arXiv:2603.01126v1 [cs.RO] 1 Mar 2026

Pro-HOI：感知式根部引导人形机器人-物体交互

Yuhang Lin[1,2] Jiyuan Shi[1] Dewei Wang[1,3] Jipeng Kong[1,4]Yong Liu [2] Chenjia Bai[1†] Xuelong Li[1†]人工智能研究院（TeleAI）、中国电信 [2]
浙江大学[3]中国科学技术大学 [4]上海科技大学

†通讯作者



图 1：Pro-HOI 使人形机器人能够仅凭 Unitree G1 上的机载感知与计算资源，执行具备泛化能力且鲁棒的全身物体交互。a) **风格化行为**：在动作跟踪奖励的引导下，机器人展现出风格化技能，例如搬运时奔跑。b)**导航与避障**：根节点引导接口便于与高层规划器无缝集成，从而同时实现搬运和避障。c) **自主重新抓取**：机器人能够自主检测意外掉落，并利用数字孪生模块进行状态估计；即使物体掉出摄像头视场，也能将其找回。d)**连续作业**：机器人成功完成了连续 15 次以上的搬运循环，展现出长时域鲁棒性。

摘要——由于缺乏通用的控制接口和稳健的闭环感知机制，人形机器人难以可靠地执行人形机器人-物体交互（HOI）任务。在本研究中，我们提出了感知式根节点引导人形机器人-物体交互（Pro-HOI），这是一个面向稳健人形机器人移动操作的通用框架。首先，我们收集适用于真实世界部署的箱子搬运动作，并通过有符号距离优化穿透伪影

场损失。其次，我们提出了一种新颖的训练框架，以期望的根节点轨迹为策略提供条件，同时仅将参考动作用作奖励。这一设计不仅免去了复杂的奖励调优，还将根轨迹确立为高层规划器的通用接口，从而实现导航与运动操作的同步进行。此外，为确保运行可靠性，我们引入了持久化物体估计模块。通过融合真实世界的

表 I： 不同方法的能力比较

方法	无外部设备	全身控制	可泛化交互	可控	故障恢复
OmniRetarget[43]	✓	✓	×	×	×
HDMI[37]	×	✓	×	×	×
DemoHLM	✓	✓	✓	×	×
PhysHSI[34]	✓	✓	✓	×	×
Falcon[47]	✓	✓	✓	✓	×
OURS	✓	✓	✓	✓	✓

实时检测与数字孪生，该模块使机器人能够自主检测滑移并触发重新抓取动作。在 Unitree G1 机器人上的实证验证表明，Pro-HOI 在泛化能力和鲁棒性方面显著优于基线方法，并能在复杂的真实世界场景中可靠地完成长时程任务。项目页面可在项目页面查看。

I. 引言

人形机器人具有高自由度（DoFs）和拟人化构型，因此被广泛视为通用机器人的理想平台。尽管最先进的方法已展现出执行高动态动作[16, 39, 8, 48] 和穿越复杂地形 [49, 18, 33],的卓越能力，但执行搬箱子等全身操作任务仍是一项重大挑战。这一难题源于在外部负载下保持平衡的严格要求、通过闭环感知与物体进行稳健交互 [7],，以及与高层规划器集成的可扩展性[5]。弥合孤立的运动原语与集成式感知控制框架之间的差距，对于实现可靠的搬运能力至关重要，也是推动人形机器人从实验室示范走向以人为中心环境中的实际应用的关键里程碑。

现有的人形机器人搬运任务研究面临诸多局限。传统的基于模型的方法通常需要针对特定任务设计优化目标，并且依赖精确的动力学建模 [6, 1, 17], ，因此难以泛化到未知场景。强化学习（RL）已成为一种可扩展的替代方案，可用于发现复杂且接触丰富的行为 [9, 41, 12]。然而，端到端强化学习方法往往需要繁琐的奖励设计，并针对不同任务阶段分别调整参数以确保收敛 [46]。近期研究尝试通过引入对抗运动先验（AMP） [25] 和由粗到精的感知方案 [34]，来提升动作的自然度和泛化能力。HDMI [37]提出从视频中学习交互技能，并设计统一的交互任务，从而构建了统一的人形物体交互训练框架。然而，该方法不仅依赖难以迁移的光学动作捕捉系统，还将物体交互位置限制在训练动作所涵盖的范围内。此外，现有方法通常缺乏供高层规划器使用的通用命令接口 [40]，也未能在任务执行过程中融入闭环感知 [43]。例如，如果机器人在搬运过程中掉落箱子，实用的系统应能使机器人自主检测到故障并将箱子拾起。这些局限使当前方法难以应用于真实世界的搬运任务。

在本文中，我们提出**Pro-HOI** （**P**感知式**Ro**根部引导的**H**人形-**O**物体**I**交互），一种面向人形机器人-物体交互任务的可泛化且鲁棒的强化学习控制器。首先，我们将运动跟踪目标与观测空间解耦，使策略仅以期望的根轨迹为条件，而仅将全身跟踪用作奖励。这使得策略能够学习根轨迹与任务状态之间的动态关联，从而在无需全身参考动作的情况下实现可泛化的物体交互。此外，该设计将根轨迹确立为通用接口，便于与多样化的高层规划器无缝集成，从而同时执行搬运、导航和避障任务。其次，为应对意外掉落等真实世界中的不确定性，我们引入了持久化物体估计模块。该模块将实时物体检测与数字孪生相结合，使机器人能够检测掉落并自主重新获取物体，从而显著提升系统可靠性。表 I 对我们的框架与最先进的人形机器人-物体交互方法进行了定性比较。尽管 OmniRetarget [43] 和 HDMI [37] 等早期方法为全身控制奠定了基础，但它们要么泛化能力较差，要么依赖外部设备。近期基于模仿的方法，如 PhysHSI [34] 和 Falcon [47], ，显著提升了交互的泛化能力与可控性。然而，一个关键缺口仍然存在：这些基线方法均未考虑执行鲁棒性。相比之下，我们的方法是唯一将可控、可泛化的交互与独特的失败恢复机制相结合的整体解决方案，从而在无需外部仪器设备的情况下确保可靠部署。

为评估我们的方法，我们针对箱子搬运任务开展了大量实验，涵盖多种实际挑战，例如在不同位置的箱子间进行连续多周期运输、搬运时避障，以及在扰动下自主重新抓取。结果证实，Pro-HOI 不仅支持无缝高层集成，还具有出色的鲁棒性，从而能够可靠地执行长时程任务。

总而言之，本工作的主要贡献如下：

- 我们提出了一种用于人形机器人-物体交互任务的根轨迹引导的强化学习框架。该框架无需大量

参考动作或复杂的奖励设计，即可实现具备泛化能力的人形机器人-物体交互能力。此外，根轨迹充当了一种灵活的接口，可与高层规划器无缝集成，从而同时实现导航与避障。

- 我们引入了一个持久化物体估计模块，将实时物体检测与数字孪生相结合。该模块能够自主重新抓取掉落的物体，显著提升系统在真实世界场景中的鲁棒性和实用性。
- 我们将完整系统部署在一台 Unitree G1 人形机器人上，所有模块均运行于单台机载 Jetson NX 上。大量真实世界实验验证了我们方法的泛化能力、可扩展性和鲁棒性。

II. 相关工作

A. 人形机器人控制的强化学习

通过大规模仿真训练和仿真到现实迁移技术，强化学习使人形机器人能够实现鲁棒的移动 [27, 38, 18], 、高度敏捷的机动动作 [49, 31],以及恢复行为 [11, 13]。为了使机器人能够表现出类似人类的行为，模仿人类示范的模仿式方法 [23] 已扩展至高度拟人化的动态人形机器人全身控制[10, 39, 16, 48, 8, 43]。尽管这些方法展现出卓越的能力，但其泛化能力有限，尤其是在需要广泛适应不同场景的现实世界人形机器人-物体交互任务中。诸如对抗运动先验（AMP） [33, 34] 这样的替代范式通过从数据集中学习风格分布，减少了对精确跟踪的需求。然而，这些方法通常需要在风格保真度与任务泛化能力之间权衡。

与以往工作不同，我们提出了一种根节点引导的人形物体交互训练范式。通过将根节点指令与运动跟踪奖励解耦，策略不仅展现出拟人化行为，还具备较强的泛化能力。

B. 人形机器人-物体交互

人形机器人-物体交互（HOI） [20, 35] 面临着复合挑战，因为其涉及高维状态空间和复杂的接触动力学，在真实世界部署中尤为如此。早期采用轨迹优化（TO） [29, 6, 17] 的方法能够针对特定任务生成物理上可行的运动，但往往难以泛化至未见过的场景。

