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

ExBody2: Advanced Expressive Humanoid Whole-Body Control

Mazeyu Ji*,1 Xuanbin Peng*,1 Fangchen Liu[2] Jialong Li[1]
Ge Yang[3] Xuxin Cheng†,1 Xiaolong Wang†,1
[1]UC San Diego [2]UC Berkeley [3]MIT
*Equal contribution †Equal advising
https://exbody2.github.io

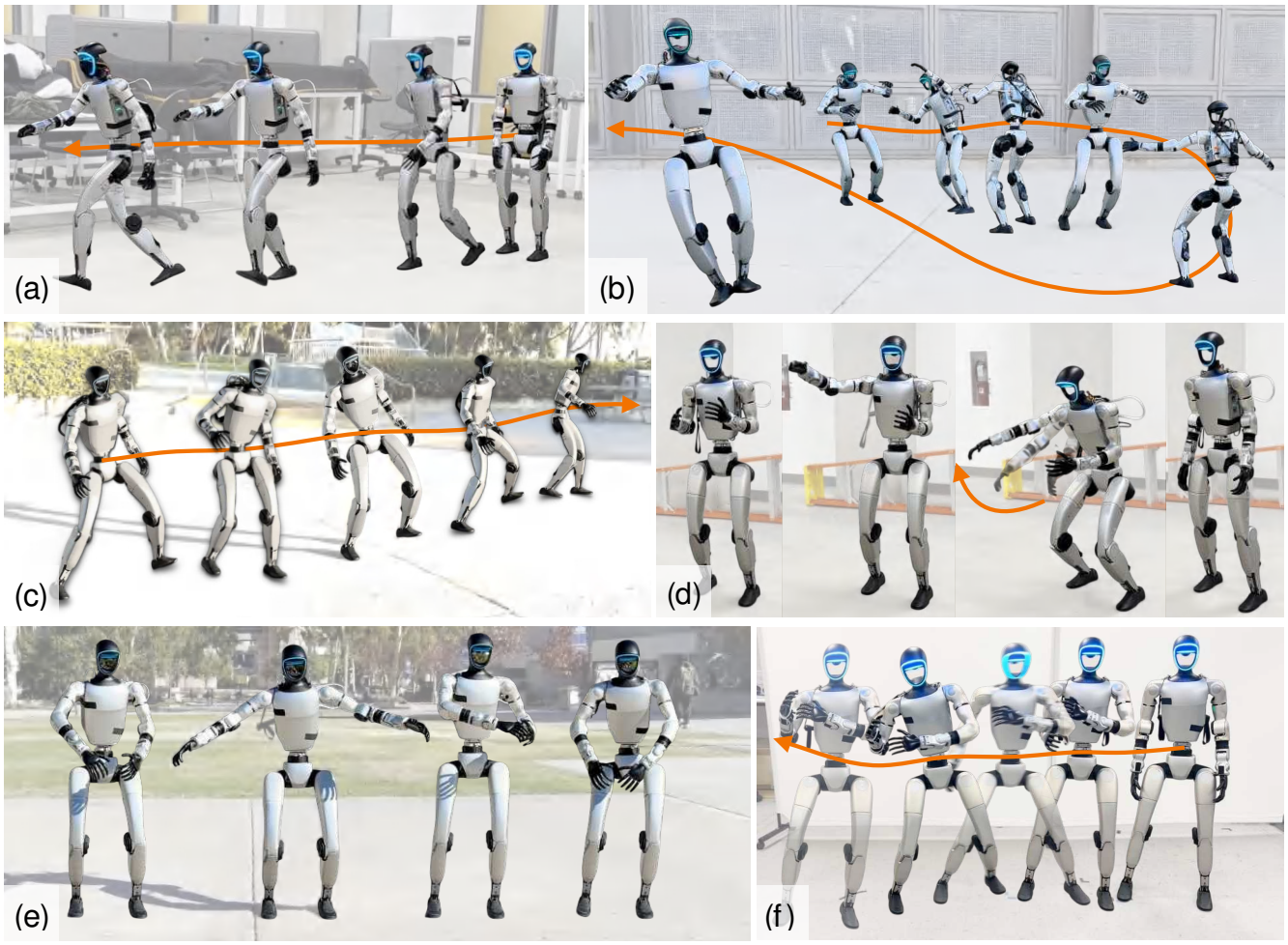

Fig. 1: Humanoid robot executing various expressive whole-body motions in the real world. The robot can (a) walk with a large stride from static standing, (b) dance along a long horizon choreography, (c) dynamic sidestep with fluid weight shifts, (d) punch with different height configurations, (e) express 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 identification of an empirical trade-off between 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 tasks to achieve superior tracking accuracy. Extensive experiments show that Exbody2 achieves consistently better performance than strong baselines and provides insights that may inform future work on whole-body humanoid control.**

