Summary of Alpha Go

The paper introduces a new method to compute Go which has been the most challenging game in the world. The method utilizes several techniques including value network, policy network, neural network, reinforcement learning and searching.

There are three stages for training the pipeline. In the first stage of the training pipeline, Supervised learning is utilized to build on previous work on predicting expert moves. A SL policy network is trained in this stage with 57 % accuracy. In the second stage, they use policy gradient reinforcement learning (RL) to try to improve the policy network. By using the RL policy network, the rate of winning reaches 85% against Pachi without using search. In the final stage of the training pipeline, a value function $v^p(s)$ is estimated that predicts the outcome from position s of games played by using policy p for both players. AlphaGo combines the policy and value networks in an MCTS algorithm that selects actions by lookahead search.