

Disturbance-Aware Adaptive Compensation in Hybrid Force-Position Locomotion Policy for Legged Robots

Yang Zhang¹, Zhanxiang Cao^{1,2}, Buqing Nie¹, Yangqing Fu¹, Haoyang Li^{1,2},
Zheng Zhang^{1,2}, Yizhi Chen², Xiaokang Yang¹, and Yue Gao^{1,2†}

Abstract— Reinforcement Learning (RL)-based methods have significantly improved the locomotion performance of legged robots. However, these motion policies face significant challenges when deployed in the real world. Robots operating in uncertain environments struggle to adapt to payload variations and external disturbances, resulting in severe degradation of motion performance. In this work, we propose a novel Hybrid Force-Position Locomotion Policy (HFPLP) learning framework, where the action space of the policy is defined as a combination of target joint positions and feedforward torques, enabling the robot to respond rapidly to payload variations and external disturbances. In addition, the proposed Disturbance-Aware Adaptive Compensation (DAAC) provides compensation actions in the torque space based on external disturbance estimation, enhancing the robot’s adaptability to dynamic environmental changes. We validate our approach in both simulation and real-world deployment, demonstrating that it outperforms existing methods in carrying payloads and resisting disturbances.

I. INTRODUCTION

Compared to wheeled and tracked robots, legged robots demonstrate advanced adaptability in unstructured environments, enabling more complex locomotion behaviors such as walking [1], jumping [2], and climbing [3]. This remarkable flexibility allows them to navigate narrow and obstacle-dense terrains more effectively, making them highly suitable for applications in rescue missions [4] and space exploration [5]. In recent years, Reinforcement Learning (RL)-based methods have made significant progress in the field of motion control of legged robots [6], [7]. Unlike traditional control methods that rely on precise modeling [8], [9], RL can automatically acquire highly dynamic motion skills through extensive simulation training. However, RL-based motion policies are highly sensitive to payload variations and external disturbances in the real world, posing significant challenges for robots operating in complex environments [10], [11]. Due to the complexity of working environments and the diversity of tasks, robots inevitably encounter external disturbances during the execution of the task [5], [12]. Therefore, enhancing the disturbance rejection capability of motion policies is essential to ensure the efficient and reliable operation of robots in practical applications [13], [14].

External disturbances typically act on the robotic system in the form of forces, leading to changes in system dynamics

and affecting its stability [15]. However, existing RL-based motion policies for legged robots primarily use target joint positions as the action space [1]–[3]. The changes in joint positions caused by disturbances can only indirectly reflect the influence of disturbances, but cannot accurately capture the impact of force and torque changes on system stability [16]. This inherent delay undermines the real-time responsiveness and effectiveness of disturbance rejection [17]. In contrast, operating in force space provides a more direct representation of the sources of disturbances and their influence on robot dynamics [18]. Inspired by the classical model-based control framework, which integrates feedforward and feedback mechanisms to compute target joint torques [8], [9], we propose a novel learning framework. This framework incorporates feedforward torques and defines the action space of the policy as a combination of the target joint positions and the feedforward torques. Compared to policies that rely on joint positions [6], feedforward torques directly account for the impact of robot dynamics and the external environment. This enables the robot to respond rapidly to payload variations and external disturbances, improving its adaptability to dynamic environmental changes [19].

In addition, to improve the robustness of motion policies under environmental disturbances, existing methods typically introduce random external forces during training to learn robust policies applicable to various disturbance conditions [20], [21]. However, policies trained using this approach lack the ability to dynamically perceive and actively compensate for unknown disturbances [22]. Furthermore, for the sake of global optimality, these policies generally exhibit conservatism, resulting in a significant decrease in motion performance in undisturbed environments [23]. In contrast, an effective solution is to use proprioceptive sensory data to estimate external disturbance information and develop an adaptive compensation policy that actively adjusts the control inputs to compensate for the disturbance [24]. This approach has been extensively studied in the model-based control field, demonstrating promising performance [25]. However, in practical applications, model-based approaches face several challenges, such as the complexity of constructing non-linear models, high computational cost, and limited scalability [26]. Therefore, we propose a learning-based approach that leverages the powerful function approximation capability of Deep Neural Networks (DNNs) to explicitly estimate external disturbance and train an adaptive compensation policy in the joint torque space, enabling real-time perception and active compensation of unknown disturbances.

This work was supported by the National Natural Science Foundation of China (Grant No. 62373242 and No. 92248303).

¹ MoE Key Lab of Artificial Intelligence and AI Institute, Shanghai Jiao Tong University, Shanghai, China.

² Shanghai Innovation Institute, Shanghai, China.

† Corresponding author. Email: yuegao@sjtu.edu.cn

In this paper, we propose a novel Hybrid Force-Position Locomotion Policy (HFPLP) learning framework to improve the robustness of legged robots under external disturbances. By incorporating feedforward torques into the action space, HFPLP reduces the reliance on the accuracy of the actuator PD control model [11] and significantly improves its dynamic response to payload variations and external impacts. In addition, we design a Disturbance-Aware Adaptive Compensation (DAAC) mechanism, which explicitly estimates external disturbance and incorporates an active compensation policy in the joint torque space to further optimize the robot's adaptability to unknown disturbances. Extensive evaluations on the Unitree Go2 quadruped robot [27] across diverse rough terrains under concurrent perturbations demonstrate that our method markedly enhances locomotion robustness in complex environments.

In summary, the contributions of this work are as follows:

- A novel hybrid force-position locomotion policy learning framework is proposed, which incorporates feedforward torque into the action space to improve the dynamic response of the policy to external disturbances.
- A disturbance-aware adaptive mechanism is designed, integrating explicit disturbance estimation with active torque-space compensation to improve robot adaptability against unknown disturbances.
- Extensive experimental results on the real robot demonstrate that our method significantly enhances the robot's payload capacity and improves its robustness against external disturbances in complex environments.