近期基于强化学习的方法在人形机器人-物体交互方面展现出更强的鲁棒性和泛化能力 [46, 37,34, 47, 45]。通过精心设计奖励和任务，基于强化学习的控制器能够学习涉及特定接触序列的全身操作技能 [46]。然而，这类方法需要复杂的奖励设计，并且由于缺乏参考轨迹，往往难以约束所学习运动的风格质量。一种颇具前景的替代方案是直接从大规模人类视频中学习交互技能 [37]。尽管这种范式下的数据获取具有可扩展性，但该接近方式依赖外部动作捕捉（MoCap）系统进行部署，并且在物体位置方面通常缺乏泛化能力。VisualMimic[45]采用分层全身控制框架，训练人形机器人执行特定的人形机器人-物体交互任务；然而，所学习的策略仅限于与始终处于视场内的物体进行交互。FALCON[47]通过双智能体学习框架实现力自适应人形物体交互，并可与高层规划器结合以执行复杂任务。然而，这些方法仅适用于物体始终处于机器人视场内的场景。不同于运动跟踪或分层方法，PhysHSI[34] 引入 AMP 来学习端到端人形物体交互策略，并显著提升泛化能力，但这也不可避免地带来了繁琐的任务奖励调优负担。

相比之下，我们的方法将目标根轨迹观测与运动跟踪奖励解耦。这一设计使机器人在保留类人运动风格的同时，能够执行具有泛化性的HOI任务。此外，根轨迹可作为一种通用接口，与高层规划器兼容，从而支持导航和避障等复杂任务。另外，通过利用数字孪生模块，在部署期间对视野之外的物体进行状态估计，我们的系统在真实世界中成功连续完成了15次以上的搬箱循环，展现出卓越的泛化能力和鲁棒性。

III. 方法

A. 问题定义

我们将人形机器人-物体交互建模为目标条件强化学习问题，其中智能体根据策略 π 在物理环境中与物体交互，以最大化累积奖励。我们采用 Unitree G1 机器人作为智能体，其总自由度为 29 个。在每个时间步 t，策略接收状态 $\boldsymbol{s}_t$和目标状态 $\boldsymbol{g}_t$，并输出动作 $\boldsymbol{a} \in \mathbb{R}^{29}$，该动作表示目标关节位置，随后通过 PD控制器转换为电机扭矩。状态转移由环境动力学 $p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\ \boldsymbol{a}_t)$决定，并采用稠密奖励函数 $r(\boldsymbol{s}_t,\ \boldsymbol{g}_t,\ \boldsymbol{a}_t)$进行任务性能评估和正则化。优化目标是最大化期望折扣回报：$J = \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right]$。

策略采用非对称演员-评论家架构 [26]，并使用近端策略优化（PPO） [30] 进行训练，以确保其能够部署在物理硬件上。训练期间，评论家可以访问特权信息（例如基座速度），而演员则使用真实机器人可获取的带噪观测。

B. 数据准备与优化

尽管现有大量人-物交互数据集 [15, 3], ，但其中的高动态动作（如大幅弯腰）往往难以由人形机器人模仿，从而增加了过热风险。为提升交互性能，我们首先使用 Xsens 惯性动作 [22] 捕捉系统，以 SMPL 格式采集了一个人类搬箱动作自定义数据集

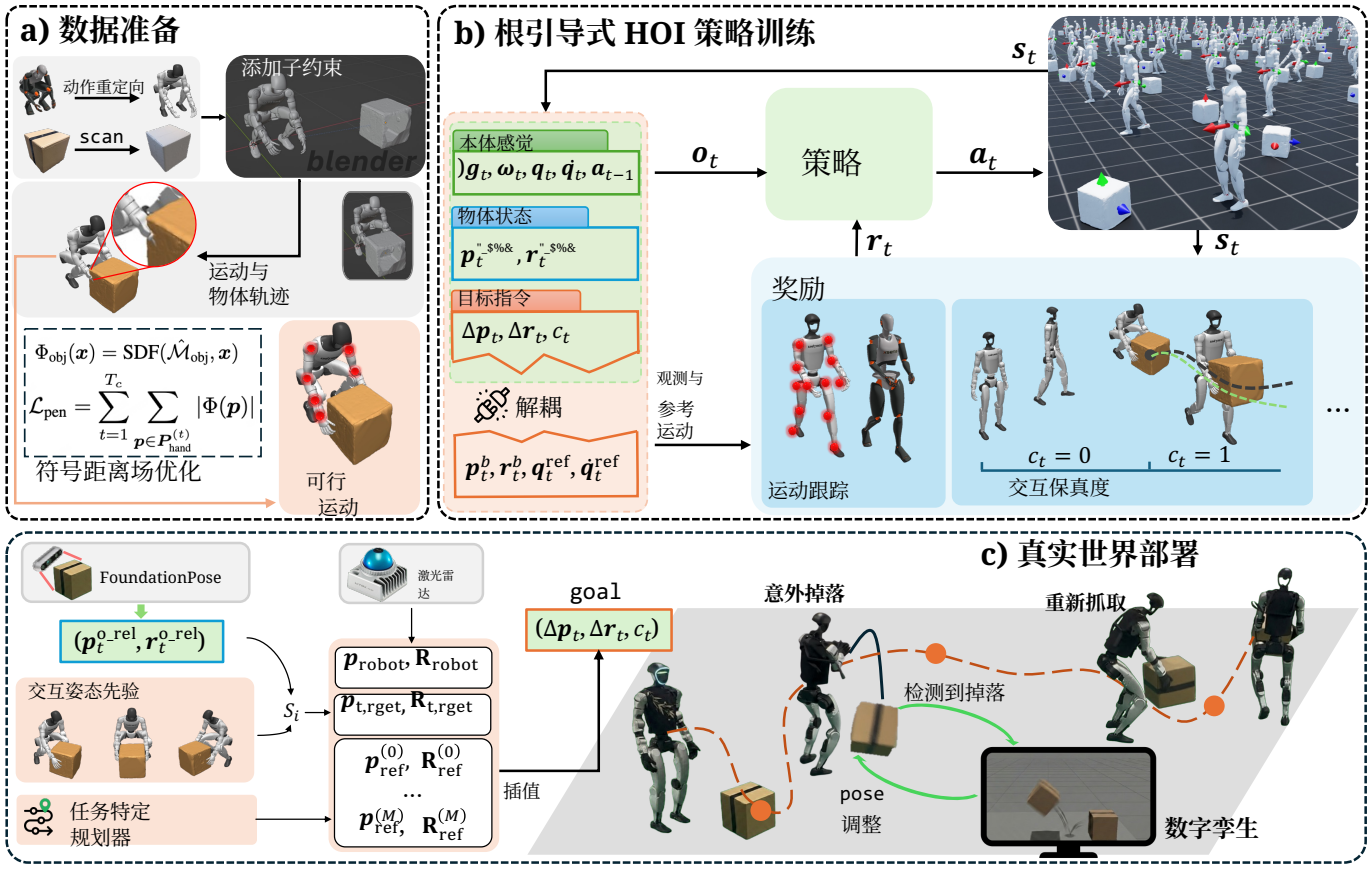

图 2：Pro-HOI 概览。a) **数据准备**：我们首先使用动作捕捉系统采集人体运动片段，并在 *Blender* 中为其添加物体几何形状。随后，我们使用基于 SDF 的优化生成物理上可行的参考动作。b) **根轨迹引导的策略学习**：训练强化学习策略，使其在期望根轨迹的条件下执行全身交互技能，并利用参考动作作为奖励。c) **真实世界部署**：我们集成 FoundationPose 进行六维物体姿态估计，并使用 FAST-LIO2 [42] 进行根姿态估计，结合交互姿态先验和任务专用规划器生成目标根轨迹。整个系统完全在搭载 D435i 相机和 Mid-360 激光雷达的 Jetson NX 上运行，实现了稳健的仿真到现实迁移。

捕捉系统。随后，我们使用通用动作重定向（GMR）[2]将 SMPL 动作重定向至人形机器人。由于没有捕捉物体状态，现有的人-物交互重定向方法 [43]并不适用。为解决这一问题，我们在 Blender [4]中使用“Child Of”约束，将物体与机器人的末端执行器手动关联。

精确的人形机器人与物体对齐对于模仿学习框架至关重要。我们基于 SDF [21, 44]构建了穿透损失，以消除机器人末端执行器与物体之间的网格穿透伪影。

具体而言，我们采用基于体素的下采样来获得简化网格 $\mathcal{M}_{\text{hand}}$ 和 $\mathcal{M}_{\text{obj}}$，并根据物体网格构建符号距离场 $\Phi_{\text{obj}}(\cdot)$。随后，我们在接触阶段 $T_c = [\phi_1, \phi_2]$内优化机器人上半身关节的位置 $\boldsymbol{q}_{\text{upper}}$，该阶段由物体的线速度计算得出。穿透损失定义为：

$$\mathcal{L}_{\text{pen}} = \sum_{t=\phi_1}^{\phi_2} \sum_{\boldsymbol{p} \in \boldsymbol{P}_t^{\text{hand}}} |\Phi_{\text{obj}}(\boldsymbol{p})| \tag{1}$$

