen, the fuel cell system, and the power battery pack, respectively. So they can be measured uniformly by money. According to the market survey, we set ρ_1, ρ_2, ρ_3 as 55 CNY/kg, 1,500 CNY/kWh, 5,000 CNY/kW, respectively. While ρ_4 is the weight coefficient to adjust the importance of SOC margin, which is set as 100 according to our previous study [2]. $SOC_{ref} = 0.5$ is the SOC reference value.

There are three major optimization objectives of automated vehicles during driving: 1) Safety, avoiding collisions with other vehicles and avoiding stepping out of the road boundary [46]; 2) Comfort, vehicles should maintain a gentle acceleration/deceleration and steering process; 3) Efficiency, vehicles should run as fast as possible as they are in the highway scenario. The reward about motion trajectory planning is:

$$r_{AD}(t) = \gamma_1 r_{safety}(t) + \gamma_2 r_{comfort}(t) + \gamma_3 r_{speed}(t) \tag{27}$$

where γ_1, γ_2, γ_3 are the weight coefficients, and each reward item is defined as:

$$\begin{cases} r_{safety} = \{1, \text{ if safe} \quad -2, \text{ otherwise} \\ r_{comfort} = 2 - \left|\dfrac{v_e^x \cdot \delta_e}{4}\right| \\ r_{speed} = \dfrac{v_e^x - v_{min}}{v_{max} - v_{min}} \end{cases} \tag{28}$$

where r_{safety} takes the value of 1 only when the ego vehicle is neither in collision nor out of road bounds. v_e^x denotes the longitudinal speed of the ego vehicle. $v_{max} = 30$ m/s, $v_{min} = 25$ m/s are the lane speed limit.

4. Results and discussion

Here we present a series of simulation experiments to evaluate the proposed eco-driving strategy. The simulation settings and evaluation metrics are firstly described. Next, in order to achieve the synergistic optimal performance of the two task modules of trajectory planning and energy management, multiple sets of comparative experiments are conducted, so as to find the appropriate values of γ_3 and ω. Then, the optimality of the proposed eco-driving strategy is verified by comparison experiments with a variety of DRL algorithms. Finally, the learned strategy is implemented in different traffic flow scenarios to verify its robustness.

4.1. Experiment settings

4.1.1. Simulation setup

The proposed eco-driving strategy was trained and tested in a highway scenario with three straight lanes in the same direction, and the width of each lane is 4 m, as shown in Fig. 2(a). In order to simulate the randomness and dynamics of real traffic scenarios, the spawn points of vehicles are randomized and the initial speeds are generated by sampling from the Gaussian distribution $N(25, 1)$. The density of traffic flow is 70 v/km/lane when training. The simulation and control frequencies are 100 and 50 Hz respectively. At each time step, the acceleration a_e and steering angle δ_e of the ego vehicle are generated by the DRL agent. Once the ego vehicle collides or exceeds the road boundary, the current episode is terminated immediately and the simulation scenario is reset to start the next episode of training. Then, the trained strategy is tested in 1,000 time steps in a new generated road segment with randomly status.

The computing platform for this work is the Ubuntu 20.04, equipped with a CPU of i7 13700 KF and a GPU of GeForce RTX 4090. The Convolution Neural Network is employed as the feature extractor for state information, whose architecture, kernel size, and stride are [64, 64, 64], 2, and 1, respectively. The actor and critic networks of SAC are both consist of linear layers, whose architectures are [32, 64, 32]. The size of experience replay pool is 50,000, the size of mini-batch is 128, and the total training step is 40,000.

4.1.2. Evaluation metrics

There are three groups of metrics for evaluation, namely, DRL training performance, driving performance, and energy management performance.

a) *Training:* *"Episode average reward"*, which calculates the average reward of a training episode, is used to assess the DRL training performance. In this work, once the ego vehicle collides or exceeds road boundary, it resets the simulation environment to start the next training episode, so the number of steps per episode is variable. *"Episode average length"* calculates the average of step number of past episodes, which depicts the convergence of driving safety during training.
b) *Driving:* *"Self-driving trajectory"*, shows the trajectory of the DRL-based eco-driving strategy. *"Average speed"*, which calculates the average speed during one episode of trip, indicates the traffic efficiency. *"Comfort indicator"*, $|v_e^x \cdot \delta_e|$, is the absolute value of the product of the longitudinal velocity value and the steering angle, and the higher the value, the worse the comfort performance.
c) *Energy:* *"H_2 consumption"*, which calculates the hydrogen consumption during the trip, shows the performance of energy-saving. *"SOH_{bat}"*, which is health state of the power battery pack, indicates the performance of preventing health degradation. *"Money cost of driving"*, consists of the money cost from both hydrogen consumption and power battery degradation. The above metrics are utilized show the performance of EMS.

4.2. Multi-objective trade-off for driving

Vehicle speed has a significant impact on the safety, comfort, and efficiency of the driving process in highway scenarios. To ensure that the proposed eco-driving strategy can better balance the optimization objectives of comfort and efficiency while maintaining safety, this section first determines the weight coefficient γ_3 of the speed reward in Eq. (27) through comparative experiments. To highlight the correlations between various objectives of the motion trajectory planning task, the weight coefficient ω of relevant reward for the energy management strategy is temporarily set to 0.01, to weaken its influence on overall optimization objectives.

