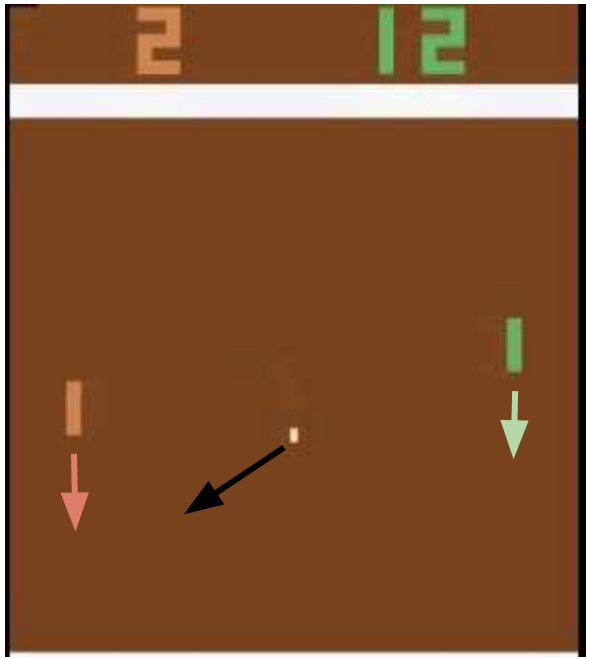
DeepMDP

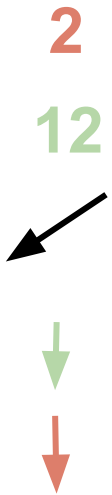
Learning Latent Space Continuous Models for Representation Learning

Carles Gelada, Saurabh Kumar, Jacob Buckman,
Ofir Nachum, Marc G. Bellemare

Simple Representations for RL



≈



DeepMDP

Latent Space Model:

$$\bar{\mathcal{M}} = \langle \bar{\mathcal{S}}, \mathcal{A}, \bar{\mathcal{R}}, \bar{\mathcal{P}}, \gamma \rangle$$

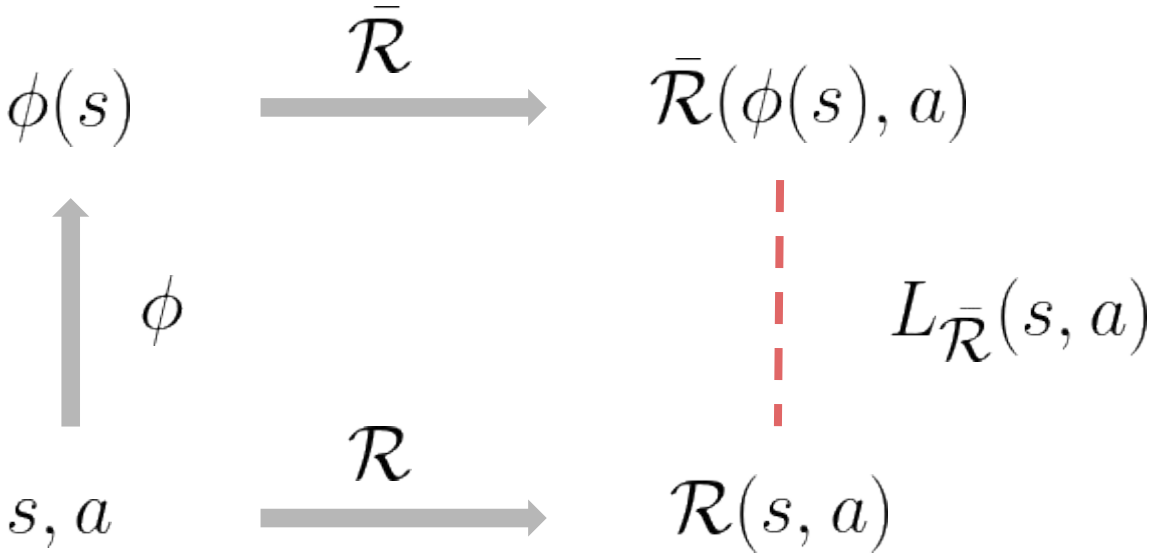
$\phi \longrightarrow$ Neural networks

MDP:

