Summary of: “Mastering the game of Go with deep neural networks and tree search”

This paper describes the AI behind AlphaGo, a program that plays the game Go and is able to beat all previous Go AIs and some of the best player in the world.
All previous AIs are based on Monte Carlo Search Trees - a stochastic method which tries to solve algorithmic problems by running many game simulations. Each simulation starts at a given state and stops when the game is over. Later simulations can traverse the previous played games and know which action is tending to increase the possibility to win the game. Furthermore, they are using complex handmade heuristics which are developed with professional Go players to find good solutions in key moments of the game.
AlphaGo instead is using deep learning algorithms. It is based on two components, a tree search procedure and a convolutional network which consists of three layers. The convolutional network guides the tree search procedure to rate the given state like the handmade heuristic did in the other Go AIs.
The convolutional network consists of three parts: two policy networks and a value network. Both types of networks take as input the current game state, represented as an image.
The value network provides an estimation of the current game state and tries to calculate the probability of winning the game. So you can see it as a kind of an evaluation function, besides the fact that it is learned instead of designed. “Learned” means that is was trained on 30 million game positions from separate games obtained while the policy network played against itself. The policy networks are used for the decision which action is to choose to lead the player to the win. One of the policy networks was learned with 30 million positions from the KGS Go Server. The network predicted expert moves with an accuracy of 57%. Am smaller but faster policy network was also trained. After these trainings these networks were improved by playing against each other. The outcome of a game was used as training signal to improve their behavior. This is called deep reinforcement learning. Summing it up the value network is evaluating board positions and the policy networks are selecting moves. The results of all three networks are added together in a tree search algorithm which is similar to the Monte Carlo Search trees procedure in the other AI programs (like mentioned before).

Results:

It becomes apparent that this combination is highly competitive compared to the other GO programs. Even more it seems to be stronger than that. Its winning rate against other Go programs is at 99.8 %! Furthermore, in October 2015 AlphaGo was the first AI that was able to beat a human professional player with the impressive result of 5-0 wins. Fan described the program as "very strong and stable, it seems like a wall. ... I know AlphaGo is a computer, but if no one told me, maybe I would think the player was a little strange, but a very strong player, a real person."