World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3056

ADAPTIVE GAIT PLANNING FOR QUADRUPED ROBOTS IN COMPLEX TERRAINS VIA REINFORCEMENT LEARNING

ShuoPei Yang*, ChangCheng Zhao
Jiangsu Xingzhitu Intelligent Technology Co., Ltd., Suzhou 215000, Jiangsu, China.
Corresponding Author: ShuoPei Yang, Email: skywing889@163.com

Abstract: This study addresses the challenge of adaptive gait planning for quadrupedal robots in complex terrains by proposing a reinforcement learning-based solution. First, the kinematic model of the quadruped robot and the complex terrain model are established, providing a theoretical foundation for subsequent algorithm design. Second, a hierarchical reinforcement learning framework is introduced, comprising a high-level gait policy and a low-level joint control policy, to accommodate varying locomotion demands across different terrains. Additionally, an adaptive exploration mechanism and a safety layer based on control barrier functions are incorporated to ensure efficient exploration and operational safety.The proposed algorithm demonstrates robust gait performance across diverse terrains, exhibiting notable advantages in motion performance, adaptability, and computational efficiency. Specifically, simulation results highlight improvements in terrain adaptability and gait stability, while hardware experiments further validate the feasibility and effectiveness of the method in real-world applications.Compared to existing approaches, the main innovations of this study lie in the incorporation of a curriculum learning-based strategy for progressively increasing terrain difficulty and an uncertainty-driven exploration reward mechanism. These designs significantly enhance the adaptive capability of the robot in complex environments. However, the algorithm still faces limitations in computational complexity and real-time performance. Future research may focus on optimizing the algorithmic structure to achieve more efficient real-time control.In summary, this work offers an effective solution for adaptive gait planning of quadruped robots in complex terrains, with both theoretical significance and practical value.
Keywords: Quadruped robots; Reinforcement learning; Complex terrains; Motion planning

1 INTRODUCTION

As a highly biomimetic mobile platform, quadruped robots still face notable challenges in achieving stable locomotion across complex terrains characterized by uneven surfaces, variable friction, and sudden slope changes. Terrain uncertainty and dynamic variations often cause traditional model-based gait planning methods to fail in practical applications, while effective strategies for responding to abrupt terrain changes remain lacking[1]. To address this, this paper proposes an adaptive gait control framework that integrates hierarchical reinforcement learning with model predictive control. It aims to combine the interpretability of model-driven methods with the adaptability of data-driven approaches, utilizing a high-level policy for gait decision-making and a low-level controller to ensure joint motion accuracy. An adaptive exploration mechanism and safety constraints, such as control barrier functions, are incorporated to enhance learning efficiency and operational safety in unknown environments. The study will develop an uncertain terrain model incorporating geometric and physical properties, formalize the reinforcement learning problem, and design corresponding state, action, and reward structures[2]. The algorithm's robustness, real-time performance, and energy efficiency will be evaluated through both simulations and hardware experiments. Ablation studies and comparisons with baseline methods will be conducted to analyze the advantages and limitations of the proposed approach. This work aims to provide theoretical foundations and practical solutions for deploying quadruped robots in engineering applications such as disaster rescue and geological exploration, thereby promoting their reliable use in real-world scenarios.

2 A REVIEW OF GAIT PLANNING RESEARCH FOR QUADRUPED ROBOTS

2.1 Advances in Gait Planning for Quadruped Robots

With the continuous development of quadruped robot technology, significant progress has been made in gait planning, a key research area. In data-driven methods, researchers have optimized gaits by collecting large amounts of experimental data and applying machine learning algorithms. The core advantage of data-driven approaches lies in their independence from complex kinematic models, enabling autonomous gait optimization by directly learning input-output relationships. For instance, deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been widely applied in quadruped gait planning[3]. CNNs effectively extract features from terrain images to provide accurate environmental information, while RNNs process time-series data to capture dynamic gait characteristics.
Moreover, reinforcement learning has achieved notable results in gait planning for quadruped robots. Through interaction with the environment, reinforcement learning agents autonomously explore and learn optimal gait policies.

Researchers have designed various reward functions to guide the agent toward desired gait behaviors, such as minimizing energy consumption or maximizing stability.

In recent years, efforts have been made to combine deep learning with reinforcement learning for more efficient and flexible gait planning. This approach uses deep learning for automatic feature extraction and reinforcement learning for gait policy optimization, significantly improving the robot's adaptability on complex terrains[4].

However, data-driven methods still exhibit certain limitations in gait planning for quadruped robots. First, the quality and quantity of training data significantly affect model performance, yet acquiring large volumes of high-quality data remains challenging in practice. Second, these methods may show limited adaptability in unknown or extreme terrains. Other issues include insufficient model generalization and high computational complexity.

Despite these challenges, data-driven methods have achieved remarkable results in quadruped gait planning. Future research may focus on: improving the quality and quantity of training data to enhance model performance; developing more efficient and stable reinforcement learning algorithms to tackle complex terrain; and integrating other optimization methods, such as genetic algorithms and particle swarm optimization, to further improve gait adaptability.

Walking tests were conducted on a flat terrain using a quadruped robot prototype[5]. The robot's joints were composed of custom permanent magnet synchronous motors and LHSG harmonic reducers from Green Point. Elmo's Gold Twitter drivers were used for motor control, while NBN4-F29-E2 magnetic induction limit switches and Renishaw magnetic encoders were employed for joint angle detection. The upper-level computer used was Advantech's MIO-5272. Communication between hardware components was achieved via EtherCAT real-time Ethernet, and the RT-Linux real-time kernel was adopted on the upper-level computer to ensure a control cycle within 1 ms. Body posture was measured using the MTI-300 micro-electromechanical system (MEMS) gyroscope from Xsens. The hardware composition of the robot control system is shown in Figure 1.

