

CoRe-MoE: Contrastive Reweighted Mixture of Experts for Multi-Terrain Humanoid Locomotion with Gait Adaptation

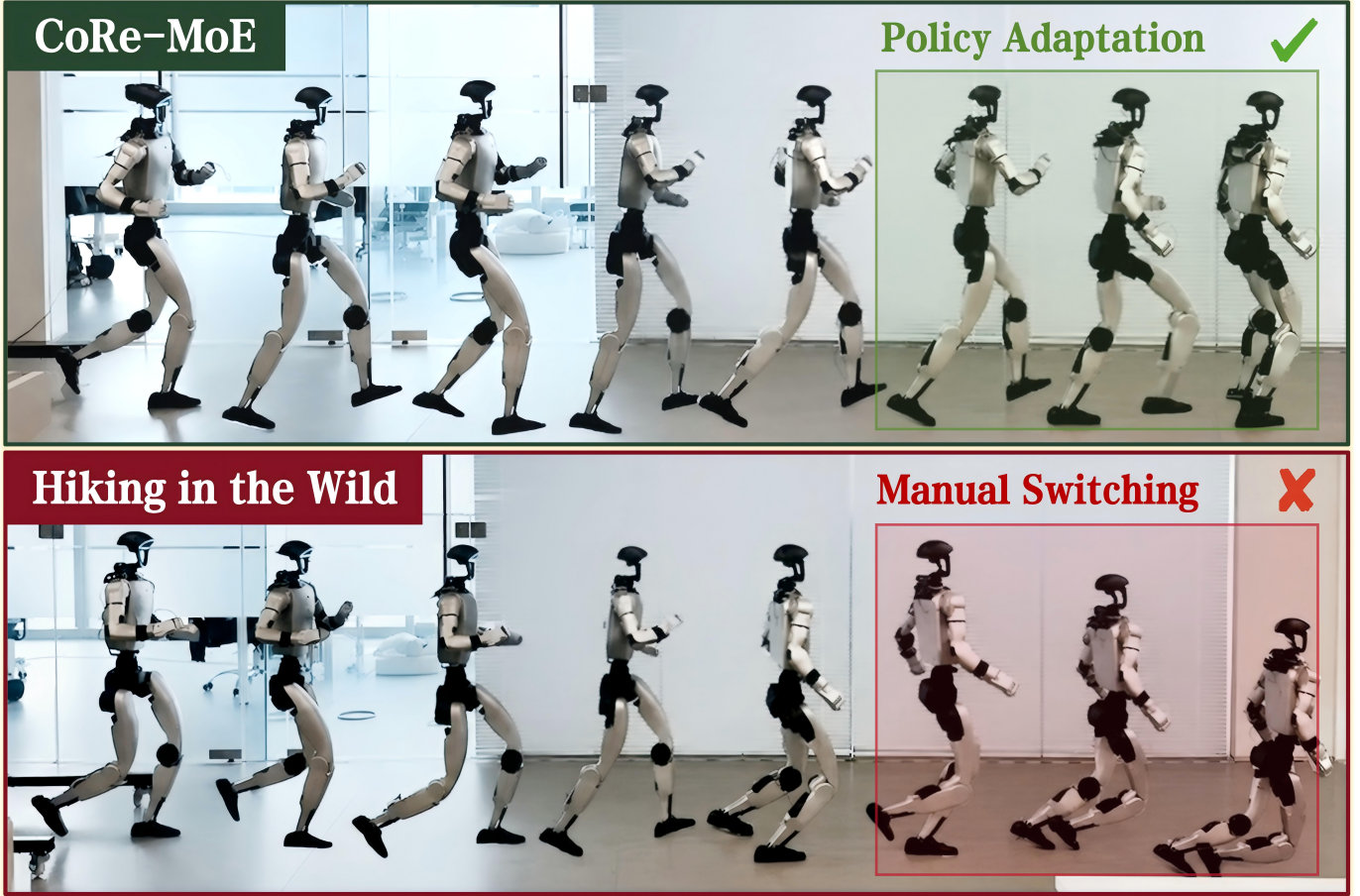


Fig. 1: We propose **CoRe-MoE** enables smooth, natural, and continuous transitions between walking and running using a single unified policy, while maintaining stable adaptability to complex terrain. In contrast, **Hiking in the Wild** relies on separately trained walking and running policies that must be manually switched during execution, leading to abrupt transitions, reduced continuity, and noticeably unnatural motion, especially under terrain variations. **Project Page:** <https://core-moe.github.io/>.

Abstract—Humans primarily rely on walking and running to traverse complex terrains. Similarly, humanoid robots should be able to smoothly transition between walking and running while maintaining natural and stable locomotion. However, unifying gait transition and multi-terrain adaptation within a single policy remains challenging due to gradient interference between tasks and the distribution shift caused by terrain variations. Although Mixture-of-Experts (MoE) architectures can mitigate multi-skill interference, direct joint training often fails to achieve clear expert specialization. To address these challenges, we propose **CoRe-MoE**, a two-stage reinforcement learning framework that decouples gait generation from terrain adaptation. In the first stage, a stable locomotion policy is learned to produce natural walking and running behaviors with smooth transitions. In the second stage, a terrain-aware MoE branch is introduced, and the gating network is trained with a contrastive objective to learn structured terrain representations and promote expert specialization. The final action is obtained through weighted fusion of the base gait policy and the terrain-aware branch, enabling the policy to preserve stable locomotion while adapting

to complex terrains. Extensive simulation results demonstrate that the proposed method outperforms baseline approaches in terms of success rate, locomotion stability, and multi-terrain adaptability. Furthermore, zero-shot deployment on a Unitree G1 humanoid robot validates the effectiveness of our framework, achieving robust walking and running across stairs, slopes, steps, obstacles, and unstructured outdoor terrains while maintaining accurate foothold control and dynamic stability.

Index Terms—humanoid locomotion, gait adaptation, reinforcement learning, contrastive learning, mixture of experts.

I. INTRODUCTION

HUMANOID robots are increasingly required to perform tasks in real-world environments, which often contain abrupt terrain variations such as stairs, slopes, and steps. Compared with wheeled robots [1], [2], humanoids can step over obstacles and traverse discontinuous terrains, making them more suitable for deployment in human-centered environments

[3], [4], [5], [6]. To achieve robust and human-like locomotion in such scenarios, robots must not only maintain basic stability but also proactively adjust their actions according to upcoming terrain changes [7], [8].

Humans rely primarily on two fundamental gaits: walking and running to adapt to complex terrains [9], [10], rarely employing additional complex or ineffective gaits to cross stairs, slopes, or gaps. This observation provides a design insight for humanoid robots: control strategies should focus on smooth switching between walking and running, rather than learning numerous artificial gait types [11], which contribute little to practical adaptability.

Despite the significant progress in reinforcement learning-based humanoid locomotion, key challenges remain. Pure proprioception-based “blind locomotion” methods can resist moderate disturbances through contact feedback [12], [13], [14], [15], but their reactive nature makes it difficult for the robot to anticipate hazards such as high steps or deep gaps. On the one hand, this can lead to instability or falls on critical terrains; on the other hand, jointly achieving gait switching and multi-terrain adaptation within a single policy often induces gradient conflicts, while the visual and dynamic variations introduced by complex terrains further increase learning difficulty.

In addition, Mixture-of-Experts (MoE) architectures [16] have shown promising potential in multi-skill learning and have been gradually applied to humanoid locomotion and multi-terrain adaptation [17], [11]. However, during direct training, the experts often fail to develop clear specialization: their activation patterns tend to be dispersed and highly overlapping across different terrains or gait conditions, preventing the policy from fully exploiting the advantages of the MoE structure and limiting its multi-terrain adaptability and gait control precision.

