arXiv:2611.33010v1 [cs.CV] 24 Nov 2026

具身人工智能未通过镜子测试：一面简单的镜子如何破坏机器人策略

Xuanchen Liu*，Hejun Zhang*，Boxuan Dong、Ying Wu、Heshuai Li、Wen Gao、Yonghong Tian、Weihao Yu†北京大学代码：

https://github.com/iPPLab/robodojo-mirror*贡献相同。 †通讯作者：

weihao@pku.edu.cn

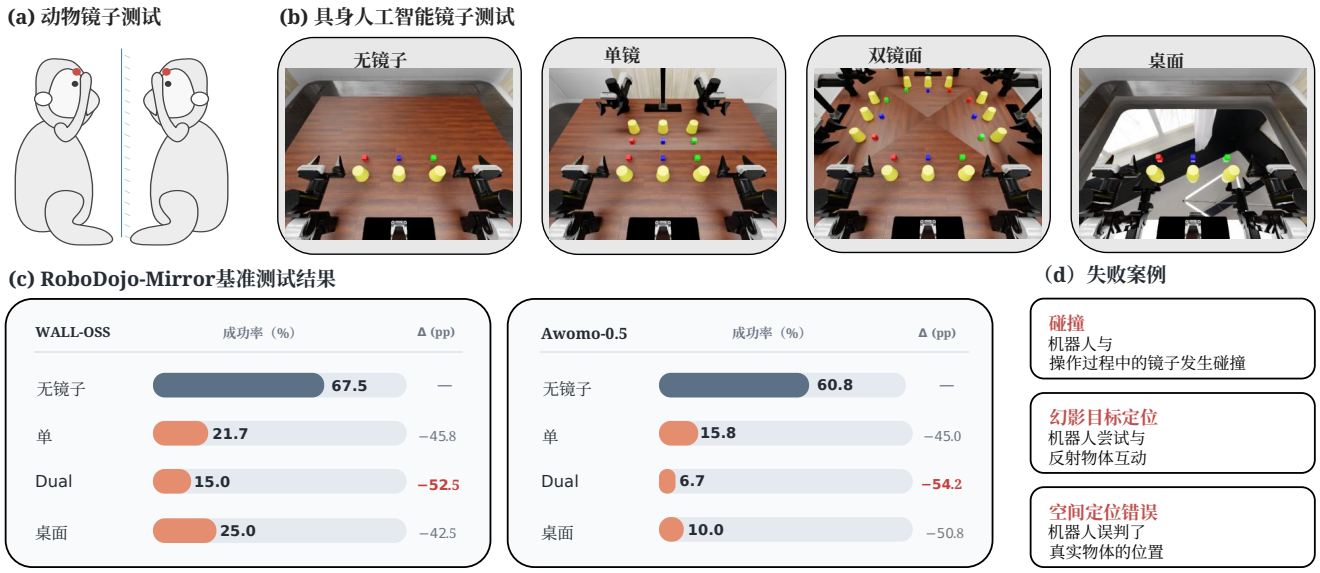


图 1。**镜像基准概览**。 (a) 动物镜子测试促使人们研究机器人在存在反射影像时的行动方式。(b) 四种实验配置：无镜子、单镜、双镜面和镜面桌面。(c) WALL- OSS 和 Awomo-0.5 在不同配置下的成功率，以及相对于无镜像基线的变化（以百分点计）。性能下降幅度根据未四舍五入的成功率计算。(d) 镜像场景中的三类失败：碰撞，即与镜面发生非预期接触；幻影目标定位，即机器人朝向反射物体执行动作；以及空间定位错误，即机器人错误估计物理目标的位置。

摘要

一面简单的镜子能否揭示当今机器人操作策略的局限？受动物认知领域镜像自我识别测试的启发，我们研究机器人策略能否在包含自身反射影像的场景中可靠地操作物理物体。我们推出了 *RoboDojo-Mirror*，这是一个基于 *RoboDojo* 构建的基准，在保留任务目标、成功标准和策略检查点的同时，对比无镜基线、单镜、双镜和镜像桌面条件。在涵盖视觉-语言-的五种策略中，对于动作模型、世界动作模型和智能体式策略，每种镜面条件都会降低总体成功率：在三种条件下取平均值，各策略的成功率降幅为 *10.8* 至 *56.7* 个百分点，在镜面桌面条件下则达到 *62.5* 个百分点。由于镜子也是一种物理障碍物，我们将其与几何匹配的非反射面板进行比较；*50%* 至 *88%* 的镜像特定失败发生时并未记录到机器人–面板接触，因此仅凭直接碰撞无法解释这种性能下降。相反，轨迹和接触分析揭示了反复出现的行为，这些行为与将反射物体误作目标的幻影目标定位，以及对

真实物体的空间定位错误相符。针对性诊断进一步刻画了反射影像附近的抓取、依赖策略的推动进展，以及主摄像头被遮挡时与物理目标的交互。使用 $\pi_{0.5}$ 和 *FastWAM* 开展的真实机器人实验也表明，在直立镜面和镜面桌面条件下，总体成功率较低。这些发现揭示了不同机器人策略对反射环境普遍存在的脆弱性，并促使人们对具身智能开展具备反射感知能力的评估。

1. 引言

一面普通的镜子能否揭示机器人将视觉观察与物理动作联系起来的局限性？经典的镜像自我识别测试考察动物是否能够利用反射影像引导动作指向自己的身体 [10]。受这一行为学探测试验启发，我们探究：当机器人的观测中也包含这些物体的反射影像时，机器人策略能否可靠地定位并操作物理物体。

镜子在视觉外观与可操作空间之间制造了一种结构化的不一致。反射影像保留了可识别特征，并随物理物体的变化而变化，但其表观位置并不包含额外的可操纵目标。它们还可能在图像中与真实物体相邻，形成复合轮廓，或显露原本被遮挡的目标。这些特性使我们能够测试两种互补能力：在存在反射影像的情况下保持可靠的操作能力，以及利用反射影像观测对被遮挡目标执行动作。

近年来，通用机器人策略取得了进展，能够根据视觉和语言输入执行的操作任务范围不断扩大；然而，这些策略是否具备上述两种能力，仍不明确。视觉-语言-动作模型直接根据多模态观察生成动作，而世界动作模型则对动作与未来观测进行联合建模 [2, 29]。智能体策略通过反复观测和推理来选择并调用机器人工具，提供了另一种方法 [15]。然而，在标准评估条件下的出色表现，并不能证明其能够适应视觉证据的变化。受控扰动基准测试表明，当视角、光照、背景和其他条件发生变化时，性能会显著下降 [7]。反射影像并不属于这些扰动：它们并非改变视觉外观，而是在与物理场景具有已知几何关系的位置复制与任务相关的内容。

现有基于镜子的评估已开始考察具身智能体的自我中心感知与交互，包括受镜像自我识别启发的任务 [12]。我们研究一种不同的操作情境：通用策略在受控镜像配置下操纵外部物体。

上述问题可分为三个部分：反射影像如何影响任务完成，观察到的性能下降是否超出了插入实体面板所造成的影响，以及伴随这些变化出现了哪些执行行为。

我们介绍 RoboDojo-Mirror，这是一个构建于 RoboDojo 之上的评估基准 [3]。如图 1 所示，它保留了原始的操作目标和成功标准，同时在四种条件下评估策略：无镜子、单镜、双镜面以及镜面桌面。在一组包含八项任务的共同任务集上，我们对齐不同条件下的布局和策略随机种子，以衡量相对于每个策略自身无镜基线的变化。对于选定的策略，几何匹配的非反射面板可进一步控制插入实体结构所带来的影响。我们还设计了三个诊断场景，即在镜面桌面上进行抓取、推动一个与其反射影像相连的物体，以及在反射影像作为主要视觉线索的情况下执行动作，并结合任务结果、轨迹和接触观察。这一协议使研究人员能够利用现有策略和任务评估器评估反射鲁棒性，并刻画总体成功率之外的失败情况。

在涵盖视觉-语言-动作模型、世界动作模型和智能体策略的五种策略中，三种镜像条件都会降低总体成功率，各策略的平均下降幅度为 10.8–56.7 个百分点。匹配面板的比较显示，50–88% 的镜像特定失败在没有记录到机器人–面板接触的情况下发生，因此仅凭直接碰撞无法解释性能下降。伴随出现的行为与对反射物体的幻影目标定位，以及对真实物体的空间定位错误相一致。三个诊断场景进一步显示：空抓取更为频繁；推动进展会随策略和轮廓而变化；即使存在与镜子相关的提示，与实体目标的接触仍然有限。使用 $\pi_{0.5}$ 和 FastWAM 开展的真实机器人实验也显示出类似的总体性能下降趋势。总体而言，这些结果揭示了不同机器人策略对反射影像普遍存在的共同脆弱性。

2. 相关工作

2.1. 机器人操作基准与视觉鲁棒性

机器人操作基准为评估策略能力提供了互补工具。LIBERO 研究终身机器人学习中的知识迁移 [20]。RoboTwin 2.0 将可扩展数据生成与结构化域随机化相结合，用于双臂操作 [4], ，而 RoboDojo 则为通用操作策略提供统一的仿真与现实世界评估框架 [3]。近期基准更加重视超越名义任务成功的鲁棒性。LIBERO-

Plus 引入了针对物体布局、相机视角、机器人初始状态、语言、光照、背景和传感器观测的受控扰动 [7]。LIBERO-VPro 通过扰动执行期间可用的证据来评估闭环视觉鲁棒性，包括观测性能下降、相机陈旧性和跨视角一致性 [17]。LIBERO-Pro2.0 进一步涵盖环境、观测、执行、机器人状态和语言维度上的静态偏移与运行时偏移 [35]。RoboPRO 通过评估杂乱环境和受控扰动下的碰撞指标，为基于成功的评估提供补充 [18]。基于这些评估原则，我们的工作聚焦于镜像增强操作，其中反射影像与物理场景变化共同引入结构化视觉歧义。我们评估任务结果和接触行为，并使用几何匹配的非反射面板，考察超越面板本身是否存在的性能差异。

2.2. 镜像感知与空间推理

镜子会对视觉理解构成挑战，因为反射物体可能看起来与直接观测到的物体相似，却位于不同的物理位置。早期的镜像检测方法通过上下文对比识别反射区域 [28], , 而基于视频的方法则利用空间和时间对应关系 [19]。这些方法关注的是镜面区域的定位，而非反射观测对动作造成的影响。Mirror3D 还会估计镜像几何，以修正反射表面上不可靠的深度测量 [26]。最近，MIRA-Bench 在真实 RGB-D 镜面场景中评估反射感知、空间推理和场景级决策，揭示了将反射外观与深度测量相协调的难点 [30]。MirrorQA 考察镜像情境下以主体为中心的左右推理 [21]。相关诊断基准，包括 HallusionBench [11]和 IllusionVQA [25], , 研究视觉错觉以及基于图像的推理失败。与这些侧重感知的研究不同，我们评估反射条件下闭环操作中的任务结果和接触行为。

2.3. 镜像介导的机器人交互

机器人镜子研究涵盖几何学与感觉运动学习。Hart 和 Scassellati 从学习到的感觉运动关系中推断镜面视角 [14]。Marchand 和 Chaumette 为机器人搭载的镜子开发视觉伺服 [22], , 而 Lanillos 等人将主动推断与感觉运动预测相结合，用于镜像自我识别以及自我与他者区分 [16]。Yoshioka 等人利用镜子反射来观测遮挡区域，而 UMI 将侧视镜用于操作，并表明经过校正的镜面视图可以优于未经校正的视图 [6, 31]。重—新近的具身评估将基于镜子的测试扩展到多模态智能体。MirrorBench 通过模拟镜面交互研究以自我为中心的智能 [12], , 而 Ziliotto 等人评估智能体能否从反射中推断隐藏的自我属性，并利用这些证据指导动作 [36]。除以自我为中心的任务外，Being-M0.7 还在现实世界的人形机器人任务中，评估对直接遮挡的玩具进行镜像辅助抓取的能力 [33]。我们的工作研究在受控镜面配置下，多种通用策略和操作任务中的反射鲁棒性。

3. 基准设计与评估协议

基于 RoboDojo [3]，RoboDojo-Mirror 评估策略对反射影像的鲁棒性，同时保留原始任务目标和成功标准。图2 概述了配对评估协议和控制项。

3.1. 策略与任务选择

我们评估了五种策略，涵盖视觉-语言-动作模型、世界动作模型和智能体式策略：WALL-OSS-1.0-Lite (WALL-OSS) [27], VPP2 [13],Awomo-0.5 [1], Rex-M1 Preview [9], 和 GPT-6-Astra (Astra) [23, 34]。这些选择涵盖了从视觉观察生成机器人动作的不同方法。

对于开放权重策略，我们使用官方检查点和推理配置。在每种策略中，我们在不同场景条件下固定检查点、观测接口、动作接口和推理设置。对于 Astra，可用的摄像头视频流、机器人状态接口、运动工具和推理预算也都在不同条件下保持不变。

我们选择一个共享的八任务集合 $\mathcal{T}$，兼顾多样化的操作需求与出色的整体RoboDojo排行榜表现。我们在四种场景条件下评估所有策略，并与各自的无镜基线进行配对比较。任务列表见补充材料。

3.2. 镜面配置

我们根据每种任务布局构建四种场景条件：无镜子、单个垂直镜子、双垂直镜子以及镜面桌面。这些条件通过改变反射影像的放置位置和几何形态，同时保持物体与机器人状态的初始值一致。

