DeepMind

Deep RL with Plasticity Injection


Junhyuk Oh


Georg Ostrovski


Clare Lyle


Razvan Pascanu


Will Dabney

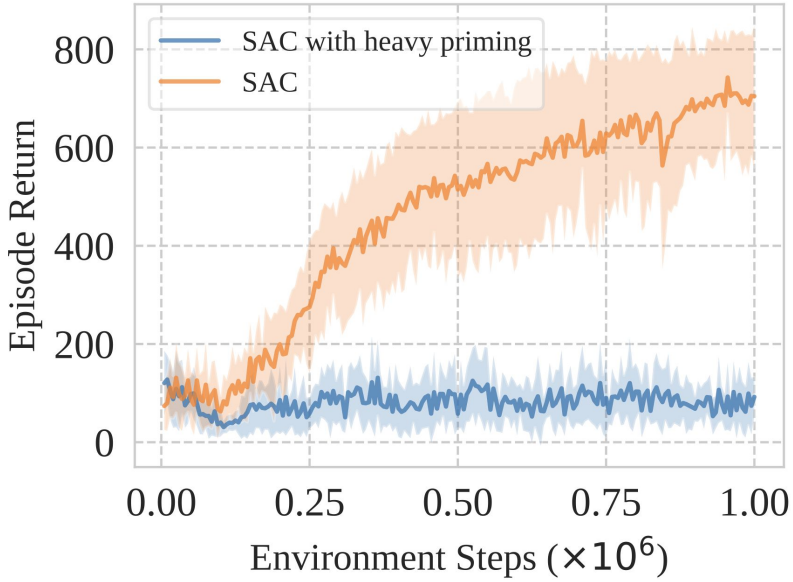

André Barreto

Evgenii Nikishin

Sep 15, 2023

The Primacy Bias Phenomenon

- A tendency to rely excessively on early experiences that damages the rest of the learning process
- An illustration: too many updates might unrecoverably impact the agent



Nikishin*, Schwarzer*, D'Oro* et al., 2022 "The Primacy Bias in Deep Reinforcement Learning"

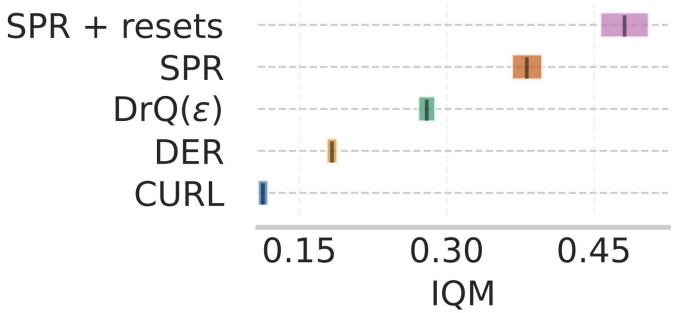
Resets Alleviate the Primacy Bias

- Re-initialize last layers of a network while keeping the replay buffer
- Resetting gives algorithmic-comparable improvements across domains

SAC: DMC dense

Resets	IQM
Yes	**656** (549, 753)
No	501 (389, 609)

SPR: Atari 100k

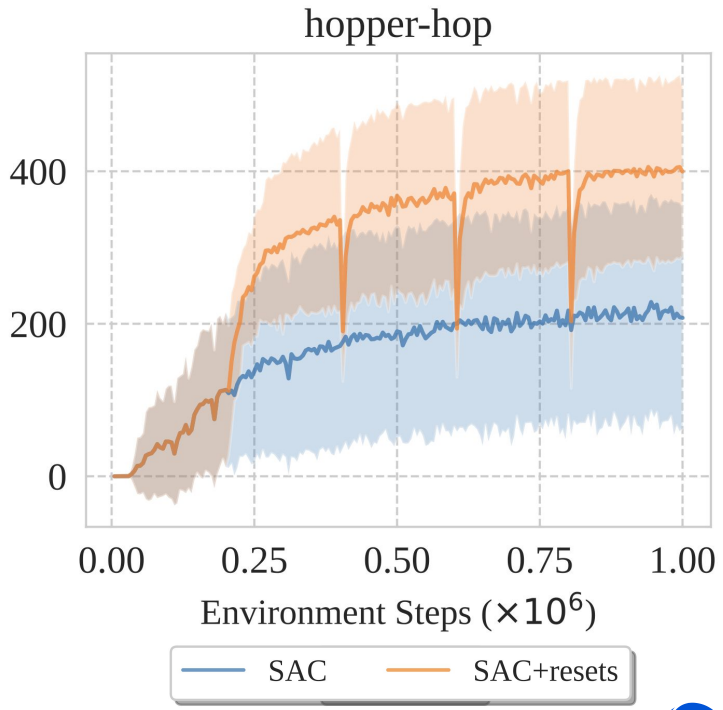


DrQ: DMC pixels

Resets	IQM
Yes	**762** (704, 815)
No	569 (475, 662)

What Resets are Helping with?

- Evidence that networks in RL lose plasticity [Dohare 2021]
- Resets restore the plasticity
- The post-reset policy is random -> can't attribute success to addressing plasticity only, an exploration confounder



Dohare et al., 2021 "Continual Backprop: Stochastic Gradient Descent with Persistent Randomness"

Challenges with Plasticity

- Plasticity = ability to learn
- The definition is broad. Existing proxies like
  - Weight Norm
  - Feature Rank [Kumar 2021, Lyle 2022]
  - Dead Units

  are incomplete [Gulcehre 2022]

Kumar*, Agarwal*, et al., 2021 “Implicit Under-Parameterization Inhibits Data-Efficient Deep Reinforcement Learning”

Lyle et al., 2022 “Understanding and Preventing Capacity Loss in Reinforcement Learning”

Gulcehre et al., 2022 “An Empirical Study of Implicit Regularization in Deep Offline RL”

