Sample-efficient Imitation Learning for Deformable Object Manipulation Using Diffusion Models

Author: Andrea Ritossa — ritossa@kth.se
Supervisor: Alberta Longhini, Marco Moletta, Danica Kragic
Location: Robotics, Perception and Learning Lab, KTH

1 INTRODUCTION

The main motivation behind this project is to improve the sample efficiency of imitation learning frameworks. Collecting demonstrations of robot manipulation tasks is tedious, time-consuming, and constrained by physical limitations, particularly for deformable objects like cloth. Therefore, this research is looking for approaches that could allow current models to operate more efficiently, requiring fewer demonstrations to potentially achieve higher performance.

Privileged Information: There has been various approaches to improve policy performance through the use of privileged information. Rapid Motor Adaptation Kumar et al. (2021) uses privileged data to pre-train a base policy and afterwards trains the module used at test time through regression on the feature space outputted by the privileged information encoder. Similarly, TraKDis Chen and Rojas (2024) proposes a framework for exploiting privileged information through Knowledge Distillation by pretraining the actor policy from receiving in input the privileged information. This, coupled with a CNN encoder for privileged information estimation results in an action policy that consumes only images at test time. Learning Deformable Object Manipulation from Expert Demonstrations (DMFD) Salhotra et al. (2022) proposed exploiting privileged information in a Reinforcement Learning paradigm, where the actor is image-based and the critic is state-based.

While these approaches demonstrate baseline techniques, they lack addressing the sample efficiency question that will be provided in this project.

Furthermore we develop a different approach in exploiting privileged information, by developing a shared feature space between privileged information and images for policy learning. This framework is flexible and applicable across a broad range of architectures, without constraints on input types, enabling information encoding beyond just images and text. Our work aims to develop an approach with higher performance than existing baselines and conduct a thorough sample efficiency study across all methods.

2 METHODOLOGY

Starting from our objective of improving sample efficiency in imitation learning for deformable object manipulation, we designed experiments to evaluate the effectiveness of incorporating privileged information during training with the purpose of finding the winning recipe for exploting privileged information.

- **Hypothesis**: Incorporating privileged information during training improves sample efficiency and performance compared to policies trained solely with image inputs.
- **Setup**: We conducted experiments on simulated environments (PushT and SoftGym Lin et al. (2021)) comparing three training regimes:
  1. Image-based policy
  2. State-based policy
  3. Image-based policy trained with privileged information (state features)
- **Implementation of Privileged Information**: We concatenate features from the image encoder and state encoder, applying dropout (ex: rate 0.37) to the state information during training to prepare the model for inference conditions when states are unavailable. These concatenated features are further processed by a shared encoder (e.g., a transformer) before being used by the policy.

3 EXPERIMENTS AND RESULTS

In the initial experiments, we focused on shaping our approach in simulation environments, evaluating policy performance and sample efficiency.

PushT with UNet-Based Diffusion Policy: We first experimented on the PushT task Chi et al. (2023), which involves pushing a T-shaped block into a target configuration. We employed a UNet-based diffusion policy and trained three variants based on input types and feature encoding:

1. An image-based policy, using a ResNet encoder for image features.

2. A state-based policy, using an MLP encoder for state features.

3. A privileged policy: This policy is image-based at inference. During training, it utilized features from both a ResNet image encoder and an MLP state encoder. These features were then processed by a shared transformer encoder before being fed to the UNet-based diffusion policy. State information was subject to dropout (rate 0.37) during training and set to zero at inference.

This experiment provided a base, offering a first positive result: an improved performance of the image-only policy obtained through the usage of privileged information during training (Figure 1). However, further experiments are necessary to evaluate the optimal setup for how to exploit at the best this privileged information and to understand how much we can bridge the performance gap between the image-only policy and a policy with access to privileged information at test time.

