Let Humanoids Hike! Integrative Skill Development on Complex Trails

Kwan-Yee Lin
University of Michigan

Stella X. Yu
University of Michigan

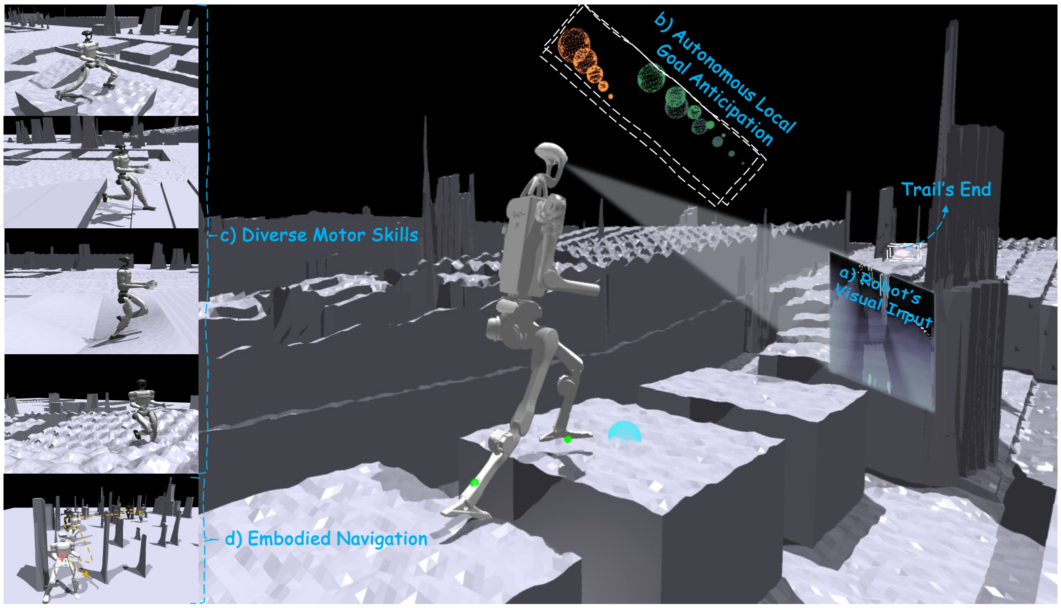

Figure 1: **We propose training humanoids to hike complex trails, driving integrative skill development across visual perception, decision-making, and motor execution. Center:** The humanoid robot (H1) **a)** equipped with vision, learns to **b)** anticipate near-future local goals to guide locomotion along the trail with self-autonomy. Bubble size (large → small) indicates anticipated goal direction; color shows temporal order (orange → green → forest). **Left:** Our LEGO-H framework is universal to different humanoid robots (*e.g.,* G1, a smaller robot) to adaptively **c)** emerge diverse motor skills, and **d)** develop embodied path exploration strategies to hike on trails with varied terrains and obstacles.

Abstract: Hiking on complex trails demands balance, agility, and adaptive decision-making over unpredictable terrain. Current humanoid research remains fragmented and inadequate for hiking: locomotion focuses on motor skills without long-term goals or situational awareness, while semantic navigation overlooks real-world embodiment and local terrain variability. We propose training humanoids to hike on complex trails, driving integrative skill development across visual perception, decision making, and motor execution.

We develop a learning framework, LEGO-H, that enables a vision-equipped humanoid robot to hike complex trails autonomously. We introduce two technical innovations: **1)** A temporal vision transformer variant - tailored into Hierarchical Reinforcement Learning framework - anticipates future local goals to guide movement, seamlessly integrating locomotion with goal-directed navigation. **2)** Latent representations of joint movement patterns, combined with hierarchical metric learning - enhance Privileged Learning scheme - enable smooth policy transfer from privileged training to onboard execution. These components allow LEGO-H to handle diverse physical and environmental challenges without relying on predefined motion patterns. Experiments across varied simulated trails and robot morphologies highlight LEGO-H's versatility and robustness, positioning hiking as a compelling testbed for embodied autonomy and LEGO-H as a baseline for future humanoid development.

RSS 2025 Workshop on Whole-Body Control and Bimanual Manipulation (RSS2025WCBM).

Keywords: Hiking, Benchmarks, Humanoid Robots, Sensing and Perception

1 Introduction

Hiking [1, 2] challenges humans to master diverse motor skills and adapt to complex, and unpredictable terrain – such as steep slopes, wide ditches, tangled roots, and sudden elevation changes (Fig. 1). It demands continuous balance, agility, and real-time decision-making, making it an ideal testbed for advancing humanoid autonomy and the integration of vision, planning, and motor control. Hiking-capable robots could explore remote areas, assist in rescue missions, and guide individuals along rugged paths.

Hiking poses challenges beyond traditional navigation, blind locomotion, or single motor pattern learning. To succeed, humanoid robots must master three core capabilities: **1) Locomotion versatility** – The ability to handle mixed terrains like dirt, rocks, stairs, and streams, adapting dynamically with skills like jumping and leaping while maintaining balance. **2) Perceptual awareness** - The ability to sense and respond to complex 3D environments, such as stepping over logs or navigating around trees. **3) Body awareness** – The ability to adjust in real time to local obstacles, terrain changes, and body states by coordinating vision and motor control for adaptive foot placement and movement.

Current humanoids struggle to meet these demands due to the lack of a unified framework that integrates low-level motor skills with high-level navigation (. **1) Locomotion methods lack adaptability to terrain variation**. They treat terrain as a fixed, homogeneous, and passive background, focusing narrowly on walking [3, 4], quasi-periodic motion patterns [5], or mimicry [6]. Advanced frameworks for complex skills like parkour [7, 8], often depend heavily on user commands or engineered behaviors. Such isolated training paradigms and abstraction overlook the embodied interaction essential for real-world locomotion, limiting generalization beyond curated environments. **2) Navigation methods struggle with real-time adaptability.** Traditional research efforts rely on scene mapping [9] or rigid world geometry [10]. While LLMs and VLMs can plan behaviors and correct execution failures from textual instructions [11], they often lack the physical grounding needed for real-world adaptability. A robot may know it needs to *step over the log*, but without real-time perception and fine-grained motor control, it cannot adjust mid-swing if the log shifts or the ground gives way. Reflexive foot placement on uneven terrain demands fast, sensor-driven adaptation - not just faster planning - which symbolic planners struggle to provide. Bridging motor skills and navigation remains challenging due to their inherently different response levels (fast, reactive control vs. slower, deliberative planning) requiring tight coordination for context-sensitive execution in complex environments.

We introduce LEGO-H, a perceptual-aware, end-to-end learning framework for acquiring situational visual-motor skills and path exploration strategies that enable humanoids to traverse complex trails autonomously (Fig. 1). It unifies navigation and locomotion by advancing Hierarchical Reinforcement Learning (HRL) and enhancing Privileged Learning (PL) for effective skill development.

Our first technical contribution is task-grounded HRL for situational visual-motor control, reformulating navigation as a sequential local goal anticipation problem to guide locomotion policy learning. While HRL can unify navigation and locomotion via multi-level abstraction, existing methods often oversimplify environments [12], or restrict low-level control to basic skills like walking [13], limiting adaptability. We address this caveat by proposing TC-ViT, a temporal vision transformer variant tailored for HRL that combines tokenization with embodied reinforcement learning. Instead of treating the navigation target as a static token, TC-ViT models **1)** navigation goals and **2)** temporal-spatial relations, considering the robot's past, present, and future states for sequential anticipation. The locomotion policy network then integrates these latent features with proprioceptive inputs and partial anticipated navigation goals to produce motor actions, enabling tight coordination between perception and control for navigating complex, dynamic trails.