II. RELATED WORK

A. Action Space of RL-based Locomotion Policy

In recent years, methods that use target joint positions as the action space have been widely applied in the motion learning of legged robots and have demonstrated outstanding performance in various tasks [2], [3], [6], [7]. However, this paradigm is highly dependent on the accuracy of the actuator model [11]. Discrepancies in PD gains and the sim-to-real actuator model mismatch significantly impact real-robot locomotion performance [23], [28]. On the other hand, researchers have explored policies that use target joint torques as the action space, which directly output torques to bypass PD controllers [18], [29]. This approach allows for a better capture of the dynamic characteristics of the robot and enhances its robustness to external disturbances [30]. However, the learning process in the torque space typically suffers from low sample efficiency and does not consistently converge to natural gaits, which limits the practical application of this method and has prevented its widespread adoption [31]. In addition, some methods define the action space of the policy as the adjustment of the parameters of the model-based controller, combining RL with the model-based controller to adapt to complex environments and tasks [32]. Bellegarda et al. [33] develop a policy to regulate the parameters of Central Pattern Generators (CPGs) to generate robust quadrupedal locomotion. Miki et al. [1] encode

temporal components in the action space to guide periodic gaits. The literature [34] introduces an RL-augmented model predictive control (MPC) framework, where the agent learns acceleration compensation to enhance the performance of the MPC. Unlike the above methods, we propose a novel HFPLP that combines the efficiency of position control in trajectory generation with the flexibility of force control in disturbance compensation [19], [35].

B. Robust Locomotion Controller for Legged Robots

Legged robot motion controllers exhibit vulnerabilities when responding to payload variations and unpredictable environmental disturbances, significantly limiting their practical application in complex scenarios [14]. To address this issue, researchers have proposed various methods to improve the robustness of motion controllers, aiming to improve the locomotion performance of legged robots in complex environments [19]. Classical model-based adaptive control methods achieve adaptability to system variations and uncertainties by estimating uncertain parameters in the robot's dynamic model online and designing adaptive laws to adjust the controller parameters accordingly [24]–[26]. Furthermore, Chen et al. [15] propose a framework based on capturability analysis to synthesize push recovery controllers, enhancing the robot recovery performance in response to impact disturbances during dynamic motion. However, these model-based approaches encounter high computational burdens and potential stability issues in practical applications [13]. In recent years, RL-based robust policies have attracted significant attention [21], [22]. Using large-scale training data and simulation environments, these approaches aim to learn optimal control policies for robots under varying disturbance conditions [36]. Hartmann et al. [20] incorporate weights for disturbance recovery capabilities into the reward function, successfully training control policies capable of adapting to external disturbances. Moreover, Robust Adversarial Reinforcement Learning (RARL) has emerged as an effective method to improve policy robustness, improving disturbance rejection performance of legged robots on complex terrains [21], [22]. However, the design of disturbance-related rewards relies heavily on expert knowledge. This paper implements explicit estimation and adaptive compensation of external disturbances in the force space, further improving the robot's adaptability to unknown disturbances. This approach offers new perspectives and possibilities for the application of legged robots in complex environments.

III. PRELIMINARIES

A. Reinforcement Learning

Learning-based motion control problems are typically formulated as a Markov Decision Process (MDP), which enables robots to learn optimal motion policies through interaction with the environment. The MDP is defined by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma)$, where \mathcal{S} represents the state space, \mathcal{A} is the action space, $\mathcal{P}(s'|s, a)$ is the state transition probability, $R(s, a)$ denotes the reward function, and $\gamma \in [0, 1)$ is the

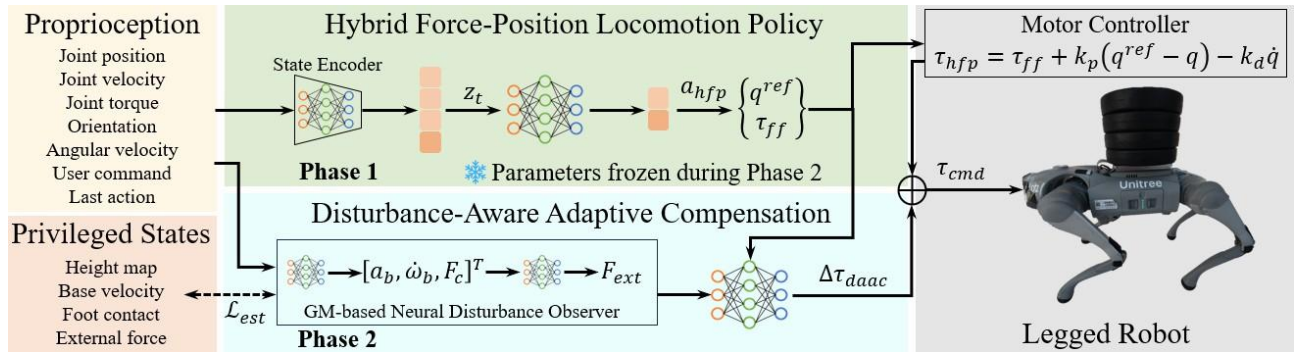


Fig. 1. Overview of the training framework. We first employ an asymmetric actor-critic architecture to train the HFPLP, enabling it to output target joint positions and feedforward torques based on proprioception and user commands. These actions are then converted into target joint torques via the motor controller to achieve robust locomotion. Subsequently, we freeze the parameters of the well-trained HFPLP and train the DAAC with a GM-based neural disturbance observer from scratch. The DAAC is designed to generate compensation torques in the joint torque space based on disturbance information, enhancing the robot’s adaptability to external disturbances.

