Mastering the Game of Go without Human Knowledge

06/15/18

Presented by: Henry Chen

CS885 Reinforcement Learning

UNIVERSITY OF WATERLOO

Introduction

Image source:
https://medium.com/syncedreview/alphago-zero-approaching-perfection-d8170e2b4e48

Introduction

The Game of Go

- ancient board game
- 19 x 19 grid
- complexity: **~10^{170}**

Image source:
https://medium.com/@karpathy/alphago-in-context-c47718cb95a5

Challenging AI problem

- How to search through an intractable search space?
- Breakthrough: AlphaGo

Background

AlphaGo

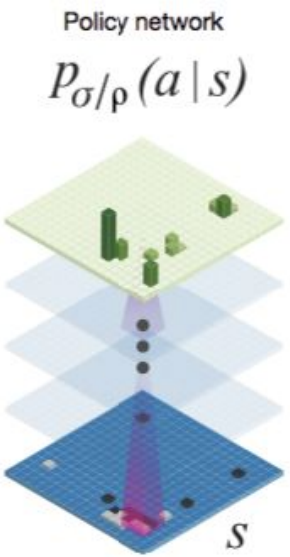
- March 2016: defeated 18-time world champion Lee Sedol 4-1

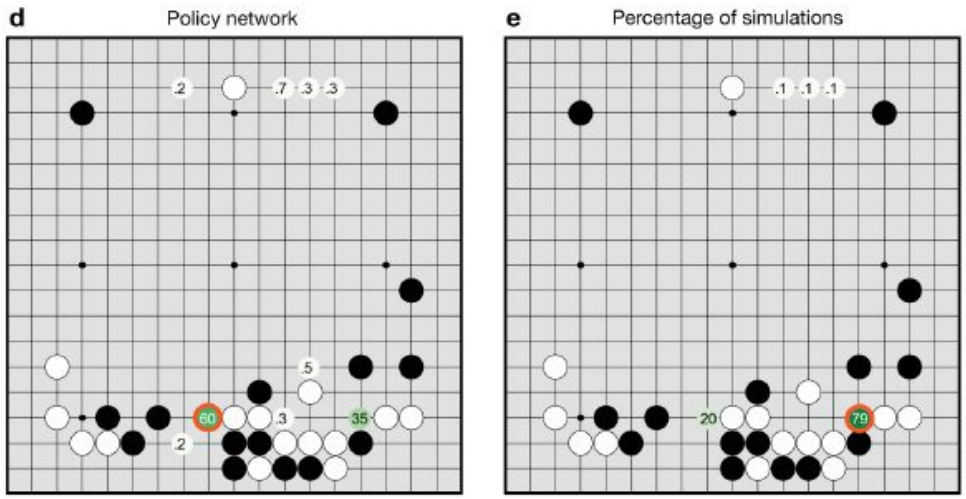
Image source: https://www.tastehit.com/blog/google-deepmind-alphago-how-it-works/

Background

AlphaGo - Architecture

1. **Policy Network**
   - Purpose: decide next best move
   - *Convolution Neural Network* (13 hidden layers)
   - Stage 1: *Supervised Learning* to predict human expert moves (57%)
   - Stage 2: Improve network by *Policy Gradient Reinforcement Learning* through self-play using roll-out policy (80% > stage 1)





Source:
Google DeepMind,
Mastering the Game of
Go with Deep Neural
Networks and
Tree Search

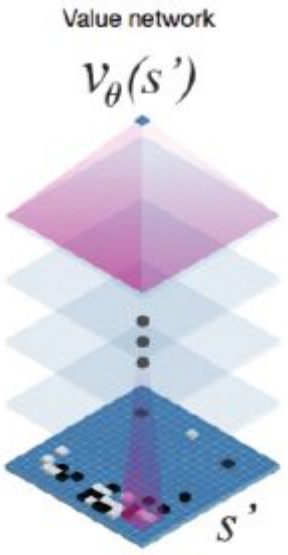
UNIVERSITY OF WATERLOO

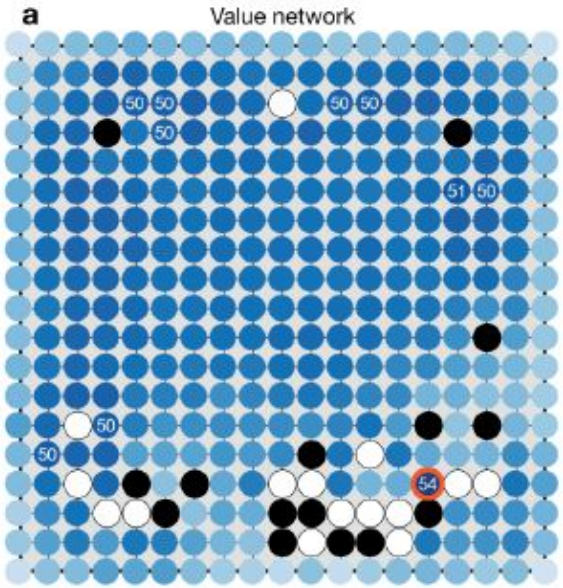
Background

AlphaGo - Architecture

2. Value Network

- Purpose: evaluate chances of winning
- Convolution Neural Network (14 hidden layers)
- Train network by regression on state-outcome pair sampled from self-play data using policy network

