ICLR 2024 Spotlight

“Learning to Act without Actions”

Latent Action Policies (LAPO)

Dominik Schmidt
Weco AI

Minqi Jiang
FAIR at Meta AI

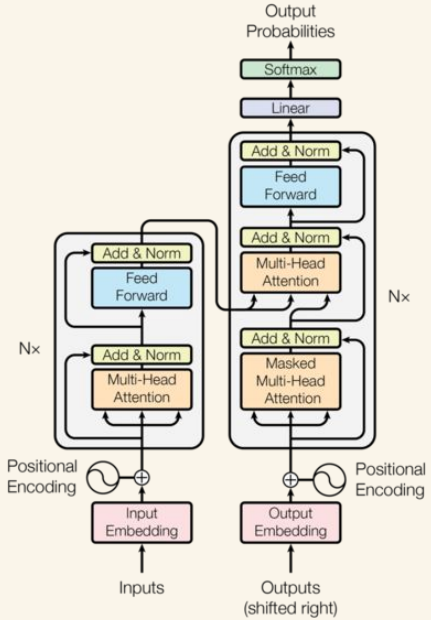
Paper: arxiv.org/abs/2312.10812
Code: github.com/schmidtdominik/LAPO

ICLR

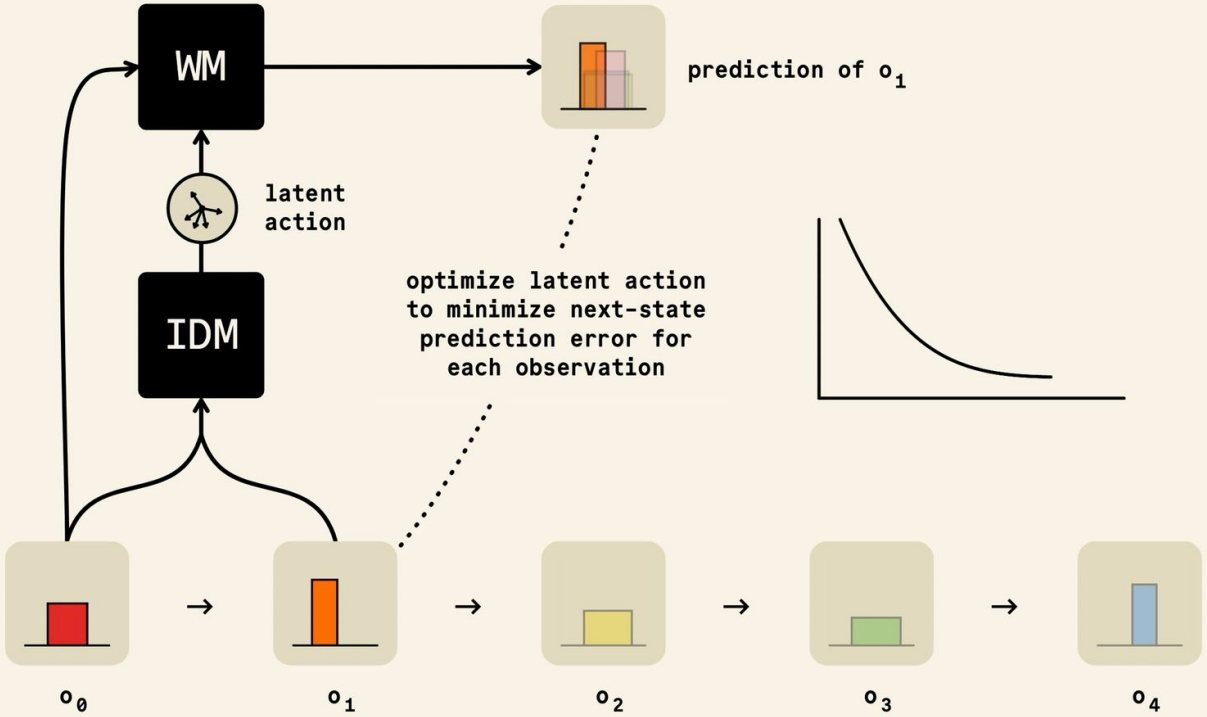
Motivation

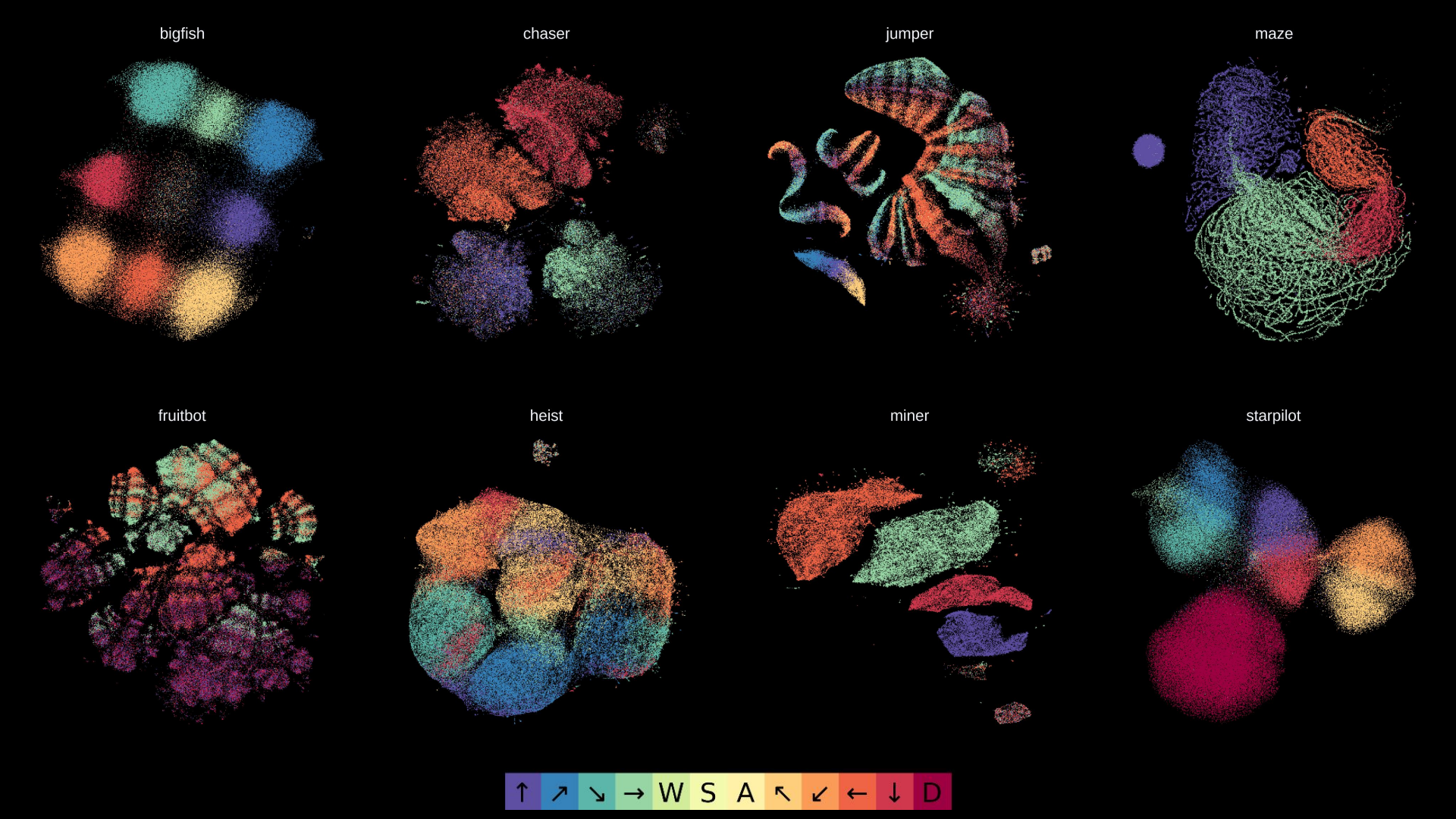
- Last few years: pretraining large *foundation models* on vast amounts of internet data
- There are vast amounts of observational data (videos) available, but RL can't use it
  - *offline RL* needs observations + reward + action labels
  - *imitation learning* needs observations + action labels
- Humans don't need action labels for imitation
- LAPO: recover latent action information from unlabeled video

  → train policies, world models,...

