StairMaster: Learning to Conquer Risky Hollow Stairs for Agile Quadrupedal Robots

Xincheng Tang[1], Youhan Xie[1], Zhengjie Shu[1], Wanyu Li[1], Lai Jiang[1], Wenkang Hu[1], Yitong Li and Ruigang Yang[1,†]

***Abstract*— Climbing hollow stairs remains a challenging problem for quadruped robots due to the high risk of leg trapping, severe depth sparsity, and high-frequency depth-sensing noise. In this paper, we propose StairMaster, a novel three-stage reinforcement learning framework for stable locomotion on such extreme discontinuous terrains. Our architecture integrates a Cross-Attention mechanism to extract structural features from noisy depth data, alongside a Spatial-aware Recurrent Unit (SRU) that maintains robust spatio-temporal memory to mitigate perception blind spots. To bridge the sim-to-real gap in depth perception, we propose a high-fidelity sim-to-real depth sensor modeling pipeline that faithfully replicates real-world sensor artifacts. Additionally, we employ a 3D waypoint-guided active perception reward for proactive sensing, alongside hollow gap kinematic and stair edge penalties to ensure precise foothold placement. We successfully deployed StairMaster on a Unitree Go2 robot, demonstrating its ability to conquer hollow stairs with an unprecedented incline of up to 55° through zero-shot transfer. To the best of our knowledge, this is the first RL-based policy to achieve such steep hollow stair climbing in real-world environments.**

Project Website: https://sivan666666.github.io/StairMaster/.

I. INTRODUCTION

In nature, quadrupedal animals exhibit a remarkable ability to stably navigate highly complex terrains. Inspired by this biological agility, deep reinforcement learning (DRL) has empowered quadruped robots to go through unstructured environments by fusing visual perception with proprioception [1], [2], [3], [4]. In industrial facilities such as power plants and construction sites, hollow stairs are ubiquitous to reduce structural weight. While quadruped robots boast extensive application potential across diverse industrial scenarios, climbing hollow stairs remains a challenging problem in legged robotics.

Unlike standard solid stairs, hollow stairs feature massive physical gaps between threads, introducing three critical challenges. First, the absence of vertical risers creates a severe risk of leg trapping; even a minuscule foothold error can cause the robot's leg to plunge into the void, leading to catastrophic hardware damage. Second, these stairs are often constructed from reflective or grating materials, inducing severe depth sparsity and causing depth cameras to suffer from massive pixel dropouts and extreme noise. Finally, due to the limited field of view of forward-facing cameras, the narrow threads completely disappear as they pass beneath the robot, forcing it to execute extremely precise hind-leg placements under total visual occlusion.

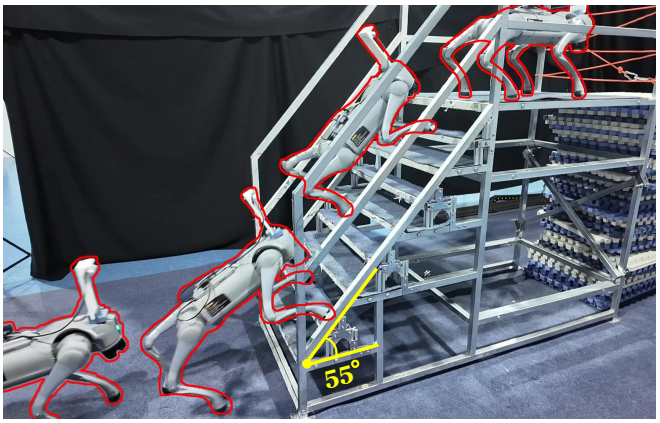


Fig. 1: StairMaster enables a quadrupedal robot to climb real-world 55° hollow stairs.

Existing visuo-motor control frameworks struggle significantly when confronted with the compounding challenges of real-world deployments [5], [6], [7]. First, they must overcome severe depth sensor noise, which is heavily exacerbated by violent camera vibrations induced by intense posture changes during stair climbing [8], [9], [10], [11]. Second, they suffer from critical partial observability; due to the forward-facing camera, the robot completely loses sight of the terrain beneath its feet and the hollow gaps as the threads pass out of view [10], [11]. Navigating such prolonged blind spots necessitates robust, simultaneous temporal and spatial memory. While traditional memoryless policies or simple frame-stacking fail to maintain accurate 3D spatial representations, standard RNNs also exhibit limitations in spatial awareness. They primarily encode one-dimensional temporal dependencies and lack the capability to implicitly align global spatial topology under the robot's ego-motion [12]. Furthermore, conventional reward designs and 2D waypoint-guided frameworks fail to encourage active visual perception. They cannot proactively guide the robot to adjust its pitch angle to observe upcoming stairs, which is a critical capability for conquering exceptionally steep inclines.

To overcome the aforementioned challenges, we propose StairMaster, an end-to-end three-stage reinforcement learning framework designed to conquer high-risk hollow stairs. To address visual blind spots and perception degradation, we introduce a novel visuospatial encoder that combines a Cross-Attention mechanism to extract sparse structural

[1]Shanghai Jiao Tong University
[†]Corresponding author.

features with a Spatial-Aware LSTM [12]. A high-fidelity depth sensor modeling pipeline is proposed to bridge the depth sim-to-real gap. Furthermore, we introduce three customized reward functions tailored for hollow stairs, including a 3D waypoint-guided active perception reward for proactive sensing, alongside hollow gap kinematic and stair edge penalties to ensure precise foothold placement. We deployed StairMaster on a Unitree Go2 robot and validated across diverse real-world hollow stairs, achieving robust performance through zero-shot transfer. Our technical contributions are:

