

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

(The New Action Value = The Old Value) + The Learning Rate \times (The New Information — the Old Information)



AlphaGo, AlphaZero, AlphaStar - Google DeepMind

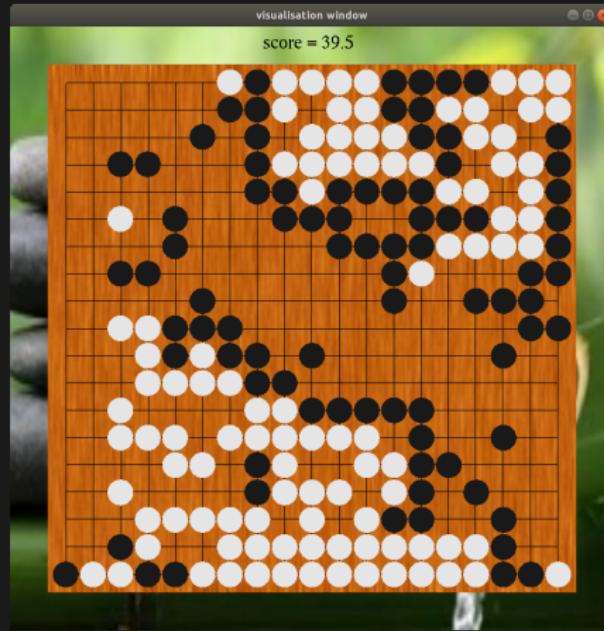
Michal CHOVANEC, PhD



End of Deep learning middle age

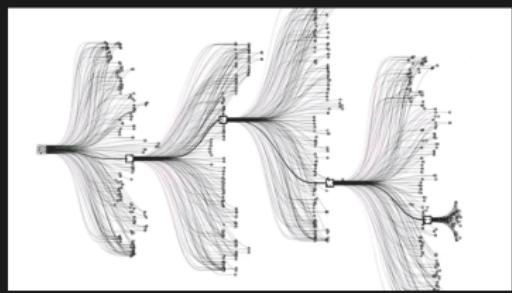
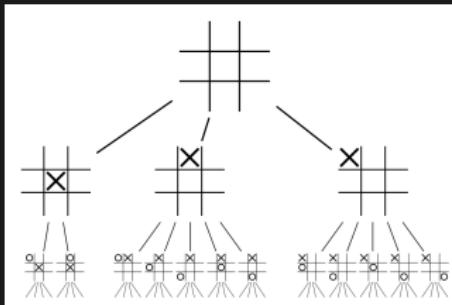
- **GO** : AlphaGO vs Lee Sedol, 4 - 1, 9th .. 15th Match 2016
https://deepmind.com/documents/119/agz_unformatted_nature.pdf
- **Chess** : AlphaZero vs StockFish, W290 - D886 - L24,
December 2017 <https://arxiv.org/pdf/1712.01815.pdf>
- **StarCraft** : AlphaStar vs MaNa, 5-0, 19th December 2019
<https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>

GO



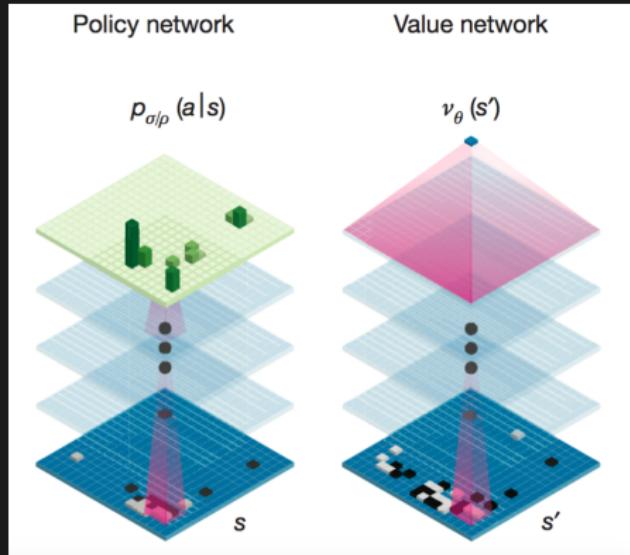
Game complexity

- 10^{170} legal positions, chess 10^{120} , atoms in universe 10^{80}
- 200 moves per position (average) (20 in chess)
- best (old) programs using MonteCarlo tree search
 - can't beat professional Dan players



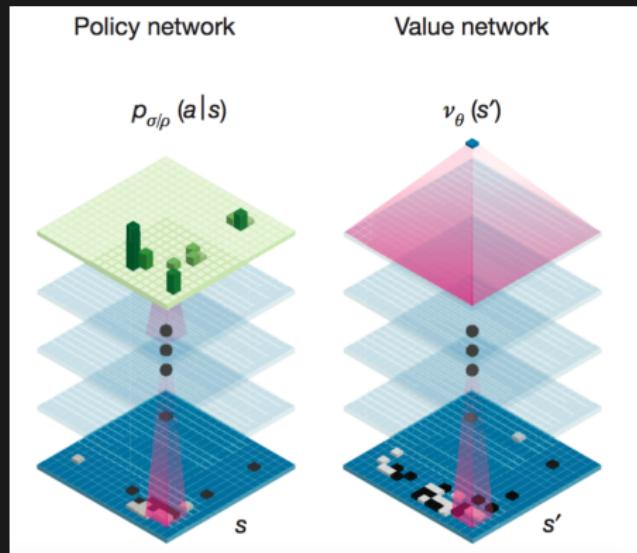
AlphaGo architecture

- Two neural networks
 - Policy network, Value network
- Policy network - give probabilities of good / wrong moves
- Value network - give probability of winning / loosing



