TWIST: Teleoperated Whole-Body Imitation System

Yanjie Ze*, Zixuan Chen*, João Pedro Araújo* , Zi-ang Cao
Xue Bin Peng, Jiajun Wu†, C. Karen Liu†

humanoid-teleop.github.io

Abstract: Teleoperating humanoid robots in a whole-body manner marks a fundamental step toward developing general-purpose robotic intelligence, with human motion providing an ideal interface for controlling all degrees of freedom. Yet, most current humanoid teleoperation systems fall short of enabling coordinated whole-body behavior, typically limiting themselves to isolated locomotion or manipulation tasks. We present the Teleoperated Whole-Body Imitation System (TWIST), a system for humanoid teleoperation through whole-body motion imitation. We first generate reference motion clips by retargeting human motion capture data to the humanoid robot. We then develop a robust, adaptive, and responsive whole-body controller using a combination of reinforcement learning and behavior cloning (RL+BC). Through systematic analysis, we demonstrate how incorporating privileged future motion frames and real-world motion capture (MoCap) data improves tracking accuracy. TWIST enables real-world humanoid robots to achieve unprecedented, versatile, and coordinated whole-body motor skills—spanning whole-body manipulation, legged manipulation, locomotion, and expressive movement—using a *single* unified neural network controller.

Keywords: Humanoids, Whole-Body Teleoperation, Learning-Based Control

1 Introduction

Humans naturally master versatile, coordinated whole-body skills essential for everyday tasks. For example, when entering a room, a person with objects in both hands can use their foot to open the door. When cleaning, they can bend down to reach under the bed and tidy up the hidden lower space. In sports like soccer, humans can balance on one leg while using the other to kick a ball. Enabling humanoid robots to perform such coordinated, whole-body actions is crucial for developing general-purpose robots that can live and work alongside humans in household environments.

One promising way to empower humanoid robots with versatile whole-body dexterity is by imitating human movements. However, due to the embodiment gap between humanoids and humans, simply imitating offline human motion data is insufficient for humanoid visuomotor control; we must build a whole-body teleoperation system to acquire humanoid observation-action data. Building a capable whole-body teleoperation system, however, has long been a challenge [1, 2, 3, 4, 5, 6, 7, 8, 9]. Classical methods [10, 1, 2, 3, 4, 5] use modular model-based controllers to separately handle teleoperation and balance, limiting the system's whole-body capabilities and robustness. Recent learning-based controllers have shown great promise [6, 7, 8, 9], but the coordinated whole-body skills, *e.g.*, crouching down to lift a box on the ground, remain limited, mainly due to the lack of accurate, real-time whole-body tracking targets and robust controllers for tracking such diverse, real-time motions.

In this work, we propose the **Teleoperated Whole-Body Imitation System (TWIST)**, a humanoid teleoperation system that achieves versatile, coordinated whole-body skills by imitating whole-body human motions in real time. A key requirement for such a system is a controller that autonomously converts arbitrary, real-time human movements into *balanced* robot motions, while closely matching the human's behaviors. To this end, we formulate whole-body teleoperation as a real-time motion re-

9th Conference on Robot Learning (CoRL 2025), Seoul, Korea.

Figure 1: **Our Teleoperated Whole-Body Imitation System (TWIST) presents versatile, coordinated, and human-like whole-body skills on real-world humanoid robots.** Our robot can perform whole-body manipulation (*e.g.*, lifting boxes from the ground), legged manipulation (*e.g.*, kicking the football), locomotion (*e.g.*, ducking under the obstacle and walking backwards), and expressive motions (*e.g.*, Waltz dance).

targeting and tracking problem. We first derive tracking targets—humanoid joint positions and root velocities—by retargeting arbitrary human motions captured by motion capture (MoCap) devices. We then train a *single* policy with reinforcement learning (RL) in large-scale simulated environments, combined with human motion data. The resulting controller can robustly and accurately track target robot joint positions and root velocities at each timestep while maintaining whole-body balance. To address the practical challenges of real-time, whole-body motion tracking and teleoperation, our training pipeline introduces several critical techniques, including:

- To ensure low latency of teleoperation, RL policies can only observe the reference motion for one current time step, which results in more conservative and hesitant behaviors compared to policies including future reference motions as observations. This hesitant behavior is aggravated by the real-time teleoperation system because the human demonstrator tends to compensate their own movements when experiencing the hesitant behaviors on the humanoid, leading to ineffective teleoperation control. To alleviate this issue, we propose a two-stage teacher-student framework: the teacher policy is trained with privileged access to future motion frames to learn smoother behaviors, and subsequently guides the student policy, which tracks only a single frame.
- Offline human motion datasets are usually high-quality and smooth, while the real-time human motions and the real-time retargeting are not that stable and smooth, causing a distribution shift for online teleoperation. Therefore, we collect a small-scale MoCap human dataset (150 clips) using the online MoCap and retargeting settings, combined with 15K offline motion clips as the training set for training the RL controller. Surprisingly, despite only a small set of online motions we use, the controller performs significantly better and more stably on unseen test motions and in real-world teleoperation.
- During offline retargeting of human motions, we can ensure high-quality motion data through many iterations of optimization. However, for online retargeting during teleoperation, fast inference is critical, often at the expense of smoothness. We find that jointly optimizing 3D joint positions and orientations helps mitigate this offline-to-online gap, compared to optimizing orientations alone [11].
- As the learning objective of the controller is simply motion tracking, tasks requiring force exertion (*e.g.*, lifting a box) rather than reaching target positions represent out-of-distribution scenarios, causing the controller to produce jittery behaviors occasionally. To enable the controller to learn to apply force, we propose to train controllers with large end-effector perturbations, which significantly improves robustness in tasks requiring contact and force.

With all these critical components integrated, TWIST achieves remarkable whole-body teleoperation capabilities on real-world humanoid robots. As shown in Figure 1, TWIST enables the Unitree G1—a medium-sized humanoid robot with 29 degrees of freedom (DoF)—to perform a wide range of diverse, human-like skills. All teleoperation tasks are accomplished using a *single* neural network controller. **We will release the code.**

2 Related Works

Learning-Based Whole-Body Control for Humanoid Robots. In recent years, learning-based methods, particularly sim-to-real reinforcement learning, have made notable progress in developing whole-body controllers for humanoid robots, enabling a wide range of skills such as walking [12, 13, 14, 15, 16], jumping [17, 7], parkour [18], dancing [19, 20], hopping [21], and fall recovery [22, 23, 24]. However, most of these works focus on developing controllers for a single specific task, limiting the generality of their methods. In contrast, our work aims to train a controller capable of performing a diverse set of real-world tasks with real-time human teleoperation.

Teleoperation Systems for Humanoid Robots. Teleoperation is key to enabling humanoid robots to interact with complex real-world environments and perform manipulation tasks. Prior works have explored teleoperation modalities such as sparse VR keypoints [25, 26, 27, 8], exoskeletons [28, 29, 30], motion capture suits [2, 31, 5, 10], and camera-based pose estimation [7, 32].

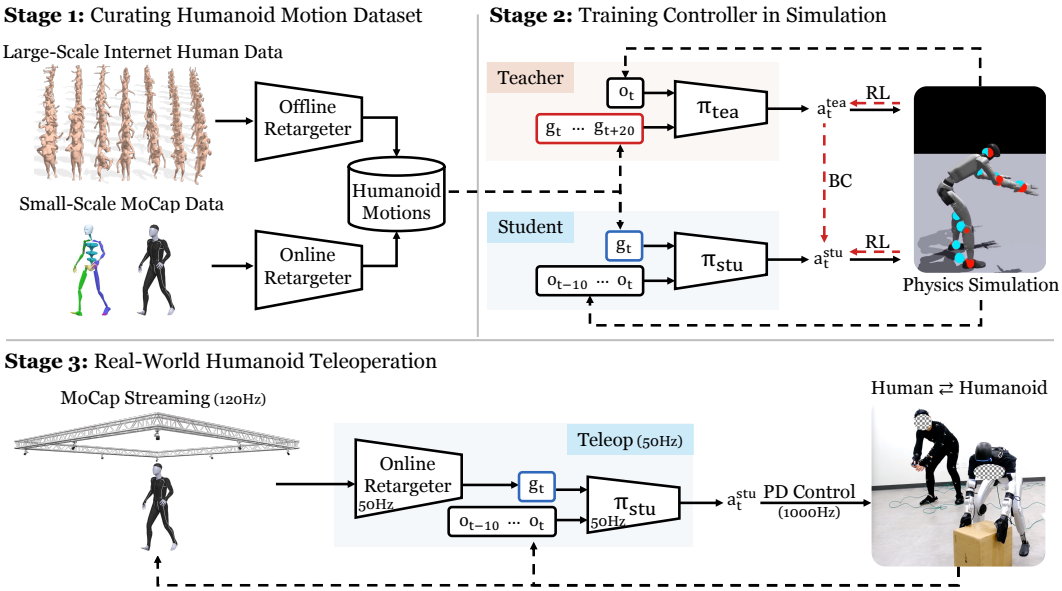

Figure 2: **The Teleoperated Whole-Body Imitation System (TWIST) consists of 3 stages**: 1) curating a humanoid motion dataset by retargeting Internet human data and our MoCap data, 2) training a single whole-body controller in simulation, 3) teleoperating real-world humanoid robots with MoCap devices.

