TAGA: Terrain-aware Active Gaze Learning for Generalizable Agile Humanoid Locomotion

LI Peizhuo[1,*], **Hongyi LI**[2,*], **Mingfeng FAN**[1,*], **Fangzhou XU**[1], **Shuhao LIAO**[1], **Yuxuan MA**[1], **Zicheng ZENG**[3], **Ze WANG**[2], **Yongbin JIN**[2,†], **Yuhong CAO**[1,†], **Hongtao WANG**[2], **Guillaume SARTORETTI**[1]
[1]MarmotLab, National University of Singapore [2]Center of X-Mechanics, Zhejiang University
[3]South China University of Technology [*]Equal contribution [†]Corresponding authors
Project Page: https://marmotlab.github.io/taga-humanoid/

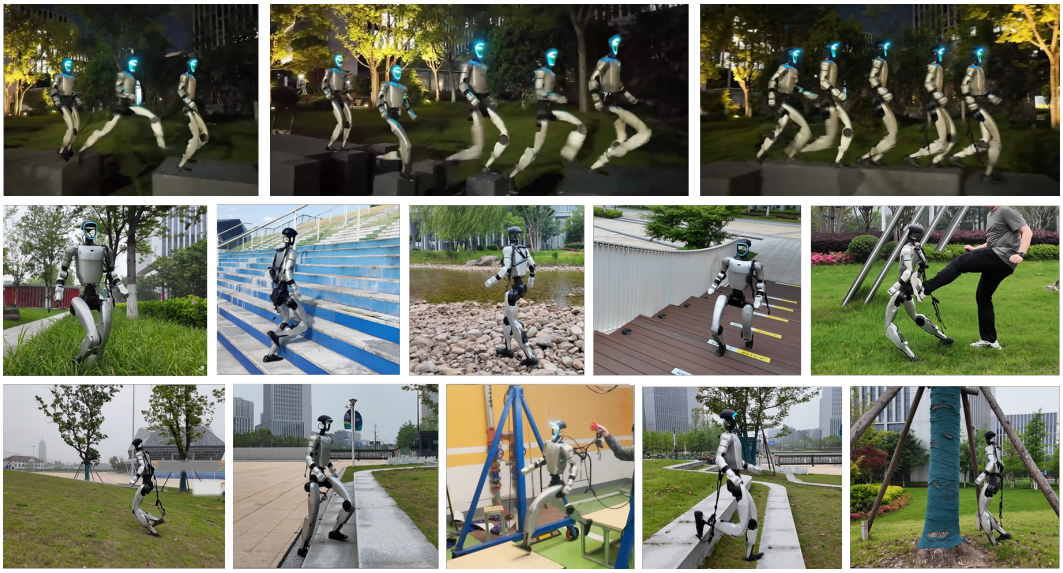

Figure 1: TAGA enables agile and robust humanoid locomotion across diverse challenging terrains. Deployed on a Unitree G1 with onboard Jetson Orin inference, the robot traverses up to $1.2\,\mathrm{m}$ gaps, narrow beams, sparse stepping stones, stairs, and outdoor terrain. TAGA uses egocentric vision and proprioception to predict task-relevant regions in the height scan, selectively routing these local terrain observations to the downstream locomotion policy, while remaining robust to severe perceptual disturbances and environmental interference.

Abstract: Agile humanoid locomotion across diverse challenging terrain demands both wide perceptual coverage and precise local geometry understanding. Motivated by the way humans selectively look at relevant terrain during locomotion, we introduce TAGA, a Terrain-aware Active Gaze learning framework for Attention-based humanoid control. By fusing vision, proprioception, and motion commands, our framework guides the model to learn anticipatory cues and actively attend to specific areas of the height scan, selectively using these informative regions for the downstream network. This adaptively increases the information density of observations under tight onboard computational constraints, thus enabling fine-grained perceptive locomotion over larger-scale terrains. We find that such gaze behaviors can naturally emerge through reinforcement learning alone, without requiring additional supervision or explicit guidance, significantly improve training efficiency. As a result, the trained policy demonstrates robust and generalizable locomotion in simulation and on hardware, including reliable terrain-aware foothold selection, elevated-platform traversal, competitive sparse-foothold traversal, and the largest reported real-world gap traversal distance of $1.2\,\mathrm{m}$ among perceptive humanoid locomotion systems, while maintaining stability under severe perceptual disturbances and environmental interference.

Keywords: Humanoid Locomotion, Gaze Mechanism, Multimodal Perception

1 Introduction

Humanoids have shown significant advantages over wheeled robots in crossing obstacles, traversing discontinuous terrain, and navigating complex spatial structures. While recent motion-tracking methods enable dynamic whole-body behaviors such as dancing, backflips, and human-motion imitation [1, 2, 3, 4, 5, 6, 7], tracking predefined motions is fundamentally different from locomotion in complex terrain, from cluttered indoor scenes to unstructured outdoor landscapes. The former resembles reproducing a memorized trajectory, whereas the latter requires the robot to actively perceive its surroundings, reason about terrain traversability, and adapt its motion strategy in real time [8]. This makes perception a central bottleneck: the robot must obtain look-ahead awareness of upcoming terrain while retaining precise local geometry for reliable foothold placement.

Existing perceptive locomotion methods can generally be divided into two categories: mapping-based methods and vision-based methods. Mapping-based approaches use point clouds or reconstructed height scans as compact terrain representations for locomotion [8, 9, 10, 11]. While effective, these methods often incur increasing computational cost as the perception range expands. In contrast, vision-based methods directly map raw depth images to actions, reducing reliance on explicit terrain reconstruction [12, 13, 14, 15, 16]. However, forward-facing depth images often miss the terrain near or beneath the robot's feet, and recurrent memory struggles to preserve fine-grained geometry over long horizons [17].

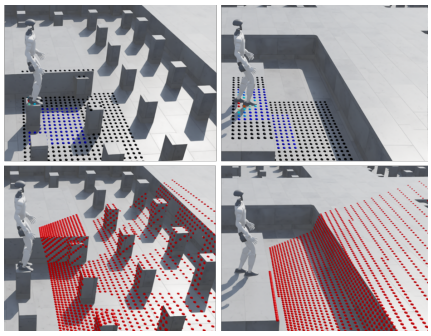

Figure 2: Comparison between local height scan and depth image perception.

These two perceptual paradigms actually provide complementary information. As shown in Fig. 2, vision offers look-ahead awareness of distant terrain, while a height scan provides accurate local geometry for foot placement and contact-rich motion control. However, many existing approaches either rely on a single source or couple the two only loosely. As a result, they fail to fully exploit the complementary strengths of multimodal perception and cannot effectively align look-ahead visual cues with local geometric details. In contrast, humans and animals actively direct their gaze toward task-relevant regions, such as nearby footholds, gaps, or distant obstacles, based on the situation. This suggests that robust humanoid locomotion requires not only multimodal perception, but also an active perception mechanism for deciding *where to look* and *which perceived information matters most for the next step(s)*.

To this end, we propose **TAGA**, an active perception framework for generalizable agile humanoid locomotion. TAGA is built around a hierarchical gaze mechanism: vision provides long-range terrain preview, height scans supply precise local geometry, and a learned active gaze policy selectively attends to the most locomotion-relevant regions, adaptively fusing these complementary signals. Concretely, TAGA consists of two core components. A *Task-Relevant Active Gaze Module* fuses vision, proprioception, and motion commands to predict which terrain region is most relevant for the next movement and crops the corresponding patch from the height scan. A *Visuomotor Fusion Encoder* then applies cross-attention over this selected region, emphasizing geometric structures critical for foothold placement and terrain-aware decision making. This hierarchical design increases the effective information density of observations, reduces interference from irrelevant terrain, and simplifies downstream policy learning. The main contributions of this work are summarized as follows:

1. We present TAGA, a perceptive locomotion system that integrates depth vision, height scan, and proprioception into a unified sensing and control pipeline for humanoids operating in challenging terrain with onboard computation, while improving training efficiency over full-context.
2. At the core of TAGA is an emergent hierarchical active gaze mechanism learned without explicit gaze supervision: the policy learns to select a task-relevant terrain patch via visual and proprioceptive cues, then applies fine-grained attention to that region for precise understanding.

3. We deployed TAGA on Unitree G1 and demonstrated state-of-the-art (SOTA) performance across gaps, stepping stones, narrow beams, stairs, and outdoor terrain. Notably, the robot achieves a 120 cm gap crossing, surpassing the best reported result by 50%.

