Analysis of the Application of Reinforcement Learning in Motion Control of Robots in Unstructured Environments

Weihan Wang

{2361910105@hhu.edu.cn}

School of Mechanical and Electrical Engineering, Hohai University, Changzhou, 213200, China

Abstract: Currently, motion control for robots in unstructured environments suffers from insufficient adaptability and robustness, and traditional methods rely on precise models that are prone to failure. Therefore, this paper focuses on the application and optimization of reinforcement learning in this scenario. Existing research has shown that balancing exploration efficiency and stability can be achieved through basis functions, teacher-student models, and domain randomization; the reward function can be optimized using an improved SAC algorithm to construct more accurate state modeling; and multi-layer convolutional networks and attention mechanisms can be used to fuse multi-source perceptual features. A complete technical framework has been constructed, research gaps have been identified, and a reference has been provided for theoretical innovation and engineering implementation of motion control for robots in unstructured environments.

Keywords: Unstructured environment; robot motion control; reinforcement learning; SAC algorithm

1 Introduction

As robotics technology extends to specialized operations, the demand for robot applications in unstructured environments such as disaster relief, outdoor inspections, and mining operations has surged. These environments present challenges such as varied terrain, strong sensory interference, and complex dynamic constraints, posing stringent challenges to the adaptability, robustness, and real-time performance of robot motion control. Traditional control methods, which rely on precise models and manual rules, are prone to failure under dynamic disturbances and struggle to meet operational requirements.

Reinforcement learning, with its model-free characteristics of "interactive trial and error", provides a new path to solve this problem. Deep reinforcement learning has broken through the bottleneck of high-dimensional state modeling and performed well in tasks such as robot gait adaptation. It is moving from theory to engineering. S. Iida et al. proposed an adaptive assignment method based on basis functions for reinforcement learning to process the state

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365230

space to achieve smooth control of humanoid robots[1]. K. Yang et al. designed a hierarchical strategic planning structure, which combines a global search strategy and local fine decision strategy to achieve fine control of robot motion and long-term sequence planning, ensuring stable performance under high-speed motion and complex obstacle environment conditions[2]. Yan, C. et al. proposed a large-scale technology-based parallel end-to-end teacher-student learning network framework to improve the flexibility and stability of reinforcement learning-based quadruped robots in these environments[3].

2 Classification of Reinforcement Learning Methods in Robot Control

Currently, reinforcement learning methods in robot control can be classified according to algorithm type and control task dimension. Classified by algorithm type, they can be divided into model-independent reinforcement learning and model-based reinforcement learning.

In model-free reinforcement learning, mainstream algorithms such as PPO (proximal policy optimization), SAC (soft Actor-Critic), and TD3 have shown strong adaptability in unstructured environments. Among them, the PPO algorithm has a high success rate in static target achievement scenarios and is suitable for basic control tasks of continuum robots [4]; the SAC algorithm shows better adaptability in dynamic target tracking environments, especially suitable for robot motion control in complex obstacle environments [2,4]. These algorithms do not rely on dynamic environmental models but optimize policies through direct interaction with the environment, which is consistent with the characteristics of unstructured environments that are difficult to model accurately.

Model-based reinforcement learning can improve sample efficiency by constructing environmental dynamics models, but it suffers from insufficient generalization. This type of method reduces invalid trial and error in the exploration process by modeling environmental features, but its performance is highly dependent on the accuracy of the model and is easily affected in dynamic, unstructured scenarios such as terrain changes. Reinforcement learning methods in robot control can be classified into gait adaptive control, obstacle avoidance control, and posture stabilization control according to the control task dimension. The core of gait adaptive control is to generate and adjust the robot's gait for terrain changes, such as gait optimization of multi-legged robots in complex terrains such as gravel, grass, and slopes [3,5], and gait mode selection (trotting/medium speed/galloping) based on power budget for micro robots [6]. Obstacle avoidance control focuses on decision optimization in dynamic/static obstacle scenarios. Through the real-time decision-making ability of reinforcement learning algorithms, it enables robots can pass safely in environments with moving obstacles and sudden terrain changes [2,7]. Posture stabilization control focuses on solving the problem of trunk posture adjustment of robots in response to external disturbances, such as the action control of humanoid robots standing up from chairs [1], and point stabilization control of nonholonomic constrained mobile robots [8].

3 Core Research Questions and Related Research Progress

3.1 The problem of balancing exploration efficiency and motion stability in reinforcement learning algorithms

The dynamic and uncertain nature of unstructured environments requires reinforcement learning algorithms to maintain robot motion stability while ensuring exploration efficiency to adapt to complex scenarios. Existing research achieves this balance through various technical approaches:

For the local minima problem in high-dimensional state space, some studies have proposed an adaptive allocation method based on basis functions. By evaluating the activity trace of the basis functions, the basis functions that have the greatest impact on local minima are eliminated. This reduces the number of basis functions while avoiding exploration into invalid regions, and finally achieves smooth motion control for humanoid robots to stand up from chairs [1].

