2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 979-8-3503-7770-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IROS58592.2024.10801925

# LEEPS: Learning End-to-End Legged Perceptive Parkour Skills on Challenging Terrains

Tangyu Qian, Hao Zhang, Zhangli Zhou, Hao Wang, Mingyu Cai, and Zhen Kan

***Abstract*— Empowering legged robots with agile maneuvers is a great challenge. While existing works have proposed diverse control-based and learning-based methods, it remains an open problem to endow robots with animal-like perception and athleticism. Towards this goal, we develop an End-to-End Legged Perceptive Parkour Skill Learning (LEEPS) framework to train quadruped robots to master parkour skills in complex environments. In particular, LEEPS incorporates a vision-based perception module equipped with multi-layered scans, supplying robots with comprehensive, precise, and adaptable information about their surroundings. Leveraging such visual data, a position-based task formulation liberates the robot from velocity constraints and directs it toward the target using innovative reward mechanisms. The resulting controller empowers an affordable quadruped robot to successfully traverse previously challenging and unprecedented obstacles. We evaluate LEEPS on various challenging tasks, which demonstrate its effectiveness, robustness, and generalizability. Supplementary and videos are available at: `https://sites.google.com/view/leeps`**

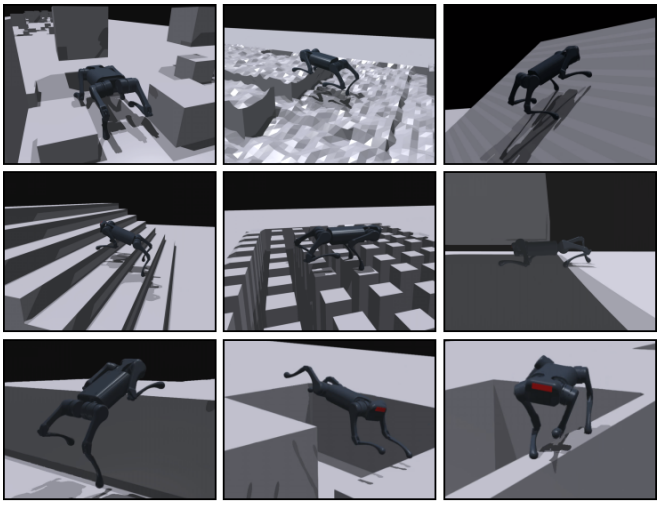

**Fig.1: Application of LEEPS.** The quadruped robot A1 traversing diverse challenging terrains and making agile parkour maneuvers.

## I. INTRODUCTION

Equipping quadruped robots with animal-like athleticism is a grand challenge in robotics. Over the decades, control-based methods have enabled quadruped robots to work in a variety of complex scenarios through model predictive control [1]–[3] and trajectory optimization [4]. Relying on predefined heuristics and model simplifications for real-time control [5], these methods struggle with highly dynamic tasks, where most assumptions and simplifications no longer apply. Moreover, the traditional hierarchical perception pipeline [6] necessitates extensive computation and is susceptible to estimation drift. Consequently, traditional methods often fall short of achieving desired parkour performance.

Reinforcement learning (RL) models uncertain dynamic systems as a Markov decision process (MDP) and learns the optimal policy from experience [7], [8]. Although RL has given rise to significant advancements in legged robots for dynamic maneuvers [9] and traversing challenging terrains [10], notable challenges persist: 1) Unlike humans who instinctively adapt decision-making to environment, the actions of quadruped robots often rely on remote control. How can we empower them with greater autonomy and meanwhile enhance athleticism? 2) Compared to animals, born with multi-modal perception and sophisticated understanding capacity, mobile robots are constrained by limited hardware resources. How can a quadruped robot gather complete and accurate environment information through a low-cost RGB-D camera?

### A. Related Work

By modeling the quadruped system as a MDP and learning through trial and error, RL manages to incorporate more complex dynamics without sacrificing real-time performance. The use of RL for continuous control was originally proposed in [11] to solve locomotion problems without human intervention, which has been proven effective in learning agile gaits. To encourage diversity, [12] presents an imitation learning framework, enabling legged robots to learn locomotion skills by imitating real-world animals and managing to synthesize controllers capable of various behaviors. However, solely imitation prevents robots from autonomously selecting skills through environment perception. [13] manages to adapt to different scenarios like changing terrains and payloads by estimating environment configurations from state history. [14] leverages an attention-based recurrent encoder that integrates proprioceptive and exteroceptive input to design a controller with superior robustness.