- We present StairMaster, a novel three-stage reinforcement learning framework specifically designed for high-risk hollow stairs. Extensive simulations and real-world experiments on a Unitree Go2 robot demonstrate that StairMaster achieves robust, zero-shot sim-to-real transfer. More importantly, to the best of our knowledge, this is the first RL-based strategy that enables a quadruped robot to conquer hollow stairs with an unprecedented incline of up to 55°.
- We introduce a visuospatial encoder integrating a Cross-Attention mechanism with a Spatial-Aware LSTM to extract structural features and maintain robust spatio-temporal memory, overcoming the limited field of view from sensors.
- We carefully design a depth sensor simulation module to account for sensing artifacts caused by abrupt motion and challenging environments, which include thin structures, hollow spaces, and reflective metallic surfaces. It greatly reduces the sim-to-real domain gap, enabling zero-shot transfer.

II. RELATED WORK

A. Proprioceptive Locomotion

Robust locomotion using purely proprioceptive sensing has laid the foundation for legged robot control. Traditionally, model-based optimization methods enabled basic walking but often struggled with the unpredictability of complex terrains [13]. To overcome this, massive parallel DRL in simulation has allowed robots to learn fundamental walking skills within minutes [1]. Building on this, researchers have introduced various frameworks to estimate latent environmental properties without exteroception, such as motor adaptation [2], [14], [15], implicit terrain imagination [16], collision detection [17] and hybrid internal models [18]. Moreover, specific hook was designed to handle climb ladders [19]. While these proprioceptive-only policies achieve fairly robustness, which allows them to traverse uneven terrains and standard solid stairs [20], [21], they remain inherently blind. This reactive nature fails when encountering sparse and discontinuous structures like hollow stairs because the absence of predictive awareness and early postural adjustment leads to catastrophic missteps into the gaps between threads.

B. Vision-guided Locomotion

The integration of exteroception, such as depth vision and LiDAR, has significantly pushed the boundaries of quadrupedal agility. Early LiDAR-based methods using elevation maps achieved robust locomotion in complex and confined spaces [22], [23], [24], yet they often suffer from high computational overhead, latency, and heavy reliance on precise state estimation. To address these issues, vision-based end-to-end frameworks have gained prominence, enabling robots to perform challenging tasks such as solid stair climbing and gap crossing [5], [7], [25], [26]. Yang et al. [6] further integrated RL with motion tracking, but the policy is limited by a fixed gait. While methods like Extreme Parkour [7] employ a two-stage teacher-student distillation framework, their resilience to real-world sensor noise remains limited. Conversely, single-stage frameworks such as PIE [27] and WMP [28] eliminate the need for privileged teachers but often face slow convergence and prolonged training cycles. More recently, PLANC [29] introduced a three-stage training pipeline to address these perception-action challenges; however, its application is primarily optimized for humanoid robots and does not account for the specific geometric hazards of quadrupedal hollow-stair climbing.

Despite these successes, current vision-guided policies are primarily optimized for solid stairs and fail on hollow ones due to three key deficiencies: critical blind spots beneath the torso caused by front-facing cameras, a lack of spatio-temporal memory to maintain the spatial connectivity of sparse stair threads over time, and high sensitivity to depth noise during violent oscillations. Furthermore, existing methods lack active visual perception to capture anticipated upcoming structures, nor do they explicitly address the severe scarcity of valid footholds caused by the large vertical and horizontal gaps in such discontinuous terrains.

III. METHOD

The proposed StairMaster framework is an end-to-end multi-stage learning-based framework that derives desired joint angle commands directly from raw depth images and onboard proprioception. Our framework introduces a specialized three-stage pipeline tailored to the unique challenges of extreme hollow stairs. This approach enhances the robot's robustness by integrating high-fidelity depth noise modeling for realistic perception, and building spatio-temporal memory to bridge observation gaps. An overview of the framework and its technical details are provided below.

A. Overview

As shown in Figure 2, the StairMaster framework is structured around a three-stage training pipeline designed to enable stable quadrupedal locomotion on hollow stairs.

1) Stage 1: Privileged Teacher Policy Training: In the initial stage, we train a teacher policy which has access to proprioception and privileged heightmaps through the Proximal Policy Optimization (PPO) algorithm. Leveraging our customized reward functions, this stage yields a highly capable teacher policy that acts as an expert demonstration.

Throughout our framework, the proprioceptive state of the robot is explicitly defined as $o_t = [\omega_t, g_t, c_t, \theta_t, \dot{\theta}_t, a_{t-1}]^T$. This observation vector encompasses the base angular velocity

