

START: Traversing Sparse Footholds with Terrain Reconstruction

Ruiqi Yu¹, Qianshi Wang¹, Hongyi Li², Zheng Jun³, Zhicheng Wang¹, Jun Wu¹, and Qiuguo Zhu*¹

Abstract—Traversing terrains with sparse footholds like legged animals presents a promising yet challenging task for quadruped robots, as it requires precise environmental perception and agile control to secure safe foot placement while maintaining dynamic stability. Model-based hierarchical controllers excel in laboratory settings, but suffer from limited generalization and overly conservative behaviors. End-to-end learning-based approaches unlock greater flexibility and adaptability, but existing state-of-the-art methods either rely on heightmaps that introduce noise and complex, costly pipelines, or implicitly infer terrain features from egocentric depth images, often missing accurate critical geometric cues and leading to inefficient learning and rigid gaits. To overcome these limitations, we propose START, a single-stage learning framework that enables agile, stable locomotion on highly sparse and randomized footholds. START leverages only low-cost onboard vision and proprioception to accurately reconstruct local terrain heightmap, providing an explicit intermediate representation to convey essential features relevant to sparse foothold regions. This supports comprehensive environmental understanding and precise terrain assessment, reducing exploration cost and accelerating skill acquisition. Experimental results demonstrate that START achieves zero-shot transfer across diverse real-world scenarios, showcasing superior adaptability, precise foothold placement, and robust locomotion.

Index Terms—Legged robots, reinforcement learning, deep learning for visual perception.

I. INTRODUCTION

LEGGED animals demonstrate exceptional locomotion capabilities on diverse unstructured and complex terrains. Even in scenarios with extremely sparse footholds, such as stepping across streams on irregular stones or balancing on narrow branches, they traverse at high speed while maintaining stability. Unlike well-structured continuous terrains, where animals could rely solely on rapid proprioceptive feedback to perform dynamic locomotion, these discrete scenarios leave

Manuscript received: August, 18, 2025; Revised November, 14, 2025; Accepted December, 8, 2025.

This paper was recommended for publication by Editor Abhinav Valada upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the "Leading Goose" R&D Program of Zhejiang (Grant No. 2023C01177), the National Key R&D Program of China (Grant No. 2022YFB4701502), and the 2035 Key Technological Innovation Program of Ningbo City (Grant No. 2024Z300).

¹Authors are with Institute of Cyber-Systems and Control, Zhejiang University, 310027, China. {yrq, niujiao, 3160105273}@zju.edu.cn, jwu@iipc.zju.edu.cn, qgzhu@zju.edu.cn.

²Author is with MARMoT Lab, National University of Singapore, 119077. hongyi.li@u.nus.edu.

³Author is with the 19th Asian Games Hangzhou 2022 Organising Committee, 310020. zj558@126.com.

*Qiuguo Zhu is the corresponding author.

Supplementary video is available at https://youtu.be/fj85qVv_kVyI.

Digital Object Identifier (DOI): see top of this page.

©2026 IEEE

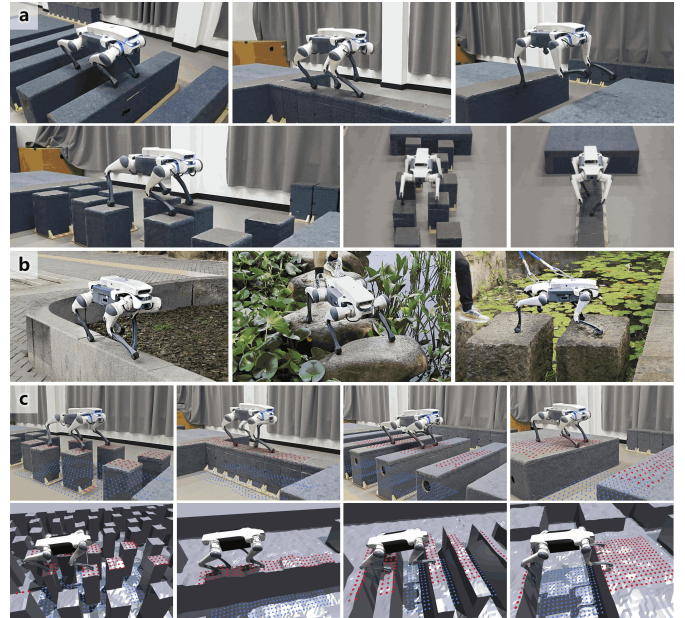


Fig. 1. Our framework enables robot to perform agile and robust locomotion on challenging sparse foothold terrains, including stepping stones, balance beams, stepping beams, and gaps, with only onboard view-limited egocentric vision. The learned policy achieves zero-shot transfer to complex indoor and outdoor real world environments. (c) Precise local terrain reconstructions from TR-Net in the real world and simulation, accurately reproducing the geometric features essential for traversing sparse footholds.

virtually no margin for error, even a slight misstep can lead to a fall [1]. Therefore, to achieve similar performance on risky terrains with sparse footholds, legged robots need the ability to leverage exteroception to accurately perceive their surroundings and employ tightly coupled feedforward control to adjust gait and body posture, ensuring each footfall lands within a safe region to maintain stable, agile locomotion [2].

Model-based hierarchical controllers have been developed to tackle this challenge [3]–[7], decomposing sparse foothold locomotion tasks into three stages: perception, planning, and control [8]. Robots utilize simultaneous localization and mapping (SLAM) to generate heightmaps, solve motion trajectories via nonlinear optimization or graph search, and track them with low-level controllers [8]–[12]. Although effective in laboratory settings, these methods demand extensive hand-crafted design or parameter tuning, and the tradeoff between model fidelity and computational cost limits scalability to complex real-world scenarios [8], [9], [12]. Recent studies integrate learning-based planning and control methods to overcome these challenges [8], [12]–[14]. For instance, Jenelten et al. [15] combined model-based trajectory optimization for foothold planning with reinforcement-learning-based tracking, achieving state-of-the-art performance in challenging environ-

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

ments. Nevertheless, such decoupled planning-control architectures often require simplified, conservative assumptions to mitigate model mismatch, which could constrain each module’s performance and yield suboptimal, overly conservative behaviors [9], [16], [17].

