CSE 546 FINAL PROJECT

Solving reinforcement algorithms on environments under uncertainty

Team 22: Antarpreet Kaur, Kanupriya Pandey, Veda Sai Rochishna Eli

05/07/2023

University at Buffalo The State University of New York

Project Overview

- The project explores performance of multiple popular algorithms on Gym or Gymnasium environments under uncertainty.
- The gymnasium discrete space environments are deterministic implying that an agent's actions have consistent effects on the environment's state. The project augments the gymnasium environments with stochastic factors such as aerodynamic disturbances, such as wind noise and random engine failures.
- The objective of introduction of noise is the environments is to incorporate a degree of unpredictability into the agent's decision-making algorithm, thereby rendering the environment more demanding and less deterministic.

Project Overview

- The project gymnasium environment modified by adding noise are **Cat Pole v1** environment which is part of the classic control environments, and **Lunar Lander v2** which is part of the Box2d environments.
- The algorithms evaluated are **Deep Q-Learning Network (DQN)**, **Double Deep Q-Learning Network (DDQN)**, **Advantage Actor Critic (A2C)**, and **Asynchronous Advantage Actor Critic (A3C)**.

Implementation – Cart Pole Environment

- Wind effects are applied to the Cart Pole environment to emulate real-world scenario and add uncertainty to the environment.
- At every step wind can blow in a random direction of left or right enabling the cartpole to move a step based on the wind direction.

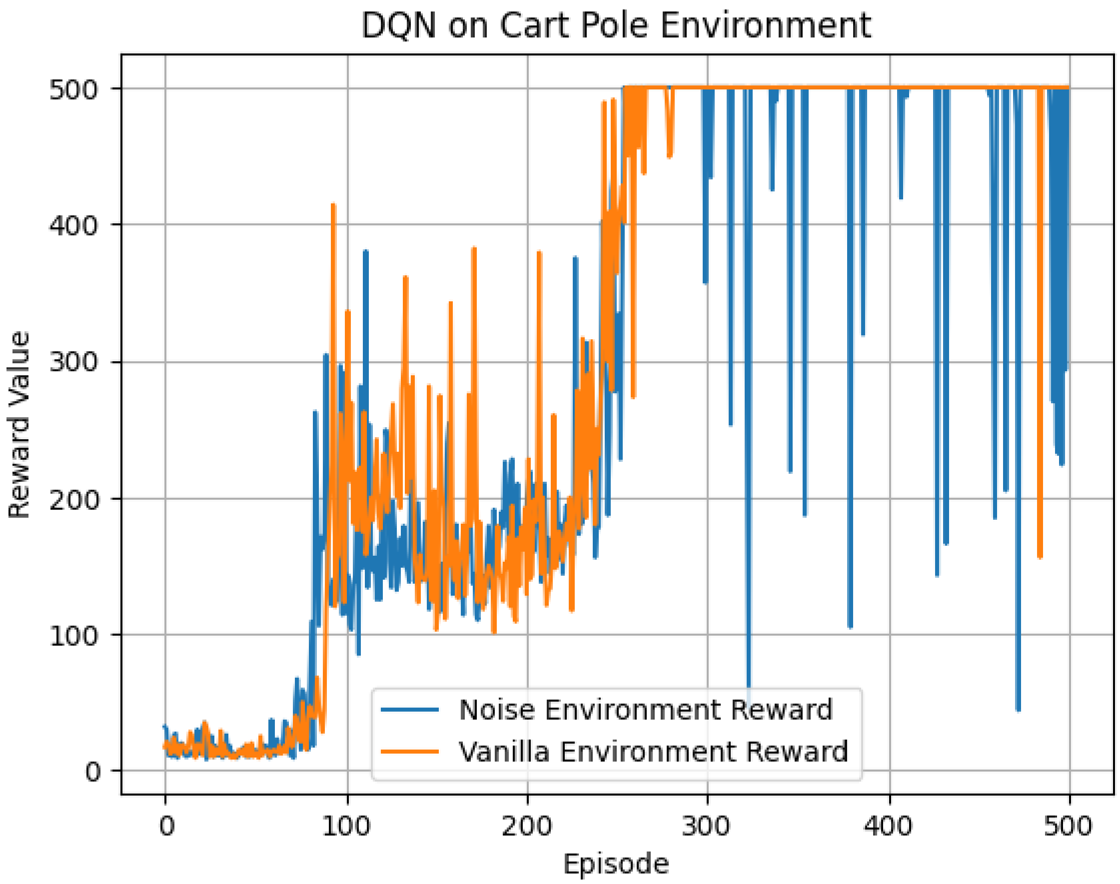
```python
# Wind noise function
def wind_noise(action):
  wind_direction = np.random.randint(0, 3)
  double_action = False

  if wind_direction == 0:  # no wind
    wind_action = action

  # Wind blows in the left direction and left action is taken
  elif wind_direction == 1 and action == 0:
    wind_action = 0  # left action
    double_action = True

  # Wind blows in the right direction and right action is taken
  elif wind_direction == 2 and action == 1:
    wind_action = 1  # Right action
    double_action = True

  else:
    wind_action = action # no wind_action

  return double_action, wind_action
```