As shown in Fig. 5, it can be observed that the agent with $\gamma_3 = 2.0$ achieves a stable curve of episode average reward with the highest convergence reward during the training process. In terms of the convergence of driving safety, although the agent with $\gamma_3 = 0.5$ achieves the highest episode average length, its episode average reward is not good enough which is an implication of worse overall performance. While the agent with $\gamma_3 = 2.0$ ranks second in terms of episode average length with the best episode average reward. This indicates that when $\gamma_3 = 2.0$, the SAC agent achieves the best convergence performance during the training process.

Fig. 6 shows the replayed spatial–temporal motion trajectories and their projections on the 2D motion plane for different speed reward weight coefficients γ_3, and Table 2 presents the collision rates of each strategies during testing. It can be observed that when $\gamma_3 = \{1.5, 2.0\}$, the agent can drive safely for the longest time with relatively lower collision rates than others, exhibiting better safety performance. It is consistent with the results shown in the episode average reward curves in Fig. 5.

Next, the efficiency and comfort of the trajectories are analyzed. Fig. 7 shows the distribution of speed and comfort indicator for different γ_3. It shows that the average speeds for the trajectories with $\gamma_3 = \{0.5, 1.0, 1.5, 2.0\}$ are very close while only the average speed for $\gamma_3 = 2.5$ is significantly small compared with others, suggesting consistent performance in the efficiency metric. Meanwhile, Fig. 7 demonstrates that the trajectory with $\gamma_3 = 2.0$ has the best lateral-longitudinal comfort, with the smallest average value and the east fluctuation of the comfort index, given the constraints of safety and efficiency. The reason is that, to ensure equal efficiency and safety, when $\gamma_3 = 2.0$, the changing magnitude of the steering angle is the most gradual, which improves comfort.

In summary, when $\gamma_3 = 2.0$, the SAC agent can achieve a good balance among various objectives related to motion trajectory planning,

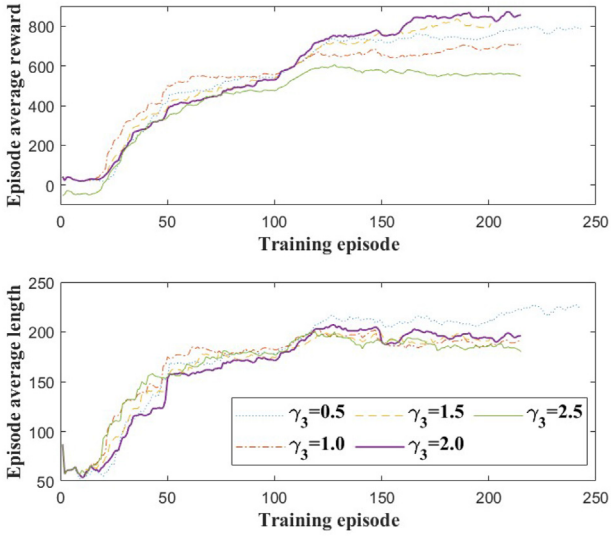


Fig. 5. The training performance for different speed weight coefficients γ_3.

