arXiv:2602.02396v1 [cs.RO] 2 Feb 2026

PRISM：用于单次多感官模仿学习的 Performer RS-IMLE

Amisha Bhaskar[1], Pratap Tokekar[1],Stefano DiCairano[2],Alexander Schperberg[2]
马里兰大学，帕克分校，马里兰州帕克市 20742，美国amishab@umd.edu, tokekar@umd.edu[2] 三菱电机研究实验室（MERL），马萨诸塞州剑桥市 02139, 美国dicairano@merl.com, schperberg@merl.com

摘要—**机器人模仿学习通常需要模型能够捕捉多模态动作分布，同时满足实时控制频率要求并兼容多种感知模态。尽管扩散模型、流匹配和隐式最大似然估计（IMLE）等近期生成式方法已取得可喜成果，但它们往往只能满足上述要求中的一部分。为此，我们提出 PRISM，这是一种基于批次级全局拒绝采样 IMLE 变体的单次前向策略。PRISM 将时序多感官编码器（整合 RGB、深度、触觉、音频和本体感觉）与采用 Performer 架构的线性注意力生成器相结合。我们在多样化的真实硬件平台上展示了 PRISM 的有效性，包括使用搭载 7 自由度机械臂 D1 的 Unitree GO2 进行移动操作，以及使用 UR5 机械臂进行桌面操作。在操作前停靠、高精度插入和多物体抓取与放置等具有挑战性的物理任务中，PRISM 的成功率比最先进的扩散策略高出 10–25%，同时保持高频（30–50Hz）闭环控制。我们还在包括 CALVIN、MetaWorld 和 Robo mimic 在内的大规模仿真基准上验证了该方法。在 CALVIN（10% 数据划分）中，PRISM 的成功率比扩散方法提高了 ∼25%，比流匹配提高了 ∼20%，同时将轨迹跃度降低了 20×–50×。这些结果表明，PRISM 是一种快速、精准的多感官模仿策略，既能保留对多模态动作的覆盖能力，又避免了迭代采样带来的延迟。**

I. 引言

模仿学习（IL） [2, 1, 26, 3, 36, 6] 是一种强大的框架，可直接从演示中习得复杂的视觉运动策略。模仿学习尤其适用于安全性和精度至关重要的真实世界机器人应用。这些优势在涉及多感官观测的场景中进一步放大，例如 RGB 图像、深度、触觉反馈和本体感觉；这些模态难以精确仿真，且在实践中往往仅部分可观测。因此，理想的模仿学习策略应满足三项关键标准：（i）以**实时推理频率** 运行，以便实现闭环物理控制；（ii）表征**多模态分布** 所对应的专家行为（例如多种有效抓取或接近路径）；以及（iii）在**部分感官输入**下稳健运行，通过有效融合多条感官流。

近期的生成式策略应对了这一挑战的部分难题，但鲜少能面面俱到。扩散模型 [11, 38, 31] 展现出了捕捉复杂多模态动作分布的卓越能力，但其依赖迭代去噪（每次推理通常∼10需要 100 步）这一特点，严重限制了实时部署 [22, 25]。基于流的方法 [18, 15, 40, 35] 通过连续时间积分减少了采样步数，但在多模态保真度方面仍可能表现不佳 [29]。此外，由于采用迭代采样，这些方法在推理时计算开销较大，而且通常需要大型数据集才能有效训练。这些权衡凸显了对兼具表达能力与效率的方法的需求。

最近，一种名为IMLE策略的新模型被提出，用于条件模仿学习 [27] ，该方法基于隐式最大似然估计（IMLE） [20]。IMLE通过最小化每个专家数据点与其最近生成样本之间的距离进行训练，从而确保完整覆盖专家分布，并实现**单次推理**，即并行生成多个合理动作，并在无需迭代精化的情况下选择其一。IMLE策略使用拒绝采样 IMLE（RS-IMLE） [33]，通过丢弃与真实值过于相似的候选项进一步提高候选质量，从而促进训练时采样中的多样性。

[27] 报告称，与基于扩散和流的策略相比，IMLE策略在视觉运动场景中具有有竞争力或更优的准确性和样本效率，同时由于无需迭代去噪或进行常微分方程（ODE）积分，测试时间显著更短。然而，将 IMLE 扩展到条件式策略面临两项关键挑战，因为这类策略的输出分布取决于感觉输入上下文。首先，先前的方法 [27]通常会针对每个样本单独进行拒绝采样，仅将生成的轨迹与配对的真实值进行比较。这违背了 IMLE 的核心原则——**批次全局覆盖**，往往导致策略退化为平均或中间行为，无法呈现多模态场景中的不同模式。

其次，闭环机器人任务要求各动作序列具备较强的**时间平滑性**（即随时间执行的有序动作序列）。局部平滑启发式方法（例如，选择与近期动作历史最接近的候选项）已被

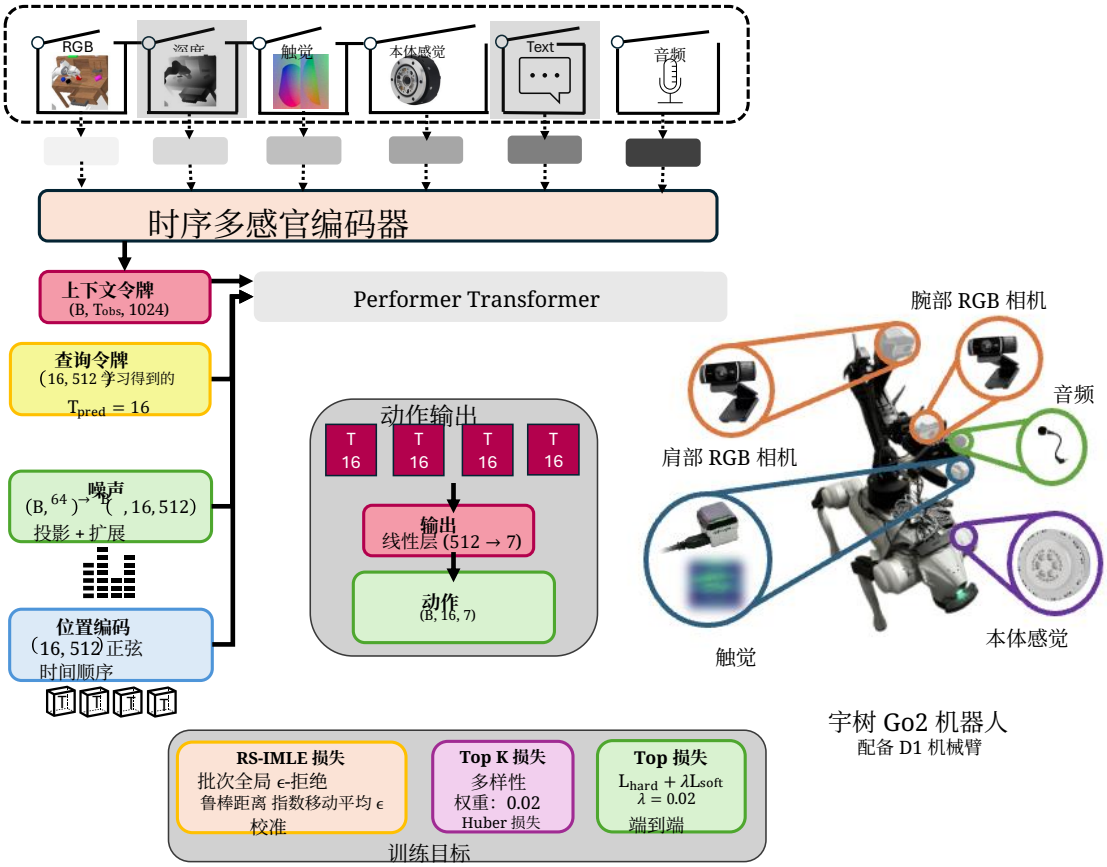

图 1：**多感官模仿 Performer RS–IMLE（PRISM）概览**。将可用传感器在每个时间步采集的特征融合为时序上下文令牌。在所有基准测试中，我们都使用一个或多个*RGB* 相机；数据集中存在时，还会使用深度相机、触觉、本体感觉、文本令牌和音频。一个双向FAVOR$^+$ （Performer）生成器利用学习型查询令牌，在单次处理中输出完整的动作序列。训练采用批次全局RS-IMLE 目标（鲁棒 Charbonnier距离， ε-拒绝机制并结合指数移动平均校准，以及可选的小型覆盖项），以保留动作的多模态性，无需迭代采样。我们针对每个基准测试分别训练模型（MetaWorld、CALVIN、Robomimic、真实硬件），仅使用相应基准测试中可用的模态（各类数据集所用模态列表见图2 ）；未使用的传感器以灰色显示。推理时，我们通过一次批处理生成k 条潜变量条件轨迹，并选取其中一条用于硬件实验的滚动时域控制。

用于缓解模式切换[1] 在执行期间 [27]，但此类方法仍然脆弱。它们会在执行早期引入模式承诺，缺乏全局轨迹推理，并且对初始次优选择或重新规划重置很敏感。

我们介绍**用于多感官模仿的 Performer RS–IMLE**（PRISM），这是一种**实时多感官模仿策略**，已在多种硬件平台上进行验证，包括用于移动操作的**UnitreeGo2 搭载 7 自由度机械臂 D1**，以及用于高精度桌面任务的**UR5 机械臂**。PRISM 通过批次全局 RS-IMLE 目标函数和时间注意力主干网络，解决了以往工作的局限性。该模型观测一段较短历史中的感官令牌（RGB、深度、触觉、本体感觉），并使用一组学习得到的动作查询令牌[2] ，通过**双向**线性注意力（FAVOR$^+$ [12]）关注这些历史信息。在一次批次化前向传播中，这些查询会生成 K 潜变量条件化的动作序列。训练期间，经 EMA 校准的批次全局拒绝掩码可确保每个候选项与小批量中的所有真实序列充分不同，从而在无需迭代采样的情况下保持多样性。**平滑且对硬件安全的轨迹生成，关键在于我们的滚动时域推理策略**：在每个时间步，模型生成 K 个动作序列，选出最连贯的序列，仅执行前 T_a 个动作，然后向前滑动观测窗口并重新规划。这种连续

[1]这里，模式对应于专家分布中的一种独特高概率策略（例如，不同的抓取方式或路径）；而不理想的模式切换则指策略随时间在此类有效策略之间突然跳转 [28, 8, 27]。

[2]学习得到的动作查询是策略内部可训练的向量，每个查询对应一个未来时间步，并通过自注意力和交叉注意力进行细化。

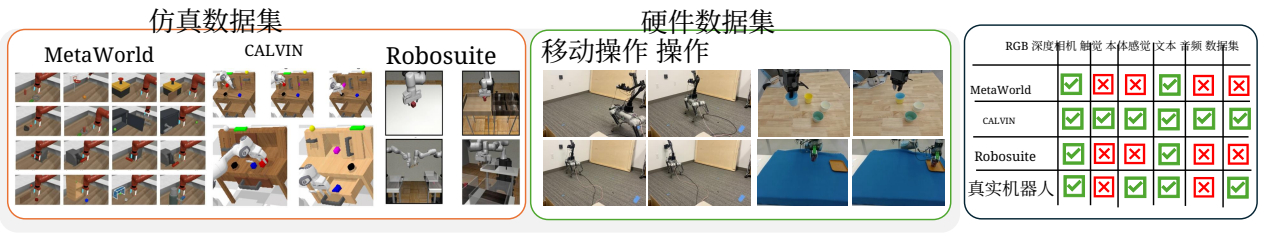

图 2：**基准数据集**。 右侧展示了各数据集所用的传感器模态。*RGB* 表示 RGB 相机，深度相机 表示深度相机，触觉 表示触觉，本体感觉 表示本体感觉，文本 表示文本令牌，音频 表示音频。

重新规划机制生成的轨迹相较于迭代采样方法加加速度更低，同时在实体机器人上保持**30-50Hz 控制**。

我们的主要贡献如下：

- 我们提出了 PRISM（用于多感官模仿的 Performer RS–IMLE），这是一种**单次前向策略**，通过结合时间交叉注意力、线性时间复杂度的 Performer 注意力和双向自注意力来生成完整的动作序列。
- 我们提出了一种**批次全局 RS-IMLE**训练目标，该目标采用经 EMA 校准的 ε 和鲁棒序列距离，鼓励覆盖小批次目标，同时不增加测试时采样步骤。
- 保时融合编码器整合了 RGB、深度、触觉、本体感觉和音频。模态丢弃研究确定了实现稳健物理性能的关键信号，使策略在传感器缺失时仍能平稳退化。
- 我们报告了一个涵盖多种任务的真实机器人测试套件的结果：**使用 Unitree Go2（7自由度机械臂）进行运动操作**，以及**使用 UR5 进行桌面操作**。我们评估了具有挑战性的任务，包括操作前停靠、高精度插销插入和多物体抓取与放置。
- 在 CALVIN、Meta-World 和 Robomimic 上，PRISM 的表现与最先进策略持平或更优。值得注意的是，**真实世界测试显示，其成功率提高了 10–25%**，相比扩散策略显著加快了推理速度。在 CALVIN 上，PRISM 的成功率比扩散策略高出 ∼25%，同时**将运动加加速度降低了20**×；这对延长硬件寿命和保障安全性至关重要。

本文其余部分的组织结构如下：第II 节介绍相关工作。第III 节将 PRISM 置于生成式模仿领域的研究版图中。第 IV 节对我们的方法进行形式化描述，从第一性原理推导批次全局 RS-IMLE 目标函数，并建立其理论一致性。第 V 节展示仿真和硬件平台上的实验结果，分析性能、计算效率和失效模式。理论证明和全面的消融实验详见附录。

II. 相关工作

行为克隆 (BC) 通过监督学习将观测映射为动作 [41, 14], , 从而实现快速的单次前向推理，以支持实时控制。然而，在多模态或接触丰富任务中，采用均方误差的朴素行为克隆往往表现不佳，会将多样化行为平均为精度较低的输出。离散化将控制重新表述为分类问题 [39, 34],, 但在高维空间中扩展性较差。混合密度网络和行为 Transformer [23, 30] 能更好地捕捉多模态性，但会受到模式坍塌和不稳定性 [13]的困扰。基于能量的隐式模型提高了多值映射 [13],的精度，但显式行为克隆难以实现良好的模式覆盖。PRISM 通过批次全局 RS-IMLE 解决这一问题，在保留单次前向效率的同时，促进多样化模式的生成，避免将其平均化。

深度生成模型，尤其是扩散和基于流的方法，通过对复杂的多模态动作进行建模，推动了视觉运动策略学习的发展 [11, 31, 21]。扩散模型通过迭代去噪实现稳定性和多模态性，其中 Diffusion Policy 在操作任务中取得了当前最优结果 [11], , 但其多步推理会引入不适用于实时控制的延迟。DDIM [31],、整流流 [21], 、SE(3) 等变变换器 [15], 以及自适应求解器 [18] 等加速方法提高了效率，而 Falcon diffusion [9] 和一致性策略 [25] 则通过蒸馏或复用步数来降低成本。然而，在较少步数下同时确保平滑度和模式覆盖仍然困难。相比之下，PRISM 通过批次全局拒绝采样实现单次推理，在无需迭代去噪的情况下保留了多模态性和时间连贯性。

