arXiv:2603.07988v1 [cs.CV] 9 Mar 2026

TeamHOI：学习适用于任意团队规模的协作人-物交互统一策略

Stefan Lionar[1,2,3] Gim Hee Lee[3]
[1]Garena [2]Sea AI Lab [3]新加坡国立大学

https://splionar.github.io/TeamHOI

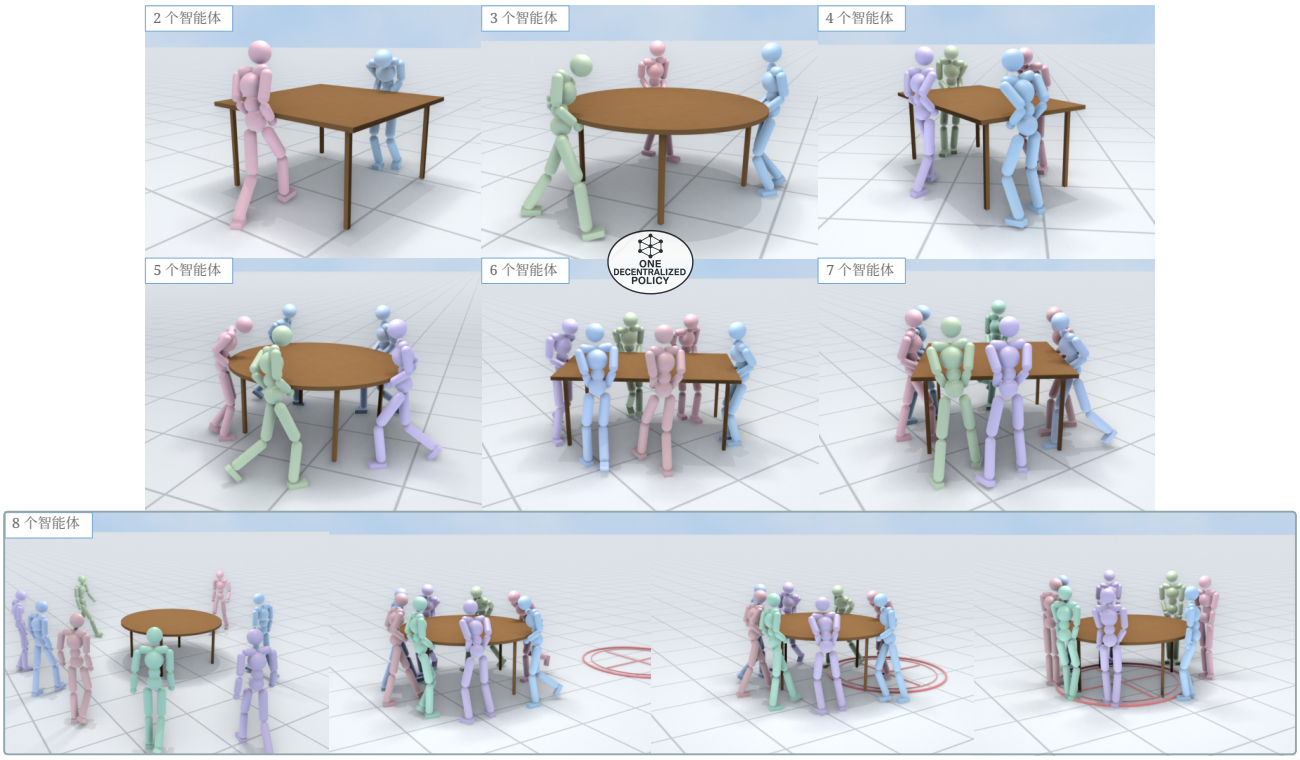

图1。我们提出了**TeamHOI**，这是一个框架，旨在学习统一的去中心化策略，以适用于不同团队规模和物体配置下的协作式人-物交互（HOI）。我们的框架能够实现有效协作：每个人形机器人都基于局部观测独立行动，同时通过单一共享策略与其他成员协调。视频示范见我们的网页。

摘要

基于物理的人形机器人控制在实现逼真且高性能的单智能体行为方面取得了显著进展，但要将这些能力扩展到协作式人-物交互（*HOI*）仍然颇具挑战。我们提出了 *TeamHOI*，这是一个框架，使单一去中心化策略能够处理任意数量协作智能体之间的协作式人-物交互。每个智能体都利用局部观测进行操作，并通过带有队友令牌的基于*Transformer*的策略网络关注其他队友，从而实现不同团队规模下的可扩展协调。为了保证运动真实性并应对协作式人-物交互数据稀缺的问题，我们进一步提出一种掩码对抗运动先验（*AMP*）策略，该策略使用单人参考动作，并在训练期间对与物体交互的身体部位进行掩码处理。随后，利用任务奖励引导这些掩码区域，从而产生多样且符合物理规律的协作行为。我们在一项具有挑战性的协作搬运任务中评估 *TeamHOI*，该任务涉及 *2* 至 *8* 个 *humanoid* 智能体以及形状各异的物体。最后，为促进稳定搬运，我们设计了一种不受团队规模和物体形状限制的队形奖励。*TeamHOI* 取得了较高的成功率，并通过单一策略，在多种配置下展现出协调一致的协作能力。

1. 引言

基于物理的人形机器人控制和人-物交互（HOI）发展迅速，使虚拟人和机器人能够以逼真的动作行走、抓取和操控物体 [10, 52, 61]。然而，许多日常任务（例如抬起大型重物）需要多个智能体协调各自的身体动作。打造能够在此类场景中协作的人形机器人，是迈向更强大、更智能系统的关键一步。除了机器人领域，这种能力也为新一代创意 AI 应用开辟了令人期待的方向，例如多角色动画和交互式游戏世界；在这些场景中，虚拟人形机器人必须彼此自然地协调。

然而，现有的基于物理的人形机器人运动框架在应用于协作式人-物交互时，在可扩展性和数据多样性方面仍存在重大局限。大多数方法依赖固定输入尺寸的多层感知机策略来生成控制动作，而将此类架构用于多智能体交互，会使策略受限于固定的团队规模 [37]。另一种方法完全省略了智能体之间的显式通信，转而仅依靠共享物体动力学作为间接通信渠道 [7]。这类设计无法捕捉人类协作的适应性本质：个体会持续感知队友的存在，并根据团队的组成和规模调整协调方式。

另一项关键局限在于数据来源。许多基于物理的人-物交互框架利用对抗运动先验（*AMP*），通过使学习到的动作符合参考动作数据来确保动作保持自然。然而，多人协同活动的参考动作大多无法获取，因此协作人-物交互框架不得不依赖单人演示。这一限制约束了可实现的协作行为。协调模式只能与一名示范者的动作挂钩，从而降低了处理智能体数量较多的群体协作式人-物交互时的灵活性。

为解决这些局限性，我们提出了*TeamHOI*，这是一个框架，使单一去中心化策略能够泛化至任意数量的协作智能体。每个智能体利用局部观察独立运行，同时共享同一策略网络参数。为建模不同团队规模下的协作，我们采用基于Transformer的策略网络，有效消除了MLP策略的固定输入规模限制，并将其他智能体的状态作为队友令牌纳入其中。策略在配置了不同团队规模的环境中进行训练，从而接触到多样的协调模式，并能够在无需重新训练或微调的情况下，顺畅适应不同团队规模及其相应的协调需求。

TeamHOI 还解决了运动先验所带来的数据多样性限制。为了拓展可行协作行为的多样性，我们在 AMP 监督过程中使用单个人类演员的参考动作，同时屏蔽与物体交互的身体部位。随后，我们通过任务奖励引导被屏蔽区域实现理想的交互。例如，调整手-物交互奖励后，侧向行走的参考动作便可用于侧向搬举。这种掩码策略有效拓宽了可行的人-物交互技能范围，使多样化的协作行为能够从单人参考动作中涌现。

作为评估所提框架的具体测试平台，我们通过一项具有挑战性的协作搬运任务展示TeamHOI，如图1所示。在该任务中，一组人形智能体必须接近并运输一个物体，具体来说，是一张形状为正方形、长方形或圆形的桌子。智能体必须相互协调，组成有利于稳定抬升的队形，并共同将桌子运送至指定目标位置。为完成这项任务，我们设计了一种队形奖励，该奖励与桌子的形状和协作智能体的数量均无关，可引导智能体分布到适合搬运的稳定位置。通过大量实验，我们表明，所提出的框架使单一策略能够顺畅适应由两到八个协作智能体组成的各种配置，同时实现较高的成功率和协调一致的协作行为。

总之，我们的贡献 ibuti如下： follows:

- 我们提出了*TeamHOI*，这是一个框架，使单一的去中心化策略能够让任意数量的智能体执行协作式人-物交互。
- 我们采用基于Transformer的策略网络，从多样化的团队规模配置中学习，并通过队友令牌适应所需的协调方式。
- 我们提出了一种掩码AMP策略，克服了以往运动先验方法在数据多样性方面的局限，拓展了可行且多样化的协作人-物交互行为范围。
- 我们展示了TeamHOI在具有挑战性的搬桌任务中涌现出的协作行为，这些任务涉及形状各异的物体和不同规模的团队。
- 我们设计了一种编队奖励，以促进稳定搬运，且不受桌子形状和协作智能体数量的影响。

2. 相关工作

2.1. 基于物理的人体-场景交互

基于物理的人体-场景交互（HSI）侧重于使人形智能体能够在真实物理条件下与物体和环境进行交互，包括接触、摩擦和动态约束。这些交互通常通过基于物理的控制，在Isaac Gym [39] 和MuJoCo [59]等现代模拟器中实现。一种

常见的训练范式是结合深度强化学习（RL）进行参考跟踪 [3, 30, 40, 43, 71, 78], 通常由对抗运动先验（AMP）加以增强 [44], 其风格奖励使合成运动与参考运动对齐。AMP已被证明能够有效提升各类下游任务中的运动真实性 [18, 45, 55] 尤其是HSI [11, 41, 70]。这一系列工作催生了广泛的高接触技能，例如在体育环境中 [28, 29, 37, 62, 64, 75, 76], 以及日常物体交互中 [2, 7, 11, 25, 26, 32, 35, 41, 57, 63, 69, 70, 73, 74]。

除了掌握单项技能之外，近期研究也在探索更广泛、可复用的能力。运动流形和技能库方法学习通用先验，以便在不同任务间组合 [1, 4, 15, 45, 48], ，而统一控制器则旨在通过单一策略捕捉多样化行为 [13, 34, 36, 56, 65, 67, 68]。最近，TokenHSI [42]引入了基于 Transformer 的策略 [60], 和任务标记化，从而实现多技能统一，支持通用 HSI 并灵活适应新任务。

2.2. 多 humanoid 交互与协作

与单智能体运动合成相比，多人形机器人交互研究仍相对有限。目前已推出若干多 humanoid 数据集，涵盖日常人际活动 [6, 21, 22, 33,50, 72, 77]和编舞式多人运动 [23, 51]。基于这些数据源，研究者提出了许多基于运动学的多角色动画方法 [5, 8, 9, 16, 27, 49, 53, 54, 66, 80, 81]。尽管这些方法在视觉上颇具吸引力，但它们高度依赖高质量的交互数据，且无法保证物理合理性。

为超越纯运动学合成，近期研究探索了利用运动学生成器 [14, 19, 58, 79] 和 PHC [34],实现基于物理的多角色交互。PHC 将多人运动学运动转换为状态-动作对，用于在物理仿真中进行策略学习。这使多个智能体之间能够产生交互式物理行为 [17, 24, 31], ，但不涉及物体。其他基于物理的方法则着重于人群导航 [12, 46] ，对碰撞规避和群体运动进行建模。

