Robotics: Science and Systems 2026
Sydney, Australia, July 13–17, 2026

HiWET: Hierarchical World-Frame End-Effector Tracking for Long-Horizon Humanoid Loco-Manipulation

Zhanxiang Cao[1,2], Liyun Yan[1,2], Yang Zhang[1], Sirui Chen[1], Jianming Ma[1,2], Tianyue Zhan[1,2], Shengcheng Fu[3,2], Yufei Jia[4], Cewu Lu[1,2], Yue Gao[1,2,†]
[1]Shanghai Jiao Tong University [2]Shanghai Innovation Institute [3]Tongji University [4]Tsinghua University
[†]Corresponding author. Email: yuegao@sjtu.edu.cn

***Abstract*—Humanoid loco-manipulation requires precise world-frame end-effector tracking while maintaining dynamic stability amid base motion and impacts. Existing approaches typically formulate commands in body-centric frames and fail to directly correct cumulative world-frame drift induced by legged locomotion. We reformulate the problem as world-frame end-effector tracking and propose HiWET, a hierarchical reinforcement learning framework that decouples global reasoning from dynamic execution. The high-level policy generates subgoals that jointly optimize end-effector accuracy and base positioning in the world frame, while the low-level policy executes these commands under stability constraints. We introduce a Kinematic Manifold Prior (KMP) that embeds a manipulation-relevant kinematic manifold into the action space via residual learning, reducing exploration dimensionality and mitigating kinematically invalid behaviors. Extensive simulation and ablation studies demonstrate that HiWET achieves precise and stable end-effector tracking in long-horizon world-frame tasks. We validate zero-shot sim-to-real transfer of the low-level policy on a physical humanoid, demonstrating stable locomotion under diverse end-effector tracking commands. These results indicate that explicit world-frame reasoning combined with hierarchical control provides an effective solution for world-frame end-effector tracking during humanoid locomotion.**

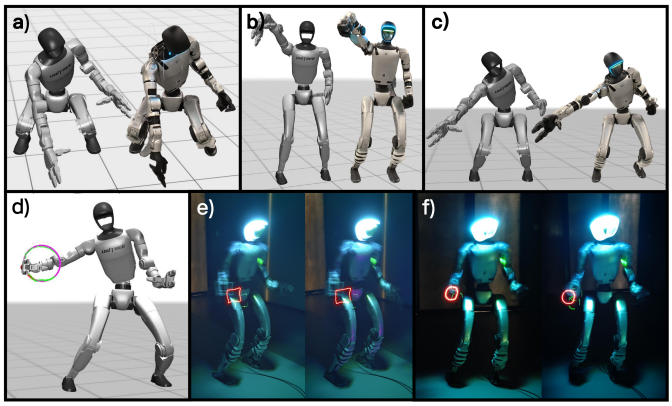

Fig. 1. **HiWET capabilities in simulation and real-world deployment.** (a)-(c) Whole-body redundancy for diverse reaching tasks (left: simulation, right: real robot): (a) lowest/farthest, (b) highest, (c) outermost semi-squat. (d) MuJoCo sim-to-sim world-frame tracking (red: target, green: actual). (e)-(f) Real-world long-exposure square/circle trajectories traced by an end-effector LED.

I. Introduction

Recent advances in reinforcement learning (RL) and imitation learning have driven rapid progress in humanoid whole-body motion control [28, 7, 43, 33]. Motion retargeting addresses the human-robot embodiment gap for expressive, balanced imitation [11, 10, 13, 5, 15, 3, 40, 12], while diffusion models, large-scale motion datasets, and visual perception further broaden whole-body control [41, 24, 23, 42, 46, 38, 34]. These results demonstrate that learning-based methods can produce stable, expressive humanoid behaviors.

Moving beyond motion imitation, command-driven loco-manipulation enables task-level control through abstract interfaces [20, 2, 44]. Operators can specify end-effector poses or joint targets via optimization, teleoperation, or planners, yielding interpretable geometric control with clear task semantics. Hierarchical architectures decouple high-level planning from low-level motor control through latent skills [45, 16] or explicit trajectory commands [13, 35, 29, 19]; combined with end-effector stabilization [21], they have advanced practical humanoid loco-manipulation.

Whole-body motion imitation optimizes joint-space errors rather than explicit task-space end-effector accuracy [1, 17] and typically requires costly dense reference trajectories [27]. Meanwhile, most command-driven methods operate in body-centric frames, where cumulative base drift and high-frequency oscillations can significantly degrade end-effector precision in the world frame [22]. The tight coupling between upper and lower body dynamics further compounds this challenge: aggressive arm motions can destabilize the gait, while locomotion-induced impacts propagate to the end-effectors [8]. When task trajectories extend beyond the static reachable workspace, the robot must actively transport its base to maintain reachability—a coordination that body-centric formulations do not explicitly address [31]. We formulate the task as *world-frame end-effector tracking* to expose this coupling, allowing the controller to compensate for base disturbances while reshaping the workspace through locomotion. We assume a usable global pose estimate and focus on whole-body control under this assumption; localization accuracy ultimately bounds tracking precision.

To address these challenges, we propose HiWET, a hi-

erarchical reinforcement learning framework that enables precise world-frame end-effector tracking through explicit upper-lower body coordination. The high-level *world-frame command policy* reasons in global coordinates, generating structured subgoals—base velocity, body height, and local end-effector targets—to guide the robot toward task objectives while maintaining reachability. The low-level *whole-body tracking policy* operates at the same or higher frequency, translating these subgoals into joint commands while ensuring dynamic stability. This decomposition separates *where to go* (global planning) from *how to move* (dynamic execution). To improve learning and precision, we introduce a Kinematic Manifold Prior (KMP) that provides kinematically consistent upper-body references, enabling residual corrections rather than absolute joint targets.

Our main contributions are as follows:

- We propose a **world-to-body hierarchical control scheme** that coordinates end-effector tracking and locomotion through an explicit spatial interface, enabling world-frame consistency via active base transport and height adjustment.
- We introduce an **efficient Kinematic Manifold Prior (KMP)** within a residual action space, providing high-speed kinematic references that ground the policy in valid kinematic manifolds while preserving dynamic adaptability.
- We perform **extensive validation in simulation and on a physical humanoid**, demonstrating 12.4 mm world-frame tracking error in simulation and robust zero-shot sim-to-real transfer across diverse limb configurations.

II. Related Works

A. Humanoid Whole-Body Motion Control

Reinforcement learning and imitation learning have enabled humanoids to achieve expressive whole-body behaviors. The seminal work DeepMimic [28] demonstrated that RL can track motion capture references with physical realism, inspiring benchmarks such as Humanoid-Gym [7] and HumanoidBench [33]. Motion retargeting methods address the embodiment gap: H2O [11] and OmniH2O [10] enable real-time teleoperation, ExBody [5, 15] extends imitation to expressive upper-body motions, and SONIC [24] scales motion tracking with large-scale datasets. Generative approaches such as BeyondMimic [23] and MaskedManipulator [35] leverage diffusion models to synthesize versatile motions from sparse goals.

