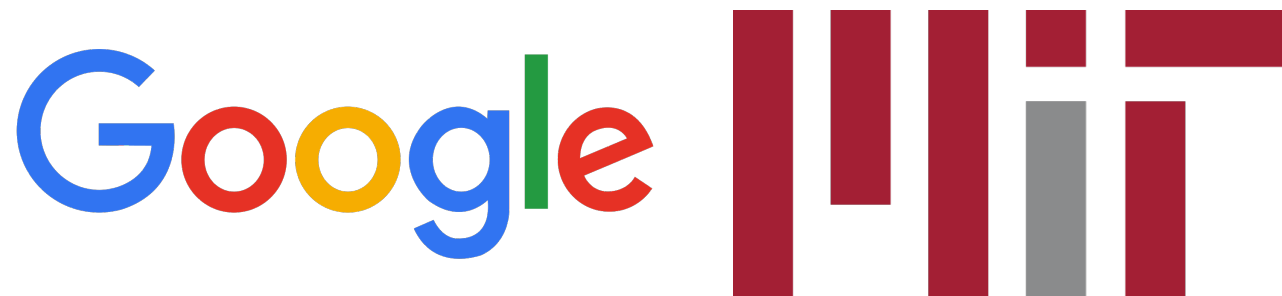


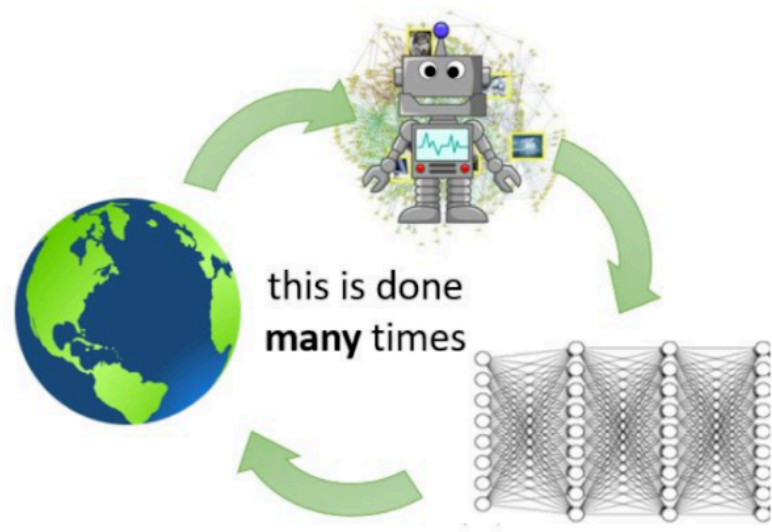
OPAL: Offline Primitive Discovery for Accelerating Offline Reinforcement Learning

Anurag Ajay

(w/ Aviral Kumar, Pulkit Agrawal, Sergey Levine, Ofir Nachum)



Reinforcement Learning



Not always practical



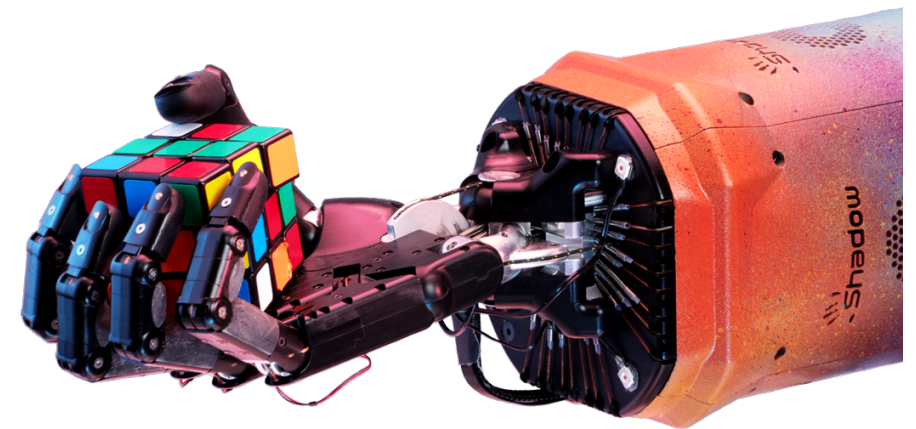
(Silver et al)



(Vinyals et al)