2 Related Works

Terrain Mapping-Based Perceptive Locomotion. A major class of perceptive locomotion methods relies on explicit geometric terrain representations such as height scans, elevation maps, or voxel grids to guide locomotion policies [18, 19, 20, 21]. These approaches generally achieve strong local terrain awareness and precise foothold placement in challenging tasks such as stepping stones, sparse footholds, and gap traversal [22, 23]. Recent methods further improve terrain understanding through learned geometric encodings, including multi-layer height scans [24], 3D terrain representations [25], and Attention-Based Map Encoding (AME) [8]. However, mapping-based approaches are usually constrained by the size and resolution of the terrain representation, which limits the tradeoff between local detail and broader terrain coverage [24, 25]. Their performance also depends on map quality and can degrade under localization uncertainty, occlusions, or deficient perception [26].

Vision-Based Perceptive Locomotion under Partial Observations. Vision-based methods reduce reliance on explicit mapping by learning policies from depth images, RGB-D inputs, or egocentric vision [13, 27, 28]. They have enabled agile locomotion across sparse footholds, discontinuous terrains, complex obstacle traversal, and humanoid stepping tasks [12, 29, 30], with recent humanoid-oriented works further improving performance through internal models, depth reconstruction, voxel-grid navigation, and limited-view omnidirectional control [10, 31, 32]. However, these policies often encode terrain reasoning implicitly [15, 28] and remain vulnerable to partial observability, especially when forward-facing sensors miss underfoot geometry needed for sparse foothold placement [33, 34]. Recent methods address this issue using memory or reconstruction, including implicit-explicit learning, visuospatial or volumetric memory, world-model perception, spatial recurrent memory, neural scene representations, sparse-terrain reconstruction, and diffusion-based occupancy synthesis [35, 36, 37, 38, 39, 40]. Nevertheless, they can still suffer from viewpoint changes, motion uncertainty, and drift between stored terrain representations and the robot's current pose, particularly when latent memory lacks explicit geometric constraints [41, 42].

Active Perception for Locomotion. Recent studies suggest that robust locomotion depends not only on perceiving the environment, but also on selecting task-relevant terrain regions during traversal. Prior attention-based locomotion methods have shown that selective terrain encoding and exteroceptive–proprioceptive fusion can improve robustness and generalization [8], while Cross-modal Transformers integrate visual and proprioceptive representations for terrain reasoning [15]. More recent adaptive perception methods, such as ADAPT, mainly improve perception robustness by adaptively clipping noisy observations and suppressing perception noise [43], while CART selects relevant temporal context for terrain adaptation [44]. Nevertheless, most existing methods still emphasize short-horizon terrain reasoning, while proactive allocation of perception across both nearby footholds and future terrain structures remains underexplored.

3 TAGA Framework

3.1 Problem Formulation

We formulate humanoid perceptive locomotion as a partially observable Markov decision process (POMDP; see Appendix A). The policy $\pi(a_t|o_t)$ maps the observation $o_t = \{p_t^H, d_t, h_t^{xyz}\}$ to an action $a_t \in \mathbb{R}^{29}$ representing target joint positions, where $d_t \in \mathbb{R}^{1\times 36\times 64}$ is a forward-facing depth image and $h_t^{xyz} \in \mathbb{R}^{3\times 21\times 21}$ is a local height scan. The proprioceptive input is a 5-frame history $p_t^H = \{p_{t-4}, \ldots, p_t\}$, where each frame is $p_t = \{\omega_{b,t}, g_{b,t}, q_t, \dot{q}_t, a_{t-1}, c_t\}$. Here, $\omega_{b,t} \in \mathbb{R}^3$ is the measured torso angular velocity, $g_{b,t} \in \mathbb{R}^3$ is the projected gravity vector, $q_t, \dot{q}_t \in \mathbb{R}^{29}$ are joint positions and velocities, and $a_{t-1} \in \mathbb{R}^{29}$ is the previous action. The command $c_t = (v_{x,t}^{\mathrm{cmd}}, v_{y,t}^{\mathrm{cmd}}, \dot{\psi}_t^{\mathrm{cmd}}) \in \mathbb{R}^3$ specifies the desired forward velocity, lateral velocity, and yaw rate.

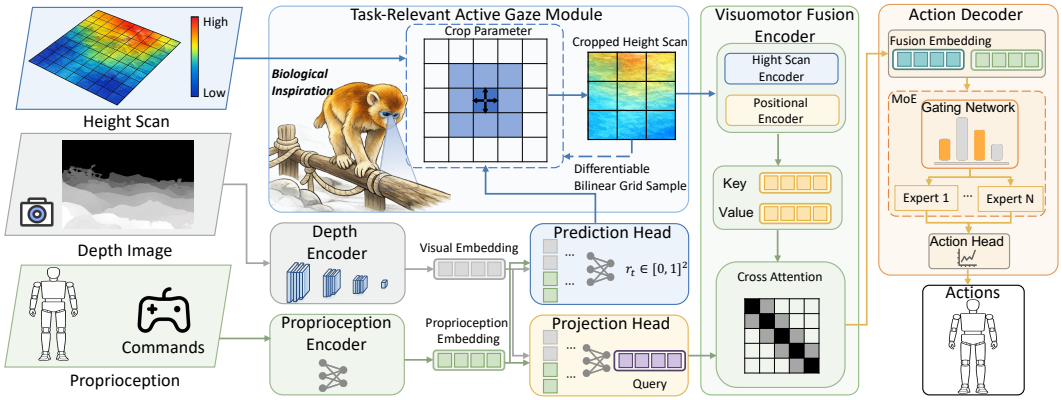

Figure 3: The architecture of TAGA.

3.2 Neural Network Design

Overview. As shown in Fig. 3, the TAGA policy $\pi(a_t|o_t)$ is parameterized by a neural network that takes the multimodal observation o_t as input and outputs a vector action a_t. TAGA first encodes the depth image d_t and proprioception history p_t^H using a CNN-based depth encoder ϕ_d and an MLP-based proprioception encoder ϕ_p, producing a visual embedding $e_t^d \in \mathbb{R}^{128}$ and a proprioceptive embedding $e_t^p \in \mathbb{R}^{128}$, respectively. The visual embedding captures distant terrain awareness, while the proprioceptive embedding encodes the robot's dynamic state and command context. Given these embeddings and the height scan h_t^{xyz}, TAGA employs a hierarchical gaze mechanism to extract locomotion-relevant terrain information. The first stage, the *task-relevant active gaze module*, predicts a region of interest (ROI) in the height scan. The second stage, the *visuomotor fusion encoder*, further emphasizes terrain cues relevant to the next locomotion decision by producing a fusion embedding e_t^{pg} based on the cropped ROI. Finally, a mixture-of-experts (MoE) based *action decoder* maps the fusion embedding e_t^{pg} and proprioceptive embedding e_t^p to the action output a_t. This design enables vision and proprioception to guide where the robot should look, while height scans provide precise local geometry, thereby supporting terrain-aware agile locomotion.

Task-Relevant Active Gaze Module. TAGA enables a robot to actively focus its gaze on the ROI via the task-relevant active gaze module. Conditioning on visual preview and proprioceptive state, this module filters irrelevant terrain and directs perception to the most informative region for the next locomotion step. Concretely, a lightweight prediction head f_{roi} takes the visual and proprioceptive embeddings as input and predicts a normalized two-dimensional gaze location: $r_t = f_{\text{roi}}([e_t^d, e_t^p]), r_t \in [0,1]^2$. The predicted location is then mapped onto the full local height-scan grid and used to crop a compact terrain patch: $\tilde{h}_t^{xyz} = \text{Crop}(h_t^{xyz}, r_t), \tilde{h}_t^{xyz} \in \mathbb{R}^{3\times K\times K}$, where $K = 11$ denotes the crop size. The crop is implemented using differentiable bilinear grid sampling, allowing this module to be optimized end-to-end with the policy. To discourage degenerate solutions near the height-scan boundary, we apply a boundary penalty $\mathcal{L}_{\text{roi}}$ (detailed in Appendix D).

Visuomotor Fusion Encoder. TAGA further identifies which perceived information is most relevant through the visuomotor fusion encoder. Given the cropped patch $\tilde{h}_t^{xyz}$, the encoder first extracts terrain features using a lightweight CNN, while the corresponding spatial coordinates are embedded by an MLP. These features are then fused to obtain pointwise terrain embeddings $E_t^m \in \mathbb{R}^{K\times K\times 128}$, which serve as the key and value vectors. Meanwhile, the query vector is generated from the visual and proprioceptive embeddings through a projection head f_{proj}. Finally, the visuomotor fusion encoder applies a multi-head cross-attention layer to obtain the fusion embedding: $e_t^{pg} = \text{MHA}\big(f_{proj}([e_t^p, e_t^d]), E_t^m, E_t^m\big)$. The resulting embedding $e_t^{pg} \in \mathbb{R}^{128}$ forms a terrain-aware visuomotor representation for downstream action decoding.

