arXiv:2611.30209v1 [cs.RO] 23 Nov 2026

RoboAda：在测试时使冻结的机器人策略适应其无法感知的物理规律

Yunbei Zhang[1] Zijian Jin[2] Shuaicheng Niu[3] BaianChen[4]WeiruiYe[4] Jihun Hamm[1] Yilun Du[5]
Tulane大学 [2]纽约大学 [3]南洋理工大学 [4]麻省理工学院 [5]哈佛大学**项目页面与代码：** https://RoboAda.github.io



图 1. **RoboAda 推断执行冻结策略规划的机体**。在视觉上无法察觉的 35 牛顿载荷下，冻结策略仍按训练时的方式执行规划，导致奶油奶酪未能到达目标位置（左）。每个动作块结束后，RoboAda 将本体感觉转化为事后标签 ❶，并对明确的机体假设进行加权 ❷（以增益 0.5 为例），然后依据后验分布采取行动 ❸。在十一种物理偏移下，这使五种冻结策略的平均成功率提高了一倍以上（右）。照片为模拟回合的渲染图。

摘要

预训练机器人策略根据相机图像决定采取何种行动，但决定其规划能否成功的某些物理规律是不可见的。物体看起来可能很轻，实际却重 *100* 倍；关节性能下降或控制器滞后也不会在图像中留下任何痕迹。当部署时的物理规律与训练数据中的不同时，规划通常仍然是正确的，但机器人却仿佛什么都没有改变一样执行它，最终无法完成任务。现有方法要么预先针对这类变化进行训练，要么在部署后进行适应。前者无法处理训练中未见过的变化；后者会改变机器人的规划，使校准错误的指令继续存在，或者通过梯度步更新机器人的模型。这类更新并不知道自身有多可靠，因此即使什么都没有改变，也可能让机器人表现得更差。我们提出了 *RoboAda*，它保持策略及其规划不变，并通过推断机器人如何将指令转化为运动来校正执行规划的指令。我们的关键观察是，机器人的自身运动揭示了发生的变化。每次动作之后

动作块中，在原始机器人上训练的逆动力学模型会将观察到的运动转换为原始机器人为实现该运动所需的指令，从而在每一步生成免费的标签。系统根据各种竞争性假设对这些标签的预测效果为其赋予权重，其中包括机器人没有发生任何变化这一假设；随后，系统会发送下一个动作块，并采用在加权假设下最能实现规划的指令。例如，如果机器人的移动距离只有指令要求的一半，*RoboAda* 就会推断机器人能力较弱，并相应加大下一步指令的幅度。*RoboAda* 无需奖励、无需梯度更新步，也无需来自发生变化的机器人的数据。我们在 *LIBERO* 上使用五种冻结策略、两个世界动作模型和三个视觉-语言-动作

模型评估 *RoboAda*。在十一种物理偏移下，它使这些模型的平均成功率从 *27.8%* 提高到 *62.9%*，增幅超过一倍。在 *Cosmos Policy* 上，它比七种测试时自适应方法中表现最佳的一种高出 *17.3* 个百分点；而在情况未发生变化时，它基本保留了策略的成功率。在采用关节空间指令控制的双臂机器人 *RoboTwin 2.0* 上，它将关节目标偏移下的成功率从 *24.9%* 提高到 *78.8%*。

1. 引言

世界动作模型（WAM）和视觉-语言-动作模型（VLA）让预训练机器人策略具备了广泛的能力。WAM 会预测场景将如何演变，并解码出能够产生这一未来的动作：它既可以基于预测的帧 [10, 19, 24, 53] ，也可以通过联合对视频和动作进行去噪 [23, 30, 60, 77]。VLA 则直接将观测映射为动作 [5, 6, 21, 43]。我们将这两类策略输出的动作块称为其规划。这两类模型都是基于训练数据中的物理规律学习规划的，并且执行规划时也默认这些物理规律依然成立。然而，决定规划能否成功的某些物理规律是不可见的。装满东西的马克杯看起来和空马克杯一样，外表像木头的方块可能由铁制成，而磨损的关节或响应滞后的控制器看起来与正常状态无异。在每种情况下，部署时的物理规律都不同于训练数据中的物理规律，而摄像头无法揭示这种差异。

机器人离开实验室后，往往会遇到此类隐蔽的物理偏移：执行器会磨损，不同机器人上的控制器调校各异，待处理物体的重量也不尽相同。这些偏移还会造成高昂代价。在图1中，机器人搬运一盒奶油奶酪时，受到一个隐蔽的35牛顿负载作用。策略仍像在训练机器人上一样规划相同的运动并发送相同的指令，结果奶油奶酪盒没能到达篮子。在 RoboTwin-Phys 上，同时随机化物理条件、桌面几何形状和相机位姿，会使 Fast-WAM 的成功率从 91.8% 降至 44.2%，使 $\pi_{0.5}$ 的成功率从 76.8% 降至 31.6%；而仅进行视觉与布局随机化，成功率最多只下降 6 个百分点 [66]。

现有的补救方法分为两类（表1）。第一类是在训练前准备好策略，要么通过在随机化的动力学条件下训练 [16, 42]，要么根据近期交互推断潜在物理上下文 [26, 27, 61, 65, 75]。这两种方法都只能应对训练期间见过的变化。第二类是在部署时进行适应。其中一些方法会改变机器人的规划或预测，例如根据观测结果调整世界模型 [41, 54, 62, 70] ，或者采样多个规划并保留验证器更青睐的那个 [44, 64, 68],，但规划仍由同一组校准不准的指令执行。另一些方法通过梯度步更新动力学模型 [36, 52] 或指令残差 [49, 50] 。这些更新不会估计自身的可靠性，因此即使一切未变也会进行调整，还可能让状态未变的机器人表现更差。

我们认为，面对这类分布偏移时，规划通常仍然正确，出错的只是执行规划的指令。如果机器人每次执行指令都只完成一半，那么它实际完成的运动也只有 Cosmos Policy 设想运动的一半（图5）。因此，我们将测试时适应视为推断机器人如何将指令转化为运动，并据此校正指令。我们把执行器、控制器以及机器人携带的任何负载统称为机器人的机体；它们共同将指令转化为运动。这一问题很难解决，因为这种分布偏移不可见，测试时没有奖励或标签可用，而且若校正过程追逐执行噪声，状态未变的机器人可能表现更差。

我们提出 RoboAda 来解决这一问题（图1）。每个动作块执行后，使用在标称机器人（即策略训练所用的机器人）上训练的逆动力学模型，反推出标称机器人为了实现实际发生的运动本应发送的指令，这类似于事后重新标注 [1, 15]。如果机器人对每条指令都只执行一半，反推出的标签就是所发送指令的一半，因此每一步都会免费提供一个标签，说明机器人如何响应。对于每个动作轴，RoboAda 会根据这些标签拟合三个竞争性假设：机器人没有变化；机器人会对每条指令进行缩放并添加偏移量；或者机器人的响应取决于其状态，例如抓取时出现负载。每个假设的权重取决于它对每个新标签的预测效果；发送下一个动作块时，则采用在这些权重下最能实现规划的指令。系统性偏移会在几个动作块内降低“机器人没有变化”这一假设的权重，而纯噪声则会使权重保持分散，因此指令不会追逐噪声。预先学习的只有这些假设的先验和特征，且仅使用标称机器人数据。拟合与加权遵循多模型自适应控制 [35, 37], ；我们将其应用于一个冻结的基础策略周围，以该策略自身的规划为目标，且不需要机器人模型或奖励。我们使用 LIBERO 上的五种冻结策略评估 RoboAda [32], , 包括两个 WAM 和三个参数规模从 2B 到 7B 的视觉-语言-动作模型；还在双臂 RoboTwin 2 上使用 Fast-WAM 进行评估。0 [7]。对所有策略采用相同设置时，在十一种物理偏移下，其平均成功率从 27.8% 提升至 62.9 %，增幅超过一倍（表2）。我们的贡献有三点：

• **测试时适应即推断机体**。我们将针对未见过的物理规律对冻结机器人策略进行适应，重新定义为：在策略自身的动作空间中，根据机器人自身运动的事后标签，推断关于机器人机体的显式假设的后验分布。该方法不需要奖励、重新训练，也不需要来自发生变化的机器人的数据。

• **RoboAda**。 我们根据后验分布下的预测证据与指令对机体假设进行加权，其先验是在标称机器人的合成变化上通过元学习获得的。不同于潜在上下文推断只能识别训练中见过的变化，RoboAda 根据机器人自身的运动拟合当前机体，无需来自

表 1. **只有 RoboAda 能在每种分布偏移下提升表现，同时不损害未发生分布偏移的机器人。**测量列标示了在 Cosmos Policy 和使用基线方法测试的全部四种策略上，各类分布偏移下相较于冻结策略的增益，以及未发生分布偏移时没有损害，即至多比冻结策略低 5 个百分点。不确定性：明确的身体后验分布。成功率：Cosmos Policy 在全部分布偏移下的成功率（冻结策略为 33.0%）。–：不适用。

	设计				测量值					
方法	冻结策略	测试时适应	无梯度	不确定性	驱动	接触/负载	关节级	全部策略	无损害	成功率（%）
RobustVLA [16]	✗	✗	–	✗	–	–	–	–	–	–
RMA 风格 [26]	✓	✓	✓	✗	✓	✓	✓	✓	✗	44.8
AdaJEPA [56]	✓	✓	✗	✗	✓	✓	✓	✓	✗	39.8
JEPA-TTT [70]	✓	✓	✗	✗	✗	✓	✓	✓	✗	45.5
GrBAL [36]	✓	✓	✗	✗	✗	✓	✓	✓	✗	52.3
TTT层 [52]	✓	✓	✗	✗	✗	✓	✗	✗	✗	30.2
ALRA [49]	✓	✓	✗	✗	✓	✓	✓	✓	✗	46.8
Sandwich-Residuals [50]	✓	✓	✗	✗	✗	✓	✓	✗	✗	42.5
RoboAda（我们的方法）	✓	✓	✓	✓	✓	✓	✓	✓	✓	**69.6**

变化后机器人的数据。不同于世界模型适应和基于验证器的引导，它改变的是规划的执行方式，而不是选择哪个规划。

• **在五种策略和两个模拟器上的证据**。在十一种物理分布偏移下，RoboAda使平均成功率提高了一倍以上；对于重量比外观所示重100倍的物体，它恢复了因物体差异而损失的大部分成功率；在感知、语言和世界变化下，它仍保持增益；其性能超过七种测试时自适应方法，并迁移到通过关节空间指令控制的双手机器人上。其后验分布在四到六个动作块内恢复了真实的指令增益。

2. 相关工作

世界动作模型与推理时引导。基于视频的策略在像素空间中进行规划，并通过逆动力学或跟踪恢复动作 [10, 11, 19, 24, 53]。联合模型在一个网络中对视频和动作进行去噪 [17,23, 29, 30, 60, 77], , 混合视觉语言动作模型则为动作预测增添前瞻能力 [67, 72, 74]。多项研究利用这类模型在推理时构建行为。它们在预测性世界模型中为策略提案排序 [44], 以视觉语言模型作为提议者和评判者来搜索末端执行器目标 [68], 从校准回合中推断设置的动作到像素映射 [59], 在失败后优化潜在物理嵌入 [25], 用逆模型验证世界模型 [34], 或在线学习外部力的内部模型 [14]。WAM 的测试时选择会借助验证器、意外度、几何信息或逆一致性，对想象出的未来进行评分 [47, 48, 64, 73]。与这些选择或优化规划内容的方法不同，RoboAda 推断执行规划的机体，并且可以与其中任意方法互补。

