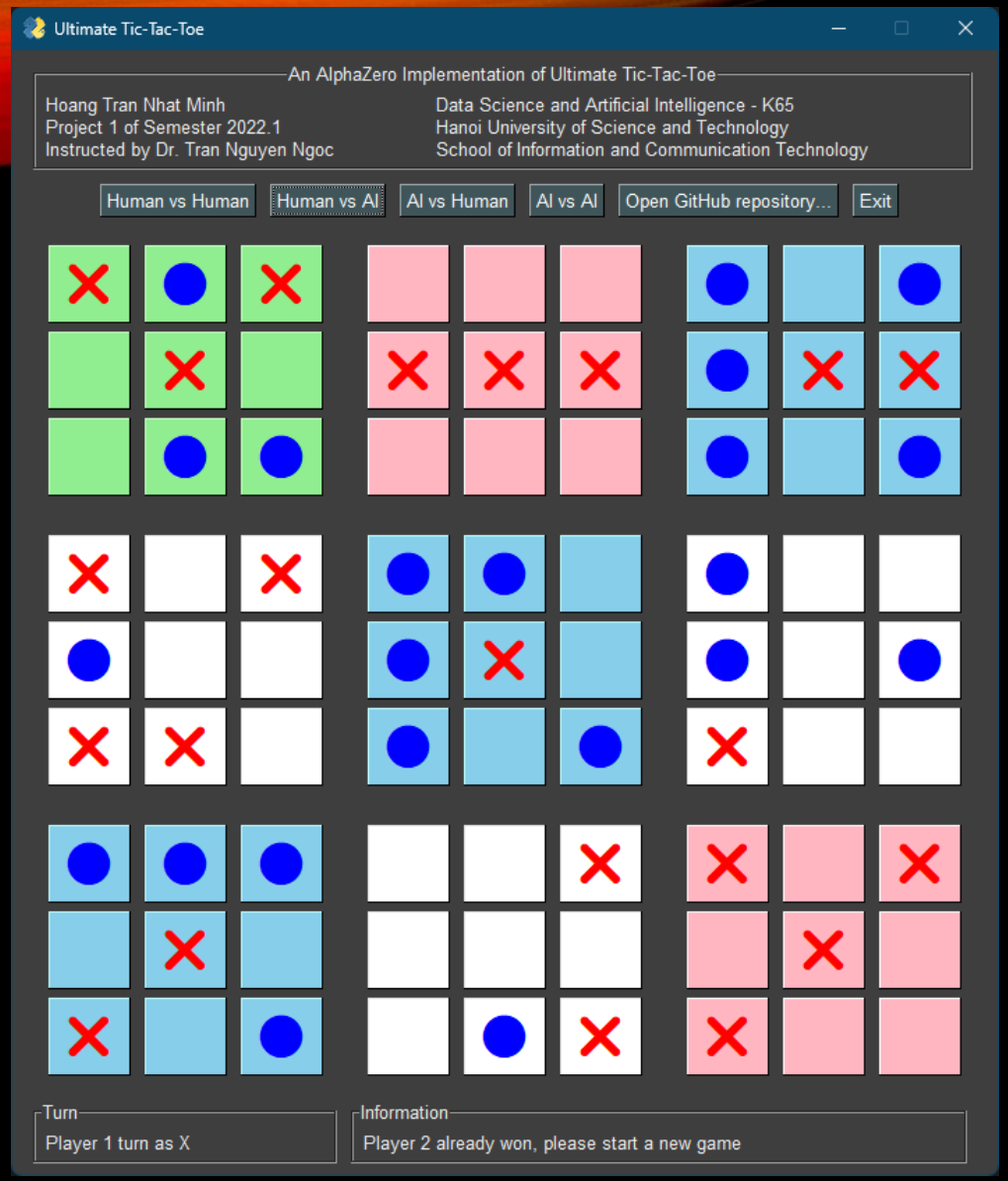
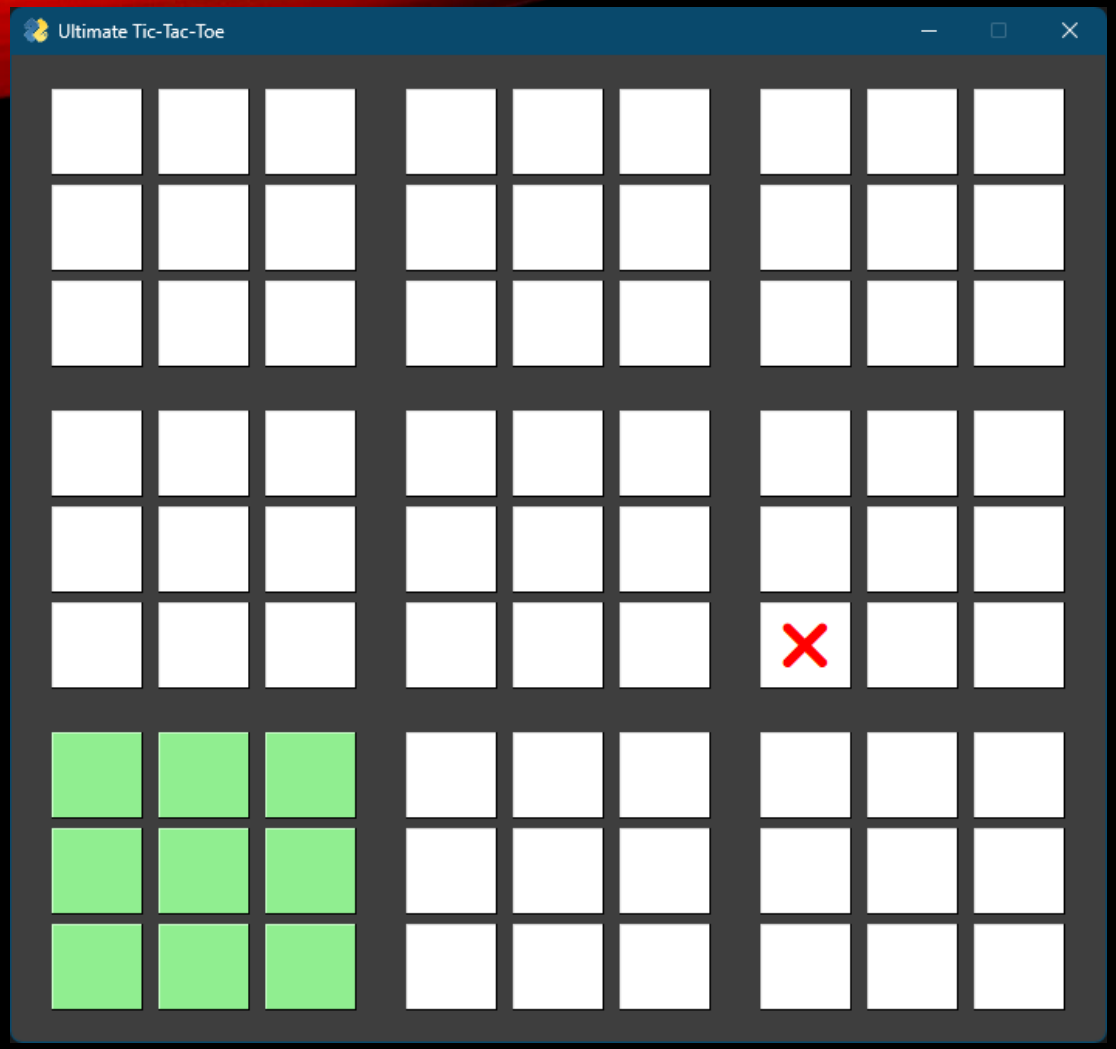
AN ALPHAZERO IMPLEMENTATION OF ULTIMATE TIC-TAC-TOE

