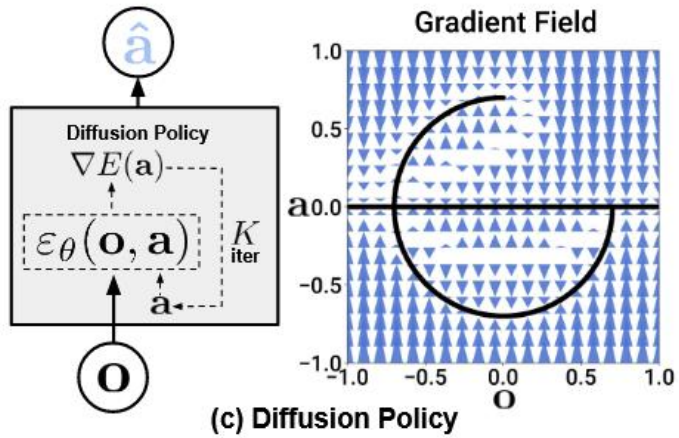
# 论文分享： Diffusion Policy

李佩泽 2025.08.01

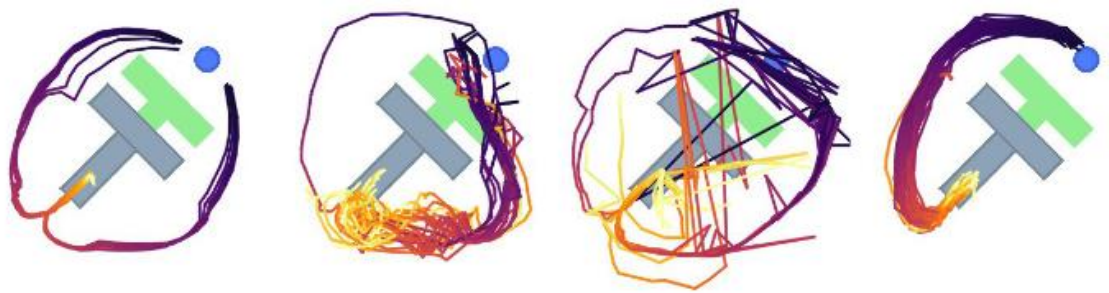
# Diffusion Policy:
## Visuomotor Policy Learning via Action Diffusion

Cheng Chi[1], Siyuan Feng[2], Yilun Du[3], Zhenjia Xu[1], Eric Cousineau[2], Benjamin Burchfiel[2], Shuran Song[1]
[1] Columbia University [2] Toyota Research Institute [3] MIT
https://diffusion-policy.cs.columbia.edu

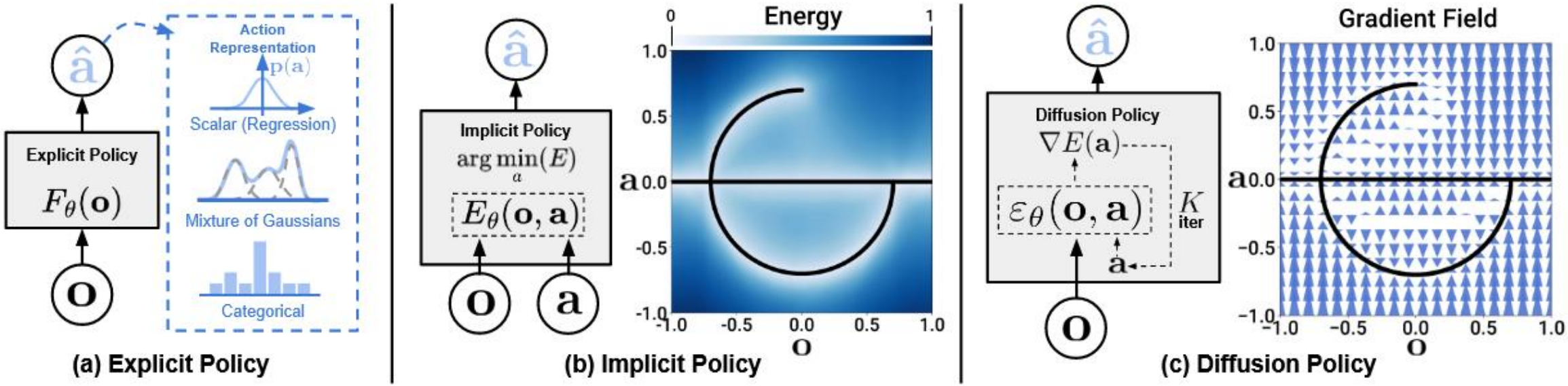
***Abstract*—This paper introduces Diffusion Policy, a new way of generating robot behavior by representing a robot's visuomotor policy as a conditional denoising diffusion process. We benchmark Diffusion Policy across 12 different tasks from 4 different robot manipulation benchmarks and find that it consistently outperforms existing state-of-the-art robot learning methods with an average improvement of 46.9%. Diffusion Policy learns the gradient of the action-distribution score function and iteratively optimizes with respect to this gradient field during inference via a series of stochastic Langevin dynamics steps. We find that the diffusion formulation yields powerful advantages when used for robot policies, including gracefully handling multimodal action distributions, being suitable for high-dimensional action spaces, and exhibiting impressive training stability. To fully unlock the potential of diffusion models for visuomotor policy learning on physical robots, this paper presents a set of key technical contributions including the incorporation of receding horizon control, visual conditioning, and the time-series diffusion transformer. We hope this work will help motivate a new generation of policy learning techniques that are able to leverage the powerful generative modeling capabilities of diffusion models. Code, data, and training details will be publicly available.**



