

Article

Adaptive Hybrid Control for Bridge Cranes Under Model Mismatch and Wind Disturbance

Yulong Qiu , Weimin Xu and Wangqiang Niu * 

Key Laboratory of Transport Industry of Marine Technology and Control Engineering, Shanghai Maritime University, Shanghai 201306, China; 202330510007@stu.shmtu.edu.cn (Y.Q.); wmxu@shmtu.edu.cn (W.X.)

* Correspondence: wqniu@shmtu.edu.cn

Abstract

Addressing the challenge of balancing high-precision positioning with strict safety constraints for underactuated bridge cranes subject to model parameter mismatch and stochastic wind disturbances, an adaptive hybrid control framework is presented integrating a Safety-Aware Dynamic Gain Sliding Mode Controller (DG-SMC) with a TD3-based residual deep reinforcement learning network. By designing a gain scheduling mechanism based on swing angle amplitude, the proposed method physically limits trolley acceleration to strictly constrain the payload swing angle within a safe range ($\pm 7^\circ$). Simultaneously, a TD3 agent is introduced as a residual compensator to adaptively learn system dynamics through environmental interaction, generating real-time compensatory control forces to counteract unmodeled dynamics arising from system parameter deviations and continuous wind resistance. Numerical simulations demonstrate that, under conditions involving payload mass deviations of up to 25% and stochastic wind disturbances, the proposed control method effectively reduces steady-state positioning errors, suppresses payload swing during operation, and significantly enhances the system's energy dissipation efficiency and global robustness in uncertain environments.

Keywords: bridge crane; model mismatch; wind disturbance rejection; adaptive hybrid control; deep reinforcement learning; sliding mode control

1. Introduction

Bridge cranes act as pivotal components in modern heavy industries, encompassing sectors such as logistics, construction, and port transportation [1,2]. Characterized as typical underactuated mechanical systems [3], bridge cranes possess fewer control inputs (i.e., trolley driving force) than degrees of freedom (i.e., trolley position and payload swing angle). The fundamental control objective is to achieve rapid and precise payload transportation to a target position while simultaneously suppressing payload swing. However, fulfilling this task is challenging due to the system's inherent nonlinearities, exacerbated by inevitable external disturbances (e.g., wind disturbances) [4,5] and internal parameter uncertainties (e.g., variations in payload mass and cable length) [6,7]. Excessive payload swing not only degrades operational efficiency but also poses severe safety hazards to both personnel and equipment.

To address these challenges, numerous control strategies have been proposed [8,9]. Conventional linear control approaches, such as Proportional–Integral–Derivative (PID) control and Linear Quadratic Regulator (LQR), have been widely adopted due to their structural simplicity [10–12]. However, relying on linearized models around equilibrium points,



Academic Editor: Wei Gao

Received: 23 December 2025

Revised: 8 February 2026

Accepted: 10 February 2026

Published: 12 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

these methods often suffer from significant performance degradation when the system undergoes large transient fluctuations or encounters strong external disturbances [13–15]. To handle system nonlinearities and uncertainties, robust control techniques—particularly Sliding Mode Control (SMC)—have garnered considerable attention [6,16]. SMC is renowned for its invariance to matched uncertainties and its capability to force system states to slide along a predefined surface. Despite the theoretical robustness of conventional SMC, it often faces limitations in practical crane applications: (1) it typically requires prior knowledge of the upper bounds of uncertainties, and inaccurate parameters (e.g., unknown payload mass) may lead to steady-state errors; (2) high switching gains are usually employed to suppress external disturbances, thereby inducing the chattering phenomenon [4]; (3) it is difficult to explicitly incorporate hard safety constraints (such as strict limits on the maximum swing angle) directly into the control law [17].

In recent years, Deep Reinforcement Learning (DRL) has emerged as a promising data-driven approach for solving complex control problems. Algorithms such as Deep Deterministic Policy Gradient (DDPG) [18] and Twin Delayed DDPG (TD3) [19] enable agents to learn optimal control policies through interaction with the environment, without the need for explicit mathematical models. While DRL has exhibited superior performance in handling nonlinearities and adaptability, applying DRL to safety-critical systems—such as cranes—remains problematic. Pure DRL control typically necessitates extensive trial-and-error exploration, which may lead to dangerous states (e.g., excessive swing angles) during the training phase [20]. Furthermore, compared with model-based controllers, DRL policies often lack theoretical stability guarantees and interpretability [21].

In light of the respective strengths and limitations of model-based and data-driven control methods, hybrid control architectures have emerged as a promising solution. The core philosophy is to synergize the stability guarantees of conventional controllers with the adaptability of learning algorithms [22]. Often referred to as residual control or compensatory control, this approach employs a traditional controller to establish baseline performance, while leveraging a learning agent to compensate for model mismatches and external disturbances [23,24].

This paper proposes a robust hybrid control framework that integrates a Dynamic Gain Sliding Mode Controller (DG-SMC) with a TD3-based Deep Reinforcement Learning residual compensator. The specific contributions and innovations are summarized as follows:

- (1) **Safety-Aware Dynamic Gain Scheduling:** Distinct from conventional SMC strategies, a novel gain scheduling mechanism is designed within the SMC law. By introducing a swing-angle-based suppression factor, the controller dynamically attenuates the position tracking gain while boosting the anti-swing gain when the swing angle is substantial. This mechanism prioritizes system safety over positioning speed during transient disturbances, thereby physically constraining the swing amplitude.
- (2) **Residual Compensation for Model Mismatch:** Parameter uncertainty is addressed by employing the SMC as a baseline controller operating on nominal parameters. A TD3 agent is trained to observe state deviations arising from discrepancies between the nominal model and the physical environment (e.g., varying payload masses and friction). Subsequently, the agent generates compensatory control actions to eliminate steady-state errors and suppress continuous external wind disturbances.
- (3) **Frequency-Domain Decoupling of Control Tasks:** The proposed architecture achieves a decoupling of control objectives in the frequency domain. The high-gain SMC component dominates the transient phase, providing rapid response to suppress high-frequency oscillations and prevent the violation of maximum swing angle constraints. Conversely, the TD3 agent dominates the steady-state phase, offering precise low-

frequency compensation against persistent wind disturbances and parameter drifts. This complementary mechanism reconciles the inherent conflict between the high gain required for disturbance rejection and the gentle actions necessary for precise positioning.

2. Model Establishment

The gantry crane primarily consists of a trolley, a cable, and a payload. As illustrated in Figure 1, the system is modeled in a two-dimensional plane. In this model, M and m denote the masses of the trolley and the payload, respectively. Furthermore, $x(t)$ represents the displacement of the trolley, l is the length of the cable connecting the trolley to the payload, $\theta(t)$ denotes the payload swing angle, and F_x represents the lateral force applied to the trolley, F_{wind} denotes the lateral wind disturbance force.

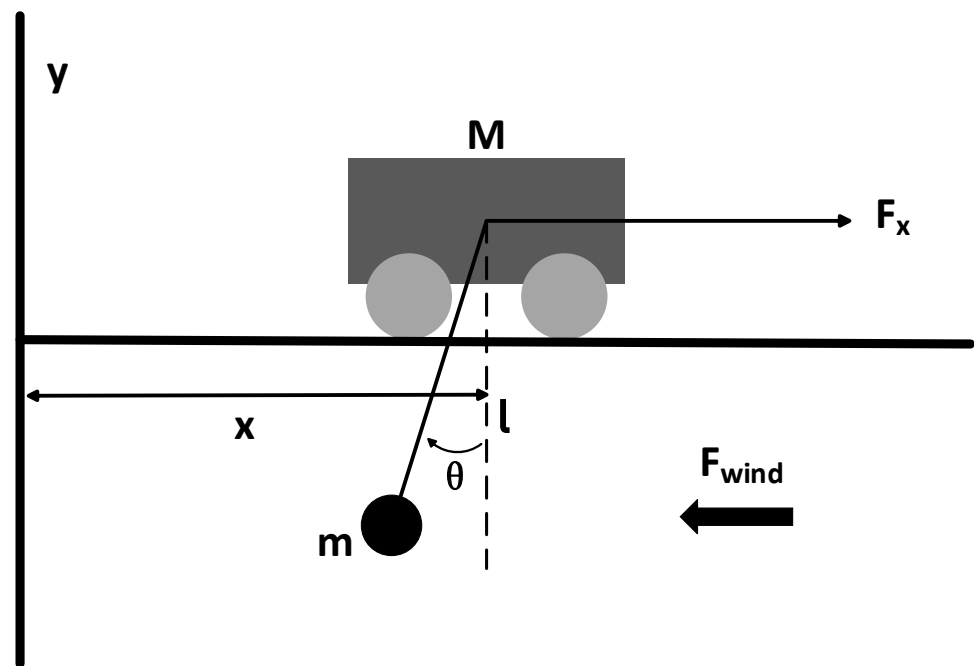


Figure 1. Two-dimensional model diagram of the bridge crane.

