SkillMimic: Learning Basketball Interaction Skills from Demonstrations

Yinhuai Wang[1,2*]   Qihan Zhao[1*]   Runyi Yu[1,2*]   Hok Wai Tsui[1]   Ailing Zeng[6†]   Jing Lin[5]
Zhengyi Luo[7]   Jiwen Yu[3]   Xiu Li[4]   Qifeng Chen[1]   Jian Zhang[3†]   Lei Zhang[5]   Ping Tan[1]

[1]Hong Kong University of Science and Technology   [2]Unitree Robotics   [3]Peking University Shenzhen Graduate School
[4]Tsinghua University   [5]International Digital Economy Academy   [6]Tencent   [7]Carnegie Mellon University
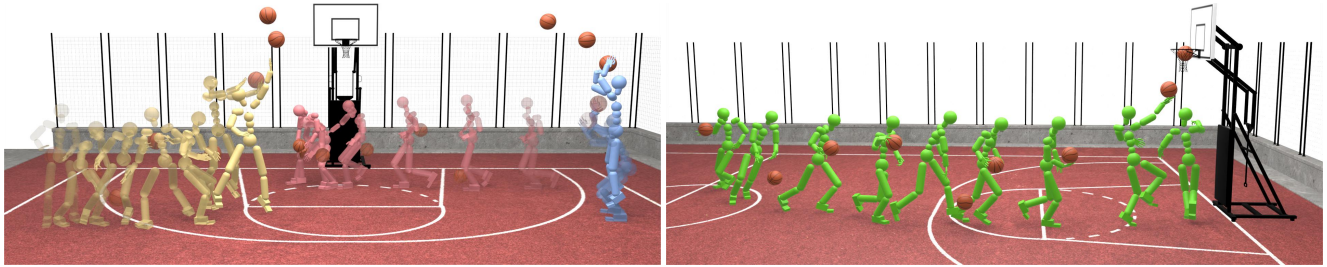yinhuai.wang@connect.ust.hk   pingtan@ust.hk

Figure 1. We propose a novel approach that for the first time enables physically simulated humanoids to learn a variety of basketball interaction skills from Human-Object Interaction (HOI) data, including but not limited to shooting (blue), retrieving (red), and turnaround layup (yellow). Once acquired, these interaction skills can be composed to accomplish complex tasks, such as consecutive scoring (green).

Abstract

Traditional reinforcement learning methods for human-object interaction (HOI) rely on labor-intensive, manually designed skill rewards that do not generalize well across different interactions. We introduce SkillMimic, a unified data-driven framework that fundamentally changes how agents learn interaction skills by eliminating the need for skill-specific rewards. Our key insight is that a unified HOI imitation reward can effectively capture the essence of diverse interaction patterns from HOI datasets. This enables SkillMimic to learn a single policy that not only masters multiple interaction skills but also facilitates skill transitions, with both diversity and generalization improving as the HOI dataset grows. For evaluation, we collect and introduce two basketball datasets containing approximately 35 minutes of diverse basketball skills. Extensive experiments show that SkillMimic successfully masters a wide range of basketball skills including stylistic variations in dribbling, layup, and shooting. Moreover, these learned skills can be effectively composed by a high-level controller to accomplish complex and long-horizon tasks such as consecutive scoring, opening new possibilities for scalable and generalizable interaction skill learning. Project page: https://ingrid789.github.io/SkillMimic/

1. Introduction

Human-Object Interaction (HOI) represents one of the most fundamental yet challenging capabilities in robotics and character animation [10, 12, 16, 23, 28, 46, 47]. Among various HOI scenarios, basketball serves as an exemplary testbed that demands mastery of complex, dynamic, and precise interactions. Despite significant progress, existing physics-based character animation methods predominantly focus on imitation learning of locomotion skills [8, 32–34, 42] through Reinforcement Learning (RL), leaving a significant gap for HOI imitation. While some methods can learn interaction skills, such as striking pillars [34], playing tennis [50], climbing ropes [2], and carrying boxes [15], they typically involve handcrafted reward functions for each different interaction skill. However, designing skill-specific rewards is labor-intensive and fundamentally limits generalization. This limitation becomes particularly acute in basketball, where the complexity and diversity of skills would demand extraordinary engineering effort for reward design. Consequently, existing methods [2, 8, 15, 32–34, 42, 50] fail to provide a unified learning framework for a single policy to learn diverse interaction skills, let alone achieving complex long-term tasks such as consecutive basketball scoring.

How can a simulated humanoid learn a wide variety of interaction skills in a simple and scalable manner? Consider how humans naturally acquire complex skills: amateur basketball players often develop their expertise by watching

and imitating game footage, even without formal coaching. This self-directed learning process involves synchronizing their body movements with both the observed player motions and ball trajectories through practice.

Inspired by these observations, we propose a data-driven method called SkillMimic, which mimics HOI through RL to learn interaction skills. SkillMimic can learn various interaction skills purely from HOI datasets, such as diverse styles of basketball shooting, layup, and dribbling skills, and even a robust pickup skill that enables picking up balls in random locations and motions. Notably, SkillMimic uses the exact same configuration to learn different interaction skills, with identical hyperparameters. This allows us to train a single Interaction Skill (IS) policy to learn multiple interaction skills and achieve smooth skill switching. By training a high-level policy to reuse these learned interaction skills, we can accomplish challenging high-level tasks such as scoring, which requires the humanoid to composite diverse basketball interaction skills to score accurately.

To advance basketball skill learning, we introduce two HOI datasets: BallPlay-V, which captures player and ball motion from RGB videos across eight basic skills, and BallPlay-M, which uses optical motion capture to record 35 minutes of comprehensive basketball interactions. Using these datasets, SkillMimic successfully learns diverse basketball skills through a unified IS policy. With an additional high-level controller for skill composition, our system can execute complex tasks like consecutive scoring. Our approach demonstrates superior efficiency in learning both interaction skills and multi-phase, long-horizon tasks compared to existing methods, opening new possibilities for scalable and generalizable interaction skill learning.

