Curating Non-Expert Demonstrations for Safer Imitation Learning

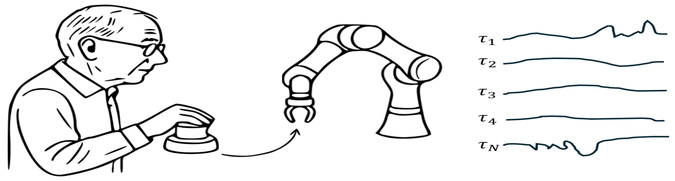
Noushad Sojib (noushad.sojib@unh.edu) Ola Ghattas Momotaz Begum

Computer Science, University of New Hampshire

Cognitive Assistive Robotics Lab

Introduction

We consider learning from non-expert human demonstrations.



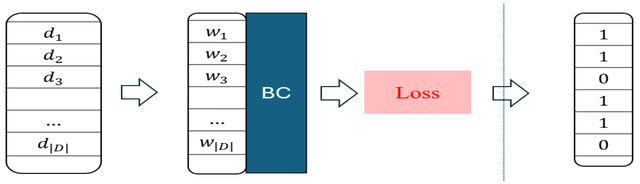
- Demonstrations from human user is costly and often erroneous.
- Lay user's errors are different than noise added data.
- Removing entire failed trajectories discards potentially useful segments.

Preliminaries on Imitation Learning

We assume a dataset $\mathcal{D} = \{\tau_1, \ldots, \tau_N\}$ of N trajectories that include both optimal and suboptimal demonstrations, each $\tau_i = \{(s_1, a_1), \ldots, (s_{T_i}, a_{T_i})\}$ containing state-action pairs with $s_t \in \mathcal{S}$ and $a_t \in \mathcal{A}$. Behavior Cloning (BC) learns a policy $\pi_\theta : \mathcal{S} \to \mathcal{A}$ by minimizing $\mathbb{E}_{(s,a)\sim\mathcal{D}}[-\log \pi_\theta(a|s)]$.

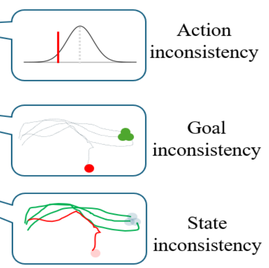
States: Two RGB images ($\mathbb{R}^{84\times84\times3}$), end-effector position ($\mathbb{R}^3$), rotation ($\mathbb{R}^4$), and gripper status ($\mathbb{R}^1$). **Actions:** $\mathbb{R}^7$ vector with six delta end-effector motions and a binary gripper command.

Stage 1: Behavior Cloning For Error Discovery (BED)



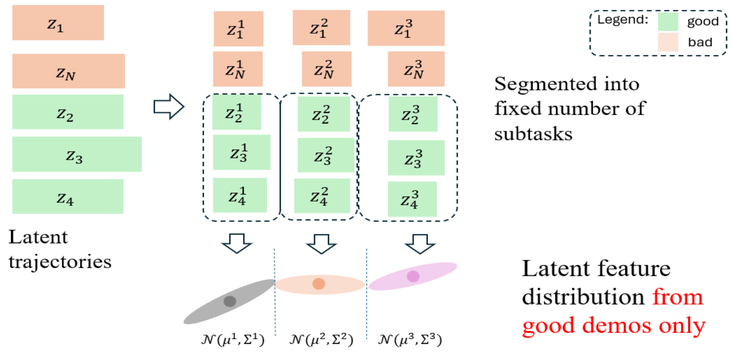
Penalizing inconsistent demos

$$L(D,w,m) = c \cdot \sum_{i\in N} w_i \cdot \frac{1}{T_i} \sum_{(s,a)\sim\tau_i} (\hat{\pi}_\theta(s) - a)^2$$
$$+h \cdot \sum_{i\in N} w_i \cdot ||G - g_i||$$
$$+q \cdot \sum_{i\in N} w_i \cdot ||Z - z_i||$$
$$+k \cdot \left(m \cdot N - \sum_{j=1}^{N} w_j\right)^2$$