(c) Diffusion Policy



# 基于显式与隐式策略的模仿学习




**显式模仿学习**
以观测为条件
直接生成动作

**隐式模仿学习**
以观测为条件
生成能量分布
在能量分布中采样动作

**扩散模型模仿学习**
以观测为条件
对随机动作去噪
来生成条件动作

# 去噪扩散概率模型(DDPM)

- 去噪过程模型：

$$\mathbf{x}^{k-1} = \alpha(\mathbf{x}^k - \gamma\varepsilon_\theta(\mathbf{x}^k, k) + \mathcal{N}(0, \sigma^2 I)), \qquad (1)$$

- $x^k$ 第k步时得到的带噪声状态，k=K, K-1, ⋯, 0
- $\epsilon_\theta$ 去噪模型基于上一步的状态观测$x^k$ （和当前步数k）
- $\mathrm{N}(0, \sigma^2 I)$ 去噪过程需要引入的随机噪声 防止局部最优
- $\gamma$ 学习率
- $\alpha$ 缩放因子 用于稳定去噪过程 调节引入噪声后的期望均值
- 去噪过程不引入噪声时 可简化为：

$$\mathbf{x}' = \mathbf{x} - \gamma\nabla E(\mathbf{x}), \qquad (2)$$

# 以图像生成为例理解DDPM

**Training**

$\bar{\alpha}_1, \bar{\alpha}_2, \ldots \bar{\alpha}_T$

$x_0$

$\bar{\alpha}_1, \bar{\alpha}_2, \ldots \bar{\alpha}_T$

$\varepsilon$

Sample $t$

$\sqrt{\bar{\alpha}_t}$  $+\sqrt{1-\bar{\alpha}_t}$ $\bar{\alpha}_1, \bar{\alpha}_2, \ldots \bar{\alpha}_T$ $=$ 