979-8-3315-8160-2/26/$31.00 ©2026 IEEE

I. 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 can further improve 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. 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.

II. 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]–[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]–[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]–[36], portable devices [37]–[40] and learn from human videos with hand tracking or motion retargeting [41]–[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]–[53], transferring diverse motions to real robots [1], [3], [32], [54]–[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 handcrafted 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.

III. 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.

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

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.

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.

B. 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

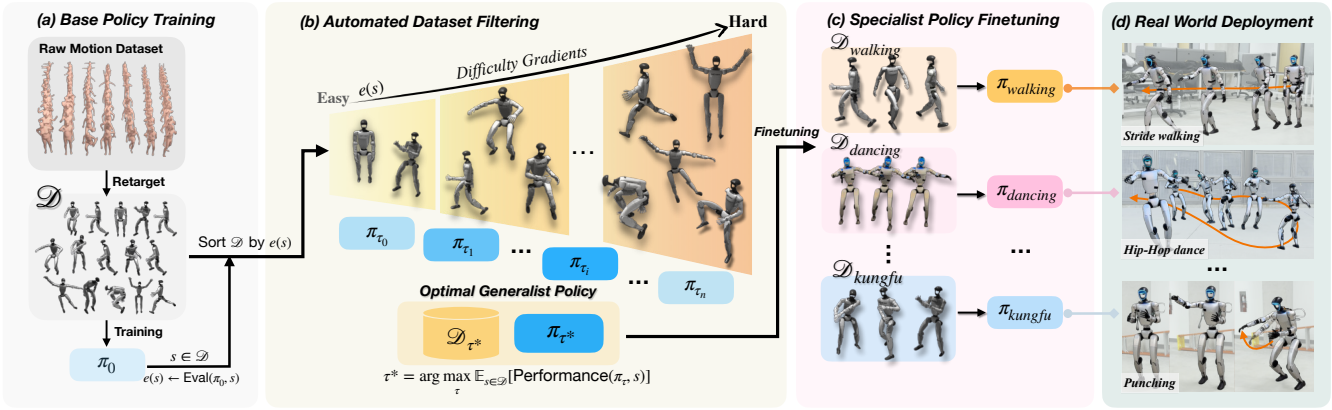

Fig. 2: Exbody2's framework. (a) Motion retargeting adapts raw human motion datasets to fit the humanoid robot's morphology, generating a diverse set of training samples. (b) Automated dataset filtering ranks motions based on tracking errors and selects an optimal subset to train a generalist policy, 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.

world, and distill the teacher policy to a deployable student policy.

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.

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

IV. 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 IV-B) Does Exbody2 generalist policy achieve higher tracking accuracy in both simulation and real-world deployment compared to prior methods?

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

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

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 I: 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 II: Comparisons with baselines on selected motions for Unitree G1 in real world.

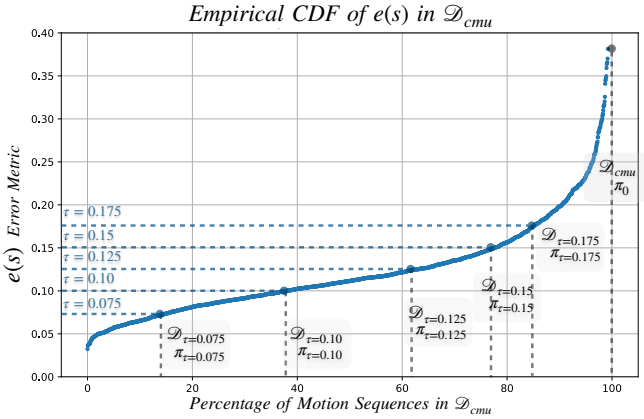

Fig. 3: Empirical CDF of the base policy's error metric $e(s)$ on the entire $\mathcal{D}_{\text{CMU}}$ dataset. The horizontal axis indicates the percentile of motion sequences from 0% (lowest error) to 100% (highest error), while the vertical axis shows $e(s)$. We overlay dashed horizontal lines at 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.

