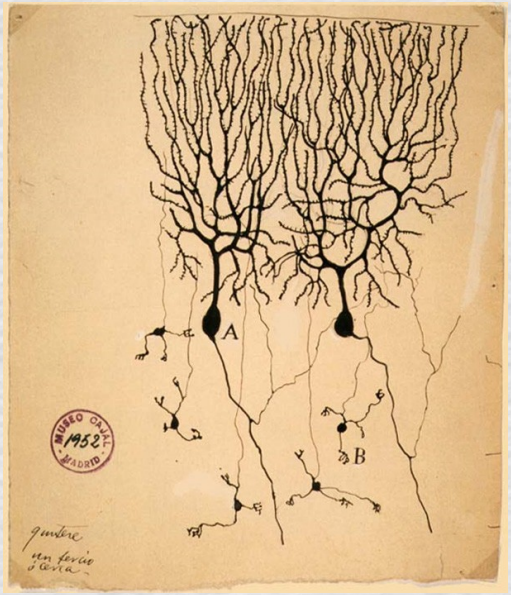
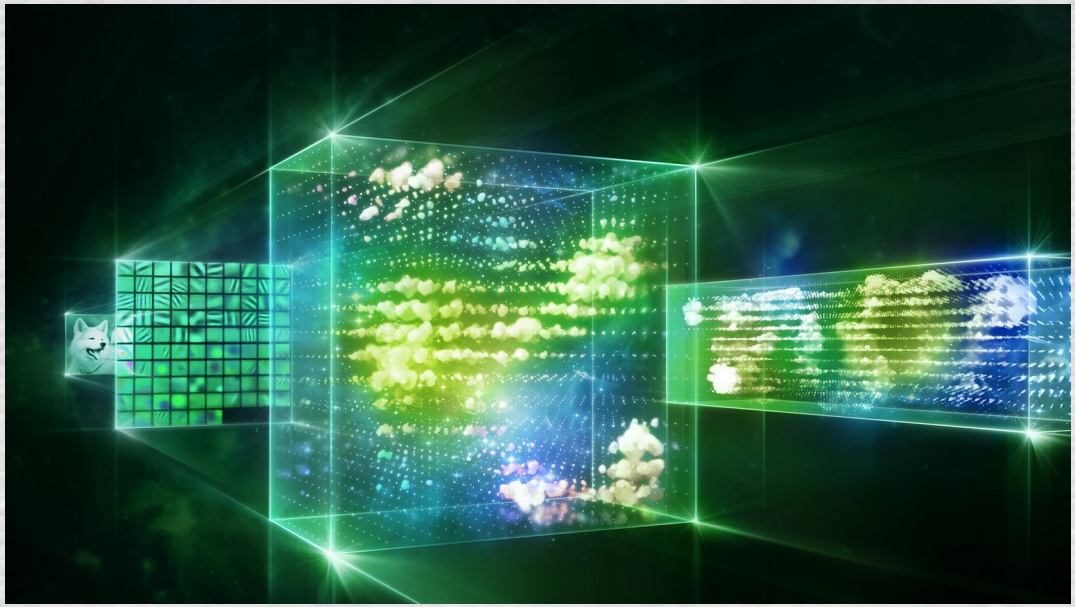




How To Build A Brain

Perception

Beating humans

Gary Kasparov

Deep Blue

Feng-Hsiung Hsu

1997: Deep Blue beats Gary Kasparov at chess.

Moravec’s paradox (1988)

"It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility"

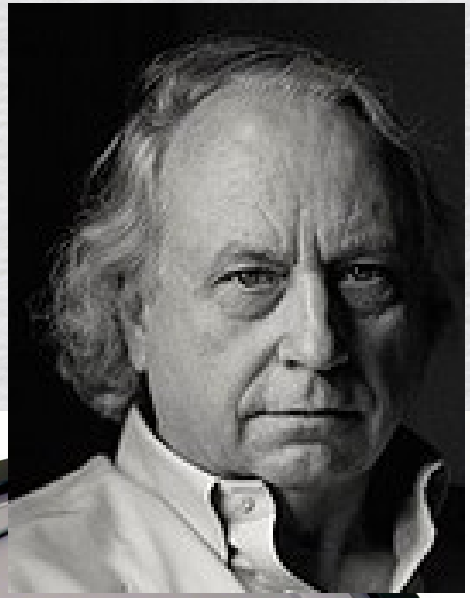
Rodney Brooks

In early AI research, intelligence was characterized as

“the things that highly-educated male scientists found challenging”

while

"things that children of four or five years could do effortlessly, such as visually distinguishing between a coffee cup and a chair...were not thought of as activities requiring intelligence"

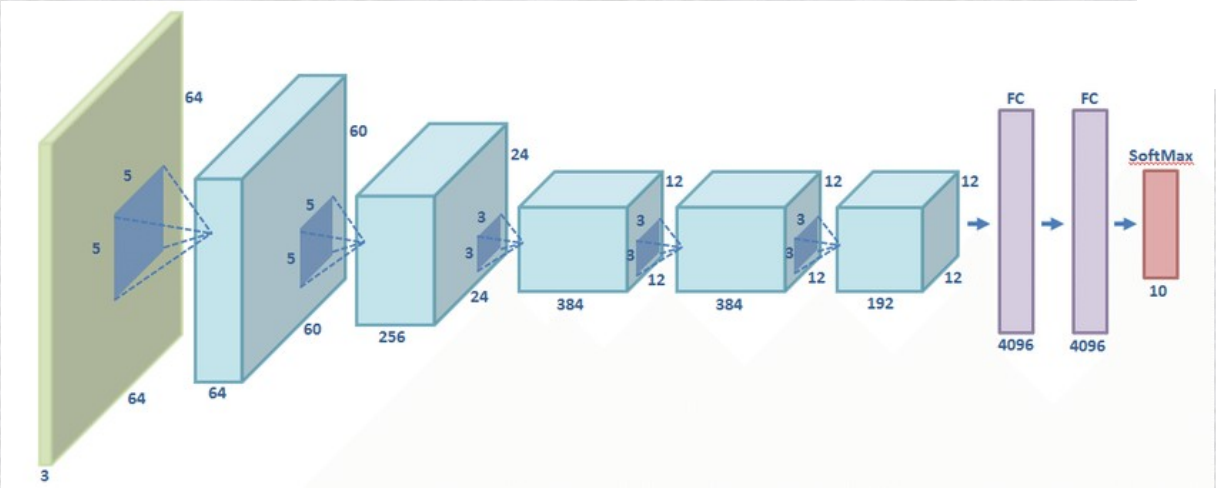


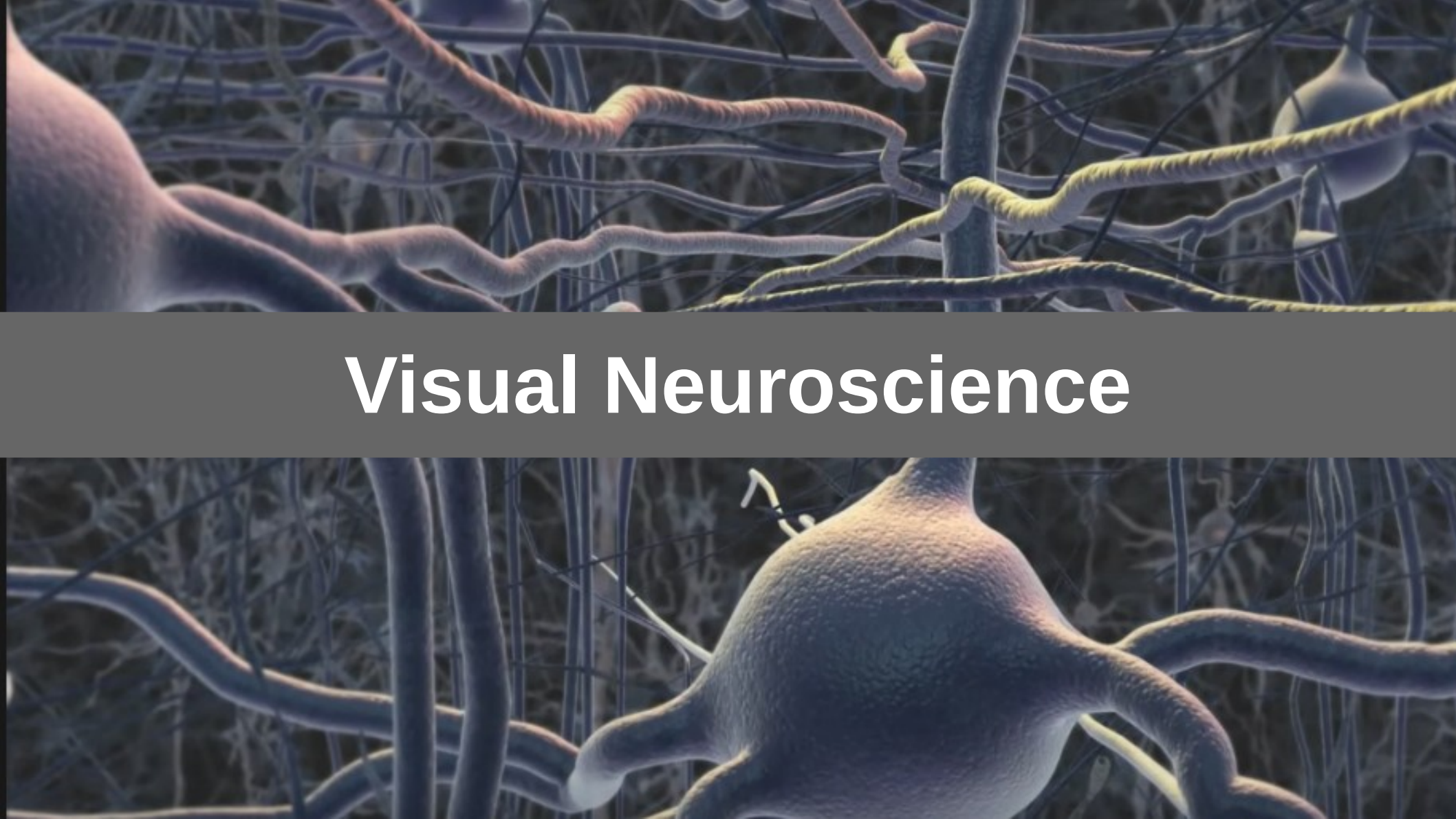
Stephen Pinker (1994)

“The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard”

AlexNet (2012)

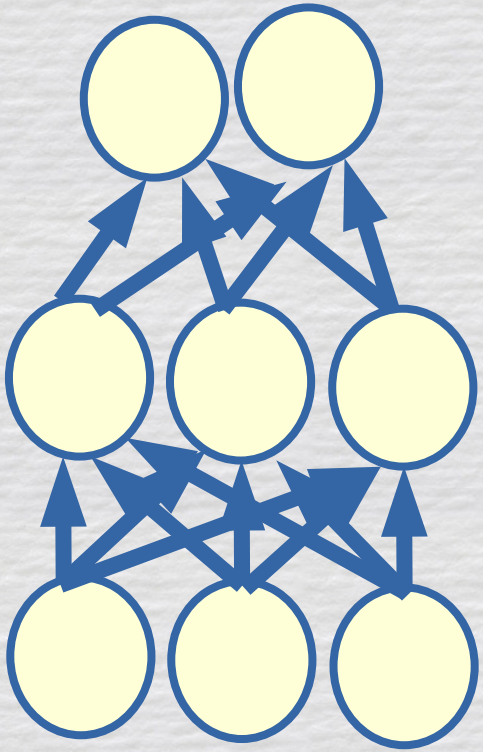
Ilya Sutskever, Alex Krizhevsky, Geoffrey Hinton

