Reinforcement Learning Models for playing Super Mario Bros

- Varadh Kaushik, Anirudha Shastri

PRESENTATION OUTLINE

- Introducing the topic
- Approaches/ algorithms implemented
- Comparison of algorithms
- Conclusion

SUPER MARIO BROS

We are using OpenAI Gym environment for Super Mario Bros on The Nintendo Entertainment System (NES) using the nes-py emulator.

The goal of this game is to reach the end of each level while avoiding pitfalls and enemies.

Proximal Policy Optimization [PPO]

- Proximal Policy Optimization is a model-free, on-policy, reinforcement learning algorithm
- Uses a “policy function” to define a probability distribution over actions
- A key feature is its use of “proximal” objective function, which helps stabilize the learning process

Deep Q Networks [DQNs]

- Deep Q Networks are a model-free, reinforcement learning algorithm
- They use a neural network to approximate the optimal action-value/Q function
- Can handle high-dimensional state spaces, that are found in complex, real-world environments, as they use neural networks

Double Deep Q Networks [DDQNs]

- Double Deep Q Networks are a variant of DQNs
- They use two separate networks, called the “target” and “online” networks
- In DQNs use the same network to both select the action of a given state and to evaluate the quality of that action
- Here, we use the “online” network to select actions and the “target” network to evaluate the quality of those actions