The Payload position (x_m, y_m) in the Cartesian coordinate system can be expressed as

$$\begin{cases} x_m = x + l \sin \theta \\ y_m = -l \cos \theta \end{cases} \quad (1)$$

Taking the time derivative yields the velocity components of the payload

$$\begin{cases} \dot{x}_m = \dot{x} + l\dot{\theta} \cos \theta \\ \dot{y}_m = l\dot{\theta} \sin \theta \end{cases} \quad (2)$$

The total kinetic energy of the system is composed of the kinetic energies of the trolley and the payload as follows:

$$T = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}m(\dot{x}_m^2 + \dot{y}_m^2) \quad (3)$$

Expanding this expression yields

$$\dot{x}_m^2 + \dot{y}_m^2 = \left(\dot{x}^2 + 2l\dot{x}\dot{\theta}\cos\theta + l^2\dot{\theta}^2\cos^2\theta \right) + \left(l^2\dot{\theta}^2\sin^2\theta \right) = \dot{x}^2 + l^2\dot{\theta}^2 + 2l\dot{x}\dot{\theta}\cos\theta$$

Substituting it into T yields

$$T = \frac{1}{2}(M + m)\dot{x}^2 + \frac{1}{2}ml^2\dot{\theta}^2 + ml\dot{x}\dot{\theta}\cos\theta \quad (4)$$

Taking the horizontal plane of the trolley as the zero potential energy reference, the total potential energy of the system is

$$V = mgy_m = -mgl\cos\theta \quad (5)$$

The Lagrange function of the system is

$$L = T - V = \frac{1}{2}(M + m)\dot{x}^2 + \frac{1}{2}ml^2\dot{\theta}^2 + ml\dot{x}\dot{\theta}\cos\theta + mgl\cos\theta \quad (6)$$

The Euler–Lagrange equations are obtained as

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = Q_i \quad (7)$$

where q_i denotes the generalized coordinates $\{x, \theta\}$, and Q_i represents the corresponding generalized forces.

Calculating the Lagrange equation for the generalized coordinate x yields

$$\frac{\partial L}{\partial \dot{x}} = (M + m)\dot{x} + ml\dot{\theta}\cos\theta \quad (8)$$

$$\frac{\partial L}{\partial x} = 0 \quad (9)$$

Differentiating with respect to time yields

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = (M + m)\ddot{x} + ml\ddot{\theta}\cos\theta - ml\dot{\theta}^2\sin\theta \quad (10)$$

The generalized force Q_i includes the control force F , the friction force $-b\dot{x}$, and the horizontal wind force F_{wind} as follows:

$$Q_x = F - b\dot{x} + F_{wind} \quad (11)$$

Combining the above equations yields

$$(M + m)\ddot{x} + (ml\cos\theta)\ddot{\theta} = F - b\dot{x} + ml\dot{\theta}^2\sin\theta + F_{wind} \quad (12)$$

The Lagrange equation with respect to the θ -coordinate is calculated as

$$\frac{\partial L}{\partial \dot{\theta}} = ml^2\dot{\theta} + ml\dot{x}\cos\theta \quad (13)$$

$$\frac{\partial L}{\partial \theta} = -ml\dot{x}\dot{\theta}\sin\theta - mgl\sin\theta \quad (14)$$

Differentiating with respect to time yields

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = ml^2 \ddot{\theta} + ml \ddot{x} \cos \theta - ml \dot{x} \dot{\theta} \sin \theta \quad (15)$$

The generalized force Q_θ represents the torque generated by the horizontal wind force F_{wind} as follows:

$$Q_\theta = F_{wind} \cdot (l \cos \theta) \quad (16)$$

Combining the above equations yields

$$(ml \cos \theta) \ddot{x} + (ml^2) \ddot{\theta} = -mgl \sin \theta + F_{wind} l \cos \theta \quad (17)$$

The dynamic equations of the system can be written in matrix form as

$$\begin{bmatrix} M + m & ml \cos \theta \\ ml \cos \theta & ml^2 \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} F - b\dot{x} + ml\dot{\theta}^2 \sin \theta + F_{wind} \\ -mgl \sin \theta + F_{wind} l \cos \theta \end{bmatrix} \quad (18)$$

3. Formulation of the Hybrid Control Strategy

3.1. Safety-Aware Dynamic Gain Sliding Mode Controller (DG-SMC)

To guarantee the fundamental stability of the underactuated bridge crane system, a sliding mode controller is formulated based on the system's dynamic model. Distinct from conventional SMC controllers with fixed gains, this paper proposes a safety-aware dynamic gain scheduling mechanism to reconcile the trade-off between trolley positioning speed and payload swing suppression.

The objective of the control system is to simultaneously achieve the convergence of the trolley position error $e_x = x_d - x$ and the payload swing angle θ to zero. The composite sliding surface s is defined as a linear combination of the system error states

$$s = \dot{e}_x + k_1 e_x + k_2 \theta \quad (19)$$

where k_1 and k_2 denote the position tracking gain and the swing suppression gain, respectively.

A critical limitation of conventional SMC controllers lies in the fact that a fixed position tracking gain k_1 may still generate significant acceleration even when the payload swing is at a dangerous amplitude, thereby exacerbating the instability of the swing angle. To address this issue, a nonlinear gain scheduling strategy driven by the swing angle amplitude $|\theta|$ is introduced.

To prevent over-actuating the system during large-angle swings—which would further amplify the payload swing—the position tracking gain k_1 is designed to decay exponentially as the swing angle increases: $k_1 = K_{p,norm} \cdot \exp(-\alpha \theta^2)$.

Where $K_{p,norm}$ represents the nominal position tracking gain, and α is the sensitivity coefficient. As $|\theta|$ increases, $k_1 \rightarrow 0$ holds. This behavior effectively suppresses the transportation task to dampen the payload swing, thereby prioritizing the safety of the system operation.

Conversely, to enhance system damping during unstable states, the swing suppression gain k_2 is linearly boosted: $k_2 = K_{\theta,norm} \cdot (1 + \beta |\theta|)$.

Where $K_{\theta,norm}$ denotes the nominal swing suppression gain, and β represents the enhancement factor.

This strategy ensures that as the system approaches the boundaries of safety constraints, the controller automatically shifts its priority from trolley transportation to stabilizing the payload swing angle.

Differentiating the sliding surface with respect to time yields

$$\dot{s} = \ddot{e}_x + \frac{d}{dt}(k_1 e_x) - \frac{d}{dt}(k_2 \theta) \quad (20)$$

Assuming k_1 and k_2 are slowly varying parameters, their time derivatives are ignored to simplify the calculation. Given that the target acceleration $\ddot{x}_d = 0$ (which implies $\ddot{e}_x = \ddot{x}_d - \ddot{x} = -\ddot{x}$), we have

$$\dot{s} \approx -\ddot{x} + k_1 \dot{e}_x - k_2 \dot{\theta} \quad (21)$$

Setting $\dot{s} = 0$ and substituting the system dynamics yields the equivalent control u_{eq} :

$$u_{eq} = M(k_1 \dot{e}_x - k_2 \dot{\theta}) + b\dot{x} - mg\theta \quad (22)$$

To ensure that the system states can reach the sliding surface from any initial position within a finite time, the exponential reaching law is adopted:

$$\dot{s} = -\eta \operatorname{sgn}(s) - ks \quad (23)$$

Consequently, the corresponding switching control force u_{sw} is derived as

$$u_{sw} = M(\eta \operatorname{sgn}(s) + ks) \quad (24)$$

By summing the calculated components, the final control law is obtained as

$$F = M \left[\eta \operatorname{sgn}(s) + ks + \left(k_1 \dot{e}_x - k_2 \dot{\theta} \right) + \frac{b}{M} \dot{x} - \frac{mg}{M} \theta \right] \quad (25)$$

Define a Lyapunov candidate function as $V = \frac{1}{2}s^2$. For $s \neq 0$, $V(s) > 0$, which satisfies the positive definiteness condition.