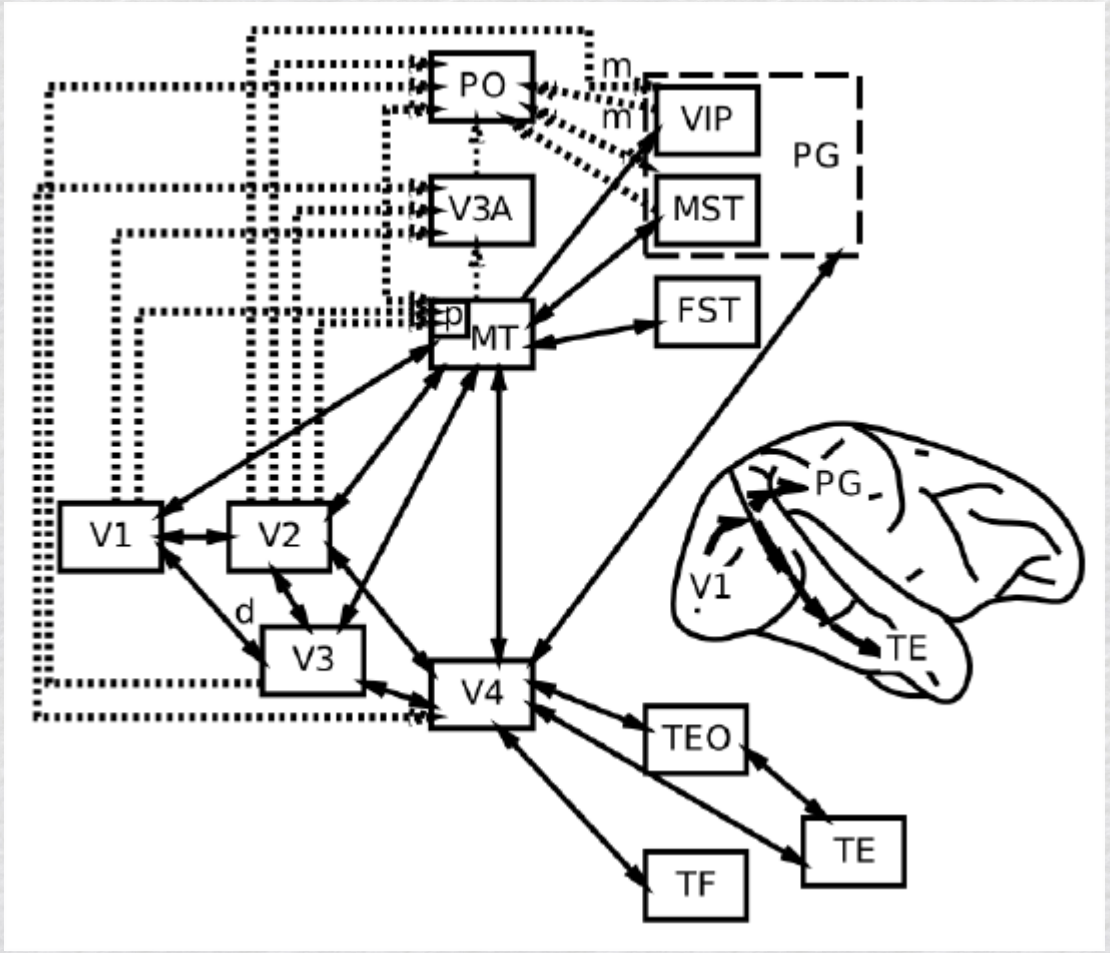




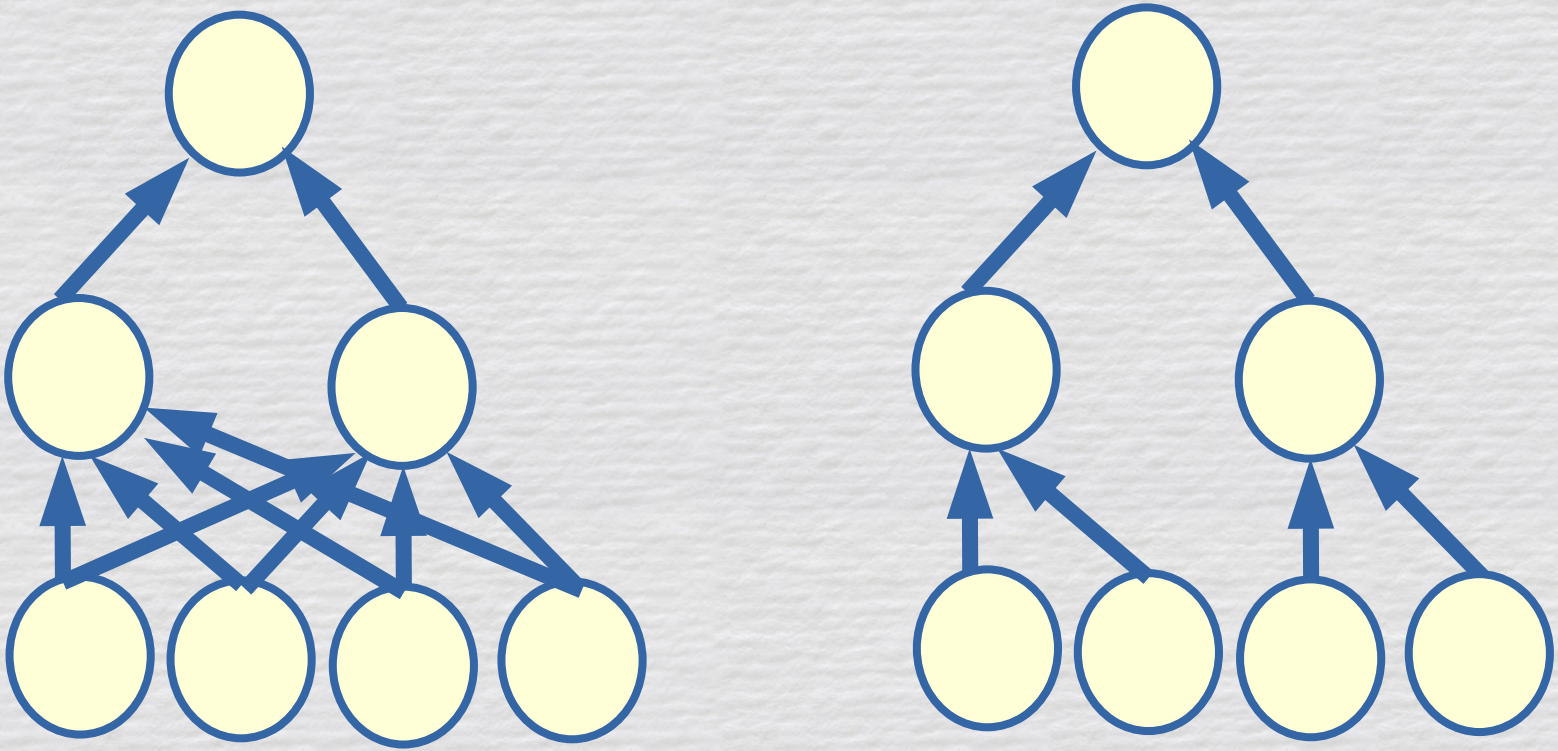
Visual Neuroscience

Deeper

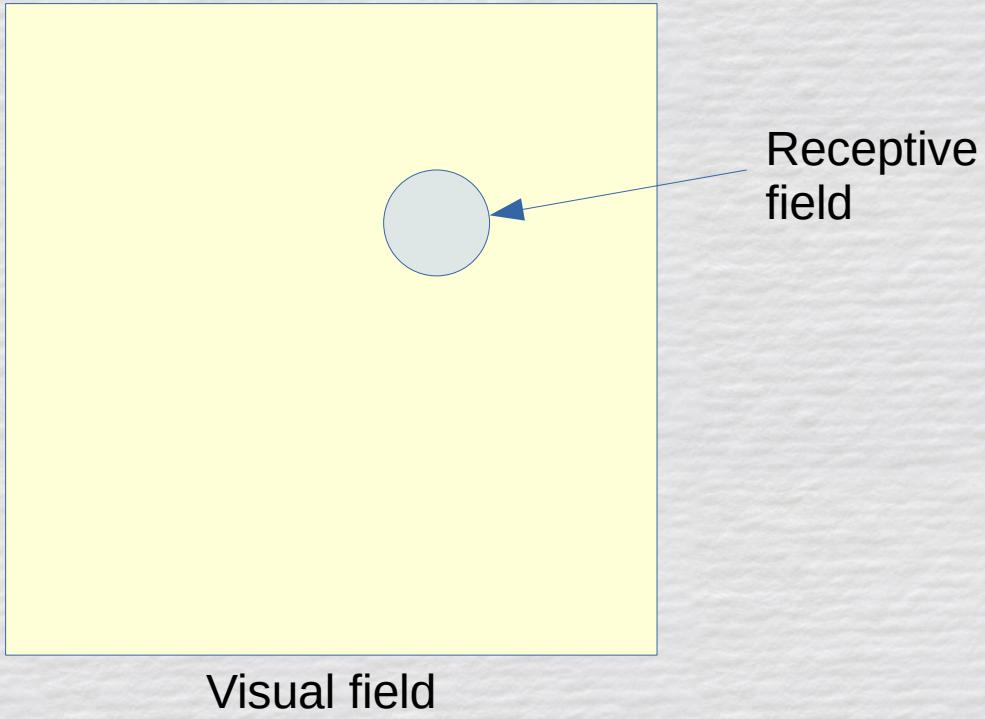




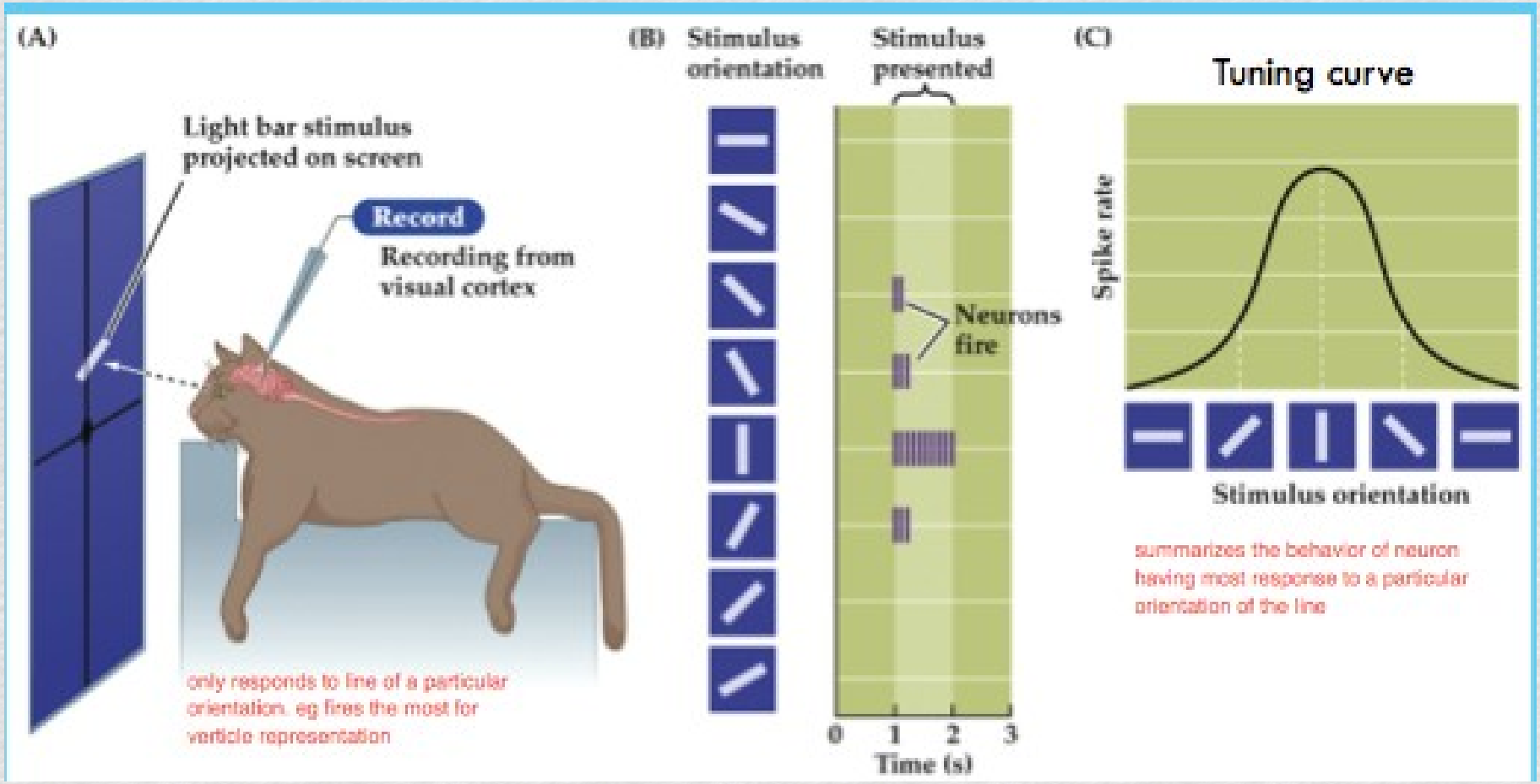
More sparse



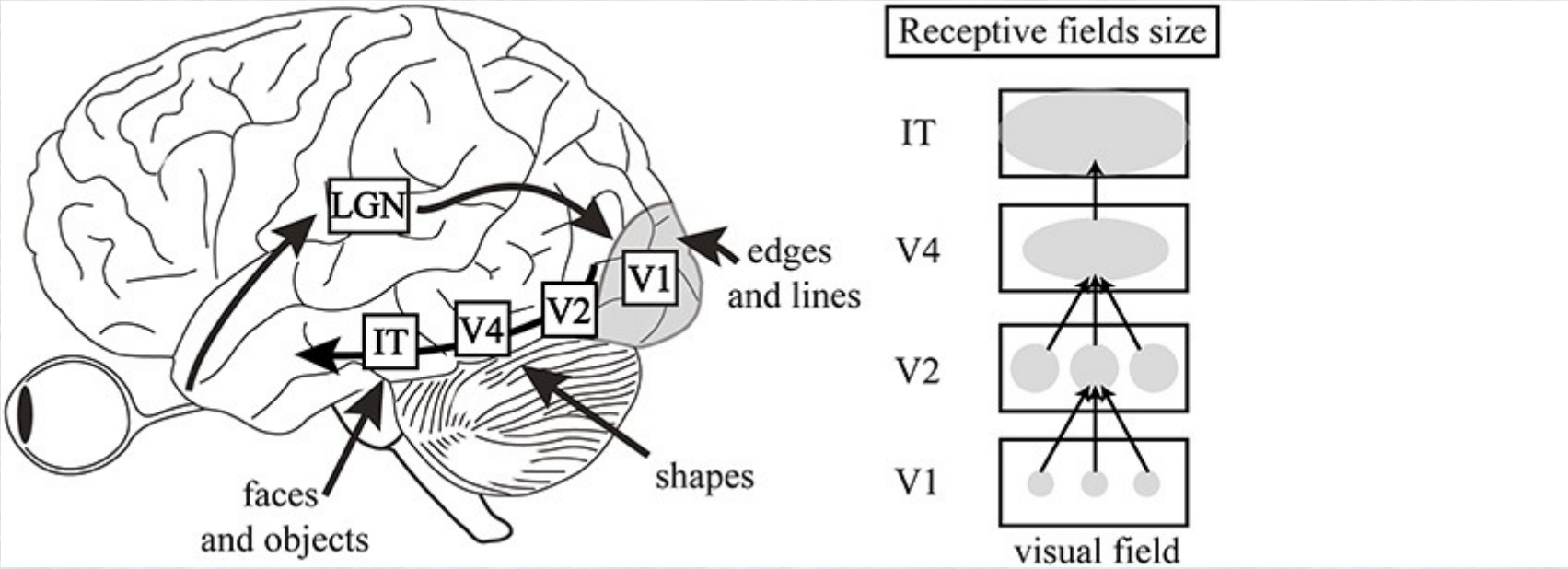
Receptive fields



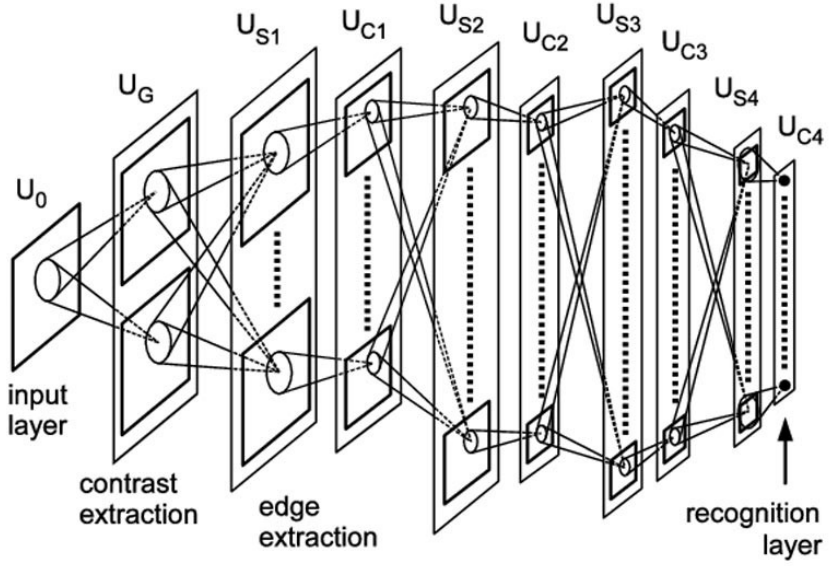
Neural tuning



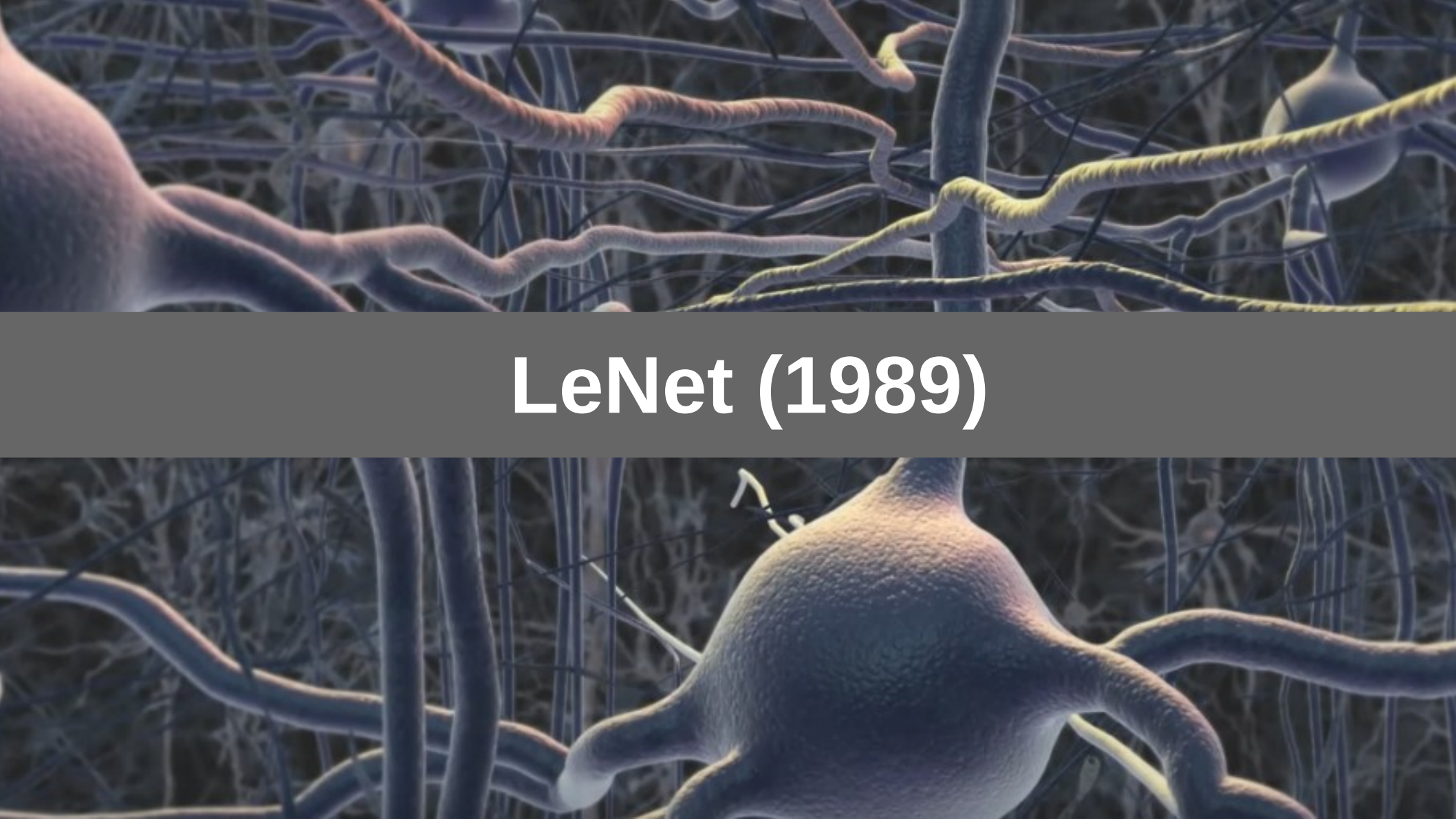
Simple to complex



Neocognitron



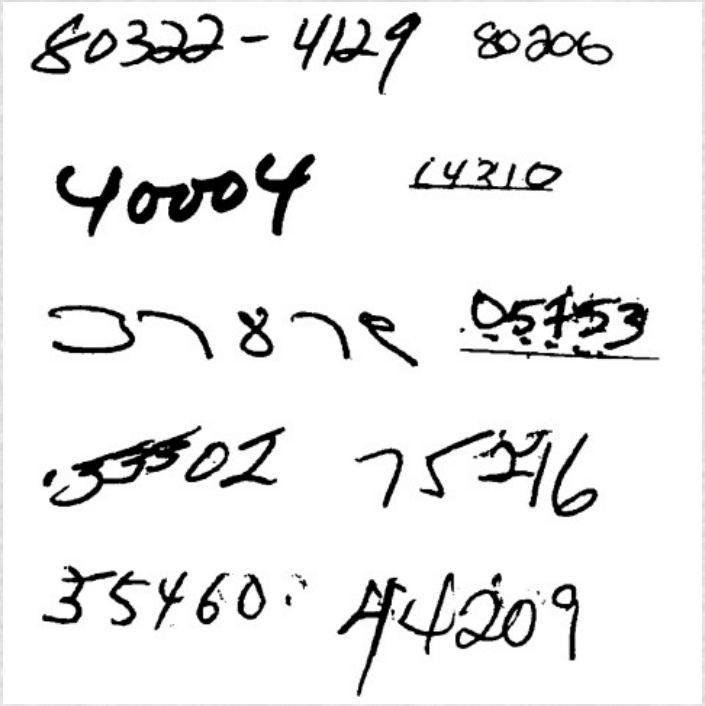