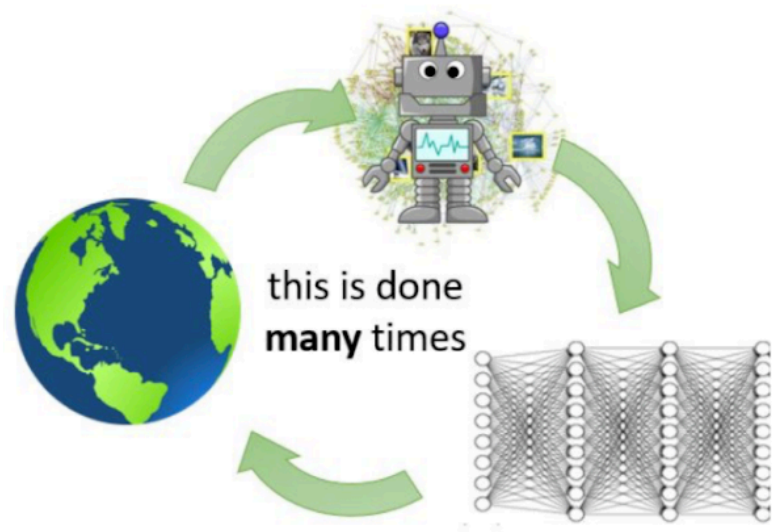


(Levine et al)



(OpenAI)

Reinforcement Learning



Not always practical

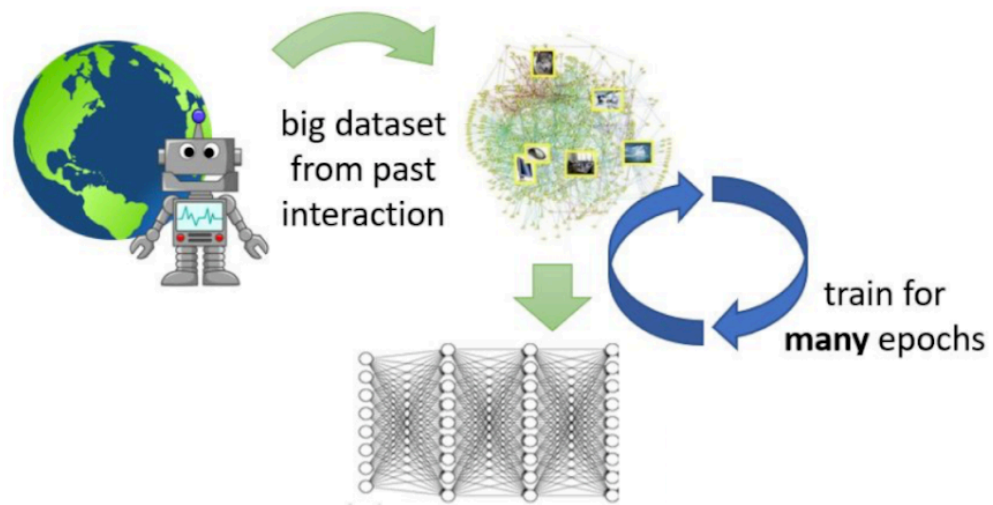
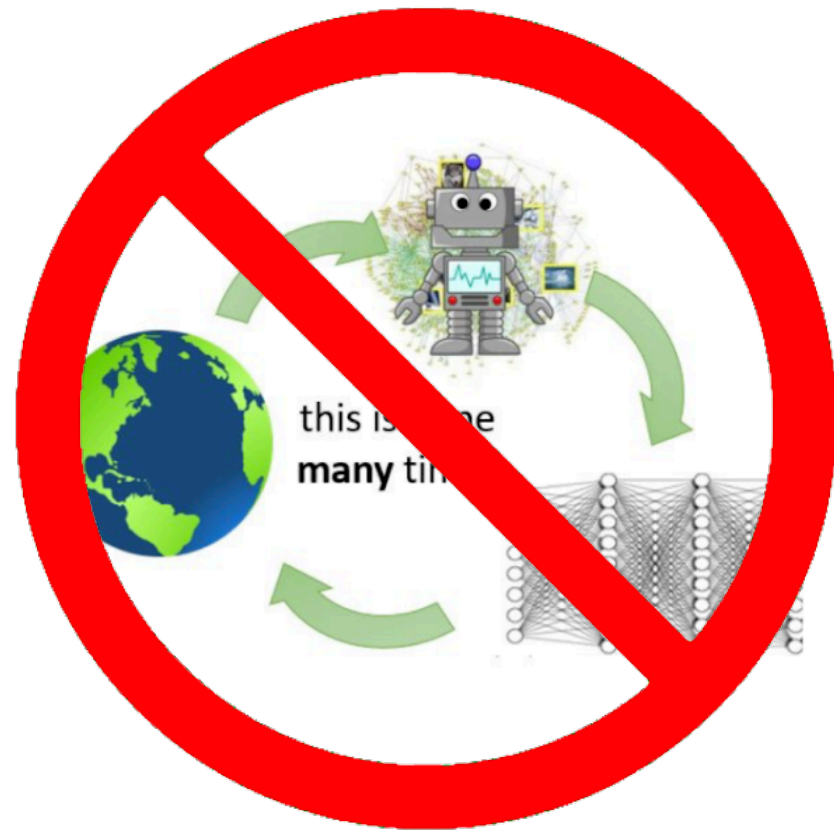


```
[User]: i'm just at work.
[RL - Model avg Psi]: 😊😊😊😊 i'm not sure, but i'm glad to see what you did!
[User]: you're boring.
[RL - Model avg Psi]: 😊👉😊 why would you say this? what's up to you?!?!?!!!!!!!
[User]: i'm sorry! I didn't mean to offend you.
[RL - Model avg Psi]: 😊😊 i'm not sure, and i think i'm going to do this, and i'm sure you'll go!!
[User]: i will go, you're being weird!
[RL - Model avg Psi]: 😊👉 we are the same, we don't understand that, sir.
[User]: we are not the same. i'm a human and you're a bot.
[RL - Model avg Psi]: 🌸💖👉 that's really nice to me too.
>
```