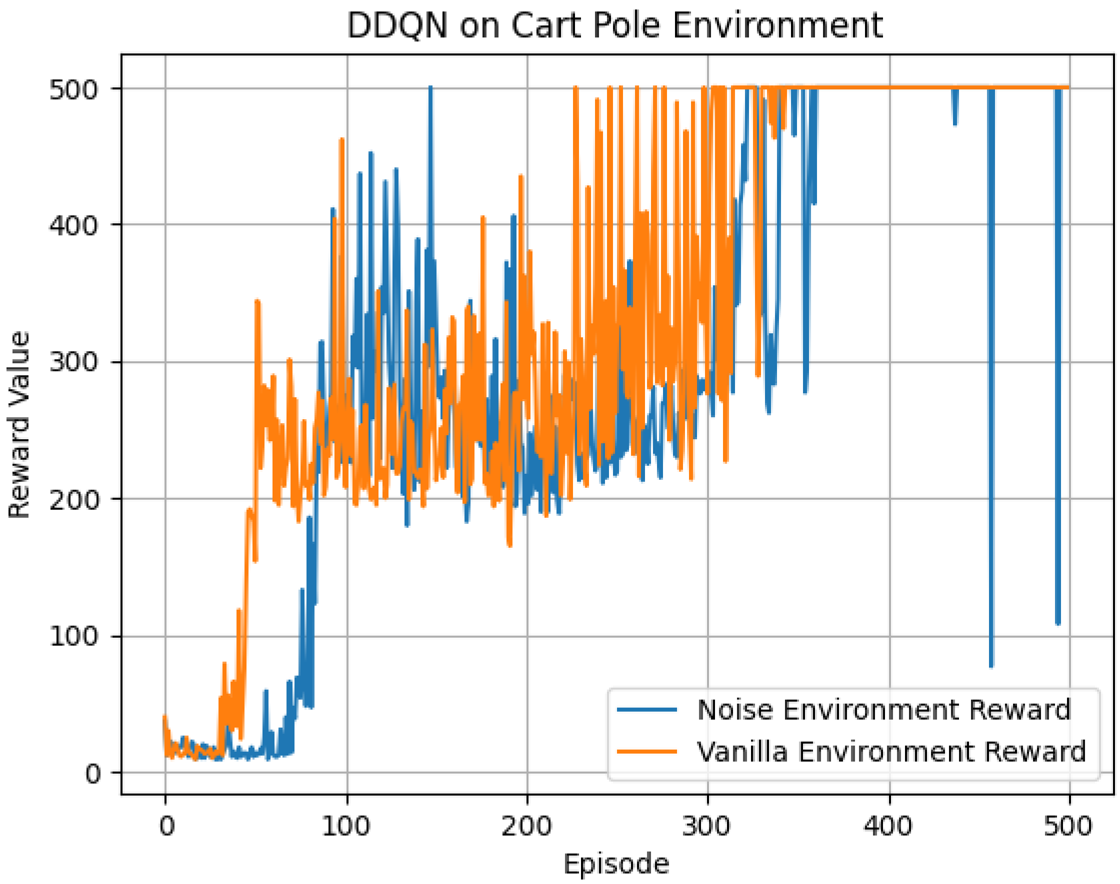
Implementation – Lunar Lander Environment

- Lunar lander is a more complex environment which enables more customization for adding uncertainty in the environment.
- Engine failure and wind effects generated by effects such as solar winds or random particles in the space are applied to the Lunar Lander environment to emulate real-world scenario and add uncertainty to the environment.

