Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.20268187

A Review of Robot Adaptive Control Driven by Deep Reinforcement Learning

Fute Xiao*

School of Computer Science (School of Artificial Intelligence), South-Central Minzu University, Wuhan 430000, China

**Corresponding author: Fute Xiao.*

Abstract

In complex settings like smart manufacturing and human-robot teamwork, robots face internal disturbances and external uncertainties. Traditional control methods depend on accurate models and manual parameter tuning, leading to complex adjustment, weak disturbance rejection, and poor generalization. Deep reinforcement learning (DRL) combines deep learning's feature extraction with reinforcement learning's sequential decision-making. Through end-to-end learning, it removes the need for exact system models and has become a key approach for robot adaptive control. This paper systematically reviews DRL-driven robot adaptive control. It first outlines core concepts and theory, building the technical framework that brings together DRL and adaptive control. It then analyzes mainstream DRL algorithm improvements and hybrid methods for adaptive control, explores main issues in Sim-to-Real transfer, and discusses safe DRL control modeling under constraints. The paper also introduces typical robot applications, examines current challenges and bottlenecks, and points out future trends. This review aims to clarify the development path of DRL-driven robot adaptive control, offering reference for theory, algorithm design, and engineering practice. It is hoped that this survey will help new researchers quickly grasp the landscape of the field.

Keywords

deep reinforcement learning, robot adaptive control, sim-to-real transfer, safe control

1. Introduction

The rapid growth of smart manufacturing, human-robot collaboration, and swarm operations demands high robot adaptability in dynamic environments. Robots need to handle internal changes like parameter drift and sudden load variation, as well as external uncertainties such as task switching and unknown obstacles. Traditional control methods work well when the system dynamics model is known and the environment remains stable, but their performance relies heavily on precise modeling and careful manual tuning. When facing complex nonlinearity, unmodeled dynamics, or sudden environmental changes, they suffer from difficult tuning, weak disturbance rejection, and poor generalization. These drawbacks make them unable to meet the requirements of flexible production and humanoid robot collaboration [1].

Deep reinforcement learning (DRL) merges the feature extraction power of deep learning with the sequential decision-making strength of reinforcement learning. It achieves model-free adaptive control through a "trial-error-feedback-optimization" closed loop, creating a new path for complex robot control [2]. Robot adaptive control aims to let the system automatically adjust its control strategy through real-time sensing and decision-making in uncertain, unstructured environments. DRL enables agents to learn optimal policies directly from high-dimensional raw sensor data. This end-to-end approach breaks the traditional reliance on accurate models and has succeeded in motion control, manipulation, and navigation [3].

However, transferring DRL policies from simulation to reality remains difficult. Mismatches in dynamics and sensor properties cause performance drops when deploying simulated policies on real robots. This "reality gap" has become a major bottleneck for practical use [4]. In safety-critical tasks, DRL agents must also satisfy critical safety limits while exploring to maximize long-term rewards. Safe reinforcement learning has therefore become an urgent problem in the robotics field [5]. This paper provides a systematic review of robot adaptive control driven by deep reinforcement learning, covering core techniques, algorithm advances, Sim-to-Real transfer, safe control, typical applications, and remaining challenges. It is intended to serve as a comprehensive reference for both researchers and engineers working in this field, helping them understand current capabilities and identify open problems. The remainder of this paper is structured as follows: Section 1 introduces core concepts, Section 2 describes the technical system, and subsequent sections cover algorithms, transfer, safety, challenges, and future directions.

2. Core Concepts and Theoretical Foundations

Robot adaptive control targets uncertain, unstructured environments; deep reinforcement learning combines deep learning and reinforcement learning. Merging DRL with adaptive control essentially means learning the system's inverse dynamics or optimal policy online from data, thereby compensating for unknown dynamics. The theoretical basis is that adaptive control provides stability and convergence frameworks, while DRL provides nonlinear approximation and model-free decision-making. The two complement each other, forming a key pathway for complex environments. This combination enables control systems to handle situations where traditional model-based approaches would fail due to unmodeled effects or rapidly changing conditions, such as a drone encountering sudden wind gusts or a manipulator picking up objects of unknown weight.

