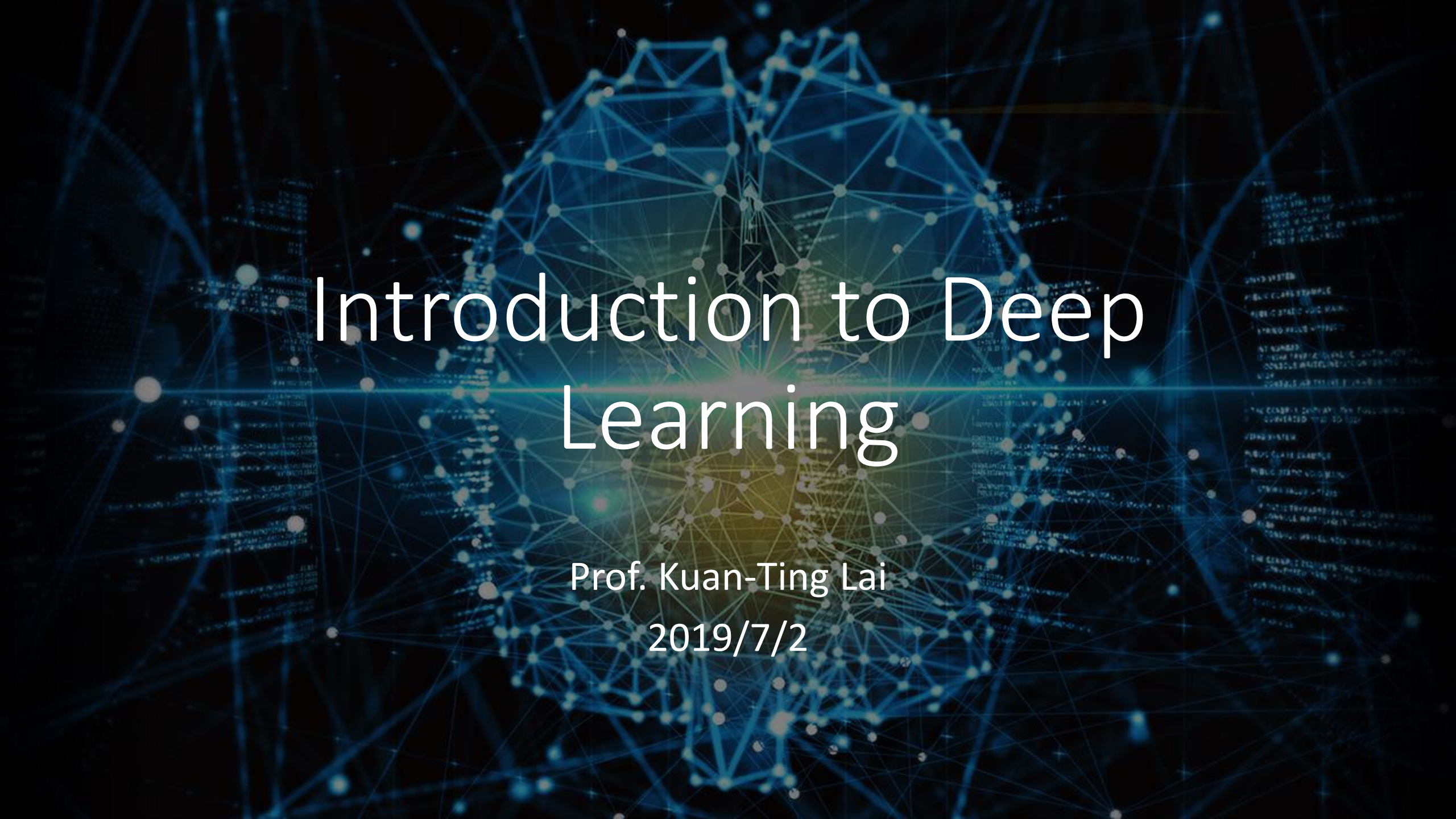
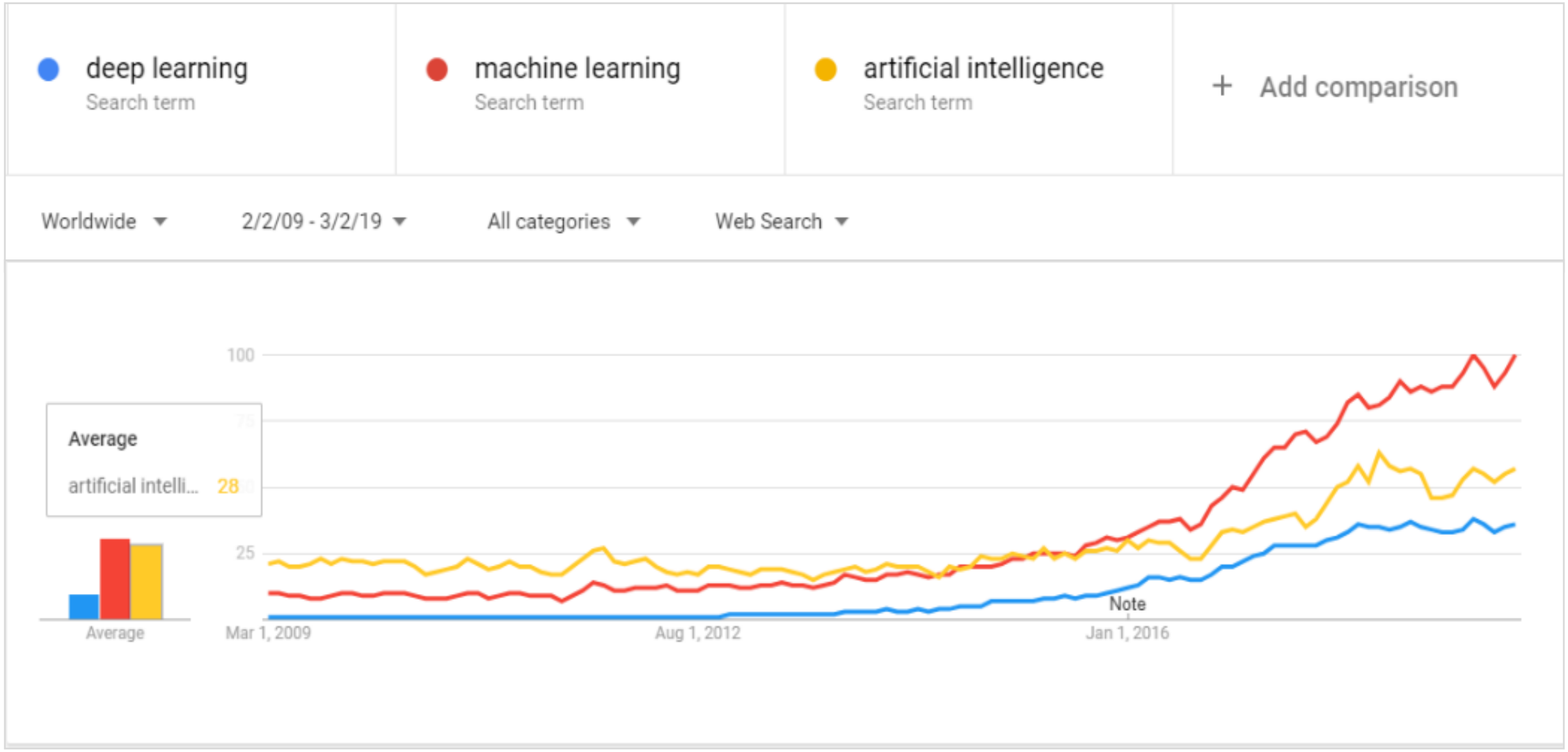


Introduction to Deep Learning

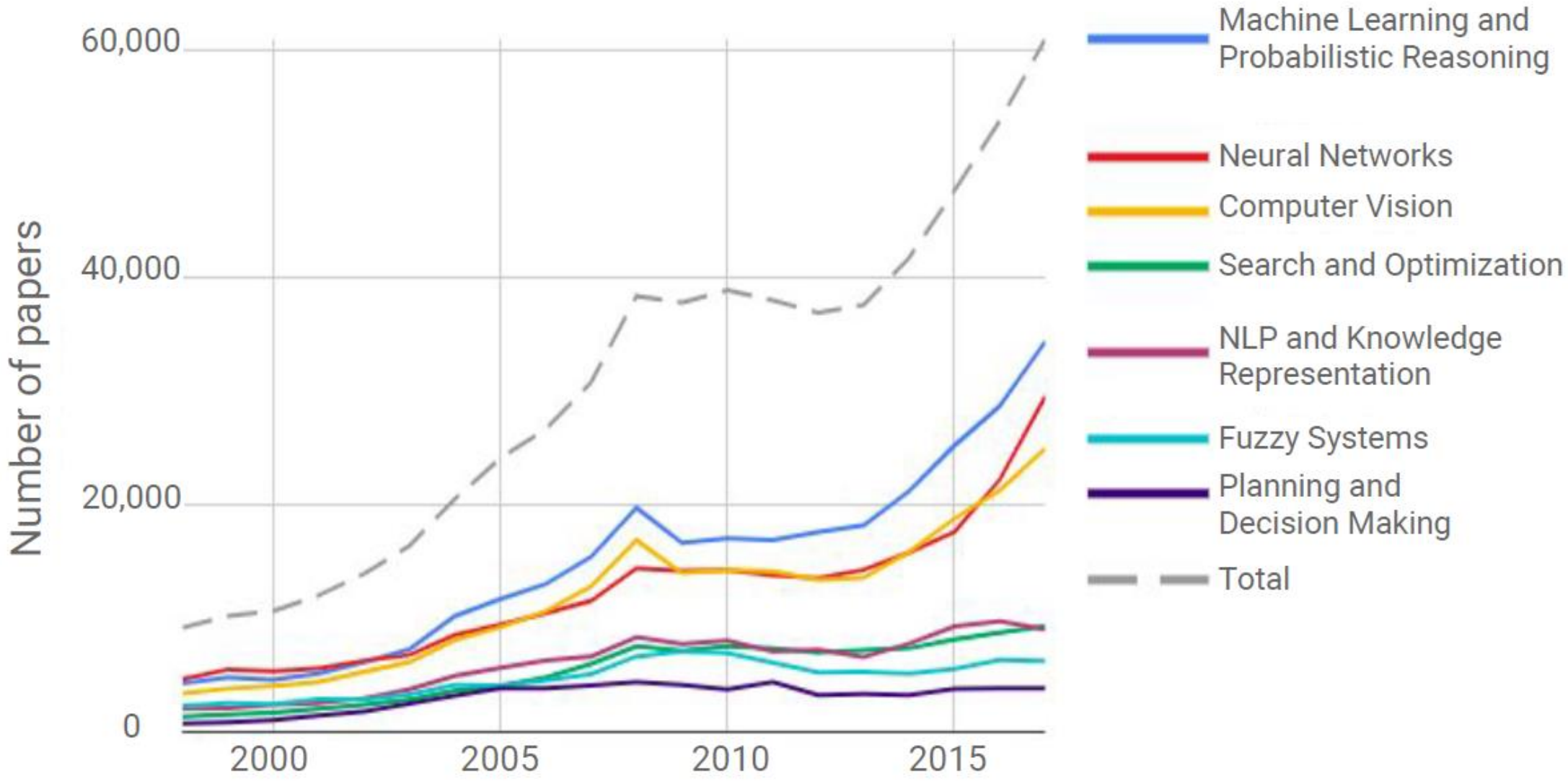
Prof. Kuan-Ting Lai

2019/7/2

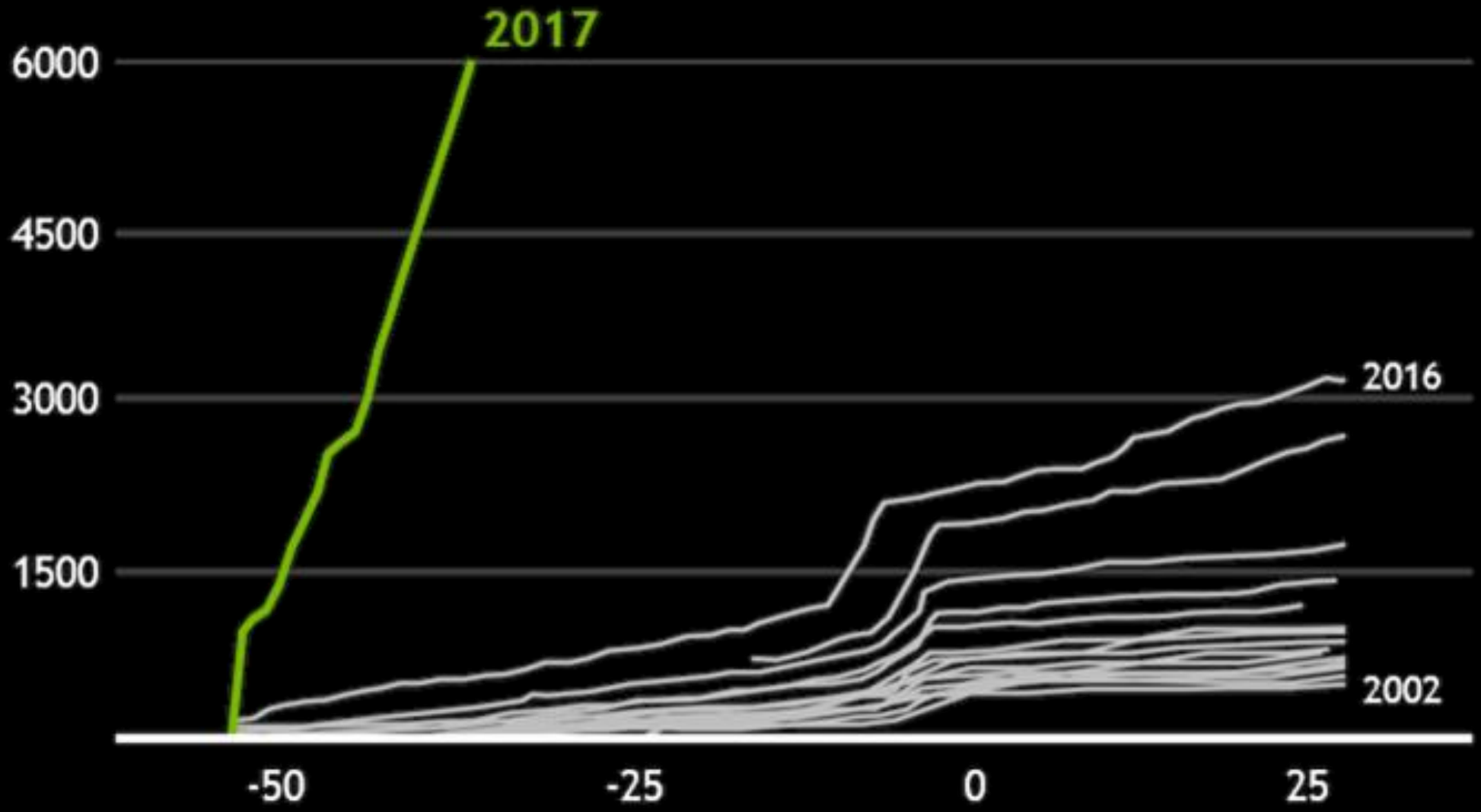
Deep Learning – a new Buzzword



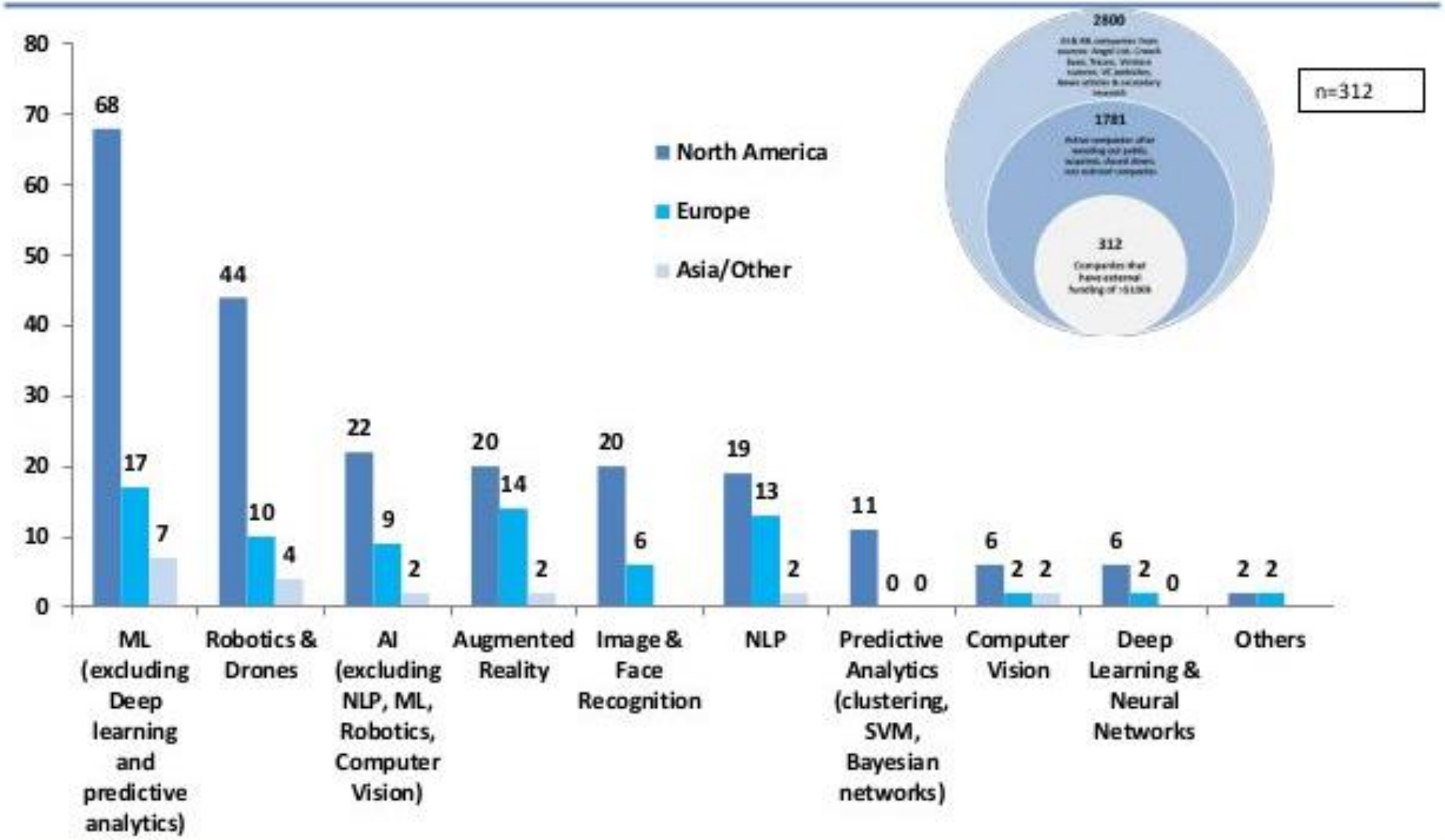
AI Papers


60,000
40,000
20,000
0
Number of papers
2000
2005
2010
2015
Machine Learning and Probabilistic Reasoning
Neural Networks
Computer Vision
Search and Optimization
NLP and Knowledge Representation
Fuzzy Systems
Planning and Decision Making
Total

Registration of NIPS

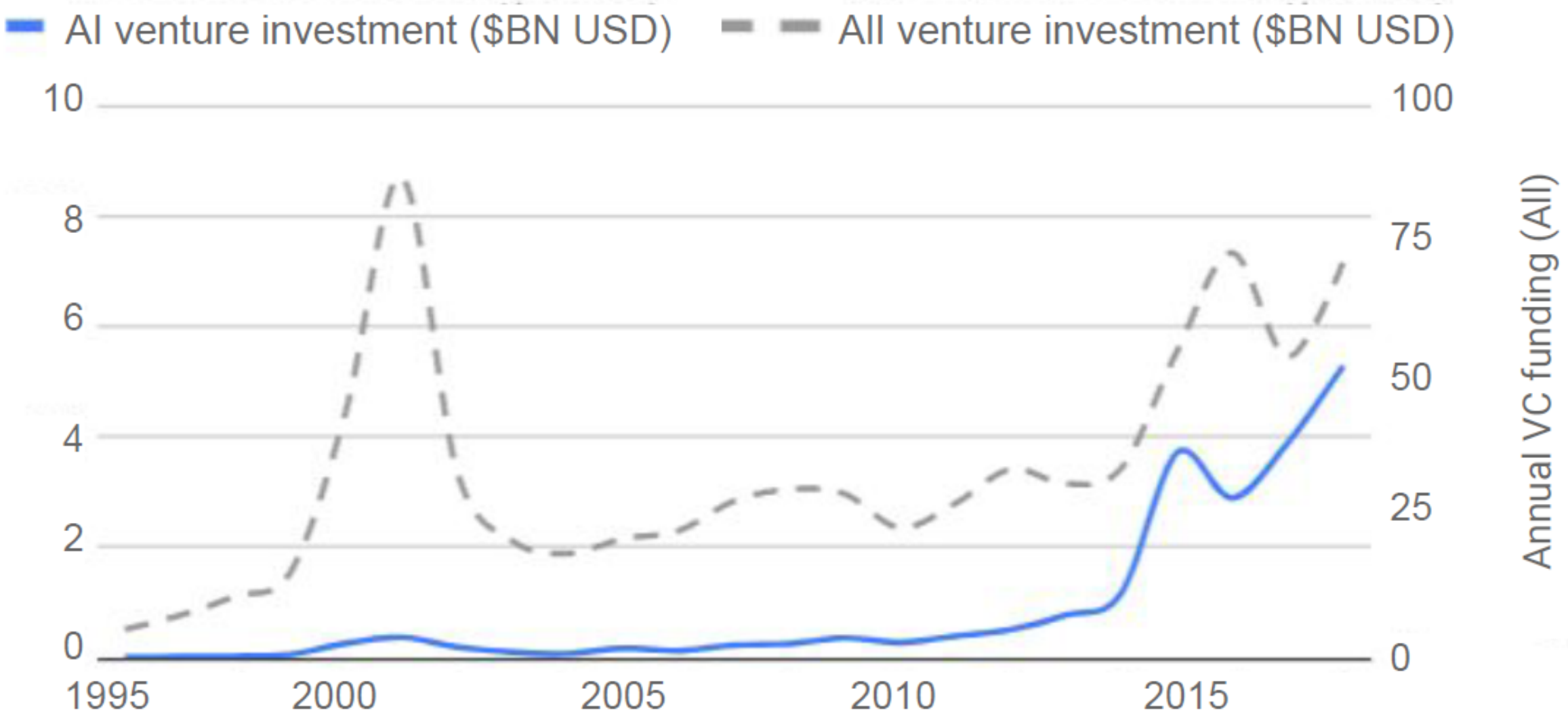


Major Fields for AI / ML Venture Investment by Geography



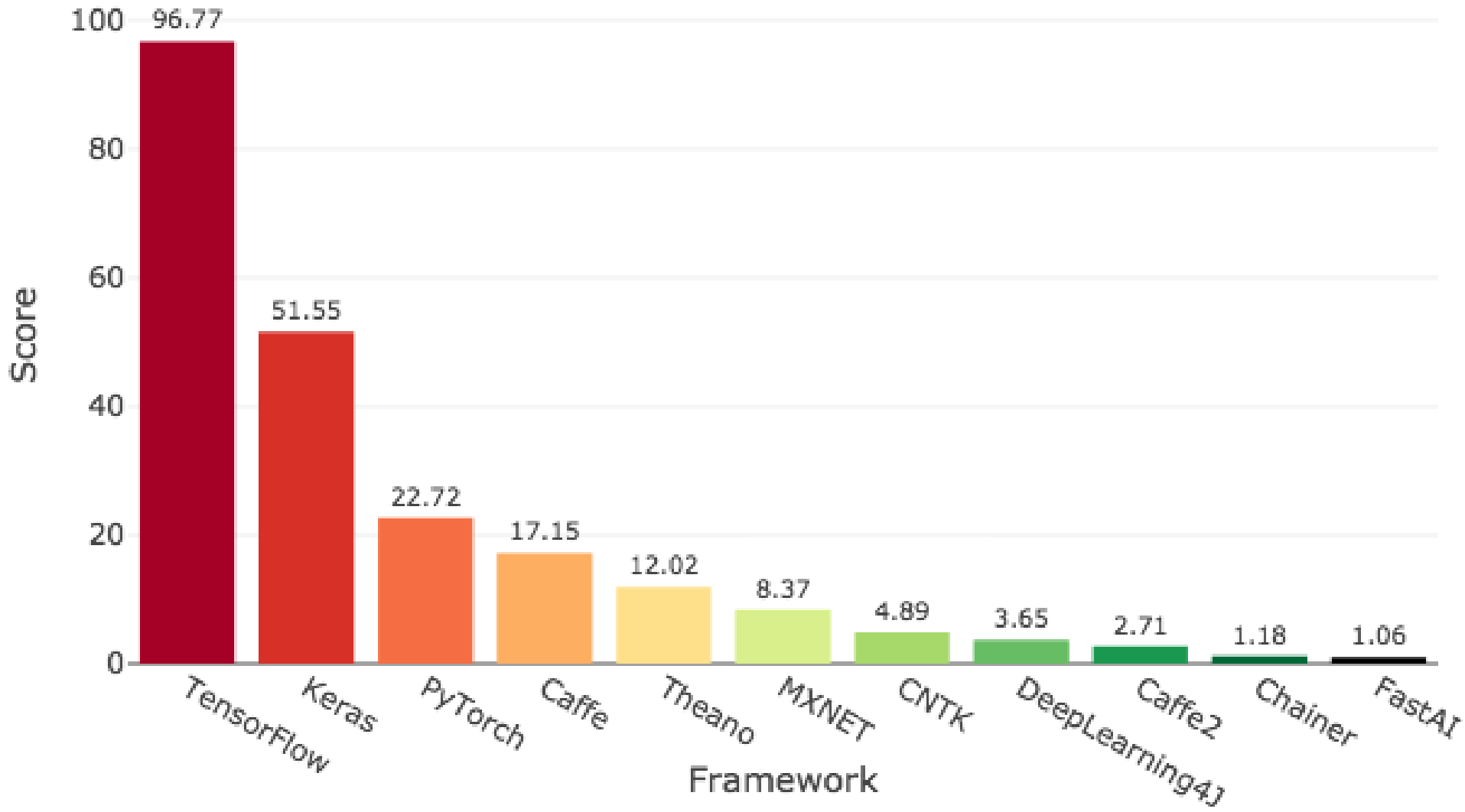
North America is leading the way but Europe has significant presence in AI & AR areas