另一类工作将物体交互纳入多个人形机器人场景。一种基于模仿的方法 [78] 展示了多个角色围绕共享物体进行交互，但仍受限于高质量多人动作捕捉数据的匮乏。与此同时，SMPLOlympics [37] 展示了物理仿真中的多个人形机器人体育行为，但团队规模固定且较小。与我们工作最相关的是 CooHOI [7], ，它通过共享物体动力学实现隐式通信，从而对协作式人-物交互进行建模。然而，该方法未纳入其他智能体的状态；对于人类协作而言，这一局限并不现实，因为智能体会持续感知并回应彼此。此外，该方法依赖单个行动者的全身参考运动，这限制了可实现的协作行为的多样性。例如，智能体仅展示了向前和向后的抬举动作，无法适应更多样的协作式人-物交互策略。

3. 方法

我们的目标是为协作式人-物交互开发一种统一的去中心化策略，使其能够适应不同的团队规模及相应的协调需求。每个智能体根据局部观测独立行动，同时将其他智能体的状态纳入考量，通过单一策略实现有效协作。

3.1. 初步介绍

我们基于 AMP框架 [44], 构建，该框架通过加入运动先验来增强强化学习，从而确保动作的真实性。在 AMP 中，策略 π_θ 与判别器 $D_\phi(s,\ s')$ 一同训练；判别器用于区分参考动作数据中的短状态转移 $(s,\ s')$ 与策略生成的短状态转移。判别器提供基于风格的反馈信号，鼓励策略生成逼真的动作转移。

强化学习设置： 每个智能体观测本体感知状态 s_t（关节角度、速度和根部位姿）、可选的目标状态 g_t（例如物体目标位置），并输出动作 a_t（目标关节位置或力矩）。环境根据模拟器动力学演化，产生下一状态和任务奖励 r_t^{task}。策略$\pi_\theta(a_t|s_t,\ s_g)$通过近端策略优化（PPO） [47], 进行优化，以最大化期望折扣回报： $J(\theta)=\mathbb{E}_{\pi_\theta}[\sum_t \gamma^t r_t]$。

风格奖励： 为融入运动真实性，AMP 从判别器引入额外的风格奖励：

$$r_t^{\text{style}} = -\log(1 - D_\phi(s, s')), \tag{1}$$

其中（$s,\ s'$）表示当前状态和下一状态，用于捕捉短期运动动力学。总奖励结合这两项，表示为：

$$r_t = r_t^{\text{task}} + \lambda_{\text{AMP}}\, r_t^{\text{style}}, \tag{2}$$

其中 λ_{AMP} 用于平衡任务性能和运动真实性。策略使用r_t通过 PPO 目标函数进行优化，同时判别器经过训练，以区分来自策略的转移和来自参考数据集的转移：

$$\begin{aligned}\mathcal{L}_D = &-\mathbb{E}_{(s,s')^{\text{ref}}}[\log D_\phi(s, s')] \\ &-\mathbb{E}_{(s,s')^{\pi}}[\log(1 - D_\phi(s, s'))].\end{aligned} \tag{3}$$

3.2. TeamHOI 框架

我们的 TeamHOI 框架（图 2）在灵活的多智能体强化学习设置中重新构建了 AMP，使其能够扩展到任意数量的人形智能体。

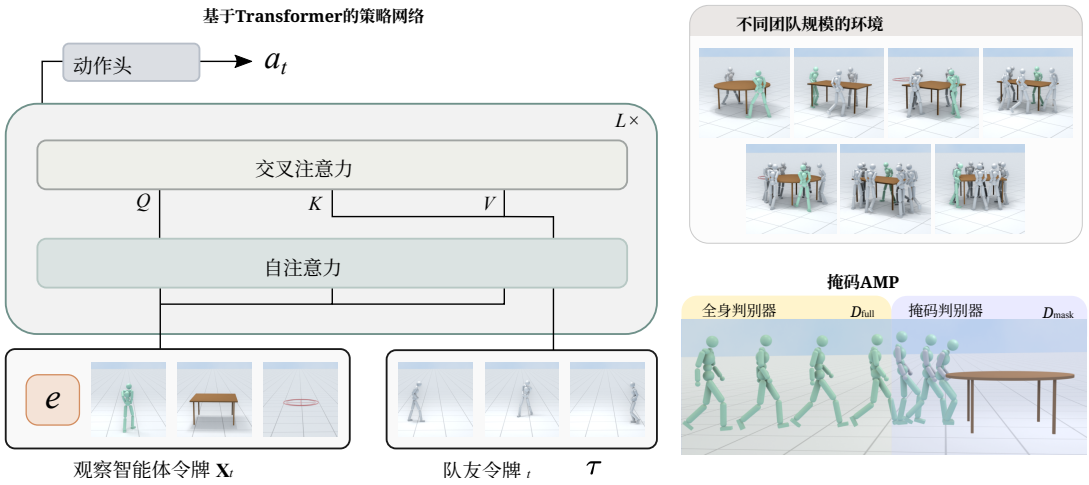

图 2. **TeamHOI 框架概览**。基于Transformer的策略网络通过交替使用自注意力层和交叉注意力层，实现观测智能体（绿色人形机器人）与其队友（灰色人形机器人）之间的协调。通过在不同团队规模的环境中进行训练，该框架学会了适用于不同团队配置的统一策略。为保持运动真实性并提升技能多样性，掩码AMP策略根据物体交互情况，将全身判别器与掩码判别器结合起来。

策略网络： 为了实现可扩展的协调，我们采用了基于 Transformer 的架构作为策略网络，其设计受到 TokenHSI [42]的启发。每个人形智能体都会获得观测 $o_t \triangleq (s_t,\ g_t)$，该观测由其本体感知状态 s_t 和目标状态 g_t组成。每个观测组件首先由专用标记器处理，以生成主要观测智能体的标记序列，$\mathbf{X}_t = [e, \mathbf{T}_t^s、\mathbf{T}_t^g]$，其中 $\mathbf{T}_t^s$和 $\mathbf{T}_t^g$ 分别表示本体感知标记和目标标记，而 e 是位于动作头之前的可学习嵌入。

为实现不同团队规模下的协调，观测智能体的策略会关注一组队友令牌 $\{\mathcal{T}_t^i\}_{i=1}^{N-1}$，每个令牌都编码了另一智能体的线索（例如位置、朝向），并以观测智能体的局部坐标系表示。Transformer 主干网络由 L 组交替排列的自注意力层和交叉注意力层组成。自注意力作用于观测智能体令牌$\mathbf{X}_t$，而交叉注意力使这些令牌能够高效地关注队友令牌，即使团队规模较大也不例外。更新后的嵌入 e 会传入动作头，以预测控制输出 a_t。整体策略定义为 $a_t = \pi_\theta(a_t \mid \mathbf{X}_t,\ \{\mathcal{T}_t^i\}_{i=1}^{N-1})$，其中控制输出对应于比例微分控制器中每个驱动自由度的目标关节旋转。

训练统一策略：在强化学习框架中，每个环境都代表一个独立的仿真实例，智能体在其中与世界交互、观察状态、采取行动并获得奖励。为了学习能够泛化到不同协作场景的单一策略，我们并行实例化多个环境，每个环境配置不同数量的协作智能体及其各自独特的协作奖励。通过这种设置，策略网络在多智能体配置上接受训练，从而接触到不同团队规模下的各种交互动力学。为确保在混合团队配置下训练稳定，我们针对每种团队规模分别归一化 PPO优势。更多详情请参阅补充材料。

掩码 AMP： 将 AMP 扩展到协作式人-物交互的一大挑战，是缺少多人参考动作数据。虽然可以使用单人参考动作，但直接让策略向这些动作靠拢，会限制可能涌现的协作行为的多样性，因为协作任务通常需要比单个示范者所展示的更丰富的运动技能。

为解决这一限制，我们提出一种掩码 *AMP*策略，在保持风格真实性的同时，允许出现多样的人-物交互行为。具体而言，我们训练两个判别器网络：一个全身 AMP 网络 D_{full}，用于评估完整的参考动作；另一个掩码 AMP 网络D_{mask}，用于排除直接与物体交互的身体部位（例如手和前臂）。在物体交互过程中，风格奖励 r_t^{mask}由 D_{mask}得出；而在人形机器人未与物体交互时，则采用来自 D_{full} 的 r_t^{full}。混合后的整体风格奖励为：

$$r_t^{\text{style}} = \sigma(\alpha_t)\, r_t^{\text{mask}} + (1 - \sigma(\alpha_t))\, r_t^{\text{full}}, \tag{4}$$

其中，σ 是作用于连续交互指示量 α_t（例如智能体与物体的距离）的 Sigmoid 函数。

我们的表述与分段运动先验（Pmp）框架在概念上有相似之处 [1]，，该框架从不同身体部位组合运动技能，以丰富物理交互。然而，与直接学习分段先验的 Pmp 不同，我们的方法通过任务奖励，在被遮蔽区域中促使动作呈现多样性，从而利用有限的单身体参考数据实现灵活的物体交互。

3.3. 协作搬运任务

作为协作人-物交互框架的测试平台，我们设计了一项协作搬运任务，要求多个智能体之间实现以物理交互为基础的协调。如图 1 所示，多个人形智能体必须共同与一个大型物体进行交互并将其运走，具体来说，是搬运一张几何形状各异（如圆形、长方形、正方形）的桌子。该任务包含多个连续阶段，共同考验智能体的协作能力：1). 协调队形：智能体从随机设定的初始位置出发，必须自主前往桌子所在位置。与 Coo HOI 假定每个智能体的手部目标位置由预先指定不同，我们的设置不提供明确的协调分配方案。相反，我们会沿桌面周界的下边缘均匀采样 64 个候选接触点，智能体必须据此自行推断合适的位置，以确保抬升过程中的稳定性。2). 协作运输：智能体共同将桌子搬运至指定的目标位置，然后将其放下。

为完成此任务，我们的整体奖励函数由以下几个部分组成：走向物体、接触、抬升、运输和放下——其详细表达式见补充材料。

3.3.1. 队形奖励

这项任务成功的关键在于智能体之间如何协调，在走向物体的同时扩散成有利于稳定抬升的编队。

角度分散奖励： 为此，我们引入了角度分散奖励。它提供了连续的学习信号，促使智能体围绕桌子均匀分布，从而提供稳定的支撑。对于每个智能体，我们找到其最近的左侧和右侧智能体，并计算其相对于桌子中心与这两个邻居之间的二维角度间隔，分别记为 $\Delta\phi_i^{\text{ccw}}$和 $\Delta\phi_i^{\text{cw}}$。对于 m 个协作智能体，理想间距为 $2\pi/m$。奖励函数随后表示为：

$$r_{\text{ang}} = \exp\Big(-k_\theta \tfrac{1}{2}\big[(\Delta\phi_i^{\text{ccw}} - \tfrac{2\pi}{m})^2 + (\Delta\phi_i^{\text{cw}} - \tfrac{2\pi}{m})^2\big]\Big). \quad (5)$$

主轴覆盖奖励： 人类最自然的移动方式是向前或向后行走，并采用对称步态；也可以通过协调的侧向步向侧方移动。他们更倾向于沿自身局部身体轴方向移动，而非沿对角线方向。在协作运输中，这种倾向体现为智能体会采用相应队形，以最大限度地沿物体的主轴提供支撑轴（即其旋转稳定性的自然轴），并沿这些方向行走。

