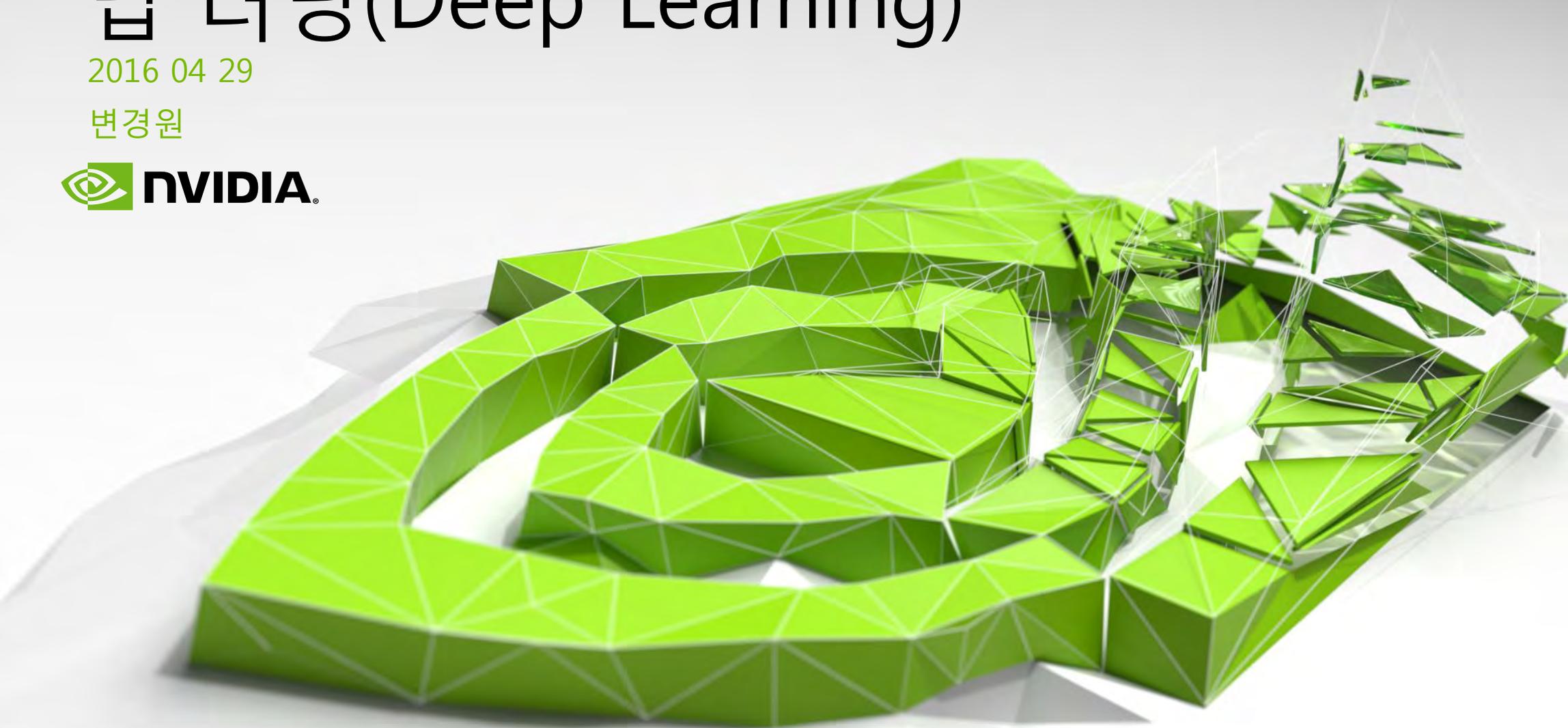


# 딥 러닝(Deep Learning)

2016 04 29

변경원



# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 딥러닝이란 무엇인가?

## 딥러닝과 인공지능

딥러닝은 세상을 이해하고 감지하는 인공지능을 개발하는데 가장 촉망받는 기술이 되고 있음.

최근에는 여러가지 인지능력 프로젝트에 집중되고 있고 수많은 성공 사례들이 발표되고 있음.

인터넷 포탈 회사들이(구글, 바이두, 페이스북) 가장 앞서서 개발을 하고 있으며 GPU를 사용한 다양한 딥러닝 연구들과 성과들이 적용되고 있음.



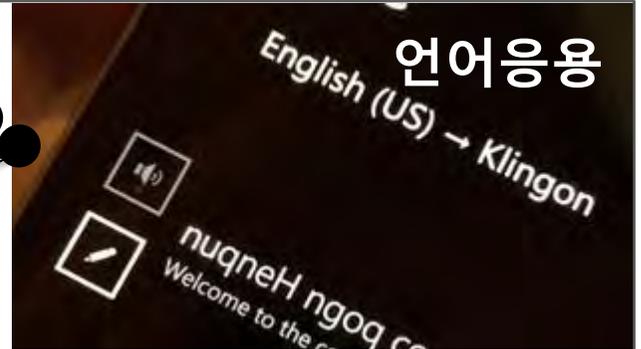
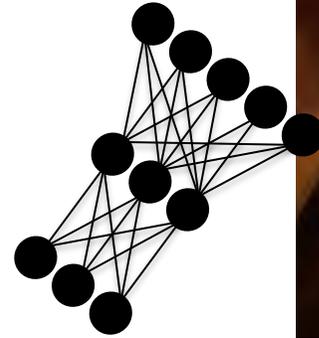
# 딥러닝이란 무엇인가?

## 적용 가능한 딥러닝 예제들

### 이미지분야



이미지 분류, 물체 감지, 위치와 움직임 인식, 장면 이해



음식인식, 음성번역, 자연어처리

### 자동차응용



보행자 감지, 교통신호 및 표지판 인식



유방암 조직 체세포 분열 감지,  
체적의 뇌 이미지 분할

# 딥러닝이란 무엇인가?

인공지능을 구성하기 위한 인공신경망(ANN, Artificial Neural Networks)에 기반하여 컴퓨터에게 사람의 사고 방식을 가르치는 방법.

사람이 가르치지 않아도 컴퓨터가 스스로 사람처럼 학습할 수 있는 인공지능 기술.



# Agenda

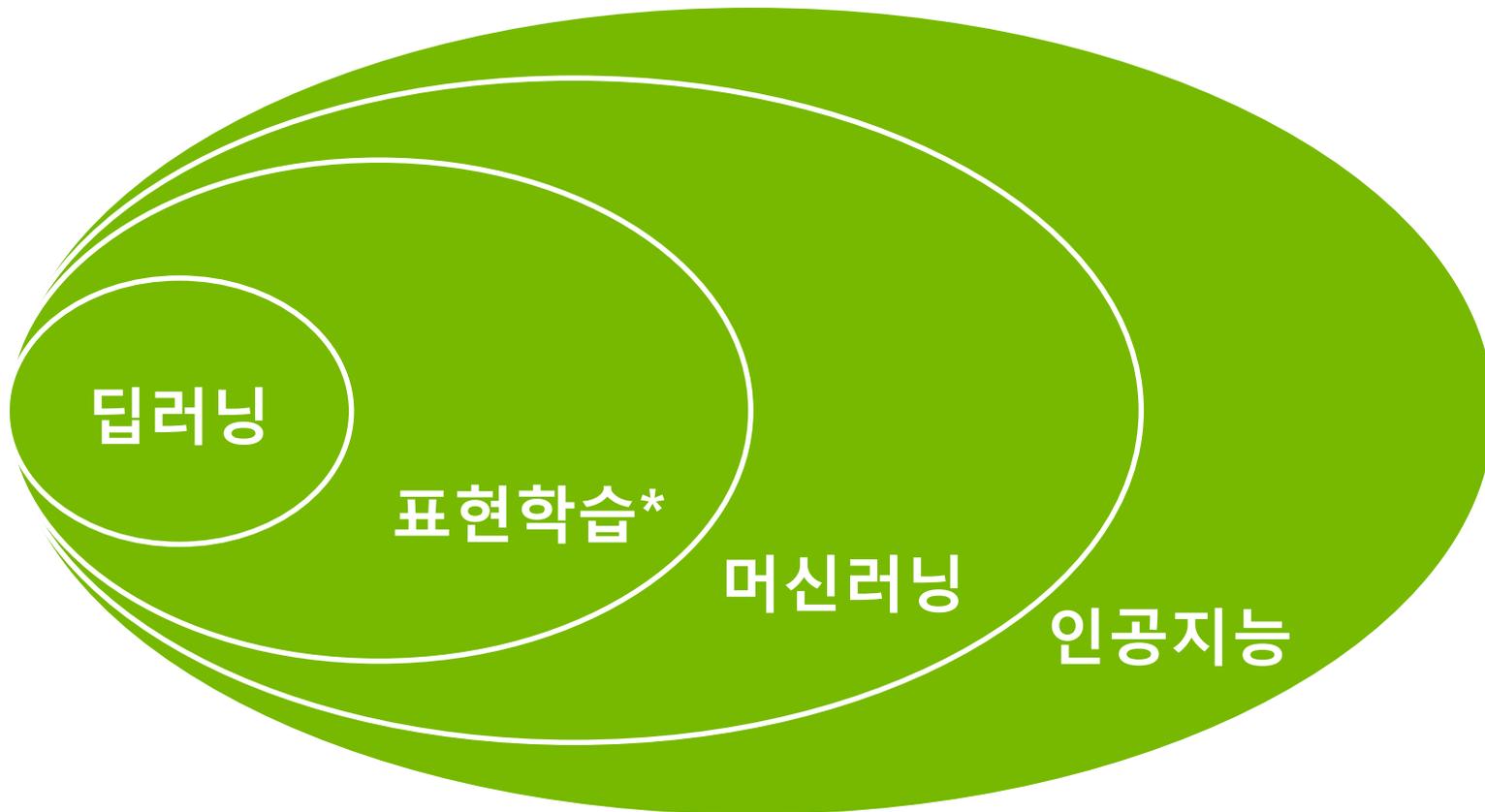
1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 인공지능이란 무엇인가?

지능적 기계 특히 지능적 컴퓨터 프로그램을 만드는 과학과 공학이다. 컴퓨터 (Computer) 를 사용해서 인간 지능 (Intelligence) 을 이해하는 작업들과 관련되어있으며 생물학적인 방법에만 국한되지는 않는다.



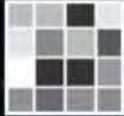
# 인공지능이란 무엇인가?



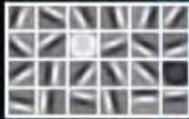
\*표현학습 : 기계가 데이터로부터 유용한 정보를 추출하는 방법을 연구하는 분야

## 인공지능의 얼굴 인식 과정

딥러닝 신경망은 얼굴과 같은 복잡한 형태를 단순한 형태에서 시작해 점점 복잡한 형태로 올라가며 그분화하면서 구분한다



**제1층** 컴퓨터가 밝고 어두운 픽셀을 구분한다



**제2층** 컴퓨터가 테두리와 단순한 형태를 구분한다



**제3층** 컴퓨터가 좀더 복잡한 형태와 사물을 구분한다



**제4층** 컴퓨터가 인간의 얼굴을 규정하는 형태를 파악한다

자료: Nature

## 인공신경망

뇌 신경망의 뉴런(신경세포)에서 신호가 전달되는 과정을 모방해 수식으로 표현한 것이 인공신경망이다



각각의 뉴런은 인공신경망에서 숫자로 표시된 데이터를 분류하는 수식으로 표현된다



뉴런과 뉴런을 잇는 연결지점으로 신경 신호가 전달되는 시냅스는 인공신경망에서 원으로 표시된 수식과 수식을 연결하는 선이 된다



한 수식에서 특정 조건에 맞는 정보를 추려내고, 그 결과가 다시 다음 단계 수식의 기본 자료가 된다. 이처럼 수식들이 서로 다양하게 연결되면서 작은 변화들을 모아 원하는 조건에 맞는 결과를 찾아낸다

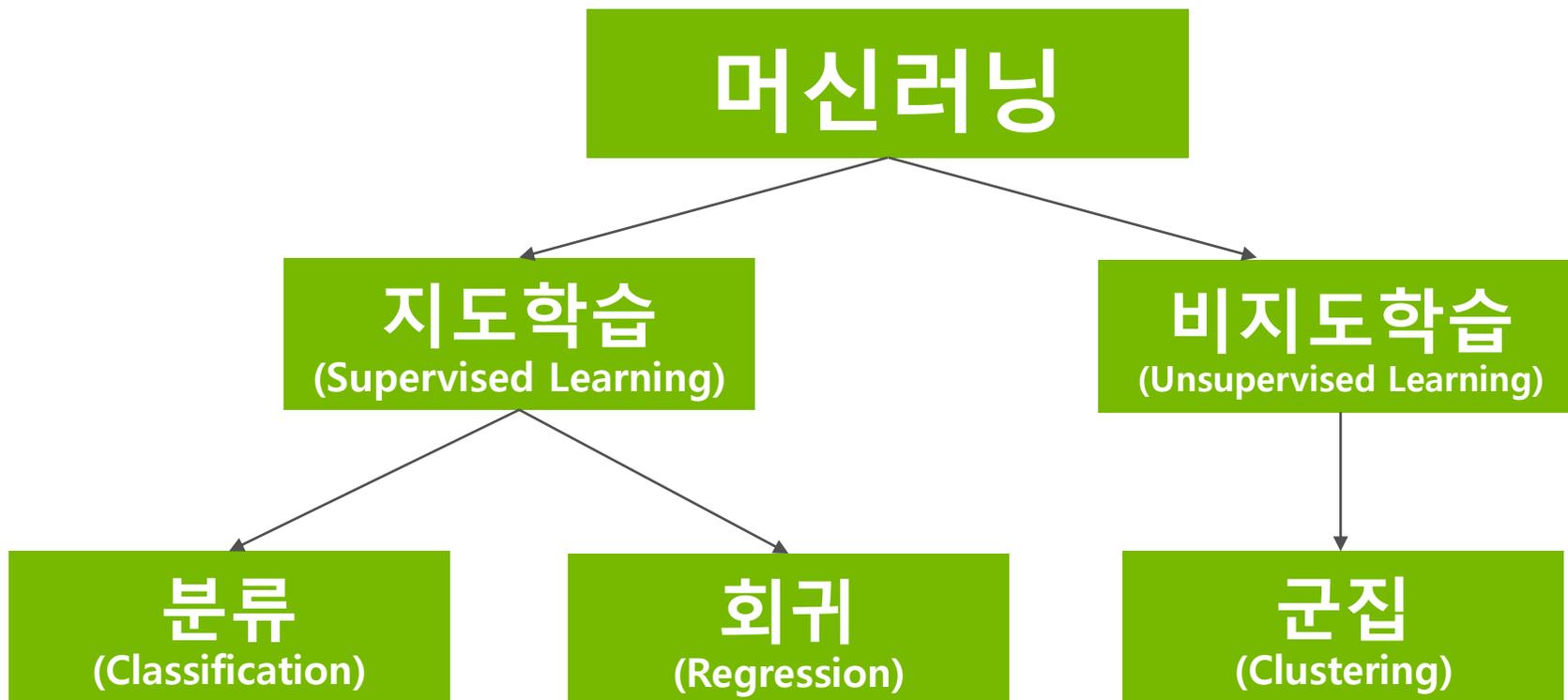
# 머신러닝이란 무엇인가?

컴퓨터에게 사람이 직접 로직(Logic)을 지시하지 않아도 데이터를 통해 컴퓨터가 '학습'을 하고 그것을 사용해 컴퓨터가 자동으로 문제를 해결하도록하는 것을 의미한다.

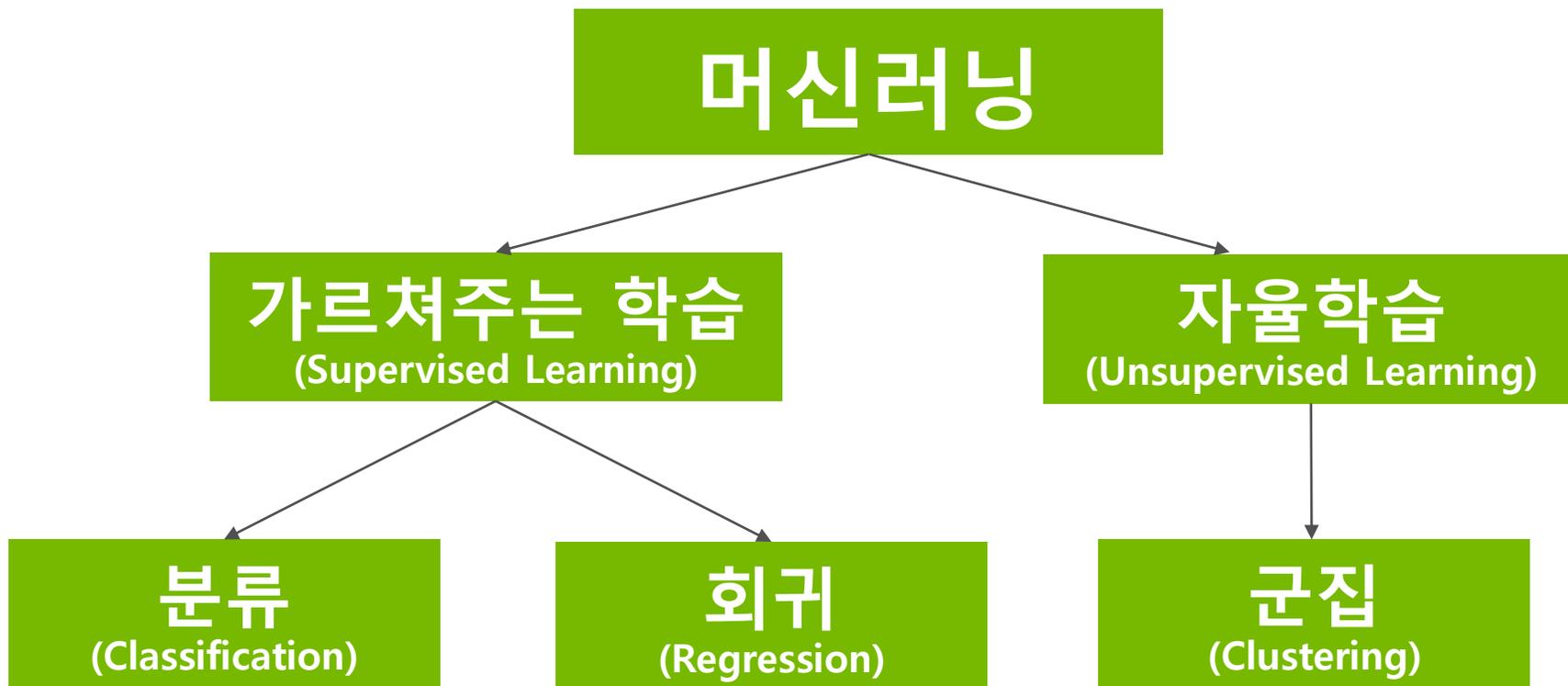
머신러닝은 기계학습으로 사용됨.



# 머신러닝이란 무엇인가?



# 머신러닝이란 무엇인가?



꽃

타블렛

레몬과 얼음물이  
들어있는 컵

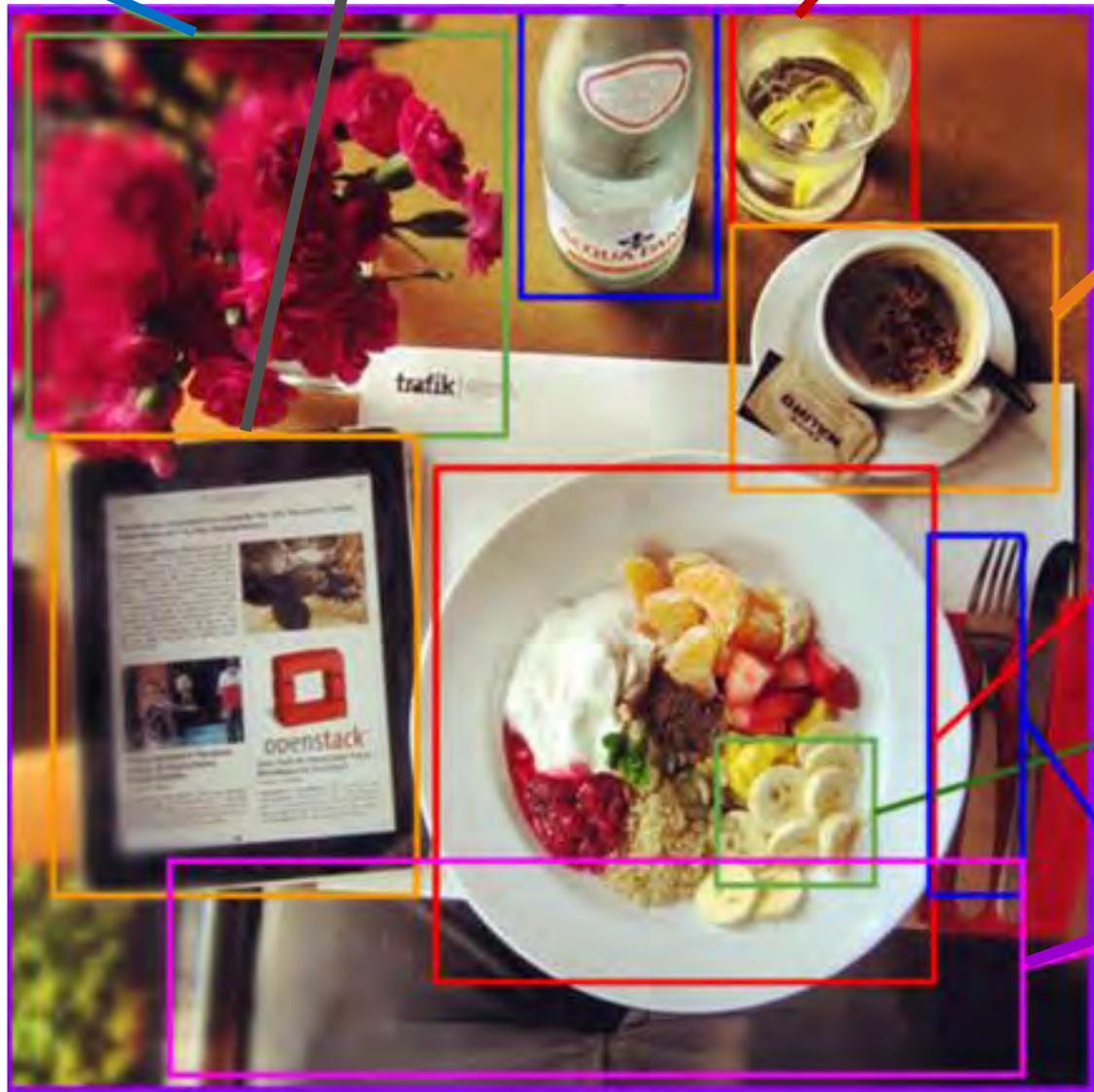
커피 컵

아침식사가  
차려진 식탁

어떤 사람이  
식탁에 앉아  
있음

지도 학습  
(Supervised  
learning)

비지도 학습  
(Unsupervised  
learning)



# 머신러닝이란 무엇인가?

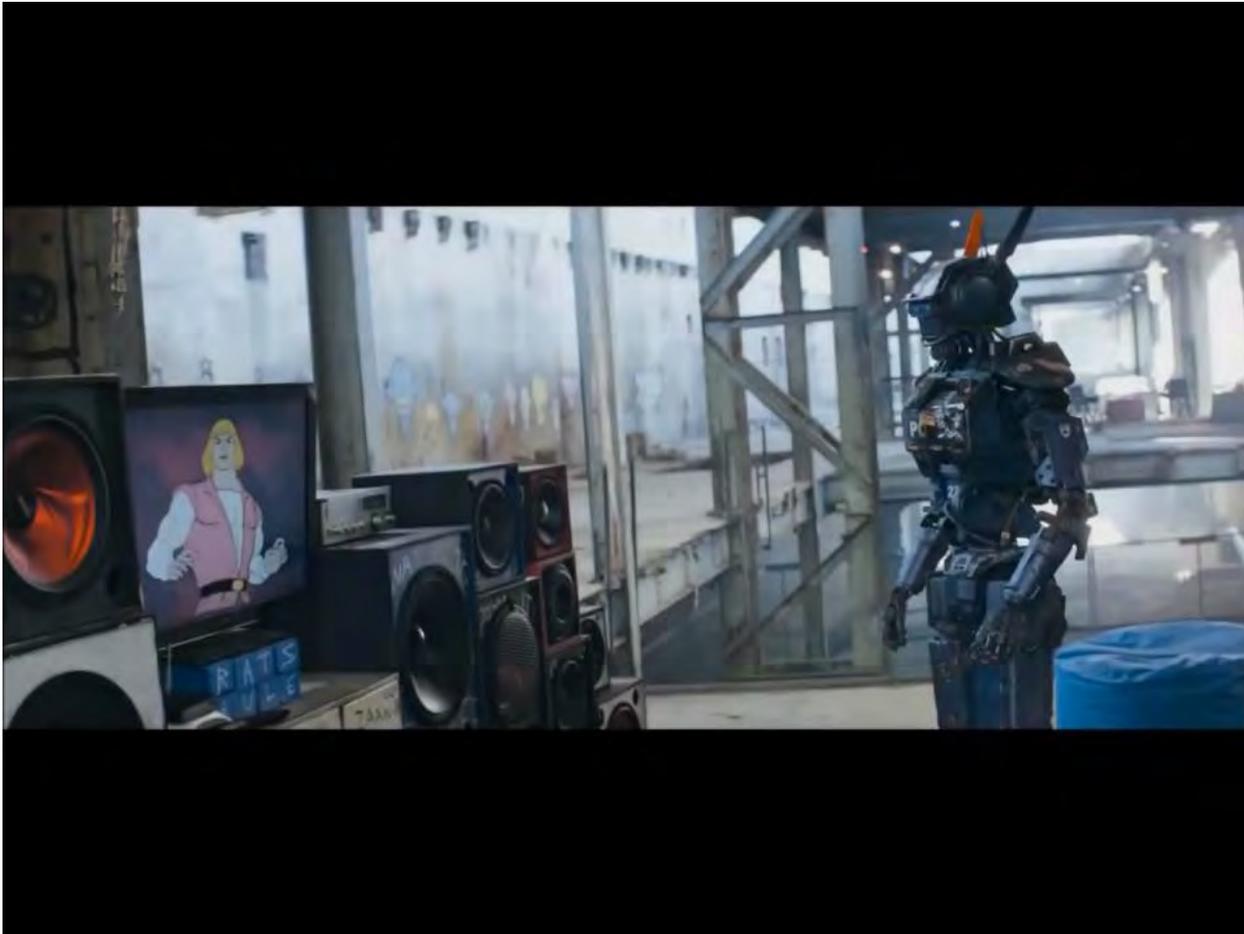
가르쳐주는 학습

(Supervised Learning)



# 머신러닝이란 무엇인가?

자율학습(Unsupervised Learning)



# 머신러닝이란 무엇인가?

자율학습(Unsupervised Learning)





# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 딥러닝은 왜 필요한가?

## 기존 머신러닝 관점

원본데이터

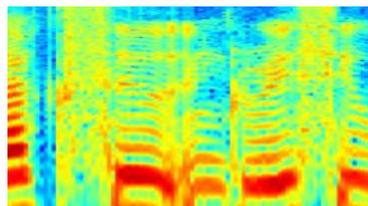
특징 추출

분류기

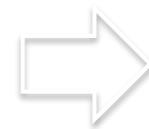
결과



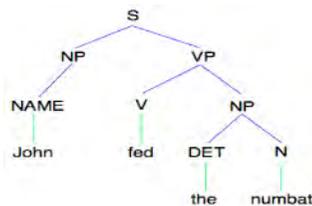
SVM,  
shallow neural net,  
...



HMM,  
shallow neural net,  
...



말하는 사람의 신분,  
음성을 문자로...



Clustering, HMM,  
LDA, LSA  
...



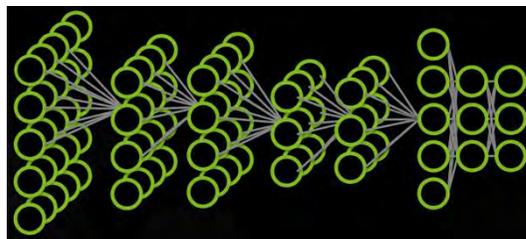
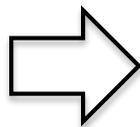
준제분류,  
번역,  
내포된 의미 분석...