Action Decoder. The action decoder is implemented as an MoE module [45] to increase policy expressiveness through adaptive action composition. The actor receives the proprioceptive embedding e_t^p and the fusion embedding e_t^{pg} as $e_t^\pi = [e_t^p, e_t^{pg}]$. A gating network $g(\cdot)$ computes soft expert weights $\alpha_t^i = \text{Softmax}(g(e_t^\pi))_i, i = 1, \ldots, N_e$, and the final action is given by a weighted combination of expert outputs: $a_t = \sum_{i=1}^{N_e} \alpha_t^i \mathcal{E}_i(e_t^\pi)$, where $\mathcal{E}_i$ denotes the i-th expert and $N_e = 5$ in our implementation. Through soft routing, all experts contribute to action generation with input-

dependent weights. This allows the decoder to adaptively compose expert outputs based on TAGA's task-relevant perceptual and proprioceptive embeddings, improving policy expressiveness for diverse locomotion conditions.

4 Training TAGA Policy via Reinforcement Learning

We train TAGA using asymmetric actor-critic PPO in massively parallel simulation [46]. The actor is described in Sec.3.2, while the critic additionally receives privileged information, including the ground-truth base linear velocity and the full uncropped height scan. To support progressive policy learning over risky terrains, we construct multiple terrain types with increasing difficulty and train the policy using curriculum learning. To encourage natural gait and stable control, we further apply AMP-style regularization and safety-oriented termination. To improve both skill acquisition and hardware robustness, we adopt a two-stage training procedure. Finally, we define the full training objective with symmetry augmentation[47] (detailed in Appendix D).

Terrain Design and Curriculum. We design a training terrain set composed of multiple challenging terrain types and train TAGA across these environments. This terrain set covers representative traversal skills, including gap crossing, stair climbing, sparse-foothold traversal, narrow-beam walking, obstacle crossing, and slope locomotion (details in Appendix C.1), allowing the robot to acquire a broad set of terrain-adaptive locomotion behaviors. Each terrain type is divided into 10 difficulty levels by varying geometric parameters such as gap width, step height, and foothold size. During training, each robot is assigned to a terrain level according to its locomotion performance: successful traversal promotes it to harder terrains, while failure moves it back to easier ones.

Task Reward and Safety Constraints. The base task reward r_t^{env} combines command tracking, posture and joint regularization, and contact-related terms (detailed in Appendix C.2). An AMP-style reward r_t^{AMP} is added to encourage human-like gait, produced by a discriminator trained to distinguish policy motions from motion capture data [48]. Since extreme agile locomotion may encourage unstable policies, we apply early termination on physically unsafe states, including illegal non-foot-body contacts, excessive torso tilt, insufficient base height, and abnormally large hip-link acceleration during foot contact. These termination conditions serve as hard safety boundaries, filtering out irrecoverable failure cases and focusing learning on stable, feasible locomotion.

Two-stage Training. Following AME [8], we use a two-stage procedure to balance skill acquisition and real-world robustness. In the first stage, the policy is trained in a clean setting without observation noise or domain randomization, allowing the robot to acquire core locomotion skills efficiently. Starting from this checkpoint, the second stage fine-tunes the policy with reduced entropy and deployment-oriented randomization, including actuation variations, visual degradation, height-scan noise, external pushes, and terrain perturbations. This improves robustness to noisy real-world sensing and dynamics while preserving the learned active gaze behavior of TAGA.

Loss Function. Two auxiliary objectives regularize the learned representations: a contrastive loss $\mathcal{L}_{\text{con}}$ aligning MoE gating latents with full height-scan context to encourage expert specialization [45], and a gaze boundary penalty $\mathcal{L}_{\text{roi}}$ preventing TAGA from degenerately fixating on height-map edges. Putting everything together, the full training objective combines the PPO surrogate loss on the augmented reward $\tilde{r} = r_t^{env} + \eta r_t^{AMP}$ with a value loss $\mathcal{L}_{\text{value}}$, an entropy bonus $\mathcal{H}(\pi_\theta)$, and the auxiliary terms:

$$\mathcal{L}_{\text{policy}} = \mathcal{L}_{\text{PPO}}(\tilde{r}) + c_v \mathcal{L}_{\text{value}} - c_e \mathcal{H}(\pi_\theta) + \lambda_c \mathcal{L}_{\text{con}} + \lambda_b \mathcal{L}_{\text{roi}}, \tag{1}$$

where η is the AMP reward coefficient, and c_v, c_e, λ_c, and λ_b are loss weights.

5 Experiments

We train TAGA in Isaac Lab [49] with 8,000 parallel environments on four NVIDIA GeForce RTX 5090 GPUs. The policy is trained over 30k iterations, and then fine-tuned for 10k iterations, requiring about 17 RTX-5090 GPU-days in total. It executes actions at $50\,\text{Hz}$, tracked by a low-level

Table 1: Comparison of TAGA with baseline methods and ablation variants across challenging terrain types. We report GPU count, total training cost in GPU-days, and success rates. Each method is evaluated over 1000 trials per terrain when applicable. Bold indicates the best success rate in each terrain column and results within 0.5% of the best.

Method	# GPUs ↓	GPU-days ↓	Gaps ↑	Stepping Stones ↑	Beam ↑	High Platform ↑	Terrain C1 ↑	Terrain C2 ↑
CReF [50]	2	∼10	97.40%	52.30%	96.50%	98.70%	85.20%	43.10%
TAGA-HSOnly	4	∼17	93.10%	92.50%	**98.30%**	**99.60%**	90.50%	91.50%
TAGA-InactiveGaze	4	∼14	57.10%	83.20%	95.60%	**100.00%**	72.70%	48.80%
TAGA-FullScan	8	∼49	**99.50%**	**98.00%**	97.50%	**100.00%**	**93.40%**	92.50%
TAGA-NoAMP	4	∼16	96.40%	97.20%	97.60%	**99.50%**	89.80%	91.60%
TAGA (Ours)	4	∼17	98.30%	**97.90%**	**98.50%**	**100.00%**	**93.70%**	**93.90%**

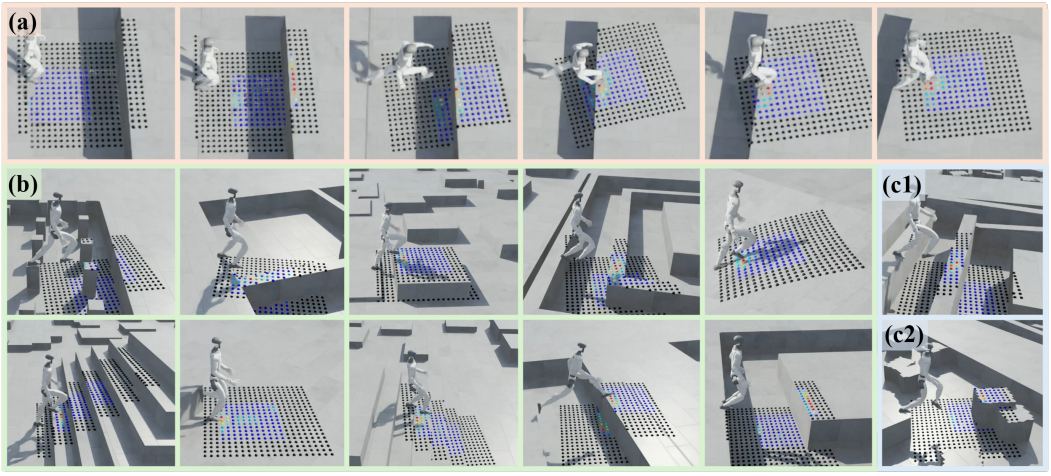

Figure 4: Visualization of the learned active gaze regions and attention-weight distributions of TAGA in simulation. Black points denote the full local height-scan grid, while colored points indicate the predicted ROI. Red and blue points indicate higher and lower attention weights, respectively. (a) shows how the active gaze regions shift during a $1.2\,\mathrm{m}$ gap crossing, and (b) demonstrates attention shifts across diverse terrains. (c) illustrates out-of-distribution testing terrains from training

200 Hz PD controller. Details are provided in Appendix C.3. For real-world experiments, TAGA is deployed on a Unitree G1 humanoid robot with onboard inference on an NVIDIA Jetson Orin.

5.1 Simulation Evaluation

We first evaluate TAGA in simulation, where terrain conditions and sensing ablations can be controlled systematically. The goal is to characterize its locomotion capability and robustness before real-world deployment. Specifically, we study three questions: ***Q1:*** What is the complementary nature of the information provided by visual preview and height scans during challenging terrain traversal? ***Q2:***What are the advantages of our gaze module dynamically selecting task-relevant terrain regions instead of processing the full height scan? ***Q3:*** Do AMP priors improve motion naturalness, whole-body coordination, and dynamic stability? We conduct a comprehensive ablation study for TAGA and compare it with a vision-based baseline. **TAGA-HSOnly** removes depth input and uses only height scans with proprioception. **TAGA-InactiveGaze** deactivates the Gaze Module and uses a fixed ROI. **TAGA-FullScan** removes the Gaze module to only use the full height scan. **TAGA-NoAMP** removes AMP priors. Since TAGA is designed around local height-scan selection, directly removing height scans would yield an ill-matched vision-only baseline. Instead, we compare TAGA with **CReF** [50], a vision-based humanoid locomotion method.

