2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Reinforcement Learning-based Robust Wall Climbing Locomotion Controller in Ferromagnetic Environment

Yong Um[1,2], Young-Ha Shin[2], Joon-Ha Kim[2], Soonpyo Kwon[1,2], and Hae-Won Park[1]

***Abstract*— We present a reinforcement learning framework for quadrupedal wall-climbing locomotion that explicitly addresses uncertainty in magnetic foot adhesion. A physics-based adhesion model of a quadrupedal magnetic climbing robot is incorporated into simulation to capture partial contact, air-gap sensitivity, and probabilistic attachment failures. To stabilize learning and enable reliable transfer, we design a three-phase curriculum: (1) acquire a crawl gait on flat ground without adhesion, (2) gradually rotate the gravity vector to vertical while activating the adhesion model, and (3) inject stochastic adhesion failures to encourage slip recovery. The learned policy achieves a high success rate, strong adhesion retention, and rapid recovery from detachment in simulation under degraded adhesion. Compared with a model predictive control (MPC) baseline that assumes perfect adhesion, our controller maintains locomotion when attachment is intermittently lost. Hardware experiments with the untethered robot further confirm robust vertical crawling on steel surfaces, maintaining stability despite transient misalignment and incomplete attachment. These results show that combining curriculum learning with realistic adhesion modeling provides a resilient sim-to-real framework for magnetic climbing robots in complex environments.**

I. Introduction

Climbing robots are a promising solution for inspection and maintenance of large-scale steel infrastructure, where manual operations are costly and hazardous [1]–[15].A variety of adhesion mechanisms have been explored, including suction [8], gecko-inspired dry adhesion [3], [9], microspines [2], [4]–[7], [11], electroadhesion [12], and magnetic attachment [1], [10], [13]–[15]. Among them, magnetic adhesion is particularly well-suited for steel structures, offering strong and repeatable attachment with minimal energy cost, while remaining robust to coatings, dust, and moderate surface irregularities. Combined with the versatility of legged locomotion, magnetic climbing legged robots can traverse complex geometries, step over obstacles, and maintain redundant points of contact for stability, making them attractive for deployment in industrial environments [13]–[15].

Control of such platforms has primarily relied on Model Predictive Control (MPC) [14], [15], which optimizes motion over a finite horizon under dynamics and kinematic constraints. By incorporating nominal adhesion forces, MPC has enabled stable crawling and trotting on vertical surfaces. However, this reliance on accurate models exposes limitations. First, MPC typically assumes perfect adhesion; in practice, paint, dust, or partial foot placement often reduce holding forces or cause detachment [14]. Since adhesion uncertainty is not modeled, such disturbances often lead to slippage or falls. Second, solving constrained optimizations at every cycle incurs a high computational cost, limiting responsiveness to sudden slip events. Third, mismatches between modeled and actual adhesion forces can destabilize the controller, reducing robustness in unstructured environments.

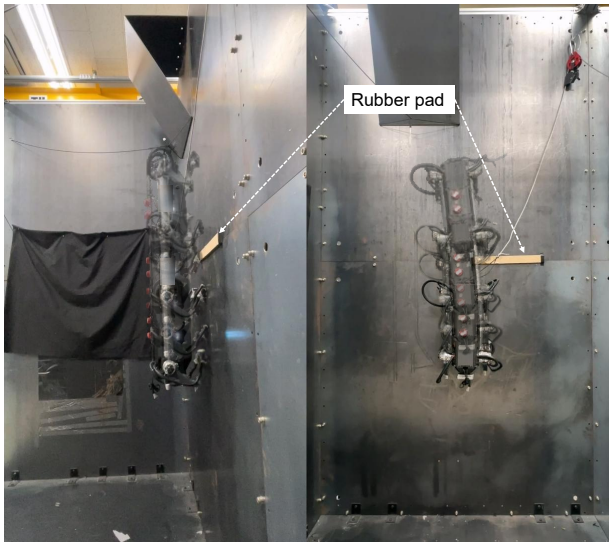

Fig. 1. Snapshots of performing robust vertical climbing. The learned controller maintains adhesion and recovers from slips even with non-ferromagnetic patches, enabling the robot to continue crawling without detaching from the wall.

Reinforcement learning (RL) offers a complementary paradigm by enabling robots to acquire robust control policies directly through interaction with their environment [16]–[24]. RL-trained controllers have demonstrated unprecedented adaptability over challenging terrestrial terrains [19], [20], [23], often exceeding model-based approaches in handling unmodeled perturbations [21]. Extending RL to vertical climbing is particularly promising, since an RL policy could, in principle, learn recovery from adhesion failures without explicit adhesion models. Yet, climbing introduces unique challenges: adhesion must be accurately represented in simulation to avoid a sim-to-real gap, and adhesion loss often leads to catastrophic falls. Moreover, naive RL training frequently produces policies that fail to generalize to vertical settings. Prior work has rarely addressed RL-based climbing, and almost none have explicitly considered adhesion uncertainty during training.

In this work, we propose a reinforcement learning framework for a quadrupedal magnetic climbing robot [14]. Our

[1]Korea Advanced Institute of Science and Technology, Yuseong gu, Daejeon 34141, Republic of Korea. haewonpark@kaist.ac.kr
[2]DIDEN Robotics, Seongdong gu, Seoul 04799, Republic of Korea.
*This research was supported by the National Research Foundation of Korea (NRF)(NRF-2022R1A2C2011927)

contributions are threefold: (1) a multi-phase curriculum that gradually rotates the gravity vector from horizontal to vertical for stable ground-to-wall transitions; (2) a physics-based adhesion model of EPM feet that captures both partial and failed contacts; and (3) the introduction of stochastic adhesion failures during training to encourage recovery under uncertainty. Simulation and hardware experiments show our learned controller achieves robust vertical climbing, maintains adhesion despite failures, and recovers from slip events, outperforming an MPC baseline (Fig. 1).

II. METHODS

We propose a reinforcement learning–based control framework for robust vertical crawling on steel surfaces with uncertain magnetic adhesion. A realistic magnetic foot adhesion model is incorporated into the simulation to reflect the behavior of electropermanent-magnetic feet. The training process follows a curriculum that gradually rotates the gravity vector from horizontal to vertical, enabling the policy to adapt to climbing conditions. To improve robustness, we introduce stochastic foot adhesion failures during training. An overview of the overall training and deployment framework is illustrated in Fig. 2, and the learned policy is deployed on the quadrupedal magnetic climbing to demonstrate stable vertical crawling in real-world experiments.

