SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Redesigning Common Mind & Business Towards Excellence

Build an Entrepreneurial Mindset Through Our Design Thinking FrameWork

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23AMB201 - MACHINE LEARNING

II YEAR IV SEM

UNIT V – REINFORCEMENT LEARNING

TOPIC 5 – Model Based Learning – Model Free Learning

Introduction to AlphaGo

- Developed by DeepMind
- First AI to defeat a world champion in the game of Go
- Combined Deep Learning and Reinforcement Learning
- Massive breakthrough in AI history

Why Go is Challenging for AI

- State space: 10^170 (more than atoms in the universe)
- Requires long-term strategy and intuition
- Reward is sparse (only at game end)
- Traditional brute-force search (like in chess) is ineffective

AlphaGo's Core Components

- **Policy Network (π):** Suggests strong moves
- **Improved Policy Network (π'):** Learned through self-play
- **Value Network (V):** Predicts win probability from a board state
- **Monte Carlo Tree Search (MCTS):** Efficiently explores move sequences

Step 1 - Supervised Learning (Policy Network)

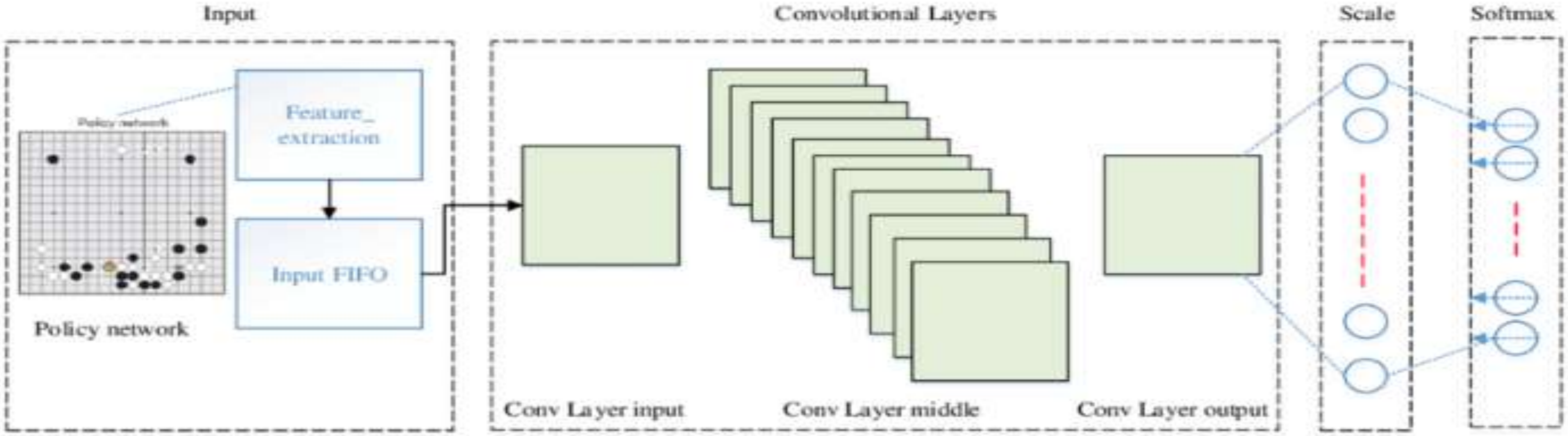
- Trained on human expert games
- Learns to predict next move from a board state
- Gives AlphaGo an expert-like base to start from
- Technique: Deep CNN + cross-entropy loss

Step 2 - Reinforcement Learning (Improved Policy)

- Self-play: AlphaGo plays against itself
- Learns which moves lead to more wins
- Improves policy beyond human-level
- Technique: Policy Gradient RL

