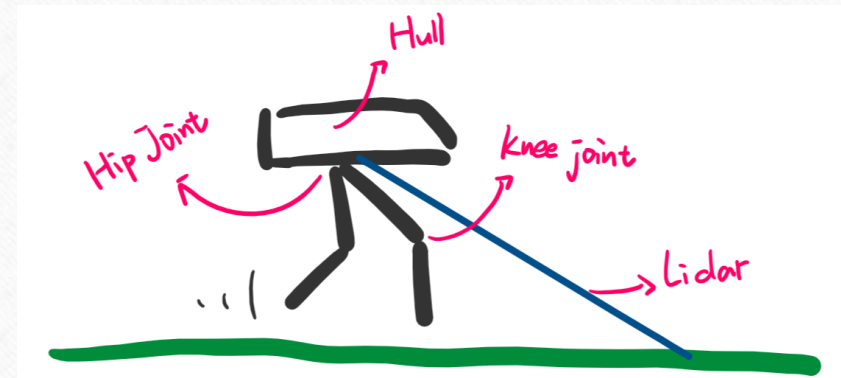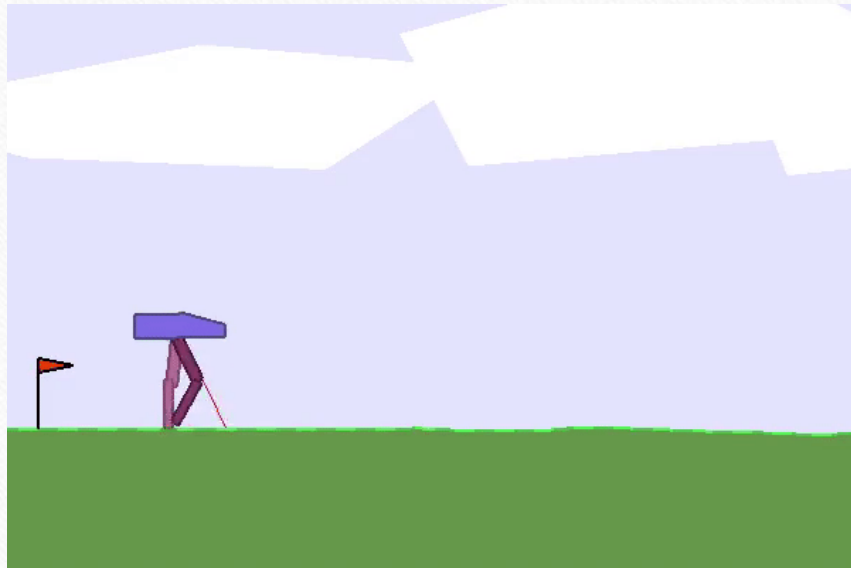# CS 5180 Final Project

- Bipedal Walker -

Presented by

Meishan Li, Zhida Zhang

Project Problem

BipedalWalker-v3:

The main problem to be solved in this project is to train the robot to choose the correct action to stand up and start walking a certain distance in an environment with terrain that is not perfectly horizontal.
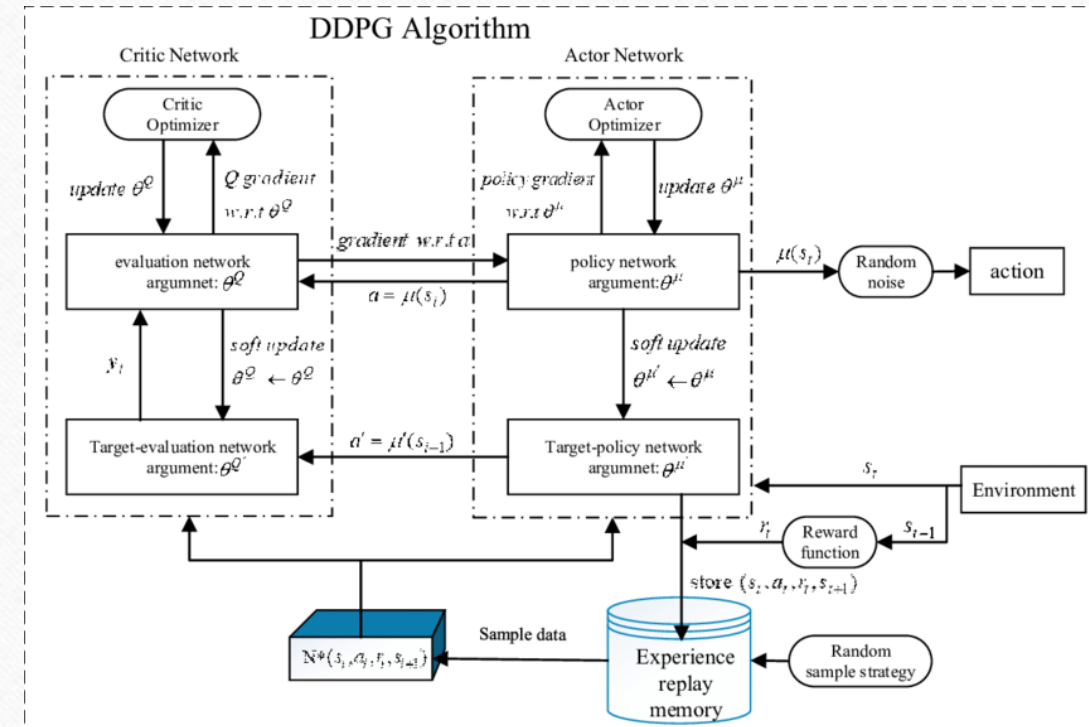
Project Approach/Algorithms