A. 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.

B. Generalist Policy Performance

As shown in Table I, *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 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 II), 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.

C. Impact of Automatic Data Curation

To experimentally examine the selection criteria for constructing an optimal human motion dataset that enhances generalist policy learning, we conducted the following experiment, reconstructing the full pipeline of our automated data curation method and evaluating its effect on policy performance.

1) **Base Policy Training**: We first trained a base policy on the unfiltered $\mathcal{D}_{CMU}$ dataset, which encompasses a broad range of human motions, including both stable and highly dynamic sequences, as well as infeasible movements that exceed the robot's physical limitations.
2) **Dataset Filtering**: The base policy's tracking performance was evaluated on $\mathcal{D}_{CMU}$, and each motion sequence was assigned a tracking error score. Based on these scores, we sorted the sequences and applied filtering thresholds to create progressively refined datasets:

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

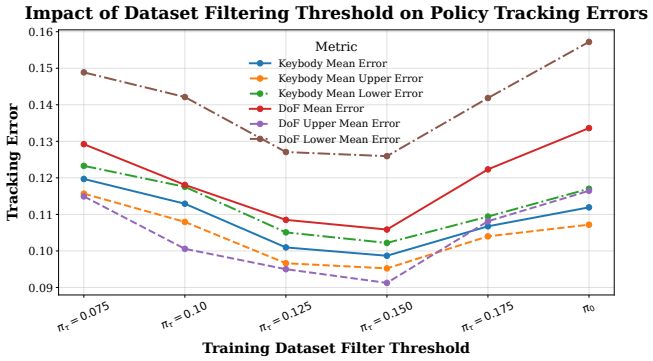

Fig. 4: Impact of dataset filtering thresholds on tracking errors. Policies trained with balanced thresholds (e.g., $\pi_{\tau=0.150}$) achieve the lowest errors, while unfiltered data and overly strict ($\pi_{\tau=0.075}$) or loose ($\pi_{\tau=0.175}$) thresholds degrade performance. We compute the 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.

$\mathcal{D}_{\tau=0.075}$, $\mathcal{D}_{\tau=0.1}$, $\mathcal{D}_{\tau=0.125}$, $\mathcal{D}_{\tau=0.15}$, $\mathcal{D}_{\tau=0.175}$. A lower threshold retains only motions with the lowest tracking errors, ensuring high stability but reducing diversity. Conversely, a higher threshold allows more challenging and diverse motions but introduces tracking instability.

These thresholds were determined based on the error distribution of the base policy, as shown in 3. The distribution-informed selection ensures that the filtering process is not arbitrary but rather grounded in empirical data. This selection strategy is not limited to a specific dataset; rather, it generalizes to other datasets with similar policy architectures and training conditions. When applied to new datasets, it effectively filters motions while maintaining a balance between feasibility and diversity.

3) **Filtered Policy Resume:** For each filtered dataset $\mathcal{D}_\tau$, we resume training from the base policy π_0 to obtain a new policy π_τ. Concretely, thresholds $\tau \in \{0.075, 0.1, 0.125, 0.15, 0.175\}$ each yield a distinct subset $\mathcal{D}_\tau$, and thus a correspondingly refined policy π_τ. This resumption process leverages the base model's learned prior, enabling faster adaptation to each subset's motion characteristics and improving overall training efficiency.
4) **Evaluation**: All the resulting policies were evaluated on the full $\mathcal{D}_{CMU}$ dataset, measuring tracking performance across multiple metrics. The results are visualized in Figure 4.

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.

D. 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 finetuning 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 III, 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 5). 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.

