Diffusion Policy:

Visuomotor Policy Learning via Action Diffusion

Group: [3]

Presenter: [Junlin Wang]

Date: [2024.4.12]

AncoraSIR.com

Main Problem and Motivation

Description of the problem

Object manipulation through imitation learning is a newly developing field in robotics. Concretely, robots learn to acquire diverse manipulation skills, such as table wiping or cooking shrimp, by taking advantage of human demonstrations.

Main Problem and Motivation

Challenges

- High-precision and robust closed-loop control is demanded.
- The high-dimensional action space makes it difficult for models to infer time-consistent action responses.
- Real-time control is indispensable, calling for computionally efficient models.

Main Problem and Motivation

Motivation

Diffusion models (DM) have shown great effectiveness in handling high-dimensional data while capturing multi-modal distributions. Applying DM in object manipulation may boost models' capabilities of inferring multi-modal actions in the high-dimensional action space.

Related Work

1. Explicit Policy

- Directly maps from world state or observation to action [1] [2].

2. Implicit Policy

- Define distributions over actions by using Energy-Based Models [3] [4].

[1] Rahmatizadeh, Rouhollah et al. “Vision-Based Multi-Task Manipulation for Inexpensive Robots Using End-to-End Learning from Demonstration.” 2018 IEEE International Conference on Robotics and Automation (ICRA) (2017): 3758-3765.

[2] Zhang, Tianhao et al. “Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation.” IEEE International Conference on Robotics and Automation (2017).

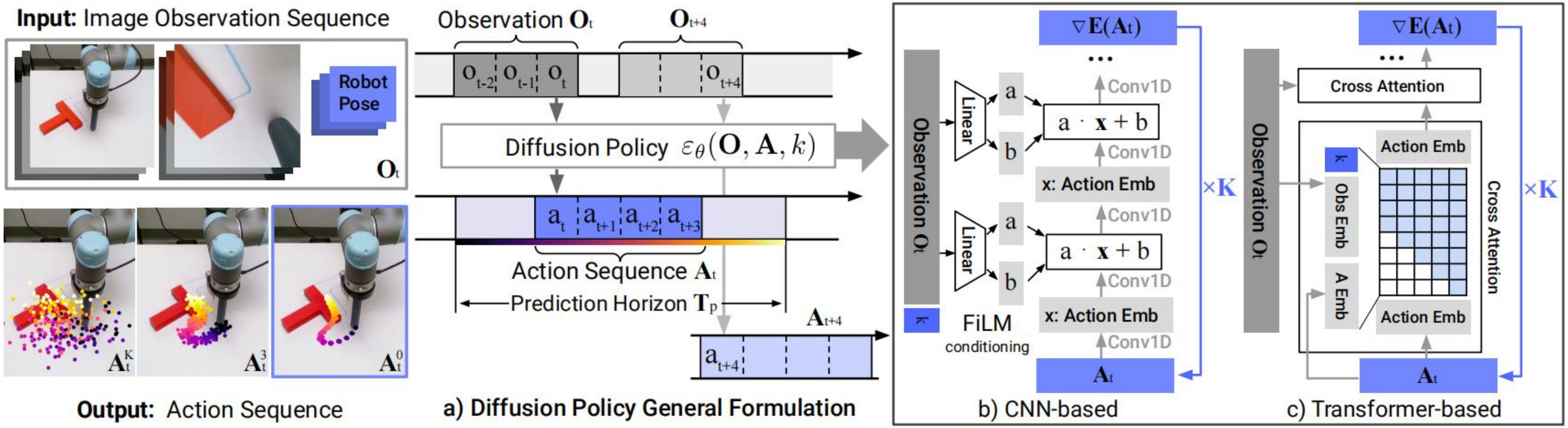
[3] Florence, Peter R. et al. “Implicit Behavioral Cloning.” ArXiv abs/2109.00137 (2021): n. pag.

[4] Jarrett, Daniel et al. “Strictly Batch Imitation Learning by Energy-based Distribution Matching.” ArXiv abs/2006.14154 (2020): n. pag.

