2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

KungfuBot2: Learning Versatile Motion Skills for Humanoid Whole-Body Control

Jinrui Han[1,2] Weiji Xie[1,2] Jiakun Zheng[1,3] Jiyuan Shi[1] Weinan Zhang[2] Ting Xiao[3] Chenjia Bai[†1]

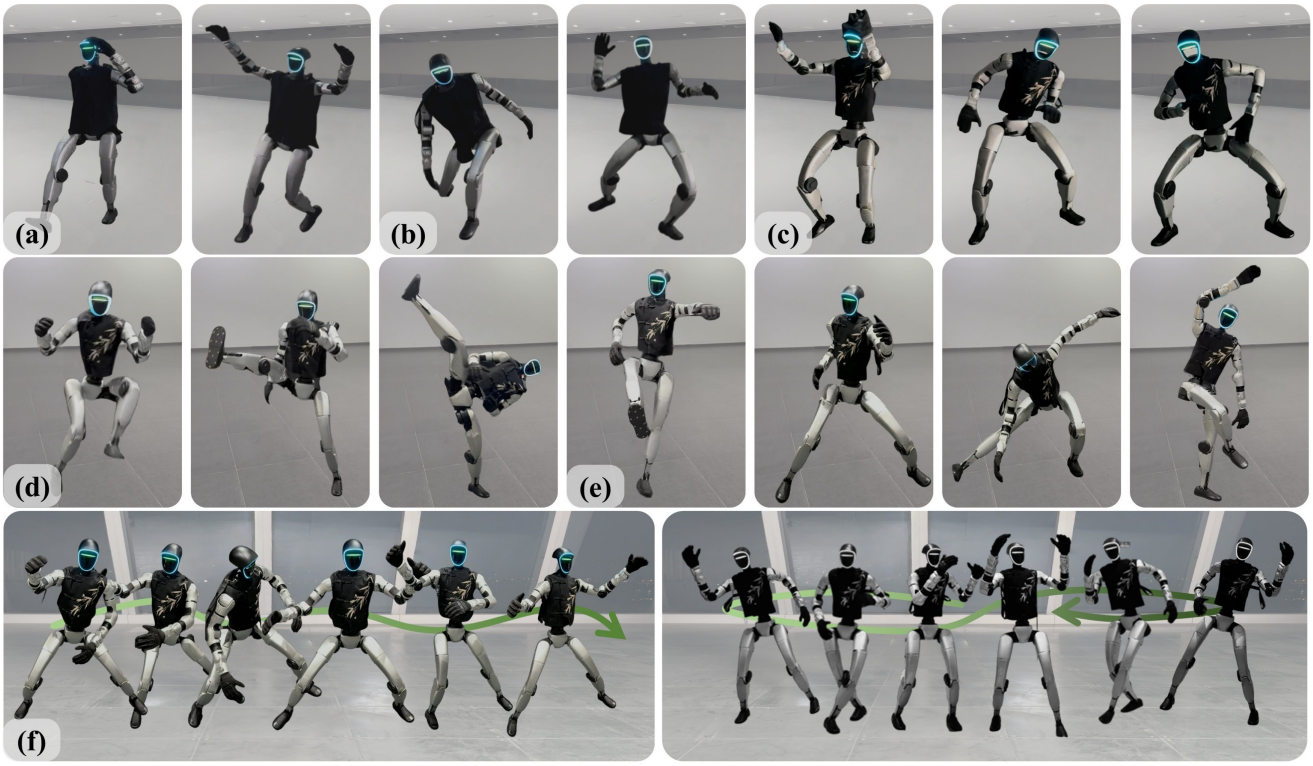

Fig. 1: **Humanoid learning versatile motion skills.** We deploy VMS on the Unitree G1 humanoid robot, demonstrating its capability to perform a broad category of motion skills with strong stability and generalization. The repertoire includes (a) walking and running, (b) ball throwing and racket swinging, (c) dancing, (d) diverse kicking, (e) Kung Fu and (f) long sequences of martial arts and dance.

***Abstract*— Learning versatile whole-body skills by tracking various human motions is a fundamental step toward general-purpose humanoid robots. This task is particularly challenging because a single policy must master a broad repertoire of motion skills while ensuring stability over long-horizon sequences. To this end, we present VMS, a unified whole-body controller that enables humanoid robots to learn diverse and dynamic behaviors within a single policy. Our framework integrates a hybrid tracking objective that balances local motion fidelity with global trajectory consistency, and an Orthogonal Mixture-of-Experts (OMoE) architecture that encourages skill specialization while enhancing generalization across motions. A segment-level tracking reward is further introduced to relax rigid step-wise matching, enhancing robustness when handling global displacements and transient inaccuracies. We validate VMS extensively in both simulation and real-world experiments, demonstrating accurate imitation of dynamic skills, stable performance over minute-long sequences, and strong generalization to unseen motions. These results highlight the potential of VMS as a scalable foundation for versatile humanoid whole-body control. The project page is available at kungfubot2-humanoid.github.io.**

†Corresponding Author
[1]Institute of Artificial Intelligence (TeleAI), China Telecom, [2]Shanghai Jiao Tong University, [3]East China University of Science and Technology

I. Introduction

Humanoid robots hold great potential for imitating human behaviors, spanning stable locomotion to agile, complex motions. Realizing this capability requires a universal whole-body controller that can generalize across versatile skills. Recent advances in motion capture (Mocap) systems have enabled the collection of large-scale human motion datasets, offering rich resource for developing such controllers.

Existing research [1], [2], [3], [4] in physics-based character animation has explored learning-based methods for building controllers that mimic versatile, human-like behaviors in simulation. Motivated by these advances, recent work has extended this paradigm to humanoid robots [5], [6], [7],

[8], [9], addressing challenges such as partial observability [5], [10], physical plausibility [11], and the sim-to-real gap [12]. However, achieving high-fidelity imitation often requires training separate policies for each motion [11], [12], [13], which limits generalization. Several methods attempt to learn a single policy for multiple motions, but these approaches are constrained by limited policy expressiveness, typically relying on a single MLP network [14], [15], and lack mechanisms to balance local and global tracking objectives [16], [17]. Specifically, local tracking (e.g., velocities or relative keybody poses) can reduce error accumulation but may compromise global stability [17], whereas global tracking ensures overall coherence but is prone to long-horizon drift [11]. Thus, two central challenges remain: enhancing policy expressiveness for versatile skill learning, and reconciling local and global tracking to achieve both motion fidelity and long-horizon stability.

