ExBody2: Advanced Expressive Humanoid Whole-Body Control

Mazeyu Ji[*] UC San Diego
Xuanbin Peng[*] UC San Diego
Fangchen Liu UC Berkeley
Jialong Li UC San Diego
Ge Yang MIT
Xuxin Cheng[†] UC San Diego
Xiaolong Wang[†] UC San Diego

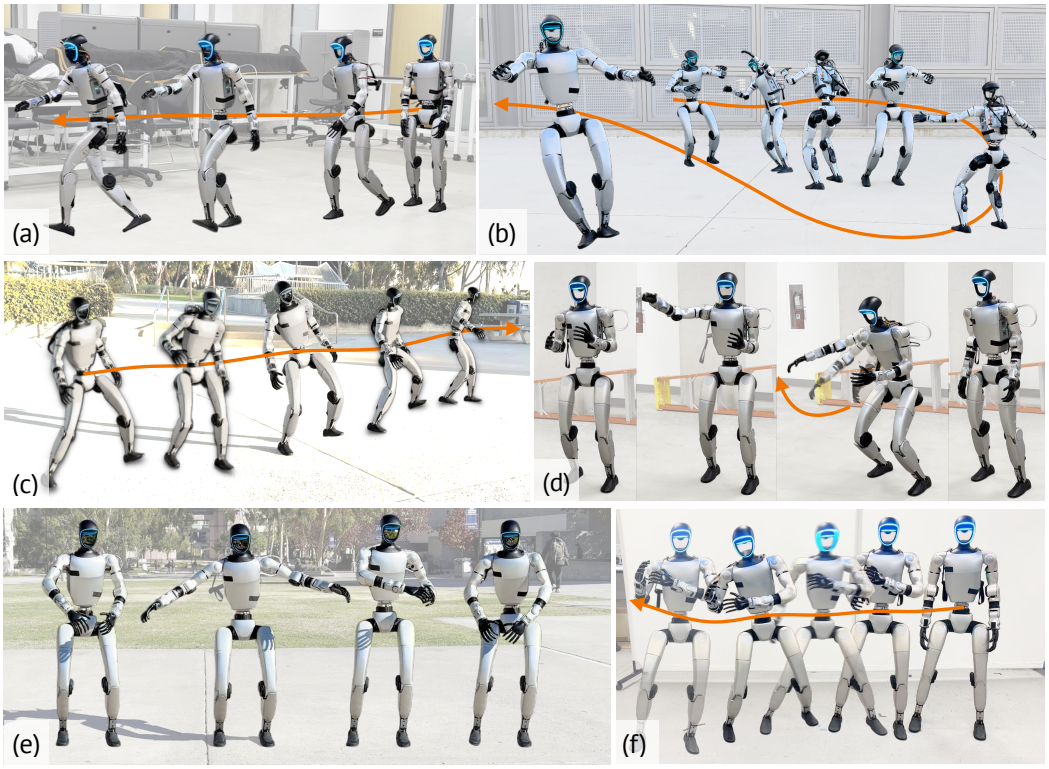

Figure 1: Humanoid robot executing various expressive whole-body motions in the real world. The robot can (a) walk with a large stride from static standing, (b) dance along a long horizon choreography, (c) dynamic sidestep with fluid weight shifts, (d) punch with different height configurations, (e) express various upper-body movements while maintaining balance, (f) powerful rightward body hook with dynamic shifts.

Abstract: This paper tackles the challenge of enabling real-world humanoid robots to perform expressive and dynamic whole-body motions while maintaining stability. We propose Advanced Expressive Whole-Body Control (ExBody2), a whole-body tracking framework trained in simulation with Reinforcement Learning and then transferred to the real world. The framework decouples keypoint tracking from velocity control and leverages a privileged teacher policy to distill precise mimic skills into the student policy, enabling robust, high-fidelity reproduction of complex motions such as walking, crouching, and dancing. A significant contribution is the discovery of a fundamental principle for balancing feasibility and diversity in motion datasets, which guides the development of an automatic dataset curation method. This principle facilitates pretraining a versatile model generalizing well across diverse motions and can be fine-tuned for specific

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

tasks to achieve superior tracking accuracy. Extensive experiments demonstrate that Exbody2 outperforms existing baselines, establishing new benchmarks and provides valuable insights for the advancement of whole-body humanoid control.

Keywords: Humanoid Robot, Reinforcement Learning, Whole-body Control

1 Introduction

The premise of humanoid robots is to enable human-like motions while occupying human living spaces. However, a humanoid robot with human-level *expressiveness* and *versatility* that is also robust in maintaining stability and control remains elusive. Inherent to this problem is the dynamic and kinematic gap between robots and biological body structures and the need for the controller to make a trade-off between expressiveness and stability. How to let robots imitate human whole-body motion across this gap and achieve both is a key challenge.

This paper introduces Advanced Expressive Whole-Body Control (Exbody2), a framework that enables humanoid robots to perform expressive, human-like full-body motions with grace. At its core, Exbody2 features both a generalist and a specialist policy. The generalist policy, trained on diverse motion datasets, **outperforms previous approaches** by achieving high adaptability across a wide range of motions with a single policy. Building on this, we further fine-tune the policy for specific motion groups, producing specialist policies that ensure **even higher fidelity in targeted behaviors**. To enable robust and expressive motion reproduction, the framework decouples global tracking into velocity control and local keypoint imitation, and incorporates well-designed rewards and network architecture. Together, these components allow Exbody2 to successfully reproduce expressive whole-body humanoid motions in the real world—to our knowledge, the first RL-based system to do so. The framework is composed of three core components:

(i) *Generalist policy with automated data curation*. Human motion datasets often contain movements beyond a robot's physical limits, making tracking difficult and reducing performance. Some methods refine datasets, like ExBody [1] filtering motions via language labels, though ambiguous terms (e.g., "dance") may still include infeasible actions. Others [2, 3] use SMPL avatars to simulate motions, but these can exceed real robot capabilities, impacting training. We identify the trade-off between dataset feasibility and diversity and develop an automated curation method that removes unsuitable lower-body motions while preserving diversity, enabling the policy to learn broad, expressive behaviors. Experiments validate that our method optimally balances feasibility and diversity, leading to improved stability and accuracy across diverse motion tasks.

(ii) *Specialist policy with finetuning for targeted motions*. While the generalist policy enables broad motion coverage, finetuning enhances precision for specific motion groups. Motions with similar patterns are easier to learn under a shared policy, as they require consistent control strategies and constraints. Instead of training from scratch, we refine the generalist policy, leveraging its learned priors for efficient adaptation. This allows the policy to better capture fine-grained motion details and improve tracking accuracy for specialized tasks. Additionally, motion labels or an action recognition model can classify input motions, enabling dynamic selection of the most suitable specialist policy.

(iii) *Tracking design and policy architecture.* Unlike H2O [2] and OmniH2O [3], which rely on global keypoint tracking and often struggle with long-horizon or dynamic motions, Exbody2 adopts a modular design that decouples tracking into velocity control and local keypoint imitation. This improves stability while preserving expressive motion details. Training follows a teacher-student framework: the teacher is optimized with PPO [4] using privileged information (e.g., root velocity, body positions), and the student is distilled via DAgger [5]-style learning to function without such inputs, enabling real-world deployment.

We evaluate Exbody2 on the Unitree G1 against the state-of-the-art baselines, achieving higher fidelity in both simulation and real-world tests. The curated generalist policy outperforms prior methods across diverse motions, while fine-tuning further improves quality for specific tasks. These

results demonstrate Exbody2's potential to bridge the gap between human-level expressiveness and robust whole-body control.

2 Related Work

Humanoid Whole-Body Control. Whole-body control for humanoid robots remains a complex and challenging problem due to the system's high non-linearity and degrees of freedom. Traditional approaches predominantly rely on dynamics modeling and control [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], which often require accurate system identification and physical modeling, and intensive online computation for real-time control to handle different external perturbations for locomotion stability. Recent advances in reinforcement learning (RL) and sim-to-real transfer have demonstrated promising results in enabling complex whole-body skills for humanoid robots [20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. These approaches typically rely on training RL policies in simulation using task-specific rewards and environment randomization before transferring them to the real world. Notably, recent works such as [1, 2, 32] have advanced real-world humanoid whole-body control for expressive motion by incorporating human motion datasets [33] to guide RL training, with real-world applications such as motion imitation. However, these approaches still exhibit limitations in expressiveness and maneuverability, highlighting the untapped potential of humanoid robots. In contrast, our method enables the learning of more expressive and dynamic motions, enhancing the robot's ability to perform complex whole-body movements.