A. Hardware

We utilize an untethered quadrupedal climbing robot designed for dynamic locomotion on vertical steel surfaces. The robot weighs 8 kg and is equipped with four magnetic feet, each weighing approximately 0.2 kg, integrated with electropermanent magnets (EPMs) that provide switchable magnetic adhesion [14]. The EPMs used in this study generate a maximum normal holding force of approximately 697 N and support rapid magnetic switching within 5 milliseconds, allowing timely attachment and detachment during locomotion. While prior designs incorporated magnetorheological elastomer (MRE) footpads to enhance shear adhesion, we employ a simplified foot design without MRE to better isolate and model adhesion uncertainty during learning.

B. Magnetic Foot Adhesion Model

To realistically capture the behavior of electropermanent-magnetic (EPM) feet, we implement a magnetic adhesion model in simulation. EPMs generate holding force by magnetizing an AlNiCo core through a coil pulse, which requires forming a closed magnetic circuit with the steel surface. Stable adhesion is obtained only when the EPM is in proper contact with the surface; if the magnet is activated without contact, or if an air gap exists between the EPM and the plate, the resulting force is significantly reduced. Fig. 3 illustrates how the adhesion force drops as the gap increases, highlighting the sensitivity of EPMs to partial contact conditions.

In simulation, adhesion is considered only if the following conditions are satisfied in sequence:

(1) Contact recognition. The state estimator outputs a contact confidence $\tilde{c}_{\text{foot}} \in [0, 1]$, which is interpreted as the probability of foot–wall contact. A contact is recognized when $\tilde{c}_{\text{foot}} \geq 0.5$, corresponding to the binary contact indicator $c_{\text{foot}} \in \{0, 1\}$. This c_{foot} is a privileged value obtained only in simulation, where $c_{\text{foot}} = 1$ denotes that the foot is in contact with the wall and $c_{\text{foot}} = 0$ otherwise:

$$\tilde{c}_{\text{foot}} \geq 0.5. \tag{1}$$

(2) Magnet activation. The policy outputs an activation signal a_{magnet} trained to regress the binary contact indicator $c_{foot} \in \{0, 1\}$. Since c_{foot} takes only 0 or 1, we use 0.5 as a threshold:

$$a_{\text{magnet}} \geq 0.5, \tag{2}$$

meaning that the magnet is regarded as ON when the output exceeds this value.

(3) Stochastic adhesion. Even when both conditions hold, adhesion occurs only probabilistically to model real-world uncertainties. A random variable $X \sim \mathcal{U}(0, 1)$ is sampled only when the foot is in swing (not contact with the wall), and adhesion succeeds only if

$$X \leq \text{Prob}_{\text{attach}}. \tag{3}$$

This stochastic component represents environment-dependent failures such as non-ferrous surfaces, uneven or painted walls, or partial contacts. The details of $\text{Prob}_{\text{attach}}$ scheduling are given in Section II-C.

(4) Geometric alignment. Finally, proper adhesion requires full geometric contact between the magnet and the wall surface:

$$S_{\text{EPM}} = S_{\text{wall}}, \tag{4}$$

where S_{EPM} and S_{wall} represents the contact surface of EPM and wall, respectively.To generate a stable adhesion force, the EPM surface must be completely aligned with the wall to form a closed magnetic circuit. In cases where even a small gap remains, the magnetic circuit is incomplete and the effective adhesion force is drastically reduced. By enforcing these sequential conditions, the policy is exposed to realistic adhesion failures and learns robust recovery strategies that transfer effectively to real hardware. This modeling choice reduces the sim-to-real gap by capturing imperfect contact and slippage, enabling reinforcement learning to produce policies robust to unreliable adhesion. Further details and empirical analyses of this strategy are presented in Section III. In hardware implementation, the electropermanent magnet (EPM) is controlled via current pulses. When the contact recognition (Eq. 1) and magnet activation (Eq. 2) conditions are satisfied, a current pulse is applied to switch the EPM *on* and generate adhesion force. Conversely, when these conditions become violated, a reverse current pulse is applied to switch the EPM *off*, thereby releasing adhesion.

C. Multi-Phase Learning Strategy for Vertical Locomotion under Adhesion Uncertainty

To enable robust and adaptive climbing locomotion, we adopt a multi-phase learning strategy that progressively advances from basic gait learning to climbing-specific challenges. As shown in Fig. 2, the training is organized into