3. Core Technical System of DRL-Driven Robot Adaptive Control

3.1 Fundamentals of Robot Adaptive Control

The goal is enabling the system to adjust its control strategy online when dynamics change or disturbances occur, maintaining desired performance. Two key technical dimensions exist.

(1) State Representation and Multi-Source Perception Fusion

State representation is the basis of DRL policy learning and directly determines control performance. Choosing the right representation can greatly speed up training and improve final performance. Wang Xuesong [6] introduced representation learning-based offline reinforcement learning, which learns policies only from historical data without interacting with the environment. Features from offline datasets become low-dimensional vectors used for training. This approach significantly reduces the need for costly online interactions. Han Shijie [7] proposed GraphIDPre, a multi-agent framework that reconstructs state representations using competition and cooperation relationships and then makes decisions based on these relationship-based representations, enhancing coordination among multiple robots.

(2) Adaptive Reward Function and Auxiliary Rewards

The reward function guides the agent's learning behavior, defining the optimization goal. The core idea is to dynamically adjust reward weights based on state, task phase, or control feedback, letting the agent switch objectives flexibly. Zhang Shuzhong et al. [8] noted that the reward function bridges goals and decisions, and proposed one that helps the Critic network identify factors relevant to robot action generation. Siyuan Li et al. [9] designed two auxiliary reward types for single tasks and long-term skill chains. Such auxiliary rewards can provide denser learning signals, helping the agent progress even when the main task reward is sparse. Designing effective reward functions remains a challenging art, as small changes can lead to drastically different learned behaviors. At the optimization level, algorithms update the policy network by maximizing expected cumulative reward. SAC uses maximum entropy regularization to balance reward and randomness, improving exploration and robustness under uncertainty. PPO limits policy update step size through clipping, enhancing training stability. These mechanisms couple with adaptive rewards to drive robot adaptation.

3.2 Fundamentals of Deep Reinforcement Learning

DRL combines deep neural networks with reinforcement learning's sequential decision-making. The agent learns optimal strategies through trial and error to maximize cumulative reward. Deep networks let reinforcement learning overcome traditional limits. DQN first merged convolutional networks with Q-learning, using experience replay and target networks to learn directly from raw images, greatly improving training stability. This end-to-end approach provides a strong model-free tool for adaptive control, capable of processing high-dimensional sensory inputs without manual feature engineering.

3.3 Fusion Logic and Adaptability of DRL and Robot Adaptive Control

The fusion of DRL and adaptive control comes from deep complementarity in goals and functions, not simple combination. Doukhi et al. [10] proposed a multimodal NMPC strategy using DRL to enhance online optimization, enabling robots to handle complex nonlinear and high-dimensional problems beyond traditional methods. Structurally, the Actor-Critic architecture resembles the identification-control dual loop in adaptive control: the Actor network acts as the controller, generating commands in real time; the Critic evaluates performance like a reference model or performance monitor. Updating the policy network is essentially adaptive adjustment of controller parameters online, which aligns closely with the principle of self-tuning regulators in classical adaptive control and provides a natural bridge between the two fields.

4. Advances in DRL Algorithms for Robot Adaptive Control

DRL combines perception from deep learning with decision-making from reinforcement learning. Recent research on algorithm improvement and fusion with classical control has produced a range from basic optimization to hybrid architectures.

4.1 Application and Optimization of Mainstream DRL Algorithms

Robot control tasks involve continuous state and action spaces, demanding real-time and stable algorithms. Three main categories exist: value-based, policy gradient, and Actor-Critic methods.

(1) Value-based methods, represented by DQN, use deep networks to approximate the Q-function Q(s,a;θ). The network minimizes the Bellman error:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')}\left[\left(y^{DQN} - Q(s,a;\theta)\right)^2\right] \tag{1}$$

Experience replay and target networks stabilize training, solving dimensionality issues in high-dimensional state spaces. DQN suits discrete action tasks like waypoint selection and behavior mode switching but faces output dimension explosion with continuous actions, limiting use in precision control like joint torque regulation.