Specifically, our contributions are as follows:

- **SkillMimic**: a data-driven paradigm for learning diverse humanoid interaction skills through RL. It supports joint learning and switching of diverse basketball interaction skills, with skill diversity and generalization improving as the demonstration data grows, showcasing the scalability.
- **Contact graph**: We propose a simple and general contact modeling method that applies to diverse skills, called the contact graph. A contact graph reward is designed to enable precise contact imitation, which proves to be critical for learning precise interaction skills.
- **Unified HOI imitation reward**: We propose a set of important designs that form a unified reward configuration for imitation learning of various interaction skills.
- **A hierarchical solution for reusing interaction skills**: We propose training a high-level controller to reuse the interaction skills acquired by SkillMimic to accomplish challenging high-level tasks, which for the first time achieves continuous basketball scoring.
- **BallPlay Datasets**: We introduce two basketball datasets to facilitate research on interaction skill learning.

2. Related Work

Locomotion Imitation Learning. Mimicking human motion for robot control is an efficient method of learning humanoid skills. DeepMimic [32], a pioneer in this field, uses imitation learning to perform a variety of highly dynamic locomotion skills. The introduction of Generative Adversarial Imitation Learning (GAIL) [18] into humanoid imitation learning by AMP [33] lessens constraints on data alignment, thus enhancing its versatility. ASE [34] further amplifies motion diversity in GAIL training, incorporates a pre-trained low-level policy to acquire locomotion skills, and deploys a high-level policy to repurpose these locomotion skills for specific tasks, such as striking pillars. These methodologies are subsequently expanded in numerous efforts to augment condition and text control [8, 22, 29, 31, 35, 39, 42, 44, 53], as well as to design automated rewards [7]. However, a limitation of these studies is their exclusive focus on learning isolated locomotion skills through imitation. This requires additional pipeline and reward design for learning interaction skills, even for relatively simple ones, such as moving boxes [15] or striking pillars [34]. In contrast, our approach enables unified learning of diverse interaction skills, eliminating the need for skill-specific reward engineering.

Learning Human-Object Interaction (HOI). Learning HOI in simulation has long posed challenges in the fields of graphics and robotics [14, 20, 21, 23, 24, 37, 38, 45, 47, 48]. Early studies often relied on manually designed control structures. For example, Hodgins et al. [19] implement a variety of sports interactions through the use of manually designed state machine controllers. Yin et al. [49] use finite-state machines to generate a large variety of gaits. Coros et al. [6] utilize an inverted pendulum model to control foot placement. Recently, deep reinforcement learning [1] has been widely employed in character control, giving rise to an array of methods based on policy network control. These methods have been applied to a range of challenging activities, such as basketball [26], skateboarding [25], and cycling [41], etc. The use of imitation rewards further lessens the reward design for locomotion styles. Zhang et al. [50] propose a framework to learn diverse tennis skills from broadcast videos. Bae et al. [2] incorporate multiple part-wise motion priors for diverse whole-body interactions. Similarly, Braun et al. [4] develop a framework for synthesizing whole-body grasps.

Applying imitation learning for HOI is intuitive, however, effectively learning diverse interaction skills in a unified framework remains unresolved. Naively extending locomotion imitation methods [32, 33] to HOI imitation often yields unstable and unnatural results, as these methods do not address the unbalanced rewards inherent in HOI, nor do they properly account for the crucial aspects of relative mo-

Figure 2. Concept of SkillMimic. We define an interaction skill as a set of Human-Object Interaction (HOI) state transitions that align with the intended skill semantics. These state transitions can be derived from captured HOI motion clips. If a simulated humanoid can manipulate objects such that the resulting HOI state transitions closely match those of the reference, we consider the humanoid to have successfully learned the interaction skill.

tion and contact in HOI. Modeling interaction through interaction graphs is a well-established approach [17]. While this method has proven useful for HOI imitation [52], it remains limited by considering only kinematic relationships, failing to ensure proper contact and becoming unstable in learning diverse scenarios. In contrast, our contact graph imitation reward explicitly prioritize physical contact learning, significantly improving the success rate of HOI imitation and enabling unified HOI imitation of diverse complex interaction skills for the first time.

This work builds upon and significantly extends ideas initially explored in our early technical report [43].

3. SkillMimic

In this section, we introduce a unified data-driven framework for simulated humanoids to learn various basketball interaction skills via simulation and reinforcement learning (RL). We begin by introducing how we define interaction skills with Human-Object Interaction (HOI) data (Sec. 3.2). Next, we explain how to learn diverse interaction skills by imitating HOI data (Sec. 3.3) and discuss the designing of a unified HOI imitation reward (Sec. 3.4). Finally, we demonstrate how the learned interaction skills can be reused to perform high-level tasks (Sec. 3.5).

3.1. The BallPlay Dataset

To address the scarcity of basketball HOI data and facilitate research on interaction skill learning, we introduce two datasets: BallPlay-V, based on monocular vision estimation, containing eight clips of different basketball skills; and BallPlay-M, based on multi-view optical motion capture systems, containing 35 minutes of diverse basketball interactions. We provide more details in the appendix.

3.2. Defining Interaction Skills Using HOI Data

We propose a data-driven approach to defining interaction skills as collections of Human-Object Interaction (HOI)

state transitions that align with the intended skill semantics. For instance, the interaction skill of "picking up balls from various positions" can be encapsulated by various HOI motion clips, each capturing a unique instance of the action. These clips collectively form a collection of HOI state transitions that represent the interaction skill.

3.3. Learning Interaction Skills by HOI Imitation

Considering an interaction skill defined by a set of reference HOI state transitions, if a humanoid can manipulate objects such that its HOI state transitions closely resemble the reference, we consider the humanoid to have successfully learned the interaction skill. Based on this concept, we propose a method for learning interaction skills by imitating HOI state transitions through RL, which we call SkillMimic. Fig. 2 illustrates the concept of SkillMimic.

In contrast to Behavior Clone [3, 5, 11] methods which rely on precise state-action data, SkillMimic learns actions through RL using state-only trajectories, demonstrating better tolerance to data noise and higher data efficiency. Unlike previous RL methods [33, 34, 50] that require manually designed rewards for each interaction skill, SkillMimic is fully data-driven, skill-agnostic, and scalable, making it capable of learning a wide range of basketball skills within a unified solution. Fig. 6 shows diverse basketball skills acquired using SkillMimic. Fig. 7 illustrates how the performance of the pickup skill improves as the reference data increases.