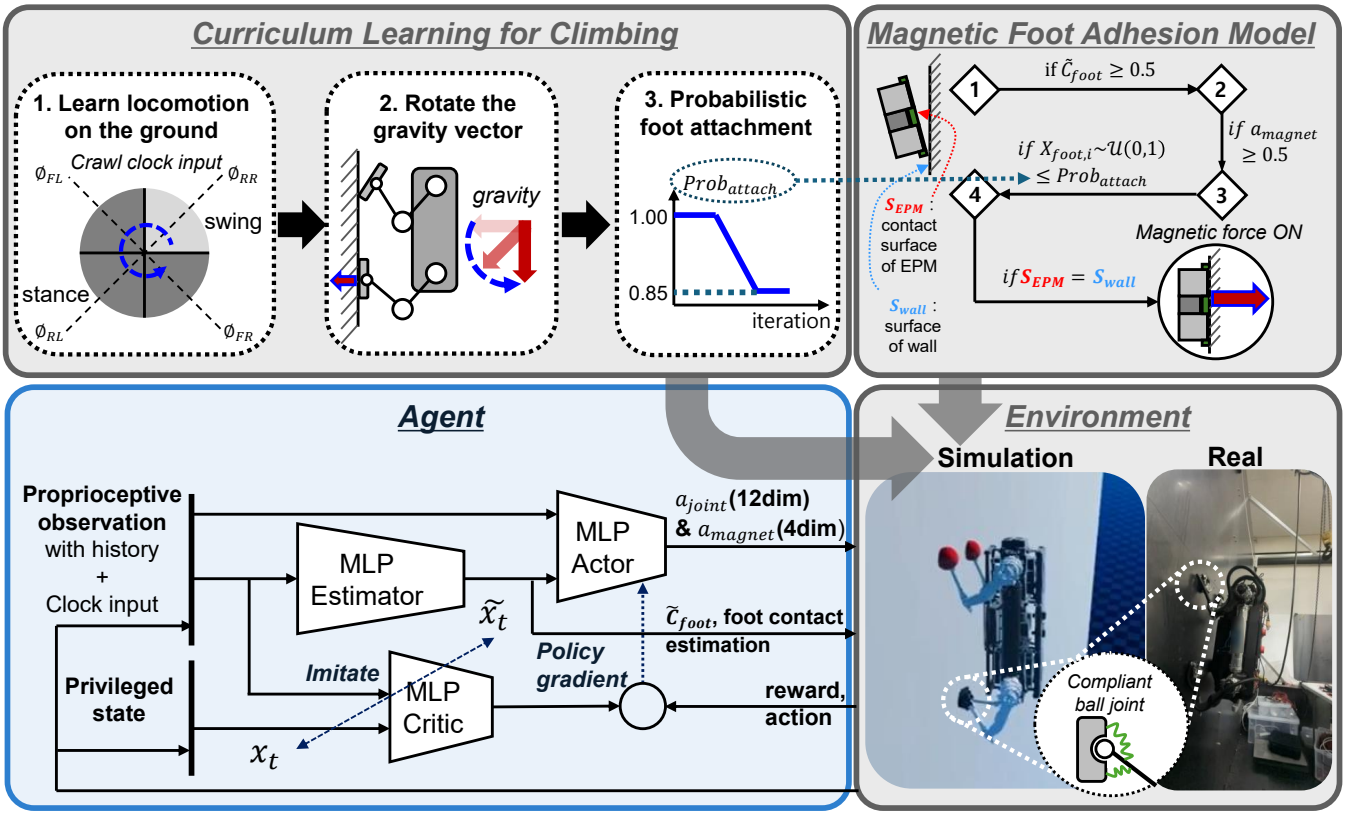

Fig. 2. Overview of the proposed learning framework for vertical locomotion under adhesion uncertainty. The process consists of four main components: (1) ground locomotion pre-training, (2) curriculum-based gravity rotation to adapt to climbing orientation, (3) probabilistic foot adhesion modeling to simulate imperfect magnetic contact, and (4) integration of simulation-to-real transfer with a compliant magnetic foot model. The policy receives proprioceptive observations with history and a crawl clock input, while an estimator predicts foot contacts. The actor–critic architecture is trained via reinforcement learning and imitation signals to produce joint torques and magnetic adhesion actions.

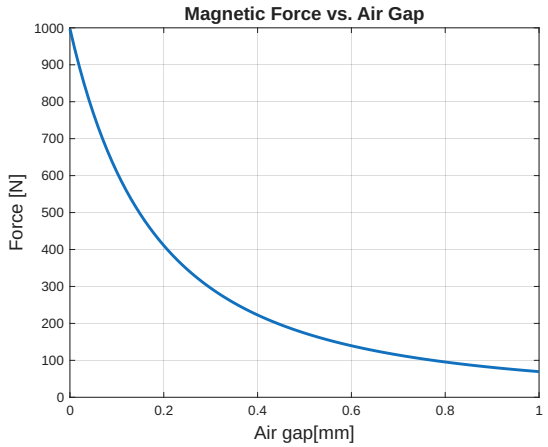

Fig. 3. Magnetic force of the magnetic feet with respect to air gap. Even with a 1 mm air gap, the adhesion force drops to about 7% of the maximum value.

three sequential phases: (1) gait acquisition on flat terrain, where the robot first learns a stable crawl without adhesion to avoid frequent early failures, (2) adaptation to altered gravitational orientations during wall climbing, and (3) robustness learning under probabilistic foot adhesion failures that mimic imperfect magnetic contact in real hardware.

a) Phase 1: Gait Acquisition on Flat Ground: The first phase aims to establish a stable crawl gait on flat ground before introducing gravitational transitions or adhesion uncertainty. We begin training on flat ground because directly attempting to learn wall climbing from scratch leads to frequent falls: Untrained initial policies, generated from random initialization, are unable to reliably lift and place their feet while maintaining body support. As a result, they frequently fall off the wall during early training, which prevents the collection of sufficient high-quality samples for stable learning. On flat ground, by contrast, the robot can safely acquire a baseline crawling gait without magnetic adhesion. During this stage, the policy is allowed to output magnetic foot actions, but the adhesion model is deliberately disabled so that no actual adhesive force is applied. This setup ensures that the controller first concentrates on learning smooth and coordinated crawling, without relying on adhesion. At the same time, an auxiliary reward is introduced to guide when magnets should ideally be switched on or off: it penalizes the policy if magnet activation occurs during swing or if magnets are off during stance. Through this mechanism, the policy learns a natural timing of adhesion—turning magnets ON during stance and OFF during swing—even before real adhesion forces are simulated. This implicit learning of magnet timing provides a foundation for stable climbing behaviors, which will be leveraged in later phases when adhesion is physically modeled. Training follows the concurrent learning approach in [18], where an estimator network is trained concurrently with the actor to predict

privileged base velocity, foot height, and contact probability by minimizing MSE loss against simulator ground truth. This is combined with an 8-dimensional clock input to encode each leg's gait phase and promote periodic behaviors such as crawling. Detailed mechanisms of how the clock input promotes crawling behavior, along with the specific gait reward formulations, are provided in Section II-D.

b) Phase 2: Adaptation to Gravitational Rotation: After learning locomotion on flat ground, the robot is gradually exposed to inclined and vertical surfaces by rotating the direction of gravity in the simulation. This curriculum-based gravitational transition enables the policy to adapt its balance and posture control to non-horizontal terrains. During this phase, the magnetic adhesion model is activated, and foot contact becomes critical for maintaining stability on climbing surfaces. The timing strategy for magnet activation, which was shaped in Phase 1 through auxiliary rewards, is now coupled with actual adhesion forces so that the robot can reliably attach its feet during stance and release them during swing.

