Focus On What Matters: Separated Models For Visual-Based RL Generalization

Di Zhang Bowen Lv Hai Zhang Feifan Yang Junqiao Zhao* Hang Yu Chang Huang Hongtu Zhou Chen Ye Changjun Jiang

Department of Computer Science, Tongji University, Shanghai, China MOE Key Lab of Embedded System and Service Computing, Tongji University, Shanghai, China

NEURAL INFORMATION PROCESSING SYSTEMS

What Matters in a Reinforcement Learning Task?

Deploying visual-based RL algorithms in real-world applications requires a high generalization ability due to numerous factors that can induce distribution shifts between training and deployment scenarios, such as variations in lighting conditions, camera viewpoints, and backgrounds.

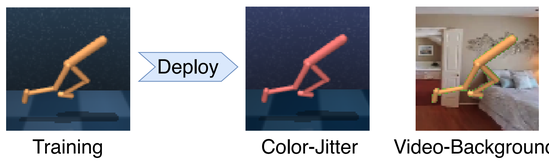


Figure 1. A robotic manipulation task explanation for task-relevant parts in the environment.

In contrast, humans can accurately figure out what matters visually when transferring to a new environment. Considering a robotic manipulation task where the agent must move the arm to the red target, despite variations in background colors and textures across four test scenarios on the left, only the arm's orientation and the target position should be focused on this task. We aim for our RL agent to learn an optimal policy that solely relies on these task-relevant features while disregarding irrelevant regions.

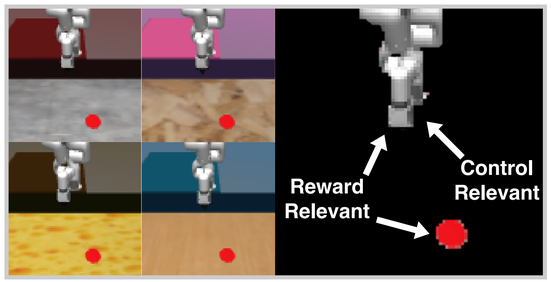


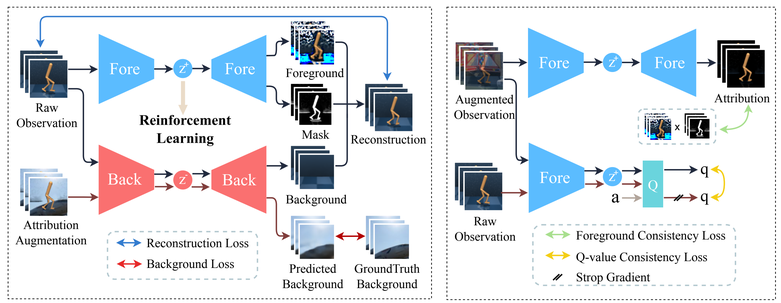
Figure 2. A robotic manipulation task explanation for task-relevant parts in the environment.

Contribution

Can we apply the way humans handle generalization problems to RL agents?

- We present SMG (Separated Models for Generalization), a novel approach that aims to enhance the zero-shot generalization ability of RL agents. SMG is designed as a plug-and-play method that seamlessly integrates with existing standard off-policy RL algorithms.
- SMG emphasizes the significance of task-relevant features in visual-based RL generalization and successfully incorporates a reconstruction loss into this setting.
- Extensive experimental results demonstrate that SMG achieves state-of-the-art performance across various visual-based RL tasks, particularly excelling in video-background settings and robotic manipulation tasks.

Learning Task-Relevant Representations with Separated Models



(a) Learning Task-Relevant Representations With SMG (b) Improving GeneralizationWith SMG

Figure 3. Architecture of SMG. One-way arrows represent different types of data flows with the same input. Two-way arrows represent different types of loss.

We introduce a separated models architecture, which reconstructs the observation via $o'_t = o^+_t \odot M_t + o^-_t \odot (1 - M_t)$. This approach allows us to extract task-relevant and task-irrelevant representations in a self-supervised manner.