Robot Motion Imitation. Robot motion imitation can be broadly categorized into manipulation and locomotion areas. For manipulation tasks, robots are often wheeled or tabletop, prioritizing precise control over balancing and ground contact, making humanoid morphology unnecessary. Such robots can imitate the motion through direct teleoperation [34, 35, 36], portable devices [37, 38, 39, 40] and learn from human videos with hand tracking or motion retargeting [41, 42, 43, 44]. In contrast, motion imitation for locomotion primarily aims to learn lifelike, natural behaviors from human or animal motion capture data. It requires precise control over contact dynamics, balance, and coordination across multiple degrees of freedom to achieve stable and realistic movement. While prior methods have enabled physics-based character motion imitation in simulation [45, 46, 47, 48, 49, 50, 51, 52, 53], transferring diverse motions to real robots [1, 3, 32, 54, 55, 56, 57] remains a significant challenge due to the hardware constraints. Previous methods [1, 3, 32, 54] typically rely on manually filtering feasible motion data with human effort or hand-crafted heuristics. However, manually filtered datasets may still contain infeasible motions or lack diversity, limiting the robot's ability to fully utilize its hardware potential. Our method overcomes this challenge by automatically curating a diverse and feasible motion dataset, enabling more effective real-world deployment.

3 Exbody2: Learning Expressive Humanoid Whole-Body Control

We propose Advanced Expressive Whole-Body Control (Exbody2), a motion mimic framework for expressive and robust whole-body control. As shown in Figure 2, Exbody2 first retargets human motion data to fit the robot's morphology, then trains a generalist policy using an automated dataset curation strategy to balance feasibility and diversity. To improve precision on specific motion groups, we further fine-tune specialist policies and deploy it onto real humanoid robots. In the following sections, we detail our generalist-specialist training pipeline, and our policy structure design, the two main contributions of our work.

3.1 Data-driven Generalist-specialist Training Pipeline

We adopt a Generalist–Specialist pipeline to balance adaptability and precision in whole-body motion tracking. This approach is guided by our **Feasibility–Diversity Principle**, which emphasizes retaining diverse upper-body motions to support task generalization, while filtering out extreme or unstable lower-body motions that hinder training. Based on this principle, we construct a pruned dataset that preserves motion diversity without compromising feasibility. A generalist policy is

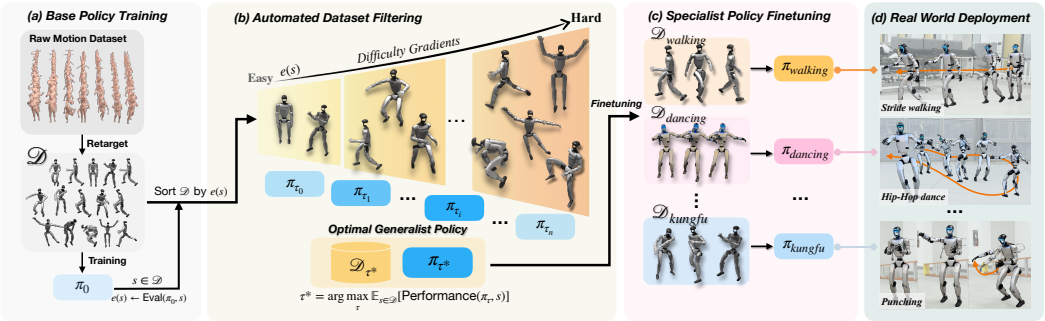

Figure 2: Exbody2's framework. (a) Motion retargeting adapts raw human motion datasets to fit the humanoid robot's morphology, generating a diverse set of training samples. (b) Automated dataset filtering ranks motions based on tracking errors and selects an optimal subset to train a generalist policy, balancing feasibility and diversity. (c) Specialist policy finetuning refines the generalist model for specific motion categories, improving precision for targeted tasks. (d) The trained policies are deployed on a real humanoid robot, demonstrating expressive, dynamic, and stable whole-body motions in real-world environments.

trained on this dataset and subsequently fine-tuned on specific tasks to obtain specialist policies with higher precision.

3.1.1 Generalist policy with automated data curation

To obtain a policy π that performs well across diverse motion inputs, we first train an initial policy π_0 on a comprehensive, unfiltered motion dataset $\mathcal{D}$, which is highly diverse with a lot of infeasible motions. After training π_0, we evaluate its tracking accuracy for each motion sequence $s \in \mathcal{D}$, obtaining a tracking error metric $e(s)$ that focuses on the lower body. The lower body plays a central role in dynamic feasibility and balance; thus, we focus on its tracking error for filtering. Specifically, we define

$$e(s) \;=\; \alpha\, E_{\text{key}}(s) \;+\; \beta\, E_{\text{dof}}(s),$$

where $E_{\text{key}}(s)$ is the mean keybody position error for the lower body (preventing extreme deviations such as flipping or rolling), and $E_{\text{dof}}(s)$ measures the mean joint-angle tracking error. The coefficients α and β weight these two terms according to their relative importance for lower-body stability and precision. Once $e(s)$ is computed for each sequence, we rank the motions by their tracking errors and derive the empirical distribution $P(e)$.

The objective is to determine an error threshold τ such that the subset of motion sequences with $e(s) \leq \tau$, denoted as $\mathcal{D}_\tau = \{s \in \mathcal{D} \mid e(s) \leq \tau\}$, enables the training of a new policy π_τ that maximizes performance across the full dataset $\mathcal{D}$. Formally, we seek:

$$\tau^* = \arg\max_{\tau} \mathbb{E}_{s\in\mathcal{D}}[\text{Performance}(\pi_\tau, s)],$$

where the performance is evaluated on the whole dataset. In practice, we divide $P(e)$ into evenly spaced error intervals to evaluate the performance of policies trained on subsets corresponding to different thresholds τ. Although we use a greedy search to identify the optimal threshold τ^*, subsequent experiments reveal a strong trend in how the policy's performance changes with τ. When τ is too small, the filtered motions are overly simple, limiting the policy's ability to generalize across the full dataset. Conversely, when τ is too large, the inclusion of many infeasible motions introduces noise, degrading the training effectiveness. The best-performing policy is consistently obtained at a moderate τ, balancing diversity and feasibility.

The optimal threshold τ^*, identified through this process, exhibits generalizability and can be effectively applied to other motion datasets, ensuring robust training and improved performance.

3.1.2 Specialist policy with finetuning for targeted motions

After obtaining the generalist policy π_{τ^*}, which balances motion diversity and feasibility, we refine it into a specialist policy for high-precision tasks through finetuning rather than training from scratch.

This approach is more efficient and effective, as specialist policies track fewer but more challenging motions, which are often difficult to learn without a strong prior.

To retain generalization and avoid overfitting to small specialized datasets, we apply a balanced sampling strategy during finetuning. Instead of limiting training to the specialist subset alone, we continue sampling from a broader motion distribution based on the difficulty gradients used in generalist training. This ensures the policy still encounters sufficient motion variety, helping it remain robust under complex real-world conditions.

In addition to improving adaptability and robustness, this finetuning approach significantly reduces training time and computational cost, making it a practical strategy for building high-fidelity controllers tailored to specific tasks.

3.2 Policy Objective and Architecture

Exbody2 aims at tracking a target motion more expressively in the whole body. To this end, Exbody2 adopts a two-stage teacher-student training procedure as in [58, 59]. Specifically, the oracle teacher policy is first trained with an off-the-shelf reinforcement learning (RL) algorithm, PPO [4], with privileged information that can be obtained only in simulators. For the second stage, we replace the privileged information with observations which are aligned with the real world, and distill the teacher policy to a deployable student policy.

3.2.1 Teacher Policy Training

We can formulate the humanoid motion control problem as a *Markov Decision Process* (MDP). The state space $\mathcal{S}$ contains privileged observation $\mathcal{X}$, proprioceptive states $\mathcal{O}$ and motion tracking target $\mathcal{G}$. A policy $\hat{\pi}$ takes $\{p_t, o_t, g_t\}$ as input, and outputs action $\hat{a}_t$. The predicted action $\hat{a}_t \in R^{23}$ is the target joint positions of joint proportional derivative (PD) controllers. We use off-the-shelf PPO [4] algorithm to maximize expectation of the accumulated future rewards $E_{\hat{\pi}}[\sum_{t=0}^{T} \gamma^t \mathcal{R}(s_t, \hat{a}_t)]$, which encourages tracking the demonstrations with robust behavior. The predicted $\hat{a}_t \in R^{23}$, which is the target position of joint proportional derivative (PD) controllers.

