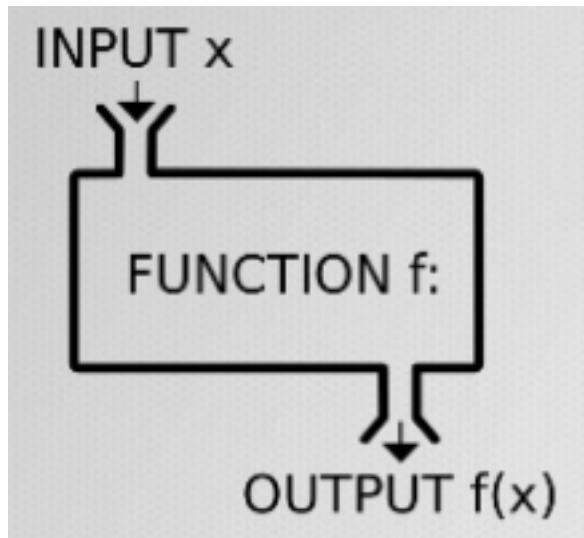


알파고 학습 이해하기

(주)엑셈 / 임도형

신경망(Neural Network)

입출력의 함수를 근사화 할 수 있다.



□ Speech Recognition

$f(\text{audio waveform}) = \text{"你好"}$

□ Image Recognition

$f(\text{cat image}) = \text{cat}$

□ Go Playing

$f(\text{Go board state}) = \text{5-5 (next move)}$

□ Dialogue System

$f(\text{"台積電怎麼去"}) = \text{"地址為... 現在建議搭乘計程車"}$

신경망의 학습

입력과 출력을 반복적으로 주고, 함수를 근사화 해가는 과정.

상세 방법은 설명하지 않지만, 하여간에 학습됩니다.

데이터

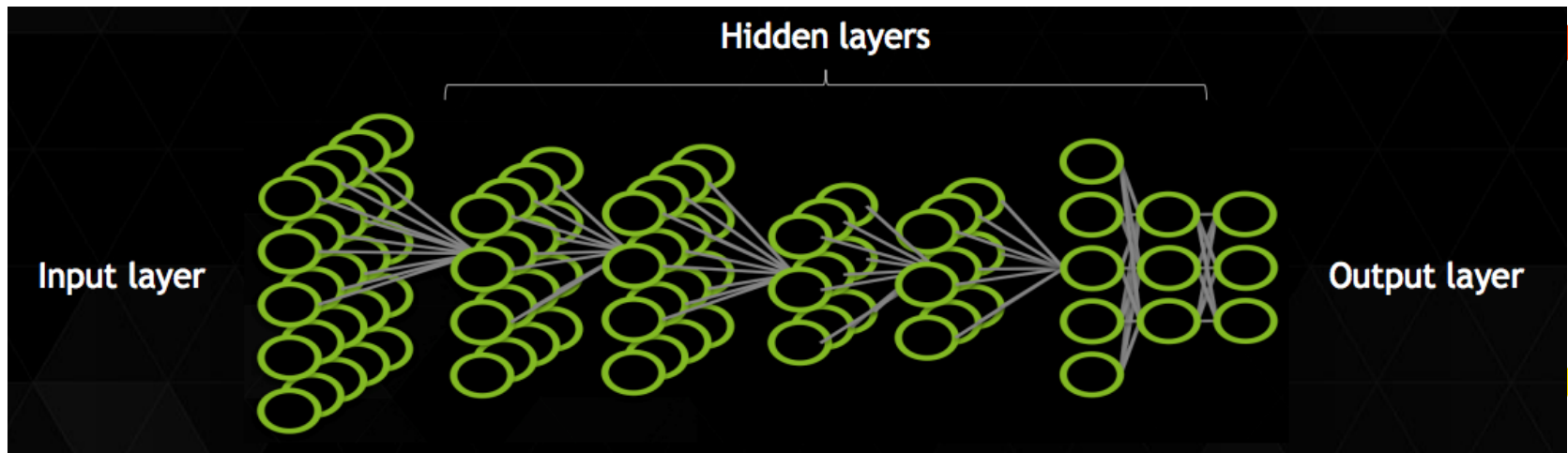
학습에 사용할 데이터가 있어 합니다.

학습된 후에는 처음 본 것도 인식 가능합니다.