其中，$\boldsymbol{P}_t^{\text{hand}}$表示时间步 t 时物体坐标系中的手部顶点。我们使用 Adam优化器 [14] 最小化 $\mathcal{L}_{\text{pen}}$，使双手与物体表面保持接触

。

C. 根引导式 *HOI* 策略训练

1) 人-物交互过程编码：基于模仿的人形机器人-物体交互面临的一项核心挑战，是动作保真度与泛化能力之间的权衡。现有的模仿方法通常会将全身参考动作（例如关节位置和速度）[16]或标量相位变量（例如 $\phi \in [0, 1]$）[10] 纳入策略观测。然而，直接观测高维参考动作会给学习身体姿态与物体位置之间复杂映射带来沉重的数据负担[19]。由于交互力学会随物体位置发生显著变化，要实现全面的数据集覆盖在实践中几乎不可行。相反，低维相位变量缺少足够的空间上下文，难以引导诸如搬箱子等复杂的多阶段任务。

为弥合这一差距，我们提出一种根节点引导的观测范式。通过利用期望的根节点运动趋势，我们能够有效编码任务的语义阶段。以搬运箱子为例，垂直运动表示蹲下或站立，水平运动则表示移动或搬运。将这一空间趋势与二元期望接触状态 $c_t \in \{0, 1\}$ 相结合，便构成了一个极简但足够的观测空间。

期望接触状态 c_t 表示根据参考动作，机器人的末端执行器是否应与物体接触。这种设计促使策略通过奖励信号学习期望运动、接触状态与不同任务阶段之间的关联，而非仅仅模仿参考动作。值得注意的是，这种精简的设计具有较高的数据效率，只需有限的人-物交互示范数据即可实现有效训练。此外，将根轨迹用作观测，有助于与高层路径规划器无缝集成，从而执行复杂的物体交互任务。

具体而言，我们将上述观测状态定义为目标 $\boldsymbol{g}_t = [\Delta \boldsymbol{p}_t^{\text{root}}, \Delta \boldsymbol{r}_t^{\text{root}}, c_t]$。 $\Delta \boldsymbol{p}_t^{\text{root}}$和 $\Delta \boldsymbol{r}_t^{\text{root}}$表示当前状态与预期状态之间根位置和方向的偏差。

2）状态空间：评论家和演员使用不同的信息集。在每个时间步 t，评论家都可以访问完整状态 $\boldsymbol{s}_t$，其中包括本体感知状态、物体状态 $\boldsymbol{s}_t^{\text{o}}$和目标 $\boldsymbol{g}_t$：

$$\boldsymbol{s}_t = \left[\boldsymbol{p}_t^{\text{body}}, \boldsymbol{r}_t^{\text{body}}, \tilde{\mathbf{g}}, \boldsymbol{v}_t, \boldsymbol{\omega}_t, \boldsymbol{q}_t, \dot{\boldsymbol{q}}_t, \boldsymbol{a}_{t-1}, \boldsymbol{p}_t^{\text{o_rel}}, \boldsymbol{r}_t^{\text{o_rel}}, \boldsymbol{g}_t\right] \tag{2}$$

其中， $\boldsymbol{p}_t^{\text{body}} \in \mathbb{R}^{14\times 3}$ 和 $\boldsymbol{r}_t^{\text{body}} \in \mathbb{R}^{14\times 6}$ 表示机器人的关键身体部位的位置和旋转。根状态由投影重力 $\tilde{\mathbf{g}} \in \mathbb{R}^3$、线速度 $\boldsymbol{v}_t \in \mathbb{R}^3$和角速度 $\boldsymbol{\omega}_t \in \mathbb{R}^3$组成。关节状态包括位置 $\boldsymbol{q}_t \in \mathbb{R}^{29}$、速度 $\dot{\boldsymbol{q}}_t \in \mathbb{R}^{29}$和先前的动作 $\boldsymbol{a}_{t-1} \in \mathbb{R}^{29}$。物体状态包括其在机器人坐标系中的位置 $\boldsymbol{p}^{\text{o_rel}}_t \in \mathbb{R}^3$和旋转 $\boldsymbol{r}^{\text{o_rel}}_t \in \mathbb{R}^6$。

相比之下，演员 π 接收部分观测 $\boldsymbol{o}_t$，以确保部署可行性。具体而言， $\boldsymbol{o}_t$ 不包含难以在物理硬件上估计的特权信息——即机器人的关键身体姿态和根部线速度 $\boldsymbol{v}_t$。因此，观测被定义为一个子集：$\boldsymbol{o}_t = \boldsymbol{s}_t \setminus \{\boldsymbol{p}_t^{\text{body}}, \boldsymbol{r}_t^{\text{body}}, \boldsymbol{v}_t\}$。此外，为弥合仿真到现实之间的差距，我们在策略的观测中注入高斯噪声，以模拟传感器不确定性；而评论家则基于无噪声状态进行运算，以确保价值估计稳定。

3）奖励项：在不进行任务特定奖励设计的情况下，我们采用简单的基于跟踪的奖励项 $\mathcal{R}(\cdot,\ \cdot)$ 来指导训练过程。训练期间，对于每个回合，我们从数据集中随机采样一个参考人-物交互运动状态 $\boldsymbol{s}_t^{\text{ref}}$。机器人和物体被初始化为该序列的初始状态。在每个时间步 t，我们将总奖励 r_t 计算为以下两个组成部分之和：

$$r_t = \mathcal{R}(\boldsymbol{s}_t, \boldsymbol{s}_t^{\text{ref}}) + \mathcal{P}_{\text{reg}}(\boldsymbol{s}_t, \boldsymbol{a}_t) \tag{3}$$

具体而言，跟踪奖励 $\mathcal{R}(\boldsymbol{s}_t,\ \boldsymbol{s}_t^{\text{ref}})$ 用于最小化全局根姿态的偏差，并激励智能体复现参考关节运动学，从而隐式地引导必要的全身协调。对于物体，该奖励优先考虑操作保真度，鼓励精确跟踪参考轨迹，并确保模拟接触状态与参考接触状态保持一致。最后，惩罚项 $\mathcal{P}_{\text{reg}}(\boldsymbol{s}_t^{\text{p}},\ \boldsymbol{a}_t)$ 作为正则化分量，用于惩罚较大的时间变化动作，以确保控制平滑性，抑制关节限位越界，并防止非期望碰撞。

4）仿真到现实迁移：为确保成功实现仿真到现实迁移，我们在训练期间采用域随机化（DR） [24]，以增强策略鲁棒性。具体而言，我们对一组物理参数进行随机化，包括摩擦、恢复系数、关节位置偏移、智能体的质心（CoM）以及物体的质量。此外，为模拟外部扰动，我们对机器人的根部线速度与角速度施加随机扰动。

D. 真实世界部署系统

仿真与真实世界之间的一个主要差异在于，无法获取机器人和物体的精确全局位姿。准确获取这些状态对于实现稳健的物体交互和精确搬运至关重要。一方面，由于视场受限、感知延迟以及计算资源有限，依赖机载传感器进行感知面临诸多挑战。在发生遮挡或无约束运动（例如自由落体）时，这一困难会进一步加剧。另一方面，训练基于模仿的策略通常依赖数据集中的参考根轨迹。然而，这些轨迹往往暗示了物体的位置，限制了策略在真实世界部署中与随机位置物体进行交互的泛化能力。为应对这些挑战，我们提出了物体状态估计模块（III-D1），以确保获取稳定的物体状态；同时提出了以物体初始位置和机器人路径点为条件的根轨迹规划器（III-D2），以实现可泛化交互。

1）物体状态估计模块：为确保交互过程中在部分遮挡条件下实现稳健的6D位姿跟踪，我们采用了 FoundationPose [36]。这一统一的基础模型无需微调，仅凭网格模型即可实现对新物体的稳定跟踪。

物体一旦掉落，往往会离开机载传感器有限的视场。将物体视为在重力作用下运动的质点并不足够，因为这忽略了物体与机器人机身及地面之间的几何碰撞。为此，我们提出了一个由 MuJoCo 仿真器[32]驱动的数字孪生模块。借助连续状态估计，我们设计了一种用于检测操作失败的启发式方法。根据接触目标与物体相对位姿之间的不一致性来识别掉落事件。具体而言，如果接触状态持续激活（$c_t = 1$），且相对物体位置的偏差在滑动时间窗 N内超过安全阈值 δ_{pos}，系统便会触发掉落警报。

该数字孪生模块运行于外部计算单元，并与机器人和物体的全局状态实时同步。检测到物体掉落后，模拟器会利用物体最后已知的姿态和速度初始化一个并行环境。随后，模拟器通过高保真物理步进模拟碰撞动力学，并预测最终静止位姿。根据这一预测，机器人会自动调整注视方向，重新捕捉物体，从而实现无缝恢复和重新抓取。

