Research Review
Article: **Mastering the game of Go with deep neural networks and tree search**

a) Goals and Techniques introduced:

Traditionally in game playing, exhaustive game tree searching and Monte Carlo tree search are widely used. But recent developments in deep convolution neural network yields very great performance in visual domains including image classification and playing Atari games. The goal of this paper is to take advantage of the recent development in deep convolutional neural network, and employ a similar architecture to play the game of Go.

In this paper, the board position is pass as a 19 by 19 images, and then a convolutional layer is used to construct a representation the board position. There are two major network in this architecture: a value network for evaluating board positions, and a policy network to sampling actions.

The training of this network consists several stages of machine learning. 1) A supervised learning policy network trained from expert human moves. 2) A reinforcement policy network optimizing the final outcome in self-playing games. 3) A value network predicts the winner of games played by the reinforcement policy network against itself. Finally these policy and value networks are combined using Monte Carlo Trees.

b) Results:

As an evaluation of this architecture, variants of AlphaGo competes with several others Go playing programs. All of these other Go playing programs are based on high-performance Monte Carlo Tree Search algorithm. All programs are allowed a computation time of 5 second per move. As a result, single machine AlphaGo wins (494 out of 495 games (99.8%) against other Go playing programs. In games with free moves for opponent, AlphaGo still win 77%, 86%, and 99% against Crazy Stone, Zen and Pachi respectively. As a significantly stronger version, the distributed version of AlphaGo wins 77% games against single-machine AlphaGo and 100% against other programs.

While assessing variants of AlphaGo that evaluates positions using just the value network or just the rollouts, AlphaGo out-perform all other Go program even without rollouts. This shows the value network can be a viable alternative for the Monte Carlo evaluation. However the mixed evaluation performed best, winning >=95% against all other variants. This indicates the two position-evaluation networks are complementary: value network approximates the outcome played by strong but slow reinforcement learning policy network, while the rollouts evaluate the outcome played by weaker by faster supervised learning policy network.

Finally, AlphaGo completes in a formal five-game match against Fan Hui, who is the winner of 2013, 2014 and 2015 European Go championships. AlphaGo wons the match 5 games to 0.