

COMP 767 Project - Gym Car Racing



Bhavya Patwa & Yifan Bai



Deep RL + Policy Gradient

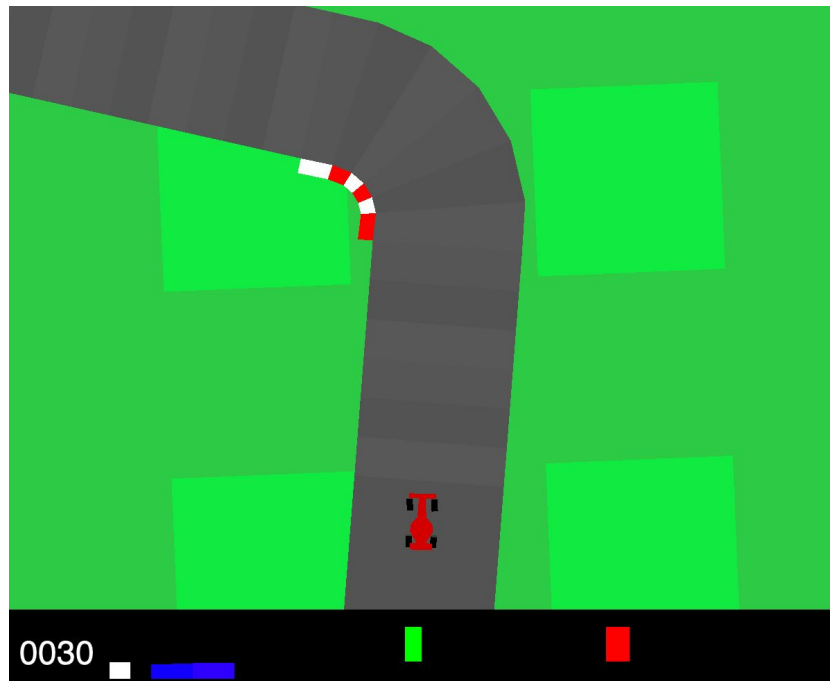
On-policy: PPO

Off-policy: DDPG, TD3

Continuous Control

OpenAI Gym CarRacing-v0

1. 96-96-3
2. [steer, brake, throttle]
3. -0.1/frame, 1000/N(tiles)
4. Finish all tiles
5. -100 if go off track; terminate





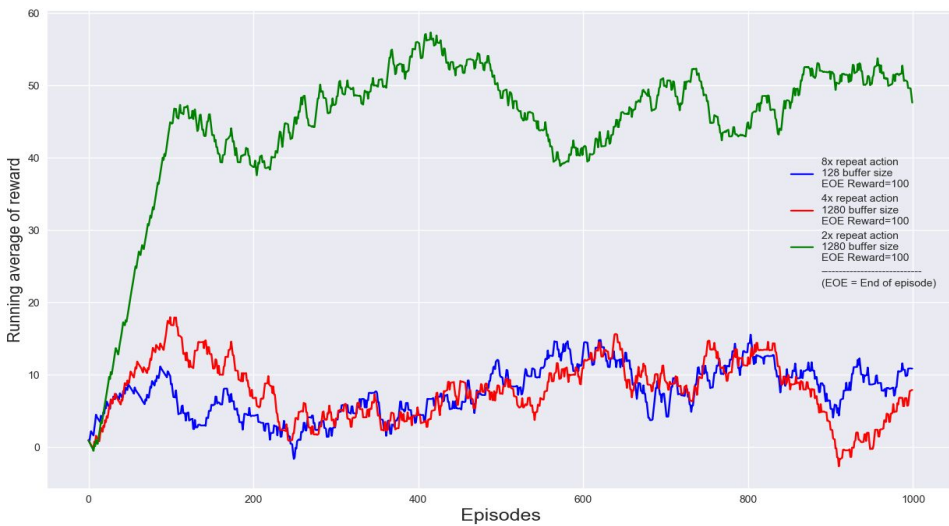
Experiments

1. Resource restricted, only run 1000 max episode (usually 100k+)
2. Different hyperparameter sets, tweak shared ones for direct comparison

