

RL experiment - widowx reach-v3

1 Introduction and methods

The aim of this document is plot the results of the following RL experiment:

- Algorithms: A2C, ACKTR, DDPG, PPO2, SAC, TD3, TRPO
- Environment: widowx reach-v3
- Number of time steps: 1M
- Number of initialisation seeds: 2
- Number of parallel environments: 8 for ACKTR and PPO2 and 1 for SAC and TD3 (parallelisation not supported).

The performance metrics are defined as follows:

- Train time (min) : Wall time to train.
- Success ratio : number of successful episodes / number of reachable episodes
An episode is successful if the distance between the finger tip and the target is less than or equal to 0.03.
- Average reaching time : sum (number of time steps of all successful episodes) / number of successful episodes
An episode has a maximum of 100 time steps.
- Efficiency: mean reward / mean training walltime.

2 Results

2.1 Raw results

2.2 Learning curves

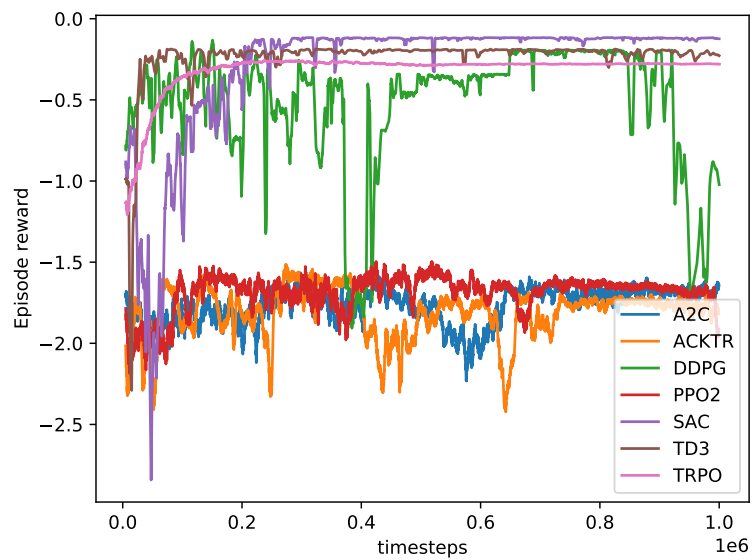


Figure 1: All learning curves.

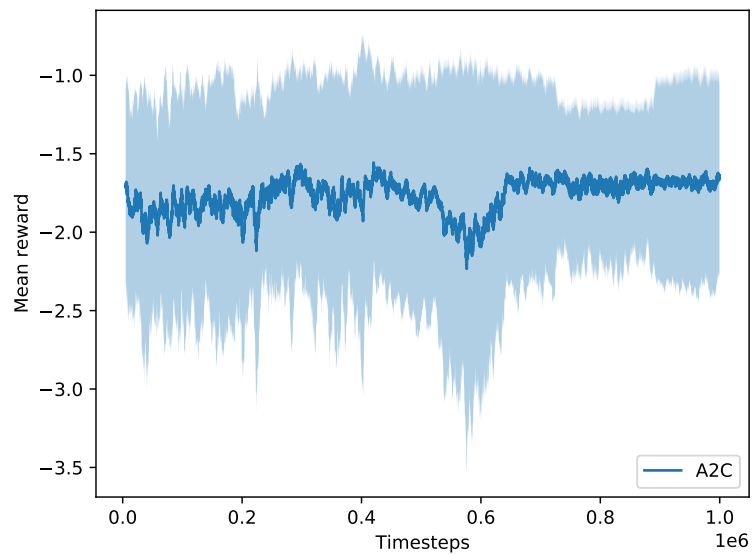


Figure 2: Learning curve A2C.

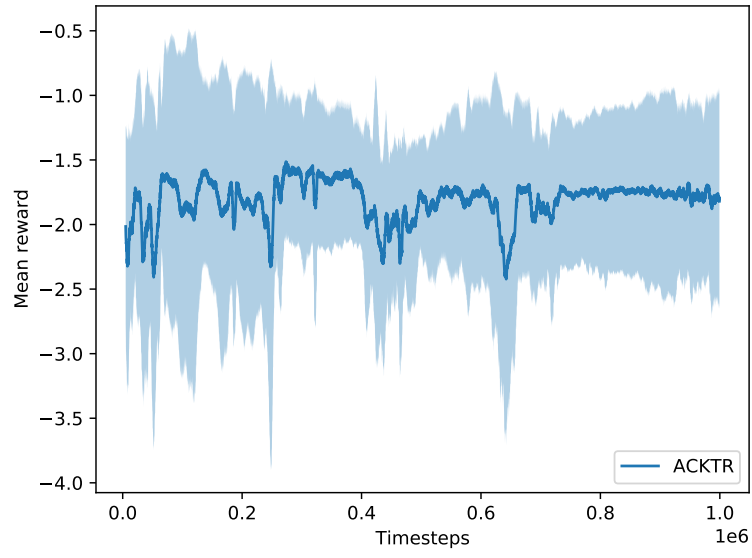


Figure 3: Learning curve ACKTR.

