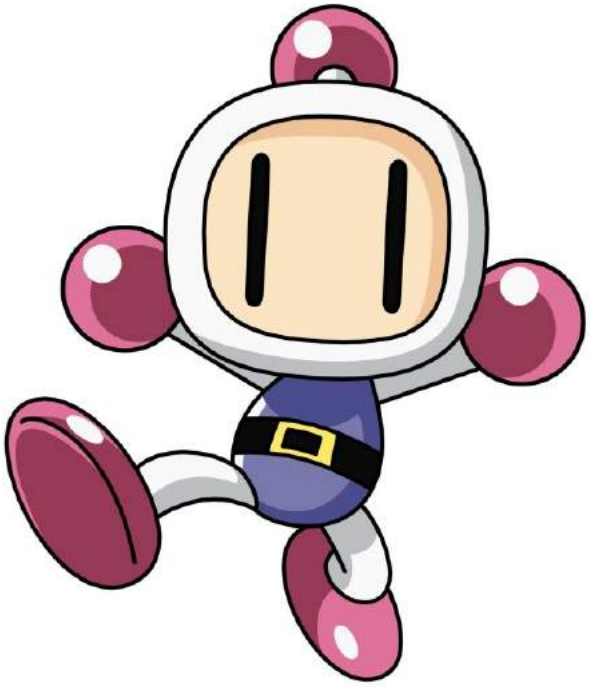
BomberMan Tournament



https://blogs.forbes.com/olliebarder/files/2018/06/bomberman_mizuno.jpg

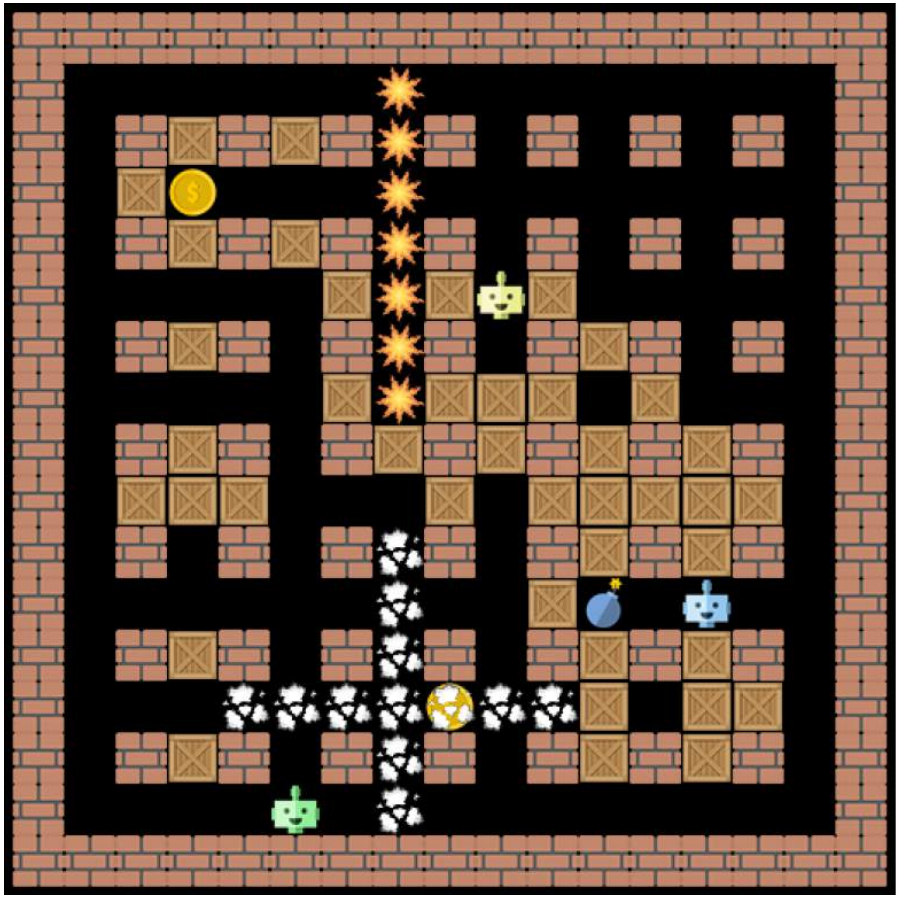
Mustafa Fuad Rifet Ibrahim

Overview

- Game + Tournament
- 3 Agents

Game + Tournament

- Environment (crates, walls, coins,...)
- Agents (move, drop bombs)
- score

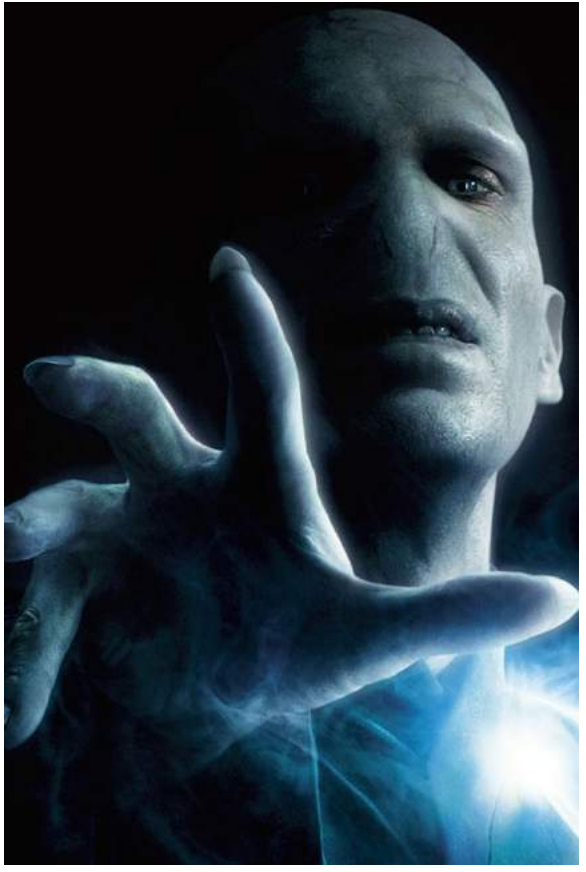


Game + Tournament

- Multiple eps
- 400 steps
- 0.5s thinking time
- Slow thinking penalty
- No multiprocessing
- One core of i7-8700K, up to 8GB RAM
- Simple agents

