Dear Area Chair N8rm and Reviewers 2Dtf, ziut, huqu :

Thank you for the effort in reviewing our manuscript ID [303], titled "OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning". We deeply appreciate the constructive feedback and have carefully considered each point raised. In response, we have revised our manuscript to address the concerns and clarify the ambiguities. Notably, we have

- Provided additional experiments to substantiate our claims and improved clarity in comparisons to prior work (N8rm).
- **Comment 1.2** Detailed differences between OmniH2O and previous work H2O, including tracking design, training design, state space design, and reward design (2Dtf).
- **Comment 1.3** Analyzed the source of OmniH2O's performance gains, highlighting the impact of state space design and training design (2Dtf).
- **Comment 1.4** Clarified the fair comparison by retraining H2O under the same settings as OmniH2O (2Dtf).
- **Comment 1.5, Comment 2.7** Enhanced clarity on the statistical significance of results with added confidence intervals (2Dtf, ziut).
- **Comment 1.6, Comment 1.7** Elaborated on the importance of historical observations in improving real-world performance (2Dtf).
- **Comment 1.8, Comment 2.11** Revised claims regarding the robustness and terrain testing conditions (2Dtf, ziut).
- **Comment 1.9** Explained the method of integrating VLMs for motion goal selection (2Dtf).
- **Comment 1.10** Provided detailed comparisons with prior works and outlined the advantages of whole-body control (2Dtf).
- **Comment 2.2** Addressed the selection process and augmentation strategy for squatting and standing poses (ziut).
- **Comment 2.3** Clarified the description and motivation of reward term "max feet height for each step" (ziut).
- **Comment 2.4** Explained the dynamic reward function curriculum and the rationale behind its design (ziut).
- **Comment 2.5** Clarified the limitations and future integration plans for hand control in the overall framework (ziut).
- **Comment 2.6** Detailed the testing constraints and future plans for non-standing real-world sequences (ziut).
- **Comment 2.8** Added detailed descriptions of the policy architecture earlier in the manuscript (ziut).
- **Comment 2.9** Explained the way to handle input sparsity levels using DAgger (ziut).
- **Comment 2.10** Clarified the process and success rate evaluation of real-world experiments (ziut).
- **Comment 3.2** Detailed the methodological improvements and contributions of our work over prior research (huqu).
- **Comment 3.3, Comment 3.5** Provided comprehensive details on learning from demonstrations and the OmniH2O-6 dataset (huqu).
- **Comment 3.4** Clarified the investigation of humanoid LfD (huqu).
- **Comment 3.6** Detailed the goal specification during outdoor locomotion test (huqu).
- **Comment 3.7** Explained the necessity of distillation process for the teleoperation policy to address input sparsity and partial observability (huqu).
- **Comment 3.8** Discussed the potential use of more accurate odometry systems and their implications for performance (huqu).

We believe these revisions enhance the clarity and impact of our work.

Yours sincerely,
Authors

Response to Meta-Review of Area Chair N8rm

Comment

The reviewers talk positively about the paper. The reviewers highlight that this is one of the first papers doing full-body mimicry for humanoids. Furthermore, the paper does many different qualitative experiments, including tennis, drawing, watering plants, and picking objects in indoor and outdoor environments. The main weaknesses are 1) that the paper makes unsubstantiated claims about improving over previous work and the experiment complexity. The statements mainly exaggerate the contributions. 2) the clarity of the paper and the comparisons to prior works. Both could be improved. The authors put too much content into a single paper, which makes the methods and comparisons partially unclear and leaves many questions compared to prior work.

Response:

We thank the area chair for the meta-review and the recommendation. We have clarified the requested details in our response and have accordingly revised our paper.

Specifically, we have substantiated our claims with additional experimental data and provided clearer comparisons to prior work. And we will continue streamlining the content to enhance clarity and focus, ensuring that our contributions are more precisely articulated and easier to understand.

1 Response to Reviewer 2Dtf

Comment 1.1

Summary Of Contribution:
The paper presents a humanoid whole-body control RL policy, OmniH2O, and integrates it with various teleoperation methods. The RL training reward, observation, and architecture are modified to remove the body velocity observation while improving tracking performance compared to prior work. The release includes a teleoperation dataset recorded using the controller.

Summary Of Strengths:
The universal interface supporting various input modalities (mocap, RGB, language) is a successful integration of many diverse components. The OmniH2O-6 dataset will provide robot data to the community for task/embodiment combinations that are not currently accessible. Compared to H2O, removing the requirement for mocap tracking during deployment allows the policy to be deployed outside of the lab.

Response:

We appreciate the reviewer's feedback and recommendation. Your acknowledgement is sincerely valued. We have carefully reviewed your comments and responded to the concerns raised.

Comment 1.2

Differences between OmniH2O and previous work H2O: Is it possible to enumerate the exact difference between OmniH2O and H2O in a table with an ablation for each difference?

Response:

Table 1: Differences between OmniH2O and H2O.

	H2O	OmniH2O
Tracking Design		
Number of Tracking Keypoints	8	3
Tracking Keypoints	Shoulders, Elbows, Hands, Ankles	Head, Hands
Supported Teleoperation Interface	RGB	RGB, VR
Training Design		
Training pipeline	RL	RL+DAgger
State Space Design		
Root Linear Velocity	Needed	Not Needed
Deployment Requirment	Indoor Motion Capture	Outdoor/Indoor
History utilization	Not Utilized	Utilized
Motion Dataset Design		
Motion Dataset	Retargeted AMASS	Retargeted AMASS + Standing&Squatting Variants
Reward Design		
Penalty Terms	3	4 (Added DOF velocity limits)
Regularization Terms	8	13 (Added More Regularizations for Better Sim2Real)
Task Rewards Terms	6	7 (Added Hand/Head Tracking Rewards)
Domain Randomization Design		
Dynamics Randomization Terms	7	8 (Added Randomized Motion Offset)

Thank you for your insightful observation. The exact difference between OmniH2O and H2O are summarized in the following Table 1. Note that H2O has specific data input and sensor requirement (MoCap and RGB) where OmniH2O aims at deploy stable humanoid loco-manipulation in the wild.

Our reported numbers for H2O are retrained using the same setting (rewards, domain randomization, and motion dataset) as OmniH2O for a fair comparison. The main differences behind the experimental results are 1) 3-point tracking vs. 8-point tracking; 2) RL vs. DAgger distillation 3) state space design (linear velocity, history). We will further breakdown the source of improvement in the next comment.

Comment 1.3

The source of the OmniH2O performance gain: A key claim of the paper is that OmniH2O "significantly improves upon prior art" H2O. Right now, it's hard to interpret which of the differences account for most of the performance. As far as I can tell, the ablations (Table 1) already include the differences: [(1) number of keypoints, (2) lin vel observation], but there are also [(3) modified reward function, (4) modified domain randomization, (5) standing trajs added to the motion dataset, (maybe others?)]. Ablations 1 and 2 don't seem to explain much of the performance gap; can the improvement can be attributed to a specific design choice like a reward term?

Response:

It is important to note that most design choices for OmniH2O aim to achieve better real-world hardware performance rather than simulation performance, as reported in Table 1. Thus, the performance gain of OmniH2O over H2O should be discussed separately by **real-world** and **simulation**. Based on real-world performance-driven design, we reduced the 8 tracking points of H2O to **3 tracking points** to make the policy compatible with a more precise and robust VR teleoperation interface. Additionally, we **eliminated the requirement for linear velocity observation** to overcome the challenge of estimating the velocity in the real world. However, these designs also raise more challenges for the RL training process, as the motion goals become much more sparse. To address these challenges, we leverage a **teacher-student training pipeline** and incorporate **historical observations** to overcome the limitations brought by 3-point tracking and the absence of linear velocity input.