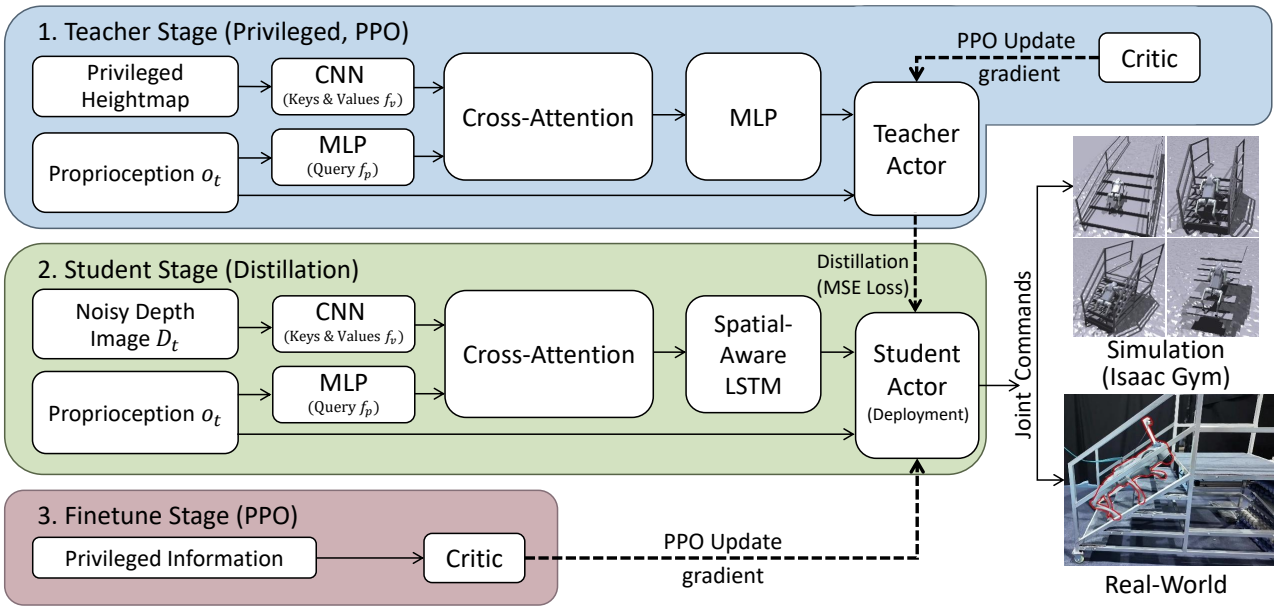


Fig. 2: Overview of StairMaster training framework.

$\omega_t \in \mathbb{R}^3$, the projected gravity vector in the body frame $g_t \in \mathbb{R}^3$, the velocity command $c_t \in \mathbb{R}^3$, the joint positions $\theta_t \in \mathbb{R}^{12}$, the joint velocities $\dot{\theta}_t \in \mathbb{R}^{12}$, and the previous actions $a_{t-1} \in \mathbb{R}^{12}$.

2) Stage 2: Student Policy Distillation: In real-world deployment, the robot can only rely on noisy egocentric depth images and proprioceptive states. To transfer the expert behaviors to the real robot, this second stage trains a student policy to imitate the teacher's actions via a distillation process using Mean Squared Error (MSE) loss. To handle the depth sparsity of hollow stairs, the student network incorporates a Cross-Attention mechanism to dynamically fuse depth and proprioceptive data, a Spatial-Aware LSTM to maintain a robust memory of the terrain and a high-fidelity sim-to-real depth noise modeling.

3) Stage 3: Fine-Tuning: Distillation alone often results in a performance gap. To maximize the robot's agility and robustness on challenging hollow stairs, the third stage fine-tunes the distilled student policy using the PPO algorithm. This final RL-based fine-tuning stage allows the student network to interact directly with the environment, exploring and correcting the sub-optimal behaviors from the distillation, ultimately outputting the final target joint actions for zero-shot real-world deployment.

B. Visuospatial Encoder Architecture

Relying on a single front-mounted depth camera inherently introduces severe visual blind spots during locomotion. Consequently, critical terrain features, such as narrow stair threads, frequently disappear from the sensor's field of view prior to the hind legs engaging with them. To mitigate the challenges posed by this partial observability, we introduce a visuospatial encoder designed to seamlessly fuse multimodal exteroceptive and proprioceptive inputs. This module not only infers immediate terrain geometry but also actively maintains a robust, spatio-temporal memory of the traversed environment to guide stable stepping behaviors in completely occluded regions.

1) Multimodal Feature Extraction and Cross-Attention: At each timestep, the encoder processes the most recent depth frame D_t through a CNN to extract a dense visual feature map f_v. Concurrently, the robot's current proprioceptive state o_t is embedded via a MLP into a high-dimensional latent vector f_p. Recognizing that simple feature concatenation often incorporates a significant amount of task-irrelevant depth pixels, we achieve multimodal integration through a specialized Multi-head Cross-Attention mechanism. By formulating the proprioceptive embedding f_p as the query, and the visual feature map f_v as the corresponding keys and values, the model computes state-dependent attention weights. This architectural design enables the encoder to dynamically focus on task-critical geometric structures such as sparse stair edges that are most relevant to the robot's current kinematic posture, ultimately yielding a highly distilled, state-aware latent representation.

2) Spatio-Temporal Memory with Spatial-Aware LSTM: Due to the camera's forward-facing configuration, the Cross-Attention mechanism only extracts instantaneous local features, lacking the spatio-temporal memory required to track stair threads into visual blind spots. To construct a comprehensive environmental representation, we process the fused features through a Spatial-Aware LSTM (SRU). Unlike standard RNNs that struggle with spatial registration, this module integrates spatial snapshots over time without requiring explicit ego-motion inputs [12].

Specifically, it introduces a learnable spatial transformation gate, s_t, to implicitly align historical memory features with the current perspective. Given the current fused feature

f_t, s_t is computed as:

$$s_t = \sigma(W_s f_t + b_s), \tag{1}$$

where σ is the sigmoid function, and W_s, b_s are learnable parameters. This spatial gate is then applied to the previous hidden state h_{t-1} and cell state c_{t-1} via the Hadamard product ($*$). The complete recurrence is formulated as:

$$h_t, c_t = \text{LSTM}(f_t, s_t * h_{t-1}, s_t * c_{t-1}). \tag{2}$$

By continuously registering frame-by-frame representations, this mechanism builds a robust full-scale spatio-temporal memory buffer. It empowers the downstream actor network to accurately infer the 3D positions of completely occluded hollow threads and generate reliable joint commands.

