Height Map informed Distillation of Visual Latents for Quadruped Locomotion

Hang Liu
University of Michigan
Ann Arbor, MI
hangliu@umich.edu

Jingyu Liu
University of Michigan
Ann Arbor, MI
juliusl@umich.edu

1. Introduction

Our work addresses the fundamental gap between biological locomotion and current robotic control paradigms. In complex settings such as stair descent or parkour, humans rely on vision to anticipate obstacles, plan foot placement, and adjust gait before contact, whereas most legged robots operate as blind systems driven purely by proprioception—adequate on flat terrain but incapable of anticipatory planning, since gaps or steps are only detected upon contact, often causing unrecoverable instability. To close this gap, we integrate ego-centric vision into a reinforcement learning framework using a teacher–student training strategy that distills high-dimensional visual input into a compact height-map representation, enabling the control policy to perceive terrain geometry and achieve precise traversal that remains out of reach for proprioception-only baselines.

2. Related Work

Proprioception-only RL has reached significant maturity: sota history-based encoders enable robust motion tracking [3] and unstructured terrain traversal [1], yet these blind policies remain inherently reactive, treating terrain irregularities as disturbances rather than planning targets and thus lacking the precision needed for discrete obstacles like stairs. To transcend these limits, recent work [5] integrates visual feedback for anticipatory planning. use an ego-vision world model to predict latent environmental states for contact planning, while [4] fuse distinct policy outputs to handle sensor uncertainties. In simulation, ground-truth terrain is easily available as dense height maps, and incorporating **attention mechanisms** [2] further improves efficiency by focusing on critical footholds; however, obtaining such noise-free, dense maps in the real world is often infeasible due to occlusions and sensor limitations.

3. METHODS

Overview. We adopt a one-stage **Teacher-Student** training framework, as illustrated in Figure 1. First, we train a privileged teacher policy encoder E_T that has access to the ground-truth height map H. The teacher encodes the terrain information into a compact latent representation z_T. At same time, we train a student encoder E_S that relies solely on noisy ego-centric depth images I and proprioception to predict a latent representation z_S that aligns with the teacher's geometric latent z_T. The final policy (actor) consumes the fused features $(z_S/z_T, o_h^{prop})$ to output joint-space targets for the PD controller. Every 20 iterations, we apply Dagger that the actor network uses student's feature z_S instead of the teacher's feature for better generalization. Training includes domain randomization over terrains, illumination, and camera parameters to ensure sim-to-real robustness [5, 4].

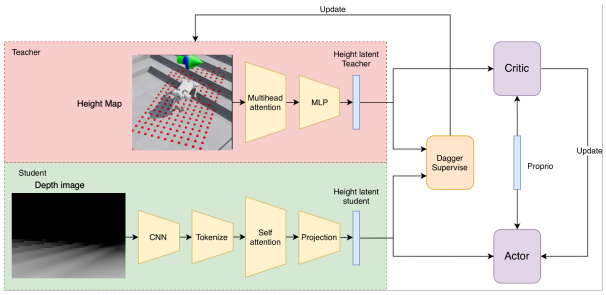


Figure 1. **Overview of the Proposed Teacher-Student RL Framework.** The training process consists of two stages using a privileged **Teacher** (top, red) and a deployable **Student** (bottom, green). The Teacher encodes ground-truth height maps into a latent representation using multi-head attention. The Student, designed for deployment, processes noisy depth images through a CNN and a self-attention mechanism to predict a matching latent representation. We align the student's latent features to the teacher's via **DAgger supervision**, allowing the policy to infer terrain geometry from onboard vision without explicit height map reconstruction.

Teacher–Student Distillation. The whole reinforcement learning process is trained via PPO. The student optimizes a joint loss function consisting of a latent distillation alignment loss (matching visual features):

$$\mathcal{L}_{latent} = \mathbb{E}_{\mathcal{D}} \left[\|z_S - z_T\|_2^2 \right], \tag{1}$$

where $z_T = E_{teacher}(H)$ and $z_S = E_\phi(I)$. This explicitly

forces the student to learn a representation containing the same geometric reasoning capabilities as the teacher.

