Removal of variance reduction term from update rule, as it would no longer be beneficial to the Policy Gradient Theorem [2].

No use of Monte Carlo Tree Search (MCTS) in training, allowing for the policy to be learned on larger board sizes. If successful, this suggests an effective strategy using RL for approaching large state space problems, for which we, as humans, are not experts (and cannot generate any good dataset).

A number of different design choices were made compared to Alpha Go:

- Removal of variance reduction term from update rule, as it would significantly increase the computation time.
- No use of Monte Carlo Tree Search (MCTS) in training, allowing for the policy to be learned on larger board sizes. If successful, this suggests an effective strategy using RL for approaching large state space problems, for which we, as humans, are not experts (and cannot generate any good dataset).

Transfer method 1 - convolution, or 'local' use:

- The network is applied to every possible window of the 9x9 board (as if the 5x5 network were a convolutional filter).
- The output is 5x5, by using a 5x5 convolution, and no padding.
- The output is used as input to the next layer.

Transfer method 2 - scaling, or 'global' use:

- A convolution is applied to the 9x9 board, such that the output is 5x5, by using a 5x5 convolution, and no padding.
- The output is used as input to the next layer.

References