Utilizing neural networks for image processing facilitates the handling of complex, high-dimensional data and has opened up a myriad of applications [15], [16]. The work [17] implements a self-supervised foothold classifier based on a convolutional neural network (CNN) and results in up to 200

T. Qian, H. Zhang, Z. Zhou (corresponding author), H. Wang, and Z. Kan are with the Department of Automation, University of Science and Technology of China, Hefei, China, 230026.

M. Cai is with the Department of Mechanical Engineering, University of California, Riverside, CA, USA, 92521.

This work was supported in part by the National Natural Science Foundation of China under Grant U2013601 and 62173314.

times faster computation compared to the heuristics. Since only extracting footholds from the entire image potentially overlooks other crucial details in the frame, [14] encodes the terrain scan map as a latent and then feeds to the policy, delivering more favorable outcomes. [18] adopts depth image as an alternative and can traverse a large variety of terrains while being robust to external perturbations and estimation drifts. Despite recent progress, previous works either rely on domain randomization for training blind policies or merely focus on decoding terrain features, three-dimensional environment perception and overhanging obstacle avoidance, which is often the case in parkour scenarios, remains a challenge. Recent work [19] adopts the properties of the obstacle as privileged information for the policy, enabling the quadruped robot to crawl beneath low barriers of 0.2m. [20] proposes a 3D reconstruction model for point clouds to complete the scene from context. [21] augments this approach via a multi-resolution scheme, demonstrating the application to complex scenes with overhanging obstacles.

### B. Contribution

The common approach of learning locomotion requires tracking a constant velocity, which deprives agents of the autonomy to adapt the speed to the surroundings. [22] constructs a position-tracking reward to free the agent from velocity constraints and manages to empower the robot with more complex behaviors. We develop a position-based task formulation with novel reward mechanisms based on [22], resulting in superior performance with a more natural gait. Recent advancements in legged perception [19]–[21] either limit to structured obstacles or require expensive hardware to guarantee performance. To overcome the limitations, we design a perception module with multi-layered and multi-resolution environment scan dots. The sparse scans are encoded as latent and can be distilled to a recurrent depth backbone through teacher-student training. Consequently, the robot can acquire comprehensive environment information through a low-cost RGBD camera with limited computation. Our contributions are summarized as follows:

1) We propose a novel framework to Learn End-to-End legged perceptive Parkour Skills (LEEPS) that enables advanced parkour maneuvers through enhanced vision perception and end-to-end reinforcement learning.

2) We develop a computationally efficient perception module with multi-layered and multi-resolution environment scans that can be distilled to flexible depth latents. Based on the visual input, we design a position-based task formulation with novel reward mechanisms to train a robust controller with extreme athleticism.

3) We evaluate LEEPS across nine complex terrains. The quadruped robot successfully overcomes previously challenging and unprecedented obstacles. The superior experimental performance demonstrates the effectiveness of the proposed method.

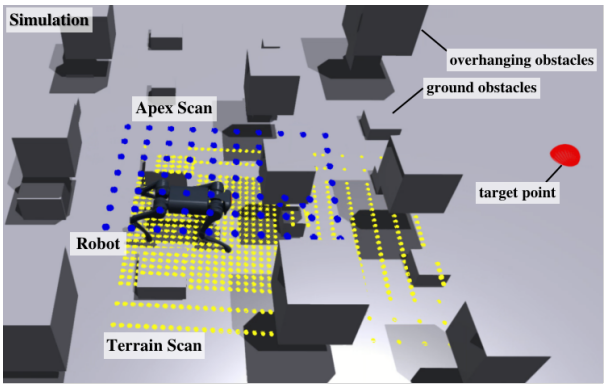


**Fig.2: Example of LEEPS.** The quadruped robot A1 navigates through rock jumbles while avoiding collisions with ground and overhanging obstacles through perception.