To this end, we present VMS, a universal whole-body controller that enables humanoid robots to learn versatile motion skills. VMS first introduces a hybrid tracking objective that preserves local motion pose while mitigating global drift, addressing the instability of purely local or global tracking. Second, an Orthogonal Mixture-of-Experts (OMoE) architecture disentangles skill representations, improving policy expressiveness and reducing overlap between skill representations. Third, a segment-level tracking reward further improves robustness by relaxing rigid step-wise matching, enabling stable long-horizon motion execution. Extensive experiments demonstrate that VMS performs dynamic skills with high fidelity and sustains stable tracking over minute-level sequences. Our main contributions are:

- We propose an OMoE architecture that disentangles motion representations, improving policy expressiveness and generalization across diverse skills.
- We introduce a hybrid tracking objective together with a segment-level reward, balancing motion style fidelity and long-horizon stability.
- We validate VMS on extensive simulated and real-world experiments, achieving robust tracking of both high-dynamic and minute-level motion sequences.

II. Related Work

A. Humanoid Whole-Body Control

Traditional model-based approaches to humanoid whole-body control rely on accurate dynamics models for precise task execution [18], [19], but they demand complex modeling effort and extensive manual tuning across diverse skills. Learning-based approaches, in contrast, typically depend on manually crafted, task-specific rewards. While such methods have been successfully applied to locomotion on challenging terrain [20], [21], [22], jumping [23], parkour [24], and fall recovery [25], [26], each task demands extensive reward engineering, and generating human-like motions remains difficult [27]. To handle the distinct objectives of upper and lower body control, some works decompose the solution into separate policies [28], [29], [30], though this limits coordination and generalization. Others tackle complex tasks such as table tennis via hierarchical planning and learning [31].

In contrast, whole-body motion tracking directly leverages human motion data as a reference, providing a unified control objective for the entire body: reproducing the reference motion. This eliminates the need for task-specific reward design and naturally encourages human-like coordination and expressiveness across a wide range of skills.

B. Humanoid Motion Tracking

Humanoid motion tracking aims to learn lifelike behaviors directly from human motion data. DeepMimic [32] pioneers a phase-based tracking framework that combines random state initialization with early termination to imitate individual motions. To bridge the sim-to-real gap, ASAP [12] proposes a multistage training pipeline with a delta-action model for dynamic skills. HuB [33] and KungfuBot [11] employ elaborate motion processing and tracking mechanisms, achieving accurate imitation of highly dynamic single motions.

For learning diverse humanoid motions within a single policy, OmniH2O [5] introduces a universal controller that inspired subsequent humanoid works. ExBody2 [10] improves expressiveness by decomposing tracking targets and applying motion filtering. TWIST [14] and CLONE [16] enable high-quality tracking but are tailored to teleoperation settings and focus on relatively low-dynamic motions. BumbleBee [34] adopts a two-stage strategy: clustering motions and training separate experts, followed by policy distillation. GMT [17] achieves robust tracking of highly dynamic motions by emphasizing root velocity and pose over global positions. UniTracker [15] supports dynamic movements, though its reliance on global targets limits stability in executing long-sequence motions. More recently, BeyondMimic [13] demonstrates high-fidelity tracking of single motions through well-designed objectives and precise system identification, and uses a distilled unified diffusion policy for task-specific control. Building on these advances, our work focuses on learning a single universal policy capable of reproducing diverse, long-sequence motions with both local style fidelity and global stability.

III. Method

A. Problem Definition

In this work, we adopt the Unitree G1 robot [35], controlling 23 degrees of freedom (DoF), excluding the three DoFs of each wrist. We formulate humanoid motion tracking as a goal-conditioned reinforcement learning (RL) problem, where the agent interacts with the environment according to a policy π to maximize cumulative reward. At each timestep t, the policy receives the state $\boldsymbol{s}_t$ and target state $\boldsymbol{g}_t$, and outputs an action $\boldsymbol{a}_t \in \mathbb{R}^{23}$, which is subsequently converted into motor torques via a PD controller. This defines the policy as $\pi(\boldsymbol{a}_t|\boldsymbol{s}_t, \boldsymbol{g}_t)$. The environment dynamics $p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t)$ determine the next state, and a dense reward $r(\boldsymbol{s}_t, \boldsymbol{g}_t, \boldsymbol{a}_t)$ evaluates tracking performance while providing regularization. The agent's objective is to maximize the expected discounted return $J = \mathbb{E}\Big[\sum_{t=0}^{T-1} \gamma^t r_t\Big]$.

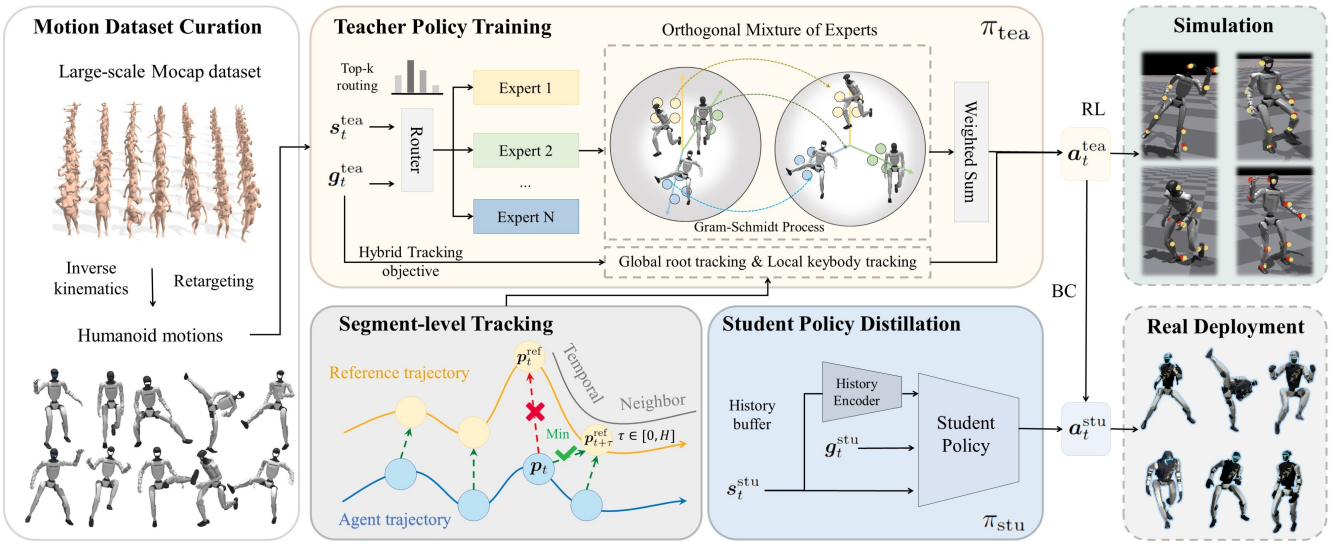

Fig. 2: **Framework of VMS.** The large-scale Mocap dataset is first retargeted to the humanoid skeleton using an IK-based method. A teacher policy π_{tea} is trained with a hybrid tracking objective, enhanced by a segment-level reward for long-horizon robustness. The student policy π_{stu} is distilled from the teacher policy through behavior cloning and deployed on the real humanoid robot.

