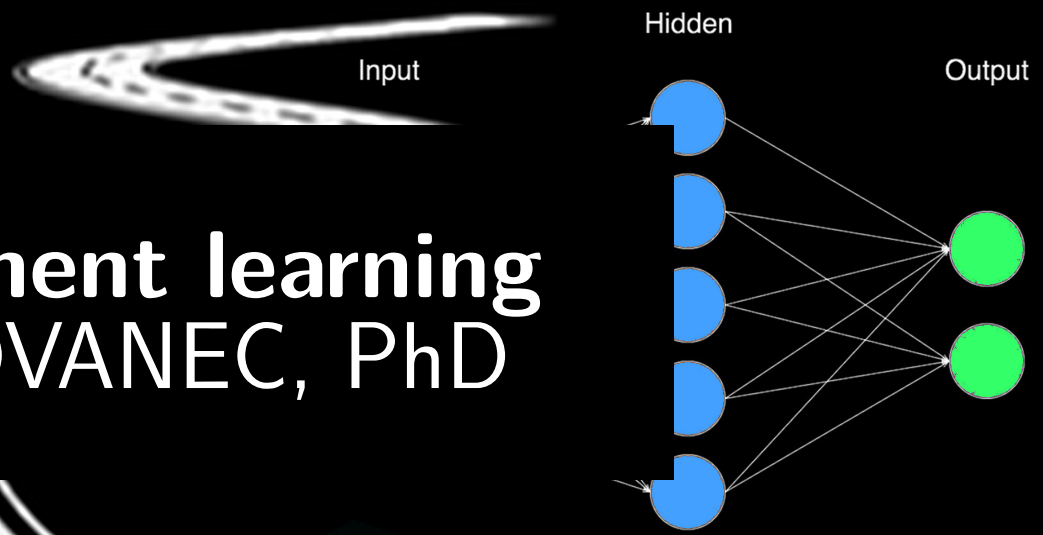
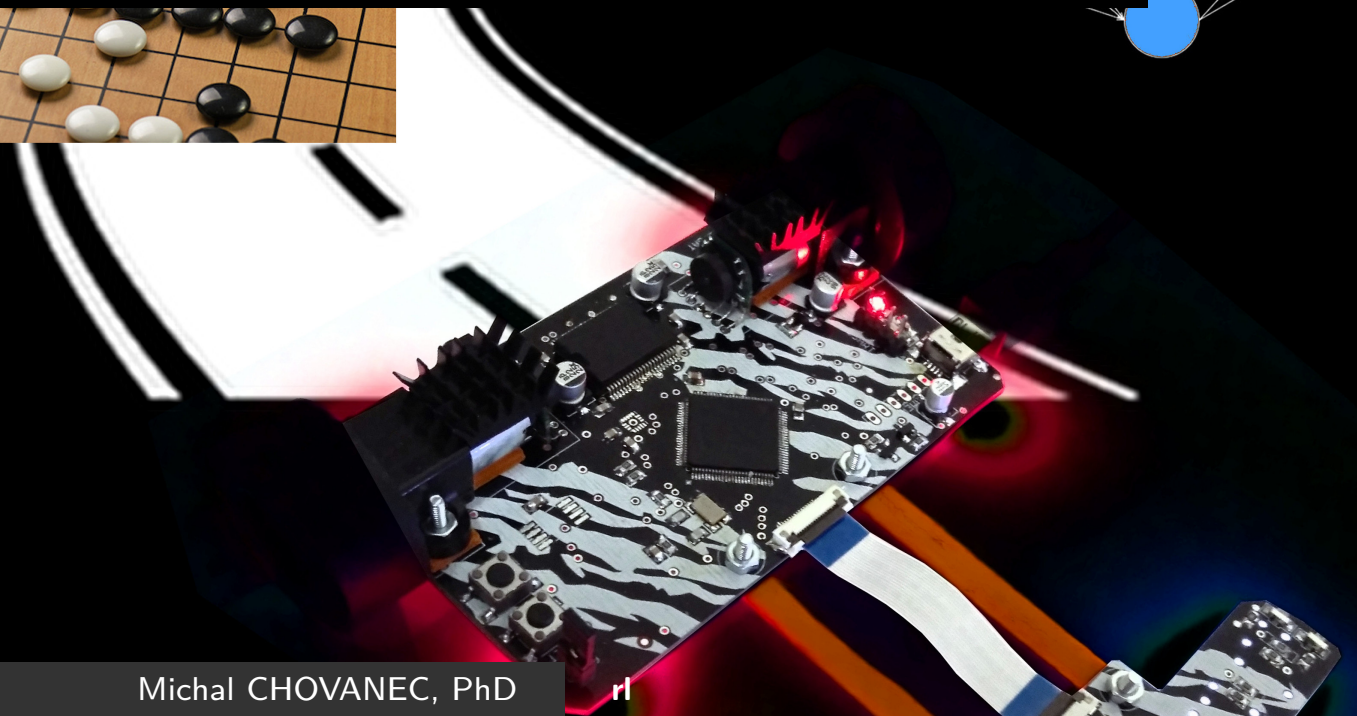
$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_{t+1} + \lambda \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\right]$$