角度跨度奖励虽然能有效促进平衡覆盖以实现稳定抬升，但并不能强制这些编队。因此，我们引入一项额外奖励，用于衡量智能体的支撑区域跨越

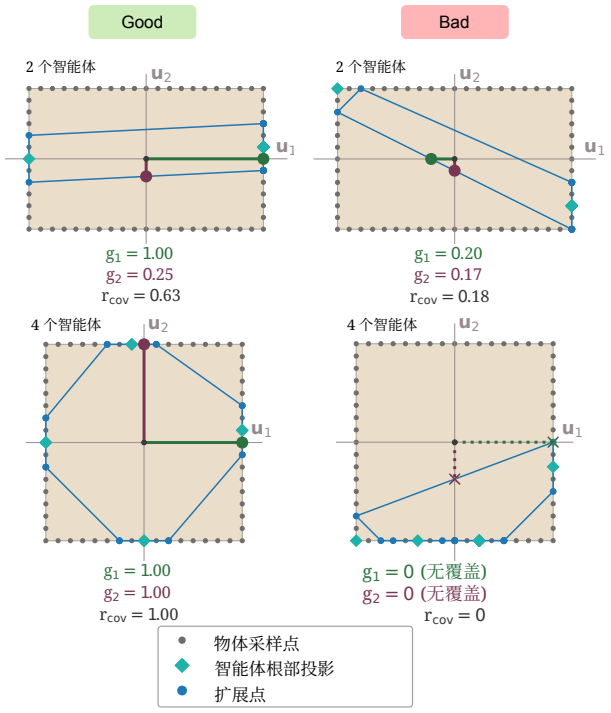

图3. 主轴覆盖奖励示意图。

物体主轴的程度，如图 3所示。首先，将每个智能体的根位置投影到沿物体周界最近的采样点。这组投影点通过凸包定义一个支撑多边形。为避免两个智能体只能形成一条直线的退化情形，每个投影点被替换为沿周界位于其左右两侧的两个点（索引 ±2）。令 c表示桌子在 x、y 平面中的质心。我们首先计算以 $\mathbf{u}_1$,$\mathbf{u}_2$ 为中心、位于 c 的主轴，并测量从 c到支撑多边形沿每条轴两个方向的距离，(d_i^+, d_i^-)。当支撑多边形未跨越桌子质心（完全位于一侧）时，距离变为负值，表明覆盖不平衡或位于外侧。因此，我们将它们裁剪为 $\tilde{d}_i^{\pm} = \max(0, d_i^{\pm})$。以 (ℓ_i^+, ℓ_i^-) 表示从桌子中心沿相同轴到桌边界的距离，逐轴覆盖为：

$$g_i = \min\left(\frac{\tilde{d}_i^+}{\ell_i^+}, \frac{\tilde{d}_i^-}{\ell_i^-}\right), \quad i \in \{1, 2\}, \qquad (6)$$

由此得到的奖励为 $r_{\text{cov}} = \frac{1}{2}(g_1 + g_2)$。在补充材料中，我们提供了 r_{cov} 的广义形式，支持不规则几何形状和非均匀质量分布。

r_{cov} 和 r_{ang} 的设计互为补充。虽然r_{cov} 在不规则几何和质量分布下仍然有效，但当智能体聚集在物体的一侧时，奖励信号可能会很稀疏（导致奖励为零）。相比之下，r_{ang} 则

提供连续信号，以鼓励智能体尽早分散，并使 r_{cov} 变为非零。最终的队形奖励结合了这两种效果，并更加侧重于 r_{cov}：

$$r_{\text{form}} = 0.25\, r_{\text{ang}} + 0.75\, r_{\text{cov}}. \tag{7}$$

4. 实验

4.1. 实现细节

在此，我们描述本框架的关键实现细节，包括针对第3.3节所述协作搬运任务定制的观测状态、架构和数据集。更多实现细节见补充材料。

观测状态：每个观测智能体接收以下观测分量，其中每个分量均在该智能体的局部坐标系中表示：

- **自身本体感知** $\in \mathbb{R}^{223}$：观测智能体的关节状态与根节点运动学，与标准AMP中一致。
- **物体中心** $\in \mathbb{R}^3$：桌子中心的三维位置。
- **候选接触点** $\in \mathbb{R}^{64\times 3}$：沿桌子下边缘的周界均匀采样 64 个点。第一个索引对应离智能体根部最近的点，其余点按逆时针顺序排列。
- **最近手物接触点** $\in \mathbb{R}^{2\times 3}$：分别是最靠近智能体两只手的候选接触点。
- **目标物体位置** $\in \mathbb{R}^3$：x，y 桌子中心的目标位置，以及目标高度 z 或用于指示抬升还是放下的二元标记。
- **队友提示** $\in \mathbb{R}^{(n-1)\times 9}$：对于每位 $n-1$ 队友，包括其 x，y 根位置（$\mathbb{R}^2$）、朝向（六维旋转表示，$\mathbb{R}^6$），以及观测智能体的根部与队友根部在水平面上相对于桌子中心的夹角（$\mathbb{R}^1$）。这些提示会进一步编码为队友令牌。

架构： 策略网络和评论家网络使用相同的观测状态和 Transformer 主干网络，仅最终输出不同。每个观测组件首先通过独立的三层多层感知机分词器编码为一个 64 维令牌，其隐藏层大小为$[256, 128, 64]$。输入令牌包括 64 维可学习嵌入 e、自身本体感知令牌、物体令牌（结合物体中心、候选接触点和最近手物接触点）、目标位置标记，以及数量可变的队友令牌。

Transformer由三组交替排列的自注意力层和交叉注意力层组成，每层包含两个注意力头和一个512维前馈模块。更新后的嵌入 e 经过一个隐藏层大小为 $[1024, 512, 28]$ 的MLP，以预测策略网络中用于比例微分控制的目标关节旋转；而评论家则使用 $[1024, 512, 1]$。两个风格判别器 D_{full} 和 D_{mask} 均按照标准AMP实现为隐藏层大小为 $[1024, 512, 1]$ 的MLP。

人形机器人和物体模型：我们采用MuJoCo人形机器人模型，并简化其手部结构，使用没有手指的球形手掌。我们为协同搬运实验中使用的三种桌子几何形状（正方形、长方形和圆形）设计了URDF模型。桌子质量根据形状不同，介于50至70千克之间。正方形桌子的尺寸为1.60 m $\times$ 1.60 m，长方形桌子的尺寸为2.00 m $\times$ 1.20 m，圆形桌子的直径为2.00 m。桌面高度固定为0.82米，略低于人形机器人默认站立姿势时的手部位置。这一设置使我们的实验能够聚焦于多智能体协调，而非复杂的接触交互。

参考运动：我们的参考运动来自AMASS数据集 [38]。遵循CooHOI，我们采用ACCAD子集中的9个行走相关动作及其时间反转版本用于后退行走，以及来自CMU子集的3个侧向行走动作。为使上半身能够向桌子方向降低并再抬起，我们使用ACCAD子集中的3个拾取动作。这些序列在达到过低姿态之前被裁剪，然后反转以生成相应的抬升动作。

4.2. 评估

仿真设置：我们使用TeamHOI框架训练一个面向2–8个智能体的统一策略。在每个回合开始时，智能体被随机初始化在以物体为中心、半径为8 m的圆上，目标位置则放置在距离桌子中心 $[3, 10]$ m、方向随机的位置。每个评估回合运行600个仿真时间步。

基线方法：为了获得非平凡的基线，我们对CooHOI [7]框架进行了大幅改造，记为CooHOI*。首先，我们引入掩码AMP策略，以使多智能体搬桌具备多样且真实的运动技能。其次，我们设计了专门的奖励函数，使任意智能体都能从任意初始位置到达指定接触点，且不与桌子发生碰撞。

训练遵循两阶段流程。在第一阶段，训练单智能体以掌握基础技能：接近给定接触点、抬升桌子，并将其推向或拖向目标。在第二阶段，同一策略被扩展到多智能体场景，其中协调通过物体动力学隐式涌现。我们训练三个变体，记为 CooHOI*-n，其中 $n \in \{2, 4, 8\}$ 表示协作智能体的数量。在每种配置中，智能体被分配固定接触点，这些接触点经人工选择，以沿物体主轴提供最大覆盖。需要注意的是，该设置消除了原始任务中对协调队形的需求，因为智能体由 oracle 定义的接触点引导。因此，我们不对 CooHOI* 施加队形奖励，而其他奖励与我们的方法保持一致。CooHOI* 的更多细节见补充材料。

表1. **不同团队规模（2A、4A、8A）下的定量比较。** 我们的方法使用单一统一策略，在所有设置中均持续取得较高的成功率、集体协作和运动平滑度。与智能体编队为预定义的 CooHOI* 基线方法不同，我们的智能体必须推断协作方式，以自主建立稳定编队，从而使协调要求更为苛刻。在重载设置（5× 桌子重量）下，只有我们的方法展现出八个智能体之间的有效协作。所有结果均在 10,000 个仿真回合上取平均。

模型	智能体编队	成功率（%）↑ / d (m)↓			t_{coop} （%）↑			$\lvert J\rvert$ (米/秒^3) ↓			成功率$_{5\times}$（%）↑ / $d_{5\times}$ (m)↓	
		2A	4A	8A	2A	4A	8A	2A	4A	8A	4A	8A
CooHOI*-2	预定义	97.5 / 0.19	73.2 / 1.49	10.1 / 4.25	90.3	54.6	1.0	48.3	85.0	189.9	0.0 / 6.38	0.4 / 6.00
CooHOI*-4	预定义	95.5 / 0.18	94.5 / 0.36	61.5 / 1.83	**96.0**	92.1	27.2	40.7	38.6	96.7	1.2 / 5.34	14.2 / 4.26
CooHOI*-8	预定义	29.4 / 3.60	52.4 / 2.68	42.2 / 3.83	93.8	93.6	81.6	**39.5**	**36.5**	45.2	0.1 / 6.22	4.1 / 5.73
Ours	学习到的协作	**99.1/ 0.06**	**99.2/ 0.08**	**97.5 / 0.18**	95.2	**96.1**	**90.1**	51.0	44.7	**34.2**	**3.5 /4.78**	**81.1 /0.49**

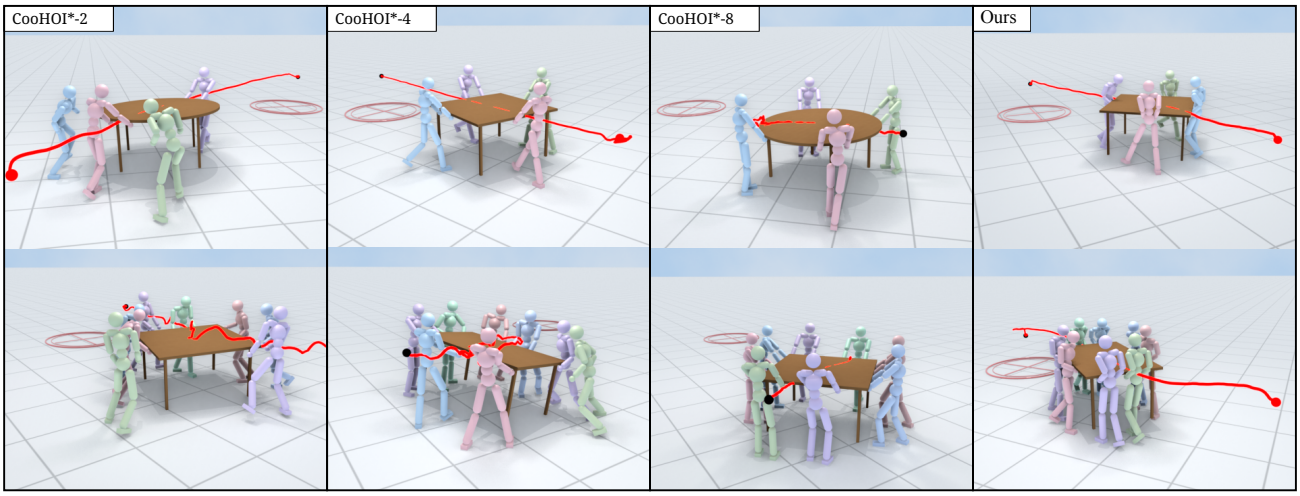

图4. **4智能体（上）和8智能体（下）配置的定性比较。** 我们的方法在两种情况下都能实现同步且稳定的团队协作，而 CooHOI* 基线则表现出有限或无效的协作。红线表示桌子的运动轨迹，黑点标记每个回合结束时桌子的最终位置。

