Bomberman AI opponents

Dawid Herc, Wojciech Francke, Tomasz Knura

1. Introduction

The goal of the project was to create fully functional AI opponent for simple unity bomberman game, which includes smooth moving, laying down bombs, running away from explosions, collecting power ups etc.

With the last version AI should be capable of being a challenging opponent for a human player

Learning process was planned to be conducted using Imitation Learning and Self-Play, starting with IL to teach an AI basics of the game

Imitation Learning and Self-Play

Imitation learning techniques aim to mimic human behavior in a given task. An agent (in this case - AI opponent) is trained to perform a task from demonstrations by learning a mapping between observations and actions.

The next step from that would be self-play. **Self-play** is where the algorithm learns by playing by itself without requiring any direct supervision while the performance is controlled by system of rewards (**Reinforcement learning**).

ML-Agents

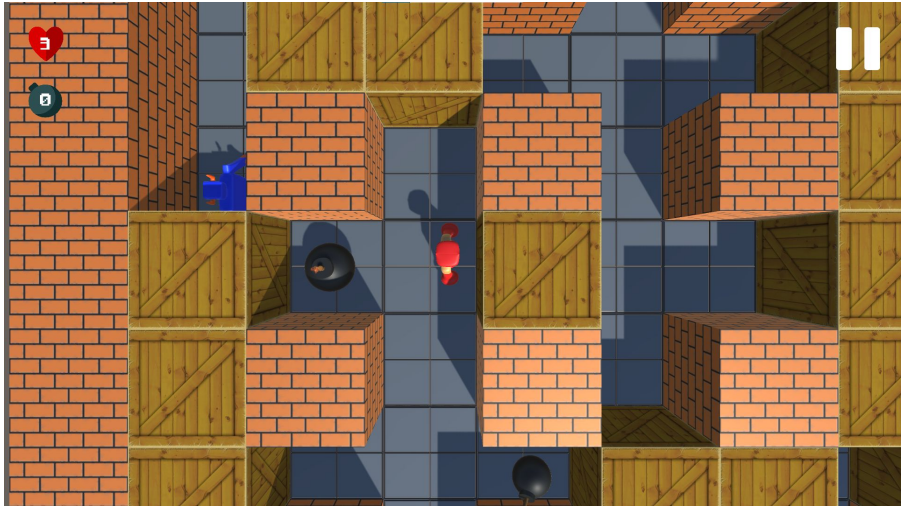
As this is a unity game, so as environment we choose **Unity Machine Learning Agents Toolkit (ML-Agents)**, which is an open-source project that enables games and simulations to serve as environments for training intelligent agents.

It provides Python API for training using reinforcement learning and imitation learning which we planned to use.

2. State at the beginning

The base for this project was the aforementioned simple unity bomberman game that we created for an earlier project.

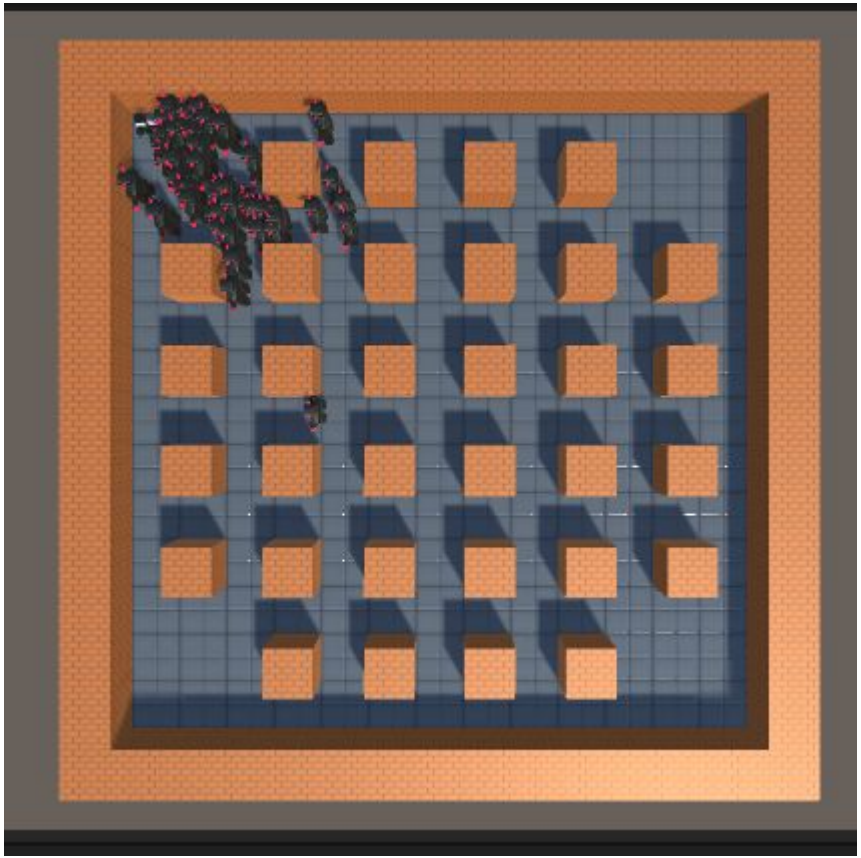
However, the game had to be severely cut down and edited to provide a suitable environment for agent training



