APPLIED DEEP LEARNING

GROUP 5

Landing Lunar Modules

Reinforcement Learning

GROUP MEMBER

Feng Jiaxu

Zeng Shaoyu

Ziya Shaheer

Zhang Xichen

Methods & Techniques

POLICY GRADIENT PRO

Adaptive Baseline Policy Gradient

- Modify the final reward by reward subtracting the baseline score
- Lower variance & converges faster compared with Policy Gradient but less stable performance

Advantage Actor Critic (A2C)

- Two networks, actor network and critic network.
- Less random variables and lower variance, but need to balance variance and bias

Proximal Policy Optimization (PPO)

- Clip the policy update in a trust region thus prevent large deviation
- More stable, converge faster, higher score

DQN PRO

Deep Q Network (DQN)

- Use Neural Networks to approximate Q(s,a) function when action space is large
- Experience Replay Buffer to reduce temporal correlation in learning, with greedy epsilon search

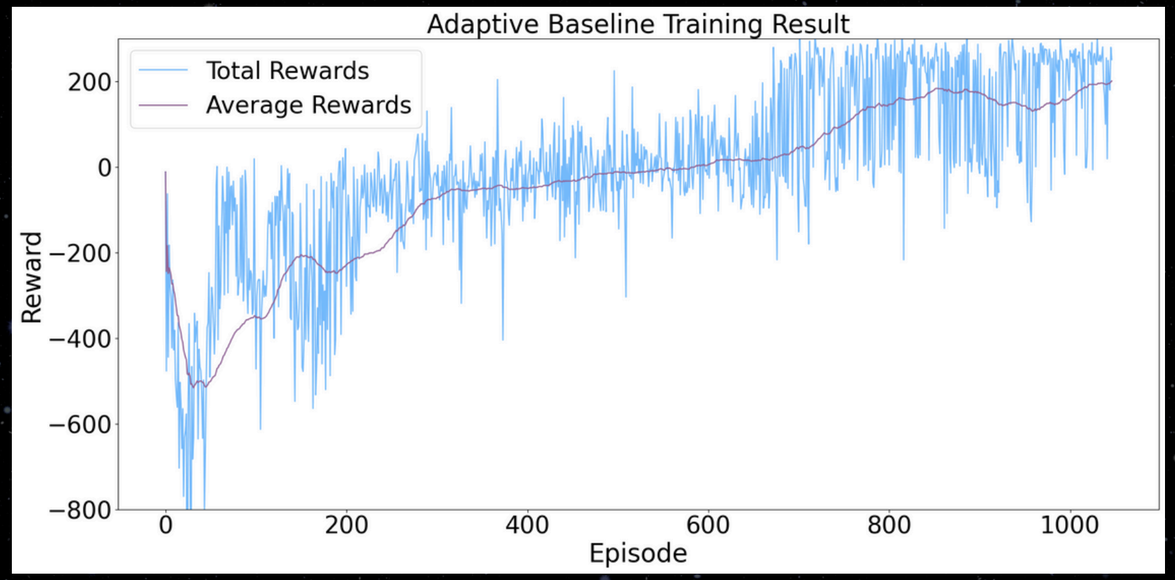
Double DQN

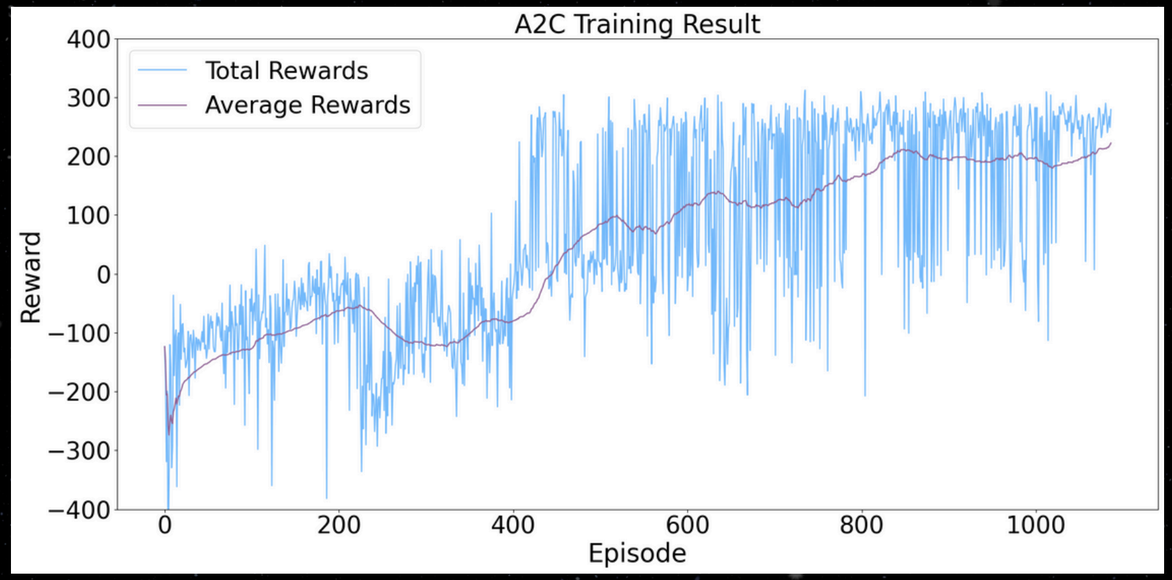
- Two networks, action value estimation and target network
- DQN overestimates Q-function, DDQN fixes a target making it more stable, but slower at learning