Building on this, we define additional loss terms to enhance the model's ability to distinguish between two types of representations:

- Mask ratio loss.

$$L_{mask} = (\frac{\sum_{i,j} M_t(i,j)}{\text{image\_size}^2} - \rho)^2 \tag{1}$$

- Background reconstruction loss.

$$L_{\text{back}} = -\mathbb{E}_{o_t \sim \mathcal{D}}[\mathbb{E}_{z_t^- \sim f^-(\tau_{\text{attrib}}(o_t))}[\log g^-(\epsilon | z_t^-)]] \tag{2}$$

- Empowerment loss.

$$L_{\text{action}} = -I(a_t, z^+_{t+1} | z^+_t) \leq -\mathbb{E}_{p(a_t, z^+_{t+1}, z^+_t)}[\log q(a_t | z^+_{t+1}, z^+_t)] \tag{3}$$

Generalize Task-Relevant Representations with Separated Models

Since the agent still lacks the ability to generalize effectively and may struggle to extract meaningful features from scenarios with transformed styles, we treat the task-relevant representation under raw observations as the ground truth and train SMG on more diversely augmented samples:

- Foreground consistency loss.

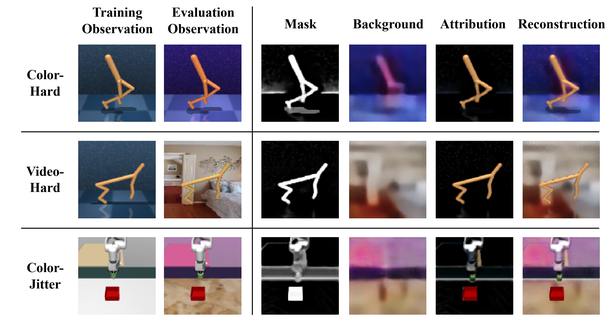
$$L_{\text{fore\_consist}} = \mathbb{E}_{o_t \sim \mathcal{D}}[|Attrib(\tau(o_t)) - \text{sg}(Attrib(o_t))|] \tag{4}$$

- Q-value consistency loss.

$$L_{\text{q\_consist}} = \mathbb{E}_{o_t, a_t \sim \mathcal{D}}[[Q(f^+(\tau(o_t)), a_t) - \text{sg}(Q(f^+(o_t), a_t))]^2] \tag{5}$$

Experiments

- SMG successfully extracts different types of features underlying in the observation.



- SMG successfully tackles the most challenging video-hard setting in DMControl.

DMControl (*video-hard*)	SAC	DrQ	SODA	SVEA (overlay)	SRM	SGQN	SMG (ours)	Δ
`cartpole, swingup`	156 ±16	168 ±35	346 ±59	510 ±177	254 ±69	599 ±112	764 ±32	+165 28%
`finger, spin`	22 ±10	54 ±44	310 ±72	353 ±71	131 ±89	710 ±159	910 ±61	+200 28%
`walker, stand`	212 ±41	278 ±79	406 ±68	814 ±57	558 ±139	870 ±78	955 ±9	+85 10%
`walker, walk`	132 ±26	110 ±33	175 ±31	348 ±80	165 ±99	634 ±136	814 ±51	+180 28%
`cheetah, run`	56 ±30	38 ±26	118 ±40	105 ±13	87 ±24	135 ±44	303 ±46	+168 124%

- SMG demonstrates superior stability in robotic manipulation tasks.

Robtic-Manipulation (*peg-in-box*)	SAC	DrQ	SODA	SVEA (overlay)	SRM	SGQN	SMG (ours)	Δ
`train`	31 ±73	233 ±14	232 ±20	212 ±39	227 ±15	232 ±19	237 ±16	+4 2%
`test1`	−33 ±25	63 ±99	34 ±143	−18 ±59	55 ±98	−67 ±28	237 ±18	+174 276%
`test2`	−42 ±31	−40 ±77	76 ±119	85 ±68	11 ±54	194 ±51	219 ±37	+25 13%
`test3`	−8 ±46	15 ±107	66 ±147	67 ±73	147 ±114	198 ±34	237 ±15	+39 20%
`test4`	−42 ±51	72 ±28	80 ±122	109 ±98	112 ±123	−51 ±46	237 ±17	+125 112%
`test5`	−52 ±31	−54 ±30	−104 ±51	−26 ±102	143 ±122	−108 ±24	237 ±15	+94 66%