## II. END-TO-END PARKOUR SKILL LEARNING

The objective of our method is to enable the quadruped robot to traverse a variety of complex terrains autonomously by perceiving its surrounding environment. An overview of our framework is in Fig. 3. We first explain the perceptive parkour policy training in Sec. II-A. Then, we present our novel position-based task setting in Sec. II-B. Finally, the terrain curriculum training is introduced in Sec. II-C

### A. Perceptive Parkour Policy Training

To navigate through complex environments, robots must possess the capability of perceiving and avoiding obstacles. We train a neural network that maps onboard sensing to joint position commands. To circumvent training complexity and enhance policy performance, we employ a teacher-student architecture [23] and introduce key designs in the perception module to enable a comprehensive and precise perception of the surroundings.

**Reinforcement Learning with Privileged Information.** The observations of the teacher policy at time $t$ include full robot state $s_t$, terrain scan dots $m_t$, apex scan dots $n_t$, and environment factor $e_t$. The full robot state consists of onboard proprioception $\mathbf{x}_t$ and privileged state $\mathbf{x}_t^p$.

Terrain scan dots $\mathbf{m}_t$ is a matrix containing robo-centric terrain measurements, which indicates the vertical distance from the sampled point to the base. Different from previous works [18], [19] employing a grid with a fixed resolution, we draw inspiration from point cloud techniques [21] and implement a multi-resolution grid. The resolution is progressively higher in proximity to the robot and decreases with distance. Similar to $\mathbf{m}_t$, apex scandots $\mathbf{n}_t$ represent the height of the overhanging sampling points to the base. Our work is the first to employ multi-layered scan dots to deal with three-dimensional perception and obstacle avoidance. By encoding nearby obstacles' height as a latent, our method provides better generalization than incorporating a fixed obstacle height to the observation [19].

To mitigate computation burden, the environment factor $\mathbf{e}_t$ is processed with a privileged information encoder, and the multi-layered scan dots are processed with scan encoders

