CVPR
#*****

Learning to Ball: Composing Policies for Long-Horizon Basketball Moves

Anonymous CVPR submission

Paper ID *****

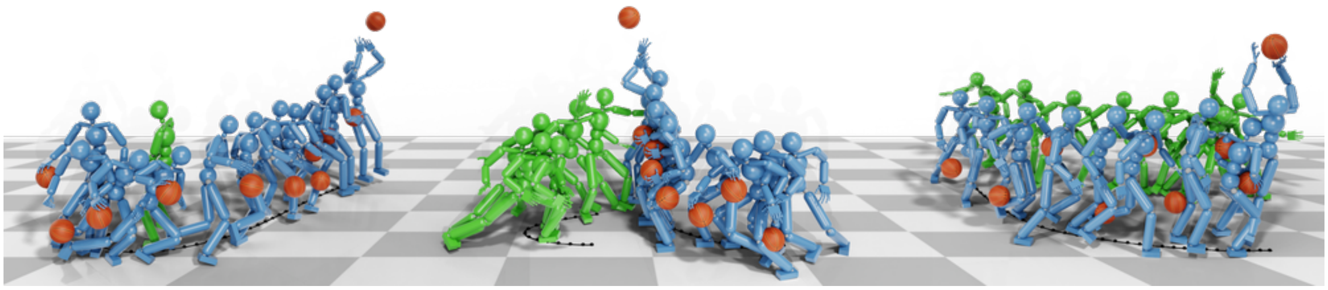

Figure 1. We introduce a novel policy integration framework to enable the composition of drastically different motor skills in multi-phase, long-horizon tasks, among them, shoot-off-the-dribble, catch-and-shoot, and board-and-bang.

Abstract

Learning a control policy for a multi-phase, long-horizon task, such as basketball maneuvers, remains challenging for reinforcement learning approaches due to the need for seamless policy composition and transitions between skills. A long-horizon task typically consists of distinct subtasks with well-defined goals, separated by transitional subtasks with unclear goals but critical to the success of the entire task. Existing methods like the mixture of experts and skill chaining struggle with tasks where individual policies do not share significant commonly explored states or lack well-defined initial and terminal states between different phases. In this paper, we introduce a novel policy integration framework to enable the composition of drastically different motor skills in multi-phase long-horizon tasks with ill-defined intermediate states. Based on that, we further introduce a high-level soft router to enable seamless and robust transitions between the subtasks. We evaluate our framework on a set of fundamental basketball skills and challenging transitions. Policies trained by our approach can effectively control the simulated character to interact with the ball and accomplish the long-horizon task specified by real-time user commands, without relying on ball trajectory references.

1. Introduction

Many real-world tasks consist of multiple subtasks that require heterogeneous skills and seamless transitions between them. Basketball exemplifies this challenge: a maneuver such as *shoot-off-the-dribble* involves dribbling, ball gathering, and shooting in sequence. While *dribble* and *shoot* have clear standalone objectives, *gather* primarily serves as a transition phase with poorly defined initial and terminal states. Thus, executing such multi-phase tasks challenges control methods proposed to date.

Reinforcement learning (RL) has achieved promising results for individual physics-based control skills [3, 7, 17, 24, 40], but composing heterogeneous policies into coherent long-horizon behaviors remains challenging. Mixture-of-experts approaches [18, 30] rely on overlapping state distributions between skills, while skill-chaining methods [2, 4, 6, 8, 10] require well-defined terminal states for each subtask. These assumptions break down for intermediate subtasks such as gathering, whose objective is implicitly defined by the requirements of the succeeding policy.

We introduce a policy integration framework for composing drastically different skills in long-horizon tasks with ill-defined intermediate phases. Given subtasks A, B, and C, where A and C have well-defined objectives but B does not, we train B using states generated by A as its initial-state distribution and the state value function of C as a reward-shaping signal. Simultaneously, policy C is adapted to states produced by B, enabling robust transitions between subtasks. On top of the primitive skills, we further train a high-level soft-routing policy that coordinates skill execution according to real-time user commands.