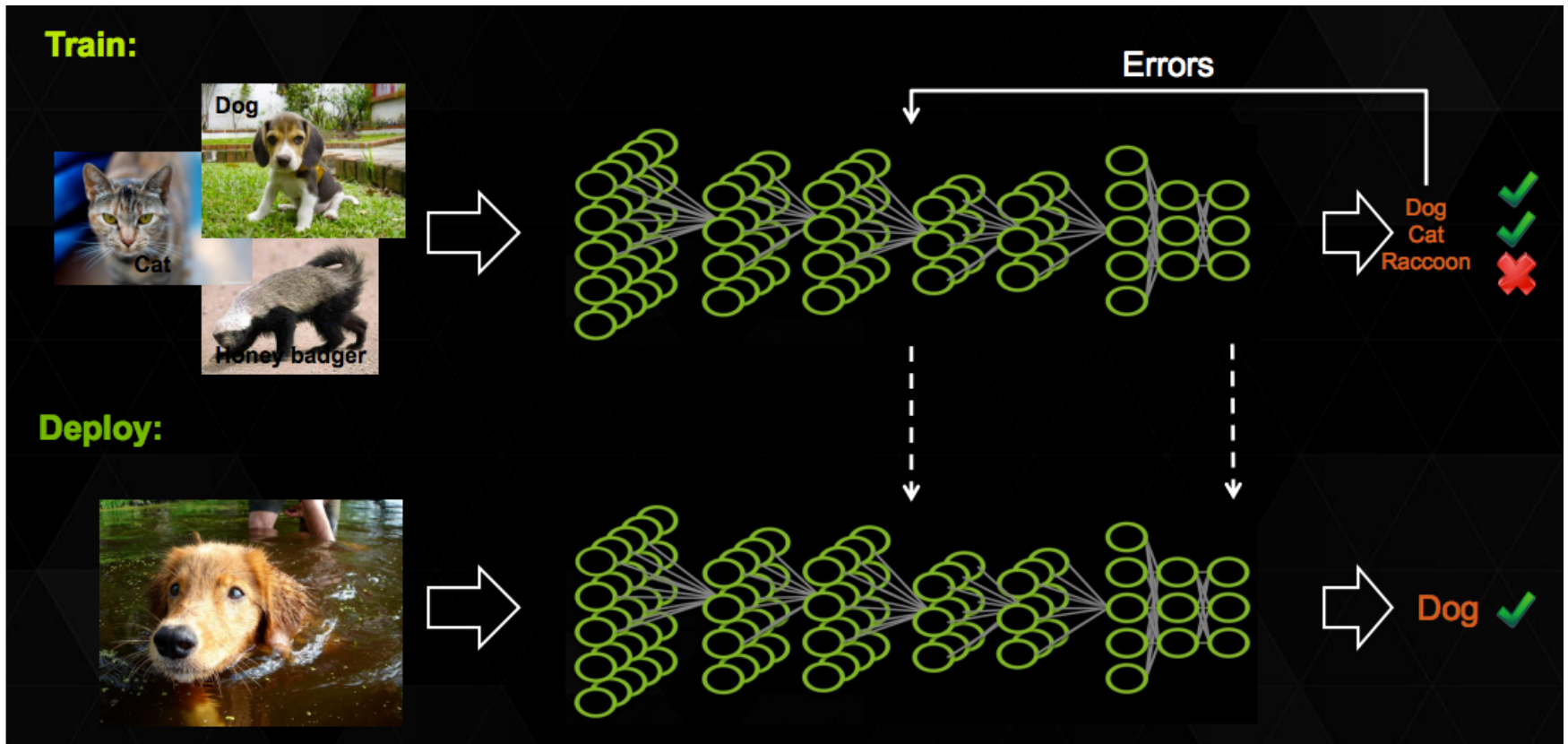
딥러닝

신경망인데, 가운데있는 층이 많은 것을 DNN(Deep NN)라고 합니다.

그리고 DNN이 학습하는 것을 Deep Learning이라고 합니다.