However, most systems show limited coordinated whole-body skills, making them insufficient for general household tasks. For example, Mobile-TV [27] and HOMIE [28] decouple upper- and lower-body control, guiding movement with external commands like joysticks or foot pedals, but limiting whole-body tasks like kicking or obstacle traversal. More closely related to our work, HumanPlus [7] and OmniH2O [8] attempt full-body teleoperation by training low-level controllers to track human motions. HumanPlus relies on camera-based pose estimation [33], which struggles with root position accuracy, affecting locomotion fidelity. OmniH2O uses VR keypoints that only captures upper-body movements, still lacking full whole-body control. In contrast, we introduce a whole-body teleoperation system capable of performing coordinated skills, leveraging motion capture for high-quality human data input and a robust training pipeline.

3 Method

We frame the problem of whole-body humanoid teleoperation as a real-time motion tracking task [7, 8, 32, 4, 34, 35], where a whole-body controller receives real-time retargeted human motions and outputs target joint positions to make the humanoid mimic the human. This setup raises two key questions: first, how to obtain reliable real-time reference motions, and second, how to build a robust controller that can accurately track them.

TWIST answers these questions through two innovations. First, we collect accurate real-time whole-body motion from a motion capture (MoCap) device; second, we develop a robust, *unified* neural network controller capable of tracking diverse real-time motions through a teacher-student framework, trained with large-scale simulation and human motions. The resulting system, the Teleoperated Whole-Body Imitation System (TWIST), is shown in Figure 2. In the following sections, we first describe how to curate human motion data for training the controller (Section 3.1). We then explain how to train a unified controller using this dataset (Section 3.2). Finally, we show how to deploy the controller for real-world teleoperation of the humanoid robot (Section 3.3).

3.1 Curating Humanoid Motion Datasets

To train a controller that presents human-like movements and accurately tracks human motions, we first curate a training dataset to serve as the fuel [34]. The majority of our data comes from publicly available MoCap datasets—AMASS [36] and OMOMO [37]—which together include over

Table 1: **Reward terms and their weights.** The left table lists tracking rewards, while the middle table lists penalty terms.

Tracking Reward Terms	Weights
KeyBody Position Tracking	2.0
Joint Position Tracking	0.6
Joint Velocity Tracking	0.2
Root Pose Tracking	0.6
Root Velocity Tracking	1.0

Penalty Terms	Weights
Feet Contact Penalty	-5e-4
Feet Slipping Penalty	-0.1
Joint Velocities Penalty	-1e-4
Action Rate Penalty	-0.01
Feet Air Time	5.0

Table 2: **Domain randomization parameters.**

Domain Rand Params	Range
Base Mass (kg)	$[-3, 3]$
Friction	$[0.1, 2.0]$
Motor Strength	$[0.8, 1.2]$
Gravity Change (m/s^2)	$[-0.1, 0.1]$
Push Robot Base (m/s)	$[-0.1, 0.1]$
Push End-Effector (N)	$[0, 20]$

15,000 clips (around 42 hours), with unfeasible motions such as climbing stairs being filtered out. Additionally, we collect a small in-house dataset using our own MoCap system, consisting of 150 clips (around 0.5 hours). This in-house data better reflects the conditions of real-world teleoperation, such as noise and imperfect calibration. Notably, we do not design these motions to match the teleoperation task—they are collected randomly.

Due to the embodiment gap between humanoids and humans, tracking raw human motions directly is non-trivial. Instead, we retarget them into the humanoid format to more easily compute tracking errors. For large-scale public datasets, we apply an offline retargeter similar to PHC [35], which optimizes key body positions. We additionally optimize temporal smoothness. While offline retargeting yields higher-quality motion, it involves iterative optimization, making it impractical for real-time teleoperation.

To simulate the real-time teleoperation setting, we use an online retargeter [11] on our small in-house dataset, which efficiently adjusts body orientation and foot placement using an Inverse Kinematics (IK) method [38]. Although this approach is significantly faster and suited for real-time use, it typically results in less smooth motion, potentially impacting teleoperation performance. We mitigate this quality gap by enhancing the optimization objective of the online retargeter to jointly optimize 3D joint positions and orientations. This improvement reduces the discrepancy between offline and online motion quality, as shown in Figure 7 (left).

3.2 Training A Whole-Body Controller in Simulation

After curating a humanoid motion dataset, we aim to train a unified whole-body controller capable of tracking arbitrary retargeted humanoid motions through large-scale simulation [39]. Naive approaches [7] typically use single-stage reinforcement learning (RL) to train a deployable tracking policy that tracks only the current motion frame. However, this often leads to artifacts such as foot sliding and fails to produce smooth control suitable for teleoperation, largely due to the lack of access to full motion sequences. In contrast, our system adopts a two-stage approach: first, a privileged expert (teacher) policy is trained via RL with access to future reference motions. Then, a deployable student policy is jointly optimized using RL and behavior cloning (BC), relying only on proprioception and a single reference frame at each time step.

