

Octo:一款开源通用机器人策略

Octo 模型团队

Dibya Ghosh*,1 Homer Walke*,1 Karl Pertsch*,1,2 Kevin Black*,1 Oier Mees*,1

Sudeep Dasari3 Joey Hejna2 Tobias Kreiman1 Ria Doshi1 Charles Xu1 Jianlan Luo1 You Liang Tan1 Lawrence Yunliang Chen1
Pannag Sanketi4 Quan Vuong4 Ted Xiao4 Dorsa Sadigh2 Chelsea Finn2 Sergey Levine1 1加州大学伯克利分校 2斯坦福大学 3卡内基梅隆大学4谷歌

DeepMind <https://octo-models.github.io>

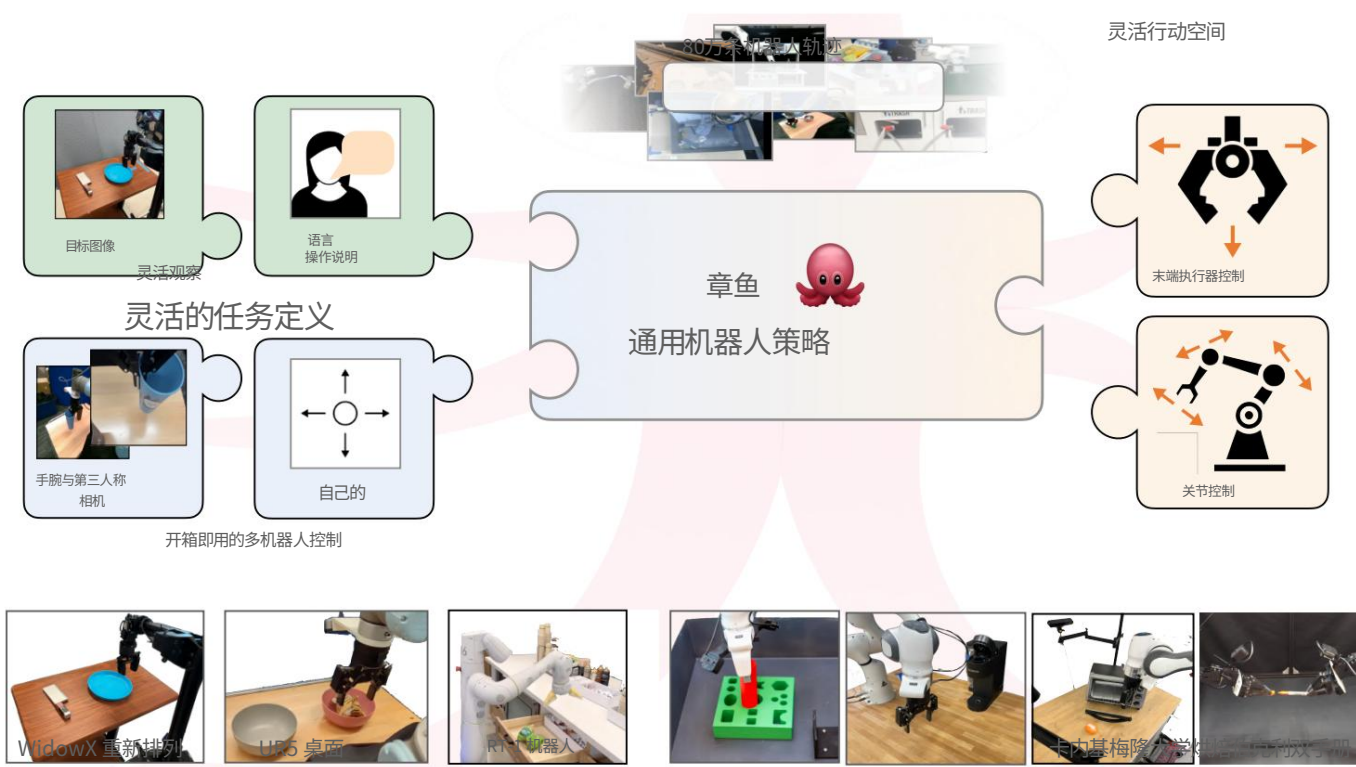


图 1:我们介绍 Octo,一种用于机器人操作的开源通用策略。Octo 是一种基于 Transformer 的策略,已在 Open X-Embodiment 数据集[67] 中 80 万个不同的机器人动作场景上进行了预训练。它支持灵活的任务和观察定义,并且可以快速微调以适应新的观察和动作空间。

摘要 在多样化的机器人数据集上预训练的大型策略有望彻底改变机器人学习:无需从头开始训练新策略,这类通用机器人策略只需少量领域内数据即可进行微调,并具有广泛的泛化能力。然而,为了使其能够广泛适用于各种情况,

利用新的观察和行动空间进行高效微调

我们在9个机器人平台上进行了实验,证明Octo是一种通用的策略初始化方法,可以有效地微调以适应新的观测和动作空间。我们还对Octo模型的设计决策进行了详细的消融分析,从架构到训练数据,旨在指导未来构建通用机器人模型的研究。

针对机器人学习场景、环境和任务,此类策略需要处理各种传感器和动作空间,兼容各种常用机器人平台,并能快速高效地微调以适应新的领域。本文旨在阐述……

本文旨在为开发开源、广泛适用的通用机器人操作策略奠定基础。作为第一步,我们推出了 Octo,这是一个基于 Transformer 的大型策略,它基于 Open X-Embodiment 数据集 (迄今为止最大的机器人操作数据集)中的 80 万条轨迹进行训练。Octo 可以通过语言命令或目标图像进行指令控制,并且可以在标准消费级 GPU 上于数小时内有效地微调,以适应具有新的传感器输入和动作空间的机器人设置。

伯克利插入斯坦福咖啡信

机器人学习的常用方法是利用为特定机器人和当前任务收集的数据集来训练策略。从零开始学习这种方法需要为每个任务投入大量数据收集工作,而且最终得到的策略通常泛化能力较弱。原则上,收集到的数据

*主要作者按字母顺序排列,请参阅A 部分了解贡献列表。
通讯作者: {dibya.ghosh, homer_walke, pertsch, kvaback, oier.mees}@berkeley.edu

来自其他机器人和任务的经验提供了一种可能的解决方案,使模型能够接触到各种各样的机器人控制问题,从而提高模型的泛化能力和下游任务的性能。然而,尽管通用模型在自然语言处理[68, 88]和计算机视觉[76, 44]领域已变得无处不在,但构建能够控制众多机器人执行多种任务的类似“通用机器人模型”仍然极具挑战性。在机器人领域训练统一的控制策略面临着独特的挑战,需要处理不同的机器人形态、传感器配置、动作空间、任务规范、环境和计算资源。

为此,一些研究提出了机器人基础模型,这些模型直接将机器人观测结果映射到动作,并能以零样本或少样本的方式泛化到新的领域和机器人。我们通常将这些模型称为“通用机器人策略”(GRP),强调它们能够在不同的任务、环境和机器人系统中执行低级视觉运动控制[75, 9, 23, 103, 10, 81, 1, 91, 35, 94, 45]。

例如,GNM模型[80]能够泛化到不同的机器人导航场景,RoboCat模型[9]能够处理不同机器人形态的目标导向任务,而RT-X模型[67]则能够在五种机器人形态上执行语言导向的操作。尽管这些模型代表着向真正的“通用机器人模型”迈出的重要一步,但它们在多个重要方面仍存在局限性:它们通常将下游用户限制在预定义且往往具有限制性的输入观测集内,例如,单个摄像头数据流;它们缺乏对新领域进行有效微调的支持;更重要的是,这些模型中规模最大的那些尚未向公众开放。

我们设计了一个系统,用于预训练更适合下游机器人应用中各种接口的通用机器人策略。我们模型的核心是一个Transformer架构,它将任意输入标记(由观测数据和任务数据生成)映射到输出标记(然后解码为动作),该架构可以在包含各种机器人和任务的数据集上进行训练。无需额外训练,该策略即可接受不同的摄像头配置(例如,工作区摄像头或腕部摄像头),控制不同的机器人,并可通过语言命令或目标图像进行引导。所有这些都只需简单地改变输入模型的标记即可实现。最重要的是,通过添加合适的适配器并使用小型目标域数据集和可接受的计算资源进行微调,该模型可以适应具有新传感器输入、动作空间或形态的新机器人设置。

我们的主要贡献是Octo,这是一个基于Transformer的策略,它在迄今为止最大的机器人操作数据集上进行了预训练:来自Open X-Embodiment数据集的80万个机器人演示[67]。Octo是第一个可以有效微调以适应新的观测和动作空间的通用机器人操作策略(GRP),也是第一个完全开源的通用机器人操作策略,包括训练流程、模型检查点和数据。

最后,虽然构成Octo的各个组件Transformer主干、对语言和目标图像规范的支持,以及用于模拟表达性动作分布的扩散头已经在之前的工作中讨论过,但

将这些组件以特定方式组合成强大的通用机器人策略是独特而新颖的。

我们通过在四个机构的九台机器人上进行的大量实验证明,我们的组合系统在单臂和双臂操作任务中实现了最先进的开箱即用型多机器人控制性能,并且Octo可以作为有效的初始化模型,用于在新的观测和动作空间中对未见过的设置进行微调。在此过程中,我们仔细研究了预训练GRP时不同设计决策的影响;我们评估了数据分布、模型架构和策略制定方式的选择如何影响预训练GRP的质量。我们的评估突显了规模和灵活性的重要性:我们最好的模型是在最广泛的数据混合上训练的,具有限制性最小的归纳偏置,并且其策略目标能够适应预训练数据中行为的多样性。

本文发布的同时,我们也提供了训练、使用、复现和微调Octo模型所需的所有资源。我们提供了预训练的Octo模型检查点,分别包含2700万和9300万个参数,开箱即用,支持多种RGB摄像头输入,以及语言和目标图像任务规范。此外,我们还提供了用于在新领域微调这些模型的脚本,以及完整的预训练流程,包括优化的数据加载器、用于多模态输入的Transformer实现,以及用于监控训练进度的工具。

二、相关工作

许多研究工作都利用从机器人收集的大型轨迹数据集来训练策略,从早期利用自主数据收集来扩展策略训练规模的研究[71, 48, 41, 19, 27, 30]到近期探索将基于Transformer的现代策略与大型演示数据集相结合的研究[10, 40, 98, 28, 83, 86]。这些研究主要集中于单一机器人模型,而Octo则利用跨多个机器人模型构建的数据集来训练策略,从而有效扩大训练数据集的规模,并允许针对各种机器人配置进行微调。

最近,论文的重点是拓宽机器人策略的泛化能力。许多工作利用各种非机器人数据或预训练的视觉语言基础模型来增强策略对新场景和任务的泛化能力[86, 103, 96, 16, 38, 11, 84, 36, 4, 37, 7, 3, 46, 15, 23]。

与Octo更为密切相关的是近期一些利用来自多种机器人形态的数据训练机器人策略的研究:GNM模型[81, 80]能够泛化到不同的机器人导航设置,而RoboCat[9]和RT-X[67]则能够控制多个单臂操作机器人。尽管这些模型取得了令人瞩目的策略学习成果,但一个关键问题是它们缺乏灵活性:它们通常要求用户坚持使用预训练期间的传感器输入和动作空间,而不支持适应新的观察和动作空间。此外,规模最大的模型尚未公开。Octo与这些研究在多个方面有所不同:它使用更大、更多样化的机器人数据集进行训练,通过高效的微调来适应新的机器人设置,从而支持更广泛的下游应用,并且它是完全开源且可复现的。