单步生成策略无需承担迭代采样的成本。基于生成对抗网络的方法 [16, 10] 能够高效推理，但往往会发生模式坍塌，从而限制视觉运动控制中的多模态覆盖。相比之下，IMLE 将每个演示样本与最近的采样结果相匹配，无需对抗训练即可确保分布覆盖 [20],，而 RS-IMLE 则通过拒绝采样提升对齐效果 [33]。然而，现有的 IMLE 变体依赖局部启发式方法来实现时间连贯性，因此在长时程任务中容易出现模式切换和不稳定性。PRISM 通过批次全局 RS-IMLE 目标函数和针对多感官输入的时间交叉注意力解决了这一问题，生成多个连贯的轨迹候选项，并避免过早的模式承诺或代价高昂的迭代拒绝采样。

RGB、深度、本体感觉、触觉和音频等多模态输入通过缓解部分可观测性与噪声来提升策略鲁棒性 [3, 36]。先前的工作通过拼接或注意力融合各模态，但很少分析它们在传感器丢失条件下的相对重要性。时间平滑性对于防止模式切换引起的抖动也至关重要 [28, 8]。基于注意力的策略增强了长时程连贯性，Performer 等线性时间变体能够实现高效的 $\mathcal{O}(T, m)$ 推理 [12]。PRISM 将每个时间步的模态融合为固定宽度标记，并应用双向

线性注意力，用于实时、非自回归生成；消融实验表明，夹爪 RGB 和本体感觉至关重要，而深度可能是冗余的。

III. 问题陈述

设定。 我们考虑基于演示进行模仿学习，使用时间对齐的多感官观测。每个回合都会提供同步的传感器流，包括一个或多个 RGB 相机（始终包含腕部相机；可选包含静态视角相机），以及可选的深度、触觉、本体感觉和音频。令 $\mathcal{D} = \{(\mathbf{O}^{(n)}, \mathbf{A}^{(n)})\}_{n=1}^{N}$ 表示 N 回合，其中$\mathbf{O}^{(n)}$ 和$\mathbf{A}^{(n)}$ 分别是观测和动作序列。

观测和动作。 在离散时间 t，观测$\mathbf{o}_t$ 拼接了所有可用的传感器数据。动作$\mathbf{a}_t \in \mathbb{R}^{D_a}$ 的每个维度均归一化至 $[-1, 1]$，其中$D_a = 7$ 适用于桌面操作（末端执行器增量、 +夹爪），而 $D_a = 14$ 适用于移动操作（绝对位姿、底座速度 +、夹爪）。完整的参数化细节见附录A7。

窗口。 对于观测时域 T_o 和预测时域 T_p:

$$\mathbf{O}_{t-T_o+1:t} = \{\mathbf{o}_{t-T_o+1}, \dots, \mathbf{o}_t\},$$
$$\mathbf{A}_{t+1:t+T_p} = \{\mathbf{a}_{t+1}, \dots, \mathbf{a}_{t+T_p}\}.$$

我们的目标是学习一种通过单次前向传播运行的多模态策略:

$$\hat{\mathbf{A}}_{t+1:t+T_p} \sim \pi_\theta(\cdot \mid \mathbf{O}_{t-T_o+1:t}, z), \tag{1}$$

其中 $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 是一个潜变量。

假设。 (i) 传感器已完成时间对齐；可能会缺失某些模态。(ii) 动作有界且经过归一化。(iii) 演示具有多模态性：同一上下文可能对应多个有效动作序列。

IV. 所提出的方法

我们的目标是学习一种**多模态**（能够捕捉多样化行为）、**时间一致**（能够生成平滑轨迹）且**高效**（能够实时运行）的策略。为此，我们提出 PRISM，这是一种单次处理框架，将问题分解为两个阶段：a）**时序多感官编码器**，在保留时间维度的同时融合异构输入（B，T，特征）；b）**单次前向生成器**，利用双向线性注意力（FAVOR$^+$ [12]）并行生成完整动作序列；以及 c）**批次全局 RS-IMLE**训练目标，它提供了一种具有理论依据的模式覆盖机制，且无需承担迭代扩散带来的推理开销。图1 展示了整体概览。训练和滚动时域推理的伪代码见算法1和算法2；稳定化的 FAVOR$^+$ 特征详见算法 3 （附录 B ）。

A. 时序多感官编码器

令 T_o 为上下文时域， d为模型宽度。对于每个时间步 $t \in \{1, \dots, T_o\}$，我们构建各模态嵌入（腕部RGB、静态 RGB、深度、触觉、本体感觉、音频、文本；取决于数据集），记作 $\{\mathbf{v}_t^{(m)}\}$，其维度为 d_m。我们按时间步进行融合，以保留时序结构:

$$\mathbf{c}_t = \mathrm{MLP}_{(\sum_m d_m)\to 1024\to d}\Big(\big[\mathbf{v}_t^{(m)}\big]_{m\in\mathcal{M}_t}\Big) \in \mathbb{R}^d, \tag{2}$$

并堆叠 $\{\mathbf{c}_t\}$ 以获得上下文令牌 $\mathbf{C} \in \mathbb{R}^{T_o\times d}$ 并加入学习型绝对位置嵌入。逐时间步融合避免纠缠时间统计量，将长程一致性交由生成器处理。各基准测试的模态使用情况及详细消融实验见（表 IV）和（图 3；第 V 节验证了腕部RGB与本体感觉至关重要，而深度有时是冗余的。

B. 单次线性注意力生成器（双向）

给定 C，生成器在一次前向传播中生成长度为T_p 的动作序列。我们初始化 T_p 个可学习查询令牌 Q $\in \mathbb{R}^{T_p\times d}$，加入投影潜变量 $z \sim \mathcal{N}(0, \mathrm{I})$（用于产生多样化轨迹候选项的随机种子）和位置编码 PE$_{pos}$，然后应用 L 个 Transformer 块，其中包括：(i) 对 Q 的双向自注意力（无因果掩码）；(ii) 对完整上下文 C 的交叉注意力（同样不使用因果掩码）。两种注意力均使用 FAVOR$^+$ （线性化 softmax） [12] ，并采用稳定的正特征，将每个头的开销从 $\mathcal{O}(T^2)$ 降至 $\mathcal{O}(T\,m)$（完整细节见算法 3、附录 B）：

$$\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \approx \frac{\Phi(\mathbf{Q})\big(\Phi(\mathbf{K})^\top \mathbf{V}\big)}{\Phi(\mathbf{Q})\big(\Phi(\mathbf{K})^\top \mathbf{1}\big) + \varepsilon_a},$$
$$\Phi(\mathbf{X}) = \frac{\exp\big(\mathbf{X}W - \mathrm{rowmax}(\mathbf{X}W)\big)}{\sqrt{m}}. \tag{3}$$

其中 $W \sim \mathcal{N}(0,\ 1/\sqrt{d})$、 m 随机特征，并由 $\varepsilon_a > 0$对分母进行截断。线性头将最终 token 映射为动作 $\hat{}\mathbf{A} \in \mathbb{R}^{T_p\times D_a}$，其中 $D_a \in \{7, 14\}$ 与第 III节相符。

a）为何采用双向和线性结构： 该策略是 非自回归的，通过对 **Q** 进行双向自注意力，并对 **C**进行完全交叉注意力，从而联合选择整个序列。FAVOR$^+$ 支持实时控制：Choromanski 等人的定理 1 [12] 证明了 softmax 估计无偏，且方差为 $O(1/m)$，可防止误差在长时域内累积（形式化分析见附录 N）。消融实验（图 7b，附录 F）表明，增加随机特征预算可提升准确性；在墙钟时间相同的情况下，FAVOR$^+$ 实现了 1.5–2×倍于标准注意力的加速（表 VII，附录H）。

C. 鲁棒序列距离

我们使用 Charbonnier 惩罚 ε_c 和逐维权重 w_d （经验尺度倒数）比较预测序列与目标序列，并随时间进行归一化以保持值 O（1）：

$$D_\rho(\hat{\mathbf{A}}, \mathbf{A}) = \frac{1}{T_p}\sum_{t=1}^{T_p}\sum_{d=1}^{\mathbf{D_a}} w_d \sqrt{(\hat{a}_{t,d} - a_{t,d})^2 + \varepsilon_c^2}, \tag{4}$$

与 $\varepsilon_c = 10^{-6}$。该指标既用于训练，也用于评估时的候选项选择（当学习型代理不可用时），其 $1/T_p$ 归一化使拒绝采样阈值的尺度在各时域上保持一致。

Charbonnier公式在保持可微性的同时对异常值具有鲁棒性，这对于IMLE目标中稳定的梯度流至关重要 [5]。

D. 批次全局 *RS–IMLE* 目标

对于批次 B中的每一项 i，我们抽取 K 个潜变量，从而得到候选项 $\{\hat{\textbf{音频}}_i^{(k)}\}_{k=1}^{K}$。令$D_{i,k} = D_\rho(\hat{\textbf{音频}}_i^{(k)}, \textbf{音频}_i)$，并将到所有目标$j$的批次全局距离定义为$D_{i,k\to j} = D_\rho(\hat{\textbf{音频}}_i^{(k)}, \textbf{音频}_j)$。我们拒绝与任何目标过于接近的候选项：

$$\mathbb{I}_{i,k}^{\text{rej}} \;=\; \mathbb{I}\Big[\min_{j} D_{i,k\to j} < \varepsilon_{\text{RS}}\Big], \tag{5}$$

并在回退机制下最小化硬 IMLE 损失：

$$\begin{aligned} \mathcal{L}_{\text{hard}} &= \frac{1}{B}\sum_{i=1}^{B} \min_{k\in\mathcal{K}_i} D_{i,k}, \\ \mathcal{K}_i &= \begin{cases} \{\,k : \mathbb{I}_{i,k}^{\text{rej}} = 0\,\}, & \text{if nonempty}, \\ \{1,\dots,K\}, & \text{otherwise}. \end{cases} \end{aligned} \tag{6}$$

我们通过批次分位数 $q \in [0.2, 0.35]$的指数移动平均来校准 ε_{RS}，并施加硬钳位（$\varepsilon_{\min}, \varepsilon_{\max}$）=（$10^{-4}$,0.2）：

$$\begin{aligned} \tilde{\varepsilon} &= \text{Quantile}_q\big(\{D_{i,k\to j}\}\big), \\ \varepsilon_{\text{RS}} &\leftarrow \text{clip}\big(\alpha\,\varepsilon_{\text{RS}} + (1-\alpha)\,\tilde{\varepsilon}, \varepsilon_{\min}, \varepsilon_{\max}\big). \end{aligned} \tag{7}$$

引理A.2 （附录L）证明该估计量的方差为$O(1/N)$，依据格利文科-坎特利定理，并在图11中得到经验验证。

一个较小的 top-K' 软覆盖项可促进多样性：

$$\mathcal{L}_{\text{soft}} = -\frac{1}{B}\sum_{i=1}^{B} \log \sum_{k\in \text{Top}K'(D_{i,\cdot})} \exp(-D_{i,k}/\tau), \tag{8}$$

产出 $\mathcal{L} = \mathcal{L}_{\text{hard}} + \lambda_{\text{soft}}\mathcal{L}_{\text{soft}}$，并结合 $\lambda_{\text{soft}} \ll 1$。我们使用$K' \ll K$ （通常为 $K' = 3$，用于 $K = 16$），使梯度集中于相关模式。引理 A.3 （附录 M）证明，这等价于熵正则化的 k中心化，可防止模式坍塌。批次全局拒绝可避免候选项覆盖多个相近目标，从而在不增加测试时间开销的情况下，为其他模式保留梯度信号。理论基础（IMLE 等价于核平滑似然最大化，以及对软覆盖和批次全局拒绝的详细论证）见附录 J。

a) 为什么采用软覆盖项：标准 IMLE 中使用的硬最小值目标函数仅通过单个最近候选项传播梯度。当 K 有限且数据集包含多个模式时，这可能会使优化偏向频繁或密集的模式，因为这些模式更有可能产生最小距离。为缓解这种优化偏差，我们在排名前K' 个候选项上引入温和的软对数和指数运算。这不会改变底层的 IMLE 形式，但提供了一种受控的松弛方式，使多个邻近候选项能够贡献梯度信号，而不会在模式之间进行平均。在图6a（附录F）中可以观察到，减小该项会在多模态设置下持续降低性能。因此，我们将软覆盖视为一种实用的稳定化机制，而非理论上的必要条件。

b) 为什么采用批次全局拒绝：通过公式 5进行掩码处理，可以防止单个候选项在整个批次范围内轻易“覆盖”许多相近的目标，否则会使梯度偏向单峰中间位置。这在不增加测试时开销的情况下，为其他模式保留了梯度信号。对 K 和 ε_{RS} 的敏感性体现在消融实验中（图6，附录 F），结果显示性能在 $K \approx 16$个候选项处达到饱和。由于 $D_{i,\ k\to j}$ 是针对批次中的所有目标计算的，因此当多模态簇不同时出现时，RS–IMLE 仍能保持稳定。在这种情况下，分位数会自然收缩至批次局部密度，从而避免抑制来自缺失模态的候选项。

E. 滚动时域推理（单次通过）

在测试时间，我们观测到 $\mathbf{O}_{t-T_o+1:\ t}$，对 **C**进行编码，采样 K潜变量，并在一次通过中生成 K 动作序列。我们使用以下两种仅依赖观测的规则之一，选取单条轨迹：**(1) ProxyScore** （可用时）。如果本体感觉或腕部位姿可观测（例如 CALVIN、硬件），我们会计算一个简单的平滑度代理指标：每个候选项的首个动作会导出一个预测的下一时刻末端执行器位姿，我们选择其导出位姿最接近观测位姿（或先前执行的动作）的候选项。这可以减少突兀的模式切换，且无需动作标签。**(2) 确定性平局裁决** （代理指标不可用时）。如果数据集未提供本体感觉或末端执行器位姿（例如 Robomimic），则无法计算 ProxyScore。此时，我们选择首个动作与上一次执行的动作最接近（L2）的候选项。这有助于保持时间一致性，而无需真实动作。两种规则都仅使用观测，从不使用目标动作。我们执行所选轨迹的前 T_a个动作，并将窗口滑动 T_a。由于生成是非自回归且基于线性层注意力，在现代加速器上进行批量评估时，适度增大 K 也能获得近乎恒定的墙钟时间。完整的超参数、预处理和稳定性设置见附录 C和 A6。