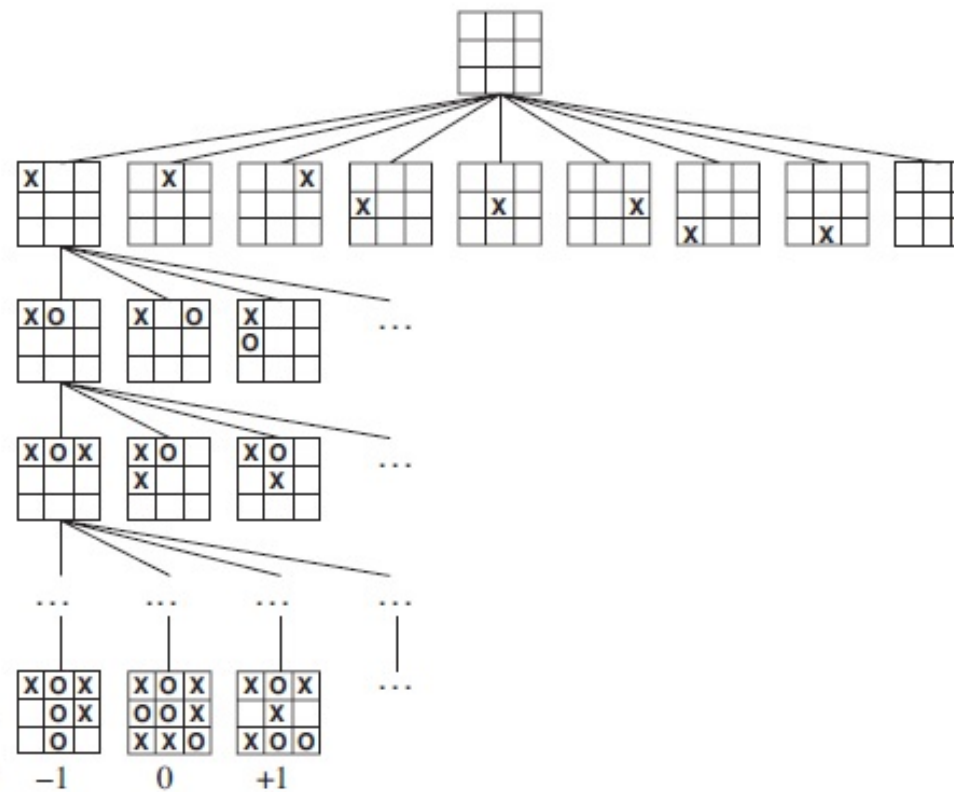


학습과 사용



틱택토

모든 경우의 수 탐색가능
 $9 \times 8 \times \dots \times 2$

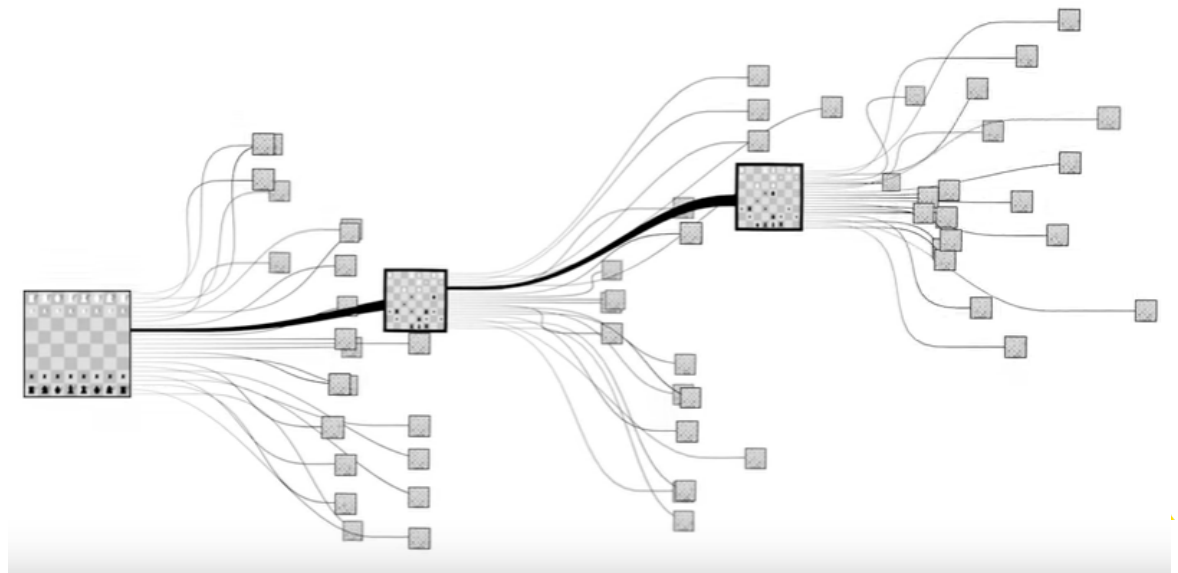


<http://blog.naver.com/PostView.nhn?blogId=best999&logNo=220680107733&parentCategoryNo=&categoryNo=&viewDate=&isShowPopularPosts=false&from=postView>

체스

1996년 딥블루

경우의 수는 탐색할 만 하다.

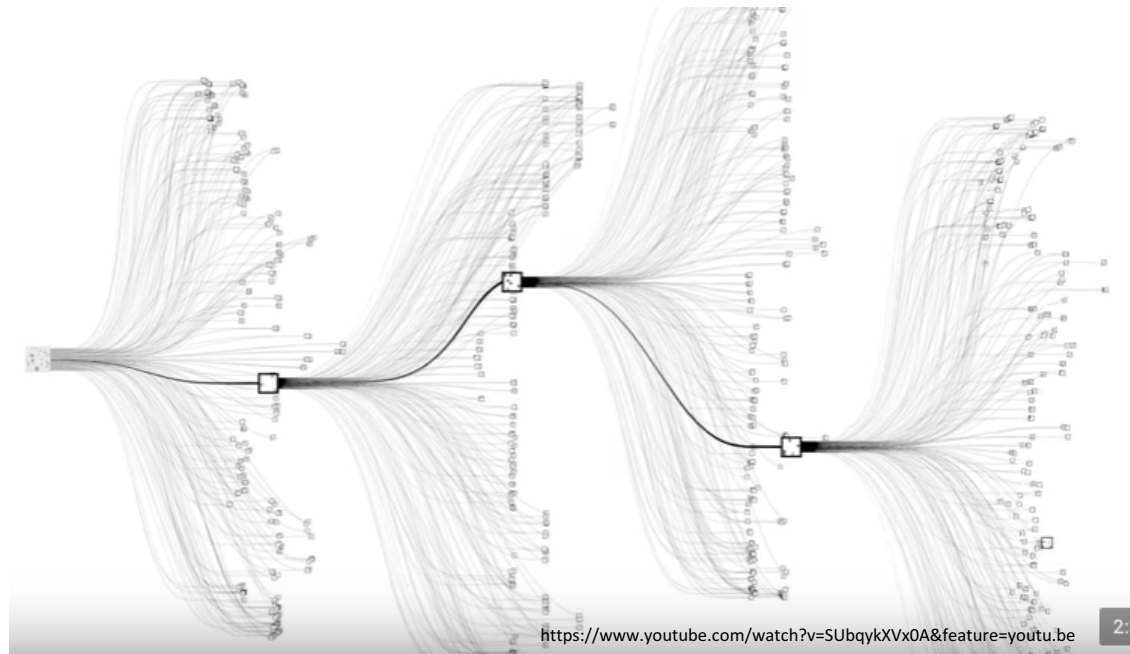


<https://www.youtube.com/watch?v=SUBqykXVx0A&feature=youtu.be>

바둑

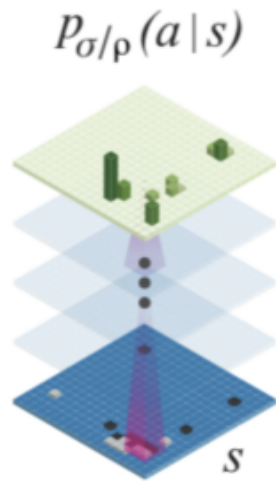
모든 경우의 수 탐색 불가.

