# Summary of the paper covering the following:

• <u>AlphaGo</u> by the DeepMind Team.

DeepMind Team's paper "Mastering the game of Go with deep neural networks and tree search", published on Nature 529,484–489 (28 January 2016).

Orginated in 4th century BC China ,The game of Go is probabliy one of the most ancient game played by humans world wide.

objective: Is simply occupy or surround a larger teritory on the board

why Go board game is difficult?

Go possesses more configurations than the number of atoms in the universe! often requres intiution and creative thinking Go was considered a challenge in Artifical Intelligence.

It all begin with the search trees where the AI selectes the best possible move at any given state. As exhaustive search of every possible move in Go game is infeasible.

Different approches were developed to reduce search space by using evaluation fuctions to replace a subtree below a node.And further the breadth is reduced by only considering the highest probable moves.

Even with these solutions computer performance was not upto the mark.

**Monte Carlo tree search** (MCTS) turned the tables for AI, The idea was to run many game simulations with random moves for both the players. Using the out come of these games the randomness also called the action value is reduced in the succesive simulations over time converging to a optimal play.

Alpha go uses neural network to guide the tree search operation. In total four networks are trained: A small and fast rollout policy network , two deep policy networks (SL policy and RL policy networks) used for selcting moves and finally a Value Network used to evaluate board positions.

**policy network** takes the current state of the game as input & outputs the probabilty of the game for every legal move. Authors use two levels of polciy network: **supervised learning polciy network** is trained from expert human moves, using stochastic gradient decent to maximize the probaility of human move a selected in a state s in p(a/s). A SLPN shows a 57% of accuracy for correctly predicting a human move. A **reinforcement learning policy network** is built upon an SLPN using a large number of self-play games as training data set.This excelen flow allows the system to improve itself without the need for the set of games played by humans. **value network** predicts the likelihood of the win based on the current board state.A VN is trying to approximate the value function of the RLPN. Training is done to minimize the MSE between the predicted value and the actual outcome.

lets undersatnd how these fit into place and improve the MCTS alogorithm. A move in MCTS simulation is made by maximizing the action value plus a bonus which ensures exploration of different routes during the simulations. Every leaf is evaluated by a mixture of output of value network and the result of first policy network. This step suggest a blend of intuition in the Alphago program, finally the action and bonous values at all the edges are updated after simulations we get a optimal play.

Alphago could acheive a 99.8% of win rate.

Finally i conclude that Alpha go has opened new doors to reach Super Artifical Intelligence,The question that remain is not if it is possible but when it is possible?