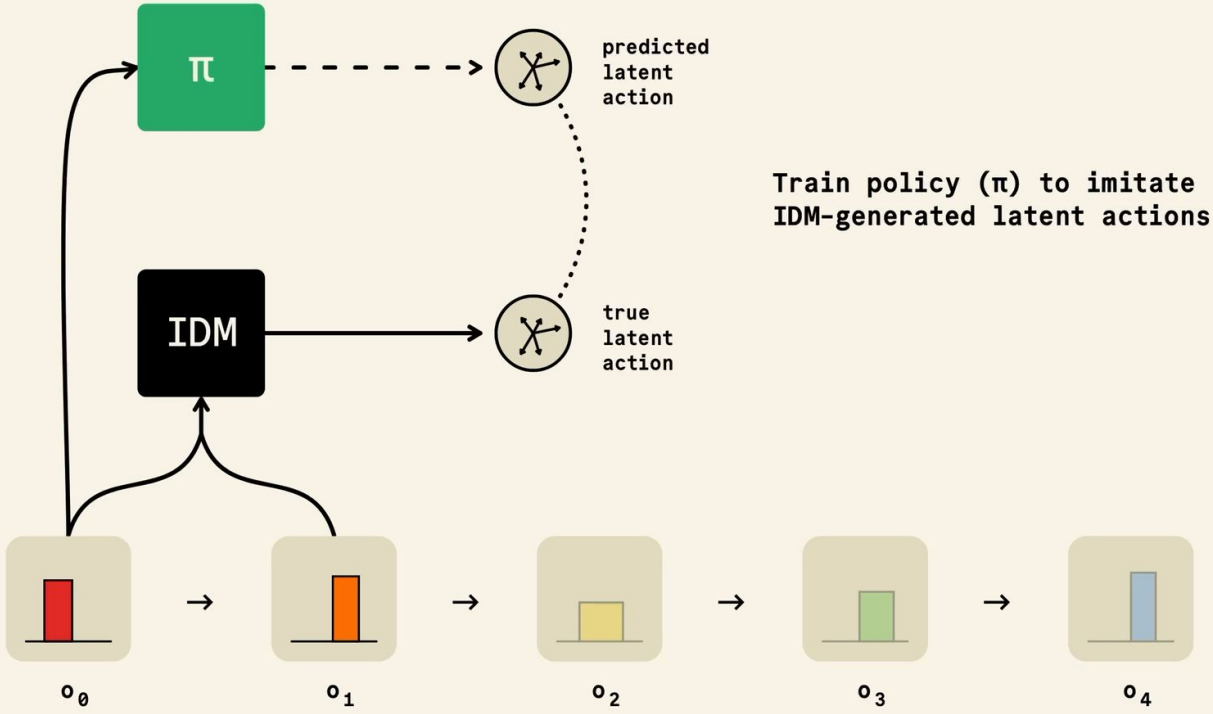


LAPO (stage 1) Training an inverse dynamics model (IDM) via a world model (WM)

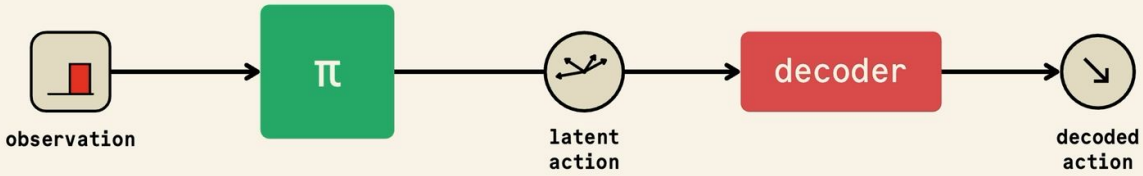



bigfish
chaser
jumper
maze
fruitbot
heist
miner
starpilot
↑ ↗ ↘ → W S A ↖ ↙ ← ↓ D