Our second technical contribution is enhanced PL that distills diverse motor skills while preserving action rationality. In PL, a teacher policy leverages privileged signals such as known foothold locations to develop diverse, optimal behaviors efficiently and safely. A student policy then learns to replicate these behaviors using only proprioception and onboard perception, enabling deployment in unstructured environments without privileged information. It improves skill acquisition but complicates action learning when integrating visual inputs, increasing the risk of errors and damage from unexpected actions. Existing distillation approaches supervise global behaviors [14] or per-joint accuracy [15], often ignoring inter-joint dependencies. We address this by proposing a Hierarchical Latent Matching (HLM) metric that distills policy based on action rationality. HLM utilizes structured latent representations and masked reconstruction via VAEs [16] to enforce relational consistency across joints. This task-agnostic HLM loss set improves policy learning across motor tasks. Crucially, the latent prior is derived from oracle policy, *not* human demonstrations, allowing robot to learn self-reliant behaviors suited to its own morphology.

To summarize, our work makes three key contributions: **1)** We propose hiking as a testbed for integrative skill development in humanoid robots. **2)** We introduce LEGO-H, a learning framework for autonomous humanoid hiking. **3)** We demonstrate LEGO-H's robustness and versatility across diverse simulated trails and humanoid morphologies, establishing hiking as a compelling testbed for embodied autonomy and LEGO-H as a baseline for future humanoid research.

2 Related Work

Humanoid locomotion. Existing approaches to low-level motor skill learning typically simplify environmental interactions, abstracting terrains into static patterns at a *momentary* scale, which neglects occlusions caused by obstacles or dynamic environmental disruptions. Research in this domain has primarily focused on learning specific locomotion skills such as walking [3, 4, 17, 18, 19], running [20, 21], and soccer-playing behaviors [22]. These approaches often rely on highly engineered designs optimized for specific lower-body tasks. Other works employ imitation learning [5, 6, 23, 24, 25] to generate human-like behaviors from large-scale motion datasets, but this comes at the cost of reduced embodiment. Some frameworks attempt to push the boundaries of robotic motor skills by exploring tasks like parkour [8, 7], acrobatic flipping [26], or cliffside climbing [27]. While impressive, these methods are often bogged down by complex engineering, reliance on user commands for motion planning, or lack of perceptual awareness.

Humanoid navigation. Research on this direction often struggles to address *real-time* environmental constraints while accounting for the unique mechanisms and actions of humanoid robots. These limitations frequently lead to suboptimal navigation plans in complex terrains. Conventional methods typically rely on scene mapping [9, 28] or structured world assumptions [10], which restrict adaptability in dynamic and unstructured environments. Contact-aware approaches [29, 30] attempt to bridge robot configurations with environmental constraints, but they often depend on pre-generated trajectories, limiting responsiveness. Similarly, mapless methods [31] leverage visual inputs for navigation but are typically constrained to basic locomotion capabilities such as walking. Recent advancements in large language and vision-language models have shown potential for complex high-level planning [11], yet remain uncoupled from motor control systems, failing to achieve autonomous perceptual awareness and last-step feasibility required for navigating diverse, fine-grained environments, like hiking.

Joint learning of navigation and locomotion. Integrating navigation and locomotion into a unified framework remains a significant challenge. In the realm of wheeled-legged and quadruped robots, several studies [32, 33, 34, 35] have explored paradigms that unify local navigation and locomotion. While these approaches provide valuable insights, tailoring them to humanoid robots as a baseline for hiking tasks reveals several critical gaps. First, humanoid robots possess significantly more degrees of freedom (DoF) than quadrupeds or wheeled-legged robots, complicating the development of stable locomotion policies. Achieving balance across diverse lower-body motor skills (*e.g.,* walking, jumping, and leaping *etc.*) within a single framework remains an open problem. Second, the greater body height of humanoid robots introduces challenges in visual perception, expanding their field of view and capturing a broader range of distances. This increased perceptual complexity exacerbates the misalignment between environmental sensing and physical contact, further complicating decision-making, navigation, and motor execution processes.

3 Problem Formulation

Drawing from human hiking paradigm [36], we consider a humanoid robot equipped with vision and GPS. A hiking trail is specified by start and end points (P_A, P_B) in GPS, optionally with M intermediate waypoints along the trail. We define the basic task of *humanoid hiking* as follows: *traversing a trail to reach the trail's end P_B with safety, efficiency, and all-level autonomy*.

The robot receives the following inputs: **1)** GPS-based 2D vector D_{rb} from robot's current projected 2D root position $P_R[:2]$ to end $P_B[:2]$, which may not be visible from start P_A. This vector provides the distance and direction of the endpoint relative to the robot. **2)** GPS-based 2D vectors $\{D_{rm}\}_{m=1}^{M}$ from P_R to M optional intermediate waypoints. We use $M=1$ to study the basic trail structure and disambiguate forks. These points provide guidance but need not be strictly followed. **3)** The onboard proprioceptive input $\mathcal{X}_{pro}$, like joint velocities and angles, reflects the robot's internal physical state. **4)** K forward-facing depth images $\{C_k\}_{k=1}^{K}$ from a head-mounted camera. Unlike prior quadruped approaches [35, 32] assuming full local 3D information, our setup limits vision to a frontal field, making perceptual-motor learning more realistic and challenging. Humanoids, being taller, see farther - enabling look-ahead planning but complicating near-term action learning.

For ideal hiking, whole-body control would allow coordinated use of arms and legs to maintain balance and support denser contact points with trails. However, as a baseline prototype for this new task – and noting that many trails can still be traversed with leg movement alone – this study simplifies the task by freezing humanoid's upper-body pose, focusing on lower-body functionality.

4 LEGO-H for Integrative Skill Learning

4.1 LEGO-H System Overview

In our setup, the robot is only given the relative position of the endpoint. Thus, it must *autonomously* determine how to traverse unknown, but locally observable trail with various terrain changes to reach the destination *safely*. From a framework perspective, a humanoid system must fulfill two core requisites to succeed: **1) learn embodied path exploration that is both target-driven and locally adaptive** – the robot must autonomously assess and adapt its local path based on immediate sensory observations and current executable motor skills, while maintaining alignment with the overall goal; **2) enable emergent, context-aware, and safe motor execution** – the robot must learn a diverse set of motor skills and execute actions that are not only safe for its body but also feasible under local environmental constraints, like clearance and terrain support. To this end, we propose an end-to-end, embodied learning framework, LEGO-H (Fig. 2), short for *Let Humanoids Go Hiking*.

To fulfill the first requisite, LEGO-H employs two levels of modules within a unified policy learning pipeline (Fig. 2b), combing a high-level navigation module ($\mathcal{H}$) that encodes trail's latent representation and anticipates local goals, with a low-level motor skill module ($\mathcal{E}$) that learns reactive motor policy in real time. Specifically: **1)** The high-level navigation module $\mathcal{H}$, implemented via TC-ViT (Sec. 4.2), acts as a *trail scout*, looking ahead and proposing local directions based on visual cues, global goal, and motor execution. It receives the state $\mathbf{s}_{real}$ (depth images $\{C_k\}_{k=1}^{K}$, proprioception $\mathcal{X}_{pro}$, endpoint P_B, and one middle waypoint M), generates a latent trail representation $\mathbf{z}_{uni}$, anticipates a sequence of N *future* local navigation goals $G=\{g_n\}_{n=1}^{N}$, and calculates a goal residual δg_0 capturing the execution mismatch from the *previous step*. Each $g_n \in [0, 2\pi]$ represents a goal direction as a yaw angle relative to the robot's root. **2)** Then, the latent trail representation $\mathbf{z}_{uni}$, proprioception $\mathcal{X}_{pro}$, residual δ_{g_0}, and the *next* anticipated goal g_1, flow to the low-level motor skill module $\mathcal{E}$ to guide *softly*. $\mathcal{E}$ plays the role of an agile *trail runner*, reacting in real time to proprioceptive feedback and terrain conditions to decide how best to execute each step. It predicts an