2）根轨迹规划器：我们的策略的核心观测机制依赖于当前状态与参考状态之间的根位姿偏差，该偏差引导着整个交互过程。在训练期间，策略依赖一组有限的固定参考轨迹。然而，在推理期间，可以灵活定义参考轨迹以满足任务需求，例如利用现成的二维规划器或生成式运动模型。为确保物体交互稳定，我们提出一种混合根轨迹规划方法，利用数据集中的抓取先验来引导策略。具体而言，该规划器由三个部分组成：交互姿态先验、任务专用规划器，以及受预期线速度和角速度约束的 6DoF 轨迹插值 $\xi(t)$。

a) 交互姿态先验：为实现稳健的物体交互，我们利用从训练数据集中提取的抓取先验。具体而言，我们记录每个接触阶段开始时机器人与物体的相对位姿。部署期间，检测到新物体后，系统会通过评估当前检测到的物体相对位姿与已存储先验之间的相似度，检索最优交互配置。

b) 任务专用规划器：确定全局交互目标后，策略可以灵活地由不同的轨迹规划器引导，以满足特定任务要求。在本研究中，我们将定时弹性带（TEB） [28]局部规划器作为导航模块的示例，用于执行具备避碰能力的长时域搬运任务。给定目标位置和障碍物地图，规划器会生成一条平滑的平面轨迹，其中包含一系列二维位置和偏航角。随后，我们将这些平面路径点转换为六自由度轨迹，作为期望的根位姿。

c) 六自由度轨迹插值：为策略生成连续的参考根轨迹，我们对构建的路径点序列进行六自由度插值。此过程将机器人的当前状态、特定交互目标以及规划器生成的轨迹融合为统一的时间参数化轨迹，同时根据交互目标输出预期的接触指令。

IV. 评估

本节将在高保真仿真和真实环境中，对我们提出的框架进行全面评估。实验旨在回答以下问题：

- **Q1:** 我们的根轨迹引导策略在搬箱任务中的表现如何？
- **Q2:**我们的策略在跟踪参考根轨迹和接触指令方面的表现如何？
- **Q3:** 该系统能否支持复杂的长时程动态任务？
- **Q4：** 策略和感知系统能否可靠地迁移到物理硬件上？

A. 我们的根轨迹引导策略在箱子搬运任务中的表现如何？

实验设置：任务分为四个阶段：接近、抓取、搬运和放置。为严格评估我们方法的鲁棒性和泛化能力，我们进行了全面测试，并将测试分为两组：分布内（*ID*）和分布外（*OOD*）。分布内测试采用与训练分布完全相同的机器人与物体配置。分布外测试则通过在工作空间内随机采样初始物体位置和目标放置位置来检验策略。具体而言，我们在由 $x \in [0.4, 1.4]$ m 和 $y \in [-x, x]$ m 定义的梯形区域内构建了网格，间距为 0.2m。对于每个网格点，我们评估了从 $\{0°, 15°, 30°, 45°\}$ 开始的物体偏航朝向。此外，以机器人为圆心、半径为 4m 的圆周上，每隔 $10°$ 采样一个目标放置位置。这一全面的评估方案共涵盖 5,756 个不同的任务场景。

表 II 展示了在 MuJoCo 模拟器中，将我们的根引导策略与以下基线方法进行比较所得的性能评估结果：

- **不使用 SDF：** 我们使用未经 SDF 优化的原始重定向运动训练了一个策略变体，同时保持所有其他超参数不变。
- **无接触：** 该基线方法在训练时未使用与接触相关的观测和奖励。
- **PhysHSI：** 这是一种最先进的基线方法，将 AMP 与为搬运箱子任务手动设计的奖励相结合。
- **我们的方法（含故障恢复）：** 这是我们提出的框架的一个变体，其中包含失败恢复策略。

使用以下指标评估性能：

1) 抓取成功率：评估抓取的稳定性，并以 (R_{succ}) 表示。其定义为机器人成功将物体抬升至阈值高度以上，并在至少 1 秒内保持抓取且不发生滑脱的试验所占百分比。我们的方法展现出卓越的抓取稳定性，尤其是在泛化场景中。基线方法在分布外设置下性能显著下降（例如，PhysHSI 从 96.60% 降至 82.54%），而我们的完整方法（**我们的方案（含故障恢复）**）在分布外任务中实现了近乎完美的 99.93% 成功抬升率。这一稳健表现证实，我们的根轨迹引导策略结合失败恢复机制，能够有效应对物体位置的变化，同时不影响抓取稳定性。图 3 展示了抓取成功率的详细可视化结果。

2) 放置精度：用于量化操作精度，并以 R_{prec} 表示。其计算方式为物体最终静止位置与目标中心之间的欧几里得距离的平均值。该指标仅针对成功回合计算，以排除抓取失败产生的异常值。从量化结果来看，我们的方法取得了最高的放置精度。在分布内设置中，**我们的策略（含故障恢复）**。将放置误差降至仅 6.46 厘米，显著优于基线方法，

表 II：成功率、放置精度和执行速度的定量比较（在 MuJoCo 中）。

方法	分布内				分布外			
	R_{succ}（%↑）	R^{prec}（cm↓）	R_{task}(%↑)	V_{exec}(米/秒↑)	R_{succ}(%↑)	R^{prec}(cm↓)	R_{task}(%↑)	V_{exec}(m/s↑)
不使用 SDF	97.90	7.54	55.76	0.29	82.52	8.44	64.26	0.436
无接触	94.99	7.86	31.13	0.29	88.43	8.28	71.65	0.41
PhysHSI [34]	96.60	7.62	**90.59**	**0.49**	82.54	7.25	70.17	**0.72**
Ours	97.90	6.32	73.07	0.29	91.90	**6.83**	76.15	0.42
我们的方法（含 FR）	**100.00**	**5.80**	73.67	0.29	**99.93**	6.84	**88.38**	0.27

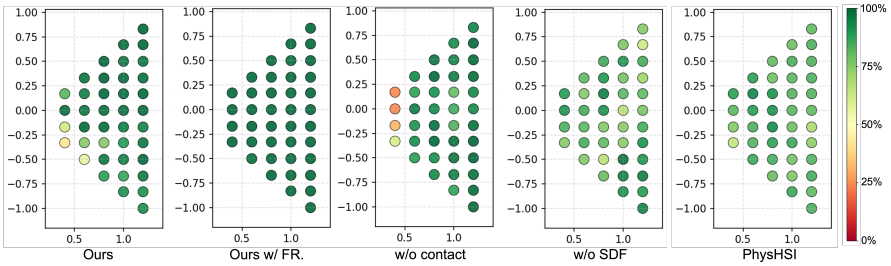

图 3：分布外（OOD）评估集上的抓取成功率空间分布。颜色刻度表示成功率，其中较深的绿色表示更高的稳定性，红色/黄色表示失败。

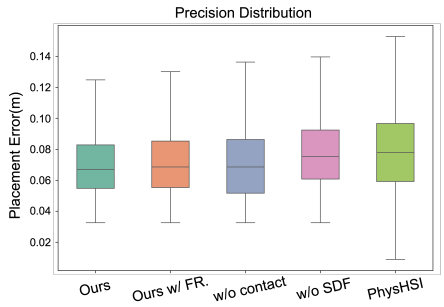

图 4：不同方法的放置精度统计分布。

基线方法的误差约为 9.5 厘米。这一趋势在分布外场景中仍然存在，我们的基础策略和带故障恢复的版本都保持了较高的精度。这表明，精简的观测空间并未牺牲控制精度；相反，通过专注于根部轨迹，该策略实现了更一致的全局空间规划。图 4 展示了容差阈值为 < 0.2 米时放置精度的箱线图。

*3)*任务成功率：任务完成率指标表示为 R_{task}。只有机器人完成整个抓取与放置序列，且最终物体位置与目标的距离在容差阈值（< 0.1m）以内，才视为试验成功。结果凸显了我们框架的泛化能力。尽管 PhysHSI 在分布内设置下取得了较高的成功率（90.59%），这可能是由于对参考动作过拟合，但其在分布外任务中的表现下降至 70.17%。相比之下，**我们的方法（含 FR）**展现出卓越的鲁棒性，在具有挑战性的分布外设置下取得了最高的任务成功率 88.38%。这验证了我们的解耦观测设计能够使机器人适应新颖的物体位置，并完成传统运动跟踪方法难以胜任的复杂多阶段任务。

表 III： 跟踪保真度与同步的定量比较。指标以**均值** ± **标准差**报告。**RPE**：根位置误差（m），**ROE**：根方向误差（rad），**MPL**：运动相位滞后（s），**CPL**：接触相位滞后（s）。最佳结果以**粗体**标出。

方法	RPE ↓	ROE ↓	MPL ↓	CPL ↓
无接触	0.23 ±0.03	6.39 ±0.53	18.37 ±1.54	20.50 ±7.00
不使用 SDF	0.26 ±0.03	6.33 ±0.79	22.11 ±1.35	**14.25** ±3.28
Ours	**0.22** ±0.02	**5.77** ±0.41	**16.21** ±1.61	14.80 ±3.89

