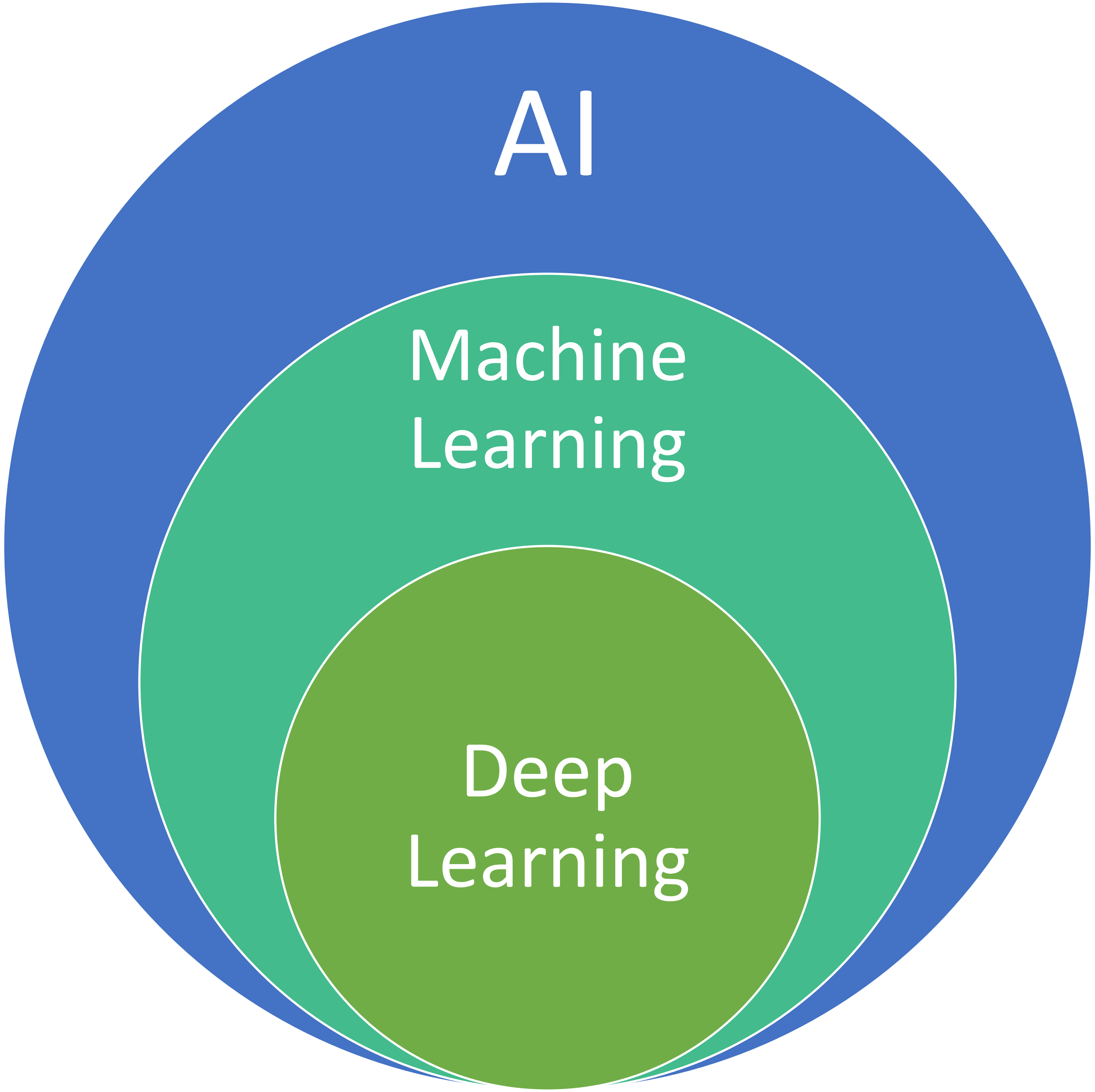
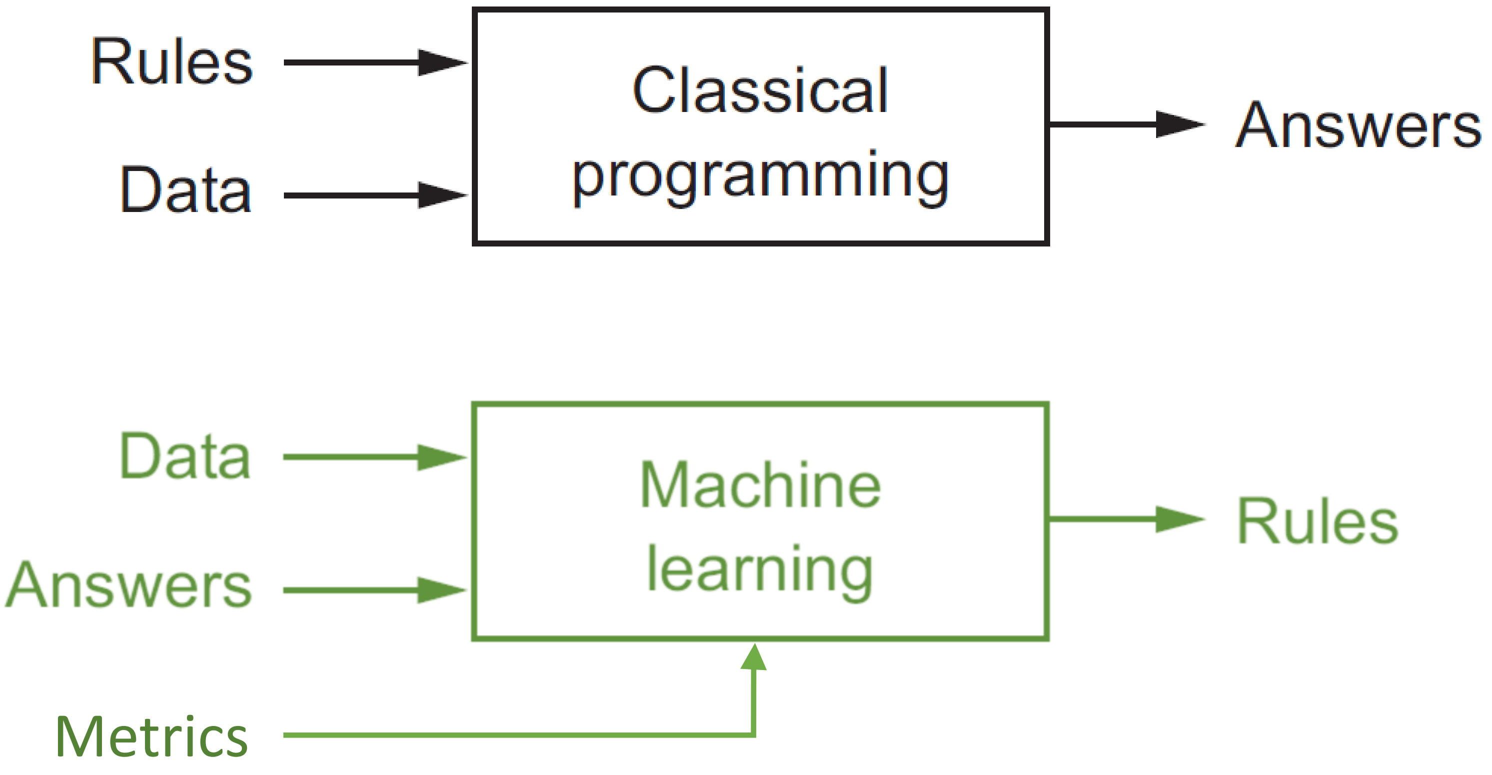
Introduction to Deep Learning

Prof. Kuan-Ting Lai

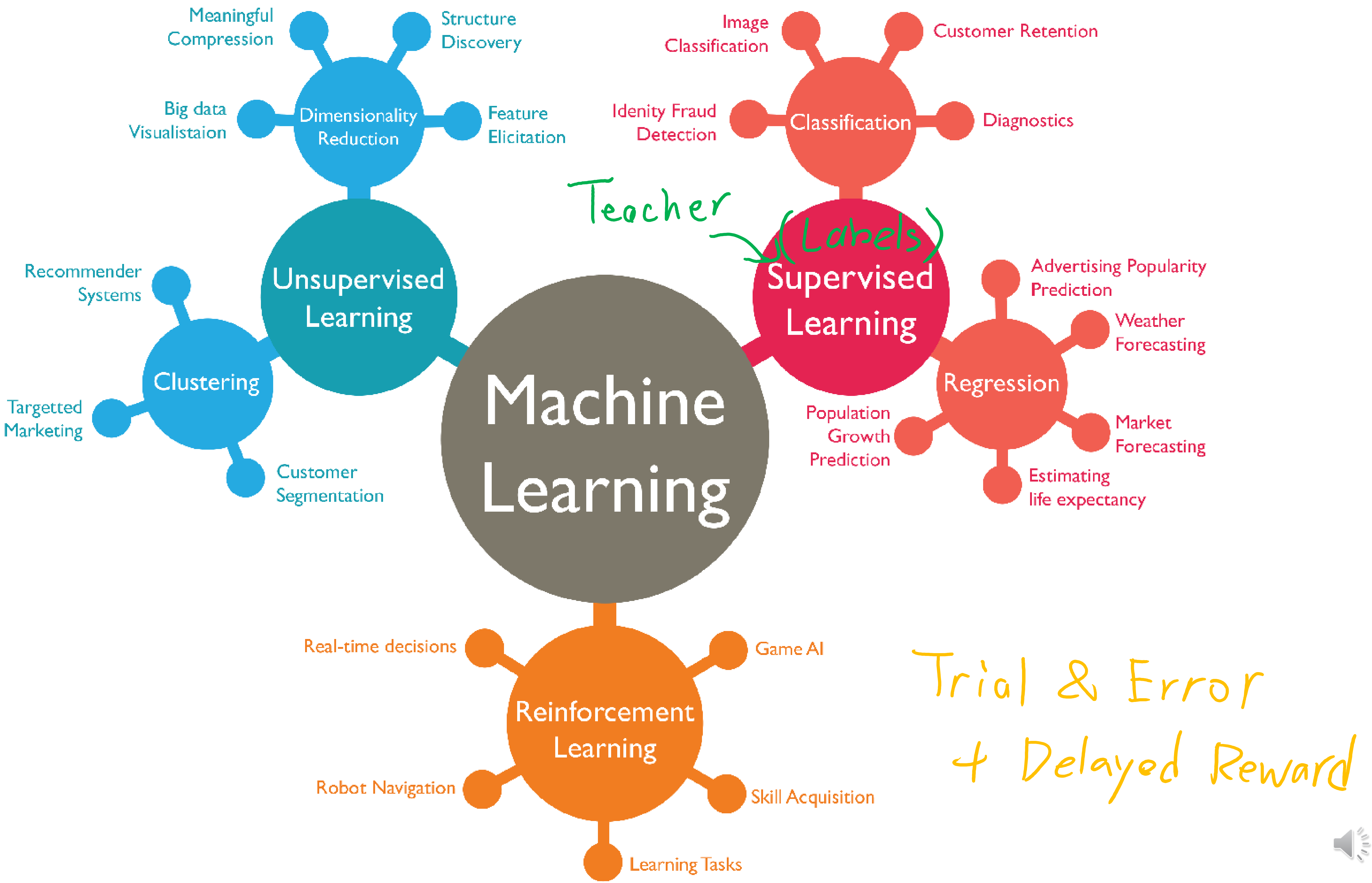
2021/9/28

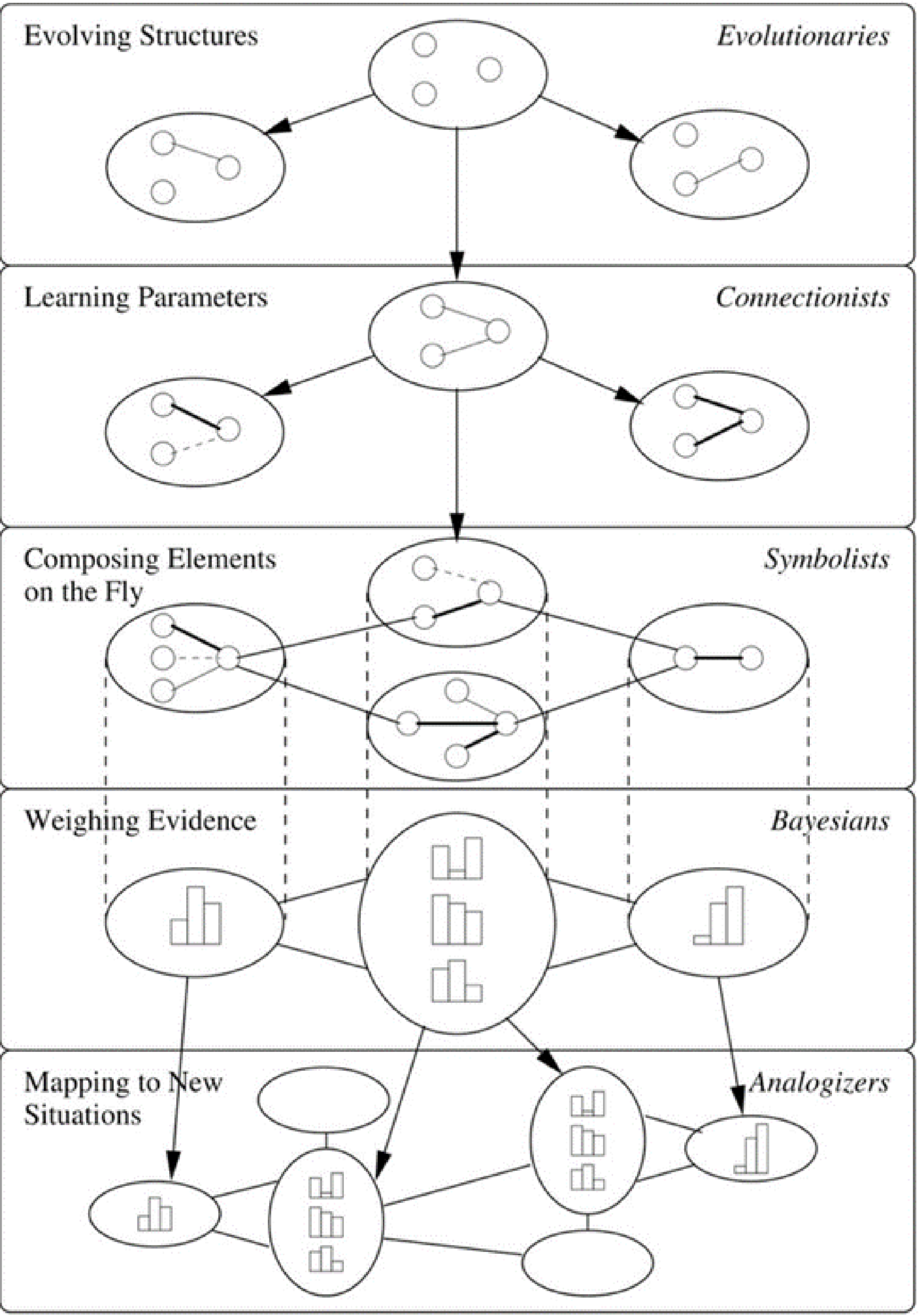

AI
Machine Learning
Deep Learning

Machine Learning (Statistical Learning)



Francois Chollet, “Deep Learning with Python,” Manning, 2017


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

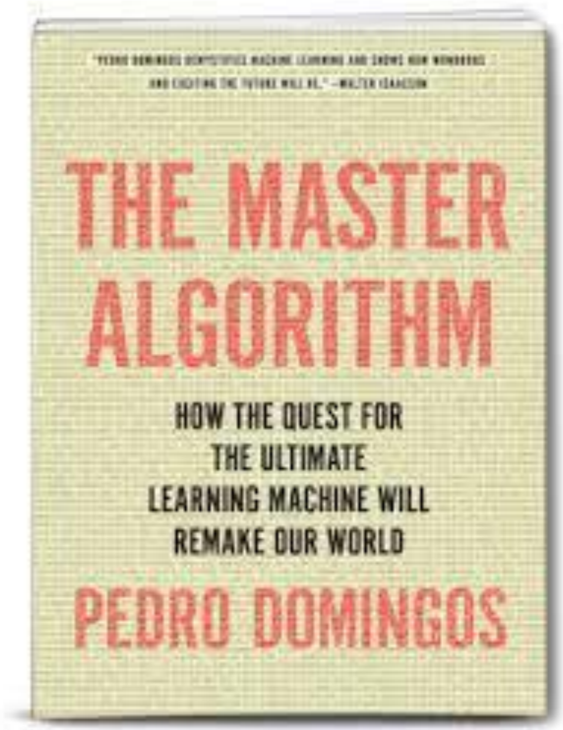


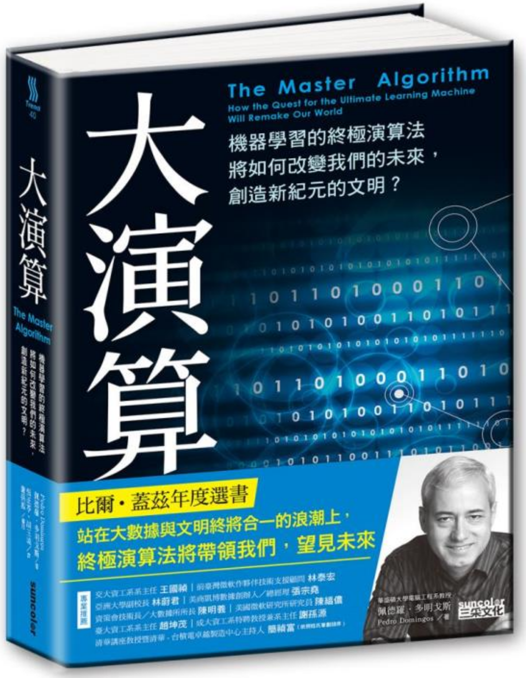
Five Tribes of Machine Learning

- Evolutionaries
- Connectionists
- Symbolists
- Bayesians
- Analogizers

Pedro Domingos

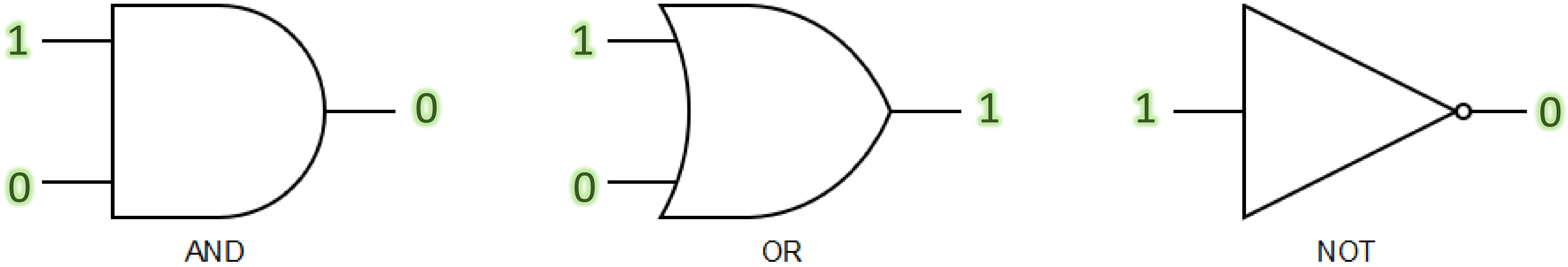
The Master Algorithm – Pedro Domingos



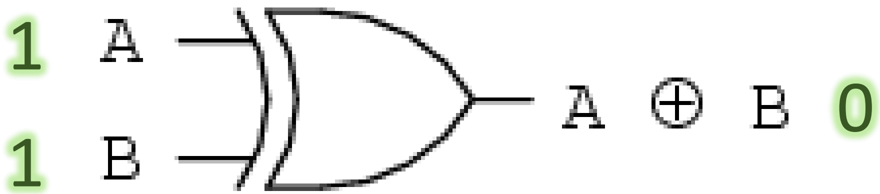


3 Basic Operations of Algorithms

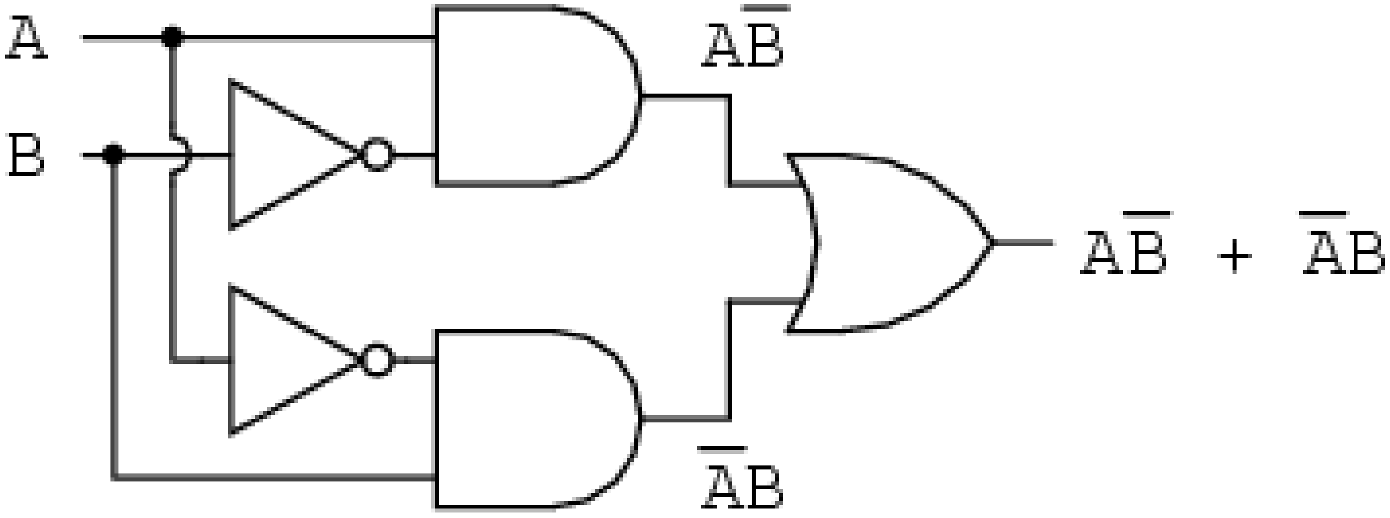
- All Algorithms can be Reduced to 3 Operations



