Solving Self-Driving Car Racing with DQN and PPO

CentraleSupélec

Raphael Faure, Adrien Fu

Introduction

Autonomous vehicles (AVs) are a key focus in today's industry, with promises of improving road safety by reducing human error, optimizing traffic flow, and lowering emissions through electric engines powered by clean energy. This project demonstrates how Reinforcement Learning (RL) can be leveraged to develop AVs. We design two agents based on Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO), each trained to drive efficiently in a simulated environment with partial observability.

Proposed method & model

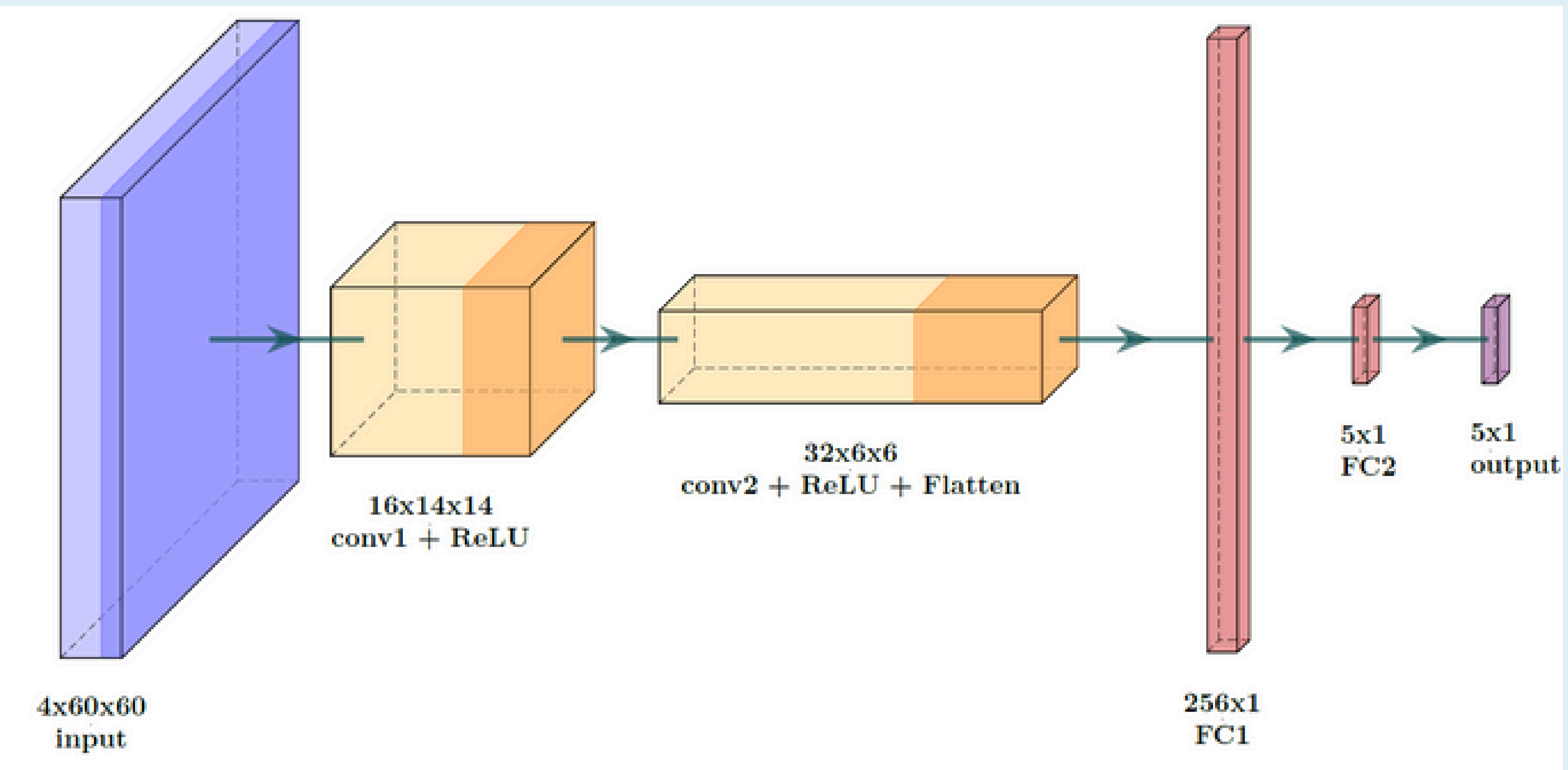
Environment & PreProcessing