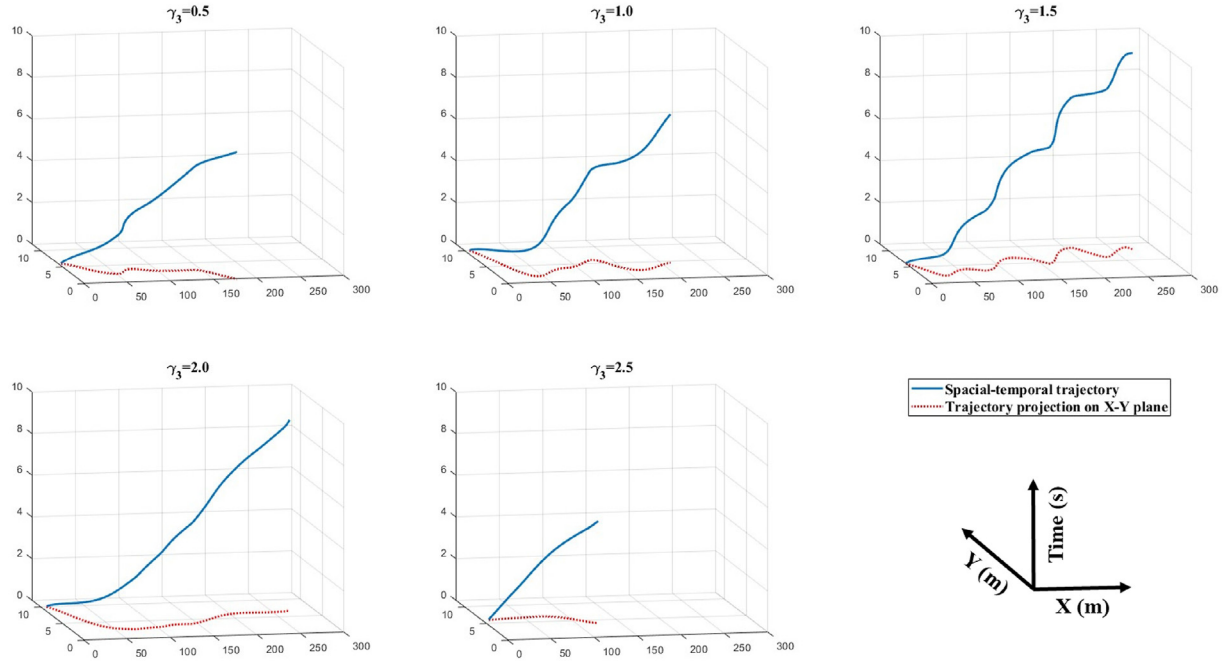


Fig. 6. The eco-driving trajectories for different γ_3. The spatial–temporal trajectories during driving and their projections on the ground (X–Y plane), with the five different speed weight coefficients γ_3.

Table 2
The collision rate during testing for different speed coefficient weight γ_3.

Value of γ_3	0.5	1.0	1.5	2.0	2.5
Collision rate	0.33%	0.52%	0.27%	0.31%	0.58%

including safety, comfort, and efficiency. The value $\gamma_3 = 2.0$ is adopted in subsequent experiments.

4.3. Multi-objective trade-off for energy

This section continues the determination of the optimal value for the reward weight coefficient, ω, associated with the energy management strategy. Fig. 8 depicts the episode average length during training for varying ω values. It is shown that when $\omega = \{5, 10\}$, the episode average length is the greatest, and consequently, the total number of training episodes is the lowest. However, Fig. 8 also reveals that the episode average reward curve corresponding to $\omega = \{5, 10\}$ is sub-optimal, failing to exploit superior strategies. The agent corresponding to $\omega = 1$ not only achieves the optimal episode average reward curve, but also shows the superior performance of episode average length. From the two metrics, it is concluded that when $\omega = 1$, the DRL agent exhibits superior optimization and convergence performance, as well as the ability to learn more effective strategies.

In order to more fully determine the optimal value of ω, the performance of the hybrid power system EMS is analyzed. As shown in Table 3, in terms of hydrogen consumption, the agent corresponding to $\omega = 0.5, 1$ is optimal, followed by the agent corresponding to $\omega = 5$; and in terms of the SOC consumption of power battery pack, the agent corresponding to $\omega = 1.5$ is optimal. In terms of SOH of the power battery pack, the agent corresponding to $\omega = 5$ is optimal, and the rest agents are not very different from each other. In terms of money cost of driving per kilometer

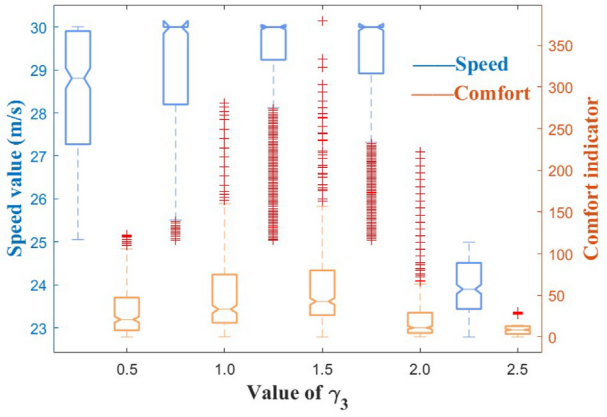


Fig. 7. The distribution of speed and comfort indicator for different γ_3. The data distributions of the speed and the comfort indicator of eco-driving strategies with the five different speed weight coefficients γ_3.

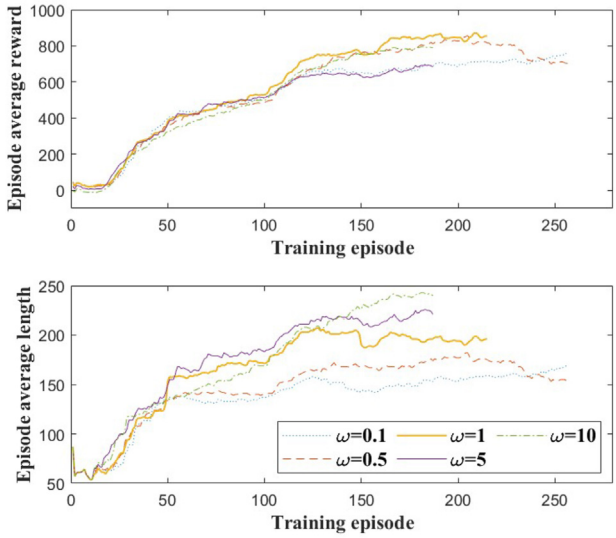


Fig. 8. The training performance for different EMS weight coefficients ω.

Table 3
EMS statistics for different ω values.