指标： 我们使用四项定量指标评估协作搬运任务，这些指标刻画任务成功、协作质量与运动平滑度：

- **成功率（SR）：** 成功完成任务的回合占比，其中物体与目标的距离达到0.03米。越高越好。
- **到目标的距离（d）：** 回合结束时桌子中心与目标位置之间的欧几里得距离（单位：米）。对于成功回合，由于放下后不再施加奖励，且基线模型会出现不稳定的运动， d 报告为0.03。越低越好。
- **协同时间比率（$t_{\text{协同}}$）：** 运输时长中所有智能体均与64个接触点中的任意一个保持接触的时间占比，表明各方持续作出贡献。越高越好。
- **平均绝对加加速度（$|J|$）：** 64 条接触点轨迹的三阶导数平均幅值，用于衡量运输运动的平滑度。数值越低越好。

结果： 我们在包含 2、4 和 8 个协作智能体的协作搬运任务上测试每个模型。表 1 给出了每种协作场景下对 10,000 次仿真取平均的定量结果。我们的方法在所有配置中始终取得较高的成功率、集体协作以及平滑的运动，证明了由我们的框架得到的单一统一策略的有效性。相比之下，CooHOI* 基线方法对其训练时所针对的特定团队规模表现出强烈依赖。CooHOI*-2 仅在两个智能体时表现良好，但扩展到更大团队时无法展现协调行为。同样，CooHOI*-4 在最多四个智能体时仍能保持良好表现，但超出该配置后性能急剧下降，而 CooHOI*-8 即使在其自身设置下也难以建立有效协调。我们进一步在重载设置（5× 桌子重量）下评估这些模型。该任务对较小团队而言过于困难，它们几乎无法抬升桌子。值得注意的是，只有我们的方法在八个智能体之间展现出有意义的协作，它们共同应对增加的负载并取得较高的成功率。

图4 中的定性结果进一步凸显了

我们的方法与 CooHOI* 基线方法在行为上的差异。CooHOI*-2 中的智能体对表现出相互竞争的行为，各对智能体试图独立移动桌子，导致运动不均匀且频繁失去接触。CooHOI*-4 在规模更大的团队中未能实现同步协作，造成运动不稳定。CooHOI*-8 在将桌子移向目标时难以有效协调，常常导致智能体之间产生冲突力。相比之下，我们的方法实现了全局协调一致的运动：所有智能体协同组成统一团队，抬起、稳定并运输物体。这些结果表明，借助在我们的框架中训练得到的统一去中心化策略，可以实现不同团队规模和协调需求下的可扩展协作。补充材料中还展示了更全面的实验结果，包括对未见过的设置的鲁棒性，以及向 16智能体配置的零样本泛化。

4.3. 消融研究

为了更好地理解掩码AMP策略和所提出的队形奖励各自的贡献，我们进行了消融研究，以检验每个组成部分的效果。

掩码AMP： 如图5所示，引入掩码AMP显著提高了抬升阶段的总体成功率。通过在判别器更新期间屏蔽与物体交互的身体部位，策略学会通过任务奖励来控制手与物体的交互，而不再受过度受限的单人全身参考动作的限制。如果不进行屏蔽，运动真实性与物体交互之间相互冲突的目标往往会导致抬升过程中的协调失败。掩码AMP通过实现更加多样化的手与物体的交互，缓解了这一问题。在后续阶段中，即使仅依赖单智能体参考动作，它也能使策略学习更加多样的协调模式，例如在搬运桌子时向不同方向行走或迈步。

队形奖励： 图 6比较了使用和不使用所提出的主轴覆盖奖励训练的策略。当策略仅使用角度分散奖励进行训练时，智能体并不总是沿着物体的主轴进行分布。在训练中期，这种设置经常会产生绕垂直轴过度旋转的物体轨迹。该策略最终趋于稳定

但会收敛到与物体主轴不对齐的不自然对角迈步模式。这些不自然的迈步模式在不同场景和团队规模下都会持续出现。

相比之下，引入主轴覆盖奖励会促使智能体沿物体旋转稳定性的自然轴对齐其编队。因此，团队学会以自然的对称步态和平衡支撑沿协调一致的方向行走。

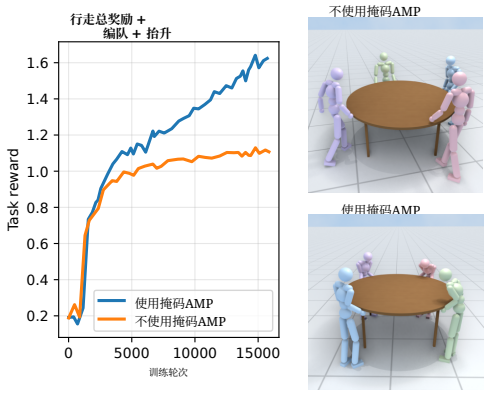

图5. **掩码AMP策略的消融实验。** 比较使用与不使用掩码AMP训练的模型，结果表明应用掩码后任务奖励得到提升，并能成功实现手与物体的交互。

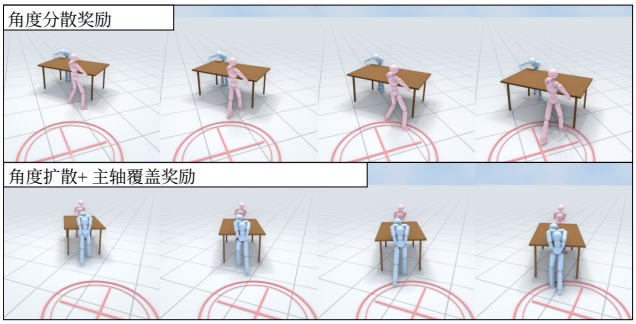

图 6. **编队奖励比较。** 添加主轴覆盖奖励后，形成了与物体主轴对齐的稳定编队，有助于习得自然运动能力。

5. 结论

我们提出了 TeamHOI，这是一个用于可扩展协作人-物交互的统一框架，使单一的去中心化策略能够泛化到不同的团队规模和物体配置。通过引入基于 Transformer 的策略架构，并使用队友令牌，我们的方法能够以可扩展的方式高效地融入队友提示，从而实现不同团队规模下有效的智能体间协调。掩码AMP策略拓展了从单人参考数据中可实现的运动多样性，而主轴覆盖奖励则鼓励协作运输过程中形成稳定、自然的编队。通过在协作搬运任务中开展全面实验，我们证明 TeamHOI 在广泛的多智能体场景下实现了连贯、稳定且多样的协调行为。我们认为，这项工作为可扩展的、基于物理的人形机器人控制奠定了基础，并为虚拟环境中的具身智能和多角色动画开辟了新的机遇。

致谢。本研究由新加坡国家研究基金会（NRF）通过其 NRF研究员计划（奖励编号：NRF-NRFI09-0008）以及新加坡教育部的 Tier 2 项目 MOET2EP20124-0015 资助。

参考文献

[1] Jinseok Bae、Jungdam Won、Donggeun Lim、Cheol-Hui Min 和 Young Min Kim。Pmp：利用分部位运动先验学习与环境进行物理交互。发表于 ACM SIGGRAPH, 2023。3、4[2] Jona Braun、Sammy Christen、Muhammed Kocabas、Emre Aksan 和 Otmar Hilliges。物理上合理的全身手物交互合成。发表于 3D 视觉国际会议（3DV），2024。3[3] Yu-Wei Chao、Jimei Yang、Weifeng Chen 和 Jia Deng。学习坐下：通过分层控制合成人椅交互。发表于 AAAI 人工智能会议，2021。3[4] Zhiyang Dou、Xuelin Chen、Qingnan Fan、Taku Komura 和 Wenping Wang。C·ase：为基于物理的角色学习条件对抗技能嵌入。发表于 ACM SIGGRAPH Asia, 2023。3[5] Ke Fan、Junshu Tang、Weijian Cao、Ran Yi、Moran Li、Jingyu Gong、Jiangning Zhang、Yabiao Wang、Chengjie Wang 和 Lizhuang Ma。Freemotion：用于无数量限制文本到动作合成的统一框架。发表于欧洲计算机视觉会议（ECCV），2024。3[6] Mihai Fieraru、Mihai Zanfir、Elisabeta Oneata、Alin-Ionut Popa、Vlad Olaru 和 Cristian Sminchisescu。人类交互的三维重建。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2020。3[7] Jiawei Gao、Ziqin Wang、Zeqi Xiao、Jingbo Wang、Tai Wang、Jinkun Cao、Xiaolin Hu、Si Liu、Jifeng Dai 和 Jiangmiao Pang。CooHOI：学习结合受操控物体动力学的协作式人物交互。神经信息处理系统进展（NeurIPS），2024。2、3、6、16[8] Anindita Ghosh、Rishabh Dabral、Vladislav Golyanik、Christian Theobalt 和 Philipp Slusallek。Remos：面向双人交互的三维动作条件反应合成。发表于欧洲计算机视觉会议（ECCV），2024。3[9] Anindita Ghosh、Bing Zhou、Rishabh Dabral、Jian Wang、Vladislav Golyanik、Christian Theobalt、Philipp Slusallek 和 Chuan Guo。Duetgen：通过分层掩码建模生成音乐驱动的双人舞蹈。发表于 ACM SIGGRAPH，2025。3[10]Zhaoyuan Gu、Junheng Li、Wenlan Shen、Wenhao Yu、Zhaoming Xie、Stephen McCrory、Xianyi Cheng、Abdulaziz Shamsah、Robert Griffin、C Karen Liu 等。人形机器人运动与操作：控制、规划与学习的当前进展和挑战。arXiv 预印本 arXiv:2501.02116，2025。2[11] Mohamed Hassan、Yunrong Guo、Tingwu Wang、Michael Black、Sanja Fidler 和 Xue Bin Peng。合成物理角色与场景交互。在ACM SIGGRAPH，2023。3[12] Brandon Haworth、Glen Berseth、Seonghyeon Moon、Petros Faloutsos和Mubbasir Kapadia。物理人形机器人控制与人群导航的深度融合。在ACM SIGGRAPH，2020。3[13] Tairan He、Wenli Xiao、Toru Lin、Zhengyi Luo、Zhenjia Xu、Zhenyu Jiang、Jan Kautz、Changliu Liu、Guanya Shi、Xiaolong Wang等。Hover：面向人形机器人的多功能神经全身控制器。在机器人与自动化国际会议（ICRA），2025。3[14] Xiaoyu Huang、Takara Truong、Yunbo Zhang、Fangzhou Yu、Jean Pierre Sleiman、Jessica Hodgins、Koushil Sreenath和Farbod Farshidian。Diffuse-cloc：用于基于物理的角色前瞻控制的引导扩散。ACM图形学汇刊（TOG），2025。3[15] Yiming Huang、Zhiyang Dou和Lingjie Liu。Modskill：物理角色技能模块化。在IEEE/CVF国际计算机视觉会议（ICCV），2025。3[16] Muhammad Gohar Javed、Chuan Guo、Li Cheng和Xingyu Li。Intermask：通过协作式掩码建模实现三维人体交互生成。在国际学习表征会议（ICLR），2025。3[17] Kaiyang Ji、Ye Shi、Zichen Jin、Kangyi Chen、Lan Xu、Yuexin Ma、Jingyi Yu和Jingya Wang。迈向沉浸式人类-X交互：一种用于合成符合物理规律的运动的实时框架。在IEEE/CVF国际计算机视觉会议（ICCV），2025。3[18] Jordan Juravsky、Yunrong Guo、Sanja Fidler和Xue Bin Peng。PADL：语言指导的基于物理的角色控制。在ACM SIGGRAPH Asia，2022。3[19] Korrawe Karunratanakul、Konpat Preechakul、Supasorn Suwajanakorn和Siyu Tang。用于可控人体运动合成的引导运动扩散。在IEEE/CVF国际计算机视觉会议（ICCV），2023。3[20] Diederik P Kingma和Jimmy Ba。Adam：一种随机优化方法。在国际学习表征会议（ICLR），2015。16[21] Kaen Kogashi、Anoop Cherian和Meng-Yu Jennifer Kuo。MMHOI：复杂三维多人与多物体交互建模。在IEEE/CVF冬季计算机视觉应用会议（WACV），2026。3[22]Jogendra Nath Kundu、Himanshu Buckchash、Priyanka Mandikal、Anirudh Jamkhandi、Venkatesh Babu Radhakrishnan等。用于长期合成人与人之间的人体运动交互的交叉条件循环网络。在IEEE/CVF冬季计算机视觉应用会议（WACV），2020。3[23] Nhat Le、Thang Pham、Tuong Do、Erman Tjiputra、Quang D Tran和Anh Nguyen。音乐驱动的群体编舞。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2023。3[24] Bin Li、Ruichi Zhang、Han Liang、Jingyan Zhang、Juze Zhang、Xin Chen、Lan Xu、Jingyi Yu和Jingya Wang。Interagent：通过交互图上的扩散实现基于物理的多智能体指令执行。arXiv预印本 arXiv:2512.07410，2025。3