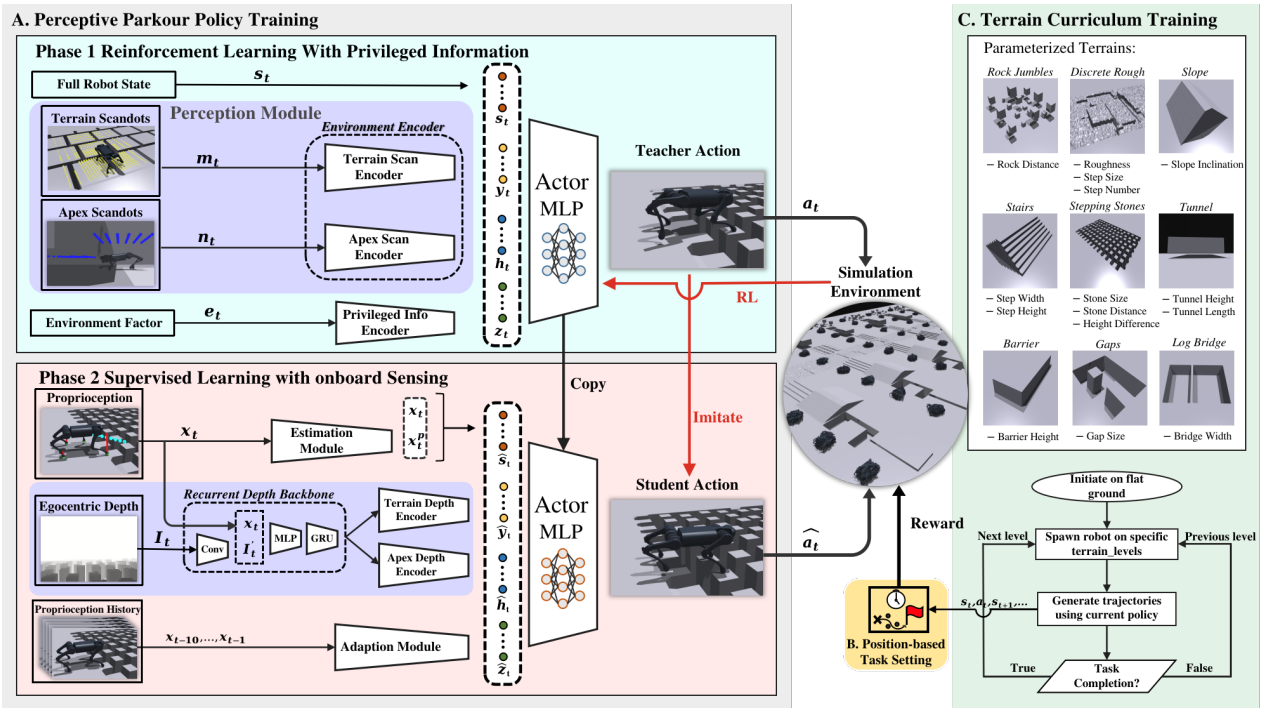


**Fig.3: Overview of the presented framework.** (**A**) Perceptive Parkour Policy Training. First, a perceptive teacher policy is trained using RL with access to privileged multi-layered scans. Next, a proprioceptive student policy imitates the teacher through a stream of onboard sensory input. (**B**) Position-based task setting. An end-to-end task setting provides the training process with innovative reward terms. (**C**) Terrain Curriculum Training. An adaptive terrain curriculum generates terrains of varying difficulty during training.

to obtain environment latent $\mathbf{z}_t$, terrain scan latent $\mathbf{y}_t$, and apex scan latent $\mathbf{h}_t$. Subsequently, these latents along with the full robot state $\mathbf{s}_t$ are concatenated and fed to the actor as input by

$$\mathbf{a}_t = \text{MLP}(\mathbf{s}_t, \mathbf{y}_t, \mathbf{h}_t, \mathbf{z}_t). \tag{1}$$

The actor network outputs a 12-dimensional vector $\mathbf{a}_t$, representing residual joint angles relative to the default pose $\boldsymbol{\theta}_{default}$. A proportional-derivative (PD) controller is adopted to generate control torque for each motor at a higher frequency.

The teacher policy network is trained via the Proximal Policy Optimization (PPO) algorithm [24]. The algorithm maximizes the following expected return of the policy $\pi$:

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right], \tag{2}$$

where $r_t$ is the reward to be defined in Sec. II-B, $\gamma$ is the discount factor, and $\pi$ represents the parkour policy.

**Supervised Learning with Onboard Sensing.** We use supervised learning to distill the teacher policy into the student policy that only has access to onboard sensing: proprioception $\mathbf{x}_t$ and egocentric depth $\mathbf{I}_t$. We train with DAgger [25] to minimize the mean squared error between the predicted and the ground truth actions $||\hat{\mathbf{a}}_t - \mathbf{a}_t||^2$. We use regularized online adaption [26] to train the estimation module and the adaption module.

### *B. Position-based Task Formulation*

We introduce a position-based task formulation with novel reward terms to empower the quadruped robot with advanced skills. Our work develops a task reward $R_t^{\text{task}}$ to free the robot from velocity constraints, enabling simultaneous mastery of navigation and locomotion skills. The expanded search space, which slows training, is counteracted by exploration reward $R_t^{\text{explo}}$ and parkour reward $R_t^{\text{parkour}}$. An additional gait-shaping reward $R_t^{\text{gait}}$ facilitates the learning of natural gaits. The normalization reward $R_t^{\text{norm}}$ is used for sim-to-real transfer.

**Task reward.** The task reward $R_t^{\text{task}}$ guides the robot towards the target within a specified time by rewarding the minimization of the distance between the robot and the target point [22]. It is defined as:

$$R_t^{\text{task}} = \begin{cases} \frac{1}{\mathbf{T}_{\text{episode}} - \mathbf{T}_{\text{task}}} \cdot \frac{1}{1+\left\|\mathbf{x}_b - \mathbf{x}_b^*\right\|}, & \text{if } t > \mathbf{T}_{\text{task}} \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

which ensures that the reward is applied only at the end of each episode.

**Exploration reward.** We introduce a reward for base velocity towards the goal, which is distinct from the typical velocity tracking reward as it does not restrict the direction and magnitude of speed. To prevent the policy from inaction and early termination, a stalling penalty and a termination penalty are designed, promoting the policy for continuous exploration.