Command-driven frameworks enable task-level control: AMO [20] employs motion optimization, HOMIE [2] uses isomorphic teleoperation, and HOVER [13] consolidates locomotion, standing, and manipulation into a unified policy. Hierarchical architectures separate planning from motor control: R^2S^2 [45] and WholeBodyVLA [16] encode commands as latent skill codes, while FALCON [44] and VisualMimic [41] transmit trajectory-level commands with clearer geometric semantics.

B. Spatial Consistency in Humanoid Loco-Manipulation

Most existing methods reference commands to the robot base, assuming a stable coordinate frame [36, 37, 31]. Wheeled mobile manipulators such as Mobile ALOHA [6] largely maintain this assumption, achieving bimanual coordination via imitation on stable bases. However, legged humanoids must contend with cumulative drift and high-frequency oscillations during dynamic locomotion [22], compounded by tight upper-lower body coupling [8]. DexMan [14] explores floating-base manipulation without explicitly modeling leg-arm coupling. FALCON [44] and RAMBO [4] improve tracking through force adaptation and hybrid model-based RL, but still operate in body-centric coordinates.

In summary, imitation-based methods optimize joint-space errors rather than task-space precision [1] and require dense references [27], while command-driven methods operate in body-centric frames that cannot maintain world-frame consistency. HiWET addresses these limitations through world-frame end-effector tracking with an explicit Cartesian interface, enabling active workspace reshaping through locomotion while ensuring dynamic stability via a modular low-level policy.

III. Problem Formulation

We propose a hierarchical reinforcement learning (HRL) framework that decouples global spatial reasoning from dynamic execution. The overall architecture is illustrated in Fig. 2. The *world-frame command policy* functions as a task planner, interpreting global end-effector targets and generating local subgoals—specifically base velocity, body height, and base-relative hand poses. The *whole-body tracking policy* operates at the same or higher frequency to translate these subgoals into joint-space targets while ensuring physical feasibility and dynamic balance. By formulating this interplay as a semi-Markov decision process (Semi-MDP), the hierarchy effectively reconciles long-horizon end-effector tracking objectives with instantaneous stability constraints.

We model the humanoid control problem as a hierarchical reinforcement learning process

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}^H, \mathcal{A}^L, P, \mathcal{R}, \gamma \rangle, \tag{1}$$

where $\mathcal{S}$ is the state space, $\mathcal{A}^H$ and $\mathcal{A}^L$ denote the high-level and low-level action spaces, P is the transition dynamics, $\mathcal{R}$ is the reward function, and $\gamma \in (0, 1)$ is the discount factor.

In the general Semi-MDP formulation, π^H updates the subgoal command every K low-level steps while π^L executes at the same or higher frequency; in our implementation, we set $K = 1$, so both policies update at every policy control step. Both policies are optimized to maximize the expected discounted return $J = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t]$, where rewards are designed to balance global tracking precision with local dynamic stability.

IV. Whole-Body Tracking Policy

The whole-body tracking policy translates high-level motion commands into joint-space targets for stable whole-body execution. It operates at high frequency and interfaces directly

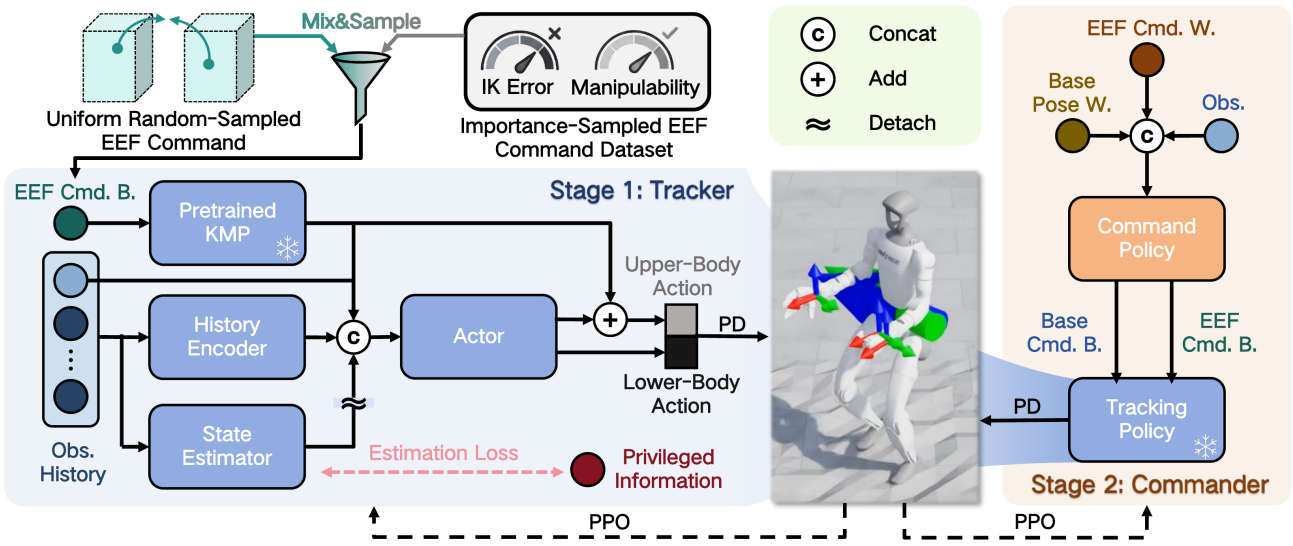

Fig. 2. **HiWET architecture and two-stage training procedure. Stage 1: Tracker** (blue): The tracking policy learns base-relative end-effector commands sampled from uniform and importance-sampled distributions filtered by IK error and manipulability. A pretrained KMP provides upper-body references refined by residual actions; the History Encoder extracts temporal context, and the State Estimator reconstructs privileged information via an auxiliary estimation loss. **Stage 2: Commander** (orange): The command policy maps world-frame EEF targets and base pose into base-frame subgoals.

with the PD controller. The humanoid is decomposed into an upper-body subsystem (arms and waist) and a lower-body subsystem (legs). This policy serves as a **standalone whole-body tracker** with a general-purpose command interface, capable of following arbitrary base velocity and end-effector subgoals independently of the high-level task planner.

A. State and Action Representation

1) State Space: The whole-body tracking policy is conditioned on both the humanoid proprioceptive state s_t and the structured command $\mathbf{u}_t$ received from the world-frame command policy. The observation at time t is defined as

$$s_t = \left[\boldsymbol{\omega}_t,\ \mathbf{g}_t,\ \mathbf{q}_t,\ \dot{\mathbf{q}}_t,\ \mathbf{a}_{t-1}^L\right], \tag{2}$$

where $\boldsymbol{\omega}_t \in \mathbb{R}^3$ is the base angular velocity, $\mathbf{g}_t \in \mathbb{R}^3$ is the gravity projection in the base frame, $\mathbf{q}_t$ and $\dot{\mathbf{q}}_t$ denote joint positions and velocities, and $\mathbf{a}_{t-1}^L$ is the previous action.