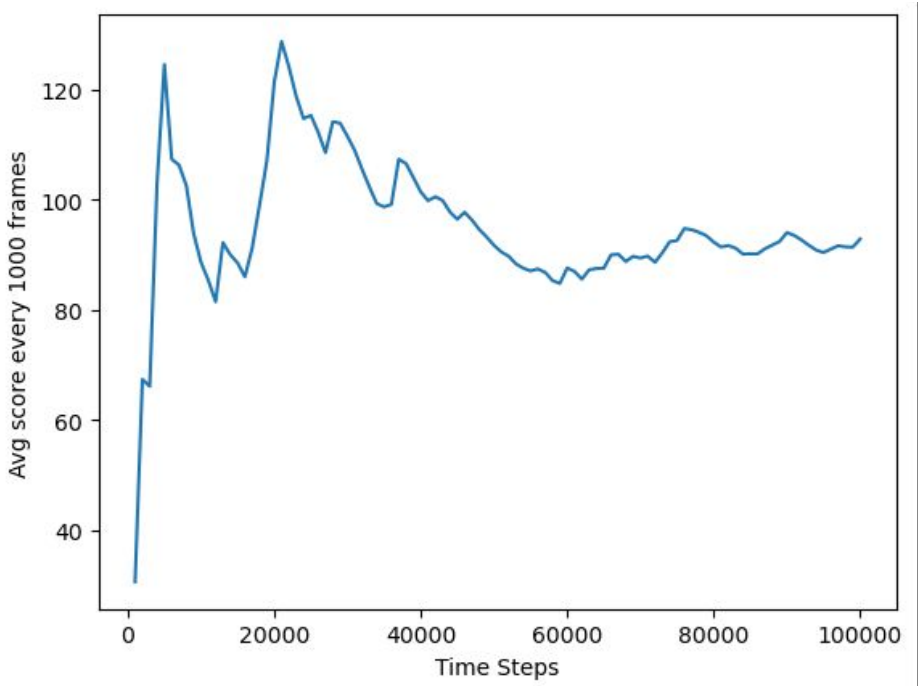
Performance comparison: PPO

Train: 3,000,000 time steps

Learning Rate: 0.000001

Model save intervals: 10,000

Training time: 6 hrs



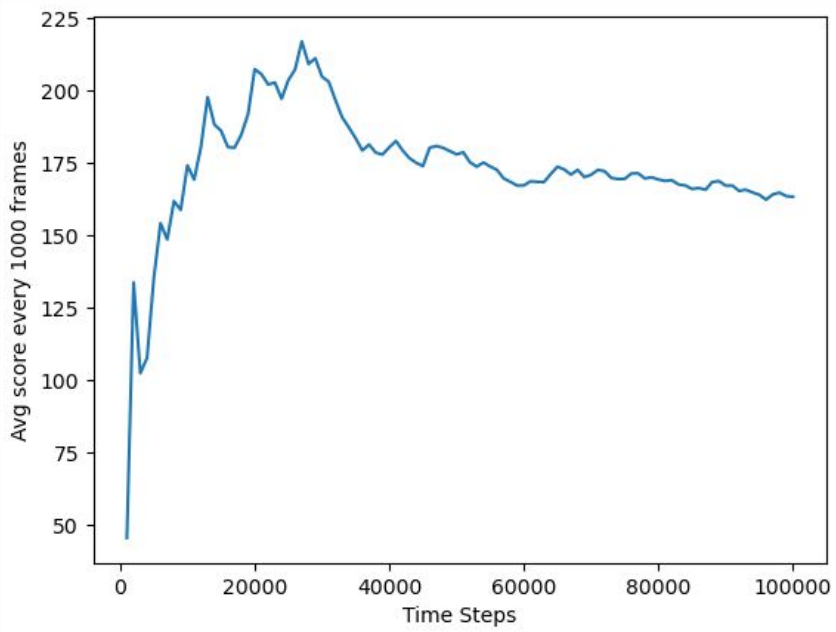
Performance comparison: DQN

Train: 3,000,000 time steps

Learning Rate: 0.000001

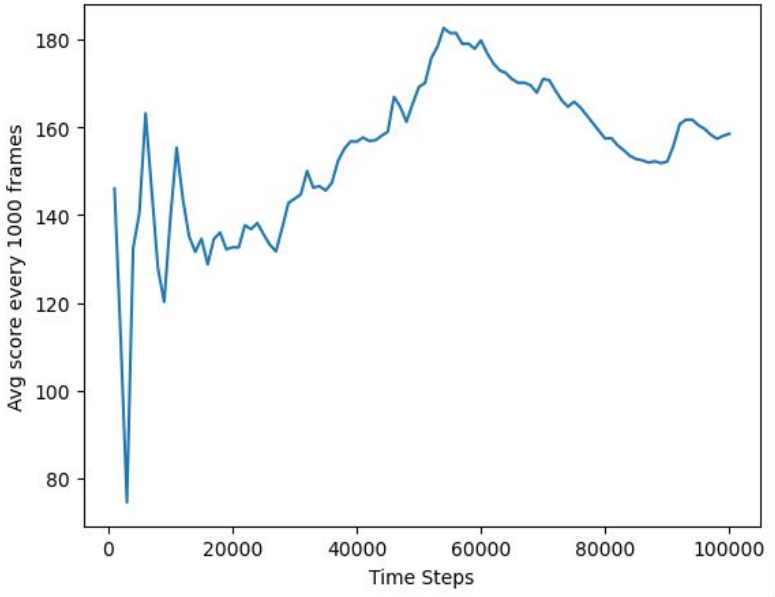
Model save intervals: 10,000

Training Time: 6 hrs

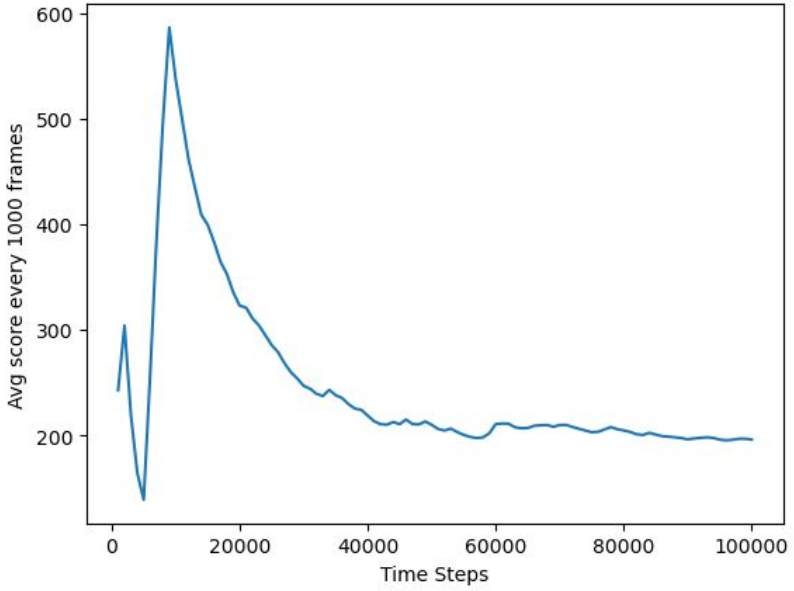


1,050,000 time steps model

Performance comparison: DQN (Cont)



1,740,000 time steps model



2,560,000 time steps model

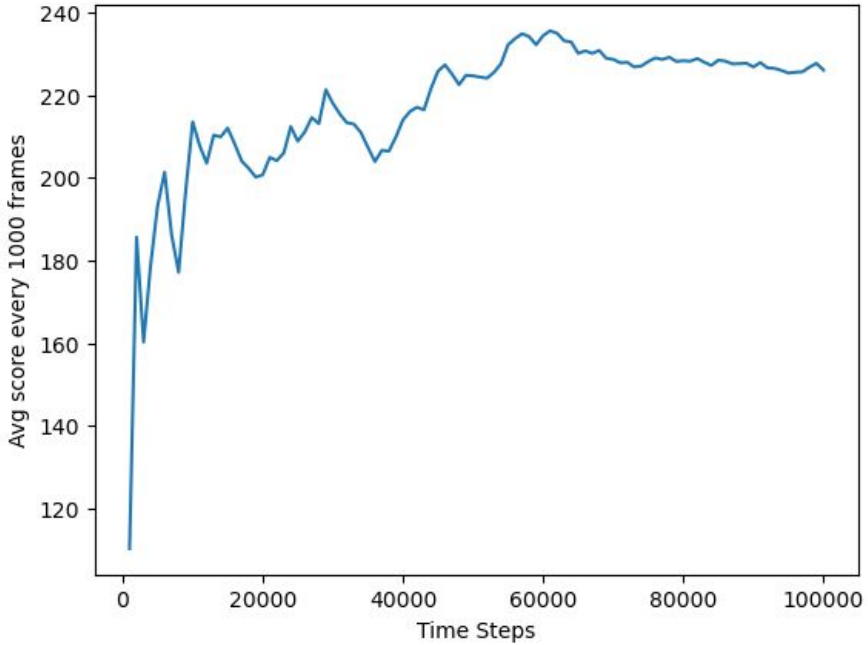
Performance comparison: DDQN

Train: 2000 Episodes

Learning Rate: 0.00025

Batch size:128

Training Time: <3 hrs



POTENTIAL IMPROVEMENTS

- Training for longer on better hardware
- More complicated CNN models for (D/)DQNs
- Other Reinforcement algorithms: Alpha GO Zero

CONCLUSION

1. PPO: Needs to lots of training and will have lower performance than DQN and DDQN for same amount of training.
2. DQN: In game it extracts more features like collecting coins as well. Slightly better performance than PPO.
3. DDQN : Performs extremely well for little training. It is able extract more details and have a higher score in the same traversed distance.

REFERENCES

- Firoiu, Vlad, William F. Whitney, and Joshua B. Tenenbaum. "Beating the world's best at Super Smash Bros. with deep reinforcement learning." *arXiv preprint arXiv:1702.06230* (2017).
- Schulman, John, et al. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017).