$x_0$ $\varepsilon$

t

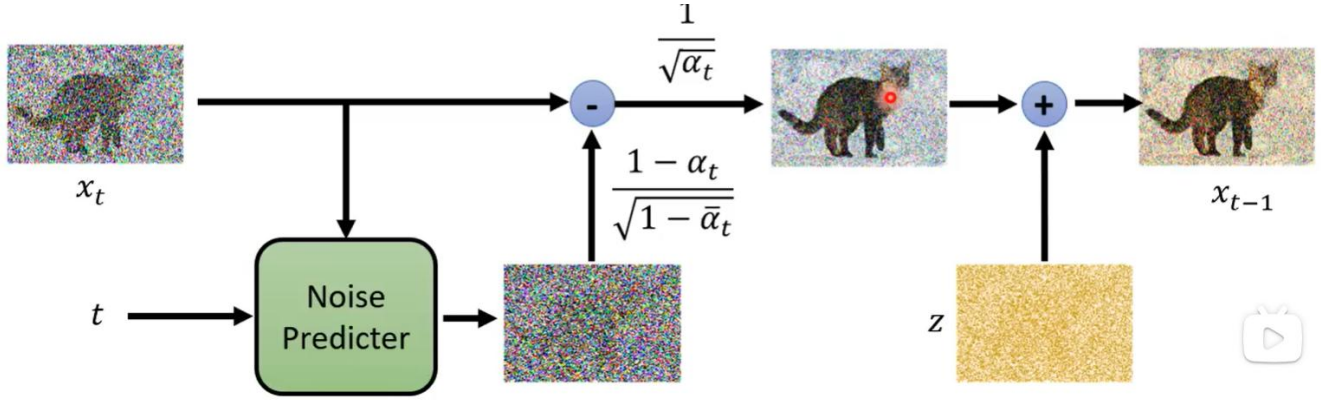
Noise Predicter

**Inference**

$x_T$

**Algorithm 2** Sampling

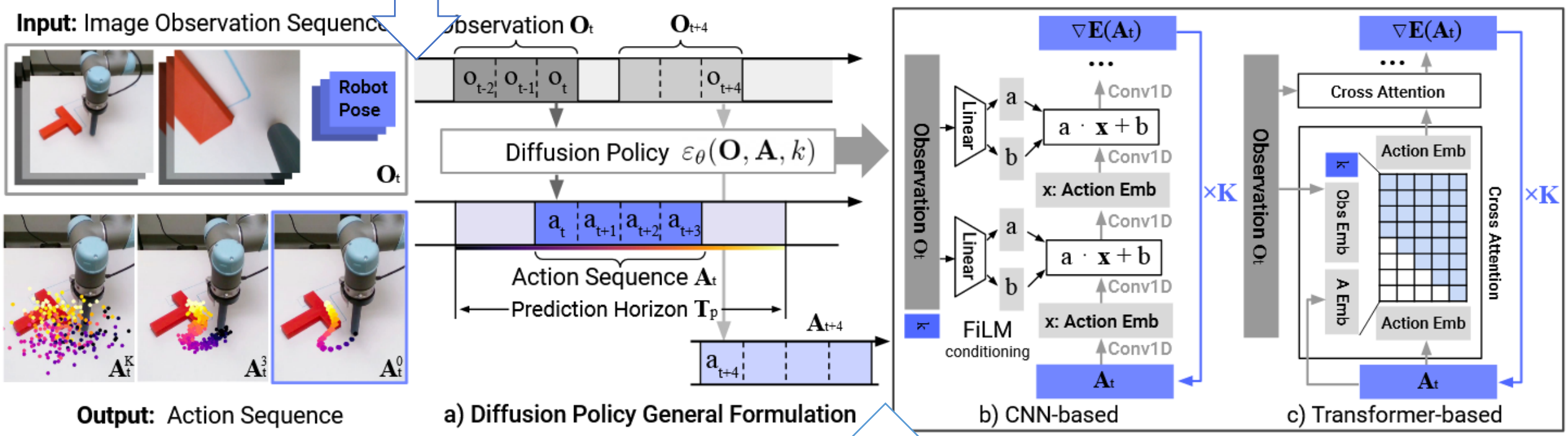
1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$



# 时序处理设计

**Recommendations.** In general, we recommend starting with the CNN-based diffusion policy implementation as the first attempt at a new task. If performance is low due to task complexity or high-rate action changes, then the Time-series Diffusion Transformer formulation can be used to potentially improve performance at the cost of additional tuning.