Hoang Tran Nhat Minh
Data Science & Artificial Intelligence
Instructed by Dr. Tran Nguyen Ngoc



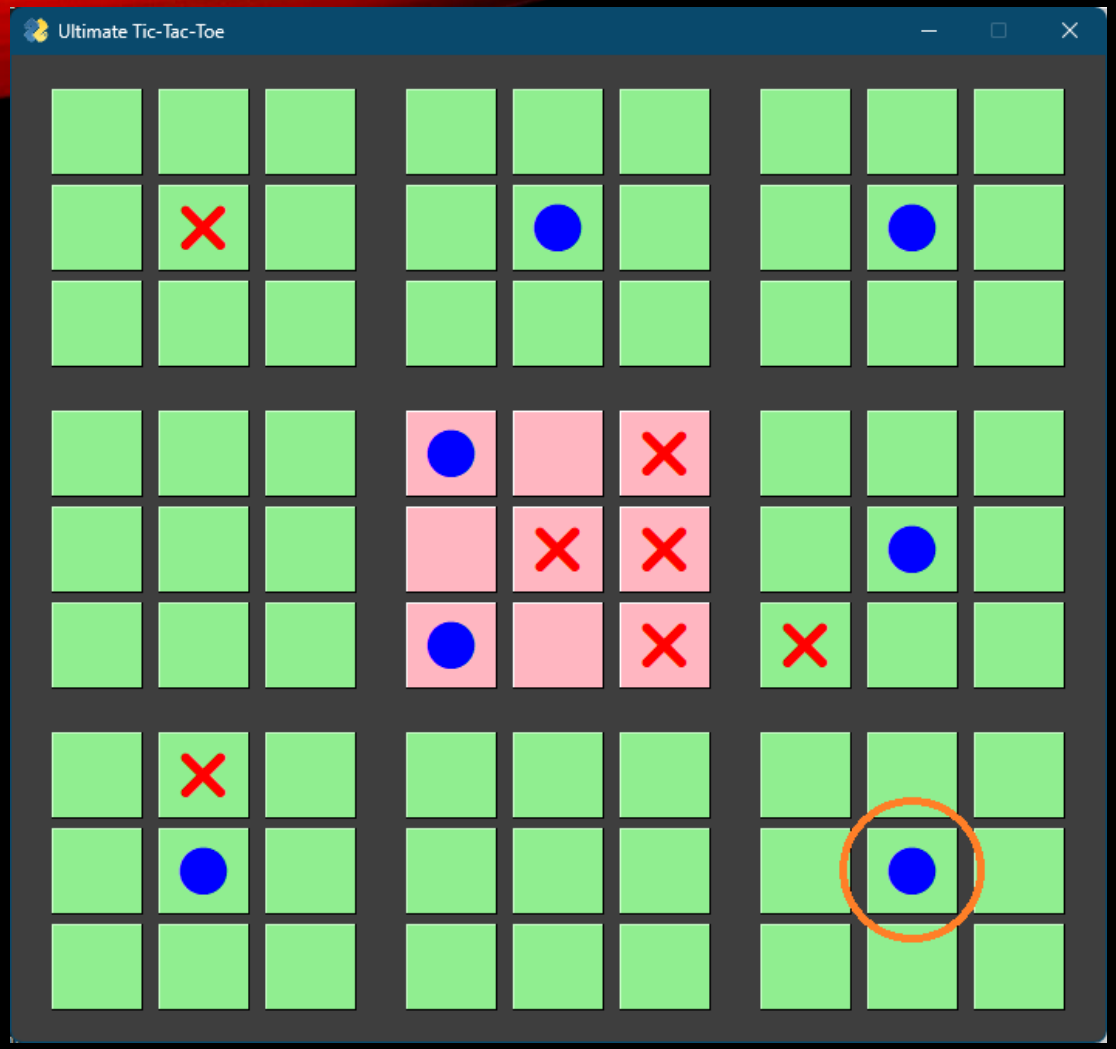
ABSTRACT

A modified version of the AlphaZero algorithm is used to train a new artificial intelligence program that plays ultimate tic-tac-toe, one of the most difficult variations of tic-tac-toe, while making use of human-extracted features using a convolutional neural network.



