SMAP 🤖: SELF-SUPERVISED MOTION ADAPTATION FOR PHYSICALLY PLAUSIBLE HUMANOID WHOLE-BODY CONTROL

Anonymous authors
Paper under double-blind review

Figure 1: Our framework enables humanoid robot execute various expressive whole-body motions. The robot can (a) turn around and walk forward, (b) wave hello, (c) swing arms while advancing, (d) jump on one leg, (e) walk fast.

ABSTRACT

This paper presents a novel framework that enables real-world humanoid robots to maintain stability while performing human-like motion. Current methods train a policy which allows humanoid robots to follow human body using the massive retargeted human data via reinforcement learning. However, due to the heterogeneity between human and humanoid robot motion, directly using retargeted human motion reduces training efficiency and stability. To this end, we introduce **SMAP**, a novel whole-body tracking framework that bridges the gap between human and humanoid motion spaces, enabling accurate motion mimicry by humanoid robots. The core idea is to use a vector-quantized periodic autoencoder to capture generic atomic behaviors and adapt human motion into physically plausible humanoid motion. This adaptation accelerates training convergence and improves stability when handling novel or challenging motions. We then employ a privileged teacher to distill precise mimicry skills into the student policy with a proposed decoupled reward. We conduct experiments in simulation and real world to demonstrate the superiority stability and performance of SMAP over SOTA methods, offering practical guidelines for advancing whole-body control in humanoid robots.

1 INTRODUCTION

Humanoid robots, with their human-like morphology, have long been a focal point in robotics due to their potential to perform diverse daily tasks Zhao et al. (2024b); Li et al. (2025). Designed for

human environments, tools, and interactions, human-sized humanoids serve as ideal platforms for general-purpose robotics, naturally adapting to tasks suited for humans. However, achieving this versatility requires precise and robust whole-body control, enabling humanoid robots to coordinate high-degree-of-freedom movements for safe and effective interaction with their surroundings Zhao et al. (2025).

Traditional approaches, which decompose the problem into perception, planning, and tracking while modularizing arm and leg control separately Chestnutt et al. (2005); Feng et al. (2014); Kuindersma et al. (2016), are time-consuming to design, limited in scope. These limitations make it challenging to scale humanoids to the diverse range of tasks and environments. Human motion capture datasets Mahmood et al. (2019); Guo et al. (2022); Punnakkal et al. (2021); Carnegie Mellon University (2007) provide a rich source of reference motion, enabling the imitation of human activities Zhao et al. (2024c;a). With the growing availability of large-scale human motion datasets, recent approaches Fu et al. (2024); Cheng et al. (2024); He et al. (2024a;b); Ji et al. (2024); Lu et al. (2024); He et al. (2025) leverage Reinforcement Learning (RL) to track and mimic retargeted human motion, allowing humanoid robots to learn versatile behaviors. For example, HumanPlus Fu et al. (2024) presents a system that enables humanoids to learn and imitate human motion and skills in real time using RL. However, a major challenge lies in the significant heterogeneity of retargeted human motion data. Given that humanoid robots and humans have entirely distinct action spaces, directly using human motion data as an imitation goal often results in physically implausible motion, leading to low training efficiency and instability. This poses a compelling research question: *How to formulate imitation goals that ensure both physical plausibility and human-like motion for humanoid robots*?

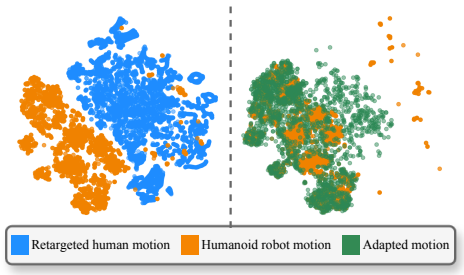

Figure 2: **t-SNE visualization** of the distribution of **retargeted human motion**, **humanoid robot motion** (recorded within the simulator), and **motion adapted by Humanoid-Adapter** on the CMU MoCap dataset Carnegie Mellon University (2007).

To address the aforementioned challenges, we propose **SMAP**, an effective framework for **S**elf-supervised **M**otion **A**daptation, enabling **P**hysically plausible whole-body control for humanoid robots. Unlike previous methods that operate within the heterogeneous retargeted human action space, we train and perform inference within the physically plausible action space for humanoid robots. Specifically, we introduce **Humanoid-Adapter**, a vector-quantized periodic autoencoder that maps human motion sequences to physically plausible humanoid robot actions. By encoding human motion sequences into a shared codebook, we decompose motion into generic atomic behaviors and then decoding them into corresponding motion, our method enables an efficient transformation between human and humanoid robot movements, as shown in Fig. 2. The adapted motion, serving as an imitation goal, significantly improves policy stability and accelerates training convergence. Additionally, through teacher-student distillation and a decoupled reward that separately optimizes upper and lower body dynamics, we further enhance both motion performance and stability. Extensive experiments on humanoid platforms demonstrate that our method achieves superior performance in full-body tracking accuracy and velocity tracking while maintaining stability in dynamic environments. In summary, our work makes the following contributions:

- We propose a novel framework for training a robust whole-body control policy that addresses the heterogeneity between human and humanoid action spaces.
- We propose a vector-quantized periodic autoencoder that adapts human motion into physically plausible motion for training and inference.
- Experiments demonstrate SMAP's superior motion imitation and stability.

2 Related Work

2.1 Humanoid Whole-Body Control

Humanoid robots have great potential to unlock the full capabilities of humanoid systems, but remains a long-standing challenge due to their high degrees of freedom (DoF) and non-linearity Grizzle et al. (2009); Hirai et al. (1998). Traditional approaches often rely on human motion capture suits Darvish et al. (2019); Dragan et al. (2013); Ben et al. (2025), haptic feedback devices Brygo et al. (2014); Peternel & Babič (2013); Ramos & Kim (2019), and dynamics modeling and control Dariush et al. (2008); Miura & Shimoyama (1984); Ramos & Kim (2019); Westervelt et al. (2003); Yin et al. (2007). Recent advances in sim-to-real reinforcement learning (RL) and sim-to-real transfer show promising results in enabling complex whole-body skills for humanoid robots such as walking Agarwal et al. (2023); He et al. (2024b); Ito et al. (2022); Cheng et al. (2023); Escontrela et al. (2022); Fu et al. (2023); Fuchioka et al. (2023); Yang et al. (2023a); Radosavovic et al. (2024a;b), jumping He et al. (2025), parkour Zhuang et al. (2024), dancing Fu et al. (2024); Cheng et al. (2024); He et al. (2024a), and hopping Ji et al. (2024). For example, H2O He et al. (2024b) presents an RL-based teleoperation framework using a third-person RGB camera to capture the human teleoperator's full-body keypoints. Exbody Cheng et al. (2024) focuses on imitating upper-body reference motion (retargeted from human data) while allowing the legs to robustly follow a given velocity. However, these methods directly use retargeted human motion, which leads to inefficient training and unstable performance due to the significant heterogeneity between human and humanoid robot motion. To address this, we propose Humanoid-Adapter that adapt human motion into the humanoid robot action space, facilitating more efficient and stable learning.

2.2 Motion Manifold Learning

Motion manifold learning has the primary goal of comprehending the fundamental structures inherent in human movement and dynamics Li et al. (2024b); Starke et al. (2022); Li et al. (2024a). Its distinctive ability to generate human movement patterns presents numerous opportunities to comprehend intrinsic motion dynamics, manage nonlinear relationships in motion data, and acquire contextual and hierarchical representations Yang et al. (2023b); Raab et al. (2023); Zhao et al. (2024d); Watanabe et al. (2025). Holden et al.Holden et al. (2016) generate character movements by mapping high-level parameters to the human motion manifold, enabling diverse motion. Recently, DeepPhaseStarke et al. (2022) propose Periodic Autoencoder to learn a low-dimensional motion manifold. With vector quantized periodic autoencoder, we learn a shared phase manifold for human and humanoid robot. The discrete amplitude vectors serve as a narrow bottleneck to regularize unsupervised learning of semantic motion clusters.

