Yuxi HAN, Dequan LI, Yang YANG, 2025. Significance extraction based on data augmentation for reinforcement learning. *Frontiers of Information Technology & Electronic Engineering*, 26(3):385-399.
http://doi.org/10.1631/FITEE.2400406

Significance extraction based on data augmentation for reinforcement learning

Front Inform Technol Electron Eng

Key words: Deep reinforcement learning; Visual tasks; Generalization; Data augmentation; Significance; DeepMind Control generalization benchmark

Corresponding author: Dequan LI
E-mail: leedqcpp@126.com
ORCID: https://orcid.org/0000-0002-4329-864X

Motivation

- Sample inefficiency and poor generalization are common issues in reinforcement learning. Researchers have attempted to enhance generalization capabilities through data augmentation. However, existing methods suffer from the following problem: agents cannot effectively distinguish between noise signals and important pixels in images, causing their attention to be diverted to non-critical features.

Front Inform Technol Electron Eng

Main idea

- To address the issue of agents' inability to effectively distinguish between noise signals and important pixels in images, we propose using attention mechanisms to identify and extract key features, and employ a mask decoder network to map these features onto the input states as masks.
- To address the issue of insufficient training data and prevent suboptimal exploration, we propose introducing noise perturbations into both observation states and state rewards.
- To enhance the stability of Q values, reduce estimation bias, and improve generalization capability, we propose applying consistency regularization to both the augmented images and the original state images.

Front Inform Technol Electron Eng

Method

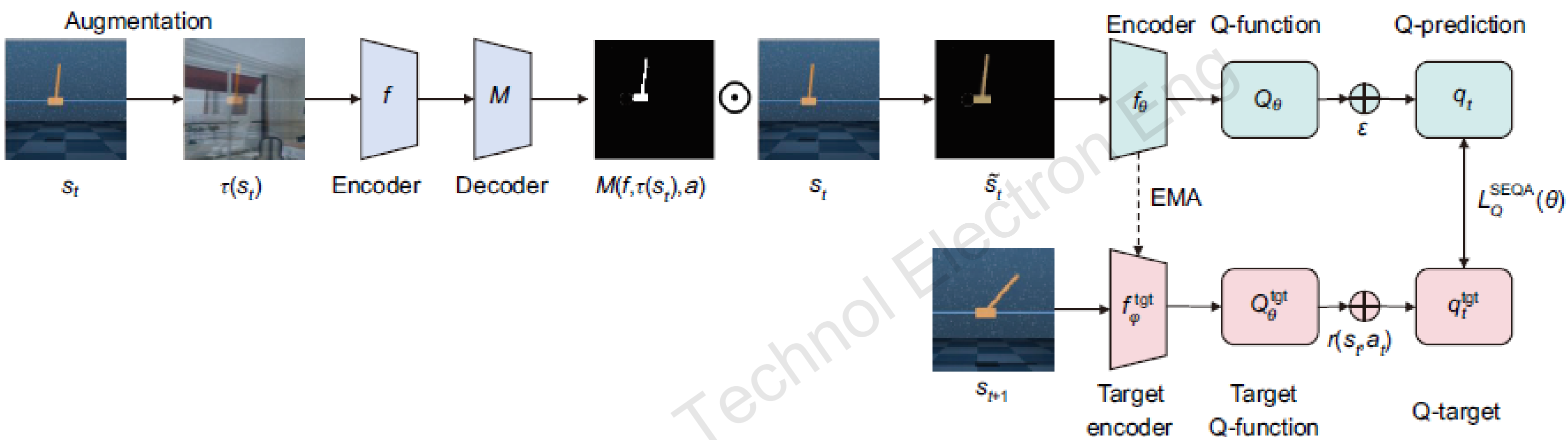


Fig. 1 An overview of the saliency-extracted Q-value by augmentation (SEQA) framework. The augmented data $\tau(s_t)$ are encoded and decoded to generate $M(f, \tau(s_t), a)$, and the Q-functions of the augmented state $\tilde{s}_t$ and the unaugmented state s_{t+1} are regularized consistently to obtain $L_Q^{\mathrm{SEQA}}(\theta)$. Here, $\odot$ represents the Hadamard product and $\oplus$ represents addition to Q. EMA: exponential moving average

Front Inform Technol Electron Eng

$$L_Q^{SEQA}(\theta) = \frac{1}{KN}\sum_{k=1}^{K}\sum_{t=1}^{N}[||Q_\theta(f_\theta(s \odot M(f, \boldsymbol{\tau}(s_t), a_t)), a_k) + \varepsilon - q_t^{tgt}||^2]$$