C. High-Fidelity Sim-to-Real Depth Noise Modeling

Pristine depth maps rendered in physics simulators fundamentally differ from real-world depth sensor outputs, which are notoriously plagued by scattering, absorption, and motion-induced artifacts. To successfully deploy the vision-guided student policy on physical hardware, bridging this severe visual sim-to-real gap is imperative. As illustrated in Figure 3, we implement a comprehensive depth noise modeling pipeline during the training of the visuospatial Encoder to bridge the sim-to-real gap.

1) Spatial and Sensor Artifact Modeling: Real-world depth sensors frequently fail on reflective surfaces or sharp geometric transitions, which are highly prevalent in hollow stair environments. We corrupt the ideal simulated depth images with a mixture of spatial noise models:

- Gaussian and Uniform Noise: We apply a combination of additive Gaussian and uniform noise to simulate the baseline measurement uncertainty and thermal noise inherent in IR-based depth hardware.
- Hole Noise: Hole noise (random pixel dropout) is utilized to simulate regions where infrared rays fail to return such as the reflective surfaces.
- Edge Noise: Specifically for the thin, discontinuous geometry of open-riser stairs, we apply localized edge noise to simulate the bleeding effect and scattering typical of depth discontinuities.

2) Disparity and Dynamic Vibration Noise: Beyond static artifacts, we explicitly model dynamic noise arising from impulsive foot–ground contacts and impact-induced vibrations during locomotion.

- Stereo Matching Noise: We simulate quantization and matching errors typical of binocular depth estimation. This noise is treated as a dynamic artifact because rapid perspective changes and motion blur during climbing lead to transient matching ambiguities on the sparse and narrow stair threads.
- Gaussian Shift: Aggressive impacts induce severe vibrations in the head-mounted camera. We introduce a Gaussian shift noise that randomly translates depth pixels spatially across consecutive frames. This forces the Cross-Attention and Spatial-Aware LSTM modules to learn vibration-invariant geometric representations rather than overfitting to unstable pixel-wise locations.

3) Depth Preprocessing: We apply a unified preprocessing pipeline to both simulated and real depth before they are fed into the neural network. Specifically, the depth images undergo size cropping, depth value clipping, spatial resize, and Gaussian blur. This standardized alignment ensures consistent inputs for a robust transfer to physical hardware.

D. Customized Reward Design for Hollow Stairs

Standard rewards struggle on steep hollow stairs due to severe foot-trapping risks and poor visual awareness [18], [7]. To address this, we introduce three customized reward terms in addition to the standard reward set during the Stage 1 and Stage 3 training phases to safely navigate these geometric hazards.

1) 3D Waypoint-Guided Active Perception Reward (r_{pitch}): Previous waypoint-based locomotion frameworks predominantly utilize 2D waypoints to guide the robot's yaw angle for directional heading [7]. However, such 2D guidance remains insufficient for the extreme geometry of steep hollow stairs. As illustrated by the yellow dashed line in Figure 4(a), we introduce a 3D waypoint tracking mechanism that targets the center of the second upcoming stair thread p_{target}. By calculating the relative vector between the robot's base and this elevated forward waypoint, we directly supervise both the yaw and, crucially, the pitch angle (θ_{pitch}).

This 3D look-ahead formulation serves a dual purpose:

- Kinematic Optimization: Proactively pitching the trunk upwards shifts the center of mass distribution to a more stable configuration and significantly facilitates hind-leg clearance during steep ascents.
- Active Perception: This mechanism acts as an implicit active sensing strategy. By guiding the robot to look up toward the distant waypoint, the onboard egocentric camera captures the structural features of upcoming threads much earlier than a horizontal baseline. This foresight is critical for the visuospatial encoder to populate its spatial memory before the threads enter the camera's near-field blind spots.

To prevent the robot from pitching its base upwards prematurely while still traversing flat ground, we introduce a distance-based activation threshold d_{th}. The active pitch guidance is only triggered when the Euclidean distance d between the robot and the upcoming waypoint falls below this threshold. We formulate this reward using a piecewise function to penalize the deviation between the current pitch and the target pitch:

$$r_{\text{pitch}} = \begin{cases} \exp\left(-4(\theta_{\text{pitch}} - \theta_{\text{pitch}}^{\text{target}})^2\right), & \text{if } d < d_{\text{th}}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

2) Hollow Gap Kinematic Penalty (r_{hollow}): The most catastrophic failure mode on hollow stairs is kinematic locking, which occurs when a robot's foot slips into the empty space between two threads, severely trapping the leg. To strictly prohibit such behaviors, we design a hard penalty

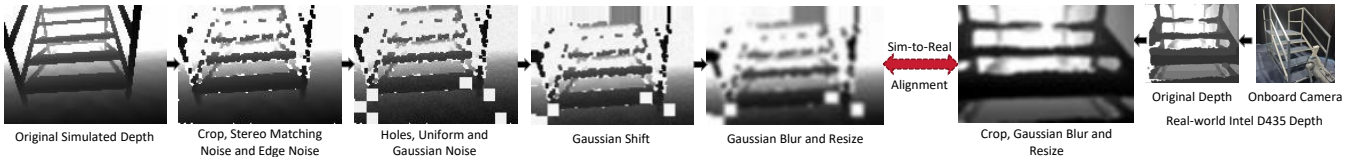
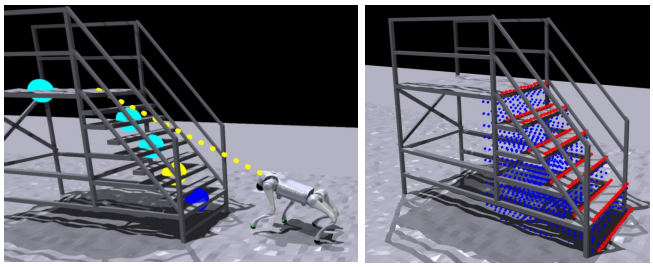


Fig. 3: Pipeline of high-fidelity sim-to-real depth noise modeling.