B. 我们的策略在跟踪参考根轨迹和接触指令方面表现如何？

为严格量化策略跟踪参考根位姿并遵循接触时序的能力，我们采用多种轨迹生成方法，在一系列目标速度下进行了广泛评估。我们记录了生成的参考轨迹 ξ_{ref}，以及机器人执行的根轨迹 ξ_{real}。性能通过以下四项指标进行衡量：

- **根位置误差（RPE）：** 参考根位置与实际根位置之间的平均欧几里得距离。
- **根方向误差（ROE）：** 参考根方向与实际根方向之间的平均角偏差。
- **运动相位滞后（MPL）：** 定义为机器人根部到达参考轨迹所指定空间位置所需的时间延迟。该指标反映了策略的动态响应速度。
- **接触相位滞后（CPL）：** 参考接触时间戳（即 c_t 切换为 1 时）与实际接触开始时间之间的平均时间差。

我们将我们的方法与上述两个消融变体进行基准比较：**不使用 SDF**和**无接触**。

表 III 中的定量结果表明，我们的完整框架在大多数指标上始终优于消融基线方法，在空间精度和时间同步之间实现了最佳平衡。

在空间跟踪保真度方面，我们的方法取得了最低的 RPE 和根方向误差。与不使用 *SDF* 基线相比，我们的方法显著减少了跟踪偏差，这表明基于 SDF 的优化能够有效消除参考动作中的物理不一致性，从而使策略能够学习更精确的运动学跟踪。

在动态响应方面，我们的方法取得了最低的 MPL，优于无接触 和不使用 *SDF*。这表明，纳入显式接触状态和高质量运动先验，使策略能够更有效地预判高动态转换，从而缩短参考轨迹与执行轨迹之间的时间延迟。

在接触同步 CPL 方面，无接触 变体表现出最大的延迟，证实若缺少明确的接触观测，策略便难以在正确时机启动交互。虽然 不使用 *SDF* 的 CPL 略低于**我们的方法**，但考虑到标准差，这一差异微乎其微。重要的是，我们的方法在保持接触时序竞争力的同时，显著提升了空间跟踪和运动响应，验证了所提出框架中各组件的必要性。

C. 系统是否支持复杂、长时程且动态的任务？

为展示集成系统的多功能性和鲁棒性，我们设计了四种具有挑战性的任务场景，如图1所示：

a) 任务*I*：连续长时程运输： 我们验证了系统执行连续重新布置任务的能力。在根轨迹规划器的引导下，机器人放置箱子后会自主返回起点，开始下一轮运输循环。该测试证实了状态机和规划器的循环稳定性。

b) 任务 *II*：避障：我们在运输路径中引入了静态障碍物。借助 TEB 规划器，我们的系统成功生成了无碰撞轨迹，并调整机器人的偏航角和路径，使其在搬运负载的同时绕过障碍物。

c) 任务 *III*：通过数字孪生实现故障恢复：我们模拟了这样一种情形：箱子在搬运过程中从机器人的手中掉落。物体状态估计模块检测到箱子掉落，并利用基于物理的模拟器估计物体的新位置。机器人随后成功调整注视方向、重新规划接近路径并重新抓取物体，展现了应对意外故障的韧性。

d) 任务 *IV*：高动态敏捷性：为探索我们方法的性能上限，我们模拟了高速“奔跑与搬运”场景。结果表明，我们的根轨迹引导策略能够泛化至高度动态的动作，这表明其在未来高速人形机器人操作方面具有潜力。

D. 策略和感知系统能否可靠地迁移到物理硬件上？

我们将系统部署在 Unitree G1 人形机器人上，以评估其仿真到现实迁移能力。

1）真实环境操作与动态鲁棒性：我们将系统部署在实体 Unitree G1 机器人上，以评估其在不同动态条件下的鲁棒性。为系统地评估策略在不同运行速度下的表现，我们将引导指令分为三种速度档位：慢速（$v = 0.2$m/s, $\omega = 0.6$ rad/s）、中速（$v = 0.4$ m/s, $\omega = 1.0$ rad/s），以及快速（$v = 0.6$ m/s, $\omega = 1.3$rad/s）。

定量结果汇总于表 IV。如表所示，系统在不同速度曲线下均表现出稳定的可靠性，所有场景中的抓取成功率都超过 60%。我们特别针对中速曲线进行了扩展评估，共连续试验 28 次。值得注意的是，这组试验中有多次故意使物体脱落的情况。尽管遭遇了这些严重干扰，我们的方法仍成功恢复并完成了任务，彰显了系统在物理硬件上的卓越鲁棒性和稳定性。

表 IV： 不同动态指令曲线下的真实世界表现。**抓取成功率**：抓取成功率（%）；**放置精度**：放置精度（误差，单位：m）。速度表示指令线速度 (v) 和角速度 (ω) 的上限。

曲线	指令速度		抓取成功率 (↑)	放置精度 (m) (↓)
	v (m/s)	ω (rad/s)		
Slow	0.2	0.6	7 / 10	0.15
中速	0.4	1.0	21 / 28	0.16
Fast	0.6	1.3	6 / 10	0.25

2）物体状态估计评估：为了验证我们的数字孪生掉落预测模块的有效性，我们建立了一个以机器人底座为中心、空间分辨率为 0.2×0.2 m 的离散网格坐标系。我们通过在机器人的视场内以任意朝向释放物体，进行了受控掉落实验。该模块的主要功能是在 MuJoCo 中通过下落仿真估计恢复姿态，并在物体掉落后重新调整机器人的注视方向。因此，我们将每个网格单元的成功率定义为：机器人根据 MuJoCo 仿真预测的掉落位置调整姿态后，在其视场内重新获取物体的成功率。图 5 展示了原始相机视场和由物体估计引导的调整后视场内的检测成功率。结果表明，该模块有效扩展了交互过程中的原始检测视场，尤其是以机器人为中心的圆形区域。

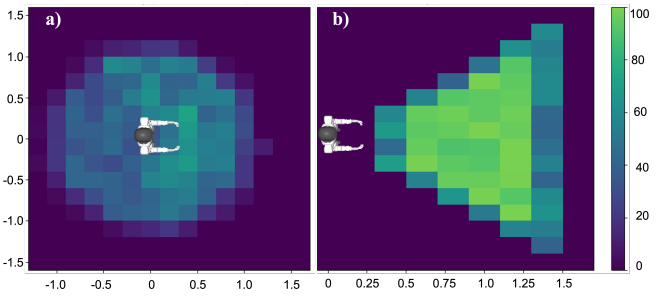

图 5： 这些百分比表示在各自区域内检测到掉落物体的可能性。(a) 通过物体估计模块模拟的检测概率分布。(b) 完全依赖机载相机的基线检测概率分布。

V. 结论

在本研究中，我们提出了 Pro-HOI（感知式根部引导人形机器人-物体交互），这是一种具有泛化能力且稳健的人形机器人-物体交互框架。通过提出一种以根轨迹和接触信息编码交互过程的低维表示，我们的方法克服了传统基于模仿的学习方法固有的泛化局限，同时使人形机器人-物体交互过程可控。大量实验表明，与最先进的基线方法相比，Pro-HOI 在分布外场景中实现了更高的任务成功率、更高的操作精度，并具备可靠的仿真到现实迁移能力。

致谢

本工作得到中国国家自然科学基金（项目编号：62306242）、中国科协青年人才托举工程（项目编号：2024QNRC001）以及上海市扬帆计划（项目编号：23YF11462200）的资助。

参考文献

[1] Alphonsus Adu-Bredu、Grant Gibson 和 Jessy Grizzle。探索用于欠驱动人形机器人的反应式全身控制的动力学织物。载于*2023 IEEE/RSJ*智能机器人与系统国际会议（*IROS*），第10397–10404页。IEEE，2023。[2] Joao Pedro Araujo、Yanjie Ze、Pei Xu、Jiajun Wu 和 C. Karen Liu。动作重定向至关重要：用于人形机器人运动跟踪的通用运动重定向。*arXiv*预印本 *arXiv:2510.02252*，2025。[3] Bharat Lal Bhatnagar、Xianghui Xie、Ilya Petrov、Cristian Sminchisescu、Christian Theobalt 和 Gerard Pons-Moll。Behave：用于跟踪人-物体交互的数据集与方法。载于*IEEE*计算机视觉与模式识别会议（*CVPR*）。IEEE，2022年6月。[4] Blender基金会。Blender，2024。网址：https://www.blender.org。版本4.0。[5]Pengxiang Ding、Han Zhao、Wenjie Zhang、Wenxuan Song、Min Zhang、Siteng Huang、Ningxi Yang 和 Donglin Wang。Quar-VLA：面向四足机器人的视觉-语言-动作模型。载于欧洲计算机视觉会议，第352–367页。Springer，2024。[6]Nadia Figueroa、Salman Faraji、Mikhail Koptev 和 Aude Billard。一种用于人形机器人自适应抓取、导航和协同操作的动力系统方法。载于*2020 IEEE*机器人与自动化国际会议（*ICRA*），第7676–7682页。IEEE，2020。[7] Zhaoyuan Gu、Junheng Li、Wenlan Shen、Wenhao Yu、Zhaoming Xie、Stephen McCrory、Xianyi Cheng、Abdulaziz Shamsah、Robert Griffin、C Karen Liu 等。人形机器人移动与操作：控制、规划和学习方面的当前进展与挑战。*IEEE/ASME*机电一体化汇刊，2025。