In addition, our framework learns basketball behaviors from heterogeneous and unstructured motion data, includ-

CVPR
#*****

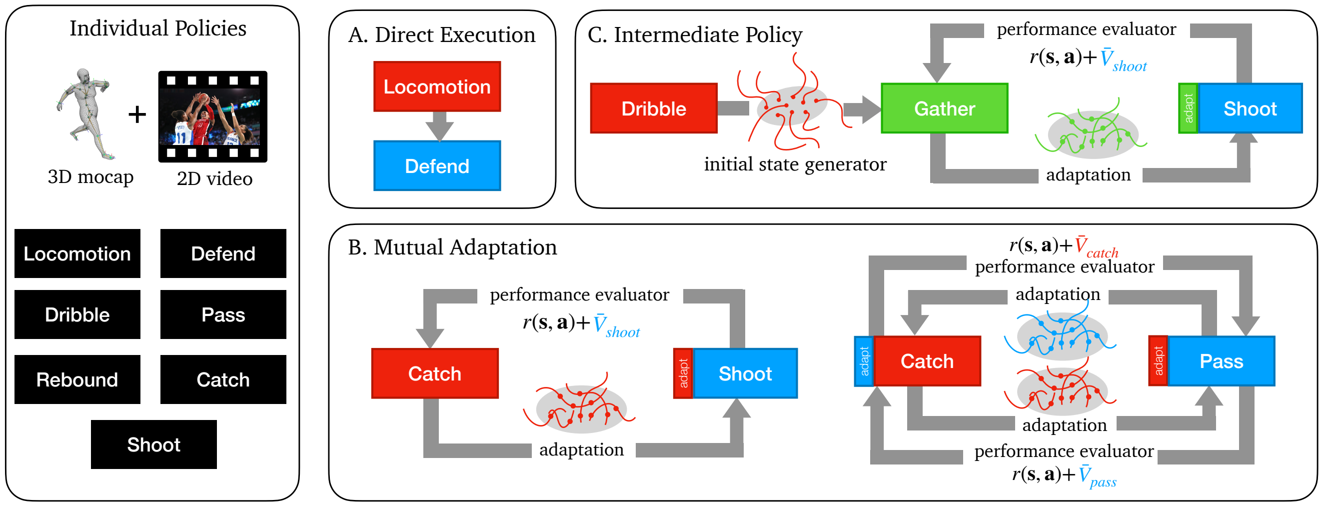

Figure 2. Overview of our policy integration framework. Type A transitions are achieved through direct execution. Type B transitions use mutual adaptation and value-based reward shaping. Type C transitions additionally require an intermediate policy to bridge incompatible subtasks.

ing full-body mocap without hands, hand-only motion capture, and motions extracted from videos, without requiring correspondence between datasets or ball trajectory supervision. The resulting controller enables interactive basketball behaviors such as dribbling at arbitrary speeds and directions, transitioning smoothly into jump shots with a shooting accuracy of 91.8%. We further demonstrate multi-agent interactions including passing, catching, rebounding, and defending.

2. Related Work

Creating robust controllers for long-horizon, multi-phase tasks remains a core challenge in RL and character animation. Our work mainly builds on two areas of research: deep RL for physics-based character control and policy composition for executing multi-phase tasks.

Recent deep RL approaches have demonstrated strong performance for physics-based character control across diverse domains, including locomotion [16, 17, 33], sports [5, 11, 26, 29, 42], instrument performance [12, 27, 36, 41], and object manipulation [1, 31, 34, 38]. Adversarial imitation learning methods [9, 14, 20, 21, 30, 35, 37, 39, 43] further enable robust motion synthesis from diverse motion sources, including mocap and monocular videos [18]. Our approach follows this paradigm and learns basketball behaviors from heterogeneous and partially observable motion data without requiring ball trajectory supervision or correspondence across datasets.

Long-horizon behaviors often require chaining multiple primitive skills into coherent multi-phase executions. Existing methods typically assume overlapping state distributions between consecutive policies [10, 15, 28, 29, 32, 35] or rely on well-defined terminal states through skill chaining and state-distribution matching [2, 4, 6, 8]. However, these assumptions break down for intermediate subtasks with ill-defined objectives, such as gathering between dribbling and shooting. Hierarchical control approaches further explore weighted blending [19, 22] or hard routing [25, 26, 29] between primitive policies, but often suffer from unnatural transitions. In contrast, our method introduces value-guided intermediate policy learning together with a soft-routing mechanism for seamless transitions between heterogeneous basketball skills.

3. Method

We present a policy integration framework for composing heterogeneous basketball skills into long-horizon behaviors. As illustrated in Fig. 2, we consider three transition types: direct execution, mutual adaptation, and intermediate-policy transitions. We focus on the most challenging case, where an additional policy is required to bridge incompatible subtasks such as dribbling and shooting.