**Gait-shaping reward.** Several gait-shaping rewards are designed to aid the robot in learning a natural gait. The

**Table I: Comparison of parkour tasks**. The numbers in each column represent barrier height, gap size, tunnel height, log bridge width, and the size and distance of stepping stones, respectively. Notably, our work not only breaks the best metrics with the same robot but also is capable of traversing terrains that previous works failed to achieve. The numbers of different methods are taken from the original papers.

| Method(Robot Platform) | Barrier | Gap | Tunnel | Log Bridge | Stepping Stones | Rock Jumbles |
|---|---|---|---|---|---|---|
| Rapid Motor Adaption(Unitree A1) | 8 | × | × | × | × | × |
| Advanced Skills Learning(ANYmal) | 85 | 120 | × | × | × | × |
| Vision Locomotion(Unitree A1) | 17 | 26 | × | × | 30×15 | × |
| Robot Parkour Learning(Unitree A1) | 40 | 60 | 20 | × | × | × |
| Extreme Legged Parkour(Unitree A1) | 50 | 80 | × | × | × | × |
| ANYmal Parkour(ANYmal) | 100 | 120 | 45 | × | × | × |
| LEEPS(Unitree A1) | **75** | **100** | **17** | **20** | **20 × 20** | ✓ |

contact forces reward and the slip reward encourage the robot to distribute the load with its hind legs. The swing reward and check contact reward access the phase of each foot, penalizing excessively long contact and too short swing action. The stumble reward penalizes horizontal foot forces, thereby encouraging the robot to lift its legs.

**Parkour reward.** Advancing [27], our approach penalizes foot contacts on edges where the height difference between adjacent coordinates exceeds a threshold.

**Normalization reward.** To aid sim-to-real transfer, we incorporate normalization terms into the overall reward, which impose constraints on the robot's motion. To prevent aggressive actions, the action rate is penalized. To ensure the robot operates within a stable configuration, we impose penalties for configurations that deviate significantly from the default. We also take into account the energy consumption by accessing the total torques.

**Episode process.** In the beginning, the robot is reset randomly to prevent the policy from overfitting. Since the task reward only takes effect at the end of each episode, the episode length in this work is set to $\mathbf{T}_{\text{episode}} = 8s$ with the task reward provided after $\mathbf{T}_{\text{task}} = 6s$ for a duration of 2s. Finally, the progress of robot is evaluated to update the terrain curriculum. To enhance robustness, we sample environment randomization on the robot mass, motor strength as well as terrain friction and model system latency during training.

### C. Terrains Curriculum Training

Besides common terrains like discrete rough, slope, and stairs, we create six additional challenging terrains, i.e., rock jumbles, stepping stones, tunnel, barrier, gap and log bridge. Rock jumbles are formed of randomly placed obstacles, demanding powerful sensing for traversal. Stepping stones and log bridge require precise foot placement and strong balance. Tunnel accesses the robot's capability to perceive overhanging obstacles and move in restricted spaces. Barrier and gap are designed to test extreme athleticism. Similar to [28], different sets of terrain are generated with varying difficulty. The training follows a game-inspired curriculum where robots are initialized at the easiest level and are promoted to harder ones if they traverse more than 60% of the length. Conversely, they are demoted to simpler terrain if they fail.

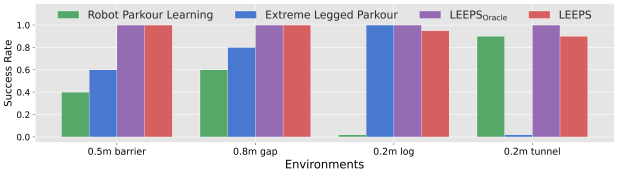


**Fig.4:** Succeess rate of different methods in four tasks.

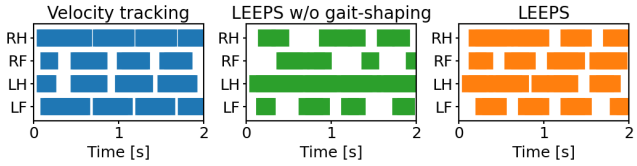


**Fig.5:** Emerging gait pattern in three settings.

## III. Simulation Experiments

### A. Experimental Setup and Baselines

A Unitree A1 robot is employed in the simulation, which is equipped with a Realsense D435i camera to capture depth images. Nvidia IsaacGym [29] is used to simulate the training environment due to its capability of massively parallel training on GPU.