http://randomhoohaas.flyingomelette.com/bomb/gba-st/img/art-multiplay.JPG

Lord_Voldermort



https://cdn.wallpapersafari.com/6/42/hSU3zx.jpg

Lord_Voldermort

	action1	action2	action3	...
state1				
state2				
state3				
⋮				

$$Q(s,a) = (1-\eta)Q(s,a) + \underbrace{\eta}_{\text{learning rate}} \left[\underbrace{r}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \underbrace{\max_{a'} Q(s',a')}_{\text{estimated optimal future reward}} \right]$$

Lord_Voldermort

$$state = (Left, Up, Right, Down, Self, Self\_Bomb)$$

Lord_Voldermort

$$state = (Left, Up, Right, Down, Self, Self\_Bomb)$$

Self:

- Empty
- Danger
- Bomb

Lord_Voldermort

$$state = (Left, Up, Right, Down, Self, Self\_Bomb)$$

Self:
- Empty
- Danger
- Bomb

Self_Bomb:
- True
- False

Lord_Voldermort

$$state = (Left, Up, Right, Down, Self, Self\_Bomb)$$

Left,Up,Right,Down:

- Wall
- Enemy
- Crate
- Coins
- Bombs
- Danger
- Empty
- Priority

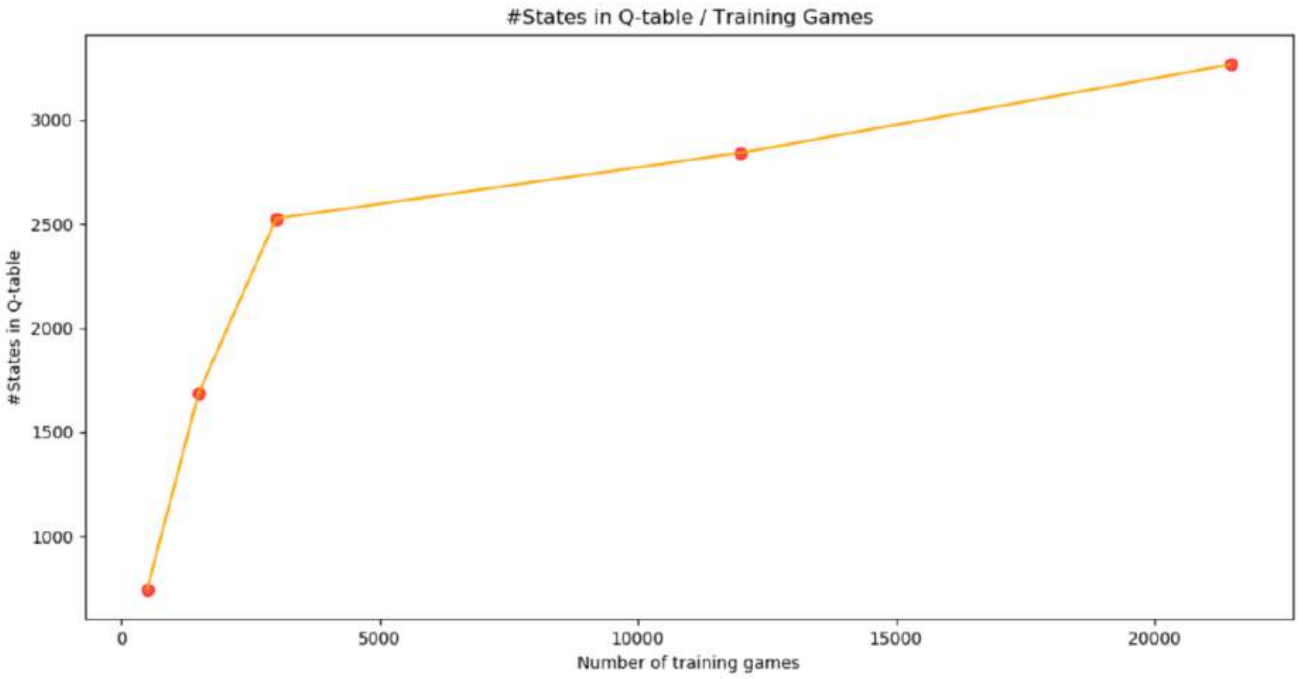
Self:

- Empty
- Danger
- Bomb

Self_Bomb:

- True
- False

Lord_Voldermort

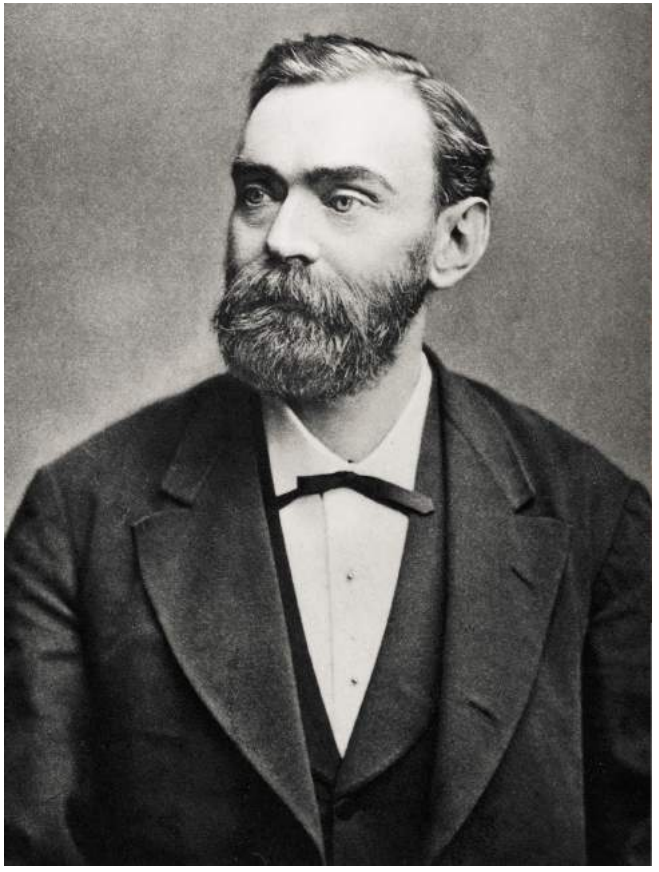


Lord_Voldermort

MATCH	ROUNDS	KILLS	SUICIDES	COINS	SCORE	RANK
Group B | 01	25	13	2	132	197	1
Group B | 02	25	24	3	129	249	1
Group B | 03	25	4	1	134	154	1
Group B | 04	25	27	0	221	356	1
Octofinal J	100	61	21	607	912	1
Quarterfinal R	100	35	22	297	472	1
Semifinal V	100	48	14	319	559	1

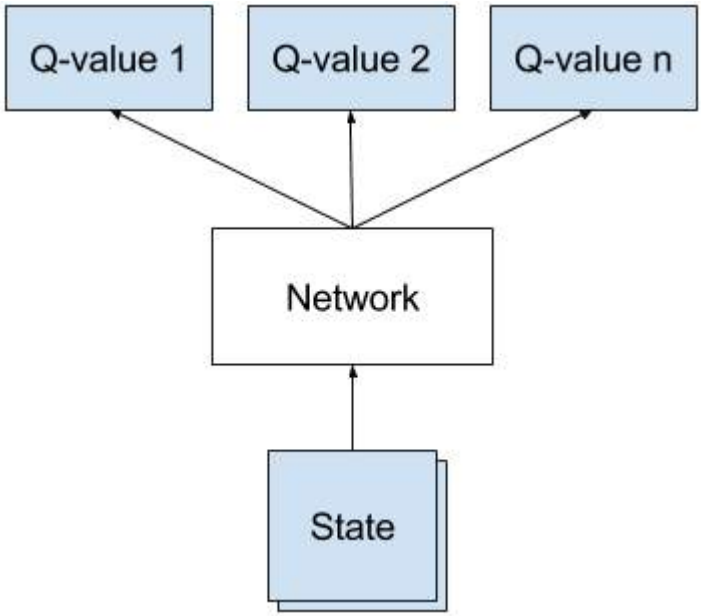
https://hci.iwr.uni-heidelberg.de/vislearn/HTML/teaching/courses/FML/bomberman_rl/agent.php?id=Lord_Voldemort

NOBEL



https://upload.wikimedia.org/wikipedia/commons/0/07/Alfred_Nobel3.jpg

NOBEL

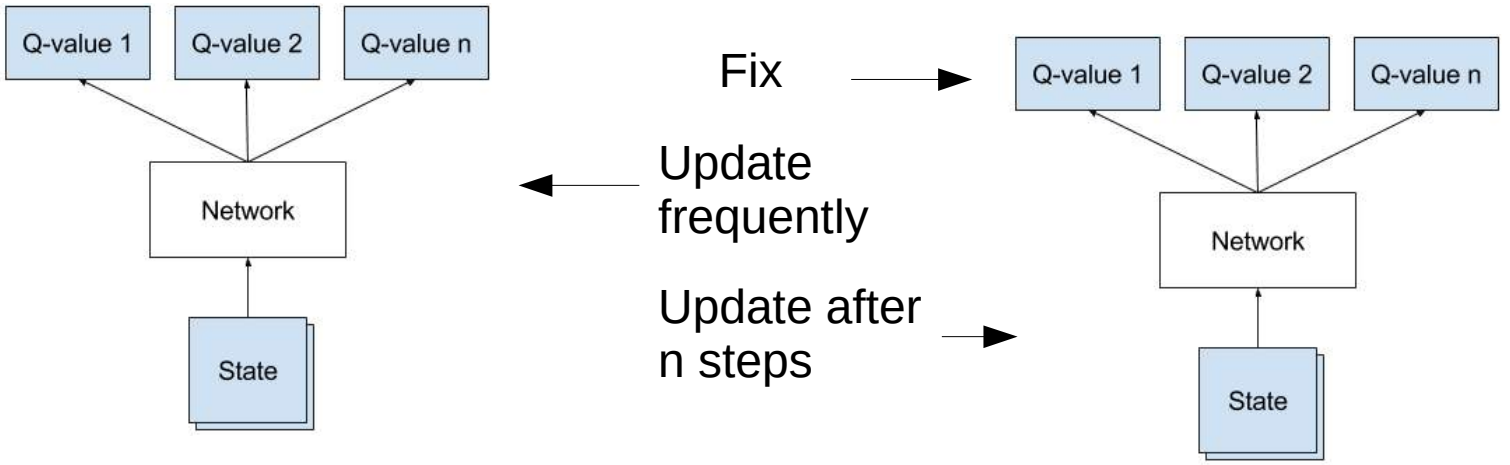


https://neuro.cs.ut.ee/wp-content/uploads/2015/12/dqn.png

$$L = \frac{1}{2}[\underbrace{r + \gamma max_{a'} Q(s', a')}_{\text{target}} - \underbrace{Q(s, a)}_{\text{prediction}}]^2$$