Specifically, the **real-world performance gain** of OmniH2O over H2O reported in Table 2 come from the following factors:

- **State Space Design**: The linear velocity information used in H2O (estimated using MoCap) is difficult to obtain reliably in the real-world via onboard processing (see Appendix G). As a result, OmniH2O excludes linear velocity in the state space and include history information to facilitate better in-the-wild real-world deployment. Table 2 shows the real-world comparison between H2O (uses ZED SDK to estimate linear velocity) and OmniH2O (uses history) when mimicing standing motions. Note that H2O has 1.86x root-relative per-joint error E_{mpjpe} compared to OmniH2O. We provide qualitative comparison in Figure 1 and video in the anonymous link https://anonymous-omni-h2o.github.io/resources/rebuttal/OmniH2O_vs_H2O_short.mp4 where OmniH2O remains stable during the test, while H2O shakes while taking small steps forward and backward. We also compare OmniH2O with or without history utilization in Table 2 and Figure 2. We can see that in the real world OmniH2O-NoHistory continues to take small steps forward and backward. A detailed video comparison of these ablations is linked in this anonymous link https://anonymous-omni-h2o.github.io/resources/rebuttal/NoHistory.mp4, validating the source of the real-world performance gain coming from state space design of linear velocity and history.

Table 2: Real-world motion tracking evaluation on 20 standing motions.

		Tested sequences			
Method	State Dimensions	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$\text{E}_{\text{acc}} \downarrow$	$\text{E}_{\text{vel}} \downarrow$
H2O	$\mathcal{S} \subset \mathcal{R}^{138}$	87.33	53.32	6.03	5.87
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	**47.94**	**41.87**	**1.84**	**2.20**
(a) Ablation on Real-world Linear Velocity Estimation					
OmniH2O-w-linvel(VIO)[1,2]	$\mathcal{S} \subset \mathcal{R}^{1743}$	N/A	N/A	N/A	N/A
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	**47.94**	**41.87**	**1.84**	**2.20**
(b) Ablation on History steps/Architecture					
OmniH2O-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	83.26	46.00	4.86	4.45
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	**47.94**	41.87	**1.84**	**2.20**

[1] Use ZED SDK to estimate the linear velocity.
[2] Unable to finish the real-world test due to falling on the ground.

The state space design mainly contributes to the **real-world performance gain**. Regarding the **simulation performance gain** over H2O reported in Table 1 of the original manuscript, there are two primary factors: the **number of tracking points** and **training design** (teacher-student distillation):

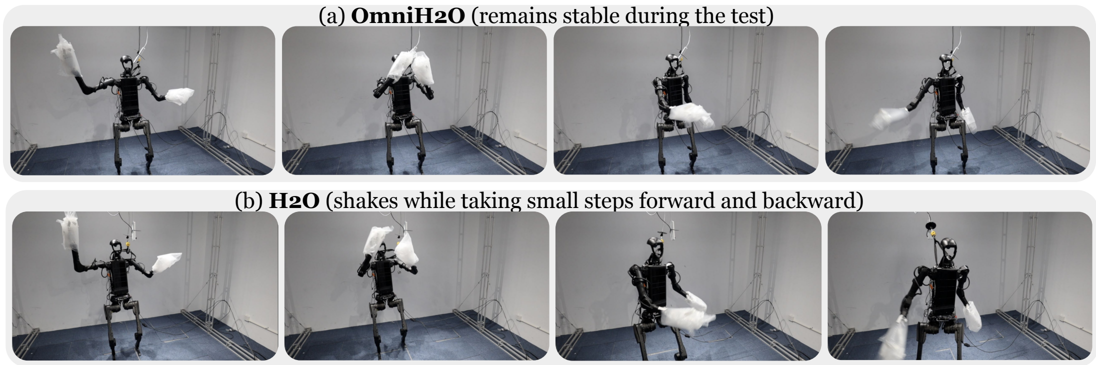

Figure 1: Comparison between H2O and OmniH2O during the real-world tests. OmniH2O remains stable during the test, while H2O shakes while taking small steps forward and backward. The performance gain comes from elimating root linear velocity in the state space to history utilization. Note that both polices are trained with the same rewards, domain randomizations and motion datasets.

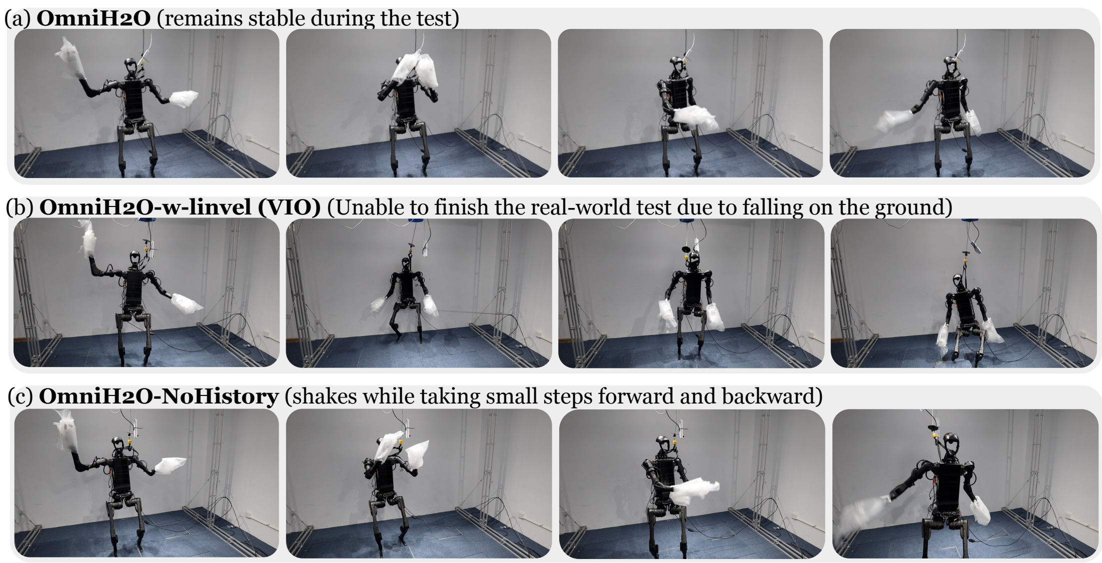

Figure 2: Ablation study on OmniH2O with linear velocity observation and without history, validating the performance gain of real-world comes from elimating root linear velocity in the state space and history utilization. Note that all polices are distilled from the exact same teacher policy.

- **Tracking Design (8-point tracking -> 3-point tracking)**: As shown in Table 3 (b), while the tracking precision metrics ($E_{\text{g-mpjpe}}$ and E_{mpjpe}) downgrade from 8-point to 3-point tracking, the impact on performance is not substantial.

- **Training Design (DAgger)**: The key training difference between H2O and OmniH2O is that OmniH2O uses a teacher-student framework to distill a student policy for sim2real from a privileged policy, whereas H2O directly trains the sim2real policy using RL. The partially observed state space in H2O hinders RL's ability to find an optimal policy. In contrast, OmniH2O mitigates this issue through supervised learning in the distillation process. As shown in Table 3 (a), there is a significant performance drop when using H2O-style sim2real RL training.

In conclusion, **state space design** (no linear velocity & history utilization) accounts for most of the **real-world performance gain**, and **training design** (teacher-student DAgger distillaion) accounts for most of the **simulation performance gain**.

Table 3: Simulation motion imitation evaluation of OmniH2O and baselines on dataset. Note that all the variants are trained with exact same rewards, domain randomizations and motion dataset $\hat{\boldsymbol{Q}}$.