3 Preliminaries

Goal-conditioned Reinforcement Learning. We formulate the whole-body control task as a goal-conditioned Markov Decision Process (MDP). Our policy, $\pi(a_t|s_t, s_t^g)$, is trained with Proximal Policy Optimization (PPO) (Schulman et al., 2017) to imitate a reference motion. The state s_t includes the robot's proprioception (e.g., joint positions and velocities), while the goal s_t^g is a sequence of future target poses adapted by our Humanoid-Adapter. The policy aims to maximize the expected discounted return $\mathbb{E}[\sum_t \gamma^t r_t]$, where the reward $r_t = \mathcal{R}(s_t, s_t^g)$ measures the mimicry performance. The action $a_t \in \mathbb{R}^n$ represents the target joint positions for a PD controller, where $n = 23$ is the number of actuated DoFs for our Unitree H1 robot.

4 Method

We introduce **SMAP**, an effective sim-to-real framework for robust whole-body humanoid control, as illustrated in Fig. 3. To deal with the heterogeneity between human and humanoid robot motion, we propose **Humanoid-Adapter** to adapt human motion action space into physically plausible humanoid robot action space. We then propose **Progressive Control Policy Learning**, which leverages teacher-student distillation to incorporate privileged inputs and employs a decoupled reward to enhance both the performance and stability of the humanoid robot's motion.

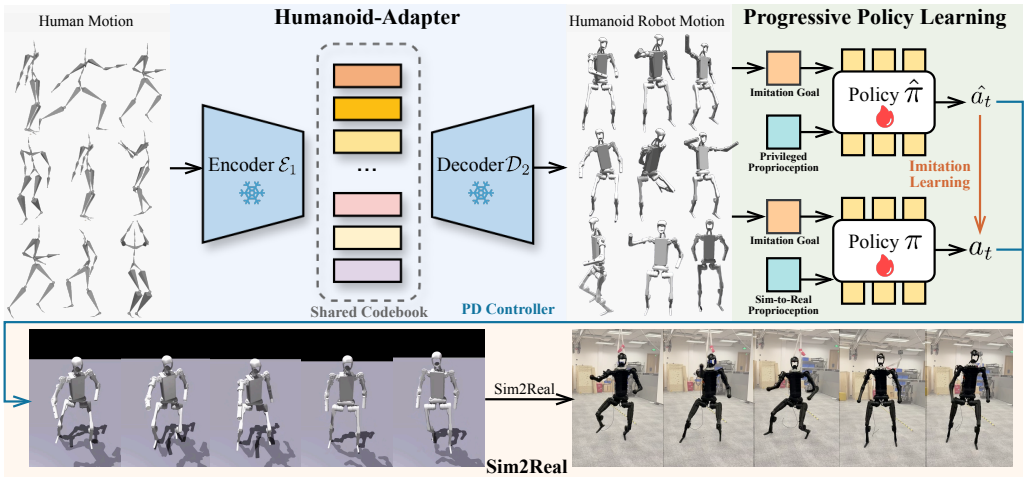

Figure 3: **Pipeline of SMAP**. Given human motion, we use the proposed **Humanoid-Adapter** (details shown in Fig. 4), pre-trained (❄) to adapt human motion into corresponding, physically plausible humanoid robot motion. Our sim-to-real policy (🔥) is distilled via imitation learning from an RL-trained privileged teacher policy that leverages privileged information with proposed decoupled reward. The policy is transferred to the real world.

4.1 HUMANOID-ADAPTER

To transform human motion action space into humanoid robot motion action space, we pre-train Humanoid-Adapter, built upon the Periodic Autoencoder (PAE) Starke et al. (2022), to learn a continuous phase manifold, and semantically align motion, as shown in Fig. 4.

Inspired by Li et al. (2024b); Starke et al. (2022), we aim to learn a generative phase representation for both human and humanoid robot motion, enabling indefinite motion synthesis while preserving temporal coherence and dynamic consistency. To build a humanoid robot motion dataset, $\mathcal{S}^r$, we train a goal-conditioned RL-based policy Cheng et al. (2024) and record the humanoid robot's motion data within the simulator. Using both the humanoid robot motion dataset $\mathcal{S}^r$ and the human motion dataset Carnegie Mellon University (2007) $\mathcal{S}^h$, we learn a shared phase manifold for human and humanoid robot characters without any supervision.

Phase Manifold. The Humanoid-Adapter models periodic motions as closed curves within a latent phase manifold. For an entire input motion sequence, our encoder predicts a single constant amplitude vector α, an initial phase ϕ_0, and a constant frequency f. This design offers a principled method to disentangle a motion's content (the "what" and "how") from its temporal progression (the "when"). The amplitude α represents the physical style of the motion, which is often kinematically and dynamically different between humans and humanoids. This disentanglement is a principled choice: we adapt the physical form via a discrete amplitude vocabulary while preserving the crucial temporal continuity of phase and frequency. Conversely, the phase ϕ and frequency f represent the rhythm and timing, which we aim to preserve to maintain the original motion's intent.

The latent representation $p(t)$ for each frame at time step t is generated by a point on this curve, determined by its time-dependent phase $\phi(t)$:

$$p(t) = \Psi(\alpha, \phi(t)) = \alpha_0 \sin(2\pi\phi(t)) + \alpha_1 \cos(2\pi\phi(t)), \tag{1}$$

where α is the amplitude vector, composed of two orthogonal components α_0 and α_1 that define the axes of the elliptical manifold. The phase $\phi(t)$ evolves linearly over time, governed by the initial phase ϕ_0 and frequency f:

$$\phi(t) = \phi_0 + f \cdot t. \tag{2}$$

This formulation allows us to model a continuous motion sequence by traversing a simple, structured latent curve, where f controls the speed of the motion and ϕ_0 aligns the sequence's starting point on

the curve. To learn a discrete amplitude space, we employ vector quantization to cluster the vector amplitude α into a learnable codebook $\mathcal{C}$, represented as:

$$\mathcal{C} = (c_1, c_2, c_3, \ldots, c_n), \tag{3}$$

where n is the space size and each c_i is a vector embedding that represents an atomic behavior. By quantizing only the amplitude, our VQ codebook learns a shared, abstract vocabulary of motion primitives that bridges the domain gap between human and humanoid motions. Keeping the phase continuous is crucial, as it allows us to adapt the physical form of the motion (via the quantized amplitude) while retaining the original, smooth temporal flow. Motion with similar physical characteristics are thus mapped to the same codebook entry.

Structure of Humanoid-Adapter. Humanoid-Adapter learns a shared, discrete latent space for both human and humanoid robot motion without requiring paired data. As illustrated in Fig. 4, this is achieved by training two separate Vector-Quantized Periodic Autoencoders (VQ-PAEs)—one for human motion ($\mathcal{S}^h$) and one for humanoid motion ($\mathcal{S}^r$)—that share a single, unified codebook $\mathcal{C}$.

For any given motion sequence, the corresponding encoder predicts its amplitude α, initial phase ϕ_0, and frequency f. The amplitude is then quantized by selecting the nearest embedding from the shared codebook $\mathcal{C}$. Using the quantized amplitude and the predicted temporal parameters (ϕ_0, f), a continuous latent trajectory is generated according to Eq. 1 and Eq. 2. A corresponding decoder then reconstructs the motion sequence from this latent trajectory. This shared discrete structure of the amplitude space ensures that semantically similar motions from both domains are mapped to the same latent curve, thus creating a common ground for motion adaptation.