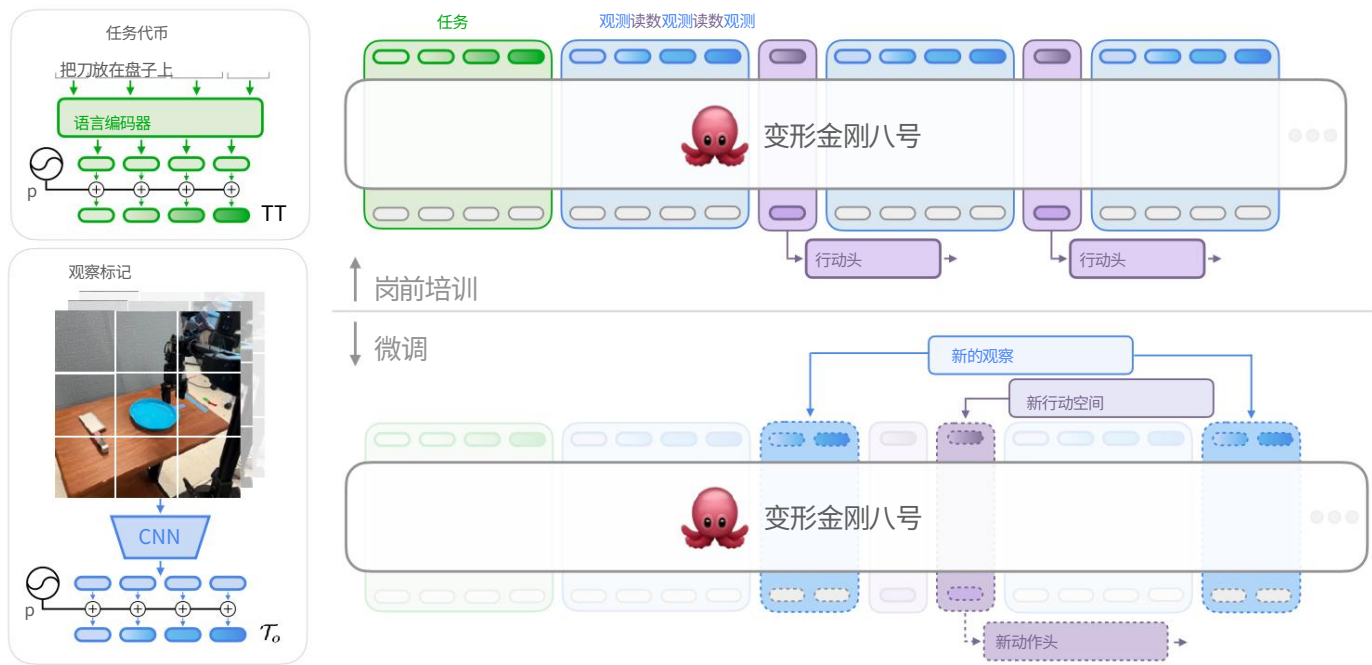


图 2:模型架构。左图： Octo分别使用预训练语言模型和轻量级 CNN 对任务描述（绿色）和输入观测值（蓝色）进行分词。上图： Transformer 主干网络处理任务和观测值序列,并生成输出标记（紫色）,这些输出标记传递给输出头以生成动作。下图： Transformer 的分块注意力结构允许我们在微调期间添加或移除输入和输出:例如,我们可以添加新的观测值（蓝色虚线）或动作空间（紫色虚线）,而无需修改任何预训练参数。

Octo 的设计灵感来源于机器人模仿学习和可扩展 Transformer 训练领域的最新进展,包括使用去噪扩散目标[34]进行动作解码[17, 31, 85]、预测“动作块”(即未来动作序列) [98, 17, 28],以及借鉴可扩展视觉 Transformer 训练文献[22, 97] 的模型布局和学习率调度。我们的工作首次将这些方法应用于跨具身通用策略的学习,并发现它们能够显著提升性能。在评估过程中,我们展示了消融实验来评估这些组件的重要性,并在附录 E 中提供了更全面的重要性（或不重要性）因素列表;我们希望我们的发现能够对未来通用策略学习的研究有所助益。

训练通用机器人策略的关键要素是机器人训练数据。与可以从网络抓取的视觉和语言数据不同,大规模获取机器人数据极具挑战性,通常需要大量的硬件和人力投入。目前已存在多个大型机器人导航和自动驾驶数据集[29, 95, 13, 87, 80, 43, 89]。近年来,人们也致力于构建规模和多样性日益增长的机器人操作数据集,这些数据集既可以通过脚本和自主策略收集[19, 41, 42, 12, 71, 30],也可以通过人类远程操作收集[59, 60, 25, 90, 39, 10, 26, 6, 77, 63, 79]。Octo 使用 Open X-Embodiment 数据集[67]进行训练,该数据集汇集了上述多个机器人数据集。Open-X 数据集包含约 150 万个机器人操作片段,其中

我们精选了 80 万个训练集用于 Octo 的训练。我们注意到 RT-X 模型[67]使用的是更为精简的 35 万个训练集子集,因此据我们所知,Octo 是迄今为止在最大的机器人操作演示数据集上训练的。

三、八爪鱼模型

本节将介绍 Octo 模型,这是一个开源的通用机器人策略,可通过微调适应新的机器人和任务,包括新的传感器输入和动作空间。我们将讨论关键的设计决策、训练目标、训练数据集和基础设施。Octo 模型的设计强调灵活性和可扩展性:它支持各种常用机器人、传感器配置和动作,同时提供了一个通用且可扩展的方案,可以基于大量数据进行训练。它还支持自然语言指令、目标图像、观测历史以及通过扩散解码进行的多模态分块动作预测[17]。此外,我们专门设计了 Octo 模型,使其能够高效地微调以适应新的机器人设置,包括具有不同动作空间以及不同摄像头和本体感觉信息组合的机器人。这种设计旨在使 Octo 成为一个灵活且应用广泛的通用机器人策略,可用于各种下游机器人应用和研究项目。

A. 建筑学

Octo 的核心是一个基于 Transformer 的策略 π 。它由三个关键部分组成:输入分词器,用于转换输入和输出。

语言指令 ℓ 、目标 g 和观察序列 o_H 被转换为词元 $T\ell$ 、 Tg 、 To (图2,左);一个 o_1 , ..., Transformer主干网络处理这些词元并生成嵌入,例如 $e_o = T(T\ell, Tg, To)$ (图2,上);以及读出头。其他产生所需输出的 $R(e)$,即行动 a 。

任务和观察标记器:我们使用特定模态标记器将任务定义(例如,语言指令 ℓ 和目标图像 g)和观察 o (例如,手腕和第三人称摄像机流)转换为通用的“标记化”格式(参见图2,左):

语言输入被分词,然后通过预训练的Transformer模型生成语言嵌入词元序列。我们使用5-base (111M)模型[74]。图像观测和目标经过浅层卷积层处理,然后被分割成一系列扁平化的图像块[22]。

我们通过将可学习的位置嵌入 p 添加到任务和观察标记,然后按顺序排列它们 $TT, To, 1, To, 2, \dots$ 来组装 transformer 的输入序列。

Transformer 主干和读出头:输入被转换为统一的标记序列后,由 Transformer 进行处理(见图2,顶部)。这与先前在观测和动作序列上训练基于 Transformer 的策略的工作类似[92, 73]。Octo Transformer 的注意力模式是分块掩码的:观测标记只能因果地关注来自相同或更早时间步 $To, 0:t$ 的标记以及任务标记 TT (绿色)。

对应于不存在的观测值的标记会被完全屏蔽(例如,没有语言指令的数据集)。这种模块化设计使我们能够在微调过程中添加和移除观测值或任务(见下文)。除了这些输入标记块之外,我们还插入了学习到的读取标记 TR, t (紫色)。位于 TR, t 的读取标记会关注序列中位于其之前的观测值和任务标记,但不会被任何观测值或任务标记关注。因此,它们只能被动地读取和处理内部嵌入,而不会对其产生影响。

读取标记的作用类似于BERT中的 $[CLS]$ 标记,它们作为迄今为止观测序列的紧凑向量嵌入。一个轻量级的“动作头”实现了扩散过程,应用于读取标记的嵌入向量。该动作头预测一个包含多个连续动作的“块”,类似于之前的研究[98, 17]。

我们的设计允许我们在下游微调过程中灵活地向模型添加新的任务和观测输入或动作输出头。当向下游添加新的任务、观测或损失函数时,我们可以完全保留Transformer的预训练权重,仅根据规范的变更添加新的位置嵌入、新的轻量级编码器或新头的参数(参见图2底部)。这与之前的架构[10, 81]形成鲜明对比,在之前的架构中,添加或移除图像输入或更改任务规范需要重新初始化或重新训练预训练模型的大部分组件。

这种灵活性对于使 Octo 成为真正的“通用”模型至关重要:因为我们无法涵盖所有可能的机器人传感器和

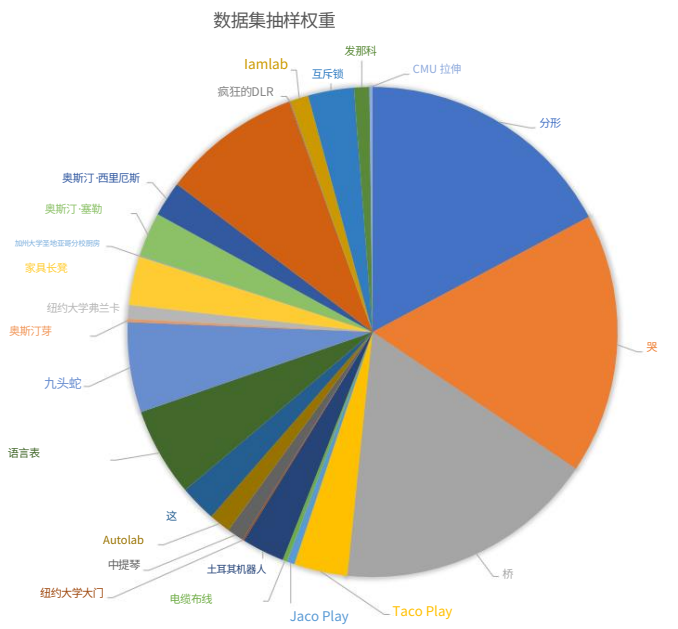


图 3:训练数据集构成。我们从 Open X-Embodiment 数据集中精选了 25 个数据集子集,这些数据集包含图像观测数据、末端执行器动作,并展现出多样化的行为。饼图可视化了每个数据集平均在每个训练批次中所占的比例。数据集权重由每个数据集中的样本数量决定,并进行少量调整以平衡数据集的大小和多样性(详见第三部分 B 节)。

Octo在预训练期间能够配置动作,并在微调期间调整其输入输出,这使其成为机器人领域一款用途广泛的工具。以往使用标准Transformer骨干网络或将视觉编码器与MLP输出头融合的设计,会锁定模型预期输入的类型和顺序。相比之下, Octo切换观测或任务无需重新初始化模型的大部分内容。

B. 训练数据

我们使用来自 Open X-Embodiment 数据集[67]的 25 个数据集的混合数据集来训练 Octo,该数据集是一个多样化的机器人学习数据集集合。我们的训练混合数据集包括来自多个机器人形态和场景的各种任务的演示数据。

