Learning Human-to-Humanoid Real-Time Whole-Body Teleoperation

Tairan He† Zhengyi Luo† Wenli Xiao Chong Zhang Kris Kitani Changliu Liu Guanya Shi
Carnegie Mellon University
†Equal Contributions https://human2humanoid.com

Fig. 1: The humanoid robot is teleoperated in real-time using an RGB camera by the human teleoperator. (a) The humanoid mimics the human teleoperator, advancing one step while delivering a punch to displace a box, followed by a victory gesture. (b) The humanoid executes a precise sidestep to align with a ball and delivers a controlled kick using its right foot. (c) The humanoid demonstrates forward walking while pushing a stroller. (d) The operator teleoperates the humanoid to catch a box, rotate its waist, and drop the box into a waste bin. **Videos:** see the website.

***Abstract*— We present Human to Humanoid (H2O), a reinforcement learning (RL) based framework that enables real-time whole-body teleoperation of a full-sized humanoid robot with only an RGB camera. To create a large-scale retargeted motion dataset of human movements for humanoid robots, we propose a scalable "sim-to-data" process to filter and pick feasible motions using a privileged motion imitator. Afterwards, we train a robust real-time humanoid motion imitator in simulation using these refined motions and transfer it to the real humanoid robot in a zero-shot manner. We successfully achieve teleoperation of dynamic whole-body motions in real-world scenarios, including walking, back jumping, kicking, turning, waving, pushing, boxing, *etc*. To the best of our knowledge, this is the first demonstration to achieve learning-based real-time whole-body humanoid teleoperation.**

I. Introduction

We aim to enable real-time teleoperation of a full-sized humanoid robot by a human teleoperator using an RGB camera. Humanoid robots, with their physical form closely mirroring that of humans, present an unparalleled opportunity for real-time teleoperation. This alignment of the embodiment allows for a seamless integration of human cognitive skills with versatile humanoid capabilities [1]. Such synergy stimulated by human-to-humanoid teleoperation is crucial for complex tasks (*e.g.*, household chores, medical assistance, high-risk rescue operations) that are yet too challenging for a fully autonomous robot, but possible for existing hardware teleoperated by humans [2, 3]. In this paper, we transfer human motions to humanoid behaviors in a real-time fashion using an RGB camera. This system also has the potential to enable large-scale and high-quality data collection of human operations for robotics [2, 4], where human-teleoperated actions can be used for imitation learning.

However, whole-body control of full-sized humanoids is a long-standing problem in robotics [5], and complexity

increases when controlling the humanoid to replicate free-form human movements in real-time [1]. Existing work on whole-body humanoid teleoperation has achieved remarkable results via model-based controllers [6–9], but they all use simplified models due to the high computational cost of modeling the full dynamics of the system [10, 11], which limits the scalability to dynamic motions. Furthermore, these works are highly dependent on contact measurement [12, 13], leading to reliance on external setups such as the exoskeleton [9] and force sensors [8, 14] for teleoperation.

Recent advances in reinforcement learning (RL) for humanoid control provide a promising alternative. First, in the graphics community, RL has been used to generate complex human movements [15, 16], perform a variety of tasks [17], and track real-time human motions captured by a webcam [18] in simulation. However, due to unrealistic state-space design and partial disregard of the hardware limit (*e.g.* torque / joint limit), it remains a question whether these methods can be applied to a full-sized humanoid. On the other hand, RL has achieved robust and agile biped locomotion in the real world [19–21]. To date, however, there has been no existing work on RL-based whole-body humanoid teleoperation. The most closely related effort is a concurrent study [22], which focuses on learning to replicate upper-body motions and uses root velocity tracking for the lower body, from offline human motions rather than real-time teleoperation.

In this paper, we design a complete system for humanoid teleportation in real time. First, we identify one of the primary challenges in whole-body humanoid teleoperation as the lack of a dataset with feasible motions tailored to the humanoid, which is essential for training a controller that can track diverse motions. Although direct human-to-humanoid retargeting has been explored in previous locomotion-focused efforts [23–25], retargeting a large-scale human motion dataset to the humanoid presents new challenges. That is, the significant dynamics discrepancy between humans and humanoids means that some human motions could be infeasible for the humanoid (*e.g.* cartwheeling, steps wider than the leg lengths of the humanoid). In light of this, we introduce an automated "sim-to-data" process to retarget and refine a large-scale human motion dataset [26] into motions that are feasible for real-world humanoid embodiment. Specifically, we first retarget the human motions to the humanoid via inverse kinematics, and train a humanoid controller with access to privileged state information [18] to imitate the unfiltered motions in simulation. Afterwards, we remove the motion sequences that the privileged imitator fails to track. By doing so, we create a large-scale humanoid-compatible motion dataset.

After obtaining a dataset of feasible motions, we develop a scalable training process for the real-world motion imitator that incorporates extensive domain randomization to bridge the sim-to-real gap. To facilitate real-time teleoperation, we design a state space that prioritizes the inputs available in the real world using an RGB camera, such as the keypoint positions. During inference, we use an off-the-shelf human pose estimator [27] to provide global human body positions for the humanoid to track.

In summary, we demonstrate the feasibility of an RL-based real-time Human-to-Humanoid (H2O) teleoperation system. Our contributions include:

1) A scalable retargeting and "sim-to-data" process to obtain a large-scale motion dataset feasible for the real-world humanoid robot;
2) Sim-to-real transfer of the RL-based whole-body tracking controller that scales to a large number of motions;
3) A real-time teleoperation system with an RGB camera and 3D human pose estimation, demonstrating fulfillment of various whole-body motions including walking, pick-and-place, stroller pushing, boxing, hand-waving, ball kicking, etc.

II. Related Works

A. Physics-Based Animation of Human Motions

Physics-based simulation has been used to generate realistic and natural motions for avatars [15–18, 28–33]. With motion capture as the main source of human motion data [26], RL is often used to learn avatar controllers that can mimic these motions, offering distinctive styles [15, 31], scalability [16, 18, 34], and reusability [17, 32].

However, realistic animation in physics-based simulators does not guarantee real-world applicability, especially for humanoids. Simulated humanoid avatars typically have high degrees of freedom and large joint torques [35], and sometimes need non-physical assistive external forces [36]. In this work, we demonstrate that, with carefully designed sim-to-real training, approaches in the humanoid animation community can be applied to a real-world humanoid robot.