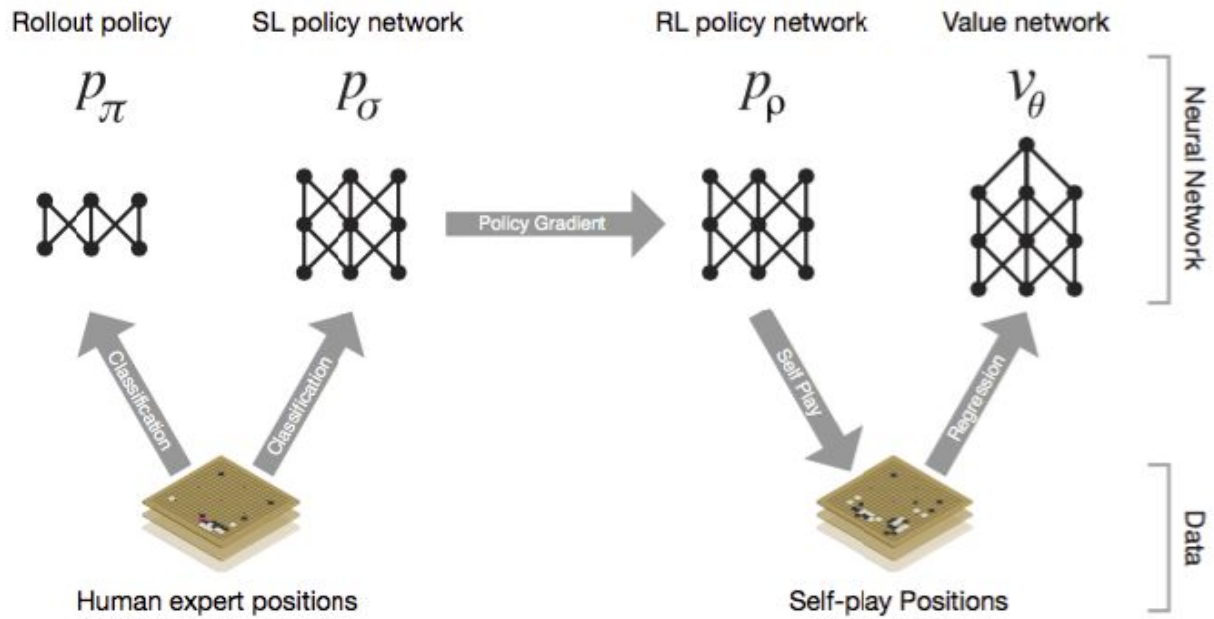




Source:
Google DeepMind,
Mastering the Game of
Go with Deep Neural
Networks and
Tree Search

UNIVERSITY OF WATERLOO

Background



Source:
Google DeepMind,
Mastering the Game of
Go with Deep Neural
Networks and
Tree Search

Policy Network (stage 1):

- 30 millions position from 160,000 human games
- 50 GPUs
- 3 weeks

Policy Network (stage 2):