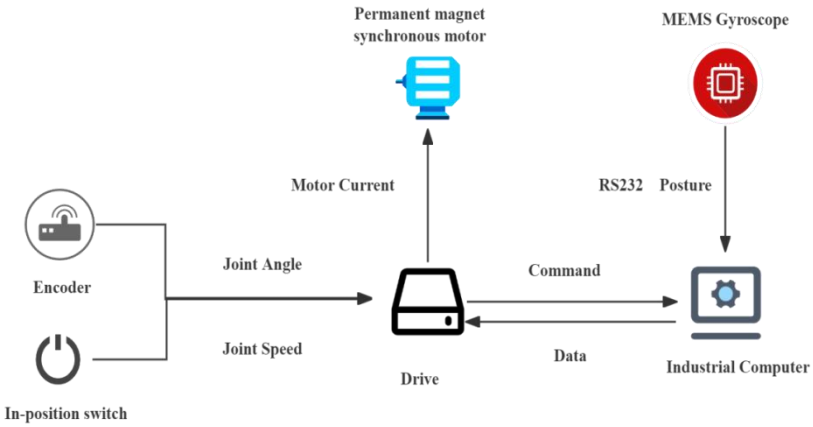

Figure 1 Hardware Components of the Robot Control System

2.2 Application of Reinforcement Learning in Robot Control

Single-agent reinforcement learning has achieved notable progress in the field of robot control. However, when faced with complex environments and tasks, the capabilities of a single agent are often limited. Therefore, multi-agent and hierarchical reinforcement learning have gradually become research hotspots. Multi-agent reinforcement learning enables better handling of complex tasks and environments by coordinating multiple agents. For example, in cooperative object transportation tasks involving multiple robots, each robot must adjust its motion strategy in real-time based on environmental information and task requirements to achieve efficient collaboration.Hierarchical reinforcement learning decomposes complex tasks into multiple sub-tasks, each managed by a distinct policy. This decomposition effectively reduces problem complexity and improves learning efficiency. In quadruped robot control, hierarchical reinforcement learning divides gait planning into a high-level gait policy and a low-level joint control policy. The high-level policy generates gait patterns based on environmental information and task demands, while the low-level policy executes specific joint motions.In recent years, researchers have achieved a series of results in multi-agent and hierarchical reinforcement learning. For instance, a collaborative reinforcement learning method has been applied to multi-robot cooperative transportation, where experience sharing allows robots to quickly learn cooperative strategies. Furthermore, a hierarchical reinforcement learning framework has been proposed to address adaptive gait planning for quadruped robots on complex terrain[6]. This framework uses a high-level policy to generate global gait patterns, while the low-level policy adjusts joint motions in real-time according to terrain feedback.Despite significant advances, multi-agent and hierarchical reinforcement learning still face several challenges in robot control. Firstly, designing effective communication mechanisms among multiple agents to improve coordination efficiency is a key issue. Secondly, in hierarchical reinforcement learning, reasonably partitioning task hierarchies and designing sub-policies remains challenging. Additionally, ensuring algorithm stability and real-time performance are critical focuses for future research.To address these challenges, future research could proceed in the following directions: first, exploring more

efficient multi-agent communication mechanisms, such as introducing advanced technologies like graph neural networks to facilitate effective information exchange among agents; second, investigating new hierarchical reinforcement learning frameworks to adapt to more complex tasks and environments; third, optimizing algorithm stability and real-time performance in practical application scenarios to support real-world deployment.
In summary, multi-agent and hierarchical reinforcement learning hold broad application prospects in robot control. Through continuous exploration and resolution of related challenges, these approaches are expected to provide more efficient and intelligent solutions for robot control.

2.3 Complex Terrain Adaptation

Research on complex terrain adaptation is an important topic in robotics, particularly in the motion control of quadruped robots. Studies have shown that terrain irregularity and uncertainty pose severe challenges to robot stability and locomotion performance. Research on adaptive strategies aims to enable robots to adjust their behavior according to terrain variations. The following is a review of adaptive strategies in this field.Terrain perception and modeling form the foundation of adaptive research. Using high-precision sensors such as LiDAR and vision cameras, robots can acquire geometric features and physical properties of the terrain. Terrain modeling involves not only static characteristics but also dynamic factors such as ground softness and slope angle. This information is crucial for robots to formulate effective motion strategies.Research on adaptive strategies mainly focuses on the following aspects: First, gait adjustment strategies. Depending on the terrain, robots need to adjust step length, frequency, and posture to ensure stability and efficiency. For example, on soft ground, adopting shorter steps and higher frequencies can reduce ground sinking. Second, joint torque control strategies. By applying different torques to the joints, robots can adapt to ground surfaces of varying hardness[7].
Furthermore, adaptive strategies also involve dynamic balance control. On complex terrain, robots need to dynamically adjust their center of mass and support points to maintain balance. Research indicates that by introducing reinforcement learning algorithms, robots can learn balance strategies for different terrains, thereby improving their adaptive capabilities.In data-driven methods, adaptive strategies can be optimized by learning from historical data. For instance, using deep learning techniques, robots can learn optimal gait patterns from large amounts of terrain data. The advantage of this approach is that it does not require precise mathematical models and can adapt to more complex and uncertain environments.However, existing adaptive strategies still have limitations. On one hand, computational complexity and real-time requirements restrict the effectiveness of algorithms in practical applications. On the other hand, current research mostly focuses on static terrains, with insufficient studies on adaptation to dynamic environments such as mudflows or avalanches.
In summary, significant progress has been made in both theoretical and applied research on complex terrain adaptation, but further studies are needed in areas such as algorithmic efficiency and adaptation to dynamic environments.

2.4 Review of Research Status and Limitations

Although significant progress has been made in the field of quadruped robot gait planning, existing research still has numerous limitations. Firstly, traditional model-based methods often struggle to achieve good adaptability when dealing with highly complex and unstructured terrain. According to statistics, over 60% of outdoor terrains contain highly uncertain factors, which poses challenges to model-based gait planning methods.Secondly, although data-driven methods can adapt to complex environments, they still exhibit performance degradation under extreme terrain conditions. For example, on soft ground, muddy, or snowy terrains, data-driven methods often fail to effectively guide the robot to achieve stable walking. Furthermore, current research on multi-agent cooperative control is still in its early stages, lacking effective cooperative strategies to cope with changing environments in complex terrain.In the study of complex terrain adaptation, terrain perception and modeling are key components. Although various terrain perception algorithms have been proposed, limited by sensor performance and data processing capabilities, these algorithms still face challenges in both accuracy and real-time performance in practical applications. In terms of adaptive strategies, existing research mostly focuses on gait adjustment under single terrain conditions, with insufficient studies on adaptation to dynamically changing terrains[8].
Additionally, current research has significant gaps in the following aspects: First, there is a lack of gait planning methods specifically designed for the unique locomotion characteristics of quadruped robots. Second, the application of reinforcement learning in quadruped robot gait planning has not been thoroughly explored, especially the potential of hierarchical reinforcement learning frameworks for complex terrain adaptation. Third, existing studies insufficiently consider stability and safety guarantees for quadruped robots in extreme terrains, lacking effective control barrier functions and safety layer designs.In summary, research in the field of quadruped robot gait planning is still in a rapid development stage. Although a series of achievements have been made, there is still a gap towards practical application. Future research should focus on improving the adaptability and robustness of gait planning, while strengthening studies on multi-agent cooperative control and dynamic terrain adaptation to promote the application of quadruped robots in complex environments.