这些数据集不仅在机器人类型上存在异质性,而且在传感器(例如,是否包含腕部摄像头)和标签(例如,是否包含语言指令)方面也存在异质性。有关详细的混合模型,请参见图 3和附录C。为了创建我们的训练混合模型D,我们首先对数据进行筛选,移除所有不包含图像流的 Open-X 数据集,以及未使用 Delta 末端执行器控制的数据集。我们还移除了重复性过高、图像分辨率过低或任务过于小众的数据集。对于剩余的数据集,我们根据任务和环境将其大致分为“更多样化”和“多样性较低”两类,然后在训练过程中将多样性较高的数据集的权重加倍。此外,我们还降低了一些重复性较高的数据集的权重,以避免其在训练中占据主导地位。

最后,我们对缺失的摄像头通道进行零填充,并对齐数据集之间的夹爪动作空间,使得夹爪指令 +1 表示“夹爪打开”, 0 表示“夹爪关闭”。虽然我们发现由此得到的训练混合数据效果良好,但未来的工作应该对数据混合质量进行更深入的分析,以用于预训练通用机器人策略。

C. 训练目标

我们使用条件扩散解码头来预测连续的多模态动作分布[34, 17]。重要的是,每次动作预测仅需执行一次Transformer主干网络的前向传播,之后多步去噪过程完全在小型扩散头内完成。我们发现,在零样本评估和微调评估中,这种策略参数化方法均优于使用均方误差 (MSE)动作头或离散化动作分布[10]训练的策略。

为了生成动作,我们对高斯噪声向量 $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 进行采样,并用 \mathbf{x}^k 用K步去噪,去噪参数为学习得到的 $\theta(\mathbf{x}^k, \mathbf{e}, k)$, 该参数以先前去噪步骤的去噪网络 $\theta(\mathbf{x}^k)$ 输出、步骤索引k和为条件。

Transformer 动作读出的输出嵌入 \mathbf{e} :

$$\mathbf{x}^{k-1} = \alpha(\mathbf{x}^k - \gamma \theta(\mathbf{x}^k, \mathbf{e}, k) + \mathbf{N}(0, \sigma^2 \mathbf{I})). \tag{1}$$

超参数 α 、 γ 和 σ 对应于噪声调度:我们使用文献[66]中的标准余弦调度。我们使用文献 [34]中首次提出的标准 DDPM 目标函数训练扩散头,其中我们向数据集动作添加高斯噪声,并训练去噪网络 $\theta(\mathbf{x})$ 以重建原始动作。有关扩散策略训练的详细说明,请参见 Chi 等人[17]。所有超参数列于附录D中。

$$\theta(\mathbf{x}^k, \mathbf{e}, k)$$

我们在微调过程中使用相同的扩散训练目标,并更新整个模型,这种方法优于冻结预训练参数子集的方法。在所有微调实验中,我们都采用相同的方法:给定一个包含约 100 条轨迹的小型目标域数据集,我们使用余弦衰减学习率和线性预热进行 5 万步微调。

D. 培训详情

我们训练了模型的两个变体:Octo-Small,其 Transformer 主干网络与 ViT-S 的大小相仿; Octo-Base,其 Transformer 主干网络与ViT-B 的大小相仿[22]。

我们使用 AdamW 优化器[51],采用反平方根衰减学习率策略[97],权重衰减值为 0.1,梯度裁剪值为 1.0。ViT-B 模型在 TPU v4-128 pod 上以2048的批大小训练了30万步,耗时 14 小时。在配备 24GB 显存的单个 NVIDIA A5000 GPU 上对同一模型进行微调大约需要 5 小时,并且可以通过多 GPU训练来加速。

我们使用两帧观测历史进行训练;在初步实验中,我们发现超过第一帧后,收益显著递减。我们使用后见之明目标

重新标记[2]从轨迹的未来状态中均匀地选择一个状态作为目标图像,类似于之前的研究[54, 90, 81, 77, 63]。我们在训练过程中应用常见的图像数据增强方法,并随机将每个训练样本的语言指令或目标图像置零,以使Octo能够基于语言指令或目标图像进行条件化。对于没有语言标注的数据集,我们始终使用目标图像进行条件化。这使得我们的模型能够通过自监督的视觉观察来学习控制,并减轻了语言标注的负担,类似于之前关于多上下文模仿学习的研究[54, 62, 61, 63]。有关超参数选择的更多详细信息,请参见附录D。

E. 模型检查点和代码

我们将训练、微调和所需的所有资源开源。运行我们的模型 (参见 <https://octo-models.github.io>) :

- 针对 Octo-Small (2700 万参数)和 Octo-Base (9300 万参数)的预训练 Octo 模型检查点。
- 用于 Octo 模型的微调脚本 (JAX 编写)。
- 用于在 Open X-Embodiment 数据集上进行 Octo 模型预训练的 JAX模型预训练流程。
- 兼容 JAX 和 PyTorch 的独立Open X-Embodiment 数据加载器。

我们在附录B中提供了加载和运行预训练 Octo 模型的简单示例。

四、实验

我们的实验对 Octo 进行了实证分析,评估了它作为通用机器人基础模型在多个维度上的能力:

- 1) Octo 能否控制多个机器人实例,并开箱即用地解决语言和目标任务?
- 2) Octo 权重是否能作为数据高效的微调以适应新任务和机器人的良好初始化,并且它是否比从头开始训练和常用的预训练表示有所改进?
- 3)在 Octo 中,哪些设计决策对于构建通用机器人策略最为重要?

评估设置:我们在4个机构的9种具有代表性的机器人学习设置中评估了Octo的性能 (见图 4)。我们测试了Octo在语言和目标图像任务中对不同机器人进行开箱即用 (“零样本”)控制的能力,使用的机器人设置与预训练数据相匹配,所有机器人均采用增量末端执行器控制动作,观测空间为RGB图像。我们还评估了Octo在新环境和任务中高效微调的能力,包括新的观测数据 (“伯克利插入”中的力矩输入)、新的动作空间 (“伯克利拾取”中的关节位置控制)以及新的机器人形态 (“伯克利可乐”和 “伯克利双手动”)。每种微调设置均使用约100个领域内演示,并在NVIDIA A5000 GPU上于5小时内完成微调,所有设置均使用相同的超参数 (见附录D)。我们的评估任务旨在测试 Octo 与各种对象 (例如 “WidowX BridgeV2”)交互以及解决长期任务的能力。



图 4:评估任务。我们在 4 个机构的 9 个真实机器人设置上对 Octo 进行了评估。我们的评估涵盖了各种物体交互（例如，“WidowX BridgeV2”）、较长的任务周期（例如，“斯坦福咖啡”）和精确操作（例如，“伯克利插钉”）。我们评估了 Octo 在开箱即用的情况下,利用预训练数据控制机器人的能力,以及利用小型目标数据集高效地微调以适应新任务和环境的的能力。我们还测试了使用新观测值（“伯克利插钉”的力矩输入）、动作空间（“伯克利拾取”和“伯克利双手动”中的关节位置控制)和新的机器人形态（例如，“伯克利双手动”和“伯克利可乐”）进行微调的能力。

（例如，“斯坦福咖啡”）并进行精确操作（例如，“伯克利插入”）。有关每种评估设置的更多详细信息,请参见附录F。

比较:我们使用已发布的检查点,将 Octo 开箱即用的多机器人控制能力与目前公开可用的最佳通用机器人策略RT-1-X [67]进行比较。与 Octo 类似,RT-1-X 也基于 Open X-Embodiment 机器人数据集进行预训练,旨在零样本控制多机器人,因此提供了一个天然的比较基准。

我们还将 Octo 的零样本能力与RT-2-X 进行了比较, RT-2-X是一个拥有 550 亿个参数的视觉语言模型,在 Open X-Embodiment 数据集上进行了微调,以生成机器人动作。

RT-1-X 和 RT-2-X 模型[67]的训练数据集更为精简,仅为 35 万个回合（相比之下,Octo 模型使用了 80 万个回合）。我们进一步将 Octo 作为数据高效微调策略初始化方法的性能与两种常用方法进行比较:（1）从头开始在目标演示数据集上进行训练;（2）使用预训练的视觉表征。

尽管一些先前的研究提出了其他用于模仿微调的预训练方案[25, 24, 26],但据我们所知,目前尚无任何方法能够提供已被证明可以成功微调到新的观察和动作空间的预训练策略。然而,诸如VC-1 [56]之类的预训练视觉表征已被用于此目的,因此我们将这些方法作为另一个比较点。

为了进行微调,我们发现从头开始训练我们的大型 Transformer 架构,在小数据集上会很快过拟合。

相反,我们使用一种被许多先前工作采用的典型策略架构,从零开始获得了更好的结果:一个带有 FiLM [70]语言条件的 ResNet 视觉编码器,结合一个使用扩散目标训练的小型 Transformer 动作解码器,类似于[10, 98, 17, 55]。我们实现的该架构具有 2800 万个参数（类似于RT-1 [10]）。我们将其作为从零开始的基线（“ResNet+Transformer Scratch”）。我们还将其与预训练的视觉表示进行了比较,具体步骤如下:

Majumdar等人[57]提出了一种基于ViT-B视觉编码器的模型,该编码器初始化为VC-1权重[57],VC-1是一种最先进的视觉表征模型,已在4000小时的以自我为中心的视频和ImageNet数据集上进行了预训练。该模型与一个多层感知器（MLP）动作解码器相结合,并使用均方误差（MSE）损失（“VC-1”）来预测专家的动作。

A. Octo 开箱即可控制多个机器人

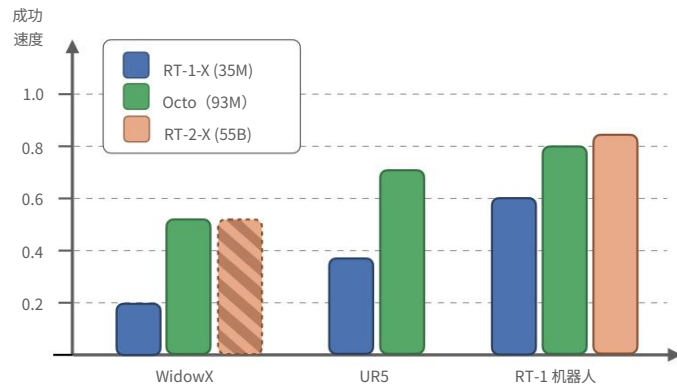


图 5:零样本评估。Octo开箱即用,能够根据预训练数据在各种环境中控制多个机器人。

在使用自然语言指定任务时,Octo在三种不同的机器人实现和设置中均优于 RT-1-X [67], RT-1-X是目前公开可用的最佳通用机器人策略。

在测试的WidowX 和 RT-1 Robot 任务上,Octo 的性能也与 RT-2-X [103]类似。1

我们在图5 中比较了 Octo、RT-1-X 和 RT-2-X的零样本操作能力。我们使用从预训练数据集中选择的几个任务进行了评估,包括拾取和

1对于 WidowX,由于 RT-2-X 未公开提供,我们报告了[7]中的 RT-2-X 数据（虚线条）,并使用与 Octo 和 RT-1-X 评估相同的任务。对于 RT-1 机器人,RT-2-X 的作者慷慨地为我们进行了评估。

	伯克利插入*	斯坦福咖啡 CMU烘焙	伯克利自取†	伯克利可乐	伯克利双语版†	平均
ResNet+Transformer Scratch	10%	45%	25%	0%	20%	20%
VC-1 [57]	5%	0%	30%	0%	10%	15%
Octo (我们的)	70%	75%	50%	60%	100%	72%