$361 \times 360 \times \dots > \text{우주 원자 수}$

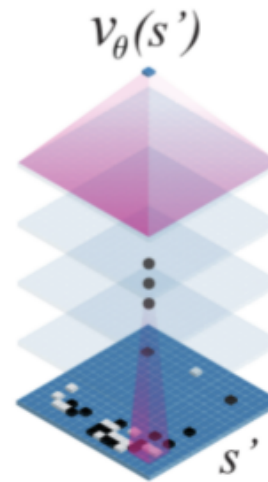


AlphaGo를 요약하면

- 2개의 네트워크 + MCTS



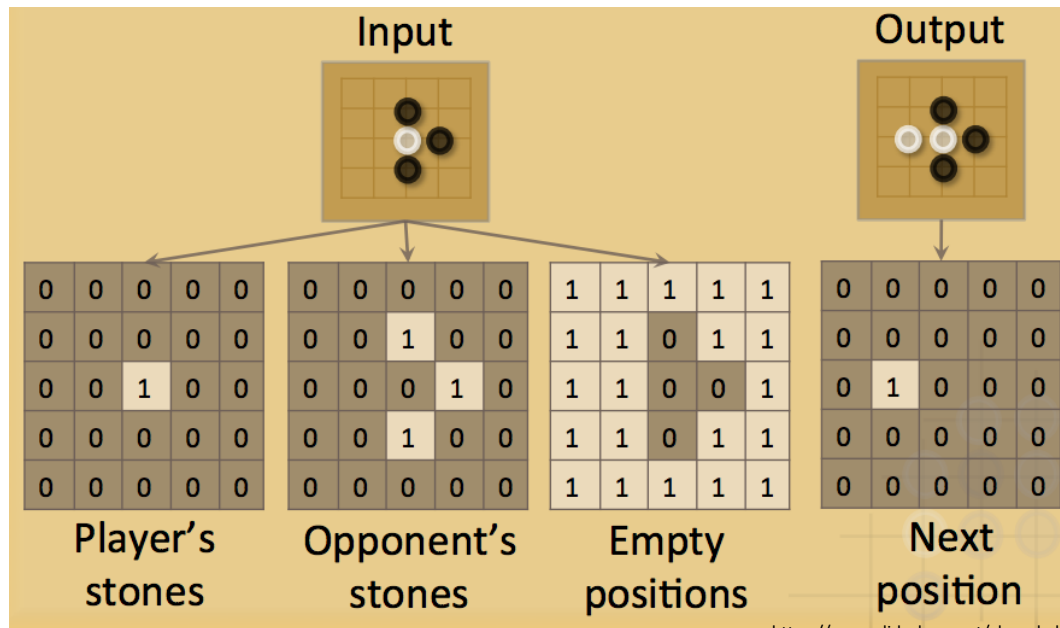
Policy Network



Value Network

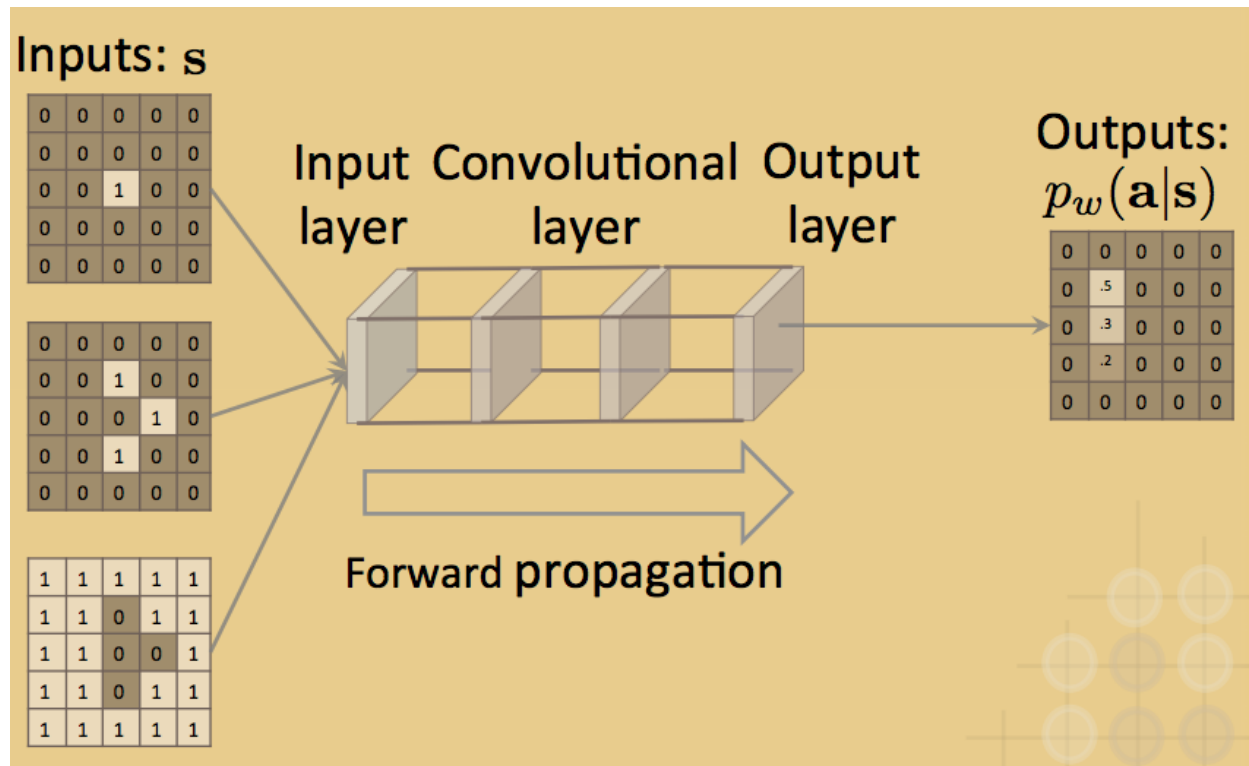
Policy Network

현재의 판에 대한 다음 수를 학습
프로기사의 기보를 사용.



<https://www.slideshare.net/ckmarkohchang/alphago-in-depth>

Policy Network

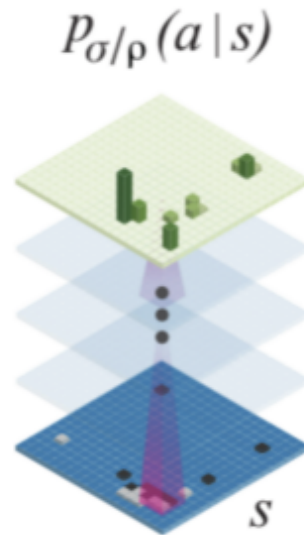


Policy Network - 강화학습

- 학습이 된 Policy Network를 가지고 스스로 대국을 두게 함.
- 그 결과를 가지고 네트워크를 업데이트.