NOBEL

Fixed Q-Targets



https://neuro.cs.ut.ee/wp-content/uploads/2015/12/dqn.png

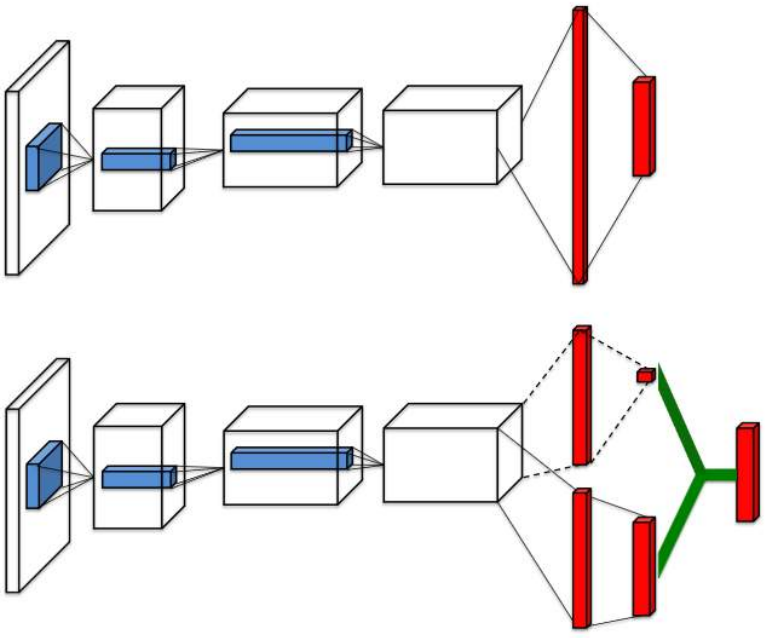
$$L = \frac{1}{2}[\underbrace{r + \gamma max_{a'} Q(s', a')}_{\text{target}} - \underbrace{Q(s, a)}_{\text{prediction}}]^2$$

NOBEL

Dueling DQN

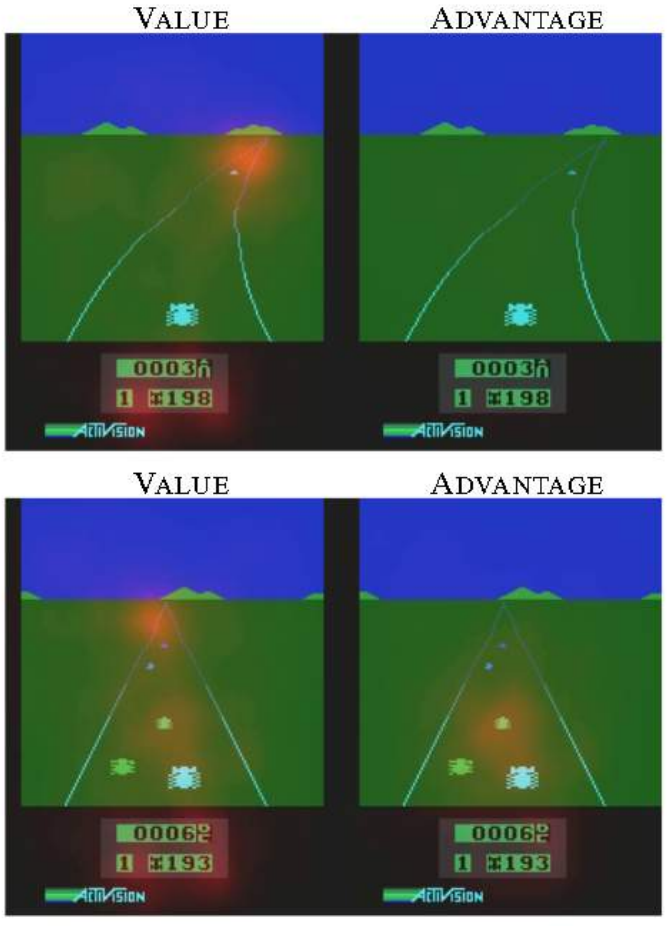
$$Q(s,a) = V(s) + [A(s,a) - \max_{a'} A(s,a')]$$