表一:微调评估。Octo能够以数据高效的方式微调模型以适应新的领域,并且性能优于训练模型。Scratch 以及最先进的预训练视觉表征。每个领域使用约 100 个目标演示,并且相同。微调超参数。在每个域中,成功率取 20 次试验的平均值。*:新的观测输入(力-力矩)本体感觉)。†:新的动作空间(关节位置控制)。

摆放物品、用抹布擦拭桌子、打开和关闭我们从抽屉中为每个机器人选择了两个语言任务。使用相应的 OXE 数据集,并对每个任务进行 10 次试验。初始条件各不相同(详情见附录F)。所选任务与预训练数据“分布一致”,但评估需要能够推广到新对象的方法位置、光照条件、背景和干扰因素对象。虽然所有方法在各项任务中都表现得相当合理,但在预训练环境中,我们发现 Octo 平均有比 RT-1-X (3500 万个参数)的成功率高 29%。在对 WidowX 和 RT-1 机器人进行评估时,我们也进行了比较。到 RT-2-X (550 亿个参数) [103]并发现 Octo 表现类似。

此外,RT-1-X 和 RT-2-X 仅支持基于语言指令的条件反射,而 Octo 还支持基于目标图像的条件反射。我们在 WidowX 上评估了我们的模型。使用目标图像条件反射的任务,发现它达到了与使用语言评估相比,成功率高出 25%。条件反射。这可能是因为目标图像提供了更多关于如何完成任务的信息。在 BridgeV2 中在该领域,我们对零样本进行了精细分析。表七中的功能;测量所见设置的性能在数据集中,以及针对新环境、场景和技能。虽然 Octo 模型在新奇物体上取得了很高的成功率,在新的场景中,零帧性能略有下降,而且对于翻转或精确等新颖行为,性能会大幅下降。插入。

B. Octo 助力在新领域实现数据高效学习

我们报告了针对新领域的高效数据微调结果如表一所示,我们发现对 Octo 进行微调可以带来更好的结果。比从零开始或使用预训练的VC-1更合适的策略权重。在六种评估设置中,平均值(详细信息如下)在附录F中),Octo 的性能优于次优基准模型。52%。重要的是,我们使用了相同的配方和超参数。为了在所有评估任务上微调 Octo (参见第三节 C 部分),这是一个不错的默认配置。

结果还突显了 Octo 的容纳能力新的观测结果 (“伯克利插入”的力矩输入),动作空间 (“伯克利拾取”的关节位置控制)以及新的机器人形态 (“伯克利可乐”和 “伯克利”)。双手操作)。这使得 Octo 适用于广泛的应用场景。单臂和双臂机器人操作问题除了单个摄像头输入和末端执行器位置控制之外。

	总体表现
Octo-Small (我们的)	83%
RT-X 数据集混合[67]	60%
单机器人数据集 (桥梁数据)	43%
离散化动作预测[67]	18%
连续动作预测 (MSE)	35%
Resnet-50 + Transformer[67]	70%

表二:模型消融。我们取得了最佳性能当使用 ViT 架构、扩散作用头时,广泛的训练数据混合。所有评估均在以下数据集上进行: WidowX 设置。成功率是根据 40 次试验的平均值计算得出的。两项语言条件反射任务和两项目标条件反射任务。

C. 通用机器人策略训练的设计决策

我们已证明 Octo 作为零样本多机器人控制器和策略初始化的有效性。

微调。接下来,我们将分析不同设计方案的效果。关于 Octo 政策执行情况的决策。具体而言,我们重点关注以下几个方面: (1)模型架构,(2)训练数据,(3)训练目标,以及 (4)模型规模。除非另有说明,由于计算预算的限制,我们在 Octo-Small 模型上执行所有消融。

模型架构:以往基于Transformer的策略设计通常使用大型 ResNet风格[32]编码器对输入图像进行编码,并将得到的图像特征与……融合。

相对较小的变压器[10, 67, 81, 17, 98, 61, 83]。相反,我们选择了一种 “Transformer优先”的架构,该架构使用非常浅层的 CNN 块编码器,并将大部分精力集中在 Transformer 主干网中的参数和 FLOPS,类似到规范的视觉Transformer架构[22]。见表II我们证明,这种可扩展架构能够带来显著的收益。使用完整的 Open X-Embodiment 数据集进行训练时,性能有所提升。重要的是,我们发现基于 ResNet 的模型性能有所提高。在小规模数据集上训练时,架构的性能优于 ViTs。例如,在我们的 “从零开始”的比较中,数据集,下划线大型 Transformer 策略特别适合可扩展性。在多样化的数据集上进行训练。

训练数据:Octo 使用迄今为止最多多样化的跨具身机器人数据集进行训练,该数据集混合了我们收集的 25 个数据集。根据 Open X-Embodiment 数据集[67]手动整理 (参见第三部分B节)。我们消除了这种训练组合的影响。

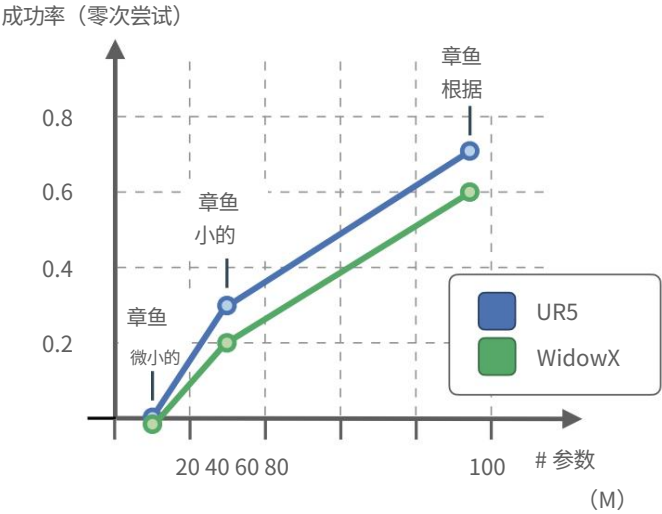


图 6:模型扩展。在 UR5 和 WidowX 任务上，Octo 的性能均随着模型尺寸的增大而提升。成功率是每个机器人在单一语言条件任务上进行 10 次试验的平均值。

通过与使用 11 个数据集（用于训练 RT-X 模型[67]）训练的较小混合数据集训练的 Octo 模型以及仅使用目标机器人领域数据训练的基线模型进行比较，我们得出了以下结论。表II显示，随着训练数据集数量的增加，Octo 的性能也随之提高。这表明，扩展到更多数据集可能会进一步提升策略性能。我们将把这部分工作留待未来，并更深入地研究数据管理的最佳实践。

训练目标:我们将 Octo 的扩散解码训练目标（见第三节 C 部分）与先前工作中常用的替代目标进行比较:简单的均方误差 (MSE) 损失[8, 47]和离散化动作上的交叉熵损失[10, 103]。表II显示，Octo 的扩散训练目标显著提高了性能。这种改进可能是因为扩散头能够对多模态动作分布进行建模（与 MSE 头不同），同时保持连续动作的精度（与离散头不同）。从定性角度来看，该策略比使用 MSE 训练的策略更果断，也比使用离散化动作训练的策略更精确。

模型规模:我们比较了三种不同规模的 Octo 模型，其规模遵循常见的视觉 Transformer 模型阶梯[97]: Octo-Tiny (10M)、Octo-Small (27M) 和 Octo-Base (93M)。图6显示，策略的零样本性能随着模型规模的增加而提升。我们发现，Base 模型比 Small 模型对初始场景配置更具鲁棒性，并且更不容易出现早期抓取尝试，这表明更大的模型具有更好的视觉场景感知能力。

五、讨论与未来工作

我们推出了 Octo，这是一个基于 Transformer 的大型策略模型，它在这迄今为止最大的机器人操作数据集（包含 80 万条机器人轨迹）上进行了预训练。我们证明了 Octo 可以开箱即用解决各种任务，并展示了 Octo 的组合式策略如何实现高效执行。

Octo 的设计使其能够针对新的输入和动作空间进行微调，从而成为解决各种机器人控制问题的通用初始化模型。除了模型本身，我们还发布了完整的训练和微调代码，以及一些工具，方便在大规模机器人数据集上进行训练。

尽管 Octo 在零样本评估和微调评估中均取得了优异的性能，但我们发现当前模型仍然存在一些不足，这很大程度上归因于训练数据的特性。首先，我们发现当前的 Octo 模型难以充分处理腕部摄像头信息。通常情况下，仅使用第三人称摄像头比结合第三人称和腕部摄像头进行微调效果更好。此外，我们注意到语言条件策略的性能与目标条件策略的性能存在显著差异。在这两种情况下，训练数据中相应模态的缺失很可能是造成这种差异的原因:只有 27% 的数据包含腕部摄像头信息，而预训练数据中只有 56% 包含语言标注。

扩展用于训练 Octo 的数据是改进的自然途径。由于 Open X-Embodiment 数据集包含最优机器人演示，当前模型通过模仿进行训练；未来的工作可以考虑从次优或在线交互数据中学习，这需要不同的训练目标。此外，虽然我们仅使用单臂和双臂机械臂训练和评估了 Octo，但扩展到执行导航或移动操作的更广泛的机器人类型将是一个极具发展潜力的方向。

Octo 代表着在构建可在各种机器人配置下开箱即用的通用机器人策略方面迈出了重要一步，但仍需进一步改进模型，包括优化语言条件化、增强对腕部摄像头的支持以及整合最优演示之外的数据。我们希望 Octo 能为研究人员和实践者提供一个简单的平台，以便他们访问更大的机器人数据集，并利用预训练的机器人模型高效学习新任务并实现广泛的泛化能力。

参考

[1] Scale AI.介绍 Scale 的汽车基础模型,2023 年。网址: <https://scale.com/blog/afm1>。

[2] Marcin Andrychowicz,Filip Wolski,Alex Ray,Jonas Schneider, Rachel Fong,Peter Welinder,Bob McGrew, Josh Tobin, Pieter Abbeel 和 Wojciech Zaremba。事后经验重现。发表于 NeurIPS,2017 年。

[3] Shikhar Bahl,Abhinav Gupta 和 Deepak Pathak。人机自然模仿。arXiv 预印本arXiv:2207.09450,2022 年。

[4] Shikhar Bahl,Russell Mendonca,Lili Chen,Unnat Jain 和 Deepak Pathak。“从人类视频中提取的可供性作为机器人学的多功能表示”。CVPR,2023。

[5] Suneel Belkhale,Yuchen Cui 和 Dorsa Sadigh。Hydra:用于模仿学习的混合机器人动作。archiv,2023。

[6] Homanga Bharadhwaj,Jay Vakil,Mohit Sharma,Abhinav Gupta,Shubham Tulsiani 和 Vikash Kumar。

Roboagent:通过语义增强和动作分块实现机器人操作的泛化和效率。
arXiv预印本 arXiv:2309.01918,2023。

[7] Kevin Black,Mitsuhiro Nakamoto,Pranav Atreya、Homer Walke,Chelsea Finn、Aviral Kumar 和 Sergey Levine。基于预训练图像编辑扩散模型的零样本机器人操作。
arXiv 预印本arXiv:2310.10639,2023 年。