Our method is extensively tested and benchmarked against prior methods, emphasizing superior performance across diverse challenging tasks. The baselines for comparison include: 1) **Rapid Motor Adaption** [13], which adapts in real-time to unseen scenarios, 2) **Advanced Skills Learning** [22], which constructs a position-tracking reward, 3) **Vision Locomotion** [18], which presents the first vision-based locomotion system capable of traversing obstacles, 4) **Robot Parkour Learning** [19], which empowers robots to autonomously select parkour skills, 5) **Extreme Legged Parkour** [27], which outputs precise control behaviors end-to-end, 6) **ANYmal Parkour** [21], which utilizes a hierarchical learning formulation.

### B. Main Results

*1) Terrain Difficulty:* The difficulty of terrains is parameterized to reflect the performance of various controllers. As indicated in Table I, LEEPS not only achieves state-of-the-art performance across terrains that are accessible

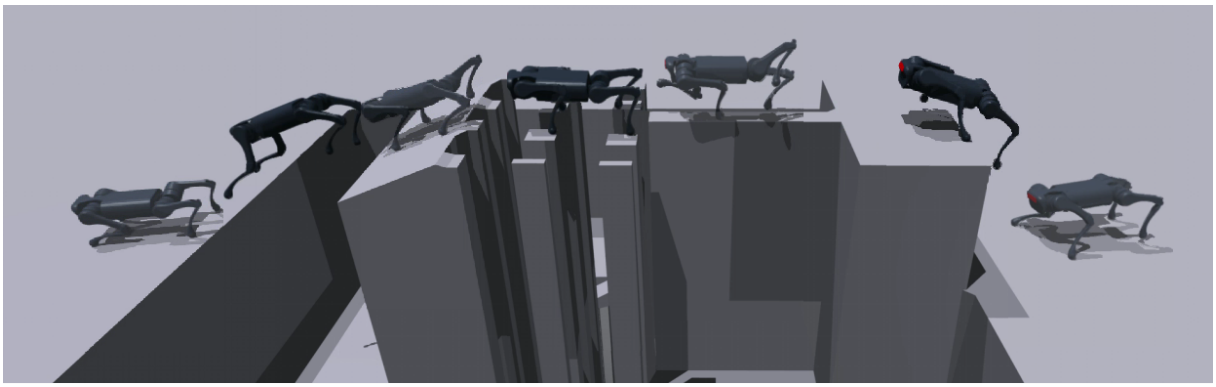

**Fig.6:** Snapshots of LEEPS traversing across an unstructured terrain, consisting of 0.7m high step, 0.2m wide log bridge, 0.2m wide, 0.2m spacing stepping stones, 0.4 rad slope and 1.0m wide gap. The traversal demonstrates LEEPS's generalizability to various obstacles.

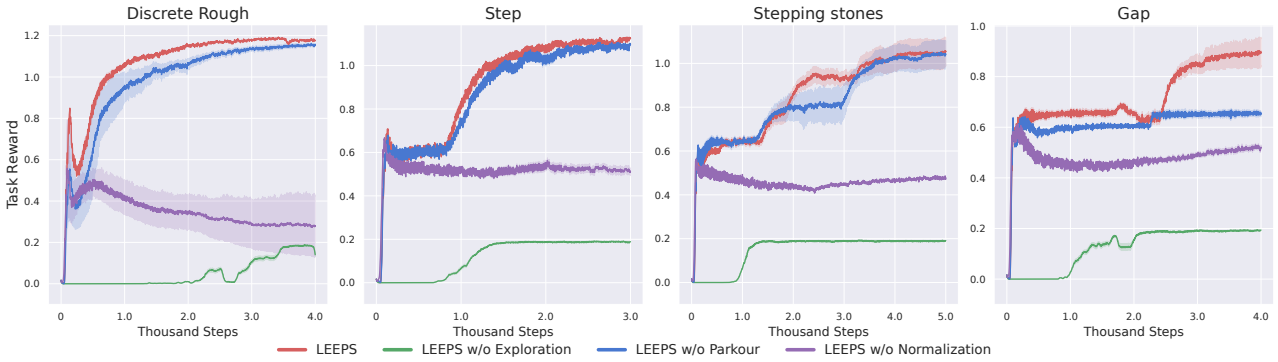


**Fig.7:** The plots of task reward in ablation studies on four terrains.

to most existing approaches but also successfully traverses terrains that are impassable by prior work. In the context of recent advancements in robot parkour, LEEPS demonstrates superior performance on the same robot platform compared to Zhuang et.al [19] and Cheng et.al [27]. Despite inferior metrics versus Holler et. al [21], ANYmal used in these works is approximately double the size of unitree A1. The enhanced performance can be credited to the position-based task formulation which frees the quadruped robot from velocity constraints, thus uncovering a more extensive set of viable solutions and augmenting the robot's athleticism.