To further unlock policy flexibility and adaptability, recent studies have developed model-free end-to-end reinforcement learning (RL) approaches to learn direct mappings from observations to joint actions, achieving significant progress in various structured environments [18]–[24]. Nonetheless, locomotion over complex terrains with sparse, discrete footholds imposes stricter demands on timely, comprehensive, and precise terrain perception to inform action decisions. Since heightmaps theoretically offer a compact representation with clear physical meaning [25], Zhang et al. [16] and He et al. [26] have adopted local heightmaps as exteroception, enabling agile locomotion on risky sparse footholds. However, such clean heightmaps rely on external motion-capture (MoCap) systems for state estimation, restricting real-world applicability. To extend these methods outdoors, [27], [28] perform model-based heightmap construction by fusing data from odometry and camera or LiDAR, but suffer from drift and estimation errors. Learning-based reconstruction methods address this by building heightmaps from noisy multi-view observations [24], [29], demonstrating superior results yet still requiring the robot’s estimated global pose to align consecutive frames. Moreover, these approaches often involve complicated hardware and introduce heavy computational load [4], [22], unsuitable for low-cost robots.

In response to these limitations, researchers turn to egocentric vision, extracting environmental information directly from single-view depth images to infer and represent surrounding terrain. Most existing work, however, focuses on continuous terrains with height variations [21], [30], such as stairs, slopes, or parkour obstacles, while research explicitly targeting extremely sparse footholds remains limited. We find that current approaches often use implicit vectors to encode terrain information [19], [20], [22], [23], [31]. Although this strategy increases information density and facilitates efficient signal propagation, the lack of explicit geometric supervision hampers the policy’s ability to accurately and comprehensively capture critical features necessary for traversing sparse footholds, impairing terrain evaluation. This imprecise feature representation often leads policies learning one single averaged gait pattern, restricting their adaptability to terrain sparsity and variability. Additionally, camera occlusions and limited viewpoints introduce partial observability, requiring the policy to maintain precise memory of past observations for reliable terrain assessment [22].

In this paper, we propose **START**, a single-stage end-to-end learning framework that enables robust locomotion on terrains with sparse footholds such as stepping stones, balance beams, and stepping beams, using only view-limited egocentric vision, as shown in Fig. 1. To meet the strict demand for precise terrain perception, we introduce local terrain heightmap as an explicit intermediate representation between perception and output actions to encode critical geometric features (e.g., safe foothold areas and terrain edges) for stable locomotion. We

develop a memory-augmented Terrain Reconstruction Network (TR-Net) that fuses proprioception with single-view depth images to reconstruct a local heightmap covering areas beneath and behind the robot, which serves as input to the subsequent locomotion policy. Experimental results indicate that, compared to implicit embeddings, our high-precision reconstruction captures and retains terrain features more comprehensively, enabling the policy to infer both its motion state and surroundings accurately. This facilitates safe and efficient locomotion on highly complex terrains with reduced exploration costs. Furthermore, local terrain reconstruction obviates the need for global mapping and localization to attain heightmaps, eliminating associated noise and drift. We also propose an Adaptive Sampling (AdaSmpl) method that tightly integrates the training of all network components into one single stage, significantly improving sample efficiency.

We deploy our policy on a low-cost DEEP Robotics Lite3 quadruped robot, and conduct extensive simulation and real-world experiments on challenging sparse footholds, demonstrating agile locomotion, precise foot placement, and robust stability. In summary, our main contributions are:

- **Efficient Learning for Risky Sparse Footholds:** We present a novel single-stage training framework that enables quadruped robot equipped solely with view-limited egocentric vision to learn safe locomotion over highly variable sparse footholds.
- **Terrain Reconstruction:** We develop a memory-augmented TR-Net that generates precise local heightmaps to enhance environmental understanding and accurate foot placement, combined with AdaSmpl to boost sample efficiency and policy performance.
- **Real World Generalization:** Our learned policy demonstrates high-agility locomotion in the real world and achieves zero-shot transfer to diverse outdoor environments, validating its robustness and generalizability.

II. METHOD

Our end-to-end learning-based framework START utilizes a set of neural networks to generate joint targets directly from proprioception and raw depth images, with local heightmap serving as an explicit intermediate representation. It comprises two core components that are jointly optimized in a single-stage pipeline, as shown in Fig. 2. The locomotion policy takes the reconstructed local heightmap and proprioception to generate joint actions (Section II-A). Upstream, TR-Net reconstructs an accurate local heightmap for the policy from view-limited vision and proprioceptive features, clearly encoding the critical geometric cues of sparse footholds with high fidelity (Section II-B). We also integrate the AdaSmpl method to enhance learning efficiency (Section II-C).

A. Terrain-Aware Locomotion Policy

The locomotion policy consists of an Implicit-Explicit Estimator (I-E Estimator) and an Actor Policy. The I-E Estimator, trained via supervised learning, produces both implicit and explicit estimation vectors that feed into the actor policy for action generation. We optimize the actor policy using proximal policy optimization (PPO) [32]. To avoid potential information

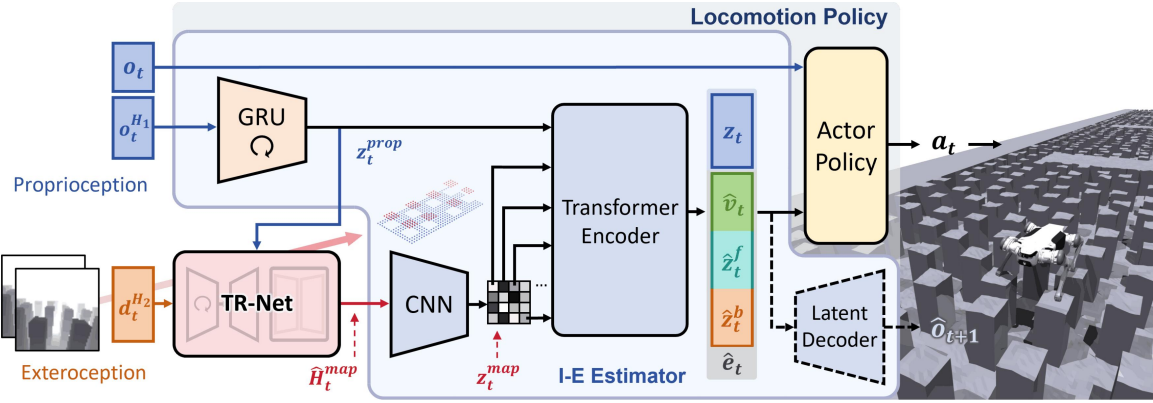


Fig. 2. Overview of the START framework. The single-stage training framework is composed of: a **TR-Net** that receives proprioceptive features and depth images to reconstruct local heightmap in real time, conveying precise terrain features of sparse footholds to the Locomotion Policy; a **Locomotion Policy** leverages an Implicit-Explicit Estimator (**I-E Estimator**) to further enhance environmental understanding and generate actions with safe foot placements.

gap introduced by the two-stage training paradigm [33]–[35], we adopt an asymmetric actor-critic architecture following [34] for single-stage training.

