

Cloth Diffusion: Halftime Report

May 10, 2025

Student: Andrea Ritossa

Project: Sample Efficient Imitation Learning for Deformable Object Manipulation Using Diffusion Models

Project description

The research project aims at developing a sample-efficient imitation learning framework for deformable object manipulation (DOM), particularly cloth manipulation tasks. The approach proposed by the research team is to leverage privileged information (state representations of the object to manipulate) to improve policy learning in diffusion models. The goal is to enhance sample efficiency of the policies.

1 First Set(s) of Experiments

The first experiments were executed in simulation. Using environment PushT (first easy task to familiarize with diffusion policies training) and SoftGym (classical simulation environment used by baselines). The research question that was tried to achieve was "is there some juice, in the approach of improving sample efficiency this way?", to get a preliminary result to the question.

Therefore I executed experiments:

- on PushT, with a Unet-Based diffusion policy: training image based policy — state based policy — privileged policy: image based but trained including privileged information through concatenation and random dropout of the states
- on SoftGym, baseline results of paper "DMFD", expanding with a sample efficiency study of the framework there proposed
- on SoftGym, with a Transformer-Based diffusion policy: training image based policy — state based policy — privileged policy: image based but trained including privileged information through concatenation and random dropout of the states

2 Results

Experiment on PushT demonstrated a performance improvement of the image based policy by using privileged information at training time.

3 Lessons Learned

We learned that privileged information is a valuable asset for accelerating learning and improving policy robustness in cloth manipulation tasks. However, the architecture for effectively fusing multimodal data remains a challenge. Simpler tasks and policies helped identify bottlenecks and informed the design of more complex experiments. The importance of careful benchmarking and ablation studies became evident to isolate the contributions of each component.

4 Plan for Final Experiments

The final experiments will focus on scaling up the complexity of tasks and models. We plan to evaluate advanced architectures that better integrate privileged and non-privileged modalities, such as transformers and contrastive learning frameworks. Experiments will be conducted both in simulation and, where possible, on real robotic platforms. Additional ablations will be performed to assess the impact of each modality and architectural choice.

5 Metrics to be Measured

We will measure sample efficiency (number of demonstrations required for convergence), task success rate, generalization to unseen tasks or objects, and computational efficiency during training and inference. Additional metrics include robustness to domain shifts and the quality of learned representations (e.g., via contrastive loss).

6 Evaluation Strategy

Results will be evaluated against established baselines from the literature, including models trained without privileged information and state-of-the-art diffusion policies. Benchmarks such as SoftGym, Garment Lab, and DaXBench will be used. Statistical significance will be assessed through repeated trials, and qualitative analysis will be performed via visualization of policy behavior.

7 Expected Outcome

We expect that models leveraging privileged information during training will demonstrate superior sample efficiency and generalization compared to base-

lines. The final system should provide insights into effective multimodal fusion strategies for DOM tasks and set a new benchmark for cloth manipulation in simulation environments.

8 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.

References

- [1] MIT, “Flow Matching and Diffusion Models,” 2025. [Online]. Available: <https://diffusion.csail.mit.edu/>