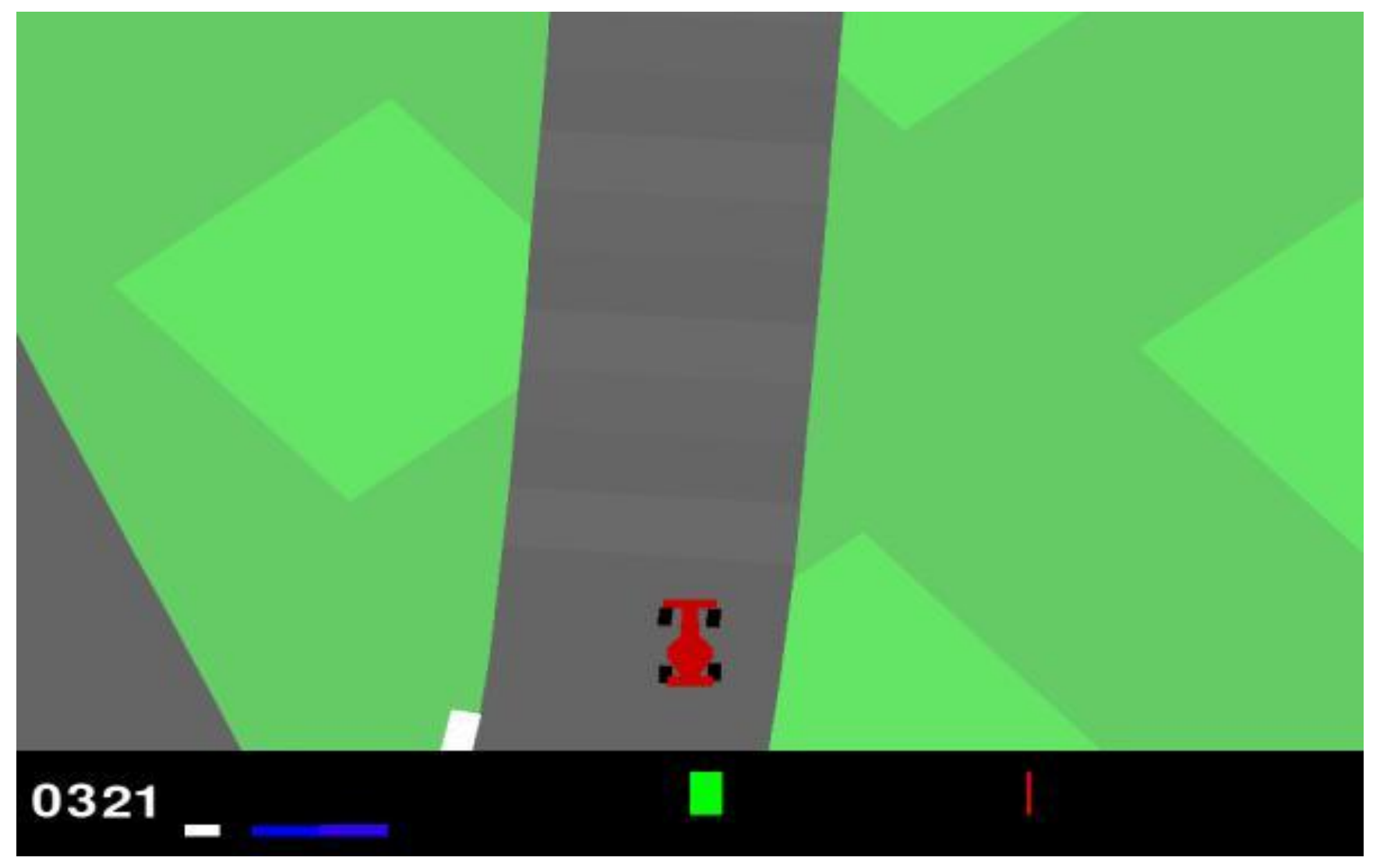
Comparison of Discrete and Continuous Action Spaces in Car Racing Environment

Piyush Malpure, Mithulesh Ramkumar, Rushikesh Deshmukh

Motivation

Many of the real-world applications do not have strictly discrete or continuous action space. They usually have hybrid action space and this hybrid action space needs to be either converted into discrete actions space or continuous action space. This project took a simple environment available with *Gym library* that had hybrid action space - CarRacing-v0. This project tried to compare the impact of choosing different action space on the effectiveness of the RL Agent.