The gravity rotation is scheduled per training iteration t as:

$$\theta(t) = \min\Big\{ \tfrac{\pi}{2},\ \max\Big\{0,\ \tfrac{\pi}{2} \cdot \tfrac{t-1200}{20000} \Big\} \Big\}, \tag{5}$$

where $\theta(t)$ is the tilt angle from the ground ($0°$) to the wall ($90°$). Let $g_0 \in \mathbb{R}^3$ denote the nominal gravity vector in the ground frame. To gradually shift the environment from flat ground to a vertical wall, g_0 is rotated by $\theta(t)$ around an axis parallel to the ground surface. In this study, gravity is defined along the $-z$ direction and rotated about the $+y$ axis, yielding

$$g(t) = R_y\big(\theta(t)\big)\, g_0, \qquad R_y(\theta) \in \mathrm{SO}(3). \tag{6}$$

Thus, for $t \leq 1200$, the robot trains on flat ground with magnetic forces already applied; from $t = 1201$ to $t = 21200$, the gravity vector is linearly tilted until it reaches $90°$; and for $t > 21200$, the gravity vector remains at $90°$ for the remainder of training.

c) Phase 3: Robustness to Adhesion Uncertainty: In the final phase, we introduce stochastic disturbances to simulate real-world adhesion uncertainty. Even when a magnetic foot is commanded ON, the attachment may fail with a certain probability, reflecting realistic failure modes such as partial foot placement, non-ferromagnetic surface coatings, or surface contamination. By exposing the policy to these perturbations during training, the robot learns to maintain balance and recover from unexpected slippage or detachment.

The adhesion success probability $\mathrm{Prob}_{\text{attach}}(t)$ is kept at 1.0 until iteration $t = 21200$ and then decreased linearly to 0.85 by iteration $t = 35000$:

$$\mathrm{Prob}_{\text{attach}}(t) = 1.0 - 0.15 \cdot \frac{\min\big(\max(t - 21200,\ 0),\ 13800\big)}{13800}. \tag{7}$$

During training, adhesion is not triggered solely by stance phase, but only when all activation conditions are satisfied: the contact estimator must indicate a valid foot contact ($\tilde{c}_{\text{foot}} \geq 0.5$, Eq. 1), the commanded magnet action must exceed the activation threshold ($a_{\text{magnet}} \geq 0.5$, Eq. 2), and the foot must be in contact with the wall surface ($S_{\text{EPM}} = S_{\text{wall}}$, Eq. 4). Once these conditions hold, the adhesion model samples a random number uniformly from $[0, 1]$ and compares it with the scheduled success probability $\mathrm{Prob}_{\text{attach}}(t)$ (Eq. 3). If the sample is smaller, the foot attaches successfully and produces magnetic force; otherwise, the attempt fails and no holding force is applied. This mechanism injects probabilistic failures into training, encouraging the policy to learn recovery strategies that maintain stability even under imperfect adhesion.

D. Learning Framework

We adopt a reinforcement learning framework based on Proximal Policy Optimization (PPO) to train a robust climbing locomotion policy in the RaiSim simulator [25]. The framework consists of three neural networks: an actor (policy) network, a critic (value function) network, and a state estimator network that predicts privileged states such as base velocity, foot height, and foot contact probabilities [18]. All networks are implemented as multilayer perceptrons (MLPs). The actor and critic share the same architecture with three hidden layers of sizes $[256, 128, 64]$, while the state estimator is a smaller MLP with two hidden layers of sizes $[256, 128]$. The overall structure is illustrated in Fig. 2.

a) Observation and Action Space: The observation vector o_t aggregates proprioceptive states, historical information, and gait phase encoding. Specifically, it contains joint positions q_t and velocities $\dot{q}_t$; previous joint position targets for the last two time steps; base orientation ϕ_t and angular velocity ω_t; Cartesian foot positions relative to the base; estimated base linear velocity, foot height, and foot contact probabilities from the state estimator; and an 8-dimensional clock input o_{clock} for gait phase encoding. The clock input is defined as $o_{\text{clock}} = [\sin\phi_i, \cos\phi_i]_{i=1}^{4} \in \mathbb{R}^8$, where $i \in \{1, 2, 3, 4\}$ corresponds to the right-rear (RR), right-front (FR), left-rear (RL), and left-front (FL) legs. The phase variables are given by $\phi_i = \frac{2\pi}{T}t + \frac{\pi}{2}i$, where t is the time index, $T = 1.2\,\text{s}$ denotes the gait cycle period, and $\frac{\pi}{2}i$ introduces a quarter-phase shift between consecutive legs. All observation signals are further processed through a first-order low-pass filter before being fed into both the actor and the estimator networks. The filter is implemented as $obs_{\text{filter}} = (1 - \alpha)obs_{\text{old}} + \alpha obs_{\text{new}}$, with the smoothing coefficient set to $\alpha = 0.35$. This filtering suppresses high-frequency noise while retaining essential state information for stable policy learning.

b) Reward Design: The reward function is designed to produce stable, efficient, and robust climbing behaviors while discouraging unsafe or inefficient actions. It is composed of multiple components that address different aspects of locomotion, such as velocity tracking, posture stability, foot placement, and action smoothness. Table I lists all reward terms and their mathematical expressions. To facilitate curriculum learning in the climbing task, several rewards are dynamically scaled over training iterations. We define the scheduling factor $\kappa = 0.99975^{\max(iter-1200,0)}$. The velocity

tracking rewards (R_{lv}, R_{av}) are scaled by $(1.5 - 0.5\kappa)$, which gradually increases their weight to emphasize accurate command following when the robot transitions to vertical climbing. Conversely, the foot slip (R_{fs}) and joint torque (R_τ) penalties are scaled by $(0.5 + 0.5\kappa)$, which gradually decreases their weight, allowing more torque usage and occasional foot slip during wall climbing without overly penalizing the agent. Furthermore, the action smoothness terms (R_{as1}, R_{as2}) are disabled during the early stage of Phase 1 ($t < 1000$) to prevent over-constraining exploration. Throughout Phases 2 and 3, action smoothness terms are evaluated as specified in Table I.

Total reward:

$$R_{\text{tot}} = (R_{lv} + R_{av} + R_g + R_{fh} + R_{sc}) \cdot \exp\Big\{ -0.2\big(R_{fs} + R_{fc} + R_o + R_\tau + R_{jp} + R_{js} + R_{ja} + R_{as1} + R_{as2} + R_{bm} + R_{am}\big)\Big\}. \quad (8)$$

This formulation promotes forward crawling with correct gait timing while suppressing undesirable behaviors such as excessive slipping, abrupt motions, or unstable body orientation. The exponential term acts as a multiplicative penalty, amplifying the effect of violations in stability, contact, and actuation smoothness.