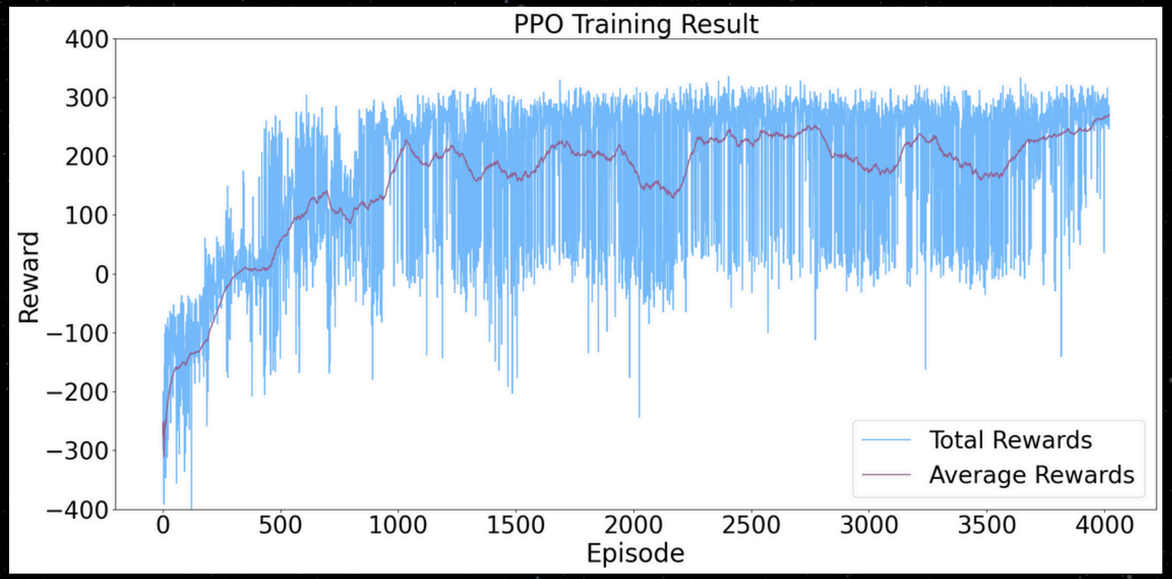
Duelling DQN

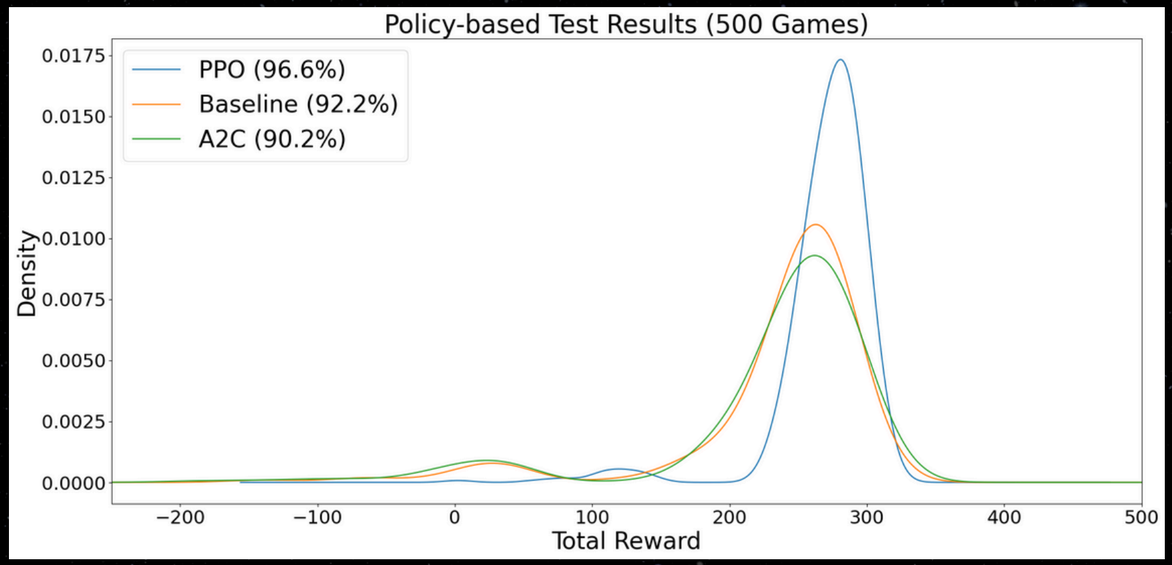
- Two networks, action selection and action evaluation
- DQN overestimates Q-function, D2QN fixes by sampling Stable learning and performance

Results (Policy Gradient Pro X)