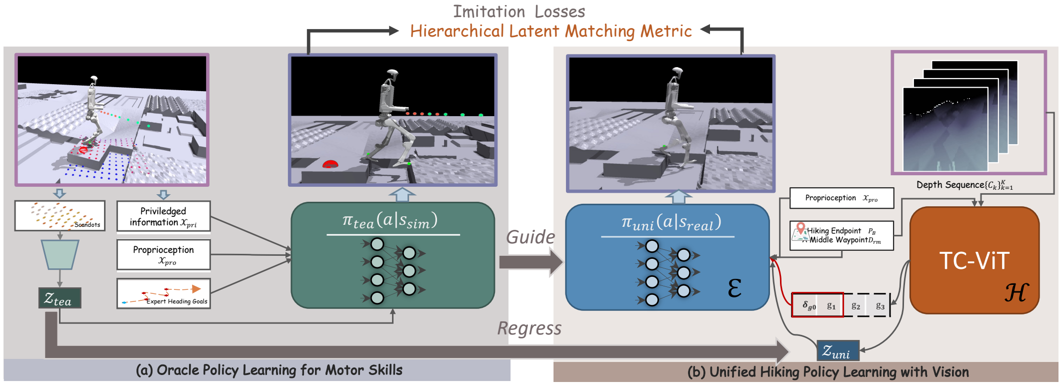

Figure 2: **LEGO-H framework overview.** LEGO-H equips humanoid robots with adaptive hiking skills by integrating navigation $\mathcal{H}$ and locomotion $\mathcal{E}$ in a unified, end-to-end learning framework (b). To foster the versatility of motor skills, we train the unified policy via privileged learning from the oracle policy (a).

executable action $\mathbf{a}_t$. Rather than strictly tracking the sequence of local goals from $\mathcal{H}$, $\mathcal{E}$ adapts to local terrain and robot state to safely progress toward the endpoint.

By seamlessly leveraging visual and proprioceptive feedback within an RL framework, this unified pipeline reflects HRL's abstraction, where local goal anticipation and reactive control jointly enable the robot to autonomously adapt local paths within traversable regions, avoiding entrapment and collisions in challenging trail terrains, while maintaining steady progress toward the trail's end.

LEGO-H achieves the second requisite by enhancing privileged learning scheme wrt structural rationality of actions: **1)** It first trains an oracle motor skill policy $\pi_{tea}(\mathbf{a}|\mathbf{s}_{sim})$ (Sec. 4.3) with privileged information $\mathcal{X}_{pri}$ (*e.g.,* terrain type, ground friction, precise state measurements) and expert navigation goals as inputs (Fig. 2a). While vision is not used at this stage, scandots and $\mathcal{X}_{pri}$ provide clean, informative signals for high-quality skill acquisition. **2)** Then, in the unified pipeline training, the teacher policy is distilled into $\mathcal{E}$ to initialize it. Aside from basic imitation losses and rewards (Sec. 4.4), LEGO-H uses a Hierarchical Latent Matching metric (Sec. 4.5) to learn the final policy $\pi_{uni}(\mathbf{a}|\mathbf{s}_{real})$ that balances robustness and behavioral diversity across diverse trail terrains.

4.2 TC-ViT: Autonomous Local Goal Anticipation

The navigation module $\mathcal{H}$ is implemented via TC-ViT, a variant of Temporal Information Conditioned Vision Transformer. It serves as a central mechanism to achieve unified policy learning with visual perception, by addressing four critical aspects to navigation module: 1) cognize surroundings with balance of short-time reactivity and final goal alignment, adapt anticipation of local goals to local terrain with 2) spatial precision and 3) embodied awareness, and 4) produces representations with synchronized perception and action (shown in Fig. 3).

1) Cognize surroundings with final goal. A common strategy for environment perception assumes Markovian observations and processes adjacent depth images via methods like 3D modeling [32]/reconstruction [37], temporal features [38], or semantic traversability [39]. However, hiking poses two key challenges: **1)** *Time scale*: short-term dynamics and long-term environmental dependencies must be handled jointly. **2)** *Specificity*: Visual features must directly support execution of immediate next step while aligning with final goal.

Thus, a direct solution is to integrate local perception with a distant global goal P_B, where we employ a temporal vision transformer with goal conditioning (Fig.3a), adapted from classic ViViT's encoder[40]. It captures the information with both spatial and long-range dependencies via processing 16-frame depth sequences (downsampled to 4) into spatio-temporal tokens (from 16×16 patches) using 6 transformer layers with spatial and temporal attention. The final goal P_B is tiled as an additional $(1,H,W)$ channel ($H = W = 128$) and fused at tokenization. This early fusion ensures goal awareness is preserved throughout spatio-temporal reasoning, yielding more task-aligned predictions.

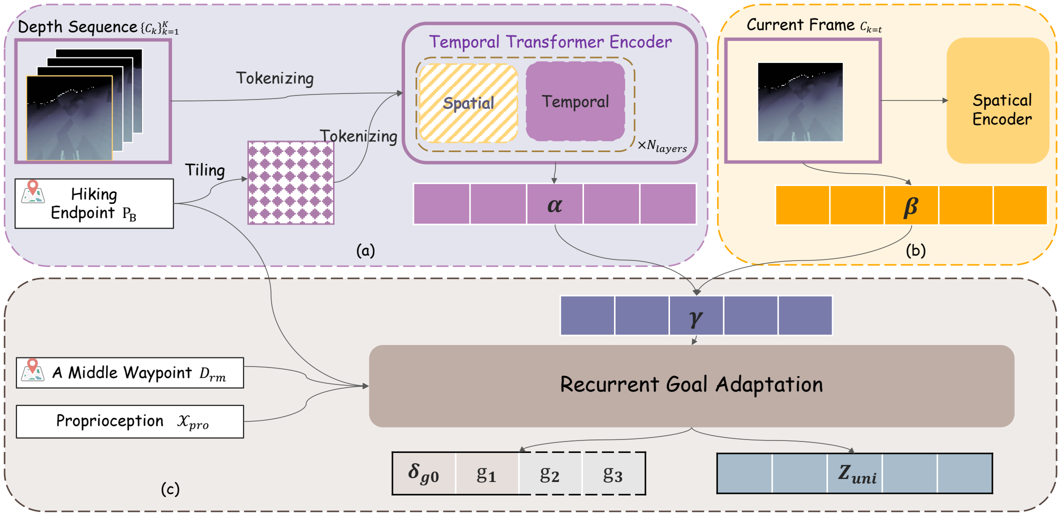

Figure 3: **TC-ViT Architecture.** Three key components: **a)** a goal-orientated temporal transformer encoder for robots cognizing surroundings with the final goal; **b)** a parallel process on the current depth frame for integrating spatially precise information to reflect the current state **c)** a recurrent goal adaptation mechanism that integrates visual awareness, goal information, and proprioception.

The encoder outputs a flattened feature vector $\boldsymbol{\alpha}(\{C_k\}_{k=1}^K, P_B)$. Intuitively, this part of TC-ViT serves as a trail scout with a map in hand: it interprets what's immediately ahead through sequences of depth images, while constantly factoring in the direction of the final destination. Embedding the goal early - before visual abstraction — ensures the robot always "looks" with intent, allowing it to anticipate terrain-compatible moves that remain globally purposeful.