B. Transferring Human Motions to Real-World Humanoids

Before the emergence of RL-based humanoid controllers, traditional methods typically employ model-based optimization to track retargeted motions while maintaining stability [1]. To this end, these methods minimize tracking errors under the constraints of stability and contacts, requiring predefined contact states [6, 10, 12, 13, 37–40] or estimated contacts from sensors [7, 9, 14, 41, 42], hindering large-scale deployment outside the laboratory. Zhang *et al.* [11] use contact-implicit model predictive control (MPC) to track motions extracted from videos, but trajectories must first be optimized offline to ensure dynamic feasibility. Furthermore, the model used in MPC needs to be simplified due to computational burden [6, 11, 14], which limits the capability of trackable motions.

RL-based controllers may provide an alternative that does not require explicit contact information. Some works [43, 44] use imitation learning to transfer human-style motions to the controller, but do not accurately track human motions. Cheng *et al.* [22] train whole-body humanoid controllers that can replicate upper body movements from offline human motions, but the lower body relies on root velocity tracking and does not track precise lower body movements. In comparison, our work achieves real-time whole-body tracking of human motions.

C. Teleoperation of Humanoids

Teleoperation of humanoids can be categorized into three types: 1) task-space teleoperation [45, 46], 2) upper-body-retargeted teleoperation [47, 48], and 3) whole-body teleoperation [6, 7, 13, 42, 49]. For the first and second types, the shared morphology between humans and humanoids is not fully utilized, and whole-body control must be solved in a task-specified way. This also raises the concern that, if tracking lower body movement is not necessary, the robot could opt for designs with better stability, such as a quadruped [50] or wheeled configuration [51].

Our work belongs to the third type and is the first to achieve learning-based whole-body teleoperation. Moreover, our approach does not require capture markers or force sensors on the human teleoperator, as we directly employ an RGB camera to capture human motions for tracking, potentially paving the way for collecting large-scale humanoid data for training autonomous agents.

III. Preliminaries

The whole-body real-time humanoid teleoperation we are tackling is formulated as a goal-conditioned RL problem, which tracks versatile human motion with a single RL control policy. In Section III-A, we set up the preliminary for our control framework. Section III-B describes the human model and dataset we use in the RL policy training. As a notation convention, we use $\tilde{\cdot}$ to represent kinematic quantities (without physics simulation) from pose estimator/keypoint detectors, $\hat{\cdot}$ to denote ground truth quantities from Motion Capture (MoCap), and normal symbols without accents for values from the physics simulation.

A. Goal-conditioned RL for Humanoid Control

We formulate our problem as goal-conditioned RL where π is trained to track real-time human motions. We formulate the learning task as a Markov Decision Process (MDP) defined by the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$ of state $\boldsymbol{s}_t \in \mathcal{S}$, action $\boldsymbol{a}_t \in \mathcal{A}$, transition dynamics $\mathcal{T}$, reward function $\mathcal{R}$, and discount factor γ. The state $\boldsymbol{s}_t$ contains the proprioception $\boldsymbol{s}_t^{\text{p}}$ and the goal state $\boldsymbol{s}_t^{\text{g}}$. The goal state $\boldsymbol{s}_t^{\text{g}}$ is a unified representation of the whole-body motion of the human teleoperator, which we will discuss in detail in Section V-A. Based on proprioception $\boldsymbol{s}_t^{\text{p}}$ and goal state $\boldsymbol{s}_t^{\text{g}}$, we define the reward $r_t = \mathcal{R}\left(\boldsymbol{s}_t^{\text{p}}, \boldsymbol{s}_t^{\text{g}}\right)$ for the policy training. The action $\boldsymbol{a}_t \in \mathbb{R}^{19}$ specifies the joint target positions that the PD controller will use to actuate the degrees of freedom. We apply the proximal policy gradient (PPO) [52] to maximize the cumulative discounted reward $\mathbb{E}\left[\sum_{t=1}^{T} \gamma^{t-1} r_t\right]$. We formulate the teleoperation task as the motion imitation/tracking/mimicking task, where we train the humanoid to track the reference motion at every frame.

B. Parametric Human Model and Human Motion Dataset

Popular in the vision and graphics community, parametric human models such as SMPL [53] are easy to work with representations of human shapes and motions. SMPL represents the human body as body shape parameters $\boldsymbol{\beta} \in \mathcal{R}^{10}$, pose parameters $\boldsymbol{\theta} \in \mathcal{R}^{24\times3}$, and root translation $\boldsymbol{p} \in \mathcal{R}^{24\times3}$. Given $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{p}$, $\mathcal{S}$ denotes the SMPL function, where $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{p}) : \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{p} \rightarrow \mathcal{R}^{6890\times3}$ maps the parameters to the position of the vertices of a triangular human mesh of 6890 vertices. The AMASS [26] dataset contains 40 hours of motion capture expressed in the SMPL parameters.

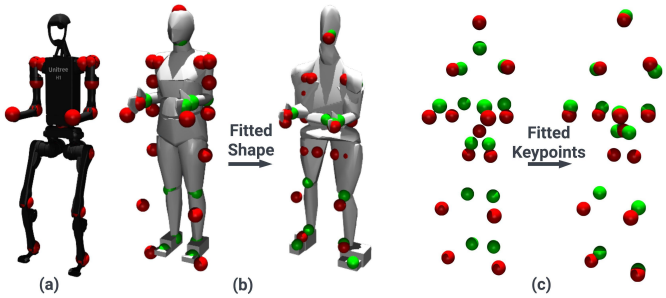

Fig. 2: Fitting the SMPL body to the H1 humanoid. (a) Visualization of the humanoid keypoints (red dots) (b) Humanoid keypoints vs SMPL keypoints (green dots and mesh) before and after fitted SMPL shape $\boldsymbol{\beta}'$. (c) Corresponding 12 joint position before and after fitting.