XOR

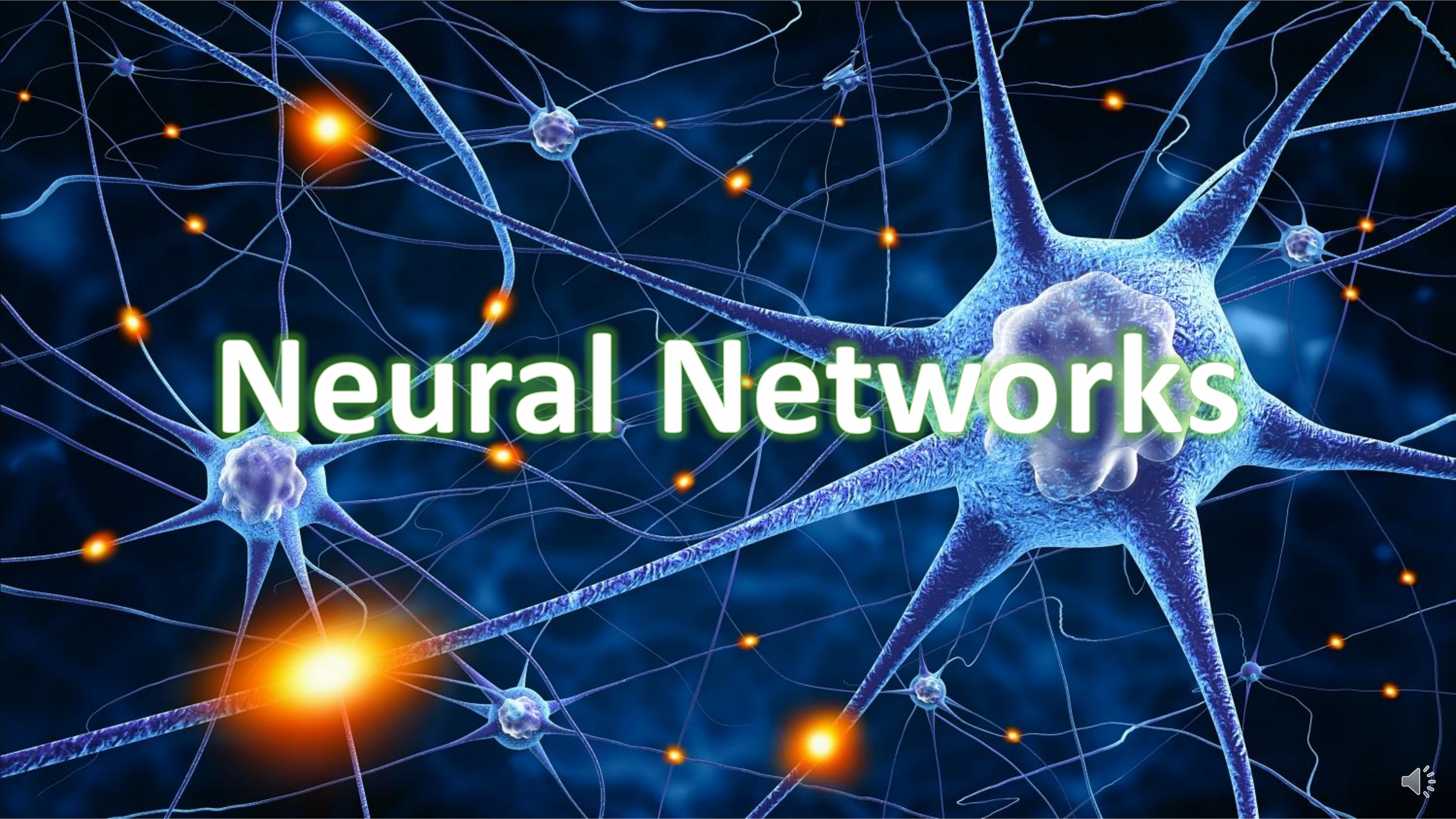


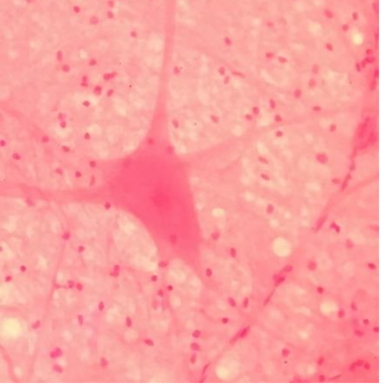
. . . is equivalent to . . .

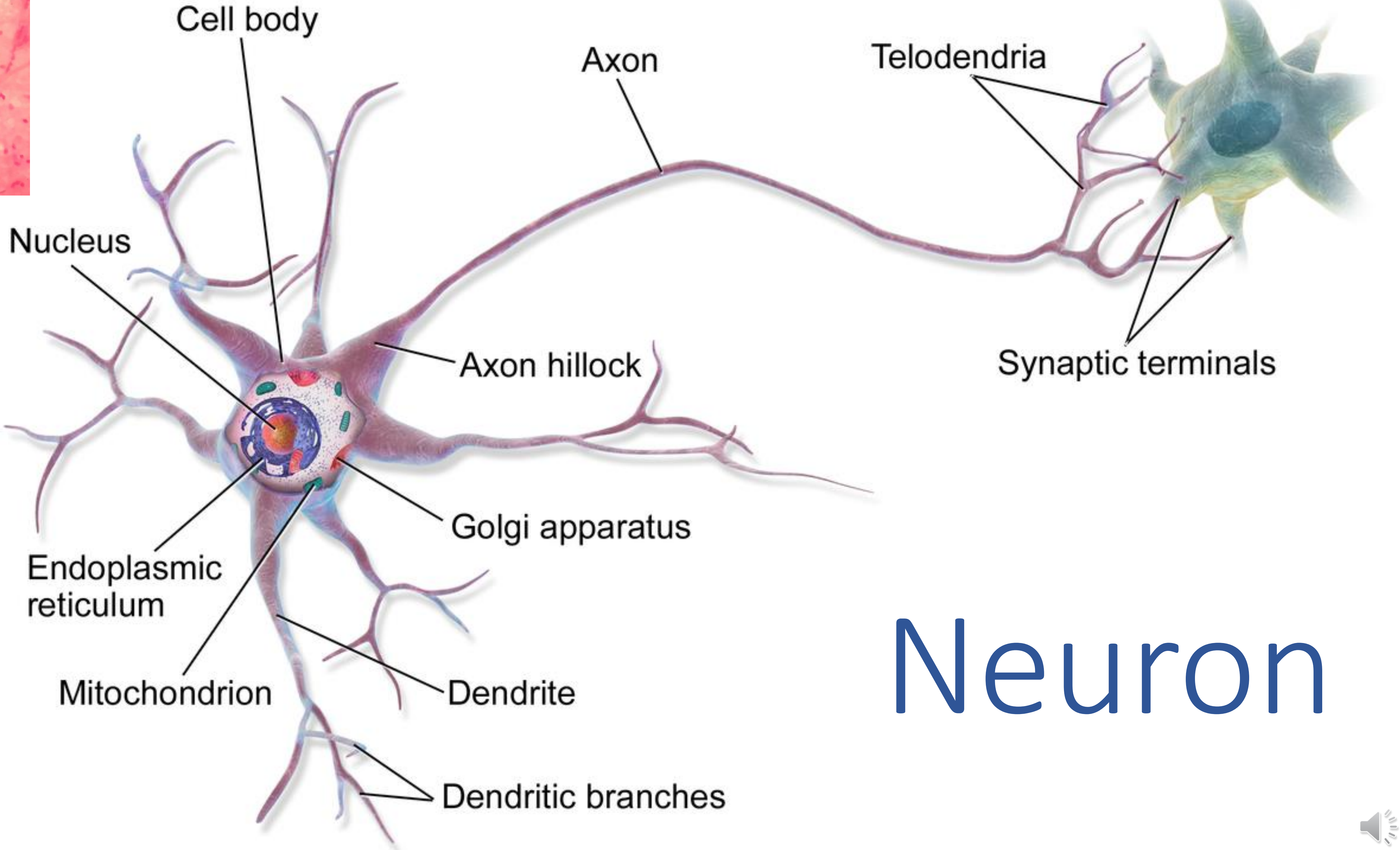


$$A \oplus B = A\overline{B} + \overline{A}B$$

Neural Networks






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

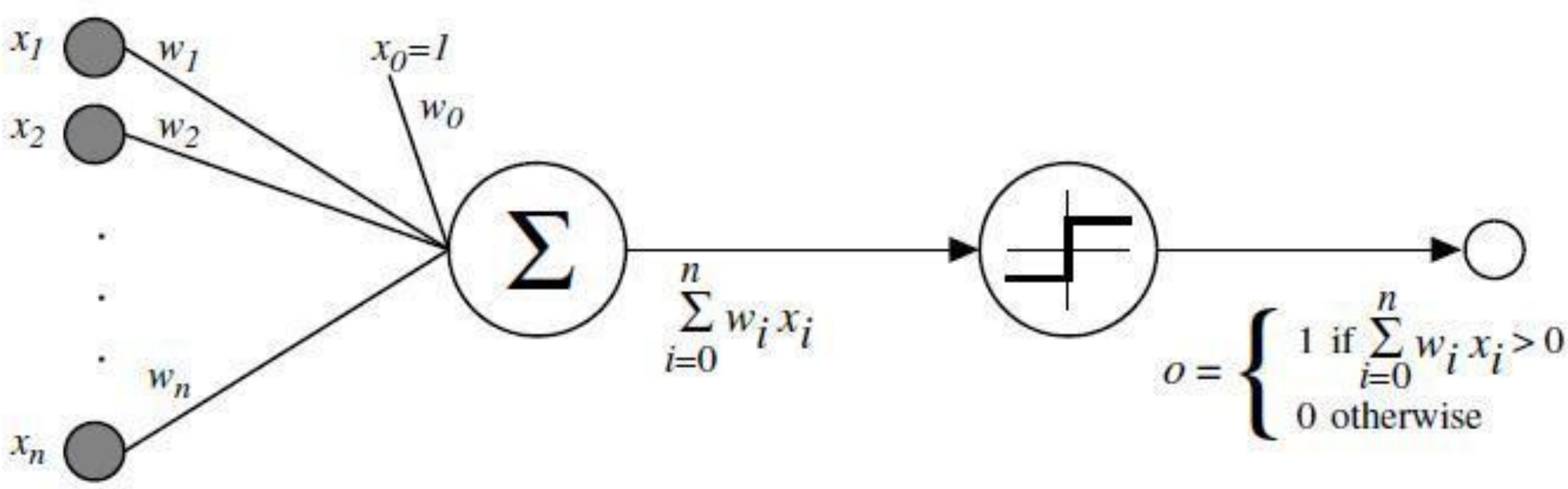
Neuron

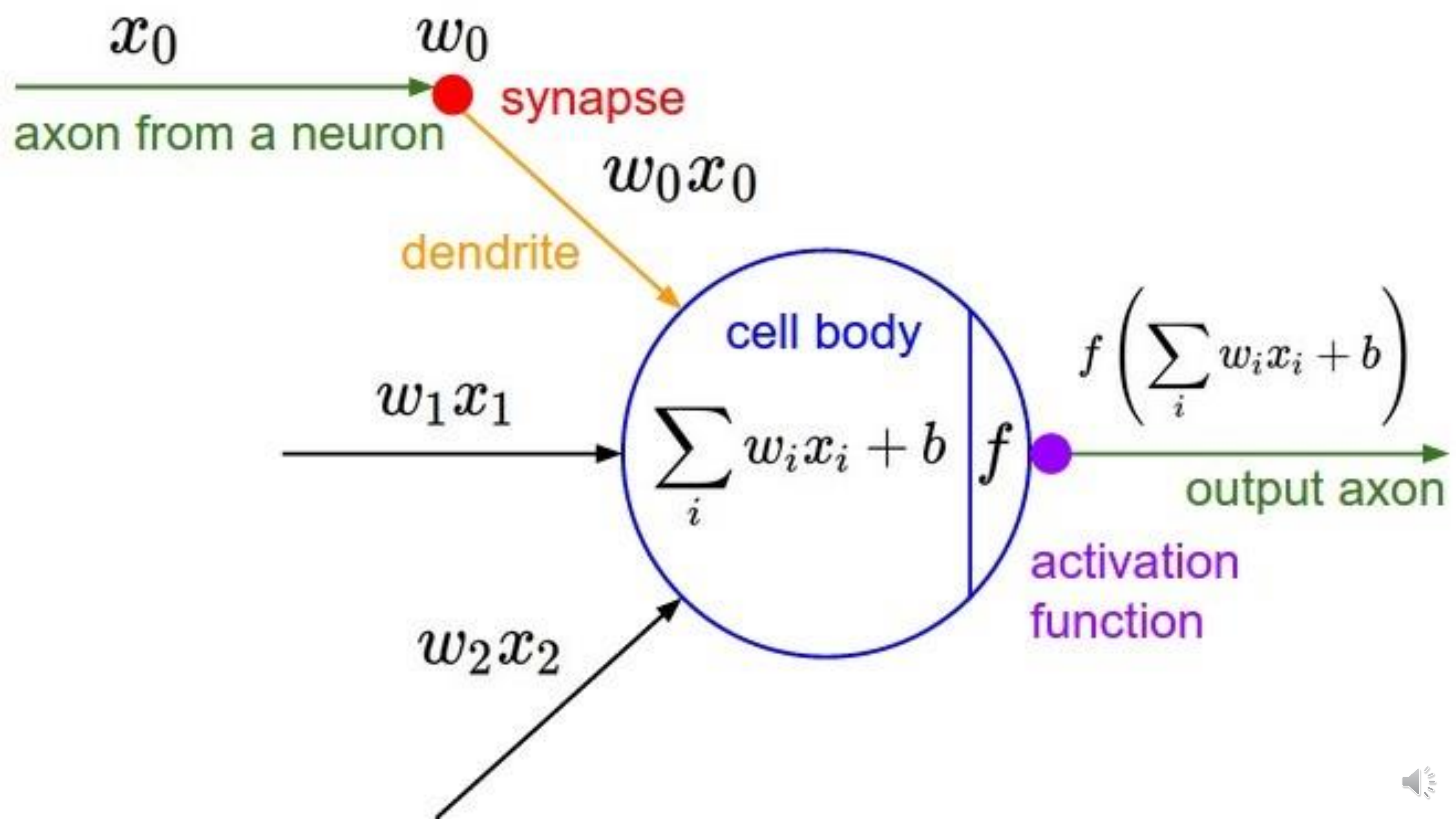
Number of Connections in the Brain

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

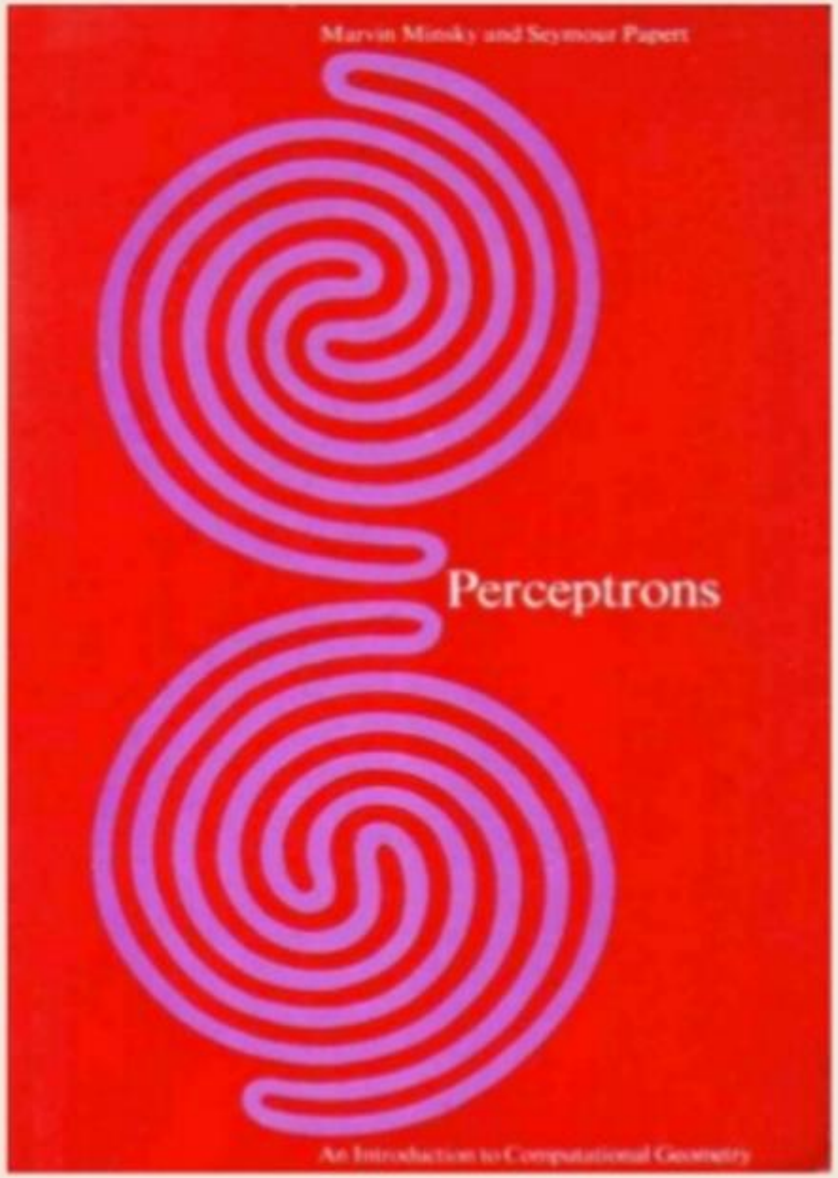
Synapses (based on 1000 per neuron):
10^{14}, or 100 trillion, 100000000000000

Frank Rosenblatt's Perceptron (1957)

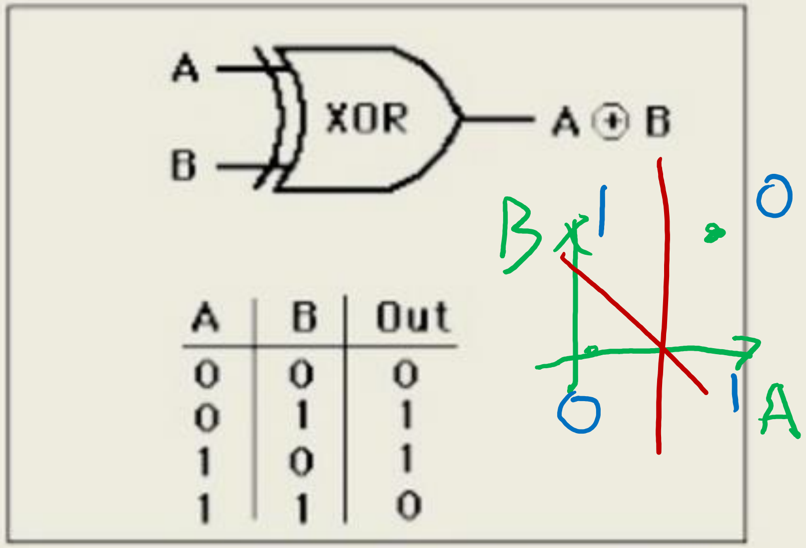



x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$f\left(\sum_i w_i x_i + b\right)$
$w_1 x_1$
$\sum_i w_i x_i + b$
f
output axon
activation
function
$w_2 x_2$