(The New Action Value = The Old Value) + The Learning Rate × (The New Information − the Old Information)

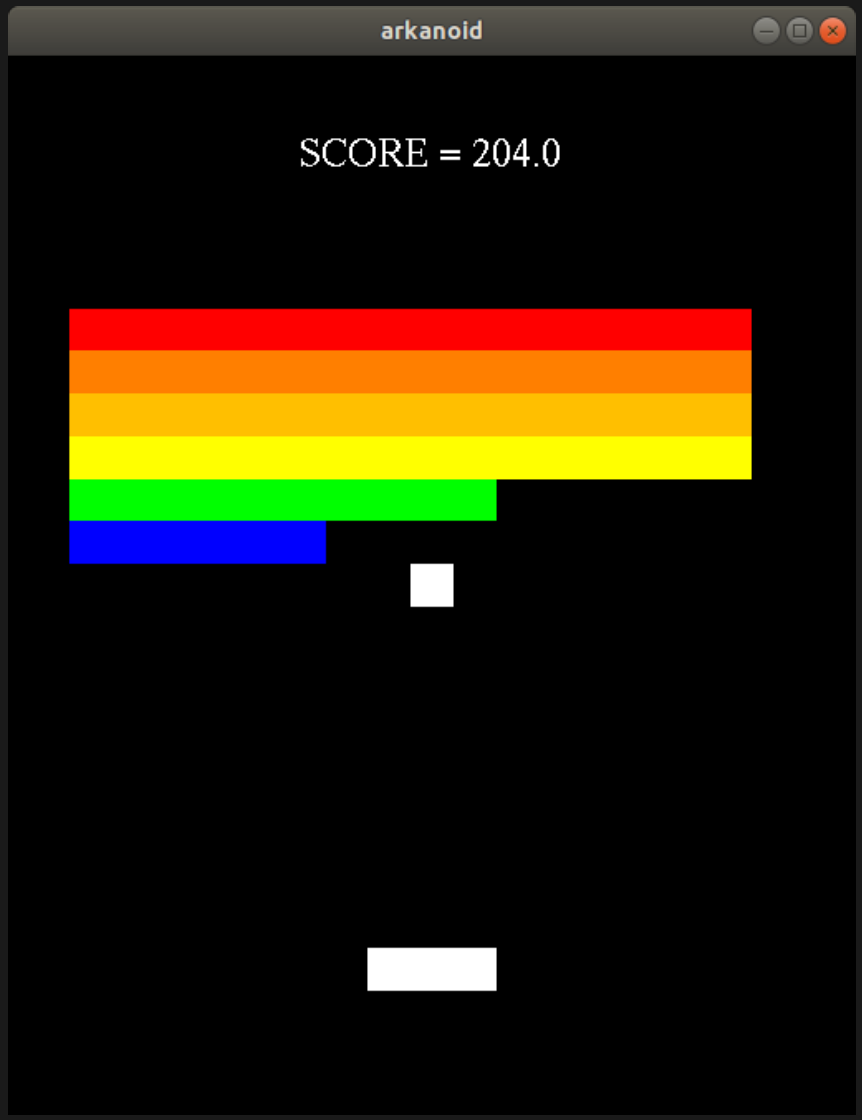
Deep Reinforcement learning

Michal CHOVANEC, PhD





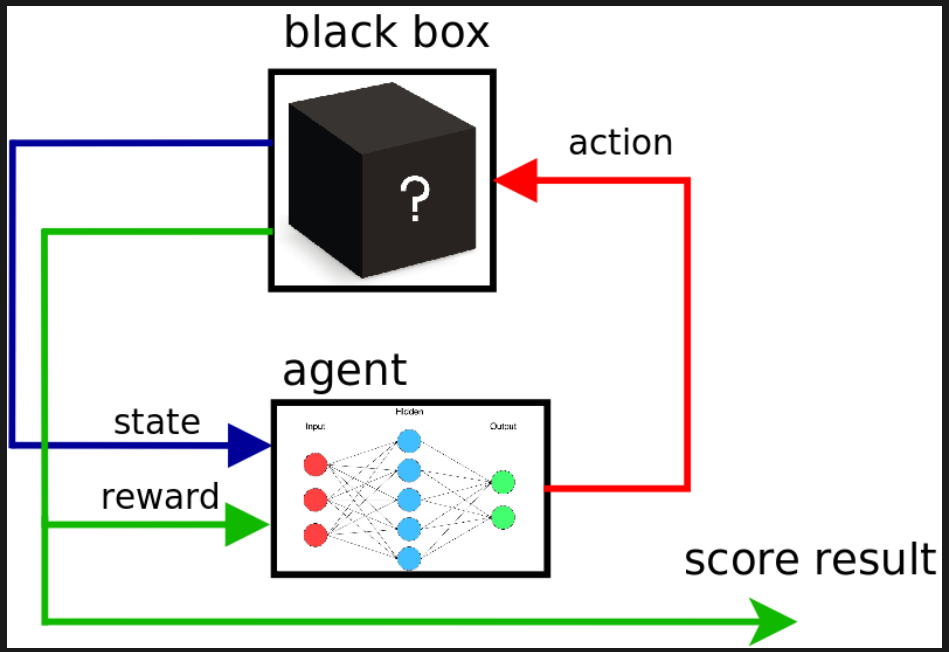
Atari Bbreakout (arkanoid)



- **state** : screen pixels 16x20x3
- **actions** : left, right, no-move
- **rewards** :
  - hit $+0.5$
  - out -1.0
  - win $+1.0$
- **learn from experiences**, $Q(s, a)$

Reinforcement learning

- learning from punishments and rewards
  - obtain **state**
  - choose **action**
  - **execute** action
  - obtain **reward**
  - learn from **experiences**, $Q(s, a)$



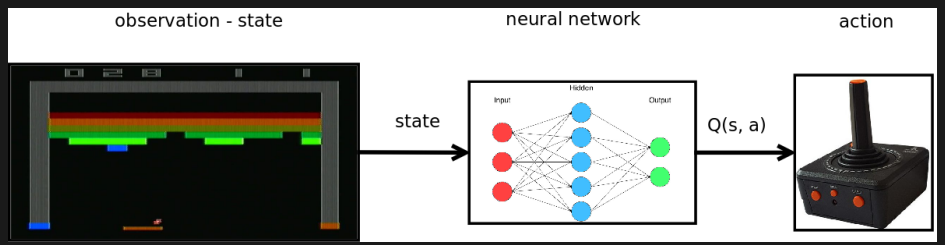
What is $Q(s, a)$

- $Q(s, a)$ value of action a, executed in state s
- Q-learning algorithm

$$Q'(s, a) = R(s, a) + \gamma \max_{a'} Q(s', a')$$

for real problems numbers of states is **too high**

- chess 10^{120}, go 10^{180}, starcraft 10^{500}
- atoms in observable universe 10^{80}
- neural networks - **deep Q network**



deep Q network - GO playing network example

total **412 121 216** computing operations

approx. **8 000 000** parameters to learn

layer	net 3	net 5
