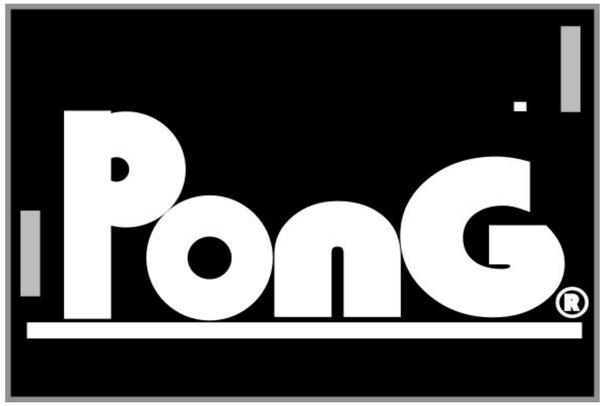
Deep Reinforcement Learning Applied to Atari Pong Game

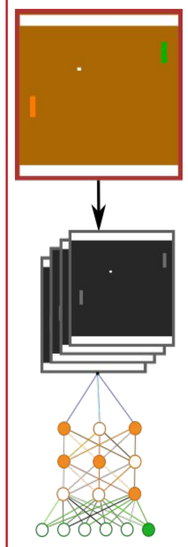
Mathias Stefan Carius Larsen - s164029,
Dennis Jensen - s155629

Introduction

- The goal is to design an AI agent that learns to play the game Pong based on raw pixels and rewards.
- Pong is a good choice of game since its concept is simple, but introduces some complex DQN algorithms.



Deep-Q-Network



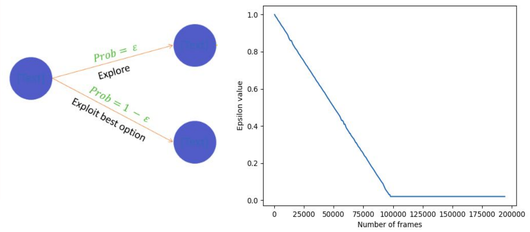
1. Loading Pong game using Gym package from OpenAI
   - Dimension: 210x160 pixels
   - RGB
2. Preprocessing the game:
   - Downscale to 84x84 pixels
   - Grayscaling
   - Frame skipping
3. Feeding the Neural Network
   - Output: Estimated Q-values
   - Replay memory
   - Calculate loss
   - Backpropagation
   - Optimizer step

Experience Replay

- Reduces correlation of training examples used to update the networks [5]
- Increases learning speed and performance [6]
- Stores memory of old transitions

Epsilon-Greedy

- ϵ is the percent of the time to take a random action
- Exploration decreases as experience increases → **exploration/exploitation**
- ϵ-decay tries to smoothen the transition between exploration and exploitation [3]



DQN

- Single Q-value giving the value of an action
- Basic Q-network leads to overestimation[4]

Double DQN

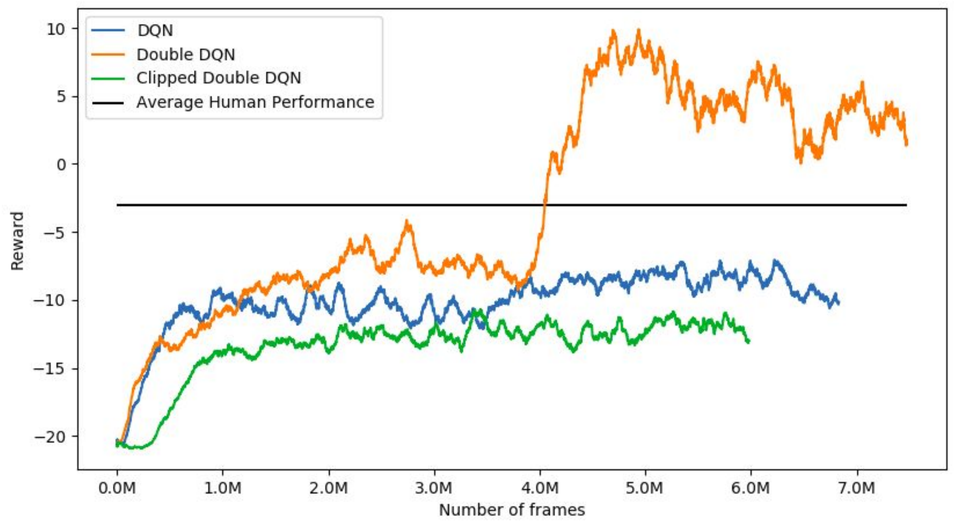
- Target-network, Training-network
- Copies the training network with a frequency over into the Target-network [1]
- Solving overestimation issues [1][4]

Clipped Double DQN

- Two independent networks
- Both networks estimates Q-values
- Takes the greater Q-value estimation of the two networks and reduces it to the minimum Q-value estimation of the other network [1]
- Solves overestimation issues [1]

Results - Qualitative

- The reward on the y-axis is the mean of the last 100 games played
- The average human performance gets a reward of -3 [2]
- Overestimation for the DQN does not happen in the figure to the right. This could be because the amount of frames trained on are not high enough. If the overestimation problem occured, the DQN would become unstable and get a worse reward when learning on more frames [4]



Results - Quantitative

Model	Best performance reward
DQN	-7.05
Double DQN	+9.91
Clipped Double DQN	-10.71
Average Human	-3

References

[1] Double-Deep-Q-Network, Chris Yoon, Tackling maximization bias in deep learning: https://towardsdatascience.com/double-deep-q-networks-905dd8325412
[2] Playing Atari with Deep Reinforcement Learning V. Mnih et al., NIPS Deep Learning Workshop, December 2013
[3] Epsilon-Greedy Algorithm, Imad Dabbura, Marts 31, 2018, https://imaddabbura.github.io/post/epsilon-greedy-algorithm/
[4] Deep Reinforcement Learning with Double Q-learning, Hado van Hasselt and Arthur Guez and David Silver, Google Deepmind, December 8, 2015, https://arxiv.org/pdf/1509.06461.pdf
[5] Reinforcement Learning with Hindsight Experience Replay, Or Rivlin, January 31, 2019, https://towardsdatascience.com/reinforcement-learning-with-hindsight-experience-replay-1fee5704f2f8
[6] Welcome to Deep Reinforcement Learning Part 1: DQ, Takuma Seno, October 2017, https://towardsdatascience.com/welcome-to-deep-reinforcement-learning-part-1-dqn-c3cab4d41b6b