3.1. Primitive Skill Learning

We train primitive basketball skills using PPO [23] combined with adversarial imitation learning [35, 37]. To support diverse basketball behaviors, we leverage heterogeneous motion sources including monocular videos, full-body mocap, and hand-only mocap. Following prior work [11, 37], body and hand motions are decoupled during training, enabling learning without requiring correspondence across datasets or ball trajectory supervision.

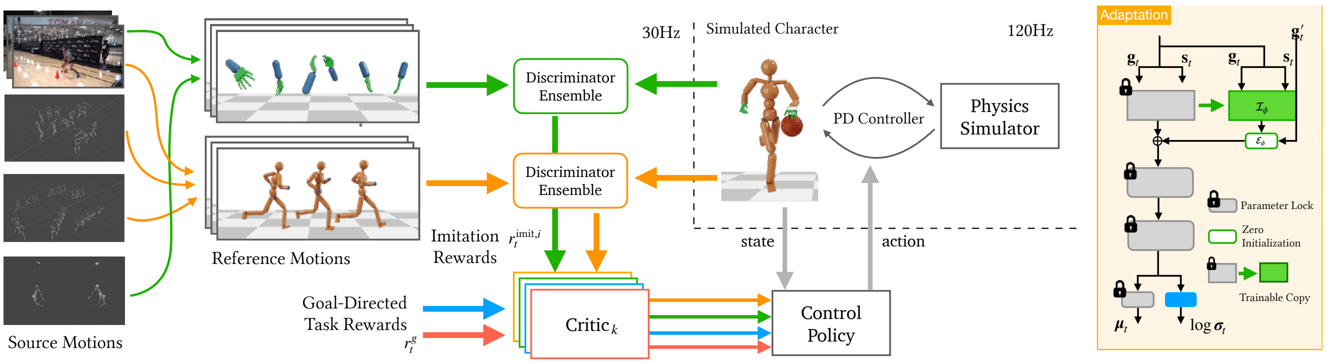

Figure 3. Primitive skill learning from heterogeneous motion sources. We combine monocular videos, full-body mocap, and hand-only mocap using adversarial imitation learning and body-part decomposition.

We train primitive policies for dribbling, shooting, passing, catching, locomotion, rebounding, and defending. The dribbling policy receives the current character and ball states together with a target velocity command, and is trained using imitation objectives and task rewards for navigation and ball control. The shooting policy is trained from full-body shooting demonstrations with rewards encouraging shot accuracy and stable ball holding.

3.2. Intermediate Policy Learning

Transitions such as *shoot-off-the-dribble* require an additional gathering policy to bridge incompatible dribbling and shooting states. Instead of defining explicit terminal states for dribbling, we sample random states from dribbling rollouts as initialization for the gathering policy. To encourage the gathering policy to reach states preferred by the shooting policy, we use the shooting value function as reward shaping:

$$r_{\text{gather}} = r_{\text{pose}} + 0.25\,\textsc{Clip}(\bar{V}_{\text{shoot}}(\mathbf{s}_t, \mathbf{g}_t), -v, v) \qquad (1)$$

Here r_{pose} encourages stable ball holding and body orientation toward the hoop, while $\bar{V}_{\text{shoot}}$ estimates future shooting performance from the current state.

To improve robustness, we simultaneously adapt the shooting policy during gathering-policy training. States generated by the gathering policy are reused as initialization states for shooting-policy adaptation, enabling both policies to gradually align their state distributions.

3.3. High-Level Policy Composition

To coordinate primitive skills, we introduce a high-level routing policy that softly combines actions from multiple policies:

$$\boldsymbol{\omega}_t = \mathbf{c}_t + \pi_c(\mathbf{s}_t, \mathbf{g}_t, \mathbf{c}_t), \qquad \mathbf{a}_t = \boldsymbol{\omega}_t \cdot \mathcal{A}_t, \qquad (2)$$

where $\mathbf{c}_t$ is a reference command, π_c is the routing policy, and $\mathcal{A}_t$ contains the actions from primitive policies. Unlike hard policy switching, the router performs soft blending while encouraging one policy to dominate at each timestep, producing stable and natural transitions.