$$Q(s,a) = V(s) + [A(s,a) - \frac{1}{|\mathcal{A}|}\sum_{a'} A(s,a')]$$

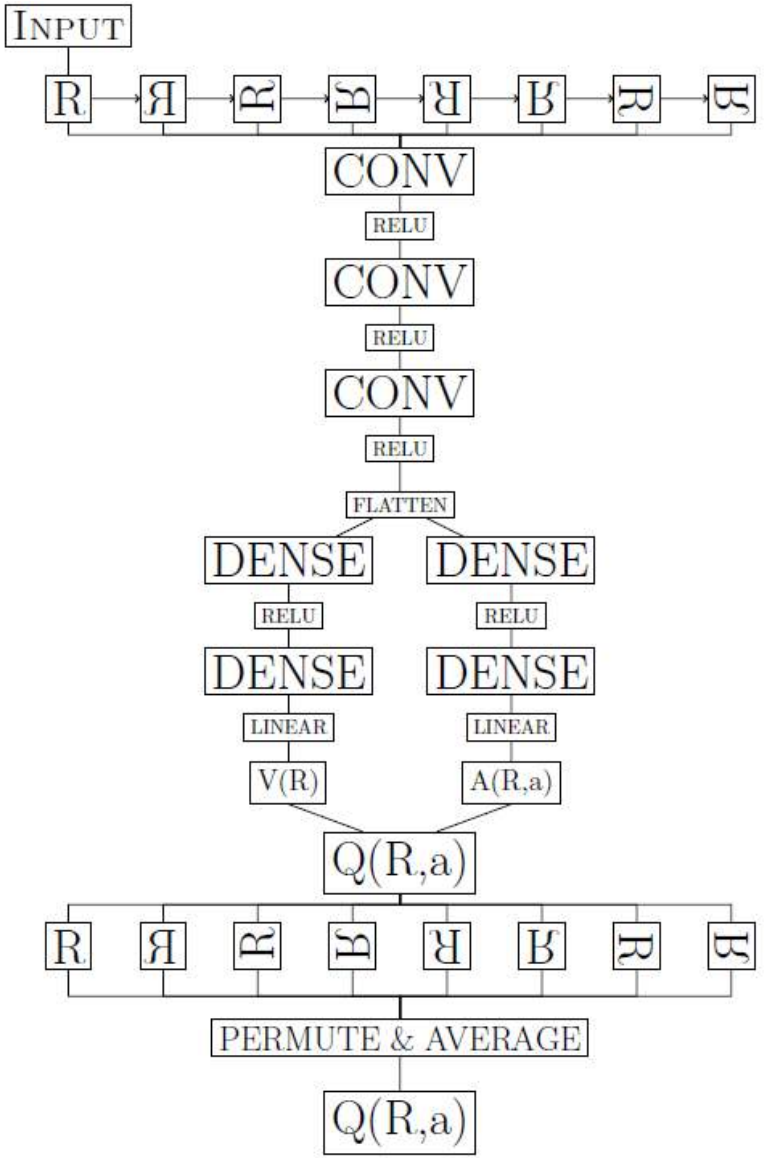


Wang et al, Dueling Network Architectures for Deep Reinforcement Learning, 2016

NOBEL



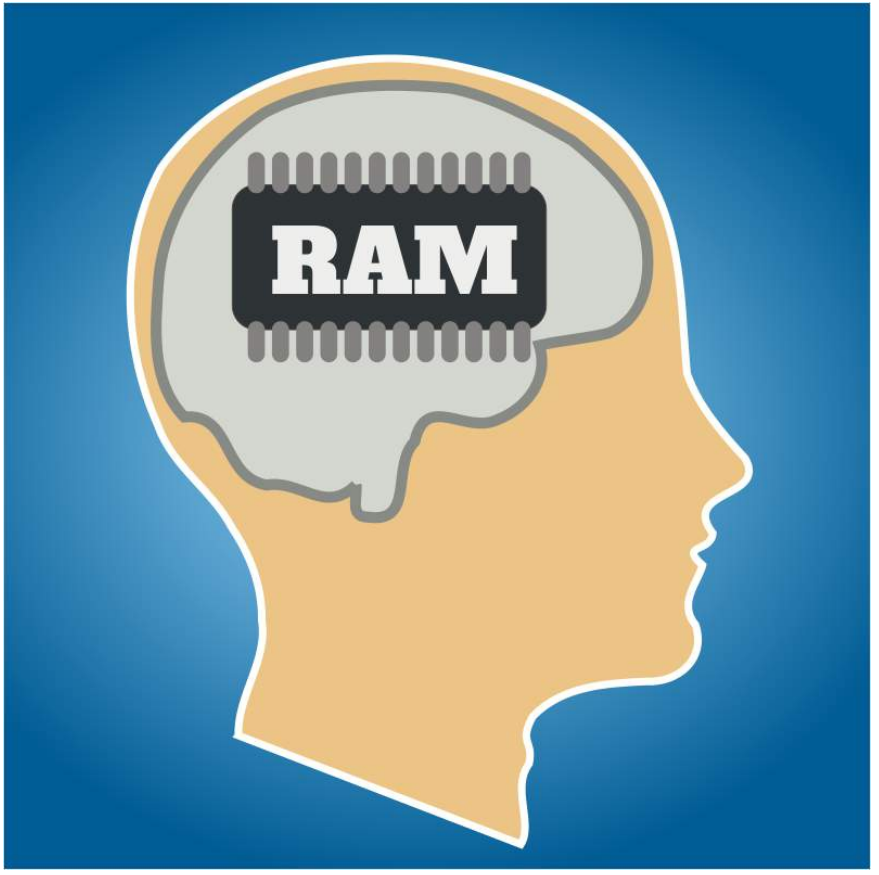
Wang et al, Dueling Network Architectures for Deep Reinforcement Learning, 2016


INPUT
R
CONV
RELU
CONV
RELU
CONV
RELU
FLATTEN
DENSE
DENSE
RELU
RELU
DENSE
DENSE
LINEAR
LINEAR
V(R)
A(R,a)
Q(R,a)
PERMUTE & AVERAGE
Q(R,a)

NOBEL

(Prioritized) Experience Replay:

- Data efficiency
- Learning the whole game
- Valuable experiences



http://clipart-library.com/data_images/449124.png

NOBEL

MATCH	ROUNDS	KILLS	SUICIDES	COINS	SCORE	RANK
Group E | 01	10	1	4	38	33	2
Group E | 02	10	1	5	46	51	1
Group E | 03	10	5	5	39	54	2
Group E | 04	10	0	3	41	41	2
Group E | 05	10	1	4	35	30	2
Group E | 06	10	2	5	40	50	2
Group E | 11	10	1	6	54	59	1
Group E | 12	10	1	5	56	51	1
Group E | 13	10	1	4	64	69	1
Group E | 14	10	0	2	69	69	1
Octofinal L	100	32	25	481	641	1
Quarterfinal R	100	41	38	255	360	3

https://hci.iwr.uni-heidelberg.de/vislearn/HTML/teaching/courses/FML/bomberman_rl/agent.php?id=NOBEL