Environment



Types of Action Spaces

1. Discrete Action Space: ✅

[Steer,Acc,Braking]
=
[0.0,1.0,0.0],
[1.0,0.3,0.0],
[-1.0,0.3,0.0],
[0.0,0.0,0.8]

2. Continuous Action Space: ✅

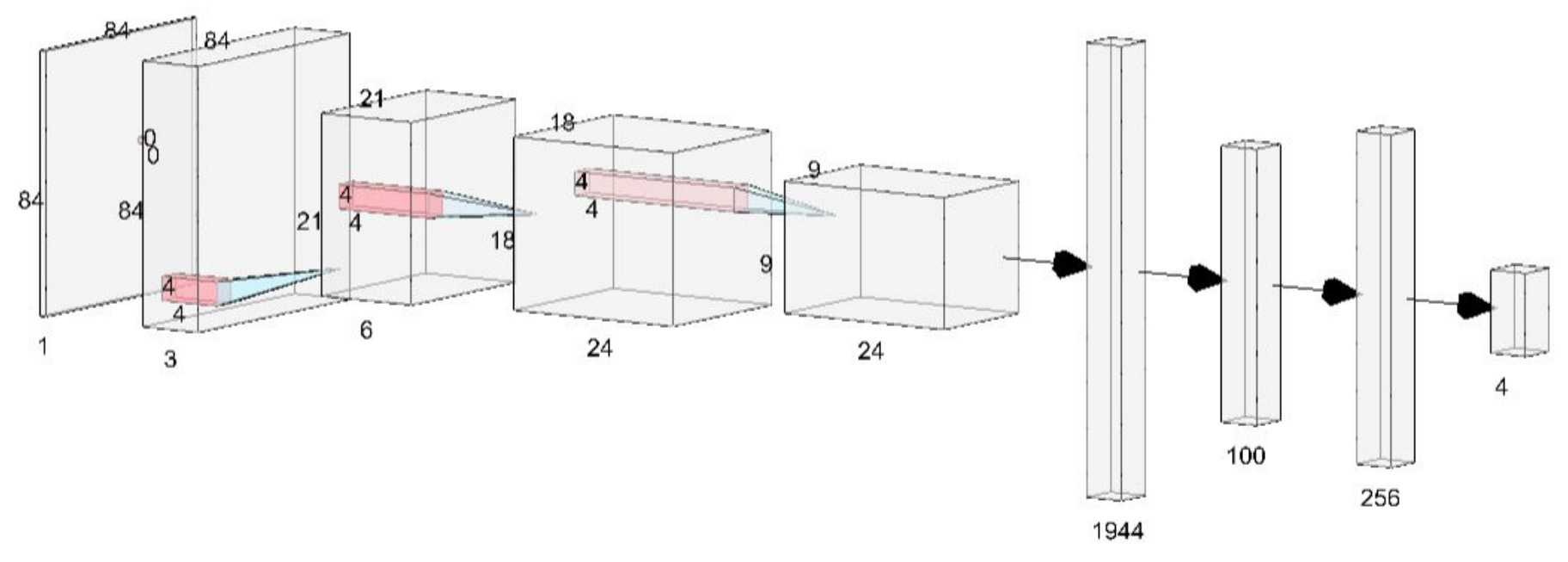
[Steer,Acc,Braking]
=
[log(-1:1), 0:1,0:1]