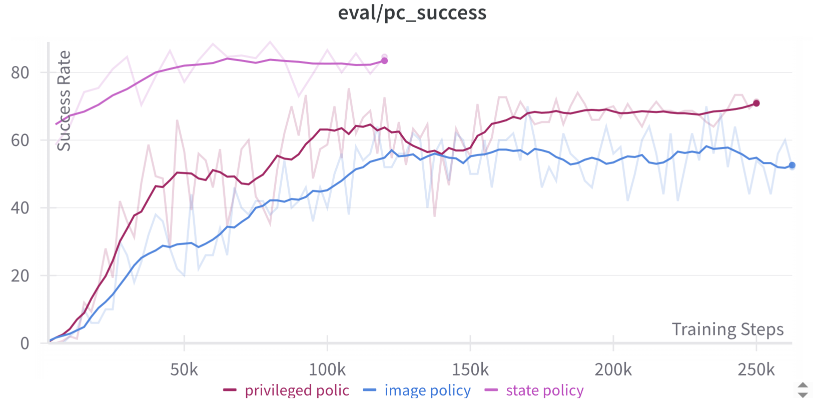

Figure 1: Performance on the PushT task. The plot shows success rate (y-axis) versus training steps (x-axis) for three policies: a privileged policy using state information directly (pink line, top), an image-based policy trained with access to privileged state information (maroon line, middle), and a baseline state-only policy (blue line, bottom).

Baseline Replication - DMFD: We then moved to the SoftGym environment Lin et al. (2021), focusing on the ”ClothFold” task, which involves folding a piece of cloth in half with variable initial size, position, and orientation. We reproduced the results from the DMFD paper Salhotra et al. (2022) and extended them with a sample efficiency study of their framework. This aimed at providing a solid benchmark for a different approach to utilizing privileged information. We conducted a sample efficiency study by training with datasets of varying sizes. The results, reported in Figure 2, use the performance metrics defined in Equations equation 1 and equation 2. To evaluata the sample efficiency differenta training dataset was used, maintaining constant to 8:1 the ratio of demonstrations to variations: including random cloth sizes and initial rotations ($\pm\pi/4$ radians) of the cloth initialization.

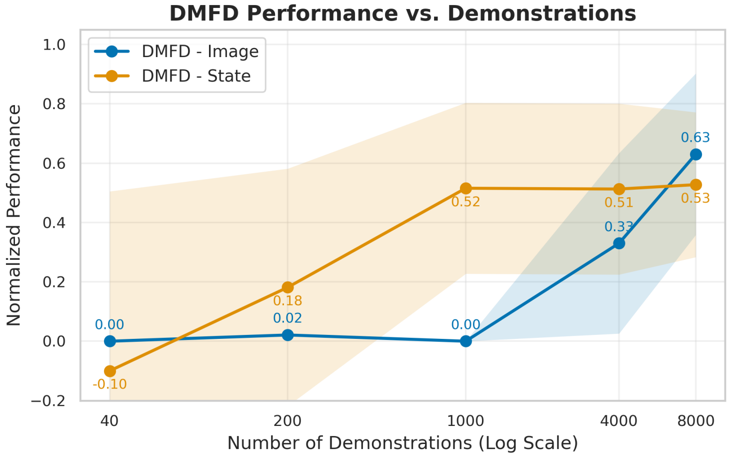

Figure 2: DMFD Performance vs. Number of Demonstrations on the ClothFold task. The plot shows Normalized Performance (y-axis, higher is better, as defined in Equation equation 2) against the number of training demonstrations (logarithmic x-axis). The DMFD - State policy (orange) generally outperforms the DMFD - Image policy (blue), particularly with fewer demonstrations. Both policies show improved performance with an increasing number of demonstrations. Shaded regions indicate standard deviation. Each data point represents statistics from 100 evaluation episodes (20 episodes across 5 distinct random seeds).

Behavioral Cloning with Transformer-Based Diffusion Policy on SoftGym's ClothFold Task: At the moment another experiment is underway: to develop policies for the ”ClothFold” task in SoftGym Lin et al. (2021) with a transformer-based diffusion policy architecture. Drawing inspiration from Diffusion Policy Chi et al. (2023) we are employing a behavioral cloning approach. We want to evaluate the same three policy variants as in the PushT experiments (image-based, state-based, and privileged image-based trained with state features). We also are executing the sample efficiency study for this approach, as a key part of the research. Once again, the performance metrics are consistent with those defined in Equations equation 1 and equation 2, with results presented in Figure 3.