- **Deep deterministic policy gradient algorithm(DDPG)**

  DDPG is a reinforcement learning algorithm based on the AC(Actor-Critic) framework, which maximizes the total reward by predicting a deterministic policy by updating the policy in a single step through the off-policy method.
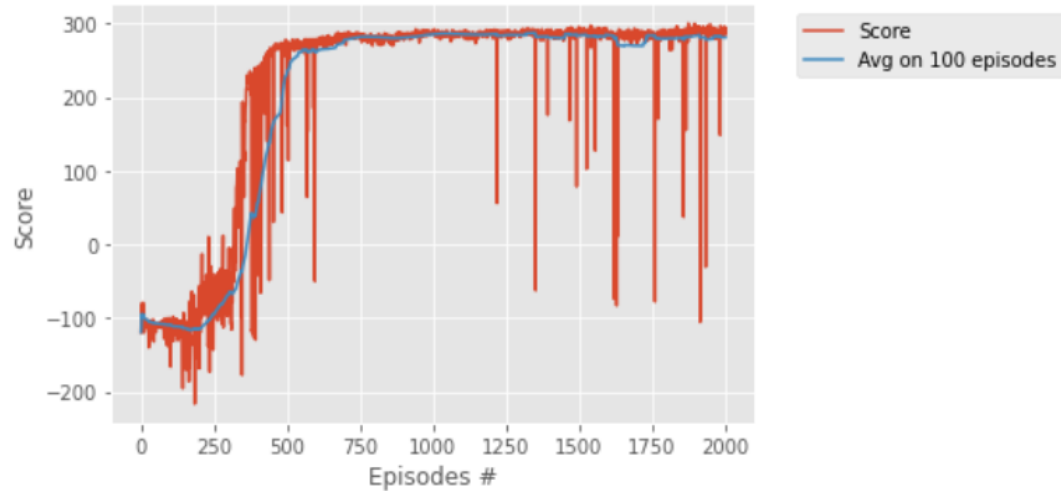
- **Twin Delayed DDPG(TD3)**

  In TD3, two sets of Critic networks are used to represent different Q values and to suppress persistent overestimation by selecting the smallest one as the Target Q Value.

DDPG Algorithm

Project Experiments/comparisons/results

length of scores: 2000 , len of avg_scores: 2000

Results of TD3 algorithm

```
100%|████████████| 1990/2000 [3:14:36<00:55,  5.59s/it]

Ep. 1991, Timestep 1903222,  Ep.Timesteps 885, Score: 288.01, Avg.Score: 281.06, Time: 03:14:36

100%|████████████| 1991/2000 [3:14:41<00:50,  5.61s/it]

Ep. 1992, Timestep 1904107,  Ep.Timesteps 885, Score: 291.61, Avg.Score: 281.10, Time: 03:14:42

100%|████████████| 1992/2000 [3:14:47<00:44,  5.62s/it]

Ep. 1993, Timestep 1904981,  Ep.Timesteps 874, Score: 293.97, Avg.Score: 281.12, Time: 03:14:48

100%|████████████| 1993/2000 [3:14:53<00:39,  5.61s/it]

Ep. 1994, Timestep 1905901,  Ep.Timesteps 920, Score: 281.15, Avg.Score: 281.05, Time: 03:14:53

100%|████████████| 1994/2000 [3:14:59<00:34,  5.69s/it]

Ep. 1995, Timestep 1906805,  Ep.Timesteps 904, Score: 288.70, Avg.Score: 280.94, Time: 03:14:59

100%|████████████| 1995/2000 [3:15:04<00:28,  5.72s/it]

Ep. 1996, Timestep 1907668,  Ep.Timesteps 863, Score: 294.81, Avg.Score: 281.00, Time: 03:15:05

100%|████████████| 1996/2000 [3:15:10<00:22,  5.65s/it]

Ep. 1997, Timestep 1908553,  Ep.Timesteps 885, Score: 289.06, Avg.Score: 280.98, Time: 03:15:11

100%|████████████| 1997/2000 [3:15:15<00:16,  5.66s/it]

Ep. 1998, Timestep 1909412,  Ep.Timesteps 859, Score: 295.10, Avg.Score: 281.04, Time: 03:15:16

100%|████████████| 1998/2000 [3:15:21<00:11,  5.61s/it]

Ep. 1999, Timestep 1910319,  Ep.Timesteps 907, Score: 284.18, Avg.Score: 280.98, Time: 03:15:22

100%|████████████| 1999/2000 [3:15:27<00:05,  5.66s/it]

Ep. 2000, Timestep 1911208,  Ep.Timesteps 889, Score: 288.87, Avg.Score: 280.90, Time: 03:15:27

100%|████████████| 2000/2000 [3:15:32<00:00,  5.87s/it]
```
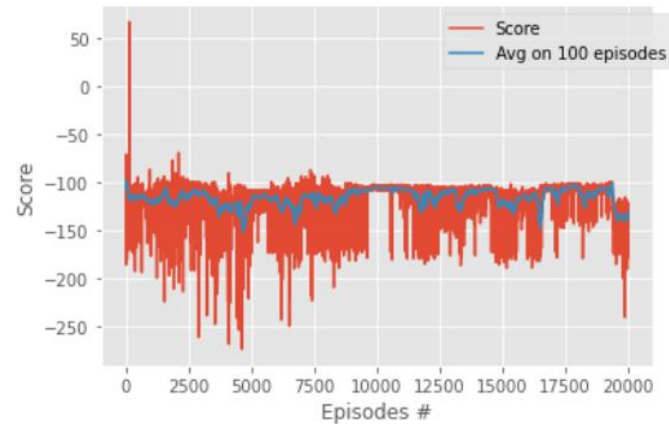
Project Experiments/comparisons/results

Only one critic network(1)

Only one critic network(2)