Following prior work [5], [14], we adopt a two-stage teacher–student learning paradigm to address the challenge of partial observability. In the first stage, an oracle teacher policy π_{tea} is trained with Proximal Policy Optimization (PPO) [36], utilizing full state information. In the second stage, a student policy π_{stu} is trained via behavior cloning (BC) to imitate the teacher, relying only on observations available at deployment. The overall framework of VMS is illustrated in Fig. 2.

B. Motion Dataset Curation

To develop a versatile humanoid motion controller, we first construct a large-scale, high-quality human motion dataset as the training source. Starting from the publicly available AMASS Mocap dataset [37] in SMPL format, we retarget human motions to the humanoid skeleton using a differentiable inverse kinematics (IK)-based method [38], [39]. This procedure formulates a differentiable optimization problem, ensuring end-effector trajectory alignment under the constraints of joint limits, yielding over 13,000 sequences.

The raw retargeted dataset, however, contains non-flat-ground motions for the humanoid (e.g., stair climbing) and sequences with corrupted frames (e.g., severe penetrations or discontinuities [40]). To curate the data, we train an oracle policy (see Section III-D) on the full set and evaluate each sequence, filtering out invalid ones. The final dataset comprises 9,770 high-quality motions, totaling 30.41 hours of humanoid-compatible data.

C. Policy Objective

1) Hybrid Tracking Targets: To guide the policy toward both expressive motion style and accurate spatial alignment, we adopt a hybrid tracking objective that combines global root tracking with local keybody tracking [10], [13].

For global root tracking, we denote the root's position and rotation in the world frame as $\boldsymbol{p}$ and $\boldsymbol{r}$. The policy seeks to minimize deviations from the reference root trajectory, ensuring overall spatial alignment across the motion.

Specifically, we define the set of keybody parts as $\mathcal{K}$, including the head, hands, elbows, knees, and ankles. The local tracking targets, denoted as keybody positions $\boldsymbol{p}^{\mathcal{K}}$ and orientations $\boldsymbol{r}^{\mathcal{K}}$, are obtained by aligning the positions and orientations of these keybodies relative to the root between the current robot state and the reference motion [13]. This encourages the policy to capture the local motion style demonstrated in the reference. By defining both global and local tracking targets in this way, the policy is guided to leverage precise reference information, balancing local motion fidelity and global spatial consistency.

2) State Space Design: At each timestep t, the teacher policy π_{tea} observes a state $\boldsymbol{s}_t^{\text{tea}}$ and a goal $\boldsymbol{g}_t^{\text{tea}}$:

$$\boldsymbol{s}_t^{\text{tea}} = \left[\boldsymbol{q}_t, \dot{\boldsymbol{q}}_t, \boldsymbol{v}_t, \boldsymbol{\omega}_t, \boldsymbol{p}_t^{\mathcal{K}}, \boldsymbol{r}_t^{\mathcal{K}}, \Delta\boldsymbol{p}_t, \Delta\boldsymbol{r}_t, \boldsymbol{a}_{t-1}\right], \quad (1)$$

$$\boldsymbol{g}_t^{\text{tea}} = \left[\boldsymbol{q}_{t+1:t+H}^{\text{ref}}, \dot{\boldsymbol{q}}_{t+1:t+H}^{\text{ref}}, \boldsymbol{p}_{t+1:t+H}^{\mathcal{K},\text{ref}}, \boldsymbol{r}_{t+1:t+H}^{\mathcal{K},\text{ref}}\right], \quad (2)$$

where $\boldsymbol{q}_t \in \mathbb{R}^{23}$ and $\dot{\boldsymbol{q}}_t \in \mathbb{R}^{23}$ denote the joint positions and velocities, $\boldsymbol{v}_t \in \mathbb{R}^3$ and $\boldsymbol{\omega}_t \in \mathbb{R}^3$ are the root's linear and angular velocities, and $\boldsymbol{a}_{t-1} \in \mathbb{R}^{23}$ is the previous action. $\Delta\boldsymbol{p}_t \in \mathbb{R}^3$ and $\Delta\boldsymbol{r}_t \in \mathbb{R}^6$ [41] represent the root's position and rotation tracking errors. The goal $\boldsymbol{g}_t^{\text{tea}}$ provides a preview of reference joint positions, velocities, and relative keybody poses over the future H timesteps.

In real-world deployment, global root position and velocity are unavailable. Therefore, the student policy π_{stu} relies on proprioceptive states, augmented with a short history of past observations. Its state $\boldsymbol{s}_t^{\text{stu}}$ and goal $\boldsymbol{g}_t^{\text{stu}}$ are defined as:

$$\boldsymbol{s}_t^{\text{stu}} = \left[\boldsymbol{q}_{t-K:t}, \dot{\boldsymbol{q}}_{t-K:t}, \boldsymbol{\omega}_{t-K:t}, \Delta\boldsymbol{r}_{t-K:t}, \boldsymbol{a}_{t-K:t-1}\right], \quad (3)$$

$$\boldsymbol{g}_t^{\text{stu}} = \left[\boldsymbol{q}_{t+1:t+H}^{\text{ref}}, \dot{\boldsymbol{q}}_{t+1:t+H}^{\text{ref}}, \boldsymbol{r}_{t+1:t+H}^{\mathcal{K},\text{ref}}\right], \quad (4)$$

where K is the number of past timesteps included in the student state, enabling the policy to leverage historical proprioceptive information to compensate for the partial observability. By defining goal states for both the teacher and the student, policies can plan actions conditioned on future motion, improving tracking accuracy and expressiveness.

D. Policy Learning Framework

1) Orthogonal Mixture of Experts: Learning versatile motion skills can be considered as a multi-task RL problem, where the policy must imitate diverse motions simultaneously. A single MLP often struggles, as it mixes multiple motions into overlapping representations, hindering effective skill learning. For instance, AMASS dataset [37] span categories such as walking, running, kicking, squatting, and dancing, each with distinct dynamics patterns. Collapsing them into a single shared representation often leads to instability and limited expressiveness [17]. To address this, we introduce an *Orthogonal Mixture-of-Experts* (OMoE) architecture for the teacher policy. The OMoE module consists of multiple expert networks whose outputs are constrained to be orthogonal, along with a router network that dynamically selects experts based on the current state.