To address these issues, we propose the Contrastive Reweighted Mixture of Experts (CoRe-MoE), a two-stage training framework that integrates adaptive gait policies, visual perception, and contrastive MoE learning. Depth images [3], [18] are incorporated to enable proactive, terrain-aware motion adjustment. In the first stage, the policy learns stable walking and running on flat terrains and acquires natural and smooth gait transition capabilities. In the second stage, we introduce a terrain-aware MoE branch guided by SwAV-style contrastive learning [19]. This encourages the network to learn structured terrain representations and promotes expert specialization. The final action is produced by a weighted fusion of the base gait policy and the terrain-aware branch, enabling robust multi-terrain adaptability while preserving stable fundamental gait behaviors.

As illustrated in Figure 1, CoRe-MoE achieves continuous, natural transitions between walking and running within a single policy while maintaining stable performance on complex terrain. In contrast, existing methods—such as Hiking in the Wild—train walking and running separately and require manual switching during execution, leading to abrupt transitions, poor continuity, and degraded stability when terrain conditions

change.

Simulation results demonstrate that CoRe-MoE outperforms baseline methods in terms of success rate, gait stability, and multi-terrain adaptability. Further zero-shot deployment experiments on the Unitree G1 robot show that the policy achieves robust locomotion across stairs, slopes, and outdoor natural terrains, while maintaining accurate foot placement and dynamic stability under external disturbances, highlighting its potential for real-world applications.

The main contributions of this work are summarized as follows:

- We propose CoRe-MoE, a two-stage end-to-end training framework that enables smooth switching between walking and running while achieving multi-terrain adaptability and robustness.
- We design a terrain-aware MoE module based on contrastive learning that generates terrain-conditioned action adjustments and promotes clear expert specialization in different terrains.
- We perform a zero-shot deployment on the Unitree G1 robot, and the results demonstrate robust locomotion and dynamic stability across stairs, slopes, and natural outdoor terrains.

TABLE I: Comparison of locomotion capabilities across different methods. Unlike prior approaches that support limited gaits or require policy switching, CoRe-MoE is the first unified policy achieving walking, running, smooth transitions, and robust multi-terrain performance.

Method	Walk	Run	Auto Transition	Terrain Adapt
Beamdojo [6]	✓	✗	✗	✓
Now You See That [20]	✓	✗	✗	✓
MoRE [11]	✓	✓	✗	✓
Hiking in the Wild [3]	✓	✓	✗	✓
CoRe-MoE	✓	✓	✓	✓

II. RELATED WORK

A. Learning-based Humanoid Robot Control

Reinforcement learning (RL) has recently achieved substantial progress in humanoid locomotion. RL-based policies can learn stable walking and running behaviors directly from proprioceptive feedback and maintain robustness on moderately challenging terrains [21], [22], [12], [23]. Compared to traditional model-based controllers [24], [25], [26], these learning-based approaches are capable of handling unknown dynamics and unstructured environments, leading to significantly improved adaptability.

Although prior works have explored multi-gait or multi-terrain locomotion [11], [17], [27], their policies often rely on manually triggered gait switches and cannot autonomously adjust gait according to speed or terrain commands. Their generalization across fundamentally different terrain types such as stairs, gaps, and obstacles also remains limited, as shown in Table I.

Human locomotion studies show that people primarily rely on only two fundamental gaits—walking and running—to

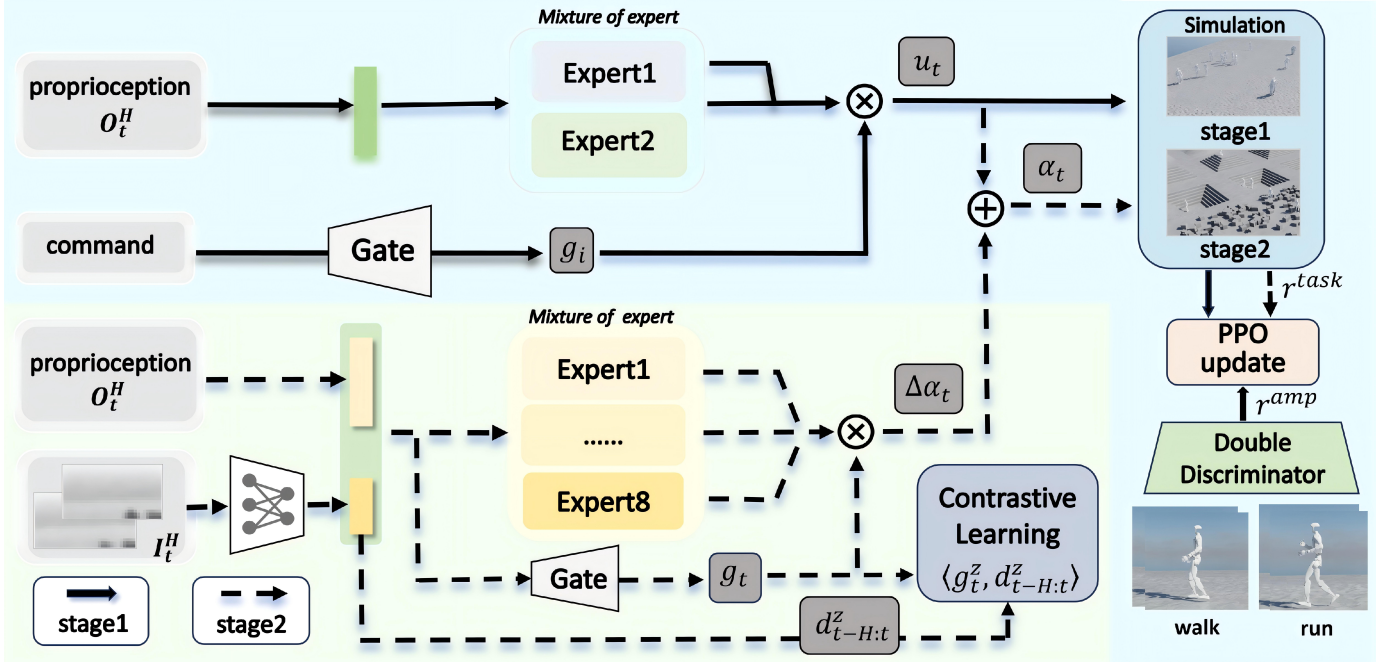


Fig. 2: Overview of our framework. In Stage 1, we learn a flat-terrain adaptive gait policy using MoE and AMP, enabling stable walking–running transitions. In Stage 2, we incorporate depth perception and employ MoE with contrastive learning to acquire terrain-aware control. During inference, the flat-terrain policy provides a gait prior and the multi-terrain policy applies terrain adjustments; their weighted fusion produces the final action, achieving natural gait and multi-terrain adaptability.

traverse diverse terrains, without resorting to additional artificial or inefficient gait patterns. This observation suggests that practical humanoid locomotion should emphasize autonomous and smooth transitions between these two core gaits while simultaneously maintaining strong multi-terrain adaptability.

B. Perception for Locomotion

In recent years, researchers have incorporated exteroceptive perception into humanoid locomotion policies [20], [28], [1], [6], [29] to enhance proactive adaptation to complex terrains. Blind locomotion methods, which rely solely on proprioceptive signals, can maintain stability on limited terrains but cannot anticipate upcoming terrain features. Traditional LiDAR-based approaches can provide height information [30], [7], [31], but their limited field-of-view, especially under the robot’s feet, combined with state estimation drift, low update frequency, and motion distortion, restricts their effectiveness in high-speed or dynamic scenarios [32], [6]. In contrast, depth cameras provide high-frequency perception and can be integrated with policy networks for end-to-end control [3], [33], [18], enabling the robot to proactively adjust its actions to handle complex terrains, thereby improving safety and adaptability.

C. Multi-terrain Adaptation and Mixture-of-Experts

Mixture-of-Experts architectures [16], [34] offer an effective approach for multi-skill learning and multi-terrain adaptation. By dividing the policy into specialized experts, the robot can learn distinct skills such as walking, running, and gap-crossing [11], [17]. However, during direct training, expert activations are often dispersed and lack specialization, limiting

the policy’s ability to adapt across different terrains. Contrastive learning methods, such as SwAV [19] and SimCLR [35], can guide experts to specialize according to terrain conditions, enhancing both policy robustness and multi-terrain adaptability. Our proposed CoRe-MoE framework combines weighted action correction with contrastive MoE learning, preserving gait stability while enabling proactive adaptation and flexible control over complex terrains.

