专题 11 / 具身策略 / 条件流匹配

π0：让动作成为条件流

A reading of π0: A Vision-Language-Action Flow Model for General Robot Control

原作：Kevin Black 等，Physical Intelligence · 初稿 2024，底本 arXiv:2410.24164v4（2026-01-08）

> 中英逐节深读。两种语言均为编者按原文证据重新撰写的解读；图号、试验条件与原版页码随段标注。适合与同目录英文论文一起读，不是逐段全文译本。

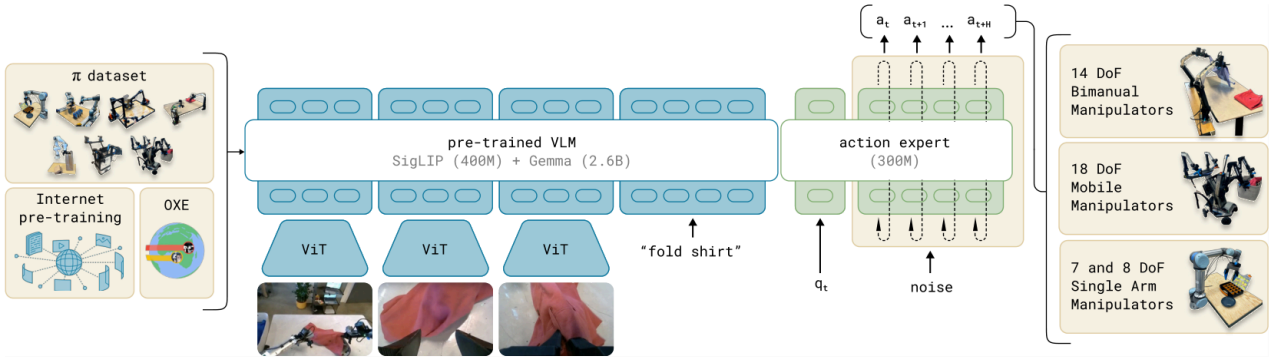
摘要与引言｜机器人究竟在生成什么　模型｜从观测到动作块　数学｜两种时间，一个监督目标
数据｜覆盖状态，再教学策略　实验｜四种能力不能混为一项分数
局限与工程读法｜50 Hz 不等于 20 ms 重规划

摘要与引言｜机器人究竟在生成什么

S001 · p.1–2 同一台机器人既要判断盘子该放哪里，又要在抓取边缘时连续调节手腕。前一问关乎语义，后一问关乎可执行的动作轨迹。π0 用预训练视觉语言模型编码情境，再让动作专家在这一情境下产生整段连续控制量。论文同时考察一种训练顺序：先扩大机器人经验的覆盖面，再用精选示教教会它怎样把一件事做好。

S002 · p.2–4 从一个平台、一个背景和一种成功策略获得的状态有限；示教之外的状态恰恰是部署中需要恢复的地方。通用 VLA 常以离散符号表示动作，专用模仿策略更擅长精细操作，却少有大规模预训练带来的语义先验。本文让文本与连续动作通过注意力交换信息，而不让语言模型逐个输出代表关节位置的离散编码。

模型｜从观测到动作块



S003 · p.4 图 3｜先看数据如何汇合，再看计算如何分工。共享注意力不等于视觉词元和动作词元共用每组权重。

读图 / 左侧示教与开放数据共同参与机器人预训练。中部两条计算路径对应 VLM 主干与动作专家；右侧多机型输出不是未来相机画面。

$$o_t = (I_t^1, \ldots, I_t^n, \ell_t, q_t), \qquad A_t = (a_t, \ldots, a_{t+49}).$$

S004 · p.4–5 在机器人时刻 t，观测包含两或三路 RGB 图像、语言指令和关节状态。策略面向的是条件分布 $p(A_t \mid o_t)$：块中的 50 个动作对应未来 50 个物理时刻。块内位置能相互注意，无需逐维量化，也不需要自回归 50 次才见到整条轨迹。重复计算发生在整块的流场更新上。

数学｜两种时间，一个监督目标

$$A_t^\tau = (1-\tau)\epsilon + \tau A_t, \quad \epsilon \sim N(0, I)$$
$$\partial_\tau A_t^\tau = A_t - \epsilon, \quad L = E\,\|v_\theta(A_t^\tau, o_t, \tau) - (A_t - \epsilon)\|_2^2$$

S005 · p.5 取一条示教动作块和同形的独立高斯噪声，以伪时间 τ 作线性插值。给定两端点，对插值式求导便得到训练速度。一次训练样本只需中间位置和目标速度，不必先运行 ODE 求解器。直线属于给定端点的条件样本；模型只看到中间点和观测，最小二乘解是相容速度的条件均值，不保证实际积分轨迹都直。

$$A_t^{\tau+\delta} = A_t^\tau + \delta\, v_\theta(A_t^\tau, o_t, \tau), \qquad \delta = 0.1.$$

S006 · p.5、16 采样从纯噪声出发，模型执行十步前向欧拉，每步 0.1。t 由机器人时钟决定，τ 只是合成一个动作块的积分变量；步长 0.1 不以秒为单位。附录 B 使用偏向较小 τ 的移位 Beta 分布训练，此时输入噪声较大，动作预测更依赖观测。训练分布的上界取 0.999，而积分网格仍是均匀的。

S007 · p.5 值得校勘的是，v4 同页把条件高斯的协方差写成 $(1-\tau)I$，实际采样公式却把噪声乘以 $(1-\tau)$。由后式计算，给定动作端点后的均值为 τA_t，协方差应为 $(1-\tau)^2 I$。复现时应让插值、速度标签和积分方向自洽。

S008 · p.15–16 注意力划分为图像与文本、机器人状态、噪声动作三个块。每块内部双向，后块可见前块，前块不看后块，所以同一次观测的键值可跨去噪步缓存。PaliGemma 主干约 30 亿参数，新动作专家约 3 亿，总计约 33 亿。小模型约 4.7 亿且结构也不同；两者的差距不能直接解释成 VLM 预训练的净收益。

数据 | 覆盖状态，再教学策略

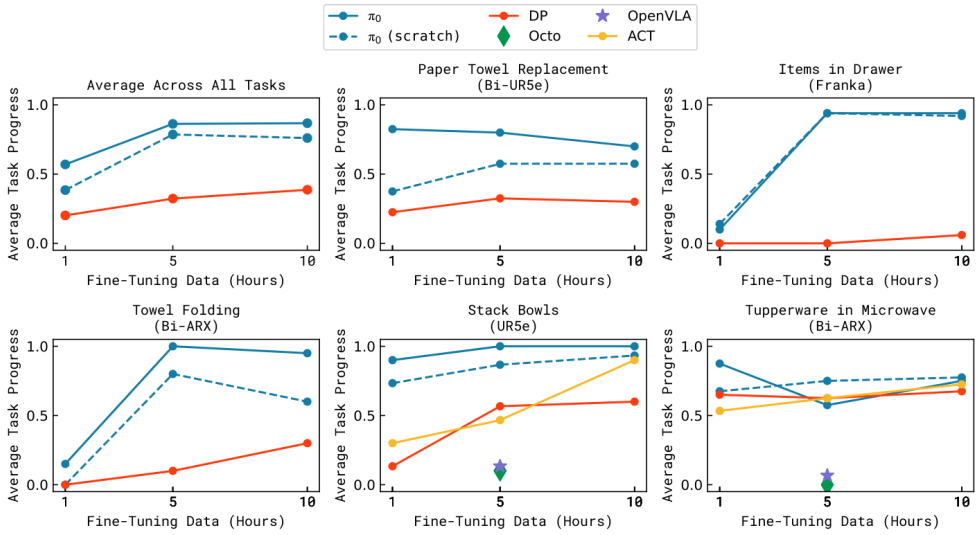
S009 · p.5–6 自采的 9.03 亿个时间步中，单臂占 1.06 亿，双臂占 7.97 亿，分布在 7 种配置、68 个任务。开放数据占训练混合权重的 9.1%，不能读成原始帧数的 9.1%。图 4 区别了样本量与采样权重。任务—机型组合按样本数 $n^{0.43}$ 加权：样本多的组合仍较常出现，但不再按原始数量线性支配训练。

S010 · p.6 跨机型先统一张量接口：关节状态和动作向量补至 18 维，不足三路的相机输入加掩膜。形状统一不代表不同关节的物理含义自动对齐。语言标注既可为任务名，也可指向约两秒的子任务片段。预训练重在覆盖状态，后训练重在示教动作的流畅和一致；只保留“漂亮”的成功片段，可能把部署时需要的恢复行为一并筛掉。

实验 | 四种能力不能混为一项分数

S011 · p.7–8 直接部署考察叠平铺 T 恤、两档清桌、装购物袋和从烤面包机取面包。这五项属于预训练覆盖范围，不是五项新任务的零样本泛化。完整版训练 70 万步；另有 16 万步 π_0 与 16 万步 OpenVLA、32 万步 Octo 对照。完整版总体领先，短训练版也优于这两组基线。训练步数接近，不意味着每步算力、墙钟时间或表达能力相同。

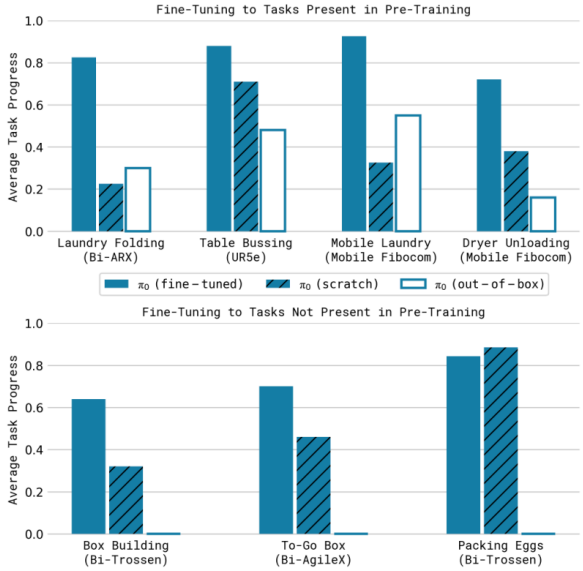
S012 · p.8–9 语言实验区分三种指令：全程只给总目标、人按进度给中间指令、由上层视觉语言策略给中间指令。清桌、摆桌和装购物袋都要求选对物体并按正确次序执行。“抓得住但拿错了”是语义失误；“拿对目标但掉落”则是运动失误。上层提示改善结果，并不说明底层动作专家独自学会了全任务的规划。



S013 · p.9–10 图 11 | 横轴为特定任务的示教小时数，纵轴为按任务细则归一化的平均进度。应在相同小时数、相同任务的纵截面上比较曲线。

读图 / 逐列读 1、5、10 小时；不是每种基线在每个预算上都有点。某任务在 1 小时上领先，不代表 5 小时上的相对差距会更大。

S014 · p.9–10 转移难度并不均匀：叠碗与叠毛巾接近源任务，微波炉引入新物体，抽屉收纳与换纸巾涉及新动作。ACT 和 Diffusion Policy 只用目标示教；Octo 和 OpenVLA 从开放预训练权重微调。同系列 π₀ 在有无机器人预训练时的比较，才更接近对机器人预训练效果的检验。有限次数的试验和不齐全的基线点不足以推导普适的样本效率定律。



S015 · p.11 图 13 | 复杂多阶段任务上的十次试验均分。完整模型在所有任务上均超过该任务最高可能得分的一半。

读图 / 按同一任务比较 out-of-box、scratch 和完整方案。一次未完成的任务也可以取得部分分数，柱高不等于整任务成功率。

局限与工程读法 | 50 Hz 不等于 20 ms 重规划

S016 · p.10–12 长任务包括叠衣、移动叠衣、从烘干机取衣、清理复杂餐桌和折盒，执行可持续 5–20 分钟。有些已见于预训练，另一些仅在后训练出现。两阶段训练对困难任务尤有帮助；但尚不能据此预测为达到近满分可靠性需要多少、何种示教。多种机械臂之间的迁移，也不能外推为对导航、自动驾驶或腿式运动的普适性。

S017 · p.16 在 RTX 4090 上对三相机观测生成一个动作块，耗时 73 ms：图像编码 14 ms、观测前缀 32 ms、十次动作前向合计 27 ms。远程再加 13 ms。50 Hz 平台每 0.5 s 重规划、执行 25 个动作；20 Hz 平台每 0.8 s 重规划、执行 16 个动作。系统能每 20 ms 发出一个已算好的动作，但并非每 20 ms 根据新观测决策一次。

预算对象	原文值	如何解释
图像 / 观测前缀	14 / 32 ms	每次新观测各付一次
动作专家 ×10	共 27 ms	不是每次 27 ms
板上 / 远程	73 / 86 ms	远程含 13 ms 网络开销
50 Hz 平台	25 / 50 = 0.5 s	重规划间隔，不是模型延迟

S018 · p.17 再查附录 E 的分母：简单清桌有 7 件物品，困难版有 12 件；叠衬衫按成功或失败计，其他若干任务按动作阶段得分。均值同为 0.8，在不同任务中可能意味着完全不同的失误。评价自己的控制器时，宜同时保留整任务成功率、分步进度、从传感器到首个新动作的延迟以及实际执行前缀的长度。

原文及所引图表署名 Physical Intelligence / Black 等。arXiv 页面列出的版本许可为非独占分发许可。 arXiv 原文