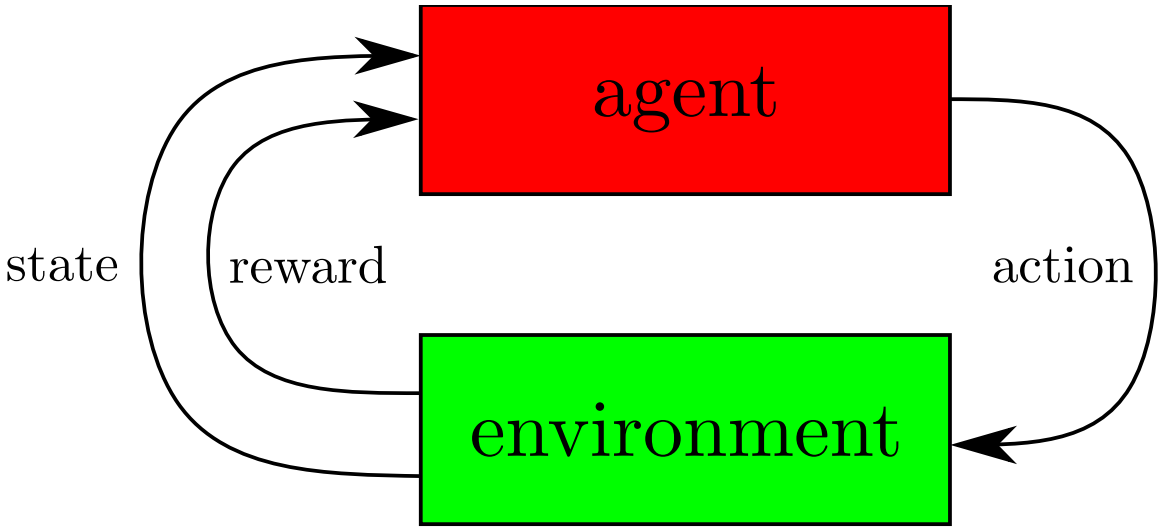
Deep Q-network

Ondřej Podsztavek

Let's Talk ML Prague, 14. 12. 2017



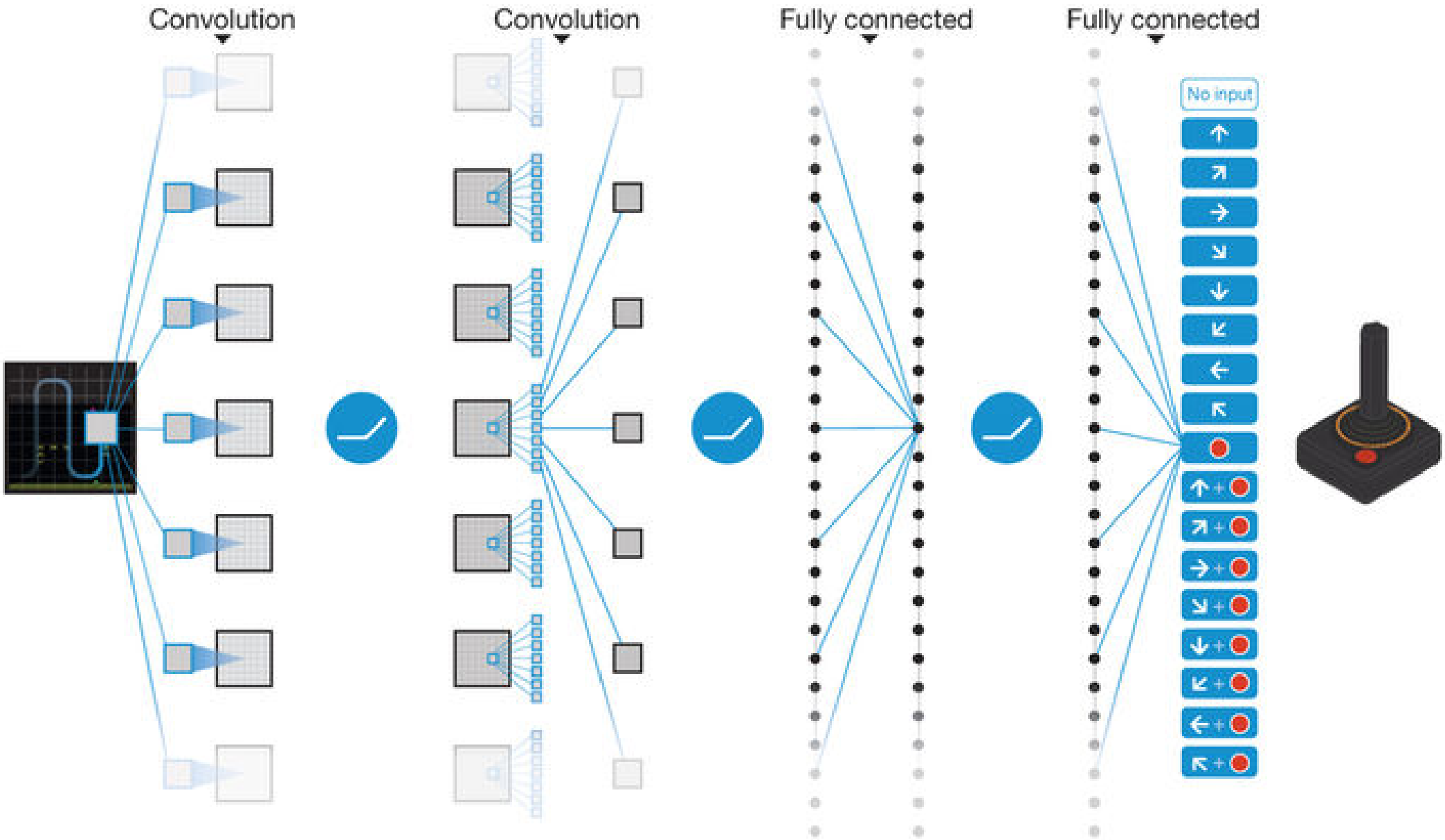
Diverse Array of Tasks

- 49 Atari 2600 games
- the same network architecture and hyperparameters
- pixels and the game score as inputs

Q function

$$Q^*(s,a) = max_\pi \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | s_t = s, a_t = a, \pi]$$