3.3.1. Training Pipeline

Fig. 3 (b) shows the training pipeline of SkillMimic. Given an HOI dataset with diverse skill-labeled clips (Fig. 3 (a)), SkillMimic trains an Interaction Skill (IS) policy by randomly selecting a frame from a clip corresponding to a chosen skill to initialize the humanoid and object states. The state s_t (Sec. 3.3.2) and skill label c_j are input into the IS policy, which predicts actions (Sec. 3.3.3) that are then simulated to produce the next state s_{t+1}. A unified HOI imitation reward (Sec. 3.4) is designed to measure the consistency between the simulated state transitions and the reference transitions. When the humanoid falls or reaches the maximum simulation length, the environment resets, and the process repeats. After training converges, the IS policy enables the humanoid to execute interaction skills that closely mirror the reference demonstrations.

3.3.2. HOI Observation

The state observed by the IS policy is not under direct supervision and may theoretically encompass any information that is available in the simulation environment. Importantly, this information does not necessarily have to be included in the reference HOI data. Similar to prior arts [34], we transform all coordinates into the root local coordinate of the humanoid, which aligns the data distribution and benefits the generalization performance. For the humanoid, we observe
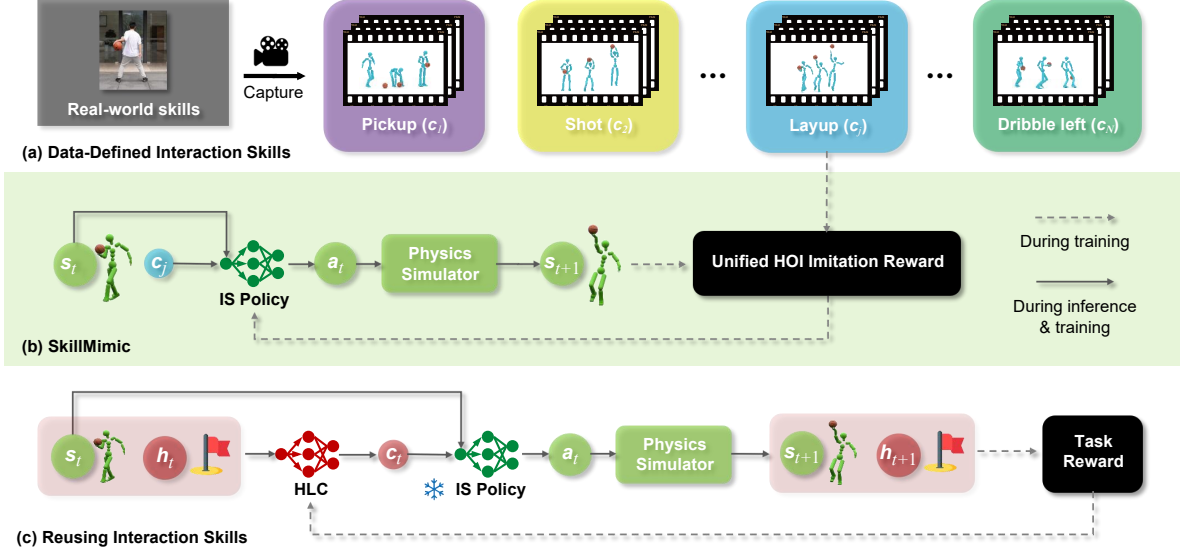
Figure 3. Our system consists of three parts. (a) First, we capture real-world basketball skills to create a large Human-Object Interaction (HOI) motion dataset. (b) Second, we train an Interaction Skill (IS) policy to learn interaction skills by imitating the corresponding HOI data through reinforcement learning. Specifically, the IS policy takes as input the HOI state s_t and skill label c_j and predicts the action a_t. The new state s_{t+1} is calculated by the simulator. A unified HOI imitation reward is designed to imitate diverse HOI state transitions. (c) The third part involves training a High-Level Controller (HLC) to reuse the learned interaction skills for complex tasks. The HLC takes as input s_t and extra task observations h_t, e.g., the basket position, and predicts the skill label c_t to drive a pre-trained IS policy.

its global root height, local body position, rotation, position velocity, and angular velocity. These representations form the humanoid proprioception o_t^{prop}. In addition, we detect net contact forces o_t^f for all fingertips, which helps to sense contact and accelerate training. For the object, we observe its local position, rotation, velocity, and angular velocity, which form the object observation o_t^{obj}. Finally, the state perceived by the IS policy is

$$s_t = \{o_t^{prop}, o_t^f, o_t^{obj}\}. \tag{1}$$

Skill labels c_j are also used as inputs to the IS policy to differentiate different skills. We use one-hot encodings to represent the skill label.

3.3.3. IS Policy and Action

The policy output is modeled as a Gaussian distribution with dimensions equal to the DOF number of the humanoid robot, featuring constant variance. The mean is modeled by a three-layer MLP consisting of [1024, 512, 512] units with ReLU [30] activations. We use the action a_t sampled from the policy as the target joint rotations for a full set of PD controllers. The PD controllers adjust and output the joint torques to reach the target rotations.

3.4. Unified HOI Imitation Reward

This HOI imitation reward function is pivotal as it must not only accurately gauge the efficacy of interaction skill ac-
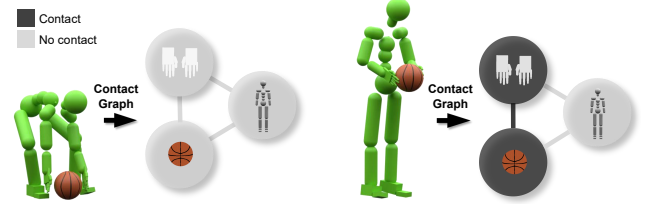

Figure 4. We propose the Contact Graph (CG) to model general contacts within an explicitly defined scene. The node stores a binary value that denotes whether it contacts other nodes. Each edge stores a binary value indicating whether the two connected nodes are in contact. The node definition is unified for a certain scene and shared between diverse interactive skills. For example, we define three nodes: hands, hands-exclusive body, and ball, to form a simple CG to model contacts for diverse basketball skills.

quisition but also avoid the incorporation of skill-specific settings to maintain robust cross-skill generalizability.

Classic imitation rewards often focus solely on imitating locomotions [32], lacking an established baseline for HOI imitation. Although adversarial imitation rewards [33, 34] are effective in locomotion tasks, we find that they perform poorly in HOI imitation. This may be because adversarial imitation rewards are relatively coarse-grained, making it difficult for them to provide the precise guidance needed for accurate interaction learning.