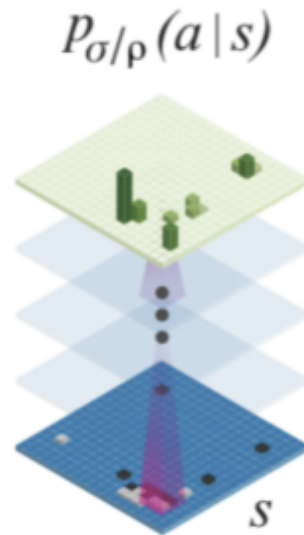
Policy Network

- 바둑 프로그램 Pachi를 상대로 85% 승율

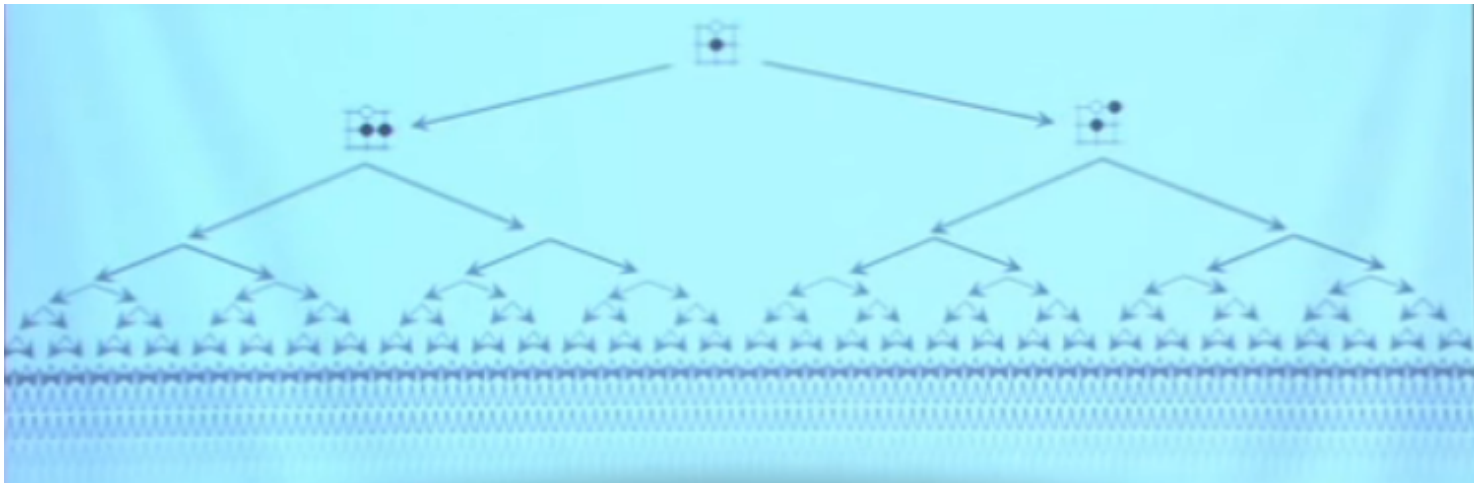


Policy Network의 역할

- 검색 폭을 줄인다.
- 출력된 수만을 고려한다.



