arXiv:2601.01948v1 [cs.RO] 5 Jan 2026

从原始技能中学习用于机器人操作的扩散策略

Zhihao Gu[1]*、**Ming Yang**[2]、**Difan Zou**[1]、**Dong Xu**[1]†

[1]香港大学计算与数据科学学院计算机科学系 [2]北京航空航天大学软件学院 ellerygzh@gmail.com, viv@buaa.edu.cn, {dzou, dongxu}@cs.hku.hk

摘要

扩散策略（DP）近期在机器人操作动作生成方面展现出巨大潜力。然而，现有方法通常依赖全局指令来生成短期控制信号，这可能导致动作生成出现错位。我们推测，原始技能——即细粒度、短时域的操作，例如“向上移动”和“打开夹爪”——能为机器人学习提供更直观、更有效的接口。为弥合这一差距，我们提出 SDP，这是一种技能条件扩散策略，将可解释的技能学习与条件动作规划相结合。SDP 从跨任务的操作中抽象出八种可复用的原始技能，并利用视觉语言模型从视觉观测和语言指令中提取离散表示。在此基础上，我们设计了一种轻量级路由网络，为每个状态分配所需的原始技能，从而构建单技能策略以生成技能对齐动作。通过将复杂任务分解为一系列原始技能，并选择单技能策略，SDP 确保了不同任务中行为与技能的一致性。在两项具有挑战性的仿真基准和真实世界机器人部署中的大量实验表明，SDP 的表现始终优于最先进方法，为基于技能的扩散策略机器人学习提供了新范式。

引言

让机器人执行多样化的真实世界任务，一直是机器人学和人工智能领域的长期目标。一种很有前景的方法是通过示范教机器人，使其能够直接从示范中学习。然而，这一过程具有独特的挑战：不同于标准的预测问题，机器人控制要求精确且具备上下文感知能力的动作（Chi et al. 2025）。为应对这一问题，先前的研究主要聚焦于改进动作表示（Mandlekar et al. 2021；Shafiullah et al. 2022）以及构建更丰富的机器人行为内部模型（Florence et al. 2022；Wu et al. 2020）。

近年来，扩散模型（Ho、Jain 和 Abbeel 2020；Mokady et al. 2023）作为一类能够学习的生成模型，
通过逆转逐步添加噪声的过程，在高保真图像生成方面取得了显著成功（Ho、Jain 和 Abbeel 2020年；Jo、Lee 和 Hwang 2022年；Rombach

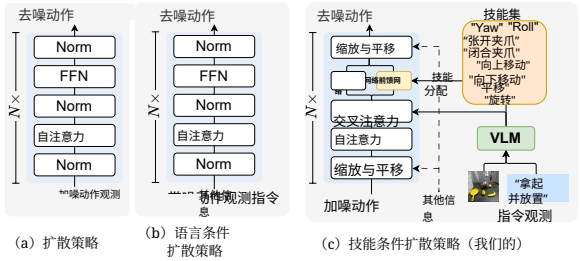
*本文工作完成于 Zhihao Gu 担任 Dong Xu 教授的博士后期间。
†通讯作者。版权所有 © 2026，人工智能促进协会 (www.aaai.org)。保留所有权利。



图 1：一项任务由一系列短期操作组成，我们将其抽象为八种共享的原始技能，为任务提供具体指引。先前的扩散策略将高层指令直接映射为动作。相比之下，我们学习原始技能，并将其融入条件动作生成中，从而实现更精确的控制。

等人 2022年）。在这一成功的基础上，扩散模型已被应用于机器人学，用于生成动作序列。扩散策略（DP）（Chi 等人，2025年）是一项开创性工作，通过扩散模型的条件去噪过程生成机器人行为。它不直接输出动作，而是在多次去噪迭代中推断动作评分梯度，从而显著提升性能。然而，仅以视觉观测为条件，难以同时学习多个任务，这严重限制了其在真实场景中的部署。

为克服这些局限，近期研究引入了自然语言指令作为条件，使机器人能够执行范围更广的任务（Ha、Florence 和 Song 2023；Reuss 等人 2024b；Liu 等人 2024；Wang 等人 2024b；Reuss等人 2024a）。通常，语言编码器会将指令转换为嵌入，然后将这些嵌入与带噪声的动作序列一同输入扩散策略，以生成动作。这一范式的研究沿着三个维度不断发展（Song 等人 2025）：机器人数据表示（Wang 等人 2024b；Ze 等人 2024；Wang 等人 2024a）、模型架构（Team 等人 2024；Reuss 等人 2024b；Ye 等人 2024）以及扩散策略（Ren 等人 2024；Reuss 等人 2024b；Liang 等人 2024）。尽管取得了这些进展，现有的大多数方法仍将高层指令直接映射为短期动作，这可能导致行为含糊或与指令不符。例如，如果机器人要闭合夹爪，那么高层任务描述——



（a）扩散策略　（b）语言条件扩散策略　（c）技能条件扩散策略（我们的）

图2：扩散策略（DP）、语言条件扩散策略和我们的技能条件扩散策略之间的比较。我们的方法借助分配的原始技能提供更精准的引导，从而执行抽象指令。

“拿起柠檬并把它放进平底锅里”——就过于抽象，无法提供明确的指令，而应采用更细粒度的指令。这促使我们设计一种方法，生成具体的短期指令（称为原始技能），例如“闭合夹爪”，并学习单技能扩散策略，以生成更准确的动作。

为弥补这一差距，我们提出了 SDP，这是一种技能条件扩散策略，将细粒度技能学习与条件式低层动作生成相结合。我们的方法基于两个关键思路：（1）根据当前观测，将模糊的高层指令分解为可学习的短期技能；（2）训练扩散策略，使其根据这些技能生成动作。具体而言，我们首先将不同任务中的短期操作抽象为八种原始技能（见图1），这些技能可以进一步组合以构成复杂任务；然后利用视觉语言模型，将视觉观测和高层指令转换为离散表示。接着，轻量级路由网络会为每种状态动态分配恰当的技能。分配的技能会合成扩散策略中前馈网络（FFN）的参数，而本体感知等附加信息则通过 AdaLN操作注入。由此得到的单技能扩散策略能够生成与该技能一致、连贯且精准的行为。与先前的方法（见图2(a)和(b)）相比，我们的 SDP 能够端到端地理解并执行复杂指令，准确度更高，控制也更加精准。在两个具有挑战性的仿真基准以及真实世界机器人部署上开展的大量实验，证明了 SDP 的卓越性能。总之，我们的主要贡献如下：

