CoorDex: Coordinating Body and Hand Priors for Continuous Dexterous Humanoid Loco-Manipulation

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

***Abstract*—Humanoid loco-manipulation is often simplified as walking to an object, stopping to manipulate it, and then resuming locomotion. It also commonly relies on low degree-of-freedom (DoF) end effectors that behave like an open-close grasp primitive. We introduce CoorDex, a learning pipeline that converts high-dimensional body and dexterous hand control into coordinated latent residual control, enabling high-DoF dexterous loco-manipulation. Starting from simulated whole-body and hand demonstrations, CoorDex trains privileged motion tracking teachers for the humanoid body and dexterous hand, distills them into proprioception conditioned latent priors, and uses the frozen priors as the action space for downstream residual reinforcement learning. A *Coordinated Latent Residual Policy* composes these priors through shared task context and separate body-hand residual heads, preserving natural whole-body motion while improving finger-level contact reliability. CoorDex enables a Unitree G1 humanoid with a 20-DoF WUJI hand to perform several dexterous loco-manipulation skills, including non-stop bottle grasping and carrying, fridge door opening, and cube pick-and-turn. Controlled ablations on the walk-grasp-carry task show that joint-space PPO, joint-space hand control, and monolithic latent prediction all fail under the same reward budget, while the latent-prior interface and coordinated residual structure make high-dimensional contact-rich loco-manipulation trainable. Anonymous Project Page: https://coordex-anonymous.github.io/**

I. Introduction

Humanoid robots equipped with high DoF dexterous hands promise a single embodiment that can navigate human environments and manipulate everyday objects. Realizing this promise requires more than placing an end effector near a target. The robot must maintain balance, keep the object reachable, coordinate wrist and finger poses before contact, close the hand without disturbing the object, and continue transporting the object while hand-object interaction perturbs whole-body dynamics. Learning such behaviors directly over all actuated joints creates a difficult high-dimensional exploration problem.

Recent progress has advanced complementary parts of humanoid loco-manipulation. Whole-body tracking enables stable locomotion and expressive upper-body motion [24, 19, 22, 8, 4, 12, 10]. Physics-based priors reduce exploration through structured latent action spaces [25, 26, 31, 21, 30]. Teleoperation, residual learning, hierarchical control, force-adaptive policies, and vision conditioned systems further improve task-level interaction and sim-to-real transfer [7, 6, 11, 45, 5, 15, 43, 29, 9, 39, 13]. However, many systems still use end-effector-centric, low-dimensional, or primitive-based manipulation interfaces, leaving high-DoF multi-finger contact during continuous locomotion less isolated and less understood.

This gap becomes more pronounced when a dexterous hand is mounted on a moving humanoid. Hand-centric methods have made strong progress in grasp synthesis, hand-object representation, cross-embodiment transfer, and demonstration generation [34, 18, 42, 35, 36, 17, 14, 2]. Yet they often assume that the wrist or arm trajectory is provided by teleoperation, planning, a fixed-base setup, or a stationary whole-body controller. On a walking humanoid, the wrist pose instead emerges from foot timing, root motion, torso posture, and whole-body reaching. A hand prior that must also explain 6D wrist motion therefore spends much of its capacity on body-side placement rather than on finger coordination. The key challenge is to decompose and coordinate body-side wrist placement with hand-side finger motion during continuous locomotion.

We introduce **CoorDex**, a learning pipeline that makes high-DoF dexterous humanoid loco-manipulation trainable through coordinated latent residual control. CoorDex constructs separate latent priors for the humanoid body and dexterous hand. Each prior is trained from a privileged motion tracking teacher and distilled into a proprioception conditioned student prior and decoder using a PULSE-style variational bottleneck [21]. The body prior supports locomotion, reaching, and wrist placement, while the wrist-stabilized hand prior controls only the active finger joints and captures reusable finger coordination. This factorization replaces direct full-joint control with residual control over two compact latent spaces.

To compose these priors, CoorDex uses a *Coordinated Latent Residual Policy*. At each control step, the frozen body and hand priors first encode their respective proprioception into latent means. These means, together with object-relative geometry, contact features, and the current proprioceptive state, are passed to a shared coordination trunk to produce a task-level coordination feature. Two residual heads then predict body and hand latent corrections, which are added to the corresponding prior means before decoding. The body residual adjusts stepping, torso motion, reaching, and wrist placement, whereas the hand residual refines finger preshape, closure, and contact. This design couples the two subsystems through shared task state without collapsing them into a single monolithic action head. It preserves the structural separation between whole-body motion and finger-level dexterity.

Our contributions are threefold. **1)** We present a complete learning pipeline in Isaac Lab [23] for high-DoF dexterous humanoid loco-manipulation, including simulated whole-body

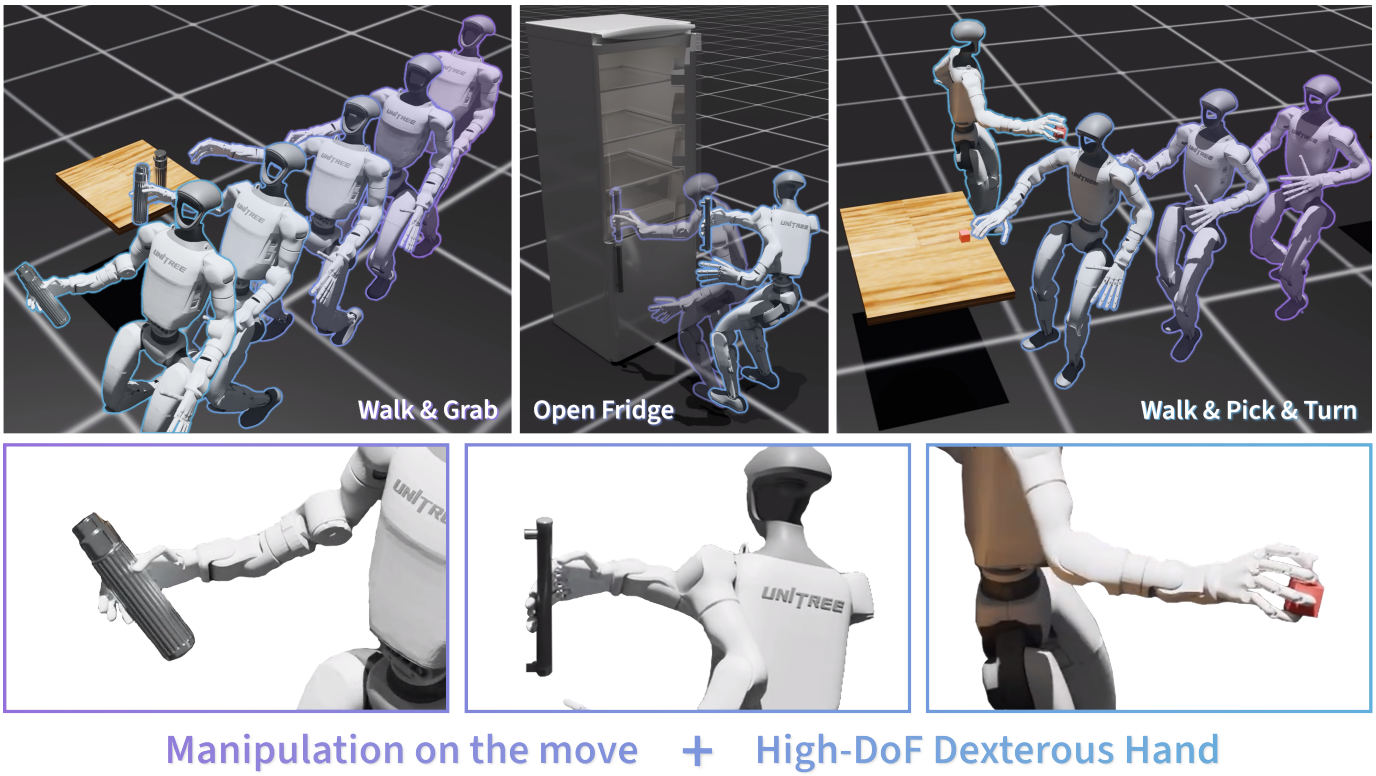

Fig. 1: **Dexterous loco-manipulation on the move.** CoorDex enables a humanoid equipped with high-DoF dexterous hands to perform continuous loco-manipulation tasks that require simultaneous coordination between locomotion and dexterous hand control, including walk-grasp-carry, fridge opening while stepping back, and walk-pick-turn.

and hand demonstration collection, motion tracking policies, prior distillation, and downstream residual RL. **2)** We propose an asymmetric body-hand prior composition for high-dimensional control. A task-aligned body prior handles locomotion and wrist placement, while a wrist-stabilized hand prior captures reusable finger coordination. A *Coordinated Latent Residual Policy* adapts the frozen priors through shared task context and separate body-hand residual heads. **3)** We provide controlled ablations on the walk-grasp-carry task showing that raw joint-space PPO, the body prior with joint-space hand control, and monolithic latent prediction all fail under the same reward budget, highlighting the importance of both latent-prior actions and structured body-hand coordination for dexterous loco-manipulation.

II. Related Work

Motion Imitation and Motion Priors. Physics-based motion imitation enables high-dimensional humanoid control. DeepMimic [24] combines reference motion tracking with task rewards to produce robust simulated skills. Recent humanoid trackers scale this idea to full-body robots by improving motion retargeting, adaptive tracking, privileged training, and sim-to-real deployment, enabling expressive locomotion, whole-body teleoperation, and highly dynamic behaviors [22, 8, 4, 12, 19, 38, 10, 41, 40, 16, 3]. A related line learns latent motion priors that replace raw joint actions with more structured action spaces. AMP [25] and ASE [26] use adversarial objectives to regularize or parameterize downstream control, while CALM [31], PULSE [21], and SLMP [30] further study conditional or compact latent representations for reusable physics-based skills. Building on prior work that learns motion priors for humanoid body control, we extend the same principle to dexterous hand control by constructing a separate hand prior. CoorDex then studies how to compose the body and hand priors for downstream contact-rich loco-manipulation tasks.

Dexterous Hand Manipulation. Dexterous hand research provides tools for grasp synthesis, hand-object representation, and manipulation data generation. Large-scale datasets and generative methods synthesize diverse multi-finger grasps over objects and hand morphologies [34, 18, 42]. Cross-embodiment representations such as $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [35] and OHRA [36] encode hand-object geometry or hand morphology in forms that transfer across different dexterous hands. Demonstration generation and transfer systems such as ManipTrans [17], DexMimicGen [14], and Jacta [2] use human motions, planners, or few demonstrations to bootstrap dexterous manipulation through tracking, imitation, and residual refinement. Our hand prior follows the same broad principle that retargeted hand-object motion can provide useful supervision. The key difference is how the hand is embedded in a walking humanoid. Instead of letting the hand latent control

the wrist pose, we train a wrist-stabilized prior whose latent action controls only finger motion. Wrist placement is handled by the humanoid body prior, and the two priors are coupled downstream by the *Coordinated Latent Residual Policy*.

