arXiv:2611.33013v1 [cs.CV] 24 Nov 2026

RFW-W：为生成式具身仿真扩展物理资产宇宙

Jiawei Weng[1,*], **Rui Zhang**[1,2,*], **Saining Zhang**[2,3,*,†], **Changjian Pan**[2,4,*], **Yuanbo Lyu**[1,*], **Jiarui Guo**[1,2,*], **Yongzhi Li**[3], **刘国林**[2], **Jialu Liu**[2], **王建坤**[2], **Yi Yang**[1], **Peishuo Li**[3], **Zhiheng Li**[2], **Jiaying Wan**[2], **Junhao Chen**[2], **WeiqingXiao**[5], **JingrunDu**[1,2], **Wenyi Li**[1,2], **Weiyi Lin**[1], **Zenan Meng**[1,‡], **Hu Wei**[1,‡], **Lin Qu**[1], **Zongzheng Zhang**[2], **Hao Zhao**[2], **Chaojian Li**[6,‡]

[1]阿里巴巴集团 [2]清华大学 [3]南洋理工大学 [4]北京大学[5]南京大学 [6]香港科技大学

*核心贡献者，†项目负责人，‡通讯作者。

电子邮箱：{mengzenan.mzn,kongwang}@alibaba-inc.com, chaojian@ust.hk**项目主页：** https://github.com/RoboFlywheel-com/RoboFlywheel-World

摘要

扩展具身仿真需要多样化的物体及其可执行物理表示。我们提出了 RFW-W，这是一个物理资产宇宙，将刚体、关节式和可变形资产来源与任务条件化场景组合相连接。其报告的资产清单包含 1,010,386 条资产记录：524,611 个刚体资产，涵盖 1,212 个类别（RFW-R）；402,410 个关节式资产，涵盖 531 个类别（RFW-A）；以及 83,365 个经过联合验证的可变形资产，涵盖 485 个类别（RFW-S）。这些分类体系共包含 2,228 个数据集–类别条目。RFW-R 将类别引导生成与碰撞几何体、组合物理属性和交互标注相结合；RFW-A 使用可复用的智能体编写的程序化生成器，并明确表示部件和关节结构；RFW-S 则采用结构感知仿真表示、校准和独立测试。通用接口在检索和场景组装过程中保留了各资产类型特有的物理要求。定量评估表明，与现有数据集相比，RFW-R 具有更广泛的资产覆盖范围和更强的场景任务完成能力。在五项 RoboTwin 任务中，每种资产来源的每项任务均使用 100 个演示数据进行训练；在干净条件下，基于 RFW-R 训练的策略宏平均成功率达到 71.2%，在随机化条件下达到 41.8%，分别比最强基线高出 15.8 和 10.6 个百分点。RFW-W 为生成以物理为基础、任务条件化的具身世界提供了可扩展且统一的资产基础。

1 引言

通用具身智能要求机器人能够在多样化的物体、布局和交互中执行任务，
物理仿真为真实世界数据采集提供了一种可扩展的替代方案，能够为具身学习提供可复用且多样化的环境 [23, 52, 77, 107, 108, 114]。然而，其有效范围取决于底层资产库：场景生成和操作不仅需要多样化的物体，还需要物理仿真器能够可靠执行的表示。

大规模 3D 资源库和生成模型可提供数百万种资产，并支持几何形状与外观的开放类别合成 [9, 24, 25, 56, 59, 95, 116, 117, 134]。然而，仅凭视觉网格并不能指定



图1 RFW-W物理资产宇宙。RFW-W汇集了三个互补的资产来源：可变形物体（RFW-S，左侧）、刚性物体（RFW-R，中间）和可动对象（RFW-A，右侧），并通过通用接口将它们连接起来，以支持任务条件化具身场景构建。这些来源共包含 2,228 个数据集–类别条目中报告的 1,010,386 条资产记录。每种资产类型都保留各自的构建和验证流程，以维持其特定的物理属性和仿真要求。

公制尺度、碰撞几何体、质量、摩擦、惯性或交互语义。YCB 和 Google Scanned Objects 等集合提供了用于操作和仿真的物理对象模型 [5, 26]，但要将这些表示扩展到广泛的类别和数十万种资产，仍然面临挑战。

近期的物理资产数据集和关节建模方法为三维内容补充了物理、功能及关节信息 [6–8, 45, 57, 124, 126]。ManiTwin 增添了面向操作的数字孪生，而 Seed3D 和 EmbodiedGen 则将资产与更广泛的仿真生成系统相连接 [40, 101, 107]。如表1所示，现有具身资产库的资产规模和类别覆盖范围有限，部件级物理标注不完整，对灵巧操作的支持也有限。这些局限促使我们开发大规模物理资产宇宙，将多种物体类型与可用于具身仿真的可执行物理表示统一起来。

我们推出了**RoboFlywheel-World（RFW-W）**，这是一个用于具身仿真的大规模物理资产宇宙，涵盖刚体、关节式和可变形物体。RFW-W 包含三个资产集合：**RFW-R**，含有 1,212 个类别中的 524,611 个刚性资产；**RFW-A**，含有 531 个类别中的 402,410 个关节式资产 [71]；以及**RFW-S**，含有 485 个类别中的 83,365 个经联合验证的可变形资产 [42]。这些集合合计包含**1,010,386 条资产记录**，涵盖 2,228 个数据集–类别条目。

RFW-W 通过针对不同类型的构建流程来组织这些资产。RFW-R 是本文的重点，它将类别引导生成与部件分割、碰撞几何体、部件级物理属性，以及针对六种灵巧手的经仿真验证的抓取标注结合起来。RFW-A 利用智能体编写的程序化生成器创建具有明确部件与关节结构的可动关节资产 [71]，，而 RFW-S 则采用结构感知仿真表示、校准和独立物理验证 [42]。通用接口将这三个集合连接到任务条件化资产检索和场景构建。

我们进行了广泛评估，以考察 RFW-R 的覆盖范围、物理保真度和下游效用。评估涵盖资产和部件级覆盖范围、物理标注质量、跨仿真器一致性、灵巧抓取执行、场景构建和策略学习。在场景构建方面，RFW-R 满足所有

表1 具身资产库对比。 R/A/D 分别表示刚体、可动关节物体和可变形物体； ✓ 表示支持， × 表示不支持， △ 表示部分支持。#资产和#类别分别表示资产和类别的数量，Med. inst. 表示每个类别的实例数量中位数。场景填充度量指标表示在 16 个固定场景布局中，对 265 项刚体物体需求的满足率。Seg. 表示部件分割；Part phys. 表示数值化的部件级物理属性，包括经组合和验证的质量、质心和惯性；Hollow 表示空心感知质量建模。Dexterous manip. 指灵巧操作，其中 Hands 表示支持的灵巧手模型数量，Grasp ann. 表示抓取标注。短横线（–）表示结果不可用。RFW-W 汇总了来自 RFW-R、RFW-A 和 RFW-S 的资产记录及数据集特定类别条目，其场景任务完成度和灵巧操作结果则继承自 RFW-R。

数据集	R/A/D	覆盖范围与场景使用				部件级物理			灵巧操作。	
		#资产	#类别	Med. 实例	场景 fill	Seg.	Part 物理	中空	手	抓取 ann.
EmbodiedGen v2 [107]	✓/×/△	4,168	200	11	97.4%	✓	×	×	0	×
PhysX-3D [6]	✓/×/△	32,047	450	1	–	✓	密度	×	0	×
ManiTwin [101]	✓/×/×	4,574	485	6	90.6%	×	×	×	0	✓
RFW-R	✓/×/×	**524,611**	**1,212**	**356**		✓	✓	✓	6	✓
RFW-W	✓/✓/✓		**2,228**	**438**		✓	✓	✓	6	✓

针对 16 个固定场景基础（表 1），我们提出了 265 项物体请求。此外，在示范预算相同的条件下，基于 RFW-R 训练的策略在五项 RoboTwin 任务中取得了最高的宏观成功率（表 5）。这些结果证明了 RFW-R 在支持具有物理依据的场景构建和下游机器人操作方面的有效性。

我们的贡献总结如下：

- 我们提出了 **RFW-W**，这是一个大规模物理资产宇宙，包含 1,010,386 条资产记录，涵盖刚体、可动关节物体和可变形物体，并通过统一接口组织，以支持任务条件化资产检索和场景构建。
- 我们构建了**RFW-R**，这是一个大规模刚体资产集合，包含类别引导生成、部件级语义与物理属性标注、碰撞几何体，以及经过仿真验证、覆盖六种灵巧手的抓取标注。
- 我们对 RFW-R 进行了全面评估，涵盖资产和部件级覆盖范围、物理属性标注质量、跨仿真器一致性、灵巧抓取执行、场景实现度，以及五项 RoboTwin 任务中的下游策略学习。

2 相关工作

2.1 三维资产集合与物理资产数据集

大规模 3D 资源库，包括 ShapeNet、3D-FUTURE、ABO、OmniObject3D、Objaverse、Objaverse- XL 和 Objaverse++，大幅拓展了可用三维资产的规模、类别覆盖范围和视觉多样性 [9, 20, 24, 25, 32, 59, 113]。近期的生成模型，如 CLAY、TRELLIS、TRELLIS.2、Hunyuan3D 2.1 等，进一步支持高质量几何形状、纹理和 PBR材质的开放类别合成 [11, 14, 56, 95, 116, 117, 128, 134]。这些资源主要聚焦于视觉外观和几何重建，而具身仿真所需的物理属性和交互能力则基本未得到解决。

早期面向操作的资源（如 YCB）提供了用于机器人操作的物理对象模型 [5], ，而 Google Scanned Objects 则提供了用于仿真的扫描家居资产。较新的物理资产数据集在此基础上进一步纳入了公制尺度、材料属性、物理属性、可供性和适用于仿真的表示 [6, 7, 57, 126]。ManiTwin 引入了大规模、面向操作的数字资产，并提供抓取标注和经验证的交互提案 [101], ，而

Seed3D 和 EmbodiedGen 进一步将物理资产生成与更广泛的具身仿真和世界生成系统联系起来 [30, 40, 106, 107]。

如表 1所总结，现有物理资产库在资产规模和类别覆盖范围方面差异显著。此外，这些数据集并未一致地提供详细的部件级物理属性和灵巧操作支持。RFW-W 通过在统一的具身仿真框架中整合刚性、可动和可变形资产，解决了这些局限。它包括用于大规模刚性资产的 RFW-R、用于智能体编写的可动对象程序化生成的 RFW-A [71], , 以及具备特定配置物理验证和运行时验证的日常可变形资产 RFW-S [42]。

2.2 物理化与面向交互的资产表示

语义和部件级表示为物理资产构建奠定了结构基础。形状感知零样本语义分割利用自监督像素特征中的谱线索，将语义识别与空间划分相结合 [67]。PartNet 为三维物体建立了细粒度的层次化部件标注 [76], , 而 SAPIEN 则为可动对象提供了基于部件的交互式仿真环境 [115]。GAPartNet 进一步研究了跨类别的通用可操作物体部件，将部件感知与机器人操作联系起来 [37]。这些工作彰显了识别有意义的物体部件的重要性，而非仅依赖整体几何表示。

除语义结构外，物理化还需要赋予物理属性，并构建可由物理仿真器执行的碰撞表示。CoACD 在考虑几何凹性的同时生成凸碰撞分解 [109]。PUGS 结合形状感知高斯重建、视觉语言属性推断、基于特征的传播和体积积分，实现零样本物理理解和物体质量估计 [90]。在这些方向的基础上，RFW-R 为语义部件赋予材料和摩擦属性，估计其物理属性，并将部件级质量、质心和惯性组合成显式刚体表示。此外，它还构建凸碰撞几何体，并保留视觉几何体、碰撞几何体与物理属性之间的对应关系。

可用于交互的资产还需要标注和可执行的抓取表示。ACRONYM 和 DexGraspNet 提供了面向一般物体的大规模仿真抓取数据集 [27, 105], , 而 GraspXL 则研究了多种物体的抓取运动生成 [132]。ManiTwin 和 EmbodiedGen 进一步将物理资产标注与面向操作的交互及具身仿真联系起来 [101, 107]。然而，仅有抓取标注或生成的抓取候选方案，并不能保证在仿真中成功执行。RFW-R 将语义部件表示与灵巧抓取生成相结合，并通过对六种灵巧手模型进行接触、穿透、闭合、抬升、保持和扰动检查来验证候选抓取。它的物理表示还在多个物理仿真器中接受评估，以检验其仿真行为的一致性。

2.3 从物理资产到生成式具身世界

程序化和语言驱动的场景生成，借助结构化生成、资产检索、学习得到的布局和物理约束，构建出日益多样且可扩展的具身环境 [23, 31,34, 60, 86, 87, 122, 125]。相关工作进一步探索了交互式环境、场景理解、可控场景编辑，以及根据真实世界观测重建可仿真场景 [18,33, 52, 98, 120]。近期的智能体方法则融入多模态推理、迭代规划、工具使用以及视觉或仿真反馈，以构建满足任务条件、可编辑且可在物理上执行的世界 [43, 61, 83, 107, 114, 123, 139]。与此同时，仿真平台和数据引擎不断扩展环境、任务和演示数据，为机器人学习提供支持 [16, 41, 53, 54, 74, 77, 82, 93, 94, 102, 108]。这些工作主要聚焦于生成、重建和扩展下游场景及交互数据，而 RFW-W 聚焦于上游物理资产基础，通过统一检索接口管理刚性、可动和可变形资产，并将其与任务条件化场景组合及下游机器人学习相连接。

2.4 机器人策略与训练数据

扩散策略通过条件扩散过程对视觉运动动作建模 [19], ，而 RT-2 将视觉语言预训练与机器人动作预测相结合 [4]。Open X-Embodiment 汇总了来自多个机器人平台的演示数据 [80], ，Octo 则研究了基于这些数据训练并适配到新环境中的通用策略 [79]。这些工作为利用多样化经验进行可扩展的策略学习提供了启发。我们的策略实验考察了这些经验的另一种输入：用于收集任务演示的物理资产库。实验在不同资产库之间固定策略系列和演示数量，同时将类别匹配、轨迹和资产准备方面的差异纳入训练来源的比较。

3 方法

RFW-W 采用刚性、关节式和可变形资产来源，用于任务条件检索和场景构建（图 1 和 3）。这些来源保留了不同的物理表示和验证记录；通用接口提供任务所需的字段，同时保留适用于每种资产类型的仿真模型。

3.1 物理资产来源与构建流程

刚性资产：类别覆盖范围与可执行交互。 RFW-R 提供 1,212 个叶类别中的 524,611 个保留物体。类别条件生成和多视角检查确保了覆盖范围；语义部件、凸碰撞体、组合物理属性和交互标注使这些资产可用于刚体仿真。第 3.2 和 3.3 节介绍了这一构建过程，第 4 节评估其在资产、抓取、场景和策略方面的结果。

关节式资产：可复用程序与运动感知检查。 RFW-A 提供涵盖 531 个类别的 402,410 个关节式资产 [71]。智能体将可复用组件及其依赖关系组织成类别蓝图，实现可执行生成器，并针对采样配置和关节状态修复故障。经过验证的生成器随后可在无需逐资产模型推理的情况下采样新资产。每个样本均保留其部件层级、关节参数、视觉与碰撞几何体、物理属性标注、生成器版本和种子。配套研究报告称，在其发布版本中，资产级运动学通过率达到 100%；在 200 个资产的子集上，四个经测试的仿真器均实现了一秒被动展开的 99.5% 通过率，且采用根部固定设置。这些检查针对关节结构和所述被动状态；它们不同于 RFW-R 的跌落响应协议。关节结构估计和真实物体操作研究为关节式资产提供了额外证据。

可变形资产：结构适配与独立测试。RFW-S 在 485 个类别中提供了 83,365 个经过联合验证的资产，涵盖软固体、布片、壳体、服装和袋子 [42]。结构感知适配器选择兼容的基于位置的动力学（PBD）或物质点法（MPM）表示，并将推断出的参数映射到 Genesis 1.3.3 中实际执行的控制项。校准支持有界修正；另一种物理探测用于测试冻结候选项，独立检查则用于度量指标运行时行为。配置匹配交集保留了 83,365 个资产和全部 485 个类别。在经过评估的仿真配置下，所得资产同时满足物理响应标准和运行时标准。此外，针对服装和抓取的研究还评估了变形行为、操作性能和物体完整性。

统一资产组织与检索。 RFW-W 通过统一接口整合刚性、可动和可变形资产，以支持任务驱动检索和场景构建。该接口组织语义、几何、物理和交互信息，并纳入类型特定表示，包括刚性物体的碰撞几何体和抓取标注、可动对象的部件层级结构和关节参数，以及可变形物体的变形模型和仿真参数。这种统一组织方式能够根据任务要求和物理兼容性，高效检索资产并构建场景。



图2 RFW-R 刚性资产构建流程。 L0–L2 阶段分别执行类别引导生成、多视角渲染和基于视觉语言模型的质量控制。L3–L4 阶段构建语义部件和碰撞表示，估计物理属性，并导出物理化资产；L5 总结下游仿真评估结果。蓝色反馈回路支持类别重新分配和重新检查，绿色反馈回路则引导物理属性改进。L4 后导出的资产支持 URDF、MJCF 和 USD 表示。

3.2 类别引导资产生成

图 2 展示了 RFW-R 构建流程，涵盖资产生成与质量控制（L0–L2）、物理化与导出（L3–L4），以及下游仿真评估（L5）。在 L0 阶段，我们采用人工整理的分类体系来指导大规模资产生成。Qwen3.7-plus 生成多样化、以类别为条件的物体描述，再使用 FLUX.2-klein 将其转换为图像 [2]，并辅以 Wan2.7 生成流程。RMBG-2。0 [3] 提取物体前景，TRELLIS。2 [117] 重建带纹理的三维网格。