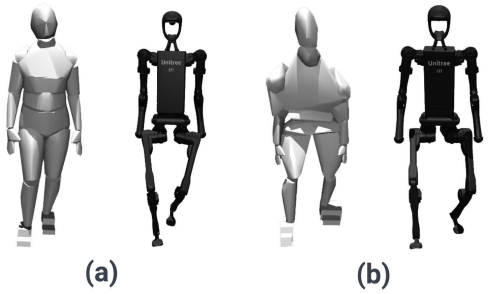

Fig. 3: Effect of using a fitted SMPL shape $\boldsymbol{\beta}'$ instead of mean body shape on position-based retargeting. (a) Retargting without using $\boldsymbol{\beta}'$, which results in unstable "in-toed" humanoid motion. (b) Retargeting using $\boldsymbol{\beta}'$, which result in balanced humanoid motion.

IV. Retargeting Human Motions for Humanoid

To enable humanoid motion imitation for unscripted human motion, we require a large amount of whole-body motion to train a robust motion imitation policy. Since humans and humanoids also have a nontrivial difference in body structure, shape, and dynamics, naively retargeted motion from a human motion dataset can result in a large number of motions impossible for our humanoid to perform. These imfeasible motion sequences can hinder imitation training as observed in prior work [32]. To resolve these issues, we design a "sim-to-data" approach to complement traditional retargeting to convert a large-scale human motion dataset to feasible motions for humanoids.

A. Motion Retargeting

As there is a non-trivial difference between the SMPL kinematic structure and the humanoid kinematic tree, we perform a two-step process for the initial retargeting. First, since the SMPL body model can represent different body

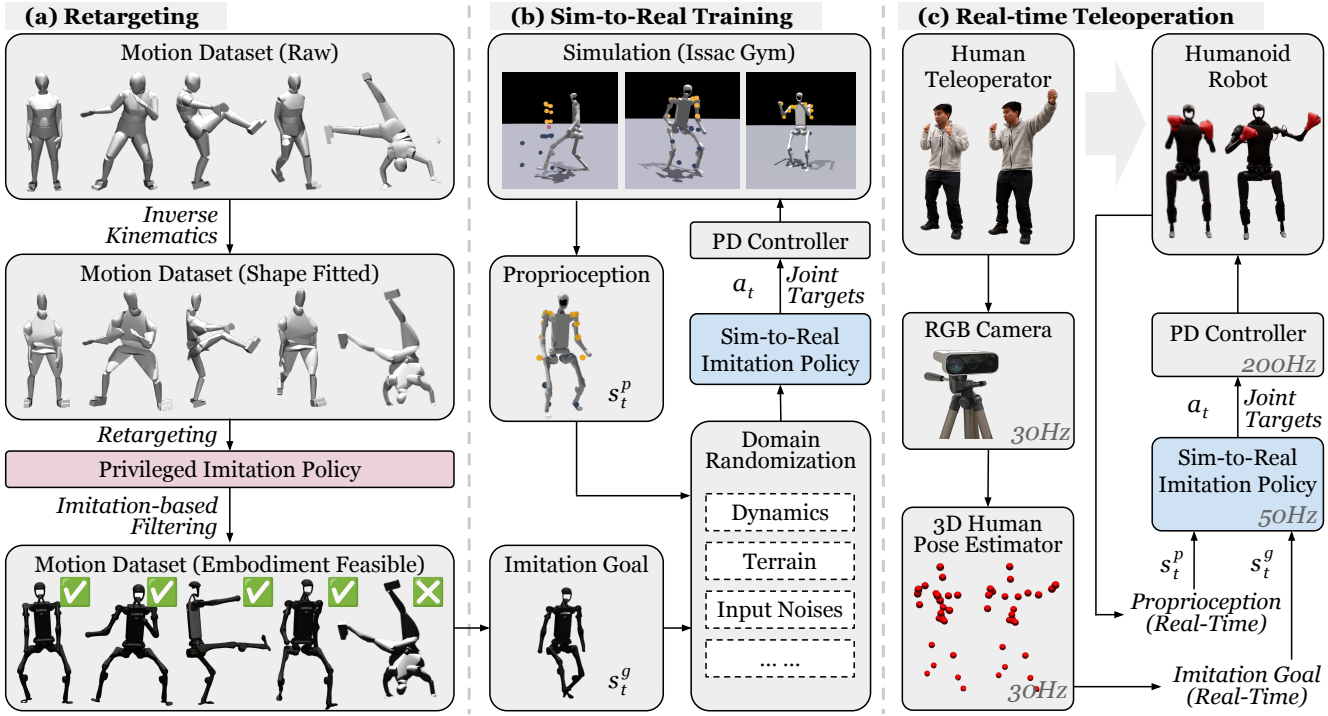

Fig. 4: Overview of H2O: (a) **Retargeting** (Section IV): H2O first aligns the SMPL body model to a humanoid's structure by optimizing shape parameters. Then H2O retargets and removes the infeasible motions using a trained privileged imitation policy, producing a clean motion dataset. (b) **Sim-to-Real Training**: (Section V) An imitation policy is trained to track motion goals sampled from a cleaned dataset. (c) **Real-time Teleoperation Deployment** (Section VI-B): The real-time teleoperation deployment captures human motion through an RGB camera and a pose estimator, which is then mimicked by a humanoid robot using the trained sim-to-real imitation policy.

proportions, we first find a body shape β' closest to the humanoid structure. We choose 12 joints that have a correspondence between humans and humanoids, as shown in Fig.2 and perform gradient descents on the shape parameter s to minimize the joint distances using a common rest pose. After finding the optimal β', given a sequence of motions expressed in SMPL parameters, we use the original translation p and pose θ, but the fitted shape β' to obtain the set of body keypoint positions. Then we retarget motion from human to humanoid by minimizing the 12 joint position differences using Adam optimizer [54]. Notice that our retargeting process try to match the end effectors of the human to the humanoid (*e.g.* ankles, elbows, wrists) to preserve the overall motion pattern. Another approach is direct copying the local joint angles from human to humanoid, but that approach can lead to large differences in end-effector positions due to the large difference in kinematic trees. During this process, we also add some heuristic-based filtering to remove unsafe sequences, such as sitting on the ground. The motivation to find β' before retargeting is that in the rest pose, our humanoid has a large gap between its feet. If naively trying to match the foot movement between the human and the humanoid, the humanoid motion can have an in-toed artifact. Using β', we can find a human body structure has a large gap between its rest pose (as shown in Fig.2). Using β' during fitting can effectively create motion that is more feasible for the humanoid, as shown in Fig.3. From the AMASS dataset $\hat{Q}$ that contains 13k motion sequences, this process computes 10k retargted motion sequences $\hat{Q}^{\text{retarget}}$.