3. Work progress after week 1

After the environment was prepared we could begin training. However, before we could start imitation learning or self-play we mostly spent time making him move properly or stop him from breaking the game.

By the end of week 1 we taught agent how to follow a player controlled character.

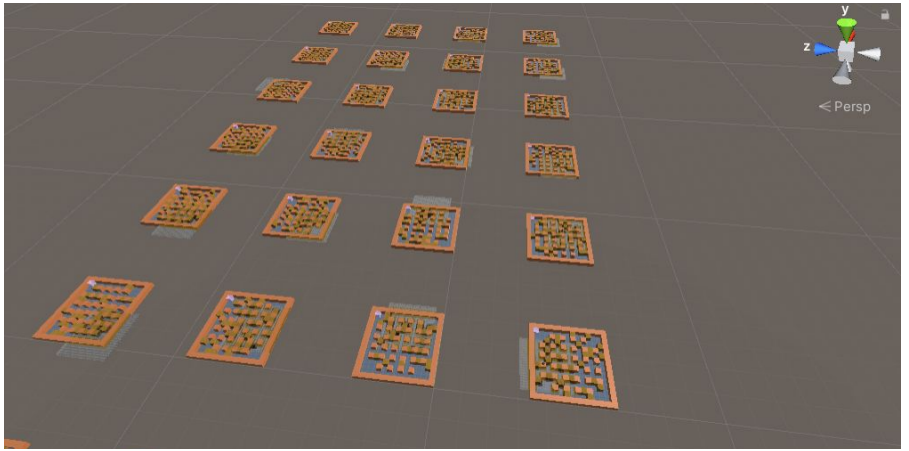


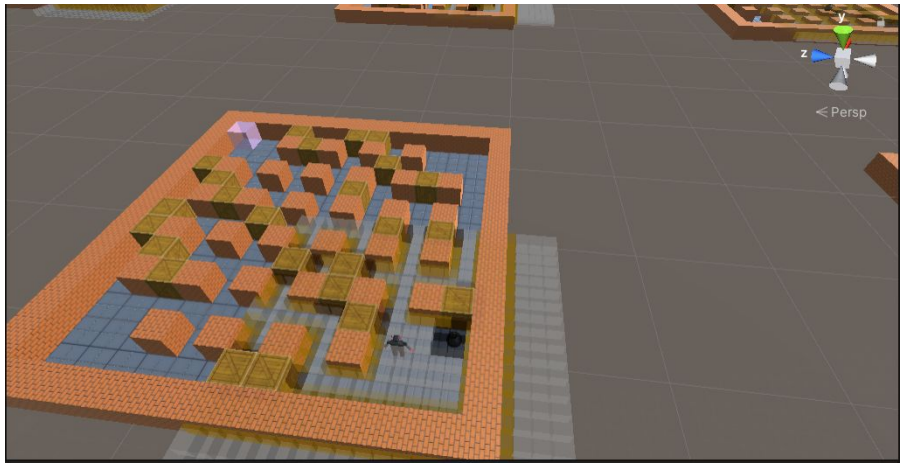
4. Work progress after week 2

After that we decided to teach them how to actually play the game by giving them ability to plant bombs to move through blocked map.

