Revisiting Plasticity in Visual RL: Data, Modules and Training Stages

Guozheng Ma[1*], Lu Li[1*], Sen Zhang[2], Zixuan Liu[1], Zhen Wang[2]

Yixin Chen[3], Li Shen[4]✉, Xueqian Wang[1]✉, Dacheng Tao[2]

[1] Tsinghua University, [2] The University of Sydney

[3] Washington University in St. Louis, [4] JD Explore Academy

Revisiting Plasticity in Visual RL:
Data, Modules and Training Stages

Contributions:

- We conduct a systematic exploration focusing on **three primary underexplored facets about the VRL's plasticity** and derive the following insightful conclusions:
  - Data augmentation (DA) is essential in maintaining plasticity **(why DA works?)**
  - **Critic's plasticity loss** is the principal bottleneck impeding efficient training
  - Maintaining plasticity in **early stages** is crucial to prevent irrecoverable loss
- Armed with these insights, we propose ***Adaptive RR*** to address the **high RR dilemma** that has perplexed the VRL community for a long time.

Plasticity (Related Work)

- What does '**plasticity**' in neural networks signify?
- The phenomenon of **plasticity loss in DRL** and its catastrophic consequences.
- How to mitigate the plasticity loss for achieving sample-efficient DRL?

Data

DA notably outperforms Reset and other interventions in preserving plasticity, highlighting the inherent mechanisms driving its effectiveness.

Modules

The primary bottleneck hindering sample-efficient VRL is the **critic's plasticity loss**, rather than the previously widely attributed **poor representation**.

Training Stages

Without timely intervention to recover critic's plasticity in the **early stages**, its loss becomes catastrophic and irrecoverable.

Plasticity (Related Work)

- What does '**plasticity**' in neural networks signify?
- The phenomenon of **plasticity loss in DRL** and its catastrophic consequences.
- How to mitigate the plasticity loss for achieving sample-efficient DRL?

Data

DA notably outperforms Reset and other interventions in preserving plasticity, highlighting the inherent mechanisms driving its effectiveness.

Modules

The primary bottleneck hindering sample-efficient VRL is the **critic's plasticity loss**, rather than the previously widely attributed **poor representation**.

Training Stages

Without timely intervention to recover critic’s plasticity in the **early stages**, its loss becomes catastrophic and irrecoverable.

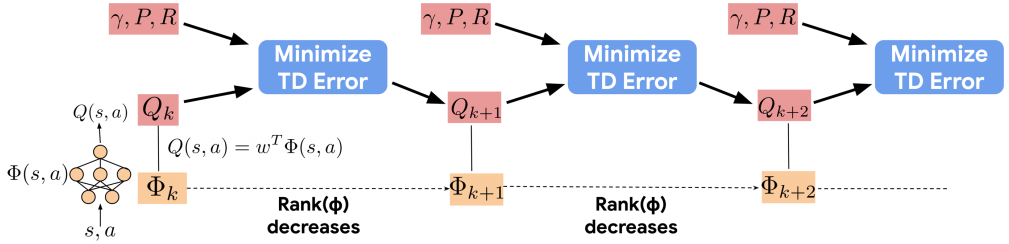
What does 'plasticity' in neural networks signify?

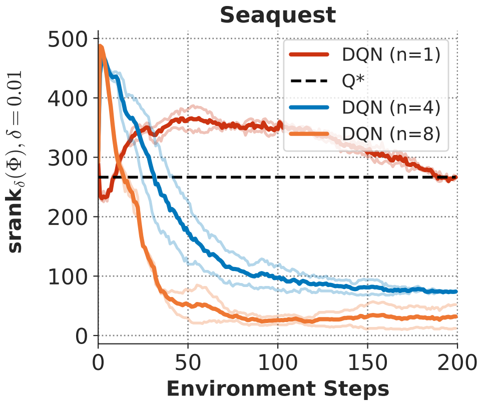
- Plasticity, the ability of a neural network to **quickly change its predictions in response to new information**; broadly construed, it refers to a neural network's ability to **learn new things**.[1]
- "You cannot teach an old dog new tricks" an old proverb says. ... ... DRL agents that use neural networks may gradually lose the ability to **learn from new experiences**. [2]
- **Dormant neuron phenomenon** in DRL: an agent's network suffers from an increasing number of inactive neurons, **thereby affecting network expressivity**.[3]
- ... ... capacity loss, whereby networks trained on a sequence of target values **lose their ability to quickly update their predictions over time**.[4]
- ... ... However, these multiple updates often lead to overfitting, which **decreases the network's ability to adapt to new data**.[5]
- ... ... plasticity refers to the ability of an agent to **quickly adapt to new information**.[6]
- Loss of plasticity is ... ... neural networks **lose their ability to learn from new experience**.[7]

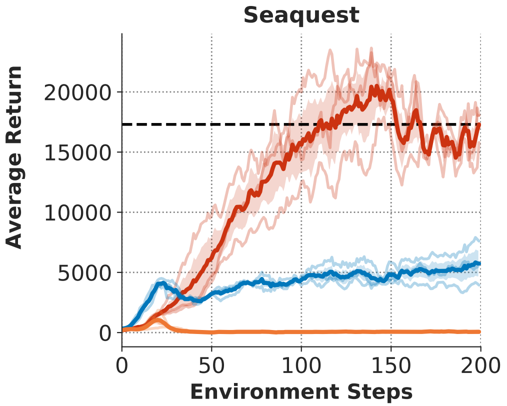
[1] Lyle C, Zheng Z, Nikishin E, et al. Understanding plasticity in neural networks[J]. ICML 2023.
[2] Nikishin E, Oh J, Ostrovski G, et al. Deep Reinforcement Learning with Plasticity Injection[J]. NeurIPS 2023.
[3] Sokar G, Agarwal R, Castro P S, et al. The dormant neuron phenomenon in deep reinforcement learning[J]. ICML 2023
[4] Lyle C, Rowland M, Dabney W. Understanding and preventing capacity loss in reinforcement learning[J]. ICLR 2022
[5] Lee H, Cho H, Kim H, et al. Enhancing Generalization and Plasticity for Sample Efficient Reinforcement Learning[J]. NeurIPS 2023
[6] Kumar S, Marklund H, Van Roy B. Maintaining plasticity via regenerative regularization[J]. arXiv preprint.
[7] Anonymous. Curvature Explains Loss of Plasticity. Submitted to The Twelfth International Conference on Learning Representations.

The phenomenon of plasticity loss in DRL and its catastrophic consequences.

➢ Implicit Under-Parametrization [1]:

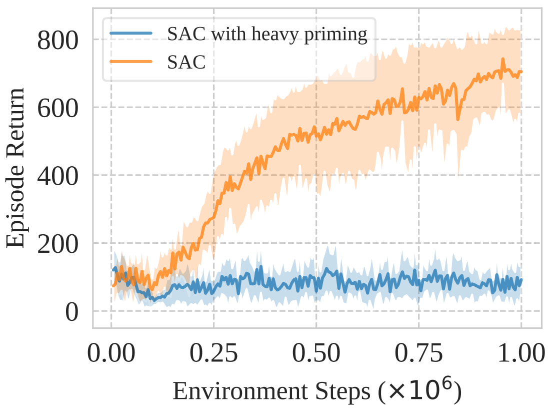




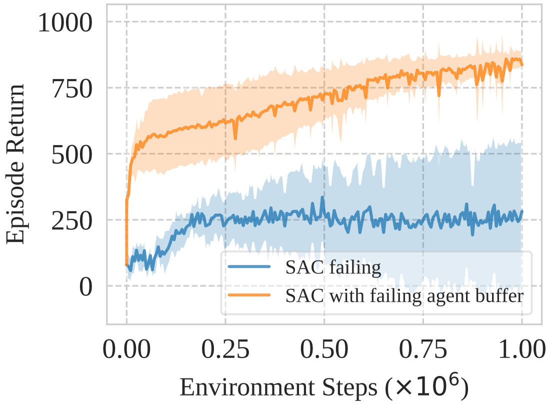


The **excessive aliasing of features** learned by the value network across states causes an otherwise **expressive value network** to **implicitly behave as an under-parameterized network**, often resulting in poor performance.