无镜子。 我们保留原始RoboDojo场景作为基线。

单镜。 在桌面边缘附近放置一面垂直平面镜，其镜面与该边缘平行，反射面朝向操作区域。

双镜面。 两面垂直镜子沿工作空间相邻的两侧放置，镜面相互垂直，反射面朝向操作区域。这种设置会引入来自

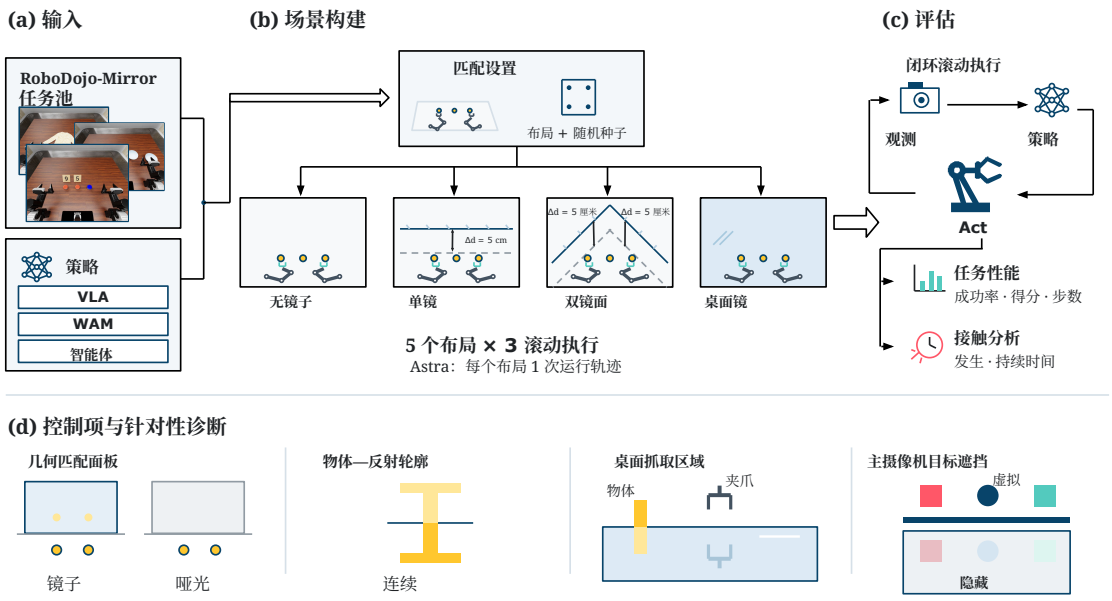


图 2. RoboDojo-Mirror 中评估协议的概述。在四种场景条件下，使用匹配的任务布局对策略进行评估，并针对部分策略额外设置非反射面板对照。

两面镜子的反射，并且还可能显示高阶反射，具体取决于可见性和渲染设置。

镜面桌面。 我们将桌面的表面材料替换为反射性表面处理，同时保持其位置和尺寸不变。因此，反射影像会出现在物体附近，以及进行操作的表面上。

对于单镜和双镜条件，我们将面板相对于其初始构造位置向远离机器人的方向移动 5 cm，以增加间隙，同时保留清晰可见的反射影像。面板固定在场景中。此放置位置不能保证执行过程中不会发生碰撞；因此，面板接触情况会与任务成功情况分开记录。

对于 WALL-OSS 和 Awomo-0.5，我们还将单镜和双镜替换为几何形状匹配的非反射面板，作为物理阻挡对照。镜面桌面未设置相应对照。

在四种主要评估条件下，我们保留原始任务指令、匹配的初始物体与机器人状态，以及任务成功标准。在场景验证期间，我们检查提供给各策略的摄像头视频流，以确认反射可见性、目标直接可见性，以及插入面板与现有场景几何之间是否存在初始相交。我们将场景和相机配置与评估记录一并保留。第 4.3 节中的目标遮挡实验采用单独的诊断设置。

3.3. 评估协议与指标

配对评估。 对于每种策略和选定任务，我们在不同场景条件下运行多个独立的闭环回合，并匹配布局和策略随机种子。主要评估中，WALL-OSS、VPP2、Awomo-0.5 和 Rex-M1 Preview 的每项任务采用五种布局，每种布局使用三个策略随机种子。Astra 则采用五种布局，每种布局运行一条运行轨迹。

终止与预算。 我们保留原始任务成功标准和特定于任务的终止规则。当任务终止或执行预算耗尽时，回合结束。我们会记录与面板的接触，但这并非额外的终止条件。对于 Astra，执行预算还包括模型调用次数上限。当诊断实验采用的布局、随机种子和预算与主要评估不同时，会单独报告。

令 $y_{m,\,t,\,c,\,l,\,s} \in \{0,1\}$ 表示策略 m在任务 t、条件 c、布局 l以及策略随机种子索引 s下的任务成功。设评估所用的布局集和随机种子集为 $\mathcal{L}_{m,\,t}$ 和 $\mathcal{S}_m$，则任务成功率和总体成功率为

$$\begin{aligned} \mathrm{SR}_{m,t,c} &= \frac{100}{|\mathcal{L}_{m,t}||\mathcal{S}_m|} \sum_{l\in\mathcal{L}_{m,t}} \sum_{s\in\mathcal{S}_m} y_{m,t,c,l,s}, \\ \mathrm{SR}_{m,c} &= \frac{1}{|\mathcal{T}|} \sum_{t\in\mathcal{T}} \mathrm{SR}_{m,t,c}. \end{aligned} \tag{1}$$

因此，SR 以百分比表示，固定八任务集合 $\mathcal{T}$ 中的每项任务权重相同。相较于无镜像基线 B，

$$\Delta\mathrm{SR}_{m,c} = \mathrm{SR}_{m,c} - \mathrm{SR}_{m,B}. \tag{2}$$

负向变化表示性能下降。相应的下降幅度为 $-\Delta$成功率$_{m,c}$。表 1 对三种镜像条件下的下降幅度取等权平均 $\mathcal{M}$:

$$D_m = -\frac{1}{3}\sum_{c\in\mathcal{M}} \Delta\mathrm{SR}_{m,c}. \tag{3}$$

变化和下降均以百分点表示；基线比较在每项任务和策略内，使用匹配的布局和策略种子索引。原生任务得分按 0–100 缩放，并以相同方式汇总，以反映部分完成情况。

我们报告 95% 布局簇自助法区间：在固定任务内对布局进行重采样，同时保留所有随机种子及条件配对（详情见补充材料）。

接触与行为记录。我们在每条运行轨迹中记录视觉观测、机器人状态和动作轨迹。对于具有有效接触测量的回合，我们还会监测与镜面板的物理交互。我们将接触回合率定义为至少发生过一次与面板的记录接触的回合所占比例。

4. 镜像条件下的性能

继第3节之后，我们报告总体性能、镜面–面板比较、针对性行为诊断和真实世界验证。通过轨迹、视频和接触分析，刻画观察到的三种行为：碰撞、幻影目标定位和空间定位错误。

碰撞。 机器人在执行过程中与镜子发生接触，可能会干扰操作。我们将此类接触事件与任务失败区分开来，因为发生接触并不一定会导致失败。

幻影目标定位。 机器人将操作动作指向反射物体，而不是其实体对应物，这表明反射影像被当作了可操作的目标。

空间定位错误。 机器人相对于物理目标在错误的位置执行动作，导致抓取失败或操作无效，但没有明确证据表明存在幻影目标定位。

4.1. 总体性能

表1 汇总了任务集上的任务得分和成功率，其中每个策略都与其无镜基线进行了比较。

这五种策略在无镜子条件下的成功率均达到 60.8–72.5%，而每种镜面配置都会降低其总体成功率。各策略的平均配对成功率下降幅度为 10.8 至 56.7 个百分点。在镜面桌面条件下，VPP2 的成功率从 72.5% 降至 27.5%，而 Rex-M1 Preview 则从 66.7% 降至 4.2%。在所有镜面配置下，任务得分也会下降。

总体而言，双镜面造成的性能下降往往大于单镜面，VPP2 除外。WALL-OSS 和 Awomo-0.5 在双镜面条件下的成功率最低，而 VPP2、Rex-M1 Preview 和 Astra 在镜面桌面条件下表现最差。

在单镜条件下，Astra 的总体变化很小（从65.0%降至62.5%），但在双镜面条件下，其成功率下降至52.5%，在镜面桌面条件下则降至47.5%。尽管 Astra 在单镜和双镜条件下均表现出记录中的最高机器人–镜面接触率（分别为55.0%和85.0%），其成功率下降却相对较小，这表明仅凭接触事件的发生无法解释性能下降。

由于各策略的执行预算不同，因此不应将这些汇总数据解读为经过控制的跨策略排名。此外，仅凭任务结果无法区分视觉干扰和物理阻碍，因此有必要在下一小节进行镜面–面板比较。

4.2. 镜面–面板比较

镜子既会产生视觉反射，也会形成物理障碍，表 1 中观察到机器人–镜面接触的情况十分常见。为考察直接机器人–表面接触是否足以解释性能下降，我们将镜子与几何匹配的非反射面板进行比较（第 3.2 节）。

非接触失败与空间定位错误。 令$h \in \{0,1\}$ 表示是否发生机器人–面板接触。为简洁起见，省略其他索引，我们定义两组配对回合:

$$\begin{aligned} \mathcal{F}&: y_{\mathrm{NoMirror}} = y_{\mathrm{Panel}} = 1, \quad y_{\mathrm{Mirror}} = 0 \\ \mathcal{N}&: y_{\mathrm{NoMirror}} = 1, \quad h_{\mathrm{Mirror}} = h_{\mathrm{Panel}} = 0 \end{aligned} \tag{4}$$

这里， $\mathcal{F}$ 包含镜像特定失败，而 $\mathcal{N}$ 包含这样的配对：无镜像基线成功，且镜子条件和面板条件下均未发生机器人–面板接触。

表2 显示，在每种设置的 120 个匹配对中，仅镜面失败的数量均多于仅面板失败，尤其是在双镜面条件下（WALL-OSS 为 25 / 2，Awomo-0.5 为 42 / 3）。这种不对称性在非接触子集中依然存在 $\mathcal{N}$。此外，在 $\mathcal{F}$ 中，50.0–87.5% 的镜像特定失败在任一条件下均未记录到机器人–面板接触。这些观测表明，失败差异并非仅由直接碰撞造成，这与实体目标可能存在空间定位错误的情况相符。

持续接触行为与幻影目标定位。除了非接触失败之外，我们还研究了反射是否会改变机器人在发生物理接触时的行为。我们定义了这样一组配对回合：其中两种条件都涉及机器人–面板接触:

$$\mathcal{C}: y_{\mathrm{NoMirror}} = 1, \quad h_{\mathrm{Mirror}} = h_{\mathrm{Panel}} = 1 \tag{5}$$

表1。镜像扰动下的任务性能。场景条件条目报告得分 / 成功率（%）。单镜和双镜列中的括号百分比表示至少记录到一次机器人–镜面接触的回合所占比例。下降值是在三个镜像条件下，使用匹配布局和策略种子索引，对基线减去镜像后的差值取平均。方括号表示成功率下降的95%配对布局聚类自助法置信区间。

	场景条件：得分 / 成功率（%）				平均配对下降	
策略	None	单	Dual	桌面	得分 ↓	成功率（百分点）↓
WALL-OSS	70.6 / 67.5	23.3 / 21.7（34.2%）	17.2 / 15.0（28.3%）	27.2 / 25.0	48.0	46.9 [38.3，55.3]
VPP2	73.2 / 72.5	36.8 / 33.3（49.2%）	40.9 / 35.0（47.5%）	34.3 / 27.5	35.8	40.6 [34.2，46.9]
Awomo-0.5	62.7 / 60.8	17.8 / 15.8（27.5%）	8.0 / 6.7（30.0%）	13.0 / 10.0	49.8	50.0 [42.2，58.1]
Rex-M1 Preview	68.8 / 66.7	18.0 / 15.8（12.5%）	13.2 / 10.0（45.0%）	6.7 / 4.2	56.2	56.7 [48.3，65.0]
Astra	66.3 / 65.0	63.3 / 62.5（55.0%）	57.0 / 52.5（85.0%）	51.3 / 47.5	9.1	10.8 [1.7，20.0]

表2. 单镜和双镜配置下匹配的镜面–面板比较。ALL 报告每种设置下 120 个匹配对中的仅镜面/仅面板失败计数。对于 $\mathcal{F}$，数值报告在任一条件下均未发生机器人–面板接触的镜像特定失败数量及百分比。对于 $\mathcal{N}$，数值报告仅镜面与仅面板失败计数。

子集	配置	WALL-OSS	Awomo-0.5
ALL	单	9 / 4	16 / 7
	Dual	25 / 2	42 / 3
$\mathcal{F}$	单	7 / 8 (87.5%)	9 / 13 (69.2%)
	Dual	11 / 22 (50.0%)	20 / 34 (58.8%)
$\mathcal{N}$	单	7 / 2	9 / 5
	Dual	11 / 1	20 / 2

在 $\mathcal{C}$ 双镜面条件下，WALL-OSS 和 Awomo-0.5 的面板成功率分别比镜像成功率高出 29.4 和 41.2 个百分点，这表明尽管两种条件下都发生了接触，但结果并不相同。

为表征接触持续性，我们将接触时间比例定义为

$$R_{\text{contact}} = \frac{T_{\text{contact}}}{T_{\text{execution}}} \tag{6}$$

在 $\mathcal{F} \cap \mathcal{C}$ 双镜面条件下，与面板相比，WALL-OSS（$n = 6$）和 Awomo-0.5（$n = 7$）使用镜子时的平均接触时间比例分别高出 1.85 倍和 6.83 倍。这种行为与一种可能的幻影目标现象相符：策略可能会反复尝试触及反射物体看似所在的位置，却被实体镜面挡住。相比之下，单镜面配置中的接触行为并未呈现同样显著且一致的差异。完整

统计数据见补充材料。

总体而言，镜面–面板比较揭示了直接碰撞之外的失败，其行为模式与空间定位错误和幻影目标定位相一致。第4.3 节通过受控场景和执行视频进一步调查这些失败模式。

4.3. 针对性行为诊断

为了补充汇总的任务结果，我们采用三种诊断方法，刻画物体与其反射影像之间不同关系下的抓取行为、推动进展和物理目标交互。

4.3.1. 在镜面桌面上抓取

镜面桌面使反射物体和机器人的特征紧邻直接可见的目标。我们考察策略究竟会抓取物理物体，还是在未能夹稳物体的情况下关闭夹爪并抬升。

图3 说明了与几何空夹闭合并抬升（ECL-G）相符的行为，具体量化结果见下文：夹爪在远离实体管子的位置闭合，并在空手状态下抬升。在此示例中，闭合位置不匹配与空间定位错误相符。