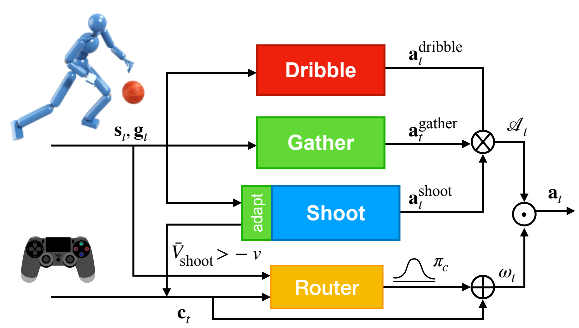

Figure 4. High-level routing policy for composing primitive skills.

The routing policy is optimized using rewards from the active primitive policy together with an additional switching reward encouraging successful transitions from gathering to shooting. The same framework generalizes to other basketball transitions including passing, catching, rebounding, and defending.

4. Experiments

We evaluate our policy integration framework on long-horizon basketball control tasks including shoot-off-the-dribble, passing, catching, rebounding, and defending. While we provide qualitative results for all transition types, our quantitative evaluation focuses on the most challenging Type C transition from dribbling to shooting, which requires an intermediate gathering policy (Fig. 2). All experiments are conducted in IsaacGym [13] using PPO [23].

We compare our method against three baselines:

- **DirectExecution**: directly switching from dribbling to the pretrained shooting policy.
- **NoAdapt**: introducing a gathering policy without simultaneous shooting-policy adaptation.
- **SequentialChaining**: jointly training gathering and

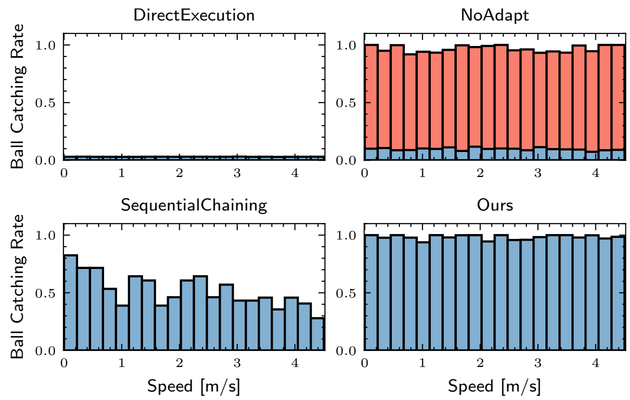

Figure 5. Ball catching rates under different dribbling speeds. Our method achieves consistently robust transitions across dynamic dribbling states.

Table 1. Shot percentage under different approaching directions. The overall shot percentage is 91.8%.

	Facing	Opposite	Left Orth.	Right Orth.
Shot Percentage	95.4%	90.4%	92.7%	92.4%

shooting as a single policy.

4.1. Quantitative Results

Ball Catching Rate. Figure 5 compares ball catching rates across methods. DirectExecution fails almost entirely due to the incompatibility between dribbling and shooting states, achieving only a 0.7% catching rate. SequentialChaining improves transition robustness but struggles to separate gathering and shooting behaviors effectively. NoAdapt successfully gathers the ball in many cases, but the pretrained shooting policy fails to generalize after policy switching. In contrast, our method achieves a 98.3% catching rate across all dribbling speeds. The result highlights the importance of explicit phase decomposition, value-guided reward shaping, and simultaneous policy adaptation for robust long-horizon skill transitions.

Shot Percentage. We evaluate shooting performance across different court positions and approaching directions. Similar to training, valid shooting locations are sampled between 2.5m and 7.5m from the hoop. DirectExecution, NoAdapt, and SequentialChaining achieve overall shot percentages of 1.3%, 6.1%, and 12.7%, respectively, demonstrating the difficulty of transitioning from dynamic dribbling states to shooting. In contrast, our method achieves 91.8% overall shooting accuracy, closely matching the pretrained shooting policy initialized from reference motions. Table 1 further shows robust performance across different approaching directions, including challenging cases where the character initially faces away from the hoop. The learned gathering policy enables dynamic pose adjustment before shooting, including pivoting, spinning, and body reorientation toward the hoop.

Figure 6. Real-time 2-on-2 basketball interaction generated by our framework. Offensive and defensive players are simultaneously controlled through composed primitive policies.

4.2. Qualitative Results

Beyond shoot-off-the-dribble, our framework supports multi-agent basketball interactions including passing, catching, rebounding, and defending, as shown in Figure 1. Figure 6 demonstrates real-time 2-on-2 gameplay generated entirely through composed primitive skills. The learned policies support arbitrary user-controlled dribbling directions and speeds while maintaining stable ball interaction and smooth transitions between offensive and defensive behaviors.