1) *Observations and Action*: The actor policy consumes the latest proprioception o_t , the explicit estimation \hat{e}_t , and the implicit encoding z_t from I-E Estimator, and outputs joint position targets $a_t \in \mathbb{R}^{12}$ for the 12 joints of the quadruped robot. The proprioception $o_t \in \mathbb{R}^{45}$ is measured directly from joint encoders and IMU, defined as:

$$o_t = [\omega_t \quad g_t \quad cmd_t \quad \theta_t \quad \dot{\theta}_t \quad a_{t-1}]^T, \quad (1)$$

where $\omega_t \in \mathbb{R}^3$, $g_t \in \mathbb{R}^3$, $cmd_t \in \mathbb{R}^3$, $\theta_t \in \mathbb{R}^{12}$, $\dot{\theta}_t \in \mathbb{R}^{12}$, and $a_{t-1} \in \mathbb{R}^{12}$ denote body angular velocity, gravity vector in the robot frame, velocity command, joint angle, joint angular velocity, and the previous action, respectively. The critic policy receives not only o_t but also privileged environmental information s_t , which is defined as:

$$s_t = [v_t \quad c_t \quad H_t^b \quad H_t^f]^T, \quad (2)$$

where v_t , c_t , H_t^b , and H_t^f denote the ground-truth base velocity, foot-ground contact state, and heightmaps around the robot's body and four feet, respectively.

2) *I-E Estimator*: Robots can infer their own state and surroundings from observations [36]. By guiding both implicit and explicit estimation, they can better extract privileged information to enhance environmental awareness and adaptability [34], [37]. Based on this insight, we adopt a multi-head autoencoder-based I-E Estimator that fuses proprioception and exteroception to estimate robot state and privileged terrain information. Within I-E Estimator, a gate recurrent unit (GRU) processes temporal proprioception observations o_t^{H1} to produce proprioceptive features z_t^{prop} , while a convolutional neural network (CNN) encoder extracts terrain features z_t^{map} from reconstructed local heightmap \hat{H}_t^{map} . A set of transformer encoder layers then fuse these features into cross-modal estimation vectors, implicit latent z_t and explicit encoding \hat{e}_t , for the actor policy. Since each terrain token corresponds to a specific local region, self-attention enables spatial reasoning and focus on relevant areas [21].

The explicit encoding is defined as $\hat{e}_t = [\hat{v}_t \quad \hat{z}_t^b \quad \hat{z}_t^f]^T$, where the base velocity \hat{v}_t provides the robot's motion state, \hat{z}_t^b and \hat{z}_t^f represents the explicit descriptors for body-centric and foot-centric heightmaps to enhance terrain awareness, and will be further decoded into \hat{H}_t^b and \hat{H}_t^f (within 0.1m), respectively. In contrast to [31], which predicts only foot clearance, our per-

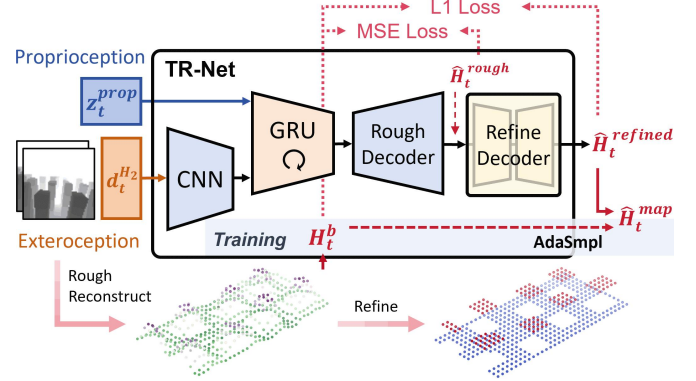


Fig. 3. TR-Net extracts and retains essential visual information, fusing them with proprioceptive features to reconstruct and refine the local terrain heightmap that includes regions beneath and behind the robot. During training, we employ AdaSmpl method to enhance the exploration efficiency of the locomotion policy.

foot heightmap estimation captures each footstep's positions and detailed terrain context for edge-avoiding foot placements. During training, all estimations are supervised against privileged ground-truth with a hybrid loss:

$$\mathcal{L}_{IE} = D_{KL}(q(z_t | o_t^{H1}, \hat{H}_t^{map}) || p(z_t)) + \text{MSE}(\hat{o}_{t+1}, o_{t+1}) + \text{MSE}(\hat{v}_t, v_t) + \text{MSE}(\hat{H}_t^f, H_t^f) + \text{MSE}(\hat{H}_t^b, H_t^b), \quad (3)$$

where the implicit encoding z_t is obtained by a variational autoencoder (VAE) structure, and \hat{o}_{t+1} is predicted from all estimation vectors to match the next-step proprioception o_{t+1} .

B. TR-Net

TR-Net receives proprioceptive feature z_t^{prop} and temporal depth images d_t^{H2} ($H_2 = 2$) as inputs, and reconstructs a precise heightmap in the robot's local frame. The heightmap spans from 0.5m behind to 1.1m in front of the robot, with a width of 0.8m and a spatial resolution of 5cm. Examples of reconstruction results in the real world and simulation are illustrated in Fig. 1(c).

As shown in Fig. 3, TR-Net first applies a CNN encoder to extract visual features from depth images d_t^{H2} , which are concatenated with proprioceptive features z_t^{prop} that encodes the robot's motion state. The combined vector is then fed to a GRU module to retain information about terrain currently occluded beneath and behind the robot. A multi-layer perceptron (MLP) decoder subsequently maps the integrated representation to a rough heightmap \hat{H}_t^{rough} . To further improve reconstruction accuracy, inspired by Duan *et al.* [38], we incorporate a

TABLE I
DETAIL REWARD TERMS.

Reward Term	Equation	Weight
Lin. vel. tracking	$\exp(-\ \min(v, v^{\text{cmd}}) - v^{\text{cmd}}\ ^2 / 0.25)$	1.5
Ang. vel. tracking	$\exp(-\ \omega_{yaw} - \omega_{yaw}^{\text{cmd}}\ ^2 / 0.25)$	0.5
Lin. vel. (z)	v_z^2	-2.0
Ang. vel. (xy)	$\ w_{xy}\ ^2$	-0.05
Orientation	$g_x^2 + g_y^2$	-1.0
Torques	$\sum_{j \in \text{joints}} \tau_j ^2$	-1.0×10^{-5}
Action rate	$\sum_{j \in \text{joints}} a_t - a_{t-1} ^2$	-0.01
Smoothness	$\sum_{j \in \text{joints}} a_t - 2a_{t-1} + a_{t-2} ^2$	-0.01
Joint power	$\sum_{j \in \text{joints}} \tau_j \dot{q}_j $	-2.0×10^{-5}
Joint accelerations	$\sum_{j \in \text{joints}} \ddot{q}_j ^2$	-2.5×10^{-7}
Joint error	$\sum_{j \in \text{joints}} q_j - q_j^{\text{default}} ^2$	-0.01
Collision	$\sum_{i \in \text{contact}} \mathbf{1}\{F_i^z > 0.1\}$	-10.0
Stumble	$\mathbf{1}\{\exists i, F_i^{xy} > 4 F_i^z \}$	-1.0
Feet edge	$\sum_{i \in \text{feet}} c_i \cdot \sum_{d \in \text{dist}} \omega_d E_d[p_i]$	-1.0