# 딥러닝은 왜 필요한가?

딥러닝

훈련

강아지



강아지(O)

# 딥러닝은 왜 필요한가?

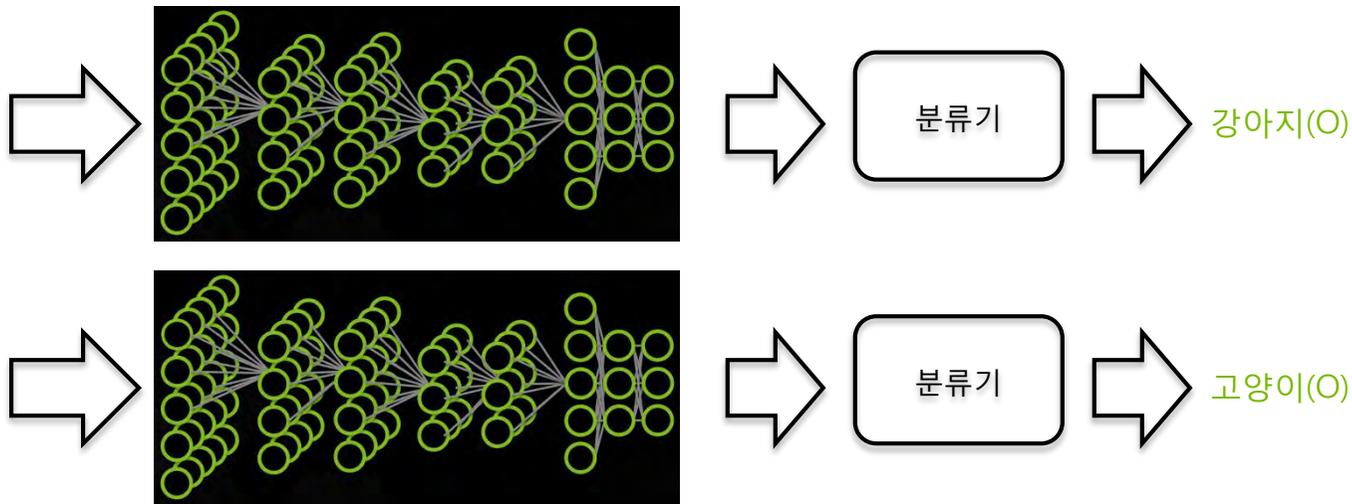
## 딥러닝

훈련

강아지



고양이

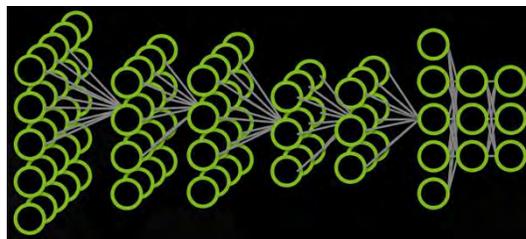
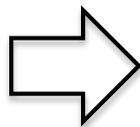


# 딥러닝은 왜 필요한가?

## 딥러닝

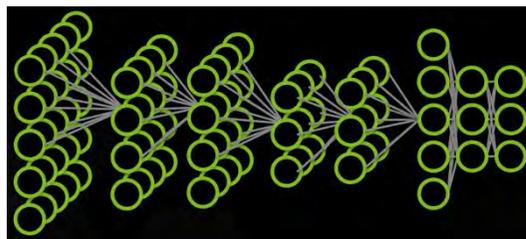
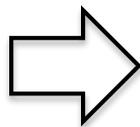
### 훈련

강아지



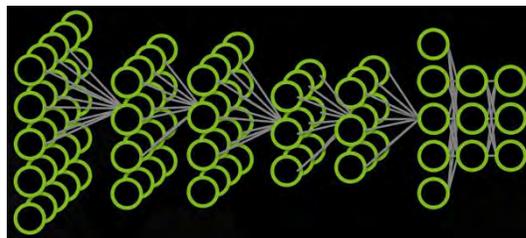
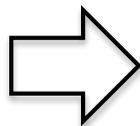
강아지(O)

고양이



고양이(O)

고래



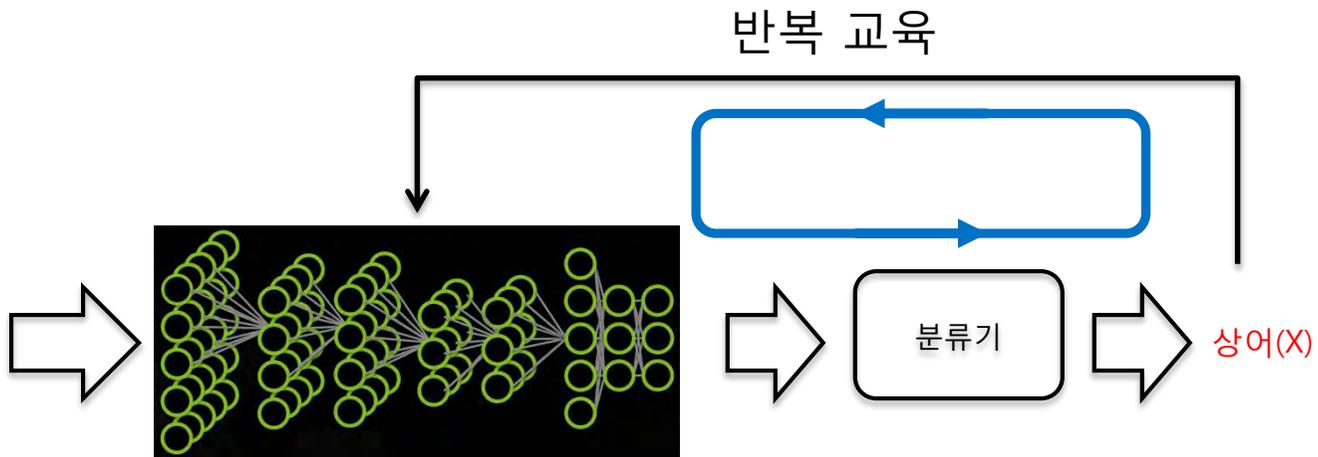
상어(X) !!!!!

# 딥러닝은 왜 필요한가?

딥러닝

후련

고래

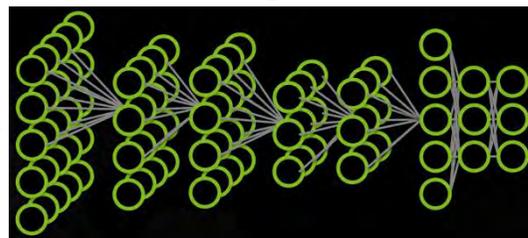
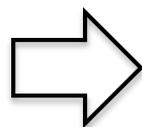


# 딥러닝은 왜 필요한가?

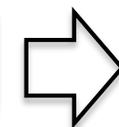
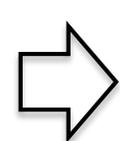
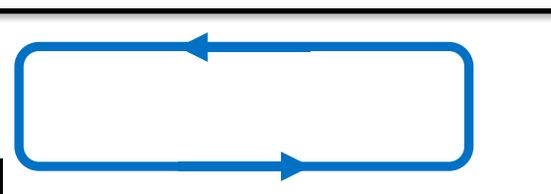
## 딥러닝

훈련

고래

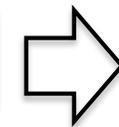
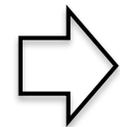
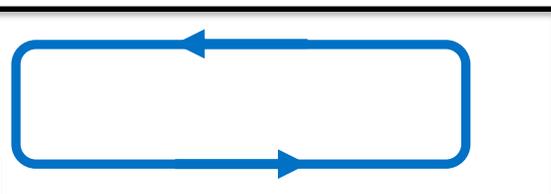


반복 교육



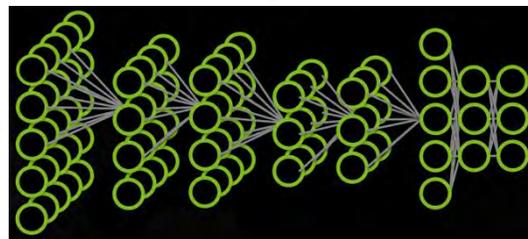
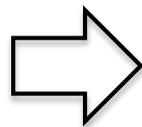
상어(X)

반복 교육



고래(O)

고래

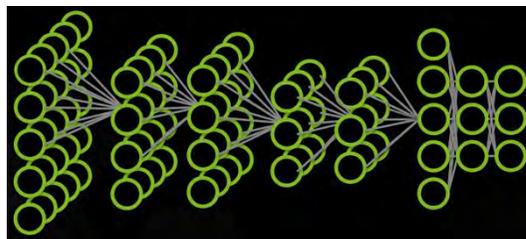
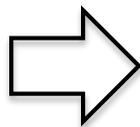


# 딥러닝은 왜 필요한가?

## 딥러닝

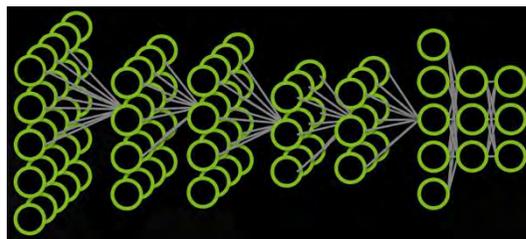
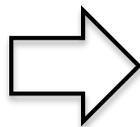
훈련 완료

강아지



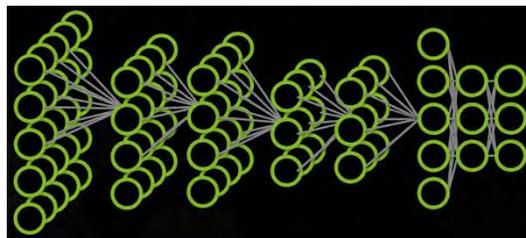
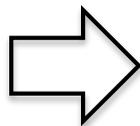
강아지(O)

고양이



고양이(O)

고래



고래(O)

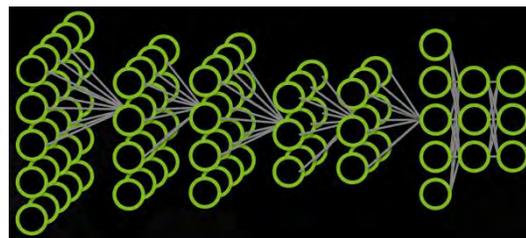
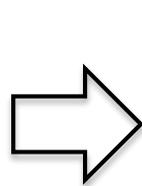
# 딥러닝은 왜 필요한가?

딥러닝

테스트



임의의 사진 입력



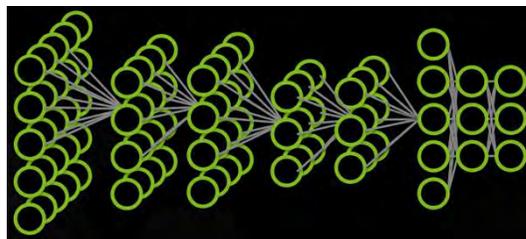
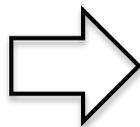
강아지(O)

# 딥러닝은 왜 필요한가?

## 딥러닝

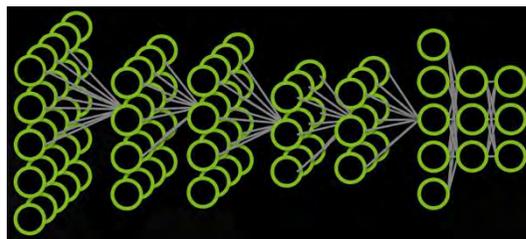
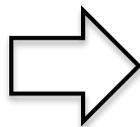
### 훈련

아이유



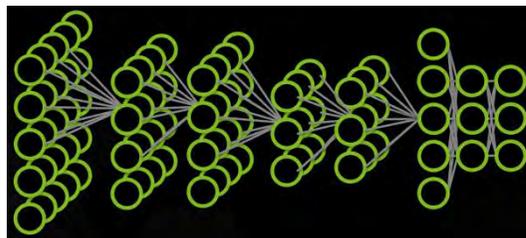
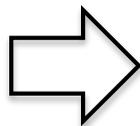
아이유(O)

혜리



혜리(O)

황정음



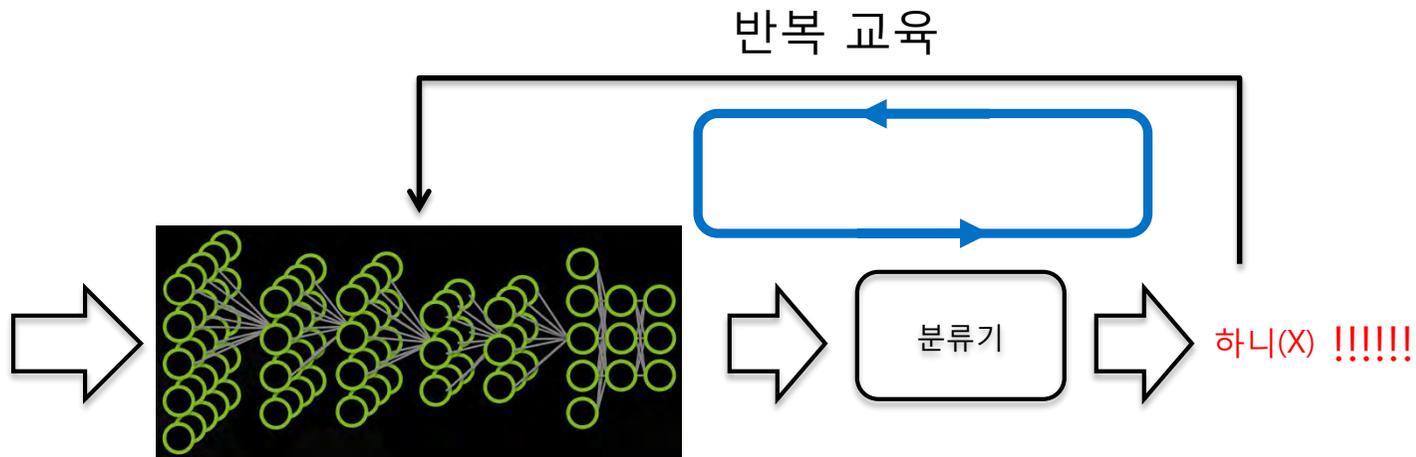
하니(X) !!!!!

# 딥러닝은 왜 필요한가?

딥러닝

훈련

황정음

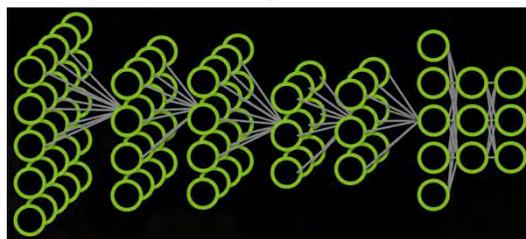
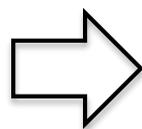


# 딥러닝은 왜 필요한가?

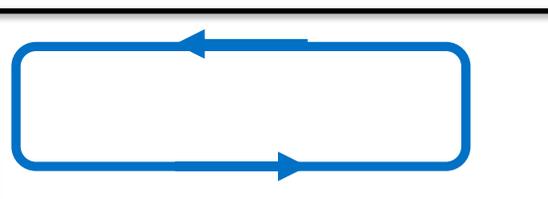
## 딥러닝

훈련

황정음

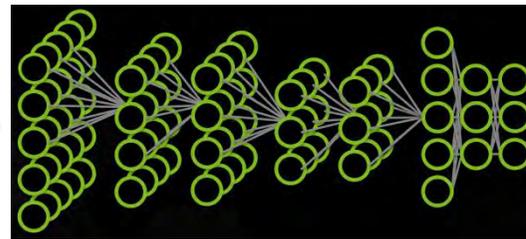
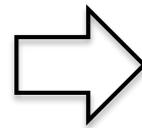


반복 교육

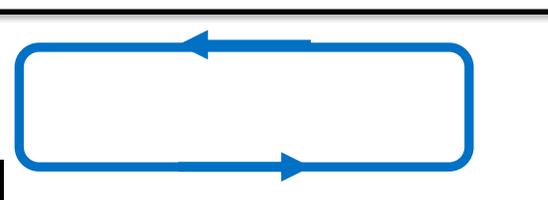


하니(X) !!!!!

황정음



반복 교육

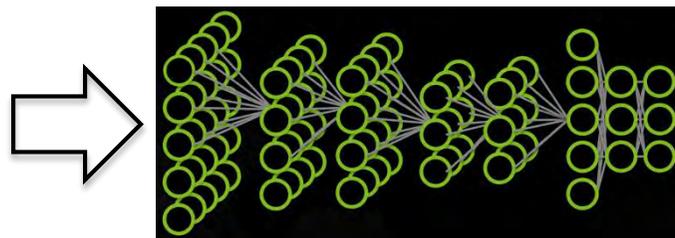


황정음(O)

# 딥러닝은 왜 필요한가?

딥러닝

테스트 임의의 사진 입력



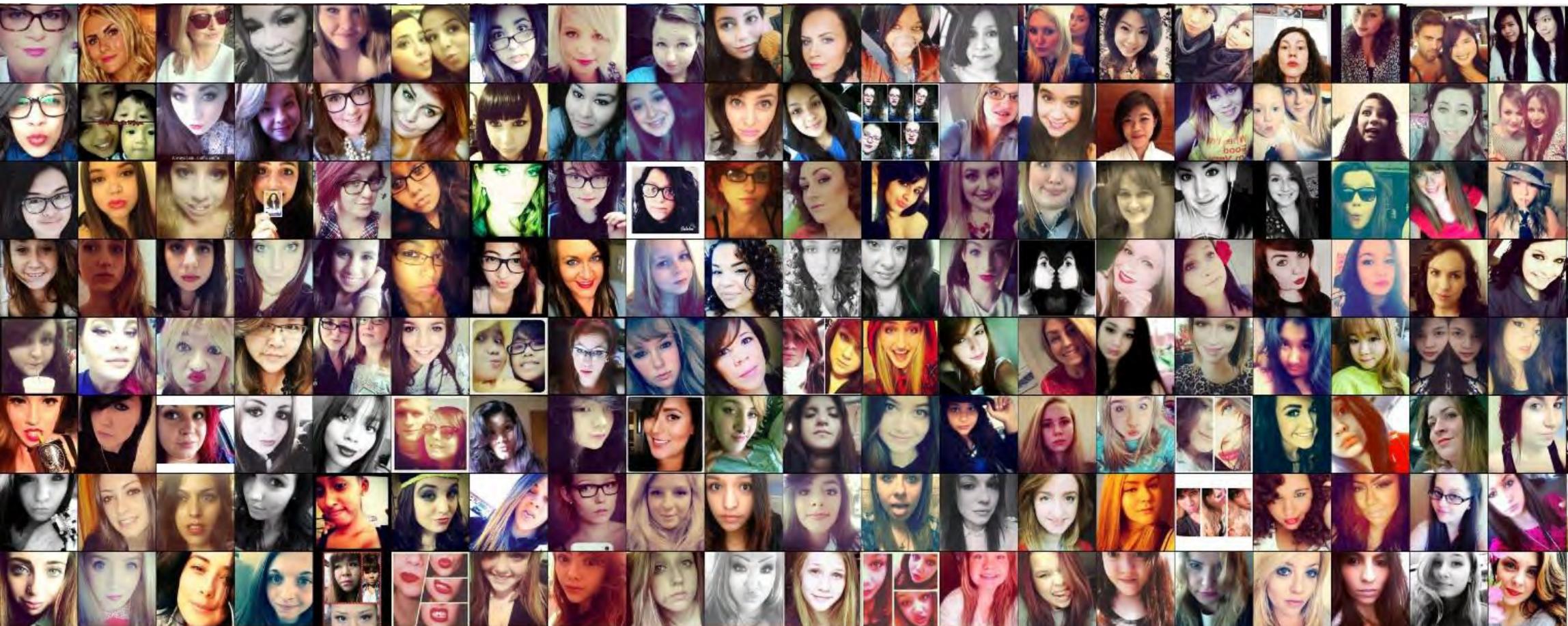
아이유(O)

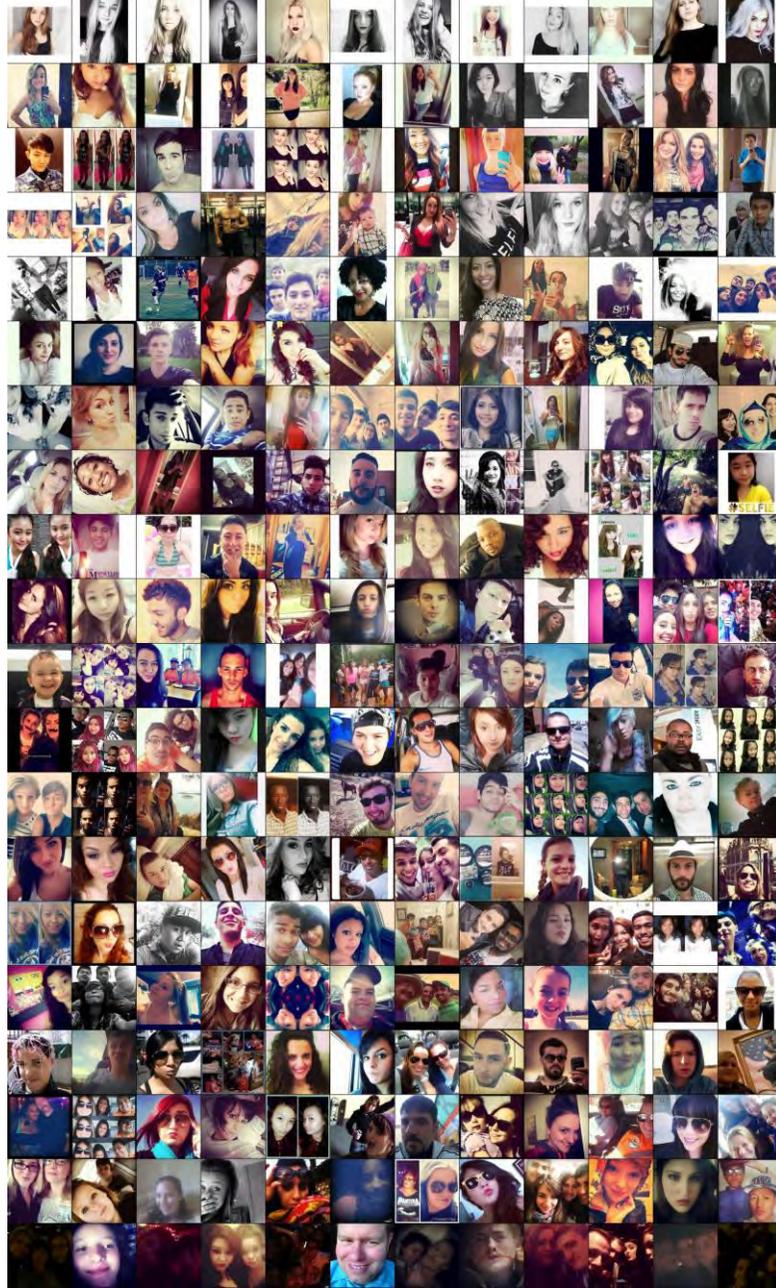
황정음(O)

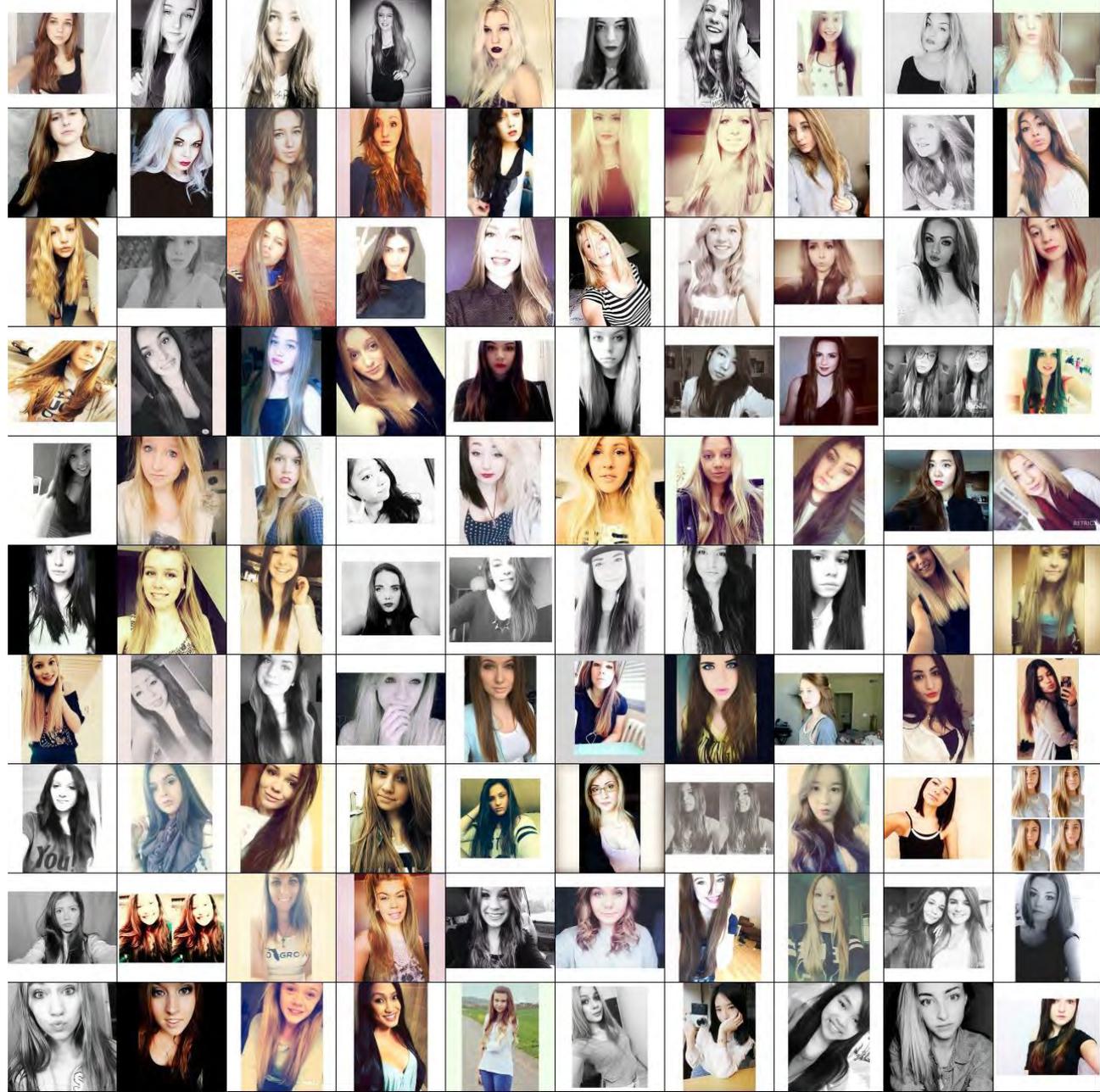
헤리(O)

# 딥러닝은 왜 필요한가?

## 딥러닝

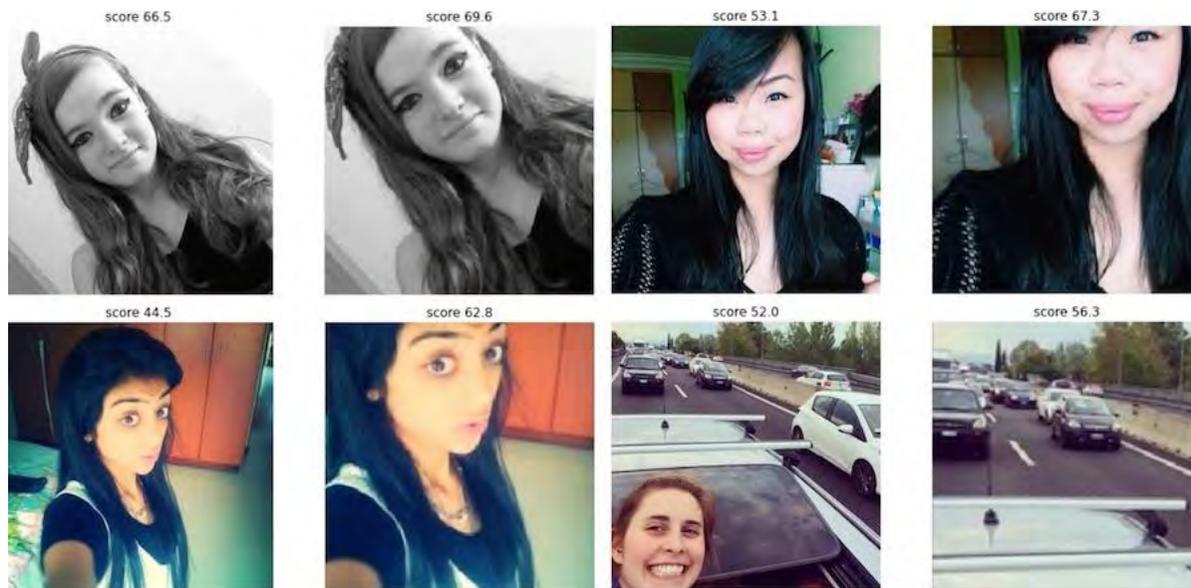






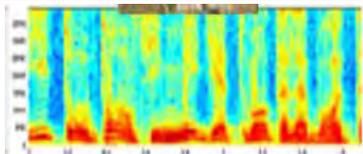
# 딥러닝은 왜 필요한가?