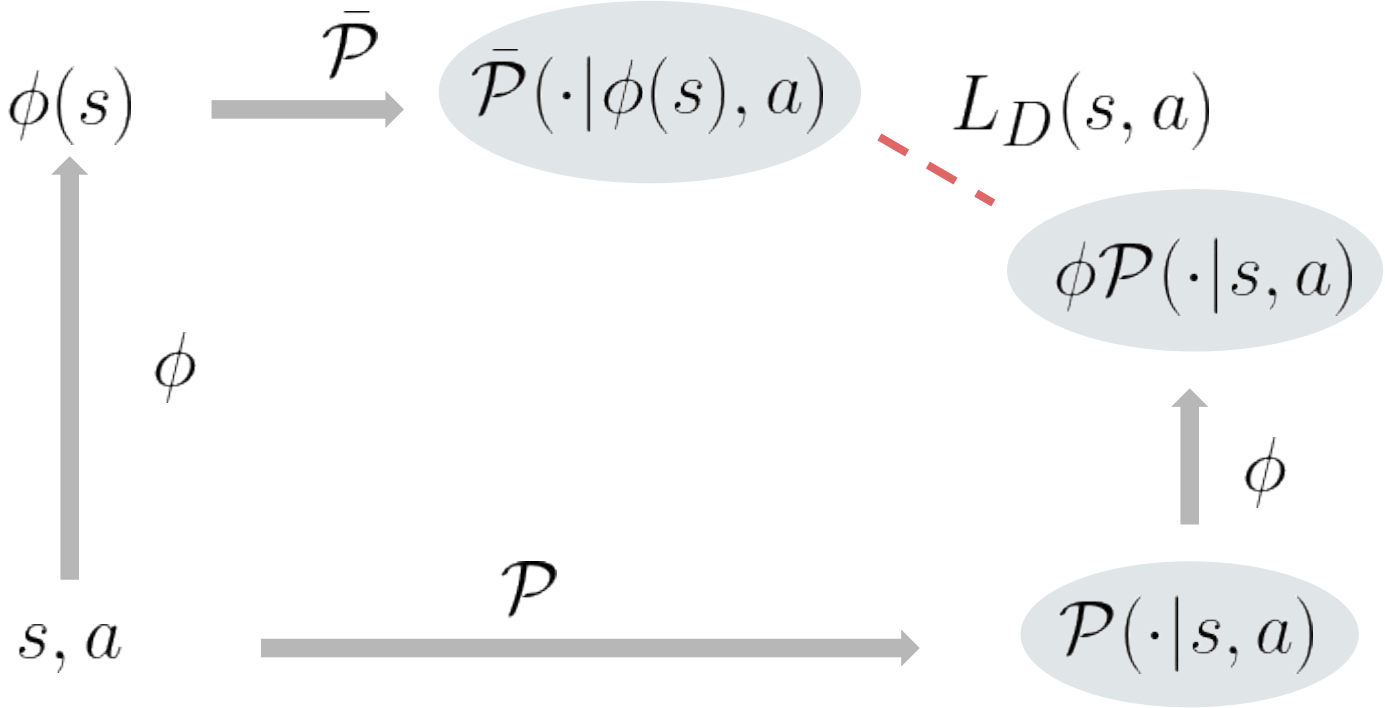
$$\mathcal{M} = < \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma >$$

& trained via the following two losses:

Reward Loss


$\bar{\mathcal{R}}$
$\phi(s)$
$\bar{\mathcal{R}}(\phi(s), a)$
ϕ
$L_{\bar{\mathcal{R}}}(s, a)$
$\mathcal{R}$
s, a
$\mathcal{R}(s, a)$

Transition Loss



Tractable Losses

$$L^{\xi}_{\bar{\mathcal{R}}} = \mathop{\mathbb{E}}_{s,a\sim\xi} |\mathcal{R}(s,a) - \bar{\mathcal{R}}(\phi(s),a)|$$

$$L^{\xi}_{\mathcal{D}} = \mathop{\mathbb{E}}_{s,a\sim\xi} \left[\mathcal{D}(\phi\mathcal{P}(\cdot|s,a), \bar{\mathcal{P}}(\cdot|\phi(s),a))\right]$$