在 L1 中，每个生成的资产都会从五个视角（前面、后面、左侧、右侧和底部）进行渲染。在 L2 中，Qwen3.7-plus [84, 121] 会从几何完整性、物体可识别性、刚体适用性和类别一致性等方面评估多视角渲染结果。智能体质量控制回路会接受合格资产、拒绝不合适的候选资产，或对分类错误的物体重新分配类别，随后进行重新检查。该生成与筛选过程最终在 1,212 个类别中保留了 524,611 个刚性资产。更多实现细节和提示词规范见附录 A.2。

3.3 物理化与交互标注

完成视觉质量控制后，L3 会构建语义部件和碰撞表示。P3-SAM [72]识别功能部件，而 CoACD [109] 则将物体几何形状分解为凸碰撞组件，用于物理仿真。在 L4 中，Qwen3.7-plus 会分析对齐的RGB图像和部件掩码，以估算部件材料、摩擦系数、公制尺寸和整体物体质量先验。系统会根据估算出的材料密度和体积，为每个部件分配质量$m_p = \rho_p V_p$。随后，将部件级属性合成为物体的总质量、质心和惯性张量。当合成质量与整体物体质量估计值之比超出 $[0.5, 2]$ 时，系统会按比例统一调整各部件质量，同时保持它们之间的相对比例。

完成 L4 物理化后，资产将导出为 URDF，并转换为 MJCF 和 USD 表示，以便在不同仿真格式中保留其几何和物理属性。L5 表示对导出资产进行下游仿真评估，具体任务协议和结果详见实验与附录。这些评估用于表征资产行为，并为物理属性改进提供反馈。详细的物理属性标注和组合流程见附录 A.2和 A.4。

为进一步支持机器人操作，我们针对六种机器人手模型生成部件条件化灵巧抓取标注。在 TRO-Grasp [29], 的基础上，我们利用部分点云将抓取生成条件化为目标语义部件，同时保留其手部条件化姿态表示。候选抓取需经过语义部件接触、网格穿透、多指接触，以及模拟闭合、抬升、保持和扰动检查等筛选。我们对所有受支持的手采用相同的标注与筛选流程，从而生成可执行的抓取标注，并将其与每项资产的语义和物理表示相关联。更多详情见附录A.6。

3.4 任务驱动的场景构建

在构建物理资产集合的基础上，RFW-W 为任务条件化资产检索和场景生成提供统一接口。该接口组织语义、几何、物理和交互信息，同时保留刚性物体、铰接物体和可变形物体所需的表示。给定任务描述后，语言模型会将所请求的物体及其空间关系转换为结构化场景规范。随后，根据语义相关性和物理兼容性，从相应的资产集合中检索资产。

如图 3所示，场景构建遵循任务核心优先策略。几何求解器首先在独立于房间的坐标系中组装任务相关物体，包括目标、支撑物体及其所需的空间关系。任务核心确定后，其物体成员关系和相对位姿即被冻结。随后，使用 Infinigen [86, 87], 构建周围房间；该过程会生成墙壁、地板、门和窗户，同时遵循已冻结任务核心的约束。之后，在保护区域约束下添加周边家具、材质和光照，并将程序化材质烘焙为 PBR 纹理。

生成的场景在导出前会经过功能和视觉验证。确定性检查会评估几何一致性、任务约束和语义正确性，而 Qwen3.7-plus 视觉语言评估器则会评估视觉逼真度和场景连贯性。检查未通过时，会通过彼此独立的功能修复和视觉修复通道（R0–R5）进行有界细化，同时不修改冻结的任务核心。最终场景将以结构化描述和独立完整的 USD 包形式导出，供 Isaac Sim 使用 [78]。图 4 展示了不同室内环境中的任务对齐示例。本研究中的场景任务完成度和下游策略评估均使用 RFW-R 资产进行。

4 实验

我们从资产构建到下游应用对 RFW-R 进行评估，考察语义和物理标注质量、跨模拟器行为、灵巧抓取执行、机器人策略学习以及任务驱动的场景构建。跨库比较将 RFW-R 用作刚性资产来源，而 RFW-A 和 RFW-S 则在其配套研究中进行评估。评估队列、指标和流程详见附录 A.1。

4.1 资产标注与物理一致性

我们评估 RFW-R 是否提供可靠的语义标注和物理一致的资产描述。在统一的视觉语言评估流程下，RFW-R 的类别一致性（0.981）和功能一致度（0.860）在所比较的资源库中均为最高；如表2所示，其语言一致度（0.864）与 PhysX-3D（0.870）接近。指标定义和物体级诊断见附录A.3。

表 2 采用共享视觉-语言评判协议评估的部件级语义标注一致性。类别、语言和功能分别表示三个标注维度；分数越高表示一致性越强。更多诊断结果见附录 A.3。

度量指标	RFW-R	EmbodiedGen v2	PhysX-3D	ManiTwin
类别 ↑	**0.981**	0.93	0.96	0.65
语言 ↑	0.864	0.67	**0.87**	0.50
功能 ↑	**0.860**	0.80	0.79	0.64

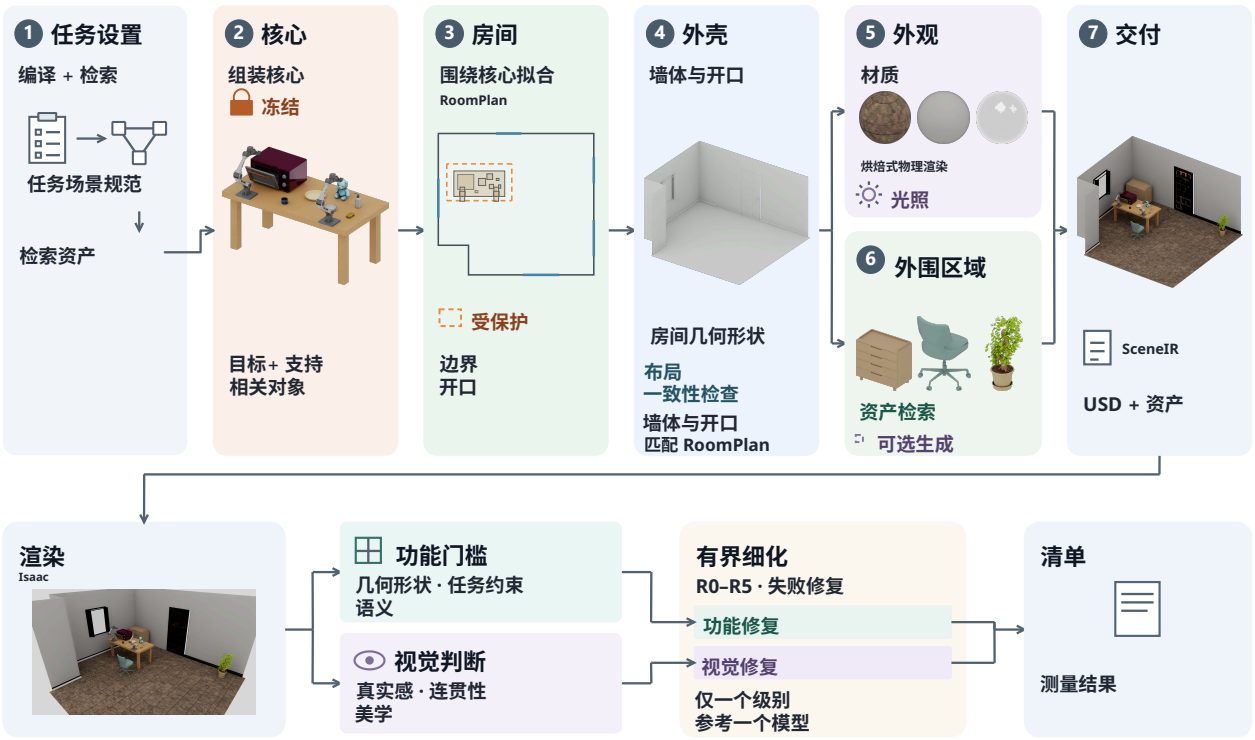

图 3 任务驱动的场景构建。任务指令被编译为结构化规范，随后进行资产检索和任务核心组装。冻结的核心为房间生成、外观合成和场景构建提供指导，而功能与视觉检查则支持在导出前进行有界细化。

物理表示。我们检查了 2,424 个资产的物理参数一致性。导出的 URDF 和 MJCF 质量与已分配的物体质量相符，最大相对误差为 1.5×10^{-6}；在部件质量总和与存储的物体质量之间，相对容差为 10^{-3} 时，98.5% 的资产符合要求。更多质心和惯性检查结果见附录A.4。

实心填充建模的影响。我们进一步比较了 31 个保留资产的原生物理表示与实心填充建模。将原生表示替换为实心填充后，MuJoCo、Genesis 和 Isaac Sim 中的平均跌落响应标准通过率分别降低了 9.25、9.25 和 7.53 个百分点。各仿真引擎中出现的一致下降凸显了原生表示所编码的质量分布的重要性。

4.2 跨仿真器的跌落响应一致性

为考察相同物理资产在不同仿真引擎中的表现，我们对 PyBullet [21], MuJoCo [96], Genesis [36], 和 Isaac Sim [78]中的 2,257 个公共资产进行了标准化跌落响应测试。所有引擎均采用相同的估算物理属性和多凸包碰撞几何体。每个资产都在三个下落高度和五种初始方向下进行测试。我们计算满足共同位移、旋转和穿透判据的试验比例，在一致性分析中以 0.5 为阈值将其二值化，并使用 Pearson r比较连续比例。评估标准详见附录 A.5。

在四种引擎中，80.77% 的资产获得相同的阈值化标签（Fleiss’ $\kappa = 0.481$），其中包括共同通过和共同未通过的情况。MuJoCo 和 Genesis 的成对一致性最高（94.20%，Cohen’s $\kappa = 0.741$），且各资产判据通过比例的相关性最强（$r = 0.919$；表 3）。这些度量指标量化了 RFW-R 的物理行为在不同仿真后端之间的一致性。



图 4 任务对齐的场景示例。 十条任务指令及其对应的场景渲染图。每个示例都展示了生成场景中所请求的动作、目标物体和支撑表面。

表 3 在 2,257 个公共资产上的跨模拟器掉落响应一致性。由标准通过比例二值化（阈值为 0.5）后计算一致性和 Cohen's κ，而 Pearson r 则比较连续比例。四个引擎在 80.77% 的资产上达成一致（Fleiss' $\kappa = 0.481$）。

引擎对	一致性（%）	↑ Cohen's κ	↑ Pearson r ↑
PyBullet–MuJoCo	87.55	0.430	0.688
PyBullet–Genesis	86.71	0.398	0.670
PyBullet–Isaac Sim	87.90	0.331	0.650
MuJoCo–Genesis	94.20	0.741	0.919
MuJoCo–Isaac Sim	90.87	0.507	0.838
Genesis–Isaac Sim	89.68	0.450	0.830

4.3 灵巧抓取执行

我们评估部件条件化抓取标注能否由 Genesis 中的六种灵巧手模型执行。对于每个物体–手部组合，最多有 64 个抓取候选接受闭合、抬升、保持和六方向扰动测试。当物体在 0.15 m 的指令运动后上升超过 0.075 m 时，即视为抬升成功。稳健成功还要求在三种质量、摩擦和姿态设置下成功保持物体并通过扰动测试。我们报告对象级任一成功覆盖率：如果某个物体至少有一个经过评估的候选通过相应测试，则该物体计为成功。

每个对象有 64 个候选项时，宏平均对象级覆盖率在抬升任务中达到 79.5%，在鲁棒抓取任务中达到 73.7%（表4）。Ability 和 Inspire 在这两项指标上的数值均超过 90%，而 Shadow 的提升覆盖率为 42.3%，稳健覆盖率为 27.3%。图 5 展示了六种手部形态下的成功示例。不同手之间的差异凸显了基于执行的抓取评估的必要性，而不能仅依赖生成的姿态；候选项预算和接触分析见附录 A.6。

4.4 策略学习与真实机器人轨迹回放

我们评估使用 RFW-R 资产采集的演示数据能否提升下游操作策略。我们针对四个资产库中的每一个，分别在五个 RoboTwin 仿真任务上训练独立的 $\pi_{0.5}$ [81] 策略；每个资产库 — 任务组合使用 100 条演示数据，并采用训练步数为 19,999 时的检查点。每个策略均在 100 个干净条件专家可解场景和 100 个随机化条件专家可解场景中进行测试，共产生 4,000 次闭环仿真运行。随机化涉及视觉上下文、桌面高度和语言指令；完整的训练和评估

表 4 六种手部模型的仿真灵巧抓取结果。 N_L 和 N_R 分别表示用于抬升和稳健成功评估的对象数量；这两项指标采用不同的对象群组。宏平均表示各只手的未加权平均值。置信区间见附录 A.6。

Hand	N_L	抬升（%）↑	N_R	稳健（%）↑
Ability	2,372	92.5	2,367	90.4
Inspire	2,334	92.3	2,353	90.4
Allegro	2,372	91.3	2,372	81.6
Barrett	2,353	84.4	2,372	81.5
XHand	2,315	74.1	2,315	71.2
Shadow	2,353	42.3	2,372	27.3
宏平均		79.5		73.7



(a) Allegro Hand (b) Barrett Hand (c) Shadow Hand (d) Ability Hand (e) Inspire Hand (f) XHand

图5 六种手部模型的灵巧抓取执行。使用 RFW-R 资产的代表性成功抓取与抬升示例，在 Genesis 中根据规定的抬升标准进行评估。

设置见附录A.7。

表5 仿真中的闭环RoboTwin策略成功率（%）。每个资产库为每项任务提供100条训练示范；每个任务—条件单元均在100个专家可解场景上进行评估。干净条件和随机化条件分别表示两种评估条件，宏平均为五项任务的未加权平均值。

Mode	训练资产	A2B-L ↑	支架 ↑	盘子 ↑	碗-2 ↑	Pillbottle-Pad ↑	宏平均 ↑
干净条件	RFW-R	**47**	**63**	**96**	**98**	**52**	**71.2**
	EmbodiedGen v2	3	0	24	0	0	5.4
	PhysX-3D	22	51	93	42	6	42.8
	ManiTwin	33	50	81	80	33	55.4
随机化条件 R		**21**	**33**	**76**	**73**	6	**41.8**
	EmbodiedGen v2	1	0	10	0	0	2.2
	PhysX-3D	8	16	48	24	**9**	21.0
	ManiTwin	8	29	56	59	4	31.2

使用 RFW-R 训练的策略在干净条件（71.2%）和随机化条件（41.8%）评估中均取得最高的宏观成功率，分别比最强基线 ManiTwin 高出 15.8 和 10.6 个百分点（表 5）。RFW-R 在全部五项无扰任务以及五项随机化任务中的四项上均领先；PhysX-3D 在随机化 Pillbottle-Pad 任务上的表现更好（9% 对比 6%）。在评估的任务设置下，匹配的示范预算体现了 RFW-R 训练源所带来的下游价值。

跨具身真实机器人轨迹回放。 为评估仿真生成动作在物理环境中的可执行性，我们在填充了 RFW-R 资产的 RoboTwin 场景中运行训练得到的策略，并记录由此产生的动作轨迹。由于仿真机器人与物理机器人的具身不同，我们将相应的末端执行器（EEF）轨迹重定向至真实机器人，以进行开环执行，而非直接迁移关节空间指令。受 Wang等人提出的仿真到硬件轨迹回放启发， [100] 以及 Kappel等人提出的方法， [50]，我们展示了五个任务族中的真实机器人执行情况。图6展示了成对的仿真与真实机器人画面，呈现抓取、抬升和放置阶段。策略推理完全在仿真中进行，而物理机器人则在没有在线策略反馈的情况下执行重定向后的轨迹。这些定性演示与闭环仿真成功率



图6 仿真生成轨迹的跨具身真实机器人回放。针对 RoboTwin 的五个任务族，策略使用 RFW-R 资产在仿真中运行，随后在物理机器人上执行其记录的轨迹。仿真和真实环境分别展示了仿真环境与硬件上的抓取、抬升和放置阶段。真实机器人阶段是轨迹回放，而非闭环策略推理。

（见表5）是分开报告的。更多实现细节见附录A.7.4。

4.5 场景构建与实现

我们将 RFW-R 资产满足任务驱动场景规格的能力，与 EmbodiedGen v2 和 ManiTwin 进行比较。所有三家资产提供方均在相同的 16 个固定房间基础场景上接受测试：一个客厅、两个厨房、七个办公室、两间卧室、三间浴室和一个餐厅。它们使用相同的请求物体、目标尺寸、支撑关系和放置规则。总计有 265 个请求实例，组成 153 个组；未解决的请求仍计入满足率的分母。我们使用共享的相机目标点、方位角和距离候选值来选择信息丰富的视点。每个提供方与房间基础场景的组合渲染五个视图，共在 48 个条件下生成 240 张图像。

表6 在16个固定房间基础场景上的任务驱动场景实现度。实例填充率和组填充分别报告265个实例和153个组中已实现的请求数；不同资产数表示所使用的唯一资产数量；放置表示成功放置的请求实例所占的比例。

库	实例填充率 ↑	组填充 ↑	不同资产数 ↑	放置 ↑
RFW-R	**265/265**	**153/153**	**80**	**100%**
EmbodiedGen v2	258/265	149/153	66	97.4%
ManiTwin	240/265	135/153	53	90.6%

RFW-R 满足了全部 265 个所需实例和全部 153 个组；相比之下，EmbodiedGen v2 满足了 258 个实例和 149 个组，ManiTwin 满足了 240 个实例和 135 个组（表 6）。此外，RFW-R 按照共享放置规则放置了每个所需实例，并使用了 80 种不同资产，这表明与所比较的资源库相比，它对固定任务要求的覆盖更全面。