Humanoid Loco-Manipulation. Humanoid loco-manipulation couples locomotion, reaching, contact, and object transport. Existing approaches address this coupling from several directions. Optimization and residual-learning methods generate or refine dynamically feasible whole-body behaviors for object interaction [20, 45, 5]. Skill composition and unified whole-body controllers widen the range of reachable tasks, while force-adaptive policies make humanoids more robust to payloads and contacts [15, 29, 43]. Teleoperation systems provide high-quality humanoid demonstrations and have begun to include dexterous hands [7, 1, 11]. Vision-based systems push loco-manipulation toward large-scale sim-to-real transfer and language- or video-conditioned control [9, 39, 13]. We isolate the control problem of continuous high-DoF dexterous loco-manipulation, where grasp formation and object transport occur during ongoing walking rather than in a stationary manipulation phase. This setting is closer to finger-level humanoid dexterity than to standard mobile manipulation, where navigation and manipulation can be separated in time.

III. Method

We present **CoorDex**, a modular pipeline that converts high-dimensional humanoid body and dexterous hand control into coordinated latent residual control for dexterous loco-manipulation, as illustrated in Fig. 2. CoorDex first constructs separate latent priors for the humanoid body and the dexterous hand, and then trains a downstream residual RL policy to coordinate the two frozen priors. Sec. III-A describes prior construction through teacher tracking and proprioceptive distillation. Sec. III-B introduces the *Coordinated Latent Residual Policy*. Sec. III-C presents residual PPO [28] training and the environment design.

A. Prior Construction

We construct two separate motion priors, one for the humanoid body and one for the dexterous hand. The two priors are trained on different tracking tasks but share the same distillation pipeline. Let $x \in \{b, h\}$ denote the body or hand subsystem.

Demonstration collection in simulation We collect task reference motions in Isaac Lab through a simulated teleoperation pipeline. Lower-body locomotion is generated by an AGILE-based locomotion controller [44], while the operator provides right wrist and hand motion through Apple Vision Pro using the CloudXR-based XR teleoperation interface in Isaac Lab [23]. The tracked human wrist pose serves as the end-effector target for a Pink inverse-kinematics solver, and human hand motion is retargeted to the target dexterous hand with optimization-based dex-retargeting [27]. The resulting whole-body trajectories provide the reference motions for body tracking training.

Humanoid body prior. For the body subsystem, we first train a general whole-body motion tracking teacher π_T^b [19]. The teacher takes body proprioception $\mathbf{s}_t^{b,p}$ and a reference goal $\mathbf{s}_t^{b,g}$ as input, and outputs joint position targets $\mathbf{a}_t^{b,T}$ for the body joints. Before training, we preprocess each reference motion on the humanoid model equipped with the corresponding dexterous hands. The resulting reference trajectories thus match the morphology and kinematic structure used in our downstream loco-manipulation tasks.

Dexterous hand prior. For the hand subsystem, we train a privileged hand tracking teacher π_T^h in a floating-hand environment using ManipTrans-style [17] retargeted hand-object motions. The teacher observes hand proprioception $\mathbf{s}_t^{h,p}$ and a reference goal $\mathbf{s}_t^{h,g}$, and outputs joint position targets $\mathbf{a}_t^{h,T}$ for the active finger joints. During hand-prior training, the reference wrist pose and velocity are written directly into simulation, so the teacher and the learned prior control only finger motion. This wrist-stabilized design keeps the hand latent space from spending most of its capacity on 6D wrist motion, and makes the learned latent command directly useful for finger coordination.

Distillation. After training the tracking teachers, we distill each teacher into a proprioception conditioned latent prior [21]. For each subsystem $x \in \{b, h\}$, the student contains an encoder $\mathcal{E}_x(\mathbf{z}_t^x \mid \mathbf{s}_t^{x,g}, \mathbf{s}_t^{x,p})$, a proprioceptive prior $\mathcal{R}_x(\mathbf{z}_t^x \mid \mathbf{s}_t^{x,p})$, and a decoder $D_x(\mathbf{s}_t^{x,p}, \mathbf{z}_t^x)$. The training objective combines teacher-action reconstruction, temporal smoothness of the encoder means, and KL regularization between the encoder and the prior:

$$\mathcal{L}_{\text{distill}}^x = \mathcal{L}_{\text{action}}^x + \alpha_x \mathcal{L}_{\text{regu}}^x + \beta_x \mathcal{L}_{\text{KL}}^x. \tag{1}$$

After distillation, $\mathcal{R}_x$ and D_x are frozen, and the prior mean $\boldsymbol{\mu}_t^{x,p}$ provides the default latent command for downstream residual RL. Full definitions are given in Appendix A.

B. Coordinated Latent Residual Policy

Given the frozen body and hand priors, a downstream residual PPO [28] policy coordinates and adapts them for the loco-manipulation task. At each control step, we obtain the proprioception conditioned prior means from the frozen prior networks:

$$\begin{aligned} \boldsymbol{\mu}_t^{b,p} &= \text{Mean}\left[\mathcal{R}_b\left(\mathbf{z}_t^b \mid \mathbf{s}_t^{b,p}\right)\right], \\ \boldsymbol{\mu}_t^{h,p} &= \text{Mean}\left[\mathcal{R}_h\left(\mathbf{z}_t^h \mid \mathbf{s}_t^{h,p}\right)\right]. \end{aligned} \tag{2}$$

The actor predicts residuals in the two latent spaces rather than joint-space targets:

$$\Delta \mathbf{z}_t = \left[\Delta \mathbf{z}_t^b, \Delta \mathbf{z}_t^h\right], \qquad \Delta \mathbf{z}_t^b \in \mathbb{R}^{d_b}, \quad \Delta \mathbf{z}_t^h \in \mathbb{R}^{d_h}. \tag{3}$$

The actor first computes a shared coordination representation $\mathbf{c}_t$ to capture body-hand coupling, and then predicts residual commands through a specialized body head f_b and hand head f_h:

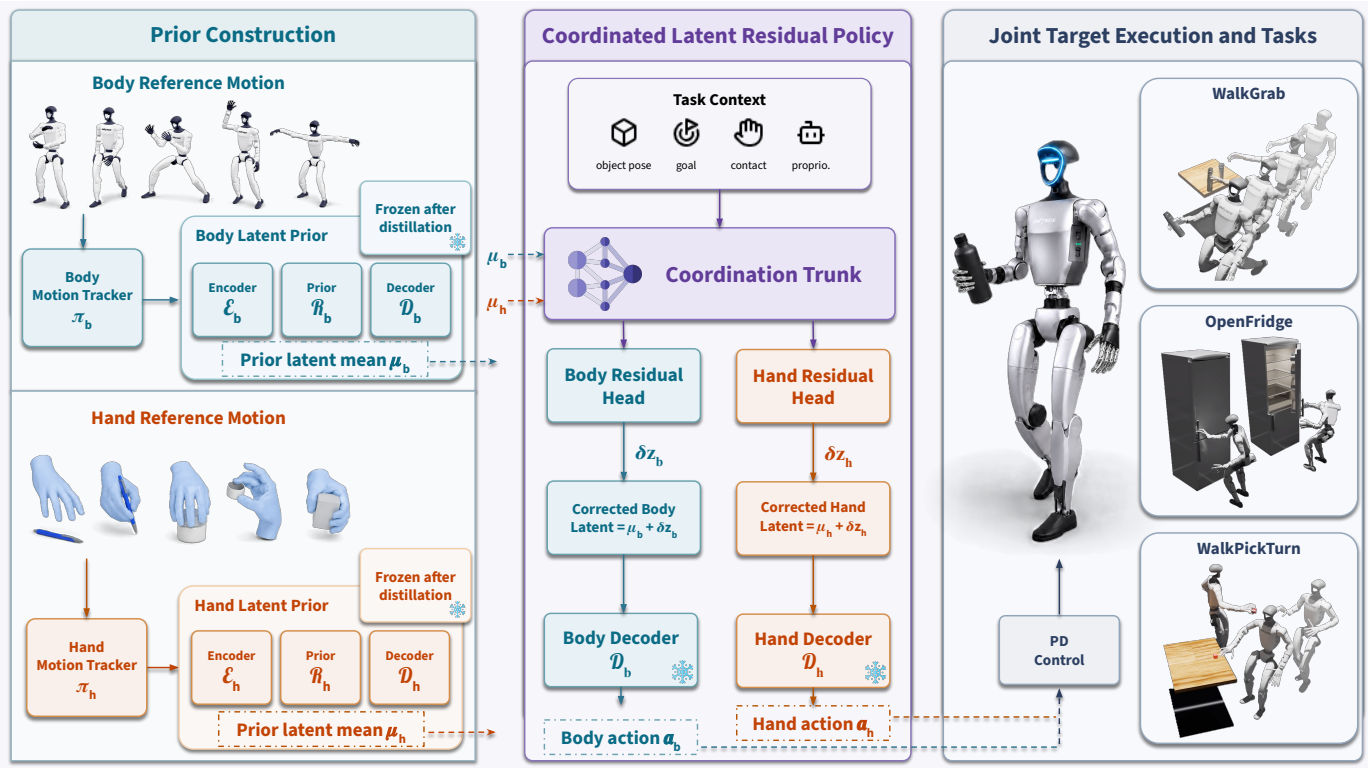

Fig. 2: **Overview of CoorDex**. Body and hand reference motions are tracked by privileged teachers and distilled into separate proprioception conditioned latent priors. During downstream RL, a coordinated residual policy uses task context and prior means to predict body and hand latent residuals. The frozen decoders map the corrected latents to joint-position targets for loco-manipulation.

$$
\begin{aligned}
\mathbf{c}_t &= f_{\text{coord}}\big(\mathbf{s}_t^{b,p}, \mathbf{s}_t^{h,p}, \mathbf{s}_t^{\text{task}}, \mathbf{s}_t^{\text{hand-object}}, \\
&\qquad \boldsymbol{\mu}_t^{b,p}, \boldsymbol{\mu}_t^{h,p}, \Delta\mathbf{z}_{t-1}\big), \\
\Delta\mathbf{z}_t^b &= \tanh\big(f_b(\mathbf{c}_t, \mathbf{s}_t^{b,p}, \boldsymbol{\mu}_t^{b,p})\big), \\
\Delta\mathbf{z}_t^h &= \tanh\big(f_h(\mathbf{c}_t, \mathbf{s}_t^{h,p}, \boldsymbol{\mu}_t^{h,p}, \mathbf{s}_t^{\text{hand-object}})\big).
\end{aligned} \tag{4}
$$

Here $\mathbf{s}_t^{\text{task}}$ contains task-level state such as object pose, goal information, projected gravity, and contact features. The hand-object state $\mathbf{s}_t^{\text{hand-object}}$ includes the object pose in the hand frame and fingertip-object contact features.

The corrected latent commands are

$$
\tilde{\mathbf{z}}_t^b = \boldsymbol{\mu}_t^{b,p} + \Delta\mathbf{z}_t^b, \qquad \tilde{\mathbf{z}}_t^h = \boldsymbol{\mu}_t^{h,p} + \Delta\mathbf{z}_t^h. \tag{5}
$$

The final joint position targets are produced by the frozen decoders:

$$
\mathbf{a}_t^b = D_b\left(\mathbf{s}_t^{b,p}, \tilde{\mathbf{z}}_t^b\right), \qquad \mathbf{a}_t^h = D_h\left(\mathbf{s}_t^{h,p}, \tilde{\mathbf{z}}_t^h\right). \tag{6}
$$

The body decoder outputs targets for all humanoid body joints. The hand decoder outputs targets for the active finger joints of the selected dexterous hand. These targets are inserted into the corresponding joint slots and executed by a low-level PD controller. This architecture keeps downstream exploration low-dimensional while preserving separate control authority for locomotion and grasping. The shared coordination trunk lets the policy reason about task phase and contact state. The separate heads let the body residual adapt stepping, torso motion, reaching, and wrist placement, and let the hand residual adapt finger preshape, closure, and contact refinement.

