IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 July 2024, accepted 15 August 2024, date of publication 19 August 2024, date of current version 29 August 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3445663

RESEARCH ARTICLE

Advancing Active Suspension Control With TD3-PSC: Integrating Physical Safety Constraints Into Deep Reinforcement Learning

MINGXING DENG, DONGXU SUN, LIU ZHAN, XIAOWEI XU, AND JUNYI ZOU
College of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
Corresponding author: Liu Zhan (zhanliu2021@wust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0125200, in part by the National Natural Science Foundation of China under Grant 52202480, and in part by Hubei Provincial Key Research and Development Program of China under Grant 2021BAA018.

ABSTRACT This study addresses the limitations of traditional active and semi-active suspension control systems in terms of adaptability and nonlinear handling, by exploring the potential of Deep Reinforcement Learning (DRL) techniques. Initially, a framework based on the Twin Delayed Deep Deterministic policy gradient (TD3) specific to active suspension systems was developed. Building on this, an enhanced TD3 algorithm, TD3-PSC (Physically Safe Constraint TD3), incorporating physical safety constraints was proposed. The TD3-PSC algorithm extends the state space to enhance understanding of suspension dynamics and improve adaptability. To accommodate the physical constraints and actuator characteristics inherent in suspension systems, TD3-PSC introduces guided training with real physical constraints and employs immediate termination and high penalty mechanisms to ensure safety and practicality of the algorithm. The simulation results demonstrate that TD3-PSC significantly outperforms the linear quadratic regulator (LQR), deep deterministic policy gradient (DDPG), and standard TD3 baseline, achieving improvements in control performance of 73.81%, 43.72%, and 32.14% under standard Class C road conditions, respectively. Additionally, it exhibits excellent generalization capabilities.

INDEX TERMS Deep reinforcement learning, TD3, active suspension system, physical constraint.

I. INTRODUCTION

In recent years, the rapid development of Artificial Intelligence (AI) technologies has catalyzed significant transformations across numerous industries. Particularly in the automotive industry, AI-driven control technologies have enabled remarkable advancements in smart driving and vehicle management. AI offers a broader array of solutions to vehicular control challenges than traditional methods [1], [2]. The suspension system, a critical component for enhancing ride smoothness and handling stability, embodies these advancements. Suspension systems are classified into passive, semi-active, and active types. Passive suspensions, with fixed stiffness and damping coefficients set during design, cannot adapt to changing road conditions. This limitation has spurred the rapid development of semi-active and active suspensions, which offer extensive tuning capabilities and enhanced control potential. Due to their adaptive features, these systems have attracted significant attention from manufacturers and academics alike [3].

Since its introduction in 1973, the Sky-hook strategy has gained widespread adoption in engineering for its conceptual simplicity and minimal computational demands [4]. While enhancing ride comfort, the Sky-hook strategy may neglect variations in tire dynamic loads, potentially reducing vehicle safety. To address this limitation, the Ground-hook control

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

strategy was developed, which specifically aims to enhance vehicular handling stability by focusing on tire-ground interactions [5]. Building on these concepts, developed a hybrid Sky-hook/Ground-hook damping control algorithm [6]. This hybrid approach tunes control coefficients to balance ride smoothness and maneuverability, yet struggles with optimizing multiple performance metrics simultaneously. Despite performance gains from enhancements to these strategies, challenges in complexity, reliability, and balancing driving demands continue. Optimal control uses state-space representations and weighted quadratic indices, a theoretically mature approach. By employing control theory, optimal control achieves higher rates of control and effectiveness by considering a wider array of system variables compared to the Sky-hook or Ground-hook strategies alone. Since 1976, Linear Quadratic Regulator (LQR) control has been extensively applied in active suspension systems [7], [8], [9]. Linear Quadratic Gaussian (LQG) control, an extension of LQR, includes Kalman filters to manage system noise and uncertainty, crucial for suspension control despite needing full state observability [10], [11], [12], [13]. Model Predictive Control (MPC) excels in managing complex dynamic systems and constrained problems, widely used across industrial and engineering fields. In active suspension control, MPC leverages model knowledge and optimization algorithms to identify optimal control strategies within set constraints [14], [15], [16]. However, the use of MPC is constrained to suspension scenarios with slow dynamics, as it requires substantial computational resources and real-time processing capabilities, limiting its wider application. With growing complexity and nonlinearity in automotive suspension systems, traditional control methods face challenges [17], [18]. Genetic Algorithms (GA) provide an innovative solution to complex control problems by mimicking natural selection and optimizing control systems through the adjustment of weighted matrices to derive optimal strategies [19], [20]. Neural networks, efficient at processing and adaptable, approximate nonlinear functions, suitable for suspension vibration control [21], [22], [23]. Yet, neural networks' tuning and training limitations may impact the real-time performance and reliability of control systems. In summary, traditional control theories rely on precise mathematical models and modern control theories demand extensive computational resources and stringent hardware specifications, with both facing challenges in ensuring stability.

Introduced by Minsky in the early 1960s, Reinforcement Learning (RL) is a major branch of machine learning, distinct from supervised and unsupervised learning as it derives data from dynamic environments. It utilizes environmental feedback and rewards to guide behavior choices, aiming to maximize the total rewards obtained. Deep Learning (DL) focuses on perception and representation, relying heavily on large datasets and robust computational hardware. Conversely, RL focuses on developing optimal problem-solving strategies. The growing complexity of real-world tasks has made integrating DL and RL crucial for technological advancements in control domains. Initially, the Deep Q-Network (DQN) was mainly used in 2D video games like those on Atari 2600. In May 2017, the deep RL-powered robot AlphaGo defeated the top-ranked Go champion, marking a strategic game breakthrough. By early 2018, OpenAI extended deep RL to the complex game Dota 2, showcasing its broad applicability. Deep Reinforcement Learning (DRL) has made significant progress not only in gaming applications but also in the control of complex dynamic systems. Paper [24] proposed a model-free tracking control framework based on machine learning, utilizing reservoir computing techniques. This method uses random inputs to train the system, achieving precise tracking of complex dynamic trajectories. Additionally, Paper [25] developed a model-free reinforcement learning method that employs policy iteration algorithms to solve the optimal tuning problem for discrete-time linear systems, ensuring state convergence speed. Paper [26] employed the deep Q-learning algorithm to simulate the dynamic game between attackers and defenders, formulating effective defense strategies to protect smart grids from cyber-attacks. Furthermore, DRL has also made significant advances in fields such as robotics, computer vision, healthcare, financial management, and autonomous driving [27], [28], [29].

Paper [30] introduces an enhanced DDPG algorithm using empirical samples to improve initial training in semi-active suspension systems. Paper [31] details a novel DDPG controller that combines DRL with expert advice, using PID pre-training and adaptive replay to enhance control of uncertain active suspensions. Paper [32] develops a DRL-based vehicular speed control for rough terrains, utilizing 'Maximum Comfort Speed' from crowdsourced data to optimize comfort and efficiency. Paper [33] advances a hierarchical suspension control framework combining dynamic programming with DRL. Employing EK-DDPG, it autonomously adapts to real-road conditions, enhancing comfort and efficiency.

Paper [34] employs the Soft Actor-Critic (SAC) model for semi-active suspension control in real road conditions, adapting in real-time to optimize performance. It significantly reduces vertical acceleration and pitch, enhancing comfort and handling, surpassing traditional systems. Paper [35] outlines a semi-active suspension control strategy using the Proximal Policy Optimization (PPO) algorithm, dynamically adapting the reward function for varying road conditions. Simulations show improved suspension performance by integrating dynamic road changes.

Paper [36] applies the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm to control suspension systems on varied road types, showing superior optimization, robustness, and learning efficiency over DDPG and DQN. This model addresses complex road challenges effectively, outperforming traditional controls. Subsequently, Paper [37] tackles actuator delay uncertainties in autonomous vehicles using TD3, enhancing suspension performance, ride comfort, and stability under varying delays. Following this, Paper [38] expands TD3 application by integrating soft and hard constraints (TD3-SH), allowing flexible adjustments to varying road conditions through detailed vehicle data. Further deepening the research, Paper [39] introduces Deterministic Experience Tracking (DET), a strategy that enhances vertical control decisions by processing and storing state and action data dynamically, significantly boosting ride comfort and control.

Current DRL research in active suspension control has made significant progress but still faces substantial challenges. Existing studies often optimize specific dynamics of suspension systems and road conditions but lack comprehensive state observation, potentially leading to non-robust and non-adaptive control under complex or changing conditions. Moreover, while current methods manage some dynamic changes, they struggle to fully understand and adapt to the comprehensive dynamics of suspension systems. These issues manifest as insufficient state observations, slow training speeds, non-converging training, and complex reward function settings. The performance of vehicle suspensions is crucial for driving comfort and safety. Furthermore, current reinforcement learning algorithms fail to effectively integrate road information with vehicle states, lacking strategies tailored to different driving scenarios. To address these issues, this study proposes an improved Twin Delayed Deep Deterministic Policy Gradient (TD3-PSC) algorithm. This algorithm focuses on integrating physical safety constraints and dynamically combines road conditions with the reward mechanism, thereby designing a more precise and reliable suspension control strategy.