We train a teacher policy using privileged information p_t that is only available in simulation, including ground-truth root velocity, and keybody differences. This improves sample efficiency and is commonly used to obtain high-performing policies [60]. The policy learns to track full-body motions composed of joint angles, 3D keypoints, and root velocity and pose, while also responding to joystick commands for high-level control. The reward function is carefully designed to balance motion fidelity and stability, combining terms for root motion, keypoint and joint tracking, and regularization terms to improve sim-to-real transfer. Following this, we train a student policy without privileged information by using DAgger [5]: the student observes long-horizon histories and is supervised by the teacher's actions via an MSE loss. Training proceeds through iterative rollouts and updates until convergence. Full details, including reward definitions, observation structures, and policy architecture, are provided in the appendix.

3.2.2 Local Keybody Tracking Strategy

Motion tracking comprises two objectives: tracking DoF (joint) positions and keypoint (body keypoint) positions. Keypoint tracking usually plays a crucial role in tracking motions during training stage, as joint DoF errors can propagate to the whole body, while keypoint tracking is directly applied to the body. Existing work like H2O, OmniH2O [2, 3] learns to follow the trajectory of global keypoints. However, this global tracking strategy usually results in suboptimal or failed tracking behavior, as global keypoints may drift through time, resulting in cumulative errors that eventually hinder learning. To address this, we map global keypoints to the robot's current coordinate frame, and instead utilize velocity-based global tracking. The coordination of velocity and motion allows tracking completion with maximal expressiveness, even if slight positional deviations arise. Moreover, to further enhance the robot's capabilities in following challenging keypoint motions, we allow

Method	$E_{\text{vel}} \downarrow$	$E_{\text{mpkpe}} \downarrow$	$E^{\text{upper}}_{\text{mpkpe}} \downarrow$	$E^{\text{lower}}_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E^{\text{upper}}_{\text{mpjpe}} \downarrow$	$E^{\text{lower}}_{\text{mpjpe}} \downarrow$
Exbody†	0.4195	0.1150	0.1106	0.1198	0.1496	0.1416	0.1607
OmniH2O*	0.3725	0.1253	0.1266	0.1240	0.1681	0.1564	0.1843
Exbody2-w/o-Filter	**0.2787**	0.1133	0.1087	0.1182	0.1355	0.1192	0.1579
Exbody2(Ours)	0.2930	**0.1000**	**0.0960**	**0.1040**	**0.1079**	**0.0953**	**0.1253**

Table 1: Comparison on $\mathcal{D}_{CMU}$ using Unitree G1. Each motion is looped 10 times in simulation, and we report the per-step average error. Lowest errors per group are bolded.

Method	$E_{\text{mpjpe}} \downarrow$	$E^{\text{upper}}_{\text{mpjpe}} \downarrow$	$E^{\text{lower}}_{\text{mpjpe}} \downarrow$
Exbody†	0.1465	0.1314	0.1672
OmniH2O*	0.1396	0.1273	0.1533
Exbody2-w/o-Filter	0.1361	0.1254	0.1481
Exbody2(Ours)	**0.1074**	**0.1092**	**0.1054**

Table 2: Comparisons with baselines on selected motions for Unitree G1 in real world.

a small global drift of keypoints during training stage and periodically correct them to the robot's current coordinate frame.

4 Experiments

In this section, we present several experiments to evaluate Exbody2. We first introduce the experimental setup and the baselines, followed by a detailed analysis addressing the following questions:

Q1. (Section 4.2) Does Exbody2 generalist policy achieve higher tracking accuracy in both simulation and real-world deployment compared to prior methods?

Q2. (Section 4.3) What selection criteria lead to the optimal subset of a human motion dataset for learning a better generalist policy?

Q3. (Section 4.4) Does finetuning a specialist policy for specific motion groups further improve tracking performance?

4.1 Experimental Setup

Baselines. We compare three baselines on the CMU dataset [61], which features diverse action types. **Exbody**† is a whole-body version of Exbody [60] that tracks full-body poses from human motion data. **OmniH2O*** is our reproduction of OmniH2O [3], using global keypoint tracking and the original observation space, adapted to our local tracking setup for fair comparison. **Exbody2**, our method, adopts local keypoint tracking and curated training data, with additional techniques to boost motion fidelity and sim-to-real performance. All methods use the same regularization rewards to ensure that improvements stem from our training system rather than auxiliary factors.

Metrics. We evaluate policy performance using several metrics over all motion sequences. The *mean linear velocity error* E_{vel} (m/s) reflects the difference between the robot's and demonstration's root velocity. The *Mean Per Keybody Position Error* (MPKPE) E_{mpkpe} (m) measures keypoint tracking accuracy, with $E^{\text{upper}}_{\text{mpkpe}}$ and $E^{\text{lower}}_{\text{mpkpe}}$ (m) evaluating upper and lower body regions, respectively. The *Mean Per Joint Position Error* (MPJPE) E_{mpjpe} (rad) quantifies joint tracking, with $E^{\text{upper}}_{\text{mpjpe}}$ and $E^{\text{lower}}_{\text{mpjpe}}$ (rad) reported for finer analysis.

4.2 Generalist Policy Performance

As shown in Table 1, *Exbody2* outperforms prior baselines (*Exbody*†, and *OmniH2O**) across all simulation metrics when trained on the full dataset without motion filtering. With motion filtering, tracking performance improves further—especially in the lower body, enhancing global stability and

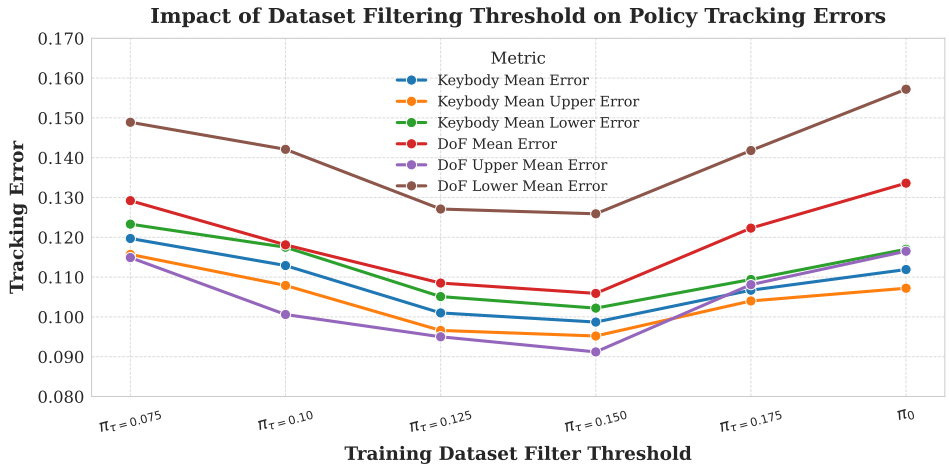

Figure 3: Impact of dataset filtering thresholds on tracking errors. Policies trained with balanced thresholds (e.g., $\pi_{\tau=0.150}$) achieve the lowest errors, while unfiltered data and overly strict ($\pi_{\tau=0.075}$) or loose ($\pi_{\tau=0.175}$) thresholds degrade performance. We compute the error as $e(s) = \alpha\, E_{\text{key}}(s) + \beta\, E_{\text{dof}}(s)$, with $\alpha = 0.1$, $\beta = 0.9$, giving higher weight to joint-angle accuracy.

indirectly benefiting upper-body precision. The only trade-off is a slight increase in velocity error due to reduced exposure to diverse velocity patterns, which is outweighed by gains in stability.

In real-world experiments (Table 2), we evaluate a diverse CMU motion subset covering postures, walking, squatting, and dancing. Results align with simulation trends: *Exbody2* achieves higher tracking accuracy across all body regions, with automated data curation playing a key role in ensuring robustness under real-world disturbances.

Overall, our generalist policy demonstrates strong tracking performance and stability across both simulation and real-world settings, outperforming baseline methods in accuracy and robustness.

4.3 Impact of Automatic Data Curation

To study how dataset composition affects generalist policy learning, we reconstruct the full pipeline of our automated data curation method and evaluate its impact on tracking performance. We first train a base policy π_0 on the unfiltered CMU dataset $\mathcal{D}_{\text{CMU}}$, which contains a wide range of motions, including physically infeasible or unstable sequences.