Value of ω	H_2 cost /(g·km^{-1})	SOC cost per 1 km	SOH_{FCS} degradation per 1 km	SOH_{bat} degradation per 1 km	Money cost/(CNY·km^{-1})	Money comparison
0.1	378.1	4.87e−3	4.5e−9	2.0e−4	2.14	84.64%
0.5	**62.3**	6.09e−3	2.5e−9	2.7e−4	1.55	61.18%
1	**61.6**	2.59e−3	4.0e−9	**2.1e−5**	**0.20**	**8.06%**
5	180.7	2.50e−3	2.9e−9	1.5e−4	0.67	26.58%
10	290.8	4.97e−3	1.4e−9	2.0e−4	2.53	100%

consumed during the segment, the agent corresponding to $\omega = 1$ is optimal, at only 8.06% of the most expensive, while the agent corresponding to $\omega = 5$ is the second.

In summary, in terms of both motion trajectory planning and EMS of the hybrid power system, $\omega = 1$ shows the optimal performance during training and trajectory replay, achieving a good multi-objective synergistic optimization between different task modules. At this point, the values of the two weight coefficients of the reward function that have the greatest impact on optimization performance have been determined and will be used in subsequent experiments: $\gamma_3 = 2.0$ and $\omega = 1$.

4.4. Collaborative optimization of driving and energy

This section discusses the collaborative optimization process of motion trajectory planning and EMS. In Section 4.2, the weight coefficient ω is set to 0.01 to eliminate the influence of EMS on the overall optimization objective, and is therefore used as the baseline without EMS. In contrast, Section 4.3 determines the optimal value of ω to be 1.0, and thus is used as the baseline with EMS. Fig. 9 illustrates the driving trajectories with and without EMS, it can be found that even with a significant increase in the weight of the EMS reward in the total reward function, the SAC agent can still learn and perform safe trajectory planning.

Fig. 10 illustrates the efficiency and comfort metrics corresponding to the two trajectories. When considering EMS, the average speed of the ego vehicle decreases by 4.11 m/s, which is 14.07%, indicating that the driving efficiency is negatively affected. The mean values of the comfort metrics are very similar, with a difference of only 14.02%, but the standard deviation of the comfort metrics decreases by 50.17% after considering EMS, which is a large improvement.

Fig. 11(a) shows the distributions of acceleration, vehicle total demand power, and output power of the fuel cell system for the two trajectories, respectively. It can be clearly seen that after considering EMS, the mean value and the standard variance of acceleration of the eco-driving strategy are much smaller, decreasing by 52.46% and 93.30%, respectively. With the decreasing of acceleration, the total demand power decreases, whose mean value decreases from 72.64 to 66.30 kW, a decrease of 8.73%; the output power of FCS decreases, whose mean value decreases from 18.65 kW, a decrease of 63.27%. The above results indicate that EMS helps the SAC agent to plan a more moderate spatial–temporal trajectory, which tends to help the hybrid power system achieve better energy economy.

The decrease in total demand power and FCS output power means less fuel consumption, as shown in Table 4, the hydrogen consumption decreases significantly from 445.4 to 61.6 g/km, which is a decrease of 86.17%. The reduction in power output of FCS is smaller than that in total demand power, which means that the power battery pack has to output more power, which tends to bring about a more severe degradation of the health state of the power battery pack. However, as shown in Fig. 11(b), after considering EMS, the operating conditions of the power battery pack are significantly improved, and the two parameters that have the greatest impact on the health state of the power battery, c-rate c and average temperature T_a, have both their mean and standard deviation decreased significantly. The better energy consumption operating condition is also reflected in statistics, as shown in Table 4, after considering EMS, the degradation of SOH_{bat} is only 12.35% of the one without EMS, i.e., the health performance of the power battery pack has been improved by 87.65%. Correspondingly, the money cost of driving decreases by 89.58%.

In summary, this section analyzes the co-optimization process of EMS and the motion trajectory planning. First, the EMS does not negatively affect the driving safety. Second, although the average speed decreases by 14.07% after considering EMS, the comfort indicator does not get

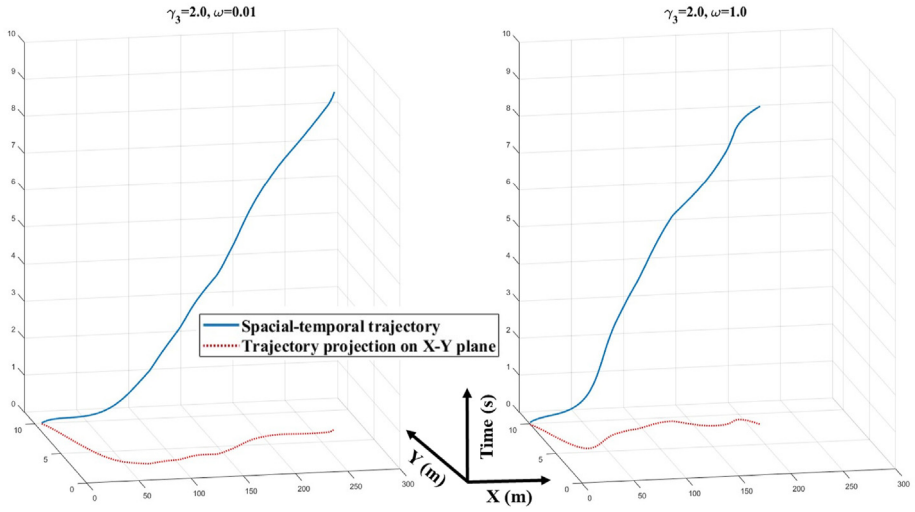


Fig. 9. Comparison of driving trajectories (Left): we set $\omega = 0.01$ to weaken the influences of EMS on the overall optimization performance as much as possible (Right): we set $\omega = 1.0$ to show the influences of EMS, demonstrating that the proposed eco-driving strategy can still learn and perform safe trajectory planning when significantly considering EMS.

