IDEAL: Interaction Dynamics and Force-Aware Learning for Dual-Humanoid Collaborative Manipulation

Zhaoyang Li[1,2*] **Chiyu Zhang**[1,2*] **Chaoyue Li**[1,2*] **Xiao Zhang**[1,2]

Wanting Li[1,2†] **Ziyu Chen**[1,2†]

[1]National Key Laboratory of Multimodal Artificial Intelligence System,
Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
{lizhaoyang2025,zhangchiyu2022,lichaoyue2025,zhangxiao2024,wanting.li,ziyu.chen}@ia.ac.cn

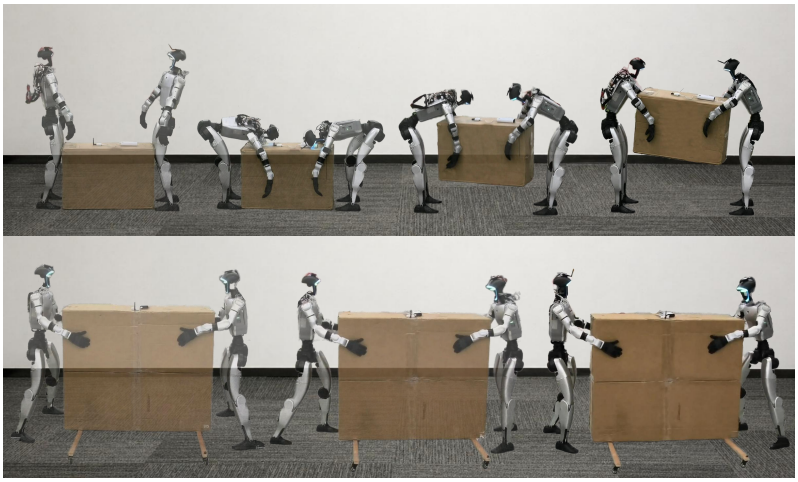


Figure 1: Real-world demonstrations of DHOI. Upper panel illustrates two humanoid robots collaboratively lifting a suitcase, while lower panel shows them jointly pushing a cabinet.

Abstract

Dual-humanoid collaborative manipulation extends the payload and workspace capabilities of a single robot, enabling applications such as large-object carrying and cooperative transportation. However, imitation learning for Dual-Humanoid-Object Interaction (DHOI) remains underexplored, particularly in real-world settings where policies must coordinate motion and maintain force consistency under closed-chain coupling and restricted observations. We propose IDEAL, an imitation learning framework for DHOI that combines interaction-consistent kinematic references with explicit dynamics priors. IDEAL first recovers dual-human motions from monocular videos and refines them for interaction consistency. It then formulates collaborative manipulation as a rigid object grasping problem and solves for desired force allocations under dynamics, contact, and friction constraints. Finally,

*Equal contribution. † Corresponding Authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

IDEAL learns control policies through an interaction-dynamics and force-aware reinforcement learning framework guided by both kinematic and dynamic references. Experimental results show that IDEAL successfully achieves DHOI in both simulation and real-world deployments. As far as we know, **IDEAL is the first force-aware learning pipeline for DHOI deployed on real dual-humanoid robots**, offering a new paradigm for complex real-world humanoid collaboration.

1 Introduction

In real-world settings, humanoid robots are often required to perform manipulation tasks that exceed the payload capacity and workspace of a single robot, such as jointly lifting, carrying, or pushing large objects. While such tasks generally require multiple robots, we define the class of collaborative manipulation problems involving two humanoid robots and a shared rigid object as Dual-Humanoid-Object Interaction (DHOI). DHOI requires not only stable whole-body control, but also physically consistent force coordination mediated by the object. Recent advances in reinforcement learning based whole-body control have enabled humanoid robots to reproduce complex human motions and transfer learned behaviors from simulation to real hardware through reference motion tracking, adversarial motion priors, and large-scale simulation training [1, 2]. Building on this, Humanoid-Object Interaction (HOI) extends humanoid skills to contact-rich object interactions, including grasping, carrying, manipulation, and imitation of interactive behaviors [3–5].

However, existing methods cannot be directly applied to DHOI. First, high-quality DHOI data are scarce. Motion capture provides accurate 3D motion but is expensive, equipment-dependent, and difficult to scale across diverse dual-agent interactions. Monocular videos are more accessible, but existing methods focus on single-human modeling [6, 7] and lack explicit constraints on multi-agent relationships and contact consistency. In dual-human manipulation, occlusion and depth ambiguity often lead to end-effector penetration and inconsistent interactions, making visually estimated motions insufficient as physical DHOI references. Second, prevailing motion tracking methods are mostly limited to kinematic tracking [3, 8]. Unlike the open-chain robot-object system in conventional HOI, two robots in DHOI are coupled through a shared rigid object, forming a closed-chain system [9]. Kinematic references alone are insufficient to constrain physically reasonable force allocation, and policies may generate redundant internal forces that degrade stability and energy utilization.

To address these challenges, we propose **IDEAL** (**I**nteraction **D**ynamics and Forc**E**-**A**ware **L**earning), an imitation learning framework for dual-humanoid collaborative manipulation of non-articulated rigid objects, jointly driven by kinematic data and interaction dynamics. IDEAL comprises three core components. For data construction, we propose Physical Prior-constrained Interaction Motion Optimization (PPIMO), which extends video-based single-human extraction [6] into a DHOI pipeline by mitigating penetration and contact inconsistency using object geometry and contact-aware optimization. The motions are then retargeted to Unitree G1 robots [10]. For dynamics modeling, we propose Dynamics-Constrained Force Allocation (DCFA) to introduce an explicit dynamics prior into the kinematics-driven tracking. By formulating DHOI as a rigid object grasping problem, DCFA converts discrete contact forces into a physically consistent desired force allocation by minimizing closed-chain internal forces [11]. Finally, for policy learning, we propose the Interaction Dynamics and Force-Aware Reinforcement Learning Framework (IDRL). Guided by both kinematic and dynamic references, IDRL incorporates the desired force distributions into multi-agent policy learning through an Interaction Dynamics Encoder and force tracking rewards.

As far as we know, IDEAL is the first force-aware learning pipeline for dual-humanoid collaborative object manipulation deployed on real hardware. Our contributions are fourfold:

(1) We develop a physics-constrained dual-agent motion extraction and retargeting pipeline that produces contact-consistent robot motions from monocular videos.

(2) We formulate a dynamics-constrained force allocation method that generates physically feasible force targets while minimizing closed-chain internal forces.

(3) We introduce a force-aware multi-agent imitation-learning framework that integrates interaction dynamics and force-tracking objectives into policy learning.

(4) We validate IDEAL through simulation and real-world experiments, demonstrating improved task success, motion-tracking accuracy, and force coordination.

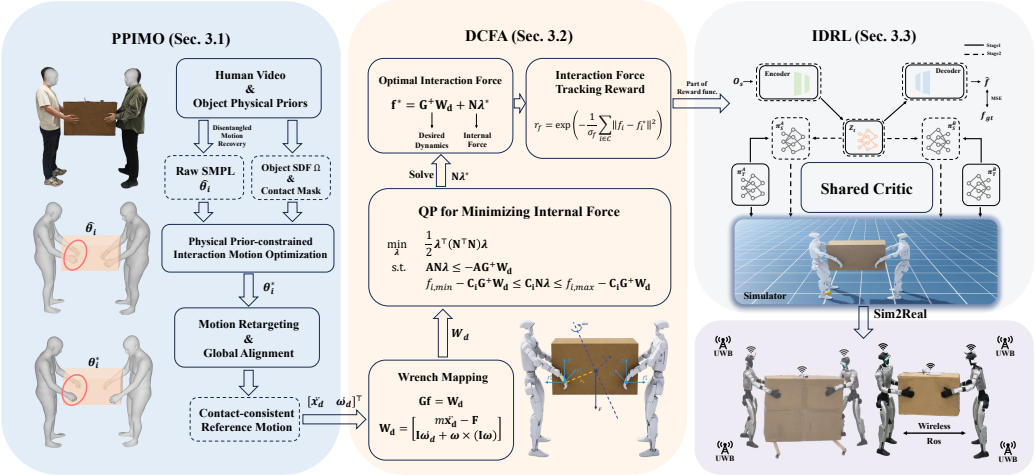


Figure 2: Overview of IDEAL. IDEAL tackles DHOI through motion-and-force guided imitation learning. PPIMO builds contact-consistent motion references from monocular demonstrations. DCFA computes desired interaction forces as dynamics references, and IDRL learns Sim-to-Real cooperative control guided by both motion and force references.

2 Related Work

Human Motion Estimation and Robot Retargeting. Extracting human motion from videos and retargeting it to robots is a common low cost way to obtain reference motions. SMPL [12] provides a standardized parametric representation of the human body, while later monocular motion estimation methods [6, 7, 13] improve the recovery of human motion in the world frame. For robot retargeting, classical inverse kinematics [14] is simple and widely used, but its adaptability is limited when the robot morphology or task setting changes. Recent learning-based retargeting methods [10, 15, 16] and object-aware designs [17] improve generalization across bodies and interaction settings. Nevertheless, these methods are mainly developed for single-person motions and do not explicitly model physical contact in dual-agent interaction. Based on this line of work, our method introduces physics-based contact constraints to post-optimize dual-agent poses and reduce interaction errors.

Physics-Based Motion Imitation via Reinforcement and Imitation Learning. Physics-based motion imitation has become an important paradigm for learning humanoid robot skills. Reference motion tracking [18], adversarial motion priors [1], reusable skill embeddings [19, 20], diffusion-based models [21], and unified motion inpainting [22] reduce the dependence on frame-wise tracking and allow policies to learn natural motion styles from large-scale motion data. Residual force control [23] and universal humanoid motion representations [24, 25] further increase motion diversity and robustness. Recent studies [8, 26–32] have also shown that these techniques can be transferred to real humanoid hardware. For HOI, physics-based interaction frameworks [4, 33], unified grasping policies [34], and video-driven skill learning pipelines [3] have been proposed. However, these studies still focus on single-robot open-chain settings and do not address force coupling in multi-robot closed-chain interaction. Our work follows the motion imitation paradigm, but further incorporates dynamics priors and interaction force modeling for dual-agent tasks.

Multi-Agent Control for Cooperative Manipulation. Multi-agent cooperative control provides useful tools for dual-robot collaboration. The CTDE paradigm [35, 36] and teacher-student distillation [37, 38] offer effective mechanisms for cooperative optimization and sim-to-real transfer. In computer graphics, multi-character interaction [39] and adversarial interaction priors [40] model coupled multi-agent dynamics, although their main goal is visual realism rather than robotic deployment under physical constraints. In dual-humanoid robotics, It Takes Two [41], Rhythm [42], and RoboStriker [43] demonstrate real-world dual-humanoid control for handshaking, dancing, and boxing, but they mainly consider human-human style interaction without object manipulation. For cooperative object manipulation, CooHOI [44] and SynAgent [45] study multi-humanoid collaboration only in simulation, while COLA [46] realizes real-world collaborative carrying with a compliant human

partner. DA-VIL [9] and H2-COMPACT [47] further investigate closed-chain force coordination in dual-arm systems. Despite these advances, existing methods either rely on implicit dynamic propagation without explicit force constraints, or consider force coordination mainly at the control level. In contrast, our work explicitly injects physically optimized force distributions into multi-agent policy learning.