Kunihiko Fukushima



LeNet (1989)

Recognition of handwritten digits

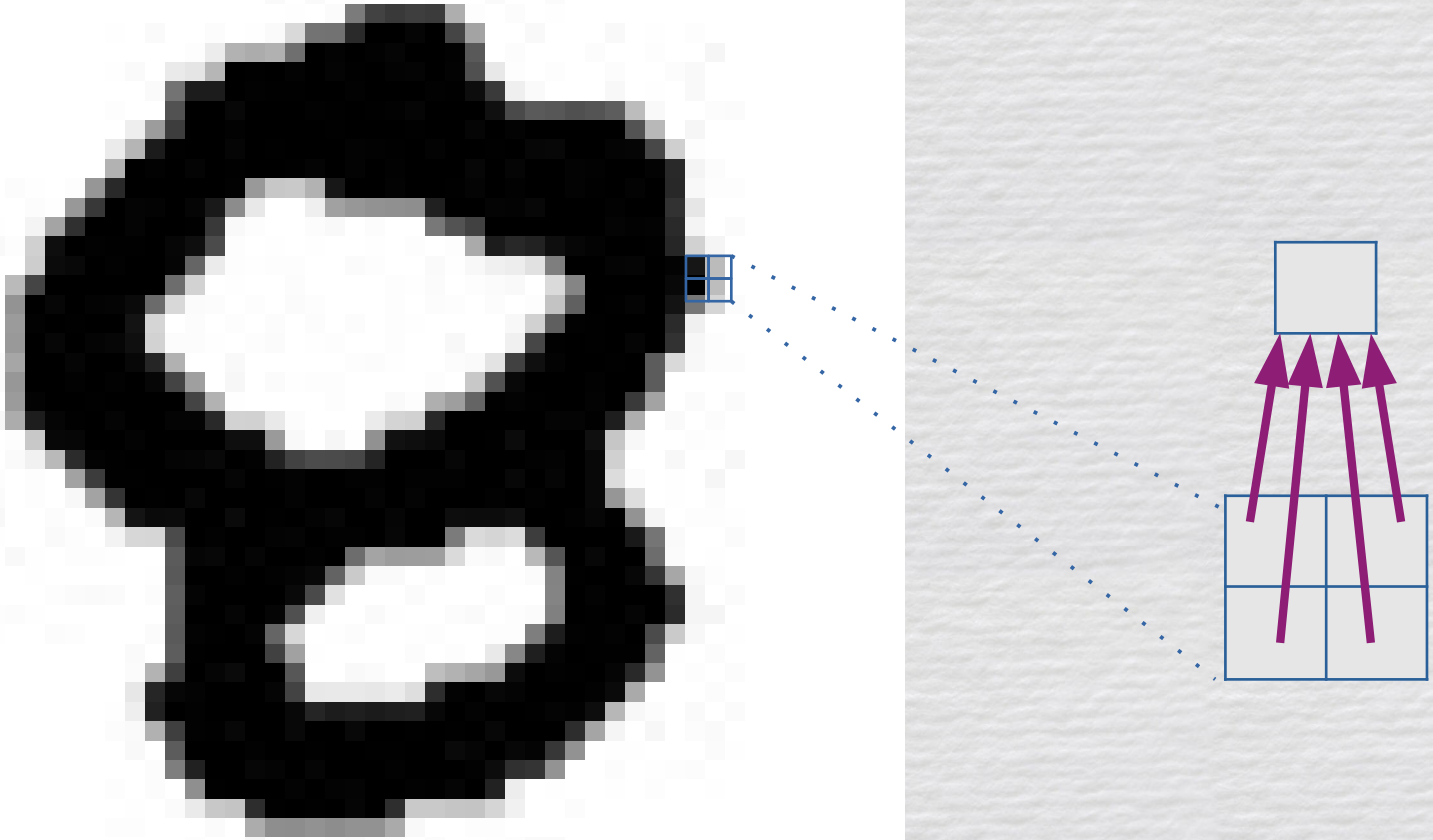
Yann LeCun



SUN-4/260

AT&T DSP 32C

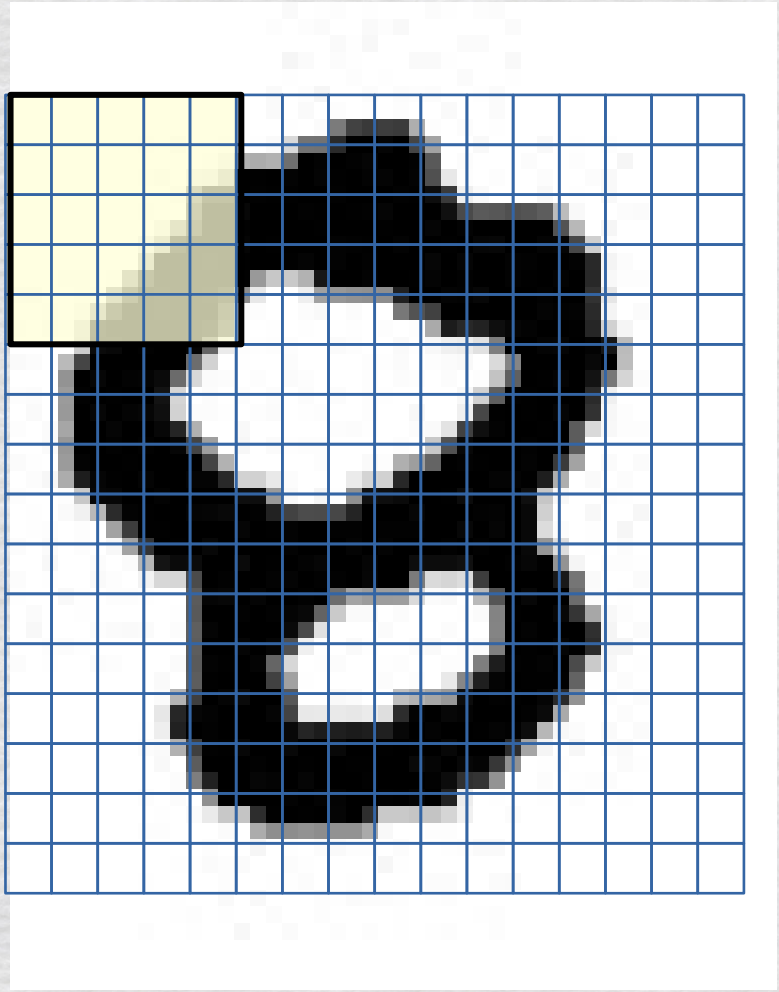
Convolutional filter



Input	Weights	Multiply	Add	Output
0 1 0 0 1 0 0 1 0	0 1 -1 -1 1 -1 -1 1 0	0 1 0 0 1 0 0 1 0	3	3
1 0 0 0 1 0 0 0 1	0 1 -1 -1 1 -1 -1 1 0	0 0 0 0 1 0 0 0 0	1	1
1 1 1 1 1 1 1 1 1	0 1 -1 -1 1 -1 -1 1 0	0 1 -1 -1 1 -1 0 1 0	-1	0