Network Architecture. To handle the domain gap between simulation and reality, we design lightweight yet effective encoders (Figure 1).

- **Teacher Encoder:** The teacher processes the ground-truth height scan using a Multi-Head Attention module followed by an MLP, extracting a geometric latent vector z_T that encodes critical footholds.
- Student Encoder: The student processes the depth image I using a multi-layer CNN to extract local features. These features are tokenized and passed through a Self-Attention mechanism to capture global spatial dependencies. A final projection layer maps the output to the shared latent space z_S.

This design avoids the computational overhead of explicit map reconstruction while preserving the teacher's terrain reasoning capabilities through latent distillation.

4. EXPERIMENTS

Phase 1: Baseline and Sim-to-Real Framework. First, we trained a **Proprioception-only** baseline as a control group. To ensure actionable Sim-to-Real transfer, we built a unified framework (Figure 2) before introducing vision. Key features include: 1) a unified LCM interface providing a standardized message bus and identical protocols for Sim-to-Sim and Sim-to-Real; 2) observation standardization, where raw sensor data is normalized to align input distributions; and 3) a control hierarchy where the policy outputs position commands converted to torque by low-level controllers. We validated this framework via stable Sim-to-Real deployment with the baseline, confirming pipeline reliability.

Phase 2: Hardware Integration. For the vision-based policy, we modified the robot hardware. We 3D-printed custom mounts for the NUC unit and depth camera. As shown in Figure 3, the camera mount is angled $45°$ downwards. This design ensures the camera captures terrain immediately in front of the feet, critical for step planning.

Phase 3: Comparative Analysis. Figure 4 (top) shows the *Height Latent Estimation Loss*. Periodic spikes every 20 iterations correspond to **DAgger** switching control from teacher to student for data collection. Despite fluctuations, the student converged, mimicking the teacher's representation. Figure 4 (bottom) shows the *PPO Surrogate Loss*, rising during initial exploration then decreasing, indicating stable learning. We evaluated the policy against two baselines: (1) Proprioception-only and (2) End-to-End Vision (MLP without distillation). **Total Reward:** In Figure 5 (left), our policy achieves the highest reward, slightly outperforming End-to-End; the Proprioception baseline is

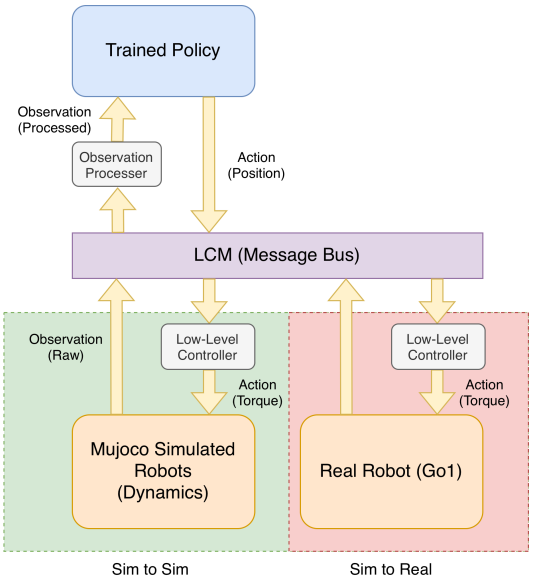
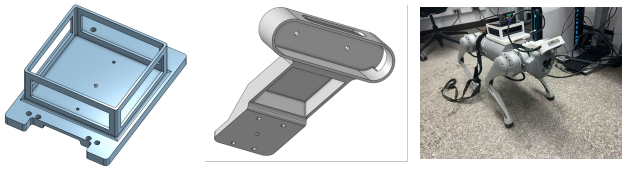


Figure 2. **Unified Sim-to-Real Framework.** Validated using the proprioception baseline to ensure high sim-real alignment.



(a) NUC Mount (b) Cam Mount (c) Assembly

Figure 3. **Hardware Setup.** Custom frames fix the camera at $45°$ for optimal terrain viewing.

