Efficient Meta-RL from Pixels

Kelly Marshall and Chingun Khasar 2021

High Level overview

Kelly & Chingun Fall 2021

Recap: Meta- Reinforcement Learning

Recap: Meta- Reinforcement Learning

M1 M2 M1

Mtest

Recap: Meta- Reinforcement Learning

$$L(\pi_\theta) = \mathbb{E}_{\tau_i \sim p(\tau)}[\mathbb{E}_{s_t, a_t \sim \pi_{\theta_i}}[\sum_{t=1}^{H} \gamma^t r_t]]$$

Meta-RL structure PEARL

Probabilistic Embeddings for Actor-Critic Learning

Algorithm 1 PEARL Meta Training

Require: Batch of training tasks $\{\tau\}_{i=1..T}$ from p(τ), learning rates $\alpha_1, \alpha_2, \alpha_3$

Initialize replay buffer $\mathcal{B}^i$ for each training task
while not done **do**
  for each τ_i **do**
    Initialize context $\mathbf{c}^i = \{\}$
    for $k = 1...K$ **do**
      Sample $z \sim q_\phi(z|c^i)$
      Gather data from $\pi_\theta(a|s,z)$ and add to $\mathcal{B}_i$
      Update $c_i = \{(s_j, a_j, r_j)\}_{j=1..N}$
    end for
  end for
  for each step in training steps **do**
    for each τ_i **do**
      Sample context $c^i \sim \mathcal{S}_c(\mathcal{B}_i)$ and RL batch $b^i \sim \mathcal{B}_i$
      Sample $z \sim q_\phi(z|c^i)$
      $\mathcal{L}^i_{actor} = \mathcal{L}_{actor}(b^i, \mathbf{z})$
      $\mathcal{L}^i_{critic} = \mathcal{L}_{critic}(b^i, \mathbf{z})$
      $\mathcal{L}^i_{KL} = \beta D_{KL}(q_\phi(\mathbf{z}|\mathbf{c^i})||r(\mathbf{z}))$
    end for
    $\phi \leftarrow \phi - \alpha_1 \nabla_\phi \sum_i (\mathcal{L}^i_{critic} + \mathcal{L}^i_{KL})$
    $\theta_\pi \leftarrow \theta_\pi - \alpha_2 \nabla_\theta \sum_i \mathcal{L}^i_{actor}$
    $\theta_Q \leftarrow \theta_Q - \alpha_3 \nabla_\theta \sum_i \mathcal{L}^i_{critic}$
  end for
end while

Algorithm 2 PEARL Meta Testing

Require: Test task $\tau \sim p(\tau)$

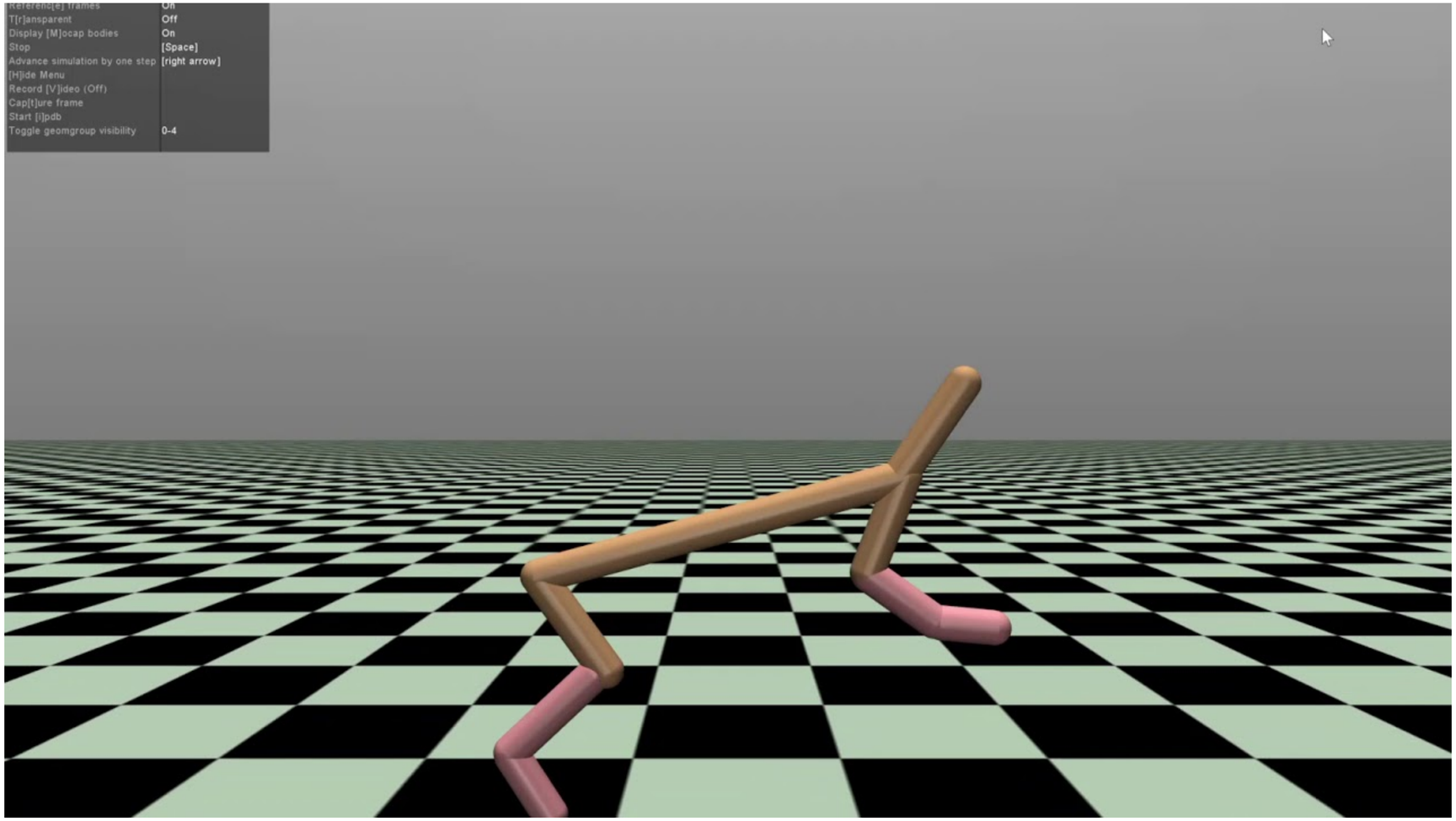
Initialize context $\mathbf{c}^\tau = \{\}$
for k=1,..,K **do**
  Sample $z \sim q_\phi(\mathbf{z}|c^\tau)$
  Roll out policy $\pi_\theta(\mathbf{a}|\mathbf{s},\mathbf{z})$ to collect data $D^\tau_k = \{\mathbf{s}_j, \mathbf{a}_j, \mathbf{s}', r_j\}$
  Accumulate context $\mathbf{c}^\tau = \mathbf{c}^\tau \bigcup D^\tau_k$
end for

