Introduction:

I picked the Alpha Go paper after completing the Isolation assignment because I was interested to know how the deep minds team approached a complex problem and how it would be similar/different from what we were being thought in this course. After going through the paper I could quickly relate to the problems we encountered in isolation: Like having a huge branching and depth factors (b~250, d~150) I could immediately see how hard a problem it would be to solve.

The alpha go team used policy networks and value networks similar to how we used iterative deepening and alpha beta pruning to encounter depth and branching factor issues. The alpha go team used MCTS(Monte Carlo tree Search) ,MCTS is nothing but a simulation based search tree which simulates a lot of random trees for each move or state on the board. The Alpha go team combined MCTS with policy and value networks to solve the GO problem.

Designing and Training the Policy Neural Network:

The alpha Go team used supervised learning policy networks which is a 13 layer CNN with a soft max layer to predict the probabilities on each move, the SL network is trained on 30 million expert human positions, alpha go team achieved an accuracy of 57% on the held out test set using all input features and an accuracy of 55.7% using just the raw board positions. The accuracy achieved by alpha go team was way better than other research groups.

The Reinforcement learning policy network is similar to the SL network, this network uses same parameters (weights) of the SL network at initialization and is pitted against itself from the previous iteration, so it learns from playing with itself instead of being trained on human moves. The RL policy network when evaluated against the SL Networks won more than 80% of the times.

Designing the Value Neural Network:

The Value Neural Networks are the final set of Reinforcement Learning Neural Network implementation for validating if the selected move is the best move or not i.e. a win or not a typical classification problem. It was first trained against the expert human positions data and when evaluated against the test set there was an overfitting problem, when trained on the data generated from RL policy network it generalized very well that is the test and train accuracies where close.

Results:

AlphaGo algorithm achieved strongest human player level using a combination of neural network evaluations with Monte Carlo rollouts.

Alhpha Go algorithm reached a major milestone in AI by beating a human player in GO. This feat was believed to be at least a decade away.

The Alpha Go algorithm outperforms any other go- playing algorithms just by using the SL network of its implementation.