[8] Jinrui Han、Weiji Xie、Jiakun Zheng、Jiyuan Shi、Weinan Zhang、Ting Xiao 和 Chenjia Bai。Kungfubot2：学习用于人形机器人全身控制的多功能运动技能。*arXiv*预印本 *arXiv:2509.16638*，2025。[9] Mohamed Hassan、Yunrong Guo、Tingwu Wang、Michael Black、Sanja Fidler 和 Xue Bin Peng。合成物理角色-场景交互。载于*ACM SIGGRAPH 2023*会议论文集，第1–9页，2023。[10] Tairan He、Jiawei Gao、Wenli Xiao、Yuanhang Zhang、Zi Wang、Jiashun Wang、Zhengyi Luo、Guanqi He、Nikhil Sobanbab、Chaoyi Pan 等。Asap：通过对齐仿真与现实世界物理，学习敏捷人形机器人全身技能。*arXiv*预印本 *arXiv:2502.01143*，2025。[11]Xialin He、Runpei Dong、Zixuan Chen 和 Saurabh Gupta。为真实世界人形机器人学习起身策略。*arXiv*预印本 *arXiv:2502.12152*，2025。[12] Dianyong Hou、Chengrui Zhu、Zhen Zhang、Zhibin Li、Chuang Guo 和 Yong Liu。高效学习用于全身操作和运动技能的统一策略。载于*2025*年*IEEE/RSJ*国际智能机器人与系统会议（*IROS*），第5455–5461页。IEEE，2025。[13] Tao Huang、Junli Ren、Huayi Wang、Zirui Wang、Qingwei Ben、Muning Wen、Xiao Chen、Jianan Li 和 Jiangmiao Pang。学习适用于多种姿势的人形机器人起身控制。*arXiv*预印本 *arXiv:2502.08378*，2025。[14] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于学习表征国际会议，2015。[15] Jiaman Li、Jiajun Wu 和 C Karen Liu。物体运动引导的人体运动合成。*ACM*图形学汇刊，42（6），2023。[16]Qiayuan Liao、Takara E Truong、Xiaoyu Huang、Yuman Gao、Guy Tevet、Koushil Sreenath 和 C Karen Liu。Beyondmimic：通过引导扩散，从运动跟踪迈向多功能人形机器人控制。*arXiv*预印本 *arXiv:2508.08241*，2025。[17] Fukang Liu、Zhaoyuan Gu、Yilin Cai、Ziyi Zhou、Hyunyoung Jung、Jaehwi Jang、Shijie Zhao、Sehoon Ha、Yue Chen、Danfei Xu 等。Opt2skill：模仿动态可行的全身轨迹，实现多功能人形机器人移动操作。*IEEE*机器人与自动化快报，2025。[18]Junfeng Long、Junli Ren、Moji Shi、Zirui Wang、Tao Huang、Ping Luo 和 Jiangmiao Pang。利用感知内部模型学习人形机器人移动。载于*2025*年*IEEE*机器人与自动化国际会议（*ICRA*），第9997–10003页。IEEE，2025。[19]Zhengyi Luo、Ye Yuan、Tingwu Wang、Chenran Li、Sirui Chen、Fernando Castañeda、Zi-Ang Cao、Jiefeng Li、David Minor、Qingwei Ben、Xingye Da、Runyu Ding、Cyrus Hogg、Lina Song、Edy Lim、Eugene Jeong、Tairan He、Haoru Xue、Wenli Xiao、Zi Wang、Simon Yuen、Jan Kautz、Yan Chang、Umar Iqbal、Linxi Fan 和 Yuke Zhu。

Sonic：扩展运动跟踪规模，实现自然的人形机器人全身控制。*arXiv*预印本 *arXiv:2511.07820*，2025。[20] Liang Pan、Zeshi Yang、Zhiyang Dou、Wenjia Wang、Buzhen Huang、Bo Dai、Taku Komura 和 Jingbo Wang。Tokenhsi：通过任务标记化统一合成物理人景交互。载于计算机视觉与模式识别会议论文集，第5379–5391页，2025。[21] Jeong Joon Park、Peter Florence、Julian Straub、Richard Newcombe 和 Steven Lovegrove。DeepSDF：学习用于形状表示的连续有符号距离函数。载于*CVPR*，2019。[22] Georgios Pavlakos、Vasileios Choutas、Nima Ghorbani、Timo Bolkart、Ahmed A. A. Osman、Dimitrios Tzionas 和 Michael J. Black。富有表现力的人体捕捉：从单张图像中重建3D手部、面部和身体。载于*IEEE*计算机视觉与模式识别会议论文集（*CVPR*），第10975–10985页，2019。[23] Xue Bin Peng、Pieter Abbeel、Sergey Levine 和 Michiel Van de Panne。DeepMimic：通过示例引导的基于物理的角色技能深度强化学习。*ACM*图形学汇刊（*TOG*），37(4)：1–14，2018。[24]Xue Bin Peng、Marcin Andrychowicz、Wojciech Zaremba 和 Pieter Abbeel。通过动力学随机化实现机器人控制的仿真到现实迁移。载于*2018*年*IEEE*机器人与自动化国际会议（*ICRA*），第3803–3810页。IEEE，2018。[25] Xue Bin Peng、Ze Ma、Pieter Abbeel、Sergey Levine 和 Angjoo Kanazawa。AMP：用于风格化基于物理的角色控制的对抗运动先验。*ACM*图形学汇刊（*TOG*），40(4)：1–20，2021。[26] Lerrel Pinto、Marcin Andrychowicz、Peter Welinder、Wojciech Zaremba 和 Pieter Abbeel。用于基于图像的机器人学习的非对称演员-评论家。*arXiv:1710.06542*，2017。网址：https://arxiv.org/abs/1710.06542。arXiv预印本。[27] Ilija Radosavovic、Tete Xiao、Bike Zhang、Trevor Darrell、Jitendra Malik 和 Koushil Sreenath。基于强化学习的真实世界人形机器人移动。科学机器人学，9(89)：eadi9579，2024。[28]Christian R¨osmann、Felix Hoffmann 和 Torsten Bertram。独特拓扑中的集成式在线轨迹规划与优化。机器人与自主系统，88：142–153，2017。[29] Francesco Ruscelli、Matteo Parigi Polverini、Arturo Laurenzi、Enrico Mingo Hoffman 和 Nikos G Tsagarakis。使用人形机器人执行物理交互任务的多接触运动规划与控制策略。载于*2020*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*），第3869–3876页。IEEE，2020。[30] John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford 和 Oleg Klimov。近端策略优化算法。*arXiv*预印本 *arXiv:1707.06347*，2017。[31] Ruiyu Tan、Xueai Li、Fenglei Ni、Dong Zhou、Yi Ji 和 Xiangyu Shao。通过课程辅助强化学习实现人形机器人的多样化跳跃。载于*2024*年中国自动化大会 *(CAC)*，第2502–2508 页。IEEE，2024 年。[32] Emanuel Todorov、Tom Erez 和 Yuval Tassa。MuJoCo：用于基于模型控制的物理引擎。载于*2012*年*IEEE/RSJ*国际智能机器人与系统会议，第 5026–5033 页。IEEE，2012 年。[33]Dewei Wang、Xinmiao Wang、Xinzhe Liu、Jiyuan Shi、Yingnan Zhao、Chenjia Bai 和 Xuelong Li。MORE：用于在复杂地形上学习人形机器人类生命步态的残差专家混合。*arXiv* 预印本 *arXiv:2506.08840*，2025 年。[34] Huayi Wang、Wentao Zhang、Runyi Yu、Tao Huang、Junli Ren、Feiyu Jia、Zirui Wang、Xiaojie Niu、Xiao Chen、Jiahe Chen 等。PhysHSI：迈向真实世界可泛化的自然人形机器人—场景交互系统。*arXiv* 预印本 *arXiv:2510.11072*，2025 年。[35]Wenjia Wang、Liang Pan、Zhiyang Dou、Jidong Mei、Zhouyingcheng Liao、Yuke Lou、Yifan Wu、Lei Yang、Jingbo Wang 和 Taku Komura。SIMS：通过检索增强脚本生成模拟风格化的人类—场景交互。载于*IEEE/CVF*国际计算机视觉会议论文集，第 14117–14127 页，2025 年。[36] Bowen Wen、Wei Yang、Jan Kautz 和 Stan Birchfield。FoundationPose：新物体统一六维姿态估计与跟踪。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 17868–17879 页，2024 年。[37]Haoyang Weng、Yitang Li、Nikhil Sobanbabu、Zihan Wang、Zhengyi Luo、Tairan He、Deva Ramanan 和 Guanya Shi。HDMI：从人类视频中学习交互式人形机器人全身控制。*arXiv* 预印本 *arXiv:2509.16757*，2025 年。[38] Weiji Xie、Chenjia Bai、Jiyuan Shi、Junkai Yang、Yunfei Ge、Weinan Zhang 和 Xuelong Li。通过动态平衡与强化学习实现人形机器人在狭窄地形上的全身移动。载于*2025*年*IEEE/RSJ*智能机器人与系统国际会议 *(IROS)*，2025 年。[39] Weiji Xie、Jinrui Han、Jiakun Zheng、Huanyu Li、Xinzhe Liu、Jiyuan Shi、Weinan Zhang、Chenjia Bai 和 Xuelong Li。KungFuBot：用于学习高动态技能的基于物理的人形机器人全身控制。载于第三十九届神经信息处理系统年会，2025 年。[40] Zhanteng Xie 和 Philip Dames。DRL-VO：利用速度障碍学习如何在拥挤动态场景中导航。*IEEE*机器人学汇刊，39(4):2700–2719，2023 年。[41] Sirui Xu、Hung Yu Ling、Yu-Xiong Wang 和 Liang-Yan Gui。InterMimic：迈向基于物理的人—物交互通用全身控制。载于计算机视觉与模式识别会议论文集

