Unsupervised Domain Adaptation with Shared Latent Dynamics for Reinforcement Learning

p(B|A)yesgroup.ru

SAMSUNG AI Center - Moscow

Evgenii Nikishin Arsenii Ashukha Dmitry Vetrov

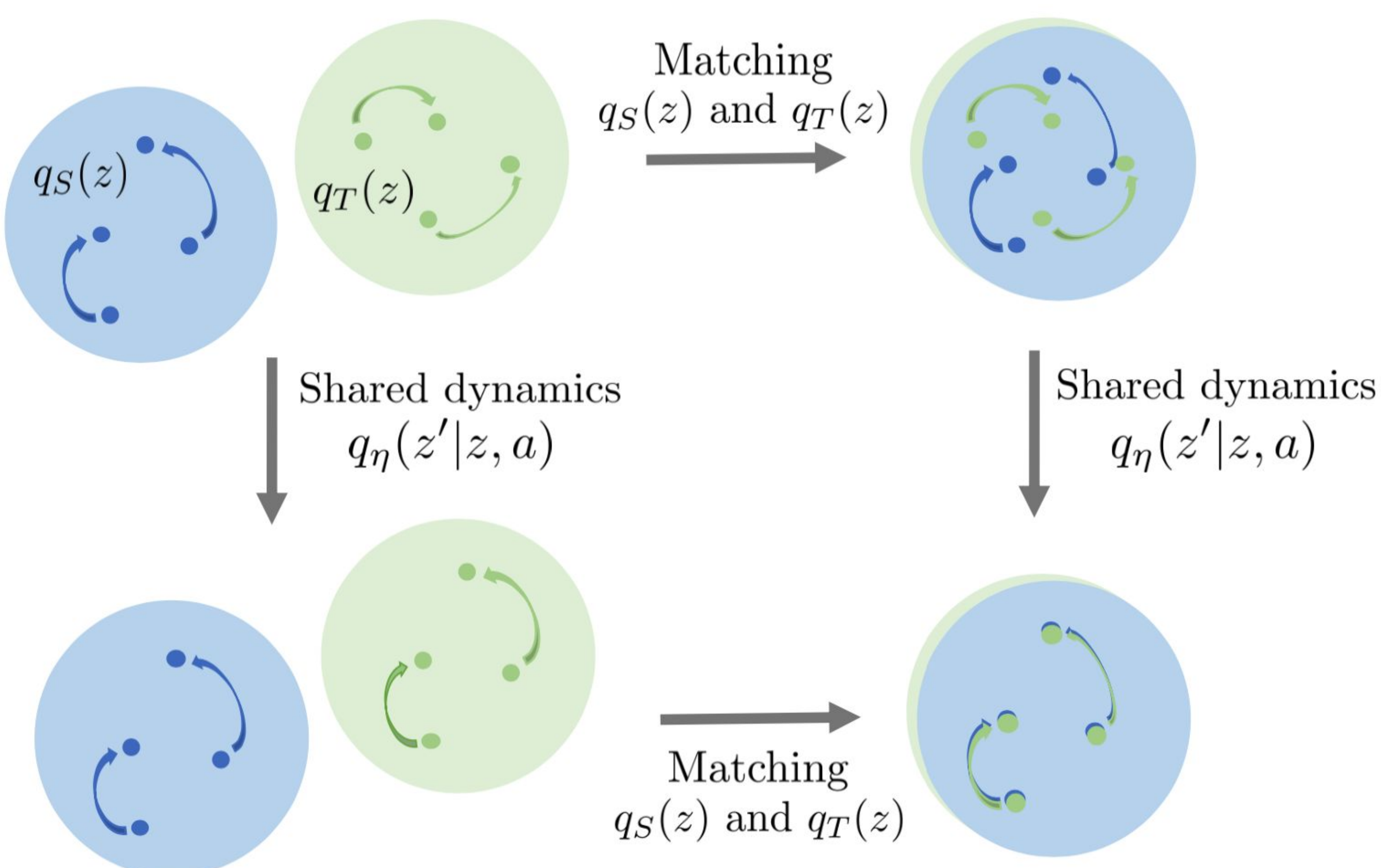
Paper cutt.ly/DynAAE_p

Code cutt.ly/DynAAE_c

Main points

- Given source and target environments that have similar underlying dynamics
- The goal is to adapt a policy trained on the source
- No access to 1-to-1 correspondence between observations

Alignment of latent codes: shared dynamics and adversarial loss



Training pipeline

1. Learning encoding and dynamics on source

$$\mathbb{E}_{q_{\phi^S}(z|x)}\left[\log p_{\theta^S}(x|z) + \log D_S(z) + \mathbb{E}_{q_\eta(z'|z,a)} \log p_{\theta'^S}(x'|z')\right] \to \max_{\phi^S,\theta^S,\theta'^S,\eta}$$

