Alpha Go Research Review
Luwei Zhang

Summary of goals and techniques:
The goal here was to create a Go playing AI capable of playing at or exceeding the level of top Go professionals (9 dan pros). The DeepMind team was able to achieve this by combining various techniques, including convolutional neural networks and Monte Carlo tree search.

A convolutional network was trained on million of grandmaster games to create a supervised learning (SL) policy network. Convolutional neural networks are typically used for image classification problems. In this case, the Go board was treated as a 19x19 image which passes through the neural network to generate predictions. This policy network was able to predict expert human moves given a Go position. However, simply training on the grandmaster games itself was not enough. There is not nearly enough data to generalize to all go positions. Therefore, the supervised learning policy network will overfit and not predict as well in real gameplay compared to the test set.

Therefore, the SL policy network is improved upon by training a reinforcement learning (RL) based policy network through self play. The RL policy network plays against previous versions itself. By playing against itself, it generates even more training data and is able to improve upon itself.

Finally, a value function is derived from the dataset generated RL based policy network via regression. The value function differs from the policy network in that it generates a single scalar value which represents the likelihood of winning the game. It is important to note that RL policy network actually does not perform as well as the SL policy network created by training on grandmaster games. This is probably because the SL policy network of GM games contains a diverse set of promising moves. However, the value network generated from the reinforcement learning generated training set performs much better from the one derived simply from the SL policy network.

During gameplay, Monte Carlo tree search is used in place of minimax search. The Monte Carlo tree search is combined with the value function to significantly reduce the space of the search tree. The Monte Carlo tree search works by simulating moves recommended by the value network down the game tree, and backwards propagating the result back to determine the optimal move.

Results:
By using these techniques, the playing strength of Alpha Go was extremely high. It was able to win 99.8% of the time against other Go programs. The distributed version of Alpha Go was able to win against Fan Hui, a 2 dan professional, 5-0 in tournament. In addition, even without rollouts (MC tree search), Alpha Go was able to beat other Go programs with the policy network alone.