			All sequences					Successful sequences			
Method	State Dimension	Sim2Real	Succ ↑	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓
Privileged policy	$\mathcal{S} \subset \mathcal{R}^{913}$	✗	94.77%	126.51	70.68	3.57	6.20	122.71	69.06	2.22	5.20
H2O	$\mathcal{S} \subset \mathcal{R}^{138}$	✓	87.52%	148.13	81.06	5.12	7.89	133.28	75.99	2.40	5.75
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%**	**141.11**	**77.82**	**3.70**	**6.54**	**135.49**	**75.75**	**2.30**	**5.47**
(a) Ablation on DAgger/RL											
OmniH2O-w/o-DAgger	$\mathcal{S} \subset \mathcal{R}^{1665}$	✗	47.11%	223.27	128.90	15.03	16.29	182.13	119.54	5.47	9.10
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%**	**141.11**	**77.82**	**3.70**	**6.54**	**135.49**	**75.75**	**2.30**	**5.47**
(b) Ablation on Tracking Points											
OmniH2O-8points	$\mathcal{S} \subset \mathcal{R}^{1710}$	✓	**94.31%**	**129.30**	**71.70**	3.78	**6.39**	**125.14**	**70.07**	**2.22**	**5.26**
OmniH2O-3points (Ours)	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	94.10%	141.11	77.82	**3.70**	6.54	135.49	75.75	2.30	5.47

Comment 1.4

Fair Comparison: Relatedly in Table 1, H2O and OmniH2O are evaluated on dataset. Is this a fair comparison since OmniH2O was trained on but H2O was trained on a different dataset and therefore being evalauted out-of-distribution?

Response:

Thank you for pointing this out. For all the evaluation results reported, H2O and all the other ablations are re-trained **under the exact same reward/randomization/motion setting**. Thus, this is a fair comparison between all the baselines and ablations, and we have added more details to clarify the confusion in the revised manuscript.

Comment 1.5

Can you report the statistical significance in Table 1?

Response:

Thank you for pointing out the missing statistical significance of the evaluation results. We added the statistical variance that is not included in our submission. The standard deviation is calculated by 5 runs for each motion in the dataset.

Table 4: Simulation motion imitation evaluation of OmniH2O and baselines on dataset. Note that all the variants are trained with exact same rewards, domain randomizations and motion dataset. The standard deviation is calculated over 5 random seeds.

			All sequences					Successful sequences			
Method	State Dimension	Sim2Real	Succ ↑	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓
Privileged policy	$\mathcal{S} \subset \mathcal{R}^{913}$	✗	94.77% ±0.32%	126.51 ±0.66	70.68 ±0.35	3.57 ±0.06	6.20 ±0.07	122.71 ±0.49	69.06 ±0.28	2.22 ±0.03	5.20 ±0.04
H2O	$\mathcal{S} \subset \mathcal{R}^{138}$	✓	87.52% ±0.19%	148.13 ±0.75	81.06 ±0.40	5.12 ±0.04	7.89 ±0.08	133.28 ±0.55	75.99 ±0.36	2.40 ±0.04	5.75 ±0.05
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	**141.11** ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	**135.49** ±0.68	**75.75** ±0.31	**2.30** ±0.02	**5.47** ±0.05
(a) Ablation on DAgger/RL											
OmniH2O-w/o-DAgger-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✗	90.62% ±0.25%	163.44 ±1.01	91.29 ±0.58	5.12 ±0.07	8.80 ±0.09	153.31 ±0.79	87.59 ±0.39	3.15 ±0.03	7.27 ±0.05
OmniH2O-w/o-DAgger	$\mathcal{S} \subset \mathcal{R}^{1665}$	✗	47.11% ±0.14%	223.27 ±1.47	128.90 ±0.81	15.03 ±0.09	16.29 ±0.12	182.13 ±1.22	119.54 ±0.69	5.47 ±0.04	9.10 ±0.08
OmniH2O-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	93.80% ±0.33%	141.21 ±0.73	78.52 ±0.39	3.74 ±0.03	6.62 ±0.08	**134.90** ±0.59	76.11 ±0.27	**2.25** ±0.02	5.48 ±0.04
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	**141.11** ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05
(b) Ablation on History steps/Architecture											
OmniH2O-History50	$\mathcal{S} \subset \mathcal{R}^{3240}$	✓	93.56% ±0.34%	141.51 ±0.78	78.51 ±0.41	4.01 ±0.05	6.79 ±0.06	135.04 ±0.63	76.07 ±0.33	2.36 ±0.02	5.55 ±0.05
OmniH2O-History5	$\mathcal{S} \subset \mathcal{R}^{405}$	✓	93.60% ±0.23%	**139.23** ±0.67	**77.82** ±0.35	3.91 ±0.04	6.66 ±0.07	132.67 ±0.55	75.33 ±0.26	2.24 ±0.02	5.41 ±0.04
OmniH2O-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	93.80% ±0.33%	141.21 ±0.73	78.52 ±0.39	3.74 ±0.03	6.62 ±0.08	**134.90** ±0.59	76.11 ±0.27	**2.25** ±0.02	5.48 ±0.05
OmniH2O-GRU	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	92.85% ±0.22%	147.67 ±0.75	80.84 ±0.38	4.05 ±0.04	6.93 ±0.09	142.75 ±0.67	79.10 ±0.32	2.38 ±0.02	5.66 ±0.03
OmniH2O-LSTM	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	91.03% ±0.21%	147.36 ±0.74	80.34 ±0.41	4.12 ±0.06	7.04 ±0.11	142.64 ±0.66	78.59 ±0.34	2.37 ±0.01	5.72 ±0.05
OmniH2O-History25 (Ours)	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	141.11 ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05
(c) Ablation on Tracking Points											
OmniH2O-22points	$\mathcal{S} \subset \mathcal{R}^{1836}$	✓	**94.72%** ±0.35%	**127.71** ±0.54	**70.39** ±0.34	**3.62** ±0.04	**6.25** ±0.05	**123.87** ±0.45	**68.92** ±0.27	**2.22** ±0.02	**5.24** ±0.04
OmniH2O-8points	$\mathcal{S} \subset \mathcal{R}^{1710}$	✓	94.31% ±0.28%	129.30 ±0.63	71.70 ±0.36	3.78 ±0.03	6.39 ±0.06	125.14 ±0.54	70.07 ±0.32	2.22 ±0.02	5.26 ±0.03
OmniH2O-3points (Ours)	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	94.10% ±0.35%	141.11 ±0.77	77.82 ±0.45	3.70 ±0.03	6.54 ±0.06	135.49 ±0.68	75.75 ±0.31	2.30 ±0.02	5.47 ±0.05
(d) Ablation on Linear Velocity											
OmniH2O-w-linvel	$\mathcal{S} \subset \mathcal{R}^{1743}$	✓	93.80% ±0.33%	**138.18** ±0.71	78.12 ±0.38	3.94 ±0.05	6.61 ±0.07	**132.44** ±0.54	75.98 ±0.30	**2.29** ±0.02	**5.40** ±0.03
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	141.11 ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05

Comment 1.6

History Utilization: The difference between the student with no history (OmniH2O-History0) vs. training with history (OmniH2O) is only 0.3%. This difference seems small; is it statistically significant across multiple seeds of training? If not, is it valid to assert that history utilization is an essential element? Following on this point, Table 2 suggests that omniH2O outperforms omniH2O-History0 by a wide margin in the real world despite nearly identical simulation performance. Does this suggest that using a history of observations helps mitigate the sim2real gap? What would explain this since there does not seem to be an obvious mechanism for the history length to influence the sim2real quality while both are enjoying high performance on the training data?

Response:

Yes, the history of observations largely helps mitigate the sim2real gap; we hypothesize that the history provides linear velocity information that is hard to estimate in the real world. In Figure 3, each subfigure presents the behavior of OmniH2O under varying conditions: (a) **OmniH2O**: The robot remains stable during the test, indicating robust performance without any shaking or instability. (b) **OmniH2O-History-0**: The robot shakes while taking small steps forward and backward, highlighting instability and less reliable performance. (c) **OmniH2O-History-5**: The shaking is less severe compared to OmniH2O-History-0, suggesting an improvement in stability with a history length of 5. (d) **OmniH2O-History-50**: The robot remains stable during the test, similar to OmniH2O, demonstrating that a longer history length (50) contributes to stable and reliable performance. All policies tested in this study are distilled from the same teacher policy. The results indicate that the history utilizaion significantly impacts the stability and performance of the robot in the real world, which echoes with prior works (https://arxiv.org/pdf/2401.16889, https://arxiv.org/pdf/2303.03381).