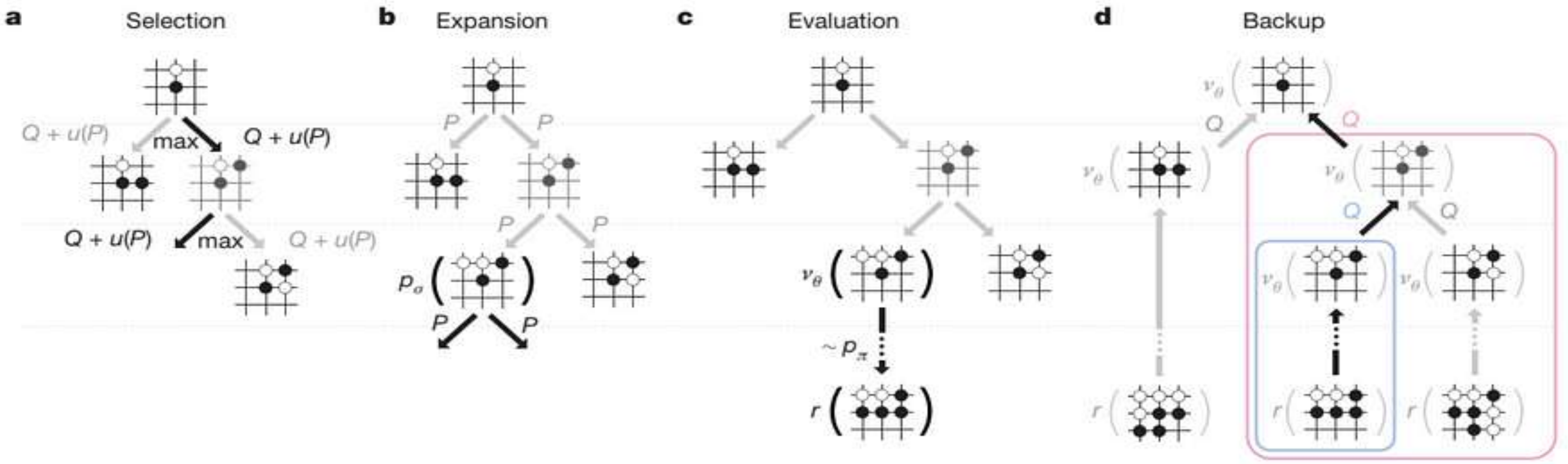
Step 3 - Value Network

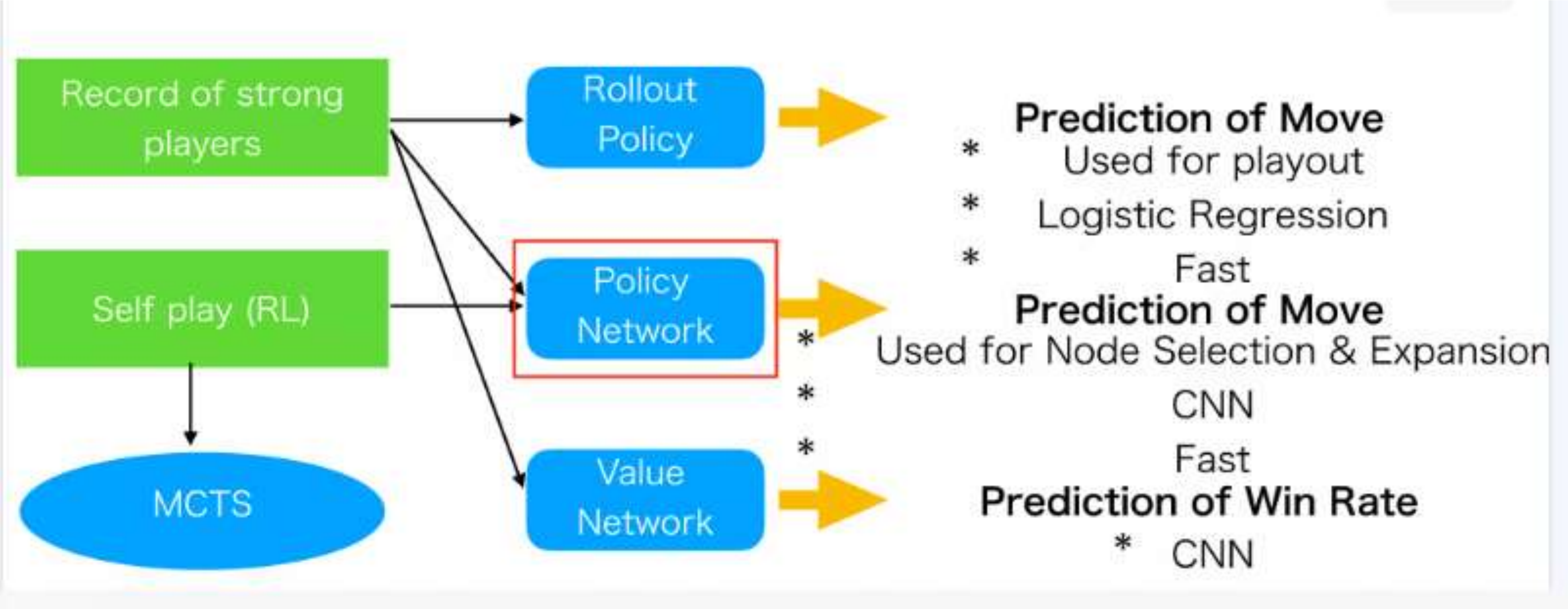
- Predicts expected outcome (win/loss) from any position
- Trained on outcomes of self-play games
- Eliminates need to simulate till end
- Technique: Deep regression with reinforcement signals



Step 4 - Monte Carlo Tree Search (MCTS)