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

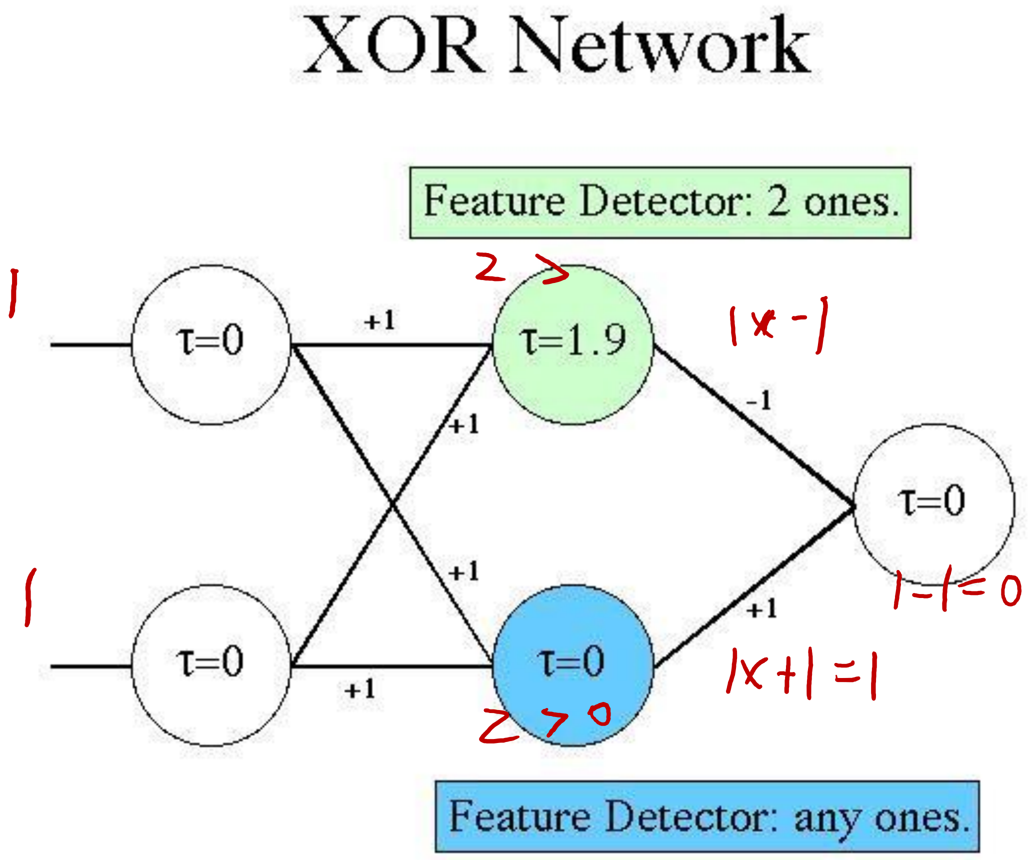
AI Winter
1969 - 1990

Deep Learning

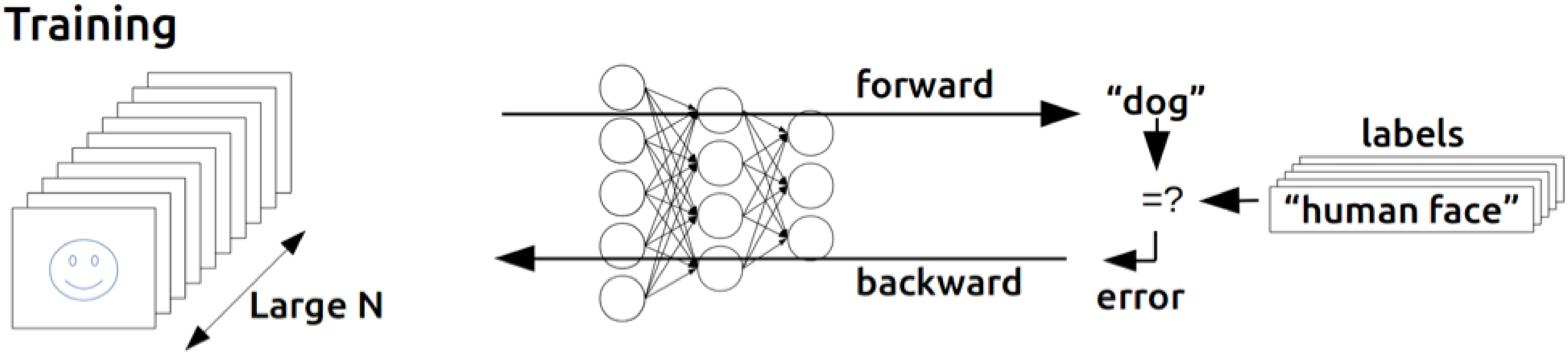
Geoffrey Hinton (Toronto, Google) | Yann LeCun (NewYork, Facebook) | Yoshua Bengio (Montreal)

Learning XOR (1986)

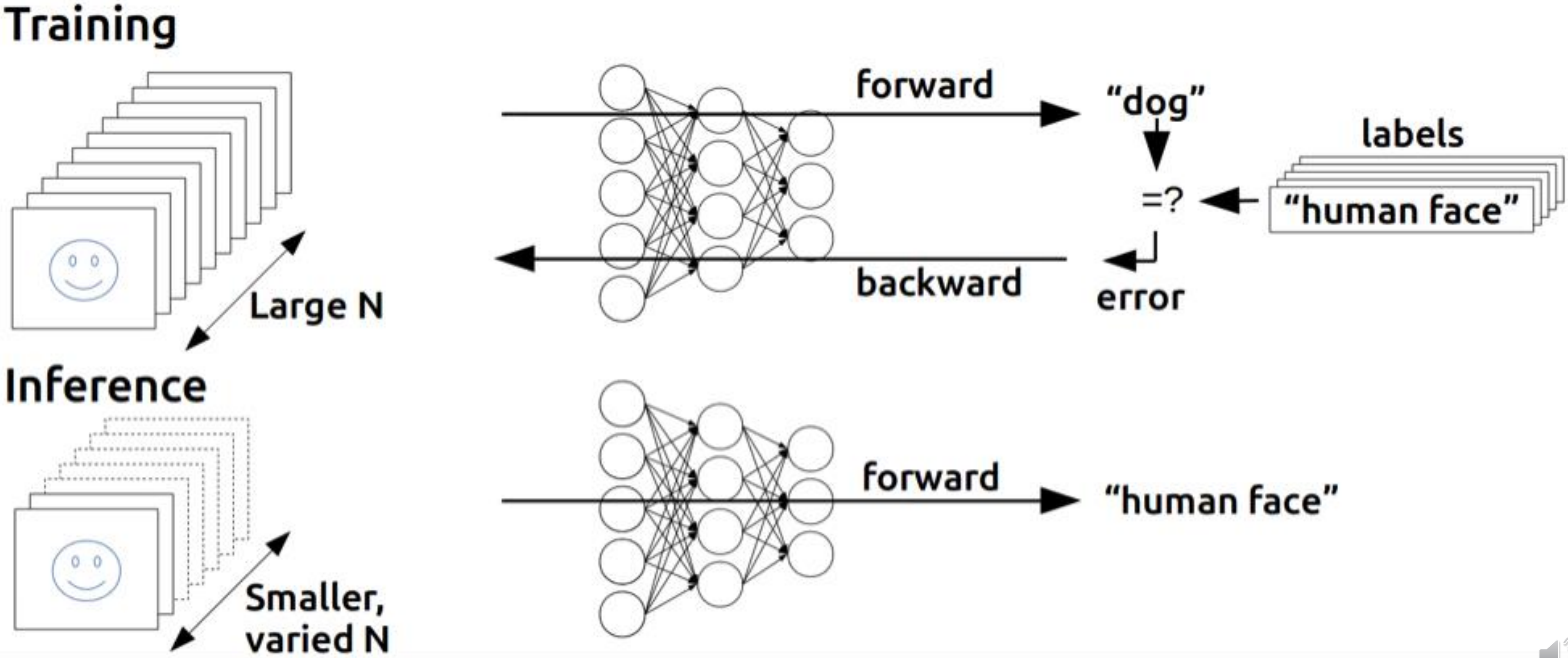
Geoffrey Hinton



Backpropagation



Inference


Training
forward
“dog”
labels
=?
“human face”
backward
error
Large N
Inference
forward
“human face”
Smaller,
varied N

Chain Rule

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

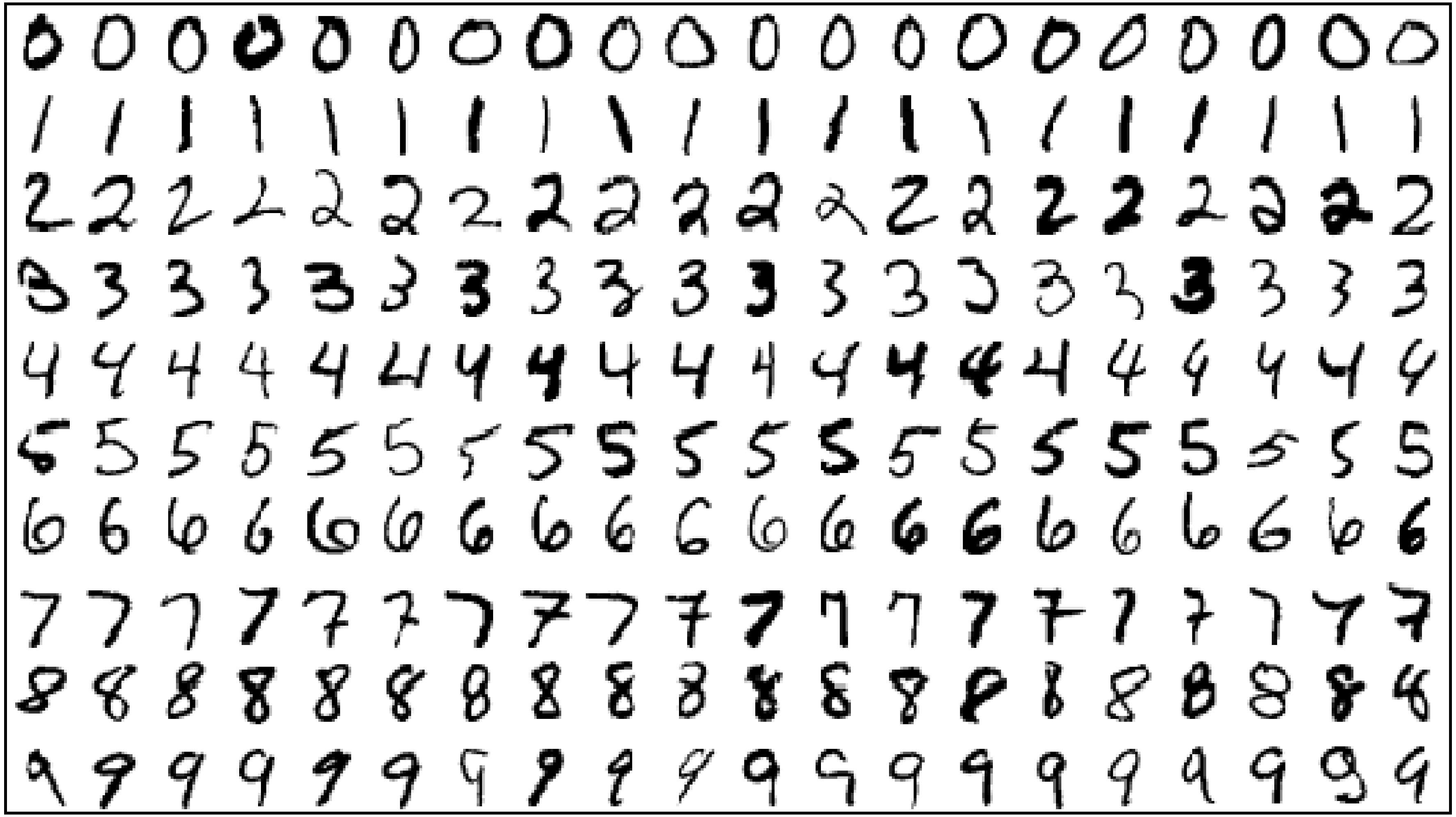
$$\frac{d^2y}{dx^2} = \frac{d^2y}{du^2}\left(\frac{du}{dx}\right)^2 + \frac{dy}{du}\frac{d^2u}{dx^2}$$

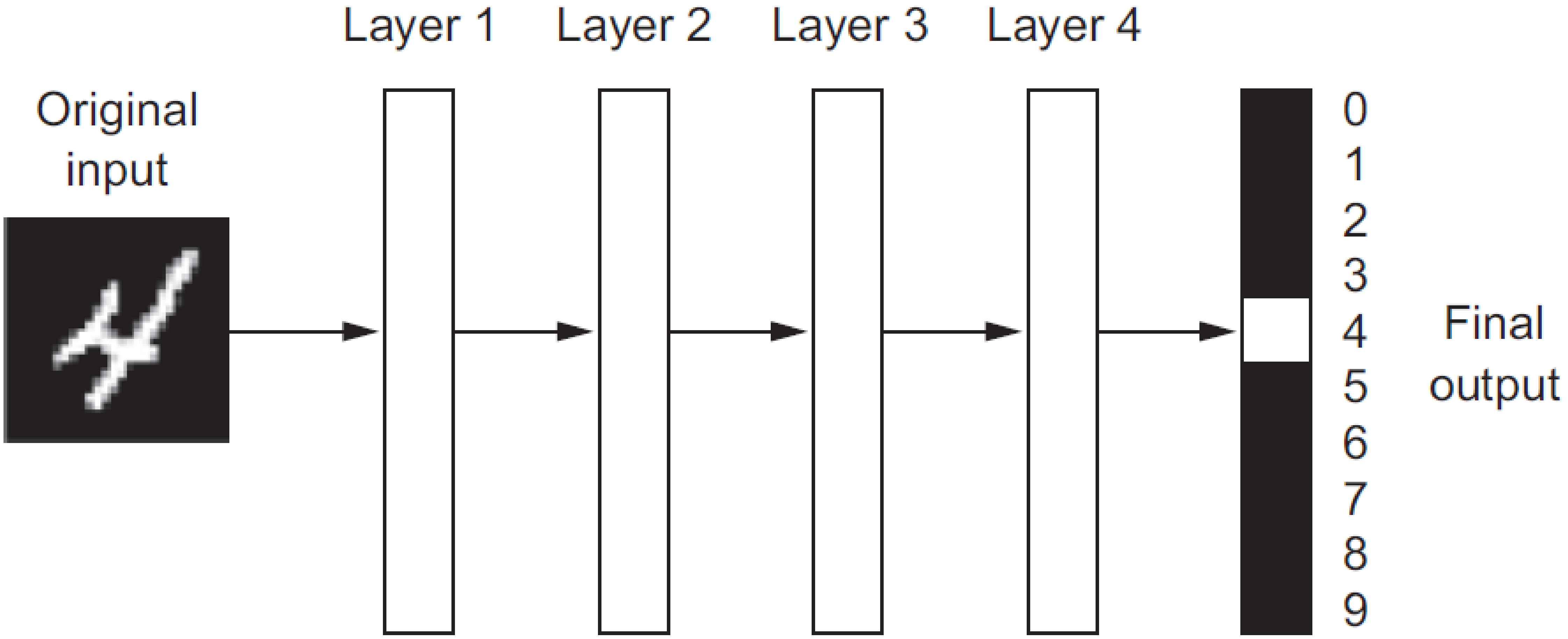
$$\frac{d^3y}{dx^3} = \frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^3 + 3\,\frac{d^2y}{du^2}\frac{du}{dx}\frac{d^2u}{dx^2} + \frac{dy}{du}\frac{d^3u}{dx^3}$$

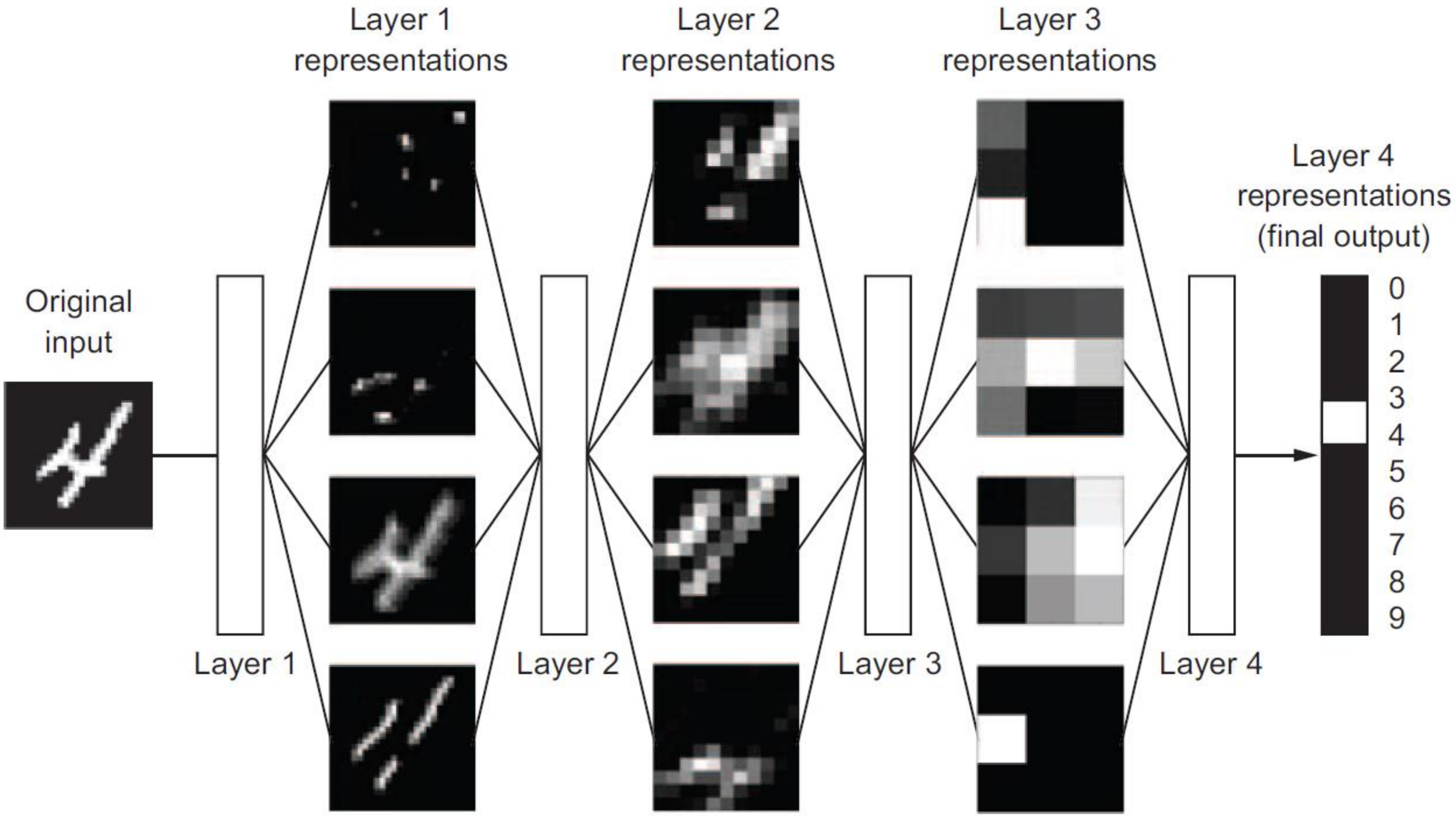
$$\frac{d^4y}{dx^4} = \frac{d^4y}{du^4}\left(\frac{du}{dx}\right)^4 + 6\,\frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^2\frac{d^2u}{dx^2} + \frac{d^2y}{du^2}\left(4\,\frac{du}{dx}\frac{d^3u}{dx^3} + 3\left(\frac{d^2u}{dx^2}\right)^2\right) + \frac{dy}{du}\frac{d^4u}{dx^4}.$$