(a) 3D waypoint-guided active perception reward (b) Hollow stair foothold reward

Fig. 4: Designed reward for hollow stairs.

for any foot trajectory that intersects with the predefined hollow volumes as illustrated by the blue area in Figure 4(b). If the position of the i-th foot, denoted as $\mathbf{p}_{\text{foot},i}$, enters the bounding box of the hollow gap region B_{hollow}, a penalty of $-c_{\text{hollow}}$ is applied for that specific foot. The reward is formulated as:

$$r_{\text{hollow}} = \sum_{i=1}^{4} \begin{cases} -c_{\text{hollow}}, & \text{if } \mathbf{p}_{\text{foot},i} \in B_{\text{hollow}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

This penalty forces the robots to prioritize high and safe swing trajectories that completely clear the gaps.

3) Stair Edge Penalty (r_{edge}): Even if a foot lands on the thread, stepping too close to the stair edge highly increases the risk of slipping off, especially given the noisy depth perception in the real world. As illustrated in the edge condition constraints, we introduce a penalty based on the distance $d_{\text{edge},i}$ from the i-th foot's contact point to the stair edge depicted by the red dots in Figure 4(b). We define a safe margin threshold d_{safe}. The policy is penalized if the foot lands closer to the edge than this safety margin:

$$r_{\text{edge}} = \sum_{i \in \text{contact}} \begin{cases} -c_{\text{edge}}, & \text{if } d_{\text{edge},i} < d_{\text{safe}}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

This encourages the robot to consistently target the center depth of the stair threads, maximizing stable footholds and significantly boosting sim-to-real robustness.

E. Training Details

1) Terrain Curriculum Design: To facilitate stable policy learning, we implement a terrain curriculum that progressively scales the terrains' complexity.

- Progressive Geometric Scaling: The curriculum initializes on flat ground (0°) for basic locomotion learning. As performance improves, it progressively increases the step height while decreasing the thread depth and width, eventually reaching an extreme 55° incline.
- Structural Noise Injection: To enhance sim-to-real robustness, we inject random noise into vertical elevations and horizontal gap distances. This non-uniformity prevents gait memorization and forces reliance on real-time visuospatial perception.
- Update and Resampling Mechanism: Robots advance upon reaching the top waypoint. To prevent catastrophic forgetting, masters of the highest level are reassigned to randomly sampled difficulties across the curriculum.

2) Domain Randomization: To facilitate sim-to-real transfer, we apply comprehensive domain randomization. This includes perturbing static and dynamic parameters (friction, mass, motor strength), applying external pushes and action delays, and randomizing the depth camera's intrinsic FOV and mounting poses.

IV. EXPERIMENTS

A. Experimental Setup

We train our policy in the Isaac Gym simulator using a single NVIDIA RTX 4090 GPU. For real-world deployment, we utilize a Unitree Go2 quadruped robot. To perceive the environment, an Intel RealSense D435 depth camera is mounted via a custom-designed bracket, capturing depth streams at 10 Hz. The trained policy is executed entirely onboard an NVIDIA Jetson Orin NX computing module. This module outputs target joint positions at a control frequency of 50 Hz, which are subsequently converted into motor torques by a low-level PD controller with stiffness and damping gains set to $K_p = 40$ and $K_d = 1$, respectively.

B. Simulation Comparison and Ablation Study

To evaluate the effectiveness of the StairMaster framework and the necessity of our proposed components, we conduct extensive comparative experiments in simulation. We compare our method against these baselines:

- Extreme Parkour [7]: A two-stage visual parkour framework that distills a depth-based student policy from a privileged teacher.
- HIMLoco [18]: A state-of-the-art blind locomotion policy using hybrid internal model.
- Ours w/o r_{pitch}: Our method without the waypoint-based pitch tracking rewards.
- Ours w/o r_{foothold}: Our method without both the hollow gap penalty (r_{hollow}) and the thread edge penalty (r_{edge}).
- Ours w/o depth noise: Our method trained without sim-to-real depth noise modeling, which is compared in depth noise ablation experiments.

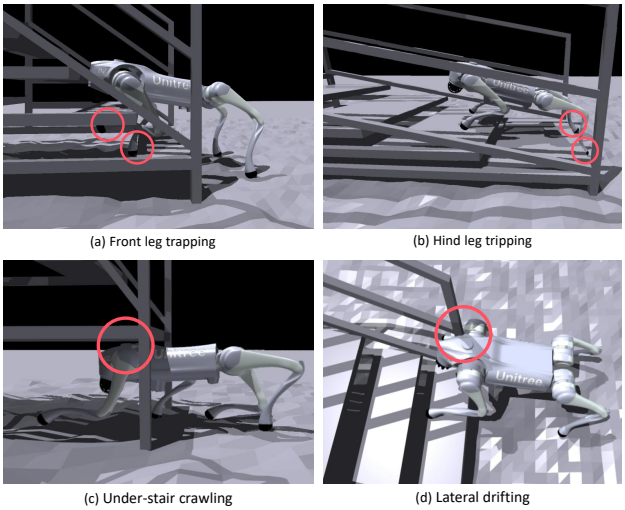


Fig. 5: Classical failure cases in simulation.