TABLE I
REWARD COMPONENTS USED IN THE CLIMBING LOCOMOTION TASK.

Reward	**Expression / Description**
Linear velocity (R_{lv})	$(1.5 - 0.5\kappa)\ 3.0 \exp\big(-5.0 \lVert v_{xy}^{\text{desired}} - v_{xy} \rVert^2\big)$
Angular velocity (R_{av})	$(1.5 - 0.5\kappa)\ 3.0 \exp\big(-5.0 (\omega_z^{\text{desired}} - \omega_z)^2\big)$
Standing command (R_{sc})	$0.5 \sum_{i=1}^{4} r_i$
Gait (R_g)	$0.5 \sum_{i=1}^{4} g_i$
Foot height (R_{fh})	$0.5 \exp\big(-\sum_i f_i\,(p_{z,i}^{\text{des}} - p_{z,i})^2\big)$
Foot slip (R_{fs})	$(0.5 + 0.5\kappa)\ 0.5 \sum_i c_i \lVert v_{xy,i} \rVert^2$
Foot clearance (R_{fc})	$140 \sum_i (1 - c_i)\,(p_{z,i}^{\text{des}} - p_{z,i})^2 \lVert v_{z,i} \rVert^{0.5}$
Orientation (R_o)	$3 \cdot \text{angle}\big(\phi_{\text{body},z},\ v_{\text{world},z}\big)$
Joint torque (R_τ)	$(0.5 + 0.5\kappa)\ 0.003 \lVert \tau_t \rVert^2$
Joint position (R_{jp})	$\alpha_{jp} \lVert q_t - q^{\text{nominal}} \rVert^2$
$0.003 \lVert \dot{q} \rVert^2$	
Joint acceleration (R_{ja})	$0.003 \lVert \ddot{q}_t \rVert^2$
Action smoothness 1 (R_{as1})	$\begin{cases} 0, & \text{if Phase 1 and } t < 1000, \\ 2.5 \lVert a_t - a_{t-1} \rVert^2, & \text{otherwise,} \end{cases}$
Action smoothness 2 (R_{as2})	$\begin{cases} 0, & \text{if Phase 1 and } t < 1000, \\ 1.2 \lVert a_t - 2a_{t-1} + a_{t-2} \rVert^2, & \text{otherwise,} \end{cases}$
Base motion (R_{bm})	$3.0 \exp\big(-0.5 \lVert \omega_{x,y} \rVert^2 + 0.2 \lvert v_z \rvert\big)$
Action magnet (R_{am})	$0.15 \sum_i \big(c_i - a_{magnet,i}\big)^2$

Notation.

$$g_i = \begin{cases} 1, & \phi_i \in (0, \frac{\pi}{2}),\ \text{foot } i \text{ not in contact}, \\ -1, & \phi_i \in (0, \frac{\pi}{2}),\ \text{foot } i \text{ in contact}, \\ 1, & \phi_i \notin (0, \frac{\pi}{2}),\ \text{foot } i \text{ in contact}, \\ -1, & \text{otherwise}, \end{cases} \quad (9)$$

$$f_i = \begin{cases} 1, & \phi_i \in (0, \frac{\pi}{2}), \\ 0, & \text{otherwise}, \end{cases} \quad (10)$$

$$p_{z,i}^{\text{des}} = \begin{cases} 0.08, & \phi_i \in (0, \frac{\pi}{2}), \\ 0, & \text{otherwise}, \end{cases} \quad (11)$$

$$c_i = \begin{cases} 1, & \text{if foot } i \text{ in contact}, \\ 0, & \text{otherwise}, \end{cases} \quad (12)$$

$$r_i = \begin{cases} 1, & \text{if foot } i \text{ in contact when } v^{\text{desired}} = 0, \\ -1, & \text{otherwise}, \end{cases} \quad (13)$$

$$\alpha_{jp} = \begin{cases} 3, & \text{if } v^{\text{desired}} = 0 \text{ (standing command)}, \\ 0.75, & \text{otherwise}. \end{cases} \quad (14)$$

$p_{z,i}$: vertical position of foot i; $v_{xy,i}$: horizontal velocity of foot i; $v_{z,i}$: vertical velocity of foot i; ϕ_i: gait phase of leg i; τ_t: joint torques; q_t: joint positions; $\ddot{q}_t$: joint accelerations; $a_{magnet,i}$: magnet action for leg i.

c) Domain Randomization: To ensure that the policy trained in simulation can be reliably deployed on the real robot, we incorporate domain randomization and hardware-aware modeling throughout the training process. This approach mitigates the sim-to-real gap by exposing the policy to a range of environmental and actuation variations during training. At the beginning of each episode, the following parameters are uniformly randomized within predefined ranges:

- **Joint PD gains:** $[0.4, 0.6]$ to represent joint P gain, and $[0.12, 0.18]$ to represent joint D gain.
- **Ground friction coefficient:** $[0.3, 0.5]$ to represent the friction between various steel surfaces and EPM feet.
- **Observation noise:** Uniform noise applied to observation channels: orientation (± 0.05 rad plus an offset bias sampled within ± 0.05 rad), joint angles (± 0.1 rad), body angular velocity (± 0.1 rad/s), joint velocities (± 0.5 rad/s), joint position histories (± 0.1 rad), joint velocity histories (± 0.5 rad/s), and foot positions (± 0.015 m).
- **Action delay:** Random delay in the range $[0, 0.008]$ s to emulate hardware switching latency.

d) Compliant 3-DOF Joint Ankle Modeling: To realistically capture the mechanical compliance of the ankle, the calf link (shank) and the magnetic foot are connected via a ball joint with three rotational degrees of freedom. The nominal ankle orientation is fixed as the position target, and elastic compliance is modeled by applying position and velocity gains that mimic the effect of the elastic band connecting the foot to the shank. To improve robustness, these gains are randomized during training within a predefined range. The nominal values of the target angle and the controller gains

are summarized below:

- Nominal ankle orientation: roll–pitch–yaw $[0, 0.523599, 0]$ rad $([0°, 30.0°, 0°])$, which follows the definition of the nominal configuration reported in [14].
- Position gain (K_p): 0.05 ± 0.01
- Velocity gain (K_d): 0.001 ± 0.0005

III. RESULTS

A. Learning Progress Across Phases

Fig. 4 shows the learning progress of the proposed curriculum. During Phase 1, the policy rapidly improves reward as it acquires a stable crawl gait on flat ground. After $t > 1000$, the action smoothness regularization terms (R_{as1}, R_{as2}) are activated, which temporarily lowers the total reward due to the additional penalty. However, the policy quickly adapts, and the reward continues to increase as smoother and more stable motions emerge.