➢ Primacy Bias[2]:



Undiscounted returns for SAC with and without **heavy priming on the first 100 transitions.**
An agent extremely affected by the **primacy bias** is **unable to learn even after collecting hundreds of thousands of new transitions.**



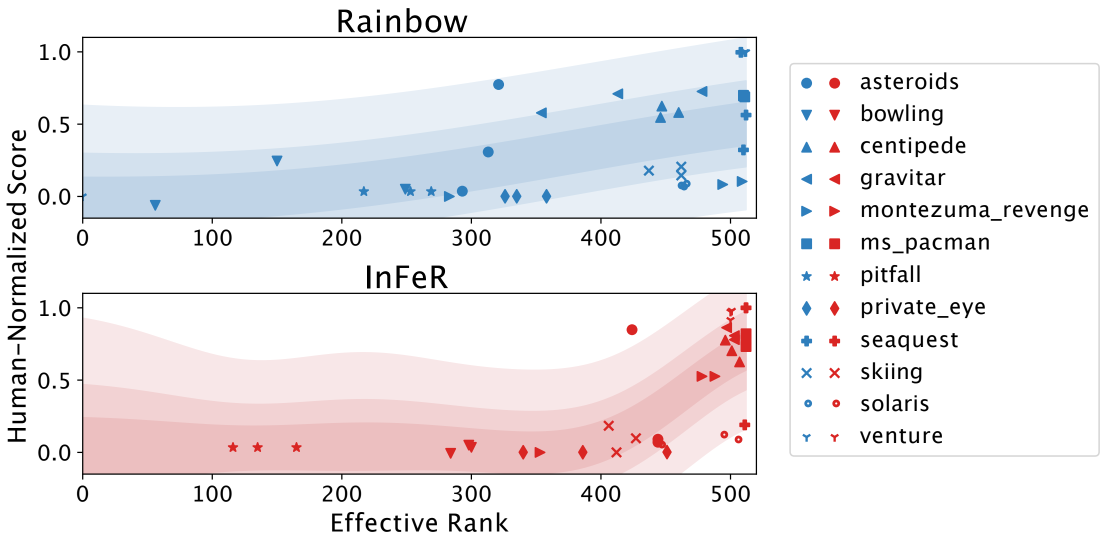
SAC with failing agent (RR=9) buffer is initialized with the replay buffer of the failed SAC agent.
This experiment articulates that the primacy bias is not a failure to collect proper data per se, but rather a failure to learn from it.

The Primacy Bias in Deep RL: a tendency to overfit early experiences that damages the rest of the learning process.

[1] Kumar A, Agarwal R, Ghosh D, et al. Implicit under-parameterization inhibits data-efficient deep reinforcement learning[J]. ICLR 2021.
[2] Nikishin E, Schwarzer M, D'Oro P, et al. The primacy bias in deep reinforcement learning[C]. ICML 2022.

The phenomenon of plasticity loss in DRL and its catastrophic consequences.

➢ Capacity Loss[1]:

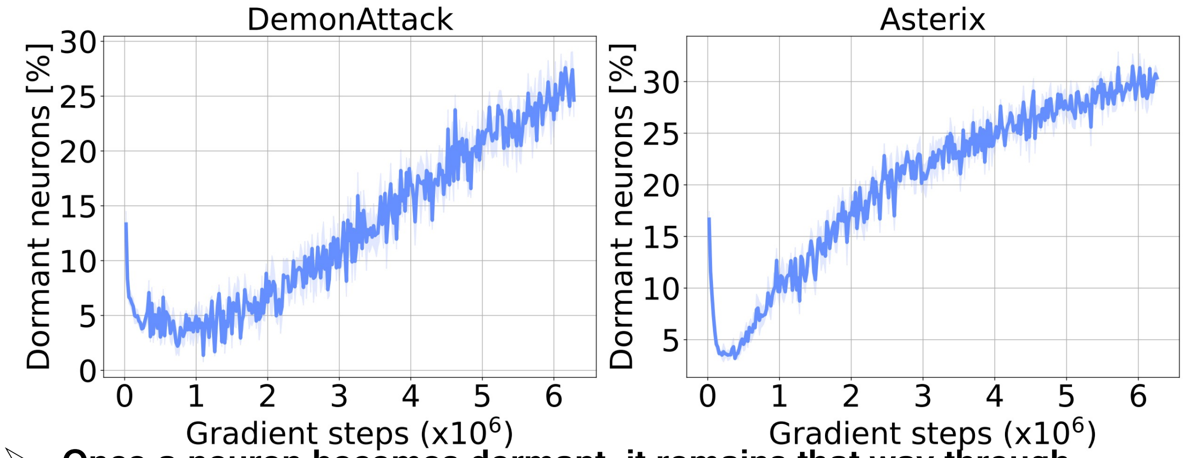


While **feature rank** does not appear to solely determine agent performance, there is a positive correlation between feature rank and human-normalized score.

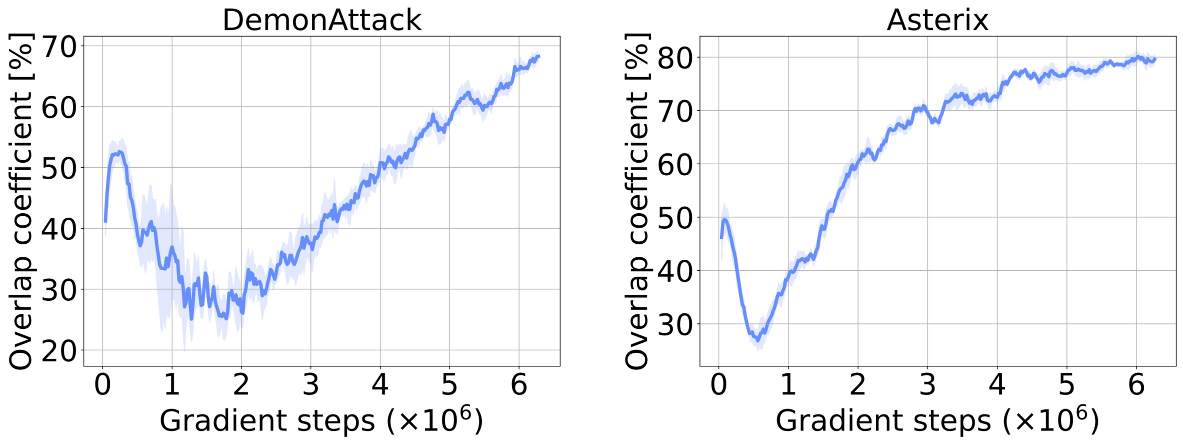
A fundamental challenge facing deep RL agents: **loss of the capacity to distinguish states and represent new target functions over the course of training.**

➢ Dormant Neuron Phenomenon[2]:

- **The percentage of dormant neurons increases throughout training.**



- **Once a neuron becomes dormant, it remains that way through training.**



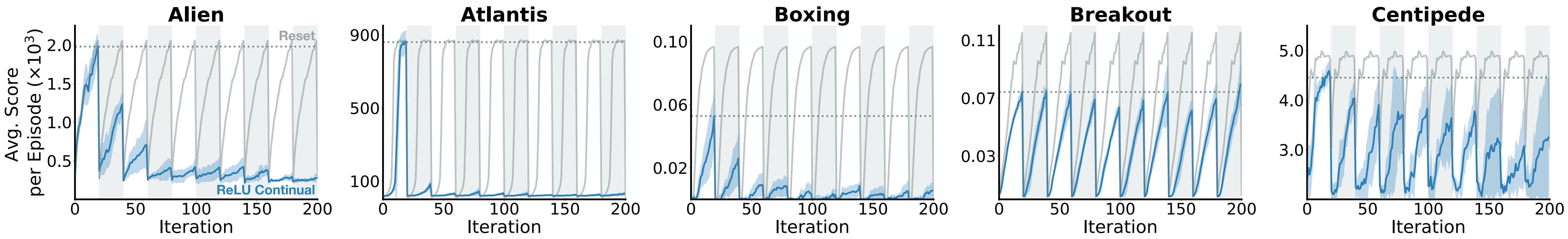
[1] Lyle C, Rowland M, Dabney W. Understanding and preventing capacity loss in reinforcement learning[J]. ICLR 2022

[2] Sokar G, Agarwal R, Castro P S, et al. The dormant neuron phenomenon in deep reinforcement learning[J]. ICML 2023

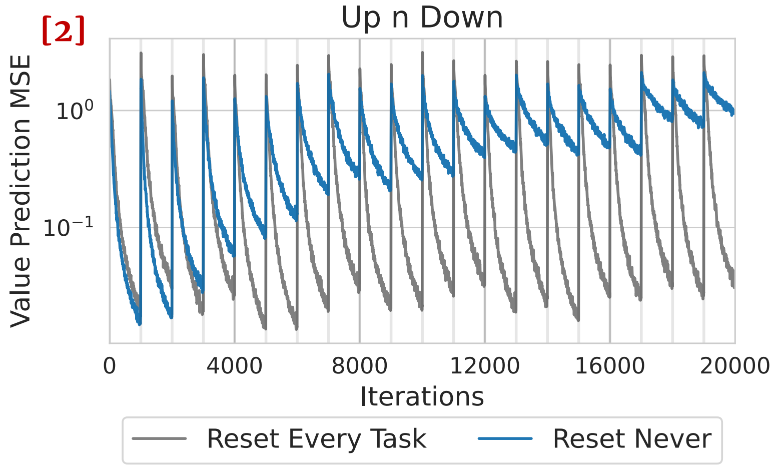
The phenomenon of plasticity loss in DRL and its catastrophic consequences.

