InterPrior: Scaling Generative Control for Physics-Based Human-Object Interactions

Anonymous CVPR submission

Paper ID *****

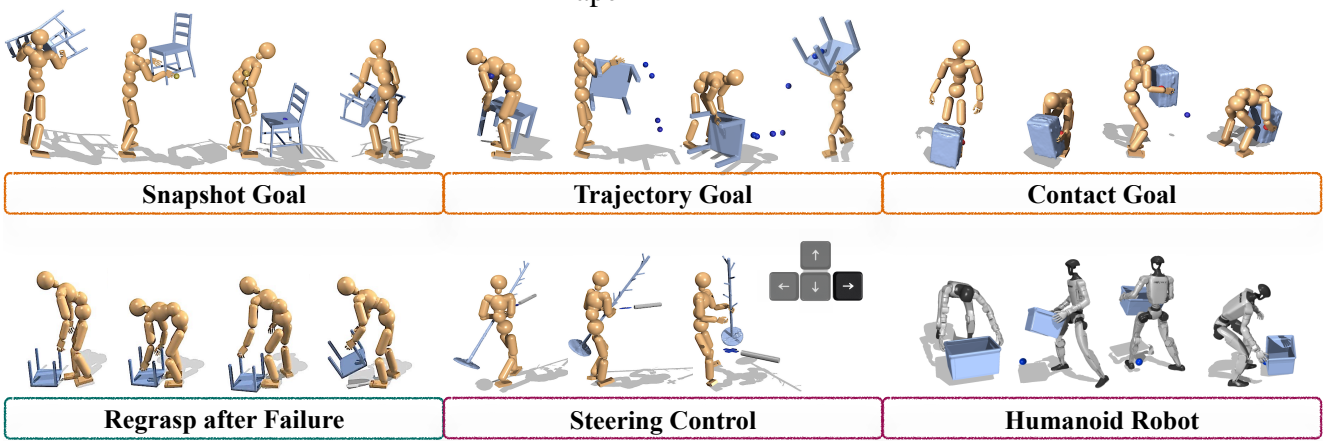

Figure 1. InterPrior is a versatile generative controller that drives a simulated humanoid to follow sparse goals and interact with objects in physics. It supports (**I**) long-horizon snapshot goals, (**II**) trajectory goals, and (**III**) contact goals (*Top*). *Yellow*, *blue*, and *red* dots respectively denote *human*, *object*, and *contact* goals. It demonstrates failure recovery (*Bottom Left*) and steering control on humanoid robot embodiments (*Bottom Right*).

Abstract

Humans rarely plan whole-body interactions with objects at the level of explicit whole-body movements. High-level intentions, such as affordance, define the goal, while coordinated balance, contact, and manipulation can emerge naturally from underlying physical and motor priors. Scaling such priors is key to enabling humanoids to compose and generalize loco-manipulation skills across diverse contexts while maintaining physically coherent whole-body coordination. To this end, we introduce InterPrior, a scalable framework that learns a unified generative controller through large-scale imitation pretraining and post-training by reinforcement learning. InterPrior first distills a full-reference imitation expert into a versatile, goal-conditioned variational policy that reconstructs motion from multimodal observations and high-level intent. While the distilled policy reconstructs training behaviors, it does not generalize reliably due to the vast configuration space of large-scale human-object interactions. To address this, we apply data augmentation with physical perturbations, and then perform reinforcement learning finetuning to improve competence on unseen goals and initializations. Together, these steps consolidate the reconstructed latent skills into a valid manifold, yielding a motion prior that generalizes beyond the training data, e.g., it can incorporate new behaviors such as interactions with unseen objects. We further demonstrate its effectiveness for user-interactive control and its potential for real robot deployment.

1. Introduction

Human-object interaction (HOI) is hierarchical: humans plan with sparse intent while limb coordination, balance, and contact emerge through fast motor responses [16]. Imitation policies [21] scale to large HOI skill repertoires but rely on explicit planners providing dense full-body and object references. A more useful *interaction motor prior* should sample feasible loco-manipulation behaviors from a distribution conditioned on sparse goals, rather than mimicking deterministic, fully-specified trajectories.