III. METHOD

To address the coupled challenges of terrain disturbances and gait transitions, we aim for humanoid robots to maintain stable dynamics while achieving natural and continuous gait switching in complex environments. We adopt a two-stage training pipeline: the first stage focuses on learning stable and controllable fundamental gaits such as walking and running; the second stage extends the objective to enable multi-terrain adaptation while preserving a consistent anthropomorphic motion style, as illustrated in Figure 2. This staged design improves training stability and allows the robot to perform smooth and coherent gait transitions across diverse terrains.

A. Problem Description

We formulate humanoid multi-terrain locomotion as a Partially Observable Markov Decision Process (POMDP) with a tuple $\langle S, A, P, R, \gamma \rangle$, where S is the state space, A is the action space, $P(s, a)$ is the transition function. The agent receives a reward from $R(s, a)$. The discount factor $\gamma \in [0, 1]$ balances immediate and future rewards. We use Proximal

Policy Optimization (PPO) [36] to learn the optimal policy π^* that maximizes the expected discounted return:

$$\pi^* = \arg \max \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right]. \quad (1)$$

We adopt an asymmetric actor-critic architecture. The policy network relies solely on deployable, non-privileged proprioceptive observations, defined at time t as:

$$O_t = [\omega_t, g_t, c_v^t, \theta_t, \dot{\theta}_t, a_{t-1}], \quad (2)$$

including robot angular velocity ω_t , gravity direction g_t , velocity command c_v^t , joint angle θ_t , velocity $\dot{\theta}_t$, last action a_{t-1} . In addition, the policy also receives the perception observation depth image I_t . The critic receives the same observation as the actor but is additionally provided with privileged base linear velocity v_t .

B. Adaptive Gait Locomotion Policy

In the first training stage, we aim to learn a stable and controllable locomotion policy that enables the humanoid robot to adaptively switch between walking and running on flat terrain. We employ two separate Adversarial Motion Priors (AMP) [37] networks, one for walking and one for running. Additionally, the walking trajectories are augmented with gait samples such as stair climbing and obstacle crossing to further enhance gait diversity and adaptability.

The policy adopts MoE architecture with historical frame inputs O_t^H to capture temporal dependencies. It is important to note that the depth image I_t is not used in this stage; locomotion control on flat terrain relies solely on proprioceptive observations to achieve gait switching. The MoE gating network takes the commanded velocity c_v^t as input and dynamically activates different experts, enabling the walking AMP under low-speed commands and the running AMP under high-speed commands. The final action is computed as a weighted sum of expert outputs through softmax gating:

$$u_t = \sum_{i=1}^N \text{softmax}(g_i) \cdot u_i, \quad (3)$$

where z_i means the i -th expert outputs and g_i is the expert activation.

The training reward consists of three terms, combined as

$$r_t = r_{task} + r_{amp} + r_{reg}, \quad (4)$$

where r_{amp} is provided by the activated AMP network to enforce realistic walking or running motions, r_{task} ensures locomotion stability and that the commanded velocity is achieved, r_{reg} is a regularization term that penalizes large or abrupt joint motions to promote smooth and physically plausible actions. By optimizing this combined reward, the policy learns to smoothly transition between walking and running, achieving natural and continuous gait adaptation while maintaining dynamic stability on flat terrain. After completing Stage 1 training, the entire policy network is frozen.

C. Adaptive Multi-Terrain Policy

In the second training stage, we introduce a multi-terrain policy on top of the adaptive gait policy learned in Stage 1 to enhance locomotion performance on complex terrains. This policy is built using a MoE architecture, where each expert receives not only historical proprioceptive observations O_t^H but also the corresponding historical depth image sequence I_t^H , allowing the model to capture both temporal dynamics and external terrain geometry. During this stage, the Stage 1 policy network is frozen, and the actions generated by the Stage 2 policy are added to those from Stage 1, enabling terrain-aware fine-grained adjustments while preserving the stable gait foundation learned previously. The reward function in this stage retains only the task reward r_{task} and regularization reward r_{reg} , ensuring the policy focuses on terrain adaptation while maintaining smooth and physically consistent actions.

To fully activate experts and encourage terrain-specific specialization, we adopt a SwAV-based contrastive learning objective that aligns the gating output with the encoded representation of the history of the depth image. Compared with single-frame information, the historical depth sequence provides more stable and structured cues about terrain geometry, enabling the gating network to form clearer terrain clusters. We align the gating embedding g_t^z with the depth history representation $d_{t-H:t}^z$:

$$\langle g_t^z, d_{t-H:t}^z \rangle \quad (5)$$

During each iteration, we collect batch trajectories from all environments. If a historical observation and its corresponding historical depth sequence belong to the same trajectory, they form a positive pair; otherwise, they are negative. The gate network output g_t serves as the source vector I_{source}^t , while the encoded representation of the historical depth sequence $d_{t-H:t}^z$ serves as the target vector I_{target}^z . Both vectors are normalized before computing the contrastive objective.

To predict the cluster assignment probabilities, we first normalize the prototype matrix:

$$\mathbf{E} = \{\bar{\mathbf{e}}_1, \dots, \bar{\mathbf{e}}_K\}, \quad (6)$$

and compute the source and target probabilities:

$$\begin{aligned} \mathbf{p}_{source}^t &= \frac{\exp(\frac{1}{\tau} I_{source}^t \mathbf{E}_k^\top)}{\sum_{k'} \exp(\frac{1}{\tau} I_{source}^t \mathbf{E}_{k'}^\top)} \\ \mathbf{p}_{target}^t &= \frac{\exp(\frac{1}{\tau} I_{target}^z \mathbf{E}_k^\top)}{\sum_{k'} \exp(\frac{1}{\tau} I_{target}^z \mathbf{E}_{k'}^\top)} \end{aligned} \quad (7)$$

where τ is a temperature parameter. To avoid trivial solutions, the Sinkhorn-Knopp algorithm is applied to both source and target probabilities to obtain target distributions q_{target}^t q_{source}^t . The final SwAV contrastive loss is:

$$\mathcal{J} = -\frac{1}{2H} \sum_{t=1}^H (q_{target}^t \log \mathbf{p}_{source}^t + q_{source}^t \log \mathbf{p}_{target}^t). \quad (8)$$

By aligning the embedding of the gating with the historical depth representation, the SwAV objective enables the gating

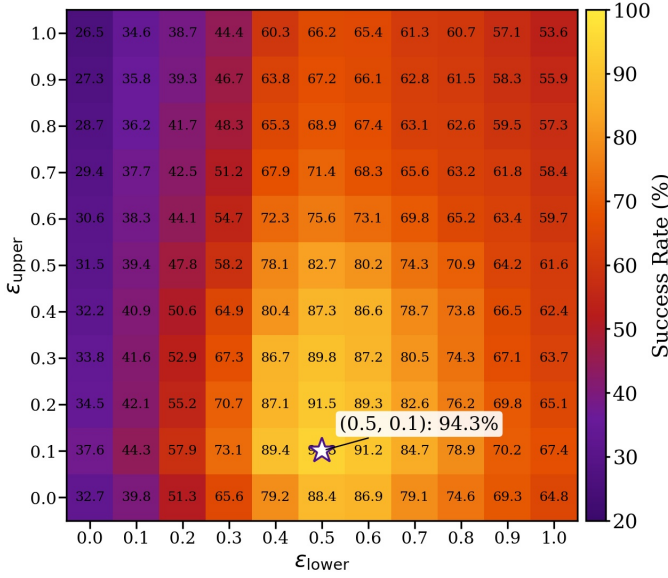


Fig. 3: Sensitivity analysis of the action fusion coefficients ϵ_{lower} and ϵ_{upper} . Colors indicate the success rate under multi-terrain evaluation.

network to better discriminate variations in terrain. As a result, different terrain types induce distinct cluster assignments, which in turn encourage the MoE experts to develop specialized behaviors. With more accurate expert activation, the policy is able to produce terrain-aware corrections rather than generic modifications. The MoE-weighted action:

$$\Delta a_t = \sum_{i=1}^N \text{softmax}(g_t) \cdot a_i, \quad (9)$$

which reflects the specific adjustments required by the upcoming terrain.