During Phase 2, the gradual rotation of the gravity vector toward the wall leads to a reduction in the average episode success rate. In Phase 3, the introduction of stochastic adhesion increases the likelihood of detachment from the wall, which in turn induces a gradual decline in overall performance. However, the performance eventually saturates, and despite the adhesion probability being reduced to 85%, the policy consistently shows an average episode success rate of approximately 90%. This shows that the curriculum design stabilizes training and allows the policy to be generalized to increasingly challenging climbing conditions.

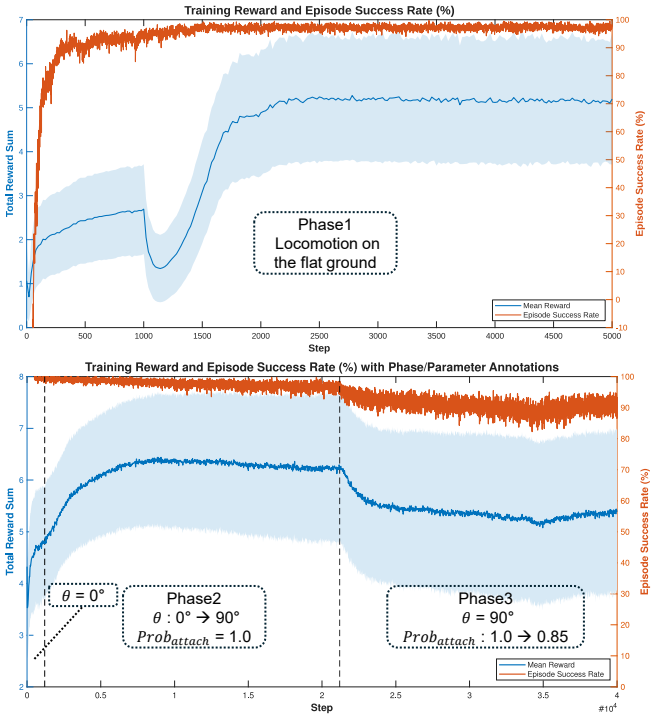

Fig. 4. Training curves of climbing policy optimization (Top: Phase 1, Bottom: Phases 2–3). The vertical dashed lines indicate changes of the multi-phase learning curriculum. Phase 1 trains locomotion on flat ground. In Phase 2, the wall inclination θ increases from $0°$ to $90°$ with perfect adhesion ($\text{Prob}_{\text{attach}} = 1.0$). Phase 3 fixes $\theta = 90°$ while gradually reducing the adhesion probability from 1.0 to 0.85, exposing the policy to adhesion failures.

B. Ablation Study

a) Training and Evaluation Settings: Each ablation policy was trained under identical reinforcement learning settings, with the only difference being the removed component. All policies were trained using PPO in RaiSim with proprioceptive observations, clock inputs, and domain randomization. The reward structure and training duration were kept consistent across all runs. The following summarizes the conditions for each ablation:

- **Full (ours):** Multi-phase curriculum learning was applied, gradually rotating the gravity vector from $0°$ to $90°$, with realistic magnetic foot modeling enabled and stochastic adhesion failures introduced according to Eq. 3.
- **w/o Curriculum:** The gravity vector was fixed at $90°$ from the beginning of training, and the reward formulation was identical to that used in Phases 2 and 3. Realistic adhesion modeling and stochastic adhesion failures were still applied.
- **w/o Probabilistic Adhesion:** The probability of foot adhesion was fixed to $\text{Prob}_{\text{attach}} = 1.0$ during training, disabling stochastic failures. Gravity curriculum and realistic modeling remained enabled.
- **w/o Modeling:** Realistic adhesion modeling, including alignment and air-gap effects, was removed. In this setting, adhesion was considered ideal whenever the magnet command(a_{magnet}) exceeded 0.5, regardless of the contact geometry. Thus, even partial contact between the EPM and the wall surface was treated as full adhesion. For fair comparison, $\text{Prob}_{\text{attach}}$ was fixed to 1.0 to isolate the effect of modeling.

b) Evaluation Metrics: We define the following metrics to evaluate climbing performance, using a fixed evaluation horizon of $T_{\text{hor}} = 10\,\text{s}$ and $N = 100$ episodes.

- **Velocity Tracking RMSE:** quantifies how accurately the robot follows the commanded velocity profile. Lower values indicate better tracking performance. The velocity command is defined as $v^{\text{desired}}(t) = [v_x^{\text{desired}}(t), v_y^{\text{desired}}(t), \omega_z^{\text{desired}}(t)] \in [-0.5, 0.5] \times [-0.3, 0.3] \times [-0.5, 0.5]$, where $v(t) = [v_x(t), v_y(t), \omega_z(t)]$ is the measured velocity, and the commands $v^{\text{desired}}(t)$ are randomly sampled within the specified ranges.
- **Early Termination Rate:** the percentage of episodes that ended before the evaluation horizon T_{hor} due to failure conditions. A lower rate indicates more reliable climbing. Early termination was triggered when any of the following occurred: (i) the robot detached from the wall and fell to the ground, or (ii) all feet remained attached for more than $5.0\,\text{s}$ indicating that the robot did not move for at least half of the episode.
- **Average Walking Time:** the mean duration that episodes lasted (up to T_{hor}). Higher values indicate the policy can sustain crawling for longer without failure.