The command $\mathbf{u}_t$ transmits high-level spatial objectives to the low-level policy:

$$\mathbf{u}_t = \left[\mathbf{v}_b^{des},\ h^{des},\ {}^b\mathbf{T}_L^{des},\ {}^b\mathbf{T}_R^{des},\ \alpha_t\right], \tag{3}$$

where $\mathbf{v}_b^{des} = (v_x, v_y, \omega_z)$ is the desired base velocity, h^{des} is the target body height, ${}^b\mathbf{T}_{\{\cdot\}}^{des}$ denote the desired end-effector poses, and α_t is the waist regularization weight (uniformly sampled from $[0.1, 10]$ during training). We use the superscript $\{\cdot\}^{des}$ to distinguish these local subgoals from global world-frame targets.

2) Hybrid Action Space: Given the proprioceptive observation s_t and the command $\mathbf{u}_t$, the policy outputs joint position targets for all actuated joints:

$$a_t^L = \mathbf{q}_t^{des} = \left[\mathbf{q}_{t,up}^{des},\ \mathbf{q}_{t,low}^{des}\right] \sim \pi^L(\mathbf{q}_t^{des} \mid s_t, \mathbf{u}_t), \tag{4}$$

where $\mathbf{q}_{t,up}^{des}$ denotes the desired joint positions of the upper body and $\mathbf{q}_{t,low}^{des}$ denotes those of the lower body.

Instead of learning a monolithic policy, we adopt a hybrid action space that leverages kinematic structure. For the upper body, the policy predicts a residual correction $\Delta\mathbf{q}_{t,up}$ which is added to the KMP reference $\hat{\mathbf{q}}_{t,up}$ to form the final command:

$$\mathbf{q}_{t,up}^{des} = \hat{\mathbf{q}}_{t,up} + \Delta\mathbf{q}_{t,up}. \tag{5}$$

This residual formulation enables the controller to make subtle dynamic adjustments for end-effector stability and accuracy while remaining within the valid kinematic manifold. Conversely, the lower body is controlled via absolute joint targets $\mathbf{q}_{t,low}^{des}$ to maintain direct authority over gait generation and base propulsion.

B. Network Architecture Overview

The whole-body tracking policy uses a modular network architecture with three components for partial observability and kinematic consistency (Stage 1 in Fig. 2).

To ensure accurate end-effector tracking during motion and to provide a strong reference for learning, we introduce the Kinematic Manifold Prior (KMP) for the upper body. The KMP predicts a kinematically consistent joint configuration given desired end-effector poses and a waist motion preference. The input to the KMP is defined as

$$\mathbf{z}_t = \left[{}^b\mathbf{T}_L^{des},\ {}^b\mathbf{T}_R^{des},\ \alpha_t\right], \tag{6}$$

where ${}^b\mathbf{T}_{\{\cdot\}}^{des}$ are the commanded end-effector poses, and $\alpha_t \in [0.1, 10]$ is a scalar waist regularization weight (the fixed weight for other upper-body joints is 0.2). This parameter serves as a hierarchical interface that modulates torso engagement. Since humanoid stability is highly sensitive to the Center of Mass (CoM) position, explicitly allowing the global planner

to control waist motion enables the framework to trade off between reachability and balance; lower values of α_t encourage the KMP to utilize waist redundancy to extend the effective workspace, while larger values prioritize a stable CoM to maintain dynamic balance during locomotion. The KMP produces a reference joint configuration for the upper body,

$$\hat{\mathbf{q}}_{t,up} = f_{KMP}(\mathbf{z}_t), \tag{7}$$

where f_{KMP} is a neural approximation of the kinematic mapping for the arms and the waist. The KMP is pretrained offline and frozen during policy optimization, enabling the policy to learn residual corrections rather than absolute joint targets.

1) History Encoder: To improve robustness under partial observability, we construct a history buffer of length $H = 5$ containing tuples $(s_t, \mathbf{u}_t)$. The History Encoder (a CNN) processes this buffer $\mathcal{H}_t = \{(s_{t-i}, \mathbf{u}_{t-i})\}_{i=0}^{H-1}$ and outputs a latent representation $\mathbf{e}_t$ encoding temporal dependencies.

2) State Estimator: The State Estimator uses a history encoder followed by an MLP to infer privileged state variables from the observation history. It predicts base linear velocity and denoised, control-oriented estimates of the end-effector poses ${}^b\mathbf{T}_L, {}^b\mathbf{T}_R$. Although body-frame end-effector poses can be recovered through forward kinematics, the history-based estimator improves robustness to noisy proprioception, locomotion-induced oscillations, and partial observability. The estimator is trained jointly with the policy using a mean squared error loss.

3) Actor and Critic Inputs: The final input to the actor network is formed by concatenating the raw observation, the command, the privileged estimates from the State Estimator, the latent embedding from the History Encoder, and the kinematic reference from the frozen KMP:

$$\mathbf{o}_t^{actor} = \big[s_t,\ \mathbf{u}_t,\ \hat{\mathbf{p}}_t,\ \mathbf{e}_t,\ \hat{\mathbf{q}}_{t,up}\big]. \tag{8}$$

The critic is trained with access to full privileged information to reduce variance in value estimation. Its input is defined as

$$\mathbf{o}_t^{critic} = \big[s_t,\ \mathbf{u}_t,\ \mathbf{p}_t,\ \mathbf{h}_t,\ \hat{\mathbf{q}}_{t,up}\big], \tag{9}$$

where $\mathbf{p}_t$ includes the true base linear velocity and end-effector poses ${}^b\mathbf{T}_L, {}^b\mathbf{T}_R$, and $\mathbf{h}_t$ denotes terrain elevation measurements from the height map sensor. Both actor and critic are optimized using Proximal Policy Optimization (PPO) [32, 30].

C. Command Importance Sampling for Policy Learning

Uniform Cartesian command sampling yields many infeasible or ill-conditioned targets near workspace boundaries, degrading learning efficiency and stability. To mitigate this, we curate a dataset using importance sampling (top-left of Fig. 2).

To ensure that the policy focuses on functionally relevant and reachable configurations, we construct a prioritized end-effector command dataset. Two metrics are used to curate this dataset. First, the inverse kinematics reconstruction error is employed to prune unreachable poses for which the solver fails to converge. Second, the manipulability index is used to weight sampling probability, favoring comfortable configurations that enable high-dexterity motion.

To preserve coverage of the full command space, the actual command fed into the policy during training is drawn from a mixture distribution,

$$\mathbf{c} \sim \beta\, p_{uniform}(\mathbf{c}) + (1-\beta)\, p_{prior}(\mathbf{c}), \tag{10}$$

where $\beta \in [0,1]$ is the mixing weight, $p_{uniform}$ samples uniformly from the full Cartesian volume, and p_{prior} samples from the curated dataset with added small perturbations. This mixture ensures that the policy observes both realistic commands and boundary cases, improving robustness without sacrificing feasibility.

D. Kinematic Manifold Prior Learning

1) Dataset Generation: Training data for the KMP are generated by solving constrained optimization-based inverse kinematics problems using PyRoki [18]. For each sampled end-effector command, an optimization problem is constructed to minimize Cartesian pose errors subject to joint limits and a regularization term on waist motion weighted by α, which is uniformly sampled from $[0.1, 10]$. This enables the prior to capture a spectrum of limb configurations ranging from rigid-torso reaching to whole-body lunges. Approximately ten million samples are generated.