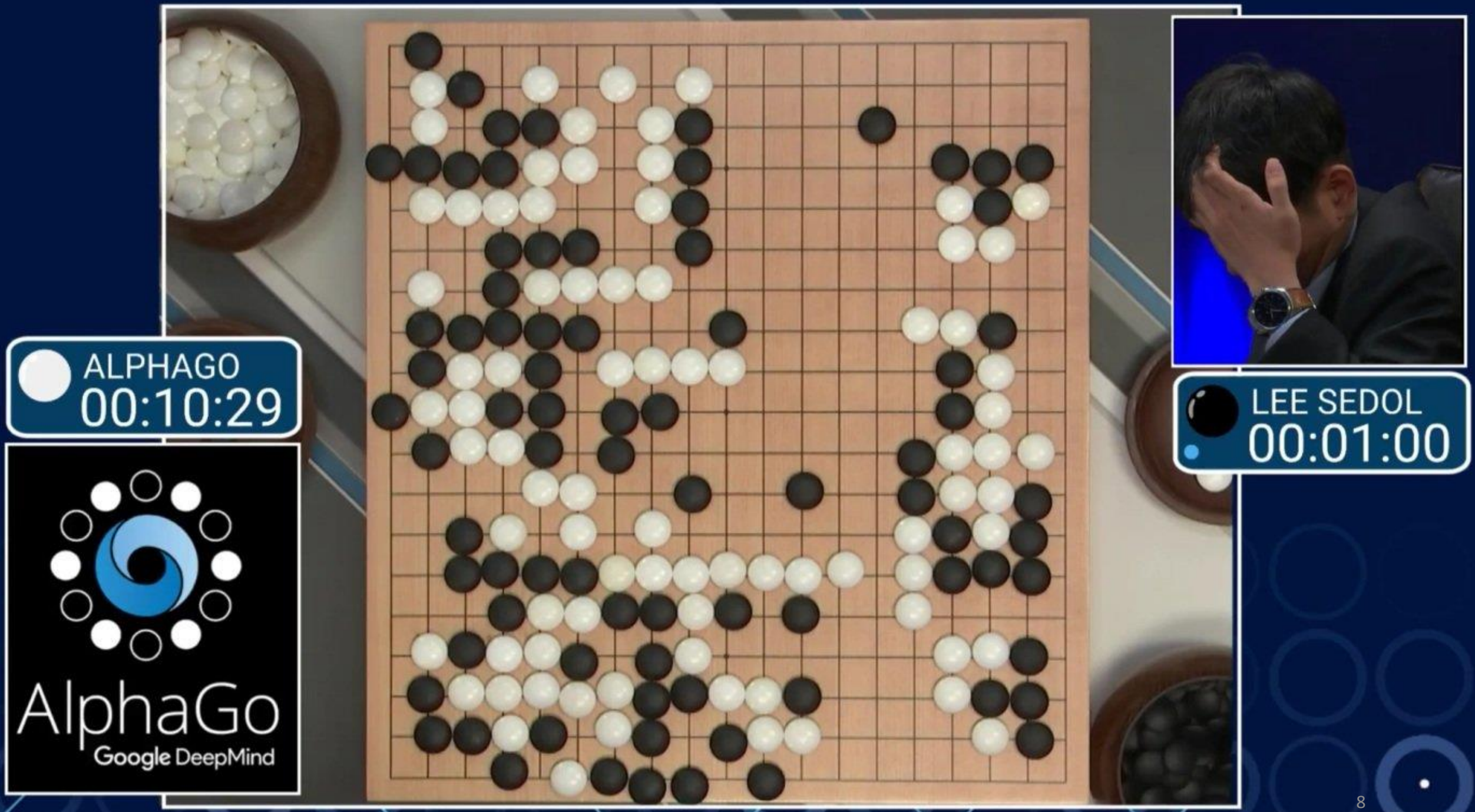
Source: Sand Hill Econometrics



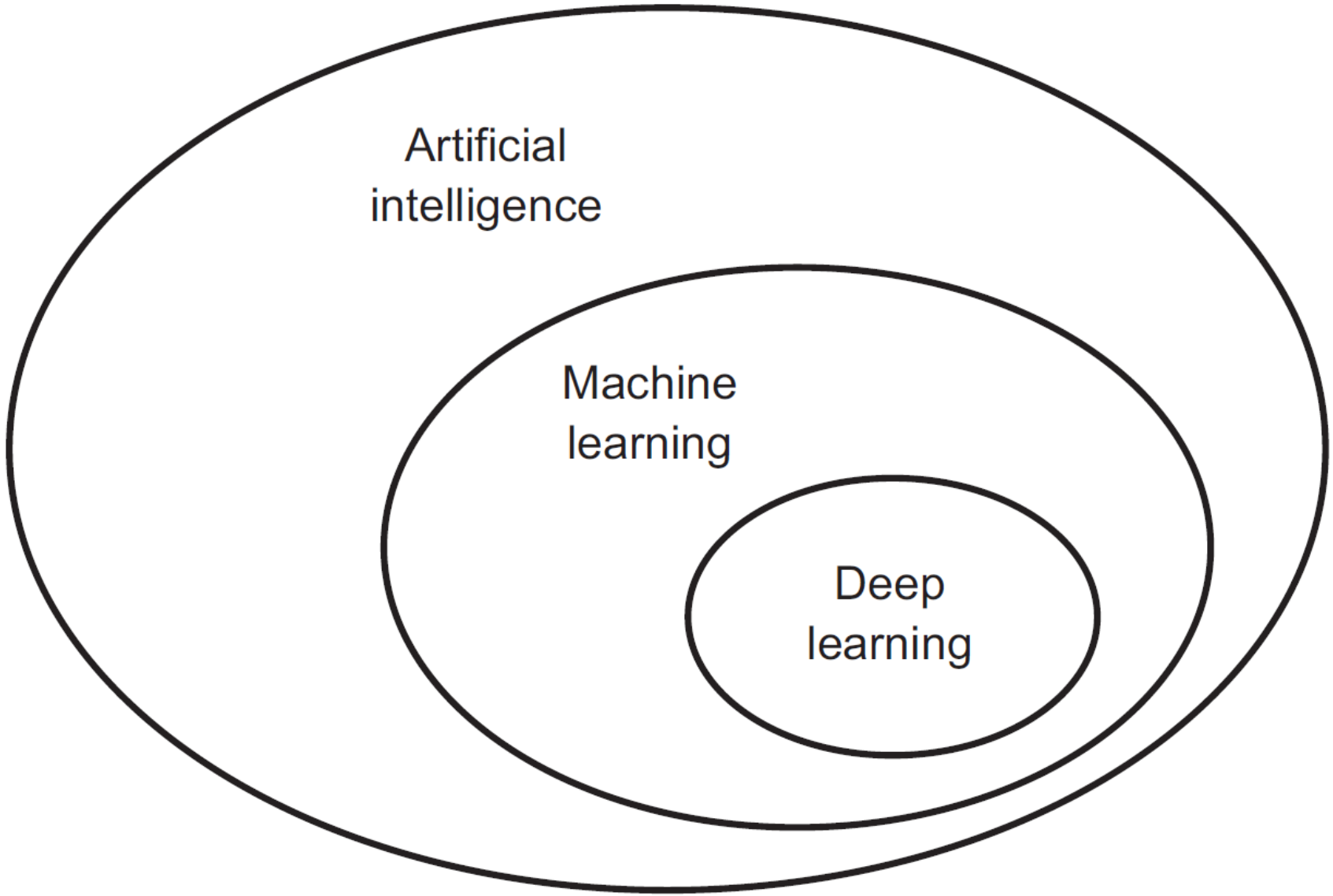
Source: Sand Hill Econometrics

Deep Learning Framework Power Scores 2018

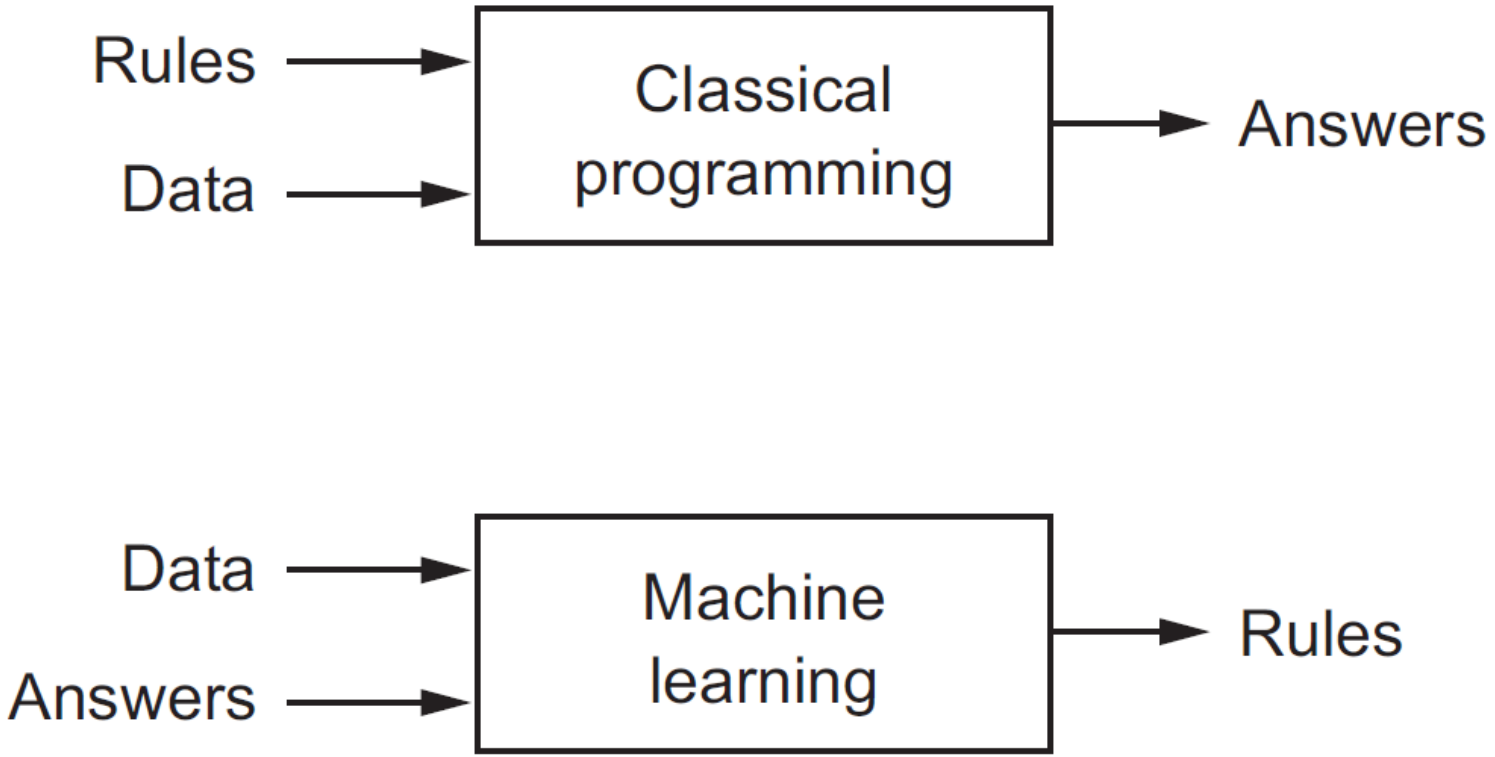


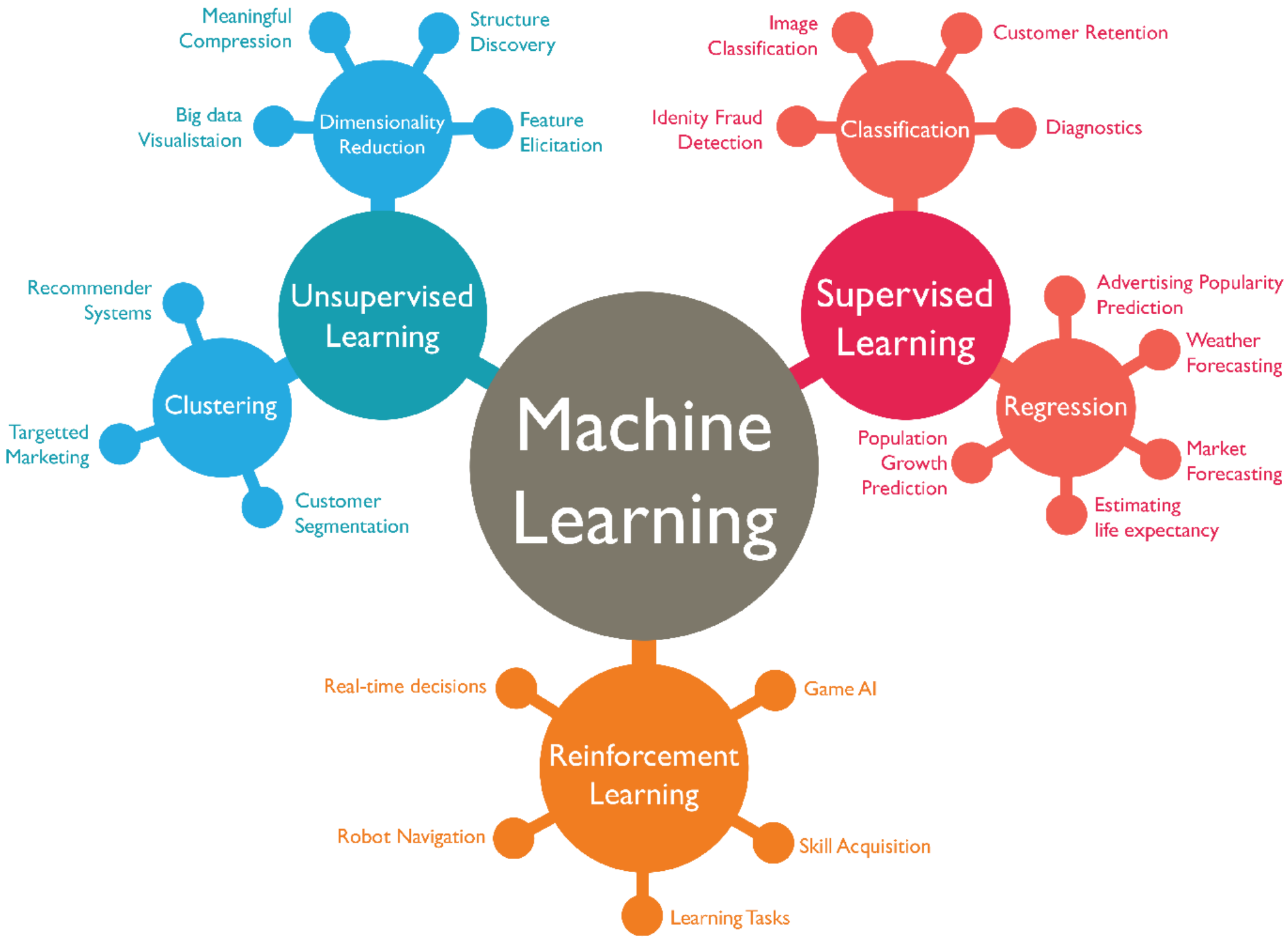

ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

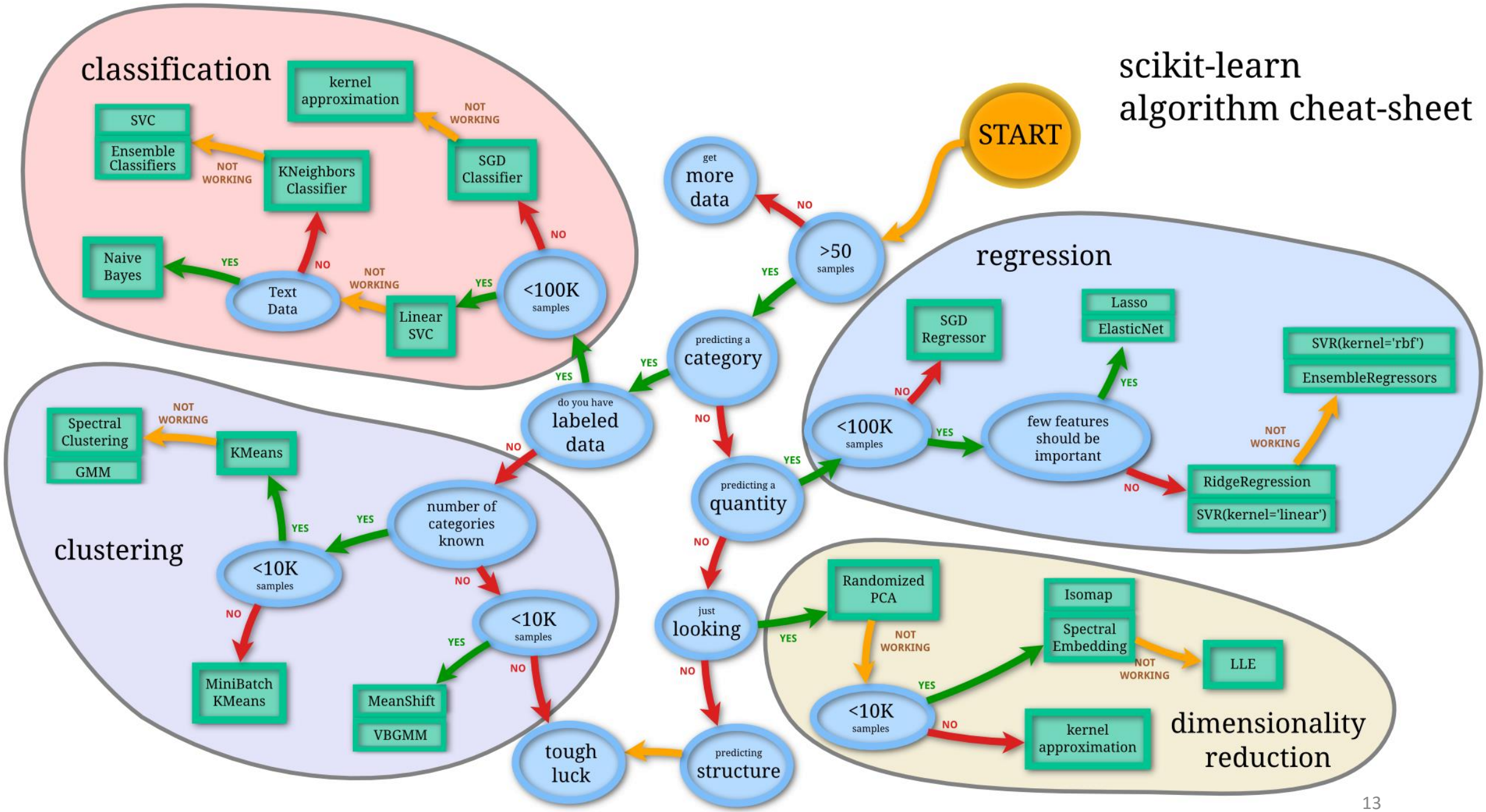
So, what is Deep Learning?