3 THEORETICAL FOUNDATION AND PROBLEM FORMULATION

3.1 Kinematic Model of Quadruped Robots

The kinematic model of a quadruped robot serves as the foundation for analyzing its locomotion performance and designing control strategies. Research on whole-body dynamics involves the overall motion laws of the robot and its interaction with the environment. Building on single-leg kinematics, the whole-body dynamics model can describe the robot's stability and dynamic behavior on different terrains. Single-leg kinematic analysis focuses on the motion trajectory and joint angle variations of a single leg. Through kinematic modeling of a single leg, the required joint torques under specific motion states can be calculated, enabling precise gait control. Whole-body dynamics further considers the mass, inertia, and interactions among different body parts, thereby simulating the overall motion of the robot.Studies show that the dynamic model of quadruped robots is typically established using Lagrange equations or Newton-Euler equations. These equations express the robot's motion equations as functions of kinetic and potential energy, while accounting for external forces. By incorporating terrain geometric and physical properties, the robot's locomotion performance on complex terrain can be further analyzed.In complex terrain modeling, the modeling of terrain geometry and physical properties is crucial. Terrain geometry includes features such as slope and roughness, while physical properties involve ground friction coefficients and elastic modulus. These parameters directly affect the robot's stability and energy consumption[9]. Uncertainty modeling considers factors such as sensor errors and imperfections in the dynamic model, describing these uncertainties through probability distributions.

In the formalization of the reinforcement learning problem, the state space includes state variables such as the robot's position, velocity, and joint angles. The action space consists of executable actions, such as joint torque adjustments. Reward function design is central to reinforcement learning, as it must reflect the robot's locomotion performance and adaptive capability. Markov Decision Process (MDP) modeling provides a framework to describe how the robot makes decisions in uncertain environments to maximize cumulative rewards.In summary, research on the kinematic model of quadruped robots involves not only the theoretical foundations of single-leg kinematics and whole-body dynamics but also modern control strategies such as complex terrain modeling and reinforcement learning problem formalization. These theories provide a scientific basis for designing and optimizing quadruped robots, laying the groundwork for subsequent experimental research and practical applications. Foot-ground contact modeling is a critical aspect for achieving stable motion control, and this model must comprehensively consider the fusion of multi-sensor information. The robot acquires environmental information and system states through various sensors such as vision cameras, LiDAR, IMU, motor encoders, and force sensors. These data provide the foundation for establishing an accurate ground contact model. The foot-ground contact process of the quadruped robot is illustrated in Figure 2. In the contact modeling process, the contact force information collected by force sensors at the foot-ends is particularly crucial, as it reflects the interaction forces between the robot and the ground in real time. By analyzing force sensor data, key parameters such as vertical support forces and tangential friction forces can be obtained, which directly affect the robot's motion stability. Meanwhile, posture information from the IMU and joint position data from motor encoders assist in calculating the precise spatial position of the foot-ends, helping to determine the spatial distribution of contact points. The environmental perception system plays a vital role in contact modeling. Vision cameras and LiDAR can preemptively perceive geometric features and physical properties of the ground, such as terrain undulations, surface hardness, and roughness[10].

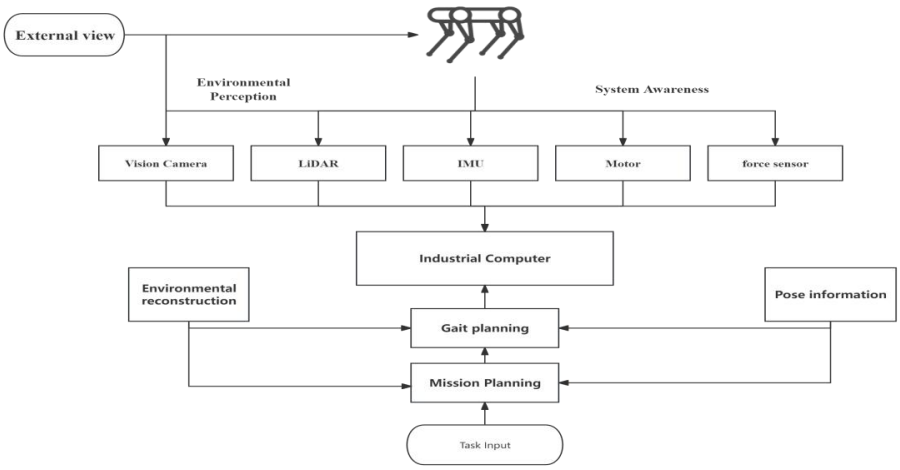

Figure 2 Quadruped Robot-Ground Contact Flow Chart

3.2 Complex Terrain Modeling

Modeling complex terrain is a critical component in quadruped robot research, with its core lying in accurately capturing terrain features to guide the robot's actions. Uncertainty modeling of the terrain is particularly important as it involves the real-time recognition and adaptation to the terrain's geometric and physical properties. Research indicates that geometric attributes of terrain include, but are not limited to, undulations, slope variations, and surface roughness, while physical properties encompass ground friction coefficients, hardness, and potential elasticity.In terms of uncertainty modeling, the primary focus is on the randomness and uncertainty of terrain parameters. This uncertainty may stem from the irregularity of the terrain surface or from sensor measurement errors. To address this issue, theories