discount factor. The robot’s state $s \in \mathcal{S}$ encodes proprioceptive information, while the action $a \in \mathcal{A}$ corresponds to the control input. The objective of the RL agent is to learn an optimal policy $\pi^*(a|s)$ that maximizes the expected cumulative reward:

$$J(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right]. \quad (1)$$

B. Dynamics of Legged Robots

Let q represents the generalized coordinates, the basic dynamics of the robot can be written as:

$$M\ddot{q} + C\dot{q} + G = S^T \tau + J_c^T F_c, \quad (2)$$

where M , $C\dot{q}$, G , S^T , τ , J_c^T , F_c are the inertia matrix, Coriolis force, gravity force, selection matrix, joint torque input, contact Jacobian matrix, and contact force, respectively. When the payload varies or external forces are applied, the system dynamics deviate from the basic model. To maintain system stability, it is typically necessary to apply additional torques to compensate for the effects of unmodeled external disturbances. The corresponding system dynamics can be represented as:

$$M\ddot{q} + C\dot{q} + G = S^T(\tau + \Delta\tau) + J_c^T F_c + J_{ext}^T F_{ext}, \quad (3)$$

where $\Delta\tau$ represents the compensation torque, F_{ext} and J_{ext} represent the external disturbances acting on the robot’s body and the corresponding Jacobian matrix. In the RL-based control framework, the policy π receives the state s as input and outputs an action a , which is then transformed by the actuators into joint torques τ to accomplish the specific task.

IV. METHODS

This section provides a detailed description of the construction and training of HFPLP and DAAC. As illustrated in Fig. 1, the framework employs a two-stage training approach. In the first stage, the HFPLP is trained using an asymmetric actor-critic structure, enabling the robot to simultaneously adjust foot position trajectories and generate

appropriate feedforward forces. In the second stage, the well-trained HFPLP parameters are frozen and integrated as part of the environment dynamics. Subsequently, the DAAC is trained from scratch. This policy employs a Generalized Momentum(GM)-based neural disturbance observer to perceive external disturbances and outputs compensation torques in the joint torque space, improving the robot’s adaptability to external disturbances.

A. Hybrid Force-Position Locomotion Policy

Legged robots are complex underactuated systems that precisely control the dynamic interaction between the foot and the environment by applying appropriate joint torques through joint actuators, achieving stable locomotion tasks. Classic model-based control methods integrate trajectory planning with robot dynamics to achieve swing leg trajectory tracking and stance leg support force optimization [35]. The robot’s joint actuator model can be expressed as:

$$\tau_i = \tau_{i,ff} + J_i^T [K_p(p_i^{ref} - p_i) + K_d(v_i^{ref} - v_i)], \quad (4)$$

where $\tau_{i,ff}$ represents the feedforward torque, J_i denotes the foot Jacobian, K_p and K_d are proportional and derivative gains, p_i^{ref} and v_i^{ref} indicate the reference foot position and velocity, p_i and v_i represent the actual foot position and velocity, $i \in \{1, \dots, n_l\}$, n_l denotes the number of legs. Typically, stance legs rely on feedforward torques to offset gravity and disturbances, while swing legs prioritize PD tracking of foot positions.

However, as discussed in Section II-A, existing RL-based motion policies focus primarily on joint positions or joint torques. These policies couple foot position tracking and foot force control within a single action space, which limits the system’s adaptability to environmental disturbances and results in poor interpretability of the actions output by motion policies [19]. In this work, we propose a novel HFPLP that combines the advantages of swing leg trajectory tracking and stance leg support force optimization. In order to implement this objective, the action space for HFPLP policy is formulated as:

$$a_{hfplp} = [q^{ref}, \tau_{ff}]^T, \quad (5)$$

where \mathbf{q}^{ref} represents the target joint position and $\boldsymbol{\tau}_{ff}$ represents the feedforward torque. The joint actuator model corresponding to this action space is:

$$\boldsymbol{\tau}_{hfp} = \boldsymbol{\tau}_{ff} + \mathbf{K}_p(\mathbf{q}^{ref} - \mathbf{q}) - \mathbf{K}_d\dot{\mathbf{q}}, \quad (6)$$

where $\boldsymbol{\tau}_{hfp}$ represents the joint torque generated by HFPLP.

B. Disturbance-Aware Adaptive Compensation

In complex environments, legged robots face external disturbances, such as unexpected foot contact events or payload variations. These disturbances alter system dynamics, causing actual trajectories to deviate from the desired paths [19]. In model-based control, the GM-based disturbance observer is widely employed to estimate external contact forces. In Eq. (3), the sum of all external forces is represented as the disturbance vector:

$$\boldsymbol{\tau}_d = \mathbf{J}_c^T \mathbf{F}_c + \mathbf{J}_{ext}^T \mathbf{F}_{ext}. \quad (7)$$

As described in [37], the disturbance estimation $\hat{\boldsymbol{\tau}}_d$ after passing through a discrete-time low-pass filter can be calculated as follows:

$$\begin{aligned} \hat{\boldsymbol{\tau}}_d &= \beta \mathbf{p} - \frac{1-\gamma}{1-\gamma z^{-1}} (\beta \mathbf{p} + \mathbf{S}^T \boldsymbol{\tau} + \mathbf{C}^T \dot{\mathbf{q}} - \mathbf{G}), \\ \mathbf{p} &= \mathbf{M} \dot{\mathbf{q}}, \beta = \frac{(1-\gamma)\gamma^{-1}}{\Delta t}, \gamma = e^{-\lambda \Delta t}, \end{aligned} \quad (8)$$

where z is the z -domain variable, Δt is the sampling period, and λ is the cutoff frequency of the filter. This method avoids the direct measurement of acceleration $\ddot{\mathbf{q}}$ by employing summation by parts and generalized momentum \mathbf{p} .

In this work, we propose a GM-based neural disturbance observer to estimate unknown external forces in real time without a precise model. These disturbances are modeled as forces at the robot's Center of Mass (CoM), neglecting torques from application point offsets. Initially, the proprioceptive history is used to estimate the acceleration vector $[\mathbf{a}_b, \dot{\boldsymbol{\omega}}_b]$ and the foot contact force \mathbf{F}_c . The acceleration vector is calculated by differencing the velocity in the simulation. Subsequently, these estimates, along with current proprioceptive observation, are fed into the disturbance observer to estimate the external disturbance force \mathbf{F}_{ext} .