The initial end-effector command space is defined inside two symmetric Cartesian volumes around the torso center with full orientation coverage over the $SO(3)$ manifold. However, not all commands in this space are kinematically feasible. To curate the dataset, we apply importance sampling based on two metrics: the inverse kinematics reconstruction error, which prunes unreachable poses, and the manipulability index, which favors configurations that admit dexterous motion. Specifically, fivefold oversampling is performed and filtered according to these criteria as detailed in the appendix.

2) Network Architecture: The KMP adopts a symmetric encoder-decoder architecture with residual connections. It is based on a ResNet backbone [9] and incorporates layer normalization, dropout, and GELU activations to ensure stable training and generalization.

E. Reward Formulation

The reward function is designed to encourage accurate command tracking, stable locomotion, and precise end-effector tracking, while preserving smooth and physically consistent motions. In addition to standard tracking terms for base linear velocity, angular velocity, body height, and end-effector poses, several task-specific shaping rewards are introduced.

1) KMP Reference Tracking: To exploit the kinematic prior provided by the KMP, we include a reward that encourages the executed upper-body configuration to stay close to the reference. Let $\hat{\mathbf{q}}_{t,up}$ denote the reference joint configuration from the KMP and $\mathbf{q}_{t,up}$ denote the current upper-body joint positions. The tracking error is defined as

$$r_{kmp,t} = \exp\left(-\frac{1}{N_{up}\sigma_{kmp}^2}\|\mathbf{q}_{t,up} - \hat{\mathbf{q}}_{t,up}\|_2^2\right), \tag{11}$$

where σ_{kmp} controls the sensitivity. A time-dependent scaling factor is applied after command resampling to avoid overly constraining early transient responses.

2) Base Height Tracking with Knee Awareness: To stabilize standing and squatting, we introduce a base height tracking term with asymmetric knee-aware weighting. Let h_t be the measured base height and h_t^{des} the commanded height. The height error is weighted by knee joint margins: when above target, by flexion margin; when below, by extension margin:

$$e_{h,t} = \begin{cases} (h_t - h_t^{des}) \cdot \bar{w}_{knee}^{flex} & \text{if } h_t > h_t^{des}, \\ (h_t - h_t^{des}) \cdot \bar{w}_{knee}^{ext} & \text{if } h_t < h_t^{des}, \end{cases} \tag{12}$$

where $\bar{w}_{knee}^{flex}$ and $\bar{w}_{knee}^{ext}$ are the average margins for knee flexion and extension, respectively. This term is activated only under zero velocity conditions to avoid interference with dynamic locomotion.

Additional regularization terms penalize excessive joint velocity, action rate, and torque variation to ensure smooth and hardware-friendly motions. The complete reward formulation and weighting is provided in the appendix.

V. World-Frame Command Policy

The world-frame command policy π^H functions as a global task planner, translating world-frame end-effector targets into dynamically feasible whole-body commands (see Stage 2: Commander in Fig. 2). It reasons over the coupled nature of locomotion and end-effector tracking, effectively enlarging the robot's operational workspace through active base transport. During this training stage, the whole-body tracking policy from Stage 1 is frozen.

A. State and Command Representation

1) Observation and Task Command: The world-frame command policy observes the proprioceptive state alongside global information and the task-level command $\mathbf{c}_t$. The observation is defined as

$$s_t^H = \left[s_t,\ {}^w\mathbf{T}_b^t,\ \mathbf{c}_t\right], \tag{13}$$

where ${}^w\mathbf{T}_b^t \in SE(3)$ denotes the base pose in the world frame. The task command $\mathbf{c}_t$ is defined as

$$\mathbf{c}_t = \left[{}^w\mathbf{T}_L^*,\ {}^w\mathbf{T}_R^*,\ \mathbf{m}_t\right], \tag{14}$$

where ${}^w\mathbf{T}_{\{\cdot\}}^* \in SE(3)$ specify the desired end-effector poses in the absolute world frame, and $\mathbf{m}_t \in \{[1,0],[0,1],[1,1]\}$ is a binary mask that indicates which end-effector targets are active, enabling compatibility with both unilateral and bilateral tracking tasks.

We employ an asymmetric actor-critic architecture: while the actor observes the above state, the critic is augmented with privileged information:

$$\mathbf{o}_t^{H,\text{critic}} = \left[s_t^H,\ \mathbf{v}_b^{w,t},\ {}^w\mathbf{T}_L^t,\ {}^w\mathbf{T}_R^t\right], \tag{15}$$

where $\mathbf{v}_b^w$ is the base linear velocity and ${}^w\mathbf{T}_{\{\cdot\}}^t$ are the current end-effector poses in the world frame. The policy is optimized using PPO [32, 30].

2) Action Space: The world-frame command policy outputs a structured motion command that serves as the input to the whole-body tracking policy:

$$\begin{aligned} a_t^H = \mathbf{u}_t = \left[\mathbf{v}_b^{des},\ h^{des},\ {}^b\mathbf{T}_L^{des},\ {}^b\mathbf{T}_R^{des},\ \alpha_t\right] \\ \sim \pi^H(\mathbf{u}_t \mid s_t^H). \end{aligned} \tag{16}$$

This abstraction allows the policy to focus on geometric consistency and reachability while delegating dynamic stability and contact management to the whole-body tracking policy. The inclusion of α_t in the action space enables the high-level policy to regulate the robot's posture and CoM stability based on the global task context and locomotion state.

B. Spatial Curriculum Strategy

Training the world-frame command policy directly on large workspace ranges is challenging, as distant targets require long-horizon locomotion that complicates credit assignment. We adopt a curriculum learning strategy that progressively expands the world-frame command range based on tracking performance.

Initially, world-frame targets ${}^w\mathbf{T}_{\{\cdot\}}^*$ are sampled within a small neighborhood around the robot's current base position. As training progresses, we monitor the end-effector tracking error $e_{ee} = \|{}^w\mathbf{T}_{\{\cdot\}} - {}^w\mathbf{T}_{\{\cdot\}}^*\|$ and expand the sampling range when the error falls below a threshold. Specifically, the planar command range R is updated as:

$$R_{k+1} = R_k + \Delta R \cdot \mathbb{1}[e_{ee} < \epsilon_{th}], \tag{17}$$

where ΔR is the range increment and ϵ_{th} is the error threshold. This curriculum yields staged behavior: base transport for far targets, then whole-body coordination once targets are reachable.

C. Reward Formulation

The high-level reward function is designed to reconcile absolute tracking precision with optimal base positioning.