- **Retention:** the fraction of stance phase time during which magnetic adhesion force is actively applied. Retention quantifies how reliably the robot maintains magnetic contact during stance. A higher value indicates that the feet remain attached whenever required or promptly re-attach after a slip, thereby minimizing periods without support.
- **Recovery Rate:** the percentage of adhesion failures (caused by $prob_{attach} < 1$) that were successfully recovered within ΔT without episode termination. A higher rate indicates that the policy can better withstand adhesion disturbances.

c) Ablation Analysis on Training Components: Table II summarizes the ablation study results, highlighting the contribution of curriculum learning, probabilistic adhesion, and adhesion modeling. Without curriculum learning, the robot tends to remain attached to the wall without detaching its feet, resulting in little actual movement and consequently a relatively low velocity tracking error compared to the case without adhesion modeling. However, this behavior results in extremely high early termination rates, since the robot fails to move forward and effectively remains stationary. Consequently, retention is artificially high because the feet stay attached throughout the episodes. In contrast, removing probabilistic adhesion shows performance comparable to the full model when $p_{\text{attach}} = 1.0$, but under $p_{\text{attach}} = 0.85$ the retention and recovery rates significantly degrade. This demonstrates that introducing stochastic adhesion improves robustness against attachment failures. Finally, training without adhesion modeling yields the worst performance across metrics, as the mismatch with the real magnetic mechanism leads to frequent detachment from the wall, large velocity tracking errors, and unstable climbing behavior.

C. Hardware Validation

We further validate the proposed framework on the untethered quadrupedal magnetic climbing robot. In vertical wall experiments, the RL policy successfully traverses over non-ferromagnetic surfaces while recovering from induced adhesion failures. As a baseline, we compare against the Model Predictive Control (MPC) controller proposed in prior work [14], which assumes perfect adhesion. Snapshots in Fig. 5 show a representative hardware trial. When a front foot fails to attach, the RL policy shifts weight, reattempts adhesion, and resumes crawling. In contrast, the MPC baseline is unable to recover once adhesion fails, leading to immediate instability.

In addition to visual demonstrations, we also log internal signals during hardware trials. Fig. 6 illustrates the time traces of (i) the policy output for magnet activation, (ii) the estimated foot contact confidence, and (iii) the measured on/off state of the EPMs. While contact estimation occasionally fluctuated around the threshold due to mismatches with simulation, we adopted a simple rule in hardware: the magnet remained ON unless the contact probability stayed below 0.5 for more than 0.02 s. These traces confirm that the policy correctly synchronizes magnet actions with actual

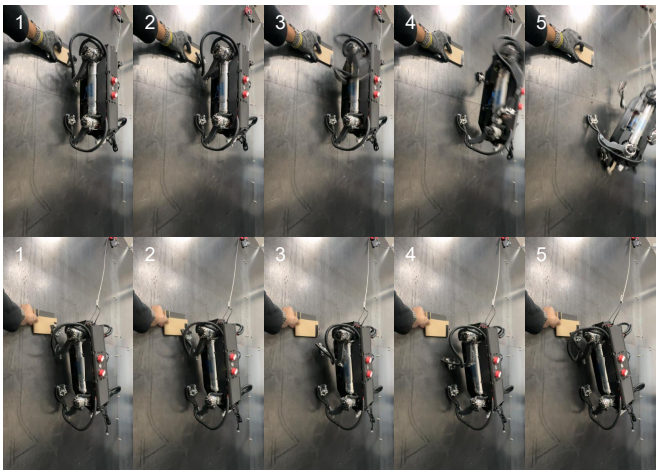

Fig. 5. Snapshots from a hardware trial. (Top) MPC fails once adhesion is lost. (Bottom) Our RL policy recovers from a failed adhesion and resumes stable crawling.

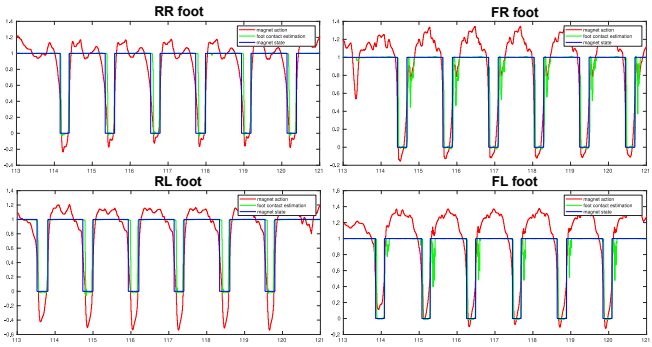

Fig. 6. Logged signals from a hardware climbing trial. For each leg (RR, FR, RL, FL), the plots show magnet activation command, foot contact estimation, and the measured EPM state, illustrating their synchronization during climbing.

contact events, and that the EPMs are reliably switched in accordance with the commanded activation.

To further ensure reproducibility, multiple walking tests were conducted to verify that the learned controller can consistently achieve stable climbing locomotion, and the results are provided in the supplementary video.

IV. CONCLUSION AND FUTURE WORK

We proposed a reinforcement learning framework for robust wall-climbing locomotion with the quadrupedal magnetic robot. The framework combines (i) a three-phase curriculum that transitions from ground crawling to vertical climbing, (ii) a realistic adhesion model of electropermanent-magnetic (EPM) feet that captures partial and failed contacts, and (iii) stochastic adhesion failures during training to promote recovery strategies. Through simulation and hardware experiments, the learned controller demonstrated stable vertical crawling, high adhesion retention, and effective slip recovery, while a model predictive control (MPC) baseline failed immediately under adhesion loss. These results highlight the benefits of learning-based control trained with explicit adhesion uncertainty.

Looking ahead, we plan to extend this framework to more diverse ferromagnetic environments, including curved

TABLE II

ABLATION OF CURRICULUM AND PROBABILISTIC ADHESION IN SIMULATION. METRICS ARE COMPUTED OVER $T_{\mathrm{hor}} = 10\,\mathrm{s}$ AND $N = 100$ EPISODES

Results with $p_{\mathrm{attach}} = 1.0$

Condition	**Vel. RMSE (mean ± std).** (m/s)	**Early Term.** (%)	**Avg. Time** (s)	**Retention** (%)
Full (ours)	0.1021 ± 0.0890	0.00	10.0000 ± 0.0000	79.67 ± 8.00
w/o Curriculum	0.2055 ± 0.1803	88.00	5.8724 ± 1.7620	98.50 ± 9.26
w/o Probabilistic	0.0844 ± 0.0728	3.00	9.7639 ± 1.3505	76.33 ± 14.70
w/o Modeling	5.1106 ± 26.2067	48.00	6.6430 ± 3.7370	44.36 ± 35.53