- **First 50 steps are ignored** due to a zoom effect that disturbs learning.
- To restore **Markov property**, we stack the last k = 4 grayscale frames (cropped to 60 x 60) with frame-skipping (every 4 frames).
- The state includes temporal info to help the agent infer motion.

Actions Space & Rewards

- **Action Space :** 5 possible actions (do nothing, turn right/left, accelerate, brake).
- **Rewards :** -0.1 per frame, +1000/N per tile visited, episode ends at 1000 frames or if the car goes off track.

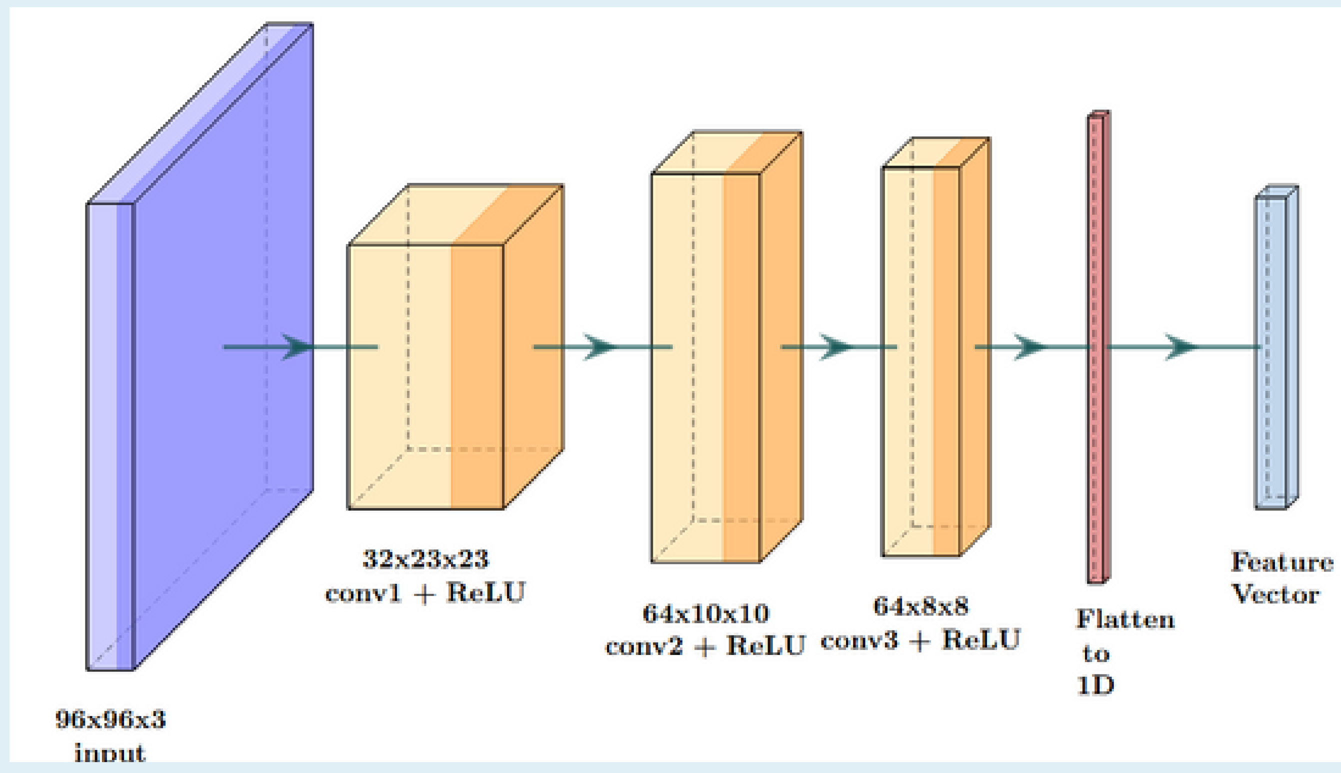
DQN Agent



- **Approximating Q* function with this NN**
- **We used the RMSprop optimizer with a learning rate of 0.0004 and a ϵ-greedy strategy with a decay rate**