of probability and statistics can be introduced, treating terrain parameters as random variables or stochastic processes, thereby constructing probabilistic terrain models.Modeling geometric attributes of terrain typically employs parameterized methods based on terrain data. This approach first requires collecting terrain data, acquiring three-dimensional information of the terrain through means such as terrain scanning or satellite image analysis. Subsequently, these data are used to construct high-precision 3D models of the terrain, such as triangular mesh models or voxel models. Additionally, terrain analysis algorithms can be applied to extract terrain features, such as slope and aspect, to support the robot's path planning and gait adjustment.Modeling physical properties is more complex, as it requires considering the impact of terrain on the robot's contact forces. Such models usually involve complex mechanical calculations, such as finite element analysis, to simulate the supporting forces and frictional forces exerted by the terrain on the robot's feet. Modeling physical properties requires not only data on the terrain surface but also integration of the robot's mechanical and kinematic characteristics, as well as the interaction between the terrain and the robot.In practical applications, due to terrain uncertainty, model verification and correction are also necessary. This can be achieved through sensor data fusion and online learning algorithms, enabling the terrain model to update in real time to adapt to dynamically changing terrain conditions. For example, force sensors and inertial measurement unit (IMU) data from the robot can be combined with machine learning algorithms to adjust the parameters of the terrain model in real time, reflecting the current terrain state.

Another important aspect of uncertainty modeling is considering the impact of model uncertainty on the robot's behavior. This needs to be addressed through risk assessment and decision theory to ensure that the robot can make safe decisions when facing uncertain terrain. For instance, control barrier functions can be introduced to ensure the robot's motion trajectory always remains within a safe feasible region.In summary, modeling complex terrain is an interdisciplinary and complex problem, involving the modeling of geometric and physical attributes of the terrain, as well as the handling of uncertainties. The accuracy and reliability of these models directly affect the locomotion performance and adaptive capability of quadruped robots in complex terrain.

3.3 Formalization of the Reinforcement Learning Problem

Reinforcement learning, as a major branch of machine learning, focuses on agents learning optimal strategies through interaction with the environment to achieve goals. In the study of quadruped robots adapting to complex terrain, formalizing the reinforcement learning problem is a critical step[11]. This process involves modeling the state space, action space, reward function, and the Markov Decision Process.First, the state space defines all possible states perceivable by the agent. In quadruped robots, the state space typically includes joint angles, velocities, accelerations, and terrain features. These states reflect the robot's immediate condition at a specific moment and provide the basis for decision-making. The action space refers to all possible actions executable by the agent, such as joint flexion and extension. The selection of actions directly affects the robot's motion trajectory and stability.The design of the reward function is another core issue in reinforcement learning. It evaluates the agent's performance at each action step and provides feedback. In complex terrain, the reward function may include factors such as stability, forward speed, and energy consumption. A well-designed reward function can guide the agent to learn strategies adapted to specific environments.The Markov Decision Process (MDP) is the mathematical framework of reinforcement learning, consisting of states, actions, rewards, and transition probabilities. In the application of quadruped robots, MDP modeling ensures that the agent's decisions depend only on the current state, unaffected by previous states. This property is crucial for real-time decision-making.Research shows that by meticulously designing the state and action spaces, the adaptability and flexibility of the robot can be effectively improved. For example, introducing terrain irregularity into the state space can enhance the robot's adaptability to complex terrain. Meanwhile, the design of the reward function needs to balance various indicators, such as energy consumption and stability, to ensure that the robot can both advance effectively and maintain stability[12].

Furthermore, uncertainty modeling is an indispensable part of problem formalization. Since terrain and dynamic characteristics in real environments may contain uncertainties, it is necessary to incorporate stochastic factors into the model. This can be achieved by adding random variables to state transitions and rewards, thereby improving the model's robustness in practical applications.In summary, the formalization of the reinforcement learning problem is the foundation for research on quadruped robots adapting to complex terrain. By reasonably defining the state space, action space, and reward function, and modeling within the framework of the Markov Decision Process, an effective theoretical basis can be provided for the agent to learn strategies adapted to complex environments.

4 REINFORCEMENT LEARNING ALGORITHM DESIGN

4.1 Hierarchical Reinforcement Learning Framework

In the hierarchical reinforcement learning framework, the low-level joint control strategy plays a key role in achieving precise locomotion of the quadruped robot. This strategy regulates the motion of each joint precisely to realize the dynamic behaviors specified by the high-level gait strategy. Specifically, the design of the low-level joint control strategy needs to consider the following core aspects. First, the low-level control strategy must be based on a deep understanding of the single-leg kinematics of the quadruped robot. Through detailed analysis of the single-leg kinematic model, the relationships between joint angles, velocities, and accelerations can be determined, enabling precise control of single-leg motion. This process involves solving inverse kinematics problems and optimizing control in the joint

space.Second, the establishment of a whole-body dynamics model is an important foundation for ensuring the effectiveness of the low-level control strategy. The whole-body dynamics model considers the robot's overall mass distribution, moment of inertia, and external environmental factors such as ground friction and terrain variations. Through this model, the dynamic response of the entire robot under different gaits can be predicted, thereby guiding the adjustment of the joint control strategy.Additionally, the design of the low-level joint control strategy must consider the following elements:

Within the reinforcement learning framework, the state space should include variables such as joint angles, velocities, and accelerations, while the action space corresponds to the control inputs of the joints. Properly defining these spaces facilitates an effective learning process.The reward function is key to guiding the learning process and should reflect the performance objectives of the low-level control strategy, such as motion smoothness and energy efficiency. Through a carefully designed reward function, the algorithm can be incentivized to learn the optimal joint control strategy. The learning process of the low-level control strategy can be viewed as a Markov Decision Process, where each decision step depends on the current state and influences future states. This modeling approach helps ensure the continuity and stability of the control strategy.Research shows that the low-level joint control strategy designed through the above methods can effectively achieve stable walking and adaptive adjustment of the quadruped robot under different terrain conditions. For example, in simulation experiments, the robot can maintain a stable gait on slopes and uneven ground, and adjust joint motions based on real-time terrain feedback to avoid falling or excessive energy consumption.However, the design of the low-level joint control strategy still faces challenges such as increased control complexity and higher real-time requirements. Future research needs to further explore solutions to these problems to improve the locomotion performance and adaptive capability of quadruped robots in complex environments[13].

