Upload Slides and start recording!!!!

Advanced Topics in Reinforcement Learning

Lecture 17: Deep Reinforcement Learning II

Josiah Hanna
University of Wisconsin — Madison

Announcements

- Literature review due tonight at 11:59PM Central.
- Homework 4 due November 17 (two weeks from today).
- Start reading Chapter 13 (skip 13.6)

Midterm Survey

- Thank you to everyone who filled it out!!!
- Suggestions for this semester:
  - Late policy
  - Pre-class slide uploads
  - Lecture coverage vs reading
  - Adding additional resources and reading
  - Project page on course website
  - Faster grading

Andrew’s Presentation

- Slides

DQN Overview

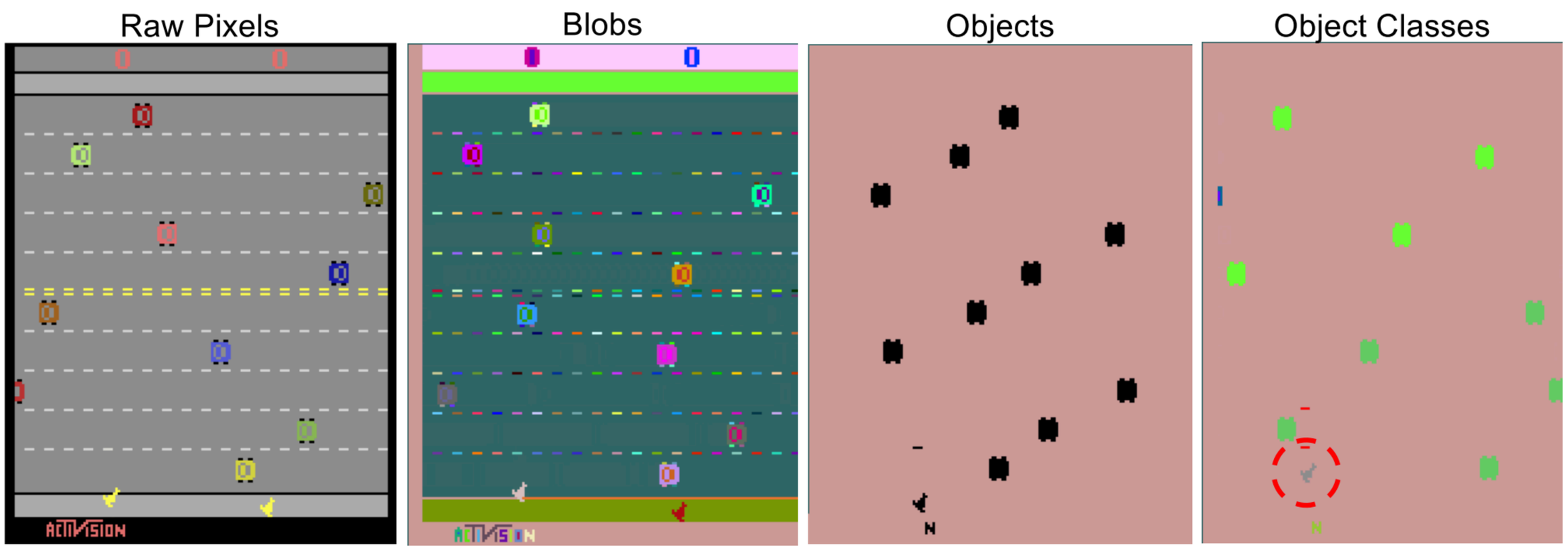
- Single algorithm, neural network architecture, and hyper-parameter setting that played 49 Atari video games at human-level.
- Training and evaluation is independent for each game.
  - The final neural network from training on “Breakout” cannot play “Pong.”
- Landmark result for deep reinforcement learning.

The Atari 57 Benchmark

- 57 Atari video games turned into RL benchmarks
- Why were Atari games hard for reinforcement learning algorithms?
  - Representation learning; hyper-parameter robustness
  - Prior state-of-the-art: neuroevolution and then Deepmind's predecessor to deep Q-learning.
- With a suitable representation, some games are simpler than others.