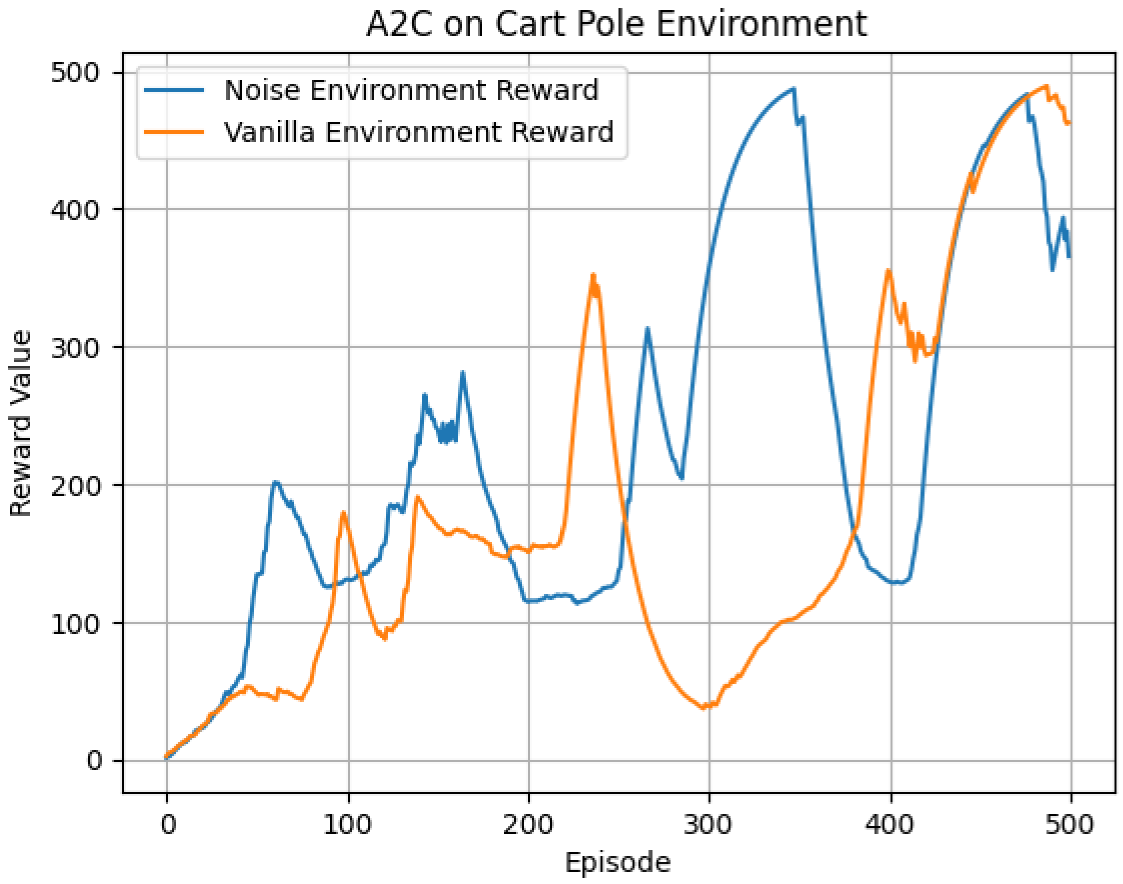
```
threshold = 0.1 # 10% engine failure, 90% engine functions properly
# Engine Failure Function
def engine_failure(threshold):
    if np.random.random() <= threshold:
        return True
    return False

# Wind Noise Function
def wind_noise(action):
  wind_direction = np.random.randint(0, 3)
  double_action = False

  if wind_direction == 0: # no wind
    wind_action = action

  elif wind_direction == 1: # wind blows in left direction
    if action == 3:
      wind_action = 0 # no action
    else:
      wind_action = 1 # left action
      double_action = True

  elif wind_direction == 2: # wind blows in right direction
    if action == 1:
      wind_action = 0 # no action
    else:
      wind_action = 3 # right action
      double_action = True

  return double_action, wind_action
```