Multimodal Perception (Q1). Comparing TAGA with CReF and TAGA-HSOnly highlights the complementary roles of visual preview and height scans. TAGA matches or surpasses the vision-based CReF on gaps, beams, and high platforms, and substantially outperforms it on stepping stones (Table 1). This suggests vision alone is insufficient for sparse footholds, where precise foot placement demands accurate local geometry. In contrast, TAGA-HSOnly, which relies solely on height scans and proprioception, struggles with gaps. This indicates that height-scan-only perception lacks sufficient preview to anticipate distant discontinuities and landing regions. Together, these results support the complementary roles of vision and height scans: vision provides anticipatory preview,

while height scans provide local geometric precision for robust terrain traversal. Under OOD (out of distribution) terrain challenges (C1, C2), TAGA achieves 93.70% and 93.90%, demonstrating stronger robustness and generalization across diverse terrain combinations.

Gaze Module (Q2). We assess the effectiveness of our gaze module through visualization (Fig. 4) and ablation (Table 1). Instead of using a fixed crop, TAGA predicts task-relevant regions guided by fused visual and proprioceptive cues. During gap crossing, the gaze shifts from the current support region forward to the opposite edge as the robot prepares to cross, resembling the anticipatory gaze behavior observed in humans and animals during locomotion (Fig. 4(a)). On stepping stones and beams, it moves toward sparse footholds or the traversable strip; on continuous terrains, the robot's gaze remains local, to cover nearby height changes and contact regions (Fig. 4 (b)). In our ablations, TAGA achieves performance comparable to TAGA-FullScan with 65.2% lower training cost than TAGA-FullScan. TAGA-InactiveGaze, with the same compact input but a fixed crop, degrades sharply on gaps and stepping stones where distant footholds lie outside its window.

Motion Prior (Q3). Removing the AMP priors does not substantially hurt task completion, with TAGA-NoAMP performing close to TAGA across all terrains (Table 1). However, Fig. 5 shows a clear drop in motion quality. Without AMP, the robot exhibits unnatural motions, including inward knee collapse and short, shuffling steps, especially during turning and terrain transitions. These high-frequency foot motions also made real-world deployment brittle: TAGA-NoAMP proves far less robust to sim-to-real gap and small disturbances.

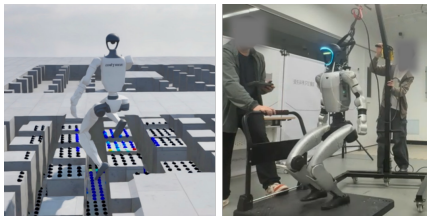

Figure 5: Robot stance without AMP. Without motion guidance, the robot constantly walk with its leg bent.

5.2 Real-World Evaluation

We further evaluate our policy in real-world environments of increasing complexity, moving from controlled indoor trials with designed challenging terrains to outdoor tests involving perception degradation, terrain variation, and physical disturbances. Table 2 compares TAGA with reported real-world perceptive humanoid and biped locomotion methods, while Fig. 1 and Fig. 6 show representative indoor and outdoor trials. TAGA covers the widest range of real-world terrain categories among currently reported methods: it **achieves the largest gap-crossing distance** and can traverse sparse-foothold areas, significantly outperforming previous reported SOTA results on both metrics.

Table 2: Real-world traversal capabilities reported in perceptive humanoid and biped locomotion works. n.r.indicates that the behavior was demonstrated but the terrain geometry was not numerically reported; –indicates that the behavior was not evaluated. Bold and underline indicate the best and second-best results, respectively.

Work	Robot	Gap	Platform	Sparse foothold	Beam	Stairs
HPL [16]	Unitree H1	**80 cm**	<u>42 cm</u>	–	✗	✗
PIM [10]	Unitree H1	<u>70 cm</u>	**50 cm**	–	✗	✓
GA-PHL [14]	LimX Oli	46 cm	–	–	✗	✓
AME-1 [8]	Fourier GR-1	n.r.	–	n.r., uneven	✓	✓
AME-2 [17]	LimX Tron1	60 cm	**48 cm**	40 cm spacing, uneven	✓	✓
Vel Tracking AME-2	Unitree G1	<u>90 cm</u>	<u>40 cm</u>	40 cm, uneven	✓	✓
Now You See That [51]	Honor	45 cm	<u>40 cm</u>	–	✗	✓
BeamDojo [29]	Unitree G1	50 cm	–	45 cm spacing, flat	✓	✗
MoRE [52]	Unitree G1	40 cm	30 cm	–	✗	✓
Hiking in the Wild [53]	Unitree G1	50 cm	32 cm	–	✗	✗
Gallant [32]	Unitree G1	40 cm	30 cm	–	✗	✓
CReF [50]	Agibot X2	80 cm	<u>40 cm</u>	–	✗	✓
RPL [54]	Unitree G1	–	–	60 cm spacing, flat	✗	✓
CMoE [45]	Unitree G1	80 cm	30 cm	–	✗	✓
Ours	Unitree G1	**120 cm**	<u>40 cm</u>	**70 cm spacing, uneven**	✓	✓

Indoor Evaluation. As shown in Fig. 6(**A**), we evaluate TAGA on highly discontinuous indoor terrains with only sparse, separated footholds—long gaps, narrow blocks, and stepping stones—where

Figure 6: Real-world evaluation of TAGA from controlled indoor terrains to unstructured outdoor environments. **(A)** Indoor trials test limit traversal over discontinuous terrains such as platform transitions, large steps, and sparse stepping blocks. **(B)** Outdoor trials evaluate robustness to terrain variation, perception degradation, and external disturbances. Our results show that TAGA maintains stable locomotion across discrete foothold selection, terrain variation, and physical perturbations.

the robot must carefully plan its footholds and generate sufficient momentum between contacts. TAGA handles these near-limit conditions, crossing $120\,\mathrm{cm}$ gaps, balancing on narrow footholds, and traversing sparse blocks with 50–$70\,\mathrm{cm}$ spacing and $10\,\mathrm{cm}$ height variation. Beyond these metrics, TAGA exhibits emergent terrain-adaptive behaviors. When crossing long gaps (Fig. 6**(A2)**), the robot steps onto the block edge and pushes off for extra momentum rather than taking off from the center of the support region. Upon landing, it bends its knees to absorb impact and quickly recovers balance. These strategies emerge without any hand-crafted motion, indicating that TAGA learns practical locomotion skills for challenging discrete terrains.

Outdoor Evaluation. We further evaluate TAGA in outdoor and low-light environments (Fig. 1 and Fig. 6), which introduce varied ground materials and challenging perception conditions, including changing illumination, background clutter, depth degradation, sensor noise, as well as occluded footholds in tall grass and confined spaces. Despite these challenges, the robot maintains stable locomotion across terrain types, demonstrating robustness to real-world perception noise and contact variations. Under external disturbances such as kicks and pushes, TAGA absorbs the perturbations, recovers balance, and continues walking without falling. These results indicate that TAGA learns a robust terrain-aware control strategy effective under both perceptual uncertainty and physical disturbances, enabling reliable real-world deployment.

6 Conclusion

We presented TAGA, a terrain-aware active gaze framework for agile humanoid locomotion. TAGA improves traversability across gaps, narrow beams, stairs, elevated platforms, and sparse stepping stones in both simulation and real-world experiments. The learned gaze behavior emerges without explicit supervision, and hardware deployment demonstrates stable locomotion under low-light conditions and strong external disturbances. By fusing vision, proprioception, and motion commands, TAGA selectively attends to task-relevant terrain regions, combining look-ahead visual with local geometric precision for anticipatory, terrain-aware decision-making.

7 Limitations

The dynamic maneuvers that TAGA achieves, including explosive jump, precise foothold targeting, and dynamic balance recovery, exert a high thermal load on the actuators during extended operation. This can progressively let actuators overheat, gradually damaging control accuracy and causing reduced jump distance or missed footholds. Additionally, poor height scan quality on complex terrain can cause improper gaits or failures. Future directions include improving tolerance to actuator degradation through online adaptation, better height scan reconstruction, and incorporating uncertainty-aware control to maintain precise foothold selection under degraded hardware and sensing conditions.

References

[1] M. Sombolestan and Q. Nguyen. Adaptive-force-based control of dynamic legged locomotion over uneven terrain. *IEEE Transactions on Robotics*, 40:2462–2477, 2024.

[2] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang. Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[3] Y. Zhang, Y. Yuan, P. Gurunath, I. Gupta, S. Omidshafiei, A.-a. Agha-mohammadi, M. Vazquez-Chanlatte, L. Pedersen, T. He, and G. Shi. Falcon: Learning force-adaptive humanoid loco-manipulation. *arXiv preprint arXiv:2505.06776*, 2025.