5. Conclusion

We presented a policy integration framework for composing heterogeneous skills in long-horizon basketball tasks. By combining value-guided intermediate policy learning, simultaneous policy adaptation, and high-level soft routing, our method enables robust transitions between drastically different subtasks such as dribbling and shooting. Our framework further supports multi-agent basketball interactions including passing, catching, rebounding, and defending. Our results demonstrate that complex basketball behaviors can be learned from heterogeneous motion sources without requiring ball trajectory supervision or correspondence across datasets. Future work includes improving motion realism through richer motion data and biomechanical character models, as well as extending the framework toward autonomous multi-agent gameplay.

References

[1] 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. 2

[2] Yuanpei Chen, Chen Wang, Li Fei-Fei, and C Karen Liu. Sequential dexterity: Chaining dexterous policies for long-horizon manipulation. *arXiv preprint arXiv:2309.00987*, 2023. 1, 2

[3] Nuttapong Chentanez, Matthias Müller, Miles Macklin, Viktor Makoviychuk, and Stefan Jeschke. Physics-based motion capture imitation with deep reinforcement learning. In *Proceedings of the 11th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–10, 2018. 1

[4] Alexander Clegg, Wenhao Yu, Jie Tan, C Karen Liu, and Greg Turk. Learning to dress: Synthesizing human dressing motion via deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 37(6):1–10, 2018. 1, 2

[5] Minsu Kim, Eunho Jung, and Yoonsang Lee. Physicsfc: Learning user-controlled skills for a physics-based football player controller. *arXiv preprint arXiv:2504.21216*, 2025. 2

[6] George Konidaris and Andrew Barto. Skill discovery in continuous reinforcement learning domains using skill chaining. *Advances in neural information processing systems*, 22, 2009. 1, 2

[7] Ariel Kwiatkowski, Eduardo Alvarado, Vicky Kalogeiton, C Karen Liu, Julien Pettré, Michiel van de Panne, and Marie-Paule Cani. A survey on reinforcement learning methods in character animation. In *Computer Graphics Forum*, pages 613–639. Wiley Online Library, 2022. 1

[8] Youngwoon Lee, Joseph J Lim, Anima Anandkumar, and Yuke Zhu. Adversarial skill chaining for long-horizon robot manipulation via terminal state regularization. *arXiv preprint arXiv:2111.07999*, 2021. 1, 2

[9] Hung Yu Ling, Fabio Zinno, George Cheng, and Michiel van de Panne. Character controllers using motion vaes. *ACM Trans. Graph.*, 39(4), 2020. 2

[10] 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. 1, 2

[11] 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

[12] Chaoyi Luo, Pengbin Tang, Yuqi Ma, and Dongjin Huang. Learning to play guitar with robotic hands. In *Computer Graphics Forum*, page e15166. Wiley Online Library, 2024. 2

[13] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. Isaac gym: High performance gpu-based physics simulation for robot learning, 2021. 3

[14] Josh Merel, Yuval Tassa, Dhruva TB, Sriram Srinivasan, Jay Lemmon, Ziyu Wang, Greg Wayne, and Nicolas Heess. Learning human behaviors from motion capture by adversarial imitation. *arXiv preprint arXiv:1707.02201*, 2017. 2

[15] 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

[16] Xue Bin Peng, Glen Berseth, KangKang Yin, and Michiel Van De Panne. Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *Acm transactions on graphics (tog)*, 36(4):1–13, 2017. 2

[17] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018. 1, 2

[18] Xue Bin Peng, Angjoo Kanazawa, Jitendra Malik, Pieter Abbeel, and Sergey Levine. Sfv: Reinforcement learning of physical skills from videos. *ACM Transactions On Graphics (TOG)*, 37(6):1–14, 2018. 1, 2

[19] Xue Bin Peng, Michael Chang, Grace Zhang, Pieter Abbeel, and Sergey Levine. Mcp: Learning composable hierarchical control with multiplicative compositional policies. In *Advances in Neural Information Processing Systems 32*, pages 3681–3692. Curran Associates, Inc., 2019. 2

[20] 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 (ToG)*, 40(4):1–20, 2021. 2

[21] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions On Graphics (TOG)*, 41(4):1–17, 2022. 2