Main Points

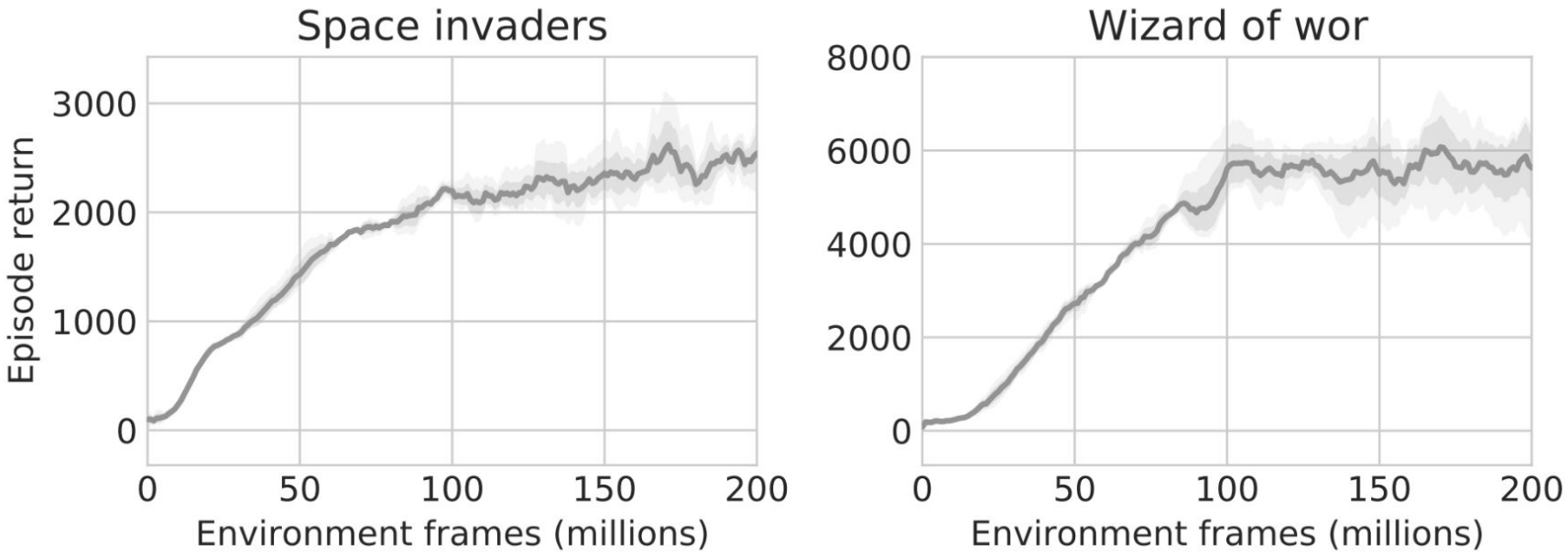
A Diagnostic Tool

How to disentangle plasticity loss from the rest of RL complexities

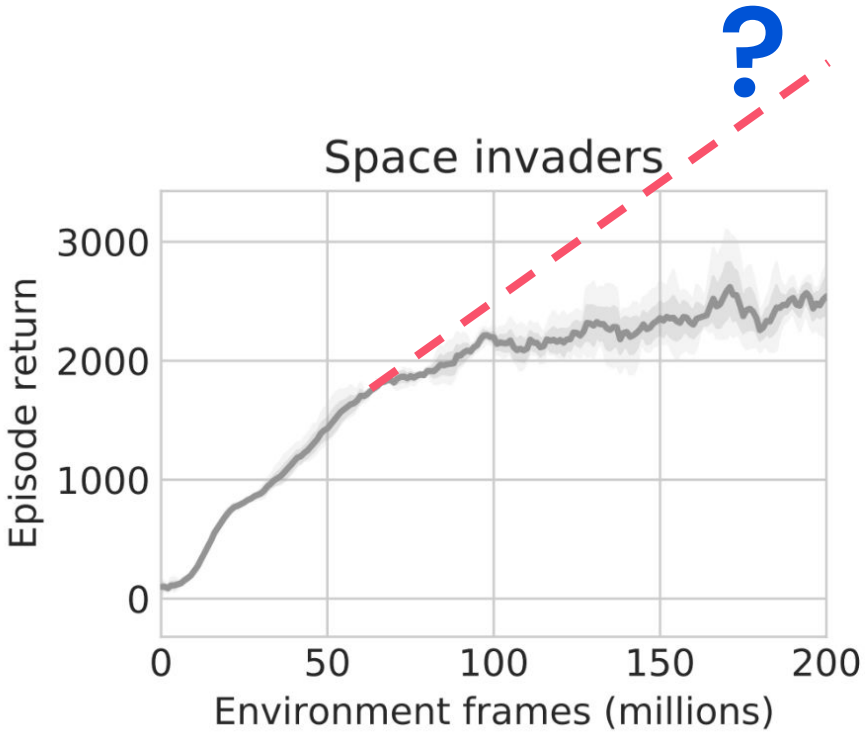
Dynamic plasticity addition

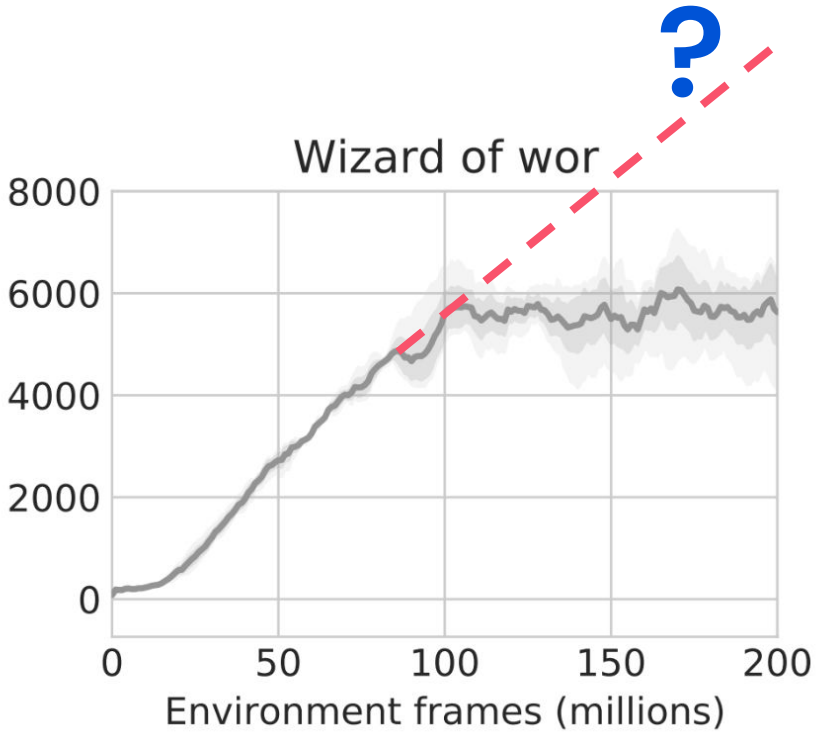
A way to save computations in large-scale RL

Example


Space invaders
Episode return
3000
2000
1000
0
0
50
100
150
200
Environment frames (millions)
Wizard of wor
8000
6000
4000
2000
0
0
50
100
150
200
Environment frames (millions)

Example

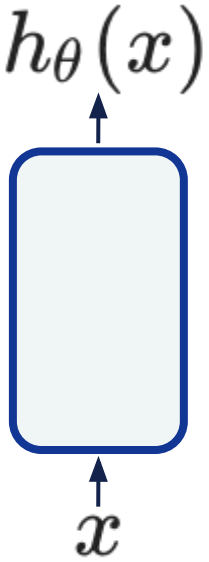

?
Space invaders
Episode return
3000
2000
1000
0
0
50
100
150
200
Environment frames (millions)