The video evaluation, available at https://anonymous-omni-h2o.github.io/resources/rebuttal/HistoryLength.mp4, clearly demonstrates that incorporating historical observations helps bridge the sim2real gap. Without history, the student policy can only learn a mixture of teacher policy actions for identical partial observations, leading to degraded performance in real-world scenarios (even though this mixture is not severe enough to hurt the simulation performance). Conversely, with historical observations, the student policy can achieve a more accurate action mapping from past observations, thereby enhancing stability and reliability. This empirical evidence underscores the importance of proper history lengths for effective policy deployment in real-world applications.

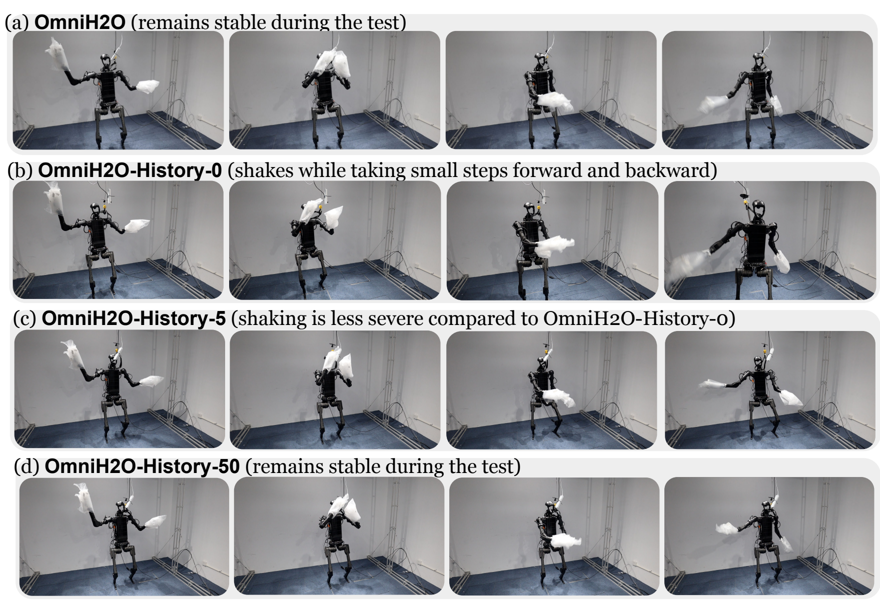

Figure 3: Ablation study on OmniH2O with different history lengths. Note that all polices are distilled from the exact same teacher policy.

Comment 1.7

As it stands, I would be inclined to draw the opposite conclusion from this table and say that history length seems to have no trend / irrelevant for the OmniH2O policy. Therefore, the role of linear velocity observation would be mainly in easing the learning dynamics, and it does not need to be reconstructed from the history for humanoid locomotion. This is somewhat surprising since other works on bipedal (https://arxiv.org/pdf/2401.16889) and even quadruped locomotion (e.g. https://arxiv.org/abs/2301.10602, https://arxiv.org/abs/2202.05481) found a history module important for velocity estimation. Can you comment on this interpretation of the results?

Response:

In reference to Comment 1.6, the primary advantage and necessity of utilizing historical data lie in enhancing sim-to-real performance. Our empirical results highlight the critical importance of linear velocity (whether explicitly estimated or implicitly encoded) in ensuring the robustness of humanoids under disturbances, and in accurately determining the transition between standing and walking. The privileged teacher policy benefits from observing the velocities of all rigid body links, and distilling actions from this teacher implicitly encodes the velocity estimators. Unlike prior works that explicitly estimate root velocity, our approach does not reconstruct velocity from historical data directly. Instead, we enforce loss at the action-level distillation.

While we acknowledge the necessity of historical observations for improved legged locomotion, our findings suggest that explicitly reconstructing velocity from historical data is not the only viable approach. We compared the performance of using explicitly reconstructed velocities from historical data with MLP or GRU against the ZED camera VIO module. As shown in Table 5, the former performed better empirically, yet it did not surpass the implicit version used in OmniH2O. The video evaluation, available at this anonymous link https://anonymous-omni-h2o.github.io/resources/rebuttal/LinearVelocity.mp4, demonstrates that our method of implicit velocity estimation via action-level distillation offers superior performance, reinforcing the value of our approach.

Table 5: Real-world motion tracking evaluation on 20 standing motions.

		Tested sequences			
Method	State Dimensions	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$\text{E}_{\text{acc}} \downarrow$	$\text{E}_{\text{vel}} \downarrow$
(a) Ablation on Real-world Linear Velocity Estimation					
OmniH2O-w-linvel(VIO)[1,2]	$\mathcal{S} \subset \mathcal{R}^{1743}$	N/A	N/A	N/A	N/A
OmniH2O-w-linvel(MLP)	$\mathcal{S} \subset \mathcal{R}^{1743}$	50.93	42.47	2.16	2.26
OmniH2O-w-linvel(GRU)	$\mathcal{S} \subset \mathcal{R}^{1743}$	49.75	42.38	2.20	2.31
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	**47.94**	**41.87**	**1.84**	**2.20**

[1] Use ZED SDK to estimate the linear velocity.
[2] Unable to finish the real-world test due to falling on the ground.

Comment 1.8

"OmniH2O demonstrates great robustness under disturbances and unstructured terrains." "OmniH2O shows superior robustness against human strikes and different outdoor terrains." Superior to what? I am not sure how the reader is supposed to evaluate these claims. Is the robustness better than Unitree's default controller, better than H2O, or other or cited works like https://arxiv.org/abs/2402.16796 or https://arxiv.org/pdf/2303.03381 ? The disturbance rejection looks somewhat similar to me. The locomotion in all videos takes place on flat surfaces and slight inclines which I do not believe should be referred to as unstructured terrains. What would you consider a structured terrain?

Response:

Thank you for highlighting the need for clarity in our claims regarding robustness and terrain. We have revised the manuscript to avoid overstatements and to better contextualize our findings.

OmniH2O advances H2O's robustness in the following crucial ways: (1) OmniH2O demonstrates more real-world movement types than H2O, and achieves higher success rate in motion imitation; (2) OmniH2O no longer relies on Motion Capture (MoCap) devices and demonstrates outdoor capabilities.

When compared to other cited works such as https://arxiv.org/abs/2402.16796 and https://arxiv.org/pdf/2303.03381, OmniH2O demonstrates the unique capability to maintain stability and return to a standing position after disturbances. These prior works tend to trade off the ability to stand still for enhanced robustness. The ability to remain stable is vital for tasks requiring precise loco-manipulation, especially in scenarios involving dexterous manipulation.

Regarding terrain, we have updated our description to more accurately reflect the testing conditions. Our tests mainly involved flat surfaces and slight inclines, which should not be classified as unstructured terrains. Structured terrains in our context refer to environments with well-defined and consistent features, such as flat floors or uniformly inclined surfaces, as opposed to irregular or unpredictable outdoor terrains.

We hope that this clarification provides a better understanding of our contributions and the specific contexts in which our approach excels.

Comment 1.9

Does the teleoperation from verbal instruction w/ gpt-4 have significant differences from SayTap (https://arxiv.org/pdf/2306.07580) that prompts a VLM to output motion parameters for a quadruped and Code as Policies https://arxiv.org/abs/2209.07753 that prompts a VLM to output code?

Response:

We connect OmniH2O policy with GPT-4 by prompting it to select from pre-defined motion sequences. Detailed prompts are provided in Appendix M of our submission. We chose this motion-selection method due to the high-latency I/O of VLMs like GPT-4. Unlike SayTap, which prompts a VLM to output motion parameters for a quadruped, and Code as Policies, which prompts a VLM to output code, our approach leverages VLMs in a more straightforward manner. This method is a proof of concept demonstrating the potential of combining VLMs with humanoid control. We anticipate future advancements in VLMs to address I/O latency and enhance reasoning capabilities.

Comment 1.10

Is there a fundamental advantage of whole-body control over a decoupled approach? It seems like all behaviors could potentially be teleoperated with an interface like https://arxiv.org/abs/2402.16796 which separates control of the arms and legs while also allowing the user to specify a base height command for squatting. It would strongly motivate the coupled training if OmniH2O is capable of some more dynamic movements involving more evident coordination between the upper and lower body. The "embodiment feasible" motion dataset appears to contain several motions like kicking and jumping that involve this type of coordination. Was sim-to-real a limitation for these motions? If so, maybe one could add videos of the robot doing tracking these moves in simulation?