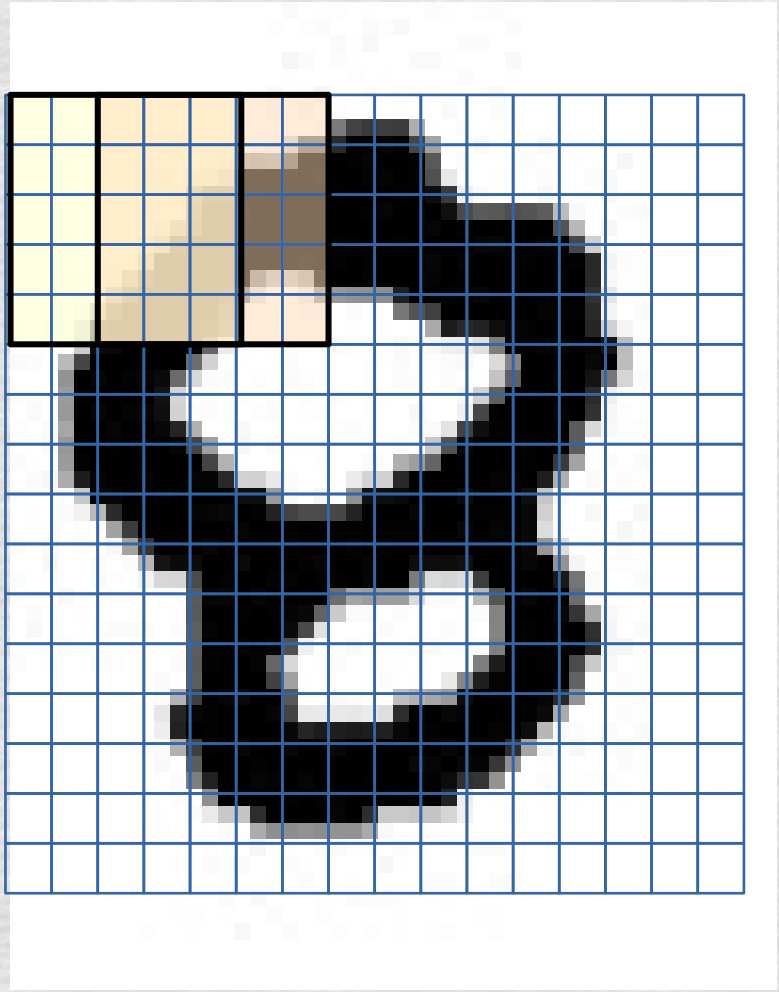
Convolutional filter

Filter applied to
top-left of image



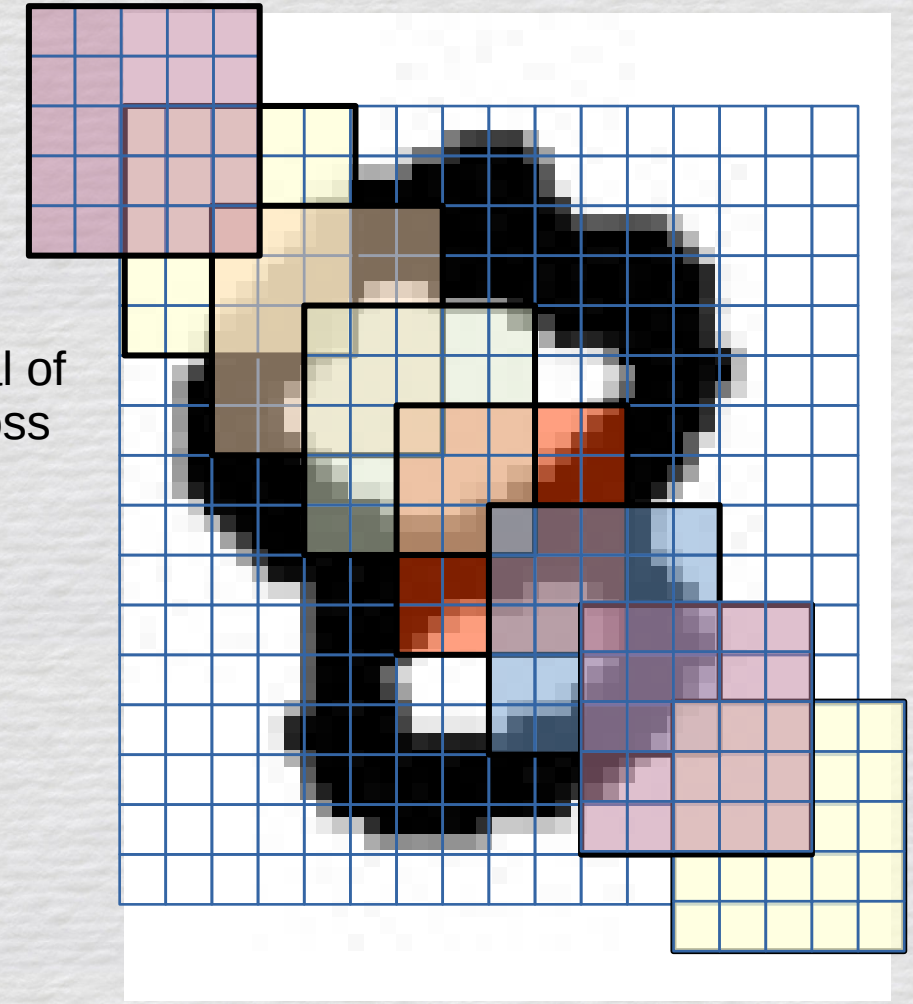
Convolutional filter

Next filter overlaps.

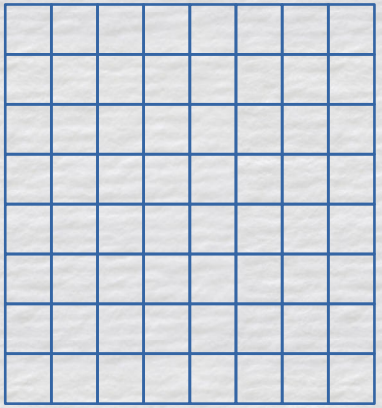


Convolutional filter

The filter is applied a total of 64 times across the image (8 examples shown here).



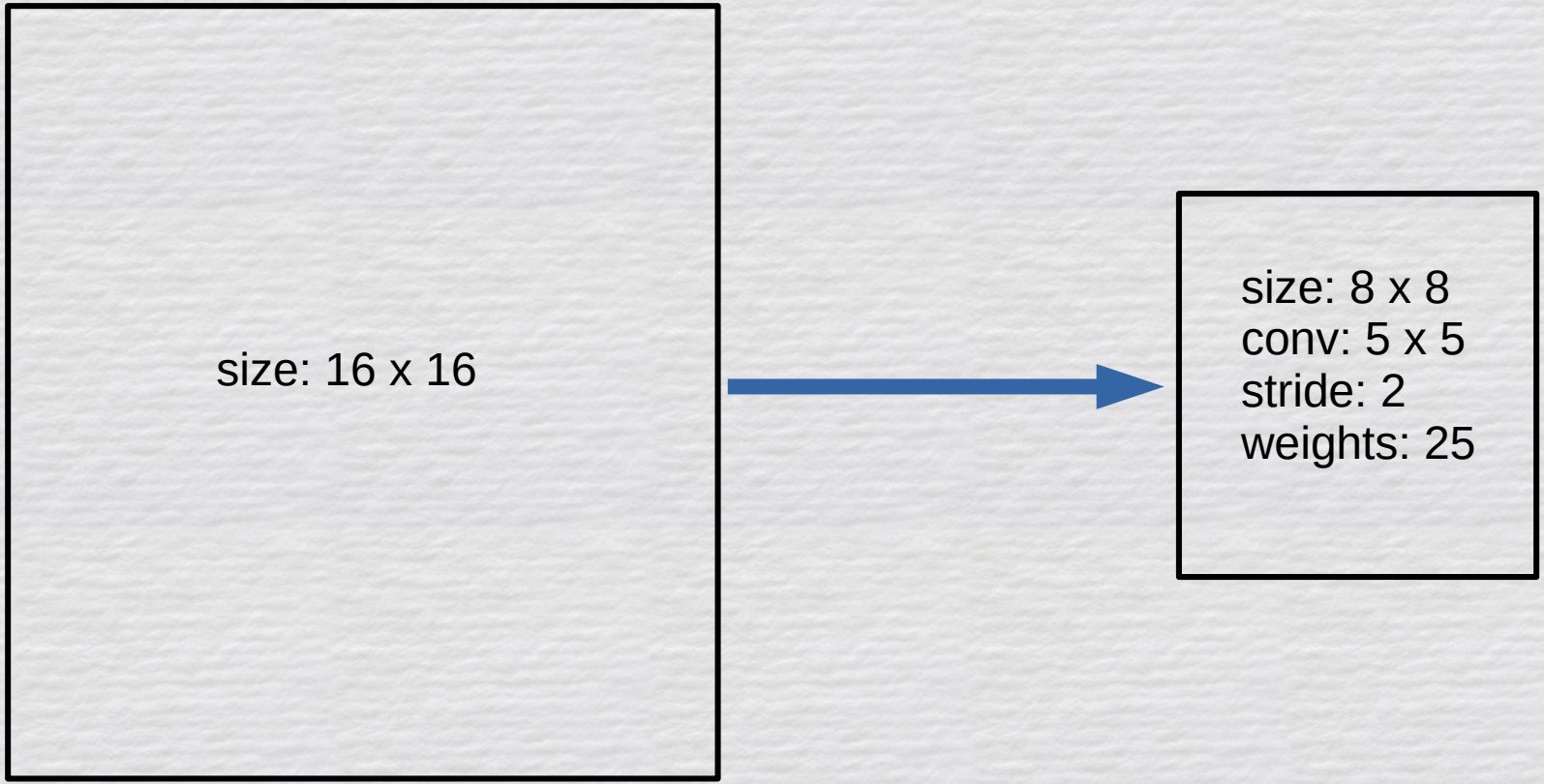
...and produces output activations for 64 hidden units, in an 8x8 grid.



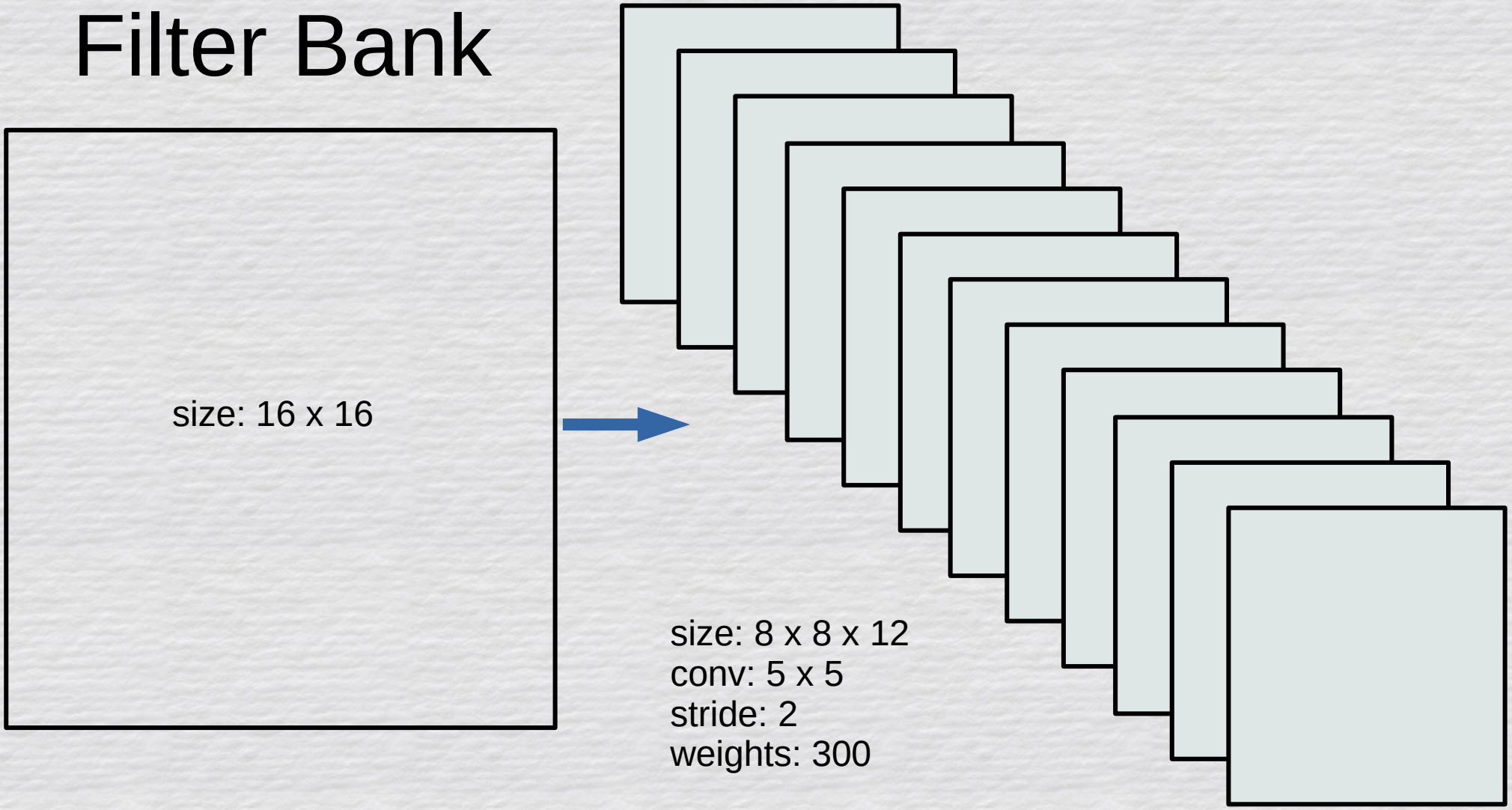
Weights: 25

(vs. a dense network: 16,384!)