?
Wizard of wor
8000
6000
4000
2000
0
0
50
100
150
200
Environment frames (millions)

Plasticity Injection

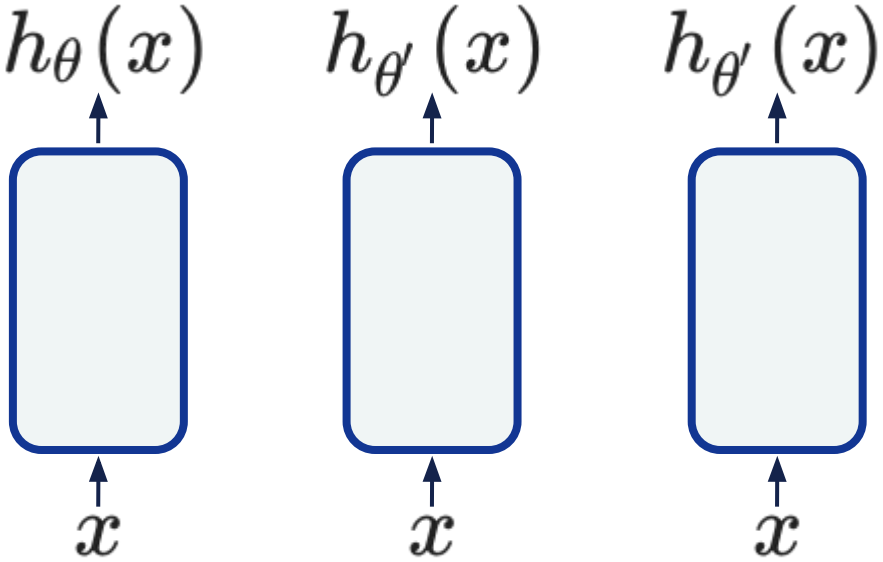
Plasticity Injection

- Take a network at, say, 50M steps



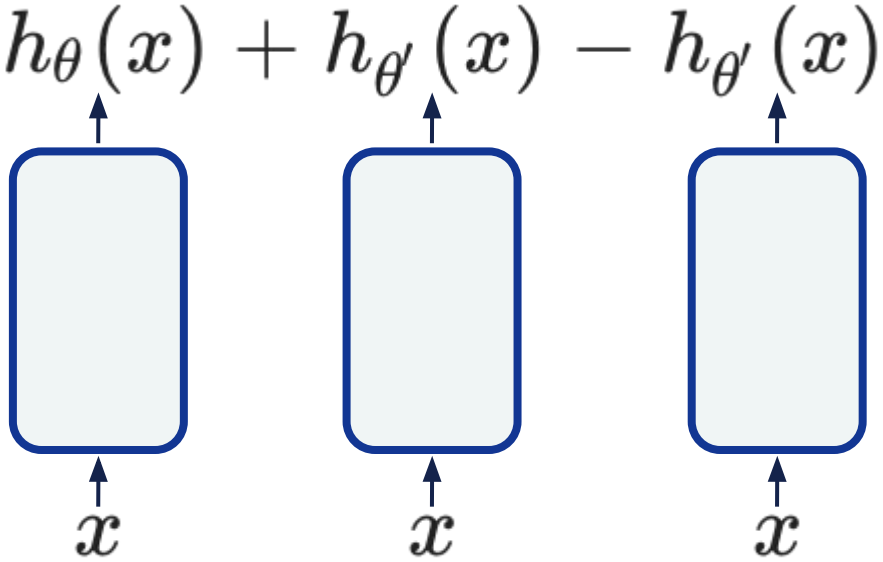
Plasticity Injection

- Take a network at, say, 50M steps
- Create 2 copies of a new network with randomly initialized parameters



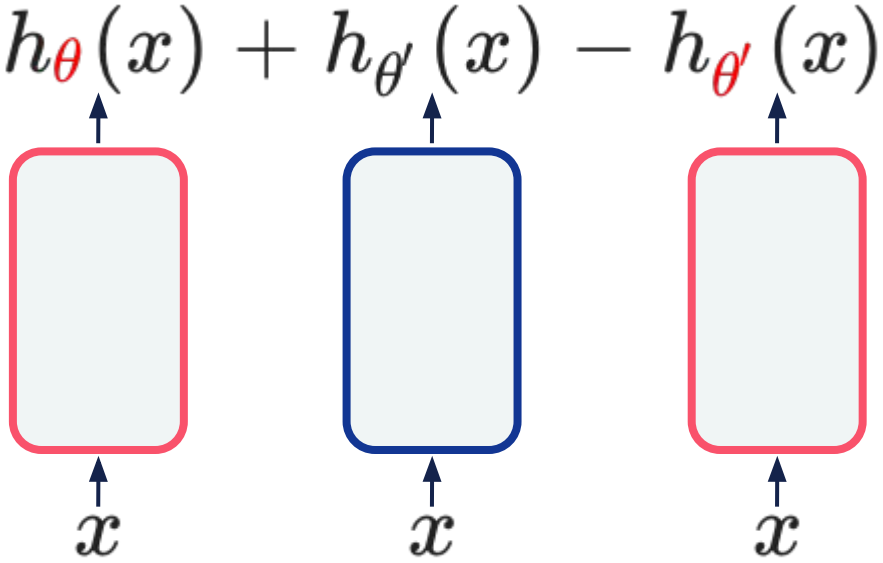
Plasticity Injection

- Take a network at, say, 50M steps
- Create 2 copies of a new network with randomly initialized parameters
- Add and subtract the outputs



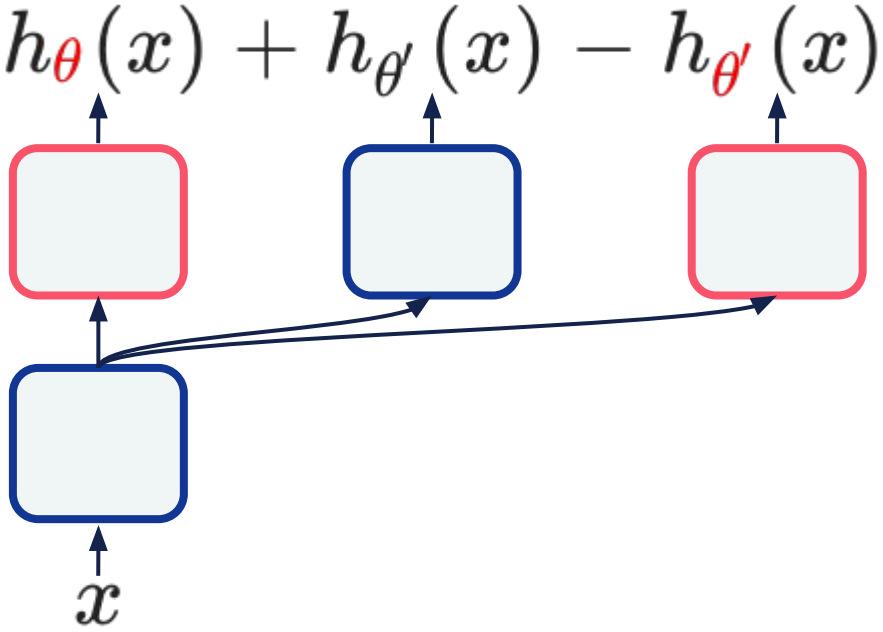
Plasticity Injection

- Take a network at, say, 50M steps
- Create 2 copies of a new network with randomly initialized parameters
- Add and subtract the outputs
- Freeze terms 1 and 3