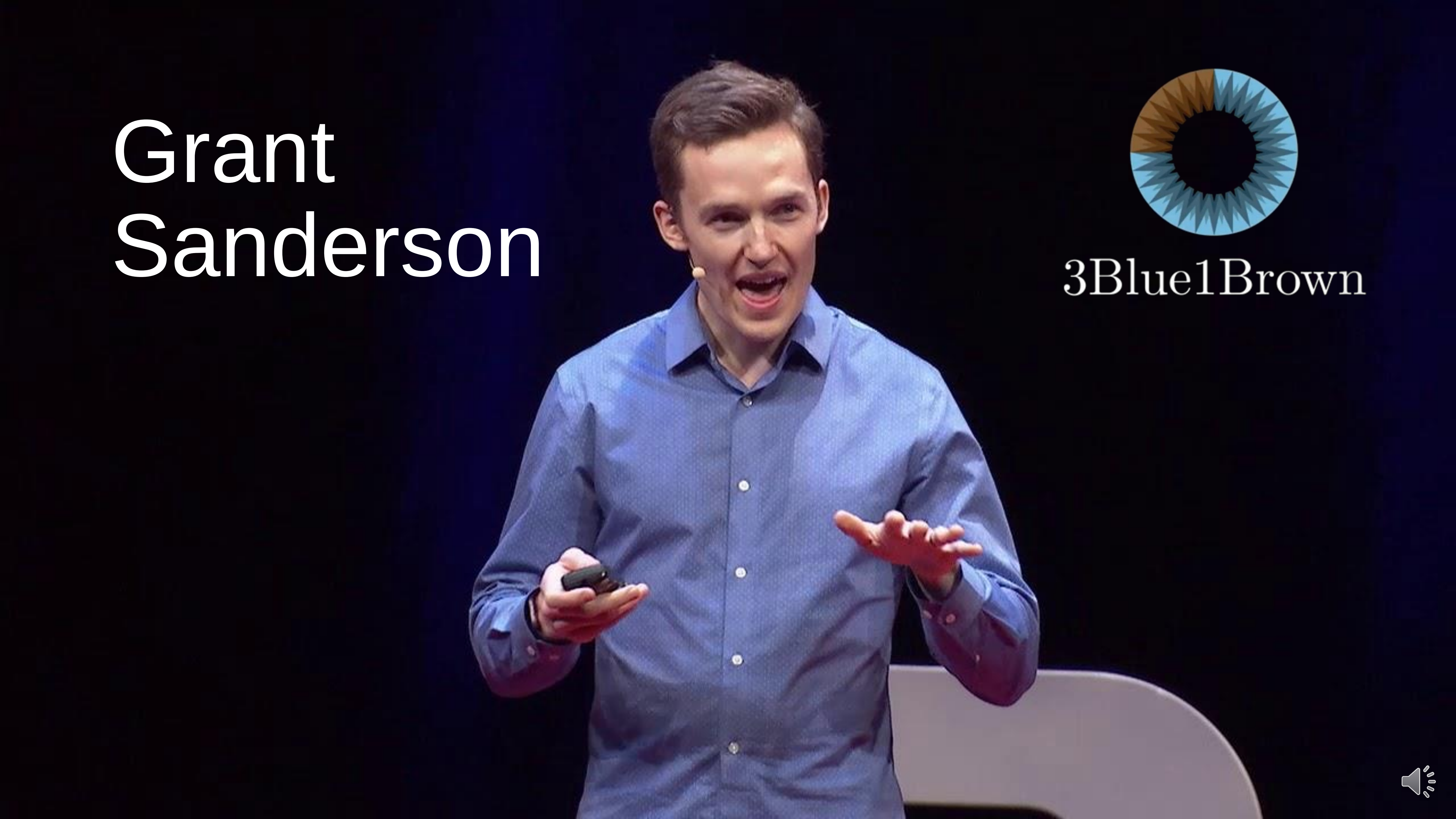
Example: Recognizing Handwritten Digits

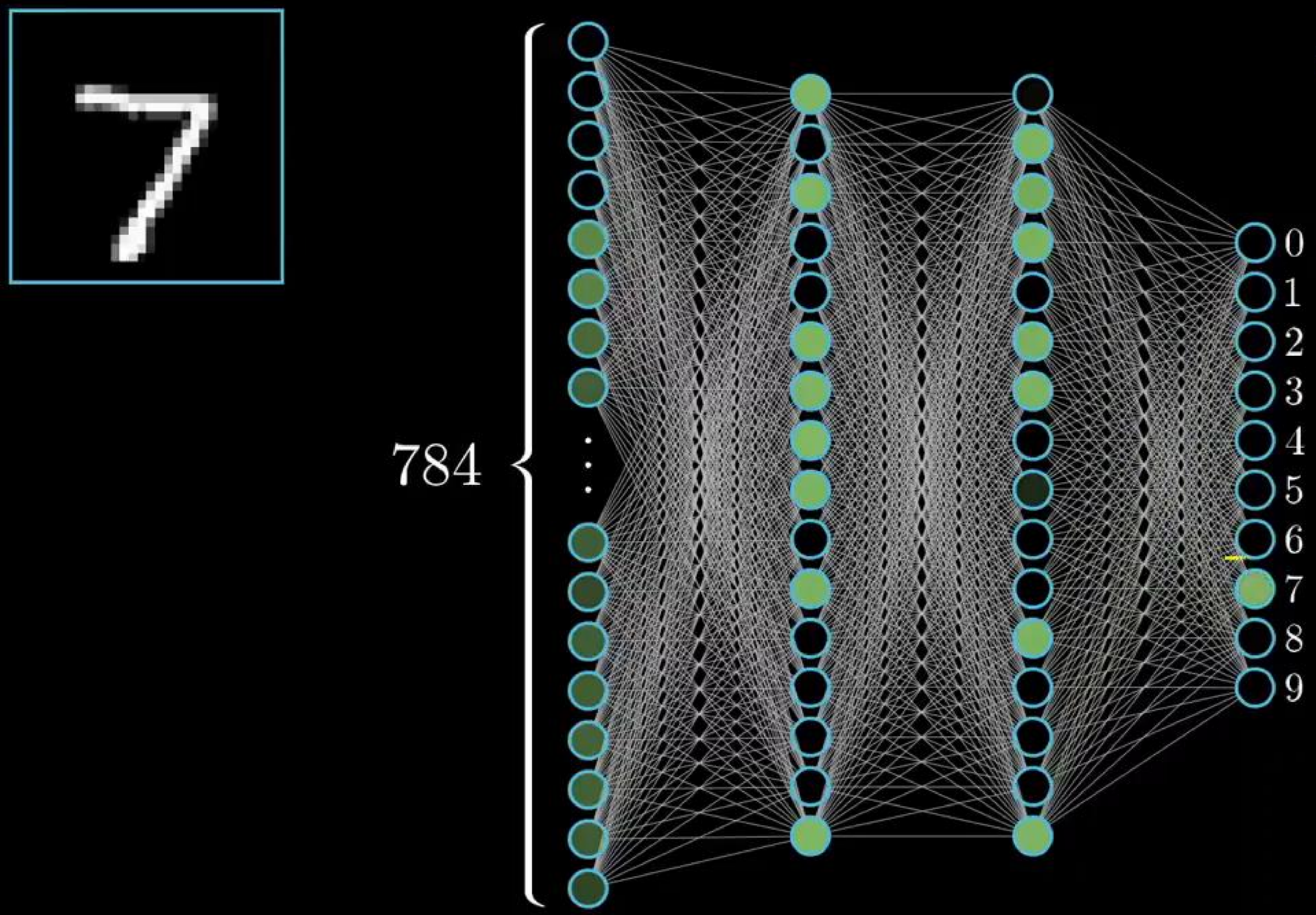
- MNIST dataset



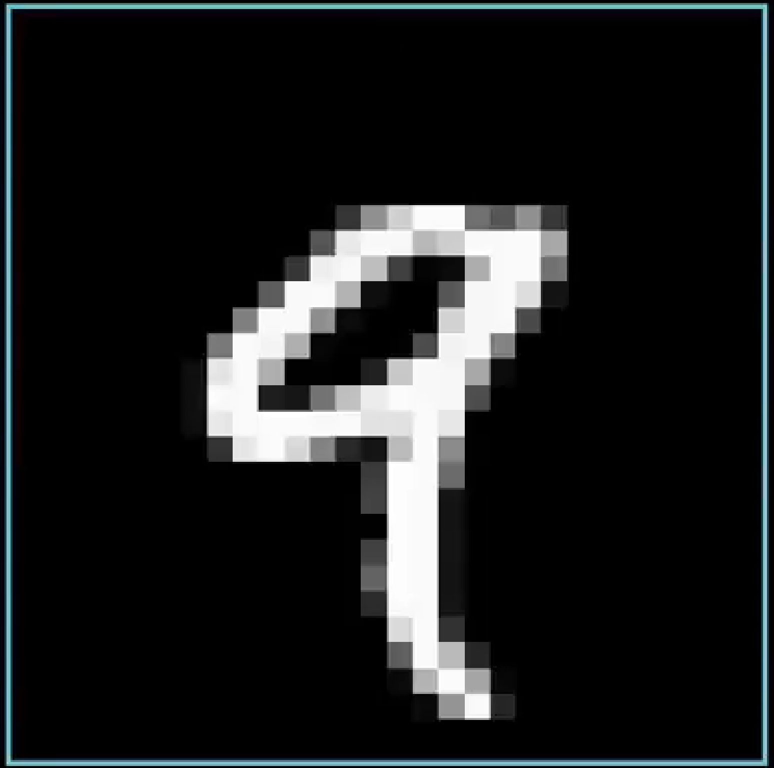

Layer 1
Layer 2
Layer 3
Layer 4
Original input
0
1
2
3
4
5
6
7
8
9
Final output


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

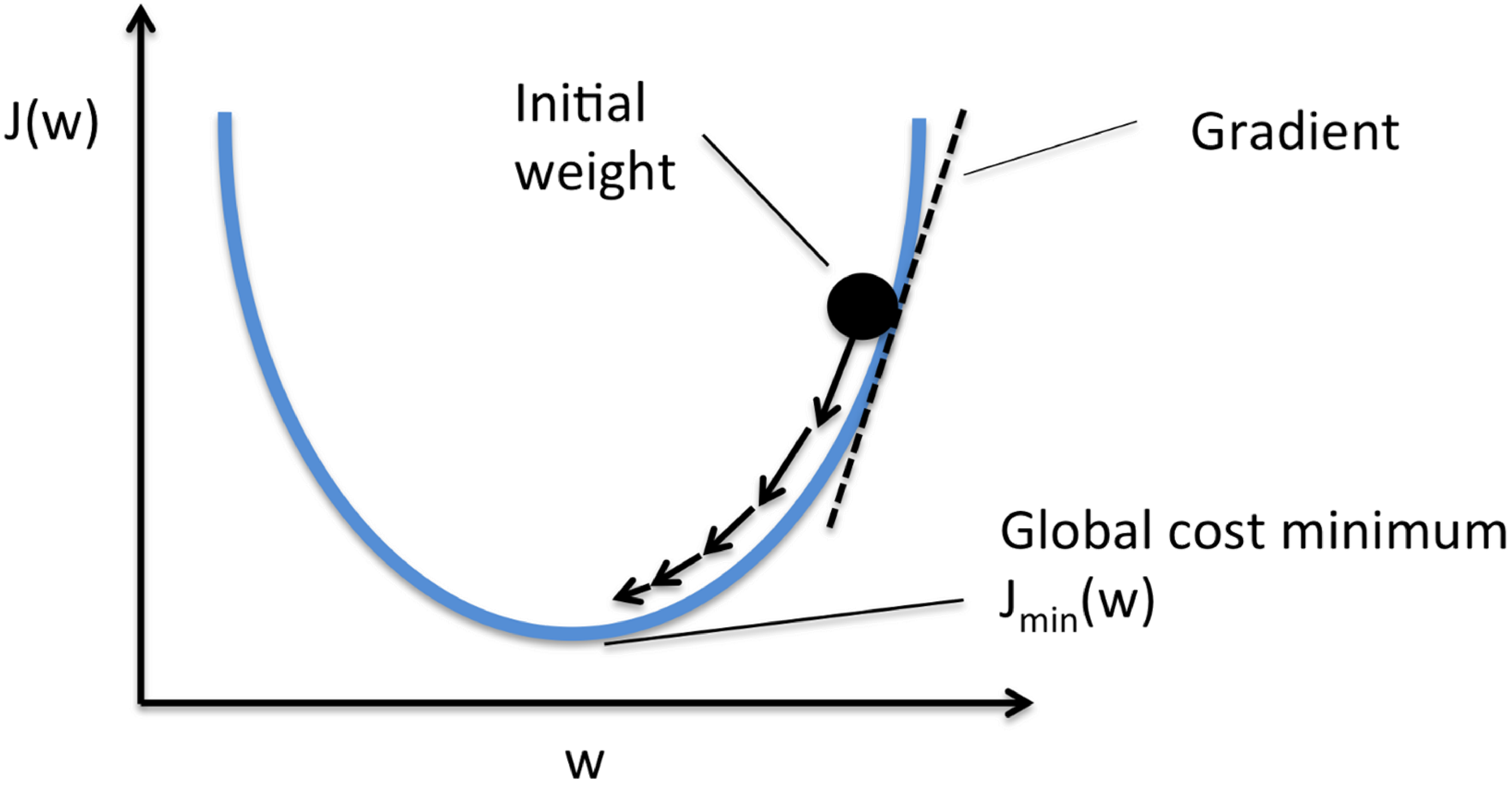

Grant
Sanderson
3Blue1Brown

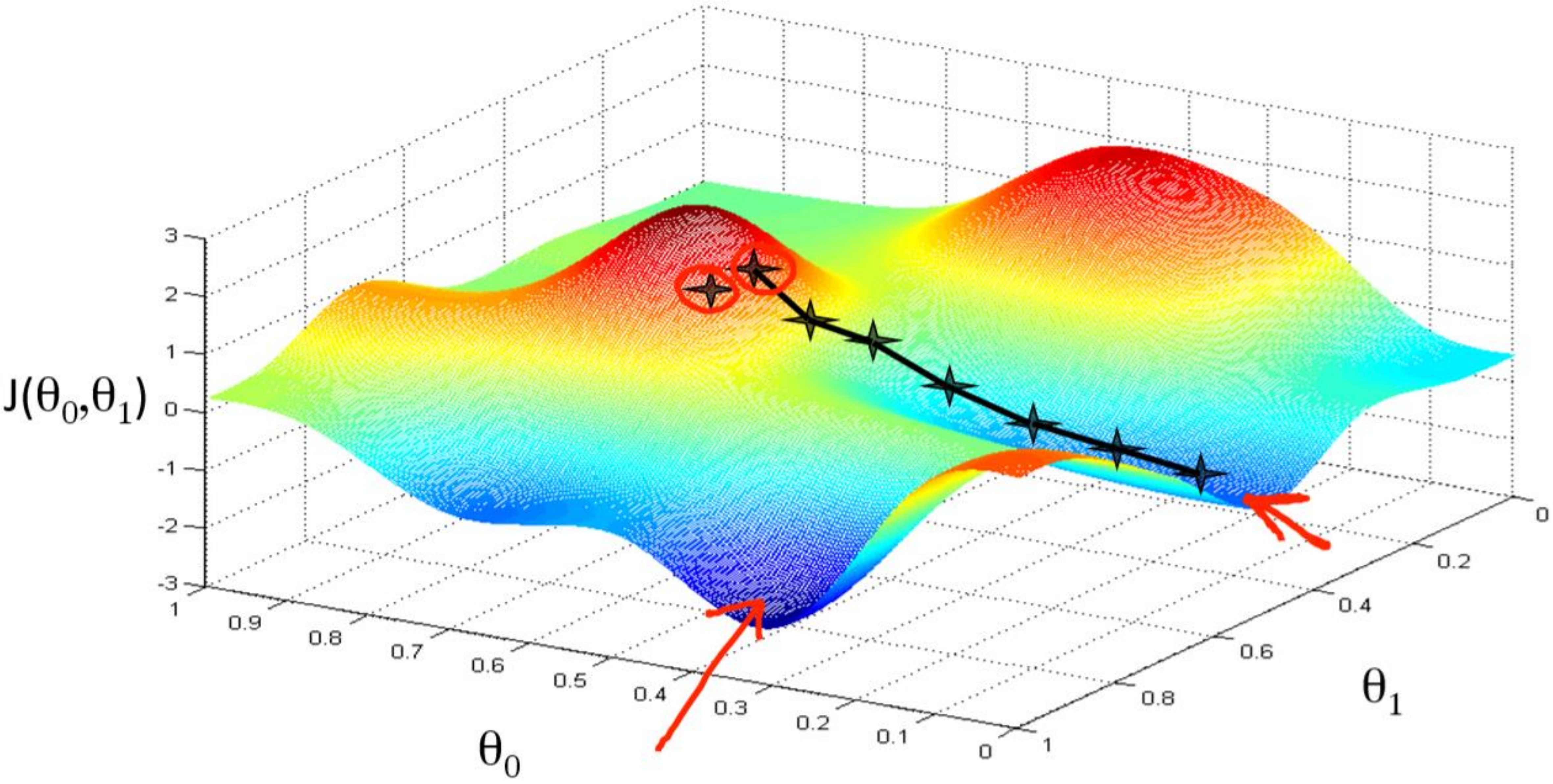


3Blue1Brown



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

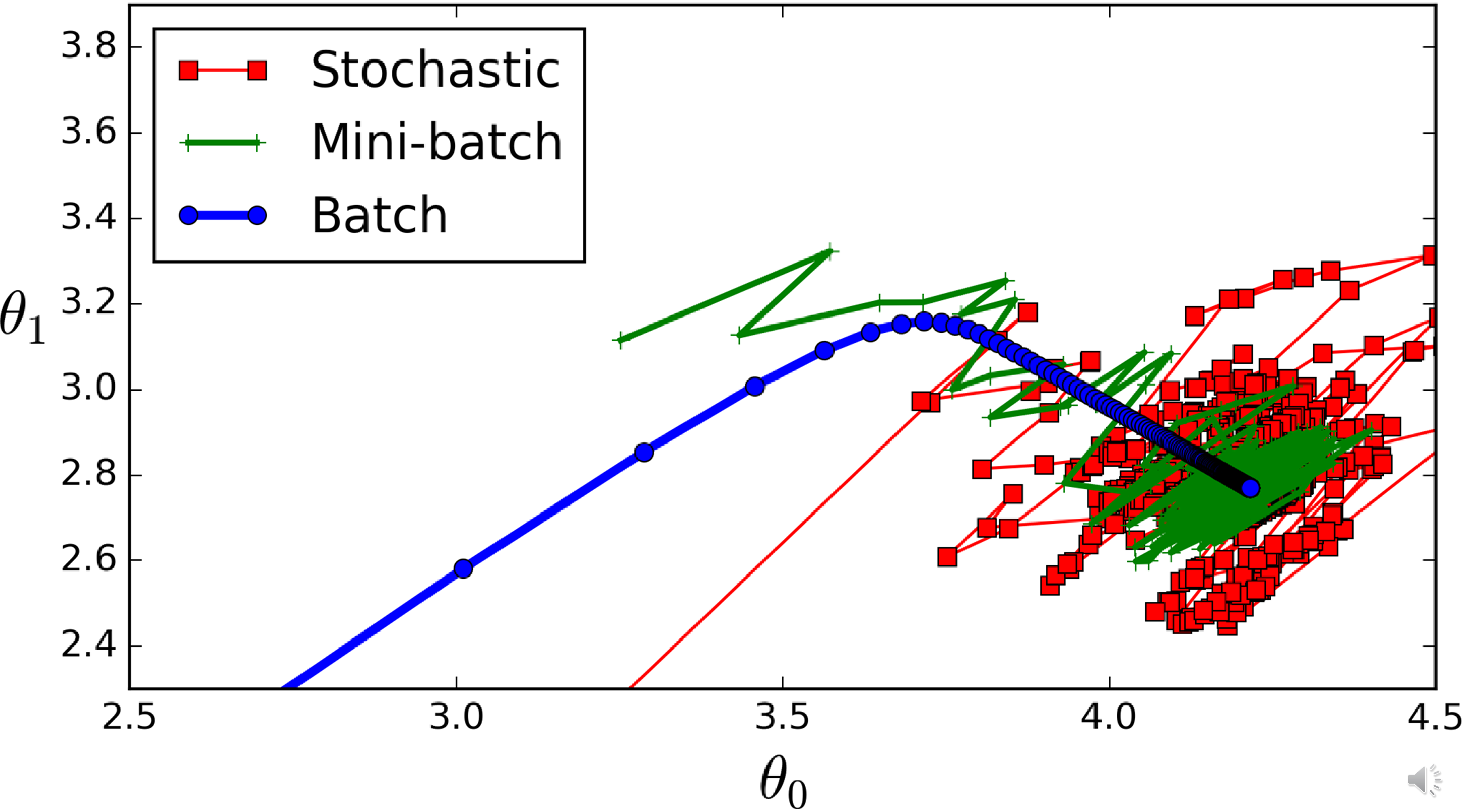
- Gradient of MSE

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

- 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

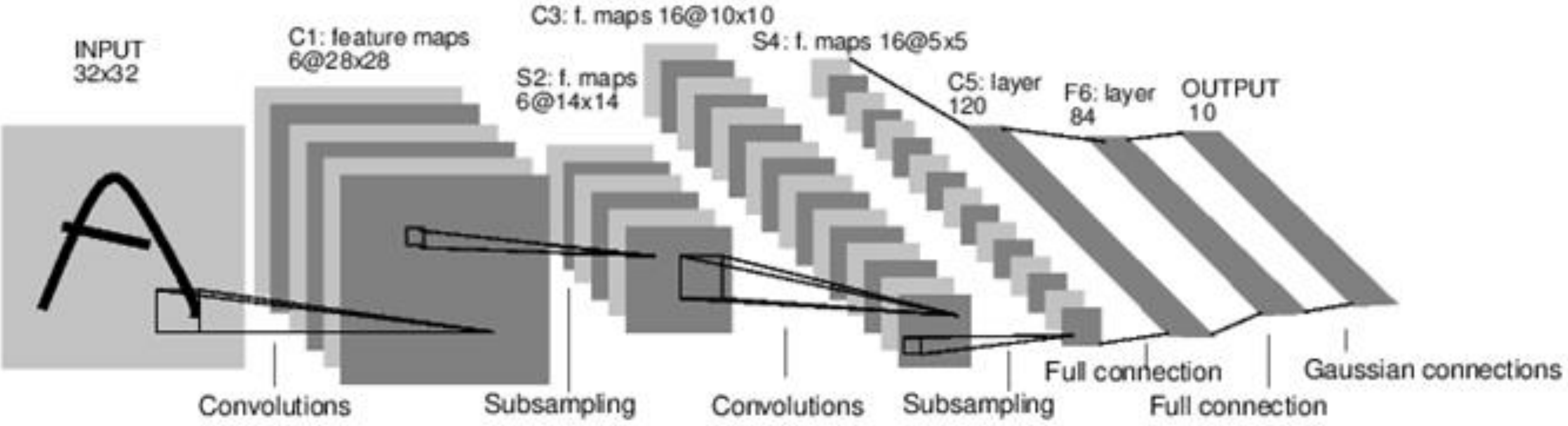
$$C(w_1, w_2, \ldots, w_{13,002})$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{Weights and biases}}$$

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs)

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



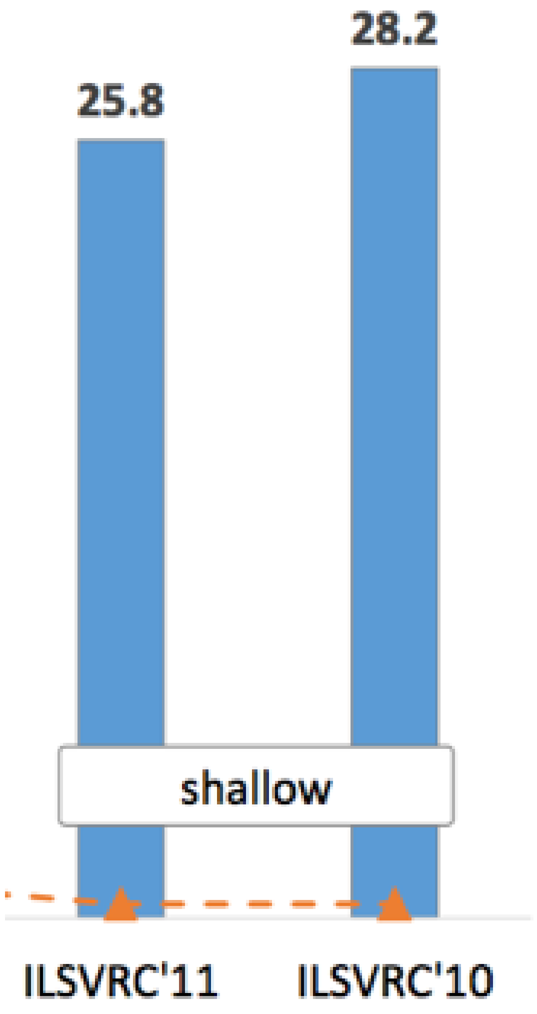
A Full Convolutional Neural Network (LeNet)

14,197,122 images, 21841 classes
(2021/9/21)

ImageNet Large Scale Visual Object Recognition Challenge (ILSVRC)