Close Chat and Rate

SEND

(Offline) Reinforcement Learning

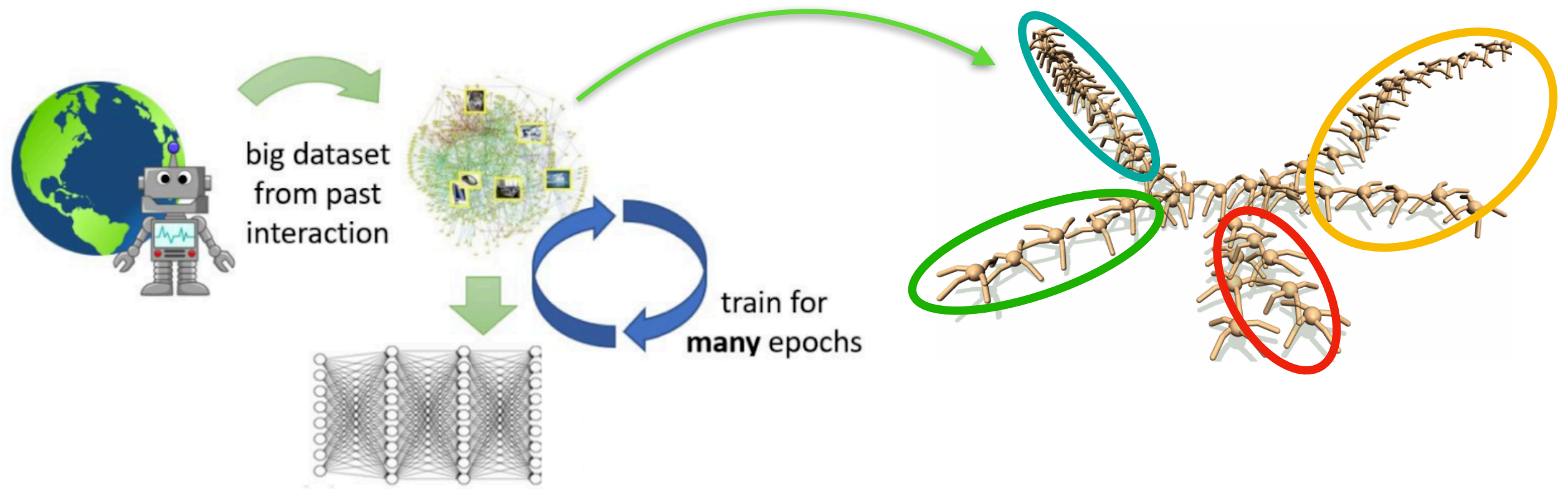


```
[User]: i'm just at work.
[RL - Model avg Psi]: 😊😊😊😊 i'm not sure, but i'm glad to see what you did!
[User]: you're boring.
[RL - Model avg Psi]: 😊👉😊 why would you say this? what's up to you?!?!?!!!!!!!
[User]: i'm sorry! I didn't mean to offend you.
[RL - Model avg Psi]: 😊😊 i'm not sure, and i think i'm going to do this, and i'm sure you'll go!!
[User]: i will go, you're being weird!
[RL - Model avg Psi]: 😊👉 we are the same, we don't understand that, sir.
[User]: we are not the same. i'm a human and you're a bot.
[RL - Model avg Psi]: 🌸💖👉 that's really nice to me too.
>
```