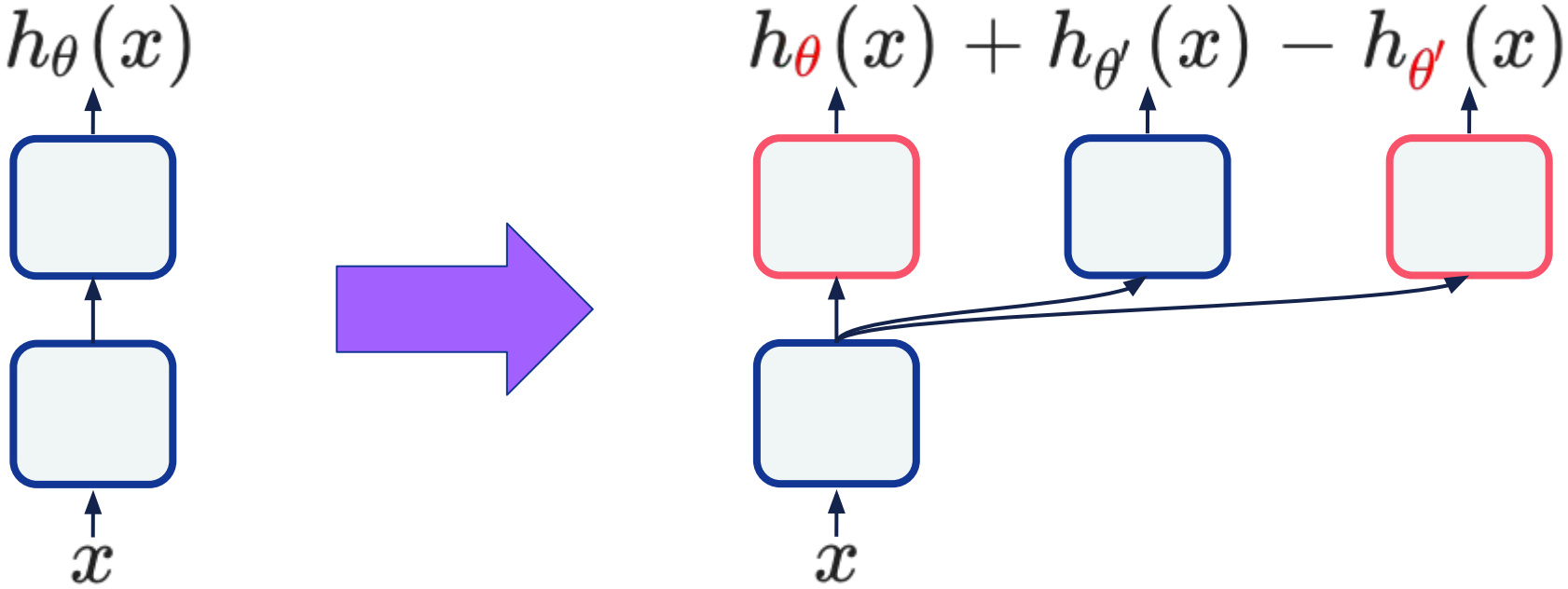


Plasticity Injection

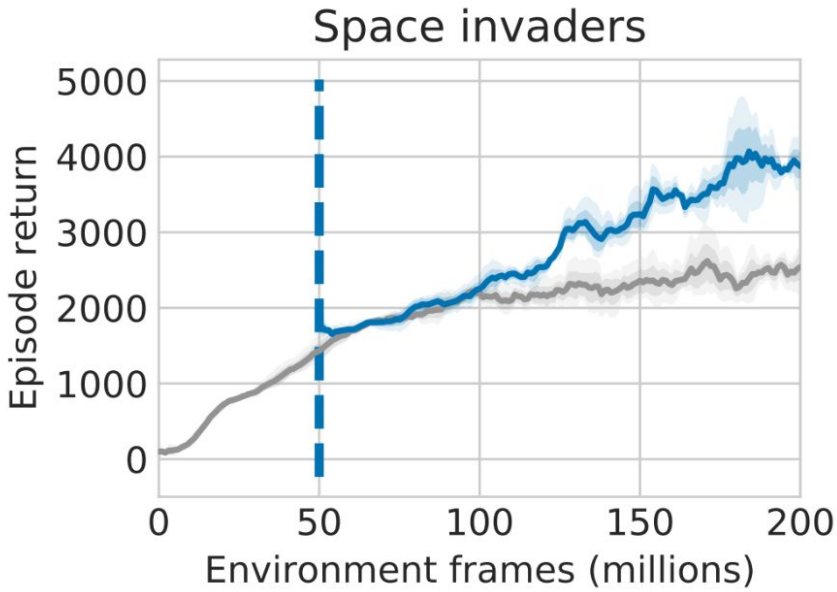
- Take a network at, say, 50M steps
- Create 2 copies of a new network with randomly initialized parameters
- Add and subtract the outputs
- Freeze terms 1 and 3
- Share the encoder to transfer representations