[25] Jianan Li、Tao Huang、Qingxu Zhu和Tien-Tsin Wong。根据人体运动生成基于物理的场景布局。在ACM SIGGRAPH，2024。3[26] Yitang Li、Mingxian Lin、Zhuo Lin、Yipeng Deng、Yue Cao和Li Yi。利用简短的行走参考学习基于物理的全身人体伸手与抓取。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2025。3[27] Han Liang、Wenqian Zhang、Wenxuan Li、Jingyi Yu和Lan Xu。Intergen：复杂交互条件下基于扩散的多人运动生成。国际计算机视觉杂志（IJCV），2024。3[28] Libin Liu和Jessica Hodgins。利用深度Q学习为基于物理的角色学习控制片段调度。ACM图形学汇刊（TOG），

2017 3。

[29]

Libin Liu和Jessica Hodgins。利用轨迹优化和深度强化学习学习篮球运球技能。ACM图形学汇刊（TOG），2018。3[30] Libin Liu、KangKang Yin和Baining Guo。改进基于采样的运动控制。在《计算机图形论坛》，2015。3[31] Yunze Liu、Changxi Chen、Chenjing Ding和Li Yi。Phys-reaction：通过正向动力学引导的4D模仿，实现符合物理规律的实时类人反应合成。在ACM多媒体国际会议，2024。3[32] Yun Liu、Bowen Yang、Licheng Zhong、He Wang和Li Yi。Mimicking-bench：一个通过人体模仿学习可泛化的类人与场景交互的基准测试。arXiv预印本 arXiv:2412.17730，2024。3[33] Yun Liu、Chengwen Zhang、Ruofan Xing、Bingda Tang、Bowen Yang和Li Yi。Core4d：用于协作式物体重排的4D人-物-人交互数据集。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2025。3[34] Zhengyi Luo、Jinkun Cao、Kris Kitani、Weipeng Xu等。面向实时模拟化身的持续人形机器人控制。在IEEE/CVF国际计算机视觉会议（ICCV），2023。3[35] Zhengyi Luo、Jinkun Cao、Sammy Christen、Alexander Winkler、Kris Kitani和Weipeng Xu。Omnigrasp：利用模拟类人角色抓取多种物体。神经信息处理系统进展（NeurIPS），2024。3[36] Zhengyi Luo、Jinkun Cao、Josh Merel、Alexander Winkler、Jing Huang、Kris Kitani和Weipeng Xu。用于基于物理的控制的通用人形机器人运动表征。在国际学习表征会议（ICLR），2024。3[37] Zhengyi Luo、Jiashun Wang、Kangni Liu、Haotian Zhang、Chen Tessler、Jingbo Wang、Ye Yuan、Jinkun Cao、Zihui Lin、Fengyi Wang等。SMPLOlympics：面向物理模拟人形机器人的体育环境。arXiv预印本 arXiv:2407.00187，2024。2，3[38] Naureen Mahmood、Nima Ghorbani、Nikolaus F Troje、Gerard Pons-Moll和Michael J Black。AMASS：以表面形状为形式的动作捕捉档案。在IEEE/CVF国际计算机视觉会议（ICCV），2019。6[39] Viktor Makoviychuk、Lukasz Wawrzyniak、Yunrong Guo、Michelle Lu、Kier Storey、Miles Macklin、David Hoeller，

Nikita Rudin, Arthur Allshire, Ankur Handa 等。Isaac Gym：用于机器人学习的高性能基于GPU的物理仿真。arXiv预印本 arXiv:2108.10470，2021。2，13[40] Josh Merel, Saran Tunyasuvunakool, Arun Ahuja, Yuval Tassa, Leonard Hasenclever, Vu Pham, Tom Erez, Greg Wayne 和 Nicolas Heess。Catch & Carry：用于视觉引导全身任务的可复用神经控制器。ACM图形学汇刊（TOG），2020。3[41] Liang Pan, Jingbo Wang, Buzhen Huang, Junyu Zhang, Haofan Wang, Xu Tang 和 Yangang Wang。在三维场景中合成符合物理规律的人体运动。载于国际三维视觉会议（3DV），2024。3[42] 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生成物理仿真人体场景交互。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），2025。3，4[43] Xue Bin Peng, Pieter Abbeel, Sergey Levine 和 Michiel van de Panne。DeepMimic：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），2018。3[44] 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），2021。3[45] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine 和 Sanja Fidler。ASE：面向物理仿真角色的大规模可复用对抗性技能嵌入。ACM图形学汇刊（TOG），2022。3[46] Davis Rempe, Zhengyi Luo, Xue Bin Peng, Ye Yuan, Kris Kitani, Karsten Kreis, Sanja Fidler 和 Or Litany。Trace and Pace：通过引导轨迹扩散实现可控行人动画。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），2023。3[47] 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。3，13[48] Agon Serifi, Ruben Grandia, Espen Knoop, Markus Gross 和 Moritz B¨ acher。VMP：用于稳健跟踪物理角色运动的通用运动先验。载于计算机图形学论坛，2024。3[49] Yonatan Shafir, Guy Tevet, Roy Kapon 和 Amit H Bermano。作为生成式先验的人体运动扩散。载于国际学习表征会议（ICLR），2024。3[50] Yijun Shen, Longzhi Yang, Edmond SL Ho 和 Hubert PH Shum。基于交互的人类活动比较。IEEE可视化与计算机图形学汇刊（TVCG），2019。3[51] Li Siyao, Tianpei Gu, Zhitao Yang, Zhengyu Lin, Ziwei Liu, Henghui Ding, Lei Yang 和 Chen Change Loy。DuoLando：采用离策略强化学习进行舞蹈伴奏的跟随者GPT。载于国际学习表征会议（ICLR），2024。3

[52] Kewei Sui, Anindita Ghosh, Inwoo Hwang, Bing Zhou, Jian Wang 和 Chuan Guo。人类交互动作生成综述。国际计算机视觉杂志（IJCV），2026。2[53] Wenhui Tan, Boyuan Li, Chuhao Jin, Wenbing Huang, Xiting Wang 和 Ruihua Song。先思考后反应：迈向不受约束的人类动作到反应生成。载于国际学习表征会议（ICLR），2025。3[54] Mikihiro Tanaka 和 Kent Fujiwara。根据文本描述生成角色感知的交互。载于IEEE/CVF国际计算机视觉会议（ICCV），2023。3[55] Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik 和 Xue Bin Peng。Calm：面向可操控虚拟角色的条件对抗潜变量模型。载于ACM SIGGRAPH，2023。3[56] Chen Tessler, Yunrong Guo, Ofir Nabati, Gal Chechik 和 Xue Bin Peng。Maskedmimic：通过掩码动作修复实现统一的基于物理的角色控制。ACM图形学汇刊（TOG），2024。3[57] Chen Tessler, Yifeng Jiang, Erwin Coumans, Zhengyi Luo, Gal Chechik 和 Xue Bin Peng。Maskedmanipulator：通用全身操控。载于ACM SIGGRAPH Asia，2025。3[58] Guy Tevet, Sigal Raab, Setareh Cohan, Daniele Reda, Zhengyi Luo, Xue Bin Peng, Amit H Bermano 和 Michiel van de Panne。Closd：弥合仿真与扩散之间的循环，实现多任务角色控制。载于国际学习表征会议（ICLR），2025。3[59] Emanuel Todorov, Tom Erez 和 Yuval Tassa。MuJoCo：用于基于模型的控制的物理引擎。载于国际智能机器人与系统会议（IROS），2012。2[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser 和 Illia Polosukhin。注意力就是你所需要的一切。神经信息处理系统进展（NeurIPS），2017。3[61] Huayi Wang, Wentao Zhang, Runyi Yu, Tao Huang, Junli Ren, Feiyu Jia, Zirui Wang, Xiaojie Niu, Xiao Chen, Jiahe Chen 等。Physhsi：迈向可泛化至真实世界的自然人形机器人场景交互系统。arXiv预印本 arXiv:2510.11072，2025。2[62] Jiashun Wang, Jessica Hodgins 和 Jungdam Won。基于物理的乒乓球动画中的策略与技能学习。载于ACM SIGGRAPH，2024。3[63] Wenjia Wang, Liang Pan, Zhiyang Dou, Zhouyingcheng Liao, Yuke Lou, Lei Yang, Jingbo Wang 和 Taku Komura。Sims：通过现实世界脚本规划模拟人类场景交互。arXiv预印本 arXiv:2411.19921，2024。3[64] Yinhuai Wang, Qihan Zhao, Runyi Yu, Hok Wai Tsui, Ailing Zeng, Jing Lin, Zhengyi Luo, Jiwen Yu, Xiu Li, Qifeng Chen 等。Skillmimic：从示范中学习篮球交互技能。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），2025。3[65] Jungdam Won 和 Jehee Lee。基于物理的角色中的身体形状变化学习。ACM图形学汇刊（TOG），2019。3[66] Jungdam Won、Kyungho Lee、Carol O'Sullivan、Jessica K Hodgins和Jehee Lee。生成并排序多样化的多角色交互。ACM图形学汇刊（TOG），2014。3[67] Jungdam Won、Deepak Gopinath和Jessica Hodgins。一种可扩展的方法，用于控制物理仿真角色的多样化行为。ACM图形学汇刊（TOG），2020。3[68] Yan Wu、Korrawe Karunratanakul、Zhengyi Luo和Siyu Tang。Uniphys：基于扩散的统一规划器与控制器，用于灵活的基于物理的角色控制。见IEEE/CVF国际计算机视觉会议（ICCV），2025。3[69] Zhen Wu、Jiaman Li、Pei Xu和C Karen Liu。基于人类级指令的人-物交互。见IEEE/CVF国际计算机视觉会议（ICCV），2025。3[70] Zeqi Xiao、Tai Wang、Jingbo Wang、Jinkun Cao、Wenwei Zhang、Bo Dai、Dahua Lin和Jiangmiao Pang。通过提示式接触链实现统一的人-场景交互。见国际学习表征会议（ICLR），2024。3[71] Zhaoming Xie、Jonathan Tseng、Sebastian Starke、Michiel van de Panne和C Karen Liu。面向箱体移动操作的分层规划与控制。ACM计算机图形学与交互技术，2023。3[72] Liang Xu、Xintao Lv、Yichao Yan、Xin Jin、Shuwen Wu、Congsheng Xu、Yifan Liu、Yizhou Zhou、Fengyun Rao、Xingdong Sheng等。Inter-x：迈向通用的人人交互分析。见IEEE/CVF计算机视觉与模式识别会议（CVPR），2024。3[73] Sirui Xu、Hung Yu Ling、Yu-Xiong Wang和Liang-Yan Gui。Intermimic：迈向基于物理的人-物交互的通用全身控制。见IEEE/CVF计算机视觉与模式识别会议（CVPR），2025。3[74] Sirui Xu、Samuel Schulter、Morteza Ziyadi、Xialin He、Xiaohan Fei、Yu-Xiong Wang和Liangyan Gui。Interprior：扩展基于物理的人-物交互的生成式控制。arXiv预印本arXiv:2602.06035，2026。3[75] Runyi Yu、Yinhuai Wang、Qihan Zhao、Hok Wai Tsui、Jingbo Wang、Ping Tan和Qifeng Chen。Skillmimic-v2：从稀疏且含噪的示范中学习稳健且可泛化的交互技能。见ACM SIGGRAPH，2025。3[76] Haotian Zhang、Ye Yuan、Viktor Makoviychuk、Yunrong Guo、Sanja Fidler、Xue Bin Peng和Kayvon Fatahalian。从转播视频中学习物理仿真网球技能。ACM图形学汇刊（TOG），2023。3[77] Juze Zhang、Jingyan Zhang、Zining Song、Zhanhe Shi、Chengfeng Zhao、Ye Shi、Jingyi Yu、Lan Xu和Jingya Wang。Hoi-mˆ3：在上下文环境中采集多人与物体交互。见IEEE/CVF计算机视觉与模式识别会议（CVPR），2024。3[78] Yunbo Zhang、Deepak Gopinath、Yuting Ye、Jessica Hodgins、Greg Turk和Jungdam Won。复杂多角色交互的仿真与重定向。见ACM SIGGRAPH，2023。3