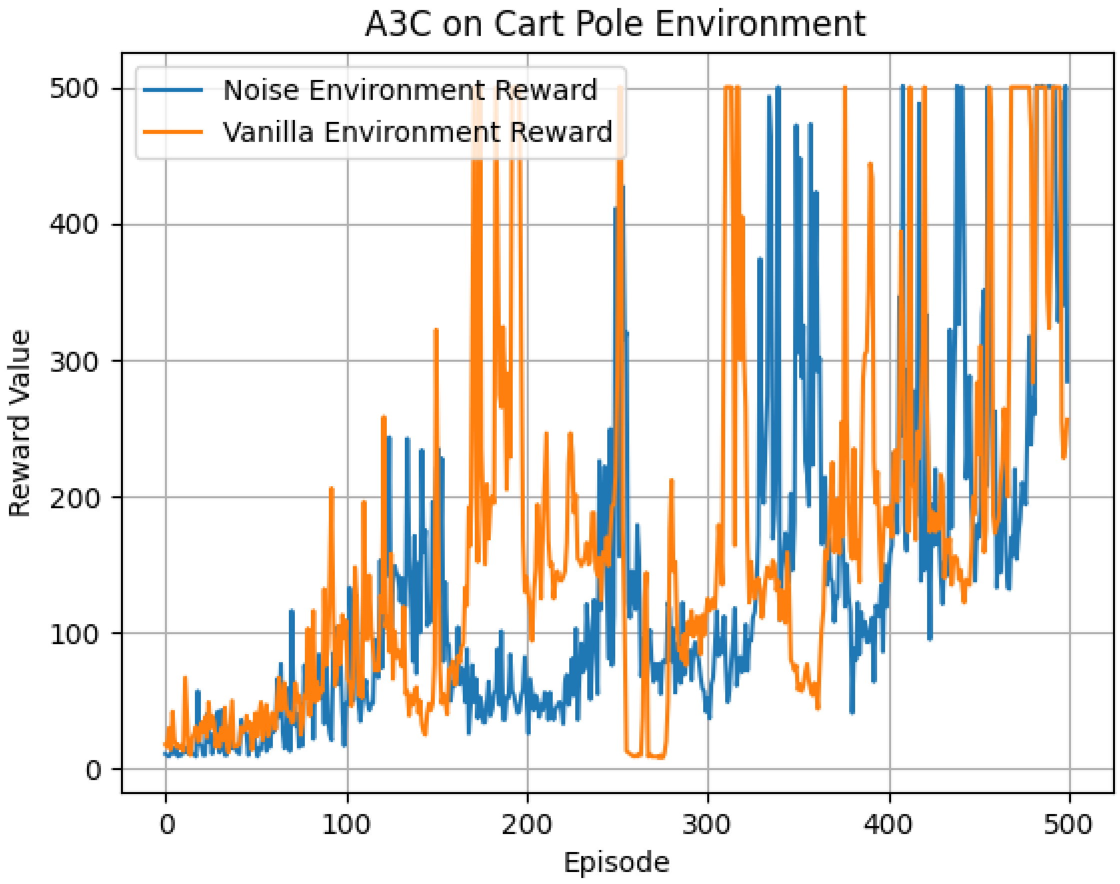
Results - Cart Pole vanilla vs noise environment rewards