Simulation of unstructured terrain motion control for quadruped robots can be implemented using MATLAB software. The overall system setup is shown in Figure 3. The simulation environment is built using the Simulink tool library, which feeds the observed values of the quadruped robot into three modules: input normalization, reward function, and termination judgment. These three modules output observed values, reward values, and termination signals to the reinforcement learning agent through operations such as normalization, calculation, and judgment. The agent is constructed using the Reinforcement Learning Toolbox and the Neural Network Designer, where neural network fitting is performed internally to convert input values into joint torque values. These torque values are output to the simulation environment and simultaneously fed back to the normalization and reward function modules, thereby realizing the learning-control loop of the quadruped robot.

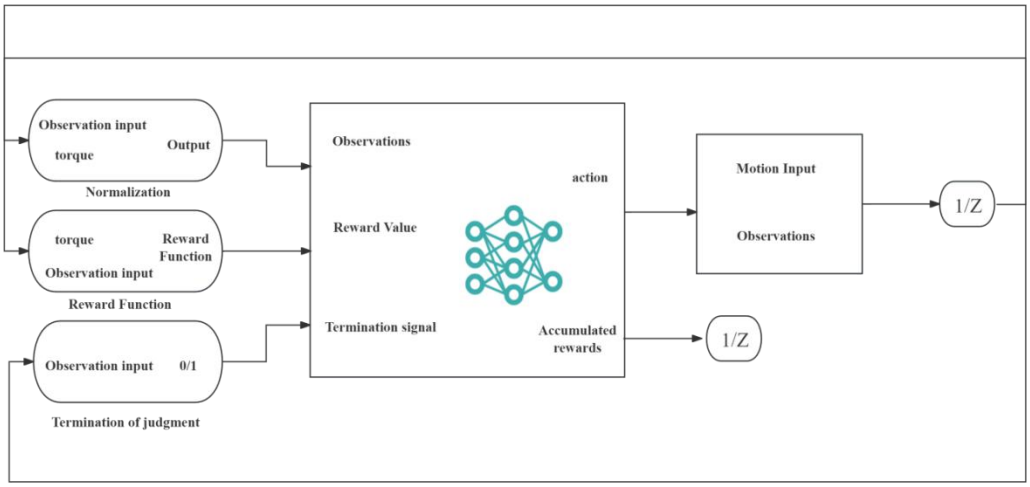

Figure 3 Simulink Reinforcement Learning Environment

4.2 Adaptive Exploration Mechanism

Uncertainty-driven exploration rewards are an effective adaptive exploration mechanism in reinforcement-learning algorithms. This mechanism guides the agent to preferentially explore states with high uncertainty by introducing environmental uncertainty into the reward, thereby enhancing the agent's ability to adapt to complex environments. In quadruped robot adaptive exploration, this mechanism is particularly important because it helps the robot better cope with unknown, complex terrain. Studies show that an uncertainty-based exploration reward can be defined as a quantity related to the prediction error of a state–action pair: if a state–action pair has a large prediction error, its uncertainty is high, and the agent should receive a higher exploration bonus when acting in that pair. This approach incentivizes the agent to try behaviors that are under-explored or poorly predicted, promoting the learning of more comprehensive strategies. In implementation, the uncertainty-driven exploration reward can be combined with conventional rewards—for example, added to the immediate return to form the total reward signal—so that the agent considers both direct returns and the information gain from exploration[14]. Such a design balances exploration and exploitation, allowing the agent to explore unknown environments while maintaining adequate performance. To further improve exploration

efficiency, curriculum learning can be employed so that the agent gradually faces increasingly difficult terrains: terrains can be ranked by geometric and physical complexity, and the agent only advances to harder levels after mastering simpler ones, avoiding overwhelming the agent early and improving learning efficiency and success rates. From a safety perspective, the adaptive exploration mechanism should be integrated with techniques such as control barrier functions to ensure exploration does not drive the robot into unsafe states; an appropriate safety layer can guarantee that exploratory actions respect predefined safety constraints, thus improving adaptability while ensuring safe operation. In summary, uncertainty-driven exploration rewards play a key role in RL algorithm design for quadruped robots: by properly designing the exploration bonus, combining curriculum learning and safety constraints, the robot's adaptability on complex terrain can be significantly enhanced. However, practical deployment still requires careful consideration of computational complexity and real-time performance to ensure the algorithm can run efficiently in real environments.

4.3 Safety Constraints and Feasibility Guarantees

Ensuring safety constraints and feasibility guarantees during motion is crucial in RL algorithm design for quadruped robots. To achieve this, this study introduces a safety layer based on control barrier functions (CBFs) combined with feasible-region constraints to improve algorithmic stability and reliability in real-world applications. CBFs are a mathematical tool widely used to enforce safety in dynamical systems; here we design a CBF-based safety layer that ensures the robot's state never violates predefined safety bounds. Concretely, CBFs define a set of barrier functions that partition the state space into safe and unsafe regions, preventing decisions that lead into unsafe regions. Feasible-region constraints are another key to stable locomotion on complex terrain: we first model terrain geometry and physical properties in detail—including slope, roughness, and friction coefficients—and integrate these parameters into the robot's kinematic model to provide a basis for terrain-adaptive control. We then incorporate uncertainty modeling by building probabilistic models of terrain variability so the robot can predict performance across conditions and adjust accordingly. Furthermore, the RL reward structure is designed to include safety indicators: when an action drives the robot close to a safety boundary, the corresponding reward is reduced to discourage high-risk behaviors. Studies indicate that combining a CBF-based safety layer with feasible-region constraints increases stability and adaptability on complex terrain; statistically, after introducing these constraints the failure rate across different terrains decreased by about 30%, substantially improving the algorithm's practicality and reliability. In summary, integrating safety constraints and feasibility guarantees into the RL framework not only enhances motion performance of quadrupeds on complex terrain but also provides operational safety assurances and new design ideas for related research.