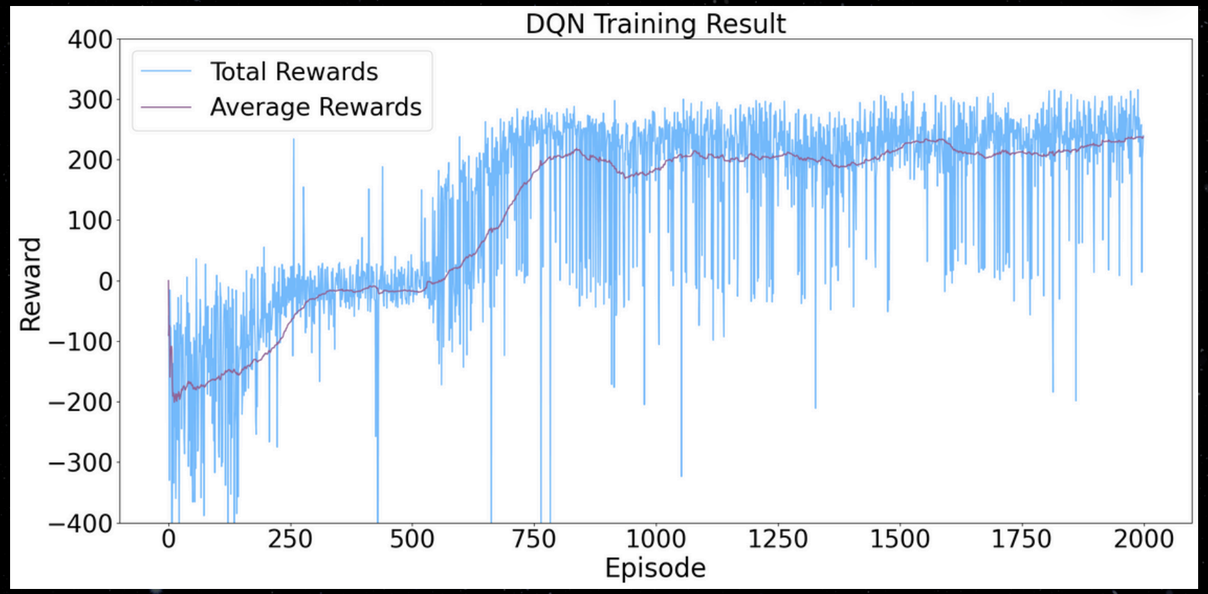


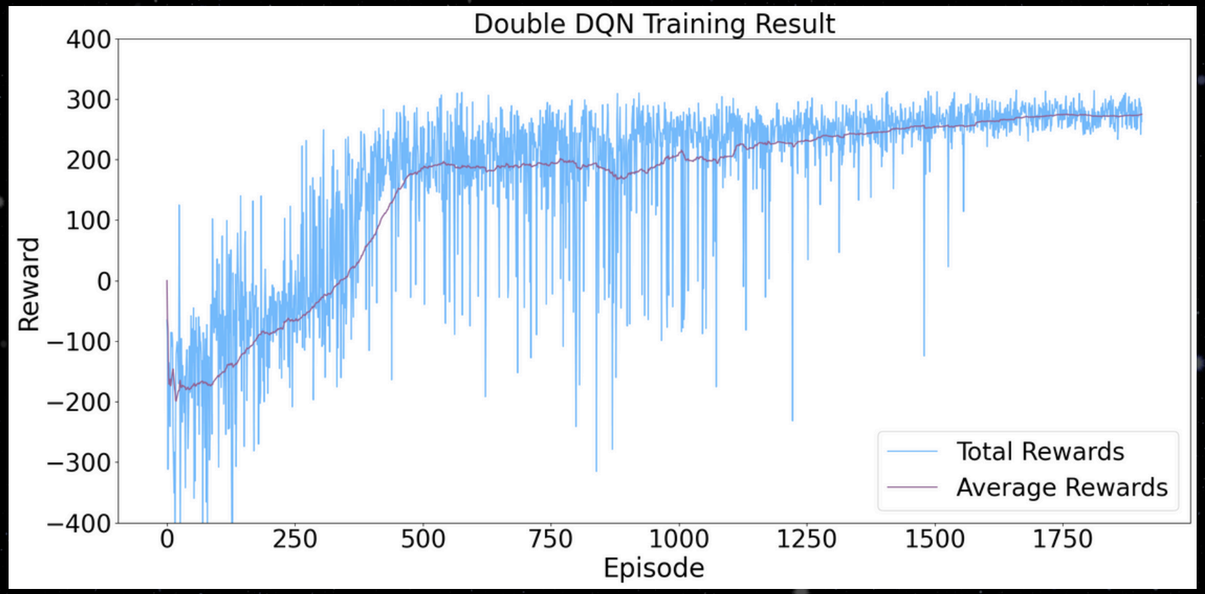