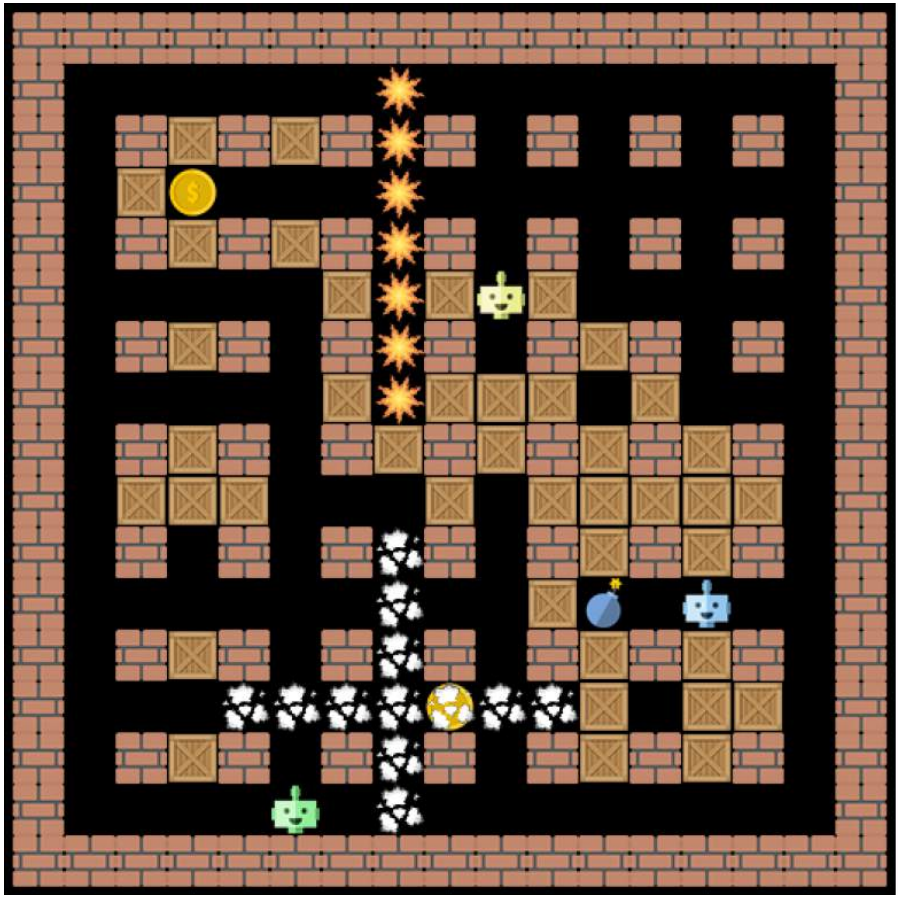
LaranTu

https://upload.wikimedia.org/wikipedia/commons/5/5c/Bronzetti_etrusco-romani%2C_laran_%28ares-marte%29_05.JPG