在视觉评估中，使用 Qwen3.7-plus 视觉语言评估器，根据真实感、视觉功能性、布局、完整性、提示对齐和风格，对每种提供方–基础场景条件进行评分。每种条件均在温度为零时评估三次；先计算该条件内的平均分，再对 16 个房间基础场景的评分取平均值。我们通过对这些基础场景进行 10,000 次自助重采样，估算配对均值分差。

RFW-R 获得了最高的平均视觉得分（3.878），在布局、完整性和风格方面均处于领先地位（表 7）。其得分比 ManiTwin 高 0.993 分（95% 自助法置信区间 $[0.611, 1.444]$）。与 EmbodiedGen v2 相比，差异很小（0.035 分；95% 置信区间 $[-0.108, 0.191]$），表明两者总体视觉评分相当。结合场景实现度结果，这些发现凸显了 RFW-R 在提供所需物体和构建任务对齐场景方面的优势。

表7 对48种提供方–基础场景条件下场景构建的视觉评估（0–10分）。视觉功能性表示视觉功能性，完整性表示完整性，提示表示提示对齐。每种条件评分三次；先对各条件的得分取平均，再对16个基础模型取平均。

库	真实感	视觉功能性	布局	完整性	提示	风格		Mean
RFW-R	3.52	4.19	**4.88**	**3.02**	<风格 id=			
EmbodiedGen v2	**3.67**	**4.25**	4.71	2.92	**3.38**	4.15	3.844	
ManiTwin	2.38	3.27	3.38	2.29	2.58	3.42	2.885	

5 结论

我们提出了 RFW-W，这是一个大规模物理资产宇宙，涵盖超过一百万条刚性、铰接式和可变形资产记录，分布于 2,228 个数据集特定类别条目中。通过将类别引导资产生成、物理表示和交互标注与任务驱动的场景构建相结合，RFW-W 为物理基础具身仿真提供了统一基础。在 RFW-R 上进行的大量评估表明，其具备广泛的资产覆盖范围、可靠的物理表示以及可执行的灵巧交互。此外，在我们的评估中，RFW-R 实现了完整的场景实现度，并且相比现有资产库显著提升了下游机器人操作性能。我们相信，RFW-W 是迈向可扩展物理资产生成以及构建多样化交互式具身世界的重要一步。

参考文献

[1] Sherwin Bahmani、Ivan Skorokhodov、Victor Rong、Gordon Wetzstein、Leonidas Guibas、Peter Wonka、Sergey Tulyakov、Jeong Joon Park、Andrea Tagliasacchi 和 David B. Lindell。4d-fy：使用混合分数蒸馏采样进行文本到四维生成。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集》，第 7996–8006 页，2024年。[2] Black Forest Labs。FLUX。

2 [ ] https://bfl.ai/blog/flux2-klein-towards-interactive-visual-intelligence 2026

klein：迈向交互式视觉智能。，1月。[3] BRIA AI。BRIA背景移除 v2.0 模型卡。

https://huggingface.co/briaai/RMBG-2.0，2024年。[4] Anthony Brohan、Noah Brown、Justice Carbajal、Yevgen Chebotar、Xi Chen、Krzysztof Choromanski、Tianli Ding、Danny Driess、Avinava Dubey、Chelsea Finn、Pete Florence、Chuyuan Fu、Montse Gonzalez Arenas、Keerthana Gopalakrishnan、Kehang Han、Karol Hausman、Alexander Herzog、Jasmine Hsu、Brian Ichter、Alex Irpan、Nikhil Joshi、Ryan Julian、Dmitry Kalashnikov、Yuheng Kuang、Isabel Leal、Lisa Lee、Tsang-Wei Edward Lee、Sergey Levine、Yao Lu、Henryk Michalewski、Igor Mordatch、Karl Pertsch、Kanishka Rao、Krista Reymann、Michael Ryoo、Grecia Salazar、Pannag Sanketi、Pierre Sermanet、Jaspiar Singh、Anikait Singh、Radu Soricut、Huong Tran、Vincent Vanhoucke、Quan Vuong、Ayzaan Wahid、Stefan Welker、Paul Wohlhart、Jialin Wu、Fei Xia、Ted Xiao、Peng Xu、Sichun Xu、Tianhe Yu 和 Brianna Zitkovich。RT-2：视觉-语言-动作模型将网络知识迁移至机器人控制，2023年。URL https://arxiv.org/abs/2307.15818。[5] Berk Calli、Aaron Walsman、Arjun Singh、Siddhartha Srinivasa、Pieter Abbeel 和 Aaron M. Dollar。操作研究中的基准测试：YCB物体与模型集及基准测试协议，2015年。URL https://arxiv.org/abs/1502.03143。[6] Ziang Cao、Zhaoxi Chen、Liang Pan 和 Ziwei Liu。PhysX-3D：物理基础三维资产生成，2025年。URL https://arxiv.org/abs/2507.12465。[7] Ziang Cao、Fangzhou Hong、Zhaoxi Chen、Liang Pan 和 Ziwei Liu。PhysX-Anything：从单张图像生成可用于仿真的物理三维资产，2025年。URL https://arxiv.org/abs/2511.13648。[8] Ziang Cao、Yinghao Liu、Haitian Li、Runmao Yao、Fangzhou Hong、Zhaoxi Chen、Liang Pan 和 Ziwei Liu。PhysX-Omni：适用于刚体、可变形物体和可动对象的统一可用于仿真的物理三维生成，2026年。URL https://arxiv.org/abs/2605.21572。[9] Angel X. Chang、Thomas Funkhouser、Leonidas Guibas、Pat Hanrahan、Qixing Huang、Zimo Li、Silvio Savarese、Manolis Savva、Shuran Song、Hao Su、Jianxiong Xiao、Li Yi 和 Fisher Yu。ShapeNet：信息丰富的三维模型库，2015年。URL https://arxiv.org/abs/1512.03012。[10] Honghua Chen、Yushi Lan、Yongwei Chen 和 Xingang Pan。Artilatent：通过结构化潜变量生成逼真的关节式三维物体。载于《SIGGRAPH Asia 2025会议论文集》，2025年。doi：10.1145/3757377.3763922。[11] Junhao Chen、Xiang Li、Xiaojun Ye、Chao Li、Zhaoxin Fan 和 Hao Zhao。Idea23D：协作式大型多模态模型智能体通过交错多模态输入实现三维模型生成。载于第31届国际计算语言学会议论文集，第 4149–4166 页。计算语言学协会，2025年。URL https://aclanthology.org/2025.coling-main.280/。[12] Junhao Chen、Mingjin Chen、Henghaofan Zhang、Minglin Chen、Liaoyuan Fan、Boran Zhang、Saining Zhang、Mingze Sun、Hao Zhao、Ruqi Huang、Zhihao Li 和 Yufei Wang。作为原生四维渲染器的视频模型：基于动画网格的世界条件。arXiv预印本 arXiv:2608 00094, 2026 https://arxiv.org/ 。。abs/2608.00094 URL 。[13] Junhao Chen、Xinghao Chen、Henghaofan Zhang、Zihao Qiao、Saining Zhang、Yongzhi Li、Ruqi Huang、Sisi Li、Yimin Sheng、Jianyi Zhu 和 Hao Zhao。引擎原生可编辑三维世界重建：包含物体与光照，2026年。[14] Junhao Chen、Jing Kong、Mingjin Chen、Boran Zhang、Saining Zhang、Hao Zhao、Ruqi Huang、Fei Ma 和 Qi Tian。UniGarment：拓扑引导的纹理规范化，用于可用于仿真的服装数字化。载于《ACM国际多媒体会议论文集》，2026年。[15] Junhao Chen、Boran Zhang、Mingjin Chen、Henghaofan Zhang、Saining Zhang、Congcong Zhu、Hao Zhao、Ruqi Huang、Zhihao Li 和 Yufei Wang。一段视频，一个世界：将单目视频转换为物理四维场景，

2026年。

[16] Tianxing Chen、Zanxin Chen、Baijun Chen、Zijian Cai、Yibin Liu、Zixuan Li、Qiwei Liang、Xianliang Lin、Yiheng Ge、Zhenyu Gu、Weiliang Deng、Yubin Guo、Tian Nian、Xuanbing Xie、Qiangyu Chen、Kailun Su、Tianling Xu、Guodong Liu、Mengkang Hu、Huan-ang Gao、Kaixuan Wang、Zhixuan Liang、Yusen Qin、Xiaokang Yang、Ping Luo 和 Yao Mu。《RoboTwin 2.0：一种具备强域随机化能力、用于鲁棒双臂机器人操作的可扩展数据生成器与基准测试》，2025年。网址 https://arxiv.org/abs/2506.18088。[17] Tianxing Chen、Yue Chen、Zixuan Li、Junyuan Tang、Kailun Su、Weijie Wan、Baijun Chen、Haoran Lu、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 等。《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测试》，2026年。[18] Xiaoxue Chen、Hao Zhao、Guyue Zhou 和 Ya-Qin Zhang。《PQ-Transformer：从点云联合解析三维物体与布局》。IEEE机器人与自动化快报，7(2):2519–2526，2022年。doi：10.1109/LRA. 2022.3143224。网址 https://arxiv.org/abs/2109.05566。[19] Cheng Chi、Zhenjia Xu、Siyuan Feng、Eric Cousineau、Yilun Du、Benjamin Burchfiel、Russ Tedrake 和 Shuran Song。《扩散策略：通过动作扩散进行视觉运动策略学习》，

2023 https://arxiv.org/abs/ 。网

2303.04137

址。[20] Jasmine Collins、Shubham Goel、Kenan Deng、Achleshwar Luthra、Leon Xu、Erhan Gundogdu、Xi Zhang、Tomas F. Yago Vicente、Thomas Dideriksen、Himanshu Arora、Matthieu Guillaumin 和 Jitendra Malik。《ABO：真实世界三维物体理解数据集与基准测试》，2021年。网址 https://arxiv.org/abs/2110.06199。[21] Erwin Coumans 和 Yunfei Bai。《PyBullet：用于游戏、机器人学和机器学习物理仿真的 Python 模块》。 https://github.com/bulletphysics/bullet3，2016—2021年。[22] Rongxu Cui、Zongzheng Zhang、Jingrui Pang、Haohan Chi、Jinbang Guo、Saining Zhang、Shaoxuan Xie、Xin Jin、Yao Mu、Jiaolong Yang、Guocai Yao、Xianyuan Zhan、Ya-Qin Zhang 和 Hao Zhao。《Libero-safety：视觉-语言-动作模型物理与语义安全的综合基准测试》，2026年。[23] Matt Deitke、Eli VanderBilt、Alvaro Herrasti、Luca Weihs、Jordi Salvador、Kiana Ehsani、Winson Han、Eric Kolve、Ali Farhadi、Aniruddha Kembhavi 和 Roozbeh Mottaghi。《ProcTHOR：利用程序化生成实现大规模具身人工智能》，2022年。网址 https://arxiv.org/abs/2206.06994。[24] Matt Deitke、Ruoshi Liu、Matthew Wallingford、Huong Ngo、Oscar Michel、Aditya Kusupati、Alan Fan、Christian Laforte、Vikram Voleti、Samir Yitzhak Gadre、Eli VanderBilt、Aniruddha Kembhavi、Carl Vondrick、Georgia Gkioxari、Kiana Ehsani、Ludwig Schmidt 和 Ali Farhadi。《Objaverse-XL：一个包含 1000 万+ 3D 物体的宇宙》，2023年。网址 https://arxiv.org/abs/2307.05663。[25] Matt Deitke、Dustin Schwenk、Jordi Salvador、Luca Weihs、Oscar Michel、Eli VanderBilt、Ludwig Schmidt、Kiana Ehsani、Aniruddha Kembhavi 和 Ali Farhadi。《Objaverse：带标注的三维物体宇宙》。载于 IEEE/CVF计算机视觉与模式识别会议论文集（CVPR），第

13142 13153, 2023 https://openaccess.thecvf.com/content/CVPR2023/html/Deitke_Objaverse_ 页起 —

A_Universe_of_Annotated_3D_Objects_CVPR_2023_paper.html

June。网址。[26] Laura Downs 、 Anthony Francis 、 Nate Koenig 、 Brandon Kinman 、 Ryan Hickman 、 Krista Reymann、Thomas B. McHugh 和 Vincent Vanhoucke。《Google Scanned Objects：高质量三维扫描家居用品数据集》，2022年。网址 https://arxiv.org/abs/2204.11918。[27] Clemens Eppner、Arsalan Mousavian 和 Dieter Fox。《ACRONYM：基于仿真的大规模抓取数据集》，2020年。网址 https://arxiv.org/abs/2011.09584。[28] 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-动作模型的深入鲁棒性分析》，2025年。[29] Xin Fei、Zhixuan Xu、Huaicong Fang、Tianrui Zhang 和 Lin Shao。《T(R,O) Grasp：用于跨具身灵巧抓取的机器人-物体空间变换高效图扩散》，

2025 https://arxiv. 。网

org/abs/2510.12724

址。[30] Jiashi Feng、Xiu Li、Jing Lin、Jiahang Liu、Gaohong Liu、Weiqiang Lou、Su Ma、Guang Shi、Qinlong Wang、Jun Wang、Zhongcong Xu、Xuanyu Yi、Zihao Yu、Jianfeng Zhang、Yifan Zhu、Rui Chen、Jinxin Chi、Zixian Du、Li Han、Lixin Huang、Kaihua Jiang、Yuhan Li、Guan Luo、Shuguang Wang、Qianyi Wu、Fan Yang、Junyang Zhang 和 Xuanmeng Zhang。《Seed3D 1.0：从图像到高保真、可用于仿真的三维资产》，2025年。网址 https://arxiv.org/abs/2510.19944。

[31] Weixi Feng、Wanrong Zhu、Tsu-jui Fu、Varun Jampani、Arjun Akula、Xuehai He、Sugato Basu、Xin Eric Wang 和 William Yang Wang。LayoutGPT：利用大语言模型进行组合式视觉规划与生成，2023年。URL https://arxiv.org/abs/2305.15393。[32] Huan Fu、Rongfei Jia、Lin Gao、Mingming Gong、Binqiang Zhao、Steve Maybank 和 Dacheng Tao。3D-FUTURE：带纹理的三维家具形状，2020年。URL https://arxiv.org/abs/2009.09633。[33] Huan Fu、Bowen Cai、Lin Gao、Ling-Xiao Zhang、Jiaming Wang、Cao Li、Qixun Zeng、Chengyue Sun、Rongfei Jia、Binqiang Zhao 和 Hao Zhang。3D-FRONT：带布局和语义的三维家具布置房间。发表于 IEEE/CVF国际计算机视觉会议（ICCV）论文集，第10933–10942页。IEEE，2020年10月 2021 https://openaccess.thecvf.com/content/ICCV2021/html/Fu_3D-FRONT_3D_Furnished_Rooms_With_layOuts_and_semaNTics_ICCV_2021_paper.html。

URL 。[34] Rao Fu、Zehao Wen、Zichen Liu 和 Srinath Sridhar。AnyHome：结构化且带纹理的三维住宅开放词汇生成，2023年。URL https://arxiv.org/abs/2312.06644。[35] Daoyi Gao、Yawar Siddiqui、Lei Li 和 Angela Dai。Meshart：利用结构引导变换器生成可动网格。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第618–627页，2025年。[36] Genesis Authors。Genesis：面向机器人及更多领域的生成式通用物理引擎，2024年12月。URL https://github.com/Genesis-Embodied-AI/genesis-world。[37] Haoran Geng、Helin Xu、Chengyang Zhao、Chao Xu、Li Yi、Siyuan Huang 和 He Wang。GAPartNet：通过可泛化且可操作部件实现跨类别、跨领域可泛化的物体感知与操作。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第7081–7091页，2023年6月。URL https://openaccess.thecvf.com/content/CVPR2023/html/Geng_GAPartNet_Cross-Category_Domain-Generalizable_Object_Perception_and_Manipulation_via_Generalizable_and_CVPR_2023_paper.html。[38]

Haoran Geng、Feishi Wang、Songlin Wei、Yuyang Li、Bangjun Wang、Boshi An、Charlie Tianyue Cheng、Haozhe Lou、Peihao Li、Yen-Jen Wang、Yutong Liang、Dylan Goetting、Chaoyi Xu、Haozhe Chen、Yuxi Qian、Yiran Geng、Jiageng Mao、Weikang Wan、Mingtong Zhang、Jiangran Lyu、Siheng Zhao、Jiazhao Zhang、Jialiang Zhang、Chengyang Zhao、Haoran Lu、Yufei Ding、Ran Gong、Yuran Wang、Yuxuan Kuang、Ruihai Wu、Baoxiong Jia、Carlo Sferrazza、Hao Dong、Siyuan Huang、Yue Wang、Jitendra Malik 和 Pieter Abbeel。Roboverse：面向可扩展且可泛化机器人学习的统一平台、数据集与基准。发表于机器人学：科学与系统（RSS）论文集，2025年。doi：10.15607/RSS.2025.XXI.022。[39] Ran Gong、Jiangyong Huang、Yizhou Zhao、Haoran Geng、Xiaofeng Gao、Qingyang Wu、Wensi Ai、Ziheng Zhou、Demetri Terzopoulos、Song-Chun Zhu、Baoxiong Jia 和 Siyuan Huang。Arnold：用于在真实三维场景中开展语言 grounding 任务学习的连续状态基准。发表于 IEEE/CVF国际计算机视觉会议（ICCV）论文集，第20483–20495页，2023年。[40] Diandian Gu、Jing Lin、Gaohong Liu、Jiahang Liu、Su Ma、Guang Shi、Jun Wang、Qinlong Wang、Qianyi Wu、Zhongcong Xu、Xuanyu Yi、Zihao Yu、Jianfeng Zhang、Zhuolin Zheng、Yifan Zhu、Rui Chen、Hengkai Guo、Xiaoyang Guo、Mingcong Han、Xu Han、Xiu Li、Yixun Liang、Weiqiang Lou、Junzhe Lu、Guan Luo、Minghan Qin、Shuguang Wang 和 Yuang Wang。Seed3D 2.0：推进高保真可仿真三维内容生成，2026年。URL https://arxiv.org/abs/2605.13862。[41] Jiayuan Gu、Fanbo Xiang、Xuanlin Li、Zhan Ling、Xiqiang Liu、Tongzhou Mu、Yihe Tang、Stone Tao、Xinyue Wei、Yunchao Yao、Xiaodi Yuan、Pengwei Xie、Zhiao Huang、Rui Chen 和 Hao Su。ManiSkill2：可泛化操作技能的统一基准，2023年。URL https://arxiv.org/abs/2302.04659。[42] Jiarui Guo、Saining Zhang、Jiaying Wan、Jiawei Weng、Junhao Chen、Rui Zhang、Yuanbo Lyu、Zhiheng Lee、Peishuo Li、Jingrun Du、Weiqing Xiao、Weiyi Lin、Zenan Meng、Hu Wei、Lin Qu、Zongzheng Zhang、Hao Zhao 和 Chaojian Li。RFW-S：带独立仿真验证的日常可变形物体。arXiv预印本 arXiv:2611.23056，2026年。URL https://arxiv.org/abs/2611.23056。[43] Yumeng He、Eachen Soong、Xiaotian Yang、Weijia Zhang、Zanwei Zhou、Junru Gong、Xiaokang Yang 和 Yunbo Wang。Neoworld-pro：从单目图像编程交互式场景，用于具身仿真，2026年。URL https://arxiv.org/abs/2608.24212。[44] Zihao Huang、Tianqi Liu、Zhaoxi Chen、Shaocong Xu、Saining Zhang、Lixing Xiao、Zhiguo Cao、Wei Li、Hao Zhao,