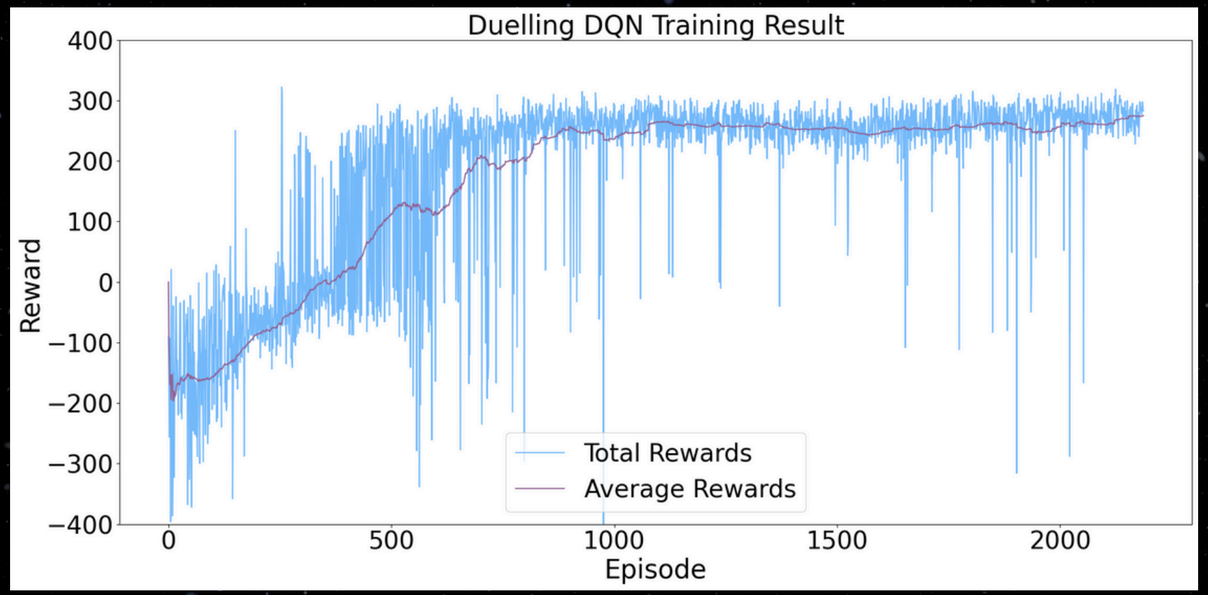