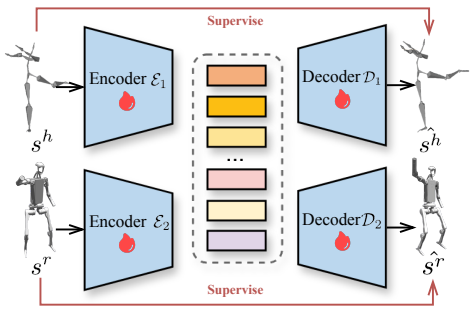

Figure 4: **Humanoid-Adapter**. To align heterogeneous human motion $\mathcal{S}^h$ and humanoid robot motion $\mathcal{S}^r$, we train two VQ-PAEs (🔥) that share a single codebook to learn a common phase manifold.

Training Objective. The network is trained by minimizing a VQ-VAE-style loss, which encourages accurate reconstruction while regularizing the latent space. The total loss is:

$$\mathcal{L} = \mathcal{L}_{\text{VQ-PAE}}(s^h, \hat{s}^h) + \mathcal{L}_{\text{VQ-PAE}}(s^r, \hat{s}^r), \tag{4}$$

where $L_{\text{VQ-PAE}}$ is the complete loss for a single domain, including reconstruction, codebook, and commitment terms. Specifically, for a sequence s and its reconstruction $\hat{s}$:

$$\mathcal{L}_{\text{VQ-PAE}}(s, \hat{s}) = \underbrace{\|s - \hat{s}\|^2}_{\text{Reconstruction}} + \underbrace{\|\text{sg}(z_e(s)) - c_k\|^2}_{\text{Codebook}} + \underbrace{\|z_e(s) - \text{sg}(c_k)\|^2}_{\text{Commitment}}. \tag{5}$$

Here, $z_e(s)$ is the encoder output, c_k is the chosen codebook vector, sg is the stop-gradient operator. A common issue in VQ-based models is codebook underutilization. To mitigate this, we adopt the reinitialization strategy from CVQ-VAE (Zheng & Vedaldi, 2023), which dynamically replaces inactive codes during training to ensure the codebook remains fully utilized for both domains.

Inference for Motion Adaptation. During inference, to adapt a human motion s^h into a physically plausible humanoid motion, we use the encoder trained on human data ($\mathcal{E}_h$) and the decoder trained on humanoid data ($\mathcal{D}_r$). This cross-domain reconstruction effectively translates the motion style while preserving the original timing, bridging the gap between human motion and executable humanoid actions (as visualized in Fig. 2).

4.2 Progressive Control Policy Learning

During real-world whole-body control of a humanoid robot, much of the information (*e.g.* global linear/angular velocity, positions of each link, and physical properties) which is available in simulation,

Method	Trained Motion Sample				Novel Motion Sample			
	$E_{vel} \downarrow$	$E_{mpkpe} \downarrow$	$E_{mpjpe} \downarrow$	$fail \downarrow$	$E_{vel} \downarrow$	$E_{mpkpe} \downarrow$	$E_{mpjpe} \downarrow$	$fail \downarrow$
Privileged policy	0.1002	0.0531	0.0888	798	0.1923	0.0732	0.1121	199
HumanPlus Fu et al. (2024)	0.3103	0.1011	0.1989	3009	0.4299	0.1598	0.2612	633
HumanPlus + Humanoid-Adapter	0.2991	0.0982	0.1862	2683	0.3816	0.1374	0.2325	500
H2O He et al. (2024b)	0.2333	0.0831	0.1989	2762	0.3613	0.1385	0.2212	501
H2O + Humanoid-Adapter	0.2291	0.0808	0.1822	2499	0.3566	0.1301	0.2160	421
OmniH2O He et al. (2024a)	0.1791	0.0619	0.1250	1899	0.2591	0.0912	0.1481	387
Exbody Cheng et al. (2024)	0.2160	0.0766	0.1783	2264	0.3002	0.1070	0.1868	432
Exbody† Cheng et al. (2024)	0.2285	0.0770	0.1592	2322	0.3239	0.1099	0.1749	489
Exbody + AMP Peng et al. (2021)	0.2499	0.0732	0.1531	1993	0.3487	0.1002	0.1604	412
Exbody + Humanoid-Adapter	0.2109	0.0752	0.1682	1982	0.2998	0.0999	0.1632	361
SMAP	**0.1698**	**0.0608**	**0.1181**	**1731**	**0.2331**	**0.0893**	**0.1458**	**266**
SMAP w/o Humanoid-Adapter	0.1743	0.0612	0.1221	1851	0.2465	0.0921	0.1442	392
SMAP w/o teacher-student distillation	0.2038	0.0732	0.1521	1751	0.2765	0.0941	0.1641	389
SMAP w/o progressive	0.1712	0.0610	0.1191	1889	0.2387	0.0914	0.1468	299
SMAP w/o decoupled reward	0.1739	0.0659	0.1283	1775	0.2389	0.0903	0.1499	291

Table 1: **Quantitative Comparisons and Ablation Study**. Simulation-based motion imitation evaluation of our method and state-of-the-art (SOTA) approaches on the CMU MoCap dataset Carnegie Mellon University (2007) for the Unitree H1 robot.

is not accessible. To address this, our control policy employs a progressive two-stage teacher-student training. In the first stage, the teacher policy is trained using privileged information that can only be obtained in simulation. In the second stage, we replace this privileged information with real-world observations, and distill the teacher policy into a student policy.

Curriculum-based Teacher Policy Training. The teacher policy $\hat{\pi}$ takes privileged proprioception and imitation goals as inputs and outputs the action $\hat{a}_t$. The privileged information $s_t^{privileged}$ includes ground-truth states of the humanoid robot and environment (e.g., root velocity, body link positions, and physical properties). For details on privileged information, please refer to the Appendix 7.5.

The training process follows a progressive curriculum strategy to effectively balance stability and expressiveness. While the physically plausible motion from Humanoid-Adapter provides stable and easy-to-learn initial targets, they can be overly conservative, filtering out expressive nuances from the original human data. Therefore, we first train the policy on the adapted data to build a robust foundation for balance. Subsequently, we gradually introduce the original retargeted human motion, allowing the now-stable policy to learn more dynamic and expressive details without failing. This curriculum learning approach not only ensures convergence stability but also enables the policy to ultimately achieve higher faithfulness to the complex human motions.

Student Policy Distilling. In this stage, we remove privileged proprioception and leverage sim-to-real proprioception (a longer history of observations) to train the student policy. The policy is supervised using the teacher policy's action $\hat{a}_t$ with loss followed:

$$\mathcal{L}_{\text{distill}} = \|a_t - \hat{a}_t\|_2. \tag{6}$$

To train the student policy, we follow the DAgger Ross et al. (2011). At each visited state, the teacher policy $\hat{\pi}$ generates an oracle action as the supervision signal. The student policy π is iteratively refined by minimizing the loss $\mathcal{L}_{\text{distill}}$.

Decoupled Reward Design Our reward function is designed to enhance both motion tracking and stability. It consists of two main components: tracking rewards and regularization terms. The tracking rewards encourage the policy to mimic the reference motion by minimizing errors in root velocity, direction, orientation, as well as keypoint and joint positions. To strike a crucial balance between precision and stability, we employ a decoupled reward strategy: the upper body receives a higher weight to ensure accurate task execution, while the lower body is weighted less to prioritize balance and stability. Several regularization terms are included to prevent unnatural motions and enhance generalization. More details about reward components are provided in the Appendix 7.4.