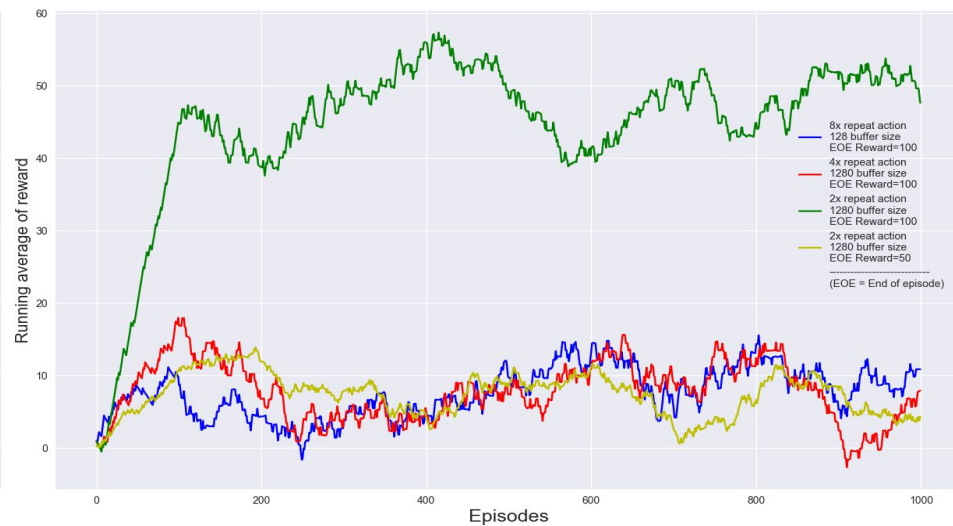


PPO

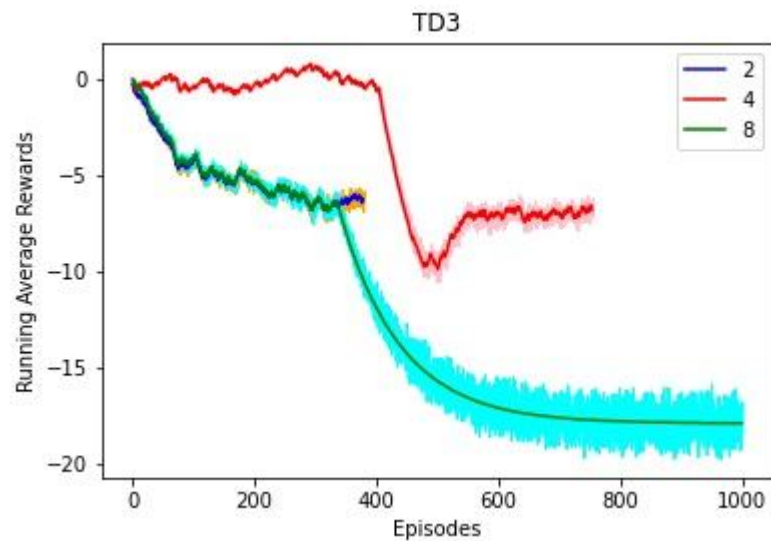
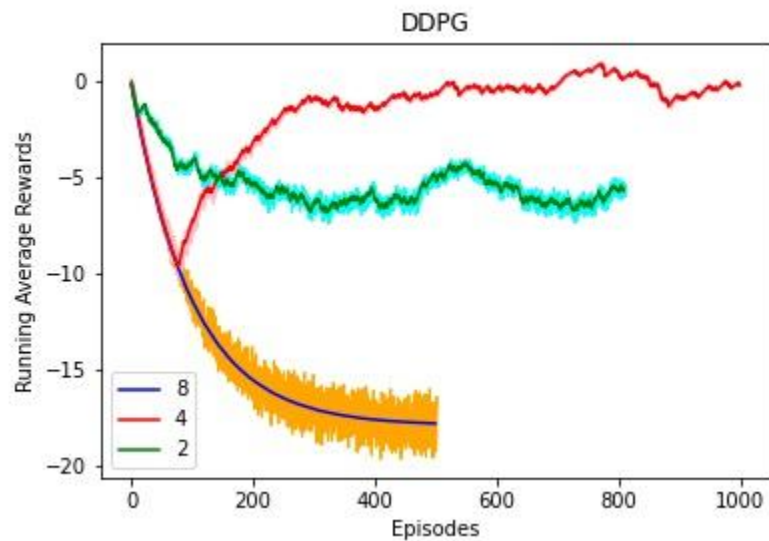
Learning curves using PPO