For the problem of complex terrain adaptation of quadruped robots, researchers designed a large-scale parallel end-to-end teacher-student learning network framework, which combines the height potential estimation of gated loop units with omnidirectional terrain learning courses, enabling the robot to move smoothly in any command direction, and realize zero-sample transfer from the simulation model to the real Unitree Go1 robot through the state machine, which significantly improves the robot's adaptability and mobility on various urban unstructured terrains [3].

In the locomotion control of the hexapod robot, the adversarial discriminator is trained by introducing a motion prior dataset, which guides the reinforcement learning algorithm to learn the natural gait. This not only improves the efficiency of gait generation during the exploration process, but also ensures the robustness of the real robot's movement in complex terrain [5].

To address the resource constraints of microrobots, the robustness of the strategy is improved by domain randomization and a resource-aware gait scheduling perspective is proposed. Under the premise of meeting the device power budget, the gait mode that maximizes the expected reward is selected, thus balancing exploration efficiency and motion stability [6].

3.2 Reinforcement Learning Reward Function Design and State Space Modeling Optimization Problem

The dynamic nature of unstructured environments (such as abrupt terrain changes, obstacle movement, and payload variations) places higher demands on the design of reward functions and state space modeling in reinforcement learning. The relevant optimization strategies are as follows.

The reward function is dynamically optimized. An adaptive reward module and an LSTM (Long Short-Term Memory) target prediction unit are embedded in the SAC algorithm framework. The action value estimation is dynamically adjusted by combining historical state information and uncertainty in the training process, so that the reward function can adapt to environmental changes in real time and improve the convergence speed and control accuracy of the model in dynamic scenarios [2].

Accurate modeling of the state space is carried out. For differentially driven mobile robots with variable payloads, an 11-dimensional state vector is constructed as the input of the reinforcement learning agent, and the wheel torque is used as the output vector and integrated into the robot dynamic model. By considering kinematics and dynamic constraints, a target navigation success rate of 98.2% is achieved under the condition of unseen load mass [7]. For nonholonomic constraint mobile robots, a memory storage module containing the current state, control action, reward, and next state is constructed. The parameters are updated based on the value function loss of the real-time network and the target network to optimize the accuracy of state modeling and effectively realize the point stability control of the robot [8].

3.3 Research Progress on the Fusion Control of Reinforcement Learning and Visual/LiDAR Sensing Data

Robot motion control in unstructured environments relies heavily on the effective use of perception data, and significant progress has been made in the fusion mechanism of reinforcement learning and visual/LiDAR perception data.

In terms of multi-source signal feature fusion, based on the SAC algorithm framework, a multi-layer convolutional network and attention mechanism are used to extract and compress representative features of multi-source signals such as cameras, force sensors, and lidar. The fused features are then input into the actor network to achieve fine control of robot motion and long-term sequence planning, ensuring stable performance in high-speed motion and complex obstacle environments [2].

In terms of environmental perception and gait adaptation, in the parallel end-to-end teacher-student learning network framework of quadruped robots, the gated loop unit realizes the potential estimation of the height around the robot by fusing environmental perception data. Combined with the omnidirectional terrain learning course, the robot can adjust the angle of the motion joints according to the perceived terrain information to achieve smooth output [3].

4 Application and Comparison of Reinforcement Learning Gait Adaptive Control for Robots with Different Configurations

Reinforcement learning exhibits differentiated application characteristics in gait adaptive control of robots with different configurations. Typical applications and performance comparisons for each configuration are as follows:

First, there are multi-legged robots, which can be divided into quadruped robots, hexapod robots, and micro quadruped robots. Quadruped robots can achieve zero-sample transfer from simulation models to real Unitree Go1 robots by combining large-scale parallel end-to-end teacher-student learning networks with state machines, and exhibit excellent adaptability and mobility on various unstructured urban terrains such as gravel, grass, slopes, and steps [3].

The hexapod robot is based on a deep reinforcement learning method of motion prior. By generating an optimized motion prior dataset to train an adversarial discriminator, the robot is guided to learn a natural gait and successfully realizes the walking of a real hexapod robot in complex terrain, demonstrating excellent robustness of non-visual information [5].

The miniature quadruped robot adopts a compact FP32 multilayer perceptron (MLP) strategy, improves robustness through domain randomization, adapts to hardware with limited resources through Int8 quantization, and achieves stable movement on uneven terrain based on resource-aware gait scheduling [6].

Then there are mobile robots, which can be divided into nonholonomic constraint mobile robots and differential drive mobile robots. Nonholonomic constraint mobile robots are based on deep reinforcement learning to build kinematic models and dual network (real-time network and target network) update mechanisms, which effectively achieve point stability control and show good stability in simulation and experiment [8].

The differential-driven mobile robot achieves a target navigation success rate of 98.2% without training on payload quality, thanks to 11-dimensional state vector modeling and a reinforcement learning motion planner adapted to variable payload [7].

The continuum robot uses PPO and SAC algorithms for basic control. PPO has the highest success rate in static target achievement scenarios, while SAC is more adaptable in dynamic target tracking environments, laying the foundation for multi-arm collaborative control [4].