At each timestep t, the input to OMoE is the concatenation of the state and goal $(\boldsymbol{s}_t, \boldsymbol{g}_t)$. A set of M experts $\{h_i\}_{i=1}^{M}$ maps the input to feature vectors:

$$U_t = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_M] \in \mathbb{R}^{d\times M}, \quad \boldsymbol{u}_i = h_i(\boldsymbol{s}_t, \boldsymbol{g}_t). \tag{5}$$

To encourage diversity, we impose an orthogonality constraint on the output features of different experts:

$$U_t^\top U_t = I_M, \tag{6}$$

ensuring that each $\boldsymbol{u}_i$ represents an independent direction in feature space. Directly solving this constrained optimization is challenging, so we instead approximate it via the Gram–Schmidt (GS) process [42], [43]. GS maps U_t into an orthogonal basis $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M\}$ by sequentially removing projections onto previously obtained vectors:

$$\boldsymbol{v}_i = \boldsymbol{u}_i - \sum_{j=1}^{i-1} \frac{\langle \boldsymbol{v}_j, \boldsymbol{u}_i\rangle}{\langle \boldsymbol{v}_j, \boldsymbol{v}_j\rangle}\, \boldsymbol{v}_j, \quad i = 1, \ldots, M. \tag{7}$$

This ensures that different experts provide mutually diverse representations. In practice, the differentiable nature of GS encourages the experts to learn diverse features, while normalization after projection further stabilizes training.

A router network computes the expert weights as $\boldsymbol{\alpha} = \mathrm{Router}(\boldsymbol{s}_t, \boldsymbol{g}_t)$ and assigns the resulting coefficients α_i to their orthogonalized features. The weighted combination is passed through a shared output head f, producing the teacher action $\boldsymbol{a}_t^{\mathrm{tea}}$:

$$\boldsymbol{a}_t^{\mathrm{tea}} = f\left(\sum_i \alpha_i \boldsymbol{v}_i\right). \tag{8}$$

With the design of OMoE, the teacher policy decomposes diverse motion skills into orthogonal representations, and the router adaptively recombines them, allowing flexible composition of versatile skills.

2) History Encoding and Student Distillation: During teacher training, a short temporal history of the robot's proprioceptive states is encoded by a convolutional network into a compact latent embedding. This embedding is trained to approximate privileged, simulation-only randomized physical parameters, such as motor strengths, friction coefficients, and base mass variations [44], [45], [46], [47]. In this way, the latent captures environment dynamics and physical variations that are unavailable during real-world deployment.

In the student training stage, this history encoder is frozen and applied to real-world history observations, producing latent embeddings that enhance robustness to system noise and parameter uncertainty. Leveraging this representation enables the student to more effectively imitate the teacher. The student policy is then distilled via DAgger [48], minimizing the $\mathcal{L}_2$ loss between the student's output actions $\boldsymbol{a}_t^{\mathrm{stu}}$ and the teacher's actions via $\|\boldsymbol{a}_t^{\mathrm{tea}} - \boldsymbol{a}_t^{\mathrm{stu}}\|_2^2$.

E. Segment-level Tracking Reward

Strict step-wise tracking of the reference trajectory often fails in long-horizon motion tracking due to error accumulation or infeasible reference states. For example, when the reference requires overly aggressive velocities or the robot is disturbed by external forces, one-to-one state matching may destabilize the policy. While global tracking emphasizes overall trajectory coverage and local tracking preserves short-term style, both objectives can be overly restrictive when the reference becomes impractical to follow.

To mitigate this, we utilize a *segment-level tracking reward* that relaxes rigid step-wise tracking. As illustrated in Fig. 2, instead of enforcing alignment at a single timestep t, the agent is rewarded according to the minimum discrepancy between its state and the candidate reference states within a short temporal neighborhood. For global tracking, the reward r_t^{global} is defined as

$$r_t^{\mathrm{global}} = \exp\Big(-\min_{\tau\in[0,H]} d_{\mathrm{global}}(\boldsymbol{p}_t, \boldsymbol{p}_{t+\tau}^{\mathrm{ref}})\Big), \tag{9}$$

where H is the small future horizon and d_{global} measures root position deviation. This allows the agent to catch up within a short future window, reducing sensitivity to transient errors.

Similarly, for local keybody tracking, we define

$$r_t^{\mathrm{local}} = \exp\Big(-\min_{\tau\in[0,H]} d_{\mathrm{local}}(\boldsymbol{p}_t^{\mathcal{K}}, \boldsymbol{p}_{t+\tau}^{\mathcal{K},\mathrm{ref}})\Big), \tag{10}$$

where d_{local} measures deviations in local keybody positions.

The segment-level design mitigates the limitations of strict step-wise matching by allowing the policy to align with the most feasible reference within a short horizon. This alleviates excessive penalties from temporary deviations or infeasible local targets, while jointly balancing local style preservation and global trajectory consistency, leading to more robust long-horizon tracking. The detailed reward terms and weights are summarized in Table I.

F. Sim2Real Transfer

To improve policy robustness and facilitate effective sim-to-real transfer, we employ domain randomization (DR) [49]

TABLE I: Reward terms and weights.

Tracking terms	Weights	Regularization terms	Weights
Root position	1.5	Action rate	-0.1
Root velocity	1.5	Feet slip	-0.1
Root rotation	1.5	Torque limits	-5.0
Key body position	3.0	Joint limits	-10
Key body velocity	2.0	Joint velocities	$-1e{-}4$
Key body rotation	2.0	Joint accelerations	$-3e{-}7$

during both teacher and student training. Additionally, we randomize default joint positions and perturb the robot root with additional linear and angular velocities.

Furthermore, we adopt a noised reference state initialization (RSI) strategy [32], [50], injecting Gaussian noise into the root and joint states of the sampled motion states. This enhances the policy's resilience to imperfect intermediate states and environmental variations. The detailed configurations are provided in Table II.

TABLE II: Randomization parameters.