[4] N. Fey, G. B. Margolis, M. Peticco, and P. Agrawal. Bridging the sim-to-real gap for athletic loco-manipulation. *arXiv preprint arXiv:2502.10894*, 2025.

[5] M. Murooka, K. Chappellet, A. Tanguy, M. Benallegue, I. Kumagai, M. Morisawa, F. Kanehiro, and A. Kheddar. Humanoid loco-manipulations pattern generation and stabilization control. *IEEE Robotics and Automation Letters*, 6(3):5597–5604, 2021.

[6] K. Bouyarmane, K. Chappellet, J. Vaillant, and A. Kheddar. Quadratic programming for multirobot and task-space force control. *IEEE Transactions on Robotics*, 35(1):64–77, 2018.

[7] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. Humanplus: Humanoid shadowing and imitation from humans. *arXiv preprint arXiv:2406.10454*, 2024.

[8] J. He, C. Zhang, F. Jenelten, R. Grandia, M. Bächer, and M. Hutter. Attention-based map encoding for learning generalized legged locomotion. *Science Robotics*, 10(105):eadv3604, 2025.

[9] 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*, 7(62):eabk2822, 2022.

[10] J. Long, J. Ren, M. Shi, Z. Wang, T. Huang, P. Luo, and J. Pang. Learning humanoid locomotion with perceptive internal model. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9997–10003. IEEE, 2025.

[11] D. Hoeller, N. Rudin, D. Sako, and M. Hutter. Anymal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics*, 9(88):eadi7566, 2024.

[12] X. Cheng, K. Shi, A. Agarwal, and D. Pathak. Extreme parkour with legged robots. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11443–11450. IEEE, 2024.

[13] A. Agarwal, A. Kumar, J. Malik, and D. Pathak. Legged locomotion in challenging terrains using egocentric vision. In *Conference on robot learning*, pages 403–415. PMLR, 2023.

[14] H. Song, H. Zhu, T. Yu, Y. Liu, M. Yuan, W. Zhou, H. Chen, and H. Li. Gait-adaptive perceptive humanoid locomotion with real-time under-base terrain reconstruction. *IEEE Robotics and Automation Letters*, 2026.

[15] R. Yang, M. Zhang, N. Hansen, H. Xu, and X. Wang. Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=kFdPX1VdgXx.

[16] Z. Zhuang, S. Yao, and H. Zhao. Humanoid parkour learning. In *8th Conference on Robot Learning*, 2024. URL https://openreview.net/forum?id=fs7ia3FqUM.

[17] 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.

[18] P. Fankhauser, M. Bjelonic, C. D. Bellicoso, T. Miki, and M. Hutter. Robust rough-terrain locomotion with a quadrupedal robot. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5761–5768. IEEE, 2018.

[19] F. Jenelten, T. Miki, A. E. Vijayan, M. Bjelonic, and M. Hutter. Perceptive locomotion in rough terrain–online foothold optimization. *IEEE Robotics and Automation Letters*, 5(4):5370–5376, 2020.

[20] Z. Wang, Y. Li, L. Xu, H. Shi, Z. Ma, Z. Chu, C. Li, F. Gao, K. Yang, and K. Wang. Sf-tim: A simple framework for enhancing quadrupedal robot jumping agility by combining terrain imagination and measurement. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10676–10683. IEEE, 2025.

[21] 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)*, pages 2273–2280. IEEE, 2022.

[22] Y. Dong, J. Ma, L. Zhao, W. Li, and P. Lu. Marg: Mastering risky gap terrains for legged robots with elevation mapping. *IEEE Transactions on Robotics*, 2025.

[23] 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)*, pages 11864–11871. IEEE, 2024.

[24] Y. Chen, J. Ma, Z. Luo, Y. Han, Y. Dong, B. Xu, and P. Lu. Learning autonomous and safe quadruped traversal of complex terrains using multi-layer elevation maps. *IEEE Robotics and Automation Letters*, 2025.

[25] T. Miki, J. Lee, L. Wellhausen, and M. Hutter. Learning to walk in confined spaces using 3d representation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8649–8656. IEEE, 2024.

[26] P. Fankhauser, M. Bloesch, and M. Hutter. Probabilistic terrain mapping for mobile robots with uncertain localization. *IEEE Robotics and Automation Letters*, 3(4):3019–3026, 2018.

[27] 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 *5th Annual Conference on Robot Learning*, 2021.

[28] 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)*, pages 56–62. IEEE, 2024.

[29] H. Wang, Z. Wang, J. Ren, Q. Ben, T. Huang, W. Zhang, and J. Pang. Beamdojo: Learning agile humanoid locomotion on sparse footholds. *arXiv preprint arXiv:2502.10363*, 2025.

[30] 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.

[31] J. Sun, G. Han, P. Sun, W. Zhao, J. Cao, J. Wang, Y. Guo, and Q. Zhang. Dpl: Depth-only perceptive humanoid locomotion via realistic depth synthesis and cross-attention terrain reconstruction. *arXiv preprint arXiv:2510.07152*, 2025.

[32] Q. Ben, B. Xu, K. Li, F. Jia, W. Zhang, J. Wang, J. Wang, D. Lin, and J. Pang. Gallant: Voxel grid-based humanoid locomotion and local-navigation across 3d constrained terrains. *arXiv preprint arXiv:2511.14625*, 2025.

[33] S. Li, S. Luo, J. Wu, and Q. Zhu. Move: Multi-skill omnidirectional legged locomotion with limited view in 3d environments. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7647–7653. IEEE, 2025.

[34] 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.

[35] 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*, 9(11):9986–9993, 2024.

[36] 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*, pages 1430–1440, 2023.

[37] 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)*, pages 11531–11537. IEEE, 2025.

[38] C. Zhang, J. Jin, J. Frey, N. Rudin, M. Mattamala, C. Cadena, and M. Hutter. Resilient legged local navigation: Learning to traverse with compromised perception end-to-end. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 34–41. IEEE, 2024.

[39] D. Hoeller, N. Rudin, C. Choy, A. Anandkumar, and M. Hutter. Neural scene representation for locomotion on structured terrain. *IEEE Robotics and Automation Letters*, 7(4):8667–8674, 2022.

[40] R. Yu, Q. Wang, H. Li, Z. Jun, Z. Wang, J. Wu, and Q. Zhu. Start: Traversing sparse footholds with terrain reconstruction. *IEEE Robotics and Automation Letters*, 11(2):2194–2201, 2025.

[41] 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*, page 02783649251401926, 2025.

[42] A. Reed, B. Crowe, D. Albin, L. Achey, B. Hayes, and C. Heckman. Scenesense: Diffusion models for 3d occupancy synthesis from partial observation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7383–7390. IEEE, 2024.

[43] S. Shao, T. Huang, W. Gao, and S. Zhang. Adapt: Adaptive dual-projection architecture for perceptive traversal. *arXiv preprint arXiv:2603.16328*, 2026.

[44] K. Singh, Y. Kim, Y. Turkar, and K. Dantu. Cart: Context-aware terrain adaptation using temporal sequence selection for legged robots. *arXiv preprint arXiv:2604.14344*, 2026.

[45] S. Ma, H. Chen, Z. Xu, Y. Zhao, K. Wu, R. Yang, L. Zou, Z. Gan, and W. Ding. Cmoe: Contrastive mixture of experts for motion control and terrain adaptation of humanoid robots. *arXiv preprint arXiv:2603.03067*, 2026.

[46] C. Schwarke, M. Mittal, N. Rudin, D. Hoeller, and M. Hutter. Rsl-rl: A learning library for robotics research. *arXiv preprint arXiv:2509.10771*, 2025.

[47] M. Mittal, N. Rudin, V. Klemm, A. Allshire, and M. Hutter. Symmetry considerations for learning task symmetric robot policies. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7433–7439. IEEE, 2024.

[48] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa. Amp: adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics*, 40(4):1–20, 2021. ISSN 1557-7368. doi:10.1145/3450626.3459670. URL http://dx.doi.org/10.1145/3450626.3459670.