[8] Mariusz Bojarski.Davide Del Testa,Daniel Dworakowski,Bernhard Firner,Beat Flepp,Prasoon Goyal,Lawrence D Jackel,Mathew Monfort,Urs Muller,Jaikai Zhang 等。面向自动驾驶汽车的端到端学习。arXiv 预印本 arXiv:1604.07316,2016年。

[9] Konstantinos Bousmalis,Giulia Vezzani,Dushyant Rao、Coline Devin,Alex X Lee、Maria Bauza,Todor Davchev,周宇翔,Agrim Gupta,Akhil Raju 等。Robocat:用于机器人操作的自我改进基础代理。arXiv 预印本 arXiv:2306.11706, 2023。

[10] Anthony Brohan,Noah Brown,Justice Carbajal,Yev-gen Chebotar、Joseph Babis,Chelsea Finn,Keerthana Gopalakrishnan,Carol Hausman,Alex Herzog,Jasmine Hsu 等人。Rt-1:机器人技术转变为现实世界的控制规模。arXiv 预印本 arXiv:2212.06817,

[11] Anthony Brohan,Yevgen Chebotar,Chelsea Finn,Karol Hausman,Alexander Herzog,Daniel Ho,Julian Ibarz、Alex Irpan,Eric Jang,Ryan Julian 等。“按我的能力做事,而不是按我说的做事:将语言建立在机器人可供性之上”。载于机器人学习会议论文集,第 287–318 页。PMLR, 2023 年。

[12] Serkan Cabi,Sergio Gómez Colmenarejo,Alexander Novikov,Ksenia Konyushkova、Scott Reed,Rae Jeong、Konrad Zolna,Yusuf Aytar,David Budden,Mel Vecerik、Oleg Sushkov,David Barker,Jonathan Scholz,Misha Denil,Nando de Freang 和 Ziyu Wyang。利用奖励草图和批量强化学习扩展数据驱动机器人技术。arXiv 预印本 arXiv:1909.12200,2019。

[13] Holger Caesar.Varun Bankiti,Alex H Lang,Sourabh Vora,Venice Erin Liong,Qiang Xu,Anush Krishnan、Yu Pan,Giancarlo Baldan 和 Oscar Beijbom。nuscenes:用于自动驾驶的多模态数据集。载于IEEE/CVF 计算机视觉与模式识别会议论文集,第 11621–11631 页, 2020 年。

[14]劳伦斯·云亮·陈,西蒙·阿德博拉,肯·戈德堡。伯克利 UR5 演示数据集。<https://www.//sites.google.com/view/berkeley-ur5/home>。

[15] William Chen,Oier Mees,Aviral Kumar 和 Sergey Levine。视觉语言模型为强化学习提供可提示的表示。arXiv 预印本arXiv:2402.02651,2024 年。

[16] Zoey Chen,Sho Kiani,Abhishek Gupta 和 Vikash Kumar,Genaug:通过生成式增强将行为重定向到未见过的情况。arXiv 预印本arXiv:2302.06671,2023 年。

[17] Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel 和 Shuran Song。扩散策略:通过动作扩散学习视觉运动策略。载于 2023 年机器人学:科学与系统 (RSS) 会议论文集。

[18] Zichen Jeff Cui,Yibin Wang,Nur Muhammad Mahi Shafiullah 和 Lerrel Pinto。“从游戏到策略:基于未经整理的机器人数据的条件行为生成”。第十一届国际学习表征会议,2022 年。

[19] Sudeep Dasari,Frederik Ebert,Stephen Tian,Suraj Nair,Bernadette Bucher,Karl Schmeckpeper,Siddharth Singh,Sergey Levine 和 Chelsea Finn。“Robonet:大规模多机器人学习”。载于机器人学习会议论文集,第 885–897 页。PMLR,2020 年。

[20] Shivin Dass,Julian Yapeter,Jesse 张,Jiahui 张、Karl Pertsch、Stefanos Nikolaidis 和 Joseph J. Lim。CLVR jaco play 数据集,2023 年。网址: https://github.com/clvrai/clvr_jaco_play_dataset。

[21] Pim de Haan,Dinesh Jayaraman 和 Sergey Levine。模仿学习中的因果混淆。NeurIPS,2019。

[22] Alexey Dosovitskiy,Lucas Beyer,Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai,Thomas Unterthiner、Mostafa Dehghani,Matthias Minderer、Georg Heigold、Sylvain Gelly 等。一张图像相当于 16x16 个单词:用于大规模图像识别的 Transformer。arXiv预印本 arXiv:2010.11929, 2020。

[23] Danny Driess,Fei Xia,Mehdi SM Sajjadi,Corey Lynch、Aakanksha Chowdhery,Brian Ichter,Ayzaan Wahid、Jonathan Thompson、Quan Vuong,Tianhe Yu,等。Palm-e:一种具体的多模式语言模型。arXiv预印本 arXiv:2303.03378, 2023。

[24] Maximilian Du,Suraj Nair,Dorsa Sadigh 和 Chelsea Finn。“行为检索:通过查询未标记数据集进行少样本模仿学习”。ArXiv,abs/2304.08742,2023年。URL: <https://api.semanticscholar.org/CorpusID:258186973>。

[25] Frederik Ebert,Yanlai Yang,Karl Schmeckpeper,Bernadette Bucher、Georgios Georgakis,Kostas Daniilidis,Chelsea Finn 和 Sergey Levine。“桥接数据:利用跨域数据集提升机器人技能的泛化能力”。arXiv 预印本 arXiv:2109.13396,2021 年。

[26] Hao-Shu Fang, Hongjie Fang, Zhenyu Tang, Jirong Liu, Chenxi Wang, Junbo Wang, Haoyi Zhu, and Cewu Lu。Rh20t:一个用于一次性学习多种技能的综合机器人数据集。迈向通用机器人:可扩展技能习得的学习范式@ CoRL2023,3:5,2023。

[27] Chelsea Finn 和 Sergey Levine。用于规划机器人运动的深度视觉预测。载于 2017 年 IEEE 国际机器人与自动化会议 (ICRA),第2786–2793 页。IEEE,2017 年。

[28] Zipeng Fu,Tony Z Zhao 和 Chelsea Finn。“移动阿罗哈:利用低成本全身远程操作学习双手移动操作”。arXiv 预印本arXiv:2401.02117,2024 年。

- [29] Andreas Geiger, Philip Lenz 和 Raquel Urtasun. “我们准备好迎接自动驾驶了吗? kitti 视觉基准测试套件” 。载于 2012 年 IEEE 计算机视觉与模式识别会议论文集, 第 3354–3361 页。IEEE, 2012 年。
- [30] 阿比纳夫 古普塔、阿迪提亚瓦万·穆拉里、迪拉杰·普拉卡什昌德·甘地和莱雷尔·平托。家庭机器人学习: 提高泛化能力并减少数据集偏差。神经信息处理系统进展, 31, 2018。
- [31] Huy Ha, Pete Florence 和 Shuran Song. “规模化和精简: 语言引导的机器人技能习得” 。在机器人学习会议论文集第 3766–3777 页。PMLR, 2023。
- [32] 何凯明, 张翔宇, 任绍庆, 孙健。用于图像识别的深度残差学习。IEEE 计算机视觉与模式识别会议论文集, 第 770–778 页, 2016 年。
- [33] Minh Heo, Youngwoon Lee, Doohyun Lee 和 Joseph J. Lim. Furniturebench: 用于长时程复杂操作的可复现真实世界基准测试。载于《机器人学: 科学与系统》, 2023 年。
- [34] Jonathan Ho, Ajay Jain 和 Pieter Abbeel. 扩散概率模型去噪。神经信息处理系统进展, 33: 6840–6851, 2020。
- [35] Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton 和 Gianluca Corrado. Gaia-1: 自动驾驶的生成式世界模型, 2023 年。
- [36] Chenguang Huang, Oier Mees, Andy Zeng 和 Wolfram Burgard. “用于机器人导航的视觉语言地图” 。载于 2023 年 IEEE 国际机器人与自动化会议 (ICRA), 第 10608–10615 页。IEEE, 2023 年。
- [37] Chenguang Huang, Oier Mees, Andy Zeng 和 Wolfram Burgard. 用于机器人导航的视听语言地图。载于 2023 年泰国清迈国际实验机器人研讨会 (ISER) 会议录。
- [38] Wenlong Huang, Chen Wang, Ruohan Zhang, Yunzhu Li, Jiajun Wu, and Li Fei-Fei. Voxposer: Composable 3d value maps for robotic manipulation with language models. arXiv preprint arXiv:2307.05973, 2023.
- [39] Eric Jang, Alex Irpan, Mohi Khansari, Daniel Kappler, Frederik Ebert, Corey Lynch, Sergey Levine 和 Chelsea Finn. Bc-z: 基于机器人模仿学习的零样本任务泛化。载于机器人学习会议论文集, 第 991–1002 页。PMLR, 2022 年。
- [40] 江云帆, Agrim Gupta, 张子琛, 王冠志, 窦永强, 陈彦军, 李飞飞, Anima Anandkumar, 朱玉珂, 范林熙. VIMA: 基于多模态提示的机器人操作。载于 Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato 和 Jonathan Scarlett 编辑的《第 40 届国际机器学习会议论文集》, 《机器学习研究论文集》第 202 卷, 第 14975–15022 页。PMLR,
- 2023 年 7 月 23 日至 29 日。网址 <https://proceedings.mlr.press/v202/jiang23b.html>。
- [41] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke 等. QT-Opt: 用于基于视觉的机器人操作的可扩展深度强化学习。arXiv 预印本 arXiv:1806.10293, 2018。
- [42] Dmitry Kalashnikov, Jake Varley, Yevgen Chebotar, Benjamin Swanson, Rico Jonschkowski, Chelsea Finn, Sergey Levine 和 Karol Hausman. “扩展多任务机器人强化学习” 。第五届机器人学习年会, 2021 年。
- [43] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Sören Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas 和 Peter Stone. “社交合规导航数据集 (scand): 用于社交导航的大规模演示数据集” 。IEEE 机器人与自动化快报, 7(4): 11807–11814, 2022。
- [44] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Shaw, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo 等. 细分任何内容, 2023 年 4 月。
- [45] Vikash Kumar, Rutav Shah, Gaoyue Zhou, Vincent Moens, Vittorio Caggiano, Abhishek Gupta 和 Ar-avind Rajeswaran. “Robohive: 机器人学习的统一框架” 。载于第三十七届神经信息处理系统会议数据集和基准测试专题, 2023 年。网址: <https://openreview.net/forum?id=0H5fRQcpQ7>。
- [46] Teyun Kwon, 诺曼·迪·帕洛和爱德华·约翰斯。语言模型作为零样本轨迹生成器。arXiv 预印本 arXiv:2310.11604, 2023 年。
- [47] Sergey Levine, Chelsea Finn, Trevor Darrell 和 Pieter Abbeel. 深度视觉运动策略的端到端训练。机器学习研究杂志, 17(1): 1334–1373, 2016。
- [48] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz 和 Deirdre Quillen. 利用深度学习和大规模数据收集学习机器人抓取的手眼协调。《国际机器人研究杂志》, 37(4-5): 421–436, 2018。
- [49] Yixin Lin, Austin S. Wang, Giovanni Sutoanto, Akshara Rai 和 Franziska Meier. 波吕墨提斯。 <https://facebookresearch.github.io/fairo/polymetis/>, 2021 年。
- [50] Huihan Liu, Soroush Nasiriany, Lance Zhang, Zhiyao Bao, and Yuke Zhu. Robot learning on the job: Human-in-the-loop autonomy and learning during deployment. 在机器人学: 科学与系统 (RSS) 中, 2023 年。
- [51] Ilya Loshchilov 和 Frank Hutter. 解耦权重衰减正则化。2018 年国际学习表征会议论文集。
- [52] Jianlan Luo, Charles Xu, Xinyang Geng, Gilbert Feng, Kuan Fang, Liam Tan, Stefan Schaal, and Sergey Levine. 通过分层模仿学习实现多阶段电缆布线。arXiv 预印本 arXiv:2307.08927, 2023 年。

