AlphaZero to SigmaZero

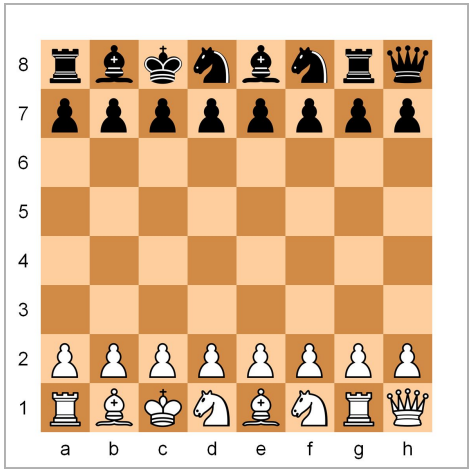
Group 13

Benjamin, Chuan Shao, Raymond, Abram, Shaun

Project Description

SOTA Vanilla chess

Chess 960



Project Description

Training Duration

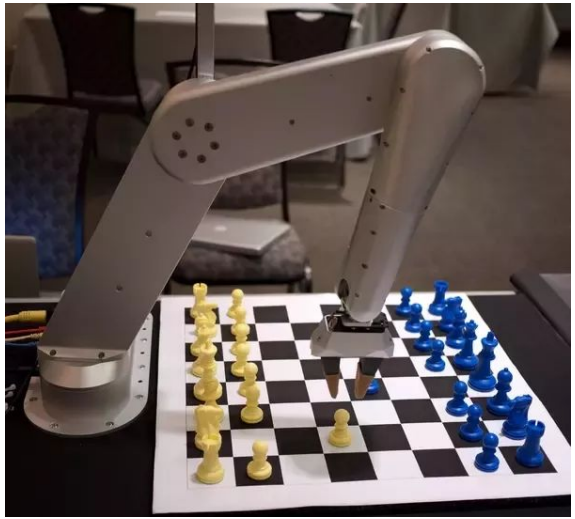
Elo Rating

Dataset

Non-Generated

Chess.com

Generated (self play)



cs.washington.edu

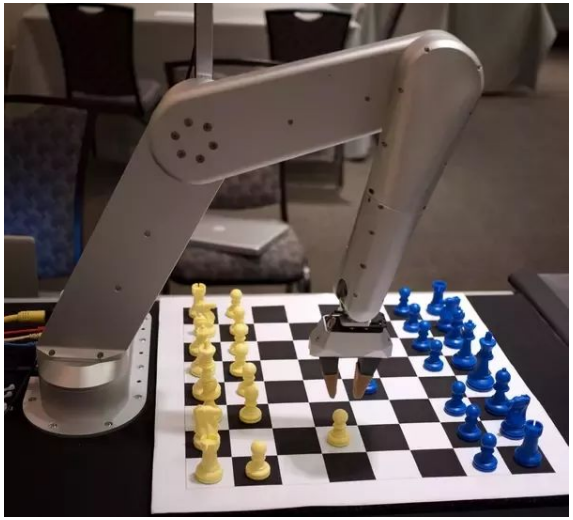
Non-Generated Dataset

- Free Internet Chess Server
- Played by real humans
- An estimated ELO of >2000
- Expert level

Chess.com

Generated Dataset

- Self-play against itself
- Generated through games
- Explores strategies that regular chess players would not think of

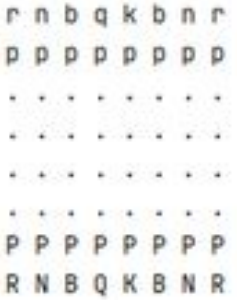


cs.washington.edu

Data Composition

- Utilized the Python Chess library
- Converted the board to a tensor
- Adapted from the original AlphaZero paper

- P1 & P2 Pieces (12 planes)
- Repetitions (2 planes)
- Colour (1 plane)
- Total moves (1 plane)
- Castling rights (4 planes)
- No progress count (1 plane)



ALPHA ZERO Architecture

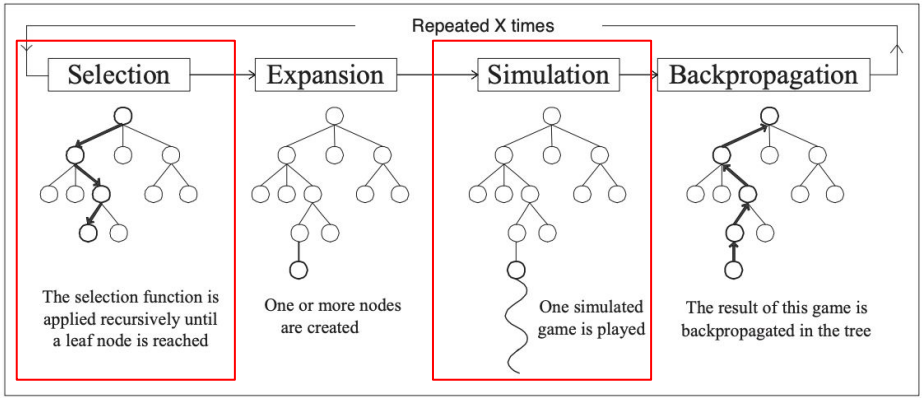


Figure 1: Outline of a Monte-Carlo Tree Search.