AlphaGo architecture

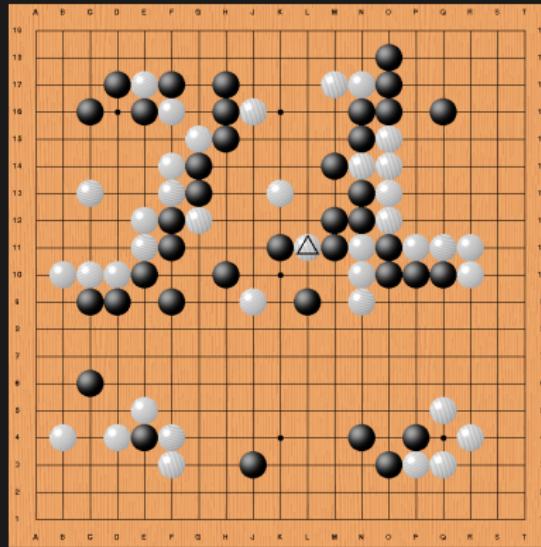
- Two neural networks
 - Policy network, Value network
- Policy network - give probabilities of good / wrong moves
- Value network - give probability of winning / loosing



- **input**
 - stones positions with feature planes
 - black, white, legal positions
 - actual player on move
 - ...
- Tensor $19 \times 19 \times \text{feature_planes}$
- **output**
 - Policy network 19×19 probabilities - Value network 2 black and white win probabilities

Game 4, Lee Sedol wins

Move 78 - "divine move" Gu Li, 15% win drop



My GO net

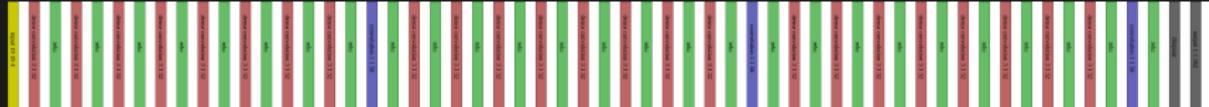


- **supervised training** - train game using Masters games
- **reinforcement learning** - let play two networks against each other

Network architecture

we need to go much deeper for GO

- **27 convolutional layers**
3 blocks with 8 dense conv + 1 conv layer
- **input**
4 matrices 19×19 : black stones, white stones, empty fields, active player
- **output**
recommended moves $19 \times 19 + 1$ for pass = 362 outputs



Network architecture

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

Supervised results

89	60	56	61	62	60	61	60	57	66	67	62	64	60	61	57	55	40	82
46	55	60	66	59	54	56	60	61	51	52	55	56	63	53	60	59	58	38
59	52	62	81	55	60	49	49	52	54	48	49	52	64	70	64	63	62	56
61	57	66	83	62	55	50	53	47	41	46	48	49	56	57	63	51	60	52
70	55	68	57	55	53	56	56	54	55	54	53	57	53	53	60	51	57	62
53	57	62	62	55	53	56	59	50	60	49	57	49	54	60	60	65	55	58
69	51	59	57	55	49	55	52	55	48	51	56	51	52	53	58	47	55	57
66	52	55	58	58	54	54	51	56	54	53	56	53	58	56	51	55	60	48
68	62	51	50	56	55	51	55	58	57	57	55	54	55	57	48	55	56	51
61	53	57	48	51	57	57	54	56	60	59	55	56	57	57	46	59	55	61
66	60	52	51	54	55	53	52	53	53	55	55	56	53	59	51	52	52	61
67	59	54	55	52	57	52	58	59	56	56	59	56	48	56	52	52	55	61
67	52	50	55	57	54	56	55	51	53	53	56	56	54	57	64	59	56	54
61	56	68	65	53	52	52	59	55	50	58	57	57	56	56	62	70	61	58
63	57	67	63	58	53	53	55	55	62	59	54	51	51	54	63	60	52	62
56	60	69	82	60	60	55	55	49	38	53	54	54	55	52	73	78	60	58
61	63	65	76	58	70	55	55	50	57	50	51	51	62	69	78	65	66	57
69	58	65	71	57	59	59	56	56	52	57	58	58	57	60	65	57	63	50
90	55	69	57	64	58	56	59	58	69	56	55	67	57	52	65	58	58	86
															0			

Q&A



michal chovanec (michal.nand@gmail.com)
www.youtube.com/channel/UCzVvP2ou8v3afNiVrPAHQGg
github <https://github.com/michalnand>