➢ Loss of plasticity substantially hinders continual DRL[1]:

Rainbow's loss of plasticity on a repeating sequence of 5 Atari 2600 games. [1]



[2]



The CRL Problem (Informal) [3]

An RL problem is an instance of CRL if the best agents never stop learning.

[1] Abbas Z, Zhao R, Modayil J, et al. Loss of plasticity in continual deep reinforcement learning[J]. arXiv preprint.
[2] Nikishin E, Oh J, Ostrovski G, et al. Deep Reinforcement Learning with Plasticity Injection[J]. NeurIPS 2023.
[3] Abel D, Barreto A, Van Roy B, et al. A Definition of Continual Reinforcement Learning[J]. NeurIPS 2023.

How to mitigate the plasticity loss for achieving sample-efficient DRL?

- **Reset-based Methods:**

  Central Concept: naturally enhancing plasticity by integrating or re-initializing networks anew.

  Representative Methods: Reset, ReDo, Plasticity Injection

- **Regularization-based Methods:**

  Central Concept: regularizing neural network parameters to prevent overfitting and model collapse.

  Representative Methods: Shrink & Perturb, L2-Init, InFeR

- **Architectural-enhanced Methods:**

  Central Concept: enhancing model capacity and optimizing its architectural structure.

  Representative Methods: Normalization, C-ReLU, Wider Network, Rational Activations

- **Optimizer-enhanced Methods:**

  Central Concept: designing superior optimizers tailored to the requirements of RL training.

  Representative Methods: resetting the optimizer

Plasticity (Related Work)

- What does '**plasticity**' in neural networks signify?
- The phenomenon of **plasticity loss in DRL** and its catastrophic consequences.
- How to mitigate the plasticity loss for achieving sample-efficient DRL?

Data

DA notably outperforms Reset and other interventions in preserving plasticity, highlighting the inherent mechanisms driving its effectiveness.

Modules

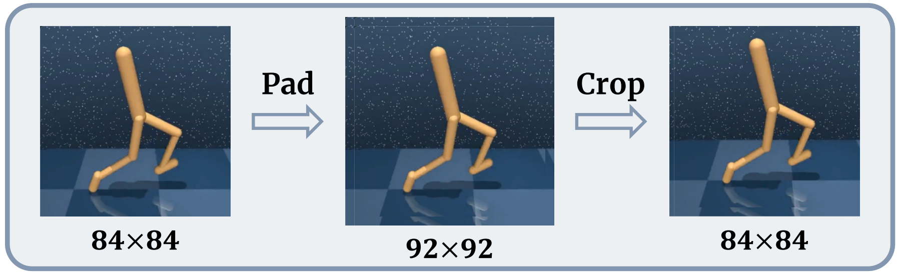
The primary bottleneck hindering sample-efficient VRL is the **critic's plasticity loss**, rather than the previously widely attributed **poor representation**.

Training Stages

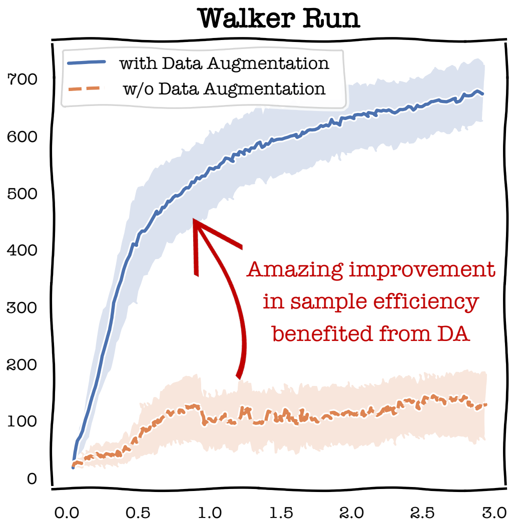
Without timely intervention to recover critic’s plasticity in the **early stages**, its loss becomes catastrophic and irrecoverable.

A Factorial Examination of Data Augmentation and Reset.

➢ Data Augmentation (DA):



Yarats D, Fergus R, Lazaric A, et al. Mastering visual continuous control: Improved data-augmented reinforcement learning[J]. ICLR 22



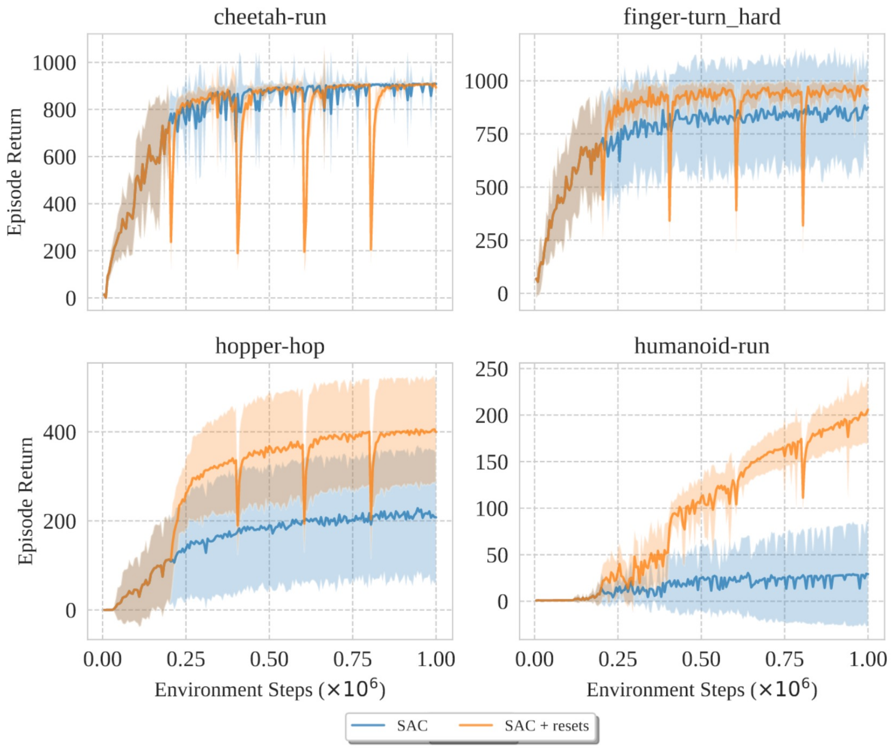
DA has become an **indispensable** component in achieving sample-efficient VRL applications. However, the **mechanisms driving DA's notable effectiveness** remain largely unclear.

➢ Reset:

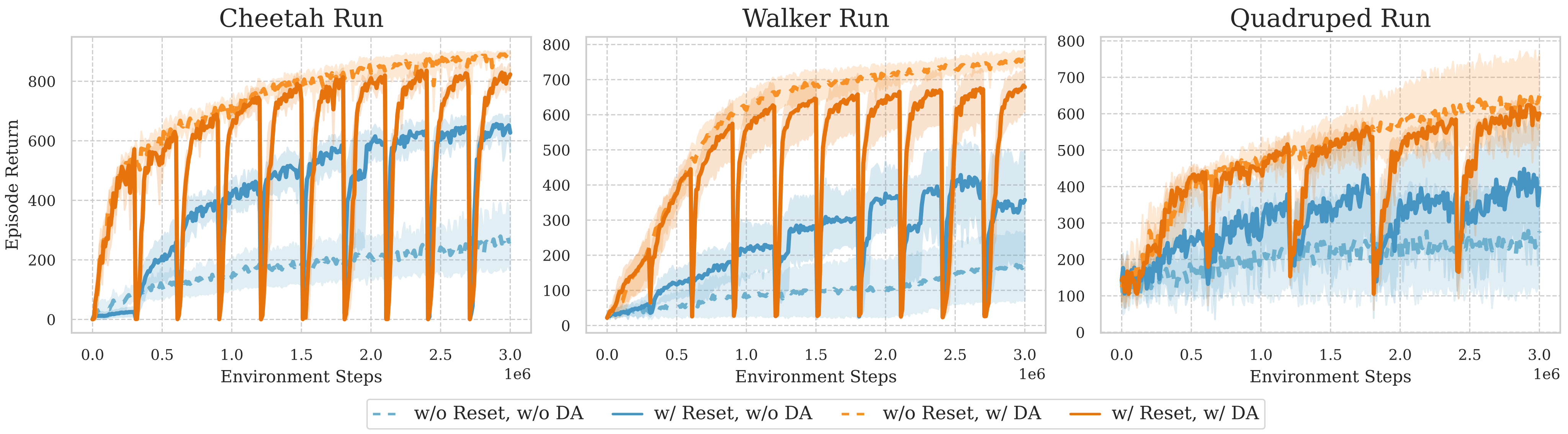
Given an agent's neural network, periodically re-initialize the parameters of its last few layers while preserving the replay buffer.