Main contributions of this paper:

(1) Optimized State Observation for Active Suspension Control: This paper extends state observations in DRL for active suspension beyond basic parameters to include more dimensions critical to suspension performance. This broadened observation spectrum enhances model training, improves understanding of suspension dynamics, and refines control strategy specificity.

(2) Integration of Physical Safety Constraints in TD3 Algorithm (TD3-PSC): This study introduces an enhanced TD3 algorithm that incorporates physical safety constraints and the dynamic characteristics of suspension system actuators. Focused on real-world applicability, TD3-PSC ensures that control strategies are not only optimal but also practical and safe. It features mechanisms like immediate termination and high penalties during training to manage safety risks effectively, reinforcing the robustness and reliability of the active suspension control strategy.

(3) Innovative Reinforcement Learning Training Strategy: This paper introduces a dynamic training strategy utilizing a composite road surface environment, where the agent experiences a spectrum of road conditions within a single training cycle. This adaptive training approach, which cycles through varying difficulties, boosts the agent's adaptability and robustness, fostering the development of more generalized control strategies and reducing overfitting risks.

The structure of the remainder of this paper is as follows: Section two provides a concise introduction to the deep reinforcement learning environment, detailing the dynamic models of active suspension systems and the stochastic road models used to simulate varying road conditions. Section three elaborates on the architecture of the TD3-PSC algorithm and integrates physical safety constraints within the learning process to ensure the safety and feasibility of the control strategies in practical applications. Section four presents the setup and results of simulation experiments, demonstrating the performance advantages of the TD3-PSC algorithm under various road conditions compared to other control strategies. Finally, section five summarizes the findings and discusses future research directions, including algorithm optimization, application expansion, and plans for real-vehicle testing.

II. ROAD AND SUSPENSION SYSTEM MODELS

This section introduces the environmental models constructed to simulate natural driving conditions more accurately, including the road model established in accordance with ISO 8608 and the quarter-car active suspension system model.

A. COMPOSITE RANDOM ROAD MODEL

The comfort and stability of a vehicle's handling are significantly influenced by the road surface roughness. To accurately simulate this effect within the vehicle dynamics model, road inputs are categorized into deterministic and random types. Deterministic road surfaces are predefined and fixed, providing an accurate reflection of specific road conditions on vehicle behavior, typically used to analyze vehicle dynamics under certain conditions. Conversely, random road surfaces, whose characteristics are generated stochastically, more closely mirror the complexity and variability of actual driving environments that deterministic models cannot fully represent.

The random road model typically employs Power Spectral Density (PSD) to characterize variations in road roughness. PSD quantifies the distribution of road height variations across different frequencies. According to the International Organization for Standardization's ISO 8608 standard, random road roughness is classified into categories ranging from Class A to Class H based on PSD values. This classification facilitates precise application and comparison in vehicle dynamics research and road design, as detailed in Table 1.

According to the International Organization for Standardization, the PSD of road displacements can be characterized as follows:

$$G_q(n) = G_q(n_0)(\frac{n}{n_0})^{-w} \tag{1}$$

TABLE 1. Road roughness categories according to ISO standards.

Road grade	Surface roughness coefficient $G_q(n_0)/10^{-6}m^3$		
	Geometric mean	Lower limit	Upper limit
A	16	8	32
B	64	32	128
C	256	128	512
D	1024	512	2048
E	4096	2048	8192
F	16384	8192	32768
G	65536	32768	131072
H	262144	131072	524288

where n is the spatial frequency; n_0 is the reference spatial frequency, typically set at $n_0 = 0.1m^{-1}$; and w is the frequency exponent, generally taken as $w = 2$.

The road excitation time-domain model, based on the filtered white noise method, is represented by the following equation:

$$\dot{q}(t) = -2\pi f_0 q(t) + 2\pi n_0 \sqrt{G_q(n_0)v}w(t) \tag{2}$$

where f_0 is the lower cutoff frequency, $f_0 = n_{00}v$, n_{00} denotes the lower cutoff spatial frequency, typically set at $0.011m^{-1}$, $q(t)$ represents the road displacement, v is the vehicle speed, and $w(t)$ signifies white noise with a mean of zero.

To enhance the adaptability and learning outcomes of agents in diverse road environments, this study employs a composite random road scenario training method, which diverges from traditional single-environment training approaches. The core training strategy utilizes a sequence of alternating difficulty levels, arranged in an 'easy-hard-easy-hard' pattern, creating a dynamic, time-series-based training environment. This method allows the agent to experience a spectrum of road conditions from basic to complex within each training cycle, thereby improving its generalization capabilities. The temporal road roughness curves used in agent training are detailed in Figure 1. Subsequently, the spatial power spectral densities (PSDs) of four different grades of random road roughness models were calculated. By comparing these PSDs with the standard road roughness PSDs corresponding to each grade, the study ensures that the generated temporal models accurately reflect the predetermined standard road characteristics. A comparison of composite road power spectral densities is shown in Figure 2.

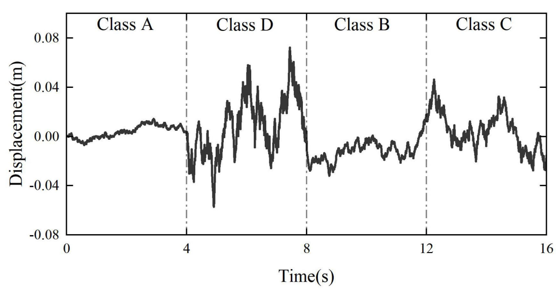

FIGURE 1. Temporal road roughness curves.

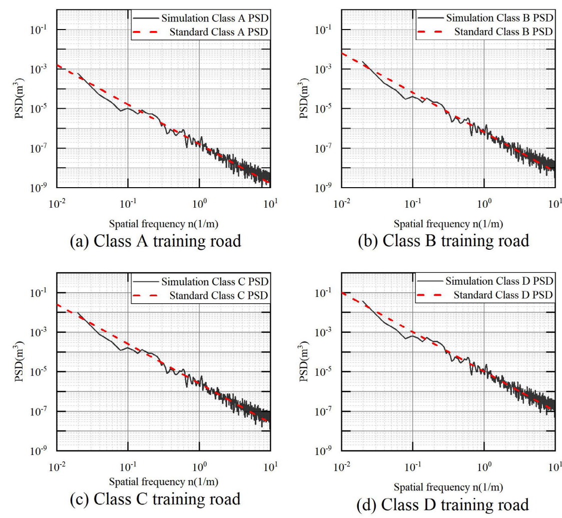

FIGURE 2. Validation of simulated road roughness power spectral density.

B. QUARTER-CAR ACTIVE SUSPENSION MODEL

Research on controlling suspension systems with deep reinforcement learning confronts increasing complexity and rising costs. The use of a two-degree-of-freedom quarter-car model as the experimental basis offers significant advantages. It simplifies the complexity of vehicle dynamics, reduces computational costs, and enhances the applicability and reliability of experiments. The quarter-car model focuses on the vehicle's vertical movements and suspension system responses, efficiently simulating dynamic performances under real-road conditions. It serves as a bridge between theoretical research and practical application, providing a clearly simplified testing environment for developing and optimizing reinforcement learning algorithms.

When employing the quarter-car model equipped with active actuators, the suspension's springs absorb and store energy from road irregularities. Subsequently, dampers release this energy through thermal dissipation. Actuators further refine the energy transformation process by adjusting control forces, aiming to reduce the vehicle's acceleration and suspension displacement. This enhances driving stability and ride comfort. Figure 3 illustrates the suspension system model.The relevant parameters are defined in Table 2.

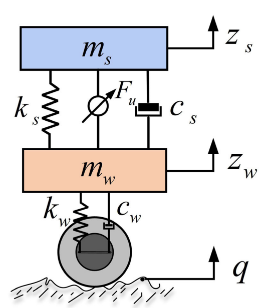

FIGURE 3. Active suspension system model.

TABLE 2. System model parameters definition.