V. 实验

我们在四个成熟的基准测试上评估 PRISM：Meta-World（50 项任务） [37]，CALVIN（多感官操作） [24]，Robomimic（PH） [23]，以及一套真实机器人运动操作测试，以回答以下问题：**RQ1** 在最佳候选项之外，序列级一致性是否能减少加加速度和模式切换？K？**RQ2** 随着时域增长，单次前向、线性注意力策略能否达到 30 赫兹？**RQ3** 多感官融合（RGB、深度、触觉、本体感觉）能否在遮挡/数据丢失情况下维持性能？**RQ4**在墙钟时间相同的情况下，PRISM 与扩散/流方法相比如何？**RQ5**批次全局拒绝、 ε校准和候选数量K有何影响？**RQ6** 仿真中的趋势能否延续到硬件上？完整的设置详情、公平性控制以及全面的逐任务细分结果见附录 D–H。

a) MetaWorld（*RQ1*、*RQ2*、*RQ4*）：在相同的单步推理（NFE= 1）条件下，PRISM 的成功率提升了 **4–5%**

算法1 PRISM 训练：采用批次全局 RS-IMLE 的单次线性注意力生成器

要求： 数据集 $\mathcal{D}=\{(\textbf{本体感觉}^{(n)},\textbf{音频}^{(n)})\}$；时域 (T_o,T_p)；候选项 K；模型宽度 d；注意力头 h；FAVOR$^+$ 特征 m
要求： 动作权重 $\{w_d\}_{d=1}^{D_a}$；Charbonnier ε_c；线性注意力截断 ε_a
要求： RS 分位数 $q\in[0.2,0.35]$、指数移动平均动量 $\alpha\in[0,1)$、截断值 $(\varepsilon_{\min},\varepsilon_{\max})=(10^{-4},0.2)$
要求： 可选软覆盖（K'，τ，$\lambda_{\text{soft}}\ll 1$）
初始化编码器 E_ϕ、生成器 G_θ（固定 h，m），优化器；初始化 $\varepsilon_{\text{RS}}>0$
对于每个 小批量 $\{(\mathbf{O}_i,\mathbf{A}_i)\}_{i=1}^{B}\sim\mathcal{D}$ **执行**
  构建窗口： 对于每个 i，取上下文 $\mathbf{O}_{i,\,t-T_o+1:\,t}$ 和目标 $\mathbf{A}_{i,t+1:\,t+T_p}$
  预处理： RGB/触觉 $\to[0,1]$；深度/本体感觉/音频 $\to$ float32
  编码： $\mathbf{C}_i\leftarrow E_\phi(\mathbf{O}_{i,\,t-T_o+1:\,t})\in\mathbb{R}^{T_o\times d}$
  采样潜变量： 抽取 $z_{i,\,k}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ 用于 $k=1..K$
  单次批量前向传播： $\hat{\mathbf{A}}_i^{(k)}\leftarrow G_\theta(\mathbf{C}_i,\,z_{i,\,k})\in\mathbb{R}^{T_p\times D_a}$ ▷ 双向自注意力与交叉注意力；FAVOR$^+$
  逐项鲁棒距离： $D_{i,k}\leftarrow\frac{1}{T_p}\sum_{t=1}^{D_a}\sum_{d=1}^{D_a}w_d\sqrt{(\hat{a}_{i,t,d}^{(k)}-a_{i,t,d})^2+\varepsilon_c^2}$
  批次全局距离： 计算 $D_{i,k\to j}\leftarrow D_\rho(\hat{\textbf{音频}}_i^{(k)},\textbf{音频}_j)$ 对所有 i,k,j
  分位数校准： $\tilde{\varepsilon}\leftarrow$ 分位数$_q(\{D_{i,k\to j}\})$；$\varepsilon_{\text{RS}}\leftarrow$ 裁剪$(\alpha\,\varepsilon_{\text{RS}}+(1-\alpha)\,\tilde{\varepsilon},\ \varepsilon_{\min},\ \varepsilon_{\max})$ ▷ 引理 A.2
  RS 掩码： $\mathbb{I}_{i,\,k}^{\text{rej}}\leftarrow\mathbb{I}[$最小值$_j\,D_{i,\,k\to j}<\varepsilon_{\text{RS}}]$
  硬 IMLE 损失： $\mathcal{K}_i\leftarrow\{k:\mathbb{I}_{i,\,k}^{\text{rej}}=0\}$；如果 $\mathcal{K}_i=\emptyset$ 设置 $\mathcal{K}_i=\{1..K\}$；$\mathcal{L}_{\text{hard}}\leftarrow-\frac{1}{B}\sum_{i=1}^{B}$最小$_{k\in\mathcal{K}_i}D_{i,\,k}$ ▷ 引理 A.1
  若 软覆盖已启用 **则**
    Soft 项： $\mathcal{L}_{\text{soft}}\leftarrow-\frac{1}{B}\sum_i$对数$\sum_{k\in\text{顶部}K'(D_{i,\cdot})}\exp(-D_{i,k}/\tau)$；$\mathcal{L}\leftarrow\mathcal{L}_{\text{困难}}+\lambda_{\text{软}}\mathcal{L}_{\text{软}}$ ▷ 引理A.3
  else
    $\mathcal{L}\leftarrow\mathcal{L}_{\text{hard}}$
  结束 if
  更新： $\theta,\ \phi\leftarrow\theta,\ \phi-\eta\nabla_{\theta,\,\phi}\mathcal{L}$ ▷ 梯度裁剪；AMP 安全
  日志： 拒绝率 $\frac{1}{BK}\sum_{i,\,k}\mathbb{I}_{i,\,k}^{\text{rej}}$，当前 ε_{RS}，软/硬损失比
结束循环

算法 2 采用单遍候选项选择的滚动时域推理

需要： 已训练的 E_ϕ，G_θ；时域（T_o，T_p，T_a）；候选项 K；当前时间索引 t
观测并预处理： $\mathbf{O}_{t-T_o+1:t}$；RGB/触觉$\to[0,1]$，深度/本体感觉/音频$\to$float32
编码上下文： $\mathbf{C}\leftarrow E_\phi(\mathbf{O}_{t-T_o+1:t})\in\mathbb{R}^{T_o\times d}$
单次批量生成： 采样 $z_k\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ 并计算 $\hat{\mathbf{A}}^{(k)}\leftarrow G_\theta(\mathbf{C},z_k)$，用于 $k=1..K$ ▷双向自注意力与交叉注意力；偏好+
选择轨迹： $k^\star\leftarrow\arg\min_k$ ProxyScore$(\hat{\textbf{音频}}^{(k)},\textbf{观测}_{t-T_o+1:t})$ 如果代理可用；否则使用确定性平局裁决；参见第IV-E
执行并滑动： 应用 $\hat{\mathbf{A}}_{1:T_a}^{(k^\star)}$；设置 $t\leftarrow t+T_a$；追加新的观测；重复

在简单/中等难度划分上提升了 **5–12%**，在困难/非常困难划分上，相较于此前最强的单步流基线提升了（表 I）。表 VIII（附录 I）给出了完整的 50 项任务细分，包括与 DP3 [38], ManiCM [22], SDM [19], 和 AdaFlow [18],的比较，结果显示 PRISM 在所有难度类别中都取得了最高的成功率。相较于采用 10 步去噪的扩散策略，在最困难的任务上，提升幅度达到 ∼**12%**。这些差距与更低的加加速度/更少的模式切换相一致（RQ1）：双向

表 I：Meta-World 上使用三维输入的性能基准（50 项任务）。除非另有说明，单步推理（NFE= 1）。输入：腕部RGB +本体感觉；深度由 MuJoCo 渲染。成功率（%）越高越好。逐任务细分请参见表 VIII（附录 I），了解完整的逐任务细分。

方法	NFE↓	简单↑	中等↑	困难↑	V-困难↑
扩散策略 [11]	10	83.6	31.1	9.0	26.6
流匹配策略 [7]	1	61.4	20.6	13.4	36.0
IMLE策略 [27]	1	75.6	60.4	43.5	76.0
PRISM	**1**	**96.4**	**85.5**	**58.0**	**85.8**

序列生成避免了短视的逐步修正，而迭代采样器在高接触、多阶段目标中往往需要这种修正。由于 PRISM 采用单次前向和线性注意力，因此在 T_o/T_p 增加（RQ2）的情况下仍能保持 30 赫兹，而多步策略则必须在准确性和延迟之间权衡（见表 VII、附录 H中的延迟分析）。

b) CALVIN（RQ1–RQ3–RQ4）：仅在环境 D 的 **10%**数据上训练，PRISM 的成功率比 IMLE策略高 ∼**10%**，比扩散策略高 ∼**28%**，比流匹配策略高 ∼**20%**（在匹配预算下，表II）。运动质量显著提升：加加速度降低了 ∼**20**×（从 1.05 降至 0.05），模式切换减少了 ∼**3**×（从 0.28 降至 0.10），表明批次全局 RS-IMLE 在保持时间连贯性的同时保留了多模态性（RQ1/RQ4）。

PRISM 对传感器退化也具有鲁棒性。在遮挡和单模态缺失情况下，相较于正常条件，性能仅下降 **3-5%**，而基线方法下降 8-15%（表 II）。图 3 给出了详细分析。移除腕部 RGB 或本体感觉会导致最大的性能下降（成功率分别为 41.5% 和 15.8%），而移除深度的影响很小（成功率为 66.2%，与无缺失基线的 65.2% 相近）。成对消融进一步证实，腕部 RGB 和本体感觉构成了关键的感知骨干：同时移除二者会导致几乎完全失败。这些结果验证了 RQ3：逐时间步的多模态融合能够可靠地利用腕部 RGB 和本体感觉，而在我们的设置中，深度有时是冗余的。与实际运行时间匹配的多步采样器（NFE 为= 10的扩散策略，以及 NFE 为= 5的流匹配策略）仍然速度较慢，且即使控制了实际运行时间，加加速度仍然更高（表 VII）。

c) Robomimic（PH）（RQ4）：在基于图像的任务中，PRISM 仍然是单次前向，但其平均成功率与若干多步策略相当或更高（表III）。与强大的单步基线（一致性策略 [25]）及先前的 IMLE变体相比，PRISM 平均领先**2–5%**；与 AdaFlow（NFE=∼1.2）相比，PRISM 也具有竞争力。表IX（附录I）给出了完整的比较，包括 Recti- fiedFlow [21], ActionFlow [15],Falcon-DDPM [9], 以及其他多步基线方法。这表明，在墙钟时间相当的情况下，连贯的序列级生成可以替代迭代精化。值得注意的是，AdaFlow（NFE=∼1.2）在 PH 任务上的表现略优于 PRISM（95.6% 对 92.2%

表II：CALVIN：成功率、鲁棒性和运动质量（均值 ± 标准误； $K=8$）。输入：腕部RGB + 静态 RGB + 深度 +触觉 + 本体感觉（环境 D 的 10%）。模态丢弃结果如图 3所示。

方法	无 RGB↑	深度↑	触觉↑	本体感知.↑	加加速度↓	切换↓
扩散策略 [11]	36.4±9.2	38.0±3.5	34.4±7.5	35.2±7.2	3.783	0.515
流匹配 [7]	44.4±3.7	45.6±2.7	45.6±4.1	42.0±3.5	1.148	0.586
IMLE策略 [27]	56.2±6.5	55.2±6.2	51.2±10.3	54.2±5.7	1.053	0.276
PRISM	**65.2±4.3** 6		**66.2±4.8**	**62.4±5.7** (		**0.101**

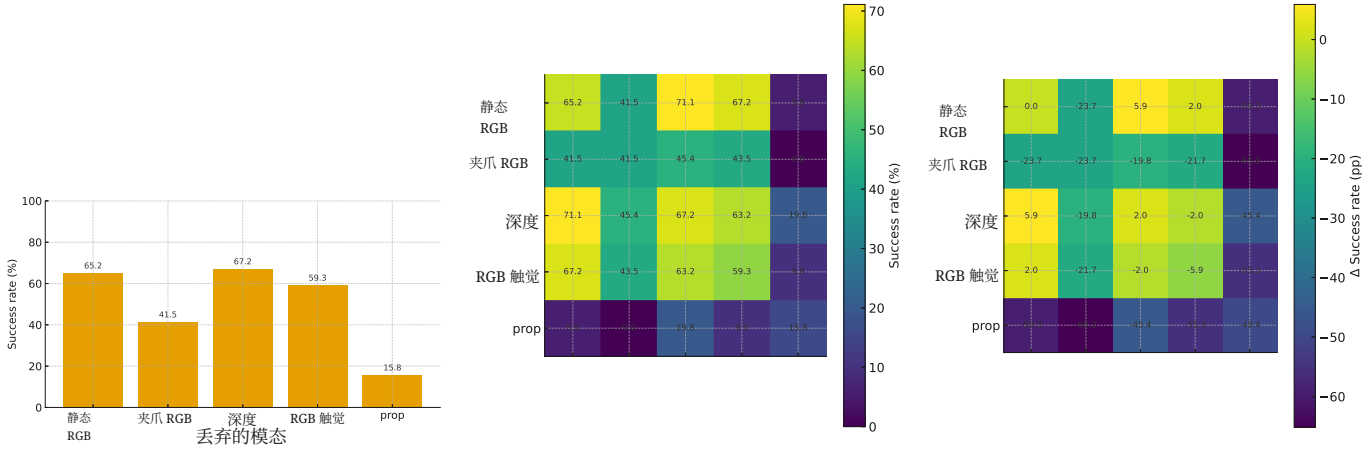

图3：**CALVIN上的模态鲁棒性。左：**评估期间的单模态缺失。**中：**成对模态缺失。**右：**相对于全模态基线的性能变化。腕部RGB和本体感觉是最关键的模态；移除两者会导致几乎完全失败，而PRISM在移除单个模态时性能会平稳下降。

平均成功率）。我们认为，这是因为 PH 演示数据近乎单峰，且缺少深度、触觉或多视角 RGB，这限制了 PRISM 多模态编码器的优势。AdaFlow 的方差自适应步长专为单模态精度优化；而当存在多模态性时，PRISM 表现出色，CALVIN（表II）和 MetaWorld 困难/极难划分（表I）上的优异表现便证明了这一点。

d) 真实世界评估（*RQ2–RQ6*）：我们在 Unitree Go2 上部署 PRISM，搭载 D1 机械臂和平行夹爪，使用历史长度 $H=8$、预测时域 $T=16$、动作块 $T_a=8$，以及每步 $K=5$ 个候选项。PRISM 在操作前停靠、插孔任务和抓取与放置任务中均优于基线方法**10–25%，**同时保持较高的推理速度（**RQ2、RQ6**）。性能随演示数量增加而提升：如图 +104 所示，将演示数量从 15 增至 35 时，停车任务提升 +20%，插孔任务提升 +30%，抓取与放置任务提升 %。此外，桌面操作实验（叠杯、罐子抓取）表明，PRISM 的表现始终比基线方法高出 15–20%（图 4），且在所有场景中均保持 50Hz 的推理频率。硬件设置详情，包括传感器配置和任务规格，见附录 I （图 8–9）。PRISM 在所有真实世界任务中均稳定提升 10–25%，这表明

PRISM 的优势能够从仿真迁移到硬件上；其持续保持 50Hz 的性能可在真实环境限制下满足**RQ2** （单次通过的效率）要求。演示数量增加带来的扩展性（性能提升 10–30%）表明，PRISM 能够有效利用更多训练数据。

架构（RQ2、RQ5）。全面消融实验（图 7–6，附录 F）表明，性能在 $H=8$个注意力头时趋于饱和；增加随机特征数量至 $m=512$会带来收益，超过 $K\approx16$个候选项后则收益递减。采用指数移动平均进行批次全局拒绝 ε校准可减少中点坍塌（降低模式切换率）并提高成功率；少量的顶部K' 覆盖项（$\lambda_{\text{soft}}\approx0.02$）能进一步提升性能，同时不会损害单候选项成功率。正如引理 A.2（附录 L）所证明的，批次全局分位数估计量方差会随 $O(1/N)$而降低。图11 通过实验证实了这一点：PRISM在批次大小为8–512时均能保持稳定的 ε_{RS} 校准；批次大小为64时，变异系数降至5%以下。请注意，我们在附录G中提供了有关PRISM效率的更多见解和失败分析。