Based on the tracking errors produced by π_0, each motion sequence is assigned a score. We sort the sequences and apply thresholds $\tau \in \{0.075, 0.1, 0.125, 0.15, 0.175\}$ to construct progressively filtered datasets $\mathcal{D}_\tau$, each representing a different trade-off between feasibility and diversity. Lower thresholds preserve only the most stable motions, while higher thresholds allow more diverse but potentially unstable examples. These thresholds are derived from the error distribution of π_0, and the method generalizes to other datasets with similar policy architectures.

For each dataset $\mathcal{D}_\tau$, we resume training from the base policy to obtain a refined policy π_τ, leveraging the learned prior to improve training efficiency and adaptability. All resulting policies are then evaluated on the full dataset $\mathcal{D}_{\text{CMU}}$ using standard tracking metrics, as shown in Figure 3.

The results show that dataset quality significantly impacts tracking performance and generalization. Policies trained on low-threshold datasets (e.g., $\tau = 0.075$) are overly stable but lack diversity, leading to limited generalization. High-threshold datasets (e.g., $\tau = 0.175$ or unfiltered) introduce instability and noise, degrading accuracy. In contrast, the dataset $\mathcal{D}_{\tau=0.15}$ achieves the best balance between feasibility and diversity, resulting in the most robust and accurate policy $\pi_{\tau=0.15}$.

Method	$E_{\text{vel}} \downarrow$	$E_{\text{mpkpe}} \downarrow$	$E^{\text{upper}}_{\text{mpkpe}} \downarrow$	$E^{\text{lower}}_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E^{\text{upper}}_{\text{mpjpe}} \downarrow$	$E^{\text{lower}}_{\text{mpjpe}} \downarrow$
(a) $\mathcal{D}_{Moderate}$							
Specialist	**0.0991**	**0.0571**	**0.0582**	**0.0559**	**0.0760**	**0.0636**	**0.0930**
Scratch	0.1188	0.0676	0.0688	0.0663	0.0924	0.0794	0.1103
Generalist	0.1217	0.0741	0.0727	0.0755	0.1092	0.0914	0.1337
(b) $\mathcal{D}_{Hard}$							
Specialist	0.1712	**0.0827**	**0.0829**	**0.0826**	**0.1047**	**0.0911**	**0.1234**
Scratch	0.1631	0.0886	0.0898	0.0873	0.1188	0.1067	0.1354
Generalist	**0.1452**	0.0890	0.0867	0.0912	0.1181	0.1011	0.1414
(c) $\mathcal{D}_{ACCAD}$							
Specialist	0.4021	**0.1149**	**0.1079**	**0.1215**	**0.1402**	**0.1290**	**0.1557**
Scratch	0.4153	0.1246	0.1154	0.1332	0.1609	0.1490	0.1771
Generalist	**0.3361**	0.1268	0.1156	0.1391	0.1716	0.1532	0.1967

Table 3: Comparison of three training strategies across three dataset groups of different difficulties.

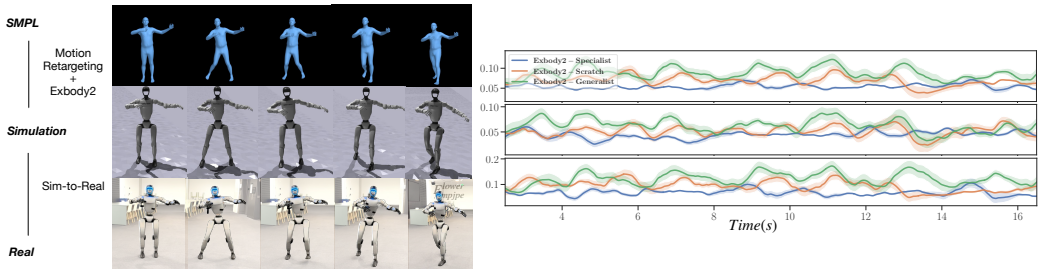

Figure 4: A sequence of a robot performing the Cha-Cha dance. Left three rows: reference avatar, simulation result, real robot. Right three rows: per-frame errors for whole-body, upper-body, and lower-body DoF. Blue: *Exbody2-Specialist* (finetuned on $\mathcal{D}dancing$), orange: *Exbody2-Scratch* (trained from scratch), green: *Exbody2-Generalist* (trained on filtered $\mathcal{D}CMU$).

4.4 Specialist Policy finetuning

We evaluate the effectiveness of the pretrain-finetune paradigm by comparing three training strategies: (1) a generalist policy $\pi_{\tau=0.15}$ trained on a curated dataset for broad motion coverage; (2) a specialist policy obtained by fine-tuning the generalist on task-specific data; and (3) a scratch-trained policy with the same total training steps as (2).

Experiments are conducted on three datasets: $\mathcal{D}_{\text{moderate}}$, and $\mathcal{D}_{\text{hard}}$, and $\mathcal{D}_{\text{ACCAD}}$ (out-of-distribution). As summarized in Table 3, the specialist policy consistently outperforms others across all datasets. The advantage of finetuning becomes more pronounced as motion complexity increases, and on the OOD dataset, it achieves the highest generalization performance. While the generalist policy shows better velocity tracking in dynamic cases, the specialist achieves higher overall precision.

To illustrate this, we present a case study on Cha-Cha dance motions (Figure 4). The specialist policy, fine-tuned on the dance set, achieves significantly lower tracking errors than both the generalist and scratch policies, capturing nuanced motion details while maintaining stability.

In summary, pretraining on diverse motions followed by task-specific finetuning yields robust, high-precision policies. This strategy is especially effective in challenging and unseen scenarios, combining generalization with specialization.

5 Conclusion

This paper introduces Advanced Expressive Whole-Body Control (Exbody2), a new framework for humanoid whole-body control that achieves high tracking accuracy, stability, and adaptability. It integrates automated dataset filtering, a generalist-specialist training pipeline, and local keybody

tracking using a teacher-student architecture. Experiments show that Exbody2 outperforms prior methods by balancing motion diversity and feasibility, enabling robust tracking and better generalization. Specialist finetuning further enhances precision for challenging tasks, validating the effectiveness of the structured pretrain-finetune paradigm.

Acknowledgments

If a paper is accepted, the final camera-ready version will (and probably should) include acknowledgments. All acknowledgments go at the end of the paper, including thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

References

[1] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for humanoid robots, 2024. URL https://arxiv.org/abs/2402.16796.

[2] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi. Learning human-to-humanoid real-time whole-body teleoperation. *arXiv preprint arXiv:2403.04436*, 2024.

[3] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning, 2024. URL https://arxiv.org/abs/2406.08858.

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

[5] 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*, 2011.

[6] H. Miura and I. Shimoyama. Dynamic walk of a biped. *IJRR*, 1984.

[7] K. Yin, K. Loken, and M. Van de Panne. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics*, 2007.

[8] M. Hutter, C. Gehring, D. Jud, A. Lauber, C. D. Bellicoso, V. Tsounis, J. Hwangbo, K. Bodie, P. Fankhauser, M. Bloesch, et al. Anymal-a highly mobile and dynamic quadrupedal robot. In *IROS*, 2016.

[9] F. L. Moro and L. Sentis. Whole-body control of humanoid robots. *Humanoid Robotics: A reference, Springer, Dordrecht*, 2019.

[10] B. Dariush, M. Gienger, B. Jian, C. Goerick, and K. Fujimura. Whole body humanoid control from human motion descriptors. In *2008 IEEE International Conference on Robotics and Automation*, pages 2677–2684. IEEE, 2008.

[11] S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa. The 3d linear inverted pendulum mode: A simple modeling for a biped walking pattern generation. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, volume 1, pages 239–246. IEEE, 2001.

[12] E. R. Westervelt, J. W. Grizzle, and D. E. Koditschek. Hybrid zero dynamics of planar biped walkers. *IEEE transactions on automatic control*, 48(1):42–56, 2003.

[13] I. Kato. Development of wabot 1. *Biomechanism*, 2:173–214, 1973.

[14] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka. The development of honda humanoid robot. In *Proceedings. 1998 IEEE international conference on robotics and automation (Cat. No. 98CH36146)*, volume 2, pages 1321–1326. IEEE, 1998.

[15] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim. The mit humanoid robot: Design, motion planning, and control for acrobatic behaviors. In *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, pages 1–8. IEEE, 2021.

[16] A. Dallard, M. Benallegue, F. Kanehiro, and A. Kheddar. Synchronized human-humanoid motion imitation. *IEEE Robotics and Automation Letters*, 8(7):4155–4162, 2023. doi:10.1109/LRA.2023.3280807.