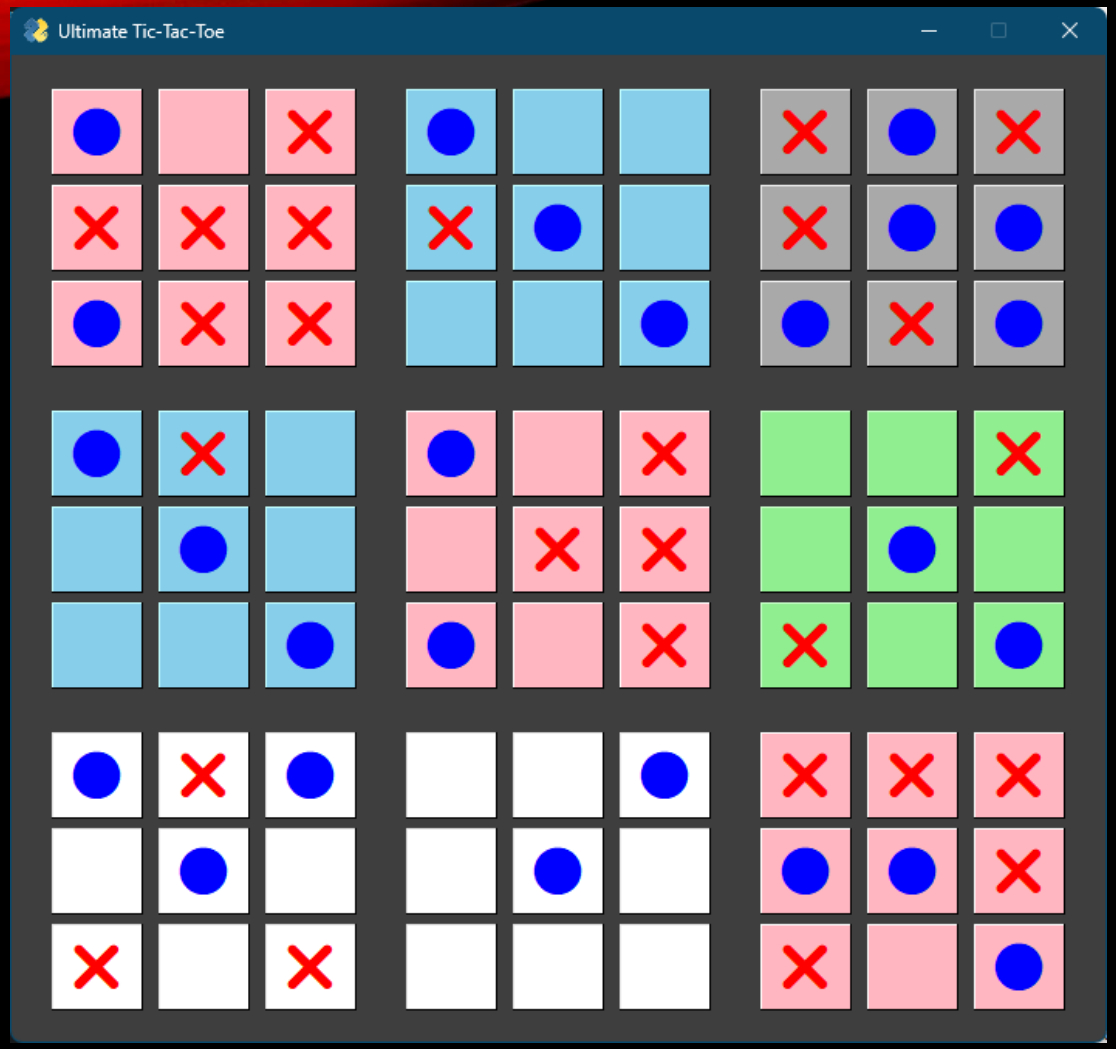
INTRODUCTION TO ULTIMATE TIC-TAC-TOE

- “Send” opponent to the relative local board in the next move



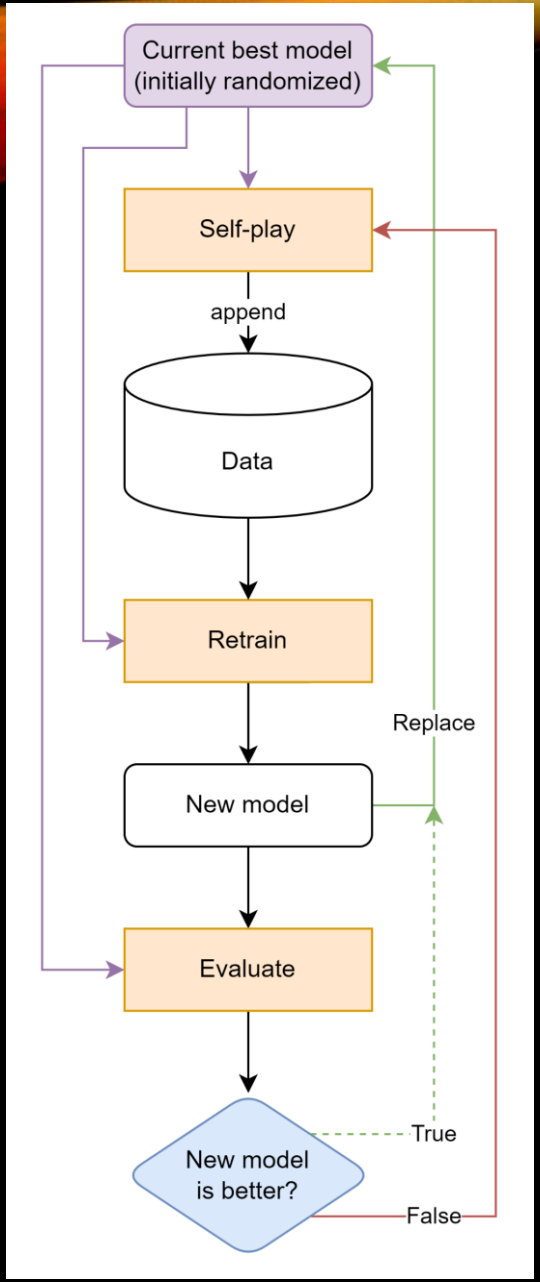
INTRODUCTION TO ULTIMATE TIC-TAC-TOE

- If a local board is determined, it is not playable anymore
- If sent to it, play anywhere else (but determined boards)