[49] M. Mittal, P. Roth, J. Tigue, A. Richard, O. Zhang, P. Du, A. Serrano-Muñoz, X. Yao, R. Zurbrügg, N. Rudin, L. Wawrzyniak, M. Rakhsha, A. Denzler, E. Heiden, A. Borovicka, O. Ahmed, I. Akinola, A. Anwar, M. T. Carlson, J. Y. Feng, A. Garg, R. Gasoto, L. Gulich, Y. Guo, M. Gussert, A. Hansen, M. Kulkarni, C. Li, W. Liu, V. Makoviychuk, G. Malczyk, H. Mazhar, M. Moghani, A. Murali, M. Noseworthy, A. Poddubny, N. Ratliff, W. Rehberg, C. Schwarke, R. Singh, J. L. Smith, B. Tang, R. Thaker, M. Trepte, K. V. Wyk, F. Yu, A. Millane, V. Ramasamy, R. Steiner, S. Subramanian, C. Volk, C. Chen, N. Jawale, A. V. Kuruttukulam, M. A. Lin, A. Mandlekar, K. Patzwaldt, J. Welsh, H. Zhao, F. Anes, J.-F. Lafleche, N. Moënne-Loccoz, S. Park, R. Stepinski, D. V. Gelder, C. Amevor, J. Carius, J. Chang, A. H. Chen, P. de Heras Ciechomski, G. Daviet, M. Mohajerani, J. von Muralt, V. Reutskyy, M. Sauter, S. Schirm, E. L. Shi, P. Terdiman, K. Vilella, T. Widmer, G. Yeoman, T. Chen, S. Grizan, C. Li, L. Li, C. Smith, R. Wiltz, K. Alexis, Y. Chang, D. Chu, L. J. Fan, F. Farshidian, A. Handa, S. Huang, M. Hutter, Y. Narang, S. Pouya, S. Sheng, Y. Zhu, M. Macklin, A. Moravanszky, P. Reist, Y. Guo, D. Hoeller, and G. State. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025. URL https://arxiv.org/abs/2511.04831.

[50] Y. Hao, R. Yu, S. Luo, G. Zhang, J. Wu, and Q. Zhu. Cref: Cross-modal and recurrent fusion for depth-conditioned humanoid locomotion. *arXiv preprint arXiv:2603.29452*, 2026.

[51] W. Sun, Y. Su, L. Huang, A. Zhang, D. Wei, M. San, D. Tian, E. Cao, B. Cao, Y. Liu, et al. Now you see that: Learning end-to-end humanoid locomotion from raw pixels. *arXiv preprint arXiv:2602.06382*, 2026.

[52] D. Wang, X. Wang, X. Liu, J. Shi, Y. Zhao, C. Bai, and X. Li. More: Mixture of residual experts for humanoid lifelike gaits learning on complex terrains. *arXiv preprint arXiv:2506.08840*, 2025.

[53] 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.

[54] Y. Zhang, Y. Seo, J. Chen, Y. Yuan, K. Sreenath, P. Abbeel, C. Sferrazza, K. Liu, R. Duan, and G. Shi. Rpl: Learning robust humanoid perceptive locomotion on challenging terrains. *arXiv preprint arXiv:2602.03002*, 2026.

[55] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, Oct. 2019.

A The Details of POMDP

We formulate humanoid perceptive locomotion as a partially observable Markov decision process (POMDP), denoted as a 6-tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$. Here, $\mathcal{S}$ is the state space, $\mathcal{O}$ is the observation space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the transition kernel, $\mathcal{R}$ is the reward function, and γ is the discount factor. At each timestep t, the robot has an underlying state $s_t \in \mathcal{S}$, which is not fully observable by the policy. The state includes the full robot configuration, velocity, contact state, actuator state, and surrounding terrain geometry. Instead of accessing s_t directly, the policy receives a multimodal observation $o_t \in \mathcal{O}$ composed of proprioception, egocentric depth perception, and local height-scan geometry:

$$o_t = \{p_t^H, d_t, h_t^{xyz}\}. \tag{2}$$

Here, $d_t \in \mathbb{R}^{1\times 36\times 64}$ denotes the forward-facing depth image, and $h_t^{xyz} \in \mathbb{R}^{3\times 21\times 21}$ denotes the egocentric height scan, where each grid cell stores an (x, y, z) terrain point. The proprioceptive observation is a five-frame history

$$p_t^H = \{p_{t-4}, \ldots, p_t\}, \tag{3}$$

where each frame is defined as

$$p_t = \{\omega_{b,t}, g_{b,t}, q_t, \dot{q}_t, a_{t-1}, c_t\}. \tag{4}$$

Specifically, $\omega_{b,t} \in \mathbb{R}^3$ is the torso angular velocity, $g_{b,t} \in \mathbb{R}^3$ is the projected gravity vector, $q_t, \dot{q}_t \in \mathbb{R}^{29}$ are joint positions and velocities, $a_{t-1} \in \mathbb{R}^{29}$ is the previous action, and $c_t = (v_{x,t}^{\mathrm{cmd}}, v_{y,t}^{\mathrm{cmd}}, \dot{\psi}_t^{\mathrm{cmd}}) \in \mathbb{R}^3$ is the velocity command.

The policy $\pi_\theta(a_t|o_t)$ maps the observation to an action $a_t \in \mathcal{A}$, where $a_t \in \mathbb{R}^{29}$ represents target joint positions at 50 Hz, tracked by a low-level PD controller at 200 Hz. After the action is executed, the environment transitions via $\mathcal{P}(s_{t+1}|s_t, a_t)$ and yields a reward $\tilde{r}_t = \mathcal{R}(s_t, a_t)$ that encourages velocity tracking, stable posture, safe contacts, and human-like motion (see Appendix C.2 for the full reward specification). The complete observation and action spaces are detailed in Table 3. The objective is to maximize the expected discounted return:

$$\max_\theta \ \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T-1} \gamma^t \tilde{r}_t\right]. \tag{5}$$

Table 3: Action and observation space specifications.

Modality	**Symbol**	**Description**
Action		
Joint position targets	$a_t \in \mathbb{R}^{29}$	50 Hz control output, tracked via PD at 200 Hz
Proprioception		
Base angular velocity	$\omega_{b,t} \in \mathbb{R}^3$	Torso angular velocity
Projected gravity	$g_{b,t} \in \mathbb{R}^3$	Gravity direction in torso frame
Joint positions	$q_t \in \mathbb{R}^{29}$	Measured joint angles
Joint velocities	$\dot{q}_t \in \mathbb{R}^{29}$	Measured joint velocities
Previous action	$a_{t-1} \in \mathbb{R}^{29}$	Last commanded joint targets
Velocity command	$c_t \in \mathbb{R}^3$	$(v_x^{\mathrm{cmd}}, v_y^{\mathrm{cmd}}, \dot{\psi}^{\mathrm{cmd}})$ target
Depth		
Depth image	$d_t \in \mathbb{R}^{1\times 36\times 64}$	Forward-facing, long-range terrain preview
Height Scan		
Height map	$h_t^{xyz} \in \mathbb{R}^{3\times 21\times 21}$	Egocentric (x, y, z) terrain grid, local surroundings

B Definition of Gaze Learning

We define *gaze* as a combination of a normalized gaze location $r_t = (r_t^x, r_t^y) \in [0,1]^2$ and an ROI centered at the gaze location. The ROI is defined as a terrain patch $\tilde{h}_t^{xyz} \in \mathbb{R}^{3\times K\times K}$ over the local height scan $h_t^{xyz} \in \mathbb{R}^{3\times M\times M}$ with $M = 21$. We crop the ROI based on the gaze location, where the normalized gaze coordinates are first mapped to the height scan grid as $(u_t, v_t) = (\lfloor r_t^x M \rfloor, \lfloor r_t^y M \rfloor)$, and the ROI is obtained by extracting terrain points through $\tilde{h}_t^{xyz} = \left\{ h_t^{xyz}(i,j) \;\middle|\; |i-u_t| \le \frac{K}{2},\ |j-v_t| \le \frac{K}{2},\ i,j \in \mathbb{Z} \right\}$. Consequently, we define *gaze learning* as training the gaze prediction head $f_{\text{roi}}(\cdot)$ to output the gaze location $r_t = f_{\text{roi}}([e_t^d, e_t^p])$, which is then used to extract an ROI $\tilde{h}_t^{xyz} = \text{Crop}(h_t^{xyz}, r_t)$, where $\text{Crop}(\cdot)$ denotes the ROI cropping operation defined above. The extracted ROI $\tilde{h}_t^{xyz}$ is subsequently provided to the downstream locomotion policy for decision-making.