会议，第 12266–12277 页，2025 年。[42]Wei Xu、Yixi Cai、Dongjiao He、Jiarong Lin 和 Fu Zhang。FAST-LIO2：快速直接激光雷达惯性里程计。*IEEE*机器人学汇刊，38(4):2053–2073，2022 年。[43] Lujie Yang、Xiaoyu Huang、Zhen Wu、Angjoo Kanazawa、Pieter Abbeel、Carmelo Sferrazza、C Karen Liu、Rocky Duan 和 Guanya Shi。OmniRetarget：用于人形机器人全身移动操作和场景交互的交互保持型数据生成。*arXiv* 预印本 *arXiv:2509.26633*，2025 年。[44] Jianglong Ye、Keyi Wang、Chengjing Yuan、Ruihan Yang、Yiquan Li、Jiyue Zhu、Yuzhe Qin、Xueyan Zou 和 Xiaolong Wang。Dex1B：利用 10 亿条示范数据学习灵巧操作。载于机器人学：科学与系统 *(RSS)*，2025 年。[45] Shaofeng Yin、Yanjie Ze、Hong-Xing Yu、C. Karen Liu 和 Jiajun Wu。VisualMimic：通过运动跟踪与生成实现视觉人形机器人移动操作。*arXiv* 预印本 *arXiv:2509.20322*，2025 年。[46] Chong Zhang、Wenli Xiao、Tairan He 和 Guanya Shi。WoCoCo：通过序列接触学习人形机器人全身控制。载于机器人学习会议，第 455–472 页。PMLR，2025 年。[47]Yuanhang Zhang、Yifu Yuan、Prajwal Gurunath、Ishita Gupta、Shayegan Omidshafiei、Ali-akbar Agha-mohammadi、Marcell Vazquez-Chanlatte、Liam Pedersen、Tairan He 和 Guanya Shi。Falcon：学习力自适应人形机器人移动操作。*arXiv* 预印本 *arXiv:2505.06776*，2025 年。[48]Zhikai Zhang、Jun Guo、Chao Chen、Jilong Wang、Chenghuai Lin、Yunrui Lian、Han Xue、Zhenrong Wang、Maoqi Liu、Jiangran Lyu 等。在任意扰动下跟踪任意运动。*arXiv* 预印本 *arXiv:2509.13833*，2025 年。[49] Ziwen Zhuang、Shenzhe Yao 和 Hang Zhao。人形机器人跑酷学习。载于机器人学习会议，第 1975–1991 页。PMLR，2025 年。

补充附录

VI. 实现细节

A. 训练与仿真设置

我们使用 NVIDIA Isaac Lab 作为强化学习仿真环境。策略训练在一台高性能计算平台上进行，该平台配备八块 NVIDIA RTX 3090 GPU 和两颗 AMD CPU。我们还使用搭载 M3 芯片的 Apple MacBook Pro，在 MuJoCo 中开展基于仿真的验证实验。

B. 感知细节

FastLIO2 机器人状态估计器、FoundationPose 物体位姿估计器和机器人控制器均完整部署在机器人搭载的 NVIDIA Jetson Orin NX 模块上。为实现实时感知，我们使用 isaac-ros 框架对接 FoundationPose 模型，并训练 YOLOv8 模型来检测物体。YOLO 的检测结果会转换为掩码，并在初始时刻提供给 FoundationPose。物体模型由 iPhone 14 Pro 使用 AR Code Object Capture 构建，并通过在线工具转换为 OBJ 格式。外部台式工作站接收机器人的实时状态估计结果，用于可视化和记录。高层运动控制器与 Unitree 底层控制接口之间通过轻量级通信与编组（LCM）协议进行通信。最终，我们实现了约 40 Hz 的物体感知、200 Hz 的机器人状态估计和 50 Hz 的机器人控制。

C. 物体状态估计模块详述

掉落检测逻辑依赖于空间安全阈值δ_{pos} 和时间窗口 N。具体而言，我们将安全区域定义为机器人基坐标系中的一个长方体，其尺寸为 $0.4 \times 0.2 \times 0.4$ 米（长 × 宽 × 高），中心位于机器人前向轴上的 $x = 0.35$ 米处。用于持续监测状态的滑动窗口大小设为 $N = 10$。数字孪生仿真部署在同一局域网内的一台外部台式工作站上，并通过 ZeroMQ 协议接收机器人的物体状态流。该模块实时计算物体的速度；触发掉落事件后，物理仿真的初始速度取最近 3 帧速度的平均值，以降低噪声影响。模拟器预测出最终静止位置后，该模块会计算所需的视线调整位姿，并将其发送给机器人以进行恢复。

D. 根轨迹规划器详述

在本附录中，我们给出第 III-D2 节所述混合根轨迹规划器的详细数学公式。这包括交互先验的定量检索指标、从 2D 导航路径到六自由度根参考的显式变换，以及插值序列的构造。

1) 交互姿态先验：为检索最优抓取先验，我们首先使用刚体变换矩阵 $\mathbf{T} \in SE(3)$ 对空间位姿进行形式化表示，其中每个 $\mathbf{T}$ 均包含一个旋转 $\mathbf{R} \in SO(3)$ 和一个平移 $\mathbf{p} \in \mathbb{R}^3$。令 $\mathcal{P} = \{(\mathbf{T}_{\text{obj}}^{(i)}, \mathbf{T}_{\text{int}}^{(i)})\}_{i=1}^{N}$ 表示从训练数据集中提取的先验位姿库。这些位姿分别表示在接触阶段开始时记录的物体位姿 $\mathbf{T}_{\text{obj}}^{(i)}$ 以及相应的机器人交互位姿 $\mathbf{T}_{\text{int}}^{(i)}$（相对于初始坐标系的机器人坐标系）。给定新检测到的物体相对位姿 $\mathbf{T}_{\text{obj}}^{\text{cur}}$，我们定义加权误差函数 S_i，用于量化当前状态与存储的先验之间的相似度：

$$S_i = w_T \|\mathbf{p}_{\text{obj}}^{\text{cur}} - \mathbf{p}_{\text{obj}}^{(i)}\|_2 + w_R \cdot \Phi(\mathbf{R}_{\text{obj}}^{\text{cur}}, \mathbf{R}_{\text{obj}}^{(i)}) \tag{4}$$

其中，$\Phi(\cdot)$计算角距离，而 w_T，w_R 分别是平移误差和旋转误差的加权系数。系统通过$k = \text{argmin}_i S_i$确定最优先验索引。随后，通过补偿当前物体位置与先验物体位置之间的偏差，计算目标交互位姿：

$$\mathbf{p}_{\text{target}} = \mathbf{R}_{\text{root}}^{z} \left(\mathbf{p}_{\text{int}}^{(k)} + (\mathbf{p}_{\text{obj}}^{\text{cur}} - \mathbf{p}_{\text{obj}}^{(k)}) \right) + \mathbf{p}_{\text{root}} \tag{5}$$

$$\mathbf{R}_{\text{target}} = \mathbf{R}_{\text{root}} \mathbf{R}_{\text{int}}^{(k)} \tag{6}$$

