# Deep Reinforcement Learning for Computer Games

Sean Yin, Chung-En Yu

Dec 13, 2022

University of Minnesota
Driven to Discover®

# Motivation / Goal

## Motivation

Neural Networks + Reinforcement learning → DQN, AC

The first deep reinforcement learning is to play Atari game.

Playing Atari with Deep Reinforcement Learning, Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller

## Goal

Implement the code from scratch: build and train DRL agents based on Pytorch.

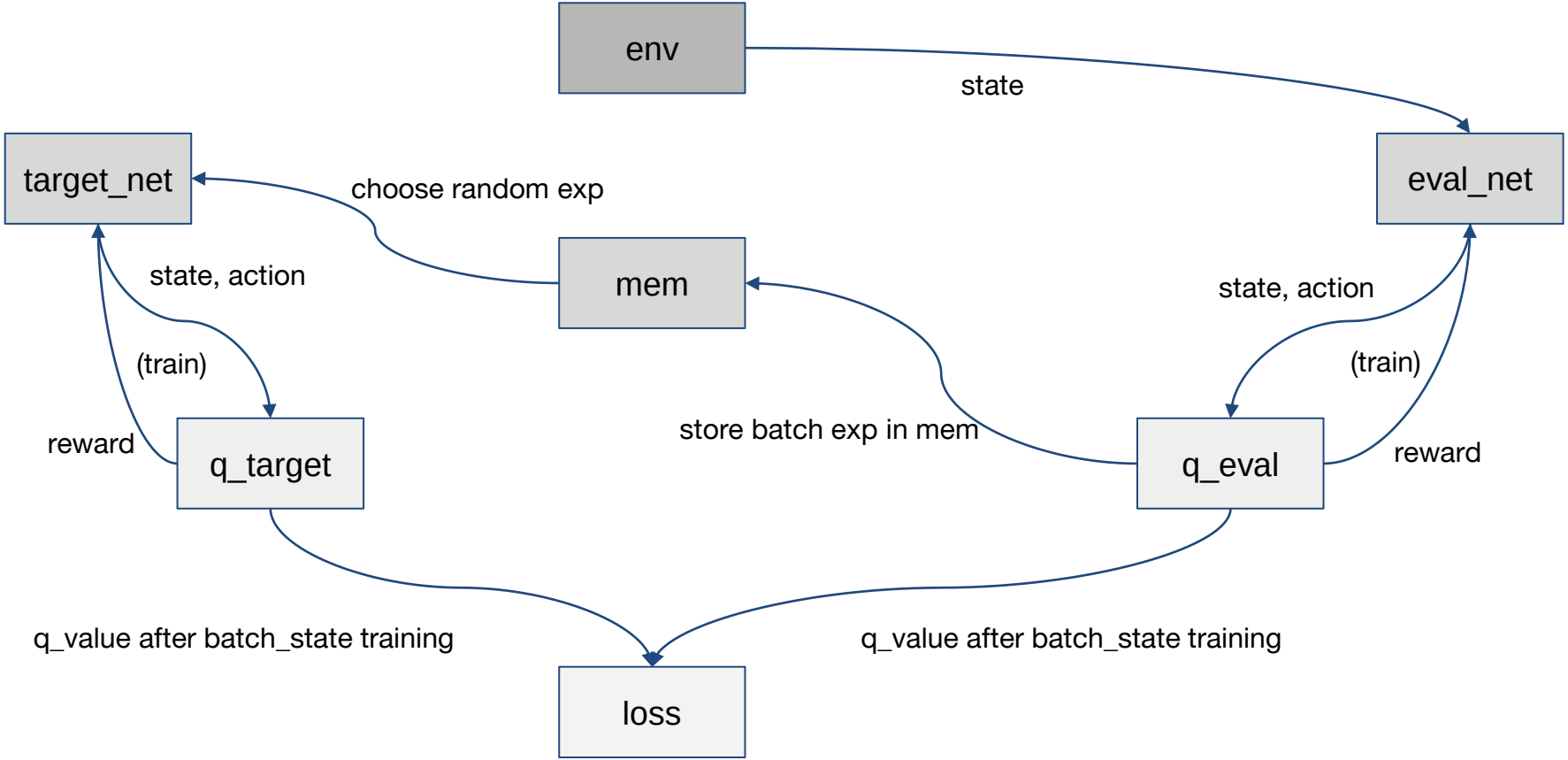
# Deep Reinforcement Learning

| On-Policy | Off-Policy |
| --- | --- |
| Policy Gradient,<br>Actor Critic (AC)<br>Advantage-Actor Critic (A2C) | DQN,<br>Double DQN<br>Proximal Policy Optimization (PPO) |

**Data**: (State, Policy, Reward, Next State) pairs along a trajectory.

**Off-policy**: can train agents based on data collected from different policy.

# Double DQN




* **exp** includes: state, action, reward

# AC

# Actor Critic

- Critic:

  Value Function served as a baseline to reduce the variance.
- Shared Common features map between actor and critic to reduce the bias of the critic.
- On-policy: PPO is an Off-policy version using importance sampling to deal with the policy distribution shift with the clip as the constraint for importance sampling.

# Simulation Setups

# Atari - Assault v4


0: NOOP
1: FIRE,
2: UP
3: RIGHT
4: LEFT
5: RIGHTFIRE
6: LEFTFIRE

| | |
|---|---|
| **State** | Image (210 x 160) x RGB (3) |
| **Action** | Discrete (7) |
| **Reward (lower, upper)** | Score (-inf, inf) |

# Results

# Fair Comparison

- Fixed Random Seed:
  Random seed of open ai gym environment and torch are both fixed for all agents.
- Fixed Feature Map Network Architecture:
  Feature Map Layers are fixed for all agents.
- Fixed Learning Rate, Optimizer, Horizon
- Except for the difference between algorithms, we fixed every hyper-parameters for every agent to make a fair comparison.

# Results - Double DQN vs. AC


**Double-DQN**


**AC**

- No Significant Difference between Agents, even though they both seem to learn the optimal policy.

Resources:
RTX 3080 10 G

# Conclusion

# Conclusion

- **Project novelty and difficulty:**
  We built and trained two of the most popular deep reinforcement learning algorithm from scratch.

- **Demo:**
  Two type of agents are trained to learn the optimal policy.
  The losses converge for both agents
  No Significant Difference between Agents even when we test on 1000 testing episodes with trained agents.

# Q&A

# UNIVERSITY OF MINNESOTA

## Driven to Discover®

Crookston Duluth Morris Rochester Twin Cities