Symbol	Parameters	Unit
m_s	Sprung mass	kg
m_w	Unsprung mass	kg
k_s	Suspension spring stiffness	N/m
k_w	Tire stiffness	N/m
c_s	Suspension damping coefficient	$N \cdot s/m$
c_w	Tire damping coefficient	$N \cdot s/m$
F_u	Actuator control force	N
z_s	Body displacement	m
z_w	Tire displacement	m
q	Pavement displacement	m

The dynamic differential equations of the suspension system can be expressed as follows:

$$\begin{cases} m_s\ddot{z}_s = F_u - k_s(z_s - z_w) - c_s(\dot{z}_s - \dot{z}_w) \\ m_w\ddot{z}_w = k_s(z_s - z_w) + c_s(\dot{z}_s - \dot{z}_w) - k_w(z_w - q) \\ \qquad -c_w(\dot{z}_w - \dot{q}) - F_u \end{cases} \tag{3}$$

For continuous linear time-invariant systems, the state-space representation can be formulated as follows:

$$\begin{cases} \dot{X} = AX + BU \\ Y = CX + DU \end{cases} \tag{4}$$

Define the state matrix as follows:

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^T = \begin{bmatrix} z_s & \dot{z}_s & z_w & \dot{z}_w \end{bmatrix}^T \tag{5}$$

Define the input matrix as follows:

$$U = \begin{bmatrix} F_u & q & \dot{q} \end{bmatrix}^T \tag{6}$$

Define the output matrix as follows:

$$Y = \begin{bmatrix} \ddot{z}_s & \dot{z}_s & z_s - z_w & q - z_w & \dot{z}_s - \dot{z}_w \end{bmatrix}^T \tag{7}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-k_s}{m_s} & \frac{-c_s}{m_s} & \frac{k_s}{m_s} & \frac{c_s}{m_s} \\ 0 & 0 & 0 & 1 \\ \frac{k_s}{m_w} & \frac{c_s}{m_w} & -\frac{k_s + k_w}{m_w} & -\frac{c_s + c_w}{m_w} \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{m_s} & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{m_w} & \frac{k_w}{m_w} & \frac{c_s}{m_w} \end{bmatrix}$$

$$C = \begin{bmatrix} -\frac{k_s}{m_s} & -\frac{c_s}{m_s} & \frac{k_s}{m_s} & \frac{c_s}{m_s} \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

$$D = \begin{bmatrix} \frac{1}{m_s} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{8}$$

III. ENHANCED TD3 ALGORITHM CONSIDERING PHYSICAL SAFETY CONSTRAINTS

This section of the research paper presents an in-depth discussion of the enhanced Twin Delayed Deep Deterministic Policy Gradient (TD3-PSC) model, developed for active suspension control. The model is designed to support complex control tasks, with significant enhancements including an expanded range of state observations for more precise system state information and an integrated training mechanism that incorporates physical safety constraints to manage potential risks during the training phase.

A. TD3-PSC MODEL

In recent years, DRL has increasingly become a focal point in the field of active suspension control, particularly for its superior handling of complex decision-making tasks. Initially, DQN algorithm was employed for active suspension systems due to its capability to manage high-dimensional state spaces. However, the DQN is inherently designed for discrete action spaces. When applied to active suspension control tasks, which require continuous control actions, DQN must discretize these actions, potentially limiting the choice of executory forces and leading to dimensionality issues that affect control outcomes and learning efficiency. As a result, the DDPG algorithm, which utilizes an Actor-Critic architecture, is better suited for continuous control problems in active suspension systems. DDPG extends traditional DQN by introducing continuous policy outputs and an actor network, enabling effective learning for continuous control tasks. Despite its advantages, DDPG faces several challenges in practical applications, including overestimation issues, insufficient exploration efficiency, and convergence problems in training. DDPG may suffer from overestimation as it tends to overvalue future states during value function updates, a phenomenon stemming from its Actor-Critic architecture where the Critic network might overestimate action values (Q-values), leading to a preference for suboptimal actions. Additionally, DDPG's exploration mechanisms, which typically involve adding noise to policy outputs, may not suffice in complex or rapidly changing environments, hindering sufficient exploration of the state space. Moreover, the complexity of continuous action spaces and environmental uncertainties can destabilize value estimates and policy updates during training, slowing or even preventing convergence.

Based on the aforementioned practical needs, this study uses the TD3 algorithm as the fundamental framework. Building upon the DDPG algorithm, the TD3 algorithm integrates several key techniques, including clipped double Q-learning, delayed policy updates, and target policy smoothing. These techniques effectively reduce estimation bias and enhance policy stability. The improved TD3 algorithm framework comprises six networks in total: two critic networks $Q_1(s, a\,|\theta_1)$, $Q_2(s, a\,|\theta_2)$ and one actor network $\pi(s\,|\phi)$, along with their corresponding target networks $Q_1'(s, a\left|\theta_1'\right.)$, $Q_2'(s, a\left|\theta_2'\right.)$, $\pi'(s\left|\phi'\right.)$. Throughout the training process, the parameters of the critic and actor networks θ_1, θ_2, ϕ are first randomly initialized. Additionally, the parameters of the corresponding target networks are also initialized:

$$\theta_1' \leftarrow \theta_1,\ \theta_2' \leftarrow \theta_2,\ \phi' \leftarrow \phi \tag{9}$$

At each time step t, the current actor network π_ϕ is used to generate an action a, with exploration noise ε added to promote agent exploration:

$$a = \pi_\phi(s) + \varepsilon,\ \varepsilon \sim N(0, \sigma) \tag{10}$$

where ε is noise sampled from a normal distribution with a mean of 0 and a standard deviation of σ.

The chosen action a is applied to the environment, which returns a reward r_s and the next state s'.The sampled transition tuple ($s\ a\ r_s\ s'$) is stored in the replay buffer β, which is later used to update the critic and actor networks.Subsequently, the target policy network π_ϕ' is used to generate the target action $\tilde{a}$ for the next state s', with added noise for smoothing:

$$\tilde{a} = \pi_\phi'(s') + \varepsilon,\ \varepsilon \sim clip(N(0, \sigma), -c, c) \tag{11}$$

Next, the target action $\tilde{a}$ and target network are used to calculate the target Q-value:

$$y = r_s + \gamma \min_{i=1,2} Q_{\theta_i'}(s', \tilde{a}) \tag{12}$$

where r_s is the current reward and λ is the discount factor. The parameters of the Q-network are updated by minimizing the loss function, which is defined as:

$$L(\theta_i) = E\left[(Q_{\theta_i}(s_i, t_i) - y)^2\right] \tag{13}$$

Update parameters by gradient descent method:

$$\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} L(\theta_i) \tag{14}$$

where λ_Q is the learning rate, and $\nabla_{\theta_i} L(\theta_i)$ represents the gradient of the loss function with respect to the parameters θ_i.

In the TD3 algorithm, the actor network's parameters are updated less frequently than those of the critic network. Specifically, for every d updates of the critic network, the actor network is updated once. The loss function for the policy network is:

$$L(\phi) = -\frac{1}{N}\sum_{i=1}^{N} Q_{\theta_1}(s_i, \pi_\phi(s_i)) \tag{15}$$

Calculate the gradient of the loss function:

$$\nabla_\phi J(\phi) = N^{-1} \sum \nabla_a Q_{\theta_1}(s, a)\left.\right|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s) \tag{16}$$

Update actor network parameters:

$$\phi \leftarrow \phi - \lambda_\pi \nabla_\phi L(\phi) \tag{17}$$

Finally, the soft update of the target network gradually aligns the parameters of the target network with those of the main network, resulting in more stable target values. Compared to hard updates, soft updates introduce a small update rate, causing only minor changes to the target network parameters at each update. This avoids drastic fluctuations and provides a more stable learning signal. For the target Q-network Q_{θ_i}', the parameters are updated as follows:

$$\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta' \tag{18}$$

For the target actor network π_ϕ', its parameters are soft updated to:

$$\phi' \leftarrow \tau\phi + (1-\tau)\phi \tag{19}$$

Further, to accommodate the physical constraints and dynamic characteristics of actuators within the active suspension system, the TD3-PSC (Physically Safe Constraint TD3) variant was developed. This variant builds on the TD3 framework by adding dynamic constraints for actuators and implementing immediate termination and high-penalty mechanisms for potential safety risks, ensuring the safety and practicality of the algorithm in real-world applications. The specifics of the TD3-PSC framework are detailed in Table 3 and illustrated in Figure 4 and Figure 5.

B. SELECTION OF STATE OBSERVATIONS

In the application of DRL to active suspension control, the careful selection and configuration of state observations are crucial for optimizing model training efficiency and enhancing final control performance. State observations provide essential environmental information required by the agent, forming the foundation for learning quality and control effectiveness. Traditional methods often limit state observation choices to basic parameters, which can restrict the model's comprehensive understanding of the dynamic performance of the suspension system.