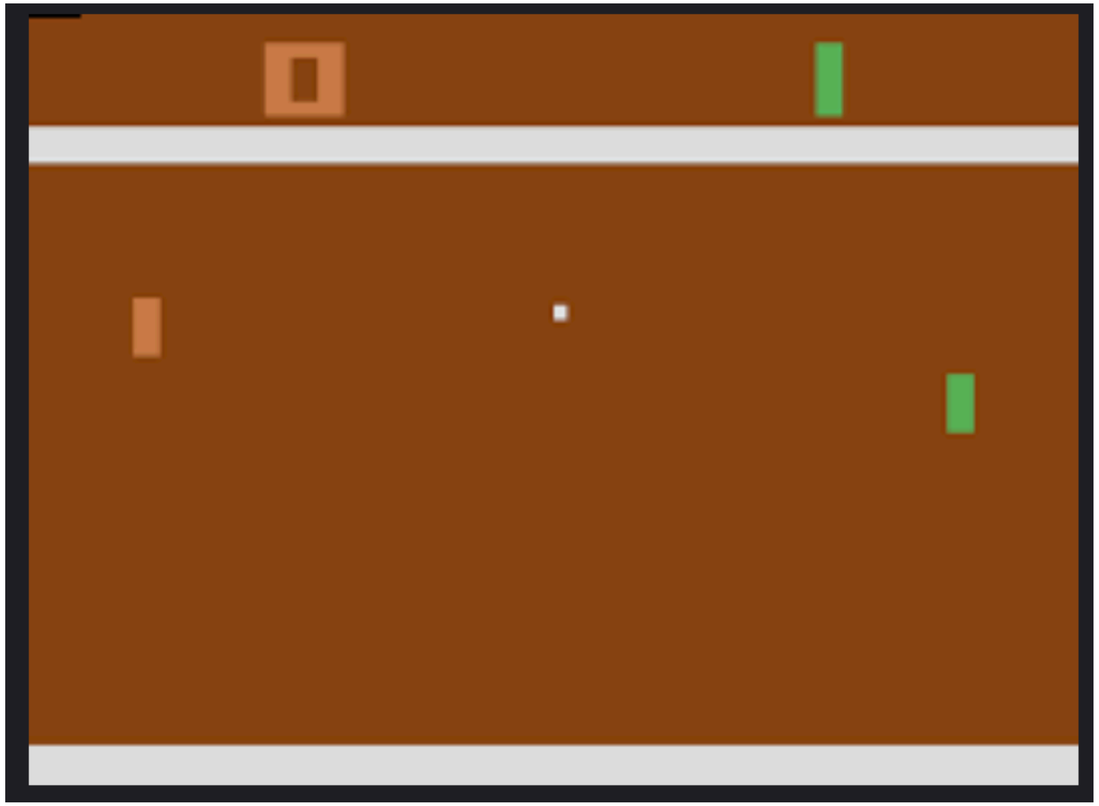
"HyperNEAT-GGP: A HyperNEAT-based Atari General Game Player." Hausknecht, Khandelwal, Miikkulainen, and Stone. 2012.

Feature Engineering



“HyperNEAT-GGP: A HyperNEAT-based Atari General Game Player.” Hausknecht, Khandelwal, Miikkulainen, and Stone. 2012.

Easy and Hard Games in Atari



Pong

Montezuma’s Revenge

DQN Architecture

- Core algorithm is semi-gradient Q-learning with a convolutional neural network as the function approximator.
- Key techniques for effective training across tasks:
  - Pre-processing
  - Experience replay
  - Target networks
  - Reward clipping

Pre-processing

- Large RGB images take a lot of memory.
  - Solution: downsample and turn the image to greyscale.
- Images are non-Markovian observations of state.
  - Solution: frame-stacking, i.e., concatenate past four frames together.
  - The agent repeats the same action for four consecutive frames and then can choose a new action.

Experience Replay

- The basic semi-gradient Q-learning algorithm processes (s, a, s', r) transitions as they are experienced and then discards them.
- Experience replay: keep around the most recent transitions (in DQN, the past 1 million) and use a random subset to update the action-value function.
  - Increased data-efficiency
  - Reduces correlation between samples.
- Other choices besides random subset can improve performance.

Prioritized Experience Replay. Schaul et al. 2015.

Target Networks

- The basic semi-gradient Q-learning algorithm always uses the most recent parameters to form the training target $R_{t+1} + \gamma \max_{a'} \hat{q}(S_{t+1}, a', \mathbf{w})$

- DQN uses a separate **target network** to compute $\gamma \max_{a'} \hat{q}(S_{t+1}, a', \tilde{\mathbf{w}})$.