Artificial
intelligence
Machine
learning
Deep
learning

Machine Learning



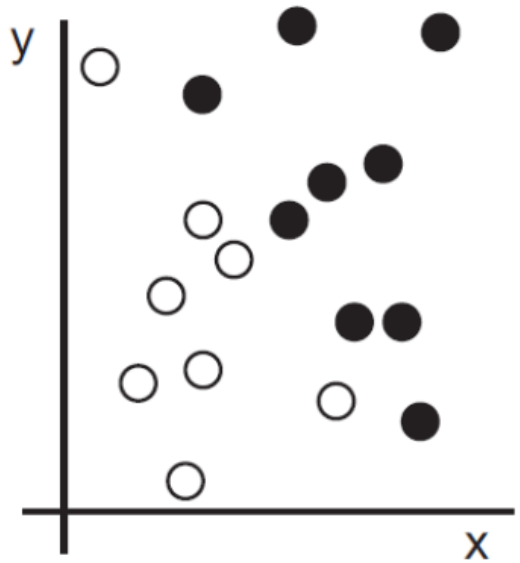

Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Dimensionality Reduction
Clustering
Classification
Regression
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Recommender Systems
Targetted Marketing
Customer Segmentation
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks


scikit-learn
algorithm cheat-sheet
START
>50 samples
NO
get more data
YES
predicting a category
YES
NO
do you have labeled data
YES
NO
<100K samples
NO
YES
Naive Bayes
YES
NO
Text Data
NOT WORKING
Linear SVC
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
KNeighbors Classifier
NOT WORKING
kernel approximation
NO
SGD Classifier
classification
predicting a quantity
YES
NO
<100K samples
NO
YES
SGD Regressor
few features should be important
YES
Lasso
ElasticNet
NO
NOT WORKING
SVR(kernel='rbf')
EnsembleRegressors
RidgeRegression
SVR(kernel='linear')
regression
just looking
YES
Randomized PCA
NOT WORKING
<10K samples
YES
Isomap
Spectral Embedding
NOT WORKING
LLE
NO
kernel approximation
dimensionality reduction
NO
predicting structure
tough luck
number of categories known
YES
NO
<10K samples
YES
NO
MeanShift
VBGMM
<10K samples
YES
NO
MiniBatch KMeans
Spectral Clustering
GMM
NOT WORKING
KMeans
clustering

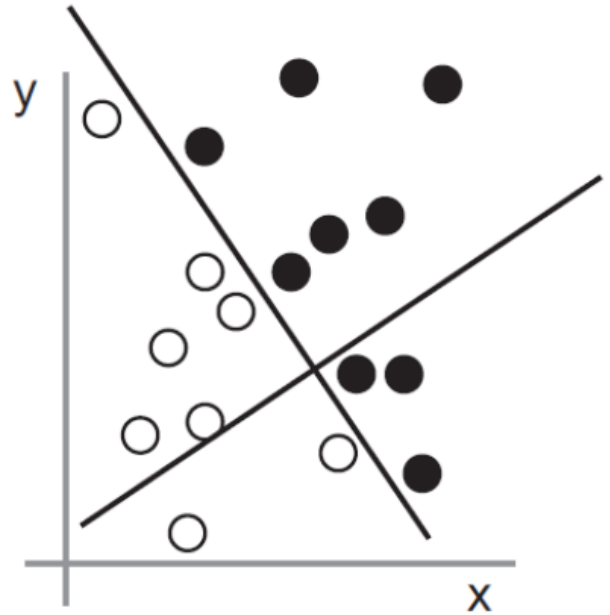
Learning Representation

- Objective: Classify white & black
- Input: (x, y)
- Output: Black or White

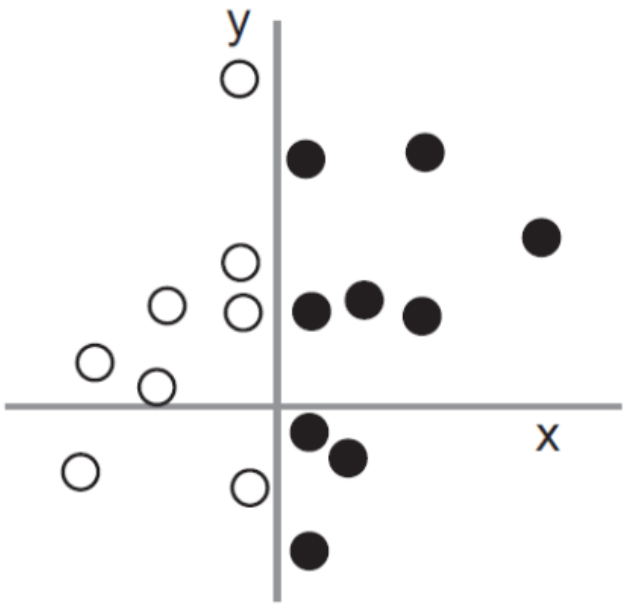
1: Raw data



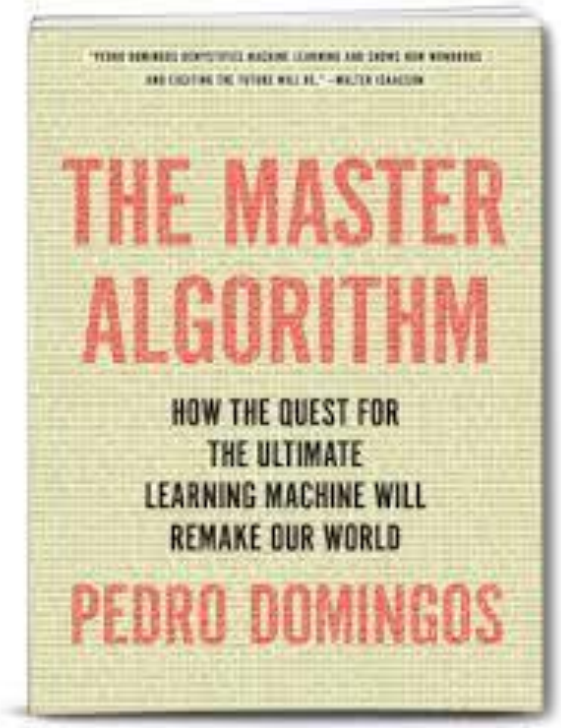
2: Coordinate change

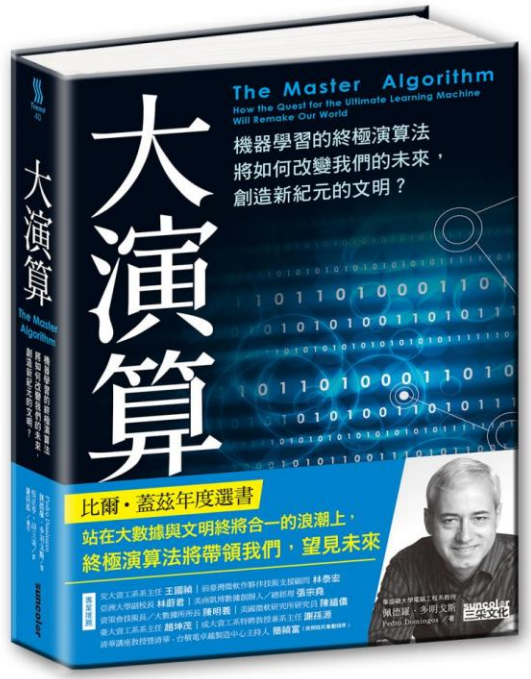


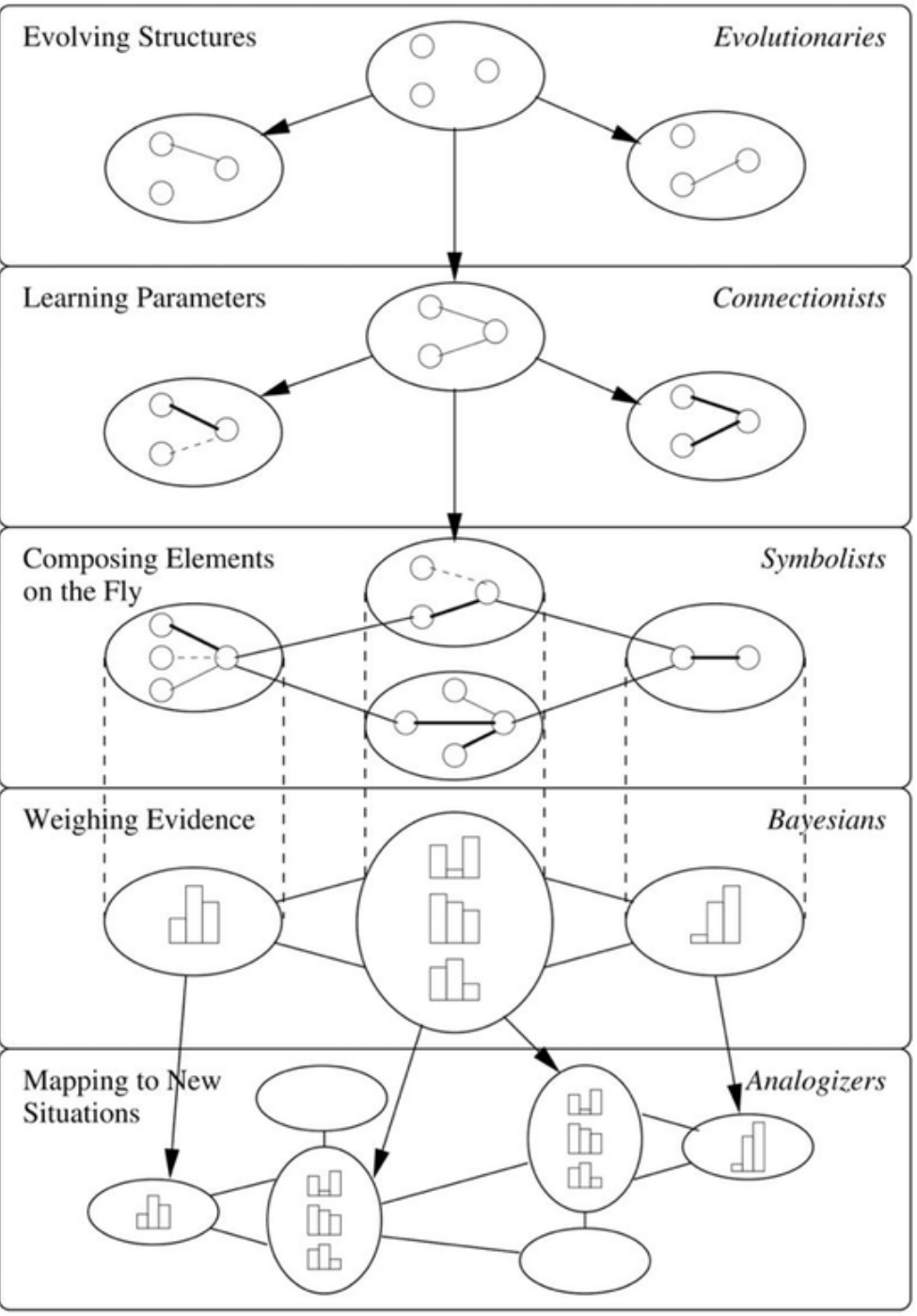
3: Better representation



The Master Algorithm – Pedro Domingos







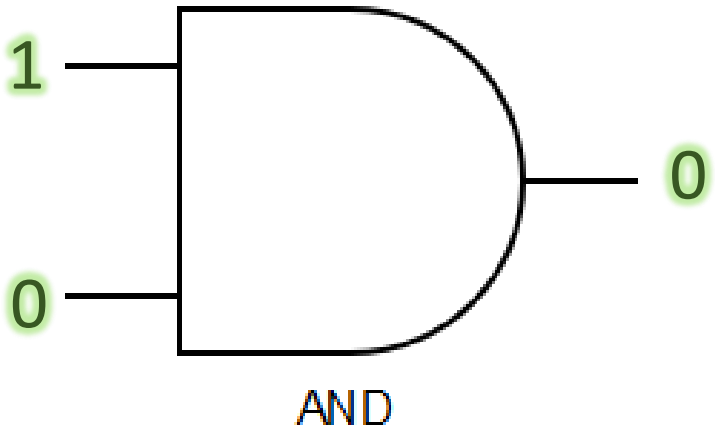
Five Tribes of Machine Learning

- Evolutionaries (基因演化法)
- Connectionists (類神經網路)
- Symbolists (歸納法)
- Bayesians (貝氏機率)
- Analogizers (類比近似)

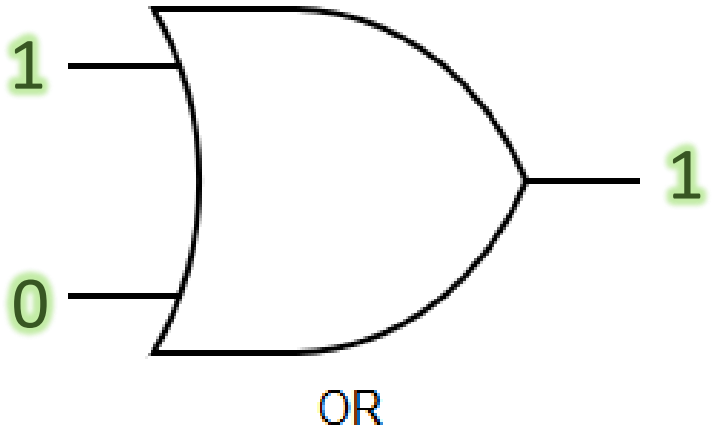
Five Tribes of Machine Learning

- Symbolists: Decision Trees, Random Forest
- Bayesians: Naïve Bayesians
- Analogizers: SVM, k-NN
- Evolutionaries: Gene algorithms
- Connectionists: Deep Learning

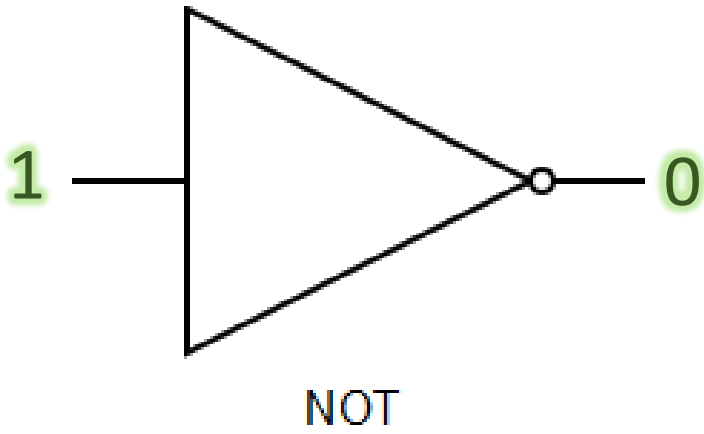
All Algorithms can be Reduced to 3 Operations



AND

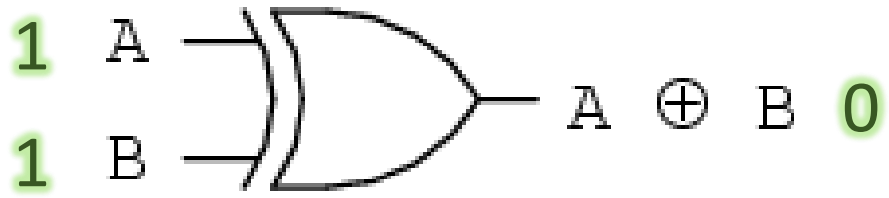


OR

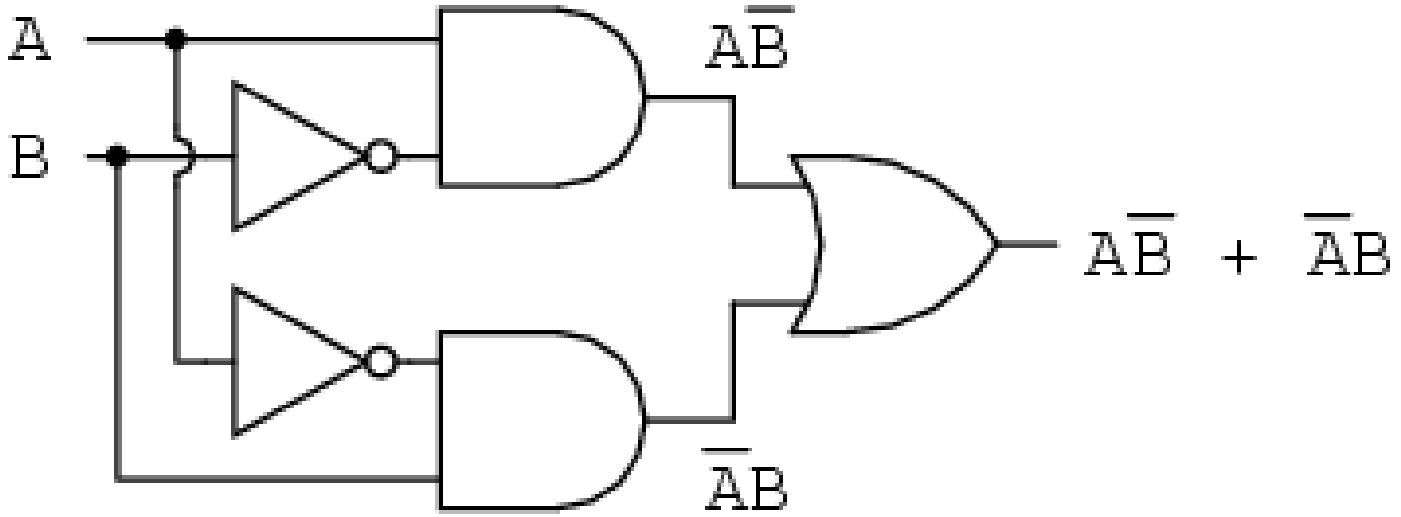


NOT

XOR

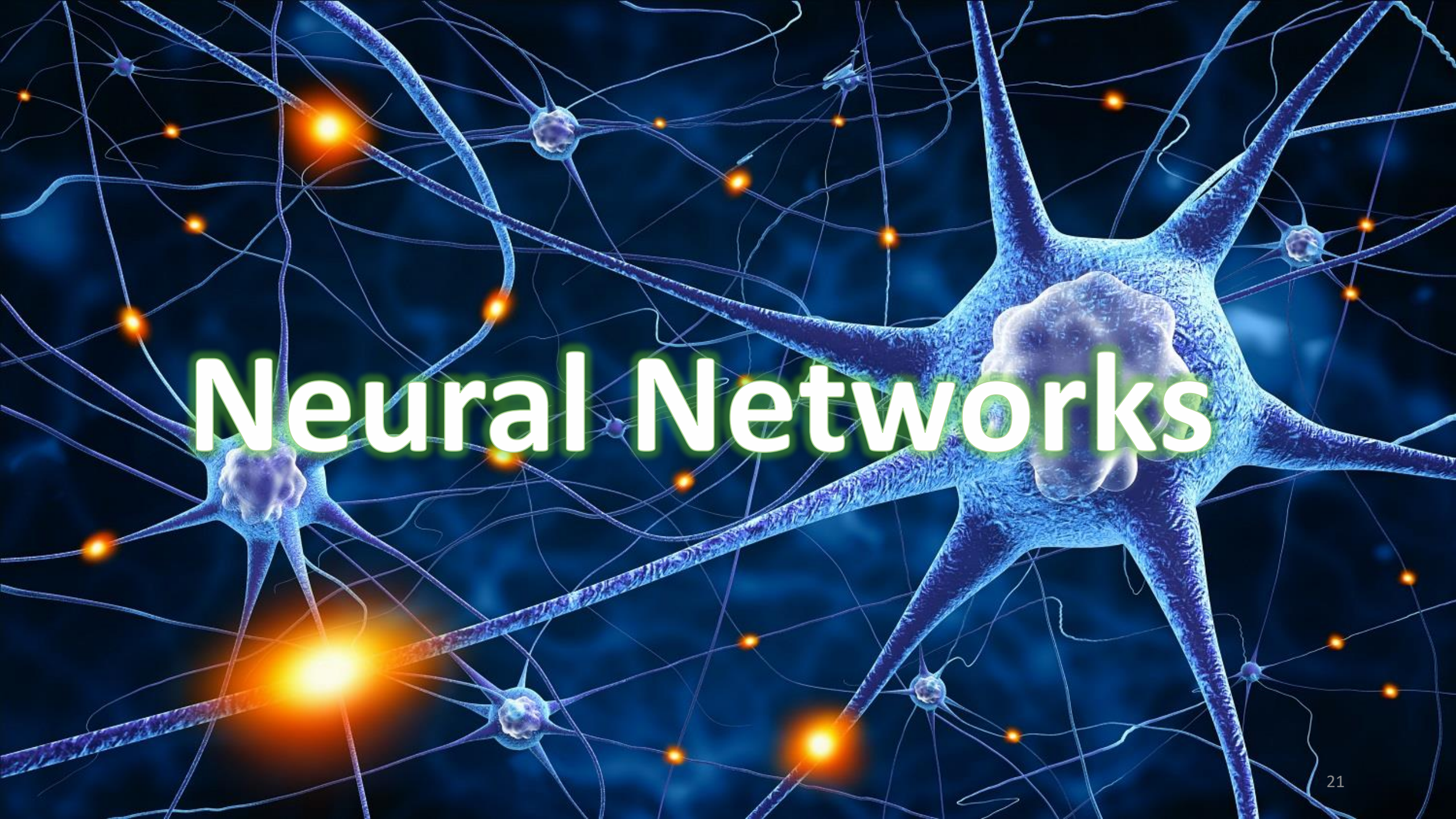


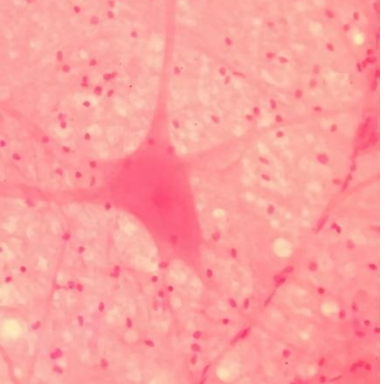
. . . is equivalent to . . .

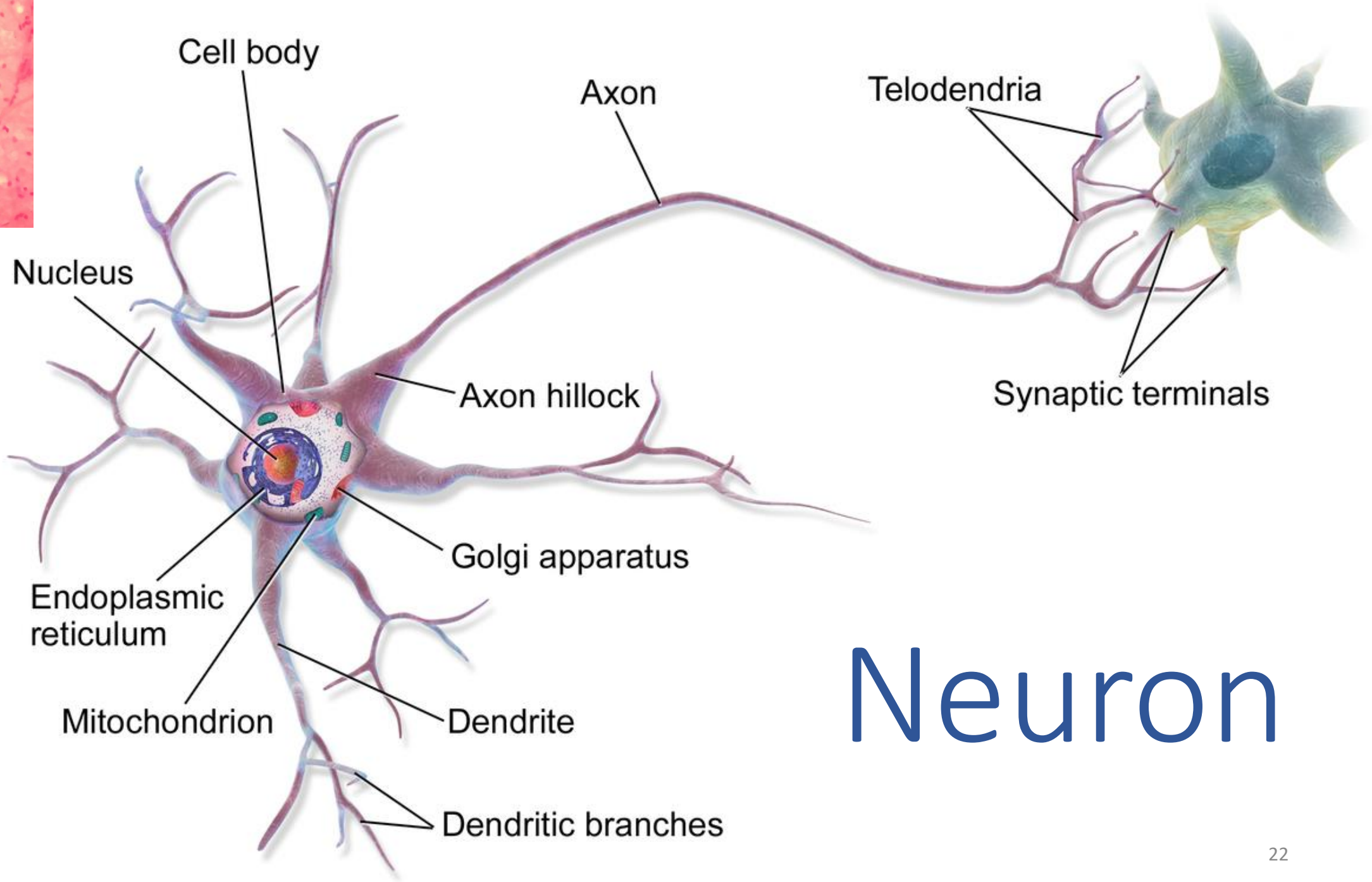


$$A \oplus B = A\overline{B} + \overline{A}B$$

OK, machine learning is cool. But what is Deep Learning?