3. Hybrid Action Space: ❌

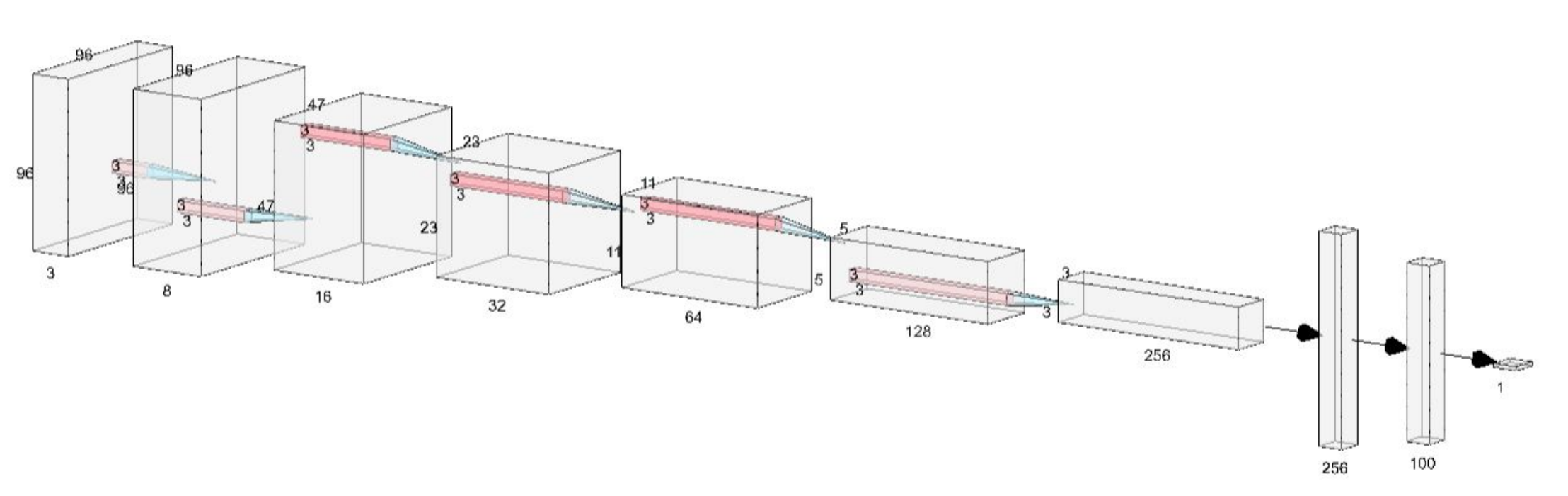
[Steer,Acc,Braking]
=
[-1 or 0 or 1, 0:1, 0:1]

Methodology

$$\mathcal{L}_{\theta_k}^{CLIP}(\theta) = \underset{\tau\sim\pi_k}{\mathrm{E}}\left[\sum_{t=0}^{T}\left[\min(r_t(\theta)\hat{A}_t^{\pi_k}, \mathrm{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t^{\pi_k})\right]\right]$$



