



# INTERMIMIC: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions

## Supplementary Material

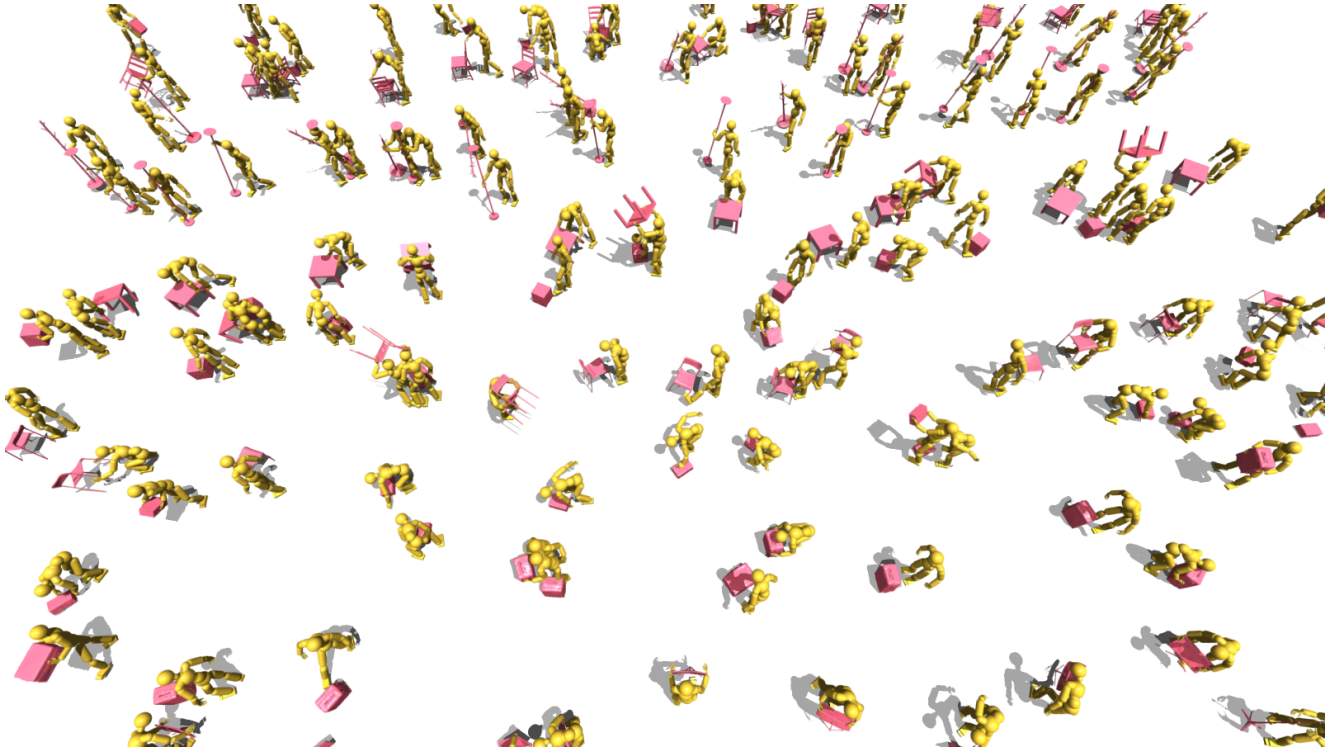


Figure A. InterMimic enables simulated humans to perform physical interactions, featuring scalable skill learning covering diverse objects.

In this supplementary, we provide additional method details and experimental setups:

- (i) **Demo Video.** A demonstration video (with a screenshot in Figure A) is provided at [demo.mp4](#), as described in Sec. A.
- (ii) **Simulation Setup.** The environment configuration for physical HOI simulations is introduced in Sec. B.
- (iii) **Reference Contact Labels.** Additional information on obtaining the reference contact label  $\hat{c}_t$  is detailed in Sec. C.
- (iv) **Reward Formulation.** A comprehensive explanation of the reward design is provided in Sec. D.
- (v) **Physical State Initialization & Interaction Early Termination.** Further insights into these mechanisms are discussed in Sec. E.
- (vi) **Implementation Details.** This section (Sec. F) covers reframing our method for interaction prediction and text-guided interaction generation, as well as translating MoCap interactions into humanoid robot

skills.