C Training Details

C.1 Terrain Design

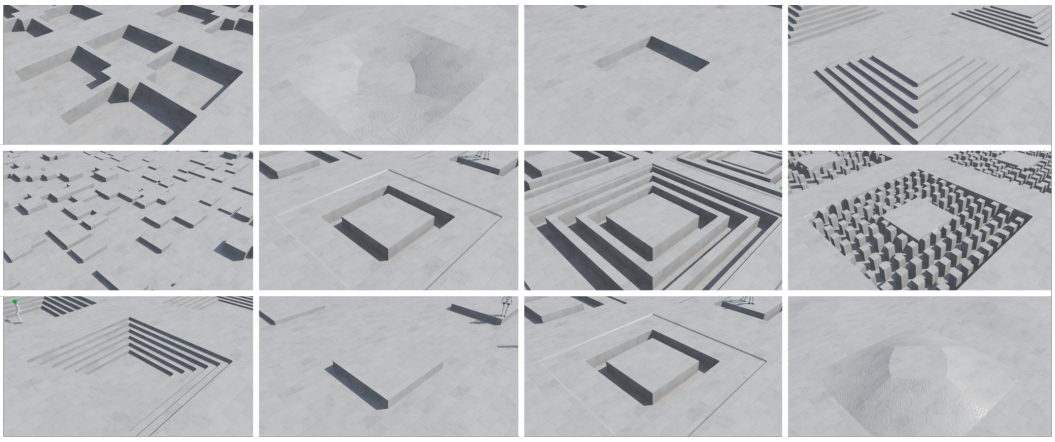

Figure 7: Training Terrains for TAGA

To improve traversability across diverse terrain conditions, we construct a broad set of representative terrain types during training, including ascending and descending stairs, gaps, stepping stones (sparse footholds), box obstacles, elevated platforms, and sloped surfaces, as illustrated in Fig. 7. This terrain mixture exposes the policy to diverse geometric features, including elevation changes, discontinuous support regions, sparse footholds, obstacle negotiation, and inclined surfaces, thereby encouraging the emergence of generalizable terrain-adaptive locomotion strategies.

Terrain sampling during training follows a probabilistic strategy to balance exposure across terrain types. Stepping stones are sampled with a probability of 0.3, as they most directly exercise precise foothold selection and sparse-terrain reasoning. Ascending stairs, descending stairs, upslope terrains, and downslope terrains are each sampled with a probability of 0.05, since they mainly introduce structured elevation variations and are less demanding in terms of sparse foothold selection. The remaining probability mass is evenly distributed among gaps, box obstacles, and elevated platforms. This sampling strategy allows the policy to experience a wide range of terrain geometries and improves its adaptability to challenging deployment scenarios.

C.2 Reward Terms

The locomotion policy is trained with an augmented reward $\tilde{r}_t = r_t^{\text{env}} + \eta r_t^{\text{AMP}}$, where r_t^{env} denotes the base task reward, r_t^{AMP} denotes the Adversarial Motion Prior (AMP) reward, and η controls the contribution of the AMP term. The base task reward combines command-tracking objectives, pos-

ture regularization, joint-level regularization, contact-aware locomotion terms, and safety-oriented penalties. It is designed to encourage agile yet stable locomotion with coordinated contacts and physically plausible motion patterns.

The primary task objective is velocity tracking, where the robot is rewarded for matching the commanded linear and angular velocities. To improve stability and motion quality, we additionally penalize excessive torso rotation, body tilt, joint torques, joint velocities, joint accelerations, action-rate changes, and deviations from nominal body configurations. These regularization terms encourage smooth and energy-efficient motions while reducing unnecessary body oscillations.

To promote robust locomotion behaviors, several contact-related rewards and penalties are introduced. These include rewards for maintaining appropriate foot swing durations and penalties for foot sliding, stumbling, undesired body contacts, loss of ground contact, and excessively narrow foot placement. Together, these terms encourage coordinated stepping behaviors and stable support transitions on challenging terrain. During the second-stage fine-tuning process, additional penalties are activated to improve robustness under deployment-oriented disturbances and terrain variations.

The detailed components and coefficients of r_t^{env} are summarized in Table 4. Here, $c_t^{xy} = (v_{x,t}^{\mathrm{cmd}}, v_{y,t}^{\mathrm{cmd}})$ and $c_t^{\omega} = \dot{\psi}_t^{\mathrm{cmd}}$ denote the commanded linear and angular velocity components; $\mathcal{F}$ denotes the set of feet, $\mathcal{B}$ denotes the set of robot links, q^0 is the default joint configuration, f_t^b is the contact force acting on body b, and $[\cdot]_+ = \max(\cdot, 0)$ denotes the positive-part operator. The gating function $G_{\mathrm{flat}}(h_t)$ activates the torso-orientation penalty only when the local terrain is sufficiently flat, preventing unnecessary penalization on uneven terrain that naturally requires body inclination.

In addition to these handcrafted reward terms, an AMP reward is used throughout training to encourage human-like locomotion styles. The discriminator is trained using motion-capture data and provides an auxiliary reward that regularizes the learned behaviors toward natural gait patterns. Specifically, we use retargeted AMASS [55] motions as the reference motion distribution. Both the simulated robot motion and the reference motion are encoded into a compact motion descriptor containing body motion, body orientation, and joint-state information. The discriminator is trained to distinguish policy-generated motion descriptors from AMASS reference descriptors, while the policy is rewarded when its motion is classified as closer to the reference distribution. Unlike frame-wise imitation, this adversarial formulation provides a distribution-level style prior, allowing the policy to preserve task performance while producing smoother and more natural locomotion patterns.

C.3 Training Hyperparameters

Table 5 summarizes the main PPO optimization hyperparameters used for training TAGA. Unless otherwise specified, the same hyperparameters are used throughout all experiments. Values shown in parentheses correspond to the second-stage fine-tuning configuration when different from the first-stage training setup.

C.4 Termination Conditions

To facilitate stable training and avoid unsafe failure modes, we use the following episode termination conditions during training:

1. **Timeout and terrain boundary.** An episode is reset after the maximum episode length of $20\,\mathrm{s}$, or when the robot moves outside the valid terrain region.
2. **Illegal non-foot contact.** We terminate the episode when non-foot bodies make contact with the environment. The monitored links include the torso, pelvis, waist, shoulder links, elbow links, hip links, and knee links. Contacts are detected using a force threshold of $1\,\mathrm{N}$.
3. **Bad torso orientation.** The episode is terminated when the torso orientation deviates excessively from upright. In implementation, this is measured by the angle between the torso-frame projected gravity direction and the upright axis, with a threshold of $0.8\,\mathrm{rad}$.

Table 4: Reward terms used for locomotion training. The formulas denote the unweighted reward terms. Values in parentheses indicate coefficients used during the second-stage fine-tuning.

Category	Term	Formula	Weight
Task Objective	Alive	$\mathbb{I}_{\text{alive}}$	3.0
	Linear velocity tracking	$\exp\!\left(-\lVert v^{xy}_{\text{torso},t} - c^{xy}_t \rVert^2 / \sigma_v^2\right)$	2.0 (2.5)
	Yaw velocity tracking	$\exp\!\left(-(\omega^z_{\text{torso},t} - c^{\omega}_t)^2 / \sigma_\omega^2\right)$	3.0
	Yaw command regularization	$\lvert c^{\omega}_t \rvert$	-1.0
	Forward progress	$\mathbb{I}(c^x_t > 0.3)\left[\mathbb{I}(v^x_{\text{torso},t} < 0.15) + \mathbb{I}(v^x_{\text{torso},t} < 0) + \mathbb{I}(v^x_{\text{torso},t} < -0.15)\right]$	-0.5
Posture Regularization	Torso angular velocity	$\lVert \omega^{xy}_{\text{torso},t} \rVert^2$	-0.05
	Torso orientation	$G_{\text{flat}}(h_t)\, \lVert g^{xy}_{\text{torso},t} \rVert^2$	-2.0
	Pelvis orientation	$\lVert g^{xy}_{\text{pelvis},t} \rVert^2$	-0.5
Joint Regularization	Joint torque	$\lVert \tau_t \rVert^2$	-1.5×10^{-7}
	Joint velocity	$\lVert \dot{q}_t \rVert^2$	-5.0×10^{-4}
	Joint acceleration	$\lVert \ddot{q}_t \rVert^2$	-1.25×10^{-7}
	Link acceleration	$\frac{1}{\lvert\mathcal{B}\rvert}\sum_{b\in\mathcal{B}} \lVert \dot{v}^b_t \rVert$	-0.01
	Hip deviation	$\lVert q^{\text{hip}}_t - q^{\text{hip},0} \rVert^2$	-0.1
	Arm deviation	$\lVert q^{\text{arm}}_t - q^{\text{arm},0} \rVert_1$	-0.3
	Waist deviation	$\lVert q^{\text{waist}}_t - q^{\text{waist},0} \rVert_1$	-1.0
	Joint position limit	$\sum_j \operatorname{dist}(q^j_t, \mathcal{Q}^j)$	-5.0
	Joint velocity limit	$\sum_j \left[\lvert \dot{q}^j_t \rvert - \rho_{\dot{q}} \dot{q}^j_{\max}\right]_+$	-1.0
	Joint torque limit	$\sum_j \left[\lvert \tau^j_t \rvert - \rho_\tau \tau^j_{\max}\right]^2_+$	-0.01
	Action rate	$\lVert a_t - a_{t-1} \rVert^2$	-0.005
Contact and Gait	Undesired contact	$\sum_{b\notin\mathcal{F}} \mathbb{I}(\lVert f^b_t \rVert > \epsilon_f)$	-1.0
	Foot air time	$\mathbb{I}_{\text{cmd}}\mathbb{I}_{\text{single}} \min_{f\in\mathcal{F}} T^f_{\text{mode}}$	0.25
	Air/contact time variance	$\operatorname{Var}_{f\in\mathcal{F}}(\text{clip}(T^f_{\text{air}}, 0.5)) + \operatorname{Var}_{f\in\mathcal{F}}(\text{clip}(T^f_{\text{contact}}, 0.5))$	$-0.7\,(-2.0)$
	Foot stumble	$\mathbb{I}(\exists f \in \mathcal{F} : \lVert f^{xy}_t \rVert > 4 \lvert f^z_t \rvert)$	$-2.0\,(-5.0)$
	Foot slide	$\sum_{f\in\mathcal{F}} \mathbb{I}^f_{\text{contact}} \lVert v^{xy,f}_t \rVert$	-0.1
	Foot orientation	$\sum_{f\in\mathcal{F}} \mathbb{I}^f_{\text{contact}} \lVert g^{xy,f}_t \rVert$	-0.5
	No-fly	$\mathbb{I}\!\left(\sum_{f\in\mathcal{F}} \mathbb{I}^f_{\text{contact}} = 0\right)$	-2.0
	Feet too near	$\left[d_{\min} - \lVert x^L_t - x^R_t \rVert\right]_+$	$-1.0\,(-5.0)$
Fine-tuning Only	Volume penetration	$\sum_{x\in\mathcal{V}} \mathbb{I}(\delta_x > 0)(\lVert v_x \rVert + \epsilon)\delta_x$	$0.0(-1.0)$
	Stand still	$\mathbb{I}(\lVert c^{xy}_t \rVert < \epsilon_c)\mathbb{I}(\lvert c^{\omega}_t \rvert < \epsilon_c)(\lVert q_t - q^0 \rVert_1 - b)$	$0.0(-0.3)$