## 딥러닝



# 딥러닝은 왜 필요한가?

## 딥러닝이 사용되는 사례

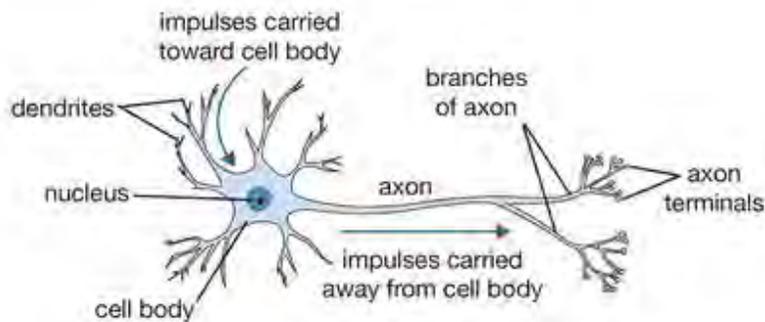
	입력	출력
사진		강아지
음성		안녕하세요 환영합니다
문자	"안녕하세요. 어떻게 지내세요?"	"Bonjour, comment allez-vous?"
사진		조그만 아이가 인형을 안고 자고 있어요

# 딥러닝은 왜 필요한가?

신경망(ANN)\*

시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반

## 생물학적 뉴런

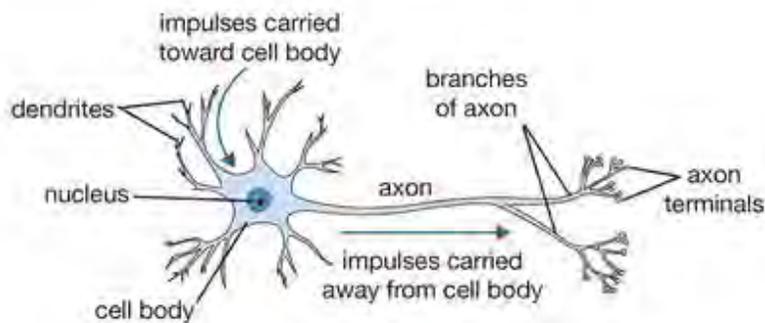


# 딥러닝은 왜 필요한가?

신경망(ANN)\*

시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반

생물학적 뉴런



$$Y = ax + b$$

\*\*w : 가중치(weight)

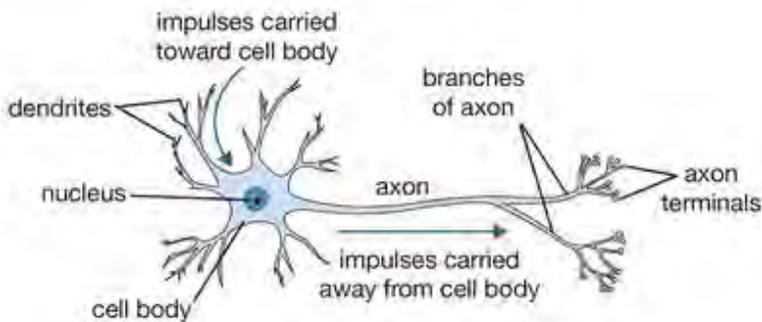
\*ANN : Artificial Neural Network

# 딥러닝은 왜 필요한가?

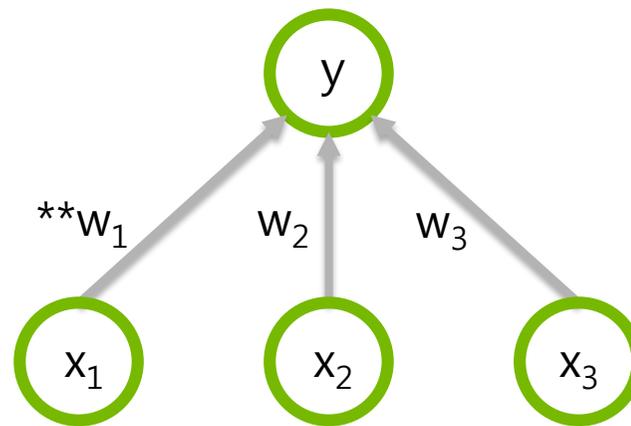
## 신경망(ANN)\*

시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반

생물학적 뉴런



인공 신경



$$y = F(w_1x_1 + w_2x_2 + w_3x_3)$$

$$F(x) = \max(0, x)$$

\*\*w : 가중치(weight)

\*ANN : Artificial Neural Network

# 딥러닝은 왜 필요한가?

신경망(ANN)\*



충분히 훈련이 된 인공신경망은 내부의 복잡한 기능들을 통해서 출력을 예측할 수 있음.

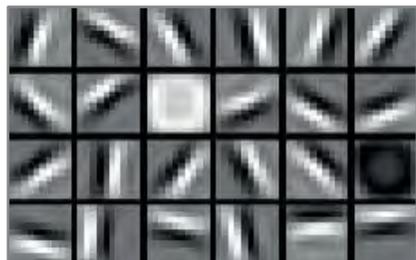
# 딥러닝은 왜 필요한가?

심층신경망(DNN)\*

원본 데이터



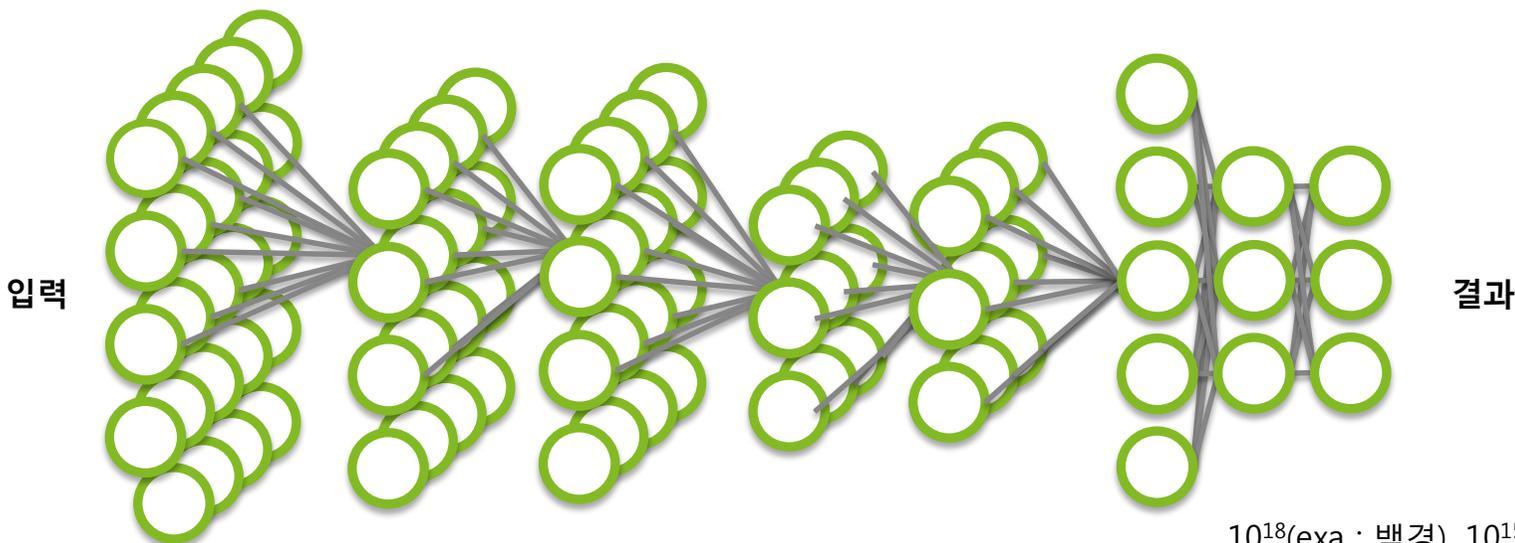
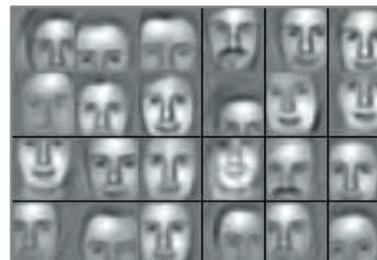
하위레벨 특징들



중간레벨 특징들



상위레벨 특징들



어플리케이션 구성:

목적 : Identify face

훈련데이터  
10-100M images

망 아키텍처  
~10 층  
1B 매개변수

훈련 알고리즘  
~30 Exaflops  
~30일 GPU

\*DNN : Deep Neural Network

$10^{18}$ (exa : 백경),  $10^{15}$ (peta : 천조),  $10^{12}$ (tera : 조),  $10^9$ (giga : 십억)

# 딥러닝은 왜 필요한가?

## 특징 표현 방법

과거:

입력

사진, 문자, 음성, 기타



수작업을 통한  
특징 추출



분류 / 통합 / 변화

연산자



출력  
결과 값

# 딥러닝은 왜 필요한가?

## 특징 표현 방법

과거:

입력 →  
사진, 문자, 음성, 기타

수작업을 통한  
특징 추출

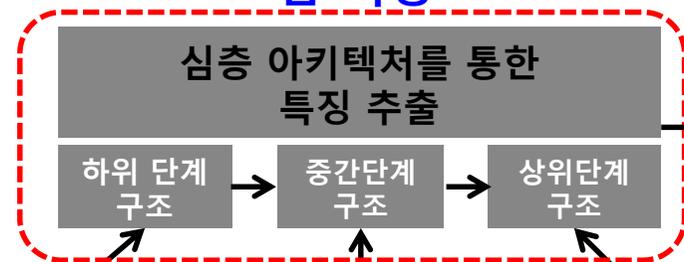
분류 / 통합 / 변화

연산자

출력  
결과 값

현재:

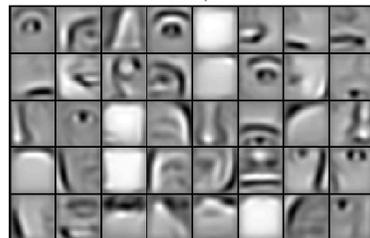
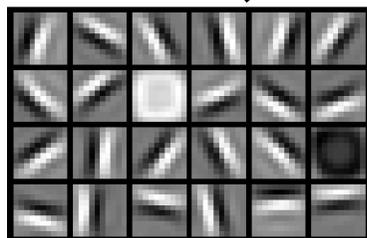
입력 →  
사진, 문자, 음성, 기타



분류 / 통합 / 변화

연산자

출력  
결과 값



### 계층형 특징 표현

- 이미지 인식  
화소 → 모서리 → 물체의 일부분 → 물체
- 문자 인식  
글자 → 단어 → 구절 → 문장
- 음성인식  
스펙트럴 대역 → 단음 → 음소 → 단어

딥러닝은 왜 필요한가?

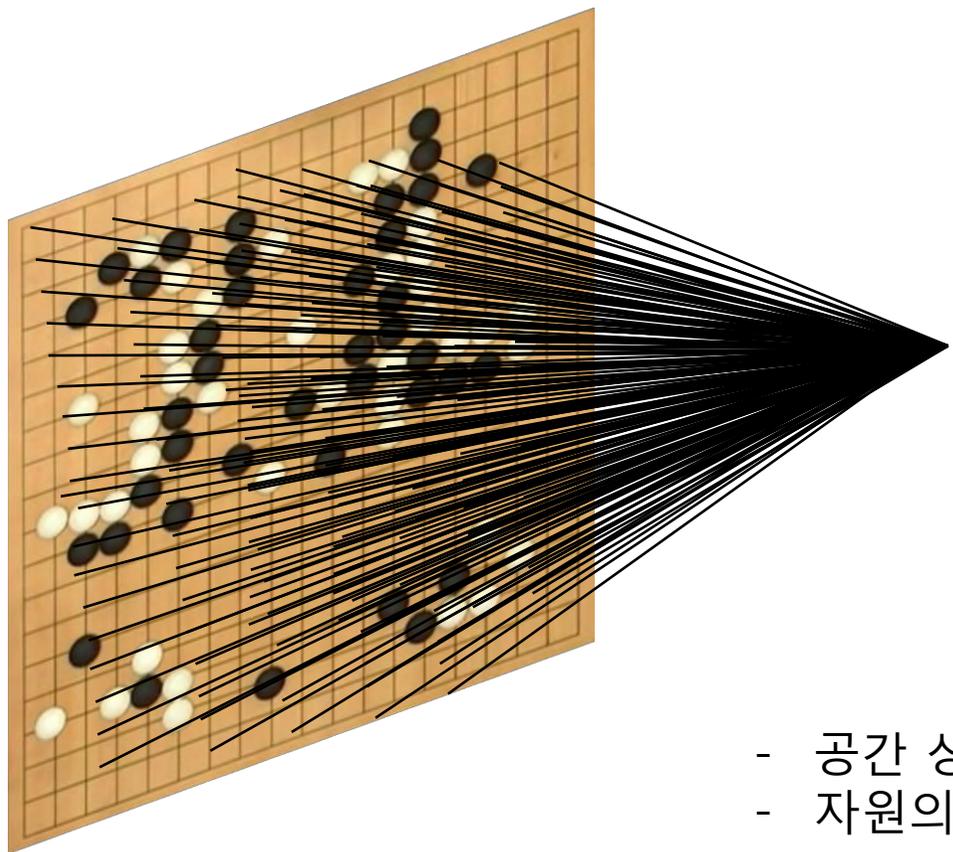
**충분히 많은 데이터를 가지고  
반복적인 학습을 통해  
올바른 결론을 도출해 내기 위해**

# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 딥러닝은 어떤 역할을 하는가?

나선구조적인신경망(CNN)\*



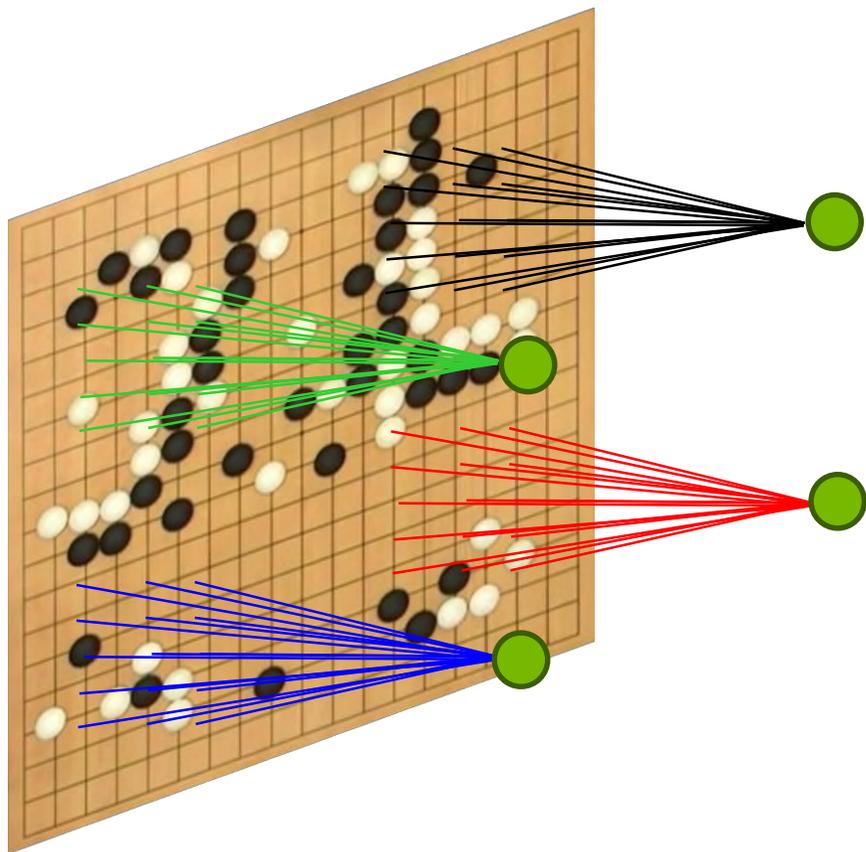
완전 연결 단계

➤ 예 : 200 x 200 이미지  
구성단위 : 40,000  
매개변수 : ~ 20억

- 공간 상관 관계
- 자원의 낭비 + 충분히 훈련이 되지 못함

# 딥러닝은 어떤 역할을 하는가?

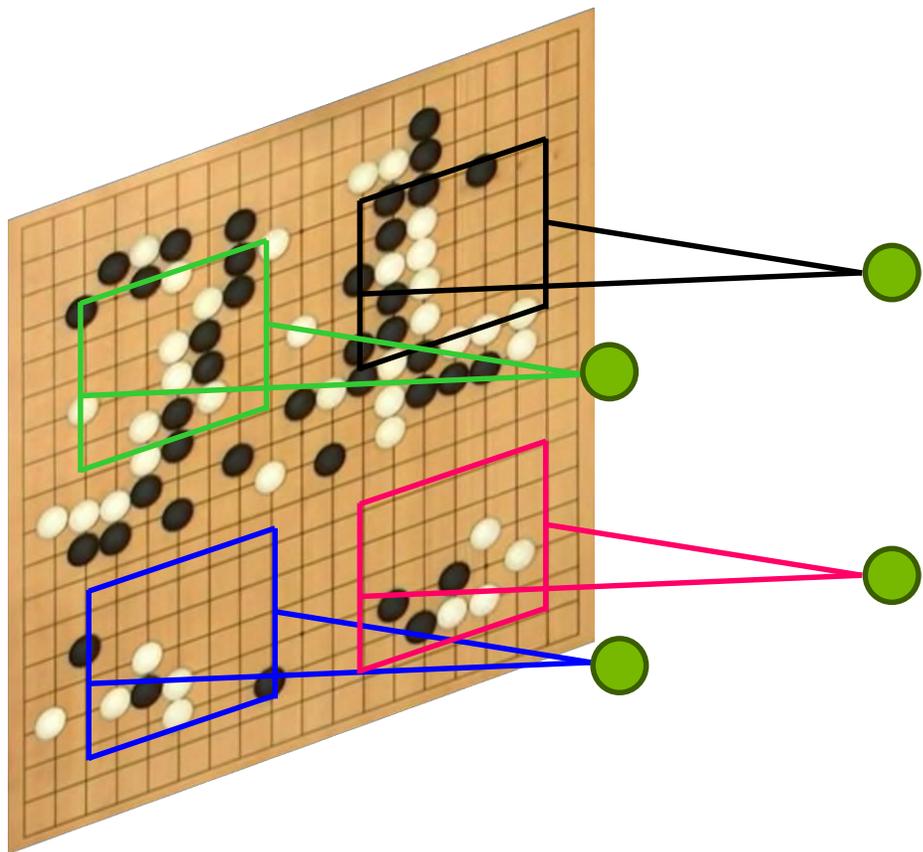
나선구조적인신경망(CNN)\*



- 예 : 200 x 200 이미지  
구성단위 : 40,000  
필터 사이즈 : 10 x 10  
매개변수 : 4M

# 딥러닝은 어떤 역할을 하는가?

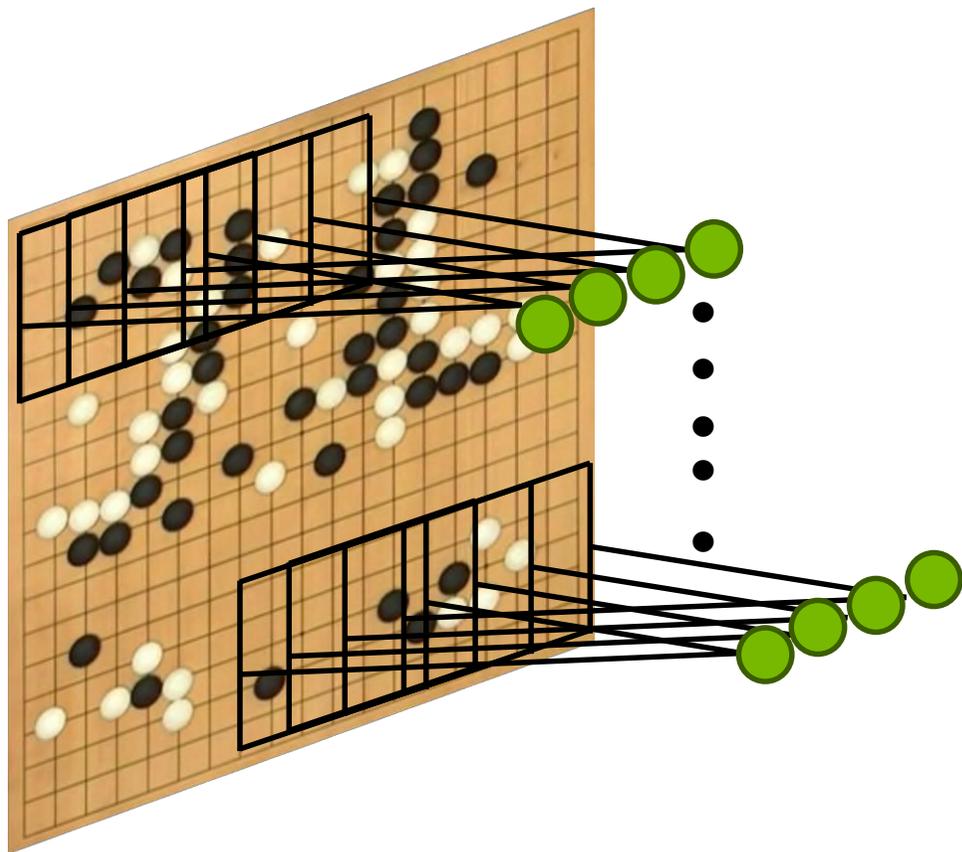
나선구조적인신경망(CNN)\*



- 예 : 200 x 200 이미지  
구성단위 : 40,000  
필터 사이즈 : 10 x 10  
매개변수 : 4M

# 딥러닝은 어떤 역할을 하는가?

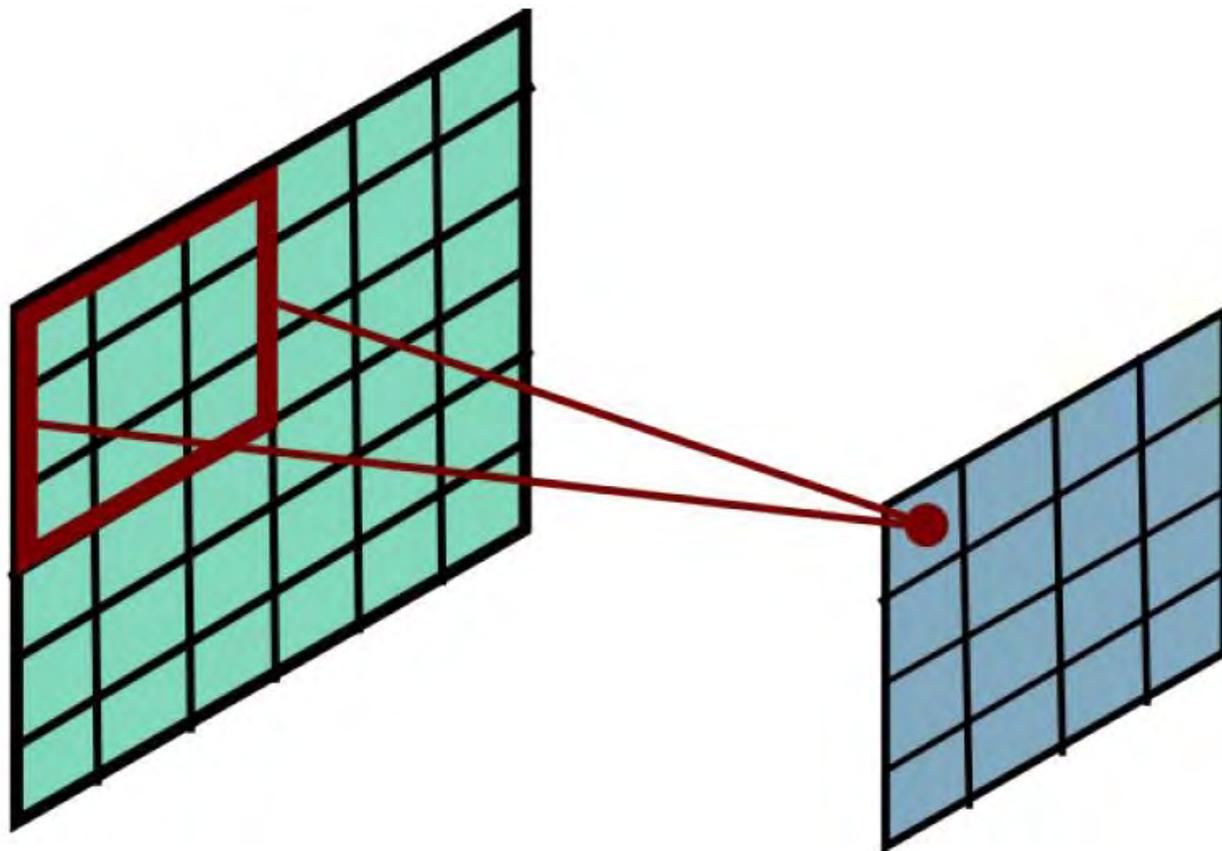
나선구조적인신경망(CNN)\*



➤ 인접영역과 매개변수를 공유

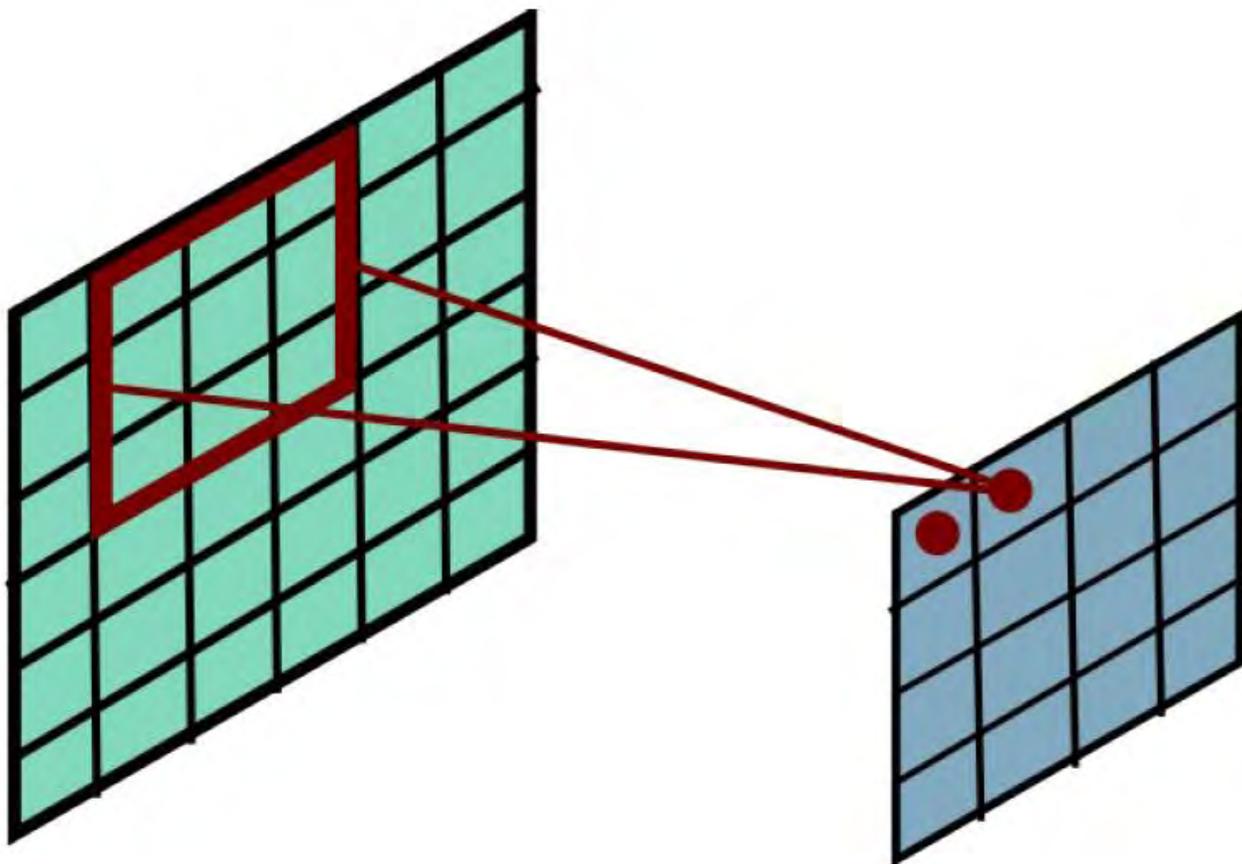
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



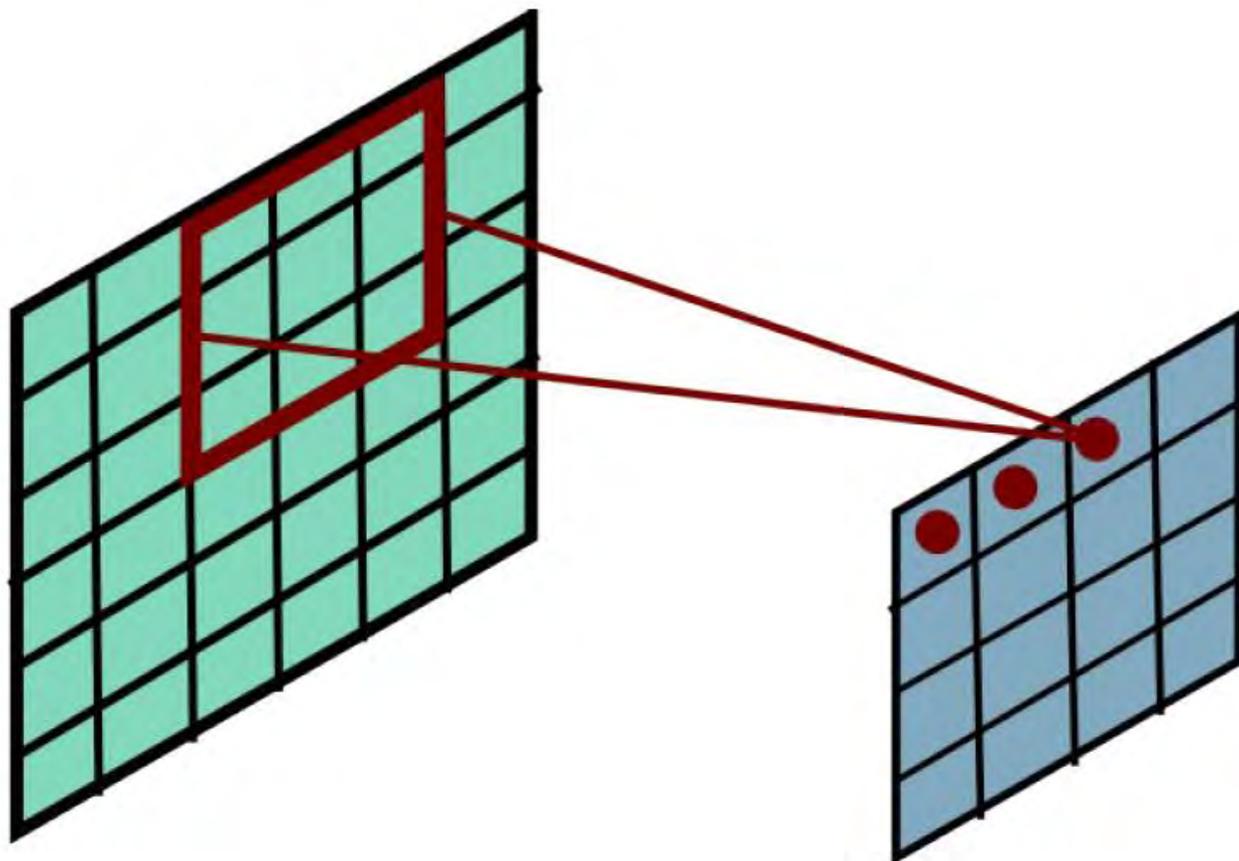
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