(2) Policy gradient methods parameterize policies directly, naturally fitting continuous action spaces. They optimize parameters via gradient ascent to maximize expected reward. PPO balances stability and sample efficiency well, making it widely used. Actor-Critic methods combine value estimation and policy optimization using dual networks: Actor outputs actions, Critic evaluates them. Key algorithms include DDPG, which uses deterministic policy to simplify sampling and handle continuous actions; TD3, which adds twin critics and delayed updates to reduce value overestimation; and SAC, which uses entropy regularization to balance reward and exploration diversity, showing strong robustness in complex tasks. These algorithms have become the backbone of many robot learning systems.

4.2 Research on Hybrid DRL Adaptive Control Algorithms

(1) Mixture of Experts (MoE) combined with DRL solves multimodal control tasks. For two-wheeled legged robots needing both rolling and stepping gaits, researchers integrated MoE into PPO to switch smoothly

between flat-surface rolling and obstacle stepping, enabling seamless transition between heterogeneous locomotion modes [11].

(2) Hybrid architectures of deep networks and DRL benefit perception-decision loops. In Human-Robot Collaboration task allocation, a CNN and Double DQN framework jointly optimizes target rewards using multimodal sensor information, leading to more adaptive task distribution and improved human-robot team efficiency.

(3) Model Predictive Control (MPC) combined with DRL represents another hybrid path. Researchers paired MPC with neural network estimators, using DRL to learn decisions under uncertainty while MPC ensures feasible control inputs and respects system constraints. This approach surpasses traditional PID and rule-based controllers in path accuracy and energy efficiency for manipulator and autonomous vehicle planning [12], showing promise for industrial adoption where performance and safety must be simultaneously guaranteed.

5. Sim-to-Real Transfer and Safe Adaptive Control

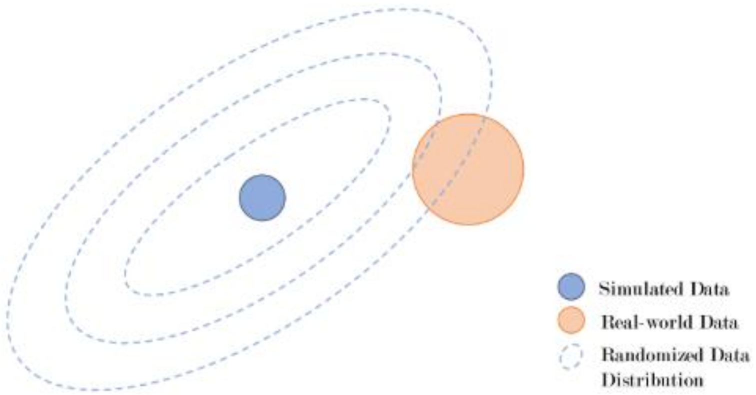
Sim-to-Real transfer and safe reinforcement learning are two hurdles for deploying DRL on real robots. The first handles differences between simulation and reality; the second ensures safety during physical interaction.

5.1 Core Problems and Implementation Methods of Sim-to-Real Transfer

DRL algorithms gather abundant training data in simulation, but policy performance drops sharply on real robots due to three types of mismatch: dynamics parameters, sensor observation shifts, and environmental randomness differences. These discrepancies are often interrelated and can amplify each other, making transfer highly unpredictable. Three main transfer approaches exist.

Domain Randomization is the most common strategy. During simulation, physical parameters, sensor noise, and target positions vary randomly over wide ranges, forcing the policy to learn robust behavior. High-level scene randomization improves generalization. In the ANYmal quadruped robot case, randomizing terrain friction and joint damping in simulation achieved stable walking on real rough ground, as shown in Figure 1.

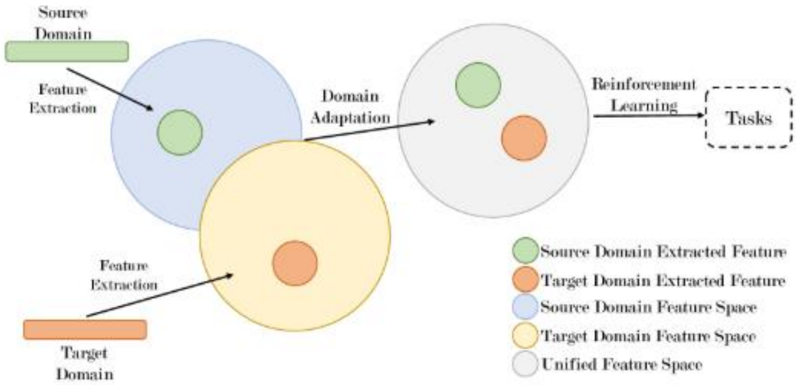
Figure 1: Intuition behind the Domain Randomization paradigm