U-Net–based refinement decoder to recover high-frequency geometric details while suppressing noise and irregular, blurry edges, producing a refined heightmap $\hat{H}_t^{\text{refined}}$. We supervise \hat{H}_t^{rough} with mean squared error (MSE) loss against the ground-truth H_t^b , while the refined reconstruction $\hat{H}_t^{\text{refined}}$ is optimized with an L1 loss to mitigate the over-smoothing tendency of MSE and to improve accuracy at edges and flat surfaces, thereby reducing the burden on the locomotion policy for terrain feature extraction. The loss for TR-Net is defined as:

$$\mathcal{L}_{\text{TR}} = \text{MSE}(\hat{H}_t^{\text{rough}}, H_t^b) + \text{L1}(\hat{H}_t^{\text{refined}}, H_t^b). \quad (4)$$

C. Adaptive Sampling

Although TR-Net enhances accurate feature representation, its early outputs often remain noisy and incomplete, with invalid data obscuring clear geometric cues. To reduce early-stage exploration cost of the locomotion policy and improve sample efficiency, we introduce Adaptive Sampling (AdaSmpl), which probabilistically substitutes reconstructed heightmaps with ground-truth as inputs to the locomotion policy during training. The sampling probability is adapted based on training performance:

$$p_{\text{smp}} = \tanh(\text{CV}(R)), \quad (5)$$

where CV denotes the coefficient of variation of the episode rewards R . High CV indicates unstable performance and triggers more frequent ground-truth sampling, providing privileged terrain information to ease sparse-reward exploration in complex environments. As training stabilizes and CV decreases, the policy gradually relies more on reconstructed heightmaps. This schedule progressively exposes the policy to realistic, unstructured reconstruction noise, thereby enhancing robustness against noise and preventing performance degradation upon sim-to-real transfer. During deployment, the locomotion policy uses only reconstructed heightmaps.

D. Training Environment

1) *Training Details*: We conduct single-stage training on Isaac Gym [39] using 3072 parallel instances of DEEP Robotics Lite3 quadruped robot on one NVIDIA RTX A6000 GPU for approximately 10000 iterations/ 16.6 hours. To accelerate training, depth images at 60×60 resolution are rendered with NVIDIA Warp instead of the default camera interface [40]. We reset any robot that flips beyond predefined angular limits or whose foot height drops below a specified threshold beneath the ground plane. We also apply domain randomization following [31] to improve policy robustness.

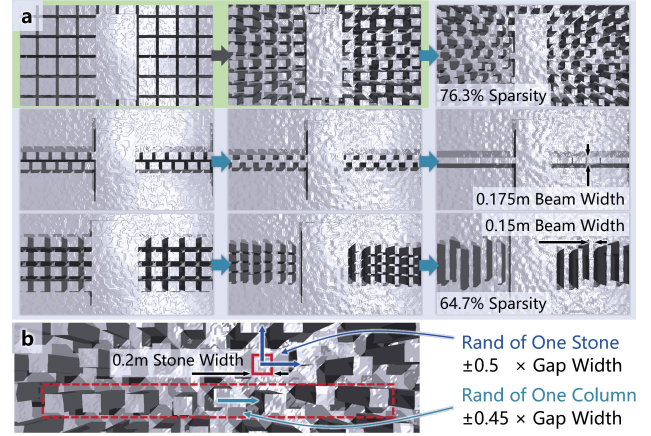


Fig. 4. (a) Robots initially train on lower-randomness stepping stones (light green) to acquire basic motor skills and then progressively transition to more challenging sparse foothold terrains (light blue). (b) For higher-randomness stepping stones, we implement two levels of stone position randomization.

Since terrains align along the x axis, we sample forward velocity commands uniformly from $[0.0, 1.5]$ m/s in the robot frame, with the lateral velocity fixed at zero. To enable turning, we clamp forward commands below 0.3 m/s to zero, then sample an angular velocity from $[-1.2, 1.2]$ rad/s when the command is zero. This configuration trains a policy capable of simultaneous forward walking and turning.

2) *Rewards*: We adopt relatively simple reward functions following [41] to demonstrate our framework enhances policy performance on sparse footholds. To encourage foot placements within safe regions, we introduce an edge-penalty:

$$r_{\text{feetedge}} = - \sum_{i \in \text{feet}} c_i \cdot \sum_{d \in \text{dist}} \omega_d \cdot E_d[p_i], \quad (6)$$

where c_i is the contact state, and E_d is a boolean mask indicating whether foot position p_i lies within d cm from an edge. We set $\text{dist} = [2.5, 5.0]$ cm with penalty weights $\omega_d = [1.0, 0.5]$. The complete rewards are listed in Table I.

3) *Terrains and Terrain Progressive Curriculum*: We generate four types of terrain during training, each with 10 difficulty levels, and increase difficulty following the curriculum in [41]. These terrains include three archetypal sparse foothold scenarios: stepping stones, balance beams, and stepping beams, as shown in Fig. 4. The stepping stones feature up to 76.3% sparsity¹, with stones 0.2m wide and gaps of 0.25m along x and 0.175m along y . Balance beams are as narrow as 0.175m. Stepping beams reach up to 64.7% sparsity, with 0.15m beam widths and 0.275m gaps. We also include gaps up to 0.7m wide ($2 \times$ robot length). Each patch measures $8\text{m} \times 4\text{m}$. To promote adaptability, we sample terrain depths from $[0.2, 0.7]$ m and apply height variations up to ± 5 cm to all stones and beams.

To further reduce the difficulty of early-stage exploration and accelerate adaptive behaviors learning, we introduce an automatic terrain progressive curriculum, inspired by Zhang et al. [16]. We categorize stepping stones into low- and high-randomness groups. Next, we incorporate balance beams and stepping beams as progressive transitions beyond stepping stones, as shown in Fig. 4(a). Initially, robots on sparse foothold terrains train on low-randomness stepping stones and auxiliary flat ground to acquire basic locomotion skills.

¹Sparsity is defined as the ratio of the non-steppable area to the total terrain area [16].