For action fusion, we apply different blending coefficients to different body parts to control the influence of terrain-aware corrections: $\epsilon_{upper} = 0.1$ for the upper body and $\epsilon_{lower} = 0.5$ for the lower body. The final action is defined as:

$$a_t = u_t + \epsilon \cdot \Delta a_t, \quad (10)$$

where ϵ is applied per body region accordingly. This design preserves the stable gait structure learned in Stage 1 while adding fine-grained, terrain-dependent refinements.

To investigate the influence of the fusion coefficients, we further performed a sensitivity analysis on ϵ_{upper} and ϵ_{lower} , as shown in Fig. 3. The results indicate that the policy is more sensitive to ϵ_{lower} , highlighting the dominant role of terrain-aware corrections of the lower-body in the adaptation of the foothold and the traversal of the terrain. When ϵ_{lower} is too small, the policy cannot respond effectively to terrain variations, whereas excessively large values may introduce overly aggressive corrections that compromise gait stability. In contrast, ϵ_{upper} has a relatively smaller impact on overall performance. The results show that the adopted configuration lies within a high-performance region, indicating that the proposed action fusion mechanism achieves a favorable balance between locomotion stability and terrain adaptability.

Through this mechanism, the multi-terrain branch serves as a lightweight yet effective adaptation layer, driven by contrastively shaped gating behavior, enabling robust locomotion across diverse and uneven terrains.

IV. EXPERIMENT CONFIGURATIONS

A. Training Parameters

We train all policies in NVIDIA Isaac Sim and Isaac Lab [38] 4096 parallel environments to ensure stable gradient estimates and rapid throughput. Motion priors are constructed from selected AMASS [39] and LAFAN [40] sequences, retargeted to the Unitree G1 model and used to supervise the AMP discriminator. The first-stage adaptive gait policy is trained for 25k iterations on an RTX 4090 GPU with a 2-expert MoE, while the second-stage residual multi-terrain policy is trained for 20k iterations with the MoE capacity expanded to 8 experts. For the contrastive learning module, we set the number of prototypes at 16 and the temperature at 0.1. Depth image preprocessing, including downsampling and spatial cropping, follows the procedure referenced in [3], ensuring consistent perceptual inputs across simulation and real-world deployment.

B. Reward Functions

Our reward formulation is based on standard locomotion objectives while incorporating two essential safety considerations. The first focuses on preventing missteps in unsupported terrain regions [6], ensuring that the policy avoids foot placements that could lead to immediate loss of balance. The second integrates a scalable terrain-edge detection module with foot-volume contact evaluation [3], encouraging the robot to land within a safe foothold margin and reducing the likelihood of edge-related slips. A complete description of all reward terms is provided in Table II.

C. Domain Randomization

To improve the real-world generalization of the learned policy, we introduce moderate domain randomization during training across dynamics, depth perception, and command inputs. This includes small perturbations to physical properties such as mass and friction, staged noise applied to ray-cast depth observations, and mild variations in velocity and heading commands. Periodic external forces are also applied to enhance disturbance robustness. Together, these randomizations strengthen the policy’s stability in real-world settings, with the detailed configurations summarized in Table III.

D. Curriculum Learning

To improve the policy’s adaptability to complex terrains, we adopt a curriculum learning strategy. During training, the environment is divided into 10 levels, and the curriculum level is dynamically increased based on the achieved velocity reward. Higher levels correspond to larger commanded target velocities and progressively more challenging terrains. This design guides the policy to learn robust locomotion behaviors in a gradual and structured manner.

TABLE II: Reward Functions for Humanoid Locomotion

Term	Equation	Weight
Velocity tracking	$\exp(-\ \mathbf{v}_{xy} - \mathbf{v}_{xy}^c\ _2^2/\sigma^2)$	2.0
Yaw tracking	$\exp(-\ \omega_z - \omega_z^c\ _2^2/\sigma^2)$	2.0
Heading error	$\ \psi - \psi^c\ _2^2$	-1.0
Is alive	1	3.0
Stand still	$\mathbb{I}(\ \mathbf{v}^c\ < \epsilon) \ \mathbf{v}\ _2^2$	-0.3
Don't wait	$\mathbb{I}(\ \mathbf{v}^c\ > \epsilon) \mathbb{I}(\ \mathbf{v}\ < \delta)$	-0.5
Orientation	$\ \mathbf{g}_{xy}\ _2^2$	-3.0
Pelvis orientation	$\ \mathbf{q}_{err,pelvis}\ _2^2$	-3.0
Angular velocity	$\ \omega_{xy}\ _2^2$	-0.05
Joint deviation hip	$\sum_{i \in \text{hip}} (\theta_i - \theta_i^{\text{def}})^2$	-0.5
Freeze upper body	$\sum_{i \in \text{upper}} \theta_i - \theta_i^{\text{def}} $	-0.004
Feet air time	$\sum_i t_{\text{air},i}$	0.5
Feet slide	$\sum_i \ \mathbf{v}_{i,xy}\ _2^2 \mathbb{I}(\text{contact}_i)$	-0.4
Feet at plane	$\sum_i (z_i - z_{\text{plane}})^2$	-0.1
Feet flat orientation	$\sum_i \ \mathbf{q}_{i,err}\ _2^2 \mathbb{I}(\text{contact}_i)$	-0.4
Feet close	$\exp(-d_{xy}^2/\sigma^2)$	0.4
Foothold penalty	$\sum_{i \in \{L,R\}} \mathbb{I}(c_i = 1) \mathbb{I}(h_i < h_{\text{thr}})$	-2.0
Volume penetration	$\sum \max(0, z_{\text{terrain}} - z_{\text{point}})$	-5.0
Undesired contacts	$\sum_i \mathbb{I}(\ \mathbf{F}_i\ _2 > F_{\text{thr}})$	-1.0
Joint torques	$\sum_i \tau_i^2$	$-1.5e^{-7}$
Joint acceleration	$\sum_i \ddot{\theta}_i^2$	$-1.25e^{-7}$
Joint velocity	$\sum_i \dot{\theta}_i^2$	$-1.0e^{-4}$
Action rate	$\ \mathbf{a}_t - \mathbf{a}_{t-1}\ _2^2$	-0.005
Energy	$\sum_i P_i^2$	$-5.0e^{-5}$
CoT penalty	$\sum_i \tau_i \dot{q}_i / (mg\ \mathbf{v}\)$	-0.05
Joint pos limits	$\sum_i \text{ReLU}(\theta_i - \theta_{\text{max}}) + \text{ReLU}(\theta_{\text{min}} - \theta_i)$	-1.0
Joint vel limits	$\sum_i \text{ReLU}(\dot{\theta}_i - \dot{\theta}_{\text{max}})$	-1.0
Torque limits	$\sum_i \text{ReLU}(\tau_i /\tau_{\text{max}} - 0.8)$	-0.01
AMP reward	$-\log(1 - D_\phi(\mathbf{s}_t))$	0.25