以及 Ziwei Liu。Arthoi：通过利用视频先验进行4D重建，合成关节式人-物交互。IEEE模式分析与机器智能汇刊，2026年。doi：10.1109/TPAMI.2026.3747767。[45] Denys Iliash、Jiayi Liu、Egor Fokin、Qirui Wu、Ali Mahdavi-Amiri、Manolis Savva 和 Angel X. Chang。Artiverse：面向可动对象的多样化物理基础数据集。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第8932–8942页，2025年6月 2026 https://openaccess.thecvf.com/content/CVPR2026/html/Iliash_Artiverse_A_Diverse_and_Physically_Grounded_Dataset_for_Articulated_Objects_CVPR_2026_paper.html 。URL 。[46] Stephen James、Zicong Ma、David Rovick Arrojo 和 Andrew J. Davison。RLBench：机器人学习基准与学习环境。IEEE机器人与自动化快报，5(2)：3019–3026，2020年。doi：10.1109/LRA.2020.2974707。[47] Yanqin Jiang、Li Zhang、Jin Gao、Weiming Hu 和 Yao Yao。Consistent4d：从单目视频一致地生成360度动态物体。发表于国际学习表征会议（ICLR），2024年。[48] Zhenyu Jiang、Cheng-Chun Hsu 和 Yuke Zhu。Ditto：通过交互构建可动对象的数字孪生。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第5616–5626页，2022年。[49] Bu Jin、Weize Li、Songen Gu、Yupeng Zheng、Yuhang Zheng、Zhengyi Zhou 和 Yao Yao。Uniart：用于生成具有开放集关节结构的三维可动对象的统一三维表示，2025年。[50] Mathilde Kappel、Clémence Grislain、Mohamed Chetouani、Olivier Sigaud、Louis Annabi、Faïz Ben Amar、Stéphane Doncieux 和 Mahdi Khoramshahi。用于训练和部署基于块的VLA操作策略的仿真到现实集成流水线。arXiv预印本 arXiv: 2609 21817, 2026 https://arxiv.org/abs/2609.21817 。URL 。[51] Alexander Khazatsky、Karl Pertsch、Suraj Nair、Ashwin Balakrishna、Sudeep Dasari、Siddharth Singh、Siddharth Karamcheti、Soroush Nasiriany、Mohan Kumar Srirama、Lawrence Yunliang Chen 等。DROID：大规模野外机器人操作数据集。发表于机器人学：科学与系统（RSS）论文集，2024年。doi：10.15607/RSS.2024.XX.120。[52] Eric Kolve、Roozbeh Mottaghi、Winson Han、Eli VanderBilt、Luca Weihs、Alvaro Herrasti、Matt Deitke、Kiana Ehsani、Daniel Gordon、Yuke Zhu、Aniruddha Kembhavi、Abhinav Gupta 和 Ali Farhadi。AI2-THOR：面向视觉人工智能的交互式三维环境，2017年。URL https://arxiv.org/abs/1712.05474。[53] Chengshu Li、Fei Xia、Roberto Martín-Martín、Michael Lingelbach、Sanjana Srivastava、Bokui Shen、Kent Vainio、Cem Gokmen、Gokul Dharan、Tanish Jain、Andrey Kurenkov、C. Karen Liu、Hyowon Gweon、Jiajun Wu、Li Fei-Fei 和 Silvio Savarese。iGibson 2.0：面向日常家庭任务机器人学习的以物体为中心的仿真，2021年。URL https://arxiv.org/abs/2108.03272。[54] Chengshu Li、Ruohan Zhang、Josiah Wong、Cem Gokmen、Sanjana Srivastava、Roberto Martín-Martín、Chen Wang、Gabrael Levine、Michael Lingelbach、Jiankai Sun、Mona Anvari、Minjune Hwang、Manasi Sharma、Arman Aydin、Dhruva Bansal、Samuel Hunter、Kyu-Young Kim、Alan Lou、Caleb R Matthews、Ivan Villa-Renteria、Jerry Huayang Tang、Claire Tang、Fei Xia、Silvio Savarese、Hyowon Gweon、Karen Liu、Jiajun Wu 和 Li Fei-Fei。BEHAVIOR-1K：包含1,000项日常活动和逼真仿真的具身人工智能基准。发表于第六届机器人学习会议论文集，《机器学习研究论文集》第205卷，第80–93页。PMLR，2023年12月14–18日。URL https://proceedings.mlr.press/v205/li23a.html。[55] Gen Li、Nan Wang、Yunlong Li、Ziyang Yan、Yinghao Shuai、Saining Zhang、Shu Han、Shaocong Xu、Baijun Ye、Lu Zhang、Luan Zhang、Paul Barsch、Zebang Shen、Chaojian Li、Hang Zhao、Ya-Qin Zhang 和 Hao Zhao。Gen-ncap：用于端到端自动驾驶极端情况基准测试的生成式模拟器。发表于IASEAI会议论文集，2026年。[56] Weiyu Li、Xuanyang Zhang、Zheng Sun、Di Qi、Hao Li、Wei Cheng、Weiwei Cai、Shihao Wu、Jiarui Liu、Zihao Wang、Xiao Chen、Feipeng Tian、Jianxiong Pan、Zeming Li、Gang Yu、Xiangyu Zhang、Daxin Jiang 和 Ping Tan。Step1X-3D：迈向高保真、可控的带纹理三维资产生成，2025年。URL https://arxiv.org/abs/2505.07747。[57] Xian Li、Rong Wei、Lujie Yang、Haolin Huang、Junyuan Fang、Siliang Tang、Jun Xiao、Rui Tang 和 Juncheng Li。UniPhysGen：面向仿真就绪 3 2026 https://arxiv.org/abs/2607.13586 D资产的统一物理落地。URL 。

[58] Zhibing Li、Mengchen Zhang、Tong Wu、Jing Tan、Jiaqi Wang 和 Dahua Lin。Ss4d：通过结构化时空潜变量实现原生4D生成模型。ACM 图形学汇刊，44(6)，2025年。doi：10.1145/3763302。[59] Chendi Lin、Heshan Liu、Qunshu Lin、Zachary Bright、Shitao Tang、Yihui He、Minghao Liu、Ling Zhu 和 Cindy Le。Objaverse ++ 3 2025 https：带 //arxiv.org/abs/2504.07334 质量标注的精选D物体数据集，。URL 。[60] Chenguo Lin 和 Yadong Mu。InstructScene：基于语义图先验的指令驱动的3D室内场景合成，2024年。URL https://arxiv.org/abs/2402.04717。[61] Lu Ling、Chen-Hsuan Lin、Tsung-Yi Lin、Yifan Ding、Yu Zeng、Yichen Sheng、Yunhao Ge、Ming-Yu Liu、Aniket Bera 和 Zhaoshuo Li。Scenethesis：用于3D场景生成的语言与视觉智能体驱动框架，2025年。URL https://arxiv.org/abs/2505.02836。[62] Bo Liu、Yifeng Zhu、Chongkai Gao、Yihao Feng、Qiang Liu、Yuke Zhu 和 Peter Stone。Libero：机器人终身学习中的知识迁移基准测试。载于神经信息处理系统进展（NeurIPS），第36卷，2023年。[63] Isabella Liu、Zhan Xu、Wang Yifan、Hao Tan、Zexiang Xu、Xiaolong Wang、Hao Su 和 Zifan Shi。Riganything：面向多样化3D资产的无模板自回归绑定。ACM 图形学汇刊，44(4)，2025年。doi：10.1145/3731149。[64] Jiayi Liu、Ali Mahdavi-Amiri 和 Manolis Savva。Paris：可动对象的部件级重建与运动分析。载于IEEE/CVF国际计算机视觉会议（ICCV）论文集，第352–363页，2023年。[65] Jiayi Liu、Denys Iliash、Angel X. Chang、Manolis Savva 和 Ali Mahdavi-Amiri。Singapo：从单张图像控制对象中可动部件的生成。载于国际学习表征会议（ICLR），2025年。[66] Tianqi Liu、Zhaoxi Chen、Zihao Huang、Shaocong Xu、Saining Zhang、Chongjie Ye、Bohan Li、Zhiguo Cao、Wei Li、Hao Zhao 和 Ziwei Liu。Light-x：带相机与光照控制的生成式4D视频渲染。载于国际学习表征会议，2026年。URL https://arxiv.org/abs/2512.05115。[67] Xinyu Liu、Beiwen Tian、Zhen Wang、Rui Wang、Kehua Sheng、Bo Zhang、Hao Zhao 和 Guyue Zhou。深入探究形状感知零样本语义分割。载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第2999–3009页，2023年。doi：10.1109/ CVPR 52729 2023 00293 https://openaccess.thecvf.com/content/CVPR2023/html/Liu_Delving_Into_ 。 。 Shape-Aware_Zero-Shot_Semantic_Segmentation_CVPR_2023_paper.html URL 。[68] Yu Liu、Baoxiong Jia、Ruijie Lu、Chuyue Gan、Huayu Chen、Junfeng Ni、Song-Chun Zhu 和 Siyuan Huang。Videoartgs：从单目视频构建可动对象的数字孪生，2025年。[69] Yu Liu、Baoxiong Jia、Ruijie Lu、Junfeng Ni、Song-Chun Zhu 和 Siyuan Huang。通过高斯泼溅构建复杂可动结构物体的可交互复制品。载于国际学习表征会议（ICLR），2025年。[70] Ruijie Lu、Yu Liu、Jiaxiang Tang、Junfeng Ni、Yuxiang Wang、Diwen Wan、Gang Zeng、Yixin Chen 和 Siyuan Huang。从单张图像生成具有部件关节结构的物体。载于SIGGRAPH Asia 2025会议论文集，2025年。[71] Yuanbo Lyu、Changjian Pan、Saining Zhang、Yi Yang、Peishuo Li、Zhiheng Li、Jialu Liu、Junhao Chen、Jiawei Weng、Yupeng Gao、Weiqing Xiao、Wenyi Li、Jiarui Guo、Rui Zhang、Jingrun Du、Weiyi Lin、Zenan Meng、Hu Wei、Lin Qu、Zongzheng Zhang、Hao Zhao 和 Chaojian Li。RFW-A：智能体程序化生成覆盖500个类别的40万个可动关节资产。arXiv预印本 arXiv:2610.40333，2026年。URL https://arxiv.org/abs/2610.40333。[72] Changfeng Ma、Yang Li、Xinhao Yan、Jiachen Xu、Yunhan Yang、Chunshi Wang、Zibo Zhao、Yanwen Guo、Zhuo Chen 和 Chunchao Guo。P 3 3 2025 https://arxiv.org/abs/2509.- 06784 SAM：原生D部件分割，。URL 。[73] Ajay Mandlekar、Danfei Xu、Josiah Wong、Soroush Nasiriany、Chen Wang、Rohun Kulkarni、Li Fei-Fei、Silvio Savarese、Yuke Zhu 和 Roberto Martin-Martin。离线人类示范学习对机器人操作的重要因素。载于第5 1690，2022年 164，1678 届机器人学习会议，卷 页 – 。

[74] Ajay Mandlekar、Soroush Nasiriany、Bowen Wen、Iretiayo Akinola、Yashraj Narang、Linxi Fan、Yuke Zhu 和 Dieter Fox。MimicGen：利用人类示范实现可扩展机器人学习的数据生成系统，2023年。URL https://arxiv.org/abs/2310.17596。[75] Oier Mees、Lukas Hermann、Erick Rosete-Beas 和 Wolfram Burgard。Calvin：面向长时域机器人操作任务的语言条件策略学习基准。IEEE机器人与自动化快报，7(3)：7327–7334，2022年。doi：10.1109/LRA.2022.3180108。[76] Kaichun Mo、Shilin Zhu、Angel X. Chang、Li Yi、Subarna Tripathi、Leonidas J. Guibas 和 Hao Su。Part-Net：细粒度分层部件级三维对象理解的大规模基准测试。载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第909–918页，2025年6月。URL 。[77] Soroush Nasiriany、Abhiram Maddukuri、Lance Zhang、Adeet Parikh、Aaron Lo、Abhishek Joshi、Ajay Mandlekar 和 Yuke Zhu。RoboCasa：面向通用机器人的日常任务大规模仿真，2024年。URL https://arxiv.org/abs/2406.02523。[78] NVIDIA。Isaac Sim。 https://developer.nvidia.com/isaac/sim/，未注明日期。[79] Octo模型团队、Dibya Ghosh、Homer Walke、Karl Pertsch、Kevin Black、Oier Mees、Sudeep Dasari、Joey Hejna、Tobias Kreiman、Charles Xu、Jianlan Luo、You Liang Tan、Lawrence Yunliang Chen、Pannag Sanketi、Quan Vuong、Ted Xiao、Dorsa Sadigh、Chelsea Finn 和 Sergey Levine。Octo：开源通用机器人策略，2024年。URL https://arxiv.org/abs/2405.12213。[80] Open X-Embodiment协作组、Abby O'Neill、Abdul Rehman 等。Open X-Embodiment：机器人学习数据集与RT-X模型，2023年。URL https://arxiv.org/abs/2310.08864。[81] 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。URL https://arxiv.org/abs/2504.16054。[82] Xavier Puig、Eric Undersander、Andrew Szot、Mikael Dallaire Cote、Tsung-Yen Yang、Ruslan Partsey、Ruta Desai、Alexander William Clegg、Michal Hlavac、So Yeon Min、Vladimír Vondruš、Theophile Gervet、Vincent-Pierre Berges、John M. Turner、Oleksandr Maksymets、Zsolt Kira、Mrinal Kalakrishnan、Jitendra Malik、Devendra Singh Chaplot、Unnat Jain、Dhruv Batra、Akshara Rai 和 Roozbeh Mottaghi。Habitat 3.0：人类、化身与机器人的协同栖居环境，2023年。URL https://arxiv.org/abs/2310.13724。[83] Minghan Qin、Yuang Wang、Xiuyu Yang、Yushi Long、Yujian Zhang、Ruihuan Wang、Kai Ye、Yangang Zhang 和 Hang Li。Lucida：用于可组合真实到仿真场景建模的解析、生成与放置，2026年。URL https://arxiv.org/abs/2608.30821。[84] Qwen团队。Qwen 3 7 2026 https://qwen.ai/blog?id=qwen3. -7-plus

Plus：多模态智能体智能，5月。URL 。[85] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark、Gretchen Krueger 和 Ilya Sutskever。从自然语言监督中学习可迁移视觉模型。载于第38届国际机器学习会议论文集，第139卷《机器学习研究论文集》，第8748–8763页。PMLR，2021年7月18–24日。URL https://proceedings.mlr.press/v139/radford21a.html。[86] Alexander Raistrick、Lahav Lipson、Zeyu Ma、Lingjie Mei、Mingzhe Wang、Yiming Zuo、Karhan Kayan、Hongyu Wen、Beining Han、Yihan Wang、Alejandro Newell、Hei Law、Ankit Goyal、Kaiyu Yang 和 Jia Deng。使用程序化生成构建无限逼真世界，2023年。URL https://arxiv.org/abs/2306.09310。[87] Alexander Raistrick、Lingjie Mei、Karhan Kayan、David Yan、Yiming Zuo、Beining Han、Hongyu Wen、Meenal Parakh、Stamatis Alexandropoulos、Lahav Lipson、Zeyu Ma 和 Jia Deng。Infinigen Indoors：使用程序化生成构建逼真室内场景，2024年。URL https://arxiv.org/abs/2406.11824。[88] Jiawei Ren、Liang Pan、Jiaxiang Tang、Chi Zhang、Ang Cao、Gang Zeng 和 Ziwei Liu。Dreamgaussian4d：生成式四维高斯泼溅。arXiv预印本 arXiv:2312.17142，2023年。