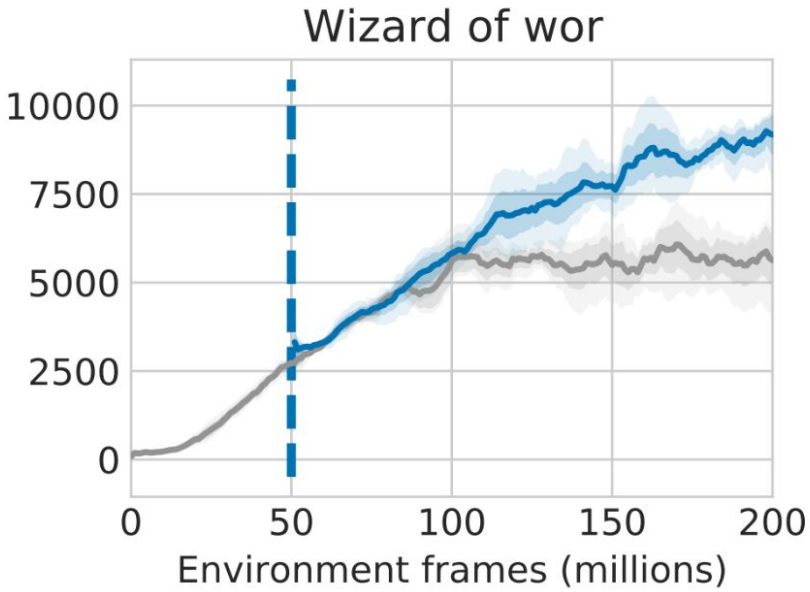


Plasticity Injection

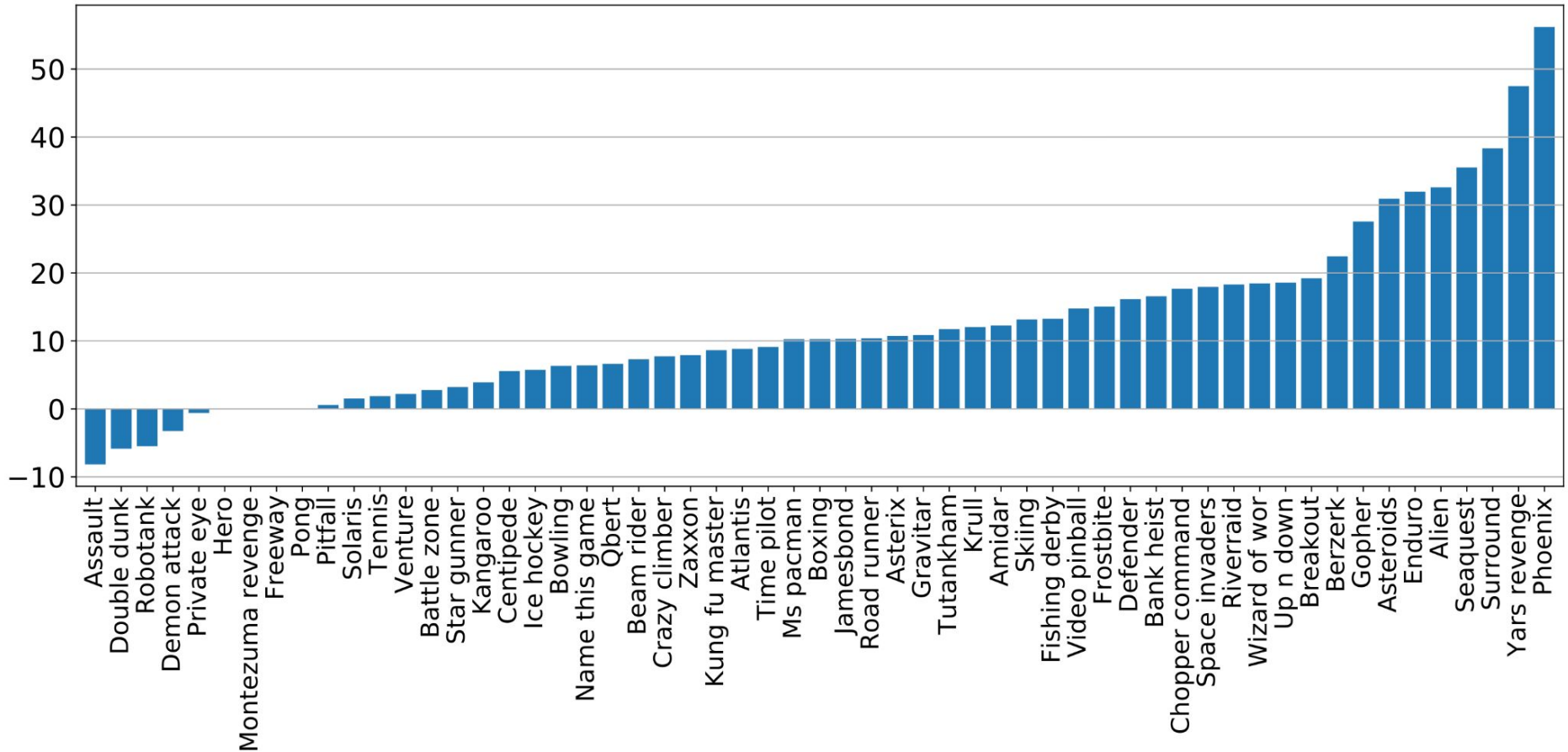


Back to the Example

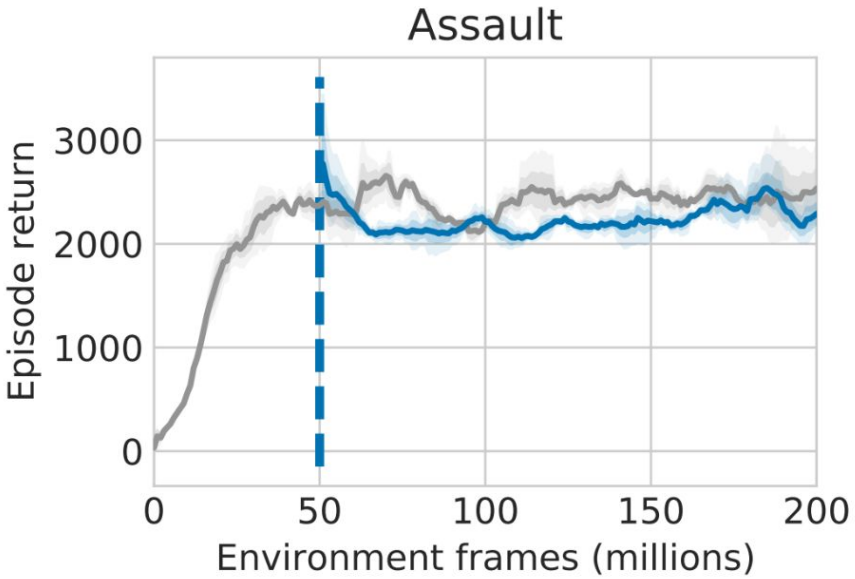




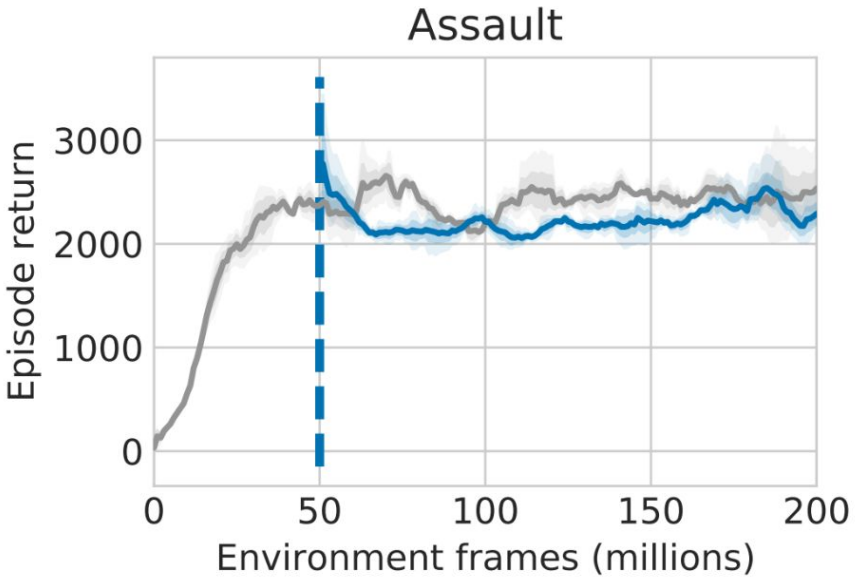
Generality of the Finding

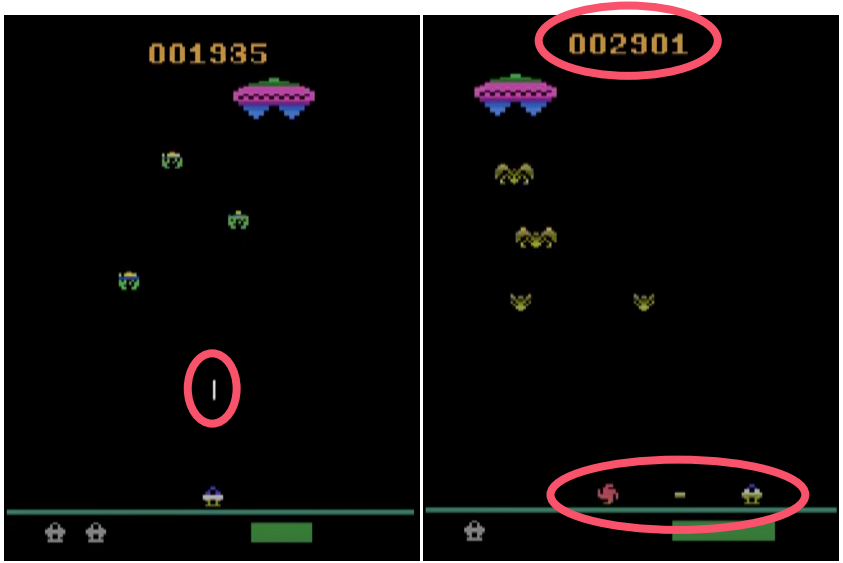


Negative Examples

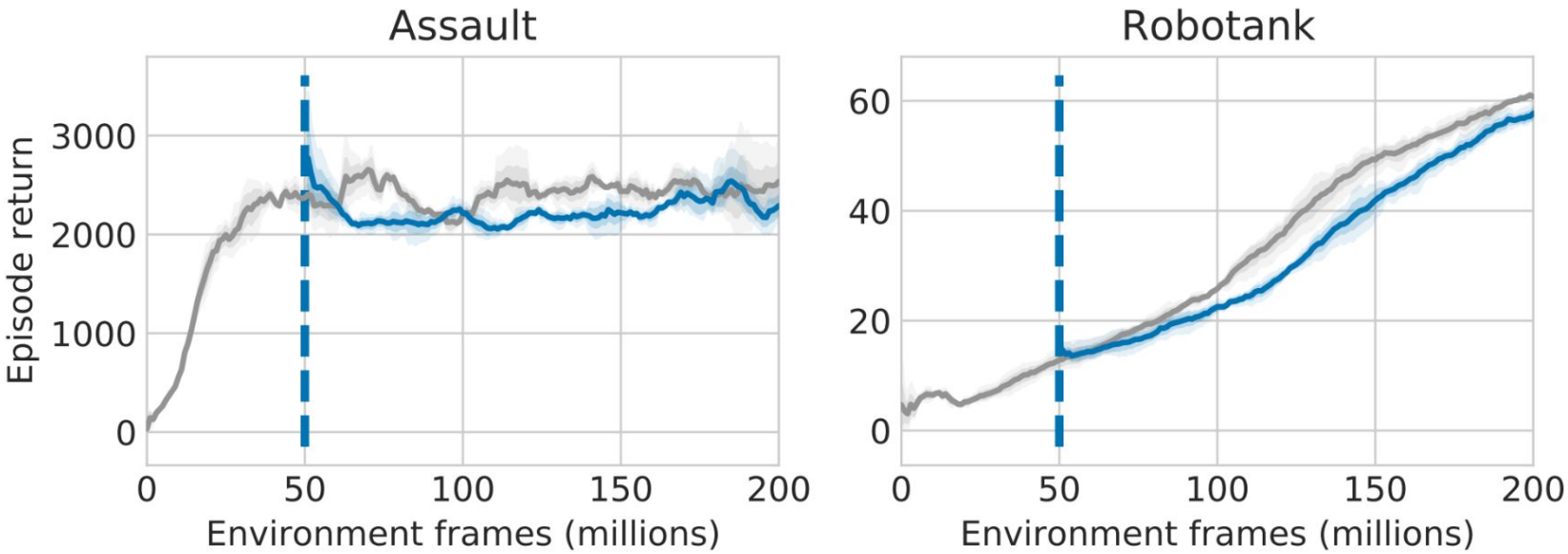


Negative Examples