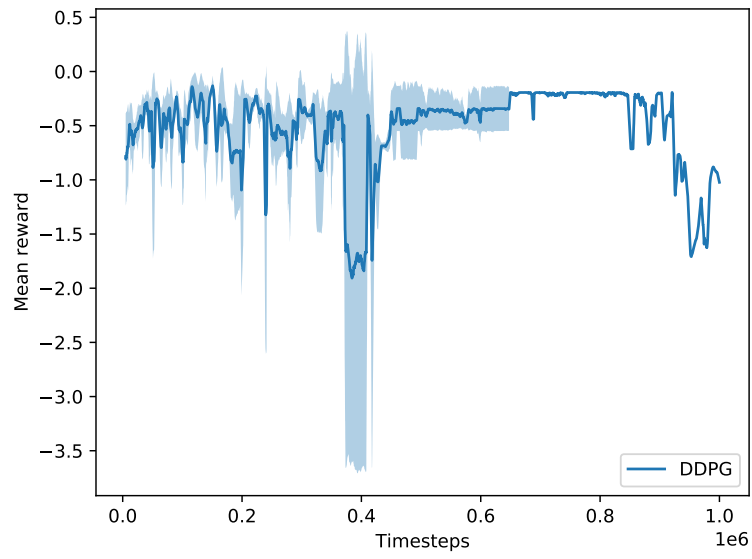


Figure 4: Learning curve DDPG.

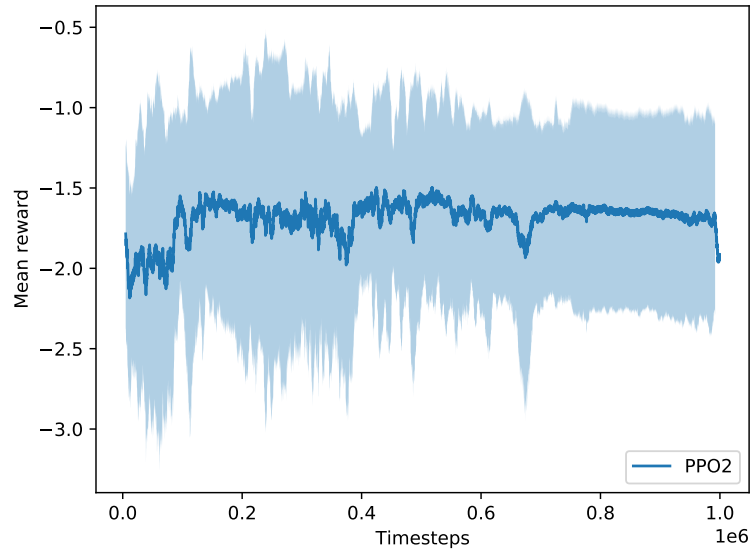


Figure 5: Learning curve PPO2.

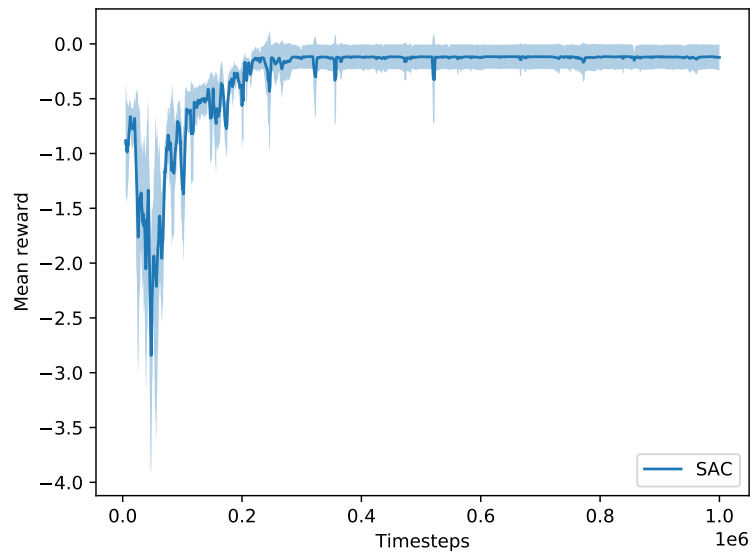


Figure 6: Learning curve SAC.