Neural Network

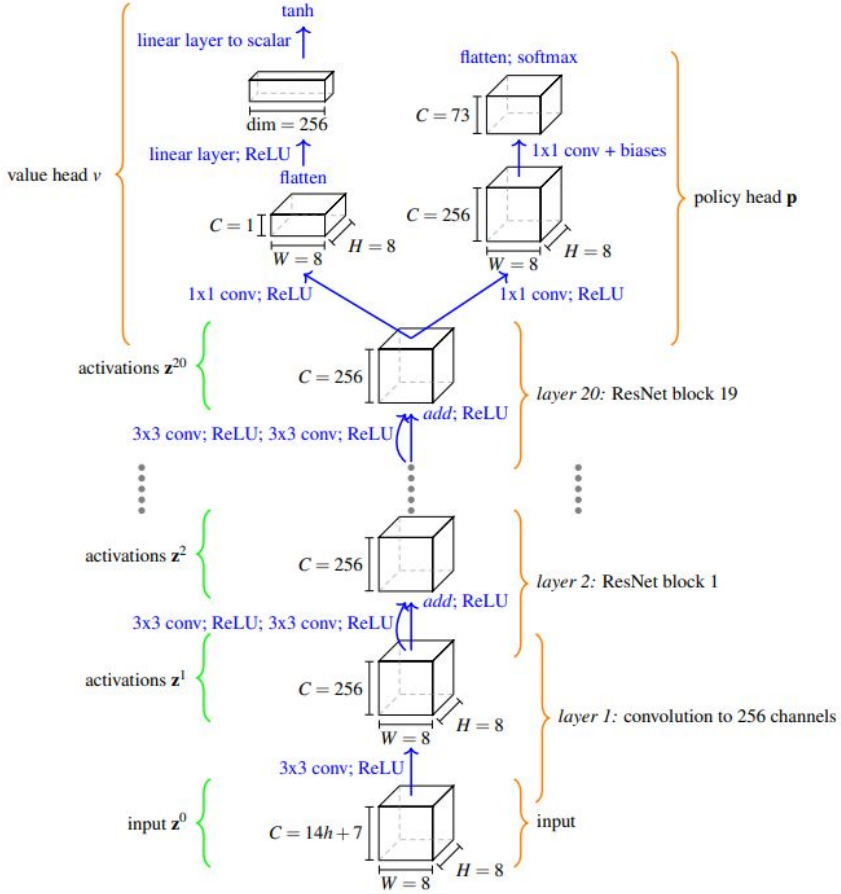
Inputs: Game State

↓

Convolutional and Residual Layers

↓

Outputs: Policy (Actions that can be taken) & Value (Expected Outcome)


tanh
linear layer to scalar
dim = 256
linear layer; ReLU
flatten
C = 1
W = 8
H = 8
value head v
flatten; softmax
C = 73
1x1 conv + biases
C = 256
W = 8
H = 8
policy head p
1x1 conv; ReLU
1x1 conv; ReLU
activations z^20
C = 256
add; ReLU
3x3 conv; ReLU; 3x3 conv; ReLU
layer 20: ResNet block 19
activations z^2
C = 256
add; ReLU
3x3 conv; ReLU; 3x3 conv; ReLU
layer 2: ResNet block 1
activations z^1
C = 256
W = 8
H = 8
3x3 conv; ReLU
layer 1: convolution to 256 channels
input z^0
C = 14h + 7
W = 8
H = 8
input