PPO Agent

$$\mathcal{L}^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t,\ \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]$$

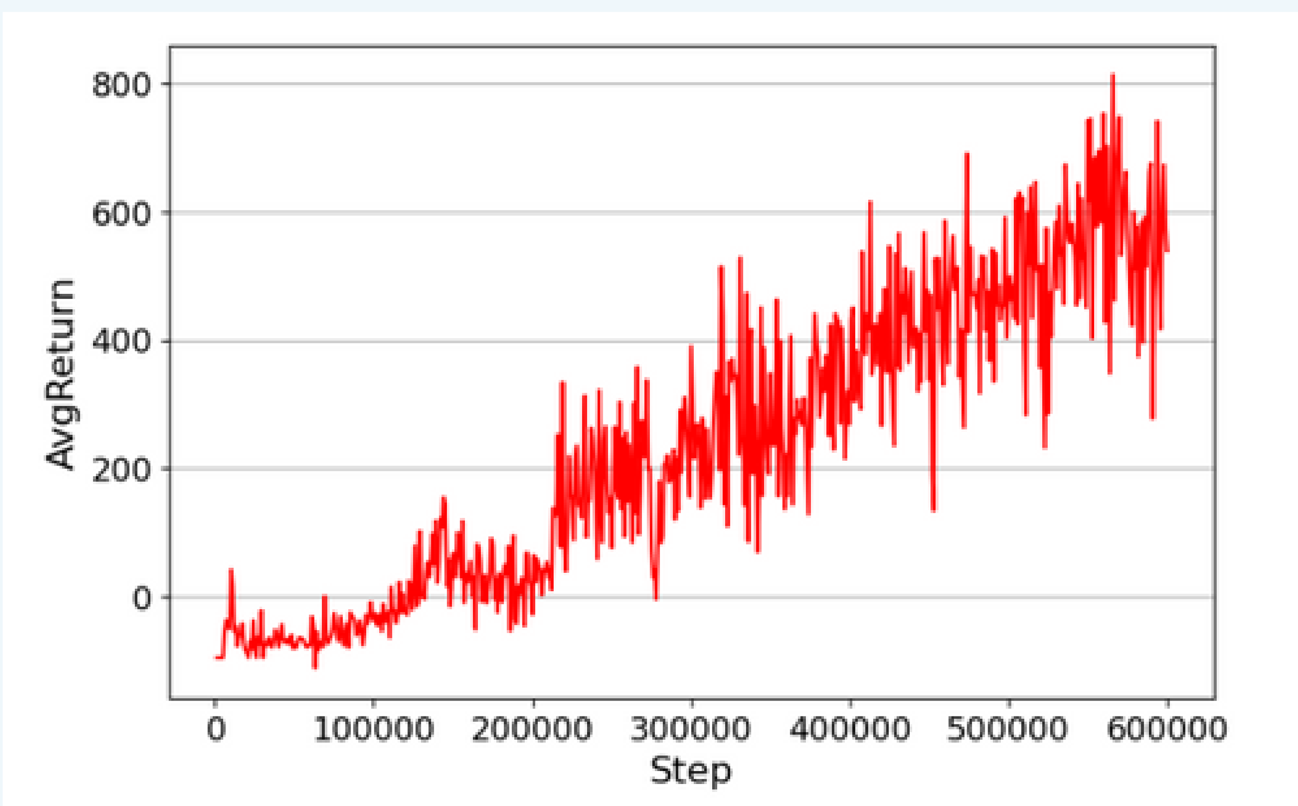


- **Adam optimizer** with a **learning rate of 0.0004** and a **batch size of 32**