The body kinematics imitation reward r_t^b encourages the

alignment of body movements by considering joint position, rotation, and their velocities. Similarly, the object kinematics imitation reward r_t^o ensures consistent object movements with the reference. However, solely relying on these two rewards often leads to unstable learning and unnatural motions. We incorporate a relative motion reward r_t^{rel} to highlight the relative relationships in interactions, which is crucial for interaction naturalness. Recognizing the interlocking relationship between interaction and contact, we propose a straightforward and universal contact modeling approach called Contact Graph (CG). A corresponding CG imitation reward r_t^{cg} is designed to enhance the precision of interaction imitation. Additionally, we propose an adaptive velocity regularization term r_t^{reg} to suppress high-frequency jitters. These sub-rewards are multiplied rather than added to encourage balanced learning, thereby avoiding local optima, as justified in Tab. 1. The combination of these innovations forms a unified imitation reward, enabling the learning of diverse basketball skills in BallPlay-M with the same configuration. To conclude, the complete HOI imitation reward is

$$r_t = r_t^b * r_t^o * r_t^{rel} * r_t^{reg} * r_t^{cg}. \qquad (2)$$

Next, we will discuss the methodology involved in crafting the CG reward r_t^{cg}. Details about the kinematic imitation reward $r_t^b * r_t^o * r_t^{rel} * r_t^{reg}$ are in the appendix.

3.4.1. Contact Graph

We observe that existing imitation rewards [33, 34, 52] only measure kinematic properties and are insufficient to measure precise contacts. To tackle this problem, we propose the Contact Graph (CG) to model the contact in general interactions and design a CG imitation reward.

As illustrated in Fig. 4, the CG is a complete graph where every pair of distinct nodes is connected by a unique edge, defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} \in \{0,1\}^k$ is the set of k nodes and $\mathcal{E} \in \{0,1\}^{k(k-1)/2}$ is the set of edges. A CG node stores a binary value indicating whether it contacts other nodes. Each CG edge stores a binary label that denotes the contact between two nodes, where 1 represents contact, and 0 means no contact. CG node and edge values are calculated frame by frame, explicitly describing the mutual contact relationship between CG nodes at different moments. Contacts inside a node are not considered. We use the edge set $\mathcal{E}$ for imitation learning.

The definitions of nodes are flexible. For example, a node can be a single part (e.g., a fingertip), or an aggregation of multiple parts (e.g., the whole left hand). Hinged objects can also be broken up into finer nodes [9, 13, 51]. The definition of CG is unified for a certain scene and shared between diverse interaction skills. For example, in basketball scenarios, we can aggregate two hands as a node, the rest of the bodies as a node, and the ball as a node, as illustrated in
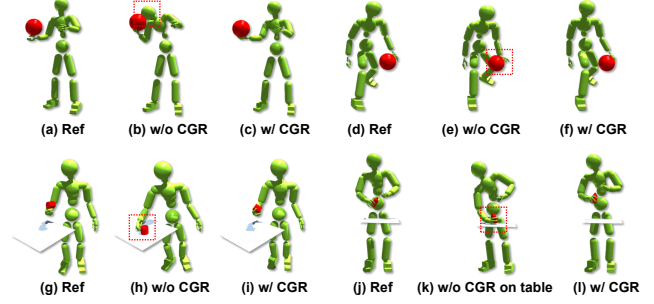

(a) Ref   (b) w/o CGR   (c) w/ CGR   (d) Ref   (e) w/o CGR   (f) w/ CGR

(g) Ref   (h) w/o CGR   (i) w/ CGR   (j) Ref   (k) w/o CGR on table   (l) w/ CGR

Figure 5. The HOI imitation falls into kinematic local-optimal solutions without Contact Graph Reward (CGR): (**b**) use the head to help control the ball; (**e**) use the wrist to contact the ball; (**h**) fail to catch the object; (**k**) support the table to keep balance. In comparison, the guidance of CGR effectively yields precise interactions, as shown in (**c**, **f**, **i**, **l**).

Fig. 4. This simple CG is effective enough for learning all the basketball skills covered in BallPlay-M.

3.4.2. Contact Graph Reward

Kinematic imitation rewards [32, 33, 52] fall short in measuring contacts and thus yield poor performance in HOI imitation. For example, during the toss skill training (Fig. 5 (a)), the contact between the right hand and the ball usually cause the ball to drop, making the kinematic rewards drastically smaller. In this case, the policy may learn a local-optimal solution that uses the head and left hands to stabilize the ball, as demonstrated in Fig. 5 (b). To tackle this problem, we propose the Contact Graph Reward (CGR) as a critical complement of kinematic imitation rewards to learn precise contact imitation. The CG error is defined as $e_t^{cg} = |s_t^{cg} - \hat{s}_t^{cg}|$, where $| \cdot |$ calculates element-wise absolute value, s_t^{cg} and $\hat{s}_t^{cg}$ denotes the simulated and reference CG state respectively, which is the edge set $\mathcal{E}_t \in \{0,1\}^{k(k-1)/2}$. The CGR is measured by CG error, with independent weights on different elements:

$$r_t^{cg} = \exp(-\sum_{j=1}^{J} \boldsymbol{\lambda}^{cg}[j] * \boldsymbol{e}_t^{cg}[j]), \qquad (3)$$

where $\boldsymbol{e}_t^{cg}[j]$ is the jth element of $\boldsymbol{e}_t^{cg} \in \{0,1\}^J$, representing a contact error; $\boldsymbol{\lambda}^{cg}[j]$ is the jth element of $\boldsymbol{\lambda}^{cg} \in \mathbb{R}^J$, a hyperparameter controls the sensitivity of a contact. Fig. 5 and Tab. 1 justify the effectiveness of CGR, where experiments without CGR show inaccurate interactions while using CGR effectively eliminates such problems.