- (vii) **Additional Experiments.** Sec. G presents further qualitative results and analyzes failure cases.
- (viii) **Limitations and Societal Impact.** Finally, we examine the limitations of InterMimic and its potential societal implications in Sec. H.

We will release the code for this project at [our webpage](#).

## Contents

<b>A Demo Video</b>	<b>2</b>
<b>B Setup of Physical Interaction Simulation</b>	<b>2</b>
<b>C Reference Contact</b>	<b>2</b>
C.1. Inferring Reference Dynamics . . . . .	3
<b>D Additional Details on Reward</b>	<b>3</b>
D.1. Embodiment-Aware Reward . . . . .	3
D.2. Contact Reward . . . . .	3

D.3 Hand Interaction Recovery . . . . .	4
D.4 Energy Reward . . . . .	4
D.5 Reward Aggregation . . . . .	4
<b>E Additional Details on Trajectory Collection</b>	<b>4</b>
E.1. Interaction Early Termination . . . . .	4
E.2. Physical State Initialization . . . . .	4
<b>F. Additional Implementation Details</b>	<b>5</b>
<b>G Additional Experimental Results</b>	<b>5</b>
<b>H Discussion</b>	<b>5</b>

## A. Demo Video

In addition to the qualitative results presented in the main paper, we provide a demo video ([demo.mp4](#)) for more detailed visualizations of the tasks, further illustrating the efficacy of our approach. The demo video conveys the following key points:

- (i) Our teacher policy can imitate highly dynamic and long-term interactions, both of which are inherently challenging.
- (ii) We visualize the effectiveness of our teacher policy in *HOI retargeting*. Given MoCap references for humans, we successfully transfer these tasks to a humanoid robot, tolerating embodiment differences.
- (iii) Our method corrects errors in reference interactions, addressing contact penetration, floating, and jittering issues. This demonstrates how teacher-based reference distillation can provide cleaner data for student policy training.
- (iv) The baseline method PhysHOI [88] *fails* on sequences our approach successfully imitates, complementing Figure 4 in the main paper.
- (v) Our student policy exhibits strong scalability, effectively learning from hours of data across diverse objects and interaction skills.
- (vi) The framework grants the student policy *zero-shot* generalizability, enabling direct application to text-to-HOI, interaction prediction, and interactions with new skills or objects – even multiple objects not present in the training set.

## B. Setup of Physical Interaction Simulation

The reference data represent humans using the SMPL models [61, 68]. For simulation, we convert these models into box and cylindrical rigid bodies following [50]. Objects are also converted into simulation models through convex decomposition, as illustrated in Figure B. We summarize the physics parameters for our task in Table A. We follow the physics parameters for the human as specified in [88, 89],

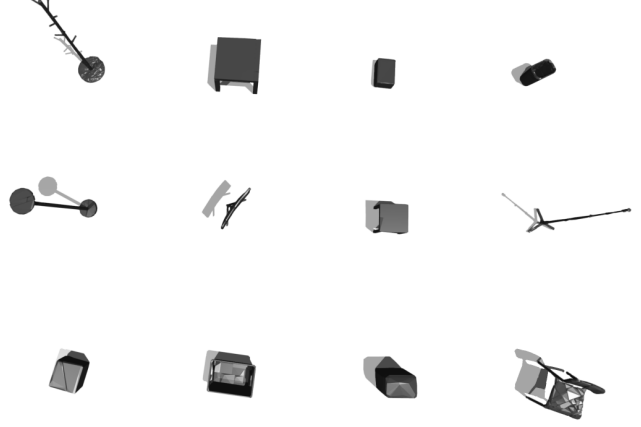


Figure B. Visualization of the objects from OMOMO [39], each decomposed into 64 convex hulls for simulation.