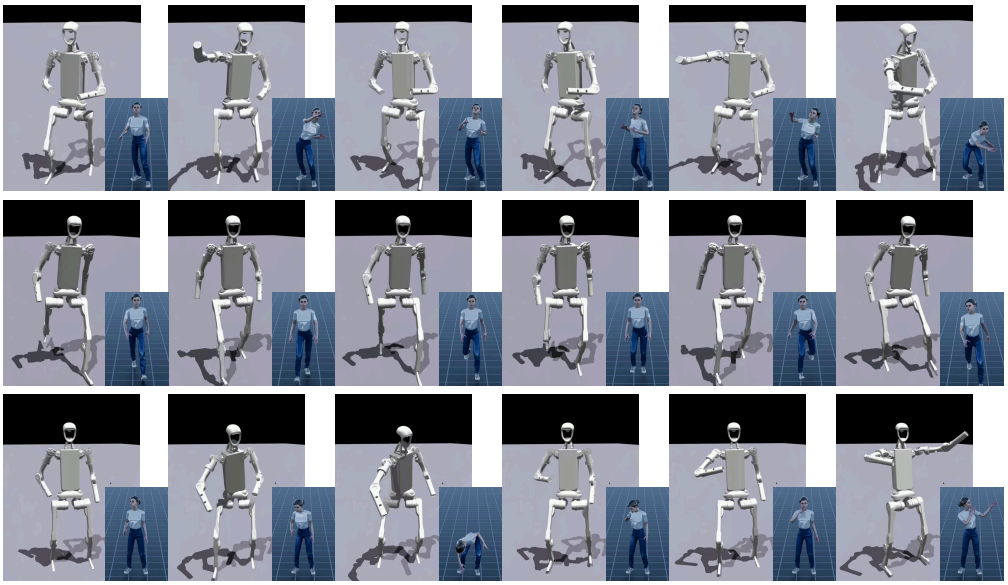

Figure 5: **Qualitative results** on the H1 robot in simulation.

5 EXPERIMENTS

Implementation Details. We conduct experiments in IsaacGym Makoviychuk et al. (2021) simulator. During training, 4096 environments are simulated in parallel on a NVIDIA RTX 4090 GPU.

Dataset. Following Exbody Cheng et al. (2024), we use a portion of the CMU MoCap Carnegie Mellon University (2007), excluding motion involving physical interactions with other individuals, heavy objects, or rough terrain. This diversity not only enhances the expressiveness of humanoid motion but also improves locomotion stability in unseen scenarios. To further evaluate robustness and generalization, we also test on **novel motion** from the CMU dataset that are not used during training.

Evaluation Metrics. We evaluate the policy's performance using several metrics. **The mean linear velocity error** (E_{vel}) measures the discrepancy between the robot's root linear velocity and the demonstration, reflecting velocity tracking accuracy. **The mean per key point position error** (E_{mpkpe}) and **the mean per joint position error** (E_{mpjpe}) evaluate motion accuracy, with E_{mpkpe} assessing keypoint tracking and E_{mpjpe} capturing joint tracking accuracy. **Failure** ($fail$) counts the number of failure terminations across all motion sequences. Lower failure rates suggest greater robustness and control consistency across diverse motions.

5.1 COMPARISON WITH SOTA METHODS

We evaluate our method on motion tracking in simulation across Unitree H1, comparing it with some SOTA methods: HumanPlus Fu et al. (2024), H2O He et al. (2024b), OmniH2O He et al. (2024a), and Exbody Cheng et al. (2024). This teacher policy (privileged policy) leverages all privileged environment information as mentioned in its observations. For a fair comparison, we reimplement the H2O He et al. (2024b) and OmniH2O He et al. (2024a) using global keypoint tracking and the same observation space. Since Exbody Cheng et al. (2024) only tracks upper-body motion, we introduce a whole-body tracking variant (Exbody†) to enable direct comparison. This version tracks full-body movements based on human motion data. Exbody + AMP Peng et al. (2021) uses an AMP reward to encourage the policy's transitions to be similar to those in the retargeted dataset.

We evaluate our method on the retargeted CMU MoCap dataset Carnegie Mellon University (2007) in simulation, following Exbody Cheng et al. (2024), as shown in Tab. 1. Our method achieves high full-body tracking accuracy and excels in velocity tracking. This improvement stems from the teacher-student distillation, where distilling privileged information into historical observations enables the student policy to track velocity more effectively. Additionally, our method has the fewest failure

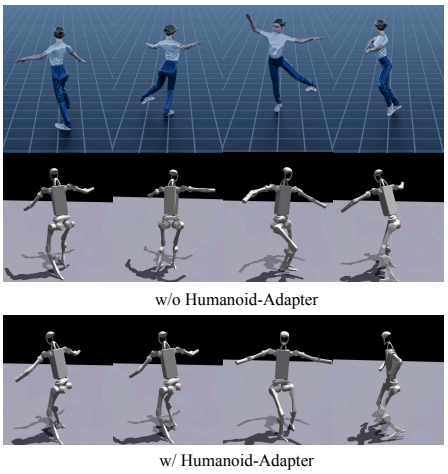

Figure 6: **Ablation study**. Visualization performance in the simulation on challenging motion sample.

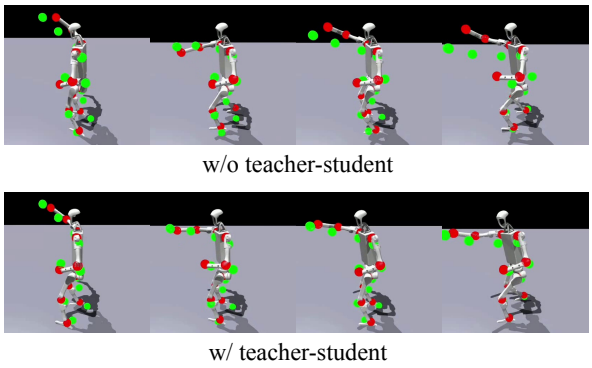

Figure 7: **Ablation study** of teacher-student distillation. The **green** points represent the imitation goal, while the **red** points correspond to the DOF position.

Figure 8: **Qualitative result** on the H1 robot in real world.

cases, demonstrating its robustness. This is largely attributed to the proposed Humanoid-Adapter, which transforms human motion into physically plausible motion, making it easier for the humanoid robot to track while maintaining stability. Notably, *the motion adapted by the proposed Humanoid-Adapter also enhances the stability of other RL-based methods*, further demonstrating its strong generalizability. To further illustrate the training efficiency and stability benefits that lead to these results, Figure. 9 compares the training reward curves of SMAP and $Exbody^{\dagger}$ *Chenget al.* (2024). Fig.5 presents qualitative results in simulation, where the humanoid robot successfully replicates various human motions, such as fast walking and squatting, while maintaining stability. Additionally, Fig.8 showcases real-world experimental results, further demonstrating the robustness of our policy in real-world settings.

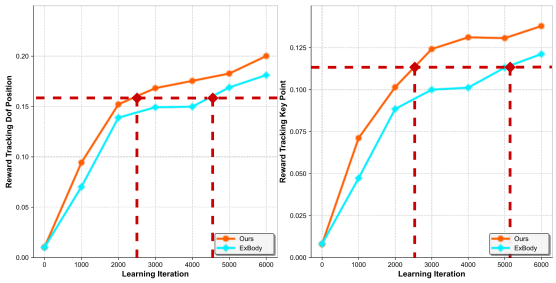

Figure 9: **Training Curves Comparison** between SMAP and Exbody†

	$E_{vel} \downarrow$	$E_{mpkpe} \downarrow$	$E_{mpjpe} \downarrow$	$fail \downarrow$
Codebook Size $\mathcal{C}$				
16	0.1732	0.0611	0.1192	1798
32	**0.1698**	**0.0608**	**0.1181**	**1731**
64	**0.1698**	0.0610	0.1190	1799
History Length				
0	0.2831	0.0751	0.1591	1973
5	0.1751	0.0623	0.1289	1787
10	**0.1698**	**0.0608**	**0.1181**	1731
32	0.1781	**0.0608**	0.1188	**1719**

Table 2: **Ablation Study Results** (Best values in bold)

5.2 Ablation Study

Humanoid-Adapter. The critical role of Humanoid-Adapter is demonstrated by comparing SMAP with a variant that uses directly retargeted motion (SMAP w/o Humanoid-Adapter). Unlike our adapter, standard Inverse Kinematics (IK) based retargeting is purely geometric and often produces dynamically infeasible references, leading to frequent policy failures.