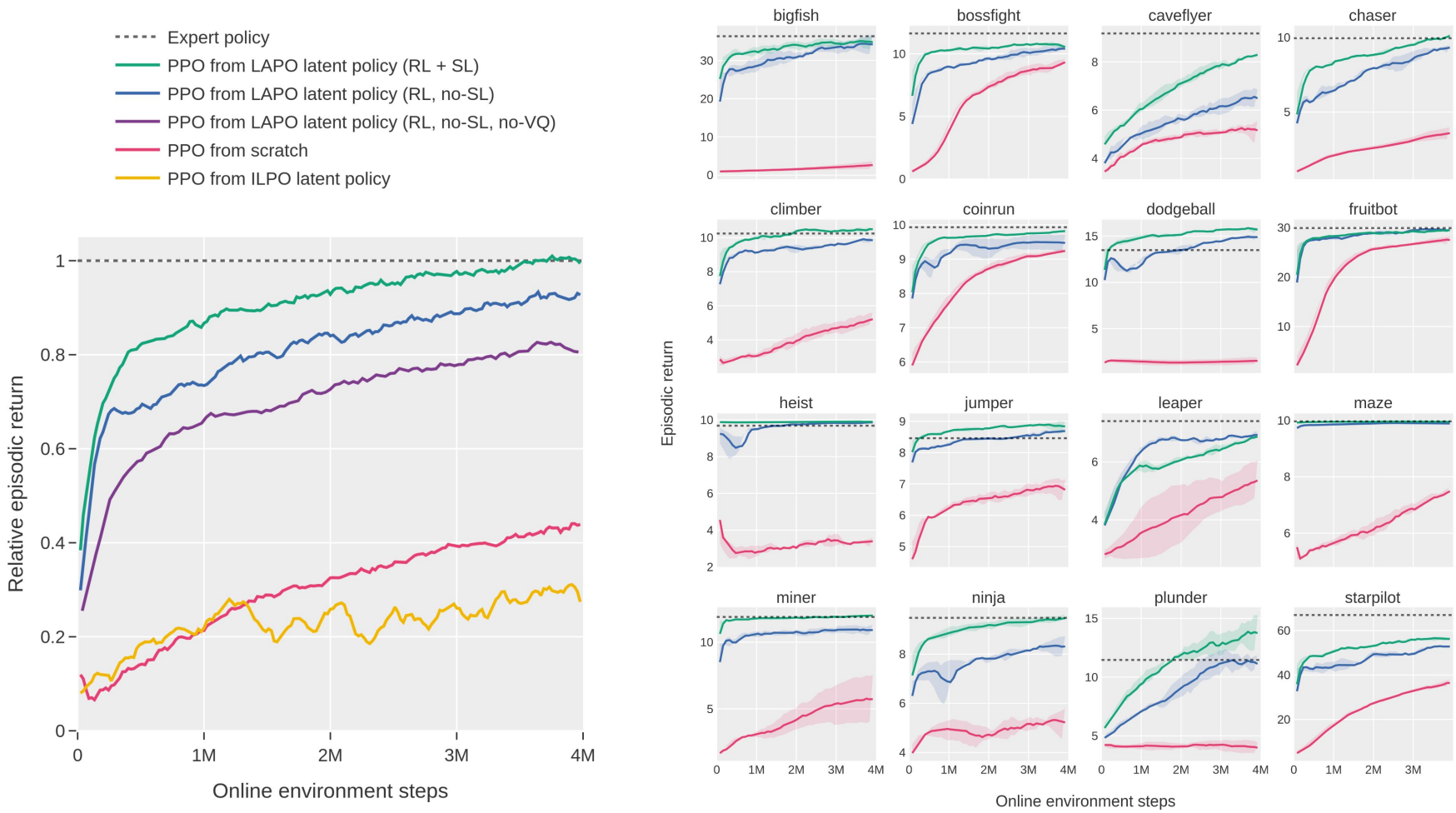
LAPO (stage 2) Behavior cloning a latent policy (π)