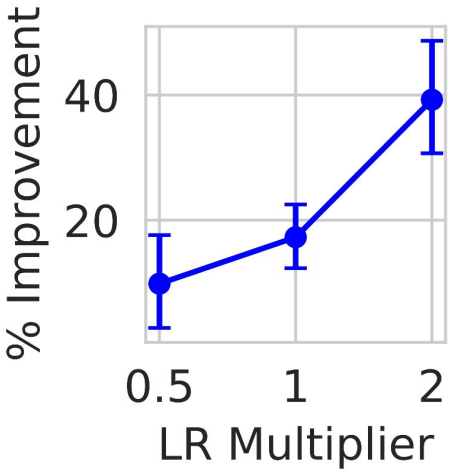


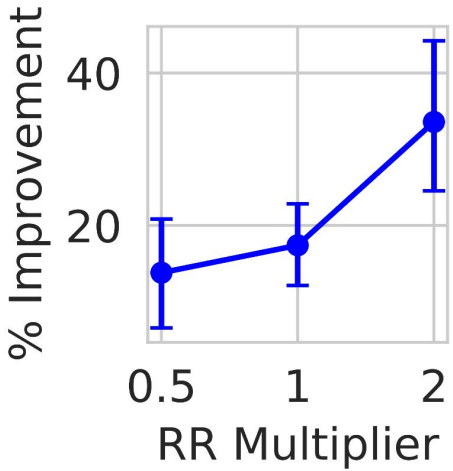
Negative Examples

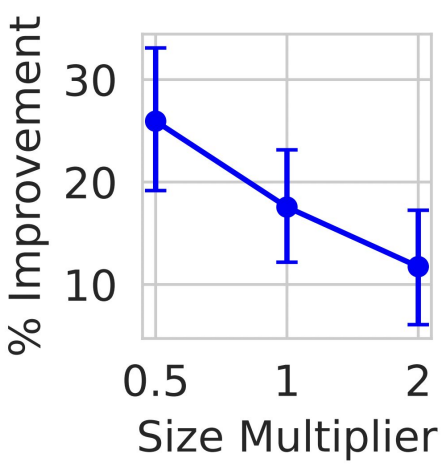


Extra Plasticity under Larger Plasticity Loss

- Learning rates, replay ratio, and network size control the plasticity loss pace
- Improvements from the injection grow with the amount of plasticity loss