Differentiating V with respect to time yields $\dot{V} = s\dot{s}$.

Substituting the control law into the equation results in $\dot{V} = -\eta|s| - ks^2$. Given that $\eta, k > 0$, it follows that $\dot{V} < 0$ for all $s \neq 0$. Consequently, the system is asymptotically stable with respect to the sliding surface. Moreover, the term $-\eta|s|$ guarantees the finite-time convergence characteristic of the system.

3.2. Design of a TD3-Based Intelligent Residual Compensator

Although the Safety-Aware DG-SMC guarantees the baseline stability of the system operation, its performance relies on nominal system parameters, rendering it incapable of eliminating steady-state errors induced by model mismatches and external wind disturbances. To address this limitation, a model-free Deep Reinforcement Learning (DRL) agent based on the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm is introduced. The working mechanism of the TD3 agent is shown in Figure 2.

The objective of the residual compensator is to learn an optimal policy π_ϕ that minimizes tracking errors and control costs in the presence of parameter uncertainties and external disturbances. This problem is formulated as an infinite-horizon discounted Markov Decision Process (MDP), characterized by the tuple $\mathcal{M} = \langle S, A, P, R, \gamma \rangle$. S denotes the states space representing the bridge crane's kinematics; A is the continuous action space for the residual control force; P represents the state transition probability governing the system dynamics; R is the reward function guiding the agent; and γ is the discount factor determining the priority of future rewards.

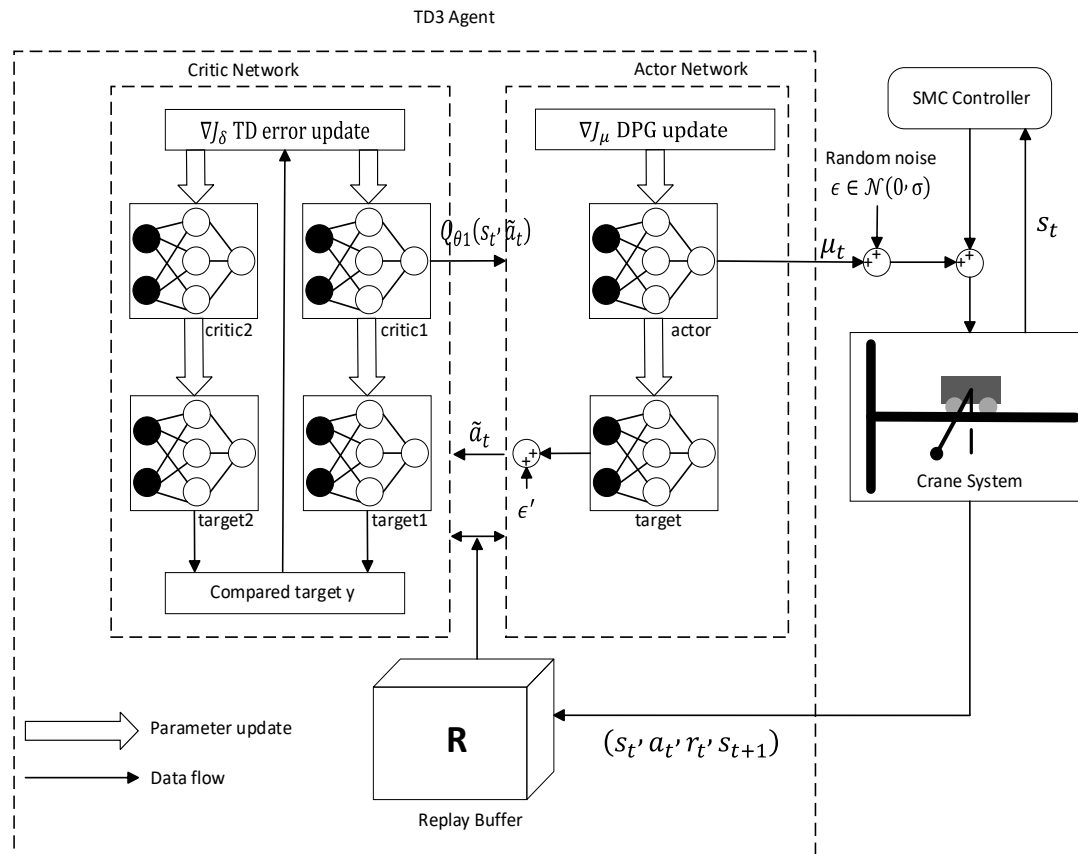


Figure 2. Schematic diagram of the TD3-based residual learning architecture.

The state space $S \subset \mathbb{R}^4$ encompasses the various kinematic variables of the system. To mitigate the adverse effects of dimensional discrepancies among different physical quantities on neural network convergence, the state vector s_t is processed using standard Z-score normalization as follows:

$$s_t = \text{Norm} \left(\begin{bmatrix} x(t), \dot{x}(t), \theta(t), \dot{\theta}(t) \end{bmatrix}^T \right) \quad (26)$$

Let $u_{norm} : S \rightarrow \mathbb{R}$ denote the nominal control law of the DG-SMC. The TD3 agent learns a residual mapping $\pi_\phi : S \rightarrow [-1, 1]$. The final composite control input $u(t)$ applied to the system is synthesized via a weighted superposition of both components as follows:

$$u(t) = \text{clip}(u_{norm}(s_t) + \lambda \cdot F_{max} \cdot \pi_\phi(s_t), -F_{max}, F_{max}) \quad (27)$$

where $\lambda \in (0, 1]$ serves as the agent control authority factor. By setting λ , the exploration space of the agent is constrained, ensuring that its learnable component acts solely as a local correction to the nominal law; $\text{clip}(\cdot)$ is the saturation function; $u_{norm}(s_t)$ represents the nominal control law, corresponding to the control force calculated by the DG-SMC; while $F_{max} \cdot \pi_\phi(s_t)$ constitutes the intelligent compensation term generated by the agent. This mechanism prevents the agent from compromising the global stability guarantee provided by the DG-SMC.

The design of the reward function $R(s_t, a_t)$ aims to minimize system energy dissipation and ensure constraint satisfaction. This function comprises three distinct reward sub-terms:

$$r = r_{track}(s_t) + r_{barrier}(s_t) + r_{sparse}(s_t) \quad (28)$$

where $r_{track}(s_t)$, $r_{barrier}(s_t)$, and $r_{sparse}(s_t)$ represent the tracking accuracy reward, the safety constraint reward, and the terminal convergence reward, respectively.

The term $r_{track}(s_t)$ is a quadratic cost function used to penalize state deviations from the equilibrium point:

$$r_{track}(s_t) = -\left(s_t^T W s_t\right) \quad (29)$$

where $W = \text{diag}(w_x, 0, w_\theta, w_\theta)$. In light of the underactuated nature of the crane system, we set $w_\theta \gg w_x$ to prioritize the suppression of the swing angle.

To incorporate the swing angle constraint $|\theta| \leq \theta_{max}$ into the policy, a discrete barrier penalty function is constructed:

$$r_{barrier}(s_t) = -\sum_{i=1}^N p_i \cdot \mathbb{I}(|\theta(t)| > \zeta_i) \quad (30)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, $\{\zeta_i\}$ represents the hierarchical thresholds, and $\{p_i\}$ denotes the corresponding penalty values.

To eliminate steady-state errors induced by friction, a sparse reward based on the Radial Basis Function (RBF) is applied near the target:

$$r_{sparse}(s_t) = \beta \cdot \exp\left(-\frac{\|x - x_d\|^2}{2\sigma^2}\right) \cdot \mathbb{I}(s_t \in \Omega_{goal}) \quad (31)$$

where Ω_{goal} defines a high-precision neighborhood; β is a scaling coefficient that determines the peak amplitude of the sparse reward. Tuning β can regulate the incentive level for the agent to eliminate residual steady-state errors within the target neighborhood. This term effectively sharpens the policy gradient near the equilibrium point, facilitating the precise positioning of the system.

The core task of the TD3 residual compensator is to search for the optimal policy network parameters ϕ^* within the parameter space, such that the cumulative reward obtained during the interaction with the environment is maximized. The optimization objective function $J(\phi)$ is defined as follows:

$$J(\phi) = \mathbb{E}_{\tau \sim \pi_\phi} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] \quad (32)$$

where $\tau = (s_0, a_0, s_1, a_1, \dots)$ denotes the state-action trajectory generated under the current residual policy π_ϕ , $\mathbb{E}_{\tau \sim \pi_\phi}$ denotes the expectation over trajectories τ generated by the policy π_ϕ , ensuring that the agent optimizes the average performance under stochastic wind disturbances rather than a specific deterministic episode, and $\gamma \in [0, 1)$ represents the discount factor.