Hyperparameter	Value
Sim $dt$	1/60s
Control $dt$	1/30s
Number of envs	8192
Number of substeps	2
Number of pos iterations	4
Number of vel iterations	0
Contact offset	0.02
Rest offset	0.0
Max depenetration velocity	100
Object & ground restitution	0.7
Object & ground friction	0.9
Object density	200
Object max convex hulls	64

Table A. Simulation hyperparameters used in Isaac Gym [54].

with the exception of the specialized range of motion (RoM) for hands, detailed in Table B. Our range of motion (RoM) setting is biologically inspired: finger flexion and extension (bending and straightening) are fully activated. However, unlike the real human, the abduction and adduction of the Metacarpophalangeal (MCP) joint are constrained to minimize the risk of finger interpenetration, in the absence of the correct reference hand pose for guidance. The rationale for these RoM settings is discussed in Sec.3.2 of the main paper and Sec.D.3 of the supplementary.

## C. Reference Contact

In this section, we detail how we extract the reference contact that formulates the state and the reward as discussed in Sec. 3.1 of the main paper. One solution involves loading the HOI data into the simulation, replaying the data, and using the force detector in Isaac Gym [54] to identify contact,

Joint	x-dim	y & z-dim
MCP	$[-55.625^\circ, 55.625^\circ]$	$[-5.625^\circ, 5.625^\circ]$
PIP	$[-55.625^\circ, 55.625^\circ]$	
DIP	$[-5.625^\circ, 90.000^\circ]$	
CMC	$[-55.625^\circ, 55.625^\circ]$	$[-55.625^\circ, 55.625^\circ]$
MCP	$[-5.625^\circ, 5.625^\circ]$	$[-5.625^\circ, 5.625^\circ]$
IP	$[-5.625^\circ, 90.000^\circ]$	$[-5.625^\circ, 5.625^\circ]$

Table B. We constrain the Range of Motion (RoM) for the joints in the index, middle, ring, and pinky fingers including: the MCP (Metacarpophalangeal) joint where the finger meets the hand, the PIP (Proximal Interphalangeal) joint as the middle joint, and the DIP (Distal Interphalangeal) joint closest to the fingertip. For the thumb, we consider the CMC (Carpometacarpal) joint at the base in the palm, the MCP connecting the thumb to the hand, and the IP (Interphalangeal) joint within the thumb. The coordinates for describing these RoMs are based on the human model from [50].

as suggested by [88]. However, this approach is ineffective for imperfect MoCap data; for instance, the force detector fails to capture contact when floating artifacts occur. To address this limitation, we propose solutions tailored differently for teacher and student training:

**Reference contact for the student.** We query the force detector from distilled reference in the simulation rather than from MoCap data replay, as the teacher policy is capable of correcting artifacts.

**Reference contact for teachers.** To account for contact distance variances, we determine reference contact based on inferred dynamics from kinematics, as outlined below.

### C.1. Inferring Reference Dynamics

By analyzing the object’s acceleration over time, we can approximate external forces without depending on simulated dynamics. We assume human-object interaction occurs if any of these conditions hold: **(i)** The object is airborne, but its acceleration deviates significantly from gravitational acceleration, indicating that an external force, *e.g.*, human interaction is acting upon it. **(ii)** The object is on the ground but not static, and its acceleration significantly differs from what is expected due to friction alone, suggesting additional force input. **(iii)** The minimum distance between the human and object vertices is below 0.01 meters.

When any condition is met, we define the contact threshold  $\sigma$  as the minimum distance between the human and object vertices, plus 0.005 meters. This adaptive threshold is essential for accommodating contact distance variations in the ground truth MoCap data. For example, the contact promotion marker is defined as  $\hat{c}_b[i] = \|\hat{\mathbf{d}}[i]\| < \sigma$ , where  $i$  is the index of human rigid bodies. We integrate  $\hat{c}_b$  into the contact promotion reward  $R_b^c$ , as introduced in Sec. 3.2 of the main paper and detailed in Sec. D.2 of supplementary.

$\hat{\mathbf{d}}$  is the joint-to-object vectors as defined in Sec. 3.1.

