DeepMind

Offline Learning from Demonstrations and Unlabeled Experience

Konrad Żołna Alexander Novikov Ksenia Konyushkova Caglar Gulcehre Ziyu Wang Yusuf Aytar Misha Denil Nando de Freitas Scott Reed

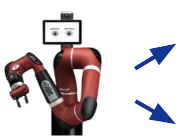
Abstract

Behavior cloning (BC) is often practical for robot learning because it allows a policy to be trained offline without rewards. However, BC does not effectively leverage what we will refer to as unlabeled experience: data of mixed and unknown quality without reward annotations. We introduce Offline Reinforced Imitation Learning (ORIL) that can use this unlabeled experience.

ORIL first learns a reward function by contrasting observations from demonstrator and unlabeled trajectories, then annotates all data with the learned reward, and finally trains an agent via offline reinforcement learning.

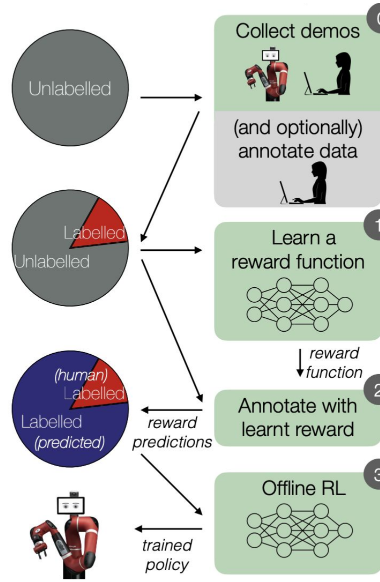
Across a diverse set of continuous control and simulated robotic manipulation tasks, we show that ORIL consistently outperforms comparable BC agents by effectively leveraging unlabeled experience.

Challenges in applying deep RL in real-world



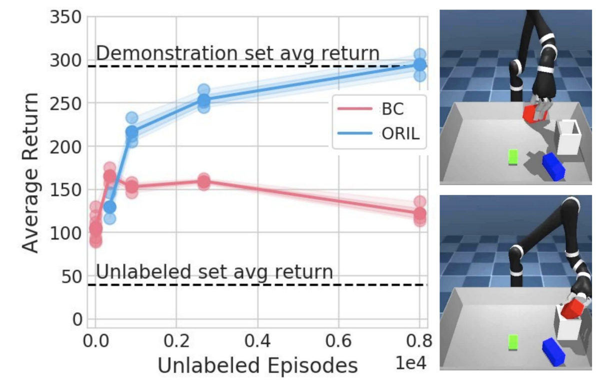
Requirement		Problem		Our solution
need a lot of data	→	expensive to collect	→	**use offline RL (CRR)**
need a reward signal	→	impractical or impossible	→	**learn reward function**

Our approach



Improvement from unlabeled episodes

ORIL achieves expert level using 200 demos + unlabeled data



Results for robotic manipulation tasks

Task	BC_{all}	BC_{pos}	ORIL	*CRR*
Box	158 ± 5	180 ± 7	**305** ± 3	*325 ± 4*
Insertion	146 ± 8	139 ± 5	**260** ± 3	*302 ± 12*
Slide	103 ± 2	181 ± 5	**214** ± 13	*312 ± 9*
Stack Banana	210 ± 12	129 ± 7	**257** ± 7	*300 ± 3*

- **Two variations of BC**: trained on all data or trained on demonstrations only
- **ORIL** (our method)
- Performance upper-bounds obtained with **CRR trained with ground-truth rewards**