Sample-Efficient Reinforcement Learning by Breaking the Replay Ratio Barrier

Pierluca D'Oro* Max Schwarzer* Evgenii Nikishin Pierre-Luc Bacon Marc G. Bellemare Aaron Courville

Replay Ratio Scaling

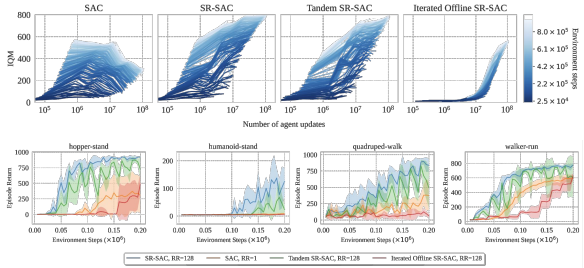
Change in an agent's performance caused by doing more updates for a fixed number of environment interactions

- In principle, intuitive way to be sample-efficient
- In practice, related to performance collapse

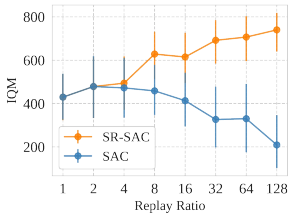
Resets for Replay Ratio Scaling

- The more updates, the more NNs lose ability to learn and generalize (Berariu et al, 2021)
- High replay ratio consumes this ability before enough data is collected
- Resetting at higher frequencies for larger replay ratios restores it

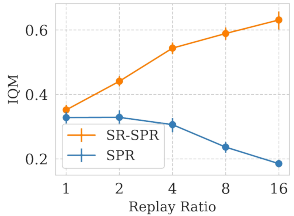
The importance of online RL



DeepMind Control Suite



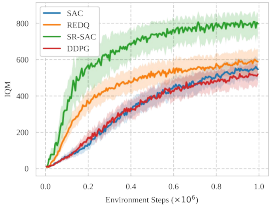
Atari 100k



Resets and replay ratio scaling allow new levels of sample efficiency in model-free RL

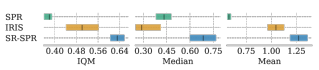
Contributions

- Design simple replay ratio-scalable algorithms based on resets (SR-SAC and SR-SPR)
- Obtain state-of-the-art model-free efficiency
- Analyze requirements for replay ratio scaling
- Study the tradeoffs behind this paradigm

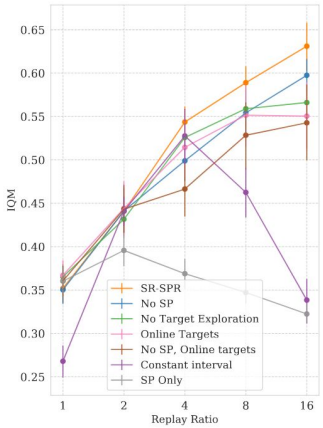


Atari 100k

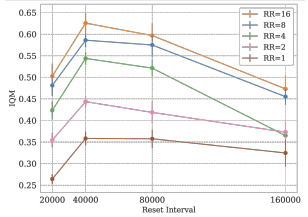
Method	IQM	Median	Mean
SR-SPR	**0.632** (0.60, 0.66)	**0.685** (0.60, 0.77)	**1.272** (1.18, 1.37)
IRIS	0.501 (0.44, 0.56)	0.289 (0.25, 0.41)	1.046 (0.96, 1.13)
SPR	0.380 (0.36, 0.39)	0.433 (0.38, 0.48)	0.578 (0.56, 0.60)
DrQ(ε)	0.280 (0.27, 0.29)	0.304 (0.28, 0.33)	0.465 (0.46, 0.48)
DER	0.183 (0.18, 0.19)	0.191 (0.18, 0.21)	0.351 (0.34, 0.36)



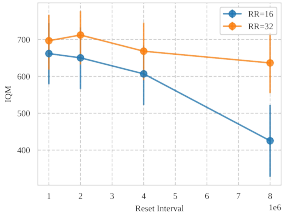
How to get replay ratio scaling?



Atari 100k



DeepMind Control Suite



Dealing with New Tradeoffs