[89] Licheng Shen、Saining Zhang、Honghan Li、Peilin Yang、Zihao Huang、Zongzheng Zhang 和 Hao Zhao。Gaussianart：面向可动对象的几何与运动统一建模，2025年。[90] Yinghao Shuai、Ran Yu、Yuantao Chen、Zijian Jiang、Xiaowei Song、Nan Wang、Jv Zheng、Jianzhu Ma、Meng Yang、Zhicheng Wang、Wenbo Ding 和 Hao Zhao。PUGS：利用高斯泼溅实现零样本物理理解。载于 2025年 IEEE机器人与自动化国际会议（ICRA），第4478–4485页，2025年。doi：10.1109/ICRA55743.2025.11128683。URL https://arxiv.org/abs/2502.12231。[91] Yuan Shui、Yandong Guan、Zhanwei Zhang、Juncheng Hu、Jing Zhang、Dong Xu 和 Qian Yu。Articad：通过多智能体代码生成进行铰接式CAD装配设计，2026年。[92] Jiayi Su、Youhe Feng、Zheng Li、Jinhua Song、Yangfan He、Botao Ren 和 Botian Xu。Artformer：可控生成多样化三维铰接物体。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第1894–1904页，2025年。[93] Andrew Szot、Alex Clegg、Eric Undersander、Erik Wijmans、Yili Zhao、John Turner、Noah Maestre、Mustafa Mukadam、Devendra Chaplot、Oleksandr Maksymets、Aaron Gokaslan、Vladimir Vondrus、Sameer Dharur、Franziska Meier、Wojciech Galuba、Angel Chang、Zsolt Kira、Vladlen Koltun、Jitendra Malik、Manolis Savva 和 Dhruv Batra。Habitat 2 0 2021 https：训练家庭助手重新整理其栖息环境，。URL 。[94] Stone Tao、Fanbo Xiang、Arth Shukla、Yuzhe Qin、Xander Hinrichsen、Xiaodi Yuan、Chen Bao、Xinsong Lin、Yulin Liu、Tse-kai Chan、Yuan Gao、Xuanlin Li、Tongzhou Mu、Nan Xiao、Arnav Gurha、Viswesh Nagaswamy Rajesh、Yong Woo Choi、Yen-Ru Chen、Zhiao Huang、Roberto Calandra、Rui Chen、Shan Luo 和 Hao Su。ManiSkill3：面向可泛化具身人工智能的GPU并行机器人仿真与渲染，2024年。URL https://arxiv.org/abs/2410.00425。[95] Team Hunyuan3D、Shuhui Yang、Mingxin Yang、Yifei Feng、Xin Huang、Sheng Zhang、Zebin He、Di Luo、Haolin Liu、Yunfei Zhao、Qingxiang Lin、Zeqiang Lai、Xianghui Yang、Huiwen Shi、Zibo Zhao、Bowen Zhang、Hongyu Yan、Lifu Wang、Sicong Liu、Jihong Zhang、Meng Chen、Liang Dong、Yiwen Jia、Yulin Cai、Jiaao Yu、Yixuan Tang、Dongyuan Guo、Junlin Yu、Hao Zhang、Zheng Ye、Peng He、Runzhou Wu、Shida Wei、Chao Zhang、Yonghao Tan、Yifu Sun、Lin Niu、Shirui Huang、Bojian Zheng、Shu Liu、Shilin Chen、Xiang Yuan、Xiaofeng Yang、Kai Liu、Jianchen Zhu、Peng Chen、Tian Liu、Di Wang、Yuhong Liu、Linus、Jie Jiang、Jingwei Huang 和 Chunchao Guo。Hunyuan3D 2.1：从图像生成具有生产级PBR材质的高保真三维资产，2025年。URL https://arxiv.org/abs/2506.15442。[96] 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智能机器人与系统国际会议，第5026–5033页。IEEE，2012年。doi：10.1109/IROS.2012.6386109。[97] Homer Rich Walke、Kevin Black、Tony Z. Zhao、Quan Vuong、Chongyi Zheng、Philippe Hansen-Estruch、Andre Wang He、Vivek Myers、Moo Jin Kim、Max Du、Abraham Lee、Kuan Fang、Chelsea Finn 和 Sergey Levine。Bridgedata v2：用于大规模机器人学习的数据集。载于第七届机器人学习会议论文集，第229卷，第1723–1736页，2023年。[98] Boyuan Wang、Yue Zhang、Xutao Xue、Xueyu Song 和 Yu Sun。TableVerse：具有基于真实世界布局的大规模桌面数据集，用于可泛化操作，2026年。URL https://arxiv.org/abs/2607.21017。[99] Jiawei Wang、Dingyou Wang、Jiaming Hu、Qixuan Zhang、Jingyi Yu 和 Lan Xu。Kinematify：高自由度铰接物体的开放词汇合成，2025年。[100] Jun Wang、Ying Yuan、Haichuan Che、Haozhi Qi、Yi Ma、Jitendra Malik 和 Xiaolong Wang。从学习旋转“笔”中获得的经验。载于机器人学习会议，2024年。URL https://arxiv.org/abs/2407.18902。[101] Kaixuan Wang、Tianxing Chen、Jiawei Liu、Honghao Su、Shaolong Zhu、Minxuan Wang、Zixuan Li、Yue Chen、Huan-ang Gao、Yusen Qin、Jiawei Wang、Qixuan Zhang、Lan Xu、Jingyi Yu、Yao Mu 和 Ping Luo。ManiTwin：将可用于数据生成的数字对象数据集扩展至 100 2026 https://arxiv.org/abs/2603. K，。16866 URL 。[102] Lirui Wang、Yiyang Ling、Zhecheng Yuan、Mohit Shridhar、Chen Bao、Yuzhe Qin、Bailin Wang、Huazhe Xu 和 Xiaolong Wang。GenSim：通过大语言模型生成机器人仿真任务，2023年。URL https://arxiv.org/abs/2310.01361。[103] Nan Wang、Yuantao Chen、Lixing Xiao、Weiqing Xiao、Bohan Li、Zhaoxi Chen、Chongjie Ye、Shaocong Xu、Saining Zhang、Ziyang Yan、Pierre Merriaux、Lei Lei、Tianfan Xue 和 Hao Zhao。统一外观编码

以及用于驾驶场景高斯泼溅的双边网格。载于《神经信息处理系统进展》，第38卷，2025年。URL https://arxiv.org/abs/2506.05280。[104] Penghao Wang、Siyuan Xie、Hongyu Yan、Xianghui Yang、Jingwei Huang、Chunchao Guo 和 Jiayuan Gu。Artllm：通过3D 大语言模型生成关节式资产。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第34281–34291页，2026年。[105] Ruicheng Wang、Jialiang Zhang、Jiayi Chen、Yinzhen Xu、Puhao Li、Tengyu Liu 和 He Wang。DexGraspNet：基于仿真的面向一般物体的大规模机器人灵巧抓取数据集，

2022 https: 。

//arxiv.org/abs/2210.02697

URL 。[106] Xinjie Wang、Liu Liu、Yu Cao、Ruiqi Wu、Wenkang Qin、Dehui Wang、Wei Sui 和 Zhizhong Su。EmbodiedGen：迈向具身智能的生成式

3 2025 https://arxiv.org/abs/2506. D世

10600

界引擎，。URL 。[107] Xinjie Wang、Liu Liu、Taojun Ding、Andrew Choi、Chaodong Huang、Mengao Zhao、Ziang Li、Jackson Jiang、Chunlei Yu、Shengxiang Liu、Wei Xu 和 Zhizhong Su。EmbodiedGen V2：面向具身人工智能的智能体驱动、仿真就绪3D世界引擎，2026年。URL https://arxiv.org/abs/2607.07459。[108] Yufei Wang、Zhou Xian、Feng Chen、Tsun-Hsuan Wang、Yian Wang、Katerina Fragkiadaki、Zackory Erickson、David Held 和 Chuang Gan。RoboGen：通过生成式仿真释放自动化机器人学习的无限数据，2023年。URL https://arxiv.org/abs/2311.01455。[109] Xinyue Wei、Minghua Liu、Zhan Ling 和 Hao Su。采用碰撞感知凹度与树搜索的三维网格近似凸分解，2022年。URL https://arxiv.org/abs/2205.02961。[110] Jiawei Weng、Saining Zhang、Zhenxin Diao、Peishuo Li、Henghaofan Zhang、Junhao Chen 和 Hao Zhao。Feedforward 3 2026，2026年，三维编辑从语义部件变换中学习。载于亚洲计算机图形学会议论文集。[111] Yijia Weng、Bowen Wen、Jonathan Tremblay、Valts Blukis、Dieter Fox、Leonidas Guibas 和 Stan Birchfield。用于构建未知可动对象数字孪生的神经隐式表示。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第3141–3150页，2024年。[112] Di Wu、Liu Liu、Anran Huang、Yuyan Liu、Qiaojun Yu、Shaofan Liu、Liangtu Song 和 Cewu Lu。Reartgs++：通过平面高斯泼溅和时序几何约束实现可泛化关节结构重建。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第1177–1186页，2026年。[113] Tong Wu、Jiarui Zhang、Xiao Fu、Yuxin Wang、Jiawei Ren、Liang Pan、Wayne Wu、Lei Yang、Jiaqi Wang、Chen Qian、Dahua Lin 和 Ziwei Liu。OmniObject3D：用于真实感知、重建与生成的大词汇量3D物体数据集，2023年。URL https://arxiv.org/abs/2301.07525。[114] Hongchi Xia、Xuan Li、Zhaoshuo Li、Qianli Ma、Jiashu Xu、Ming-Yu Liu、Yin Cui、Tsung-Yi Lin、Wei-Chiu Ma、Shenlong Wang、Shuran Song 和 Fangyin Wei。SAGE：面向具身人工智能的可扩展智能体驱动3D场景生成，2026年。URL https://arxiv.org/abs/2602.10116。[115] Fanbo Xiang、Yuzhe Qin、Kaichun Mo、Yikuan Xia、Hao Zhu、Fangchen Liu、Minghua Liu、Hanxiao Jiang、Yifu Yuan、He Wang、Li Yi、Angel X. Chang、Leonidas J. Guibas 和 Hao Su。SAPIEN：一个基于模拟部件的交互式环境。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第

11097 11107 2020 https://openaccess.thecvf.com/content_CVPR_2020/ –页。

html/Xiang_SAPIEN_A_SimulAted_Part-Based_Interactive_ENvironment_CVPR_2020_paper.html

IEEE，六月。URL 。[116] Jianfeng Xiang、Zelong Lv、Sicheng Xu、Yu Deng、Ruicheng Wang、Bowen Zhang、Dong Chen、Xin Tong 和 Jiaolong Yang。用于可扩展且多用途D生成的结构化

3 3 2024 https: D潜

//arxiv.org/abs/2412.01506

变量，。URL 。[117] Jianfeng Xiang、Xiaoxue Chen、Sicheng Xu、Ruicheng Wang、Zelong Lv、Yu Deng、Hongyuan Zhu、Yue Dong、Hao Zhao、Nicholas Jing Yuan 和 Jiaolong Yang。用于3D生成的原生紧凑结构化潜变量，2025年。URL https://arxiv.org/abs/2512.14692。[118] Yiming Xie、Chun-Han Yao、Vikram Voleti、Huaizu Jiang 和 Varun Jampani。Sv4d：具有多帧和多视图一致性的动态3D内容生成。载于国际学习表征会议（ICLR），2025年。

[119] Haoran Xu、Saining Zhang、Peishuo Li、Baijun Ye、Xiaoxue Chen、Huan-ang Gao、Jv Zheng、Xiaowei Song、Ziqiao Peng、Run Miao、Jinrang Jia、Yifeng Shi、Guangqi Yi、Hang Zhao、Hao Tang、Hongyang Li、Kaicheng Yu 和 Hao Zhao。《Cruise：利用高斯泼溅技术在 V2X 场景中进行协同重建与编辑》。载于 IEEE/RSJ智能机器人与系统国际会议（IROS）论文集，2025年。doi：10.1109/IROS60139.2025.11246201。[120] Ziyang Yan、Yihua Shao、Minwen Liao、Siyu Chen、Nan Wang、Muyuan Lin、Jenq-Neng Hwang、Hao Zhao、Fabio Remondino 和 Lei Li。3DSceneEditor：利用高斯泼溅进行可控3D场景编辑。载于 IEEE/CVF冬季计算机视觉应用会议（WACV）论文集，第1852–1863页，2026年。doi：

10 1109 61042 2026 00185 https://openaccess.thecvf.com/content/WACV2026/html/Yan_ 。/3DSceneEditor_Controllable_3D_Scene_Editing_with_Gaussian_Splatting_WACV_2026_paper.html WACV . . . URL .[121] An Yang、Anfeng Li、Baosong Yang、Beichen Zhang、Binyuan Hui、Bo Zheng、Bowen Yu、Chang Gao、Chengen Huang、Chenxu Lv、Chujie Zheng、Dayiheng Liu、Fan Zhou、Fei Huang、Feng Hu、Hao Ge、Haoran Wei、Huan Lin、Jialong Tang、Jian Yang、Jianhong Tu、Jianwei Zhang、Jianxin Yang、Jiaxi Yang、Jing Zhou、Jingren Zhou、Junyang Lin、Kai Dang、Keqin Bao、Kexin Yang、Le Yu、Lianghao Deng、Mei Li、Mingfeng Xue、Mingze Li、Pei Zhang、Peng Wang、Qin Zhu、Rui Men、Ruize Gao、Shixuan Liu、Shuang Luo、Tianhao Li、Tianyi Tang、Wenbiao Yin、Xingzhang Ren、Xinyu Wang、Xinyu Zhang、Xuancheng Ren、Yang Fan、Yang Su、Yichang Zhang、Yinger Zhang、Yu Wan、Yuqiong Liu、Zekun Wang、Zeyu Cui、Zhenru Zhang、Zhipeng Zhou 和 Zihan Qiu。Qwen3技术报告。arXiv预印本 arXiv:2505.09388，2025年。URL https://arxiv.org/abs/2505.09388。[122] Yandan Yang、Baoxiong Jia、Peiyuan Zhi 和 Siyuan Huang。PhyScene：面向具身人工智能的物理可交互3D场景合成，2024年。URL https://arxiv.org/abs/2404.09465。[123] Yandan Yang、Baoxiong Jia、Shujie Zhang 和 Siyuan Huang。SceneWeaver：配备可扩展的自我反思智能体的一体化3D场景合成，2025年。URL https://arxiv.org/abs/2509.20414。[124] Yi Yang、Hao Pan、Yijing Cui、Alla Sheffer 和 Changjian Li。Sketch2Arti：基于草图的CAD对象关节建模。《ACM 图形学汇刊》，

45 4 2026 10 1145 3811375 https: ( )， 。//arlo-yang.github.io/Sketch2Arti/ doi：。/。URL。[125] Yue Yang、Fan-Yun Sun、Luca Weihs、Eli VanderBilt、Alvaro Herrasti、Winson Han、Jiajun Wu、Nick Haber、Ranjay Krishna、Lingjie Liu、Chris Callison-Burch、Mark Yatskar、Aniruddha Kembhavi 和 Christopher Clark。Holodeck：语言引导生成具身人工智能环境，。URL。[126] Yunhan Yang、Chunshi Wang、Junliang Ye、Yang Li、Zanxin Chen、Zehuan Huang、Yao Mu、Zhuo Chen、Chunchao Guo 和 Xihui Liu。PhysForge：为交互式虚拟世界生成物理基础 3D 资产，2026 年。URL https://arxiv.org/abs/2605.05163。[127] Baijun Ye、Minghui Qin、Saining Zhang、Moonjun Goon、Shaoting Zhu、Zebang Shen、Luan Zhang、Lu Zhang、Hao Zhao 和 Hang Zhao。Gs-occ3d：利用高斯泼溅扩展纯视觉占用重建。载于 IEEE/CVF国际计算机视觉会议论文集，

2025 https://gs-occ3d. 。github.io/

URL。[128] Chongjie Ye、Yushuang Wu、Ziteng Lu、Jiahao Chang、Xiaoyang Guo、Jiaqing Zhou、Hao Zhao 和 Xiaoguang Han。Hi3dgen：通过法线桥接从图像生成高保真 3D 几何。载于 IEEE/CVF 国际计算机视觉会议（ICCV）论文集，第25050–25061页，2025年。doi：10.1109/ ICCV51701.2025.02323。[129] Tianhe Yu、Deirdre Quillen、Zhanpeng He、Ryan Julian、Karol Hausman、Chelsea Finn 和 Sergey Levine。Meta-world：多任务与元强化学习的基准和评估。载于机器人学习会议论文集，第100卷，第1094–1100页，2019年。[130] Tianjiao Yu、Vedant Shah、Muntasir Wahed、Ying Shen、Kiet A. Nguyen 和 Ismini Lourentzou。部件^{2}gs：利用3D高斯泼溅进行可动对象的部件感知建模。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第18913–18923页，2026年。[131] Chuanrui Zhang、Minghan Qin、Yuang Wang、Baifeng Xie、Hang Li 和 Ziwei Wang。Simart：利用多模态大语言模型将整体网格拆解为仿真就绪的可动资产，2026年。[132] Hui Zhang、Sammy Christen、Zicong Fan、Otmar Hilliges 和 Jie Song。GraspXL：大规模生成多样物体的抓取动作，2024年。URL https://arxiv.org/abs/2403.19649。