B. Simulation-based Data Cleaning

As shown in Figure 4, $\hat{Q}^{\text{retarget}}$ contain a large number of implausible motions for the humanoid due to the significant gap between the capabilities of a human and a motor-actuated humanoid. Manually finding these data sequences from a large-scale dataset can be a rather cumbersome process. Thus, we propose a "sim-to-data" procedure, where we train a motion imitator $\pi_{\text{privileged}}$ (similar to PHC [18]) with access to privileged information and no domain randomization to imitate all uncleaned data $\hat{Q}^{\text{retarget}}$. Without domain randomization, $\pi_{\text{privileged}}$ can perform well in motion imitation, but is not suitable for transfer to the real humanoid. However, $\pi_{\text{privileged}}$ represents the upper bound of motion imitation performance, and sequences which $\pi_{\text{privileged}}$ fails to imitate represent implausible ones. Specifically, we train $\pi_{\text{privileged}}$ following the same state space, control parameters, and hard-negative mining procedure proposed in PULSE [32], and train a single imitation policy to imitate the entire retargeted dataset. After training, ∼8.5k out of 10k motion sequences from AMASS turn out to be plausible for the H1 humanoid, and we denote the obtained clean dataset as $\hat{Q}^{\text{clean}}$.

Privileged Motion Imitation Policy. To train $\pi_{\text{privileged}}$, we follow PULSE [32] and train a motion imitator with access

to the full rigid body state of the humanoid. Specifically, for the privileged policy $\pi_{\text{privileged}}$, its proprioception is defined as $\boldsymbol{s}_t^{\text{p-privileged}} \triangleq [\boldsymbol{p}_t, \boldsymbol{\theta}_t, \boldsymbol{v}_t, \boldsymbol{\omega}_t]$, which contains the global 3D rigid body position $\boldsymbol{p}_t$, orientation $\boldsymbol{\theta}_t$, linear velocity $\boldsymbol{v}_t$, and angular velocity $\boldsymbol{\omega}_t$ of all rigid bodies in the humanoid. The goal state is defined as $\boldsymbol{s}_t^{\text{g-privileged}} \triangleq [\hat{\boldsymbol{\theta}}_{t+1} \ominus \boldsymbol{\theta}_t, \hat{\boldsymbol{p}}_{t+1} - \boldsymbol{p}_t, \hat{\boldsymbol{v}}_{t+1} - \boldsymbol{v}_t, \hat{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_t, \hat{\boldsymbol{\theta}}_{t+1}, \hat{\boldsymbol{p}}_{t+1}]$, which contains the one-frame difference between the reference and current simulation result for all rigid bodies on the humanoid. It also contains the next frame's reference rigid body orientation and position. All values are normalized to the humanoid's coordinate system. Notice that all values are global, and values such as global rigidbody linear velocity $\boldsymbol{v}_t$ and angular velocity $\boldsymbol{\omega}_t$ are hard to obtain accurately in the real world.

V. Whole-Body Teleoperation Policy Training

A. State Space

To achieve real-time teleoperation of humanoid robots, the state space of RL policy must contain only quantities available in the real world. This differs from the simulation-only approaches, where all the physics information (*e.g.*, foot contact force) is available. For example, in the real world, we have no access to each joint's precise global angular velocity due to the lack of IMUs, but the privileged policy $\pi_{\text{privileged}}$ requires them.

In our state space design, proprioception is defined by $\boldsymbol{s}_t^{\text{p}} \triangleq [\boldsymbol{q}_t, \dot{\boldsymbol{q}}_t, \boldsymbol{v}_t, \boldsymbol{\omega}_t, \boldsymbol{g}_t, \boldsymbol{a}_{t-1}]$, with joint position $\boldsymbol{q}_t \in \mathbb{R}^{19}$ (DoF position), joint velocity $\dot{\boldsymbol{q}}_t \in \mathbb{R}^{19}$ (DoF velocity), root linear velocity $\boldsymbol{v}_t \in \mathbb{R}^3$, root angular velocity $\boldsymbol{\omega}_t \in \mathbb{R}^3$, root projected gravity $\boldsymbol{g}_t \in \mathbb{R}^3$, and last action $\boldsymbol{a}_{t-1} \in \mathbb{R}^{19}$. The goal state is $\boldsymbol{s}_t^{\text{g}} \triangleq [\hat{\boldsymbol{p}}_t^{\text{kp}}, \hat{\boldsymbol{p}}_t^{\text{kp}} - \boldsymbol{p}_t{}^{\text{kp}}, \hat{\dot{\boldsymbol{p}}}_t^{\text{kp}}]$. $\hat{\boldsymbol{p}}_t^{\text{kp}} \in \mathbb{R}^{8\times 3}$ is the position of eight selected reference body positions (shoulders, elbows, hands, ankles); $\hat{\boldsymbol{p}}_t^{\text{kp}} - \boldsymbol{p}_t{}^{\text{kp}}$ is the position difference between the reference joints and humanoid's own joints; $\hat{\dot{\boldsymbol{p}}}_t^{\text{kp}}$ is the linear velocity of the reference joints. All values are normalized to the humanoid's own coordinate system. As a comparison, we also consider a reduced goal state $\boldsymbol{s}_t^{\text{g-reduced}} \triangleq (\hat{\boldsymbol{p}}_t^{\text{kp}})$, where only the reference position $\hat{\boldsymbol{p}}_t^{\text{kp}}$ but not the position difference. The action space of the agile policy consists of 19-dim joint targets. A PD controller tracks these joint targets by converting them to joint torques: $\tau = K_p(\boldsymbol{a}_t - \boldsymbol{q}_t) - K_d\dot{\boldsymbol{q}}_t$.

B. Reward Design

We formulate the reward function r_t with the summation of three terms: 1) penalty; 2) regularization; and 3) task rewards, which are summarized in detail in Table I. Note that while we only have eight selected body positions $\hat{\boldsymbol{p}}_t^{\text{kp}}$ in our state space, we provide six *full-body* reward terms for all joints (DoF position, DoF velocity, body position, body rotation, body velocity, body angular velocity) for the imitation task. These expressive rewards give more dense reward signals for efficient RL training.