INTRODUCTION TO ULTIMATE TIC-TAC-TOE

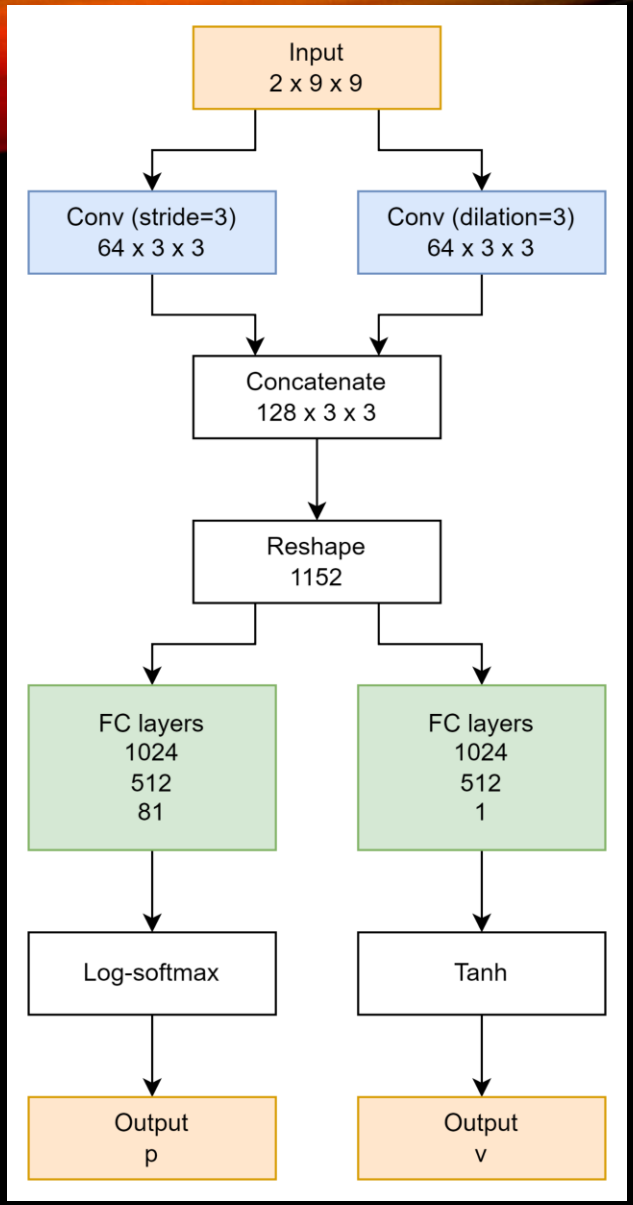
Player 1 won this game



ALPHAZERO: TRAINING LOOP

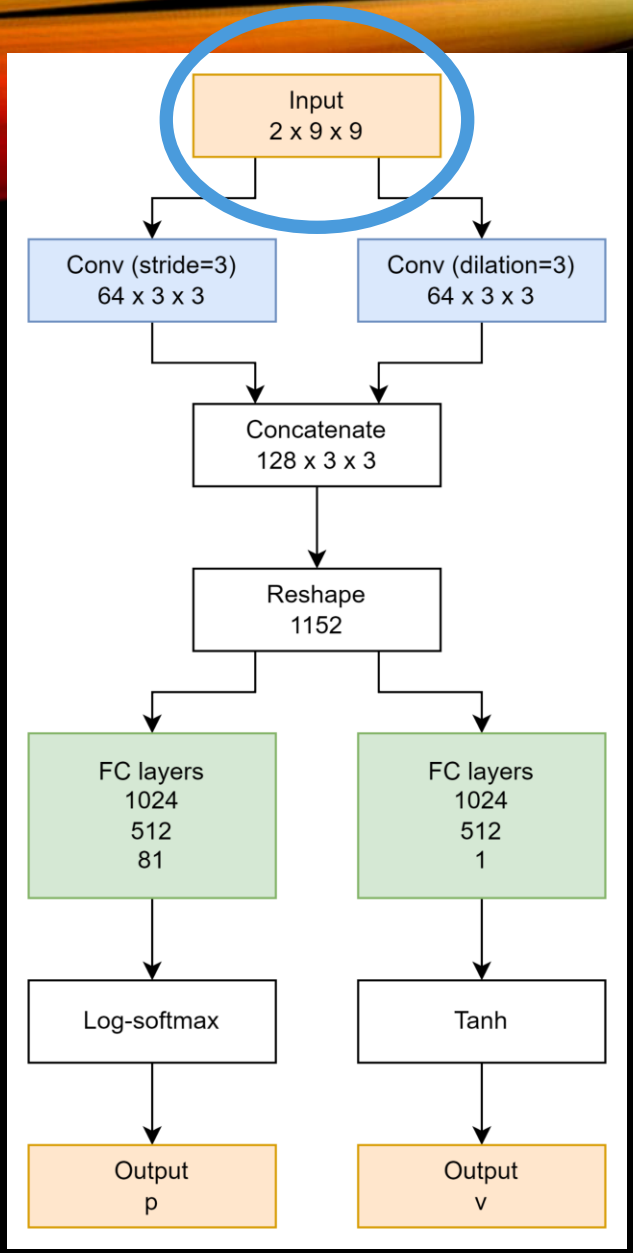
- Self-play: to generate new data
- Retrain: to get new model
- Evaluate: to decide replacing the best model or not

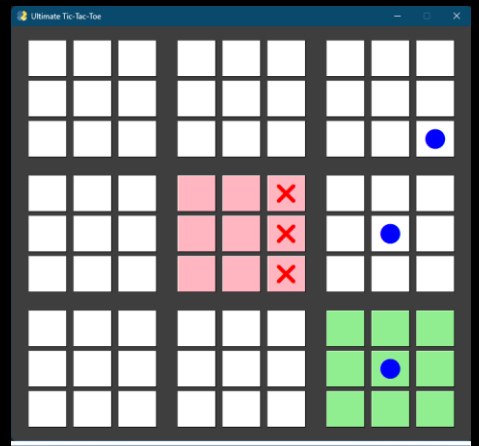
In self-play and evaluate: Monte Carlo tree search decision guided by a convolutional neural network (model)



