Background

Representation learning for RL

Encoder $z_t = e_\theta(o_t)$

Dynamics $\hat{z}_{t+1} = z_t + d_\phi(z_t, a_t)$

Reward $\hat{r}_{t+1} = r_\phi(z_t, a_t)$

Critic $q_t = Q_\psi(z_t, a_t)$

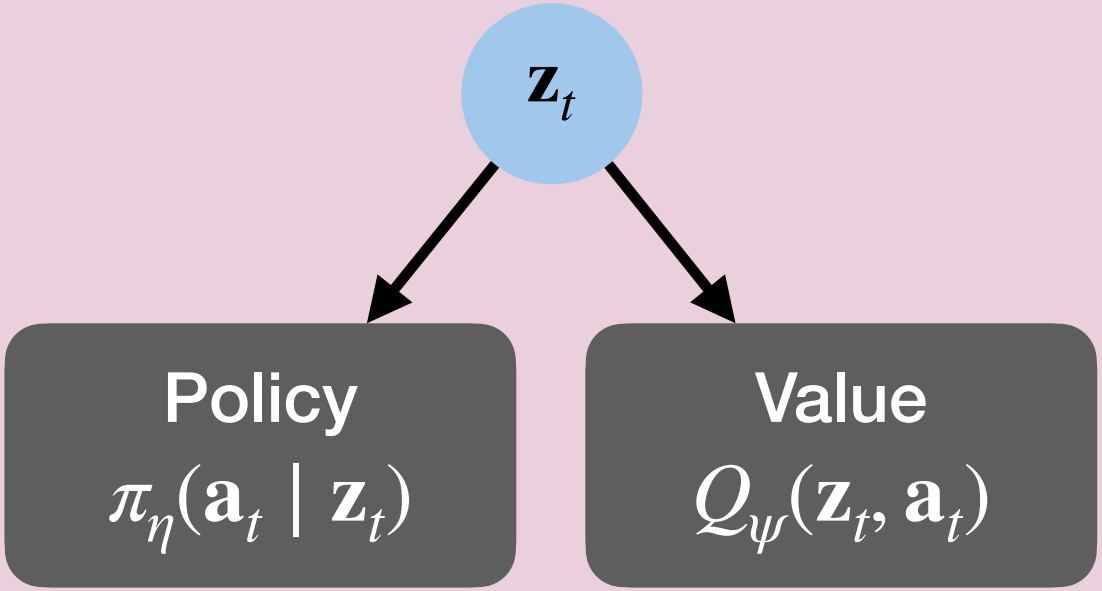
Policy $a_t \sim \pi_\eta(a_t \mid z_t)$

Latent-state consistency loss (representation learning)