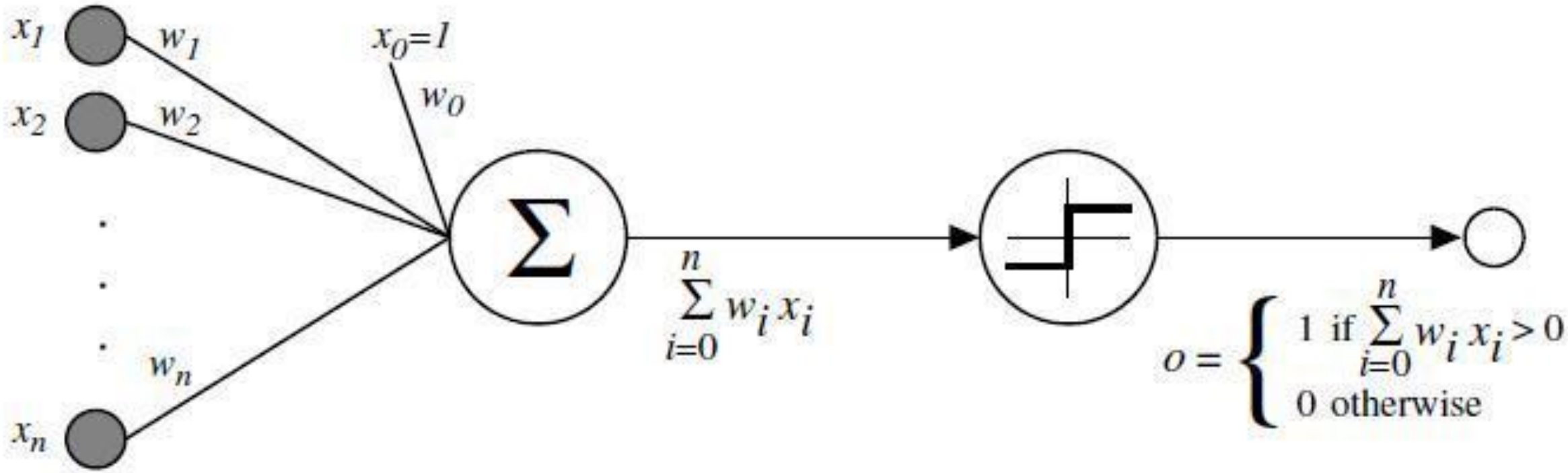

Neural Networks

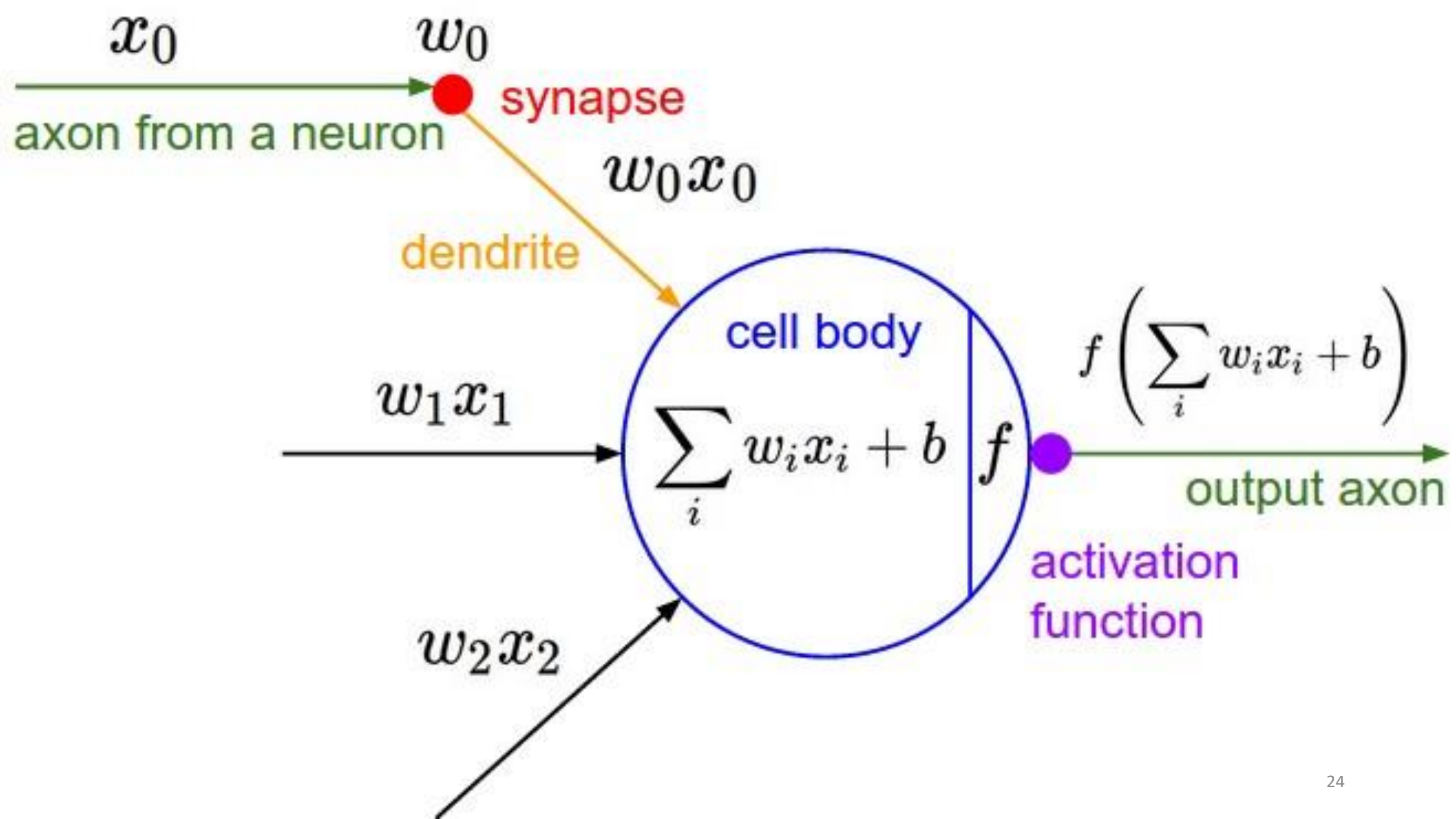



Cell body
Axon
Telodendria
Nucleus
Axon hillock
Synaptic terminals
Golgi apparatus
Endoplasmic reticulum
Mitochondrion
Dendrite
Dendritic branches

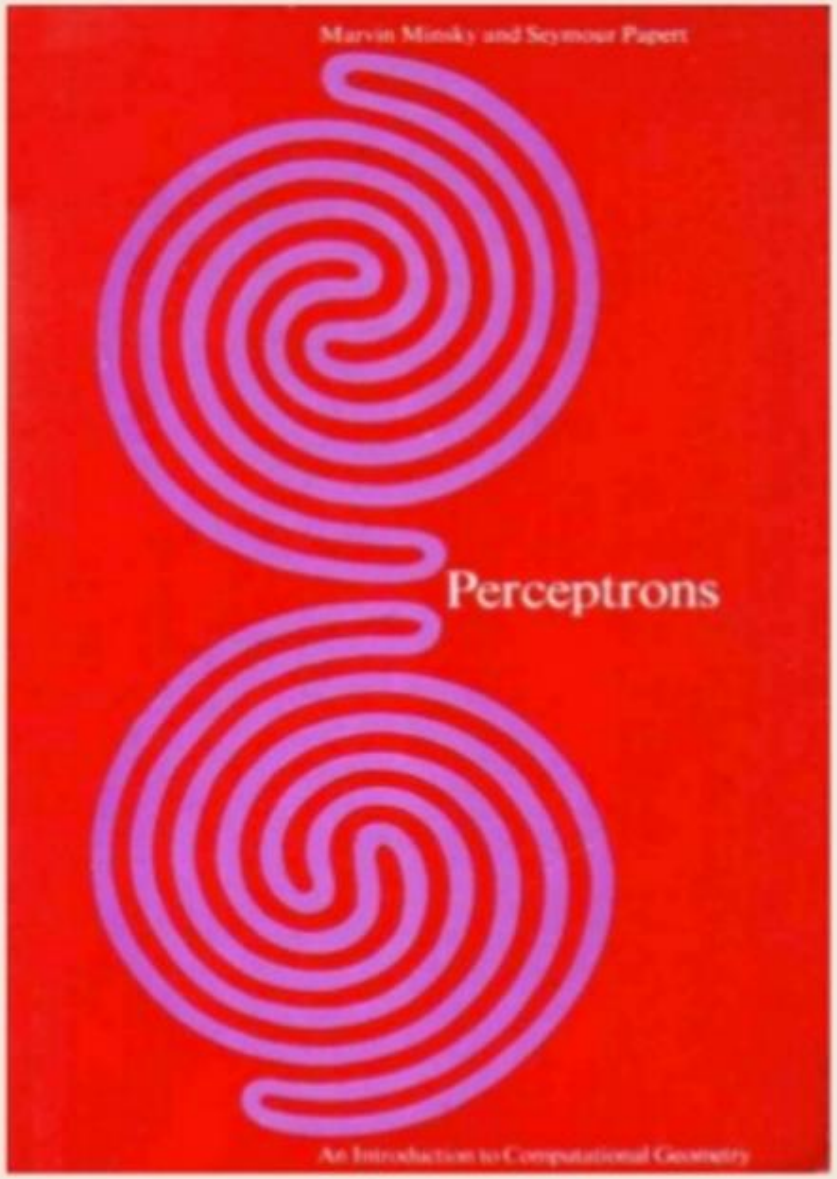
Neuron

Frank Rosenblatt's Perceptron (1957)

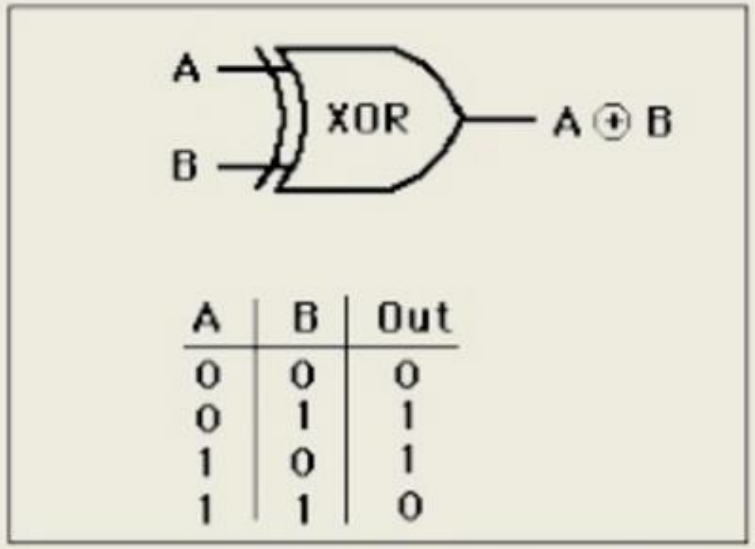



x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$w_1 x_1$
$\sum_i w_i x_i + b$
f
$f\left(\sum_i w_i x_i + b\right)$
output axon
activation function
$w_2 x_2$

1969: Perceptrons can't do XOR!



http://www.i-programmer.info/images/stories/BabBag/AI/book.jpg



A	B	Out
0	0	0
0	1	1
1	0	1
1	1	0

http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/ietron/xor.gif

Minsky & Papert

https://constructingkids.files.wordpress.com/2013/05/minsky-papert-71-csolomon-x640.jpg


AI Winter
1969 - 1990

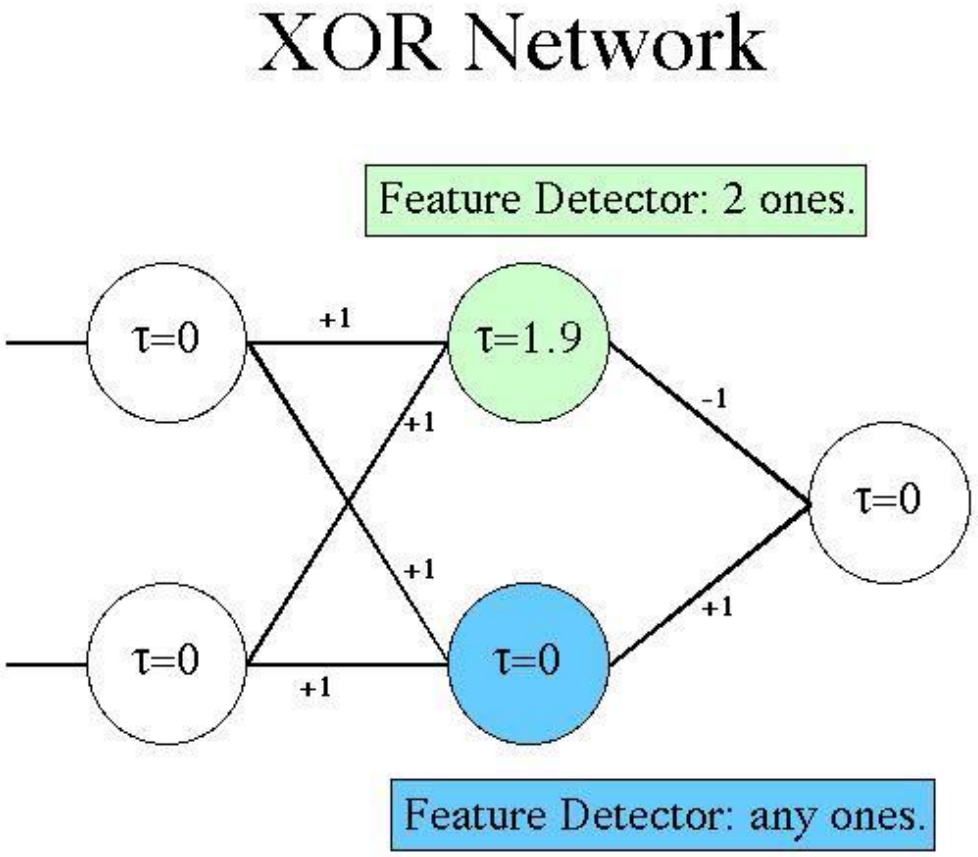
Deep Learning

Geoffrey Hinton
(Toronto, Google)

Yann LeCun
(NewYork, Facebook)

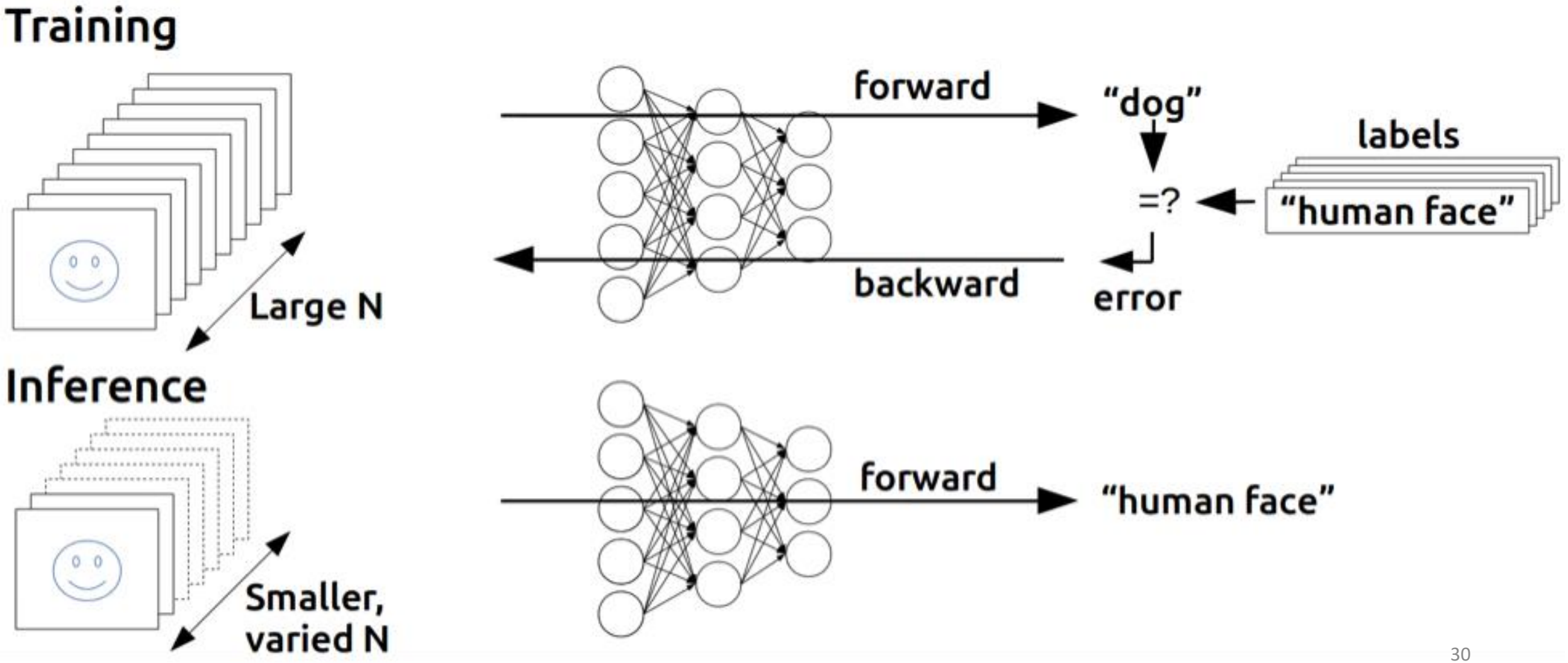
Yoshua Bengio
(Montreal)

Learning XOR (1986)



Geoffrey Hinton

Backpropagation



Chain Rule

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$$\frac{d^2y}{dx^2} = \frac{d^2y}{du^2}\left(\frac{du}{dx}\right)^2 + \frac{dy}{du}\frac{d^2u}{dx^2}$$

$$\frac{d^3y}{dx^3} = \frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^3 + 3\,\frac{d^2y}{du^2}\frac{du}{dx}\frac{d^2u}{dx^2} + \frac{dy}{du}\frac{d^3u}{dx^3}$$

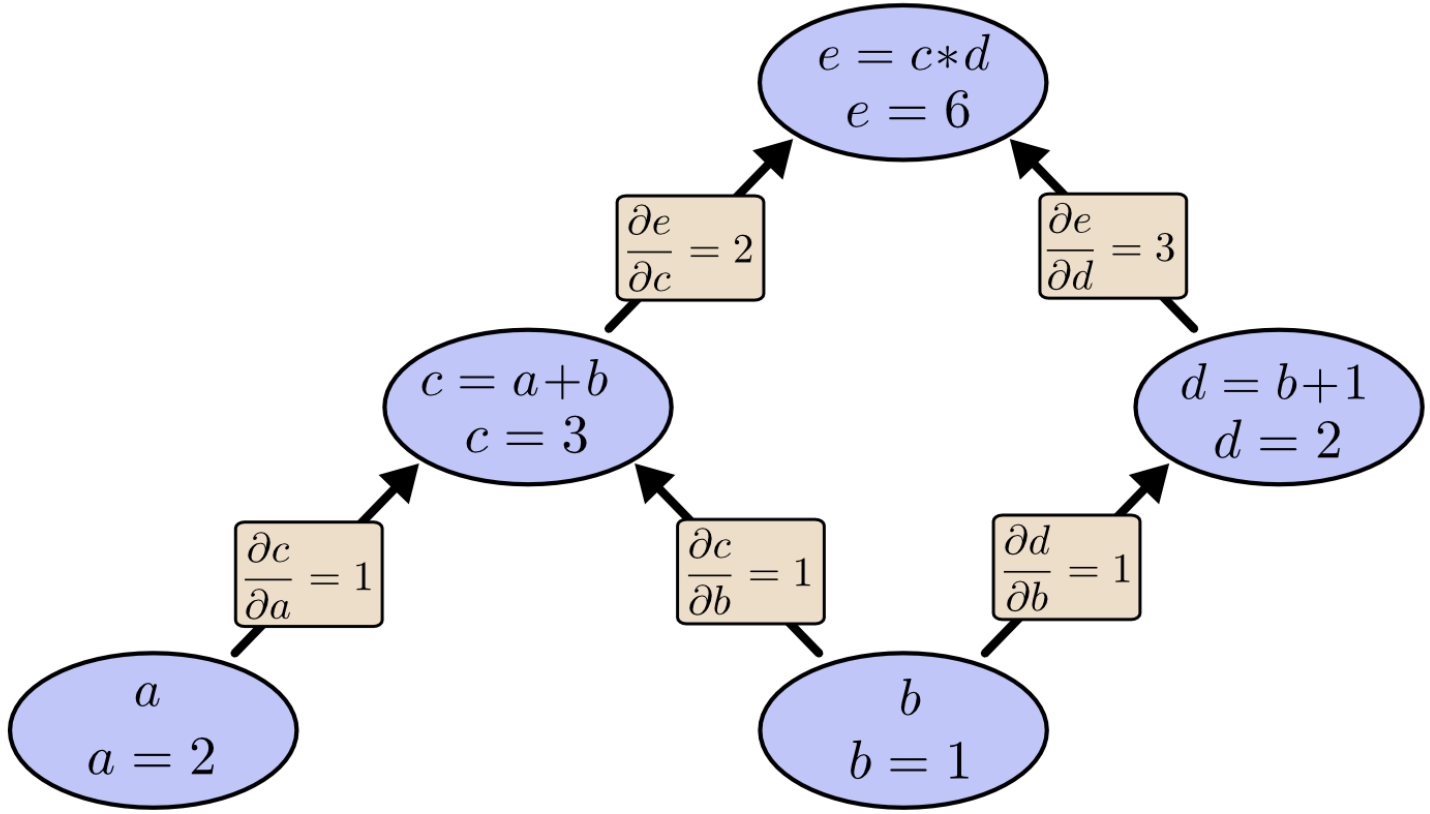
$$\frac{d^4y}{dx^4} = \frac{d^4y}{du^4}\left(\frac{du}{dx}\right)^4 + 6\,\frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^2\frac{d^2u}{dx^2} + \frac{d^2y}{du^2}\left(4\,\frac{du}{dx}\frac{d^3u}{dx^3} + 3\left(\frac{d^2u}{dx^2}\right)^2\right) + \frac{dy}{du}\frac{d^4u}{dx^4}.$$