In the context of the current system, maximizing $J(\phi)$ is equivalent to seeking a control law that minimizes the cumulative tracking error and maximizes the rate of payload swing energy dissipation during motion, while strictly adhering to the system's safety constraints. The introduction of the discount factor γ ensures that the optimization process balances the immediate rewards of current actions with the long-term stability of the system.

The TD3 agent adopts an Actor-Critic architecture for solution derivation. A Critic network $Q_\psi(s, a)$ is introduced to approximate the state-action value function. The network parameters ψ are updated by minimizing the Temporal Difference (TD) error, with the loss function defined as

$$L(\psi) = \mathbb{E}_{(s, a, r, s') \sim D} \left[(y - Q_\psi(s, a))^2 \right] \quad (33)$$

where $y = r + \gamma_{i=1,2}^{min} Q_{\psi'_i}(s', \pi_{\phi'}(s') + \epsilon)$ represents the target value incorporating clipped double Q-learning and target policy smoothing, and D denotes the experience replay buffer. By minimizing the Bellman residual, the Critic network can accurately evaluate the long-term efficacy of the current residual policy in suppressing swing and eliminating steady-state errors.

The update of the Actor network π_{ϕ} aims to maximize the Q-value estimated by the Critic. According to the Deterministic Policy Gradient Theorem, the gradient of the objective function $J(\phi)$ with respect to the policy parameters ϕ can be approximated as

$$\nabla_{\phi} J(\phi) \approx \mathbb{E}_{S \sim D} \left[\nabla_a Q_{\psi}(s, a) \big|_{a=\pi_{\phi}(s)} \cdot \nabla_{\phi} \pi_{\phi}(s) \right] \quad (34)$$

This expression utilizes the chain rule to backpropagate the gradient of the Q-value with respect to the action to the policy network, the term $\nabla_a Q_{\psi}$ indicates the direction in the action space that maximizes the estimated value, while $\nabla_{\phi} \pi_{\phi}$ represents the sensitivity of the policy output to parameter changes. This mechanism enables the Actor network to adjust the residual output along the direction of value function ascent, thereby learning the optimal control force capable of actively compensating for model mismatches and counteracting external wind disturbances.

3.3. Hybrid Control Architecture

To synergize the stability of model-based control with the adaptability of data-driven learning, a parallel hybrid control architecture is proposed. Control authority is allocated between the Dynamic Gain Sliding Mode Controller (DG-SMC) and the TD3 agent. The final control input $u(t)$ applied to the bridge crane system is expressed as

$$u(t) = \text{clip}(u_{\text{norm}}(s_t) + \lambda \cdot F_{\text{max}} \cdot \pi_{\phi}(s_t), -F_{\text{max}}, F_{\text{max}}) \quad (35)$$

where $\lambda \in (0, 1]$ serves as the control authority factor for the agent. By configuring λ , the exploration space of the agent is constrained, ensuring that its learnable component functions solely as a local correction to the nominal control law. This mechanism prevents the agent from compromising the global stability guarantee provided by the DG-SMC. Additionally, $\text{clip}(\cdot)$ denotes the saturation function, and F_{max} represents the maximum output limit of the actuator. The DG-SMC operates based on fixed physical laws. Its primary function is to guarantee that the system states remain within a bounded stable region and to ensure rapid convergence during the initial transient phase. Specifically, the dynamic gain scheduling mechanism fundamentally constrains the trolley acceleration when the swing angle approaches its safety limit.

The TD3 agent detects discrepancies between the nominal system model and the actual dynamics (subject to parameter mismatches and external wind disturbances). By injecting precise compensatory forces, the agent eliminates steady-state errors—which the DG-SMC with fixed nominal parameters fails to address—and suppresses external disturbances.

To ensure the robustness of the learned policy against parameter uncertainties, a domain randomization strategy is employed during the training phase of the TD3 agent. Specifically, the following is completed for each training episode:

- (1) Controller Setup: The DG-SMC is initialized with fixed nominal parameters.
- (2) Environment Randomization: Physical environment parameters are sampled from uniform distributions to simulate model mismatches.
- (3) External Disturbances: Random wind forces containing both mean components and Gaussian noise are continuously applied.

By training in highly variable environments while keeping the baseline controller fixed, the TD3 agent learns a generalized policy capable of identifying and compensating

for diverse system dynamics, rather than overfitting to a specific set of parameters. The hybrid control architecture is shown in Figure 3.

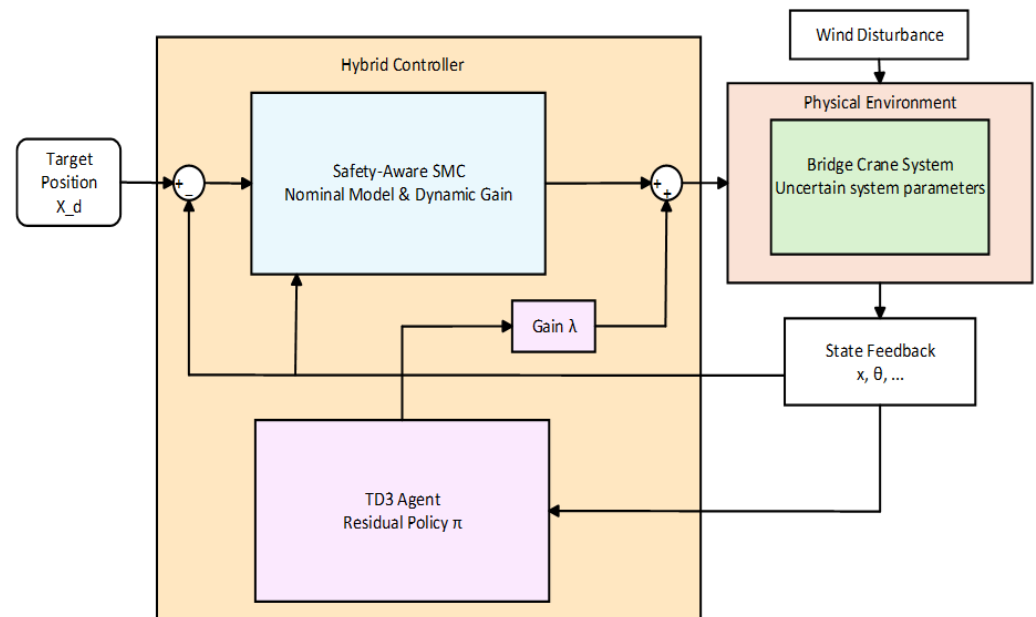


Figure 3. Block diagram of the proposed adaptive hybrid control framework.

4. Simulation Experiments and Results Analysis

This section aims to comprehensively validate the effectiveness of the proposed hybrid control strategy for underactuated bridge crane systems through a series of simulation experiments. The experimental design focuses on evaluating the comprehensive performance of the control system under conditions of model parameter mismatches and persistent stochastic external disturbances. The discussion unfolds from three perspectives: a comparative analysis of transient and steady-state performance during system operation, and an assessment of global stability on the phase plane.

4.1. Simulation Setup

The simulation environment is constructed based on the nonlinear coupled dynamic equations derived in Chapter 2. The fourth-order Runge–Kutta method (with a time step of $dt = 0.02$ s) is employed for numerical integration to ensure the precision of the physical simulation. To simulate authentic industrial scenarios and examine the algorithm’s generalization capability, significant model uncertainties are introduced between the controller and the physical environment:

The Dynamic Gain Sliding Mode Controller (DG-SMC) is designed using fixed nominal parameters (payload mass $m_{norm} = 1.0$ kg, friction coefficient $b_{norm} = 0.1$). However, in the testing environment, the actual physical parameters are set to deviate significantly from the nominal values. Under these conditions, the controller severely underestimates the system’s inertia and friction. Furthermore, the maximum control force is constrained to $F_{max} = 5$ N. To evaluate the disturbance rejection capability of the proposed method, persistent stochastic wind disturbances F_{wind} are applied to the payload, simulating external wind environments encountered in outdoor operations. Regarding the TD3 agent, both the Actor and Critic networks utilize a dual hidden layer fully connected structure with [400, 300] units. The Rectified Linear Unit (ReLU) is employed as the activation function, while the output layer uses the Tanh function. The Adam optimizer is selected with a learning rate of 3×10^{-4} . The discount factor is set to $\gamma = 0.99$, and the soft update coefficient is $\tau = 0.005$.