Limitations of Prior Work

- Not suitable for modeling multi-modal demonstrated behavior.
- Struggles with high-precision tasks.
- Unstable to train.

Methodology



Preliminary

1. Imitation Learning

Algorithm 1 Imitation Learning

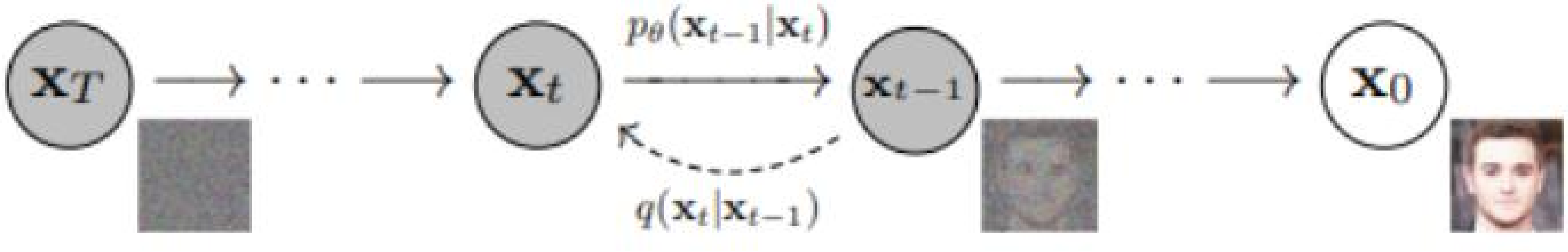
Given: Observation $\mathcal{O}$, human demonstration $\mathcal{H}$.

1: Randomly initialize policy $\mathcal{P}$.
2: **while** not converge **do**
3:     Sample $o_t \in \mathcal{O}$, $h_t \in \mathcal{H}$
4:     $a_t = \mathcal{P}(o_t)$
5:     minimize $MSE(a_t, h_t)$
6: **return** $\mathcal{P}$