Method

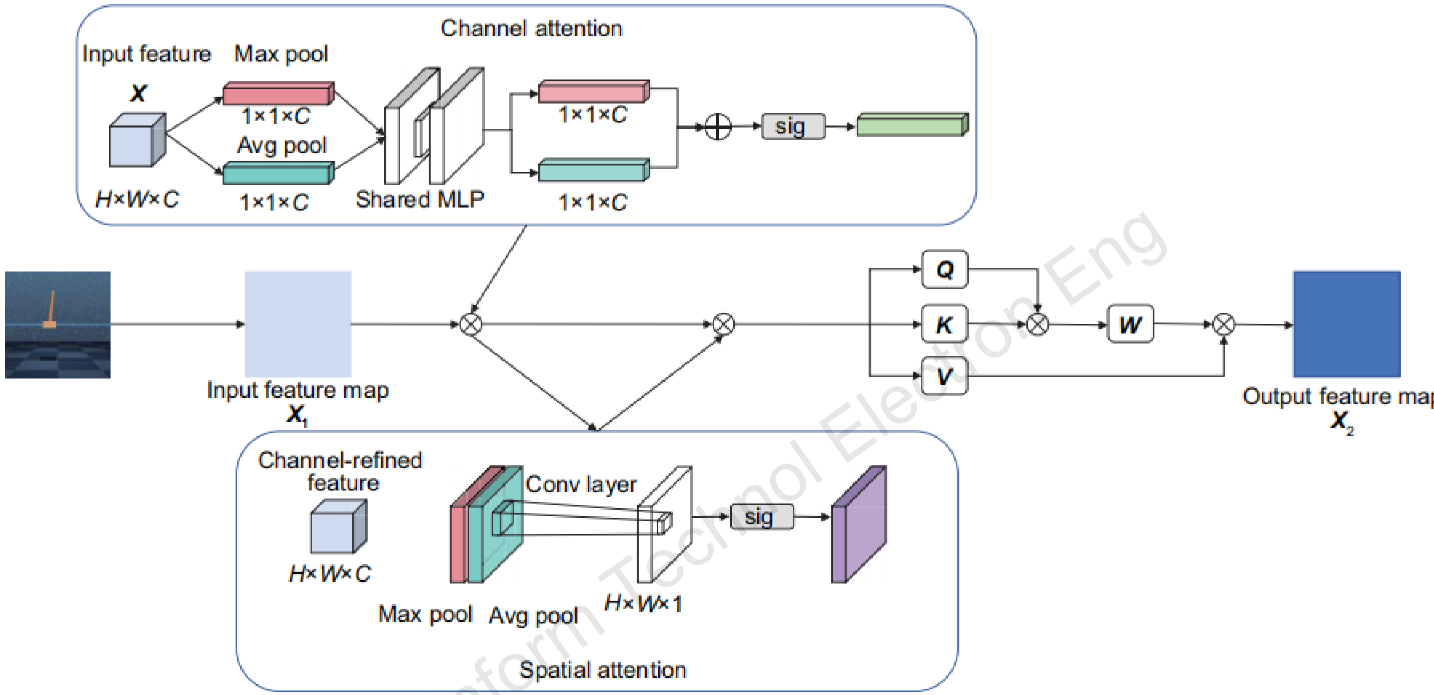


Fig. 2 Attention model. The detailed design of the attention mechanism is given, including how channel attention, spatial attention, and mixed attention are implemented

