

Hypernetwork-PPO for Continual Reinforcement Learning

Philemon Schöpf, Sayantan Auddy,
Jakob Hollenstein, Antonio Rodriguez-Sanchez

Intelligent and Interactive Systems Group
University of Innsbruck
Innsbruck, Austria

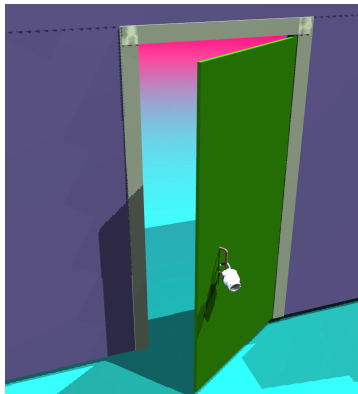
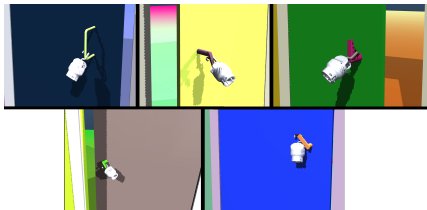
2022-11-10

Motivation

- Goal: Learn multiple tasks sequentially
- Cannot revisit old environment when learning new tasks
- Do not forget old skills

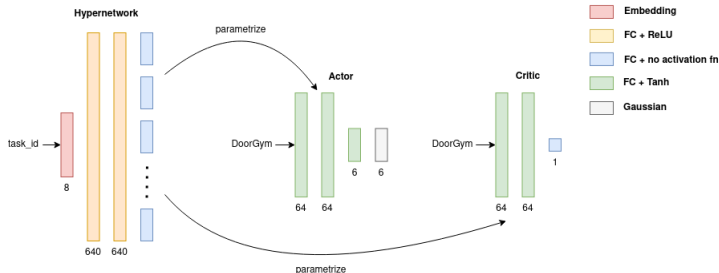
DoorGym

- Based on OpenAI Gym
- Robot arms try to open doors
- Multiple handles, opening directions
- Our experiments
 - “Floating hook” robot
 - 6 different kinds of doors

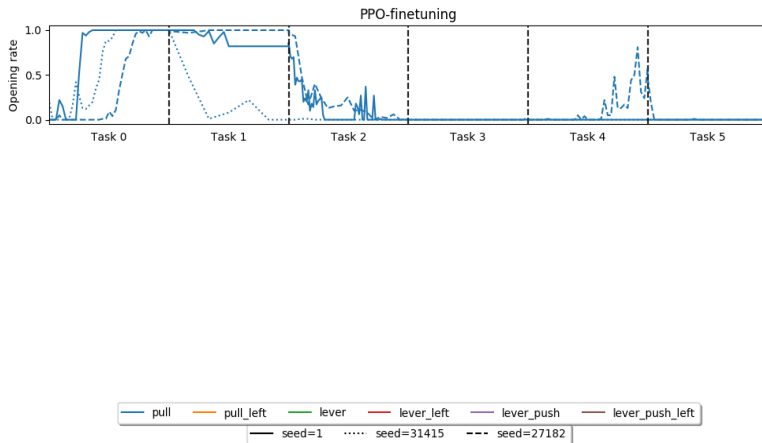


Hypernetworks

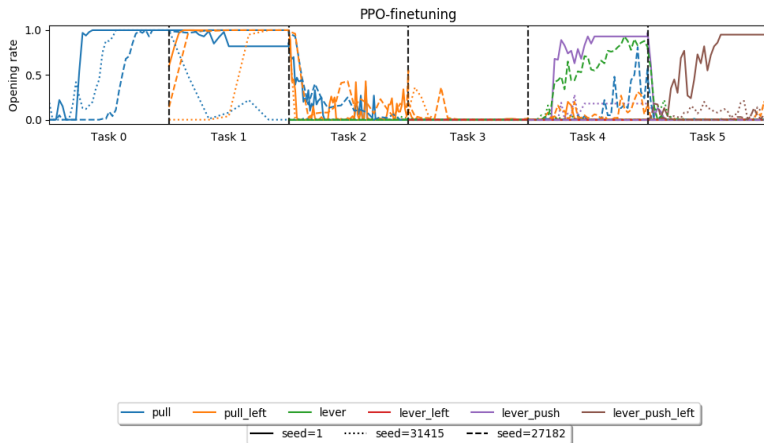
- Network that outputs a network
- Task ID as input
- Target networks determine policy/dynamics
- Regularization on changes of outputs for old tasks



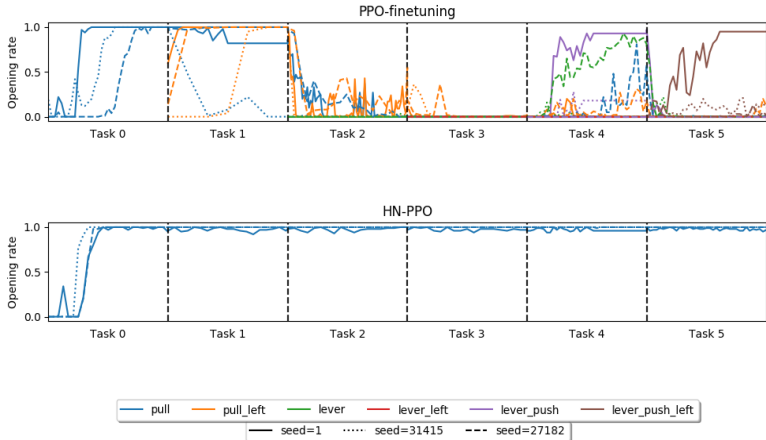
HN-PPO protects against catastrophic forgetting



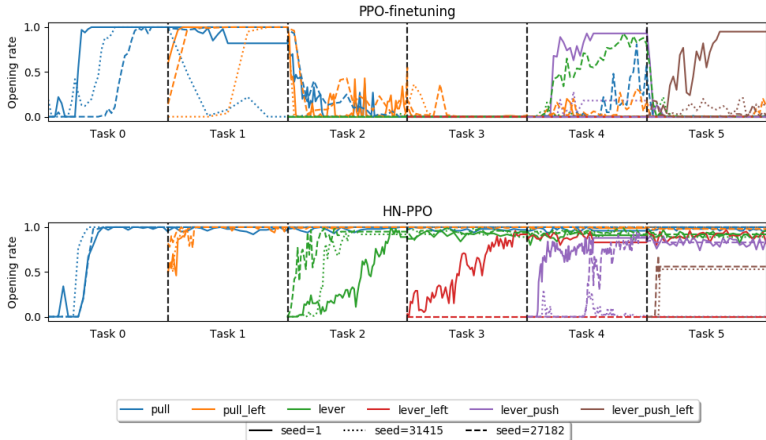
HN-PPO protects against catastrophic forgetting



HN-PPO protects against catastrophic forgetting



HN-PPO protects against catastrophic forgetting



HN-PPO protects against catastrophic forgetting

Metric	PPO-finetuning	HN-PPO
Accuracy	0.20 ± 0.035	0.81 ± 0.041
Remembering	0.47 ± 0.060	1.00 ± 0.0024