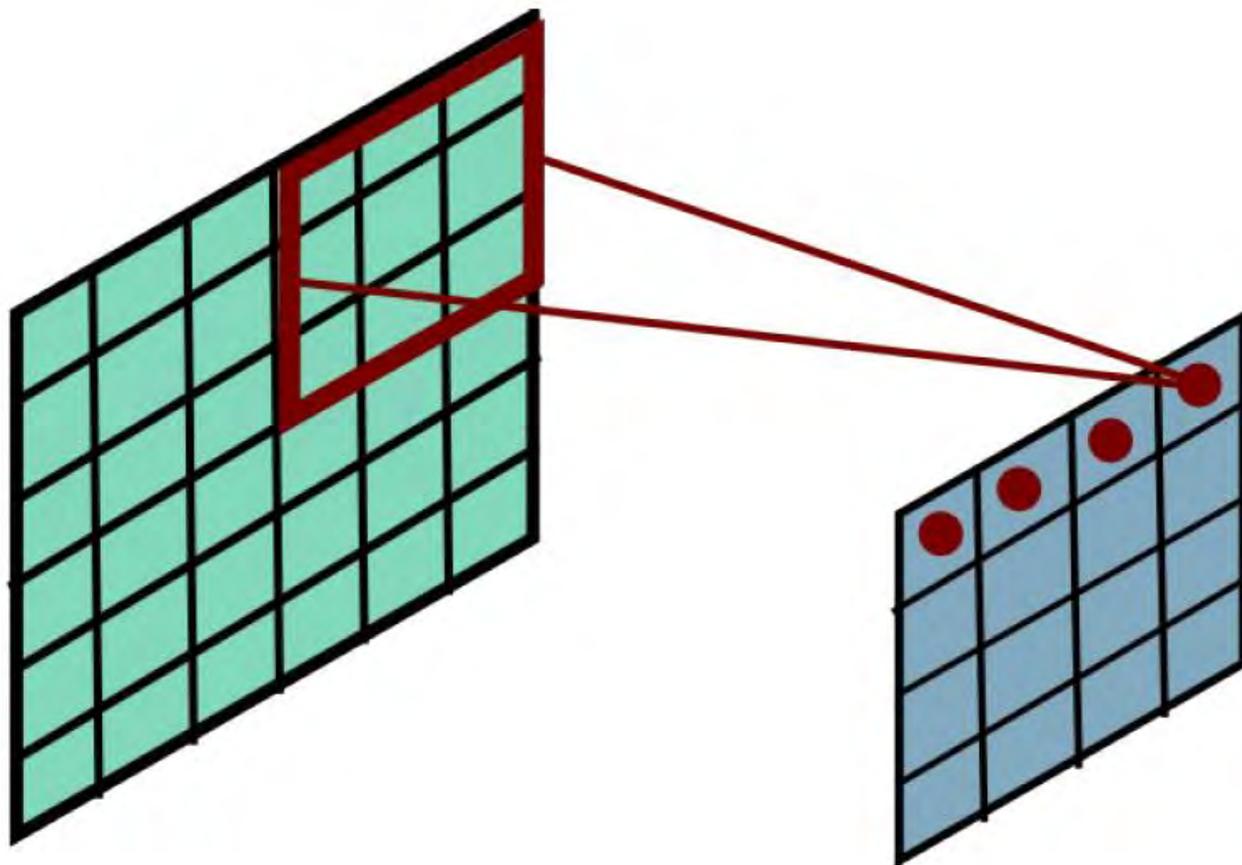
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



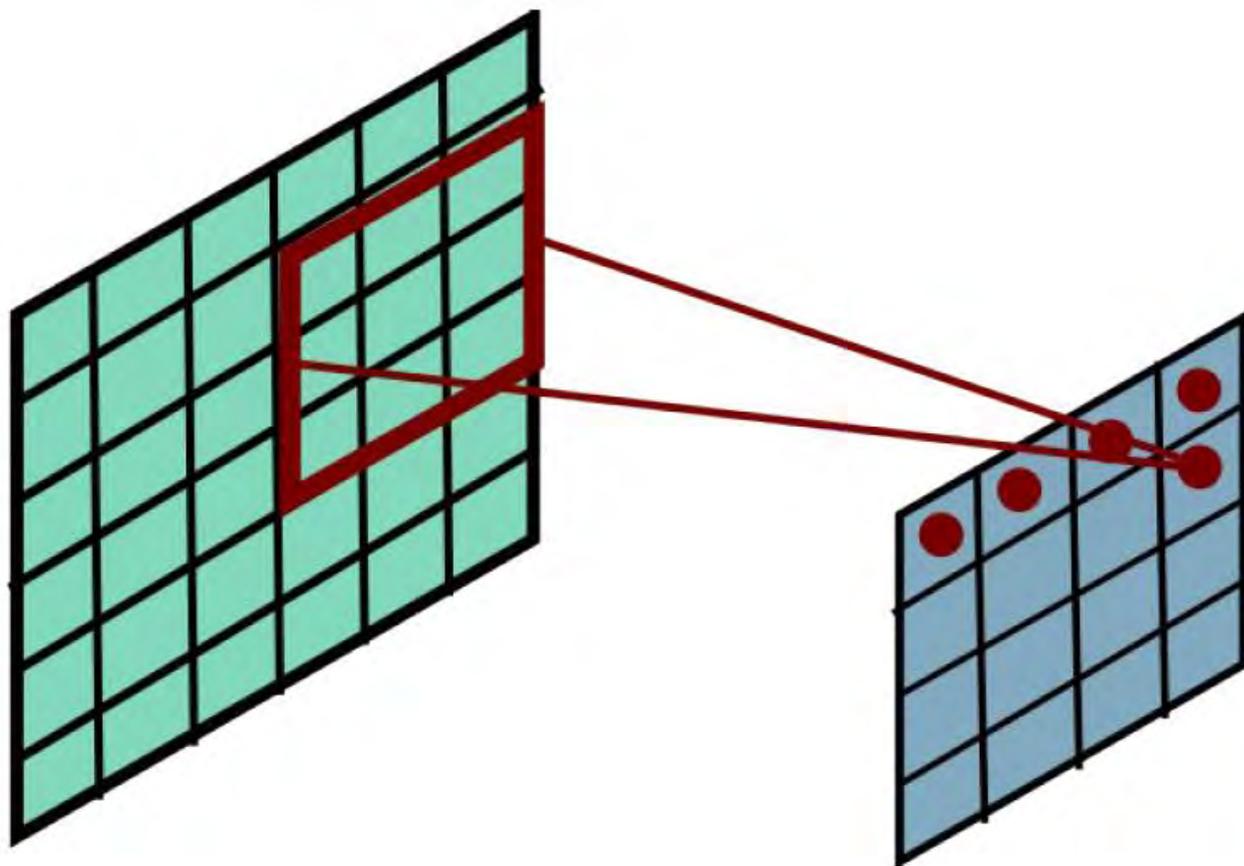
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



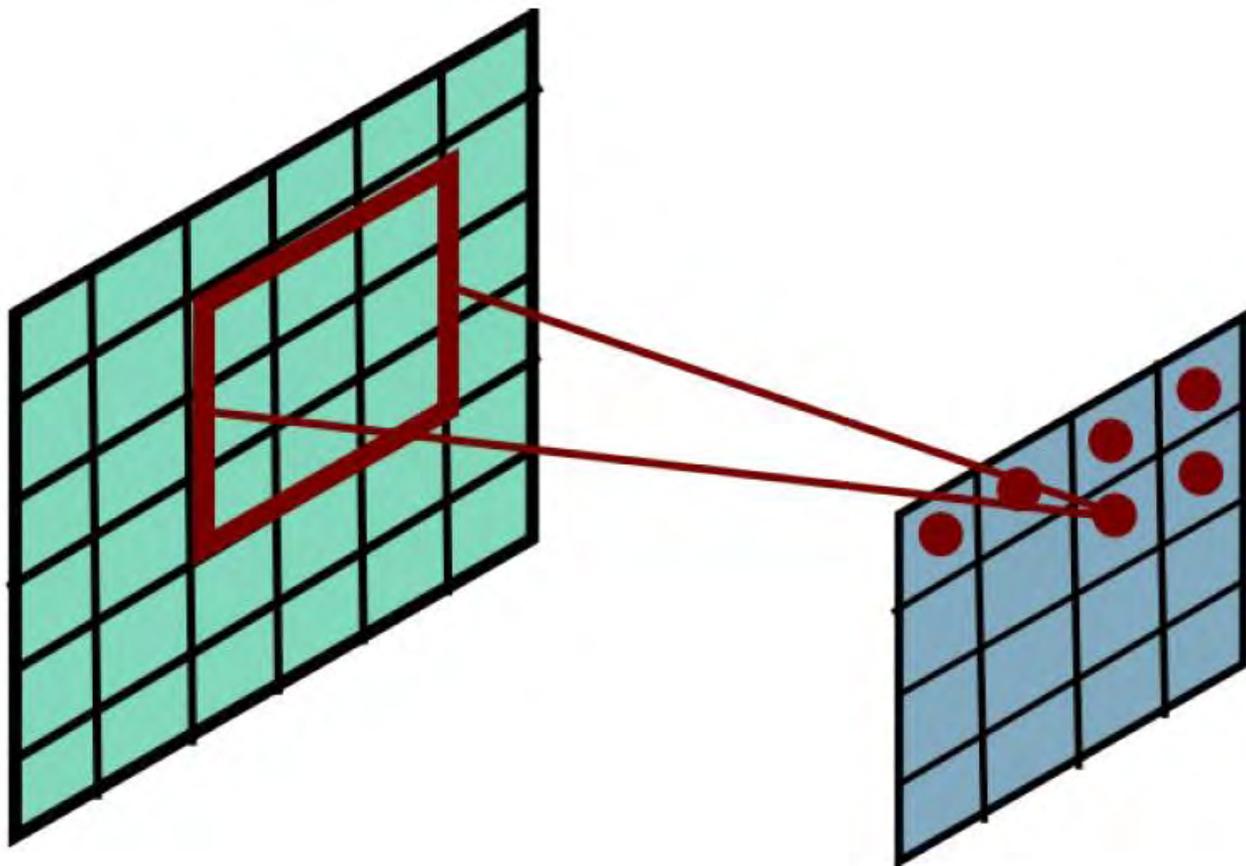
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



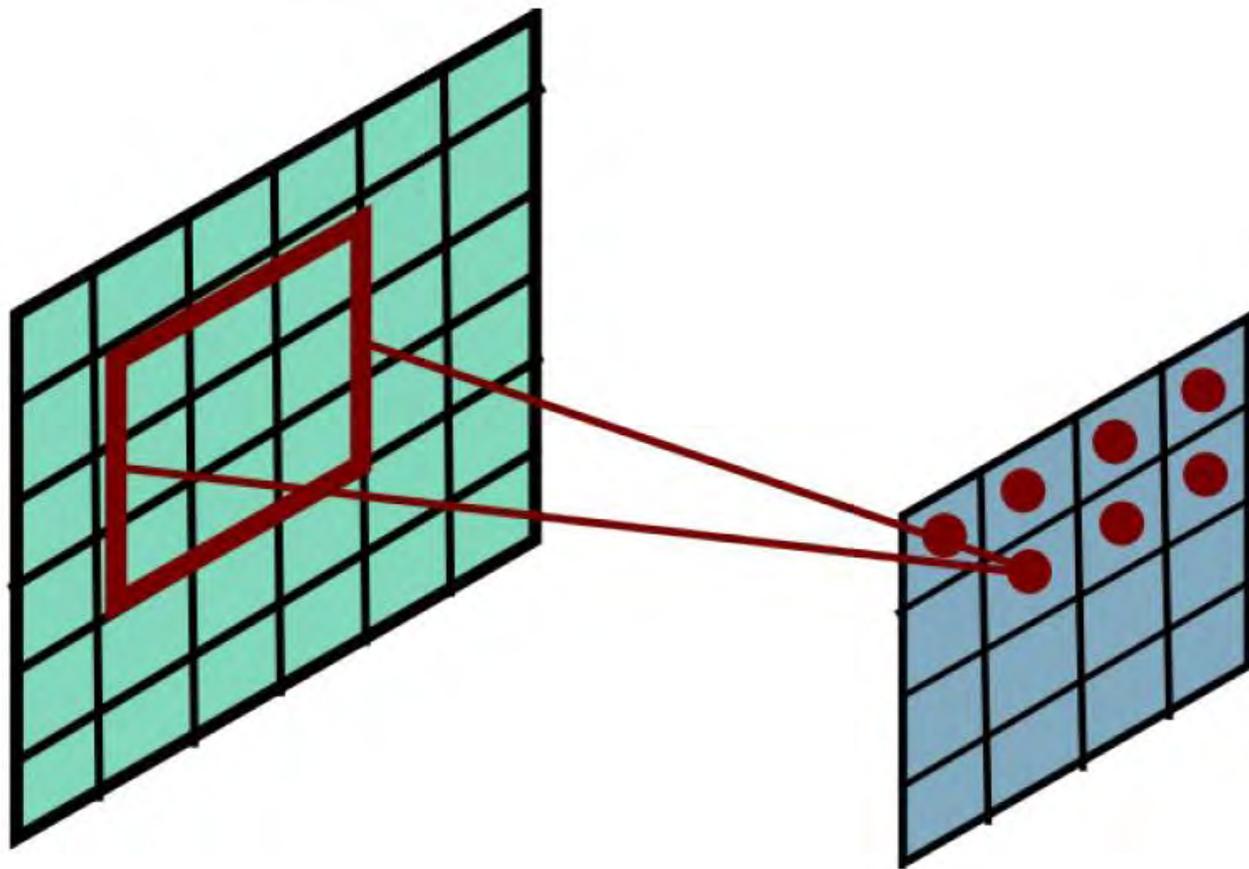
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



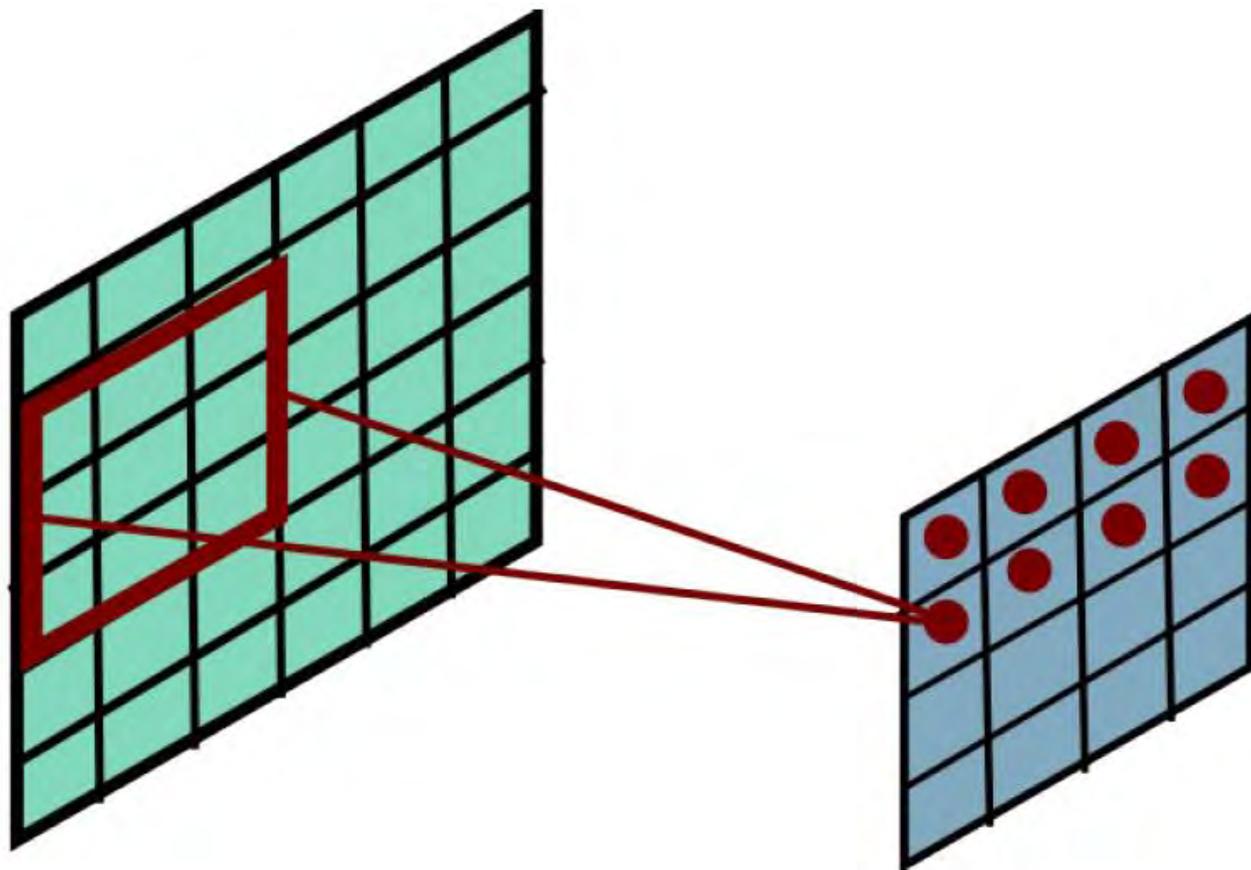
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



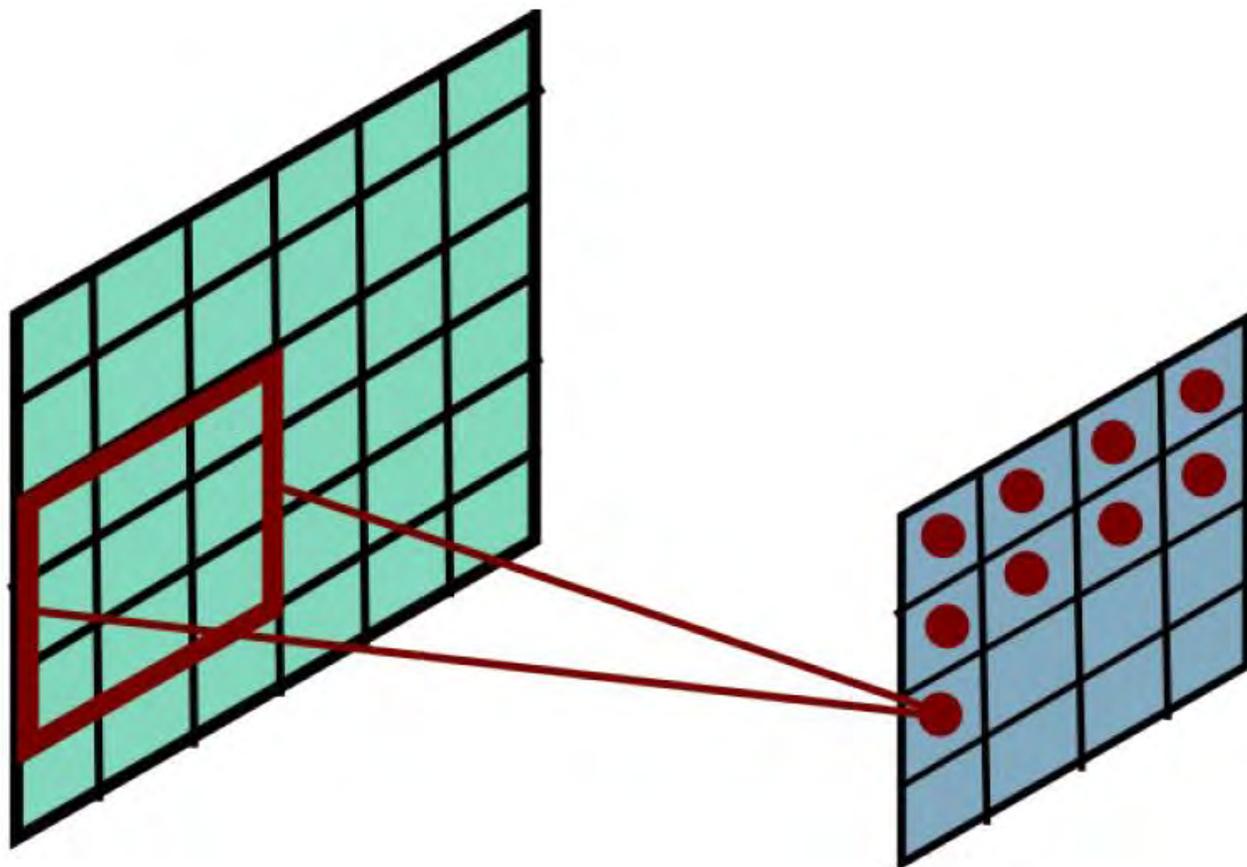
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



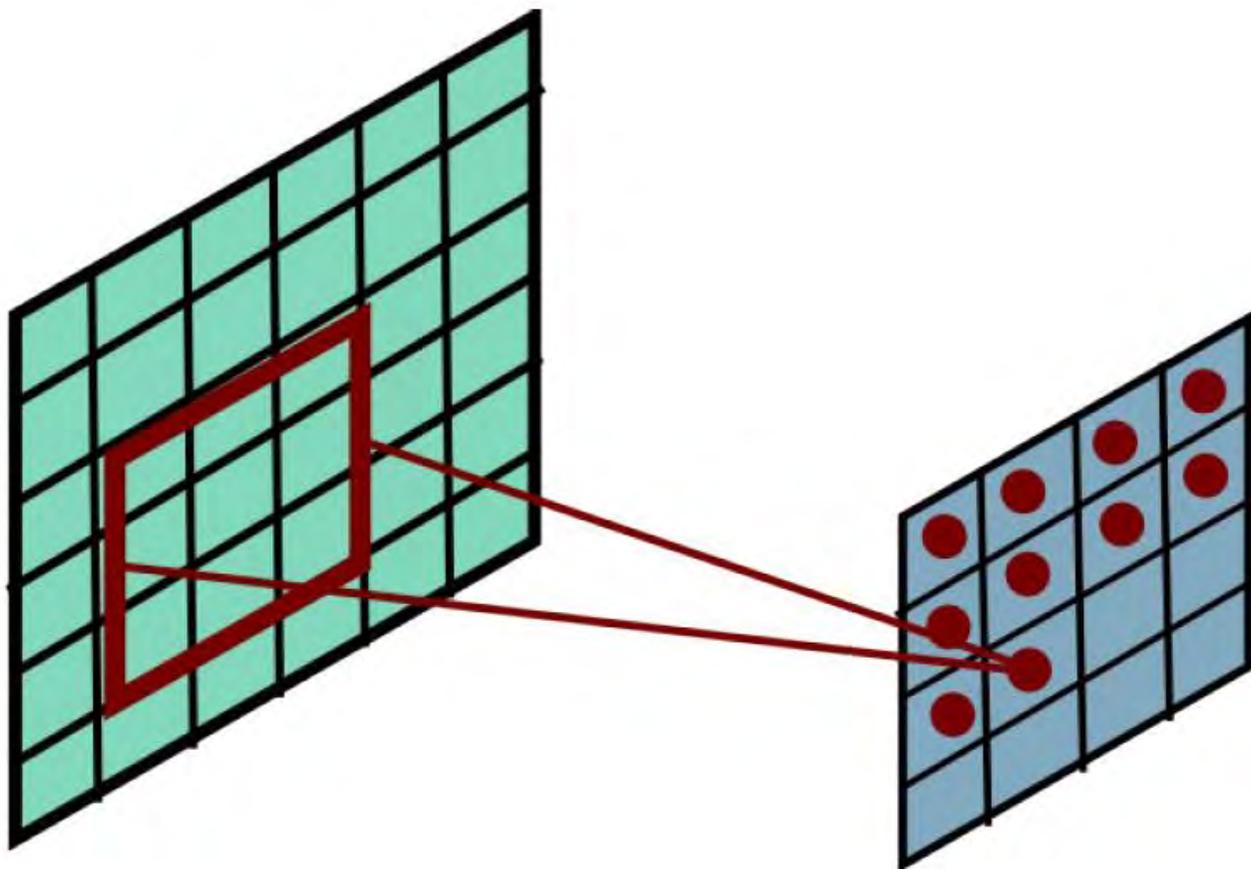
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



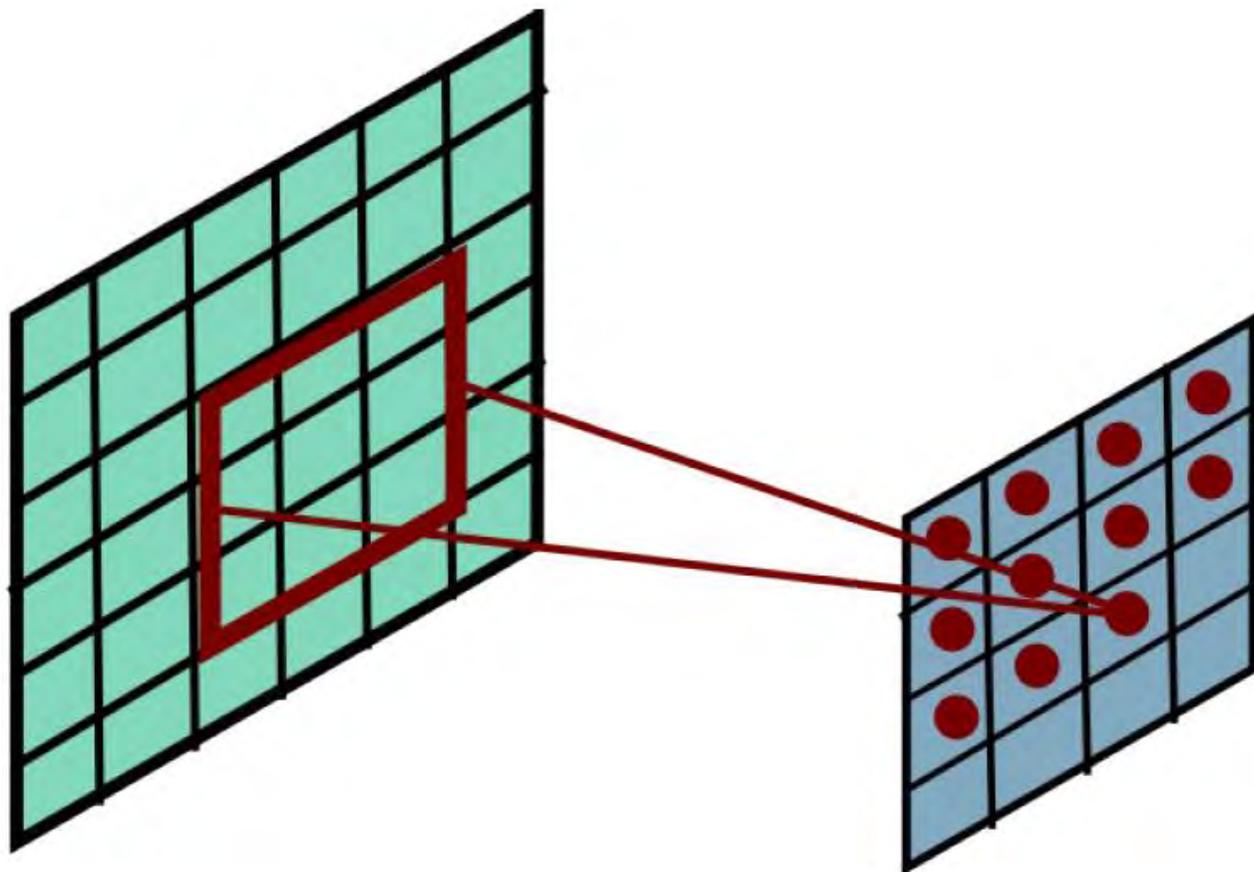
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



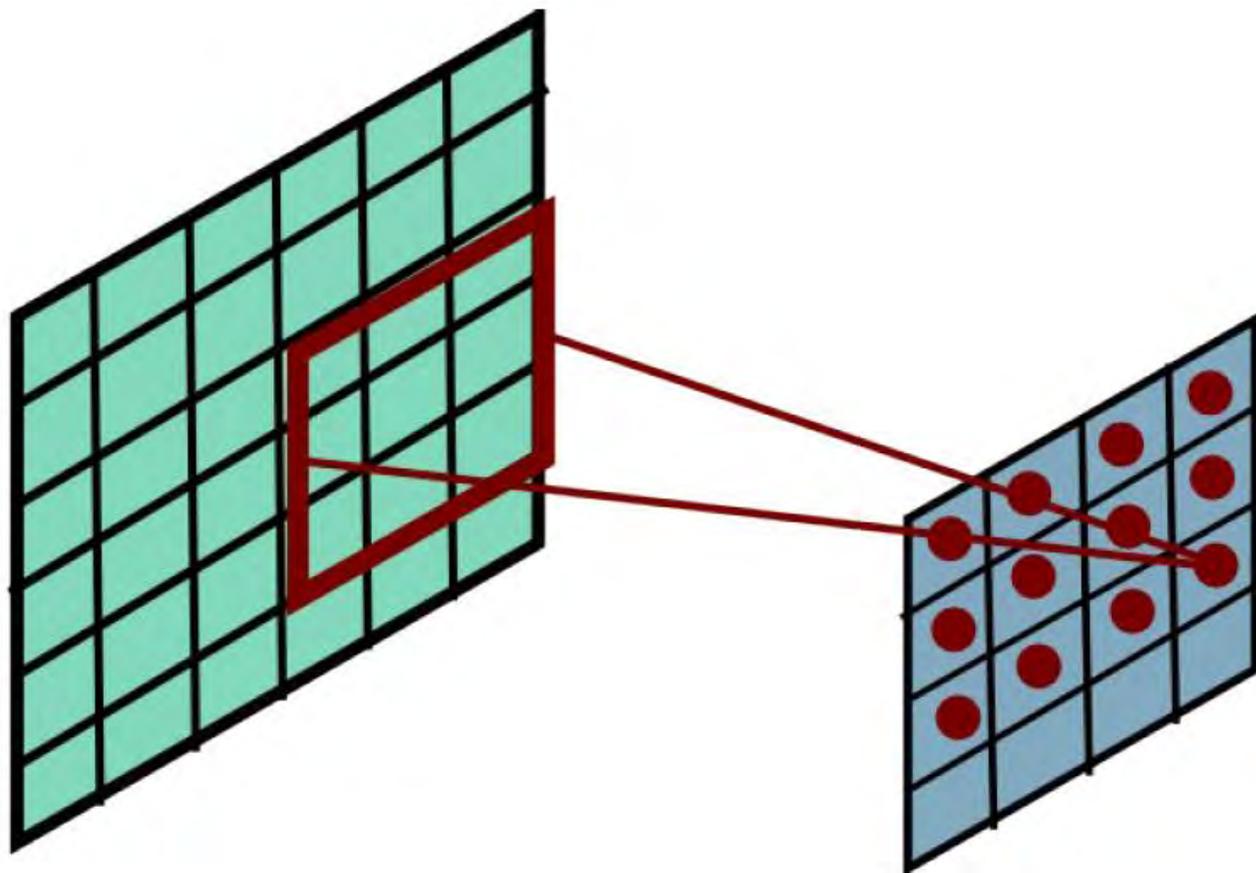
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