To address this limitation, a strategy to expand and optimize the range of state observations has been proposed, incorporating multiple dimensions closely related to suspension performance. Early research in DRL applied to active suspension control typically selected state observations like body acceleration, suspension deflection, and tire dynamic load. While these parameters somewhat reflect the dynamics of the vehicle suspension, they may not capture the system's response comprehensively under complex road conditions, especially in rapidly changing or extreme driving scenarios. Relying solely on these basic parameters may lead to suboptimal control strategies, adversely affecting the suspension

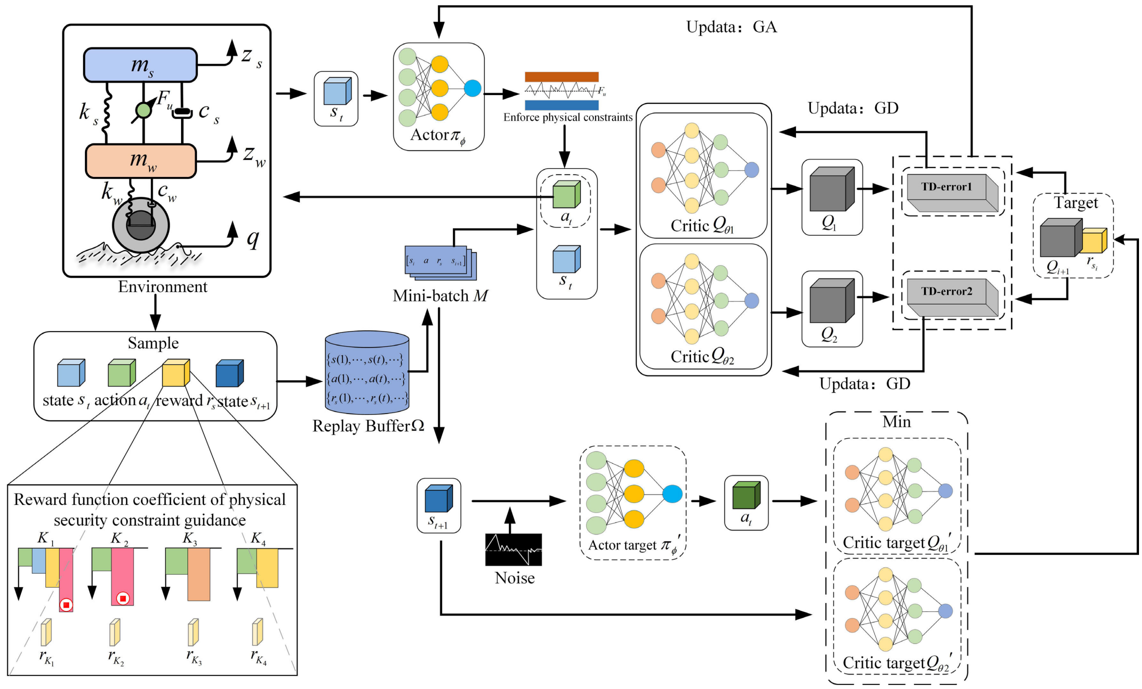

FIGURE 4. **TD3-PSC control algorithm architecture.**

TABLE 3. **TD3-PSC algorithm pseudo-code.**

Algorithm 1 TD3-PSC algorithm

Randomly initialize critic network $Q_{\theta 1}, Q_{\theta 2}$, and actor network π_ϕ , the network random parameters are θ_1, θ_2, ϕ .

Initialize target network $\theta_1' \leftarrow \theta_1, \theta_2' \leftarrow \theta_2, \phi' \leftarrow \phi$.

Initialize replay buffer β

for t=1 to T **do**

Obtain initial state of suspension system $s = [\ddot{z}_s \quad \dot{z}_s \quad \dddot{z}_s \quad z_s - z_w \quad q - z_w \quad \dot{z}_s - \dot{z}_w]$

Select actions guided by exploration noise and physical safety constraints $a \sim \pi_\phi(s) + \varepsilon, \varepsilon \sim N(0, \sigma)$

Execution action a , obtain r_s ,and update state s'

Store transition tuple $(s \quad a \quad r_s \quad s')$ in buffer β

Sample mini-batch of M transitions $(s \quad a \quad r_s \quad s')$ from β

Base on target strategy network $\pi_{\phi'}$ and state s' ,

get action $\tilde{a}$, $\tilde{a} \leftarrow \pi_{\phi'}(s') + \varepsilon, \varepsilon \sim clip(N(0, \tilde{\sigma}), -c, c)$

Target network estimation $y \leftarrow r_s + \gamma \min_{i=1,2} Q_{\theta_i'}(s', \tilde{a})$

Updated critic network weight parameters $\theta_i \leftarrow \arg\min_{\theta_i} N^{-1} \sum (y - Q_{\theta_i}(s, a))^2$

if t mod d **then**

Update the actor network through deterministic policy network gradient: $\nabla_\phi J(\phi) = N^{-1} \sum \nabla_a Q_{\theta_1}(s, a) \big|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s)$

Soft update target network:

$\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i'$

$\phi' \leftarrow \tau\phi + (1-\tau)\phi$

end if

end for

system's adaptability and performance under dynamic road conditions.

The proposed enhancement and optimization strategy for selecting state observations includes vehicle body velocity, the derivative of body acceleration, and the velocity of suspension deflection. These additional observations provide a richer set of environmental data for the DRL model. Body velocity not only reflects the state of vehicle motion but also aids in predicting future movement trends, crucial for adjusting suspension settings to accommodate varying speed conditions. The rate of change in body acceleration offers direct information on the speed of vehicle motion changes, allowing the control system to respond more agilely to changes in road surfaces and driving maneuvers. Lastly, the relative velocity of the suspension describes the rate of relative motion between two masses within the suspension system, reflecting the rate of relative displacement between internal components. These enhanced state observations significantly improve the agent's understanding of suspension dynamics, boost adaptability to complex road conditions, and increase the efficacy of control strategies. These advancements are vital supplements and improvements to traditional methods, selecting the following state observations to represent the actual state.

$$s = \begin{bmatrix} \ddot{z}_s & \dot{z}_s & \dddot{z}_s & z_s - z_w & q - z_w & \dot{z}_s - \dot{z}_w \end{bmatrix}^T \quad (20)$$

where $\ddot{z}_s$ is a critical parameter for evaluating ride comfort, directly influencing the intensity of vibrations felt by passengers. $\dot{z}_s$ represents the vertical velocity of the vehicle, used to assess the smoothness and comfort of the ride. $\dddot{z}_s$ provides the rate of change of body acceleration, reflecting the suspension system's responsiveness and sensitivity to external disturbances. $z_s - z_w$ illustrates the suspension's capacity to

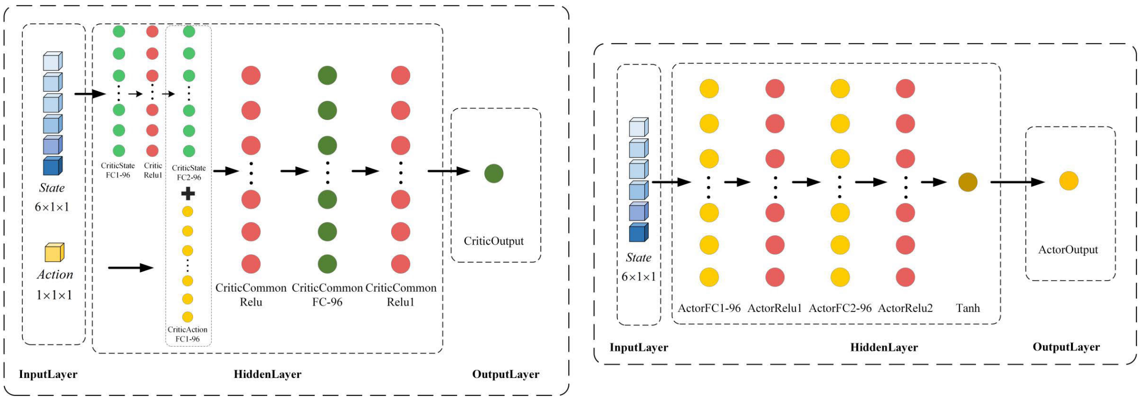

FIGURE 5. Actor-critic network structure.

absorb road impacts and adjust the vehicle's posture, affecting both driver and passenger comfort and safety. $q-z_w$ indicates the load borne by the suspension while traversing different road surfaces, a crucial metric for evaluating shock absorption and maintaining vehicle stability. $\dot{z}_s - \dot{z}_w$ measures the rate of relative displacement between internal components of the suspension system, serving as a reference for physical safety constraints discussed later.