To incentivize the agent to prioritize payload swing suppression over rapid trolley positioning, the weight coefficients in the quadratic reward term were determined through empirical tuning. Specifically, the weights were set as $w_x = 0.5$, $w_\theta = 40$ and $w_{\dot{\theta}} = 0.1$. This configuration ensures that $w_\theta \gg w_x$, thereby enforcing effective suppression of the payload swing. Sensitivity analyses indicated that a lower w_θ resulted in the swing angle exceeding safety limits, whereas an excessively high w_θ led to overly conservative system behavior, prolonging the transportation time.

4.2. Performance at a Heavier Load and a Weaker Wind Disturbance

Two cases are reported here, one with a heavier load, $m = 1.5$ kg, and a weaker wind disturbance, $F_{wind} \approx 0.1$ N, Figure 4, and another with a lighter load, $m = 1.25$ kg, and a stronger wind disturbance, $F_{wind} \approx 0.3$ N, Figures 5 and 6.

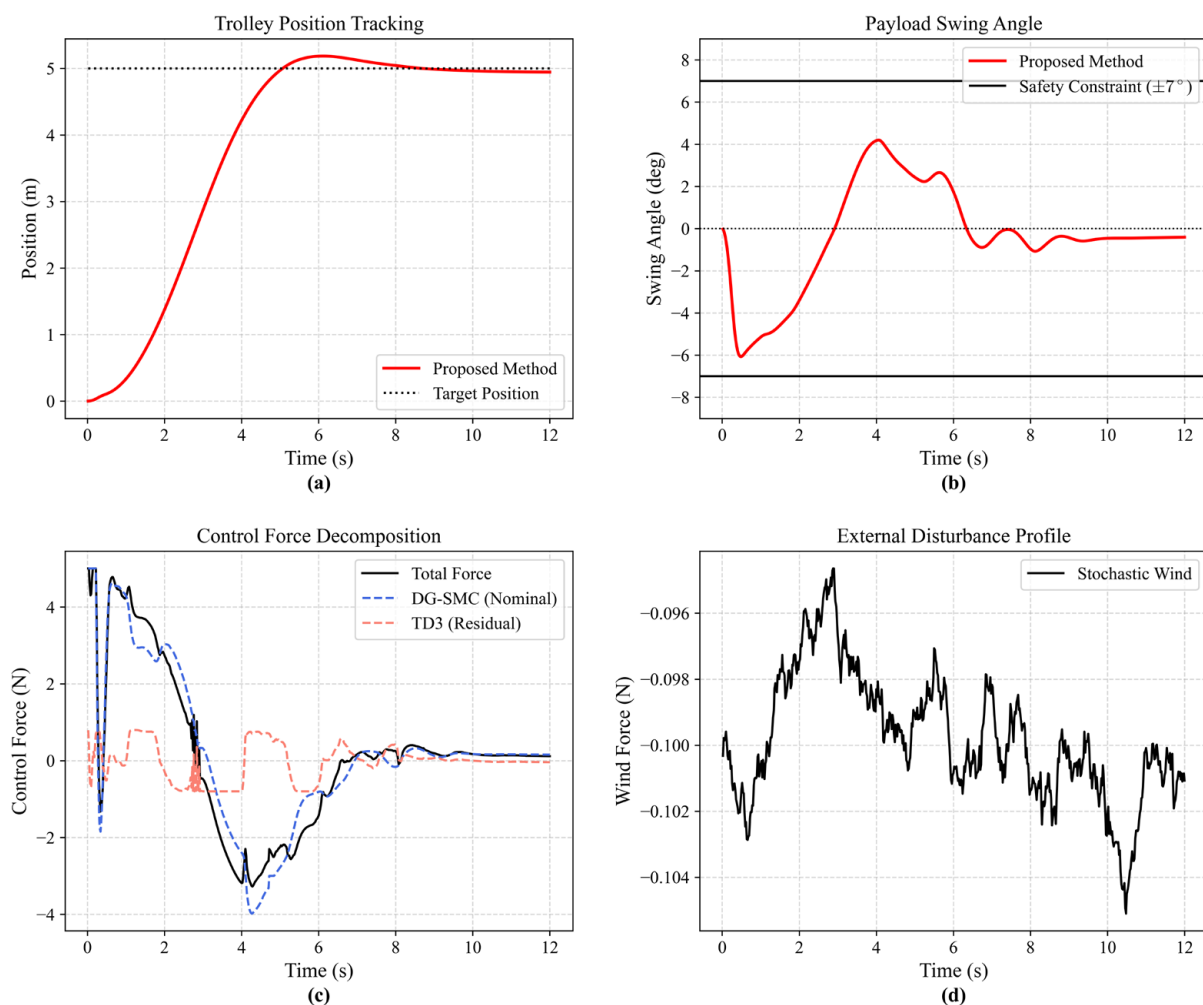


Figure 4. Detailed performance analysis and control force decomposition of the proposed hybrid strategy under a weak wind disturbance. (a) Trolley position tracking trajectory converging to the target; (b) payload swing angle response; (c) time histories of control force inputs; (d) profile of the applied wind disturbance.

Figure 4 illustrates the operational performance of the proposed control method under the actual system parameters of $m = 1.5$ kg, $b = 0.15$, and a wind disturbance of $F_{wind} \approx 0.1$ N. The simulation results demonstrate that under persistent stochastic wind disturbances with a mean of approximately -0.1 N, the proposed hybrid control strategy exhibits a superior comprehensive performance. Specifically, the position tracking curve is smooth and rapidly converges to the target, achieving zero steady-state error and demonstrating exceptional disturbance rejection stiffness. Throughout the dynamic process, the payload swing angle is strictly confined within the safety range of $\pm 7^\circ$ (with a maximum transient amplitude of approximately -6°). Furthermore, the steady-state residual swing angle effectively counteracts the wind impact, which aligns with physical principles. The decomposition of control forces clearly elucidates the synergistic mechanism: the DG-SMC maintains baseline stability, while the TD3 agent actively suppresses unmatched disturbances by outputting precise residual signals. This validates the method's distinct advantages in achieving high-precision positioning, guaranteeing strict safety constraints, and ensuring strong robustness in uncertain environments.