Close Chat and Rate

SEND

Improving Offline RL with *primitives*



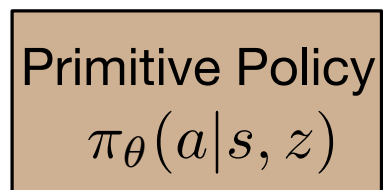
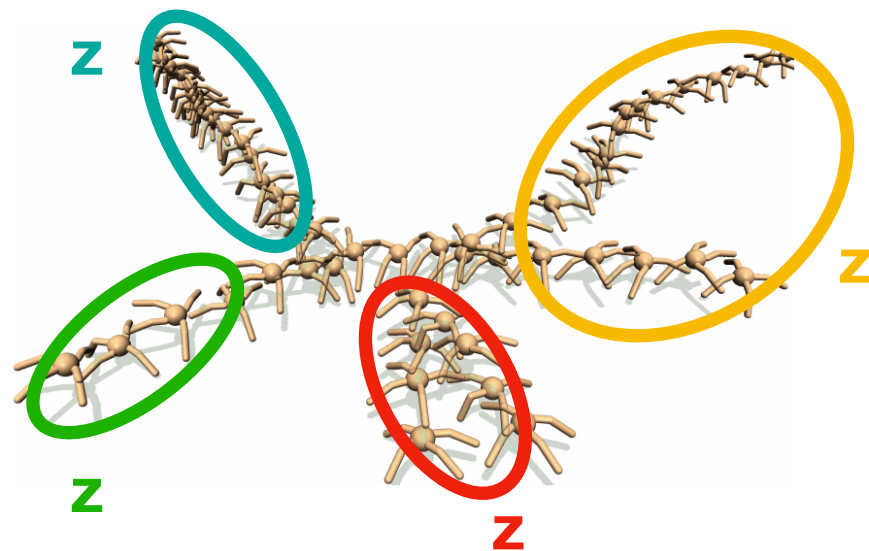
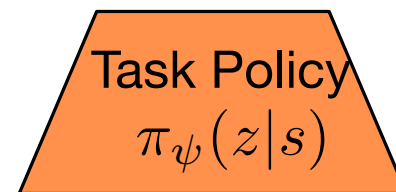
Idea: Improve Offline RL by leveraging *primitives* present in a *multi-modal* dataset

Intuition for OPAL



+1 for reaching goal

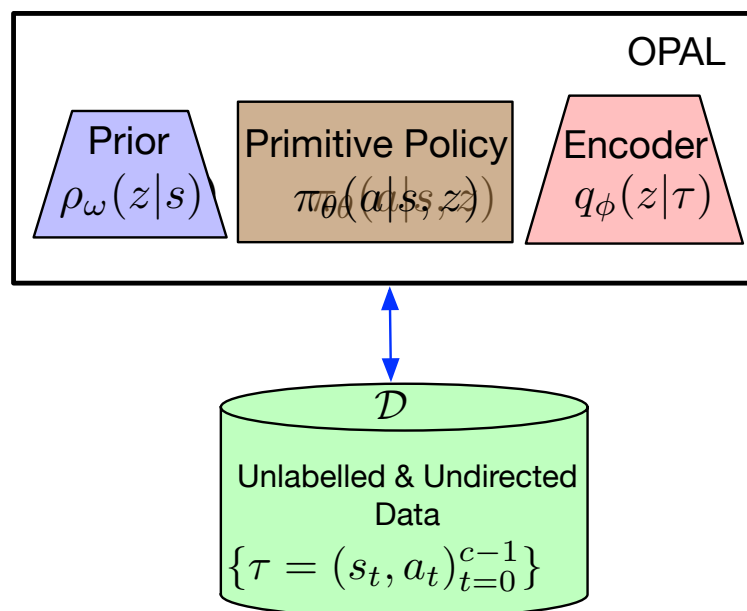
Given a task
(i.e. a reward
function)