Computation Graph

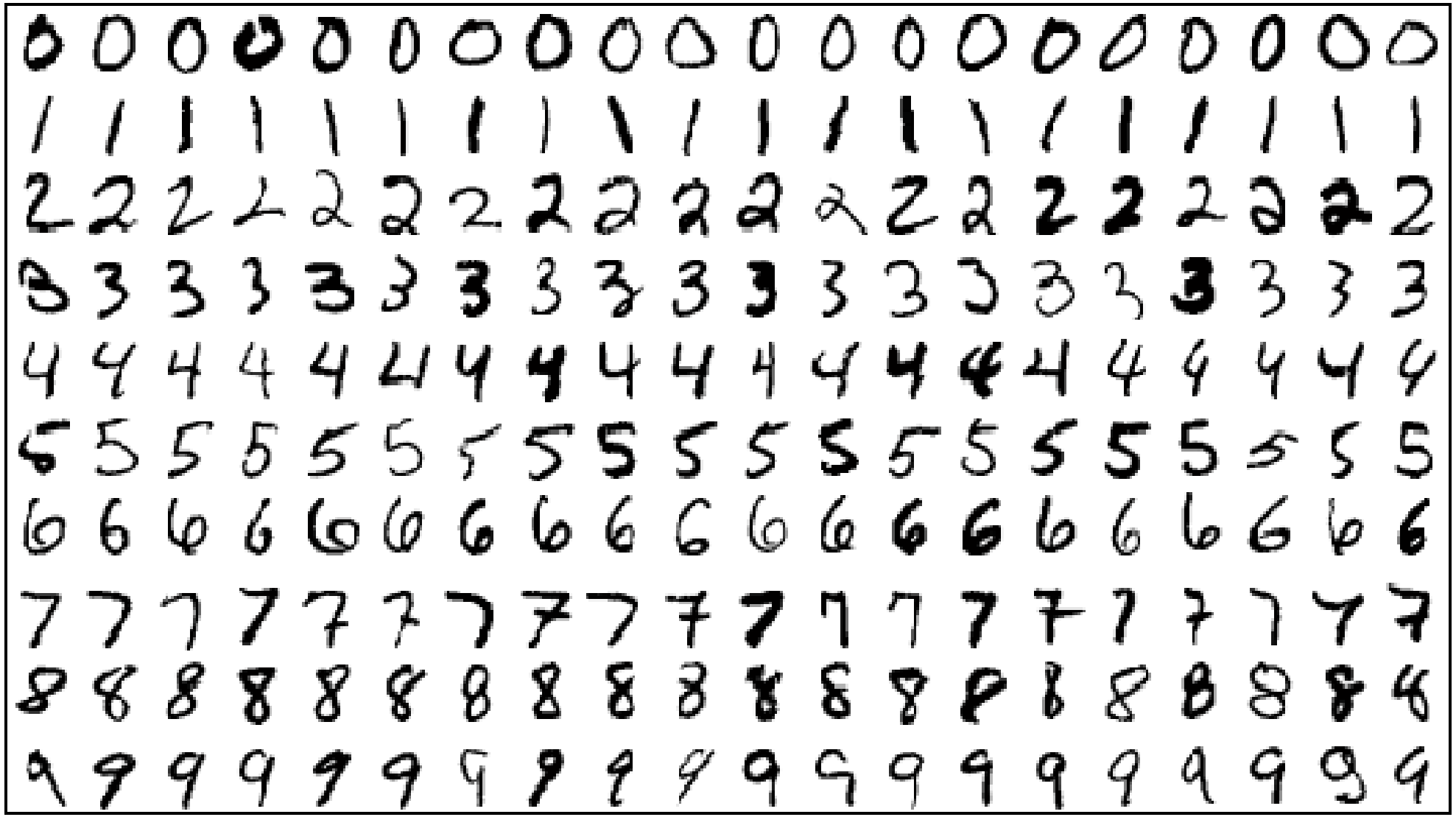
c = a + b

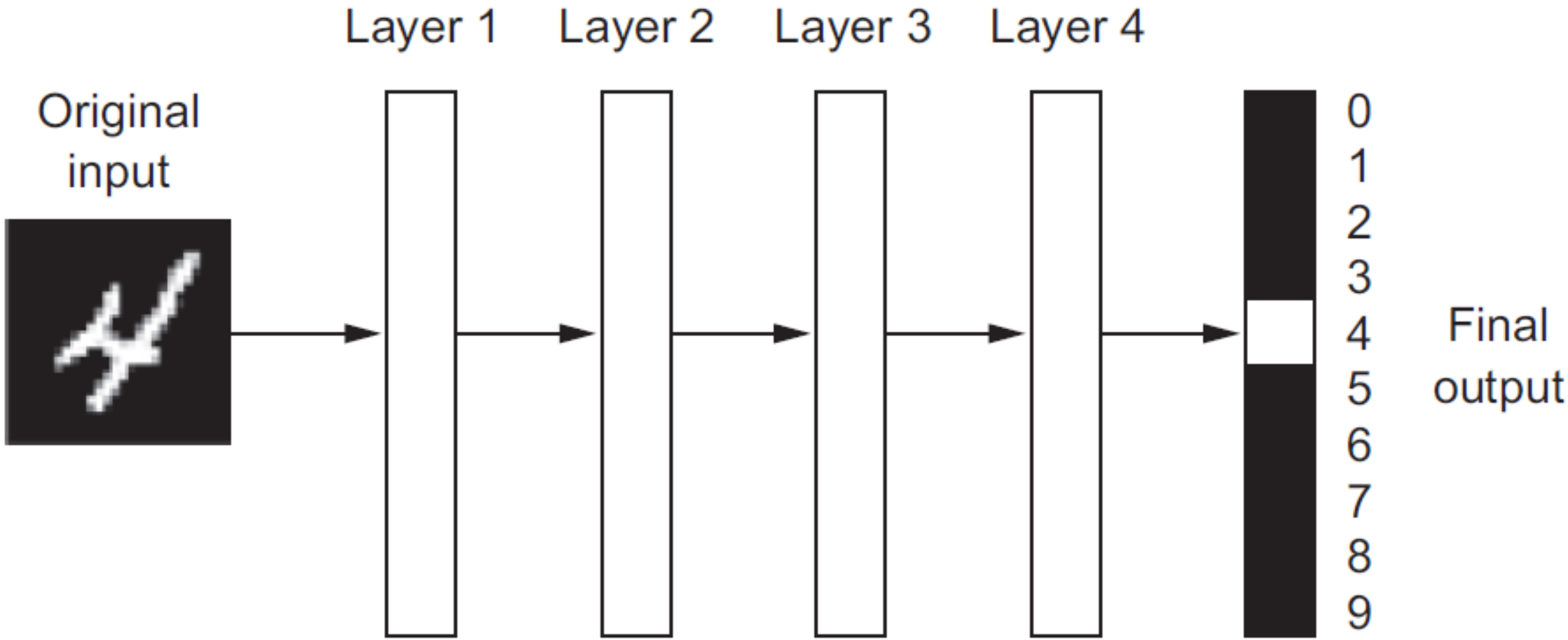
d = b + 1

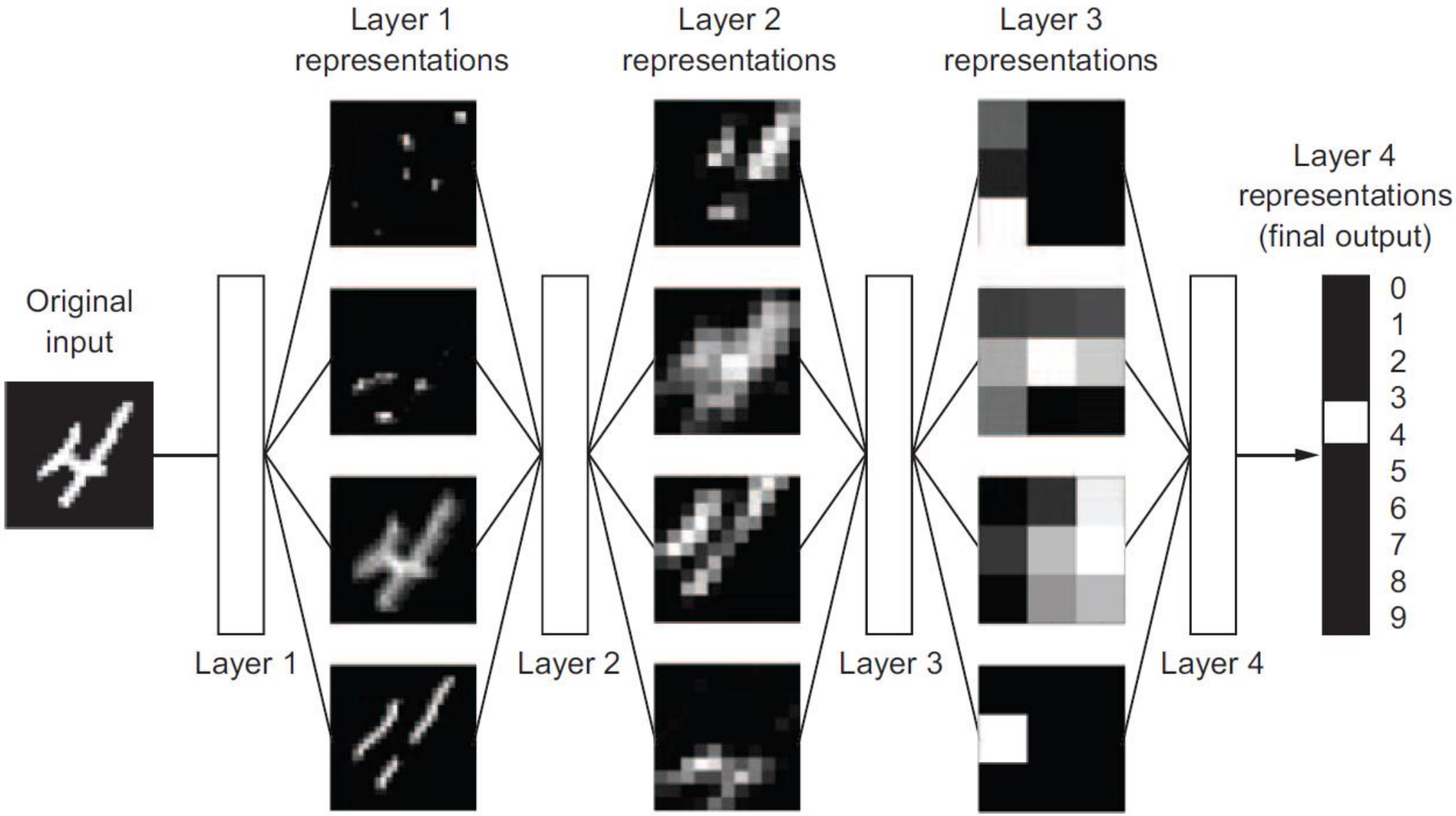
e = c*d

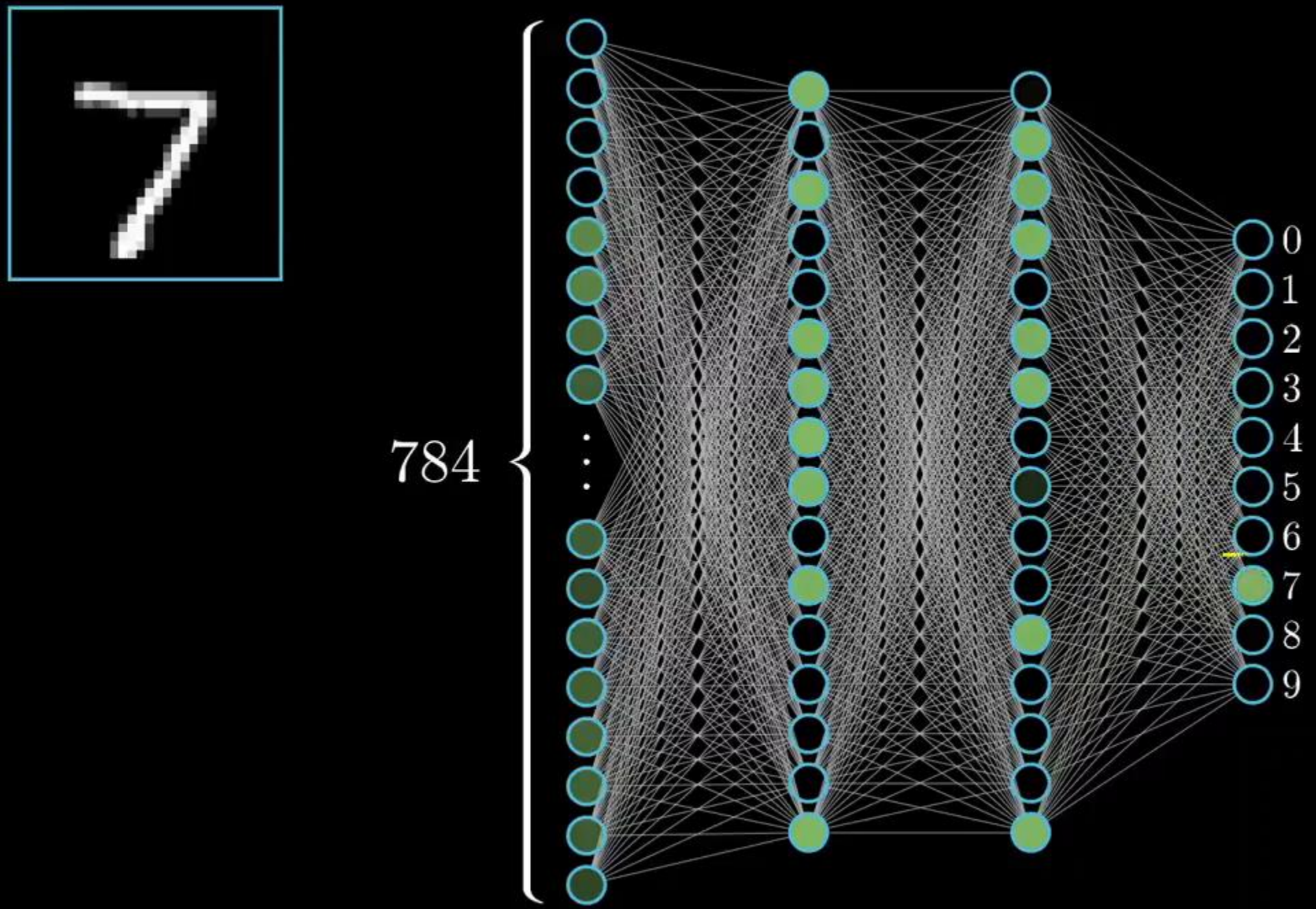


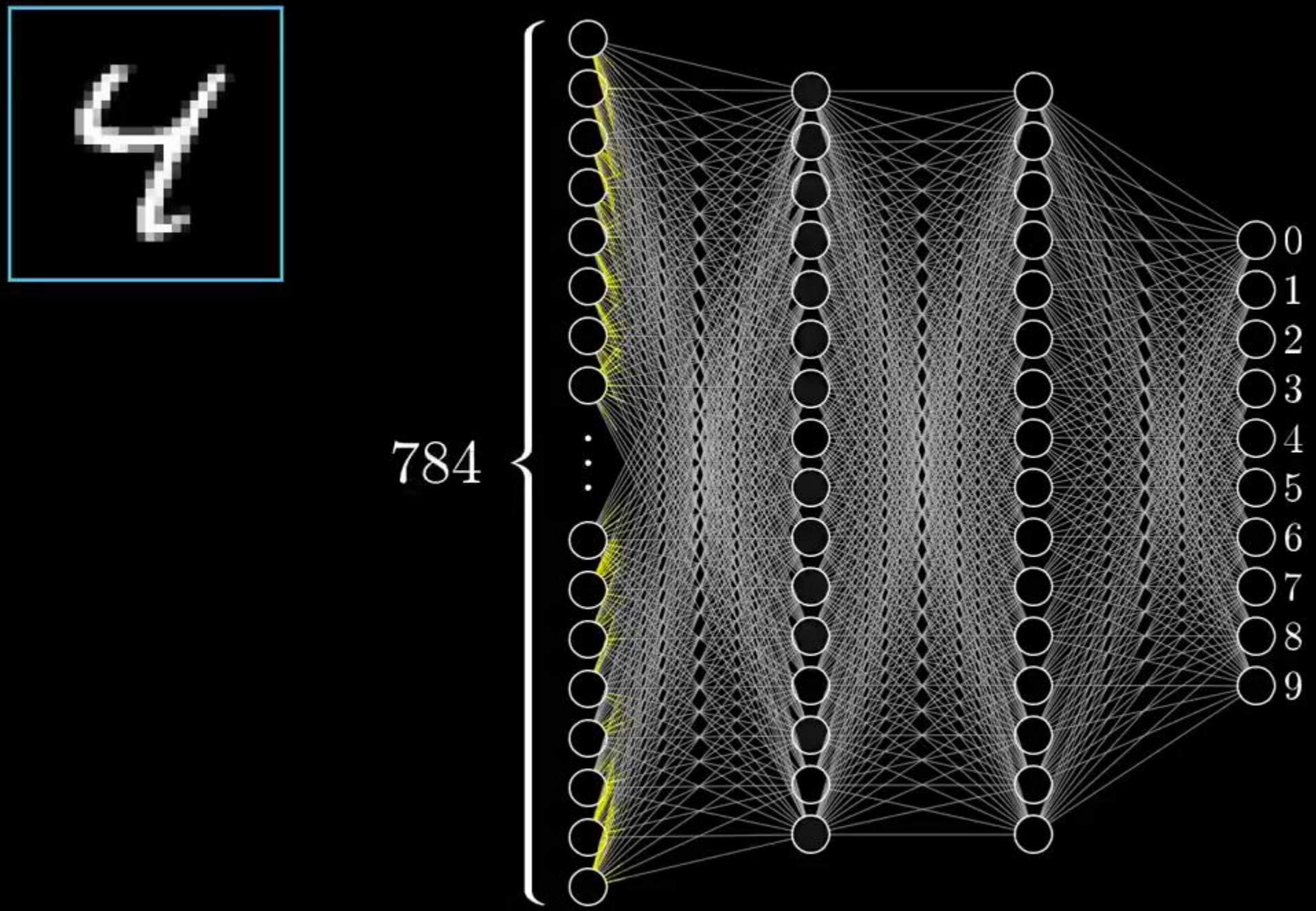
MNIST database of Handwritten Digits



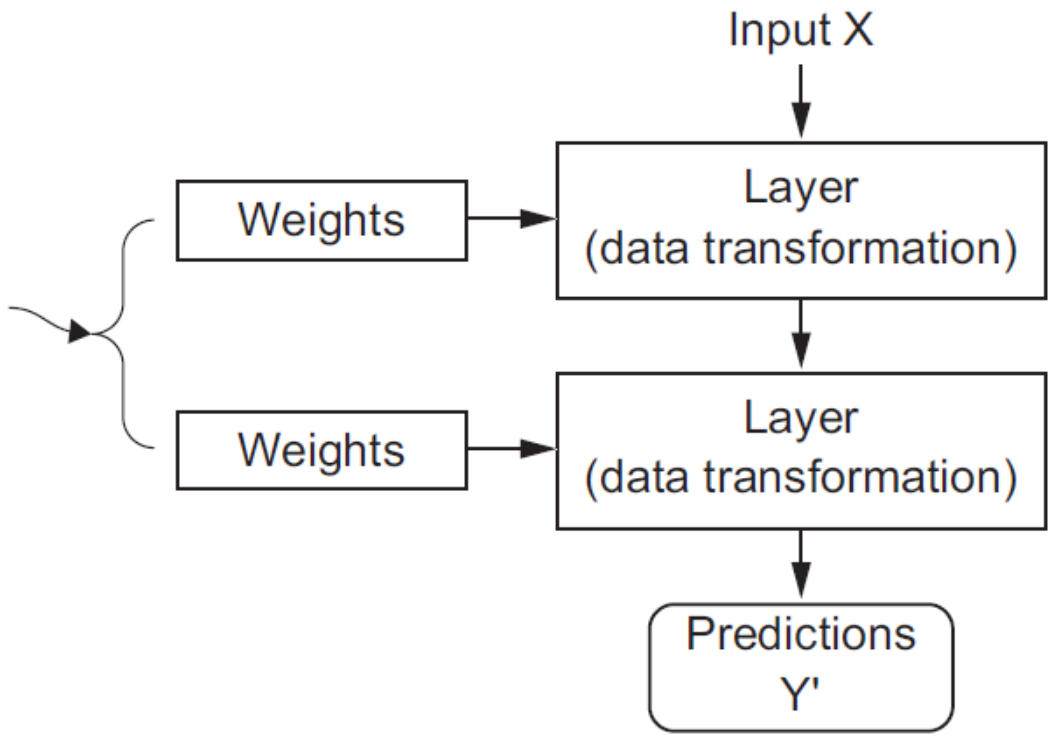

Layer 1
Layer 2
Layer 3
Layer 4
Original input
0
1
2
3
4
5
6
7
8
9
Final output

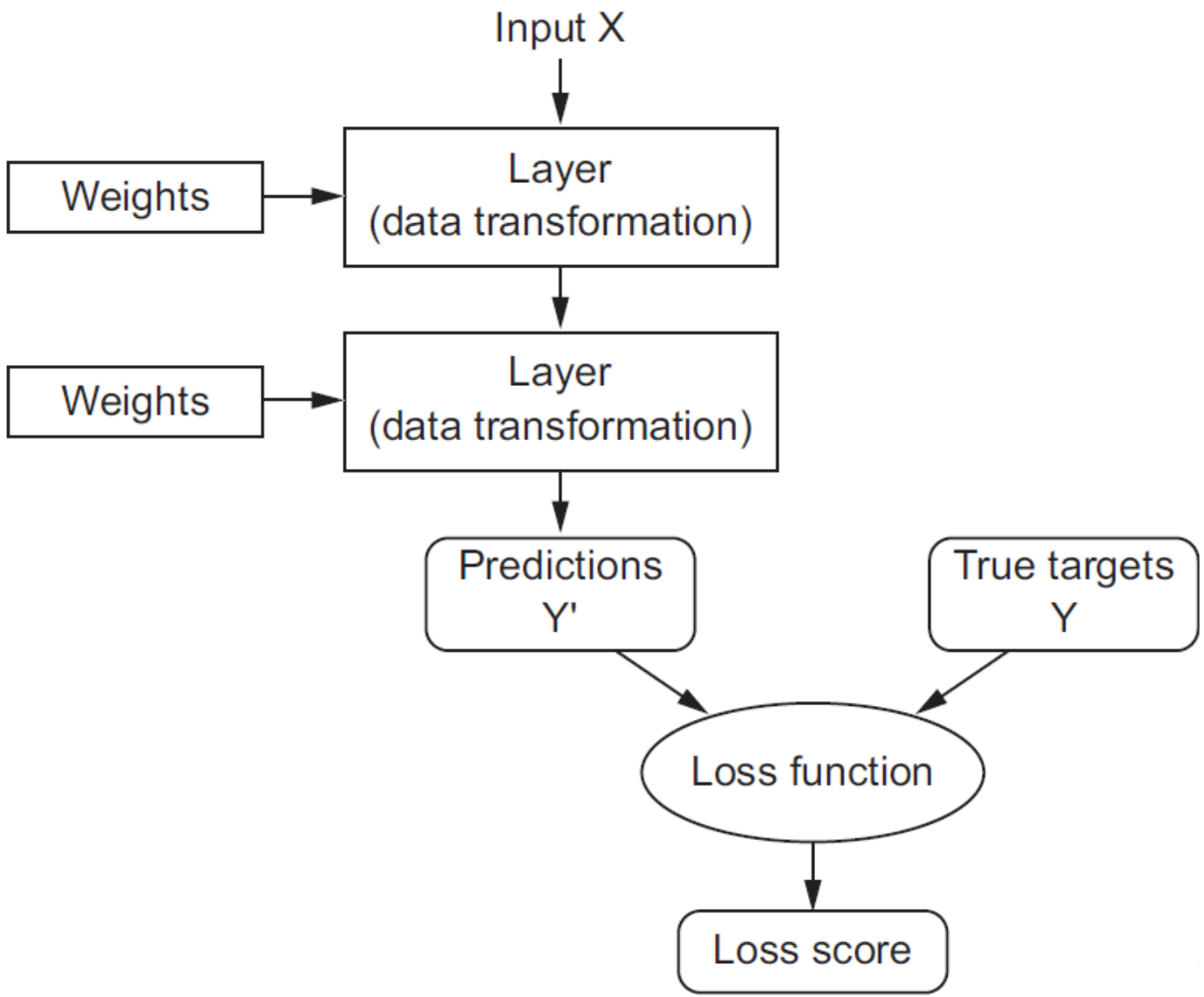

Layer 1 representations
Layer 2 representations
Layer 3 representations
Layer 4 representations (final output)
Original input
Layer 1
Layer 2
Layer 3
Layer 4
0
1
2
3
4
5
6
7
8
9