- Monte Carlo Tree Search (MCTS) is the algorithm we use to prioritize and build this search tree. It composes of 4 steps below.
- Simulates future sequences of moves
- Policy Network guides exploration (prioritizes good moves)
Value Network evaluates board states at tree leaves
Smart balance between exploration and exploitation





Record of strong players
Self play (RL)
MCTS
Rollout Policy
Policy Network
Value Network
Prediction of Move
* Used for playout
* Logistic Regression
* Fast
Prediction of Move
* Used for Node Selection & Expansion
* CNN
* Fast
Prediction of Win Rate
* CNN

Policy Network: Overview

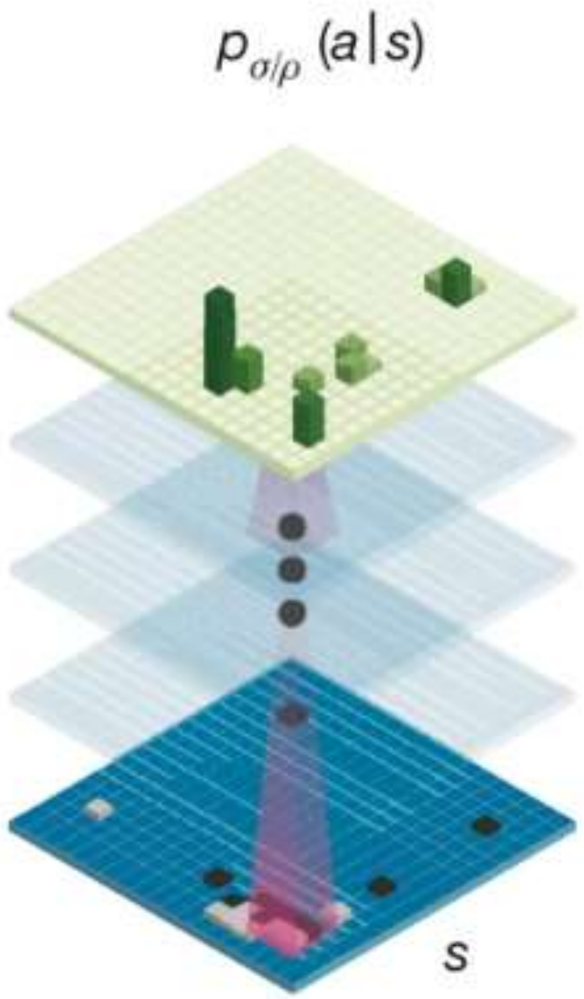
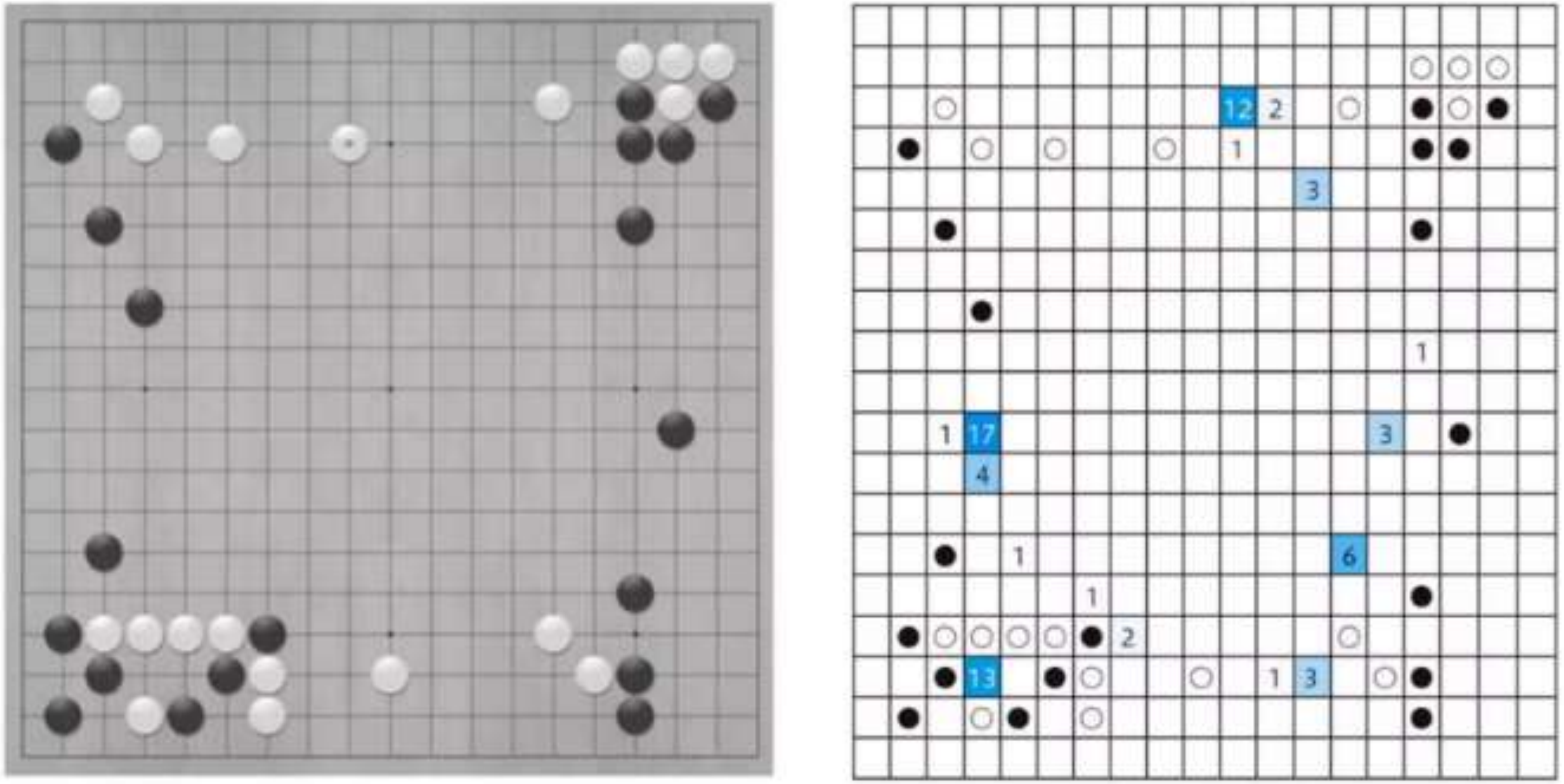