VI. 局限性

尽管 PRISM 为单次、多感官视觉运动策略学习提供了强大的框架，并在准确性、平滑度和实时控制方面取得了显著提升，但仍存在若干局限。为确保与扩散、流和 IMLE 基线方法进行公平比较，我们没有使用预训练

表III：Robomimic（PH）：按任务划分的成功率（均值 ± 标准误）。除非特别注明，均采用单步推理（显示NFE）。输入：腕部RGB + 本体感觉（基于图像的控制，每项任务200个演示）。与多步基线方法的扩展比较见表IX （附录I）。

方法	NFE↓	Lift↑	Can↑	Sq.↑	转移↑	工具↑	平均值↑
扩散策略 [11]	15	1.00	0.99±0.01	0.92±0.03	0.79±0.04	0.55±0.05	0.85
流匹配 [7]	1	0.99±0.01	0.96±0.01	0.79±0.02	0.66±0.05	0.86±0.05	0.85
IMLE策略 [27]	1	1.00	0.94±0.01	0.81±0.01	0.85±0.05	0.75±0.06	0.87
PRISM	**1**	**1.00**	**1.00**	**0.84±0.02**	**0.91±0.02**	**0.86±0.03**	**0.922**

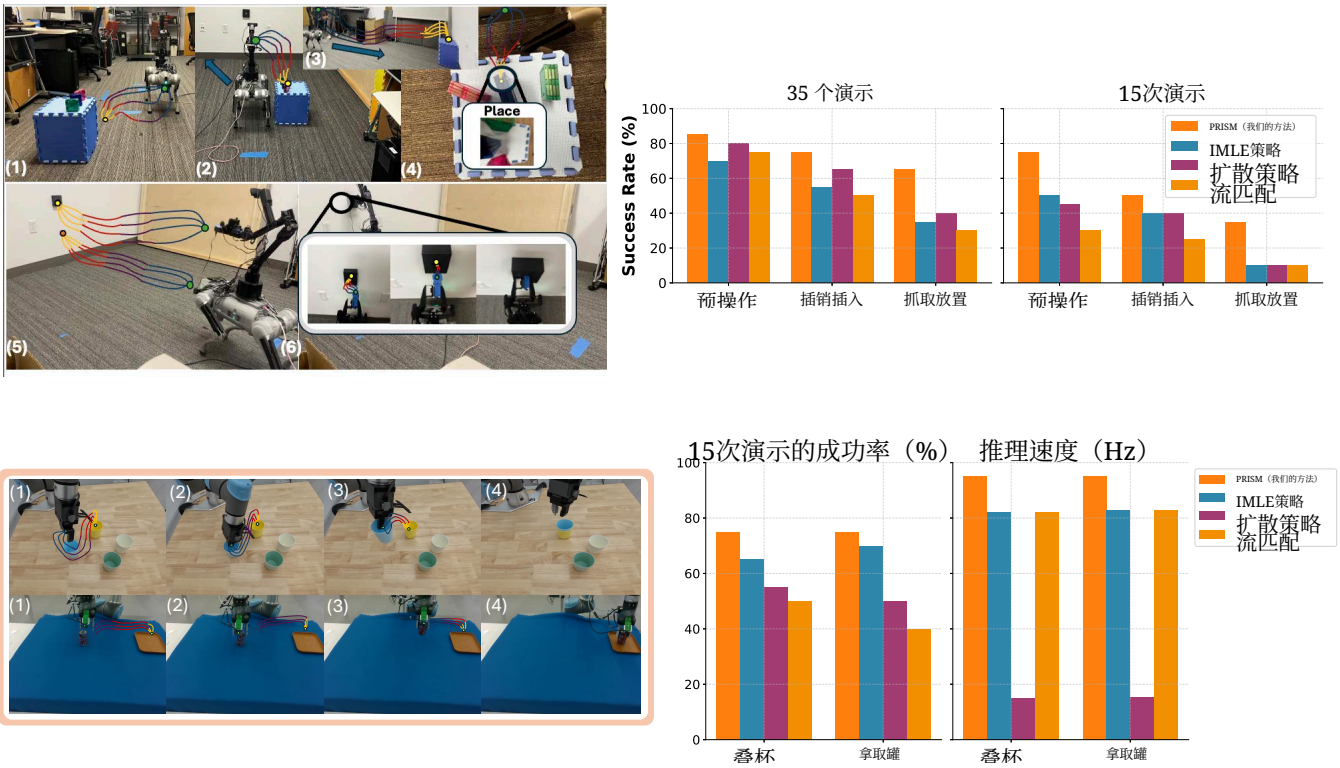

图 4：**真实世界硬件设置与性能。左上：**移动操作平台（宇树 Go2+D1）及停车、抓取与放置和插入等任务。**右上：**移动操作成功率（每种方法 50 次试验）；PRISM 在保持 30 赫兹的同时取得更高成功率。**左下：**桌面操作任务，展示了嵌套容器和精准的罐子放置。**右下：**桌面操作成功率，以及针对任务 (1) 和 (2) 的推理速度消融实验。

编码器。采用预训练视觉或多模态编码器是未来值得探索的方向，尤其适用于语言或音频条件任务。PRISM 依赖一个根据小批次距离校准的批次全局 ε_{RS}，这会使样本相互耦合，并且可能对批次大小和数据集异质性较为敏感；校准失准可能导致拒绝不足（模式平均）或拒绝过度（优化速度变慢）。未来可通过开发自适应、数据驱动的拒绝阈值来缓解这种敏感性。最后，尽管编码器能够稳健应对间歇性丢失，但持续的传感器故障仍带来挑战；可利用我们快速的推理速度，整合显式的不确定性估计通过集成编码器、多次潜变量采样或在线传感器诊断，还可进一步增强鲁棒性。

VII. 结论

我们提出了 PRISM，这是一种单次前向视觉运动策略，可融合多感官历史信息，利用双向 Performer 注意力生成完整的动作序列，并通过批次全局 RS–IMLE 保持多模态覆盖。在各项基准测试中，PRISM 准确、平滑且快速：在 MetaWorld（NFE= 1）上，其平均表现比强大的流模型高**4-5%**，在困难/极困难任务上最高提升 **∼5%**，并且优于扩散变体 **∼20%**；在 CALVIN 上，仅使用**10%** 的数据，其表现就比 IMLE-Policy 高 **∼10%**，比流模型高 **∼20%**，并且

比扩散模型高 ∼**25%** ，同时加加速度显著更低；在真实移动操作和操作任务中，其成功率提高**10-25%** ，并保持实时推理能力。未来的工作包括自适应拒绝控制、用于线性注意力的学习得到的特征映射，以及通过强化学习微调来提升泛化能力。

参考文献

[1]Brenna D Argall、Sonia Chernova、Manuela Veloso 和 Brett Browning。基于示范的机器人学习综述。《机器人与自主系统》，57（5）：469–483，2009年。[2] Christopher G Atkeson 和 Stefan Schaal。基于示范的机器人学习。载于 ICML，第 97 卷，第 12–20 页，1997年。[3] Yahav Avigal、Lars Berscheid、Tamim Asfour、Torsten Kr¨oger 和 Ken Goldberg。Speedfolding：学习高效的服装双手折叠。载于 2022 IEEE/RSJ 智能机器人与系统国际会议（IROS），第 1–8 页。IEEE，2022年。[4] R Raj Bahadur。关于大样本中分位数的注记。《数理统计年刊》，37（3）：577–580，1966年。[5] Jonathan T Barron。一种通用且自适应的鲁棒损失函数。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 4331–4339 页，2019年。[6] Amisha Bhaskar、Rui Liu、Vishnu D Sharma、Guangyao Shi 和 Pratap Tokekar。LAVA：基于长时域视觉动作的食物获取。载于 2024 IEEE/RSJ 智能机器人与系统国际会议（IROS），第 8929–8935 页。IEEE，2024年。[7] 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，2026年。URL https://arxiv.org/abs/2410.24164。[8]Yuning Chai、Benjamin Sapp、Mayank Bansal 和 Dragomir Anguelov。Multipath：用于行为预测的多种概率锚点轨迹假设。载于机器人学习会议，第 86–99 页。PMLR，2020年。[9]Haojun Chen、Minghao Liu、Chengdong Ma、Xiao-jian Ma、Zailin Ma、Huimin Wu、Yuanpei Chen、Yifan Zhong、Mingzhi Wang、Qing Li 和 Yaodong Yang。Falcon：通过部分去噪实现快速视觉运动策略。载于第四十二届国际机器学习会议，2025年。URL https://openreview.net/forum?id=fiv2M4P5vk。[10] Yiting Chen、Junnan Jiang、Ruiqi Lei、Yasemin Bekiroglu、Fei Chen 和 Miao Li。Graspada：通过领域迁移实现深度抓取适应。载于 2023 IEEE机器人与自动化国际会议（ICRA），第 10268–10274 页。IEEE，2023年。[11] 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》。《国际机器人研究杂志》，44(10-11):1684–1704，2025年。[12]Krzysztof Choromanski、Valerii Likhosherstov、David Dohan、Xingyou Song、Andreea Gane、Tamas Sarlos、Peter Hawkins、Jared Davis、Afroz Mohiuddin、Lukasz Kaiser 等。《用 Performers 重新思考注意力机制》。arXiv预印本 arXiv:2009.14794，2020年。[13]Pete Florence、Corey Lynch、Andy Zeng、Oscar A Ramirez、Ayzaan Wahid、Laura Downs、Adrian Wong、Johnny Lee、Igor Mordatch 和 Jonathan Tompson。《隐式行为克隆》。载于《机器人学习会议》，第158–168页。PMLR，2022年。[14] Peter Florence、Lucas Manuelli 和 Russ Tedrake。《视觉运动策略学习中的自监督对应关系》。IEEE机器人与自动化快报，5(2):492–499，2019年。[15]Niklas Funk、Julen Urain、Joao Carvalho、Vignesh Prasad、Georgia Chalvatzaki 和 Jan Peters。《动作流：借助空间对称流匹配实现等变、准确且高效的策略》。arXiv预印本 arXiv:2409.04576，2024年。[16] Ian J Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。《生成对抗网络》。神经信息处理系统进展，27，2014年。[17] Dorit S Hochbaum 和 David B Shmoys。《k-中心问题的最优启发式算法》。《运筹学数学》，10(2):180–184，1985年。[18] Xixi Hu、Qiang Liu、Xingchao Liu 和 Bo Liu。《AdaFlow：利用方差自适应流策略进行模仿学习》。神经信息处理系统进展，37:138836–138858，2024年。[19] Bofang Jia、Pengxiang Ding、Can Cui、Mingyang Sun、Pengfang Qian、Siteng Huang、Zhaoxin Fan 和 Donglin Wang。《分数与分布匹配策略：通过匹配蒸馏实现先进的加速视觉运动策略》。arXiv预印本 arXiv:2412.09265，2024年。[20] Ke Li 和 Jitendra Malik。《隐式最大似然估计》。arXiv预印本 arXiv:1809.09087，2018年。[21]Xingchao Liu、Chengyue Gong 和 qiang liu。《流的直行与快速生成：通过整流流学习生成和迁移数据》。载于《第十一届国际学习表征会议》，2023年。网址 https://openreview.net/forum?id=XVjTT1nw5z。[22] Guanxing Lu、Zifeng Gao、Tianxing Chen、Wenxun Dai、Ziwei Wang、Wenbo Ding 和 Yansong Tang。《ManiCM：通过一致性模型实现用于机器人操作的实时三维扩散策略》。arXiv预印本 arXiv:2406.01586，2024年。[23]Ajay Mandlekar、Danfei Xu、Josiah Wong、Soroush Nasiriany、Chen Wang、Rohun Kulkarni、李飞飞、Silvio Savarese、Yuke Zhu 和 Roberto Mart´ín-Mart´ín。

《从离线人类示范中学习对机器人操作有何重要性》。载于《机器人学习会议》，第1678–1690页。PMLR，2022年。[24]Oier Mees、Lukas Hermann、Erick Rosete-Beas 和 Wolfram Burgard。《CALVIN：用于长时程机器人操作任务的语言条件策略学习基准测试》。IEEE机器人与自动化快报，7(3):7327–7334，2022年。[25] Aaditya Prasad、Kevin Lin、Jimmy Wu、Linqi Zhou 和 Jeannette Bohg。《一致性策略：通过一致性蒸馏加速视觉运动策略》。载于《机器人学：科学与系统》，2024年。[26] Rouhollah Rahmatizadeh、Pooya Abolghasemi、Ladislau B¨ol¨oni 和 Sergey Levine。《面向低成本机器人的基于视觉的多任务操作：通过示范进行端到端学习》。载于2018年IEEE机器人与自动化国际会议（ICRA），第3758–3765页。IEEE，2018年。[27] Krishan Rana、Robert Lee、David Pershouse 和 Niko Suenderhauf。《IMLE策略：通过隐式最大似然估计，实现快速且样本高效的视觉运动策略学习》。arXiv预印本 arXiv:2502.12371，2025年。[28] Nicholas Rhinehart、Rowan McAllister、Kris Kitani 和 Sergey Levine。《Precog：在视觉多智能体场景中基于目标进行预测》。载于IEEE/CVF国际计算机视觉大会论文集，第2821–2830页，2019年。[29] Quentin Rouxel、Andrea Ferrari、Serena Ivaldi 和 Jean-Baptiste Mouret。《用于多支撑面操作的流匹配模仿学习》。载于2024年IEEE-RAS第23届国际人形机器人会议（Humanoids），第528–535页。IEEE，2024年。[30] Nur Muhammad Mahi Shafiullah、Zichen Jeff Cui、Ari-untuya Altanzaya 和 Lerrel Pinto。《行为变换器：一石克隆k种模式》。载于 Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho 编，《神经信息处理系统进展》，2022年。网址 https://openreview.net/forum?id=agTr-vRQsa。[31] Jiaming Song、Chenlin Meng 和 Stefano Ermon。《去噪扩散隐式模型》。载于《国际学习表征会议》，2021年。网址 https://openreview.net/forum?id=St1giarCHLP。[32] Aad W Van der Vaart。《渐近统计学》。剑桥大学出版社，1998年。[33] Chirag Vashist、Shichong Peng 和 Ke Li。《拒绝采样IMLE：为更优的小样本图像合成设计先验》。载于《欧洲计算机视觉会议》，第441–456页。施普林格，2024年。[34]Jimmy Wu、Xingyuan Sun、Andy Zeng、Shuran Song、Johnny Lee、Szymon Rusinkiewicz 和 Thomas Funkhouser。《用于移动操作的空间动作图》。载于《机器人学：科学与系统》（RSS）论文集，2020年。[35]Ling Yang、Zixiang Zhang、Zhilong Zhang、Xingchao Liu、Minkai Xu、Wentao Zhang、Chenlin Meng、Stefano Ermon 和 Bin Cui。一致性流匹配：通过速度一致性定义直流。*arXiv*预印本 *arXiv:2407.02398*，2024年。[36] Peihong Yu、Amisha Bhaskar、Anukriti Singh、Zahiruddin Mahammad 和 Pratap Tokekar。从草图到技能：利用人工绘制的轨迹草图引导机器人学习。*arXiv*预印本 *arXiv:2503.11918*，2025年。[37] Tianhe Yu、Deirdre Quillen、Zhanpeng He、Ryan Julian、Karol Hausman、Chelsea Finn 和 Sergey Levine。Meta-world：用于多任务和元强化学习的基准测试与评估。载于机器人学习会议，第1094–1100页。PMLR，2020年。[38] Yanjie Ze、Gu Zhang、Kangning Zhang、Chenyuan Hu、Muhan Wang 和 Huazhe Xu。三维扩散策略：通过简单的三维表示实现可泛化的视觉运动策略学习。*arXiv*预印本 *arXiv:2403.03954*，2024年。[39] Andy Zeng、Pete Florence、Jonathan Tompson、Stefan Welker、Jonathan Chien、Maria Attarian、Travis Armstrong、Ivan Krasin、Dan Duong、Vikas Sindhwani 等。Transporter网络：重组视觉世界以实现机器人操作。载于机器人学习会议，第726–747页。PMLR，2021年。[40] Qinglun Zhang、Zhen Liu、Haoqiang Fan、Guanghui Liu、Bing Zeng 和 Shuaicheng Liu。FlowPolicy：通过一致性流匹配实现快速稳健的三维流策略，应用于机器人操作。载于*AAAI*人工智能会议论文集，第39卷，第14754–14762页，2025年。[41] Tianhao Zhang、Zoe McCarthy、Owen Jow、Dennis Lee、Xi Chen、Ken Goldberg 和 Pieter Abbeel。通过虚拟现实遥操作对复杂操作任务进行深度模仿学习。载于*2018*年电气电子工程学会机器人与自动化国际会议（*ICRA*），第5628–5635页。电气电子工程学会，2018年。