Privileged Teacher Policy. The teacher policy π_{tea} takes a sequence of future reference motion frames (spanning 2 seconds) as part of the input, enabling the teacher to anticipate and plan for upcoming tracking goals, further leading to smooth locomotion gaits. Besides, we choose to track joint positions and root velocities expressed in the robot's local frame rather than the world frame, to 1) alleviate the accumulated tracking error and imperfect retargeting [20, 9]; 2) ensure better consistency with the real-world teleoperation setup. The teacher policy is optimized by PPO [40, 41], with a reward structure r_{tea} that emphasizes accurate tracking while penalizing artifacts such as jitter and foot slippage (see Table 4): $r_{\text{tea}} = r_{\text{track}} + r_{\text{penalty}}$.

Deployable Student Policy. Since privileged information (*e.g.*, future motion frames) is unavailable during deployment, the expert policy π_{tea} must be distilled into a deployable student policy π_{stu} using only proprioceptive inputs and immediate reference motion targets. The observational gap created by the differing input modalities poses a challenge that standard imitation methods (*e.g.*, DAgger [42]) cannot fully address. Therefore, we adopt a hybrid RL and BC approach [15], optimizing the student policy with the following loss:

$$L(\pi_{\text{stu}}) = L_{\text{RL}}(\pi_{\text{stu}}) + \lambda D_{\text{KL}}(\pi_{\text{stu}} \parallel \pi_{\text{tea}}), \tag{1}$$

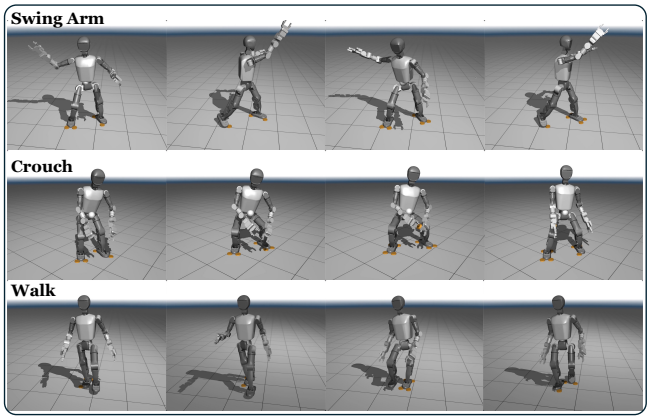

Figure 3: **Booster T1 sim2sim results.** The whole-body controller is trained in IsaacGym [39] and evaluated in MuJoCo [44]. The tracking goals are sampled from training data.

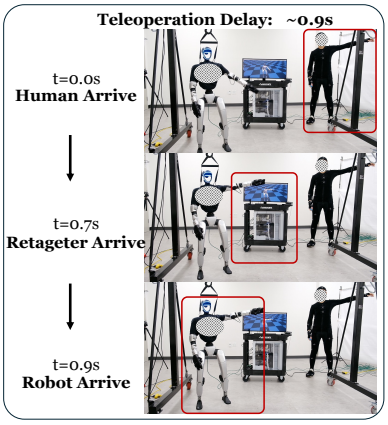

Figure 4: **Teleoperation delay** is roughly measured by the video, around 0.9 seconds.

where L_{RL} denotes the PPO loss using the same reward r_{tea} as the teacher, D_{KL} denotes KL divergence for encouraging imitation of the expert, and the weight λ is gradually reduced during training. Our results demonstrate that this RL+BC strategy significantly outperforms pure BC and pure RL, resulting in smoother motions and better generalization.

3.3 Real-World Humanoid Teleoperation

After obtaining the whole-body controller π_{stu} though our two-stage training pipeline, our policy can be zero-shot deployed to the real robot, due to the careful tuning of domain randomization (see Table 2). To achieve precise real-time motion tracking in the real world, we establish an online streaming and retargeting pipeline, which captures high-quality human motions at 120Hz using OptiTrack [43] and retargets into humanoid motions at 50Hz using the retargeting approach detailed in Section 3.1. The high-quality tracking goals in the real world ensure that our robot can perform diverse tasks across manipulation and locomotion in a unified manner. Our policy π_{stu} then takes the retargeted motions as input and outputs joint position commands at 50Hz on an Nvidia RTX 4090 GPU, sending to the robot's PD controller running at 1000Hz.

4 Experiments

In this section, we demonstrate the teleoperation capabilities of the Teleoperated Whole-Body Imitation System (TWIST) on Unitree G1 [45], and analyze key factors affecting TWIST's performance. We also show that TWIST can transfer to other humanoid robots such as Booster T1 [46]. Additionally, we analyze TWIST's reachability, teleoperation delay, and failure cases.