随后，根据机器人的当前根状态，将该目标转换到世界坐标系中，以引导策略。

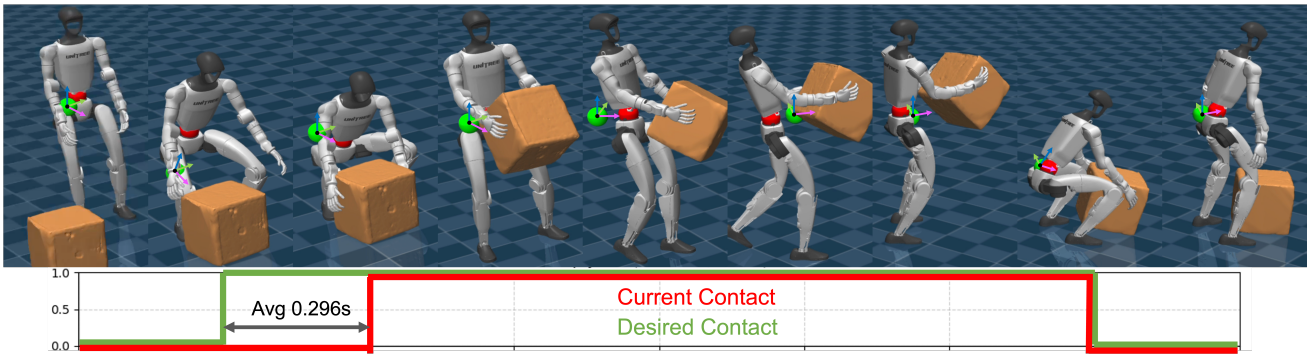

图 6： 结合接触信息和模拟跟踪接触进行轨迹生成的示例。

表 V： 与 PPO 相关的超参数。

超参数	值
优化器	Adam
批次大小	4096
小批次	4
学习轮数	5
熵系数	0.01
价值损失系数	1.0
裁剪参数	0.2
最大梯度范数	1.0
初始噪声标准差	1.0
学习率	1e-3
期望 KL 散度	0.01
GAE λ	0.95
GAE γ	0.99
演员多层感知机	[512,256,128]
评论家多层感知机	[512,256,128]
激活函数	ELU

表 VI： 奖励项及权重。

Term	表达式	权重
	Task	
全局机器人位置	$\exp(-\|\boldsymbol{p}_t - \hat{\boldsymbol{p}}_t\|^2/\sigma_{\text{rpos}})$	0.5
全局机器人朝向	$\exp(-\|\boldsymbol{r}_t - \hat{\boldsymbol{r}}_t\|^2/\sigma_{\text{rrot}})$	0.5
全局躯干位置	$\exp(-\|\boldsymbol{p}_t^b - \hat{\boldsymbol{p}}_t^b\|^2/\sigma_{\text{bpos}})$	1.0
全局躯干朝向	$\exp(-\|\boldsymbol{r}_t^b - \hat{\boldsymbol{r}}_t^b\|^2/\sigma_{\text{brot}})$	1.0
躯干线速度	$\exp(-\|\boldsymbol{v}_t^b - \hat{\boldsymbol{v}}_t^b\|^2/\sigma_{\text{bvel}})$	1.0
躯干角速度	$\exp(-\|\boldsymbol{\omega}_t^b - \hat{\boldsymbol{\omega}}_t^b\|^2/\sigma_{\text{bang}})$	1.0
物体位置	$\exp(-\|\boldsymbol{p}_t^o - \hat{\boldsymbol{p}}_t^o\|^2/\sigma_{\text{opos}})$	1.0
	正则化	
关节限位	$\mathbb{I}(\boldsymbol{q} \notin [\boldsymbol{q}_{\min}, \boldsymbol{q}_{\max}])$	-10.0
非期望接触	$\sum \mathbb{I}(\|\mathbf{F}\| > 1.0)$	-0.1
动作速率	$\|\boldsymbol{a}_t - \boldsymbol{a}_{t-1}\|^2$	-0.3

2) 二维到三维航点提升：对于长时域导航，TEB局部规划器生成一条平面轨迹 $\mathcal{T}_{\text{2D}} = \{(x_i,\ y_i,\ \psi_i)\}_{i=1}^{M}$，其中 $(x_i,\ y_i)$ 表示二维位置，ψ_i 表示偏航角。为使其适用于我们的全身控制器，我们使用以下公式将这些二维路径点提升为六自由度参考轨迹 $\mathcal{T}_{\text{waypoints}} = \{\mathbf{T}^{(i)}_{\text{ref}}\}_{i=1}^{M}$：

$$\mathbf{p}_{\text{ref}}^{(i)} = [x_i, y_i, h_{\text{stand}}]^\top \tag{7}$$

$$\mathbf{R}_{\text{ref}}^{(i)} = \mathbf{R}_z(\psi_i) \tag{8}$$

这里，h_{stand} 表示机器人默认的站立高度，而$\mathbf{R}_z(\psi_i)$是对应于偏航角ψ_i的旋转矩阵，假设滚转角和俯仰角为零，以保持稳定行走。

3) 轨迹插值序列：为确保导航与操作阶段之间平滑过渡，我们通过六自由度插值构建连续的时间参数化轨迹 $\xi(t)$。该插值序列明确包含根据站立高度（$\mathbf{T}^{\text{up}}$）调整的中间"途经点"，以避免接近和搬运过程中的碰撞：

$$\begin{aligned}\xi(t) = \text{Interp}\big(&\mathbf{T}^{\text{curr}}, \mathbf{T}^{\text{up}}_{\text{target}}, \mathbf{T}_{\text{target}}, \mathbf{T}^{\text{up}}_{\text{target}}, \mathbf{T}^{(1)}_{\text{ref}}, \ldots,\\ &\mathbf{T}^{(M)}_{\text{ref}}, \mathbf{T}^{\text{up}}_{\text{end}}, \mathbf{T}_{\text{end}}, \mathbf{T}^{\text{up}}_{\text{end}}\big)\end{aligned} \tag{9}$$

在区间 $[\mathbf{T}^{\text{target}}, \mathbf{T}_{\text{end}}]$内，手动将接触信息 c_t 设为1，其他情况下设为0。示例如图6所示。该序列确保机器人在与目标$\mathbf{T}^{\text{target}}$或最终目标$\mathbf{T}_{\text{end}}$交互之前和之后，先将根部抬升至安全高度。

VII. 补充表格

本节通过四张表详细介绍了 Pro-HOI框架的实现。这些表格分别列出了奖励项的权重、观测空间的维度、域随机化设置以及训练所用的超参数。

A. 奖励项及权重

所有奖励函数详见表 VI。整体奖励设计分为两类：跟踪奖励和正则化奖励。非期望接触 项用于惩罚机器人各连杆之间的自碰撞。具体而言，集合 $\mathcal{B}$ 包含除脚踝和手腕以外的所有机器人身体部位。相对身体包括骨盆、髋部、膝部、脚踝、肩部、肘部、手腕和躯干连杆。

B. 观测

演员和评论家的具体观测项及维度详见表 VII。

表 VII： 演员与评论家的观测状态空间。

状态项		演员维度	评论家维度
全局根位置		3	3
全局根方向		6	6
相对物体位置		3	3
相对物体方向		6	6
接触信息		1	1
根部投影重力		3	3
根部角速度		3	3
关节位置		29	29
关节速度		29	29
动作		29	29
根线速度	y	–	3
相对身体位置		–	45
相对身体姿态		–	90
总维度		112	250

C. 域随机化

由于模拟器的物理特性无法准确反映真实情况，我们在训练过程中引入域随机化，以提高训练策略向真实世界或其他仿真环境的迁移能力。具体设置见表 VIII。

D. PPO 超参数

详细的 PPO 超参数见表 V。

域随机化	采样分布
物理参数	
静摩擦系数	$\mu_s \sim \mathcal{U}[0.3, 1.6]$
动摩擦系数	$\mu_d \sim \mathcal{U}[0.3, 1.2]$
恢复系数	$e \sim \mathcal{U}[0, 0.5]$
默认关节位置	$\boldsymbol{q}_0 \sim \boldsymbol{q}_0 + \mathcal{U}[-0.01, 0.01]$
基座质心偏移（x、y、z）	$\Delta x \sim \mathcal{U}[-0.025, 0.025]$, $\Delta y \sim \mathcal{U}[-0.05, 0.05]$, $\Delta z \sim \mathcal{U}[-0.05, 0.05]$
根速度扰动（外部推力）	
根线速度（x、y、z）	$v_x \sim \mathcal{U}[-0.5, 0.5]$, $v\ y \sim \mathcal{U}[-0.5, 0.5]$, $v_z \sim \mathcal{U}[-0.2, 0.2]$
推送持续时间	$\Delta t \sim \mathcal{U}[1, 3]$秒
根节点角速度	$\omega_{\text{roll}} \sim \mathcal{U}[-0.52, 0.52]$, $\omega_{\text{pitch}} \sim \mathcal{U}[-0.52, 0.52]$, $\omega_{\text{yaw}} \sim \mathcal{U}[-0.78, 0.78]$
任务相关	
物体质量	缩放$_o \sim \mathcal{U}[0.5, 2.0]$

表 VIII： 训练期间应用的域随机化参数。 $\mathcal{U}[\cdot]$: 均匀分布。