Research Review

References:

- Research Paper - Mastering the game of Go without human knowledge
- DeepMind article

Summary

As the research paper mentions, it has been a long-standing goal of the AI field to learn from a blank slate, a *tabula rosa* and gain superhuman proficiency in solving a problem. In this paper, a new model called AlphaZero goes on to achieve this using reinforcement learning.

AlphaGo Fan, which defeated the European champion Fan Hui, used two deep neural networks:

- *Policy network* that outputs move probabilities. It was trained initially by supervised learning to accurately predict human expert moves, and was subsequently refined by policy-gradient reinforcement learning.
- *Value network* that outputs a position evaluation. It was trained to predict the winner of games played by the policy network against itself. Once trained, these networks were combined with a Monte Carlo tree search (MCTS) to provide a lookahead search, using the policy network to narrow down the search to high-probability moves, and using the value network (in conjunction with Monte Carlo rollouts using a fast rollout policy) to evaluate positions in the tree.

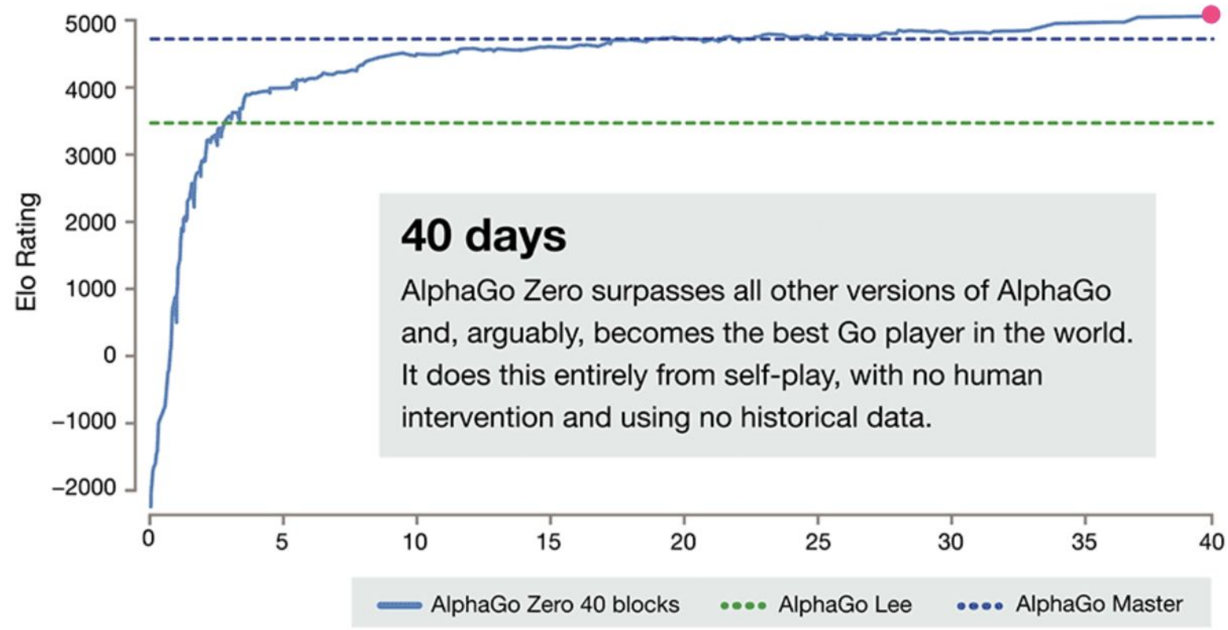
A subsequent version, AlphaGo Lee, used a similar approach and defeated Lee Sedol, the winner of 18 international titles, in March 2016.

How AlphaZero differs:

- It is trained solely by self-play reinforcement learning, starting from random play, without any supervision or use of human data
- It uses only the black and white stones from the board as input features
- It uses a single neural network, rather than separate policy and value networks
- It uses a simpler tree search that relies upon this single neural network to evaluate positions and sample moves, without performing any Monte Carlo rollouts

