Sample-Efficient Automated Deep Reinforcement Learning

Jörg K.H. Franke[1], Gregor Köhler[2], André Biedenkapp[1], Frank Hutter[1,3]

[1]University of Freiburg, Germany
[2]German Cancer Research Center (DKFZ), Heidelberg, Germany
[3]Bosch Center for Artificial Intelligence, Renningen, Germany

UNI
FREIBURG

Problem Setting

- We want to train an off-policy Reinforcement Learning Agent on a new task.
- We don’t know the optimal hyperparameters.
- We assume, environment interactions are expensive or limited.

Our Approach:

Sample-Efficient Automated Deep Reinforcement Learning (SEARL)

Key idea:

Use an evolutionary routine with shared replay memory

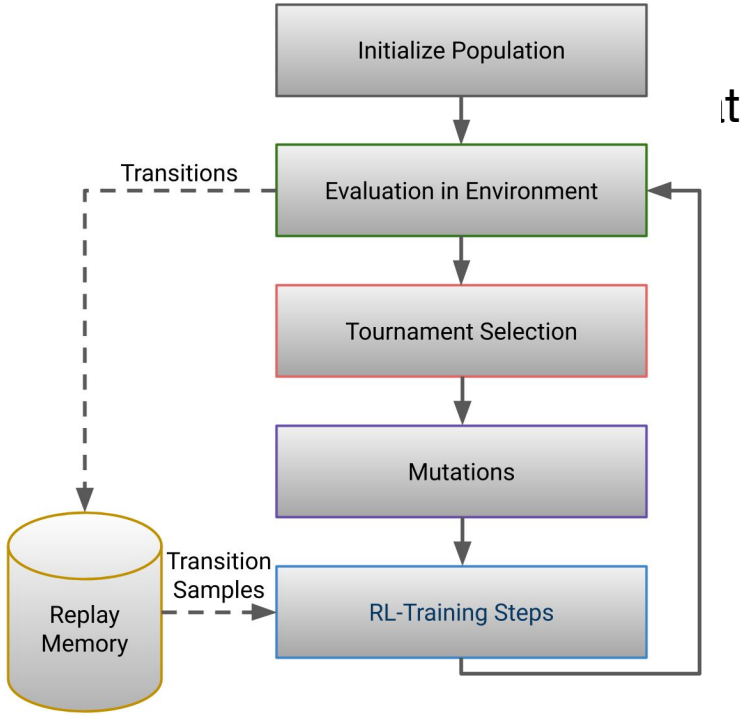
to do RL training and population-based hyperparameter and

neural architecture optimization.

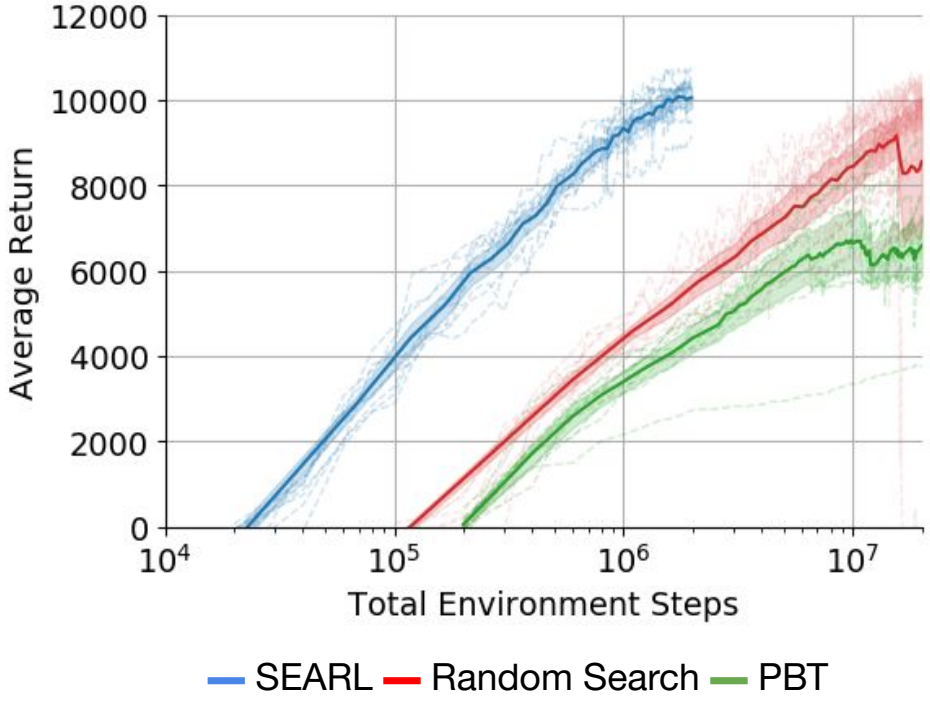
SEARL in a nutshell

- Neuroevolution for RL-training and onli t the same time
- Neuroevolution allows to find the optim
- Agent evaluations fill the shared replay
- Agents are trained using the shared rep

→ Automatically finds optimal HP during
→ Dynamic adaptation instead of fix optir
→ Saves environment interactions