Results with $p_{\mathrm{attach}} = 0.85$

Condition	**Vel. RMSE** (m/s)	**Early Term.** (%)	**Avg. Time** (s)	**Retention** (%)	**Recovery Rate (%)** ΔT=1.2s	ΔT=2.4s	ΔT=3.6s
Full (ours)	0.5567 ± 7.3437	4.00	9.7300 ± 1.3699	72.93 ± 18.46	100.00 ± 0.00	98.79 ± 10.08	97.87 ± 13.75
w/o Probabilistic	6.4677 ± 29.1904	66.00	6.5077 ± 3.2063	42.51 ± 32.31	96.91 ± 16.99	58.74 ± 47.79	47.93 ± 48.36

surfaces, gaps, and irregular steel structures. We also aim to realize richer climbing skills, such as transitions across orientations (floor–wall–ceiling), enabling the robot to traverse a broader range of industrial and exploratory settings.

REFERENCES

[1] S. Hirose, A. Nagakubo, and R. Toyama, "Machine that can walk and climb on floors, walls and ceilings," in *Fifth International Conference on Advanced Robotics Robots in Unstructured Environments*. IEEE, 1991, pp. 753–758.

[2] S. Kim, A. T. Asbeck, M. R. Cutkosky, and W. R. Provancher, "Spinybotii: climbing hard walls with compliant microspines," in *International Conference on Advanced Robotics*. IEEE, 2005, pp. 601–606.

[3] K. Autumn, M. Buehler, M. Cutkosky, R. Fearing, R. J. Full, D. Goldman, R. Groff, W. Provancher, A. A. Rizzi, U. Saranli, *et al.*, "Robotics in scansorial environments," in *Unmanned ground vehicle technology VII*, vol. 5804. International Society for Optics and Photonics, 2005, pp. 291–302.

[4] A. T. Asbeck, S. Kim, M. R. Cutkosky, W. R. Provancher, and M. Lanzetta, "Scaling hard vertical surfaces with compliant microspine arrays," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1165–1179, 2006.

[5] A. Saunders, D. I. Goldman, R. J. Full, and M. Buehler, "The rise climbing robot: body and leg design," in *Unmanned Systems Technology VIII*, vol. 6230. SPIE, 2006, pp. 401–413.

[6] T. Bretl, "Motion planning of multi-limbed robots subject to equilibrium constraints: The free-climbing robot problem," *The International Journal of Robotics Research*, vol. 25, no. 4, pp. 317–342, 2006.

[7] B. Kennedy, A. Okon, H. Aghazarian, M. Badescu, X. Bao, Y. Bar-Cohen, Z. Chang, B. E. Dabiri, M. Garrett, L. Magnone, *et al.*, "Lemur IIb: a robotic system for steep terrain access," *Industrial Robot: An International Journal*, 2006.

[8] W. Brockmann, "Concept for energy-autarkic, autonomous climbing robots," in *Climbing and Walking Robots*. Springer, 2006, pp. 107–114.

[9] S. Kim, M. Spenko, S. Trujillo, B. Heyneman, D. Santos, and M. R. Cutkosky, "Smooth vertical surface climbing with directional adhesion," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 65–74, 2008.

[10] P. Ward and D. Liu, "Design of a high capacity electro permanent magnetic adhesion for climbing robots," in *IEEE International Conference on Robotics and Biomimetics*. IEEE, 2012, pp. 217–222.

[11] A. Parness, N. Abcouwer, C. Fuller, N. Wiltsie, J. Nash, and B. Kennedy, "Lemur 3: A limbed climbing robot for extreme terrain mobility in space," in *IEEE International conference on Robotics and Automation*. IEEE, 2017, pp. 5467–5473.

[12] S. D. de Rivaz, B. Goldberg, N. Doshi, K. Jayaram, J. Zhou, and R. J. Wood, "Inverted and vertical climbing of a quadrupedal microrobot using electroadhesion," *Science Robotics*, vol. 3, no. 25, p. eaau3038, 2018.

[13] T. Bandyopadhyay, R. Steindl, F. Talbot, N. Kottege, R. Dungavell, B. Wood, J. Barker, K. Hoehn, and A. Elfes, "Magneto: A versatile multi-limbed inspection robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2018, pp. 2253–2260.

[14] S. Hong, Y. Um, J. Park, and H.-W. Park, "Agile and versatile climbing on ferromagnetic surfaces with a quadrupedal robot," *Science Robotics*, vol. 7, no. 73, p. eadd1017, 2022.

[15] S. Leuthard, T. Eugster, N. Faesch, R. Feingold, C. Flynn, M. Fritsche, N. Hürlimann, E. Morbach, F. Tischhauser, M. Müller, *et al.*, "Magnecko: Design and control of a quadrupedal magnetic climbing robot," in *Climbing and Walking Robots Conference*. Springer, 2024, pp. 55–67.

[16] 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.

[17] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[18] G. Ji, J. Mun, H. Kim, and J. Hwangbo, "Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4630–4637, 2022.

[19] S. Choi, G. Ji, J. Park, H. Kim, J. Mun, J. H. Lee, and J. Hwangbo, "Learning quadrupedal locomotion on deformable terrain," *Science Robotics*, vol. 8, no. 74, p. eade2256, 2023.

[20] I. M. Aswin 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)*, 2023, pp. 5078–5084.

[21] D. Youm, H. Jung, H. Kim, J. Hwangbo, H.-W. Park, and S. Ha, "Imitating and finetuning model predictive control for robust and symmetric quadrupedal locomotion," *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7799–7806, 2023.

[22] 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.

[23] G. Kim, Y.-H. Lee, and H.-W. Park, "A learning framework for diverse legged robot locomotion using barrier-based style rewards," *arXiv preprint arXiv:2409.15780*, 2024.

[24] Y.-H. Shin, T.-G. Song, G. Ji, and H.-W. Park, "Reinforcement learning for high-speed quadrupedal locomotion with motor operating region constraints: Mitigating motor model discrepancies through torque clipping in realistic motor operating region," *IEEE Robotics & Automation Magazine*, 2024.

[25] J. Hwangbo, J. Lee, and M. Hutter, "Per-contact iteration method for solving contact dynamics," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 895–902, 2018.