DR terms	Range	Noised RSI	Noise scale
Base Mass	$[-3, 3]$	Joint position	0.1
Friction	$[0.1, 1.6]$	Root position	0.05
Motor Strength	$[0.9, 1.1]$	Root velocity	0.2
Default joint pos	$[-0.01, 0.01]$	Root rotation	0.1
Push Robot	$[-0.5, 0.5]$	Root angular vel	0.5

IV. EXPERIMENTS

In our experiments, we aim to answer the following key research questions:

- **Q1.** Can VMS outperforms baseline methods when trained on large-scale motion datasets?
- **Q2.** Does OMoE enables effective learning and generalization of versatile motion skills?
- **Q3.** Can the segment-level reward improve long-horizon motion performance?
- **Q4.** How well does VMS perform in real-world?

A. Experiment Setup

We evaluate the performance of VMS in both simulation and real world. In simulation, we train policies on the curated training dataset and assess generalization on a test dataset, including the AMASS test dataset [40], the LAFAN1 dataset [13], and additional in-house Mocap motions. Each policy is trained in IsaacGym [51] simulator with 4,096 parallel environments, and for evaluation we sample 1,000 rollout trajectories per motion.

1) Baselines: We compare VMS with two baselines, ExBody2 [10] and GMT [17]. For fairness, we re-implement both methods on our curated training dataset and conducted evaluations on the teacher policies, as the student policies are purely distilled from the teacher through behavior cloning.

2) Metrics: Tracking performance is evaluated using the following metrics: Mean Per Keybody Position Error (E_{mpkpe}, mm), Mean Per Joint Position Error (E_{mpjpe}, rad), Mean Global Root Position Error (E_{pos}, mm) and Mean Root Linear Velocity Error (E_{vel}, m/s). For the test dataset, we report the success rate (%), where a motion is considered successful if the humanoid completes the entire sequence.

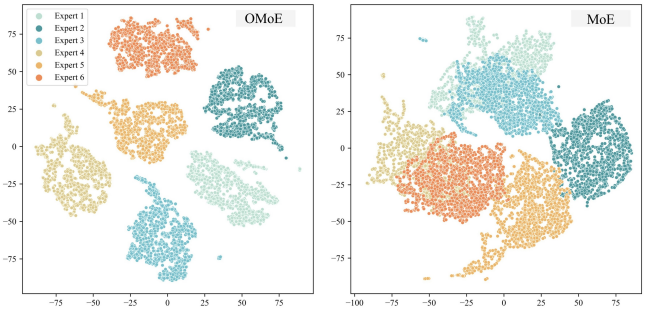

Fig. 3: **t-SNE visualization of experts' output features.** Standard MoE exhibits substantial overlap across experts, while OMoE produces more diverse and specialized skill subspaces.

B. Main Results

To address **Q1**, Table III shows that VMS outperforms ExBody2 and GMT across all metrics. It achieves notably better local tracking performance, while the inclusion of a global root objective effectively reduces root position errors. On the test dataset, although ExBody2 and GMT report relatively high success rates, they replace global position tracking with root velocity objectives, which amplifies global drift and leads to larger position errors. In contrast, VMS substantially reduces both global and local tracking errors, demonstrating its accuracy and stability.

C. Ablation Studies

1) Ablations on OMoE: To investigate **Q2**, we conduct ablation studies by comparing OMoE with a standard soft MoE [52] and an MLP baseline of the same parameter size. As shown in Table III, OMoE consistently achieves the lowest tracking errors, followed by MoE, while MLP performs the worst. This demonstrates that introducing orthogonalized experts leads to more effective motion tracking and better generalization. In addition to quantitative metrics, we further analyze the learned representations from the experts. Specifically, we visualize the expert output features using t-SNE, as shown in Fig. 3. For the standard MoE, these features exhibit substantial overlap, indicating redundancy and limited differentiation among experts. In contrast, OMoE structure produces more dispersed and well-separated features, suggesting that the orthogonalization constraint encourages experts to capture distinct skill subspaces, thereby improving diversity and representation efficiency.

To further probe this specialization, we examine expert activation frequencies across four representative motion categories from AMASS: walking, squatting, kicking, and dancing. As shown in Fig. 4, walking mainly activates a small subset of experts, reflecting its repetitive structure, while dancing exhibits a more evenly distributed activation pattern due to its higher variability. Squatting and kicking fall in between, each with distinct but less uniform patterns. These

TABLE III: Main results on training and test datasets. Results are reported as mean ± one standard deviation.

	Training Dataset				Test Dataset				
Method	$E_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E_{\text{pos}} \downarrow$	$E_{\text{vel}} \downarrow$	$Succ \uparrow$	$E_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E_{\text{pos}} \downarrow$	$E_{\text{vel}} \downarrow$
Baseline									
ExBody2 [10]	$53.03_{\pm 2.48}$	$0.684_{\pm 0.020}$	$373.9_{\pm 28.1}$	$0.256_{\pm 0.020}$	$90.97_{\pm 0.18}$	$62.78_{\pm 0.805}$	$0.654_{\pm 0.012}$	$468.2_{\pm 40.2}$	$0.416_{\pm 0.028}$
GMT [17]	$45.28_{\pm 1.21}$	$0.526_{\pm 0.011}$	$292.1_{\pm 14.2}$	$0.187_{\pm 0.006}$	$90.07_{\pm 0.15}$	$58.89_{\pm 0.932}$	$0.574_{\pm 0.023}$	$397.1_{\pm 28.7}$	$0.354_{\pm 0.022}$
VMS (ours)	$\mathbf{42.59}_{\pm 1.10}$	$\mathbf{0.499}_{\pm 0.009}$	$\mathbf{48.15}_{\pm 1.26}$	$\mathbf{0.176}_{\pm 0.004}$	$\mathbf{92.68}_{\pm 0.20}$	$\mathbf{52.95}_{\pm 0.497}$	$\mathbf{0.568}_{\pm 0.003}$	$\mathbf{57.92}_{\pm 1.16}$	$\mathbf{0.216}_{\pm 0.002}$
Ablations on OMoE									
VMS-MLP	$49.12_{\pm 1.43}$	$0.583_{\pm 0.020}$	$59.86_{\pm 1.43}$	$0.199_{\pm 0.010}$	$86.67_{\pm 0.34}$	$56.73_{\pm 0.811}$	$0.653_{\pm 0.006}$	$67.21_{\pm 3.92}$	$0.285_{\pm 0.007}$
VMS-MoE	$45.03_{\pm 1.78}$	$0.558_{\pm 0.014}$	$56.97_{\pm 1.56}$	$0.188_{\pm 0.007}$	$88.52_{\pm 0.31}$	$54.86_{\pm 0.517}$	$0.628_{\pm 0.003}$	$59.90_{\pm 1.25}$	$0.250_{\pm 0.003}$
VMS (ours)	$\mathbf{42.59}_{\pm 1.10}$	$\mathbf{0.499}_{\pm 0.009}$	$\mathbf{48.15}_{\pm 1.26}$	$\mathbf{0.176}_{\pm 0.004}$	$\mathbf{92.68}_{\pm 0.20}$	$\mathbf{52.95}_{\pm 0.497}$	$\mathbf{0.568}_{\pm 0.003}$	$\mathbf{57.92}_{\pm 1.16}$	$\mathbf{0.216}_{\pm 0.002}$
Ablations on Segment-level Tracking									
VMS-Global	$50.28_{\pm 1.55}$	$0.562_{\pm 0.022}$	$\mathbf{44.09}_{\pm 2.12}$	$0.180_{\pm 0.010}$	$60.52_{\pm 0.24}$	$69.57_{\pm 0.969}$	$0.576_{\pm 0.005}$	$60.53_{\pm 1.12}$	$0.243_{\pm 0.005}$
VMS-Step-wise	$43.42_{\pm 1.01}$	$0.512_{\pm 0.015}$	$51.15_{\pm 2.43}$	$0.182_{\pm 0.010}$	$84.68_{\pm 0.27}$	$\mathbf{50.63}_{\pm 0.312}$	$0.577_{\pm 0.004}$	$63.92_{\pm 1.00}$	$0.266_{\pm 0.002}$
VMS (ours)	$\mathbf{42.59}_{\pm 1.10}$	$\mathbf{0.499}_{\pm 0.009}$	$48.15_{\pm 1.26}$	$\mathbf{0.176}_{\pm 0.004}$	$\mathbf{92.68}_{\pm 0.20}$	$52.95_{\pm 0.497}$	$\mathbf{0.568}_{\pm 0.003}$	$\mathbf{57.92}_{\pm 1.16}$	$\mathbf{0.216}_{\pm 0.002}$