In addition, we develop an adaptive compensation policy to mitigate the impact of disturbance forces on the system. Benefiting from the HFPLP introduced in Section IV-A, which exhibits sensitivity to feedforward forces, the adaptive policy can directly perform disturbance compensation in the joint torque space. As illustrated in Fig. 1, the parameters of the well-trained HFPLP are frozen, and the DAAC is learned from scratch to output the compensation torque $\Delta \boldsymbol{\tau}_{daac}$. This compensation torque is then combined with the torque output from the HFPLP to serve as the joint torque command for the robot:

$$\boldsymbol{\tau}_{cmd} = \boldsymbol{\tau}_{hfp} + \Delta \boldsymbol{\tau}_{daac}. \quad (9)$$

C. Training Details

All policies are trained using the Proximal Policy Optimization (PPO) algorithm in the Isaac Gym simulation [6] and zero-shot transferred to the Unitree Go2 quadruped robot. In the simulation, the magnitude and duration of the disturbance force are randomly sampled to mimic external forces in the real world. The 3-dim disturbance is sampled in $([-100, 100] \times [-100, 100] \times [-100, 0]) N$, and the duration time is sampled from $[1.0, 4.0] s$. Additionally, in each episode, 60% of the environments are randomly selected to experience disturbance forces.

1) *Learning the HFPLP*: The proposed HFPLP uses proprioception and base velocity command as inputs, and outputs target joint positions and feedforward torques to achieve the desired velocity tracking for the legged robot. The proprioceptive observation $\mathbf{o}_t \in \mathbb{R}^{69}$ contains body angular velocity $\boldsymbol{\omega}_t$, projected gravity vector \mathbf{g}_t , base velocity command \mathbf{c}_t , joint positions \mathbf{q}_t , joint velocities $\dot{\mathbf{q}}_t$, joint torques $\boldsymbol{\tau}_t$, and previous action $\mathbf{a}_{t-1,hfp}$. The history length $H = 5$. The critic network incorporates additional privileged information, including base velocity \mathbf{v}_b , height map \mathbf{h}_t , foot contact \mathbf{c}_f , foot clearance \mathbf{p}_f , and disturbance forces \mathbf{F}_{ext} . The action $\mathbf{a}_{t,hfp} \in \mathbb{R}^{24}$ includes target joint position \mathbf{q}^{ref} and feedforward torque $\boldsymbol{\tau}_{ff}$, which are transformed into the joint torque by Eq. (6) with $K_p = 20$ and $K_d = 0.5$. The reward functions follow the work of DreamWaQ [7], where a new reward function $(\mathbf{q}^{ref} - \mathbf{q})^2$ is added to encourage joint position tracking with a weight of -0.2 . Curriculum learning [38] is used to gradually increase the difficulty of tasks, allowing the robot to adapt to rough terrain.

2) *Learning the DAAC*: The disturbance observer is composed of two concatenated 3-layer MLP with ReLU activations. The first MLP takes in historical observations $\mathbf{o}_t^H \in \mathbb{R}^{345}$ as input and outputs $[\mathbf{a}_b, \dot{\boldsymbol{\omega}}_b, \mathbf{F}_c]^T \in \mathbb{R}^{18}$, with hidden layer [256 128 64]. The second MLP receives $[\mathbf{o}_t, \mathbf{a}_b, \dot{\boldsymbol{\omega}}_b, \mathbf{F}_c]^T \in \mathbb{R}^{87}$ as input and outputs $\mathbf{F}_{ext} \in \mathbb{R}^3$, with hidden layer [256 128 64]. The observation of the DAAC $\mathbf{o}_{t,daac} \in \mathbb{R}^{96}$ includes proprioception \mathbf{o}_t , disturbance force \mathbf{F}_{ext} , and action $\mathbf{a}_{t,hfp}$. The action $\mathbf{a}_{t,daac} \in \mathbb{R}^{12}$ represents the compensation torque $\Delta \boldsymbol{\tau}_{daac}$, which is transformed into the joint torque command by Eq. (9). The reward functions are consistent with HFPLP training.

V. EXPERIMENTS

To evaluate the locomotion performance and disturbance adaptation of the proposed method, the state-of-the-art RL-based locomotion method DreamWaQ [7] is utilized as the baseline. Comparative experiments and ablation studies are conducted to demonstrate the improvements over the baseline. The following are the baseline methods in the experiment, including ablation studies to demonstrate the necessity of some submodule of our method:

- **DreamWaQ-FT**: Fine-tuning the DreamWaQ in environments with random disturbances.
- **DreamWaQ-DAAC**: Learning the DAAC based on the DreamWaQ.

- **HFPLP-FT**: Fine-tuning the HFPLP in environments with random disturbances.
- **HFPLP-DAAC w/o DO**: Learning the DAAC without disturbance observer based on the HFPLP.
- **HFPLP-DAAC (Ours)**: Learning the DAAC based on the HFPLP.

A. Simulation Results

1) *Action Tracking*: To demonstrate the advantages of incorporating feedforward torque in HFPLP, we compare the joint position tracking curves of DreamWaQ-FT and HFPLP-DAAC. As shown in Fig. 2, DreamWaQ-FT compensates for contact forces by adjusting joint position tracking errors of the stance leg. In contrast, HFPLP-DAAC incorporates feedforward torque into its actions, significantly improving joint position tracking accuracy, demonstrating that HFPLP combines the advantages of swing leg position control and stance leg force control to effectively provide compensation torque in the stance phase.

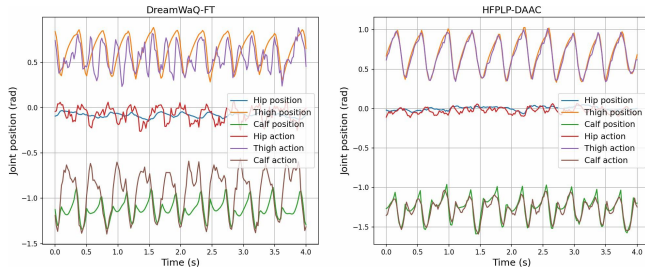


Fig. 2. The joint position tracking of the front right leg during robot motion at $v_x = 1 \text{ m/s}$. Due to the additional contact forces required by the stance leg to provide gravity compensation, the baseline exhibits significant joint position tracking errors. In contrast, our proposed method significantly improves the tracking accuracy by providing feedforward torques.