Response:

Thank you for bringing up this important distinction. Thank you for your insightful question. It's essential to distinguish between the architecture and goals of whole-body control (WBC) in humanoid robots. WBC, as an overarching framework, integrates all degrees of freedom (DoFs) to achieve task-specific objectives, enabling complex, coordinated movements that a decoupled approach may struggle to handle.

ExBody, despite being a WBC framework, adopts a decoupled control strategy—separating upper and lower body objectives, which inherently restricts dynamic movements like kicking or bending that require tight coordination between the upper and lower body.

Conversely, OmniH2O's goal specification enables comprehensive whole-body kinematic pose tracking, facilitating the demonstration of complex and dynamic motions. The teacher policy in OmniH2O, which targets whole-body motion tracking, successfully demonstrates good tracking for such coordinated actions. The absence of these behaviors in certain demonstrations can be attributed to two main factors:

1) **sim2real**: the sim2real regularization rewards and randomizations tend to limit the execution of whole-body motions deemed 'risky' for real-world transfer. These constraints ensure safer and more reliable transitions from simulation to physical robots, but can also restrict more dynamic motions;

2) **VR headset sparse input**: the VR headset provides only the pose estimation of the head and hands, making it challenging for the policy to infer complex leg movements, such as kicking. Despite this, the teacher policy, even under stringent sim2real regularizations, can showcase several dynamic whole-body motions.

To illustrate the potential of OmniH2O, we have compiled a video demonstrating the teacher policy executing various dynamic whole-body motions like bending, quick turning etc. This video is available at the following anonymous link: https://anonymous-omni-h2o.github.io/resources/rebuttal/TeacherSimulation.mp4. We believe that this will substantiate the capability of OmniH2O in handling complex coordinated movements.

2 Response to Reviewer ziut

Comment 2.1

Summary Of Contribution:
The paper describes an approach for training a humanoid controller that can mimic human motions. Simulated and real-world experiments are provided to demonstrate the controller's performance relative to a number of ablations. The controller is used to produce imitation learning data for several tasks and a policy is learned on top of the controller for each task.

Summary Of Strengths:
Overall this is one of the first (among a couple others) works on sim2real learning of full-body humanoid mimicry. Thus, from an application perspective the novelty is high.

Response:

We are grateful for your feedback and recommendation. Your recognition holds great value to us. We have carefully reviewed your comments and responded to the concerns raised.

Comment 2.2

How did you select the squatting and standing pose used in the training? This could bias the system to less stable poses, which should ultimately depend on the task. More discussion of this choice would be useful.

Response:

Thank you for your insightful question. Here's how we approached it:

- **Motion Selection:** We uniformly sampled 2000 poses from the AMASS dataset to contruct the standing/squatting variances. We tested the impact of not using augmentation and found that it significantly reduced the model's performance. As shown in Figure 9 in Appendix H, augmentation is necessary to achieve stable and reliable results.
- **Robustness Testing:** To ensure our approach is robust, we tested different sizes of random augmentation of the dataset (20% and 40%). And the resulting polices both demonstrate the robustness of stable standing as shown in the anonymous video https://anonymous-omni-h2o.github.io/resources/rebuttal/MotionAugmentation.mp4

In summary, our pose selection and enhancement strategies help build a robust and reliable training process. We acknowledge that task-specific tuning can further optimize performance, and we encourage future work to explore this aspect in more detail.

Comment 2.3

The main paper should provide at least some description of the reward term "max feet height for each step." since that is claimed to be crucial and is not self-explanatory. I tried to parse Appendix E that the authors pointed to, but it mostly provided factual details without motivation or definition of the symbols.

Response:

Thank you for your valuable feedback. The "max feet height for each step" reward is designed to encourage the robot to take higher steps by rewarding it based on the maximum height achieved by the feet during the air phase of each step. Here's a detailed explanation:

- **Contact Detection:** We first detect when the feet make contact with the ground.
- **Tracking Maximum Height:** During the air phase of each step, we track the maximum height achieved by the feet.
- **Reward Calculation:** The reward is calculated based on the difference between a desired maximum height and the actual maximum height achieved during the air phase. This reward is only applied at the first contact with the ground after the air phase.

The reward encourages the robot to achieve a higher trajectory of the foot during the step. This is essential for better sim2real. We have added an explanation in the revised manuscript to clarify the details of this reward term.

Comment 2.4

I was not able to understand the rationale for the dynamic reward function described in the appendix. Is there a fundamental principle here that can be formalized more soundly?

Response:

Thank you for your question. The fundamental principle behind our dynamic reward function is the concept of curriculum learning, which is widely recognized in machine learning to facilitate gradual learning from easier to harder tasks. Starting training with less severe penalties allows the agent to learn basic behaviors without being overwhelmed by high penalties. As the agent's performance improves, the penalties are gradually increased to encourage learning more complex and precise behaviors.

In summary, our dynamic reward function leverages the fundamental principle of curriculum learning, providing a structured and incremental approach to training. This makes the learning process more effective and easier.

Comment 2.5

Why didn't you try to include the hand control in the overall learning framework? The videos showing hand movement are a bit misleading and this should be made clear early on in the paper.

Response:

Thank you for your observation. Including hand control in the overall learning framework is indeed possible and would enhance the coordination of full-body movements. However, there are specific reasons for our current approach:

Lack of Whole-Body to Finger Motion Dataset: We currently lack a comprehensive dataset that captures the intricate coordination between whole-body movements and fine finger actions. Such datasets are essential for training models that can seamlessly integrate both aspects.

Task Requirements: For most of the tasks we are targeting, dynamic coordination between the body and fingers is not critically required. Decoupling the hand control from the overall body movement allows us to simplify the model and focus on achieving robust whole-body locomotion and manipulation.

We acknowledge that the videos showcasing hand movements might give an impression of integrated hand control within the overall learning framework. To address this, we have made it clear early in the paper that hand control is currently decoupled due to the reasons mentioned above.

In future work, we aim to include hand control as more comprehensive datasets become available, allowing for better integrated whole-body and finger coordination.

Comment 2.6

It would have been very interested to have at least a couple of non-standing real-world sequences. The additional dynamics of such sequences are non-trivial to deal with and it is important to know if the current approach has any fundamental limitations in that respect. Even a negative result would be interesting. The excuse given for not including such sequences (not enough lab space) is not very reasonable given videos already shown.

Response:

Thank you for your valuable feedback. Here's a more detailed explanation of the challenges we faced:

To conduct non-standing real-world tests, we need to move off the gantry system due to the limited size of our gantry. However, when testing off the gantry, the OmniH2O policy is the only one we have confidence in for ensuring stability and safety. Other policies may result in severe falls, potentially damaging the hardware.

After our paper was submitted, the hardware suffered damage, which we have not been able to repair. This constraint limited our ability to include non-standing sequences at this time.

We acknowledge the importance of testing under more dynamic conditions and aim to address these limitations in future work by enhancing our setup and policies to facilitate safe and comprehensive real-world testing.

Comment 2.7

Many of the values in Table 1 are quite similar, including those that are marked in bold as somehow significantly better. I'm not sure I can agree with conclusions regarding the relative performance of some of these choices. These are also hard to evaluate since they are apparently averaged over all 14,000 sequences. Was there any randomization in the experiments (e.g. multiple randomizations for each sequence)? If so, the confidence intervals could be given?

Response:

Thank you for your observation. We understand the concern regarding the similarities in the values presented in Table 1. Here is a detailed explanation to address your concerns:

Metrics: While success rate is straightforward, tracking precision is measured by mean joint position error (MJPE), which is more informative and distinctive. We believe that MJPE provides a clearer picture of the model's performance.

Randomization and Confidence Intervals: The experiments were conducted by randomizing the environment with five random seeds for each motion sequence. This approach ensures that the results are not biased by a specific random seed and provides a more robust evaluation.