TABLE III: Parameters for Domain Randomization During Training

Term	Randomization Range
Static Friction	[0.1, 3.0]
Dynamic Friction	[0.1, 3.0]
Ground Restitution	[0.05, 0.5]
Base Mass Offset	[-5.0, 5.0]
K_p Scale	[0.9, 1.1]
K_d Scale	[0.9, 1.1]
Base Orientation Reset	[-0.1, 0.1]
Linear Velocity Reset	[-0.2, 0.2]
Angular Velocity Reset	[-0.2, 0.2]
Linear Velocity Perturbation	[-0.2, 0.2]
Initial Joint Positions	[-0.15, 0.15]
Initial Joint Velocities	[0.0, 0.1]
Crop & Resize	[18, 0, 16, 16]
Gaussian Noise	$\sigma = 0.02$, range [0.12, 1.2]
Disparity Artifact Synthesis	$P_{\text{mask}} = 0.05\%$
Depth Gaussian Blur	kernel size = 3×3 , $\sigma = 1.0$
Depth Range Normalization	[0.0, 2.5] $\rightarrow$ [0.0, 1.0]
OOD Perturbation Probability	$P_{\text{ood}} = 0.02\%$

V. EXPERIMENTS

In this section, we evaluate the robustness and performance of our two-stage adaptive locomotion framework through extensive experiments in both simulation and real-world environments. The policy is tested across flat and complex terrains, including stairs, platforms, slopes, and discrete gaps, summarized in Table IV, using both walking and running gaits. All quantitative results are reported as the mean and standard deviation (mean $\pm$ std) over multiple evaluation runs. Our

evaluation aims to answer the following three questions:

- **Q1:** In Stage 1, can the adaptive gait locomotion policy learn natural walking and running gaits on flat terrain, while enabling smooth transitions between them?
- **Q2:** In Stage 2, can the depth perception policy branch with contrastive learning enable stable locomotion and terrain-aware motion adjustment on complex terrains through action fusion with the Stage 1 policy?
- **Q3:** Can the trained policy be reliably deployed on physical hardware in a zero-shot manner, without additional fine-tuning?

TABLE IV: Terrain Types and Corresponding Parameter Ranges

Terrain Type	Terrain Feature	Range
Flat Ground	Flat terrain with mild variations	0.0–0.1m
Up Slope	Upward inclined terrain	0.0–35.0°
Down Slope	Downward inclined terrain	0.0–35.0°
Up Stairs	Standard stair ascent terrain	0.0–0.2m
Down Stairs	Standard stair descent terrain	0.0–0.2m
High Up Stairs	Steep and tall stair ascent terrain	0.0–0.45m
High Down Stairs	Steep and tall stair descent terrain	0.0–0.45m
High Obstacle	Tall single-step obstacle terrain	0.1–0.4m
Random Obstacle	Randomly distributed obstacle terrain	0.1–0.3m

A. Evaluation of the Adaptive Gait Locomotion Policy

We compared the proposed Adaptive Gait Locomotion Policy with three baseline methods: **(1) Mixed AMP:** which merges walking and running datasets to train a single AMP network; **(2) Sel AMP:** which selects walking or running references based on the commanded velocity to train a single AMP network; **(3) w/o MoE:** a single actor-critic network without using the MoE architecture.

To evaluate the performance of the Stage-1 adaptive gait policy on flat terrain, the robot receives distance-based velocity commands while navigating toward target points. When far from the target, the commanded velocity increases in the range of 1 to 2.5 m/s to induce a running gait, while near the target the velocity gradually decreases in the range of 0 to 1 m/s to produce a walking gait. This setup examines whether the policy can generate natural walking and running gaits and achieve smooth gait transitions on flat ground.

Locomotion performance was evaluated using four metrics: **(1) Velocity Tracking Score (Vel. Score.):** An averaged exponential kernel score computed from the squared error between the commanded and actual base linear velocities in the xy-plane. **(2) Success Rate (Succ.):** The percentage of trials in which the robot successfully reaches the specified target point; **(3) Average Traversal Ratio (Dist.):** The fraction of the target distance completed before reaching the goal or termination; **(4) Side-view Gait Sequencess:** Visual inspection of the robot's side-view motion sequences to assess the naturalness and smoothness of the walking and running gaits.

As shown in Table V, CoRe-MoE achieves the highest velocity tracking score, success rate, and traversal ratio. Moreover, the relatively small standard deviations indicate that the proposed policy maintains consistent performance across different evaluations.

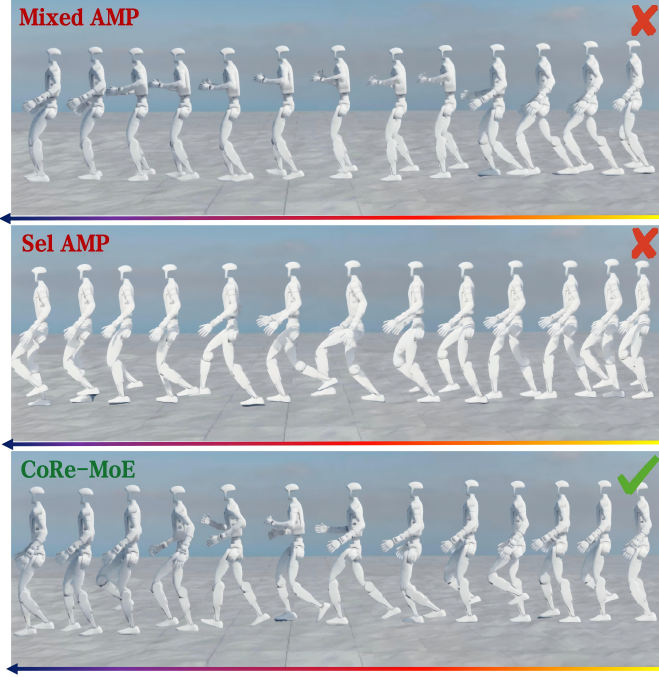


Fig. 4: Side-view gait sequence visualization results. CoRe-MoE exhibits the most natural gait.

The side-view gait sequences in Figure 4 further reveal the differences among the policies. For Mixed AMP, merging walking and running data into a single network causes it to learn an intermediate shuffling gait between walking and running, resulting in unnatural motion. Sel-AMP, although selecting the corresponding reference data based on the commanded velocity, often incorrectly increases stride length under high-speed commands, producing an unnatural over-stride behavior. This leads to unstable mixed gait characteristics and prevents the emergence of clear walking or running patterns. In addition, both baseline methods exhibit unnatural arm movements.

In contrast, CoRe-MoE employs separate AMP experts for walking and running and dynamically selects them based on the commanded velocity, enabling the policy to learn distinct walking and running gaits at low and high speeds. Moreover, modeling each gait with its own expert results in more natural and coherent arm swings, yielding significantly better gait continuity and adaptability compared to the baselines. We strongly recommend watching the demonstration videos on our project homepage for a more intuitive understanding.

TABLE V: Performance Comparison of Adaptive Gait Policy and Baseline Methods on Flat Ground

Method	Vel. Score.	Succ.	Dist.
Mixed AMP	0.796 ± 0.008	$96.12 \pm 0.37\%$	$97.60 \pm 0.29\%$
Sel AMP	0.812 ± 0.015	$54.20 \pm 1.84\%$	$54.50 \pm 1.67\%$
w/o MoE	0.817 ± 0.007	$97.06 \pm 0.31\%$	$98.41 \pm 0.24\%$
CoRe-MoE	0.853 ± 0.004	$99.13 \pm 0.15\%$	$99.85 \pm 0.08\%$

B. Evaluation of the Adaptive Multi-Terrain Policy

We compare the proposed Adaptive Multi-Terrain Policy with five baseline variants to analyze the contribution of each

key component: (1) **w/o MoE**: removing the MoE architecture; (2) **Masked Expert**: zeroing out the outputs of activated experts; (3) **w/o Fusion**: removing the second-stage policy branch and directly fine-tuning the Stage 1 policy on multi-terrain tasks; (4) **w/o SwAV**: removing SwAV-based contrastive learning; (5) **SimCLR**: using instance-level SimCLR contrastive learning instead of SwAV.

To evaluate the policy under complex terrain conditions, the robot is deployed on diverse terrains including stairs, slopes and obstacle. We retain the same distance-dependent velocity setting used in Stage 1 to create mixed walking–running scenarios. The key distinction is that this stage focuses on assessing stability under terrain-induced disturbances and determining whether the policy can leverage the second-stage policy branch and the contrastively trained MoE gating mechanism to generate terrain-aware corrective actions for robust traversal over challenging terrains.