Domain Adaptation narrows the feature gap between simulation and reality. Güitta-López et al. proposed SICGAN, an improved CycleGAN, to convert simulation images into realistic styles, letting the DRL agent train with “real appearance” in a virtual setting [13], as shown in Figure 2.

Progressive Neural Networks (PNNs) transfer knowledge from pretrained networks to new tasks through lateral connections, reusing knowledge without catastrophic forgetting. Meta-reinforcement learning encodes reward and action history into hidden states via LSTM, enabling quick adaptation to unseen tasks. These structured approaches continue to show promise for policy transfer [14] and offer ways to build more flexible robot learning systems.

Figure 2: Intuition behind the Domain Adaptation paradigm



5.2 Modeling and Application of DRL Adaptive Control under Safety Constraints

Despite progress, unpredictable policies and lack of safety guarantees limit deployment in safety-critical tasks like human-robot collaboration and precision assembly. A single unsafe action can cause damage or injury, so ensuring safety is not just desirable but mandatory.

Constrained Reinforcement Learning embeds safety constraints directly into policy optimization to achieve "safe training." Corsi et al. proposed λ-PPO, a Lagrangian-based PPO variant, jointly optimizing cost signals from Scenario-Based Programming with standard rewards. Safe filter methods allow free exploration during training but correct actions in real-time during execution using Quadratic Programming (QP). Tests on the Kinova Gen3 arm and Unitree H1 humanoid show these runtime filters work directly on policies not trained for safety, demonstrating good generality and practical value [15].

Control Barrier Functions (CBF) combined with safe reinforcement learning is another path. Traditional CBF guarantees safe set invariance by modifying control inputs, but combining with DRL faces conservatism and feasibility challenges. A joint training framework using Differential High-Order CBF and DRL introduces a neural parameterized penalty function, optimizing it alongside the policy network to reduce conservatism while maintaining safety guarantees [15]. This integrated approach represents an important step toward verifiably safe learning-based control, though many open issues remain in scaling to high-dimensional systems.

6. Existing Challenges and Technical Bottlenecks

Despite progress in DRL-based adaptive control, real-world deployment still faces systematic challenges.

6.1 Sample Efficiency and Learning Cost Dilemma

DRL algorithms need massive environmental interactions to learn effective policies. Real robot interactions are limited by physical time, hardware wear, energy cost, and safety risks. Simulation can run thousands of trials per second, while the same training on a real robot could take weeks or months, forcing heavy reliance on simulation. Sparse rewards worsen the problem. Tasks often only give positive feedback upon completing a specific sub-goal. For example, a grasping task rewards only when the object is lifted, leaving vast exploration without useful gradient signals. This causes poor convergence and extremely low random exploration efficiency. Reward design is also complex: overly simple rewards may cause local optima or unintended behavior; overly complex multi-objective rewards need careful weight tuning and may conflict with each other, making it hard to specify the right behavior a priori. Better exploration strategies and intrinsic motivation methods are being studied to address these issues.

6.2 Persistent Challenges of Sim-to-Real Transfer

Sim-to-Real transfer remains a core bottleneck. The compound effect of dynamics mismatches means parameter differences between simulation and reality couple within the control loop, causing sharp performance drops during real deployment [16]. The visual domain gap is another issue: visual transfer and policy learning are often handled separately, preventing joint optimization. Domain randomization improves robustness by widening parameter variation, but its effectiveness has limits. Determining optimal randomization ranges and which parameters to randomize still lacks theory and depends on designer experience and trial-and-error. Furthermore, some real-world effects like wear and tear or lighting changes are hard to model and randomize effectively.