3 Method

As illustrated in Fig. 2, the **IDEAL** framework addresses DHOI tasks through three core modules: motion extraction, force allocation, and policy learning. Section 3.1 introduces **Physical Prior-constrained Interaction Motion Optimization (PPIMO)**, a pipeline that extracts dual-human motions from videos and refines them via contact-aware optimization to ensure physical plausibility. Section 3.2 presents **Dynamics-Constrained Force Allocation (DCFA)**, which computes an optimal, internal force minimizing contact force distribution based on rigid object dynamics. Finally, Section 3.3 details the **Interaction Dynamics and Force-Aware Reinforcement Learning (IDRL)** framework, where a multi-agent policy is trained to track both the optimized kinematic references and the analytically computed force targets.

3.1 Physical Prior-constrained Interaction Motion Optimization

To extract reference motions for dual robots from monocular videos, a natural extension of single-robot methods involves decoupling the agents, estimating local motions, and aligning them globally. However, this decoupled mapping pipeline degrades substantially in complex human-object interactions. First, existing single-robot methods lack explicit physical contact modeling. Second, video acquisition inherently introduces a trade-off. Capturing agents separately makes global alignment unreliable, while capturing them together forces a viewpoint compromise: an unbalanced perspective causes severe occlusion for one agent, whereas a balanced view introduces projection degeneracy, complicating depth estimation along the optical axis. This systematic depth bias, combined with a lack of physical constraints, ultimately leads to end-effector penetration and contact inconsistency.

To address these issues, we adopt the balanced side-view perspective to minimize occlusion and resolve the resulting depth bias through a physically constrained post-optimization framework. Following the decoupling strategy, we first independently model each agent $i \in \{1, 2\}$ and estimate its initial SMPL parameters $\hat{\theta}_i$ using GVHMR [6]. We additionally annotate the object trajectory and contact signals. To correct the initial depth biases and ensure physical plausibility, we optimize each $\hat{\theta}_i$ over the entire sequence by explicitly modeling human-object contact.

The core of this optimization is the contact-aware interaction energy, $E_C(\theta)$, which is based on the object's signed distance field (SDF) [48]. The object is assumed to occupy a domain $\Omega \subset \mathbb{R}^3$ with boundary $\partial\Omega$, where the signed distance for a spatial point $\mathbf{x}$ is positive when $\mathbf{x} \notin \Omega$. Given the end-effector position computed by forward kinematics $K(\theta)$ and a contact indicator $\mathbb{I}_{\mathcal{C}}$, we define:

$$E_C(\theta) = \begin{cases} \left(\max\left(0, \epsilon - \mathrm{SDF}(K(\theta), \partial\Omega)\right)\right)^2, & \text{if } \mathbb{I}_{\mathcal{C}} = 0 \\ \left(\max\left(0, |\mathrm{SDF}(K(\theta), \partial\Omega)| - \delta\right)\right)^2, & \text{if } \mathbb{I}_{\mathcal{C}} = 1 \end{cases} \tag{1}$$

Here, θ denotes the optimization variable θ_i for a single subject. This term prevents penetration by penalizing distances below ϵ when there is no contact ($\mathbb{I}_{\mathcal{C}} = 0$), and anchors the end effector within a contact thickness δ when contact occurs ($\mathbb{I}_{\mathcal{C}} = 1$). Further details are in Appendix B.1.

To guarantee that the optimized poses remain feasible and realistic, we introduce two additional energy terms. First, a fidelity energy, $E_D(\theta) = \|\theta - \hat{\theta}\|^2$, adaptively restricts deviations from the initial network prediction for each joint. Second, a regularization energy is applied at each temporal frame to ensure temporal smoothness and kinematic feasibility, defined as

$$E_R(\theta) = \|\theta_t - \theta_{t-1}\|_2^2 + \alpha \|\theta_t - 2\theta_{t-1} + \theta_{t-2}\|_2^2, \tag{2}$$

where t denotes the current temporal frame. This regularization softly penalizes abrupt inter-frame changes in velocity and acceleration.

The overall post-optimization objective balances these constraints:

$$J(\theta) = \lambda_C E_C(\theta) + \lambda_D E_D(\theta) + \lambda_R E_R(\theta) \tag{3}$$

where λ_C, λ_D, and λ_R control the relative importance of contact consistency, data fidelity, and temporal smoothness.

By solving $\theta_i^* = \arg\min_\theta J(\theta)$ across all frames for both agents, we obtain corrected and physically plausible pose sequences, as shown in Fig. 2. Finally, these optimized human sequences are retargeted to the robot joint space using GMR [10] and aligned in the global coordinate frame, successfully yielding the final feasible reference motions for the dual-robot system.

3.2 Dynamics-Constrained Force Allocation for DHOI

DHOI inherently involves multi-point contact and force-mediated collaboration. Consider a representative rigid-object manipulation scenario in which two humanoid robots jointly carry a bulky object by establishing four hand-object contact points, as illustrated in Fig. 2. High-quality manipulation requires satisfying two simultaneous control objectives: first, generating the resultant net force and moment required to compensate for gravity and track a desired object trajectory; second, allocating these contact forces such that the robots minimize their reliance on closed-chain internal forces, thereby preserving independent upright balance and individual postural stability.

We model the carried box as a rigid object with mass m, inertia tensor $\mathbf{I}$, and center-of-mass (CoM) position $\mathbf{p}_{\text{CoM}} \in \mathbb{R}^3$. The four stable contact points are denoted by $\mathbf{p}_i \in \mathbb{R}^3$ ($i \in \{1, 2, 3, 4\}$), and the corresponding interaction forces exerted by the robots are denoted by $\mathbf{f}_i \in \mathbb{R}^3$. The stacked contact-force vector is defined as $\mathbf{f} = \left[\mathbf{f}_1^\top, \mathbf{f}_2^\top, \mathbf{f}_3^\top, \mathbf{f}_4^\top\right]^\top$.

To track a specified kinematic reference motion, the box must achieve the desired linear acceleration $\ddot{\mathbf{x}}_d \in \mathbb{R}^3$ and angular acceleration $\dot{\boldsymbol{\omega}}_d \in \mathbb{R}^3$. According to the Newton-Euler equations, the corresponding object-level motion requirement can be compactly represented by a desired spatial wrench $\mathbf{W}_d$ applied at the CoM:

$$\mathbf{W}_d = \begin{bmatrix} m\ddot{\mathbf{x}}_d - \mathbf{F} \\ \mathbf{I}\dot{\boldsymbol{\omega}}_d + \boldsymbol{\omega} \times (\mathbf{I}\boldsymbol{\omega}) \end{bmatrix} \tag{4}$$

where $\mathbf{F} = [F_x \quad F_y \quad F_z]^\top \in \mathbb{R}^3$ denotes the resultant external force exerted on the object by sources other than the dual-robot system, and $\boldsymbol{\omega} \in \mathbb{R}^3$ is the angular velocity of the box. The detailed derivation is provided in Appendix B.2.

The mapping from the distributed contact forces to this resultant CoM wrench is governed by the grasp matrix $\mathbf{G}$, such that $\mathbf{G}\mathbf{f} = \mathbf{W}_d$ [49]. Here, $\mathbf{G}$ is constructed by concatenating four contact-level submatrices, $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4]$, where each submatrix is defined as $\mathbf{G}_i = \begin{bmatrix} \mathbf{I}_{3\times3} \\ \hat{\mathbf{r}}_i \end{bmatrix}$. In this formulation, $\mathbf{I}_{3\times3}$ is the identity matrix, $\mathbf{r}_i = \mathbf{p}_i - \mathbf{p}_{\text{CoM}}$ is the contact position relative to the center of mass, and $\hat{\mathbf{r}}_i$ is its skew-symmetric matrix.

Because four contact points provide a 12-dimensional force input for a 6-dimensional wrench requirement, this mapping is underdetermined. We therefore decompose the solution space into a task-space particular solution and a null-space internal-force component:

$$\mathbf{f} = \mathbf{G}^+\mathbf{W}_d + \mathbf{N}\boldsymbol{\lambda} \tag{5}$$

where $\mathbf{G}^+$ is the pseudo-inverse of $\mathbf{G}$, $\mathbf{N}$ is a basis of the null space of the grasp matrix, and $\boldsymbol{\lambda}$ denotes the internal-force coefficient vector. The particular solution $\mathbf{G}^+\mathbf{W}_d$ is uniquely determined by the desired inertial motion. Conversely, the null-space component $\mathbf{N}\boldsymbol{\lambda}$ represents internal squeezing or pulling forces between the robots and the shared object. Relying excessively on these closed-chain internal forces can induce unstable postures, such as severe backward leaning. To avoid this, we seek to minimize the internal forces using the objective function $J(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\lambda}^\top(\mathbf{N}^\top\mathbf{N})\boldsymbol{\lambda}$.

The allocated forces must also adhere to strict physical interaction boundaries. First, non-slip contact is enforced through a linearized Coulomb friction cone, $\mathbf{A}\mathbf{f} \leq \mathbf{0}$, where $\mathbf{A}$ is the friction constraint matrix. Second, normal contact forces are strictly bounded to ensure physical validity and hardware safety: $f_{i,\min} \leq \mathbf{C}_i\mathbf{f} \leq f_{i,\max}$, where $\mathbf{C}_i$ extracts the normal component of the contact force [50]. Substituting the decomposition from Eq. 5 into these bounds elegantly transforms the physical constraints over $\mathbf{f}$ into linear inequalities strictly over $\boldsymbol{\lambda}$.

By integrating the force decomposition with these transformed constraints, low-level interaction force allocation is reformulated as a quadratic program (QP), with $\boldsymbol{\lambda}$ as the sole decision variable:

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & \frac{1}{2}\boldsymbol{\lambda}^\top(\mathbf{N}^\top\mathbf{N})\boldsymbol{\lambda} \\ \text{s.t.} \quad & \mathbf{AN}\boldsymbol{\lambda} \leq -\mathbf{AG}^+\mathbf{W}_d \\ & f_{i,\min} - \mathbf{C}_i\mathbf{G}^+\mathbf{W}_d \leq \mathbf{C}_i\mathbf{N}\boldsymbol{\lambda} \leq f_{i,\max} - \mathbf{C}_i\mathbf{G}^+\mathbf{W}_d \end{aligned} \tag{6}$$

In this formulation, the equality constraint $\mathbf{Gf} = \mathbf{W}_d$ is implicitly satisfied by the null-space projection, substantially reducing the dimensionality of the optimization. Frame by frame, given the desired motion from kinematic data and the contact-position prior, we construct the generalized force $\mathbf{W}_d$, update the constraints, and solve for the optimal coefficient $\boldsymbol{\lambda}^*$. The corresponding optimal interaction force $\mathbf{f}^*$ is recovered via Eq. 5, yielding reference forces that are dynamically consistent, frictionally stable, and perfectly suited to provide dynamics-level guidance for subsequent reinforcement learning.

3.3 Interaction Dynamics and Force-Aware Learning Framework

We formulate the DHOI task as a multi-agent Markov decision process (MA-MDP). To achieve coordinated control in this contact-rich environment, we propose the IDRL framework. Unlike standard methods relying solely on kinematics, IDRL explicitly models and constrains interaction forces, integrating them into decision-making to better exploit contact dynamics.