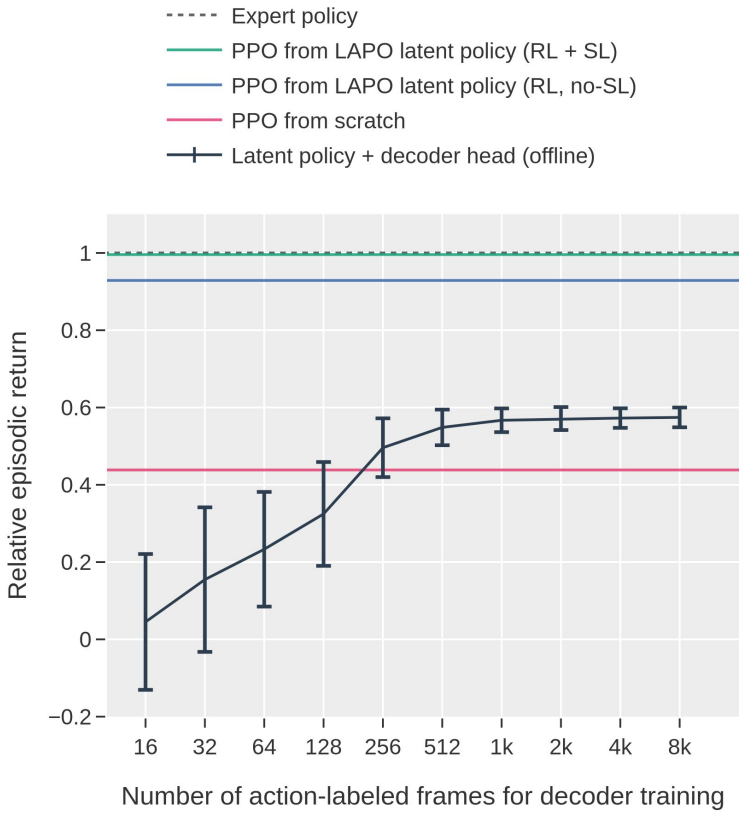


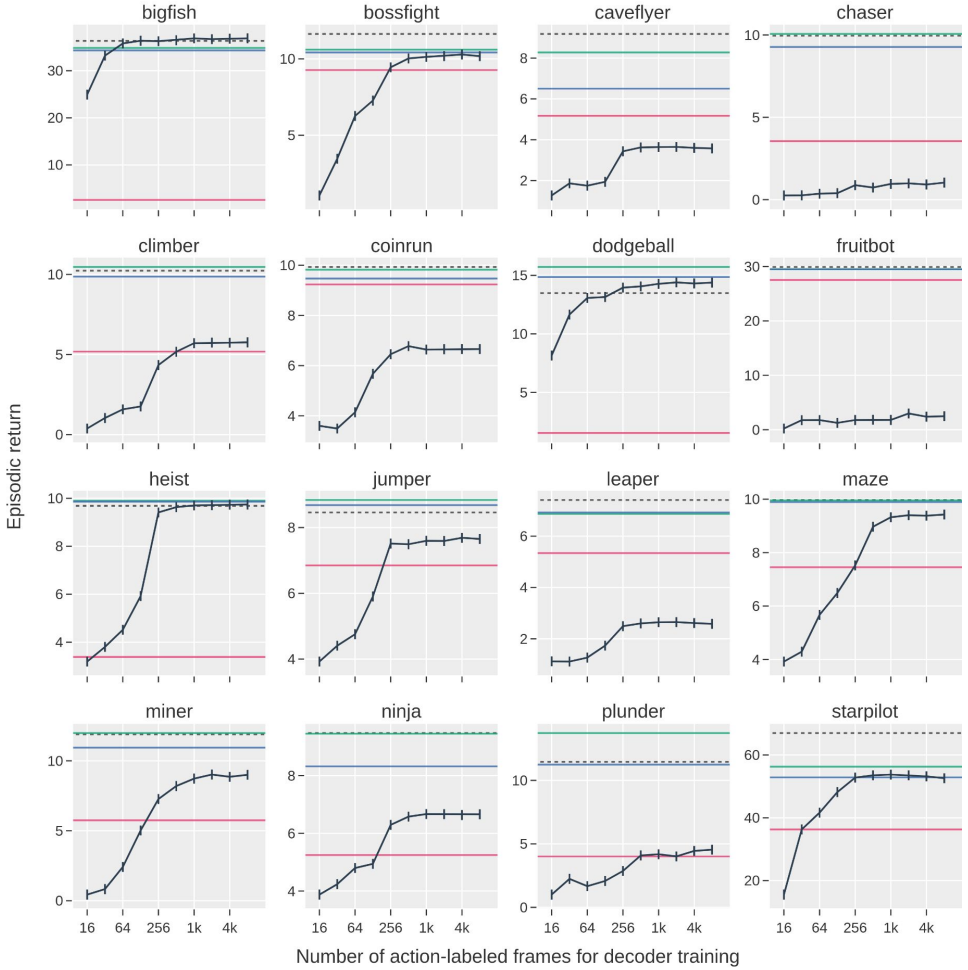
LAPO (stage 3) Decoding a latent policy (π)



Train decoder to decode latent actions into true actions


Expert policy
PPO from LAPO latent policy (RL + SL)
PPO from LAPO latent policy (RL, no-SL)
PPO from LAPO latent policy (RL, no-SL, no-VQ)
PPO from scratch
PPO from ILPO latent policy
Relative episodic return
Online environment steps
Episodic return
bigfish
bossfight
caveflyer
chaser
climber
coinrun
dodgeball
fruitbot
heist
jumper
leaper
maze
miner
ninja
plunder
starpilot


Expert policy
PPO from LAPO latent policy (RL + SL)
PPO from LAPO latent policy (RL, no-SL)
PPO from scratch
Latent policy + decoder head (offline)
Relative episodic return
1
0.8
0.6
0.4
0.2
0
−0.2
16
32
64
128
256
512
1k
2k
4k
8k
Number of action-labeled frames for decoder training


bigfish
bossfight
caveflyer
chaser
climber
coinrun
dodgeball
fruitbot
heist
jumper
leaper
maze
miner
ninja
plunder
starpilot
Episodic return
16
64
256
1k
4k
Number of action-labeled frames for decoder training

ICLR 2024 Spotlight

“Learning to Act without Actions”

Latent Action Policies (LAPO)

Dominik Schmidt
Weco AI

Minqi Jiang
FAIR at Meta AI

Paper: arxiv.org/abs/2312.10812
Code: github.com/schmidtdominik/LAPO

ICLR