Locomotion performance was quantified using multiple metrics: (1) **Success rate (Succ.)**: The percentage of trials in which the robot successfully reached the target point; (2) **Average Traversal Ratio (Dist.)**: The fraction of the target distance completed before reaching the goal or termination; (3) **Convergence time (Conv. Time)**: The number of training iterations required for the policy to reach stable performance; (4) **Discriminability of terrain features**: evaluated via t-SNE [41] visualization of MoE gating outputs to assess how well the embeddings differentiate between different terrain types.

As shown in Table VI and Figure 5, (1) w/o MoE exhibits a significant drop in success rate and traversal ratio across most terrains, indicating that MoE plays a central role in expert decomposition and policy specialization; (2) Masked Expert severely degrades performance by zeroing out activated expert outputs, further validating the importance of correct expert selection for terrain-aware control; (3) w/o Fusion shows reduced stability in complex multi-terrain scenarios and performs reasonably only on structurally simple terrains such as stairs, demonstrating that this fusion mechanism is essential for cross-terrain generalization; (4) w/o SwAV leads to lower success rate and slower convergence, suggesting that SwAV contrastive learning helps the MoE gating network better distinguish terrain features and accelerate policy optimization; (5) SimCLR performs better than w/o SwAV on most terrains but still underperforms CoRe-MoE overall, indicating that although instance-level contrastive learning improves representation quality, it lacks structured terrain semantic modeling capability and cannot form effective cluster-level alignment

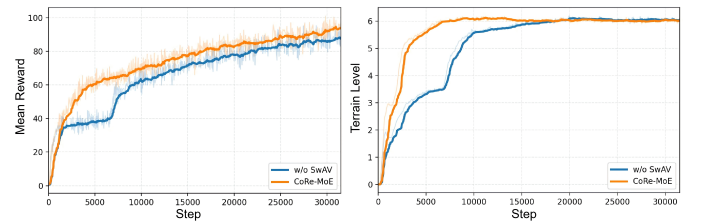


Fig. 5: Evolution of Mean Reward and Terrain Curriculum Level Across Training Iterations

TABLE VI: Ablation Study of CoRe-MoE with Multi-Terrain Locomotion Performance Analysis. All results are reported as mean values over evaluation episodes. Standard deviations are omitted for clarity.

Method	Flat Ground		Up Slope		Down Slope		Up Stairs		Down Stairs		High Up Stairs		High Down Stairs		High Obstacle		Random Obstacle	
	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.	Succ.	Dist.
w/o MoE	97.6%	99.5%	76.4%	82.6%	79.1%	86.1%	35.6%	54.7%	40.7%	63.2%	12.2%	30.1%	16.7%	38.9%	23.5%	41.3%	29.8%	46.2%
Masked Expert	19.2%	32.6%	4.9%	19.3%	15.6%	27.4%	2.1%	6.7%	9.4%	24.7%	0.2%	2.3%	4.3%	15.6%	7.6%	21.0%	3.5%	12.5%
w/o Fusion	98.7%	99.8%	81.2%	87.6%	86.3%	93.7%	36.3%	58.6%	51.8%	70.4%	9.4%	24.8%	17.5%	42.3%	26.1%	43.4%	35.4%	58.5%
w/o SwAV	99.5%	99.9%	96.4%	98.5%	98.2%	99.4%	84.6%	92.3%	86.3%	93.6%	72.3%	85.1%	79.5%	87.2%	90.2%	96.9%	81.5%	89.6%
SimCLR	99.7%	99.9%	97.3%	99.1%	98.6%	99.7%	87.6%	94.3%	89.4%	94.2%	76.8%	89.3%	81.6%	90.3%	93.4%	97.5%	83.6%	91.4%
CoRe-MoE	99.9%	99.9%	98.7%	99.4%	98.9%	99.9%	94.3%	96.8%	96.5%	97.9%	86.4%	91.6%	87.1%	93.7%	97.6%	98.7%	89.3%	92.8%

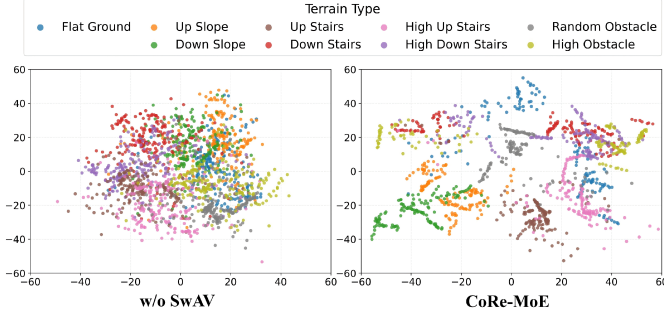


Fig. 6: T-SNE visualization of MoE gating features across different terrains. The left figure (w/o SwAV) shows mixed and overlapping feature distributions, while the right figure (CoRe-MoE) presents more compact and clearly separated clusters.

between terrains and experts, thereby limiting its effectiveness in MoE expert selection.

In contrast, CoRe-MoE consistently achieves the highest success rates and traversal ratios across all evaluated terrains, demonstrating superior robustness and generalization capability. Nevertheless, a small number of failures still occur in the most challenging scenarios, particularly on High Up Stairs and Random Obstacle terrains. We attribute these failures primarily to the abrupt terrain height variations and limited foothold margins in highly irregular environments, where small perception or foot-placement errors can accumulate and eventually lead to loss of balance.

Figure 6 presents t-SNE visualizations of MoE gating outputs across different terrain types. We specifically conducted this experiment to verify the effect of SwAV contrastive learning on terrain feature encoding. The embeddings with SwAV show clearly separated clusters for each terrain type, while embeddings without SwAV exhibit more overlap, indicating reduced discriminability. These results demonstrate that contrastive learning enables the gating network to more effectively distinguish terrain features and activate the corresponding experts for terrain-aware motion adjustments.

Overall, the proposed CoRe-MoE demonstrates superior robustness and stability in multi-terrain locomotion tasks, and achieves more reliable locomotion performance under complex terrain conditions.

C. Comparison with Prior Locomotion Methods

To further evaluate the proposed method, we compared CoRe-MoE with representative humanoid locomotion frameworks, including Hiking in the Wild [3] and MoRE [11]. For concise presentation, the reported stair, slope, and obstacle success rates are averaged over the corresponding terrain

variants, including both ascending and descending cases as well as different difficulty levels.

As shown in Table VII, Hiking in the Wild trains separate walking and running policies and requires manual switching during deployment, while MoRE employs a single policy but still relies on manually specified gait motion modes. Such approaches struggle to continuously adapt locomotion behaviors according to motion demands, limiting both adaptability and energy efficiency across a wide range of velocities. For example, using only a walking policy becomes inefficient at high speeds, whereas relying solely on running introduces unnecessary energy expenditure at low speeds. Moreover, manually specified motion modes cannot always guarantee optimal gait selection under varying terrain and velocity conditions.

In contrast, CoRe-MoE achieves unified walking and running control within a single policy and automatically performs gait transitions according to commanded velocities. In addition to the simplified policy structure, CoRe-MoE achieves the best overall locomotion performance, obtaining the highest velocity tracking score and success rates across multiple terrain categories. Furthermore, CoRe-MoE achieves the lowest Cost of Transport (CoT), indicating superior energy efficiency during locomotion. This advantage stems from the automatic walk run transition mechanism, which enables the policy to employ energy-efficient walking at low speeds, switch to more suitable running behaviors at high speeds, and smoothly transition between the two regimes in intermediate velocity ranges. As a result, CoRe-MoE avoids the efficiency loss associated with relying on a single locomotion mode across all velocities. These results demonstrate that the proposed framework effectively combines unified gait generation, automatic gait transition, energy-efficient locomotion, and robust multi-terrain traversal within a single policy architecture.