[133] Jia-Peng Zhang、Cheng-Feng Pu、Meng-Hao Guo、Yan-Pei Cao 和 Shi-Min Hu。《一个模型绑定所有角色：使用 Unirig 进行多样化骨架绑定》。《ACM 图形学汇刊》，44(4):1–18，2025年。doi：10.1145/3730930。[134] Longwen Zhang、Ziyu Wang、Qixuan Zhang、Qiwei Qiu、Anqi Pang、Haoran Jiang、Wei Yang、Lan Xu 和 Jingyi Yu。CLAY：用于创建高质量3D资产的可控大规模生成模型，2024年。URL https://arxiv.org/abs/2406.13897。[135] Saining Zhang、Baijun Ye、Xiaoxue Chen、Yuantao Chen、Zongzheng Zhang、Cheng Peng、Yongliang Shi 和 Hao Zhao。利用跨视角不确定性的无人机辅助道路高斯泼溅。载于英国机器视觉会议，2024年。URL https://bmvc2024.org/proceedings/452/。[136] Shiduo Zhang、Zhe Xu、Peiju Liu、Xiaopeng Yu、Yuan Li、Qinghui Gao、Zhaoye Fei、Zhangyue Yin、Zuxuan Wu、Yu-Gang Jiang 和 Xipeng Qiu。Vlabench：面向语言条件机器人操作与长时程推理任务的大规模基准测试。载于 IEEE/CVF国际计算机视觉会议（ICCV）论文集，第11142–11152页，2025年。[137] Zongzheng Zhang、Jingrui Pang、Zhuo Yang、Kun Li、Minwen Liao、Saining Zhang、Guoxuan Chi、Jinbang Guo、Huan-ang Gao、Modi Shi、Dongyun Ge、Yao Mu、Jiayuan Gu、Rui Chen、Hao Dong、Huazhe Xu、Li Yi、Yixin Zhu、Hang Zhao、Pengwei Wang、Shanghang Zhang、Guocai Yao、Jianyu Chen、Hongyang Li 和 Hao Zhao。Dexora：用于高自由度双臂灵巧操作的开源视觉-语言-动作模型。arXiv预印本 arXiv:2605.18722，2026年。[138] Zongzheng Zhang、Jijun Wang、Saining Zhang、Shuo Wang、Yiru Wang、Hai Yang、Yang Chen、Yuwen Heng、Hao Sun、Anqing Jiang 和 Hao Zhao。面向端到端自动驾驶的即插即用交通元素感知。载于《计算机视觉——ECCV 2026》，第96–117页。施普林格，2026年。doi：10.1007/978-3-032-36839-3_6。URL https://arxiv.org/abs/2608.18035。[139] Yang Zhao、Shizhao Sun、Meisheng Zhang、Yingdong Shi、Xubo Yang 和 Jiang Bian。SceneReVis：通过多轮强化学习进行三维室内场景合成的自反思视觉基础框架，2026年。URL https://arxiv.org/abs/2602.09432。

附录目录

详细方法、评估流程和补充结果。

A.1 数据集与评估集.24 全集普查与计数约定.24 清单与评估范围.24 数据集覆盖范围.24 类别均衡评估.25 发布范围与类别内多样性.26 A.2 资产构建详情.26 生成与类别重新分配.26 模型与提示词规范.27 实体化流程.27 几何清理.27 记录的处理成本.28 A.3 几何与语义评估.28 几何度量指标.28 语义标注指标.29 成对几何与碰撞保真度.29 A.4 物理表示与组合.30 部件级物理属性.30 物理一致性.30 实心填充建模的影响.31 外观、部件与碰撞几何体.32 A.5 跨仿真器评估.32 跌落响应协议.32 一致性指标.32 A.6 灵巧抓取：协议与诊断.33 候选对象生成与执行.33 候选预算.34 鲁棒性与闭合位移.35 A.7 策略数据、训练与评估.36 数据预算与资产选择.36 训练与评估设置.36 聚合与任务级结果.37 跨具身真实机器人轨迹回放.

A.1 数据集与评估集

A.1.1 全集普查与计数惯例

表 S1 记录了用于描述 RFW-W 的清单。RFW-R 纳入其保留的刚性版本。RFW-A 纳入其版本比较中报告的 402,410 项资产版本和 531 个类别 [71]。RFW-S 区分了 100,000 项已完成的生成资产与 83,365 项同时通过物理验证和运行时检查的配置匹配交集中的资产 [42]。其五个类别组包含 168 个软固体类别、89 个布片类别、63 个外壳类别、94 个服装类别和 71 个包袋类别，总计 485 个；通过两项检查后，这些类别仍全部有所体现。

表 S1 RFW-W 资产来源清单。资产总数为各来源的源数据集记录之和，未进行跨来源去重；类别总数按数据集–类别条目计数。各来源的纳入标准分别列出。

来源 / 统计口径	资产	类别 纳入标准
RFW-R	524,611	1,212 个保留的刚性版本
RFW-A	402,410	531 个可动关节版本 [71]
RFW-S（联合验证）	83,365	485 个相同配置的物理与运行时通过项
RFW-W ，数据集总和	**1,010,386**	**2,228**

验证按资产类型进行：RFW-R 使用刚体诊断和跌落响应测试；RFW-A 报告完整版本的运动学检查和一项包含 200 个资产的被动展开研究；RFW-S 则结合冻结的 Genesis PBD/MPM 配置下的物理验收与单独的运行时检查。以下评估使用刚体资产队列，分母见表S3。

A.1.2 清单和评估范围

RFW-R 刚体物体数据集包含 524,611 个资产，涵盖 1,212 个叶类别。表 S2 将其与 EmbodiedGen v 2 [107], PhysX-3D [6],和 ManiTwin [101]进行了比较。我们报告各库分类体系内的熵等效覆盖率$N_{\text{eff}} = \exp(-\sum_c p_c \log p_c)$，其中 $p_c = n_c / \sum_j n_j$。平衡评估从每个符合条件的类别中抽取两个资产（共 2,424 个物体），且不受仿真和操作结果影响。

表 S2 共享评估中使用的刚体发布版本。 N_{eff} 表示各库分类体系内的熵等效类别覆盖率。

数据集	资产	类别	N_{eff}	每类中位数
RFW-R	**524,611**	**1,212**		**356**
EmbodiedGen v2	4,168	200	139.2	11
PhysX-3D	32,047	450	7.8	1
ManiTwin	4,574	485	285.6	6

RFW-R 的有效覆盖范围为 1,212 个类别中的 905.1 个，相比之下，PhysX-3D 为 450 个类别中的 7.8 个。因此，RFW-R 清单分布于众多类别，而 PhysX-3D 清单则集中在其分类体系的一小部分中。RFW-R 每个类别的资产中位数为 356，这一数值通过描述类别内的典型可用性，对该覆盖度量指标进行了补充。

A.1.3 数据集覆盖范围

该目录兼具广泛的类别覆盖范围，并为每个类别提供多个实例。

图 S1 总结了刚性物体目录的广度，展示了来自八个功能组的示例，以及不同领域、类别和估算物理属性的分布情况。



图 S1 RFW-R 数据集覆盖范围。 上：来自八个功能组的 48 个代表性资产，分别调整大小以便清晰展示。下：30 个领域的取整后物体数量、按频率加权的 1,212 个类别词云，以及估计质量、公制高度和部件材料的分布。最终的 RFW-R 发布版本包含 524,611 个保留资产，涵盖 1,212 个类别。

A.1.4 类别均衡评估

我们从每个具有可用几何形状和物理描述的类别中选取两个物体，共计 2,424 个物体。选择过程不取决于仿真或操作结果。这些类别涵盖 30 个领域，包括 121 个头部类别、364 个中部类别和 727 个尾部类别，分别贡献 242、728 和 1,454 个入选物体。

表 S3 评估设置。

评估	Size Unit	设置
刚性物体清单	524,611 资产	1,212 个类别
类别均衡样本	2,424 个资产	每个类别两个
四引擎一致性	2,257 个资产	共同可评估子集
策略训练	2,000 个示范，每个来源和任务 100 个	
策略评估	4,000 回合	每种策略和条件 100 个
场景构建	16 个基础模型	265 个实例，三家提供方

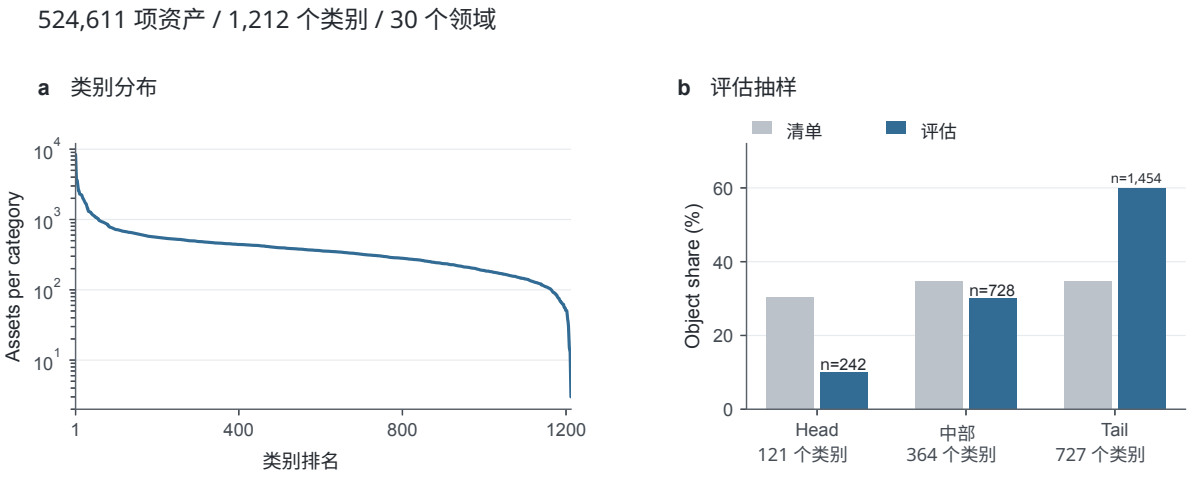

图 S2 刚性物体清单与类别均衡抽样。 类别数量曲线使用对数纵轴。柱形图对比清单占比与评估占比； n 标签给出评估对象数量。抽样时每个类别选取两个物体。

A.1.5 发布范围与类别内多样性

保留清单包含 524,611 项资产；其中 524,562 项（99.99%）已完成仿真导出。本快照中剩余 49 项保留资产尚未完成仿真导出。所有 2,424 项类别均衡评估资产均具有完整的仿真表示。

在包含 2,424 个物体的子集中，图像特征筛查以 0.9782 的余弦相似度识别出 24 个候选近重复簇，涉及 55 个物体（2.27%）。阈值根据同类别和跨类别代理样本对进行校准，以筛选出供目视检查的候选项。在 1,212 个类别内样本对中，图像特征距离的中位数为 0.1108，第 5–95 百分位数为 0.0525–0.1764。

这些特征统计描述了均衡样本内部的变化：采样物体中仅有 2.27% 进入候选近重复簇，而类别内距离的跨度较大。候选簇是用于目视检查的筛选结果，所测比例适用于 2,424 个采样物体。

A.2 资产构建详情

A.2.1 生成与类别重新分配

图像生成、背景去除、网格重建、五视图检查和类别重新分配共同构成保留清单。视觉评估标准检查单物体内容、刚体适用性、完整性、悬浮碎片、可识别性和类别一致性。仅类别一致性不合格的物体会通过词汇匹配和图像条件匹配进行重新分配，随后再次检查破损情况和多物体组合。符合条件但无法匹配的物体则归入新的尺寸受限类别。

表 S4 视觉筛选和类别重新分配后的候选对象生成与保留。

阶段或处置方式	资产
生成的候选项	599,439
保留原标签	357,117
重新分类后保留	147,573
在12个归纳类别中保留	19,921
保留总数	524,611
未保留在此清单中	74,828

通过重新标注或新类别归纳，类别重新分配保留了 167,494 个候选物体，占保留清单的 31.9%。这将语义不匹配与不可用的几何形状区分开来：视觉上合格的物体在修正标签后仍可用于扩大覆盖范围，而另有 74,828 个候选物体未纳入保留清单。

A.2.2 模型和提示词规范

类别条件生成（*L0*）。 Qwen3.7-plus 接收领域名称以及英语和中文类别名称，并以 JSON 格式生成批量英语物体描述。提示词要求每条描述使用 15–30 个词组成一个句子，并改变物体子形态、材料和表面处理、风格和时代、颜色和图案、状态和磨损，以及显著的几何形状。十二种轮换使用的多样性提示规定了材料、风格、状态和颜色的组合。生成使用温度 0.9 和 top-p 0.95。描述会根据 0.85 的词元集 Jaccard 相似度进行去重，并筛查特定领域的材料合理性。FLUX.2-klein 使用六个推理步数和 1.0 的引导尺度，根据选定的描述渲染图像；补充的 Wan2.7 路径则对每个类别中的描述进行采样。

视觉质量和类别提示词（*L2*）。 Qwen3.7-plus [84]接收前面、后面、左侧、右侧和底部渲染图。其有序评估准则要求判断资产是否为单个物体、刚体、完整、没有悬浮碎片、可识别，并且与所要求的领域和类别一致。响应会记录这些判断，以及用于类别重新分配的物体猜测。对于模棱两可的分配，会根据渲染视图和候选类别列表进行判定，并提供弃答选项。后续提示词会对可见破损以及渲染图中独立物体的数量进行评分。

物理属性标注提示词（*L4*）。Qwen3.7-plus 接收对齐的纹理化 RGB 图像和颜色编码的部件掩码视图。提示词利用掩码颜色定位部件，并利用 RGB 视图识别材料类别和子类型。对于每个部件，提示词要求给出其与橡胶之间的静摩擦和动摩擦；对于完整的刚性物体，则要求给出公制尺寸和质量估计。提示词要求针对每种掩码颜色提供一个答案，并结合整个物体的上下文处理较小或被遮挡的部件。随后，材料密度和部件体积决定合成质量、质心和惯性，详见附录 A.4。

A.2.3 实体化流程

图 2 连接了三个实体化阶段。L3 接收通过审核的网格，并生成语义部件标签和凸碰撞组件。二者发挥不同作用：部件用于索引材料和交互标注，而碰撞组件则规定接触几何形状。L4 接收对齐的 RGB 视图和部件掩码，估计物理属性，并在共同物体坐标系中合成质量、质心和惯性（见附录 A.4）。完成 L4 后导出资产，并在不同仿真格式之间保留几何形状和物理量。L5 表示对这些导出资产进行的下游仿真评估。评估结果为物理属性改进提供诊断反馈。跨仿真器评估采用附录 A.5中明确定义的有界运动和穿透协议。

A.2.4 几何清理

干净条件针对重合顶点、碎片化组件和不一致的部件标注，同时保留开口和可抓取结构。类别重新分配与几何清理服务于不同目的：前者恢复类别覆盖范围，后者改进用于标注和导出的表示。对构建阶段进行配对审计后发现，坐标对齐后视觉表面基本保持不变；附录 A.3.3 报告了度量指标。物理归一化在共同物体坐标系中对齐物理属性和几何形状，同时保留视觉表面。

图 S3 区分了三种诊断结果。将 Coupon Box 重新归类为马克杯，可纠正语义索引。残留的开放边界表明，保守清理能够保留视觉拓扑，但不保证形成闭合体积；因此，基于体积的测量需要通过附录 A.3.3中的资格检查。锤子在归档的 15 次跌落响应试验中无一通过，表明该资产始终违反

至少一项响应判据。这一汇总结果表明，该资产未能通过下游评估协议；具体违反了哪项判据及其物理原因，仍需通过试验级诊断确定。

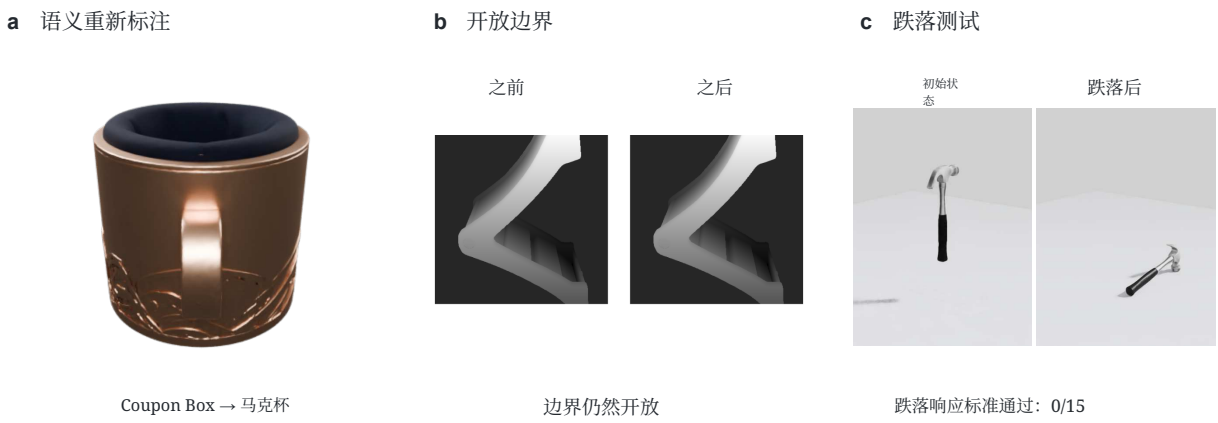

图 S3 观察到的诊断案例。一个源类别从 Coupon Box 重新分配为马克杯；处理后边界仍然开放；一把羊角锤在全部 15 次归档跌落试验中均未达到指定的跌落响应标准。这些面板展示了三个相互独立的诊断示例。

A.2.5 记录的处理成本

表 S5 报告物理归一化、修复和碰撞重建的 CPU 时间。每个阶段均在表中列出的资产集上独立测量。

表 S5 记录的资产处理成本。CPU 小时表示累计计算时间。资产数量对应于各阶段可用的计时记录。

阶段	记录的资产数	CPU 小时	平均秒数/资产
物理归一化	524,611	10,723.2	73.6
修复	2,424	35.4	52.6
碰撞重建	1,291	358.8	1,000.5

各阶段的计时队列有所不同，因为表格汇总的是可用的阶段特定记录：规范化涵盖保留清单，修复涵盖 2,424 个资产，而碰撞重建涵盖 1,291 个资产。在记录的资产中，碰撞重建的平均 CPU 成本最高（1,000.5 秒），表明分解是该测量批次中一项重要的成本。这些累计 CPU 小时反映的是计算资源消耗，而非端到端墙钟延迟。

A.3 几何形状与语义评估

A.3.1 几何度量指标

我们对连通分量、边界边比率、视觉体积相对于凸包体积的比例，以及视觉到碰撞表面距离进行表征。深度侵入用于衡量位于视觉体积内部超过 1 cm 的碰撞采样点，并在水密网格上进行评估。原始诊断设置采用的组件阈值为 5,000、边界比率为 0.30、实心度为 0.20、视觉到碰撞P95距离为 2 cm，侵入比例为 0.02。