C. Domain Randomization

Domain randomization has been shown to be the key source of robustness and generalization to achieve successful sim-to-real transfers [19, 58]. All the domain randomization we use in `H2O` are listed in Table II, including ground friction coefficient, link mass, Centor-of-Mass (CoM) position of the torso link, PD gains of the PD controller, torque noise on the actually applied torques on each joint, control delay, terrain types. The link mass and PD gains are independently randomized for each link and joint, and the rest are episodic randomized. These domain randomization together can effectively facilitate the sim-to-real transfer for the real-world dynamics and hardware gaps.

TABLE I: Reward components and weights: penalty rewards for preventing undesired behaviors for sim-to-real transfer, regularization to refine motion, and task reward to achieve successful whole-body tracking in real-time.

Term	Expression	Weight
Penalty		
Torque limits	$\mathbb{1}(\boldsymbol{\tau}_t \notin [\boldsymbol{\tau}_{\min}, \boldsymbol{\tau}_{\max}])$	$-2e^{-1}$
DoF position limits	$\mathbb{1}(\boldsymbol{q}_t \notin [\boldsymbol{q}_{\min}, \boldsymbol{q}_{\max}])$	$-1e^{2}$
Termination	$\mathbb{1}_{\text{termination}}$	$-2e^{2}$
Regularization		
DoF acceleration	$\|\ddot{\boldsymbol{q}}_t\|_2^2$	$-8.4e^{-6}$
DoF velocity	$\|\dot{\boldsymbol{q}}_t\|_2^2$	$-3e^{-3}$
Action rate	$\|\boldsymbol{a}_t - \boldsymbol{a}_{t-1}\|_2^2$	$-9e^{-1}$
Torque	$\|\boldsymbol{\tau}_t\|$	$-9e^{-5}$
Feet air time	$T_{\text{air}} - 0.25$ [55]	$8e^{2}$
Feet contact force	$\|F_{\text{feet}}\|_2^2$	$-1e^{-1}$
Stumble	$\mathbb{1}(F_{\text{feet}}^{xy} > 5 \times F_{\text{feet}}^{z})$	$-1e^{3}$
Slippage	$\|\boldsymbol{v}_t{}^{\text{feet}}\|_2^2 \times \mathbb{1}(F_{\text{feet}} \geq 1)$	$-3e^{1}$
Task Reward		
DoF position	$\exp(-0.25\|\hat{\boldsymbol{q}}_t - \boldsymbol{q}_t\|_2)$	$2.4e^{1}$
DoF velocity	$\exp(-0.25\|\hat{\dot{\boldsymbol{q}}}_t - \dot{\boldsymbol{q}}_t\|_2^2)$	$2.4e^{1}$
Body position	$\exp(-0.5\|\boldsymbol{p}_t - \hat{\boldsymbol{p}}_t\|_2^2)$	$4e^{1}$
Body rotation	$\exp(-0.1\|\boldsymbol{\theta}_t \ominus \hat{\boldsymbol{\theta}}_t\|)$	$1.6e^{1}$
Body velocity	$\exp(-10.0\|\boldsymbol{v}_t - \hat{\boldsymbol{v}}_t\|_2)$	$6e^{1}$
Body angular velocity	$\exp(-0.01\|\boldsymbol{\omega}_t - \hat{\boldsymbol{\omega}}_t\|_2)$	$6e^{1}$

TABLE II: The range of dynamics randomization. Describing simulated dynamics randomization, external perturbation, and randomized terrain, which are important for sim-to-real transfer and boost robustness and generalizability.

Term	Value
Dynamics Randomization	
Friction	$\mathcal{U}(0.2, 1.1)$
Base CoM offset	$\mathcal{U}(-0.1, 0.1)$m
Link mass	$\mathcal{U}(0.7, 1.3) \times$ default kg
P Gain	$\mathcal{U}(0.75, 1.25) \times$ default
D Gain	$\mathcal{U}(0.75, 1.25) \times$ default
Torque RFI [56]	$0.1 \times$ torque limit N $\cdot$ m
Control delay	$\mathcal{U}(20, 60)$ms
External Perturbation	
Push robot	interval $= 5s$, $v_{xy} = 0.5$m/s
Randomized Terrain	
Terrain type	flat, rough, low obstacles [57]

D. Early Termination Conditions

We introduce three early termination conditions to make the RL training process more sample-efficient: 1) low height: the base height is lower than 0.3m; 2) orientation: the projected gravity on x or y axis exceeds 0.7; 3) teleoperation tolerance: the average link distance between the robot and reference motions is further than 0.5m.

TABLE III: Quantitative motion imtiation results the uncleaned retargeted AMASS dataset $\hat{\boldsymbol{Q}}^{\text{retarget}}$.

			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	$\mathcal{S} \subset \mathcal{R}^{778}$	✗	85.5%	50.0	43.6	6.9	7.8	46.0	40.9	5.2	6.2
H2O-reduced	$\mathcal{S} \subset \mathcal{R}^{90}$	✓	53.2%	200.2	115.8	11.2	13.8	182.5	111.0	3.0	8.1
H2O-w/o-sim2data	$\mathcal{S} \subset \mathcal{R}^{138}$	✓	67.9%	176.6	95.0	10.2	12.2	163.1	93.8	3.0	7.5
H2O	$\mathcal{S} \subset \mathcal{R}^{138}$	✓	**72.5%**	**166.7**	**91.7**	**8.9**	**11.0**	**151.0**	**88.8**	**2.9**	**7.0**

VI. Experimental Results

A. Simulation Experiments

Baselines. To reveal the effect of different retargeting, state space designs, and sim-to-real training techniques on whole-body teleoperation performance, we consider four baselines:

1) Privileged policy $\pi_{\text{privileged}}$: The privileged policy (trained without any sim-to-real regularizations or domain randomizations) is used to filter the dataset to find infeasible motion. It has no sim-to-real capability and has a much higher input dimension.
2) H2O-w/o-sim2data: H2O without the "sim-to-data" retargeting, trained on the $\hat{\boldsymbol{Q}}^{\text{retarget}}$;
3) H2O-reduced: H2O with a state space of goal state consisting only of selected body positions $\boldsymbol{s}_t^{\text{g-reduced}}$.
4) H2O: Our full H2O system, with all the retargeting process introduced in Section IV and the state space design introduced in Section V-A, trained on $\hat{\boldsymbol{Q}}^{\text{clean}}$;