4.1 Main Results on Whole-Body Humanoid Teleoperation

We showcase the teleoperation capabilities of TWIST on Unitree G1, a 29 DoF 1.3m humanoid robot. As shown in Figure 1 and the supplementary videos, our robot can be teleoperated to perform diverse human-like whole-body skills, including: **whole-body manipulation**, such as uprighting a fallen trash can, crouching to lift a box from the ground, and carrying a "Minions" toy from a table to a human; **legged manipulation**, such as kicking a door shut/open while carrying objects with hands, delivering a powerful kick to a soccer ball, and transporting the box with its feet; **locomotion**, such as executing sidesteps and backward walking, and crouching to navigate under obstacles; **expressive motion**, such as boxing and performing Waltz dance steps with a human partner.

To demonstrate that TWIST functions as a general framework for diverse embodiments, we further evaluate it on Booster T1 [46]. Figure 3 displays the sim-to-sim evaluation results. The controller successfully tracks diverse motions, including arm swinging with coordinated whole-body joints, deeply crouching down, and walking.

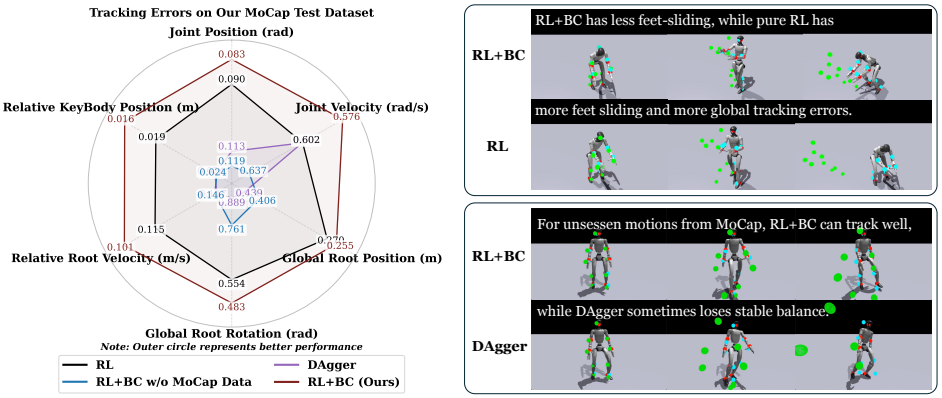

Figure 5: (left) **Tracking errors of different controllers,** measured on our MoCap test data. (right) **Controller behaviors.** Our RL+BC controller produces smooth and robust behaviors.

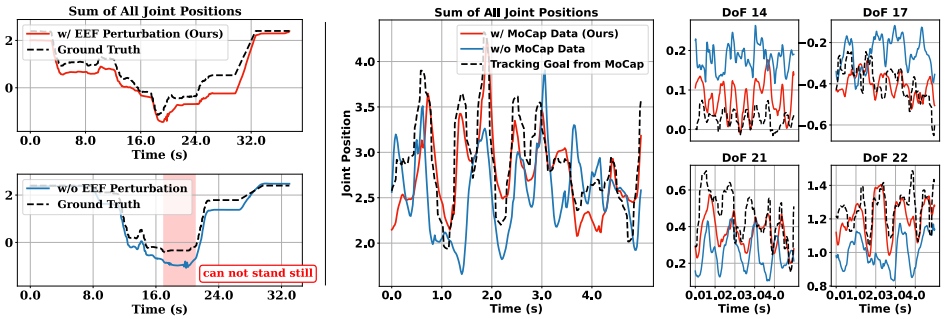

Figure 6: (left) **Rollout curves** in the real world when the robot holds a box. (right) **Rollout curves** in MuJuCo from different controllers when tracking MoCap data.

4.2 Ablation Experiments

To further evaluate the impact of different components of TWIST on its effectiveness in teleoperation, we collect another small-scale MoCap motion dataset (50 clips) for evaluation, all of which are not used in training. The following experimental results are evaluated on this test dataset.

Key Finding 1: RL+BC >> RL >> BC (DAgger). Unlike HumanPlus [7], which uses single-stage RL, and OmniH2O [8], which employs DAgger, we find that our combined pipeline, RL+BC, achieves superior tracking accuracy (Figure 5 (left)) and motion smoothness (Figure 5 (right)). Pure RL approaches frequently exhibit feet sliding artifacts due to their inability to anticipate future motion goals. Meanwhile, DAgger can not stably and robustly track unseen motions occasionally, due to the lack of task reward guidance like RL. In summary, while RL demonstrates better generalization than BC [47], the combination of both approaches yields significant performance improvements.