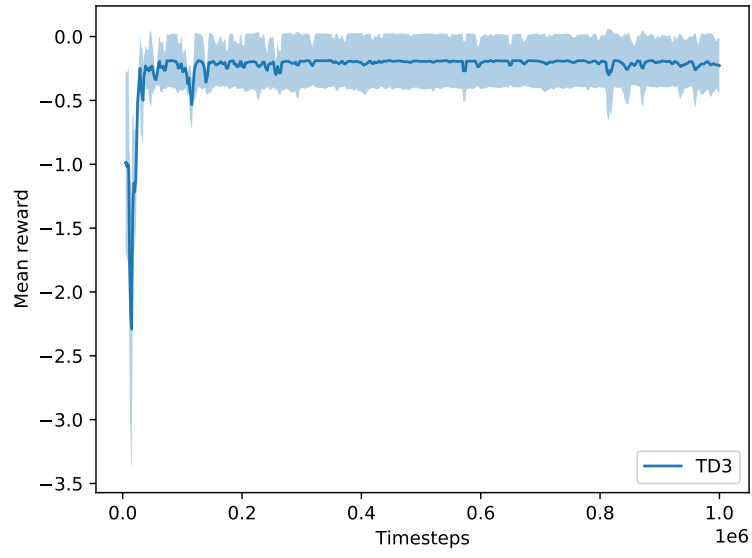


Figure 7: Learning curve TD3.

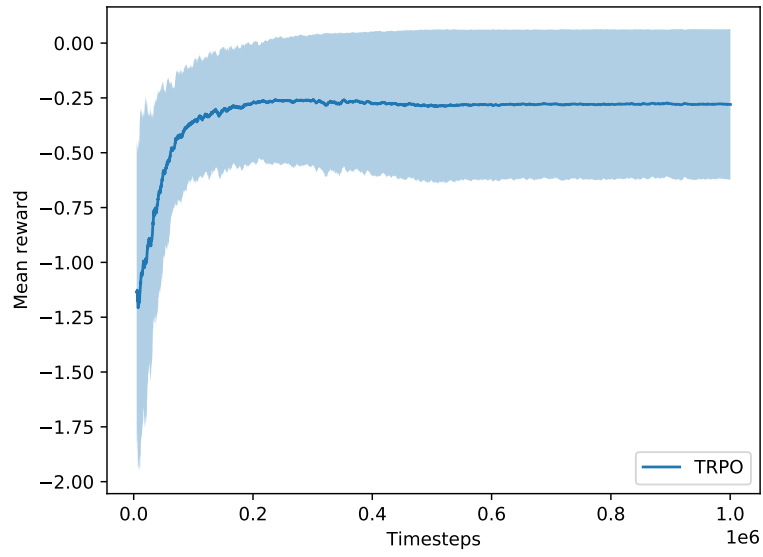


Figure 8: Learning curve TRPO.

2.3 Evaluation

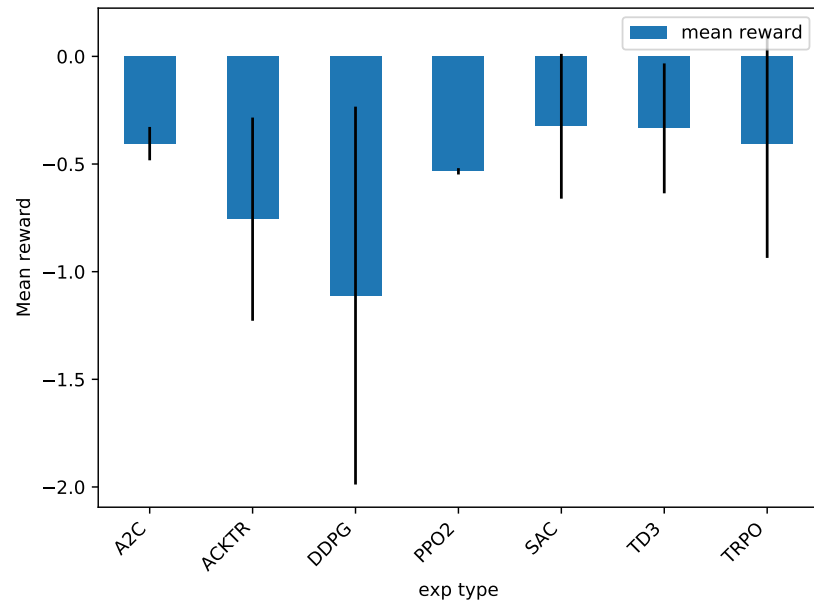


Figure 9: Reward vs algorithms.