Reset stands as the most direct strategy to **recover an agent's plasticity**. In this approach, a re-initialized network manifests a complete restoration of its plasticity. By retaining the original buffer, not only can the agent rapidly resume its pre-Reset performance, but it also showcases heightened plasticity.

Nikishin E, Schwarzer M, D'Oro P, et al. The primacy bias in deep reinforcement learning[C] ICML 2022



A Factorial Examination of Data Augmentation and Reset.



- **In the absence of DA**, the implementation of Reset consistently yields marked enhancements. This underscores the evident plasticity loss when training is conducted devoid of DA.
- **With the integration of DA**, the introduction of Reset leads to only slight improvements, or occasionally, a decrease. This indicates that applying DA alone can sufficiently preserve the agent's plasticity, leaving little to no room for significant improvement.
- Comparatively, **the performance of Reset without DA lags behind that achieved employing DA alone**, underscoring the potent effectiveness of DA in preserving plasticity.

Comparing DA with Other Interventions.

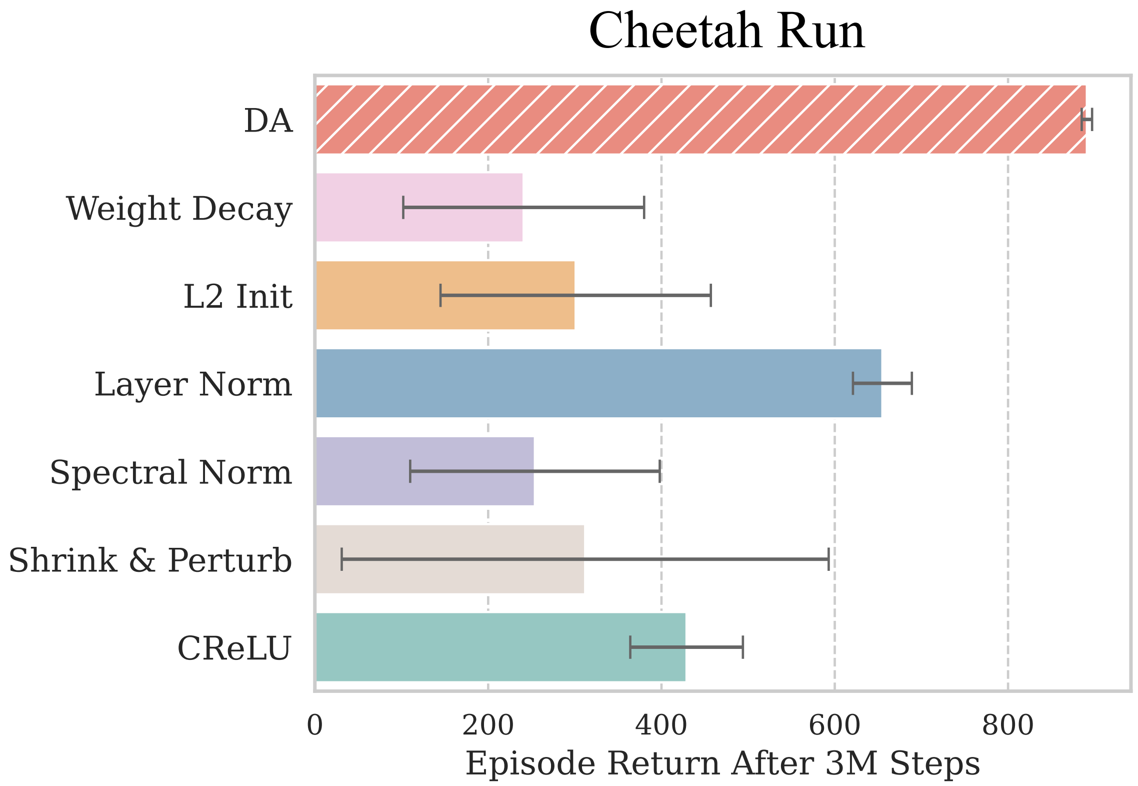
➢ **L2 Init[1]** integrates L2 regularization toward initial parameters instead of the origin.

$$\mathcal{L}_{\text{reg}}(\theta) = \mathcal{L}_{\text{train}}(\theta) + \lambda \|\theta - \theta_0\|_2^2$$

➢ **CReLU[2]:** concatenates the input with its negation and applies ReLU to the result, allowing a filter to be activated in both positive and negative direction.

$$\text{CReLU}(x) \doteq [\text{ReLU}(x), \text{ReLU}(-x)]$$

[1] Saurabh Kumar, Henrik Marklund, and Benjamin Van Roy. Maintaining plasticity via regenerative regularization. arXiv preprint arXiv:2308.11958, 2023.
[2] Zaheer Abbas, Rosie Zhao, Joseph Modayil, Adam White, and Marlos C Machado. Loss of plasticity in continual deep reinforcement learning. arXiv preprint arXiv:2303.07507, 2023.



The pronounced difference in **plasticity due to DA's presence or absence** provides compelling cases for comparison, allowing a deeper investigation into the **differences and developments of plasticity across different modules and stages.**

Plasticity (Related Work)

- What does '**plasticity**' in neural networks signify?
- The phenomenon of **plasticity loss in DRL** and its catastrophic consequences.
- How to mitigate the plasticity loss for achieving sample-efficient DRL?

Data

DA notably outperforms Reset and other interventions in preserving plasticity, highlighting the inherent mechanisms driving its effectiveness.

Modules

The primary bottleneck hindering sample-efficient VRL is the **critic's plasticity loss**, rather than the previously widely attributed **poor representation**.

Training Stages

Without timely intervention to recover critic’s plasticity in the **early stages**, its loss becomes catastrophic and irrecoverable.

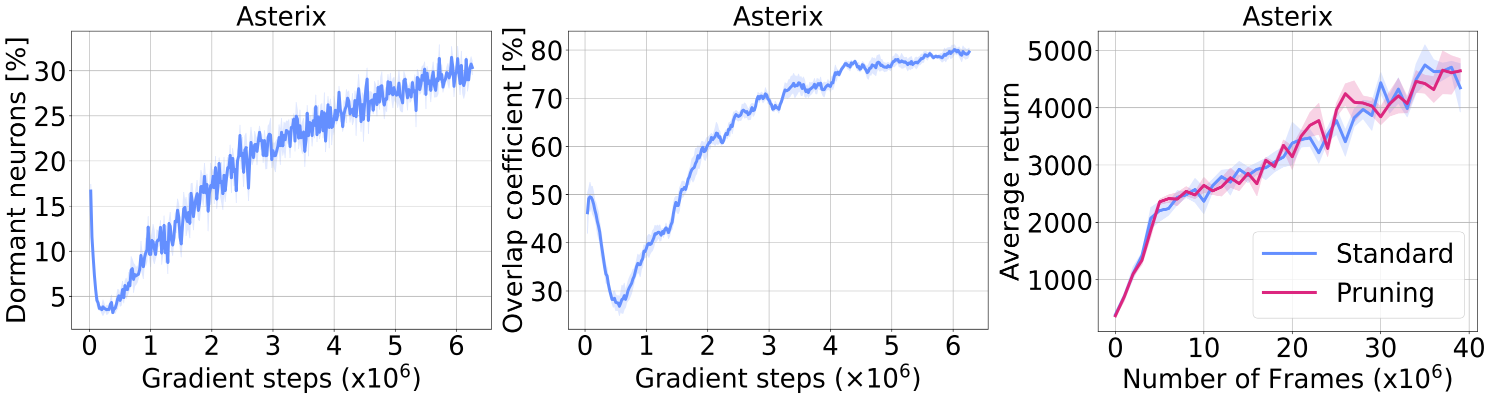
Fraction of Active Units (FAU).

Although the complete mechanisms underlying plasticity loss remain unclear, a reduction in the number of active units within the network has been identified as a principal factor contributing to this deterioration.

