Mastering the game of Go without human knowledge

Authors:
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

Cons presented by:
John Fang

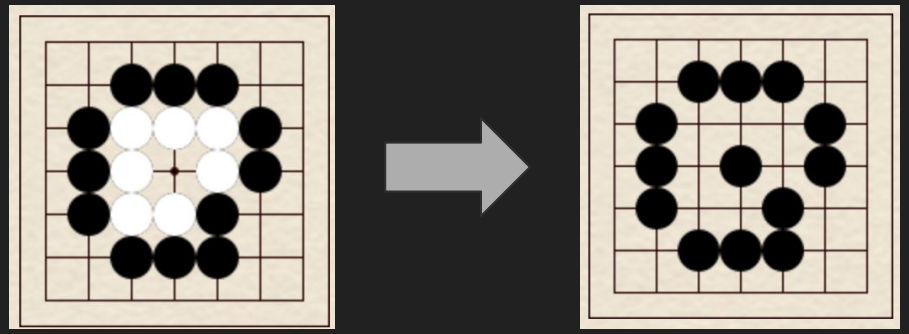
1. lack of Go context

- didn't fully address some game concepts that are important for context
  - could have reworded it to be more clear without prior knowledge
  - could have easily explained it with a diagram

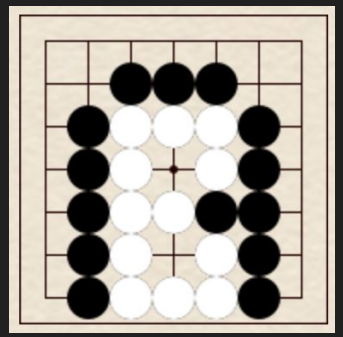
the MCTS is not augmented by any other heuristics or domain-specific rules. No legal moves are excluded—even those filling in the player's own eyes (a standard heuristic used in all previous programs[67]).

1. lack of Go context - continued

- one eyes



- two eyes

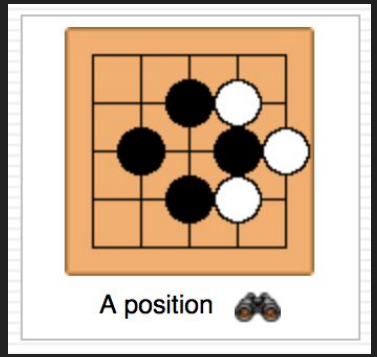


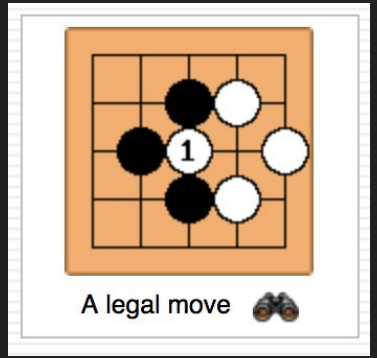
2. unexplained design decisions

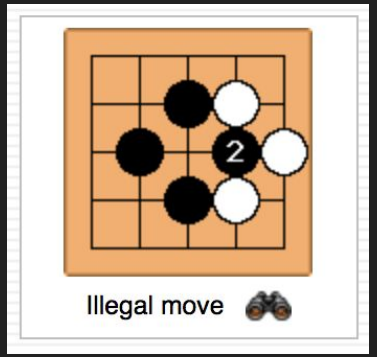
Neural network architecture. The input to the neural network is a $19 \times 19 \times 17$ image stack comprising 17 binary feature planes. Eight feature planes, X_t, consist of binary values indicating the presence of the current player's stones ($X_t^i = 1$ if intersection i contains a stone of the player's colour at time-step t; 0 if the intersection is empty, contains an opponent stone, or if $t < 0$). A further 8 feature planes, Y_t, represent the corresponding features for the opponent's stones. The final feature plane, C, represents the colour to play, and has a constant value of either 1 if black is to play or 0 if white is to play. These planes are concatenated together to give input features $s_t = [X_t, Y_t, X_{t-1}, Y_{t-1}, \ldots, X_{t-7}, Y_{t-7}, C]$. History features X_t, Y_t are necessary, because Go is not fully observable solely from the current stones, as repetitions are forbidden; similarly, the colour feature C is necessary, because the *komi* is not observable.

2. unexplained design decisions - continued

- repetitions







- rules don't imply 8 time steps are necessary

The major division in rules to prevent repetition is between the simple ko rule and the super ko rule: the simple ko rule (typically part of the Japanese ruleset) prevents repetition of the *last* previous board position; while the superko rule (typically part of Chinese derived rulesets, including those of the AGA and the New Zealand Go Society) prevents repetition of *any* previous position. In both cases, the rule does not however prohibit passing.

3. unclear terminology

- unclear use of word "leaf"
- "and finishes when the simulation reaches a leaf node s_L at time-step L."