附录

本附录详细介绍了所有 PRISM 实验中使用的基准测试、模态、基线方法实现以及不同基准测试之间的差异。本节汇总了此前分散在第4–5节中的信息，并回应审稿人提出的要求，即统一、透明地说明如何针对各数据集调整基线方法。

A. 基准测试规范

我们在四个广泛使用的视觉运动模仿学习基准测试上评估 PRISM：MetaWorld、CALVIN、Robomimic（PH）以及一个真实世界运动操作套件。本节提供正文中省略的实现细节，包括数据集预处理、模态配置和基线复现协议。我们的目标是使不同设置下的实验比较具备可复现性和一致性。

1) MetaWorld（50 项任务）：我们采用官方 MetaWorld MT50 基准测试 [37], , 遵循扩散、流匹配和 IMLE 视觉运动策略所采用的标准数据采集流程。所有实验均使用腕部RGB图像和本体感觉；深度由 MuJoCo 渲染，但除非另有说明，否则不用于训练。所有演示均进行帧对齐，并下采样至 20–30 赫兹。观测窗口使用 $T_o = 4$，预测时域使用 $T_p = 16$。在基线复现方面，我们遵循扩散策略、流匹配策略、IMLE策略、ManiCM、AdaFlow 和 DP3 中的设置。每个基线均使用相同的数据集划分和动作参数化方式进行训练（7D 末端执行器增量 + 夹爪）。对于单步基线（流匹配、IMLE、Consistency-FM），我们核实其报告的 NFE 配置。多步扩散基线使用 10步 DDPM 或扩散策略，除非另有明确说明。这确保了公平的实际运行时间比较。

2) CALVIN（多模态操作）：CALVIN 提供腕部相机 RGB、深度、静态相机 RGB、触觉和本体感觉 [24]。除非消融实验指定丢弃某些模态，否则 PRISM 和所有基线方法均使用所有可用模态。基线方法（扩散策略、流匹配、IMLE策略）使用与 PRISM 相同的观测和控制频率重新训练。所有传感器流均以 30 赫兹进行时间对齐，并在所有基线方法中采用相同的归一化方案进行处理。PRISM 使用时序多感官编码器融合每个时间步的所有可用模态；我们不引入任何特定于数据集的架构变更。默认包含深度，除非在消融实验中将其移除。由于 CALVIN 提供本体感觉数据，PRISM 在滚动时域推理期间使用 ProxyScore 候选选择规则，从而实现平滑且一致的轨迹 rollout，而无需访问真实动作。基线方法（扩散策略、流匹配策略和 IMLE 策略）在匹配的观测空间、控制频率和动作参数化设置下重新训练，以确保公平比较。

3) Robomimic（熟练人类演示）：我们使用 Lift、Can、Square、Transport 和 Toolhang 任务的熟练人类（PH）数据集 [23]。所有策略均使用腕部 RGB 相机和本体感觉。演示数据按时间对齐，频率为 20Hz。我们遵循 SDM[19],一致性策略[25],采用的相同评估协议：每个任务评估 50 个初始化种子，并报告平均成功率。基线方法包括扩散策略、一致性策略、一致性流匹配、IMLE策略、SDM 和 ActionFlow。若基线代码在观测归一化方面存在差异（例如，采用逐序列统计量而非全局统计量），我们会调整预处理，使其与相应论文保持一致。

表 IV：各基准测试的模态可用性。 ✓：可用且已使用；—：不可用； Δ：取决于任务。

基准测试	腕部 RGB	静态 RGB	深度	触觉	本体感觉	音频
MetaWorld	✓	—	—	—	✓	—
CALVIN	✓	✓	✓	✓	✓	—
Robomimic	✓	—	—	—	✓	—
真实硬件	✓	✓	—	Δ	✓	Δ

4) 真实世界的移动操作与桌面操作：我们在配备 D1 七自由度机械臂、腕部 RGB、肩部 RGB、基于视觉的触觉感知、音频和本体感觉的 Unitree GO2 上评估 PRISM。我们使用 ROS时间戳对传感器数据流进行时间对齐，并同步至 30 赫兹。我们使用三项任务——操作前停靠、插销插入和抓取与放置，每项任务均使用 15-35 个遥操作示范。对于所有移动操作任务，我们都使用视觉和触觉。对于图4底行所示的桌面操作任务，我们同样使用视觉和触觉。对于图4 顶行的桌面操作任务，我们使用视觉和音频。我们还新增了图10所示的语言条件任务实验，其中使用了视觉和文本。基线方法按照相同的推理时间假设重新实现：扩散策略（10 步去噪）、流匹配（1 步）和 IMLE策略（1 步）。

5) 基线方法与缩写：为使结果表格简洁，我们在各基准测试中统一使用以下缩写：“MW”表示 MetaWorld，“CV”表示 CALVIN，“RM”表示 Robomimic（PH）。所有对比方法均在观测模态、控制频率、动作参数化和数据预算一致的条件下重新训练。对于多步基线方法，我们使用其推荐的函数评估次数（NFE）报告主要结果。对于单步基线方法，评估严格遵循 Consistency-FM 中描述的设置 [35], ，确保 PRISM 与现有最强加速方法进行比较。仅当基线方法的设置与我们的设置不同时，我们才重新实现；否则，我们使用原文报告的、条件匹配的结果。

6) 数据集详情： 表 IV 总结了各基准测试中的模态可用性。所有数据集均包含腕部安装的 RGB 相机图像，作为主要视觉输入。

a) 数据预处理：

- **视觉输入**：中心裁剪并调整大小（静态图像为 200×200 ，腕部图像为 84×84 ），归一化至 [0,1]
- **深度**：转换为 float32，不进行归一化

- **本体感觉**：使用训练统计量进行逐维度 z 标准化
- **触觉**：6 通道力/力矩读数，归一化至[0,1]
- **时间对齐**：所有模态均与相机时间戳同步（30 Hz）

7) 动作空间参数化：**桌面操作**（$D_a = 7$）：

- 平移增量：$(\Delta x, \Delta y, \Delta z)$，单位为米
- 旋转增量：$(\Delta r_x, \Delta r_y, \Delta r_z)$，单位为弧度（小角度欧拉角）
- 夹爪：连续开合指令 $\in [-1, 1]$

所有增量均以机器人基座坐标系表示。

移动操作（$D_a = 14$）：

- 夹爪指令：$\in [-1, 1]$
- 末端执行器位置：(x, y, z)，单位：米
- 末端执行器方向：(q_w, q_x, q_y, q_z)，单位四元数
- 基座线速度：(v_x, v_y, v_z)，单位：米/秒（机体坐标系）
- 基座角速度：$(\omega_x,\ \omega_y,\ \omega_z)$，单位为 rad/s（机体坐标系）

传感器观测以 30 Hz 与相机时间戳同步。坐标系约定遵循各数据集的标准（详见第 A6 节中各数据集的具体说明）。

8) 评估协议：在所有基准测试中，我们均遵循以往视觉运动策略研究采用的标准评估协议。对于 MetaWorld，我们评估全部 50 项任务，并按简单/中等/困难/极难划分进行分层；每项任务使用 10 个演示来衡量数据效率，并将成功率报告为在回合限制内完成任务的比例。CALVIN 在多任务设置下进行评估，使用环境 D 数据集的 10%，通过从 K 个候选项中选出最佳项并使用特定于任务的代理函数计算 Success@K。对于 Robomimic，我们采用熟练人类（PH）划分，每项任务使用 200 个演示，并对 Lift、Can、Square、Transport 和 Tool-Hang 进行基于图像的控制。真实世界评估中，每种方法在随机初始条件下进行 50 次试验，超时时间为 60 秒，涵盖操作前停车、插销插入，以及涉及不同物体几何形状的抓取与放置任务。

B. 模型架构

a) 时序多感官编码器：编码器在进行时序融合之前，会分别处理每种模态。我们对 RGB 输入采用 Res Net-18 主干网络，以利用预训练特征；其他模态则使用轻量级自定义编码器：

- **RGB 流**（腕部 $\mathbf{v}_{\text{rgb-w}\,t}$，静态$\mathbf{v}_{\text{rgb-s}\,t}$）：ResNet-18 主干网络（在 ImageNet 上预训练），移除最后的全连接层，并将特征投影至 d=512。
- **深度流**（腕部 $\mathbf{v}_{\text{dep-w}\,t}$，静态 $\mathbf{v}_{\text{dep-s}\,t}$）：三个卷积层（1→64→128→256 个通道），采用 ReLU 激活、自适应池化至 2×2，并通过线性投影映射至 d=256。
- **触觉**（$\mathbf{v}_{\text{tact}t}$）：六通道输入经相同的 CNN 架构（3 层）处理，输出维度为 256。
- **本体感觉**（$\mathbf{v}_{\text{本体感知}t}$）：两层多层感知机（输入→128→64），用于处理关节状态与速度。

逐时间步融合可保留时序结构：

$$\mathbf{c}_t = \mathrm{MLP}_{(\sum_m d_m)\to 1024\to d}\Big(\big[\mathbf{v}_t^{\text{rgb-s}}; \mathbf{v}_t^{\text{rgb-w}}; \mathbf{v}_t^{\text{dep-s}}; \mathbf{v}_t^{\text{dep-w}}; \mathbf{v}_t^{\text{tac}}; \mathbf{v}_t^{\text{prop}}\big]\Big). \tag{9}$$

b) 单次生成器：生成器采用 L=6 个 Transformer 块，通过双向自注意力处理 T_p个学习型查询标记，并与上下文 C进行交叉注意力。两种注意力机制均采用 FAVOR$^+$ 线性化：

算法 3 FAVOR$^+$ 线性注意力实现

要求： 查询 $\mathbf{Q}$，键 $\mathbf{K}$，值 $\mathbf{V}$ $\in \mathbb{R}^{L\times d_h}$
要求： 随机特征 $\mathbf{W} \in \mathbb{R}^{d_h\times m}$，$W_{ij} \sim \mathcal{N}(0,\ 1/\sqrt{d_h})$
函数稳定化特征(**X**)
  $\mathbf{U} \leftarrow \mathbf{XW}$
  $\mathbf{U} \leftarrow \mathbf{U}$ −行最大值(**U**)　　数值稳定化
  返回 $\exp(\mathbf{U})/\sqrt{m}$
结束函数
函数线性注意力(**Q**，**K**，**V**)
  $\phi_Q \leftarrow$ 稳定化特征(**Q**)
  $\phi_K \leftarrow$稳定化特征(**K**)
  $\mathbf{S} \leftarrow \phi_K^\top \mathbf{V}$，$\mathbf{n} \leftarrow \phi_K^\top \mathbf{1}$
  返回 $(\phi_Q \mathbf{S}) \oslash (\phi_Q \mathbf{n} + \varepsilon_a)$
结束函数

c) 评论 ε_{RS} 和 K：由于 $D_{i,\ k\to j}$ 是针对批次中的所有目标计算的，因此当多模态簇未同时出现时，RS–IMLE 仍能保持稳定。在这种情况下，分位数会自然收缩至批次局部密度，从而避免抑制来自缺失模态的候选项。

C. 训练配置

a) 批次全局 *RS-IMLE*：训练目标将硬 IMLE 损失与可选的软覆盖相结合：

$$\mathcal{L}_{\text{hard}} = \frac{1}{B}\sum_{i=1}^{B} \min_{k\in\mathcal{K}_i} D_\rho(\hat{\mathbf{A}}_i^{(k)}, \mathbf{A}_i) \tag{10}$$

$$\mathcal{L}_{\text{soft}} = -\frac{1}{B}\sum_{i=1}^{B} \log \sum_{k\in \mathrm{Top}_{K'}(D_{i,\cdot})} \exp(-D_{i,k}/\tau) \tag{11}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{hard}} + \lambda_{\text{soft}}\mathcal{L}_{\text{soft}} \tag{12}$$

其中，$\mathcal{K}_i$ 包含根据批次全局距离筛选出的未拒绝候选项，而拒绝阈值 ε_{RS} 则通过批次分位数的指数移动平均进行动态校准。

b) 优化细节：

- **优化器**：AdamW，学习率为 10^{-4}（MetaWorld、Robomimic）或 5×10^{-5}（CALVIN）
- **批次大小**：128，采用混合精度训练
- **梯度裁剪**：可选，设为 1.0 以提高稳定性
- **学习率调度**：带预热的余弦退火

表 V：数学符号及其定义。

符号	描述
$\mathcal{D} = \{(\mathbf{O}^{(n)}, \text{音频}^{(n)})\}_{n=1}^{N}$	N 示范回合数据集
$\mathbf{o}_t \in \mathbb{R}^{D_o}$	时刻 t的多模态观测
$\text{音频}_t \in [-1,1]^{D_a}$	归一化动作（桌面操作为7D，移动-操作为14D）
T_o, T_p, T_a	观测、预测与动作执行时域
$\mathbf{C} \in \mathbb{R}^{T_o \times d}$	时间融合上下文标记
$\mathbf{Q} \in \mathbb{R}^{T_p \times d}$	用于动作生成的学习型查询标记
G_θ	FAVOR$^+$ transformer生成器，带参数-参数 θ
$D_\rho(\cdot,\ \cdot)$	鲁棒 Charbonnier 序列距离
ε_{RS}	批次全局拒绝阈值（指数移动平均-校准）
K, K'	候选项数量；取前K' 项用于软覆盖
m	FAVOR$^+$的随机特征数量
	线性化

- **RS-IMLE 校准**：分位数 $q \in [0.2, 0.35]$，指数移动平均动量 $\alpha = 0.9$
- **鲁棒距离**：使用 Charbonnier 和 $\varepsilon_c = 10^{-6}$，各维度权重取自训练统计量的倒数