Convolutional filter

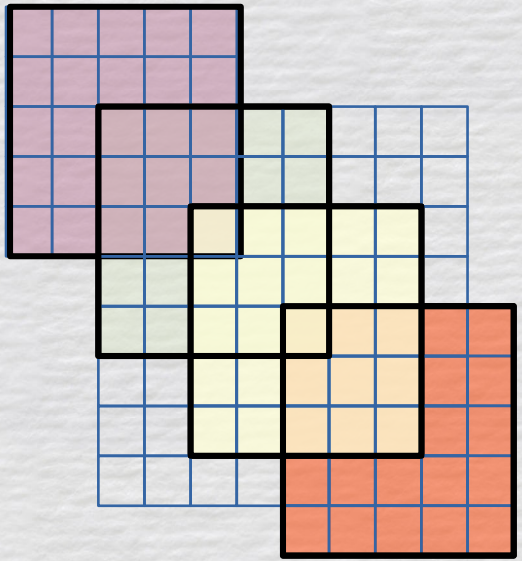


Filter Bank

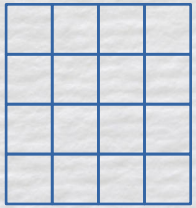


2nd Filter Bank

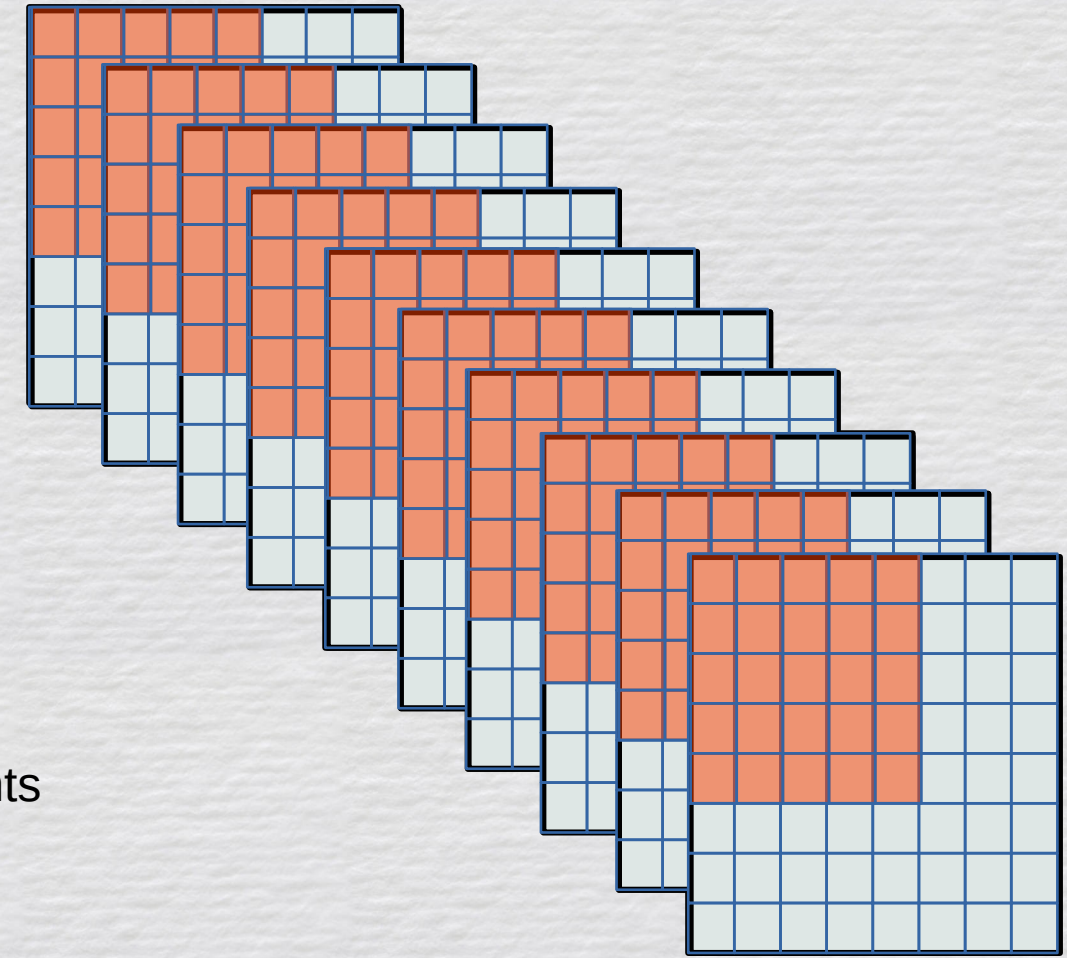
The 2nd filter is applied a total of 16 times across the 1st filter (4 examples shown here).



...and produces output activations for 16 hidden units, in an 4x4 grid.

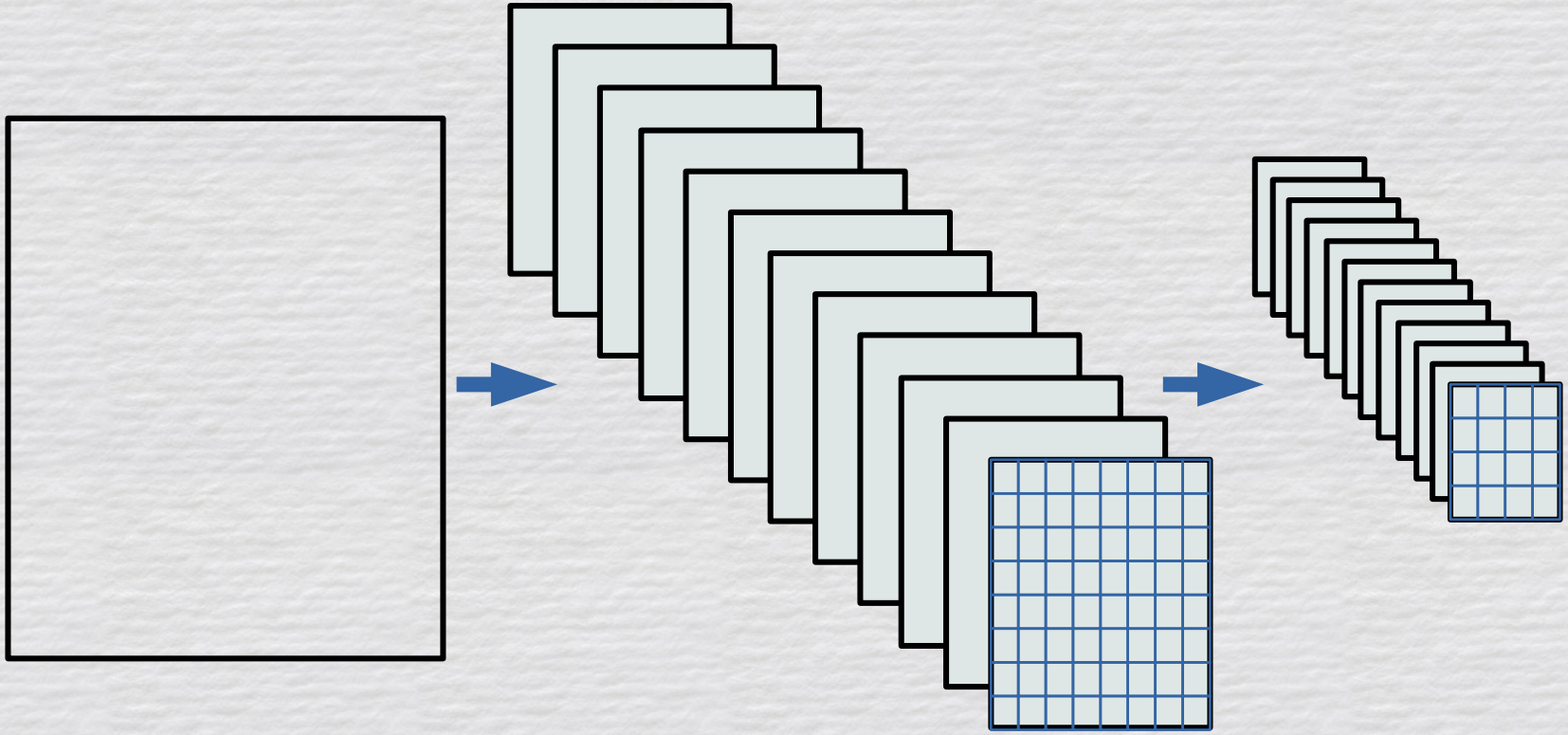


3D convolutional filter



The 3D filter has 5 x 5 x 12 = 300 weights

First Three Layers

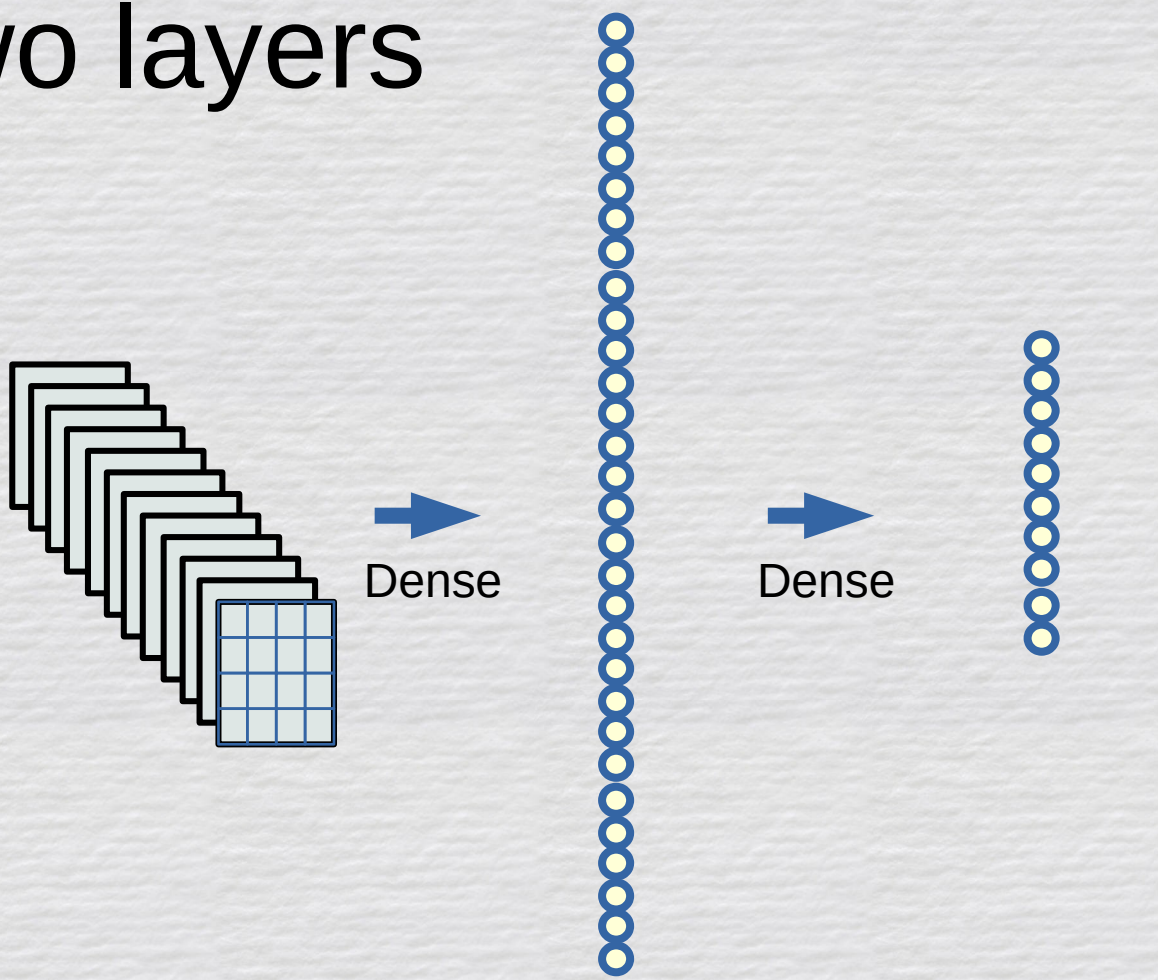


Image

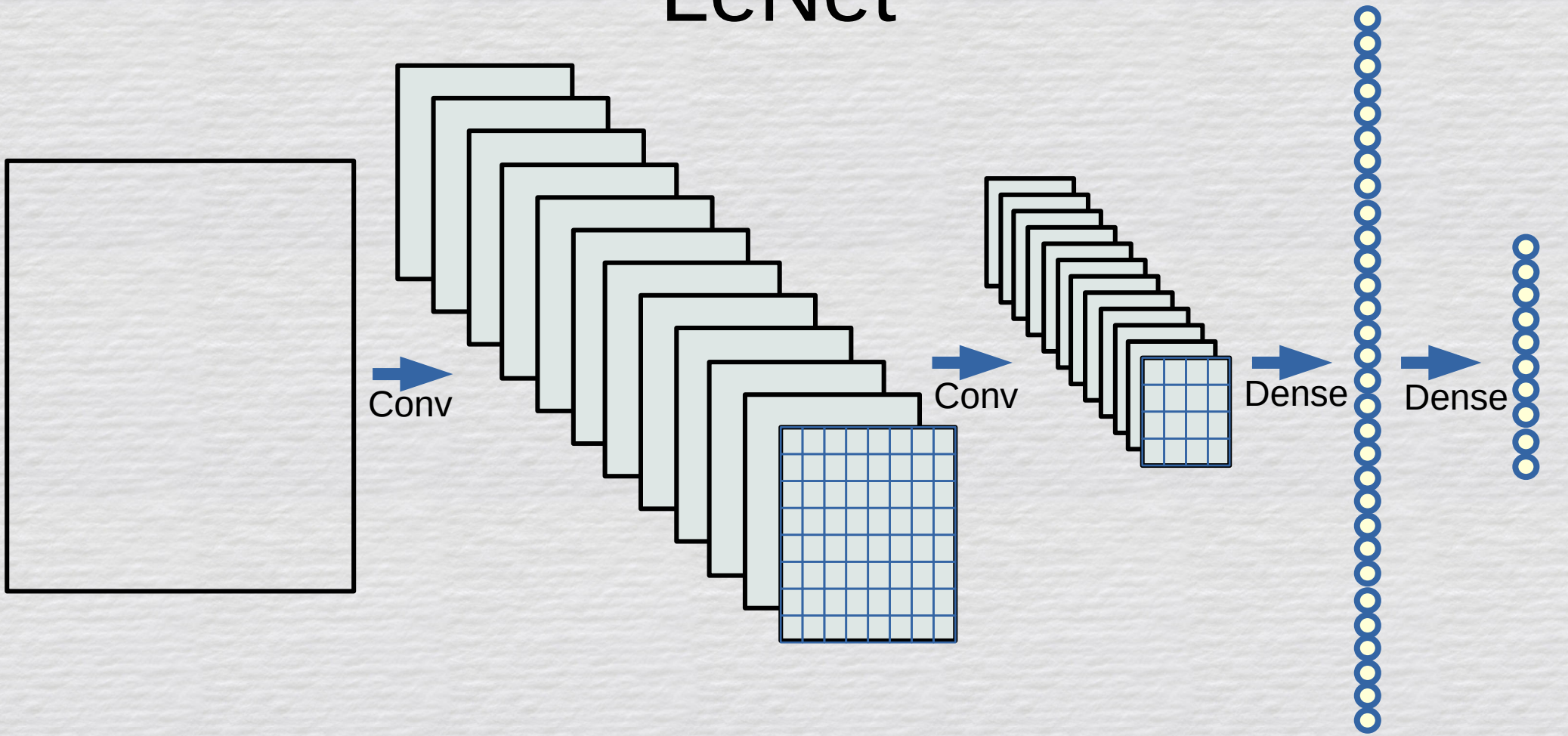
2D filters

3D filters

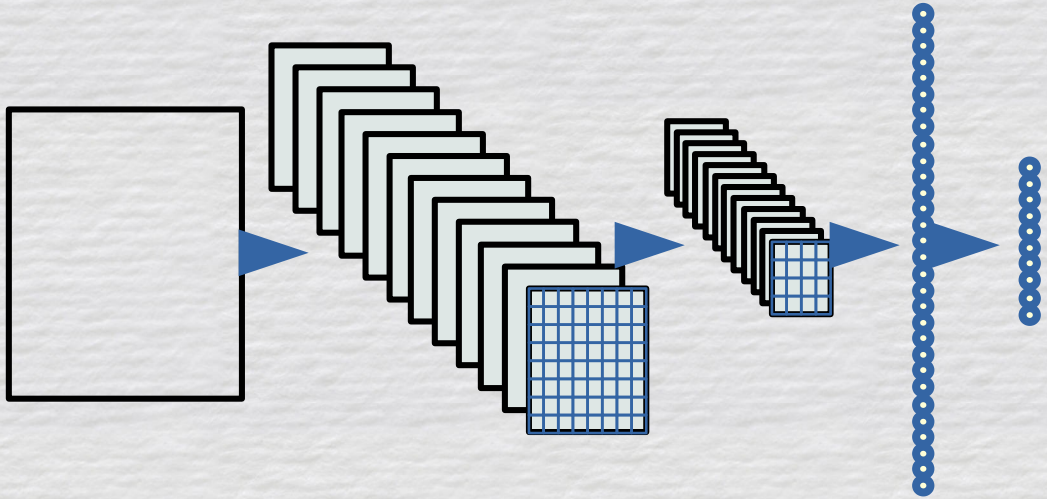
Last two layers



LeNet


Conv
Conv
Dense
Dense
Image
2D filters
3D filters
Hidden
Output

LeNet



Trained with backprop

7,291 digits, 23 times each (167,693 trials).