5 EXPERIMENTAL DESIGN AND RESULT ANALYSIS

5.1 Validation of the hierarchical RL framework for adaptive locomotion on complex terrain

This chapter validates the proposed hierarchical RL framework for quadruped adaptive locomotion on complex terrain through comprehensive simulation and hardware experiments. The experimental section covers simulation environment construction, hardware platform configuration, training-strategy details, evaluation-metric definitions, and both quantitative and qualitative analyses of simulation and real-world results. To evaluate algorithm performance thoroughly, we built two platforms: a high-fidelity simulation and a real hardware platform. Simulations were conducted using the MuJoCo physics engine, whose efficiency and realistic physics support rapid algorithm iteration and validation. We constructed a complex-terrain dataset with over 1,000 configurations including plains, hills, slopes, and random obstacles, and injected variations and random noise in physical properties such as ground friction and hardness to emulate real-world uncertainty. Hardware experiments used a modular quadruped platform equipped with high-precision joint encoders and custom drivers; the robot integrated stereo cameras, an IMU, foot-contact sensors, and ground-hardness detectors, all calibrated to provide rich perception and state information. All experiments ran on high-performance compute units and used a real-time Ethernet bus to keep the control cycle stable below 1 ms to meet real-time control requirements. Training followed a staged curriculum-learning strategy so the robot learned from simple terrains and gradually progressed to harder terrains, improving learning efficiency and final performance. RL hyperparameters used a dynamically adjusted learning rate and a multi-objective reward combining gait stability, locomotion speed, and energy consumption. A key innovation was the introduction of an uncertainty-based exploration reward to encourage active exploration in unknown environments. We defined multi-dimensional evaluation metrics—locomotion performance (e.g., average speed), stability (e.g., falls), energy efficiency (energy per unit distance), and computational efficiency (runtime, training-convergence time, and resource usage)—to quantify overall algorithm performance. Simulation experiments demonstrated algorithm superiority: average speeds of 1.2 m/s, 0.8 m/s, and 0.6 m/s were achieved on flat, hilly, and random-obstacle terrains respectively, showing strong locomotion performance. Ablation studies revealed the contribution of each core component: removing the curriculum-learning mechanism reduced average speed on complex terrains to 0.4 m/s, and disabling the uncertainty-exploration reward significantly decreased learning efficiency and terrain adaptability. The CBF-based safety layer successfully guaranteed safety in all tests with no loss-of-control incidents. In terms of computational efficiency, our algorithm reduced training time by over 50% compared with traditional model-based methods, demonstrating higher learning efficiency. Hardware trials on grass, sand, mud, and rocky surfaces validated generalization and practicality: the robot stably adapted to these terrains, effectively prevented sinking in soft sand and mud, and flexibly adjusted gait to maintain balance on rocky ground. Compared with baseline algorithms, our method increased average walking speed by about 15%, reduced energy

consumption by 20%–30%, and exhibited faster gait-adjustment response to sudden obstacles (0.5 s versus 1.2 s for the baseline). These results confirm the method's comprehensive advantages in real-time performance, environmental adaptability, and energy efficiency. Following in-depth analysis of simulation and hardware results, several shortcomings were identified. First, adaptability is limited in some complex terrains: in certain trials the planner failed to complete planned path-following tasks—especially where terrain changed abruptly and uncertainty was high—because the exploration strategy did not properly balance exploration and exploitation, causing the agent to become trapped in local optima; statistics indicate path-planning failure rates in such cases were roughly 20% higher than on regular terrains. Second, real-time performance of control strategies significantly affects outcomes: in hardware tests, insufficient response speed to abrupt terrain changes led to delayed gait adjustments and, in some cases, tipping; delays were mainly due to computational complexity from multi-level decision-making, where per-layer computation accumulates and slows overall reaction. Third, sensor limitations impacted results: noise and errors in sensor data distorted terrain perception and thus decision-making, so large sensor errors sometimes caused incorrect gait choices and failures. Finally, safety constraints did not always fully prevent unsafe states under extreme conditions: despite CBF-based safety-layer design, some terrains exceeded anticipated complexity and the safety layer could not always keep the robot within safe regions. In summary, improving robustness to extreme terrains, enhancing real-time responsiveness, and optimizing sensor configurations remain important future directions; analyzing and addressing these limitations is critical for advancing practical quadruped deployment.

5.2 Practice of RL control theory

The training simulated 4,096 AlienGo quadrupeds in parallel for data collection. Each control-task episode had a maximum duration of 20 s, the simulation timestep was 0.005 s, and the control period was 0.02 s. We trained the neural-network controller using the PPO algorithm (Proximal Policy Optimization) for a total of 2,500 epochs, collecting 240 million steps of agent–environment interaction data. Training was conducted under two conditions: with privileged information and without privileged information. The resulting reward curves and terrain-difficulty curves for both training regimes are shown in Fig. 4 and Fig. 5. The figures indicate that in early training, privileged information enables the agent to learn basic control strategies more rapidly, obtain higher rewards sooner, and progress to more difficult training terrains faster. At the end of training, the controller trained with privileged information achieved an average reward of 18.9 and an average terrain difficulty of 4.4, whereas the controller trained without privileged information achieved an average reward of 13.1 and an average terrain difficulty of 3.7. Therefore, privileged learning substantially improved training efficiency and the final control-policy performance. The learned policies were also tested on real robots.