2) Anticipate near-future goals with spatial precision. While aboves might be effective to support long-horizon goal prediction in coarse, body-agnostic navigation [41], humanoid hiking demands fine-grained, multi-scale decision-making. On uneven trails with sudden obstacles (Fig. 1), precise foot placement and rapid balance adjustments are critical - capabilities that suffer as temporal transformers abstract away fine spatial structure critical for precise control.

The second component of TC-ViT (Fig. 3b) thus introduces a parallel path focused on immediate perception. It processes the current depth image $C_{k=t}$ through a shallow CNN, producing high-resolution spatial features $\boldsymbol{\beta}(C_{k=t})$ that capture near-field terrain details. This branch omits goal conditioning, as its role is purely reactive.

The final representation $\boldsymbol{\gamma}$ combines long-range goal-informed context $\boldsymbol{\alpha}$ with fine-grained local perception $\boldsymbol{\beta}$ via feature concatenation followed by MLPs: $\boldsymbol{\gamma} = \text{MLPs}(\text{concat}(\boldsymbol{\alpha}, \boldsymbol{\beta}))$. Intuitively, this merges the foresight of a trail guide - who knows where the path leads - with the reflexes of a hiker watching their next step.

3) Adapt goals with embodied awareness. Beyond understanding environment, effective navigation must also account for how motor actions and body state affect outcomes. TC-ViT includes a third part - a recurrent goal adaptation mechanism (Fig. 3c) - that fuses visual features, proprioception, and goal information to adaptively anticipate a sequence of local goals, and produce an embodied latent representation.

Specifically, inputs including the visual representation $\boldsymbol{\gamma}$, endpoint P_B, intermediate cue $\mathcal{D}_{rm}$, and proprioception $\mathcal{X}_{pro}$ are passed through a two-layer MLP and a GRU to model temporal dependencies: $\mathbf{z}_{uni}, \delta g_0, G = \text{GRU}(\text{MLPs}(\boldsymbol{\gamma}, P_B, \mathcal{D}_{rm}, \mathcal{X}_{pro}))$. The resulting latent encodes perceptual context, physical embodiment. The residual correction δg_0 and near-future goals G provide soft guidance for the locomotion module. Intuitively, this mechanism helps the robot learn not just what it sees, but how it moves through what it sees, adapting its local goals based on how past actions played out, and staying grounded in both vision and bodily awareness (as shown in Fig 4).

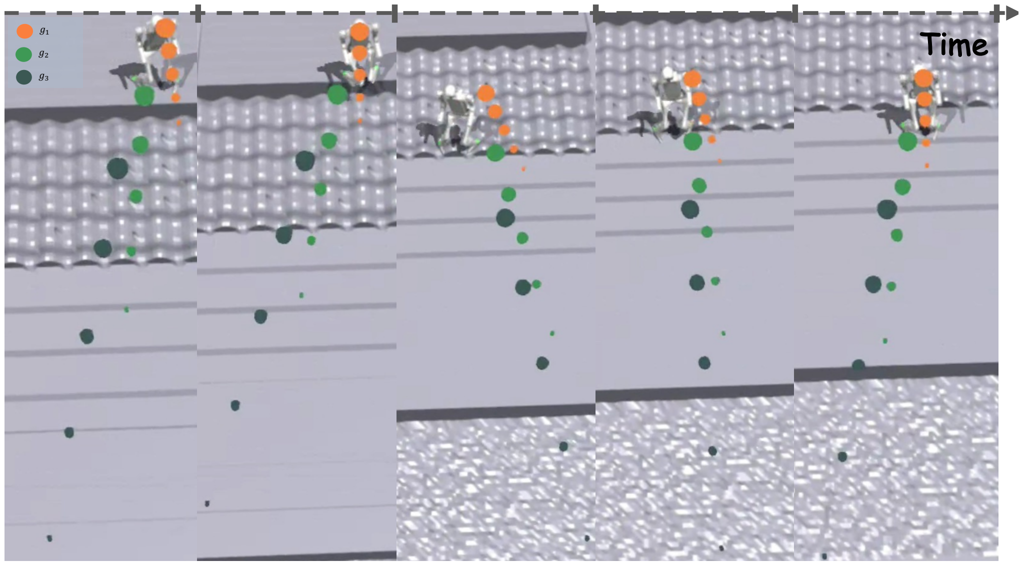

Figure 4: **Dynamic adjustments of near goal anticipation.** Snapshots from left to right show a robot traversing mixed terrains along a trail. TC-ViT does not provide a fixed trajectory that locomotion module must rigidly follow. Instead, it predicts several near-future goals (g_1, g_2, g_3), which dynamically adapt to robot's current state, reflecting real-time adjustments to its navigation decisions. Bubble size (large$\rightarrow$ small) represents predicted local navigation direction.

4) Synchronize perception and control. Real-world systems operate at mismatched time scales, e.g., Unitree H1's depth sensing runs at 10 ± 2 Hz with RealSense D435i, while control executes at 50 Hz on Jetson NX. TC-ViT addresses this latency gap with two strategies. *1. Nearest-goal forwarding:* Only the immediate goal g_1 is passed to the locomotion module, ensuring timely response and reducing drift from delayed decisions. Intuitively, this reflects the idea that – while multiple goals are anticipated, only the immediate one shapes action, as it reflects the step that matters right now. *2. Latent tiling:* The latent representation $\mathbf{z}_{uni}$ is tiled five times per control cycle to maintain a stable signal stream. Together, these mechanisms bridge asynchronous modules and allow perception and action to stay in sync despite hardware-level delays.

4.3 Oracle Policy Learning for Motor Skills

Before unifying navigation and locomotion via TC-ViT, we pretrain an *oracle* locomotion policy (Fig. 2a) to acquire diverse motor skills. The oracle takes as input proprioception $\mathcal{X}_{pro}$, current navigation goal, privileged state $\mathcal{X}_{pri}$, and latent terrain features $\mathbf{z}_{tea}$ from scandots $\mathcal{S} \in \mathbb{R}^{66\times 2}$. To encourage upright locomotion with emergent motor behavior rather than pre-defined modes, rewards in three aspects are essential in this stage: **1)** direction-aligned velocity tracking r_{tracking}, **2)** soft torso height constraint $r_{\text{base-height}}$, **3)** foot airtime accumulation $r_{\text{air-time}}$.

4.4 Unified Hiking Policy Learning with Vision

After training the oracle policy $\pi_{tea}(\mathbf{a}|\mathbf{s}_{sim})$, we distill it into a unified student policy $\pi_{uni}(\mathbf{a}|\mathbf{s}_{real})$ that jointly learns navigation and motor control from visual input (Fig. 2b). Specifically, TC-ViT encodes depth sequences into latent $\mathbf{z}_{uni}$ and predicts near-future goals. The tuple $(\mathbf{z}_{uni}, \delta_{g_0}, g_1)$ is passed to the locomotion module to compute $\pi_{uni}(\mathbf{a}|\mathbf{s}_{real})$, which outputs current action $\mathbf{a}_t$. Both policies are implemented as MLPs. Basic training losses here are RL rewards and reconstructions for imitation in goal, latent, and action levels from teacher stage:

$$\begin{aligned}\mathcal{L}_{im} = w_1\|\mathbf{z}_{tea} - \mathbf{z}_{uni}\|^2 + w_2\, \text{SmoothL1}(\mathbf{G}_{tea}, \mathbf{G}_{uni}) \\ + w_3\, \text{SmoothL1}(\mathbf{a}_{tea}, \mathbf{a}_{uni}).\end{aligned} \tag{1}$$

The oracle acts as a mentor guiding student through complex terrain. By initializing π_{uni} via imitation and optimizing it together with TC-ViT under RL framework, π_{uni} learns to align vision, planning, and control into a cohesive behavior.