- 我们提出了 SDP，这是一种技能条件扩散策略，将细粒度技能学习与低层动作生成相结合，以缓解全局指令与短期动作在粒度上的不匹配。
- 我们提出了八种可复用的原始技能，可泛化至多种操作任务，为机器人学习提供结构化且可解释的动作空间。此外，为了有效利用这些技能，我们设计了一种轻量级路由网络，能够动态评估状态的相关性并选择最优技能，从而确保生成的行为具有自适应性并与任务相匹配。
- 我们设计了一种新颖的单技能扩散策略，可生成与各项技能精确匹配的动作。通过根据分配的技能动态参数化策略的前馈神经网络层，我们能够有效捕捉原始技能与低层控制信号之间的依赖关系。
- 在模拟任务和真实世界任务中，与基线方法相比，我们展现出更强的多任务能力和泛化能力。对技能的可视化也揭示了该方法分解抽象指令并组合原始技能的能力，验证了其有效性和可解释性。

相关工作

机器人操作中的扩散策略。扩散模型（Ho、Jain 和 Abbeel 2020；Mokady 等人 2023年）近来在多个领域取得了显著成功，其在机器人操作中的潜力也日益受到关注。在机器人领域，研究人员致力于开发基于扩散的策略，使机器人能够遵循语言指令并执行多种任务（Ha、Florence 和 Song 2023年；Reuss 等人 2024b；Liu 等人 2024年；Wang 等人 2024b；Reuss等人 2024a）。这些工作主要涉及三个方面：机器人数据表示、模型架构和扩散策略。数据表示包括二维轨迹（Wang 等人 2024b）、三维点云（Ze 等人 2024年）以及感官输入的组合（Wang 等人 2024a）。模型架构通常将扩散模型与大语言模型（Team 等人 2024年）、Transformer（Reuss 等人 2024b）或变分自编码器（Ye 等人 2024年）相结合。此外，研究人员还探索了不同的训练策略，例如融合强化学习（Ren 等人 2024年）、自监督学习（Reuss 等人 2024b）和分类器引导（Liang 等人 2024年；Mete 等人 2024年）。相比之下，我们的 SDP 着重于学习可执行技能，并训练技能条件扩散策略，以实现更精准、连贯的机器人控制。

基于视觉语言模型的规划。将复杂指令分解为易于处理的子目标，有助于机器人更可靠地完成复杂任务（Dalal、Pathak 和 Salakhutdinov，2021年；Dhakan 等，2022年；Hiranaka 等，2023年；Liu 等，2025年）。然而，人工标注这些子目标既耗费人力，也难以扩展。为解决这一问题，近期的方法（Singh 等，2022年；Zhang 等，2023年；Ni 等，2024年）利用蕴含丰富现实世界知识的视觉语言模型，自动为机器人学习生成任务计划。其他方法，如 Garg 等（2022年）的方法，则从潜变量中学习子目标码本，其中每个码都可能对应多个状态。我们的 SDP 在两个关键方面有所不同。首先，我们提出一组人类可理解的原始技能，这些技能可以灵活组合，以完成各种任务。其次，不同于以往对这些技能进行隐式建模的工作，我们使用一个由视觉语言模型输出引导的轻量级神经网络，为每个状态显式分配一项技能，从而实现更透明、可控的行为。

参数合成。超网络（Ha、Dai 和 Le（2016））是一类神经网络，旨在以上下文信息为输入，生成其他网络的参数。这种方法能够高效地建模任务与最优控制策略之间的依

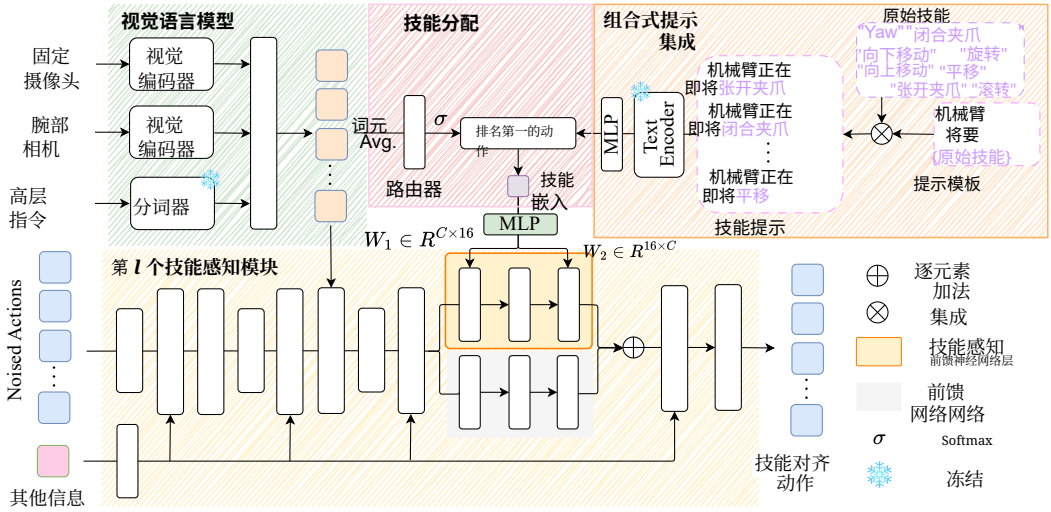


图3：所提出的技能条件扩散策略（SDP）概览。SDP 将不同任务中的短期操作抽象为八种原始技能，并引入统一提示模板来指定即将执行的操作。随后设计了一个路由器，根据由视觉语言模型（VLM）从视觉观测和语言指令生成的嵌入 z_{vl}，为所有候选技能分配重要性分数。此外，选择得分最高的技能，用于参数化扩散策略中的一个额外前馈神经网络层。其他信息（例如本体感知）由多层感知机编码，并通过 AdaLN 操作进一步注入，从而得到一个预测技能对齐动作以实现精确控制的单技能策略。

赖关系（Ren 等，2025年）。例如，HyperDistill（Xiong 等，2024年）使用超网络为具有不同物理结构的机器人学习策略，以极低的计算成本实现了出色的性能。受其启发，SDP 通过参数化扩散策略中的前馈神经网络层，建立技能与动作预测之间的依赖关系。

预备知识

机器人操作旨在学习一个能够执行多种任务的通用策略。假设我们有一组机器人示范 $\mathcal{T} = \{\tau_i\}_{i=1}^{|\mathcal{T}|}$，其中每条轨迹 $\tau_i = \{(s_n, a_{n,k}, l_i)\}_{n=1}^{N}$ 包含状态 $s_n \in \mathbb{R}^{d_s}$、长度为 k 的动作序列 $a_{n,k} \in \mathbb{R}^{7\times k}$，该序列从时间步 n 开始，以及指定任务的高级语言指令 l_i。语言条件策略旨在训练策略 $\pi_\theta(a|s, l) : (s_n, l_i) \mapsto a_{n,k}$，将时间步 n 的状态 s_n 和指令 l_i 映射为未来动作序列。