## D. Additional Details on Reward

In this section, we provide further details about the reward function used for policy training. Specifically, we describe how we balance the components of the embodiment-aware reward, formulate the contact and energy rewards, address hand interaction recovery, and explain the process of integrating all rewards into a unified scalar.

### D.1. Embodiment-Aware Reward

We formulate the weight  $w_d$ , introduced in Sec. 3.2 of the main paper, for balancing the embodiment-aware reward:

$$w_d[i] = 0.5 \times \frac{1/\|\mathbf{d}[i]\|^2}{\sum_i 1/\|\mathbf{d}[i]\|^2} + 0.5 \times \frac{1/\|\hat{\mathbf{d}}[i]\|^2}{\sum_i 1/\|\hat{\mathbf{d}}[i]\|^2}, \quad (2)$$

where  $i$  is the joint index, and  $\mathbf{d}$  and  $\hat{\mathbf{d}}$  are vectors from the human joint to the object surface for simulation and reference, respectively, as defined in Sec. 3.1 of the main paper. The value  $\|\mathbf{d}[i]\|^2$  and  $\|\hat{\mathbf{d}}[i]\|^2$  are clipped by a small positive value to prevent division by zero.

Our joint position and rotation tracking rewards,  $R_p^h$  and  $R_\theta^h$ , include both body and hand joints, even for imitating datasets such as [3, 39] which present hands always in flat or mean poses. This encourages hands to maintain a reasonable default pose when the contact reward is not activated.

### D.2. Contact Reward

The contact promotion cost function  $E_b^c$  is designed to encourage highly probable contact, as highlighted by the red regions in Figure 3(i) of the main paper. This reward utilizes the adaptive contact marker  $\hat{c}_b$ , described in Sec. C.1,

$$E_b^c = \sum \|\hat{c}_b - \mathbf{c}\| \odot \hat{c}_b, \quad (3)$$

where  $\mathbf{c}$  is the simulated contact extracted from the force detected, as introduced in Sec. 3.1 of the main paper.

Contact penalties, applied to the blue regions in Figure 3(i) of the main paper, are defined using a larger and fixed threshold of  $\sigma_p = 0.1$ . Specifically,  $\hat{c}_p[i] = (\|\hat{\mathbf{d}}[i]\| > \sigma_p) \wedge \neg \hat{c}_g[i]$ , where  $\|\hat{\mathbf{d}}[i]\|$  is the distance between joint  $i$  and the object surface in the reference interaction as defined in Sec. 3.1 of the main paper, and the negation  $\neg$  of  $\hat{c}_g[i]$  indicates the rigid body part  $i$  that is not in contact with the ground. The cost of penalty is then calculated as:

$$E_p^c = \sum \|\mathbf{c}\| \odot \hat{c}_p. \quad (4)$$

### D.3. Hand Interaction Recovery

Our hand contact guidance is defined as:

$$E_h^c = \sum \|c^{\text{lhand}} - \hat{c}^{\text{lhand}}\| \odot \hat{c}^{\text{lhand}} \quad (5)$$

$$+ \|c^{\text{rhand}} - \hat{c}^{\text{rhand}}\| \odot \hat{c}^{\text{rhand}}, \quad (6)$$

where  $c^{\text{lhand}}$  and  $c^{\text{rhand}}$  represent contact labels for rigid body components of the hands. The reference contact markers,  $\hat{c}^{\text{lhand}}$  and  $\hat{c}^{\text{rhand}}$ , are defined when any hand vertices are within an adaptive threshold distance  $\sigma$  to the objects, as described in Sec. C.1 of supplementary. To avoid overly aggressive hand contact that could lead to unrealistic poses, we impose range of motion constraints for the hand, as shown in Table B, ensuring that RL-explored grasping remains biologically realistic.

### D.4. Energy Reward

We define the energy cost as  $E_h^e = \sum \|a_h\|$ ,  $E_o^e = \sum \|a_o\|$ , and  $E_c^e = \max \|f\|$ , where  $a_h$  represents the acceleration of human joints,  $a_o$  the object’s acceleration, and  $f$  the force detected on human rigid bodies. Applying them penalizes abrupt contact and promotes smooth interactions.