D. Real-world Experiment

To validate the generalization capability of the proposed method CoRe-MoE on real robotic hardware, we directly deployed the trained policy onto the Unitree G1 humanoid robot without any additional fine-tuning. The system uses the onboard Intel RealSense D435i depth camera to capture real-time depth images, which are processed through lightweight geometric operations and fed into the policy network, enabling zero-shot transfer from simulation to the real robot.

All experiments are conducted in real outdoor environments, including uneven ground, natural steps, outdoor obstacles, natural slopes, and terrains with height discontinuities. Despite significant discrepancies in geometry, surface materials, and

TABLE VII: Comparison of Policy Structures and Multi-Terrain Locomotion Performance Across Methods.

Method	Unified Policy	Transition Mode	Vel. Score.	Flat Succ.	Slope Succ.	Stairs Succ.	Obstacle Succ.	CoT
Hiking in the Wild (Walk)	✗	Manual	0.85 ± 0.02	$99.7 \pm 0.2\%$	$76.8 \pm 0.8\%$	$82.6 \pm 0.7\%$	$84.9 \pm 0.6\%$	0.75 ± 0.02
Hiking in the Wild (Run)	✗	Manual	0.62 ± 0.03	$99.4 \pm 0.3\%$	$67.9 \pm 1.1\%$	$54.7 \pm 1.4\%$	$78.5 \pm 0.9\%$	0.67 ± 0.02
MoRE	✓	Manual	0.79 ± 0.02	$97.6 \pm 0.4\%$	$93.7 \pm 0.5\%$	$90.3 \pm 0.6\%$	$91.2 \pm 0.5\%$	0.82 ± 0.03
CoRe-MoE	✓	Automatic	0.89 ± 0.01	$99.9 \pm 0.1\%$	$98.8 \pm 0.3\%$	$91.1 \pm 0.4\%$	$93.5 \pm 0.3\%$	0.61 ± 0.01

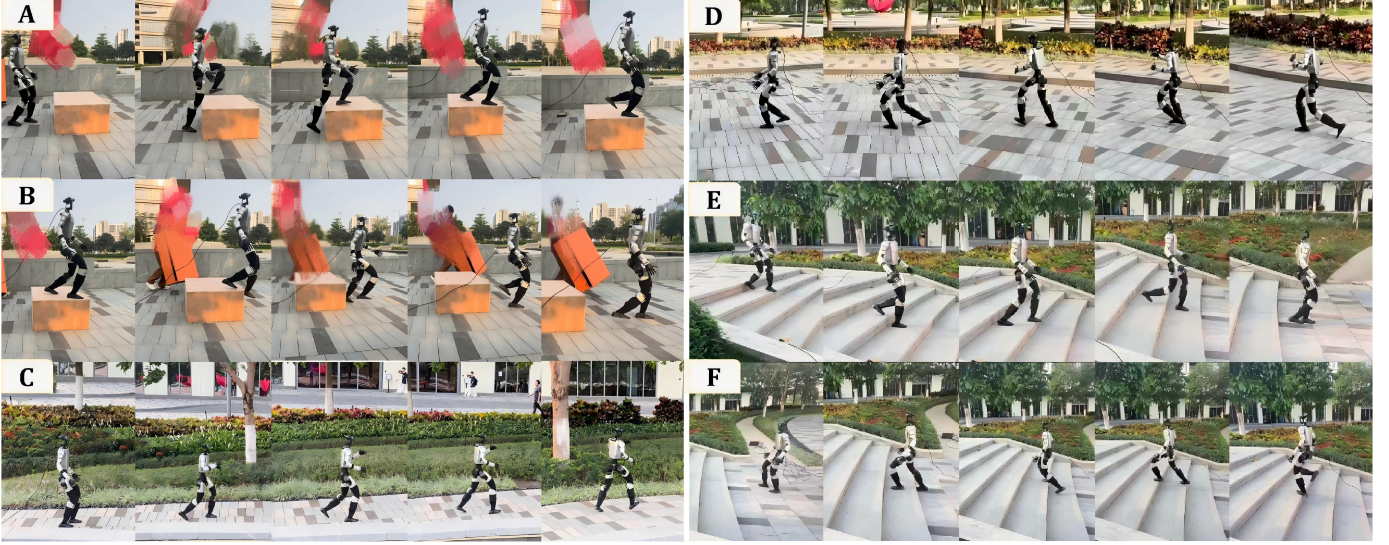


Fig. 7: The multi-terrain sequence demonstrates the robot’s continuous locomotion across six representative terrains, where A-F correspond to obstacle ascent, obstacle descent, slope ascent, flat ground, stair descent, and stair ascent. The robot smoothly transitions between different terrains while maintaining stable gait and reliable foothold control, demonstrating strong multi-terrain adaptability.



Fig. 8: Side-view gait sequence of the humanoid robot in real world deployment, showing a continuous transition from walking to running and back to walking without any mode switching.

lighting conditions compared to simulation, the robot consistently maintains stable locomotion with accurate foothold placement.

As shown in Figure 8, the robot executes a continuous sequence consisting of walking, running, and returning to walking on flat ground. This experiment highlights the natural and coherent behavior of CoRe-MoE under varying velocity commands. Since the framework is trained as a single unified policy, the robot automatically adjusts its gait as speed increases or decreases, without relying on any manually triggered mode-switching mechanism. Even during high-speed motions, where the maximum outdoor forward velocity reaches 2.5 m/s, the robot exhibits stable and coordinated dynamics with no noticeable discontinuities.

To further illustrate the advantage of unified gait control, we provide a comparison with the Hiking in the Wild baseline, as shown in Figure 1. The baseline trains walking and running policies separately, and switching between them must be performed manually during execution. Such manual switching

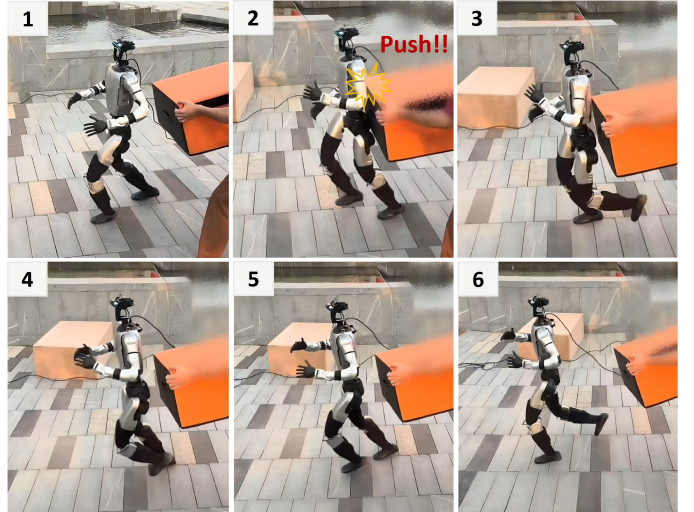


Fig. 9: The robot is subjected to a sudden external impact while running, maintaining balance and quickly recovering stable locomotion.

often introduces brief discontinuities or posture inconsistencies at the transition points, resulting in a visibly fragmented motion sequence. In contrast, CoRe-MoE preserves natural, smooth, and dynamically consistent gait transitions throughout the entire sequence, including walking, running, returning to walking. The unified policy adaptively regulates stride length, center-of-mass motion, and whole-body coordination, demonstrating superior continuity compared with methods that depend on switching between multiple independent controllers.

Furthermore, in multi-terrain traversal tasks, the robot ex-

hibits strong adaptability across different terrain types. The temporal sequence in Figure 7 illustrates the robot traversing uneven ground, natural steps, and slopes in a continuous manner. Throughout the process, the policy dynamically adjusts step length, foothold placement, and body posture in response to terrain variations, while maintaining stability under height discontinuities and varying friction conditions. These results demonstrate that the proposed method enables robust and coherent locomotion across diverse and challenging terrain compositions.