0	dense conv 5x5x32	dense conv 3x3x32
1	dense conv 5x5x32	dense conv 3x3x32
2	dense conv 5x5x32	dense conv 3x3x32
3	dense conv 5x5x32	dense conv 3x3x32
4	conv 1x1x32	dense conv 3x3x32
5	dense conv 5x5x32	dense conv 3x3x32
6	dense conv 5x5x32	dense conv 3x3x32
7	dense conv 5x5x32	dense conv 3x3x32
8	dense conv 5x5x32	conv 1x1x56
9	conv 1x1x32	dense conv 3x3x32
10	dense conv 5x5x32	dense conv 3x3x32
11	dense conv 5x5x32	dense conv 3x3x32
12	dense conv 5x5x32	dense conv 3x3x32
13	dense conv 5x5x32	dense conv 3x3x32
14	conv 1x1x32	dense conv 3x3x32
15	dense conv 5x5x32	dense conv 3x3x32
16	dense conv 5x5x32	dense conv 3x3x32
17	dense conv 5x5x32	conv 1x1x56
18	dense conv 5x5x32	dense conv 3x3x32
19	conv 5x5x64	dense conv 3x3x32
20	fc 362	dense conv 3x3x32
21		dense conv 3x3x32
22		dense conv 3x3x32
23		dense conv 3x3x32
24		dense conv 3x3x32
25		dense conv 3x3x32
26		conv 1x1x64
27		fc 362

Playing GO (October 2017)

- **supervised training** - train game using Masters games
- **reinforcement learning** - let play two networks against each other

Q&A

michal chovanec (michal.nand@gmail.com)

www.youtube.com/channel/UCzVvP2ou8v3afNiVrPAHQGg

github https://github.com/michalnand