A.Y. 2023/2024

PPO FOR PROCGEN

Autonomous & Adaptive Systems

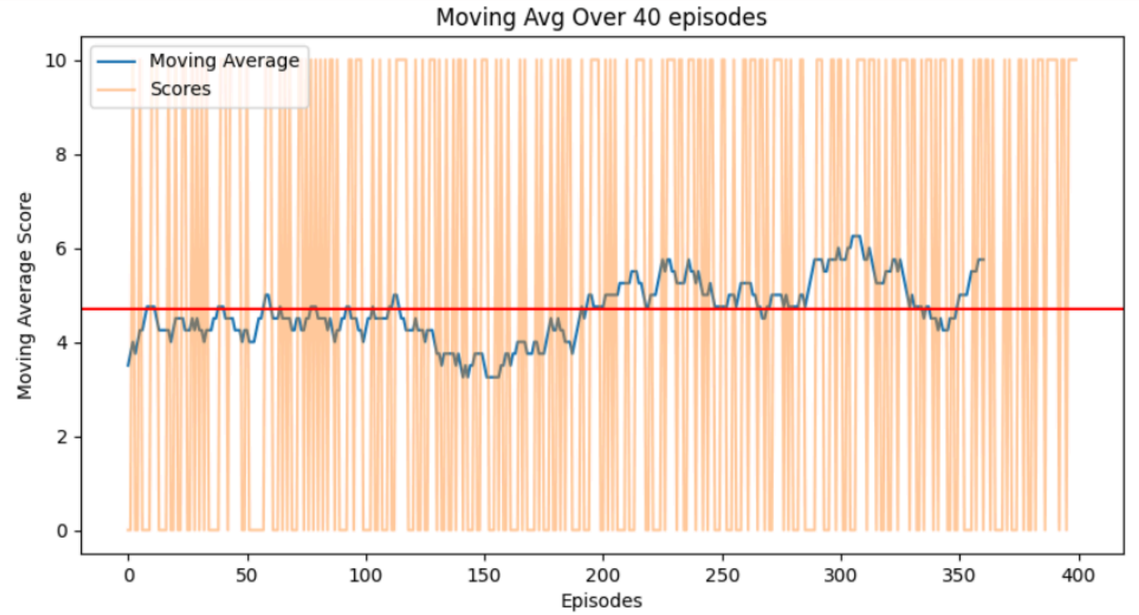
Umberto Carlucci

Choice of PPO and Experiment Setup

- Given the intrinsic difficulty of ProcGen, any actor-critic algorithm could have been beneficial for tackling it, especially on resource-constrained environments.
- PPO has been chosen for its robustness to hyperparameters and the performances already shown on ProcGen-like environments.
- To stabilize learning further, it has been implemented also a function to compute GAE (Generalised Advantage Estimation), giving good results.
- After some preliminary tests, the two games chosen for the experiments are CoinRun and BigFish.
- For each experiment, both the best model achieved during training and the fully trained model are tested.

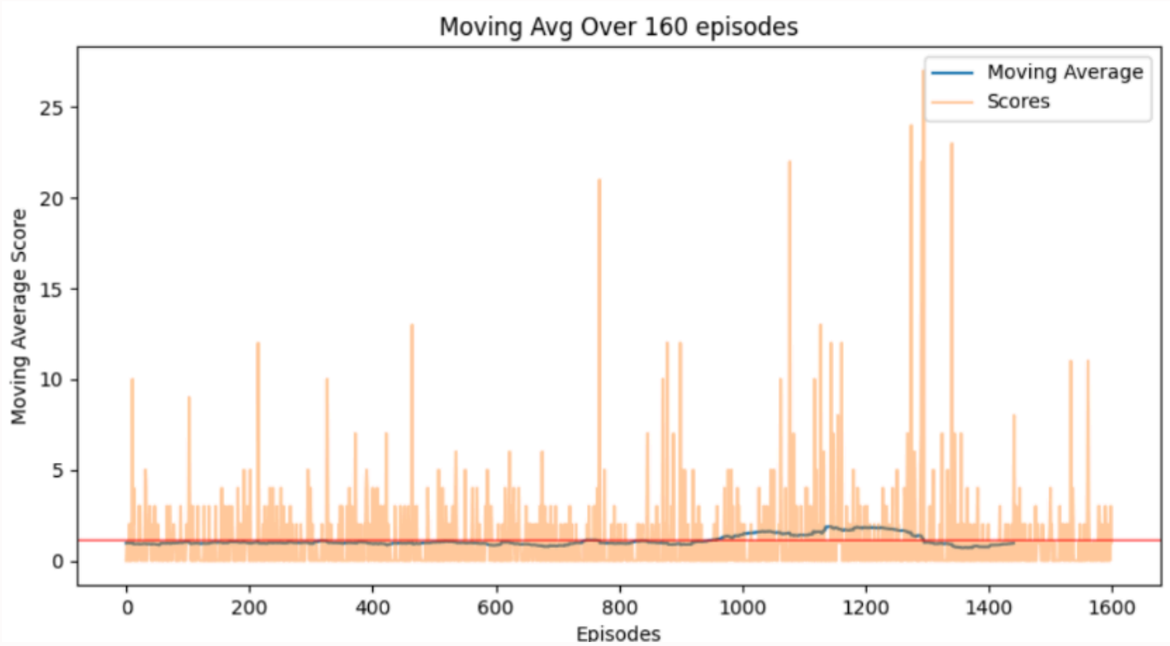
Results

CoinRun



Agent	Env	Avg	Highest
Full	Training	4.7	10
Random	Testing	2.5	10
Full	Testing	5	10
Best (344)	Testing	5.125	10

BigFish



Agent	Env	Avg	Highest
Full	Training	1.09	27
Random	Testing	0.84	10
Full	Testing	0.84	12
Best (1303)	Testing	3.14	40

Discussion & Limitations

- The experiments have shown how the PPO algorithm can generalize over games from the benchmark.
- PPO robustness to hyperparameters has been shown testing how good results were provided even with different values of lambda and gamma
- The best hyperparameters found were also tried with the backgrounds on, giving good results but only for the best scoring agent.
- An higher number of timesteps and a larger amount of training levels on both games could surely enhance performance.
- Using a bigger neural network and two different networks for the actor and the critic could be beneficial.

CONCLUSIONS

- *On CoinRun, it was possible to overcome the random agent relatively easily, probably due to the fact that the game is the easiest within the entire benchmark.*
- *On BigFish, instead, only the best performing agents succeeded in overcoming the random agent.*
- *The use of the backgrounds didn't make the results too much worse*
- *Despite challenges and limitations, the potential of the PPO agent in tackling this kind of tasks has been showcased.*

THANK YOU

FOR YOUR ATTENTION