5% errors on a test set.

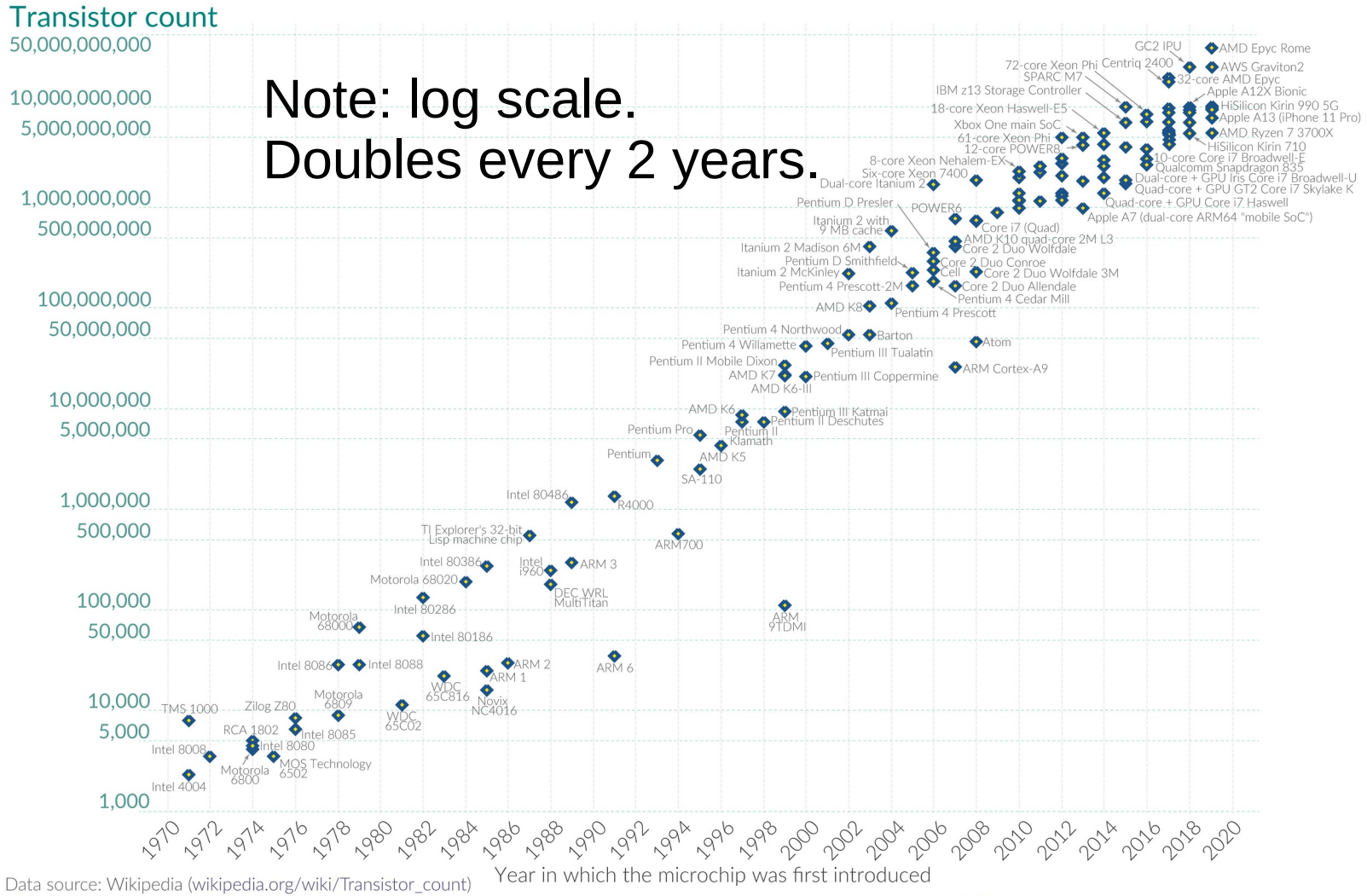
10 digits/second

AlexNet (2012)

Moore's Law: The number of transistors on microchips doubles every two years

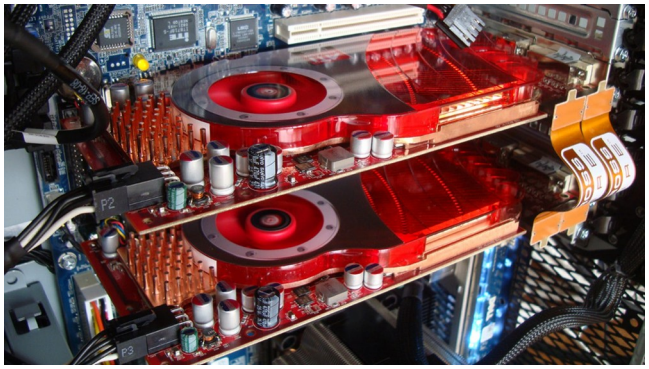
Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

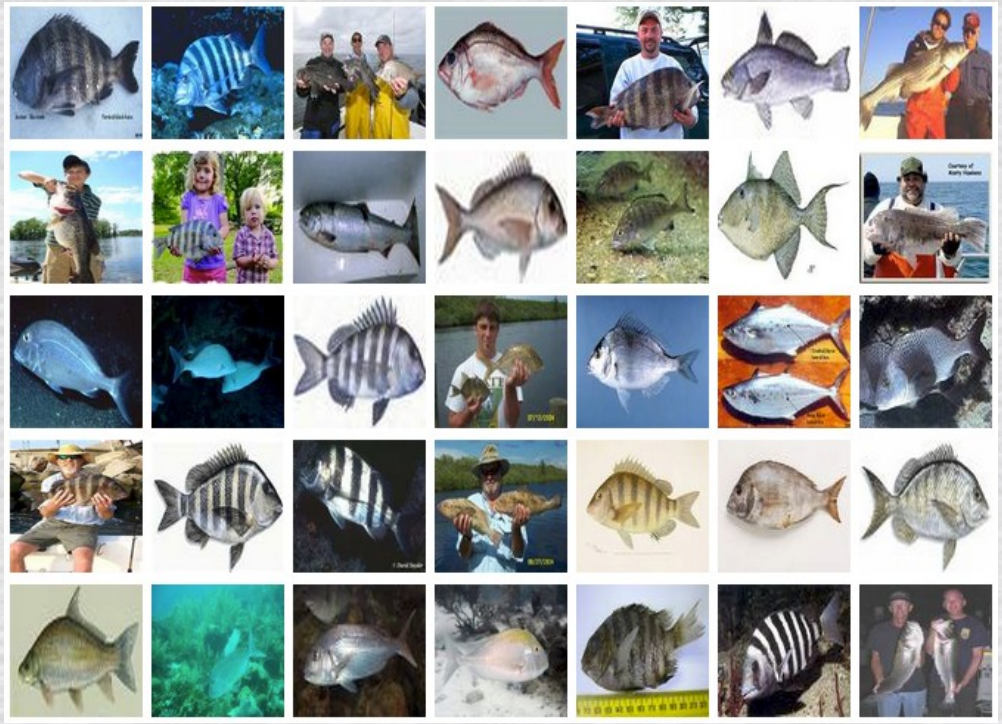


Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

GPUs and CPUs.
Note: linear scale.
Doubles every 2 years.



Imagenet (2009-)