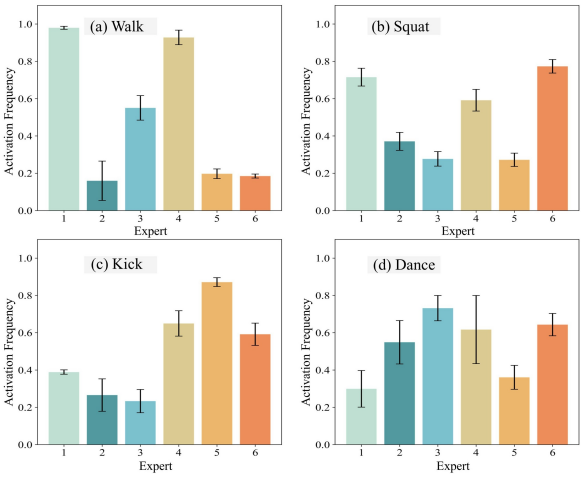

Fig. 4: **Expert activation frequencies across representative motion categories.** The OMoE architecture effectively decomposes skills and adapts expert usage to motion diversity.

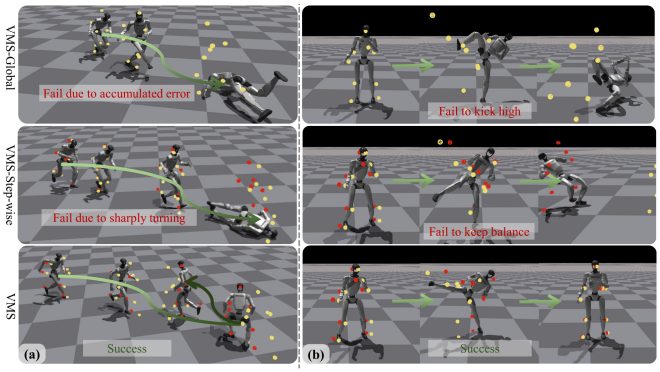

Fig. 5: **Visualizations of example motions.** (a) for a running-to-turn motion, VMS achieves a smooth transition, (b) for a side kick, it maintains balance and completes the motion stably.

results confirm that VMS adapts expert allocation to motion complexity, achieving flexible skill composition.

2) Ablations on Segment-level Reward: To answer **Q3**, we conduct an ablation study comparing VMS with two other tracking strategies: (i) VMS-Global, where all keybodies are aligned strictly to their global targets [11], [12], and (ii) VMS-Step-wise, which uses the same hybrid tracking objective as VMS but enforces strict step-wise tracking target at each timestep. As shown in Table III, while pure global tracking achieves the lowest root error on the training set, it generalizes poorly, exhibiting high errors and low success rates on the test set due to error accumulation and sensitivity to imperfect reference motions. Step-wise tracking achieves more stable performance but still underperforms our method.

To further analyze, we visualize two representative motions in Fig. 5, with yellow points denoting global target keybody positions and red points representing local tracking targets. In example (a), a fast run followed by a sharp turn, the global tracker quickly fails as accumulated errors prevent it from reaching the target, while the step-wise tracker rigidly adheres to the target path and collapses during the turn. In contrast, our segment-level reward leverages a short future window, allowing smoother transitions and preserving motion style. In example (b), a side kick motion, the global tracker overemphasizes height and falls, while the step-wise tracker completes the kick but loses balance afterward. Our method succeeds by slightly compromising on kick height to maintain overall stability. In Fig. 6, test motions are grouped by duration. While both compared strategies degrade rapidly as motion duration increases, our method remains robust, showing strong performance on long-horizon motions.

D. Real-World Experiments

For **Q4**, Fig. 1 and the supplementary videos show our robot executing diverse real-world skills, including: (1) locomotion styles such as walking, marching, and running; (2) athletic movements like racket swings and ball throws; (3) expressive dances (e.g., Charleston); (4) highly dynamic actions such as punches, side kicks, and jumping kicks; and (5) long-horizon composite skills involving martial arts, such as Tai Chi and Shaolin Kungfu. These results demonstrate VMS's versatility as a general-purpose humanoid controller.