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

As training advances, they are probabilistically allocated to high-randomness stepping stones, balance beams, and stepping beams. The probability follows a linear schedule: $p_{\text{advance}} = \max(\min(\frac{p_{\text{max}}(T - T_{\text{start}})}{T_{\text{end}} - T_{\text{start}}}, 0), p_{\text{max}})$. Gaps are excluded from this progression. Ultimately, the policy learns adaptive locomotion skills across all challenging terrains.

To ensure robustness on stepping stones with high sparsity and randomness, we randomize stone positions by shifting each stone along both axes by up to $0.5 \times$ the gap and offsetting entire columns along the x-axis by up to $0.45 \times$ the gap, as shown in Fig. 4(b).

III. EXPERIMENT

A. Ablation Studies

We compare our START framework against eight ablation methods. The first group evaluates model architecture:

- **HMap GT**: This policy feeds the locomotion policy with ground-truth heightmap as input, with all other settings identical to START, representing an upper-bound performance. It provides a comparison using identical input sources to methods that obtain heightmaps via mapping or external MoCap [16], [26].
- **PIE (START w/o TR-Net)**: This policy removes TR-Net and encodes terrain information implicitly with a 2D-GRU. Its network structure is identical to PIE [31], providing a fair comparison. It also assesses the efficacy of explicit heightmap reconstruction for conveying key terrain geometry.
- **TR-Net w/o Prop**: This policy omits proprioceptive feature input in TR-Net and relies only on depth images.
- **TR-Net w/o GRU**: This policy replaces the GRU in TR-Net with an MLP encoder to disable temporal memory.
- **TR-Net w/o Refine**: This policy skips U-Net-based refinement in TR-Net and feeds the rough heightmap directly to the locomotion policy.

The second group evaluates training pipeline components:

- **START w/o AdaSmpl**: This policy is trained without the AdaSmpl method, so the locomotion policy always uses reconstructed heightmaps as input.
- **START w/o FtEdge**: This policy is trained without foot-edge penalty term r_{feetedge} .
- **START w/o TerProg**: This policy is trained without the terrain progressive curriculum and is exposed to all highly risky sparse foothold terrains from scratch.

We exclude implicit-explicit estimation from ablation study as it was already evaluated in prior work [31], [37]. We use three metrics: (1) **success rate**: the proportion of runs that successfully complete the full distance, evaluated per difficulty level; (2) **traversing rate**: the ratio of distance covered before falling to total distance, evaluated per difficulty level; and (3) **mean edge violation (MEV)**: the proportion of steps landing on terrain edges, measured at the highest difficulty level. For methods involving TR-Net, we also measure reconstruction errors in simulation. All experiments were conducted with 500 runs over a 6m traversal on 10 randomly generated terrains.

B. Simulation Experiments

1) **Terrain Reconstruction Accuracy**: We assess the reconstruction performance of TR-Net variants. As shown in Fig.

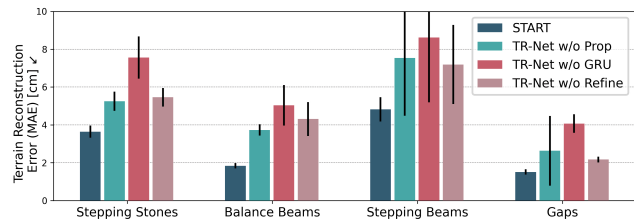


Fig. 5. Terrain reconstruction error for various TR-Nets at the highest terrain difficulty level. Higher errors indicate poorer reconstruction performance.

5, lacking temporal memory, TR-Net w/o GRU exhibits the highest error, since it can only reconstruct terrain visible from the current viewpoint. TR-Net w/o Prop, which receives only depth images, performs only slightly better than the former. It neglects important motion states (e.g., base velocity) from proprioceptive features, impairing its ability to correlate current observation with past reconstructions and update its memory accurately. TR-Net w/o Refine omits refinement process and relies solely on a rough decoder, yielding noisy outputs with irregular edges. This limitation stems from restricted spatial awareness and poor high-frequency detail modeling of a single MLP, preventing accurate edge geometry capture. In contrast, our full TR-Net fuses proprioception and vision, leverages GRU-based temporal context and U-Net spatial-detail refinement, thereby achieving superior reconstruction accuracy across all complex terrains.

Notably, the focus of terrain reconstruction in our work is to help enhance robot locomotion, rather than mapping itself. By framing terrain reconstruction as an explicit terrain-feature extraction task, we produce an intermediate representation that connects accurate perception from limited egocentric vision with rapid, safe actions. Compared to approaches like [27]–[29], START requires only single-view depth images from a low-cost onboard camera, thereby avoiding the complexities of mapping and global estimation.

2) **Policy Performance**: Results in Fig. 6 demonstrate the advantages of our START framework over all ablation methods across diverse terrains. First, it underscores the importance of explicit local heightmaps reconstructed by TR-Net for accurately encoding foothold-relevant terrain features, which enable stable traversal on sparse footholds. In contrast, PIE (START w/o TR-Net) relies on implicit feature vectors extracted directly from depth images. These vectors lack the clear physical meaning of heightmaps and thus struggle to capture precise geometric details. Consequently, the robot’s perception of rapidly changing terrain is impaired, causing it to converge to a single, rigid action pattern. Although performance remains acceptable on low-sparsity terrains, the policy cannot adapt to higher difficulty or greater randomization.

Fig. 7 further highlights these differences at the highest difficulty level of stepping stones. The foothold distribution of PIE (START w/o TR-Net) shifts toward edges, exhibits greater variability and leads to frequent failures. Conversely, START proactively places footholds near stone centers similar as the upper-bound HMap GT, producing consistently concentrated distributions even under strong domain randomization, demonstrating precise, robust foothold selection.

The accuracy of terrain reconstruction is also essential for achieving high policy performance. Both TR-Net w/o Prop