[22] Avinash Ranganath, Pei Xu, Ioannis Karamouzas, and Victor Zordan. Low dimensional motor skill learning using coactivation. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–10, 2019. 2

[23] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. 2, 3

[24] Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. An hci-centric survey and taxonomy of human-generative-ai interactions. *arXiv preprint arXiv:2310.07127*, 2023. 1

[25] Chen Tessler, Shahar Givony, Tom Zahavy, Daniel Mankowitz, and Shie Mannor. A deep hierarchical approach to lifelong learning in minecraft. In *Proceedings of the AAAI conference on artificial intelligence*, 2017. 2

[26] Jiashun Wang, Jessica Hodgins, and Jungdam Won. Strategy and skill learning for physics-based table tennis animation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 2

[27] Ruocheng Wang, Pei Xu, Haochen Shi, Elizabeth Schumann, and C. Karen Liu. FürElise: Capturing and physically synthesizing hand motions of piano performance. In *SIGGRAPH Asia 2024 Conference Papers*, New York, NY, USA, 2024. Association for Computing Machinery. 2

[28] Wenjia Wang, Liang Pan, Zhiyang Dou, Zhouyingcheng Liao, Yuke Lou, Lei Yang, Jingbo Wang, and Taku Komura. Sims: Simulating human-scene interactions with real world script planning. *arXiv preprint arXiv:2411.19921*, 2024. 2

[29] Yinhuai Wang, Qihan Zhao, Runyi Yu, Ailing Zeng, Jing Lin, Zhengyi Luo, Hok Wai Tsui, Jiwen Yu, Xiu Li, Qifeng Chen, et al. Skillmimic: Learning reusable basketball skills from demonstrations. *arXiv preprint arXiv:2408.15270*, 2024. 2

[30] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. A scalable approach to control diverse behaviors for physically simulated characters. *ACM Transactions on Graphics*, 39(4): 33–1, 2020. 1, 2

[31] Zhen Wu, Jiaman Li, and C Karen Liu. Human-object interaction from human-level instructions. *arXiv preprint arXiv:2406.17840*, 2024. 2

[32] 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. *arXiv preprint arXiv:2309.07918*, 2023. 2

[33] Zhaoming Xie, Hung Yu Ling, Nam Hee Kim, and Michiel van de Panne. Allsteps: curriculum-driven learning of stepping stone skills. In *Computer Graphics Forum*, pages 213–224. Wiley Online Library, 2020. 2

[34] Zhaoming Xie, Jonathan Tseng, Sebastian Starke, Michiel van de Panne, and C Karen Liu. Hierarchical planning and control for box loco-manipulation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(3):1–18, 2023. 2

[35] Pei Xu and Ioannis Karamouzas. A gan-like approach for physics-based imitation learning and interactive character control. *Proc. of the ACM on Computer Graphics and Interactive Techniques*, 4(3), 2021. 2

[36] Pei Xu and Ruocheng Wang. Synchronize dual hands for physics-based dexterous guitar playing. In *SIGGRAPH Asia 2024 Conference Papers*, New York, NY, USA, 2024. Association for Computing Machinery. 2

[37] Pei Xu, Xiumin Shang, Victor Zordan, and Ioannis Karamouzas. Composite motion learning with task control. *ACM Transactions on Graphics*, 42(4), 2023. 2

[38] Zeshi Yang, Kangkang Yin, and Libin Liu. Learning to use chopsticks in diverse gripping styles. *ACM Transactions on Graphics (TOG)*, 41(4):1–17, 2022. 2

[39] Heyuan Yao, Zhenhua Song, Baoquan Chen, and Libin Liu. Controlvae: Model-based learning of generative controllers for physics-based characters. *ACM Trans. Graph.*, 41(6), 2022. 2

[40] Zhiqi Yin, Zeshi Yang, Michiel Van De Panne, and KangKang Yin. Discovering diverse athletic jumping strategies. *ACM Transactions on Graphics (TOG)*, 40(4):1–17, 2021. 1

[41] Kevin Zakka, Philipp Wu, Laura Smith, Nimrod Gileadi, Taylor Howell, Xue Bin Peng, Sumeet Singh, Yuval Tassa, Pete Florence, Andy Zeng, and Pieter Abbeel. Robopianist: Dexterous piano playing with deep reinforcement learning. In *Conference on Robot Learning (CoRL)*, 2023. 2

[42] 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 Transactions on Graphics (TOG)*, 42(4):1–14, 2023. 2

[43] 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