4.5 Hierarchical Latent Matching Metric

Standard action imitation loss aggregates per-joint errors, overlooking joint coordination. Thus, we introduce Hierarchical Latent Matching (HLM) loss metric, which captures structural dependencies to bound the student's action space. We first train a masked VAE on oracle actions to learn a latent space that encodes joint coordination. During distillation, student policy is guided to match this latent structure, promoting physically coherent and well-coordinated actions despite modality and representation gaps. Analogous to feature matching in image reconstruction, this method shifts imitation from pointwise joint matching to holistic joint pattern matching, treating the body as a coordinated system rather than a set of independent joints.

Specifically, during distillation, VAE is *iteratively* trained on teacher actions with randomly masked joints, learning to reconstruct full actions from partial inputs, where:

$$\mathcal{L}_{rec} = w_4\mathcal{L}_{KL} + w_5\mathcal{L}_{self} + w_6\mathcal{L}_{mask} \tag{2}$$

$$\mathcal{L}_{KL} = KL\left(q(\mathbf{z_{vae}}|\mathbf{a}_{\text{tea}}) \parallel \mathcal{N}(0, I)\right) \tag{3}$$

$$\mathcal{L}_{self} = \text{SmoothL1}(\text{Dec}(\text{Enc}(\mathbf{a}_{\text{tea}})), \mathbf{a}_{tea}) \tag{4}$$

$$\mathcal{L}_{mask} = \text{SmoothL1}(\text{Dec}(\text{Enc}(\mathbf{a}_{\text{tmask}})), \mathbf{a}_{tea}) \tag{5}$$

Here, w_x are weighting terms, $\mathbf{z}_{vae}$ is latent vector, and $\mathbf{a}_{tmask}$ denotes masked teacher action. KL term follows VAE formulation [16]. To handle joint permutation invariance, we apply sine-cosine positional embeddings to each joint. The compact latent space, regularized by the Gaussian prior and enriched by masking, encourages learning of inter-joint dependencies and structural consistency, capturing coordination patterns aligned with the robot's physical embodiment, rather than relying on human motion priors.

Once trained, the encoder defines a structured feature space for comparing teacher and student actions. We utilize it to introduce a two-level HLM loss: full-feature alignment and masked-subset matching.

Concretely, for each student action $\mathbf{a}_{uni}$, we compute a cosine similarity loss with the teacher action:

$$\mathcal{L}_{ts} = 1 - \text{cos\_sim}(\text{Enc}(\mathbf{a}_{tea}), \text{Enc}(\mathbf{a}_{uni})) \tag{6}$$

$$= 1 - \frac{\text{Enc}(\mathbf{a}_{tea}) \cdot \text{Enc}(\mathbf{a}_{uni})}{\|\text{Enc}(\mathbf{a}_{tea})\|\|\text{Enc}(\mathbf{a}_{uni})\|} \tag{7}$$

We further apply a triplet-style consistency loss using a randomly masked student action:

$$\begin{aligned}\mathcal{L}_{trip} = c_{mt}&(1 - \text{cos\_sim}(\text{Enc}(\mathbf{a}_{tea}), \text{Enc}(\mathbf{a}_{umask}))) \\ + c_{ms}&(1 - \text{cos\_sim}(\text{Enc}(\mathbf{a}_{uni}), \text{Enc}(\mathbf{a}_{umask})))\end{aligned} \tag{8}$$

The combined hierarchical loss is:

$$\mathcal{L}hie = w_7\mathcal{L}ts + w_8\mathcal{L}_{trip} \tag{9}$$

As shown in Tab. 1, without HLM, student robots can complete the task but with frequent collisions and poor coordination. In contrast, HLM promotes robots to exhibit more refined, collision-free movements that align better with internal structural consistency.

5 Experiments

5.1 Experimental Settings

Robots. We use Unitree H1 [42] and G1 [43] humanoids, chosen for their distinct differences in body scale and mechanism: H1, at adult size (5.9 ft/47kg), contrasts with kid-sized G1 (4.26 ft/35kg), with

Table 1: **Ablation of LEGO-H'on H1.** for best goal completeness; for most safeness; for best efficiency.

Metrics	Oracle	LEGO-H	w TC-ViT	Vanilla
Success Rate (SR) (%) ↑	71.20 ± 0.72	68.40 ± 1.34	64.73 ± 2.22	42.97 ± 0.67
Trail Completion (TC) (%) ↑	77.73 ± 0.92	52.78 ± 1.30	52.50 ± 1.52	32.01 ± 0.61
Traverse Rate (TR) (%) ↑	73.60 ± 0.81	71.96 ± 2.37	72.04 ± 0.98	60.26 ± 0.94
MEV (%) ↓	7.12 ± 0.92	7.84 ± 0.92	10.40 ± 1.50	9.41 ± 1.27
TTF (s) ↑	7.25 ± 0.09	7.46 ± 0.17	7.00 ± 0.20	5.36 ± 0.10
T2R (s) ↓	4.59 ± 0.08	4.95 ± 0.12	5.13 ± 0.12	6.50 ± 0.07

notable variations in torque density and morphology. These inherent differences impact key factors like visual perception range/motor stability/overall movement complexity even within identical trails.

Implementations. *Proprioception* ($\mathcal{X}_{pro} \in \mathbb{R}^{45}$): covers lower-body joint positions, velocities, torso roll and pitch, foot contact indicators, and previous action a_{t-1} for both robots. *Actions* ($a_t \in \mathbb{R}^{10}$): the learned policy uses position control for joints, with positions converted to torque via a PD controller $\tau = K_p(\hat{q} - q) + K_d(\dot{\hat{q}} - \dot{q})$ with fixed gains (K_p and K_d follow default configuration of Unitree). *Training*: for both oracle and unified policy training, we use PPO [44], supported by Dagger [45] and Actor-Critic [46] for privileged learning. Rewards follow those introduced in method section, with additional basic elements from [38, 47]. All physics simulations perform in Isaac Gym simulator [48].

Metrics. We evaluate models based on three core criteria with levels of granularity: *goal completeness*, *safeness*, and *efficiency*. Concretely, we use 6 evaluation metrics – (1) *Goal Completeness*: Success Rate (%) measuring the percentage of episodes where robots reach the hiking endpoint; Trail Completion (%) indicating the portion of the trail route a robot passed; and Traverse Rate (%) reflecting the distance from robot's final position (if not complete goal) to endpoint relative to total trail length. (2) *Safeness*: MEV (%) assessing foot-edge collisions; and TTF ($seconds$) evaluating robot stability based on episode duration before a fall occurs. (3) *Efficiency*: Time-to-Reach ($seconds$) measuring average time required for successful episodes to reach endpoint. Unless specified, experiments are conducted with 512 randomly spawned robots over 30 seconds on 5 distinct trail types, each featuring 5 difficulty levels. Results are averaged over 5 runs to minimize random biases and verify robustness.

5.2 Ablation Study

Settings. We compare full LEGO-H with following designs: (1)*Oracle*: trained with access to privileged info and expert-designed navigation goals, representing an upper-bound performance.(2) *w TC-ViT*: LEGO-H trained without Hierarchical Latent Matching (HLM) loss metric. (3)*Vanilla*: LEGO-H variant where TC-ViT is replaced by a ConvGRU to predict latent and goal, altering the navigation mechanism.