3.5. Reusing Interaction Skills for High-level Tasks

In this section, we propose training a High-Level Controller (HLC) to reuse the interaction skills acquired through SkillMimic to accomplish high-level tasks. As depicted in Fig. 3 (c), the HLC takes as input the current HOI state s_t

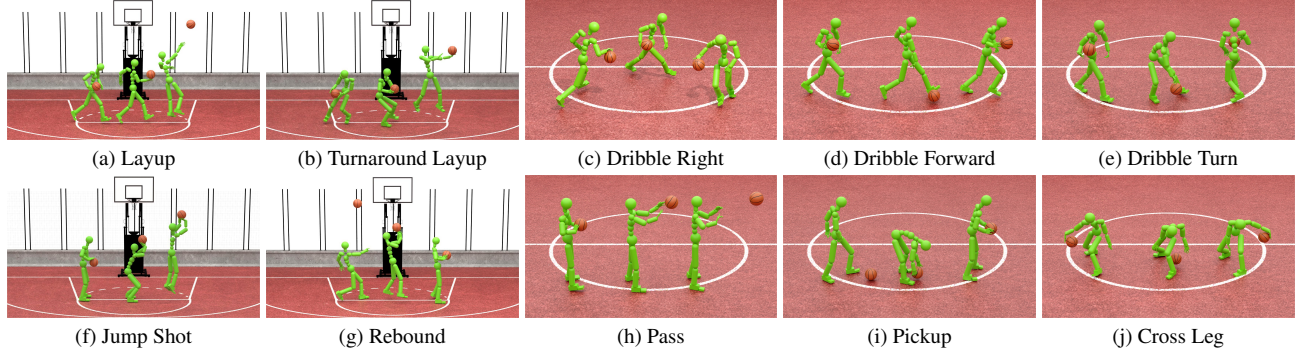| (a) Layup | (b) Turnaround Layup | (c) Dribble Right | (d) Dribble Forward | (e) Dribble Turn |
| (f) Jump Shot | (g) Rebound | (h) Pass | (i) Pickup | (j) Cross Leg |

Figure 6. Simulated humanoids exhibit comprehensive basketball skills. SkillMimic can teach humanoids a wide range of basketball skills using the same configuration in a purely data-driven manner, covering almost all fundamental basketball skills. Keyframes are placed in chronological order from left to right.

and task observation h_t, and outputs a discrete skill embedding c_t that serves as inputs to a pre-trained SkillMimic policy, which subsequently generates actions for the humanoid.

Compared to previous hierarchical approaches that are based on locomotion priors [34, 50], our method leverages interaction skills as prior, which significantly simplifies the design of task rewards and greatly accelerates training. Moreover, unlike their mandatory hierarchical structure, our approach is motivated by data efficiency - tasks can be alternatively solved through direct HOI imitation when sufficient demonstrations are available, as shown in pickup experiments in Sec. 4.2, Fig. 7, and Fig. 8.

We present four representative high-level tasks:

- **Throwing:** throw the ball to approach a certain height, grab the rebound, and keep on throwing the ball.
- **Heading:** dribble the ball to approach the target position.
- **Circling:** dribble the ball around the target position following a target radius.
- **Scoring:** dribble the ball toward the basket, timing the layup to score, retrieving the rebound, and repeating.

Utilizing interaction skills as prior significantly improves learning efficiency for complex, long-horizon tasks. For example, a decent HLC for the Scoring task only requires 3 hours of training on an NVIDIA RTX 4090 GPU.

4. Experiments

We conduct extensive experiments to evaluate the performance of SkillMimic in learning interaction skills and its application for high-level tasks. Our evaluation consists of two main parts: (1) assessing the capability of SkillMimic in learning interaction skills, including comparisons with existing methods and ablation studies, and (2) demonstrating the effectiveness of reusing acquired interaction skills to learn high-level tasks. Details of the experimental settings and additional experiments on skill switching and robustness are provided in the appendix.

4.1. Metrics

We consider the following quantitative metrics:

- **Accuracy.** The overall accuracy of HOI imitation, abbreviated as *Acc.*, is defined per frame, and deems imitation accurate when the object position and body position errors are both under the thresholds and the contacts are correct. The object threshold is defined as $0.2m$. The body threshold is defined as $0.1m$. The *Acc.* is calculated by averaging the values of all frames.
- **Position Error.** The Mean Per-Joint Position Error (MPJPE) of the body ($E_{\text{b-mpjpe}}$) and object ($E_{\text{o-mpjpe}}$) is used to evaluate the positional imitation performance (in mm) following [27].
- **Contact Error.** The contact error E_{cg}, ranging from 0 to 1, is defined as $\frac{1}{N}\sum_{t=1}^{N}\text{MSE}(s_t^{cg}, \hat{s}_t^{cg})$, where N is the total frames of the reference HOI data.
- **Success Rate.** The execution success rate of an interaction skill. Detailed definitions of success criteria for different skills are provided in the appendix.

4.2. Evaluating Interaction Skill Learning

Since our approach is the first to learn interaction skills from HOI demonstrations, there are no direct benchmarks for comparison. Therefore, we adapt reward strategies commonly used in locomotion imitation, i.e., DeepMimic [32] and AMP [33], to an object-inclusive setting. Specifically, we adapt our reward to the styles of DeepMimic [32] and AMP [33] while keeping the other components unchanged for a fair comparison. We denote these variant versions as DeepMimic* and AMP*. In Tab. 2, we present the success rate on 4 typical basketball skills from BallPlay-M, where our method achieves state-of-the-art performance in learning basketball skills compared with variants using DeepMimic or AMP style rewards. We guess that AMP-like rewards are ineffective for learning precise interactions due to their coarse-grained supervision signals.
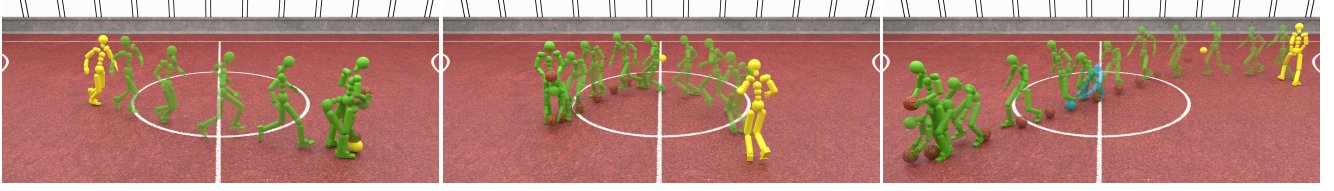
Figure 7. Demonstration of the pickup skill learned from 40 HOI motion clips. Yellow denotes the initial frame. Left: The humanoid picks up a stationary ball effortlessly. Middle: The humanoid intercepts a ball with random velocity. Right: The humanoid adjusts after missing the ball initially (the frame in blue) and successfully retrieves it on the second attempt, showcasing the potential for learning robust and generalizable skills through extensive data collection.