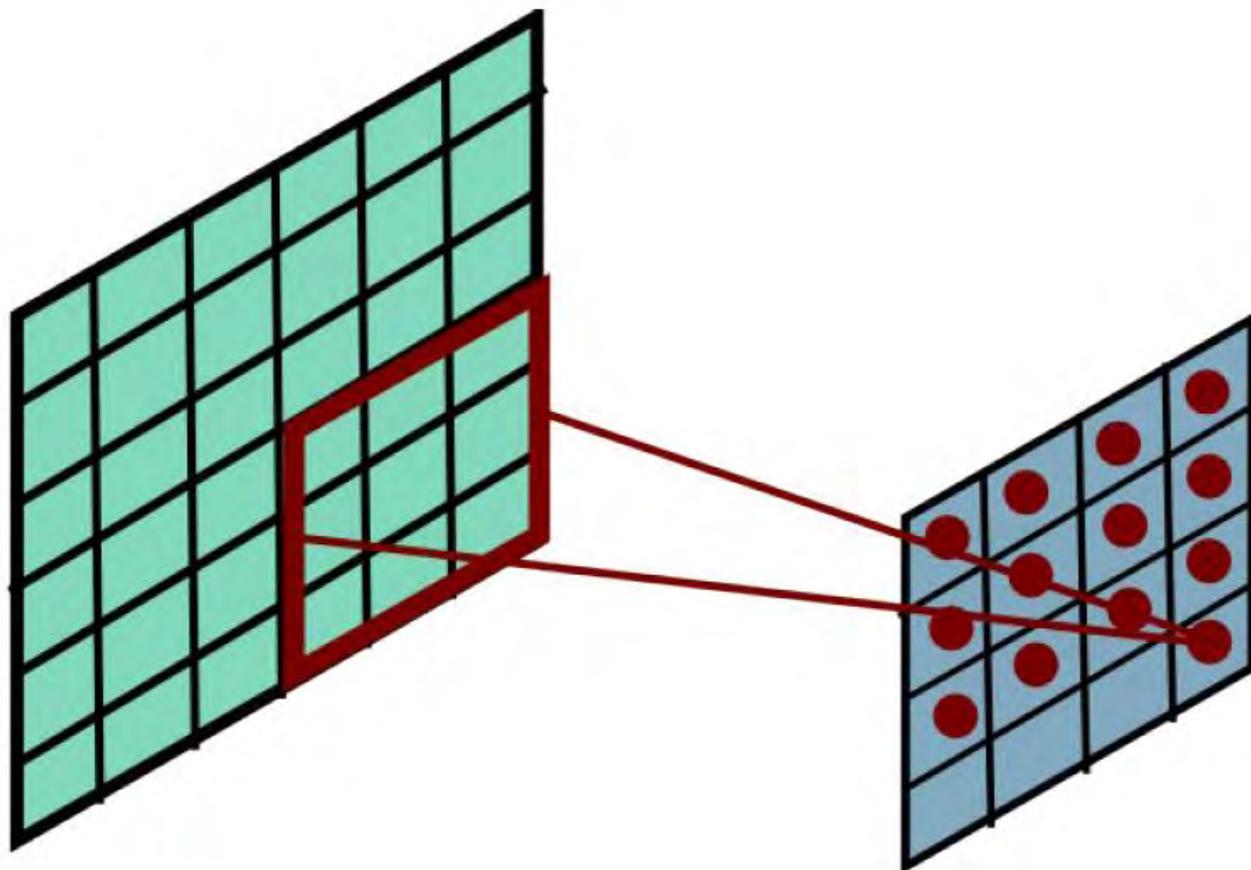
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



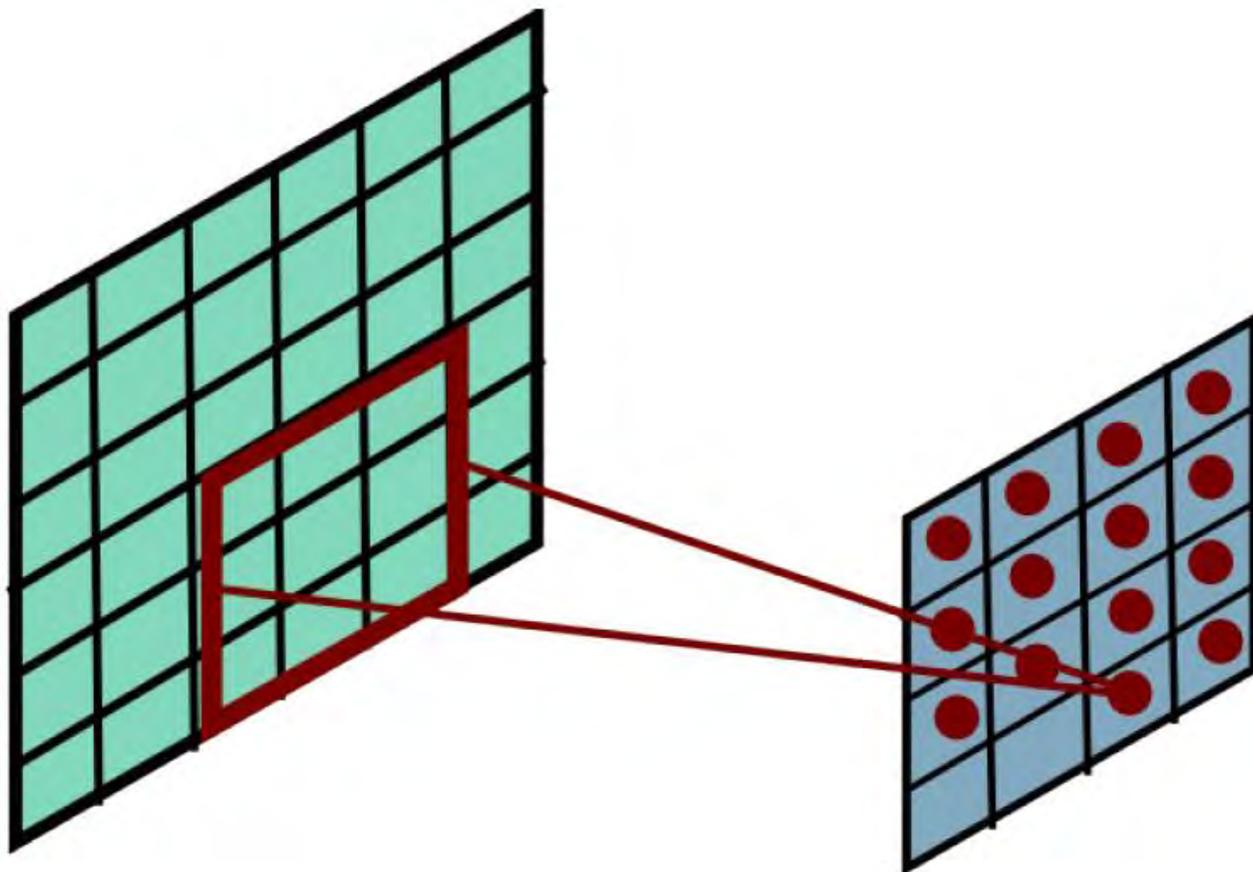
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



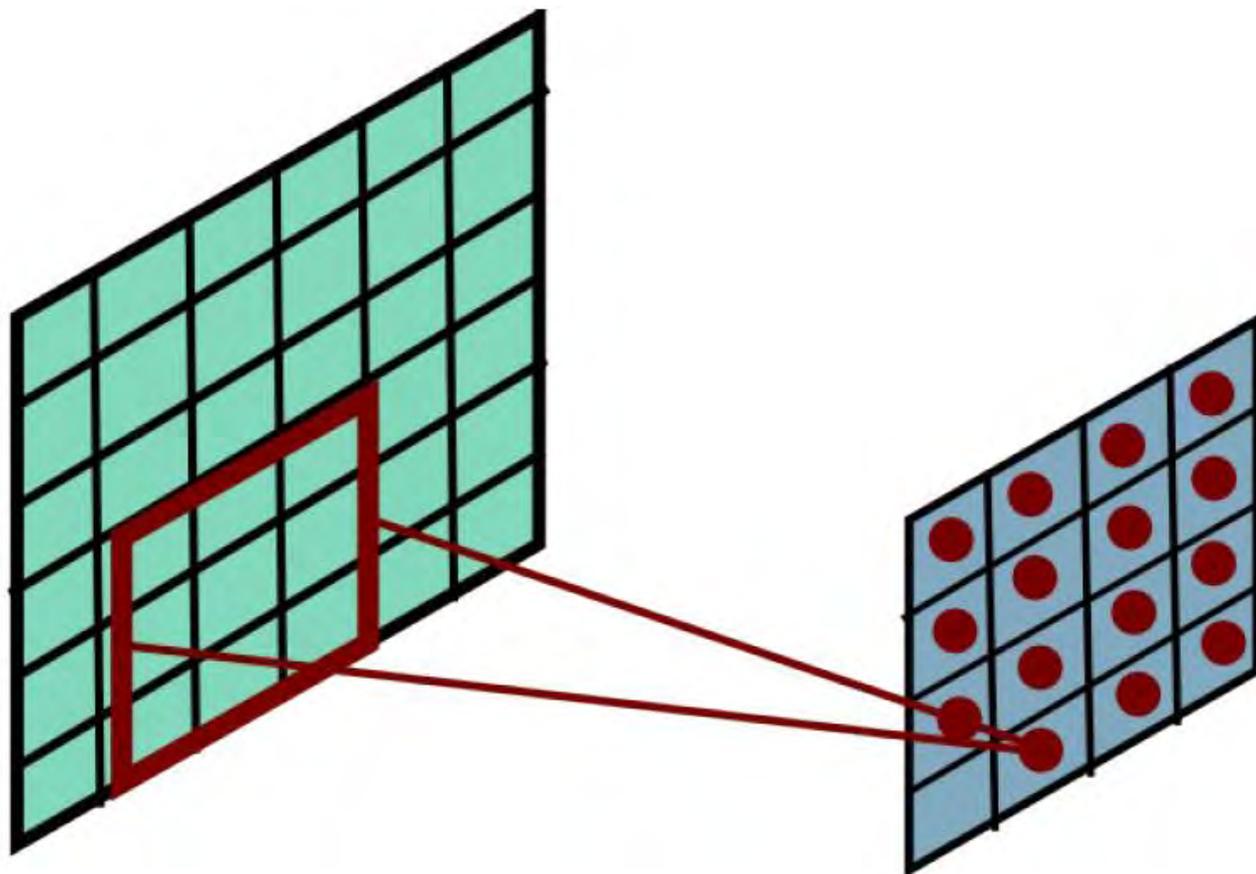
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



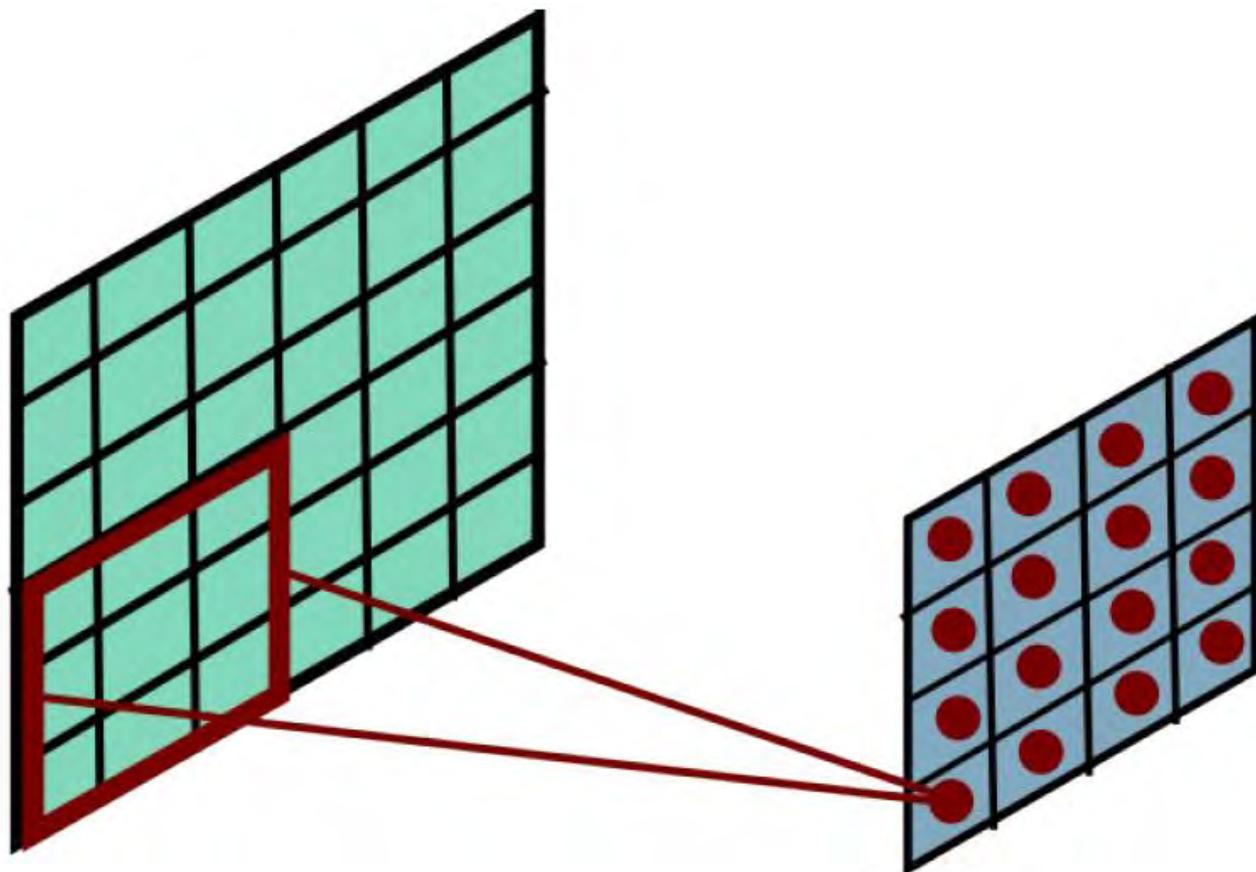
# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



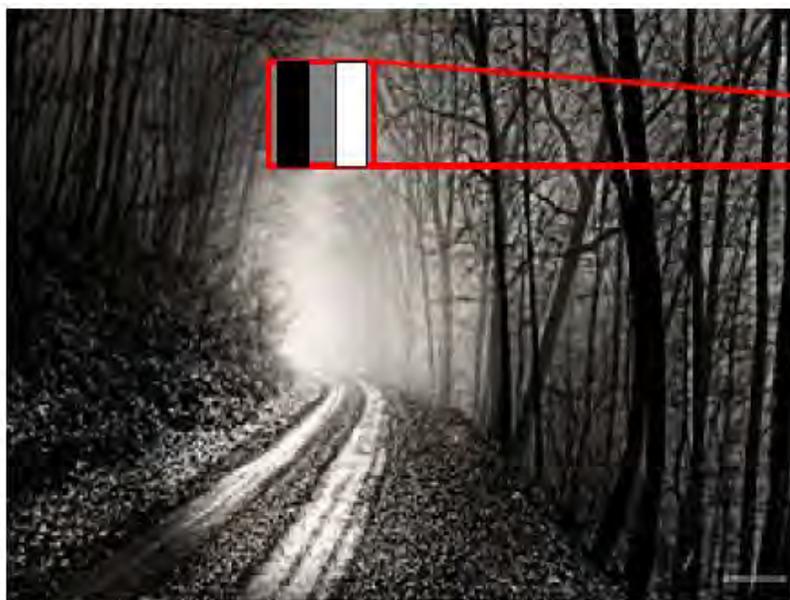
# 딥러닝은 왜 필요한가?

나선구조적인 층(Convolutional Layer)

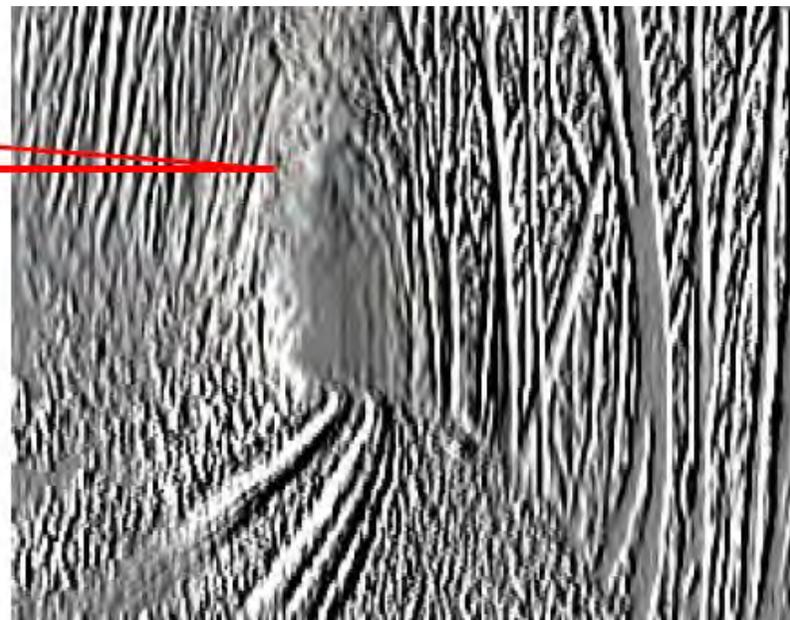


# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)

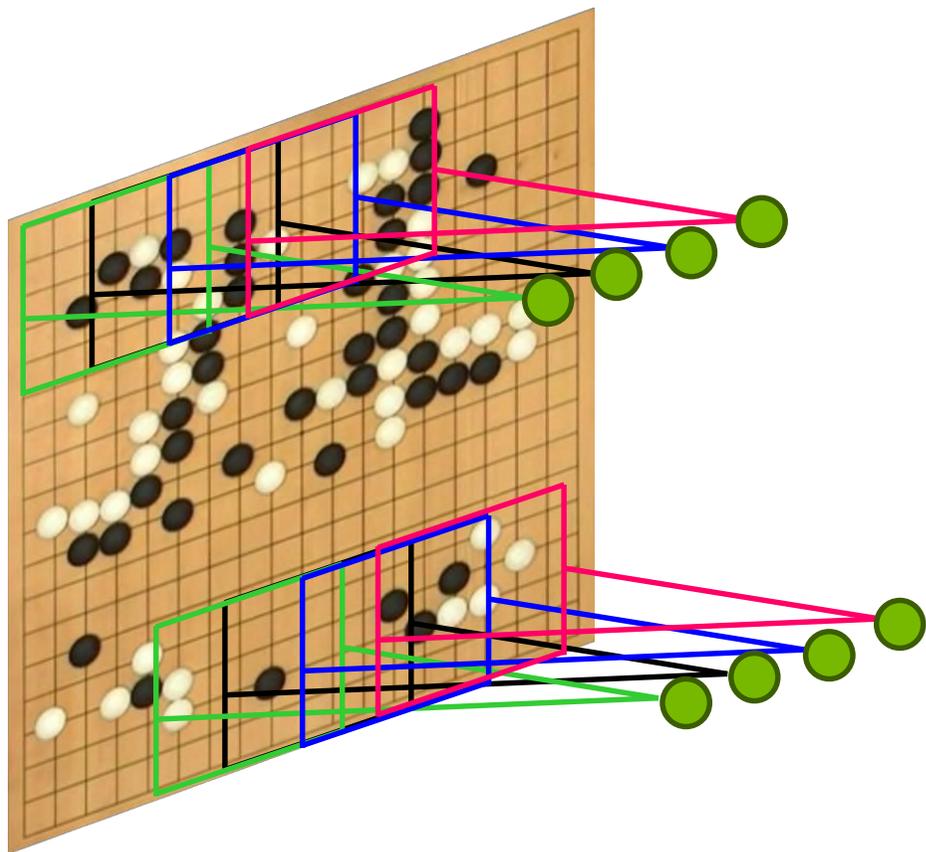


$$* \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} =$$



# 딥러닝은 어떤 역할을 하는가?

나선구조적인 층(Convolutional Layer)



## ➤ 다중 필터

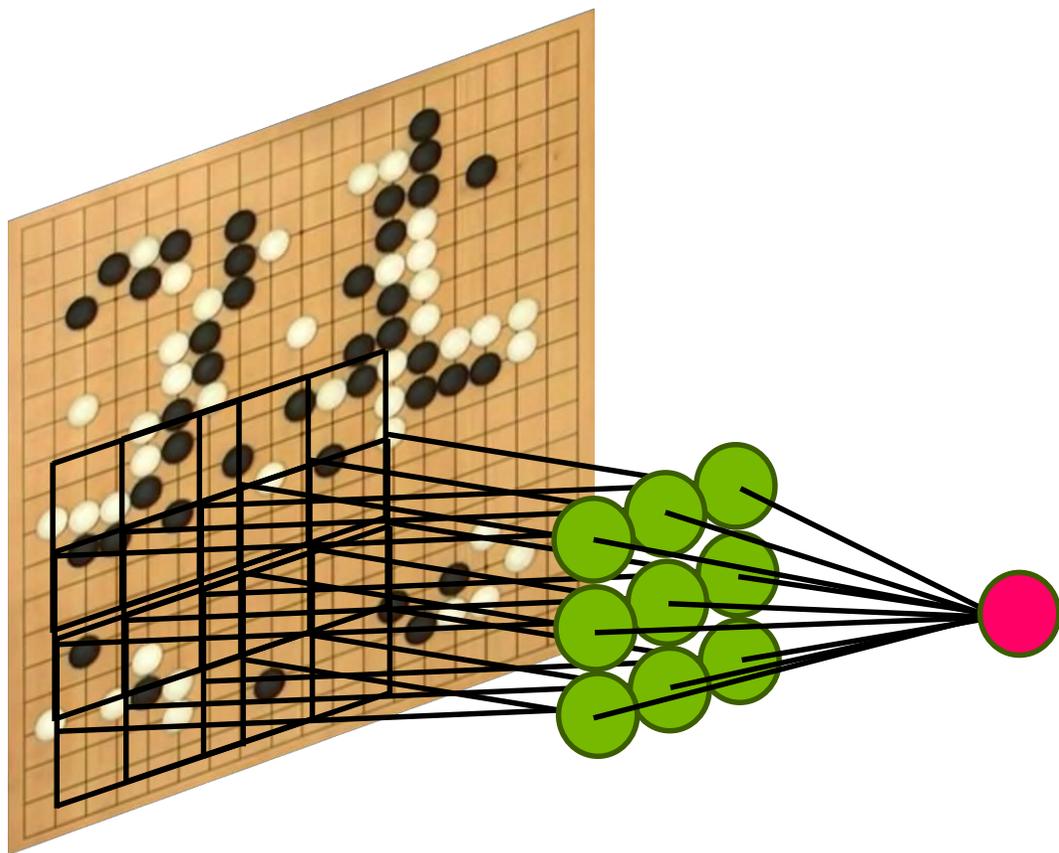
예 : 200 x 200 이미지

필터 : 100개, size : 10 x 10

매개변수 : 10K

# 딥러닝은 어떤 역할을 하는가?

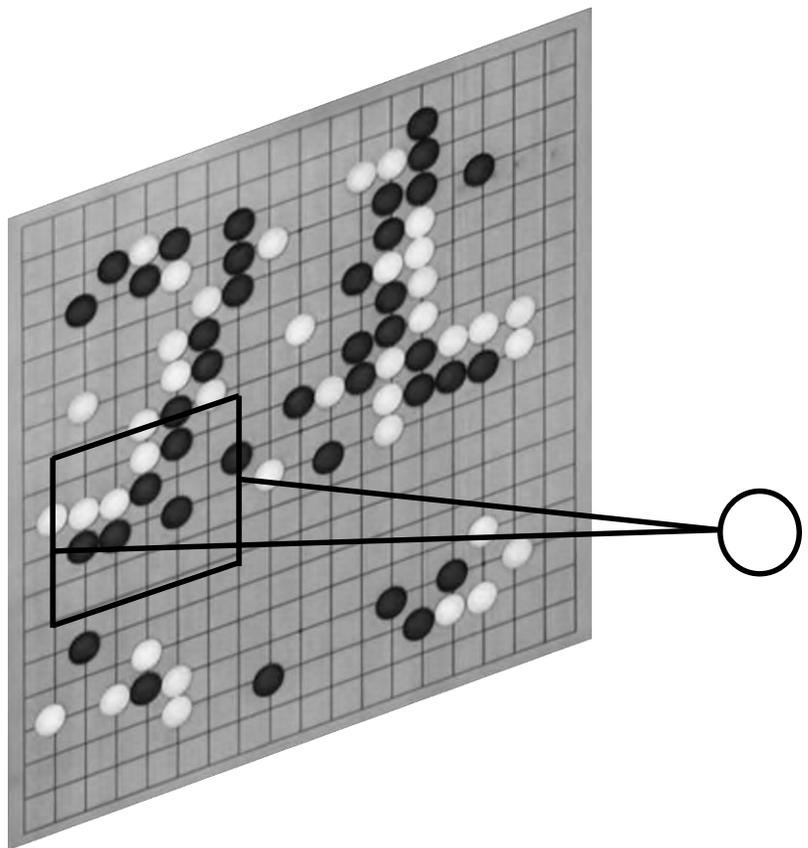
최대값 모으기(Max Pooling)



- 검은색 원(예 : 3x3 필터라면)에서 단일 3x3의 데이터를 얻기 위해 샘플링을 함.
- 방법 : 평균값, 선형조합값, 최대값.

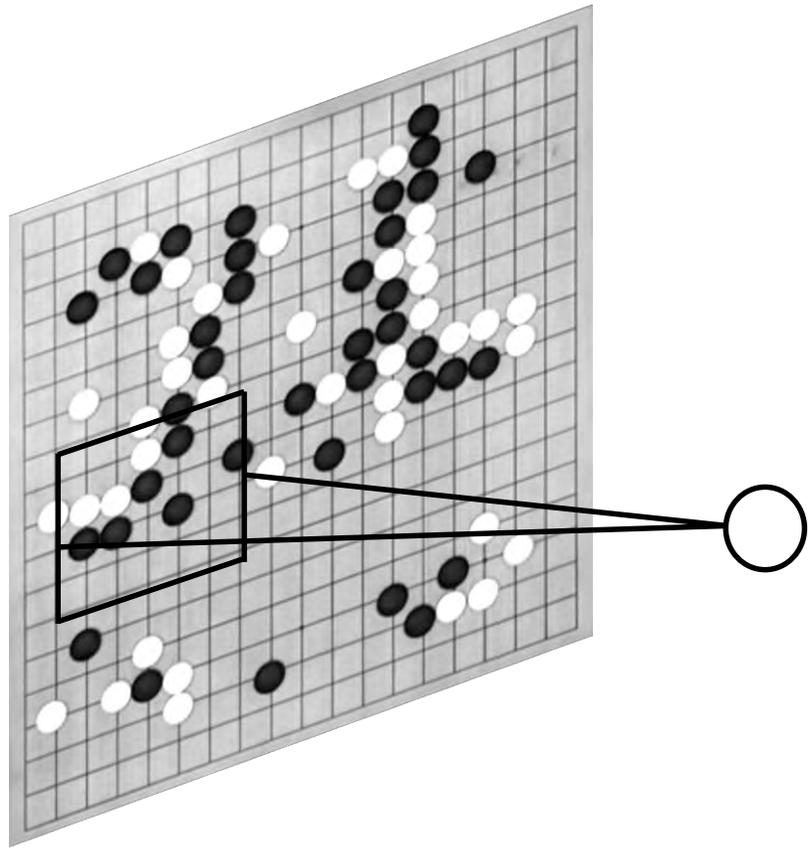
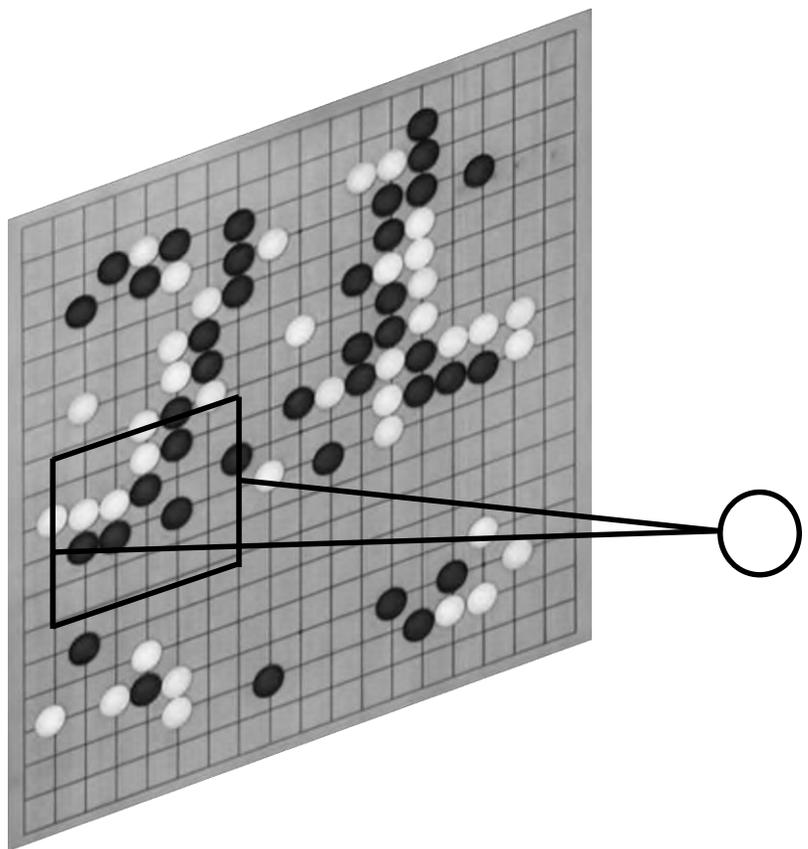
# 딥러닝은 어떤 역할을 하는가?