- 10,000 mini-batches of 128 self-play games
- 50 GPUs
- 1 day

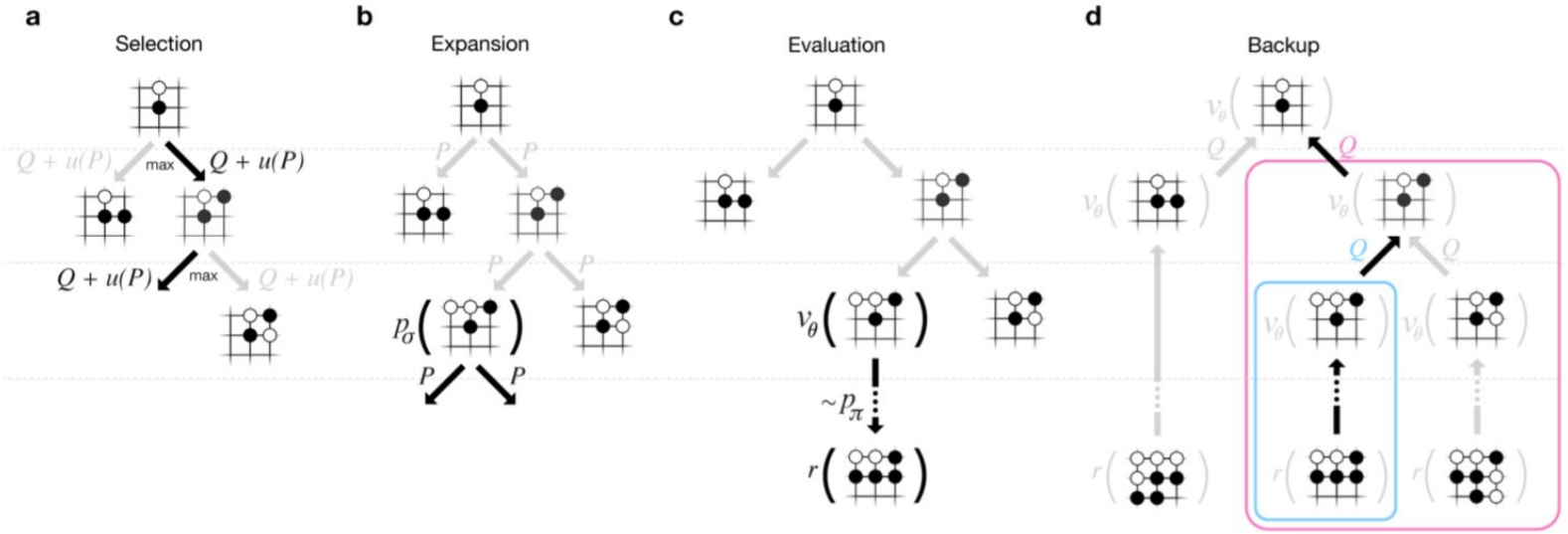
Value Network

- 30 millions unique positions
- 50 GPUs
- 1 week

Background

3. Monte-Carlo Tree Search (MCTS)

- Purpose: Combining policy and value networks to select actions by lookahead search
- Asynchronous multi-threaded search (distributed ~50 GPUs)



Source:
Google DeepMind, Mastering the Game of Go with Deep Neural Networks and Tree Search

UNIVERSITY OF WATERLOO

Background

Limitations

- Require large data-set of expert games
- Use of handcraft features

Feature	# of patterns	Description
Response	1	Whether move matches one or more response features
Save atari	1	Move saves stone(s) from capture
Neighbour	8	Move is 8-connected to previous move
Nakade	8192	Move matches a *nakade* pattern at captured stone
Response pattern	32207	Move matches 12-point diamond pattern near previous move
Non-response pattern	69338	Move matches 3×3 pattern around move
Self-atari	1	Move allows stones to be captured
Last move distance	34	Manhattan distance to previous two moves
Non-response pattern	32207	Move matches 12-point diamond pattern centred around move

Source:
Google DeepMind,
Mastering the Game of
Go with Deep Neural
Networks and
Tree Search

- Asynchronous training and computation intensive

Content of paper

Mastering the Game of Go without Human Knowledge

David Silver*, Julian Schrittwieser*, Karen Simonyan*, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, Demis Hassabis.

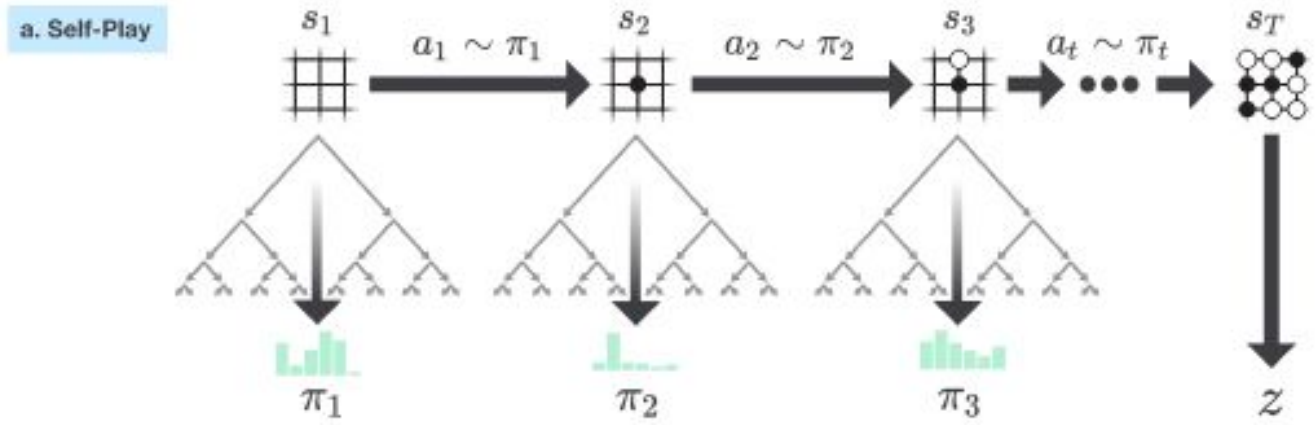
DeepMind, 5 New Street Square, London EC4A 3TW.