Asymmetric MARL Architecture IDRL employs Multi-Agent PPO (MAPPO) under the centralized training and decentralized execution (CTDE) paradigm [35, 36]. To handle partial observability during deployment, we introduce an asymmetric teacher-student mechanism [51]. A teacher policy π_T leverages full state information and reference motion based residual control to learn feasible collaborative behaviors. A student policy π_S relies solely on local observations, distilling knowledge from π_T. Training proceeds in two stages: pretraining π_T and the representation modules using full states to favorably initialize π_S, followed by online RL fine-tuning of π_S under constrained observations. The detailed design of observation space is provided in Appendix B.3.1.

Interaction Dynamics Encoder Since real interaction forces are unavailable to π_S during execution, we introduce a VAE-based [52] Interaction Dynamics Encoder (IDE). It encodes a multi-frame sequence of local observations (including proprioceptive, collaborator, and contact signals) into a latent variable z_t. A decoder reconstructs the estimated force $\hat{f}$, supervised by the ground-truth force f_{gt} during training in the simulator. The objective is:

$$\mathcal{L}_{\text{IDE}} = \text{MSE}(\hat{f}, f_{gt}) + \beta \cdot D_{\text{KL}}\big(q(z_t|\cdot) \parallel \mathcal{N}(0, I)\big) \tag{7}$$

The latent representation z_t is injected into the observation space of π_S as a predicted force feature, enhancing its perception of complex contact dynamics.

Interaction Force Tracking Reward To ensure the physical consistency of the learned policy and explicitly link the analytical optimization to the learning process, we introduce a tracking reward based on the DCFA detailed in Section 3.2. This reward penalizes deviations between the actual interaction force f_i and the optimal target force f_i^*, which is pre-computed by solving the QP formulation in Eq. 6 at contact points $i \in C$:

$$r_f = \exp\left(-\frac{1}{\sigma_f}\sum_{i\in C}\|f_i - f_i^*\|^2\right) \tag{8}$$

Here, σ_f is a sensitivity scaling factor. The detailed design of the remaining reward terms is provided in Appendix B.3.2.

4 Experiments

In this section, we conduct a series of experiments in both simulation and the real world to answer the following four questions: **Q1.** What performance gains does PPIMO bring to policy optimization

through optimized reference motions? **Q2.** Do desired interaction forces planned by DCFA improve task performance through IDE and interaction-force tracking rewards? **Q3.** Does IDRL outperform approaches that model the agents as independent entities in shared-object manipulation? **Q4.** How well does IDEAL perform when deployed in real-world environments?

4.1 Experimental Setup

Our policy training is conducted in Isaac Sim [53] using 4096 parallel environments. To facilitate sim-to-sim and sim-to-real transfer, we apply domain randomization during training, whose details are provided in Appendix B.3.3. The complete two-stage training procedure takes approximately 10 hours on a single NVIDIA RTX 4090 GPU. For evaluation, we consistently select and compare the checkpoints obtained after training convergence for each baseline.

Evaluation Metrics To quantitatively assess task and tracking performance, we employ the following metrics, whose detailed definitions are provided in Appendix C.2. All robot-related metrics are averaged across the two robots. We evaluate: Success Rate (SR, %); whole-body metrics including Global Mean Per Body Position Error ($E_{\text{g-mpbpe}}$, mm), root-relative Mean Per Body Position (E_{mpbpe}, mm) and Orientation (E_{mpboe}, 10^{-3} rad) Error; joint-level metrics comprising Mean Per Joint Position (E_{mpjpe}, 10^{-3} rad) and Velocity (E_{mpjve}, 10^{-3} rad/frame) Error; and object-tracking metrics including Mean Per Object Position (E_{mpope}, mm) and Orientation (E_{mpooe}, 10^{-3} rad) Error.

4.2 Main Results

Ablation Study on PPIMO-Optimized Reference Motions Addressing Q1, we compare policies trained with the original reference motions and those trained with PPIMO-optimized reference motions on three tasks: Box Transport (BT), Cabinet Pushing (CP) and Suitcase Lifting (SL). All policies are evaluated under the same protocol after convergence. The results are reported in Table 1. As shown in the results, utilizing the optimized reference motions consistently improves all key evaluation metrics across the three tasks. Specifically, tracking accuracy is significantly enhanced, evidenced by substantial reductions in body and joint-level errors (e.g., E_{mpboe} and E_{mpjpe}). Furthermore, the varying degrees of improvement in Success Rate (SR) across tasks demonstrate that these optimized reference motions bolster not only tracking precision but also the overall stability and consistency of task execution.

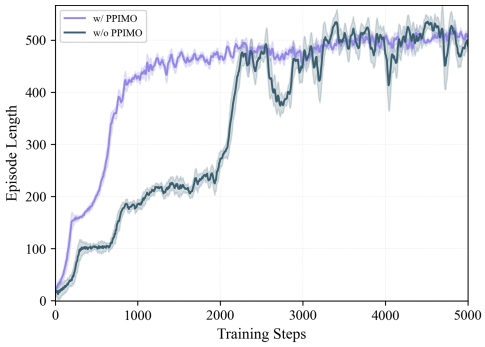


Figure 3: Episode length comparison during training on the BT task.

Notably, PPIMO significantly reduces physically implausible penetrations during object interaction, resulting in reference motions that are more temporally coherent and physically feasible. Moreover, the resulting reliable interaction states improve the effectiveness of training initialization from randomly sampled reference frames, thereby enhancing training stability and sample efficiency, as illustrated by the episode length comparison in Fig. 3.

Ablation Study on Interaction Force Constraints In response to Q2, we compare IDEAL with the multi-agent cooperative method CooHOI [44] in Table 2. CooHOI primarily targets task-level cooperative control without explicitly modeling internal forces in closed-chain interactions. In contrast, IDEAL is guided by both kinematic motion references and interaction dynamics references, enabling explicit force coordination between the two robots. Across BT, CP, and SL, IDEAL outperforms CooHOI on all robot- and object-tracking metrics, reproducing the whole-body motions of both robots and the shared-object trajectories more accurately. These results motivate a closer examination of the role of explicit interaction force constraints.

We then conduct an ablation study on the SL task to evaluate the contributions of the interaction force tracking reward and the IDE module. As illustrated in Fig. 4(a), without proper constraints, a typical internal force antagonism phenomenon occurs: the two robots exert redundant, opposing

Table 1: Effect of PPIMO-optimized reference motions on policy performance across three tasks.

Method	SR ↑	$E_{\text{g-mpbpe}}$ ↓	E_{mpbpe} ↓	E_{mpboe} ↓	E_{mpjpe} ↓	E_{mpjve} ↓	E_{mpope} ↓	E_{mpooe} ↓
Box Transport (BT)								
w/o PPIMO	88	$101.35_{\pm 6.22}$	$67.28_{\pm 2.93}$	$388.56_{\pm 9.05}$	$208.41_{\pm 6.34}$	$15.79_{\pm 0.53}$	$51.98_{\pm 5.95}$	$52.74_{\pm 7.81}$
w/ PPIMO	**100**	$\mathbf{93.37}_{\pm \mathbf{4.97}}$	$\mathbf{43.88}_{\pm \mathbf{1.71}}$	$\mathbf{252.59}_{\pm \mathbf{6.13}}$	$\mathbf{137.38}_{\pm \mathbf{2.74}}$	$\mathbf{10.35}_{\pm \mathbf{0.15}}$	$\mathbf{47.50}_{\pm \mathbf{6.24}}$	$\mathbf{34.14}_{\pm \mathbf{8.85}}$
Cabinet Pushing (CP)								
w/o PPIMO	96	$88.93_{\pm 9.03}$	$90.35_{\pm 4.61}$	$391.32_{\pm 11.24}$	$217.06_{\pm 8.73}$	$8.97_{\pm 0.97}$	$160.28_{\pm 14.92}$	$129.65_{\pm 21.33}$
w/ PPIMO	**100**	$\mathbf{81.44}_{\pm \mathbf{8.24}}$	$\mathbf{65.11}_{\pm \mathbf{2.98}}$	$\mathbf{253.42}_{\pm \mathbf{7.98}}$	$\mathbf{142.79}_{\pm \mathbf{4.39}}$	$\mathbf{5.84}_{\pm \mathbf{0.54}}$	$\mathbf{147.68}_{\pm \mathbf{10.64}}$	$\mathbf{84.46}_{\pm \mathbf{18.76}}$
Suitcase Lifting (SL)								
w/o PPIMO	93	$92.13_{\pm 4.79}$	$71.47_{\pm 1.88}$	$410.26_{\pm 5.71}$	$247.98_{\pm 4.24}$	$12.91_{\pm 0.33}$	$43.24_{\pm 3.15}$	$31.35_{\pm 2.20}$
w/ PPIMO	**100**	$\mathbf{85.09}_{\pm \mathbf{3.88}}$	$\mathbf{47.63}_{\pm \mathbf{0.37}}$	$\mathbf{268.62}_{\pm \mathbf{2.43}}$	$\mathbf{164.87}_{\pm \mathbf{1.55}}$	$\mathbf{7.99}_{\pm \mathbf{0.15}}$	$\mathbf{39.84}_{\pm \mathbf{2.57}}$	$\mathbf{20.15}_{\pm \mathbf{1.36}}$

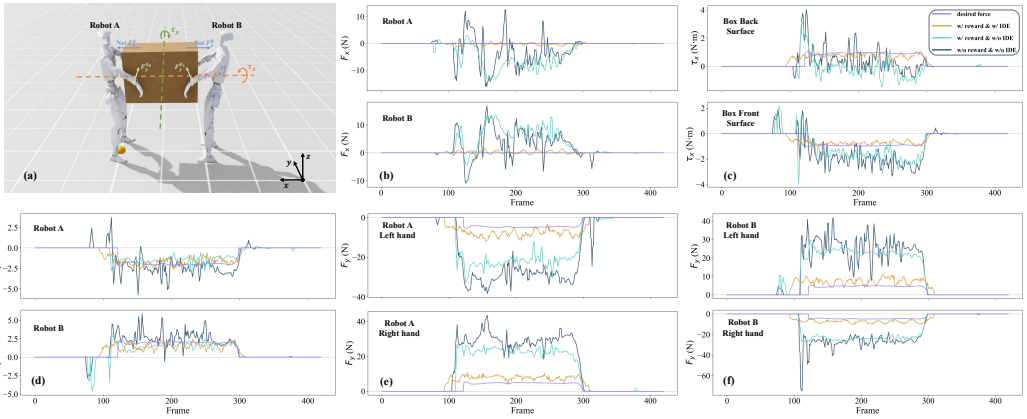


Figure 4: Ablation analysis of interaction force constraint in the SL task. (a) Illustration of internal force conflict between two robots; (b) resultant force along the x direction at both object ends; (c) resultant moment around the x axis; (d) resultant moment around the y axis; (e) y-direction hand forces of robot A; (f) y-direction hand forces of robot B.

forces at both ends of the object to maintain their individual balance, leading to ineffective energy consumption and unstable interactions.

To analyze and address this issue, we define three ablation settings: S1 (w/o reward & w/o IDE) serves as the baseline without either component; S2 (w/ reward & w/o IDE) introduces only the tracking reward; and S3 (w/ reward & w/ IDE) incorporates both. Fig. 4(b) and (c) present the resultant force along the x-direction and the resultant torque around the x-axis at the object's two ends, respectively. Under S1, evident force cancellation occurs. In contrast, S3 significantly reduces the overall magnitudes of both the resultant force and torque, with temporal trends closely matching the desired force distribution computed by DCFA. This improvement is further supported by the resultant torque around the y-axis (Fig. 4(d)) and the y-direction forces exerted by the hands of robot A and B (Fig. 4(e) and (f)). The results demonstrate that S3 yields the smallest fluctuations in the y-axis torque. Furthermore, both robots' vertical forces closely align with the DCFA-computed targets, effectively suppressing redundancy. While S2 reduces part of the redundant force compared to S1, it still produces non-negligible ineffective outputs.