Discussion

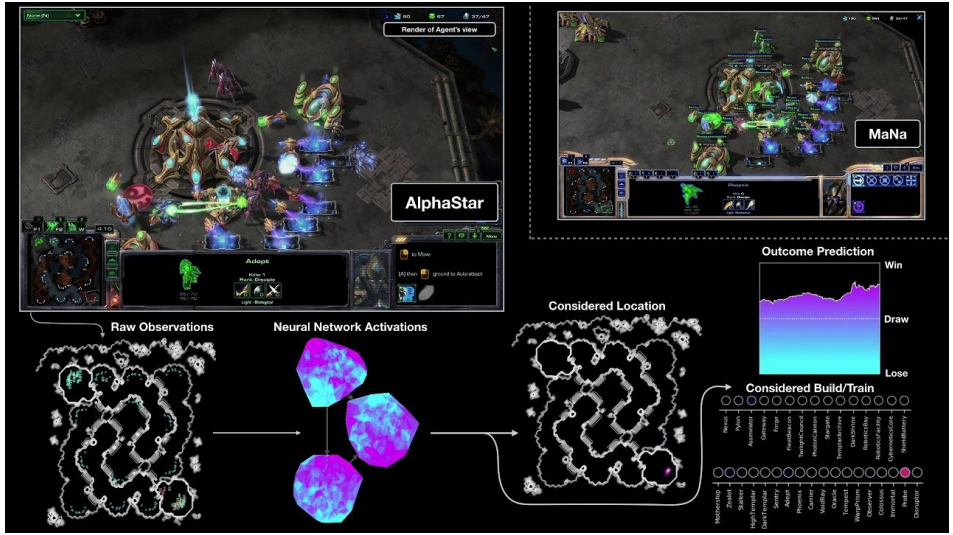
- Clean demo of the phenomenon in deep RL
- A diagnostic tool reveals non-uniform effects across games
- Additional plasticity addresses plasticity loss

Discussion

- Clean demo of the phenomenon in deep RL
- A diagnostic tool reveals non-uniform effects across games
- Additional plasticity addresses plasticity loss
- **“Can’t we just use a bigger net?”**

Large-Scale RL is Often Expensive

- Increased value of saving computations [Agarwal 2022]

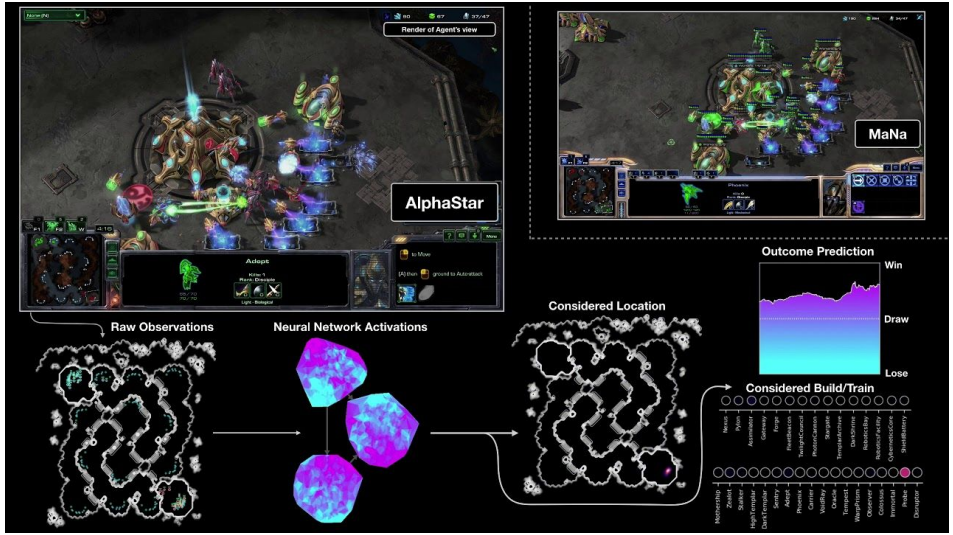


Vinyals et al., 2022 “Grandmaster level in StarCraft II using multi-agent reinforcement learning”

Agarwal et al., 2022 “Reincarnating Reinforcement Learning: Reusing Prior Computation to Accelerate Progress”

Large-Scale RL is Often Expensive

- Increased value of saving computations [Agarwal 2022]
- **Do we need all plasticity from the beginning?**
- **What if we lose plasticity but want to keep learning?**

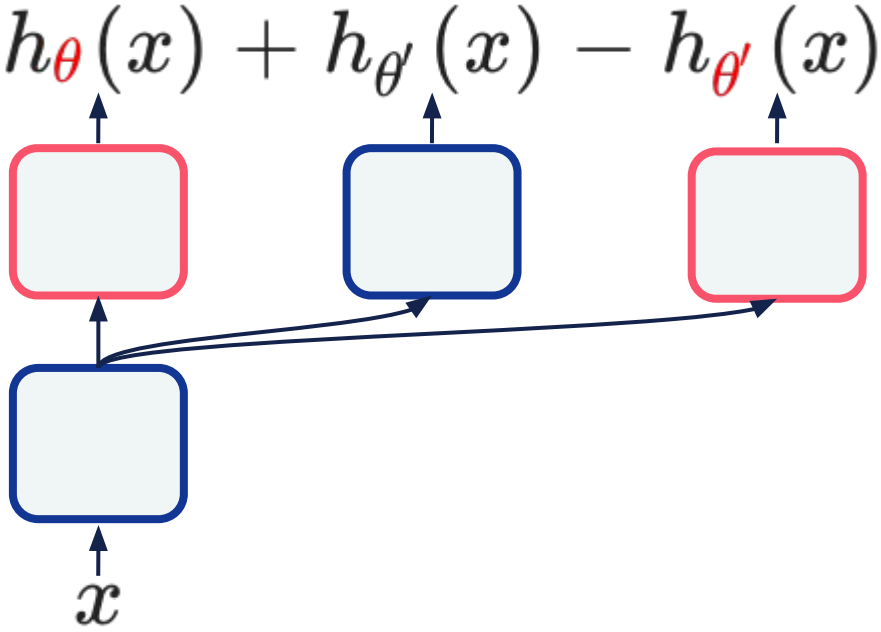


Vinyals et al., 2022 “Grandmaster level in StarCraft II using multi-agent reinforcement learning”