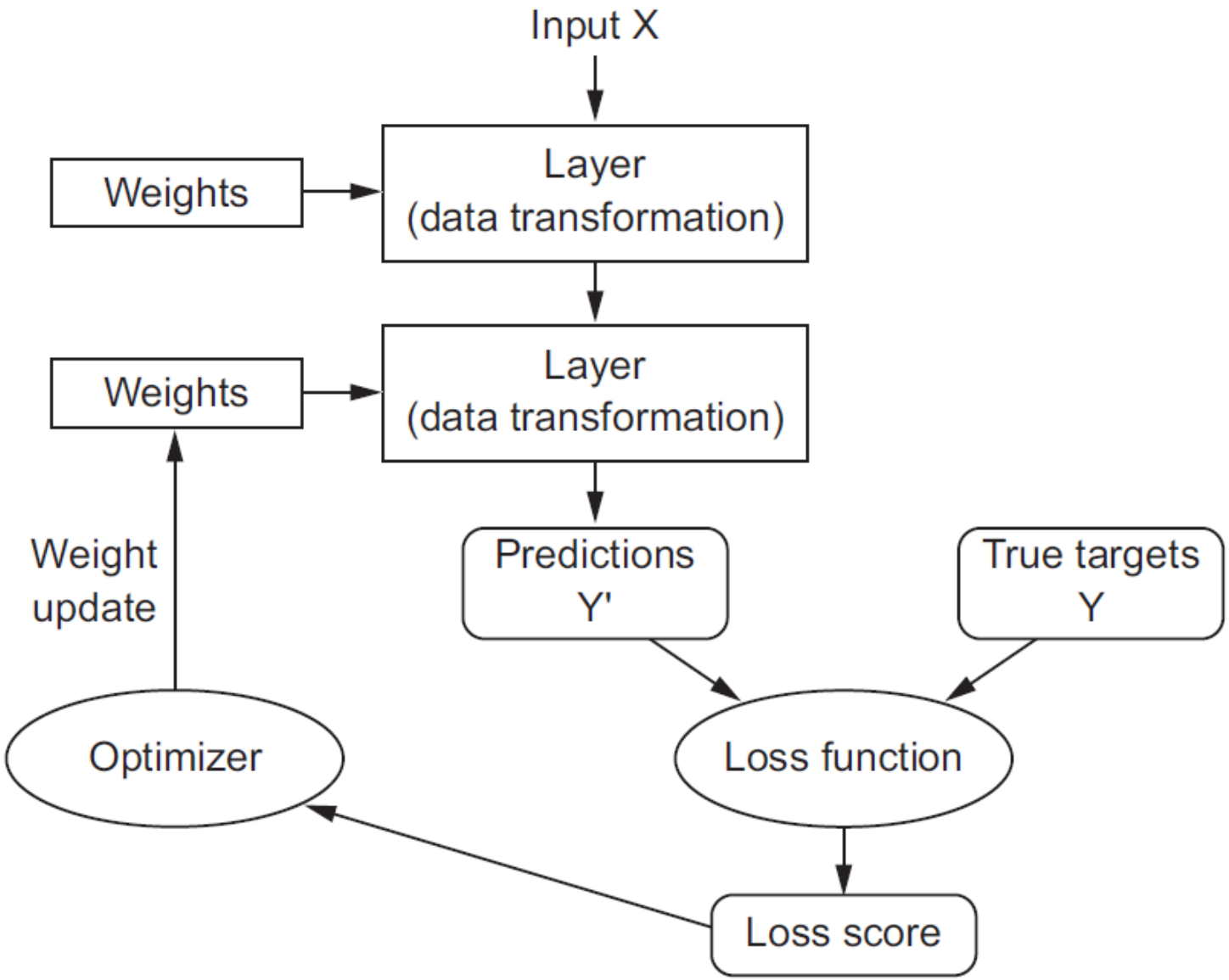

784
0
1
2
3
4
5
6
7
8
9


784
0
1
2
3
4
5
6
7
8
9

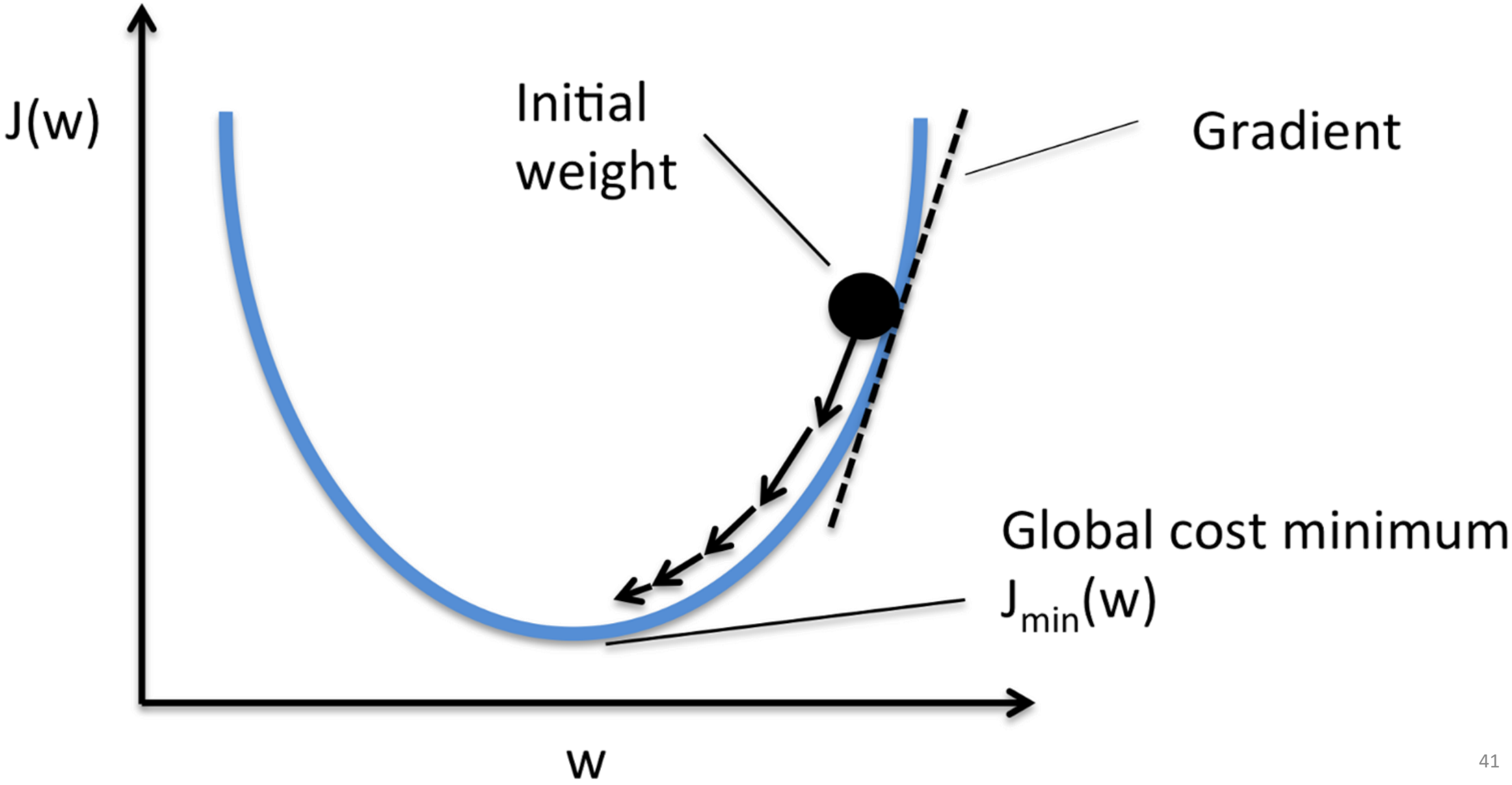
Goal: finding the right values for these weights

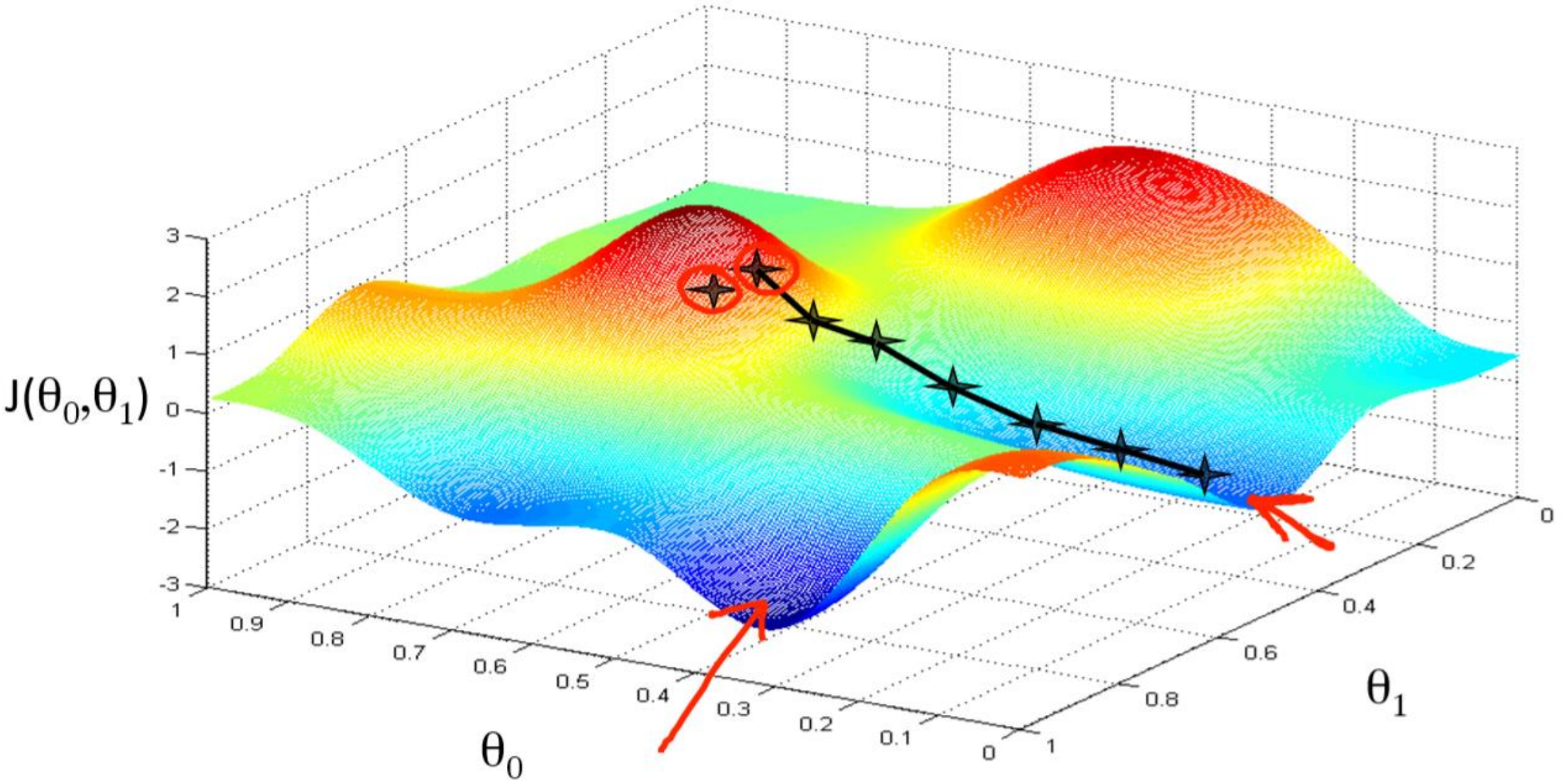



Input X
Weights
Layer
(data transformation)
Weights
Layer
(data transformation)
Predictions
Y'
True targets
Y
Loss function
Loss score


Input X
Weights
Layer
(data transformation)
Weights
Layer
(data transformation)
Weight
update
Predictions
Y'
True targets
Y
Optimizer
Loss function
Loss score

Gradient Descent





https://hackernoon.com/gradient-descent-aynk-7cbe95a778da

Cost Function

- Mean-Squared Error

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(f_\theta(x_i) - y_i)^2$$

Gradient Descent of MSE

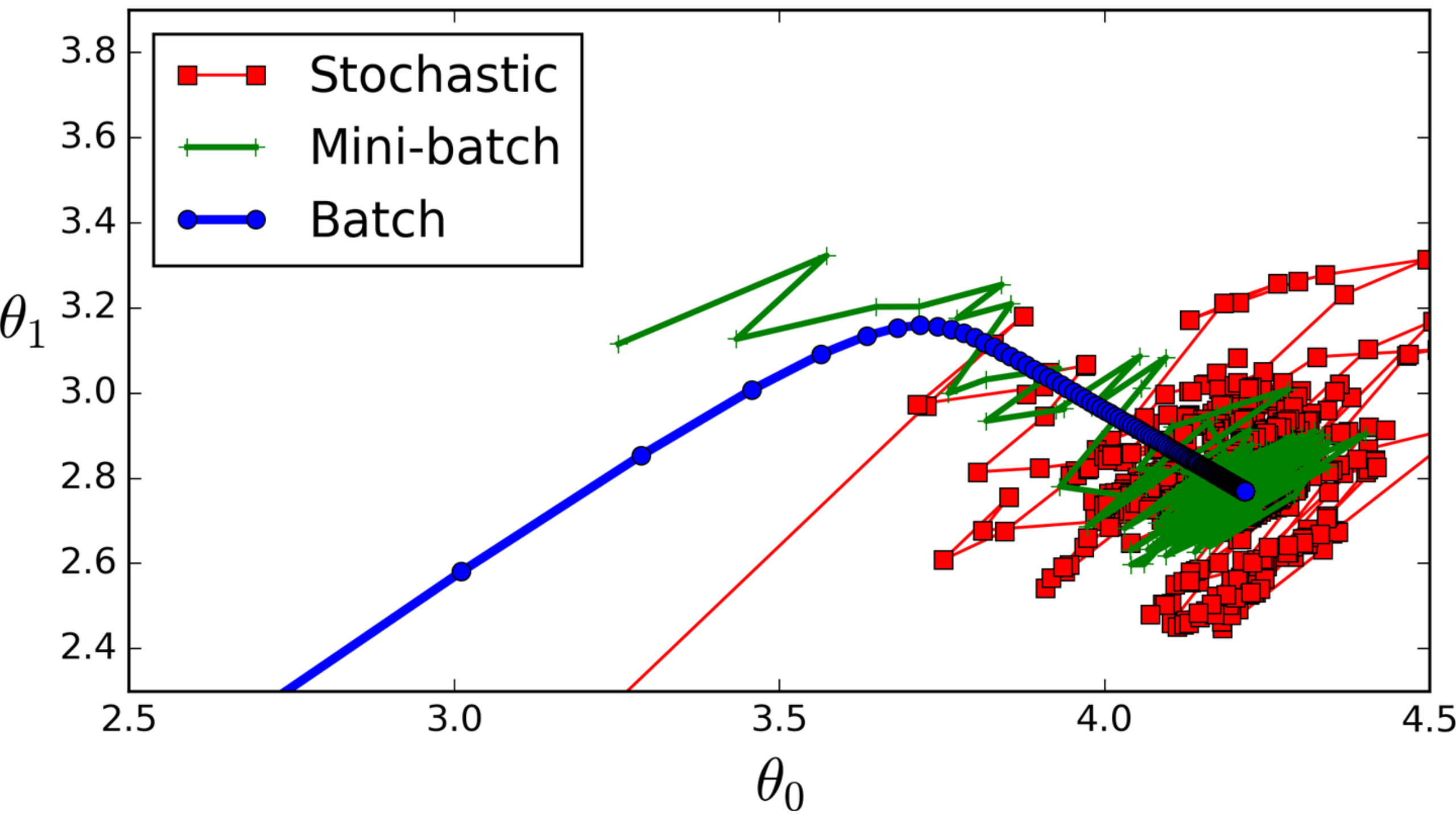
- Gradient of MSE

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{2}{N}\sum_{i=1}^{N}(f_\theta(x_i) - y_i) f'_\theta(x_i)$$

- Update

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

- Repeat until Convergence

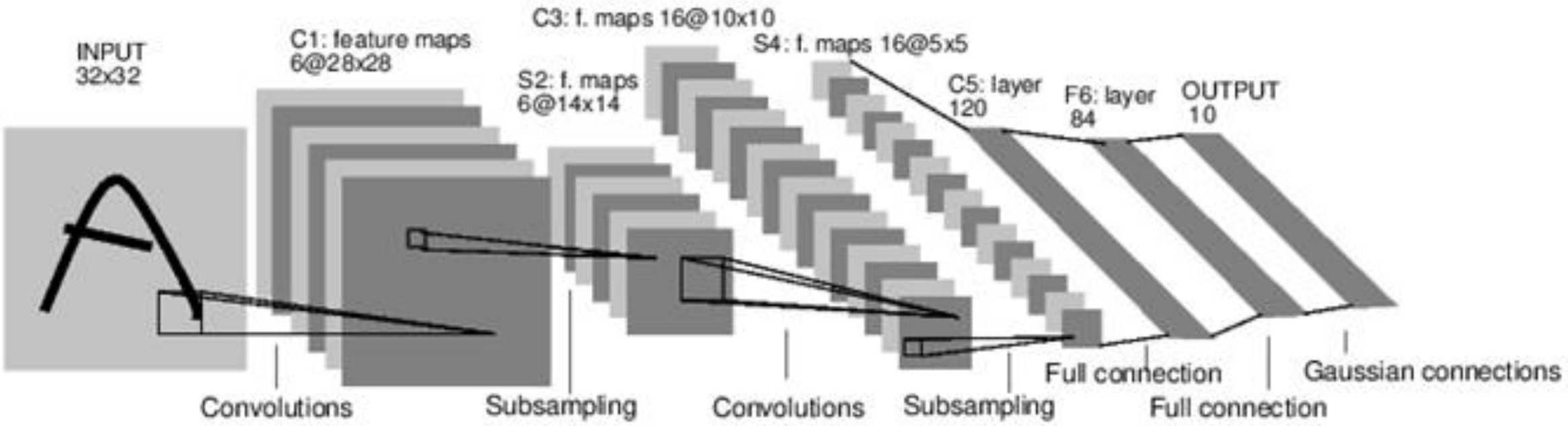

Stochastic
Mini-batch
Batch
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
θ_1
2.5
3.0
3.5
4.0
4.5
θ_0

Cost function

$$C(\underbrace{w_1, w_2, \ldots, w_{13{,}002}}_{\text{Weights and biases}})$$

Convolutional Neural Network (LeNet-5)

- https://medium.com/@sh.tsang/paper-brief-review-of-lenet-1-lenet-4-lenet-5-boosted-lenet-4-image-classification-1f5f809dbf17



A Full Convolutional Neural Network (LeNet)

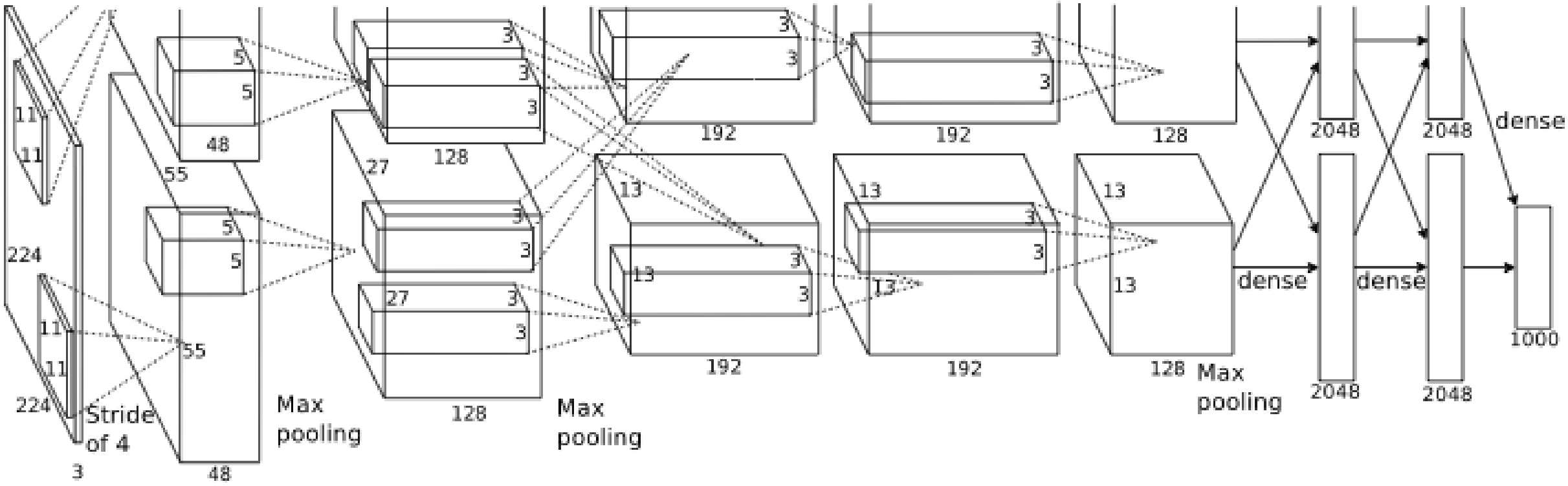

IMAGENET

ImageNet Large Scale Visual Object Recognition Challenge (ILSVRC)

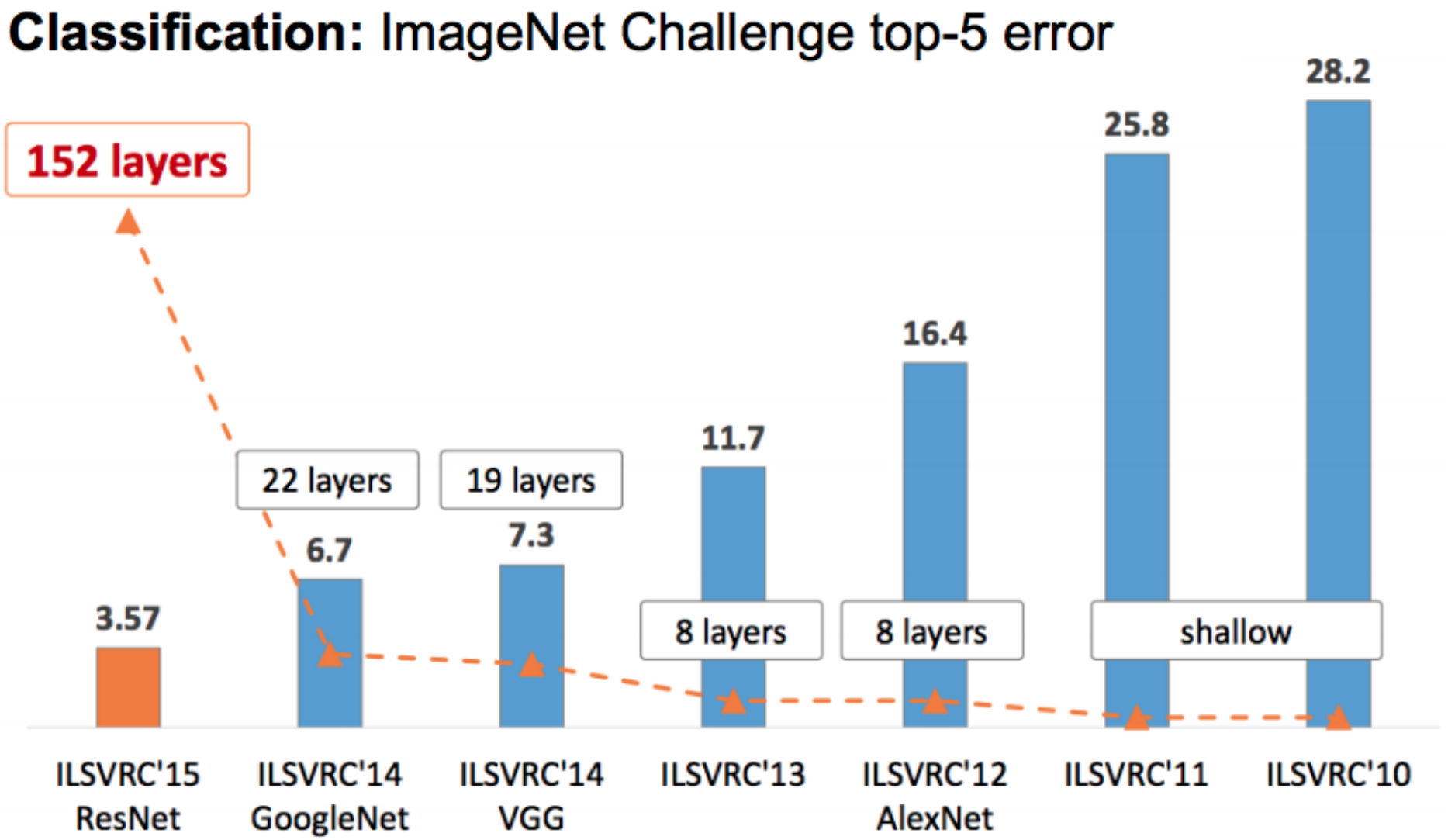
- 1000 categories
- For ILSVRC 2017
  - **Training images** for each category ranges from 732 to 1300
  - 50,000 validation **images** and 100,000 test **images**.
- Total number of images in ILSVRC 2017 is around 1,150,000