More focus went towards reinforcement learning. Agents were taught through optimization of the reward system.

Also, we expanded project by creating multiple separate environments for separate agents, whose performance was monitored through TensorBoard





5. Current work progress

Final rewards

Time: -0.001

Moving forward: 0.1

Putting bomb properly: 0.1

Being near bomb: -0.05

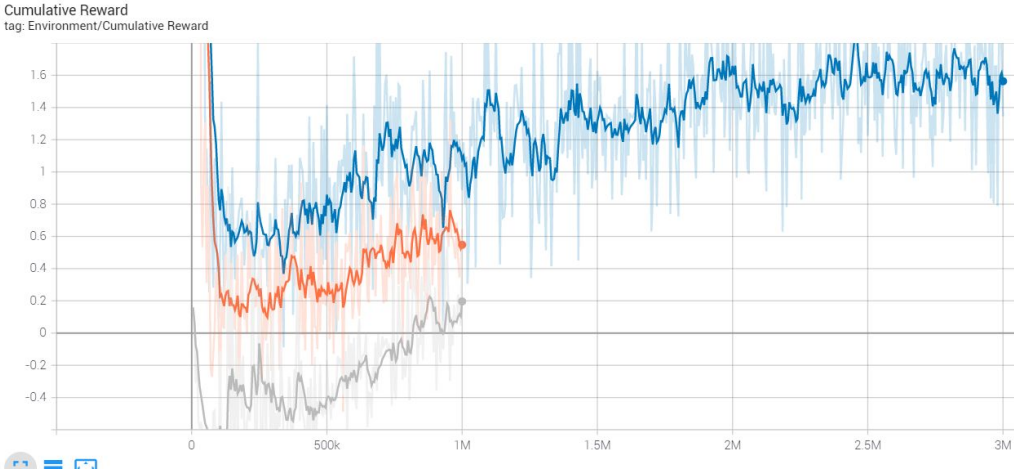