Overall, the interaction-force tracking reward partially alleviates redundant forces, while the IDE module further enhances the policy's ability to match the desired force distribution. Combining them ensures more physically reasonable force interactions, thereby minimizing redundant outputs and lowering overall system energy consumption.

Ablation Study on Coordination Mechanism During Training To tackle Q3, we conduct comparative experiments on the SL, BT, and CP tasks against HDMI [3], a single-robot baseline. Single-robot HOI forms an open-chain interaction system, in which a policy can typically maintain self-balance while manipulating an object. We investigate whether this self-balancing capability is sufficient to handle internal force conflicts when independently trained policies are directly deployed in a closed-chain DHOI system. To adapt HDMI for this evaluation, each robot is treated as an independent entity and trained to manipulate an object whose longest side is scaled down to one-quarter of its

Table 2: Comparison of IDEAL with an independently trained HDMI baseline and CooHOI across three tasks.

Method	SR ↑	$E_{\text{g-mpbpe}}$ ↓	E_{mpbpe} ↓	E_{mpboe} ↓	E_{mpjpe} ↓	E_{mpjve} ↓	E_{mpope} ↓	E_{mpooe} ↓
Box Transport (BT)								
HDMI	36	$125.23_{\pm 7.19}$	$\mathbf{38.10_{\pm 1.97}}$	$\mathbf{226.48_{\pm 8.30}}$	$\mathbf{121.37_{\pm 3.83}}$	$\mathbf{9.64_{\pm 0.32}}$	$69.28_{\pm 7.84}$	$109.11_{\pm 9.43}$
CooHOI	43	$128.96_{\pm 1.82}$	$78.62_{\pm 8.21}$	$423.38_{\pm 7.84}$	$156.61_{\pm 3.26}$	$27.32_{\pm 8.52}$	$53.62_{\pm 9.16}$	$62.51_{\pm 6.28}$
Ours	**100**	$\mathbf{93.37_{\pm 4.97}}$	$43.88_{\pm 1.71}$	$252.59_{\pm 6.13}$	$137.38_{\pm 2.74}$	$10.35_{\pm 0.15}$	$\mathbf{47.50_{\pm 6.24}}$	$\mathbf{34.14_{\pm 8.85}}$
Cabinet Pushing (CP)								
HDMI	47	$118.77_{\pm 19.20}$	$\mathbf{61.32_{\pm 3.16}}$	$\mathbf{247.81_{\pm 13.55}}$	$\mathbf{134.22_{\pm 5.62}}$	$5.96_{\pm 0.73}$	$212.45_{\pm 16.85}$	$151.58_{\pm 20.31}$
CooHOI	58	$116.82_{\pm 1.35}$	$86.28_{\pm 6.32}$	$327.81_{\pm 6.67}$	$187.70_{\pm 3.26}$	$25.07_{\pm 8.81}$	$167.66_{\pm 6.81}$	$102.43_{\pm 3.88}$
Ours	**100**	$\mathbf{81.44_{\pm 8.24}}$	$65.11_{\pm 2.98}$	$253.42_{\pm 7.98}$	$142.79_{\pm 4.39}$	$\mathbf{5.84_{\pm 0.54}}$	$\mathbf{147.68_{\pm 10.64}}$	$\mathbf{84.46_{\pm 18.76}}$
Suitcase Lifting (SL)								
HDMI	32	$123.40_{\pm 6.21}$	$\mathbf{41.59_{\pm 1.23}}$	$\mathbf{244.66_{\pm 5.16}}$	$\mathbf{151.27_{\pm 3.48}}$	$\mathbf{7.12_{\pm 0.24}}$	$58.94_{\pm 4.65}$	$98.35_{\pm 3.13}$
CooHOI	71	$100.72_{\pm 1.73}$	$63.41_{\pm 3.61}$	$373.25_{\pm 8.69}$	$216.43_{\pm 2.55}$	$29.91_{\pm 7.61}$	$67.71_{\pm 7.84}$	$51.37_{\pm 7.36}$
Ours	**100**	$\mathbf{85.09_{\pm 3.88}}$	$47.63_{\pm 0.37}$	$268.62_{\pm 2.43}$	$164.87_{\pm 1.55}$	$7.99_{\pm 0.15}$	$\mathbf{39.84_{\pm 2.57}}$	$\mathbf{20.15_{\pm 1.36}}$

Table 3: Real-world performance of IDEAL across three tasks.

Task	SR	$E_{\text{g-mpbpe}}$	E_{mpbpe}	E_{mpjpe}	E_{mpjve}	E_{mpope}
BT	95	101.71 ± 15.92	82.58 ± 7.53	177.36 ± 7.36	14.35 ± 1.97	89.17 ± 8.98
CP	95	94.76 ± 19.26	92.15 ± 8.73	161.74 ± 5.79	9.84 ± 2.56	217.70 ± 22.28
SL	100	87.81 ± 8.32	61.93 ± 3.34	190.15 ± 3.24	8.67 ± 1.31	53.87 ± 10.53

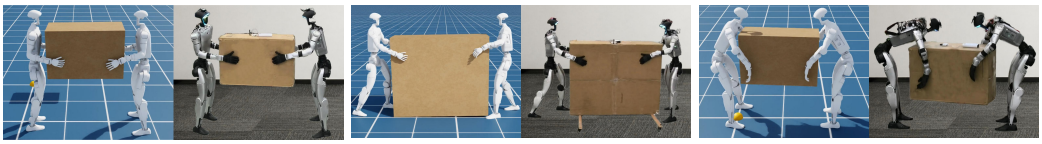
original length, without any explicit inter-robot coordination. These independently learned policies are then directly deployed in the specific DHOI tasks. The results are summarized in Table 2.

The evaluation reveals a clear divergence between robot-level and object-level performance. On robot-motion-related metrics, both methods achieve comparable results, suggesting that independent methods like HDMI can still successfully learn individual motion control and maintain accurate tracking of postures and joint trajectories. However, HDMI performs substantially worse on object-related metrics, yielding higher errors and lower Success Rates across all three tasks. This stark contrast highlights that for shared-object manipulation, independent decision-making is fundamentally insufficient; without explicit coordination, the robots fail to produce consistent object-level effects, leading to poor control over the object's state. By contrast, IDRL overcomes this limitation by explicitly modeling the interaction between the two robots during policy learning. This allows the agents to seamlessly coordinate their behaviors toward a shared objective. Consequently, IDRL significantly improves collaborative object-control performance while maintaining the high individual motion tracking accuracy achieved by independent methods.

Sim-to-Real Transfer and Real-World Deployment To answer Q4, we evaluate the sim-to-real transfer ability of the learned policies by deploying them on two Unitree G1 humanoid robots across three real-world DHOI tasks: SL, BT, and CP. The global positions of the robots and objects are tracked via a UWB system [54], while inter-robot synchronization and communication are managed over a wireless network using ROS [55].

To validate the policy's capability, each task evaluates different physical and coordination constraints. In the SL task, the robots collaboratively approach, establish contact, and lift a suitcase while adopting a forward-bending posture, highlighting motion stability under kinematic changes and dual-robot contact. In the BT task, the robots carry a box from opposite sides, maintaining consistent movement and object stability, demonstrating effective coordination. In the CP task, the robots jointly push a cabinet forward while maintaining a fixed grasp, overcoming continuous ground friction through sustained pushing forces and coordinated whole-body motion.

Overall, real-world deployments reproduce key interaction patterns observed in simulation, with the robots maintaining stable contact and coordinated, task-consistent motions. Across 20 independent real-world trials per task, IDEAL achieves success rates of 95%, 100%, and 95% on BT, SL, and CP, respectively, as summarized in Table 3. The two failed trials were caused by a temporary UWB dropout and a wireless link failure. The side-by-side comparisons in Fig. 5 further illustrate the consistency between simulated and real-world executions across all three tasks. Together, these results demonstrate effective sim-to-real transfer and validate IDEAL's ability to coordinate dual-humanoid collaborative manipulation under real-world conditions.



(a) Box Transport (b) Cabinet Pushing (c) Suitcase Lifting

Figure 5: The sim-to-real comparison of dual-humanoid robots performing coordinated object manipulation across diverse DHOI tasks.

In addition to the above main experiments, we provide further analysis in Appendix D. Specifically, we include (1) a task-level comparison to quantify the individual contributions of PPIMO and IDRL, showing that they improve performance from complementary aspects of motion quality and inter-robot coordination; (2) additional evaluations in weakly coupled scenarios (e.g., Parcel Passing and Cargo Pushing), demonstrating that IDEAL generalizes beyond strictly closed-chain settings and remains effective under transient or loose interaction constraints; and (3) qualitative real-world ablations illustrating the effects of interaction force constraints and inter-robot coordination.

5 Conclusion & Limitations

This paper introduces IDEAL, an imitation learning framework for dual-humanoid collaborative manipulation of rigid objects. Addressing DHOI data scarcity and kinematic tracking limitations, IDEAL combines kinematic motion optimization and interaction dynamics. It extracts contact-aware reference motions from monocular videos via PPIMO, minimizes closed-chain internal forces via DCFA, and incorporates force distributions into policy learning via IDRL. Extensive experiments on Unitree G1 robots show IDEAL consistently outperforms baselines, achieving superior success rates, precise tracking, and physically consistent force coordination.

However, our approach still has several limitations. (1) Reliance on privileged information restricts evaluation to non-articulated rigid objects. Future work will generalize to complex scenarios like articulated and deformable objects, dynamic contacts, and heavy occlusions. (2) Sim-to-real transfer relies on complex external setups, limiting multi-robot scalability. Consequently, future research will explore weaker supervision, tighter perception-control coupling, and scalable multi-agent coordination for unstructured environments.

Acknowledgments

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Nos. XDB16304, XDB1630400, and XDB1630402) and the National Natural Science Foundation of China (Grant No. 62303459).

References

[1] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *ACM Transactions on Graphics (ToG)*, vol. 40, no. 4, pp. 1–20, 2021.

[2] C. Zhang, W. Xiao, T. He, and G. Shi, "Wococo: Learning whole-body humanoid control with sequential contacts," *arXiv preprint arXiv:2406.06005*, 2024.

[3] H. Weng, Y. Li, N. Sobanbabu, Z. Wang, Z. Luo, T. He, D. Ramanan, and G. Shi, "Hdmi: Learning interactive humanoid whole-body control from human videos," *arXiv preprint arXiv:2509.16757*, 2025.

[4] S. Xu, H. Y. Ling, Y.-X. Wang, and L.-Y. Gui, "Intermimic: Towards universal whole-body control for physics-based human-object interactions," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2025, pp. 12 266–12 277.

[5] D. Li, X. Chen, Q. Wu, B. Chen, S. Wu, H. Wu, G. Zhang, L. Li, M. Zhou, D. Xiang *et al.*, "Haic: Humanoid agile object interaction control via dynamics-aware world model," *arXiv preprint arXiv:2602.11758*, 2026.

[6] Z. Shen, H. Pi, Y. Xia, Z. Cen, S. Peng, Z. Hu, H. Bao, R. Hu, and X. Zhou, "World-grounded human motion recovery via gravity-view coordinates," in *SIGGRAPH Asia 2024 Conference Papers*, 2024, pp. 1–11.