- **Workspace Optimization**: To maintain high manipulability, we penalize the planar distance between the robot base and the active end-effector targets. We further reward the alignment of both the base heading and the commanded velocity vector $\mathbf{v}_b^{des}$ with the vector pointing toward the target center, ensuring the policy actively "steers" the body to reshape the available workspace.
- **Precise End-effector Tracking**: Direct rewards are applied to minimize the pose error between ${}^w\mathbf{T}_{\{\cdot\}}$ and ${}^w\mathbf{T}_{\{\cdot\}}^*$ to enforce global spatial consistency.
- **Stability and Regularization**: To ensure hardware compatibility, we penalize action rates and high-frequency end-effector jitter. Additionally, when an arm is inactive (as indicated by $\mathbf{m}_t$), the policy is penalized for commanding deviations from a default neutral pose to prevent task-irrelevant limb drift.

A detailed description of reward components and weighting is provided in the appendix. Through this hierarchical formulation, the world-frame command policy learns to coordinate

locomotion and end-effector tracking in the world frame, while delegating joint-space realization to the whole-body tracking policy.

VI. Experiments And Results

The experiments aim to answer the following key questions.

1) How do the constituent components of the whole-body tracking policy—kinematic prior, privileged estimation, and importance sampling—contribute to command tracking precision and stability?
2) How does the world-frame command policy achieve spatial consistency and base positioning accuracy in long-horizon trajectory tracking tasks?
3) How does the proposed KMP compare to optimization-based inverse kinematics solvers in terms of inference efficiency and reconstruction accuracy?
4) Can the whole-body tracking policy successfully transfer to real hardware without additional fine-tuning?

A. Experimental Setup

Our experimental platform is the Unitree G1 humanoid robot; the policy controls 29 body degrees of freedom (12 for legs, 14 for dual arms, and 3 for waist), excluding the finger joints. All policies are trained in Isaac Lab [26] and deployed via zero-shot sim-to-real transfer. Sections on low-level tracking, world-frame command policy, and KMP benchmarking are evaluated in simulation unless explicitly marked as hardware experiments. Further details are provided in the Appendix.

TABLE I
Command Tracking Performance Comparison

Method	Lin. Vel. Error (m/s)	Ang. Vel. Error (rad/s)	Height Error (m)	EE Pos. Error (mm)
HiWET	0.157 ± 0.003	0.461 ± 0.006	0.018 ± 0.012	12.4 ± 2.4
HiWET w/o IS	0.165 ± 0.005	0.472 ± 0.006	0.018 ± 0.014	16.1 ± 5.3
HiWET w/o State Est.	0.169 ± 0.003	0.459 ± 0.004	0.018 ± 0.016	23.0 ± 7.2
HiWET w/o KMP	0.149 ± 0.004	0.423 ± 0.005	0.015 ± 0.010	25.2 ± 12.8
HOMIE [2]	0.194 ± 0.003	0.451 ± 0.006	0.022 ± 0.019	-

B. Low-Level Tracking Performance

We evaluate low-level tracking across whole-body commands, measuring *base linear velocity error*, *angular velocity error*, *body height error*, and *hand Cartesian error*. Commands are uniformly sampled within: linear velocity $v \in [-1, 1]$ m/s, angular velocity $\omega_z \in [-3, 3]$ rad/s, body height $h \in [0.3, 0.78]$ m, and end-effector poses within the maximal reachable Cartesian volume. We compare against HOMIE [2] and ablations in Table I.

1) Comparison with Baselines: HiWET reduces base linear velocity RMSE by ∼20% compared with HOMIE in this command-tracking setting. Because HOMIE receives joint position commands rather than end-effector poses, direct end-effector comparison is not available in Table I. Our hierarchical formulation exposes locomotion–tracking coupling, enabling more effective coordination while maintaining superior height consistency.

2) Ablation Studies and Precision Improvements: The ablation results highlight the critical contribution of each constituent component to the precision and robustness of bimanual tracking:

- **HiWET w/o KMP**: Removing the KMP reference causes the most significant drop—hand error doubles (25.2 mm) with 5× higher variance—confirming that learning Cartesian tasks without kinematic guidance is significantly more challenging.
- **HiWET w/o State Est.**: Without the state estimator, tracking degrades by nearly 10 mm, indicating that accurate end-effector feedback is essential for compensating locomotion-induced oscillations.
- **HiWET w/o IS**: Comparison with uniform sampling shows that focusing training on functional workspace regions yields more consistent bimanual coordination.

Additional ablation studies on reward components, including the knee-aware height tracking term, are provided in the Appendix.

Overall, the full HiWET framework achieves the highest precision (12.4 mm hand error) while maintaining the lowest variance, demonstrating its robustness as a whole-body end-effector tracking controller.

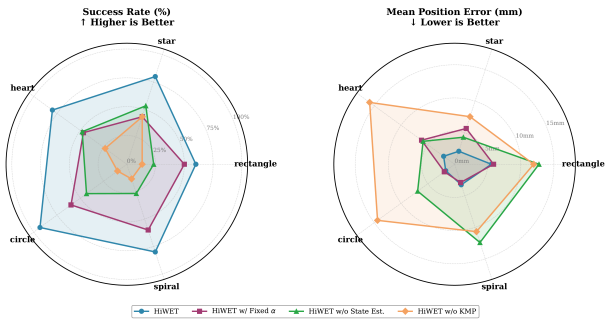

Fig. 3. **Tracking success rate and mean position error across diverse geometric trajectories.** HiWET demonstrates improved performance and consistency compared to its ablated versions (w/ Fixed α, w/o State Est., w/o KMP), particularly in complex tasks like tracing stars and hearts.

C. World-Frame Navigation and Long-Horizon Tracking

1) Trajectory Tracking Evaluation: We evaluate the capability of the high-level policy to perform spatially extended world-frame end-effector tracking tasks, measuring *tracking success rate* and *mean end-effector position error*. Geometric trajectories serve as standardized benchmarks to evaluate the controller's geometric precision and robustness across diverse motion profiles (e.g., sharp turns, velocity changes, symmetric/asymmetric movements). Five distinct geometric trajectories—star, heart, circle, spiral, and rectangle—are randomly initialized at global positions within a range of ± 5 m in both x and y axes relative to the robot's starting location. This requires the robot to first navigate to the target region and then execute precise end-effector tracking while maintaining dynamic stability. While the formulation supports bilateral commands, the high-level world-frame evaluations use a single active end-effector because bilateral global targets

impose stricter feasibility constraints on base placement, torso usage, and balance.

A trial is considered successful if the average end-effector tracking error remains below 20 mm throughout the trajectory execution. This criterion evaluates both the navigation–tracking coordination of the high-level policy and the disturbance rejection of the low-level policy.

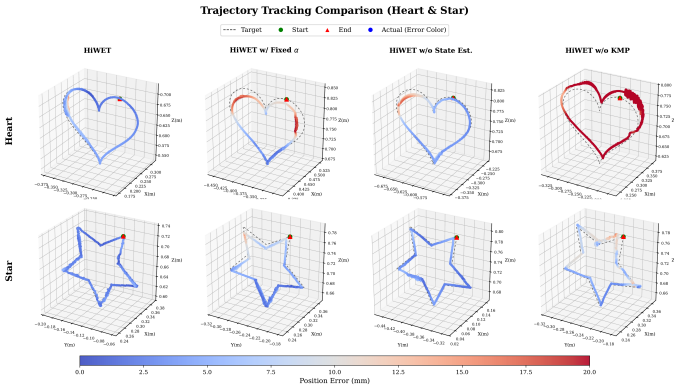

Fig. 4. **Qualitative comparison of 3D world-frame trajectory tracking for heart and star shapes.** Color indicates instantaneous Cartesian error (blue: < 2 mm, red: > 20 mm). HiWET maintains high spatial consistency, while fixed $\alpha = 1.0$, w/o State Est., and w/o KMP variants show deviation or oscillation. Additional trajectories are in the Appendix.