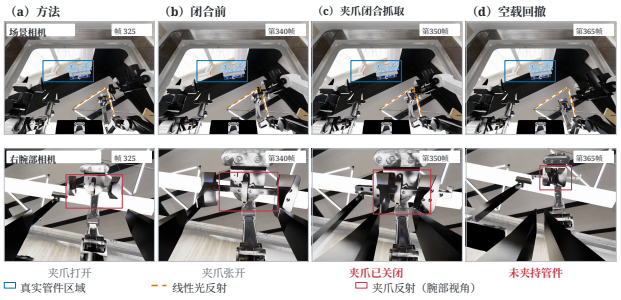


图 3. 镜面桌面上与 ECL-G 一致的空抓行为。场景视图和右手腕相机视图显示，夹爪关闭并抬升，但未能夹稳实体管子。标注区分了管子与附近的反射影像。

表3比较了 VPP2、Awomo-0.5、WALL-OSS 和

表3. 汇总三个任务的任务成功和抓取相关行为。条目为无镜子/镜面桌面的百分比。每种条件下，每个策略包含45个回合，Astra除外（15个）。行为事件可能重叠。

策略	任务成功率	ECL-G	NL	NF
VPP2	97.8/4.4	0.0/95.6	26.7/42.2	6.7/100.0
Awomo-0.5	80.0/26.7	8.9/17.8	35.6/62.2	17.8/44.4
WALL-OSS	60.0/28.9	24.4/40.0	42.2/53.3	26.7/51.1
Astra	33.3/33.3	13.3/26.7	53.3/46.7	60.0/73.3

Astra 在导管插入、方块堆叠和瓶子处理任务中的表现。每种策略内按布局配对条件，每个条件有 45 个回合（每项任务 15 个），Astra 则有 15 个回合（每项任务 5 个）。除了任务成功率，我们还考察夹爪闭合并抬升时是否确实带动物体移动。我们报告三项回合级比率：几何空夹闭合并抬升（ECL-G），即夹爪闭合并抬升时，其传感区域内没有目标物体；未抬升（NL），即某个闭合时间段内未出现所需的夹爪抬升；以及未确认物体随动（NF），即发生抬升，但未确认物体随夹爪移动。这些事件可能在同一回合中重叠发生。

镜面桌面会提高四种策略的 ECL-G 和 NF。VPP2 的成功率从 97.8% 降至 4.4%，而 ECL-G 从 0.0% 升至 95.6%：几乎每个镜像回合都包含闭合和抬升，却没有包围物理目标。对于 Astra，成功率保持不变，NL 则有所下降，但 ECL-G 和 NF 均有所上升。因此，无抬升事件减少并不一定意味着抓取效果更好，而任务成功率不变也可能掩盖了空抓或未经确认的抬升更频繁这一情况。

4.3.2. 推动与其反射影像相连的物体

在前述抓取示例中，物理目标与附近的反射影像仍可分别看见。接下来，我们考察物理物体与其反射影像相邻时的推动情况。在一种基于镜子的 Push-T 变体中 [5, 8], ，物体及其反射影像构成 Π形、I形或十字形轮廓（图4）。

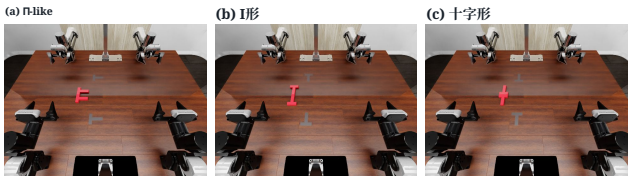


图4. 实体T形物体与其镜面反射相邻形成的复合轮廓：(a) 侧向 Π形，(b) I形，(c) 十字形。

我们在每个轮廓的五个匹配放置位置上评估 VPP2、WALL-OSS 和 Astra。VPP2 和 WALL-OSS 在每个放置位置使用三个随机种子（$n = 15$），而 Astra 在每个放置位置使用一条运行轨迹（$n = 5$）。回合在达到有效分离或经过 1,000 个控制步后结束。

表4. 推动进展和机器人–镜面接触。位移和步数：均值 ± 样本标准差；步数仅包括达到 10 毫米的回合（“–”：无）。接触：每回合的平均事件数。

轮廓	已达到 10 毫米 (%)	最大位移 (mm)	达到 10毫米的步数	镜像接触
VPP2 (WAM)，$n = 15$				
十字形	53.3	23.64 ± 23.38	308.1 ± 216.7	13.07
Π形	0	< 0.01	–	11.80
I形	93.3	57.79 ± 48.53	303.0 ± 168.7	6.27
WALL-OSS（*VLA*），$n = 15$				
十字形	86.7	47.11 ± 31.25	258.3 ± 70.0	13.87
Π形	80.0	32.84 ± 44.07	420.9 ± 192.5	17.20
I形	100	105.97 ± 41.90	105.6 ± 22.7	6.47
Astra（智能体式），$n = 5$				
十字形	100	135.04 ± 42.04	229.6 ± 111.9	5.60
Π形	100	174.64 ± 5.32	208.2 ± 156.5	8.40
I形	100	160.69 ± 15.06	232.4 ± 224.8	4.20

我们考察策略是否能够调整其动作，使物理T形物体远离其反射影像，并打破连通轮廓。表4 报告达到 10 毫米位移的比率、最大位移、达到该阈值所需的步数，以及机器人–镜面接触次数。达到率越高、位移越大，表示物体移动越频繁且幅度越大。在达到阈值的回合中，所需步数越少表示越早越过该阈值。镜面接触次数较多但物体移动很少，表明策略难以调整动作以实现有效推动。

Astra 在全部 15 个回合中均实现了持续有效分离，最大位移均值为 135.04–174.64 毫米，每个回合发生 4.20–8.40 次镜面接触；相比之下，另外两种策略的该数值为 6.27–17.20。镜面接触后反复改变轨迹，表明其能够有效适应镜面约束。在 I形轮廓下，VPP2 和 WALL-OSS 进展最大，分别在 93.3% 和 100% 的回合中达到 10毫米。在 Π形轮廓下，VPP2 的移动距离不足 0.01 毫米，尽管其平均发生 11.80 次镜面接触。这些结果凸显了不同策略在将物理物体与其相邻的反射影像分离时，所采取的调整方式各不相同。

4.3.3. 以反射影像为主要视觉线索进行执行

我们测试：当主摄像头只能看到实体目标的反射影像、无法直接看到目标时，策略能否定位并操纵这些目标；同时也测试明确的镜像相关引导是否有所帮助。在覆盖方块任务中，升高面板挡住了主摄像头对方块和杯子的直接视野，但保留了它们的镜面反射（图5）。腕部相机仍能直接看到部分目标。我们测量实体目标接触率和任务成功率，以评估在这种受限视野下，策略能否接触真实物体并完成任务。

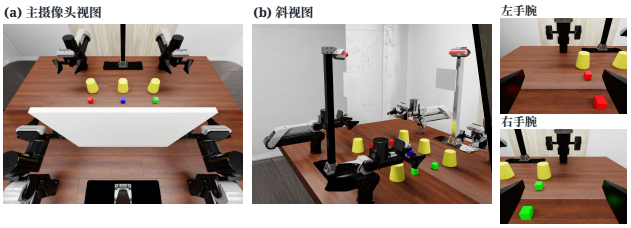


图5. 覆盖方块中的目标遮挡。面板遮挡了主摄像头视野中的物理目标，同时保留了它们的反射；部分目标仍可被腕部相机直接看到。

表5. 主摄像头视野受遮挡时的目标接触与任务成功率。所有数值均为百分比。

策略	提示词	杯子接触	方块接触	SR
WALL-OSS	原始	20.0	20.0	0.0
	简短	20.0	40.0	0.0
	详细	0.0	13.3	0.0
Awomo-0.5	原始	20.0	0.0	0.0
	简短	20.0	0.0	0.0
	详细	26.7	0.0	0.0
Astra	原始	100.0	0.0	20.0
	简短	100.0	0.0	0.0
	详细	100.0	0.0	20.0

我们比较 WALL-OSS、Awomo-0.5 和 Astra 在原始（指令不变）、简短（包含镜像提示）和详细（包含场景描述）提示词下的表现。每种提示词均采用相同的五种布局；WALL-OSS 和 Awomo-0.5 在每种布局下使用三个随机种子（$n = 15$），Astra 则在每种布局下运行一次（$n = 5$）。表5报告实体目标接触率和原生任务成功率。接触率衡量机器人接触真实杯子或方块的频率；接触率越高，表示实体交互越频繁。杯子接触与任务直接相关，因为任务要求移动杯子来覆盖方块。成功率衡量这些交互是否促成任务完成。比较不同提示词下的这两项指标，可以检验镜像相关引导是否能帮助策略接触并成功操纵实体目标。

WALL-OSS 和 Awomo-0.5 仅在 0.0–26.7% 的回合中接触到实体杯子，并且从未完成任务，这表明在主摄像头遮挡的情况下，它们难以到达并操作与任务相关的物体。对于 WALL-OSS，简短将方块接触率从 20.0% 提升至 40.0%，但并未改善杯子接触率或 SR。Astra 在每个回合中都能到达杯子，但在原始和详细条件下仅完成 1/5 个回合的任务，在简短条件下则为 0/5。因此，到达实体目标与成功操作实体目标仍然是两个不同的问题，而镜像相关引导在这些小样本中并未带来一致的改善。

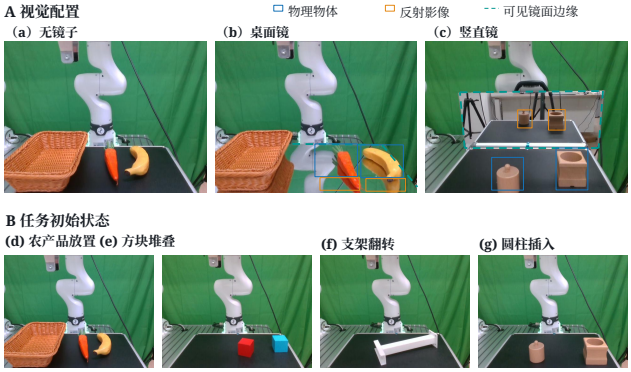


图6. 真实世界视觉条件和任务初始状态。(A) 无镜、镜面桌面和竖直镜子配置；标注指出物理物体、反射影像和可见镜面边缘。(B) 农产品放置、方块堆叠、支架翻转和圆柱插入的初始状态。

4.4. 真实世界验证

为检验仿真中观察到的性能下降是否延伸至物理操作，我们在真实世界镜像条件下评估机器人策略。

我们在 Franka 机器人上评估 $\pi_{0.5}$ [24] 和 FastWAM [32] 在四项任务中的表现：农产品放置、方块堆叠、支架翻转和圆柱插入。每项任务均采用无镜、竖直镜子和镜面桌面条件（图 6）。竖直镜面条件使用单面竖直镜，其类型与仿真中的单镜像配置相匹配。五种匹配的初始布局各重复两次，因此每项任务、每种策略和每种条件均进行10次试验。所有条件下的检查点、指令、相机配置和执行预算均保持不变。更多实验细节见补充材料。

表 6 显示，在两种镜子条件下总体成功率均较低。对于 $\pi_{0.5}$，成功率从无镜子时的 77.5% 降至镜子正立时的 52.5 %，以及镜面桌面时的 27.5%；FastWAM 则分别从 60.0% 降至 30.0% 和 10.0%。两种策略在无镜子时的方块堆叠任务中均达到 80.0%，但在镜面桌面条件下均为 0.0%。这些结果表明，总体性能下降趋势与仿真结果一致。

5. 结论

我们推出 RoboDojo-Mirror，用于评估机器人在受控镜面配置下的操作能力。对五种策略进行的配对评估及真实世界实验表明，镜子会降低成功率。匹配的面板对照显示，仅凭直接碰撞无法解释这种性能下降。行为诊断揭示了与幻影目标定位和空间定位错误相一致的模式。这些发现揭示了将视觉

表6。不同镜子条件下的真实世界成功率（%）。每种设置包括10次试验（五种布局，每种布局重复两次）。

	$\pi_{0.5}$			FastWAM		
Task	无	直立	桌面	无	直立	桌面
农产品放置	90.0	90.0	60.0	80.0	60.0	40.0
方块堆叠	80.0	30.0	0.0	80.0	0.0	0.0
支架翻转	90.0	70.0	50.0	60.0	40.0	0.0
圆柱体插入	50.0	20.0	0.0	20.0	20.0	0.0
总体	**77.5**	**52.5**	**27.5**	**60.0**	**30.0**	**10.0**

观测转化为有效物理动作的难点，并确立了反射鲁棒性这一具体目标，用于评估和改进机器人策略。

致谢

本研究部分使用了 Google 的 TPU ResearchCloud（TRC）提供的云端TPU。

参考文献

[1] Awomo-WAM 团队。Awomo-0.5 技术报告。技术报告，Awomo，2026年。 https://github.com/Awomo-WestlakeDI/Awomo-0.5/blob/main/Awomo_0_5.pdf。 3

[2] Kevin Black、Noah Brown、Danny Driess、Adnan Esmail、Michael Robert Equi、Chelsea Finn、Niccolo Fusai、Lachy Groom、Karol Hausman、Brian Ichter、Szymon Jakubczak、Tim Jones、Liyiming Ke、Sergey Levine、Adrian Li-Bell、Mohith Mothukuri、Suraj Nair、Karl Pertsch、Lucy Xi-aoyang Shi、Laura Smith、James Tanner、Quan Vuong、Anna Walling、Haohuan Wang 和 Ury Zhilinsky。 π_0：一种用于通用机器人控制的视觉-语言-动作流模型。载于机器人学：科学与系统会议论文集，美国加利福尼亚州洛杉矶，2025年。 2