AncoraSIR.com

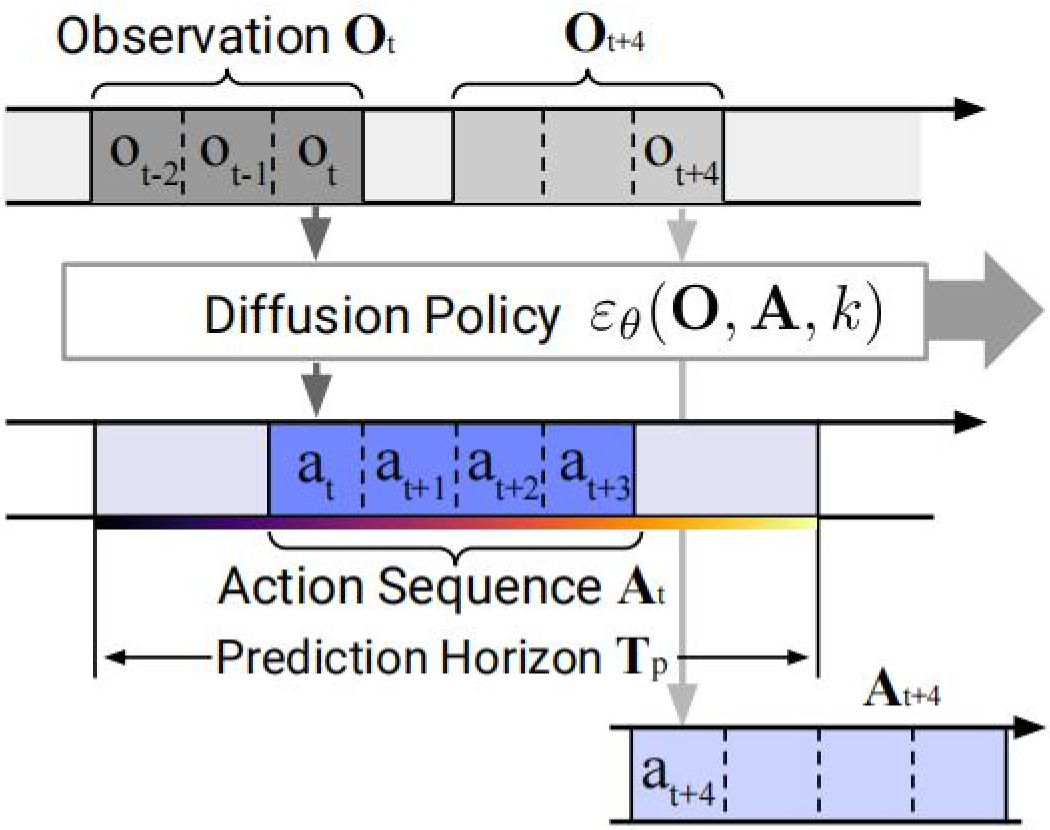
Preliminary

2. Denoising Diffusion Probabilistic Models (DDPM)



- **Forward Process:** Adding noise to the original image.
- **Reverse Process:** Recover the original image by denoising.