Deep Convolutional Network

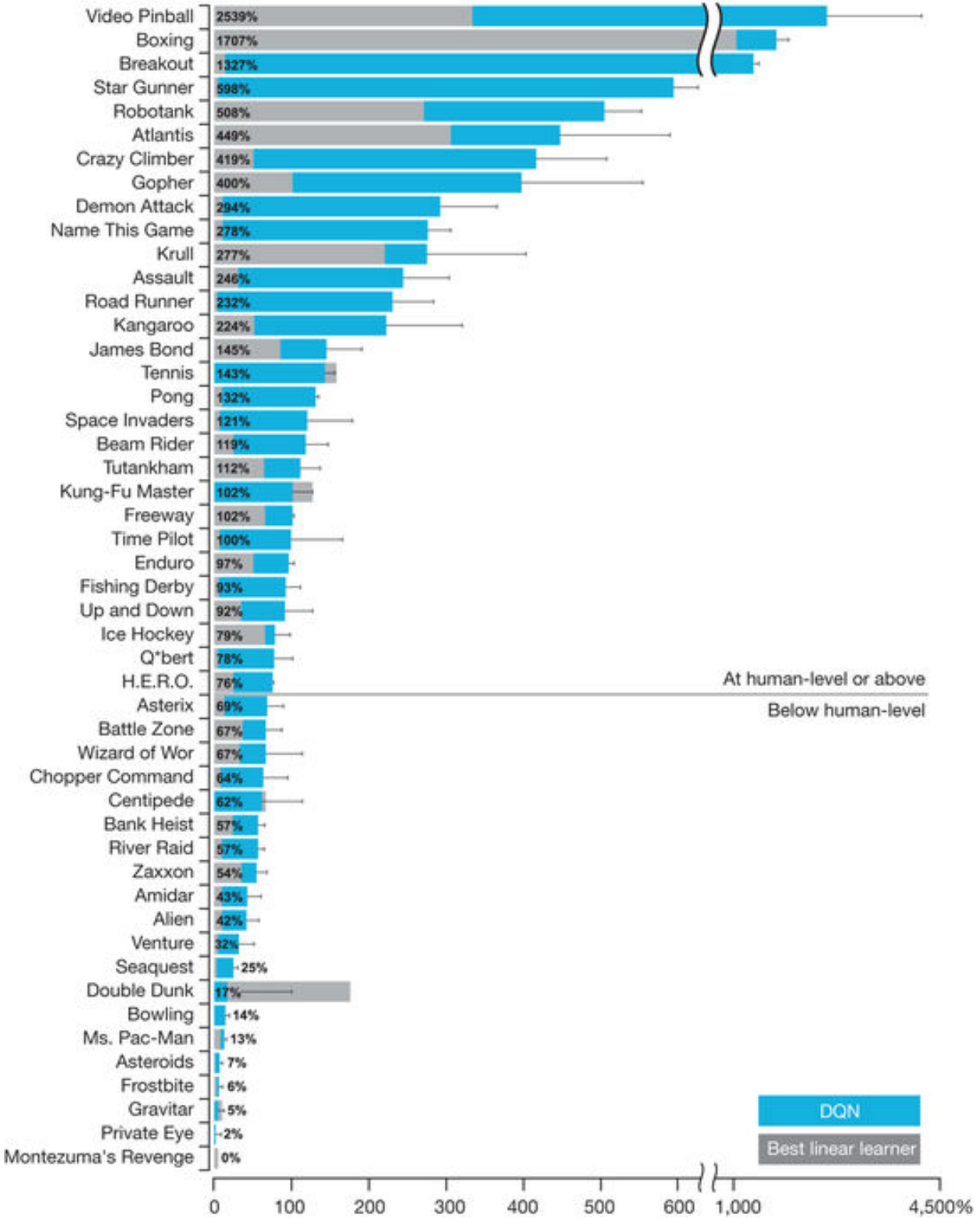


Training Algorithm Modifications

- RL is known to be unstable for a nonlinear function approximator
- separate network for generating the targets for the Q-learning update
- experience replay

$$e_t = (s_t, a_t, r_t, s_{t+1})$$

$$D_t = \{e_1, \ldots, e_t\}$$


Video Pinball 2539%
Boxing 1707%
Breakout 1327%
Star Gunner 598%
Robotank 508%
Atlantis 449%
Crazy Climber 419%
Gopher 400%
Demon Attack 294%
Name This Game 278%
Krull 277%
Assault 246%
Road Runner 232%
Kangaroo 224%
James Bond 145%
Tennis 143%
Pong 132%
Space Invaders 121%
Beam Rider 119%
Tutankham 112%
Kung-Fu Master 102%
Freeway 102%
Time Pilot 100%
Enduro 97%
Fishing Derby 93%
Up and Down 92%
Ice Hockey 79%
Q*bert 78%
H.E.R.O. 76%
Asterix 69%
Battle Zone 67%
Wizard of Wor 67%
Chopper Command 64%
Centipede 62%
Bank Heist 57%
River Raid 57%
Zaxxon 54%
Amidar 43%
Alien 42%
Venture 32%
Seaquest 25%
Double Dunk 17%
Bowling 14%
Ms. Pac-Man 13%
Asteroids 7%
Frostbite 6%
Gravitar 5%
Private Eye 2%
Montezuma's Revenge 0%
At human-level or above
Below human-level
DQN
Best linear learner
0
100
200
300
400
500
600
1,000
4,500%

Variants

- *Double Q-learning*, corrects tendency to overestimate action's values
- *Prioritized Replay*, somehow improves the experience replay
- *Dueling DQN*, one estimates the value at a time step, other calculates advantages of an action

References

- Human-level control through deep reinforcement learning, https://www.nature.com/articles/nature14236
- Deep Reinforcement Learning with Double Q-learning, https://arxiv.org/abs/1509.06461
- Learning from Delayed Rewards, http://www.cs.rhul.ac.uk/~chrisw/thesis.html
- OpenAI Baselines: DQN, https://blog.openai.com/openai-baselines-dqn/