The activation of neuron *n* given the input *x*

FAU of Module $\mathcal{M}$: $$\Phi_{\mathcal{M}} = \frac{\sum_{n \in \mathcal{M}} \mathbf{1}(\boxed{a_n(x)} > 0)}{\boxed{N}}$$

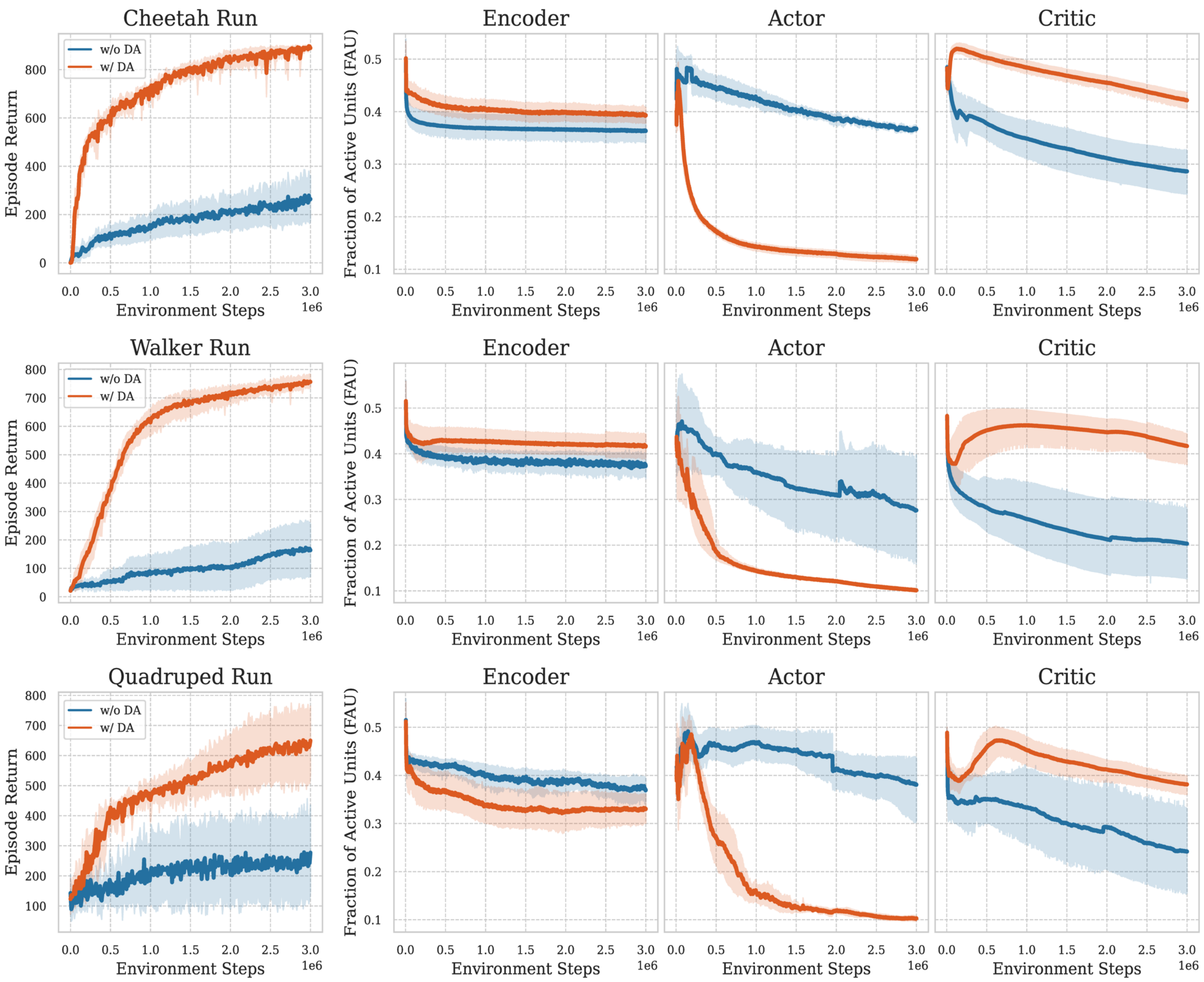
The total number of neurons within module

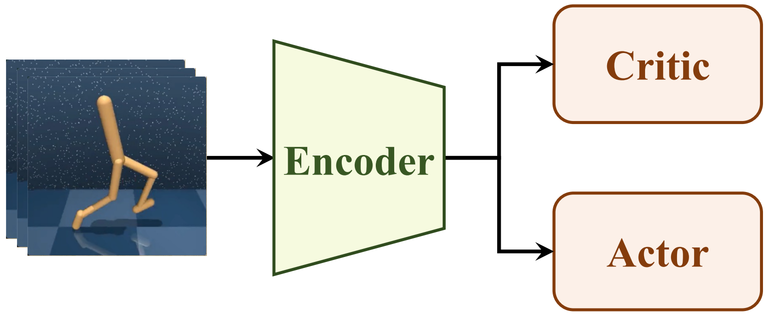


- The **overlap coefficient for dormant neurons** rises during training, indicating that neurons, once dormant, tend to remain inactive throughout.
- **Pruning dormant neurons** during training does not affect the performance of an agent, further confirming that dormant neurons remain dormant.

Sokar G, Agarwal R, Castro P S, et al. The dormant neuron phenomenon in deep reinforcement learning[J]. ICML 2023

Different FAU trends across modules throughout training.