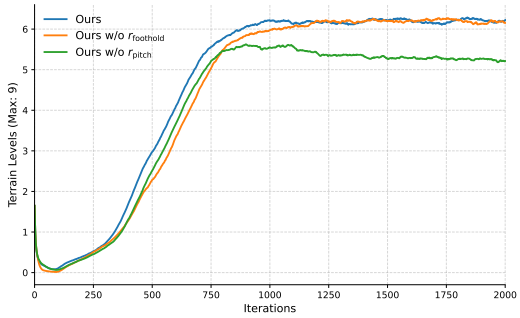


Fig. 6: Terrain level comparison during training.

For Extreme Parkour and HIMLoco, they are designed to deal with common terrains like solid stairs or uneven planes. We retrain the policy in our hollow stair terrain, but the rewards and training settings have not changed. This may cause the relatively poor performance of these two methods.

We define two primary evaluation metrics: Success Rate and Average Reached Steps. We evaluate their performance across three distinct environments: a flat-ground baseline, standardized hollow stairs, and randomized mixed stairs. The mixed environment specifically consists of consecutive threads generated with stochastic step heights and random horizontal gaps between each pair of steps, posing a significant challenge to the robot's spatial reasoning and landing precision.

- Success Rate: The average percentage of trials where the robot successfully reaches the final waypoint at the top platform on hollow stairs.
- Average Reached Steps (%): The average percentage of the steps successfully traversed per trial. A step is counted when the robot's center of mass passes the center of corresponding thread.

We report the results in Table I. In comparison experiments, our method outperforms others in all metrics.

1) Necessity of Vision in Discontinuous Terrains: As shown in Table I, the blind baseline (HIMLoco) fails completely on all hollow stairs, confirming that pure proprioception is insufficient for terrains with high geometric sparsity. As illustrated in Figure 5(c) and (d), the lack of perception often leads the policy into critical failure modes such as under-stair crawling and severe torso collisions, as the robot fails to recognize the suspended stair threads. Furthermore, these physical impacts with the stair structures frequently induce lateral drifting, where the reactionary forces from the collisions inadvertently alter the robot's base orientation and divert it from the intended heading.

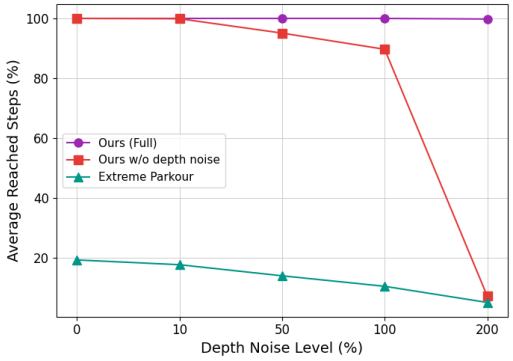


Fig. 7: Experiment results under different depth noise levels.

2) Robustness to Extreme Inclines: As shown in Table I, the visual baseline, Extreme Parkour, demonstrates a critical inability to handle the geometric discontinuity of hollow stairs. While it manages a 17.0% average reached step rate at 20°, its overall success rate remains at 0.0%. Consistent with the observations in [6], this baseline may occasionally surmount the first thread but fails to climb a second, as illustrated by the front leg trapping and hind leg tripping cases in Figure 5(a) and (b). Once threads leave the camera FOV during steep climbs, the policy cannot infer precise footholds for the sparse structures. In contrast, StairMaster maintains a 97.5% success rate even at a 55° incline, proving that our visuospatial memory and active perception effectively overcome severe occlusions and prevents the catastrophic failures observed in the baselines.

3) Impact of Proposed Rewards: To further evaluate safety, we additionally recorded the average number of collisions during stair climbing, with the detailed results presented in Table II. These internal ablations reveal that our geometric rewards are critical for both traversal success and structural safety. As shown in Figure 6, the integration of r_{pitch} significantly elevates the average terrain level by mitigating visual blind spots; without it, Table II shows a substantial surge in average collisions on steep inclines, as the robot fails to adapt its orientation and suffers from catastrophic torso-thread impacts. Similarly, while the r_{foothold}-ablated variant eventually converges, the full StairMaster framework demonstrates markedly higher sample efficiency and landing precision. The lower collision frequency of our full method in mixed-terrain scenarios confirms that explicit foothold constraints are indispensable for preventing the edge-slipping and foot-trapping failures inherent to discontinuous stairs.

TABLE I: Quantitative comparison results with other locomotion policies on hollow stairs in simulation.

Method	Success Rate (%) ↑							Average Reached Steps (%) ↑					
	Flat	Hollow Stairs						Hollow Stairs					
	0°	20°	30°	40°	50°	55°	Mixed	20°	30°	40°	50°	55°	Mixed
Ours	**100.0**	**100.0**	**100.0**	**100.0**	**98.0**	**97.5**	**86.5**	**100.0**	**100.0**	**100.0**	**99.1**	**98.4**	**93.0**
Ours w/o r_{pitch}	100.0	100.0	100.0	99.5	97.5	96.0	77.2	100.0	100.0	99.9	99.0	98.3	87.5
Ours w/o r_{foothold}	100.0	100.0	100.0	100.0	97.5	97.0	83.5	100.0	100.0	100.0	98.6	98.2	91.2
Extreme Parkour	100.0	0.0	0.0	0.0	0.0	0.0	0.0	17.0	19.4	0.0	0.0	0.0	0.5
HIMLoco	100.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	2.9	0.0	0.0	0.0	0.0