Reduced task horizon & Compact
action space → Better Credit
Assignment

OPAL: Offline Primitive Discovery for Accelerating RL

Forward $\dashrightarrow$ Autoencoding loss + KL constraint $\longleftrightarrow$ Fine-tuning with BC $\longleftrightarrow$ Offline RL $\longleftrightarrow$

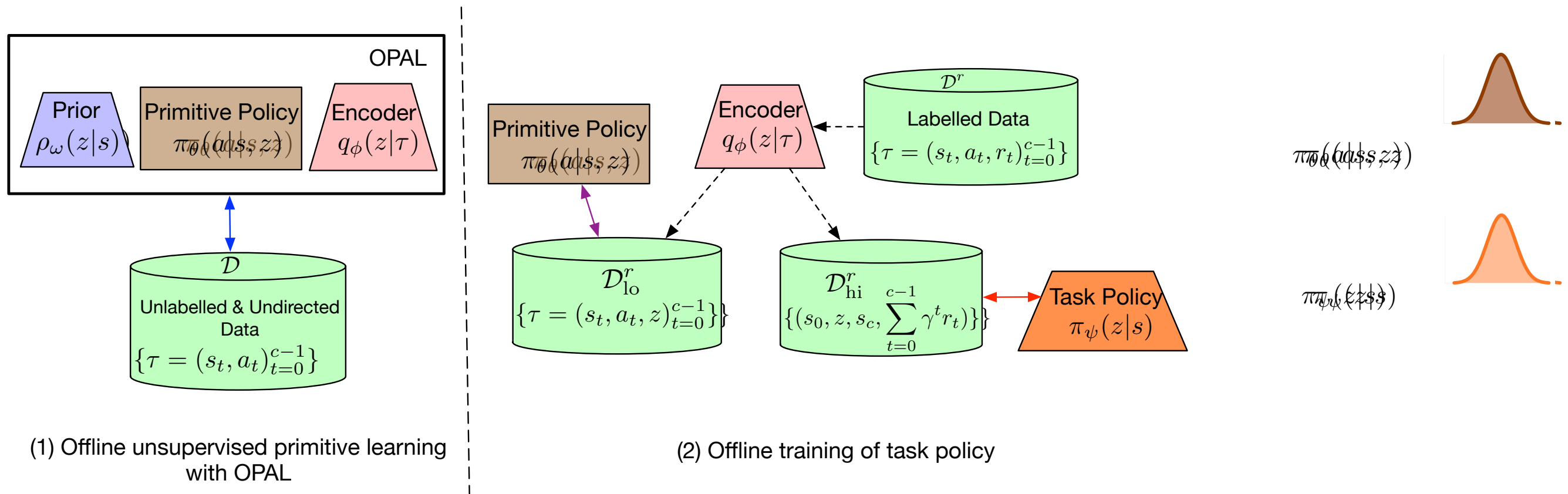


(1) Offline unsupervised primitive learning with OPAL

$$\begin{array}{ccccc}
 & & \mathcal{D}^r & & \\
 \pi_{\theta\phi}(a|s, z) & q_\phi(z|\tau) & \{\tau = (s_t, a_t, r_t)_{t=0}^{c-1}\} & \pi_{\theta\phi}(a|s, z) & \text{Graphical distribution} \\
 \mathcal{D}_{\text{do}}^r & \mathcal{D}_{\text{hi}}^r & & \pi_{\psi}(z|s) & \text{Graphical distribution} \\
 \{\tau = (s_t, a_t, z)_{t=0}^{c-1}\} & \{(s_0, z, s_c, \sum_{t=0}^{c-1} \gamma^t r_t)\} & & &
 \end{array}$$