Key Finding 2: In-House MoCap Data Matters. We find that adding even a small set of in-house MoCap sequences—retargeted online to mimic real teleoperation—substantially reduces tracking errors on unseen motions (Figure 5 (left) and Figure 6 (right)). This gain arises from two factors: (1) our in-house captures are inherently noisier and less stable, suffering from calibration drift and occlusions; and (2) our online retargeter yields less-smooth reference motions compared to the offline version. Exposing the controller to these real-world imperfections bridges the gap between clean public datasets and the variability encountered during teleoperation, improving generalization.

Key Finding 3: Learning to Apply Force. We find that training policies without end-effector perturbations leads to drift and instability during stationary poses (Figure 6 (left)). Incorporating perturbations during training significantly improves stability, especially crucial for contact-rich tasks.

Key Finding 4: Better Online Retargeter, Better Tracking. As detailed in Section 3.1, improving our online retargeter by simultaneously optimizing 3D positions and orientations yields smoother humanoid reference motions. Figure 7 (left) shows that such smoother motions benefit controllers trained both with and without MoCap data to reduce overall tracking errors.

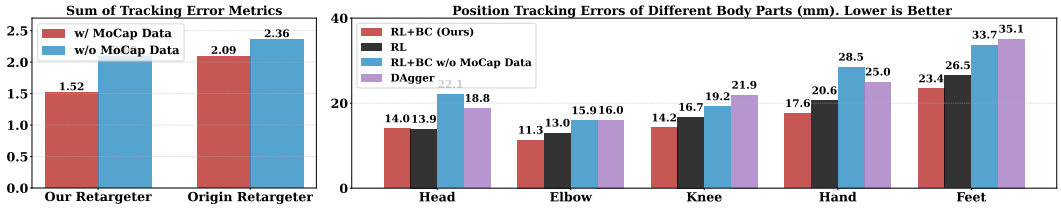

Figure 7: (left) **Sum of tracking error metrics** on different online retargeters. (right) **Tracking errors across different body parts.** The feet exhibit the largest errors, highlighting the greater difficulty in accurately tracking lower-body movements.

4.3 System Analyses

Tracking Error Distribution Across Body Parts. We are curious about how the controller's performance varies in tracking different body parts. As depicted in Figure 7 (right), we observe: (1) the end-effectors—hands and feet—exhibit the largest tracking errors, consistent with their positions at the extremities of the humanoid's kinematic tree; and (2) lower-body parts, such as feet and knees, generally incur higher tracking errors compared to upper-body parts like elbows and hands, confirming our intuition that tracking the lower body is inherently more challenging due to more complex contact dynamics.

Teleoperation Delay. The total teleoperation delay of our system is approximately 0.9 seconds, as measured in Figure 4. The major overhead arises from generating tracking goals (0.7 seconds), while policy inference remains efficient (0.2 seconds). Reducing this latency further will be a key focus in future improvements.

Reachability. By utilizing whole-body DoFs, our teleoperation system significantly enhances reachability compared to prior work [32, 7, 8]. As illustrated in Figure 8 (a), the robot can nearly reach its toes with its hands, demonstrating exceptional whole-body reachability.

Failure Cases. Most failures arise from hardware unreliability. In particular, our robot's motors tend to overheat after 5-10 minutes of continuous operation, especially during tasks that require crouching (Figure 8 (b)), which necessitates cooling periods between tests. Nevertheless, our policy exhibits robust balance control even under extreme testing conditions.

Figure 8: **(a) Extreme reachability** by TWIST. **(b) Failures** caused by motor overheating.

5 Conclusion

In this work, we introduce the Teleoperated Whole-Body Imitation System (TWIST), a system that teleoperates real-world humanoid robots using whole-body human motion data captured from MoCap devices. We provide systematic analysis of our controller training pipeline and overall system to identify critical factors for TWIST. Our system fully leverages the robot's whole-body joints, enabling the execution of versatile and coordinated whole-body skills previously unachieved in the literature. The system demonstrates exceptional reachability compared to prior work, significantly expanding the capabilities of humanoid robots. Our future work will include studying how to learn visuomotor policies with combined TWIST-collected teleoperation data and egocentric human data.

6 Limitations

While TWIST demonstrates strong capabilities in teleoperating humanoid robots, there remain several limitations and directions for future improvement:

Lack of Robotic Feedback. Currently, there is no robotic egocentric vision streamed back to the human operator. As a result, when visual occlusion occurs, it becomes challenging to teleoperate effectively. In addition, there is no tactile feedback to inform the operator when grasping actions succeed, limiting the naturalness and reliability of manipulation.

Hardware Reliability. As discussed in Section 4.3, the current generation of humanoid hardware cannot sustain long-term continuous operation. Future improvements in robot hardware are likely to significantly enhance the overall system capability and robustness.