Our Humanoid-Adapter resolves this by transforming human motion into a physically plausible, "robot-friendly" form. This directly enhances stability, as visually demonstrated in Fig. 6 and quantified by the failure count on novel motions, which drops from 392 to 266 (Table 1). The adapter's generalizability is confirmed by its plug-and-play capability. When applied to other methods like HumanPlus Fu et al. (2024) and Exbody Cheng et al. (2024), it consistently reduces their failure rates (e.g., from 432 to 361 for Exbody).

Progressive Control Policy Learning. We first analyze the impact of teacher-student distillation. Without this training (SMAP w/o teacher-student distillation), tracking accuracy decreases significantly. This is primarily due to the lack of privileged velocity guidance, which makes it challenging for the single-stage RL policy to learn velocity directly from historical data. As shown in Fig. 7, the policy without teacher-student distillation demonstrates lower tracking precision. We then examine the effectiveness of the proposed progressive learning. We find that directly using the final weight for policy training (SMAP w/o progressive) yields worse results. This is because gradually allowing the model to learn retargeted motion is crucial for improving learning performance.

Decoupled Reward. We observe that the control policy employing the decoupled reward demonstrates significantly higher precision in tracking motion targets and fewer failures. This improvement is due to the upper body's greater need for precision in task execution, while the lower body is prioritized for overall balance rather than strict positional accuracy.

Hyperparameters of SMAP. Choosing an appropriate codebook size is critical for our framework, as a small codebook size fails to capture the diverse semantics, and a large codebook reduces semantic alignment accuracy. As shown in Tab. 2, a codebook size of 32 yields the best performance.

We also test student policies trained with varying history lengths in Tab. 2. Without additional history, the policy struggles to learn effectively. History length of 10 produces the best results, which we use in our experiments. Longer history lengths increase the difficulty of fitting privileged information, ultimately reducing tracking performance.

6 Conclusion

This paper introduces **SMAP**, a novel sim-to-real framework for whole-body humanoid control. Different from previous methods, we use a network to map human motion into the humanoid robot's action space for training and inference. For superior disentanglement, we propose a vector-quantized periodic autoencoder to bridge the gap between human motion and humanoid robots. Then, we propose Progressive Control Policy Learning, leveraging teacher-student distillation and employing a

decoupled reward that separately optimizes upper and lower body dynamics. Extensive experiments in simulation and real world demonstrate that our SMAP achieves superior full-body tracking accuracy while maintaining stability.

Limitation and future work. One limitation of SMAP is the lack of explicit joint correspondence, which may cause minor mismatches in motion alignment. By sampling motion in action space of humanoid robot, Humanoid-Adapter can be a data augmentation tool to generate reliable motion, offering a promising avenue for future research.

ETHICS STATEMENT

To the best of our knowledge, our work does not present any direct ethical concerns. The research is based on publicly available datasets and does not involve sensitive personal information, human subjects, or applications with immediate potential for harm. We have strived to ensure our methods are presented transparently and that any potential societal impacts are considered within the scope of academic research.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our work, we have provided comprehensive details throughout the paper and appendix, and we commit to releasing our code upon publication.

Code: We commit to making our full source code publicly available upon acceptance of the paper. The released code will include detailed instructions for setting up the simulation environment, training the `Humanoid-Adapter` and control policies, and running evaluation scripts to reproduce the quantitative results reported in Table 1.

Datasets: Our experiments are conducted on a subset of the publicly available CMU MoCap dataset. The process for generating the physically plausible humanoid robot motion dataset (S_r) for training the `Humanoid-Adapter` is detailed in Section 4.1. Details on the motion data used are provided in Section 5.

Models and Algorithms: The overall SMAP framework is described in Section 4. The specific architecture and training objective for our proposed `Humanoid-Adapter` are presented in Section 4.1, with further implementation details available in Appendix 7.2. The Progressive Control Policy Learning strategy, which leverages teacher-student distillation, is detailed in Section 4.2.

Experimental Setup and Hyperparameters: Implementation details, including the use of the IsaacGym simulator and hardware specifications, are provided in Section 5. A complete and detailed breakdown of our decoupled reward function, including all tracking (Table 4) and regularization (Table 3) components with their respective weights, is provided in Appendix 7.4. The precise observation spaces for both the teacher and student policies are listed in Appendix 7.5. Key hyperparameters, such as the VQ codebook size and history length, are analyzed in the ablation studies in Section 5.2 and Table 2.

REFERENCES

Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Legged locomotion in challenging terrains using egocentric vision. In *Proc. of Conf. on Robot Learning*, pp. 403–415, 2023. 3

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

Anais Brygo, Ioannis Sarakoglou, Nadia Garcia-Hernandez, and Nikolaos Tsagarakis. Humanoid robot teleoperation with vibrotactile based balancing feedback. In *Haptics: Neuroscience, Devices,*

Modeling, and Applications: 9th International Conference, EuroHaptics 2014, Versailles, France, June 24-26, 2014, Proceedings, Part II 9, pp. 266–275. Springer, 2014. 3

Carnegie Mellon University. Carnegie-Mellon mocap database. http://mocap.cs.cmu.edu/, Mar 2007. [Online]. 2, 4, 6, 7

Xuxin Cheng, Ashish Kumar, and Deepak Pathak. Legs as manipulator: Pushing quadrupedal agility beyond locomotion. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 5106–5112. IEEE, 2023. 3

Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. In *Proc. of Robotics: Science and Systems*, 2024. 2, 3, 4, 6, 7, 8, 9

Joel Chestnutt, Manfred Lau, German Cheung, James Kuffner, Jessica Hodgins, and Takeo Kanade. Footstep planning for the honda asimo humanoid. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 629–634, 2005. 2

Behzad Dariush, Michael Gienger, Bing Jian, Christian Goerick, and Kikuo Fujimura. Whole body humanoid control from human motion descriptors. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 2677–2684. IEEE, 2008. 3

Kourosh Darvish, Yeshasvi Tirupachuri, Giulio Romualdi, Lorenzo Rapetti, Diego Ferigo, Francisco Javier Andrade Chavez, and Daniele Pucci. Whole-body geometric retargeting for humanoid robots. In *IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pp. 679–686. IEEE, 2019. 3

Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 301–308. IEEE, 2013. 3

Alejandro Escontrela, Xue Bin Peng, Wenhao Yu, Tingnan Zhang, Atil Iscen, Ken Goldberg, and Pieter Abbeel. Adversarial motion priors make good substitutes for complex reward functions. In *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 25–32. IEEE, 2022. 3

Siyuan Feng, Eric Whitman, X Xinjilefu, and Christopher G Atkeson. Optimization based full body control for the atlas robot. In *IEEE-RAS International Conference on Humanoid Robots*, pp. 120–127, 2014. 2

Zipeng Fu, Xuxin Cheng, and Deepak Pathak. Deep whole-body control: learning a unified policy for manipulation and locomotion. In *Proc. of Conf. on Robot Learning*, pp. 138–149. PMLR, 2023. 3

Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. HumanPlus: Humanoid shadowing and imitation from humans. In *Proc. of Conf. on Robot Learning*, 2024. 2, 3, 6, 7, 9

Yuni Fuchioka, Zhaoming Xie, and Michiel Van de Panne. Opt-mimic: Imitation of optimized trajectories for dynamic quadruped behaviors. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 5092–5098. IEEE, 2023. 3

Jessy W Grizzle, Jonathan Hurst, Benjamin Morris, Hae-Won Park, and Koushil Sreenath. Mabel, a new robotic bipedal walker and runner. In *American Control Conference*, pp. 2030–2036. IEEE, 2009. 3

Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 5152–5161, 2022. 2

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. In *Proc. of Conf. on Robot Learning*, 2024a. 2, 3, 6, 7

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 *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2024b. 2, 3, 6, 7