Agarwal et al., 2022 “Reincarnating Reinforcement Learning: Reusing Prior Computation to Accelerate Progress”

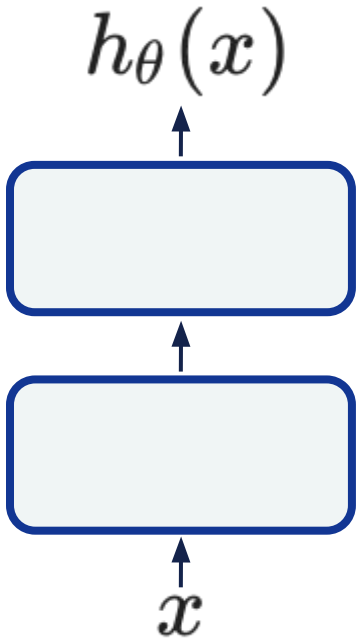
Injection vs Bigger Nets

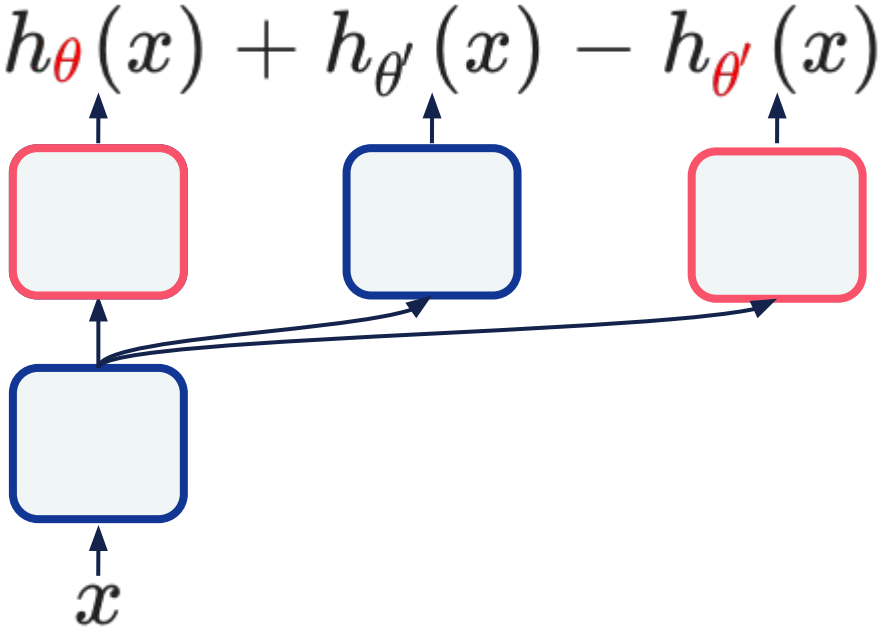
- Plasticity injection @ 50M



Injection vs Bigger Nets

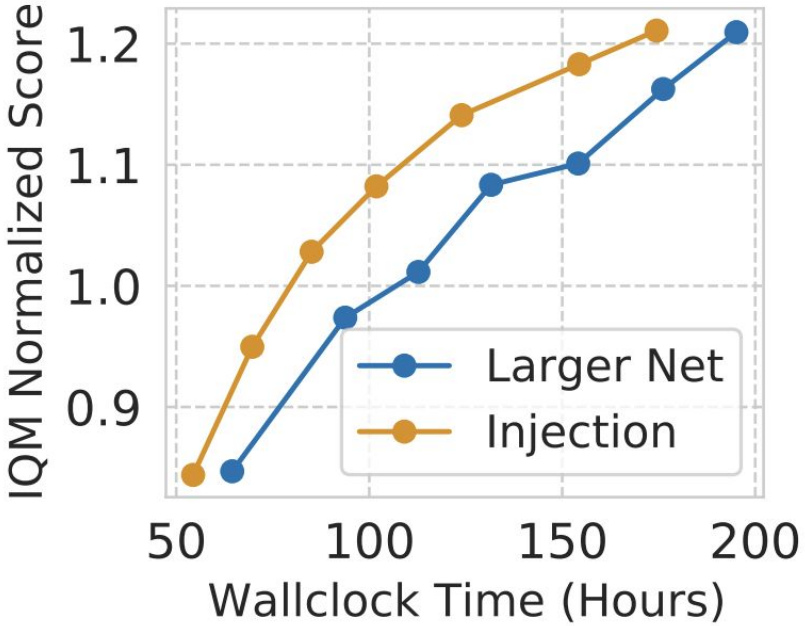
- Plasticity injection @ 50M
- Baseline: larger net





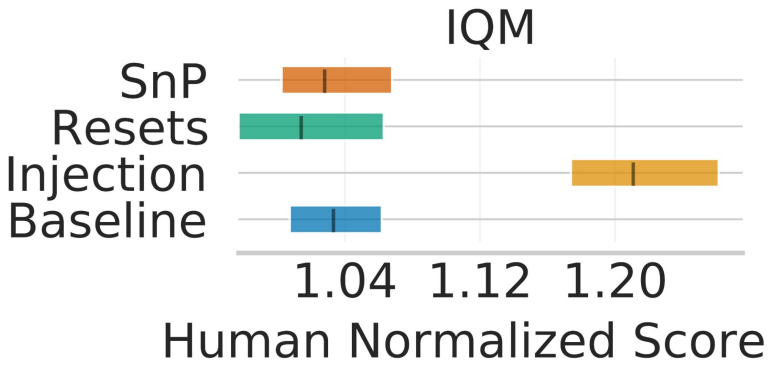
Injection vs Bigger Nets

- Extra plasticity might be unnecessary from the start
- Injection saves compute resources



Injection as a Mechanism to Keep Learning

- Shrink-and-Perturb [Ash 2020]: multiply weights by S, add noise with scale P.
  - Best hparams: S = 1, P = 0.01
- Resets
  - 4M buffer is too small in Atari 200M



Ash and Adams, 2020 “On Warm-Starting Neural Network Training”

Ablations

- Multiple times
- Injection iteration
- Without the 3rd term
- Unfrozen original network
- Adaptive criterion for injecting
- Versions without encoder sharing
- Interventions on weights vs optimizer state

Preventing plasticity loss in the first place

- The need to understand the plasticity loss causes
- Call for rethinking the deep learning foundations of deep RL

Summary

A diagnostic tool

If plasticity injection improves the learning progress, you are experiencing plasticity loss

Dynamic plasticity addition

Extra plasticity from the start might be unnecessary and costly; injection saves computations and allows keeping training