Death: -1

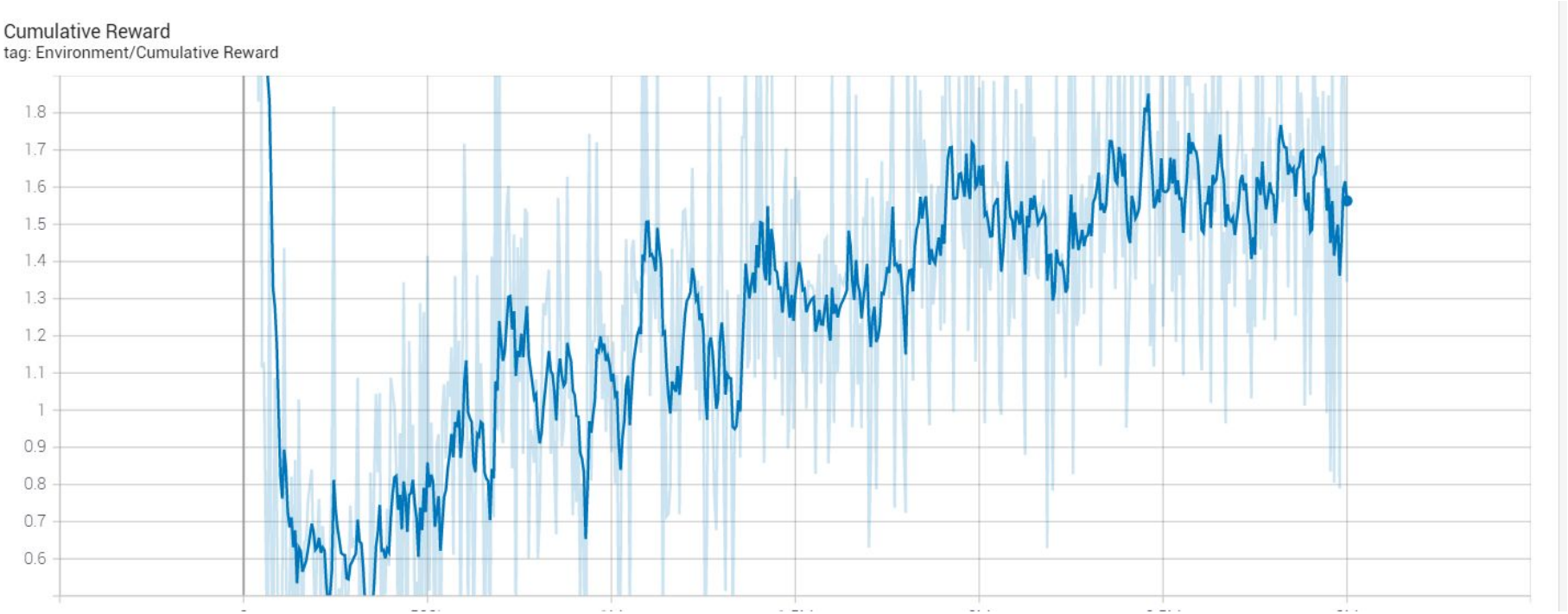
Reaching target: 1

We were able to find best working rewards. The idea of imitation learning was abandoned due to its inefficiency.

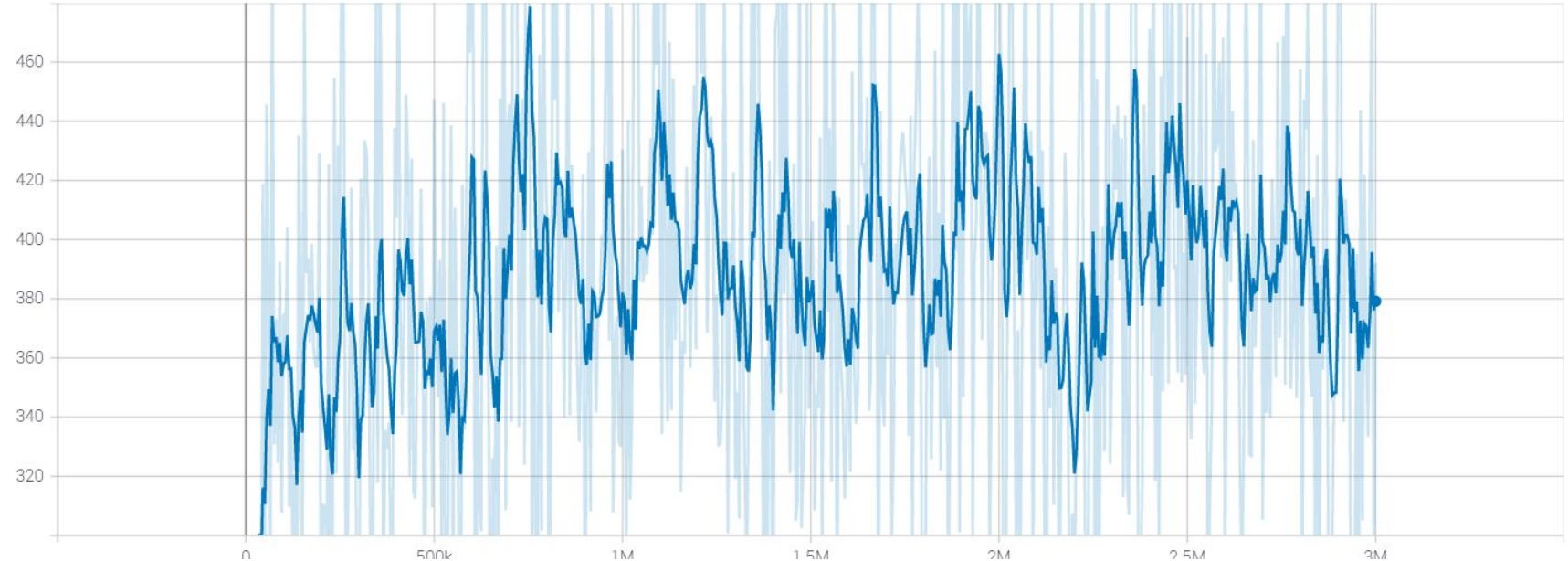
We also extended time of learning (up to 12 hours) which gave us the best results. We think we did not face the issue of overtraining, so maybe the time could be extended more for even better results.