Metrics. We evaluate these baselines in simulation on the uncleaned retargeted AMASS dataset (10k sequences $\hat{\boldsymbol{Q}}^{\text{retarget}}$). The metrics are as follows:

1) Success rate: the success rate (Succ) as in PHC [18], deeming imitation unsuccessful when, at any point during imitation, the average difference in body distance is on average further than 0.5m. Succ measures whether the humanoid can track the reference motion without losing balance or significantly lag behind.
2) E_{mpjpe} and $E_{g-\text{mpjpe}}$: the global MPJPE $E_{g-\text{mpjpe}}$ and the root-relative mean per-joint position error (MPJPE) E_{mpjpe} (in mm), measuring our imitator's ability to imitate the reference motion both globally and locally (root-relative).
3) E_{acc} and E_{vel}: To show physical realism, we also compare acceleration E_{acc} (mm/frame^2 and velocity E_{vel} (mm/frame) difference.

Results. The experimental results are summarized in Table III, where H2O significantly outperforms H2O-w/o-sim2data and H2O-reduced by a large margin, demonstrating the importance of the "sim-to-data" process and the state-space design of motion goals for RL. Note that the privileged policy and H2O-w/o-sim2data are trained on the entire retargeted AMASS dataset $\hat{\boldsymbol{Q}}^{\text{retarget}}$ while H2O and H2O-reduced are trained on the filtered dataset $\hat{\boldsymbol{Q}}^{\text{clean}}$. The success rate gap between H2O and the privileged policy comes from two factors: 1) H2O uses a much more practical and less informative observation space compared to the privileged policy; 2) H2O is trained with all sim-to-real regularizations and domain randomization. These two factors will both lead to degradation in simulation performance. This shows that while the RL-based avatar control frameworks have achieved impressive results in simulation, transferring them to the real world requires more robustness and stability. With the carefully chosen dataset and the state space, we could make H2O achieve a higher success rate compared to H2O-w/o-sim2data and H2O-reduced. By comparing H2O with H2O-w/o-sim2data, we can see that our "sim-to-data" process is effective in obtaining higher success rate, even when the RL policy is trained on *less* data. Intuitively, an implausible motion may cause the policy to waste resources trying to achieve them, and filtering them out can lead to better overall performance, as also observed in PULSE [32]. Comparing H2O with H2O-reduced, the only difference is the design of the state space of the goal, which indicates that including more informative physical information about motions helps RL to generalize to large-scale motion imitation.

TABLE IV: Quantitative results of H2O on different sizes of motion dataset for training, evaluated on the uncleaned retargeted AMASS dataset $\hat{\boldsymbol{Q}}^{\text{retarget}}$.

	All sequences				
Dataset Size	Succ ↑	$E_{\text{g-mpjpe}}$ ↓	E_{mpjpe} ↓	E_{acc} ↓	E_{vel}
0.1% of $\hat{\boldsymbol{Q}}^{\text{clean}}$	52.0%	198.0	107.9	12.4	13.7
1% of $\hat{\boldsymbol{Q}}^{\text{clean}}$	58.8%	183.8	96.4	10.7	12.0
10% of $\hat{\boldsymbol{Q}}^{\text{clean}}$	61.3%	174.3	92.3	10.8	12.1
100% of $\hat{\boldsymbol{Q}}^{\text{clean}}$	**72.5%**	**166.7**	**91.7**	**8.9**	**11.0**

Ablation on Motion Dataset Size. To show how motion tracking performance scales with the size of the motion dataset, we test H2O with different size of $\hat{\boldsymbol{Q}}^{\text{clean}}$ by randomly selecting 1%, 10% of $\hat{\boldsymbol{Q}}^{\text{clean}}$. The results are summarized in Table IV, where policies trained larger motion datasets continue to improve the tracking performance. Notice that a policy trained on only 0.1% of the data can achieve a surprisingly high success rate, most likely due to the ample domain randomization applied to the humanoid, such as push robot significantly widens the state the humanoid has encountered, improving its generalization capability.

B. Real-world Demonstrations

Deployment Details. For real-world deployment tests, we use a standard 1080P webcam as the RGB camera, and use HybrIK [27] as the 3D human pose estimator running at 30Hz. For the linear velocity estimation of the robot, we

leverage the motion capture system (50Hz), and all the other proprioception is obtained from built-in sensors (200Hz) of Unitree H1 humanoid. Linear velocity state estimation could be replaced by onboard visual/LiDAR odometry methods, though we opt in to MoCap for this work due to its simplicity.

Real-world Teleoperation Results. For real-time teleoperation, the 3D pose estimation from the RGB camera is noisy and can suffer from perspective bias, but our H2O policy shows a strong generalization ability to real-world estimated motion goals in real-time. The real-world teleoperation is shown in Figure 1, Figure 5 and Figure 6, where H2O enables precise real-time teleoperation of humanoids to do whole-body dynamic motions like ball kicking, walking, and back jumping. More demonstrations can be found on our website.

Fig. 5: The humanoid robot is able to track the precise lower-body movements of the human teleoperator.

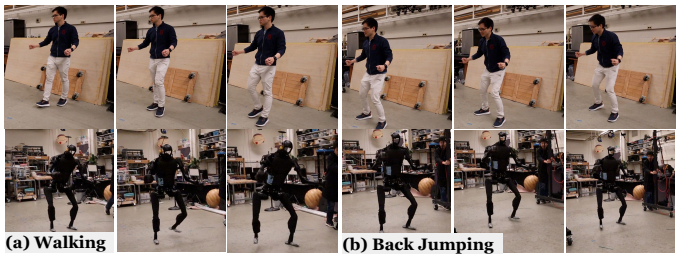

Fig. 6: The humanoid robot is able to track walking motions of human-style pace and imitate continuous back jumping.

Robustness. Our H2O system can keep balance under external force disturbances, as shown in Figure 7. These tests demonstrate the robustness of our system.