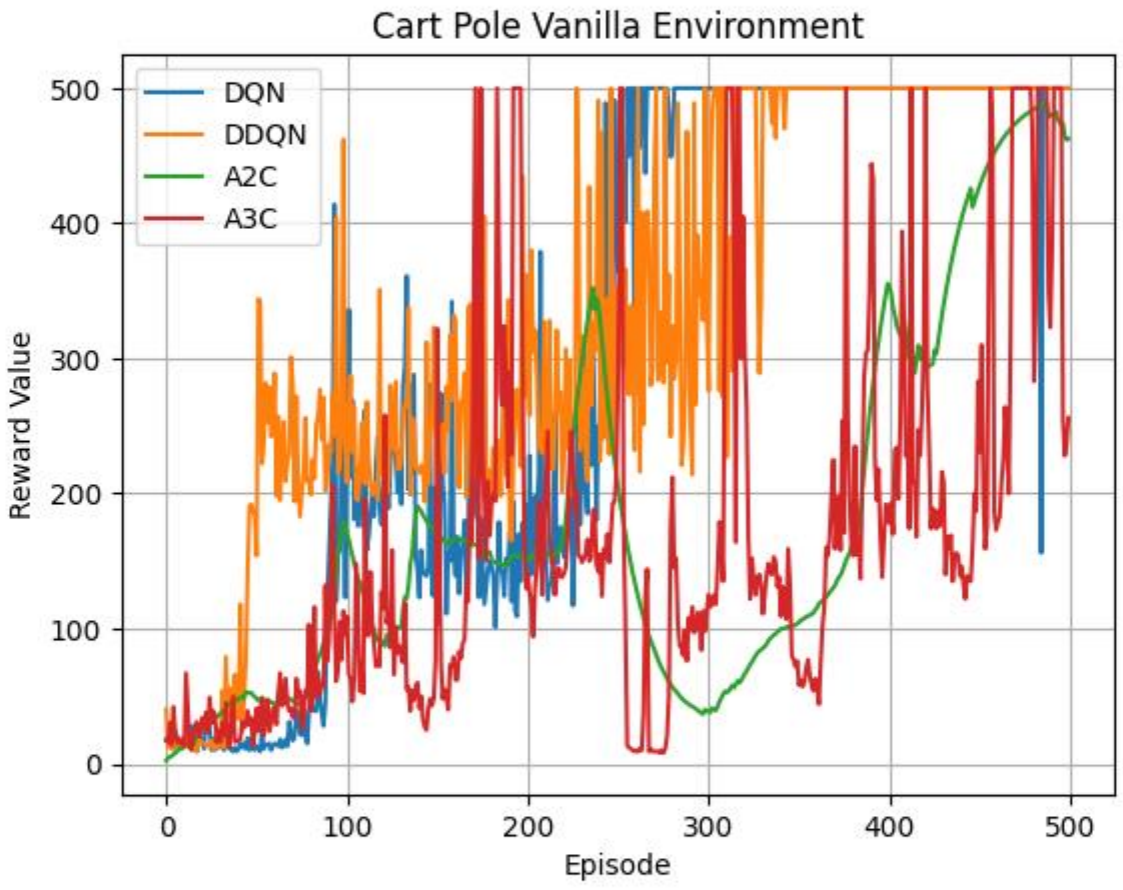


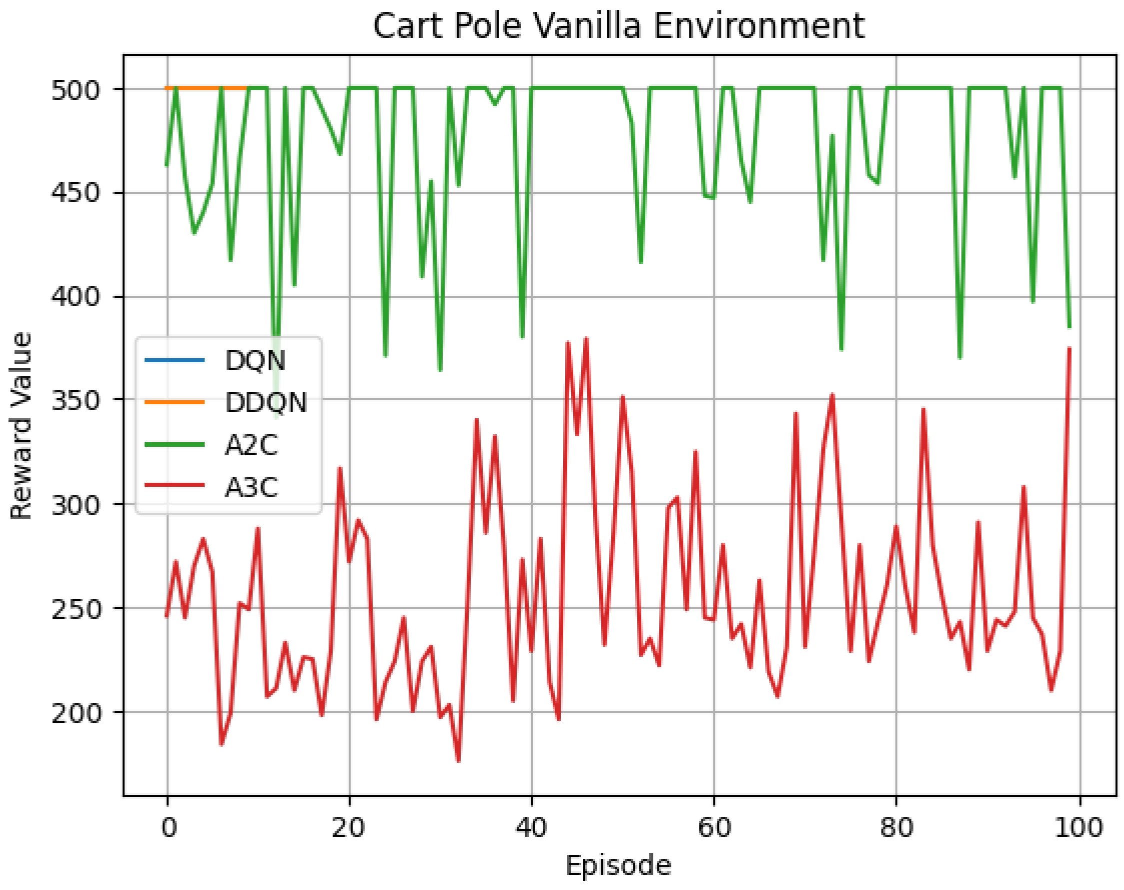
Results - Cart Pole vanilla vs noise environment rewards