Convolutional Neural Network

- Alex Krizhevsky, Geoffrey Hinton et al., 2012

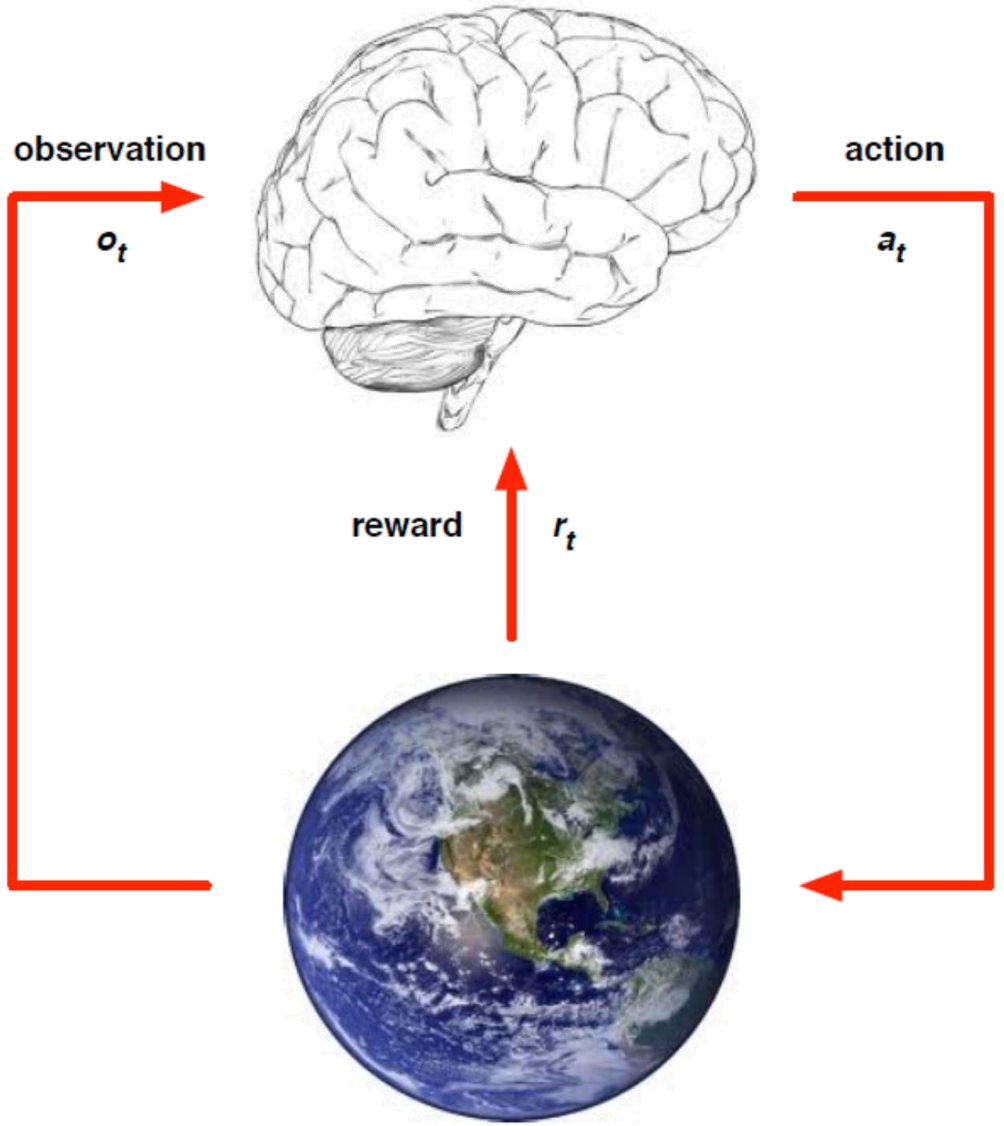


Previous Winners of ILSVRC

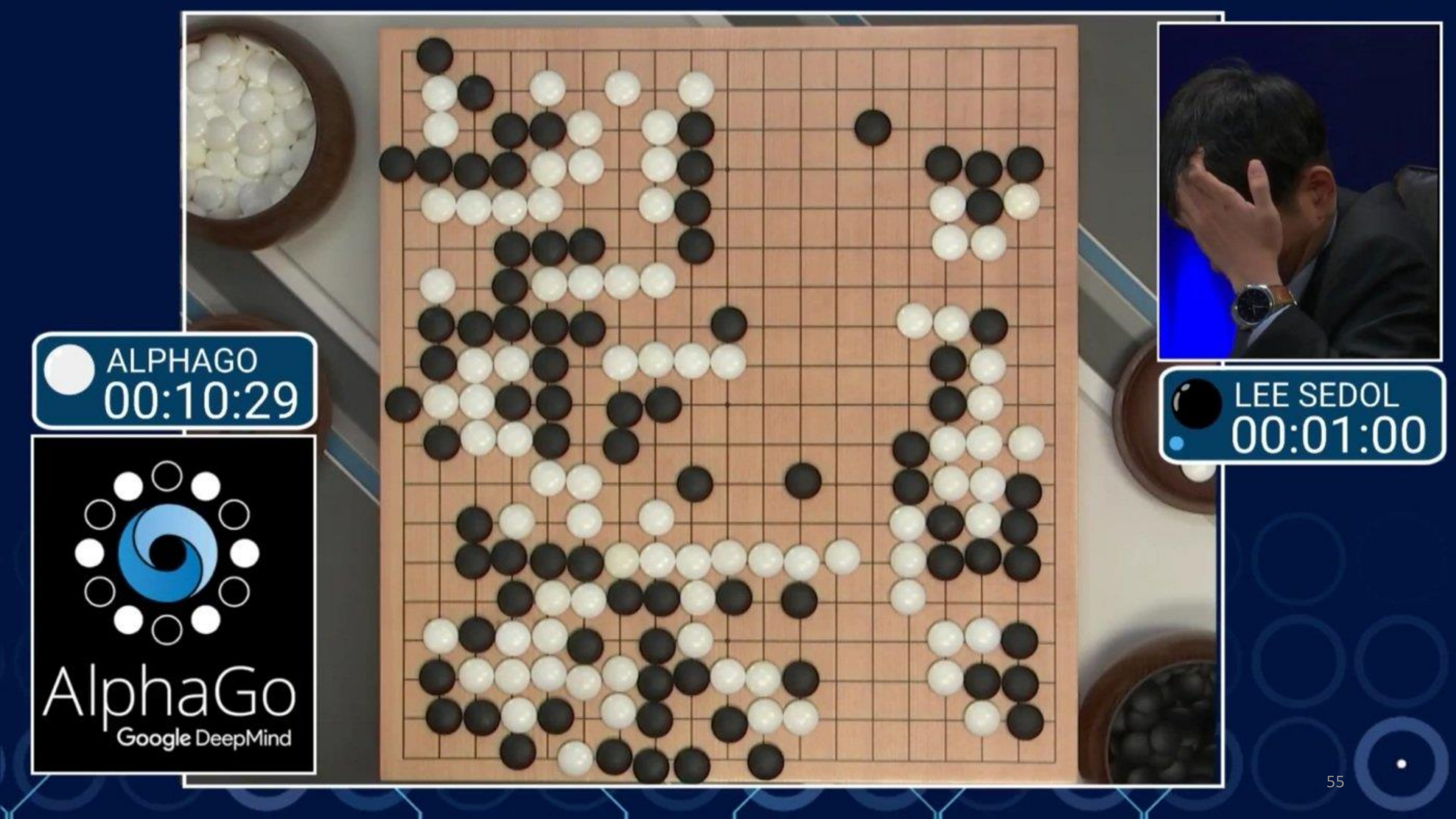


Deep Reinforcement Learning

Reinforcement Learning



Google DeepMind

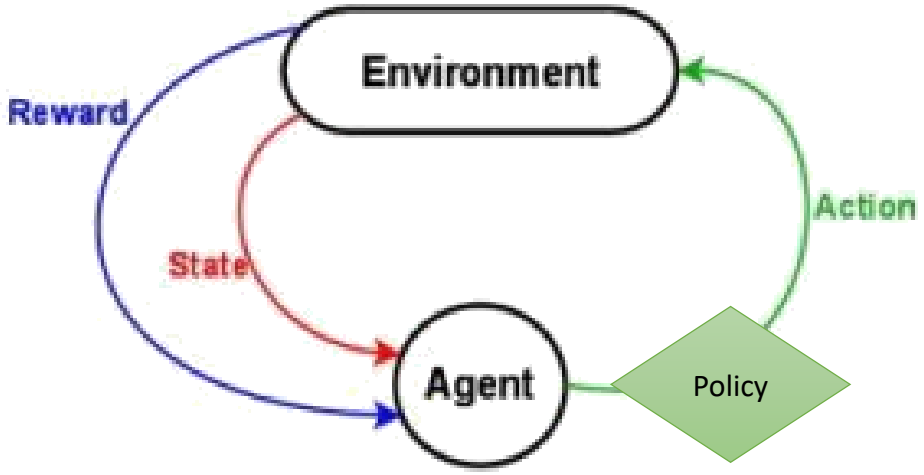

ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

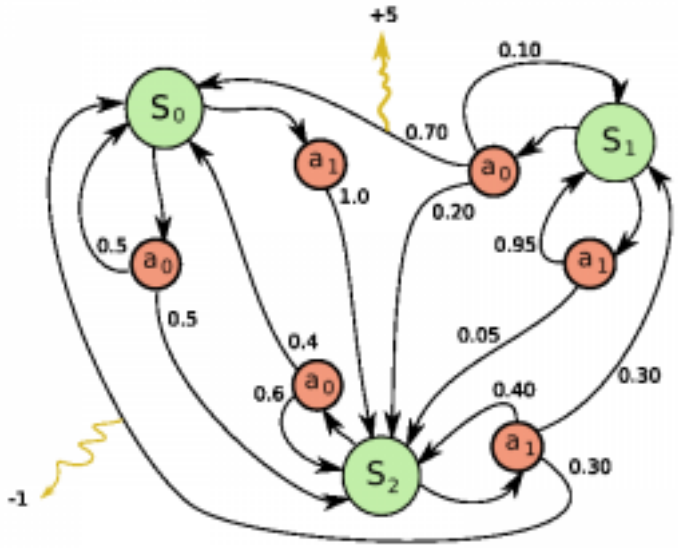
The Complexity of Go vs Chess

Game	Board size	State space	Game tree size
Go	19 x 19	10^{172}	10^{360}
Chess	8 x 8	10^{50}	10^{123}
Checkers	8 x 8	10^{18}	10^{54}

Reinforcement Learning

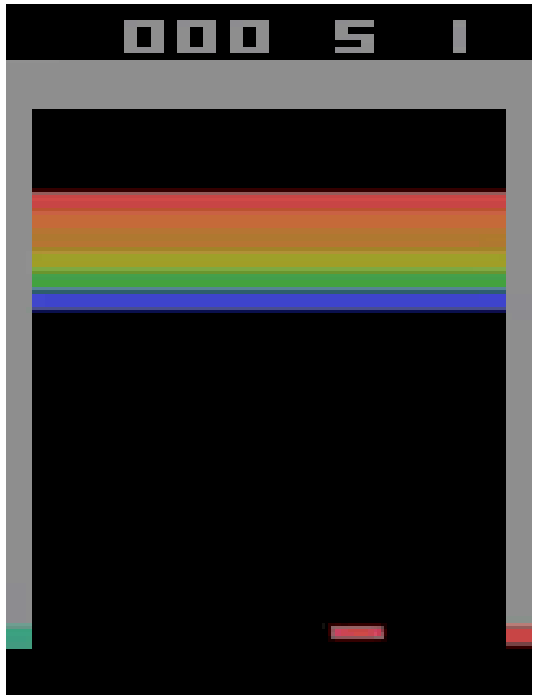
- An agent learns how to do actions a_t to achieve maximum reward R
- Policy $\pi(a_t|s_t)$: agent's behavior function
- Value function V: evaluate quality of each action/state
- Model: agent's representation of the environment

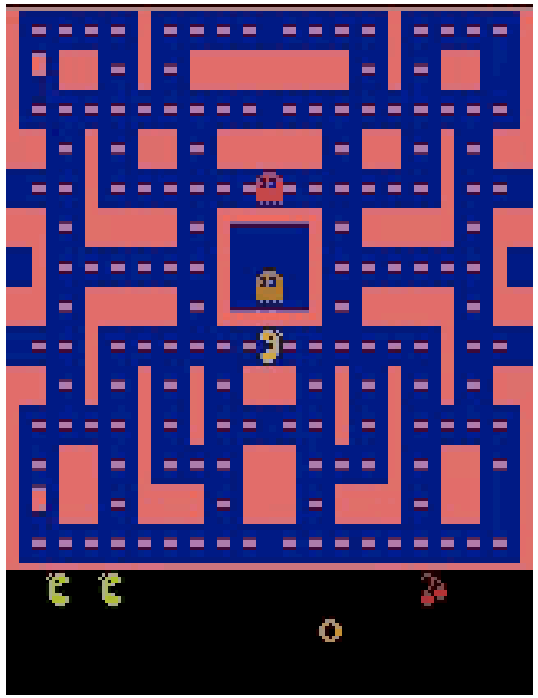




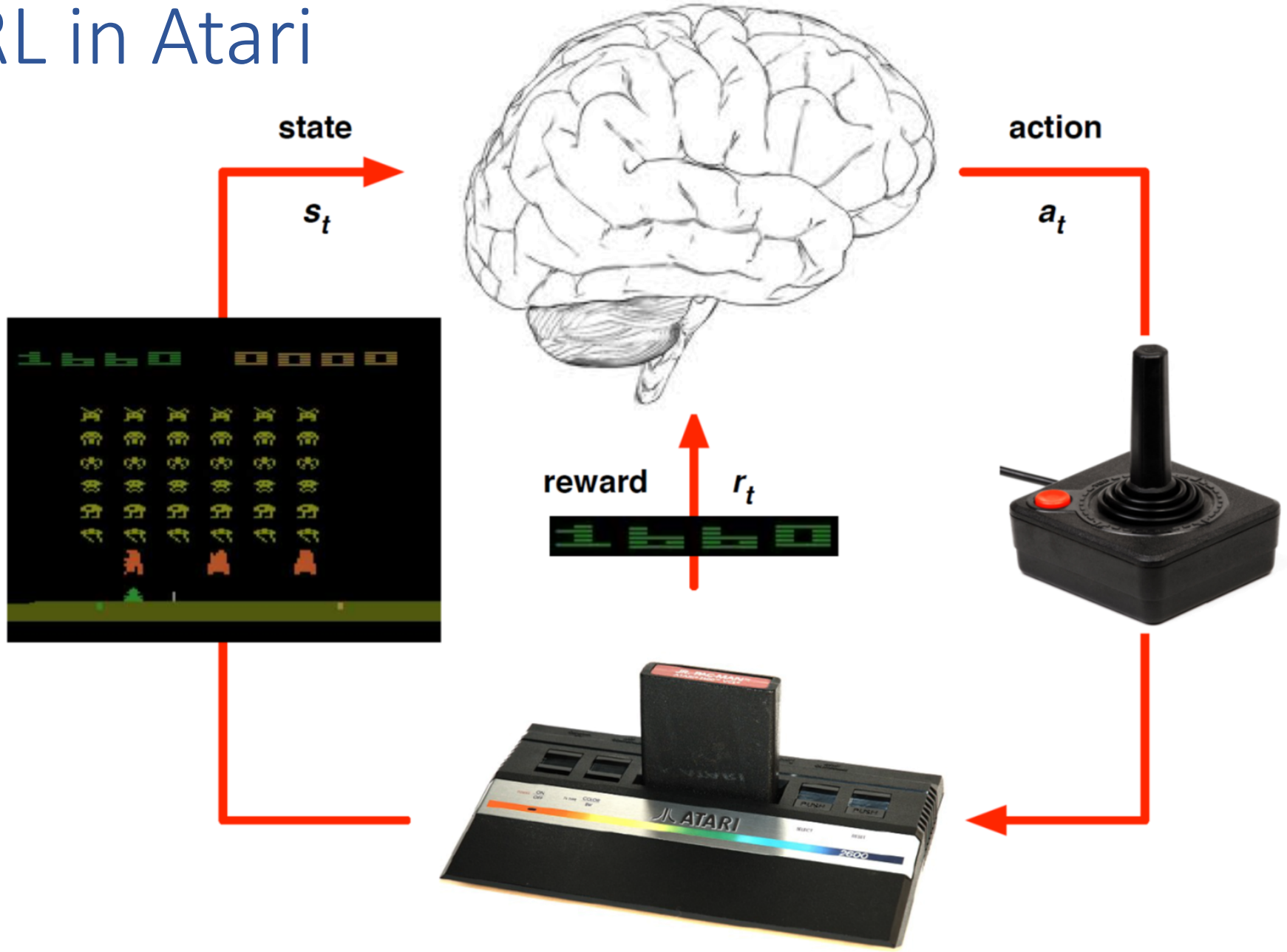
Learn to Play Atari Games

- Mnih et al., “Human Level Control through Deep Reinforcement Learning,” *Nature*, 2015

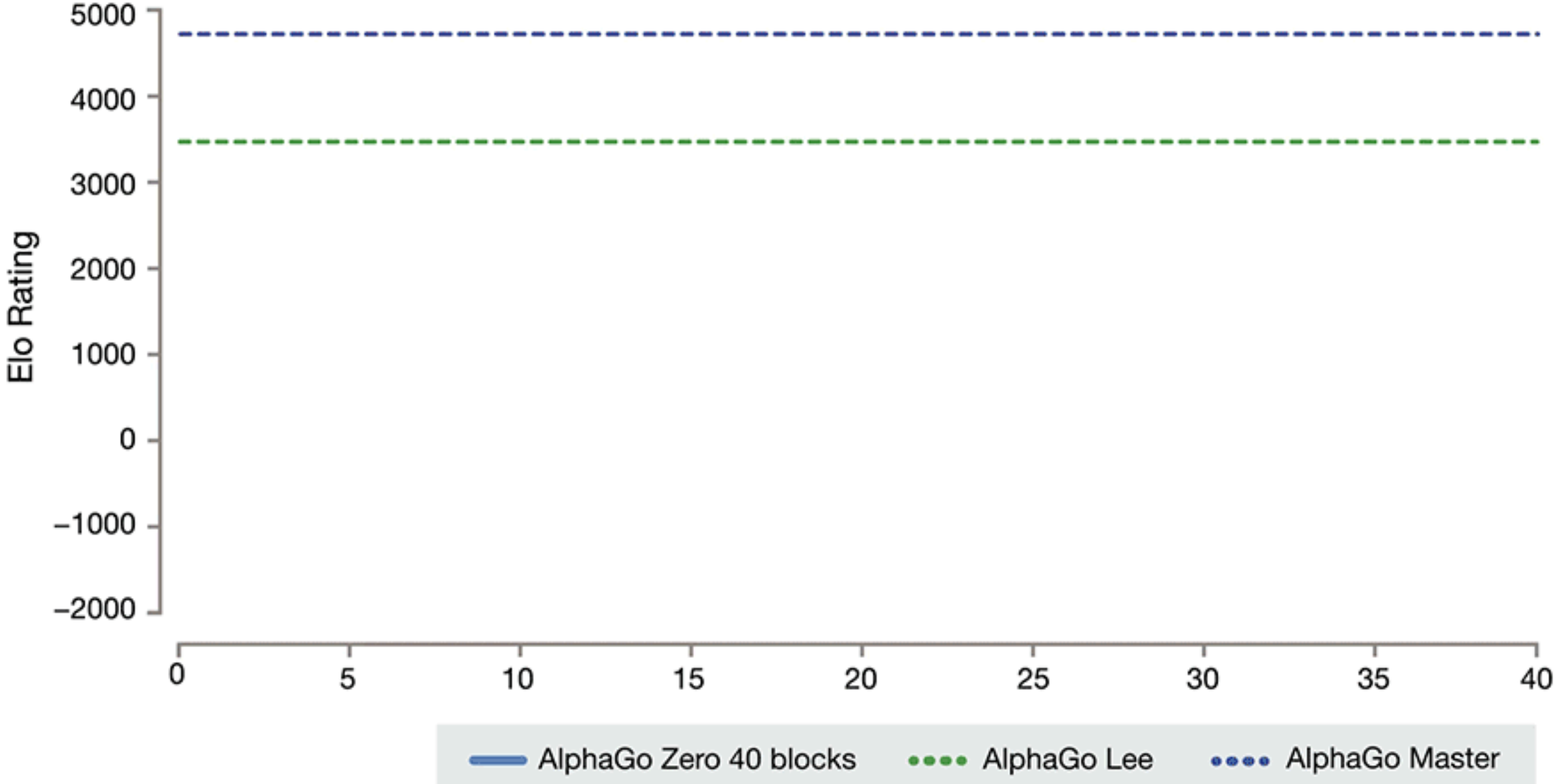





DRL in Atari




AlphaGo Zero
Starting from scratch


5000
4000
3000
2000
1000
0
−1000
−2000
Elo Rating
0
5
10
15
20
25
30
35
40
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master


Elo Rating
5000
4000
3000
2000
1000
0
−1000
−2000
0
5
10
15
20
25
30
35
40
40 days
AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master

Virtual-to-real Learning

- Inspired by DeepMind (Mnih et al., *Nature*, 2015)
  - "Human Level Control through Deep Reinforcement Learning"
- Applied to computer vision applications
  - **Image segmentation**: Armeni et al. (2016), Qiu et al., (2017)
  - **Indoor navigation**: Brodeur et al. (2017), Gupta et al. (2017), Savva et al. (2017), Wu et al. (2018)
  - **Autonomous vehicles**: Marinez et al. (2017), Muller et al. (2018), Pan et al. (2017), Shah et al. (2018)

UnrealCV

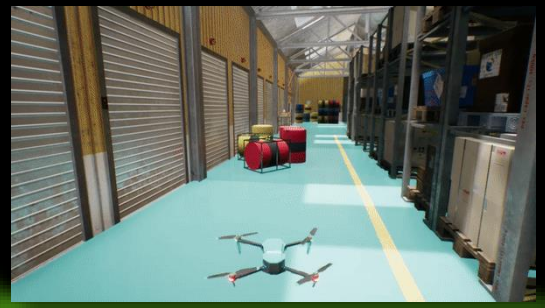
CAD^{2}Real

Semantic Segmentation

Depth Prediction

VIVID

Autonomous Navigation

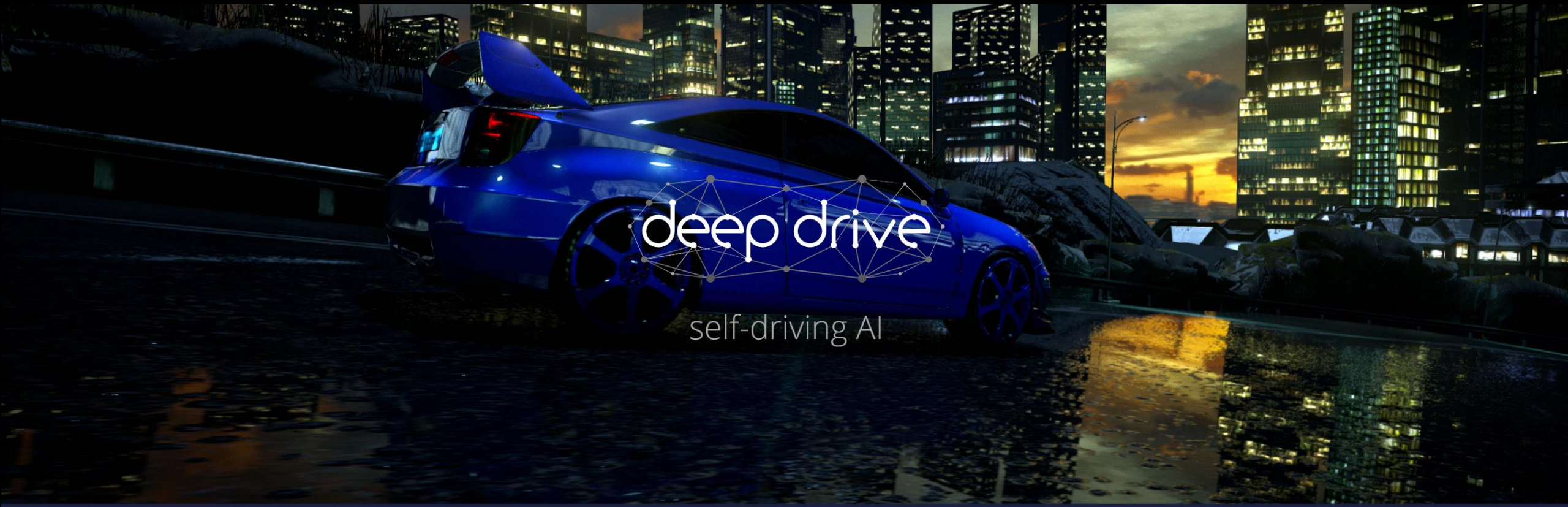


Action Recognition

Simulate Real-life Events

Searching for the Shooter

DeepDrive

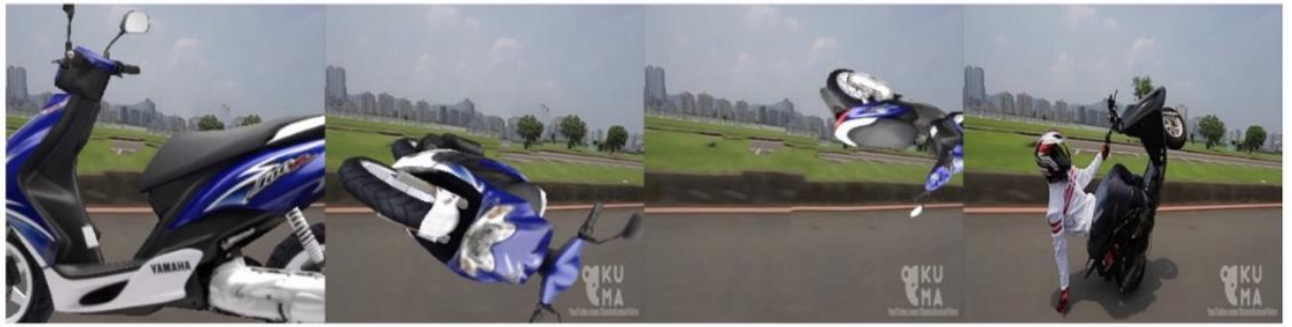


Deepdrive is a simulator that allows anyone with a PC to push the state-of-the-art in self-driving.