E. Downstream Tasks

1) Text-to-motion Generation: To further assess the generalization ability of VMS, we conduct experiments on text-to-motion generation. The MLD model [53] is used to generate reference motions from language descriptions, and

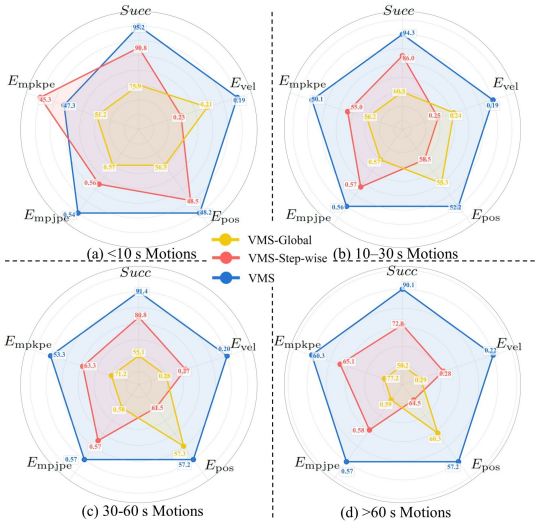

Fig. 6: **Tracking performance across motion durations.** Our segment-level tracking achieves consistently lower errors and higher robustness compared to global and step-wise tracking.

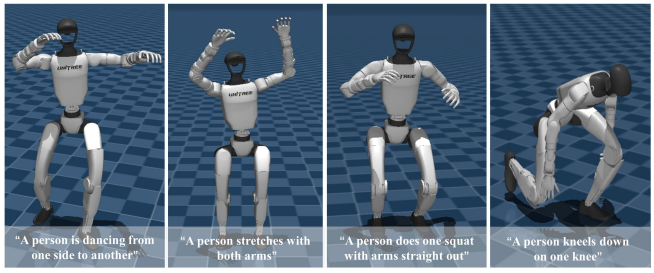

Fig. 7: **Text2motion generation results.** VMS reproduces diverse motions such as dancing, stretching, squatting, and kneeling from text descriptions, demonstrating strong zero-shot generalization.

we evaluate whether the policy can follow them. As shown in Fig. 7, VMS successfully follows diverse text-generated motion instructions, highlighting its potential to serve as a universal low-level controller for higher-level planners.

2) Finetuning on Extreme Motions: In Fig. 8, VMS also adapts well to challenging out-of-distribution (OOD) motions. Skills that demand collision ignoring or highly acrobatic movements can both be realized with minimal fine-tuning, highlighting VMS's practicality to edge cases.

V. Conclusion

In this work, we present VMS, a universal framework that enables humanoid robots to learn versatile motion skills. By combining an OMoE architecture for skill decomposition with a hybrid tracking objective and a segment-level reward, our approach achieves dynamic, human-like motions while maintaining stability over long-horizon sequences. Extensive experiments demonstrate that VMS outperforms strong baselines. Real-world deployments demonstrate that VMS can perform various dynamic motions, including minute-level sequences. The downstream tasks highlight the potential of VMS to serve as a foundation for general humanoid control.

However, VMS lacks visual perception, limiting its ability to understand complex scenes. Second, it heavily relies on large-scale Mocap datasets, which may contain uneven coverage of certain skills, affecting generalization.

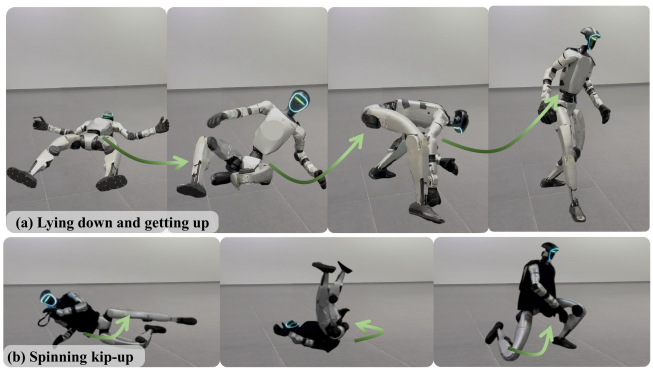

Fig. 8: **Fintuning results.** VMS adapting to OOD motions: (a) lying down and getting up, and (b) a highly dynamic spinning kip-up.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No.62306242), the Young Elite Scientists Sponsorship Program by CAST (Grant No. 2024QNRC001), and the Yangfan Project of the Shanghai (Grant No.23YF11462200). The authors acknowledge using ChatGPT solely for language polishing.

References

[1] 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.*, vol. 41, no. 4, Jul. 2022.

[2] Z. Luo, J. Cao, K. Kitani, W. Xu *et al.*, "Perpetual humanoid control for real-time simulated avatars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 895–10 904.

[3] 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.

[4] R. Yu, Y. Wang, Q. Zhao, H. W. Tsui, J. Wang, P. Tan, and Q. Chen, "Skillmimic-v2: Learning robust and generalizable interaction skills from sparse and noisy demonstrations," in *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, 2025, pp. 1–11.

[5] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. M. Kitani, C. Liu, and G. Shi, "Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning," in *8th Annual Conference on Robot Learning*, 2024.

[6] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," in *8th Annual Conference on Robot Learning*, 2024.

[7] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," in *Robotics: Science and Systems*, 2024.

[8] A. Serifi, R. Grandia, E. Knoop, M. Gross, and M. Bächer, "Vmp: Versatile motion priors for robustly tracking motion on physical characters," in *Computer graphics forum*, vol. 43, no. 8. Wiley Online Library, 2024, p. e15175.

[9] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang *et al.*, "Hover: Versatile neural whole-body controller for humanoid robots," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 9989–9996.

[10] M. Ji, X. Peng, F. Liu, J. Li, G. Yang, X. Cheng, and X. Wang, "Exbody2: Advanced expressive humanoid whole-body control," *arXiv preprint arXiv:2412.13196*, 2024.

[11] W. Xie, J. Han, J. Zheng, H. Li, X. Liu, J. Shi, W. Zhang, C. Bai, and X. Li, "Kungfubot: Physics-based humanoid whole-body control for learning highly-dynamic skills," 2025. [Online]. Available: https://arxiv.org/abs/2506.12851

[12] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan, Z. Yi, G. Qu, K. Kitani, J. Hodgins, L. J. Fan, Y. Zhu, C. Liu, and G. Shi, "Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills," 2025. [Online]. Available: https://arxiv.org/abs/2502.01143

[13] Q. Liao, T. E. Truong, X. Huang, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," 2025. [Online]. Available: https://arxiv.org/abs/2508.08241

[14] Y. Ze, Z. Chen, J. P. Araújo, Z. ang Cao, X. B. Peng, J. Wu, and C. K. Liu, "Twist: Teleoperated whole-body imitation system," *arXiv preprint arXiv:2505.02833*, 2025.

[15] K. Yin, W. Zeng, K. Fan, Z. Wang, Q. Zhang, Z. Tian, J. Wang, J. Pang, and W. Zhang, "Unitracker: Learning universal whole-body motion tracker for humanoid robots," 2025. [Online]. Available: https://arxiv.org/abs/2507.07356

[16] Y. Li, Y. Lin, J. Cui, T. Liu, W. Liang, Y. Zhu, and S. Huang, "Clone: Closed-loop whole-body humanoid teleoperation for long-horizon tasks," 2025.