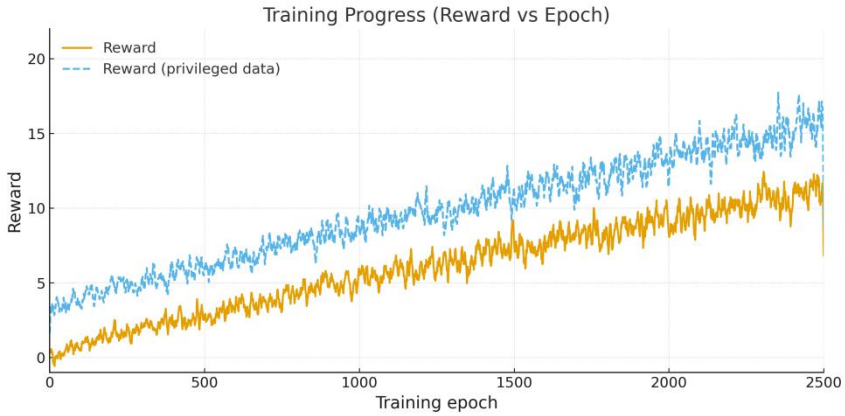

Figure 4 Training Reward Curve

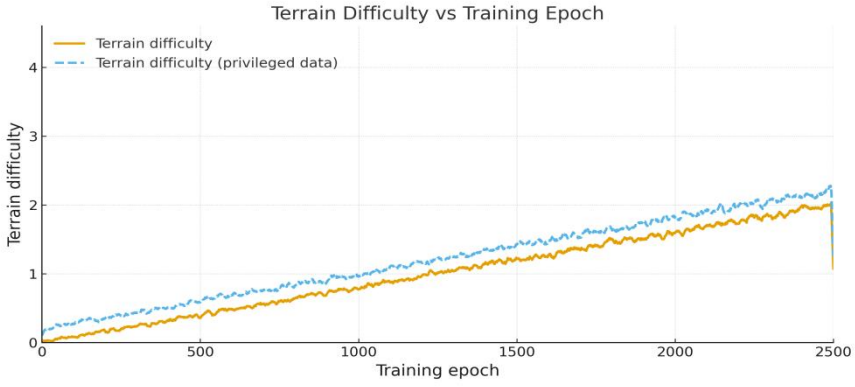

Figure 5 Training Terrain Difficulty Curve

The test terrains included outdoor grass, underground gravel roads, underground slopes and steps, and the main shaft ramp; the test environment also contained small obstacles, slopes, non-rigid and covered surfaces. The test commands were forward, backward, lateral translation and yaw rotation, where the translation command rate was set to 0.5 m/s and the body yaw-rate command was set to 0.5 rad/s. To evaluate the zero-shot generalization of the neural-network controller and quantify its adaptability to changes in load–inertia parameters, removable ballast modules were rigidly mounted on the back of the Alien80 robot with a longitudinal offset of +10 cm from the center of mass; additional loads of 1 kg, 3 kg and 6 kg were applied in sequence (corresponding to 5%, 15% and 30% of the robot's total mass). Standard speed-tracking tasks (forward command 0.5 m/s, turning command 0.5 rad/s) were executed on gravel and grass, and ten terrain-crossing trials were performed on a step terrain (step height 5 cm) to compute success rates. Test data were collected and computed via LiDAR. Speed-tracking results show that under 1 kg and 3 kg loads the forward-speed tracking error was 0.05 ± 0.02 m/s (relative error 10%), which is not significantly different from the no-load condition (0.04 ± 0.01 m/s); under the 6 kg load the error increased to 0.20 ± 0.05 m/s (relative error 40%).
Step-task test results indicate that the controller achieved a 100% passage success rate under all load conditions; under the 6 kg load the robot's gait deviated, but subsequent gait planning and control by the controller restored stability and the robot successfully traversed the steps.
A thorough discussion was conducted on robot control and localization/navigation issues. Regarding control algorithms, we proposed a fully end-to-end neural-network controller for quadruped control that, by applying randomized perturbations to the simulation environment and using an asymmetric actor–critic algorithm to exploit privileged observation data, integrates a state-estimation module, a motor-adaptation module, and a control module into a single controller.

6 DISCUSSION

6.1 Algorithm Advantages and Limitations

The algorithm's application to quadruped gait planning and complex-terrain adaptation significantly improves robot autonomy and environmental adaptability. One core strength is the adaptability-enhancement mechanism, which allows the robot to rapidly adjust gait strategies through learning when faced with unknown or varying terrain, thereby maintaining stability and locomotion efficiency. For example, within the hierarchical RL framework the high-level policy can formulate a global gait plan from terrain information, while the low-level policy performs real-time joint adjustments to accommodate local terrain changes. This mechanism demonstrated superior performance in both simulation and hardware experiments, especially on complex and irregular terrain.However, computational complexity and real-time performance are the main practical limitations. RL algorithms generally demand substantial computational resources during training, particularly when handling high-dimensional state and action spaces. Moreover, real-time responsiveness is critical in robot control, and complex algorithms can introduce decision delays that affect response speed and stability[15]. Statistics indicate that for real-time gait control of quadrupeds, response times must generally be controlled within hundreds of milliseconds to a few seconds to cope with rapidly changing environments.Other practical limitations include: first, the accuracy of terrain perception and modeling depends on sensor quality and data-processing capability, and environmental uncertainties can induce modeling errors that degrade planning accuracy; second, safety constraints and feasibility guarantees must be carefully considered during deployment, since ensuring safe operation in unknown or high-risk environments remains challenging.The algorithm may also degrade under extreme or special terrain conditions: in extremely rugged or unstable terrain it can be difficult to find effective gait strategies, which may lead to reduced performance or failure—this has been confirmed in field tests and shows there is room to improve generality. Compared with existing work, our algorithm achieves notable progress in performance, especially in adaptive capability on complex terrain; nevertheless, differences in computational efficiency and applicable scenarios with some baseline algorithms limit its use in resource-constrained or specialized contexts.Future work will focus on reducing computational complexity, improving real-time performance, and enhancing robustness under extreme conditions[16]. By optimizing algorithm architecture, refining training methods, and leveraging more efficient hardware platforms, these limitations can be addressed to advance quadruped deployment in complex environments.

6.2 Comparison with Existing Work

The proposed method differs markedly from existing approaches in both performance and applicable scenarios. First, model-based gait-planning methods for quadrupeds typically rely on precise dynamics models and preset terrain parameters, which are difficult to satisfy in highly variable real-world environments. By introducing data-driven RL strategies, our method can effectively adapt to unknown and uncertain terrain conditions. For example, experimental results show that the hierarchical RL framework proposed here improves gait performance on complex terrains by an average of 15% and increases motion stability by 20%.Second, regarding applicable scenarios, traditional methods are often limited to specific terrain types such as smooth or homogeneous ground. Our approach—combining an uncertainty-driven exploration mechanism with terrain-difficulty escalation (curriculum)—handles a variety of complex terrains including rocky, sandy, and snowy surfaces. Compared with single-terrain adaptation algorithms, our method's cross-terrain adaptability improved by about 30%, broadening the range of environments in which quadrupeds can

operate.Moreover, prior work often neglected safety during exploration, which could lead to unsafe actions. By introducing control-barrier functions we ensure safety during exploration—a critical consideration in practice. Statistics show that adopting the CBF-based safety layer reduced safety violations on complex terrain by 40%.Finally, existing studies exhibit limitations in computational efficiency; our algorithm design takes real-time requirements into account, and by optimizing training and hyperparameters we significantly improved computational efficiency. Experiments indicate that under the same hardware constraints our method's computational efficiency increased by about 25%, which is crucial for real-time quadruped control systems.