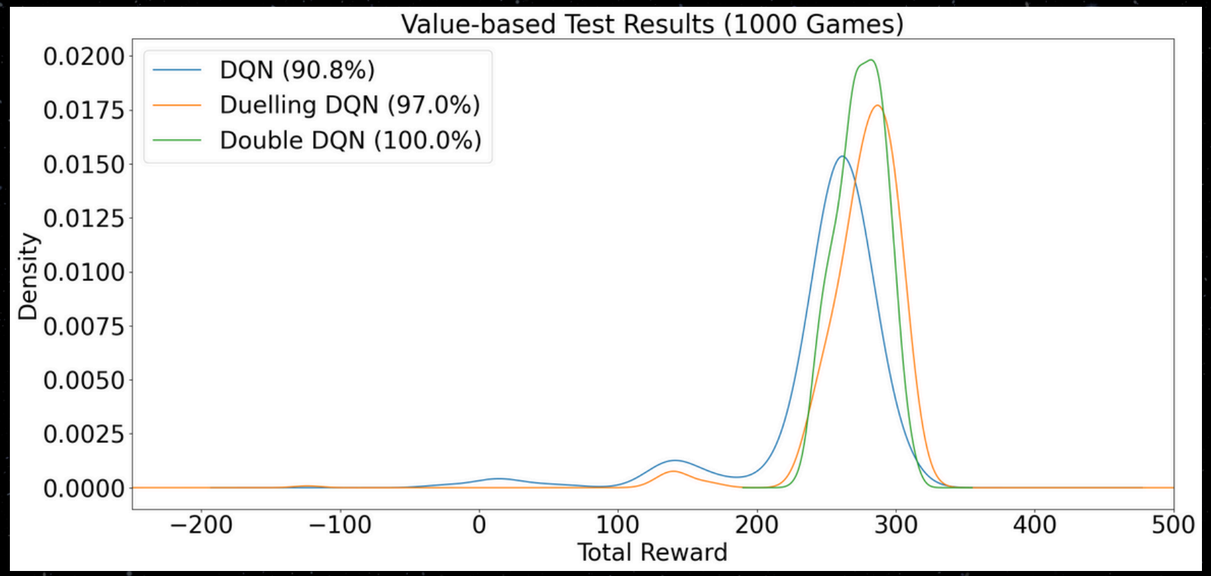


Results (DQN Pro)