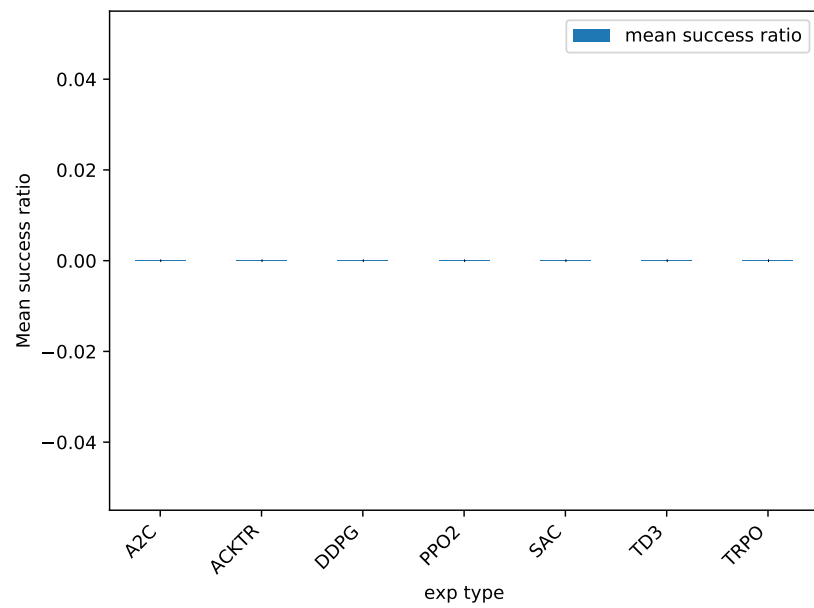


Figure 10: Success ratio vs algorithms.

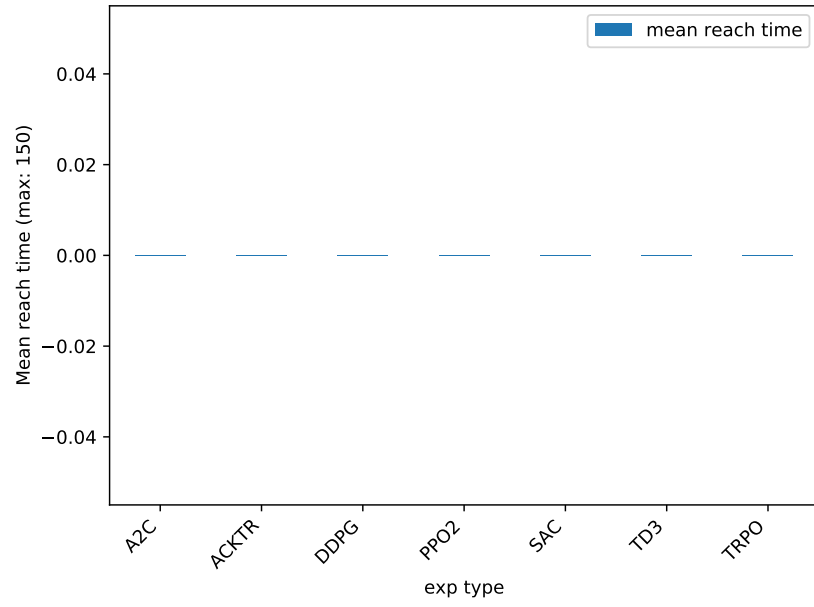


Figure 11: Reach time vs algorithms.

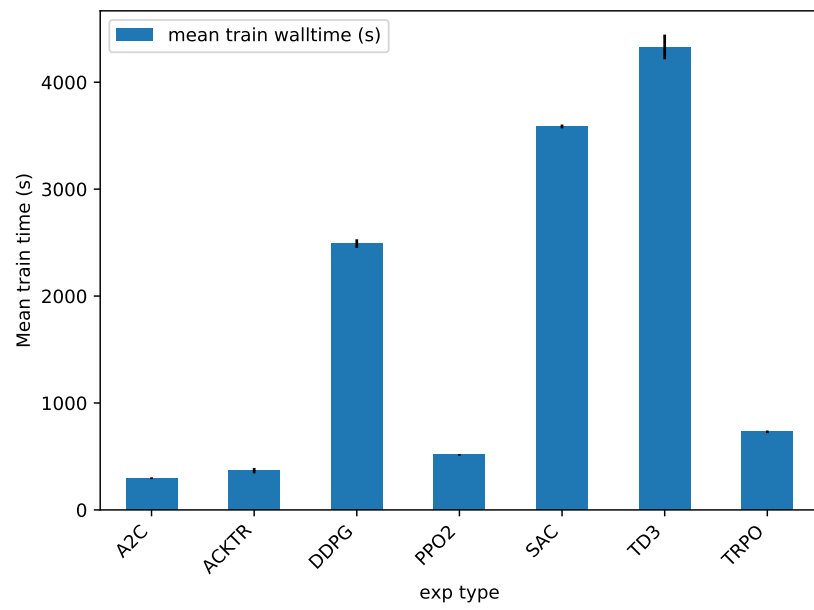


Figure 12: Train walltime vs algorithms.

3 Findings summary