[7] S. Shin, J. Kim, E. Halilaj, and M. J. Black, "Wham: Reconstructing world-grounded humans with accurate 3d motion," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2024, pp. 2070–2080.

[8] Q. Liao, T. E. Truong, X. Huang, Y. Gao, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," *Science Robotics*, vol. 11, no. 117, p. eadx8924, 2026.

[9] M. F. Karim, S. Bollimuntha, M. S. Hashmi, A. Das, G. Singh, S. Sridhar, A. K. Singh, N. Govindan, and K. M. Krishna, "Da-vil: Adaptive dual-arm manipulation with reinforcement learning and variable impedance control," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 11 896–11 903.

[10] J. P. Araujo, Y. Ze, P. Xu, J. Wu, and C. K. Liu, "Retargeting matters: General motion retargeting for humanoid motion tracking," in *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2026, pp. 6097–6104.

[11] T. Ayral, S. Aloui, and M. Grossard, "Reactive slip control in multifingered grasping: Hybrid tactile sensing and internal-force optimization," *arXiv preprint arXiv:2602.16127*, 2026.

[12] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 851–866.

[13] M. Kocabas, N. Athanasiou, and M. J. Black, "Vibe: Video inference for human body pose and shape estimation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 5252–5262.

[14] M. Gleicher, "Retargetting motion to new characters," in *Proceedings of SIGGRAPH*, 1998, pp. 33–42.

[15] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 8944–8951.

[16] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi, "Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning," *arXiv preprint arXiv:2406.08858*, 2024.

[17] J. Li, Y. Zhu, Y. Xie, Z. Jiang, M. Seo, G. Pavlakos, and Y. Zhu, "Okami: Teaching humanoid robots manipulation skills through single video imitation," *arXiv preprint arXiv:2410.11792*, 2024.

[18] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions On Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[19] X. B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, "Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters," *ACM Transactions On Graphics (TOG)*, vol. 41, no. 4, pp. 1–17, 2022.

[20] C. Tessler, Y. Kasten, Y. Guo, S. Mannor, G. Chechik, and X. B. Peng, "Calm: Conditional adversarial latent models for directable virtual characters," in *ACM SIGGRAPH 2023 conference proceedings*, 2023, pp. 1–9.

[21] J. Ren, M. Zhang, C. Yu, X. Ma, L. Pan, and Z. Liu, "Insactor: Instruction-driven physics-based characters," *Advances in Neural Information Processing Systems*, vol. 36, pp. 59 911–59 923, 2023.

[22] C. Tessler, Y. Guo, O. Nabati, G. Chechik, and X. B. Peng, "Maskedmimic: Unified physics-based character control through masked motion inpainting," *ACM Transactions On Graphics (TOG)*, vol. 43, no. 6, pp. 1–21, 2024.

[23] Y. Yuan and K. Kitani, "Residual force control for agile human behavior imitation and extended motion synthesis," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 763–21 774, 2020.

[24] Z. Luo, J. Cao, J. Merel, A. Winkler, J. Huang, K. Kitani, and W. Xu, "Universal humanoid motion representations for physics-based control," in *International Conference on Learning Representations*, vol. 2024, 2024, pp. 56 766–56 782.

[25] Z. Luo, J. Cao, K. Kitani, W. Xu *et al.*, "Perpetual humanoid control for real-time simulated avatars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 895–10 904.

[26] J. Shi, X. Liu, D. Wang, S. Schwertfeger, C. Zhang, F. Sun, C. Bai, X. Li *et al.*, "Adversarial locomotion and motion imitation for humanoid policy learning," *Advances in Neural Information Processing Systems*, vol. 38, pp. 73 918–73 949, 2026.

[27] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," *arXiv preprint arXiv:2402.16796*, 2024.

[28] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang *et al.*, "Hover: Versatile neural whole-body controller for humanoid robots," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 9989–9996.

[29] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, vol. 9, no. 89, p. eadi9579, 2024.

[30] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan *et al.*, "Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills," *arXiv preprint arXiv:2502.01143*, 2025.

[31] C. Zhang, J. Bai, Z. Chen, J. Gao, and H. Qiao, "Hierarchical control for robust standing stability and fall recovery of task-performing humanoid robots," *IEEE Transactions on Automation Science and Engineering*, 2025.

[32] C. Zhang, J. Gao, Z. Chen, S. Zhong, and H. Qiao, "Fall analysis and prediction for humanoids," *Robotics and Autonomous Systems*, vol. 190, p. 104995, 2025.

[33] Y. Wang, J. Lin, A. Zeng, Z. Luo, J. Zhang, and L. Zhang, "Physhoi: Physics-based imitation of dynamic human-object interaction," *arXiv preprint arXiv:2312.04393*, 2023.

[34] Z. Luo, J. Cao, S. Christen, A. Winkler, K. Kitani, and W. Xu, "Omnigrasp: Grasping diverse objects with simulated humanoids," *Advances in Neural Information Processing Systems*, vol. 37, pp. 2161–2184, 2024.

[35] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in neural information processing systems*, vol. 35, pp. 24 611–24 624, 2022.

[36] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine learning proceedings 1994*. Elsevier, 1994, pp. 157–163.

[37] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.

[38] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[39] Y. Zhang, D. Gopinath, Y. Ye, J. Hodgins, G. Turk, and J. Won, "Simulation and retargeting of complex multi-character interactions," in *ACM SIGGRAPH 2023 conference proceedings*, 2023, pp. 1–11.

[40] M. Younes, E. Kijak, R. Kulpa, S. Malinowski, and F. Multon, "Maaip: Multi-agent adversarial interaction priors for imitation from fighting demonstrations for physics-based characters," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 6, no. 3, pp. 1–20, 2023.

[41] Z. Liu, J. Ge, M. Xiong, J. Gu, B. Tang, W. Jing, and S. Chen, "It takes two: Learning interactive whole-body control between humanoid robots," *arXiv preprint arXiv:2510.10206*, 2025.

[42] H. Chen, W. Zhang, P. Li, S. Ma, K. Ma, Y. Jin, Z. Xu, X. Wang, Y. Zheng, Z. Wang *et al.*, "Rhythm: Learning interactive whole-body control for dual humanoids," *arXiv preprint arXiv:2603.02856*, 2026.

[43] K. Yin, Z. Cao, W. Dong, W. Zeng, T. Zhang, Q. Zhang, J. Wang, J. Pang, M. Zhou, and W. Zhang, "Robostriker: Hierarchical decision-making for autonomous humanoid boxing," *arXiv preprint arXiv:2601.22517*, 2026.

[44] J. Gao, Z. Wang, Z. Xiao, J. Wang, T. Wang, J. Cao, X. Hu, S. Liu, J. Dai, and J. Pang, "Coohoi: Learning cooperative human-object interaction with manipulated object dynamics," *Advances in Neural Information Processing Systems*, vol. 37, pp. 79 741–79 763, 2024.

[45] W. Yao, H. Ma, H. Zhang, L. Xing, Z. Yang, Y. Guo, Y. Sun, and Y. Liu, "Synagent: Generalizable cooperative humanoid manipulation via solo-to-cooperative agent synergy," *arXiv preprint arXiv:2604.18557*, 2026.

[46] Y. Du, Y. Li, B. Jia, Y. Lin, P. Zhou, W. Liang, Y. Yang, and S. Huang, "Cola: Learning human-humanoid coordination for collaborative object carrying," in *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2026, pp. 18 578–18 585.

[47] G. C. R. Bethala, H. Huang, N. Pudasaini, A. M. Ali, S. Yuan, C. Wen, A. Tzes, and Y. Fang, "H 2-compact: Human-humanoid co-manipulation via adaptive contact trajectory policies," in *2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2025, pp. 1004–1011.

[48] N. Kulkarni, D. Rempe, K. Genova, A. Kundu, J. Johnson, D. Fouhey, and L. Guibas, "Nifty: Neural object interaction fields for guided human motion synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 947–957.

[49] C. K. Verginis, D. Zelazo, and D. V. Dimarogonas, "Cooperative manipulation via internal force regulation: A rigidity theory perspective," *IEEE Transactions on Control of Network Systems*, vol. 10, no. 3, pp. 1222–1233, 2022.

[50] R. Di Leva, H. Gattringer, A. Mueller, and M. Carricato, "Force and time-optimal trajectory planning for dual-arm unilateral cooperative grasping," *Mechanism and Machine Theory*, vol. 201, p. 105729, 2024.

[51] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," *arXiv preprint arXiv:1710.06542*, 2017.

[52] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[53] NVIDIA, "Isaac Sim." [Online]. Available: https://github.com/isaac-sim/IsaacSim

[54] W. Zhao, J. Panerati, and A. P. Schoellig, "Learning-based bias correction for time difference of arrival ultra-wideband localization of resource-constrained mobile robots," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3639–3646, 2021.

[55] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

A Asset Details

A.1 Interaction Objects

Box / Suitcase The box and suitcase are modeled as standard cuboids with a mass of $1.3\,\mathrm{kg}$ and a size of $0.76\,\mathrm{m} \times 0.33\,\mathrm{m} \times 0.50\,\mathrm{m}$. They are used in the Suitcase Lifting (SL) and Box Transport (BT) tasks to evaluate dual-humanoid cooperative manipulation under sustained contact.

Movable Cabinet The movable cabinet is modeled as a wheeled cuboid with a mass of $3\,\mathrm{kg}$ and a size of $1.20\,\mathrm{m} \times 0.40\,\mathrm{m} \times 1.00\,\mathrm{m}$. To reduce collision-modeling cost, the wheels are not explicitly simulated; instead, a mean bottom sliding friction coefficient of 0.2 is used to approximate wheeled motion. This object is used in the Cabinet Pushing (CP) task to evaluate cooperative manipulation under ground-contact and friction constraints.

A.2 Task Settings

Box Transport (BT) Two robots hold the short sides of the box, establish stable contact, transport it laterally for 7 steps, and then stop while maintaining its pose. This task evaluates coordinated locomotion and object stabilization during sustained carrying.

Cabinet Pushing (CP) Two robots maintain side contact with the cabinet, push it laterally for 3 steps, and then forward for another 3 steps. Since the cabinet remains on the ground, this task evaluates cooperative pushing under non-lifting contact, ground friction, and direction changes.

Suitcase Lifting (SL) Two robots bend down to grasp the side surfaces of the suitcase, lift it while standing up, keep it stable briefly, place it back on the ground, and then return to an upright standing posture. This task evaluates cooperative lifting and placing under large posture changes and dual-humanoid contact constraints.

A.3 Data Acquisition

Human Demonstrations Human demonstrations are recorded with a monocular camera. In each trial, two volunteers perform one cooperative manipulation task following the predefined procedure, providing motion and interaction references for SL, BT, and CP.

Object Tracking UWB tags are attached to each object to record its position during demonstrations. Due to hardware limitations, object orientation is not reliably measured; the scene is therefore designed to minimize object rotation, and the orientation is fixed as the default quaternion $[1, 0, 0, 0]$. Contact frames and robot–object contact positions are manually annotated after data collection.

Scene Setup All demonstrations are conducted in a flat indoor environment with sufficient space. Initial object placements and volunteer standing positions are kept as consistent as possible across trials to ensure comparability.