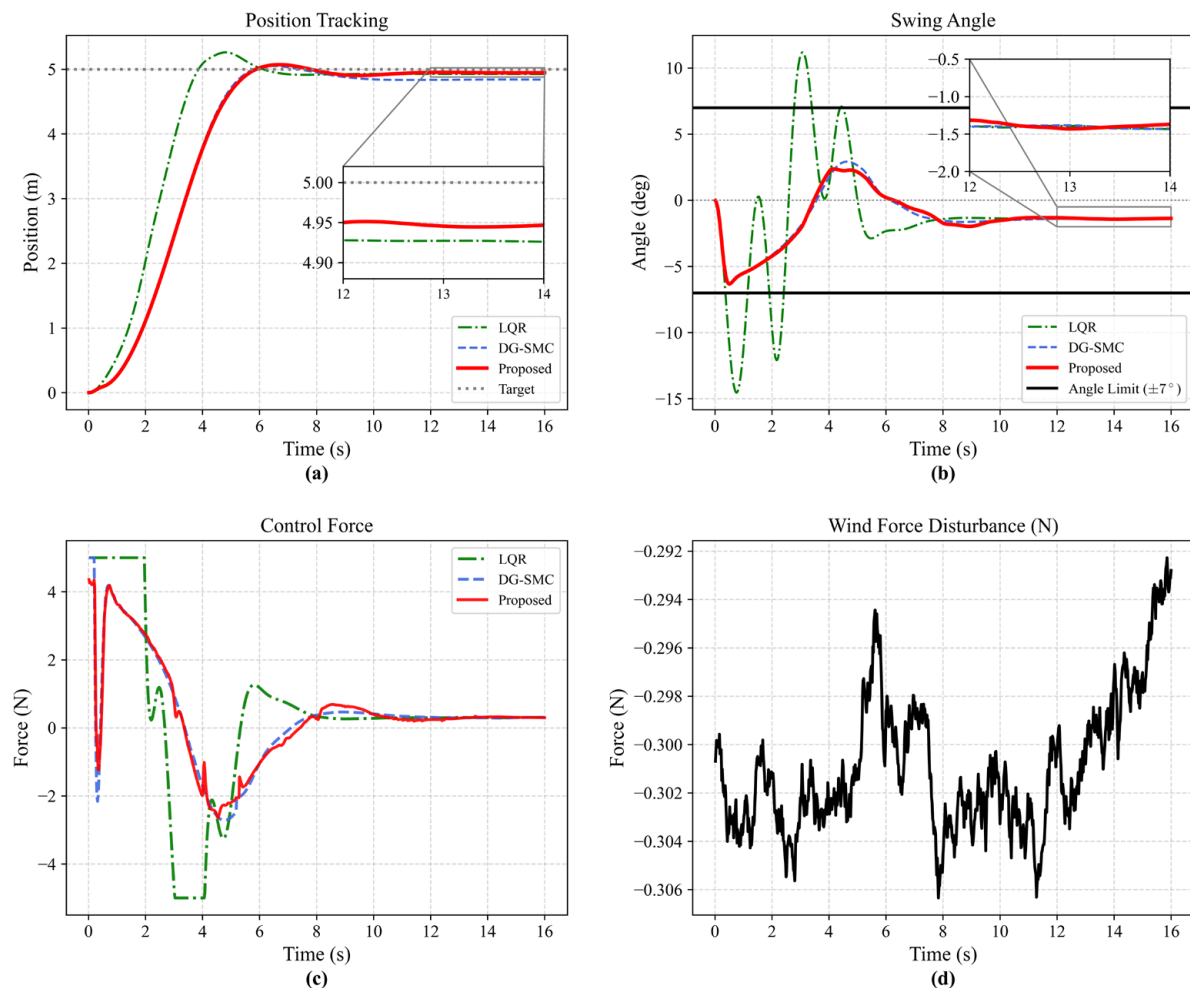


Figure 5. Comparison of system responses among LQR, DG-SMC, and the proposed hybrid method under strong wind disturbance. (a) Trolley position tracking trajectory; (b) payload swing angle response; (c) time histories of control force inputs; (d) profile of the applied strong wind disturbance.

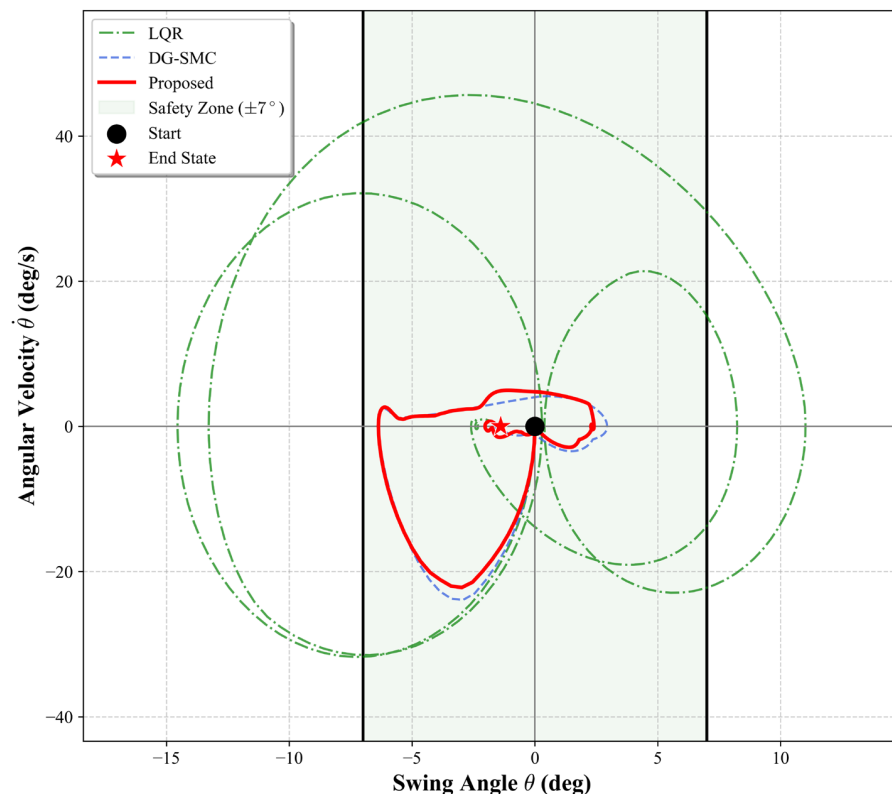


Figure 6. Comparison of state trajectories on the phase plane with safety constraints under strong wind disturbance.

4.3. Performance at a Lighter Load and a Stronger Wind Disturbance

To objectively assess control performance, comparative experiments were conducted between the proposed method and two classical control strategies. The testing conditions were uniformly set as follows: actual mass $m = 1.25$ kg, friction coefficient $b = 0.2$, wind force $F_{wind} \approx 0.3$ N, and the maximum permissible system swing angle is 7° . Figure 5 illustrates a detailed comparison of the operational performance among the three methods.

As observed in Figure 5, regarding the trolley position response, the LQR method exhibits significant overshoot due to the lack of compensation mechanisms for nonlinear friction and wind disturbances. Although the DG-SMC method demonstrates a rapid response with no overshoot, the localized magnified view clearly reveals that it eventually settles near 4.9 m, presenting a substantial steady-state error. This occurs because the control force calculated by the DG-SMC, based on nominal mass parameters, is insufficient to overcome the actual system resistance and external wind disturbances. In contrast, the proposed hybrid method exhibits superior tracking performance, remaining closest to the target value during the steady-state phase. The residual error is primarily attributed to two factors: (1) the persistent stochastic nature of the external wind disturbance renders absolute mathematical static equilibrium unattainable; (2) the Actor network approximates the optimal compensation force with finite precision. Nevertheless, compared with the other two methods, the proposed approach reduces the error by over 45%, which falls well within the acceptable tolerance range for real-world industrial crane operations.

Regarding swing angle suppression and safety constraints, the proposed method successfully limits the maximum swing angle within the safety boundaries. Upon entering the steady-state phase, the residual swing angle stabilizes at approximately -1.5° . Due to the presence of persistent external wind, -1.5° represents the physical static equilibrium angle. The ability of the proposed method to rapidly converge to this physical limit demonstrates its exceptionally high efficiency in swing suppression and energy dissipation.

To explicitly quantify the contribution of the DRL component and justify its computational cost, a comparative ablation study was conducted between the baseline DG-SMC and the proposed hybrid strategy. Table 1 summarizes the key performance metrics under conditions of parameter mismatch and stochastic wind disturbances.

Table 1. Quantitative comparison of control performance metrics.

Metric	DG-SMC	Proposed	Improvement
Steady-State error (m)	0.1780	0.0623	65.0% ↓
Max Overshoot (m)	0.3518	0.3096	12.0% ↓
Settling Time (s)	11.86	10.58	10.8% ↓
RMSE (m)	2.1564	2.0982	2.7% ↓

As shown in Table 1, while the DG-SMC guarantees fundamental system stability, the introduction of the TD3 agent significantly enhances dynamic performance. Notably, the steady-state error is reduced by 65%, confirming the agent's capability to compensate for model mismatches and external wind disturbances. Furthermore, unlike traditional integral action, which often exacerbates overshoot, the learned policy intelligently optimizes the braking phase. This results in a 12.0% reduction in maximum overshoot and a shorter settling time (reduced by 1.28 s). These improvements clearly demonstrate the efficacy of the DRL module, verifying that it refines both transient response and steady-state precision.

To further validate the superiority of the proposed control algorithm from a state-space perspective, Figure 6 depicts the evolutionary trajectory of the payload swing angle θ versus its angular velocity $\dot{\theta}$ on the phase plane. In the figure, the vertical solid black lines indicate the safety constraint boundaries set at $\pm 7^\circ$.

It can be observed from the geometric envelope of the phase trajectories that the LQR controller exhibits severe divergence, with both its maximum positive and negative swing angles significantly breaching the safety boundaries of $\pm 7^\circ$. This indicates that under conditions of model mismatch and strong wind disturbances, the LQR controller—reliant on linearization assumptions—cannot guarantee the state safety of the nonlinear system. In contrast, the phase trajectories of both the standalone DG-SMC and the proposed method are strictly confined within the green shaded region. This demonstrates that the dynamic gain mechanism, based on barrier functions, effectively constrains the system, locking the state trajectories within the safe zone.

