

Deep Q-learning on Atari Assault

CS 229 Final Project

Fabian Chan, Xueyuan Mei, You Guan
{fabianc, xmei9, you17}@stanford.edu

Predicting

Motivation: an old classic but fun and challenging game; human player vs. machine.

We built a neural network based on convolutional layers to handle the huge number of possible states
Inputs are k consecutive frames of pixels converted to black/white. This is a matrix of size $(k, 250, 160)$
Outputs are the Q values for every action
Results: deep q-learning obtains significantly higher score than normal q-learning

Features:

Raw pixels processed into black and white pixels in binary. We chose these pixels to be our features because they are simple to obtain and we can let the CNN figure out the rest.

Data:

Game frames from the OpenAI framework as 250x160 pixels, then processed into black and white binary values. Each frame is associated with a reward.

References:

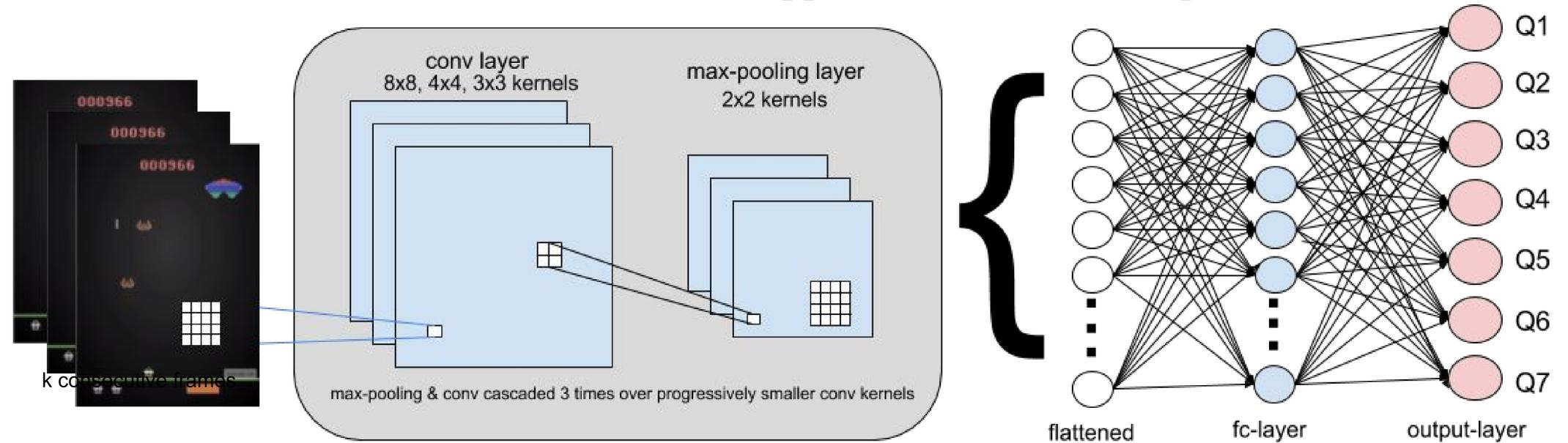
Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Wierstra, D., & Riedmiller, M. (2016). Playing Atari with Deep Reinforcement Learning. University of Toronto.
Assault (1983, Bomb) - Atari 2600 - Score 4153. (n.d.). Retrieved November 16, 2016, from <https://www.youtube.com/watch?v=Wxio1\ ytTTo>
Matiisen, B. T. (n.d.). Demystifying Deep Reinforcement Learning. Retrieved November 16, 2016, from <http://neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/>

Models:

Q-learning: MDP recurrence $Q(s^*, a) = E_{s', \epsilon} [r + \gamma \max_{a'} Q^*(s', a') | s, a]$

Weight update $w \leftarrow w - \eta [\hat{Q}_{opt}(s, a; w) - (r + \gamma \hat{V}_{opt}(s'))] \Phi(s, a)$

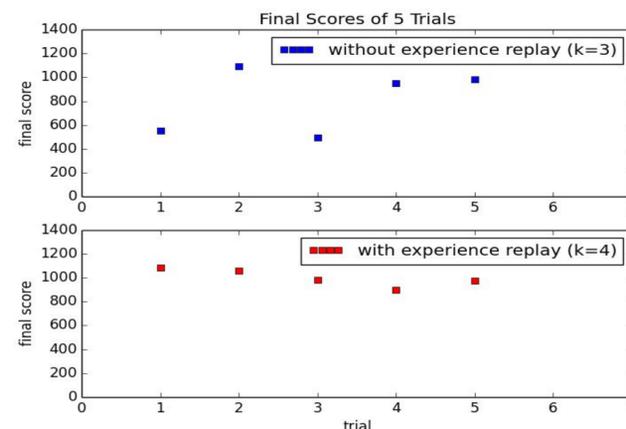
Convolutional neural networks as function approximation (GPU computation):



Loss function: $L = \frac{1}{2} (r + \gamma \max_{a'} Q(s', a') - Q(s, a))^2$

Results:

Ordinary Q-learning as Baseline: 670.7 ± 101
Deep Q-learning (5 trials): 980 ± 130



Discussion:

The baseline performance for our project is the performance of a simple Q-learning algorithm with a simple feature extractor that only indicates whether or not a pixel is black. The performance we can now achieve by our implementation of the deep Q-learning algorithm is discernibly better than the baseline, especially with the addition of experience replay. We expect deep q-learning to perform better and it did, because CNN can learn better high-level features for better performance.

Future:

We would like to investigate and ascertain whether the convolutional network layers alone are sufficient for feature extraction. If not, we have a few ideas in mind that will add additional feature extraction and would like to try them out. However, we also want to limit the number of features we add, since the more features we include, the more time it takes to train our model. Therefore we would like to prioritize our choice of feature extractors via forward search.