http://auckland-west.co.nz/wordpress/wp-content/uploads/2011/07/PICT5250aw.jpg

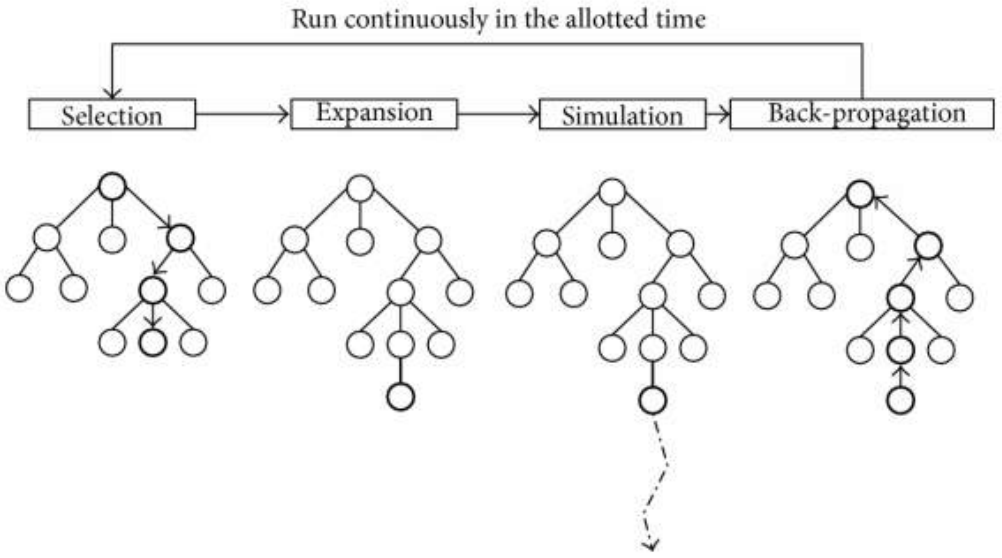
LaranTu

- Make it simple --> minigame
- Heuristic to get there



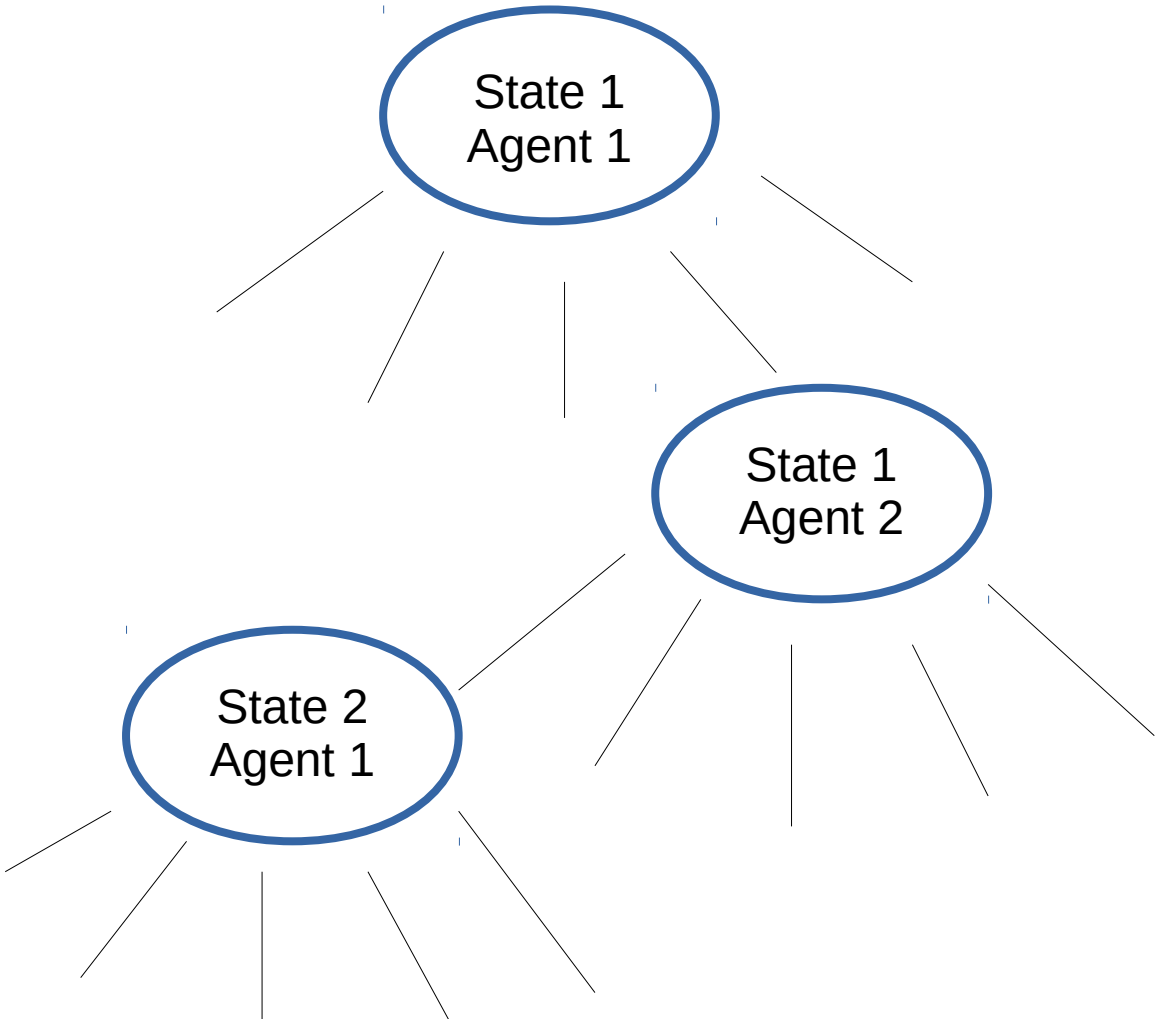
LaranTu

$$v_i + C \times \sqrt{\frac{\ln N}{n_i}}$$



https://www.researchgate.net/profile/Jacek_Mandziuk/publication/282043296/figure/fig11/AS:325106040623172@1454522728821/Four-steps-of-the-Monte-Carlo-Tree-Search-algorithm.png