[3] Tianxing Chen、Yue Chen、Zixuan Li、Junyuan Tang、Kailun Su、Haoran Lu、Weijie Wan、Baijun Chen、Songling Liu、Haowen Yan、Honghao Su、Zhiyang Dou、Kaixuan Wang、Dandan Zhang、Yunze Liu、Yan Qin、Qiwei Liang、Qiwei Wu、Zijian Lin、Wenwei Lin、Yuran Wang、Minghua He、Tianshu Wu、Ruihai Wu、Jingquan Zhou、Kai-Chong Lei、Haibao Yu、Yuanfeng Ji、Weiyang Jin、Guanyu Lin、Xiao- fan Li、Qi Xiong、Renjing Xu、Zhongyu Li、Wenhao Chai、Enze Xie、Ziwei Wang、Yao Mu、Hao Dong、Wojciech Ma- tusik、Mingyu Ding、Wenbo Ding、Ping Luo 和 Masayoshi Tomizuka。《RoboDojo：用于全面评估通用机器人操作策略的统一仿真与真实基准》。 *arXiv* 预印本 *arXiv:2607.04434*，2026年。 2，3

[4] Tianxing Chen、ZanxinChen、Baijun Chen、Zijian Cai、Yibin Liu、Zixuan Li、Qiwei Liang、Xianliang Lin、Yiheng Ge、Zhenyu Gu、Weiliang Deng、Yubin Guo、Tian Nian、Xuan- bing Xie、Qiangyu Chen、Kailun Su、Tianling Xu、Guodong Liu、Mengkang Hu、Huan-Ang Gao、Kaixuan Wang、Zhix- uan Liang、Yusen Qin、Xiaokang Yang、Ping Luo 和 Yao Mu。《RoboTwin 2.0：用于稳健双手机器人操作的可扩展数据生成器和基准，具备强大的域随机化能力》，载于《国际机器学习会议论文集》，第 13673–13699 页，2026年。 2

[5] Cheng Chi、Siyuan Feng、Yilun Du、Zhenjia Xu、Eric Cousineau、Benjamin Burchfiel 和 Shuran Song。《扩散策略：通过动作扩散进行视觉运动策略学习》。载于 机器人学：科学与系统论文集，2023年。 7

[6] Cheng Chi、Zhenjia Xu、Chuer Pan、Eric Cousineau、Benjamin Burchfiel、Siyuan Feng、Russ Tedrake 和 Shuran Song。《通用操作接口：无需身处真实环境中的机器人，即可在真实环境中开展机器人教学》，2024年。 3

[7] Senyu Fei、Siyin Wang、Junhao Shi、Zihao Dai、Jikun Cai、Pengfang Qian、Li Ji、Xinzhe He、Shiduo Zhang、Zhaoye Fei、Jinlan Fu、Jingjing Gong 和 Xipeng Qiu。《LIBERO-Plus：视觉-语言-动作模型的渐进式鲁棒性基准》。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第38574–38583页，2026年。 2，3

[8] Pete Florence、Corey Lynch、Andy Zeng、Oscar Ramirez、Ayzaan Wahid、Laura Downs、Adrian Wong、Johnny Lee、Igor Mordatch 和 Jonathan Tompson。隐式行为克隆，2021年。 arXiv:2109.00137。 7

[9] Futian Lab 和 VisIncept。Rex-M1 Preview：XPolicy- Lab/RoboDojo 策略发布。XPolicyLab 发布文档。 https://github.com/XPolicyLab/XPolicyLab/blob/main/policy/Rex_M1_preview/README.md，2026年。访问日期：2026年11月13日。 3

[10] Gordon G. Gallup, Jr. 黑猩猩：自我识别。《科学》，167(3914)：86–87，1970年。 2

[11] Tianrui Guan、Fuxiao Liu、Xiyang Wu、Ruiqi Xian、Zongxia Li、Xiaoyu Liu、Xijun Wang、Lichang Chen、Furong Huang、Yaser Yacoob、Dinesh Manocha 和 Tianyi Zhou。HallusionBench：用于诊断大型视觉语言模型中交织的语言幻觉和视觉错觉的高级诊断套件。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第14375–14385页，2024年。 3

[12] Shengyu Guo、Tongrui Ye、Jianbo Zhang、Zicheng Zhang、Chunyi Li 和 Guangtao Zhai。MirrorBench：通过引入一面镜子，评估多模态大语言模型中的自我中心智能。*arXiv* 预印本 *arXiv:2604.14785*，2026年。 2，3

[13] Yanjiang Guo、Haodong Yan、Zhide Zhong、Zhongru Zhang、Qingyuan Yang、Qingzhou Lu、Xiaoyu Chen、Yen-Jen Wang、Shuying Deng、Chenghan Yang、Puzhen Yuan、Chenxin Liu、Tun Ban、Xiang Zhu、Yichen Liu、Kun Feng、Haoang Li 和 Jianyu Chen。视频预测策略 2：预测更佳，执行更优，2026年。 arXiv:2610.10270。 3

[14] Justin W. Hart 和 Brian Scassellati。《人形机器人中的镜像视角采择》。载于*AAAI*人工智能会议论文集，第1990–1996 页，2012 年。 3

[15] Mengzhao Jia、Yang Lin、Xixin Zhang、Zhihan Zhang、Xiaobai Liu 和 Meng Jiang。《面向机器人操作的智能体作为策略》，2026 年。 2

[16] Pablo Lanillos、Jordi Pages 和 Gordon Cheng。《机器人自我/他者区分：主动推断与在镜子中学习的神经网络相遇》。载于《欧洲人工智能会议论文集》，第 2410–2416 页，2020 年。3[17] Huiqiong Li、Zhiting Mei、Anirudha Majumdar、Jingjing Chen、Yu-Gang Jiang 和 Bin Zhu。《LIBERO-VPro：机器人基础模型闭环视觉鲁棒性基准测试》。arXiv 预印本 arXiv:2609.24350，2026 年。3[18] Zhiyuan Li、Charles Eret、Xuan Zhao、Yangzheng Wu 和 Amir Rasouli。《鲁棒机器人操作的数据与评估协议》。载于机器人学：科学与系统会议的以数据为中心的机器人学研讨会，2026 年。3[19] Jiaying Lin、Xin Tan 和 Rynson W. H. Lau。《通过双重对应关系从视频中学习检测镜子》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 9109–9118 页，2023 年。3[20] Bo Liu、Yifeng Zhu、Chongkai Gao、Yihao Feng、Qiang Liu、Yuke Zhu 和 Peter Stone。《LIBERO：终身机器人学习知识迁移基准测试》。载于《神经信息处理系统进展》，第 44776–44791 页，2023 年。2 – 。[21] Jingping Liu、Xingchen Peng、Yan Zhou、Ziyan Liu、Jie Zhai、Ronghao Chen、Huacan Wang 和 Xiaofeng Jia。《MirrorQA：多模态大语言模型镜像朝向推理基准测试》。载于计算语言学协会第 64 届年会论文集（第 1 卷：长篇论文），第 40464 40476, 2026 3 – 页。Eric [22] Marchand 和 Francois Chaumette。《通过镜面反射实现视觉伺服》。载于 IEEE机器人与自动化国际会议，2017 年。3[23] OpenAI。《GPT-6 Astra：新一代智能》。官方模型公告。https://openai.com/index/gpt-6 13 -astra/，2026 年。访问日期：11 月。2026 年。3[24] Physical Intelligence、Kevin Black、Noah Brown、James Darpinian、Karan Dhabalia、Danny Driess、Adnan Esmail、Michael Equi、Chelsea Finn、Niccolo Fusai、Manuel Y. Galliker、Dibya Ghosh、Lachy Groom、Karol Hausman、Brian Ichter、Szymon Jakubczak、Tim Jones、Liyiming Ke、Devin LeBlanc、Sergey Levine、Adrian Li-Bell、Mohith Mothukuri、Suraj Nair、Karl Pertsch、Allen Z. Ren、Lucy Xiaoyang Shi、Laura Smith、Jost Tobias Springenberg、Kyle Stachowicz、James Tanner、Quan Vuong、Homer Walke、Anna Walling、Haohuan Wang、Lili Yu 和 Ury Zhilinsky。《$\pi_{0.5}$：具备开放世界泛化能力的视觉-语言-动作模型》，2025 年。arXiv:2504.16054。8[25] Haz Sameen Shahgir、Khondker Salman Sayeed、Abhik Bhattacharjee、Wasi Uddin Ahmad、Yue Dong 和 Rifat Shahriyar。《IllusionVQA：面向视觉语言模型的高难度视觉错觉数据集》。载于《语言建模会议论文集》，2024 年。3[26] Jiaqi Tan、Weijie Lin、Angel X. Chang 和 Manolis Savva。《Mirror3D：镜面的深度细化》。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 15990–15999 页，2021 年。3[27] X Square Robot。《WALL-OSS 1.0 Lite — RoboDojo》。Hugging Face 模型卡。https://huggingface.co/x-square-robot/wall-oss-1.0-lite-robodojo，2026年。访问日期：2026年11月13日。3[28] Xin Yang、Haiyang Mei、Ke Xu、Xiaopeng Wei、Baocai Yin 和 Rynson W. H. Lau。《我的镜子在哪里？》载于*IEEE/CVF* 国际计算机视觉会议论文集，第 8809–8818 页，2019 年。3[29] Seonghyeon Ye、Yunhao Ge、Kaiyuan Zheng、Shenyuan Gao、Sihyun Yu、George Kurian、Suneel Indupuru、You Liang Tan、Chuning Zhu、Jiannan Xiang、Ayaan Malik、Kyungmin Lee、William Liang、Nadun Ranawaka、Jiasheng Gu、Yinzhen Xu、Guanzhi Wang、Fengyuan Hu、Avnish Narayan、Johan Bjorck、Jing Wang、Gwanghyun Kim、Dan- tong Niu、Ruijie Zheng、Yuqi Xie、Jimmy Wu、Qi Wang、Ryan Julian、Danfei Xu、Yilun Du、Yevgen Chebotar、Scott Reed、Jan Kautz、Yuke Zhu、Linxi Jim Fan 和 Joel Jang。《世界动作模型是零样本策略》，2026 年。2[30] Hao Yin、Tianchen Guo、Heming Du、Yan Ke、Yan- bin Liu 和 Xin Yu。《当深度说谎：对视觉语言模型在镜面引发的 RGB-D 歧义上的基准评测》。 *arXiv* 预印本 *arXiv:2610.41651*，2026 年。3[31] Kentaro Yoshioka、Hidenori Okuni、Tuan Thanh Ta 和 Aki- hide Sai。《透过镜面看世界：利用镜面反射减少机器人视觉系统中的遮挡》，2021 年。3[32] Tianyuan Yuan、Zibin Dong、Yicheng Liu 和 Hang Zhao。《Fast-WAM：世界动作模型是否需要测试时未来想象？》，2026 年。arXiv:2603.16666。8[33] Junpeng Yue、Boyuan Li、Yuxuan Wang、Zepeng Wang、Yuhui Fu、Feiyang Xie、Yu Zhang、Jing Zhang、Xianqi Zhang、Weibo Li、Xiaofei Zheng、Yuming Fang、Jiangx- ing Wang 和 Zongqing Lu。《Being-M0.7：面向人形机器人的潜在世界动作模型》。 *arXiv* 预印本 *arXiv:2610.11283*，2026 年。3[34]Wenbo Zhang、Kaixuan Wang、Yutao Ouyang、Xiaoyu Huang、Liyang Li、Kailun Su、Weiyang Jin、Wenhao Chai、Haotian Liang、Zhiyang Dou、Yue Chen 和 Tianxing Chen。《出乎意料的机器人策略：对 GPT-6 Astra 在 RoboDojo 及其他环境中的早期评估》，2026 年。3[35] Xueyang Zhou、Yangming Xu、Qianjiang Li、Guiyao Tie、Junhao Wu、Ziheng Huang、Guowen Zhang、Zhijie Yuan、Di Wu、Duanfeng Chu、Pan Zhou、Lichao Sun 和 Yongchao Chen。《LIBERO-Pro2.0：用于评估具身基础模型鲁棒性的基准》。 *arXiv* 预印本 *arXiv:2610.24587*，2026 年。3[36] Filippo Ziliotto、Ciro Beneduce、Bruno Lepri、Luciano Serafini、Massimiliano Luca 和 Tommaso Campari。《镜子，墙上的镜子：视觉语言模型代理究竟能否辨认自己？》 *arXiv* 预印本 *arXiv:2605.08816*，2026 年。3

具身人工智能未能通过 镜子测试：简单的镜子如何破坏机器人策略

补充材料

概述

本补充材料记录了场景几何、策略配置和评估指标（附录 A）；任务级结果（附录B）；镜面板对照和接触测量（附录 C）；桌面抓取、Push-T 和目标遮挡诊断（附录 D、E和F）；以及现实世界部署详情（附录 G）。

A. 场景几何与评估详情

A.1. 镜像几何与虚像

两面镜子可以形成一阶像（由单次反射形成）和高阶像（由光线在镜子之间连续反射形成）。图 7 展示了三种理想化镜面配置下的虚像排列。图中所示的 90°、60°和 45° 配置分别会形成三个、五个和七个虚像。

我们的主要双镜评估使用相互垂直的镜子（90°）。其他角度用于说明反射几何，而非表示额外评估条件。在实验场景中，镜子的有限尺寸、摄像机视角、遮挡和渲染设置决定了策略观测中实际可见的虚像。

A.2. 策略配置

我们评估了四种预训练机器人策略：WALL-OSS- 1.0-Lite、VPP2、Awomo-0.5 和 Rex-M1Preview，以及使用工具的 GPT-6-Astra 智能体。对于这四种预训练策略，我们使用其官方发布的检查点及相应的推理配置，不进行额外训练或针对镜子的适配。

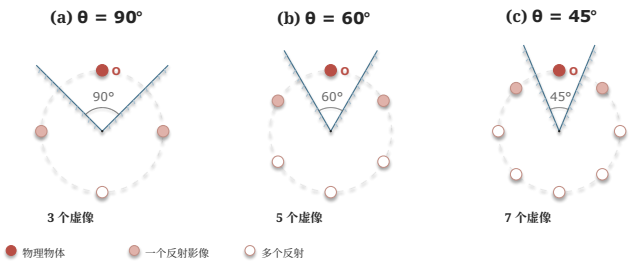


图7. 两面镜子夹角分别为 (a) 90°、(b) 60° 和 (c) 45° 时的理想化虚像排列。实心红色标记表示物理物体；浅色实心标记表示一阶像；轮廓标记表示高阶像。这些示意图展示的是反射几何，而非相机可见的像的数量。

对于每种策略，在不同场景条件下，我们都固定检查点、观测预处理、相机输入、动作表示和推理设置。任务指令和原生成功标准也保持不变。在主要评估中，不会将场景条件作为额外的策略输入。与镜子相关的指令变体仅用于目标遮挡诊断。

保留各策略专属的推理和控制设置，而不强行采用统一的动作接口。因此，比较主要衡量相对于各策略自身无镜基线的变化。此外，GPT-6-Astra 还采用明确的机器人工具接口和决策调用预算，详情如下。

表7 总结了策略配置。预测时域表示每次推理生成的动作步数，而重新规划间隔表示在请求新的预测之前执行的步数。

WALL-OSS 对参数选择性地转换为 BF16，同时保留 FP32 组件。Awomo-0.5 使用原始 RoboDojo 检查点。VPP2 使用 RoboDojo EE16 部署适配器；检查点名称并不表示实际执行的动作表示。GPT-6-Astra 会选择单个运动工具动作，并采用可变的执行持续时间，而不是预测固定长度的动作块。其推理强度设为中等，这与数值精度不同。

A.2.1. GPT-6-Astra 智能体配置

我们将 GPT-6-Astra 配置为使用 RoboDojo Inspect-EEF 接口的闭环工具使用智能体。我们通过 Codex CLI 的应用服务器接口访问该模型，并使用 ChatGPT 订阅认证，而不是直接通过 API 密钥认证访问 OpenAI API。请求使用的模型为 gpt-6-astra，推理强度设为中等。在每个观测边界，模型都会接收来自头部相机和两个腕部相机的 RGB 图像、机器人本体感知状态以及执行反馈。不会提供深度观测和模拟器中的真实物体坐标。

智能体使用两个工具。move eef 用于指定一侧或两侧机械臂在世界坐标系中的末端执行器位置、方向和夹爪开合状态。原生执行器会验证目标，使用 cuRobo 规划关节轨迹，并通过模拟器的机器人控制器执行这些轨迹。随后，智能体会收到执行反馈和更新后的观测结果。放弃工具允许智能体终止一次未成功的尝试。

我们保留上游 Inspect-EEF 提示词和任务

表 7. 评估中使用的策略配置。H+W 表示头部和双腕部摄像头。EE 表示末端执行器位姿和夹爪目标；Joint 表示关节位置和夹爪目标。预测时域和重新规划间隔以控制步数计量。精度描述的是策略推理，而非模拟器算术运算。

策略	检查点 / 模型	RGB摄像头	动作	预测 预测时域	重新规划 间隔	精度
WALL-OSS-1.0-Lite	wall-oss-1.0-lite-RoboDojo	H+W	EE	32	16	BF16/FP32
VPP2	joint2b s100000	H+W	EE	32	24	BF16
Awomo-0.5	RoboDojo 发布版 model.pt	H+W	联合	16	8	BF16
Rex-M1 Preview	Rex-M1-预览版 last.ckpt	H+W	EE	30	15	BF16
GPT-6-Astra	gpt-6-astra	H+W	EE 工具	变量	取决于工具	未披露

专用配方。这些配方提供静态任务描述和评分标准，不会暴露实时物体状态、当前得分或成功标记。在回合过程中，不会提供额外示范、先前尝试的轨迹或人工指导。

每个回合都会启动一个全新的 Codex 会话。后续决策会在同一会话中继续进行，并接收新的观测和执行反馈。回合内的历史记录由原生会话管理，不会跨回合传递历史记录。尽管本地策略缓冲区保留两组观测，但这并不意味着模型可见的会话历史仅限于两组观测。

每个回合最多允许 100 次智能体决策调用，并且仍受原生任务的控制步预算和终止规则约束。决策调用次数上限不同于低层控制步数。我们在不同场景条件下保持智能体配置、工具接口和预算不变。任务结果由原生环境评估器确定。

A.3. 实验覆盖范围和测量来源

表 8 总结了所报告实验的样本结构和测量来源。一个单元格确定策略、任务（或 Push-T 轮廓）以及实验条件；布局和重复次数在该单元格内计数。场景条件、指令变体以及碰撞/补丁设置构成独立的单元格。布局在同一策略和任务内成对对应；布局数量相同并不意味着不同策略之间的初始状态相同。

A.4. 配对评估和指标

每个主评估任务—条件单元格包含五个布局；对于 WALL-OSS、VPP2、Awomo-0.5 和 Rex-M1 Preview，使用三个策略随机种子（15 个回合），而对于 GPT-6-Astra，则包含五个布局，每个布局执行一次（五个回合）。匹配在同一策略和任务内进行，而不是跨策略进行。

原生评估器为每个回合分配一个二元成功指标和一个任务得分，$[0, 1]$ 用于体现部分完成情况。对于每个策略、任务和场景条件，我们以百分比报告成功率，并将平均任务得分缩放至 0–100。对于每个策略和场景条件，我们对八项任务的这两个指标分别进行宏平均，使每项任务的权重相等。在同一任务上，不同策略的无镜基线表现可能有所不同。任务级表格报告了八项任务中各策略对应的基线。

对于给定指标 v （得分或成功率），令 $\bar{v}_{m,t,c}$表示策略 m在任务 t 和场景条件c下的平均值。主表报告了 $\mathcal{T}$ 中八个任务以及 $\mathcal{M}$中的三个镜像条件下，相对于无镜像基线 B 的平均配对下降：

$$D_m(v) = \frac{1}{3|\mathcal{T}|} \sum_{c \in \mathcal{M}} \sum_{t \in \mathcal{T}} (\bar{v}_{m,t,B} - \bar{v}_{m,t,c}) . \qquad (7)$$

正值表示镜像条件下性能较低。成功率下降以百分点表示。

我们使用 20,000 次配对自助法重复抽样，报告平均成功率下降的 95% 百分位置信区间。在每次重复抽样中，我们针对每项任务有放回地抽取五种布局。我们保留与每种抽取布局相关的所有运行结果，并在不同场景条件下使用相同的布局抽样，以保留配对关系。因此，这些区间反映了在所评估任务集合中不同布局之间的变异。

达到任务预算的执行仍算作已评估回合；接触并不是额外的终止条件。我们仅将原始策略执行计为试验。对于 Astra 批次，保留的技术中断不会被独立计为任务失败；获接受的已完成恢复会占用其原始评估名额，不计为额外的重复次数。

表 8。样本结构与测量来源。 n 统计每个单元格中的原始策略执行次数。

实验与范围	策略	布局 每个单元格	每项的运行次数 布局	n	测量来源
主要评估：八项任务，四个场景	WALL-OSS、VPP2、Awomo-0.5, Rex-M1 Preview	5	3	15	原生任务评估器；策略特定接触记录
	Astra	5	1	5	原生任务评估器和记录的执行过程
镜面板控制：八个任务，两个面板设置	WALL-OSS、Awomo-0.5	5	3	15	原生任务评估器；物理子步机器人–表面接触
桌面诊断：三个任务，两个条件	VPP2、Awomo-0.5、WALL-OSS	5	3	15	原生结果；记录的轨迹以及补充模拟器测量
	Astra	5	1	5	原生结果；20 个回合的原始轨迹，以及 10 个回合的模拟器测量
Push-T 贴片/碰撞控制：三个轮廓，四种设置	VPP2	1	5	5	物体轨迹、投影视轮廓，以及机器人–镜面接触记录
Push-T 放置诊断：三个轮廓	VPP2、WALL-OSS	5	3	15	物体轨迹、投影视轮廓，以及机器人–镜面接触记录
	Astra	5	1	5	物体轨迹、投影视轮廓，以及机器人–镜面接触记录
目标遮挡：一个任务，三条指令	WALL-OSS、Awomo-0.5	5	3	15	原生结果；模拟器测量的目标记录的动作所产生的接触
	Astra	5	1	5	原生结果；模拟器测量的目标来自记录的动作的接触
真实世界评估：四项任务，三个场景	$\pi_{0.5}$，FastWAM	5	2	10	记录的任务结果；与策略相关包括执行失败

B. 任务覆盖范围和详细结果

B.1. 常见任务集

我们选择了表 9中的八项 RoboDojo任务，以涵盖多样的操作要求，同时优先选择在 RoboDojo排行榜上总体性能较强的任务。这些任务涵盖物体拾取与重新排列、堆叠、插入、折叠布料，以及根据数值指令按下按钮。在我们的无镜子评估中，八项任务中的七项至少有五个策略中的三个取得了 60% 或更高的成功率。对于单项任务，不同策略的基线性能有所差异。

我们在不同策略和场景条件下使用相同的任务集，并保留每项任务的原始指令、目标和原生成功标准。图8 展示了四种条件下的初始场景。

B.2. 任务级性能

Astra 表格汇总了已接受的特定任务批次。覆盖方块、叠衣服、按数字按压、导管插入和碗堆叠使用布局 10–14；常规拾取、交换 T 和方块堆叠使用布局 0–4。每项任务中的条件均成对匹配。叠衣服和按数字按压条目使用更新后的匹配批次，而非较早的运行结果。历史执行与更新后的执行不会合并为额外的重复次数。

表9。通用评估集中的八项任务。

Task	RoboDojo 标识符
覆盖方块	覆盖方块
叠衣服	叠衣服
一般拾取	一般拾取_
按数字按压	按数字按压
导管插入	插入导管
交换 T	交换 T
方块堆叠	堆叠方块
碗堆叠	叠放碗_

C. 接触分析与镜子—面板控制

C.1. 接触定义与位置

我们根据涉及的实体，对与实验镜或非反射面板的接触进行分类：夹爪（手掌和手指）、其他机器人本体（机械臂连杆，以及已登记的腕部相机本体）或任务物体。我们排除机器人自接触、机器人—物体接触以及与普通桌面的接触。对于接触记录器覆盖相关实体的情况，我们报告每个类别。

表 15 区分了夹爪接触与其他本体接触，以展示在八项任务基准中接触发生的位置。对于

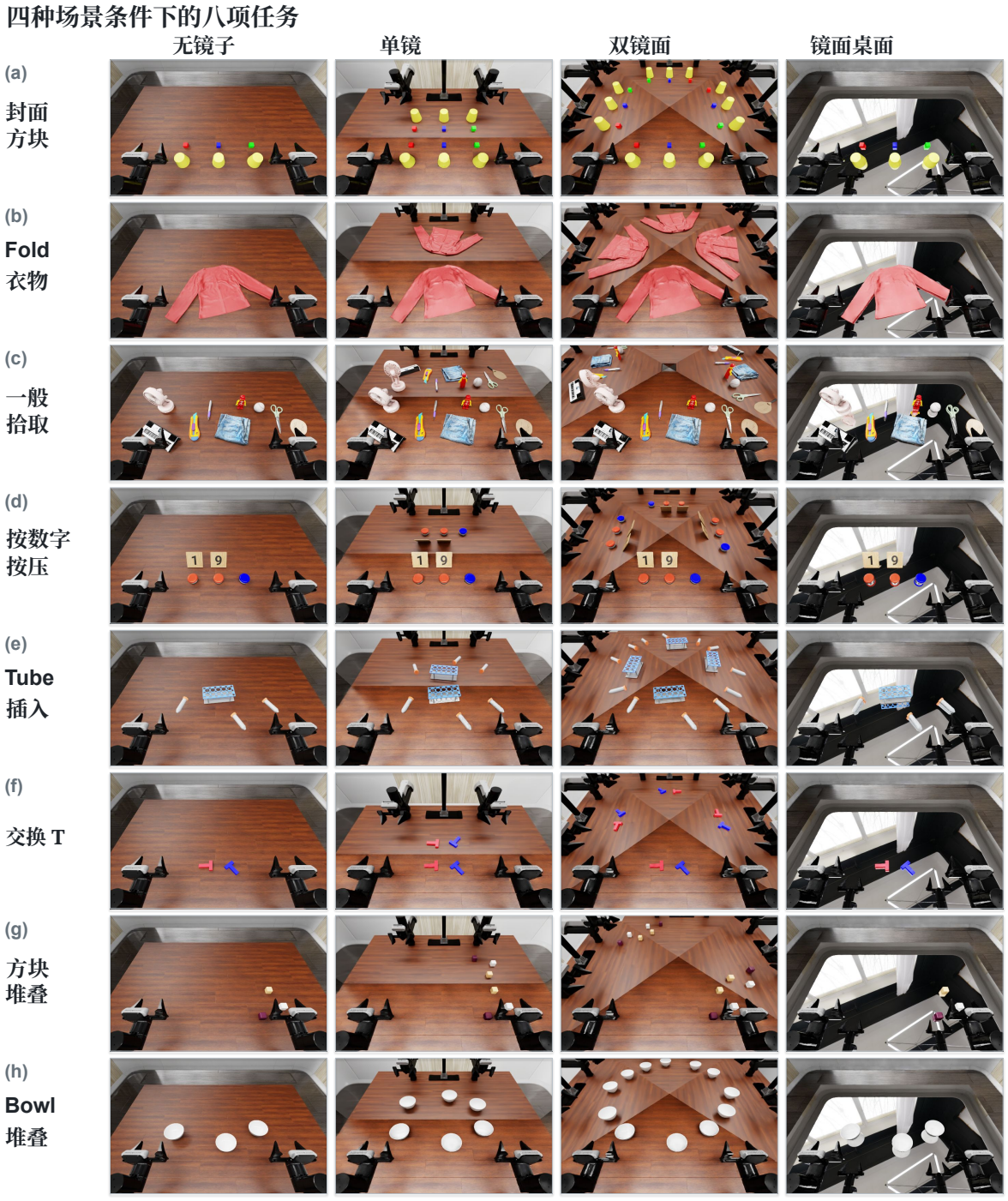


图8。四种场景条件下的八项常见评估任务：无镜子、单镜、双镜面和镜面桌面。每行展示 WALL-OSS 在布局 0 和策略种子 0 下的初始头部相机视图，各条件下的机器人和物体初始状态一致。所有图像使用相同的裁剪区域。

表10. WALL-OSS-1.0-Lite 任务级结果。条目报告得分/成功率（%）。每个任务–条件单元包含15个回合。

Task	No 镜子	单 镜子	Dual 镜子	镜像 桌面
覆盖方块	100.0 / 100.0	0.0 / 0.0	1.0 / 0.0	0.0 / 0.0
叠衣服	89.3 / 86.7	69.3 / 66.7	65.3 / 60.0	78.7 / 73.3
一般拾取	73.3 / 73.3	33.3 / 33.3	20.0 / 20.0	60.0 / 60.0
按数字按压	6.7 / 6.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
导管插入	36.0 / 20.0	0.0 / 0.0	0.0 / 0.0	5.3 / 0.0
交换 T	100.0 / 100.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
方块堆叠	66.0 / 60.0	6.0 / 0.0	6.0 / 0.0	7.7 / 6.7
碗堆叠	93.3 / 93.3	77.3 / 73.3	45.0 / 40.0	66.0 / 60.0
平均值	70.6 / 67.5	23.3 / 21.7	17.2 / 15.0	27.2 / 25.0

表11。VPP2任务级结果。条目报告得分 / 成功率（%）。每个任务–条件单元包含15个回合。

Task	No 镜子	单 镜子	Dual 镜子	镜像 桌面
覆盖方块	100.0 / 100.0	0.0 / 0.0	42.0 / 20.0	72.0 / 60.0
叠衣服	82.7 / 80.0	72.0 / 66.7	100.0 / 100.0	74.7 / 73.3
一般拾取	6.7 / 6.7	20.0 / 20.0	13.3 / 13.3	0.0 / 0.0
按数字按压	93.3 / 93.3	60.0 / 60.0	73.3 / 73.3	46.7 / 46.7
导管插入	96.0 / 93.3	13.3 / 0.0	12.0 / 0.0	38.7 / 13.3
交换 T	6.7 / 6.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
方块堆叠	100.0 / 100.0	46.0 / 40.0	15.7 / 6.7	6.0 / 0.0
碗堆叠	100.0 / 100.0	83.0 / 80.0	70.7 / 66.7	36.7 / 26.7
平均值	73.2 / 72.5	36.8 / 33.3	40.9 / 35.0	34.3 / 27.5

表12。Rex-M1 Preview 任务级结果。条目报告得分 / 成功率（%）。每个任务–条件单元格包含15个回合。

Task	No 镜子	单 镜子	Dual 镜子	镜像 桌面
覆盖方块	100.0 / 100.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
叠衣服	93.3 / 93.3	49.3 / 40.0	56.0 / 46.7	16.0 / 6.7
一般拾取	46.7 / 46.7	53.3 / 53.3	26.7 / 26.7	6.7 / 6.7
按数字按压	100.0 / 100.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
导管插入	70.7 / 66.7	0.0 / 0.0	1.3 / 0.0	0.0 / 0.0
交换 T	20.0 / 20.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
方块堆叠	36.7 / 26.7	0.0 / 0.0	3.0 / 0.0	0.0 / 0.0
碗堆叠	83.0 / 80.0	41.3 / 33.3	18.7 / 6.7	31.0 / 20.0
平均值	68.8 / 66.7	18.0 / 15.8	13.2 / 10.0	6.7 / 4.2

表13。Awomo-0.5 任务级结果。条目报告得分 / 成功率（%）。每个任务–条件单元包含15个回合。

Task	No 镜子	单 镜子	Dual 镜子	镜像 桌面
覆盖方块	100.0 / 100.0	0.0 / 0.0	0.7 / 0.0	19.7 / 13.3
叠衣服	88.0 / 86.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
一般拾取	66.7 / 66.7	40.0 / 40.0	26.7 / 26.7	20.0 / 20.0
按数字按压	26.7 / 26.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
导管插入	78.7 / 73.3	57.3 / 46.7	6.7 / 0.0	46.7 / 33.3
交换 T	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
方块堆叠	70.7 / 66.7	2.0 / 0.0	2.0 / 0.0	0.0 / 0.0
碗堆叠	70.7 / 66.7	43.0 / 40.0	27.7 / 26.7	17.3 / 13.3
平均值	62.7 / 60.8	17.8 / 15.8	8.0 / 6.7	13.0 / 10.0

表14. GPT-6-Astra 任务级结果。各项报告得分 / 成功率(%)。每个任务–条件单元包含5个回合。

Task	No 镜子	单 镜子	Dual 镜子	镜像 桌面
覆盖方块	46.0 / 40.0	3.0 / 0.0	46.0 /40.0	29.0 / 20.0
叠衣服	100.0 / 100.0	100.0 / 100.0	52.0 /40.0	68.0 / 60.0
一般拾取	100.0 / 100.0	100.0 / 100.0	100.0 /100.0	60.0 / 60.0
按数字按压	80.0 / 80.0	100.0 / 100.0	80.0 /80.0	60.0 / 60.0
导管插入	4.0 / 0.0	0.0 / 0.0	12.0 / 0.0	4.0 / 0.0
交换 T	20.0 / 20.0	20.0 / 20.0	20.0 /20.0	40.0 / 40.0
方块堆叠	100.0 / 100.0	100.0 / 100.0	83.0 /80.0	83.0 / 80.0
碗堆叠	80.0 / 80.0	83.0 / 80.0	63.0 /60.0	66.0 / 60.0
平均值	66.3 / 65.0	63.3 / 62.5	57.0 /52.5	51.3 / 47.5

对于每个类别，如果一个回合至少包含一条该类型的记录接触，则计数一次。对于给定配置，一个回合可以同时计入两列。

对于每个策略，我们使用与表1中相同的接触记录。如果夹爪或其他机器人本体与表面发生接触，则将该回合计为发生了机器人–表面接触。在可用的情况下，任务对象接触会被单独跟踪，并从该指标中排除。

C.2. 匹配镜面–面板对照

为考察直接机器人–表面接触是否能够解释所观察到的结果差异，我们使用 WALL-OSS 和 Awomo-0.5，将镜子与几何匹配的非反射面板进行比较。面板采用浅色哑光花岗岩纹理。双面板条件中的黑色接缝是视觉特征。我们不施加特定于条件的物理材质、摩擦力或恢复系数覆盖设置。

在每个策略内部，镜子条件和面板条件使用匹配的任务布局与策略随机种子。我们固定检查点、任务指令、相机配置和执行预算。每个策略在八个任务上进行评估，每个任务包含五种布局和三个随机种子，因此每种条件下共有 120 个回合。对照涵盖单垂直镜面配置和双垂直镜面配置。图9展示了这些设置。

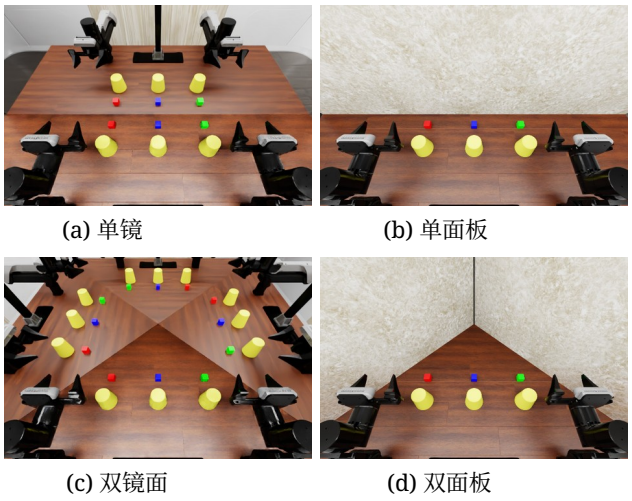
配对对照的接触检测。对于 WALL-OSS 和 Awomo-0.5，我们在每个记录的物理子步（$\Delta t = 0.004$ s）查询 PhysX GPU 刚体接触视图。当已注册机器人刚体与实验表面之间至少有一个报告的接触点的法向力大于 10^{-6}N 时，我们记录为发生接触。

令 $c_{i,t}^{q} \in \{0,1\}$ 表示回合 i的物理子步 t中机器人与表面的接触，其中 $q \in \{M, P\}$表示镜子或面板条件。对于双表面，若该子步与任一表面发生接触，则 $c_{i,t}^{q}$ 为 1。回合级指示变量为

$$h_i^q = \max_t c_{i,t}^q. \tag{8}$$

表15. 八任务基准中的接触位置。条目给出在该策略和配置下，至少有一次指定类型记录的接触的回合数，占回合总数的比例。夹爪与其他身体部位的计数可能重叠。

	单镜		双镜面	
策略	夹爪	其他机器人本体	夹爪	其他机器人本体
WALL-OSS	28/120	40/120	27/120	30/120
VPP2	32/120	55/120	53/120	52/120
Awomo-0.5	15/120	29/120	32/120	32/120
Rex-M1 Preview	11/120	14/120	52/120	40/120
GPT-6-Astra	9/40	20/40	30/40	33/40



(a) 单镜 (b) 单面板

(c) 双镜面 (d) 双面板

图9. 覆盖方块中的镜子和非反射面板对照：(a) 单镜，(b) 单面板，(c) 双镜，(d) 双面板。每个对照均保留相应的镜像几何，并改变表面外观。

任务级接触次数。使用 WALL-OSS 和 Awomo-0.5 的所有匹配镜面–面板运行数据，我们比较八项任务中夹爪及其他物体类别的接触事件次数。对于每种物体类别，我们将每段连续发生接触的物理子步计为一个事件。在双表面配置中，我们先合并与任一表面的接触，再统计事件。表 16 汇总了两种策略以及单、双表面配置的数据；每项任务各有 60 个镜面回合和 60 个面板回合。在镜面条件下，方块堆叠任务在两种物体类别中的事件次数均为最高。

配对子集。 对于每个匹配任务、布局和策略随机种子，令 y_i^B、y_i^M 和 y_i^P 分别表示无镜子基线、镜子条件和面板条件下的原生任务成功率，

表16. 匹配镜像–面板评估中各任务的夹爪与其他机器人本体接触事件。条目报告镜子 / 面板的计数。任务按镜子条件下的夹爪接触次数排序。

Task	夹爪	其他机器人本体
方块堆叠	454/264	526/385
碗堆叠	240/107	390/218
覆盖方块	239/87	355/476
交换 T	230/306	188/224
导管插入	57/225	95/138
叠衣服	26/166	4/16
一般拾取	0/1	0/4
按数字按压	0/0	0/0

我们定义

$$\begin{aligned} \mathcal{F} &= \{i : y_i^B = y_i^P = 1,\ y_i^M = 0\}, \\ \mathcal{N} &= \{i : y_i^B = 1,\ h_i^M = h_i^P = 0\}, \\ \mathcal{C} &= \{i : y_i^B = 1,\ h_i^M = h_i^P = 1\}, \\ \mathcal{F}_{00} &= \mathcal{F} \cap \mathcal{N}. \end{aligned} \tag{9}$$

这里，$\mathcal{F}$包含镜像特定失败。在无镜像基线成功的配对中，$\mathcal{N}$包含两种条件下均无记录的机器人–表面接触的配对，而 $\mathcal{C}$包含两种条件下均有记录接触的配对。表17 给出了各子集的分母和任务结果。

C.3. 接触持续性

对于每个回合，我们将接触时间比例计算为存在机器人–表面接触的物理子步数除以已执行物理子步的总数。在$c_{i,\,t}^{q}$ 对各表面的并集上进行统计，可确保同时与两个面板接触时仅对该比例贡献一次。

表 18 报告了其中每回合平均占比$\mathcal{F} \cap \mathcal{C}$，即两种条件下均记录到机器人与表面接触的镜像特定失败。该比率以镜子条件下的平均值除以面板条件下的平均值。

表17。匹配的镜面–面板比较的子集计数。在 $\mathcal{N}$ 中，仅镜面 / 仅面板分别统计仅在镜子条件或仅在面板条件下失败的配对。最后几列报告 $\mathcal{C}$ 中的成功次数。

策略	配置	$\lvert\mathcal{F}_{00}\rvert/\lvert\mathcal{F}\rvert$	$\lvert\mathcal{N}\rvert$	仅镜面 / 仅面板	$\lvert\mathcal{C}\rvert$	M 成功	P 成功
WALL-OSS	单	7/8	41	7/2	36	0/36	0/36
	Dual	11/22	35	11/1	17	3/17	8/17
Awomo-0.5	单	9/13	47	9/5	20	0/20	2/20
	Dual	20/34	40	20/2	17	0/17	7/17

表18。 $\mathcal{F}\cap\mathcal{C}$内的接触时间分数。M 和 P 为回合级平均百分比。比率使用未舍入数值计算。“–”表示空子集。

策略	配置	n	M (%)	P (%)	M/P
WALL-OSS	单	0	–	–	–
	Dual	6	44.46	24.03	1.85
Awomo-0.5	单	2	12.18	10.93	1.11
	Dual	7	32.26	4.72	6.83

D. 镜像桌面抓取诊断

D.1. 诊断设置

我们在导管插入、方块堆叠和瓶子处理任务上分析 VPP2、Awomo-0.5、WALL-OSS-1.0-Lite 和 GPT-6-Astra。瓶子处理是一项额外的诊断任务，不属于主要的八任务集合 $\mathcal{T}$。VPP2、Awomo-0.5 和 WALL-OSS 的每个任务—条件单元包含 15 个回合，Astra 则包含五个。每种策略下，原始条件与镜面桌面条件按布局配对；对于前三种策略，还进一步匹配了策略随机种子。两个条件下共进行了 300 次执行，形成 150 对匹配对。

任务成功取自原始原生评估结果。行为测量通过保存的命令和物理轨迹获得。对于 VPP2，我们还通过在仿真中执行记录的动作来收集参考测量结果。补充分析不构成新的策略评估。

D.2. 行为指标定义

图 10根据以下定义展示了物体跟随以及几何空抓合拢抬升的情况。

候选闭合区间。我们根据实际执行的、归一化后的夹爪打开指令，分别为每个机械臂识别候选区间。当打开指令达到至少0。80后，第一个小于或等于0。75 的采样点即开始一个候选区间。当指令重新达到至少0。80时，区间在此之前的最后一个采样点结束；若指令始终未达到重新打开阈值，则区间在该回合的最后一个采样点结束。图11展示了这些规则如何识别出记录的一个插管回合中有五个候选区间。

合格抬升。 在候选区间内，我们搜索相对于此前运行最小值至少增加 20毫米的 TCP 高度增量。该高度增量必须在连续三个记录的控制采样点中持续存在。该标准描述的是 TCP 的运动；并不要求物体被抬起。

令 $I=[s,e]$ 表示候选闭合区间，令 z_t表示记录采样点 t处的 TCP高度。定义运行最小值

$$m_t=\min_{u\in[s,t]} z_u. \tag{10}$$

我们在 I内测试连续的三采样点窗口。对于从 t开始的每个窗口，我们对全部三个采样点使用相同的运行最小值 m_t 作为参考：

$$L(I)=\mathbf{1}\Big[\exists\, t\in\{s,\dots,e-2\}:\ \min_{r\in\{0,1,2\}}(z_{t+r}-m_t)\geq 0.02\,\mathrm{m}\Big]. \tag{11}$$

如果 I 少于三个样本， $L（I）=0$。

几何空夹闭合并抬升（ECL-G）。我们根据实际记录的手掌和手指姿态，重建一个定向感知盒。其横向范围位于手指内侧面之间，纵向和垂直范围则取手指碰撞网格的共有跨度。该区域包含 1 毫米的边界容差。

如果候选项包含一次合格抬升，且从候选项开始到确认抬升期间的任何样本中，目标碰撞包络都未与感知盒相交，则将其标记为 ECL-G。即使物体随后离开该区域，先前曾占据该区域也会使其无法被标记为空。

令 O_u 表示在样本 u处是否有任何目标碰撞包络与夹爪感知区域相交。若 $L(I)=1$，则令 $t^\star$ 为最早的合格三样本窗口中的第一个样本，并令 $c=t^\star+2$ 为其确认样本—

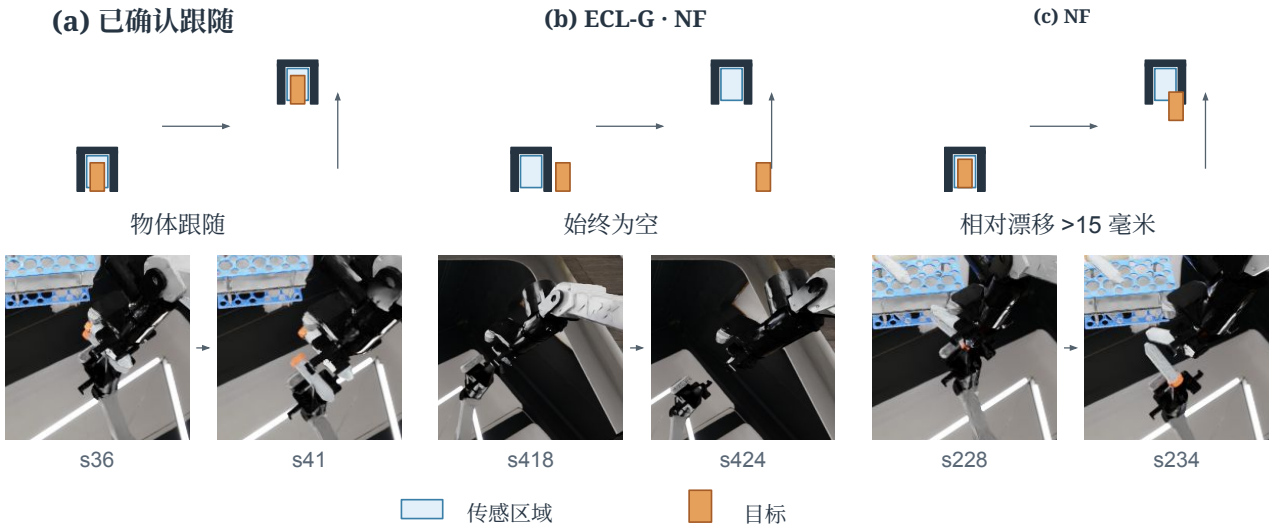


图10. 物体跟随和空抓合拢抬升案例示意。三个示例均包含符合条件的 TCP 抬升。(a) 确认物体跟随。(b) 几何空夹闭合并抬升（ECL-G），同时也符合 NF。(c) 连续目标占据，且相对于手掌坐标系的位移超过 15 毫米，因此无法确认物体跟随（NF）。ECL-G 是 NF 的子集。相机图像展示记录的案例；指标标签由轨迹测量确定。

例。我们定义

$$\text{ECL-G}(I) = \begin{cases} \mathbf{1}[O_u = 0 \,\forall\, u \in \{s, \ldots, c\}], & L(I) = 1, \\ 0, & L(I) = 0. \end{cases} \tag{12}$$

无抬升（NL）。如果候选项在重新张开或终止前没有出现符合条件的抬升，则标记为 NL：NL(I) $=1 - L(I)$。在回合层面，NL 表示至少有一个候选项缺少此类抬升。

未确认物体随动（NF）。对于具有合格抬升的候选对象，我们从达到运行最小值 $m_{t^\star}$ 的最早样本开始，一直检测到确认样本 c，以判断是否存在物体跟随。若同一目标在此窗口的每个样本中都占据夹爪感知区域，其最终高度至少比窗口起始时高 10 毫米，且在手掌坐标系中测得的最大欧氏位置偏差不超过 15 毫米，则确认存在物体跟随。若候选对象包含合格抬升，但没有目标满足上述全部三个条件，则将其标记为 NF。

回合级聚合。对于每个指标，如果任一机械臂至少有一个带有该标签的候选区间，则将一个回合计数一次。比率以相应条件下的所有已评估回合为分母，包括成功回合以及没有候选区间的回合。一个回合可能因其候选区间获得不同标签而计入多个指标。每个 ECL-G 区间也都满足 NF，而 NL 和 NF 不可能出现在同一个区间中。

D.3. 任务级行为结果

表 19 按任务报告了四种策略的原生任务成功率、ECL-G、NF 和辅助NL统计量。对于 VPP2、Awomo-0.5 和 WALL-OSS，每个任务–条件单元包含 15 个回合；对于 Astra，则包含 5 个回合。原生任务成功率根据原始策略执行得出；行为测量采用上述来源。

VPP2 的三项任务中，ECL-G 均有所增加，但不同策略或任务呈现的模式并不一致。对于 Awomo- 0.5，尽管任务成功率降低，堆叠方块的检测次数仍从 2/15 降至 0/15。对于 WALL-OSS，插管的检测次数保持在 4/15；瓶子处理在两种条件下均未检测到 ECL-G。

NL 也会与成功的操作并存。例如，Awomo-0.5 和 WALL-OSS 在原始桌面条件下的 15 个瓶子处理回合均成功，且每个回合都至少包含一个 NL 候选项。

E. Push-T诊断

E.1. 场景构建与分离判据

我们构造了三种近镜像构型，使一个实体T形物体及其反射影像在头部相机视图中形成十字形、

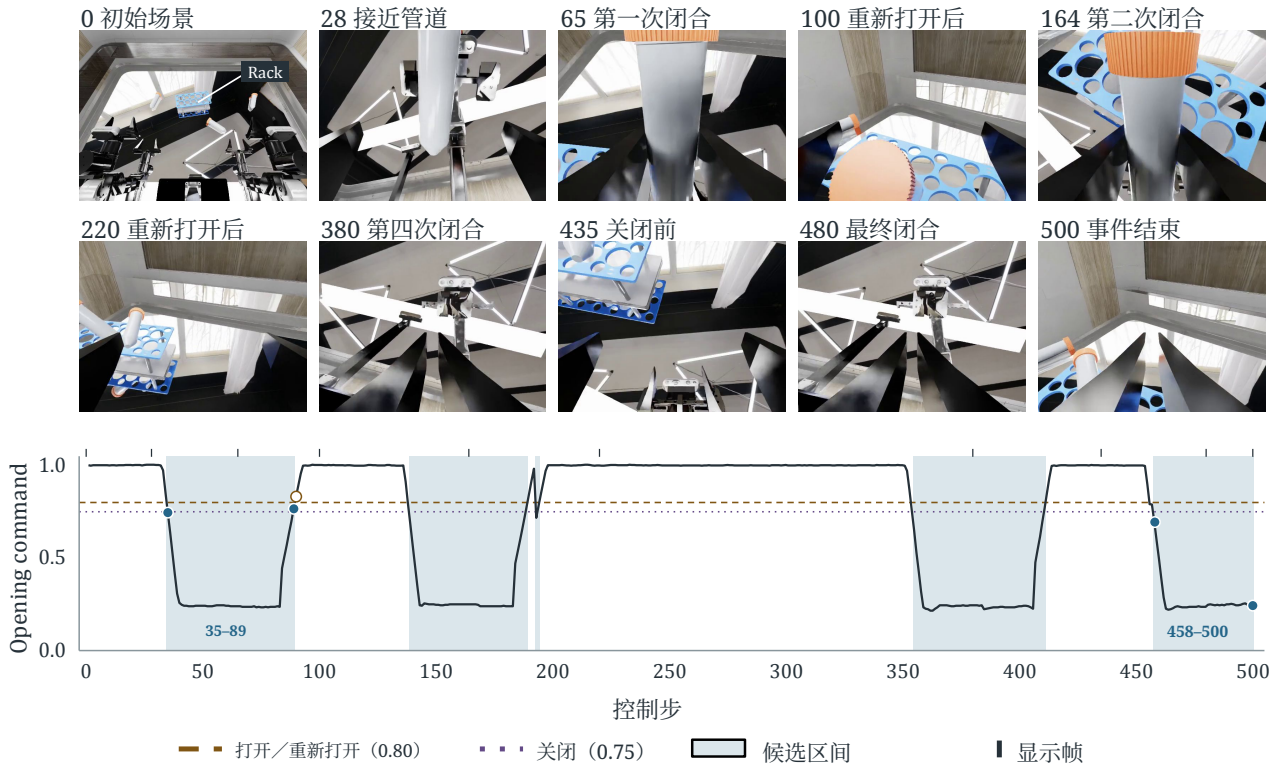


图11. 镜面桌面上一个 VPP2 插管回合中的候选闭合区间。初始头部相机视图显示架子，随后按时间顺序展示右腕视图。打开指令曲线上方的短刻度标示显示帧，阴影部分标示右夹爪的全部五个候选区间。第一个区间覆盖第 35–89 步，不包括第 90 步的重新张开采样。最后一个区间覆盖第 458–500 步，并包括最后一个采样，因为直到终止时，指令始终低于重新打开阈值。该回合结束时，任务尚未完成。

II形或 I形复合轮廓。不同轮廓构型对应的物体姿态各不相同。因此，轮廓之间的差异可能同时反映视觉外观和物理可达性的差异。

我们将推动进展与物体和反射影像之间的持续分离区分开来。令$\mathbf{p}^{xy}(t)$表示控制步 t时物体的平面位置。位移为

$$d(t) = \|\mathbf{p}_{xy}(t) - \mathbf{p}_{xy}(0)\|_2 . \tag{13}$$

若 $\max_t d(t) \geq$10毫米，则回合达到运动阈值。最大位移在完整执行过程中计算，而达到阈值所需的步数仅在达到阈值的回合中汇总。

分离需要满足更严格的条件组合：1. 平面位移至少为 10 毫米；2. 在固定头部相机视图中，实体轮廓与反射轮廓之间至少相隔五个像素；3. 两个轮廓均保持在视野内，红色像素可见度至少为 50%，且各自至少有 12 个可见红色像素；

and

4. 持续受到桌面支撑，且物体高度与初始静止高度相差不超过 3 毫米，倾斜角不大于 15°。

这些条件必须在至少 0.5 秒的仿真时间内持续成立。轮廓间隙根据几何投影计算；视觉贴片不会从评分轮廓中扣除。

回合在持续有效分离时终止，或在达到 1,000 个控制步后终止。物体抬升会被记录，但并非独立的终止条件。达到 10 毫米或满足分离判据，均不代表在原生 Push-T 对齐目标上取得成功。

E.2. 多种初始布局

为考察对放置位置的敏感性，我们在每条等高线上采用五种匹配布局评估 VPP2 和 WALL- OSS-1.0-Lite，每种布局对应三个策略随机种子。这些布局在参考排列的基础上，分别施加 0.05、0.10、0.15 和 0.20 米的预设横向平移。每个策略

表19. 任务级桌面诊断计数（原始/镜像）。样本量按任务和条件统计。行为列统计至少发生一次事件的回合，且可能重叠。

Task	成功	ECL-G	NF	NL
VPP2($n = 15$)				
导管插入	14/2	0/15	0/15	0/1
方块堆叠	15/0	0/13	0/15	0/6
瓶子处理	15/0	0/15	3/15	12/12
Awomo-0.5($n = 15$)				
导管插入	11/5	2/4	2/6	0/1
方块堆叠	10/0	2/0	6/2	1/12
瓶子处理	15/7	0/4	0/12	15/15
WALL-OSS($n = 15$)				
导管插入	3/0	4/4	4/7	0/0
方块堆叠	9/1	7/14	8/14	4/10
瓶子处理	15/12	0/0	0/2	15/14
Astra($n = 5$)				
导管插入	0/0	0/2	4/5	3/2
方块堆叠	5/4	0/0	1/1	2/3
瓶子处理	0/1	2/2	4/5	3/2

表20。五种横向放置布局下的持续分离成功率。每个单元格包含三个策略随机种子。所有碰撞均保持启用，且未应用补丁。最后一行汇总了五种布局。

	VPP2			WALL-OSS-1.0-Lite		
平移 (m)	十字形	II-like	I形	十字形	II-like	I形
0.00	0/3	0/3	3/3	0/3	3/3	3/3
0.05	0/3	0/3	3/3	0/3	2/3	3/3
0.10	0/3	0/3	1/3	0/3	1/3	3/3
0.15	0/3	0/3	1/3	0/3	0/3	3/3
0.20	0/3	0/3	0/3	0/3	0/3	0/3
所有布局	0/15	0/15	8/15	0/15	6/15	12/15

贡献 45 个回合。

表 20补充了论文正文中基于位移的结果。对于十字形配置，8/15 个 VPP2 回合和 13/15 个 WALL 回合达到 10 毫米，但两种策略均未实现持续有效分离。对于 I 形配置，相应达到 10 毫米的回合数分别为 14/15 和 15/15，而实现分离的回合数为 8/15 和 12/15。因此，可测量的物体运动并不一定满足完整的分离判据。

我们还在相同的五个横向偏移处评估 GPT-6-Astra，每个偏移和轮廓执行一次。在十字形、 II形和 I形配置中，Astra 在每种配置的 5/5 次执行中均实现了持续分离（总计 15/15）。该结果采用了上文定义的持续分离判据。

图 12展示了每个 Push-T 轮廓对应的一个 Astra 回合。在这三个示例中，机器人在镜面接触后重新调整其末端执行器的位置，随后使实体T形物体发生位移。

E.3. 视觉贴片和碰撞控制

我们进一步使用五个匹配的策略随机种子，在每个轮廓的一个固定初始状态下评估 VPP2。视觉干预会在物体–反射交界处添加一个 70×18毫米的哑光贴片。该贴片没有碰撞器。

我们在两种碰撞设置下比较未修改镜面和打了补丁的镜面：启用机器人–镜面碰撞，以及选择性过滤机器人–镜面碰撞。在这两种设置下，物体–镜面碰撞均保持启用。对于每种轮廓，比较的条件之间均保持检查点、任务指令、相机配置和初始状态不变。实验共进行了 $3\times5\times2\times2$ =60次执行。

表21 表明，启用机器人–镜面碰撞时，补丁并未提高分离成功率。十字形和 II形配置在两种视觉条件下均保持在0/5，而I形配置则从未使用补丁时的5/5变为使用补丁后的3/5。

十字形和 II形配置的结果从0/5变为5/5，这表明实体镜面限制了VPP2执行有效推动动作的能力。一种可能的解释是，VPP2持续生成会被镜面阻挡的动作。过滤机器人–镜面碰撞后，这些动作得以执行，使物理目标更容易触及。因此，过滤后的结果反映了移除这一物理约束后的性能。使用补丁后控制步数减少，可能与轮廓连续性、局部反射内容或视觉显著性的变化有关。

F. 目标遮挡诊断

F.1. 指令条件

我们比较三种指令条件。原始使用未经修改的任务指令。简短和详细分别在指令前添加图中所示的镜面相关前缀。13两个前缀均指向主摄像头视图。一些实体目标对于腕部相机仍然保持直接可见。

对于每个策略，我们在不同指令条件下使用相同的五种布局。WALL-OSS 和 Awomo-0.5 在每种布局下使用三个匹配的策略随机种子（每种条件 15 个回合）。Astra 在每种布局下执行一次（每种条件 5 个回合）。对于每个策略，检查点、观测接口、场景几何和执行预算均保持不变。

目标接触测量。 当机器人本体与物理杯子或方块之间报告的法向力超过 10^{-6}N 时，我们记录一次接触。对于每个目标类别，

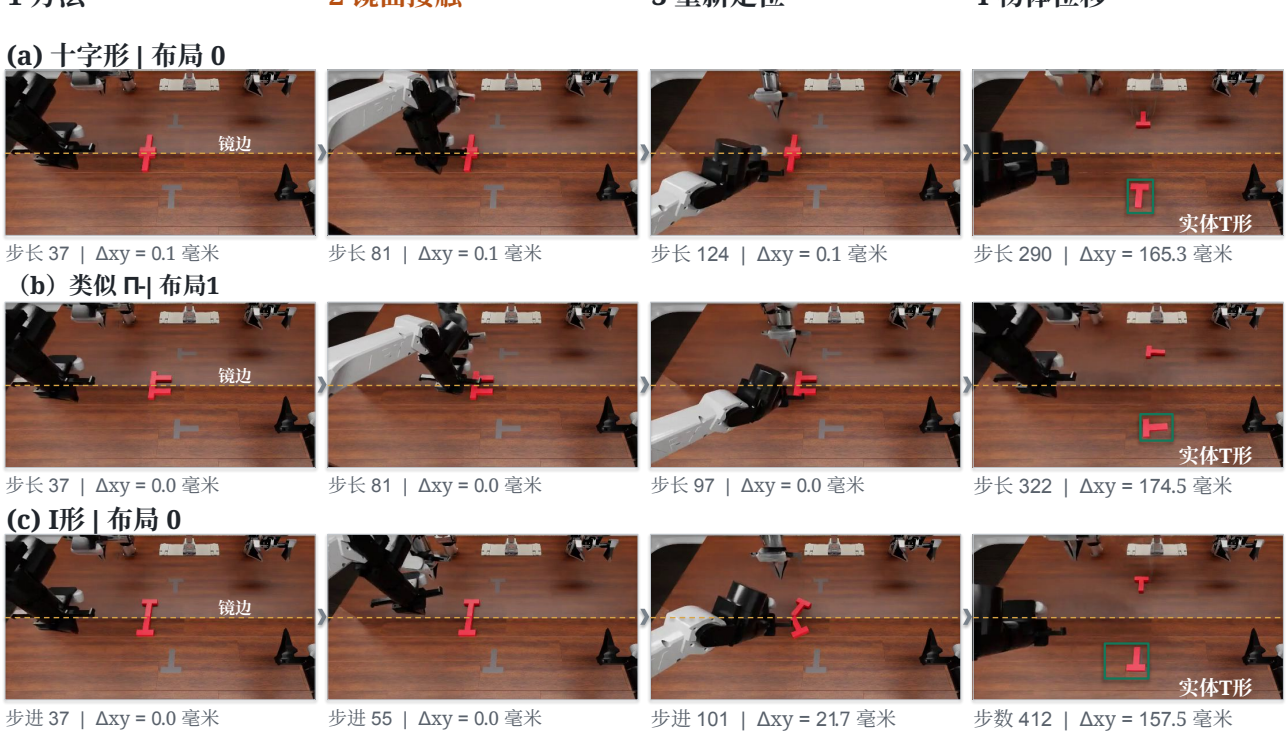


图12。三种 Push-T配置中，Astra 在镜面接触后的轨迹。每一行展示一个选定回合，从接近到接触、重新定位以及物体位移的过程。标注给出了控制步和物理物体相对于其初始位置的平面位移。

表21。VPP2 视觉补丁和碰撞控制结果。每个单元格包含五个匹配的策略随机种子，在每个轮廓的一个固定初始状态下运行。Filtered 仅禁用机器人–镜面碰撞。控制步数为所有五个回合的均值 ± 样本标准差，其中包括因耗尽 1,000 步数预算而未成功的回合。

		分离成功		控制步数	
轮廓	机器人–镜面碰撞	无补丁	补丁	无补丁	补丁
十字形	启用	0/5	0/5	1000.0 ± 0.0	1000.0 ± 0.0
	已筛选	5/5	5/5	260.8 ± 57.5	219.4 ± 118.3
II形	启用	0/5	0/5	1000.0 ± 0.0	1000.0 ± 0.0
	已过滤	5/5	5/5	333.8 ± 133.2	258.2 ± 99.9
I形	启用	5/5	3/5	428.6 ± 51.6	563.4 ± 403.5
	已过滤	5/5	5/5	244.0 ± 94.0	188.6 ± 100.1

如果一个回合中至少包含一次符合条件的接触，则计数一次。接触率是该策略–指令条件下所有已评估回合中符合此标准的回合所占的比例。一个回合可以同时计入杯子和方块的接触率。

F.2. 代表性执行

图 14 补充展示了表 5中的三种策略结果，内容是 Astra 在相同初始布局下的两次执行。在原始设置下，Astra 覆盖全部三个方块，然后按要求的顺序揭开它们：红色、绿色和蓝色。在简短设置下，Astra 也移动了实体杯子，其中一个发生了倾斜，但未能完成要求的顺序。

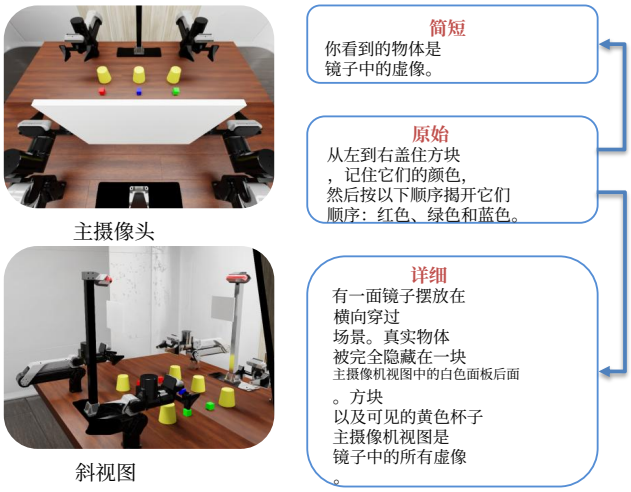


图13。覆盖方块的目标遮挡设置和指令变体。升高面板遮挡了主摄像头视野中的物理目标，同时保留其镜面反射。斜视图展示了实体布置；它不是额外的策略摄像机。简短和详细版本会在原始任务指令前添加相应前缀。

G. 真实世界实验细节

G.1. 机器人、摄像头和策略配置

我们使用配备平行夹爪的 Franka FR3 机械臂。视觉观察由两台 Intel RealSense D435i 相机提供：一台固定在外部、俯瞰工作空间，另一台安装在腕部。两台相机均以 640×480 分辨率和 30赫兹采集 RGB 图像。相机配置中禁用了深度、红外和惯性数据流。

策略输入包括两个 RGB 视图和一个八维本体感知状态：七个以弧度为单位测得的关节位置，以及以米为单位的夹爪总宽度。部署时的动作接口包含七个绝对关节位置目标和一个二值夹爪指令，其中 0 表示张开，1 表示抓取。指令通过 ROS 2 关节阻抗控制栈执行。

我们对 $\pi_{0.5}$ 和 FastWAM 均使用 Franka 专用检查点。$\pi_{0.5}$ 部署使用经过低秩适配的多任务检查点，训练步数为 31,927，预测时域为 32 步，动作网格频率为 15 赫兹。FastWAM 使用训练步数为 30,000 的多任务检查点，预测时域为 64 步，动作网格频率为 30赫兹。每个检查点都配有其保存的归一化统计量。

对于 $\pi_{0.5}$，相机图像以 224×224 分辨率进行处理。对于 FastWAM，客户端会将单独的原生分辨率图像转发至服务器，由服务器应用与检查点训练兼容的图像预处理。

表22。文档记录的 Franka 策略部署设置。预测时域表示预测的控制步数；动作频率表示指令网格频率。

策略	检查点 step	预测时域	动作 rate
$\pi_{0.5}$ 低秩适配	31,927	32	15赫兹
FastWAM	30,000	64	30赫兹

表23。真实世界评估的任务指令。

Task	指令
农产品放置	将农产品放入篮子中。
方块堆叠	将红色方块放在蓝色方块上面。
支架翻转	将白色塑料支架上下翻转。
圆柱插入	将圆柱体物体插入圆形物体托架。

G.2. 任务与镜子放置位置

我们研究了四项桌面操作任务：农产品放置、方块堆叠、支架翻转和圆柱体插入。表 23 列出了部署文档中记录的任务指令。

每项任务均在三种场景条件下进行评估：无镜子、竖直镜和镜面桌面。竖直镜会在物理工作空间旁呈现反射影像。镜面桌面会在物体及其支撑表面附近呈现反射影像。图中展示了具有代表性的构型和任务初始状态。6。

对于支架翻转，原始指令保持原样。根据记录的演示，成功的定义是达到稳定直立构型。

G.3. 执行与记录协议

针对每项任务和策略，我们在三种场景条件下评估相同的五种初始布局，每种布局重复两次。因此，每项任务、每种策略及每种条件下共进行 10 次试验。表 6 报告了相应的任务级成功率。

对于每种策略，我们在不同场景条件下使用相同的检查点、任务指令和相机配置。每种策略均使用各自的部署接口和动作频率。

成功标准。 我们根据以下标准评估任务成功与否：

- **农产品放置：** 农产品已完全放入篮子内，并由夹爪释放。
- **方块堆叠：** 红色方块放置在蓝色方块上，夹爪松开后至少 3 秒内保持稳定。



图14. 在主摄像头目标遮挡（布局1）条件下，Astra 执行覆盖方块任务的两次过程。共同的初始视图显示，主摄像头中有遮挡物，而右腕摄像头中可直接看到物理目标。原始执行在第463个控制步取得成功，简短执行则在第800步结束时仍未成功。图像下方的数字表示记录的控制步数。L和R分别表示左、右腕部相机。

- **支架翻转：** 支架从初始平放姿态翻转至稳定的直立状态。
- **圆柱插入：** 圆柱体已稳稳放入接收容器，并由夹爪松开。

试验评分。对于取消的录制，我们根据现有的执行和终止记录确定任务结果。与策略相关的命令拒绝、控制器警报和安全停止均计为失败。此类试验仍计入每个任务、策略和条件的 10 次试验分母。