As illustrated in Fig. 3, HiWET achieves high success rates and low tracking errors across all geometric patterns. The qualitative comparison in Fig. 4 highlights heart and star trajectories, which require sharp changes in direction and coordinated end-effector movement. HiWET (left column) consistently maintains errors below 5 mm, while the ablated variants exhibit significant performance degradation. Specifically, the variant without the kinematic prior (*w/o KMP*) shows severe trajectory distortions, as the high-level commands often drive the arms into kinematically ill-conditioned regions. The removal of the privileged estimator (*w/o State Est.*) leads to increased oscillations and tracking errors, highlighting the necessity of absolute state feedback for long-horizon spatial consistency. The variant with fixed $\alpha = 1.0$ (*w/ Fixed* α) exhibits degraded performance in tasks requiring large reaching motions, as the policy cannot exploit waist redundancy to control the CoM, thereby lacking the ability to dynamically trade off between reachability and stability. Results for additional geometric trajectories (circle, spiral, rectangle) are provided in the Appendix.

2) Base Mobility Assessment: To isolate the robot's navigation performance, we evaluate its ability to reposition its support base in response to task-space requirements. We initialize targets in eight cardinal and ordinal directions around the robot at a distance of 5 m. Figure 5 (left) illustrates the top-view trajectories of the robot base for HiWET and its ablated variants.

HiWET exhibits the most direct and stable paths toward the targets, demonstrating efficient whole-body coordination where locomotion is actively guided by the global task objective.

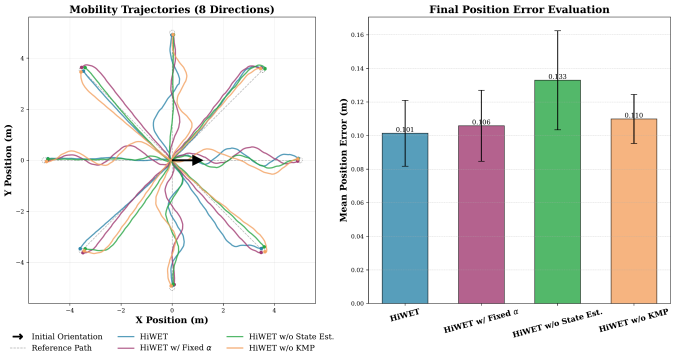

Fig. 5. **Analysis of base mobility and positioning precision.** (Left) Top-view trajectories for 8-directional base repositioning tasks. (Right) Evaluation of final base positioning error in the xy-plane at the target location. HiWET provides the most stable trajectories and the highest positioning accuracy among all variants (w/ Fixed α, w/o State Est., w/o KMP).

In contrast, variants lacking high-fidelity feedback (*w/o State Est.*), kinematic priors (*w/o KMP*), or adaptive α modulation (*w/ Fixed* α) show oscillations or "weaving" behaviors, as the high-level policy struggles to reconcile conflicting base and arm objectives.

We further quantify the positioning precision by measuring the final distance between the robot base and the target starting point in the xy-plane upon task completion. As shown in Fig. 5 (right), HiWET achieves the lowest mean error (0.101 m) and significantly less variance than other methods. This high positioning accuracy is a direct consequence of our active whole-body coordination strategy, which enables the policy to treat the base as an active degree of freedom for workspace optimization rather than an independent mobile platform.

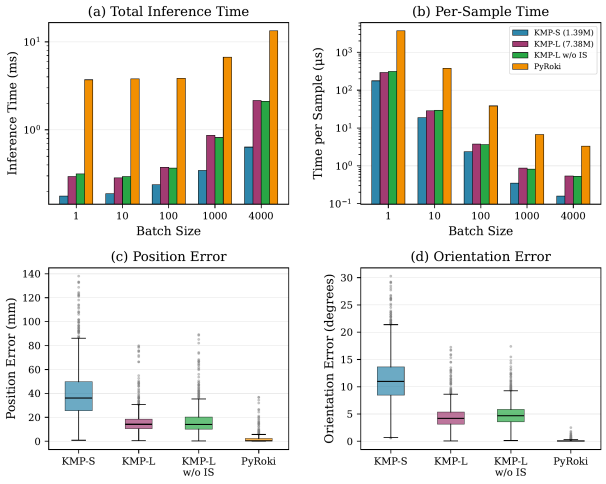

Fig. 6. **Performance comparison between the proposed KMP variants and the optimization-based PyRoki solver.** (a)-(b) show that KMP variants achieve orders of magnitude speedup in inference latency, which is critical for real-time hierarchical control. (c)-(d) demonstrate that while optimization-based methods are near-exact, our KMP-L model trained with importance sampling (IS) provides sufficiently high precision for dual-arm coordination, serving as an efficient kinematic prior.

D. Kinematic Manifold Prior Benchmarking

To evaluate the proposed kinematic prior, we benchmark the performance of the KMP against an optimization-based solver, PyRoki [18], across two primary metrics: *inference latency* for efficiency and *Cartesian position/orientation error* for reconstruction accuracy. We investigate two architectural variants: **KMP-S**, a lightweight 4-layer ResNet with smaller hidden dimensions, and **KMP-L**, a deeper 5-layer ResNet with larger hidden dimensions. Furthermore, we examine the impact of training distributions by comparing models trained on our importance-sampled (IS) dataset against those trained on a uniform Cartesian cube (w/o IS) with an equivalent number of samples.

1) Inference Efficiency: As shown in Fig. 6(a)-(b), the KMP variants achieve significant computational speedups. Compared to the PyRoki solver configured with 5 maximum iterations, KMP-L exhibits over $5\times$ speedup at a single-sample inference level. This performance gap widens drastically under parallelized execution; at a batch size of 4000, KMP-L maintains millisecond-level latency. This efficiency enables seamless integration into the reinforcement learning loop, providing high-frequency kinematic references without imposing a bottleneck on training throughput.

2) Reconstruction Accuracy: We evaluate the precision of the kinematic mapping using a test set derived from AMASS motion capture data [25] retargeted to the G1 humanoid, which captures the natural spatial distribution of human end-effector usage. Figures 6(c) and (d) summarize the Cartesian position and orientation errors. While PyRoki remains the most precise due to its iterative nature, KMP-L achieves a median position error below 15 mm and orientation error below 5 degrees. KMP-L trained with importance sampling (IS) outperforms its uniform counterpart (KMP-L w/o IS), showing the value of functional, manipulable workspace samples for bimanual coordination. These results validate that KMP-L provides a sufficiently accurate and highly efficient kinematic prior for whole-body control.

