Humanoid Everyday: A Comprehensive Robotic Dataset for Open-World Humanoid Manipulation

Zhenyu Zhao[1*], Hongyi Jing[1*], Xiawei Liu[1], Jiageng Mao[1†], Abha Jha[1],
Hanwen Yang[1], Rong Xue[1], Sergey Zakharov[2], Vitor Guizilini[2], Yue Wang[1†]
[1]University of Southern California [2]Toyota Research Institute
[*]Equal contribution. [†]Equal advising.

Abstract

From locomotion to dexterous manipulation, humanoid robots have made remarkable strides in demonstrating complex full-body capabilities. However, the majority of current robot learning datasets and benchmarks mainly focus on stationary robot arms, and the few existing humanoid datasets are either confined to fixed environments or limited in task diversity, often lacking human-humanoid interaction and lower-body locomotion. Moreover, there are few standardized evaluation platforms for benchmarking learning-based policies on humanoid data. In this work, we present Humanoid Everyday, a large-scale and diverse humanoid manipulation dataset characterized by extensive task variety involving dexterous object manipulation, human-humanoid interaction, locomotion-integrated actions, and more. Leveraging a highly efficient human-supervised teleoperation pipeline, Humanoid Everyday aggregates high-quality multimodal sensory data—including RGB, depth, LiDAR, and tactile inputs—together with natural language annotations, comprising 10.3k trajectories and over 3 million frames of data across 260 tasks in 7 broad categories. In addition, we conduct an analysis of representative policy learning methods on our dataset, providing insights into their strengths and limitations across different task categories. For standardized evaluation, we introduce a cloud-based evaluation platform that allows researchers to seamlessly deploy their policies in our controlled setting and receive performance feedback. By releasing Humanoid Everyday along with our policy learning analysis and a standardized cloud-based evaluation platform, we intend to advance research in general-purpose humanoid manipulation and lay the groundwork for more capable and embodied robotic agents in real-world scenarios. Our dataset, data collection code, and cloud evaluation website will be released upon acceptance.

1. Introduction

Recent progress in humanoid robotics has significantly reduced the embodiment gap, enabling robots to perform dynamic activities such as running, dancing, and complex full-body movements [3, 8, 11, 16]. However, collecting humanoid manipulation datasets remains challenging. It requires operating in both indoor and outdoor environments, executing a wide range of tasks, and leveraging the humanoid form through bimanual coordination, full-body motion, and human-centric interactions. Existing datasets mainly target stationary arms or mobile platforms with simple grippers and wheeled bases [2, 5, 6, 12, 15]; even egocentric efforts like Humanoid Policy [14] emphasize repetitive tasks with limited locomotion.

In this work, we introduce the Humanoid Everyday Dataset, a large-scale collection of 260 tasks across 7 categories (see Figure 1) that captures full-body locomotion, dexterous manipulation, and rich human-humanoid interactions in diverse environments. Unlike prior datasets limited in scope or simple settings [2, 6, 14], we accumulate more comprehensive tasks across various environments, retaining human supervision for data accuracy. We propose a multiprocessing teleoperation pipeline, re-engineered upon the official Unitree teleoperation library and significantly improved for large-scale, high-quality humanoid data collection. Our pipeline halves data collection time compared to the official Unitree teleoperation system, while the control delay decreases from 500 ms to 2 ms. At 30Hz, our pipeline captures high-resolution sensor and action data from every episode. We record each task with egocentric RGB video, depth information, LiDAR scans, tactile and IMU information, joint poses, joint actions, and natural language annotations.

Our contributions are threefold: (1) a large-scale multimodal humanoid manipulation dataset collected in diverse real-world scenarios with an optimized teleoperation pipeline; (2) an analysis of representative policy learning methods on Humanoid Everyday, highlighting their

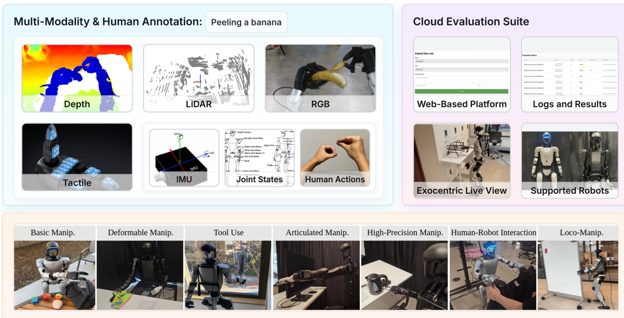

Figure 1. **Humanoid Everyday Overview.** Humanoid Everyday covers 7 distinct categories of humanoid manipulation tasks with rich multimodal information, and provides a cloud-based evaluation platform for standardized policy deployment.