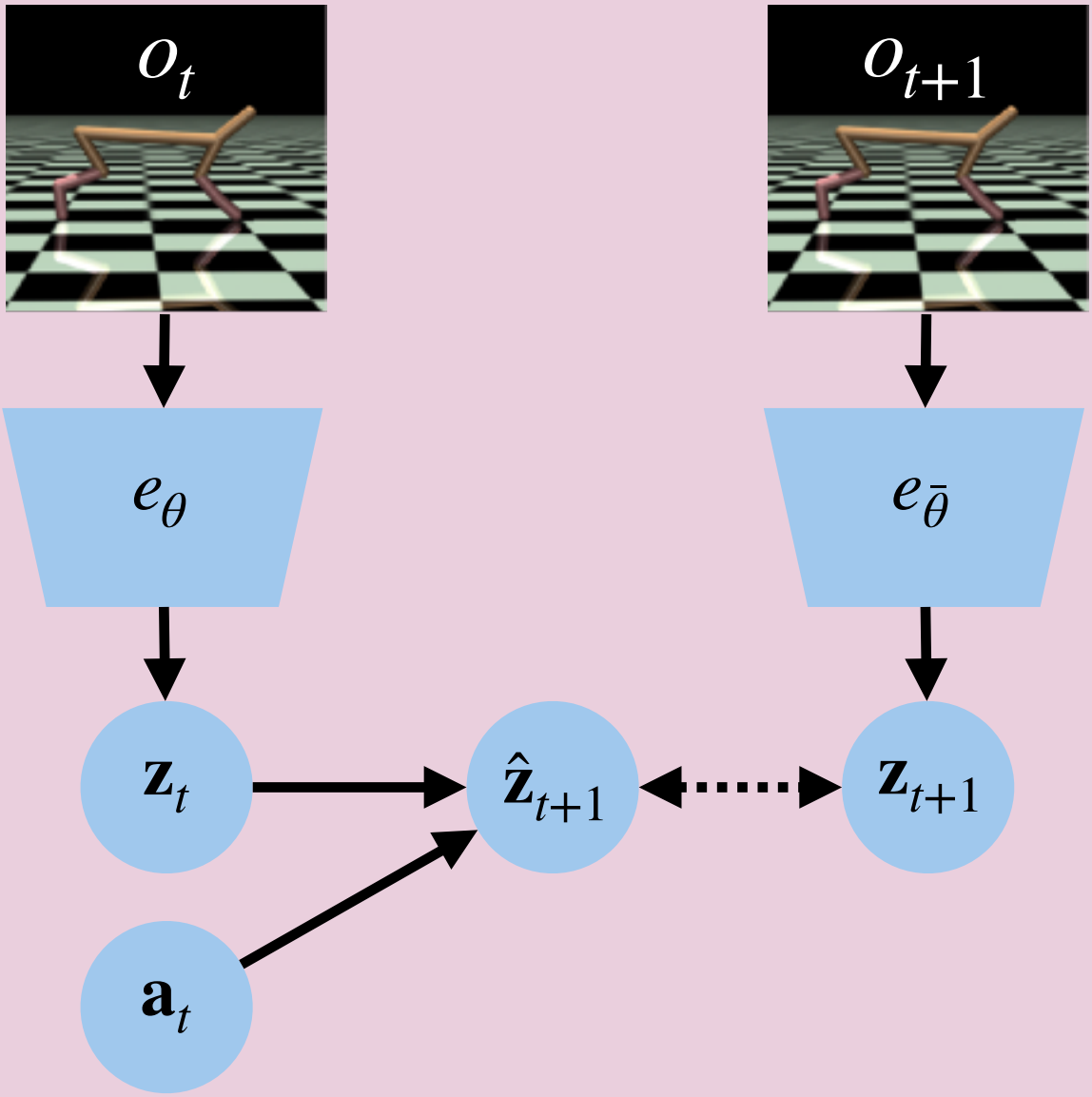
$$\arg\min_{\theta,\phi} \sum_{h=t}^{t+H} \gamma^h \left(\frac{z_h + d_\phi(e_\theta(o_h), a_h)}{\| z_h + d_\phi(e_\theta(o_h), a_h) \|_2} \right)^\top \left(\frac{e_{\bar{\theta}}(o_{h+1})}{e_{\bar{\theta}}(o_{h+1})} \right) + \left\| r_\phi(e_\theta(o_h), a_h) - r_{h+1} \right\|_2^2$$

2

2. Latent actor-critic



1. Learn representation

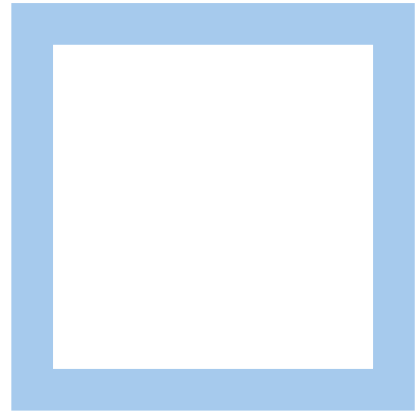


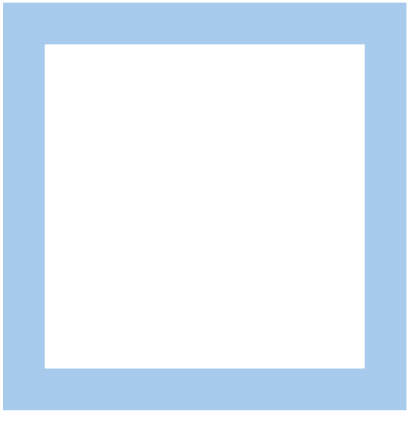
Zhao et al. (2023). Simplified Temporal Consistency Reinforcement Learning. ICML.

Representation is task specific!