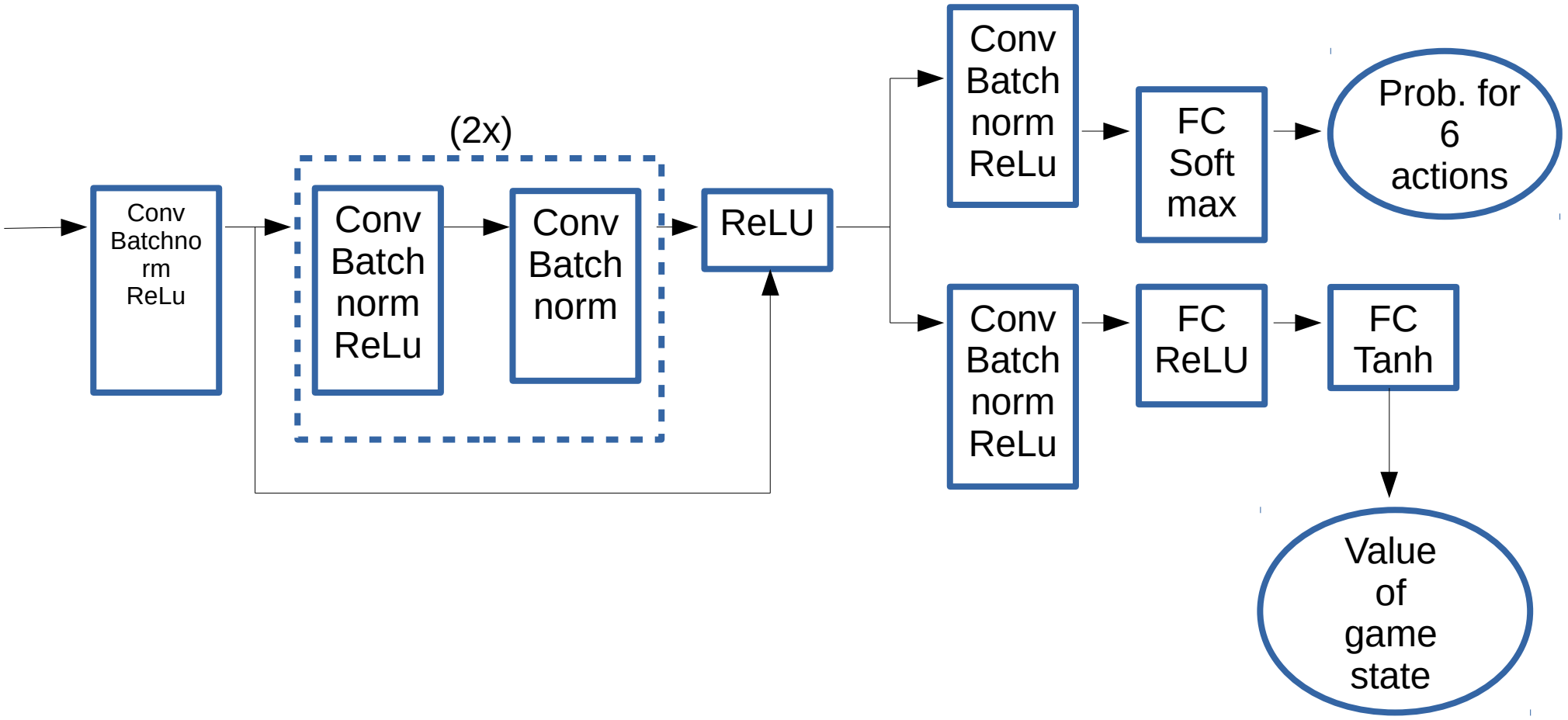
LaranTu



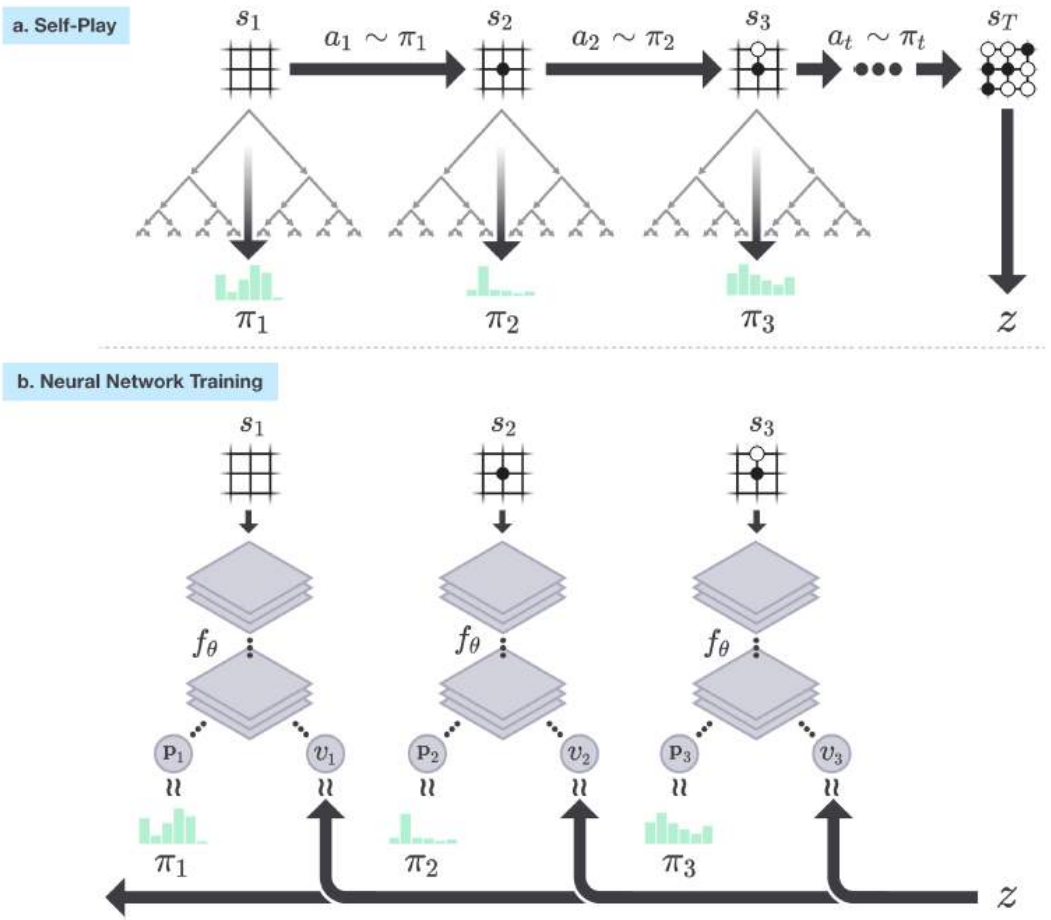
Selection based on:

- Value
- Policy prediction
- # Visited

LaranTu


Conv
Batchno
rm
ReLu
(2x)
Conv
Batch
norm
ReLu
Conv
Batch
norm
ReLU
Conv
Batch
norm
ReLu
FC
Soft
max
Prob. for
6
actions
Conv
Batch
norm
ReLu
FC
ReLU
FC
Tanh
Value
of
game
state

LaranTu



Silver et al 2017, „Mastering the game of Go without Human Knowledge“

LaranTu

MATCH	ROUNDS	KILLS	SUICIDES	COINS	SCORE	RANK
Group D | 03	10	1	3	11	8	3
Group D | 05	10	1	3	3	1	3
Group D | 06	10	0	3	4	-2	4
Group D | 08	10	2	5	8	12	2
Group D | 09	10	1	4	6	4	2
Group D | 10	10	1	4	6	7	2
Group D | 12	10	0	3	13	3	1
Group D | 13	10	2	3	6	6	3
Group D | 14	10	0	3	9	0	2
Group D | 15	10	1	5	2	-3	4
Octofinal J	100	1	25	48	-43	4

https://hci.iwr.uni-heidelberg.de/vislearn/HTML/teaching/courses/FML/bomberman_rl/agent.php?id=LaranTu

Sources

- Report: LaranTu, NOBEL, Lord_Voldemort
- neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/
- Wang et al, Dueling Network Architectures for Deep Reinforcement Learning, 2016
- Silver et al, A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, 2018