测试时适应与物理推断。测试时适应会利用无标签测试数据更新模型 [38,51, 55]。对于视觉-语言-动作模型和 WAM，这种方法主要由未来预测损失 [13, 20, 41, 54, 78] 以及测试时强化学习 [4, 33]推动。鲁棒性基准测试会改变策略所看到和接收到的信息 [12, 76], , 改变策略执行时所处的世界 [69], , 或改变物理条件 [28, 66], 而在这些变化中，只有物理变化无法通过视觉察觉。人们已通过在线系统辨识 [61],、探测性交互 [75]以及上下文感知动力学模型 [27], , 从交互中推断隐藏的物理参数；RMA、其面向机械臂的版本，以及自适应视觉-语言-动作模型，则从近期历史中推断潜在上下文 [26, 31, 65]。它们都在日后要辨识的各种变化下接受训练，动力学随机化 [42]和 RobustVLA [16]也一样。更接近我们设定的方法包括 JEPA-TTT 和 AdaWorldPolicy，它们根据想象与观测到的未来之间的差异更新前向模型或世界模型 [62, 70], ; 一种自我补偿视觉-语言-动作模型会更新其动作专家上的低秩适配器 [28], ; FADA 在线下利用目标域轨迹微调冻结的人形机器人规划器的逆动力学模型 [58], ; APEX 则调整低级控制器执行策略目标的方式 [71]。与这些方法不同，RoboAda 既不更改策略，也不更改世界模型，而且不进行梯度更新。它会维护一个涵盖显式身体模型（包括名义身体模型）的后验分布，并且不需要奖励、数据或发生变化的机器人的模拟器。

自适应控制。对于每个轴，RoboAda 使用带遗忘因子的递归最小二乘法拟合其假设 [2], , 通过温度缩放的折扣序贯证据为这些假设赋权 [9, 45] , 并按照多模型自适应估计与控制的传统发出谨慎指令 [57], [3, 35, 37]。其事后标签减去已发送的指令，便是扰动观测器的估计值 [8, 40], ; 该观测器以学习得到的逆模型取代解析式被控对象。与元学习自适应控制一样 [18, 39, 46], , 其特征和先验也是通过元学习获得的。与这一领域的相关研究不同，这里的被控对象是冻结的基础策略动作接口背后的黑盒机器人系统栈，而该策略自身的规划则作为参考。

3. 方法

3.1. 问题表述

冻结策略将观测 o_t 和指令 ℓ 映射为一组 K 规划指令 $p_{t:t+K-1}$。在 LIBERO 上，这些指令是六个机械臂轴的末端执行器运动以及

表 2. **RoboAda 在 LIBERO 的每种分布偏移下均提升了五种冻结策略的表现。** 在相同回合中，未使用和使用 RoboAda 时的成功率（%）。各列分别归类了 RobustVLA 的动作扰动、我们的物理偏移和关节级偏移，以及在三个基准测试的感知、语言和世界变化上叠加的物理偏移。RobustVLA†: π_0 重新训练以增强鲁棒性 [16]。–: 未报告。

	RobustVLA 协议				物理偏移			+感知 / 世界分布偏移		
	动作（5）	系统性零均值	推力	物理（11）	关节级（5）	物体 ×100	LIBERO-Plus	LIBERO-PRO		LIBERO-MAX
Cosmos Policy [23] (WAM, 2B)										
冻结	64.4	72.7	58.9	45.0	25.1	38.3	58.0	21.4	17.7	12.4
+ RoboAda	**74.9**	**93.0**	**62.8**	**57.1**	**73.2**	**75.3**	**76.5**	**62.6**	**46.9**	**42.4**
Fast-WAM [63] (WAM)										
冻结	51.6	51.6	51.7	40.5	23.5	25.7	55.3	16.1	13.1	8.5
+ RoboAda	**62.5**	**74.4**	**54.5**	**46.0**	**51.9**	**46.8**	**64.5**	**34.4**	**23.3**	**18.4**
π_0 [5](视觉-语言-动作，3B)										
冻结	45.0	43.5	46.0	28.5	28.8	38.3	60.3	16.3	17.7	12.3
RobustVLA†	53.3	56.1	51.5	**40.8**	–	–	–	–	–	–
+ RoboAda	**65.8**	**83.0**	**54.3**	34.8	**55.7**	**61.3**	**73.8**	**38.9**	**39.1**	**27.7**
$\pi_{0.5}$ [43](视觉-语言-动作，3B)										
冻结	62.8	64.0	62.0	38.5	35.0	45.7	63.0	32.4	23.9	21.8
+ RoboAda	**75.9**	**88.2**	**67.7**	**50.0**	**71.2**	**75.7**	**72.0**	**64.3**	**46.6**	**49.1**
OpenVLA-OFT [22](视觉-语言-动作, 7B)										
冻结	46.6	48.6	45.3	17.0	26.8	33.7	58.5	19.4	15.5	12.6
+ RoboAda	**66.8**	**87.0**	**53.2**	**19.8**	**62.3**	**66.3**	**75.0**	**59.9**	**44.3**	**40.2**

夹爪；在 RoboTwin 上，它们是绝对关节目标（第 D 节）。对于策略训练时所用的机体，发送$x = p$ 会产生策略预期的运动。部署时，机体发生了视觉上不易察觉的变化，因此相同的指令会产生不同的运动，尽管规划通常仍然适用。RoboAda 保持策略不变，并将 $x = p$ 替换为针对机器人当前机体选择的指令 $x^\star$。它只需要规划动作块和机器人的本体感觉，因此适用于任何具有此接口的策略。

概述。 RoboAda 在冻结策略周围运行一个循环（图1）。每个动作块之后，事后标签会将实际实现的运动转化为标注对（第 3.2 节）。机体假设的后验分布会将每个假设与这些标注对进行拟合，并根据其预测证据为其赋予权重（第3.3 节）。后续指令会在该后验分布下选定（第 3.4 节）。只有先验是在标称机器人上预先学习的（第3.5 节）。

3.2. 事后标签

变化后的机体无法被直接观测，但其影响可以。一个在标称机器人上训练一次的逆动力学模型 h，会将步骤 s 前后的本体感觉 q 映射为标称主体为实现已发生的运动所需的指令，$\nu_s = h(q_{s-1}, \ldots\ldots, q_{s+2})$。在标称主体上，$\nu_s \approx x_s$。在变化后的机体上，$\nu_s$ 与实际发送的指令 x_s 之间的关系描述了策略自身动作空间的变化。执行器变弱会产生$\nu_s \approx 0.5\,x_s$，恒定扰动会产生偏移量，而负载则会产生一个在抓取时出现的偏移量。因此，每个已执行的步骤都会免费产生一对带标签的数据（x_s，ν_s），无需奖励，也无需演示。

3.3. 机体假设的后验分布

单一的拟合系统体会追逐执行噪声，也无法表达自身有多确定。因此，RoboAda 为每个动作轴 j保留三个显式假设：(i) 名义主体$H_0\,\nu_{s,\,j} = x_{s,\,j}$；(ii) 仿射机体 $H_1\,\nu_{s,\,j} = a_j x_{s,\,j} + b_j$，用于描述增益变化和偏移量；以及 (iii) 状态依赖身体 $H_2\,\nu_{s,\,j} = a_j x_{s,\,j} + \mathbf{b}_j^\top [1, \psi(z_s)]$, ，其状态特征 $\psi\,z_s$（末端执行器状态、动作块内的阶段、近期指令和残差）能够捕捉负载和弱关节等因素造成的影响。

拟合。 对于某个假设在单个轴上的回归量（ϕ_s用于 H_1 的 $[x_s, 1]$、 $[x_s, 1, \psi\,(z_s)\,]$用于 H_2）以及一个以名义主体为中心的高斯先验$\mathcal{N}$（μ_0，Λ_0^{-1}），其后验分布为$\mathcal{N}$（μ，A^{-1}），其中

$$A = \beta \sum_s \gamma^{\,\mathrm{age}(s)} \phi_s \phi_s^\top + \Lambda_0,$$
$$\mu = A^{-1}\big(\beta \sum_s \gamma^{\,\mathrm{age}(s)} \phi_s \nu_s + \Lambda_0 \mu_0\big), \tag{1}$$

其中 age(s) 表示自标签 s 以来经过的数据块数量，因此旧标签会以速率 γ 被遗忘，而标签噪声 $1/\beta$ 则通过在线推断得出。该闭式更新在 CPU 上运行。

预测证据。 在用于拟合之前，每个假设 m都会根据其当前后验分布$\mathcal{N}$（$\mu_{m,\,s,\,j}$，$\sigma^2_{m,\,s,\,j}$）对其进行预测。每个数据块的预测对数似然会累积为逐轴证据，

$$L_{m,j} \leftarrow \lambda L_{m,j} + \tau \sum_s \log \mathcal{N}\big(\nu_{s,j}; \mu_{m,s,j}, \sigma^2_{m,s,j}\big), \tag{2}$$

以及假设 m 在轴 j 上的权重为$\pi_{m,\,j}$ =softmax$_m$ $L_{m,\,j}$，每 $\lambda = 0.6$ 进行一次证据遗忘

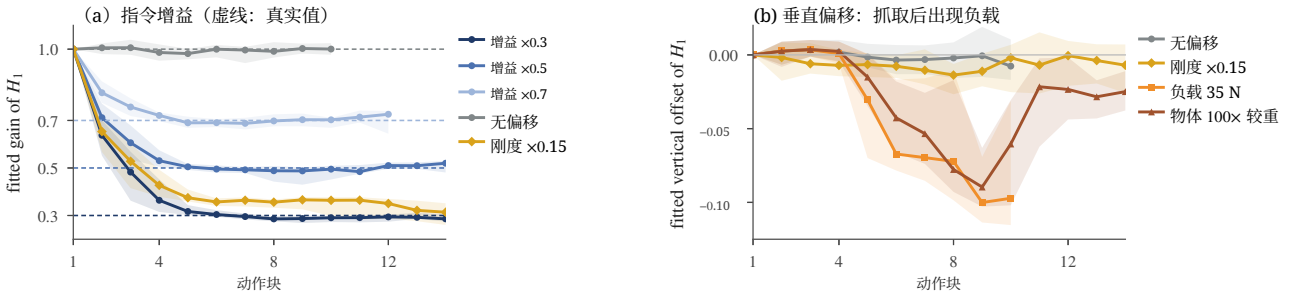


图2. **RoboAda 在几个动作块内恢复隐藏系统体。** Cosmos Policy 上每个动作块的拟合系统体（每种条件下 40 个留出回合的中位数和四分位距）。(a) 拟合增益在四到六个动作块内达到真实增益（虚线）。(b) 负载或 100×较重物体在抓取后表现为向下偏移，而更柔和的控制器则不会。

动作块与温度调节 $\tau = 0.25$。假设只有在预测出尚未见过的标签时才会获得权重，因此，灵活的机体无法通过拟合噪声取胜。这样一来，这些权重π 就成为 RoboAda 对机体的置信度。系统性变化会在几个动作块内使其偏离名义主体（图 1 和 2），而零均值执行噪声则使其保持分散。

3.4. 基于后验分布行动

规划说明名义主体应当做什么，因此指令应让当前机体做到这一点。对于每个规划步骤，指令会最小化后验分布中各机体的预期跟踪误差，

$$x^{\star}_{s,j} = \arg\min_{x} \sum_{m} \pi_{m,j}\, \mathbb{E}_m\left[(\nu_m(x) - p_{s,j})^2\right], \qquad (3)$$