NETWORK ARCHITECTURE

- Convolutional layers
- Fully connected layers



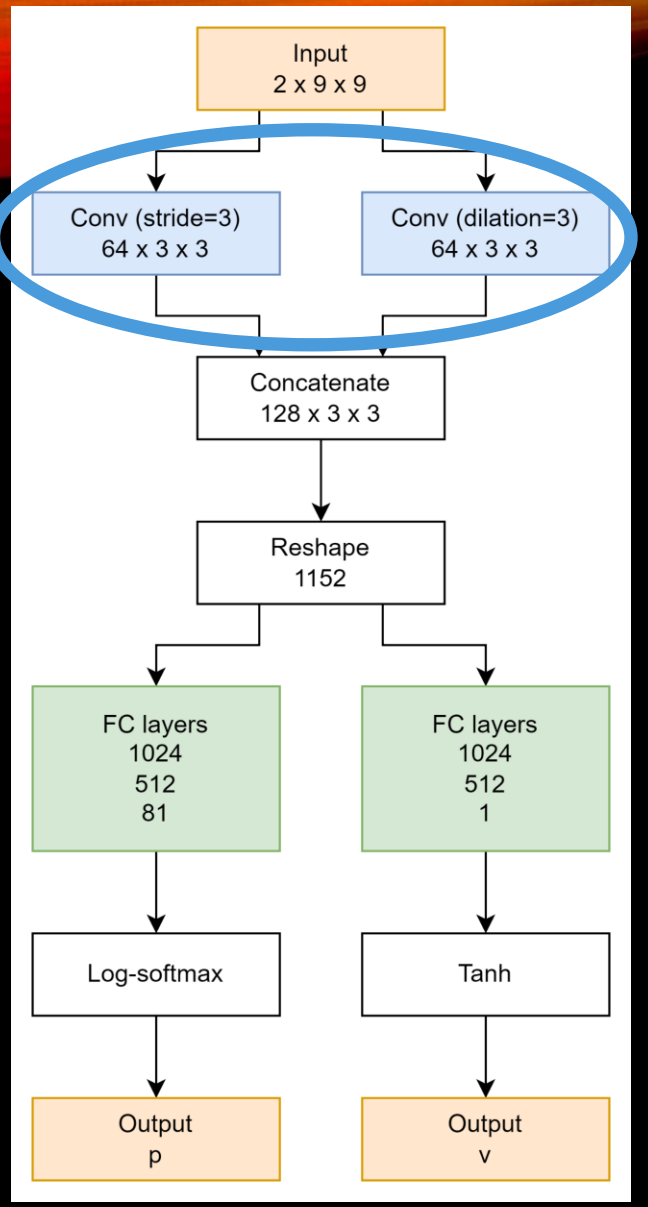


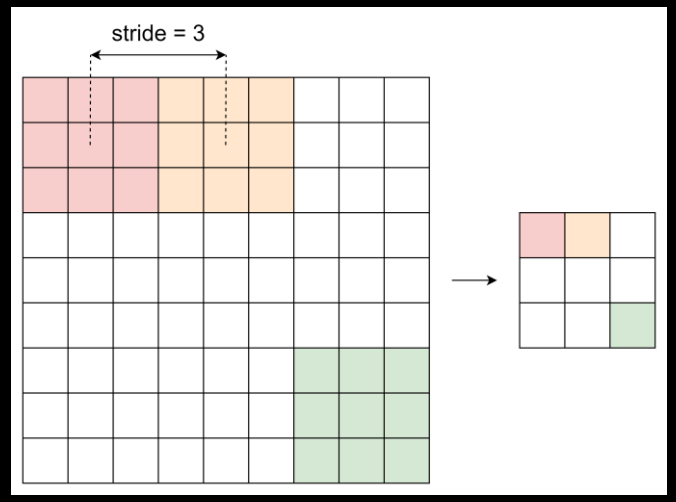
```
[[ 0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 -1]
 [ 0  0  0  0  0  1  0  0  0]
 [ 0  0  0  0  0  1  0 -1  0]
 [ 0  0  0  0  0  1  0  0  0]
 [ 0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0 -1  0]
 [ 0  0  0  0  0  0  0  0  0]]

[[0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 1 1]
 [0 0 0 0 0 0 1 0 1]
 [0 0 0 0 0 0 1 1 1]]
```

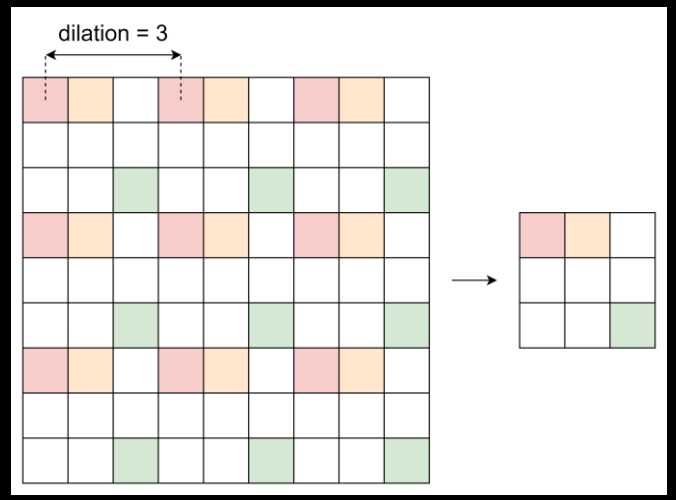
INPUT

- Channel #1: state (values: 1, -1, 0)
- Channel #2: valid-move mask (values: 1, 0)