VII. Discussions, Limitations, and Future Work

Towards Universal Humanoid Teleoperation. Our ultimate goal is to enable the humanoid to follow as many human-demonstrated motions as possible. We emphasize three key factors that can be improved in the future. 1) Closing the representation gap: as shown in Section VI-A, the state representation of the motion goals critically affects the scalability of RL training with more diverse motions, leading to a trade-off. While incorporating more expressive motion representations into the state space can accommodate finer-grained and more diverse motions, the expanded dimensionality will lead to a curse of sample efficiency in scalable RL. 2) Closing the embodiment gap: as evident in Section VI-A and prior

Fig. 7: Robustness Tests of our H2O system under powerful kicking. The policy is able to maintain balance for both stable and dynamic teleoperated motions.

work [32], training on infeasible or damaged motions might largely harm performance. The feasibility of motions varies from robot to robot due to hardware constraints, and we lack systematic algorithms to identify feasible motions. We need more efforts to close this embodiment gap: on one end, more human-like humanoids would help; on the other, more teleoperation research is expected to improve the learnability of human motions. 3) Closing the sim-to-real gap: to achieve a successful sim-to-real transfer, regularization (e.g., reward regularization) and domain randomization are needed. However, over-regularization and over-randomization will also hinder the policy from learning the motions. It remains unknown how to strike the best trade-off between motion imitation leaning and sim-to-real transfer into a universal humanoid control policy.

Towards Real-time Humanoid Teleoperation. In this work, we leverage RGB and 3D pose estimator to transform the motions of human teleoperators into humanoid robots. The latency and error from RGB cameras and pose estimation also lead to an inevitable trade-off between efficiency and precision in teleoperation. Also, in this work, the human teleoperator receives feedback from the humanoid only in the form of visual perception. More research is needed on human-robot interaction to study this emerging multimodal interaction (e.g., force feedback [59], verbal and conversational feedback [60]), which could further enhance the capability of humanoid teleoperation.

Towards Whole-body Humanoid Teleoperation. One may wonder if lower-body tracking is necessary, as the major embodiment gap between humans and humanoids is the lower-body capability. A large proportion of skillful motions of humans (e.g., sports, dancing) need diverse agile lower-body movements. We emphasize the scenarios where legged robots hold an advantage over wheeled robots, in which lower-body tracking is necessary to follow human lower-body movements, including stepping stones, kicking, spread legs, etc. In the future, a teleoperated humanoid system that learns to switch between robust locomotion and skillful lower-body tracking would be a promising research direction.

VIII. Conclusions

In this study, we introduced Human to Humanoid (H2O), a scalable learning-based framework that enables real-time whole-body humanoid robot teleoperation using just an RGB camera. Our approach, leveraging reinforcement learning and a novel "sim-to-data" process, addresses the complex challenge of translating human motion into actions a humanoid robot can perform. Through comprehensive simulation and real-world tests, H2O demonstrated its capability to perform a wide range of dynamic tasks with high fidelity and minimal hardware requirements.

ACKNOWLEDGMENT

The authors express their gratitude to Jessica Hodgins for providing assistance in conducting hardware experiments. Special thanks are extended to Ziqiao Ma, Zhongyu Li, Yiyu Chen, Xuxin Cheng, and Unitree for their valuable help on graphics design and hardware debugging. Furthermore, we acknowledge the significance of CMU Wean Hall room 1334, formerly utilized as the recording location for the CMU MoCap dataset. In the present study, this dataset is used for real-world humanoid teleoperation within the same room.

References

[1] K. Darvish *et al.*, "Teleoperation of humanoid robots: A survey," *IEEE Transactions on Robotics*, 2023.

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

[3] C. Chi *et al.*, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," *arXiv preprint arXiv:2402.10329*, 2024.

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

[5] D. Kulić, G. Venture, K. Yamane, E. Demircan, I. Mizuuchi, and K. Mombaur, "Anthropomorphic movement analysis and synthesis: A survey of methods and applications," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 776–795, 2016.

[6] F.-J. Montecillo-Puente, M. Sreenivasa, and J.-P. Laumond, "On real-time whole-body human to humanoid motion transfer," 2010.

[7] Y. Ishiguro *et al.*, "High speed whole body dynamic motion experiment with real time master-slave humanoid robot system," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 5835–5841.

[8] J. Ramos and S. Kim, "Humanoid dynamic synchronization through whole-body bilateral feedback teleoperation," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 953–965, 2018.

[9] Y. Ishiguro *et al.*, "Bilateral humanoid teleoperation system using whole-body exoskeleton cockpit tablis," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6419–6426, 2020.

[10] K. Yamane, S. O. Anderson, and J. K. Hodgins, "Controlling humanoid robots with human motion data: Experimental validation," in *2010 10th IEEE-RAS International Conference on Humanoid Robots*, IEEE, 2010, pp. 504–510.

[11] J. Z. Zhang *et al.*, "Slomo: A general system for legged robot motion imitation from casual videos," *IEEE Robotics and Automation Letters*, 2023.

[12] A. Di Fava, K. Bouyarmane, K. Chappellet, E. Ruffaldi, and A. Kheddar, "Multi-contact motion retargeting from human to humanoid robot," in *2016 IEEE-RAS 16th international conference on humanoid robots (humanoids)*, IEEE, 2016, pp. 1081–1086.

[13] K. Otani and K. Bouyarmane, "Adaptive whole-body manipulation in human-to-humanoid multi-contact motion retargeting," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, IEEE, 2017, pp. 446–453.

[14] J. Ramos and S. Kim, "Dynamic locomotion synchronization of bipedal robot and human operator via bilateral feedback teleoperation," *Science Robotics*, vol. 4, no. 35, eaav4282, 2019.

[15] 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)*, vol. 37, no. 4, pp. 1–14, 2018.

[16] J. Won, D. Gopinath, and J. Hodgins, "A scalable approach to control diverse behaviors for physically simulated characters," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 33–1, 2020.

[17] X. B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, "Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters," *ACM Transactions On Graphics (TOG)*, vol. 41, no. 4, pp. 1–17, 2022.

[18] Z. Luo, J. Cao, AlexanderWinkler, K. Kitani, and W. Xu, "Perpetual humanoid control for real-time simulated avatars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, pp. 10 895–10 904.