Dependence on MoCap Systems. Our method relies on a motion capture system, which is not portable and difficult to democratize, despite showcasing the potential of using whole-body human motion data. In the future, we plan to explore RGB-based pose estimation techniques to bridge the gap and approximate the quality of MoCap data more accessibly.

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] 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 & Automation Magazine*, 26(4):73–82, 2019.

[2] S. Dafarra, U. Pattacini, G. Romualdi, L. Rapetti, R. Grieco, K. Darvish, G. Milani, E. Valli, I. Sorrentino, P. M. Viceconte, et al. icub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots. *Science Robotics*, 9(86):eadh3834, 2024.

[3] L. Penco, B. Clément, V. Modugno, E. M. Hoffman, G. Nava, D. Pucci, N. G. Tsagarakis, J.-B. Mouret, and S. Ivaldi. Robust real-time whole-body motion retargeting from human to humanoid. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pages 425–432. IEEE, 2018.

[4] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci. Teleoperation of humanoid robots: A survey. *IEEE Transactions on Robotics*, 39(3):1706–1727, 2023.

[5] 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. IEEE, 2019.

[6] Z. He, K. Lei, Y. Ze, K. Sreenath, Z. Li, and H. Xu. Learning visual quadrupedal loco-manipulation from demonstrations. *arXiv preprint arXiv:2403.20328*, 2024.

[7] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. Humanplus: Humanoid shadowing and imitation from humans. In *Conference on Robot Learning (CoRL)*, 2024.

[8] 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. *arXiv preprint arXiv:2406.08858*, 2024.

[9] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang, et al. Hover: Versatile neural whole-body controller for humanoid robots. *arXiv preprint arXiv:2410.21229*, 2024.

[10] J. Koenemann, F. Burget, and M. Bennewitz. Real-time imitation of human whole-body motions by humanoids. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2806–2812. IEEE, 2014.

[11] A. Authors. Mind the mapping: How motion retargeting choices impact sim2real success. Anonymous submission, 2025. Under review.

[12] Z. Chen, X. He, Y.-J. Wang, Q. Liao, Y. Ze, Z. Li, S. S. Sastry, J. Wu, K. Sreenath, S. Gupta, et al. Learning smooth humanoid locomotion through lipschitz-constrained policies. *arXiv preprint arXiv:2410.11825*, 2024.

[13] X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, and J. Chen. Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning. *arXiv preprint arXiv:2408.14472*, 2024.

[14] I. Radosavovic, S. Kamat, T. Darrell, and J. Malik. Learning humanoid locomotion over challenging terrain. *arXiv preprint arXiv:2410.03654*, 2024.

[15] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024.

[16] J. Long, W. Yu, Q. Li, Z. Wang, D. Lin, and J. Pang. Learning h-infinity locomotion control. In *ArXiv*, 2024.

[17] C. Zhang, W. Xiao, T. He, and G. Shi. Wococo: Learning whole-body humanoid control with sequential contacts, 2024.

[18] Z. Zhuang, S. Yao, and H. Zhao. Humanoid parkour learning. *arXiv preprint arXiv:2406.10759*, 2024.

[19] 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.

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

[21] Y. Xue, W. Dong, M. Liu, W. Zhang, and J. Pang. A unified and general humanoid whole-body controller for fine-grained locomotion. *arXiv preprint arXiv:2502.03206*, 2025.

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

[23] Z. Zhuang and H. Zhao. Embrace collisions: Humanoid shadowing for deployable contact-agnostics motions. *arXiv preprint arXiv:2502.01465*, 2025.

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

[25] Y. Ze, Z. Chen, W. Wang, T. Chen, X. He, Y. Yuan, X. B. Peng, and J. Wu. Generalizable humanoid manipulation with 3d diffusion policies. *arXiv preprint arXiv:2410.10803*, 2024.

[26] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang. Open-television: Teleoperation with immersive active visual feedback. *arXiv preprint arXiv:2407.01512*, 2024.

[27] C. Lu, X. Cheng, J. Li, S. Yang, M. Ji, C. Yuan, G. Yang, S. Yi, and X. Wang. Mobile-television: Predictive motion priors for humanoid whole-body control. *arXiv preprint arXiv:2412.07773*, 2024.

[28] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang. Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[29] S. Yang, M. Liu, Y. Qin, D. Runyu, L. Jialong, X. Cheng, R. Yang, S. Yi, and X. Wang. Ace: A cross-platfrom visual-exoskeletons for low-cost dexterous teleoperation. *arXiv preprint arXiv:240*, 2024.

[30] Y. Jiang, R. Zhang, J. Wong, C. Wang, Y. Ze, H. Yin, C. Gokmen, S. Song, J. Wu, and L. Fei-Fei. Behavior robot suite: Streamlining real-world whole-body manipulation for everyday household activities. *arXiv preprint arXiv:2503.05652*, 2025.