B Algorithm Details

B.1 Details of PPIMO

Signed Distance Field Definition. To formulate the contact-aware interaction energy, we represent the manipulated object using a signed distance field (SDF). Let the object occupy a closed region $\Omega \subset \mathbb{R}^3$ with boundary surface $\partial\Omega$. For any spatial point $\mathbf{x} \in \mathbb{R}^3$, the signed distance function is defined as

$$\mathrm{SDF}(\mathbf{x}, \partial\Omega) = \begin{cases} \min\limits_{\mathbf{y}\in\partial\Omega} \|\mathbf{x} - \mathbf{y}\|_2, & \mathbf{x} \notin \Omega, \\ -\min\limits_{\mathbf{y}\in\partial\Omega} \|\mathbf{x} - \mathbf{y}\|_2, & \mathbf{x} \in \Omega. \end{cases} \tag{9}$$

Therefore, $\mathrm{SDF}(\mathbf{x}, \partial\Omega) > 0$ indicates that $\mathbf{x}$ lies outside the object, $\mathrm{SDF}(\mathbf{x}, \partial\Omega) < 0$ indicates penetration into the object interior, and $\mathrm{SDF}(\mathbf{x}, \partial\Omega) = 0$ corresponds to points exactly on the object surface.

In practice, the SDF is precomputed from the object mesh in the object coordinate frame and queried during optimization after transforming the end-effector position into the object frame. Specifically, given the object pose $(\mathbf{R}_o, \mathbf{t}_o)$ and the end-effector position obtained by forward kinematics $K(\theta)$ in world coordinates, the queried point is

$$\mathbf{x}_o = \mathbf{R}_o^\top \big(K(\theta) - \mathbf{t}_o\big), \tag{10}$$

and the interaction energy evaluates $\mathrm{SDF}(\mathbf{x}_o, \partial\Omega)$ accordingly.

Based on this representation, the contact-aware interaction energy is formulated as

$$E_C(\theta) = \begin{cases} \big(\max\big(0, \epsilon - \mathrm{SDF}(K(\theta), \partial\Omega)\big)\big)^2, & \text{if } \mathbb{I}_{\mathcal{C}} = 0, \\ \big(\max\big(0, |\mathrm{SDF}(K(\theta), \partial\Omega)| - \delta\big)\big)^2, & \text{if } \mathbb{I}_{\mathcal{C}} = 1, \end{cases} \tag{11}$$

where, in our approach, we empirically set $\epsilon = 5\,\mathrm{mm}$ as a safety margin to prevent penetration in non-contact states, and $\delta = 2\,\mathrm{mm}$ as the allowable contact thickness to maintain stable physical interaction during contact.

B.2 Details of DCFA

Newton–Euler Equations The desired spatial wrench $\mathbf{W}_d$ is derived from the Newton-Euler equations describing the rigid-body dynamics of the manipulated object. Specifically, the force and moment balance at the CoM must satisfy:

$$\sum_{i=1}^{4} \mathbf{f}_i + \mathbf{F} = m\ddot{\mathbf{x}}_d \tag{12}$$

$$\sum_{i=1}^{4} (\mathbf{p}_i - \mathbf{p}_{\text{CoM}}) \times \mathbf{f}_i = \mathbf{I}\dot{\boldsymbol{\omega}}_d + \boldsymbol{\omega} \times (\mathbf{I}\boldsymbol{\omega}) \tag{13}$$

where $\mathbf{F} = [F_x \quad F_y \quad F_z]^\top$ denotes the resultant external force exerted on the object by sources other than the dual-robot system. It is assumed that environmental external moments (e.g., from ground contact) are negligible, such that the dominant moment contribution is only generated by the dual-robot manipulation system.

These equations can be compactly expressed as a spatial wrench balance:

$$\mathbf{W}_d = \begin{bmatrix} m\ddot{\mathbf{x}}_d - \mathbf{F} \\ \mathbf{I}\dot{\boldsymbol{\omega}}_d + \boldsymbol{\omega} \times (\mathbf{I}\boldsymbol{\omega}) \end{bmatrix} \tag{14}$$

Grasp Matrix $\mathbf{G}$ The grasp matrix $\mathbf{G}$ maps the stacked contact forces to the resultant spatial wrench at the object center of mass (CoM). For four contact points, it takes the explicit block form:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 \\ \hat{\mathbf{r}}_1 & \hat{\mathbf{r}}_2 & \hat{\mathbf{r}}_3 & \hat{\mathbf{r}}_4 \end{bmatrix} \in \mathbb{R}^{6\times 12} \tag{15}$$

where $\mathbf{r}_i = \mathbf{p}_i - \mathbf{p}_{\text{CoM}}$ is the position vector from the CoM to the i-th contact point.

The skew-symmetric matrix $\hat{\mathbf{r}}_i$ is defined as:

$$\hat{\mathbf{r}}_i = \begin{bmatrix} 0 & -r_{i,z} & r_{i,y} \\ r_{i,z} & 0 & -r_{i,x} \\ -r_{i,y} & r_{i,x} & 0 \end{bmatrix} \tag{16}$$

such that $\hat{\mathbf{r}}_i \mathbf{f}_i = \mathbf{r}_i \times \mathbf{f}_i$.

Therefore, the grasp mapping $\mathbf{Gf}$ explicitly yields:

$$\mathbf{Gf} = \sum_{i=1}^{4} \begin{bmatrix} \mathbf{f}_i \\ \mathbf{r}_i \times \mathbf{f}_i \end{bmatrix} = \mathbf{W}_d \tag{17}$$

Friction Constraint Matrix A To ensure non-slip contact, each contact force must lie within its Coulomb friction cone. For computational efficiency, we adopt a linearized pyramidal approximation of the friction cone. Let $\mathbf{n}_i$ denote the inward surface normal at contact i, and $\mathbf{t}_{i,1}, \mathbf{t}_{i,2}$ be two orthogonal tangential directions. The contact force can be decomposed as:

$$\mathbf{f}_i = f_{i,n}\mathbf{n}_i + f_{i,t_1}\mathbf{t}_{i,1} + f_{i,t_2}\mathbf{t}_{i,2} \tag{18}$$

The Coulomb friction constraint is:

$$\sqrt{f_{i,t_1}^2 + f_{i,t_2}^2} \leq \mu f_{i,n}, \quad f_{i,n} \geq 0 \tag{19}$$

This is approximated by a linear pyramid:

$$\begin{aligned} \pm f_{i,t_1} &\leq \mu f_{i,n} \\ \pm f_{i,t_2} &\leq \mu f_{i,n} \end{aligned} \tag{20}$$

These inequalities can be written in matrix form:

$$\mathbf{A}_i \mathbf{f}_i \leq \mathbf{0} \tag{21}$$

where the constraint matrix explicitly includes the unilateral contact condition:

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{t}_{i,1}^\top - \mu \mathbf{n}_i^\top \\ -\mathbf{t}_{i,1}^\top - \mu \mathbf{n}_i^\top \\ \mathbf{t}_{i,2}^\top - \mu \mathbf{n}_i^\top \\ -\mathbf{t}_{i,2}^\top - \mu \mathbf{n}_i^\top \\ -\mathbf{n}_i^\top \end{bmatrix} \in \mathbb{R}^{5\times 3} \tag{22}$$

Stacking all contacts yields the global friction constraint:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \end{bmatrix} \in \mathbb{R}^{20\times 12} \tag{23}$$

such that:

$$\mathbf{A}\mathbf{f} \leq \mathbf{0} \tag{24}$$

Normal Force Extraction Matrix $\mathbf{C}_i$ To enforce bounds on the normal component of each contact force, we define a linear extraction operator $\mathbf{C}_i$ that projects $\mathbf{f}$ onto the normal direction at contact i.

For a single contact force $\mathbf{f}_i$, the normal component is:

$$f_{i,n} = \mathbf{n}_i^\top \mathbf{f}_i \tag{25}$$

Thus, the extraction matrix $\mathbf{C}_i \in \mathbb{R}^{1\times 12}$ is defined as:

$$\mathbf{C}_i = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{n}_i^\top & \cdots & \mathbf{0} \end{bmatrix} \tag{26}$$

where $\mathbf{n}_i^\top$ occupies the columns corresponding to $\mathbf{f}_i$.

For a system with four contact points, the complete matrix is given by:

$$\mathbf{C} = \begin{bmatrix} \mathbf{n}_1^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{n}_2^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{n}_3^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{n}_4^\top \end{bmatrix} \tag{27}$$

These matrices allow the normal force bounds to be expressed compactly as:

$$f_{i,\min} \leq \mathbf{C}_i \mathbf{f} \leq f_{i,\max} \tag{28}$$

Table 4: Details of the Observation Space.

Observation Term	Noise (σ)	Available to	Description
O_r (Reference)			
Ref. Motion Phase	-	Teacher / Student	Phase variable ϕ of the reference motion.
Ref. Joint Positions	-	Teacher / Student	Future reference joint positions.
Ref. Body Pos (Local)	-	Teacher / Student	Future reference body positions in robot local frame.
O_p (Ego State)			
Base Angular Velocity	0.05	Teacher / Student	History of base angular velocity.
Projected Gravity	0.05	Teacher / Student	Gravity vector in base frame.
Joint Positions	0.015	Teacher / Student	History of joint positions.
Previous Actions	-	Teacher / Student	Joint target history (last 3 steps).
O_c (Collaborator)			
Collaborator Joint Positions	0.015	Teacher / Student	Other agent's joint states.
Collaborator State	0.1	Teacher / Student	Other agent's pose and velocity in local frame.
Collaborator Orientation	-	Teacher Only	Orientation of the other agent in the local frame.
O_o^p (Object State)			
Object State	0.1	Teacher / Student	Object pose and velocity in local frame.
Object Orientation	0.0	Teacher Only	Orientation of manipulated object in local frame.
O_o^m (Contact)			
Ref. Contact Pos	0.0	Teacher Only	Reference contact positions on object in local frame.
Contact State	0.01	Teacher / Student	Reference contact indicators between agents and object.
O_o^f (Interaction Force)			
Interaction Force Ground Truth	0.0	Teacher Only	Ground-truth interaction forces between agents and object.
Interaction Force Embedding	-	Student Only	Predicted interaction force via IDE encoder.
O_z (Privileged)			
Clean Proprioception	0.0	Teacher Only	Noise-free full proprioceptive history.
Root Linear Velocity	0.0	Teacher Only	Base linear velocity.
Body Velocity	0.0	Teacher Only	Key body velocities.
Body Height	0.0	Teacher Only	Heights relative to ground.
Ref. Diff (Local)	0.0	Teacher Only	Difference to reference states.
$\hat{O}_z$ (Estimated Privileged)			
Estimated Privileged State	-	Student Only	Estimated counterpart of O_z, predicted by the MLP adapter.

B.3 Details of IDRL Framework

B.3.1 Observation Space Design

In our asymmetric observation setting, the teacher policy π_T and student policy π_S share O_r and O_p, but differ in collaborator, object-interaction, and privileged observations, whose definitions are provided in Table 4. During training, the teacher uses $O_T = \{O_r, O_p, O_c^T, O_o^T, O_z\}$, where O_z contains privileged information including clean proprioception, root velocity, body velocity, body height, and local reference-state differences. The student uses $O_S = \{O_r, O_p, O_c^S, O_o^S, \hat{O}_z\}$, where $\hat{O}_z$ is predicted by an MLP adapter $g_\psi(\cdot)$ from observations and supervised with an MSE loss.