Dataset	SkillMimic w/o Multiplication				SkillMimic w/o CGR				SkillMimic			
	$Acc. \uparrow$	$E_{\text{b-mpjpe}} \downarrow$	$E_{\text{o-mpjpe}} \downarrow$	$E_{\text{cg}} \downarrow$	$Acc. \uparrow$	$E_{\text{b-mpjpe}} \downarrow$	$E_{\text{o-mpjpe}} \downarrow$	$E_{\text{cg}} \downarrow$	$Acc. \uparrow$	$E_{\text{b-mpjpe}} \downarrow$	$E_{\text{o-mpjpe}} \downarrow$	$E_{\text{cg}} \downarrow$
GRAB [40]	27.0%	**44.7**	180.2	0.724	38.6%	91.0	180.1	0.337	**95.4%**	71.1	**78.0**	**0.026**
BallPlay-V	7.5%	**40.1**	1662.5	0.306	13.6%	88.5	155.3	0.412	**82.4%**	56.8	**82.9**	**0.087**

Table 1. Ablation study on imitation reward design. We experiment on two HOI datasets: GRAB and BallPlay-V. The result highlights the importance of the CG reward and multiplication in achieving precise HOI imitation.

Method	Success Rate ↑			
	Pick Up	Dribble Forward	Layup	Shot
DeepMimic*[32]	19.6%	68.5%	98.9%	97.8%
AMP*[33]	0.0%	13.6%	0.0%	1.6%
SkillMimic (ours)	**86.7%**	**79.6%**	**99.1%**	**97.9%**

Table 2. Success rates across four typical basketball skills in BallPlay-M. Our method significantly outperforms variant methods that use imitation reward styles of DeepMimic and AMP.

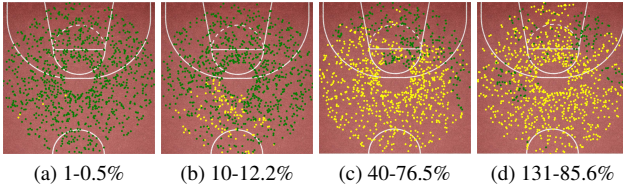(a) 1-0.5% (b) 10-12.2% (c) 40-76.5% (d) 131-85.6%

Figure 8. Pickup generalization performance with different training data scales. First number: clips; second: success rate. In each test, 1000 balls are randomly placed within 1 to 5 meters away from the center. Yellow dots indicate successful pickups and green dots represent failures. See Sec. 4.2 for details.

In Tab. 1, we do ablation on reward design, presenting detailed metrics of HOI imitation across two diverse datasets, GRAB [40] and BallPlay-V. The results indicate that the absence of reward multiplication leads to imbalanced sub-reward learning, e.g., favoring body motion ($E_{\text{b-mpjpe}}$) over interactions. The absence of CGR leads to kinematic local optima. In contrast, our complete setting (SkillMimic) yields significantly higher performance on interaction-related metrics ($Acc.$, $E_{\text{o-mpjpe}}$, and E_{cg}).

4.2.1. Ablation on Contact Graph Reward

We conduct ablation experiments on diverse datasets to comprehensively evaluate the impact of Contact Graph Reward (CGR) on interaction skill learning. Fig. 5 illustrates the qualitative results of ablating CGR on the GRAB and BallPlay-V datasets. It can be clearly observed that without CGR, the humanoid often resorts to incorrect contacts to achieve more stable object control, such as using its head to push the ball, clamping the ball with its hands and legs, or using its hands to support itself on a table for balance. These examples illustrate typical kinematic local optima. By incorporating CGR, these local optima are effectively eliminated, resulting in accurate interactions. Additionally, the result in Tab. 1 also shows that, without CGR, the contact error (E_{cg}) significantly increases, leading to kinematic local optima and causes a noticeable decline in overall imitation accuracy. These results underscore the critical importance of CGR for learning correct interactions.

4.2.2. Data Scale and Generalization Performance

To examine how the generalization performance grows as the data scale increases, we conduct ablation experiments on the pickup skill. We create four datasets by randomly selecting pickup clips from BallPlay-M, containing 1, 10, 40, and 131 clips, respectively. We train policy on each dataset with approximately 3.2 billion samples using SkillMimic.

To evaluate the generalization performance, we randomly place balls within a circular area ranging from 1 to 5 meters around the humanoid—this distribution significantly exceeds the coverage of the reference data. Fig. 7 and Fig. 8 show the results. We can see that as the data scale increases, pickup performance improves significantly due to broader state transition coverage in larger datasets. This indicates

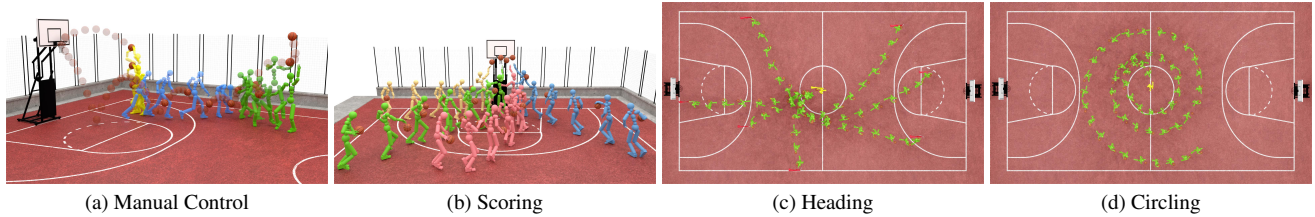| (a) Manual Control | (b) Scoring | (c) Heading | (d) Circling |

Figure 9. Our method supports training a single IS policy under a unified configuration to acquire various interaction skills. These interaction skills can be flexibly switched, as illustrated in (a), where yellow denotes shot, blue denotes pickup, and green denotes turnaround layup. Complex, long-horizon tasks can be easily achieved by training a high-level controller (HLC) to manage switching of the learned interaction skills: (b) scoring from random positions (c) dribbling to target locations, and (d) dribbling along an expanding radius.

Method	Success Rate↑			
	Heading	Circling	Throwing	Scoring
PPO[36]	0.70%	11.14%	0.00%	0.00%
ASE[34]	0.19%	4.37%	0.00%	0.00%
ASE*[34]	0.31%	7.21%	0.00%	0.00%
SkillMimic (ours)	**93.04%**	**79.92%**	**93.40%**	**80.25%**

Table 3. Success rates on 4 high-level basketball tasks. Both PPO (learn from scratch), ASE (using locomotion prior), and ASE* (using body and object motion prior) fail to converge on all tasks. In contrast, leveraging interaction skill priors acquired through SkillMimic, our method effectively learns these difficult tasks.

our method's strong scalability and potential for large-scale learning. It is worth noting that we do not incorporate any additional designs for robustness and generalization.