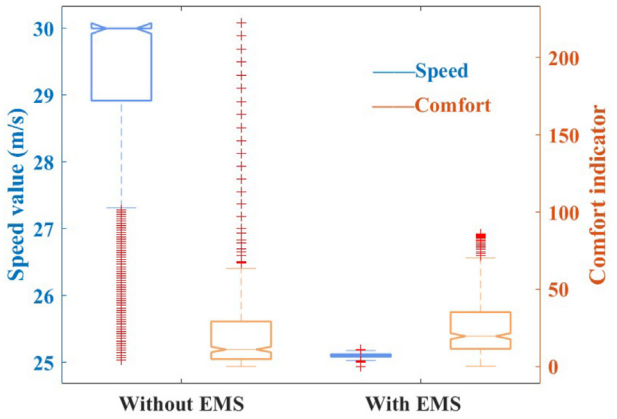
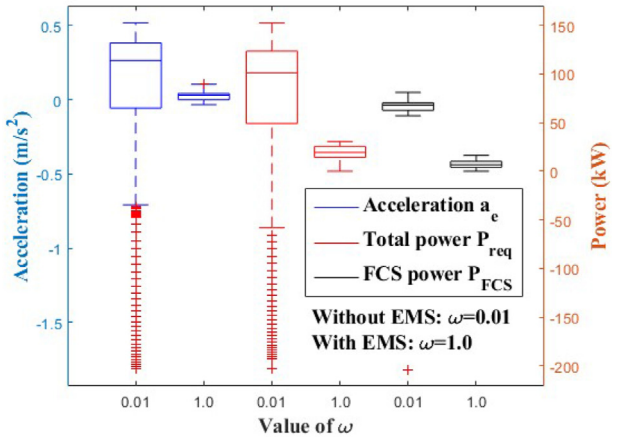
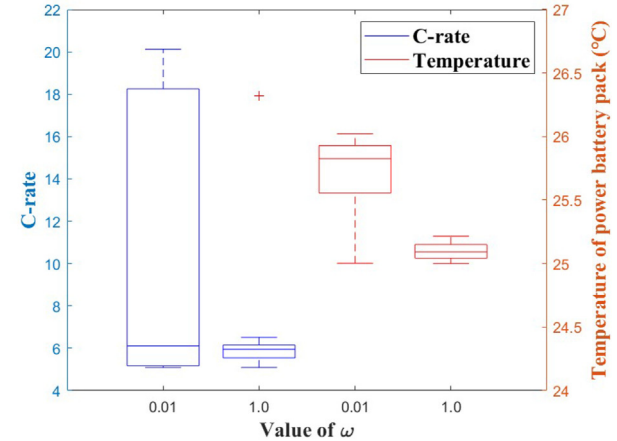


Fig. 10. Comparison of speed and comfort. The data distributions of the speed and the comfort indicator of eco-driving strategies with $\gamma_3 = 2.0$ but different values of ω, 'Without EMS' corresponds to $\omega = 0.01$ and 'With EMS' corresponds to $\omega = 1.0$.



(a)



(b)

Fig. 11. Comparison of EMS performance statistics with $\gamma_3 = 2.0$ but different values of ω. (a) The data distributions of acceleration, the total power requested, and the output power of FCS. (b) The data distributions of two important SOH indicators of the power battery pack, C-rate, and average temperature.

worse. More importantly, the hydrogen consumption decreases by 86.17%, the SOH performance of the power battery pack improves by 87.65%, and the money cost of driving decreases by 89.58%. The results demonstrate that the proposed integrated eco-driving method guides better velocities and accelerations for the motion trajectory planning module and improves operating conditions of the hybrid power system by comprehensively considering the coupling relationship between EMS and motion planning.

4.5. Optimality analysis

This section utilizes DDPG, TD3, A2C (Advantage Actor Critic), and PPO (Proximal Policy Optimization) to optimize the proposed eco-driving strategy respectively, with the same hyper parameters and $\gamma_3 = 2.0$, $\omega = 1.0$. Among them, DDPG and TD3, like SAC, are off-policy DRL algorithms; while A2C and PPO are on-policy algorithms, meaning that they do not have experience replay buffers and directly update the policy based on the data collected.

Table 5 provides a quantitative analysis for the five DRL eco-driving strategies during testing. It can be observed that the strategy based on the SAC algorithm achieved the longest travel distance of 183.71 m with the lowest collision rate, showing the optimal safety performance. Notably, the SAC strategy did not compromise driving speed for safety, as it maintains an average speed of 25.10 m/s, which is comparable to others.