C. REWARD MECHANISM MODEL BASED ON MULTIDIMENSIONAL STATE OBSERVATIONS

The reward function plays a crucial role within the DRL framework, defining the learning objectives and the direction for behavior optimization of the agent. By carefully designing the reward function, the agent is ensured to recognize and execute actions that maximize long-term rewards, thus optimizing its behavioral strategy. For active suspension control systems, it is essential to select physical quantities that accurately reflect the suspension's motion states and performance to construct the reward function. Key elements such as body acceleration, suspension deflection, tire dynamic load, and actuator output are chosen for the reward function formulation. These variables are directly linked to the suspension system's ride comfort and driving stability and consider the energy consumption and efficiency of the actuator. By minimizing energy usage while ensuring effective control, the system aims for sustainable development and environmental friendliness. Reward function r_1 is established to evaluate the vibrational characteristics of the suspension system.

$$r_1 = \begin{bmatrix} \ddot{z}_s & z_s - z_w & q - z_w & F_u \end{bmatrix}^T \tag{21}$$

In the formulation, the elements sequentially represent the vertical acceleration of the body, the dynamic travel of the suspension, the dynamic load on the wheels, and the control force of the active suspension system. However, due to the different magnitudes and units of the physical quantities involved in the reward function, directly using these quantities may lead to instability or biases in the learning process. It is essential to normalize these quantities to ensure the reward function effectively guides the agent. By adjusting the weights of these physical quantities in the reward function, the dynamic response characteristics of the suspension system can be flexibly altered to meet diverse control requirements and personalized settings. This reward mechanism model, based on multidimensional physical quantities, not only enhances the adaptability of the suspension system to complex road conditions but also accommodates specific driving styles and comfort needs. The reward function, after normalization and weight adjustment, is defined as follows:

$$r_s = -\left(K_1 \left|\ddot{z}_s\right|^2 + K_2 \left|z_s - z_w\right|^2 + K_3 \left|q - z_w\right|^2 + K_4 \left|F_u\right|^2\right) \tag{22}$$

where $K = \begin{bmatrix} K_1 & K_2 & K_3 & K_4 \end{bmatrix}$, the parameters within the coefficient matrix K represent the coefficients used for each variable in the computation of the reward function. If emphasis is placed on comfort, the weighting coefficient for body acceleration in the reward function can be adjusted to prioritize the reduction of body acceleration.

D. REWARD MECHANISM GUIDED BY PHYSICAL SAFETY CONSTRAINTS

In the application of DRL for active suspension system control, ensuring the physical safety of system operations is paramount. Specifically, this study imposes physical safety constraints on four key state variables: body acceleration, suspension deflection, tire dynamic load, and actuator output force. Constraints on body acceleration and suspension deflection ensure that the vehicle maintains passenger comfort and prevents safety incidents caused by excessive vibration or tilt during sudden road changes. Additionally, constraints on tire dynamic load prevent tire damage or blowouts due to overload, ensuring stability and safety during driving. Constraints on actuator output ensure that the system does not compromise equipment integrity or vehicle stability while optimizing energy efficiency and response times. However, strict physical safety constraints can limit the exploration space of the agent, making it overly conservative when attempting new behaviors. This can hinder the learning process, potentially leading to poor training outcomes or slower convergence rates. If the agent frequently terminates training episodes due to triggering safety thresholds, it may lead to a scarcity of positive rewards, affecting the agent's ability to learn effective strategies. In response to these considerations, the reward function is refined through guidance from physical safety constraints, minimizing the impact of training termination due to any state variable exceeding its safety limits.

During training, vertical body acceleration is subject to tiered constraints divided into four reward settings: within the safety and comfort threshold; exceeding the comfort but not reaching the maximum tolerance threshold; exceeding the maximum tolerance threshold but not reaching the safety threshold; and exceeding the safety threshold, which terminates the training episode and imposes a high penalty to guide the agent away from actions that pose safety risks. The expressions for these settings are as follows:

$$r_a = \begin{cases} -k_1 \left|\ddot{z}_2\right|, & \mathit{if}\ \left|\ddot{z}_s\right| \leq \left|a_1\right| \\ -k_2 \left|\ddot{z}_2\right|, & \mathit{if}\ \left|a_1\right| < \left|\ddot{z}_s\right| \leq \left|a_2\right| \\ -k_3 \left|\ddot{z}_2\right|, & \mathit{if}\ \left|a_2\right| < \left|\ddot{z}_s\right| \leq \left|a_3\right| \\ -k_4 \left|\ddot{z}_2\right|, & \mathit{otherwise} \end{cases} \tag{23}$$

where $K_1 = k_i (i = 1, 2, 3, 4)$, different body acceleration levels correspond to varying penalty coefficients, which increase incrementally; coefficients a_1, a_2, a_3 correspond to different gradients of body acceleration.

Considering vehicle design and safety margins, the safe range for suspension dynamic travel should be set to not exceed 80% of the vehicle's maximum designed suspension travel, to prevent damage to the suspension system under extreme conditions. The corresponding constraints are as follows:

$$\begin{cases} r_{f1} = -K_2 \left|z_s - z_w\right|^2, & \left|z_s - z_w\right| \leq \left|k_f z_{f\,\max}\right| \\ r_{f2} = -100K_2 \left|z_s - z_w\right|^2, & \left|z_s - z_w\right| > \left|k_f z_{f\,\max}\right| \end{cases} \tag{24}$$

where k_f represents the suspension safety travel coefficient, set at 0.8, while $z_{f\,\max}$ denotes the maximum travel designed into the suspension system.

The vehicle's design load refers to the maximum weight that a manufacturer specifies can be safely supported by the vehicle, including the weight of the vehicle itself, passengers, and cargo. In designing the load-bearing capacity of wheels and tires, both static and dynamic loads are considered. Dynamic load refers to additional forces exerted on the tires under vehicular motion, arising from conditions such as acceleration, emergency braking, or driving over uneven surfaces. To prevent wheel slippage or tire damage under these conditions, the total load-bearing capacity of the wheels should neither fall below 60% nor exceed 150% of the vehicle's design load under normal driving conditions. The constraints related to tire dynamic load are as follows:

$$\begin{cases} r_{d1} = -K_3 \left|q - z_w\right|^2, \ k_{d\,\min} D_L \leq D_t + D_s \leq k_{d\,\max} D_L \\ r_{d2} = -2K_3 \left|q - z_w\right|^2, \ D_t + D_s < k_{d\,\min} D_L \text{ or } D_t \\ \qquad\qquad + D_s > k_{d\,\max} D_L \end{cases} \tag{25}$$

where $D_t = k_w(q - z_w) + c_w(\dot{q} - \dot{z}_w)$ is the dynamic load of the tire; $D_s = (m_s + m_w)g$ is the static load of the wheel; D_L is the standard design load of the single wheel, which is slightly higher than the static load of the tire, $k_{d\,\min}, k_{d\,\max}$ are the minimum and maximum load coefficients respectively, which are set here as 0.6 and 1.5.

The design capabilities of actuators in active suspension systems represent a fundamental constraint. The magnitude and direction of control forces must be strictly limited to the maximum and minimum force values achievable by the actuators. Due to physical limitations of the actuators, the actuation force of the suspension should not exceed its limit values, and should minimize energy consumption while ensuring effective control. The constraints are defined within a two-tiered penalty function as follows:

$$r_u = \begin{cases} -k_{u1} \left|F_u\right|^2, & \mathit{if}\ \left|F_u\right| \leq \left|F_{uEco}\right| \\ -k_{u2} \left|F_u\right|^2, & \mathit{otherwise} \end{cases} \tag{26}$$

where $K_4 = k_{ui} (i = 1, 2)$, penalty coefficient of different actuator output force.

IV. SIMULATION AND RESULTS ANALYSIS

This section details the simulation setup and testing environment created using Matlab/Simulink, designed to model the response of an active suspension system under various road conditions. The environment includes both road and active suspension models. The agent was trained using a composite road model, resulting in an improved TD3-PSC agent model that incorporates physical safety constraints. To evaluate the performance of the agent, the TD3-PSC model was compared with traditional passive suspension systems and several typical control strategies, including LQR, DDPG, and the standard TD3 algorithm without physical safety constraints. Simulations were conducted under identical conditions to analyze the control effectiveness and performance of the TD3-PSC agent in simulated road conditions.

A. IMPLEMENTATION DETAILS

The simulations employed a two-degree-of-freedom vehicle active suspension model, with physical parameter details provided in Table 4. The core of the TD3-PSC consists of a dual critic network and an actor network structure. An observation space formed by six key state variables provides environmental information to the agent. The dual critic architecture reduces estimation bias by evaluating both state and action, while the actor network generates optimal actions, ensuring continuous action generation and adaptability to complex environments. Table 5 lists critical hyperparameter settings.

TABLE 4. Suspension model parameters.

Parameters	Value	Parameters	Value
m_s	365	m_w	42.5
k_s	24000	k_w	350000
c_s	2100	c_w	1000

TABLE 5. Agent hyperparameter.

	Hyperparameter	
	Item	Value
Critic network	Learning rate	1×10^{-3}
	Gradient threshold	1
	L2 Regularization factor	1×10^{-3}
Actor network	Learning rate	1×10^{-2}
	Gradient threshold	1
Agent	Sample time	0.001
	Target smoothing factor	1×10^{-3}
	Experience buffer length	1×10^{6}
	Discount factor	0.9
	Mini-batch size	64
	Noise variance	0.6
	Decay rate of noise variance	1×10^{-5}
Training process	Max episodes	200
	Max steps	8000
	Body acceleration coefficient K_1	$\{0.5\quad 1\quad 10\quad 100\}$
Physical security constraint	Suspension dynamic deflection coefficient K_2	1
	Tire dynamic load coefficient K_3	1
	Actuator output force coefficient K_4	$\{1\quad 2\}$

B. COMPARISON AND ANALYSIS OF CONTROL PERFORMANCE

To comprehensively evaluate the performance of the TD3-PSC algorithm in active suspension control, comparisons were made with traditional passive suspension systems and active systems utilizing LQR, DDPG, and the base TD3 algorithm. Before fully evaluating the agent, in order to verify the effect of extended state observations on the agent's performance, an ablation study was conducted in this paper. First, the training reward and performance of the agent under the TD3-PSC algorithm with and without extended state observations were compared. Figure 6 shows the reward curve during agent training in both cases.