### D.5. Reward Aggregation

We define the weights for each cost function, including  $E_p^h$ ,  $E_\theta^h$ ,  $E_d$ ,  $E_p^o$ , and  $E_\theta^o$ , as described in Sec. 3.2 of the main paper, along with  $E_b^c$ ,  $E_p^c$ ,  $E_h^c$ ,  $E_\theta^c$ ,  $E_o^c$ , and  $E_c^e$  detailed in supplementary as  $(\lambda_p^h, \lambda_\theta^h, \lambda_d, \lambda_p^o, \lambda_\theta^o, \lambda_{cb}, \lambda_{cp}, \lambda_{ch}, \lambda_e^h, \lambda_e^o, \lambda_e^f)$ . The final aggregated reward is computed as:  $R = \exp(-\lambda_\theta^h E_\theta^h - \lambda_p^h E_p^h - \lambda_\theta^o E_\theta^o - \lambda_p^o E_p^o - \lambda_d E_d - \lambda_{cb} E_b^c - \lambda_{cp} E_p^c - \lambda_{ch} E_h^c - \lambda_e^h E_e^h - \lambda_e^o E_e^o - \lambda_e^f E_e^f)$ , following a multiplication of the exponential structure, as suggested in [59, 92].

## E. Additional Details on Trajectory Collection

### E.1. Interaction Early Termination

In Sec. 3.2 of the main paper, we introduce the termination conditions defined for human-object interaction. Additionally, we use three conditions general for single human imitation as follows: (i) The joints are, on average, more than 0.5 meters from their reference. (ii) The root joint is under the height of 0.15. (iii) The episode ends after 300 frames, as the maximum episode length (also specified in Table C).

### E.2. Physical State Initialization

**Limitations of RSI.** Figure C illustrates why Reference State Initialization (RSI) [63] is suboptimal for interaction imitation with imperfect MoCap data. In single-person MoCap scenarios, where failures are less frequent, RSI performs well; however, in the presence of MoCap errors, RSI

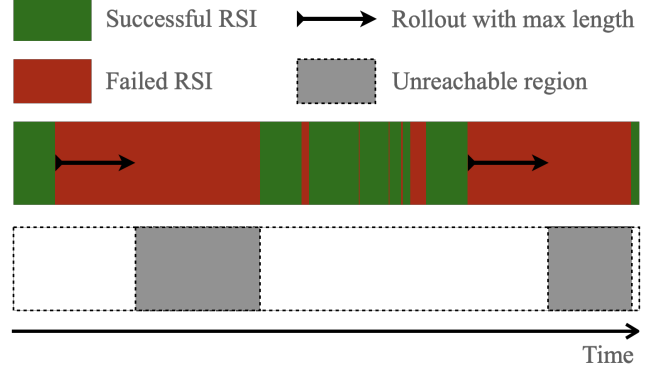


Figure C. A sanity check on why Reference State Initialization (RSI) [63] can fail: we use a bar representing the reference interaction sequence that the policy imitates, where red regions indicate that initializing in those regions leads to immediate failure, while green regions signify that successful initialization is possible. There may be periods, shown as two gray blocks, where the policy cannot collect trajectories for updates (*i.e.*, unreachable regions), as the successful rollout cannot cover large failed RSI region given the fixed length of the rollout. In real scenarios, rollouts can be suboptimal and terminated prematurely, preventing the policy from collecting sufficient trajectories for challenging periods that extend beyond the boundaries illustrated by the gray blocks.

leads to reduced experience collection, ultimately undermining performance.

**Does Interaction Early Termination Help?** While early termination can filter out poor initial states, excessive initialization failures lead to frequent simulation resets that significantly slow down training. Consequently, the agent spends more time restarting simulations rather than engaging in productive learning.