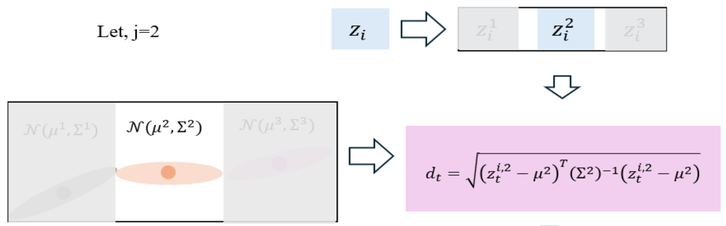


Stage 2: Identifying Suboptimal Segments

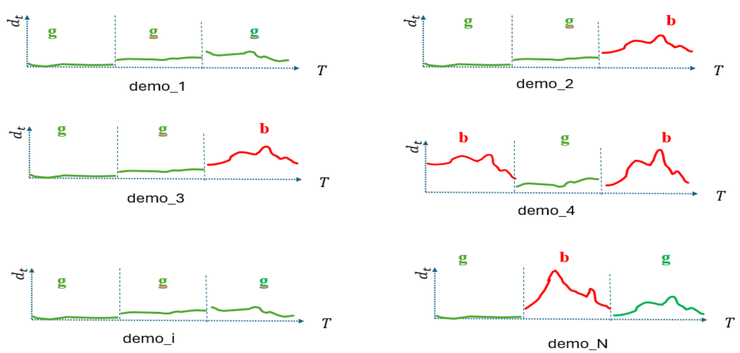
Learning Subtask Feature Distribution



Subtask Evaluation Using Mahalanobis Distance



Top r subtasks are bad



Results

Consistent Reduction in False Positives

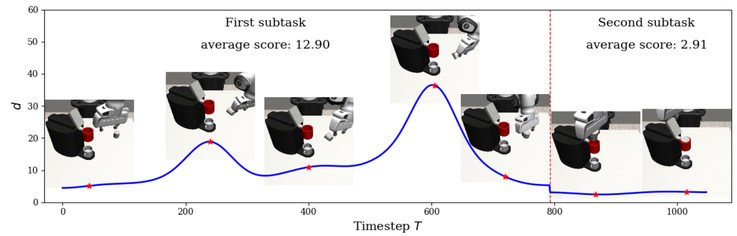
Task	Type 1	Type 2	Type 3
Lift	1/20	1/20	0/20
Can	2/30	4/30	13/30
Square	3/20	7/20	4/20
Drawer	2/14	2/14	3/14
Spoon	0/20	1/20	7/20

Average Success Rate Across 3 Seeds (150 Total Rollouts)

Policy	Task	Only good	All	BED-masked	LOF-masked	S2I-masked	GiB-masked
Diffusion	Square	0.40	0.27	0.33	0.43	0.11	**0.45**
	Coffee	0.68	0.68	0.68	0.62	0.56	**0.76**
	Mug	**0.74**	0.66	0.53	0.69	0.40	**0.74**
	Kitchen	0.61	**0.87**	0.69	0.78	0.75	0.85
BC-Transformer	Square	0.28	0.28	0.33	0.33	0.20	**0.52**
	Coffee	0.64	0.53	0.70	0.49	0.50	**0.76**
	Mug	0.51	0.41	**0.60**	0.53	0.30	**0.60**
	Kitchen	0.79	0.71	0.73	**0.87**	0.65	0.77

Distance as a Proxy for Quality

The distance plot reflects demonstration quality over time, with higher distances indicating lower-quality segments.



References

[1] Noushad Sojib and Momotaz Begum.
Self supervised detection of incorrect human demonstrations: A path toward safe imitation learning by robots in the wild.
In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2862–2869. IEEE, 2024.