Latent-state consistency with cosine similarity

EMA





Background

Representation learning for RL

Encoder $z_t = e_\theta(o_t)$

Dynamics $\hat{z}_{t+1} = z_t + d_\phi(z_t, a_t)$

Reward $\hat{r}_{t+1} = r_\phi(z_t, a_t)$

Critic $q_t = Q_\psi(z_t, a_t)$

Policy $a_t \sim \pi_\eta(a_t \mid z_t)$

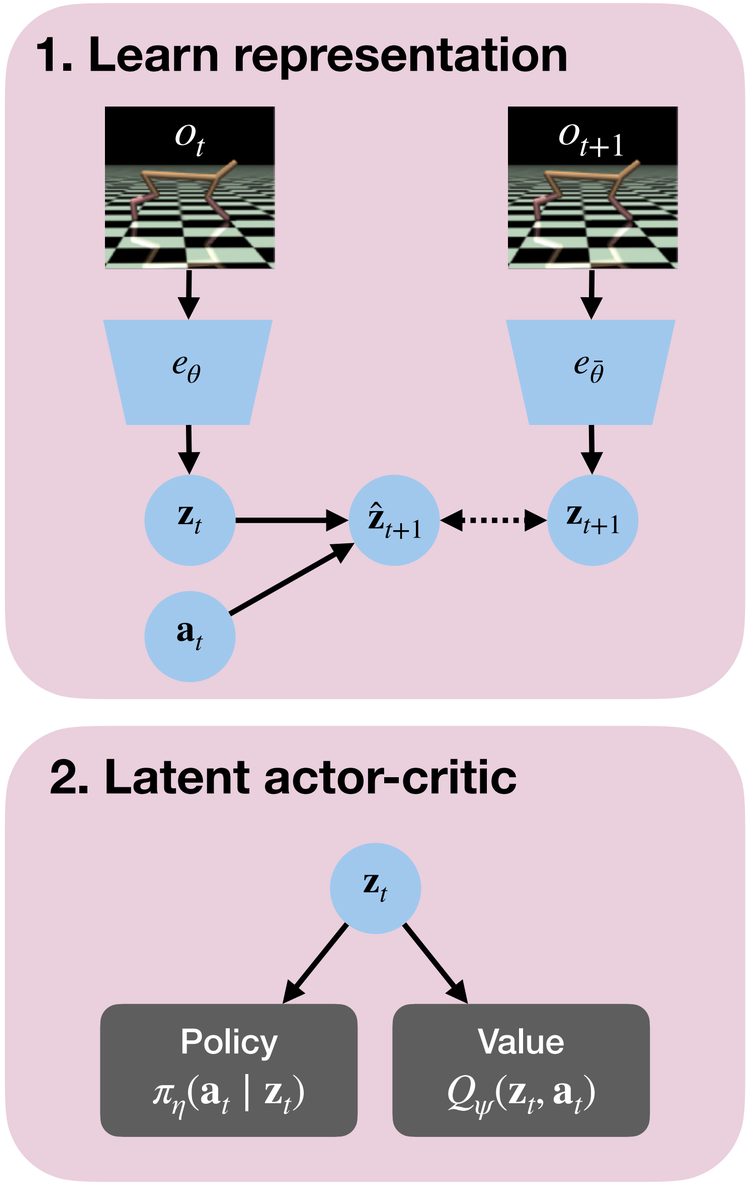
Latent-state consistency loss (representation learning)

EMA

$$\arg\min_{\theta,\phi} \sum_{h=t}^{t+H} \gamma^h \left(\frac{z_h + d_\phi(e_\theta(o_h), a_h)}{\|z_h + d_\phi(e_\theta(o_h), a_h)\|_2} \right)^\top \left(\frac{e_{\bar{\theta}}(o_{h+1})}{e_{\bar{\theta}}(o_{h+1})} \right) + \left\| r_\phi(e_\theta(o_h), a_h) - r_{h+1} \right\|_2^2$$

Representation is task specific!

Latent-state consistency with cosine similarity



FCAI

Zhao et al. (2023). Simplified Temporal Consistency Reinforcement Learning. ICML.

fcai.fi

Background

Task-agnostic representations for RL

Temporal Consistency RL (TCRL)

$$\arg\min_{\theta,\phi} \sum_{h=t}^{t+H} \gamma^h \left(\frac{z_h + d_\phi(e_\theta(o_h), a_h)}{\|z_h + d_\phi(e_\theta(o_h), a_h)\|_2} \right)^\top \left(\frac{e_{\bar{\theta}}(o_{h+1})}{\|e_{\bar{\theta}}(o_{h+1})\|_2} \right) + \boxed{\left\| r_\phi(e_\theta(o_h), a_h) - r_{h+1} \right\|_2^2}$$

Representation is task specific!