# Space Invaders and Reinforcement Learning
# A Comparison of Different Models

Benjamin Winslett
Master's in Computer Science
Software Engineer at SDL

# The Summary

- Space Invaders
  - OpenAI Gym
  - Destroy the alien invaders, get points
- Determine the best model
  - Fastest compute time
  - Highest score achieved
- Beat the average player score
  - 5,000 - human player average
- Models
  - A2C (Advantage Actor Critic)
  - DDQN (Dueling Double Deep Q Networks)
  - DQN (Deep Q Networks)
  - PG (Monte Carlo Policy Gradient)
  - PPO (Proximal Policy Gradients)
  - Rainbow
- Control variables (Present in all 6):
  - Learning rate = .99
  - Decay = .0001
  - Update Every = 100 episodes
  - Number of Episodes = 1,000, 5,000 episodes

# The Process

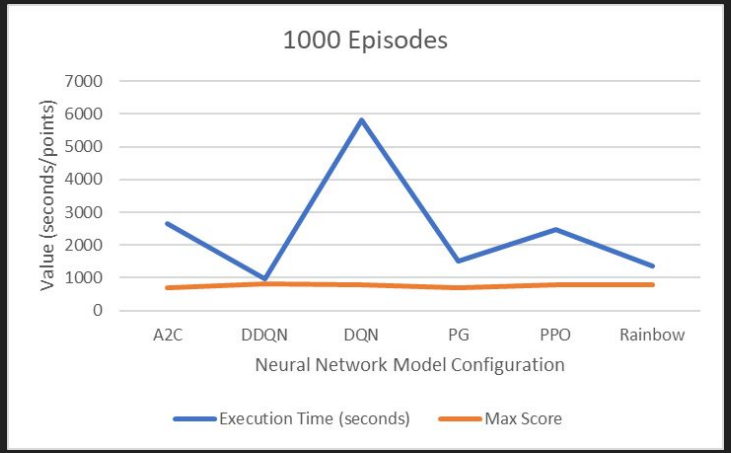
- Modified existing code
  - Generalized using similar hyper-parameters
  - Wrapped to compute time and output max score achieved
  - Applied a penalty for death
- Run code and recorded values and graphs.
  - 1,000 and 5,000 episodes
- Compared results (Tables)

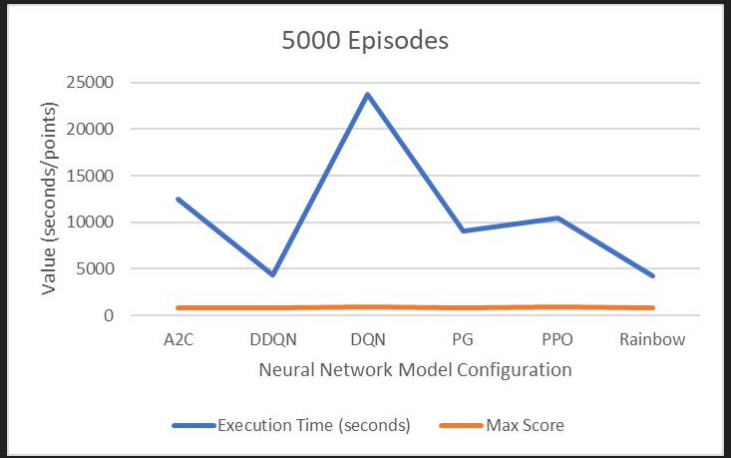
| 1000 Ep | Execution Time (seconds) | Max Score |
|---|---|---|
| A2C | 2662 | 705 |
| DDQN | 979 | 820 |
| DQN | 5802 | 785 |
| PG | 1515 | 690 |
| PPO | 2486 | 775 |
| Rainbow | 1352 | 800 |

| 5000 Ep | Execution Time (seconds) | Max Score |
|---|---|---|
| A2C | 12455 | 825 |
| DDQN | 4336 | 810 |
| DQN | 23762 | 955 |
| PG | 9019 | 855 |
| PPO | 10426 | 960 |
| Rainbow | 4279 | 825 |

# The Results

- On average, the same max score achieved, plus or minus 50
  - 762 +/- 50 points for 1,000 episodes
  - 871 +/- 50 points for 5,000 episodes
- Compared computation time instead
  - DDQN
  - Rainbow
- Decided to expand upon DDQN

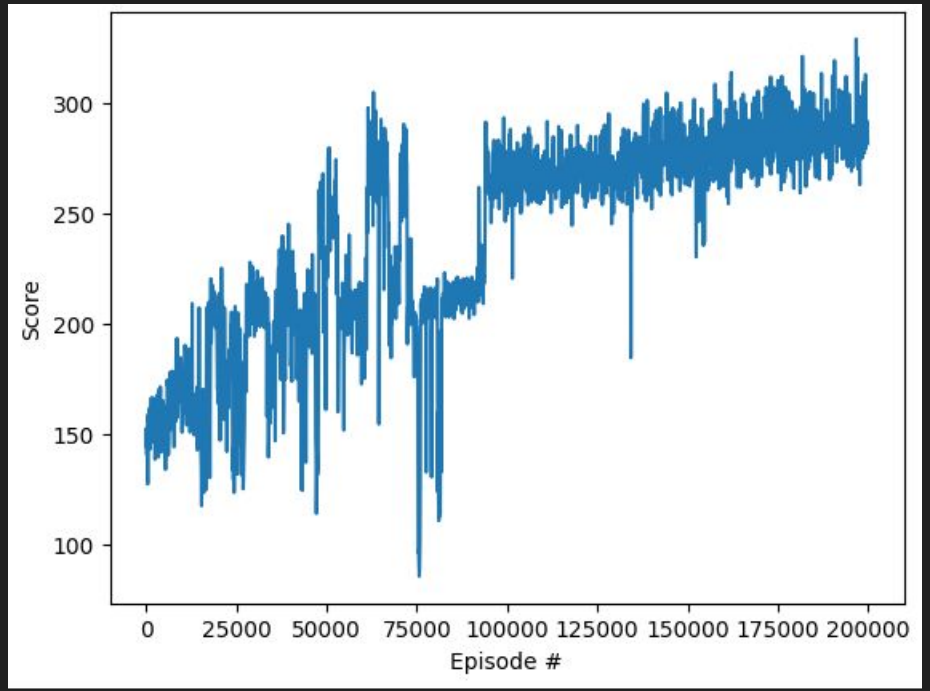


# The Results

- DDQN - Expanded
  - Slower learning rate
    - 1,000 times slower
  - 200,000 episodes
  - 18,826 seconds (5.23 hours)
  - 1,065 Max score
    - A decent score, but not close to the average

DDQN - Slower Learning Rate



# Conclusions

- Discovered the best model configuration for Space Invaders
  - Was unable to train it to beat an average score
- DDQN
  - Run for longer
  - Refine hyper parameters
    - Decay Rate
    - Starting Learning Rate
- Rainbow
  - Do similarly to DDQN

# Demo