Results. Tab 1 indicates several insights. (1) *TC-ViT is essential for basic hiking functionality.* The consistent, significant performance advantage of *w TC-ViT* over *Vanilla* across all metrics, except MEV, reveals the essence of balancing the goal, physical state, and visual perception, which is crucial for coordination between navigation and locomotion.(2) *Structural action behavior helps more efficient goal accomplishment and better stability.* The absence of HLM (*w TC-ViT)* results in behaviors that complete tasks but compromise stability, often leading to mechanical risks (worse MEV than others). Including HLM (*LEGO-H)* ensures coordinated joint actions that align with the robot's physical structure, promoting both task success (SR rises from 64.73% to 68.40%) and mechanical integrity (MEV goes from 10.40% to 7.84%, TTF increase to 7.46s), leading to more efficient task accomplishment (T2R improves from 5.13s to 4.95s). (3) *LEGO-H rivals oracle in efficiency and safety.* Compared to oracle which has perfect observation conditions and expert navigation goals, LEGO-H falls behind on success rate and trail completion. But surprising aspects are the efficiency and safeness, where LEGO-H's performances are comparable to or slightly better than oracle. This stresses again LEGO-H's effectiveness and capacity.

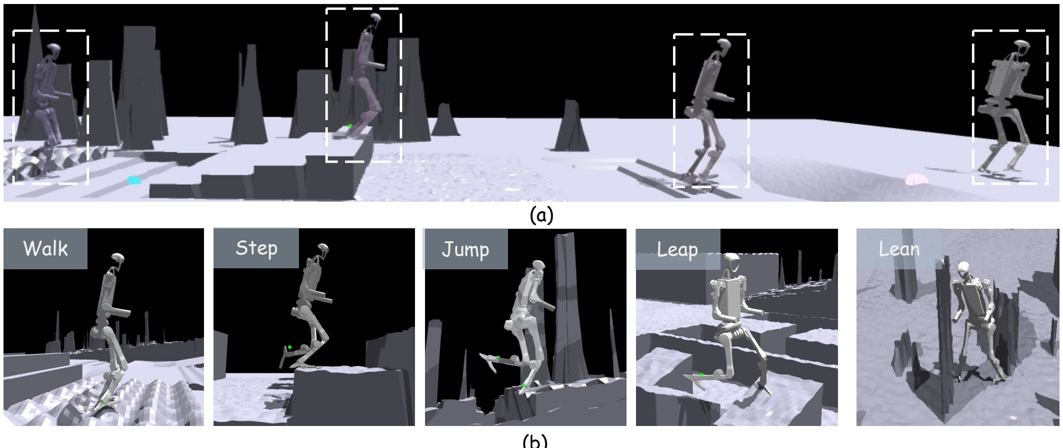

Figure 5: **Locomotion in diverse trail terrains.** Robots developed distinct motor skills to tackle different terrains, *e.g.,* walking on rough surfaces/leaping across ditches/leaning away high obstacles.

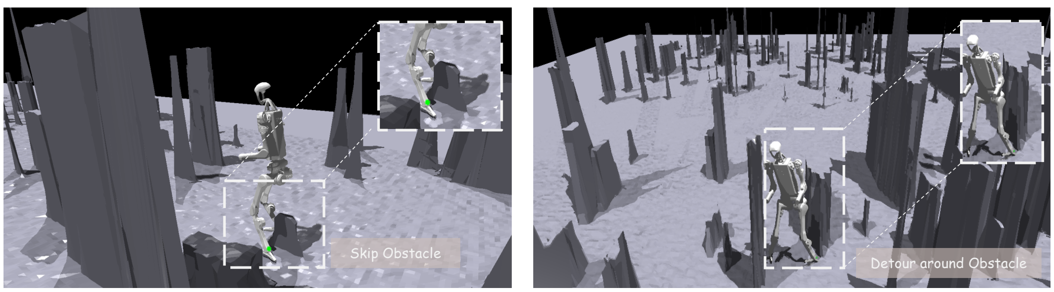

Figure 6: **Navigation in diverse situations.** Robots developed different navigation skills, such as directly skipping a small obstacle and detouring around a high obstacle to edge through.

5.3 Emerged Behaviors in Different Situations

We further explore the behaviors that emerge in humanoid robots to unfold how robots autonomously adapt their motor skills and decision-making in response to various factors.

Locomotion in diverse trail terrains. Different terrains trigger distinct locomotion behaviors, like *walking, stepping, jumping, leaping, and leaning* (Fig 5). Key observations include: (1) H1 robots typically opt for a walking gait on continuous surfaces, regardless of variations in friction, adjusting their body tilt as needed to maintain balance (Fig. 5a). (2) Irregular surfaces, like fractured or sloped terrains, prompt gaits like stepping, jumping, or leaping, depending on slope and gap size (Fig. 5b). (3) In tight spaces, such as cracks between large obstacles, H1's adapt by leaning sideways to navigate through these confined areas (*Lean* in Fig. 5b).

Navigation in blocked paths. Two key behaviors are evident from Fig 6: (1) When faced with tall or large obstacles, the robots typically choose to detour, maintaining a safe clearance from the obstacles. (2) For obstacles below hip height, the robots initially attempt to stride or step over; if unsuccessful, they then choose to detour. These phenomena reveal the embodied character in high-level decisions.

Motor behavior differences between robots. As shown in Fig 7, when encountering identical trails like transitions between platform and flat ground, H1 and G1 exhibit different behaviors. H1 navigates down smoothly, while G1 bends its knees to jump down. This difference highlights the impact of physical mechanisms on emergent motor styles.

5.4 Humanoid Hiking Benchmark

Settings. Since current research does not directly support humanoid hiking, we selected two representative quadruped pipelines, adapting them to this task using the same input structure and oracle policy as LEGO-H. This setup allows us to investigate several key factors essential for effective humanoid hiking. The first adapted pipeline, *EP-H*, represents a modified humanoid-hiking version

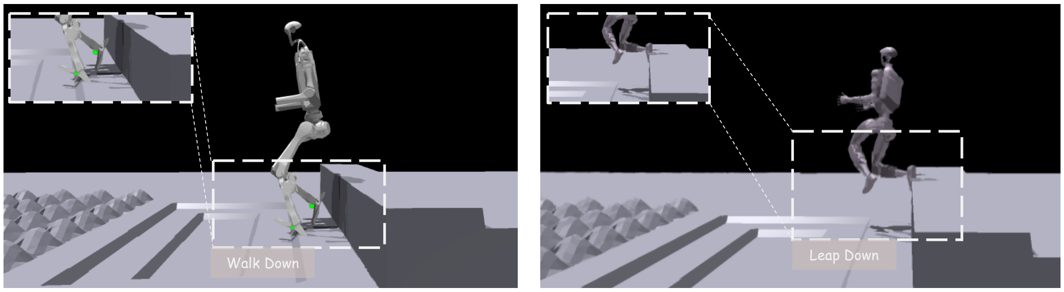

Figure 7: **Motor behavior differences between robots.** Robots with different structures developed unique skills – H1, which is higher and heavier, chooses to "walk down" step, while G1, which is shorter and more lightweight, chooses to "leap down" the step.

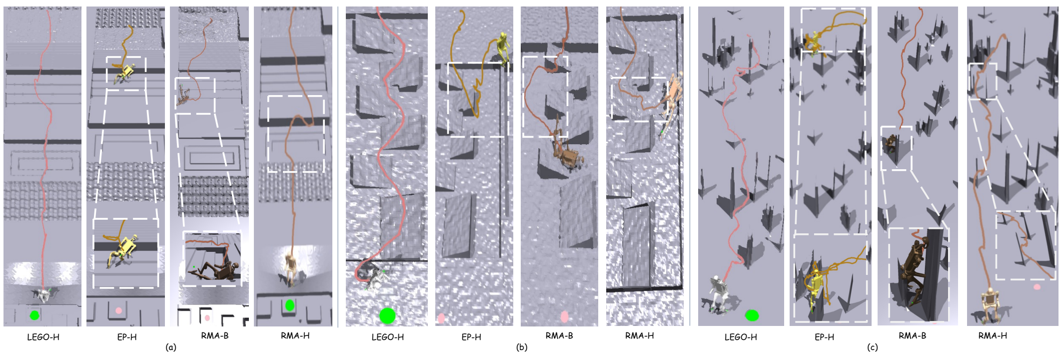

Figure 8: **Qualitative comparisons between LEGO-H and other benchmarked methods.** The trajectories, visualized through dynamically updated colored lines, depict the robots' torso position as they traverse diverse trail environments. (a) illustrates the performance on a *RandomMix* trail featuring unobstructed views with varied terrain types. (b) highlights the results on a *Ditch* trail, where uneven terrain with slopes and gaps demands quick turns and agile leaps. (c) showcases the performance on a *Forest* trail, where extensive obstacles of different sizes and heights block the robot's view. The zoom-in regions highlight the issues of the robots.