Challenges & Solutions

Convergence & Stability

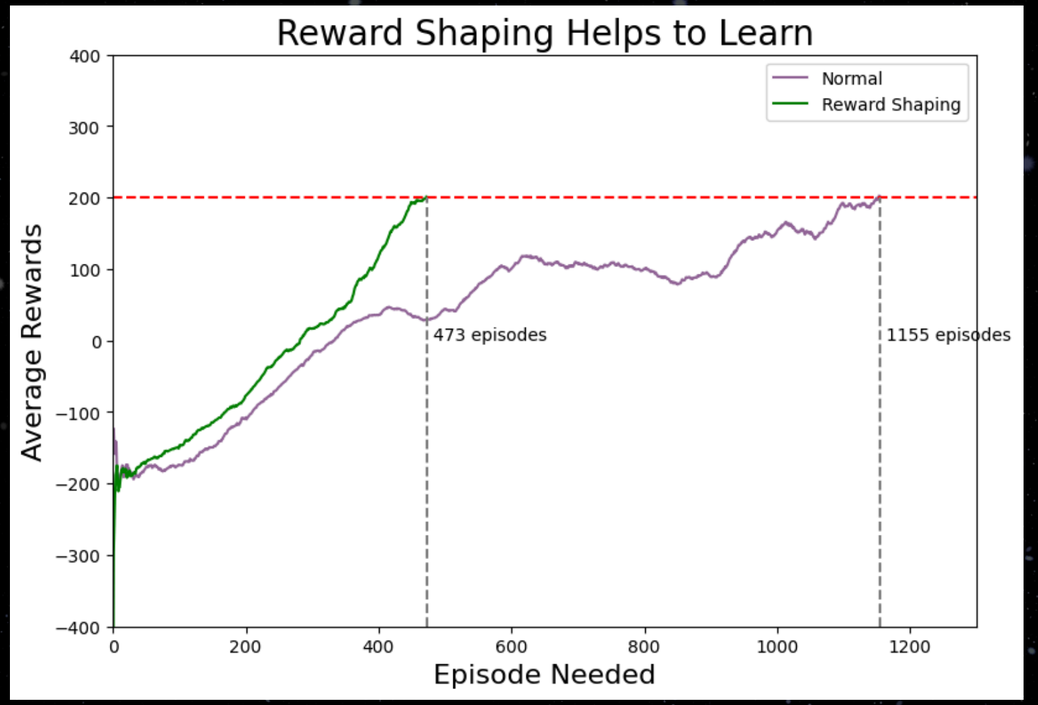
Speed up convergence and learning stability with Reward Shaping, modify rewards of the environment to encourage the agent to remain close to the center

Policy Gradient Convergence

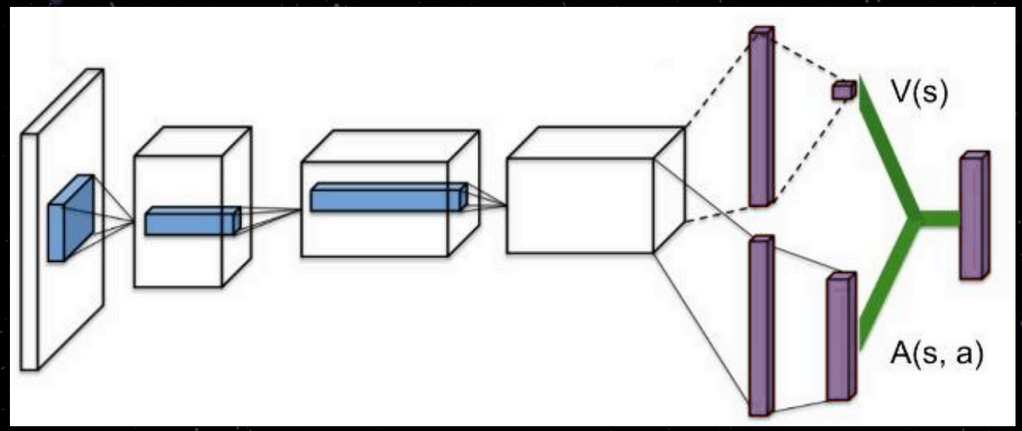
Issues with convergence due to changing environment with adaptive baseline. Fix the environment to train and test on random environments.