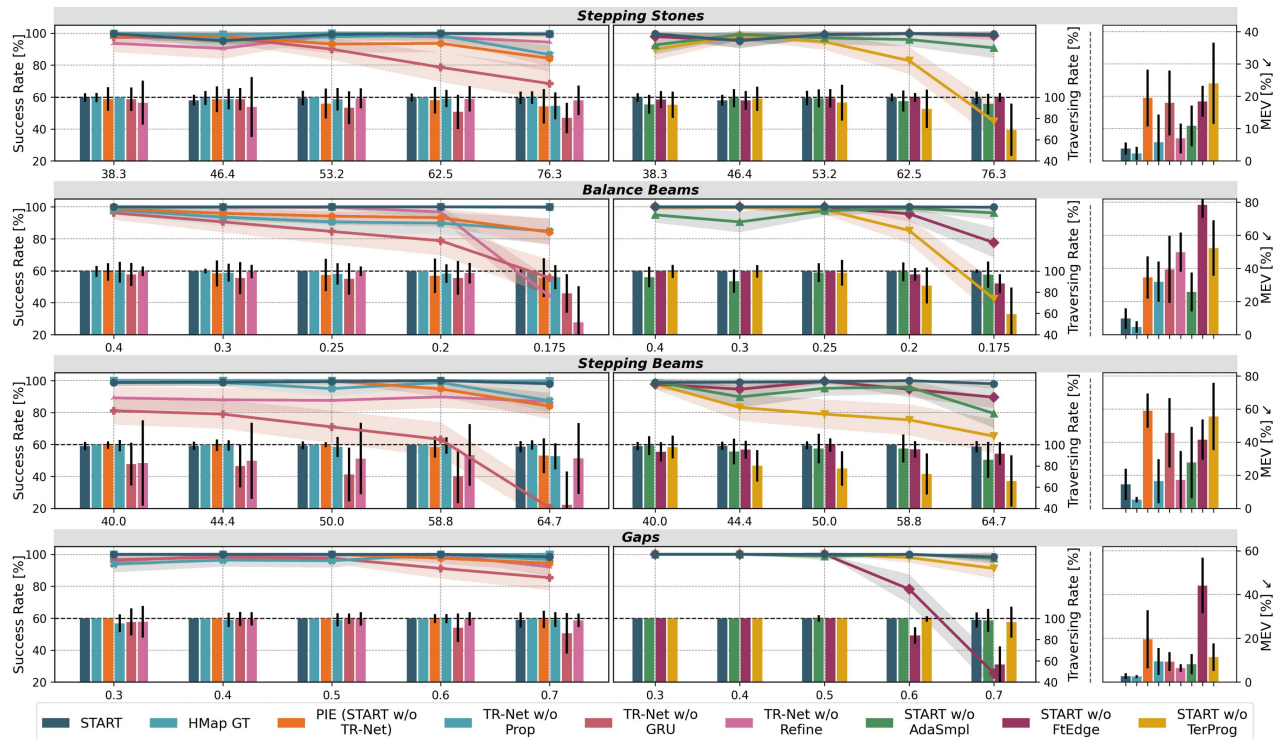


Fig. 6. Ablation studies in simulation. The **left** and **middle** columns show **success rate** (line) and **traversing rate** (bar) for model-architecture and training-pipeline groups, both plotted against terrain difficulty. Difficulty is defined by sparsity (with maximum randomization) for stepping stones and stepping beams, and by beam and gap width for balance beams and gaps. The **right** column reports **mean edge violations (MEV)** at the highest difficulty level.

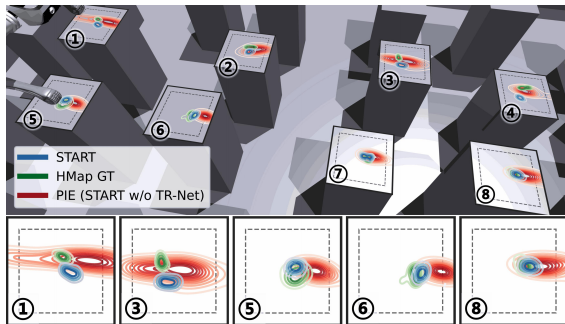


Fig. 7. Foothold locations distribution for START, HMap GT, and PIE (START w/o TR-Net) on the highest difficulty level of stepping stones. We conducted 500 runs per policy and converted recorded footholds into continuous distributions via kernel density estimation (KDE).

and TR-Net w/o GRU conduct explicit reconstruction but omit proprioceptive information or temporal memory. This prevents them from fully capturing terrain features and leads to inferior performance compared to START. Without U-Net refinement, although TR-Net w/o Refine can generate rough heightmaps that approximate safe foothold regions (e.g., stone and beam centers), it fails to delineate edges precisely. This inaccuracy increases MEV and reduces success rates on terrains like balance beams where edge precision is vital. Overall, START leverages high-fidelity local heightmaps to significantly enhance adaptability, robustness, and agility on challenging sparse footholds, even when relying solely on limited egocentric vision.

We also conducted ablation studies on different training pipelines. AdaSmpl plays a key role in improving exploration efficiency of locomotion policy. START w/o AdaSmpl is exposed to noisy, low-information reconstructions early in training, which hampers extraction and use of critical geometric cues like terrain edges. As illustrated in Fig. 8, the policy

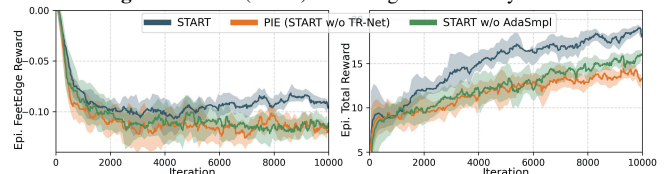


Fig. 8. r_{feetedge} and total reward in one episode for methods of START, PIE (START w/o TR-Net), and START w/o AdaSmpl. The former reflects the robot’s ability to select safe footholds. START achieves higher r_{feetedge} and faster convergence, demonstrating the efficiency of the training framework.

requires more training episodes to reach the same reward level as START, thereby degrading performance. Similarly, removing the foot-edge reward r_{feetedge} (START w/o FtEdge) or the terrain progressive curriculum (START w/o TerProg) leads to significant performance drops. Lacking foot-edge guidance, START w/o FtEdge often places feet on edges, causing instability, falls, and extremely high MEV. Without progressive exposure from easier environments, START w/o TerProg trains directly on challenging terrains, where sparse rewards and conflicting optimal gaits hinder incremental skill discovery and result in overall failure. These results highlight another strength of START: a simple yet effective progressive learning pipeline that guides the policy from easy to hard tasks, enabling mastery of demanding locomotion skills and preserving superior adaptability across all risky terrains.

C. Real-World Lab-Level Experiments

We use DEEP Robotics Lite3 quadruped robot for real-world deployment. An Intel RealSense D435i camera on its head captures raw depth at 10 Hz. We apply decimation, temporal, and hole-filling filters from the librealsense utilities to obtain 60×60 images. Network inference runs on an onboard Jetson Orin NX with TensorRT acceleration, outputting action

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE II
ABLATION STUDIES IN THE REAL WORLD. EACH METHOD WAS TESTED 5 TIMES WITH SUCCESS RATE AND TRAVERSING RATE RECORDED.