Front Inform Technol Electron Eng

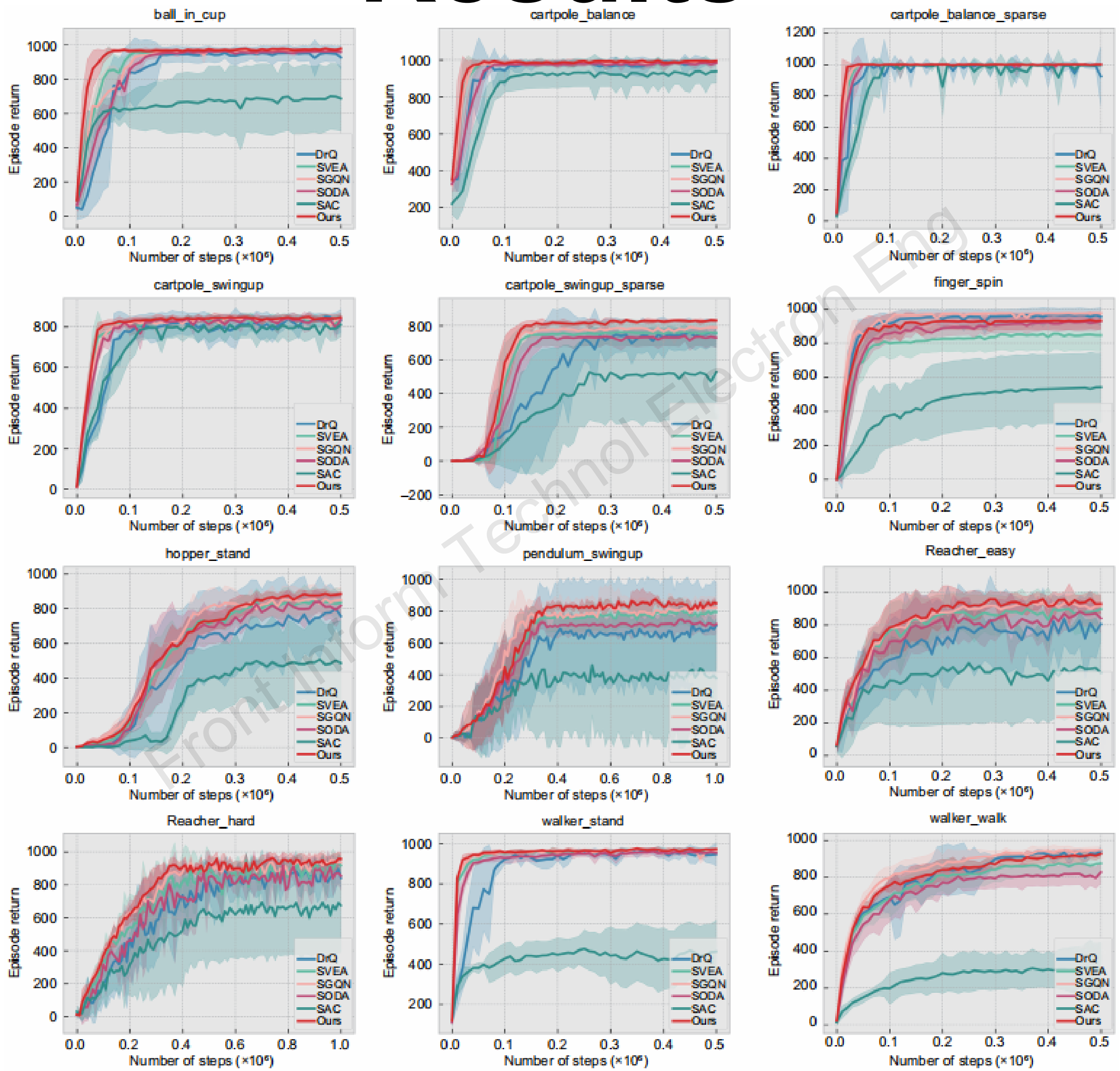
The channel attention weight matrix $\boldsymbol{M}_{\mathrm{c}}$: $\boldsymbol{M}_{\mathrm{c}} = \mathrm{sig}(\mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{X})) + \mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{X})))$

The spatial attention weight matrix $\boldsymbol{M}_{\mathrm{s}}$: $\boldsymbol{M}_{\mathrm{s}} = \mathrm{sig}(f^{7\times 7}([\mathrm{MaxPool}(\boldsymbol{X}); \mathrm{AvgPool}(\boldsymbol{X})]))$

The mixed attention weight matrix $\boldsymbol{M}_{\mathrm{f}}$: $\boldsymbol{M}_{\mathrm{f}} = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V}$

A complete feature map $\boldsymbol{X}_2$: $\boldsymbol{X}_2 = \boldsymbol{M}_{\mathrm{c}} \otimes \boldsymbol{M}_{\mathrm{s}} \otimes \boldsymbol{M}_{\mathrm{f}},$