语言条件扩散策略利用扩散模型获得策略 $\pi_\theta(a|s, l)$。为了根据历史状态嵌入 $\bar{s}$ 和指令 l 从噪声中生成新样本，它通过去噪分数匹配（Vincent2011）训练神经网络 D_θ，以近似扩散过程的得分函数：

$$\mathcal{L}_{\mathrm{SM}}(\theta; s, l) = \mathbb{E}_{\sigma,\bar{a},\epsilon}\left[\frac{1}{\sigma_t}\left\|D_\theta(\bar{a}+\epsilon, \bar{s}, l, \sigma_t) - \bar{a}\right\|_2^2\right], \quad (1)$$

其中，ϵ 是噪声，σ_t 是步骤 t 处的密度。扩散模型通过最小化来自 $\mathcal{T}$ 的状态-动作-指令元组的平均损失进行训练。训练好 D_θ 后，采用 DDIM（Zhang、Tao和Chen 2022）在 N_d 个去噪步数内采样所需的动作。我们请读者参阅 MoDE（Reuss等人 2024a）了解更多详情。

方程 (1) 将全局指令 l 直接映射为局部动作。我们认为，任务规范过于抽象，无法提供明确指令来引导扩散策略生成精确的短期动作。

提出的方法

方法概述

本文提出技能条件扩散策略，以解决高层指令导致执行不精确的问题。图示如图3所示：上半部分预测描述即将进行的操作的基本技能，下半部分则提供单技能策略，整合状态信息并生成与技能相匹配的动作。值得注意的是，我们使用视觉语言表征来分配技能，从而使任务执行过程对人类而言既可解释又易于理解。

基本技能分配

为了根据语言指令执行任务，策略会针对每个状态预测短期动作。然而，高层指令的粒度可能会给细粒度动作生成带来歧义。为此，我们提出将任务分解为基本的、不可再分的操作原语，称为 基本技能。这些技能为生成准确的短期控制提供了精确且可执行的指导。

组合式提示集成（CPE）。遵循这一思路，我们显式地学习这些原始技能。具体而言，我们将基本操作抽象为八种可复用的原始技能，记为 P，即“滚转”、“偏航”、“打开

夹爪”、“向上移动”、“平移”、“关闭夹爪”、“向下移动”和“旋转”。为了更好地描述机器人的状态，我们专门设计了一个统一的文本模板“$the\ robot\ arm\ is\ going\ to\ \{skill\}$.”。受 CLIP 中提示集成的启发（Radford 等人，2021），我们进一步提出组合式提示集成，用于为每种技能生成提示，形式如下：

$$P_{En} := \text{“}the\ robot\ arm\ is\ going\ to\ \{skill\}\text{.”} \otimes P \quad (2)$$

其中，$\otimes$ 表示集成操作。之后，冻结的 CLIP 文本编码器 CLIP$_{\text{text}}(\cdot)$后接一个多层感知机 f，将集成后的 P_{En} 编码为提示嵌入 $\boldsymbol{p}$：$\boldsymbol{p} = f(\text{CLIP}_{\text{text}}(P_{En})) \in \mathbb{R}^{8\times C_{\text{img}}}$，其中 C_{img} 是技能分配联合空间的维度。最后，将从$\{\boldsymbol{p}_i\}_{i=1}^{8}$ 中选择一个来指导动作生成。

请注意，几乎所有任务都可以由这些原始技能组合而成，而文本模板则为机器人的状态提供通用提示。因此，来自 CPE 的文本可以重复使用，我们会预先计算并存储这些文本，以提高推理效率。

视觉语言模型。由于 CPE 为每项技能提供了具体提示，我们需要确定将执行哪项技能。具体而言，我们利用视觉观测和高层指令中的视觉语言表征，为技能分配提供指导。形式化地，设$\boldsymbol{I}_s, \boldsymbol{I}_w \in \mathbb{R}^{3\times H\times W}$ 分别表示来自固定摄像头和腕部摄像头的视觉观测。共享图像编码器 $f_{\text{img}}(\cdot): \mathbb{R}^{3\times H\times W} \to \mathbb{R}^{N_{\text{img}}\times C_{\text{img}}}$将它们编码为视觉嵌入 $f_{\text{img}}(\boldsymbol{I}_s)$ 和 $f_{\text{img}}(\boldsymbol{I}_w)$，其中$N_{\text{img}}$ 和 C_{img} 分别表示视觉词元的数量和维度。同时，高层指令 $\boldsymbol{l}$ 经过分词器和词嵌入层（来自 Xiao 等人，2024年）处理，得到文本嵌入 $f_{\text{t}}(\boldsymbol{l}) \in \mathbb{R}^{N_t\times C_{\text{text}}}$。随后，将视觉词元和文本词元拼接起来并输入 TransformerΦ，以获得视觉语言表征 $\boldsymbol{z}_{vl} = \Phi([f_{\text{t}}(\boldsymbol{l})\ f_{\text{img}}(\boldsymbol{I}_s)\ f_{\text{img}}(\boldsymbol{I}_w)]) \in \mathbb{R}^{(N_t+2N_{\text{img}})\times C_{\text{img}}}$。为实现维度对齐，我们省略了对 $f_{\text{t}}(\boldsymbol{l})$的额外投影。

原子技能选择。基于技能提示的嵌入 $\boldsymbol{p}$ 和视觉语言表征 $\boldsymbol{z}_{vl}$，我们进一步设计了技能分配模块，该模块采用轻量级路由网络为每个状态选择一项技能。具体而言，我们首先对 $\boldsymbol{z}_{vl}$ 的词元维度取平均值，得到变量$\boldsymbol{z}_{\text{avg}} \in \mathbb{R}^{C_{\text{img}}}$。然后，通过一个多层感知机层将$\boldsymbol{z}_{\text{avg}}$ 映射为逻辑值，以反映每项技能的重要性；接着使用 Softmax 函数 $\sigma(\cdot)$ 和 top-1$(\cdot)$ 操作，将所有技能筛选为最合适的一项：

$$R(\boldsymbol{z}_{vl}) = \text{top-1}(\sigma(\text{MLP}(\text{Avg}(\boldsymbol{z}_{vl})))), \quad (3)$$

随后，根据技能在 $R(\boldsymbol{z}_{vl}) \in \mathbb{R}^8$中的重要性，选择得分最高的技能。最后，通过 $\boldsymbol{z} = \sum_{i=1}^{8} R(\boldsymbol{z}_{vl})_i \cdot \boldsymbol{p}_i$为每个状态选择技能嵌入。

分析。向量量化（Van Den Oord、Vinyals 等人，2017）被用于隐式地学习离散潜在编码（Garg 等人，2022；Liang 等人，2024）。不同的是，我们的 SDP显式地从不同任务中抽象出共享的原始技能，且技能分配更便于人类理解。最近，GSC（Mishra 等人，2023）显式地参数化由 aDP 预测的技能。相反，我们更细粒度的原始技能（PS）可以组合成这些技能，并表达范围更广的任务。此外，我们的技能是在统一模型中学习的，而非通过多个独立模型学习，因此效率更高。