[79] Zihan Zhang、Richard Liu、Rana Hanocka和Kfir Aberman。Tedi：面向长期运动合成的时间纠缠扩散。见*ACMSIGGRAPH*，2024。3[80] Zihang Zhang、Shoulong Zhang、Yan Wang和Shuai Li。Reactffusion：用于反应生成的物理接触引导扩散模型。见*ACM*多媒体国际会议，2025。3[81] Yiwen Zhao、Yang Wang、Liting Wen、Hengyuan Zhang和Xingqun Qi。Freedance：通过统一框架实现和谐的任意人数群体舞蹈生成。见*IEEE/CVF*国际计算机视觉会议（*ICCV*），2025。3

TeamHOI：面向任意团队规模的协作人-物交互统一策略学习

补充材料

6. 使用各种团队规模进行训练

6.1. 团队规模优势归一化

PPO [47]算法计算优势项 A_t，该项衡量某个动作相对于评论家网络估计的同一状态下策略动作期望回报的表现提升程度。优势项根据在有限时域内收集的轨迹计算得出，而有限时域决定了奖励向未来累积的范围。由于不同轨迹中的优势项尺度可能有所不同，通常会在一个轨迹批次中对优势项进行归一化，以稳定训练，具体表示为：

$$A_t \leftarrow \frac{A_t - \mu(A)}{\sigma(A) + \epsilon},$$

其中，$\mu(A)$ 和 $\sigma(A)$ 分别表示在该批次上计算得到的均值和标准差。

在我们的框架中，训练批次可以包含来自不同团队规模的团队的数据，而这些团队产生的奖励具有不同的尺度和方差。对这类异质数据的所有优势统一进行归一化，可能会扭曲它们的相对幅度，并影响策略更新信号的准确性。因此，我们针对每种团队规模分别对优势进行归一化 n：

$$A_t^{(n)} \leftarrow \frac{A_t^{(n)} - \mu_n(A)}{\sigma_n(A) + \epsilon}.$$

如图 7所示，团队规模优势归一化可带来更高的任务奖励。

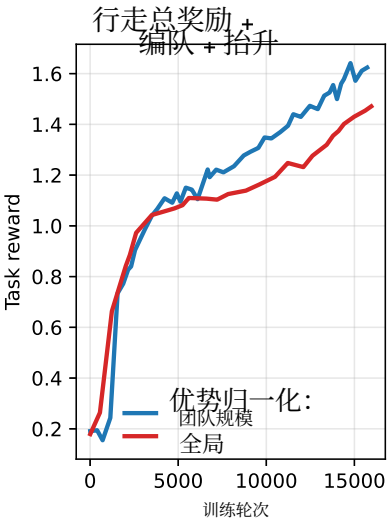

图7. 使用团队规模优势归一化和全局优势归一化训练的模型，其任务奖励曲线对比。

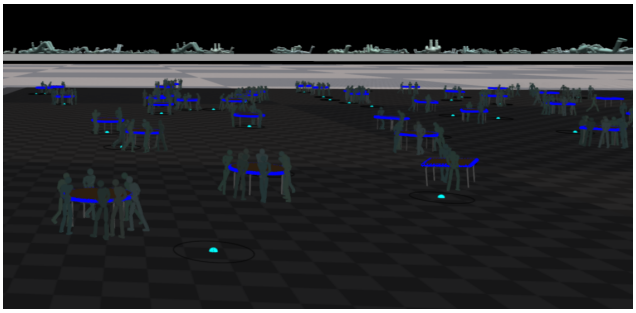

图8. 在 IsaacGym 中使用虚拟天花板搭建的环境，以支持灵活的团队规模训练。额外的智能体会被移至天花板处，并从观测、奖励和梯度更新中排除。

6.2. 环境实例化

我们使用 IsaacGym [39]模拟器训练模型。IsaacGym 当前的一个局限是，并行训练中的每个环境都必须包含相同数量的执行体，包括人形智能体和物体。为了解决这一局限并实现任意团队规模的统一策略训练，我们在每个环境中添加一个虚拟天花板平面，并实例化固定数量的智能体 N。对于所需团队规模较小的环境 n，我们将其余的 $N - n$ 智能体放置在虚拟天花板上。这些智能体不参与奖励计算、观测状态和梯度计算。此外，它们的比例微分控制器会被禁用。图8展示了相关示意图。

7. 奖励函数

这里，我们详细介绍协作搬运任务中使用的所有奖励组成部分，但不包括正文中已介绍的队形奖励 r_{form}、角度分散奖励r_{ang}和主轴覆盖奖励 r_{cov}。

7.1. 走向物体

初始化后，每个人形机器人都会从距离桌子一定距离的位置出发，并被鼓励在抬升阶段开始前走向物体。行走奖励分解为三个部分，分别用于塑造每个智能体的位置、速度和朝向。

位置： 对于每个智能体，我们在桌子周边找到最近的采样点，记为$\mathbf{p}^*$，并计算该点与智能体根部$\mathbf{x}_{\text{root}}$ 在 x-y 平面上的距离，

$d = \|\mathbf{x}_{\text{root}} - \mathbf{p}^*\|_2$。通过惩罚平方偏差 $\Delta_{\text{gap}} = (d - d_{\text{gap}})^2$，鼓励智能体站在距离$\mathbf{p}^*$ 为 $d_{\text{gap}} = 0.3$ 米的目标间距处。位置奖励为：

$$r_{\text{walk}}^{\text{pos}} = \begin{cases} \exp\big(-2.0\Delta_{\text{gap}}\big), & \Delta_{\text{gap}} > 0.04\ \text{m}, \\ 1, & \Delta_{\text{gap}} \le 0.04\ \text{m}. \end{cases} \tag{8}$$

此项作为吸引势，将每个智能体拉向桌子。必须由编队奖 r_{form} 励加以平衡，以确保智能体在仍然朝各自适当的站立区域汇聚、准备抬升的同时彼此分散。

速度： 设$\mathbf{v} \in \mathbb{R}^2$ 为 x，y 根速度。我们将期望行走方向$\mathbf{u}^* \in \mathbb{R}^2$ 定义为与最近周边点$\mathbf{p}^*$对应的 x-y 平面上的向内单位法向量。将速度 s 投影到$\mathbf{u}^*$上，即$\mathbf{u}^{*\top}\mathbf{v}$，即可计算出智能体的方向速度。接着，我们鼓励智能体以 $s^*_{\text{low}} = 1.5\,\text{m/s}$ 至 $s^*_{\text{high}} = 2.5\,\text{m/s}$ 的期望速度范围朝桌子移动。偏离此范围的程度用 ReLU 函数表示：$\delta_{\text{vel}} = \max(0,\ s^*_{\text{low}} - s) + \max(0,\ s - s^*_{\text{high}})$。速度奖励为：

$$r_{\text{walk}}^{\text{vel}} = \begin{cases} 0, & s \le 0, \\ 1, & \Delta_{\text{gap}} \le 0.04\ \text{m}, \\ \exp\big(-2.0\,\delta_{\text{vel}}^2\big), & \text{otherwise}. \end{cases} \tag{9}$$

朝向： 我们通过提取智能体根部朝向中的航向分量，计算其朝向。设$\mathbf{f} \in \mathbb{R}^2$ 为 x-y 平面内的朝向。我们定义两个目标朝向：用于近距离视角对齐的$\mathbf{u}^*$ 处的内法线$\mathbf{p}^*$，以及从智能体指向桌子中心的方向$\mathbf{c}^* = \mathbf{p}_{\text{center}} - \mathbf{x}_{\text{root}} \,/\, \|\mathbf{p}_{\text{center}} - \mathbf{x}_{\text{root}}\|_2$。朝向奖励的计算方式如下：

$$r_{\text{walk}}^{\text{face}} = \begin{cases} \max(0, \mathbf{u}^{*\top}\mathbf{f}), & d \le 1.0\ \text{m}, \\ \max(0, \mathbf{c}^{*\top}\mathbf{f}), & d > 1.0\ \text{m}. \end{cases} \tag{10}$$

7.2. 手部接触准备

当智能体靠近桌子至 $\|\mathbf{x}_{\text{root}} - \mathbf{p}^*\|_2 \le 1.0$ 米后，我们鼓励智能体将双手伸向桌子的接触点，并保持适合抬升的手部构型。