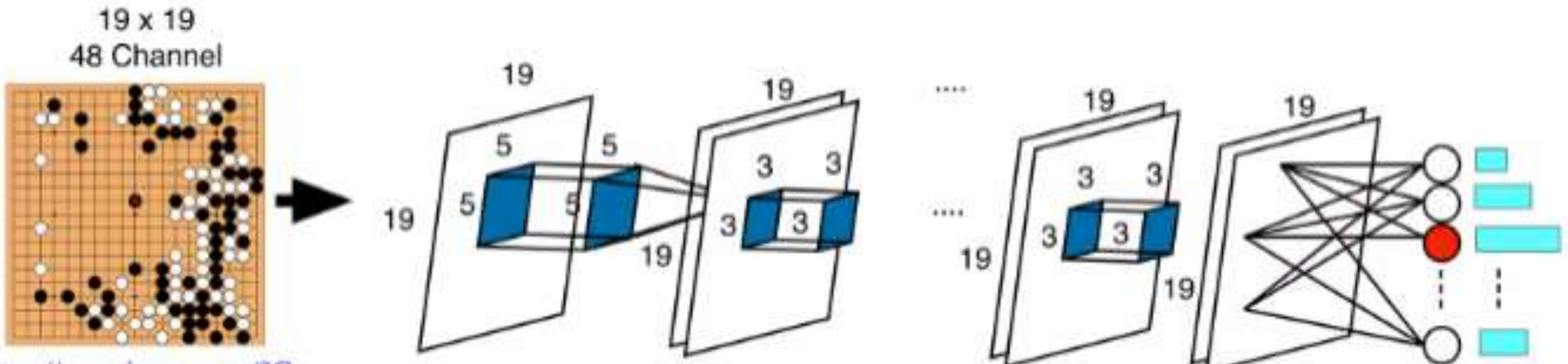
Fig1. of (Silver 2016)

- Convolutional Neural Network
- The network first is trained by supervised learning algorithm and later refined by reinforcement learning
- Trained with KGS dataset. 29.4 million positions from 160000 games played by KGS 6 to 9 dan



Output is percentage Fig. 2.18 (Otsuki 2017)

- Convolutional Neural Network
- Trained with KGS dataset. 29.4 million positions from 160000 games played by KGS 6 to 9 dan
- 48 Channels (Features) is prepared (Next slide explains details).



https://senseis.xmp.net/?Go

Output: Prob. of the next move

- They further trained the policy network by policy gradient reinforcement learning.
- Training is done by self-play
- The win rate of the RL policy network over the original SL policy network was 80%

Summary Table

Component	Technique	Purpose
Policy Network	Supervised Learning	Mimic expert moves
Improved Policy	Reinforcement Learning	Improve via self-play
Value Network	Deep RL Regression	Predict game outcomes
MCTS	Guided Tree Search	Efficient move exploration

Impact of AlphaGo

Proved RL can solve real-world complex problems

Inspired AlphaGo Zero, AlphaZero, MuZero

Techniques used in protein folding (AlphaFold)

Advanced game-playing, robotics, healthcare, and more

Key Takeaways

AlphaGo = Deep Learning + RL + Self-Play + Search

Breakthrough in strategy game AI

Set the foundation for general-purpose AI systems