*These authors contributed equally to this work.

Content of paper

AlphaGo Zero

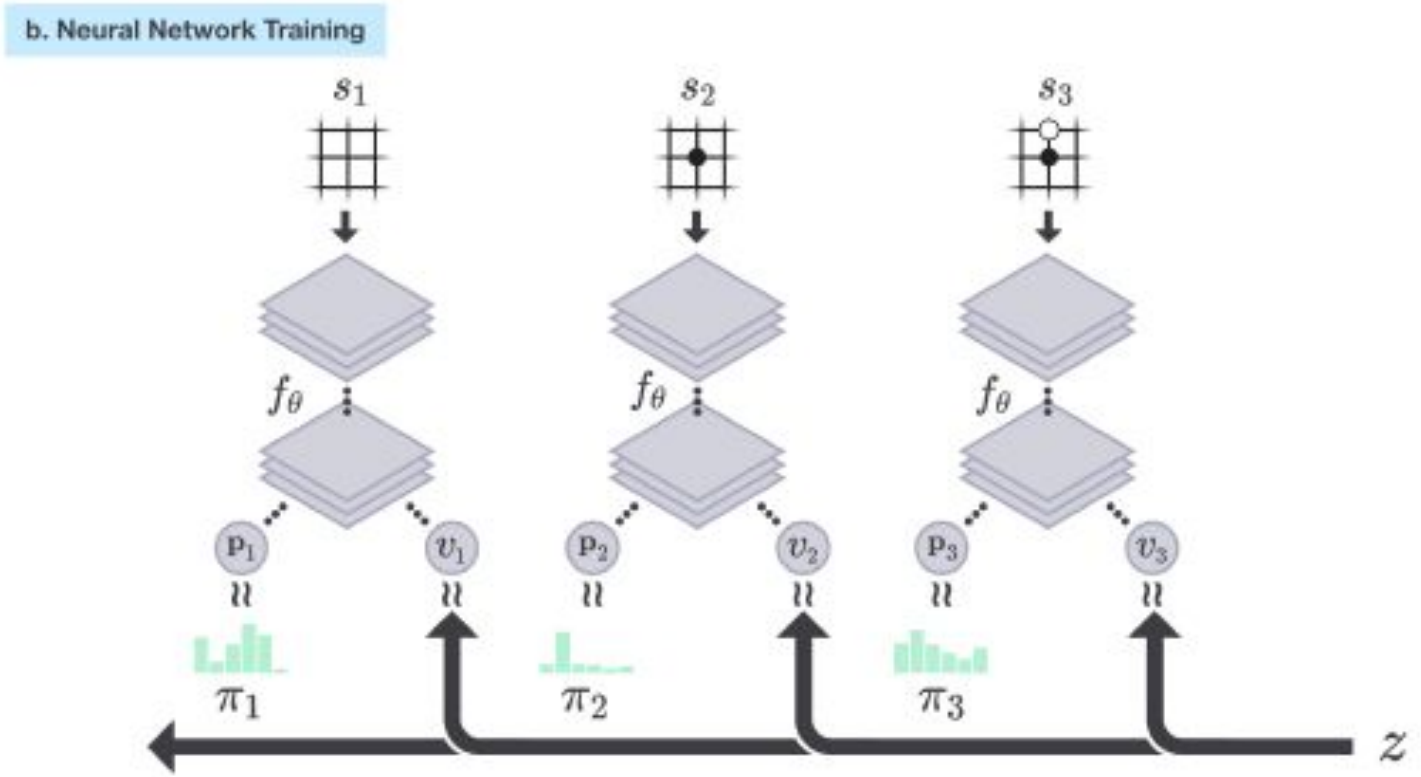
1. uses no Human Knowledge and learn only by Self-Play



Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

Content of paper

AlphaGo Zero

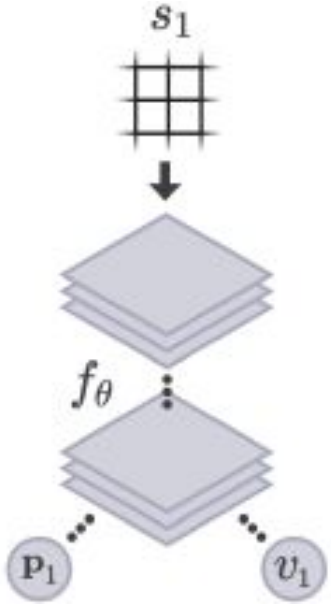


Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

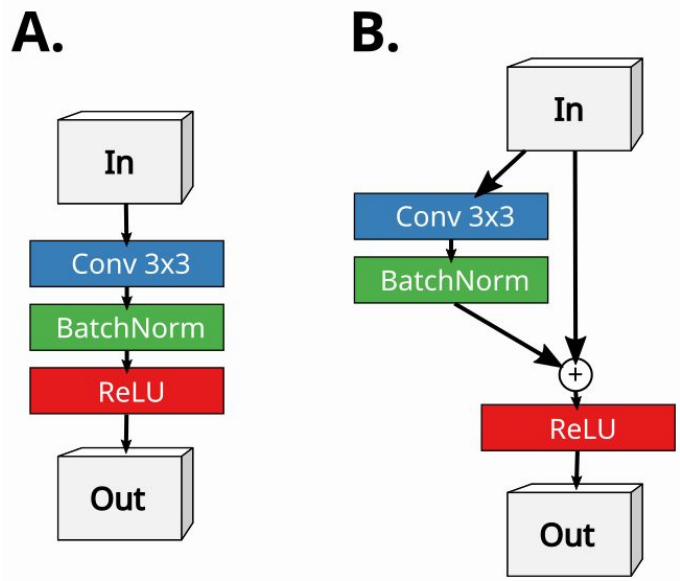
Content of paper

AlphaGo Zero

2. Single Neural Network with ResNets Structure
   - Dual purpose: decide next best move and evaluate chances of winning



Source:
Google DeepMind,
Mastering the Game of
Go without Human
Knowledge

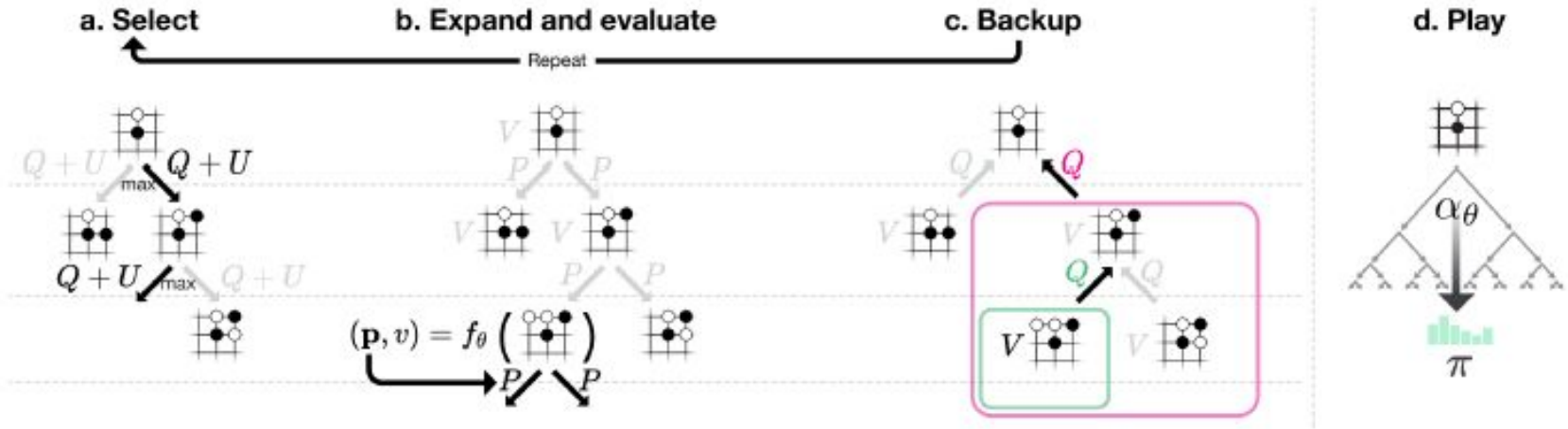


Source:
http://neural.vision/blog
/article-reviews/deep-lea
rning/he-resnet-2015/