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

[20] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Learning humanoid locomotion with transformers," *arXiv preprint arXiv:2303.03381*, 2023.

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

[22] 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.

[23] K. Darvish *et al.*, "Whole-body geometric retargeting for humanoid robots," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, IEEE, 2019, pp. 679–686.

[24] R. Cisneros-Limón *et al.*, "A cybernetic avatar system to embody human telepresence for connectivity, exploration, and skill transfer," *International Journal of Social Robotics*, pp. 1–28, 2024.

[25] I. Radosavovic *et al.*, "Humanoid locomotion as next token prediction," *arXiv preprint arXiv:2402.19469*, 2024.

[26] 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*, 2019, pp. 5442–5451.

[27] J. Li, C. Xu, Z. Chen, S. Bian, L. Yang, and C. Lu, "Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3383–3393.

[28] X. B. Peng, G. Berseth, K. Yin, and M. Van De Panne, "Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–13, 2017.

[29] T. Wang, Y. Guo, M. Shugrina, and S. Fidler, "Unicon: Universal neural controller for physics-based character motion," *arXiv preprint arXiv:2011.15119*, 2020.

[30] L. Fussell, K. Bergamin, and D. Holden, "Supertrack: Motion tracking for physically simulated characters using supervised learning," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–13, 2021.

[31] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *ACM Transactions on Graphics (ToG)*, vol. 40, no. 4, pp. 1–20, 2021.

[32] Z. Luo *et al.*, "Universal humanoid motion representations for physics-based control," in *The Twelfth International Conference on Learning Representations*, 2024.

[33] A. Winkler, J. Won, and Y. Ye, "Questsim: Human motion tracking from sparse sensors with simulated avatars," in *SIGGRAPH Asia 2022 Conference Papers*, 2022, pp. 1–8.

[34] Z. Luo, R. Hachiuma, Y. Yuan, and K. Kitani, "Dynamics-regulated kinematic policy for egocentric pose estimation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25 019–25 032, 2021.

[35] Bullet Physics, *Humanoid urdf in bullet3*, Accessed: 2024-03-01, 2023.

[36] Y. Yuan and K. Kitani, "Residual force control for agile human behavior imitation and extended motion synthesis," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 763–21 774, 2020.

[37] L. Penco *et al.*, "Robust real-time whole-body motion retargeting from human to humanoid," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, IEEE, 2018, pp. 425–432.

[38] 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)*, IEEE, 2014, pp. 2806–2812.

[39] O. E. Ramos, N. Mansard, O. Stasse, C. Benazeth, S. Hak, and L. Saab, "Dancing humanoid robots: Systematic use of osid to compute dynamically consistent movements following a motion capture pattern," *IEEE Robotics & Automation Magazine*, vol. 22, no. 4, pp. 16–26, 2015.

[40] L. Penco *et al.*, "Mixed reality teleoperation assistance for direct control of humanoids," *IEEE Robotics and Automation Letters*, 2024.

[41] K. Ayusawa and E. Yoshida, "Motion retargeting for humanoid robots based on simultaneous morphing parameter identification and motion optimization," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1343–1357, 2017.

[42] K. Hu, C. Ott, and D. Lee, "Online human walking imitation in task and joint space based on quadratic programming," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2014, pp. 3458–3464.

[43] S. Bohez *et al.*, "Imitate and repurpose: Learning reusable robot movement skills from human and animal behaviors," *arXiv preprint arXiv:2203.17138*, 2022.

[44] A. Tang *et al.*, "Humanmimic: Learning natural locomotion and transitions for humanoid robot via wasserstein adversarial imitation," *arXiv preprint arXiv:2309.14225*, 2023.

[45] M. Seo *et al.*, "Deep imitation learning for humanoid loco-manipulation through human teleoperation," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, IEEE, 2023, pp. 1–8.

[46] S. Dafarra *et al.*, "Icub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots," *Science Robotics*, vol. 9, no. 86, eadh3834, 2024.

[47] J. Chagas Vaz, D. Wallace, and P. Y. Oh, "Humanoid loco-manipulation of pushed carts utilizing virtual reality teleoperation," in *ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers, vol. 85628, 2021, V07BT07A027.

[48] M. Elobaid, Y. Hu, G. Romualdi, S. Dafarra, J. Babic, and D. Pucci, "Telexistence and teleoperation for walking humanoid robots," in *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2*, Springer, 2020, pp. 1106–1121.

[49] S. Tachi, Y. Inoue, and F. Kato, "Telesar vi: Telexistence surrogate anthropomorphic robot vi," *International Journal of Humanoid Robotics*, vol. 17, no. 05, p. 2 050 019, 2020.

[50] C. D. Bellicoso *et al.*, "Alma-articulated locomotion and manipulation for a torque-controllable robot," in *2019 International conference on robotics and automation (ICRA)*, IEEE, 2019, pp. 8477–8483.

[51] C. Lenz *et al.*, "Nimbro wins ana avatar xprize immersive telepresence competition: Human-centric evaluation and lessons learned," *International Journal of Social Robotics*, pp. 1–25, 2023.

[52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017. arXiv: 1707.06347.

[53] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, vol. 34, no. 6, 248:1–248:16, Oct. 2015.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[55] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, *Learning to walk in minutes using massively parallel deep reinforcement learning*, 2022. arXiv: 2109.11978 [cs.RO].

[56] L. Campanaro, S. Gangapurwala, W. Merkt, and I. Havoutis, *Learning and deploying robust locomotion policies with minimal dynamics randomization*, 2023. arXiv: 2209.12878 [cs.RO].

[57] T. He, C. Zhang, W. Xiao, G. He, C. Liu, and G. Shi, *Agile but safe: Learning collision-free high-speed legged locomotion*, 2024. arXiv: 2401.17583 [cs.RO].

[58] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2018, pp. 3803–3810.

[59] F. Yang, C. Ma, J. Zhang, J. Zhu, W. Yuan, and A. Owens, "Touch and go: Learning from human-collected vision and touch," *arXiv preprint arXiv:2211.12498*, 2022.

[60] J. Y. Chai, Q. Gao, L. She, S. Yang, S. Saba-Sadiya, and G. Xu, "Language to action: Towards interactive task learning with physical agents.," in *IJCAI*, 2018, pp. 2–9.