4.3. Evaluating High-level Tasks

We conduct experiments on four high-level tasks described in Sec. 3.5. We employ a single IS policy as the interaction skill prior, trained using SkillMimic for around 4.5 billion samples across seven skills: pickup, layup, turnaround layup, dribble left, dribble right, dribble forward, and shot. During training, the IS policy is fixed, and only the HLC is trained. For tasks such as throwing, heading, and circling, we train for around 0.4 billion samples. For the more challenging scoring task, we train for around 1.2 billion samples. To objectively evaluate the performance of our method, we compared it against three baseline methods: (1) learning from scratch using PPO [36], (2) learning with body motion priors using ASE [34], and (3) learning with body and object motion priors using ASE (denoted as ASE*). The low-level controllers of ASE and ASE* are trained using the same training steps as our IS policy. For task training, all methods are trained using identical task rewards and simulation steps for fair comparisons.

Fig. 9 (a) showcases an example of manually controlling the pre-trained IS policy to perform skill switching. Fig. 9 (b) shows several trajectories of the HLC performing skill controls to accurately deliver the ball to the target basket. Fig. 9 (c) & (d) present a top-down view of the HLC control-

ling the humanoid to dribble the ball toward target locations and along an expanding radius. As shown in Tab. 3, PPO and ASE yield low success rates because the tasks are too challenging to master without interaction priors. The failure of ASE* can be traced back to its low-level controller's inadequate ability to imitate interaction skills. This is primarily because adversarial rewards fail to provide effective guidance for precise interaction learning, which is similar to the failure mode observed in AMP*. In contrast, by utilizing the interaction skills learned through SkillMimic, we greatly improve the performance of learning complex high-level interaction tasks.

5. Discussion and Future Work

We presented a novel framework for learning diverse interaction skills with a unified HOI imitation reward using reinforcement learning. Our approach demonstrates multi-skill mastery, skill reusability, and clear data scaling benefits, highlighting the significant potential of HOI imitation for learning complex interaction skills. However, the full application of this paradigm requires further exploration. Several key areas deserve particular attention:

Firstly, exploring more general HOI skill learning within a single policy, such as household tasks or other types of sports, presents challenges in handling generalization across different object geometries. Moreover, perceiving multiple objects simultaneously requires a more general environmental perception approach, rather than directly obtaining privileged information for each object.

Secondly, incorporating finer-grained control conditions, such as tracking [27], could enable more flexible control and generalization capabilities as HOI data scales up.

Thirdly, the difficulty of collecting HOI data highlights the importance of exploring methods to learn robust interaction skills with limited HOI data, as well as investigating HOI data augmentation and generation techniques.

Lastly, applying this technology to real robots would be highly valuable, such as enabling real-world humanoid robots to play basketball games. This endeavor could involve various challenges, including HOI data retargeting and sim-to-real transfer issues.

Acknowledgment

References

[1] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017. 2

[2] Jinseok Bae, Jungdam Won, Donggeun Lim, Cheol-Hui Min, and Young Min Kim. Pmp: Learning to physically interact with environments using part-wise motion priors. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–10, 2023. 1, 2

[3] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. Pi0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024. 3

[4] Jona Braun, Sammy Christen, Muhammed Kocabas, Emre Aksan, and Otmar Hilliges. Physically plausible full-body hand-object interaction synthesis. *arXiv preprint arXiv:2309.07907*, 2023. 2

[5] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023. 3

[6] Stelian Coros, Philippe Beaudoin, and Michiel Van de Panne. Generalized biped walking control. *ACM Transactions On Graphics (TOG)*, 29(4):1–9, 2010. 2

[7] Jieming Cui, Tengyu Liu, Nian Liu, Yaodong Yang, Yixin Zhu, and Siyuan Huang. Anyskill: Learning openvocabulary physical skill for interactive agents. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[8] Zhiyang Dou, Xuelin Chen, Qingnan Fan, Taku Komura, and Wenping Wang. C· ase: Learning conditional adversarial skill embeddings for physics-based characters. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023. 1, 2

[9] Zicong Fan, Omid Taheri, Dimitrios Tzionas, Muhammed Kocabas, Manuel Kaufmann, Michael J Black, and Otmar Hilliges. Arctic: A dataset for dexterous bimanual hand-object manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12943–12954, 2023. 5

[10] Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. Humanplus: Humanoid shadowing and imitation from humans. In *8th Annual Conference on Robot Learning*, 2024. 1

[11] Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024. 3

[12] Jiawei Gao, Ziqin Wang, Zeqi Xiao, Jingbo Wang, Tai Wang, Jinkun Cao, Xiaolin Hu, Si Liu, Jifeng Dai, and Jiangmiao Pang. Coohoi: Learning cooperative human-object interaction with manipulated object dynamics. *Advances in Neural Information Processing Systems*, 37:79741–79763, 2024. 1

[13] Haoran Geng, Helin Xu, Chengyang Zhao, Chao Xu, Li Yi, Siyuan Huang, and He Wang. Gapartnet: Cross-category domain-generalizable object perception and manipulation via generalizable and actionable parts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7081–7091, 2023. 5

[14] Anindita Ghosh, Rishabh Dabral, Vladislav Golyanik, Christian Theobalt, and Philipp Slusallek. Imos: Intent-driven full-body motion synthesis for human-object interactions. In *Eurographics*, 2023. 2

[15] Mohamed Hassan, Yunrong Guo, Tingwu Wang, Michael Black, Sanja Fidler, and Xue Bin Peng. Synthesizing physical character-scene interactions. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–9, 2023. 1, 2

[16] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris M Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. In *8th Annual Conference on Robot Learning*, 2024. 1

[17] Edmond SL Ho, Taku Komura, and Chiew-Lan Tai. Spatial relationship preserving character motion adaptation. In *ACM SIGGRAPH 2010 papers*, pages 1–8, 2010. 3

[18] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. *Advances in neural information processing systems*, 29, 2016. 2

[19] Jessica K Hodgins, Wayne L Wooten, David C Brogan, and James F O'Brien. Animating human athletics. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 71–78, 1995. 2

[20] Nan Jiang, Tengyu Liu, Zhexuan Cao, Jieming Cui, Zhiyuan Zhang, Yixin Chen, He Wang, Yixin Zhu, and Siyuan Huang. Full-body articulated human-object interaction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9365–9376, 2023. 2

[21] Nan Jiang, Zhiyuan Zhang, Hongjie Li, Xiaoxuan Ma, Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, and Siyuan Huang. Scaling up dynamic human-scene interaction modeling. *arXiv preprint arXiv:2403.08629*, 2024. 2