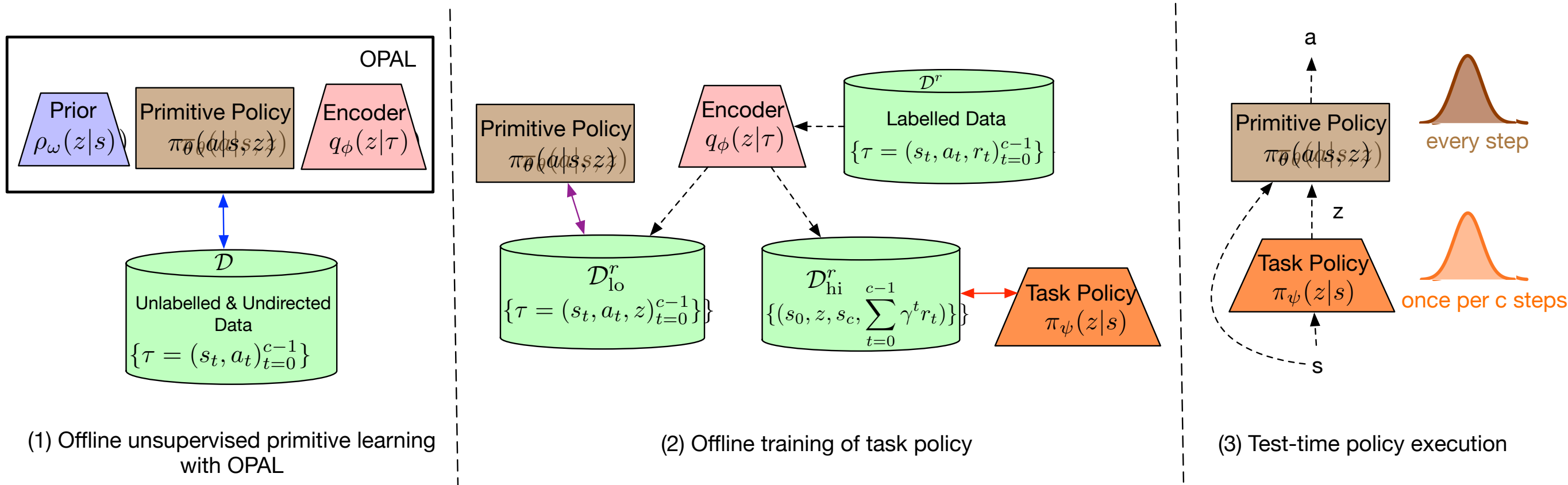
OPAL: Offline Primitive Discovery for Accelerating RL

Forward $\dashrightarrow$ Autoencoding loss + KL constraint $\longleftrightarrow$ Fine-tuning with BC $\longleftrightarrow$ Offline RL $\longleftrightarrow$



OPAL: Offline Primitive Discovery for Accelerating RL

Forward $\dashrightarrow$ Autoencoding loss + KL constraint $\longleftrightarrow$ Fine-tuning with BC $\longleftrightarrow$ Offline RL $\longleftrightarrow$



