As the paper introduces, all games with perfect information can be studied under an optimal value function that determines the outcome of the game. This function must be evaluated in every possible state of the game, it could be done by using a search tree. The problem is, that this game tree grows exponentially where, for the game of GO. Hence, this method seems to be impossible to apply. At this point, the authors are facing the typical problem that artificial intelligence have to face: challenging decision-making (especially in go where the difficulty to evaluate board positions is outstanding) an intractable search space and a complex optimal solution. Let´s see how the team face the problem.

Basically, they introduce a new approach by using deep neural networks. They relay in the power of convolutional layers to construct the representation of the position by passing a 19x19 image of the board. By using neural networks, they reduce the depth and the breadth of the search.

In order to do that, they present the “value network” which evaluates the board position and afterwards the “policy network” starts to play and select the right moves. All these combined with a new search algorithm that combines Monte Carlo simulation with value and policy networks

They start the process by trying with supervised learning the policy network () from expert human moves. Afterwards they prepare as was done before in previous works a fast policy that can sample actions during rollouts and also train a reinforcement learning policy network which optimizes the final outcome.

The 13 layer supervised policy network has been trained by using 30 million positions, predicting expert moves with an accuracy of 57% improving by 11% the current algorithms. The reinforcement learning policy network which has the same structure as the supervised one, is supposed to increase the performance by making it play against a previous state of the policy network and using a reward function (+1 for winning and -1 for losing)