*2) Success Rate:* For a more detailed analysis, we benchmark the success rate of four methods over different terrains. The results are shown in Fig. 4. $\text{LEEPS}_{\text{Oracle}}$ denotes the teacher policy utilizing privileged observation. $\text{LEEPS}_{\text{Oracle}}$ achieves a 100% success rate across all terrains, significantly outperforming the other works. While some approaches were entirely ineffective on log and tunnel terrains, LEEPS and $\text{LEEPS}_{\text{Oracle}}$ still demonstrate robust performance. The reason for the lower success rate of LEEPS compared to $\text{LEEPS}_{\text{Oracle}}$ lies in the limitations of visual input, since the camera's field of view is confined to obstacles ahead and fails to capture obstructions above or beneath. As demonstrated in Sec. III-B.1, the superior performance on barrier and gap terrains can be attributed to the position-based task formulation. LEEPS's robust performance on log and tunnel is owed to the key designs in the perception module: the multi-resolution terrain scan that enhances the ground obstacle perception, and the apex scan that endows the robot with the capability to perceive unstructured overhanging obstacles.

### C. Ablation Study

Ablation studies are conducted in this section to show the role of different modules in LEEPS and evaluate the importance of each reward component. Specifically, we compare the original LEEPS with four variants: 1) LEEPS w/o task reward, 2) LEEPS w/o exploration reward, 3) LEEPS w/o gait-shaping reward, 4) LEEPS w/o parkour reward.

*1) Task Completion:* The task reward represents the distance between the robot and the target point. Fig. 7 shows the learning curves of task reward for several methods on different terrains. The following conclusions can be drawn from the ablation plots: **Exploration reward is essential in the position-based task formulation**. Exploration reward mitigates the challenge of reward sparsity in the position-based formulation. In the absence of exploration terms, the task reward completely fails to improve, indicating that the robot is unable to reach the target. **Parkour reward helps to accelerate the convergence rate**. The parkour reward penalizes foot placements near the edges of the terrain, thus encouraging the quadruped robot to execute safer actions, which leads to better exploration and a faster convergence rate. **Normalization reward improves policy performance**. The normalization reward imposes restrictions on the robot's motion to avert excessively aggressive actions,

thus safeguarding the robot. The plotted curves demonstrate a consistent gap between LEEPS and LEEPS w/o exploration reward, indicating the crucial role of normalization reward.

*2) Emergent Gaits:* Previous works focus on velocity tracking, which often results in a trotting gait, as the symmetry facilitates the maintenance of a constant speed. Yet, a fixed-speed symmetrical gait is uncommon among legged animals, as it requires additional effort and can not apply to any terrain. Moreover, robots tend to take smaller strides which limits the performance on terrains with significant height changes. Fig. 5 compares the gaits of velocity tracking, LEEPS w/o gait-shaping and LEEPS. The quadruped learns a more adaptive gait without the constraints of constant velocity. As illustrated in [22], position tracking adopts a three-phased asymmetric gait akin to LEEPS w/o gait-shaping. Nevertheless, a skewed walking direction and asymmetric gait appear unnatural and may damage the hardware. By integrating position-based tracking rewards with gait-shaping rewards, both natural and dynamic gait can be acquired.

### D. Generalizability Evaluation

Prior methods typically evaluate the robot's performance on single terrain types. However, a powerful controller should not only excel in a single task but also adapt to multiple obstacles. Consequently, we synthesize a new terrain by combining six types of terrain in our work. Fig. 6 showcases snapshots of LEEPS traversing such an unstructured terrain, affirming LEEPS's outstanding generalizability.

