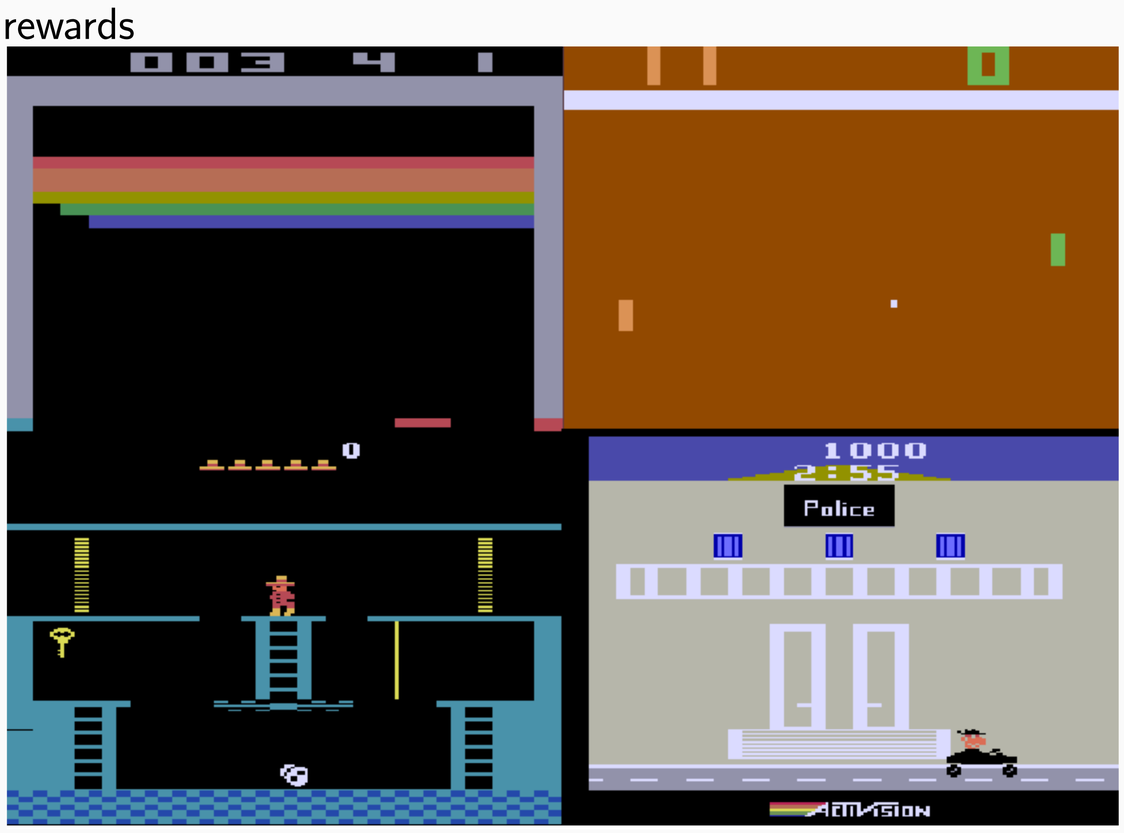
Playing Atari games with deep reinforcement learning and human checkpoint replay

Ionel-Alexandru Hosu, Traian Rebedea
Computer Science Department, Faculty of Automatic Control and Computers, Politehnica University of Bucharest

Problem Description

- Construct an architecture that is able to play any video game on the Atari 2600 platform
- Achieved for the simpler games
- Focus on two of the most difficult games, Montezuma's Revenge and Private Eye
- The biggest obstacle: Difficult exploration due to sparse rewards



Human Checkpoint Replay

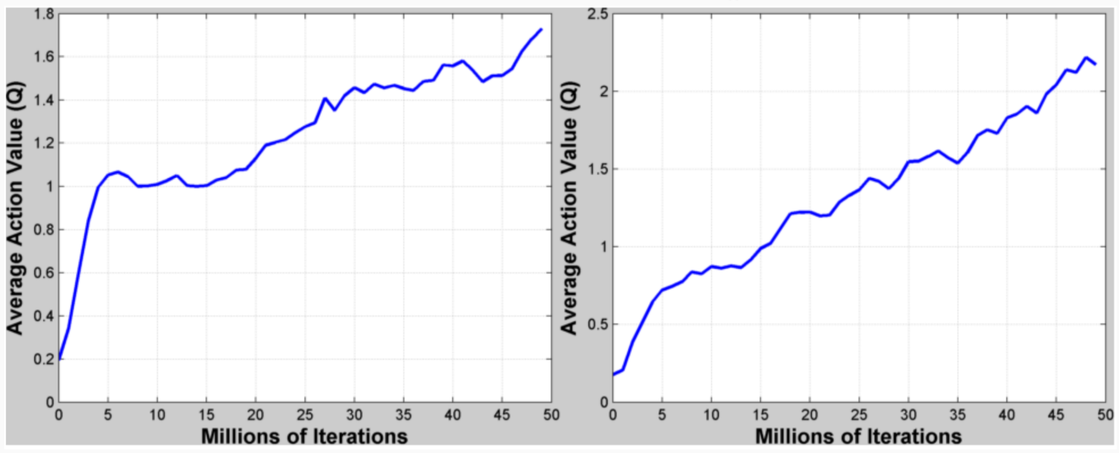
- Generated 100 checkpoints from a human players experience for each game
- Used them as starting points for the environment at training time.
- Main idea: The approach should make exploration less difficult
- Other checkpoints used for testing, in order to evaluate the agent's ability to generalize to unseen contexts

Human Experience Replay

- We also proposed training a deep reinforcement learning agent using offline human experience, combined with online agent experience
- This is meant to provide the agent with training samples that result in a positive reward, therefore making learning possible in environments that feature a sparse reward signal
- The training process consists of repeatedly sampling a minibatch composed of both human transitions and agent transitions

Experiments

- Double DQN Architecture
- Trained for 50 million iterations on each game



Results

- Evaluation: average score obtained after 100 episodes or 30 minutes or gameplay.
- Performed from a random human-generated checkpoint, not from the beginning of the game.

	Random Agent	HCR-DDQN	HER-DDQN
Montezuma's Revenge	177.1	379.1	218
Private Eye	–41	1264.4	N/A

Current Work

- Experiments with intrinsic motivation: integrating a form of artificial curiosity in the architecture of the agent, in order for it to learn to explore the relevant parts of the space
- Imitation learning: learning to follow a game trajectory that was generated by a human player
- Combination of both, which would resemble more the way humans learn to solve a task: watching someone else do it in order to get a good grasp of it, and after that performing it with slight variations.

	DeepMind (2017)	DeepMind (2018)	OpenAI (2018)	Ours
Montezuma's Revenge	4,739.6	41,098.4	74,500	8,400
Private Eye	40,908.2	98,763.2	N/A	N/A

Remaining Challenges

- Sample inefficiency
- Exploration
- State representation

Politehnica University of Bucharest