CONVOLUTIONAL LAYERS





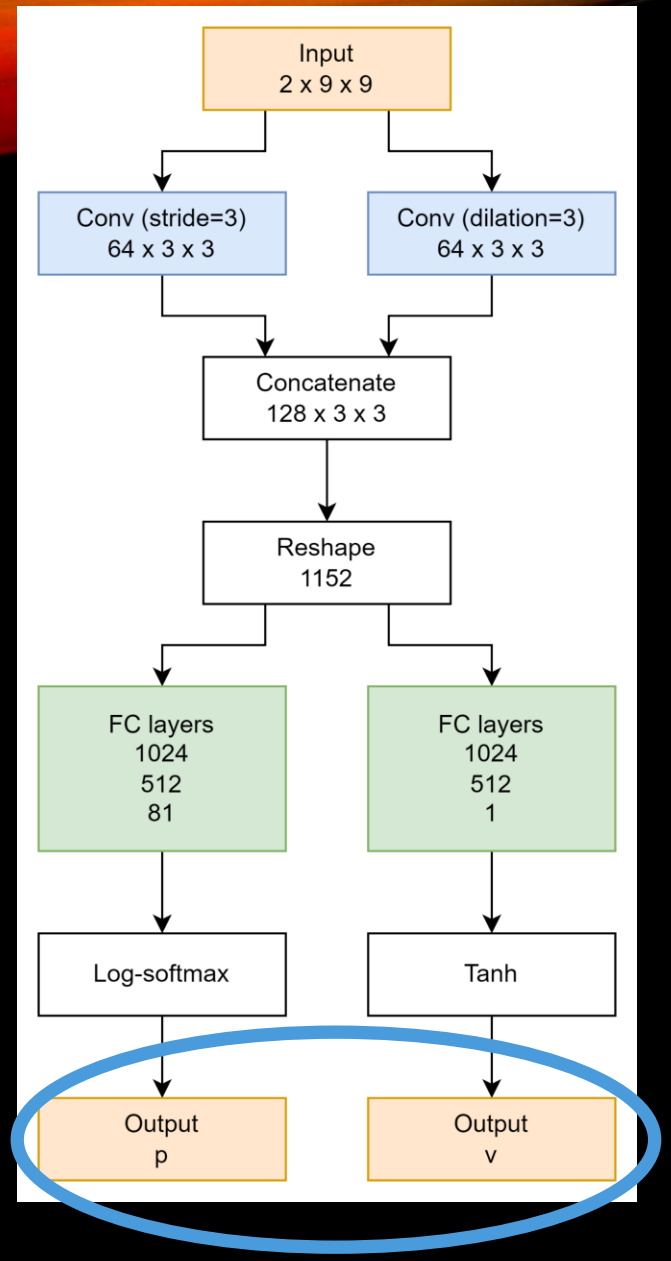


Similarities

- Kernel size: 3x3
- Output image size: 3x3

Differences

- Layer #1: stride = 3
- Layer #2: dilation = 3



OUTPUT

- p: Policy, the probabilities of moves
- v: Value, the evaluation of the input board

RESULTS

Iter.	Win	Loss	Draw	Win rate
1	10	10	4	50.0%
2	12	8	4	60.0%
3	5	8	11	38.5%
4	3	8	13	27.3%
5	5	7	12	41.7%
6	6	4	14	60.0%
7	3	4	17	42.9%
8	7	1	16	87.5%
9	4	3	17	57.1%
10	4	3	17	57.1%
11	12	11	1	52.2%
12	10	9	5	52.6%
13	10	8	6	55.6%
14	9	9	6	50.0%
15	10	9	5	52.6%
16	8	9	7	47.1%
17	11	11	2	50.0%
18	10	11	3	47.6%
19	8	8	8	50.0%
20	12	8	4	60.0%
21	11	8	5	57.9%
22	8	11	5	42.1%
23	8	9	7	47.1%
24	9	7	8	56.3%
25	9	10	5	47.4%
26	10	9	5	52.6%
27	9	7	8	56.3%
28	11	8	5	57.9%
29	10	8	6	55.6%

- Tested itself in the training loop
  - Discarding previous models: improvement
- Beginner level

AN ALPHAZERO IMPLEMENTATION OF ULTIMATE TIC-TAC-TOE

- Oversimplified, isn't it? Please see the report for more details.
- Everything is in https://github.com/htnminh/AlphaZero-Ultimate-TicTacToe