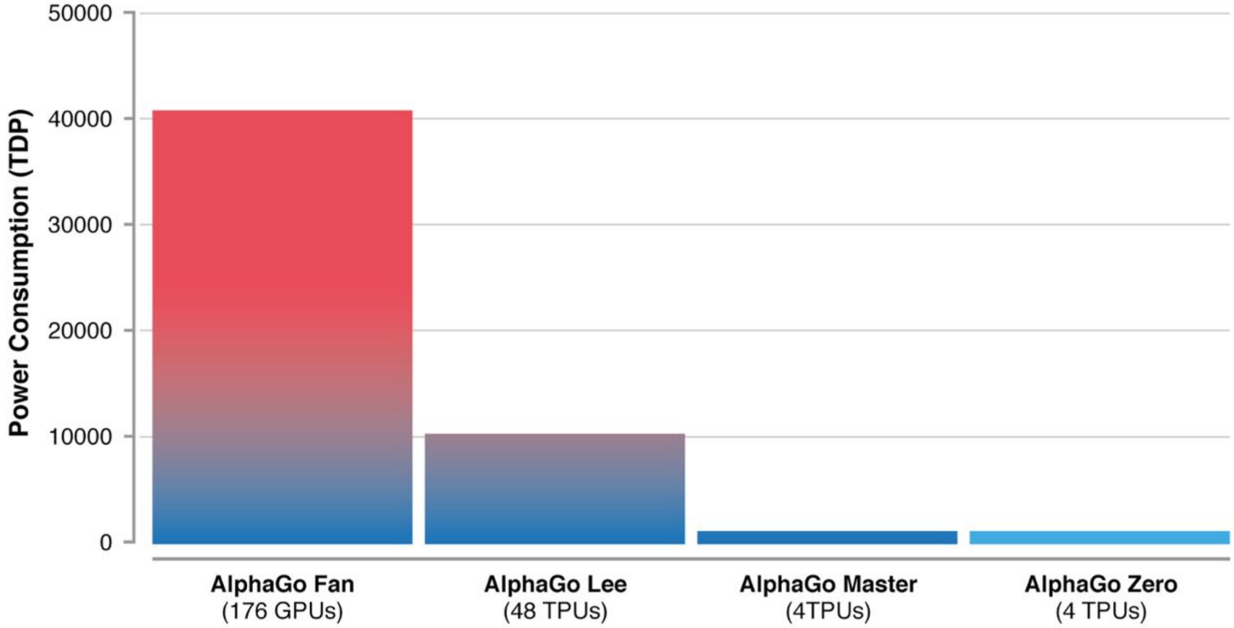
Results

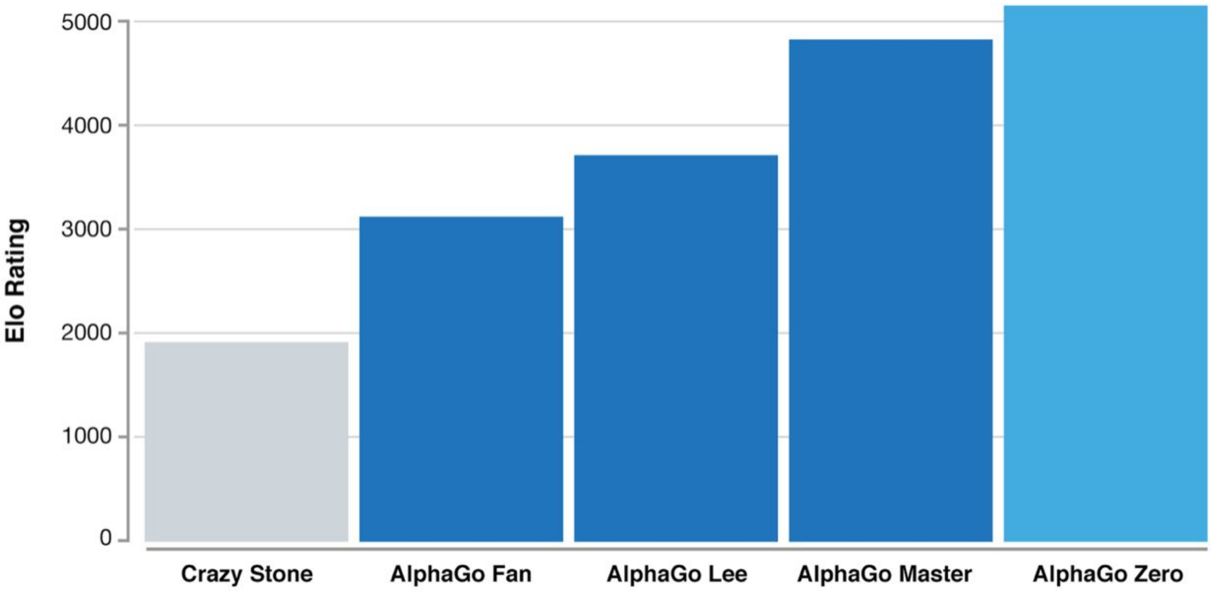
The results are astonishing across different metrics: AlphaZero surpassed AlphaGo's performance while training faster and using less resources.



40 days

AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.




Elo Rating
5000
4000
3000
2000
1000
0
Crazy Stone
AlphaGo Fan
AlphaGo Lee
AlphaGo Master
AlphaGo Zero