Table 5: Main training hyperparameters. Values in parentheses indicate the second-stage fine-tuning configuration when different from the first stage.

Parameter	**Value**
Rollout length	24
PPO epochs	5
Mini-batches	10
Learning rate	1×10^{-3}
Learning-rate schedule	Adaptive
Discount factor γ	0.99
GAE parameter λ	0.95
PPO clipping coefficient	0.2
Value loss coefficient	1.0
Entropy coefficient	0.005 (0.002)
Gradient clipping norm	1.0

4. **Low base height.** We terminate states in which the robot has effectively fallen or collapsed. This includes cases where the root height is below $0.5\,\mathrm{m}$ relative to the terrain origin, or the root clearance above the local terrain estimate is below $0.2\,\mathrm{m}$.
5. **High hip-link acceleration during foot contact.** To filter out high-impact failure cases, we terminate the episode when a foot is in contact and the corresponding hip-pitch link experiences excessive linear acceleration. A contact is considered active when the foot contact force exceeds $1\,\mathrm{N}$, and termination is triggered when the hip-link acceleration exceeds $225\,\mathrm{m/s^2}$. This condition discourages stiff, high-impact landings and encourages recoverable contact behaviors.

Triggering any unsafe termination condition resets the episode immediately. These conditions prevent the policy from exploiting unstable behaviors and focus learning on recoverable locomotion strategies.

C.5 Two-stage Training and Domain Randomization

We use a two-stage training procedure to balance skill acquisition and real-world robustness. In the first stage, TAGA is trained in a clean simulation setting without observation noise or deployment-oriented dynamics randomization, allowing the policy to efficiently acquire the core locomotion skills. In the second stage, we fine-tune from the pre-trained checkpoint with a reduced entropy coefficient and deployment-oriented domain randomization.

For robot dynamics, we randomize the added base payload within $[-1.0, 3.0]\,\mathrm{kg}$, perturb the base center of mass by $[-0.05, 0.05]\,\mathrm{m}$ along the x and y axes and along the z axis, and sample motor delays from 0–3 delay steps. Contact properties are randomized with static friction in $[0.3, 1.0]$, dynamic friction in $[0.3, 0.8]$, and restitution in $[0.0, 0.5]$. We also apply random pushes every 10–$15\,\mathrm{s}$ with planar velocity perturbations in $[-0.5, 0.5]\,\mathrm{m/s}$.

For observations, we add independently sampled uniform noise to proprioceptive channels, with maximum magnitudes of $0.2\,\mathrm{rad/s}$ for base angular velocity, 0.05 for projected gravity, $0.01\,\mathrm{rad}$ for joint positions, and $1.5\,\mathrm{rad/s}$ for joint velocities. The depth camera is updated at $30\,\mathrm{Hz}$ during fine-tuning and is degraded with contour corruption, random depth artifacts, reflections, sky artifacts, Gaussian blur, stereo failure for too-close surfaces, and robot self-occlusion. Depth values are clipped to $[0.4, 3.0]\,\mathrm{m}$ and then normalized before being fed to the policy.

For height scans, we add z-axis noise in $[-0.05, 0.05]\,\mathrm{m}$ and ray-cast drift in $[-0.05, 0.05]\,\mathrm{m}$ along each axis. We also randomly corrupt a small portion of height-scan returns to simulate missing or unreliable terrain measurements. Terrain geometry is further perturbed with Perlin roughness, randomized gap-edge transition widths, and virtual edge obstacles. These randomizations improve robustness to perception noise, actuation uncertainty, contact variation, and terrain irregularities that may arise during hardware deployment.

D Training Objectives

TAGA is trained using an asymmetric actor–critic PPO objective augmented with auxiliary regularization losses. The overall optimization objective is

$$\mathcal{L}_{\text{policy}} = \mathcal{L}_{\text{PPO}} + c_v \mathcal{L}_{\text{value}} - c_e \mathcal{H}(\pi_\theta) + \lambda_c \mathcal{L}_{\text{con}} + \lambda_b \mathcal{L}_{\text{roi}}, \tag{6}$$

where $\mathcal{L}_{\text{PPO}}$ is the PPO surrogate loss, $\mathcal{L}_{\text{value}}$ is the critic regression loss, $\mathcal{H}(\pi_\theta)$ is the policy entropy bonus, and $\mathcal{L}_{\text{con}}$ and $\mathcal{L}_{\text{roi}}$ denote the contrastive and gaze regularization losses, respectively.

D.1 MoE–Terrain Contrastive Loss

Following CMoE [45], we employ a contrastive objective to encourage terrain-aware soft routing in the MoE policy. Let B denote the batch size and $\mathbb{K}$ denote the number of learnable prototypes in a shared prototype dictionary C. The contrastive pair is formed by the actor-gate embedding e_i^g and the terrain embedding e_i^h encoded from the full height scan. Both embeddings are normalized and compared through the prototype dictionary.

For an embedding e, we define its prototype prediction distribution and balanced assignment as

$$P(e) = \mathrm{softmax}(eC^\top / T_{\text{con}}), \qquad Q(e) = \mathrm{Sinkhorn}(eC^\top), \tag{7}$$

where T_{con} is a temperature parameter. The contrastive loss is

$$\mathcal{L}_{\text{con}} = -\frac{1}{2B\mathbb{K}} \sum_{i=1}^{B} \left[Q(e_i^g)^\top \log P(e_i^h) + Q(e_i^h)^\top \log P(e_i^g) \right]. \tag{8}$$

This objective encourages the gating network to produce terrain-aware soft routing weights by aligning the gate embedding e_i^g with the terrain embedding e_i^h in the shared prototype space. This helps reduce terrain-invariant or nearly uniform expert mixtures and promotes functional differentiation among experts.

D.2 Gaze Boundary Loss

To prevent the learned gaze from degenerately selecting crops near the height-scan boundary, we penalize ROI centers that are too close to the edge of the normalized scan domain. For notational convenience, let

$$r_i = (r_i^x, r_i^y) \in [0,1]^2 \tag{9}$$

denote the normalized gaze location predicted by TAGA for the i-th sample in a mini-batch. Here, r_i is a batch-level notation for the corresponding time-indexed prediction r_t used in the policy rollout. Given a boundary margin $m = 0.05$, the boundary loss is defined as

$$\mathcal{L}_{\text{roi}} = \frac{1}{2B} \sum_{i=1}^{B} \left([m - r_i^x]_+ + [r_i^x - (1-m)]_+ + [m - r_i^y]_+ + [r_i^y - (1-m)]_+ \right), \tag{10}$$

where $[\xi]_+ = \max(\xi, 0)$. This regularizer keeps the selected crop inside the valid height-scan region and prevents unstable edge fixation.

Symmetry Augmentation. We exploit the humanoid's left-right morphological symmetry by mirroring proprioceptive observations, height scans, AMP states, and actions. For depth observations, a horizontally flipped virtual camera view is used. Both original and mirrored samples are included in PPO updates, encouraging symmetric locomotion behaviors and improving training efficiency.