2) *Disturbance Adaptation*: In the simulation, time-varying disturbance forces are introduced to evaluate the disturbance adaptation of HFPLP-DAAC. In order to analyze the response of DAAC to disturbance, the compensation torques are mapped to foot forces using the foot Jacobian: $F_{ee} = J^{-T} \Delta \tau$. The disturbance forces along the X and Y axes are applied with magnitudes of 100 N , and their directions are changed at intervals of 5 s . Fig. 3 illustrates the applied disturbance forces, the disturbance forces estimated by the disturbance observer, and the compensation foot forces generated by the DAAC. It can be observed that the disturbance observer is able to predict disturbance forces, and compensation foot forces generated by the DAAC align with the trends of the disturbance forces. Consequently, the corresponding ground reaction forces effectively counteract the impact of the disturbance forces. This is attributed to the compensation mechanism introduced in the force space, which enhances the interpretability and transparency of the DAAC. Additionally, the velocity tracking performance of the robot under disturbance forces is shown in Fig. 4, demonstrating that the HFPLP-DAAC can enable the robot to rapidly recover to the desired motion state when subjected to significant disturbances.

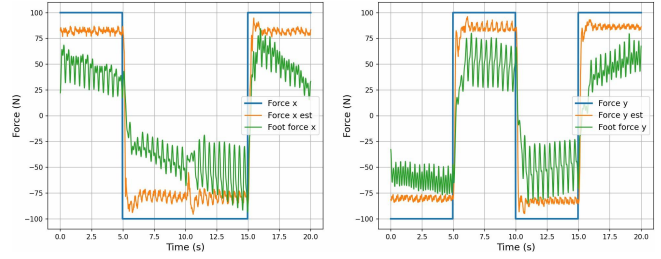


Fig. 3. Adaptation of the DAAC to disturbances. The blue line represents the applied disturbance force, the orange line represents the estimated force from the disturbance observer, and the green line represents the foot force generated by the compensation torque of the DAAC through the foot Jacobian mapping. The DAAC effectively generates compensation torque based on the disturbance force estimated by the disturbance observer.

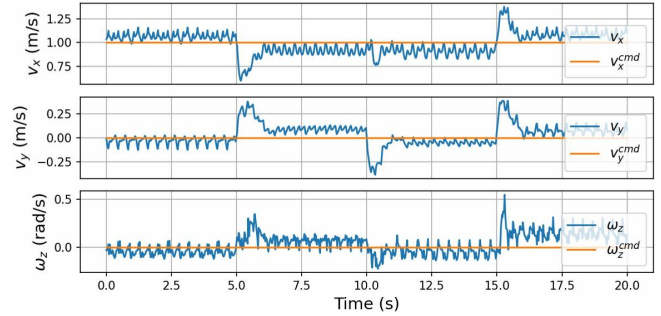


Fig. 4. Velocity tracking curve under external disturbances. The HFPLP-DAAC demonstrates the ability to rapidly recover from external disturbances and adapt to the disturbance force to maintain the desired motion state.

B. Ablation Study

To evaluate the impact of different modules on the robot's motion performance, ablation experiments are conducted. All policies are trained under identical environmental setups and hyper-parameters to ensure fairness in the results. Fig. 5 illustrates the learning curves of different policies, demonstrating that HFPLP-DAAC outperforms the baseline and other ablation versions. The comparison between DreamWaQ-FT and HFPLP-FT shows that the introduction of feedforward action space improves the response of the policy to disturbances. The comparison between HFPLP-FT and HFPLP-DAAC highlights that the DAAC leverages estimated disturbance information to output additional compensation torques, thereby enhancing the overall performance. The performance degradation observed in the absence of the disturbance observer underscores the importance of explicit disturbance estimation in improving disturbance responsiveness.

In addition, all policies are deployed on a real robot to evaluate their performance in the real world. For each policy, the Absolute Tracking Error (ATE) is calculate under three conditions: nominal condition, loaded condition, and disturbed condition with pulling disturbance. The robot follows $v_x = 1 \text{ m/s}$, $v_y = 1 \text{ m/s}$ commands sequentially, and each task is executed 5 times. The experimental setups for payload and lateral pulling disturbance are shown in Fig. 6, where the payload is 5 kg and the weight dragged by the rope is 19 kg . The average pulling force F_p exerted by the rope is approximately 40 N , measured using the

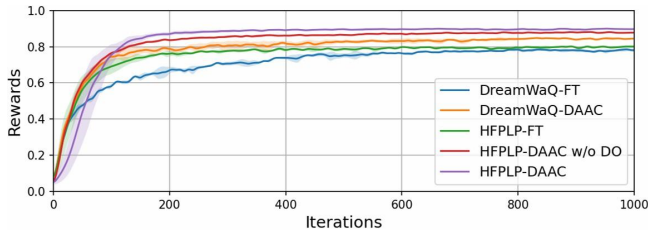


Fig. 5. Learning curves of different methods. The curves and shaded areas represent the mean and standard deviation of rewards across 10 different seeds, respectively. The HFPLP-DAAC outperforms baseline and other ablation versions on rewards, indicating higher motion performance.



Fig. 6. Diagram of the real world ablation experiment setup. Left: The robot carries a 5 kg payload. Right: The robot pulls a 19 kg weight via a rope to simulate pulling disturbance.

digital dynamometer. Fig. 7 illustrates the ATE under various environmental conditions. The experimental results demonstrate that the proposed method exhibits superior disturbance rejection capabilities compared to other policies.