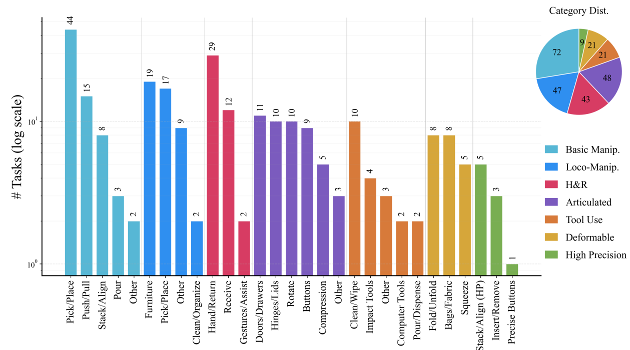

Figure 2. **Task category distribution.** Humanoid Everyday spans 7 distinct categories of humanoid manipulation tasks, each contributing to the overall diversity of the dataset.

strengths and limitations across task categories; and (3) a cloud-based evaluation platform that enables standardized, reproducible, and collaborative research in humanoid manipulation.

2. Experiments

To understand how existing imitation learning approaches perform in the context of humanoid manipulation, we evaluate a variety of policies on our Humanoid Everyday dataset, including Diffusion Policy (DP) [4], 3D Diffusion Policy (DP3) [17], Action Chunking with Transformers (ACT) [18], OpenVLA [9], π_0-FAST [13], $\pi_{0.5}$ [7], and GR00T N1.5 [1]. We train these policies on our 30Hz humanoid data until convergence. For VLA-based policies, we adopt a two-stage finetuning strategy: we first fine-tune on the full Humanoid Everyday dataset and then further adapt each model to task-specific data from individual categories.

We conduct experiments across all seven task categories in Humanoid Everyday. We present the results of these policies in Table 1, where each policy is evaluated over 10 trials on seven tasks.

Overall, all the end-to-end imitation policies struggle in humanoid manipulation tasks due to the high-dimensional action space in our dataset (which is 28 DoFs in total). In contrast, large VLA models demonstrate consistent and stable performance on humanoid manipulation, benefiting from pretraining priors that improve generalization, particularly on Deformable Manipulation and Loco-Manipulation tasks that require precision and robustness. In particular, GR00T N1.5 achieves the strongest performance overall, largely due to its extensive pretraining across multiple large-scale humanoid datasets, which provides a powerful prior well-suited for our diverse manipulation tasks.

Humanoid Everyday is a large-scale, diverse collection of humanoid manipulation tasks, comprising seven main distinct categories including Basic Manipulation (basic object pick-and-place manipulation), Deformable Manipulation (interacting with cloths or other deformable objects), Articulated Manipulation (operating hinged or jointed structures), Tool Use (utilizing external objects to achieve goals), High-Precision Manipulation (performing difficult tasks requiring high accuracy), Human-Robot Interaction (engaging in cooperative actions with humans), and Loco-Manipulation (combining locomotion and manipulation) (see Figure 2). Our tasks are performed in indoor and outdoor environments, involving complex interactions with surrounding objects and dynamic settings. Additionally, a portion of these tasks requires lower-body locomotion, adding further diversity to the dataset. This variety in environmental contexts enriches the collected data and supports the development of generalizable policies capable of adapting to diverse real-world scenarios.

To further support the community, we introduce a cloud-based evaluation platform that allows researchers to upload their policies, execute them within our standardized real-world environment, and receive detailed performance feedback. Unlike existing cloud evaluation systems that focus on robotic arms [19] or simulated agents [10], ours is the first platform designed for humanoids, which aims to lower the barrier to fair comparison and fosters collaborative progress in humanoid robotics.

Category	**Task**	**DP**	**DP3**	**ACT**	**OVLA**	π_0**F**	$\pi_{0.5}$	**GR00T**
Artic.	Rotate chair	100%	90%	100%	70%	100%	100%	100%
Tool	Use eraser to wipe desk	0%	70%	0%	30%	40%	40%	0%
Basic	Put dumpling toy into plate	30%	20%	70%	30%	60%	30%	80%
Deform.	Fold towel on desk	0%	20%	0%	40%	20%	40%	50%
HRI	Hand over dumpling toy	40%	40%	70%	60%	30%	40%	100%
Loco.	Walk to grab door handle	30%	0%	0%	30%	10%	0%	30%
Precision	Insert rose into vase	0%	0%	0%	10%	0%	0%	0%
Avg.		29%	34%	34%	39%	37%	36%	51%

Table 1. Success rates of imitation learning methods on the Humanoid Everyday Dataset.