대비(차이) 정형화 작업



# 딥러닝은 어떤 역할을 하는가?

대비(차이) 정형화 작업



딥러닝은 어떤 역할을 하는가?

충분히 많은 데이터를 가지고  
반복적인 학습을 통해

**효율적으로**

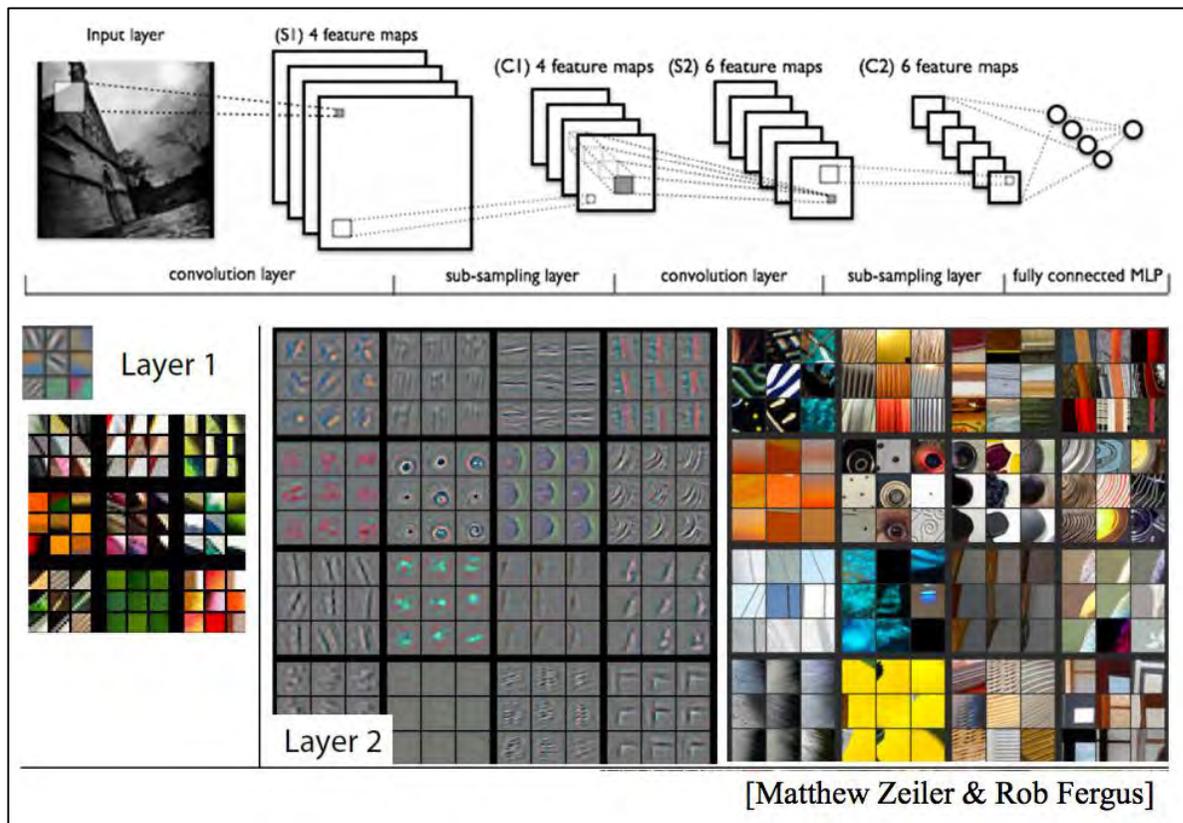
올바른 결론을 도출해 내기 위한  
과정

# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 딥러닝은 어떻게 만들어야 하는가?

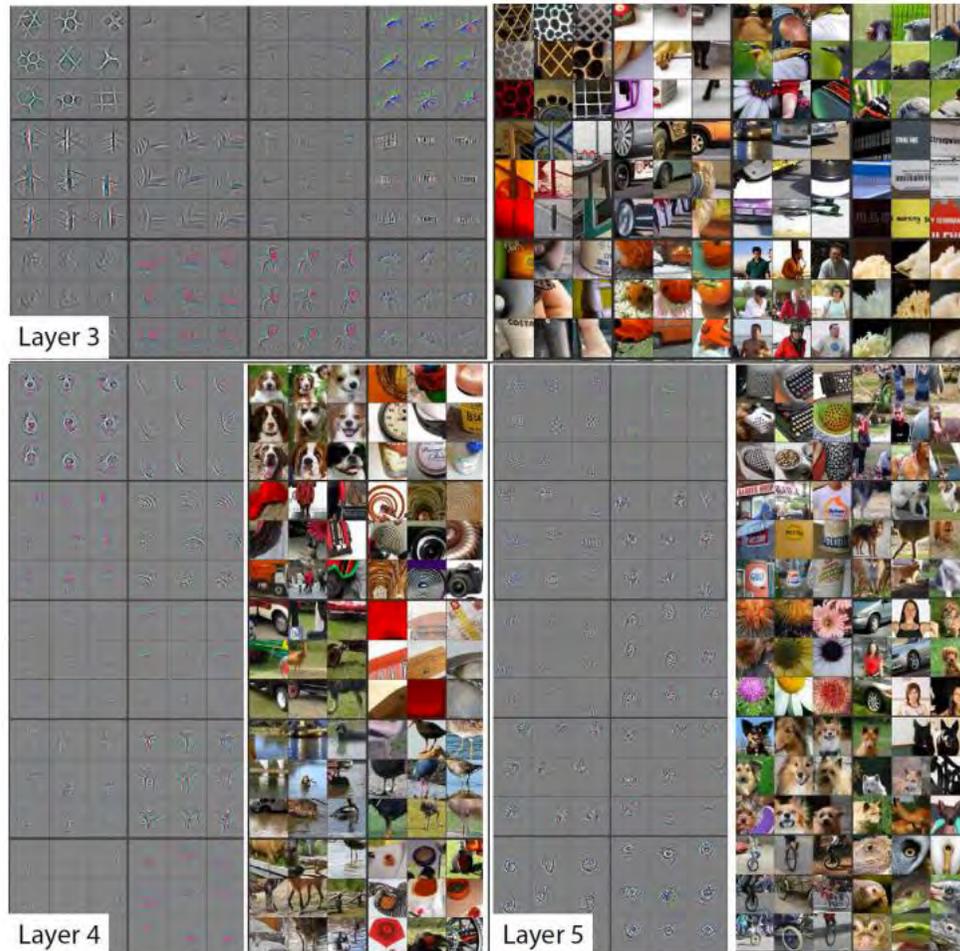
나선구조적인신경망(CNN)\*



- 인간의 시각령과 같이 매우 탁월함
- 시각 특징의 계층적 구조 가짐
- 지역 화소 특징들은 크기와 변형의 변화가 없음.
- 시각적으로 물체의 본질과 일반화가 잘 맞음

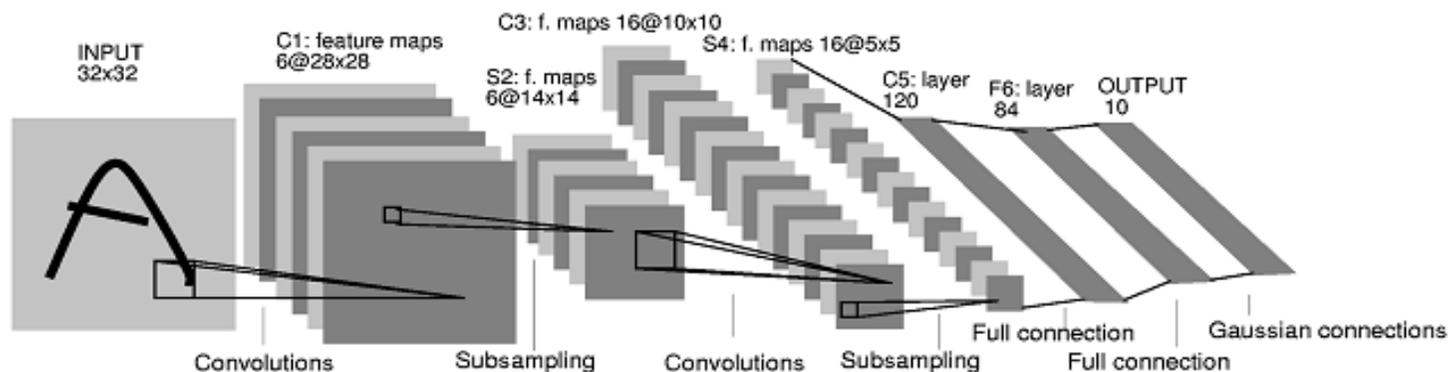
# 딥러닝은 어떻게 만들어야 하는가?

나선구조적인신경망(CNN)\*



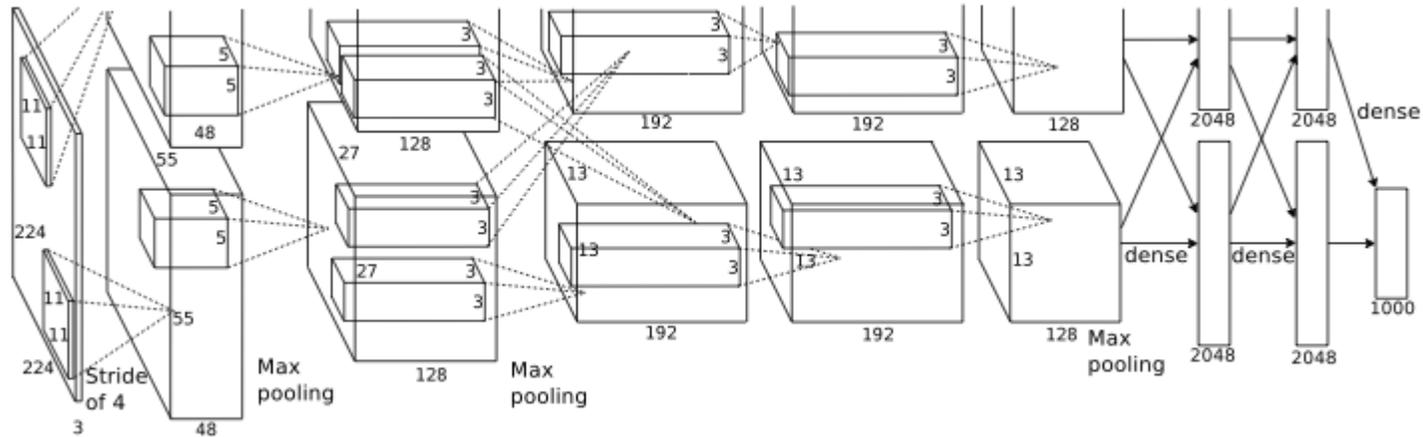
\*CNN : Convolutional Neural Network

# 딥러닝은 어떻게 만들어야 하는가?



Y. LeCun et al. 1989-1998 : Handwritten digit reading

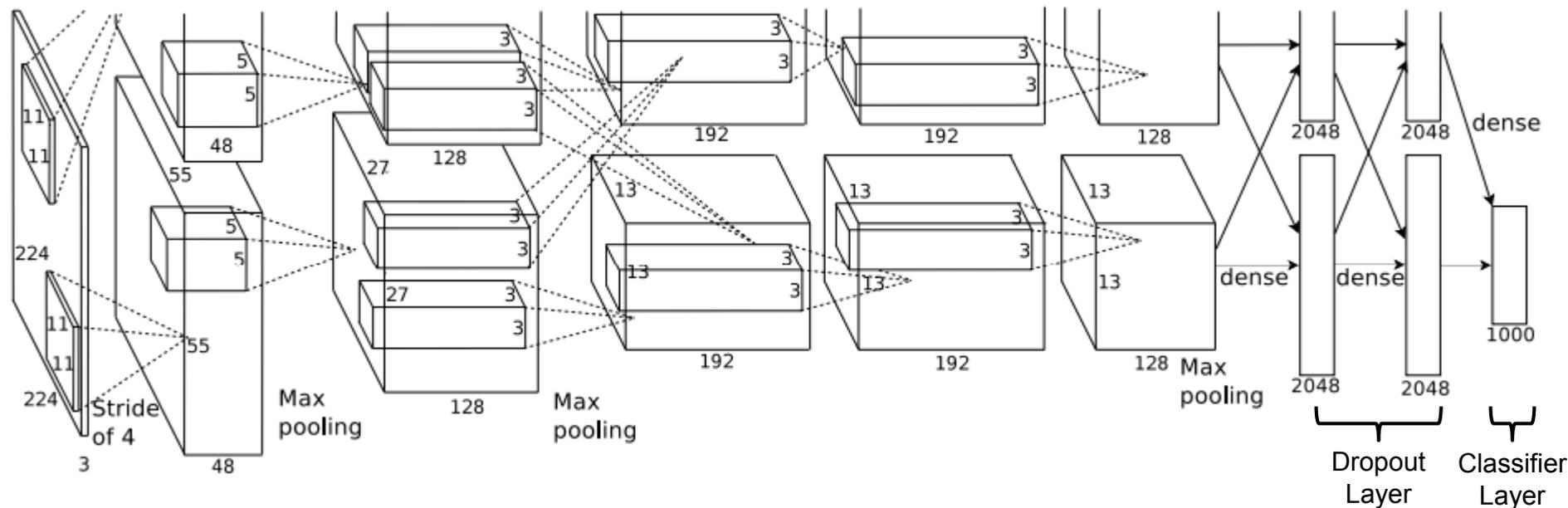
# 딥러닝은 어떻게 만들어야 하는가?



A. Krizhevsky, G. Hinton et al. 2012 : Imagenet classification winner

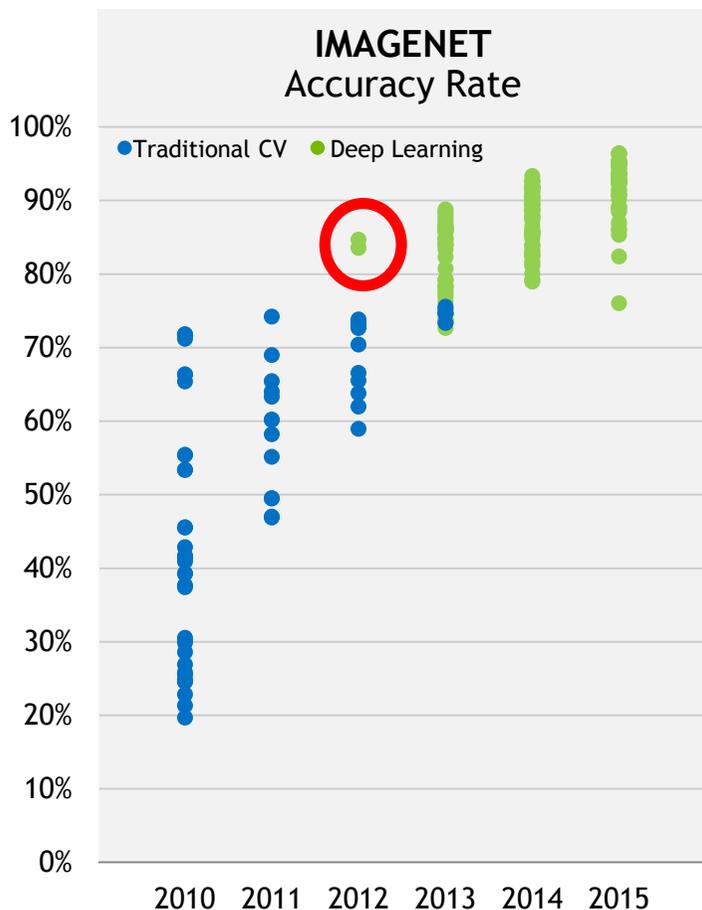
# Alexnet이란?

나선형신경망(CNN)을 이용하여 깊은 신경망(Deep Neural Network) 구현



- GTX580 3GB 2개 사용
- 8개의 Layer 구성
- 트레이닝 시간 : 5 ~ 6일 소요
- 필요한 계산량이 단일 CPU만으로 불가능에 가깝도록 커짐

# 딥러닝의 획기적인 도약



- IMAGENET : 1000개의 카테고리 와 100만개의 이미지로 구성되어 정확도를 겨루는 대회

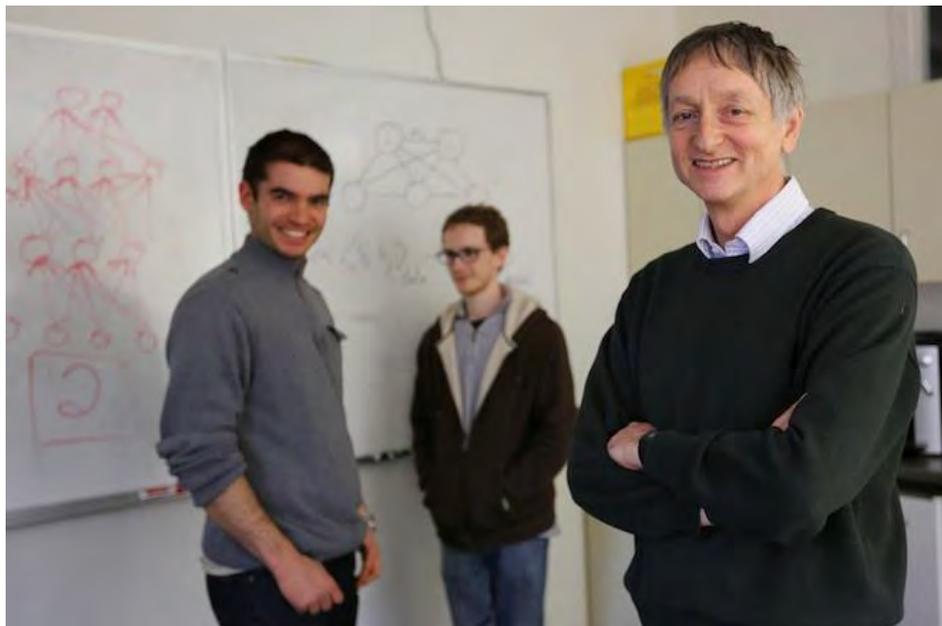
- 2012년 IMAGENET 정확도 84.7% 도달  
(알렉스 : 딥러닝 이용)

획기적인 사건 : 10여년동안 75%를 넘는 정확도를 보이지 못 함  
이후의 대부분의 참가자들이 딥러닝을 사용

알렉스 딥러닝 에러율

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
<i>SIFT + FVs [7]</i>	—	—	26.2%
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	16.4%
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	15.3%

# Alex Krizhevsky는 누구?



Ilya Sutskever

Alex Krizhevsky

Geoffrey Hinton

- IMAGENET CHALLENGE에서 딥러닝을 직접 구현
- 마의 벽 80% 정확도를 극복
- 소스코드 공개
- 딥러닝의 붐업을 이끔

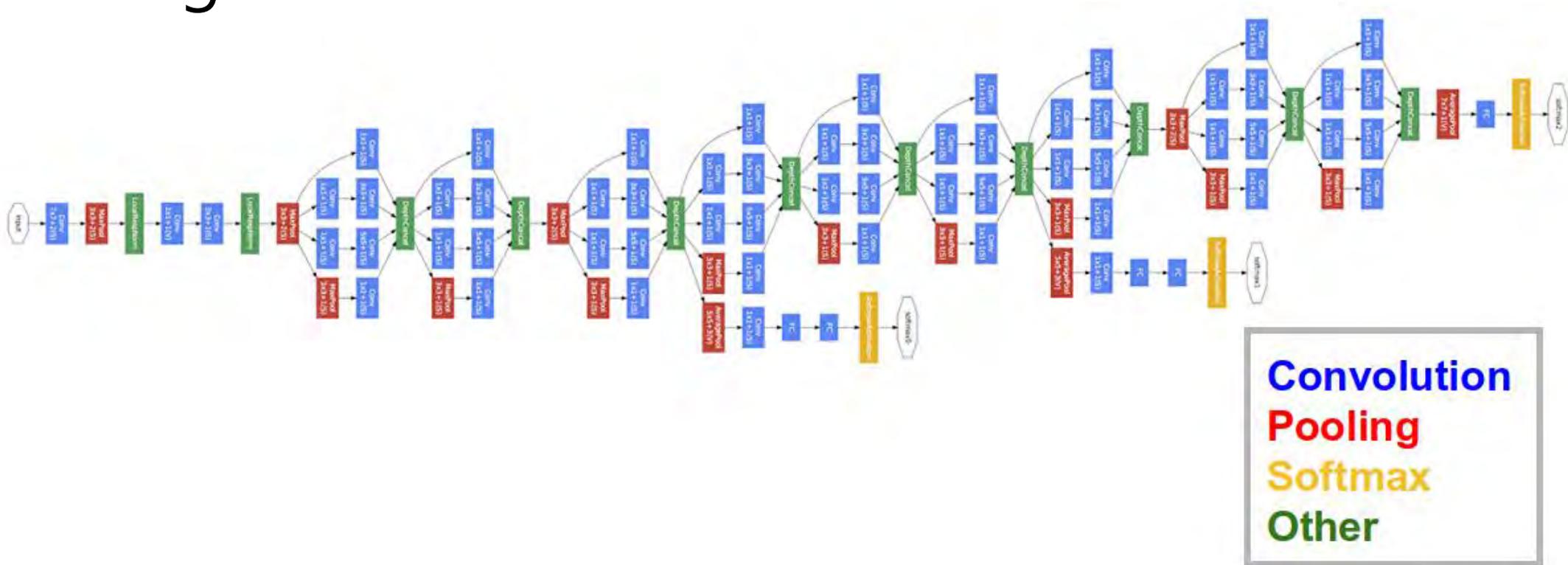
스탠포드 딥러닝 튜토리얼 : 딥러닝 대표기술저자중의 한명으로 선정



<http://deeplearning.stanford.edu/tutorial/>

# 딥러닝은 어떻게 만들어야 하는가?

## Google net



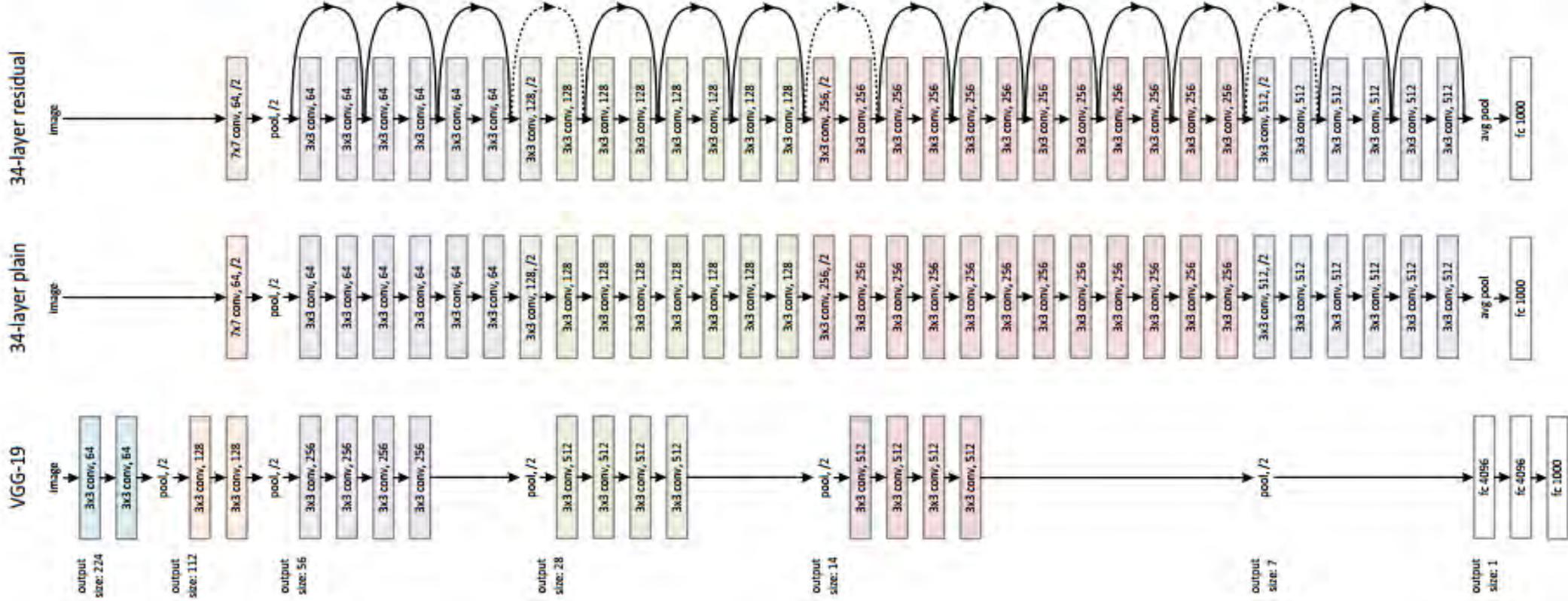
# 딥러닝은 어떻게 만들어야 하는가?

## Google net

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	-	16.4%	no
SuperVision	2012	1st	15.3%	ImageNet 22k
Clarifai	2013	-	11.7%	no
Clarifai	2013	1st	11.2%	ImageNet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

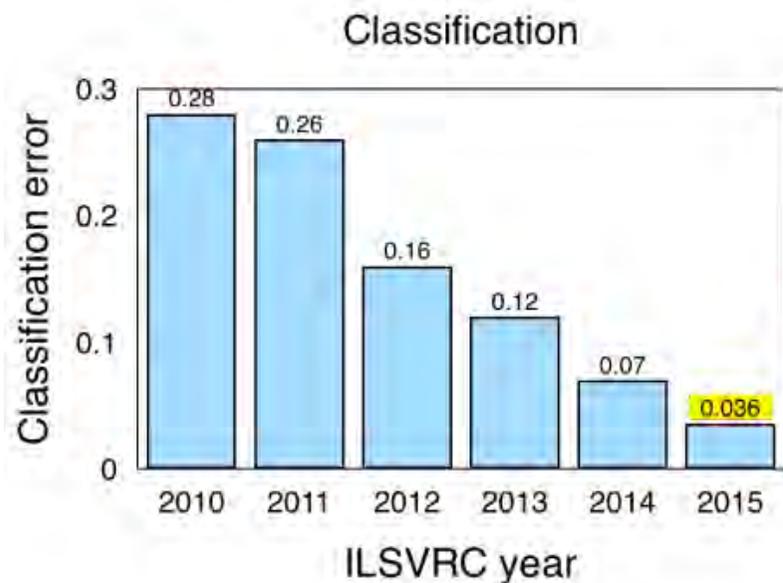
# 딥러닝은 어떻게 만들어야 하는가?

## MSRA 2015 Imagenet Challenge(152 Layer)

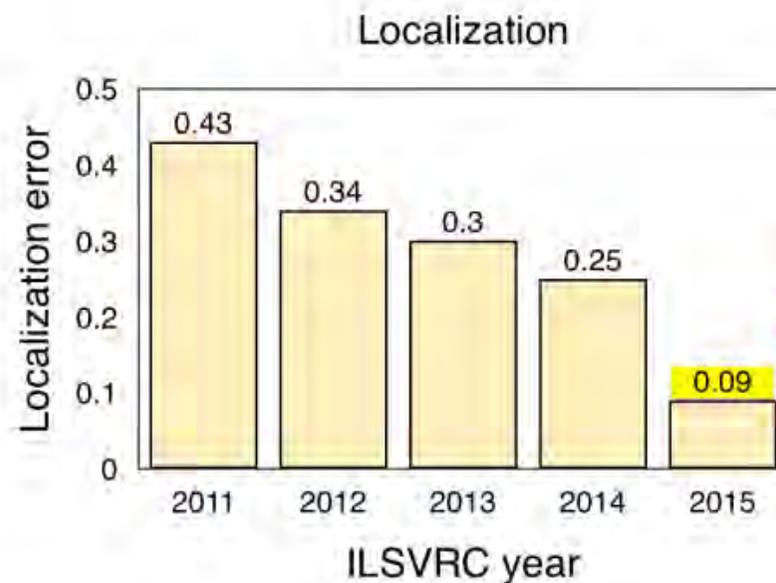


# 딥러닝은 어떻게 만들어야 하는가?

## MSRA 2015 Imagenet Challenge



1.9x



2.8x

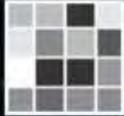


딥러닝은 어떻게 만들어야 하는가?

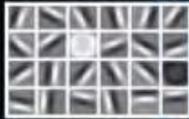
**충분히 많은 데이터를 가지고  
반복적인 학습을 통해  
효율적으로  
올바른 결론을 도출해 내기 위한  
시스템 설계**

## 인공지능의 얼굴 인식 과정

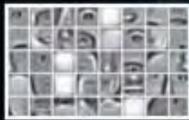
딥러닝 신경망은 얼굴과 같은 복잡한 형태를 단순한 형태에서 시작해 점점 복잡한 형태로 올라가며 그분화하면서 구분한다



**제1층** 컴퓨터가 밝고 어두운 픽셀을 구분한다



**제2층** 컴퓨터가 테두리와 단순한 형태를 구분한다



**제3층** 컴퓨터가 좀더 복잡한 형태와 사물을 구분한다



**제4층** 컴퓨터가 인간의 얼굴을 규정하는 형태를 파악한다

자료: Nature

## 인공신경망

뇌 신경망의 뉴런(신경세포)에서 신호가 전달되는 과정을 모방해 수식으로 표현한 것이 인공신경망이다



각각의 뉴런은 인공신경망에서 숫자로 표시된 데이터를 분류하는 수식으로 표현된다



뉴런과 뉴런을 잇는 연결지점으로 신경 신호가 전달되는 시냅스는 인공신경망에서 원으로 표시된 수식과 수식을 연결하는 선이 된다



한 수식에서 특정 조건에 맞는 정보를 추려내고, 그 결과가 다시 다음 단계 수식의 기본 자료가 된다. 이처럼 수식들이 서로 다양하게 연결되면서 작은 변화들을 모아 원하는 조건에 맞는 결과를 찾아낸다

왜 딥러닝이 떠 올랐는가?