Duelling Architectures

Introducing too many hidden layers in the duelling network structure causes a drop in performance

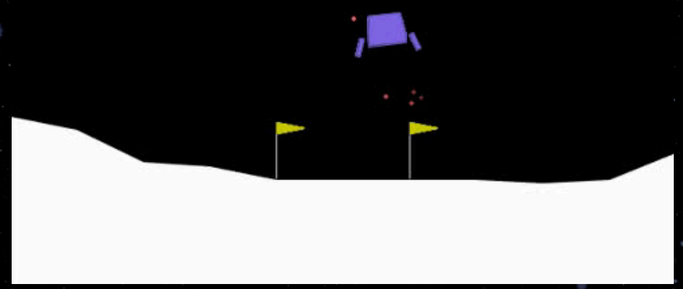


(Top) Training a PPO model with and without reward shaping
(Bottom) Network Architecture of Duelling DQN

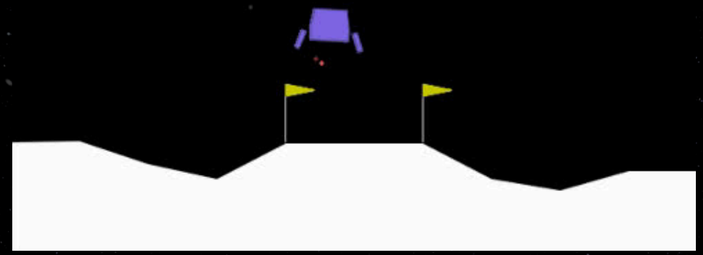


Demo

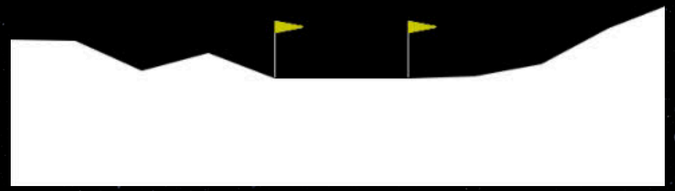
Baseline



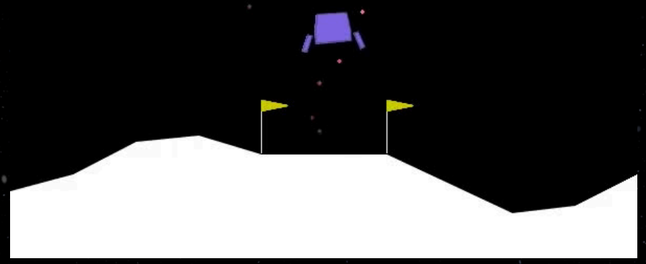
A2C



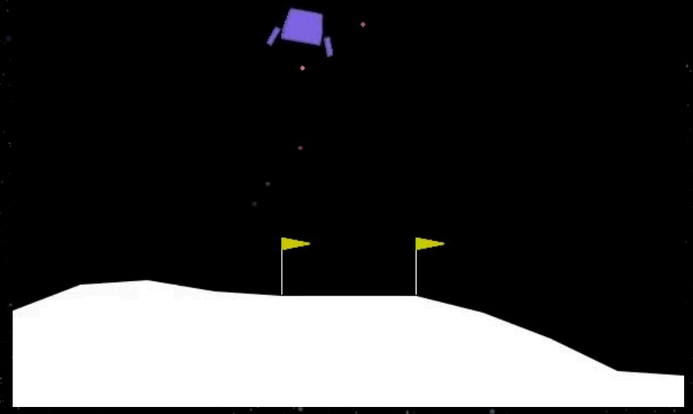
PPO



DQN



Double DQN



Duelling DQN