Table 2: **Humanoid hiking benchmark for H1 across all trail categories.** / / show best goal completeness/safeness/efficiency.

Metrics	LEGO-H	EP-H	RMA-H	RMA-B
Success Rate (%) ↑	68.40 ± 1.34	28.80 ± 0.88	65.17 ± 2.05	48.11 ± 0.72
Trail Completion (%) ↑	52.78 ± 1.30	25.98 ± 0.22	52.51 ± 1.41	41.92 ± 0.34
Traverse Rate (%) ↑	71.96 ± 2.37	64.16 ± 0.48	74.61 ± 0.93	69.85 ± 1.50
MEV (%) ↓	7.84 ± 0.92	12.44 ± 1.32	8.70 ± 1.55	10.74 ± 1.13
TTF (s) ↑	7.46 ± 0.17	4.64 ± 0.13	6.97 ± 0.17	5.22 ± 0.03
Time-to-Reach (s) ↓	4.95 ± 0.12	9.79 ± 0.16	4.98 ± 0.11	6.19 ± 0.05

of EP [38]. The main methodological difference between EP-H and LEGO-H is that EP-H handles visual-aware navigation and locomotion by processing each depth frame independently, disregarding farther depth data to avoid distributional shifts. *RMA-H* and *RMA-B* are the adapted pipeline from RMA [15] – the former has vision inputs, and the later is blind. This pipeline originally supports blind locomotion, and employs a frozen oracle policy with an adapter network to map real-world sensory data to oracle's latent space for policy adaptation.

Results. We focus on three vital questions from the benchmark: 1) *Is visual perception essential for integrated navigation and locomotion?* 2) *What type of visual information is most effective?* 3) *Is unified cross-level learning necessary?* Key findings in Tab 2 and Fig 8 revel the answers:(1) *Vision is essential.* Without vision, RMA-B struggles across all metrics, highlighting the need for visual feedback. (2) *Goal-aligned, multi-scale visual perception is critical. EP-H*, which processes each depth frame independently without continuous goal alignment, and brute-force cutoff distance information, results in frequent circles and fails to lock onto navigation paths. The performance gap between LEGO-H and EP-H across metrics underscores the importance of structured visual information. (3) *Unified learning is vital for adaptability. RMA-H* performs adequately on straight paths but fails with turns or obstacles, showing that locomotion feedback alone is insufficient for embodied-aware decision-making. A unified learning framework supports essential cross-level interaction, enabling adaption and effectiveness across all levels.

6 Conclusion

We propose *humanoid hiking* as a new testbed for advancing research in embodied autonomy. To address the challenges it poses, we introduce LEGO-H, a unified policy learning framework that highlights the importance of integrative skill development for a humanoid to autonomously accomplish complex tasks like hiking. Experiments demonstrate effectiveness of LEGO-H and also uncover promising directions for future research, like whole-body control, long-horizon exploration, and visual-motor coordination.

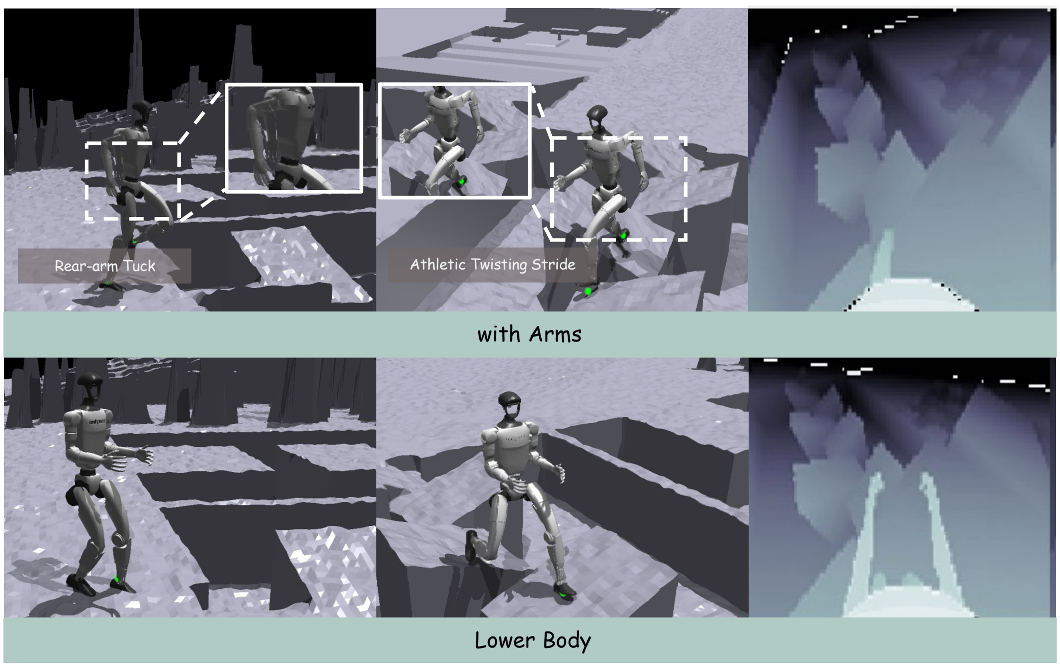

Figure 9: **Preliminary observations for future work on Whole-body Control setting.** G1 exhibits distinct motor behaviors over *with arms* vs *only lower body*. Besides, G1 emerges a rear-arm tuck posture while walking, likely to minimize arm interference with vision (see depth map).

7 Limitations

There are four limitations: (1) Kilometer-scale hiking. In this paper, we investigate humanoid robots on prototype trails to establish a baseline on the importance of integrative high-level navigation and low-level motor skills. However, real-world trails are considerably more complex, with long-distance traverse challenges. Future work could expand the framework to handle kilometer-scale trails, where sustained adaptability, energy efficiency, and long-term planning become crucial. (2) Whole-body control for integrative navigation and locomotion skills. Expanding control across the entire body would enable a wider spectrum and adaptive behaviors, enhancing the robot's flexibility in complex, obstacle-rich environments. Our preliminary results suggest that while robots exhibit distinct motor styles based on physical constraints(Fig. 9), *direct* involvement of the upper body does not significantly impact performance in a positive manner. This opens opportunities for future work on exploring how coordinated whole-body strategies can enhance performance. (3) Simulated environment upgrading. Our current simulated trails are primarily for foot contact; Future work could upgrade the simulated environment to better incorporate whole-body interactions, enabling a better testbed for future hiking studies. (4) Real-world deployment. In this paper, we conduct experiments on the simulator, enabling controlled benchmarking, rapid iteration, and reproducibility — *key prerequisites* for real-world deployment. However, applying LEGO-H to real-world scenarios remains a vital next step toward closing the sim-to-real gap and realizing field-ready humanoid hikers.

References

[1] I. Nordbø and N. K. Prebensen. Hiking as mental and physical experience. In *Advances in hospitality and leisure*, volume 11, pages 169–186. 2015.