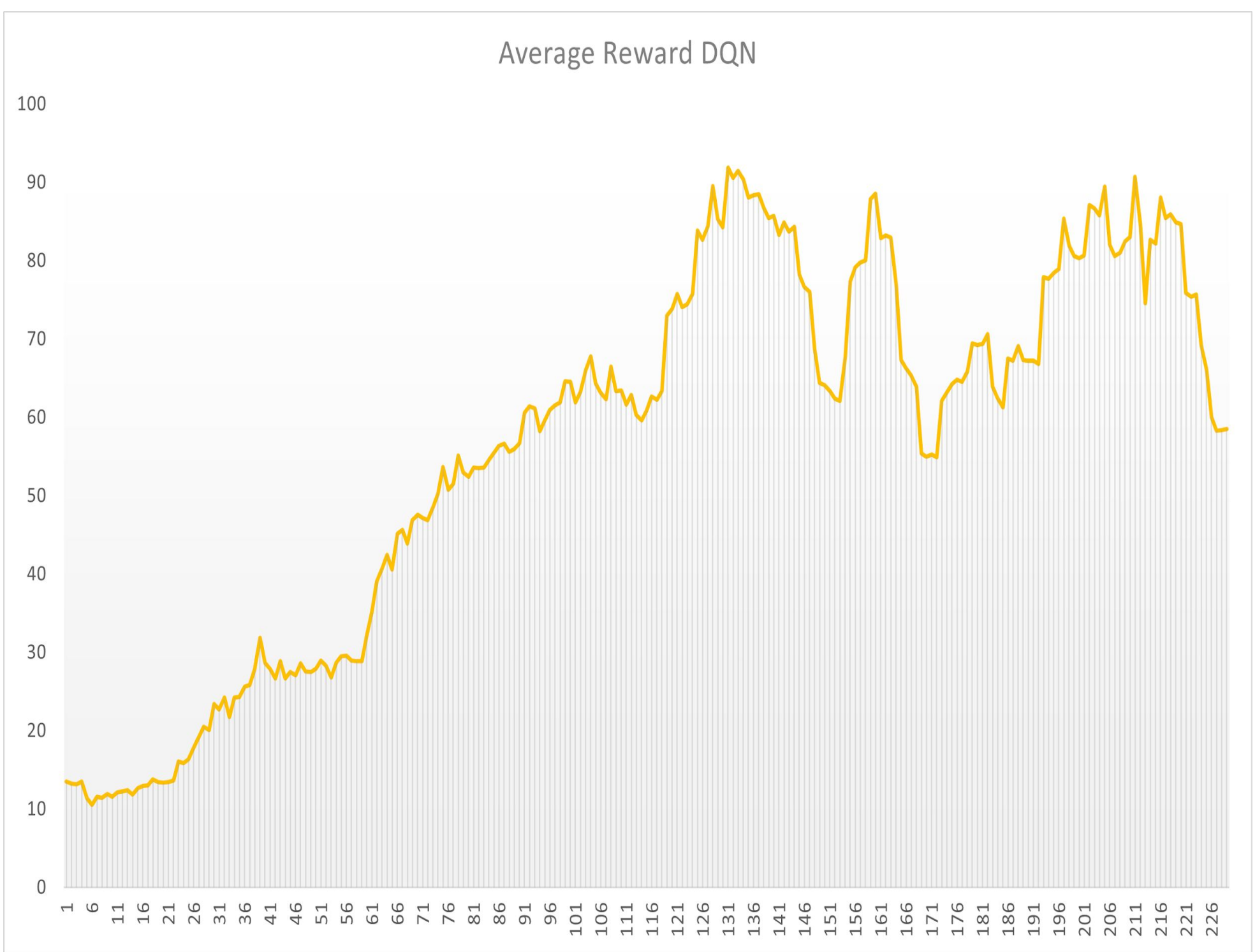
DQN

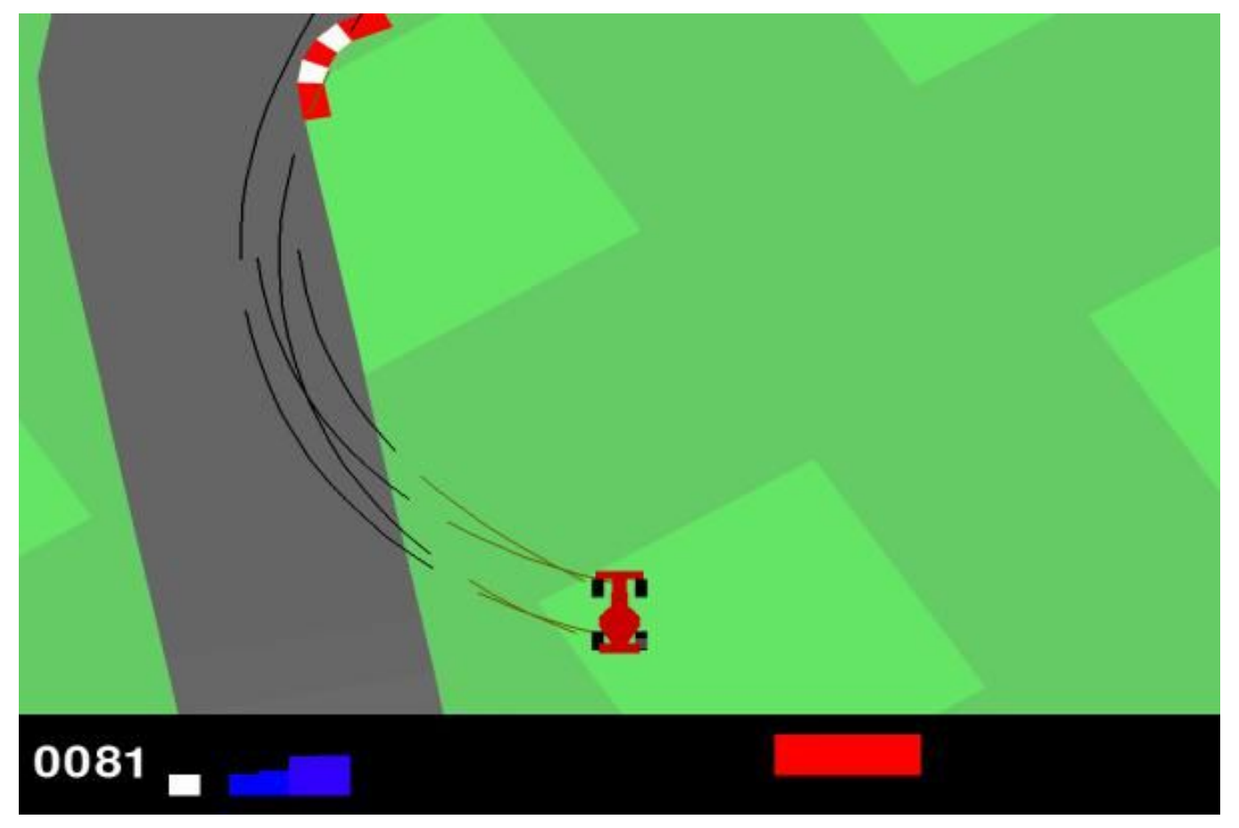


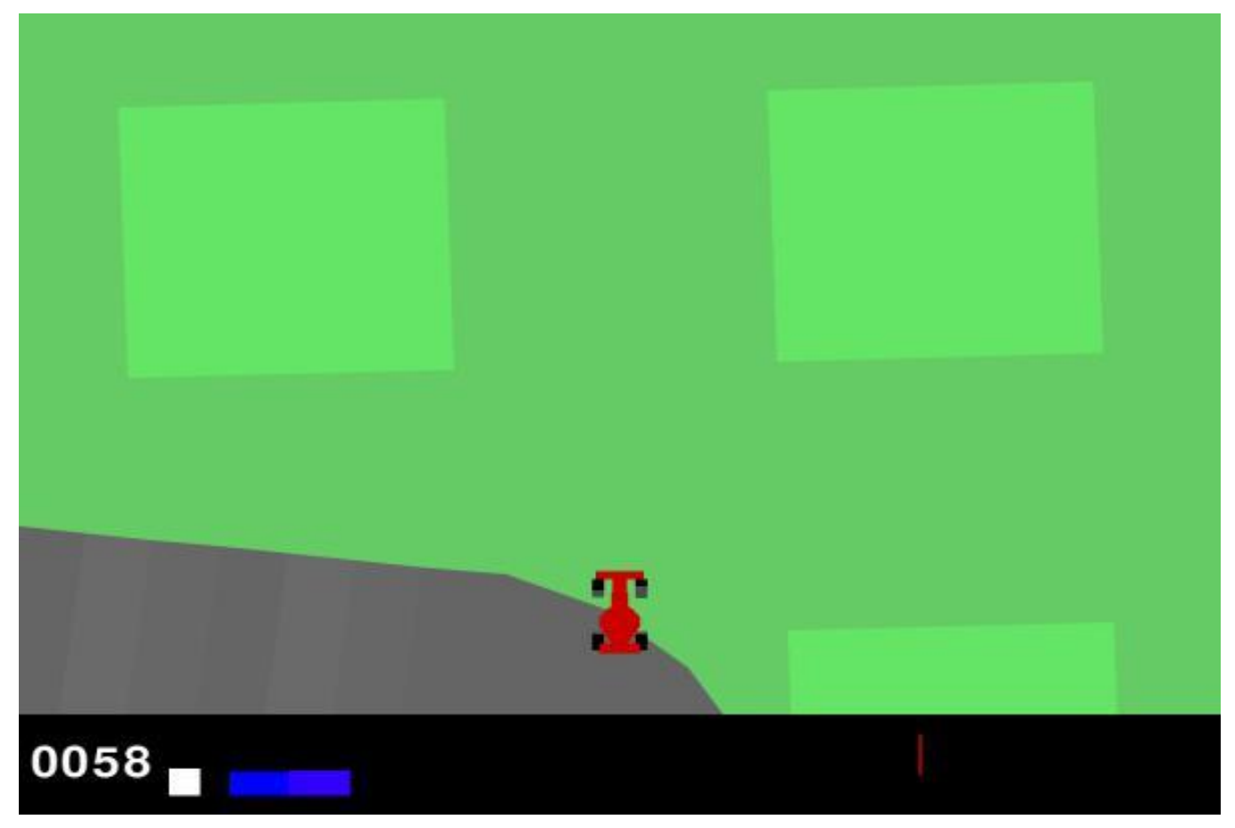
PPO

DQN Results

Avg. Testing Reward : 92.13

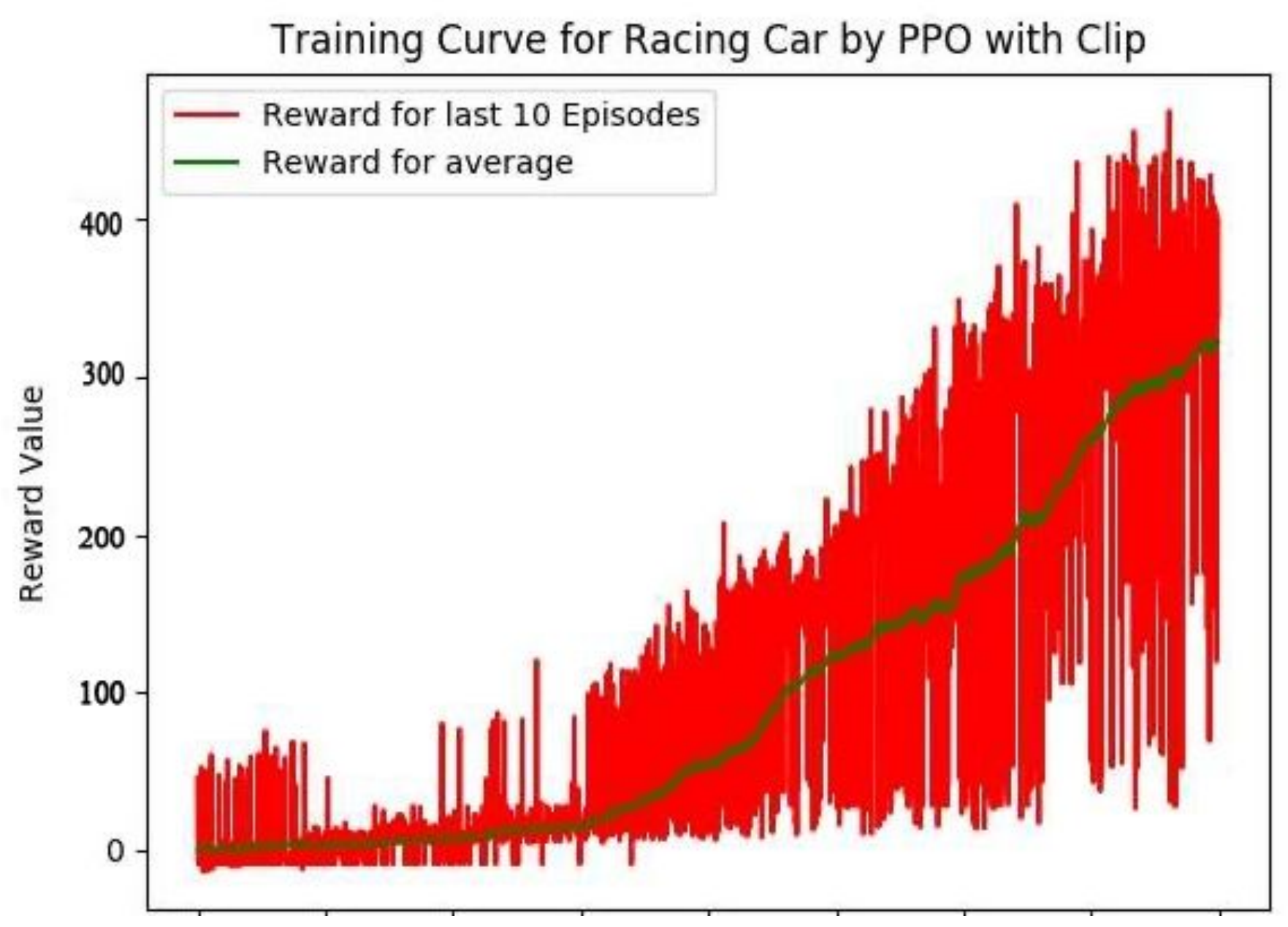


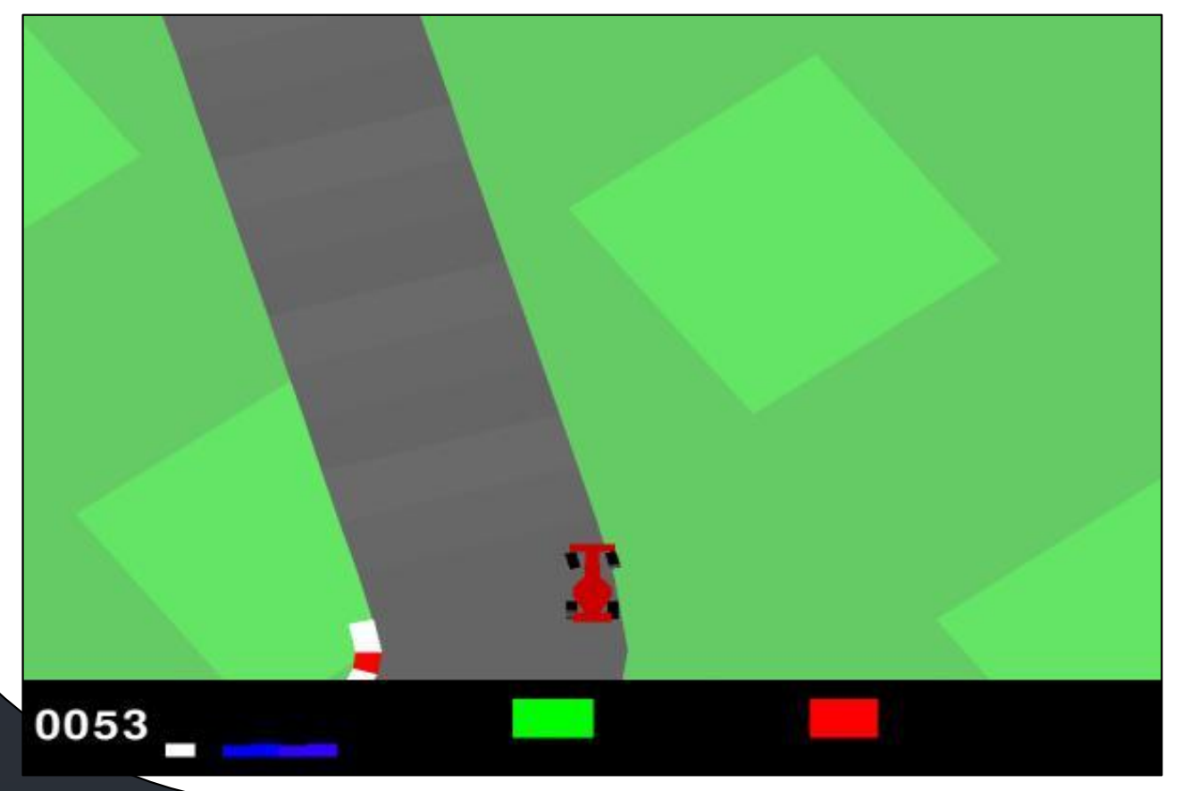


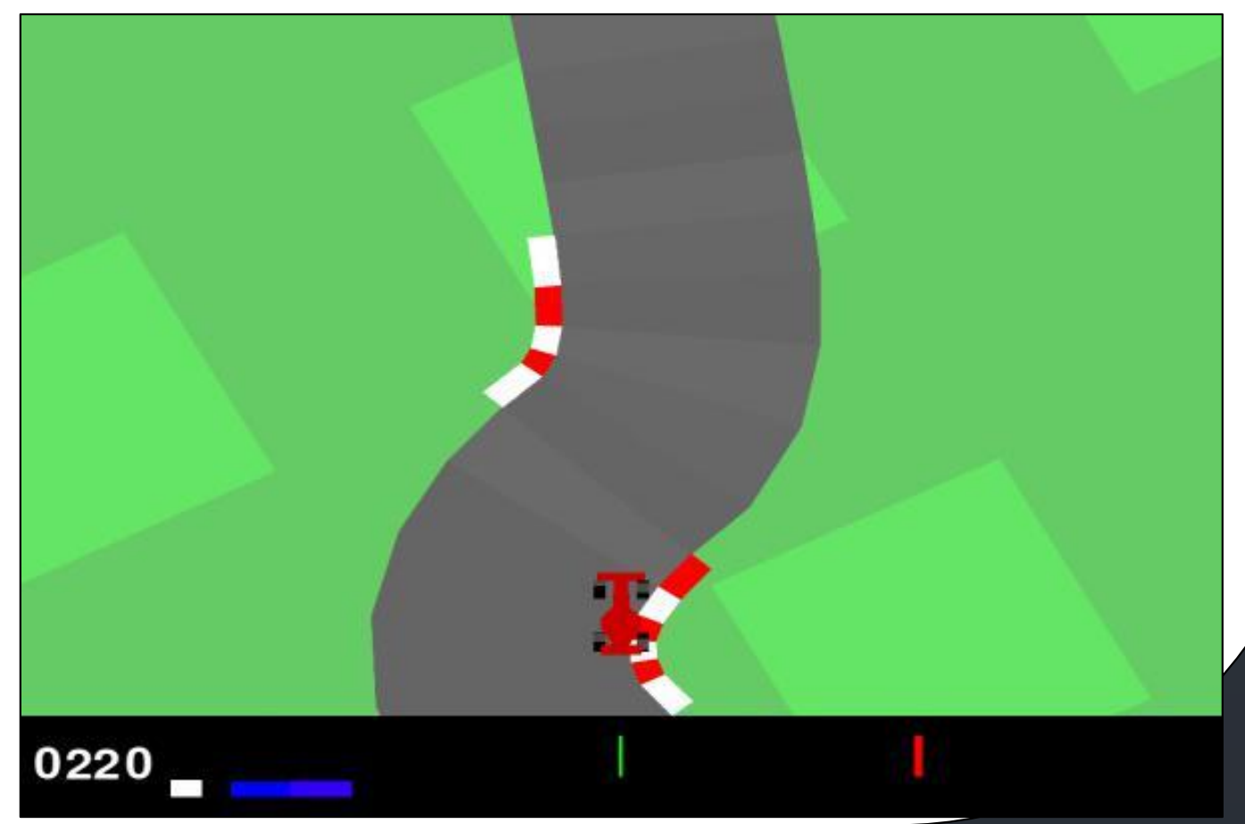


PPO Results

Avg. Testing Reward : 217.38







Discussion