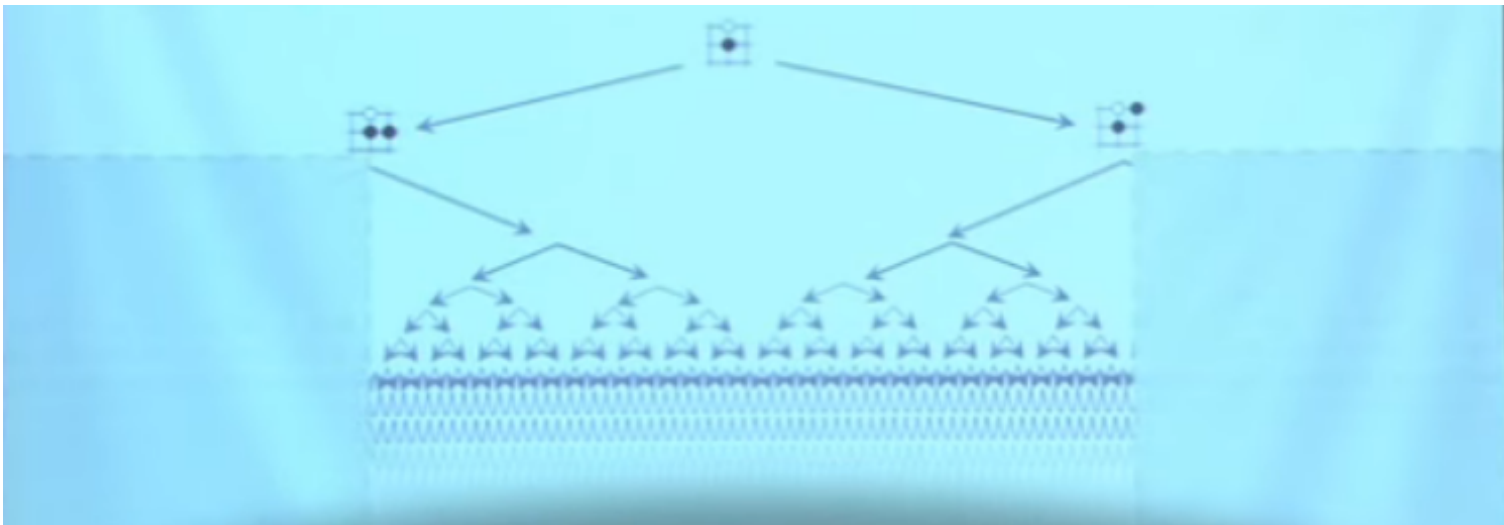
Policy Network의 역할



<https://www.youtube.com/watch?v=8Z2eLTSCuBk&t=0s>

Policy Network의 역할

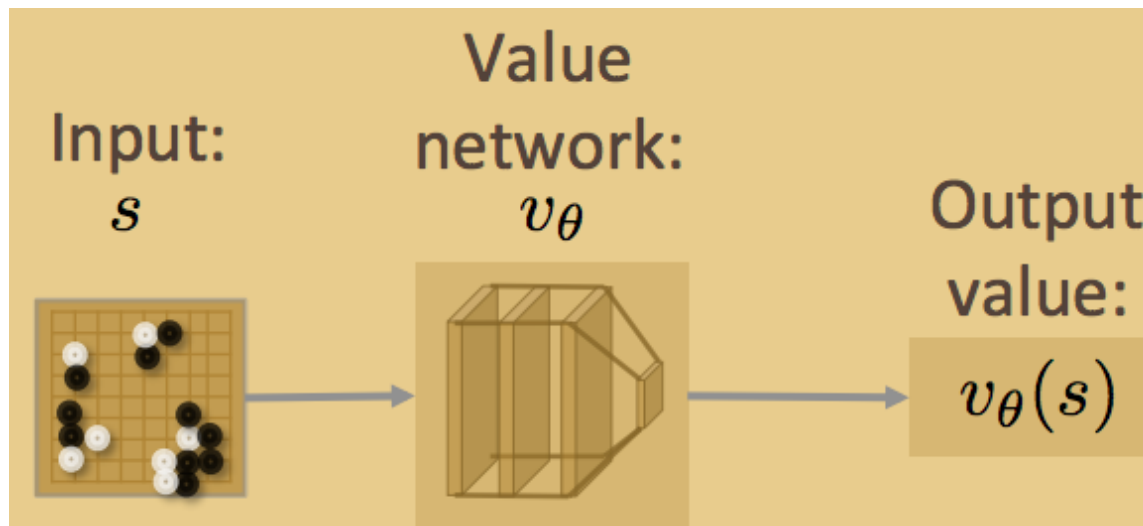
- 모든 가능한 수를 다 검색하지 않아도 된다.



<https://www.youtube.com/watch?v=8Z2eLTSCuBk&t=0s>

Value Network

- 현재의 판을 가지고 유리/불리를 학습
- 최종 결과를 출력으로 학습. 1: 승리, 0: 패배



Value Network

- Policy Network으로 스스로 플레이한 기보 데이터를 사용

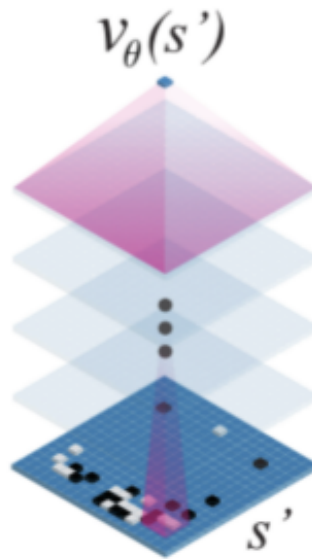


VS



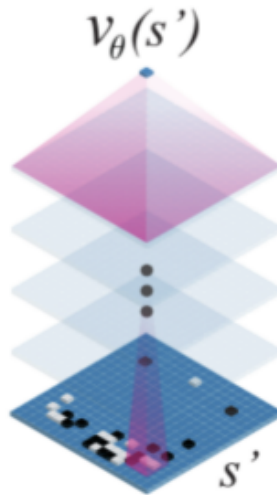
Value Network

- 현재 판을 보고 유리/불리를 출력

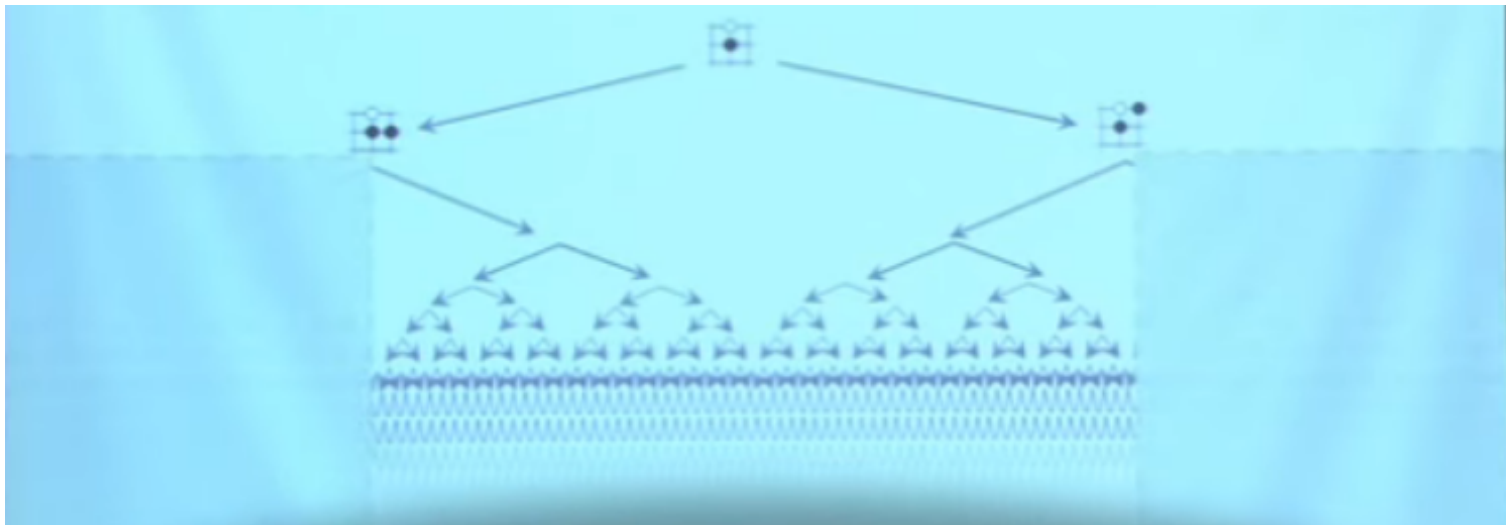


Value Network의 역할

- 검색 깊이를 줄인다.