We added curiosity to agents behavior in order to help them not being afraid of exploring the map.

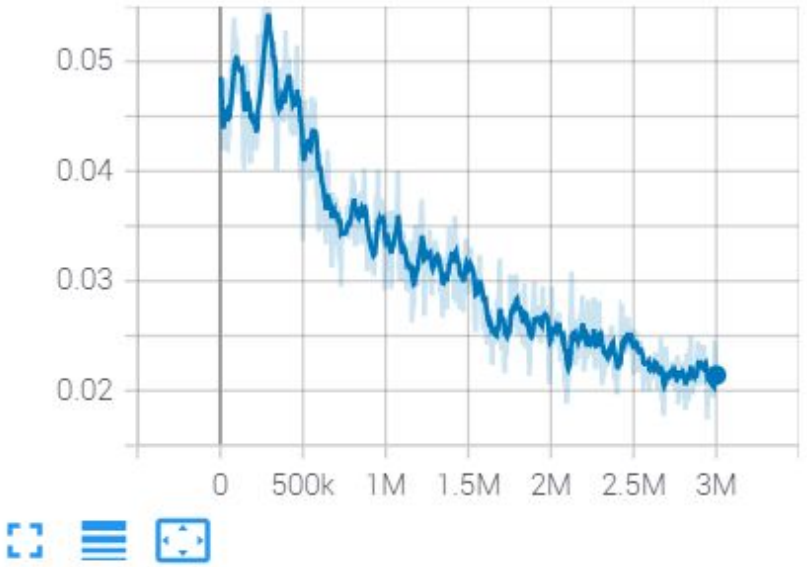



Cumulative Reward
tag: Environment/Cumulative Reward
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6

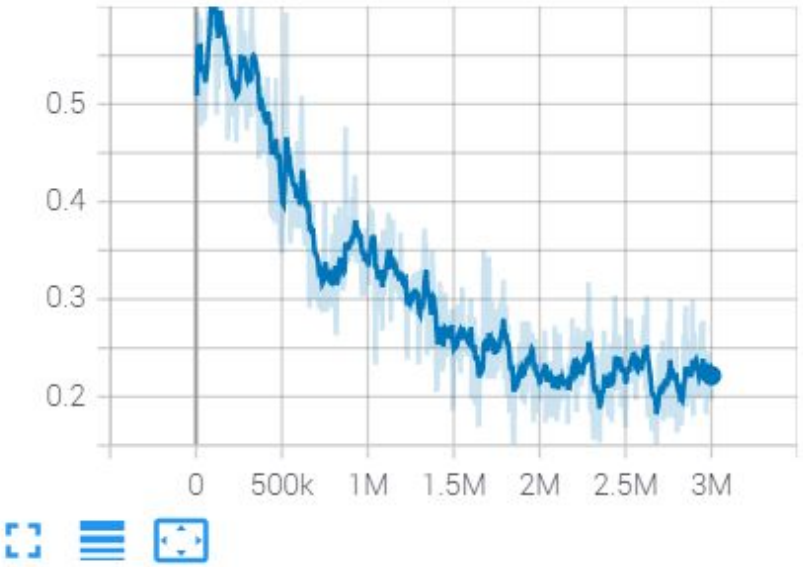
Episode Length
tag: Environment/Episode Length



Curiosity Forward Loss
tag: Losses/Curiosity Forward Loss



Curiosity Inverse Loss
tag: Losses/Curiosity Inverse Loss



5. Future improvement

Thank you