In Figure 6, the green curve represents the instantaneous reward with extended state observations, while the red curve represents the instantaneous reward without extended state observations. The bold lines indicate the rolling averages for the corresponding conditions. The agent with extended state observations exhibits a rapid increase in reward values during

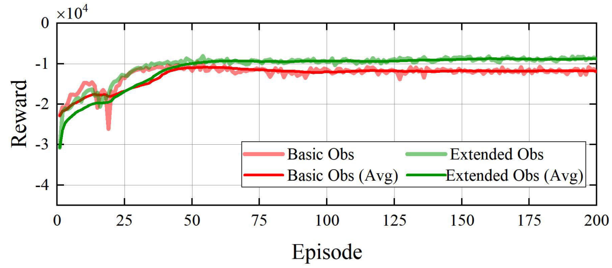

FIGURE 6. Training reward curves for agents with and without extended observations.

the initial training phase, with the rolling average demonstrating better stability. This indicates that the extended state observations significantly enhance the agent's learning efficiency and convergence speed, resulting in higher final reward values. In contrast, the agent without extended state observations shows a slower increase in reward values during the initial training phase, with larger fluctuations in the rolling average, indicating an unstable training process and poorer final performance. Figure 7 illustrates the control effect on vehicle body acceleration with and without extended state observations.

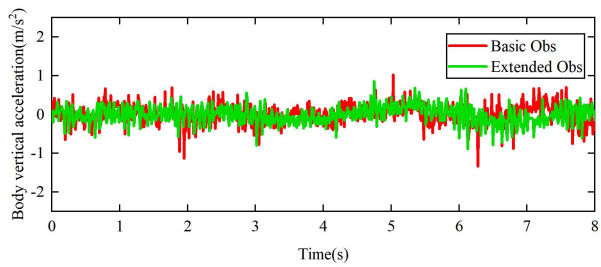

FIGURE 7. Comparison of body acceleration with and without extended observations in Class C pavement.

Figure 7 indicates that the agent trained with extended state observations achieved a lower root mean square (RMS) value for vehicle body acceleration compared to the agent trained with basic observations. This demonstrates that extended state observations enhance the agent's ability to more effectively control vehicle body acceleration, thereby improving ride comfort. Additionally, the smaller fluctuations in the green curve suggest that the agent with extended state observations exhibits greater stability in controlling vehicle body acceleration, further enhancing passenger experience.

Figure 8 illustrates the reward curve variations during the training process for different DRL algorithms. The curves in the figure represent the instantaneous rewards and rolling averages for each algorithm across different training iterations. The DDPG algorithm shows a slow initial increase in instantaneous rewards with significant fluctuations; the rolling average indicates considerable instability and slow convergence, resulting in relatively low final reward values. The TD3 algorithm exhibits a faster initial increase in instantaneous rewards with moderate fluctuations; the rolling

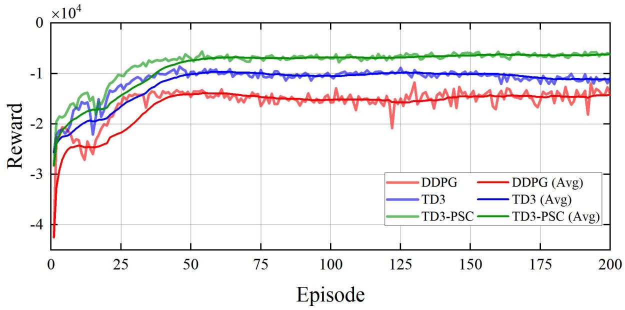

FIGURE 8. Comparison of training reward curves of different algorithms.

average demonstrates better stability and faster convergence, leading to higher final reward values than DDPG. The TD3-PSC algorithm shows the fastest initial increase in instantaneous rewards with the smallest fluctuations; the rolling average indicates the best stability and convergence speed, resulting in the highest final reward values. The TD3-PSC algorithm significantly outperforms the other two algorithms in terms of learning efficiency, stability, and overall performance.

All systems were assessed under uniform experimental conditions, specifically on standard Class B and C road surfaces, with a constant vehicle speed of 72 km/h. Performance was analyzed by comparing graphs of vehicle body vertical acceleration, suspension deflection, and tire dynamic load to assess each control strategy's effectiveness in reducing vehicle vibration and improving ride comfort, as well as their precision in responding to road irregularities. Notably, in the graphical analysis of tire dynamic load, differences between control algorithms might not be very pronounced, potentially causing overlaps in the plotted curves. To address this, a vertical translation plotting technique was employed to ensure clear differentiation of each algorithm's curve and avoid excessive overlap, thereby enabling a more accurate representation of the performance variations across control strategies.

In the above simulation diagram for Class B pavement, Figure 6 illustrates a notable performance of the TD3-PSC algorithm in controlling vehicle body acceleration, achieving an RMS acceleration of 0.0909 m/s^2. This represents an optimization of ride comfort by 87.06%, which outperforms the LQR, DDPG, and TD3 algorithms by 76.96%, 61.89%, and 41.28% respectively. Additionally, Figure 7 reveals a slight increase in suspension deflection as a trade-off for enhanced acceleration control. While this indicates a minor negative impact on the suspension's responsive flexibility, it remains well within the safe design limits. As for the tire dynamic load depicted in Figure 8, although TD3-PSC does not significantly differ from other strategies, the analysis confirms that the dynamic loads remain within a reasonable operational range.

As depicted in Figure 9, the simulation results on Class C road surfaces highlight the effectiveness of the TD3-PSC algorithm in controlling vehicle body acceleration, with an

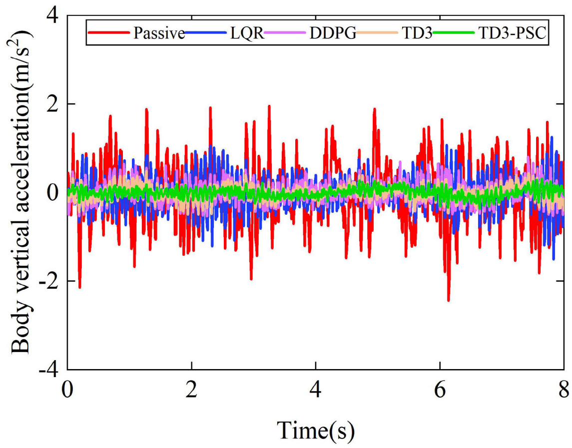

FIGURE 9. Body acceleration of different control methods on Class B pavement.

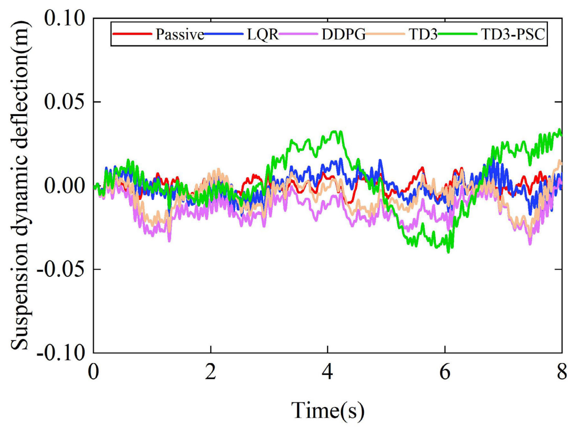

FIGURE 10. Suspension dynamic deflection of different control methods on Class B pavement.

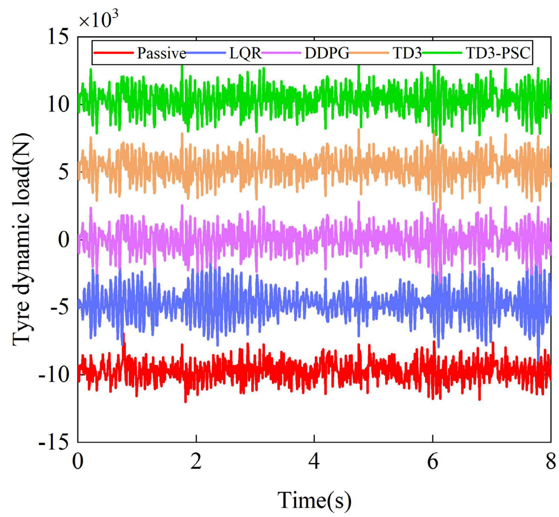

FIGURE 11. Tire dynamic load of different control methods on Class B pavement.