Content of paper

AlphaGo Zero

3. Simpler Tree Search



Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

Content of paper

AlphaGo Zero

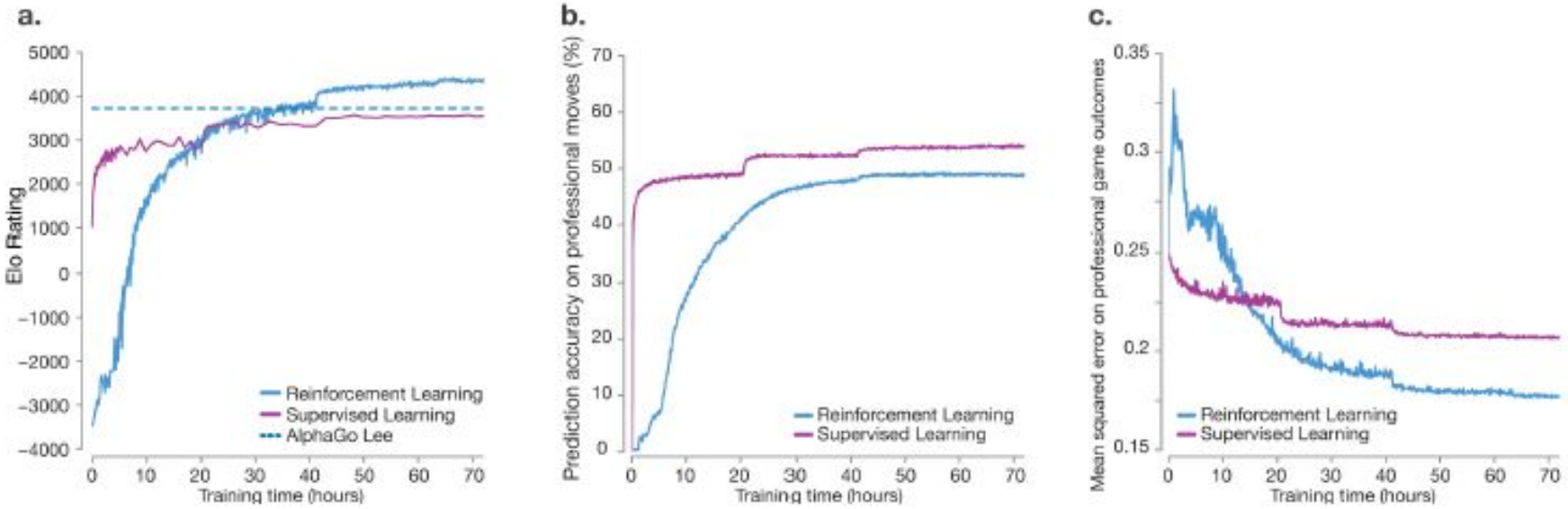
4. Requires no handcraft features
   - Only requires raw board representations and its history, plus some basic game rules as neural network input