For collaborator observations, both policies observe joint states, position, and velocity in the robot local frame, while collaborator orientation is available only to the teacher, matching deployment constraints where UWB localization provides position but not reliable orientation. For object observations, we define $O_o = \{O_o^p, O_o^m, O_o^f\}$. Both policies observe object position and velocity, while object orientation is teacher-only. For contact information, the teacher uses reference contact positions and contact states, whereas the student uses only contact states. For interaction forces, the teacher uses ground-truth forces, while the student uses force embeddings predicted by the Interaction Dynamics Encoder.

B.3.2 Reward Design

The training objective is to maximize the cumulative reward R, summed over robots $k \in \{\mathrm{A}, \mathrm{B}\}$, timesteps t, and reward terms i, where $r_{i,t}^{(k)}$ denotes each reward term and w_i its weight:

$$R = \sum_{k \in \{\mathrm{A},\mathrm{B}\}} \sum_t \sum_i w_i \, r_{i,t}^{(k)} \tag{29}$$

Table 5: Reward functions for the dual-humanoid-object interaction task.

Term	Expression	Weight	Description
Motion Tracking Reward			
Joint Position	$\exp(-\|q - q^{\text{ref}}\|^2/\sigma_q)$	0.5	Track reference joint angles.
Joint Velocity	$\exp(-\|\dot{q} - \dot{q}^{\text{ref}}\|^2/\sigma_{\dot{q}})$	0.5	Track reference joint velocities.
Upper Body Pos.	$\exp(-\|p_{\text{up}} - p_{\text{up}}^{\text{ref}}\|^2/\sigma_p)$	0.5	Track reference upper body positions.
Upper Body Ori.	$\exp(-\|\theta_{\text{up}} \ominus \theta_{\text{up}}^{\text{ref}}\|^2/\sigma_\theta)$	0.5	Track reference upper body orientations.
Lower Body Pos.	$\exp(-\|p_{\text{low}} - p_{\text{low}}^{\text{ref}}\|^2/\sigma_p)$	0.5	Track reference lower body positions.
Lower Body Ori.	$\exp(-\|\theta_{\text{low}} \ominus \theta_{\text{low}}^{\text{ref}}\|^2/\sigma_\theta)$	0.5	Track reference lower body orientations.
Root Position	$\exp(-\|p_{\text{root}} - p_{\text{root}}^{\text{ref}}\|^2/\sigma_p)$	0.5	Track reference root position.
Root Orientation	$\exp(-\|\theta_{\text{root}} \ominus \theta_{\text{root}}^{\text{ref}}\|^2/\sigma_\theta)$	0.5	Track reference root orientation.
Body Lin/Ang Vel.	$\exp(-\|v - v^{\text{ref}}\|^2/\sigma_v)$	0.5	Track reference body velocities.
Interaction Force Tracking Reward			
Interaction Force	$\exp(-\sum_{i\in C} \|f_i - f_i^*\|^2/\sigma_f)$	1.0	Track reference interaction force.
Object Interaction Reward			
Object Position	$\exp(-\|p_{\text{obj}} - p_{\text{obj}}^{\text{ref}}\|^2/\sigma_p)$	1.0	Track reference object positions.
Object Orientation	$\exp(-\|\theta_{\text{obj}} \ominus \theta_{\text{obj}}^{\text{ref}}\|^2/\sigma_\theta)$	1.0	Track reference object orientations.
Object Contact	$\frac{1}{\|\mathcal{E}\|}\sum_{e\in\mathcal{E}} \mathbb{I}_c \cdot r_{\text{pos}}^e \cdot r_{\text{force}}^e$	1.0	Track contact masks of objects.
Foot Constraints			
Feet Air Time	$\exp(\text{clip}(t_{\text{air}} - t_{\text{thre}})/\sigma) \cdot \mathbb{I}_{\text{step}}$	5.0	Reward foot air time.
Feet Slip	$-\|v_{\text{foot}}^{xy}\| \cdot \mathbb{I}_{\text{ground}}$	0.5	Penalize foot slip velocity.
Impact Force	$-\|F_{\text{impact}}\|^2$	1.0	Penalize large contact forces.
Regularization			
Action Rate	$-\|a_t - a_{t-1}\|^2$	0.1	Penalize rapid control changes.
Joint Velocity L2	$-\|\dot{q}\|^2$	5e−4	Penalize large joint velocities.
Joint Limits	$-\sum \text{clip}(q - q_{\text{limit}})$	10.0	Penalize joint limits.
Torque Limits	$-\sum \text{clip}(\tau - \tau_{\text{limit}})$	0.01	Penalize large joint torques.
Survival	1.0	1.0	Stay alive.

The reward function is organized into five components: Motion Tracking, Interaction Force Tracking, Object Interaction, Foot Constraints, and Regularization. The sensitivity scaling factor for the force tracking reward is empirically set to $\sigma_f = 1.5$ in all experiments. Details are summarized in Table 5.

B.3.3 Domain Randomization

To ensure that the learned policy can be transferred to the physical robots, we apply extensive domain randomization during training. This approach models physical uncertainties as probability distributions, thereby reducing the discrepancy between simulation and reality.

We randomize a comprehensive set of parameters, covering the robot's inertial properties and actuation characteristics, as well as the physical properties of the manipulated objects. The detailed settings are provided in Table 6.

B.3.4 Action Space and Low-level Control

At each timestep, the policy predicts a concatenated action vector $a_t \in \mathbb{R}^{46}$ for the dual-robot system. This action is partitioned into two 23-dimensional residual joint-position commands, $a_t^{(1)}$ and $a_t^{(2)}$, each corresponding to one of the two robots. For each robot $i \in \{1, 2\}$, the target joint position is calculated as $q_{target}^{(i)} = q_t^{(i)} + \Delta q^{(i)}$.

For both robots, the 23-dimensional action subspace covers the whole-body actuated joints, excluding the six wrist degrees of freedom (three per hand), as these are not required for the specific interaction tasks. The resulting target positions are then tracked independently by low-level proportional-derivative (PD) controllers, which generate joint torques for each robot i as:

$$\tau^{(i)} = k_p(q_{target}^{(i)} - q_t^{(i)}) - k_d\dot{q}_t^{(i)}, \tag{30}$$

Table 6: Domain randomization parameters for robots and objects.

Category	Parameter	Range / Distribution
Robot Dynamics		
Properties	Link Mass Scale	$\mathcal{U}(0.9, 1.1) \times$ default
	Center of Mass Offset	$\mathcal{U}(-0.02, 0.02)$ m
	Static Friction	$\mathcal{U}(0.3, 1.6)$
	Dynamic Friction	$\mathcal{U}(0.3, 1.2)$
Actuation	Joint Position Offset	$\mathcal{U}(-0.01, 0.01)$ rad
	Motor Stiffness Scale	$\mathcal{U}(0.9, 1.1)$
	Motor Damping Scale	$\mathcal{U}(0.9, 1.1)$
	Action Delay	$\mathcal{U}[40, 120]$ ms
Object		
Box/Suitcase	Mass	$\mathcal{U}(1.0, 2.5)$ kg
	Scale	$\mathcal{U}(0.9, 1.1)$
	Dynamic Friction	$\mathcal{U}(0.3, 0.8)$
	Static-to-Dynamic Ratio	$\mathcal{U}(1.0, 2.0)$
	Restitution	$\mathcal{U}(0.05, 0.3)$
Cabinet	Mass	$\mathcal{U}(3.0, 5.0)$ kg
	Scale	$\mathcal{U}(0.9, 1.1)$
	Dynamic Friction	$\mathcal{U}(0.1, 0.3)$
	Static-to-Dynamic Ratio	$\mathcal{U}(1.0, 2.0)$
	Restitution	$\mathcal{U}(0.0, 0.2)$

Table 7: Joint stiffness and damping parameters for different body parts

Joint Name	k_p	k_d
Legs		
Left/Right Hip Pitch/Yaw	40.18	2.558
Left/Right Hip Roll	99.10	6.309
Left/Right Knee	99.10	6.309
Left/Right Ankle Pitch/Roll	28.50	1.814
Waist		
Waist Yaw	40.18	2.558
Waist Roll/Pitch	28.50	1.814
Arms		
Left/Right Shoulder Pitch/Roll/Yaw	14.25	0.9072
Left/Right Elbow Pitch/Roll	14.25	0.9072
Left/Right Wrist Roll/Pitch	16.78	1.068

Table 8: Hyperparameters related to MAPPO and network architecture

Hyperparameter	**Value**
Optimizer	Adam
Number of Environments	4096
Rollout Steps	32
Mini-batches	8
Learning Epochs	3
Discount Factor (γ)	0.99
GAE Parameter (λ)	0.95
Clip Parameter (ϵ)	0.2
Entropy Coefficient	0.001
Max Gradient Norm	1.0
Desired KL	0.01
Learning Rate	3×10^{-4}
Initial Noise Std	1.0
Loss coefficients	
Policy Loss Coefficient	1.0
Value Loss Coefficient	1.0
Privileged Loss Coefficient	1.0
Distillation Loss Coefficient	1.0
IDE KL Coefficient	1×10^{-3}
Network architecture	
Number of Teacher Actors	2
Number of Student Actors	2
Teacher Actor MLP Size	[512, 256, 256]
Student Actor MLP Size	[512, 256, 256]
Shared Critic MLP Size	[512, 256, 128]
Adapter MLP Size	[256, 256]
IDE Encoder MLP Size	[256, 128]
IDE Decoder MLP Size	[128, 256]
IDE Latent Dimension	32
Activation Function	ELU

where k_p and k_d denote the proportional stiffness and damping coefficients, respectively. This residual position-control formulation ensures stable and compliant physical interaction between the two humanoid robots. The PD gains adopted for the G1 humanoid robots are summarized in Table 7.

B.3.5 MAPPO Hyperparameters and Network Architecture

We adopt the MAPPO algorithm to optimize multi-agent policies and employ the Adam optimizer for parameter updates. During training, generalized advantage estimation (GAE) is introduced to reduce the variance of advantage estimation, while a clipped surrogate objective is used to improve the stability of policy updates. In addition, an entropy regularization term is incorporated to encourage policy exploration.

In terms of network architecture, we employ an asymmetric multi-agent Actor–Critic framework under a teacher–student paradigm. Both the teacher and the student consist of two Actor networks, corresponding to two agents. All Actors share the same network architecture but do not share parameters. Each Actor is implemented as a multilayer perceptron (MLP) with identical dimensions and ELU activation functions.

A centralized Critic network is shared across all agents to estimate the joint value function using global state information or privileged information. This design provides stable value feedback for cooperative multi-agent policy optimization during centralized training.

The overall optimization objective is composed of a weighted combination of policy loss, value loss, entropy regularization, and auxiliary loss. The detailed hyperparameter settings and network dimensions are reported in Table 8.

C Experimental Details

C.1 Experimental Setup

Computing Platform All models are trained on a workstation equipped with an Intel(R) Xeon(R) Platinum 8360Y CPU, 128 GB of memory, and a single NVIDIA RTX 4090 GPU, running Ubuntu 22.04. The training time for a single model is approximately 10 hours.

Deployment on Physical Robots For real-world deployment, each robot is equipped with an onboard control board. The global positions of the robots and objects are obtained using an UWB localization system. Inter-robot communication is established via WiFi, while the global positions and joint states are shared through the ROS framework. The onboard controller outputs joint torques at $500\,\mathrm{Hz}$, whereas both communication and policy inference are executed at $50\,\mathrm{Hz}$.