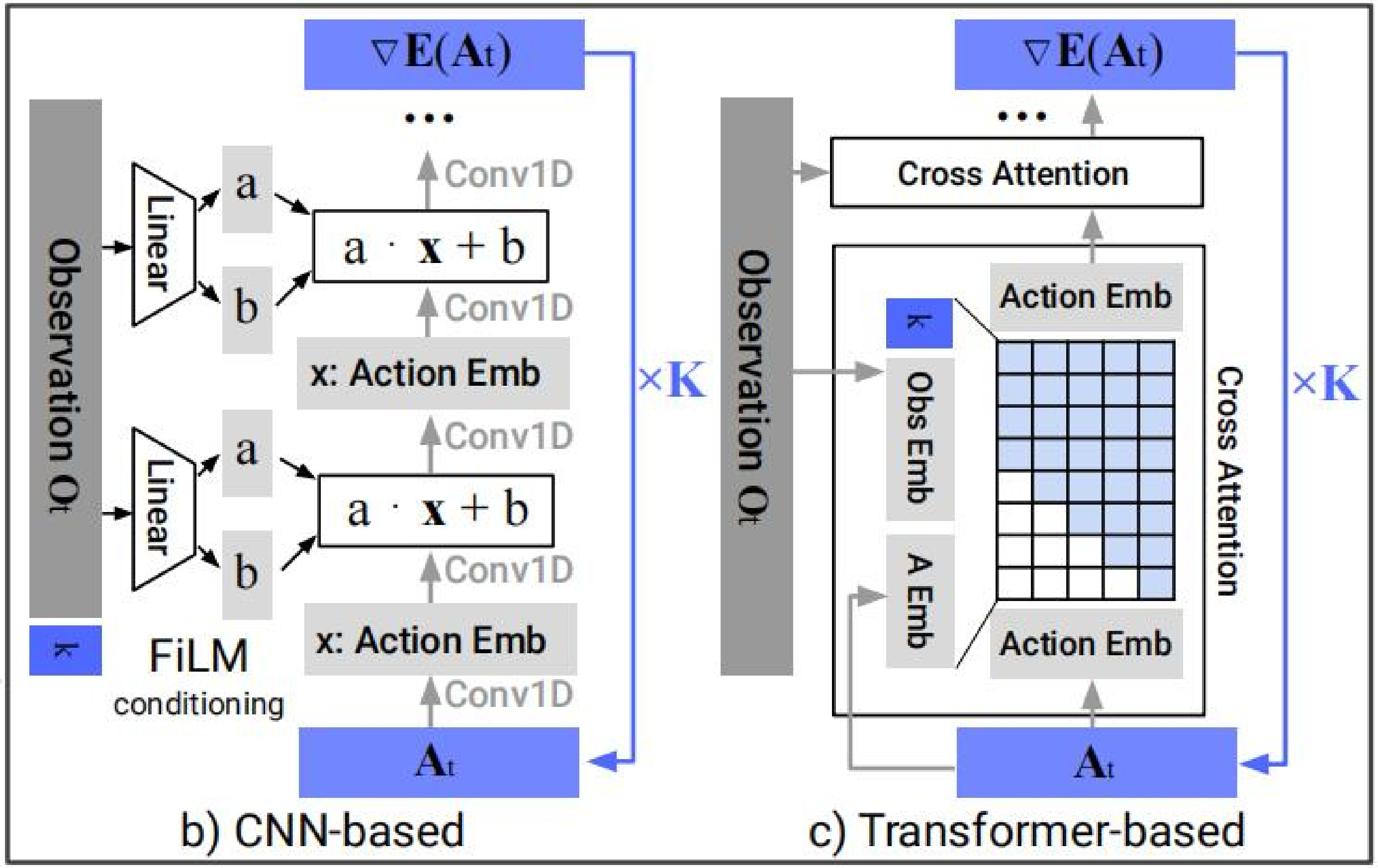
Action Chunking



At time step t, the policy takes as input the lastest T_0 steps of observation and predict T_p steps of actions, of which T_a steps of actions are executed on the robot without re-planning.

AncoraSIR.com

Diffusion Policy



Visual encoder: ResNet-18 [1]

Conditioning: FiLM [2] (CNN), cross attention (Transformer)

Backbone: UNet [3] (CNN), MinGPT [4] (Transformer)

[1] He, Kaiming et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 770-778.

[2] Perez, Ethan et al. "FiLM: Visual Reasoning with a General Conditioning Layer." AAAI Conference on Artificial Intelligence (2017).

[3] Janner, Michael et al. "Planning with Diffusion for Flexible Behavior Synthesis." International Conference on Machine Learning (2022).

[4] Shafiullah, Nur Muhammad (Mahi) et al. "Behavior Transformers: Cloning k modes with one stone." ArXiv abs/2206.11251 (2022): n. pag.

Experimental Setup - Simulation

1. Datasets

- Robomimic, Push-T, Multimodal Block Pushing, Franka Kitchen.

2. Evaluation Metrics

- The metric for most tasks is success rate, except for the Push-T task, which is target area coverage.

3. Training

- State-based tasks are trained for 4500 epochs, and image-based tasks for 3000 epochs.

Experimental Setup - Real World

1. Tasks

- Push-T, Mug Flipping, Sauce Pouring and Spreading.

2. Evaluation Metrics

- IoU, success rate, coverage rate, duration.

Limitations

- Reach suboptimal performance with inadequate demonstration data.
- High computational costs and inference latency.

Future Work

- Exploit diffusion model acceleration methods such as new noise schedules, inference solvers, and consistency models.

Extended Readings

[1] Ho, Jonathan et al. “Denoising Diffusion Probabilistic Models.” ArXiv abs/2006.11239 (2020): n. pag.

[2] Song, Jiaming et al. “Denoising Diffusion Implicit Models.” ArXiv abs/2010.02502 (2020): n. pag.

[3] Nichol, Alex and Prafulla Dhariwal. “Improved Denoising Diffusion Probabilistic Models.” ArXiv abs/2102.09672 (2021): n. pag.

[4] Rombach, Robin et al. “High-Resolution Image Synthesis with Latent Diffusion Models.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 10674-10685.

[5] Jang, Eric et al. “BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning.” ArXiv abs/2202.02005 (2022): n. pag.

[6] Ahn, Michael et al. “Do As I Can, Not As I Say: Grounding Language in Robotic Affordances.” Conference on Robot Learning (2022).

[7] Brohan, Anthony et al. “RT-1: Robotics Transformer for Real-World Control at Scale.” ArXiv abs/2212.06817 (2022): n. pag.

[8] Zhao, Tony et al. “Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware.” ArXiv abs/2304.13705 (2023): n. pag.

Summary

- This work proposed a novel approach for manipulation, dubbed **diffusion policy**, which achieved state-of-the-art performance on 4 benchmarks with an average improvement of 46.9%.
- Action trajectory generation was formulated as a reverse Gaussian denoising process conditioned on the latest observation and current iteration through **FiLM** modulation or cross attention.
- Experiments demonstrated that diffusion policy possessed strong abilities of modeling highly expressive multimodal distribution while maintaining temporal consistency and training stability.

Q & A