# 왜 딥러닝이 떠 올랐는가?

## 빅 데이터

**facebook**

매일 350  
백만개의  
이미지업로드

**Walmart** ✱

시간당 2.5  
Petabytes  
고객데이터 생성

**You Tube**

매분당 150시간의  
비디오들이 업로드

# 왜 딥러닝이 떠 올랐는가?

## 빅 데이터

**facebook**

매일 350  
백만개의  
이미지업로드

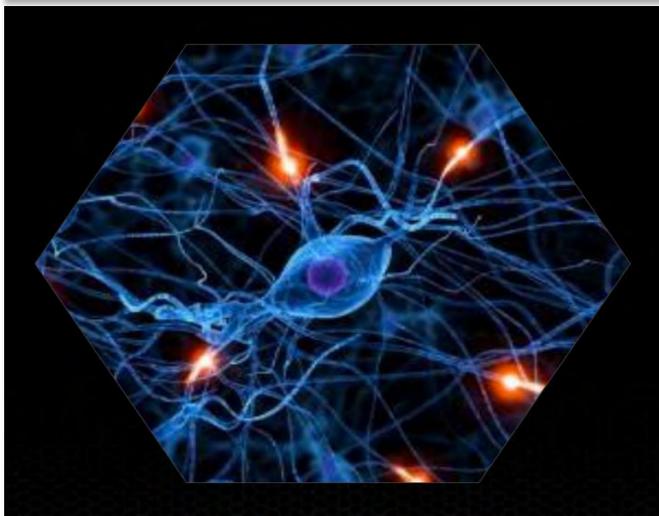
**Walmart** ✱

시간당 2.5  
Petabytes  
고객데이터 생성

**You Tube**

매분당 150시간의  
비디오들이 업로드

## 새로운 딥러닝 기술



# 왜 딥러닝이 떠 올랐는가?

## 빅 데이터

**facebook**

매일 350  
백만개의  
이미지업로드

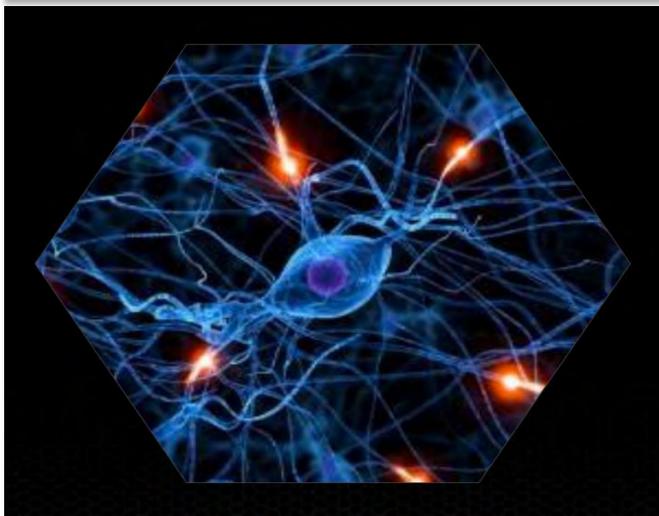
**Walmart** ✱

시간당 2.5  
Petabytes  
고객데이터 생성

**You Tube**

매분당 100시간의  
비디오들이 업로드

## 새로운 딥러닝 기술



## GPU 가속



# Agenda

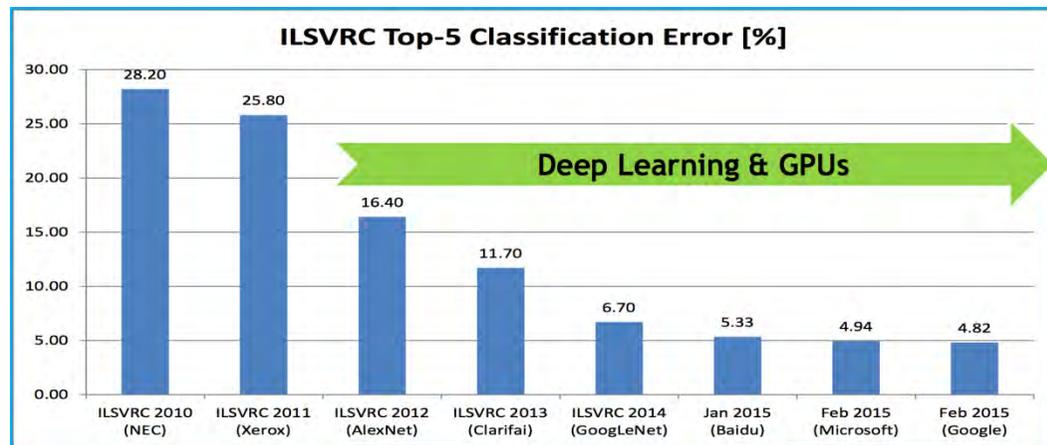
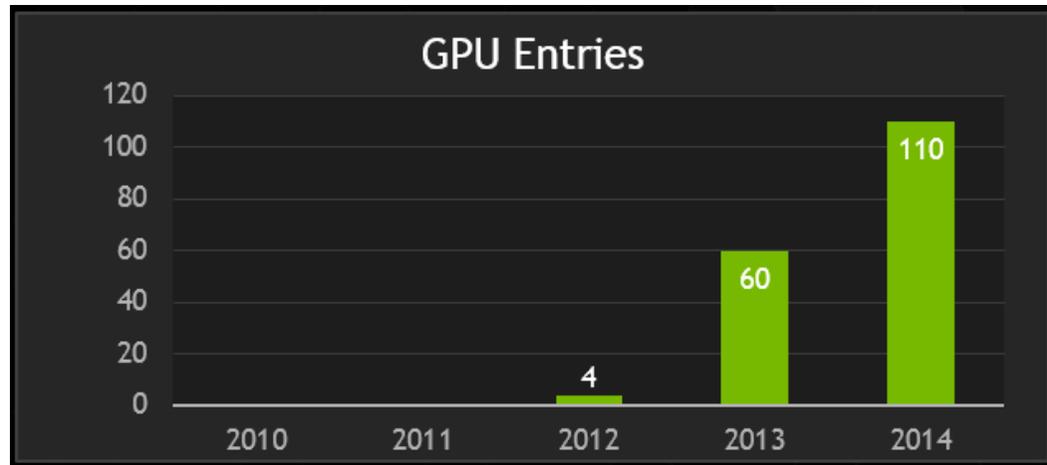
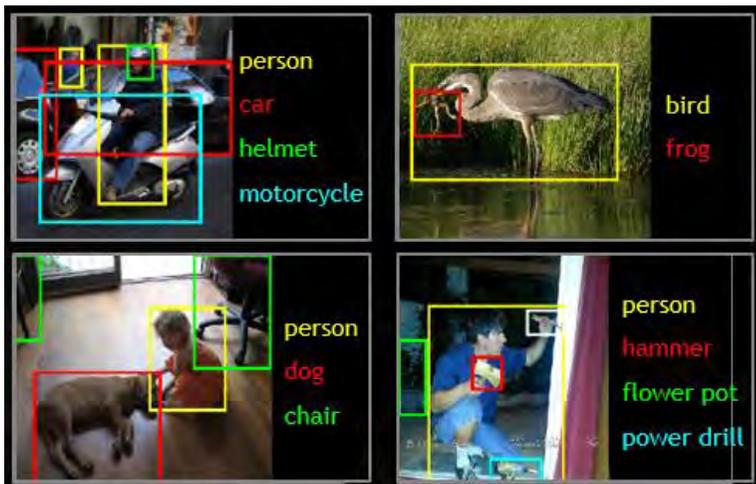
1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# GPU의 역할

## Image Recognition Challenge

1.2M training images • 1000 물체들의 범주

IMAGENET



Google Brain	스탠포드대 인공지능랩
CPU 서버 1,000 대	GPU 서버 3 대
16,000 코어	18,432 코어
50억원	3천3백만원
60 만와트	4 천와트



# GPU의 역할

왜 딥러닝에 GPU가 좋을까?

	Neural Networks	GPUs
Inherently Parallel	✓	✓
Matrix Operations	✓	✓
FLOPS	✓	✓
Bandwidth	✓	✓

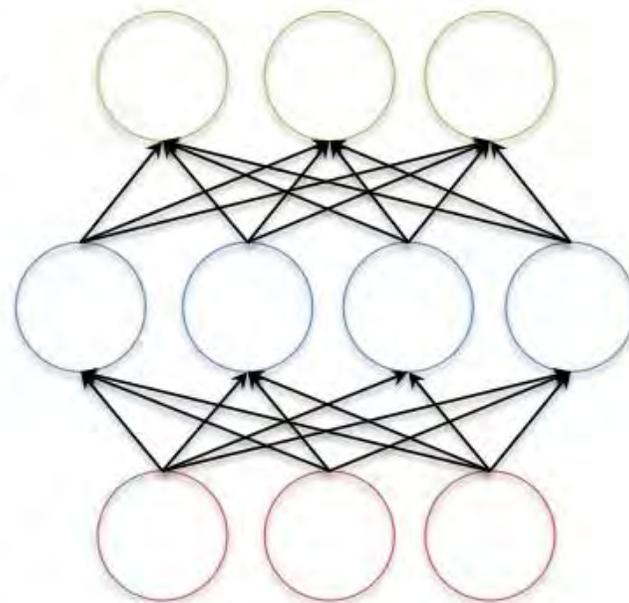
## GPU의 장점

- 예측정확도가 같거나 더 좋음
- 빠른 결과
- 차지하는 공간이 작음.
- 저전력
- 저비용

Output layer

Hidden layer

Input layer



# GPU의 역할

## GPU vs CPU 비교

Batch 사이즈	훈련시간 CPU	훈련시간 GPU	GPU 가속 능력
64 images	64 초	7.5 초	8.5배
128 images	124 초	14.5 초	8.5배
256 images	257 초	28.5 초	9.0배

- ▶ ILSVRC12 winning model: "Supervision"
- ▶ 7 layers
- ▶ 5 convolutional layers + 2 fully-connected
- ▶ ReLU, pooling, drop-out, response normalization
- ▶ Implemented with Caffe
- ▶ Training time is for 20 iterations
- ▶ Dual 10-core Ivy Bridge CPUs
- ▶ 1 Tesla K40 GPU(2880 core)
- ▶ CPU times utilized Intel MKL BLAS library
- ▶ GPU acceleration from CUDA matrix libraries (cuBLAS)

# GPU의 역할

## GPU vs CPU 비교

### Car

	Method	Setting	Code	Moderate	Easy	Hard	Runtime	Environment	Compare
1	<a href="#">sensekitti</a>			90.03 %	91.19 %	81.69 %	4.5 s	GPU @ 2.5 Ghz (Python + C/C++)	<input type="checkbox"/>
Anonymous submission									
2	<a href="#">NVDriVeNet-H</a>			89.81 %	90.92 %	83.76 %	0.15s	GPU @ 2.5 Ghz (Python + C/C++)	<input type="checkbox"/>
3	<a href="#">SDP+RPN</a>			88.85 %	90.14 %	78.38 %	0.2 s	GPU @ 2.5 Ghz (Python + C/C++)	<input type="checkbox"/>
F. Yang, W. Choi and Y. Lin: <a href="#">Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers</a> . Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition 2016.									
S. Ren, K. He, R. Girshick and J. Sun: <a href="#">Faster R-CNN: Towards real-time object detection with region proposal networks</a> . Advances in Neural Information Processing Systems 2015.									
4	<a href="#">Mono3D</a>			88.66 %	92.33 %	78.96 %	4.2 s	GPU @ 2.5 Ghz (Matlab + C/C++)	<input type="checkbox"/>
X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler and R. Urtasun: <a href="#">Monocular 3D Object Detection for Autonomous Driving</a> . CVPR 2016.									
5	<a href="#">3DOP</a>		<a href="#">code</a>	88.64 %	93.04 %	79.10 %	3s	GPU @ 2.5 Ghz (Matlab + C/C++)	<input type="checkbox"/>
X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler and R. Urtasun: <a href="#">3D Object Proposals for Accurate Object Class Detection</a> . NIPS 2015.									
6	<a href="#">SubCNN</a>			88.55 %	90.74 %	77.95 %	2 s	GPU @ 3.5 Ghz (Python + C/C++)	<input type="checkbox"/>
Anonymous submission									
7	<a href="#">SDP+CRC (ft)</a>			83.53 %	90.33 %	71.13 %	0.6 s	GPU @ 2.5 Ghz (C/C++)	<input type="checkbox"/>
F. Yang, W. Choi and Y. Lin: <a href="#">Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers</a> . Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition 2016.									
8	<a href="#">Faster R-CNN</a>		<a href="#">code</a>	81.84 %	86.71 %	71.12 %	2 s	GPU @ 3.5 Ghz (Python + C/C++)	<input type="checkbox"/>

# GPU의 역할

엔비디아가 지원하는 딥러닝 요소들

음성이해

이미지분석

언어처리

사용자 어플리케이션들

DIGITS

GPU 가속되는 딥러닝 프레임웍(Caffe, Torch, Theano)

최적의 성능을 낼 수 있는 라이브러리(cuDNN, cuBLAS)

CUDA- 최고의 병렬 프로그램 툴 키트

GPU- 세계 최고의 딥러닝 하드웨어

# GPU의 역할

GPU 가속 딥러닝 프레임워크의 종류(Tensorflow, CNTK)

	CAFFE	TORCH	THEANO	MINERVA	KALDI
Domain	Deep Learning Framework	Scientific Computing Framework	Math Expression Compiler	Deep Learning Framework	Speech Recognition Toolkit
cuDNN	4.0	4.0	4.0	4	--
Multi-GPU	via DIGITS 2	In Progress	In Progress	✓	✓(nnet2)
Multi-Node	✗	✗	✗	✗	✓(nnet2)
License	BSD-2	GPL	BSD	Apache 2.0	Apache 2.0
Interface(s)	Command line, Python, MATLAB	Lua, Python, MATLAB	Python	C++	C++, Shell scripts
Embedded (TK1)	✓	✓	✗	✗	✗
사용예	머신비전	구글, 트위터, 페이스북	다차원 배열 라이브러리		음성인식

# GPU의 역할

## EDUCATION

TORCH



CAFFE



THEANO



MATCONVNET



MOCHA.JL



PURINE



MINERVA



MXNET\*



TENSORFLOW



WATSON



CNTK



## START-UPS

CHAINER



DL4J



KERAS



OPENDEEP



NVIDIA GPU PLATFORM

\*U. Washington, CMU, Stanford, TuSimple, NYU, Microsoft, U. Alberta, MIT, NYU Shanghai

# GPU의 역할

## THE EXPANDING UNIVERSE OF MODERN AI

### "THE BIG BANG"

Big Data  
GPU  
Algorithms

### RESEARCH



### CORE TECHNOLOGY / FRAMEWORKS



### AI-as-a-PLATFORM



### START-UPS



1,000+ AI START-UPS  
\$5B IN FUNDING  
Source: Venture Scanner

### INDUSTRY LEADERS



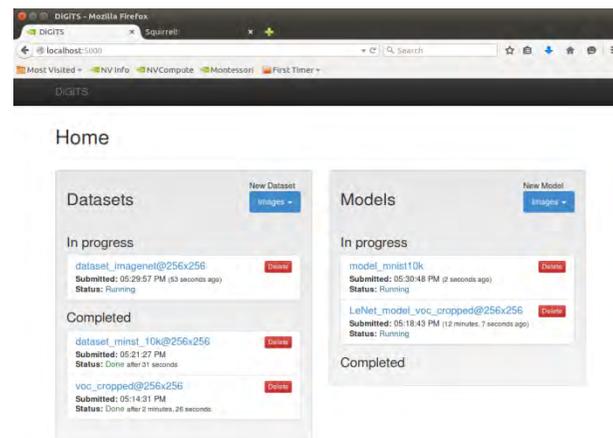
# GPU의 역할

딥러닝을 위한 엔비디아 솔루션



## DIGITS DevBox

- 4개의 TITAN X GPU
  - 7TFlops 단일정밀도
  - 336.5 GB/s 메모리 대역
  - 12GB
- Ubuntu 14.04
- w/Caffe, Torch, Theano, cuDNN v4, CUDA 7.5



## DIGITS

- NVIDIA Deep Learning GPU Training System (DIGITS)

# GPU의 역할

딥러닝을 위한 엔비디아 솔루션

## NVIDIA DGX-1

World's First Deep Learning Supercomputer



Engineered for deep learning

170 TF FP16

8x Tesla P100 in hybrid cube mesh

Accelerates major AI frameworks



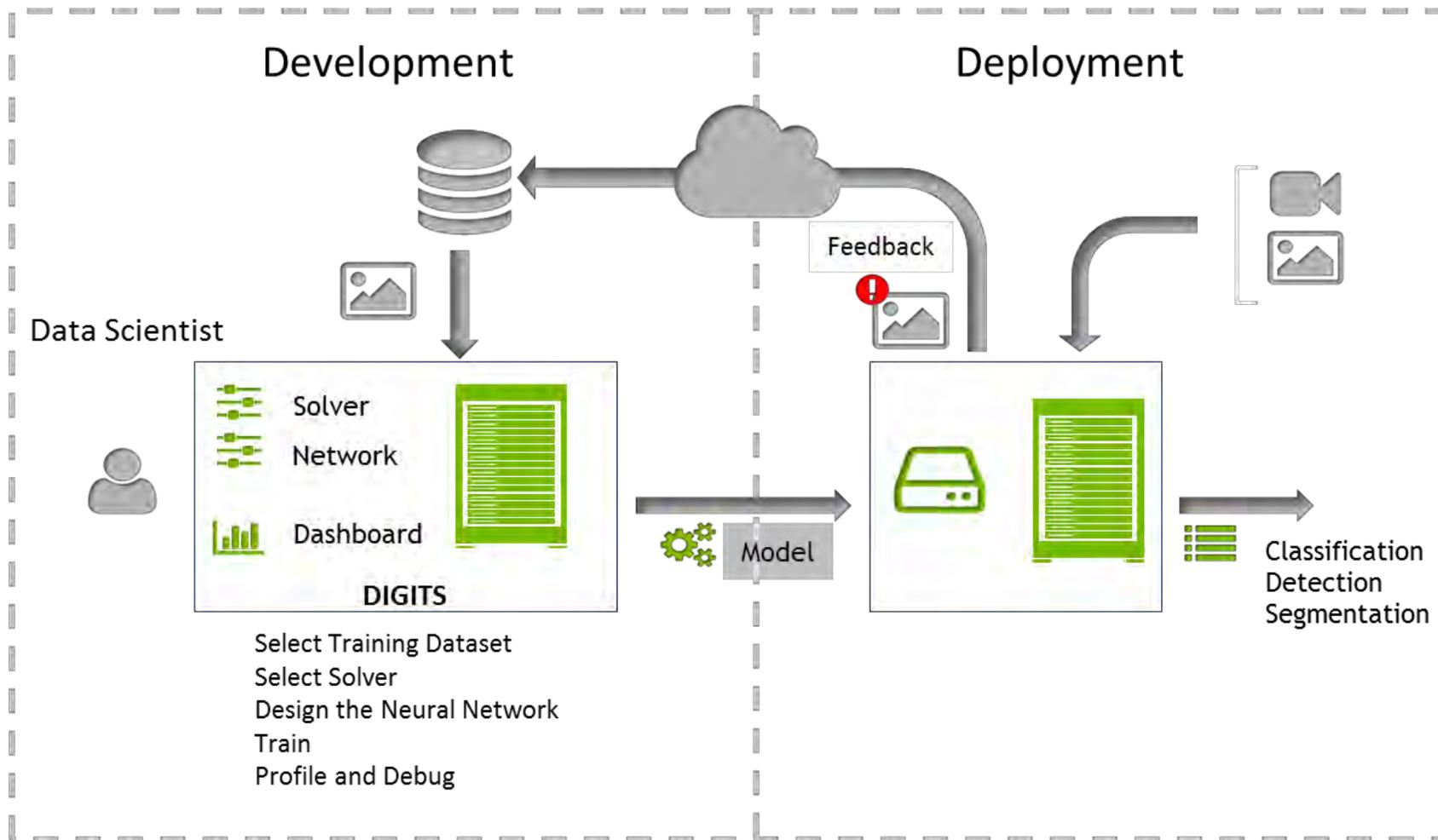
1.33 billion images/day

# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 딥러닝의 기여

## 딥러닝의 전체 흐름도



# 딥러닝의 기여

구글무인차



# 딥러닝의 기여

## 예술작품 스타일 인식과 모방

VisLAB: RESULTS

experiment `wikpaintings_mar23` setting `caffe_fc7 None vw` style `style_impresionism`  
split `test` actual\_label `all` predicted\_label `positive` confidence `decreasing` page `1`



conf: 1.56 | gt: +



conf: 1.51 | gt: +



conf: 1.49 | gt: +



conf: 1.41 | gt: +



conf: 1.41 | gt: +



conf: 1.40 | gt: +

<http://demo.vislab.berkeleyvision.org/>

The Deep Forger  
@DeepForger

Follow

#StyleNet #NeuralArt, comission from  
@FinnSiegmond and style by Pablo Picasso.



# 강화학습

## 다양한 게임



Figure 1: Screen shots from five Atari 2600 Games: (Left-to-right) Pong, Breakout, Space Invaders, Seaquest, Beam Rider



Google DeepMind

# Agenda

1. 딥러닝이란 무엇인가?
2. 인공지능이란 무엇인가?
3. 딥러닝은 왜 필요한가?
4. 딥러닝은 어떤 역할을 하는가?
5. 딥러닝은 어떻게 만들어야 하는가?
6. GPU의 역할
7. 딥러닝의 기여
8. AlphaGo와 GPU

# 이세돌 vs 알파고 비교



데뷔시기(나이)	1988년(33세)	2015년(2세)
프로경력	21년(우승 47회)	1년(우승 1회) (2016 딥마인트챌린즈매치)
단위(段位)	9단	2 ~ 5단(자체추정)
경험치	10,000경기	100,000경기
트레이닝	30,000시간	30,000시간(GPU)
검색 파워	100 위치/초	100,000 위치/초
컴퓨팅파워	본인두뇌	100 GPUs

# 알파고(AlphaGo) 컴퓨팅 파워 비교



상대자	Fan Hui(유럽 챔피언)	이세돌
시합일자	2015.10.05 ~10.09	2016. 03 09 ~ 15
대국수	5 (5:0 알파고 승)	5 (4:1 알파고 승)
트레이닝	GPU : 50개	GPU : 100개
경기	CPU : 1202개 GPU : 176개	CPU : 1920개 GPU : 280개

KGS 6 ~ 9단의 160,000경기로 부터 29,400,000위치 정보를 뽑아서 학습

<http://www.economist.com/news/science-and-technology/21694540-win-or-lose-best-five-battle-contest-another-milestone>

# 알파고(AlphaGo)의 구조와 성능

구분	탐색 쓰레드	CPUs	GPUs	Elo rating *
단일(Single)	40	48	8	2890
분산	40	1202	176	3140

탐색 쓰레드	쓰레드 개수만큼 바둑의 경기 경로를 탐색
CPU의 역할	CPU 한 개당 1초에 1000회 이상의 시뮬레이션 (Fast rollout방법)
GPU의 역할	딥러닝을 사용하여 바둑판 상태의 승률과 다음 착수 예측

\* Elo rating : 게임 플레이어 간의 실력을 상대적으로 나타낸 것으로, 단일/분산 AlphaGo의 Elo rating을 측정하기 위하여 기존 인공지능 프로그램과 대결하여 산출

# 알파고(AlphaGo)의 구조와 성능

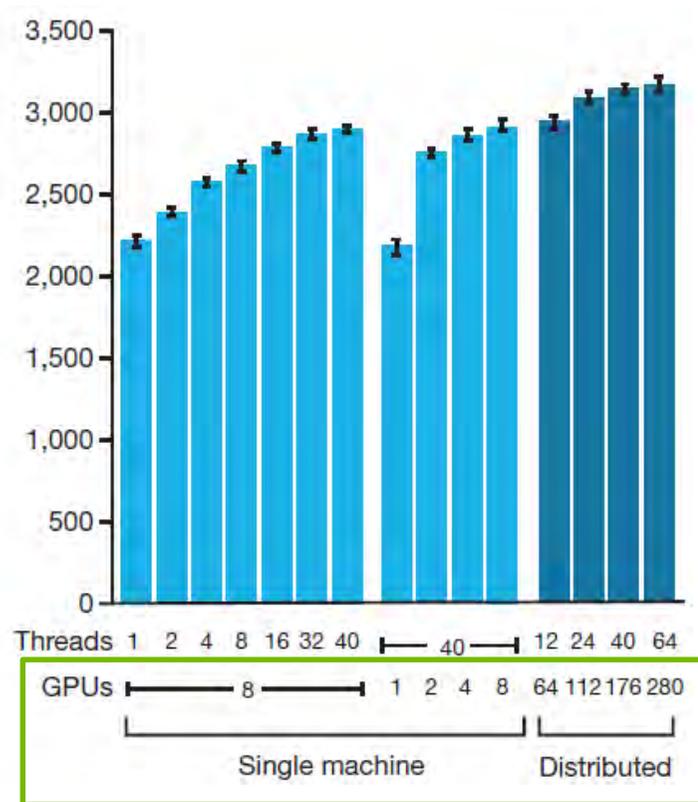
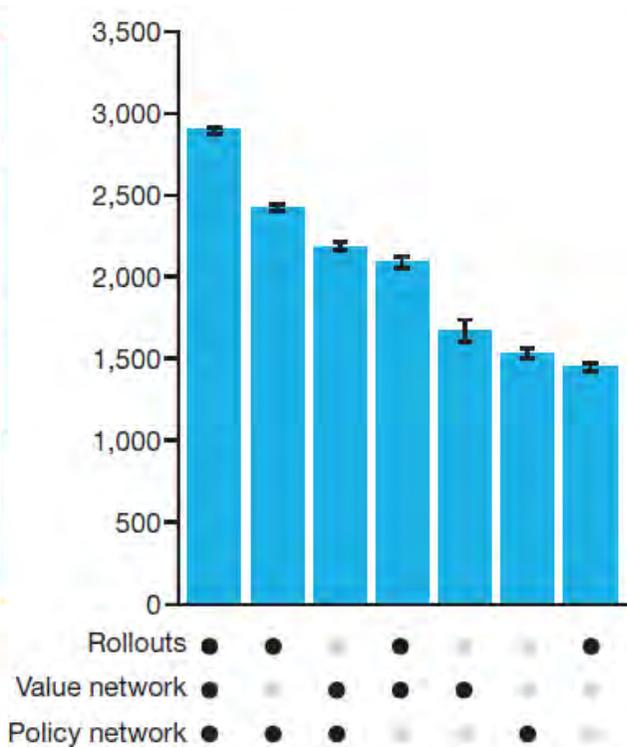
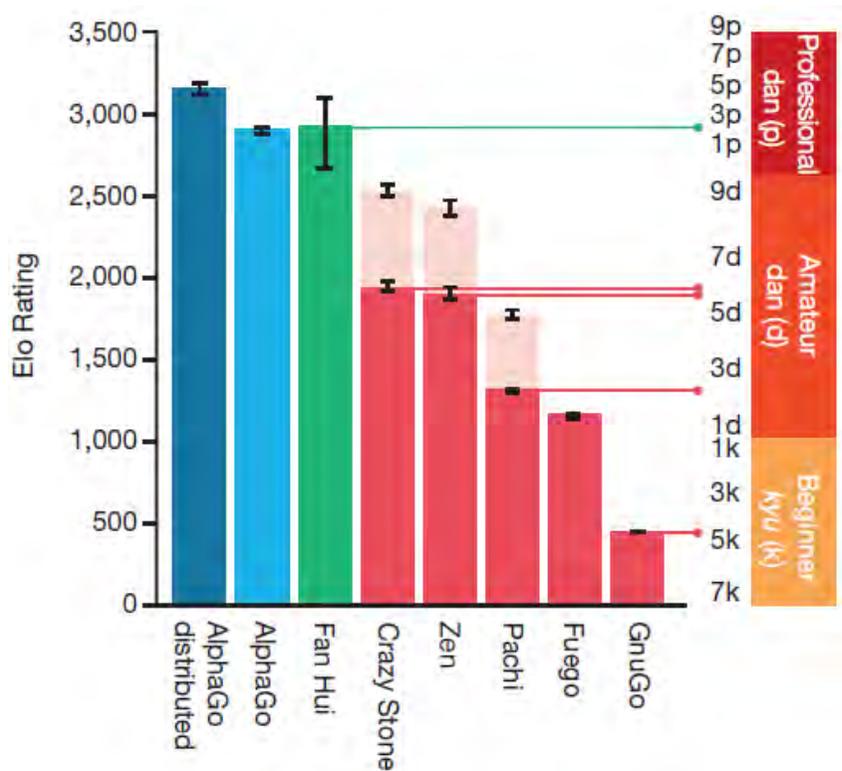
인공지능 바둑 프로그램	버전	착수 시간 설정	CPUs	GPUs	KGS 랭크	Elo rating
분산 AlphaGo		5초	1202	176	-	3140
AlphaGo		5초	48	8	-	2890
CrazyStone	2015	5초	32	-	6d	1929
Zen	5	5초	8	-	6d	1888
Pachi	10.99	5초(400,000sims)	16	-	2d	1298
Fuego	svn1989	5초(100,000sims)	16	-	-	1148
GnuGo	3.8	5초(level10)	1	-	5k	431
CrazyStone <sub>4</sub>	4접바둑	5초	32	-	-	2526
Zen <sub>4</sub>	4접바둑	5초	8	-	-	2413
Pachi <sub>4</sub>	4접바둑	5초(400,000sims)	16	-	-	1756