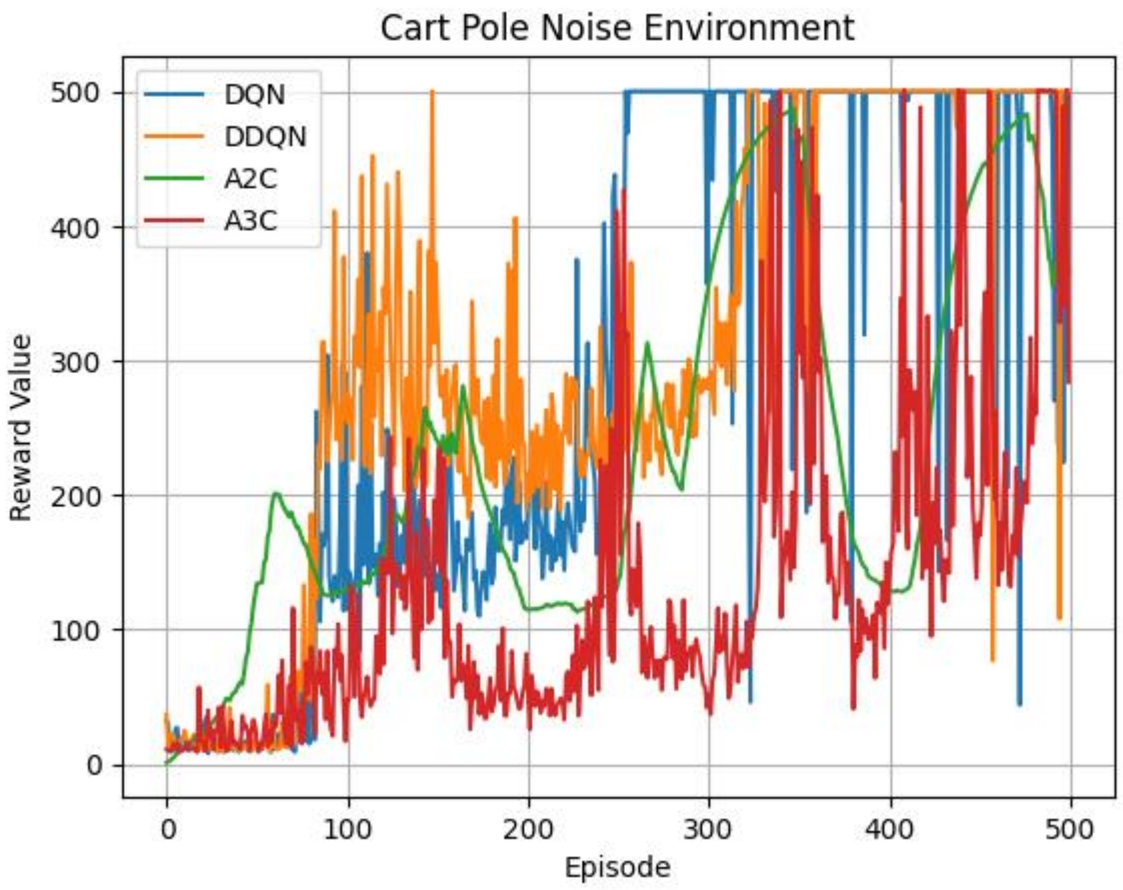


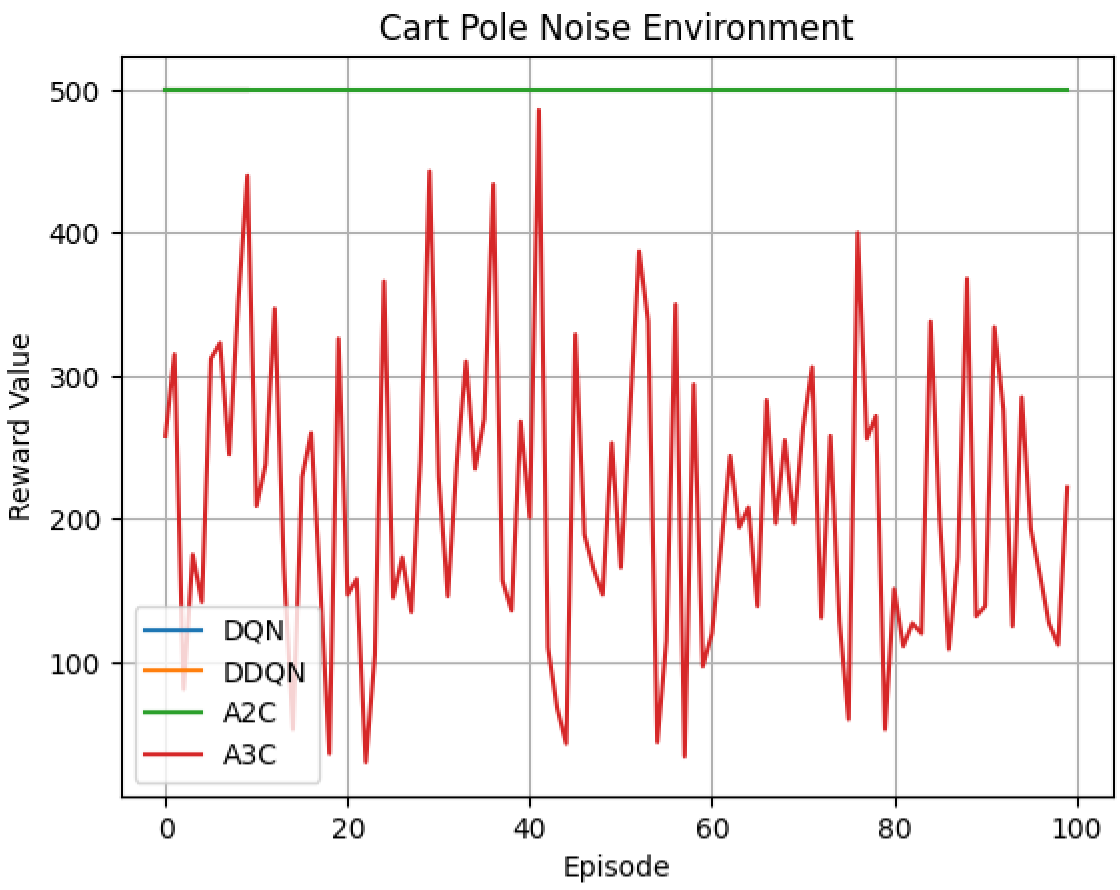
Results - Cart Pole vanilla environment summary