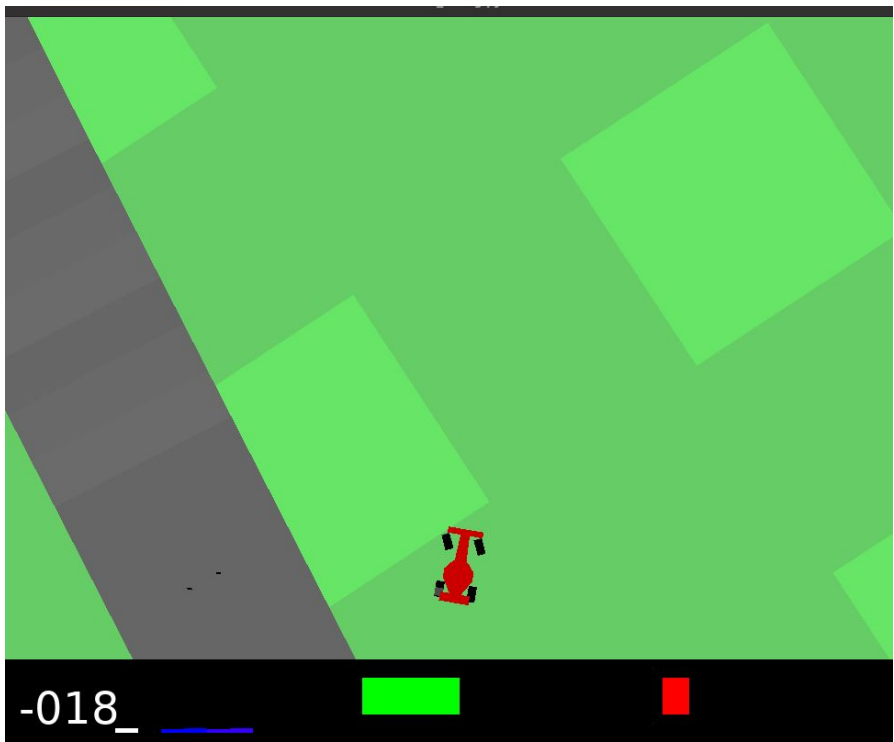


Reward curves on CarRacing-v0 using PPO

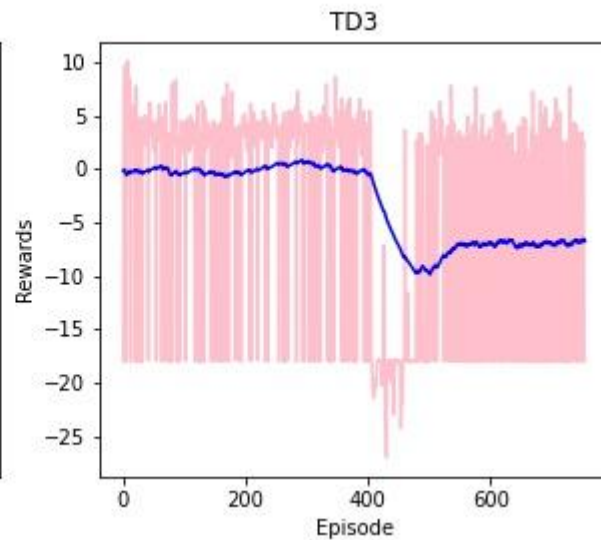
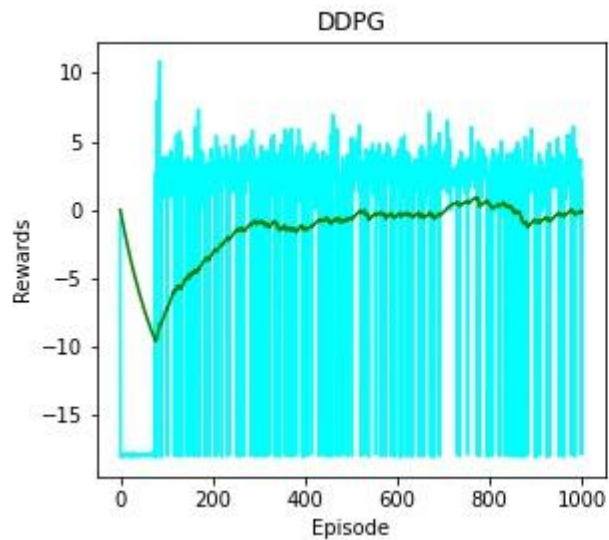


DDPG/TD3





DDPG/TD3





Conclusion and Future Work

- PPO performed well, but lots of noise
 - DDPG/TD3 handicapped by Q-function, can't find policies
-
- Run more episodes and hyperparameter search/architecture improvements
 - Feature engineering action space
 - Add additional reward conditions