References

[1] Johan Bjorck, Valts Blukis, Fernando Castaneda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Jim Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, Joel Jang, Xiaowei Jiang, Kaushil Kundalia, Jan Kautz, Zhiqi Li, Zongyu Lin, Kevin Lin, Loic Magne, Yunze Man, Ajay Mandlekar, Avnish Narayan, Soroush Nasiriany, Scott Reed, You Liang Tan, Guanzhi Wang, Jing Wang, Qi Wang, Shihao Wang, Jiannan Xiang, Yuqi Xie, Yizhen Xu, Seonghyeon Ye, Zhiding Yu, Yizhou Zhao, Zhe Zhang, Ruijie Zheng, and Yuke Zhu. Gr00t n1.5: An improved open foundation model for generalist humanoid robots. Model Card / Technical Report, NVIDIA Research, 2025.

[2] Qingwen Bu, Jisong Cai, Li Chen, Xiuqi Cui, Yan Ding, Siyuan Feng, Shenyuan Gao, Xindong He, Xu Huang, Shu Jiang, et al. Agibot world colosseo: A large-scale manipulation platform for scalable and intelligent embodied systems. *arXiv preprint arXiv:2503.06669*, 2025.

[3] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[4] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, page 02783649241273668, 2023.

[5] Hao-Shu Fang, Hongjie Fang, Zhenyu Tang, Jirong Liu, Chenxi Wang, Junbo Wang, Haoyi Zhu, and Cewu Lu. Rh20t: A comprehensive robotic dataset for learning diverse skills in one-shot. *arXiv preprint arXiv:2307.00595*, 2023.

[6] Yao Mu Fourier ActionNet Team. Actionnet: A dataset for dexterous bimanual manipulation. 2025.

[7] Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Manuel Y. Galliker, Dibya Ghosh, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Devin LeBlanc, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Allen Z. Ren, Lucy Xiaoyang Shi, Laura Smith, Jost Tobias Springenberg, Kyle Stachowicz, James Tanner, Quan Vuong, Homer Walke, Anna Walling, Haohuan Wang, Lili Yu, and Ury Zhilinsky. $\pi_{0.5}$: a vision-language-action model with open-world generalization, 2025.

[8] Mazeyu Ji, Xuanbin Peng, Fangchen Liu, Jialong Li, Ge Yang, Xuxin Cheng, and Xiaolong Wang. Exbody2: Advanced expressive humanoid whole-body control. *arXiv preprint arXiv:2412.13196*, 2024.

[9] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[10] Xuanlin Li, Kyle Hsu, Jiayuan Gu, Karl Pertsch, Oier Mees, Homer Rich Walke, Chuyuan Fu, Ishikaa Lunawat, Isabel Sieh, Sean Kirmani, et al. Evaluating real-world robot manipulation policies in simulation. *arXiv preprint arXiv:2405.05941*, 2024.

[11] Jiageng Mao, Siheng Zhao, Siqi Song, Tianheng Shi, Junjie Ye, Mingtong Zhang, Haoran Geng, Jitendra Malik, Vitor Guizilini, and Yue Wang. Learning from massive human videos for universal humanoid pose control. *arXiv preprint arXiv:2412.14172*, 2024.

[12] Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903. IEEE, 2024.

[13] Karl Pertsch, Kyle Stachowicz, Brian Ichter, Danny Driess, Suraj Nair, Quan Vuong, Oier Mees, Chelsea Finn, and Sergey Levine. Fast: Efficient action tokenization for vision-language-action models. *arXiv preprint arXiv:2501.09747*, 2025.

[14] Ri-Zhao Qiu, Shiqi Yang, Xuxin Cheng, Chaitanya Chawla, Jialong Li, Tairan He, Ge Yan, David J. Yoon, Ryan Hoque, Lars Paulsen, Ge Yang, Jian Zhang, Sha Yi, Guanya Shi, and Xiaolong Wang. Humanoid policy ˜ human policy. *arXiv preprint arXiv:2503.13441*, 2025.

[15] Homer Rich Walke, Kevin Black, Tony Z Zhao, Quan Vuong, Chongyi Zheng, Philippe Hansen-Estruch, Andre Wang He, Vivek Myers, Moo Jin Kim, Max Du, et al. Bridgedata v2: A dataset for robot learning at scale. In *Conference on Robot Learning*, pages 1723–1736. PMLR, 2023.

[16] Yufei Xue, Wentao Dong, Minghuan Liu, Weinan Zhang, and Jiangmiao Pang. A unified and general humanoid whole-body controller for fine-grained locomotion. *arXiv preprint arXiv:2502.03206*, 2025.

[17] Yanjie Ze, Gu Zhang, Kangning Zhang, Chenyuan Hu, Muhan Wang, and Huazhe Xu. 3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations. *arXiv preprint arXiv:2403.03954*, 2024.

[18] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[19] Zhiyuan Zhou, Pranav Atreya, You Liang Tan, Karl Pertsch, and Sergey Levine. Autoeval: Autonomous evaluation of generalist robot manipulation policies in the real world. *arXiv preprint arXiv:2503.24278*, 2025.