[22] Jordan Juravsky, Yunrong Guo, Sanja Fidler, and Xue Bin Peng. Padl: Language-directed physics-based character control. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 2

[23] Jiaman Li, Alexander Clegg, Roozbeh Mottaghi, Jiajun Wu, Xavier Puig, and C Karen Liu. Controllable human-object interaction synthesis. *arXiv preprint arXiv:2312.03913*, 2023. 1, 2

[24] Jiaman Li, Jiajun Wu, and C Karen Liu. Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)*, 42(6):1–11, 2023. 2

[25] Libin Liu and Jessica Hodgins. Learning to schedule control fragments for physics-based characters using deep q-learning. *ACM Transactions on Graphics (TOG)*, 36(3):1–14, 2017. 2

[26] Libin Liu and Jessica Hodgins. Learning basketball dribbling skills using trajectory optimization and deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018. 2

[27] Zhengyi Luo, Jinkun Cao, Kris Kitani, Weipeng Xu, et al. Perpetual humanoid control for real-time simulated avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10895–10904, 2023. 6, 8

[28] Zhengyi Luo, Jinkun Cao, Sammy Christen, Alexander Winkler, Kris Kitani, and Weipeng Xu. Omnigrasp: Grasping diverse objects with simulated humanoids. *Advances in Neural Information Processing Systems*, 37:2161–2184, 2024. 1

[29] Zhengyi Luo, Jiashun Wang, Kangni Liu, Haotian Zhang, Chen Tessler, Jingbo Wang, Ye Yuan, Jinkun Cao, Zihui Lin, Fengyi Wang, et al. Smplolympics: Sports environments for physically simulated humanoids. *arXiv preprint arXiv:2407.00187*, 2024. 2

[30] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010. 4

[31] Liang Pan, Jingbo Wang, Buzhen Huang, Junyu Zhang, Haofan Wang, Xu Tang, and Yangang Wang. Synthesizing physically plausible human motions in 3d scenes. In *2024 International Conference on 3D Vision (3DV)*, pages 1498–1507. IEEE, 2024. 2

[32] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. Deepmimic. *ACM Transactions on Graphics*, page 1–14, 2018. 1, 2, 4, 5, 6, 7

[33] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics*, page 1–20, 2021. 2, 3, 4, 5, 6, 7

[34] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Trans. Graph.*, 41(4), 2022. 1, 2, 3, 4, 5, 6, 8

[35] Jiawei Ren, Mingyuan Zhang, Cunjun Yu, Xiao Ma, Liang Pan, and Ziwei Liu. Insactor: Instruction-driven physics-based characters. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[36] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 8

[37] Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. Local motion phases for learning multi-contact character movements. *ACM Trans. Graph.*, 39(4), 2020. 2

[38] Sebastian Starke, Yiwei Zhao, Fabio Zinno, and Taku Komura. Neural animation layering for synthesizing martial arts movements. *ACM Trans. Graph.*, 40(4), 2021. 2

[39] Jingkai Sun, Qiang Zhang, Yiqun Duan, Xiaoyang Jiang, Chong Cheng, and Renjing Xu. Prompt, plan, perform: Llm-based humanoid control via quantized imitation learning. *arXiv preprint arXiv:2309.11359*, 2023. 2

[40] Omid Taheri, Nima Ghorbani, Michael J. Black, and Dimitrios Tzionas. GRAB: A dataset of whole-body human grasping of objects. In *European Conference on Computer Vision (ECCV)*, 2020. 7

[41] Jie Tan, Yuting Gu, C Karen Liu, and Greg Turk. Learning bicycle stunts. *ACM Transactions on Graphics (TOG)*, 33 (4):1–12, 2014. 2

[42] Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik, and Xue Bin Peng. Calm: Conditional adversarial latent models for directable virtual characters. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–9, 2023. 1, 2

[43] Yinhuai Wang, Jing Lin, Ailing Zeng, Zhengyi Luo, Jian Zhang, and Lei Zhang. Physhoi: Physics-based imitation of dynamic human-object interaction. *arXiv preprint arXiv:2312.04393*, 2023. 3

[44] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Physics-based character controllers using conditional vaes. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 2

[45] Yan Wu, Jiahao Wang, Yan Zhang, Siwei Zhang, Otmar Hilliges, Fisher Yu, and Siyu Tang. Saga: Stochastic whole-body grasping with contact. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 2

[46] Zeqi Xiao, Tai Wang, Jingbo Wang, Jinkun Cao, Wenwei Zhang, Bo Dai, Dahua Lin, and Jiangmiao Pang. Unified human-scene interaction via prompted chain-of-contacts. In *The Twelfth International Conference on Learning Representations*, 2024. 1

[47] Sirui Xu, Zhengyuan Li, Yu-Xiong Wang, and Liang-Yan Gui. Interdiff: Generating 3d human-object interactions with physics-informed diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14928–14940, 2023. 1, 2

[48] Sirui Xu, Ziyin Wang, Yu-Xiong Wang, and Liang-Yan Gui. Interdreamer: Zero-shot text to 3d dynamic human-object interaction. *arXiv preprint arXiv:2403.19652*, 2024. 2

[49] KangKang Yin, Kevin Loken, and Michiel Van de Panne. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics (TOG)*, 26(3):105–es, 2007. 2

[50] Haotian Zhang, Ye Yuan, Viktor Makoviychuk, Yunrong Guo, Sanja Fidler, Xue Bin Peng, and Kayvon Fatahalian. Learning physically simulated tennis skills from broadcast videos. *ACM Trans. Graph.*, 2023. 1, 2, 3, 6

[51] Hui Zhang, Sammy Christen, Zicong Fan, Luocheng Zheng, Jemin Hwangbo, Jie Song, and Otmar Hilliges. Artigrasp: Physically plausible synthesis of bi-manual dexterous grasping and articulation. In *2024 International Conference on 3D Vision (3DV)*, pages 235–246. IEEE, 2024. 5

[52] Yunbo Zhang, Deepak Gopinath, Yuting Ye, Jessica Hodgins, Greg Turk, and Jungdam Won. Simulation and retargeting of complex multi-character interactions. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 3, 5

[53] Qingxu Zhu, He Zhang, Mengting Lan, and Lei Han. Neural categorical priors for physics-based character control. *ACM Transactions on Graphics (TOG)*, 42(6):1–16, 2023. 2