Regarding the comfort metric, the A2C and PPO strategies exhibit notably better performance than the other three. However, the travel distances of the A2C and PPO strategies are significantly shorter than others, suggesting that their high comfort levels are achieved at the expense of safety, indicating the lack of a balanced trade-off among multiple objectives. In contrast, the off-policy strategies demonstrate better trade-offs between safety and comfort. The SAC, DDPG, and TD3 can maintain a reasonable level of comfort while ensuring a certain travel distance. Among them, the TD3 shows the best comfort performance but has the shortest travel distance. DDPG has a travel distance similar to TD3, but its comfort performance is significantly lower. The SAC strategy achieves the longest travel distance, nearly 20 m more than the other two off-policy algorithms, while its comfort performance is much closer to that of TD3.

Regarding EMS performance, the SAC-based strategy outperforms DDPG and TD3 in terms of money cost during the trip. The money cost of the SAC-based strategy is only 7.49% of that of the TD3-based strategy, while the DDPG-based strategy has a money cost of 91.19% of the TD3-based strategy, indicating that the SAC-based strategy achieves the best overall performance in EMS.

4.6. Robustness analysis

In this section, the trained neural network parameters are loaded into the network with the same structure and the traffic scene parameters are modified to verify the robustness of the learned eco-driving strategy under different traffic flow densities. As mentioned in 4.1.1, 70 v/km/lane is the scenario used for training, and Table 6 shows the performance statistics of the eco-driving strategy in different traffic densities. In the sparser (50 v/km/lane) scenario, the SAC agent demonstrates safer and more comfortable driving trajectories, with 76% improvement in self-driving distance, 34% improvement in comfort indicator, 25% reduction in money cost of driving compared to the training scenario, and the lowest collision rate. In the scenario with denser traffic (95 v/km/lane), the motion trajectory planning performance is slightly degraded, with a 39% decrease in self-driving distance compared to the training scenario, but the average speed, comfort indicator, and money cost of driving are all very close to the training scenario. Overall, the proposed SAC eco-driving method is able to show better performance than the training scenario in sparser scenarios. While in denser traffic scenarios, it is comparable to the training scenario, showing good adaptability.

Table 4
EMS statistics for different eco-driving strategies.

Strategy	H_2 cost /(g·km^{-1})	SOC cost per 1 km	SOH_{bat} degradation per 1 km	Money cost/(CNY·km^{-1})	Money comparison
Without EMS	445.4	3.73e−3	1.7e−4	1.92	100%
With EMS	61.6	2.59e−3	2.1e−5	0.20	10.42%

Table 5
Money cost of different eco-driving methods.

Method	Self-driving distance /m	Collision rate	Average speed/(m·s^{-1})	Comfort indicator	Money cost /(CNY·km^{-1})	Money comparison
SAC	**183.71**	**0.31%**	25.10	26.77	**0.20**	**7.49%**
DDPG	159.72	0.43%	28.39	36.27	2.48	91.19%
TD3	153.76	0.62%	27.18	17.02	2.72	100%
A2C	92.82	0.81%	24.62	9.27	0.75	27.65%
PPO	114.07	0.89%	25.66	5.22	0.53	19.50%

Table 6
Statistics of eco-driving strategies in different traffic density.

Traffic density /(v·(km·lane)$^{-1}$)	Self-driving distance/m	Collision rate	Average speed /(m·s^{-1})	Comfort indicator	Money cost /(CNY·km^{-1})
50	322.9	0.11%	24.82	17.74	0.15
70	183.7	0.31%	25.10	26.76	0.20
95	111.6	0.44%	25.12	27.39	0.21

5. Conclusions

This paper proposes an eco-driving framework for HEV with integrated collaborative optimization of motion trajectory planning and energy management. A joint state space is established through feature extraction and fusion, and a multi-objective reward function that considers traffic safety, efficiency, driving comfort, health states, and money cost of driving, is designed. Then the SAC algorithm is employed for simultaneous optimization of acceleration, steering angle, and the output power of FCS. Main conclusions are as follows.

1) Firstly, the appropriate values of the two weight coefficients that have the greatest impact on the co-optimization performance, i.e., the speed reward γ_3 and the energy management reward ω, are determined. Experiments show that the proposed eco-driving strategy achieves a good balance between multiple optimization targets when $\gamma_3 = 2.0$, $\omega = 1.0$.
2) Secondly, the co-optimization process of EMS and the motion trajectory planning is analyzed, and the eco-driving method achieves better energy economy by sacrificing 14.07% of the average speed, with a decrease of 86.17% in hydrogen consumption, an improvement of 87.65% in performance of the power battery pack, and a decrease of 89.58% in the money cost of driving.
3) Finally, the proposed eco-driving method based on SAC algorithm not only performs much better than that based on on-policy DRL algorithms, but also has a better overall performance compared with DDPG and TD3, which are also off-policy algorithms. Moreover, it shows similar performance in dynamic scenarios with different traffic flow densities, demonstrating good adaptability.