**Step-by-step details** to complement Sec. 3.2 of the main paper: (i) PSI begins by creating an initialization buffer that stores a collection of reference states from motion capture data and simulation states from previous rollouts. This buffer is used to select initialization states for future rollouts. (ii) For each new rollout, an initialization state is randomly selected from the buffer. (iii) Using the current policy, the agent performs rollouts in the simulation environment by taking actions, transitioning through states, and receiving rewards. (iv) After each rollout, the collected trajectories are evaluated based on their expected discounted rewards to update the critic network. Trajectories with expected rewards above a defined threshold are added to the PSI buffer, while older or lower-quality trajectories are removed to maintain the buffer’s size and quality. We apply PSI in a sparse manner to enhance training efficiency, with a probability of 0.005 for updating the buffer for each rollout.



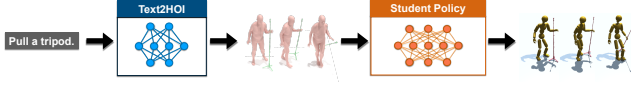


Figure D. Overview of integrating HOI-Diff [62] with InterMimic to perform text-guided interaction generation, *i.e.*, generating interaction sequences based on text input.

## F. Additional Implementation Details

In Figures D and E, we illustrate the framework that integrates the kinematic generators with our InterMimic – let the policy use the kinematic output as the input reference to imitate. Table C lists the hyperparameters used during the PPO [71]. The weights for the reward function ( $\lambda_p^h, \lambda_\theta^h, \lambda_d, \lambda_p^o, \lambda_\theta^o, \lambda_{c_b}, \lambda_{c_p}, \lambda_{c_h}, \lambda_e^h, \lambda_e^o, \lambda_e^f$ ) are set as  $(30, 2.5, 5, 0.1, 5, 5, 5, 3, 2 \times 10^{-5}, 2 \times 10^{-5}, 10^{-9})$ .

For evaluation on the OMOMO [39] dataset, we use Subject 9 as the base model, with teacher policies retargeting interactions from other subjects into this base.

Similar to existing motion imitation approaches [63], we use API in Isaac Gym [54] to initialize the first frame to match the first reference frame – whether it comes from MoCap or kinematic generation methods. The subsequent sequence is then simulated based on the starting frame.

For learning interaction skills on a humanoid robot [24, 81] from SMPL-X [61] data, we bypass external retargeting and directly learn, highlighting our framework’s integrated ability for both retargeting and imitation. Note that we model each Inspire hand with 12 actuators using PD control, without accounting for the mimic joint present in the actual setup, which could be inapplicable in real deployment. Due to the embodiment gap, the humanoid cannot be initialized to match the first SMPL-X frame. Thus, we adopt a two-stage approach: during the first 15 frames, the policy learns to stand and approach the reference’s initial pose, establishing a basis for subsequent tracking. Afterward, the policy transitions to track the reference. We rewrite the position and rotation rewards for the robot’s joints mapped to SMPL-X joints. We do not use the contact reward as we disable the self-collision, since the human reference now cannot ensure proper collision constraints for the humanoid robot. To mitigate the impact of contact artifacts in MoCap data without relying on a contact reward, we leverage teacher distillation references for training.

For interactions involving multiple objects, our framework remains unchanged except for the state and reward components related to the objects, such as  $\{\theta_t^o, p_t^o, \omega_t^o, v_t^o\}$ ,  $d_t$ , and the rewards  $R_p^o$ ,  $R_\theta^o$ , and  $R_d$ , which now include multiple components to represent multiple objects.

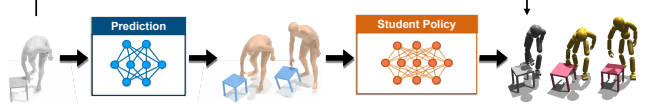


Figure E. Overview of integrating InterDiff [101] with InterMimic to perform interaction prediction, *i.e.*, generating future interactions based on past interaction frames.

Hyperparameters	value
Action distribution	153D Continuous
Discount factor $\gamma$	0.99
Generalized advantage estimation $\lambda$	0.95
Entropy regularization coefficient	0.0
Optimizer	Adam [33]
Learning rate (Actor)	2e-5
Learning rate (Critic)	1e-4
Minibatch size	16384
Horizon length $H$	32
Action bounds loss coefficient	10
Maximum episode length	300

Table C. Hyperparameters for training teacher and student policies.