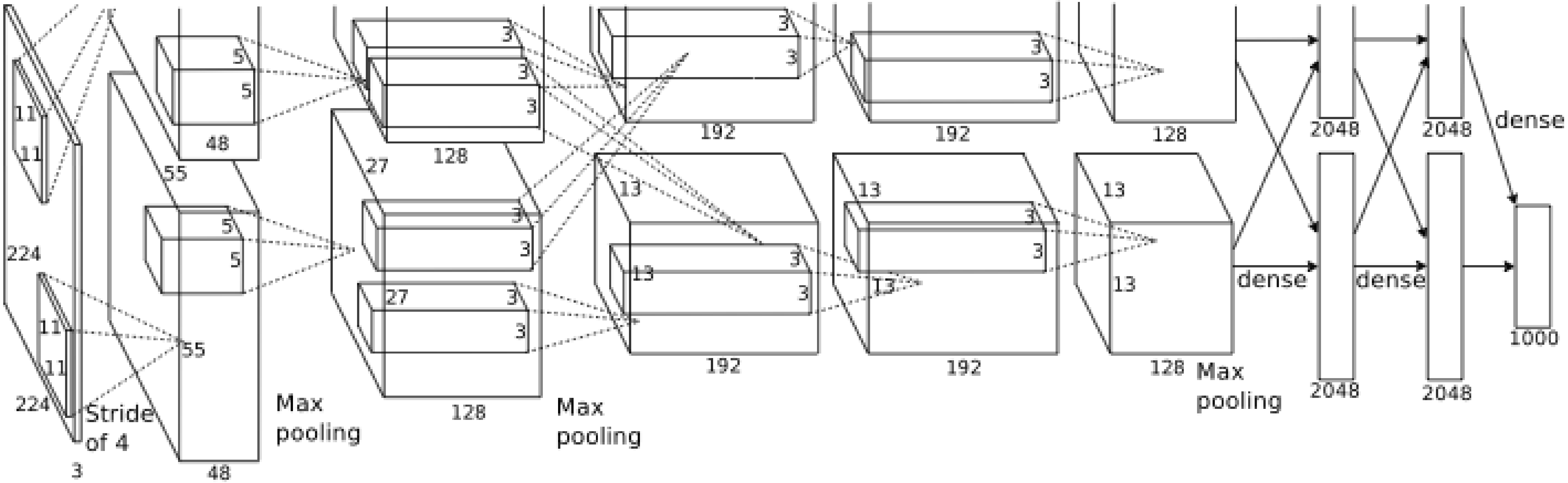
- 1000 categories
- For ILSVRC 2017
  - **Training images** for each category ranges from 732 to 1300
  - 50,000 validation **images** and 100,000 test **images**.
- Total number of images in ILSVRC 2017 is around 1,150,000

Winners' Error Rates on ImageNet Challenge

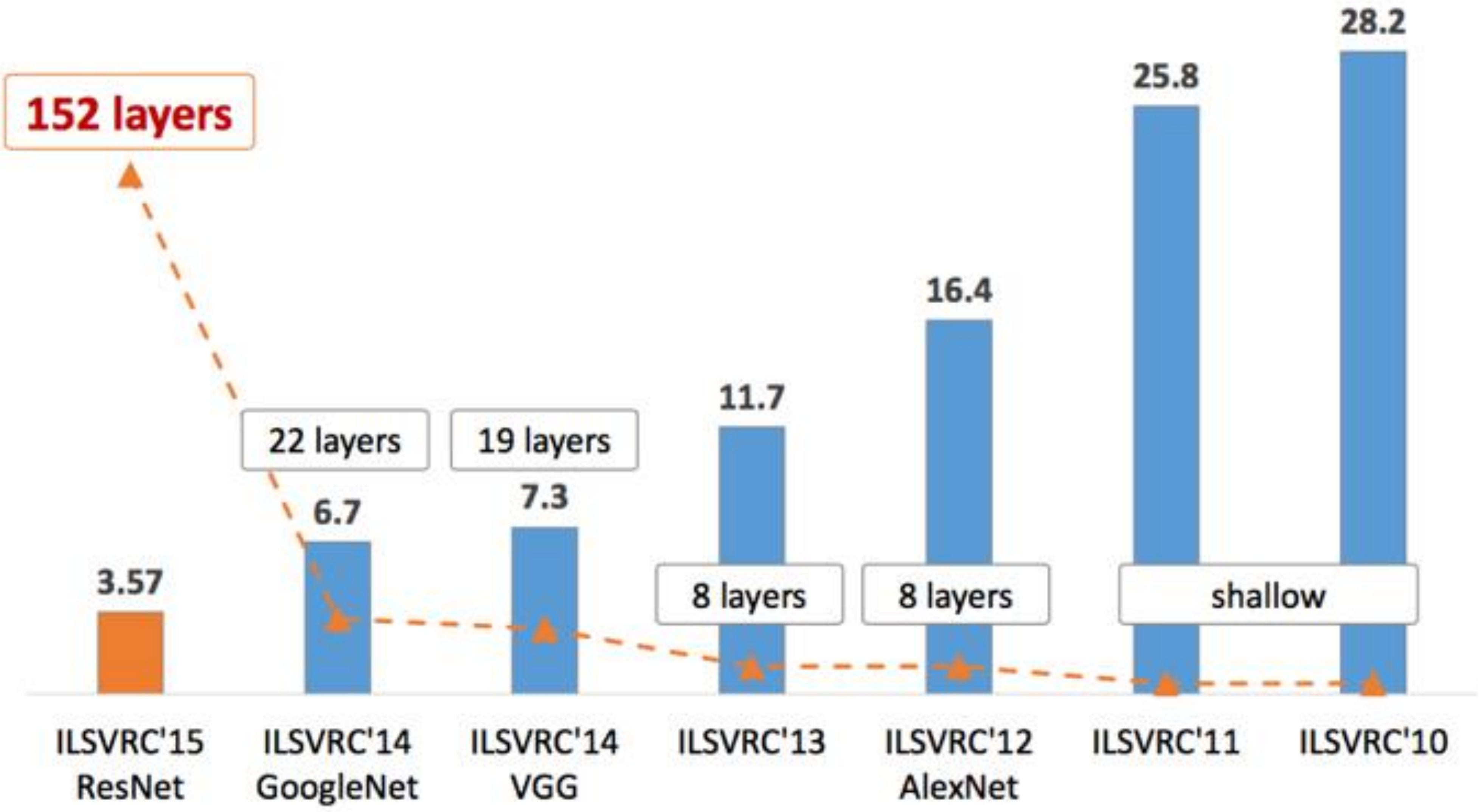


Convolutional Neural Network (AlexNet)

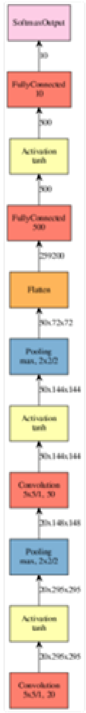
- Alex Krizhevsky, Geoffery Hinton et al., 2012



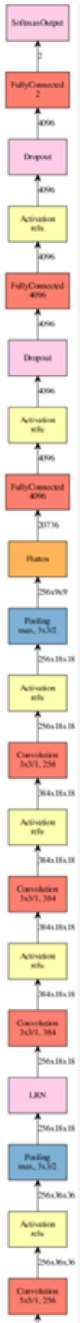
Winners' Error Rates on ImageNet Challenge



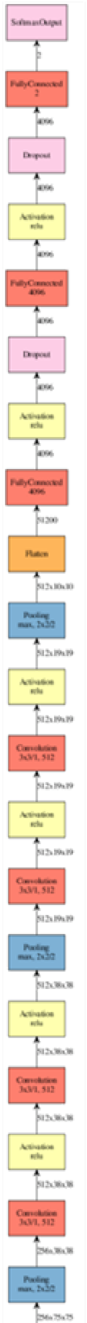
LeNet



AlexNet



VGG



GoogLeNet

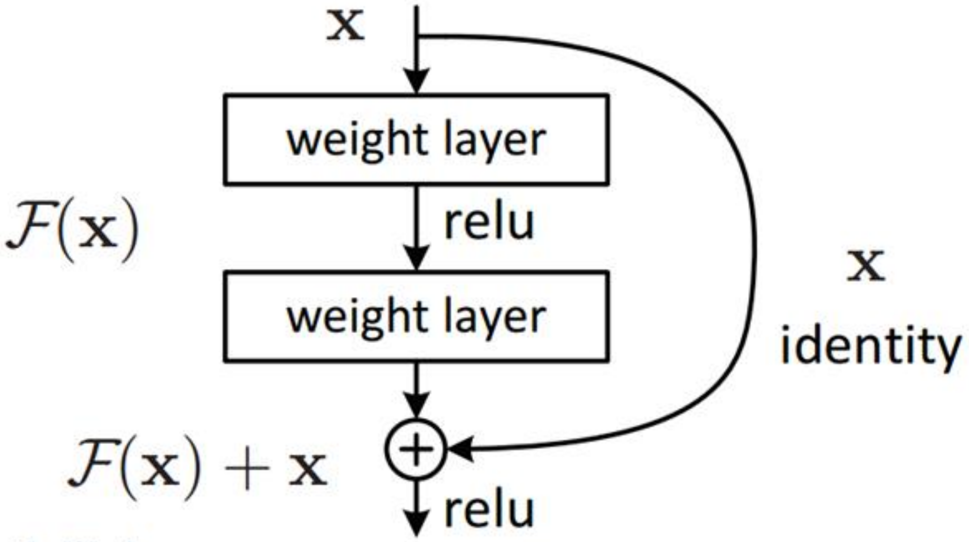
Inception V3

Inception BN

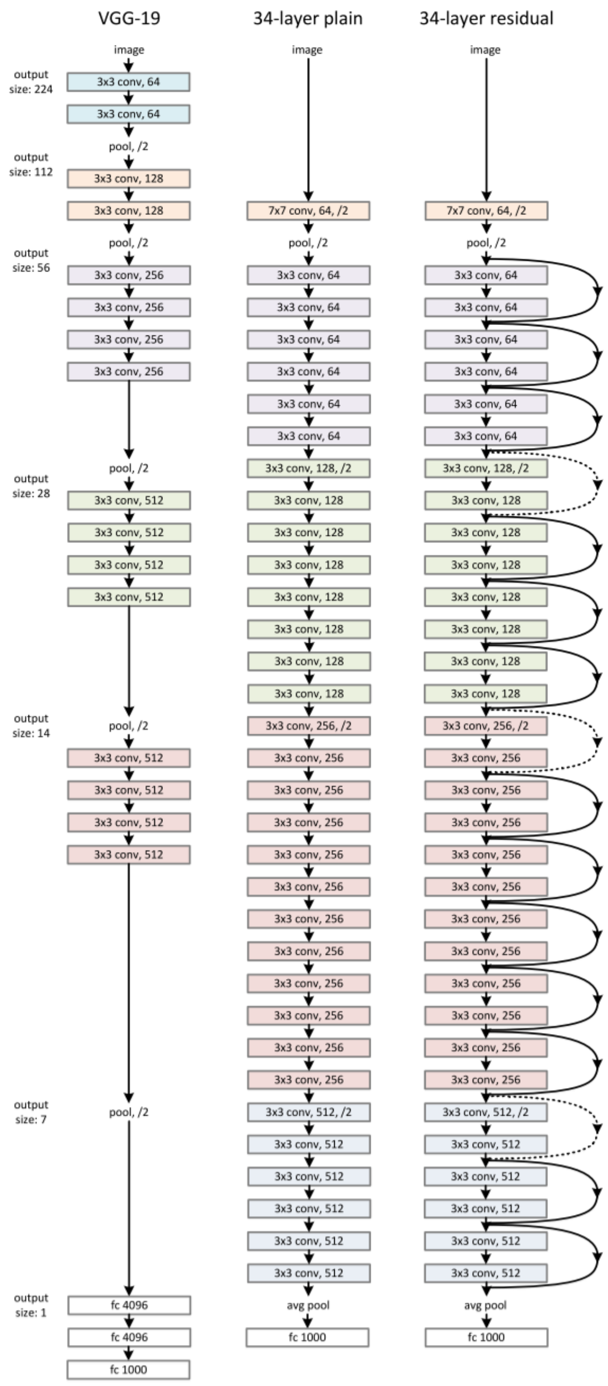

WE NEED TO GO
DEEPER
memegenerator.net

ResNet (2015)

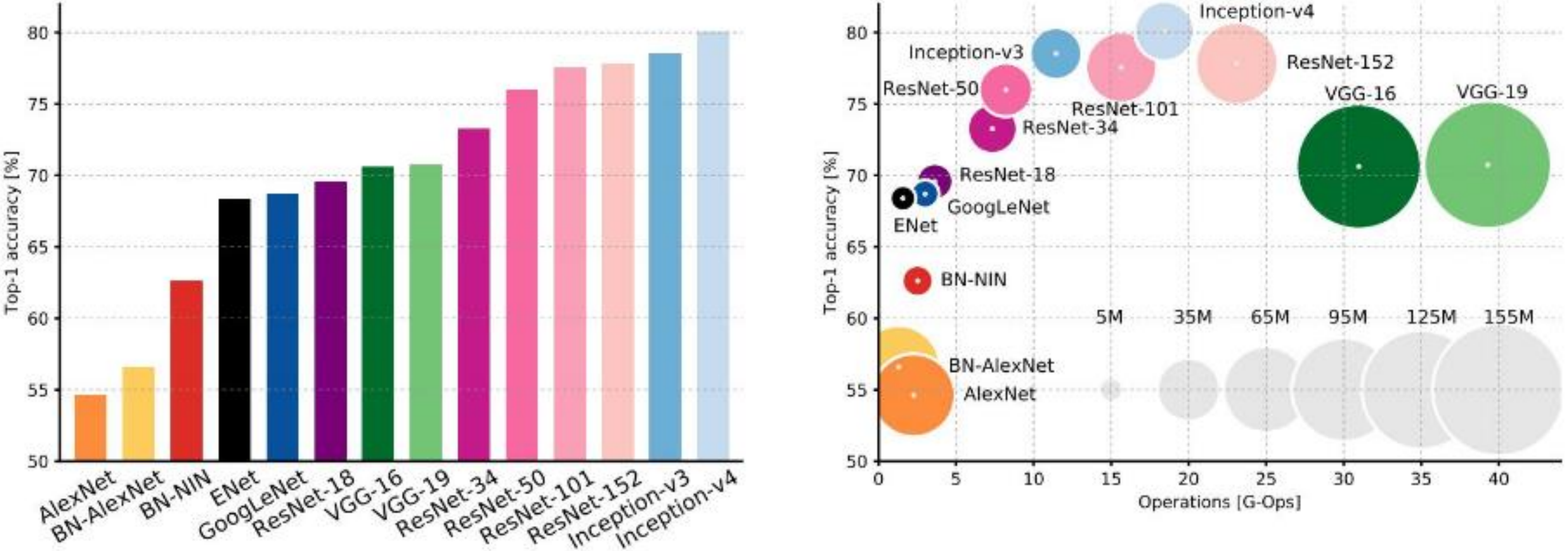
- Residual Neural Network
- Proposed “skip connection”
- 152-layer with 3.57% error rate



A residual block



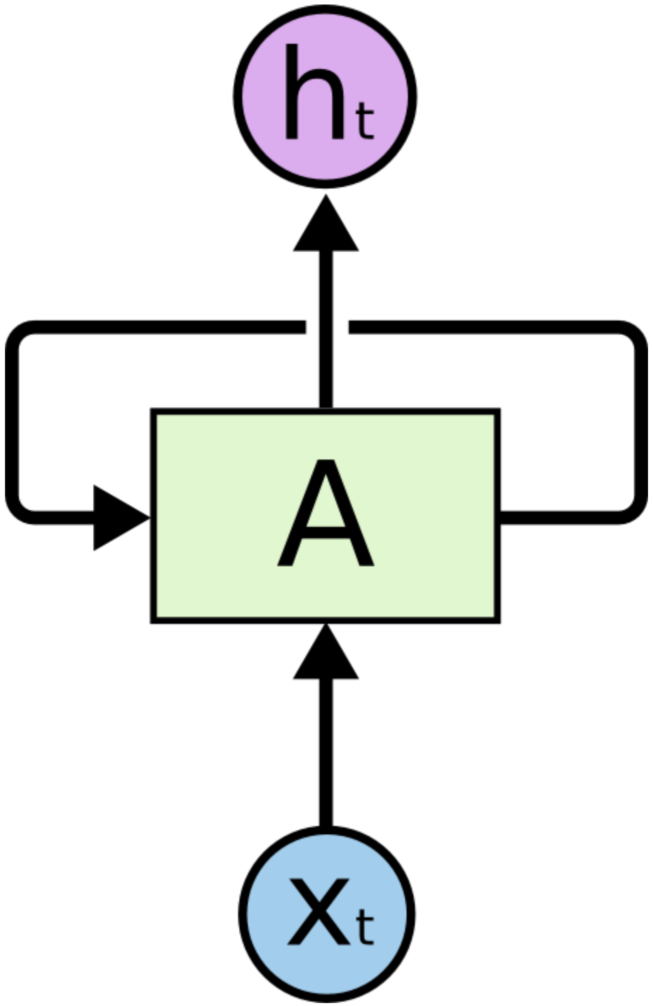
CNN Comparison



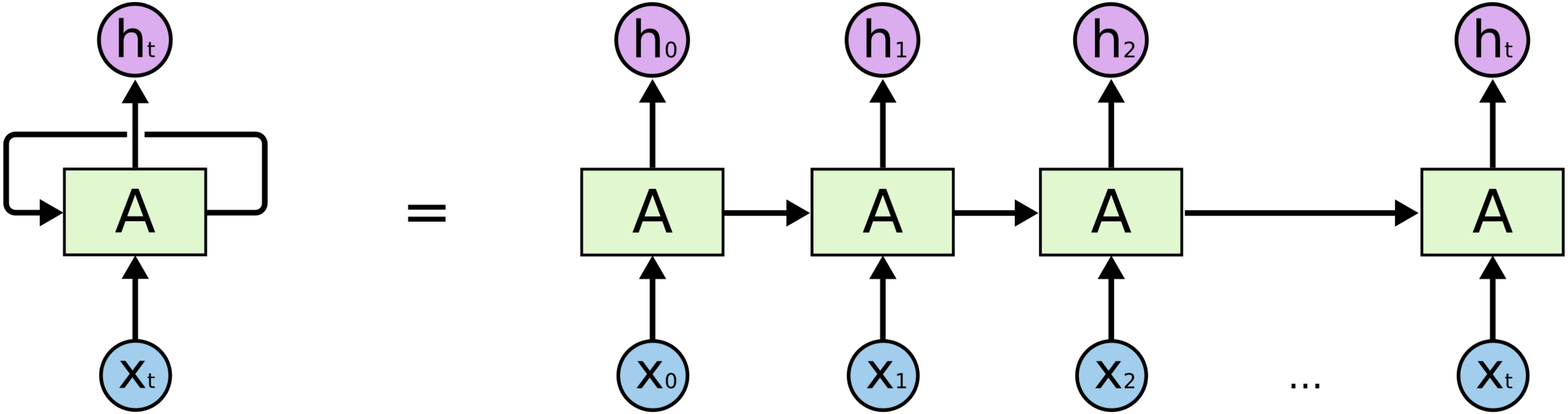
https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5

Recurrent Neural Networks (RNNs) and LSTM

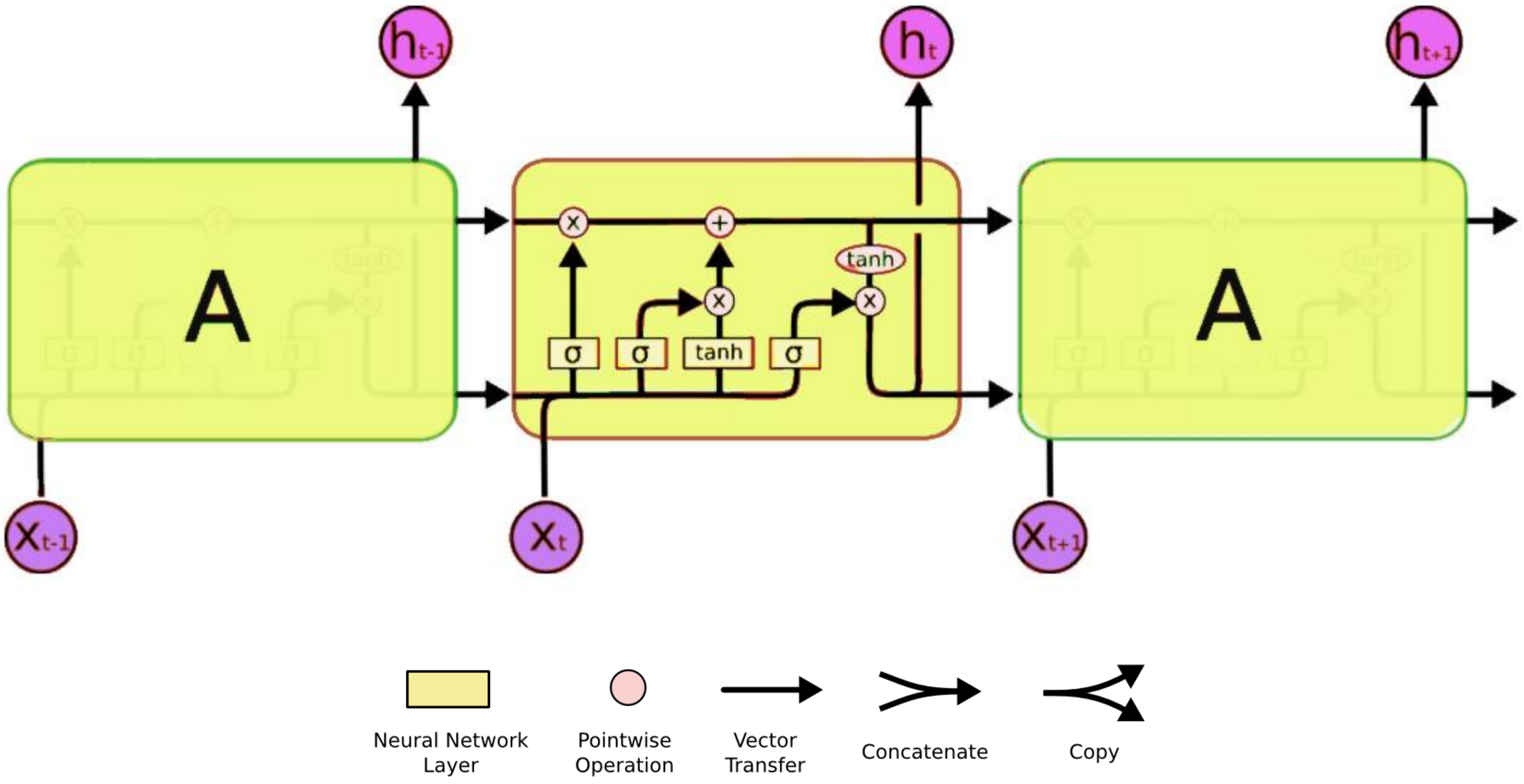
Recurrent Neural Networks (RNNs)