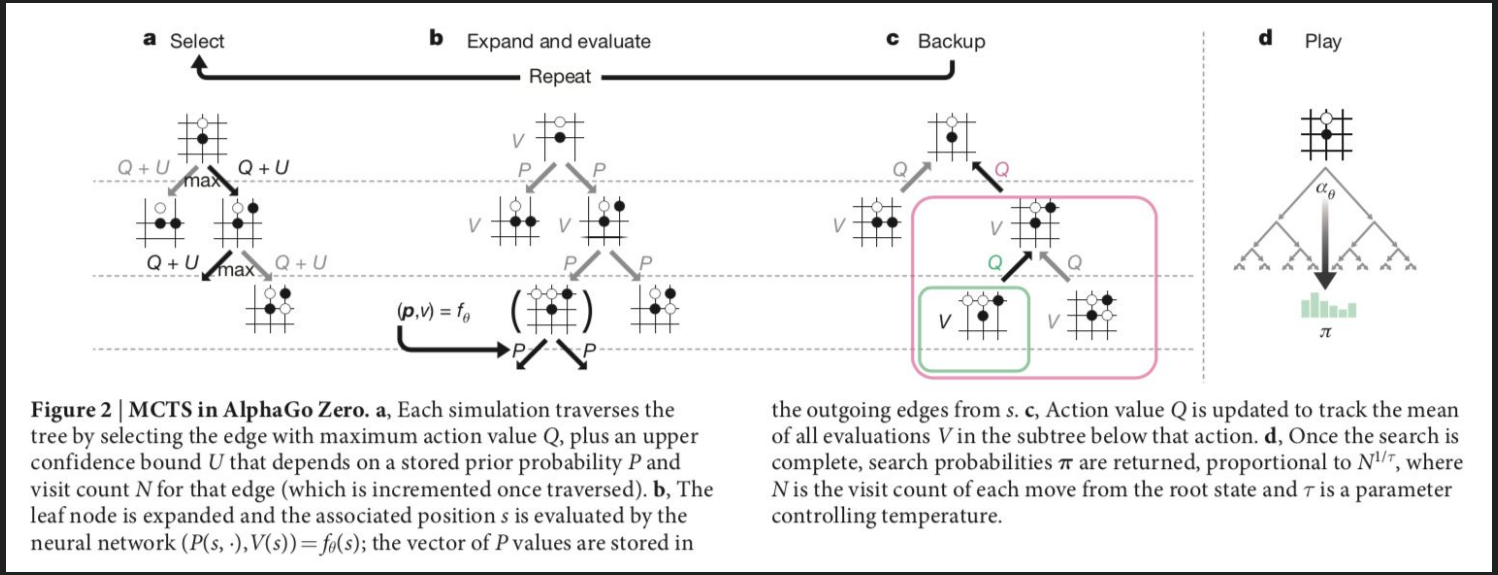


Figure 2 | MCTS in AlphaGo Zero. a, Each simulation traverses the tree by selecting the edge with maximum action value Q, plus an upper confidence bound U that depends on a stored prior probability P and visit count N for that edge (which is incremented once traversed). **b**, The leaf node is expanded and the associated position s is evaluated by the neural network $(P(s,\cdot),V(s)) = f_\theta(s)$; the vector of P values are stored in the outgoing edges from s. **c**, Action value Q is updated to track the mean of all evaluations V in the subtree below that action. **d**, Once the search is complete, search probabilities $\boldsymbol{\pi}$ are returned, proportional to $N^{1/\tau}$, where N is the visit count of each move from the root state and τ is a parameter controlling temperature.

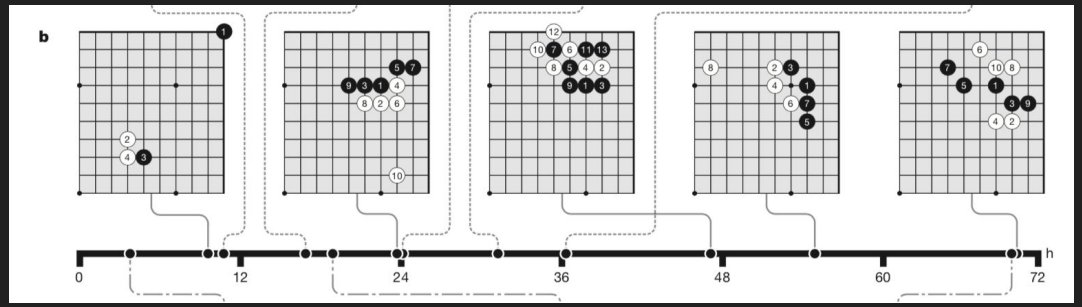
4. vague analysis

- would have been nice to describe characteristics of gameplay or what sets AlphaGo Zero apart from humans

- paper only vaguely describes differences
  - "may be learning a strategy that is qualitatively different to human play."

5. emphasis on time

- why use training time as opposed to iterations?



Humankind has accumulated Go knowledge from millions of games played over thousands of years, collectively distilled into patterns, proverbs and books. In the space of a few days, starting *tabula rasa*, AlphaGo Zero was able to rediscover much of this Go knowledge, as well as novel strategies that provide new insights into the oldest of games.