Deep Policies

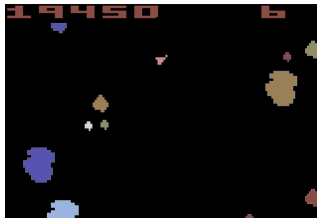
$$\bar{\Pi} \subset \Pi$$

$$\bar{\pi}(a|s) := \bar{\pi}(a|\phi(s))$$

Representation ϕ Quality

$\bar{\Pi}$ Only Discards:

$\pi($  $) =$ ⬇

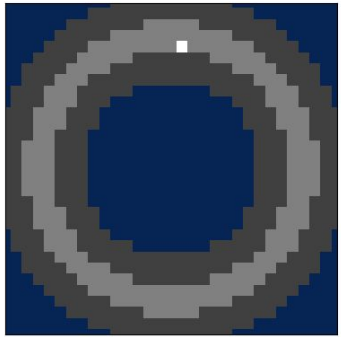
$\pi($  $) =$ ⬆

Ferns, N., Panangaden, P., and Precup, D. Metrics for Finite Markov Decision Processes. In Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI '04, pp. 162–169, 2004.

Phi as a Representation

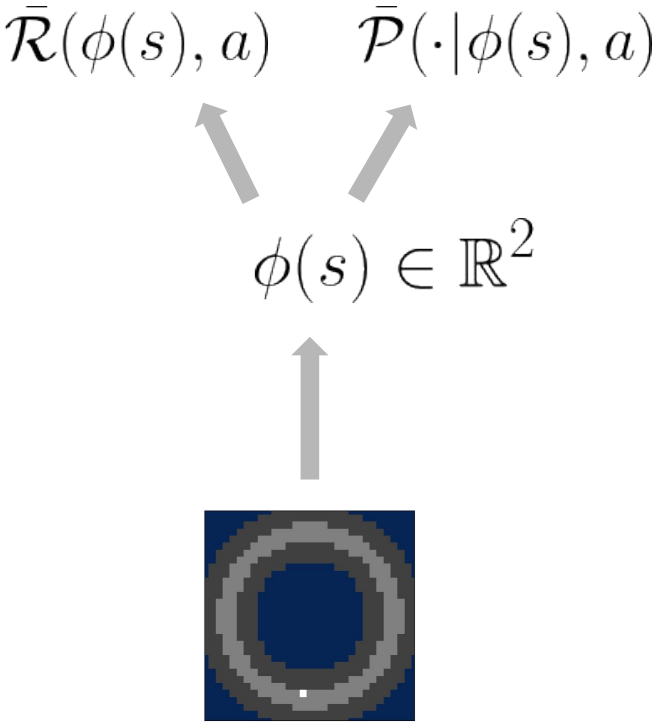
$$\mathop{\mathbb{E}}_{s,a\sim\xi_{\bar{\pi}}} \left|Q^{\bar{\pi}}(s,a) - \bar{Q}^{\bar{\pi}}(\phi(s),a)\right| \leqslant \frac{\left(L^{\xi_\pi}_{\bar{\mathcal{R}}} + \gamma K_{\bar{V}} L^{\xi_\pi}_{D}\right)}{1-\gamma}$$