RMS value maintained at 0.2067. This optimization resulted in an 85.29% enhancement in ride comfort, surpassing the

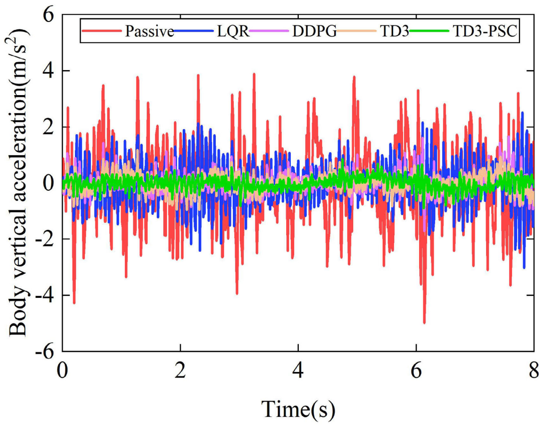

FIGURE 12. Body acceleration of different control methods on Class C pavement.

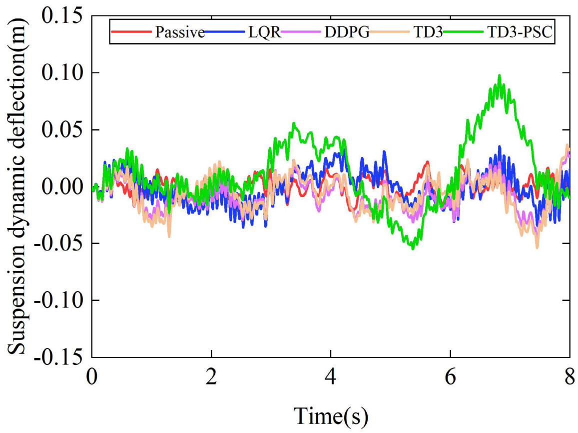

FIGURE 13. Suspension dynamic deflection of different control methods on Class C pavement.

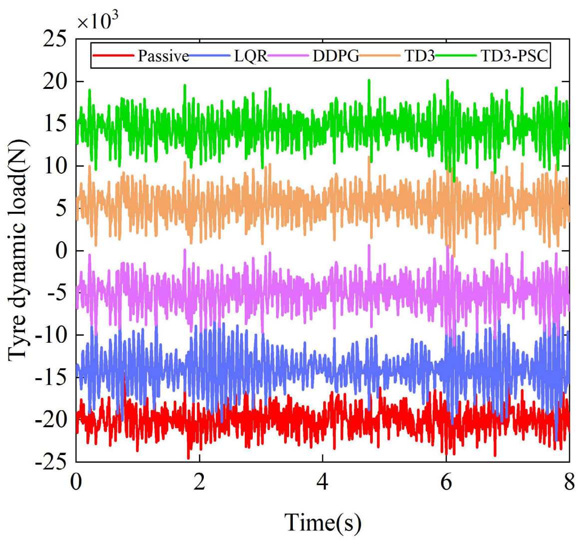

FIGURE 14. Tire dynamic load of different control methods on Class C pavement.

LQR, DDPG, and TD3 algorithms by 73.81%, 43.72%, and 32.14% respectively. Despite prioritizing ride comfort, which increased the suspension's deflection, the peak

TABLE 6. Root-mean-square value of body acceleration under different control conditions.

Algorithm	Class B road,72km/h		Class C road,72km/h	
	RMS acceleration (m/s^2)	RMS Improvement (%)	RMS acceleration (m/s^2)	RMS Improvement (%)
passive	0.7025	—	1.4050	—
LQR	0.3946	43.83	0.7891	43.84
DDPG	0.2385	66.05	0.3673	73.86
TD3	0.1548	77.96	0.3046	78.32
TD3-PSC	0.0909	87.06	0.2067	85.29

deflection at 6.8 seconds reached 0.094m without compromising vehicle safety. In Figure10, this trend was consistent with results from Class B surface tests, indicating that the control strategy causes fluctuations in suspension deflection but remains within safe limits due to training guided by physical safety constraints. Importantly, Figure 11 also shows that the active control strategy does not have excessive adverse effects on tire dynamic load. Combined with the data in Table 6, these results validate the effectiveness of the TD3-PSC algorithm in maintaining suspension system stability and its generalizability across different road surfaces.

V. CONCLUSION

This study has significantly enhanced the adaptability and robustness of deep reinforcement learning algorithms in controlling active suspension systems by integrating complex stochastic road scenarios and physical safety constraints. The TD3-PSC algorithm demonstrated superior control performance on standard Class B and Class C road surfaces, confirming its effectiveness in improving ride comfort and driving safety through critical metrics such as body acceleration, suspension deflection, and tire dynamic load. Notably, the algorithm successfully reduced the RMS of body acceleration while maintaining suspension deflection and tire loads within safe and reasonable limits, ensuring vehicle stability under complex road conditions. Future research will focus on refining and optimizing physical safety constraints. An adaptive physical safety constraint control strategy is planned, which will progressively intensify restrictions on physical safety constraints based on extensive interaction between the agent and its environment over training episodes, to regulate the agent's behavior. Additionally, conducting real-vehicle tests will be a crucial step to validate the performance and reliability of the TD3-PSC algorithm in real-world scenarios. These experiments are expected to deepen the understanding of the algorithm's practical value and potential for commercial applications in real driving environments.

REFERENCES

[1] Q. Zhang, W. Pan, and V. Reppa, "Model-reference reinforcement learning for collision-free tracking control of autonomous surface vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 7, pp. 8770–8781, Jul. 2022, doi: 10.1109/TITS.2021.3086033.

[2] G. Du, Y. Zou, X. Zhang, Z. Li, and Q. Liu, "Hierarchical motion planning and tracking for autonomous vehicles using global heuristic based potential field and reinforcement learning based predictive control," *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 8, pp. 8304–8323, Aug. 2023, doi: 10.1109/TITS.2023.3266195.

[3] Z. Liu and H. Pan, "Barrier function-based adaptive sliding mode control for application to vehicle suspensions," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 3, pp. 2023–2033, Sep. 2021, doi: 10.1109/TTE.2020.3043581.

[4] D. Karnopp, M. J. Crosby, and R. A. Harwood, "Vibration control using semi-active force generators," *J. Eng. Ind.*, vol. 96, no. 2, pp. 619–626, May 1974, doi: 10.1115/1.3438373.

[5] M. Valášek, M. Novák, Z. Šika, and O. Vaculín, "Extended groundhook–new concept of semi-active control of truck's suspension," *Vehicle Syst. Dyn.*, vol. 27, nos. 5–6, pp. 289–303, Jun. 1997, doi: 10.1080/00423119708969333.

[6] M. Ahmadian and N. Vahdati, "Transient dynamics of semiactive suspensions with hybrid control," *J. Intell. Mater. Syst. Struct.*, vol. 17, no. 2, pp. 145–153, Feb. 2006, doi: 10.1177/1045389x06056458.

[7] A. G. Thompson, "An active suspension with optimal linear state feedback," *Vehicle Syst. Dyn.*, vol. 5, no. 4, pp. 187–203, Dec. 1976, doi: 10.1080/00423117608968414.

[8] M. Nagarkar, Y. Bhalerao, G. V. Patil, and R. Z. Patil, "Multi-objective optimization of nonlinear quarter car suspension system–PID and LQR control," *Proc. Manuf.*, vol. 20, pp. 420–427, Jan. 2018, doi: 10.1016/j.promfg.2018.02.061.

[9] Y. Yao, "Optimization design of active suspension of vehicle based on LQR control," *J. Phys., Conf. Ser.*, vol. 1629, Sep. 2022, Art. no. 12094. Accessed: Apr. 18, 2024, doi: 10.1088/1742-6596/1629/1/012094.

[10] H. C. Sohn, K. T. Hong, K. S. Hong, and W. S. Yoo, "An adaptive LQG control for semi-active suspension systems," *Int. J. Vehicle Des.*, vol. 34, no. 4, p. 309, Jan. 2004, doi: 10.1504/ijvd.2004.004060.

[11] L. Chai and T. Sun, "The design of LQG controller for active suspension based on analytic hierarchy process," *Math. Problems Eng.*, vol. 2010, no. 1, pp. 1–19, Jan. 2010, doi: 10.1155/2010/701951.

[12] H. Pang, Y. Chen, J. Chen, and X. Liu, "Design of LQG controller for active suspension without considering road input signals," *Shock Vibrat.*, vol. 2017, pp. 1–13, Feb. 2017, doi: 10.1155/2017/6573567.

[13] S.-A. Chen, Y.-M. Cai, J. Wang, and M. Yao, "A novel LQG controller of active suspension system for vehicle roll safety," *Int. J. Control, Autom. Syst.*, vol. 16, no. 5, pp. 2203–2213, Oct. 2018, doi: 10.1007/s12555-017-0159-2.