C. Disturbance Adaptation on Real Robot

1) *Robustness to Actuators*: The baseline method relies on the actuator’s PD controller to convert the target joint position into the desired torques, and differences between the PD controller in simulation and the real robot will lead to sim-to-real problem. In contrast, our method reduces dependence on the PD controller through feedforward torque control, which better accommodates the actuator discrepancies between simulation and the real robot, improving motion robustness. Fig. 8 shows the motion performance of policies trained in simulation with $K_p = 20$, $K_d = 0.5$ transferred to the real robot with different PD parameters. As illustrated in the Fig. 8, our method can still maintain motion performance under large differences in parameter K_p , while the performance of the baseline method drops significantly. This indicates that our method can avoid tedious PD parameter tuning and reduce the sim-to-real gap.

2) *Heavy Payload Adaptation*: To demonstrate the adaptability of the motion policy to varying payloads, we evaluate the robot’s locomotion performance under different payloads. The robot tracks a velocity command of $v_x = 1.0 m/s$ for a duration of 10 s while carrying various payloads. For each payload, 5 trials are conducted to calculate the Success Rate (SR) and ATE. As illustrated in Fig. 9, the baseline could only handle a maximum payload of 7.5 kg, whereas our proposed method achieve remarkable performance under a payload of 20.0 kg (133% uncertainty). To the best of our knowledge, this is the first demonstration of a learning-

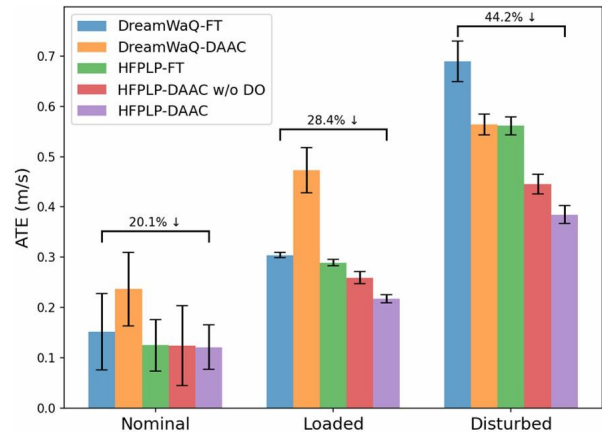


Fig. 7. Comparison of ATE under different disturbance conditions for various policies. The HFPLP-DAAC significantly improves velocity tracking performance under load and external force disturbances. Moreover, even in the absence of disturbances, our method outperforms the baseline, demonstrating superior sim-to-real performance.

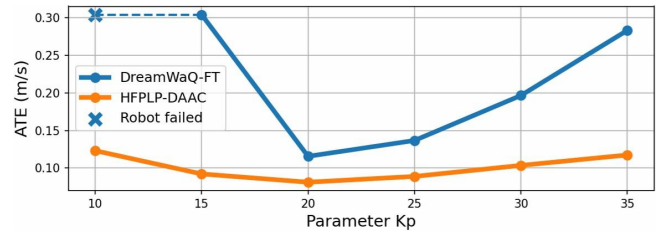


Fig. 8. The locomotion performance of the policy trained in simulation with $K_p = 20$, $K_d = 0.5$ when transferred to the real robot with different PD parameters. Compared to the baseline, HFPLP-DAAC demonstrates adaptability to different K_p values, reducing sensitivity to actuator discrepancies between simulation and real robot.

based approach enabling the real robot to handle a payload exceeding its own weight, highlighting the state-of-the-art capability of our method in payload handling. Additionally, we evaluate the joint position tracking and external force estimation under dynamically varying payloads. The robot is commanded to track a velocity of $v_x = 1.0 m/s$ while the payloads are gradually increased to [2.5, 5.0, 7.5, 10.0] kg. As shown in Fig. 10, during dynamic payload changes, joint position tracking exhibited no significant variation, and the disturbance observer effectively estimates the external force magnitude. These results demonstrate that the DAAC can accurately perceive changes in external forces and provide effective feedforward compensation torques to maintain the robot’s locomotion performance.

3) *Impact Disturbance Adaptation*: We further evaluate the adaptability of HFPLP-DAAC to impact disturbances. To ensure a controlled and quantifiable comparison, we introduce lateral impact disturbances by releasing a weight from a fixed height, allowing it to swing laterally and collide with the robot moving at a velocity of $v_x = 1.0 m/s$. Each set of experiments is conducted with 10 trials. We evaluate the performance of the algorithm using three metrics: SR, ATE, and Lateral Displacement (LD). Table I shows that our method significantly improves the ability to resist impact

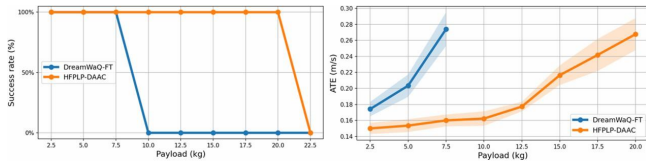


Fig. 9. The success rate and ATE of the robot when carrying different payloads. The HFPLP-DAAC demonstrates the ability to successfully track a velocity command of $v_x = 1.0 \text{ m/s}$ while carrying a payload of 20.0 kg (equivalent to 133% of its own weight), significantly enhancing the robot’s capability to perform tasks with high payloads.

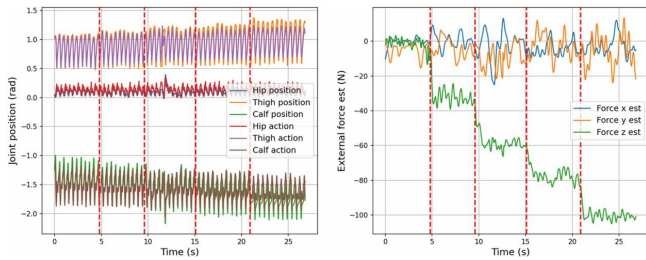


Fig. 10. Joint position tracking and external force estimation under dynamically varying payloads. The joint position tracking remains consistent despite changes in the payload, while the disturbance observer accurately estimates the external force disturbances. These results demonstrate that the HFPLP-DAAC effectively perceives variations in external forces and provides appropriate feedforward compensation torques to mitigate the impact of disturbances.

disturbances. The robot has smaller ATE and LD under larger impact disturbances, indicating that our method improves the response to impact disturbances and enables rapid recovery.