技能条件扩散策略学习

最终目标是预测技能对齐动作。我们提出学习单技能扩散策略。我们首先注入状态先验，然后在分配的基本技能与条件动作生成之间建立依赖关系。

先验注入。对于每个状态，都会提供时间步、本体感知信息、视觉观测和高层指令。遵循 Doshi 等人（2024年）的工作，使用基于小型 MLP 的编码器来处理时间步和本体感知信息。随后，通过改进的 AdaLN（Perez 等人，2017年）将它们注入网络，该 AdaLN 生成在所有层之间共享的不同调制信号。相较之下，对于视觉和语言信息，VLM 的输出词元首先经过带有 RMSNorm（Zhang 和 Sennrich，2019年）的线性层投影，然后如图3所示，通过每个模块中的交叉注意力进行注入。这些操作整合了状态先验，实现了条件注入。与为每一层分配独立参数的标准 AdaLN 相比，我们的方法在保持性能的同时减少了可学习参数。

技能依赖的前馈神经网络层。为了建立基本技能与动作生成之间的依赖关系，我们在原始 FFN$_{\text{ori}}$中额外引入了类似 LoRA（Hu 等人，2022年）的前馈神经网络（FFN）层。新的 FFN 包含一个 SwishGLU 激活函数和两个矩阵 $\boldsymbol{W}_{\boldsymbol{z}}^1 \in \mathbb{R}^{C\times 16}$和 $\boldsymbol{W}_{\boldsymbol{z}}^2 \in \mathbb{R}^{16\times C}$，它们由 MLP 根据技能嵌入 $\boldsymbol{z}$生成，并分别将维度为 C 的输入 $\boldsymbol{x}$ 映射到 16 维，以及将 16 维映射到 C。因此，最终的 FFN 层可表示为：

$$\text{FFN}(\boldsymbol{x}) = \boldsymbol{W}_{\boldsymbol{z}}^2(\text{SwishGLU}(\boldsymbol{W}_{\boldsymbol{z}}^1\boldsymbol{x})) + \text{FFN}_{\text{ori}}(\boldsymbol{x}), \quad (4)$$

这种类似低秩适配的前馈网络在特征提取时显式考虑技能，从而节省内存并减少总参数量。

训练目标。此外，我们采用正交损失 $\mathcal{L}_{\text{Orth}}(\theta)$ 来降低 $\boldsymbol{p}_{i,\ j}$ 上的成对余弦相似度，并使用超参数 γ。因此，损失函数为：

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{SM}}(\theta) + \gamma\mathcal{L}_{\text{Orth}}(\theta), \quad (5)$$

where $\mathcal{L}_{\text{Orth}} = \frac{1}{64}\sum_{i=1}^{8}\sum_{j=1}^{8}\text{Cos}(\boldsymbol{p}_i, \boldsymbol{p}_j)$ and $\gamma = 0.01$.

分析。公式（4）可视为混合专家的一种变体（Jacobs 等人，1991年），其中第一项被设计为技能依赖型专家，而第二项则是共享专家。因此，它构建了一个能够预测技能对齐动作的单技能扩散策略。本文将其称为技能条件扩散策略。

实验

本节介绍基准测试和实现的详细信息。我们进行了全面评估，以研究以下问题：

- **性能**。我们的 SDP 能否在各种设置下，相较于最先进水平的竞争方法实现出色性能?
- **有效性**。我们提出的架构设计选择会如何影响最终性能?
- **可解释性**。SDP 如何完成各种任务?

训练→测试	方法	连续指令数量（1000 条链）					
		1	2	3	4	5	**平均长度**
ABCD→D	DiffPolicy	86.3%	72.7%	60.1%	51.2%	41.7%	3.16±0.06
	RoboFlamingo	96.4%	89.6%	82.4%	74.0%	66.0%	4.09±0.00
	GR-1	94.9%	89.6%	84.4%	78.9%	73.1%	4.21±0.00
	MDT	98.6%	95.8%	91.6%	86.2%	80.1%	4.52±0.02
	MoDE†	95.4%	89.1%	83.8%	78.5%	73.4%	4.19±0.03
	SDP（我们的方法）	**99.7%**	**96.7%**	**93.8%**	**90.8%**	**86.5%**	**4.67±0.02**
ABC→D	RT-1	53.3%	22.2%	9.4%	3.8%	1.3%	0.90±0.06
	DiffPolicy	63.5%	35.3%	19.4%	10.7%	6.4%	1.35±0.05
	RoboFlamingo	82.4%	61.9%	46.6%	33.1%	23.5%	2.47±0.00
	OpenVLA†	91.3%	77.8%	62.0%	52.1%	43.5%	3.27±0.00
	GR-1	85.4%	71.2%	59.6%	49.7%	40.1%	3.06±0.00
	UniVLA	95.5%	85.8%	75.4%	66.9%	56.5%	3.80±0.07
	SkillDiffuser	94.4%	82.7%	72.1%	62.4%	55.4%	3.66±0.07
	MoDE†	95.8%	88.1%	79.3%	70.8%	62.4%	3.92±0.05
	SDP（我们的方法）	**99.3%**	**96.1%**	**90.9%**	**85.3%**	**76.9%**	**4.49±0.04**

表 1：CALVIN 上的性能。报告各项任务的成功率，以及完成连续 5 条指令的平均执行步数。 ± 0.00 表示不提供平均性能的方法，† 表示使用官方代码重新实现的方法。

实验设置与实现细节

仿真基准。 我们在**CALVIN**（Mees 等人，2022年）和 **LIBERO** （Liu 等人，2023年）基准测试上评估所提出的 SDP。**CALVIN** 由四种不同的场景配置（划分为 A-D）组成，包含 34 项不同任务和 24,000 条语言标注示范。在我们的研究中，我们采用具有挑战性的 ABC→D 评估设置：策略使用来自环境 A、B 和 C 的示范进行训练，并在环境 D 中进行零样本评估；以及 ABCD→D。评估协议包含一个由 1,000 条独特指令链组成的测试集，每条指令链均由五项连续任务构成。性能以连续完成 1 至 5 项任务的序列成功率，以及已完成任务序列的平均长度来衡量。**LIBERO** （Liu 等人，2023年）包含多个任务套件，反映机器人操作的不同方面。我们的实验聚焦于在目标套件内进行监督微调，包括针对空间关系的 LIBERO-Spatial、针对各种物体操作的 LIBERO-Object、针对不同目标的 LIBERO-Goal，以及用于延长任务时长的 LIBERO-Long；每个套件均包含 10 项任务，每项任务有 50 条人工遥操作示范。

真实世界评估。 我们设计了 9 项任务，用于评估多任务学习能力和视觉泛化能力。每项任务均通过一台 6 自由度节卡机械臂采集 30 条轨迹。报告 20 次试验的平均成功率。

- **多任务学习**。 1) 空间感知 （拿起柠檬并放入平底锅；打开微波炉并将薯片放入其中）。2) 工具使用 （将方块扫入簸箕；用杯中的勺子搅拌碗里的水）。3) 语义理解 （将杯中的水倒入碗中；将黄色方块堆叠在另一个方块上）。

- **视觉泛化**。 它包括两个方面：1）对未见物体（苹果或香蕉）进行操作。2）在复杂干扰物中拿起并放下柠檬。

方法	空间	物体	Goal	Long	平均值
DiffPolicy	78.3±0.0	92.5±0.0	68.3±0.0	50.5±0.0	72.4±0.0
Octo	78.9±1.0	85.7±0.9	84.6±0.9	51.1±1.3	75.1±0.6
MDT	78.5±0.0	87.5±0.0	73.5±0.0	64.8±0.0	76.1±0.0
OpenVLA	84.7±0.9	88.4±0.8	79.2±1.0	53.7±1.3	76.5±0.6
MaIL	74.3±0.0	90.1±0.0	81.8±0.0	78.6±0.0	83.5±0.0
UniActions	65.0±0.0	78.0±0.0	68.0±0.0	47.0±0.0	64.5±0.0
UniVLA	95.2±0.0	95.4±0.0	91.9±0.0	87.5±0.0	92.5±0.0
Ours	**98.3±1.3**	**99.8±0.4**	**95.6±0.5**	**93.8±0.8**	**96.9±0.7**

表 2：LIBERO 四个套件上的成功率（%）。标准差为零表示平均性能没有波动。

实现细节。 我们基于 12 个模块的扩散变换器（Peebles 和 Xie，2023年）构建 SDP，并按照（Team 等，2024年）的方法在 OpenX（Vuong 等，2023年）上进行预训练。对于仿真任务，模型在 4 个 A100 GPU 上微调 40 个 epoch，使用 AdamW 作为优化器，学习率为 10^{-4}。批量大小设为 64，固定摄像头和腕部摄像头拍摄的图像均缩放至 224×224。使用 $N_d = 4$ 个去噪步数生成动作，并报告 3 个随机种子下所有任务的平均性能。对于真实世界评估，我们仅使用固定摄像头拍摄的图像，并训练模型 200 个 epoch。所有结果均为 20 次试验的平均值。基线方法使用默认超参数在真实世界数据上进行微调。

模拟机器人操作性能

基线方法。 我们采用在 CALVIN 基准上报告结果的最先进扩散策略作为基线方法，包括使用 CNN 骨干网络的扩散策略（Chi 等人，2025年）（DiffPolicy）、Octo（Team 等人，2024年）、MDT（Reuss 等人，2024b）和 MoDE（Reuss等人 2024a）。Octo（Team 等人，2024年）采用统一动作表示来处理异构动作空间。MDT 利用扩散模

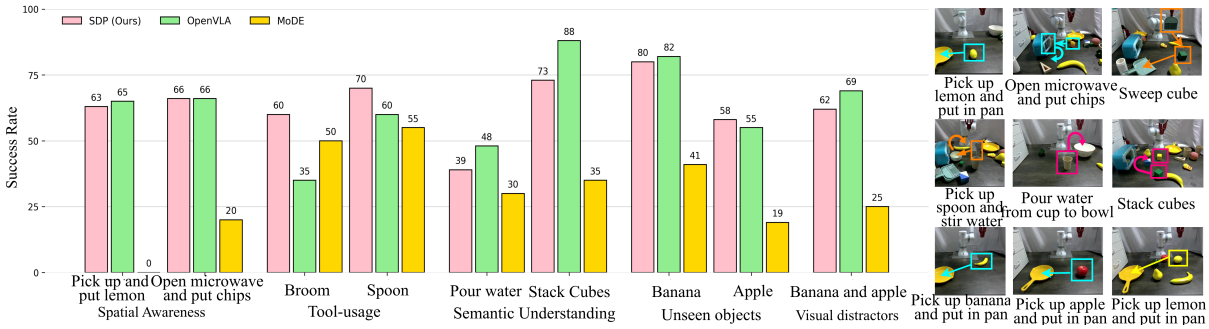


图4：真实世界机器人操作任务的成功率（%）。我们专门设计了9项任务（见右图），用于评估策略能力的两个方面：多任务学习（前六项任务）和视觉泛化（最后三项任务）。在视觉泛化设置中，我们进一步考察了对未见物体（苹果和香蕉）的泛化能力，以及对视觉干扰物的鲁棒性。所提出的 SDP（粉色）始终优于基线方法（绿色和橙色），展现出更强的跨任务和跨物体泛化能力，以及对干扰物的鲁棒性。

方法	ABCD→D	ABC→D	LIBERO-Long
基线方法（DP）	1.98±0.09	1.13±0.02	50.5±0.5%
+ 交叉注意力	4.09±0.07	3.59±0.03	81.0±0.8%
+ 先验注入	4.30±0.07	4.01±0.04	86.0±0.3%
+ 技能抽象	4.51±0.07	4.32±0.07	91.5±0.7%
+ CPE	**4.67±0.02**	**4.49±0.05**	**93.8±0.8%**

(a) 关键组件研究。

策略	ABCD→D	ABC→D	LIBERO-Long
加法	4.34±0.02	4.12±0.04	90.9±0.3%
拼接	4.41±0.04	4.24±0.06	91.8±0.5%
FiLM	4.49±0.03	4.31±0.02	92.5±0.6%
式（4）	**4.67±0.02**	**4.49±0.05**	**93.8±0.8%**

（b）技能条件化策略研究。

表3：CALVIN 和 LIBERO-Long 上的消融实验。

型，根据多模态目标生成灵活的动作序列。MoDE 将稀疏专家与噪声条件自注意力机制相结合，从而在不同噪声水平下实现更有效的去噪。其他基线方法还包括当前最先进的 VLA策略，包括 RoboFlamingo（Li 等人，2023年）、GR-1（Wu 等人，2023年）、OpenVLA（Kim 等人，2024年）和最新的 UniVLA（Bu 等人，2025年）。RoboFlamingo 提出了一种替代性的 VLA，使用连续动作头预测而非离散预测。GR-1 在预训练后学习预测未来帧和动作。OpenVLA 在大规模数据集上进行预训练，以实现通用机器人策略。UniVLA 利用潜在动作模型从视频中提取以任务为中心的动作表示。此外，还针对 LIBERO 比较了 MaIL（Ji a 等人，2024年）和 UniActions（Zheng 等人，2025年）。

在 CALVIN 上的性能。表 1 的结果表明，所提出的 SDP 在两项挑战中均持续优于所有最先进水平的策略。此外，SDP 在动作生成时仅采用四个去噪步数，显著少于 MDT 和 MoDE 等基于扩散的方法基线所用的十步。具体而言，在 ABCD→D 设置中，SDP 以显著优势超越了此前最先进的 MDT 和 MoDE。在更具挑战性的 ABC→D 设置中，SDP 连续完成全部五项任务的成功率达到 76.9%，比此前表现最佳的方法 MoDE 高出 14.5%，比近期提出的 UniVLA 高出 20.4%。连续完成任务的平均数量从 UniVLA 的 3.80 提升至 4.49。这些结果不仅证实了 SDP 的出色性能，也表明它能够泛化至未见过的环境设置，并应对长时域操作任务。

在 LIBERO 上的性能。如表2所示，我们的 SDP 在全部四个评估套件中均表现出色，取得了很高的任务完成率，并显著优于 MaIL 和 UniVLA 等强基线方法。值得注意的是，SDP 是唯一在 LIBERO-Long 套件上成功率超过 90% 的策略，而其他通用方法在应对复杂的长程任务时表现不佳，只有近期提出的 UniVLA 达到了有竞争力的性能。此外，SDP 的平均性能达到 96.9%，分别领先基于扩散的方法 MDT 和 UniVLA 13.4% 和 4.4%。总体而言，所提出的 SDP 在各种机器人操作场景中展现出多样性和鲁棒性，并在 LIBERO基准测试上取得了新的最先进水平。

真实世界机器人操作性能

基线方法。 我们将 SDP 与采用混合专家结构的最先进水平 MoDE，以及采用大型自回归架构的代表性模型 OpenVLA 进行比较。

多任务学习。 评估结果如图4所示。所提出的 SDP 始终表现最佳，在空间感知、工具使用和语义理解方面展现出明显优势。值得注意的是，在“打开微波炉并放入薯片”和“倒水”等复杂任务中，SDP 的表现显著优于其他方法，表明它能够更出色地学习并泛化到多样化的操作任务中。这进一步凸显了 SDP 在多任务学习场景中处理复杂且多样化任务的有效性。

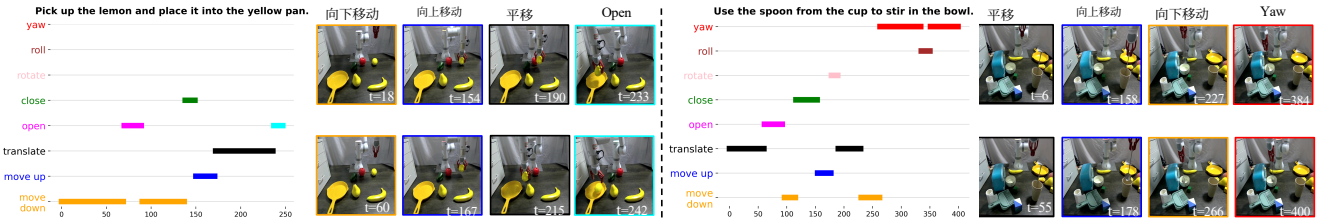


图5：分配的技能可视化。左侧图表展示了每个时间步分配的技能，其中横轴表示时间步，纵轴对应不同的原始技能。右侧图像对应于执行这些技能时的观测。SDP 在训练期间学习原始技能，并在推理过程中将它们组合起来以完成复杂任务。

	#参数	浮点运算次数	推理时间	ABC→D	ABCD→D
Diff-P-T	286M	36.3G	22.1毫秒	1.13±0.02	1.98±0.09
MoDE	780M	57.4G	30.5ms	3.92±0.07	4.19±0.03
Ours	1017M	74.5G	45.1ms	4.49±0.05	4.67±0.02

表4：CALVIN 基准上的复杂度分析。

视觉泛化能力。 此设置评估以下两方面：1）操作此前未见过的物体的能力；2）对无关物体（即视觉干扰物）的鲁棒性。对于前者，我们引入此前未见过的苹果和香蕉。对于后者，我们重复执行该任务，并使用柠檬，同时在附近放置更多物体作为干扰物。相应结果如图4所示。所提出的 SDP（粉色）能够操作苹果，而基线方法（绿色和橙色）则难以拾取和放置苹果。这是因为 SDP 会将任务分解，并生成与柠檬任务相符、具备技能感知能力的动作；柠檬与苹果的形状相似。然而，面对训练期间未见过的形状（如香蕉）时，性能会下降，这表明对于不熟悉的形状，泛化更具挑战性。更重要的是，视觉干扰物的存在仅略微影响成功率（从 75% 降至 65%），而基线方法会受到视觉干扰物的干扰，表现不佳，这凸显了 SDP 出色的鲁棒性。

设计选择的有效性

关键组件研究。 SDP 包含几个关键组件：先验注入（包括交叉注意力和 AdaLN）、技能抽象（由技能依赖型前馈网络使用）以及组合式提示集成（CPE）。表 3 (a) 评估了这些组件对 CALVIN 和 LIBERO-Long 套件的贡献。首先，基线模型在所有任务上的性能都相对较低。一方面，通过交叉注意力融入视觉语言信息显著提升了结果，表明其在特征注入方面的有效性。通过 AdaLN 进一步引入其他信息，则持续提升了性能，证明了利用先验知识的益处。另一方面，引入技能抽象带来了额外提升，尤其是在 LIBERO-Long 套件上，得分提高了 5.5%。最后，在上述结构的基础上，集成组合式提示取得了最佳性能。所有这些结果都验证了各组件的重要性。

技能条件化研究。 技能依赖型前馈网络有助于建立分配的技能与动作预测之间的依赖关系。我们研究了不同的条件化策略，包括逐元素相加、通道拼接和 FiLM（Perez 等人，2018）。如表 3 (b) 所示，我们在公式（4）中的建模方式通过分配的技能对前馈网络层进行参数化，在所有设置下始终优于其他方法。具体而言，我们的方法在 LIBERO-Long 套件上取得了最高性能，分别领先 2.9%、2.0% 和 1.3%。在另外两项任务中也观察到了类似结果。这些结果表明，我们的策略对于复杂任务是有效的。

复杂度分析。 我们分析了表4中的训练和部署成本。与其他基于扩散的策略相比，我们的 SDP 模型大小更大，计算成本更高，但在所有任务上的表现始终显著优于它们，而推理时间仅增加了 14.6 毫秒，增幅可以忽略不计。这些对比表明，尽管存在额外开销，我们的设计选择仍兼具有效性与效率。

可视化分析

图5展示了每个时间步分配的技能（左图），以及执行该技能时对应的观测（右侧部分）；边框颜色与技能相对应。可以观察到，SDP 在训练期间学会分配可复用的技能，并在推理过程中按顺序组合这些技能，以实现总体目标。在较低层级，扩散模型以这些技能为条件生成控制信号，从而实现所需的操作，完成复杂任务。尽管技能的分配没有明确的监督，视觉观测仍与分配的技能高度吻合，这体现了我们方法的有效性和可解释性。更多可视化结果见补充材料。

结论

本文提出 SDP，这是一种技能条件扩散策略，将技能学习与条件扩散规划相结合。它从不同任务中抽象出原始技能，并分配合适的技能来指导动作生成。在模拟任务和真实世界任务上的实验展示了其泛化能力和鲁棒性，广泛的研究进一步验证了其有效性和可解释性。

致谢

本项目部分由香港特别行政区研究资助局（RGC）通过一般研究基金（17203023）、协作研究基金（C5052-23G）和 NSFC/RGC 协作研究计划（CRS HKU703/24）提供支持，部分由优必选科技提供支持。本文所述研究工作是在 Zhihao Gu 担任赛马会多媒体与机器学习 STEM 实验室 Dong Xu 教授的博士后期间完成的，该实验室由香港赛马会慈善信托基金资助。

参考文献

Bu, Q.; Yang, Y.; Cai, J.; Gao, S.; Ren, G.; Yao, M.; Luo, P.; 和 Li, H. 2025年。《通过以任务为中心的潜在动作学习在任意环境中执行动作》。*arXiv:2502.14420*。Chi, C.; Xu, Z.; Feng, S.; Cousineau, E.; Du, Y.; Burchfiel, B.; Tedrake, R.; 和 Song, S. 2025年。《扩散策略：通过动作扩散进行视觉运动策略学习》。*TheInternationalJournal ofRoboticsResearch*，44(10-11)：1684–1704。Dalal, M.; Pathak, D.; 和 Salakhutdinov, R. R. 2021年。《通过参数化动作原语加速机器人强化学习》。*Advances in Neural Information Processing Systems*，34：21847–21859。Dhakan, P.; Kasmarik, K.; Vance, P.; Rano, I.; 和 Siddique, N. 2022年。《使用原始技能集成进行并发技能组合》。*IEEE Transactions on Cognitive and Developmental Systems*，15(4)：1879–1890。Doshi, R.; Walke, H.; Mees, O.; Dasari, S.; 和 Levine, S. 2024年。《扩展跨具身学习：用于操作、导航、运动和航空的一种策略》。*arXiv:2408.11812*。Florence, P.; Lynch, C.; Zeng, A.; Ramirez, O. A.; Wahid, A.; Downs, L.; Wong, A.; Lee, J.; Mordatch, I.; 和 Tompson, J. 2022年。《隐式行为克隆》。载于机器人学习会议，158–168。Garg, D.; Vaidyanath, S.; Kim, K.; Song, J.; 和 Ermon, S. 2022年。《Lisa：从语言中学习可解释的技能抽象》。*Advances in Neural Information Processing Systems*，35：21711–21724。Ha, D.; Dai, A.; 和 Le, Q. V. 2016年。《超网络》。*arXiv:1609.09106*。Ha, H.; Florence, P.; 和 Song, S. 2023年。《扩展并提炼：语言引导的机器人技能习得》。载于机器人学习会议，3766–3777。PMLR。Hiranaka, A.; Hwang, M.; Lee, S.; Wang, C.; Fei-Fei, L.; Wu, J.; 和 Zhang, R. 2023年。《基于原始技能的机器人学习：来自人类评价反馈》。载于*2023IEEE/RSJ International Conference onIntelligent Robotsand Systems (IROS)*，7817–7824。IEEE。Ho, J.; J ain, A.; 和 Abbeel, P. 2020年。《去噪扩散概率模型》。*Advances in neural information processing systems*，33：6840–6851。Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.; 等. 2022年。《低秩适配：大语言模型的低秩适配》。*ICLR*，1(2)：3。Jacobs, R. A.; Jordan, M. I.; Nowlan, S. J.; 和 Hinton, G. E. 1991. 自适应局部专家混合。神经计算，3(1)：79–87。Jia, X.; Wang, Q.; Donat, A.; Xing, B.; Li, G.; Zhou, H.; Celik, O.; Blessing, D.; Lioutikov, R.; 和 Neumann, G. 2024. Mail：利用选择性状态空间模型改进模仿学习。发表于第*8*届机器人学习年会。Jo, J.; Lee, S.; 和 Hwang, S. J. 2022. 通过随机微分方程系统实现基于分数的图生成建模。发表于国际机器学习会议，10362–10383。PMLR。Kim, M. J.; Pertsch, K.; Karamcheti, S.; Xiao, T.; Balakrishna, A.; Nair, S.; Rafailov, R.; Foster, E.; Lam, G.; Sanketi, P.; 等。2024. OpenVLA：一种开源视觉-语言-动作模型。*arXiv:2406.09246*。Li, X.; Liu, M.; Zhang, H.; Yu, C.; Xu, J.; Wu, H.; Cheang, C.; Jing, Y.; Zhang, W.; Liu, H.; 等。2023. 视觉-语言基础模型作为有效的机器人模仿器。*arXiv:2311.01378*。Liang, Z.; Mu, Y.; Ma, H.; Tomizuka, M.; Ding, M.; 和 Luo, P. 2024. SkillDiffuser：通过基于扩散的任务执行中的技能抽象，实现可解释的层级规划。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，16467–16476。Liu, B.; Zhu, Y.; Gao, C.; Feng, Y.; Liu, Q.; Zhu, Y.; 和 Stone, P. 2023. LIBERO：终身机器人学习中的知识迁移基准测试。神经信息处理系统进展，36：44776–44791。Liu, S.; Wu, L.; Li, B.; Tan, H.; Chen, H.; Wang, Z.; Xu, K.; Su, H.; 和 Zhu, J. 2024. RDT-1B：一种用于双手操作的扩散基础模型。*arXiv:2410.07864*。Liu, T.; Li, J.; Zheng, Y.; Niu, H.; Lan, Y.; Xu, X.; 和 Zhan, X. 2025. 参数空间中的技能扩展与组合。*arXiv:2502.05932*。Mandlekar, A.; Xu, D.; Wong, J.; Nasiriany, S.; Wang, C.; Kulkarni, R.; Fei-Fei, L.; Savarese, S.; Zhu, Y.; 和 Mart´ín-Mart´ın, R. 2021. 从离线人类示范中学习机器人操作时，哪些因素最重要。*arXiv:2108.03298*。Mees, O.; Hermann, L.; Rosete-Beas, E.; 和 Burgard, W. 2022. CALVIN：一个针对长时域机器人操作任务语言条件策略学习的基准。*IEEE*机器人与自动化快报，7(3)：7327–7334。Mete, A.; Xue, H.; Wilcox, A.; Chen, Y.; 和 Garg, A. 2024. QUEST：用于学习连续控制的自监督技能抽象。神经信息处理系统进展，37：4062–4089。Mishra, U. A.; Xue, S.; Chen, Y.; 和 Xu, D. 2023. 生成式技能链：利用扩散模型进行长时域技能规划。发表于机器人学习会议，2905–2925。PMLR。Mokady, R.; Hertz, A.; Aberman, K.; Pritch, Y.; 和 Cohen-Or, D. 2023. 使用引导扩散模型编辑真实图像的空文本反演。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，6038–6047。

Ni, F.; Hao, J.; Wu, S.; Kou, L.; Liu, J.; Zheng, Y.; Wang, B.; 和 Zhuang, Y. 2024. 行动前生成子目标图像：通过多模态提示，释放扩散模型在机器人操作中的思维链推理能力。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，13991–14000。Peebles, W.; 和 Xie, S. 2023. 基于变换器的可扩展扩散模型。发表于*IEEE/CVF*国际计算机视觉会议论文集，4195–4205。Perez, E.; Strub, F.; De Vries, H.; 和 Dumoulin, V. 2017. 使用通用条件层进行视觉推理，Courville。发表于*AAAI*人工智能会议论文集。Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; 和 Courville, A. 2018. FiLM：使用通用条件层进行视觉推理。发表于*AAAI*人工智能会议论文集，第32卷。Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; 等。2021. 在自然语言监督下学习可迁移视觉模型。发表于国际机器学习会议，8748–8763。PMLR。Ren, A. Z.; Lidard, J.; Ankile, L. L.; Simeonov, A.; Agrawal, P.; Majumdar, A.; Burchfiel, B.; Dai, H.; 和 Simchowitz, M. 2024. 扩散策略优化。*arXiv:2409.00588*。Ren, H.; Sun, L.; Wang, X.; Zhou, P.; Wu, Z.; Dong, S.; Zou, D.; Zheng, Y.; 和 Yang, Y. 2025. HyPoGen：面向泛化策略生成的优化偏置超网络。发表于第十三届国际学习表征会议。Reuss, M.; Pari, J.; Agrawal, P.; 和 Lioutikov, R. 2024a. 用于多任务学习的高效扩散变换器策略与专家去噪器混合。*arXiv:2412.12953*。¨Reuss, M.; Ya˘gmurlu, O. E.; Wenzel, F.; 和 Lioutikov, R. 2024b. 多模态扩散变换器：从多模态目标中学习多样化行为。*arXiv:2407.05996*。Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; 和 Ommer, B. 2022. 利用潜在扩散模型进行高分辨率图像合成。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，10684–10695。Shafiullah, N. M.; Cui, Z.; Altanzaya, A. A.; 和 Pinto, L. 2022. 行为变换器：用一块石头克隆 k 模式。神经信息处理系统进展，35：22955–22968。Singh, I.; Blukis, V.; Mousavian, A.; Goyal, A.; Xu, D.; Tremblay, J.; Fox, D.; Thomason, J.; 和 Garg, A. 2022. ProgPrompt：使用大语言模型生成情境化机器人任务计划。*arXiv:2209.11302*。Song, M.; Deng, X.; Zhou, Z.; Wei, J.; Guan, W.; 和 Nie, L. 2025. 机器人操作扩散策略综述：分类体系、分析与未来方向。*Authorea*预印本。Team, O. M.; Ghosh, D.; Walke, H.; Pertsch, K.; Black, K.; Mees, O.; Dasari, S.; Hejna, J.; Kreiman, T.; Xu, C.; 等，2024年。Octo：一种开源通用机器人策略。*arXiv:2405.12213*。Van Den Oord，A.；Vinyals，O.；等，2017年。神经离散表征学习。神经信息处理系统进展，30。Vincent，P.，2011年。分数匹配与去噪自编码器之间的联系。*Neural Computation*，23(7)：1661–1674。Vuong，Q.；Levine，S.；Walke，H. R.；Pertsch，K.；Singh，A.；Doshi，R.；Xu，C.；Luo，J.；Tan，L.；Shah，D.；等，2023年。Open X-Embodiment：机器人学习数据集与RT-X 模型。载于迈向通用机器人：可扩展技能习得的学习范式@ *CoRL2023*。Wang，L.；Chen，X.；Zhao，J.；He，K.，2024a年。利用异构预训练 Transformer 扩展本体感知-视觉学习。神经信息处理系统进展，37：124420–124450。Wang，Y.；Zhang，Y.；Huo，M.；Tian，R.；Zhang，X.；Xie，Y.；Xu，C.；Ji，P.；Zhan，W.；Ding，M.；等，2024b年。稀疏扩散策略：一种用于机器人学习的稀疏、可复用且灵活的策略。*arXiv:2407.01531*。Wu，H.；Jing，Y.；Cheang，C.；Chen，G.；Xu，J.；Li，X.；Liu，M.；Li，H.；Kong，T.，2023年。释放大规模视频生成式预训练的潜力，以实现视觉机器人操作。*arXiv:2312.13139*。Wu，J.；Sun，X.；Zeng，A.；Song，S.；Lee，J.；Rusinkiewicz，S.；Funkhouser，T.，2020年。用于移动操作的空间动作图。*arXiv:2004.09141*。Xiao，B.；Wu，H.；Xu，W.；Dai，X.；Hu，H.；Lu，Y.；Zeng，M.；Liu，C.；Yuan，L.，2024年。Florence-2：推进适用于多种视觉任务的统一表征。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，4818–4829。Xiong，Z.；Vuorio，R.；Beck，J.；Zimmer，M.；Shao，K.；Whiteson，S.，2024年。蒸馏形态条件超网络，实现高效的通用形态控制。*arXiv:2402.06570*。Ye，S.；Jang，J.；Jeon，B.；Joo，S.；Yang，J.；Peng，B.；Mandlekar，A.；Tan，R.；Chao，Y.-W.；Lin，B. Y.；等，2024年。从视频中进行潜在动作预训练。*arXiv:2410.11758*。Ze，Y.；Zhang，G.；Zhang，K.；Hu，C.；Wang，M.；Xu，H.，2024年。3D扩散策略：通过简单的三维表征实现可泛化的视觉运动策略学习。*arXiv:2403.03954*。Zhang，B.；Sennrich，R.，2019年。均方根层归一化。神经信息处理系统进展，32。Zhang，J.；Zhang，J.；Pertsch，K.；Liu，Z.；Ren，X.；Chang，M.；Sun，S.-H.；Lim，J. J.，2023年。自举习得技能：在大语言模型引导下学习解决新任务。*arXiv:2310.10021*。Zhang，Q.；Tao，M.；Chen，Y.，2022年。广义去噪扩散隐式模型。*arXiv:2206.05564*。Zheng，J.；Li，J.；Liu，D.；Zheng，Y.；Wang，Z.；Ou，Z.；Liu，Y.；Liu，J.；Zhang，Y.-Q.；Zhan，X.，2025年。用于增强具身基础模型的通用动作。载于计算机视觉与模式识别会议论文集，22508–22519。