Learning from Guided Play: A Scheduled Hierarchical Approach for Improving Exploration in Adversarial Imitation Learning

Trevor Ablett*, Bryan Chan*, Jonathan Kelly *(*equal contribution)*

University of Toronto

University of Toronto STARS Space & Terrestrial Autonomous Robotic Systems Laboratory

Code github.com/utiasSTARS/lfgp

Blog post papers.starslab.ca/lfgp

Motivation

Problem: Adversarial Imitation Learning (AIL) does not explicitly enforce good exploration.

Question: Can we use simple human-selected auxiliary tasks in a scheduled hierarchical model to improve performance?

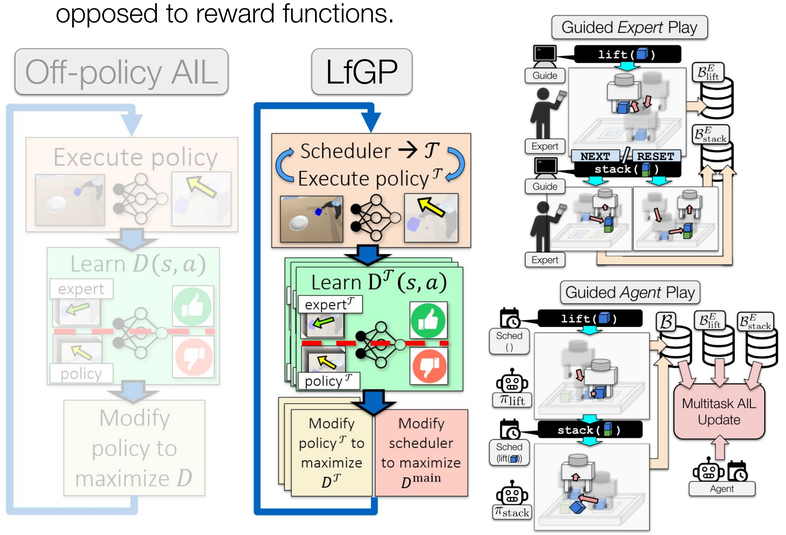
Approach

Our model combines off-policy AIL with a learned scheduler and a hierarchy of policies, discriminators, and Q-functions.

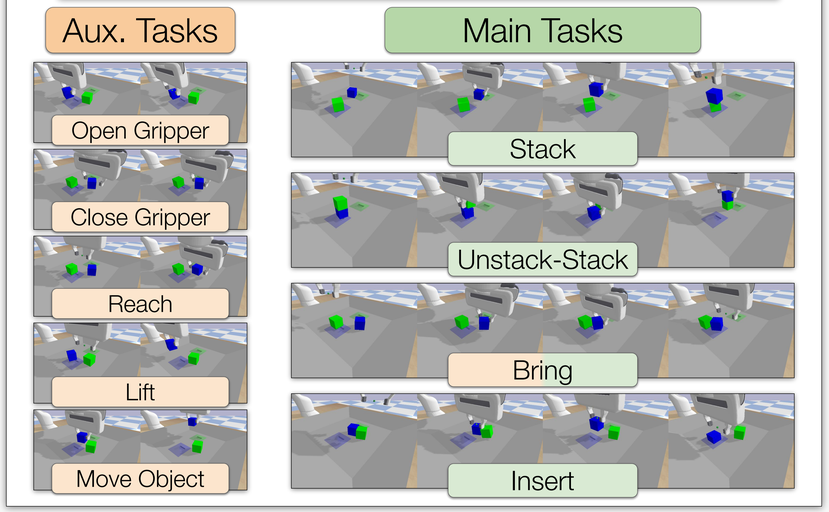
Play: Attempt multiple tasks in an environment.

Guided play:

1. *Expert* plays in environment, guided by uniform sampler.
2. *Agent* plays in environment, guided by expert data, as opposed to reward functions.

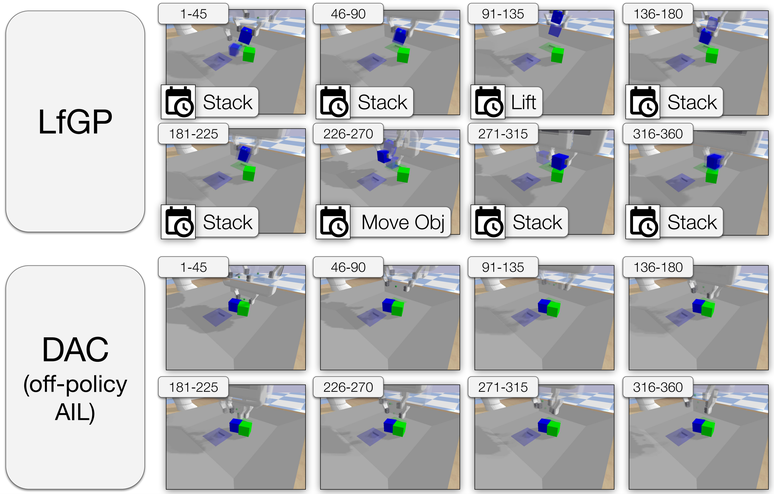


Multitask Manipulation Environment



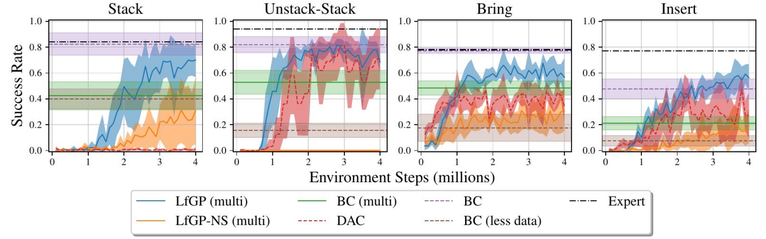
LfGP vs. DAC: Episode at 1M steps

A single episode of LfGP and DAC at 1M steps while learning Stack.



Results

Main Task Performance Comparison

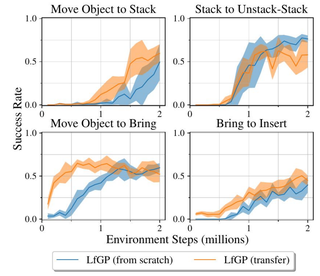


Expert Data Summary

	Task	Dataset Sizes	Reuse	Single	Total
Multi task	Stack	**OCSRLM**: 9k/task	**45k**	9k	54k
	U-Stack	**OCURLM**: 9k/task	**45k**	9k	54k
	Bring	**OCBRLM**: 9k/task	**54k**	0	54k
	Insert	**OCIBRLM**: 9k/task	**54k**	9k	63k
Single Task	Stack	S: 54k (low: 9k)	0	54k	54k
	U-Stack	U: 54k (low: 9k)	0	54k	54k
	Bring	B: 54k (low: 9k)	0	54k	54k
	Insert	I: 63k (low: 9k)	0	63k	63k

***O**pen Gripper, **C**lose Gripper, **S**tack, etc.* — ***bold:** reused*

Transfer Performance



- Single-task methods have 6-7x more main task data than multitask methods.
- Multitask methods *reuse* expert data between main tasks.
- Multitask methods can also transfer existing agents to new main tasks.

Conclusion

- LfGP outperforms both multitask baselines and single-task AIL (DAC) by enforcing exploration.
- Performs comparably to single-task BC, but allows **reusable** expert data and models.