4) *Analysis of DAAC Policy:* We employ t-distributed Stochastic Neighbor Embedding (t-SNE) to analyze the actions of the DAAC under various disturbances. In the experiments, the robot is controlled to move along the positive and negative directions of the x-axis and y-axis, respectively. Following the method illustrated in Fig. 6, the robot is subjected to pulling disturbances in different directions. As shown in Fig. 11, the actions generated by the DAAC demonstrate a clear distinction between the effects of disturbances from different directions. This indicates that the DAAC effectively integrates disturbance observer to accurately estimate external forces and provides appropriate feedforward torque compensation, enabling the robot to respond effectively to external disturbances.

5) *Traversability in Complex Environments:* We demonstrate the locomotion performance of the HFPLP-DAAC under various challenging terrains and external disturbances. As shown in Fig. 12, the robot is capable of successfully traversing unstructured terrains such as sandy surfaces, slippery gravel, variable sponge mats, and stair under payload conditions, while effectively handling disturbances such as unexpected collisions and foot slippage. The experimental results indicate that our approach significantly enhances the robot’s ability to traverse complex environments with heavy payloads and improves task execution stability. Further experimental details can be found in the supplementary video.

TABLE I
COMPARISON OF LOCOMOTION PERFORMANCE UNDER IMPACT DISTURBANCES

Methods	Weights	Metrics		
		SR (%)	ATE (m/s)	LD (m)
DreamWaQ-FT	2.5 kg	100	0.54 ± 0.14	0.39 ± 0.09
	5.0 kg	40	0.76 ± 0.17	0.94 ± 0.16
	7.5 kg	0	-	-
	10.0 kg	0	-	-
HFPLP-DAAC	2.5 kg	100	0.23 ± 0.06	0.10 ± 0.03
	5.0 kg	100	0.35 ± 0.10	0.23 ± 0.04
	7.5 kg	100	0.47 ± 0.09	0.41 ± 0.07
	10.0 kg	80	0.56 ± 0.12	0.57 ± 0.11

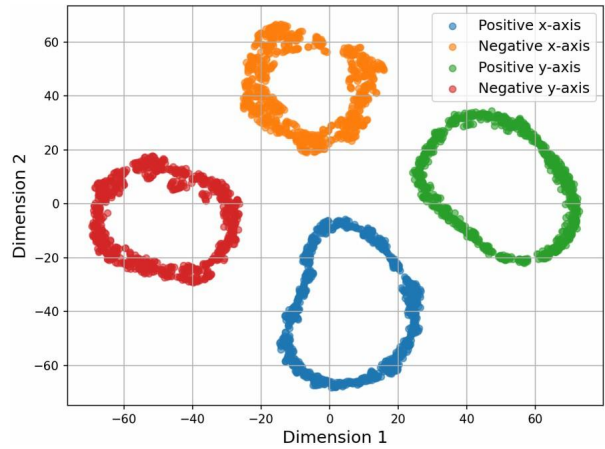


Fig. 11. The t-SNE visualization of compensation torques generated by the DAAC. The compensation torques under different directional pulling disturbances are clearly distinguishable, demonstrating that the DAAC effectively perceives the disturbances and generates appropriate compensation torques to mitigate their effects.

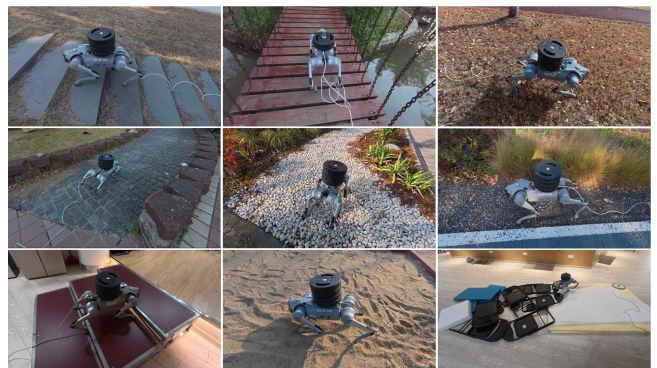


Fig. 12. Evaluating the locomotion performance on challenging terrains, including grass, stairs, gravel, soft sand, pebbles, and deformable sponge mats. HFPLP-DAAC significantly improves complex terrain traversal capabilities under payload.

VI. CONCLUSION

In this work, we propose the HFPLP, which significantly improves the dynamic response of legged robots to payload variations and external force impacts. In addition, benefiting from the sensitivity of HFPLP to the torque space, we introduce the DAAC in the torque space, further enhancing the robot’s adaptability to unknown disturbances.

Comprehensive validation using the Unitree Go2 quadruped robot demonstrates that HFPLP-DAAC exhibits superior locomotion performance under various complex terrains and disturbance conditions. However, this approach still has some limitations, such as the lack of theoretical guarantees and the disturbance model considering only forces applied to the robot's CoM. Future work will focus on exploring more complex disturbance models and further developing theoretical analysis frameworks to enhance the adaptability of legged robots to various types of disturbances.