As seen in the results it is clear that PPO implemented in continuous action space performs better than DQN implemented on a discrete space.PPO is a very robust algorithm and is more efficient as training advances, showing a very good performance in the environment. The car successfully manages to stay in the given track with a good acceleration using the trained agent. Discrete action space makes the problem easier to implement but introduces certain challenges to the agent to learn and overcome problems of sliding and slipping as seen in the results. Reasons for failure of the DQN implemented in discrete space is mainly due to the limited number actions given reducing its degree of freedom.

Citation

[1]Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., \& Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. doi:10.48550/ARXIV.1312.5602

[2]Zhang, Y. (2020). Deep reinforcement learning with mixed convolutional networks. arXiv preprint arXiv:2010.00717.

[3]Holubar, M. S., \& Wiering, M. A. (2020). Continuous-action reinforcement learning for playing racing games: Comparing SPG to PPO. arXiv preprint arXiv:2001.05270.

[4]Hu, Z., \& Kaneko, T. (2021, August). Hierarchical Advantage for Reinforcement Learning in Parameterized Action Space. In 2021 IEEE Conference on Games (CoG) (pp. 1-8). IEEE.

[5]Kakade, S., \& Langford, J. (2002). Approximately optimal approximate reinforcement learning. In In Proc. 19th International Conference on Machine Learning.