CRediT authorship contribution statement

Weiqi Chen: Writing – original draft, Project administration, Methodology, Investigation, Conceptualization. **Jiankun Peng:** Funding acquisition. **Yuhan Ma:** Conceptualization. **Hongwen He:** Resources. **Tinghui Ren:** Writing – original draft. **Chunhai Wang:** Funding acquisition.

Data availability

The data and materials used to support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interest

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

Acknowledgments

This work is supported in part by the National Key R&D Program of China (No.2022YFB4300300), the National Natural Science Foundation of China (No.52372380), and the Emission Peak & Carbon Neutrality Innovation S&T Project of Nanjing (No.202211018).

References

[1] Iea P. World energy outlook 2022. Paris, France: International Energy Agency (IEA); 2022.
[2] Chen W, Peng J, Chen J, Zhou J, Wei Z, Ma C. Health-considered energy management strategy for fuel cell hybrid electric vehicle based on improved soft actor critic algorithm adopted with beta policy. Energy Convers Manag 2023;292:117362.
[3] Chen W, Yin G, Fan Y, Zhuang W, Zhang H, Peng J. Ecological driving strategy for fuel cell hybrid electric vehicle based on continuous deep reinforcement learning. In: 2022 6th CAA international conference on vehicular control and intelligence (CVCI). IEEE; 2022. p. 1–6.
[4] Peng J, Chen W, Fan Y, He H, Wei Z, Ma C. Ecological driving framework of hybrid electric vehicle based on heterogeneous multi agent deep reinforcement learning. IEEE Transactions on Transportation Electrification 2024;10:392–406.
[5] Wang Z, He H, Peng J, Chen W, Wu C, Fan Y, et al. A comparative study of deep reinforcement learning based energy management strategy for hybrid electric vehicle. Energy Convers Manag 2023;293:117442.
[6] Thomas P, Shanmugam PK. A review on mathematical models of electric vehicle for energy management and grid integration studies. J Energy Storage 2022;55:105468.
[7] Peng J, Ren T, Chen Z, Chen W, Wu C, Ma C. Efficient training for energy management in fuel cell hybrid electric vehicles: an imitation learning-embedded deep reinforcement learning framework. J Clean Prod 2024:141360.
[8] Peng J, He H, Xiong R. Rule based energy management strategy for a series–parallel plug-in hybrid electric bus optimized by dynamic programming. Appl Energy 2017;185:1633–43.
[9] Ou K, Yuan W-W, Choi M, Yang S, Jung S, Kim Y-B. Optimized power management based on adaptive-pmp algorithm for a stationary pem fuel cell/battery hybrid system. Int J Hydrogen Energy 2018;43:15433–44.
[10] Jinquan G, Hongwen H, Jiankun P, Nana Z. A novel mpc-based adaptive energy management strategy in plug-in hybrid electric vehicles. Energy 2019;175:378–92.
[11] Lü X, Wu Y, Lian J, Zhang Y, Chen C, Wang P, et al. Energy management of hybrid electric vehicles: a review of energy optimization of fuel cell hybrid power system based on genetic algorithm. Energy Convers Manag 2020;205:112474.
[12] Li Y, He H, Khajepour A, Wang H, Peng J. Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information. Appl Energy 2019;255:113762.