Confidence Intervals: To enhance the clarity and reliability of our results, we have included the standard deviation calculated over five random seeds. The confidence intervals for the key metrics are provided in Table 6:

Comment 2.8

The OmniH20 NN architecture was not mentioned until the experimental section. This should be mentioned earlier.

Response:

Thank you for pointing out our missing details on policy architecture. The NN architecture is a 3-layer MLP (512, 256, 128) with tanh activation. For the sequential RNN networks (LSTM, GRU), it is a one-layer RNN with hidden dims 256. We have added them in the revised manuscript.

Table 6: Simulation motion imitation evaluation of OmniH2O and baselines on dataset. Note that all the variants are trained with exact same rewards, domain randomizations and motion dataset. The standard deviation is calculated over 5 random seeds.

			All sequences					Successful sequences			
Method	State Dimension	Sim2Real	Succ ↑	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel} ↓
Privileged policy	$\mathcal{S} \subset \mathcal{R}^{913}$	✗	94.77% ±0.32%	126.51 ±0.66	70.68 ±0.35	3.57 ±0.06	6.20 ±0.07	122.71 ±0.49	69.06 ±0.28	2.22 ±0.03	5.20 ±0.04
H2O	$\mathcal{S} \subset \mathcal{R}^{138}$	✓	87.52% ±0.19%	148.13 ±0.75	81.06 ±0.40	5.12 ±0.04	7.89 ±0.08	133.28 ±0.55	75.99 ±0.36	2.40 ±0.04	5.75 ±0.05
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	**141.11** ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	**135.49** ±0.68	**75.75** ±0.31	**2.30** ±0.02	**5.47** ±0.05
(a) Ablation on DAgger/RL											
OmniH2O-w/o-DAgger-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✗	90.62% ±0.25%	163.44 ±1.01	91.29 ±0.58	5.12 ±0.07	8.80 ±0.09	153.31 ±0.79	87.59 ±0.39	3.15 ±0.03	7.27 ±0.05
OmniH2O-w/o-DAgger	$\mathcal{S} \subset \mathcal{R}^{1665}$	✗	47.11% ±0.14%	223.27 ±1.47	128.90 ±0.81	15.03 ±0.09	16.29 ±0.12	182.13 ±1.22	119.54 ±0.69	5.47 ±0.04	9.10 ±0.08
OmniH2O-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	93.80% ±0.33%	141.21 ±0.73	78.52 ±0.39	3.74 ±0.03	6.62 ±0.08	**134.90** ±0.59	76.11 ±0.27	**2.25** ±0.02	5.48 ±0.04
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	**141.11** ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05
(b) Ablation on History steps/Architecture											
OmniH2O-History50	$\mathcal{S} \subset \mathcal{R}^{3240}$	✓	93.56% ±0.34%	141.51 ±0.78	78.51 ±0.41	4.01 ±0.05	6.79 ±0.06	135.04 ±0.63	76.07 ±0.33	2.36 ±0.02	5.55 ±0.05
OmniH2O-History5	$\mathcal{S} \subset \mathcal{R}^{405}$	✓	93.60% ±0.23%	**139.23** ±0.67	**77.82** ±0.35	3.91 ±0.04	6.66 ±0.07	132.67 ±0.55	75.33 ±0.26	2.24 ±0.02	5.41 ±0.04
OmniH2O-History0	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	93.80% ±0.33%	141.21 ±0.73	78.52 ±0.39	3.74 ±0.03	6.62 ±0.08	**134.90** ±0.59	76.11 ±0.27	**2.25** ±0.02	5.48 ±0.05
OmniH2O-GRU	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	92.85% ±0.22%	147.67 ±0.75	80.84 ±0.38	4.05 ±0.04	6.93 ±0.09	142.75 ±0.67	79.10 ±0.32	2.38 ±0.02	5.66 ±0.03
OmniH2O-LSTM	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	91.03% ±0.21%	147.36 ±0.74	80.34 ±0.41	4.12 ±0.06	7.04 ±0.11	142.64 ±0.66	78.59 ±0.34	2.37 ±0.01	5.72 ±0.05
OmniH2O-History25 (Ours)	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	141.11 ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05
(c) Ablation on Tracking Points											
OmniH2O-22points	$\mathcal{S} \subset \mathcal{R}^{1836}$	✓	**94.72%** ±0.35%	**127.71** ±0.54	**70.39** ±0.34	**3.62** ±0.04	**6.25** ±0.05	**123.87** ±0.45	**68.92** ±0.27	**2.22** ±0.02	**5.24** ±0.04
OmniH2O-8points	$\mathcal{S} \subset \mathcal{R}^{1710}$	✓	94.31% ±0.28%	129.30 ±0.63	71.70 ±0.36	3.78 ±0.03	6.39 ±0.06	125.14 ±0.54	70.07 ±0.32	2.22 ±0.02	5.26 ±0.03
OmniH2O-3points (Ours)	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	94.10% ±0.35%	141.11 ±0.77	77.82 ±0.45	3.70 ±0.03	6.54 ±0.06	135.49 ±0.68	75.75 ±0.31	2.30 ±0.02	5.47 ±0.05
(d) Ablation on Linear Velocity											
OmniH2O-w-linvel	$\mathcal{S} \subset \mathcal{R}^{1743}$	✓	93.80% ±0.33%	**138.18** ±0.71	78.12 ±0.38	3.94 ±0.05	6.61 ±0.07	**132.44** ±0.54	75.98 ±0.30	**2.29** ±0.02	**5.40** ±0.03
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%** ±0.35%	141.11 ±0.77	**77.82** ±0.45	**3.70** ±0.03	**6.54** ±0.06	135.49 ±0.68	**75.75** ±0.31	2.30 ±0.02	**5.47** ±0.05

Comment 2.9

I did not understand the "ablation to sparse input" section. Does your OmniH20 architecture allow inputs with different types of sparsity or do you need to train a student for each sparsity you consider. I tried to figure this out from the paper text, but couldn't piece things together.

Response:

Thank you for your question. For the "ablation to sparse input" section, we are trying to figure out "how much does sparse input influence performance". Indeed our OmniH2O architecture requires training a separate student policy for each type of input sparsity. We use the DAgger approach, which makes this training efficient and manageable. Additionally, all policies trained with DAgger are compatible with sim2real since the teacher policy is designed for sim2real, and the student policies inherits this capability.

Comment 2.10

How were the real-world experiments conducted? Was there just one trial of each motion? Why isn't there a success rate here?

Response:

For the real-world experiments, we conducted one trial for each of the 20 different motions. Instead of calculating a success rate, we opted to test all motions as long as the robot did not fall. This approach provided a practical assessment of the robot's stability and performance across a variety of tasks. The real-world test videos are attached for your reference https://anonymous-omni-h2o.github.io/resources/rebuttal/OmniH2O_vs_H2O_short.mp4.

Comment 2.11

How does Figure 5 show "superior robustness"? Compared to what? There is nothing quantitative here upon which to draw a conclusion.

Response:

Thank you for highlighting the need for clarity in our claims regarding robustness and terrain. We have revised the manuscript to avoid overstatements and to better contextualize our findings.

OmniH2O advances H2O's robustness in the following crucial ways: (1) OmniH2O demonstrates more real-world movement types than H2O, and achieves higher success rate in motion imitation; (2) OmniH2O no longer relies on Motion Capture (MoCap) devices and demonstrates outdoor capabilities.

When compared to other cited works such as https://arxiv.org/abs/2402.16796 and https://arxiv.org/pdf/2303.03381, OmniH2O demonstrates the unique capability to maintain stability and return to a standing position after disturbances. These prior works tend to trade off the ability to stand still for enhanced robustness. The ability to remain stable standing is vital for tasks requiring precise loco-manipulation, especially in scenarios involving dexterous humanoid manipulation. We hope that this clarification provides a better understanding of our contributions and the specific contexts in which our approach excels.

3 Response to Reviewer huqu

Comment 3.1

Summary Of Contribution:
This paper introduces a bi-level autonomous humanoid policy that can be commanded by vocal instructions. The low-level policy can be operated by human operator by teleoperation. The high-level policy can be deployed using GPT-4o, MGM, or a trained diffusion policy. The author demonstrates the training framework of the low-level policy and releases the dataset OmniH2O-6 dataset to train the high-level diffusion policy.