The rate at which the phase trajectory converges toward the origin reflects the energy dissipation efficiency of the system. The LQR controller exhibits underdamped characteristics, forming multiple large loops on the phase plane with a sluggish convergence process. Although the standalone DG-SMC ensures that safety limits are not exceeded, its trajectory presents considerable fluctuations during convergence, reflecting the controller's hesitation when dealing with parameter uncertainties. Conversely, the phase trajectory of the proposed method is the most compact; the curve is rapidly pulled down during the initial phase, avoiding deep undershoot, and subsequently contracts toward the equilibrium point via the fastest path. This rapid contraction attests that the TD3 agent, through learning, injects optimal equivalent damping into the system, thereby significantly enhancing energy dissipation efficiency.

It should be noted that due to the presence of the persistent -0.3 N wind disturbance, the system cannot converge to the absolute origin $(0, 0)$. The phase plane analysis compellingly demonstrates that, under severe uncertainties and external wind disturbances, the proposed hybrid control strategy not only provides deterministic safety guarantees at the physical level but also optimizes the control process through deep learning, achieving superior global stability compared to traditional methods.

4.4. Sensitivity Analysis Under Different Payload Masses and Cable Lengths

Sensitivity analysis under different payload masses and cable lengths is carried out using Monte Carlo simulations. The study comprised 500 independent runs. In each test episode, the physical parameters of the crane system were randomized to simulate the discrepancy between the nominal controller model and the actual physical plant. Specifically, the payload mass m was randomly sampled from a uniform distribution $U(0.7, 1.5)$ kg, while the cable length l varied within the range of $[0.70, 0.80]$ m. Furthermore, randomized friction coefficients and varying external wind disturbances were applied in each test.

The simulation results are shown in Figure 7. As illustrated in the figure, the control system maintains high precision despite system parameter uncertainties and persistent stochastic wind disturbances. The mean positioning error is recorded at 0.0791 m with a standard deviation of 0.0405 m, demonstrating that the TD3-based residual compensator effectively counteracts steady-state errors induced by model mismatches and external influences, thereby enhancing control accuracy. Across 500 experimental trials, the proposed strategy achieved a 100% success rate in swing suppression. The maximum observed swing angle was 6.74° , which is strictly confined within the safety limits. The histogram of swing angles exhibits a well-converged distribution with a sufficient safety margin between the peak angle and the safety boundary, confirming the reliability of the method. Furthermore, the scatter plot reveals that lighter payloads, due to their lower inertia, are more sensitive to wind disturbances and exhibit larger transient swing angles, whereas heavier payloads show the opposite trend. Despite these variations in physical characteristics, the proposed method successfully constrains the swing angle within the safety zone, demonstrating superior global stability and adaptability.

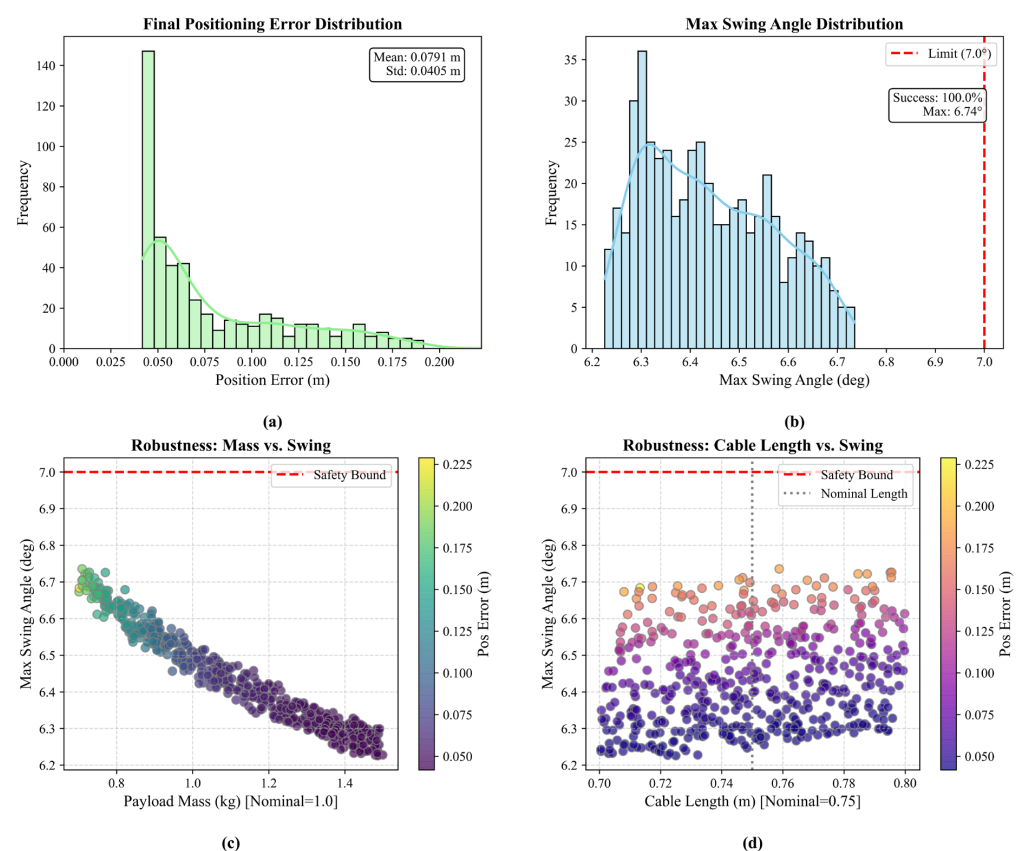


Figure 7. Statistical results of 500 Monte Carlo simulation runs under parameter uncertainty and stochastic wind disturbance. (a) Distribution of final positioning errors; (b) distribution of maximum swing angles; (c) robustness analysis scatter plot correlating payload mass with maximum swing angle; (d) robustness analysis scatter plot correlating cable length with maximum swing angle.

4.5. Performance at a Tighten Swing Angle of $\pm 5^\circ$

To further validate the performance and adaptability of the proposed control strategy, the maximum permissible swing angle was tightened to $\pm 5^\circ$, as illustrated in Figure 8. As observed, although the DG-SMC successfully confines the swing angle within the safety bounds, it exhibits significant positioning errors and fails to reach the designated target. In contrast, the proposed hybrid strategy demonstrates superior performance; it strictly constrains the swing angle within the $\pm 5^\circ$ range, while the TD3-based residual compensator counteracts dynamic losses. This enables the trolley to converge precisely to the target position, effectively eliminating steady-state errors.

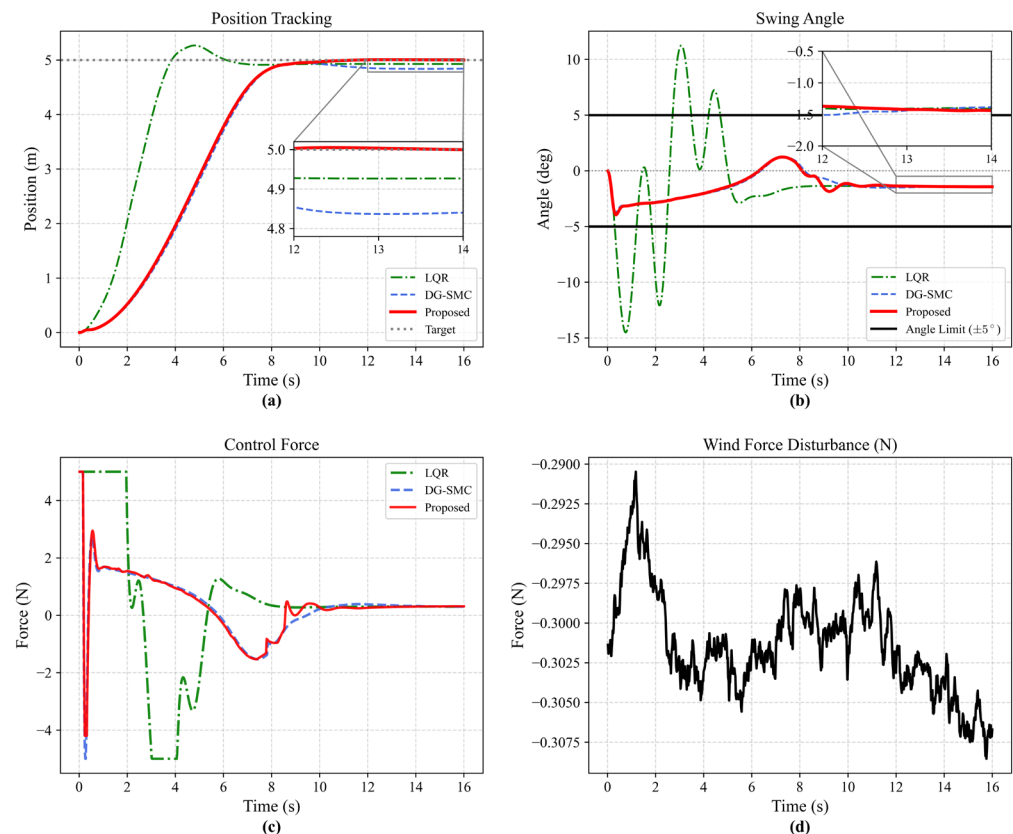


Figure 8. Sensitivity analysis results under a tightened safety constraint of $\pm 5^\circ$. (a) Trolley position tracking trajectory; (b) payload swing angle response; (c) time histories of control force inputs; (d) profile of the applied strong stochastic wind disturbance.