Existing approaches struggle to scale. Generative controllers trained with adversarial matching [3, 10] suffer from unstable optimization and handcrafted task rewards. Reference-imitation distillation [8, 14] absorbs large-scale data but is brittle when reference coverage lags the configuration space, as in loco-manipulation, where a few object DoFs combinatorially explode contact modes.

We introduce InterPrior, a generative HOI controller

(Fig. 1) scalable along four axes: *task coverage* (one policy supports sparse targets and their compositions), *skill coverage* (a single recipe scales to large HOI data), *motion coverage* (it generates expressive trajectories beyond demonstrations), and *dynamics coverage* (it sustains success under varied physics). Our key insight is that *RL finetuning is essential* for turning distillation from data reconstruction into a robust, generalizable policy. Distillation alone cannot cover the full HOI configuration space, while RL alone drifts toward unnatural reward-hacking. We use distillation as a strong, natural initialization, and RL as a local optimizer that improves robustness while remaining anchored to the pretrained model.

2. Related Work

Physics-based HOI control. Reference imitation policies [19, 21] produce high-fidelity HOI motion when dense references are available, but rely on explicit kinematic planners that scale poorly with object DoFs. Adversarial generative controllers [3, 10] broaden distributions but suffer from unstable optimization and discriminator mode collapse. Goal-conditioned distillation [15] absorbs large-scale data without task-specific design, yet remains brittle when reference coverage lags the contact-rich configuration space. InterPrior combines imitation distillation with RL post-training to consolidate sparse-goal coverage and recovery in a single policy.

Humanoid loco-manipulation. Whole-body humanoid policies have shown locomotion and pick-and-place [2, 5], often via teleoperation or staged imitation. Recent works combine retargeting with universal trackers for general loco-manipulation [4]; InterPrior acts as the underlying interaction prior such trackers can build on.

3. Method

Task formulation. We learn a policy π that operates in a physics simulator and produces HOI motion from *high-level goals* rather than full reference. Goals can be supplied by a human user, a kinematic motion generator, or sparse keypoints from MoCap. The policy conditions on the current human-object state and recent history together with these goals, and samples control signals from its learned distribution. The output is a physically simulated rollout that follows the goals where specified and remains diverse elsewhere.

Goal conditioning. At each timestep the policy receives short-horizon previews and a long-horizon snapshot, each represented as a *masked residual encoding*: a binary mask selects which goal components (joints, object pose, contacts) are active, and active entries are encoded as residuals to the current state. This unifies snapshot, trajectory, contact, and composed goals under a single masked formulation [14]. At inference, user-specified or model-generated sparse targets are supplied by filling only the informed components and zeroing the rest.

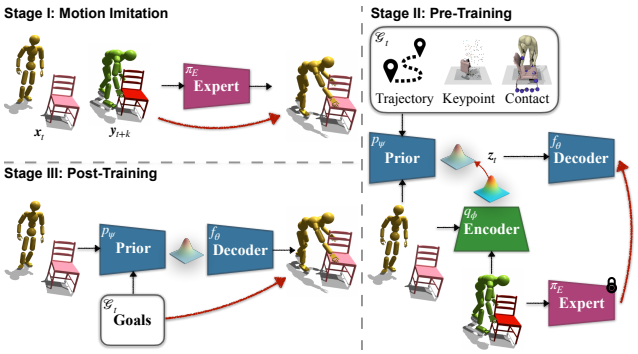

Figure 2. Overview of InterPrior. (**I**) Train an imitation expert on large-scale HOI data with augmentation, perturbations, and a contact-aware hand reward; (**II**) distill the expert into a variational policy with structured latent skills via online DAgger and ELBO loss; (**III**) RL post-training on randomized single-frame inbetweening, anchored by concurrent distillation. Blue: inference-time modules. Green/red: training-only.

Stage I: imitation expert. We train an expert π_E on large-scale HOI data with PPO [13]. Beyond reference tracking, we expand reference scope via random initialization, sparse impulse perturbations, and dynamics randomization, and we introduce a *reference-free hand reward* that encourages the hand to target and wrap the actual simulated object rather than rigidly tracking the reference. This corrects failures on thin geometries and under perturbation, where strict reference tracking becomes unreliable.

Stage II: variational distillation. We distill π_E into a masked conditional variational policy π with a Transformer prior, an MLP encoder (training-only), and an MLP decoder. Latents are projected onto a hypersphere ($\boldsymbol{z}_t \leftarrow \boldsymbol{z}_t/\|\boldsymbol{z}_t\|$) following [11] to bound out-of-distribution draws while preserving directional diversity. Online DAgger [12] minimizes a weighted ELBO [6] (imitation, masked-goal reconstruction, KL), with two auxiliary losses constraining the prior mean to unit magnitude and encouraging temporal consistency across consecutive priors. The decoder also reconstructs masked goal entries, yielding a latent that learns to *complete* intent.

Stage III: RL post-training. The distilled policy follows goals but is brittle when state or goals drift off the demonstration manifold. We RL-finetune on *single-frame in-betweening*: from randomized initial states, the policy must reach a single sparse goal sampled from the dataset, with a sparse success reward $r_{\text{goal}} = r_{\text{succ}} \cdot \mathbb{1}[\|\boldsymbol{m}_{t+L} \odot \Delta(\tilde{\boldsymbol{y}}_{t+L}, \boldsymbol{x}_t)\|_1 < \tau]$. Composed and randomized goals systematically broaden the encountered state distribution, training the policy to recover from near-failure states (e.g., regrasping after a slipped grasp) without explicit supervision. To preserve the pretrained prior, a subset of environ-

Table 1. **Goal-conditioned tasks** (% success). Snapshot/Trajectory/Contact in-distribution tasks plus Chain and Random-Init stress tests; each row adds one component.

Variant	Snap	Traj	Cont	Chain	RInit	Fail ↓
MaskedMimic [14]	64.2	88.0	52.2	29.1	31.7	12.6
+ InterMimic+ Expert	71.4	92.7	69.3	33.9	30.1	11.0
+ Latent Shaping	74.9	92.4	71.9	40.0	30.9	10.6
+ Bounded Latent & Obs.	89.1	93.6	88.5	45.1	41.1	6.0
+ RL Finetuning (**full**)	**90.0**	**94.6**	**90.7**	**68.8**	**88.6**	**3.7**

Table 2. **Full-reference tracking and adaptation.** OMOMO uses thin objects with initialization perturbations; BEHAVE/HODome report zero-shot transfer to unseen objects.

Method	OMOMO SR ↑	BEHAVE SR ↑	HODome SR ↑
InterMimic [21]	63.9	10.7	27.8
InterMimic + finetuning	–	38.9	55.5
InterPrior (Ours)	**83.2**	27.4	40.1
InterPrior + finetuning	–	**52.0**	**72.4**

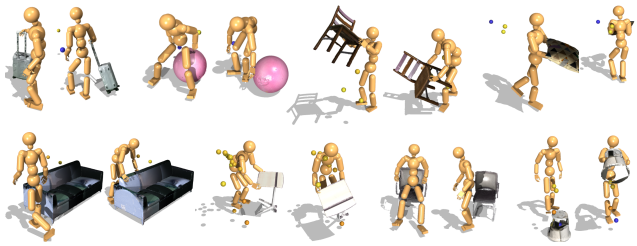

Figure 3. **Zero-shot generalization.** A single InterPrior model trained on OMOMO [7] adapts to unseen objects and interactions from BEHAVE [1] and HODome [22].

ments continues optimizing the distillation objective in parallel, anchoring the policy to its natural skill manifold without freezing weights. Skills outside demonstrations (e.g., *getting up*) are introduced via a learnable subtask token and a shaped upright-posture reward.

4. Experiments

Setup. We train on the InterAct [20] preprocessing of OMOMO [7] and evaluate generalization on selected sequences from BEHAVE [1] and HODome [22]. Policies run at 30 Hz in IsaacGym [9]. We compare InterPrior against the original InterMimic [21] (full-reference imitator) and an adapted MaskedMimic [14, 15] (sparse-goal distillation). Tasks include (i) full-reference tracking on thin objects with initialization noise; (ii) sparse goal following under snapshot, trajectory, contact, and composed goals; (iii) stress tests with long-horizon multi-goal chains and random-init lifting; and (iv) zero-shot adaptation to unseen objects/interactions.

Goal-conditioned results. Table 1 shows cumulative gains as we add each component. The largest jumps appear on stress tests: bounded latent and observation spaces roughly double random-init success, and RL finetuning further raises it from 41.1% to 88.6% while halving failure rate to 3.7%. Distillation-only policies fit the demonstration-induced state distribution; long rollouts and goal switching enter under-covered states and drift, while RL on single-frame in-betweening directly trains for sparse-target reaching from diverse initializations, improving recovery from off-distribution states. Trajectory-following is preserved because finetuning operates on snapshot goals while concurrent distillation protects trajectory-conditioned modes.

Tracking and adaptation. Table 2 shows that InterPrior achieves **83.2%** success on OMOMO with thin objects and initialization noise, versus 63.9% for InterMimic, while remaining competitive on per-joint position errors. The gain stems from the contact-aware hand reward and bounded latent: the policy makes small targeted deviations to realign contact instead of rigidly tracking the reference. On BEHAVE and HODome (held-out objects, different human shapes), InterPrior and its finetuned variant outperform InterMimic by large margins, showing the distilled prior absorbs additional interaction data more effectively even when that data is imperfect.

Qualitative behavior. Fig. 1 shows minute-long whole-body interaction with smooth transitions across approach, grasp, lift, and reposition; when contact or balance drift, InterPrior self-corrects rather than compounding errors. Fig. 3 shows zero-shot generalization: sparse snapshot goals are sufficient for InterPrior to complete unspecified DoFs and converge to feasible contacts on unseen objects.

Embodiment and sim-to-real. We retrain InterPrior on the Unitree G1 [18] with G1-specific stabilization rewards and dynamics randomization, and observe sustained loco-manipulation under sim-to-sim transfer to MuJoCo [17]. Building on this prior, a downstream controller [4] deploys autonomous loco-manipulation on a real G1 from egocentric depth and sparse task goals, indicating that scaling interaction priors in simulation is a viable path to versatile real-world humanoid behavior.

Operator-controlled steering. Beyond dataset-derived goals, InterPrior accepts on-the-fly sparse targets from a human operator via a keyboard interface that streams a small number of object-pose or end-effector commands per second. Because the policy is conditioned on the same masked goal interface used at training, no additional finetuning is needed to support this mode. In our deployments, an operator chains push, lift, and reposition commands across multiple objects, and the policy completes the unspecified DoFs while maintaining whole-body balance. Mid-trajectory command switches are absorbed gracefully, and brief contact loss triggers re-approach behaviors learned through randomized in-betweening rather than scripted primitives.

Implementation details. The imitation expert, encoder, and decoder are MLPs with hidden sizes (1024, 1024, 512); the prior is a 4-layer Transformer encoder; critics share the expert MLP architecture. Latent dimension is $d_z = 64$.

All policies are trained with PPO [13]; distillation uses online DAgger. For the G1 humanoid retraining, we additionally randomize friction, mass density, center-of-mass offsets, and inertia, and apply per-joint impulse perturbations during rollout. Episode-fixed sampling noise $\epsilon \sim \mathcal{N}(0, I)$ promotes temporal consistency across a rollout.

Failure modes. Remaining failures involve extremely thin or elongated unseen geometries, and partial completion in multi-goal chaining when canonicalization across subgoals introduces large alignment discrepancies, where the policy favors balance over precise goal matching. Integrating tactile sensing or depth-conditioned goal completion is a promising direction.

5. Conclusion

InterPrior is a generative HOI controller that combines large-scale imitation distillation with RL post-training under perturbations and randomized goals. Distillation provides a strong, natural initialization; RL anchors the policy to its pretrained skill manifold while expanding state coverage and recovery behavior. The result is a reusable interaction prior that handles snapshot, trajectory, contact, and composed goals; recovers from failures; generalizes zero-shot to unseen objects; and transfers across embodiments toward real-humanoid deployment.

References

[1] Bharat Lal Bhatnagar, Xianghui Xie, Ilya Petrov, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *CVPR*, 2022. 3

[2] Yuhui Fu, Feiyang Xie, Chaoyi Xu, Jing Xiong, Haoqi Yuan, and Zongqing Lu. DemoHLM: From one demonstration to generalizable humanoid loco-manipulation. *arXiv preprint arXiv:2510.11258*, 2025. 2

[3] Mohamed Hassan, Yunrong Guo, Tingwu Wang, Michael Black, Sanja Fidler, and Xue Bin Peng. Synthesizing physical character-scene interactions. In *SIGGRAPH*, 2023. 1, 2

[4] Xialin He, Sirui Xu, Xinyao Li, Runpei Dong, Liuyu Bian, Yu-Xiong Wang, and Liang-Yan Gui. ULTRA: Unified multimodal control for autonomous humanoid whole-body loco-manipulation. *arXiv preprint arXiv:2603.03279*, 2026. 2, 3

[5] Dvij Kalaria, Sudarshan S Harithas, Pushkal Katara, Sangkyung Kwak, Sarthak Bhagat, Shankar Sastry, Srinath Sridhar, Sai Vemprala, Ashish Kapoor, and Jonathan Chung-Kuan Huang. DreamControl: Human-inspired whole-body humanoid control for scene interaction via guided diffusion. *arXiv preprint arXiv:2509.14353*, 2025. 2

[6] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[7] Jiaman Li, Jiajun Wu, and C Karen Liu. Object motion guided human motion synthesis. *ACM Transactions on Graphics (TOG)*, 42(6):1–11, 2023. 3

[8] Zhengyi Luo, Jinkun Cao, Sammy Christen, Alexander Winkler, Kris Kitani, and Weipeng Xu. Grasping diverse objects with simulated humanoids. In *NeurIPS*, 2024. 1

[9] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. In *NeurIPS*, 2021. 3

[10] Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, and Jingbo Wang. TokenHSI: Unified synthesis of physical human-scene interactions through task tokenization. In *CVPR*, 2025. 1, 2

[11] 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

[12] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011. 2

[13] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 2, 4

[14] Chen Tessler, Yunrong Guo, Ofir Nabati, Gal Chechik, and Xue Bin Peng. Maskedmimic: Unified physics-based character control through masked motion inpainting. *ACM Transactions on Graphics (TOG)*, 43(6):1–21, 2024. 1, 2, 3

[15] Chen Tessler, Yifeng Jiang, Erwin Coumans, Zhengyi Luo, Gal Chechik, and Xue Bin Peng. MaskedManipulator: Versatile whole-body control for loco-manipulation. *arXiv preprint arXiv:2505.19086*, 2025. 2, 3

[16] Emanuel Todorov and Michael I Jordan. Optimal feedback control as a theory of motor coordination. *Nature neuroscience*, 5(11):1226–1235, 2002. 1

[17] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *IROS*, 2012. 3

[18] Unitree. Unitree g1 humanoid agent ai avatar. https://www.unitree.com/g1/. 3

[19] 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. 2

[20] Sirui Xu, Dongting Li, Yucheng Zhang, Xiyan Xu, Qi Long, Ziyin Wang, Yunzhi Lu, Shuchang Dong, Hezi Jiang, Akshat Gupta, Yu-Xiong Wang, and Liang-Yan Gui. InterAct: Advancing large-scale versatile 3d human-object interaction generation. In *CVPR*, 2025. 3

[21] Sirui Xu, Hung Yu Ling, Yu-Xiong Wang, and Liang-Yan Gui. InterMimic: Towards universal whole-body control for physics-based human-object interactions. In *CVPR*, 2025. 1, 2, 3

[22] Juze Zhang, Haimin Luo, Hongdi Yang, Xinru Xu, Qianyang Wu, Ye Shi, Jingyi Yu, Lan Xu, and Jingya Wang. NeuralDome: A neural modeling pipeline on multi-view human-object interactions. In *CVPR*, 2023. 3