PhysX-3D 和 ManiTwin 的基线转换使用视觉几何体进行碰撞检测，因此视觉几何体与碰撞表示之间的距离实际上是自我比较。几何评分因此不仅取决于碰撞表示，也取决于源资产。表S6 报告了这些特定于来源的碰撞表示的诊断结果。

表 S6 原始几何诊断通过率（%）。如正文所述，不同来源的碰撞表示有所不同。

门槛（通过率 ↑）	RFW-R	EmbodiedGen v2	PhysX-3D	ManiTwin
仅几何严格标准	79.3	71.1	53.7	**89.5**
连通性	98.6	98.2	100	100
开放边界边	98.9	96.1	99.9	100
视觉→碰撞距离。	98.8	96.4	100	100
深度穿透	99.5	100	100	100
外壳实心度	81.6	75.0	53.8	**89.5**

外壳实心度是 RFW-R 各单项门槛中通过率最低的（81.6%），接近联合仅几何通过率 79.3%；其他单项门槛的通过率均至少为 98.6%。因此，在这项诊断中，实心度是主要的边际瓶颈。较低的视觉体积与外壳体积比可能源于凹陷和开口，也可能是有缺陷的几何结构，因此该门槛刻画的是在所选阈值下的形状紧致度。PhysX-3D 和 ManiTwin 的视觉几何体与碰撞表示之间的距离通过率为 100%，这是因为比较两侧使用的是相同几何形状。

A.3.2 语义标注指标

四个库共用 Qwen3.7-plus 视觉语言评分与聚合协议，用于评估类别、语言、功能和物理相关一致性。CLIP [85]图文相似度提供了额外的对齐度量。物理相关评审分数用于衡量语义合理性，以及与估算物理标注的一致性。Qwen3.7-plus 还辅助生成标注，因此该分数反映的是基于模型的标注一致性；物理参数准确性则需要通过独立测量来确定。不适用表示已发布的对比来源未按照此协议提供可比较的物理评审字段。

表 S7 共享评审器下的部件级标注质量。四个库均采用相同的评分与聚合协议。 *EmbodiedGen v2 的物理属性标注基于占位值计算；不适用表示没有可比较的字段。

维度	RFW-R	EmbodiedGen v2	PhysX-3D	ManiTwin
类别 ↑	**0.981**	0.93	0.96	0.65
语言 ↑	0.864	0.67	**0.87**	0.50
功能 ↑	**0.860**	0.80	0.79	0.64
物理（判定） ↑	**0.895**	0.85*	N/A	N/A
CLIP 对齐 ↑	0.244	0.232	0.238	**0.261**

RFW-R 的类别（0.981）和功能（0.860）得分最高，而 ManiTwin 的 CLIP 对齐得分最高（0.261，相比之下为 0.244）。这些指标回答的是不同问题：共享评审器评估指定的标注维度，而 CLIP 评估图像–文本嵌入对齐。排名不同表明，广义的视觉–文本对齐与任务相关标注一致性应分别解读。

A.3.3 配对几何与碰撞保真度

配对审计追踪了 2,424 个资产经过几何清理和物理归一化的情况。坐标对齐后，物理归一化前后每个对象的双向表面距离 P95 中位数均为 3.87×10^{-7}m。归一化前后，水密性均约为 10.1%，这与视觉表面的保留情况一致。视觉网格闭合与模拟器执行是彼此独立的属性：接触由碰撞表示处理，而体积 IoU 则要求使用闭合视觉体积代理。

另取由 200 个资产组成的分层抽样，对比发布的碰撞几何体、标准 CoACD 和单一凸包。我们在 25 个具有有效闭合视觉体积代理的物体上计算体积 IoU。其余 175 个抽样物体缺少用于该计算的有效闭合视觉体积代理，因此符合条件的样本占比为 12.5%

。在同样的 25 个符合条件的物体上，发布的表示取得了 0.924 的 IoU 中位数，相比之下，CoACD 为 0.913，单一凸包为 0.682（表 S8）。与单一凸包之间更大的差距与多凸包几何保留凹形结构这一点相一致。该对比刻画了封闭代理子集；上文的表面距离诊断则为不具有有效体积代理的几何形状提供了补充证据。

表 S8 从 200 个资产的样本中选取的同一组 25 个可测量物体的碰撞保真度。IoU 用于比较每个碰撞体积与闭合视觉体积代理。

碰撞表示	可测量物体	体积 IoU 中位数 ↑
发布的表示	25	**0.924**
标准 CoACD	25	0.913
单个凸包	25	0.682

标注连续性。 在全部 2,424 个经审查的物体中，物理归一化前后的可供性标注和语义评估输入均保持不变。这验证了物理归一化过程中的连续性。

A.4 物理表示与组成

A.4.1 部件级物理属性

每个部件都将度量几何形状与估算的材料、密度、摩擦和语义属性关联起来。由此得到的量是根据几何形状和材料先验推导出的组合物理标注。给定部件质量 $m_p = \rho_p V_p$，物体质量和质心为

$$M = \sum_p m_p, \qquad c = \frac{1}{M} \sum_p m_p c_p. \tag{1}$$

在物体坐标系中，部件惯性张量通过平行轴定理进行组合。整体物体质量 $\widehat{\ }$ 估计值 M 通过公共因子 M/M 对所有部件质量进行锚定，其中 $M/M \notin [0.5, 2]$。均匀缩放可保持组合后的质心和相对部件质量不变。

图 S4 比较了在相同物体总质量下，已分配的部件质量与均匀密度分配的差异。这隔离了部件标注如何在每个资产内部重新分配质量。

A.4.2 物理一致性

我们在 2,424 个资产上评估已存储物理标注与模拟器导出结果之间的数值一致性。下方的质量误差用于衡量跨格式一致性，参照值为已存储／导出的数值，而非真实物体的实测质量。物体级量随几何形状一同变换至公共坐标系；对齐至共同物体坐标系后，所有经过审计的质心和惯性张量均保持一致。通过 URDF 和 MJCF 恢复的物体质量，其最大相对误差为 1.5×10^{-6}。对于 98.5% 的资产，部件质量总和与已存储物体质量之间的相对误差在容差 10^{-3} 以内。在模拟器导出过程中保留这些物理量，有助于在场景构建和操作策略学习中重复使用 RFW-R 资产。

表 S9 对 2,424 个资产进行的物理表示检查。

检查项	结果
仿真描述有效性	2,424/2,424
物体坐标系下质心与惯性一致性	2,424/2,424
URDF/MJCF 质量相对误差（最大值）↓	1.5×10^{-6}
部件质量总和一致性在 10^{-3} ↑以内	98.5%

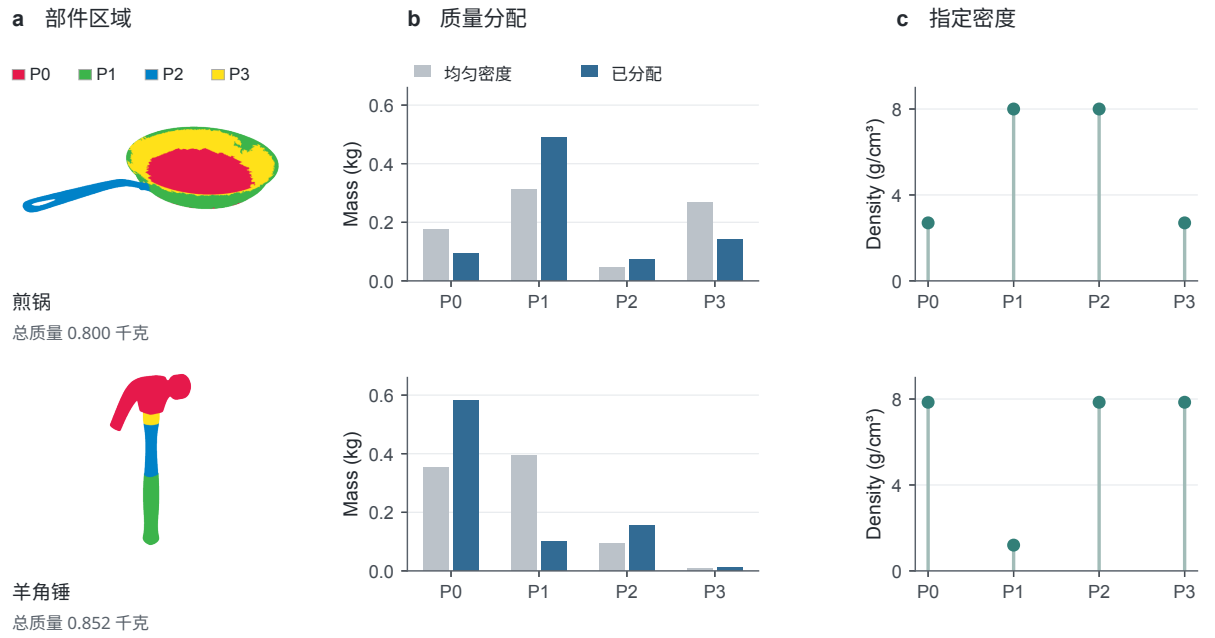

图 S4 按部件的物理分配。 对于每个资产，均匀密度基线按照存储的部件体积比例分配固定总质量。分配质量使用导出的标注；对于平底锅和锤子，两种分配方式的总质量分别均为 0.800 千克和 0.852 千克。该比较展示了部件标注如何在通用对象表示内部重新分配质量。

A.4.3 实心填充建模的影响

我们在保留的 31 个资产上比较原生表示与实心填充。一个因质量分配退化而从资产库中移除的资产，在所有模拟引擎的两种条件下均不纳入比较。比较使用 MuJoCo、Genesis 和 Isaac Sim，每个资产设置三个下落高度和五种姿态，仿真运行时长为 2.5 秒，时间步长为 0.01 秒。两种条件共进行 2,790 次试验，且没有执行失败。试验的通过／失败标签依据附录 A.5中的下落响应标准确定。每个资产根据 15 次试验计算标准通过比例，再汇总各资产的配对差值。

表 S10 隔离了实心填充比较。在共享的位移、旋转和穿透判据下，原生表示在三个引擎中的标准通过比例均高于实心填充。估计降幅介于 7.53 至 9.25 个百分点之间，且三个逐比较区间均低于零。三个引擎呈现出的共同方向表明，跌落响应标准对表示变化具有敏感性。一种可能的机制是，占据体积和质量分布的变化会改变接触响应与旋转运动；要确定几何形状与惯性各自的贡献，则需要明确的控制变量规范。

表 S10 实心填充对保留的 31 个资产的影响。变化值为实心填充减去原生方式在标准通过比例上的差异，以百分点表示。近似的逐比较 95%区间通过计算各资产间的配对均值 ± 1.96 标准误得到。

引擎	平均变化（百分点）	约 95% 区间
MuJoCo	-9.25	$[-16.01, -2.48]$
Genesis	-9.25	$[-15.88, -2.62]$
Isaac Sim	-7.53	$[-13.52, -1.54]$

A.4.4 外观、部件和碰撞几何体

单个刚体可以包含多个语义部件。图 S5 对齐展示了两个资产的外观、部件分割和碰撞表示。在这些用于说明的基于部件的分解中，煎锅和羊角锤各包含四个标注部件，分别具有 92 个和 210 个凸包。定量实验采用整对象碰撞分解：语义部件用于组织材料和交互属性，而对象级凸组件则定义仿真接触。因此，该示例中的基于部件的凸包数量描述的是可视化示例，而不是定量评估所采用的碰撞构造。



图 S5 对齐的资产表示。 展示了煎锅和羊角锤的外观、原始部件标签、凸碰撞块和分配质量。部件颜色用于标识 P0–P3；碰撞颜色用于区分凸块。质量是导出的估计值。图示的基于部件的分解不同于定量实验中使用的整对象碰撞分解。

A.5 跨仿真器评估

A.5.1 跌落响应协议

我们在全部四个引擎中评估相同的 2,257 个资产。每个物体从 0.05、0.15 和 0.30 m 的高度，以五种初始方向进行评估。仿真运行持续 2.5 秒，步长为 0.01 秒。若水平位移不超过 max(0.05 m $0.1d$)、旋转变化不超过 $8°$，且支撑穿透深度不超过 0.01 m，则该有效试验符合跌落响应标准，其中 d 是物体尺寸。所有引擎均使用估算出的物理属性和多凸包碰撞几何。这些固定阈值定义了一项统一的有界运动与穿透测试：即使物体的运动在物理上合理，只要发生倾倒或移动超过阈值，就会被标记为负标签。采用相同的标准，可在此协议下比较各引擎的响应。

A.5.2 一致性指标

对于资产 i 和引擎 e，令 s_{ie} 表示通过全部三项跌落响应标准的有效试验占比，且$b_{ie} = \mathbb{1}[s_{ie} \geq 0.5]$。成对一致性比较 b_{ie}；Cohen's κ 用于校正边际标签频率。Fleiss'κ 用于概括四引擎一致性，而 Pearson 相关系数则使用连续值 s_{ie}。

所有引擎均使用相同的 2,257 个资产。所有引擎之间的一致性为 80.77%，Fleiss' $\kappa = 0.481$。正资产标签（$b_{ie} = 1$）在各引擎中占 7.75–13.03%。一致性同时计入正标签匹配和

负标签匹配。在全部四个引擎中，令 q_+ 和 q_- 分别表示获得正标签和负标签的比例。那么，$q_+ + q_- = 80.77\%$。由于 q_+ 不可能超过正标签比例最低的引擎（7.75%），因此 $q_- \geq 73.02\%$，允许存在四舍五入误差。因此，根据有界运动准则，大多数一致标签都是负标签。确切的共同正标签和共同负标签比例需要依据资产的联合标签；报告的边际比例则确定了这些界限。Fleiss' κ 通过考虑标签频率不均衡，对原始一致率作补充；连续得分相关性则量化各引擎对响应比例的排序是否相似。

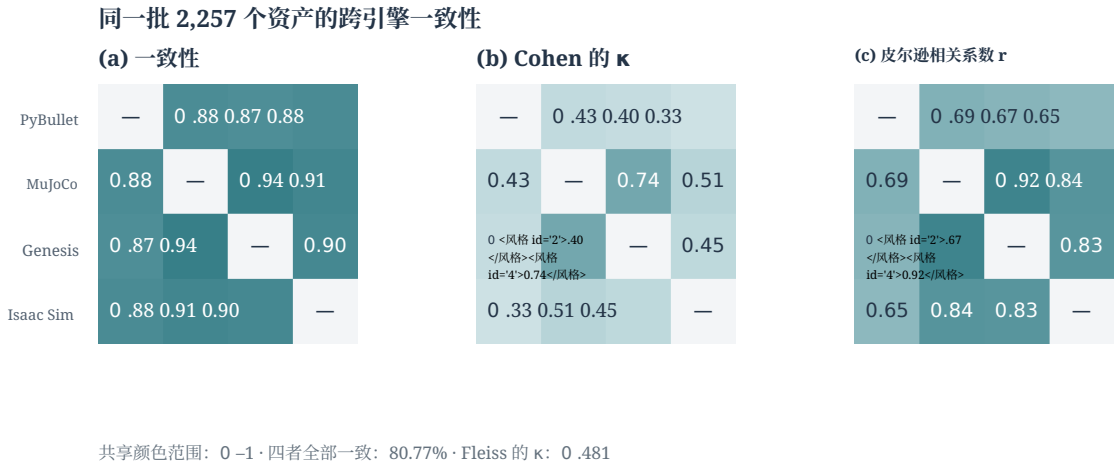

图 S6 四种物理引擎间的掉落响应一致性。两两一致性、Cohen's κ 以及判定通过率相关性均基于同一批 2,257 个资产。数值根据评估试验计算。

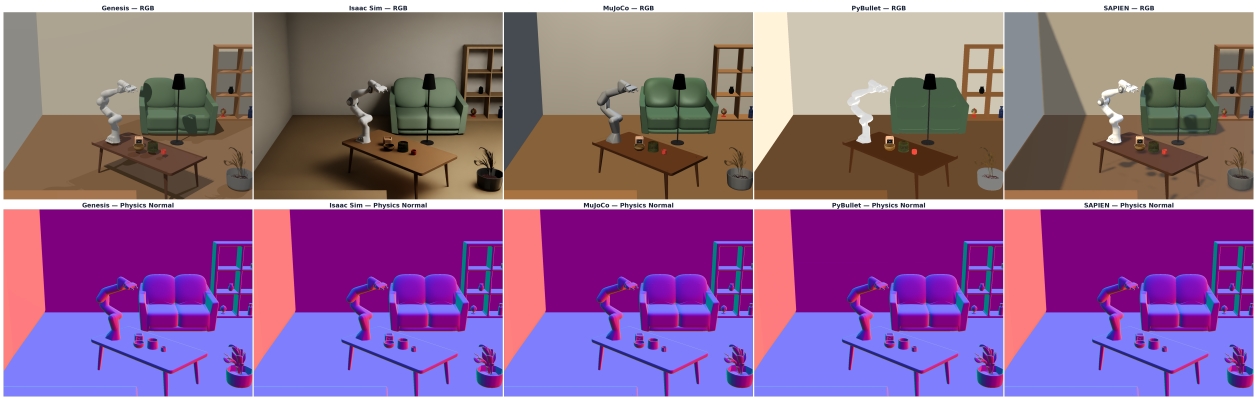

图 S7 跨模拟器场景渲染。同一组合场景通过共享资产接口在五种物理引擎中渲染。上：RGB 视图。下：表面法线视图。这些视图通过共享接口展示场景结构和依赖引擎的外观。任务—场景对应关系另见图 4；五引擎渲染展示是定性的接口演示，而行为一致性则通过上述跌落响应协议在四种引擎上单独评估。