REFERENCES

- [1] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science robotics*, vol. 7, no. 62, p. eabk2822, 2022.
- [2] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, "Extreme parkour with legged robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 11 443–11 450.
- [3] D. Vogel, R. Baines, J. Church, J. Lotzer, K. Werner, and M. Hutter, "Robust ladder climbing with a quadrupedal robot," *arXiv preprint arXiv:2409.17731*, 2024.
- [4] A. Kleiner and C. Dornhege, "Real-time localization and elevation mapping within urban search and rescue scenarios," *Journal of Field Robotics*, vol. 24, no. 8-9, pp. 723–745, 2007.
- [5] P. Arm, G. Waibel, J. Preisig, T. Tuna, R. Zhou, V. Bickel, G. Ligeza, T. Miki, F. Kehl, H. Kolvenbach *et al.*, "Scientific exploration of challenging planetary analog environments with a team of legged robots," *Science robotics*, vol. 8, no. 80, p. eade9548, 2023.
- [6] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 91–100.
- [7] I. M. A. Nahrendra, B. Yu, and H. Myung, "Dreamwaq: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5078–5084.
- [8] M. Bjelonic, R. Grandia, M. Geilinger, O. Harley, V. S. Medeiros, V. Pajovic, E. Jelavic, S. Coros, and M. Hutter, "Offline motion libraries and online mpc for advanced mobility skills," *The International Journal of Robotics Research*, vol. 41, no. 9-10, pp. 903–924, 2022.
- [9] R. Grandia, F. Jenelten, S. Yang, F. Farshidian, and M. Hutter, "Perceptive locomotion through nonlinear model-predictive control," *IEEE Transactions on Robotics*, vol. 39, no. 5, pp. 3402–3421, 2023.
- [10] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3803–3810.
- [11] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.
- [12] P. Biswal and P. K. Mohanty, "Development of quadruped walking robots: A review," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 2017–2031, 2021.
- [13] J. Lee, M. Bjelonic, A. Reske, L. Wellhausen, T. Miki, and M. Hutter, "Learning robust autonomous navigation and locomotion for wheeled-legged robots," *Science Robotics*, vol. 9, no. 89, p. eadi9641, 2024.
- [14] J. Kang, H.-b. Kim, B.-I. Ham, and K.-S. Kim, "External force adaptive control in legged robots through footstep optimization and disturbance feedback," *IEEE Access*, 2024.
- [15] H. Chen, Z. Hong, S. Yang, P. M. Wensing, and W. Zhang, "Quadruped capturability and push recovery via a switched-systems characterization of dynamic balance," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 2111–2130, 2023.
- [16] J. Queeney, X. Cai, M. Benosman, and J. P. How, "Gram: Generalization in deep rl with a robust adaptation module," *arXiv preprint arXiv:2412.04323*, 2024.
- [17] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *The International Journal of Robotics Research*, p. 02783649241285161, 2024.
- [18] S. Chen, B. Zhang, M. W. Mueller, A. Rai, and K. Sreenath, "Learning torque control for quadrupedal locomotion," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*. IEEE, 2023, pp. 1–8.
- [19] S. Lyu, X. Lang, H. Zhao, H. Zhang, P. Ding, and D. Wang, "R12ac: Reinforcement learning-based rapid online adaptive control for legged robot robust locomotion," in *Proceedings of the Robotics: Science and Systems*, 2024.
- [20] A. Hartmann, D. Kang, F. Zargarbashi, M. Zamora, and S. Coros, "Deep compliant control for legged robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 11 421–11 427.
- [21] F. Shi, C. Zhang, T. Miki, J. Lee, M. Hutter, and S. Coros, "Rethinking robustness assessment: Adversarial attacks on learning-based quadrupedal locomotion controllers," *arXiv preprint arXiv:2405.12424*, 2024.
- [22] J. Long, W. Yu, Q. Li, Z. Wang, D. Lin, and J. Pang, "Learning h-infinity locomotion control," *arXiv preprint arXiv:2404.14405*, 2024.
- [23] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.
- [24] Z. Zhu, G. Zhang, Z. Sun, T. Chen, X. Rong, A. Xie, and Y. Li, "Proprioceptive-based whole-body disturbance rejection control for dynamic motions in legged robots," *IEEE Robotics and Automation Letters*, 2023.
- [25] L. Amanzadeh, T. Chunawala, R. T. Fawcett, A. Leonessa, and K. A. Hamed, "Predictive control with indirect adaptive laws for payload transportation by quadrupedal robots," *IEEE Robotics and Automation Letters*, 2024.
- [26] M. Somboldestan and Q. Nguyen, "Adaptive force-based control of dynamic legged locomotion over uneven terrain," *IEEE Transactions on Robotics*, 2024.
- [27] Unitree Go2, <https://www.unitree.com/products/go2>, 2024, [Online; accessed 28-November-2024].
- [28] Z. Xie, X. Da, M. Van de Panne, B. Babich, and A. Garg, "Dynamics randomization revisited: A case study for quadrupedal locomotion," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4955–4961.
- [29] D. Kim, G. Berseth, M. Schwartz, and J. Park, "Torque-based deep reinforcement learning for task-and-robot agnostic learning on bipedal robots using sim-to-real transfer," *IEEE Robotics and Automation Letters*, 2023.
- [30] S. Sood, G. Sun, P. Li, and G. Sartoretti, "Decap: Decaying action priors for accelerated learning of torque-based legged locomotion policies," *arXiv preprint arXiv:2310.05714*, 2023.
- [31] J. Eßer, G. B. Margolis, O. Urbann, S. Kerner, and P. Agrawal, "Action space design in reinforcement learning for robot motor skills," in *8th Annual Conference on Robot Learning*.
- [32] S. Gangapurwala, M. Geisert, R. Orsolino, M. Fallon, and I. Havoutis, "Rloc: Terrain-aware legged locomotion using reinforcement learning and optimal control," *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 2908–2927, 2022.
- [33] G. Bellegarda and A. Ijspeert, "Cpg-rl: Learning central pattern generators for quadruped locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 547–12 554, 2022.
- [34] Y. Chen and Q. Nguyen, "Learning agile locomotion and adaptive behaviors via rl-augmented mpc," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 11 436–11 442.
- [35] G. Bledt, M. J. Powell, B. Katz, J. Di Carlo, P. M. Wensing, and S. Kim, "Mit cheetah 3: Design and control of a robust, dynamic quadruped robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2245–2252.
- [36] Z. Xiao, X. Zhang, X. Zhou, and Q. Zhang, "Pa-loco: Learning perturbation-adaptive locomotion for quadruped robots," *arXiv preprint arXiv:2407.04224*, 2024.
- [37] G. Bledt, P. M. Wensing, S. Ingersoll, and S. Kim, "Contact model fusion for event-based locomotion in unstructured terrains," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4399–4406.
- [38] Y. Zhang, B. Nie, and Y. Gao, "Robust locomotion policy with adaptive lipschitz constraint for legged robots," *IEEE Robotics and Automation Letters*, 2024.