D. 数学符号

表 V为全文所用符号提供了全面参考。

本附录提供了全面的消融研究和支持正文论断的扩展实验结果，包括架构设计选择和计算效率测量。

E. Push-T 多模态性研究

为了可视化不同策略在固有多模态性各异的状态中的行为，我们遵循 Rana etal. [27] 的 Push-T 设置。我们固定 T 形物体，并沿 T 的上边缘逐渐扫动机器人末端执行器的起始位置。靠近角落的状态实际上是单峰的（大多数演示沿一致方向推动），而靠近中心的状态则表现出高度多模态性（演示在向左推动与向右推动之间分化）。

图5比较了 PRISM、基于 U-Net 的 IMLE 基线、DiffusionPolicy 和流匹配策略。每个单元格展示了从固定初始状态出发的多条轨迹 rollout，轨迹随时间进行颜色编码。

F. 架构与超参数消融

我们在 CALVIN 上进行了系统的消融实验，以验证架构设计和超参数选择。图7 展示了对关键架构组件（注意力头、FAVOR$^+$ 特征、Transformer 深度、候选数量 K）的全面消融实验，而图6则分析了训练超参数（软覆盖权重、RS-IMLE 校准、观测时域）。这些研究为我们的默认

表 VI：模型规模与效率比较。与基线架构相比，PRISM 以更少的参数和更小的内存占用实现了更高的性能。

模型	参数	MB	成功率	类型
PRISM	**44.4**	**169.4**	**66.0**	线性层 Transformer
U-Net RS-IMLE	91.8	350.4	58.8	CNN扩散
扩散策略	79.9	305.0	62.0	CNN扩散
流匹配	79.9	305.0	61.5	CNN流

表 VII：NVIDIA A100 上不同预测时域的推理延迟（ms）。越低越好；30 Hz 要求 $\leq$33ms。

方法	T_o=8, T_p=16		T_o=16, T_p=32		30赫兹?
	Mean	p99	Mean	p99	
扩散策略 (NFE=10)	142.3	168.5	285.7	342.1	No
流匹配（NFE=5）	73.8	89.2	147.6	181.3	No
PRISM（K=8）	**15.0**	**16.0**	**31.2**	**35.8**	Yes

配置选择提供了依据，并展示了我们的方法在合理的超参数范围内具有鲁棒性。

G. 架构分析

效率。PRISM 的架构效率源于两项设计选择：通过 FAVOR 实现线性注意力复杂度 $O(T \cdot m)^+$ 以及紧凑的编码器-生成器架构。与基于 UNet 的扩散模型相比，PRISM 的参数量减少了 48%（44.4M 对比 91.8M），内存占用减少了 45%（169.4 MB 对比 350.4 MB）（表 VI，附录 H）。在线性注意力机制下，均值和 p99 推理延迟随观测时域 T_o 和预测时域 T_p 近乎线性增长，在本文报告的设置下均低于 33 ms（表 VII，附录 H），从而支持实时 30 Hz 控制。

失败分析。详细细分结果（附录 O）表明，CALVIN 中约 34% 的失败源于模式切换不稳定性（10–15%）和高急动事件（5–10%），而非无效动作生成。这证实了多模态模仿学习的主要挑战在于轨迹一致性，PRISM 通过批次全局 RS-IMLE 和时间交叉注意力应对这一问题。这些研究共同回答了 **RQ1–RQ5**；完整的消融图、逐任务细分结果及扩展的基线比较见附录D–H。

a) 消融实验与分析（*RQ1–RQ5*）：**多模态行为（RQ1）**。Push-T 评估（图 5，附录E）证实，PRISM 能够在模糊区域正确保持两种不同的轨迹模式（向左推动与向右推动），而流匹配策略则将这些模式平均为次优的直线推动，从而验证批次全局 RS-IMLE 可在不发生模式坍塌的情况下保留多模态性。

H. 计算效率

我们测量不同时间时域下的推理延迟，以验证 PRISM 的实时能力。所有测

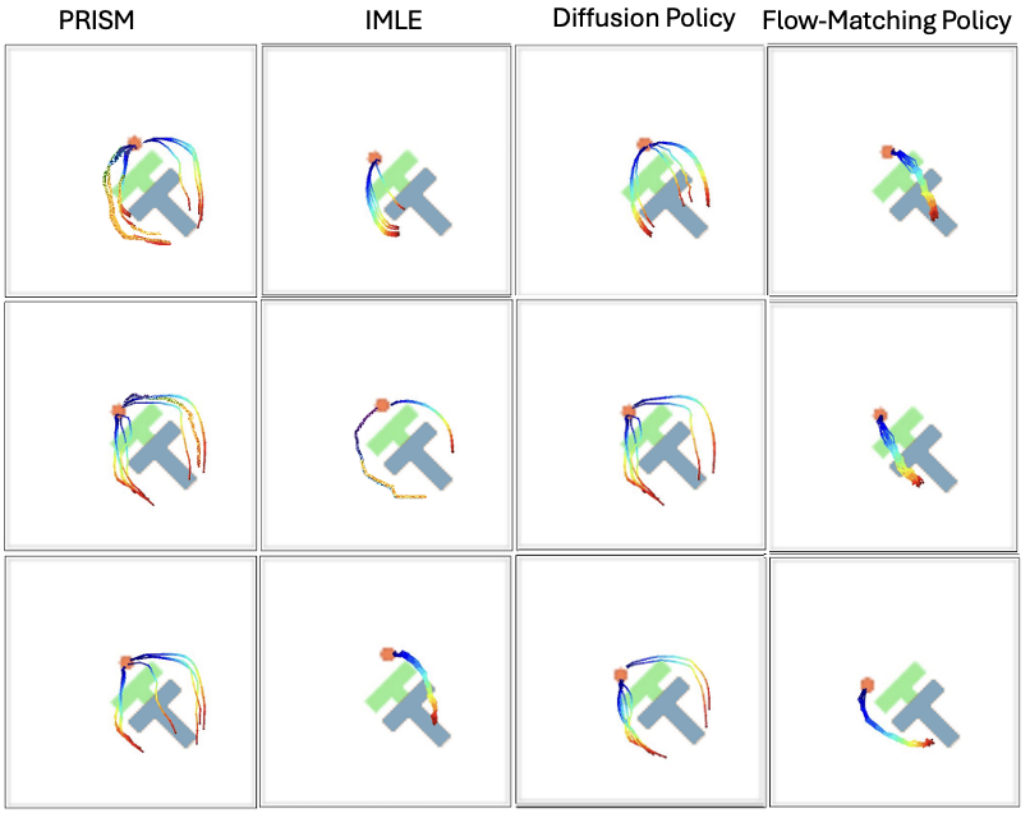

图 5：**Push-T 多模态性可视化**。各列展示末端执行器初始位置沿 T 形块上边缘从左向右扫动时的情况。每个单元格叠加了来自给定策略的多条轨迹。**PRISM** 能够保持多样且平滑的轨迹族，尤其是在含糊不明的中央区域。IMLE 基线呈现的模式较少，扩散策略产生了更多杂乱的分支，而流匹配在所有初始状态下基本保持单峰。

量均使用 NVIDIA A100 GPU，批次大小为 $K = 8$个候选项，并在预热后对 1000 次前向传播取平均值。表格VII报告了均值和第 99 百分位数（p99）延迟，表明即使在较长时域下，PRISM 仍能将推理延迟保持在 33 毫秒以下；而迭代采样方法的延迟会随去噪步数线性增长。

表VII报告了观测时域和预测时域扩展时的推理延迟。*在 $T_o = 16$，$T_p = 32$，PRISM 的运行频率为 22 Hz（p99），仍适用于许多操作任务。

I. 扩展表格

本节完整列出了正文中报告的 MetaWorld（50 项任务）和 Robomimic（5 项任务）实验的逐任务细分结果。所有结果均报告均值 ±标准误差，基于相同的随机种子，每项任务进行 5 次评估运行。表 VIII 和 IX 展示了单步（NFE=1）和多步基线比较，表明 PRISM 在各类操作挑战中始终具备性能优势。MetaWorld 逐任务结果（50 项任务）以及 Robomimic 逐任务／时间匹配比较详见补充材料（.zip），其中包含均值、标准误差和相同的种子划分。

为验证仿真结果能否在真实环境中复现，我们将 PRISM 部署到两种不同的机器人平台上，展示其在不同机器人本体和传感器套件上的通用性：

1) **移动操作（Unitree GO2）：** 一台配备 Unitree D1 七自由度机械臂的四足机器人。传感器套件包括腕部和肩部 RGB 相机、触觉传感器以及本体感觉。我们使用该平台评估诸如预操作和插销插入等移动操作任务。（图8）。

2) **桌面操作（UR3e）：** 一台固定基座的 6自由度机械臂，配备腕部安装的 RGB 相机、一个**DIGIT**高分辨率触觉传感器以及接触式麦克风。（图4）。

图4 总结了这些多样化真实世界场景中的定量成功率。

J. 语言条件操控

为了验证PRISM多传感器编码器的灵活性（图1），我们使用冻结的CLIP文本编码器（ViT-B/32）与视觉上下文流融合，开展了带文本条件的真实世界实验。

实验设置：我们针对四项不同的任务，为每条语言指令收集了 25 个示范：

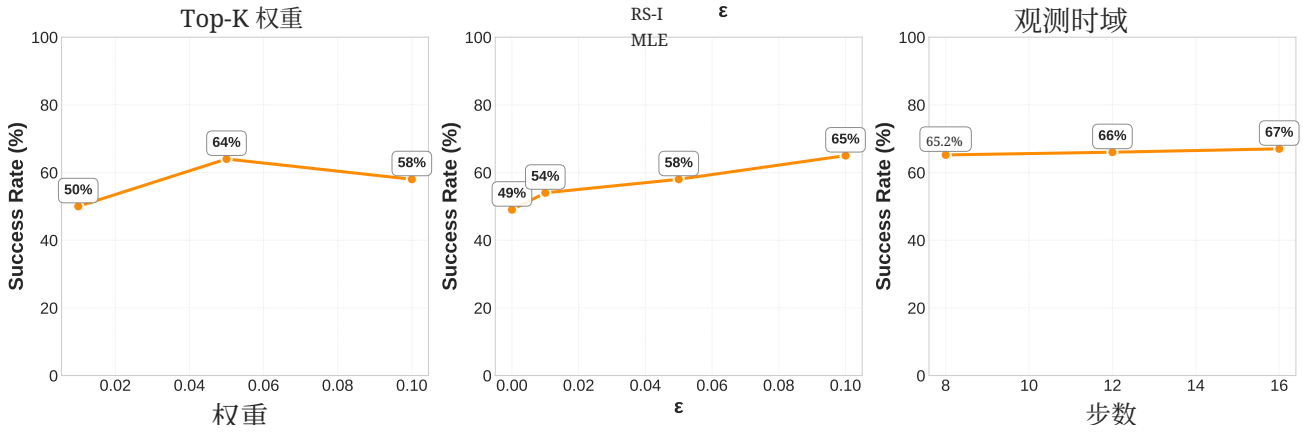

图 6：**CALVIN 上的训练超参数消融实验**。（a）软覆盖权重 λ_{soft}、（b）批次全局 RS-IMLE 阈值校准以及（c）观测时域对性能的影响。经 EMA 校准的阈值和较小的覆盖项（$\lambda_{\text{soft}} = 0.02$）在多样性与准确性之间实现了最佳权衡。

表 VI II：MetaWorld（50 项任务）。除非另有说明，均为单步（NFE= 1）。成功率（%）越高越好 ter.

方法	NFE ↓	简单（28）↑	中等 (11) ↑	困难 (6) ↑	非常困难 (5) ↑
扩散策略 [11]	10	83.6	31.1	9.0	26.6
DP3* [38]	10	89.0	72.7	38.0	75.8
ManiCM [22]	1	83.6	55.6	33.3	67.0
SDM [19]	1	86.5	65.8	35.8	71.6
FlowPolicy* [40]	1	92.1	73.6	46.2	80.0
AdaFlow* [18]	1	50.6	19.1	12.6	32.3.0
流匹配策略* [7]	1	61.4	20.6	13.4	36.0
IMLE [27]	1	75.6	60.4	43.5	76.0
PRISM	**1**	**96.4**	**85.5**	**58.0**	**85.8**

- “将蓝色杯子叠放在绿色杯子上”（结合颜色辨别的垂直堆叠）
- “将黄色杯子叠放在蓝色杯子上”（需要从初始配置重新排序）
- “将绿色球放入绿色杯子中”（形状和颜色匹配）
- “将绿色杯子挂在挂钩上”（需要精确空间推理）

结果：表 X显示，PRISM 能够成功泛化至语言条件控制，在各项任务中达到 78–92% 的成功率。策略能够正确关注颜色和空间描述符，表明 Performer 架构能够在语言标记与视觉上下文之间有效进行交叉注意。

定性观测：视频分析表明，PRISM能够根据物体描述符正确调节其抓取策略（例如，对“杯子”采用比“球”更宽的抓取开口）。该策略还展现出组合性，能够在单条语言指令中成功串联多个子目标（例如，先伸手到蓝色杯子，再将其Transport到绿色杯子）。

K. IMLE 框架与隐式似然

我们将 PRISM 建立在隐式最大似然估计（IMLE）的理论框架之上。虽然 PRISM 不会输出显式的参数化密度（如高斯分布），但它并非启发式方法，而是一个具有统计学理论依据的生成模型。

引理 A.1（隐式似然最大化与模式覆盖）。 设$\mathbf{x}$ 表示一个数据样本， $G_\theta(\mathbf{z})$ 表示一个将噪声$\mathbf{z} \sim \mathcal{N}(0, I)$ 映射到输出空间的潜变量生成器。*IMLE* 目标函数。

$$\min_\theta \mathbb{E}_{\mathbf{x}} \left[\min_{\mathbf{z}} \|G_\theta(\mathbf{z}) - \mathbf{x}\|_2^2 \right]$$