[31] A. D. Dragan, K. C. Lee, and S. S. Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

[32] 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.

[33] S. Shin, J. Kim, E. Halilaj, and M. J. Black. Wham: Reconstructing world-grounded humans with accurate 3d motion. In *Computer Vision and Pattern Recognition (CVPR)*, 2024.

[34] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018.

[35] Z. Luo, J. Cao, K. Kitani, W. Xu, et al. Perpetual humanoid control for real-time simulated avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10895–10904, 2023.

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

[37] J. Li, J. Wu, and C. K. Liu. Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)*, 42(6):1–11, 2023.

[38] K. Zakka. Mink: Python inverse kinematics based on MuJoCo, July 2024. URL https://github.com/kevinzakka/mink.

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

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

[41] N. Rudin, D. Hoeller, P. Reist, and M. Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on Robot Learning*, pages 91–100. PMLR, 2022.

[42] 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*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

[43] NaturalPoint, Inc. Optitrack - motion capture systems, 2025. URL https://www.optitrack.com/. Accessed: 2025-04-22.

[44] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE, 2012.

[45] Unitree Robotics. Unitree G1 Humanoid Robot. https://www.unitree.com/cn/g1/, 2024. Accessed: 2025-04-26.

[46] Booster Robotics. Booster t1 humanoid robot. https://www.boosterobotics.com/, 2025.

[47] T. Chu, Y. Zhai, J. Yang, S. Tong, S. Xie, D. Schuurmans, Q. V. Le, S. Levine, and Y. Ma. Sft memorizes, rl generalizes: A comparative study of foundation model post-training. *arXiv preprint arXiv:2501.17161*, 2025.

A Implementation Details

Policy Architecture. We adopt the same network structure for the student policy and the teacher policy. The policies and the critics are all MLPs, with hidden dimensions $[512, 512, 256, 128]$.

Goal Observations. The goal observation of the teacher policy consists of: 1) the root height, 2) the root rotation, 3) the root velocity, 4) the root yaw angle's angular velocity, 5) the joint positions, 6) the key body positions. We then stack 20 frames of future reference motions that span 2 seconds as the goal observations of the teacher policy. The goal observation of the student policy differs in two parts: 1) not using the key body positions to save time to compute forward kinematics, and 2) not using the future reference motions but only using the single frame.

Proprio Observations. The proprioception of the teacher policy consists of: 1) the root angular velocity, 2) the root rotation, 3) the joint positions, 4) the joint velocities, and 5) the last action. The teacher only takes 1 frame of the proprioception, while the student policy takes 10 history frames plus 1 current frame.

Policy Training. We use Adam optimizer with the learning rate 1e-4. We train the teacher policy with 100k iterations and further train the student policy with 200k iterations.

Control Parameters. The control parameters used in simulation and the real world are given in Table 3. For the lower bodies, we adopt large stiffnesses for obtaining large torque.

Table 3: Joint stiffness and damping coefficients for Unitree G1.

Joint	**Stiffness (N·m/rad)**	**Damping (N·m·s/rad)**
Hip Yaw	100	2
Hip Roll	100	2
Hip Pitch	100	2
Knee	150	4
Ankle	40	2
Waist	150	4
Shoulder	40	5
Elbow	40	5

Table 4: **Reward terms expressions.**

Tracking Reward Terms	**Weights**
target keybody positions	$\exp(-\|\mathbf{p}_{\text{keybody}} - \mathbf{p}^{\text{tg}}_{\text{keybody}}\|^2)$
target joint positions	$\exp(-\|\mathbf{q} - \mathbf{q}^{\text{tg}}\|^2)$
target joint velocities	$\exp(-\|\dot{\mathbf{q}} - \dot{\mathbf{q}}^{\text{tg}}\|^2)$
target root pose	$\exp(-\|[r, p] - [r^{\text{tg}}, p^{\text{tg}}]\|^2 - \|h - h^{\text{gt}}\|^2)$
target root velocity	$\exp(-\|\mathbf{v} - \mathbf{v}^{\text{gt}}\|^2)$
feet contact	$-(F_{\text{feet}} > 350)$
feet slipping	$-\|\mathbf{v}_{\text{feet}} \cdot \mathbf{1}[F_{\text{feet}} > 1]\|$
joint velocities	$-\|\dot{\mathbf{q}}\|$
joint acceleration	$-\|\ddot{\mathbf{q}}\|$
action rate	$-\|\mathbf{a}_t - \mathbf{a}_{t-1}\|$
feet air time	$\min(t_{\text{air time}} - 0.25, 0)$

Real-World Infrastructure. Our system integrates a joystick controller into the real-world setup to allow seamless pausing and resuming of the teleoperation process. This feature is crucial for long-horizon tasks, where the human operator may occasionally need to pause control and adjust their own position.