## IV. CONCLUSIONS

In this work, we present LEEPS, an end-to-end legged perceptive parkour skills learning framework that enables quadrupeds to traverse challenging and unprecedented terrains. A three-dimensional, vision-based perception module enables the robot to acquire comprehensive, precise, and adaptable environment information in real-time on limited hardware. Based on the visual input, a position-based task formulation with innovative reward mechanisms uncovers a more extensive set of viable solutions and augments the robot's athleticism. Future work will consider the deployment of LEEPS on real hardware and tacking more challenging tasks.

## REFERENCES


[1] 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 *IEEE Int Conf Intell Rob Syst*. IEEE, 2018, pp. 2245–2252.

[2] P. Fankhauser and M. Hutter, "Anymal: a unique quadruped robot conquering harsh environments," *Research Features*, no. 126, pp. 54–57, 2018.

[3] Z. Zhou, Z. Chen, M. Cai, Z. Li, Z. Kan, and C.-Y. Su, "Vision-based reactive temporal logic motion planning for quadruped robots in unstructured dynamic environments," *IEEE Trans Ind Electron*, 2023.

[4] 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 Robot. Autom.*, vol. 3, no. 3, pp. 1560–1567, 2018.

[5] J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim, "Dynamic locomotion in the mit cheetah 3 through convex model-predictive control," in *IEEE Int Conf Intell Rob Syst*. IEEE, 2018, pp. 1–9.

[6] T. Qian, Z. Zhou, S. Wang, Z. Li, C.-Y. Su, and Z. Kan, "Vision-based reactive planning and control of quadruped robots in unstructured dynamic environments," in *IEEE Int. Conf. Adv. Robot. Mechatronics, ICARM*. IEEE, 2023, pp. 745–750.

[7] M. Cai, S. Xiao, Z. Li, and Z. Kan, "Optimal probabilistic motion planning with potential infeasible ltl constraints," *IEEE Trans Autom Control*, vol. 68, no. 1, pp. 301–316, 2021.

[8] H. Zhang, H. Wang, and Z. Kan, "Exploiting transformer in sparse reward reinforcement learning for interpretable temporal logic motion planning," *IEEE Robot. Autom.*, 2023.

[9] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," *arXiv preprint arXiv:2205.02824*, 2022.

[10] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Sci. Robotics*, vol. 4, no. 26, p. eaau5872, 2019.

[11] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," in *Int. Conf. Mach. Learn., ICML*. PMLR, 2016, pp. 1329–1338.

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

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

[14] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Sci. Robotics*, vol. 7, no. 62, p. eabk2822, 2022.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit*, 2016, pp. 770–778.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit*, 2016, pp. 779–788.

[17] O. A. V. Magana, V. Barasuol, M. Camurri, L. Franceschi, M. Focchi, M. Pontil, D. G. Caldwell, and C. Semini, "Fast and continuous foothold adaptation for dynamic locomotion through cnns," *IEEE Robot. Autom.*, vol. 4, no. 2, pp. 2140–2147, 2019.

[18] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, "Legged locomotion in challenging terrains using egocentric vision," in *Proc. Mach. Learn. Res.* PMLR, 2023, pp. 403–415.

[19] Z. Zhuang, Z. Fu, J. Wang, C. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, "Robot parkour learning," in *Proc. Mach. Learn. Res.*, 2023.

[20] D. Hoeller, N. Rudin, C. Choy, A. Anandkumar, and M. Hutter, "Neural scene representation for locomotion on structured terrain," *IEEE Robot. Autom.*, vol. 7, no. 4, pp. 8667–8674, 2022.

[21] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, "Anymal parkour: Learning agile navigation for quadrupedal robots," *arXiv preprint arXiv:2306.14874*, 2023.

[22] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter, "Advanced skills by learning locomotion and local navigation end-to-end," in *IEEE Int Conf Intell Rob Syst*. IEEE, 2022, pp. 2497–2503.

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

[24] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[25] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, "Learning by cheating," in *Proc. Mach. Learn. Res.* PMLR, 2020, pp. 66–75.

[26] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: learning a unified policy for manipulation and locomotion," in *Proc. Mach. Learn. Res.* PMLR, 2023, pp. 138–149.

[27] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, "Extreme parkour with legged robots," *arXiv preprint arXiv:2309.14341*, 2023.

[28] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Proc. Mach. Learn. Res.* PMLR, 2022, pp. 91–100.

[29] V. Makoviychuk, 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," *arXiv preprint arXiv:2108.10470*, 2021.