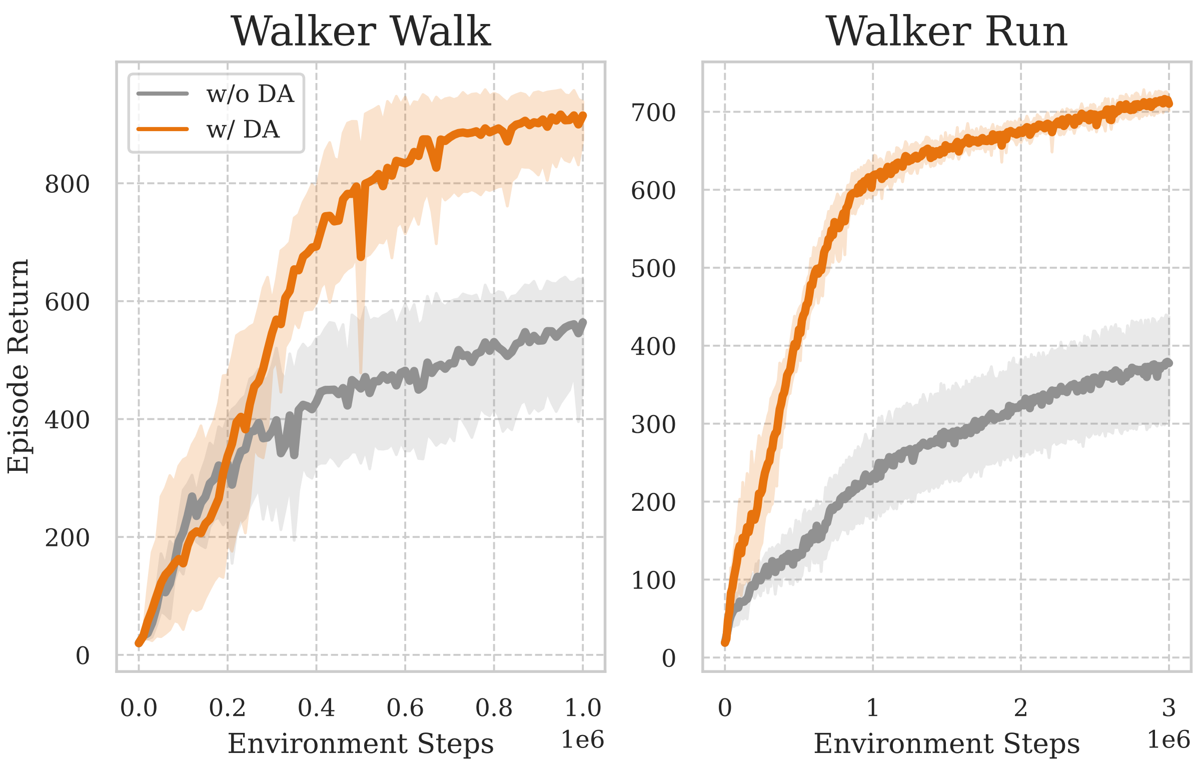


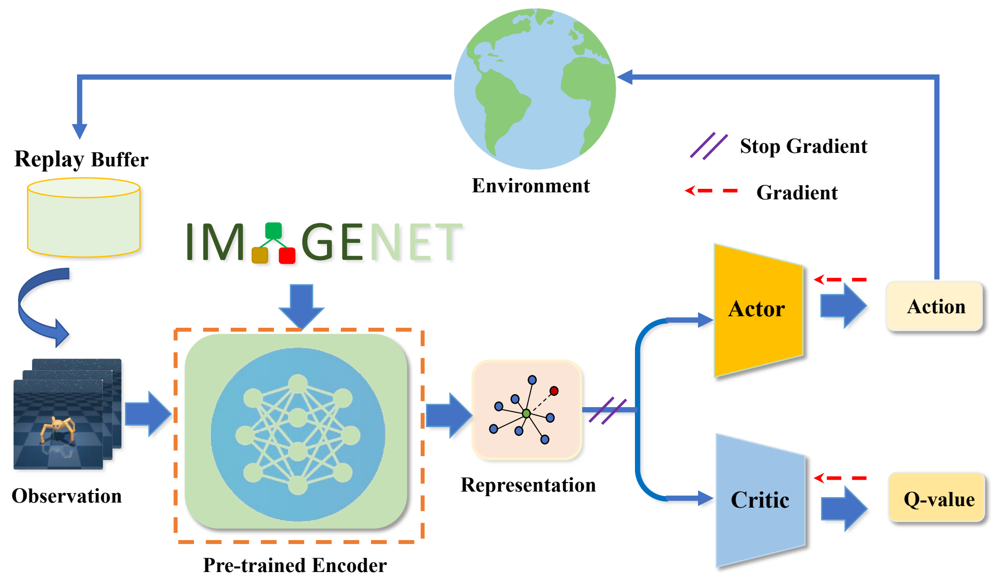


- The integration of DA leads to a substantial leap in **training performance**.
- Both the **encoder and actor's FAU** exhibit similar trends regardless of DA's presence or absence.
- DA elevates the **critic's plasticity** to a level almost equivalent to an initialized network.

This finding tentatively suggests that critic's plasticity loss is the bottleneck constraining training efficiency.

Is the sample inefficiency in VRL truly blamed on poor representation?





PIE-G: Yuan Z, Xue Z, Yuan B, et al. Pre-trained image encoder for generalizable visual reinforcement learning[J]. NeurIPS 2023.

- Instead of training the encoder from scratch, we employ an ImageNet pre-trained ResNet model as the agent's encoder and retain its parameters frozen throughout the training process.
- Employing DA consistently surpasses scenarios without DA by a notable margin.

This significant gap reveals that the sample inefficiency in VRL cannot be predominantly attributed to poor representation.

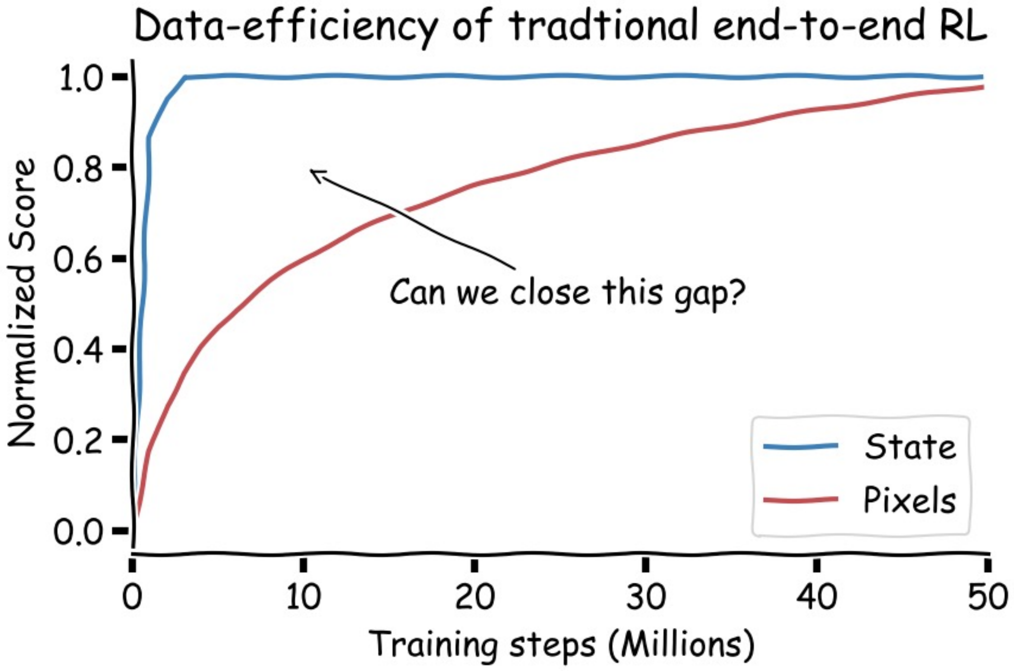
Is the sample inefficiency in VRL truly blamed on poor representation?

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Subscribe About Archive BAIR

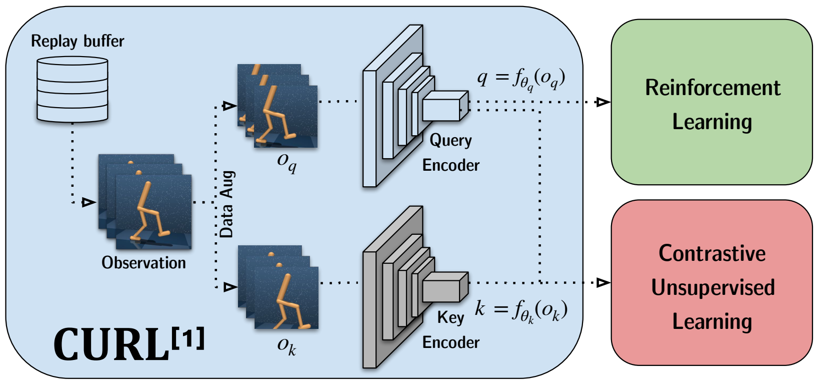
Can RL From Pixels be as Efficient as RL From State?

Misha Laskin, Aravind Srinivas, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel
Jul 19, 2020

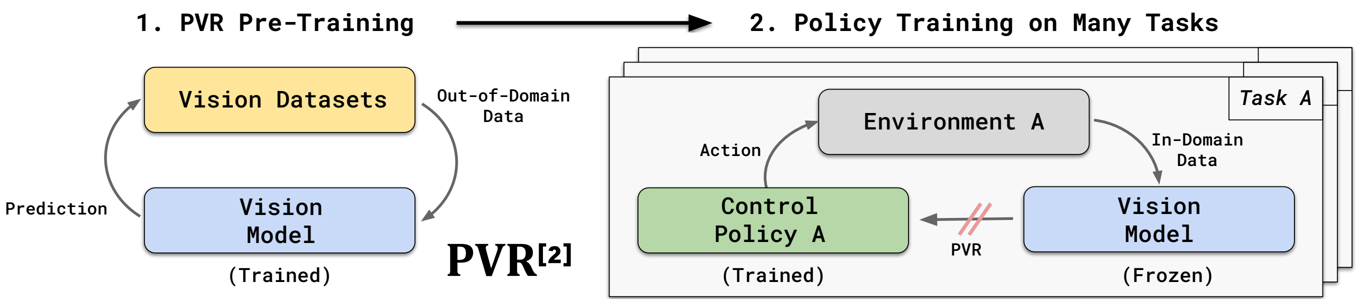


Towards Sample-Efficient VRL:

➢ Crafting auxiliary self-supervised learning tasks:



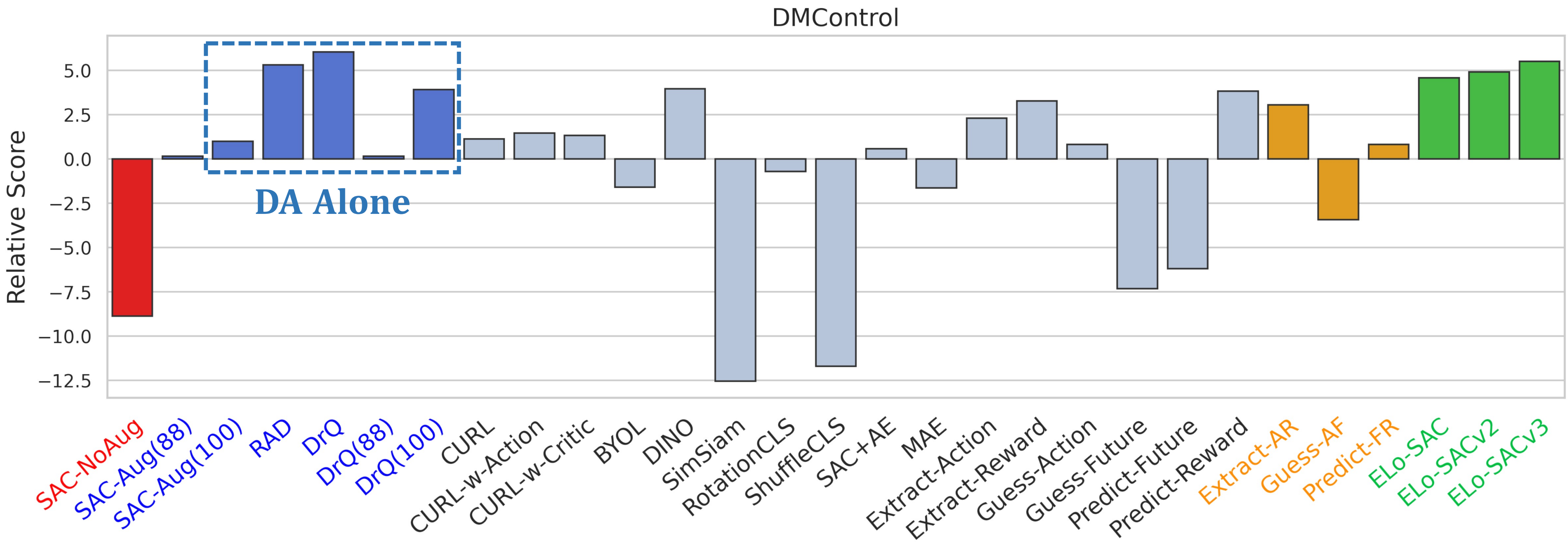
➢ Pre-training representation encoders with extra data:



[1] Laskin M, Srinivas A, Abbeel P. Curl: Contrastive unsupervised representations for reinforcement learning[C]. ICML 2020
[2] Parisi S, Rajeswaran A, Purushwalkam S, et al. The unsurprising effectiveness of pre-trained vision models for control[C]. ICML 2022.

Is the sample inefficiency in VRL truly blamed on poor representation?

➢ Does Self-supervised Learning Really Improve Reinforcement Learning from Pixels?

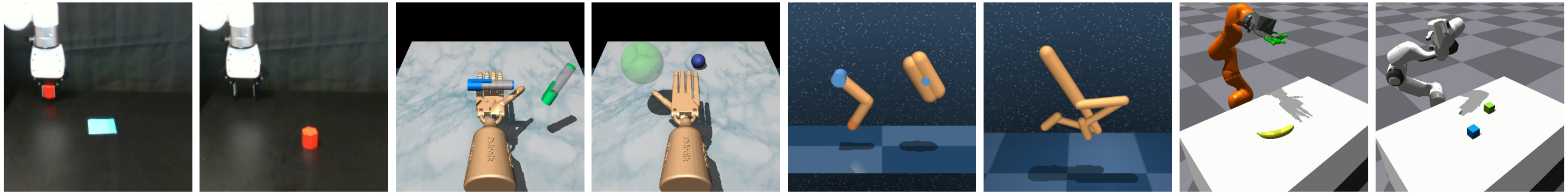


Our observations suggest that the **existing SSL framework for RL** fails to bring meaningful improvement over the baselines **only taking advantage of image augmentation** when the same amount of data and augmentation is used.

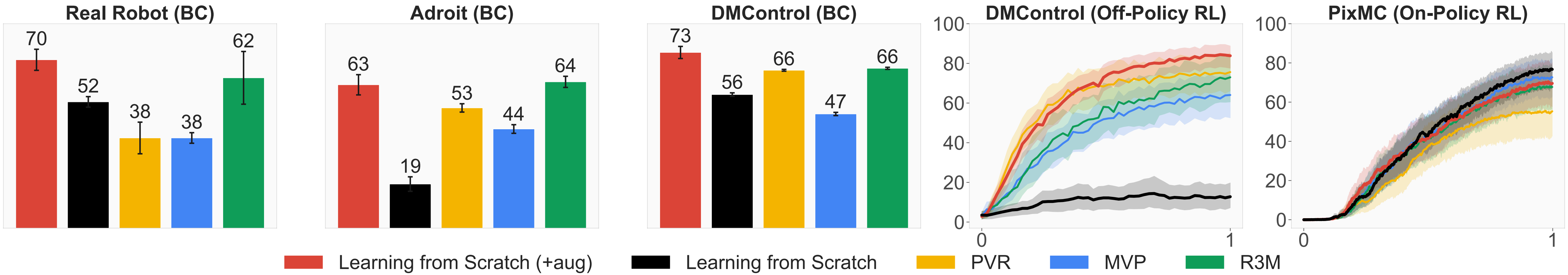
Li X, Shang J, Das S, et al. Does self-supervised learning really improve reinforcement learning from pixels?[J]. NeurIPS 2022

Is the sample inefficiency in VRL truly blamed on poor representation?

➢ Does Pre-Trained Encoder Really Improve the Sample Efficiency of VRL?



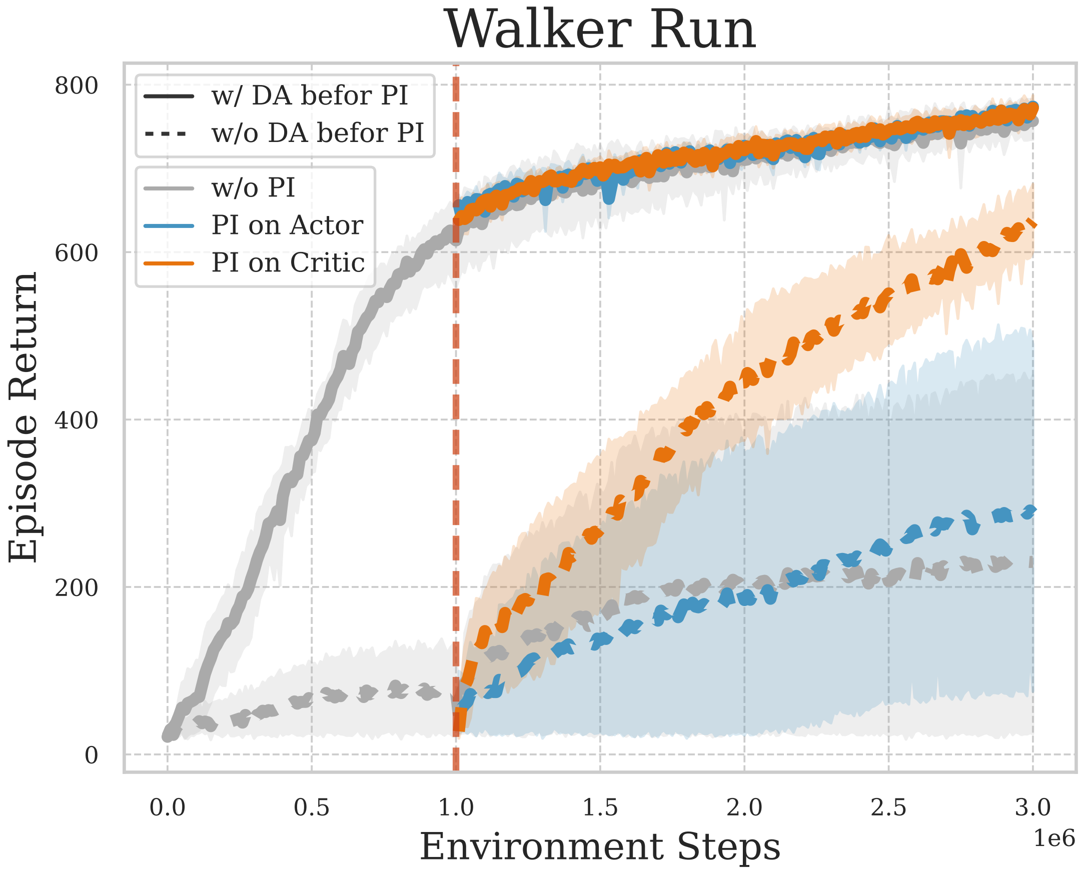
Real robot **Adroit** **DMControl** **PixMC**