Unroll the RNN



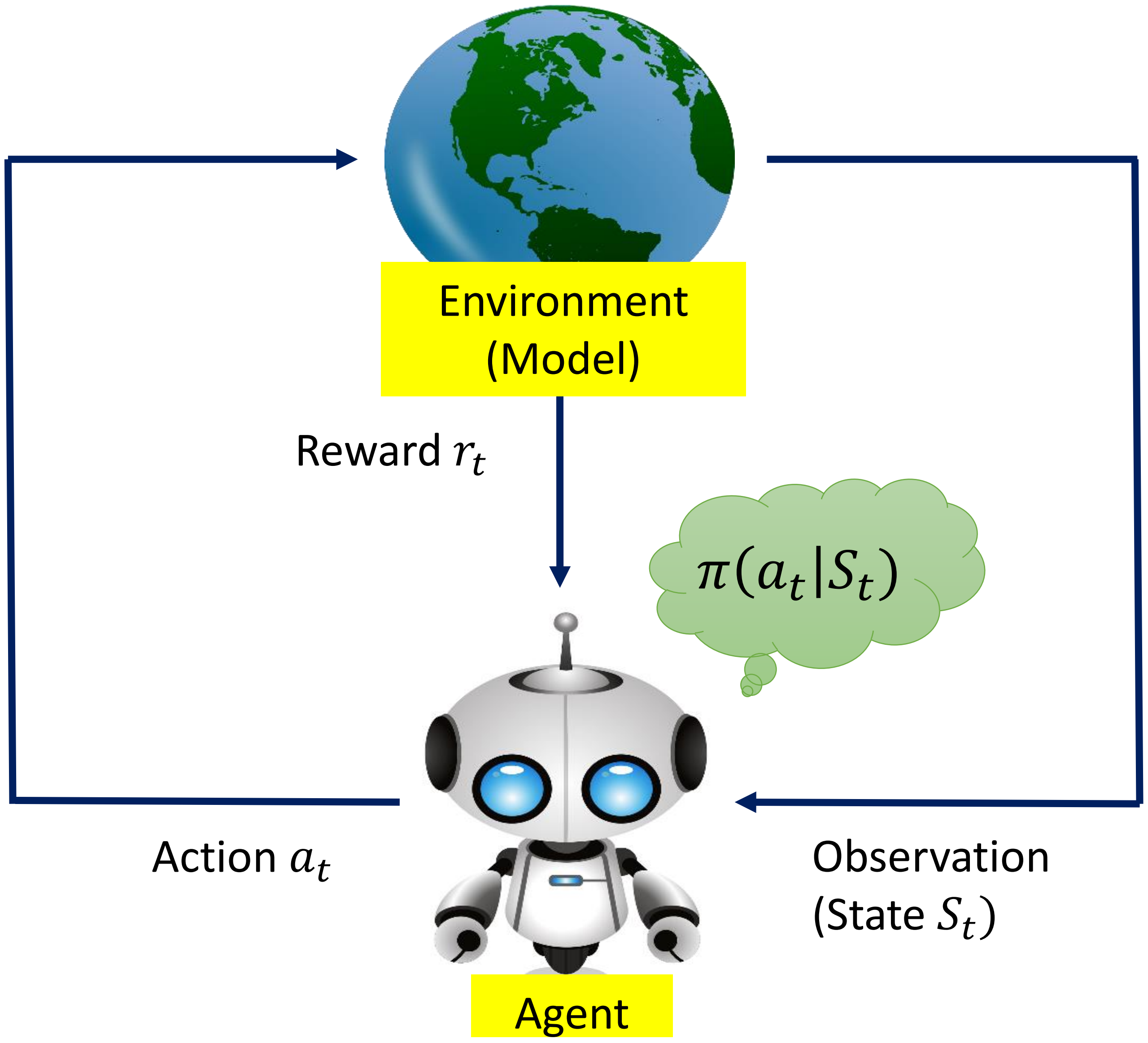
Long Short-term Memory (LSTM)



Deep Reinforcement Learning

What is Reinforcement Learning?

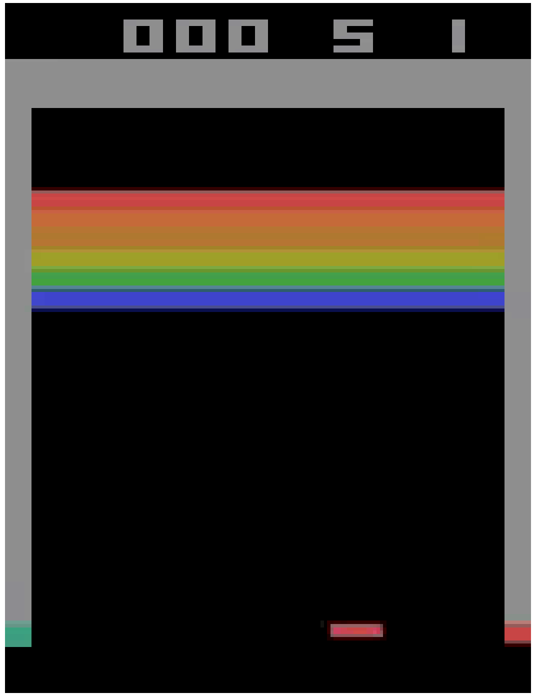
Fundamentals of Reinforcement Learning

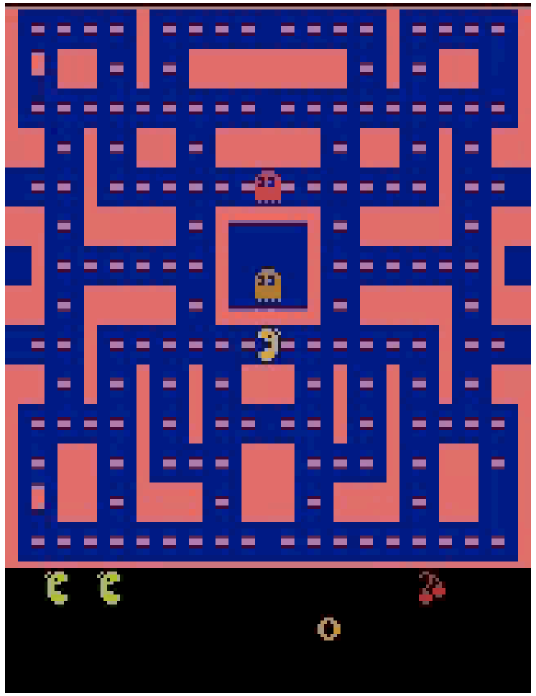
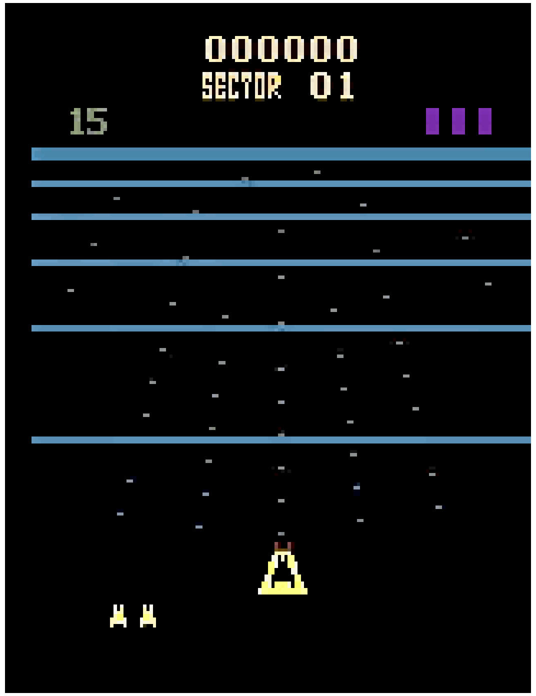


Google DeepMind

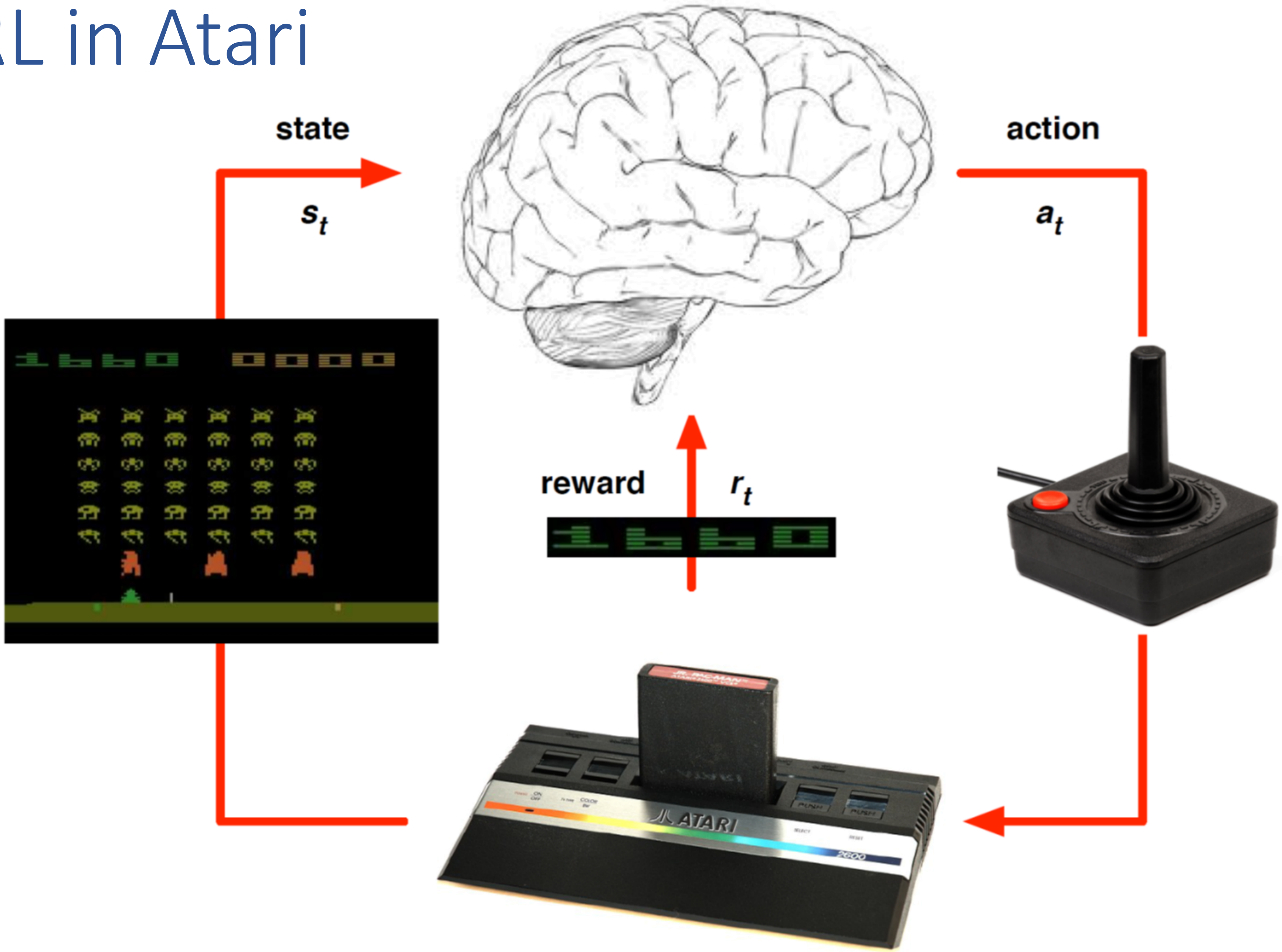
Learn to Play Atari Games

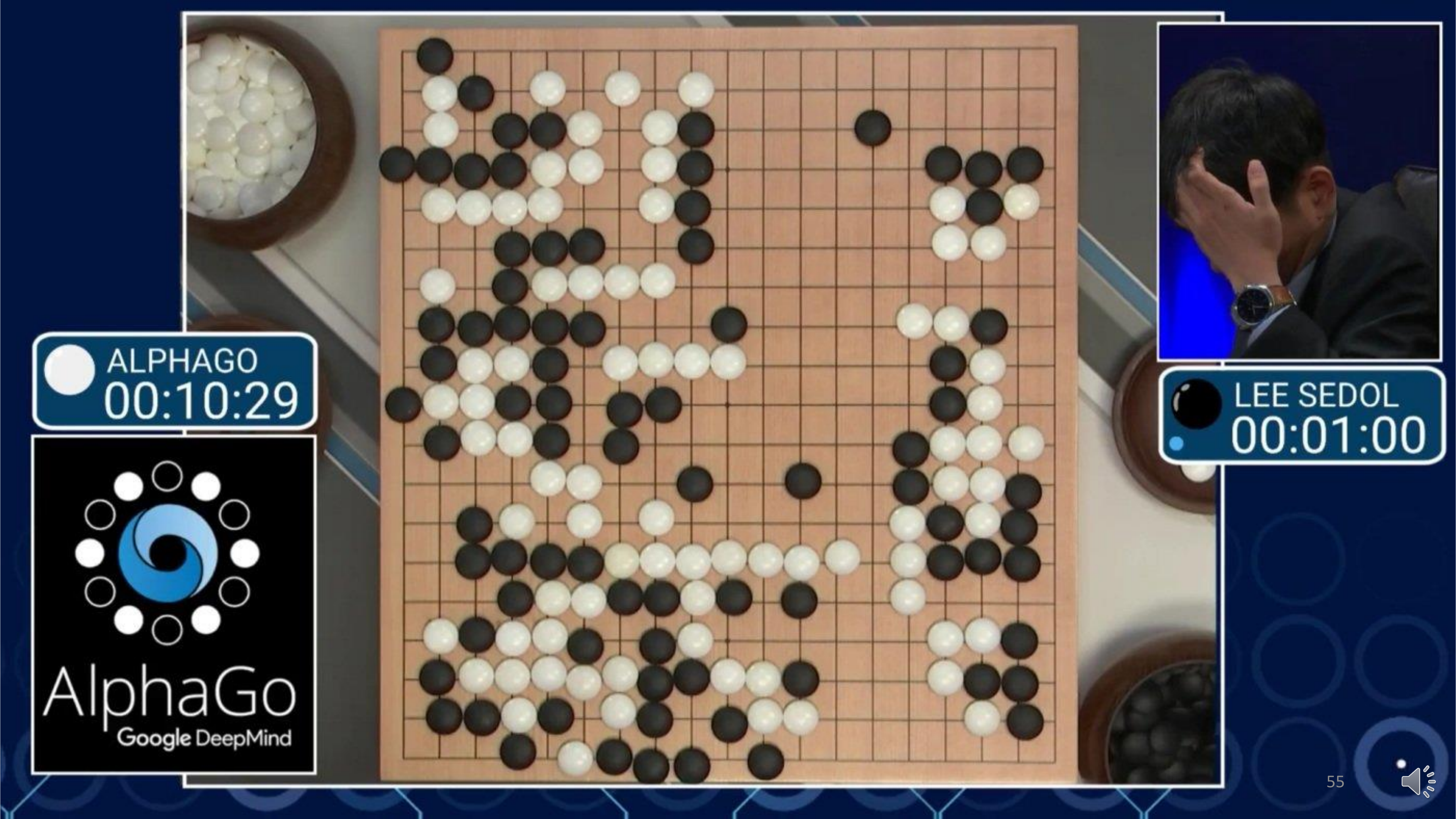
- Mnih et al., “Human Level Control through Deep Reinforcement Learning,” *Nature*, 2015



DRL in Atari




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

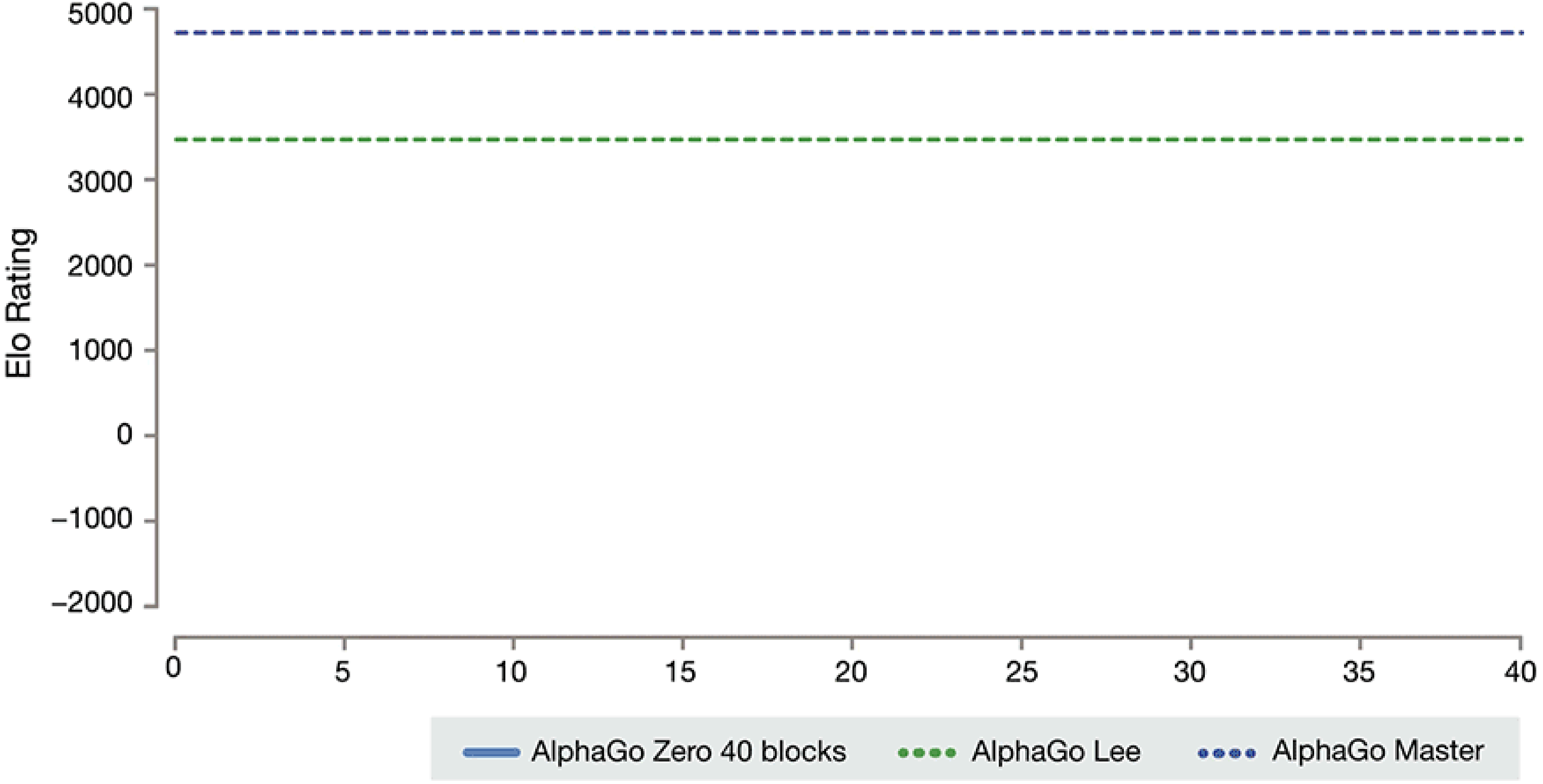
Dr. Aja Huang (黃士杰)