Vision Encoder：
魔改 ResNet-18

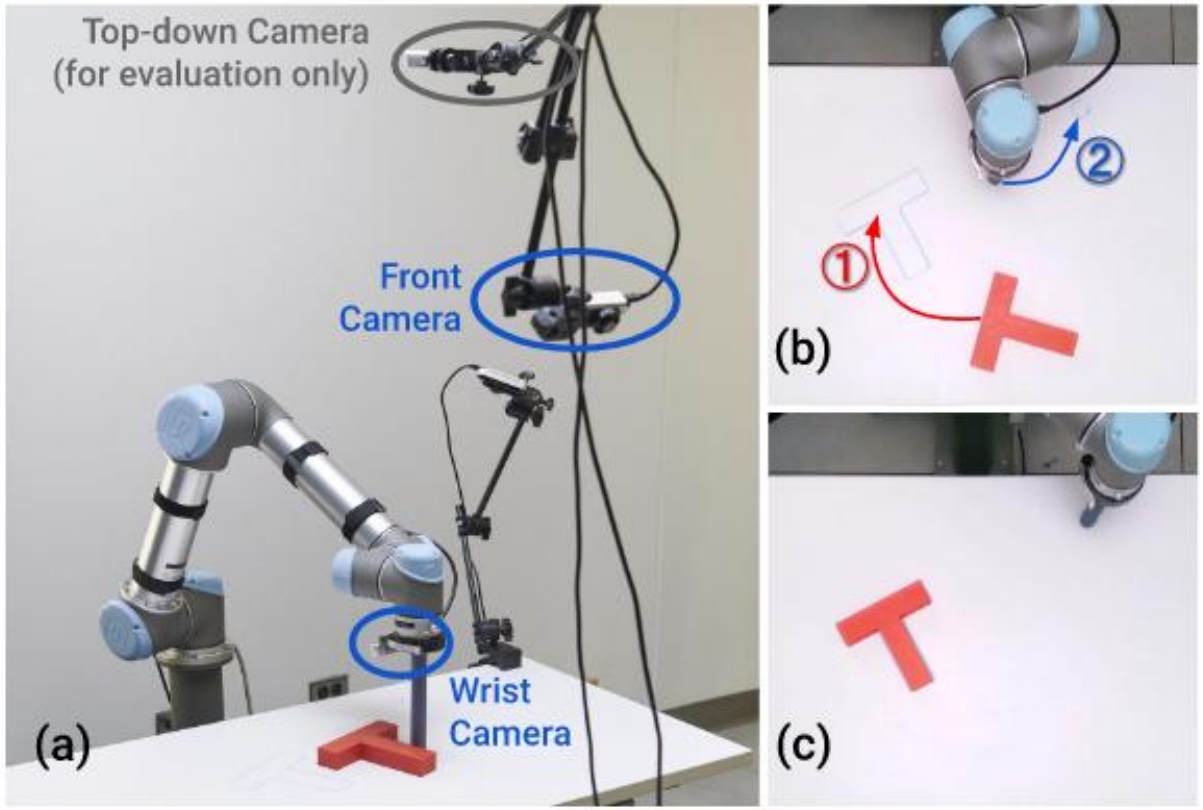


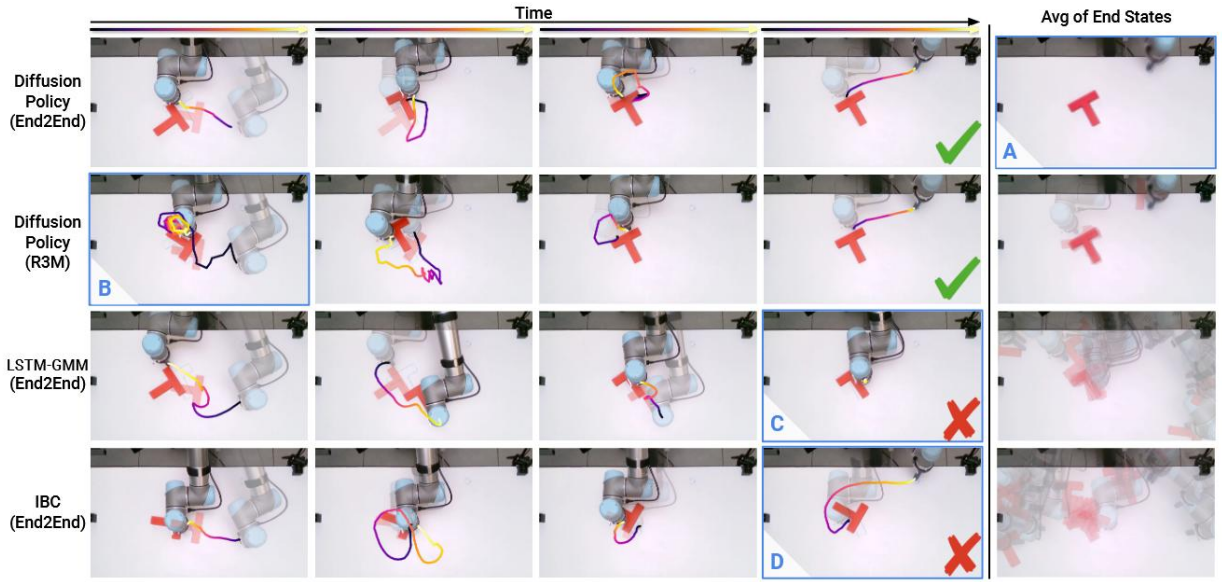

The training loss is modified from Eq 3 to:

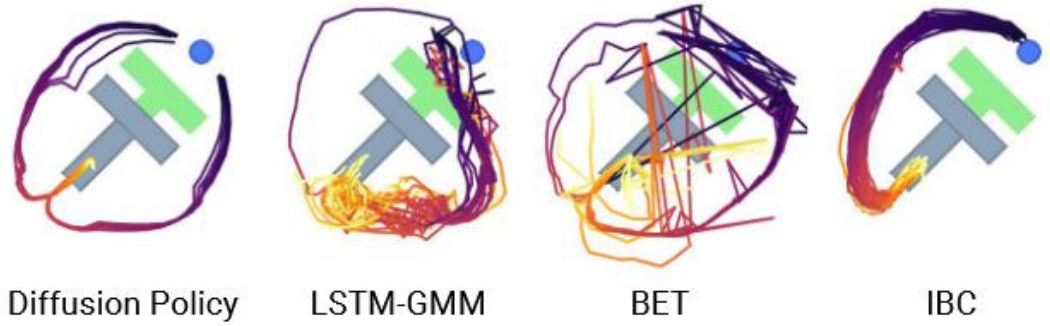
$$\mathscr{L} = MSE(\varepsilon^k, \varepsilon_\theta(\mathbf{O}_t, \mathbf{A}_t^0 + \varepsilon^k, k)) \quad (5)$$

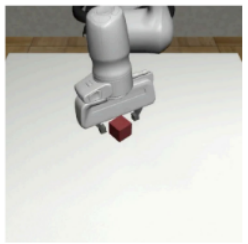
Controller:
receding horizon
control

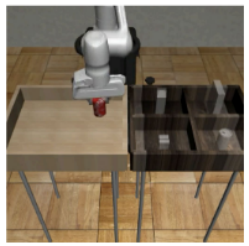
# 实验设计与结果


Top-down Camera
(for evaluation only)
Front
Camera
Wrist
Camera
(a)
(b)
(c)

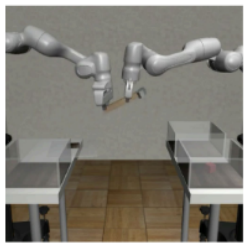

Time
Avg of End States
Diffusion
Policy
(End2End)
Diffusion
Policy
(R3M)
LSTM-GMM
(End2End)
IBC
(End2End)
A
B
C
D

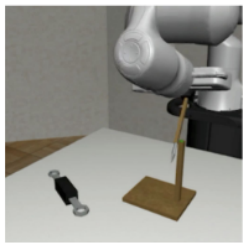

Diffusion Policy
LSTM-GMM
BET
IBC


Lift


Can


Square


Transport


Tool Hang

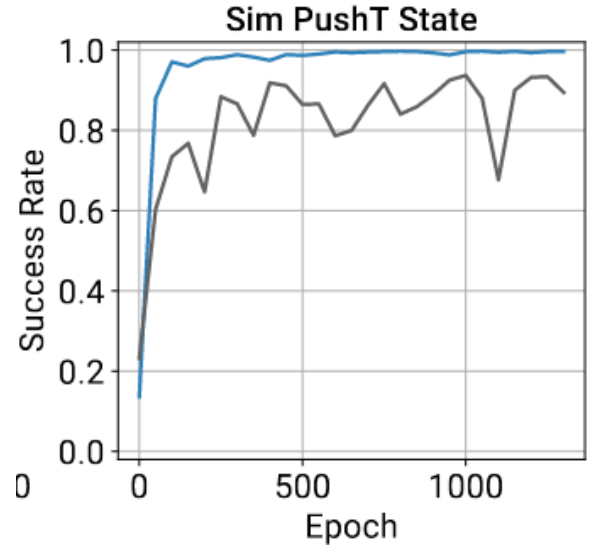

Push-T

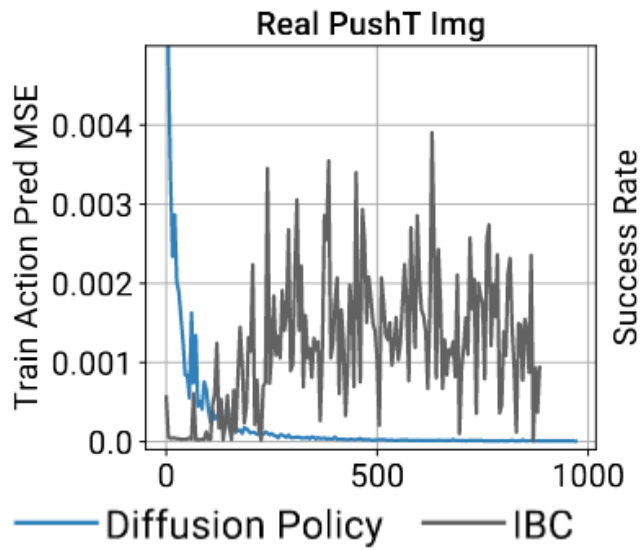
| | Lift | | Can | | Square | | Transport | | ToolHang | Push-T |
|---|---|---|---|---|---|---|---|---|---|---|
| | ph | mh | ph | mh | ph | mh | ph | mh | ph | ph |
| LSTM-GMM [29] | **1.00**/0.96 | **1.00**/0.93 | **1.00**/0.91 | **1.00**/0.81 | 0.95/0.73 | 0.86/0.59 | 0.76/0.47 | 0.62/0.20 | 0.67/0.31 | 0.67/0.61 |
| IBC [12] | 0.79/0.41 | 0.15/0.02 | 0.00/0.00 | 0.01/0.01 | 0.00/0.00 | 0.00/0.00 | 0.00/0.00 | 0.00/0.00 | 0.00/0.00 | 0.90/0.84 |
| BET [42] | **1.00**/0.96 | **1.00**/0.99 | **1.00**/0.89 | **1.00**/0.90 | 0.76/0.52 | 0.68/0.43 | 0.38/0.14 | 0.21/0.06 | 0.58/0.20 | 0.79/0.70 |
| DiffusionPolicy-C | **1.00**/0.98 | **1.00**/0.97 | **1.00**/0.96 | **1.00/0.96** | **1.00/0.93** | **0.97/0.82** | 0.94/0.82 | **0.68/0.46** | 0.50/0.30 | 0.95/**0.91** |
| DiffusionPolicy-T | **1.00/1.00** | **1.00/1.00** | **1.00/1.00** | **1.00**/0.94 | **1.00**/0.89 | 0.95/0.81 | **1.00/0.84** | 0.62/0.35 | **1.00/0.87** | **0.95**/0.79 |

TABLE I: **Behavior Cloning Benchmark (State Policy)** We present success rates with different checkpoint selection methods in the format of (max performance) / (average of last 10 checkpoints), with each averaged across 3 training seeds and 50 different environment initial conditions (150 in total). LSTM-GMM corresponds to BC-RNN in RoboMimic[29], which we reproduced and obtained slightly better results than the original paper. Our results show that Diffusion Policy significantly improves state-of-the-art performance across the board.

| | Lift | | Can | | Square | | Transport | | ToolHang | Push-T |
|---|---|---|---|---|---|---|---|---|---|---|
| | ph | mh | ph | mh | ph | mh | ph | mh | ph | ph |
| LSTM-GMM [29] | **1.00**/0.96 | **1.00**/0.95 | **1.00**/0.88 | 0.98/0.90 | 0.82/0.59 | 0.64/0.38 | 0.88/0.62 | 0.44/0.24 | 0.68/0.49 | 0.69/0.54 |
| IBC [12] | 0.94/0.73 | 0.39/0.05 | 0.08/0.01 | 0.00/0.00 | 0.03/0.00 | 0.00/0.00 | 0.00/0.00 | 0.00/0.00 | 0.00/0.00 | 0.75/0.64 |
| DiffusionPolicy-C | **1.00/1.00** | **1.00/1.00** | **1.00**/0.97 | **1.00**/0.96 | 0.98/**0.92** | **0.98/0.84** | **1.00/0.93** | **0.89/0.69** | **0.95/0.73** | **0.91/0.84** |
| DiffusionPolicy-T | **1.00/1.00** | **1.00**/0.99 | **1.00/0.98** | **1.00/0.98** | **1.00**/0.90 | 0.94/0.80 | 0.98/0.81 | 0.73/0.50 | 0.76/0.47 | 0.78/0.66 |

TABLE II: **Behavior Cloning Benchmark (Visual Policy)** Performance are reported in the same format as in Tab I. LSTM-GMM numbers were reproduced to get a complete evaluation in addition to the best checkpoint performance reported. Diffusion Policy shows consistent performance improvement, especially for complex tasks like Transport and ToolHang.



# Q&A Section