Paper Name:
AlphaGo - Mastering the game of Go with deep neural networks and tree search

Paper Goal:
A game player for Go

Techniques:

- ➔ Summary:
  - ◆ Search Algorithm: Monte Carlo simulation with value and policy networks
  - ◆ Evaluation function: value networks
  - ◆ Select move: policy networks
- ➔ Training strategy
  - ◆ Supervised learning
    - Predict human expert next move
    - Efficient learning update and high-quality gradient
  - ◆ Reinforcement learning
    - Optimize outcome (winning rate) of self-play game
    - Adjust policy
- ➔ First stage:
  - ◆ Goal: predict expert moves
  - ◆ Model: a random rollout policy + supervised learning policy
  - ◆ Optimization: maximize the likelihood of a selected human move in state s
  - ◆ Training approach: SGD
- ➔ Second stage
  - ◆ Goal: maximizing the winning rate
  - ◆ Model: RL policy network
  - ◆ Training approach: policy gradient reinforcement learning
  - ◆ Initialize weight to SL network
- ➔ Final stage
  - ◆ Goal: predict outcome from a position
  - ◆ Model: value network
  - ◆ Optimization: minimize MSE state-outcome pair
  - ◆ Training approach: regression with SGD
- ➔ Search
  - ◆ Monte Carlo search tree
  - ◆ A leaf node is evaluated by
    - Value network
    - The outcome of a random rollout played out until the terminal

Results:

- ➔ 99.8% winning rate against other Go programs
- ➔ Defeat the human European Go champion by 5 games to 0