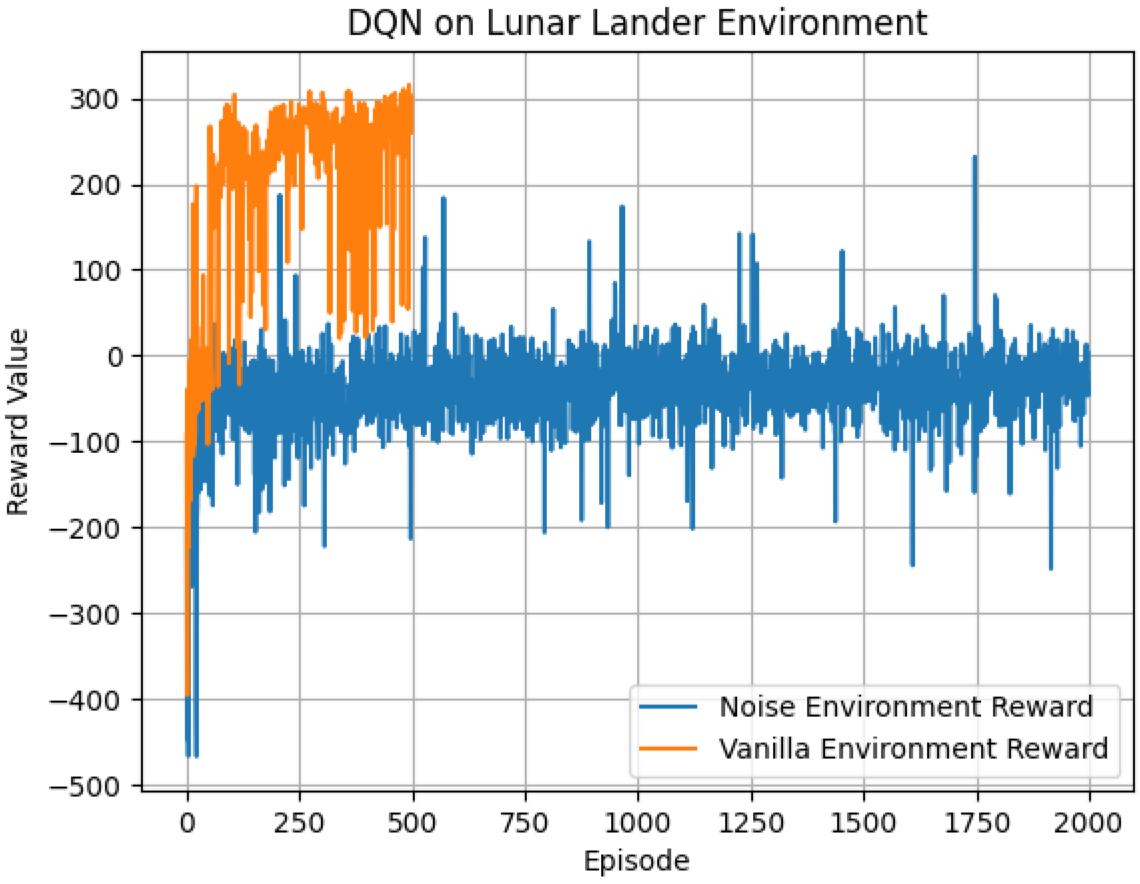
Results - Cart Pole noise environment summary