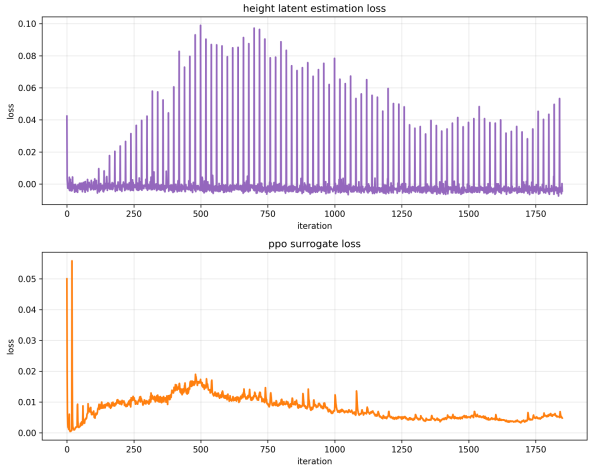
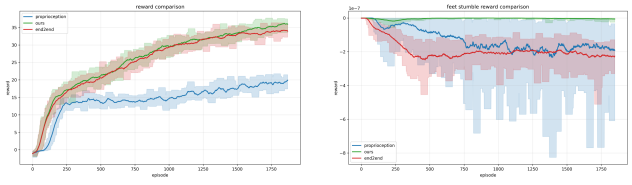


Figure 4. **Training Dynamics.** Top: Latent estimation loss showing DAgger spikes. Bottom: PPO surrogate loss showing convergence.

$\sim 50\%$ lower. **Anticipatory Capability:** The advantage is evident in the *Feet Stumble Reward* (Figure 5 right). This metric penalizes collision with step edges. Our policy maintains near-zero penalty, showing ability to "see" obstacles

and lift feet proactively. In contrast, both baselines suffer lower rewards, proving the efficacy of our distilled reasoning.



(a) Total Reward (b) Stumble Reward

Figure 5. **Results.** (a) Our method achieves highest performance. (b) Our method avoids stumbling, proving effective visual anticipation.

Limitation. Due to time constraints, full Sim-to-Real deployment of the vision policy is a work in progress, though simulation results and framework validation provide strong evidence for feasibility.

5. Conclusion

This work presents a vision-based reinforcement learning framework enabling quadrupedal robots to perform anticipatory planning on discrete terrain. By employing a teacher-student strategy with attention-based latent distillation, our method effectively bridges the gap between noisy onboard depth observations and precise geometric reasoning. Experimental results demonstrate that our policy significantly outperforms both proprioception-only and end-to-end vision baselines, particularly in minimizing foot stumbling on stairs. Furthermore, we established a robust hardware-software ecosystem that successfully validated the Sim-to-Real transfer of the baseline policy. While full real-world deployment of the vision policy remains a future objective, the presented framework lays a solid foundation. Future work will focus on closing the Sim-to-Real loop for visual navigation and evaluating generalization in unstructured outdoor environments.

References

[1] Yi Cheng, Hang Liu, Guoping Pan, Houde Liu, and Linqi Ye. Quadruped robot traversing 3d complex environments with limited perception. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9074–9081. IEEE, 2024.

[2] Junzhe He, Chong Zhang, Fabian Jenelten, Ruben Grandia, Moritz Bächer, and Marco Hutter. Attention-based map encoding for learning generalized legged locomotion. *Science Robotics*, 10(105):eadv3604, 2025.

[3] Qiayuan Liao, Takara E Truong, Xiaoyu Huang, Guy Tevet, Koushil Sreenath, and C Karen Liu. Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. *arXiv preprint arXiv:2508.08241*, 2025.

[4] Hang Liu, Yi Cheng, Rankun Li, Xiaowen Hu, Linqi Ye, and Houde Liu. Mbc: Multi-brain collaborative control for quadruped robots. *arXiv preprint arXiv:2409.16460*, 2024.

[5] Hang Liu, Yuman Gao, Sangli Teng, Yufeng Chi, Yakun Sophia Shao, Zhongyu Li, Maani Ghaffari, and Koushil Sreenath. Ego-vision world model for humanoid contact planning. *arXiv preprint arXiv:2510.11682*, 2025.