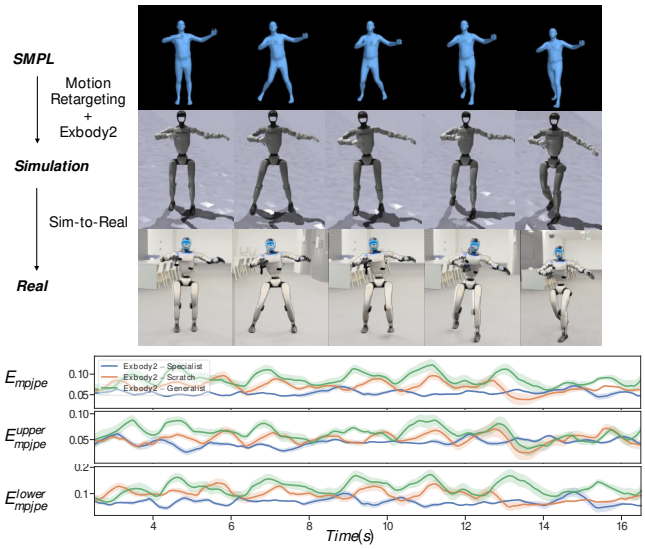

Fig. 5: A sequence of a robot performing the Cha-Cha dance. Top three rows: reference avatar, simulation result, real robot. Bottom three rows: per-frame errors for whole-body, upper-body, and lower-body DoF. Blue: *Exbody2-Specialist* (finetuned on $\mathcal{D}dancing$), orange: *Exbody2-Scratch* (trained from scratch), green: *Exbody2-Generalist* (trained on filtered $\mathcal{D}CMU$).

This strategy is especially effective in challenging and unseen scenarios, combining generalization with specialization.

V. 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.

VI. Limitations

Although our framework provides both generalist and specialist policies, it remains challenging to deploy them effectively for composite behaviors that require smooth transitions across multiple specialized skills. In real-world scenarios, humans naturally sequence diverse specialist capabilities to complete complex tasks. Enabling such policy-level composition, where a system can adaptively select and transition between specialists in real time, is a promising yet unresolved research direction.

Another open challenge lies in the definition of specialist groups. Currently, we rely on manually specified motion categories to train specialist policies. However, a more principled, data-driven approach to automatically discovering latent skill clusters—e.g., via unsupervised learning or hierarchical skill extraction—could lead to more scalable and efficient policy specialization.

VII. Acknowledgement

This project was supported, in part, by NSF CAREER Award IIS-2240014, NSF CCF-2112665 (TILOS), gifts from Amazon, Meta, and Qualcomm.

References

[1] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," 2024. [Online]. Available: 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. [Online]. Available: 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*. IEEE, 2008, pp. 2677–2684.

[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)*, vol. 1. IEEE, 2001, pp. 239–246.

[12] E. R. Westervelt, J. W. Grizzle, and D. E. Koditschek, "Hybrid zero dynamics of planar biped walkers," *IEEE transactions on automatic control*, vol. 48, no. 1, pp. 42–56, 2003.

[13] I. Kato, "Development of wabot 1," *Biomechanism*, vol. 2, pp. 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)*, vol. 2. IEEE, 1998, pp. 1321–1326.

[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)*. IEEE, 2021, pp. 1–8.

[16] A. Dallard, M. Benallegue, F. Kanehiro, and A. Kheddar, "Synchronized human-humanoid motion imitation," *IEEE Robotics and Automation Letters*, vol. 8, no. 7, pp. 4155–4162, 2023.

[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)*, 2019, pp. 679–686.

[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*, vol. 26, no. 4, pp. 73–82, 2019.

[19] J. Ramos and S. Kim, "Dynamic locomotion synchronization of bipedal robot and human operator via bilateral feedback teleoperation," *Science Robotics*, vol. 4, no. 35, p. eaav4282, 2019. [Online]. Available: 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)*. IEEE, 2021, pp. 2811–2817.

[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*, vol. 7, no. 65, p. 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)*. IEEE, 2023, pp. 1–8.

[32] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," 2024. [Online]. Available: 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. [Online]. Available: 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*, vol. 7, no. 4, pp. 10 873–10 881, 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. [Online]. Available: 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)*, vol. 40, no. 4, pp. 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.*, vol. 41, no. 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*, ser. SIGGRAPH '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: 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. [Online]. Available: 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. [Online]. Available: 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)*, vol. 39, no. 4, pp. 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.*, vol. 42, no. 4, jul 2023. [Online]. Available: 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*, 2024, pp. 1–11.

[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.

[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)*, vol. 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. [Online]. Available: 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*, vol. 5, no. 47, p. 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/, Pittsburgh, PA, USA, Mar 2007, [Online].