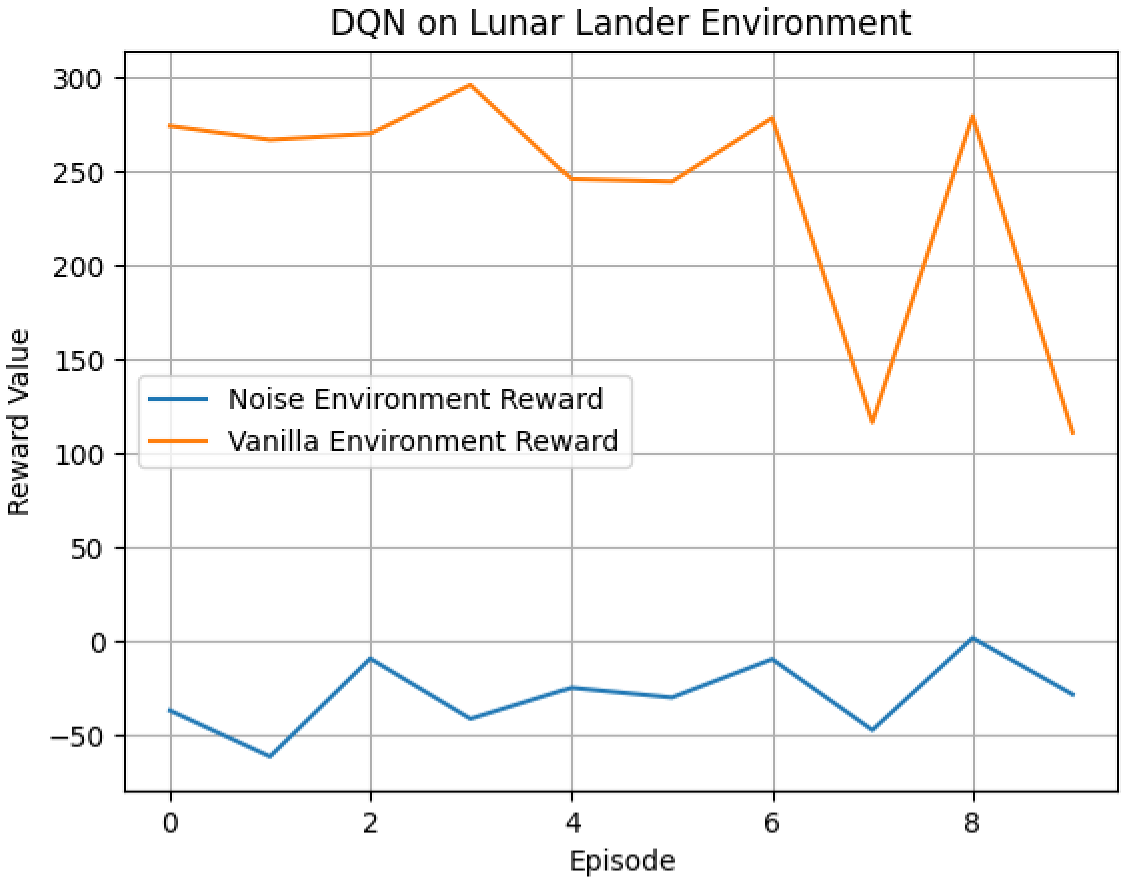




University at Buffalo The State University of New York

Results – Lunar Lander vanilla vs noise environment rewards





Summary

- The algorithms Deep Q-Learning Network, Double Deep Q-Learning Network, Advantage Actor Critic, and Asynchronous Advantage Actor Critic all performed well on Cart Pole environment.
- The stochasticity added to the Cart Pole environment was small enough and allowed the agent based on DQN, DDQN, A2C and A3C to successfully converge and learn an optimal policy.
- The lunar lander environment had a higher degree of uncertainty introduced into the environment with random engine failure and aerodynamic disturbance through wind noise. Thus, the agent following DQN algorithm failed to converge or learn the optimal policy in the modified Lunar Lander environment.

Team Contribution

Team Member	Project Contribution	Contribution %
Kanupriya Pandey	DQN Algorithm DDQN Algorithm	100%
	Noise in Environment	33%
Veda Sai Rochishna Eli	A2C Algorithm	100%
	Noise in Environment	33%
Antarpreet Kaur	A3C Algorithm	100%
	Noise in Environment	33%

University at Buffalo The State University of New York

THANK YOU