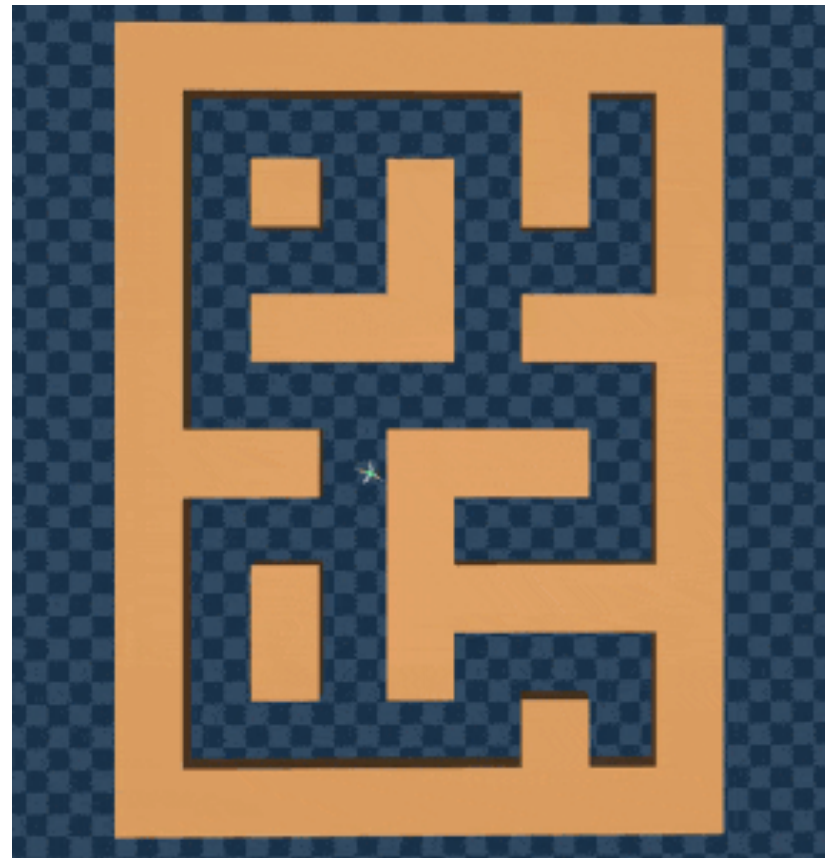
Environments



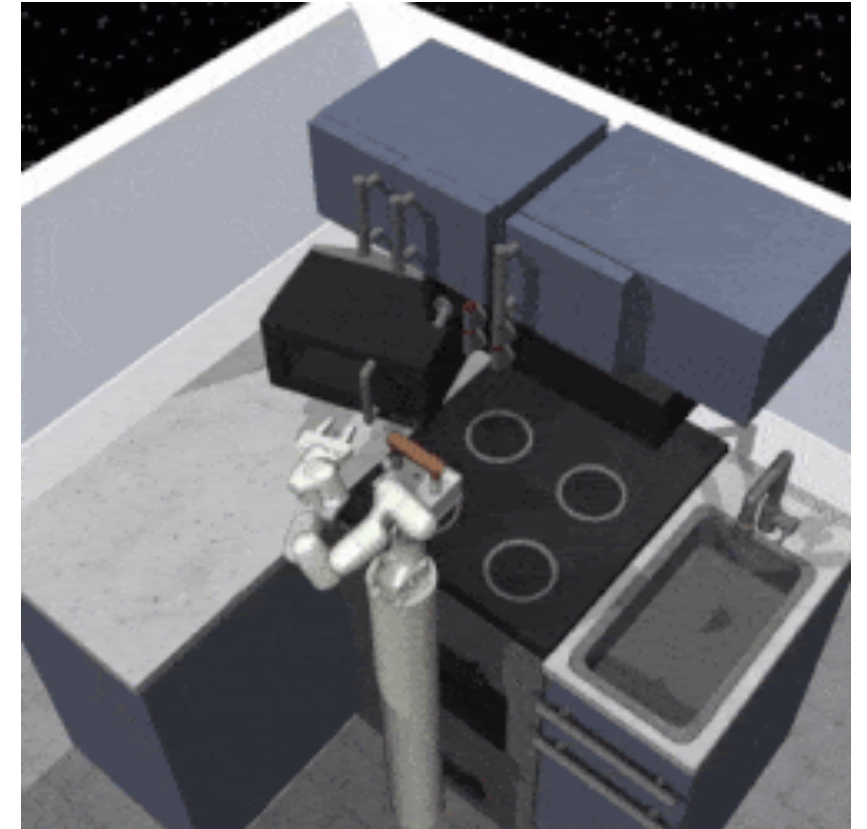
Antmaze medium

State (joint angles + xy pose): 29 dim

Action (joint torques): 8 dim



Antmaze large



kitchen

State (joint angles + xy pose): 60 dim

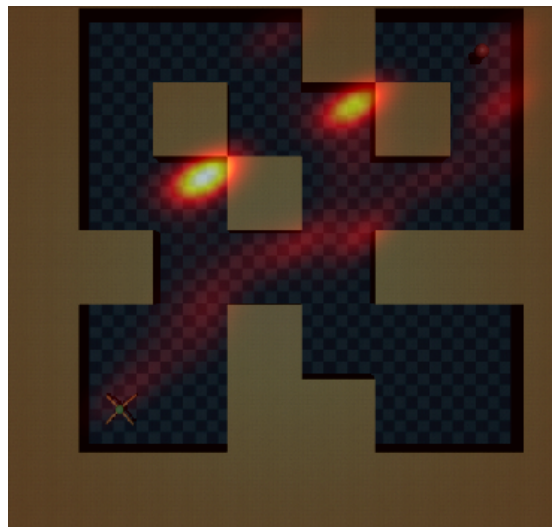
Action (joint torques): 9 dim

(from D4RL (fu et. al.))