5. Conclusions

Targeting the objectives of high-precision positioning and strict swing suppression for underactuated bridge crane systems under parameter uncertainties and external disturbances, this paper proposes a robust hybrid control framework based on a Safety-Aware Dynamic Gain Sliding Mode Controller (DG-SMC) and the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. Through the design of a dynamic gain scheduling mechanism based on swing angle amplitude, trolley acceleration is physically constrained when approaching safety limits, thereby ensuring operational safety. This mechanism is synergized with a TD3 agent to achieve precise residual compensation of control forces in the presence of model mismatches and external wind disturbances. Leveraging the capability of deep neural networks to learn and map complex nonlinear dynamics, an adaptive mechanism is developed to update control parameters. This effectively counteracts steady-state errors caused by variations in payload mass and friction coefficients while suppressing persistent external wind disturbances.

Simulation results demonstrate that the proposed hybrid control method strictly confines payload swing within the safe range ($\pm 7^\circ$) under severe parameter variations, preventing the overshoot typically observed in traditional linear controllers. Furthermore, the method significantly improves steady-state positioning accuracy, converging to the target position with negligible residual fluctuations even under stochastic wind loads. Phase plane analysis further confirms that the control force variations are smooth without abrupt mutations, and the system exhibits high energy dissipation efficiency, with states rapidly contracting toward stability. In summary, the simulation results indicate that the proposed control method achieves satisfactory control performance, reliable stability, and effective adaptability to unmodeled dynamics. Future work will focus on validating the proposed method on a physical experimental testbed and extending it to control systems involving varying cable lengths and three-dimensional motion scenarios.

Author Contributions: All the authors discussed the idea, conducted the theoretical research, and formulated the problem. Conceptualization, W.X. and W.N.; methodology, Y.Q.; simulation, Y.Q.; writing—original draft, Y.Q.; writing—review and editing, W.N. and Y.Q.; revision—W.N., Y.Q. and W.X.; supervision, W.X. and W.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Capacity Building Program of Municipal Universities of Shanghai (20040501400).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abdel-Rahman, E.M.; Nayfeh, A.H.; Masoud, Z.N. Dynamics and Control of Cranes: A Review. *J. Vib. Control* **2003**, *9*, 863–908.
2. Ramli, L.; Mohamed, Z.; Abdullahi, A.M.; Jaafar, H.; Lazim, I.M. Control strategies for crane systems: A comprehensive review. *Mech. Syst. Signal Process.* **2017**, *95*, 1–23. [\[CrossRef\]](#)
3. Fantoni, I.; Lozano, R. *Non-Linear Control for Underactuated Mechanical Systems*; Springer: New York, NY, USA, 2002.
4. Sun, N.; Fang, Y.; Chen, H. A New Antiswing Control Method for Underactuated Cranes with Unmodeled Uncertainties: Theoretical Design and Hardware Experiments. *IEEE Trans. Ind. Electron.* **2015**, *62*, 453–465. [\[CrossRef\]](#)
5. Yu, Z.; Niu, W. Flatness-based backstepping antisway control of underactuated crane systems under wind disturbance. *Electronics* **2023**, *12*, 244. [\[CrossRef\]](#)
6. Tuan, L.A.; Lee, S.-G. Sliding mode controls of double-pendulum crane systems. *J. Mech. Sci. Technol.* **2013**, *27*, 1863–1873. [\[CrossRef\]](#)
7. Zhihong, M.; O'Day, M.; Yu, X. A Robust Adaptive Terminal Sliding Mode Control for Rigid Robotic Manipulators. *J. Intell. Robot. Syst.* **1999**, *24*, 23–41. [\[CrossRef\]](#)
8. Zhang, K.; Niu, W.; Zhang, K. Phase Plane Trajectory Planning for Double Pendulum Crane Anti-Sway Control. *Algorithms* **2025**, *18*, 246. [\[CrossRef\]](#)
9. Huang, W.; Niu, W.; Gu, W. Double-step Acceleration Input Shaping Anti-sway Control Based on Phase Plane Trajectory Planning. *Int. J. Control Autom. Syst.* **2024**, *22*, 419–429. [\[CrossRef\]](#)
10. Abdulhamid, I.B.; Muhammad, M.; Khaleel, A.I. *Control of a Double Pendulum Crane System Using PSO-Tuned LQR*; IEEE: Piscataway, NJ, USA, 2019. [\[CrossRef\]](#)
11. Huang, W.; Lu, B.; Weng, C.; Niu, W. Improved waveform command shaping anti-sway control for arbitrary initial and terminal conditions. *Mech. Syst. Signal Process.* **2025**, *241*, 113547. [\[CrossRef\]](#)
12. Solihin, M.I.; Akmeliawati, R.; Legowo, A. Robust feedback control design using PSO-based optimisation: A case study in gantry crane control. *Int. J. Mechatron. Autom.* **2011**, *1*, 121–131.
13. Chen, H.; Fang, Y.; Sun, N. Optimal trajectory planning and tracking control method for overhead cranes. *IET Control Theory Appl.* **2016**, *10*, 692–699. [\[CrossRef\]](#)

14. Huang, W.; Lu, B.; Weng, C.; Niu, W. Active disturbance rejection control with phase plane trajectory planning for overhead cranes. *J. Sound Vib.* **2025**, *621*, 119451. [[CrossRef](#)]
15. Huang, W.; Niu, W.; Zhou, X.; Gu, W. Anti-sway control of variable rope length container crane based on phase plane trajectory planning. *J. Vib. Control* **2024**, *30*, 1227–1240.
16. Ngo, Q.H. Adaptive sliding mode control of container cranes. *IET Control Theory Appl.* **2012**, *6*, 662–668. [[CrossRef](#)]
17. Wang, P.-C.; Fang, Y.-C.; Jiang, Z.-Y. A Direct Swing Constraint-based Trajectory Planning Method for Underactuated Overhead Cranes. *Acta Autom. Sin.* **2014**, *40*, 2414–2419. [[CrossRef](#)]
18. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D. Continuous control with deep reinforcement learning. *J. Intell. Learn. Syst.* **2015**, *8*, A187. [[CrossRef](#)]
19. Fujimoto, S.; Van Hoof, H.; Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. *arXiv* **2018**. [[CrossRef](#)]
20. García, J.; Fernández, F. A comprehensive survey on safe reinforcement learning. *J. Mach. Learn. Res.* **2015**, *16*, 1437–1480.
21. Berkenkamp, F.; Turchetta, M.; Schoellig, A.P.; Krause, A. Safe Model-based Reinforcement Learning with Stability Guarantees. *arXiv* **2017**. [[CrossRef](#)]
22. Hewing, L.; Wabersich, K.P.; Menner, M.; Zeilinger, M.N. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annu. Rev. Control Robot. Auton. Syst.* **2020**, *3*, 269–296. [[CrossRef](#)]
23. Johannink, T.; Bahl, S.; Nair, A.; Luo, J.; Kumar, A.; Loskyll, M.; Ojea, J.A.; Solowjow, E.; Levine, S. Residual Reinforcement Learning for Robot Control. *arXiv* **2018**. [[CrossRef](#)]
24. Silver, T.; Allen, K.; Tenenbaum, J.; Kaelbling, L. Residual Policy Learning. *arXiv* **2018**. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.