手部触及： 对于每个智能体，设 $\mathbf{h}_j \in \mathbb{R}^3, j \in \{\text{L,R}\}$分别表示左手和右手的位置，$\{\mathbf{q}_k\}_{k=1}^{64}$ 表示 64 个候选接触点。我们首先为每只手找到最近的接触点及其距离，$d_j^{\text{hand}} = \min_k \|\mathbf{h}_j - \mathbf{q}_k\|_2$。邻近项鼓励双手向接触点靠近：

$$r_{\text{prox}} = \frac{1}{2}\sum_j \exp(-5.0 d_j^{\text{hand}}). \tag{11}$$

此外，我们鼓励双手触及桌子的下边缘，而不是漂移到桌面上。对于位置为 $\mathbf{h}_j \in \mathbb{R}^3$ 的每只手 $j \in \{\text{L,R}\}$，设 $\mathbf{p}_j^* \in \mathbb{R}^3$ 为桌子周界上离它最近的采样周边点。我们定义接触方向 $\hat{\mathbf{v}}_j = \mathbf{h}_j - \mathbf{p}_j^* \,/\, \|\mathbf{h}_j - \mathbf{p}_j^*\|_2$。令$\mathbf{e}_z = (0,0,1)^\top$ 表示世界向上方向。我们计算 $\cos\theta_j = \hat{\mathbf{v}}_j^\top \mathbf{e}_z$，以衡量手相对于其对应接触点向上移动的程度。每只手的垂直对齐得分定义为：

$$r_{\text{above},j} = \begin{cases} \exp\big(-3.0\cos\theta_j\big), & \cos\theta_j > 0, \\ 1, & \cos\theta_j \le 0. \end{cases} \tag{12}$$

双手的组合项为：

$$r_{\text{above}} = \frac{1}{2}\big(r_{\text{above,L}} + r_{\text{above,R}}\big). \tag{13}$$

双手间距： 我们还鼓励双手保持目标水平间距。首先，我们计算双手在 x-y 平面内的水平间距 $d_{\text{hand}} = \|(\mathbf{h}_{\text{L}} - \mathbf{h}_{\text{R}})_{xy}\|_2$。我们鼓励双手保持在$d^*_{\text{low}} = 0.4$ 米至 $d^*_{\text{high}} = 0.6$米的理想间距范围内。超出此范围的偏差通过 ReLU 函数表示：$\delta_{\text{sep}} = \max(0, d^*_{\text{low}} - d_{\text{hand}}) + \max(0, d_{\text{hand}} - d^*_{\text{high}})$。然后，我们得到双手间距奖励：

$$r_{\text{sep}} = \exp\big(-5.0\delta_{\text{sep}}^2\big). \tag{14}$$

为鼓励稳定抬升，我们对双手之间的垂直高度差进行惩罚。设 z_{L} 和 z_{R} 为双手的高度，奖励定义如下：

$$r_{\text{same-z}} = \exp\big(-20.0(z_{\text{L}} - z_{\text{R}})^2\big). \tag{15}$$

综合奖励： 手部准备的综合奖励为：

$$r_{\text{hand}} = r_{\text{prox}} \times r_{\text{above}} \times r_{\text{sep}} \times r_{\text{same-z}}, \tag{16}$$

这要求四个项同时满足。

7.3. 接触与抬升

双手放到桌沿附近后，将启用额外奖励，以便智能体与桌子建立接触，并将桌子抬升至所需高度。

接触激活： 令 d_j^{hand}为前文定义的最近手部接触距离。每只手的接触得分计算为 $\gamma_j = \max\Big(0, 1 - \frac{d_j^{\text{hand}}}{0.06\ \text{m}}\Big)$。随后，我们将接触奖励定义为双手各自接触得分中的最小值：

$$r_{\text{contact}} = \min(\gamma_{\text{L}}, \gamma_{\text{R}}), \tag{17}$$

以及每只手的接触指示器：

$$m_j = \begin{cases} 1, & d_j^{\text{hand}} < 0.04\ \text{m}, \\ 0, & \text{otherwise}, \end{cases}$$

该项用于门控后续的抬升和运输奖励。

抬升高度： 建立接触后，双手应将桌子抬升至目标高度。设 $\hat{}z_j$ 为与手 j关联的接触点高度，目标抬升高度为 $z^*_{\text{lift}} = 0.94\text{m}$。我们为每只手计算抬升奖励：

$$\rho_j = \exp(-5.0\,|\hat{z}_j - z^*_{\text{lift}}|). \tag{18}$$

只有接触有效的手才会产生贡献。因此，综合抬升奖励为：

$$r_{\text{lift}} = \frac{1}{2}\big(m_{\text{L}}\rho_{\text{L}} + m_{\text{R}}\rho_{\text{R}}\big). \tag{19}$$

7.4. 集体运输

运输： 所有智能体都用双手与桌子建立接触后，系统会鼓励它们协同搬动物体前往目标位置。设$\mathbf{x}_{\text{obj}} \in \mathbb{R}^2$为$x$，$y$ 桌子的位置，$\mathbf{x}_{\text{tar}} \in \mathbb{R}^2$ 为目标位置。我们将运输奖励定义为：

$$r_{\text{transport}} = \begin{cases} \exp\Big(-0.15\,\|\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}\|_2^2\Big), & m_{\text{L}} = m_{\text{R}} = 1 \\ & \text{for all agents}, \\ 0, & \text{otherwise}. \end{cases} \tag{20}$$

搬运对齐： 虽然运输并不严格要求这一点，但我们加入了搬运对齐奖励，鼓励至少一个智能体在搬运过程中朝向目标方向。我们将距离目标最远的智能体确定为该智能体。设$\mathbf{f} \in \mathbb{R}^2$ 为该智能体在 x-y 平面内的朝向，期望的运输方向为：

$$\mathbf{u}_{\text{tar}} = \frac{\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}}{\|\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}\|_2}.$$

我们计算对齐奖励：

$$r_{\text{align}} = \begin{cases} \max\big(0,\, \mathbf{u}_{\text{tar}}^{\top}\mathbf{f}\big), & m_{\text{L}} = m_{\text{R}} = 1 \text{ for all agents} \\ & \text{and } \|\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}\|_2 \geq 0.5\ \text{m}, \\ 1, & m_{\text{L}} = m_{\text{R}} = 1 \text{ for all agents} \\ & \text{and } \|\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}\|_2 < 0.5\ \text{m}, \\ 0, & \text{otherwise}. \end{cases} \tag{21}$$

$r_{\text{transport}}$ 和 r_{align} 由所有智能体共享。在运输过程中，当所有智能体的 $m_{\text{L}} = m_{\text{R}} = 1$时，我们设置 $r^{\text{face}}_{\text{walk}} = 1.0$，以便智能体在共同搬运物体时能够灵活调整朝向。

7.5. 放下

物体到达目标位置后，智能体必须放下桌子并松开双手。因此，当物体到达目标位置时，我们引入放下奖励：$\|\mathbf{x}_{\text{tar}} - \mathbf{x}_{\text{obj}}\|_2 < 0.03$ 米。

手部释放： 令 z_j 为手部 $j \in \{\text{L}, \text{R}\}$的高度，$z^*_{\text{put}} = 0.65$ m为放下过程中的目标手部高度。令 d_j^{hand}为前文所定义的最近手部–桌子距离。我们计算手部释放奖励如下：

$$r^{\text{release}}_{\text{put}} = \begin{cases} 1, & d^{\text{hand}}_{\text{L}} > 0.07\ \text{m} \\ & \text{and } d^{\text{hand}}_{\text{R}} > 0.07\ \text{m}, \\ \min\limits_{j\in\{\text{L},\text{R}\}} \exp\big(-5.0\,|z_j - z^*_{\text{put}}|\big), & \text{otherwise}. \end{cases} \tag{22}$$

零速度： 在放下过程中，我们还通过施加以下奖励来鼓励智能体停止移动：

$$r^{\text{vel}}_{\text{put}} = \exp\big(-2\,\|\mathbf{v}\|_2\big), \tag{23}$$

其中$\mathbf{v}$ 是智能体的 x，y 根速度。

综合奖励： 最终放下奖励是手部释放项与零速度项的加权组合：

$$r_{\text{put}} = 0.8\, r^{\text{release}}_{\text{put}} \;+\; 0.2\, r^{\text{vel}}_{\text{put}}. \tag{24}$$

7.6. 总任务奖励

协作搬运任务的任务奖励按如下方式聚合：

$$\begin{aligned} r^{\text{task}} = {} & 0.2\, r^{\text{pos}}_{\text{walk}} + 0.4\, r^{\text{vel}}_{\text{walk}} + 0.2\,\sqrt{r^{\text{face}}_{\text{walk}}\, r_{\text{ang}}} + 0.6\, r_{\text{form}} \\ & + 0.7\,(r_{\text{hand}} r_{\text{cov}}) + 0.7\, r_{\text{contact}} + 0.7\,(r_{\text{lift}} r_{\text{cov}}) \\ & + 1.0\, r_{\text{transport}} + 0.4\, r_{\text{align}} + 1.0\, r_{\text{put}}. \end{aligned} \tag{25}$$

8. 广义主轴覆盖奖励

我们详细阐述各个组成部分，以得到广义主轴覆盖奖励 r_{cov}，该奖励适用于不规则几何形状（包括 L形等凹形）和非均匀质量分布。

质心： 设 $\mathcal{X} = \{\mathbf{x}_k \in \mathbb{R}^2\}_{k=1}^N$ 表示从物体的 x–y 平面（例如桌面）采样得到的一组二维点。每个点可以选择携带一个质量权重 $w_k > 0$，用于表示局部密度。均匀密度对应于 $w_k = 1$。平面质心计算如下：$\mathbf{c} = \frac{\sum_{k=1}^N w_k \mathbf{x}}{\sum_{k=1}^N w_k}$.

主轴： 接下来，我们从真实对称物体的

平面惯性矩阵 $\mathbf{I} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}$的特征向量中，得到主轴$\mathbf{u}$和$\mathbf{u}$。令 $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \mathbf{c}$表示中心化坐标，其分量为 $(\tilde{x}_k, \tilde{y}_k)$。惯性分量计算如下：$I_{xx} = \sum_k w_k \tilde{y}_k^2$、$I_{yy} = \sum_k w_k \tilde{x}_k^2$和 $I_{xy} = -\sum_k w_k \tilde{x}_k \tilde{y}_k$。随后，我们对$\mathbf{I}$进行特征分解，并将$\mathbf{u}_1$ 定义为与最小特征值对应的特征向量，将$\mathbf{u}_2$定义为另一个正交归一特征向量。

边界范围：我们沿每个主轴$\mathbf{u}_i$ 计算边界范围 ℓ_i^+ 和 ℓ_i^-，确保该方法对不规则和凹几何形状也有明确的定义。为此，我们在平面域中计算物体边界的凸包 $\mathcal{H}$ 。边界范围是质心$\mathbf{c}$到凸包$\mathcal{H}$ 沿$\mathbf{u}_i$正向和负向的最大距离。

9. 其他实现细节

9.1. 训练策略

由于协作搬运任务具有长时域特性，直接使用多达八个智能体训练统一策略在计算上效率低下。因此，我们采用分阶段的顺序训练，逐步推进任务完成并扩大团队规模；一旦任何智能体跌倒或桌子倾倒，就会触发提前终止。以下所有阶段均使用单块 NVIDIA A100 GPU 进行训练。

阶段 1：我们首先训练由 1–4 个智能体构成的环境，使其掌握导航、接触和抬升等核心行为。在此阶段，运输、对齐和放下奖励均被禁用（即 $r_{\text{transport}} = r_{\text{align}} = r_{\text{put}} = 0$）。全身判别器 D_{full} 仅使用前进和侧向行走参考动作进行监督，而掩码判别器 D_{mask} 还会接收拾取动作。在此阶段，目标位置标记也会被屏蔽。训练持续约 1.5 天，回合长度为 400 个时间步。