渐近等价于最大化数据分布的核平滑对数似然。

证明： 我们引用 Li 和 Malik [20], 中的定理 1，该定理指出：“IMLE 估计量是一致的，并且渐近

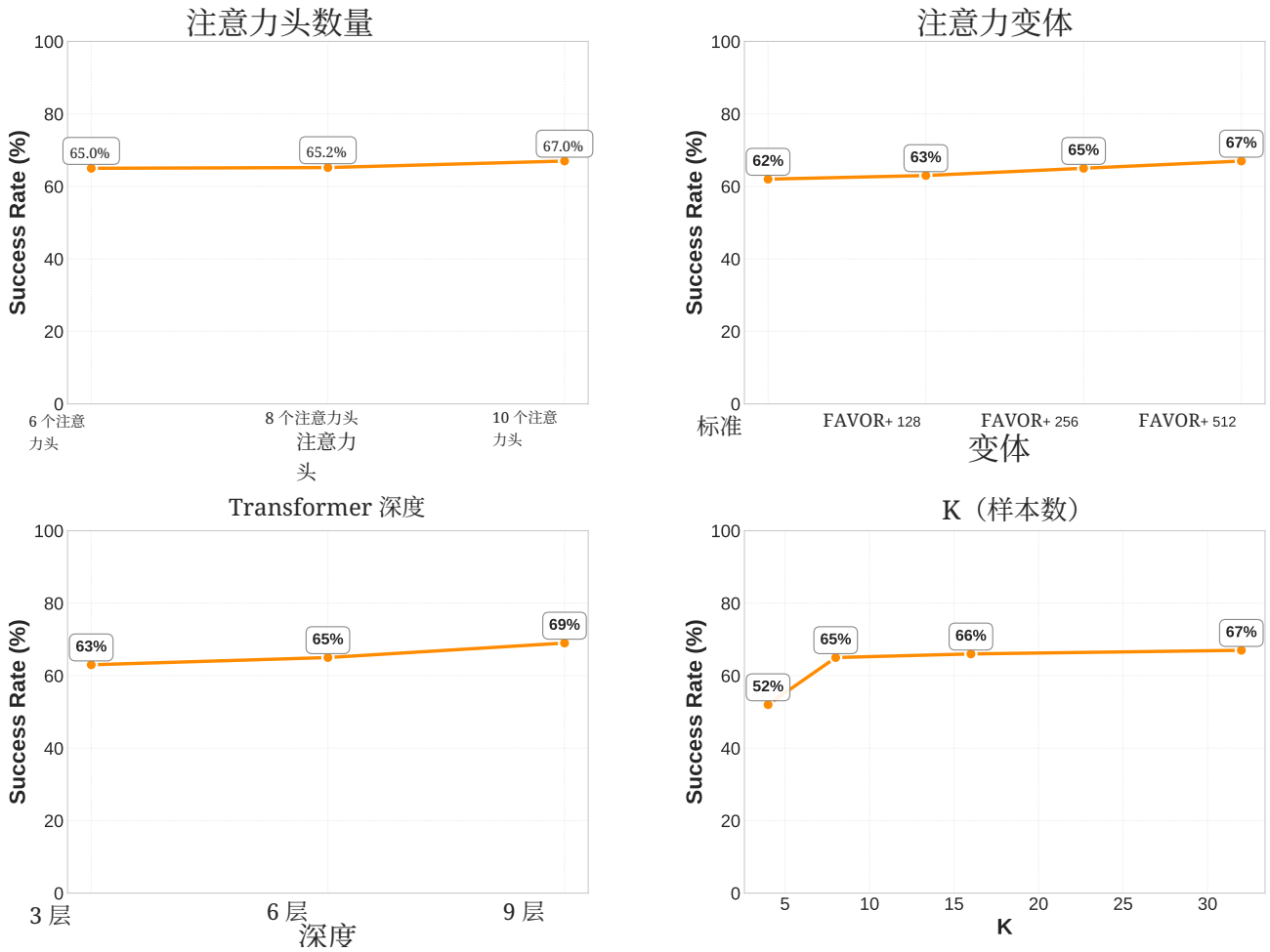

图 7：**CALVIN 上的架构消融实验。** (a) 注意力头、(b) FAVOR$^+$ 随机特征、(c) Transformer 深度以及 (d) 潜在候选项数量 K 对成功率的影响。性能在 $H = 8$ 个注意力头时趋于饱和，随机特征数量增加至 $m = 512$前可带来提升，随深度增加而改善，而候选项数量超过 $K = 16$ 后收益递减。

高效……并且等价于最大化根据生成样本构建的核密度估计器的似然。”该定理严格保证了，最小化匹配距离并非启发式方法，而是一种能够覆盖数据支撑集的有效统计估计程序。

虽然引理 A.1讨论了渐近情况，但我们还进一步分析了与反应式控制相关的有限样本行为。

命题 A.1（距离作为似然下界）。 对于有限样本，重建距离 $d^* = \min_{\mathbf{z}} \|G_\theta(\mathbf{z}) - \mathbf{x}\|_2$ 构成对数似然的几何下界。具体而言，对于带宽为 σ 的高斯核，对数似然满足：

$$\log p(\mathbf{x}) \geq -\frac{1}{2\sigma^2}(d^*)^2 + C.$$

证明：考虑核密度估计 $\hat{p}(\mathbf{x}) = \frac{1}{K}\sum_{k=1}^{K} \mathcal{K}(G_\theta(\mathbf{z}_k), \mathbf{x})$。使用高斯核 $\mathcal{K}(\mathbf{u}, \mathbf{v}) \propto \exp(-\|\mathbf{u} - \mathbf{v}\|^2/2\sigma^2)$以及最大值-求和不等式 $\log \sum_k a_k \geq \log(\max_k a_k)$，我们得到

$$\begin{aligned}\log \hat{p}(\mathbf{x}) &\geq \log\left(\frac{1}{K}\max_k \exp\left(-\frac{\|G_\theta(\mathbf{z}_k) - \mathbf{x}\|^2}{2\sigma^2}\right)\right) \\ &= -\frac{1}{2\sigma^2}\min_k \|G_\theta(\mathbf{z}_k) - \mathbf{x}\|^2 - \log K + C'.\end{aligned}$$

因此，最小化匹配距离 $(d^*)^2$ 即可最大化数据对数似然的这一下界。这证明了我们将重建距离（以及 ProxyScore）用作反应式控制的几何不确定性度量是合理的。

L. 批次全局分位数阈值的一致性

PRISM 的一项关键贡献是批次全局拒绝阈值 ε_{RS}，该阈值通过批次分位数的指数移动平均（EMA）进行动态校准。我们现在证明，该

表 IX：Robomimic（PH）—按任务划分的成功率（均值 ±标准误）；除非另有说明，均为单步（显示 NFE）。

方法	NFE	Lift	Can	Square	Transport	工具悬挂	平均值
多步策略							
扩散策略 [11]	15	1.00	0.99 ± 0.01	0.92 ± 0.03	0.79 ± 0.04	0.55 ± 0.05	85.0
RectifiedFlow* [21]	15	1.00	0.96 ± 0.02	0.90 ± 0.02	0.84 ± 0.04	**0.90 ± 0.02**	92.0
ActionFlow [15]	10	1.00	0.96 ± 0.02	**0.93 ± 0.02**	–	0.51 ± 0.02	–
Falcon-DDPM [9]	12–52	1.00	0.97 ± 0.02	**0.95 ± 0.02**	0.85 ± 0.04	0.55 ± 0.05	86.4
Falcon-DDiM [9]	6–10	1.00	**1.00 ± 0.00**	0.91 ± 0.03	0.74 ± 0.04	0.51 ± 0.05	83.2
音频 [18]	2	1.00	**1.00**	**0.98**	**0.92**	**0.88**	**95.6**
CP* [25]	3	1.00	0.95 ± 0.02	**0.96 ± 0.01**	0.88 ± 0.02	0.77 ± 0.03	91.2
单步策略							
DDiM [11]	1	0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.8
CP* [25]	1	1.00	0.98 ± 0.01	0.92 ± 0.02	0.78 ± 0.03	0.70 ± 0.03	89.6
一致性流匹配* [35]	1	1.00	0.94 ± 0.02	0.90 ± 0.01	0.84 ± 0.02	0.80 ± 0.02	89.0
流匹配策略*[7]	1	0.99± 0.01	0.96± 0.01	0.79± 0.02	0.66± 0.05	0.86± 0.05	0.85
IMLE* [27]	1	1.00	0.94 ± 0.01	0.81 ± 0.01	0.85 ± 0.05	0.75 ± 0.06	87.0
PRISM	**1**	**1.00**	**1.00**	0.84 ± 0.02	**0.91 ± 0.02**	**0.86 ± 0.03**	**92.2**

表 X：语言条件操控成功率（每项任务 50 次试验）。

语言指令	成功率（%）	失败模式
“将蓝色杯子堆叠在绿色杯子上 cup”	92 ± 4	错位和碰撞-撞（8%）
“将黄色杯子堆叠在蓝色 cup”	88 ± 5	错位和碰撞-撞（12%）
“将绿色球放进绿色杯子”	84 ± 6	抓取滑脱（16%）
“将绿色杯子挂在挂钩上”	78 ± 7	空间误差（22%）

估计量在统计上是一致的，并且其方差会随着有效样本量的增加而趋于零。

引理 A.2 (批次全局分位数估计量的一致性)。 令 $\{D_{i,k\to j}\}$ 表示大小为 N 的小批量中生成的候选项与真实目标之间的距离，$\hat{}$并令 $Q_N(q)$ 为该分布的经验 q 分位数。那么当 $N \to \infty$ 时：

(i) 经验累积分布函数 F_N 几乎必然一致收敛于真实累积分布函数 F 。 *(ii)* 分位数估计量满足

$$Var(\hat{Q}_N(q)) = \frac{q(1-q)}{Nf^2(\xi)} + o(N^{-1}),$$

其中 $\xi = F^{-1}(q)$ 为真实分位数，且 $f(\xi) > 0$ 为 ξ处的密度。

(iii) 因此，$\hat{}$ $Q_N(q) \xrightarrow{p} Q(q)$ 的方差为 $O(1/N)$。

证明：

第（i）部分： 一致收敛 上确界 $|F_N(x) - F(x)| \xrightarrow{x} 0$ 几乎必然成立，这直接来自格利文科-坎特利定理 [Van der Vaart 定理 19.1 [32]]。

第(ii)部分： 对于样本分位数，Bahadur表示 [4]建立了渐近展开

$$\hat{Q}_N(q) = Q(q) + \frac{q - F(Q(q))}{f(Q(q))} + R_N,$$

其中余项满足 $R_N = O_p(N^{-3/4}\log N)$。对两边取方差，并利用$F(Q(q)) = q$这一事实：

$$\begin{aligned}\mathrm{Var}(\hat{Q}_N(q)) &= \mathrm{Var}\left(\frac{q - F_N(Q(q))}{f(Q(q))}\right) + O(N^{-3/2}) \\ &= \frac{\mathrm{Var}(F_N(Q(q)))}{f^2(Q(q))} + o(N^{-1}) \\ &= \frac{q(1-q)}{Nf^2(\xi)} + o(N^{-1}),\end{aligned}$$

其中最后一个等式利用了 $F_N(Q(q)) \sim$二项分布$(N, q)/N$这一事实，因此$\mathrm{Var}(F_N(Q(q))) = q(1-q)/N$。

第(iii)部分： 由第（i）部分，$N\mathrm{Var}(\hat{Q}_N(q)) \to \infty$，与第（i）部分结合后确立了概率一致性。

对 PRISM 的启示： 通过对整个小批量中的距离进行聚合（而非像标准 RS-IMLE 那样逐样本计算），PRISM 最大化了有效样本量 N，从而最大限度地降低阈值方差。在典型的批次大小 $N \geq 64$ 下，变异系数 CV($\hat{}Q_N$) =

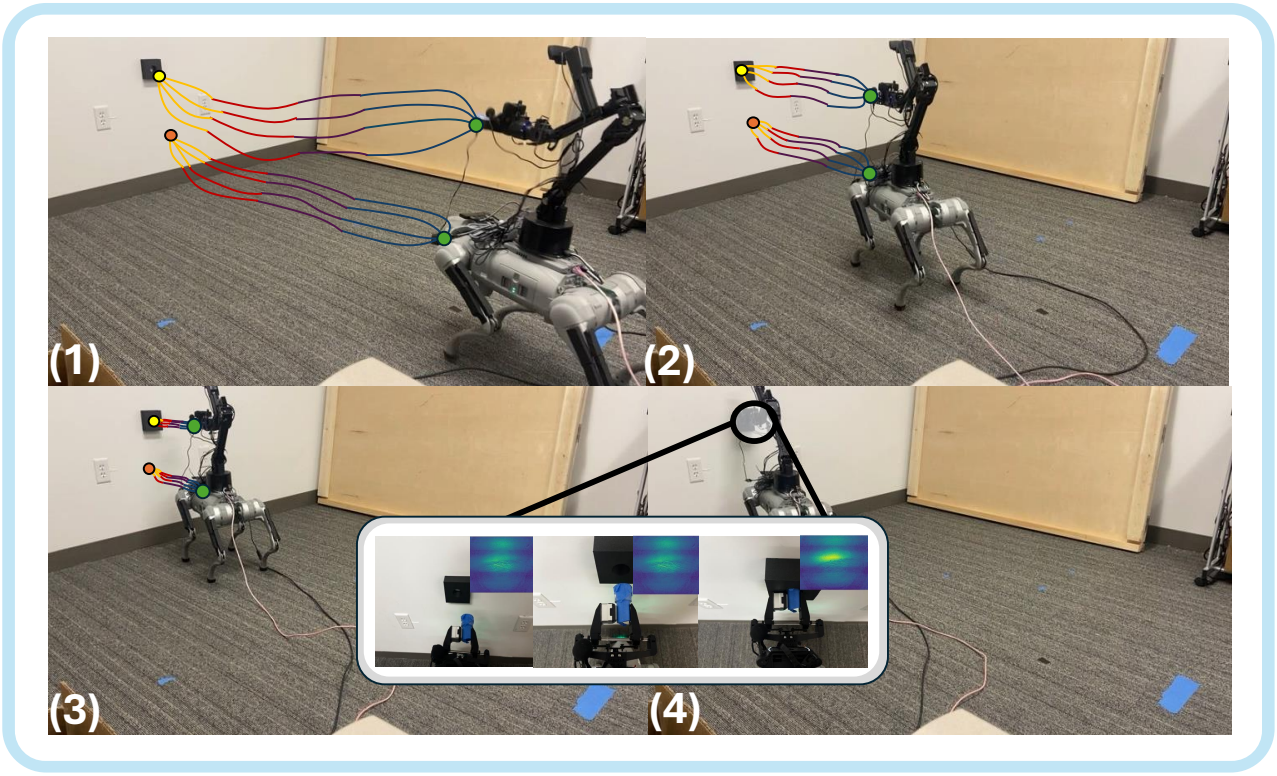

图 8：**真实世界的移动操作平台**。 配备 D1 七自由度机械臂的 Unitree GO2 四足机器人执行插销插入。面板显示：(1-3)带有底盘移动的接近阶段；(4)带有触觉反馈的插入阶段，展示了对精确操作至关重要的接触模式。

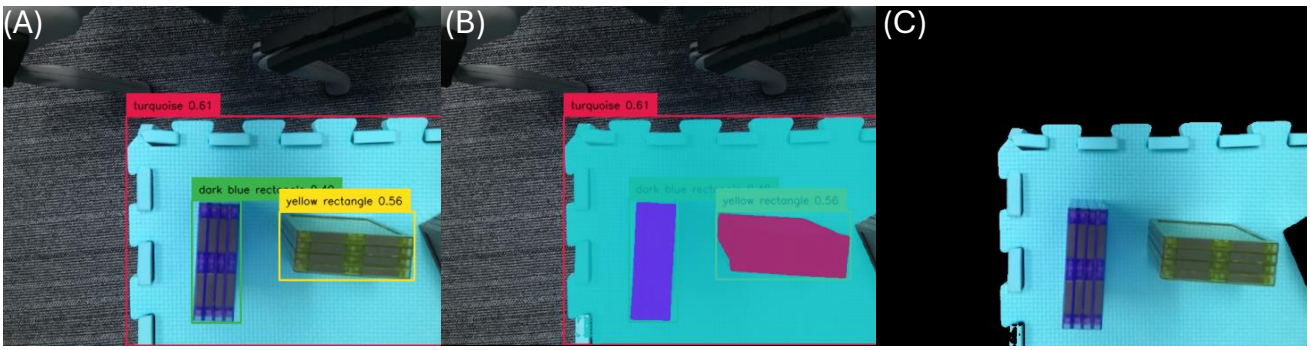

图 9：**视觉预处理流程**。 (A)GroundingDINO 根据物体关键词生成初始边界框。(B)SAM 生成分割掩码。(C)去除背景的 RGB 图像作为 PRISM 的输入，在减少干扰项的同时保持视觉保真度。

$\sqrt{\mathrm{Var}(\hat{Q}_N)/Q(q)} = O(N^{-1/2})$ 保持低于 5%（已在图 11中通过实证验证）。这种统计稳定性使其无需针对不同数据集和任务分布手动调整阈值，即可实现一致的模式覆盖。 □ ■

图 11 通过实验证明引理 A.2： PRISM 的拒绝阈值方差随 $O(1/\sqrt{N})$ 批次大小的增大而减小，并在各种训练配置下保持稳定。

M. 软覆盖作为熵正则化

引理 A.3 （软覆盖作为熵正则化）。 最小化软覆盖损失

$$\mathcal{L}_{soft} = -\sum_i \log \sum_k \exp(-d_{ik}/\tau)$$

等价于最小化 k-中心目标，同时对分配分布施加熵约束，其中d_{ik} 是候选项 k 到目标 i的距离。

证明：标准 IMLE 中的硬分配等价于度量 k-中心问题的贪心算法。

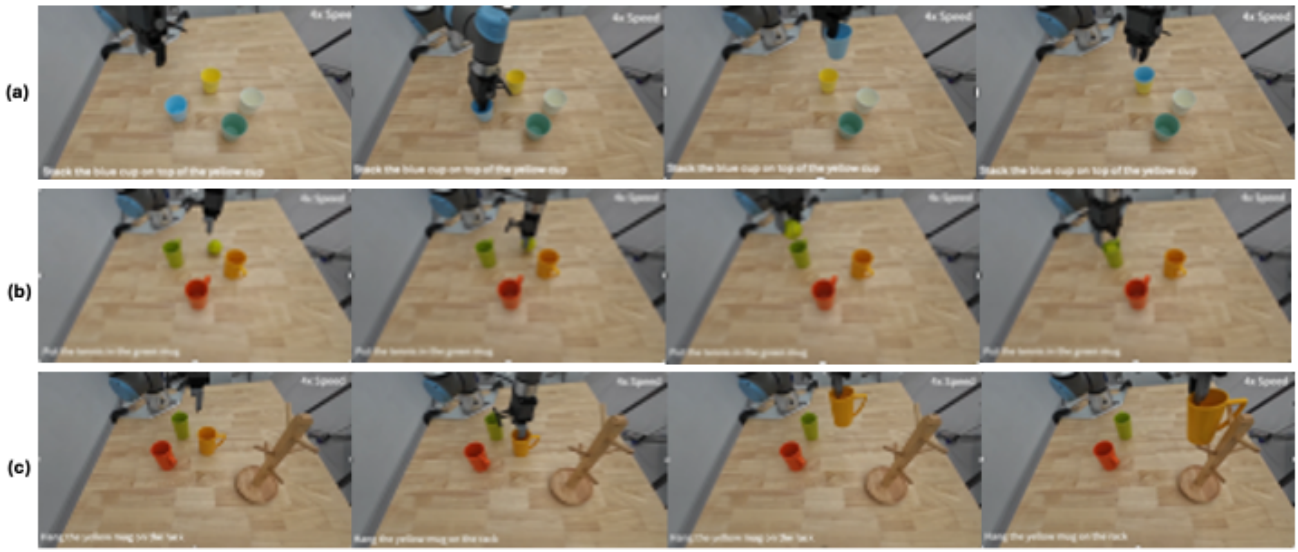

图 10：**语言条件操控示例**。三条不同语言指令的成功执行：(a) 将蓝色杯子叠放在黄色杯子上，(b) 将网球放进绿色马克杯中，(c) 将黄色马克杯挂在架子上，展示了 PRISM 将文本嵌入与视觉上下文融合以实现任务特定控制的能力。

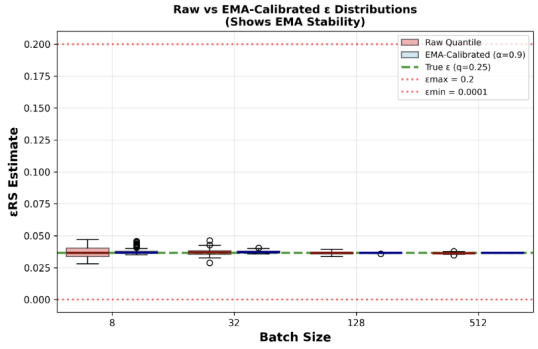

图 11：**批次全局拒绝阈值的一致性**。我们分析估计的拒绝阈值 ϵ_{RS} 随批次大小变化的方差。正如引理A.2所预测的，在整个小批量中全局汇总重建误差（批次全局）相比小样本估计，能显著降低估计量方差。这种稳定性使 PRISM 能够在无需手动调节的情况下保持一致的接受率。

Hochbaum 和 Shmoys [17] 证明了，这种贪心策略能够为最优覆盖半径提供 2-近似，而要获得更好的结果是 NP-hard 的。

我们的软覆盖项引入了温度缩放的对数和指数，它是最小值算子的平滑近似。具体而言，当 $\tau \to 0$ 时：

$$-\log \sum_k \exp(-d_{ik}/\tau) \to \min_k d_{ik}.$$

对于有限的 $\tau > 0$，这会创建一个障碍函数：当分配概率趋近于零时，损失趋于无穷大，从而惩罚“孤立”模式（分配质量为零的簇）。

从理论上讲，这确保 PRISM 会为批次中的每种不同模式至少分配一个候选项，从而防止模式坍塌，同时保持计算效率。■

a) 软覆盖的理论依据：引理A.3 （附录M）证明，最小化$\mathcal{L}_{\text{soft}}$ 等价于求解满足熵约束的 k中心问题。标准 IMLE 中的硬分配对应于度量 k中心问题的贪心算法；Hochbaum 和 Shmoys [17] 证明该算法可提供 2-近似（而求得更好的近似是 NP-hard）。我们的软覆盖项引入温度缩放的对数和指数，将其作为障碍函数；随着分配概率趋近于零，它会以无限损失惩罚“孤立”模式（分配质量为零的簇）。这一理论保证了 PRISM 会为批次中的每个不同模式至少分配一个候选项，从而防止模式坍塌。实验上，图6（附录F）显示，$\lambda_{\text{soft}} \approx 0.02$ 能够在多样性和准确性之间实现最佳平衡。

N. 通过 *Performer* 注意力实现时间一致性

我们的注意力机制的时间稳定性由 Choromanski 等人提出的定理 1 保证。该定理证明，FAVOR$^+$ 机制能够对 softmax 注意力核进行无偏估计，且方差有界：

$$\mathbb{E}[\text{FAVOR}^+(\mathbf{Q},\mathbf{K})] = \text{Softmax}(\mathbf{Q}\mathbf{K}^\top), \quad \text{Var}[\text{FAVOR}^+] = O(1/m),$$

其中，m 是随机特征的数量。由于注意力近似误差有界，且该机制是 Lipschitz 连续的，因此在时域T内的误差传播有界，为 $O(T\sqrt{\text{Var}}) = O(T/\sqrt{m})$，从而防止误差沿长轨迹发生灾难性漂移。

O. 失败案例分析

我们深入分析了失败回合（在 CALVIN 上约占试验的 34%），以识别超越简单成功率的模式。

- **高急动度**（> 0.5）：在 ∼5–10% 的回合中观察到。此现象与所选候选项索引的突然变化相关，进而导致执行轨迹出现物理不连续性。
- **模式切换不稳定性**：在 ∼10–15% 的失败案例中，我们观察到一种“闪烁”行为：策略在连续时间步之间于两种有效模式（例如绕过障碍物时选择向左或向右）反复切换。当两个候选项的 ProxyScore 几乎相同时，就会出现这种情况。
- **传感器遮挡**：约 8–12% 的失败案例发生在关键信息（例如目标物体）被遮挡时。尽管 PRISM 对单模态丢失具有鲁棒性（表II），但多个互补模态（腕部 RGB + 本体感觉）同时受到遮挡会导致任务失败。
- **启示**：失败很少是因为生成了无效动作，更多是因为在有效的多模态候选项之间犹豫不决。这是显式多模态策略中已知的权衡；我们通过时间一致性 ProxyScore 缓解了这一问题，但极端情况仍然颇具挑战。

a) 定量分解。：我们对 CALVIN 上 100+失败回合的分析揭示了以下分布：

- **模式切换不稳定性**：占失败的 10–15%（30–45 个回合）
- **高急动度事件**（> 0.5）：占失败的 5–10%（15–30 个回合）
- **传感器遮挡**：占失败的 8–12%（24–36 个回合）
- **其他因素**：占 1–2%（碰撞、超时等）

关键的是，$< 1\%$ 的失败源于生成了不符合运动学约束的动作，这证实了主要挑战在于轨迹一致性，而非动作有效性。

P. 分任务性能

我们提供了MetaWorld 50 项操作任务的细粒度逐任务成功率，并按难度类别细分（简单：28 项任务，中等：11 项任务，困难：6 项任务，非常困难：5 项任务）。这些详细结果佐证了表格中报告的汇总指标I （正文）和表格 VIII（附录 I），并表明 PRISM 在各任务类别中均持续取得改进，尤其是在多模态动作分布最为关键的困难和非常困难任务上。

Meta-World（简单）

算法/任务	俯视按按钮	俯视按墙上按钮	按墙上按钮	从侧面拔出插销	关门	锁门
扩散策略	98 ± 1	96 ± 3	97 ± 3	74 ± 3	100 ± 0	86 ± 8
三维扩散策略	99 ± 1	96 ± 3	100 ± 0	93 ± 3	100 ± 0	96 ± 3
ManiCM	100 ± 0	96 ± 2	98 ± 3	71 ± 15	100 ± 0	98 ± 2
SDM策略	98 ± 2	99 ± 1	100 ± 0	74 ± 19	100 ± 0	96 ± 2
FlowPolicy*	100 ± 0	100 ± 0	100 ± 0	93 ± 2	100 ± 0	100 ± 0
PRISM	**100 ± 0**	**100 ± 0**	**100 ± 0**	**90 ± 7**	**100 ± 0**	**100 ± 0**

Meta-World（简单）

算法/任务	开门	开门锁	关抽屉	开抽屉	关水龙头	开水龙头	按压把手	拉动把手
扩散策略	98 ± 3	98 ± 3	100 ± 0	93 ± 3	100 ± 0	100 ± 0	81 ± 4	27 ± 22
三维扩散策略	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	52 ± 8
ManiCM	100 ± 0	82 ± 16	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	10 ± 10
SDM策略	100 ± 0	100 ± 0	100 ± 0	100 ± 0	99 ± 1	100 ± 0	100 ± 0	28 ± 11
FlowPolicy*	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0	31 ± 6
PRISM	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	**100 ± 0**	**38 ± 2**

Meta-World（简单）

算法/任务	按压把手侧面	拉动把手侧面	拉动杠杆	滑动板	向后滑动板	转动旋钮	伸手	到达墙壁
扩散策略	100 ± 0	23 ± 17	49 ± 5	83 ± 4	99 ± 0	63 ± 10	18 ± 2	59 ± 7
三维扩散策略	0 ± 0	82 ± 5	84 ± 8	100 ± 0	100 ± 0	91 ± 0	26 ± 3	74 ± 3
ManiCM	0 ± 0	48 ± 11	82 ± 7	100 ± 0	96 ± 5	84 ± 2	33 ± 3	62 ± 5
SDM策略	0 ± 0	68 ± 6	84 ± 9	100 ± 0	100 ± 0	88 ± 3	34 ± 3	80 ± 1
FlowPolicy*	100 ± 0	55 ± 10	91 ± 6	98 ± 2	100 ± 0	88 ± 6	41 ± 8	78 ± 2
PRISM	**100 ± 0**	**53 ± 4**	**76 ± 2**	**100 ± 0**	**100 ± 0**	**93 ± 2**	**55 ± 7**	**83 ± 6**

Meta-World（简单）

算法/任务	盘子侧滑	关闭窗户	打开窗户
扩散策略	100 ± 0	100 ± 0	100 ± 0
三维扩散策略	100 ± 0	100 ± 0	99 ± 1
ManiCM	100 ± 0	100 ± 0	80 ± 26
SDM策略	100 ± 0	100 ± 0	78 ± 18
FlowPolicy*	100 ± 0	100 ± 0	100 ± 0
PRISM	**100 ± 0**	**100 ± 0**	**100 ± 0**

Meta-World（中等）

篮球	料箱拣选	关闭盒子	拉咖啡机把手	推咖啡机把手
85 ± 6	15 ± 4	30 ± 5	34 ± 7	67 ± 4
100 ± 0	56 ± 14	59 ± 5	79 ± 2	96 ± 2
4 ± 4	49 ± 17	73 ± 2	68 ± 18	96 ± 3
28 ± 26	55 ± 13	61 ± 3	72 ± 9	97 ± 2
93 ± 6	51 ± 22	68 ± 2	93 ± 4	98 ± 2
98 ± 2	**63 ± 20**	**70 ± 4**	**93 ± 2**	**97 ± 2**

Meta-World（中等）

算法/任务	锤击	侧向插销	推墙	足球	扫动
扩散策略	15 ± 6	34 ± 7	20 ± 3	14 ± 4	18 ± 8
三维扩散策略	100 ± 0	79 ± 4	78 ± 5	23 ± 4	92 ± 4
ManiCM	98 ± 2	75 ± 8	31 ± 7	27 ± 3	54 ± 16
SDM策略	98 ± 2	83 ± 5	83 ± 4	25 ± 2	90 ± 6
FlowPolicy*	100 ± 0	75 ± 4	61 ± 16	38 ± 10	98 ± 2
PRISM	**100 ± 0**	**88 ± 7**	**75 ± 6**	**49 ± 6**	**98 ± 2**

Meta-World（困难）

扫入	装配	手动插入	从孔中取出
10 ± 4	15 ± 1	0 ± 0	0 ± 0
38 ± 9	100 ± 0	28 ± 8	44 ± 3
37 ± 13	87 ± 3	28 ± 15	30 ± 16
32 ± 15	100 ± 0	24 ± 14	34 ± 24
33 ± 16	100 ± 0	26 ± 2	36 ± 6
39 ± 14	**100 ± 0**	**30 ± 0**	**48 ± 8**

Meta-World（困难）

算法/任务	抓取放置	Push	向后推动	货架放置
扩散策略	0 ± 0	30 ± 3	0 ± 0	11 ± 3
三维扩散策略	0 ± 0	56 ± 5	0 ± 0	47 ± 2
FlowPolicy*	66 ± 2	61 ± 16	–	46 ± 8
ManiCM	0 ± 0	55 ± 2	0 ± 0	48 ± 3
SDM策略	0 ± 0	57 ± 0	100 ± 0	51 ± 4
PRISM	**63 ± 6**	**75 ± 4**	**–**	**68 ± 10**

Meta-World（非常困难）

拆卸	拉杆	推杆	拾取放置墙	平均值
43 ± 7	11 ± 2	63 ± 3	5 ± 1	55.5 ± 3.58
91 ± 4	67 ± 0	100 ± 0	74 ± 4	76.1 ± 2.32
80 ± 4	78 ± 6	100 ± 0	95 ± 0	81.5 ± 3.84
87 ± 3	63 ± 2	100 ± 0	37 ± 16	69.0 ± 4.60
86 ± 10	68 ± 10	0 ± 0	53 ± 12	74.8 ± 4.51
83 ± 8	**78 ± 4**	**100 ± 0**	**98 ± 2**	**84.2**