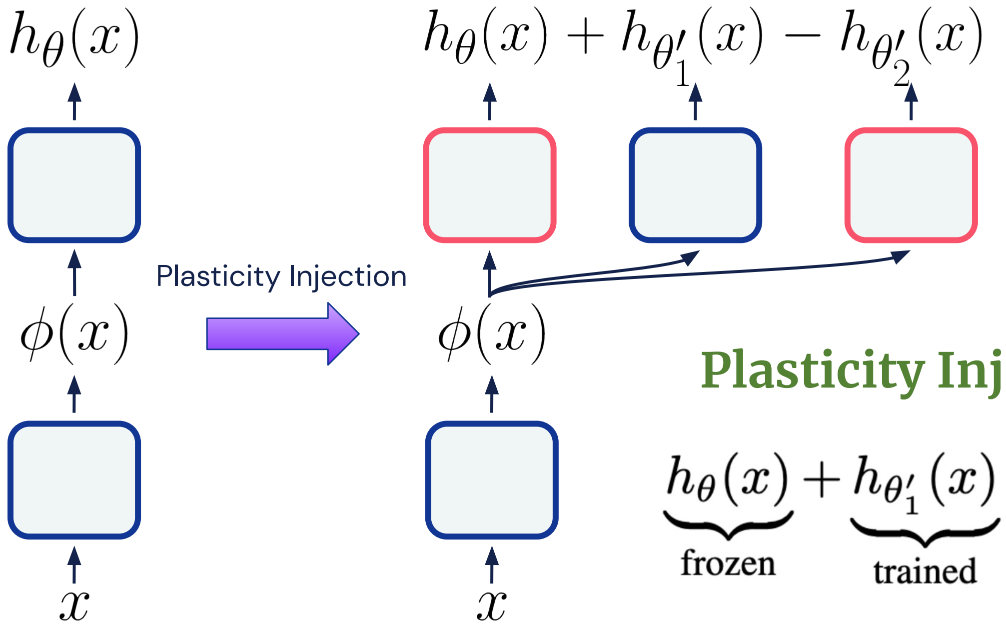


A simple Learning-from-Scratch (LfS) baseline that **incorporates data augmentation and a shallow ConvNet** is surprisingly competitive with recent approaches that **leverage frozen visual representations trained on large-scale vision datasets.**

Hansen N, Yuan Z, Ze Y, et al. On Pre-Training for Visuo-Motor Control: Revisiting a Learning-from-Scratch Baseline[J]. ICML 2023

Plasticity Injection on Actor and Critic as a Diagnostic Tool.





Plasticity Injection

$$\underbrace{h_\theta(x)}_{\text{frozen}} + \underbrace{h_{\theta'_1}(x)}_{\text{trained}} - \underbrace{h_{\theta'_2}(x)}_{\text{frozen}}$$

- same number of trainable parameters;
- gradients from the frozen heads affect the encoder too;

Nikishin E, Oh J, Ostrovski G, et al. Deep Reinforcement Learning with Plasticity Injection[J]. NeurIPS 2023

- DA alone is sufficient to maintain plasticity within the Walker Run task.
- Plasticity injection to the critic resulted in a significant performance improvement.

Critic's plasticity loss is the primary culprit behind VRL's sample inefficiency.

Plasticity (Related Work)

- What does '**plasticity**' in neural networks signify?
- The phenomenon of **plasticity loss in DRL** and its catastrophic consequences.
- How to mitigate the plasticity loss for achieving sample-efficient DRL?

Data

DA notably outperforms Reset and other interventions in preserving plasticity, highlighting the inherent mechanisms driving its effectiveness.

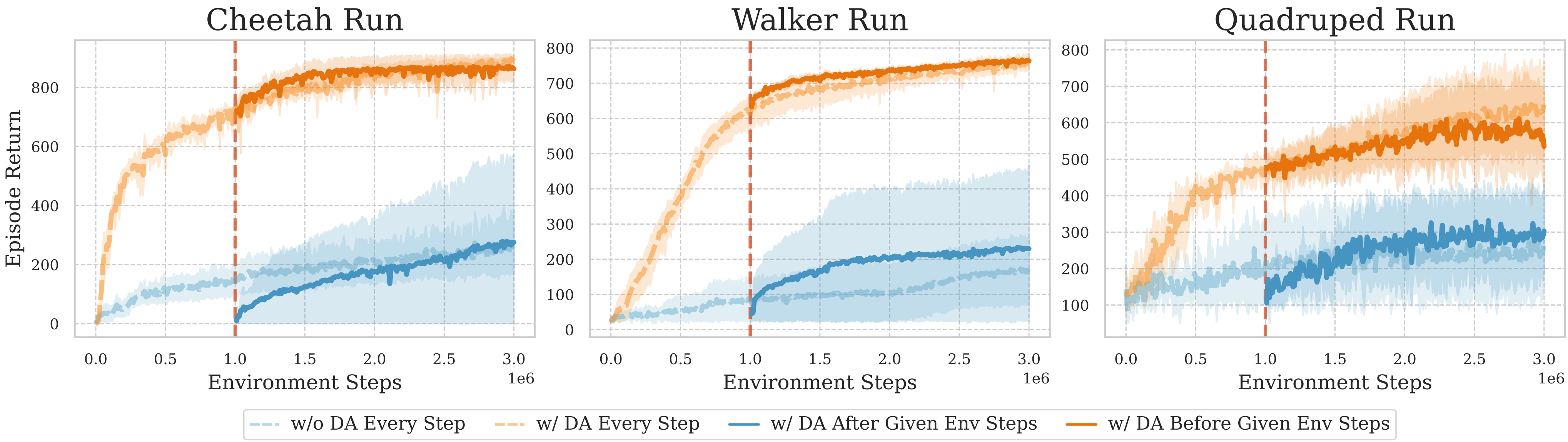
Modules

The primary bottleneck hindering sample-efficient VRL is the **critic's plasticity loss**, rather than the previously widely attributed **poor representation**.

Training Stages

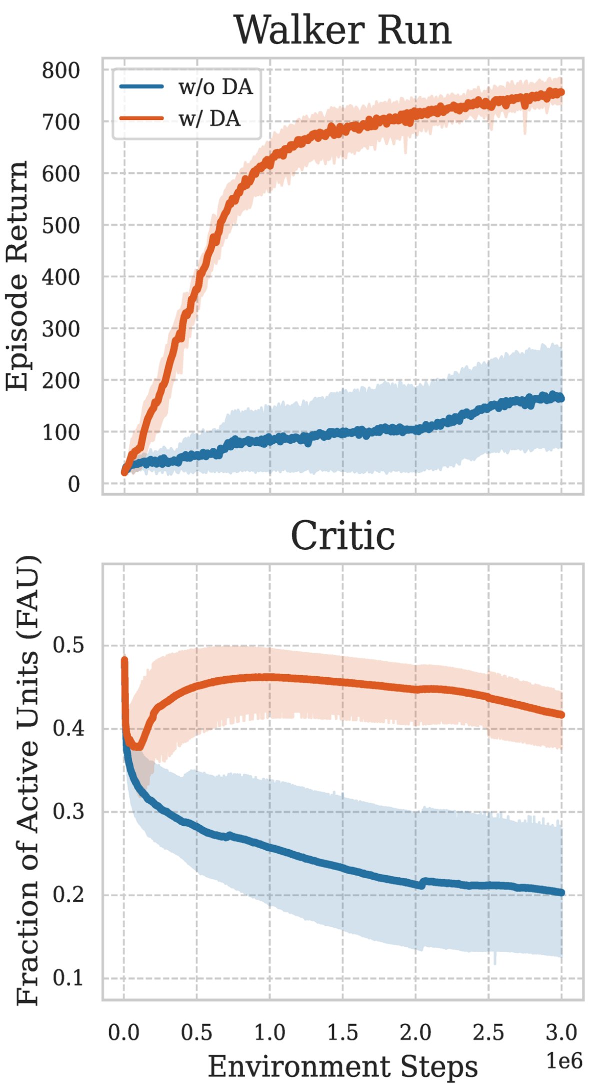
Without timely intervention to recover critic's plasticity in the **early stages**, its loss becomes catastrophic and irrecoverable.

What happens when DA is turned on or off at specific training steps?



- **Turning off DA** after the critic's plasticity has been recovered does not affect training.
  - This suggests that it is **not necessary to employ specific interventions to maintain plasticity in the later stages of training**.
- **Turning on DA** when plasticity has already been significantly lost and without timely intervention in the early stages cannot revive the agent's training performance.
  - This observation underscores **the vital importance of maintaining plasticity in the early stages**; otherwise, the loss becomes **irrecoverable**.

Catastrophic vs. Benign Plasticity Loss



Catastrophic Plasticity Loss:

- During the initial phases of training, bootstrapped target derived from **low-quality and limited-quantity** experiences exhibits **high non-stationarity** and deviates significantly from the actual state-action values.
- The severe non-stationarity of targets induces **a rapid decline in the critic's plasticity**.
- Having lost the ability to learn from newly collected data, the critic will perpetually fail to capture the dynamics of the environment, preventing the agent from acquiring an effective policy.

Benign Plasticity Loss:

- Although the critic's plasticity also experiences **a gradual decline after recovery**, this can be viewed as a process of **progressively approximating** the optimal value function for the current task.
- For **single-task VRL scenarios,** that doesn't require the agent to retain continuous learning capabilities.
- However, this loss of plasticity will have detrimental effects in the continual DRL scenarios.