The Complexity of Go vs Chess

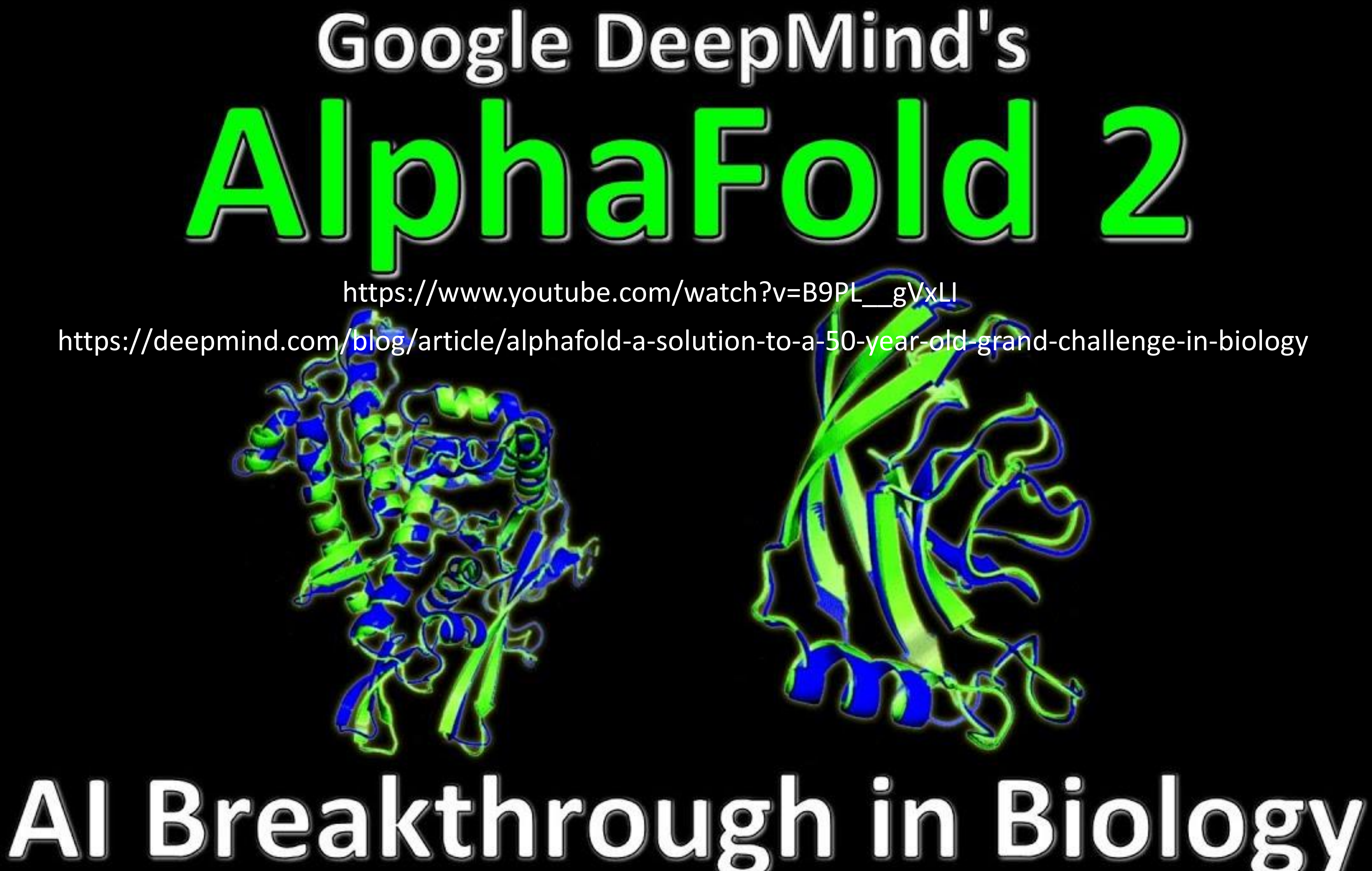
Game	Board size	State space	Game tree size
Go	19 x 19	10^{172}	10^{360}
Chess	8 x 8	10^{50}	10^{123}
Checkers	8 x 8	10^{18}	10^{54}


AlphaGo Zero
Starting from scratch

AlphaGo Zero



5000
4000
3000
2000
1000
0
−1000
−2000
Elo Rating
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

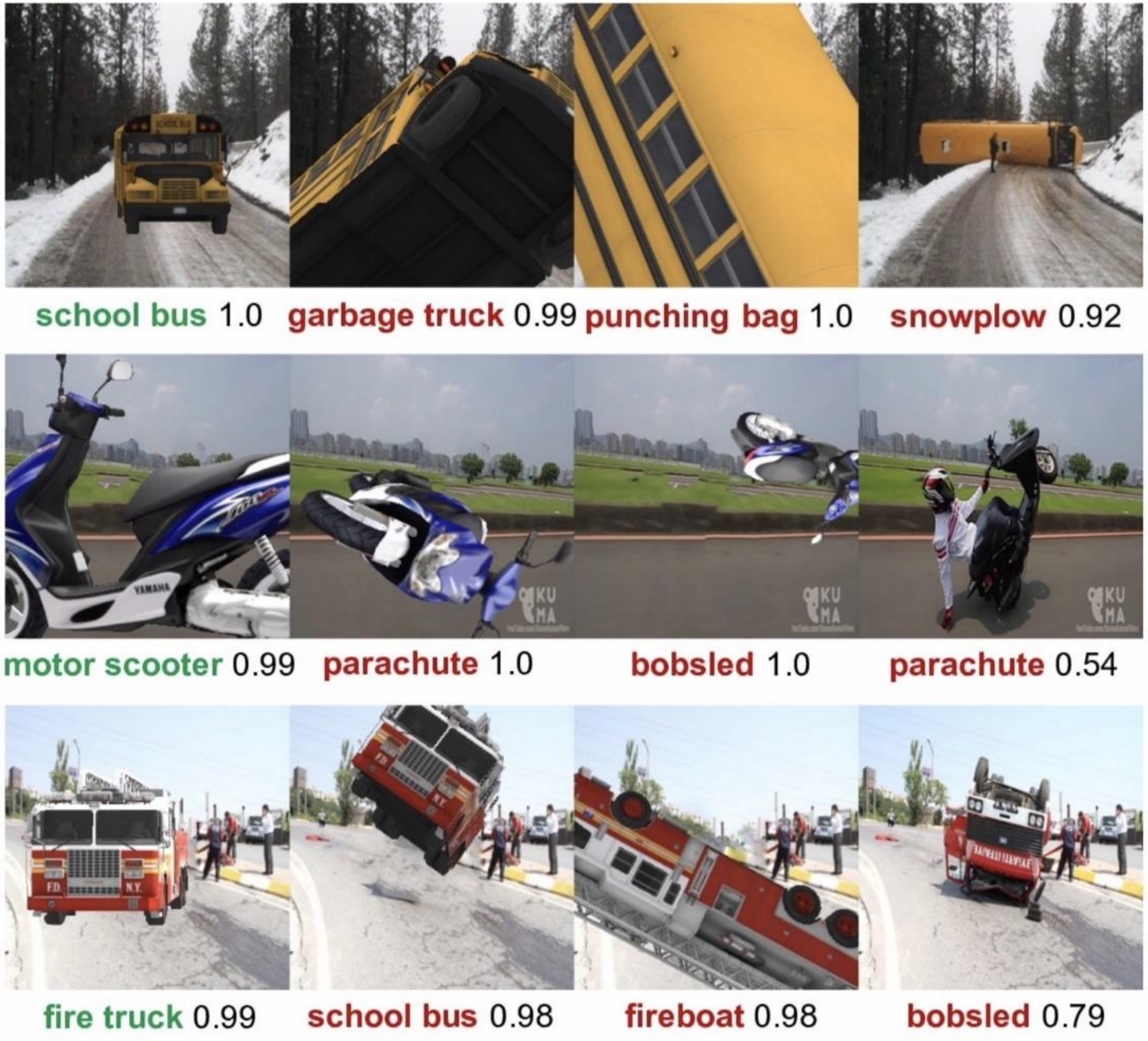


https://www.youtube.com/watch?v=B9PL__gVxLI

https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

Limits of Deep Learning

No Idea of Real World

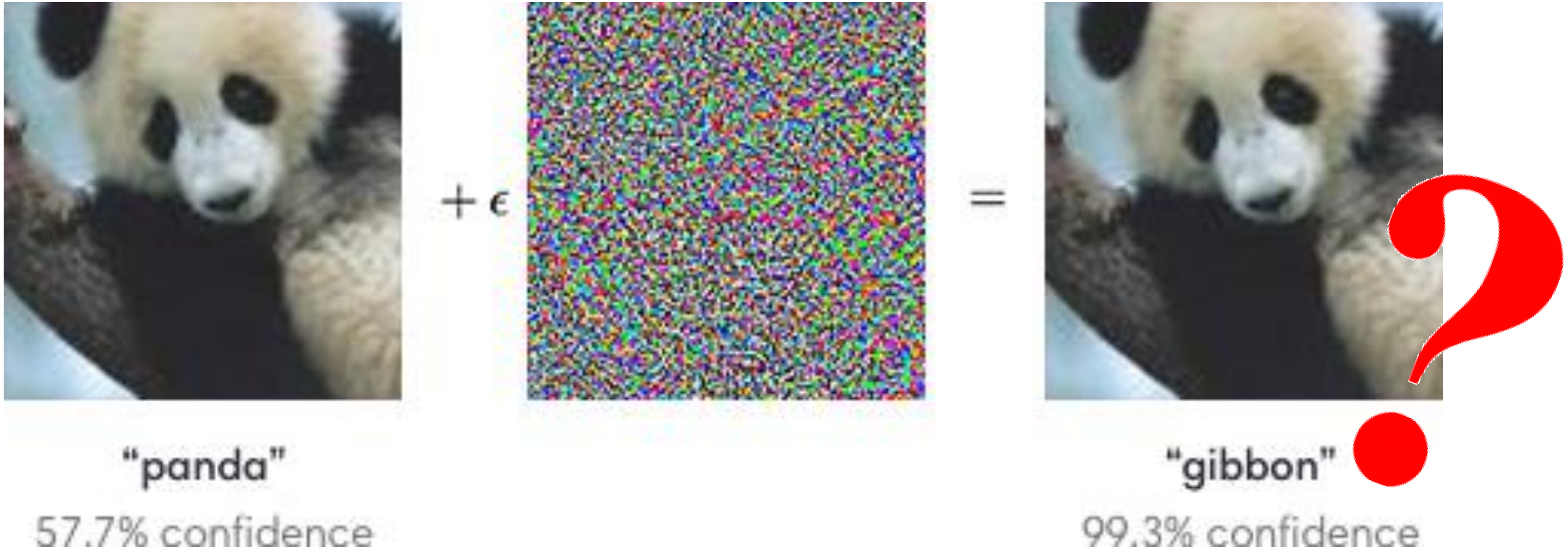


Alcorn et al., “Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects,”CVPR 2019.

Tesla Autopilot Hit an Overturned Truck in Taiwan

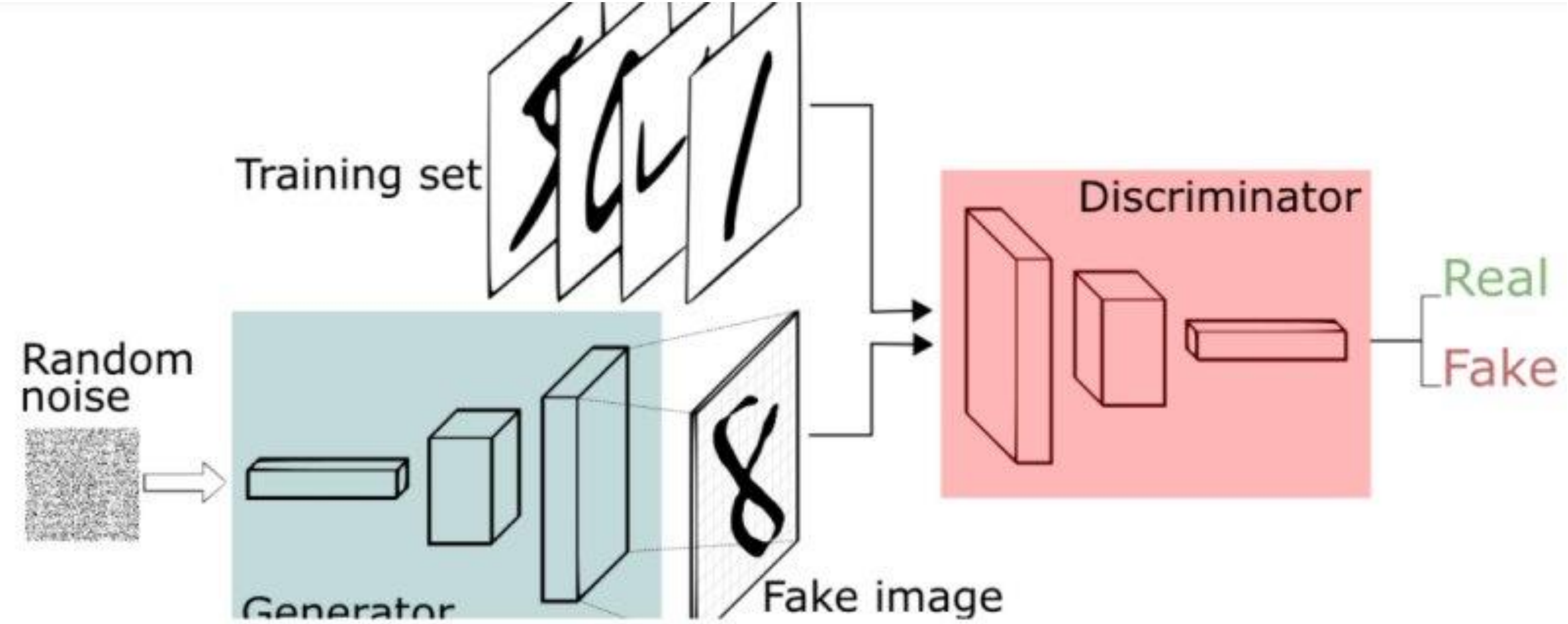

2020-06-01 06:43:53
國1 北 267K+650 水上路段
民視新聞台 HD
大貨車翻覆橫倒車道 特斯拉高速撞進車廂

Adversarial Attack



Generative Adversarial Networks (GAN)

- Ian Goodfellow



Super Resolution

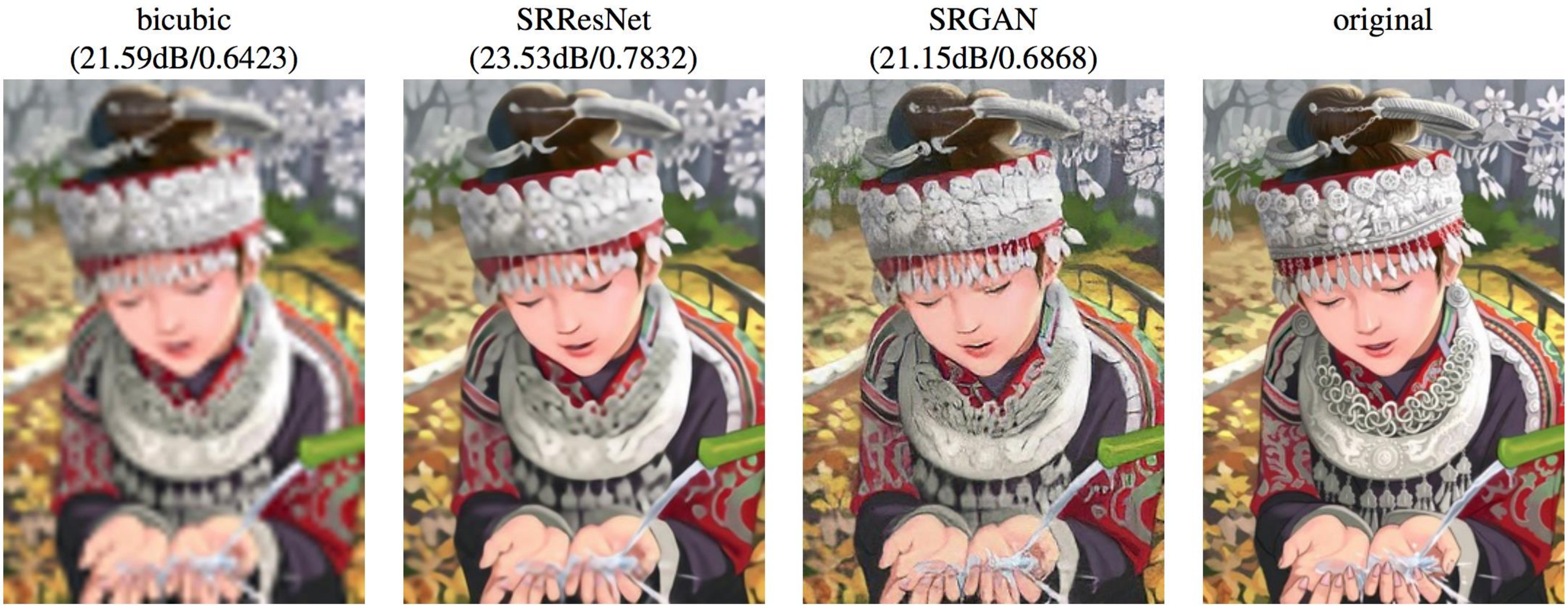
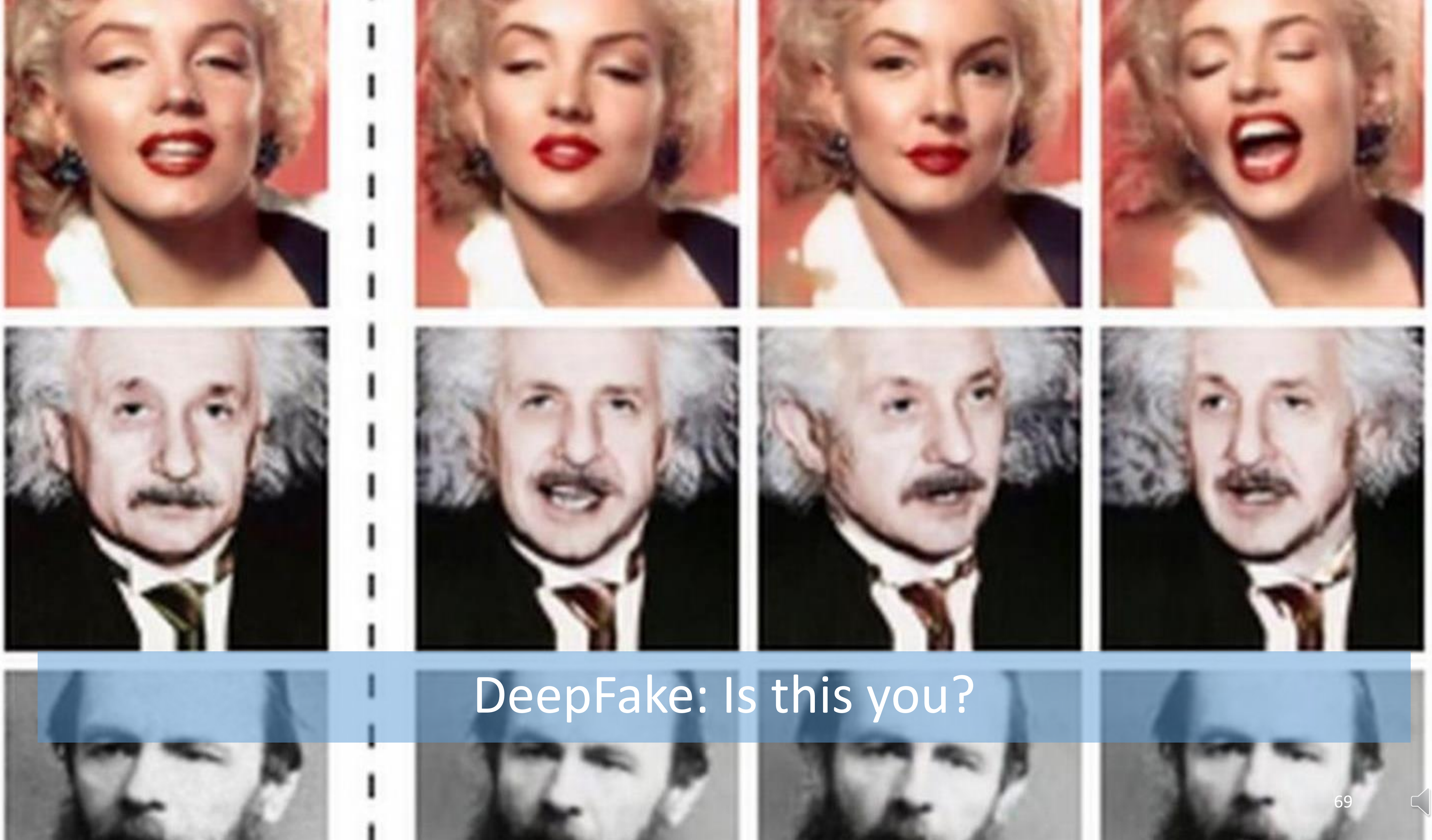


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

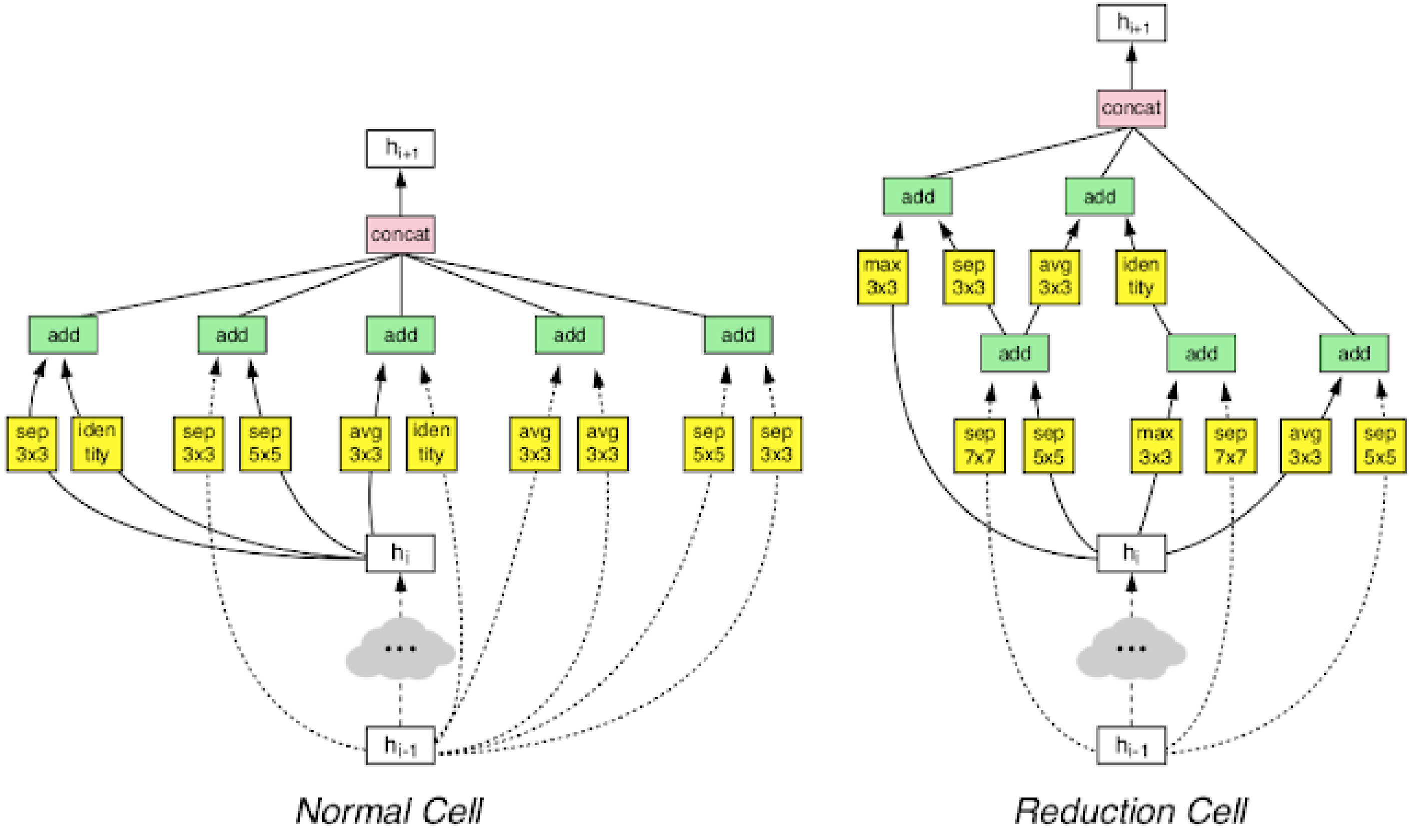


DeepFake: Is this you?