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. 2, 3

Kazuo Hirai, Masato Hirose, Yuji Haikawa, and Toru Takenaka. The development of honda humanoid robot. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, volume 2, pp. 1321–1326. IEEE, 1998. 3

Daniel Holden, Jun Saito, and Taku Komura. A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016. 3

Hiroshi Ito, Kenjiro Yamamoto, Hiroki Mori, and Tetsuya Ogata. Efficient multitask learning with an embodied predictive model for door opening and entry with whole-body control. *Science Robotics*, 7(65):eaax8177, 2022. 3

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. 2, 3

Scott Kuindersma, Robin Deits, Maurice Fallon, Andrés Valenzuela, Hongkai Dai, Frank Permenter, Twan Koolen, Pat Marion, and Russ Tedrake. Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot. *Autonomous robots*, 40:429–455, 2016. 2

Chenhao Li, Elijah Stanger-Jones, Steve Heim, and Sangbae Kim. Fld: Fourier latent dynamics for structured motion representation and learning. *arXiv preprint arXiv:2402.13820*, 2024a. 3, 16

Hangyu Li, Qin Zhao, Haoran Xu, Xinyu Jiang, Qingwei Ben, Feiyu Jia, Haoyu Zhao, Liang Xu, Jia Zeng, Hanqing Wang, et al. Teleopbench: A simulator-centric benchmark for dual-arm dexterous teleoperation. *arXiv preprint arXiv:2505.12748*, 2025. 1

Peizhuo Li, Sebastian Starke, Yuting Ye, and Olga Sorkine-Hornung. Walkthedog: Cross-morphology motion alignment via phase manifolds. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–10, 2024b. 3, 4

Chenhao Lu, Xuxin Cheng, Jialong Li, Shiqi Yang, Mazeyu Ji, Chengjing Yuan, Ge Yang, Sha Yi, and Xiaolong Wang. Mobile-television: Predictive motion priors for humanoid whole-body control. *arXiv preprint arXiv:2412.07773*, 2024. 2

Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 5442–5451, 2019. 2

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021. 7

Hirofumi Miura and Isao Shimoyama. Dynamic walk of a biped. *The International Journal of Robotics Research*, 3(2):60–74, 1984. 3

Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (TOG)*, 40(4):1–20, 2021. 6, 7

Luka Peternel and Jan Babič. Learning of compliant human–robot interaction using full-body haptic interface. *Advanced Robotics*, 27(13):1003–1012, 2013. 3

Abhinanda R Punnakkal, Arjun Chandrasekaran, Nikos Athanasiou, Alejandra Quiros-Ramirez, and Michael J Black. Babel: Bodies, action and behavior with english labels. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 722–731, 2021. 2

Sigal Raab, Inbal Leibovitch, Peizhuo Li, Kfir Aberman, Olga Sorkine-Hornung, and Daniel Cohen-Or. Modi: Unconditional motion synthesis from diverse data. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 13873–13883, 2023. 3

Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024a. 3

Ilija Radosavovic, Bike Zhang, Baifeng Shi, Jathushan Rajasegaran, Sarthak Kamat, Trevor Darrell, Koushil Sreenath, and Jitendra Malik. Humanoid locomotion as next token prediction. In *Proc. of Advances in Neural Information Processing Systems*, 2024b. 3

Joao Ramos and Sangbae Kim. Dynamic locomotion synchronization of bipedal robot and human operator via bilateral feedback teleoperation. *Science Robotics*, 4(35):eaav4282, 2019. 3

Stéphane Ross, Geoffrey Gordon, and Drew 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*, 2011. 6

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 3

Sebastian Starke, Ian Mason, and Taku Komura. Deepphase: Periodic autoencoders for learning motion phase manifolds. *ACM Transactions on Graphics (TOG)*, 41(4):1–13, 2022. 3, 4

Ryo Watanabe, Chenhao Li, and Marco Hutter. Dfm: Deep fourier mimic for expressive dance motion learning. *arXiv preprint arXiv:2502.10980*, 2025. 3, 16

Eric R Westervelt, Jessy W Grizzle, and Daniel E Koditschek. Hybrid zero dynamics of planar biped walkers. *IEEE transactions on automatic control*, 48(1):42–56, 2003. 3

Ruihan Yang, Ge Yang, and Xiaolong Wang. Neural volumetric memory for visual locomotion control. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1430–1440, 2023a. 3

Sicheng Yang, Zhiyong Wu, Minglei Li, Zhensong Zhang, Lei Hao, Weihong Bao, and Haolin Zhuang. Qpgesture: Quantization-based and phase-guided motion matching for natural speech-driven gesture generation. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2321–2330, 2023b. 3

KangKang Yin, Kevin Loken, and Michiel Van de Panne. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics (TOG)*, 26(3):105–es, 2007. 3

Haoyu Zhao, Hao Wang, Chen Yang, and Wei Shen. Chase: 3d-consistent human avatars with sparse inputs via gaussian splatting and contrastive learning. *arXiv preprint arXiv:2408.09663*, 2024a. 2

Haoyu Zhao, Hao Wang, Xingyue Zhao, Hongqiu Wang, Zhiyu Wu, Chengjiang Long, and Hua Zou. Automated 3d physical simulation of open-world scene with gaussian splatting. *arXiv preprint arXiv:2411.12789*, 2024b. 1

Haoyu Zhao, Chen Yang, Hao Wang, Xingyue Zhao, and Wei Shen. Sg-gs: Photo-realistic animatable human avatars with semantically-guided gaussian splatting. *arXiv preprint arXiv:2408.09665*, 2024c. 2

Haoyu Zhao, Linghao Zhuang, Xingyue Zhao, Cheng Zeng, Haoran Xu, Yuming Jiang, Jun Cen, Kexiang Wang, Jiayan Guo, Siteng Huang, et al. Towards affordance-aware robotic dexterous grasping with human-like priors. *arXiv preprint arXiv:2508.08896*, 2025. 2

Wenshuai Zhao, Yi Zhao, Joni Pajarinen, and Michael Muehlebach. Bi-level motion imitation for humanoid robots. *arXiv preprint arXiv:2410.01968*, 2024d. 3

Chuanxia Zheng and Andrea Vedaldi. Online clustered codebook. In *Proc. of IEEE Intl. Conf. on Computer Vision*, pp. 22798–22807, 2023. 5

Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. Humanoid parkour learning. *arXiv preprint arXiv:2406.10759*, 2024. 3

7 APPENDIX

STATEMENT ON THE USE OF LARGE LANGUAGE MODELS

During the preparation of this manuscript, we utilized a Large Language Model (LLM), specifically OpenAI's GPT-4, as a general-purpose writing assistant. The primary role of the LLM was to assist with improving the clarity, conciseness, and grammatical correctness of the text. Specific tasks included proofreading for errors, rephrasing sentences for better flow, and ensuring consistent terminology throughout the paper. The LLM was not used for core research ideation, developing the proposed methods, conducting experiments, or generating results. All scientific contributions, claims, and the final version of the text were authored by the human authors, who take full responsibility for the entire content of this paper.

7.1 REAL ROBOT SYSTEM SETUP

Our real robot is built on the Unitree H1 platform, as shown in Fig. 10, equipped with Damiao DM-J4310-2EC motors. The control policy receives motion-tracking target information as input, computes the desired joint positions for each motor, and sends commands to the robot's low-level interface. The policy's inference frequency is set at 50 Hz. The commands are sent with a delay kept between 18 and 30 milliseconds. The low-level interface operates at a frequency of 500 Hz, ensuring smooth real-time control. The communication between the control policy and the low-level interface is realized through LCM (Lightweight Communications and Marshalling).

Total weight: 47kg

Arm total length: 338 mm ×2 [4]

Thigh & shank length: 400 mm ×2 (each leg)

Battery: 15 Ah (0.864 kWh), Max voltage: 67.2 V

Key dimensions: (1520+285) mm × 570 mm × 220 mm

Joint torque limits (Knee): 360N·m

Joint torque limits (Hip): 220N·m

Joint torque limits (Ankle): 45N·m

Joint torque limits (Arm): 75N·m

Figure 10: **Details** about Unitree H1 robot.

7.2 MORE DETAILS OF HUMANOID ADAPTER

The efficiency of mapping semantically similar motions to the same discrete code is a direct consequence of the global optimization pressure imposed by a finite, shared codebook. To minimize the joint reconstruction loss across both domains, the network must use its limited codebook capacity parsimoniously. It is therefore compelled to discover an abstract representation—like "periodic forward locomotion" for walking—that maps to a single code. Humans and humanoids share a similar high-level body plan—both are bipeds with a torso, head, and two legs. Consequently, even though their low-level joint movements differ, their actions exhibit similar high-level patterns. This allows the encoder to latch onto these fundamental similarities, effectively learning the shared concept while disregarding the minor, morphology-specific differences

We assume the two properties hold for the whole input sequence $\mathbf{X}$ and extrapolate the phase linearly with the predicted frequency to the whole sequence. We calculate the embeddings using:

$$p = \Psi(\alpha, \phi) = \alpha^0 \sin(2\pi\phi) + \alpha^1 \cos(2\pi\phi), \tag{7}$$

with extrapolated phases and amplitudes. A decoder is then used to reconstruct the input motion sequence from the predicted embedding. A decent reconstruction can only be achieved if the learned mapping is close to phase linear and amplitude constant.

The encoder uses a 2-layer 1D convolutional network to map the input to an intermediate representation, which is then split into two branches: the timing branch and the amplitude branch. The timing branch predicts phase ϕ and frequency f from a temporal signal generated by a 1D convolution, while the amplitude branch predicts amplitude $\mathbf{A}$ by applying average pooling followed by an MLP, with vector quantization used to select the nearest amplitude from a finite codebook.

The phase is calculated using the relative timing $\mathcal{T}$ and the equation $\Phi = \phi + f \cdot \mathcal{T}$. The final motion embedding is obtained by $\mathbf{P} = \Psi(\mathbf{A}, \Phi)$. The decoder is a 2-layer 1D convolutional network that maps the embedding back to the original motion space.

Our decision to quantize only the amplitude is a principled choice to disentangle a motion's spatial "style" from its temporal dynamics. The amplitude vectors directly define the physical form—like the height of a step or the extent of an arm swing. Quantizing them creates a discrete vocabulary of these core motion "shapes", which is precisely the "physical style" we need to adapt between the different morphologies. In contrast, frequency and phase (or offset) govern the motion's timing. Quantizing frequency, which represents speed, would create artificial boundaries and prevent nuanced tempo control (e.g., forcing a 1.0Hz and 1.3Hz walk into different categories). Quantizing phase, which represents progress, would be even more detrimental, as it would destroy the smooth temporal flow and result in jerky, unnatural movement. Therefore, we adapt the physical form via a discrete amplitude vocabulary while preserving the crucial temporal continuity of frequency and phase. We also conducted experiments to validate this claim.

How it works. The alignment is achieved by the optimization pressure on a shared, discrete bottleneck. To minimize the joint reconstruction loss across both human and humanoid domains, the network's most efficient strategy is to map semantically similar motions (e.g., "human walking" and "robot walking") to the same discrete codebook. This shared codebook then acts as an abstract, cross-domain instruction, which each domain-specific decoder learns to interpret and reconstruct. This forces the model to discover a shared vocabulary of motion concepts without direct supervision. We will add a more detailed explanation of this mechanism to the camera-ready version.

Why VQ on Amplitude only. This choice is a principled method to disentangle a motion's "what/how" from its "when". Amplitude represents the physical style of the motion (the "what/how") [1]. This is what is physically different and often implausible between a human and a humanoid, and thus is what we need to adapt. Phase/Frequency represents the temporal dynamics—the rhythm and timing (the "when") [1]. We want to preserve this to maintain the original motion's intent. By quantizing only the amplitude, our VQ codebook learns a shared, abstract vocabulary of motion primitives. This bridges the domain gap. Keeping the phase continuous is crucial to avoid destroying the motion's natural temporal flow. This design allows us to adapt the physical form while retaining the original timing.

How to reduce retargeting effort. Previous methods use IK-retargeted human motion as the target goal. These goals can be physically implausible, forcing the policy to struggle with unrealistic targets. Our method uses the Humanoid-Adapter to first convert human motion into a physically plausible humanoid motion. This adapted motion serves as a high-quality, achievable goal. In short, while retargeting exists in our data pipeline, our policy avoids the effort of learning from physically implausible goals, which is the core problem we solve for improved stability and efficiency.

7.3 More Ablation Studies

Vector quantization over continuous latent representations. Our central contribution, the Humanoid-Adapter, is specifically designed to bridge this domain gap by adapting physically challenging or infeasible human motions into a plausible action space for the robot before the imitation

Term	Expression	Weight
DoF acceleration	$\|\ddot{d}_t\|_2^2$	$-3e^{-7}$
DoF position limits	$\mathbb{I}(d_t \notin [q_{\min}, q_{\max}])$	-10
DoF error	$\|d_t - d_0\|_2^2$	-0.5
Energy	$\|\tau_t \dot{d}_t\|_2^2$	-0.001
Linear velocity (z)	$\|v_t^{\text{lin-z}}\|_2^2$	-1
Angular velocity (xy)	$\|v_t^{\text{ang-xy}}\|_2^2$	-0.4
Action rate	$\|a_t - a_{t-1}\|_2^2$	-0.1
Torque	$\|\tau_t\|_2$	-0.0001
Feet air time	$T_{\text{air}} - 0.5$	10
Feet velocity	$\|v_{\text{feet}}\|_1$	-0.1
Feet contact force	$\|F_{\text{feet}}\|_2^2$	-0.003
Stumble	$\mathbb{I}(F_{\text{feet}}^x > 5 \times F_{\text{feet}}^z)$	-2
Hip pos error	$\|p_t^{\text{hip}} - p_0^{\text{hip}}\|_2^2$	-0.2
Waist roll pitch error	$\|p_t^{\text{wrp}} - p_0^{\text{wrp}}\|_2^2$	-1
Ankle Action	$\|a_t^{\text{ankle}}\|_2^2$	-0.1

Table 3: **Regularization rewards** Regularization rewards for preventing undesired behaviors for sim-to-real transfer.

Term	Expression	Weight
DoF Position (Upper)	$\exp(-0.7\|\mathbf{q}_{ref}^{upper} - \mathbf{q}^{upper}\|)$	3.0
DoF Position (Lower)	$\exp(-0.7\|\mathbf{q}_{ref}^{lower} - \mathbf{q}^{lower}\|)$	1.0
Keypoint Position (Upper)	$\exp(-\|\mathbf{p}_{ref}^{upper} - \mathbf{p}^{upper}\|)$	2.0
Keypoint Position (Lower)	$\exp(-\|\mathbf{p}_{ref}^{lower} - \mathbf{p}^{lower}\|)$	1.0
Linear Velocity	$\exp(-4.0\|\mathbf{v}_{ref} - \mathbf{v}\|)$	6.0
Velocity Direction	$\exp(-4.0\cos(\mathbf{v}_{ref}, \mathbf{v}))$	6.0
Roll & Pitch	$\exp(-\|\mathbf{\Omega}_{ref}^{\phi\theta} - \mathbf{\Omega}^{\phi\theta}\|)$	1.0
Yaw	$\exp(-\|\Delta y\|)$	1.0

Table 4: **Tracking Reward.** $\mathbf{q}_{ref}^{upper/lower}$ and $\mathbf{p}_{ref}^{upper/lower}$ denote the reference joint and keypoint positions for the upper and lower body, respectively. $\mathbf{v}_{ref}$ is the reference velocity of the body, while $\mathbf{\Omega}_{ref}^{\phi\theta}$ and $\mathbf{\Omega}^{\phi\theta}$ denote the reference and actual roll and pitch of the body.

learning process even begins. Our proposed SMAP explicitly addresses this challenge by introducing a Vector-Quantized (VQ) latent space. The discrete codebook in our VQ-PAE acts as a powerful information bottleneck, forcing the model to discover a shared vocabulary of abstract, cross-domain motion primitives. By compelling both human and robot motions to map to the same discrete code words, our Humanoid-Adapter achieves a robust semantic alignment that continuous-space models struggle to learn implicitly. While DFM Watanabe et al. (2025) has not yet released public source code, we have implemented a new baseline inspired by the core principles of FLD Li et al. (2024a) to directly address the reviewer's suggestion. Specifically, we replaced our VQ-PAE-based Humanoid-Adapter with a standard Periodic Autoencoder (PAE) trained using FLD's signature forward-prediction loss. This new baseline, which we term SAMP (FLD). The quantitative results, presented in the table below, strongly support our central claim that a Vector-Quantized (VQ) latent space is needed to align human and humanoid motion.

As shown in Tab. 5, continuous latent space models, used in FLD/DFM and standard VAEs, struggle with this alignment task. They lack a strong inductive bias to map semantically equivalent but physically distinct motions (e.g., a human's walk vs. a robot's walk) to the same latent region. To solve this, our Humanoid-Adapter employs Vector Quantization (VQ). The discrete codebook of VQ acts as a powerful information bottleneck, forcing the model to learn a shared, abstract vocabulary of motion primitives, thereby ensuring robust alignment. To empirically validate this, we implement a new baselines with VAE structure SMAP (VAE) as requested.

Method	Trained Motion Sample				Novel Motion Sample			
	$E_{vel} \downarrow$	$E_{mpkpe} \downarrow$	$E_{mpjpe} \downarrow$	$fail \downarrow$	$E_{vel} \downarrow$	$E_{mpkpe} \downarrow$	$E_{mpjpe} \downarrow$	$fail \downarrow$
SAMP	0.1698	0.0608	0.1181	1731	0.2331	0.0893	0.1458	266
SAMP (FLD)	0.3312	0.1121	0.1999	1920	0.3672	0.1287	0.2176	407
SAMP (VAE)	0.3472	0.1142	0.2004	1878	0.3723	0.1298	0.2131	389

Table 5: **Ablation Study**.

Codebook size	$\mathbf{E}_{mpkpe}$
16	12.3
32	10.6
64	10.7

Table 6: Mean per joint position error (cm) of Humanoid-Adapter.

7.4 More Details of Reward

In the main paper, we introduce the tracking reward. Tab. 3 and Tab. 4 shows details on the regularization reward.

7.5 More Details of Observation of RL

The student policy's observation comprises the following components: Base angular velocity, IMU measurements (roll and pitch angles), Directional difference between current yaw angle and target yaw (represented as sine and cosine terms), Joint positions and velocities, Observation history from the past n steps, and Target goal.

The teacher policy's observation extends the student policy's observation with privileged information, including: Foot contact flags, System dynamics parameters (mass, ground friction coefficients, motor strength parameters), External push forces.

7.6 More Details of Goal-conditioned Policy

In our framework, the goal state is a sequence of future target states for the humanoid robot to mimic. The key difference between our method and previous works lies in how these target goals are generated. Other Methods typically use human motion data that has been retargeted to the robot's skeleton by inverse kinematics (IK). These goals may be physically implausible for the robot to execute. Our Method uses human motion that has been processed by our Humanoid-Adapter. The Humanoid-Adapter is an network trained on a dataset of physically plausible, dynamically stable humanoid motions generated in simulation. Therefore, it transforms any input human motion into a more physically reasonable target goal for the humanoid. As shown in Table 1, removing the adapter and using only directly retargeted motion causes the failure count to jump from 266 to 392, a 47% increase in task failures. This directly demonstrates that the adapter provides more stable and executable goals.

Following Exbody, our policy is provided with target goals for the next 4 future frames. The H1 robot has a control frequency of 30Hz, which corresponds to a time step of approximately 33ms. Therefore, at any given time t, the goal provided to the policy consists of the target states for the following time steps: t+33ms, t+66ms, t+99ms, and t+132ms.

At each training step, the goal-conditioned policy receives the current proprioceptive state of the robot and the goal input. The goal input is the sequence of target motion states generated by our Humanoid-Adapter, as described above. The policy then outputs an action, and the reward is calculated based on how well the robot's state matches the target goal, guided by our decoupled reward function. This process iteratively teaches the policy to follow the sequence of physically plausible goals provided by the Humanoid-Adapter. In practice, more physically plausible target goals are generated by our Humanoid-Adapter, the pre-trained policy try to track these goals. In practice, physically plausible target goals are generated by our Humanoid-Adapter, and the pre-trained policy's objective is to track them just like training step.

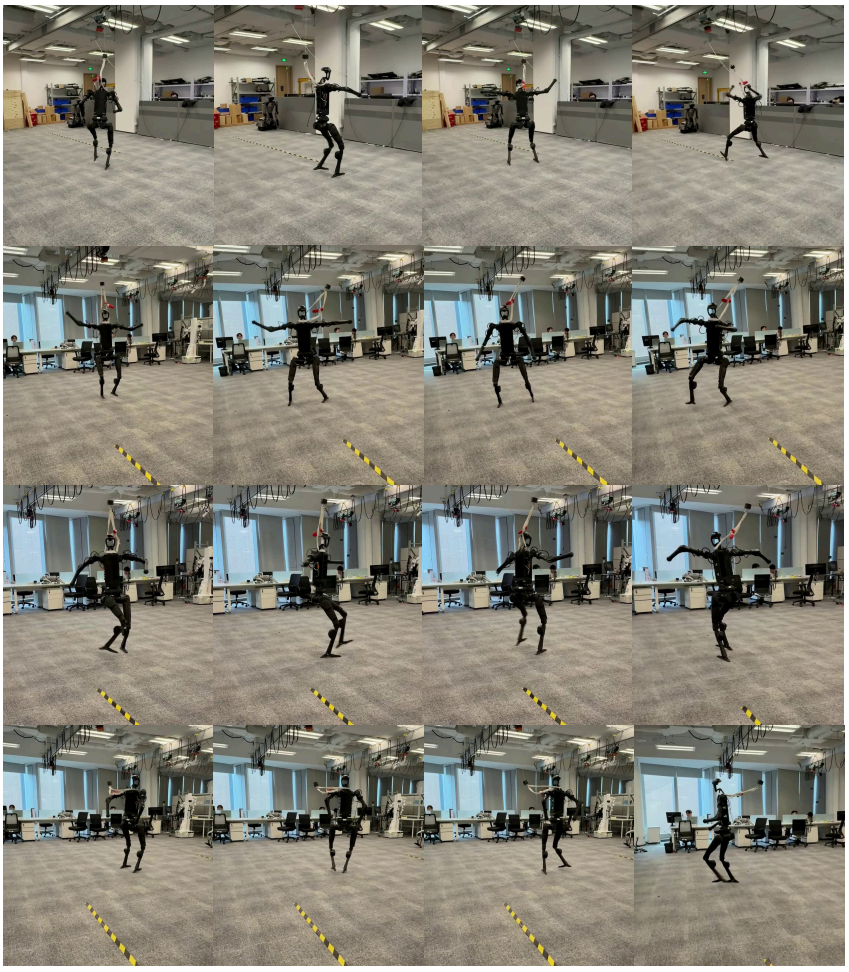

Figure 11: Expressive motion evaluation in the real world..

7.7 Performance of Humanoid-Adapter

We save a set of paired human motion and humanoid robot motion data (saved by simulator with a RL policy) to evaluate the performance of our Humanoid-Adapter, as shown in Tab. 6.

7.8 Additional Real World Results Visualization

We provide detailed visualization for some motions evaluated in the real world in Fig. 11.