Some of the fish pictures in ImageNet

14 million images

20,000 categories

100s examples per category

Legacy of AlexNet

100,000+ citations

Acquired by Google

Big Tech headhunts brain-inspired AI academics

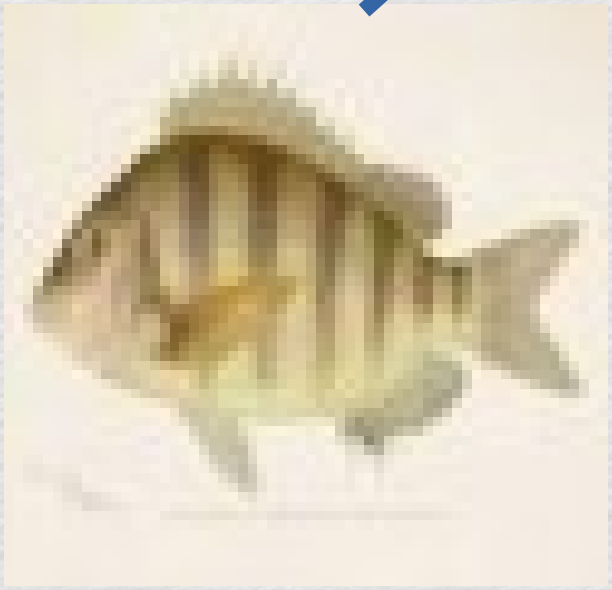
Yann LeCun
Now @ facebook

Geoff Hinton
Now @ Google

Higher-res, and colour

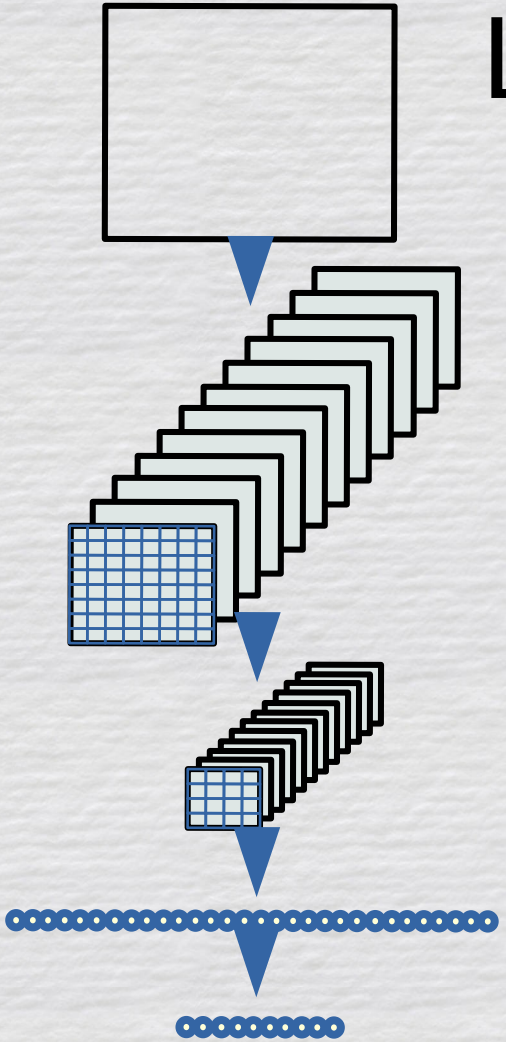
LeNet

size: 16 x 16



AlexNet

size: 256 x 256 x 3



LeNet

Image

Conv

Conv

Dense

Dense

Category

AlexNet

Image

Conv

Conv

Conv

Conv

Conv

Dense

Dense

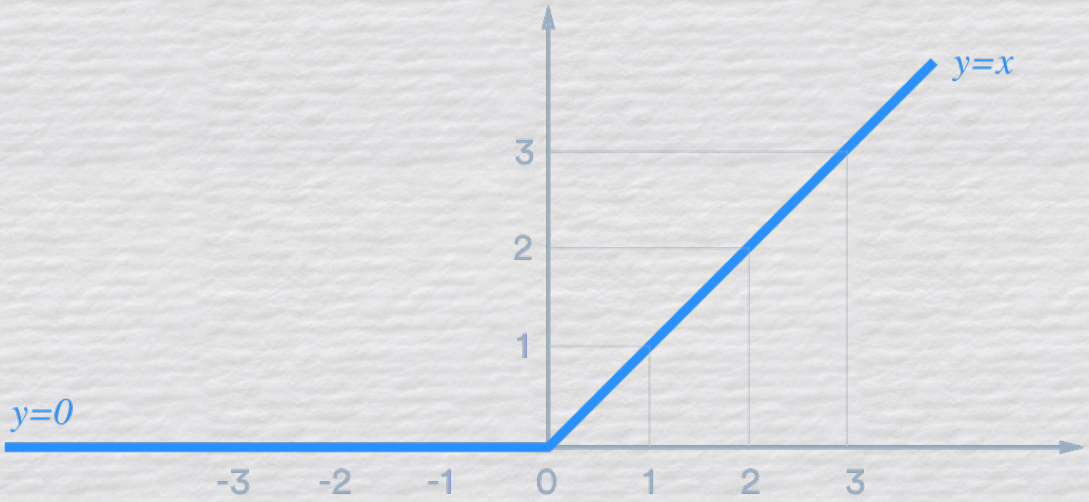
Dense

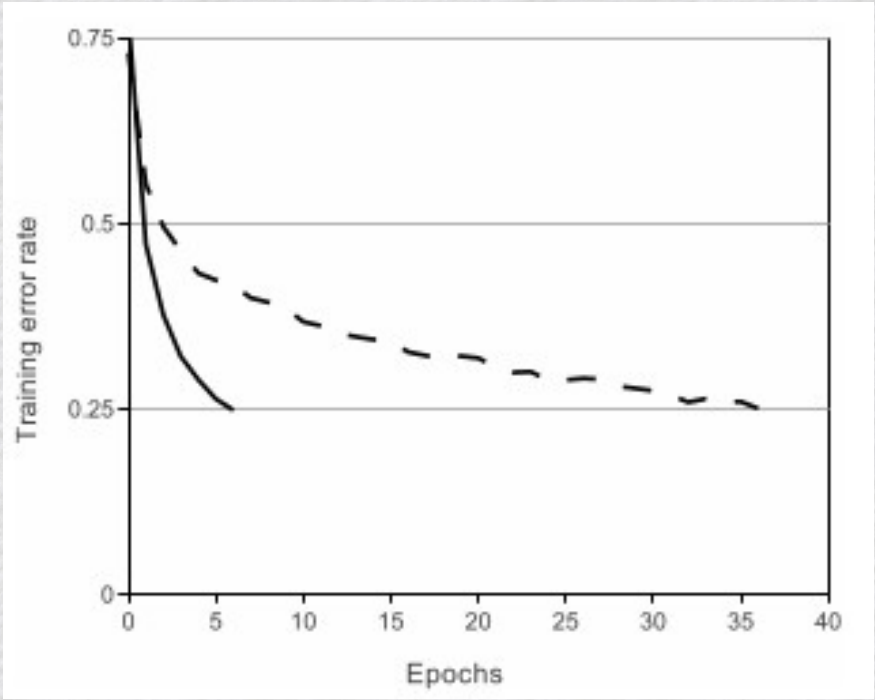
Category

Deeper

ReLU


1
0.5
0
−6
−4
−2
0
2
4
6


y=x
y=0
3
2
1
-3
-2
-1
0
1
2
3


0.75
0.5
0.25
0
Training error rate
0
5
10
15
20
25
30
35
40
Epochs

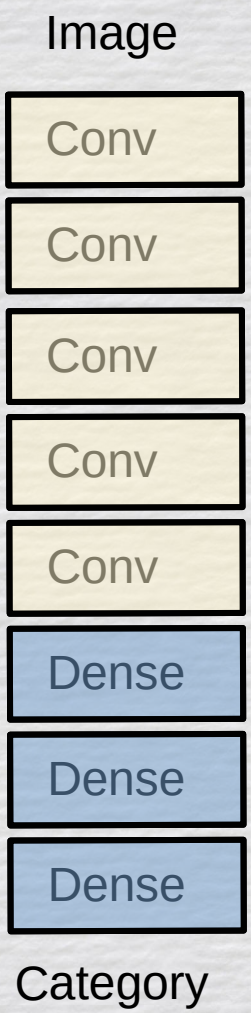


Image
Conv
Conv
Conv
Conv
Conv
Dense
Dense
Dense
Category

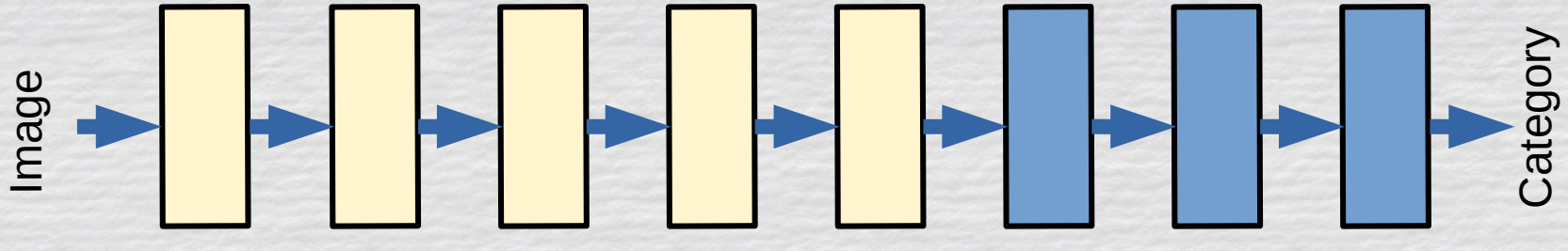


Image
Category

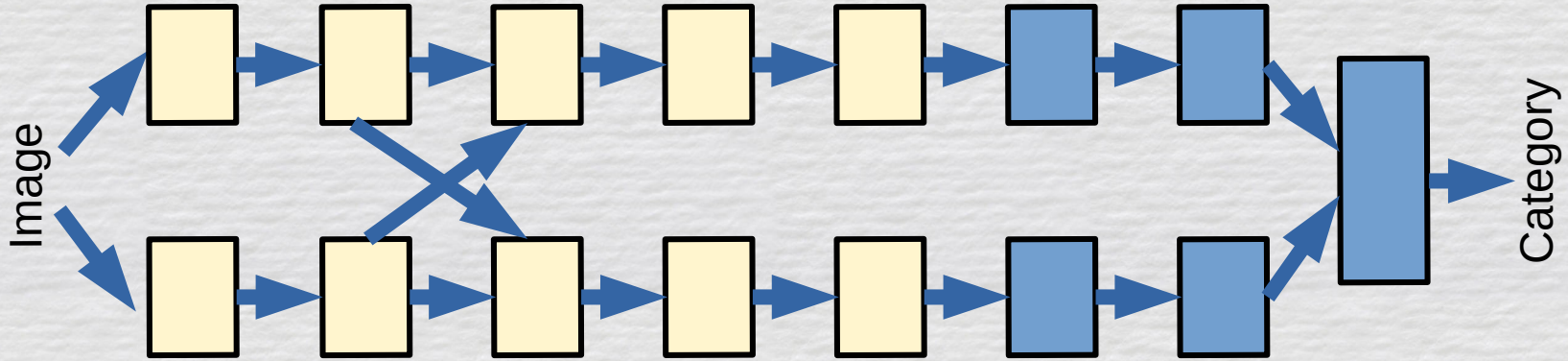
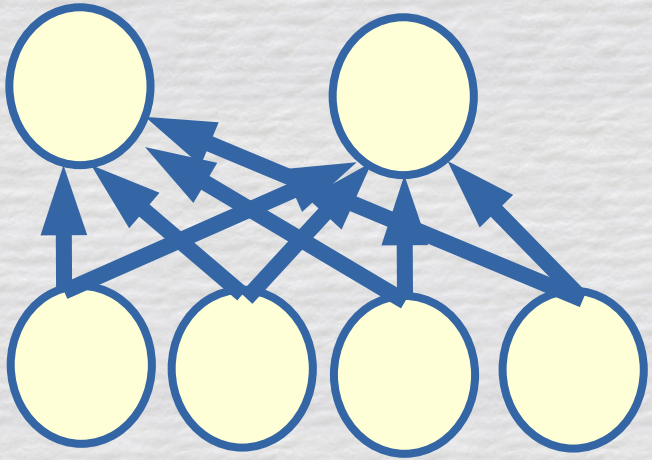
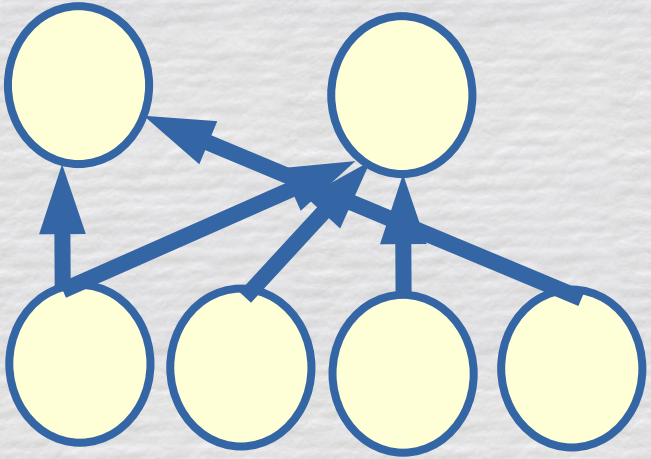


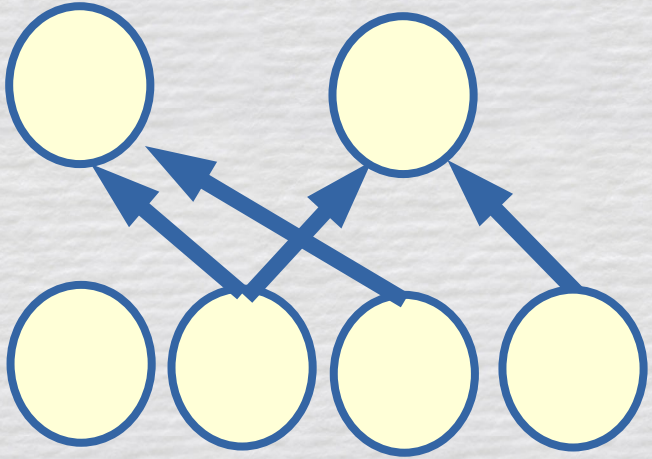
Image
Category

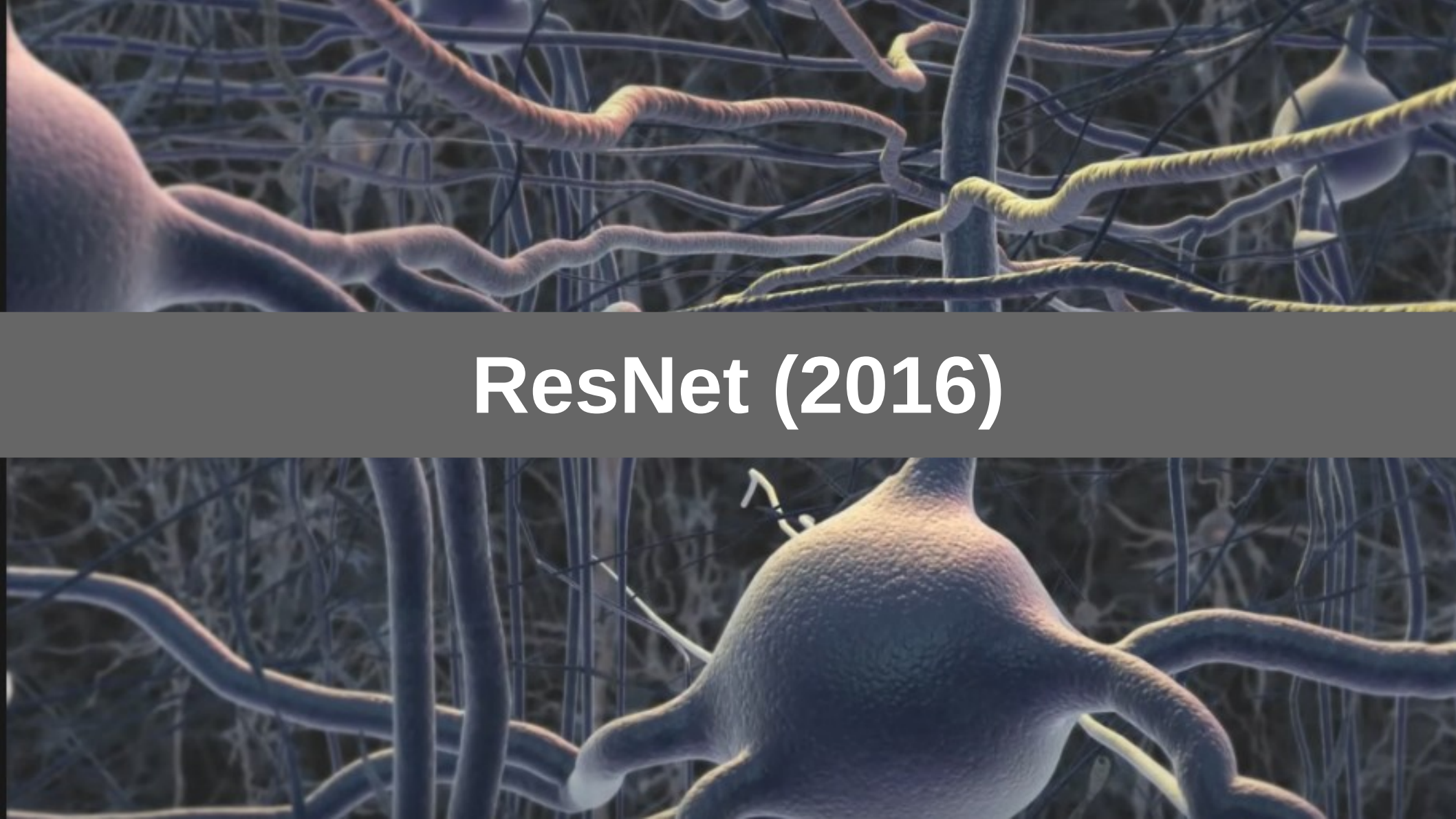
Grouping

Dropout

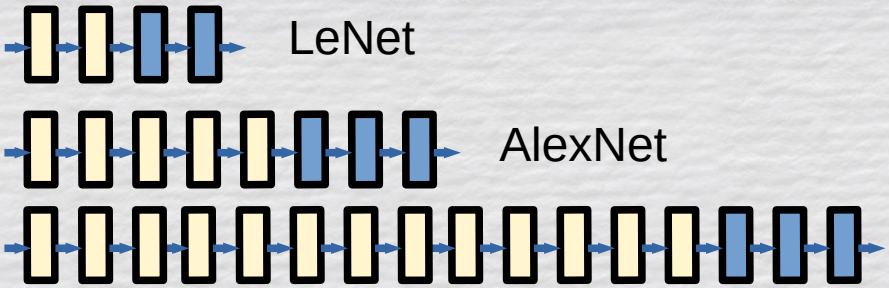








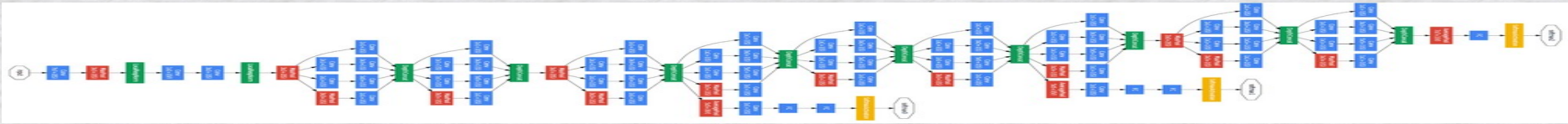
ResNet (2016)