[13] Zheng C, Zhang D, Xiao Y, Li W. Reinforcement learning-based energy management strategies of fuel cell hybrid vehicles with multi-objective control. J Power Sources 2022;543:231841.
[14] Wu C, Ruan J, Cui H, Zhang B, Li T, Zhang K. The application of machine learning based energy management strategy in multi-mode plug-in hybrid electric vehicle, part i: Twin delayed deep deterministic policy gradient algorithm design for hybrid mode. Energy 2023;262:125084.
[15] Wu J, Wei Z, Li W, Wang Y, Li Y, Sauer DU. Battery thermal-and health-constrained energy management for hybrid electric bus based on soft actor-critic drl algorithm. IEEE Trans Ind Inf 2020;17:3751–61.
[16] Ye F, Hao P, Qi X, Wu G, Boriboonsomsin K, Barth MJ. Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles. IEEE Trans Intell Transport Syst 2018;20:1378–89.
[17] Peng J, Fan Y, Yin G, Jiang R. Collaborative optimization of energy management strategy and adaptive cruise control based on deep reinforcement learning. IEEE Transactions on Transportation Electrification 2023;9:34–44.
[18] Chen W, Peng J, Ren T, Zhang H, He H, Ma C. Integrated velocity optimization and energy management for fchev: an eco-driving approach based on deep reinforcement learning. Energy Convers Manag 2023;296:117685.
[19] Li L, Coskun S, Zhang F, Langari R, Xi J. Energy management of hybrid electric vehicle using vehicle lateral dynamic in velocity prediction. IEEE Trans Veh Technol 2019;68:3279–93.
[20] Huang C, Li L, Fang S, Cheng S, Chen Z. Energy saving performance improvement of intelligent connected phevs via nn-based lane change decision. Sci China Technol Sci 2021;64:1203–11.
[21] Guo Q, Angah O, Liu Z, Ban XJ. Hybrid deep reinforcement learning based eco-driving for low-level connected and automated vehicles along signalized corridors. Transport Res C Emerg Technol 2021;124:102980.
[22] Yu C, Ni A, Luo J, Wang J, Zhang C, Chen Q, et al. A novel dynamic lane-changing trajectory planning model for automated vehicles based on reinforcement learning. J Adv Transp 2022;2022:1–16.
[23] liu X. Research on decision-making, planning and control of autonomous vehicle in high speed driving environment. Ph.D. thesis. Zhejiang University; 2022. doi: 1022779241.nh.
[24] Persson SM, Sharf I. Sampling-based a* algorithm for robot path-planning. Int J Robot Res 2014;33:1683–708.
[25] Erke S, Bin D, Yiming N, Qi Z, Liang X, Dawei Z. An improved a-star based path planning algorithm for autonomous land vehicles. Int J Adv Rob Syst 2020;17: 1729881420962263.
[26] Li Q, Xu Y, Bu S, Yang J. Smart vehicle path planning based on modified prm algorithm. Sensors 2022;22:6581.
[27] Khan AT, Li S, Kadry S, Nam Y. Control framework for trajectory planning of soft manipulator using optimized rrt algorithm. IEEE Access 2020;8:171730–43.
[28] Feraco S, Luciani S, Bonfitto A, Amati N, Tonoli A. A local trajectory planning and control method for autonomous vehicles based on the rrt algorithm. In: 2020 AEIT international conference of electrical and electronic technologies for automotive (AEIT automotive). IEEE; 2020. p. 1–6.
[29] Yang D, Zheng S, Wen C, Jin PJ, Ran B. A dynamic lane-changing trajectory planning model for automated vehicles. Transport Res C Emerg Technol 2018;95: 228–47.
[30] You C, Lu J, Filev D, Tsiotras P. Autonomous planning and control for intelligent vehicles in traffic. IEEE Trans Intell Transport Syst 2019;21:2339–49.
[31] Zheng L, Zeng P, Yang W, Li Y, Zhan Z. Bézier curve-based trajectory planning for autonomous vehicles with collision avoidance. IET Intell Transp Syst 2020;14: 1882–91.
[32] Rosolia U, De Bruyne S, Alleyne AG. Autonomous vehicle control: a nonconvex approach for obstacle avoidance. IEEE Trans Control Syst Technol 2016;25:469–84.
[33] Guo H, Shen C, Zhang H, Chen H, Jia R. Simultaneous trajectory planning and tracking using an mpc method for cyber-physical systems: a case study of obstacle avoidance for an intelligent vehicle. IEEE Trans Ind Inf 2018;14:4273–83.
[34] Dixit S, Montanaro U, Dianati M, Oxtoby D, Mizutani T, Mouzakitis A, et al. Trajectory planning for autonomous high-speed overtaking in structured environments using robust mpc. IEEE Trans Intell Transport Syst 2019;21:2310–23.
[35] Chen L, Teng S, Li B, Na X, Li Y, Li Z, et al. Milestones in autonomous driving and intelligent vehicles—part ii: perception and planning. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2023;53(10):6401–15.
[36] Cai P, Sun Y, Chen Y, Liu M. Vision-based trajectory planning via imitation learning for autonomous vehicles. In: 2019 IEEE intelligent transportation systems conference (ITSC). IEEE; 2019. p. 2736–42.
[37] Aradi S. Survey of deep reinforcement learning for motion planning of autonomous vehicles. IEEE Trans Intell Transport Syst 2020;23:740–59.
[38] Zhang L, Zhang R, Wu T, Weng R, Han M, Zhao Y. Safe reinforcement learning with stability guarantee for motion planning of autonomous vehicles. IEEE Transact Neural Networks Learn Syst 2021;32:5435–44.
[39] Ye F, Zhang S, Wang P, Chan C-Y. A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles. In: 2021 IEEE intelligent vehicles symposium (IV). IEEE; 2021. p. 1073–80.
[40] Li J, Fotouhi A, Liu Y, Zhang Y, Chen Z. Review on eco-driving control for connected and automated vehicles. Renew Sustain Energy Rev 2024;189:114025.
[41] Singh H, Kathuria A. Profiling drivers to assess safe and eco-driving behavior–a systematic review of naturalistic driving studies. Accid Anal Prev 2021;161: 106349.
[42] Leurent E. An environment for autonomous driving decision-making. 2018.
[43] Kong J, Pfeiffer M, Schildbach G, Borrelli F. Kinematic and dynamic vehicle models for autonomous driving control design. In: 2015 IEEE intelligent vehicles symposium (IV). IEEE; 2015. p. 1094–9.
[44] Ebbesen S, Elbert P, Guzzella L. Battery state-of-health perceptive energy management for hybrid electric vehicles. IEEE Trans Veh Technol 2012;61: 2893–900.
[45] Haarnoja T, Zhou A, Hartikainen K, Tucker G, Ha S, Tan J, et al. Soft actor-critic algorithms and applications. 2018. arXiv preprint arXiv:1812.05905.
[46] Tang X, Huang B, Liu T, Lin X. Highway decision-making and motion planning for autonomous driving via soft actor-critic. IEEE Trans Veh Technol 2022;71:4706–17.