5. Improved computation efficiency
   - Single machine on Google Cloud with 4 TPUs

Source:
Google DeeMind, Mastering the Game of Go without Human Knowledge

Empirical Evaluation

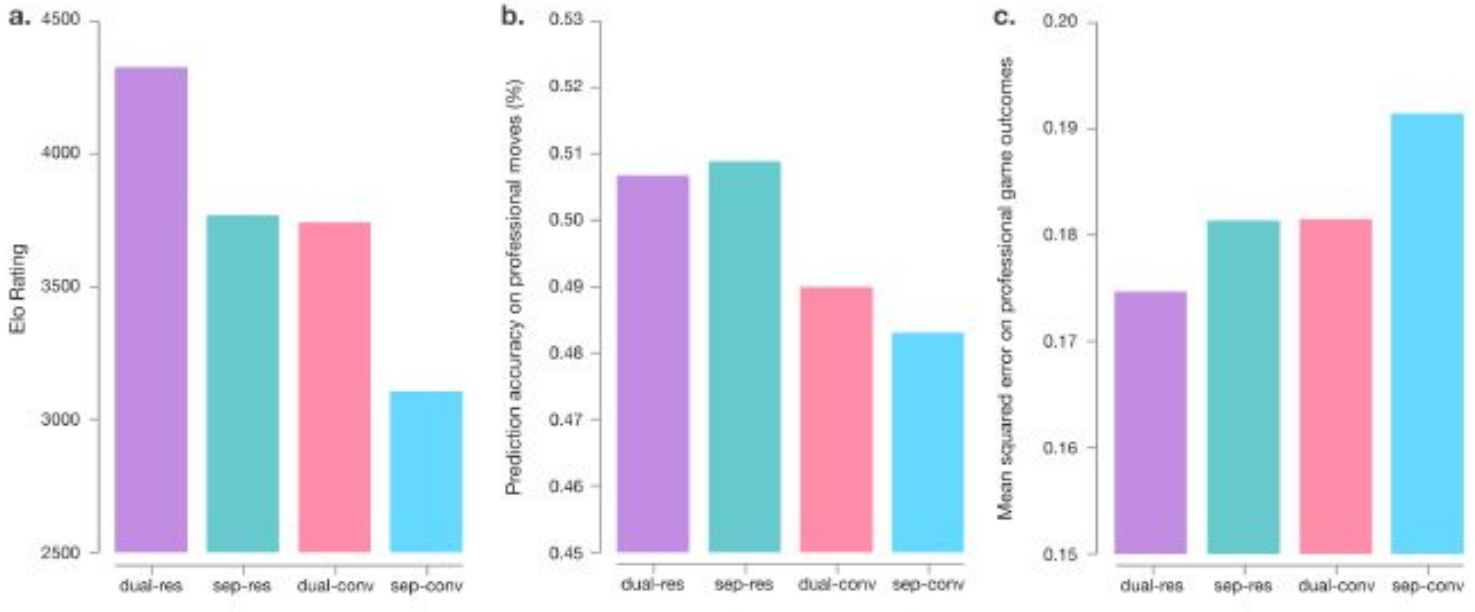
- Training for 3 days



Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

Empirical Evaluation

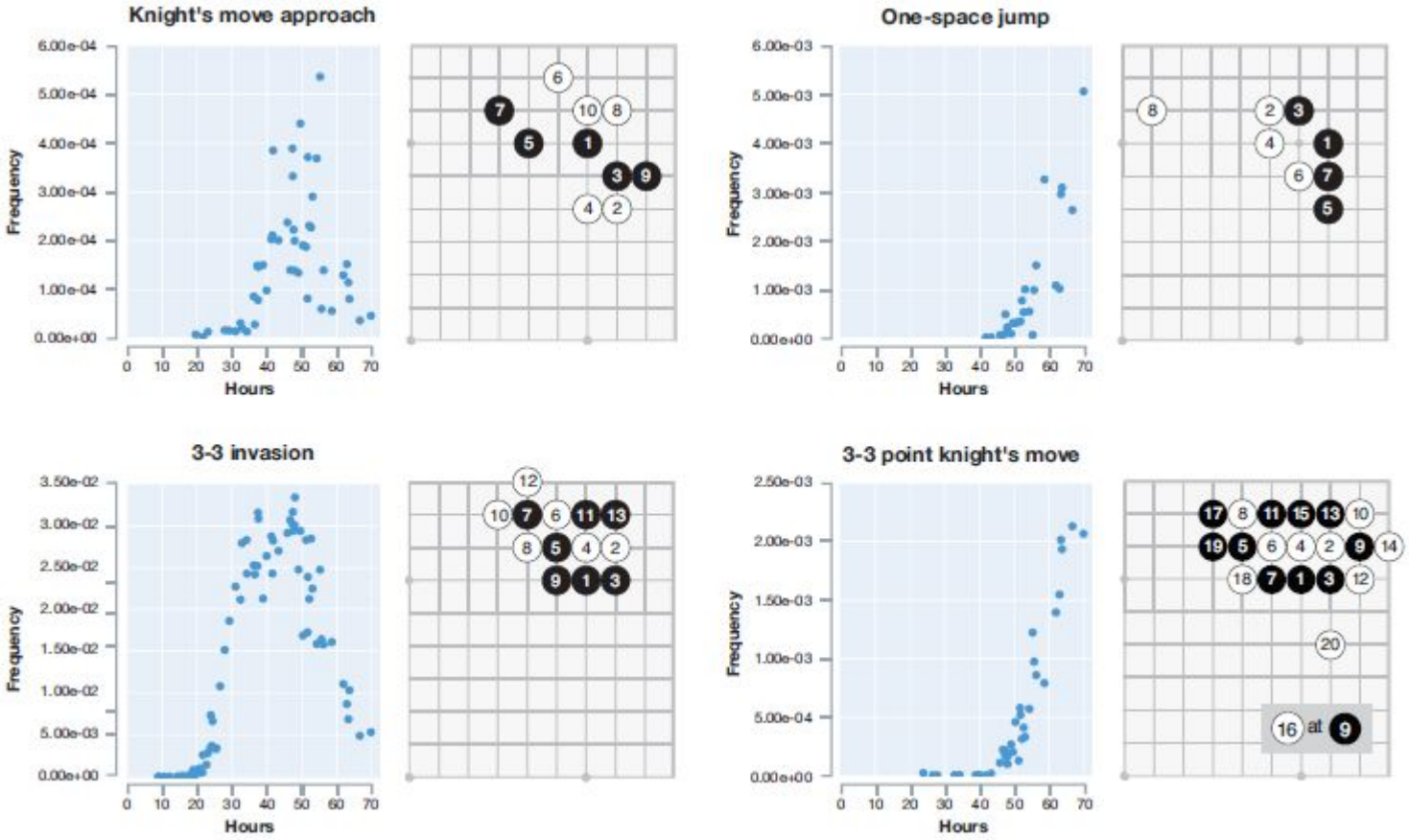
- Comparison of neural network architectures



Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

Empirical Evaluation

- Discovering existing strategies and some unknown by human

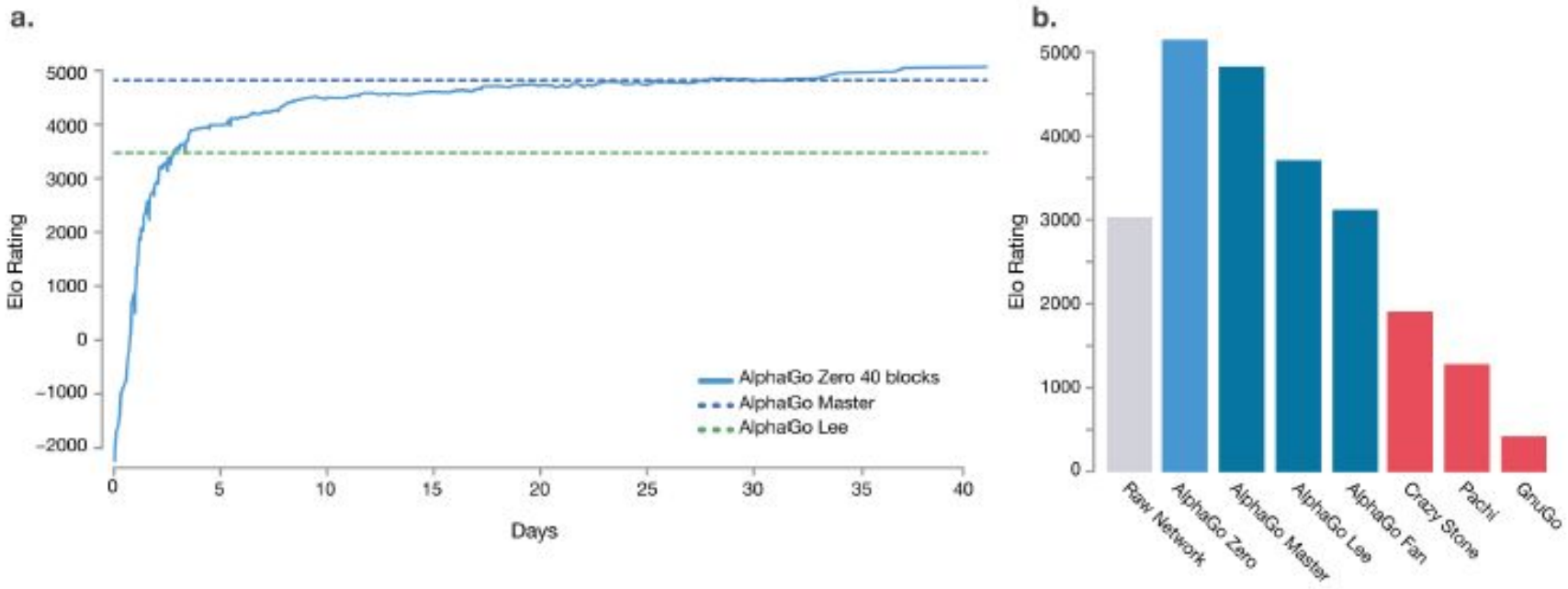


Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

UNIVERSITY OF WATERLOO

Empirical Evaluation

- Training for 40 days



Source:
Google DeepMind, Mastering the Game of Go without Human Knowledge

Conclusion

- Pure reinforcement learning is fully feasible, even in the most challenging domain
- It is possible to achieve superhuman performance, without human knowledge
- In the matter of days, AlphaGo Zero rediscover Go knowledge accumulated by human over thousands of year; it also discover new insights and strategies for the game

Discussion

- Some critics suggest AlphaGo is a very narrow AI and it rely on many properties of Go. Do you think the algorithm can be generalized for another domain?
- Did this paper inspire you in any way? Any suggestions for improvement?
- Do you think we should use AI to discover more knowledge?
- How do you feel about superintelligence AI? Are you in the Elon Musk or Mark Zuckerberg camp?

Images source:
https://jedionston.wordpress.com/2015/02/14/go-wei-chi-vs-tafl-hnafatafl/
https://www.123rf.com/photo_69824284_stock-vector-thank-you-speech-bubble-in-retro-style-vector-illustration-isolated-on-white-background.html