Ablations	<i>Stepping Stones</i> 76.3% sparsity		<i>Balance Beams</i> 0.2m width		<i>Stepping Beams</i> 58.8% sparsity		<i>Gaps</i> 0.7m width
	Success Rate	Traversing Rate	Success Rate	Traversing Rate	Success Rate	Traversing Rate	Success Rate
START	1.0	1.00±0.00	0.8	0.99±0.03	1.0	1.00±0.00	1.0
PIE (START w/o TR-Net)	0.2	0.58±0.27	0.6	0.84±0.21	0.4	0.66±0.32	0.8
TR-Net w/o Prop	0.4	0.89±0.19	0.4	0.75±0.19	0.6	0.85±0.25	0.6
TR-Net w/o GRU	0.2	0.53±0.28	0.2	0.74±0.25	0.2	0.42±0.33	0.4
TR-Net w/o Refine	0.4	0.74±0.27	0.6	0.92±0.15	0.4	0.71±0.34	0.2
START w/o AdaSmpl	0.8	0.97±0.06	0.6	0.89±0.16	0.8	0.99±0.03	0.8
START w/o FtEdge	0.6	0.75±0.34	0.4	0.78±0.21	0.4	0.78±0.31	0.0
START w/o TerProg	0.2	0.55±0.37	0.4	0.87±0.14	0.6	0.80±0.28	0.4

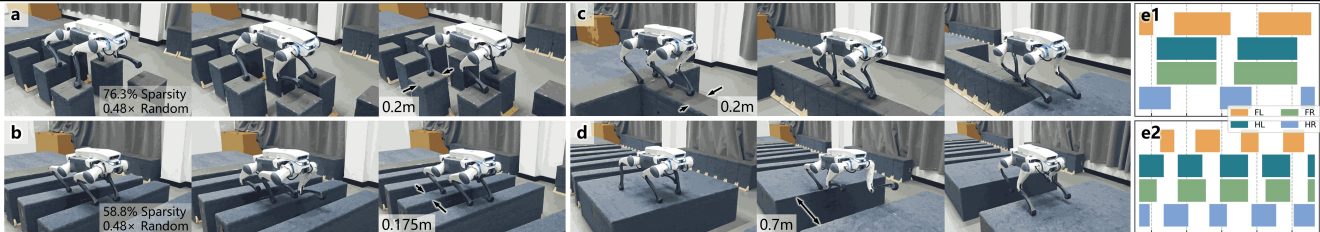


Fig. 9. (a-d): Key frames of the robot traversing sparse foothold terrains in the real world, demonstrating agile, stable locomotion and sim-to-real generalization. (e): The robot exhibits different gait patterns on stepping stones and balance beams. (a)(e1): The robot traverses stepping stones slowly at low step frequency, placing each foot centrally within the sparse foothold area. (b)(e2): It first aligns its body over the beam center, then rapidly shifting the legs into a dynamically stable inverted-triangle stance and crossing the balance beam at high step frequency to counteract external disturbances.

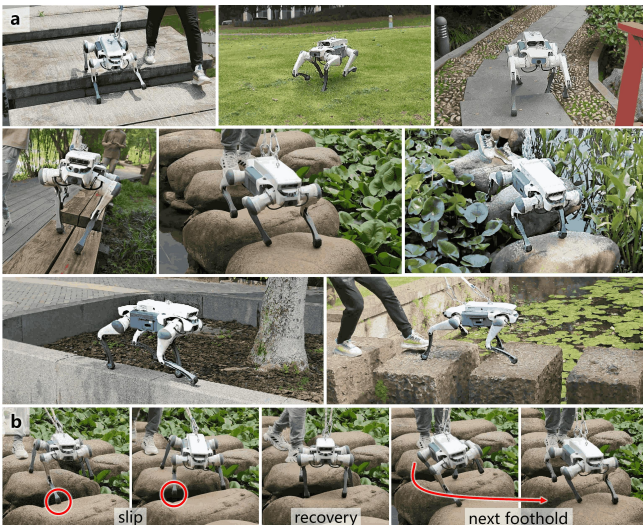


Fig. 10. (a) We conduct outdoor experiments to validate START's locomotion performance in diverse complex sparse foothold terrains, including grassy ramps, park benches, irregularly shaped single-column stepping stones over water, and narrow (0.15m wide) slippery flowerbed edges, etc. (b) When confronted with uneven-surface stones never seen during training, the robot recovers rapidly from a slip and redirects to the next safe foothold.

at 50 Hz. The learned policy can zero-shot transfer to complex real-world environments, as shown in Fig. 9.

To further validate the effectiveness of the proposed framework, we built four real-world terrains and compared START against all ablation methods (HMap GT excluded due to the lack of MoCap device), as summarized in Table II. Results show that PIE (START w/o TR-Net) suffers severe degradation in real world tests, experiencing missteps, edge placements, or directional deviations that lead to failures. Similarly, START w/o TerProg, trained directly on risky sparse foothold terrains, lacks adaptiveness and often attempts infeasible actions in the real world. In contrast, START leverages accurate explicit terrain reconstruction to match simulation performance across all terrains and robustly outperforms other methods. It traverses challenging sparse footholds at speeds up to 1.5m/s, continuously adjusting its gait to secure stable footholds. To

our knowledge, no existing quadruped robot relying solely on onboard egocentric depth images has demonstrated comparable capabilities.

D. Real-World Outdoor Experiments

We also conducted extensive experiments in outdoor natural sparse foothold scenarios, as shown in Fig. 10. Notably, although our method was never exposed to single-column stepping stones with uneven surfaces and irregular shapes, or vegetation obstructions during training, the policy's precise terrain perception enables safe foothold detection and stable foot placement, demonstrating excellent robustness and zero-shot generalization. Moreover, without specialized fine-tuning, START seamlessly adapts to rugged, non-flat terrains such as grassy ramps and descending staircases, highlighting its exceptional balance and adaptability.

IV. CONCLUSION

Rapid, accurate exteroceptive perception and dynamic gait adjustment are essential for traversing risky sparse foothold terrains. In this work, we present START, an end-to-end learning framework that, for the first time, enables agile quadruped locomotion with high traversability on highly sparse and randomized terrains using only low-cost onboard egocentric vision and proprioception. START employs local terrain heightmaps as an explicit intermediary to convey critical geometric features for stable locomotion. This representation allows the robot to infer rapidly varying surroundings and generate precise foothold placements, enhancing adaptability and stability. START also integrates the AdaSmpl method and the terrain progressive curriculum to unify all training components into a single-stage pipeline for improving exploration efficiency. Extensive simulation and real-world experiments demonstrate START's robustness and generalization across diverse challenging terrains and its zero-shot sim-to-real transfer to complex real-world environments.

Although only four representative terrains were evaluated, START can extend to more generalized complex scenarios

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

(e.g. stairs or boulders with height variations) and their combinations. In future work, we will extend START to more diverse settings to better cover real-world complexity and improve locomotion performance. To enhance environmental understanding, we will introduce finer edge-preservation mechanisms in heightmap refinement to better retain critical terrain features. We further plan to explore more expressive intermediate representations, such as voxels, that encode terrain geometry with higher fidelity and completeness.