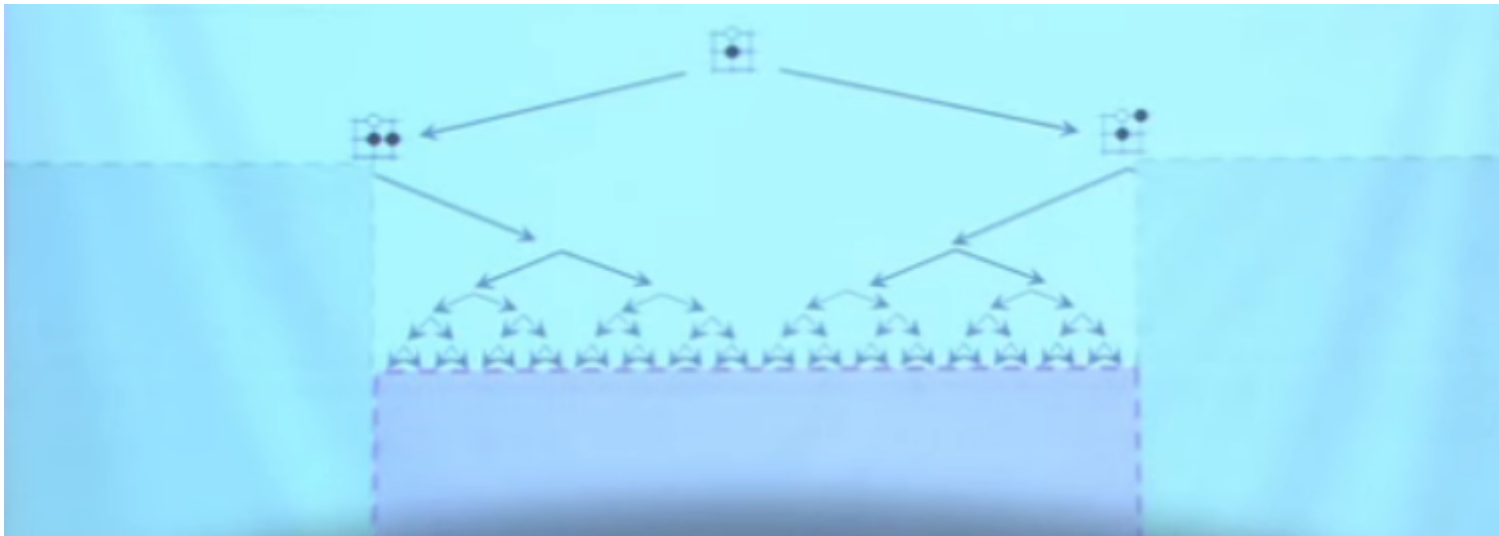
Value Network의 역할



<https://www.youtube.com/watch?v=8Z2eLTSCuBk&t=0s>

Value Network의 역할

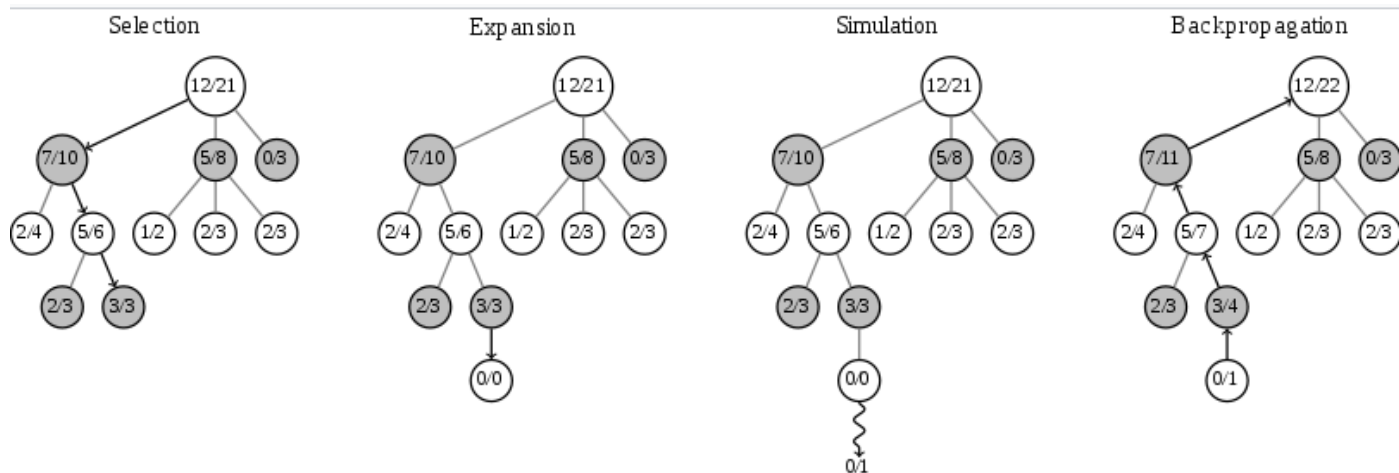
- 게임의 끝까지 가보지 않아도 된다.