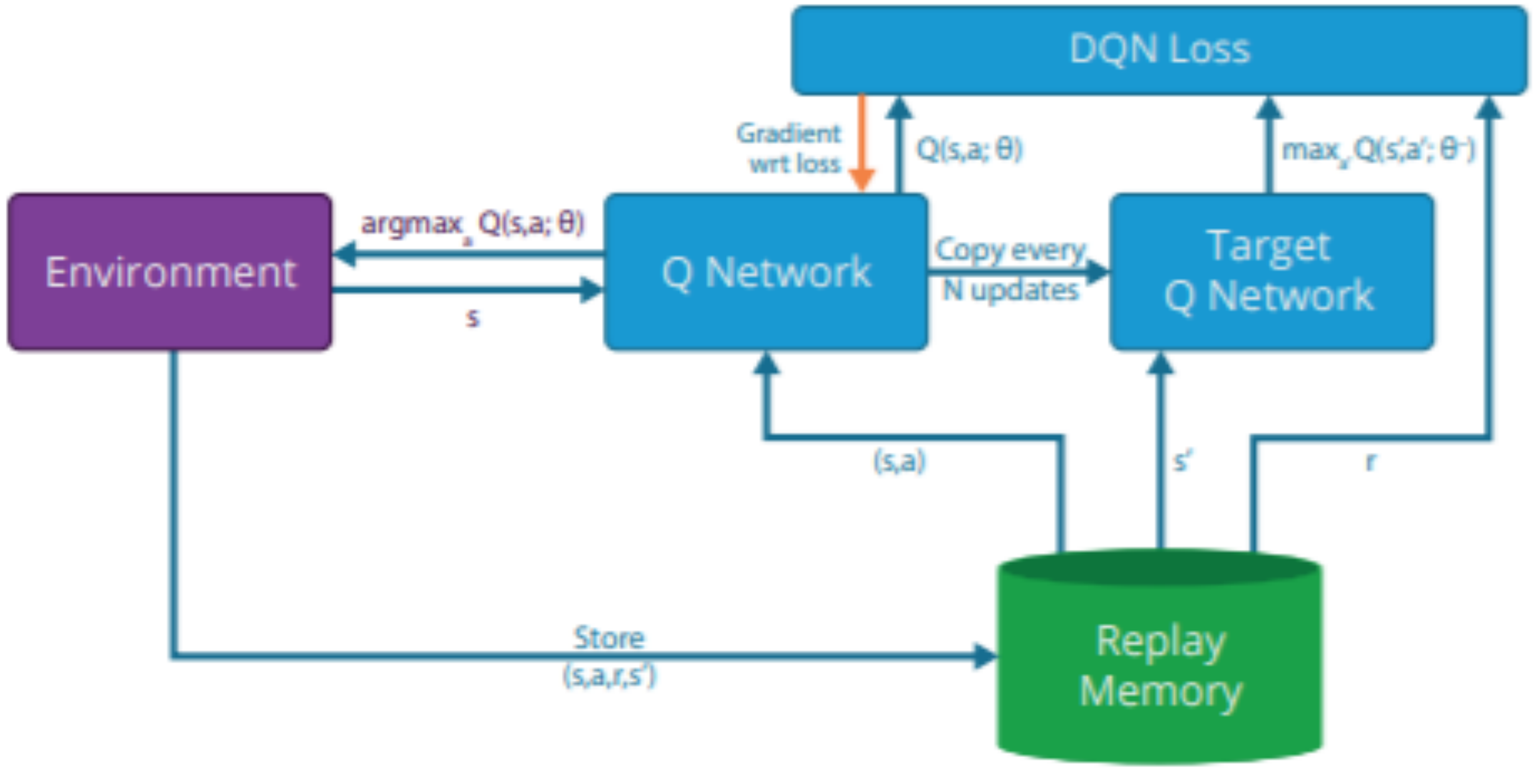
  - The target network is infrequently updated by setting the target network parameters to be the same as the main network's parameters.

  - Makes the learning target more stable as in supervised learning.