A.6 灵巧抓取：协议与诊断

A.6.1 候选对象生成与执行

我们基于 T(R,O) Grasp (TRO-Grasp) [29], , 它可生成手部条件化的物体—机器人变换。对于每个目标资产，我们保留生成的姿态表示，并通过部分点云接口，根据选定的语义部件对候选池施加条件。候选项会转换到通用资产坐标系，与语义部件和真实网格进行核验，然后传入适用于每种训练好的手部模型的同一执行协议。每个物体–手部组合最多评估 64 个候选项。手部姿态由根节点平移/方向和关节配置构成。

Genesis 执行闭合、0.15 米指令抬升、保持以及六方向扰动，采用开环位置控制。仅提升成功要求物体上升超过 0.075 米。物体级任一成功在某物体至少有一个经评估的候选成功时，即将该物体计为成功。鲁棒记录结合了三种质量/摩擦/姿态设置。队列规模详见表 S11。

图 S8 展示了保存的执行示例及由此产生的闭合–抬升–保持序列。

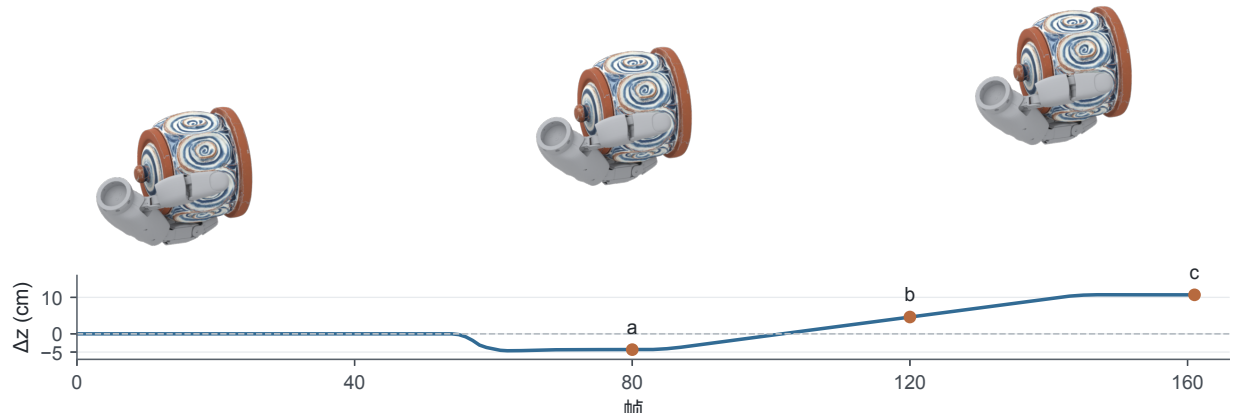

图 S8 示例抓取执行。XHand 在三个保存的姿态下闭合、抬升并保持一个调味品罐。标记 a–c 在垂直位移曲线上定位第 80、120 和 161 帧，这些位移均相对于第 0 帧测量。该序列展示了物体操作中的闭合、抬升和保持阶段。

A.6.2 候选预算

候选预算控制生成的抓取分布中有多少会被用于执行。图S9展示了记录的七种候选预算：随着预算增加，覆盖率不断提升，其中几种机械手在候选数从 16 增加到 64 之间获得了最大幅度的提升。表 S11 给出了完整候选预算下的比率以及单独的稳健队列。将上限从 16 增加到 64 后，每个物体–手部组合可执行的候选执行数量增加至四倍，从而展现了覆盖率–验证成本权衡。这些是候选搜索后的物体级任一成功率，因此候选预算属于评估协议的一部分。在完整候选预算下，Ability 和 Inspire 的提升覆盖率均超过 92%，而 Shadow 达到 42.3%，表明在共享的资产与执行流程下，不同机械手之间存在显著差异。

表 S11 完整候选预算下 Genesis 中的灵巧抓取结果。抬升率和鲁棒率使用不同的对象群组（N_L、 N_R）。方括号表示 Wilson 95% 置信区间；宏平均值涵盖全部六种机械手。抬升率与鲁棒率宏平均值之间的差异并非配对扰动效应，因为两者使用的对象群组不同。

Hand	N_L	抬升 [95% CI] ↑	N_R	鲁棒 ↑
Ability	2,372	92.5% [91.3, 93.4]	2,367	90.4%
Inspire	2,334	92.3% [91.1, 93.3]	2,353	90.4%
Allegro	2,372	91.3% [90.1, 92.4]	2,372	81.6%
Barrett	2,353	84.4% [82.8, 85.8]	2,372	81.5%
XHand	2,315	74.1% [72.3,75.8]	2,315	71.2%
Shadow	2,353	42.3% [40.3, 44.3]	2,372	27.3%
宏平均		79.5%		73.7%

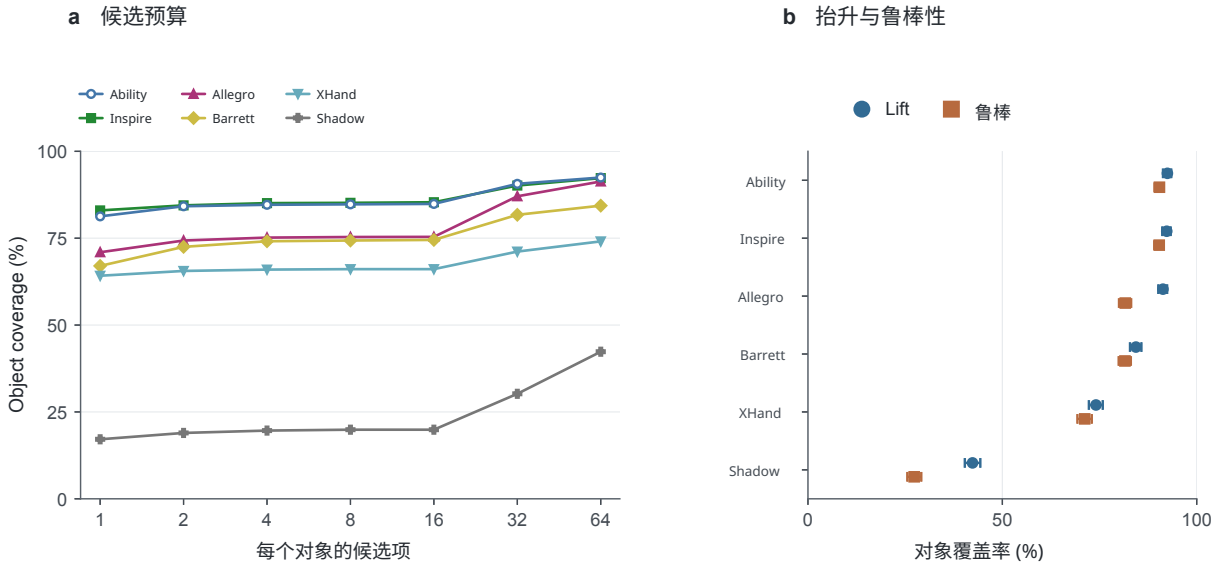

图 S9 候选预算与执行结果。左侧：七个记录预算下的提升覆盖率。右侧：全预算抬升和稳健覆盖率，并附 Wilson 95% 置信区间，使用表 S11中的不同队列。折线连接记录的观测值。

A.6.3 鲁棒性与闭合位移

鲁棒性在初始状态抬升后，通过结合保持以及对质量、摩擦和姿态的扰动进行评估。表 S12 展示了有多少稳健执行同时满足闭合平移界限，而表 S13报告了在单独的压力测试队列中，名义条件、更强扰动和组合扰动下的结果。在六种机械手中，10,432 次稳健成功中仅有五次未能满足额外的闭合平移界限。因此，对于已经通过鲁棒性测试的候选，闭合位移会排除的数量很少；这一条件计数无法确定抓取失败的原因。在单独的压力测试队列中，从名义条件到组合扰动的变化范围为 1.0 至 5.9 个百分点，其中 Allegro 的降幅最大（从 86.3% 降至 80.4%）。Ability 和 Inspire 的组合比率仍接近 90%，而 Shadow 在三种设置下均低于 41%。这些结果区分了基线抓取覆盖率与对所测试扰动的敏感性。

同一候选记录还会存储语义部件命中、穿透、接触接合、闭合位移和物理结果，使上述三个实验能够使用同一种资产侧标注格式。

表 S12 增加了闭合平移界限的三种设置下的稳健结果，该界限用于度量夹爪闭合期间的物体位移。

Hand	N	稳健成功次数 也设界限	也设界限（%）↑	
Ability	2,367	2,140	2,140	90.41
Inspire	2,353	2,126	2,125	90.31
Allegro	2,372	1,936	1,935	81.58
Barrett	2,372	1,934	1,933	81.49
XHand	2,315	1,648	1,648	71.19
Shadow	2,372	648	646	27.23

表 S13 独立压力测试队列：名义条件、更强扰动和组合压力率。 N 给出了每只手对应的评估对象数量。

Hand	N	名义条件（%）↑	更强（%）↑	合计（%）↑
Ability	2,372	91.6	91.1	90.6
Inspire	2,297	91.3	90.6	89.7
Allegro	2,372	86.3	80.7	80.4
Barrett	2,372	82.6	81.7	81.4
XHand	2,334	72.9	73.0	71.8
Shadow	2,353	40.8	40.0	38.7

A.7 策略数据、训练与评估

A.7.1 数据预算与资产选择

对于四个资产库中的每一个，我们都针对五项 RoboTwin 任务分别训练一个策略，每项任务使用 100 个成功示范。这共得到 20 个策略，总计使用 2,000 个示范数据，并且各资产库的示范数据预算相同。

匹配回合数可确保每个来源拥有相同的示范预算。由此得到的比较反映了已发布资产的覆盖范围、轨迹、类别匹配和准备方案如何共同支持策略学习。EmbodiedGen v2 为每项任务使用一个源资产，而评估中包含多种任务物体，因此，该比较衡量的是在固定示范预算下各已发布来源的下游效用。资产多样性和准备方案仍是此来源级比较的组成部分；若要将性能提升归因于某个单独的物理化模块，则需要进行受控消融实验。

A.7.2 训练和评估设置

训练。四个库均为每项任务使用 100 条演示数据，以微调 $\pi_{0.5}$ base 模型，训练随机种子为 0，全局批次大小为 32。对于每个来源–任务对，评估均使用第 19,999 步的最终检查点。每个来源–任务对均提供各自的训练数据和归一化统计数据，评估时使用最终检查点，不进行测试集选择。

仿真中的闭环评估。对 20 个策略中的每一个，均在 100 个干净条件专家可解场景和 100 个随机化条件专家可解场景中进行评估，共计 4,000 个回合。干净条件和随机化条件测试集独立采样。随机化会同时改变视觉上下文、桌面高度、光照和指令。任务成功依据 RoboTwin 原生的 RoboTwin 成功判定条件；在每个分母固定为 100 的统计中，超时、规划失败和基础设施错误均计为失败。每个库都有 10 个基础设施错误回合，且全部发生在随机化条件下。

评估采用冻结评估套件。所有库在每个任务–条件组合上均使用相同的 100 个随机种子进行评估。这 100 个评估回合用于表征每个固定的训练后检查点；这项单随机种子实验无法估计训练随机种子变异性。

A.7.3 聚合与任务级结果

宏平均值对五项任务赋予相同权重。两种模式下表现最强的基线都是 ManiTwin：干净条件下为 55.4%，随机化条件下为 31.2%；RFW-R 对应为 71.2% 和 41.8%。RFW-R 在全部五项干净条件任务和四项随机化任务上领先；在随机化条件下的“移动药瓶垫”任务中，PhysX-3D 达到 9%，RFW-R 为 6%。

RFW-R 在干净条件与随机化条件之间的宏平均差异为 29.4 个百分点。盘子和双碗的成功率分别保持在 76% 和 73%，而干净场景中的成功率分别为 96% 和 98%。A2B-L 和支架的成功率分别为 21% 和 33%；Pillbottle-Pad 则从 52% 降至 6%，任务级差异最大（46 个百分点）。对于所有来源而言，随机化条件下的 Pillbottle-Pad 都很难（0–9%），这表明存在共同的任务级泛化差距。由于干净场景和随机化场景是独立采样的，且多个因素同时发生变化，这些差异反映的是套件级敏感性；要确定外观、几何形状、指令还是规划的影响占主导，需要进行因素级评估或使用失败标签。

A.7.4 跨具身真实机器人轨迹回放

我们在 RoboTwin 仿真中使用 RFW-R 资产运行学习到的策略，并将生成的末端执行器轨迹重定向至物理 Piper 机器人，以进行跨具身回放。仿真机器人与物理机器人的具身不同；末端执行器对齐为在硬件上执行记录的运动提供了任务空间对应关系。策略推理完全在仿真中进行，真实机器人则执行重定向后的动作序列，过程中不接收来自真实世界观测的在线策略反馈。

图 S10 展示了五个任务族中的抓取、抬升和放置阶段。这些序列为跨具身运动执行提供了定性证据。垫子示例使用真实马克杯来说明放置任务族。表 5中的策略定量成功率则通过仿真中的闭环评估单独测量。

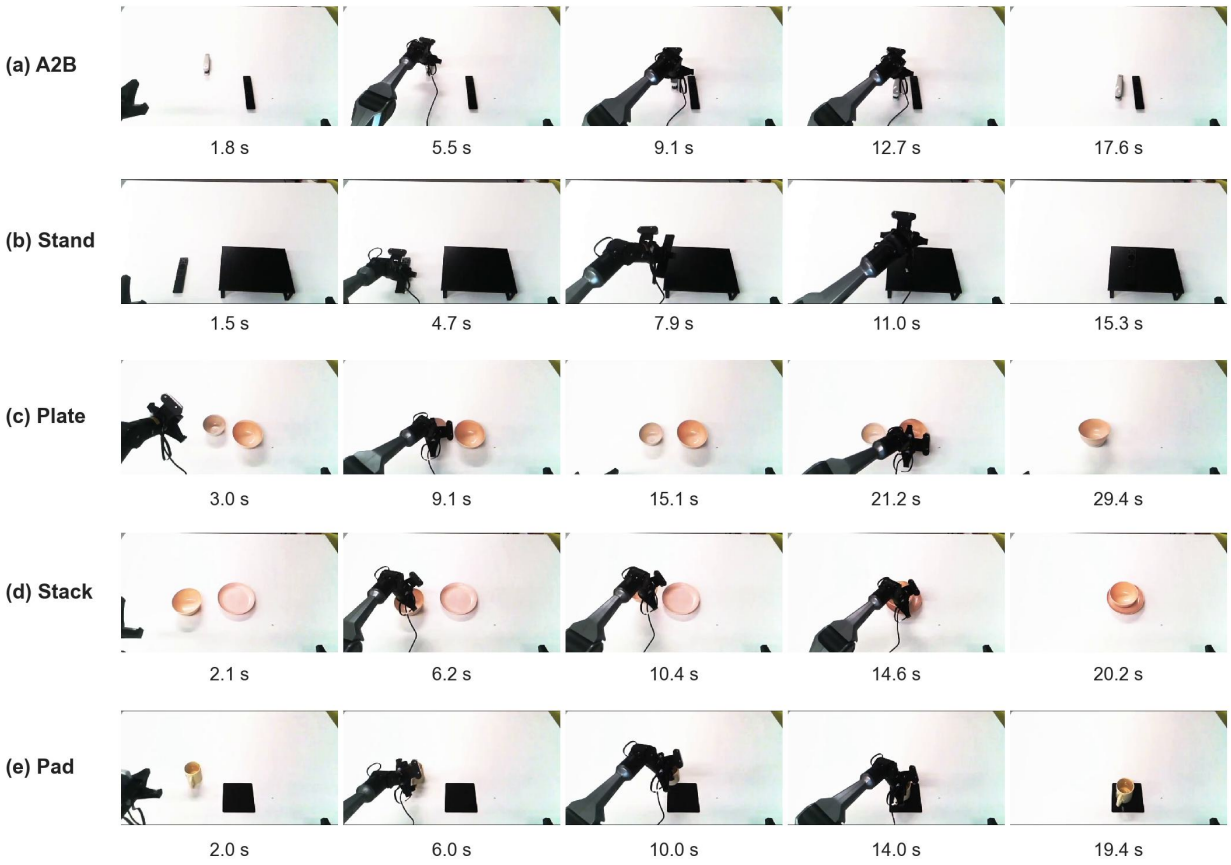

图 S10 跨具身真实机器人轨迹回放的时间序列。 实体 Piper 机器人执行从 RoboTwin 仿真中生成的策略轨迹重定向而来的末端执行器运动。各行依次展示：(a) 将 A 放在 B 左侧，(b) 放置在支架上，(c) 放在盘子上，(d) 堆叠碗，以及 (e) 放置在垫子上。每行包含其录制回放视频中 10%、30%、50%、70% 和 97% 时刻的帧。垫子示例使用了真实马克杯，展示了放置任务族。

A.8 未来工作

未来工作将聚焦于三个方向。首先，我们计划将部件级可供性和抓取标注扩展到可动对象生成与重建，并以可动对象重建和生成领域的近期进展为基础 [10, 35, 48, 49, 64, 65, 68–70, 89, 91, 92, 99, 104, 111, 112, 124, 130, 131]。其次，我们计划将智能体标注和场景生成扩展到三维编辑、自动绑定以及动态三维/四维世界生成，从而为具身人工智能和自动驾驶实现可扩展的代码到视频数据合成 [1, 11–13, 15, 44, 47, 55, 58, 63, 66, 88, 103, 110, 118, 119, 127, 133, 135, 137, 138]。最后，我们设想通过整合多样化的物理资产、交互式场景、机器人演示和多具身任务 [16, 17, 22, 28, 38, 39, 46, 51, 62, 73, 75, 97, 129, 136, 137]，构建更全面的具身基准和数据引擎。这样便能系统评估操作能力、物理一致性、鲁棒性以及仿真到现实泛化能力。