Unsupervised Curricula for Visual Meta-Reinforcement Learning

Allan Jabri, Kyle Hsu, Ben Eysenbach, Abhishek Gupta, Sergey Levine, Chelsea Finn

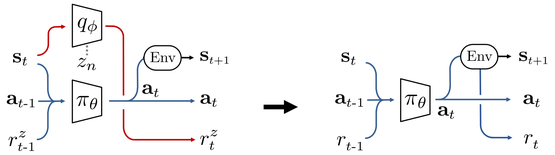
BAIR BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Berkeley UNIVERSITY OF CALIFORNIA · UNIVERSITY OF TORONTO · Carnegie Mellon University

Motivation

Meta-RL provides a formulation for transfer between different tasks, but relies on specification of a task distribution

→ Can useful meta-RL strategies be acquired by pretraining with task distribution formed through unsupervised interaction?



Unsupervised Pre-training → **Transfer** to Test Tasks

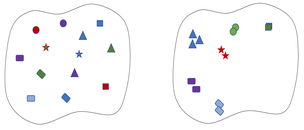
Key Challenges

1. Task acquisition with high-dimensional observations, i.e. pixels.
2. Adaptation of task distribution and meta-learner, i.e. curriculum.

Formulation

Intuition: Prepare meta-RL agent with task inference and execution strategies that transfer

→ Task dist. should balance Diversity and Structure



$H(\boldsymbol{\tau})$ $\quad -H(\boldsymbol{\tau}|\mathbf{z})$

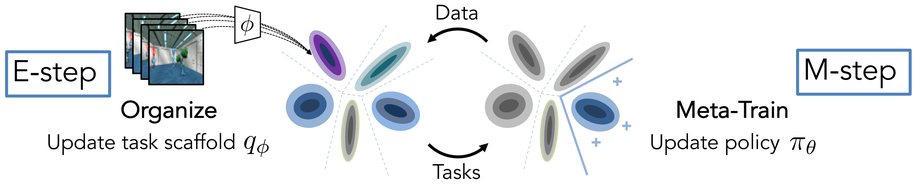
$$\max_{\theta,\phi} I(\boldsymbol{\tau};\mathbf{z})$$

π_θ Meta-RL policy
q_ϕ Learned task scaffold
$\boldsymbol{\tau}$ Post-update trajectories
$\mathbf{z}$ Task latent variable

Information Maximization via Variational EM

Method

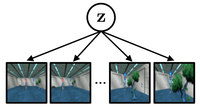
Main Idea: Jointly discover and meta-learn tasks by alternating between task acquisition and meta-reinforcement learning of updated task distribution.



(E) Trajectory-level generative model with discriminative features

$$\max_{\phi} \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}),\boldsymbol{\tau}\sim\mathcal{D}}\left[\log q_\phi(\boldsymbol{\tau}|\mathbf{z})\right]$$

w/ conditional independence $q_\phi(\mathbf{s}) = \sum_{\mathbf{z}} q_\phi(\mathbf{s}|\mathbf{z})p(\mathbf{z})$



(M) Policy meta-learns task distribution

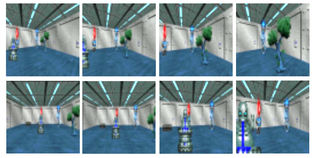
$$\max_{\theta} \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}),\mathbf{s}\sim\pi_\theta(\mathbf{s}|\mathbf{z})}\left[\log q_\phi(\mathbf{s}|\mathbf{z}) - \log q_\phi(\mathbf{s})\right]$$

i.e. learn tasks indexed by $\mathbf{z}$

$$r_{\mathbf{z}}(\mathbf{s}) = \log q_\phi(\mathbf{s}|\mathbf{z}) - \log q_\phi(\mathbf{s})$$

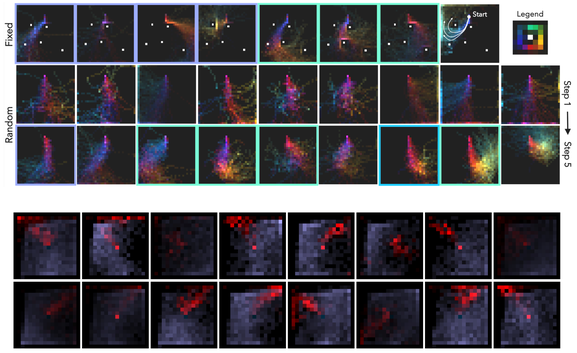
Environments and Discovered Tasks

Visual Navigation in ViZDoom



Visuomotor Control w/ MuJoCo Sawyer

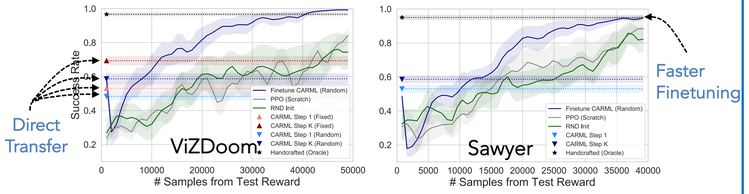
Visualization of Discovered Task Distributions



Experiments

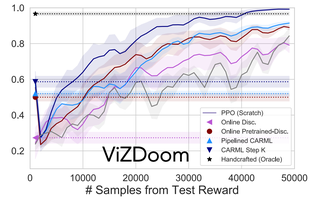
Do acquired meta-RL strategies transfer to test task distributions?

Direct Transfer: Apply policy to test tasks without finetuning
Finetuning: Update parameters of policy to specific test tasks

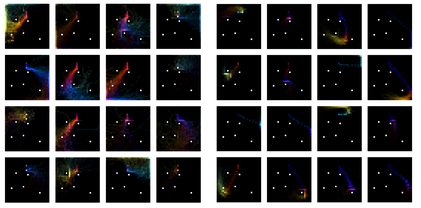


Comparing Variants: Effect of Task Acquisition and Joint Optimization

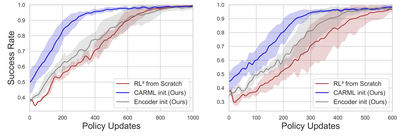
Online disc.: Discriminative q_ϕ (DIAYN)
Pipelined CARML: 1 E-step, 1 M-step



Avoiding mode collapse in task distribution



Meta-Pretraining: Accelerated Supervised Meta-RL



Allows for accelerated supervised learning of test task distributions **even when controlling for pretrained encoder**

Takeaways

- Proposed task acquisition strikes balance between generative and discriminative approaches for modeling pixel trajectories, avoiding task mode-collapse
- Task distribution supports meta-RL of strategies that transfer to related but unseen hand-crafted tasks
- Adapting tasks with the policy eases meta-learning with curriculum
- ? Need more task-specific bias for direct transfer, more complex tasks