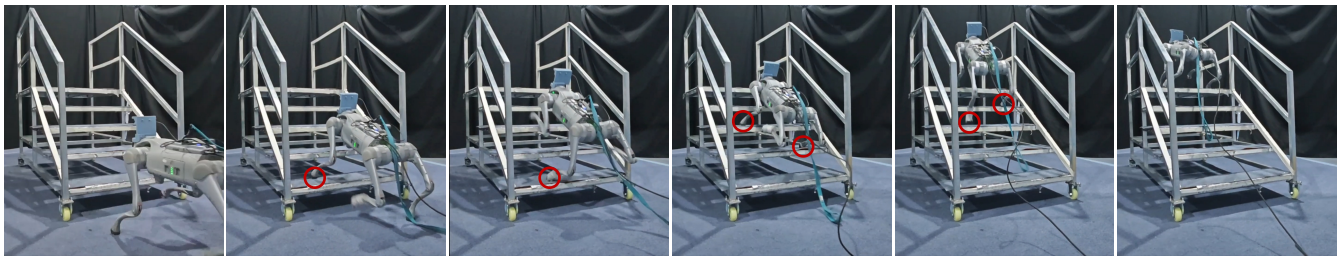


Fig. 8: Snapshots of real-world experiments on the 37° hollow stairs.

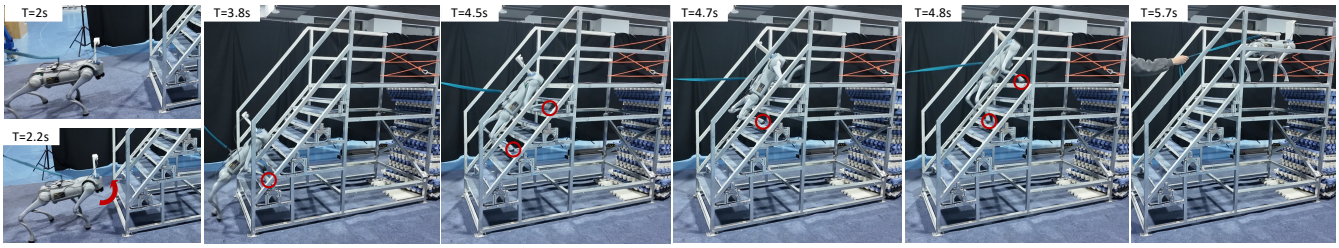


Fig. 9: Snapshots of real-world experiments on the steep 55° hollow stairs.

TABLE II: Comparison results of average collision counts in the ablation study on hollow stairs.

Method	Average Collisions ↓					
	Hollow Stairs					
	20°	30°	40°	50°	55°	Mixed
Ours	**0.03**	0.04	**0.78**	**4.62**	**5.64**	**13.25**
Ours w/o r_{pitch}	0.06	0.16	1.59	9.68	12.83	31.13
Ours w/o r_{foothold}	0.03	**0.02**	0.98	5.25	6.46	22.66

C. Robustness to Depth Noise

Real-world depth sensors often suffer from severe noise on hollow stairs due to depth sparsity. To evaluate robustness, we inject varying levels of artificial noise (0% to 200%) into the depth observations on the 30° hollow stair. Specifically, we compare three methods in this test: Ours, Ours w/o depth noise, and Extreme Parkour. The quantitative results, measured in average reached steps (%), are shown in Figure 7.

As the noise level escalates, the average reached steps of both the baseline and the ablated method drop drastically. Extreme Parkour can only traverse 14.1% of the stairs under 50% noise and fails completely under more severe noise. Similarly, our policy w/o depth noise modeling experiences a steep decline, negotiating only 7.3% of the steps at 200% noise. In contrast, our full framework demonstrates exceptional resilience, consistently completing 99.8% of the stair climb even when subjected to extreme 200% depth noise.

D. Real-World Experiments

To validate the sim-to-real transferability of the StairMaster framework, we conducted physical experiments on challenging hollow stairs with inclines of 37° and 55°. Each policy was deployed on a Unitree Go2 quadruped and tested for 10 consecutive trials per stair. We also introduce the robot's built-in MPC and built-in RL as an additional baseline for comparison. The results are shown in Table III.

TABLE III: Real-world success rates on hollow stairs.

Method	Success Rate (%) ↑	
	Hollow Stairs	
	37°	55°
Ours	**80**	**40**
Ours w/o depth noise	40	10
Extreme Parkour	0	0
HIMLoco	0	0
Built-in MPC	0	0
Built-in RL	70	0

As shown in Table III, most baselines, including HIMLoco, Extreme Parkour, and the Built-in MPC, completely fail to conquer either stair. Notably, the Built-in RL (Unitree's blind RL) is the only baseline capable of ascending the 37° stair with a 70% success rate. However, because it relies on blind proprioception, it frequently suffers from foot-trapping as it cannot perceive the gaps between threads.

For the extreme 55° stair, the Built-in RL fails entirely; its lack of height perception causes the front legs to reach directly underneath the first thread rather than stepping onto it, leading to immediate collision and failure. Similarly, the vision-based Extreme Parkour fails at both angles due to its inability to handle the sparse geometric discontinuities.

The necessity of our depth noise modeling is underscored by the Ours w/o depth noise ablation, which suffered a drastic success rate drop. Its failure to generalize to real-world depth sensor noise confirms that realistic noise injection is essential for bridging the sim-to-real gap.