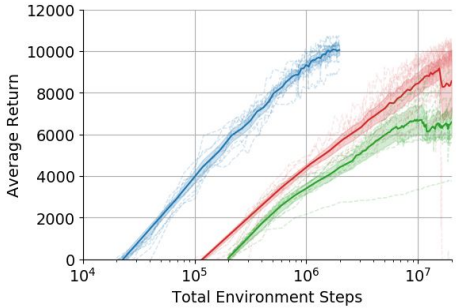


Results - SEARL for TD3 in MuJoCo “HalfCheetah”

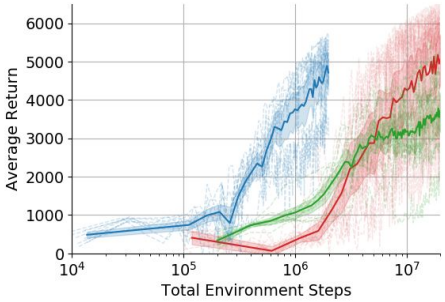


Results - SEARL for TD3 with MuJoCo Suite

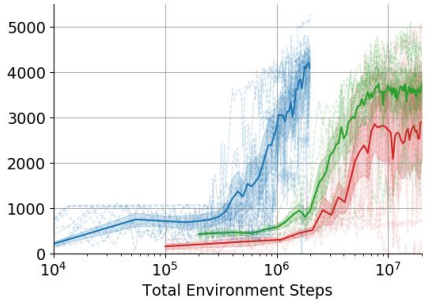
SEARL
Random Search
PBT



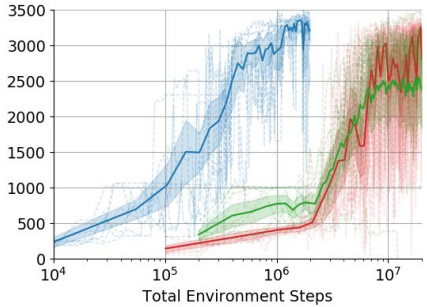
HalfCheetah



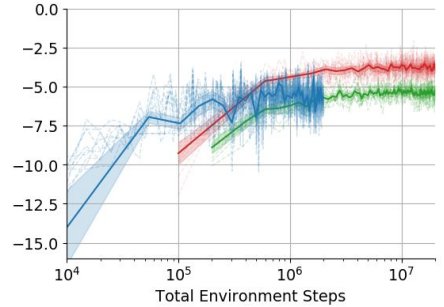
Ant



Walker2d



Hopper

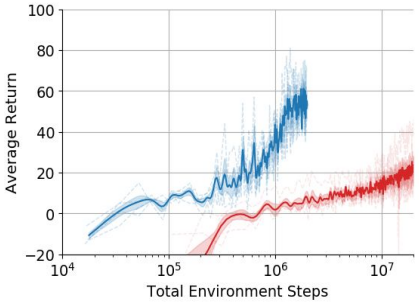


Reacher

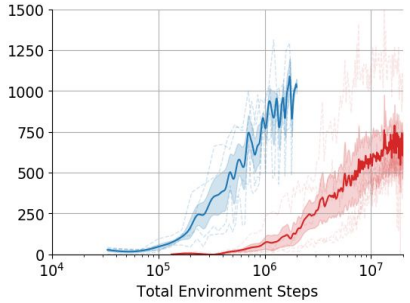
UNI FREIBURG

Results - SEARL for DQN with Atari environments

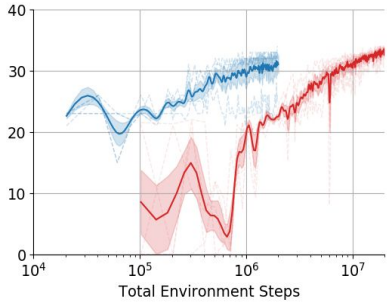
SEARL
Random Search



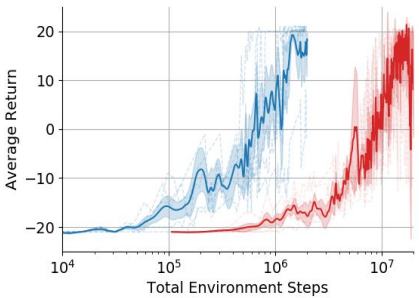
Boxing



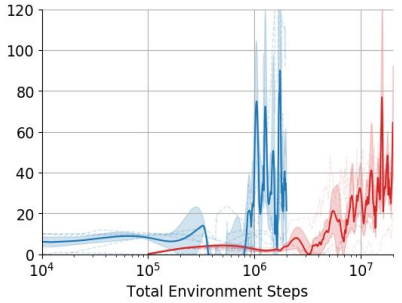
Enduro



Freeway



Pong



RoadRunner

UNI FREIBURG

Conclusion

- SEARL **jointly trains** an RL off-policy agent and **optimizes** the **hyperparameters** including the neural network **architecture**.
- SEARL is based on **neuroevolution** and a **shared replay memory**.
- The population in SEARL benefits from a **diverse set of experience** in the shared replay memory.
- SEARL requires **up to ten times fewer environment interactions** as random search or Population Based Training (PBT) and nearly the same amount of environment steps as a normal RL agent training.

Thanks for your attention!

Paper

Code

Session 4
04.05.21

Poster