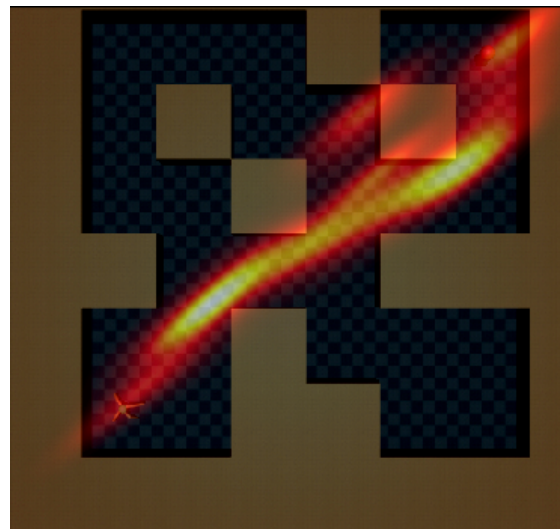
Results

Offline RL

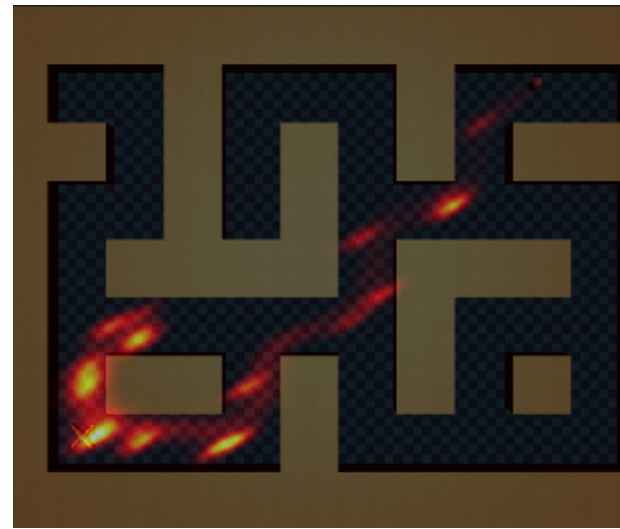
Environment	BC	BEAR (Kumar et al., 2019)	EMAQ (Ghasemipour et al., 2020)	CQL (Kumar et al., 2020)	CQL+OPAL (ours)
antmaze medium (diverse)	0.0	8.0	0.0	53.7 ± 6.1	81.1 ± 3.1
antmaze large (diverse)	0.0	0.0	0.0	14.9 ± 3.2	70.3 ± 2.9
kitchen mixed	47.5	47.2	70.8 ± 2.3	52.4 ± 2.5	69.3 ± 2.7
kitchen partial	33.8	13.1	74.6 ± 0.6	50.1 ± 1.0	80.2 ± 2.4



CQL (medium)



CQL + OPAL (medium)



CQL (large)



CQL + OPAL (large)

Results

Offline RL

Environment	BC	BEAR (Kumar et al., 2019)	EMAQ (Ghasemipour et al., 2020)	CQL (Kumar et al., 2020)	CQL+OPAL (ours)
antmaze medium (diverse)	0.0	8.0	0.0	53.7 ± 6.1	81.1 ± 3.1
antmaze large (diverse)	0.0	0.0	0.0	14.9 ± 3.2	70.3 ± 2.9
kitchen mixed	47.5	47.2	70.8 ± 2.3	52.4 ± 2.5	69.3 ± 2.7
kitchen partial	33.8	13.1	74.6 ± 0.6	50.1 ± 1.0	80.2 ± 2.4

Few-shot imitation learning

Online RL

Online multi-task transfer learning

Summary

OPAL discovers a continuous space of *primitives* from a *multi-modal* dataset which can be used to

- Improve offline RL
- Improve few-shot imitation learning
- Accelerate online RL and multi-task learning

Check out paper/poster for

- Detailed description of OPAL
- Theoretical guarantees
- Results on few-shot imitation learning, online RL and multi-task learning

Visualization & code:



<https://sites.google.com/view/opal-iclr>