Limits of Deep Learning

No Idea of Real World

school bus 1.0 garbage truck 0.99 punching bag 1.0 snowplow 0.92



motor scooter 0.99 parachute 1.0 bobsled 1.0 parachute 0.54

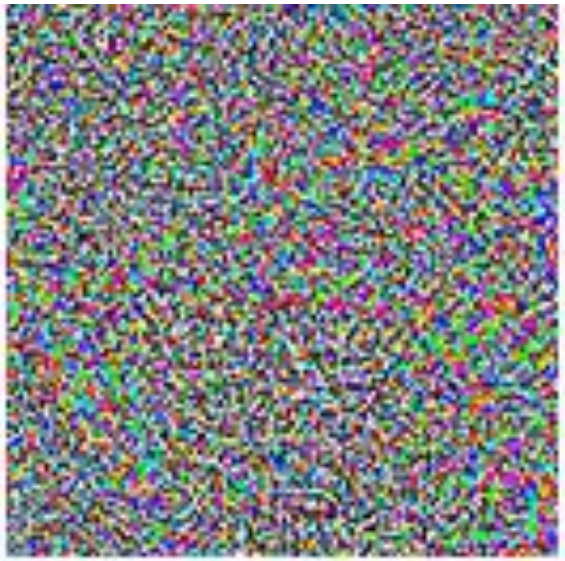
fire truck 0.99 school bus 0.98 fireboat 0.98 bobsled 0.79

Adversarial Attack

“panda”

57.7% confidence

$+\epsilon$



=

“gibbon”

99.3% confidence

Number of Connections in the Brain

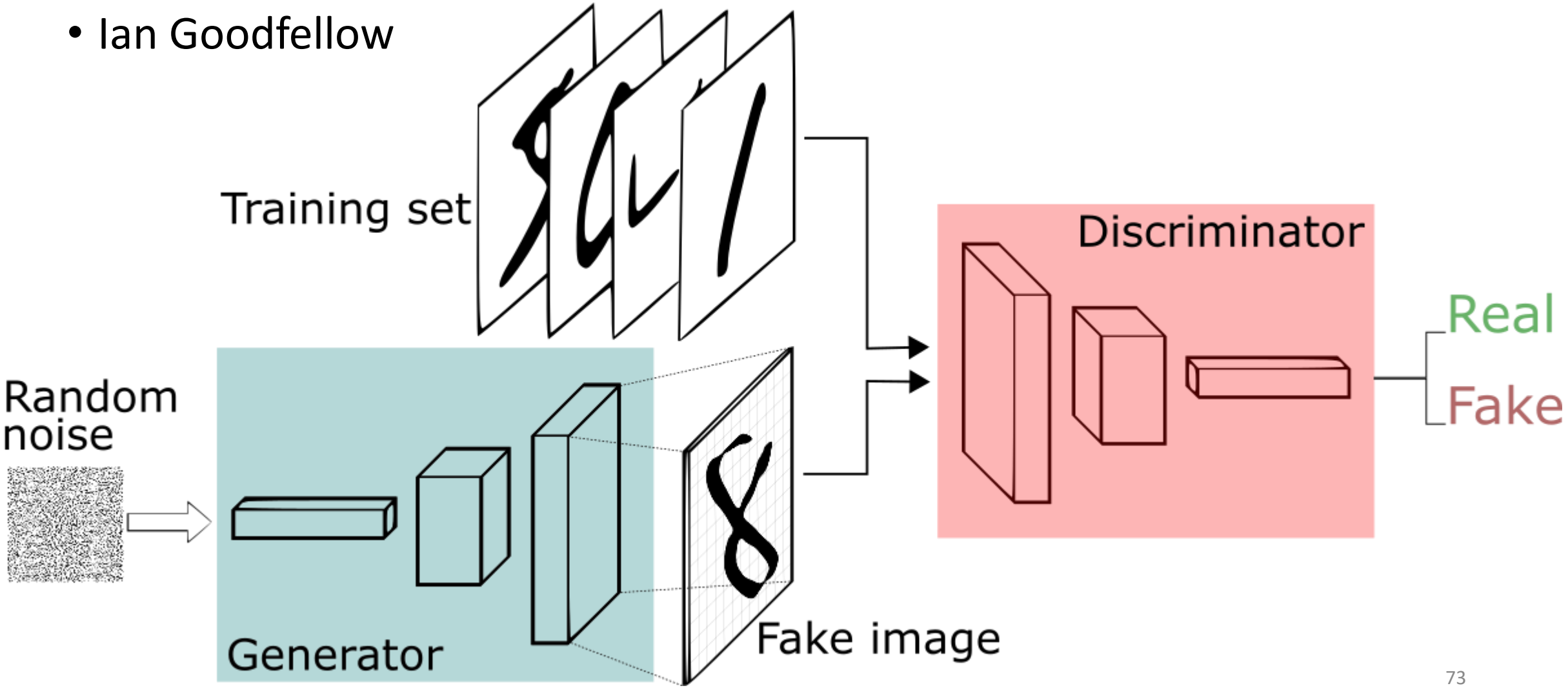
Neurons (for adults):
10^11, or 100 billion, 100000000000

Synapses (based on 1000 per neuron):
10^14, or 100 trillion, 100000000000000

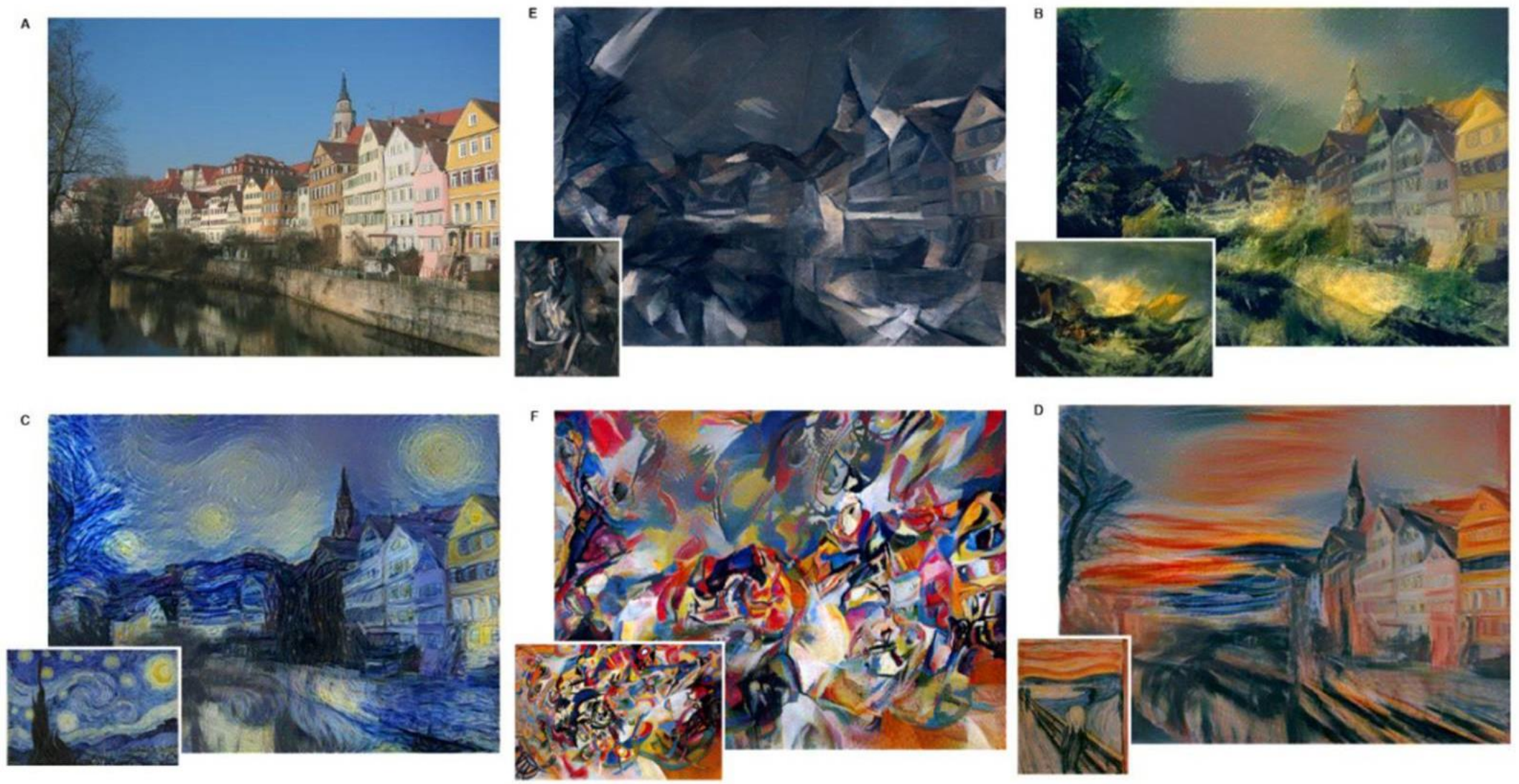
Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GAN)

- Ian Goodfellow



Painting like Van Gogh



Super Resolution

bicubic
(21.59dB/0.6423)

SRResNet
(23.53dB/0.7832)

SRGAN
(21.15dB/0.6868)

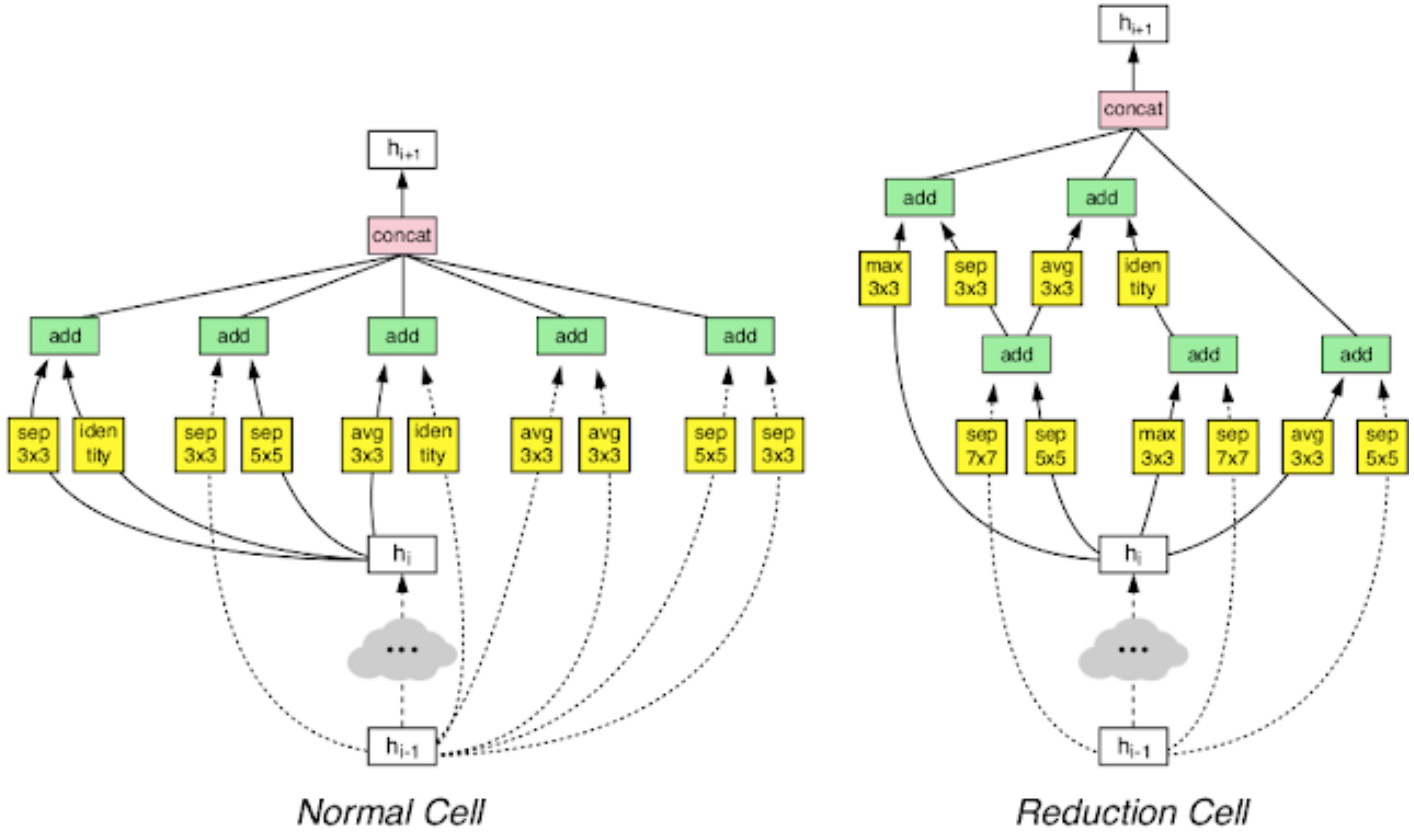
original

Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

DeepFake: Is this you?

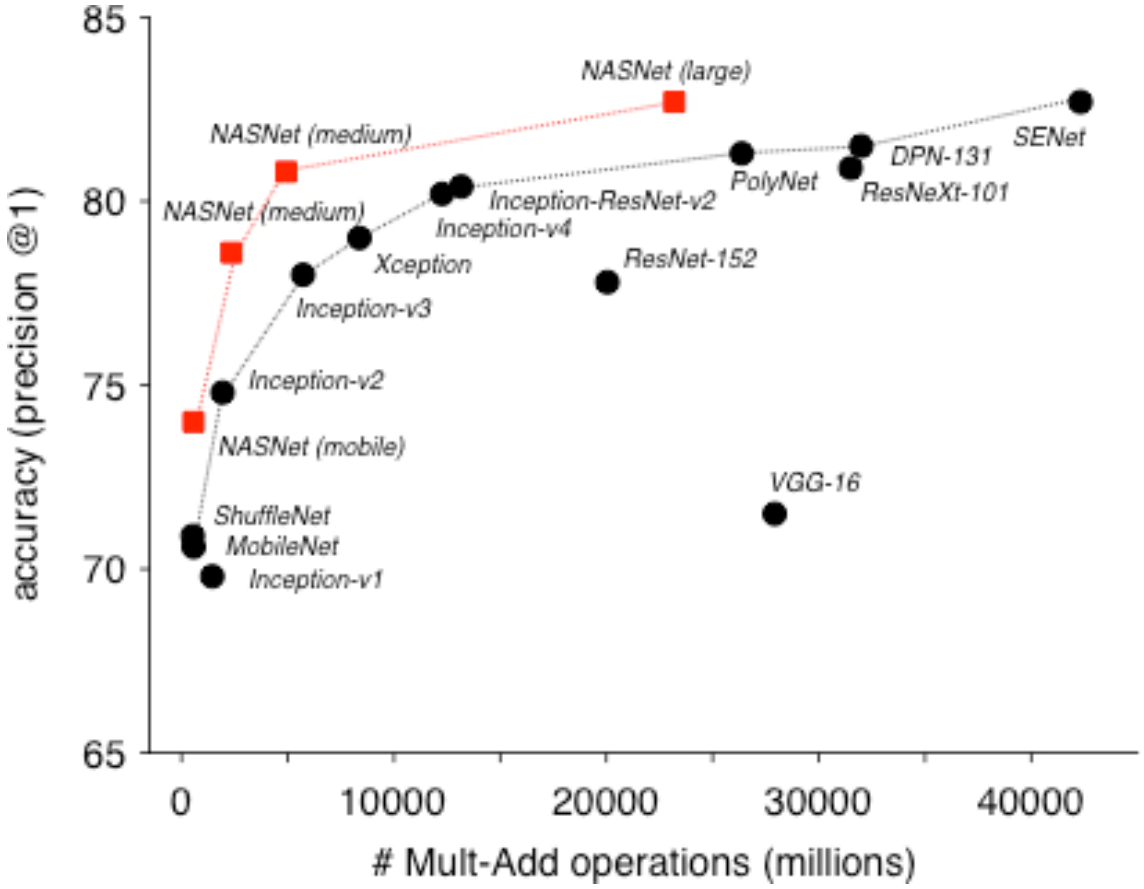
Google's AutoML

- Learning neural network cells automatically

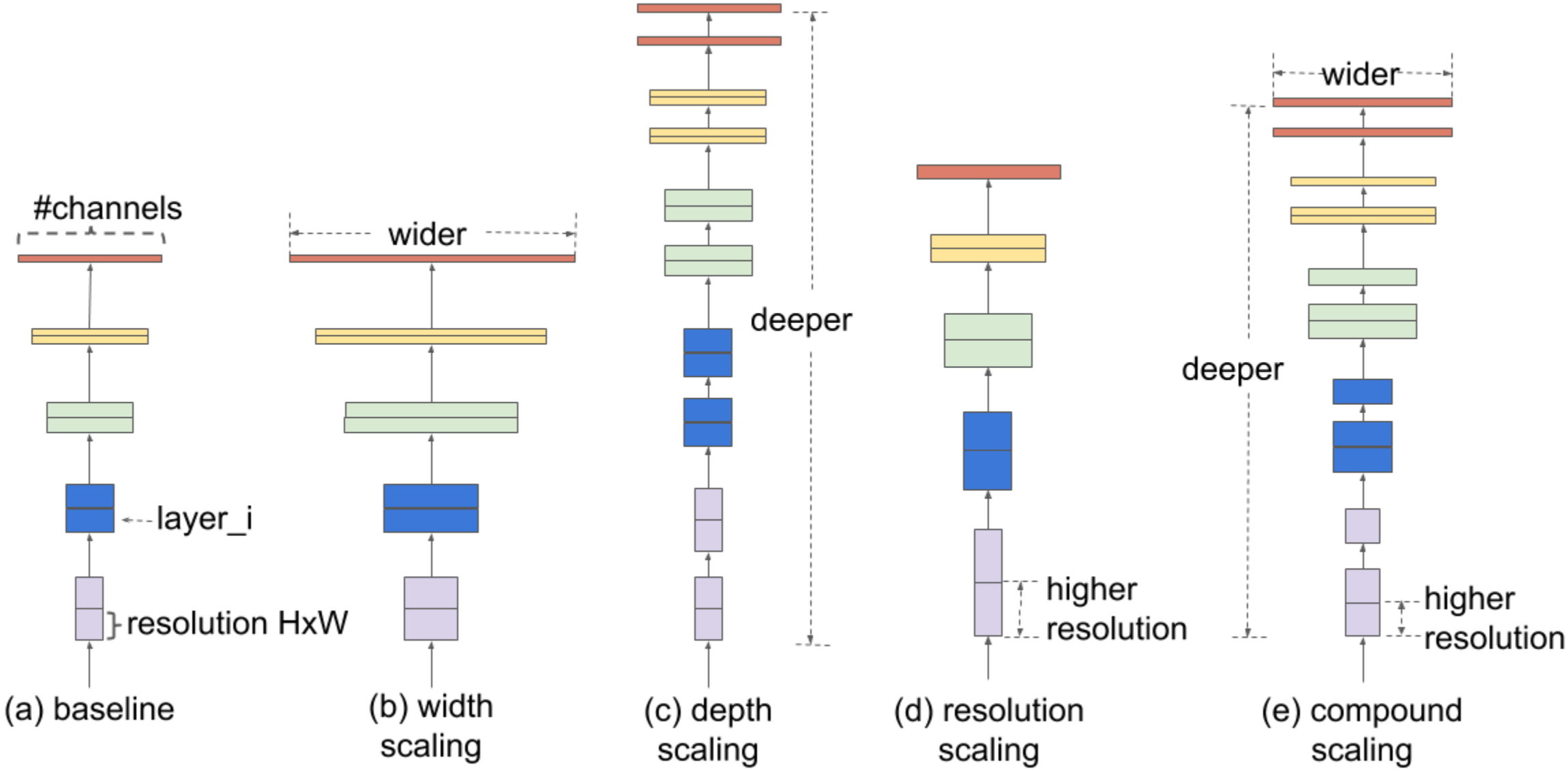


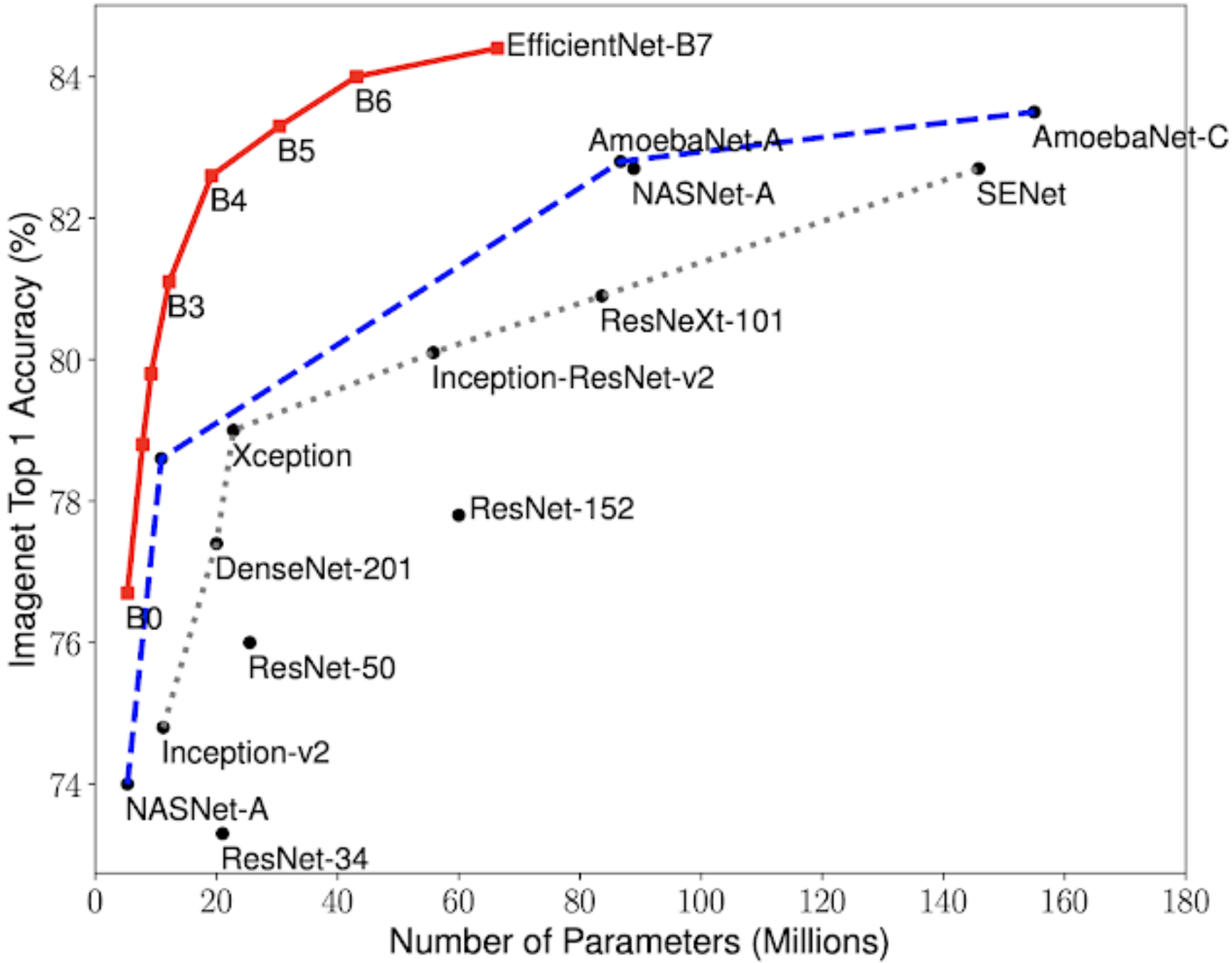
https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html

AutoML on ImageNet



EfficientNet (May, 2019)




EfficientNet-B7
B6
B5
B4
B3
B0
AmoebaNet-A
AmoebaNet-C
NASNet-A
SENet
ResNeXt-101
Inception-ResNet-v2
Xception
ResNet-152
DenseNet-201
ResNet-50
Inception-v2
NASNet-A
ResNet-34
Imagenet Top 1 Accuracy (%)
84
82
80
78
76
74
0
20
40
60
80
100
120
140
160
180
Number of Parameters (Millions)

References

- Francois Chollet, “Deep Learning with Python.” Chapter 1
- What is backpropagation really doing? ( 3Blue1Brown) https://www.youtube.com/watch?v=Ilg3gGewQ5U
- http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/
- https://pmirla.github.io/2016/08/16/AI-Winter.html
- https://tw.saowen.com/a/6cdc2f1279016e566832bb1234e06d321992dd1fabcdf4a2e0a3e16fc0dc09dc
- https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html
- https://hackernoon.com/gradient-descent-aynk-7cbe95a778da
- http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf