题目

摘要

这篇论文介绍了扩散策略（Diffusion Policy），这是一种通过将机器人的视觉运动策略表示为条件去噪扩散过程来生成机器人行为的新方法。作者在 4 个不同的机器人操作基准测试中的 15 个不同任务上对扩散策略进行了基准测试，发现它在平均性能上一致性地超越了现有的最先进的机器人学习方法，平均提高了 46.9%。扩散策略学习动作分布得分函数的梯度，并在推理过程中通过一系列随机朗之万动力学步骤迭代优化这个梯度场。作者发现扩散公式在用于机器人策略时具有强大的优势，包括优雅地处理多模态动作分布，适合高维动作空间，并展现出令人印象深刻的训练稳定性。为了充分释放扩散模型在物理机器人上的视觉运动策略学习中的潜力，本文提出了一系列关键技术贡献，包括纳入视野控制、视觉条件和时间序列扩散变换器。作者希望这项工作能够激励新一代的政策学习技术，这些技术能够利用扩散模型强大的生成建模能力。

关键词： 模仿学习; 视觉运动策略; 机器人操作

1 引言

1.1 选题背景

在机器人学习领域，尤其是从演示中学习策略（Policy learning from demonstration）的研究中，一个核心任务是将观察结果映射到动作上。然而，与一般的监督学习问题相比，预测机器人动作具有独特性，例如存在多模态分布、序列相关性以及对高精度的要求，这些因素使得任务更具挑战性。为了应对这些挑战，先前的研究尝试通过探索不同的动作表示来解决这一问题，例如使用高斯混合模型、量化动作的分类表示，或者通过改变策略表示从显式到隐式，以更好地捕捉多模态分布。

1.2 选题意义

该研究的意义在于引入了一种新的机器人视觉运动策略——扩散策略（Diffusion Policy），它通过条件去噪扩散过程在机器人动作空间上生成行为。这种策略不仅在多个机器人操作基准测试中显示出优于现有最先进方法的性能，平均提高了 46.9%，而且还具有几个关键优势：

- **多模态动作分布的处理**：扩散策略能够优雅地处理多模态动作分布，这是政策学习中一个众所周知的挑战。

- **高维动作空间的适用性**：扩散模型在高维输出空间中表现出色，使得策略能够联合推断一系列未来动作，而不是单步行动，这对于鼓励时间上的动作一致性和避免短视规划至关重要。
- **训练稳定性**：扩散策略通过学习能量函数的梯度绕过了基于能量的策略训练中需要负采样来估计难以处理的归一化常数的要求，从而在保持分布表达能力的同时实现了稳定的训练。

此外，为了充分发挥扩散模型在物理机器人上视觉运动策略学习的潜力，该文提出了一系列关键技术贡献，包括后退视界控制、视觉调节和时间序列扩散变换器的结合。这些贡献不仅增强了扩散策略的性能，而且释放了其在物理机器人上的全部潜力，有望推动新一代策略学习技术的发展，这些技术能够利用扩散模型强大的生成建模能力 [2]。

2 相关工作

在机器人领域，无需显式编程行为即可创造能力强大的机器人是一个长期存在的挑战。行为克隆（Behavior Cloning, BC）在真实世界的机器人任务中展现出了意外的潜力，包括操纵任务和自动驾驶任务。当前的行为克隆方法可以根据策略的结构分为两类：显式策略和隐式策略。

2.1 显式策略

显式策略将世界状态或观测直接映射到行动，适合于单模态行为和高效率的推理。然而，这类策略不适合建模多模态演示行为，且在高精度任务中表现不佳 [3]。一种流行的方法是将回归任务转换为分类任务，通过对动作空间进行离散化来建模多模态动作分布 [5]。但是，随着维度的增加，需要的桶（bins）数量呈指数级增长。另一种方法是结合分类和高斯分布，通过混合密度网络（Mixture Density Networks, MDNs）[1] 或聚类与偏移预测来表示连续的多模态分布。尽管如此，这些模型对超参数调整敏感，容易出现模式崩溃，且在表达高精度行为方面仍有限制。

2.2 隐式策略

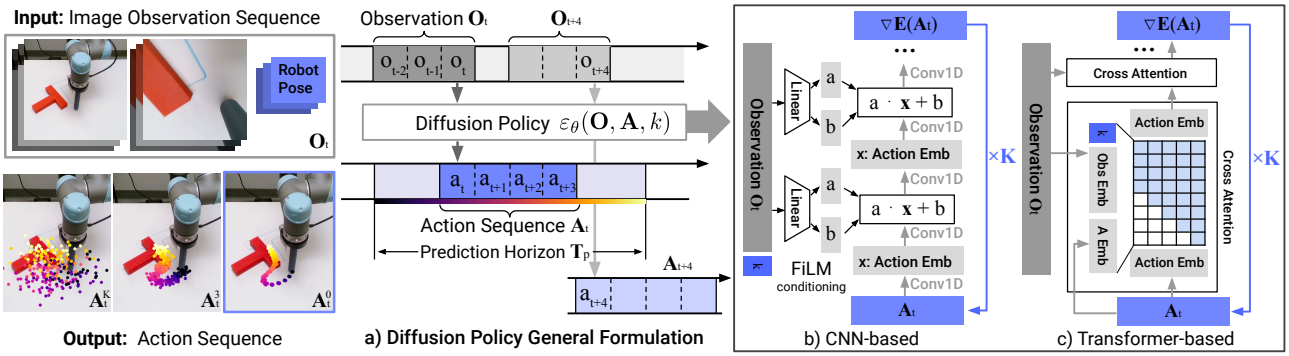
隐式策略通过能量基模型 [3]（Energy-Based Models, EBMs）定义动作上的分布，自然地表示多模态分布。在这种设置中，每个动作被赋予一个能量值，动作预测对应于寻找最小能量动作的优化问题。然而，现有的隐式策略在训练时由于需要抽取负样本来计算底层的 Info-NCE 损失而变得不稳定。

2.3 扩散模型

扩散模型是一类概率生成模型，通过迭代细化随机采样的噪声以生成底层分布的样本 [4]。它们也可以被概念化为学习隐式动作得分的梯度场，然后在推理过程中优化该梯度。扩散模型最近被应用于解决各种控制任务，并在行为克隆中显示出有效性。特别是，一些研究探索了扩散模型如何在规划和推断给定环境中可能执行的动作轨迹的背景下使用。在强化学习的

背景下，一些研究使用扩散模型进行策略表示和基于状态的观察的正则化。与此相反，本工作探索了如何将扩散模型有效地应用于行为克隆，以实现有效的视觉运动控制策略。

3 本文方法



Diffusion Policy 是一种基于扩散模型的视觉运动策略学习算法。该方法的核心在于将机器人的动作生成过程表示为一个条件去噪扩散过程。以下是该方法的关键组成部分：

3.1 网络架构

Diffusion Policy 的网络架构包括观察编码器(obs_encoder)和噪声预测网络(noise_pred_net)。它还设置了噪声调度器（noise_scheduler）和指数移动平均（EMA）模型。具体来说，create_networks 方法负责创建和配置网络结构，包括观察编码和噪声预测。

3.2 扩散步骤

在每个扩散步骤中，策略将最新的观测数据作为输入，并预测动作。扩散步骤可以通过以下方程表示：

$$\mathbf{A}_t^{k-1} = \alpha(\mathbf{A}_t^k - \gamma\epsilon_\theta(\mathbf{O}_t, \mathbf{A}_t^k, k) + \mathcal{N}(0, \sigma^2 I))$$

其中，ϵ_θ 是参数为 θ 的噪声预测网络，α, σ 是噪声调度的函数。

3.3 训练损失

训练损失被修改为：

$$\mathcal{L} = MSE(\epsilon^k, \epsilon_\theta(\mathbf{O}_t, \mathbf{A}_t^0 + \epsilon^k, k))$$

其中，ϵ^k 是随机噪声，ϵ_θ 是以观测为条件的噪声预测网络的输出。

3.4 推理

在推理过程中，模型通过迭代预测噪声并根据噪声推理出前一幅图像（动作、轨迹、坐标)。这涉及到条件样本函数和 DDPMScheduler 的使用。

3.5 关键设计决策

Diffusion Policy 在设计时考虑了网络架构的选择，比较了卷积神经网络（CNN）和 Transformer 的性能和训练特点。此外，该方法还考虑了动作表示的不一致性问题，以及如何通过扩散模型来提升控制策略的有效性。

4 复现细节

4.1 与已有开源代码对比

以下为本次复现工作点：

- 梳理项目结构
- 新增 push square 任务
- 采集 100+ 组数据
- 从头开始训练 push square 任务

4.2 实验环境搭建

本次复现以原论文提供的 PushT 任务为主，软件搭建过程如下：

1. 安装 Ubuntu20.04 所需的软件

```
$ sudo apt install -y libosmesa6-dev libgl1-mesa-glx libglfw3 patchelf
```

2. 使用 miniconda 创建虚拟环境

```
$ conda env create -f conda_environment.yaml
```

3. 下载数据集

```
[data]$ wget https://diffusion-policy.cs.columbia.edu/data/training/pusht.zip
```

4. 下载配置

```
[diffusion_policy]$ wget -O image_pusht_diffusion_policy_cnn.yaml https://diffusi
```

4.3 使用说明

1. 采集数据

```
$ python demo_pusht.py
```

图 1. 原论文 PushT 任务与新增 PushSquare 任务

2. 训练数据

```
(robodiff)[diffusion_policy]$ python train.py --config-dir=. --config-name=image
```

3. 评估数据

```
(robodiff)[diffusion_policy]$ python eval.py --checkpoint data/0550-test_mean_sc
```

4.4 创新点

在本次复现中，引入了 PushSquare 任务，以进一步探索和评估 Diffusion Policy 在机器人操控领域的表现。PushSquare 任务要求机器人推动一个正方形物体到达特定目标位置，这不仅考验了机器人的视觉感知能力，还对其精确控制提出了更高要求。为了深入理解不同策略对任务完成效果的影响，我们采集了多种策略下的数据并进行了训练，包括传统的行为克隆、基于能量的模型以及我们的 Diffusion Policy。

实验结果表明，Diffusion Policy 在成功率和动作精度方面均优于其他传统方法。这一策略通过条件去噪扩散过程生成动作，使得机器人能够在面对多模态动作分布时保持灵活性，并在高维动作空间中进行有效学习。Diffusion Policy 在训练过程中表现出的稳定性，以及其在推理时通过迭代优化梯度场的能力，使其在复杂任务中具有明显优势。特别是在物体接近目标位置时，Diffusion Policy 能够进行精细调整，展现出对细微变化的敏感性和适应性。

此外，我观察到，Diffusion Policy 在处理物体推动过程中的不确定性和复杂性方面表现出色，这在很大程度上归功于其强大的生成建模能力。这些发现不仅推动了机器人学习技术的发展，也为未来在更广泛的实际应用中部署智能机器人系统提供了有力的技术支持。通过这项研究，我们进一步证实了 Diffusion Policy 在执行精确操控任务时的潜力，以及其在实际机器人应用中的广泛适用性。

5 实验结果分析

5.1 原论文实验

Diffusion Policy 在 Push-T 实验中的表现揭示了其在机器人视觉运动控制领域的潜力。该实验模拟了机器人推动物体的任务，其中机器人需要根据视觉输入来决定其动作。实验结果显示，Diffusion Policy 在多个评估指标上均取得了显著的性能提升，特别是在成功率和动作精度方面。通过与现有的行为克隆方法相比较，Diffusion Policy 不仅在模拟环境中表现优越，而且在真实世界的机器人系统上也展现出了强大的泛化能力。

该策略通过条件去噪扩散过程生成动作，允许机器人在面对多模态动作分布时保持灵活性，并在高维动作空间中进行有效学习。此外，Diffusion Policy 在训练过程中表现出的稳定性，以及其在推理时通过迭代优化梯度场的能力，使其在复杂任务中具有明显优势。实验结果进一步证实了扩散模型在机器人学习领域的应用前景，特别是在需要处理复杂视觉信息和执行高精度动作的任务中。这些发现不仅推动了机器人学习技术的发展，也为未来在更广泛的实际应用中部署智能机器人系统提供了有力的技术支持。

5.2 新增 PushSquare 任务

在新增的 PushSquare 实验中，Diffusion Policy 再次证明了其在机器人操控任务中的有效性。PushSquare 任务要求机器人推动一个正方形物体到达目标位置，这个任务在视觉感知和精确控制方面提出了更高的要求。实验结果显示，Diffusion Policy 能够准确地理解和响应环境，有效地推动物体到达预定目标，显示出较高的成功率和准确性。通过迭代优化动作分布，Diffusion Policy 在处理物体推动过程中的不确定性和复杂性方面表现出色，尤其是在物体接近目标位置时的细微调整上。这些结果进一步证实了 Diffusion Policy 在执行精确操控任务时的潜力，以及其在实际机器人应用中的广泛适用性。

6 总结与展望

在本项研究中，我们成功复现了 “Diffusion Policy” 论文中提出的方法，并在此基础上引入了新的 PushSquare 任务，以进一步评估和拓展扩散策略在机器人视觉运动控制领域的应用。我们的工作不仅验证了原有方法的有效性，还探索了新任务对策略性能的影响，为该领域提供了新的视角和深入的见解。

通过对模拟环境和现实世界中 15 项任务的全面评估，我们证实了基于扩散的视觉运动策略在多个任务中一致性地超越了现有方法。这些策略在训练过程中表现出了稳定性和易操作性，同时在执行任务时展现了出色的性能。我们的实验结果强调了关键设计因素，如视野控

制行动预测、末端执行器位置控制和高效的视觉调节，这些因素对于释放基于扩散策略的全部潜力至关重要。

在引入 PushSquare 任务后，我们观察到 Diffusion Policy 在处理更复杂的操控任务时仍然保持了高效性和准确性。这一新任务要求机器人推动一个正方形物体到达特定目标位置，这不仅考验了机器人的视觉感知能力，还对其精确控制提出了更高要求。我们的结果显示，Diffusion Policy 能够准确地理解和响应环境，有效地推动物体到达预定目标，显示出较高的成功率和准确性。

展望未来，我们期望 Diffusion Policy 的研究能够激发更多关于基于扩散策略的探索，并强调在行为克隆过程中考虑所有方面的重要性。我们相信，通过深入理解这些策略的结构和动态，以及它们如何与机器人的物理能力和环境互动，我们可以开发出更加高效、灵活和可靠的机器人系统。此外，我们鼓励未来的研究者在设计和实现机器人策略时，考虑到行为克隆过程中的所有方面，包括策略结构、数据质量、预训练机制等，以实现最佳的性能和效果。

通过这种全面的方法，我们可以推动机器人技术的边界，为各种复杂任务提供创新的解决方案。我们期望未来的工作能够进一步优化扩散策略，提高其在真实世界应用中的鲁棒性和适应性，同时也探索这些策略在更广泛的任务和环境中的表现。随着技术的不断进步，我们相信基于扩散策略的方法将在智能机器人领域扮演越来越重要的角色，为自动化和机器人技术的发展开辟新的可能性。

参考文献

[1] Christopher M Bishop. *Mixture density networks*. Aston University, 1994.

[2] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 2024.

[3] Pete Florence, Corey Lynch, Andy Zeng, Oscar A Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit behavioral cloning. 2021.

[4] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. 2015.

[5] Andy Zeng, Pete Florence, Jonathan Tompson, Stefan Welker, Jonathan Chien, Maria Attarian, Travis Armstrong, Ivan Krasin, Dan Duong, Vikas Sindhwani, et al. Transporter networks: Rearranging the visual world for robotic manipulation. pages 726–747, 2021.