The adaptive assignment reinforcement learning method based on basis functions for humanoid robots achieves a smooth action of standing up from a chair with fewer basis functions, avoiding the problems of local minima and dimensionality curse [1].

The robotic arm plans motion on various robot structures through deep reinforcement learning algorithms, integrates OpenAI Gym and PyBullet simulators, solves the problems of static/random target achievement and collision avoidance in predefined configuration space, and has good scalability and repeatability [9,10]; the ability of reinforcement learning agents to adapt to dynamic obstacle environments has promoted their increasingly widespread application in robotic arm motion planning [11,12].

5. Conclusion

This paper adopts an overall research framework of algorithm-configuration-fusion-engineering, systematically outlining the corresponding technical system, deeply analyzing the core strategies and implementation mechanisms of configuration-specific control, and focusing on the multi-method fusion mechanism and key bottlenecks in engineering implementation. Addressing the balance between exploration efficiency and motion stability in reinforcement learning algorithms, the paper suggests removing the influence of local minima through adaptive allocation of basis functions and improving the flexibility and stability of quadruped robots in unstructured environments using a large-scale parallel end-to-end teacher-student learning framework. Regarding the design and optimization of reward functions and state space modeling, an adaptive reward module can be introduced to dynamically adjust the reward function according to environmental changes. Simultaneously, a more comprehensive high-dimensional state space model is constructed to achieve precise control of complex tasks. In terms of combining reinforcement learning with perceptual data, multi-source signal features can be fused through multi-layer convolutional networks and attention mechanisms. Finally, reinforcement learning exhibits differentiated application characteristics in gait adaptive control of robots with different configurations, each with its own advantages and disadvantages in stability and adaptability. Through a systematic review of algorithms, robot configurations, and perceptual fusion, this paper clarifies the core issues and technological frontiers in the current field. The technical framework developed in this study aims to provide a clear roadmap for future researchers and engineers.

References

[1] S. Iida, M. Kanoh, S. Kato and H. Itoh, "Reinforcement learning for motion control of humanoid robots," 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), Sendai, Japan, 2004, pp. 3153-3157 vol.4, doi: 10.1109/IROS.2004.1389902.

[2] K. Yang, "Research on Optimization of Robotic Motion Control Algorithms based on Deep Reinforcement Learning," 2025 4th International Conference on Intelligent Mechanical and Human-Computer Interaction Technology (IHCIT), Beijing, China, 2025, pp. 1-5, doi: 10.1109/IHCIT66787.2025.11198978.

[3] Yan, C., Wang, N., Gao, H. et al. An advanced reinforcement learning control method for quadruped robots in typical urban terrains. Int. J. Mach. Learn. & Cyber. 16, 3747–3757 (2025). https://doi.org/10.1007/s13042-024-02478-9.

[4] X. Liu, J. Wu, Y. Li, C. Qi, Y. Xue and F. Gao, "Learning Natural and Robust Hexapod Locomotion over Complex Terrains via Motion Priors based on Deep Reinforcement Learning," 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hangzhou, China, 2025, pp. 17101-17106, doi: 10.1109/IROS60139.2025.11247402.

[5] A. Demir and V. Sezer, "Motion Planning and Control with Randomized Payloads Using Deep Reinforcement Learning," 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 2019, pp. 32-37, doi: 10.1109/IRC.2019.00014.

[6] R. Gao et al., "Motion Control of Non-Holonomic Constrained Mobile Robot Using Deep Reinforcement Learning," 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), Toyonaka, Japan, 2019, pp. 348-353, doi: 10.1109/ICARM.2019.8834284.

[7] R. Gao et al., "Motion Control of Non-Holonomic Constrained Mobile Robot Using Deep Reinforcement Learning," 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), Toyonaka, Japan, 2019, pp. 348-353, doi: 10.1109/ICARM.2019.8834284.

[8] M. Elsaman et al., "Deep Reinforcement Learning for Foundational Control of a Continuum Robot Towards Cooperative Manipulation," 2025 7th Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 2025, pp. 500-503, doi: 10.1109/NILES68063.2025.11232046.

[9] Parák, R.; Kůdela, J.; Matoušek, R.; Juříček, M. Deep-Reinforcement-Learning-Based Motion Planning for a Wide Range of Robotic Structures. Computation 2024, 12, 116. https://doi.org/10.3390/computation12060116.

[10] arXiv:2512.24740 [cs.RO] (or arXiv:2512.24740v1 [cs.RO] for this version) https://doi.org/10.48550/arXiv.2512.24740.

[11]Elguea-Aguinaco, Íñigo, Inziarte-Hidalgo, Ibai, Bøgh, Simon, Arana-Arexolaleiba, Nestor, A Review on Reinforcement Learning for Motion Planning of Robotic Manipulators, International Journal of Intelligent Systems, 2024, 1636497, 29 pages, 2024. https://doi.org/10.1155/int/1636497

[12] S. R. Mulkana, "Safe Reinforcement Learning for Collaborative Robots in Dynamic Human Environments," 2024 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Glasgow, United Kingdom, 2024, pp. 61-65, doi: 10.1109/ACIIW63320.2024.00013.