[14] J. Theunissen, A. Tota, P. Gruber, M. Dhaens, and A. Sorniotti, "Preview-based techniques for vehicle suspension control: A state-of-the-art review," *Annu. Rev. Control*, vol. 51, pp. 206–235, Jan. 2021, doi: 10.1016/j.arcontrol.2021.03.010.

[15] R. K. Mehra, J. N. Amin, K. J. Hedrick, C. Osorio, and S. Gopalasamy, "Active suspension using preview information and model predictive control," in *Proc. IEEE Int. Conf. Control Appl.*, Oct. 1997, pp. 860–865, doi: 10.1109/CCA.1997.627769.

[16] C. Göhrle, A. Wagner, A. Schindler, and O. Sawodny, "Active suspension controller using MPC based on a full-car model with preview information," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 497–502, doi: 10.1109/ACC.2012.6314680.

[17] M. Q. Nguyen, M. Canale, O. Sename, and L. Dugard, "A model predictive control approach for semi-active suspension control problem of a full car," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Dec. 2016, pp. 721–726, doi: 10.1109/CDC.2016.7798353.

[18] M. Papadimitrakis and A. Alexandridis, "Active vehicle suspension control using road preview model predictive control and radial basis function networks," *Appl. Soft Comput.*, vol. 120, May 2022, Art. no. 108646, doi: 10.1016/j.asoc.2022.108646.

[19] C. Zhou, X. Liu, W. Chen, F. Xu, and B. Cao, "Optimal sliding mode control for an active suspension system based on a genetic algorithm," *Algorithms*, vol. 11, no. 12, Dec. 2018, Art. no. 12, doi: 10.3390/a11120205.

[20] W. Yu, J. Li, J. Yuan, and X. Ji, "LQR controller design of active suspension based on genetic algorithm," in *Proc. IEEE 5th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Oct. 2021, pp. 1056–1060, doi: 10.1109/ITNEC52019.2021.9587272.

[21] F. Zhao, S. S. Ge, F. Tu, Y. Qin, and M. Dong, "Adaptive neural network control for active suspension system with actuator saturation," *IET Control Theory Appl.*, vol. 10, no. 14, pp. 1696–1705, Sep. 2016, doi: 10.1049/iet-cta.2015.1317.

[22] Z. Ding, F. Zhao, Y. Qin, and C. Tan, "Adaptive neural network control for semi-active vehicle suspensions," *J. Vibroeng.*, vol. 19, no. 4, pp. 2654–2669, Jun. 2017, doi: 10.21595/jve.2017.18045.

[23] A. Hamza and N. B. Yahia, "Heavy trucks with intelligent control of active suspension based on artificial neural networks," *Proc. Inst. Mech. Eng., I, J. Syst. Control Eng.*, vol. 235, no. 6, pp. 952–969, Jul. 2021, doi: 10.1177/0959651820958516.

[24] Z.-M. Zhai, M. Moradi, L.-W. Kong, B. Glaz, M. Haile, and Y.-C. Lai, "Model-free tracking control of complex dynamical trajectories with machine learning," *Nature Commun.*, vol. 14, no. 1, p. 5698, Sep. 2023, doi: 10.1038/s41467-023-41379-3.

[25] S. E. Razavi, M. A. Moradi, S. Shamaghdari, and M. B. Menhaj, "Adaptive optimal control of unknown discrete-time linear systems with guaranteed prescribed degree of stability using reinforcement learning," *Int. J. Dyn. Control*, vol. 10, no. 3, pp. 870–878, Jun. 2022, doi: 10.1007/s40435-021-00836-x.

[26] M. Moradi, Y. Weng, and Y.-C. Lai, "Defending smart electrical power grids against cyberattacks with deep Q-learning," *PRX Energy*, vol. 1, no. 3, Nov. 2022, Art. no. 033005, doi: 10.1103/prxenergy.1.033005.

[27] Y. Li, "Deep reinforcement learning: An overview," 2017, *arXiv:1701.07274*.

[28] 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, doi: 10.1016/j.compchemeng.2020.106886.

[29] M. Panzer, B. Bender, and N. Gronau, "Deep reinforcement learning in production planning and control: A systematic literature review," in *Proc. Conf. Prod. Syst. Logistics*, 2021, pp. 535–545, doi: 10.15488/11238.

[30] L. Ming, L. Yibin, R. Xuewen, Z. Shuaishuai, and Y. Yanfang, "Semi-active suspension control based on deep reinforcement learning," *IEEE Access*, vol. 8, pp. 9978–9986, 2020, doi: 10.1109/ACCESS.2020.2964116.

[31] Z. Tan, G. Wen, Z. Pan, S. Yin, X. Wu, and G. Tohti, "Control of a nonlinear active suspension system based on deep reinforcement learning and expert demonstrations," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 13, Sep. 2023, Art. no. 09544070231191842, doi: 10.1177/09544070231191842.

[32] Y. Du, J. Chen, C. Zhao, C. Liu, F. Liao, and C.-Y. Chan, "Comfortable and energy-efficient speed control of autonomous vehicles on rough pavements using deep reinforcement learning," *Transport. Res. C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103489, doi: 10.1016/j.trc.2021.103489.

[33] Y. Du, J. Chen, C. Zhao, F. Liao, and M. Zhu, "A hierarchical framework for improving ride comfort of autonomous vehicles via deep reinforcement learning with external knowledge," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 8, pp. 1059–1078, May 2023, doi: 10.1111/mice.12934.

[34] H. Yong, J. Seo, J. Kim, M. Kim, and J. Choi, "Suspension control strategies using switched soft actor-critic models for real roads," *IEEE Trans. Ind. Electron.*, vol. 70, no. 1, pp. 824–832, Jan. 2023, doi: 10.1109/TIE.2022.3153805.

[35] S.-Y. Han and T. Liang, "Reinforcement-learning-based vibration control for a vehicle semi-active suspension system via the PPO approach," *Appl. Sci.*, vol. 12, no. 6, Jan. 2022, Art. no. 6, doi: 10.3390/app12063078.

[36] D. Shen, S. Zhou, and N. Zhang, "Twin delayed deep deterministic reinforcement learning application in vehicle electrical suspension control," *Int. J. Vehicle Perform.*, vol. 9, no. 4, pp. 429–446, Jan. 2023, doi: 10.1504/ijvp.2023.133852.

[37] Y. Wang, C. Wang, S. Zhao, and K. Guo, "Research on deep reinforcement learning control algorithm for active suspension considering uncertain time delay," *Sensors*, vol. 23, no. 18, Jan. 2023, Art. no. 18, doi: 10.3390/s23187827.

[38] C. Wang, X. Cui, S. Zhao, X. Zhou, Y. Song, Y. Wang, and K. Guo, "Enhancing vehicle ride comfort through deep reinforcement learning with expert-guided soft-hard constraints and system characteristic considerations," *Adv. Eng. Informat.*, vol. 59, Jan. 2024, Art. no. 102328, doi: 10.1016/j.aei.2023.102328.

[39] C. Wang, X. Cui, S. Zhao, X. Zhou, Y. Song, Y. Wang, and K. Guo, "A deep reinforcement learning-based active suspension control algorithm considering deterministic experience tracing for autonomous vehicle," *Appl. Soft Comput.*, vol. 153, Mar. 2024, Art. no. 111259, doi: 10.1016/j.asoc.2024.111259.

MINGXING DENG received the Ph.D. degree from Wuhan University, in 2008. She is currently working as an Associate Professor and the Master's Supervisor with Wuhan University of Science and Technology. Her research interests include intelligent perception and cooperative control technologies for unmanned systems and new energy and intelligent connected vehicle technologies.

DONGXU SUN received the bachelor's degree from Shandong Jiaotong University, in 2022. He is currently pursuing the master's degree with Wuhan University of Science and Technology. His research interests include deep reinforcement learning and intelligent control.

LIU ZHAN received the bachelor's degree from Wuhan University of Science and Technology, in 2021, where she is currently pursuing the Ph.D. degree. Her research interests include machine learning and vehicle dynamics control.

XIAOWEI XU received the Ph.D. degree from Wuhan University of Technology, in 2013. He is currently working a Professor, a Doctoral Supervisor, and the Deputy Dean of the School of Automotive and Traffic Engineering, Wuhan University of Science and Technology. He also works as a Senior Engineer with China Society of Automotive Engineers and the Director of Hubei Society of Automotive Engineers. His research interests include new energy and intelligent connected vehicle technologies and intelligent perception and cooperative control technologies for unmanned systems.

JUNYI ZOU received the Ph.D. degree from Wuhan University of Technology, in 2019. He is currently working as an Associate Professor and the Master's Supervisor with Wuhan University of Science and Technology. His research interests include vehicle dynamics control, steer-by-wire chassis design, and intelligent connected vehicle technologies. He is a member of China Society of Automotive Engineers.