[2] D. Mitten, J. R. Overholt, F. I. Haynes, C. C. D'Amore, and J. C. Ady. Hiking: A low-cost, accessible intervention to promote health benefits. *American journal of lifestyle medicine*, 12 (4):302–310, 2018.

[3] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath. Real-world humanoid locomotion with reinforcement learning. *Sci. Robotics*, 2024.

[4] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath. Learning humanoid locomotion with transformers. *CoRR*, abs/2303.03381, 2023.

[5] C. Li, E. Stanger-Jones, S. Heim, and S. Kim. FLD: fourier latent dynamics for structured motion representation and learning. In *ICLR*, 2024.

[6] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Trans. Graph.*, 2021.

[7] Z. Zhuang, S. Yao, and H. Zhao. Humanoid parkour learning. *CoRR*, abs/2406.10759, 2024.

[8] B. Dynamics. https://bostondynamics.com/blog/leaps-bounds-and-backflips/. 2018.

[9] O. Mohareri and A. B. Rad. Autonomous humanoid robot navigation using augmented reality technique. In *2011 IEEE International Conference on Mechatronics*, 2011.

[10] M. Missura, A. Roychoudhury, and M. Bennewitz. Polygonal perception for mobile robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*. IEEE, 2020.

[11] J. Wang, A. Laurenzi, and N. G. Tsagarakis. Autonomous behavior planning for humanoid loco-manipulation through grounded language model. *CoRR*, abs/2408.08282, 2024.

[12] J. Garimort, A. Hornung, and M. Bennewitz. Humanoid navigation with dynamic footstep plans. In *IEEE International Conference on Robotics and Automation, ICRA*, 2011.

[13] J. Ahn, S. J. Jorgensen, S. Bang, and L. Sentis. Versatile locomotion planning and control for humanoid robots. *Frontiers Robotics AI*, 2021.

[14] C. Tessler, Y. Kasten, Y. Guo, S. Mannor, G. Chechik, and X. B. Peng. Calm: Conditional adversarial latent models for directable virtual characters. In *ACM SIGGRAPH Conference Proceedings*, 2023.

[15] A. Kumar, Z. Fu, D. Pathak, and J. Malik. Rma: Rapid motor adaptation for legged robots. In *Robotics: Science and Systems*, 2021.

[16] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In Y. Bengio and Y. LeCun, editors, *International Conference on Learning Representations, ICLR*, 2014.

[17] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for humanoid robots. *RSS*, 2024.

[18] Q. Liao, B. Zhang, X. Huang, X. Huang, Z. Li, and K. Sreenath. Berkeley humanoid: A research platform for learning-based control. *CoRR*, abs/2407.21781, 2024.

[19] X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, and J. Chen. Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning. In *Robotics: Science and Systems*, 2024.

[20] S. Sovukluk, J. Englsberger, and C. Ott. Highly maneuverable humanoid running via 3d slip+foot dynamics. *IEEE Robotics Autom. Lett.*, 9(2):1131–1138, 2024.

[21] F. M. Smaldone, N. Scianca, L. Lanari, and G. Oriolo. From walking to running: 3d humanoid gait generation via MPC. *Frontiers Robotics AI*, 9, 2022.

[22] T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, J. Humplik, M. Wulfmeier, S. Tunyasuvunakool, N. Y. Siegel, R. Hafner, M. Bloesch, K. Hartikainen, A. Byravan, L. Hasenclever, Y. Tassa, F. Sadeghi, N. Batchelor, F. Casarini, S. Saliceti, C. Game, N. Sreendra, K. Patel, M. Gwira, A. Huber, N. Hurley, F. Nori, R. Hadsell, and N. Heess. Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *Sci. Robotics*, 2024.

[23] X. B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Trans. Graph.*, 2022.

[24] Z. Luo, J. Cao, J. Merel, A. Winkler, J. Huang, K. M. Kitani, and W. Xu. Universal humanoid motion representations for physics-based control. In *ICLR*, 2024.

[25] C. Tessler, Y. Guo, O. Nabati, G. Chechik, and X. B. Peng. Maskedmimic: Unified physics-based character control through masked motion inpainting. *ACM Transactions on Graphics (TOG)*, 2024.

[26] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim. The MIT humanoid robot: Design, motion planning, and control for acrobatic behaviors. In *IEEE-RAS International Conference on Humanoid Robots, Humanoids*, 2021.

[27] C. Zhang, W. Xiao, T. He, and G. Shi. Wococo: Learning whole-body humanoid control with sequential contacts. *CoRL*, 2024.

[28] O. A. Donca, C. Beokhaimook, and A. Hereid. Real-time navigation for bipedal robots in dynamic environments. *CoRR*, 2022.

[29] Y. Lin and D. Berenson. Humanoid navigation in uneven terrain using learned estimates of traversability. In *IEEE-RAS International Conference on Humanoid Robotics, Humanoids*. IEEE, 2017.

[30] Y. Lin and D. Berenson. Long-horizon humanoid navigation planning using traversability estimates and previous experience. *Auton. Robots*, 2021.

[31] A. Brandenburger, D. Rodriguez, and S. Behnke. Mapless humanoid navigation using learned latent dynamics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*. IEEE, 2021.

[32] J. Lee, M. Bjelonic, A. Reske, L. Wellhausen, T. Miki, and M. Hutter. Learning robust autonomous navigation and locomotion for wheeled-legged robots. *Sci. Robotics*, 2024.

[33] R. Yang, M. Zhang, N. Hansen, H. Xu, and X. Wang. Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers. In *ICLR*, 2022.

[34] D. Hoeller, L. Wellhausen, F. Farshidian, and M. Hutter. Learning a state representation and navigation in cluttered and dynamic environments. *IEEE Robotics Autom. Lett.*, 2021.

[35] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter. Advanced skills by learning locomotion and local navigation end-to-end. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 2022.

[36] AllTrails. https://www.alltrails.com/.

[37] R. Yang, G. Yang, and X. Wang. Neural volumetric memory for visual locomotion control. In *CVPR*, 2023.

[38] X. Cheng, K. Shi, A. Agarwal, and D. Pathak. Extreme parkour with legged robots. In *IEEE International Conference on Robotics and Automation, ICRA*, 2024.

[39] J. Frey, M. Mattamala, L. Piotr, N. Chebrolu, C. Cadena, G. Martius, M. Hutter, and M. Fallon. Wild Visual Navigation: Fast Traversability Learning via Pre-Trained Models and Online Self-Supervision. In *Arxiv*, 2024.

[40] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, and C. Schmid. Vivit: A video vision transformer. In *International Conference on Computer Vision (ICCV)*, 2021.

[41] D. Shah, A. Sridhar, N. Dashora, K. Stachowicz, K. Black, N. Hirose, and S. Levine. Vint: A foundation model for visual navigation. In J. Tan, M. Toussaint, and K. Darvish, editors, *Conference on Robot Learning, CoRL*, Proceedings of Machine Learning Research. PMLR, 2023.

[42] UnitreeH1. https://www.unitree.com/h1.

[43] UnitreeG1. https://www.unitree.com/g1/.

[44] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *CoRR*, 2017.

[45] S. Ross, G. J. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In G. J. Gordon, D. B. Dunson, and M. Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS*, JMLR Proceedings, 2011.

[46] V. R. Konda and J. N. Tsitsiklis. Actor-critic algorithms. In S. A. Solla, T. K. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems 12, [NIPS*, 1999.

[47] X. Gu, Y. Wang, and J. Chen. Humanoid-gym: Reinforcement learning for humanoid robot with zero-shot sim2real transfer. 2024.

[48] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State. Isaac gym: High performance GPU based physics simulation for robot learning. In J. Vanschoren and S. Yeung, editors, *NeurIPS Datasets and Benchmarks*, 2021.