To further evaluate the system’s robustness in real-world conditions, we conducted an impact disturbance test. During level-ground locomotion, the robot was subjected to a sudden external impact while running, as shown in Figure 9. Despite the disturbance occurring under high-speed motion, the robot maintained balance and quickly recovered stable locomotion. This experiment demonstrates that the proposed method provides strong stability and recovery capability under external perturbations during dynamic gait execution.

VI. CONCLUSIONS

In this work, we presented a humanoid locomotion framework CoRe-MoE that integrates visual perception, weighted action refinement, and contrastively enhanced mixture-of-experts. The proposed method enables terrain-aware corrections and improves expert specialization through SwAV, allowing smooth and adaptive transitions between walking and running under varying velocity commands. Extensive simulations show that our approach outperforms baseline policies in success rate, travel distance, and multi-terrain adaptability. The trained policy is deployed onto a Unitree G1 robot without any fine-tuning, achieving stable zero-shot sim-to-real transfer using only the onboard depth camera. In both indoor and outdoor experiments, the robot demonstrates strong robustness, natural gait transitions, and resilience to external disturbances. Future work will extend this framework to whole-body coordination for more dynamic and human-like locomotion behaviors.

REFERENCES

- [1] Shixin Luo, Songbo Li, Ruiqi Yu, Zhicheng Wang, Jun Wu, and Qiuguo Zhu. Pie: Parkour with implicit-explicit learning framework for legged robots. *IEEE Robotics and Automation Letters*, 2024.
- [2] I Made Aswin Nahrendra, Byeongho Yu, Minho Oh, Dongkyu Lee, Seunghyun Lee, Hyeonwoo Lee, Hyungtae Lim, and Hyun Myung. Dreamwaq++: Obstacle-aware quadrupedal locomotion with resilient multimodal reinforcement learning. *IEEE Transactions on Robotics*, 42:819–836, 2026.
- [3] Shaoting Zhu, Ziwen Zhuang, Mengjie Zhao, Kun-Ying Lee, and Hang Zhao. Hiking in the wild: A scalable perceptive parkour framework for humanoids. *arXiv preprint arXiv:2601.07718*, 2026.
- [4] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024.
- [5] Wenhao Cui, Shengtao Li, Huaxing Huang, Bangyu Qin, Tianchu Zhang, Liang Zheng, Ziyang Tang, Chenxu Hu, NING Yan, Jiahao Chen, et al. Adapting humanoid locomotion over challenging terrain via two-phase training. In *8th Annual Conference on Robot Learning*, 2024.
- [6] Huayi Wang, Zirui Wang, Junli Ren, Qingwei Ben, Tao Huang, Weinan Zhang, and Jiangmiao Pang. Beamdojo: Learning agile humanoid locomotion on sparse footholds. 2025.
- [7] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model, 2024.
- [8] Junfeng Long, ZiRui Wang, Quanyi Li, Liu Cao, Jiawei Gao, and Jiangmiao Pang. Hybrid internal model: Learning agile legged locomotion with simulated robot response. In *The Twelfth International Conference on Learning Representations*, 2024.
- [9] Dimple Sethi, Sourabh Bharti, and Chandra Prakash. A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work. *Artificial Intelligence in Medicine*, 129:102314, 2022.
- [10] E. Y. S. Chao. *Biomechanics of the Human Gait*, pages 225–244. Springer New York, New York, NY, 1986.
- [11] Dewei Wang, Xinmiao Wang, Xinzhe Liu, Jiyuan Shi, Yingnan Zhao, Chenjia Bai, and Xuelong Li. More: Mixture of residual experts for humanoid lifelike gaits learning on complex terrains, 2025.
- [12] Xinyang Gu, Yen-Jen Wang, and Jianyu Chen. Humanoid-gym: Reinforcement learning for humanoid robot with zero-shot sim2real transfer. 2024.
- [13] Jonah Siekmann, Kevin Green, John Warila, Alan Fern, and Jonathan Hurst. Blind bipedal stair traversal via sim-to-real reinforcement learning. 2021.
- [14] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024.
- [15] Ilija Radosavovic, Sarthak Kamat, Trevor Darrell, and Jitendra Malik. Learning humanoid locomotion over challenging terrain. 2024.
- [16] Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.
- [17] Runhan Huang, Shaoting Zhu, Yilun Du, and Hang Zhao. Moe-loco: Mixture of experts for multitask locomotion. 2025.
- [18] Ziwen Zhuang, Shaoting Zhu, Mengjie Zhao, and Hang Zhao. Deep whole-body parkour, 2026.
- [19] Mathilde Caron, Ishan Misra, Julien Mairal, Piotr Goyal, Piotr Bojanowski, and Armand Joulin. Swav: Unsupervised learning of visual features by contrasting cluster assignments. 2020.
- [20] Wandong Sun, Yongbo Su, Leoric Huang, Alex Zhang, Dwyane Wei, Mu San, Daniel Tian, Ellie Cao, Finn Yan, Ethan Xie, and Zongwu Xie. Now you see that: Learning end-to-end humanoid locomotion from raw pixels, 2026.
- [21] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024.
- [22] Jonah Siekmann, Kevin Green, John Warila, Alan Fern, and Jonathan Hurst. Blind bipedal stair traversal via sim-to-real reinforcement learning. *arXiv preprint arXiv:2105.08328*, 2021.
- [23] Mingqi Yuan, Tao Yu, Haolin Song, Bo Li, Xin Jin, Hua Chen, and Wenjun Zeng. Pvp: Data-efficient humanoid robot learning with proprioceptive-privileged contrastive representations. *arXiv preprint arXiv:2512.13093*, 2025.
- [24] Manuel Yves Galliker. Whole-body humanoid mpc: Realtime physics-based procedural loco-manipulation planning and control. https://github.com/1x-technologies/wb_humanoid_mpc, 2024.
- [25] Nicola Scianca, Daniele De Simone, Leonardo Lanari, and Giuseppe Oriolo. Mpc for humanoid gait generation: Stability and feasibility. *IEEE Transactions on Robotics*, 36(4):1171–1188, 2020.
- [26] Sotaro Katayama, Masaki Murooka, and Yuichi Tazaki. Model predictive control of legged and humanoid robots: models and algorithms. *Advanced Robotics*, 37(5):298–315, 2023.
- [27] Haolin Song, Hongbo Zhu, Tao Yu, Yan Liu, Mingqi Yuan, Wengang Zhou, Hua Chen, and Houqiang Li. Gait-adaptive perceptive humanoid locomotion with real time under-base terrain reconstruction. 2025.
- [28] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11443–11450. IEEE, 2024.
- [29] Nikita Rudin, Junzhe He, Joshua Aurand, and Marco 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.
- [30] 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.

- [31] Qingwei Ben, Botian Xu, Kailin Li, Feiyu Jia, Wentao Zhang, Jingping Wang, Jingbo Wang, Dahua Lin, and Jiangmiao Pang. Gallant: Voxel grid-based humanoid locomotion and local-navigation across 3d constrained terrains. *arXiv preprint arXiv:2511.14625*, 2025.
- [32] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model. 2025.
- [33] Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. Humanoid parkour learning. *arXiv preprint arXiv:2406.10759*, 2024.
- [34] Michael I. Jordan and Robert A. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6(2):181–214, 1994.
- [35] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.
- [36] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [37] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (ToG)*, 40(4):1–20, 2021.
- [38] Mayank Mittal, Pascal Roth, James Tigue, et al. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025.
- [39] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5442–5451, 2019.
- [40] Félix G. Harvey, Mike Yurick, Derek Nowrouzezahrai, and Christopher Pal. Robust motion in-betweening. 39(4), 2020.
- [41] Matthew C. Cieslak, Ann M. Castelfranco, Vittoria Roncalli, Petra H. Lenz, and Daniel K. Hartline. t-distributed stochastic neighbor embedding (t-sne): A tool for eco-physiological transcriptomic analysis. *Marine Genomics*, 51:100723, 2020.