Summary Of Strengths:
This paper describes the bi-level framework of humanoid policy that can be commanded by vocal instructions. The writing is clear and the technical details are well-organized. The author provides a clear motivation. The author provides a clear motivation and well-organized technical details. The author provides a clear description of the training framework of the teleoperation policy. Clear illustration on using GPT-4o for the full-autonomous tasks.

Response:

We are thankful for the reviewer's feedback and recommendation. We highly appreciate your recognition. We have carefully reviewed your comments and responded to the concerns raised.

Comment 3.2

The research advancement is not clear compared with previous works.

Response:

Thank you for your feedback. Our **key contribution** lies in the development of the OmniH2O humanoid control pipeline and data collection framework. The primary innovation here is not just the application of Learning from Demonstration (LfD), but the creation of a robust system that makes it feasible to apply LfD effectively to humanoid robots—something that has been straightforward in table-top manipulation but remains exceedingly challenging for complex, full-body humanoid tasks.

We have made several advancements compared to previous works, which we detail below:

We provided a clear and detailed comparison between OmniH2O and prior work H2O in Comment 1.2, and ablations on the source of performance improvement in Comment 1.3. Furthermore, we offer a comprehensive comparison of OmniH2O with previous works (H2O, ExBody) in the table below in Table 7:

Table 7: Comparison of OmniH2O with Prior Works

Feature/Metric	OmniH2O	H2O	ExBody
Whole-Body Loco-Manipulation	Yes	Limited	Limited
Dexterous Manipulation	Yes	No	No
Teleoperation	VR, RGB, Verbal	RGB	N/A
Robustness to Sparse Inputs	High	Low	Low
Stability During Manipulation	High	Low	Low
Dataset Contribution	OmniH2O-6 (First Humanoid Loco-Manipulation Dataset)	None	None
Imitation Learning from Dataset	Yes	No	No
Reward Design for RL	Regularization, Curriculum	Regularization	Regularization
History Utilization in State Space	Yes (25 Steps)	No	No
Deployment Feasibility	High	Low (MoCap Dependent)	High
Integration with VLMs	Yes	No	No

Our contributions are significant and multifaceted:

(1) We propose a robust pipeline for training a humanoid control policy that supports whole-body dexterous loco-manipulation with a universal interface enabling versatile human control and autonomy.

(2) Through extensive large-scale motion tracking experiments in both simulation and the real world, we validate the superior motion imitation capability of OmniH2O.

(3) We contribute the first-ever humanoid loco-manipulation dataset (OmniH2O-6) and demonstrate the effectiveness of imitation learning methods on this dataset, showcasing the ability to learn whole-body skills.

These advancements collectively demonstrate significant progress in the field and set a new benchmark for future research.

Comment 3.3

Details of learning from demonstrations are not sufficient in the manuscript.

Response:

Thank you for pointing out the missing details of learning from demonstration. We have added the following the details in the revised manuscript.

Our system learns desired motion goals from demonstration data by treating motion goal prediction as a conditional generation problem. This process is as follows:

Learning from Demonstrations: We use a denoising diffusion probabilistic model (DDPM) or Denoising Diffusion Implicit Models (DDIM) to predict action sequences based on input observations. With an observation horizon of 1 or 4, we predict the action sequence of the corresponding length from the current input observations. Each input observation is a collection of data from proprioception and vision. Each action is a 23-dimensional vector specifying the desired 3D positions for the two hands (2x6 dimensions), the two wrists (2x1 dimension), and three motion goals (3x3 dimensions).

Optimization Details: We use the AdamW optimizer with a learning rate of 0.0001, weight decay of 0.00001, and a batch size of 128. Following the diffusion policy, we maintain an exponential weighted average of the model weights, which is used during evaluation and deployment.

Diffusion Architecture: All encoded images and proprioceptive observations are concatenated and fed into a CNN-based diffusion model. We utilize a square cosine noise schedule and 100 diffusion steps for training. The model outputs the normalized desired motion goals (head, hands) for the low-level whole-body control policy.

Proprioception: We use joint angles of humanoid motors and dexterous hands as proprioception observations. The proprioception data is processed through a two-layer network with ReLU activation, a hidden size of 256, and an output feature size of 64.

By incorporating these detailed methodologies, our system effectively learns and predicts motion goals from demonstration data, ensuring robust and reliable performance. These enhancements are now clearly documented in the revised manuscript.

Comment 3.4

Lack of thorough investigation of the performance of different methods in learning from demonstrations.

Response:

Thank you for your comment. Our primary goal in this work is to demonstrate the feasibility of humanoid learning from demonstrations (LfD). We have conducted a series of tests to validate our approach, including

1. **Method Evaluation**: We have evaluated several methods, including Vanilla Behavior Cloning (BC), DDIM Diffusion Policy, and DDPM Diffusion Policy.

2. **Data Size Analysis**: We tested different data sizes (25%, 50%, 100%) to assess the impact of the amount of training data on performance.

3. **Sequence of Observations and Actions**: We also investigated the sequence of observations and actions to understand their effect on the learning process.

The detailed performance of 4 tasks is documented in Table 8

Table 8: Quantitative LfD autonomous agents performance for 4 tasks.

Metrics	Catch-Release			Squat			Hammer-Catch			Rock-Paper-Scissors		
(a) Ablation on Data size												
	25%data	50%data	100%data	25%data	50%data	100%data	25%data	50%data	100%data	25%data	50%data	100%data
MSE Loss	3.01E-3	3.04E-4	9.89E-5	1.25E-4	1.10E-4	7.07E-5	2.18E-2	1.56E-2	3.29E-4	2.72E-2	1.39E-2	1.60E-3
Succ rate	1/10	3/10	6/10	9/10	10/10	10/10	3/10	6/10	6/10	3/10	9/10	10/10
(b) Ablation on Sequence observation/action												
	Si-O-Si-A	Se-O-Se-A	Si-O-Se-A	Si-O-Si-A	Se-O-Se-A	Si-O-Se-A	Si-O-Si-A	Se-O-Se-A	Si-O-Se-A	Si-O-Si-A	Se-O-Se-A	Si-O-Se-A
MSE Loss	2.52E-4	1.47E-4	9.89E-5	5.18E-5	9.60E-5	7.07E-5	2.22E-4	3.62E-4	3.29E-4	1.43E-3	3.36E-3	1.60E-3
Succ rate	3/10	7/10	6/10	10/10	10/10	10/10	5/10	9/10	6/10	10/10	9/10	10/10
(c) Ablation on BC/DDIM/DDPM												
	BC	DP-DDIM	DP-DDPM	BC	DP-DDIM	DP-DDPM	BC	DP-DDIM	DP-DDPM	BC	DP-DDIM	DP-DDPM
MSE Loss	1.39E-3	4.79E-5	9.89E-5	6.24E-4	6.42E-5	7.07E-5	4.50E-3	3.41E-4	3.29E-4	1.46E-2	2.42E-3	1.60E-3
Succ rate	0/10	6/10	6/10	3/10	10/10	10/10	0/10	5/10	6/10	1/10	10/10	10/10

Our main contribution lies in developing a robust training and deployment system capable of stable whole-body locomanipulation. This system ensures the robot's performance is reliable across various tasks and conditions.

We acknowledge that a more comprehensive investigation of different methods in LfD would provide deeper insights. However, our focus is on establishing a solid foundation for stable and robust humanoid control. We hope that future research can build upon our work and explore these methods in greater detail, leading to further advancements in the field.

Comment 3.5

Lack of details on the OmniH2O-6 dataset. For example, how is the lower-limb motion captured when collecting the dataset using VR? In the training section of the teleoperation policy, the policy is designed to follow the lower-limbs motion.

Response:

Thank you for pointing this out. The lower-limb motion is not (and could not be) captured using the VR headset directly. To address this challenge of **motion goal sparsity**, we leverage a teacher-student distillation pipeline. Here's how it works:

1. **Teacher Policy Training:** We first train a teacher policy that has access to privileged information and the full whole-body motion goals. This includes detailed motion data of both the upper and lower limbs.