6.3 Dual Challenge of Policy Robustness and Safety

DRL policies are extremely sensitive to small input perturbations. This vulnerability arises from neural network nonlinearity and insufficient coverage of state space boundaries during training. Attacks on a few key states can cause catastrophic outcomes, yet training frameworks lack effective defenses [17]. The absence of hard safety guarantees means policies may produce dangerous actions like exceeding joint limits, approaching singular postures, or colliding with obstacles. Existing methods still struggle to provide formal safety proofs that are reliable under all conditions, which remains a critical barrier for safety-critical applications such as medical robotics and autonomous driving.

7. Summary and Future Development Trends

DRL-driven robot adaptive control is moving toward building general cross-morphology foundation models, achieving rapid policy transfer and few-shot adaptation through large-scale pretraining. World-model-driven imagination learning will greatly improve sample efficiency, allowing robots to optimize policies through internal simulation without costly real-world interactions. Formal safety verification combined with safe reinforcement learning will provide certifiable behavioral guarantees for high-risk scenarios, which is essential for building trust in autonomous systems. Digital twin technology enables virtual-real iterative loops, continuously reducing the simulation-reality gap. Standardized evaluation protocols and open-source ecosystems will speed up technology accumulation and industrial adoption, helping robots transition from pre-programmed machines to truly autonomous environmental adapters. The future will see a control paradigm deeply integrating data-driven and knowledge-guided approaches, driving robots from fixed program execution toward intelligent adaptation in dynamic environments. As these technologies mature, we can expect broader deployment of adaptive robots in manufacturing, service, and exploration scenarios, fundamentally changing how robots operate in the real world. Continued interdisciplinary collaboration will be key to overcoming the remaining hurdles.

References

[1] Åström, K. J., & Hägglund, T. (2006). Advanced PID Control. ISA.

[2] Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518, 529–533.

[3] Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. Journal of Machine Learning Research, 17, 1–40.

[4] Zhao, W., Queralta, J. P., & Westerlund, T. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: A survey. Proc. IEEE SSCI.

[5] Brunke, L., et al. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. Annual Review of Control, Robotics, and Autonomous Systems, 5, 411–444.

[6] Wang Xuesong, Wang Rongrong, Cheng Yuhu. (2024). A survey of offline reinforcement learning methods based on representation learning. Acta Automatica Sinica, 50(06), 1104-1128.

[7] Zhang Shuzhong, Su Jiahong, Zhang Gong, et al. (2026). Adaptive compliant control for dual-robot cooperative handling based on deep reinforcement learning. Mechanical & Electrical Engineering, 1-11.

[8] Li, Z., Zhou, Y., Wang, L., et al. (2025). An end-to-end controller with image-based visual servoing of industrial manipulators with soft-actor-critic algorithm. Knowledge-Based Systems, 311, 112980.

[9] Doukhi, O., & Lee, D. J. (2025). Sim-to-real learning-based nonlinear MPC for UAV navigation and collision avoidance in unknown cluttered environments. IEEE Access, 13, 46249-46262.

[10] (2025). Deep reinforcement learning of mobile robot navigation in dynamic environment: A review. Sensors, 25(11), 3394.

[11] (2025). Hybrid deep learning-reinforcement learning for adaptive human-robot task allocation in Industry 5.0. Journal of Systems Engineering, 13(8), 631.

[12] Tiwari, R., Khapre, S., & Singh, A. (2026). Reinforcement learning in robotic systems: A review on sim-to-real transfer. Robotics and Autonomous Systems.

[13] Güitta López, L. (2024). Efficiently transferring deep reinforcement learning experience to industrial assets [Doctoral dissertation, Universidad Pontificia Comillas].

[14] Corsi, D., Yerushalmi, R., Amir, G., et al. (2024). Enforcing specific behaviours via constrained DRL and scenario-based programming. In International Conference on Neural Information Processing (ICONIP).

[15] Ma, Y., et al. (2025). Differential high order control barrier function-based safe reinforcement learning. IEEE Robotics and Automation Letters, 10(7), 7524-7531.

[16] Zhao, et al. (2026). State-aware perturbation optimization for robust deep reinforcement learning. IEEE Transactions on Mobile Computing, 25, 992.

[17] Toquebiau, M., Jun, J. Y., Benamar, F., et al. (2025). Multi-agent deep reinforcement learning in robotics: Context and open challenges. HAL Preprint.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.