Results

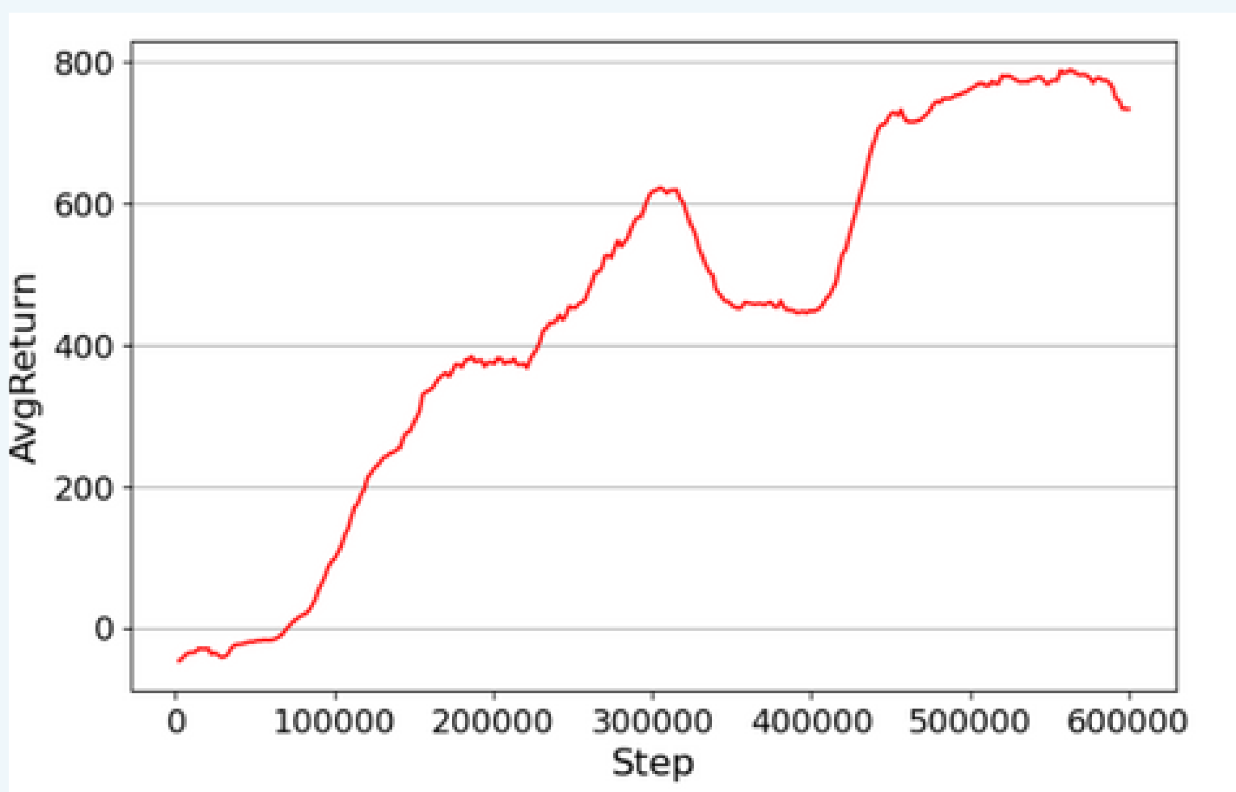
DQN Agent

- **12 hours to perform 600,000 steps**, with a final average return of **605.**
- The average return value fluctuates significantly during training but shows a general upward trend over time



PPO Agent

- **3 hours to perform 650,000 steps**, with a final average return of **733.**
- Performance decreased after 300,000 steps, suspected to be due to policy collapse, but improved after 400,000 steps.



Conclusion

- The project demonstrated **the potential and effectiveness of DQN and PPO** algorithms for **autonomous car navigation in a simulated environment.**
- **PPO achieved better performance in less time** but was **more prone to policy collapse**, while **DQN showed a general upward trend despite fluctuations.**

References

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov (28 Aug 2017). Proximal Policy Optimization Algorithms. OpenAI

V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller (19 Dec 2013). Playing Atari with Deep Reinforcement Learning. DeepMind Technologies.