Our full framework demonstrates superior robustness and precision. As illustrated by the red circles in Figure 8 and Figure 9, StairMaster achieves highly accurate foot placement. By utilizing the visuospatial memory and r_{foothold} supervision, the robot consistently lands in the center of the narrow threads, avoiding the trapping issues that plague the Built-in RL. On the challenging 55° incline shown in Figure 9, the effectiveness of r_{pitch} is clearly visible. The robot proactively raises its pitch angle before reaching the first step ($T = 2.2s$), allowing the camera to adapt to the steep environment and capture the upcoming stair structure. This look-ahead capability enables the robot to complete the 55° climb in under 4 seconds, showcasing remarkable speed and fluid motion. Ultimately, achieving an 80% success rate at 37° and a 40% success rate at 55° confirms that our proposed rewards and architecture facilitate reliable, zero-shot sim-to-real transfer on discontinuous terrains.

V. CONCLUSIONS

In this work, we presented StairMaster, a robust reinforcement learning framework that enables quadruped robots to traverse high-risk hollow stairs. Our method effectively overcomes the challenges of sensor noise and visual occlusions by combining a Cross-Attention-based visuospatial encoder with a Spatial-Aware LSTM. We introduced a three-stage training pipeline and tailored reward functions for active pitch control and precise foothold placement to ensure the robot's safety and stability on extremely sparse terrains. Real-world experiments on the Unitree Go2 robot confirm that our policy achieves remarkable robustness, successfully navigating 55° hollow stairs with zero-shot sim-to-real transfer.

Future work will focus on improving sim-to-real fidelity through motor modeling and integrating RGB images to further enhance environmental awareness. We also aim to extend our framework to handle more complex industrial constraints, pushing the boundaries of robot mobility in extreme scenarios.

REFERENCES

[1] 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.

[2] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.

[3] X. Peng, E. Coumans, T. Zhang, T. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals. arxiv 2020," *arXiv preprint arXiv:2004.00784*, 2020.

[4] J. Wu, G. Xin, C. Qi, and Y. Xue, "Learning robust and agile legged locomotion using adversarial motion priors," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4975–4982, 2023.

[5] Z. Zhuang, Z. Fu, J. Wang, C. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, "Robot parkour learning," *arXiv preprint arXiv:2309.05665*, 2023.

[6] Y. Yang, G. Shi, C. Lin, X. Meng, R. Scalise, M. G. Castro, W. Yu, T. Zhang, D. Zhao, J. Tan *et al.*, "Agile continuous jumping in discontinuous terrains," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 10 245–10 252.

[7] 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.

[8] P. Li, H. Li, Y. Ma, L. Chang, X. Yang, R. Yu, Y. Zhang, Y. Cao, Q. Zhu, and G. Sartoretti, "Kivi: Kinesthetic-visuospatial integration for dynamic and safe egocentric legged locomotion," *arXiv preprint arXiv:2509.23650*, 2025.

[9] Y. Zhang, Q. Qian, T. Hou, P. Zhai, X. Wei, K. Hu, J. Yi, and L. Zhang, "Renet: Fault-tolerant motion control for quadruped robots via redundant estimator networks under visual collapse," *IEEE Robotics and Automation Letters*, 2025.

[10] N. Rudin, J. He, J. Aurand, and M. Hutter, "Parkour in the wild: Learning a general and extensible agile locomotion policy using multi-expert distillation and rl fine-tuning," *arXiv preprint arXiv:2505.11164*, 2025.

[11] S. Zhu, Z. Zhuang, M. Zhao, K.-Y. Lee, and H. Zhao, "Hiking in the wild: A scalable perceptive parkour framework for humanoids," *arXiv preprint arXiv:2601.07718*, 2026.

[12] F. Yang, P. Frivik, D. Hoeller, C. Wang, C. Cadena, and M. Hutter, "Spatially-enhanced recurrent memory for long-range mapless navigation via end-to-end reinforcement learning," *The International Journal of Robotics Research*, p. 02783649251401926, 2025.

[13] G. Bledt, M. J. Powell, B. Katz, J. Di Carlo, P. M. Wensing, and S. Kim, "Mit cheetah 3: Design and control of a robust, dynamic quadruped robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2245–2252.

[14] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," *The International Journal of Robotics Research*, vol. 43, no. 4, pp. 572–587, 2024.

[15] 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.

[16] I. Nahrendra, B. Yu, and H. Myung, "Dreamwaq: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning," *arXiv preprint arXiv:2301.10602*, 2023.

[17] S. Zhu, R. Huang, L. Mou, and H. Zhao, "Robust robot walker: Learning agile locomotion over tiny traps," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 15 987–15 993.

[18] J. Long, Z. Wang, Q. Li, J. Gao, L. Cao, and J. Pang, "Hybrid internal model: Learning agile legged locomotion with simulated robot response," *arXiv preprint arXiv:2312.11460*, 2023.

[19] D. Vogel, R. Baines, J. Church, J. Lotzer, K. Werner, and M. Hutter, "Robust ladder climbing with a quadrupedal robot," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 7239–7244.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] C. Zhang, V. Klemm, F. Yang, and M. Hutter, "Ame-2: Agile and generalized legged locomotion via attention-based neural map encoding," *arXiv preprint arXiv:2601.08485*, 2026.

[25] R. Yang, M. Zhang, N. Hansen, H. Xu, and X. Wang, "Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers," *arXiv preprint arXiv:2107.03996*, 2021.

[26] 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.

[27] 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*, vol. 9, no. 11, pp. 9986–9993, 2024.

[28] H. Lai, J. Cao, J. Xu, H. Wu, Y. Lin, T. Kong, Y. Yu, and W. Zhang, "World model-based perception for visual legged locomotion," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 11 531–11 537.

[29] M. Dai, W. D. Compton, J. Li, L. Yang, and A. D. Ames, "Walk the planc: Physics-guided rl for agile humanoid locomotion on constrained footholds," *arXiv preprint arXiv:2601.06286*, 2026.