<https://www.youtube.com/watch?v=8Z2eLTSCuBk&t=0s>

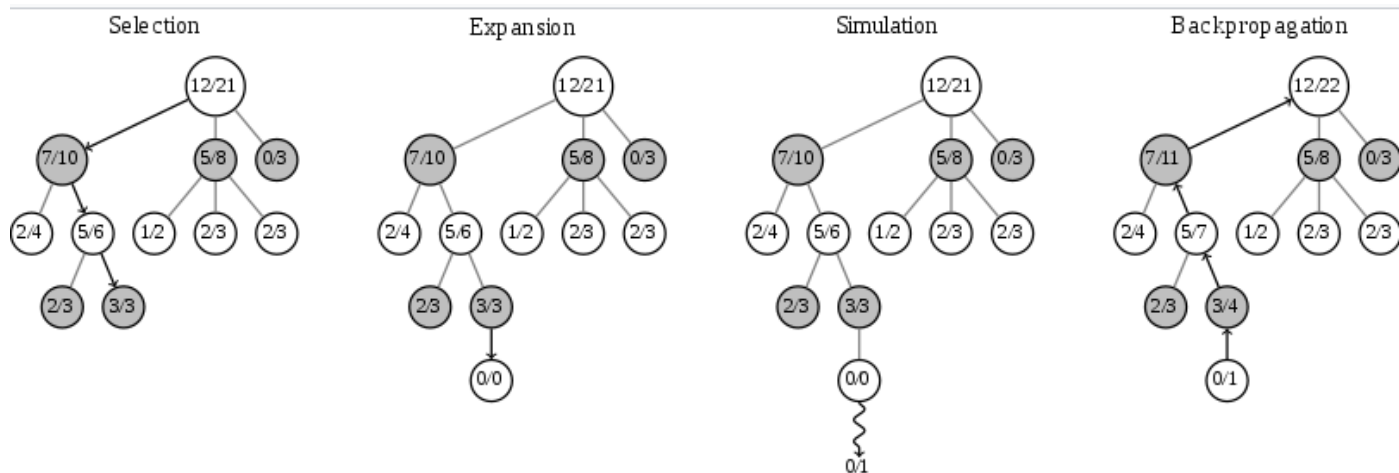
MCTS

- Monte-Carlo Tree Search
- 트리 검색 방법
- 이전 최강 프로그램 Pachi이 사용



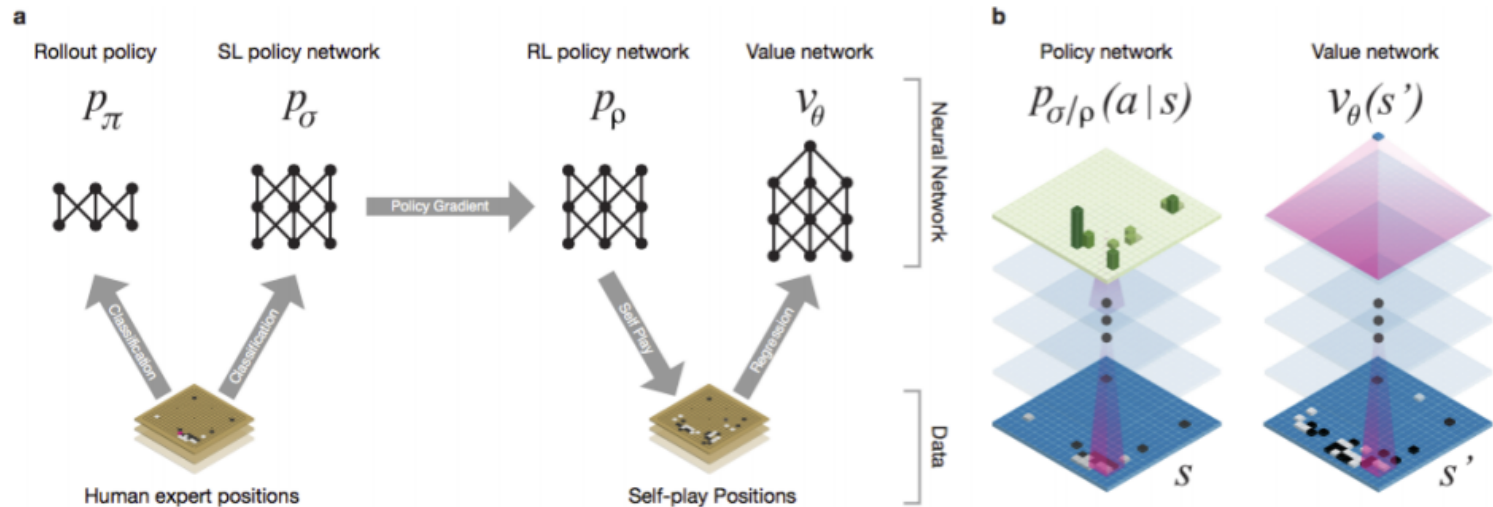
MCTS

- Alphago의 경우 끝까지 가지 않고, 특정 깊이에서 Value Network의 값을 사용.



정리하면

- Policy Network을 사용하여 검색 폭을 줄이고
- MCTS를 사용하여 최선의 가지를 찾고
- Value Network을 사용 하여 검색 깊이를 줄이고



<https://gogameguru.com/i/2016/03/deepmind-mastering-go.pdf>

Reference

- Mastering the Game of Go with Deep Neural Networks and Tree Search <https://gogameguru.com/i/2016/03/deepmind-mastering-go.pdf>
- Google DeepMind's AlphaGo: How it works : <https://www.tastehit.com/blog/google-deepmind-alphago-how-it-works/>
- Alphago in Depth : <https://www.slideshare.net/ckmarkohchang/alphago-in-depth>