

驯服虚拟现实远程操作和从演示中学习以实现多任务双手桌式服务操控

MilkDragon Team

Weize Li^{1*} Zhengxiao Han² Lixin Xu^{1,3} Xiangyu Chen^{1,4}

Harrison Bounds² Chenrui Zhang^{1,5} Yifan Xu⁶

¹AIR, Tsinghua University ²Northwestern University ³Georgia Institute of Technology

⁴HKUST (GZ) ⁵NUS (Suzhou) Research Institute ⁶University of Michigan

liweize0224@gmail.com

Abstract

本技术报告介绍了 ICRA 2025 What Bimanuals Can Do (WBCD) 竞赛中 Table Service Track 的冠军解决方案。我们在对速度、精度和可靠性有严格要求的情况下，完成了一系列具有挑战性的任务：展开桌布（可变形物体操作）、将比萨放入容器内（拾取并放置）以及打开和关闭带盖的食物容器。我们的解决方案结合了基于虚拟现实的遥操作和从演示中学习（LfD）以平衡鲁棒性和自主性。大多数子任务是通过高保真远程遥操作执行的，而比萨饼的放置则由一个基于 ACT 的策略处理，该策略是从 100 次随机初始配置的人工遥操作演示中训练出来的。通过仔细整合评分规则、任务特性以及当前技术能力，我们的方法实现了高效率 and 可靠性，最终在竞赛中获得第一名。

1. 介绍

在过去十年中，双臂操作和从演示中学习（LfD）作为有前景的框架，使机器人能够从人类操作员那里获取复杂技能。双手协调是许多现实世界任务的基础——从工业组装到家庭服务——因此受到了学术界和工业界的越来越多的关注。利用机器人硬件、可穿戴界面和策略学习算法的进步，平台如 ALOHA [8], DexCap [7], Mobile ALOHA [4] 和 Open-TeleVision [2] 展示了在越来越灵巧的场景中人到机器人技能转移的令人信服的

演示。在算法方面，取得了两个层次的进步：通用视觉语言动作（VLA）策略如 π -0 [1] 和 RDT-1B [5] 旨在实现广泛的任务覆盖，以及专注于特定操作任务的稳定高效学习的专业策略如 Diffusion Policy [3]、BC [6] 和 ACT [8]。

尽管取得了这些进步，大多数研究仍然依赖于自行设计的任务进行评估，往往忽视了诸如操作速度、可靠性及自主性等关键的行业指标。ICRA 2025 WBCD 挑战赛——餐桌服务赛道通过引入一系列结合可变形操控、精密操作和协调动作的任务直接填补了这一空白，这些任务所面临的挑战无法仅凭单一范式解决。

在此技术报告中，我们介绍了 MilkDragon 团队的冠军解决方案，该方案集成了基于 VR 的遥操作与模仿学习。我们的设计理念是在变量或力敏感子任务中利用遥操作来提高鲁棒性，同时使用基于 ACT 的策略来自动化可靠的拾取和放置。这种混合方法使我们在严格的比赛约束下能够平衡效率和可靠性，最终在挑战赛中获得第一名。我们的结果进一步突显了混合管道在连接研究原型与实际服务机器人应用方面的潜力。

2. 挑战设置：表格服务赛道

2.1. 硬件

环境设置。 我们被分配了一个尺寸为 10 英尺 × 15 英尺的工作空间，配备了一张桌子和一把椅子。这张桌子的尺寸约为 8 英尺 × 30 英寸 × 29 英寸（长 × 宽 × 高）。每个工作站都配备了电源插座（120V, 1500W）。

*Corresponding author.



图 1. 我们团队在 ICRA 2025 WBCD 挑战中的双臂操作任务流程概述——餐桌服务组别。

机器人平台。 在比赛中，我们使用了硬件赞助商 ARX Robotics 提供的比赛版 X7s 机器人。X7s 是一款人形数据采集平台，具有一个拥有 2 个自由度的头部和两个各具 7 个自由度的机械臂以及 1 个末端执行器夹爪自由度。值得注意的是，由于运输限制，比赛版的 X7s 不包括移动底座或提升机构。机器人的上半身使用 3 个 G 型夹固定在桌子上。