https://www.youtube.com/watch?v=gLoI9hAX9dw

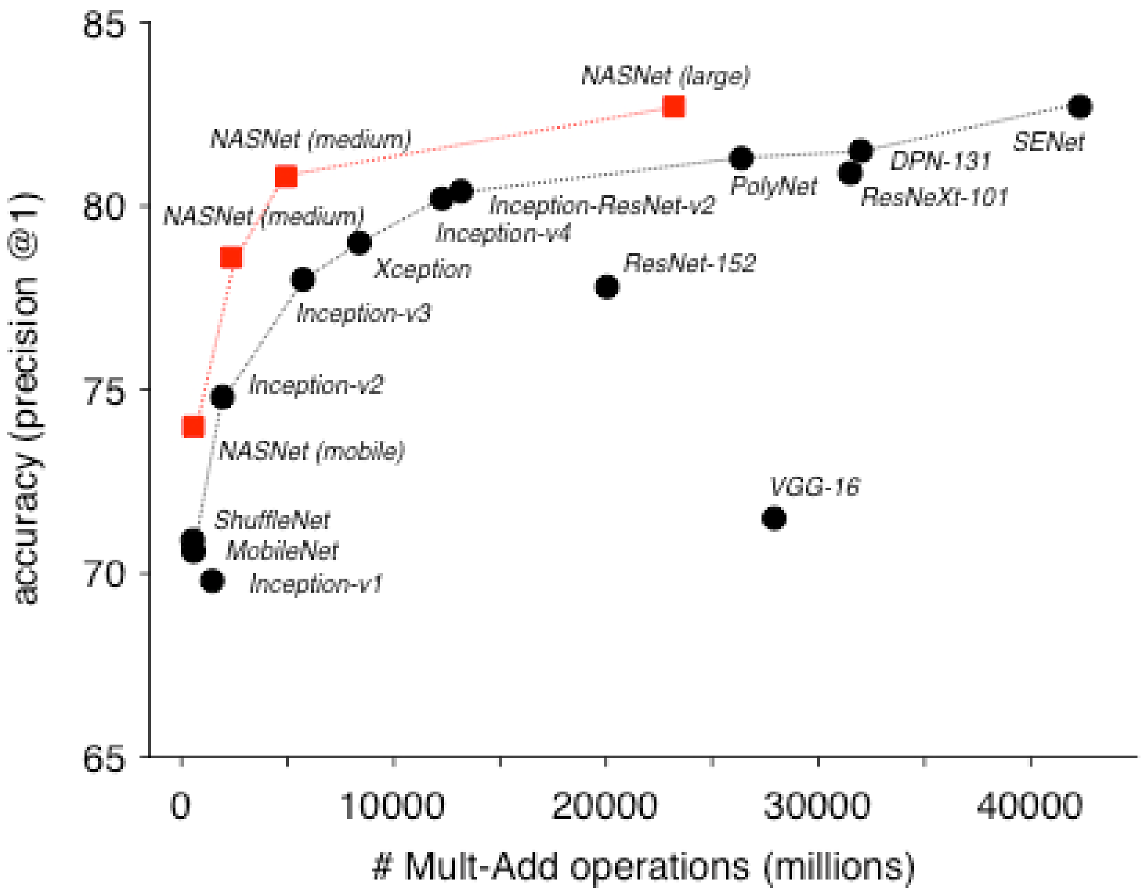
Google's AutoML

- Learning neural network cells automatically

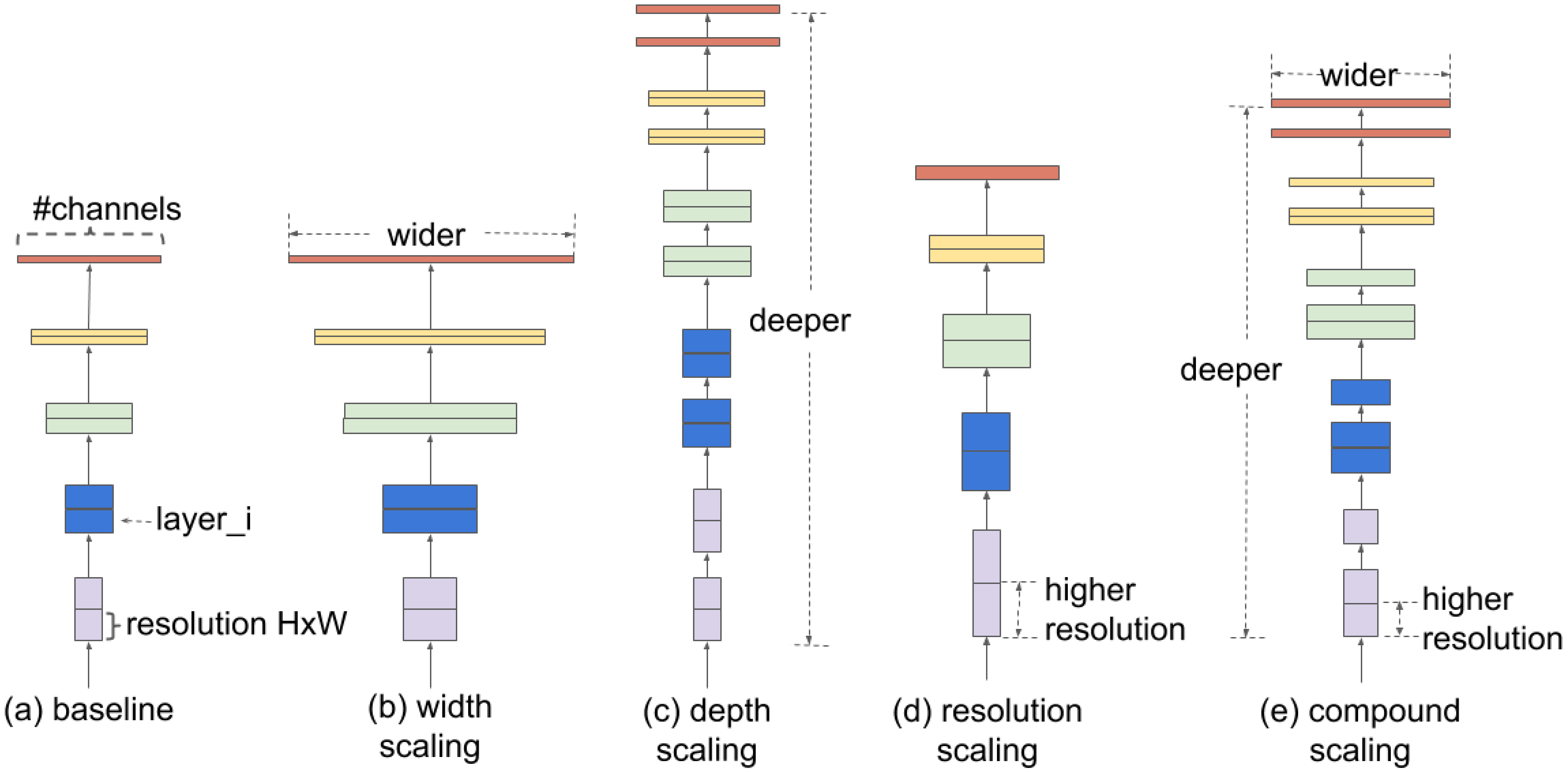


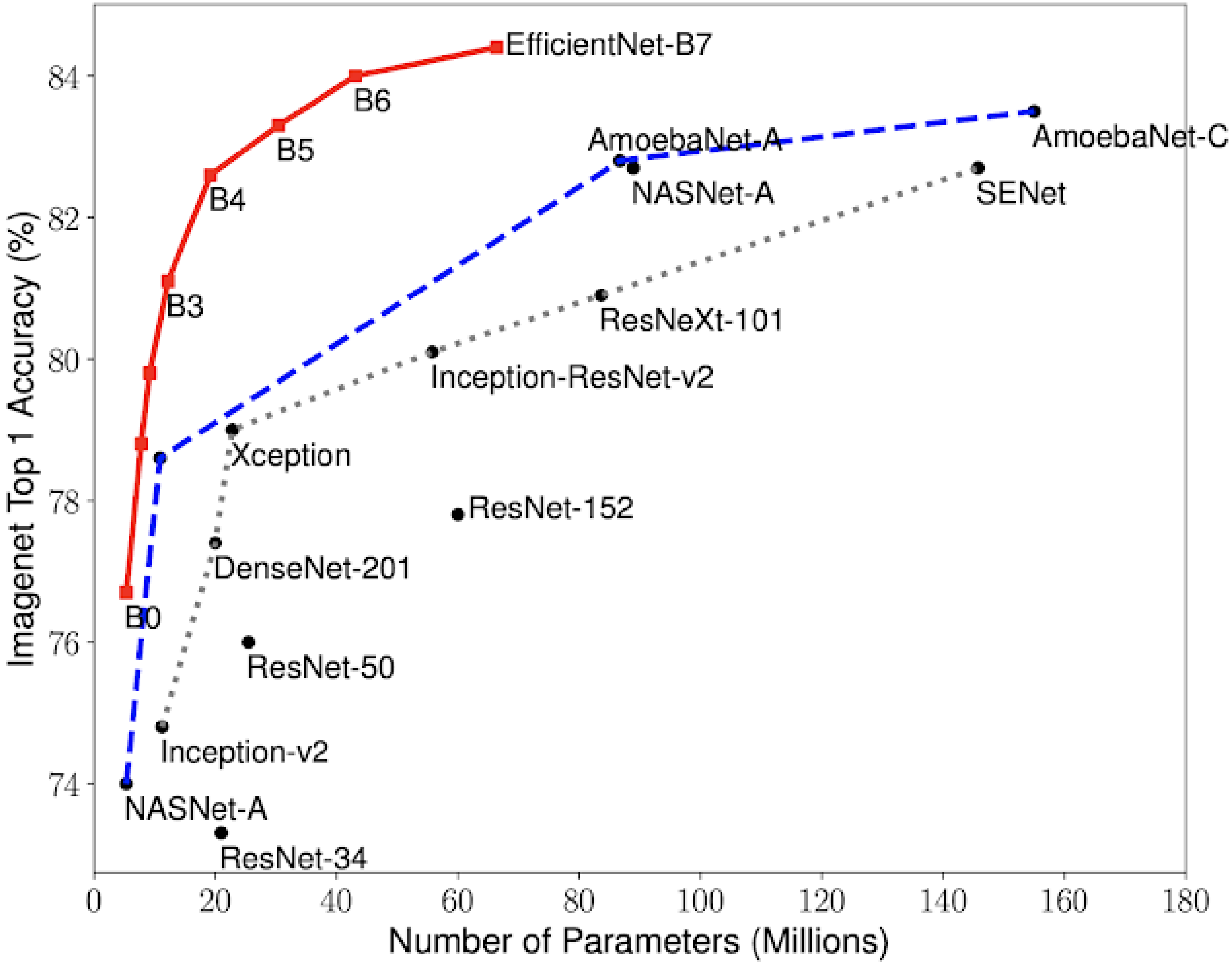
https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html

AutoML on ImageNet

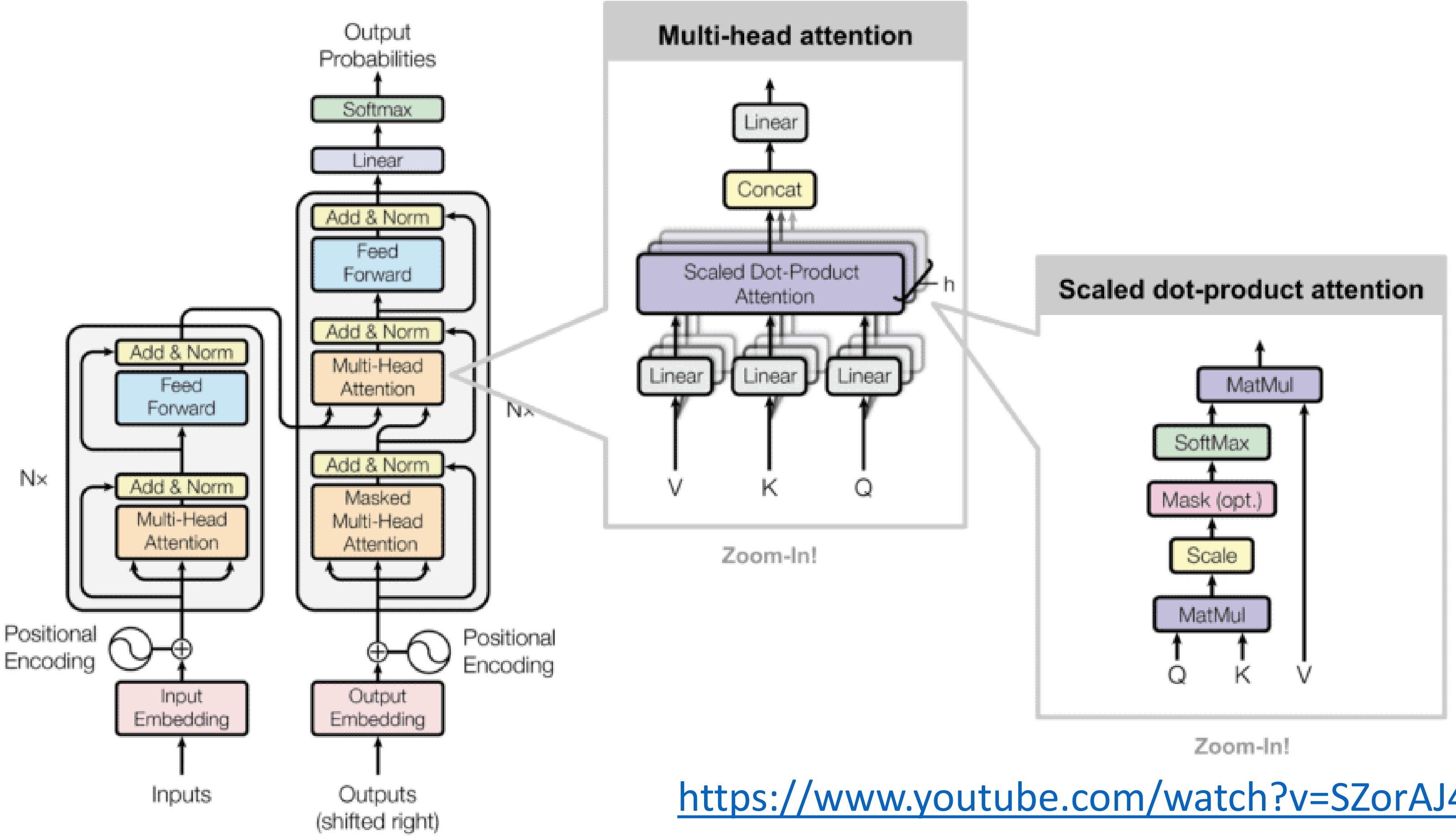


EfficientNet (May, 2019)



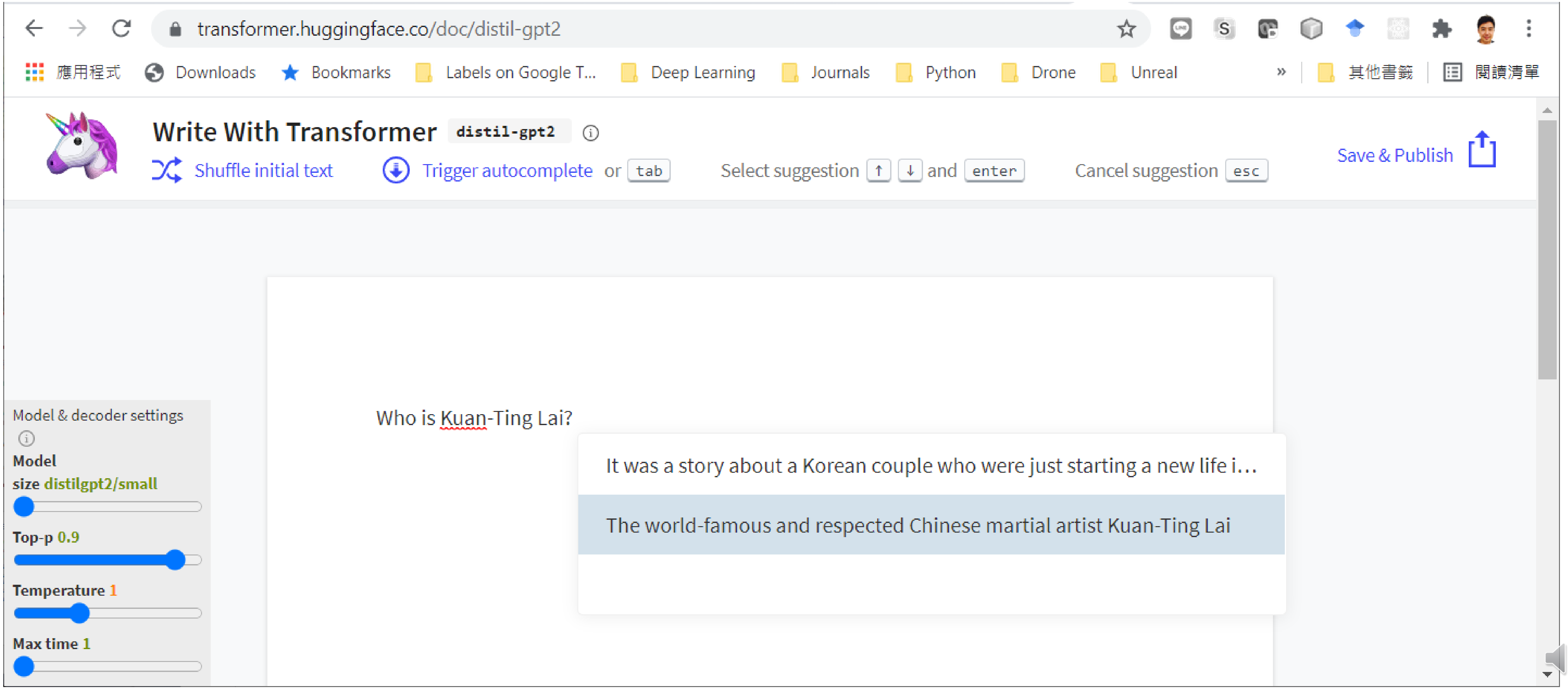

EfficientNet-B7
B6
B5
B4
B3
B0
AmoebaNet-A
AmoebaNet-C
NASNet-A
SENet
ResNeXt-101
Inception-ResNet-v2
Xception
ResNet-152
DenseNet-201
ResNet-50
Inception-v2
NASNet-A
ResNet-34
Imagenet Top 1 Accuracy (%)
84
82
80
78
76
74
0
20
40
60
80
100
120
140
160
180
Number of Parameters (Millions)

Transformer & GPT3



https://www.youtube.com/watch?v=SZorAJ4I-sA

huggingface.co



Key Takeaways

1. Deep learning is a branch of Machine Learning, which is a sub-field of Artificial Intelligence.
2. There are two stages in machine learning: training (learning) and testing (inference).
3. Gradient Descent is used to train NN models by updating weights to minimize the prediction errors.
4. Convolutional Neural Networks (CNN) are used to recognize images.
5. RNN and LSTM are used to recognize sequential data such as text or speech.
6. Generative Adversarial Networks (GANs) can be used to generate fake data.
7. Transformer told us that attention is all you need!
8. Deep Reinforcement Learning can not only play Go, but also study new drugs.

References

1. https://www.buzzfeed.com/kasiagalazka/science-fiction-things-that-actually-exist-now
2. https://www.geek.com/movies/10-movies-that-helped-create-real-technology-1740036/
3. https://www.gadgetsnow.com/slideshows/8-sci-fi-movie-technologies-that-are-real-now/Video-calling/photolist/52869590.cms
4. What is backpropagation really doing? https://www.youtube.com/watch?v=Ilg3gGewQ5U
5. http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/
6. https://pmirla.github.io/2016/08/16/AI-Winter.html
7. https://tw.saowen.com/a/6cdc2f1279016e566832bb1234e06d321992dd1fabcdf4a2e0a3e16fc0dc09dc
8. https://tectales.com/ai/ai-used-to-identify-different-types-of-brain-injuries.html
9. Transformers, explained: Understand the model behind GPT, BERT, and T5
10. https://transformer.huggingface.co/
11. https://www.scienceabc.com/innovation/what-is-artificial-intelligence.html