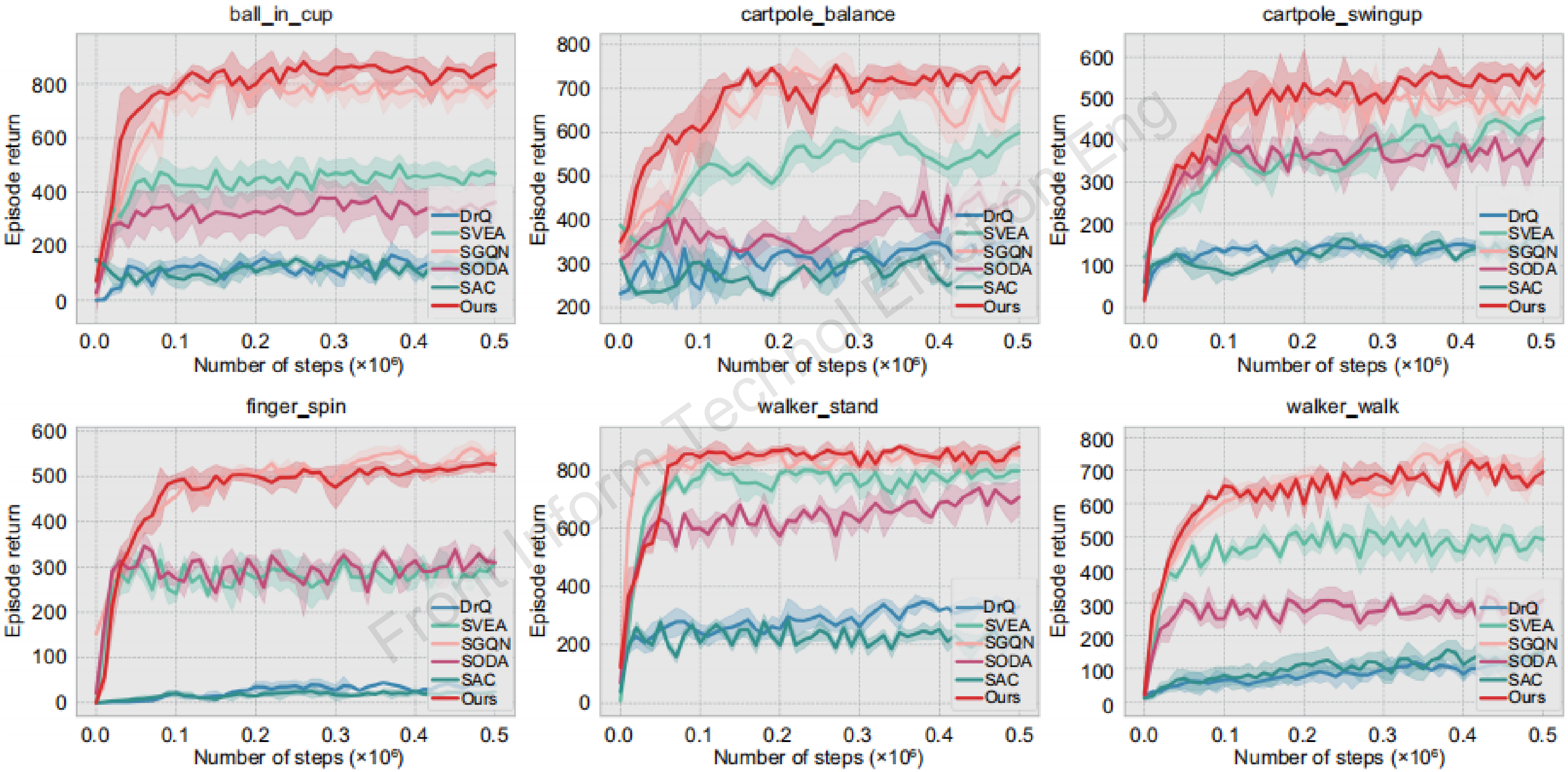
Results



Front Inform Technol Electron Eng

Fig. 4 Performance on training. SEQA achieves the best performance in almost all environments (References to color refer to the online version of this figure)

Results



Front Inform Technol Electron Eng

Fig. 5 Performance on the video_hard benchmark. SEQA performers significantly better than the other algorithms in four environments and slightly worse than the optimal algorithm SGQN in the other two environments. References to color refer to the online version of this figure

Conclusions

In this paper, we propose SEQA, an algorithm that relies primarily on three key components: noise augmentation, mask coverage, and consistency regularization. These components collectively encourage that the agent concentrates on important features during its decision-making process while disregarding distract factors.

Front Inform Technol Electron Eng

Yuxi HAN is currently pursuing the MS degree at Anhui University of Science and Technology. He obtained his BS degree from the same university in 2022. His research focus is on reinforcement learning.

Dequan LI is currently the deputy dean of the Faculty of Artificial Intelligence at Anhui University of Science and Technology. His main research interests include artificial intelligence algorithms, machine vision perception, autonomous intelligent networks, intelligent robot control, cyber-physical systems, and other related fields.

Yang YANG is currently pursuing the MS degree at Anhui University of Science and Technology. She obtained her BS degree from the same university in 2022. Her research focus is on medical image segmentation.

Front Inform Technol Electron Eng