[17] Z. Chen, M. Ji, X. Cheng, X. Peng, X. B. Peng, and X. Wang, "Gmt: General motion tracking for humanoid whole-body control," 2025. [Online]. Available: https://arxiv.org/abs/2506.14770

[18] H. Geyer, A. Seyfarth, and R. Blickhan, "Positive force feedback in bouncing gaits?" *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1529, pp. 2173–2183, 2003.

[19] K. Sreenath, H.-W. Park, I. Poulakakis, and J. W. Grizzle, "A compliant hybrid zero dynamics controller for stable, efficient and fast bipedal walking on mabel," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1170–1193, 2011.

[20] H. Wang, Z. Wang, J. Ren, Q. Ben, J. Pang, T. Huang, and W. Zhang, "Beamdojo: Learning agile humanoid locomotion on sparse footholds," *arXiv preprint arXiv:2502.10363*, 2024.

[21] W. Xie, C. Bai, J. Shi, J. Yang, Y. Ge, W. Zhang, and X. Li, "Humanoid whole-body locomotion on narrow terrain via dynamic balance and reinforcement learning," *arXiv preprint arXiv:2502.17219*, 2025.

[22] D. Wang, X. Wang, X. Liu, J. Shi, Y. Zhao, C. Bai, and X. Li, "More: Mixture of residual experts for humanoid lifelike gaits learning on complex terrains," 2025. [Online]. Available: https://arxiv.org/abs/2506.08840

[23] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Robust and versatile bipedal jumping control through reinforcement learning," in *Robotics science and systems*. RSS, 2023.

[24] Z. Zhuang, S. Yao, and H. Zhao, "Humanoid parkour learning," in *Conference on Robot Learning*. PMLR, 2025, pp. 1975–1991.

[25] T. Huang, J. Ren, H. Wang, Z. Wang, Q. Ben, M. Wen, X. Chen, J. Li, and J. Pang, "Learning humanoid standing-up control across diverse postures," *arXiv preprint arXiv:2502.08378*, 2025.

[26] X. He, R. Dong, Z. Chen, and S. Gupta, "Learning getting-up policies for real-world humanoid robots," *arXiv preprint arXiv:2502.12152*, 2025.

[27] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *ACM Transactions on Graphics (ToG)*, vol. 40, no. 4, 2021.

[28] Y. Zhang, Y. Yuan, P. Gurunath, T. He, S. Omidshafiei, A. akbar Agha-mohammadi, M. Vazquez-Chanlatte, L. Pedersen, and G. Shi, "Falcon: Learning force-adaptive humanoid loco-manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2505.06776

[29] Y. Li, Y. Zhang, W. Xiao, C. Pan, H. Weng, G. He, T. He, and G. 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.

[30] J. Shi, X. Liu, D. Wang, O. Lu, S. Schwertfeger, F. Sun, C. Bai, and X. Li, "Adversarial locomotion and motion imitation for humanoid policy learning," *arXiv preprint arXiv:2504.14305*, 2025.

[31] Z. Su, B. Zhang, N. Rahmanian, Y. Gao, Q. Liao, C. Regan, K. Sreenath, and S. S. Sastry, "Hitter: A humanoid table tennis robot via hierarchical planning and learning," 2025. [Online]. Available: https://arxiv.org/abs/2508.21043

[32] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions On Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[33] T. Zhang, B. Zheng, R. Nai, Y. Hu, Y.-J. Wang, G. Chen, F. Lin, J. Li, C. Hong, K. Sreenath, and Y. Gao, "Hub: Learning extreme humanoid balance," *arXiv preprint arXiv:2505.07294*, 2025.

[34] Y. Wang, M. Yang, Z. Ding, Y. Zhang, W. Zeng, X. Xu, H. Jiang, and Z. Lu, "From experts to a generalist: Toward general whole-body control for humanoid robots," 2025. [Online]. Available: https://arxiv.org/abs/2506.12779

[35] Unitree Robotics, "Humanoid robot G1_Humanoid Robot Functions_Humanoid Robot Price — Unitree Robotics," 2025, https://www.unitree.com/g1/. [Online]. Available: https://www.unitree.com/g1/

[36] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[37] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "Amass: Archive of motion capture as surface shapes," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5442–5451.

[38] Y. Ze, J. P. Araújo, J. Wu, and C. K. Liu, "Gmr: General motion retargeting," 2025, gitHub repository. [Online]. Available: https://github.com/YanjieZe/GMR

[39] K. Zakka, "Mink: Python inverse kinematics based on MuJoCo," Jul. 2024, https://github.com/kevinzakka/mink. [Online]. Available: https://github.com/kevinzakka/mink

[40] 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 *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=OrOd8PxOO2

[41] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5745–5753.

[42] S. J. Leon, Å. Björck, and W. Gander, "Gram-schmidt orthogonalization: 100 years and more," *Numerical Linear Algebra with Applications*, vol. 20, no. 3, pp. 492–532, 2013.

[43] A. Hendawy, J. Peters, and C. D'Eramo, "Multi-task reinforcement learning with mixture of orthogonal experts," 2024. [Online]. Available: https://arxiv.org/abs/2311.11385

[44] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.

[45] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: learning a unified policy for manipulation and locomotion," in *Conference on Robot Learning*. PMLR, 2023, pp. 138–149.

[46] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, vol. 9, no. 89, p. eadi9579, 2024.

[47] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *The International Journal of Robotics Research*, vol. 44, no. 5, pp. 840–888, 2025.

[48] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 627–635.

[49] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2018, p. 3803–3810. [Online]. Available: http://dx.doi.org/10.1109/ICRA.2018.8460528

[50] J. Wang, Y. Jiang, H. Zhang, C. Tessler, D. Rempe, J. Hodgins, and X. B. Peng, "Hil: Hybrid imitation learning of diverse parkour skills from videos," 2025. [Online]. Available: https://arxiv.org/abs/2505.12619

[51] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.

[52] J. Puigcerver, C. Riquelme, B. Mustafa, and N. Houlsby, "From sparse to soft mixtures of experts," 2024. [Online]. Available: https://arxiv.org/abs/2308.00951

[53] X. Chen, B. Jiang, W. Liu, Z. Huang, B. Fu, T. Chen, and G. Yu, "Executing your commands via motion diffusion in latent space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 18 000–18 010.