其中 $\nu_m(x)$ 是假设 m对指令 x 预测的标签，而 $\mathbb{E}_m$ 则是在其后验分布上的期望。对于仿射假设，这具有闭式解（第A节）。因此，当拟合系统体存在不确定性时，校正幅度会相应缩小；若没有变化，每个拟合系统体都会保持接近名义主体，从而使规划几乎原样通过。对于延迟，一旦识别出延迟，延迟测试便会跳过每个动作块中过时的部分，并通过一次性追赶来消除此前累积的漂移。

3.5. 来自名义机器人的先验

测试时的变化并未被提前学习。唯一经过学习的，是如何进行适应；这项学习只进行一次，并在线下完成。特征映射 ψ、先验精度 Λ_0、基础标签噪声和遗忘 γ，都是通过对可微闭环代理模型的每一次闭式求解进行反向传播而完成元学习的。在这个代理模型中，记录下来的名义指令序列代替策略的规划；适配器则驱动随机合成机体，这些机体具有增益、偏移量、滞后、死区、饱和、依赖状态的扰动、机体切换、名义机体以及零均值执行噪声（第A节）。使用的数据仅来自名义机器人，策略从未更新，所有超参数均在评估前，使用开发回合固定。

为何采用显式假设。 与使用学习得到的潜在上下文或单个适应模型不同，对少数几个显式身体模型保留后验分布会带来三方面影响。首先，后验分布是可读的：拟合得到的增益和偏移量说明了发生了什么变化，并可与真实变化进行核对（图 2）。其次，名义主体仍作为一种假设，因此由证据而非阈值来决定偏离它的程度。第三，拟合的是一个新的主体，而不是对其进行识别，因此这种变化无需与部署前见过的任何情况相似；元学习先验只让拟合更具样本效率，并更加审慎。

4. 实验

我们回答五个问题。（Q1）对于不同的策略系列（第 4.2 节），RoboAda 能否弥补冻结策略因无法感知的物理变化而损失的成功率？（Q2）它与测试时自适应方法相比表现如何（第 4.3 节）？（Q3）为什么要保留显式假设的后验分布，而不是采用一个适应后的机体（第 4.4 节）？（Q4）后验分布能否识别真实变化（第 4.5 节）？（Q5）RoboAda 能否迁移到感知、语言和世界变化，以及另一种机器人和动作空间（第 4.6 节）？

4.1. 实验设置

策略。我们使用 LIBERO [32]（包含四个任务套件的 40 项任务）和五种冻结策略，采用其公开的 LIBERO 检查点及推理设置：WAM 系列中的 Cosmos Policy [23]（2B）和 Fast-WAM [63], ，以及视觉-语言-动作模型 π_0 [5]、$\pi_{0.5}$ [43]（3B）和 OpenVLA-OFT [22]（7B）。逆动力学模型和先验均使用标称条件下的 Cosmos Policy rollout 进行训练，并在所有五种策略中保持不变。

分布偏移。 我们在以下情形下进行评估：(i) RobustVLA 协议 [16], 通过系统性地对已执行动作施加——

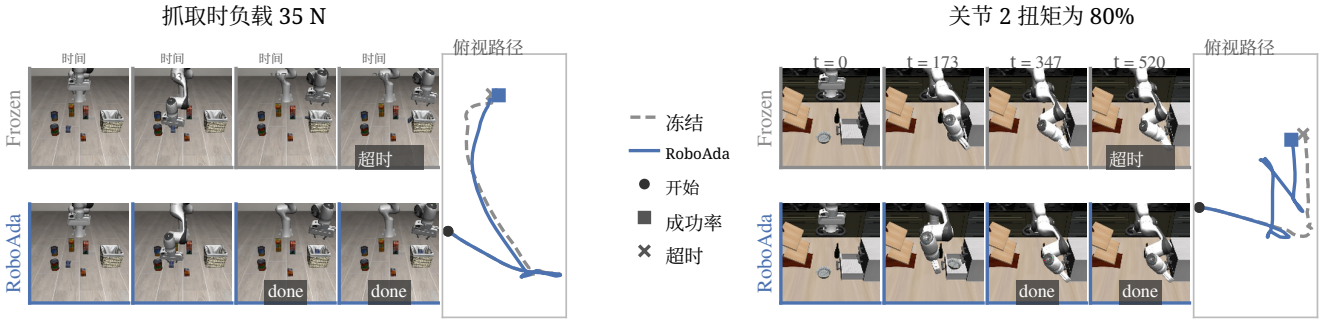


图3。**冻结策略规划出了正确路线，但无法完成任务**。Cosmos Policy 在相同初始状态下，分别搭载 35牛顿载荷（左）以及关节 2 扭矩为 80%（右）时的表现。冻结策略分别在 280 步和 520 步后超时，而 RoboAda 分别在 167 步和 217 步完成任务。

扰动（均匀偏移、偏差）或零均值扰动（高斯噪声、随机翻转、尖峰），或推动夹爪一次；(ii) 对机器人施加十一种物理偏移，即命令增益 $\times 0.3$ 和 $\times 0.5$，在 $\times 1$ 到 $\times 0.3$ 之间于回合中途骤降、控制器刚度 $\times 0.15$、六步延迟、恒定偏移和逐回合偏移、20牛顿向下作用力、抓取时承受35牛顿载荷，以及两个弱化关节；(iii) 从 RoboStress 重新实现的五种关节级退化 [28]，；(iv) 外观与之前相同、但重量增加100倍的物体；以及 (v) 在 LIBERO-Plus [12]和 LIBERO-PRO [76] 的感知与语言变化，以及 LIBERO-MAX [69]的执行中途环境变化上，额外加入四种物理偏移。机器人采取行动前，这些物理变化在视觉上均不可察觉（第B节）。

实验方案。所有方法均在相同回合上运行，开发集与留出初始状态互不重叠，并且所有设置均在留出测试运行前确定。

4.2. 策略无法感知的物理变化

主要结果。 RoboAda 在每一类分布偏移下都能提升所有策略（表 2）。在我们的物理偏移下，它使五种策略的平均成功率提高了一倍以上，从 27.8% 提升至 62.9%（每种策略 4,400 个配对回合）；在相同设置下，7B OpenVLA-OFT 的恢复效果不逊于规模较小的策略。由他人设计的关节级性能退化也呈现出相同趋势，平均成功率从 36.3% 提升至 65.1%。在 π_0上，面对动作扰动，RoboAda 无需任何训练便优于为提升鲁棒性而重新训练的 π_0。这些增益源于动力学的系统性变化，此时后验分布会偏离名义主体。在零均值扰动下，增益较小；没有扰动时，RoboAda 的表现与冻结策略相差不超过 4.0 个百分点。在两台 WAM 上，适配器使每步实际运行时间增加约 1–2%。图 3 展示了两个回合：冻结策略规划了正确路线，但其指令未能及时执行。

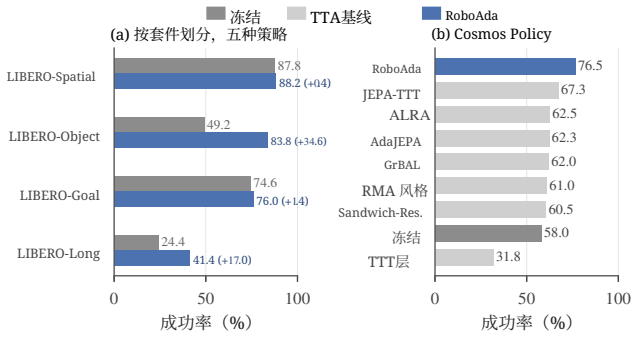


图4。**比看起来更重的物体**。物体的重量增加了 100×，但外观保持不变。（a）按 LIBERO 套件统计的成功率，取五个策略的平均值。（b）Cosmos Policy 与七种测试时适应方法在相同 400 个留出回合上的表现。RoboAda 在冻结策略失败的场景中增益最大，并且优于所有基线。

看起来不重的物体其实更重。 看起来像木头的方块，实际可能由铁制成。我们保持每个物体的外观不变，同时按比例缩放其质量和惯性。当 LIBERO 的质量放大到 100 倍时，Cosmos Policy 的成功率从 98.8% 降至 58.0%，而 RoboAda 将其提升至 76.5%（400 个配对的留出回合）。综合五种策略，平均成功率从 59.0% 提升至 72.4%。当较重的物体导致策略失败时，增益最大（图 4a）。对五种策略取平均后，在 LIBERO-Object 上，成功率从 49.2% 提升至 83.8%；该基准中的每项任务都要求将物体抬起并放入篮子。在 LIBERO-Long 上，成功率则从 24.4% 提升至 41.4%。而在 LIBERO-Spatial 和 LIBERO-Goal 上，冻结策略在大多数回合中仍能成功，RoboAda 则保持了它们的成功率。当 LIBERO 的质量放大到 200 倍时，RoboAda 仍将 Cosmos Policy 的成功率从 36.8% 提升至 45.0%（表 11）。

4.3. 与测试时适应方法的比较

我们在相同接口下重新实现了七种测试时适配方法，并使用相同的事后标签、延迟测试和追赶机制，使它们的差异仅在于机体的建模与适配方式（第 C节）。这些方法涵盖以下路径：

表 3. **RoboAda 优于七种测试时自适应方法。** Cosmos Policy 在留出初始状态上的成功率（%）。驱动：增益 $\times 0.5$、增益下降、刚度 $\times 0.15$。接触/负载：20 N 的力、35牛顿载荷、两个弱化关节。RoboAda 领先 17.3 个百分点，且在无偏移时达到与冻结策略相同的表现。

方法	驱动	接触/负载	全部分布偏移	无偏移
冻结策略	42.1	26.3	33.0	**95.0**
TTT层 [52]	21.3	36.9	30.2	47.5
AdaJEPA [56]	46.3	35.0	39.8	93.8
Sandwich-Residuals [50]	39.6	44.7	42.5	87.5
RMA 风格 [26]	57.1	35.6	44.8	90.0
JEPA-TTT [70]	35.0	53.4	45.5	86.3
ALRA [49]	56.7	39.4	46.8	**95.0**
GrBAL [36]	38.8	62.5	52.3	81.3
RoboAda（我们的方法）	**73.8**	**66.6**	**69.6**	**95.0**

第 1节：从历史中推断潜在上下文（RMA 风格）；通过梯度步适配前向模型和世界模型（Ada-JEPA、JEPA-TTT、GrBAL、TTT层）；以及残差适配器（ALRA、Sandwich-Residuals）。

我们的物理偏移。 在 Cosmos Policy 上，RoboAda 在两组分布偏移中都是最佳方法，并且在全部分布偏移上以 17.3 个百分点领先于最佳基线 GrBAL（69.6% 对比 52.3%，表 3）；与 ALRA 一样，它是唯一能够在发生分布偏移时保持冻结策略无分布偏移下成功率的自适应方法。该排名在另外三种策略上同样成立（表 9）。RoboAda 在 $\pi_{0.5}$上领先最佳基线 17.2 个百分点（70.7% 对比 53.5%），在 OpenVLA-OFT 上领先 18.3 个百分点（63.0% 对比 44.7%），在 Fast-WAM 上领先 12.1 个百分点（48.4% 对比 36.3%）。在 Cosmos Policy 上，它在另外四组留出初始状态中领先 14–20 个百分点；在物体重量增加 100 倍时，领先 9.2 个百分点（图 4b）。

来自独立台架测试的关节级退化。 为了测试我们未设计过的分布偏移，我们重新实现了 RoboStress 的关节级退化 [28]，其代码尚未发布。在每个关节中注入加倍或发生漂移的 Stribeck 摩擦，以及 20–40% 的重力补偿误差；这些退化可能持续存在，也可能仅在抓取时出现，并以关节力矩的形式施加在未改变的控制器下方（第 B 节）。在其他实验均采用相同设置的情况下，RoboAda 将 CosmosPolicy 的成功率从 38.3% 提升至 75.3%（600 个成对回合），并且平均表现领先于全部七种基线方法（表 4）。另外四种策略提升了 21–33 个百分点；在 $\pi_{0.5}$ 和 OpenVLA-OFT 上，RoboAda 再次领先于全部七种基线方法（表 9）。随着摩擦增大，增益也随之提高；当摩擦为 RoboStress 的四倍时，增益达到 38.7 个百分点。

4.4 为什么需要显式假设

表 5 分别单独运行每个身体假设，使用相同的标签、延迟测试和追赶机制，并在相同的回合上进行测试。在系统性扰动和物理偏移下，每种单一身体都能提升成功率；但在零均值噪声下，成功率会下降，因为

表4。**RoboStress 中的关节级退化。**Cosmos Policy 在留出初始状态上的成功率（%）。摩擦从 $\times 1.5$ 增至 $\times 2$（漂移），或增加一倍（老化）；抓取期间或全程关闭重力补偿，而完整组合则同时包含摩擦和重力误差。RoboAda 的平均成功率最高，且在无分布偏移时与冻结策略的差距不超过 2 个百分点。

方法	漂移	老化	抓取	重力	完整组合	平均值	无偏移
冻结策略	46.7	28.3	69.2	40.8	6.7	38.3	**97.5**
测试时训练层 [52]	34.2	35.8	40.0	43.3	21.7	35.0	46.7
AdaJEPA [56]	56.7	33.3	76.7	49.2	5.8	44.3	95.0
Sandwich-Residuals [50]	64.2	44.2	75.0	45.8	15.8	49.0	87.5
RMA 风格 [26]	60.8	42.5	76.7	53.3	6.7	48.0	92.5
JEPA-TTT [70]	60.0	56.7	71.7	55.8	30.0	54.8	84.2
ALRA [49]	60.0	52.5	71.7	50.8	11.7	49.3	95.0
GrBAL [36]	71.7	72.5	68.3	68.3	**62.5**	68.7	75.8
RoboAda（我们的）	**80.8**	**73.3**	**94.2**	**73.3**	55.0	**75.3**	95.8

表5. **单一身体假设** 对比**后验分布。** Cosmos Policy 在相同回合上的成功率（%）。 H_1，手： H_1 采用手动设定的先验。 H_1 和 H_2：仅使用仿射机体或状态依赖机体。每种变体在物理偏移下都有帮助，但只有后验分布在零均值噪声和推力下也能提升。

		单一身体			
扰动	冻结	H_1，手	H_1	H_2	RoboAda
动作（5 次的平均值）	64.4	60.6	68.2	55.6	**74.9**
系统性	72.7	93.9	**94.1**	92.1	92.9
零均值	58.9	38.3	50.9	31.2	**62.8**
Push	44.0	32.8	48.7	23.5	**61.0**
物理偏移（11）	26.7	74.9	75.6	**76.8**	75.6

它会拟合噪声，并将其传递给指令。只有证据加权后验在这种情况下也能提升成功率，而且面对推力时表现显然最佳。权重的表现符合预期：在系统性扰动下，它们会转向学习得到的身体；在零均值噪声下，它们则保持接近均匀分布；而在推力到来前，它们会维持名义主体。改变留出初始状态上的$\tau \in [0.1, 0.5]$ 或 $\lambda \in [0.3, 0.9]$，成功率最多只会变化五个百分点。

4.5. 后验分布揭示了什么

后验分布能够识别机体。当指令增益为 0.3、0.5 或 0.7 时，拟合仿射增益会在四到六个动作块内达到真实值；没有分布偏移时，它则保持在 1.0（图2a）。更柔和的控制器表现为较低的增益，而负载或较重的物体则表现为向下的偏移量，且只有在物体被拿起后才会出现（图2b）。机器人行动之前，这些变化在视觉上都无法察觉，但机器人的自身运动会在几个动作块内揭示每一种变化。

与预言机的差距。如果预言机已知真实的增益、偏移量和延迟，RoboAda 在 RobustVLA 的系统性扰动上弥合了冻结策略与预言机之间 74% 的差距（从 74.5% 提升至 91.0%，预言机为 96.8%）；在我们的仿射偏移上弥合了 91%（从 29.2% 提升至 77.9%，预言机为 82.9%）。

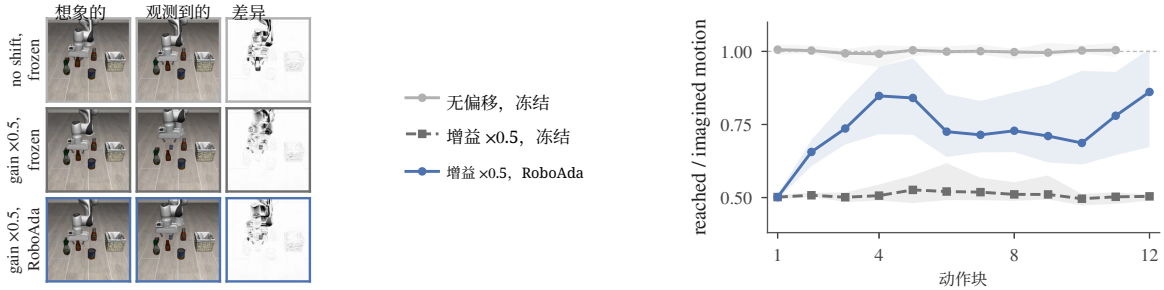


图 5。**WAM 的想象结果记录了机体变化**。左：Cosmos Policy 的一个留出回合，其中展示了为动作块 4 结束时想象的帧、观测帧及二者的差异。右：每个动作块实际达到的末端执行器运动，占想象运动的比例（每种条件下 80 个留出回合的中位数和四分位距）。在指令增益为 0.5 时，冻结机器人达到想象运动的一半，而 RoboAda 在四个动作块内恢复了其中的大部分。

WAM 的想象能够感知变化。动力学变化在机器人采取动作之前无法被观察到，但一旦机器人行动，变化便会显现。每个动作块中，WAM 都会想象摄像头帧和机器人在动作块结束时预期的状态。在冻结的 Cosmos 策略滚动轨迹中，想象与观测到的潜变量之间的差异，能够区分发生分布偏移与未发生分布偏移的回合：对于增益、刚度、力、四步或更长的延迟，以及 0.05 或更大的偏移量，曲线下面积（AUROC）为 0.97–1.00；对于仅在抓取后才起作用的负载，AUROC 为 0.78–0.83。想象出的机器人状态让这种变化具体可见（图 5）。当指令增益为 0.5 时，冻结策略控制下的机器人只完成了 WAM 想象动作的一半，而 RoboAda 在四个动作块内便恢复了其中大部分。之后剩余的差距主要出现在幅度最大的规划动作中，因为增益减半可能会要求发出超出动作范围的指令。RoboAda 本身依据本体感觉获取标签，这种方式也适用于没有想象能力的视觉-语言-动作模型。

4.6. LIBER 上超越物理偏移 O

感知、语言与世界变化。LIBERO- Plus 和 LIBERO-PRO 会改变策略所看到和接收到的内容：摄像头、光照、背景、纹理、布局、物体形状、遮挡、传感器噪声、指令和初始状态。加入四种物理偏移（延迟、负载、刚度和老化摩擦）后，RoboAda 将五种策略在 LIBERO-Plus 上的平均成功率从 21.1% 提高到 52.0%，在 LIBERO-PRO 上从 17.6% 提高到 40.0%（表 2）。LIBERO-MAX 会在机器人执行动作时改变世界。当夹爪靠近物体时，目标或容器会被移动，障碍物或干扰物会出现，或者摄像头、光照、视觉主题或传感器噪声会发生变化。加入四种物理偏移后，RoboAda 将平均成功率从 13.5% 提高到 35.6%（表 12）。没有物理偏移时，在这些基准测试上，其表现与冻结策略相差不超过 3.2 个百分点，因为它只改变规划的执行方式，不会改变感知和语言部分。

表 6. **RoboTwin 2.0 上的双手操作**。 Fast-WAM 在八项任务上的成功率（%）。每回合从 $U(-c,\ c)$ 抽取一次固定偏移量或逐关节偏移量，并将其加到机械臂关节目标上。RoboAda 将平均成功率从 24.9% 提升至 78.8%。

	固定偏移量（弧度）			**每回合偏移量**		
	$+0.03$	$+0.05$	-0.05	$c = 0.03$	$c = 0.06$	**全部五项**
冻结的 Fast-WAM	22.5	4.4	14.4	52.5	30.6	24.9
+ RoboAda	**75.6**	**78.1**	**81.9**	**77.5**	**80.6**	**78.8**

另一种模拟器、机器人和动作空间。LIBERO 的机器人是由末端执行器运动驱动的单臂机器人。RoboTwin 2.0 [7] 在 SAPIEN 中模拟双臂 Aloha-AgileX 机器人；Fast-WAM 的公开 RoboTwin 检查点则为每条机械臂的六个关节和夹爪指令绝对目标，模拟器通过规划轨迹和刚性、重力补偿关节驱动来跟踪这些目标。在关节空间中，名义逆模型是对每个关节实际关节增量进行仿射校准，并利用 53 条名义轨迹拟合；这些轨迹的初始状态均未用于任何评估。指令则是目标位置减去测得的关节位置。第三个假设采用与负载相关的增益和偏移量，而非学习到的特征（第 D 节）。我们对执行的机械臂关节目标施加偏移，这是 RobustVLA 系统性扰动在关节空间中的对应形式。施加 0.03 或 0.05rad（1.7 至 2.9 度）的固定偏移量——机器人移动前肉眼无法察觉——会使冻结策略在三个固定偏移量下的平均成功率降至 13.8%；RoboAda 则将其恢复至 78.5%（表 6），且每项任务均有所提升。没有扰动时，RoboAda 在三个采样随机种子上的平均成功率变化了 1.7 个百分点（从 81.3% 降至 79.6%），小于冻结策略在不同随机种子间的波动。

5. 结论

在这项工作中，我们提出了 RoboAda，这是一种测试时方法，通过推断执行规划的机体，使冻结的机器人策略适应其无法观测到的物理规律。事后标

签来自名义逆模型，将机器人的自身运动转化为证据；针对显式机体假设的后验分布对这些证据进行加权，并在后验分布下执行每个规划。在五种世界动作策略和视觉-语言-动作策略上，RoboAda 使物理偏移下的平均成功率提高了一倍以上，挽回了因物体实际比外观更重而损失的大部分成功率，优于七种测试时自适应方法，并成功迁移到通过关节空间指令控制的双臂机器人。我们在第F节讨论其局限性。

大语言模型使用声明

我们使用了一款基于大语言模型的助手（Claude, Anthropic）来润色本文的写作，包括语法、措辞和 LaTeX 格式。方法、实验和结果均为作者原创，作者对本文内容承担全部责任。

参考文献

[1]Marcin Andrychowicz、Filip Wolski、Alex Ray、Jonas Schneider、Rachel Fong、Peter Welinder、Bob McGrew、Josh Tobin、Pieter Abbeel 和 Wojciech Zaremba。事后经验回放。*arXiv* 预印本 *arXiv:1707.01495*，2017年。NeurIPS 2017。2[2] Karl Johan ˚Astr¨om and Bj¨orn Wittenmark。关于自整定调节器。*Automatica*，9(2):185–199，1973。3[3] M. Athans、D. Castanon、K.-P. Dunn、C. S. Greene、W. H. Lee、N. R. Sandell 和 A. S. Willsky。采用多模型自适应控制（MMAC）方法对 F-8C 飞机进行随机控制——第一部分：平衡飞行。*IEEE* 自动控制汇刊，22(5):768–780，1977。3[4] Zechen Bai、Chen Gao 和 Mike Zheng Shou。EVOLVE-VLA：基于环境反馈的视觉-语言-动作模型测试时训练。*arXiv* 预印本 *arXiv:2512.14666*，2025年。3[5] Kevin Black、Noah Brown、Danny Driess、Adnan Esmail、Michael Equi、Chelsea Finn、Niccolo Fusai、Lachy Groom、Karol Hausman、Brian Ichter、Szymon Jakubczak、Tim Jones、Liyiming Ke、Sergey Levine、Adrian Li-Bell、Mohith Mothukuri、Suraj Nair、Karl Pertsch、Lucy Xiaoyang Shi、James Tanner、Quan Vuong、Anna Walling、Haohuan Wang 和 Ury Zhilinsky。 π_0：用于通用机器人控制的视觉-语言-动作流模型。*arXiv* 预印本 *arXiv:2410.24164*，2024年。2，4，5[6] 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。arXiv 预印本 arXiv:2307.15818，2023年。CoRL 2023。2[7] 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、Xuan- bing 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 2.0：一种具有强域随机化能力、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。arXiv 预印本 arXiv:2506 18088，2025年 2, 8, 14 . .[8]Wen-Hua Chen、Jun Yang、Lei Guo 和 Shihua Li。基于扰动观测器的控制及相关方法综述。IEEE工业电子学汇刊，63 (2):1083–1095，2016年。3[9] A. Philip Dawid。统计理论：预测序贯方法。皇家统计学会会刊：A辑，147(2):278–292，1984年。3[10] Yilun Du、Mengjiao Yang、Bo Dai、Hanjun Dai、Ofir Nachum、Joshua B. Tenenbaum、Dale Schuurmans 和 Pieter Abbeel。通过文本引导视频生成学习通用策略。arXiv 预印本 arXiv:2302.00111，2023年。NeurIPS 2023。2, 3[11] Yilun Du、Mengjiao Yang、Pete Florence、Fei Xia、Ayzaan Wahid、Brian Ichter、Pierre Sermanet、Tianhe Yu、Pieter Abbeel、Joshua B. Tenenbaum、Leslie Kaelbling、Andy Zeng 和 Jonathan Tompson。视频语言规划。arXiv 预印本 arXiv:2310.10625，2023年。3[12] 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。arXiv 预印本 arXiv:2510.13626，2025年。3, 6, 14[13] Yusen Feng、Bingchen Han、Jiangran Lyu、Kai Liu、Yixin Zheng、Yuxuan Wan、Weiheng Liu、Sun Han、Ruiqin Li、Yu-long Zhang、Fangfu Liu、Xuesong Shi、Libin Liu、Yizhou Wang、Zhizheng Zhang 和 He Wang。WAM-TTT：在测试时通过观看人类操作来引导世界-动作模型。arXiv 预印本 arXiv:2607.06988，2026 年。3[14] Jiawei Gao 、 Chaoqi Liu、Peilin Wu、Haonan Chen 和 Yilun Du。IMPACT：为大力机器人操作学习内部模型预测控制。arXiv 预印本 arXiv:2606.10818，2026年。3[15] Dibya Ghosh、Abhishek Gupta、Ashwin Reddy、Justin Fu、Coline Devin、Benjamin Eysenbach 和 Sergey Levine。通过迭代监督学习掌握目标达成。arXiv 预印本 arXiv:1912.06088，2019年。ICLR 2021。2[16] Jianing Guo、Zhenhong Wu、Chang Tu、Yiyao Ma、Xiangqi Kong、Zhiqian Liu、Jiaming Ji、Shuning Zhang、Yuanpei Chen、Kai Chen、Qi Dou、Yaodong Yang、Xianglong Liu,

Huijie Zhao、Weifeng Lv 和 Simin Li。视觉-语言-动作模型抵御多模态扰动的鲁棒性。发表于国际学习表征会议（ICLR），2026年。2, 3, 4, 5, 14[17] Yanjiang Guo、Yucheng Hu、Jianke Zhang、Yen-Jen Wang、Xiaoyu Chen、Chaochao Lu 和 Jianyu Chen。基于动作的预测：通过联合去噪过程进行视觉策略学习。arXiv 预印本 arXiv:2411.18179，2024年。3[18] James Harrison、Apoorva Sharma 和 Marco Pavone。用于高效在线贝叶斯回归的元学习先验。发表于机器人算法基础研讨会（WAFR），2018年。3[19] Yucheng Hu、Yanjiang Guo、Pengchao Wang、Xiaoyu Chen、Yen-Jen Wang、Jianke Zhang、Koushil Sreenath、Chaochao Lu 和 Jianyu Chen。视频预测策略：一种具有预测性视觉表征的通用机器人策略。arXiv 预印本 arXiv:2412.14803，2024年。2, 3[20] Hongjin Ji、Guoyang Xia、Luoyang Sun、Fangxiang Feng 和 Lei Ren。VANE：通过未来视觉表征预测实现可靠的视觉-语言-动作模型测试时训练。arXiv 预印本 arXiv:2608.09448，2026年。3[21] Moo Ji n Kim、Karl Pertsch、Siddharth Karamcheti、Ted Xiao、Ashwin Balakrishna、Suraj Nair、Rafael Rafailov、Ethan Fos- ter、Grace Lam、Pannag Sanketi、Quan Vuong、Thomas Kollar、Benjamin Burchfiel、Russ Tedrake、Dorsa Sadigh、Sergey Levine、Percy Liang 和 Chelsea Finn。OpenVLA：一种开源视觉-语言-动作模型。arXiv 预印本 arXiv:2406.09246，2024年。CoRL 2024。2[22] Moo Jin Kim、Chelsea Finn 和 Percy Liang。微调视觉-语言-动作模型：优化速度与成功率。arXiv 预印本 arXiv:2502.19645，2025年。RSS 2025。4, 5[23] Moo Jin Kim、Yihuai Gao、Tsung-Yi Lin、Yen-Chen Lin、Yunhao Ge、Grace Lam、Percy Liang、Shuran Song、Ming- Yu Liu、Chelsea Finn 和 Jinwei Gu。Cosmos Policy：微调视频模型以实现视觉运动控制与规划。arXiv 预印本 arXiv:2601.16163，2026年。2, 3, 4, 5[24] Po-Chen Ko、Jiayuan Mao、Yilun Du、Shao-Hua Sun 和 Joshua B. Tenenbaum。通过密集对应关系，从无动作视频中学习行动。arXiv 预印本 arXiv:2310.08576，2023年。ICLR 2024。2, 3[25] Po-Chen Ko、Yu Chan、Yu-Hsiang Fu、Hsien-Jeng Yeh、Chu- Rong Chen、Dong-Chen Tsai、Wei-Chiu Ma、Yilun Du、Ji- ayuan Mao 和 Shao-Hua Sun。通过潜在嵌入优化与基于拒绝的采样进行视频重新规划。arXiv 预印本 arXiv:2510.17315，2025年。3[26] Ashish Kumar、Zipeng Fu、Deepak Pathak 和 Jitendra Ma- lik。RMA：面向足式机器人的快速电机适应。arXiv 预印本 arXiv:2107.04034，2021年。RSS 2021。2, 3, 7, 14, 16[27] Kimin Lee、Younggyo Seo、Seunghyun Lee、Honglak Lee 和 Jinwoo Shin。用于基于模型的强化学习泛化的上下文感知动力学模型。发表于国际机器学习大会（ICML），2020年。2, 3[28] Sohyun Lee、Yoonjae Baek、Jaesang Won、Jinnyeong Kim、Kang Hyunwoo、Seung-Hwan Baek、Ivan Laptev 和 Suha Kwak。《机器人执行误差下的视觉-语言-动作模型：自补偿与压力测试》。arXiv预印本 arXiv:2609.37334，2026年。3，6，7，13[29] Lin Li、Qihang Zhang、Yiming Luo、Shuai Yang、Ruilin Wang、Fei Han、Mingrui Yu、Zelin Gao、Nan Xue、Xing Zhu、Yujun Shen 和 Yinghao Xu。《用于机器人控制的因果世界建模》。arXiv预印本 arXiv:2601.21998，2026年。3[30] Shuang Li、Yihuai Gao、Dorsa Sadigh 和 Shuran Song。《统一视频动作模型》。arXiv预印本 arXiv:2503.00200，2025年。2，3[31] Yichao Liang、Kevin Ellis 和 Jo˜ao Henriques。《机器人机械臂的快速电机适应》。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2024年。3[32] Bo Liu、Yifeng Zhu、Chongkai Gao、Yihao Feng、Qiang Liu、Yuke Zhu 和 Peter Stone。《LIBERO：终身机器人学习的知识迁移基准测试》。arXiv预印本 arXiv:2306.03310，2023年。NeurIPS 2023 数据集与基准。2，5[33] Changyu Liu、Yiyang Liu、Taowen Wang、Qiao Zhuang、James Chenhao Liang、Wenhao Yang、Renjing Xu、Qifan Wang、Dongfang Liu 和 Cheng Han。《通过测试时强化学习实现视觉-语言-动作模型的即时适应》。arXiv预印本 arXiv:2601.06748，2026年。3[34]Yuejiang Liu、Fan Feng、Lingjing Kong、Weifeng Lu、Jinzhou Tang、Kun Zhang、Kevin Murphy、Chelsea Finn 和 Yilun Du。《世界动作验证器：通过正向-逆向不对称性实现世界模型自我改进》。arXiv预印本 arXiv:2604.01985，2026年。3[35] D. T. Magill。《采样随机过程的最优自适应估计》。IEEE自动控制汇刊，10(4):434–439，1965年。2，3[36] Anusha Nagabandi、Ignasi Clavera、Simin Liu、Ronald S. Fearing、Pieter Abbeel、Sergey Levine 和 Chelsea Finn。《通过元强化学习在动态真实环境中学习适应》。载于国际学习表征会议（ICLR），2019年。2，3，7，14，16[37] Ku mpati S. Narendra 和 Jeyendran Balakrishnan。《使用多模型的自适应控制》。IEEE自动控制汇刊，42(2):171–187，1997年。2，3[38] Shuaicheng Niu、Jiaxiang Wu、Yifan Zhang、Yaofo Chen、Shijian Zheng、Peilin Zhao 和 Mingkui Tan。《不遗忘的高效测试时模型适应》。arXiv预印本 arXiv:2204.02610，2022年。ICML 2022。3[39] Michael O’Connell、Guanya Shi、Xichen Shi、Kamyar Aziz-zadenesheli、Anima Anandkumar、Yisong Yue 和 Soon-Jo Chung。《Neural-fly：助力强风环境下敏捷飞行的快速学习》。Science Robotics，7(66)，2022年。3[40] Kiyoshi Ohishi、Masato Nakao、Kouhei Ohnishi 和 Kunio Miyachi。《用于负载不敏感位置伺服系统的微处理器控制直流电机》。IEEE工业电子汇刊，34(1):44–49，1987年。3[41] Sangwu Park、Wonjoong Kim、Yeonjun In、Sein Kim、Hongseok Kang 和 Chanyoung Park。《面向视觉预见的视觉-语言-动作模型测试时训练》。arXiv预印本 arXiv:2605.08215，2026年。2，3

[42] Xue Bin Peng、Marcin Andrychowicz、Wojciech Zaremba 和 Pieter Abbeel。《通过动力学随机化实现机器人控制的仿真到现实迁移》。载于 IEEE机器人与自动化国际会议（ICRA），2018年。2，3[43] Physical Intelligence、Kevin Black、Noah Brown、James Darpinian、Karan Dhabalia、Danny Driess、Adnan Esmail、Michael Equi、Chelsea Finn、Niccolo Fusai、Manuel Y. Gal-liker、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。arXiv预印本 arXiv:2504.16054，2025年。2，4，5[44] Han Qi、Haocheng Yin、Aris Zhu、Yilun Du 和 Heng Yang。《通过预测性世界建模增强生成式机器人策略的推理时表现》。arXiv预印本 arXiv:2502.00622，2025年。IEEE RA-L。2，3[45] Adrian E. Raftery、Miroslav K´arn´y 和 Pavel Ettler。《模型不确定性下的动态模型平均在线预测：冷轧机应用》。Technometrics，52 (1):52–66，2010年。3[46] Spencer M. Richards、Navid Azizan、Jean-Jacques Slotine 和 Marco Pavone。《面向自适应控制的非线性系统元学习》。载于机器人学：科学与系统（RSS），2021年。3[47] Bo-Kai Ruan、Teng-Fang Hsiao、Ling Lo 和 Hong-Han Shuai。《未来是否兼容？诊断世界动作模型中的动态一致性》。arXiv预印本 arXiv:2605.07514，2026年。3[48] Gawon Seo、Dongwon Kim 和 Suha Kwak。《ACID：通过逆动力学实现动作一致性，以便利用世界模型进行规划》。arXiv预印本 arXiv:2607.02403，2026年。3[49] Krishnam Soni、Aditya Sehgal、Vedant Dave 和 Elmar Rueckert。《ALRA：世界模型的动作潜变量残差适应》。载于 NeurIPS“从预训练表示到行动智能体”研讨会（PTA），2026年。2，3，7，14，16[50] Krishnam Soni、Aditya Sehgal、Vedant Dave 和 El-mar Rueckert。《Sandwich-Residuals：世界模型的参数高效测试时适应》。arXiv预印本 arXiv:2609.21740，2026年。2，3，7，14，16[51] Yu Sun、Xiaolong Wang、Zhuang Liu、John Miller、Alexei A. Efros 和 Moritz Hardt。《通过自监督进行测试时训练，以实现分布偏移下的泛化》。arXiv预印本 arXiv:1909.13231，2019年。ICML 2020。3[52] Yu Sun、Xinhao Li、Karan Dalal、Jiarui Xu、Arjun Vikram、Genghan Zhang、Yann Dubois、Xinlei Chen、Xiaolong Wang、Sanmi Koyejo、Tatsunori Hashimoto 和 Carlos Guestrin。《学习（在测试时学习）：具有表达力强的隐藏状态的循环神经网络》。arXiv预印本 arXiv:2407.04620，2024年。2，3，7，14，16[53] Yang Tian、Sizhe Yang、Jia Zeng、Ping Wang、Dahua Lin、Hao Dong 和 Jiangmiao Pang。《预测性逆动力学模型是可扩展的机器人操作学习器》。arXiv预印本 arXiv:2412.15109，2024年。ICLR 2025。2，3[54] Chi Wan、Kangrui Wang、Yuan Si、Pingyue Zhang 和 Manling Li。《WorldAgen：通过测试时世界模型训练进行统一的状态-动作预测》。arXiv 预印本 arXiv:2609.08162，2026年。2，3[55] Dequan Wang、Evan Shelhamer、Shaoteng Liu、Bruno Olshausen 和 Trevor Darrell。《Tent：通过熵最小化实现完全测试时适应》。arXiv 预印本 arXiv:2006.10726，2020年。ICLR 2021。3[56] Ying Wang、Oumayma Bounou、Yann LeCun 和 Mengye Ren。《AdaJEPA：一种自适应潜在世界模型》。arXiv 预印本 arXiv:2606.32026，2026年。3, 7, 14, 16[57] Bj¨orn Wittenmark。《自适应双重控制方法：综述》。IFAC会议论文集，28(13):67–72，1995年。3[58] Angchen Xie、Nikhil Sobanbabu、Ishayu Shikhare、Alan Wang、Max Simchowitz 和 Guanya Shi。《FADA：通过动力学对齐实现人形机器人控制的少样本域适应》。arXiv 预印本 arXiv:2606.28476，2026年。3[59] Yuncong Yang、Zhengtao Han、Furkan Ozyurt、Zeyuan Yang、Han Yang、Junyi Cao、Haoyu Zhen、Yilun Du 和 Chuang Gan。《SyncWorld：视觉校准使世界模型成为零样本模拟器》。arXiv 预印本 arXiv:2609 09155，2026 3 . . [60] Seonghyeon Ye、Yunhao Ge、Kaiyuan Zheng、Shenyuan Gao、Sihyun Yu、George Kurian、Suneel Indupuru、You Liang Tan、Chuning Zhu、Jiannan Xiang、Ayaan Malik、Kyungmin Lee、William Liang、Nadun Ranawaka、Jiasheng Gu、Yinzhen Xu、Guanzhi Wang、Fengyuan Hu、Avnish Narayan、Johan Bjorck、Jing Wang、Gwanghyun Kim、Dan- tong Niu、Ruijie Zheng、Yuqi Xie、Jimmy Wu、Qi Wang、Ryan Julian、Danfei Xu、Yilun Du、Yevgen Chebotar、Scott Reed、Jan Kautz、Yuke Zhu、Linxi ”Jim” Fan 和 Joel Jang。《世界动作模型是零样本策略》。arXiv 预印本 arXiv:2602.15922，2026年。2, 3[61] Wenhao Yu、Jie Tan、C. Karen Liu 和 Greg Turk。《为未知做好准备：通过在线系统辨识学习通用策略》。载于《机器人学：科学与系统》（RSS），2017年。2, 3[62] Ge Yuan、Qiyuan Qiao、Jing Zhang 和 Dong Xu。《Ada-WorldPolicy：用于机器人操作的世界模型驱动扩散策略与在线自适应学习》。arXiv 预印本 arXiv:2602.20057，2026年。2, 3[63] Tianyuan Yuan、Zibin Dong、Yicheng Liu 和 Hang Zhao。《Fast-WAM：世界动作模型需要测试时未来想象吗？》。arXiv 预印本 arXiv:2603.16666，2026年。4, 5, 14[64] Chuhan Zhang、Seiji Ito、Kenta Hoshino、Satoshi Ikehata 和 Ikuro Sato。《世界一致性解码：面向世界动作模型的自验证测试时规划》。arXiv 预印本 arXiv:2609.02159，2026年。2, 3[65] Hongxin Zhang、Chunru Lin、Tsun-Hsuan Wang、Zhenjia Xu 和 Chuang Gan。《用于鲁棒机器人部署的自适应视觉-语言-动作模型》。arXiv 预印本 arXiv:2609.30092，2026年。2, 3[66]Jiaqi Zhang、Feng Ye、Mingjia Yang、Zhihong Chen、Mingkang Xiang、Xinglin Yao、Yanbin Li、Siwei Ma 和 Chuanmin Jia。《RoboTwin-Phys：世界动作模型和视觉-语言-动作模型理解物理世界吗？》。arXiv 预印本 arXiv:2609.26292，2026年。2, 3

[67] Wenyao Zhang、Hongsi Liu、Zekun Qi、Yunnan Wang、Xin- qiang Yu、Jiazhao Zhang、Runpei Dong、Jiawei He、Fan Lu、He Wang、Zhizheng Zhang、Li Yi、Wenjun Zeng 和 Xin Jin。《DreamVLA：凭借全面世界知识构想的视觉-语言-动作模型》。arXiv 预印本 arXiv:2507.04447，2025年。3[68] Xiangcheng Zhang、Runhan Huang 和 Yilun Du。《World Action Planner：利用动作条件世界模型实现可泛化的机器人决策》。arXiv 预印本 arXiv:2607.27599，2026年。2, 3[69]Yunbei Zhang、Zijian Jin、Yuanzhe Liu、Janet Wang、Xilun Zhang、Yuyou Zhang、Zhenyu Zhang、Daoan Zhang、Shuaicheng Niu、Gen Li、Jianfei Yang、Jihun Hamm、Is- mini Lourentzou、Weirui Ye、Bo Liu、Peter Stone 和 Marco Pavone。《LIBERO-MAX：世界变化时，机器人策略会适应吗？》。arXiv 预印本 arXiv:2609.36518，2026年。3, 6, 14, 16[70] Zheyuan Zhang、Suyu Ye、Nakul Agarwal、Hos- sein Nourkhiz Mahjoub、Ehsan Moradi Pari、Daniel Khashabi、Tianmin Shu 和 Vaishnav Tadiparthi。《JEPA- TTT：针对动力学变化下规划的潜在世界模型持续测试时训练》。arXiv 预印本 arXiv:2610.00722，2026年。2, 3, 7, 14, 16[71] Mengfei Zhao、Chenxi Jiang、Tuo An、Jindou Jia 和 Jianfei Yang。《APEX：用于精准操作的自适应策略执行》。arXiv 预印本 arXiv:2606.16504，2026年。3[72] Qingqing Zhao、Yao Lu、Moo Jin Kim、Zipeng Fu、Zhuoyang Zhang、Yecheng Wu、Zhaoshuo Li、Qianli Ma、Song Han、Chelsea Finn、Ankur Handa、Ming-Yu Liu、Donglai Xiang、Gordon Wetzstein 和 Tsung-Yi Lin。《CoT-VLA：面向视觉-语言-动作模型的视觉思维链推理》。arXiv 预印本 arXiv:2503.22020，2025年。3[73] Zesen Zhao、Minkyoung Cho、Hui shen、Boyuan Zheng、Kunxiao Gao、Yulong Cao 和 Z. Morley Mao。《通过零样本几何评估对世界动作模型进行测试时扩展》。arXiv 预印本 arXiv:2607.17454，2026年。3[74] Ruijie Zheng、Jing Wang、Scott Reed、Johan Bjorck、Yu Fang、Fengyuan Hu、Joel Jang、Kaushil Kundalia、Zongyu Lin、Loic Magne、Avnish Narayan、You Liang Tan、Guanzhi Wang、Qi Wang、Jiannan Xiang、Yinzhen Xu、Seonghyeon Ye、Jan Kautz、Furong Huang、Yuke Zhu 和 Linxi Fan。《FLARE：通过隐式世界建模进行机器人学习》。arXiv 预印本 arXiv:2505.15659，2025年。3[75] Wenxuan Zhou、Lerrel Pinto 和 Abhinav Gupta。《环境探测交互策略》。载于《国际学习表征会议》（ICLR），2019年。2, 3[76] Xueyang Zhou、Yangming Xu、Guiyao Tie、Yongchao Chen、Guowen Zhang、Duanfeng Chu、Pan Zhou 和 Lichao Sun。《LIBERO-PRO：迈向超越记忆化的视觉-语言-动作模型鲁棒且公平的评估》。arXiv 预印本 arXiv:2510.03827，2025年。3, 6, 14[77] Chuning Zhu、Raymond Yu、Siyuan Feng、Benjamin Burch- fiel、Paarth Shah 和 Abhishek Gupta。《统一世界模型：结合视频与动作扩散，在大型机器人数据集上进行预训练》。arXiv 预印本 arXiv:2504 02792，2025 2, 3 . . [78] Yuchen Zhu、Chenyi Xu、Yulin Zhang、Gang Xu 和 Wentao Zhu. Juno：为视觉-语言-动作模型驯服预测潜变量。*arXiv* 预印本 *arXiv:2610.09940*，2026年。 3

RoboAda：在测试时让冻结的机器人策略适应其无法感知的物理规律

补充材料

A. 实现细节

标称数据。逆动力学模型和元学习参数是在冻结的 Cosmos Policy 于标称机器人上的 1,200 条滚动轨迹（约 12k 个动作块）上训练的，使用了全部 40 项任务中编号为 20–49 的初始状态。为使指令独立于规划而变化，75% 的动作块执行时采用每轴随机的指令增益 $[0.4, 2.0]$ 和偏移量 $[-0.1, 0.1]$。训练未使用任何来自已改变机器人的数据。物理偏移和关节级评估使用编号小于 20 的初始状态，这些状态从未出现在标称数据中。针对 Cosmos Policy 的 RobustVLA 协议使用全部 50 个状态，其在编号小于 20 的状态与编号从 20 开始的状态上的增益相同（分别为 $+10.6$ 和 $+10.4$ 个百分点）。

逆动力学模型。一个多层感知机将步 $s-1, \ldots, s+2$ 处的本体感觉映射为步 s的六个机械臂指令，因此两步之后即可获得一个标签。第二个多层感知机将一个 16 步窗口起始、中间和结束处的本体感觉映射为其两个半窗口的平均指令（每个轴的留出 R^2 为 0.91–0.98）；其输出用于确定步标签的 8 步块均值。同一组逆动力学模型服务于全部五个策略。

假设。特征映射 ψ H_2（由两个包含 64 个双曲正切单元的隐藏层组成，每个轴有八个特征）接收动作块起始时的末端执行器状态、动作块内的阶段、最近三个指令以及最近已标注步的平均残差；夹爪指令从不改变。在公式（1）中，先验均值 μ_0 是名义主体，而标签噪声 $1/\beta$ 按轴在线推断，且只能增大。

元学习参数。ψ，先验精度 Λ_0，基础标签噪声和 γ 每回合统计量的遗忘因子通过闭环的可微代理模型进行元学习（第3.5节）。合成机体的增益服从对数均匀分布，取值范围为 $[0.25, 1.6]$，偏移量取值范围为 $[-0.2, 0.2]$，并包含滞后、死区、饱和及状态相关扰动；四分之一的回合会切换机体，30% 则保留标称机体。步标签是机体输出与该步逆动力学模型的实际残差之和，损失为相对于规划的跟踪误差。

证据与指令。假设先验是均匀的，$\tau = 0.25$，证据遗忘为 $\lambda = 0.6$每个 16 步动作块。对于仿射假设，公式（3）具有闭式

$$x^{\star}_{s,j} = \frac{\sum_m \pi_{m,j}\big(a_m(p_{s,j} - b_m) - c_m\big)}{\sum_m \pi_{m,j}\big(a_m^2 + v_m\big)},$$

其中 a_m 和 b_m 分别是假设 m, v_m 的预测增益和偏移量，c_m 是其增益的后验方差，H_0是增益与偏移量的协方差（$a = 1$：$b = v = c = 0$）。分母下限设为 0。3$_2$，指令会裁剪至动作范围，并按规划执行第一个动作块。

表7。**实验中使用的物理偏移**。

分布偏移	实现方式
我们的十一个物理偏移	
增益 $\times 0.3$, $\times 0.5$ 进入控制器前缩放的机械臂指令	
增益下降 $1 \to 0.3$	相同操作，在 80 步后启用
刚度 $\times 0.15$	控制器刚度缩放，重新计算阻尼
延迟 6	指令延迟六个控制步到达
固定偏移量 0.1	$+0.1$ 作用于全部七个动作维度
均匀偏移 0.1	每回合从 $U(-0.1,\ 0。1)^7$ 中抽取一次
力 20 N	施加在手上的恒定向下作用力
负载 35 N	相同的力，仅在持有物体时施加
关节2较弱，关节4的扭矩分别为指令扭矩的 80% 和 70%	
关节级退化，来自 *RoboStress [28]*	
老化摩擦	每个关节中的Stribeck摩擦 $\times 2$
摩擦漂移	摩擦从 $\times 1.5$ 上升至 $\times 2$
重力误差	重力补偿偏差为20–40%
重力误差、抓取	相同，只在抓取时
Full	老化摩擦与重力误差同时存在
比看起来更重的物体	
质量 $\times 100, \times 200$	每个可移动物体；外观不变

延迟、追赶与成本。延迟测试会比较不同滞后下的指令运动与实际运动；识别出延迟后，会跳过每个动作块中过时的起始几步。校正首次生效时，会一次性消除此前累积的方向漂移。这两项机制适用于所有假设和自适应基线。所有更新均采用闭式计算，并在 CPU 上运行；在相同回合上配对测试后，每个控制步的平均耗时在 Cosmos Policy 上从 107.9 增至 109.1 毫秒，在 Fast-WAM 上从 85.9 增至 87.8 毫秒。

B. 偏移与协议

表7 列出了我们的物理偏移和 RoboStress 退化设置。所有设置都仅在测试时生效，且不改变渲染图像。LIBERO 中可移动物体的质量为 5–136 克（中位数 23 克），远低于其现实中的对应物体，

而在试验物体的质量 $\times 20$ 和摩擦 $\times 0.1$ 方面，冻结策略的成功率未发生变化；因此，我们将每个可移动物体的质量和惯性分别放大 100 倍和 200 倍（中位数物体的质量分别为 2.3 千克和 4.5 千克），并保持其外观不变。RobustVLA 实验方案遵循已发布代码和论文中的评估等级 [16]：每回合从 $U(-0.04, 0.04)^7$，中抽取一次均匀偏移量；偏差为 $+0.03$；高斯噪声的 $\sigma = 0.3$；随机翻转（每个维度以 0.05 的概率替换为 ± 1）；尖峰（以 0.05 的概率添加± 1）；沿 $+x$ 方向施加一次 200 牛顿的推力，持续5 ± 2 步。

关节级退化在每个物理步中，以关节力矩的形式注入到未更改的控制器下方；不对回差和柔顺性进行建模。在 LIBERO-Plus [12] 和 LIBERO-PRO [76], 中，每个任务均与四种物理偏移组合：延迟 6、负载 35 N、刚度$\times 0.15$以及老化摩擦。LIBERO-MAX [69] 在机器人执行动作时改变世界：一旦夹爪进入指定物体 18 cm 范围内，便会施加八种事件之一（目标或容器被重新定位、插入障碍物、干扰物出现、相机移位、光照或视觉主题切换，或传感器噪声生效）。我们使用固定随机种子从其 8,000 个案例中抽取 320 个，每种事件 40 个（其中 224 个基于其 LIBERO-Plus 基底，96 个基于其 LIBERO-PRO 基底），并将每个案例与相同的四种物理偏移组合。在没有物理偏移的情况下，我们的冻结运行结果与已发布的 LIBERO-MAX 结果相差不超过两个百分点（Cosmos Policy 61.6对比60.6%， $\pi_{0.5}$ 60.6 对比58.8%，Fast-WAM 26.9 对比25.6%）。

回合。 RobustVLA 实验方案：全部 40 项任务，每项任务为 Cosmos Policy 使用 50 个初始状态，为 Fast-WAM 使用 20 个，为 π_0 和 OpenVLA-OFT 使用 10 个，为 $\pi_{0.5}$使用 5 个。我们的物理偏移：每项任务、每种分布偏移使用 10 个初始状态，每种策略共 4,400 个配对回合（状态 0–9；对于开发此方法所用的 Cosmos Policy，则使用留出状态 10–19）。关节级退化：使用留出状态 6–8，每种策略共 600 个配对回合。LIBERO-Plus：Cosmos Policy 用于 700 项任务（每个类别 100 项，2,800 个配对回合），另外四种策略用于其中 350 项任务（每种策略各 1,400 个）。LIBERO-PRO：全部 280 个案例（每个类别 40 个），每种策略共 1,120 个配对回合。较重物体：每项任务使用 10 个初始状态，每种质量和策略各 400 个配对回合（Cosmos Policy 使用状态 10–19，其他策略使用状态 0–9）。LIBERO-MAX：320 个案例，每个案例均在四种物理偏移下及无偏移条件下进行，每种策略共 1,600 个配对回合。

C. 基线方法

这些方法都不是为输出动作块的策略而设计的（基于 JEPA 的方法会适应供规划器查询的世界模型，GrBAL 会适应模型预测控制中的动力学模型，而 RMA 使用特权训练），因此我们在与 RoboAda 相同的接口下，重新实现了各方法的适应规则（表8）。每个

表 8. **重新实现的测试时自适应基线方法**。

方法	主体模型	每个动作块的更新
ALRA [49]	仿射残差	随机梯度下降，最后五个数据块
Sandwich-Residuals [50]	输入/输出残差	一次 Adam 步骤
AdaJEPA [56]	前向网络	Adam，最后一层
JEPA-TTT [70]	前向网络	重放全部标签
GrBAL [36]	元学习网络	一步，最近16个标签
测试时训练层 [52]	线性测试时训练层	一步，学习率
RMA 风格 [26]	历史编码器	无（离线）

基线都使用相同的事后标签、延迟测试和追赶机制；我们通过对适应后的模型进行求逆，获得规划所需的指令；此外，所学习的组件使用与 RoboAda 元学习参数相同的名义数据和合成机体进行训练。作为合理性检验，除 TTT 层外，我们重新实现的每种方法都能在我们的分布偏移下优于冻结策略；TTT 层的原始论文并未涉及动力学（GrBAL $+19$、ALRA $+14$、JEPA-TTT $+13$、RMA 风格 $+12$、Sandwich-Residuals $+10$、AdaJEPA $+7$ 个百分点；TTT 层 -3）。

表 9 在另外三种策略上重复了表3 和4 的对比，使用相同的重新实现和设置。对于 Cosmos Policy，另外四组留出的初始状态配对（8–9、10–11、12–13、14–15；每组各有 560 个分布偏移回合）中，RoboAda 的表现分别为 71.2%、70.2%、70.0% 和 73.4%；每次表现最佳的基线 GrBAL 分别为 56.8%、52.1%、54.3% 和 53.9%。

D. RoboTwin 2.0

设置。我们在干净配置下使用 RoboTwin 2。0 [7] 以及公开的 Fast-WAM RoboTwin 检查点 [63]及其推理设置（三个摄像头）；不同之处在于，无论是冻结运行还是适应运行，每执行 24 个动作就预测一个新的动作块，而不是每 8 个动作预测一次。机器人为一台 Aloha-AgileX，配有两条六关节机械臂和平行夹爪；一个动作包含 12 个机械臂关节和两个夹爪的绝对目标，由轨迹跟踪这些目标，并通过重力补偿 PD 关节驱动器执行。每种方法使用相同的初始状态，每项任务均遵循 RoboTwin 的步数限制。这八项任务（表10）都是冻结策略在 60–100% 的常规回合中能够解决的任务，其中一项需要使用双臂。

偏移量。 每个已执行动作的十二个手臂关节目标上都会加上一个偏移量；夹爪目标值保持不变。固定偏移量在每个关节上都相同（$+0.03$、 $+0.05$ 或 -0.05弧度）；每回合偏移量则按回合和关节从 $U(-c, c)$ 中抽取一次， $c \in \{0.03, 0.06\}$。

RoboAda 在关节空间中。 对于机械臂关节 j 和已执行步骤 s，指令是发送的目标值减去测量位置，$x_{s,\,j} = a_{s,\,j} - q_{s,\,j}$，而标签是实际实现的

表 9. **在另外三个策略上进行测试时适应**。留出初始状态上的成功率（%），每行使用相同的回合。我们的分布偏移：在 Cosmos Policy 上使用的驱动和接触/负载偏移。关节级：RoboStress 的五种性能退化。在每种策略上，RoboAda 在所有分布偏移下都表现最佳；在没有分布偏移时，其表现与冻结策略相差不超过 4 个百分点。

	$\pi_{0.5}$				OpenVLA-OFT				Fast-WAM	
	我们的分布偏移		关节级		我们的分布偏移		关节级		我们的分布偏移	
方法	已偏移	无	已偏移	无	已偏移	无	已偏移	无	已偏移	无
冻结策略	44.9	**96.3**	45.7	**99.2**	34.6	**97.5**	33.7	95.0	27.9	**86.3**
测试时训练层	35.2	48.8	44.0	47.5	23.8	33.1	27.0	29.2	20.5	48.8
AdaJEPA	47.1	93.1	51.5	93.3	38.7	91.9	43.0	90.0	32.9	80.0
Sandwich-Residuals	41.0	79.4	50.7	85.0	37.6	80.0	47.5	76.7	31.3	70.0
RMA 风格	53.5	91.3	52.8	94.2	44.7	92.5	41.2	94.2	35.4	77.5
JEPA-TTT	45.8	80.0	54.7	81.7	44.6	83.1	49.2	75.8	34.6	65.0
ALRA	47.9	95.0	56.3	91.7	40.8	93.1	51.3	93.3	36.3	80.0
GrBAL	50.8	88.1	70.5	87.5	44.6	72.5	58.0	72.5	32.7	61.3
RoboAda（我们的方法）	**70.7**	93.8	**75.7**	96.7	**63.0**	**97.5**	**66.3**	**95.8**	**48.4**	82.5

表 10. **RoboTwin 2.0 各任务**。Fast-WAM、冻结 → RoboAda 在每个任务的 20 个初始状态下的成功率（%）。RoboAda 在五种偏移量下的平均表现均有所提升。

	固定偏移量（弧度）			**每回合偏移量**	
Task	$+0.03$	$+0.05$	-0.05	$c = 0.03$	$c = 0.06$
提起锅（双臂）	5.0 → **75.0**	0.0 → **80.0**	0.0 → **80.0**	35.0 → **75.0**	0.0 → **75.0**
将扑克牌移开	50.0 → **100.0**	15.0 → **100.0**	35.0 → **100.0**	70.0 → **100.0**	45.0 → **100.0**
将 A 放到 B（左侧）	35.0 → **80.0**	5.0 → **85.0**	25.0 → **90.0**	70.0 → **90.0**	50.0 → **80.0**
将 A 放置到 B（右侧）	35.0 → **90.0**	0.0 → **85.0**	15.0 → **90.0**	75.0 → **90.0**	50.0 → **85.0**
将面包放在煎锅上	10.0 → **65.0**	5.0 → **85.0**	0.0 → **85.0**	40.0 → **80.0**	30.0 → **90.0**
将手机放在支架上	15.0 → **45.0**	5.0 → **40.0**	20.0 → **55.0**	**60.0** → 45.0	30.0 → **55.0**
订书机	30.0 → **85.0**	5.0 → **85.0**	20.0 → **95.0**	60.0 → **90.0**	30.0 → **90.0**
印章	0.0 → **65.0**	0.0 → **65.0**	0.0 → **60.0**	10.0 → **50.0**	10.0 → **70.0**
全部任务	22.5 → **75.6**	4.4 → **78.1**	14.4 → **81.9**	52.5 → **77.5**	30.6 → **80.6**

增量，按关节校准为 $\nu_{s,\ j} = (q_{s+1,\ j} - q_{s,\ j} - b_j)/k_j$。校准基于冻结策略在 53 条名义轨迹上拟合，这些轨迹来自评估过程中不会使用的初始状态；k_j 为 0.73–0.98，因为驱动器在一步内无法完全到达目标值。假设为 H_0: $\nu = x$； H_1: $\nu = ax + o$；以及 H_2： $\nu = (a + a_g g)$ $x + o + o_g g$，其中$g = 1$ 表示该机械臂夹爪闭合时的情况，以此代替学习特征。先验以名义主体为中心，增益的标准差为 0.5，偏移量的标准差为相应关节名义指令标准差的两倍；基础标签噪声为校准后的残差标准差。每回合统计量每 16 步按 0.9 的因子遗忘；证据温度 $\tau = 0.25$、每 16 步的遗忘量 $\lambda = 0.6$，以及指令分母的界限 0.3^2 均采用 LIBERO 的设定。指令规则应用于每个规划步骤的期望运动，而动作块剩余目标值每四步根据测量位置重新计算。如果机械臂被发送到其他位置却没有移动，该步不产生标签；小于相应关节名义标签噪声的修正不予应用，且不使用延迟测试。名义校准是 RoboAda 在 RoboTwin 上使用的唯一数据；没有进行元学习。

开发集与留出回合。 这些设置是在初始状态 0–9 上选定的，未施加扰动，偏移量为 $+0.05$ 和 $c = 0.06$。在未用于任何选择的 640 个配对回合中（所有偏移量下的状态 10–19，以及 $+0.03$、 -0.05 和 $c = 0.03$ 下的状态 0–9），RoboAda 将成功率从26.3%提高到 77.7%。

E. 其他结果

表 11 按套件细分较重物体，并展示七种基线方法在相同 Cosmos Policy 回合上的表现；表12 按事件细分 LIBERO-MAX。

F. 局限性

RoboAda 修正的是规划的执行方式，而非规划内容：它无法应对视觉、语言或场景变化，并且只能抵消可通过命令变换表达的影响。其逐轴假设只能通过特征捕捉轴间耦合，且不具备闭环稳定性保证，尽管修正幅度会受到限制，后验分布也会向名义主体收缩。本体感受标签的质量决定了可识别内容的上限。所有实验均在仿真中进行；下一步是在真实机器人上开展研究。

表 11. **各套件中看起来不重、实际较重的物体**。成功率（%），每个套件和行包含 100 个配对回合。每个可移动物体的质量都更大，但外观不变。上部和中部：在 $\times 100$ 和 $\times 200$ 时，使用 Cosmos Policy 的冻结 $\rightarrow$ RoboAda；其他策略则使用 $\times 100$。下部：在相同的 Cosmos Policy 回合上测试七种基线方法，结果见 $\times 100$。在 LIBERO-Object 和 LIBERO-Long 上的增益最大，因为冻结策略在这两个套件上的失败率最高。

物体质量	空间	物体	Goal	Long	All
Cosmos Policy（冻结的 $\rightarrow$*RoboAda*）					
$\times 100$	89.0 $\rightarrow$ **90.0**	46.0 $\rightarrow$ **92.0**	79.0 $\rightarrow$ 79.0	18.0 $\rightarrow$ **45.0**	58.0 $\rightarrow$ **76.5**
$\times 200$	48.0 $\rightarrow$ **49.0**	30.0 $\rightarrow$ **49.0**	**59.0** $\rightarrow$ 58.0	10.0 $\rightarrow$ **24.0**	36.8 $\rightarrow$ **45.0**
其他策略，$\times 100$ *(冻结的 $\rightarrow$RoboAda)*					
Fast-WAM	80.0 $\rightarrow$ **87.0**	50.0 $\rightarrow$ **74.0**	**59.0** $\rightarrow$ 57.0	32.0 $\rightarrow$ **40.0**	55.3 $\rightarrow$ **64.5**
π_0	83.0 $\rightarrow$ **87.0**	51.0 $\rightarrow$ **86.0**	80.0 $\rightarrow$ **83.0**	27.0 $\rightarrow$ **39.0**	60.3 $\rightarrow$ **73.8**
$\pi_{0.5}$	**96.0** $\rightarrow$ 87.0	51.0 $\rightarrow$ **77.0**	77.0 $\rightarrow$ **81.0**	28.0 $\rightarrow$ **43.0**	63.0 $\rightarrow$ **72.0**
OpenVLA-OFT	**91.0** $\rightarrow$ 90.0	48.0 $\rightarrow$ **90.0**	78.0 $\rightarrow$ **80.0**	17.0 $\rightarrow$ **40.0**	58.5 $\rightarrow$ **75.0**
Cosmos Policy，$\times 100$，在相同回合上的基线方法					
TTT层 [52]	36.0	29.0	57.0	5.0	31.8
AdaJEPA [56]	89.0	46.0	**86.0**	28.0	62.3
Sandwich-Residuals [50]	71.0	69.0	76.0	26.0	60.5
RMA 风格 [26]	82.0	56.0	78.0	28.0	61.0
JEPA-TTT [70]	85.0	83.0	68.0	33.0	67.3
ALRA [49]	**91.0**	52.0	79.0	28.0	62.5
GrBAL [36]	62.0	91.0	68.0	27.0	62.0
RoboAda（我们的方法）	90.0	**92.0**	79.0	**45.0**	**76.5**

表 12. **在执行中途的环境变化基础上叠加物理偏移**（来自 LIBERO-MAX）[69]。按事件统计的成功率（%），取四种物理偏移的平均值（延迟为六个控制步、35牛顿载荷、刚度 $\times 0.15$ 和老化摩擦）。每当夹爪接近指定物体时，就会发生一次事件。

	Cosmos Policy		Fast-WAM		π_0		$\pi_{0.5}$		OpenVLA-OFT	
世界变化	冻结	+ 我们的	冻结	+ 我们的	冻结	+ 我们的	冻结	+我们的	冻结	+ 我们的
目标物移位	2.5	**10.0**	4.4	**8.8**	3.1	**10.0**	12.5	**30.0**	5.6	**18.1**
容器移位	5.0	**40.6**	5.0	**13.1**	7.5	**13.1**	15.6	**44.4**	5.6	**33.1**
插入障碍物	17.5	**50.6**	15.0	**31.9**	16.9	**36.9**	26.3	**53.8**	14.4	**50.0**
干扰物出现	12.5	**47.5**	10.6	**21.9**	11.3	**26.3**	15.0	**44.4**	5.0	**20.0**
相机移位	13.8	**45.0**	4.4	**7.5**	7.5	**23.8**	23.1	**44.4**	15.6	**45.6**
光照切换	19.4	**55.6**	18.1	**36.3**	21.3	**45.0**	32.5	**60.6**	20.0	**52.5**
视觉主题切换	16.3	**52.5**	10.6	**28.1**	18.8	**31.3**	24.4	**53.8**	21.9	**58.1**
传感器噪声开始出现	12.5	**37.5**	0.0	0.0	12.5	**35.0**	25.0	**61.3**	12.5	**43.8**
全部事件	12.4	**42.4**	8.5	**18.4**	12.3	**27.7**	21.8	**49.1**	12.6	**40.2**