## G. Additional Experimental Results

In this section, we introduce experimental results that are not included in the main paper due to space limit.

**Failure Cases.** In Figure F, we illustrate an example where our teacher policies fail to perform successful imitation. Despite the strong adaptability of our policies, as demonstrated in Figures 1 and 5, where they effectively correct reference errors, there are limitations when encountering too many errors. Since the reward design inherently prioritizes reference tracking, excessive errors in the reference inevitably result in failures.

**HOI Retargeting.** Figure G shows that teacher policies, trained on reference data for a specific body shape, can successfully drive a human model with a body shape different from the reference in the simulator to accomplish the same task, albeit with slightly varied trajectories. This result highlights the effectiveness of our design, which integrates retargeting into interaction imitation.

## H. Discussion

**Limitations and Future Work.** One limitation, as discussed in Sec. G and illustrated in Figure F, is that our method struggles to fully correct MoCap data with significant errors. However, it also underscores a strength of our teacher-student framework: teacher policies filter out data that are too corrupted to imitate, allowing the student policy to concentrate on learning from viable samples and avoid wasting training effort on unrecoverable data.

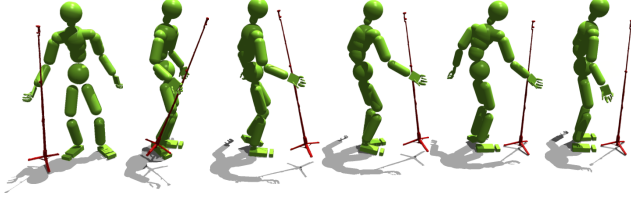


Figure F. For certain reference from OMOMO [39], the hand is incorrectly flipped, which leads to the failure of the teacher policy. We exclude such data when training the student policy.

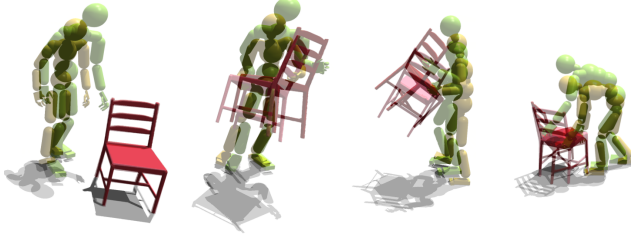


Figure G. Comparison between the reference interaction (human shown in green) and the simulated interaction (human shown in yellow) demonstrates that, despite the different body shapes, the simulated human driven by InterMimic successfully accomplishes the same task with different trajectories, highlighting the effectiveness of our imitation as retargeting.

The policy sometimes results in unnatural object support, where the human produces penetration rather than relying on friction. While we mitigate this issue by setting a high maximum depenetration velocity in simulation (See Table A) and applying a contact-based energy (See Sec. D.4) to discourage large forces that could cause penetration, it does not entirely solve the problem. A potential solution could involve using a signed distance-based penetration score as a criterion for early termination.

The hand interaction recovery method is effective for the tasks explored in this paper. For tasks requiring dexterity with detailed finger motions, its benefits may be limited.

Additionally, while our method demonstrates good scalability by effectively training on hours of MoCap data involving different objects and generalizing to unseen skills and object geometries, its performance could be further improved with a larger dataset. Incorporating more diverse objects [97] would likely further enhance InterMimic’s zero-shot generalization capabilities.

**Potential Negative Societal Impact.** Our approach has the potential to generate vivid human-object interaction sequences, which, if misused, could lead to negative societal impacts, with the risk of creating misleading content by depicting individuals in fabricated scenarios. However, our model is designed with privacy in mind – it employs an abstract representation, using simple geometric shapes like boxes and cylinders to depict different parts. This abstrac-

tion reduces the inclusion of personally identifiable features, making it less likely for our synthesized data to be misused in ways that compromise individual identities.