Training

Vanilla Chess Training

Vanilla Chess

Chess.com

Transfer Learning

Fischer Random Chess / Chess960

Chess.com

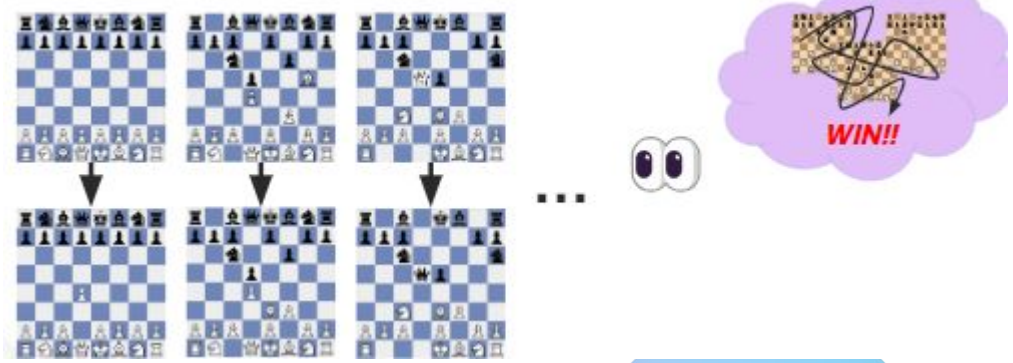
Chess960 Training

Training – Vanilla Chess

Initially reinforcement learning, but generation of games takes too long ~ 418 years

If we used reinforcement learning...

Supervised learning for 60 epochs on 15000 games of 2000+ ELO players used to train from Free Internet Chess Server database



Supervised Learning

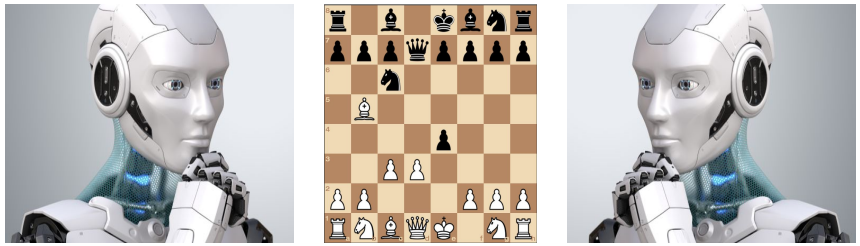
Reinforcement Learning - Loss Function

$$l = (z - v)^2 - \pi^T log\{p\} + c||\theta||^2$$

where z represents the value of the node, π signifies the action chosen, p and v denote the policy and the value yielded from the model's output, respectively, and c stands for a constant term

Training - Chess960

Generated 20 epochs of reinforcement self-play after transfer learning



AI vs Itself

Fischer Random Chess / Chess960

Chess.com

Results & Discussion

Stockfish Configuration

Level of Playing Strength	Skill Level	Time Limit	Search Depth	Estimated ELO
0	0	1	5	1376
1	1	1	5	1462
2	2	1	5	1547
3	3	1	5	1596
4	4	1	5	1718
5	5	1	5	1804
6	6	1	5	2012
7	7	1	5	1993
8	8	1	6	2127
9	9	2	7	2270
10	20	10	50	3100

Win Rate (Vanilla SigmaZero)

Model	Game Mode	Stockfish Level of Playing Strength	Estimated ELO	Model Win	Model Loss	Model Draw	Games	Points
supervised_model_15k_40.pt	Vanilla	3	1596	2	1	2	DWLDW	3.0/5.0
		4	1718	1	2	2	LDDDL	1.5/5.0
supervised_model_15k_45.pt	Vanilla	3	1596	3	2	0	WLWLW	3.0/5.0
		4	1718	1	2	2	DWDLL	2.0/5.0
supervised_model_15k_40.pt	Chess960	3	1596	1	4	0	WLLLL	1.0/5.0
supervised_model_15k_45.pt	Chess960	3	1718	2	3	0	WLLLW	2.0/5.0

Win Rate (Chess960 SigmaZero)

Model	Game Mode	Stockfish Level of Playing Strength	Estimated ELO	Model Win	Model Loss	Model Draw	Games	Points
RL_960_0.pt	Chess960	0	1376	3	1	0	WWLW	3.0/5.0
		1	1462	0	3	1	DLLL	0.5/5.0
RL_960_5.pt	Chess960	0	1376	0	3	2	LDLDL	1.0/5.0
RL_960_15.pt	Chess960	0	1376	1	0	0	WLLL	1.0/5.0
RL_960_20.pt	Chess960	0	1376	0	3	0	LLL	0.0/5.0
supervised_model_15k_45.pt	Chess960	0	1376	3	0	0	WWW	3.0/5.0
		1	1462	3	1	0	WWLW	3.0/5.0
		2	1596	1	1	3	LWDDD	2.5/5.0
		3	1718	2	3	0	WLLLW	2.0/5.0

Training Duration & ELO Rating

Supervised Learning (theoretical)

15k games
5 min/game
137 years on RTX4080

Transfer Learning

20 epochs
~10h on RTX4080

Openings

1
Ruy Lopez

2
Morphy’s Defence

3
Indian Defence

Comparison with State-of-the-art

AlphaZero	Leela Chess Zero	SigmaZero (vanilla)
>4500 ELO 44m games 9 hours 5000x TPUs	~4000 ELO >2.5b games Open-source distributed computing	~1700 ELO 10h 1x RTX4080

GUI Demo

The End