2.2. 任务描述和评分标准

在 Table Services Track 中，每个团队有一个 30 分钟的时间窗口，在此期间我们需要使用组织者提供的道具（如图 2 所示），并尽可能多地按照规定的顺序完成任务轮次。每个子任务最多可得 5 分，每一轮的最大分数为 15 分。任务描述如下：

- **子任务 1：桌布展开**
 - 完全展开一张已折叠的桌布（通常是折叠了 4-5 次）成一个平坦的矩形形状：5 分
- **子任务 2：打开食物容器**
 - 解锁容器的两侧：3 个点
 - 取下盖子：2 点
- **子任务 3：打包披萨饼**
 - 将披萨从盘子中放入容器：1 分
 - 对齐盖子与容器的位置：2 分
 - 锁定容器的两侧：2 点

对于评分标准，每个子任务的分数由三个因素决定：

- α ：反映自主性的**操作系数**。其值分别为 {0.5, 1, 4}，对应于现场遥操作、远程遥操作和自主策略。
- β ：**完成时间**（以秒为单位）。更快的执行对应较小的 β ，从而产生更高的效率。
- s ：子任务的**基础分数**，由成功完成定义（每个子任务最多 5 分）。

子任务的最终得分计算如下：

$$\text{Score}_{\text{subtask}} = \frac{s}{\beta} \times \alpha.$$

然后，轮次得分是：

$$S_{\text{round}} = \sum_{k=1}^3 \frac{s_k}{\beta_k} \alpha_k,$$

总竞赛得分则是通过将所有轮次的得分相加得到：

$$S_{\text{total}} = \sum_{\text{rounds}} S_{\text{round}}.$$

这种表述确保了评分同时捕捉到三个关键方面：任务准确率 (s)，自主级别 (α) 和执行效率 (分母中的 β)。因此，团队被激励尽快、正确地完成任务，并且以更高的自主性进行。

3. 我们的解决方案

从挑战中获得的见解。 我们的解决方案是在竞赛规则、任务特性和当前技术能力之间权衡的结果。特别

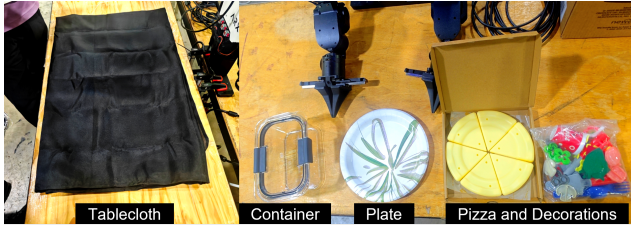


图 2. 组织者为餐桌服务赛道提供的道具和物品。

是，子任务在性质上差异很大——从长时段可变形操作（桌布展开）到力敏感的精准操作（盖子对齐）、协调锁定和标准拾取放置（披萨摆放）。这种多样性使得单一方法难以有效处理所有情况。因此，我们采用了混合策略：仅将模仿学习应用于鲁棒性的拾取放置任务，而依赖远程遥控操作来处理更多变或精细的子任务。

3.1. 亲自遥操作

现场遥控直接连接了操作员的控制设备与机器人，并允许操作员在现场观察环境。为了使放置比萨饼能够由基于模仿学习的策略执行，我们使用了现场遥控收集具有极低延迟的高质量演示。

具体来说，我们使用了 Meta Quest 3S 头戴设备来建立一个基于 VR 的远程操作界面。手持控制器被重新定向到 X7s 双臂的末端执行器（EEs），提供了人类手部动作与机器人手臂动作之间的自然映射。通过控制器按钮控制抓取状态，确保了精确且直观的抓握。这种设置使我们能够高效地收集可靠的演示样本，这些样本后来用于训练模仿学习策略。

3.2. 远程遥操作

与面对面遥操作不同，远程遥操作不允许操作员直接观察操作场景。相反，操作员依赖于机器人机载传感器，如头戴式或手眼摄像头（如图 3 所示），来感知环境并指导任务执行。这种模式更具挑战性但也更接近实际服务应用。

在我们的设置中，Meta Quest 3S 头戴设备和控制器用于姿态追踪和输入。控制器的姿态被重新定向到 X7s 双臂末端执行器，而按钮操作夹爪。视觉反馈没有通过 VR 设备传输，而是通过桌面监视器显示摄像头画面，同时操作员通过 VR 控制器操控机器人。主从控制系统基于 ROS 1 诺提克构建，所有主题都通过由 WiFi 7 路由器支持的本地网络进行传输。RGB 流被以

JPEG 压缩格式发送并在接收端解压缩，从而减少带宽需求和延迟。虽然比面对面远程操作慢，但这种配置稳定、可扩展，并分配了更高的操作系数，因此在比赛中大多数子任务中被采用。

3.3. 自主策略

采用 ACT [8] 用于双手拣选和放置。我们采用了动作分块变换器（ACT）框架作为策略主干，因为其在稳定操作方面表现出色。与标准的四个摄像头输入设置不同，我们使用了三相机配置：两个戴在手腕上的（手眼）摄像头和一个头戴式摄像头。这提供了互补的第一人称和全局视角，同时减少了数据带宽和冗余。

该政策直接预测了机器人关节命令，而不是任务空间姿态。每个 X7s 手臂拥有 7 个关节加上一个 1 自由度夹爪，使得双臂系统的总数为 16 个自由度。通过输出连贯的关节级轨迹，ACT 实现了两臂之间的平滑协调和可靠的夹爪控制。

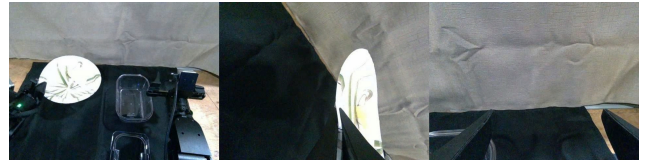


图 3. 远程遥操作过程中内置传感器观测的可视化。从左到右：头戴相机、左手腕相机和右手腕相机。

数据后处理：静态帧剪辑。 在数据收集过程中，我们观察到演示轨迹通常在开始时包含很长的静态前缀，因为操作员正在为运动做准备。这些冗余帧引入了偏差并降低了策略学习的效果，因为模型被迫模仿静止状态而不是有意义的动作。

为了解决这个问题，我们开发了一个后处理脚本，它能够根据两个末端执行器（EEF）的运动自动检测并修剪静态前缀。该脚本监控左和右 EEF 位置的帧间位移，并在位移连续几帧低于阈值时删除初始段。修剪后的数据集随后被重新保存为干净的 HDF5 文件，确保每个轨迹直接从有意义的动作开始。此外，工具生成了修剪前后 EEF 轨迹的图表，允许快速验证。这个预处理步骤显著提高了数据质量，并有助于稳定策略训练。

训练方案。 我们在 100 的披萨取放任务的真人遥控演示中训练了 ACT 策略，其中盘子、容器和披萨的初始

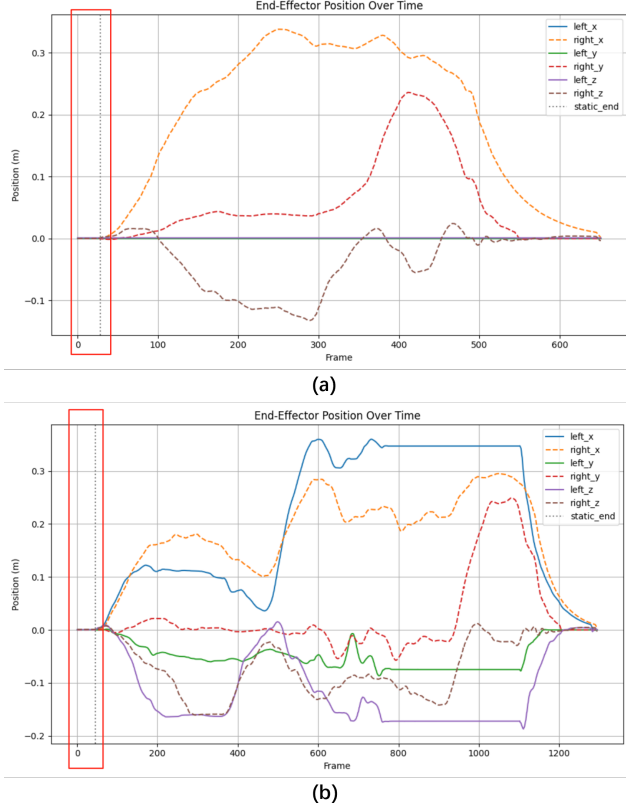


图 4. 通过静态帧剪辑进行单手 (a) 或双手 (b) 演示的数据预处理示例。红色方框突出显示了具有长时间空闲前缀的末端执行器 (EE) 位置，这些位置被自动修剪。

姿态是随机化的。训练使用了行为克隆，采用 Adam ($lr=1 \times 10^{-5}$)，块大小为 30，隐藏维度为 512，前馈维度为 3200，批大小为 4，以及 8000 个 epoch，并使用 KL 权重为 10 进行正则化。

3.4. 性能

在最后的 30 分钟环节中，我们的团队成功完成了九轮任务，在所有子任务上都取得了满分，除了最后一轮。如表 1 所示，每项任务的完成时间在各轮次之间存在显著差异：例如，任务 1（餐布展开）的时间从不到一分钟到超过两分钟不等，这取决于折叠状态的复杂程度；任务 2（容器开启）的时间波动在 10 秒至 55 秒之间，受锁是否顺利脱离的影响；而任务 3（披萨打包）由于对齐和锁定的原因显示出最大的变异性，有时会延长至 90 秒以上。尽管存在这些不稳定因素，整体轮次时间平均约为 3 分 20 秒，这使我们在严格的时限内实现了高度可靠性和竞争力的效率。

表 1. 官方记录牛奶龙团队的表现。

轮次		任务 1#	任务 2#	任务 3#
Round 1#	α	1	1	1
	β	1'47"	0'13"	1'56"
	Score	5	5	5
Round 2#	α	1	1	1
	β	1'24"	0'40"	1'15"
	Score	5	5	5
Round 3#	α	1	1	1
	β	1'03"	0'56"	1'32"
	Score	5	5	5
Round 4#	α	1	1	1
	β	2'18"	0'27"	0'30"
	Score	5	5	5
Round 5#	α	1	1	1
	β	2'10"	0'55"	0'52"
	Score	5	5	5
Round 6#	α	1	1	1
	β	0'58"	0'22"	0'59"
	Score	5	5	5
Round 7#	α	1	1	1
	β	1'43"	0'16"	1'24"
	Score	5	5	5
Round 8#	α	1	1	1
	β	1'24"	0'28"	1'42"
	Score	5	5	5
Round 9#	α	1	1	1
	β	1'24"	0'10"	0
	Score	5	5	1

4. 讨论

我们的解决方案证明，将基于 VR 的遥操作与模仿学习相结合是处理复杂双手操作任务的有效策略，在竞赛约束下平衡了鲁棒性和效率。然而，仍然存在挑战：政策需要许多演示才能达到高成功率，可扩展数据收集的硬件支持有限，并且该领域缺乏标准化的模拟

基准以进行公平评估。解决这些差距将是推动双手操作向实际部署的关键。

致谢

我们想感谢 ICRA 2025 WBCD 挑战赛的组织者：徐卓、Gu Dennis 和 Yu Peter，他们为我们团队协调硬件材料并澄清比赛细节提供了现场支持。我们也感谢决赛评委黄迪和董远哲的努力。我们还要感谢来自 ARX Robotics 的张 Josh、Ma Nina 和邹勇振，在比赛中提供硬件支持并赞助冠军奖。

参考文献

- [1] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. π_0 : A vision-language-action flow model for general robot control. arXiv preprint arXiv:2410.24164, 2024. [1](#)
- [2] Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. arXiv preprint arXiv:2407.01512, 2024. [1](#)
- [3] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. The International Journal of Robotics Research, page 02783649241273668, 2023. [1](#)
- [4] Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. arXiv preprint arXiv:2401.02117, 2024. [1](#)
- [5] Songming Liu, Lingxuan Wu, Bangguo Li, Hengkai Tan, Huayu Chen, Zhengyi Wang, Ke Xu, Hang Su, and Jun Zhu. Rdt-1b: a diffusion foundation model for bimanual manipulation. arXiv preprint arXiv:2410.07864, 2024. [1](#)
- [6] Faraz Torabi, Garrett Warnell, and Peter Stone. Behavioral cloning from observation. arXiv preprint arXiv:1805.01954, 2018. [1](#)
- [7] Chen Wang, Haochen Shi, Weizhuo Wang, Ruohan Zhang, Li Fei-Fei, and C Karen Liu. Dexcap: Scalable and portable mocap data collection system for dexterous manipulation. arXiv preprint arXiv:2403.07788, 2024. [1](#)
- [8] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual ma-

nipulation with low-cost hardware. arXiv preprint
arXiv:2304.13705, 2023. [1](#), [3](#)