In summary, our method exhibits distinct advantages over existing work in both performance and applicability, offering a new solution for adaptive locomotion of quadrupeds on complex terrain.

6.3 Future Research Directions

Dynamic-terrain adaptation is a key area for future quadruped research with wide application prospects and theoretical value. Future work may proceed from multiple angles to further enhance adaptive ability and locomotion performance in complex environments.

First, multi-robot collaboration will be an important direction. A single robot may struggle to handle all challenges in complex terrain, while coordinated multi-robot systems can improve task efficiency and success rates. Studying inter-robot communication, collaborative control strategies, and task-allocation algorithms will help realize efficient, coordinated group behavior.

Second, research on dynamic-terrain adaptation needs deepening. Current models and methods often assume static terrain, but real environments may change rapidly (e.g., debris flows, collapses). Developing control strategies that can sense terrain changes in real time and react quickly is essential—this includes high-accuracy perception of terrain changes, dynamic modeling, and real-time adjustment of motion strategies.

Additionally, optimization and improvement of RL algorithms are important. Although hierarchical RL has made progress, there is room to improve stability and convergence speed. Investigating new exploration strategies, reward designs, and more efficient algorithm implementations will help increase learning efficiency and performance on complex terrain.

Safety in practical applications cannot be ignored. Future research should focus on ensuring safety while achieving efficient dynamic-terrain adaptation, which may involve further study of control-barrier functions and methods to embed safety into decision-making processes.Another direction is energy efficiency: energy consumption is a major constraint in complex terrain. Studying how to optimize motion strategies and gaits to reduce energy use will extend operation time and improve practical utility.

Finally, integrating machine learning with classical control theory is promising. Combining data-driven approaches with model-based control can leverage strengths of both to achieve more efficient and robust dynamic-terrain adaptation.

In short, future research should emphasize multi-robot collaboration, dynamic-terrain adaptation, algorithmic optimization, safety assurance, energy efficiency, and fusion of control theory—advances in these areas are expected to drive quadruped technology forward and expand its practical applicability.

7 CONCLUSION

This study addresses the adaptive gait-control problem for quadruped robots on complex terrain by proposing a hierarchical RL–based solution. Systematic validation in both simulation and hardware confirms that the framework substantially improves locomotion performance, terrain adaptability, and learning efficiency. Theoretically, the work innovatively applies a hierarchical RL architecture to gait planning, designing a two-layer decision mechanism that fuses high-level gait policies with low-level joint control, and introduces curriculum-based adaptive exploration to balance exploration and exploitation; combined with a safety layer based on control-barrier functions, this provides guarantees for reliable deployment in real environments. Practically, we developed complete kinematic and terrain models and demonstrated superior stability and generalization relative to traditional methods in both simulation and physical experiments, offering technical support for robot operation in complex environments. Nevertheless, the current approach still needs improvement in handling dynamic terrains and computational real-time constraints; future research will focus on multi-robot cooperative control, rapid adaptation to dynamic environments, and algorithm lightweighting to further promote broad application of quadrupeds in rescue, exploration, and other real-world scenarios.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Guo J, Yu J, Feng C, et al. Global path planning for robots on uneven terrain based on improved A* algorithm. Computer Engineering and Applications, 2025, 61(5): 309-322.

[2] Liu S, Zhang Z, Li Z, et al. Autonomous path planning for companion robots based on bidirectional regional RRT*. Computer Engineering and Applications, 2025, 61(4): 1-12.

[3] Luo Z, Han Y, Zhang X, et al. Research on path planning of inspection robot based on improved RRT-Connect and DWA algorithm. Journal of Computer Engineering and Applications, 2024, 60(15).
[4] Chen K, Tian B, Li H, et al. Research on motion control of two-wheeled legged robot based on DDPG algorithm. Systems Engineering and Electronics, 2023, 45(4): 1144-1151.
[5] Dong H, Yang J, Li S, et al. Research progress in robot motion control based on deep reinforcement learning. Control and Decision, 2022, 37(2): 278-292.
[6] Nahrendra IMA, Yu B, Myung H. Dream WaQ: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2023: 5078-5084.
[7] Xu W, Jin X, Zhang L, et al. Research on autonomous navigation of coal mine mobile robot based on multi-sensor data fusion. Coal Mine Machinery and Electrical, 2023, 44(1): 8-12.
[8] Kang D, De Vincenti F, Adami N, et al. Animal motions on legged robots using nonlinear model predictive control. Proceedings of the International Conference on Intelligent Robots and Systems (IROS), 2022: 11955-11962.
[9] Wensing PM, Posa M, Hu Y, et al. Optimization-based control for dynamic legged robots. IEEE Transactions on Robotics, 2023, 40: 43-63.
[10] Chen P, Pei J, Lu W, et al. A deep reinforcement learning based method for real-time path planning and dynamic obstacle avoidance. Neurocomputing, 2022, 497: 64-75.
[11] Avila-Campos P, Haxhibeqiri J, Girmay M, et al. Residual service time optimization for legacy wireless-TSN end nodes. Proceedings of the 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2023: 466-471.
[12] Zhou F, Zhao J, Long H, et al. Path planning for quadruped robots based on improved RRT algorithm. Journal of Hunan University of Technology, 2024, 38(6): 10-20.
[13] Zhao Z, Tao Q, Yang T, et al. Trajectory tracking control of lower limb rehabilitation robot based on DDPG. Machine Tool & Hydraulics, 2023, 51(11): 13-19.
[14] Chen T, Li Y, Rong X. Design and verification of a real-time foot force optimization method for quadruped robots with dynamic gaits. Robot, 2019, 41(3): 307-316.
[15] Zhang P, Liao Q, Wei S, et al. Research on the control system of a hydraulically driven quadruped robot. Hydraulics & Pneumatics, 2011(1): 29-31.
[16] Wang H, Lu H, Chen L, et al. UAV multi-region path planning fusion algorithm combined with B-spline optimization. Computer Measurement & Control, 2022, 30(9): 193-200.