4 METRICS

ClothFold Task Performance Metrics

For the ClothFold task, we evaluate performance using two metrics. First, the raw performance P is calculated based on the particle positions:

$$P = -\left(\frac{1}{M}\sum_{i\in G_a}\|\mathbf{p}_i - \mathbf{p}_{j(i)}\|_2\right) - 1.2 \times \left(\frac{1}{M}\sum_{j\in G_b}\|\mathbf{p}_j - \mathbf{p}_{j,init}\|_2\right) \qquad (1)$$

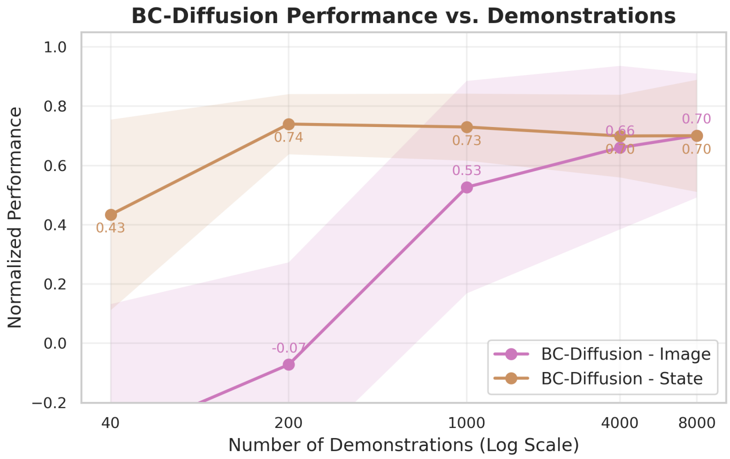

Figure 3: BC-Diffusion Performance vs. Number of Demonstrations on the ClothFold task. The plot shows Normalized Performance (y-axis, higher is better, as defined in Equation equation 2) against the number of training demonstrations (logarithmic x-axis). Shaded regions indicate standard deviation. Each data point represents statistics from 100 evaluation episodes (20 episodes across 5 distinct random seeds).

where M is the number of particles per fold group, G_a, G_b are the particle sets for the two halves being folded together, $\mathbf{p}_i$ is the current 3D position of particle i, $\mathbf{p}_{j(i)}$ is the current 3D position of the corresponding particle in the other group, and $\mathbf{p}_{j,init}$ is the initial 3D position of particle j in the fixed group. This raw performance is then normalized to a scale representing the progress from the initial state towards an ideal fold. The ideal performance $P_{ideal} = 0$ occurs when corresponding particles align ($\|\mathbf{p}_i - \mathbf{p}_{j(i)}\|_2 \to 0$) and fixed particles remain stationary ($\|\mathbf{p}_j - \mathbf{p}_{j,init}\|_2 \to 0$), causing both terms in Equation equation 1 to vanish. The normalized performance is:

$$P_{\text{norm}} = \frac{P - P_{init}}{P_{ideal} - P_{init}} = \frac{P - P_{init}}{-P_{init}} \tag{2}$$

where P_{init} is the raw performance calculated using Equation equation 1 at the start of the episode.

5 TARGETED VENUE

The targeted venue for publication is the **International Conference on Intelligent Robots and Systems (IROS)**, given its focus on cutting-edge research in robotics, learning, and manipulation. IROS provides an ideal platform for presenting innovations in sample-efficient robot learning and deformable object manipulation.

6 REFERENCES

Wei Chen and Nicolas Rojas. Trakdis: A transformer-based knowledge distillation approach for visual reinforcement learning with application to cloth manipulation, 2024.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023. Version 4, revised 1 Jun 2023.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for legged robots, 2021.

Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. *arXiv preprint arXiv:2011.07215*, 2021.

Gautam Salhotra, I-Chun Arthur Liu, Marcus Dominguez-Kuhne, and Gaurav S. Sukhatme. Learning deformable object manipulation from expert demonstrations, 2022.