[53] Jianlan Luo, Charles Xu, Fangchen Liu, Liam Tan, Zipeng Lin, Jeffrey Wu, Pieter Abbeel 和 Sergey Levine. Fmb:用于通用机器人学习的功能操作基准测试。arXiv 预印本arXiv:2401.08553, 2024。

[54] Corey Lynch 和 Pierre Sermanet.基于非结构化数据的语言条件模仿学习。发表于 RSS, 2021 年。

[55] Corey Lynch.Ayzaan Wahid,Jonathan Tompson、丁天利、James Betker、Robert Baruch、Travis Armstrong和 Pete Florence。交互式语言 :与机器人实时对话。IEEE 机器人与自动化快报,2023 年。

[56] Arjun Majumdar,Karmesh Yadav,Sergio Arnaud、 Yecheng Jason Ma,Claire Chen、 Sneha Silwal,Aryan Jain.Vincent-Pierre Berges,Pieter Abbeel,Jitendra Malik、 Dhruv Batra,Yixin Lin,Oleksandr Maksymets、 Aravind Rajeswaran 和 Franziska Meier。“在探索具身智能的人工视觉皮层方面,我们进展如何?” 2023。

[57] Arjun Majumdar,Karmesh Yadav,Sergio Arnaud、 Yecheng Jason Ma,Claire Chen、 Sneha Silwal,Aryan Jain.Vincent-Pierre Berges,Pieter Abbeel,Jitendra Malik 等。我们在探索具身智能的人工视觉皮层方面进展如何?arXiv 预印本arXiv:2303.18240, 2023 年。

[58] Ajay Mandlekar,Yuke Zhu.Animesh Garg,Jonathan Booher,Max Spero.Albert Tung,Julian Gao,John Emmons.Anchit Gupta,Emre Orbay,Silvio Savarese和李飞飞.RoboTurk:一个通过模仿进行机器人技能学习的众包平台。CoRR,abs/1811.02790,2018,URL:<http://arxiv.org/abs/1811.02790>。

[59] Ajay Mandlekar,Yuke Zhu.Animesh Garg,Jonathan Booher,Max Spero.Albert Tung,Julian Gao,John Emmons.Anchit Gupta,Emre Orbay 等。Roboturk:一个通过模仿进行机器人技能学习的众包平台。载于机器人学习会议论文集,第 879–893 页。PMLR ,2018 年。

[60] Ajay Mandlekar,Soroush Nasiriany,Bowen Wen,Ireti- ayo Akinola,Yashraj Narang, Linxi Fan,Yuke Zhu 和 Dieter Fox。Mimicgen:一种利用人类演示进行可扩展机器人学习的数据生成系统。在2023 年第七届机器人学习年会上。

[61] Oier Mees,Luke Hermann 和 Wolfram Burgard。在非结构化数据上进行语言条件机器人模仿学习时,哪些因素至关重要。IEEE机器人与自动化快报,7(4):11205–11212, 2022。

[62] Oier Mees,Lukas Hermann,Erick Rosete-Beas 和Wolfram Burgard。Calvin:用于长时域机器人操作任务的语言条件策略学习的基准。IEEE 机器人与自动化快报(RA-L), 7(3):7327–7334,2022。

[63] Oier Mees,Jessica Borja-Diaz 和 Wolfram Burgard。利用视觉可供性将语言与非结构化数据联系起来。发表于IEEE国际机器人与自动化会议 (ICRA)论文集,伦敦。

英国,2023年。

[64] Russell Mendonca,Shikhar Bahl 和 Deepak Pathak。基于人类视频的结构化世界模型。CoRL, 2023。

[65] Soroush Nasiriany,Tian Gao,Ajay Mandlekar 和 Yuke Zhu。基于技能的模仿学习中从先验数据学习和检索。在2022 年机器人学习会议 (CoRL) 上。

[66] Alexander Quinn Nichol 和 Prafulla Dhariwal。改进的去噪扩散概率模型。国际机器学习会议论文集,第8162–8171 页。PMLR,2021 年。

[67] Open X-Embodiment Collaboration, Abhishek Padalkar, Acorn Pooley, Ajinkya Jain, Alex Bewley, Alex Her-zog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anikait Singh, Anthony Brohan, Antonin Raffin, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Brian Ichter, Cewu Lu, Charles Xu, Chelsea Finn, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Chuer Pan, Chuyuan Fu, Coline Devin, Danny Driess, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Federico Ceola, Fei Xia, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Giulio Schiavi、苏浩、方浩书、史浩辰、Henri Ben Amor、Henrik I Christensen、Hiroki Furuta、Homer Walke、方洪杰、Igor Mordatch、Ilija Radosavovic、Isabel Leal、Jacky Liang、Jaehyung Kim、 Jan Schneider、Jasmine Hsu、Jeannette Bohg、Jeffrey Bingham、Jiajun Wu、Jialin Wu、Jianlan Luo、 Jiayuan Gu, Jie Tan, Jihoon Oh, Jitendra Malik, Jonathan Tompson, Jonathan Yang, Joseph J. Lim, João Silvério, Junhyek Han, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin 张, Keyvan Majd, 克里尚 拉纳,克里希南Srinivasan、Lawrence Yunliang Chen、Lerrel Pinto、 Liam Tan、Lionel Ott、Lisa Lee、Masayoshi Tomizuka、 Maximilian Du、Michael Ahn、 Mingtong 张、Mingyu Ding、Mohan Kumar Srirama、Mohit Sharma、Moo Jin Kim、 Naoaki Kanazawa、Nicklas Hansen、Nicolas Heess、 Nikhil J Joshi、Niko Suenderhauf、Norman Di Palo、Nur Muhammad Mahi Shafiullah、Oier Mees、 Oliver Kroemer、Pannag R Sanketi、Paul Wohlhart、Peng Xu、 Pierre Sermanet、 Priya Sundareshan、Quan Vuong、Rafael Rafailov、Ran Tian、Ria Doshi、Roberto Martín-Martín、 Russell Mendonca、Rutav Shah、Ryan Hoque、Ryan Julian 、 Samuel Bustamante、Sean基玛尼、谢尔盖·莱文、雪莉Moore、Shikhar Bahl、Shivini Dass、Shuran Song、 Sichun Xu、Siddhant Haldar、Simeon Adebola、Simon Guist、Soroush Nasiriany、Stefan Schaal、Stefan Welker、 Stephen Tian、Sudeep Dasari、Suneel Belkale、Takayuki Osa、Tatsuya Harada、Tatsuya Matsushima、 Ted Xiao、于天河、Tianli Ding、Todor Davchev、Tony Z. 赵、 Travis Armstrong、 Trevor Darrell、Vidhi Jain、Vincent Vanhoucke、We Zhan、Wenxuan Zhou、 Wolfram Burgard、

Xi Chen, Xiaolong Wang, Xinghao Zhu, Xuanlin Li, Yao Lu, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yueh hua Wu, Yujin Tang, Yuke Zhu, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zhuo Xu, and Zichen Jeff Cui. Open X-Embodiment: Robotic learning datasets and RT-X models. <https://arxiv.org/abs/2310.08864>, 2023.

[68] OpenAI. GPT-4 技术报告,2023 年 3 月。

[69] Jyothish Pari.Nur Muhammad Shafiullah.Sridhar Pandian Arunachalam 和 Lerrel Pinto.视觉模仿表征学习的惊人有效性,2021。

[70] Ethan Perez,Florian Strub,Harm De Vries,Vincent Dumoulin 和 Aaron Courville。“电影:具有通用条件层的视觉推理”。载于 AAAI 人工智能会议论文集,第 32 卷,2018 年。

[71] Lerrel Pinto 和 Abhinav Gupta. 超大型自监督:从 50k 次尝试和 700 个机器人小时中学习抓取。在 2016 年 IEEE 国际机器人与自动化会议 (ICRA) 中,第 3406-3413 页。IEEE,2016 年。

[72] Gabriel Quere,Annette Hagengruber,Maged Iskandar, Samuel Bustamante, Daniel Leidner,Freek Stulp 和 Joern Vogel。“辅助机器人的共享控制模板”。载于 2020 年 IEEE 国际机器人与自动化会议 (ICRA),第 7 页,法国巴黎, 2020 年。

[73] Ilija Radosavovic,Baifeng Shi,Letian Fu,Ken Goldberg、 Trevor Darrell 和 Jitendra Malik。“基于感觉运动预训练的机器人学习”。机器学习会议,2023 年。

[74] Colin Raffel,Noam Shazeer,Adam Roberts,Katherine Lee,Sharan Narang、 Michael Matena,Yanqi Zhou,Wei Li 和 Peter J. Liu。“使用统一的文本到文本转换器探索迁移学习的极限”。《机器学习研究杂志》,21(140):1-67,2020。

URL <http://jmlr.org/papers/v21/20-074.html>。

[75] Scott Reed,Konrad Zolna,Emilio Parisotto,Sergio Gómez Colmenarejo, Alexander Novikov,Gabriel Barth-maroon,Mai Giménez,Yury Sulsky,Jackie Kay,Jost Tobias Springenberg 等。多才多艺的经纪人。机器学习研究学报,2022 年。

[76] 罗宾·隆巴赫、安德烈亚斯·布拉特曼、多米尼克·洛伦茨、帕特里克·埃瑟和比约恩·奥默。使用潜在扩散模型进行高分辨率图像合成,2022 年 4 月。

[77] Erick Rosete-Beas,Oier Mees,Gabriel Kalweit,Joschka Boedecker 和 Wolfram Burgard。面向任务无关离线强化学习的潜在计划。载于 2022 年第 6 届机器人学习会议 (CoRL) 会议录。

[78] Saumya Saxena,Mohit Sharma 和 Oliver Kroemer。基于视觉语言模型的实时控制多分辨率传感。第七届机器人学习年会,2023。网址: <https://openreview.net/forum?id=WuBv9-IGDUA>。

[79] Nur Muhammad Mahi Shafiullah,Anant Rai,Haritheja Etukuru, Yiqian Liu,Ishan Misra,Soumith Chintala 和 Lerrel Pinto。关于将机器人带回家,2023 年。

[80] Dhruv Shah,Ajay Sridhar,Arjun Bhorkar,Noriaki Hirose 和 Sergey Levine。GNM:一种驱动任何机器人的通用导航模型。载于 2023 年 IEEE 国际机器人与自动化会议 (ICRA),第 7226-7233 页。IEEE,2023 年。

[81] Dhruv Shah,Ajay Sridhar,Nitish Dashora,Kyle Sta-chowicz,Kevin Black,Noriaki Hirose 和 Sergey Levine。ViNT:视觉导航的基础模型。在 2023 年第七届机器人学习年会上。URL <https://arxiv.org/abs/2306.14846>。

[82] 鲁塔夫·沙阿、罗伯特·马丁·马丁·朱宇科。MUTEX:从多模态任务规范中学习统一策略。第七届机器人学习年会,2023。网址: <https://openreview.net/forum?id=PwqiaaEzJ>。

[83] 莫希特·施里达尔、卢卡斯·曼努埃利和迪特·福克斯。感知器-执行器:用于机器人操作的多任务转换器。载于机器人学习会议论文集,第 785-799 页。PMLR,2023。

[84] Ishika Singh,Valts Blukis,Arsalan Mousavian,Ankit Goyal,Danfei Xu,Jonathan Tremblay,Dieter Fox,Jesse Thomason 和 Animesh Garg。Progprompt:使用大型语言模型生成情境机器人任务计划。2023 年 IEEE 国际机器人与自动化会议 (ICRA),第 11523-11530 页。IEEE, 2023 年。

[85] Ajay Sridhar,Dhruv Shah,Catherine Glossop 和 Sergey Levine。“Nomad:用于导航和探索的目标掩蔽扩散策略”。arXiv 预印本 arXiv:2310.07896,2023 年。

[86] Austin Stone, Ted Xiao, Yao Lu, Keerthana Gopalakrishnan, Kuang-Huei Lee, Quan Vuong, Paul Wohlhart, Sean Kirmani, Brianna Zitkovich, Fei Xia 等。使用预训练视觉语言模型进行开放世界物体操作。在 2023 年第 7 届机器人学习年会上。

[87] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine 等。自动驾驶感知的可扩展性:Waymo 开放数据集。IEEE/CVF 计算机视觉与模式识别会议论文集,第 2446-2454 页,2020 年。

[88] Hugo Touvron,Thibaut Lavril,Gautier Izacard,Xavier Martinet,Marie-Anne Lachaux,Timothée Lacroix,Baptiste Rozière,Naman Goyal, Eric Hambro,Faisal Azhar 等。LLaMA:开放高效的基础语言模型,2023 年 2 月。

[89] Samuel Triest,Matthew Sivaprakasam,Sean J Wang,Wenshan Wang, Aaron M Johnson 和 Sebastian Scherer。“Tartandrive:用于学习越野动力学模型的大规模数据集”。载于 2022 年国际机器人与自动化会议 (ICRA),第 2546-2552 页。IEEE,2022 年。

[90] Homer Walke,Kevin Black,Abraham Lee,Moo Jin Kim、 Max Du、Chongyi Zheng,Tony Zhao,Philippe Hansen- Estruch,Quan Vuong,Andre He,Vivek Myers,Kuan Fang,Chelsea Finn 和 Sergey Levine。Bridgedata v2 :用于大规模机器人学习的数据集,2023 年。

[91] Wayve. Lingo :自动驾驶的自然语言 ,2023 年。
网址<https://wayve.ai/thinking/lingo-natural-language-autonomous-driving/>。

[92] Philip Wu,Arjun Majumdar,Kevin Stone,Yixin Lin、 Igor Mordatch,Peter Abel 和 Aravind Rajeswaran。
用于预测、表示和控制的掩码轨迹模型。2023 年国际机器学习会议。

[93] Ge Yan,Kris Wu 和 Xiaolong Wang。加州大学圣地亚哥分校厨房数据集。2023 年 8 月。

[94] Jonathan Heewon Yang,Dorsa Sadigh 和 Chelsea Finn。
“Polybot :在适应变异性的同时跨机器人训练单一策略” 。第七届机器人学习年会,2023 年。网址： <https://openreview.net/forum?id=HEIRj51lcS>。

[95] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan 和Trevor Darrell. Bdd100k :用于异构多任务学习的多样化驾驶数据集。在IEEE/CVF 计算机视觉与模式识别会议论文集,第 2636–2645 页,2020 年。

[96] Tianhe Yu, Ted Xiao, Austin Stone, Jonathan Tompson, Anthony Brohan, Su Wang, Jaspier Singh, Clayton Tan, Jodilyn Peralta, Brian Ichter 等。利用语义想象经验扩展机器人学习。arXiv 预印本arXiv:2302.11550,2023。

[97]翟晓华,Alexander Kolesnikov,Neil Houlsby 和 Lucas Beyer。“扩展视觉转换器” ,IEEE /CVF 计算机视觉与模式识别会议论文集,第 12104–12113 页,2022 年。

[98] Tony Z Zhao,Vikash Kumar,Sergey Levine 和 Chelsea Finn。使用低成本硬件学习细粒度的双手动操作。arXiv 预印本 arXiv:2304.13705,2023年。

[99] Gaoyue Zhou, Victoria Dean, Mohan Kumar Srirama, Aravind Rajeswaran, Jyothish Pari, Kyle Hatch, Aryan Jain, Tianhe Yu, Pieter Abbeel, Lerrel Pinto, Chelsea Finn, 和 Abhinav Gupta. Train offline, test online: A real robot learning benchmark, 2023.

[100]朱星浩,田冉,徐晨峰,丁明宇,詹伟,富家正芳。 Fanuc 操作 :使用fanuc mate 200id 机器人进行基于学习的操作的数据集。 2023 年。

[101] Yifeng Zhu, Peter Stone 和 Yuke Zhu. 基于非分段演示的自下而上技能发现,用于长时域机器人操作. IEEE 机器人与自动化快报, 7(2):4126–4133, 2022.

[102] Yifeng Zhu, Abhishek Joshi, Peter Stone, and Yuke Zhu. Viola :基于视觉的模仿学习与对象提议先验,2023。

[103] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu,

Ted Xiao,Fei Xia,Jialin Wu,Paul Wohlhart,Stefan Welker,Ayzaan Wahid 等人。Rt-2 :视觉语言动作模型将网络知识转移到机器人控制。

在 2023 年第七届机器人学习年会上。

附录A
贡献

Dibya Ghosh:领导模型开发,提出并实施了最终模型设计的大部分内容,并全程负责。

参与了训练,接触了代码库的各个部分,并提供了帮助。

包括模型评估和技术报告撰写。

霍默·沃克:领导模型评估,设计主要桥梁评估基准,为初始模型做出了贡献。

实施并运行了本技术报告中的许多评估。

Karl Pertsch:负责整个项目,领导 Open-X 数据整合与整理工作主导了这份技术报告的撰写。

参与模型开发和实施,并负责运行技术报告的模型评估。

Kevin Black:领导数据加载和训练基础设施建设,管理TPU小组培训,创建项目网站,为模型开发和实施做出了贡献,并且协助进行机器人评估和技术报告撰写。

Oier Mees:进行了无数次消融实验以开发模型,参与模型实施,协助评估和撰写技术报告。

Sudeep Dasari:为卡内基梅隆大学的模型评估做出了贡献,尝试使用预训练编码器。

Joey Hejna:为斯坦福大学的模型评估做出了贡献。

Tobias Kreiman:对模拟模型的评估做出了贡献。

环境。

Ria Doshi:对交叉形态学的研究做出了贡献。

训练。

徐查尔斯:为伯克利大学贡献了模型评估。

插销插入件。

罗建兰:为伯克利大学贡献了模型评估。

插销插入件。

谭友良:协助诊断和解决瓶颈问题。

数据正在加载。

陈云亮:协助创建并运营UR5。

在伯克利汽车实验室进行微调实验。

Pannag Sanketi,Quan Vuong,Ted Shaw:贡献模型 Google Robot 上的评估。

多尔萨·萨迪格、切尔西·芬恩、谢尔盖·莱文:提供项目全程指导,并对写作提供反馈。

本技术报告。

附录B
OCTO代码示例

加载预训练的 Octo 模型并执行推理
所需代码量很少:

```
1 导入jax
2 从octo.model.octo_model导入OctoModel
3
4 model = OctoModel.load_pretrained( hf://rail-berkeley/octo-base )
5 print(model.get_pretty_spec()) # 打印输入输出规范
6 观察 = { image_primary :img}
7 task = model.create_tasks(texts=[ 拿起文字 ])
8 操作 = model.sample_actions(
9     观察,任务, rng=jax.random.PRNGKey(0))
```

示例 1:使用 Python 代码进行推理的示例
预训练的 Octo 模型。

附录C
数据混合

我们列出了用于训练的详细训练混合模型。

表 III 中的 Octo 模型。采样权重大多由数据集的相对大小决定,其中少数手动调整 (参见第三部分B节) 。

Octo 预训练数据集混合	
分形[10]	17.0%
哭泣[41]	17.0%
桥[25, 90]	17.0%
BC-Z [39]	9.1%
斯坦福九头蛇数据集[5]	6.0%
语言表[55]	5.9%
Taco Play [77, 63]	3.6%
家具长凳数据集[33]	3.3%
UTAustin Mutex [82]	3.0%
奥斯汀水手数据集[65]	2.9%
Roboturk [58]	2.8%
这[99]	2.4%
Austin Sirius 数据集[50]	2.3%
Berkeley Autolab UR5 [14]	1.5%
IAMLab CMU 拾取插入[78]	1.2%
中提琴[102]	1.2%
Berkeley Fanuc 操纵[100]	1.0%
纽约大学弗兰卡游戏数据集[18]	0.9%
加州大学圣地亚哥分校厨房数据集[93]	小于0.1%
Jaco Play [20]	0.6%
伯克利电缆布线[52]	0.3%
Austin Buds 数据集[101]	0.3%
CMU 拉伸[64]	0.2%
纽约大学大门开启[69]	0.1%
DLR EDAN 共享控制[72]	0.1%

表三:使用来自以下数据集的 Octo 预训练数据混合
Open X-Embodiment 数据集[67]。

附录D
训练超参数

我们主要遵循有据可查的视觉训练方法。

变压器[97]。我们使用AdamW优化器[51],并结合一个反平方根衰减学习率方案[97]和线性学习率预热阶段。我们列出了预热阶段使用的超参数。

表IV中的训练数据以及不同模型的参数。

表V中列出了尺寸。我们应用了标准图像增强技术。在训练期间。具体来说,对于第三人称视角,我们应用随机裁剪后,将图像大小调整为256×256,然后……颜色抖动。最后,我们将输入图像归一化为像素数为 10^6。浮点数值介于-1.0和1.0 之间。对于腕式摄像头,我们采用相同的步骤,只是不进行随机裁剪。改为调整尺寸为 128 × 128。

超参数	价值
学习率	3e-4
热身步骤	2000
LR 调度器	平方根倒数
体重下降	0.1
梯度裁剪阈值	1
批量大小	2048

表 IV:训练期间使用的超参数。

图像经过浅层卷积层处理后,被分割成一系列大小为 16×16 的扁平化图像块[22]。这使得第三人称视角图像生成 256 个词元,腕部视角图像生成 64 个词元。对于包含语言标注的数据集,我们使用预训练的基于 t5 的Transformer 模型 (111M) [74],该模型生成 16 个语言嵌入词元序列。

模型	图层 隐藏层大小 D MLP 大小 头部参数		
Octo-Small 12	384	1536	6 27M
Octo-Base 12	768	3072	12 93M

表 V:Octo 型号变体的架构详情。

扩散作用头由一个3层多层感知器 (MLP)构成,其隐藏层维度为256,并带有残差连接和层归一化。我们采用文献[34]提出的标准DDPM目标函数,并结合余弦噪声调度[66]和20个扩散步骤。

附录E 哪些方法有效,哪些方法 (目前)无效

我们发现以下方法可以提高性能:

·在预训练过程中添加历史数据:在零样本评估中,以一帧历史数据作为上下文的模型比未添加历史数据的预训练模型表现更好。在我们评估的少数几个任务中,我们没有观察到进一步增加历史数据长度的益处,但其他任务可能从中受益。

·使用动作分块:我们发现使用“动作分块” [98] (即预测未来的多个动作)有助于获得更连贯的策略运动。在我们的评估中,我们发现对未来动作进行时间集成并没有在滚动视界控制之外带来额外的好处。 ·减小图像块大小:将图像分块为 16×16 大小的图像块比 32×32 大小的图像块性能更好,尤其是在抓取和其他细粒度任务中。但这确实增加了计算复杂度 (标记数量是原来的 4 倍),因此如何平衡计算成本和分辨率仍然是一个值得研究的问题。

·增加混洗缓冲区大小:并行加载来自 25 个数据集的数据是一项挑战。具体来说,我们发现训练过程中帧的良好混洗至关重要 使用较小的混洗缓冲区 (20k)和轨迹级交错时,零样本性能会显著下降。我们通过了解码图像之前对来自不同轨迹的帧进行混洗和交错来解决这个问题,从而能够使用更大的混洗缓冲区 (高达 500k)。此外,在数据加载期间,我们从每个训练轨迹中随机抽取最多 100 个步骤进行子采样,以避免单个过长的片段 “过度填充”混洗缓冲区。

尚未奏效的方法:

· MSE 动作头:用简单的 L2 损失替换我们的扩散解码头,导致 “对冲”策略移动非常缓慢,例如,在WidowX 评估中无法旋转夹爪。

·离散动作头:将动作离散化到每个维度 256 个区间,并使用类似 Brohan 等人[10] 的交叉熵损失进行训练,可以得到更 “果断”的策略,但精度不足,经常无法准确把握目标。

· ResNet 编码器:在我们的评估中,它们在大规模数据集上的扩展性不佳 (见表II), 尽管在小型数据集 (约 100 个演示)上从头开始训练时,它们的性能优于我们的 ViT 架构。 ·预训练编码器: ImageNet 预训练的 ResNet 编码器在零样本评估中没有带来优势,但这可能与上述ResNet 架构性能不佳的情况有关。

·相对夹爪动作表示:在对齐不同数据集的夹爪动作表示时,我们尝试了两种方法: (A) 绝对夹爪动作表示,即夹爪打开时动作值为 +1,关闭时为 0;(B) 相对夹爪动作表示,即仅在夹爪打开/关闭的时间步中动作值为 +1/0,其他时间步为 0.5。我们发现,由于大多数训练数据代表的是 “不改变夹爪”的动作,因此相对表示方法倾向于减少夹爪的打开/关闭频率,从而略微提高了抓取成功率。但同时,相对表示方法导致抓取失败后的重试行为减少,最终效果更差。因此,我们选择了绝对夹爪动作表示。 ·添加本体感觉输入:使用本体感觉观测数据训练的策略似乎总体上效果更差,这可能是由于状态和未来动作之间存在很强的相关性。我们假设这可能是由于本体感觉信息和目标动作之间存在因果混淆[21]。 ·语言模型微调:为了提升 Octo 的视觉语言基础,我们尝试了以下方法:i) 改变 T5 编码器[74] 的大小:小型(30M)、基础(111M) 和大型(386M);以及 ii) 微调编码器的最后两层。使用冻结的基础模型获得了最佳的语言条件策略。使用更大的编码器或微调编码器均未带来改进。我们推测这可能是由于大多数数据集缺乏丰富、多样化的自由语言标注。

附录F 实验装置

A. 零样本评估

WidowX BridgeV2:采用Walke等人[90]的设置,其中Trossen WidowX 250 6自由度机器人执行各种操作任务。观察数据由单个第三人称摄像机流组成,动作空间为末端执行器位置变化量。我们评估了两个语言条件反射任务。



图 7:评估任务。为方便起见,此图直接摘自正文。我们在 4 个机构的 9 个真实机器人设置中,分别在零样本和微调场景下对 Octo 进行了评估。

其中,机器人需要“将胡萝卜放在盘子上”和“将茄子放入锅中”。虽然这些任务属于分布式任务,但策略仍需能够泛化到新的物体位置。我们对每个任务进行了10次试验,并在每次试验中改变物体的位置。

UR5:采用 Chen 等人[14]的设置,其中 UR5 机械臂执行多个桌面操作任务。观察数据来自单个第三人称摄像机流,动作空间为末端执行器位置变化量。我们评估了两个语言条件反射任务:从一个碗中取出玩具老虎并将其放入另一个碗中,以及用抹布擦拭桌子。虽然这些任务属于分布式任务,但策略仍必须能够泛化到新的物体位置、干扰物体和光照条件。由于训练数据是在几个月前收集的,并且机器人装置被拆卸并重新组装,因此策略还必须能够泛化到环境中的其他各种变化,例如略微不同的摄像机视角和背景。我们对每个任务进行了 10 次试验,并在每次试验中改变物体的位置。

RT-1 机器人:采用 Brohan 等人[10]的设置,其中一台专有机器人执行多项桌面和家具操作任务。观察数据来自单个第三人称摄像机,动作空间为末端执行器位置变化量。我们评估了拾取七喜汽水罐、苹果、蓝莓片袋或棕薯片袋的任务,以及打开或关闭橱柜抽屉的任务。虽然这些任务属于分布式任务,但策略仍必须能够泛化到新的物体位置。此外,由于机器人装置在建筑物之间移动,策略还必须能够泛化到环境中的各种变化,例如略微不同的摄像机视角和背景。我们对每个任务进行了 10 次试验,并在每次试验中改变物体的位置。

B. 模型消融

我们所有的模型消融实验均在 WidowX平台上进行评估。表 VI 详细列出了各项任务的成功率。我们评估了两项语言条件任务（将胡萝卜放在盘子上和将茄子放在花盆里）和两项目标条件任务（将面包放在盘子上和将勺子放在手套上）。目标条件任务包含 Bridge 数据集中未出现过的物体。此外,我们在表 VII 中分析了Octo 模型在多个不同维度上的泛化能力,例如新物体、新环境和新技能。

C. 微调评估

CMU烘焙任务:机器人必须拾取玩具面包,将其放入烤面包机,然后关闭烤面包机。这项任务需要机器人能够泛化初始位置（烤面包机和玩具面包的初始位置）以及目标玩具面包的形状。我们使用末端执行器增量动作空间（笛卡尔坐标位置 + 旋转增量）。观测数据来自第三人称前置Zed摄像头。动作预测频率为15 Hz,并通过R2D2 Franka控制器在机器人上执行。微调数据集包含 120 个演示,这些演示是通过专家 VR 远程操作收集的,每个策略都使用 20 次试验进行评估（4 个新的测试对象,每个对象有 5 个位置）。

斯坦福咖啡:机器人需要从四种不同的Keurig咖啡胶囊中拾取一种,并将其放入Keurig咖啡机中。这项任务既要求机器人能够识别不同初始位置和颜色的咖啡胶囊,又要求其能够精确地将其放入Keurig咖啡机中。我们使用末端执行器增量动作空间,并采用基于Polymetis [49]的开源控制器,运行频率为10 Hz（点击[此处查看](#)）。我们仅使用单次第三人称手腕视角观察。我们的训练数据集包含118个专家演示,这些演示涵盖了不同的咖啡胶囊和位置,并通过VR远程操作收集。我们评估了20个回合的策略,每个回合包含四种不同颜色的咖啡胶囊,每种颜色对应五个回合。

伯克利插销插入任务:该任务是将预先抓取的3D打印插销插入料箱内3D打印板上的匹配插槽中。插销与孔之间的匹配公差为1.5毫米,这使得该任务成为一项接触密集型的精密零件配合任务。机器人必须学习一种合适的策略,通过接触“搜索”匹配的开口,这需要使用力/扭矩测量。该策略的观测空间由单个侧视摄像头图像、末端执行器扭转以及末端执行器的力/扭矩读数组成。当机器人的末端执行器以5赫兹的频率扭转时,该策略会发送动作指令,并通过一个低级阻抗控制器以1000赫兹的频率跟踪该扭转。我们的微调数据集包含来自FMB数据集[53]的100个人类演示。我们评估了训练好的策略在20次试验中的表现,每次试验的板的位置都是随机的。

Berkeley Pick Up:我们采用 Radosavovic 等人的设置。[73]机器人需要从桌面拾取一个方块,此前它已经接受过包含 100 次不同物体拾取的训练数据集。机器人使用关节位置控制,并采用底层 Polymetis 控制栈[49]（[此处](#)）。它以腕部摄像头输入图像以及本体感觉为条件。

		语言条件		目标条件		
		把胡萝卜放在盘子里,把茄子放进锅里,把面包放在盘子里,把勺子戴在手套上。平均				
概	八角形（我们的）	80%	90%	70%	90%	83%
	RT-X 数据集混合[67]	80%	80%	40%	40%	60%
	单机器人数据集（桥梁数据）	20%	70%	60%	20%	43%
微	离散化动作预测[67]	0%	20%	10%	40%	18%
	连续动作预测（MSE）	70%	30%	0%	40%	35%
基	Resnet-50 + 变压器[67]	80%	60%	100%	40%	70%

表六:模型消融。使用 ViT 架构、扩散作用头和宽消融头时,我们获得了最佳性能。训练数据混合。所有评估均在 WidowX 平台上进行。成功率取自两个数据集中 40 次试验的平均值。语言条件反射任务和两个目标条件反射任务。

概括类型	任务	平均成功率	
分销中	将胡萝卜放在盘子里	80%	85%
	将茄子放入花盆	90%	
新奇物品	将面包放在盘子上	70%	80%
	把勺子放在手套上	90%	
新环境	将蘑菇放入锅中	20%	40%
	将勺子放在布上	60%	
新技能	将杯子侧放	10%	5%
	将方块放入插槽	0%	

表七:零样本泛化分析。使用 WidowX 设置,我们分析了泛化能力。Octo 在多个不同维度上都表现出色。Octo 在分布式任务中表现最佳,并且能够很好地泛化到新对象。以及各种环境。然而,Octo 难以泛化到 WidowX 形态中未展现的技能。成功率是取两项任务共 20 次试验的平均值。

读取机器人的数据。我们评估了其拾取任务。进行了 10 次试验,每次试验都将钢笔放置在不同的位置。黄色立方体,进行了 20 次试验。每次试验。

伯克利可乐:一台 Trossen Robotics ViperX 机器人是任务是从桌子上捡起一个可乐罐。ViperX 在训练混合数据集中完全没有出现过,因此这项任务测试模型对新机器人的泛化能力体现。这项任务还需要概括性。ViperX机器人将可乐罐放置在不同的位置采用末端执行器增量控制,频率为 5Hz。该策略使用图像来自第三人称视角和腕式摄像头的观察。我们的训练数据集包含收集到的 115 个专家演示。通过虚拟现实技术从不同位置进行远程操作。伯克利双臂机器人:该任务需要一台 ALOHA 双臂机器人[98],由两个 ViperX 机械臂组成,用于拾取用右手从工作台上拿起一支记号笔用左手取下它的帽子。由于 Octo 只用单臂机器人数据进行了预训练,我们重新初始化一个新的动作。映射到 ALOHA 的 14 维动作空间的头部 (2x 6个关节位置 + 2个夹爪位置)。成功政策需要利用来自左右两侧的图像观察结果。利用腕部摄像头预测精确的操作行为。最佳我们的方法和基线方法均取得了良好的性能。通过在训练期间预测 64 个动作片段,然后测试期间,使用递减地平线执行 12 个动作。在重新规划之前进行控制。我们评估了解除限制任务。