E. Real-world Hardware Experiments

1) Whole-body Tracking Policy: To evaluate the transferability and robustness of the proposed framework, we deploy the low-level tracking policy on the physical Unitree G1 robot. The policy is trained entirely in simulation and deployed via zero-shot transfer to the robot's onboard computer, operating at 50 Hz.

As illustrated in Fig. 7, the robot demonstrates stable locomotion under varying and asymmetric limb configurations. In Fig. 7(a), the robot maintains a steady walking gait while the right arm is commanded to a high reaching pose. In Fig. 7(b), the robot navigates while the right hand traces a circle, showing robustness to rapid upper-body motions, CoM shifts, and limb coupling. These experiments confirm that the low-level policy effectively manages the trade-off between task-space tracking and dynamic balance in real-world scenarios.

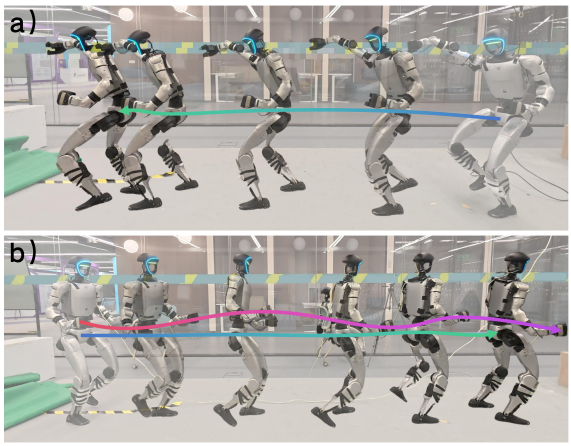

Fig. 7. **Snapshots of real-world deployment on the Unitree G1 robot.** The low-level policy maintains stable locomotion while tracking diverse arm motions: (a) unilateral right-arm reaching, and (b) dynamic upper-body motion with the right hand tracing a circular trajectory (magenta curve indicates the right end-effector path). The system successfully rejects dynamic disturbances caused by rapid limb movements.

TABLE II
REAL-WORLD END-EFFECTOR TRACKING ERRORS FOR CIRCLE AND SQUARE TRAJECTORY TASKS.

Method	Circle RMSE (m)	Square RMSE (m)
HiWET	0.012 ± 0.005	0.015 ± 0.007
HiWET w/ Fixed α	0.018 ± 0.008	0.019 ± 0.009
HiWET w/o State Est.	0.024 ± 0.011	0.028 ± 0.011
HiWET w/o KMP	0.032 ± 0.013	0.039 ± 0.015

2) World-Frame Command Policy: For world-frame tracking, the 50 Hz commander uses the latest Fast-LIO2 global pose [39], transformed to the robot base through forward kinematics; this 10 Hz localization input is held between updates. In the circle and square trajectory tasks (Fig. 1(e,f)), we compare HiWET against ablated variants over 10 repetitions. As shown in Table II, HiWET achieves the lowest tracking errors (12 mm for circles, 15 mm for squares). The fixed $\alpha = 1.0$ variant shows moderate degradation due to its inability to modulate torso engagement, while removing the state estimator or KMP leads to significantly larger errors.

VII. Conclusion And Limitations

We presented HiWET, a hierarchical reinforcement learning framework for world-frame end-effector tracking during humanoid locomotion. By decoupling global task reasoning from dynamic execution, our approach addresses the spatial and dynamic coupling between legged locomotion and precise end-effector tracking. The high-level planner operates in a world coordinate frame to ensure long-horizon consistency, while the low-level policy augmented by an efficient KMP provides the necessary dynamic stabilization. Experiments validate HiWET with 12.4 mm simulation error and robust zero-shot sim-to-real transfer.

The current framework has **limitations**: (1) world-frame tracking precision is bounded by LiDAR-based localization

accuracy; (2) evaluated trajectories are relatively small in scale; (3) bimanual global tracking depends on target feasibility, and high-level experiments involve only single-arm tracking; (4) contact-rich manipulation (e.g., grasping) remains unexplored. Future work will integrate visual perception and force control for dexterous, contact-aware manipulation.

ACKNOWLEDGMENTS

This work was supported by the New Generation Artificial Intelligence-National Science and Technology Major Project (No. 2025ZD0122901) and the National Natural Science Foundation of China (Grant No. 62373242).

REFERENCES

[1] Joao Pedro Araujo, Yanjie Ze, Pei Xu, Jiajun Wu, and C Karen Liu. Retargeting matters: General motion retargeting for humanoid motion tracking. *arXiv preprint arXiv:2510.02252*, 2025.

[2] Qingwei Ben, Feiyu Jia, Jia Zeng, Junting Dong, Dahua Lin, and Jiangmiao Pang. Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[3] Zhanxiang Cao, Yang Zhang, Buqing Nie, Huangxuan Lin, Haoyang Li, and Yue Gao. Learning motion skills with adaptive assistive curriculum force in humanoid robots. *arXiv preprint arXiv:2506.23125*, 2025.

[4] Jin Cheng, Dongho Kang, Gabriele Fadini, Guanya Shi, and Stelian Coros. Rambo: Rl-augmented model-based whole-body control for loco-manipulation. *IEEE Robotics and Automation Letters*, 2025.

[5] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[6] Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

[7] Xinyang Gu, Yen-Jen Wang, and Jianyu Chen. Humanoid-gym: Reinforcement learning for humanoid robot with zero-shot sim2real transfer. *arXiv preprint arXiv:2404.05695*, 2024.

[8] Zhaoyuan Gu, Junheng Li, Wenlan Shen, Wenhao Yu, Zhaoming Xie, Stephen McCrory, Xianyi Cheng, Abdulaziz Shamsah, Robert Griffin, C Karen Liu, et al. Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning. *arXiv preprint arXiv:2501.02116*, 2025.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. *arXiv preprint arXiv:2406.08858*, 2024.

[11] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8944–8951. IEEE, 2024.

[12] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbab, Chaoyi Pan, et al. Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills. *arXiv preprint arXiv:2502.01143*, 2025.

[13] Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, et al. Hover: Versatile neural whole-body controller for humanoid robots. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9989–9996. IEEE, 2025.

[14] Jhen Hsieh, Kuan-Hsun Tu, Kuo-Han Hung, and Tsung-Wei Ke. Dexman: Learning bimanual dexterous manipulation from human and generated videos. *arXiv preprint arXiv:2510.08475*, 2025.

[15] Mazeyu Ji, Xuanbin Peng, Fangchen Liu, Jialong Li, Ge Yang, Xuxin Cheng, and Xiaolong Wang. Exbody2: Advanced expressive humanoid whole-body control. *arXiv preprint arXiv:2412.13196*, 2024.

[16] Haoran Jiang, Jin Chen, Qingwen Bu, Li Chen, Modi Shi, Yanjie Zhang, Delong Li, Chuanzhe Suo, Chuang Wang, Zhihui Peng, et al. Wholebodyvla: Towards unified latent vla for whole-body loco-manipulation control. *arXiv preprint arXiv:2512.11047*, 2025.

[17] Xisheng Jiang, Baolei Wu, Simin Li, Yongtong Zhu, Guoxiang Liang, Ye Yuan, Qingdu Li, and Jianwei Zhang. Multi-humanoid robot arm motion imitation and collaboration based on improved retargeting. *Biomimetics*, 10(3):190, 2025.