REFERENCES

- [1] B. J. Farrell, M. A. Bulgakova, M. G. Sirota, B. I. Prilutsky, and I. N. Beloozerova, "Accurate stepping on a narrow path: mechanics, emg, and motor cortex activity in the cat," *Journal of neurophysiology*, vol. 114, no. 5, pp. 2682–2702, 2015.
- [2] V.-G. Loc, S.-g. Roh, I. M. Koo, D. T. Tran, H. M. Kim, H. Moon, H. R. Choi, and S. Park, "Control of a quadruped robot with enhanced adaptability over unstructured terrain," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2671–2676.
- [3] 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.
- [4] A. Agrawal, S. Chen, A. Rai, and K. Sreenath, "Vision-aided dynamic quadrupedal locomotion on discrete terrain using motion libraries," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4708–4714.
- [5] R. J. Griffin, G. Wiedebach, S. McCrory, S. Bertrand, I. Lee, and J. Pratt, "Footstep planning for autonomous walking over rough terrain," in *2019 IEEE-RAS 19th international conference on humanoid robots (humanoids)*. IEEE, 2019, pp. 9–16.
- [6] C. Mastalli, I. Havoutis, M. Focchi, D. G. Caldwell, and C. Semini, "Motion planning for quadrupedal locomotion: Coupled planning, terrain mapping, and whole-body control," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1635–1648, 2020.
- [7] S. Fahmi, V. Barasuol, D. Esteban, O. Villarreal, and C. Semini, "Vital: Vision-based terrain-aware locomotion for legged robots," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 885–904, 2022.
- [8] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, "Visual-locomotion: Learning to walk on complex terrains with vision," in *Conference on Robot Learning*. PMLR, 2022, pp. 1291–1302.
- [9] G. B. Margolis, T. Chen, K. Paigwar, X. Fu, D. Kim, S. bae Kim, and P. Agrawal, "Learning to jump from pixels," in *Conference on Robot Learning*. PMLR, 2022, pp. 1025–1034.
- [10] A. W. Winkler, C. D. Bellicoso, M. Hutter, and J. Buchli, "Gait and trajectory optimization for legged systems through phase-based end-effector parameterization," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1560–1567, 2018.
- [11] O. Villarreal, V. Barasuol, P. M. Wensing, D. G. Caldwell, and C. Semini, "Mpc-based controller with terrain insight for dynamic legged locomotion," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2436–2442.
- [12] 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.
- [13] Z. Xie, X. Da, B. Babich, A. Garg, and M. v. de Panne, "Glide: Generalizable quadrupedal locomotion in diverse environments with a centroidal model," in *International Workshop on the Algorithmic Foundations of Robotics*. Springer, 2022, pp. 523–539.
- [14] V. Tsounis, M. Alge, J. Lee, F. Farshidian, and M. Hutter, "Deepgait: Planning and control of quadrupedal gaits using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3699–3706, 2020.
- [15] F. Jenelten, J. He, F. Farshidian, and M. Hutter, "Dtc: Deep tracking control," *Science Robotics*, vol. 9, no. 86, p. eadh5401, 2024.
- [16] C. Zhang, N. Rudin, D. Hoeller, and M. Hutter, "Learning agile locomotion on risky terrains," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 11 864–11 871.
- [17] C. Plagemann, S. Mischke, S. Prentice, K. Kersting, N. Roy, and W. Burgard, "Learning predictive terrain models for legged robot locomotion," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008, pp. 3545–3552.
- [18] 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.
- [19] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, "Legged locomotion in challenging terrains using egocentric vision," in *Conference on robot learning*. PMLR, 2023, pp. 403–415.
- [20] 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.
- [21] R. Yang, M. Zhang, N. Hansen, H. Xu, and X. Wang, "Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers," in *Deep RL Workshop NeurIPS 2021*, 2021.
- [22] R. Yang, G. Yang, and X. Wang, "Neural volumetric memory for visual locomotion control," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1430–1440.
- [23] Z. Zhuang, Z. Fu, J. Wang, C. G. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, "Robot parkour learning," in *Conference on Robot Learning*. PMLR, 2023, pp. 73–92.
- [24] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, "Anymal parkour: Learning agile navigation for quadrupedal robots," *Science Robotics*, vol. 9, no. 88, p. eadi7566, 2024.
- [25] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, and M. Hutter, "Elevation mapping for locomotion and navigation using gpu," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2273–2280.
- [26] J. He, C. Zhang, F. Jenelten, R. Grandia, M. Bächer, and M. Hutter, "Attention-based map encoding for learning generalized legged locomotion," *Science Robotics*, vol. 10, no. 105, p. eadv3604, 2025.
- [27] P. Fankhauser, M. Bloesch, and M. Hutter, "Probabilistic terrain mapping for mobile robots with uncertain localization," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3019–3026, 2018.
- [28] P. Fankhauser and M. Hutter, "A universal grid map library: Implementation and use case for rough terrain navigation," *Robot Operating System (ROS) The Complete Reference (Volume 1)*, pp. 99–120, 2016.
- [29] D. Hoeller, N. Rudin, C. Choy, A. Anandkumar, and M. Hutter, "Neural scene representation for locomotion on structured terrain," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8667–8674, 2022.
- [30] A. Loquercio, A. Kumar, and J. Malik, "Learning visual locomotion with cross-modal supervision," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7295–7302.
- [31] S. Luo, S. Li, R. Yu, Z. Wang, J. Wu, and Q. Zhu, "Pie: Parkour with implicit-explicit learning framework for legged robots," *IEEE Robotics and Automation Letters*, 2024.
- [32] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [33] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: learning a unified policy for manipulation and locomotion," in *Conference on Robot Learning*. PMLR, 2023, pp. 138–149.
- [34] 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.
- [35] 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.
- [36] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *Robotics: Science and Systems XVII*, 2021.
- [37] Z. Wang, W. Wei, R. Yu, J. Wu, and Q. Zhu, "Toward understanding key estimation in learning robust humanoid locomotion," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 11 232–11 239.
- [38] H. Duan, B. Pandit, M. S. Gadde, B. Van Marum, J. Dao, C. Kim, and A. Fern, "Learning vision-based bipedal locomotion for challenging terrain," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 56–62.
- [39] V. Makovychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa *et al.*, "Isaac gym: High performance gpu based physics simulation for robot learning," in *NeurIPS Datasets and Benchmarks*, 2021.
- [40] M. Macklin, "Warp: A high-performance python framework for gpu simulation and graphics," in *NVIDIA GPU Technology Conference (GTC)*, vol. 3, 2022.
- [41] 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.