Reward clipping

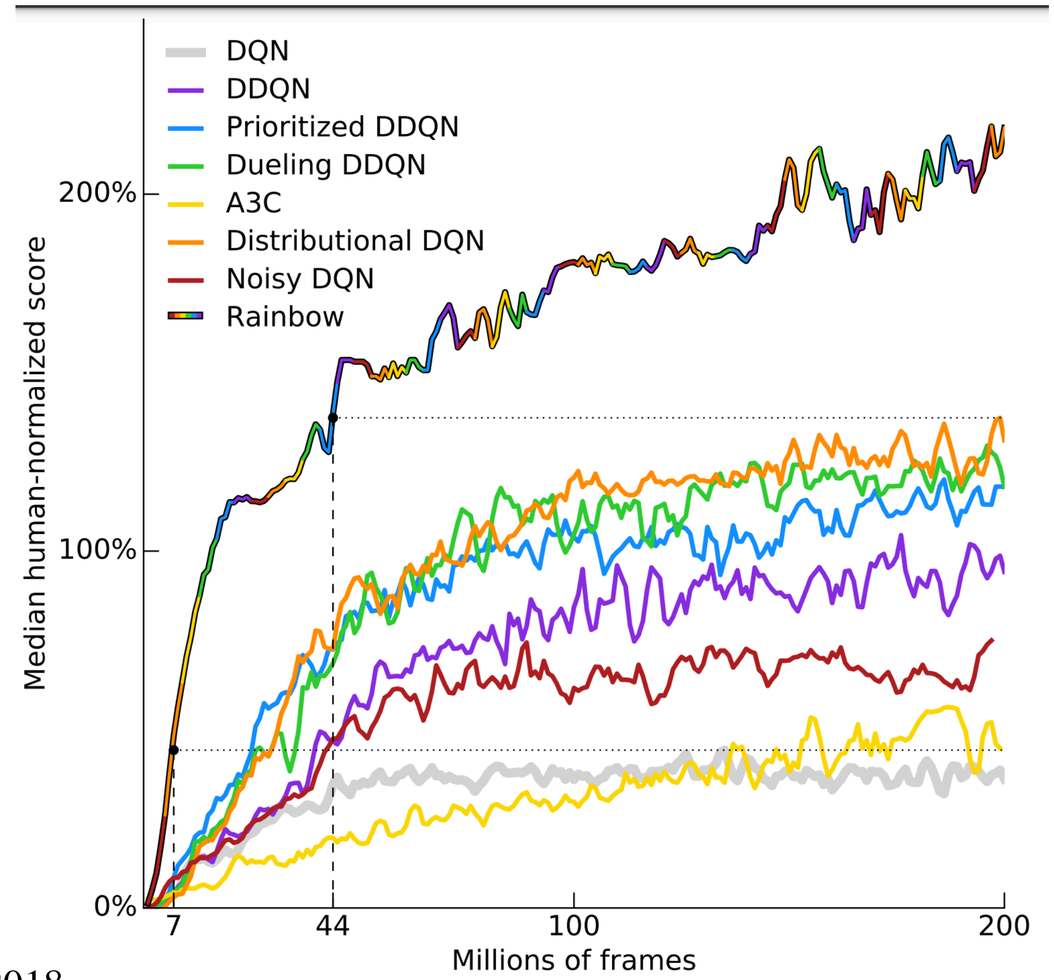
- Different Atari games have different reward magnitudes.
- Why is this a challenge?
  - Hard to tune a step-size that works across all games.
- Solution: clip rewards to be between -1 and 1.
-

DQN Architecture



Looking Forward

- DQN launched a surge of interest in deep reinforcement learning that has led to many exciting new applications and RL developments.
- DQN is widely used in practice though many improvements have been made.



Rainbow: Combining Improvements in Deep Reinforcement Learning. Hessel et al. 2018.
https://www.deepmind.com/blog/agent57-outperforming-the-human-atari-benchmark

Yixuan’s Presentation

- Slides

Summary

- Deep reinforcement learning is not just deep learning + RL.
  - It’s often deep learning + RL + new techniques and tricks for stability.
- This week focused on deep value-based RL.
  - Deep networks can be used for model-based RL.
  - Deep networks can represent policies in policy gradient RL.

Action Items

- Literature review due **tonight!**
- Begin reading chapter 13 (policy gradients)
- Homework 4 has been released.