C. Residual RL and Environment Design

We train downstream policies with PPO while keeping the body and hand priors frozen. The policy observation contains body proprioception, hand proprioception, task state, object-relative geometry, fingertip contact features, the prior means, and the previous latent residual. Since the actor outputs only $\Delta\mathbf{z}_t$, exploration happens in the learned body and hand latent spaces rather than in the full joint-position action space.

All tasks use the same latent residual interface but differ in their interaction structure and exploration difficulty. The priors used by different tasks are not identical, since each task needs different coarse motion support. For shorter interaction tasks such as walk-grasp-carry and fridge door opening, we use a single-stage reward without reference state initialization. The reward combines locomotion and balance terms, palm-object reaching and alignment terms, fingertip contact and sustained contact terms, task-completion terms such as lift, carry, or door-angle progress, and regularization on latent residuals, decoded action changes, and joint velocities. For the longer-horizon walk-pick-turn task, pure reward optimization from the initial state is unreliable. We therefore use a lightweight staged reward together with demonstration-free reference state initialization (NoDemoRSI), which resets part of the episodes from states the policy itself has reached. This curriculum is

TABLE I: Actuated degrees of freedom and latent dimensions.

Subsystem	DoF	Latent dim.
G1 body	29	16
WUJI hand	20	12

task-specific and is not part of the policy architecture. More details are given in Appendix B.

IV. Experimental Results

We evaluate the core design components of **CoorDex** in Isaac Lab [23]. Our experiments use a 29-DoF Unitree G1 humanoid [33] equipped with a 20-DoF five-finger WUJI dexterous hand [37]. We study three questions. **Q1.** Whether a unified body–hand latent residual interface supports diverse loco-manipulation skills. **Q2.** Whether latent actions improve exploration over direct joint-space control. **Q3.** Whether coordinated residual prediction improves over monolithic latent prediction.

A. Experiment Setup

Tasks. We consider three dexterous humanoid loco-manipulation tasks with different interaction structures. In WalkGrab, the humanoid must grasp and lift a bottle from a side table while continuously walking forward. In OpenFridge, it must grasp a fridge handle and open the door while stepping backward to create workspace. In WalkPickTurn, it must approach a table, pick up a cube, and complete a 180° turn while retaining the object. The three tasks respectively stress dynamic grasping during locomotion, sustained articulated-object interaction, and long-horizon skill composition. Fig. 1 visualizes the three tasks. Table I summarizes the action dimensions.

Evaluation metrics. Unless stated otherwise, all metrics are computed over 50,000 evaluation episodes collected with 10,000 parallel simulation environments. We report task success, fall rate, drop rate when applicable, and task-specific progress metrics. A rollout counts as successful only if the task-specific completion condition is satisfied and the robot stays balanced. WalkGrab requires lifting and carrying the bottle beyond a target displacement, OpenFridge requires the maximum door angle to exceed the success threshold, and WalkPickTurn requires grasping, lifting, and retaining the cube while completing a 180° turn. Exact thresholds are provided in Table XI. We additionally report diagnostic metrics, including door angle, heading error, and action rate.

Controlled variants on WalkGrab. We use WalkGrab as the main controlled ablation task because it is the cleanest test of grasping during locomotion. It uses a single-stage reward, does not use reference state initialization, and directly couples root motion, wrist placement, and finger closure. We compare CoorDex with three variants under the same environment and task-level reward. *All Joint Space* removes both latent priors and predicts joint targets directly. *Body Prior + Hand Joint Space* keeps the body prior but controls the hand in joint space. *Monolithic Latent Residual* keeps both priors but predicts one concatenated residual vector with a single MLP. Since the action parameterizations differ, we compare task-level metrics rather than raw return.

TABLE II: Task performance of CoorDex. Results are evaluated over 50,000 episodes collected with 10,000 parallel simulation environments in Isaac Lab.

Task	Success	Fall	Drop	Task-specific metric
WalkGrab	0.55	0.00	0.40	Velocity profile, Fig. 4
OpenFridge	0.66	0.00	–	Door angle: 57.76 / 60°
WalkPickTurn	0.89	0.01	0.10	Min heading err.: 9.98°

B. Task Performance Across Skills

Table II shows that the same CoorDex interface can be instantiated across all three tasks, answering **Q1**. WalkGrab is the most challenging because grasping must happen under continuous whole-body motion, whereas OpenFridge permits the robot to slow down and reposition during manipulation. WalkPickTurn is longer-horizon, but NoDemoRSI provides additional exploration support by resetting later stages from states discovered by the policy itself, without any expert states or action labels. Implementation details are given in Appendix B.

C. Action-Space Variants on WalkGrab

We next study whether the latent-prior action interface is necessary for WalkGrab, answering **Q2**. This task directly tests the definition of non-stop dexterous loco-manipulation: the robot must grasp an object while it keeps moving forward. Table III compares the action-space variants under the same environment and task reward.

Metrics. Reach measures whether the palm enters a predefined distance threshold from the bottle. Grasp measures whether the bottle is lifted above the task threshold. Stop measures whether the humanoid slows below a velocity threshold near the bottle. Fall measures robot-fall termination.

The two failed variants reveal different failure modes, also visible in Fig. 3. All Joint Space removes both learned priors and must explore the full body and hand joint space. The policy often twists the whole body into unnatural postures to avoid falling, and it never grasps the bottle.

Body Prior + Hand Joint Space isolates the hand-side difficulty. With the body prior, the humanoid can walk toward the bottle and place the wrist near the interaction region. However, the policy still has to discover high-DoF finger coordination directly in joint space. In practice, it often learns to slow down or stop near the bottle and tries to solve the problem as a stationary grasping task. This behavior shows that the body prior alone is not enough for non-stop dexterous loco-manipulation. The hand prior is also needed to make finger coordination learnable under residual RL.

For WalkGrab, we also analyze non-stop behavior with a velocity profile. For each rollout, we compute the body-frame forward velocity $v_{x,t}^{\text{body}}$ and the relative forward position $d_t = x_t^{\text{robot}} - x_t^{\text{bottle}}$. We bin samples by d_t with bin width

TABLE III: Action-space variants on WALKGRAB under the same PPO budget. Reach, grasp, stop, and fall are rates over evaluation episodes.

Method	Success	Reach	Grasp	Stop	Fall
All Joint Space	0	1.00	0.00	0.86	0.04
Body Prior + Hand Joint Space	0	0.96	0.01	0.90	0.04
CoorDex	0.55	1.00	0.55	0.00	0.00

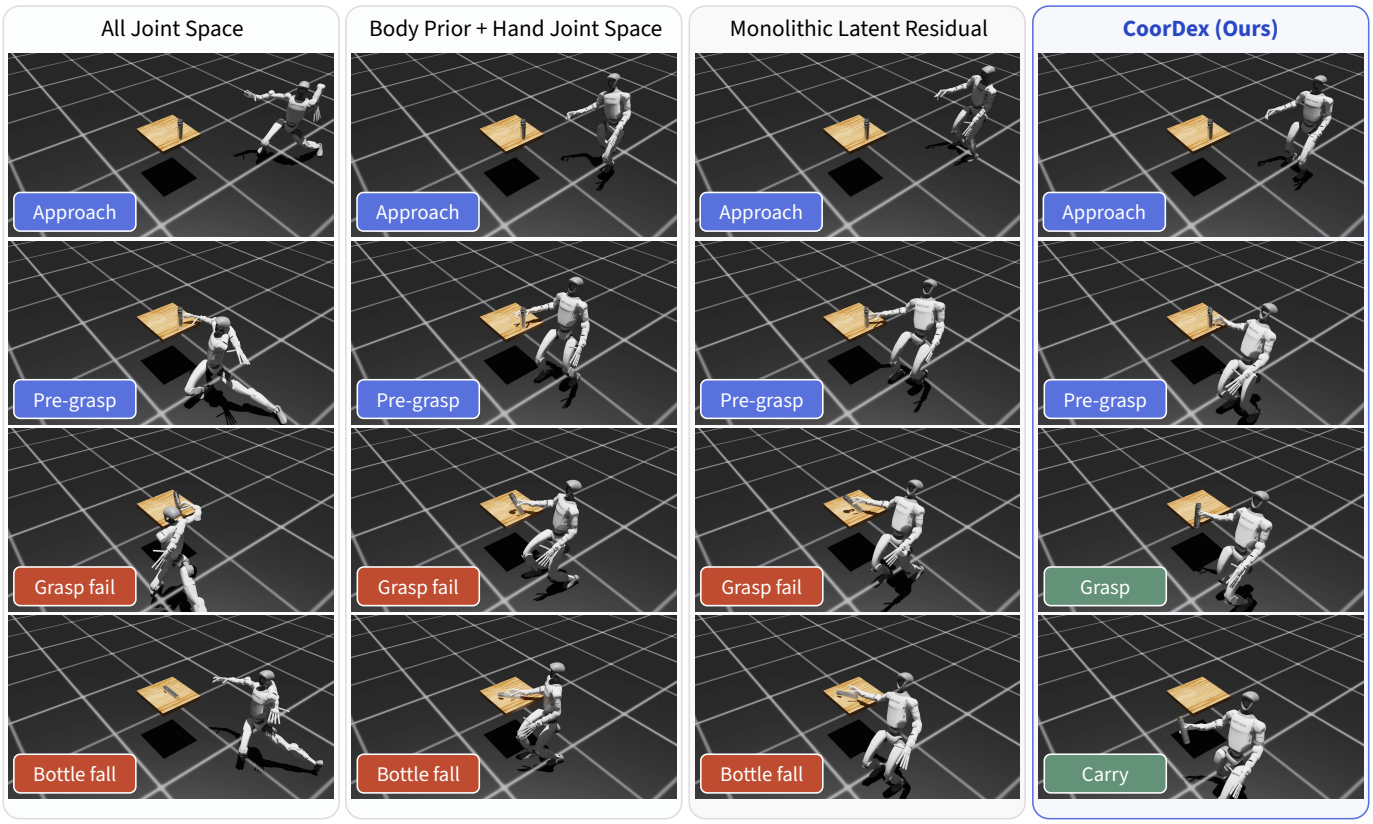

Fig. 3: **Qualitative comparison on WALKGRAB.** Each column shows sequential key frames from one rollout of the corresponding method. All Joint Space produces unstable whole-body motion. Body Prior + Hand Joint Space reaches the bottle but fails to learn a reliable grasp. Monolithic Latent Residual reaches the interaction region but produces less natural body motion and fails to complete the task. **CoorDex** completes the full sequence of approach, grasp, lift, and carry.

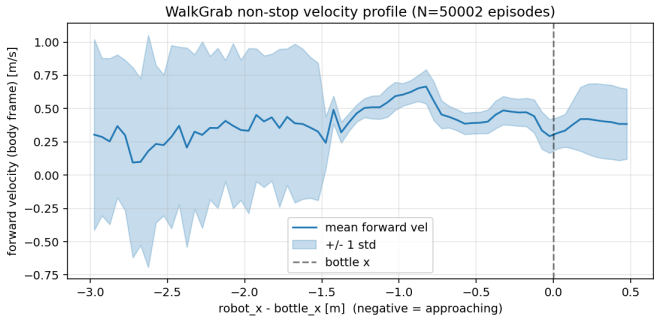

Fig. 4: Non-stop locomotion on WALKGRAB.

$0.05\,\mathrm{m}$ over the range $[-3.0, 0.5]\,\mathrm{m}$, and plot the mean and standard deviation of $v_{x,t}^{\mathrm{body}}$ in each bin. As shown in Fig. 4, CoorDex slows down slightly near the bottle but still keeps a forward velocity around $0.25\,\mathrm{m/s}$ near $d_t = 0$. This indicates that the policy does not solve the task by stopping before grasping.

D. Coordinated Residual Prediction

We finally test whether the actor structure matters once both priors are available, answering **Q3**. Monolithic Latent Residual and CoorDex use the same frozen body and hand priors, the same latent dimensions, and the same RL environment and reward. The only difference is the policy architecture. Monolithic Latent Residual uses one MLP to predict the full latent residual and then splits it into body and hand parts. CoorDex uses a shared coordination trunk with separate body and hand residual heads.

Metrics. Action rate measures the average change of decoded body joint targets between consecutive control steps. A lower action rate indicates smoother and more natural body motion.

Table IV and Fig. 3 show that the coordinated actor matters

TABLE IV: Coordination analysis on WALKGRAB.

Method	Success	Action rate	Fall
Monolithic Latent Residual	0.00	0.40	0.02
CoorDex	**0.55**	**0.22**	0.00

even when both priors are present. The monolithic actor can often follow the task direction and approach the bottle, but its body motion is less natural and more jittery. This is reflected in the higher action rate in Table IV and the unstable torso motion visible in Fig. 3. The hand failure is harder to see because the fingers are small, but the zero success rate shows that monolithic latent prediction does not produce reliable grasping under the same reward budget. CoorDex performs better because the shared trunk lets the policy reason about the same task state, while the separate heads keep body adaptation and finger adaptation from being forced through a single output pathway.

V. DISCUSSION AND LIMITATIONS

We present **CoorDex**, a modular pipeline for dexterous loco-manipulation. Rather than treating locomotion and manipulation as two separate phases, **CoorDex** exposes a structured latent action space for the humanoid body and the dexterous hand. Within this space, downstream RL can jointly adapt stepping, wrist placement, and finger-level contact. Across bottle grasping and carrying, fridge door opening, and cube pick-and-turn, the results indicate that dexterous humanoid skills benefit from priors that provide task-relevant motion support, and from policies that preserve the structure of body-hand coordination. This points to a practical direction for future loco-manipulation systems. Reusable motion priors should be paired with task-aware latent coordination, treating whole-body mobility and dexterous contact as coupled behaviors rather than stitched-together modules.

Limitations and Future Work. Although CoorDex demonstrates continuous high-DoF dexterous loco-manipulation across multiple simulated tasks, our results are still an early step toward general humanoid dexterity. The current policies use privileged state observations, including object poses and contact signals, and do not yet address perception or visual sim-to-real transfer. The experiments also focus on a fixed G1 platform with a WUJI hand, and broader deployment will require evaluation across more objects, hands, and hardware configurations. Finally, longer-horizon tasks such as WALKPICKTURN still rely on task-specific exploration support such as NoDemoRSI. Future work should combine the proposed latent residual interface with perception, automatic curricula, and broader task-conditioned priors.

APPENDIX A
MOTION PRIOR IMPLEMENTATION DETAILS

This section describes the implementation for building the body and hand priors that are later composed by the downstream residual policy. The body tracker extends the motion tracking design used by BeyondMimic [19]. References are represented as joint trajectories and Cartesian body targets, the policy tracks them through normalized joint-position setpoints, and the critic receives additional privileged tracking information. The hand tracker is trained with PPO to track MANO keypoints retargeted to robot hand bodies [17]. The distillation stage learns an encoder–decoder latent skill space together with a proprioception-conditioned prior [21].

A. Humanoid Body Tracking Teacher

The robot is a Unitree G1 [33] with fixed WUJI hands [37]. Only the 29 body joints are actuated for training efficiency. The reference bodies used by the tracking objective include the pelvis, leg links, torso, arm links, and wrist links; finger bodies are excluded from the body prior. Actions are normalized joint position setpoints with joint-specific scales derived from actuator effort and stiffness. The simulation uses decimation 4 and a physics step of $1/(60 \cdot 4)$ seconds, giving a 60 Hz control rate and 10 second episodes.

The no-state-estimator body teacher removes anchor translation and base linear velocity from the actor and teacher observations for real-world deployment. Per frame, the deployable proprioceptive suffix is

$$[\omega_{\text{base}}, q, \dot{q}, a_{t-1}] \in \mathbb{R}^{90},$$

where $q, \dot{q}, a_{t-1}$ each have 29 body-joint dimensions. The actor also receives the current motion command and relative anchor orientation during teacher training, with a five-frame history. The critic is asymmetric and additionally observes privileged reference body pose and velocity errors.

Rewards are exponential tracking rewards plus minimal regularization. The main positive terms track anchor position and orientation, relative body position and orientation, body linear and angular velocity, and palm keypoints. Penalties discourage action rate jitter, joint limit violations, and undesired contacts. Episodes terminate on timeout, excessive anchor height or orientation error, or excessive end-effector height error at the ankles and wrist yaw links. Domain randomization includes friction, joint default offsets, torso center-of-mass perturbation, and random base pushes.

The teacher is trained with PPO using 24 rollout steps per environment, five learning epochs, four minibatches, $\gamma = 0.99$, $\lambda = 0.95$, adaptive learning rate with target KL 0.01, learning rate 10^{-3}, entropy coefficient 0.005, and gradient clipping at 1.0. Actor and critic MLPs use ELU activations and hidden dimensions $[1024, 512, 256]$.

B. Dexterous Hand Tracking Teacher

Our hand motion loader reads ManipTrans-style annotations [17], evaluates SMPL-X to obtain MANO wrist and finger keypoints, converts from the MuJoCo frame to the Isaac Lab frame, and downsamples 120 Hz annotations to 60 Hz.

For the current WUJI hand prior, the tracked hand bodies are the palm and the five finger links and tips. The environment writes the reference wrist root pose and velocity into simulation both at command resampling and during every command update, so the wrist is kinematically driven by the dataset. The

policy therefore controls only the twenty active finger joints for WUJI hand. This fingers-only action space avoids spending latent capacity on high-variance wrist motion and matches the downstream residual interface, where the humanoid body prior determines the wrist motion.

The policy observation concatenates a target command, hand proprioception, and the previous action. The target command contains flattened MANO keypoint positions and velocities, plus the reference wrist pose expressed relative to the simulated wrist. Proprioception contains wrist linear and angular velocity in the wrist frame, and active joint position and velocity relative to default joint states. The critic receives the same target and proprioception plus privileged per-body keypoint position and velocity deltas.

The remaining rewards track MANO keypoints with fingertip-heavy weights, track keypoint velocities, penalize joint power through an exponential power reward, and penalize action rate. Tracking failure terminates an episode if keypoint errors exceed group-specific thresholds after a short warmup. An additional unstable state termination catches unrealistic wrist, body, or joint velocities. PPO hyperparameters match the body teacher: 24 rollout steps, five epochs, four minibatches, ELU MLPs with $[1024, 512, 256]$ hidden units, learning rate 10^{-3}, adaptive KL target 0.01, and 4096 environments.

C. VAE Distillation

The student contains three trainable networks:

$$q_\phi(z_t \mid s_t^{\text{full}}), \qquad p_\psi(z_t \mid s_t^{\text{prop}}), \qquad D_\theta(s_t^{\text{prop}}, z_t).$$

The encoder predicts the posterior diagonal Gaussian from the full student observation. The prior predicts a diagonal Gaussian from the deployable proprioceptive slice. And the decoder maps proprioception plus a latent sample to the teacher action. During training, z_t is sampled from the encoder with the reparameterization trick. At inference and downstream residual RL time, the default latent is the prior mean.

The distillation loss is

$$\begin{aligned}\mathcal{L} = \lambda_a \|\hat{a}_t - a_t^T\|_2^2 + \lambda_s \|\mu_t^q - \mu_{t-1}^q\|_2^2 \\ + \lambda_{\text{KL}}\, D_{\text{KL}}\big(q_\phi(z_t \mid s_t^{\text{full}}) \,\big\|\, p_\psi(z_t \mid s_t^{\text{prop}})\big)\,,\end{aligned} \tag{7}$$

where the smoothness term is evaluated only across valid consecutive samples in the same environment. This mirrors the PULSE [21] motivation. The decoder learns the teacher's motor actions. The prior makes the latent usable from proprioception alone. And the temporal regularizer keeps nearby states from mapping to discontinuous latent codes.

For the body prior, the latent dimension is 16. The proprioceptive prior input is the five-frame suffix

$$5 \times [\omega_{\text{base}}(3), q(29), \dot{q}(29), a_{t-1}(29)] = 450$$

dimensions. Encoder, decoder, and prior networks are MLPs with $[512, 256, 128]$. The body distillation optimizer uses learning rate $2 \cdot 10^{-4}$, action coefficient 1.0, smoothness coefficient 0.005, and a KL coefficient annealed from 10^{-3} to 10^{-4} between 15000 and 20000 updates.

For the active WUJI hand prior, the latent dimension is 12. The prior input is 66 dimensions:

$$[v_{\text{wrist}}(3), \omega_{\text{wrist}}(3), q_{\text{rel}}(20), \dot{q}_{\text{rel}}(20), a_{t-1}(20)].$$

The decoder outputs twenty fingers-only actions. The hand distillation optimizer uses learning rate $5 \cdot 10^{-4}$, action coefficient 1.0, smoothness coefficient 0.005, and a KL coefficient annealed from 10^{-2} to 10^{-3} between 15000 and 25000 updates. Both distillation runners use 24 rollout steps, five learning epochs, gradient accumulation length 16, gradient clipping at 1.0, empirical observation normalization, and up to 30000 iterations.

D. Downstream Residual Composition

Downstream loco-manipulation tasks use the frozen priors. The action term loads only the distilled prior and decoder weights, plus the observation normalizer. At each control step it constructs a body-prior observation from the current G1 proprioceptive history and a hand-prior observation from wrist-frame hand velocity, finger joint state, and the previous decoded hand action.

The downstream PPO actor outputs a residual latent

$$\Delta z_t = [\Delta z_t^b, \Delta z_t^h].$$

For each prior, the action term computes the prior mean and decodes the shifted latent:

$$a_t^b = D_b(o_t^b, \mu_t^b + \Delta z_t^b), \qquad a_t^h = D_h(o_t^h, \mu_t^h + \Delta z_t^h).$$

The decoded body actions are inserted into the 29 body-joint slots. The decoded hand actions are converted through the configured joint-action scale and offset into finger targets. An EMA with coefficient 0.4 is applied to the hand targets. The combined joint position action is then sent to the underlying Isaac Lab joint position action term.

The body prior is a 16-dimensional no-base-linear-velocity G1 prior and the hand prior is the 12-dimensional kinematic-wrist WUJI hand prior, yielding a 28-dimensional residual action space.

Appendix B
Training Details

We instantiate three downstream loco-manipulation tasks on the same Unitree G1 humanoid equipped with a WUJI five-finger right hand: **WalkGrab**, **OpenFridge**, and **WalkPickTurn**. All three share the same control stack. The policy runs at 60 Hz (physics step $1/240$ s, decimation 4) over 4096 parallel environments, and emits a 28-dim latent residual (16 body $+12$ hand) on top of two frozen priors mentioned above. WalkGrab and OpenFridge are single-stage reformulations whose phase structure is realized through predicate gating inside the reward terms, whereas WalkPickTurn keeps an explicit three stage schedule with a stage weighted reward.

A. Observation Details

For every task the actor and critic observe the same terms. The only difference is that observation noise is enabled for the actor and disabled for the critic. Each observation concatenates the proprioceptive prior stack (`humanoid_prior`, a 5-step history of the 90-dim body proprio frame $[\,v_{\text{ang}}^{\text{base}}(3),\, q(29),\, \dot{q}(29),\, a_{t-1}(29)\,]$; `right_hand_prior`, the 66-dim hand proprio frame; and the last emitted latent), the frozen prior latent means, contact forces, and task-specific object poses. Object poses are expressed as a 3-D position plus the first two columns of the rotation matrix (6-D), i.e. 9 dims each. Table V lists the per-task layouts.

B. Reward Details

a) WALKGRAB and OPENFRIDGE (single-stage).

Both tasks use a single "continuous" stage, so the per-step reward is a flat weighted sum of shaping terms,

$$r_t \;=\; \sum_k w_k\, r_t^{(k)}, \tag{8}$$

where phase structure is induced by predicate gating *inside* the terms: a grasp predicate $\mathbb{1}[\text{grasped}]$ turns approach terms off and manipulation terms on, and a contact predicate gates the door-opening / lifting rewards so the robot cannot exploit them without touching the object. Tables VI and VII list the active terms.

b) WALKPICKTURN (multi-stage).

A WALKPICKTURN episode is decomposed into stages. Table VIII instantiates $r^{(s)}$ with the stage-dependent shaping terms. More details can be found in Appendix B-E

$$r_t \;=\; \sum_{i=0}^{2} w_i\, \mathbb{1}[\,s_t = i\,]\, r_t^{(i)}, \qquad w_i > 0, \tag{9}$$

C. Hyperparameter Details

All three tasks are trained with PPO under the same actor–critic architecture and optimizer. Only the task-dependent settings (episode length, number of stages, and initial action noise) differ. The actor is a coordinated residual network with a shared coordination trunk that feeds body and hand heads that each output a latent residual on the corresponding frozen prior. The shared PPO settings are listed in Table IX and the per-task differences in Table X.

D. Task-specific success conditions and auxiliary metrics.

See Table XI

E. Demonstration-Free Reference State Initialization

Long-horizon loco-manipulation suffers from a sparse exploration bottleneck. A policy that starts every episode from a default standing pose almost never discovers the later phases of the task by chance, because reaching them requires first solving every earlier phase. A common remedy is reference state initialization (RSI) in VIRAL [9], which resets a fraction of the environments to states drawn from a demonstration so that later phases receive gradient signal from the outset. WALKPICKTURN has no such demonstrations available. We therefore use a *demonstration-free* variant (NoDemoRSI) that bootstraps its own reset distribution from states the policy visits during training, rather than from an external dataset.

a) Stage Decomposition for WALKPICKTURN

We decompose a WALKPICKTURN episode into three stages, indexed 0 to 2, and treat *completing stage 2 as task success*. The policy controls the G1 humanoid at 60 Hz over episodes of 10 s (600 control steps), and a per-environment stage index advances monotonically as the corresponding subgoal is met.

Stage 0 (approach and reach). The robot walks toward the object and brings its open right hand above it. The stage completes when the mean height of the palm and fingertips is within 0.3 m above the object, the mean hand-to-object horizontal distance is below 0.1 m, the right hand is open, both feet are in a stable stance (foot stagger below 0.4 m), and no table collision is detected.

Stage 1 (grasp and lift). The robot closes the hand on the object and lifts it. The stage completes when the object is grasped, defined as a minimum fingertip-to-object distance below 0.08 m together with a positive fingertip contact, and the object is raised more than 0.1 m above its initial height.

Stage 2 (turn). While holding the object, the robot rotates its base to face the target heading. The stage completes when the heading error to the target falls below 0.4 rad.

The transition criteria are one-way. An environment that satisfies the gate for stage $k \to k+1$ has its stage index incremented and never decreases it within an episode. This monotone structure is what lets NoDemoRSI treat the entry state of each stage as a reusable initialization point.

b) Self-Populating Snapshot Buffer

NoDemoRSI maintains one fixed-capacity ring buffer per non-initial stage. Each entry is a *snapshot* of the full simulator state needed to resume an episode from a stage boundary. It includes the robot root state (position, orientation quaternion, and linear and angular velocities), the positions and velocities of all actuated joints, the object root state, the stage index, and the object's initial height. Each buffer holds up to 512 snapshots. Once full, new snapshots overwrite the oldest in first-in-first-out order, so the buffer tracks the policy's current behaviour rather than its early, low-quality behaviour.

Snapshots are collected on the fly. When an environment first transitions into a stage $k \geq 1$, a deferred capture is queued for that environment. On the following control step, provided the environment is still in stage k, the current state is captured and inserted into buffer k, but only if it passes a set of quality filters that reject states unsuitable as reset points. A snapshot is stored only when the robot has not fallen (root height at least 0.5 m), the environment has spent a minimum dwell time in the stage (12 steps for stage 1, 6 steps for stages 2), and the motion is settled rather than mid-transient: root linear velocity below 0.4 m/s, root angular velocity below 0.75 rad/s, joint velocity norm below 10 rad/s, and object linear and angular velocities below 1.0 in their respective units.

TABLE V: Observation terms for the three downstream tasks. Policy and critic share the same layout.

WALKGRAB

State term	Dim.
Stage one-hot	1
Projected gravity	3
Body prior proprio history	450
Right-hand prior proprio	66
Last latent action	28
Fingertip forces on bottle	15
Bottle pose, root	9
Source-table pose, root	9
Bottle pose, hand	9
Body prior latent mean	16
Hand prior latent mean	12
Total	618

WALKPICKTURN

State term	Dim.
Stage one-hot	3
Projected gravity	3
Body prior proprio history	450
Right-hand prior proprio	66
Last latent action	28
Fingertip forces on object	15
Object pose, root	9
Source-table pose, root	9
Object pose, hand	9
Body prior latent mean	16
Hand prior latent mean	12
Total	620

OPENFRIDGE

State term	Dim.
Stage one-hot	1
Projected gravity	3
Body prior proprio history	450
Right-hand prior proprio	66
Last latent action	28
Fingertip forces on handle	15
Fridge pose, root	9
Handle-point pose, root	9
Door pose, root	9
Handle-point pose, hand	9
Door joint state	2
Body prior latent mean	16
Hand prior latent mean	12
Total	629

TABLE VI: WALKGRAB reward components, expressions, and weights. $\overline{d}$ denotes a mean distance over the listed bodies; *grasped* requires mean fingertip–bottle distance < 0.10 m and at least two fingertip contact sensors above 1 N, including thumb contact.

Term	Expression	Weight
Task / object-centric rewards		
Hand–bottle distance	$\frac{1}{2}\left[\exp(-8.5\,\overline{d}_{\text{palm+tips}})\right.$	50.0
Grasp by finger dir	$\dot{\theta}_{\text{thumb,others}}/\pi$	40.0
Fingertip contact	$\text{mean}_i\ \mathbb{1}[\lVert F_i\rVert > 1]$	15.0
Grasp force	$\text{clip}\big(\text{mean}_i \sum \lVert F_i\rVert,\, 0, 2\big)$	15.0
Palm facing object	$\exp\big(-((\cos\angle_{xy} - 0.47)/0.25)^2\big)$	15.0
Thumb opposition	$\sum_i w_i \exp(-\lvert q_i - q_i^*\rvert)$ (4 thumb joints)	20.0
Others open (penalty)	$[1 - \exp(-3\,\overline{\text{dev}})] + 5\,\text{clip}(\Delta\text{err}, 0, 0.02)$	-5.0
Bottle lift height	$(\min(\ell, 0.05)/0.05)\exp(-2\,v_z^2)$, grasp-gated	100.0
Hold bottle	$\mathbb{1}[\text{grasped}]$	100.0
Sustained grasp	$\text{clip}(\text{secs}(\text{grasped} \wedge \ell > 0.01)/1.5, 0, 1)$	100.0
Forward progress	$\text{clip}(v_x^{\text{bottle}}/1, 0, 1)\ \mathbb{1}[\text{grasped} \wedge \ell > 0.01]$	100.0
Locomotion / posture shaping		
Forward target	$\exp\big(-[(x-2)^2 + y^2]\big)$	25.0
Heading forward	$\exp(-(\Delta\psi)^2/0.3^2)$	10.0
Upright orientation	$\exp(-\lVert g_{xy}\rVert^2/0.2^2)$	5.0
y below zero (penalty)	$\text{clip}(-y, 0, 0.5)$	-10.0
Termination / generic penalties		
Success bonus	$\mathbb{1}[\text{success}]$	500.0
Action rate	$\lVert a_t - a_{t-1}\rVert^2$	-0.01
Undesired table contact	$\sum \mathbb{1}[\text{contact} > 5\,\text{N}]$	-20.0
Passed without grasp	$\mathbb{1}[\text{passed-without-grasp}]$	-20.0
Time-out	$\mathbb{1}[\text{time-out}]$	-20.0
Robot fall	$\mathbb{1}[\text{fall}]$	-10.0
Bottle dropped	$\mathbb{1}[\text{dropped}]$ (scaled $\times 5$ once lifted)	-1.0

TABLE VII: OPENFRIDGE reward components, expressions, and weights. $\Delta\theta$ is the door open angle, $c \in \{0, 1\}$ the fingertip-contact predicate; the success angle is 60° and the held-open angle 35°.

Term	Expression	Weight
Task / object-centric rewards		
Hand–handle distance	$\exp(-10\,\overline{d})$ over {5 tips + palm}	30.0
Door open amount	$\text{clip}(\Delta\theta/60^\circ, 0, 1) \cdot c$	30.0
Door open progress	$\text{clip}(\max(\dot{\theta}, 0)/0.02, 0, 1) \cdot c$	20.0
Door held open	$\text{clip}(\text{secs}(\Delta\theta \geq 35^\circ \wedge c)/1, 0, 1)$	100.0
Door close regression	$\text{clip}(\max(-\dot{\theta}, 0)/0.02, 0, 1)$	-20.0
Backward progress	$\text{clip}(v_{\text{back}}/0.10, 0, 1) \cdot c$	20.0
Fingertip contact	$\text{mean}_i \mathbb{1}[f_i > 1] + 0.25(\frac{\max(n_c - 2, 0)}{3})^2$	15.0
Thumb contact	$\mathbb{1}[f_{\text{thumb}} > 1]$	10.0
Termination / generic penalties		
Success bonus	$\mathbb{1}[\text{success}]$	1000.0
Action rate	$\text{clip}(\lVert a_t - a_{t-1}\rVert^2, \leq 100)$	-0.01
Joint action rate	$\text{clip}(\sum(u_t - u_{t-1})^2, \leq 50)$	-0.01
DoF velocity	$\text{clip}(\sum_j \dot{q}_j^2, \leq 200)$	-0.01
Feet slip	$\text{clip}(\sum_{\text{feet}} \mathbb{1}[F > 5]\lVert v_{xy}\rVert, \leq 2)$	-5.0
Large linear v_x	$\min(\max(\lvert v_x\rvert - 0.8, 0), 2)$	-8.0
Large linear v_y	$\min(\max(\lvert v_y\rvert - 0.35, 0), 2)$	-10.0
Large angular ω	$\min(\max(\lvert \omega_z\rvert - 0.6, 0), 5)$	-10.0
Hand far from handle	$\mathbb{1}[\text{hand far}]$	-20.0
Time-out	$\mathbb{1}[\text{time-out}]$	-20.0
Robot fall	$\mathbb{1}[\text{fall}]$	-10.0

c) Reset Sampling and Progressive Stage Unlocking

At every reset, NoDemoRSI assigns each resetting environment a target stage. An environment assigned to stage 0 is reset to the default standing pose, while an environment assigned to a later stage is reset by sampling a snapshot from that stage's buffer and writing it back into the simulator, so training resumes from a state the policy itself previously reached. If a later stage is selected but its buffer is empty, the reset falls back to stage 0.

The per-stage sampling probabilities are computed adaptively. Each stage receives a weight proportional to its difficulty times its availability. Difficulty is $\max(1 - \text{SR}_k,\ \epsilon)^p$ with floor $\epsilon = 0.1$ and power $p = 1$, where SR_k is a smoothed success rate for resets at stage k. A reset counts as a success if the resulting episode advances to the next stage (or, for the final stage, reaches the task success heading condition). The success rate is tracked as an exponential moving average with decay 0.98 and a Beta-style prior of weight 4 centred at 0.5, which keeps the estimate stable when a stage has few recent attempts. Availability is a log-compressed function of the buffer occupancy, so a stage with more stored snapshots is

TABLE VIII: WALKPICKTURN reward components, expressions, weights, and the stage(s) (0–2) where each term is applied. Stages: (0) approach & hand-prep, (1) grasp & lift, (2) turn.

Term	Expression	Weight	Stage(s)
Termination / generic penalties			
Success bonus	$\mathbb{1}[\text{success}]$	1000.0	—
Robot fall	$\mathbb{1}[\text{fall}]$	-10.0	—
Action rate	$\lVert a_t - a_{t-1}\rVert^2$	-0.01	0–2
Joint action rate	$\sum(u_t - u_{t-1})^2$	-0.01	0–2
Left-arm DoF velocity	$\sum \dot{q}^2_{\text{l-arm}}$	-0.01	0–2
DoF velocity	$\sum_j \dot{q}_j^2$	-0.01	0–2
Large linear v_x	$\text{clip}(\lvert v_x\rvert - 0.5, 0, 5)$	-10.0	0–2
Large linear v_y	$\text{clip}(\lvert v_y\rvert - 0.5, 0, 5)$	-10.0	0–2
Large angular ω	$\text{clip}(\lvert\omega_z\rvert - 0.5, 0, 5)$	-5.0	0–2
Feet slip	$\sum_{\text{feet}} \mathbb{1}[\text{contact}]\lVert v_{xy}\rVert$	-5.0	0–2
Undesired table contact	$\sum \mathbb{1}[\text{contact} > 5\,\text{N}]$	-10.0	0–2
Upright orientation	$\exp(-\lVert g_{xy}\rVert^2/0.2^2)$	5.0	0–2
Heading / command shaping			
Stage-progress bonus	$\mathbb{1}[s_t > s_{t-1}]$	500.0	0–2
Turn alignment	$-\lvert\Delta\psi_{\text{region}}\rvert$	15.0	2
Heading to object	$(\Delta\psi_{\text{obj}}/\pi)^2$	-10.0	0
Task / object-centric rewards			
Robot–object distance	$\exp(-4(d - 0.5)^2)$	15.0	0–2
Right-hand target pose	$\exp(-25\lVert\bar{p}_{\text{hand}} - p^*\rVert^2)$	30.0	0
Hand prep-region penalty	$1 - \exp(-(60\,e_{xy}^2 + 20\,e_z^2))$	-10.0	0
Hand flat (stage 0)	$1 - \exp(-(\alpha/0.8)^2)$	-5.0	0
Keep hand open	$\exp(-6\,\lVert q - q_{\text{open}}\rVert) + 5\,\text{prog}$	20.0	0
Hand–object distance	$\exp(-10\,\bar{d})$ over {5 tips + palm}	20.0	1–2
Fingertip contact	contact ratio + count bonus, lift-gated	10.0	1–2
Grasp force	$\tanh\big(\sum\lVert F\rVert_{\text{clip}}/2\big)$, lift-gated	10.0	1–2
Object lift height	$\min(\ell, 0.12)/0.12$	50.0	1–2
Hold object	$\mathbb{1}[\text{grasped}], \mathbb{1}[\ell > \ell_0]$	20.0	1–2
Object not lifted (pen.)	$1 - \text{clip}(\ell/0.1, 0, 1)$	-10.0	1–2

TABLE IX: Shared training hyperparameters (identical across all three tasks).

Hyperparameter	Value
Parallel environments	4096
Steps per environment per rollout	24
Batch size (env $\times$ steps)	98304
Mini-batches	4
Learning epochs per update	5
Discount γ	0.99
GAE λ	0.95
Clip parameter	0.2
Entropy coefficient	0.005
Value loss coefficient	1.0
Learning rate (adaptive, KL target)	1×10^{-3}
Desired KL	0.01
Max gradient norm	1.0
Empirical observation normalization	enabled
Actor / critic hidden dims	[1024, 512, 256]
Activation	ELU
Coord. trunk hidden dims	[512, 256]
Body / hand head hidden dims	[256, 128]
Body / hand residual scale	1.0
Body / hand prior latent dim	16 / 12

TABLE X: Task-dependent settings.

Setting	WALKGRAB	OPENFRIDGE	WALKPICKTURN
Episode length (s)	15.0	7.0	10.0
Control steps	900	420	600
Number of stages	1	1	3
Action dim (latent)	28	28	28
Observation dim	618	629	620
Init. action noise σ	0.22	0.22	0.22

sampled more readily without dominating the difficulty term. When all tracked stages reach a success rate of 0.80, the sampler enters a consolidation mode that spends 85% of resets on stage 0 to rehearse the full chain end to end, and exits when any stage falls back below 0.75. In the non-adaptive fallback, stage 0 instead receives a fixed share of 0.3 and the remaining mass is split across later stages by the weights $(2, 1)$ for stages 1 through 2.

Two mechanisms govern when the sampler may reset into a later stage, which is essential to avoid initializing a stage before the policy can make use of it. First, a stage must have accumulated at least 128 snapshots in its buffer before it is eligible. Second, a progressive warmup unlocks stages in order of their success. A stage becomes available only after the preceding stage is itself available, the preceding stage's success rate exceeds 0.70, and the stage's buffer satisfies the minimum-sample requirement. Immediately after a stage unlocks, its sampling probability is capped and ramped from an initial share of 0.05 up to its full adaptive value over the first 1000 cumulative attempts at that stage. The cap prevents a newly unlocked stage from abruptly absorbing the reset budget and starving the stages the policy has not yet

TABLE XI: Task-specific success conditions and auxiliary metrics.

Task	Success condition	Auxiliary metrics
WALKGRAB	Root forward displacement > 2.0 m, bottle lift > 0.02 m, and grasp is true.	Root x progress, fingertip–bottle distance, contact/grasp rate, bottle lift, carry-forward distance, bottle drop, pass-without-grasp, robot fall.
OPENFRIDGE	Door angle $\geq 60°$, held-open counter ≥ 30 control steps. The held-open counter increments only while door angle $\geq 35°$ and at least one fingertip contact sensor is above 1 N.	Final/max door angle, held-open seconds, hand–handle distance, fingertip contact, backward distance, hand-far termination, robot/joint/fridge blow-up.
WALKPICKTURN	Grasp the cube with minimum fingertip–cube distance < 0.08 m and at least one fingertip contact > 1 N, lifting the cube by > 0.10 m, and completing the turn with heading error < 0.4 rad.	Max stage reached, stage-transition rates, fingertip–cube distance, fingertip contact, cube lift, heading error, cube drop.

TABLE XII: NoDemoRSI configuration for the WALKPICKTURN run. Per-stage entries are listed for stages 0–2; stage 0 has no buffer and is always reset to the default pose.

Setting	**Value**
Buffer capacity per stage	512
Min. snapshots to unlock a stage	128 (stages 1–2)
Min. dwell steps before storing	(–, 12, 6)
Snapshot root-height gate	≥ 0.5 m
Max. root linear / angular velocity	0.4 m/s / 0.75 rad/s
Max. joint velocity norm	10 rad/s
Max. object linear / angular velocity	1.0 / 1.0
Adaptive stage sampling	enabled
Success-rate EMA decay	0.98
Success-rate prior (weight @ 0.5)	4.0
Difficulty floor / power	0.1 / 1.0
Consolidation enter / exit	0.80 / 0.75
Consolidation stage-0 share	0.85
Non-adaptive stage-0 share	0.30
Non-adaptive stage weights	(–, 2, 1)
Warmup unlock mode	previous-stage success
Previous-stage success threshold	0.70
Ramp initial share / attempts	0.05 / 1000

consolidated. Together these rules induce a curriculum that begins almost entirely from the default pose, admits stage 1 resets once the policy reliably reaches and grasps the object, and only later admits turn-stage resets, all without any external demonstration.

APPENDIX C
REAL-WORLD DEMOS

This section provides additional qualitative hardware visualizations and clarifies the hardware variant used for real-robot replay. The quantitative simulation experiments in the main paper are conducted on a Unitree G1 humanoid equipped with a 20-DoF WUJI dexterous hand. In contrast, the physical robot available for our hardware visualization uses a Unitree G1 humanoid equipped with a Dex3-1 [32] dexterous hand. The hardware results in this section should therefore be interpreted as a qualitative trajectory replay on a G1+Dex3-1 platform, rather than as the same G1+WUJI configuration used for the reported simulation success rates.

CoorDex supports such hand variants through its factorized body-hand prior design. The body prior is responsible for whole-body locomotion, reaching, and wrist placement, while the hand prior is specific to the dexterous hand morphology. When replacing the hand, the same pipeline can be instantiated by training a hand specific tracking teacher and distilling it into a hand-specific latent prior and decoder. In our implementation, we train an additional Dex3-1 hand tracking teacher and distill it into a Dex3-1 hand prior using the same procedure as the WUJI hand prior. The downstream residual interface remains structurally unchanged.

For the hardware visualization, we replay recorded joint position trajectories on the physical G1+Dex3-1 robot using the low level joint-position control interface. These trajectories are used to visualize whether the generated body-hand motions are kinematically compatible with the real hardware configuration. We include qualitative snapshots for each task, including WALKGRAB, OPENFRIDGE, and WALKPICKTURN.

REFERENCES

[1] Qingwei Ben, Feiyu Jia, Jia Zeng, Junting Dong, Dahua Lin, and Jiangmiao Pang. Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[2] Jan Brüdigam, Ali-Adeeb Abbas, Maks Sorokin, Kuan Fang, Brandon Hung, Maya Guru, Stefan Sosnowski, Jiuguang Wang, Sandra Hirche, and Simon Le Cleac'h. Jacta: A versatile planner for learning dexterous and whole-body manipulation, 2024. URL https://arxiv.org/abs/2408.01258.

[3] Zixuan Chen, Mazeyu Ji, Xuxin Cheng, Xuanbin Peng, Xue Bin Peng, and Xiaolong Wang. Gmt: General motion tracking for humanoid whole-body control, 2025. URL https://arxiv.org/abs/2506.14770.

[4] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots, 2024. URL https://arxiv.org/abs/2402.16796.

[5] Yuhui Fu, Feiyang Xie, Chaoyi Xu, Jing Xiong, Haoqi Yuan, and Zongqing Lu. Demohlm: From one demonstration to generalizable humanoid loco-manipulation, 2025. URL https://arxiv.org/abs/2510.11258.

[6] Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. Humanplus: Humanoid shadowing and imitation from humans, 2024. URL https://arxiv.org/abs/2406.10454.

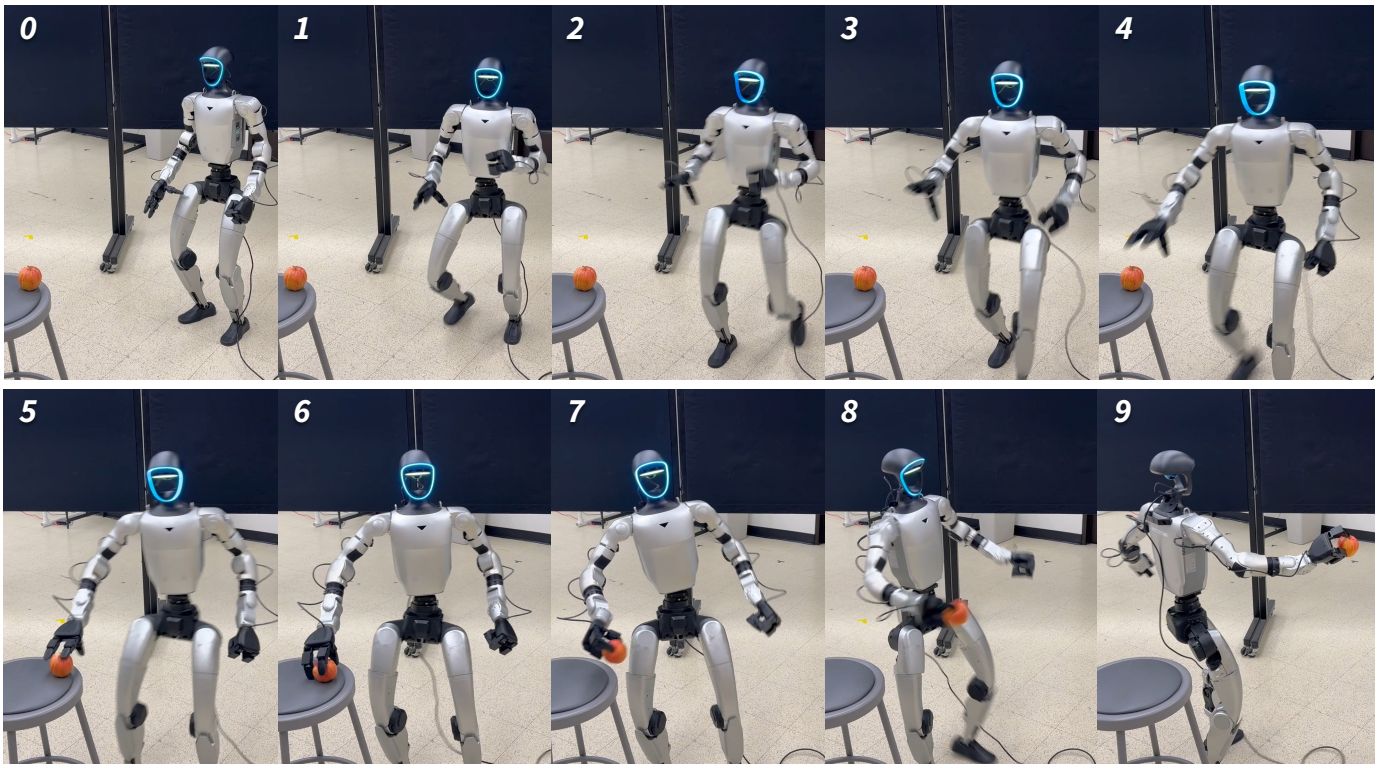

Fig. 5: WALKPICKTURN real-world demo.

[7] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning, 2024. URL https://arxiv.org/abs/2406.08858.

[8] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation, 2024.

[9] Tairan He, Zi Wang, Haoru Xue, Qingwei Ben, Zhengyi Luo, Wenli Xiao, Ye Yuan, Xingye Da, Fernando Castaneda, Shankar Sastry, Changliu Liu, Guanya Shi, Linxi Fan, and Yuke Zhu. Viral: Visual sim-to-real at scale for humanoid loco-manipulation. *arXiv preprint arXiv:2511.15200*, 2025.

[10] Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, Linxi Fan, and Yuke Zhu. Hover: Versatile neural whole-body controller for humanoid robots, 2025. URL https://arxiv.org/abs/2410.21229.

[11] Liang Heng, Yihe Tang, Jiajun Xu, Henghui Bao, Di Huang, and Yue Wang. Humdex: Humanoid dexterous manipulation made easy, 2026.

[12] Mazeyu Ji, Xuanbin Peng, Fangchen Liu, Jialong Li, Ge Yang, Xuxin Cheng, and Xiaolong Wang. Exbody2: Advanced expressive humanoid whole-body control, 2024.

[13] Haoran Jiang, Jin Chen, Qingwen Bu, Li Chen, Modi Shi, Yanjie Zhang, Delong Li, Chuanzhe Suo, Chuang Wang, Zhihui Peng, and Hongyang Li. Wholebodyvla: Towards unified latent vla for whole-body loco-manipulation control, 2025. URL https://arxiv.org/abs/2512.11047.

[14] Zhenyu Jiang, Yuqi Xie, Kevin Lin, Zhenjia Xu, Weikang Wan, Ajay Mandlekar, Linxi Fan, and Yuke Zhu. Dexmimicgen: Automated data generation for bimanual dexterous manipulation via imitation learning, 2025. URL https://arxiv.org/abs/2410.24185.

[15] Yuxuan Kuang, Haoran Geng, Amine Elhafsi, Tan-Dzung Do, Pieter Abbeel, Jitendra Malik, Marco Pavone, and Yue Wang. Skillblender: Towards versatile humanoid whole-body loco-manipulation via skill blending, 2025. URL https://arxiv.org/abs/2506.09366.

[16] Jialong Li, Xuxin Cheng, Tianshu Huang, Shiqi Yang, Ri-Zhao Qiu, and Xiaolong Wang. Amo: Adaptive motion optimization for hyper-dexterous humanoid whole-body control, 2025. URL https://arxiv.org/abs/2505.03738.

[17] Kailin Li, Puhao Li, Tengyu Liu, Yuyang Li, and Siyuan Huang. Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning, 2025. URL https://arxiv.org/abs/2503.21860.

[18] Puhao Li, Tengyu Liu, Yuyang Li, Yiran Geng, Yixin Zhu, Yaodong Yang, and Siyuan Huang. Gendexgrasp: Generalizable dexterous grasping, 2023. URL https://arxiv.org/abs/2210.00722.

[19] Qiayuan Liao, Takara E. Truong, Xiaoyu Huang, Yuman Gao, Guy Tevet, Koushil Sreenath, and C. Karen

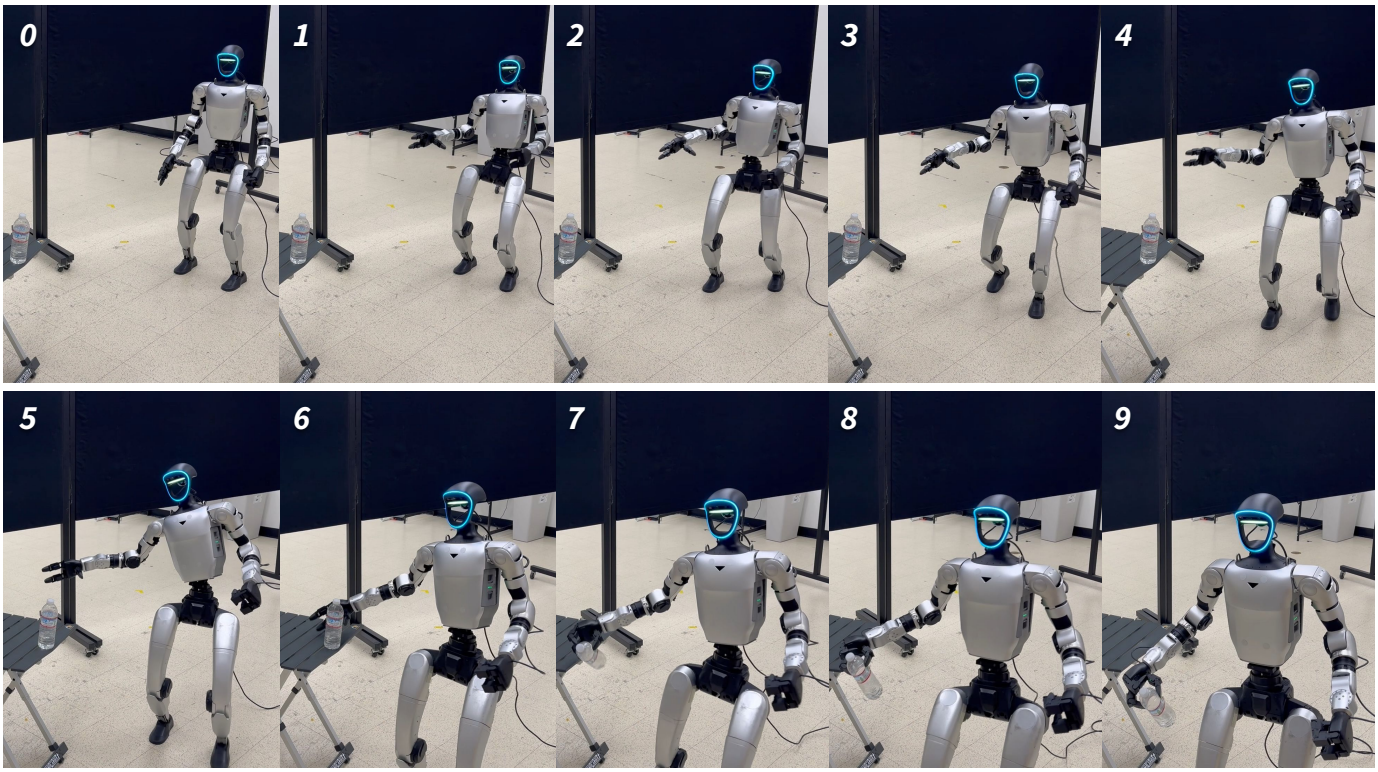

Fig. 6: WALKGRAB real-world demo.

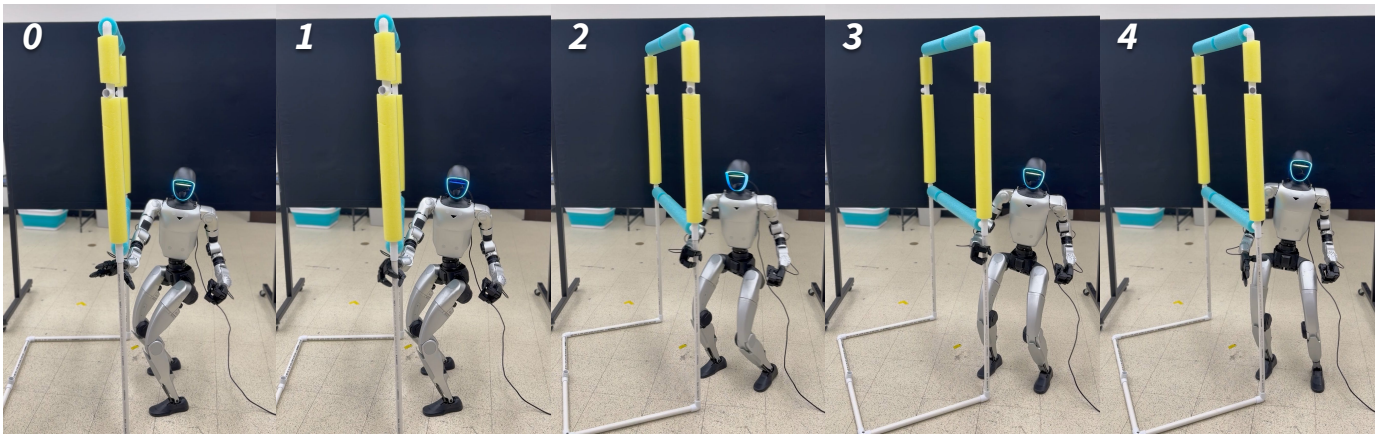

Fig. 7: OPENFRIDGE real-world demo. Due to facility constraints, we use a simplified mock-up instead of a full refrigerator door, focusing on the core behavior of maintaining a grasp while stepping backward to pull the object open.

Liu. Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion, 2025. URL https://arxiv.org/abs/2508.08241.

[20] Fukang Liu, Zhaoyuan Gu, Yilin Cai, Ziyi Zhou, Hyunyoung Jung, Jaehwi Jang, Shijie Zhao, Sehoon Ha, Yue Chen, Danfei Xu, and Ye Zhao. Opt2skill: Imitating dynamically-feasible whole-body trajectories for versatile humanoid loco-manipulation, 2025. URL https://arxiv.org/abs/2409.20514.

[21] Zhengyi Luo, Jinkun Cao, Josh Merel, Alexander Winkler, Jing Huang, Kris Kitani, and Weipeng Xu. Universal humanoid motion representations for physics-based control. In *International Conference on Learning Representations*, 2024.

[22] Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Fernando Castañeda, Sirui Chen, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, Jinhyung Park, David Sami, Zi Wang, Xingye Da, Runyu Ding, Cyrus Hogg, Lina Song, Edy Lim, Eugene Jeong, Tairan He, Haoru Xue, Wenli Xiao, Simon Yuen, Jan Kautz, Yan Chang, Umar Iqbal, Linxi "Jim" Fan, and Yuke Zhu. Sonic: Supersizing motion tracking for natural humanoid whole-body

control, 2026. URL https://arxiv.org/abs/2511.07820.

[23] Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Muñoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, Lukasz Wawrzyniak, Milad Rakhsha, Alain Denzler, Eric Heiden, Ales Borovicka, Ossama Ahmed, Iretiayo Akinola, Abrar Anwar, Mark T. Carlson, Ji Yuan Feng, Animesh Garg, Renato Gasoto, Lionel Gulich, Yijie Guo, M. Gussert, Alex Hansen, Mihir Kulkarni, Chenran Li, Wei Liu, Viktor Makoviychuk, Grzegorz Malczyk, Hammad Mazhar, Masoud Moghani, Adithyavairavan Murali, Michael Noseworthy, Alexander Poddubny, Nathan Ratliff, Welf Rehberg, Clemens Schwarke, Ritvik Singh, James Latham Smith, Bingjie Tang, Ruchik Thaker, Matthew Trepte, Karl Van Wyk, Fangzhou Yu, Alex Millane, Vikram Ramasamy, Remo Steiner, Sangeeta Subramanian, Clemens Volk, CY Chen, Neel Jawale, Ashwin Varghese Kuruttukulam, Michael A. Lin, Ajay Mandlekar, Karsten Patzwaldt, John Welsh, Huihua Zhao, Fatima Anes, Jean-Francois Lafleche, Nicolas Moënne-Loccoz, Soowan Park, Rob Stepinski, Dirk Van Gelder, Chris Amevor, Jan Carius, Jumyung Chang, Anka He Chen, Pablo de Heras Ciechomski, Gilles Daviet, Mohammad Mohajerani, Julia von Muralt, Viktor Reutskyy, Michael Sauter, Simon Schirm, Eric L. Shi, Pierre Terdiman, Kenny Vilella, Tobias Widmer, Gordon Yeoman, Tiffany Chen, Sergey Grizan, Cathy Li, Lotus Li, Connor Smith, Rafael Wiltz, Kostas Alexis, Yan Chang, David Chu, Linxi "Jim" Fan, Farbod Farshidian, Ankur Handa, Spencer Huang, Marco Hutter, Yashraj Narang, Soha Pouya, Shiwei Sheng, Yuke Zhu, Miles Macklin, Adam Moravanszky, Philipp Reist, Yunrong Guo, David Hoeller, and Gavriel State. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025. URL https://arxiv.org/abs/2511.04831.

[24] 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*, 37(4):1–14, 2018. ISSN 1557-7368. doi: 10.1145/3197517.3201311. URL http://dx.doi.org/10.1145/3197517.3201311.

[25] 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*, 40(4):1–20, 2021. ISSN 1557-7368. doi: 10.1145/3450626.3459670. URL http://dx.doi.org/10.1145/3450626.3459670.

[26] 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*, 41(4):1–17, 2022. ISSN 1557-7368. doi: 10.1145/3528223.3530110. URL http://dx.doi.org/10.1145/3528223.3530110.

[27] Yuzhe Qin, Wei Yang, Binghao Huang, Karl Van Wyk, Hao Su, Xiaolong Wang, Yu-Wei Chao, and Dieter Fox. Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. In *Robotics: Science and Systems*, 2023.

[28] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

[29] Wandong Sun, Luying Feng, Yang Liu, Baoshi Cao, Yaochu Jin, and Zongwu Xie. Ulc: A unified and fine-grained controller for humanoid loco-manipulation, 2025.

[30] Jing Tan, Weisheng Xu, Xiangrui Jiang, Jiaxi Zhang, Kun Yang, Kai Wu, Jiaqi Xiong, Shiting Chen, Yangfan Li, Yixiao Feng, Yuetong Fang, Yujia Zou, Yiqun Song, and Renjing Xu. Spherical latent motion prior for physics-based simulated humanoid control, 2026. URL https://arxiv.org/abs/2603.01294.

[31] Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik, and Xue Bin Peng. Calm: Conditional adversarial latent models for directable virtual characters. *ACM Transactions on Graphics*, 2023.

[32] Unitree Robotics. Unitree dex3-1 dexterous hand. https://www.unitree.com/Dex3-1, 2026. Accessed: 2026-05-27.

[33] Unitree Robotics. Unitree g1 humanoid robot. https://www.unitree.com/g1, 2026. Accessed: 2026-05-27.

[34] Ruicheng Wang, Jialiang Zhang, Jiayi Chen, Yinzhen Xu, Puhao Li, Tengyu Liu, and He Wang. Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation, 2023. URL https://arxiv.org/abs/2210.02697.

[35] Zhenyu Wei, Zhixuan Xu, Jingxiang Guo, Yiwen Hou, Chongkai Gao, Zhehao Cai, Jiayu Luo, and Lin Shao. $\mathcal{D}(\mathcal{R},\mathcal{O})$ grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping, 2025. URL https://arxiv.org/abs/2410.01702.

[36] Zhenyu Wei, Yunchao Yao, and Mingyu Ding. One hand to rule them all: Canonical representations for unified dexterous manipulation, 2026. URL https://arxiv.org/abs/2602.16712.

[37] WUJI TECH. Wuji hand product introduction. https://docs.wuji.tech/docs/en/wuji-hand/latest/overview/, 2026. Accessed: 2026-05-27.

[38] Weiji Xie, Jinrui Han, Jiakun Zheng, Huanyu Li, Xinzhe Liu, Jiyuan Shi, Weinan Zhang, Chenjia Bai, and Xuelong Li. Kungfubot: Physics-based humanoid whole-body control for learning highly-dynamic skills, 2025. URL https://arxiv.org/abs/2506.12851.

[39] Haoru Xue, Tairan He, Zi Wang, Qingwei Ben, Wenli Xiao, Zhengyi Luo, Xingye Da, Fernando Castañeda, Guanya Shi, Shankar Sastry, Linxi "Jim" Fan, and Yuke Zhu. Opening the sim-to-real door for humanoid pixel-to-action policy transfer, 2025. URL https://arxiv.org/abs/2512.01061.

[40] Yanjie Ze, Zixuan Chen, João Pedro Araújo, Zi ang Cao, Xue Bin Peng, Jiajun Wu, and C. Karen Liu. Twist: Teleoperated whole-body imitation system, 2025. URL

https://arxiv.org/abs/2505.02833.

[41] Yanjie Ze, Siheng Zhao, Weizhuo Wang, Angjoo Kanazawa, Rocky Duan, Pieter Abbeel, Guanya Shi, Jiajun Wu, and C. Karen Liu. Twist2: Scalable, portable, and holistic humanoid data collection system, 2025. URL https://arxiv.org/abs/2511.02832.

[42] Xinyu Zhan, Lixin Yang, Yifei Zhao, Kangrui Mao, Hanlin Xu, Zenan Lin, Kailin Li, and Cewu Lu. Oakink2: A dataset of bimanual hands-object manipulation in complex task completion, 2024. URL https://arxiv.org/abs/2403.19417.

[43] Yuanhang Zhang, Yifu Yuan, Prajwal Gurunath, Ishita Gupta, Shayegan Omidshafiei, Ali akbar Agha-mohammadi, Marcell Vazquez-Chanlatte, Liam Pedersen, Tairan He, and Guanya Shi. Falcon: Learning force-adaptive humanoid loco-manipulation, 2025. URL https://arxiv.org/abs/2505.06776.

[44] Huihua Zhao, Rafael Cathomen, Lionel Gulich, Wei Liu, Efe Arda Ongan, Michael Lin, Shalin Jain, Soha Pouya, and Yan Chang. Agile: A comprehensive workflow for humanoid loco-manipulation learning, 2026. URL https://arxiv.org/abs/2603.20147.

[45] Siheng Zhao, Yanjie Ze, Yue Wang, C. Karen Liu, Pieter Abbeel, Guanya Shi, and Rocky Duan. Resmimic: From general motion tracking to humanoid whole-body loco-manipulation via residual learning, 2025.