C.2 Evaluation Metrics

Success Rate. Success Rate (SR) evaluates the proportion of successful rollouts in the test environment. A rollout is regarded as successful only when the agent remains balanced and tracks the reference motion with acceptable accuracy. Cases in which the agent merely avoids falling but fails to follow the reference trajectory are counted as failures, since they do not satisfy the intended motion objective.

Global Mean Per-Body-Part Position Error. This metric measures the average global position deviation between each body part and its reference trajectory:

$$E_{\text{g-mpbpe}} = \mathbb{E}\left[\left\|p_t - p_t^{\text{ref}}\right\|_2\right]. \tag{31}$$

Root-Relative Mean Per-Body-Part Position Error. This metric evaluates the body-part position error after removing the global root translation, thereby focusing on the relative body configuration:

$$E_{\text{mpbpe}} = \mathbb{E}\left[\left\|(p_t - p_{\text{root},t}) - \left(p_t^{\text{ref}} - p_{\text{root},t}^{\text{ref}}\right)\right\|_2\right]. \tag{32}$$

Mean Per-Body-Part Orientation Error. This metric measures the average angular deviation between the body-part orientation and the reference orientation:

$$E_{\text{mpboe}} = \mathbb{E}\left[2\arccos\left(\left|\left\langle q_t, q_t^{\text{ref}}\right\rangle\right|\right)\right]. \tag{33}$$

Mean Per-Joint Position Error. This metric evaluates the average joint-angle deviation from the reference motion:

$$E_{\text{mpjpe}} = \mathbb{E}\left[\left\|\theta_t - \theta_t^{\text{ref}}\right\|_1\right]. \tag{34}$$

Mean Per-Joint Velocity Error. This metric measures the average deviation in joint velocity, where the velocity is approximated by the frame-to-frame joint-position difference:

$$E_{\text{mpjve}} = \mathbb{E}\left[\left\|\Delta\theta_t - \Delta\theta_t^{\text{ref}}\right\|_1\right], \quad \Delta\theta_t = \theta_t - \theta_{t-1}. \tag{35}$$

Mean Per-Object Position Error. This metric measures the average position deviation between each manipulated object and its reference trajectory:

$$E_{\text{mpope}} = \mathbb{E}\left[\left\|p_t^{\text{obj}} - p_t^{\text{obj,ref}}\right\|_2\right]. \tag{36}$$

Mean Per-Object Orientation Error. This metric evaluates the average angular deviation between the object orientation and its reference orientation:

$$E_{\text{mpooe}} = \mathbb{E}\left[2\arccos\left(\left|\left\langle q_t^{\text{obj}}, q_t^{\text{obj,ref}}\right\rangle\right|\right)\right]. \tag{37}$$

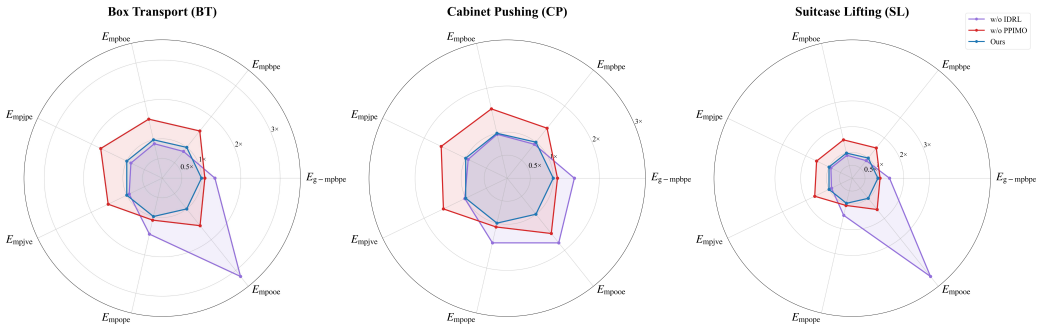


Figure 6: Task-level performance comparison across three tasks under different module settings. Results are normalized with respect to the full method.

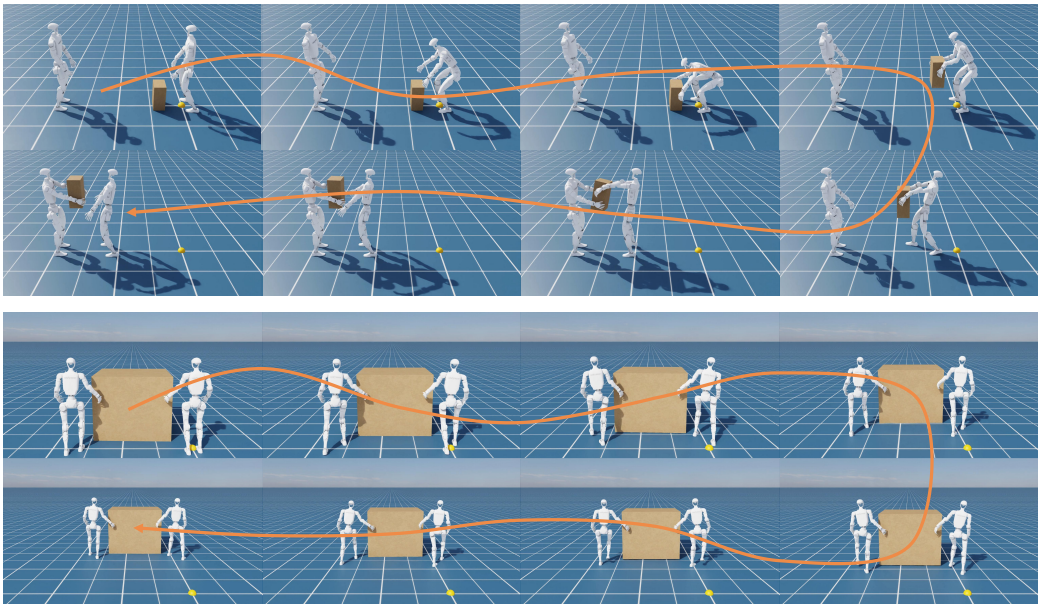


Figure 7: Key motion sequence snapshots of IDEAL in weakly coupled dual-humanoid collaborative manipulation tasks. **Upper Panel** illustrates the Parcel Passing task, where the system forms a closed-chain configuration only during the brief handover phase, while remaining largely unconstrained before and after the contact. **Lower Panel** depicts the Cargo Pushing task, where two robots form unstable contacts via a large cargo, and the system maintains a loosely coupled structure throughout the entire push. Time advances along the indicated arrows.

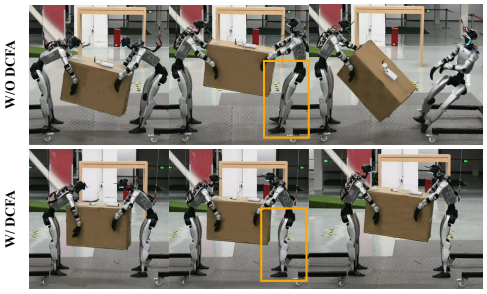


Figure 8: Representative real-world SL executions with and without DCFA-based interaction force constraints.



Figure 9: Representative real-world BT executions with IDRL and with independently trained policies (w/o IDRL).

D Additional Experimental Results

D.1 Task-Level Comparison of Module Contributions

As shown in Fig. 6, we evaluate the contributions of different modules from a task-level perspective, focusing on their impact on task execution and overall performance. Specifically, we plot performance curves for three tasks, comparing three settings: w/o PPIMO, w/o IDRL, and the full method. All results are normalized with respect to the full method to ensure cross-task comparability. The results show that removing PPIMO leads to a clear degradation in robot centric motion metrics, indicating that reference motion quality directly affects motion imitation and tracking. In contrast, removing IDRL results in significant performance drops in object related metrics, suggesting that the lack of coordination undermines stable object-level control. Overall, the two modules influence task performance from complementary perspectives, motion quality and coordinated control.

D.2 Additional Experiments in Weakly Coupled Scenarios

In Section 4, we evaluate IDEAL on three representative dual-humanoid collaborative manipulation tasks designed to cover most real-world collaborative scenarios. In these tasks, both robots firmly grasp a shared rigid object with both hands, forming a closed-chain system with coupling effects of varying intensities and durations.

However, not all DHOI tasks exhibit equally prominent closed-chain dynamics. In certain scenarios, the inter-robot contact relationship via the object is transient or unstable, resulting in weaker closed-chain constraints. Under such conditions, the system more closely resembles a loosely coupled or approximately open-chain structure. For instance, in tasks such as Parcel Passing, where one robot bends down to pick up a small parcel from the ground, walks forward, and hands it to another robot, or Cargo Pushing, where two robots stand on opposite sides of a cargo object and each use one hand to push it forward while walking, the physical interactions between the robots through the manipulated object are typically transient and weakly sustained. In these cases, the coordination of internal forces has a relatively limited impact on overall system stability.

Despite these conditions, IDEAL is not constrained by strict closed-chain assumptions. It serves as a unified framework capable of adapting to varying coupling strengths. Under strongly coupled conditions, the interaction force allocation modeled by DCFA effectively constrains closed-chain internal forces, enabling physically consistent collaborative control. Conversely, under transient contact or weakly coupled conditions, the interaction force constraints adaptively degenerate based on contact signals, the number of contact points, and contact positions, naturally reverting to a kinematics-dominated tracking mode. Leveraging this unified mechanism, IDEAL not only maintains robust performance in strongly coupled scenarios but also naturally generalizes to weakly coupled tasks. To demonstrate this, we further evaluate the performance of IDEAL in weakly coupled scenarios, specifically the Parcel Passing and Cargo Pushing tasks, and provide a series of key motion sequence snapshots, as illustrated in Fig. 7.

D.3 Qualitative Analysis of Force Constraints and Inter-Robot Coordination

To complement the quantitative ablation studies in Section 4, we provide representative real-world execution snapshots that illustrate the effects of interaction force constraints and inter-robot coordination. As shown in Fig. 8, removing the DCFA-based interaction force constraints in SL leads to pronounced backward leaning and object tilting, followed by a loss of stable manipulation. This behavior is consistent with the internal force antagonism analyzed in Fig. 4: kinematic tracking alone does not prevent the two robots from generating redundant, opposing forces through the shared object. In contrast, incorporating the desired force allocation into policy learning helps the robots maintain more stable postures and object support during lifting.

Fig. 9 further illustrates the coordination issue in BT. Here, the setting without IDRL corresponds to the independently trained HDMI policies described in the main text. Although each policy learns individual motion control, the two robots fail to manipulate the shared object synchronously, causing it to tilt and fall. With IDRL, the robots coordinate their motions and maintain stable object transport. Together, these qualitative results support the quantitative findings in the main text: dynamics-consistent force targets help suppress redundant internal forces, while interaction-aware policy learning improves shared-object control beyond what individual self-balancing can achieve.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The paper's contributions and scope are specifically claimed in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our work in Section 5.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We propose and proof our theory in Section 3 and provide more detailed analysis in Appendix A and Appendix B.3.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We fully describe our settings in simulation and real world experiments in Section 4 and Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We plan to release code and data soon.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: We specify our training and test details in Appendix A, Appendix B.3 and Appendix C.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We define the evaluation metrics in Section 4.1 and C.2. Then we report them in Section 4.2.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provided our experimental details in Appendix C.1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We read the NeurIPS Code of Ethics. Our work avoids harm, respects privacy, promotes fairness, and adheres to responsible reporting and data handling practices.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper doesn't discuss anything relative to societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Our Paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All external assets used in this work are properly credited, and their licenses and terms of use are explicitly acknowledged and fully respected.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: Our paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: This research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.