[18] Chung Min Kim, Brent Yi, Hongsuk Choi, Yi Ma, Ken Goldberg, and Angjoo Kanazawa. Pyroki: A modular toolkit for robot kinematic optimization. *arXiv preprint arXiv:2505.03728*, 2025.

[19] Yuxuan Kuang, Haoran Geng, Amine Elhafsi, Tan-Dzung Do, Pieter Abbeel, Jitendra Malik, Marco Pavone, and Yue Wang. Skillblender: Towards versatile humanoid whole-body loco-manipulation via skill blending. *arXiv preprint arXiv:2506.09366*, 2025.

[20] Jialong Li, Xuxin Cheng, Tianshu Huang, Shiqi Yang, Ri-Zhao Qiu, and Xiaolong Wang. Amo: Adaptive motion optimization for hyper-dexterous humanoid whole-body control. *arXiv preprint arXiv:2505.03738*, 2025.

[21] Yitang Li, Yuanhang Zhang, Wenli Xiao, Chaoyi Pan, Haoyang Weng, Guanqi He, Tairan He, and Guanya Shi. Hold my beer: Learning gentle humanoid locomotion and end-effector stabilization control. In *RSS 2025 Workshop on Whole-body Control and Bimanual Manipulation: Applications in Humanoids and Beyond*, 2025.

[22] Yixuan Li, Yutang Lin, Jieming Cui, Tengyu Liu, Wei

Liang, Yixin Zhu, and Siyuan Huang. Clone: Closed-loop whole-body humanoid teleoperation for long-horizon tasks. *arXiv preprint arXiv:2506.08931*, 2025.

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

[24] Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Sirui Chen, Fernando Castañeda, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, et al. Sonic: Supersizing motion tracking for natural humanoid whole-body control. *arXiv preprint arXiv:2511.07820*, 2025.

[25] 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.

[26] Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Muñoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, et al. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025.

[27] Chaoyi Pan, Changhao Wang, Haozhi Qi, Zixi Liu, Homanga Bharadhwaj, Akash Sharma, Tingfan Wu, Guanya Shi, Jitendra Malik, and Francois Hogan. Spider: Scalable physics-informed dexterous retargeting. *arXiv preprint arXiv:2511.09484*, 2025.

[28] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018.

[29] Tifanny Portela, Andrei Cramariuc, Mayank Mittal, and Marco Hutter. Whole-body end-effector pose tracking. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11205–11211. IEEE, 2025.

[30] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on robot learning*, pages 91–100. PMLR, 2022.

[31] Thushara Sandakalum and Marcelo H Ang Jr. Motion planning for mobile manipulators—a systematic review. *Machines*, 10(2):97, 2022.

[32] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[33] Carmelo Sferrazza, Dun-Ming Huang, Xingyu Lin, Youngwoon Lee, and Pieter Abbeel. Humanoidbench: Simulated humanoid benchmark for whole-body locomotion and manipulation. *arXiv preprint arXiv:2403.10506*, 2024.

[34] Yiyang Shao, Xiaoyu Huang, Bike Zhang, Qiayuan Liao, Yuman Gao, Yufeng Chi, Zhongyu Li, Sophia Shao, and Koushil Sreenath. Langwbc: Language-directed humanoid whole-body control via end-to-end learning. *arXiv preprint arXiv:2504.21738*, 2025.

[35] Chen Tessler, Yifeng Jiang, Erwin Coumans, Zhengyi Luo, Xue Bin Peng, and Gal Chechik. Maskedmanipulator: Versatile whole-body control for loco-manipulation. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*, pages 1–11, 2025.

[36] Chenzheng Wang, Qiang Huang, Xuechao Chen, Zeyu Zhang, and Jing Shi. Robust visuomotor control for humanoid loco-manipulation using hybrid reinforcement learning. *Biomimetics*, 10(7):469, 2025.

[37] Zifan Wang, Yufei Jia, Lu Shi, Haoyu Wang, Haizhou Zhao, Xueyang Li, Jinni Zhou, Jun Ma, and Guyue Zhou. Arm-constrained curriculum learning for loco-manipulation of a wheel-legged robot. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10770–10776. IEEE, 2024.

[38] Zifan Wang, Teli Ma, Yufei Jia, Xun Yang, Jiaming Zhou, Wenlong Ouyang, Qiang Zhang, and Junwei Liang. Omni-perception: Omnidirectional collision avoidance for legged locomotion in dynamic environments. *arXiv preprint arXiv:2505.19214*, 2025.

[39] Wei Xu, Yixi Cai, Dongjiao He, Jiarong Lin, and Fu Zhang. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics*, 38(4):2053–2073, 2022.

[40] Lujie Yang, Xiaoyu Huang, Zhen Wu, Angjoo Kanazawa, Pieter Abbeel, Carmelo Sferrazza, C Karen Liu, Rocky Duan, and Guanya Shi. Omniretarget: Interaction-preserving data generation for humanoid whole-body loco-manipulation and scene interaction. *arXiv preprint arXiv:2509.26633*, 2025.

[41] Shaofeng Yin, Yanjie Ze, Hong-Xing Yu, C Karen Liu, and Jiajun Wu. Visualmimic: Visual humanoid loco-manipulation via motion tracking and generation. *arXiv preprint arXiv:2509.20322*, 2025.

[42] Yanjie Ze, Siheng Zhao, Weizhuo Wang, Angjoo Kanazawa, Rocky Duan, Pieter Abbeel, Guanya Shi, Jiajun Wu, and C Karen Liu. Twist2: Scalable, portable, and holistic humanoid data collection system. *arXiv preprint arXiv:2511.02832*, 2025.

[43] Yang Zhang, Zhanxiang Cao, Buqing Nie, Haoyang Li, Zhong Jiangwei, Qiao Sun, Xiaoyi Hu, Xiaokang Yang, and Yue Gao. Keep on going: Learning robust humanoid motion skills via selective adversarial training. *arXiv preprint arXiv:2507.08303*, 2025.

[44] Yuanhang Zhang, Yifu Yuan, Prajwal Gurunath, Ishita Gupta, Shayegan Omidshafiei, Ali-akbar Aghamohammadi, Marcell Vazquez-Chanlatte, Liam Pedersen, Tairan He, and Guanya Shi. Falcon: Learning force-adaptive humanoid loco-manipulation. *arXiv preprint arXiv:2505.06776*, 2025.

[45] Zhikai Zhang, Chao Chen, Han Xue, Jilong Wang, Sikai Liang, Yun Liu, Zongzhang Zhang, He Wang, and Li Yi. Unleashing humanoid reaching potential via real-world-ready skill space. *arXiv preprint arXiv:2505.10918*, 2025.

[46] Zhikai Zhang, Jun Guo, Chao Chen, Jilong Wang, Chenghuai Lin, Yunrui Lian, Han Xue, Zhenrong Wang, Maoqi Liu, Jiangran Lyu, et al. Track any motions under any disturbances. *arXiv preprint arXiv:2509.13833*, 2025.