Donut World



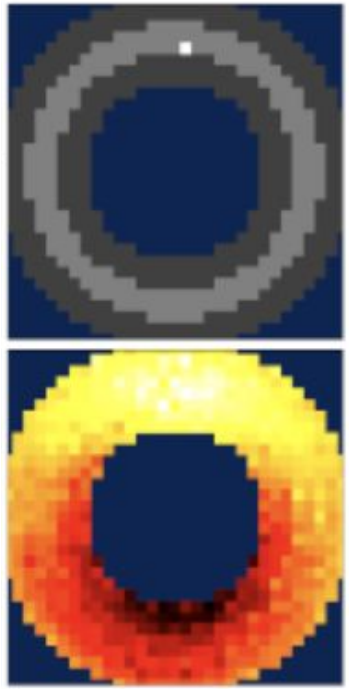
DeepMDP on Donut World

2D latent space
+
DeepMDP losses



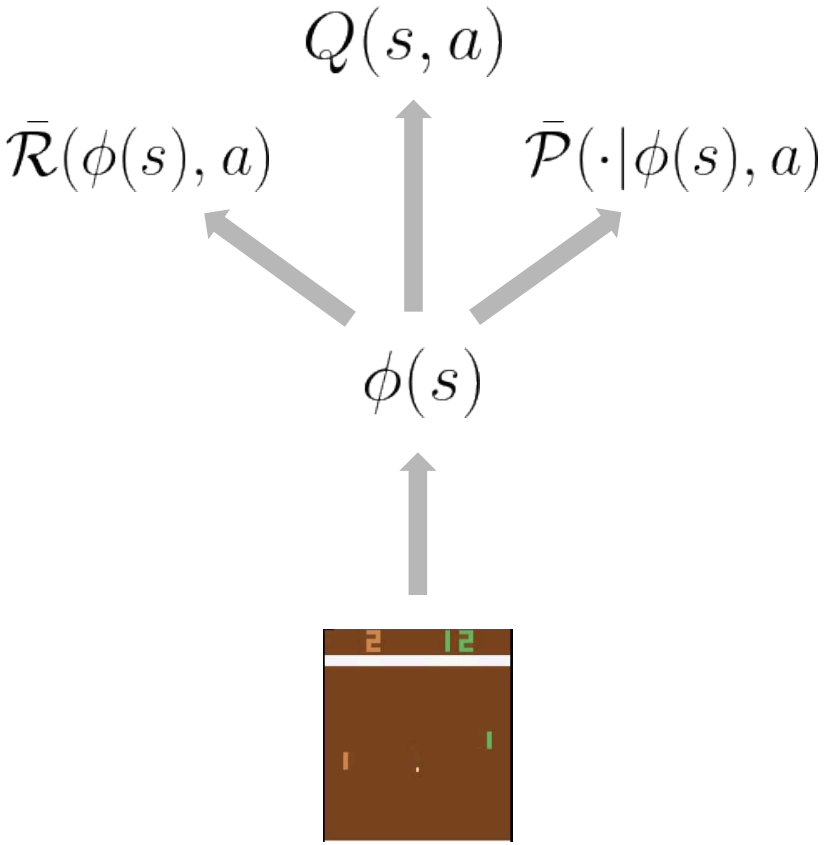
DeepMDP on Donut World

Visualization of latent distance



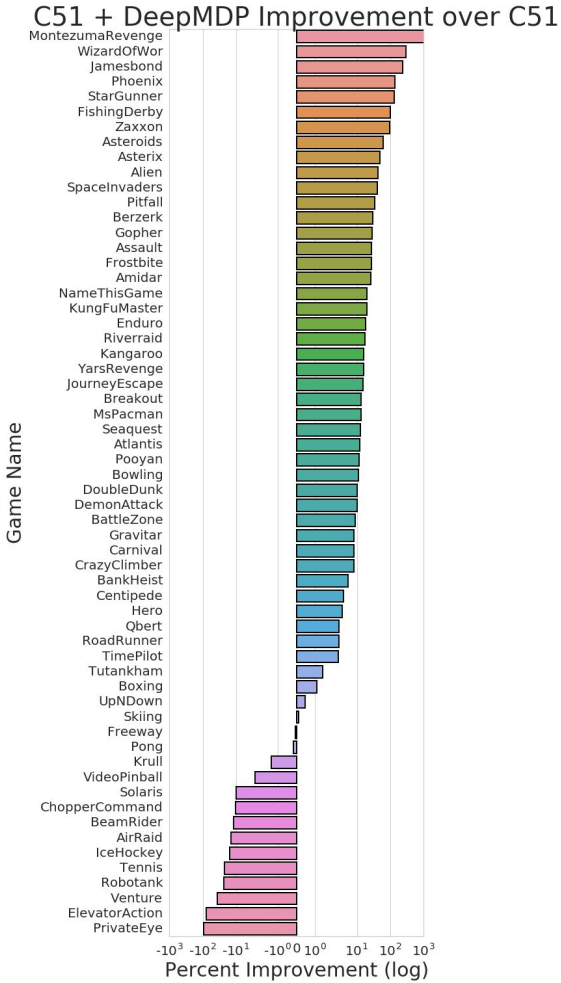
DeepMDP Auxiliary Task

Base C51 agent
+
DeepMDP losses



DeepMDP Auxiliary Task

Base C51 agent
+
DeepMDP losses



- **DeepMDPs** as Models of the Environment
- **Norm-MMD Metrics** and their Associated Smoothness

Thanks For Listening

Poster #108