<http://spri.kr/post/14725>

프로(p), 아마추어 단(d), 아마추어 급(k)

<http://www.nature.com/news/google-ai-algorithm-masters-ancient-game-of-go-1.19234>

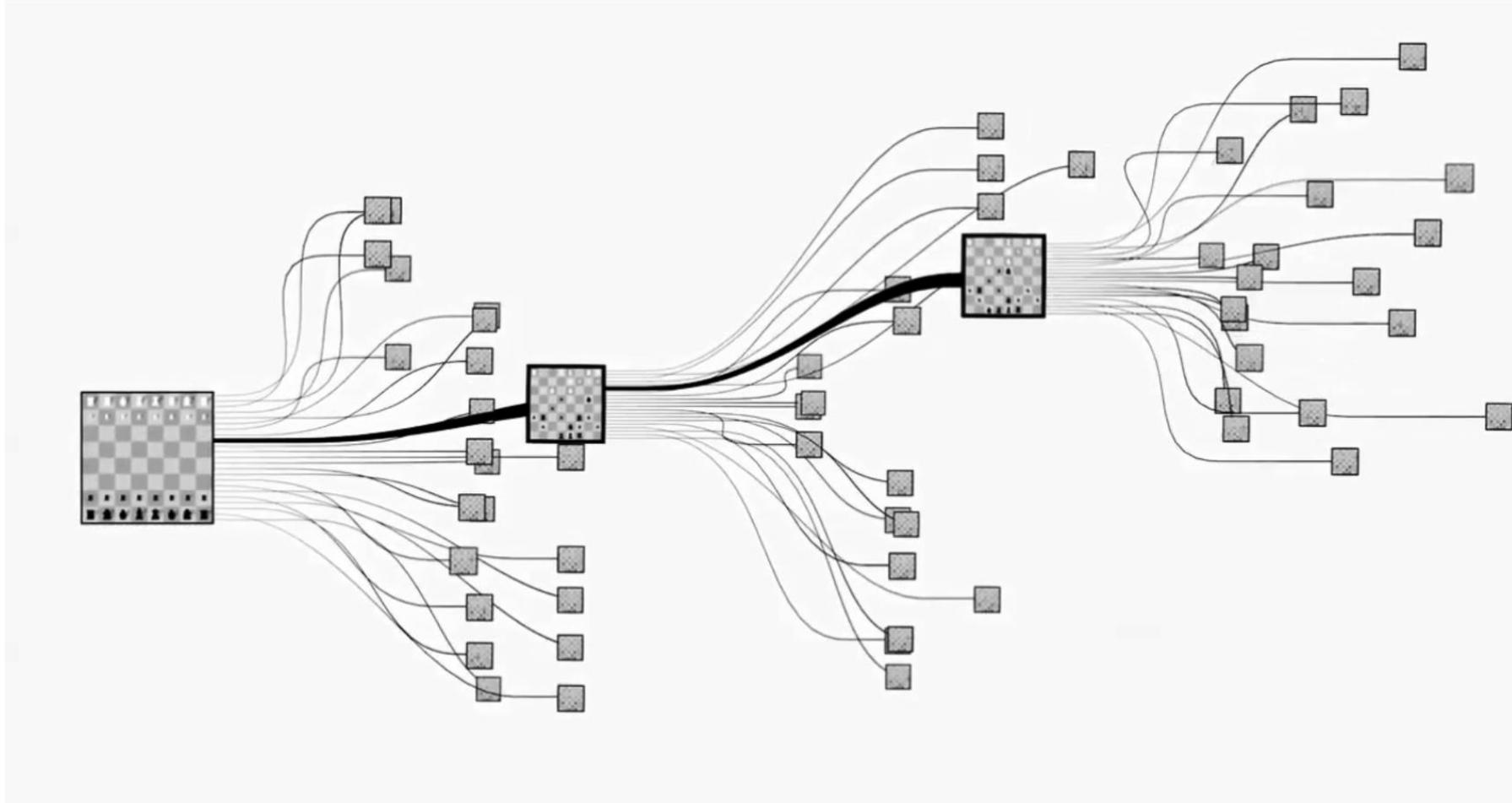
# 알파고 (AlphaGo) 수준



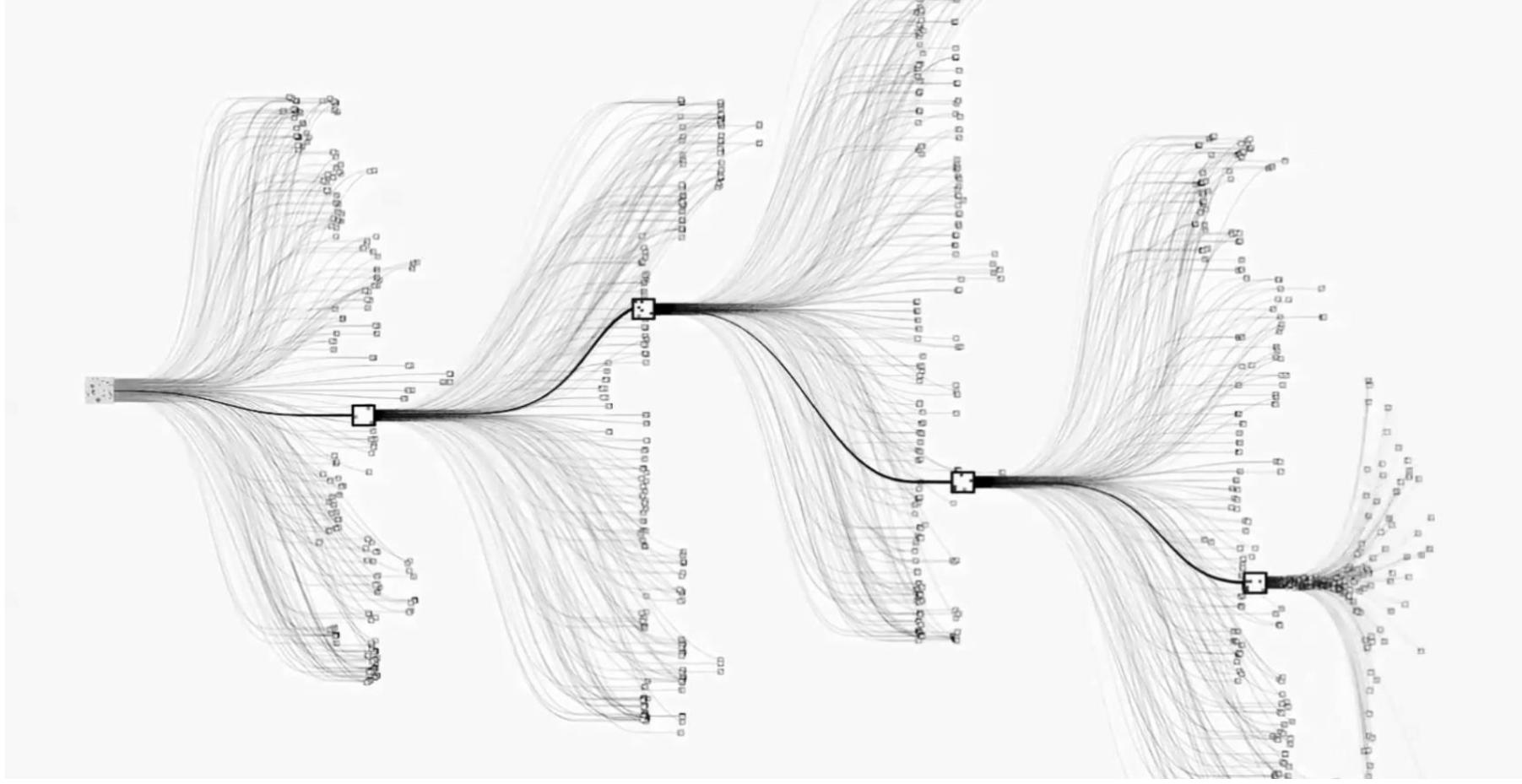
# AlphaGo가 바둑을 선택한 이유

- $10^{360}$  가지의 경우 수 존재  
바둑 규칙을 고려한 평균 경우 수 250개, 바둑 평균 약 150수 정도,  $250^{150} \approx 10^{360}$   
우주의 원자수 :  $10^{80}$
- 현존하는 컴퓨팅 파워로는 경우 수를 모두 파악하기 어려움
- 경우의 수(트리 탐색)를 줄이는 방법이 관건

# 체스 탐색 트리



# 바둑 탐색 트리



# AlphaGo 트레이닝 딥러닝

Human expert Data  
(100K games)



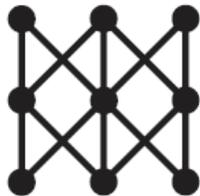
# AlphaGo 트레이닝 딥러닝

Human expert Data  
(100K games)



Classification

Supervised Learning  
Policy network



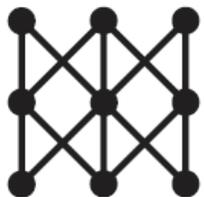
# AlphaGo 트레이닝 딥러닝

Human expert Data  
(100K games)



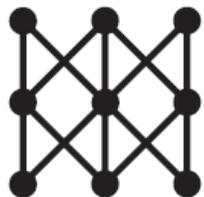
Classification

Supervised Learning  
Policy network



Self Play

Reinforcement Learning  
Policy network



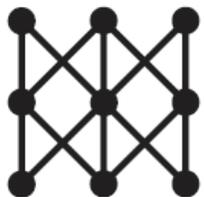
# AlphaGo 트레이닝 딥러닝

Human expert Data  
(100K games)



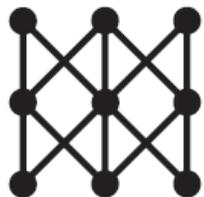
Classification

Supervised Learning  
Policy network



Self Play

Reinforcement Learning  
Policy network



Self Play

Generates New Data  
(30M Positions)



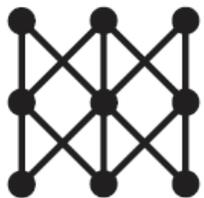
# AlphaGo 트레이닝 딥러닝

Human expert Data  
(100K games)



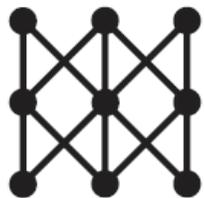
Classification

Supervised Learning  
Policy network



Self Play

Reinforcement Learning  
Policy network



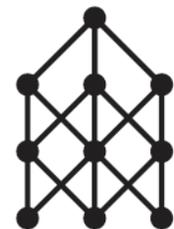
Self Play

Generates New Data  
(30M Positions)



Regression

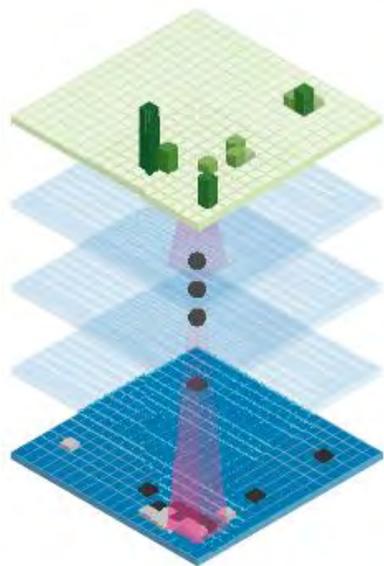
Value network



# AlphaGo 트레이닝 딥러닝

## Policy and Value Networks

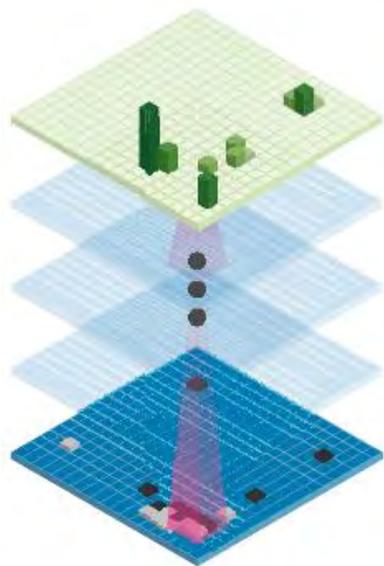
Policy Network  
(Probability Distribution over moves)



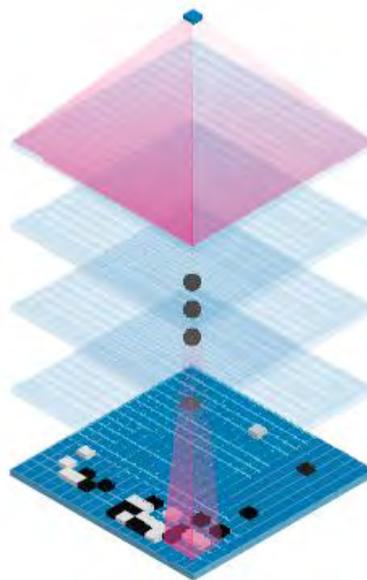
# AlphaGo 트레이닝 딥러닝

## Policy and Value Networks

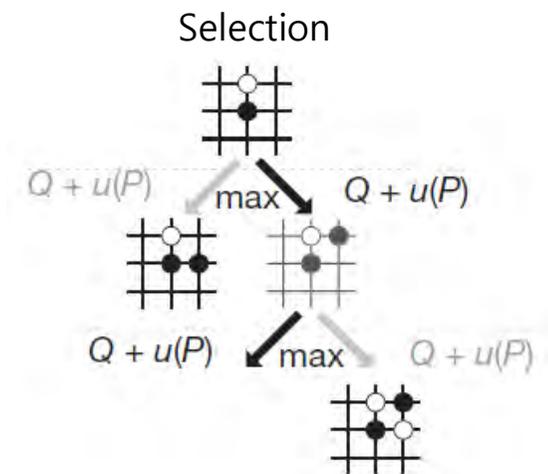
Policy Network  
(Probability Distribution over moves)



Value Network  
(Real numbers : 0 White, -1 Black)



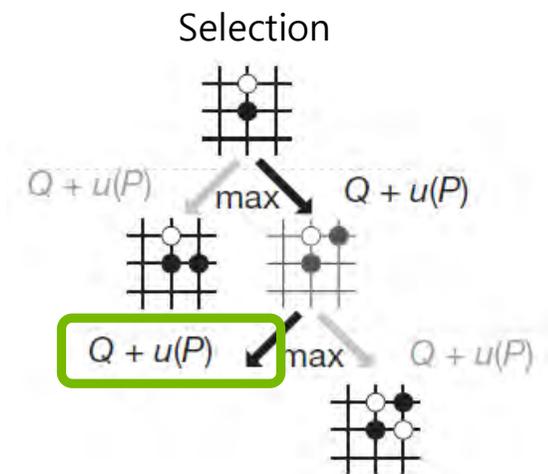
# AlphaGo Tree search and Rollouts



$Q$  : Action Value of Move

$P$  : Prior Probability of Move

# AlphaGo Tree search and Rollouts

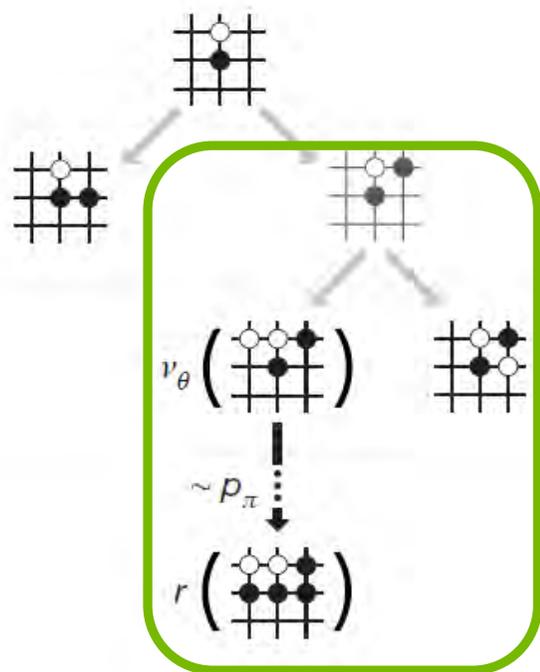


$Q$  : Action Value of Move

$P$  : Prior Probability of Move

# AlphaGo Tree search and Rollouts

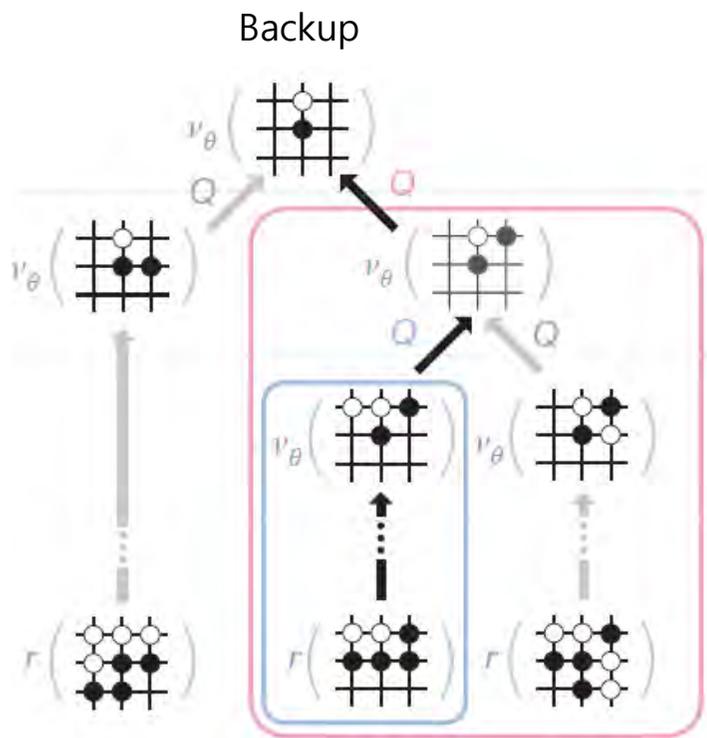
Evaluation



Q : Action Value of Move

P : Prior Probability of Move

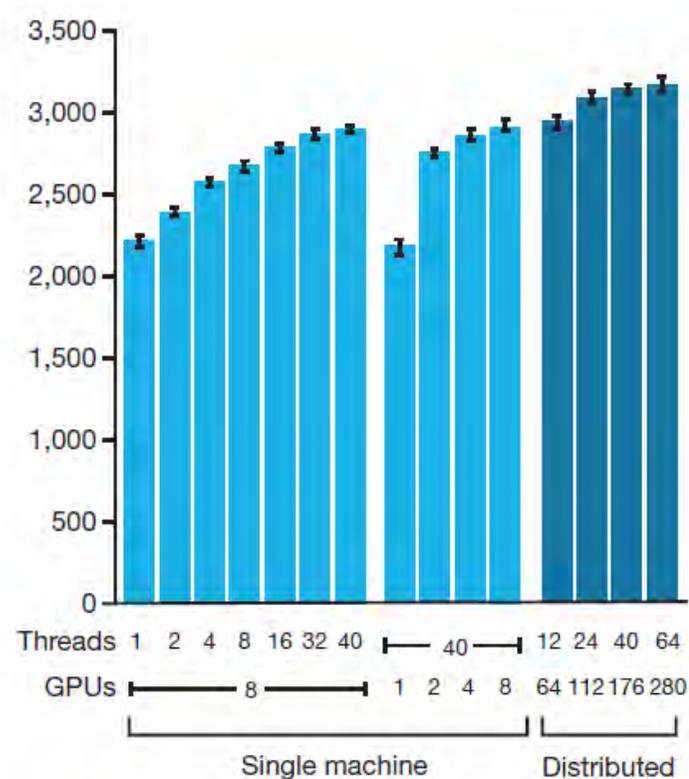
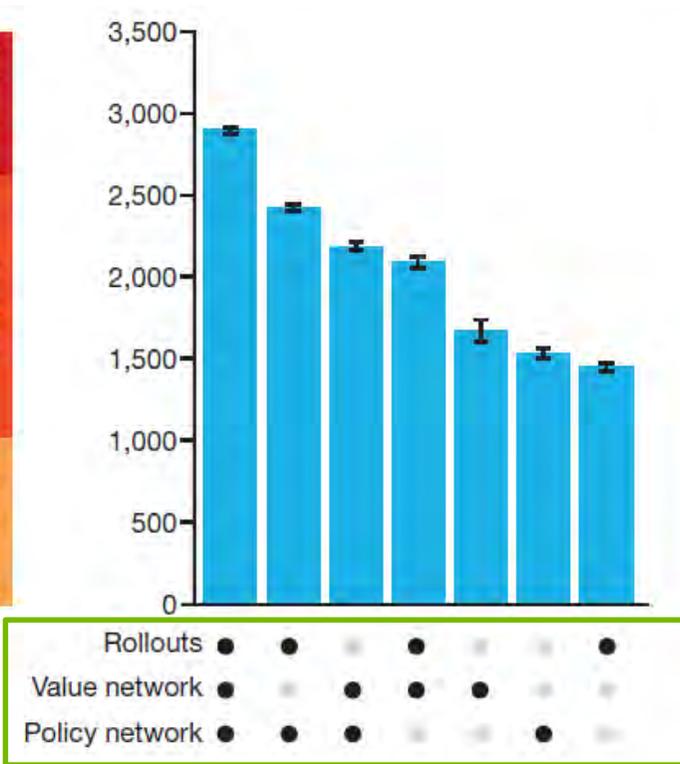
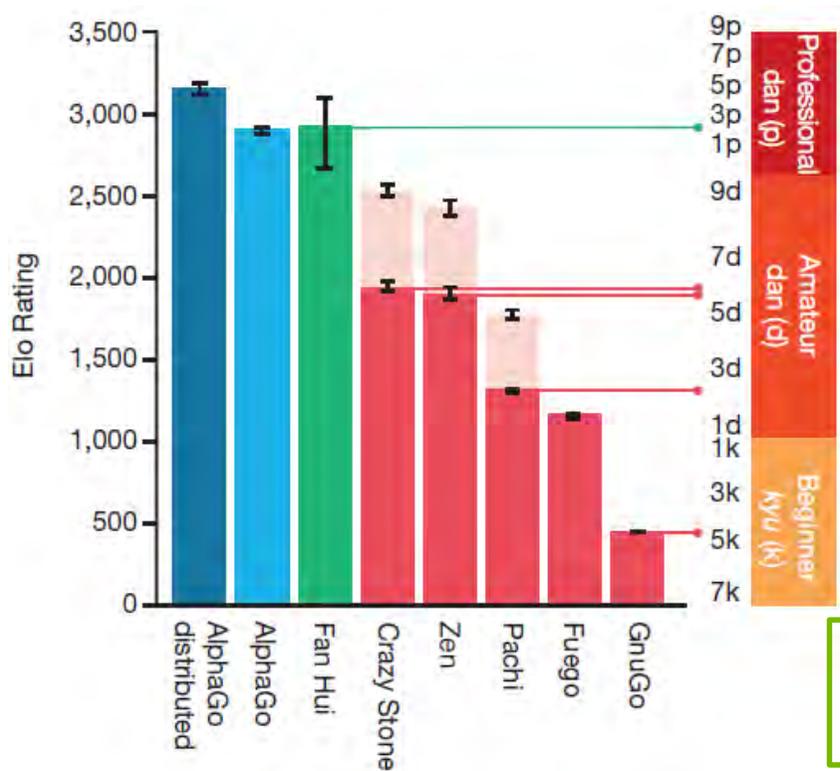
# AlphaGo Tree search and Rollouts



$Q$  : Action Value of Move

$P$  : Prior Probability of Move

# 알파고 (AlphaGo) 수준



# MACHINE INTELLIGENCE 2.0

## AGENTS

AGENTS			AUTONOMOUS SYSTEMS			
<b>PROFESSIONAL</b> Howdy! x.ai @clara KASIST DigitalGenius OVERLAP.CC meekan fuse machines PRIMER	<b>PERSONAL</b> facebook XIAOICE assistant ai large nestor @awesome Magic	<b>OS INTERFACES</b> Siri Cortana VIV moluuba api.ai COGNEA Google Now	<b>AIR</b> SDR DJI PROJECT LOON VERTICAL DroneDeploy AIRDOG SKYCATCH SKYDIO Airware LILY	<b>GROUND</b> Google UBER TESLA CRUISE Mobileye COMMA AdasWorks	<b>SEA</b> LIQUID ROBOTICS bluefin data OPENRV BluHaptics	<b>INDUSTRIAL</b> KIVA Systems fetch HARVEST CLEARPATH AVIDBOTS ENERGI rethink robotics GREYORANGE OSARO

## ENTERPRISE

SECURITY / FRAUD	HR / RECRUITING	SALES	MARKETING	CUSTOMER SUPPORT	INTERNAL INTEL	MARKET INTEL
Sentinel graphistry SIGHT feedzai AREA1 drawbridge sift science CYLANCE Brighterion	textio hi gild SpringRole entelo unitive GIGSTER	sense clari infer Preact Gainsight AVISO salespredict sentient Vidora people Prism	LiftIgniter RADIUS brightfunnel retention AIRPR	CLARABRIDGE QUANTIFIND Wiseio ACTIONIO FRAMED DigitalGenius	Alation ADATAQ Palantir sapho lucid Rainbird SKIPFLAG egolo Digital Reasoning Narrative Science	Quid mattermark Datafox bottlenose PREMISE enigma CB INSIGHTS

## PLATFORMS

RESEARCH / AGI	FULL STACK	MACHINE LEARNING	INDUSTRIAL IOT	AUDIO	VISION	DATA ENRICHMENT
OpenAI vicarious Google DeepMind Numenta Cyclic nnsense SCALED INFERENCE CURIOUS GEOMETRIC INTELLIGENCE	context relevant CognitiveScale NVIDIA TERADEEP QUALCOMM nervana	Data rapio/miner cortico.io AYASDI amazon Azure nrologics PredictionIO SKYTREE bigm blueyonder	ThingWorx UPTAKE IMUBIT Preferred Networks Alluvium xively PLANET OS	Gridspace Talkio nexidia vocaliq NUANCE Expect Labs popUP archive	ORBITAL INSIGHT Descartes Labs DEXTR0 cortica clarifai MetaMind PLANET	diffbot Pavats TRIFACTA IDIBON WorkFusion loop CrowdFlower

## INDUSTRIES

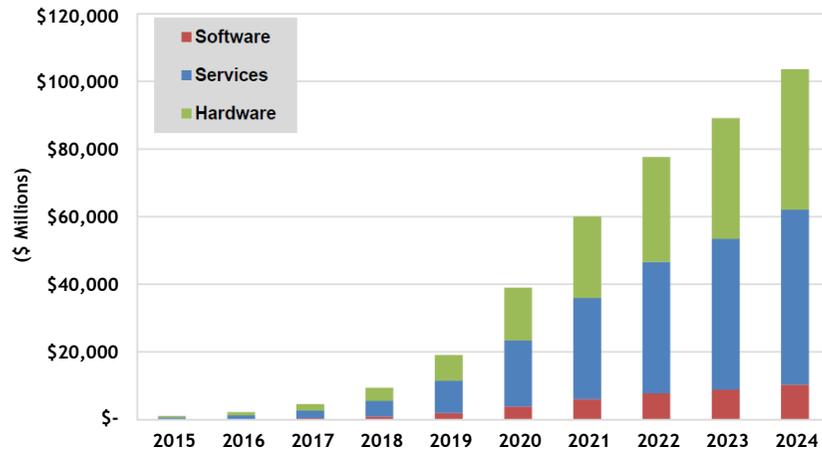
ADTECH	AGRICULTURE	FOR GOOD	RETAIL FINANCE	LEGAL	MATERIALS & MFG	HEALTHCARE
Rothement distillery BEYONDERBAL METAMARKETS TAPPD rocketfuel affectiva	BLUE RIVER tule TerraVision mavrx THE CLIMATE CORPORATION CERES HONEYCOMB	Conservation Metrics DataKind thorn BAYES IMPACT	inVenture Affirm earnest MIRADOR Lendo zen finance LendUp	Everlaw RAVEL LEGAL ROBOT Sedi BEAGLE R+SS Lex Machina	zymergen AUGMATE GINKGO BIOWORKS TRINE SIGHT MACHINE TECHNOLOGIES CALCULARIO Egen Innovations	deep genomics 3SCAN enlitic Calico Atomwise Recombine color METABIOTA GRANO ROUNDS Google Life Sciences Watson Health

## INDUSTRIES (CONT'D)

EDUCATION	TRANSPORT & LOGISTICS	INVESTMENT FINANCE	DATA SCIENCE	MACHINE LEARNING	OPEN SOURCE
KNEWTON coursera turnitin gradscope UDACITY KHANACADEMY	NAUTO taleris PRETECKT clearmetal	Bloomberg Quantopian Dotaminr KENSHC ISENTIUM NEURENSIC alphasense	DOMINO kaggle Sentend sense yseep Outlier yhat DataRobot	Cortana Analytics AlchemyAPI glowflsh Watson Platform Anodot MonkeyLearn (h [s]) HyperScience fuzzy.io SIGOPT Oxdata.ho SPARKBEYOND indico	SKYJND TensorFlow seldon Caffe theano Spark MLIB Microsoft PM spaCy DL4J SciKit CGT

# \$500B OPPORTUNITY OVER 10 YRS

**Deep Learning Total Revenue by Segment**  
World Markets: 2015-2024



**Deep Learning Software Revenue by Industry**  
World Markets: 2015-2024



Source : "Deep Learning for Enterprise Applications", 4Q 2015, Tractica



Question : [sbyun@nvidia.com](mailto:sbyun@nvidia.com)