Our Additions

Kelly & Chingun Fall 2021

Image + meta-RL

On Half-Cheetah



Design Questions of Image + meta-RL

- What loss function to use?

- How to balance the use of true state representations?

CNN + PEARL

1. Train CNN model image to state classification.
2. Train RL using state vectors.
3. During test use CNN predicted vectors

Distribution Shift from +Image

CNN error + RL error = bad Agent

CNN + PEARL

1. Train CNN model for understanding image to state classification.
2. Train RL ~~using state vectors~~ **CNN predicted vectors**.
3. During test use CNN predicted vectors.

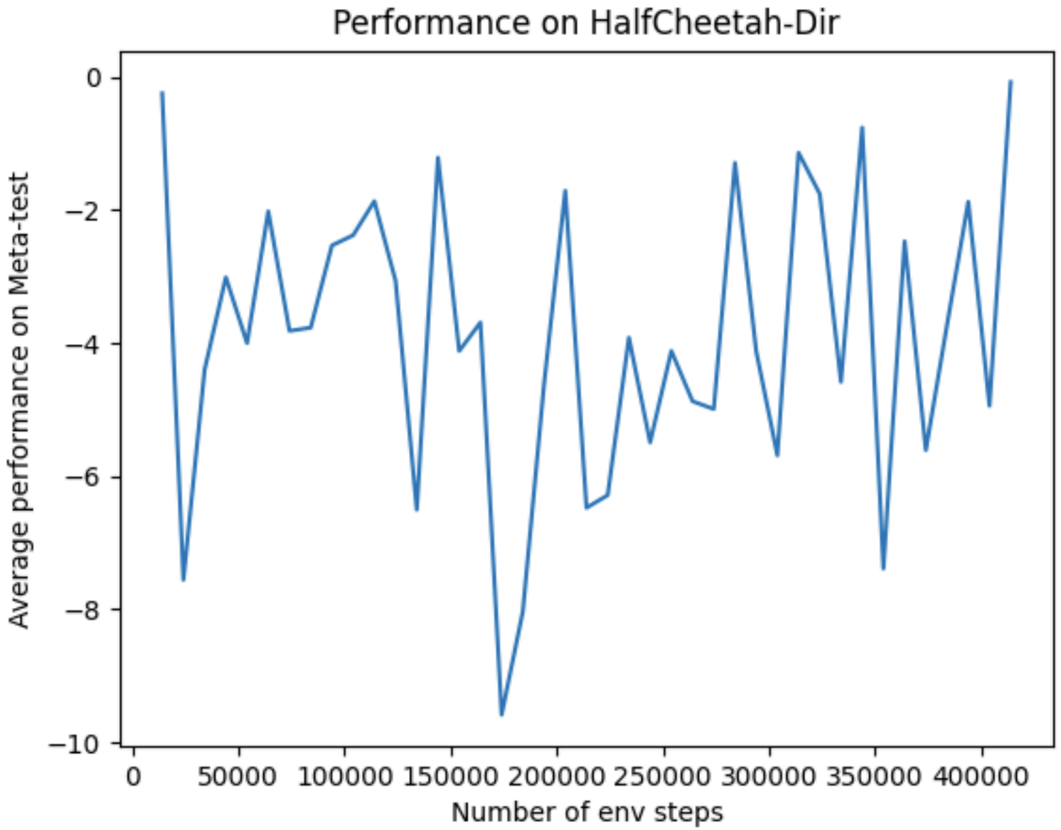
Augment CNN loss

CNN loss = image classification loss + RL reward loss

CNN + PEARL

1. Train CNN model ~~for understanding image to state classification~~ with **added signal form RL reward.**
2. Train RL ~~using state vectors~~ **CNN predicted vectors**.
3. During test use CNN predicted vectors.

Experiments



Thank you