VAE reconstruction; Matching to prior; Next states through dynamics

2. Learning policy on top of latent codes

$$\mathbb{E}_{q_{\phi^S}(z|x)} \log \pi_\xi(a|z) \to \max_\xi$$

Behavior cloning

3. Align representations of a new environment

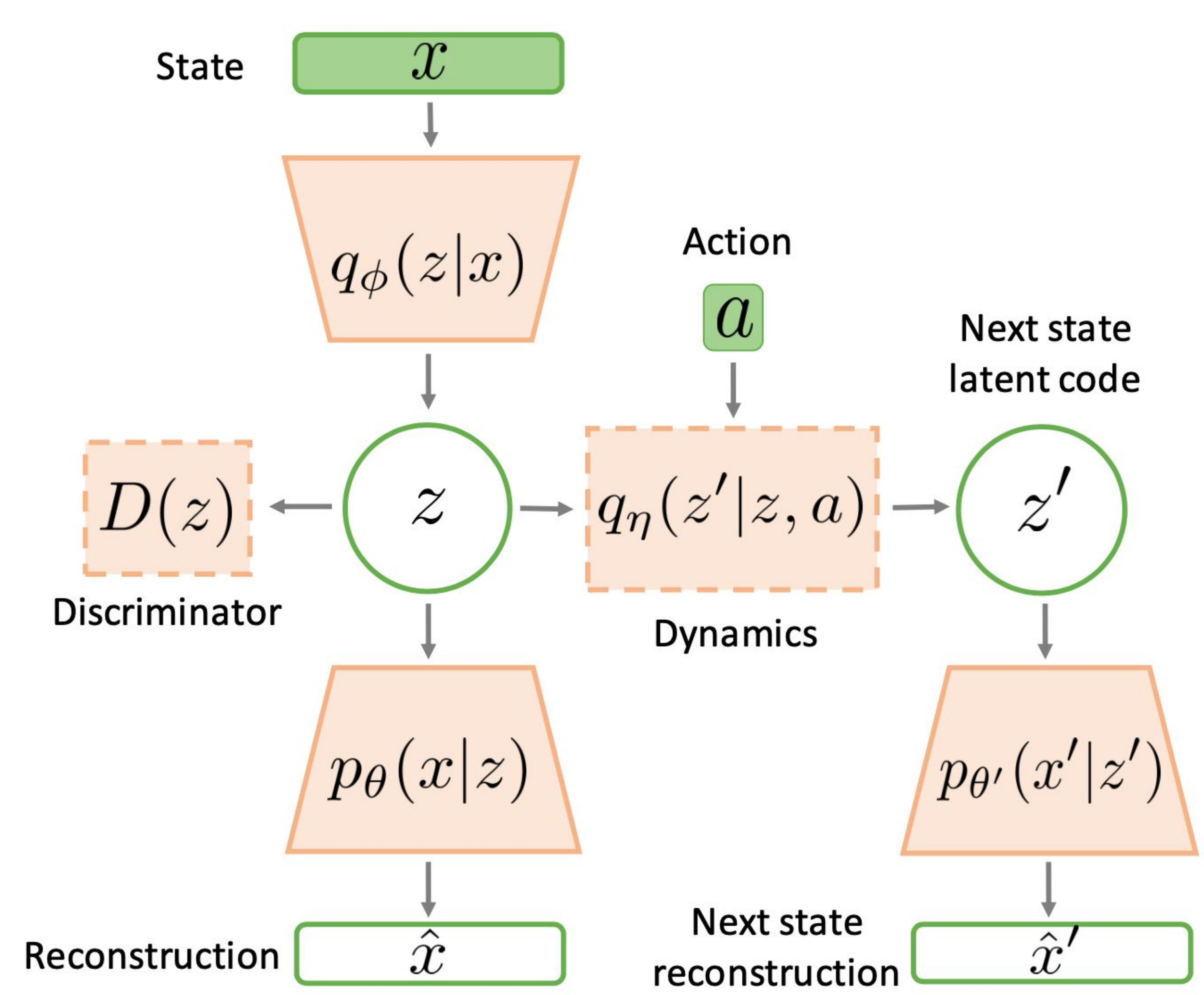
$$\mathbb{E}_{q_{\phi^T}(z|x)}\left[\log p_{\theta^T}(x|z) + \log D_T(z) + \mathbb{E}_{q_\eta(z'|z,a)} \log p_{\theta'^T}(x'|z')\right] \to \max_{\phi^T,\theta^T,\theta'^T}$$

VAE reconstruction; $q_S(z)$ and $q_T(z)$ matching; Next states, dynamics is fixed

Method Summary

- An autoencoding architecture for domain adaptation in RL
- Alignment of latent representations of states via
  - learning shared dynamics
  - matching aggregated posteriors of latent codes
- A policy trained on the latent representations acts optimally for a target environment

The proposed architecture



Experiments on a toy environment

x — MNIST $\quad a \in \{-90°, 0°, +90°\}$

$r = +1$ for a correct rotation

Target environment — digits with inverted colors

Ablation of different model parts

	Const	VAE	Adversarial	Dynamics	Model
Reward	0.40 ± 0	0.40 ± 0.03	0.45 ± 0.06	0.54 ± 0.07	0.81 ± 0.21

Cross-reconstructions for trained source and target models

observations x

$x \to q_{\phi^S}(z|x) \to p_{\theta^S}(x|z)$

$x \to q_{\phi^S}(z|x) \to q_\eta(z'|z,a) \to p_{\theta^S}(x'|z')$

$x \to q_{\phi^S}(z|x) \to p_{\theta^T}(x|z)$

$x \to q_{\phi^S}(z|x) \to q_\eta(z'|z,a) \to p_{\theta^T}(x'|z')$