阶段 1+2：接下来，我们继续使用2-4个智能体进行训练，以推进协同运输和放下。我们启用其余任务组件，包括$r_{\text{transport}}$、r_{align} 和r_{put}，并取消对目标位置标记的掩码。我们加入后退行走参考动作来监督 D_{mask}，以提升搬运物体时的运动多样性。该阶段大约在5天内收敛，回合长度为600个时间步。

最多8个智能体的微调：最后，我们在实例化为2-8个智能体的环境中对统一策略进行微调，以优化协调模式，并稳定更大规模团队的集体运输。该微调阶段大约需要3天。

9.2. 训练超参数

我们使用 1024 个并行环境训练所有阶段。对于 PPO，在训练最多四个智能体时，小批量大小设为 16384；在训练最多八个智能体时，则减小为 8192。对于 PPO 和 AMP 更新，均会将属于已停用智能体（即被放置在天花板上方的智能体）的观测从小批次中排除。对于 AMP，参考运动小批量大小为 4096，策略观测小批次设为 $1.5\times$ 4096，但不包括已停用的智能体。

除非另有说明，其余所有超参数均遵循CooHOI [7]。我们将关键取值列于表2 中，以供完整参考。

表2. 我们实验中的关键训练超参数。

超参数	值
时域长度	32
优化器	Adam [20]
学习率	2×10^{-5}
任务奖励权重	0.5
风格奖励权重	0.5
PPO裁剪阈值（ϵ）	0.2
折扣因子（γ）	0.99
广义优势估计参数（λ）	0.95

10. CooHOI* 基线

架构： CooHOI* 的策略和评论家均采用与我们的方法相同的基于 Transformer 的主干网络，但将交叉注意力替换为自注意力层，且不引入队友令牌。这一设计模仿了原始 CooHOI 公式，其中合作完全源于物体的共享动力学。

接近角奖励：我们设计了一种接近角奖励，引导每个智能体前往指定的接触点，同时避免与桌子发生碰撞。令 $\mathbf{p}_{\text{des}} \in \mathbb{R}^2$表示指定点的 x, y 坐标。我们首先计算从物体指向智能体根部的归一化 x, y 方向：$\hat{\mathbf{a}}_o = \mathbf{x}\| \overset{\text{root}}{-} \overset{\text{obj}}{\|}_{2\mathbf{xxx}}$，以及从物体指向指定点的归一化 x, y 方向：$\hat{\mathbf{p}}_o = \mathbf{p}\| \overset{\text{des}}{-} \overset{\text{obj}}{\|}_{2\mathbf{xpx}}$。接着，我们根据这两个方向之间的余弦相似度计算接近角奖励：

$$r_{\text{approach}} = \frac{\hat{\mathbf{a}}_o^\top \hat{\mathbf{p}}_o + 1}{2}. \tag{26}$$

当智能体正对其目标接触点（$\theta = 0°$）时，该值为 $r_{\text{approach}} = 1$；当它朝向相反方向时，该值为 $r_{\text{approach}} = 0$（$\theta = 180°$），从而有效避免智能体在随机位置初始化时与桌子发生碰撞。

聚合任务奖励的结构与公式25相同，只是 r_{form}、r_{ang} 和 r_{cov} 被接近角度奖励r_{approach}替换。

训练策略： 我们采用与 CooHOI 相同的两阶段训练流程。第一阶段训练一个单智能体，使其掌握基础的运动能力和操作技能，包括接近桌子、移动至指定接触点、建立接触、将桌子抬升（或倾斜）至目标高度，随后将其推向或拖向目标。为简化学习，此阶段将桌腿与地面之间的摩擦设为零，并将桌子质量减半。训练约持续 3 天。

随后在第二阶段从单智能体检查点恢复训练，引入多智能体协作。针对团队规模为 2、4 和 8 个智能体的情况分别训练独立模型，依次记为 CooHOI-2、CooHOI-4 和 CooHOI*-8。CooHOI*-2 约 2 天收敛。CooHOI*-4 需要约 6 天，而 CooHOI*-8 则从 4 智能体检查点继续训练，额外耗时 5 天。所有模型的回合长度均为 600 个时间步。

接触点分配： 为减少大型团队生成时智能体之间的碰撞，我们确保智能体与其指定接触点之间保持一致的几何映射。所有接触点均按逆时针顺序排序，从左下角开始。智能体初始化后，也按照相同的逆时针顺序进行编号，同样从左下位置开始。随后依照此顺序，将智能体编号与接触点一一对应。

11. 更多实验结果

适用于所有团队规模的统一策略： 为补充论文正文中展示的结果，表 3 报告了我们的统一策略在团队规模为 2 至 8 个智能体时的完整性能，涵盖正常（1×）和重型（5×）桌子质量配置。在论文正文所示配置（2A、4A、8A）之外，我们还纳入了中间团队规模（3A、5A、6A、7A），展示了相同的去中心化策略无需重新训练即可平稳地泛化到所有团队规模。在正常质量设置下，我们的模型在所有团队规模中都持续实现近乎完美的成功率，并展现出稳定的协作。

在重型配置（5× 桌子质量）下，增加的负载进一步凸显了协调产生合力的必要性。随着团队规模扩大，我们的统一策略能够促进有效协作，利用更大群体所带来的额外机械优势，从而使智能体数量越多，成功率越稳定地提升。

零样本泛化： 我们进一步评估统一策略在未见过的桌子几何形状和团队规模下的表现，测试训练期间学到的协调队形和搬运技能能否迁移到新场景中。我们考虑较小的桌子（圆形，直径 1.40 m；正方形，1.30 m × 1.30 m；长方形，1.60 m × 0.90 m）以及较大的桌子（圆形，直径 2.40 m；正方形，2.20 m ×2.20 m；长方形，3.0 m × 1.40 m），所有桌子的质量密度均与主要实验中的相同。

表3。在正常（1×）和重型（5×）桌子重量下，我们的统一策略在不同团队规模下的性能。

Team size	成功率（%）↑	d（米）↓	t_{coop}（%）↑	$\lvert J\rvert$（米/秒^3）↓
正常权重（1×）				
2	99.1	0.06	95.2	51.0
3	99.4	0.06	98.3	50.5
4	99.2	0.08	96.1	44.7
5	99.5	0.06	97.3	40.6
6	99.3	0.07	95.9	38.0
7	98.6	0.11	93.7	35.7
8	97.5	0.18	90.1	34.2
高权重（5×）				
4	3.5	4.77	90.9	23.4
5	18.2	2.48	79.0	28.3
6	50.1	1.04	79.0	32.0
7	71.6	0.59	81.2	31.8
8	81.1	0.49	81.5	31.7

表4。我们的统一策略在小桌子和大型桌子不同团队规模下的性能。所有结果均基于10,000个模拟回合取平均值。

Team size	成功率（%）↑	d（米）↓	t_{coop}（%）↑	$\lvert J\rvert$（米/秒^3）↓
小桌子				
2	93.1	0.37	94.8	63.2
3	97.5	0.14	97.0	64.7
4	98.4	0.12	97.1	55.5
8	96.4	0.23	85.4	45.0
大型桌子				
2	71.0	0.85	91.7	53.3
3	82.7	0.46	94.3	52.1
4	85.3	0.51	94.6	48.8
8	84.2	0.93	86.1	45.4
12	80.6	1.14	58.2	45.2
16	74.5	1.41	15.1	46.6

如表 4所示，尽管存在这种分布偏移，我们的策略仍能在所有配置下保持协调一致的协作行为。对于较小的桌子，智能体偶尔会表现出略强的抬升启动，导致加加速度略高，但运输阶段仍然稳定，成功率也始终保持在较高水平。对于较大的桌子，智能体仍能维持同步协作行为。然而，质量增加且力臂变长，使抬升和稳定桌子变得更加困难。因此，智能体有时会失去平衡、摔倒，并触发提前终止。大型桌子的设置对双智能体团队尤其具有挑战性，因为它们稳定和抬升较重桌子的机械力臂较小，导致运输速度较慢。

我们还评估了携带大型桌子的 12 智能体和 16 智能体团队的零样本泛化能力，这使策略面临的团队规模远超训练期间所遇到的范围。与变得极不稳定的基线相比，统一策略仍能产生同步且协调一致的运动，实现相对较高的成功率和较低的加加速度。然而，当团队规模变得很大时，桌面周界会变得拥挤，而且智能体尚未充分学会在狭窄的支撑间隙中定位自身，导致协作时间比降低。尽管如此，我们的策略对训练期间未见过的物体尺寸和更大团队规模总体表现出稳健的泛化能力。图 9中给出了若干定性结果。

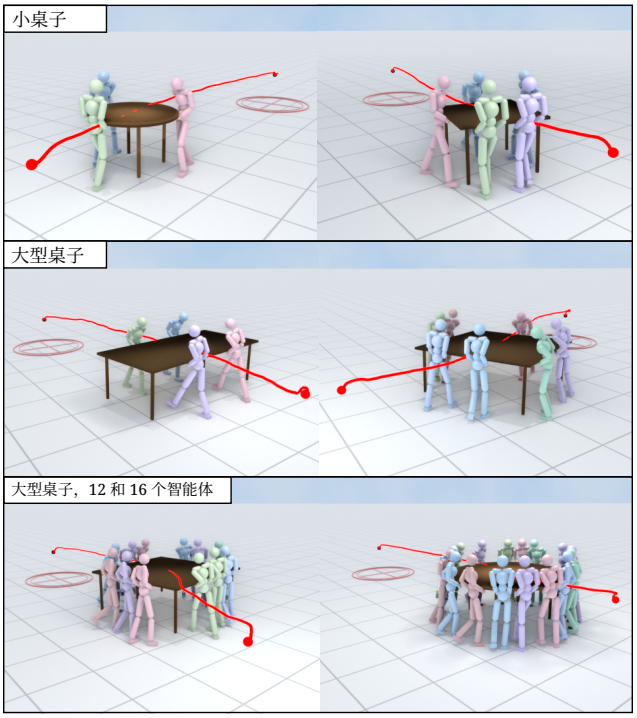

图9. 未见过的桌子几何形状和团队规模下零样本泛化的定性可视化。红线表示桌子的运动轨迹，黑点标记每个回合结束时的最终位置。

12. 多种可供性行为

虽然我们的主要实验聚焦于单一可供性行为（边缘抬起），但我们的框架可以通过调整任务奖励来支持多种可供性行为。图 10展示了这一能力：智能体会根据与相应可供性可行区域的距离，采取侧向扶持或边缘抬起行为。

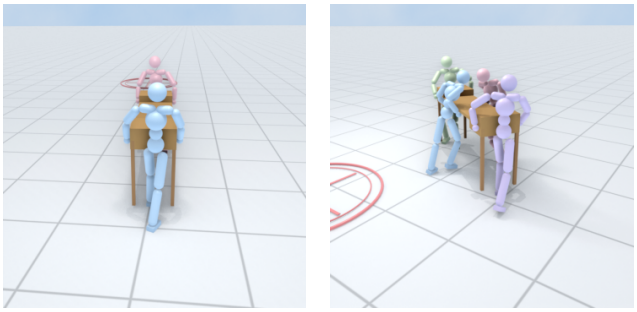

图10. 通过调整任务奖励学到的多种可供性行为示例。智能体能够适应侧边抓握或边缘抬起，同时还能朝不同方向行走。该策略使用与主要实验相同的单人参考动作进行训练。