[17] K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. J. A. Chavez, and D. Pucci. Whole-body geometric retargeting for humanoid robots. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 679–686, 2019. doi:10.1109/Humanoids43949.2019.9035059.

[18] L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, and S. Ivaldi. A multimode teleoperation framework for humanoid loco-manipulation: An application for the icub robot. *IEEE Robotics and Automation Magazine*, 26(4):73–82, 2019. doi:10.1109/MRA.2019.2941245.

[19] J. Ramos and S. Kim. Dynamic locomotion synchronization of bipedal robot and human operator via bilateral feedback teleoperation. *Science Robotics*, 4(35):eaav4282, 2019. doi:10.1126/scirobotics.aav4282. URL https://www.science.org/doi/abs/10.1126/scirobotics.aav4282.

[20] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath. Reinforcement learning for robust parameterized locomotion control of bipedal robots. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2811–2817. IEEE, 2021.

[21] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath. Robust and versatile bipedal jumping control through multi-task reinforcement learning. *arXiv preprint arXiv:2302.09450*, 2023.

[22] J. Siekmann, K. Green, J. Warila, A. Fern, and J. Hurst. Blind bipedal stair traversal via sim-to-real reinforcement learning. *arXiv preprint arXiv:2105.08328*, 2021.

[23] H. Duan, B. Pandit, M. S. Gadde, B. J. van Marum, J. Dao, C. Kim, and A. Fern. Learning vision-based bipedal locomotion for challenging terrain. *arXiv preprint arXiv:2309.14594*, 2023.

[24] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath. Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control. *arXiv preprint arXiv:2401.16889*, 2024.

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

[26] I. Radosavovic, B. Zhang, B. Shi, J. Rajasegaran, S. Kamat, T. Darrell, K. Sreenath, and J. Malik. Humanoid locomotion as next token prediction. *arXiv:2402.19469*, 2024.

[27] H. Ito, K. Yamamoto, H. Mori, and T. Ogata. Efficient multitask learning with an embodied predictive model for door opening and entry with whole-body control. *Science Robotics*, 7 (65):eaax8177, 2022.

[28] S. Jeon, M. Jung, S. Choi, B. Kim, and J. Hwangbo. Learning whole-body manipulation for quadrupedal robot. *arXiv preprint arXiv:2308.16820*, 2023.

[29] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath. Real-world humanoid locomotion with reinforcement learning. *arXiv:2303.03381*, 2023.

[30] A. Tang, T. Hiraoka, N. Hiraoka, F. Shi, K. Kawaharazuka, K. Kojima, K. Okada, and M. Inaba. Humanmimic: Learning natural locomotion and transitions for humanoid robot via wasserstein adversarial imitation. *arXiv preprint arXiv:2309.14225*, 2023.

[31] M. Seo, S. Han, K. Sim, S. H. Bang, C. Gonzalez, L. Sentis, and Y. Zhu. Deep imitation learning for humanoid loco-manipulation through human teleoperation. In *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, pages 1–8. IEEE, 2023.

[32] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. Humanplus: Humanoid shadowing and imitation from humans, 2024. URL https://arxiv.org/abs/2406.10454.

[33] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. Amass: Archive of motion capture as surface shapes. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. URL https://amass.is.tue.mpg.de.

[34] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

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

[36] Y. Qin, H. Su, and X. Wang. From one hand to multiple hands: Imitation learning for dexterous manipulation from single-camera teleoperation. *IEEE Robotics and Automation Letters*, 7(4): 10873–10881, 2022.

[37] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu. Dexcap: Scalable and portable mocap data collection system for dexterous manipulation. *arXiv preprint arXiv:2403.07788*, 2024.

[38] S. Chen, C. Wang, K. Nguyen, L. Fei-Fei, and C. K. Liu. Arcap: Collecting high-quality human demonstrations for robot learning with augmented reality feedback. *arXiv preprint arXiv:2410.08464*, 2024.

[39] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[40] H. Ha, Y. Gao, Z. Fu, J. Tan, and S. Song. Umi on legs: Making manipulation policies mobile with manipulation-centric whole-body controllers. *arXiv preprint arXiv:2407.10353*, 2024.

[41] C. Wang, L. Fan, J. Sun, R. Zhang, L. Fei-Fei, D. Xu, Y. Zhu, and A. Anandkumar. Mimicplay: Long-horizon imitation learning by watching human play. *arXiv preprint arXiv:2302.12422*, 2023.

[42] M. K. Srirama, S. Dasari, S. Bahl, and A. Gupta. Hrp: Human affordances for robotic pre-training. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, 2024.

[43] J. Li, Y. Zhu, Y. Xie, Z. Jiang, M. Seo, G. Pavlakos, and Y. Zhu. Okami: Teaching humanoid robots manipulation skills through single video imitation. In *8th Annual Conference on Robot Learning*, 2024.

[44] S. Kareer, D. Patel, R. Punamiya, P. Mathur, S. Cheng, C. Wang, J. Hoffman, and D. Xu. Egomimic: Scaling imitation learning via egocentric video, 2024. URL https://arxiv.org/abs/2410.24221.

[45] 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)*, 40(4): 1–20, 2021.

[46] 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.*, 41(4), July 2022.

[47] C. Tessler, Y. Kasten, Y. Guo, S. Mannor, G. Chechik, and X. B. Peng. Calm: Conditional adversarial latent models for directable virtual characters. In *ACM SIGGRAPH 2023 Conference Proceedings*, SIGGRAPH '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701597. doi:10.1145/3588432.3591541. URL https://doi.org/10.1145/3588432.3591541.

[48] M. Hassan, Y. Guo, T. Wang, M. Black, S. Fidler, and X. B. Peng. Synthesizing physical character-scene interactions. 2023. doi:10.1145/3588432.3591525. URL https://doi.org/10.1145/3588432.3591525.

[49] 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. URL https://openreview.net/forum?id=OrOd8PxOO2.

[50] H. Y. Ling, F. Zinno, G. Cheng, and M. Van De Panne. Character controllers using motion vaes. *ACM Transactions on Graphics (TOG)*, 39(4):40–1, 2020.

[51] H. Zhang, Y. Yuan, V. Makoviychuk, Y. Guo, S. Fidler, X. B. Peng, and K. Fatahalian. Learning physically simulated tennis skills from broadcast videos. *ACM Trans. Graph.*, 42(4), jul 2023. ISSN 0730-0301. doi:10.1145/3592408. URL https://doi.org/10.1145/3592408.

[52] J. Wang, J. Hodgins, and J. Won. Strategy and skill learning for physics-based table tennis animation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024.

[53] 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.

[54] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, C. Liu, G. Shi, X. Wang, L. Fan, and Y. Zhu. Hover: Versatile neural whole-body controller for humanoid robots. *arXiv preprint arXiv:2410.21229*, 2024.

[55] X. B. Peng, E. Coumans, T. Zhang, T.-W. E. Lee, J. Tan, and S. Levine. Learning agile robotic locomotion skills by imitating animals. In *Robotics: Science and Systems*, 07 2020. doi:10.15607/RSS.2020.XVI.064.

[56] A. Escontrela, X. B. Peng, W. Yu, T. Zhang, A. Iscen, K. Goldberg, and P. Abbeel. Adversarial motion priors make good substitutes for complex reward functions. 2022 ieee. In *International Conference on Intelligent Robots and Systems (IROS)*, volume 2, 2022.

[57] P. Dugar, A. Shrestha, F. Yu, B. van Marum, and A. Fern. Learning multi-modal whole-body control for real-world humanoid robots, 2024. URL https://arxiv.org/abs/2408.07295.

[58] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning quadrupedal locomotion over challenging terrain. *Science robotics*, 5(47):eabc5986, 2020.

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

[60] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[61] Carnegie Mellon University. Carnegie-Mellon mocap database. http://mocap.cs.cmu.edu/, Mar 2007. [Online].

[62] A. S. Huang, E. Olson, and D. C. Moore. Lcm: Lightweight communications and marshalling. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4057–4062, 2010. URL https://api.semanticscholar.org/CorpusID:10900899.

[63] J. Li, C. Xu, Z. Chen, S. Bian, L. Yang, and C. Lu. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3383–3393, 2021.

A Environments

A.1 Real-world Deployment

Our real robot employs a Unitree G1 platform, with an onboard Jetson Orin NX acting as the primary computing and communication device. 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) [62].

A.2 State Space Definition

In this section, we provide detailed information on the state space used for policy training, including proprioceptive states, privileged information, and motion tracking targets.

Robot Proprioceptive States. The robot proprioceptive states for the teacher and the student policy can be found in Table 4. Note that the student policy is trained on longer history length compared to the teacher, as it cannot observe privileged information but have to learn from a longer sequence of past observations.

Privileged Information. The teacher policy leverages privileged information to obtain accurate motion-tracking performance. The complete information about the privileged states is listed in Table 5.

Tracking Target Information. Both the teacher policy and student policy also take the motion tracking goal as part of their observations, which consists of the keypoint positions, DoF (joint) positions, as well as root movement information. The detailed components of the motion tracking target can be found in Table 6.

Action Space. The action is the target position of joint proportional derivative (PD) controllers, which is 23 dimensions for Unitree G1.

State	**Dimensions**
DoF position	23
DoF velocity	23
Last Action	23
Root Angular Velocity	3
Roll	1
Pitch	1
Yaw	1
Total Dim	**75*10**

Table 4: Proprioceptive states used in Exbody2. The rotation information is from IMU. 10 is the length of the history proprioception

State	**Dimensions**
Keybody Difference	36
Keybody Pos	36
Root velocity	3
Total dim	**75**

Table 5: Privileged information used in Exbody2.

State	Dimensions
DoF position	23
Keypoint position	36
Root Velocity	3
Root Angular Velocity	3
Roll	1
Pitch	1
Yaw	1
Height	1
Total dim	**69**

Table 6: Reference information used in Exbody2.

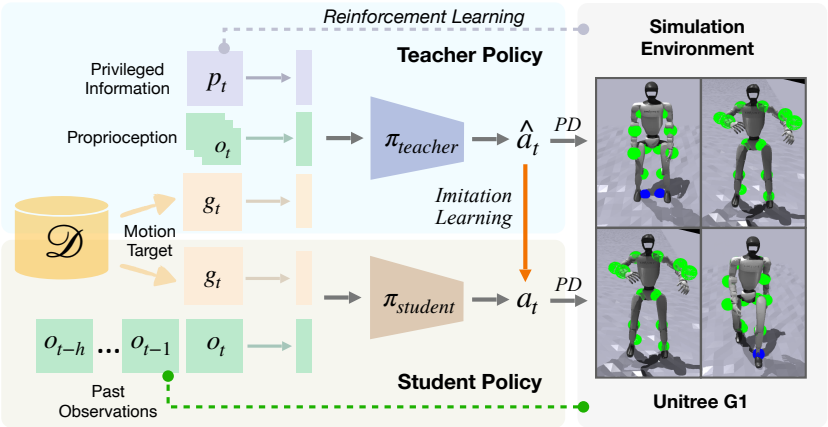

Figure 5: Teacher-student framework for humanoid motion learning, where the teacher uses privileged information, and the student learns from past observations to generate control actions.

B Model and Training Details

B.1 Policy Training Hyper-parameters

Exbody2 adopts a teacher-student training framework, as illustrated in Figure 5. The teacher policy is trained with standard PPO [4] algorithm on privileged information, tracking target and proprioceptive states. The student policy is trained with Dagger [5] without privileged information, but using longer history. For both teacher and student policies, we concatenate the corresponding inputs and feed them into MLP layers for policy learning. We provide the detailed training hyper-parameters for our teacher and student policy in Table 7.

B.2 Reward Design

Table 8 lists the tracking-based reward components, while Table 9 summarizes the additional regularization terms and their corresponding weights. The final reward is computed as a weighted sum of these components and is used to train a robust RL policy.

C Empirical Analysis of Dataset Selection

In the main paper, we propose a **Feasibility–Diversity Principle**, which posits that a good motion dataset for humanoid tracking must be:

Hyperparameter	Value
Optimizer	Adam
β_1, β_2	0.9, 0.999
Learning Rate	$1e^{-4}$
Batch Size	4096
Teacher Policy	
Discount factor (γ)	0.99
Clip Param	0.2
Entropy Coef	0.005
Max Gradient Norm	1
Learning Epoches	5
Mini Batches	4
Value Loss Coef	1
Entropy Coef	0.005
Value MLP Size	[512, 256, 128]
Actor MLP Size	[512, 256, 128]
Student Policy	
Student Policy MLP Size	[1024, 1024, 512]

Table 7: Hyperparameters related to the teacher and student policy's training.

Term	Expression	Weight
Expression Goal G^e		
DoF Position	$\exp(-0.7\|\mathbf{q}_{\text{ref}} - \mathbf{q}\|)$	3.0
Keypoint Position	$\exp(-\|\mathbf{p}_{\text{ref}} - \mathbf{p}\|)$	2.0
Root Movement Goal G^m		
Linear Velocity	$\exp(-4.0\|\mathbf{v}_{\text{ref}} - \mathbf{v}\|)$	6.0
Velocity Direction	$\exp(-4.0\cos(\mathbf{v}_{\text{ref}}, \mathbf{v}))$	6.0
Roll & Pitch	$\exp(-\|\mathbf{\Omega}_{\text{ref}}^{\phi\theta} - \mathbf{\Omega}^{\phi\theta}\|)$	1.0
Yaw	$\exp(-\|\Delta y\|)$	1.0

Table 8: Tracking rewards specification for Exbody2.

Term	**Expression**	**Weight**
DoF position limits	$\mathbb{1}(d_t \notin [q_{\min}, q_{\max}])$	-10
DoF acceleration	$\|\ddot{d}_t\|_2^2$	$-3e^{-7}$
DoF error	$\|d_t - d_0\|_2^2$	-0.1
Action rate	$\|a_t - a_{t-1}\|_2^2$	-0.1
Feet air time	$T_{\text{air}} - 0.5$	10
Feet contact force	$\|F_{\text{feet}}\|_2^2$	-0.003
Stumble	$\mathbb{1}(F_{\text{feet}}^x > 5 \times F_{\text{feet}}^z)$	-2
Waist roll pitch error	$\|p_t^{\text{wrp}} - p_0^{\text{wrp}}\|_2^2$	-0.5
Ankle Action	$\|a_t^{\text{ankle}}\|_2^2$	-0.1

Table 9: Regularization rewards for preventing undesired behaviors for sim-to-real transfer and refined motion.

Training Dataset	In dist.	Metrics						
		$E_{\text{vel}} \downarrow$	$E_{\text{mpkpe}} \downarrow$	$E^{\text{upper}}_{\text{mpkpe}} \downarrow$	$E^{\text{lower}}_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E^{\text{upper}}_{\text{mpjpe}} \downarrow$	$E^{\text{lower}}_{\text{mpjpe}} \downarrow$
(a) Eval. on $\mathcal{D}_{50}$								
$\mathcal{D}_{50}$	✓	**0.1375**	**0.0627**	**0.0571**	**0.0682**	**0.0753**	**0.0626**	**0.0928**
$\mathcal{D}_{250}$	✓	0.1454	0.0669	0.0600	0.0738	0.0870	0.0689	0.1119
$\mathcal{D}_{\text{CMU}}$	✓	0.1543	0.0767	0.0649	0.0885	0.1099	0.0854	0.1437
(b) Eval. on $\mathcal{D}_{\text{CMU}}$								
$\mathcal{D}_{50}$	✗	0.3509	0.1076	0.1074	0.1076	0.1338	0.1285	0.1410
$\mathcal{D}_{250}$	✗	0.2834	**0.1048**	**0.1021**	**0.1073**	**0.1148**	**0.1012**	**0.1335**
$\mathcal{D}_{\text{CMU}}$	✓	**0.2622**	0.1071	0.1036	0.1110	0.1291	0.1129	0.1512
(c) Eval. on $\mathcal{D}_{\text{ACCAD}}$								
$\mathcal{D}_{50}$	✗	0.4226	0.1277	0.1210	0.1330	0.1720	0.1618	0.1861
$\mathcal{D}_{250}$	✗	0.3533	**0.1234**	**0.1141**	**0.1315**	**0.1421**	**0.1223**	**0.1692**
$\mathcal{D}_{\text{CMU}}$	✗	**0.3452**	0.1267	0.1146	0.1381	0.1780	0.1635	0.1979

Table 10: Dataset Ablation Study: Evaluation on $\mathcal{D}_{50}$, $\mathcal{D}_{\text{CMU}}$, and $\mathcal{D}_{\text{ACCAD}}$ datasets with models trained on various datasets. Statistically significant results are highlighted in bold across 5 random seeds.

1. *Diverse enough* (especially in upper-body movements) to ensure the learned policy can generalize beyond very simple or repetitive actions.
2. *Feasible enough* that lower-body motions do not exceed the robot's mechanical limits, avoiding extreme samples (e.g., tumbling, handstands) that hamper training.

To illustrate how we arrived at this principle, We manually design three datasets of varying sizes, where the largest being the complete CMU dataset. The remaining datasets, with sizes 50, and 250, are subsets of the CMU dataset, each constructed with different levels of action diversity:

- **50-action dataset** ($\mathcal{D}_{50}$): A minimal set containing only fundamental and mostly static actions (e.g., standing, simple walking). While highly feasible, it lacks diversity in both upper and lower limb motions.
- **250-action dataset** ($\mathcal{D}_{250}$): A moderate-sized set extending $\mathcal{D}_{50}$ with additional upper-limb variations (e.g., arm gestures, some dance moves) and moderately dynamic lower-body actions (e.g., running, mild jumps). Crucially, it avoids highly extreme motions that are difficult for the robot to replicate.
- **Full CMU dataset** ($\mathcal{D}_{CMU}$): The complete CMU motion-capture repository of 1,919 sequences, including extreme movements like push-ups, rolling on the ground, and somersaults. Although highly diverse, it contains many infeasible actions that can introduce significant training noise.

We train separate policies with our Exbody2 framework on each dataset above and test them on three different evaluation sets:

1. $\mathcal{D}_{50}$ (in-distribution for the simplest data).
2. $\mathcal{D}_{CMU}$ (the full, more complex dataset).
3. $\mathcal{D}_{ACCAD}$, an out-of-distribution set containing actions not found in any of the training subsets.

Table 10 summarizes our findings:

- **Evaluation on $\mathcal{D}_{50}$**: Policies trained on $\mathcal{D}_{50}$ unsurprisingly achieve the highest tracking accuracy for *in-distribution* actions, as reflected in metrics across all categories. This sug-

Method	$E_{\text{vel}} \downarrow$	$E_{\text{mpkpe}} \downarrow$	$E^{\text{upper}}_{\text{mpkpe}} \downarrow$	$E^{\text{lower}}_{\text{mpkpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$E^{\text{upper}}_{\text{mpjpe}} \downarrow$	$E^{\text{lower}}_{\text{mpjpe}} \downarrow$
(a) History Length Ablation							
Exbody2-History10 (Ours)	0.2930	**0.1000**	0.0960	**0.1040**	**0.1079**	**0.0953**	**0.1253**
Exbody2-History0	0.4151	0.1047	0.1010	0.1081	0.1119	0.0986	0.1303
Exbody2-History25	0.2950	0.1032	0.0984	0.1078	0.1128	0.0965	0.1351
Exbody2-History50	**0.2648**	0.1004	**0.0956**	0.1051	0.1114	0.0967	0.1317
Exbody2-History100	0.3242	0.1063	0.1001	0.1122	0.1225	0.1050	0.1466
(b) DAgger Ablation							
Exbody2(Ours)	**0.2930**	**0.1000**	**0.0960**	**0.1040**	**0.1079**	**0.0953**	**0.1253**
Exbody2-w/o-DAgger	0.4195	0.1150	0.1106	0.1198	0.1496	0.1416	0.1607

Table 11: Self Ablation Study: Evaluation of different configurations of our method on dataset $\mathcal{D}_{CMU}$. The table is divided into two parts: (a) History Length Ablation and (b) DAgger Ablation.

gests that additional data does not necessarily benefit in-distribution tasks. While the policy trained on $\mathcal{D}_{250}$ performs similarly to $\mathcal{D}_{50}$, the policy trained on $\mathcal{D}_{CMU}$ exhibits a substantial drop in tracking accuracy.

- **Evaluation on** $\mathcal{D}_{CMU}$: Policies trained on $\mathcal{D}_{250}$ achieve the best performance on $\mathcal{D}_{CMU}$, surpassing those trained on the full $\mathcal{D}_{CMU}$ dataset. Due to the limited diversity of the $\mathcal{D}_{50}$ dataset, especially in upper limb movements, the $\mathcal{D}_{50}$-trained policy struggles to maintain high accuracy for out-of-distribution actions. Unexpectedly, the $\mathcal{D}_{250}$-trained policy generalizes better than the one trained on $\mathcal{D}_{CMU}$. This result underscores that noisy datasets degrade policy performance, as the policy may expend unnecessary effort on tracking infeasible actions, lowering the accuracy of feasible actions.
- **Evaluation on** $\mathcal{D}_{ACCAD}$: This experiment further emphasizes the importance of clean datasets. Here, the ACCAD dataset ($\mathcal{D}_{ACCAD}$) consists of actions that are entirely not in the training data. The policy trained on $\mathcal{D}_{250}$ outperforms the others, achieving the best tracking accuracy. Additionally, the $\mathcal{D}_{250}$ and $\mathcal{D}_{CMU}$-trained policies perform relatively well in velocity tracking. However, the $\mathcal{D}_{50}$-trained policy suffers from substantial tracking errors, suggesting the limitations of a small, simple dataset in handling unseen data.

In conclusion, these results validate the core insight behind our **Feasibility–Diversity Principle**. A small dataset ($\mathcal{D}_{50}$) is indeed easy for the policy to master but lacks sufficient variety to generalize well. On the other hand, a fully unfiltered large dataset ($\mathcal{D}_{CMU}$)—while highly diverse—contains many motions well beyond the robot's capabilities, introducing detrimental noise. The $\mathcal{D}_{250}$ subset thus provides the best balance between feasible lower-body motions and diverse upper-body actions, enabling our policy to learn robust and expressive whole-body control.

D Policy Ablation and Additional Results

D.1 Ablation on Policy Training

We conduct ablation studies on our policy design to highlight the effectiveness of both (i) the history length for the student policy and (ii) the teacher–student (DAgger) distillation.

History length. We test student policies trained with different history lengths in Table 11 (a). When no extra history is used, the policy struggles to learn effectively. Among the non-zero history lengths, most policies perform similarly while the history length of 10 yields the best results, which is used by us in the main experiments. Longer history lengths increase the difficulty of fitting the privileged information, ultimately reducing tracking performance.

Teacher–student distillation. Table 11 (b) shows that removing DAgger-style distillation severely degrades performance. Without privileged velocity guidance, the student policy must learn velocity

tracking directly from raw observations, making it harder to track fast or dynamic motions accurately.

D.2 Distribution-Guided Threshold Selection

To choose filtering thresholds in a principled manner, we first analyze the error distribution of the base policy across the entire dataset. Figure 6 presents the empirical cumulative distribution of $e(s)$, with the x-axis indicating the percentile of motion sequences (from lowest to highest error) and the y-axis displaying the corresponding error value.

We derive thresholds directly from the empirical distribution, ensuring a data-driven rather than arbitrary cutoff. Smaller thresholds yield mostly lower-body motions with limited dynamics, while gradually increasing the threshold admits more dynamic behaviors. Higher thresholds include samples with excessive errors that could degrade policy learning. Consequently, we select $\tau = 0.075, 0.10, 0.125, 0.15, 0.175$ to filter the dataset into subsets of varying feasibility and diversity. This data-driven approach aligns with our **feasibility-diversity principle**, yielding balanced subsets that support robust policy learning.

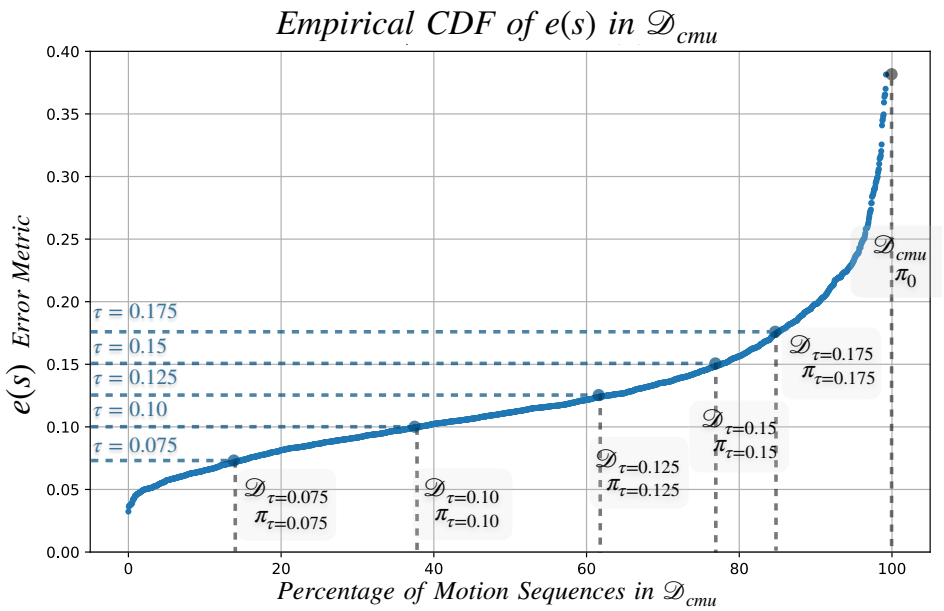

Figure 6: Empirical CDF of the base policy's error metric $e(s)$ on the entire $\mathcal{D}_{\text{CMU}}$ dataset. The horizontal axis indicates the percentile of motion sequences from 0% (lowest error) to 100% (highest error), while the vertical axis shows $e(s)$. We overlay dashed horizontal lines at key thresholds ($\tau = 0.075, 0.10, 0.125, 0.15, 0.175$) to illustrate how we systematically determine feasible versus unfeasible motions based on the empirical distribution.

D.3 Real-world Results Visualization

Figure 7 illustrates how ExBody2 successfully replicates various motions in both simulation and real-world settings. We align each frame's pose from (i) the reference SMPL animation, (ii) our simulated humanoid robot, and (iii) the real robot deployment. These snapshots confirm that our learned policy retains high fidelity to the target motion, including lower-body poses critical for balance. Additional results can be viewed in the supplementary video.

E ExBody2's Multi-source Demonstration

One key advantage of ExBody2 is its flexibility in handling multiple motion sources. In the main text, we focus primarily on motion capture data (i.e., offline datasets). Below, we highlight two

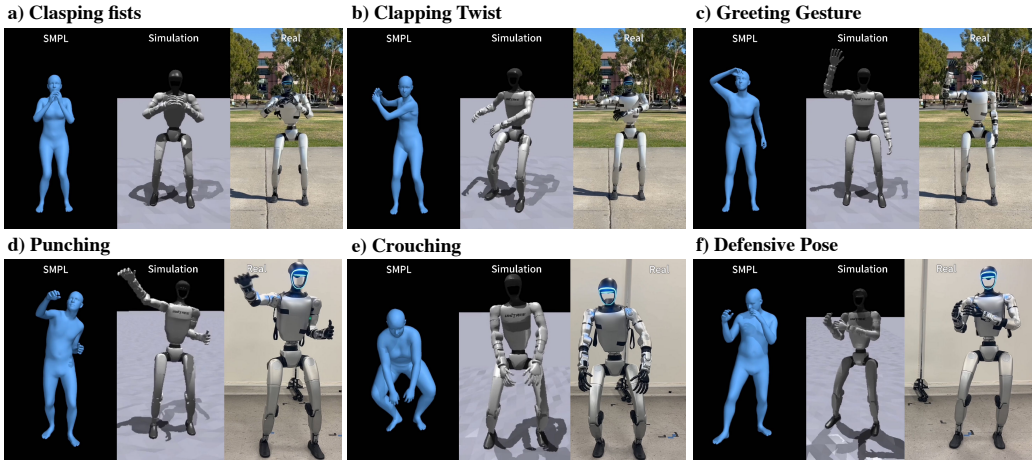

Figure 7: Sim-to-real experiment results showcasing diverse motions across SMPL, simulation, and real-world environments. Examples include: (a) Clasping Fists, (b) Clapping Twist, (c) Greeting Gesture, (d) Punching, (e) Crouching, and (f) Defensive Pose.

other sources—*Real-time Whole-body Mimic* (RGB-based) and *Motion Synthesis* (latent generative model)—that can drive ExBody2 for more interactive and long-horizon tasks. Figure 8 visually summarizes these capabilities alongside possible VR or IMU-based streams.

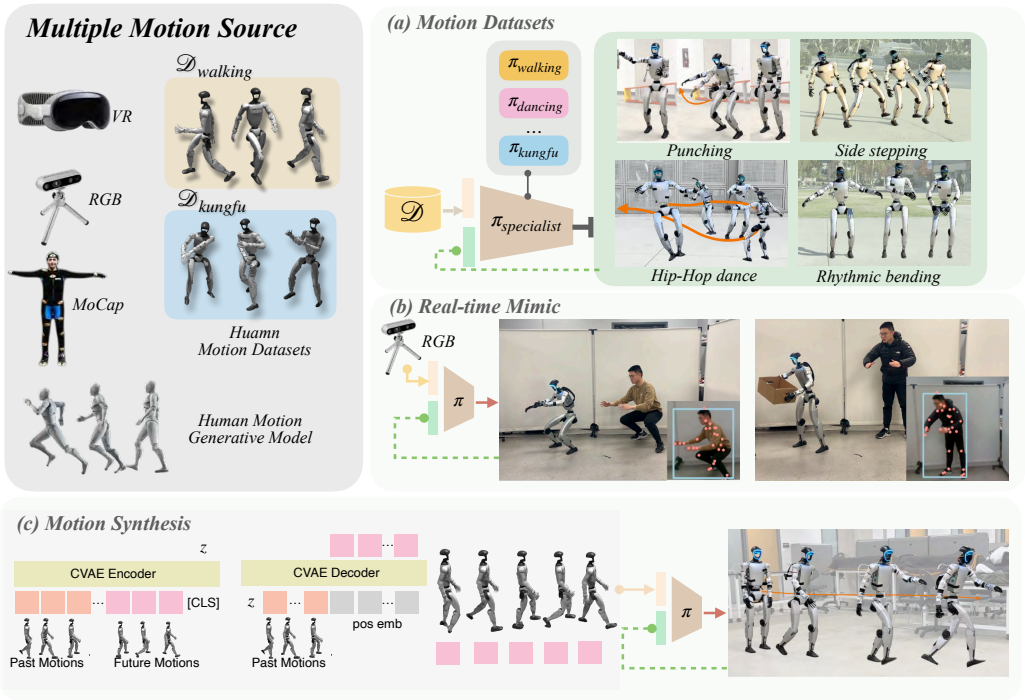

Figure 8: Illustration of ExBody2's multi-source application, demonstrating how VR, RGB, motion capture, and generative models can be combined to produce diverse humanoid behaviors. **(a) Motion Datasets:** specialized policies (e.g., kung fu, dancing) finetuned on specialist motion datasets. **(b) Real-time Whole-body Mimic:** real-time replication of human motions from monocular RGB via HybrIK. **(c) Motion Synthesis:** a CVAE-based approach for extended and varied motion generation. Experiments demonstrate ExBody2's capability to seamlessly integrate multiple motion sources in both simulation and real-world scenarios.

E.1 Real-time Whole-body Mimic (RGB Input)

We implement a real-time tracking pipeline that uses only *monocular RGB input* to mimic human movements. Our system first applies the HybrIK algorithm [63] to extract 3D human poses from each image frame. We then retarget this sequence of poses to the robot's kinematic structure and feed it into the ExBody2 whole-body policy. Because our policy is trained to be robust to partial or noisy signals, it can accommodate real-time streaming of 3D keypoints and still maintain stable lower-body tracking. Figure 8 (b) demonstrates a user controlling the robot to lift and carry an object, showcasing responsive teleoperation.

Relying on monocular pose estimates is more lightweight than requiring a full-body Mocap or multi-camera setup. Although the 3D pose can be less accurate than multi-view solutions, our control policy's robust design helps it remain stable even under potential keypoint noise.

E.2 Motion Synthesis for Extended Behaviors

We further incorporate a *Conditional Variational Autoencoder (CVAE)* to generate new motion segments based on a short sequence of past motions, as Figure 8 (c) illustrated. During inference, each latent code z is sampled (or set to the prior mean) to produce new motion trajectories that seamlessly continue from the current pose. Unlike naive random sampling, the CVAE ensures continuity by conditioning on past pose context and penalizing abrupt transitions with a smoothness loss.

Training details. The CVAE is trained on a broad set of humanoid motion clips, optimizing a reconstruction loss plus KL-divergence for the latent space. We also add a small penalty for high-frequency velocity changes, improving the realism of the generated motions.

Integration with ExBody2. The generated motion frames are retargeted in exactly the same way as a regular Mocap clip, so the policy sees no difference. This allows the robot to perform extended, varied sequences—e.g., spontaneously transitioning from walking to an upper-body gesture—without needing to rely on a fixed database of motion capture clips.