2. **Student Policy Distillation:** We then distill the teacher policy into a student policy, which only has access to partial observations and the 3-point motion goals (head and hands) provided by the VR headset. The student policy learns to approximate the teacher's behavior using this limited input.

The key here is that teacher policy can leverage comprehensive motion data, including lower-limb movements, to learn a better control strategy. This strategy is then simplified and adapted for the student policy, which operates under the constraints of sparse input data from the VR system.

Our experiments showed that using RL alone to handle sparse motion goals is challenging. Therefore, we incorporate DAgger into our pipeline to iteratively refine the student policy by incorporating feedback from the teacher. This combined RL-DAgger approach significantly improves the performance and robustness of the teleoperation policy.

This method allows us to overcome the limitations of VR-based data collection and ensures that the policy can effectively follow lower-limb motions, even though they are not directly captured during data collection.

For **more details** on the OmniH2O-6 dataset:

Here are six datasets corresponding to the six Learning from Demonstration (LfD) tasks listed in our paper, with the following correspondence relationships:

- `basket_package_file.pkl` → Basket-Pick-Place, which has 18,436 frames in total.
- `boxing_package_file.pkl` → Boxing, which has 11,118 frames in total.
- `hammer_package_file.pkl` → Hammer-Catch, which has 12,759 frames in total.
- `rps_package_file.pkl` → Rock-Paper-Scissors, which has 9,380 frames in total.
- `squat_package_file.pkl` → Squat, which has 8,535 frames in total.
- `trash_package_file.pkl` → Catch-Release, which has 13,234 frames in total.

Our dataset comprises several components including RGB images, root positions, root quaternions, motion goal positions, motion goal velocities, depth images, wrist actions, and finger actions. Below is a Python code snippet demonstrating how to utilize this dataset:

```python
import pickle

# Load dataset from the pkl file
with open('pkl_file_you_want', 'rb') as file:
    data = pickle.load(file)

# Process the data into different parts
rgb_image_set = []
root_position_set = []
root_quaternion_set = []
motion_goal_position_set = []
motion_goal_velocity_set = []
depth_image_set = []
wrist_action_set = []
finger_action_set = []

for i in range(len(data)):
    rgb_image_set.append(data[i]['img'])  # data[i]['img'] is a numpy array shaped (240, 424, 3)
    root_position_set.append(data[i]['root_pos'])  # data[i]['root_pos'] is a numpy array shaped (3,)  # (x, y, z)
    root_quaternion_set.append(data[i]['root_quat'])  # data[i]['root_quat'] is a numpy array shaped (4,)  # (x, y, z, w)
    motion_goal_position_set.append(data[i]['ref_body_pos'])  # data[i]['ref_body_pos'] is a numpy array shaped (9,)
    motion_goal_velocity_set.append(data[i]['ref_body_vel'])  # data[i]['ref_body_vel'] is a numpy array shaped (9,)
    depth_image_set.append(data[i]['depth'])  # data[i]['depth'] is a numpy array shaped (240, 424)
    wrist_action_set.append(data[i]['wrist'])  # data[i]['wrist'] is a numpy array shaped (2,)  # left wrist, right wrist
    finger_action_set.append(data[i]['finger'])  # data[i]['finger'] is a numpy array shaped (12,)  # left hand, right hand, each hand has 6 DoF where
        thumb has two DoF. (little finger, ring finger, middle finger, index finger, thumb1, thumb2)
```

This detailed explanation and the code snippet should help clarify how the lower-limb motions are indirectly managed through our teacher-student training pipeline and how to effectively use the OmniH2O-6 dataset.

Comment 3.6

What is the reference trajectory when performing outdoor locomotion tests?

Response:

Thank you for pointing this out! During outdoor locomotion tests, we set a static target tracking pose 0.5m ahead of the humanoid in the robot local frame to walk forward and 0.5 m behind to walk backward. For turning left or right, we position three tracking points 0.3m to the left or right in the robot's local frame. This approach allows the humanoid to perform outdoor locomotion using joystick commands, demonstrating the versatility of the OmniH2O policy. It can track dynamic motions from datasets or real-time teleoperation devices and also track static kinematic poses for robust locomotion.

Comment 3.7

Why distillation is needed for teleoperation policy?

Response:

Distillation is necessary when the motion goal is **sparse** (only tracking head and hands) and the humanoid observation is **partially observable** (no linear velocity observation in the real world, thus necessitating the utilization of historical data). This necessity is substantiated by the results in Table 9, where OmniH2O-w/o-DAgger achieved a success rate of only 47.11% compared to the DAgger version's 94.10%.

The challenges of **input sparsity** and **partial observability** are significant. To make the policy compatible with a more precise and robust VR teleoperation interface, we reduced the tracking points from 8 to **3 tracking points**. Furthermore, we **eliminated the requirement for linear velocity observation** to address the real-world state estimation challenge. However, these design choices introduce additional complexities for the reinforcement learning (RL) training process, as the motion goals become much sparser and the observation space is partially observable.

To overcome these challenges, we leverage a **teacher-student training pipeline** and incorporate **historical observations**. This approach allows us to effectively address the limitations brought by 3-point tracking and the absence of linear velocity in the state space. The results in Table 9 clearly demonstrate the effectiveness of this method. For instance, the privileged policy with a full state dimension ($\mathcal{S} \subset \mathcal{R}^{913}$) and no sim2real achieved a success rate of 94.77%, with a significant reduction in errors compared to other methods. When we apply the proposed method (OmniH2O) with distillation and sim2real, it achieved a success rate of 94.10%, showcasing its robustness and effectiveness.

In summary, distillation plays a crucial role in enhancing teleoperation policy by addressing the challenges of input sparsity and partial observability, thereby significantly improving performance and success rates.

Table 9: Simulation motion imitation evaluation of OmniH2O and OmniH2O without distillaion (DAgger). Note that all the variants are trained with exact same rewards, domain randomizations and motion dataset $\hat{\boldsymbol{Q}}$.

			All sequences					Successful sequences			
Method	State Dimension	Sim2Real	Succ $\uparrow$	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$\text{E}_{\text{acc}} \downarrow$	$\text{E}_{\text{vel}} \downarrow$	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	$\text{E}_{\text{acc}} \downarrow$	$\text{E}_{\text{vel}} \downarrow$
(a) Ablation on DAgger/RL											
OmniH2O-w/o-DAgger	$\mathcal{S} \subset \mathcal{R}^{1665}$	✗	47.11%	223.27	128.90	15.03	16.29	182.13	119.54	5.47	9.10
OmniH2O	$\mathcal{S} \subset \mathcal{R}^{1665}$	✓	**94.10%**	**141.11**	**77.82**	**3.70**	**6.54**	**135.49**	**75.75**	**2.30**	**5.47**

Comment 3.8

Is it possible to use a more accurate odometry system to provide the linear velocity of the robot to achieve a better performance? For example, using an in-door tracking system for a proof of concept.

Response:

Thank you for your question. Using a more accurate odometry system, such as an indoor tracking system, is indeed possible and can provide the linear velocity of the robot to potentially achieve better performance. This approach is similar to the setting used in the H2O framework.

However, our aim is to develop a humanoid robot that can operate effectively in a variety of real-world environments, not just controlled indoor settings. Therefore, we seek solutions that are robust and adaptable to different conditions.

To explore the feasibility of better odometry systems, we conducted tests of OmniH2O-w/o-linvel using the iPhone Visual Inertial Odometry (VIO) system. As shown in Figure 4, though the results are better than ZED VIO, it still much worse than OmniH2O which does not include linear velocity, showing the challenge of state estimation for humanoids.

Figure 4: OmniH2O-w/o-linvel policy fails fatally due the inaccuracy of real-world VIO (iPhone VIO module)

In conclusion, while indoor tracking systems can serve as an effective proof of concept, our focus remains on developing a versatile solution that performs well in diverse environments. We are definitely interested in including velocity from indoor tracking systems for a proof of concept when our MoCap system is fixed.