LeNet

AlexNet

VGG

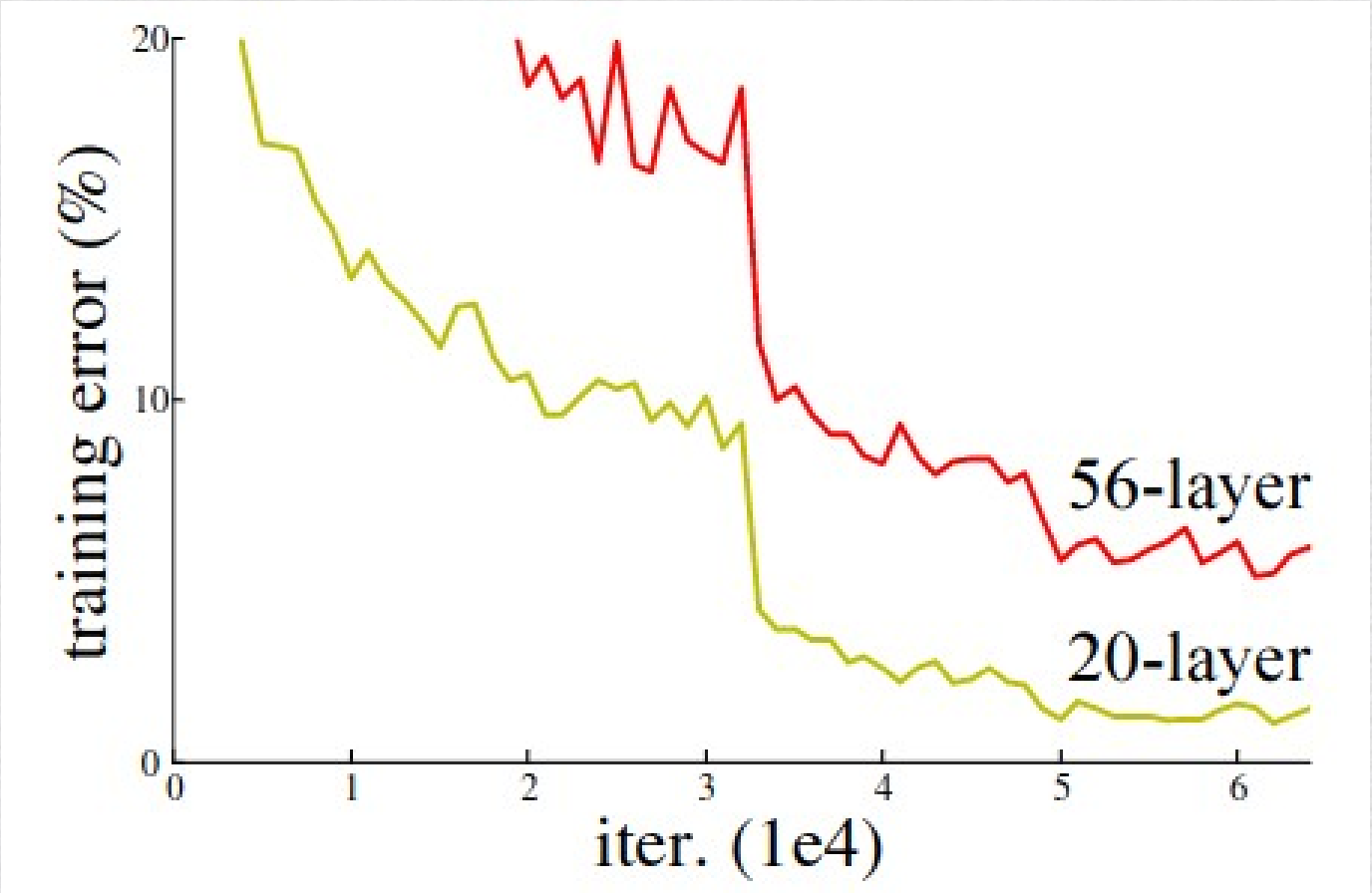


Inception

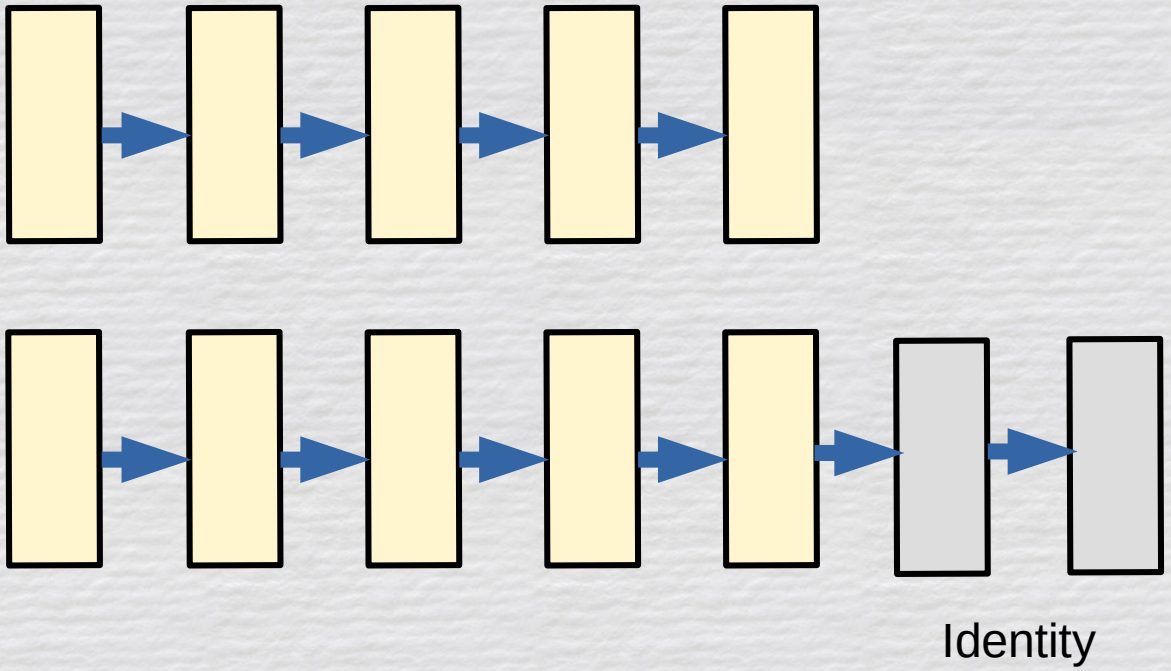
Network degradation

Kaiming He

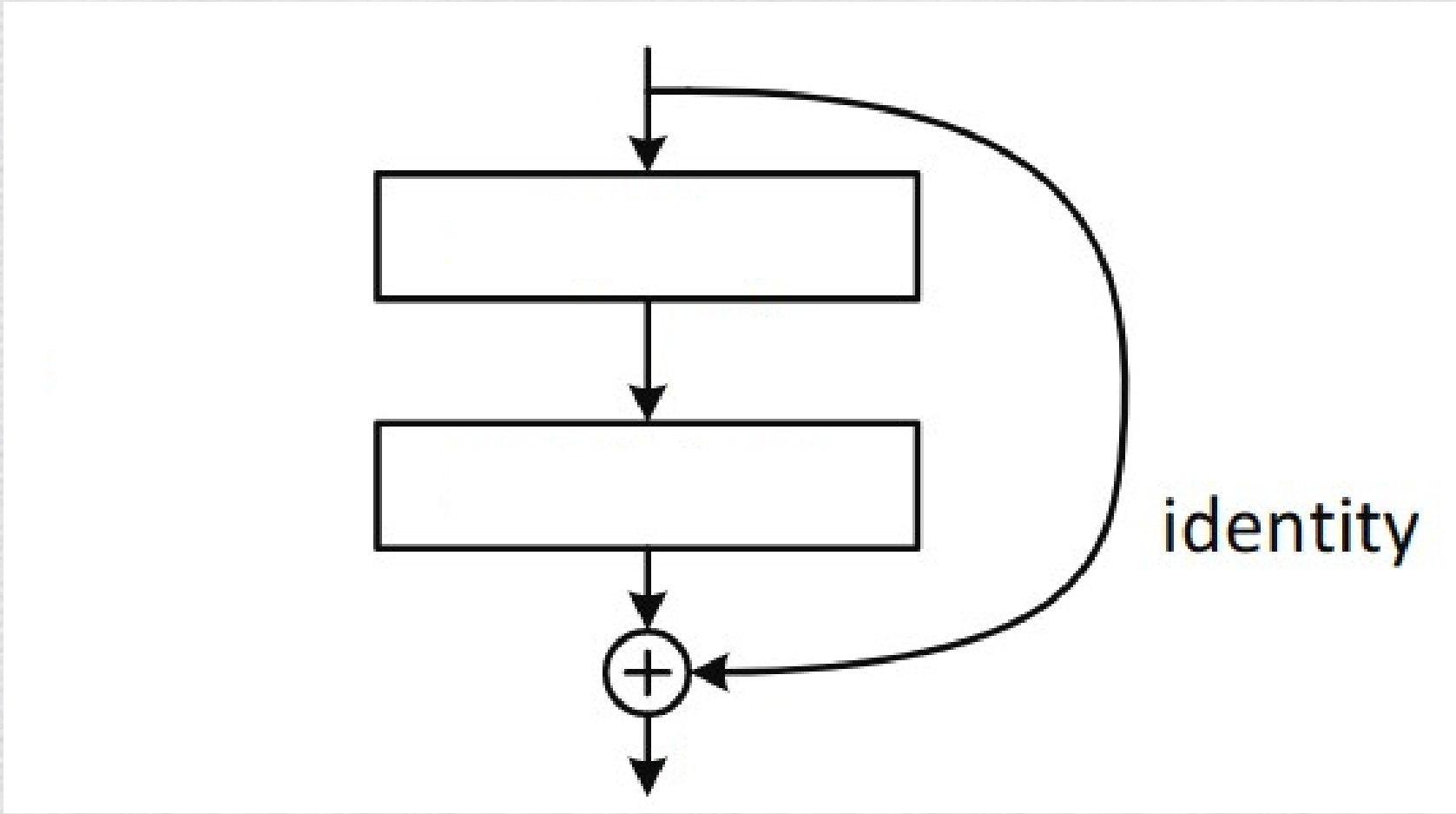
Jian Sun



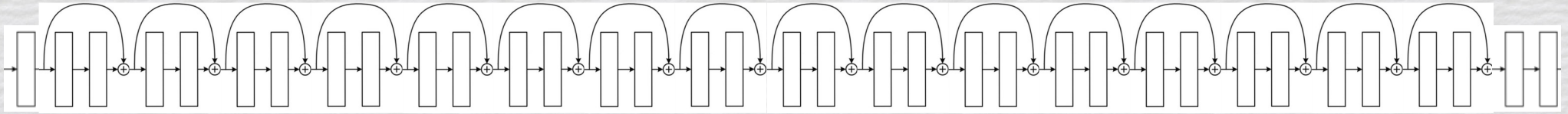
Network degradation



Residual Block

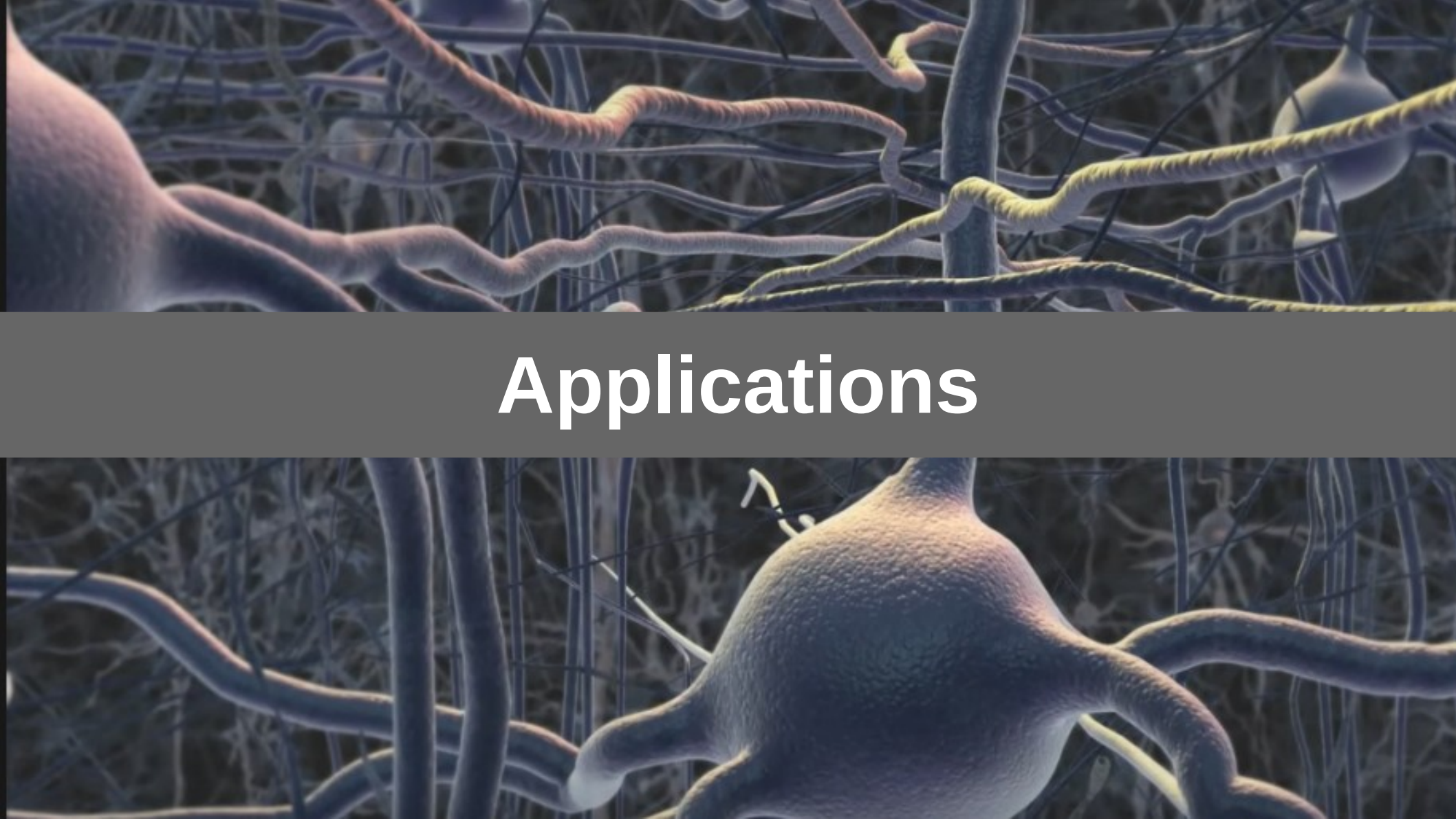


ResNet 34



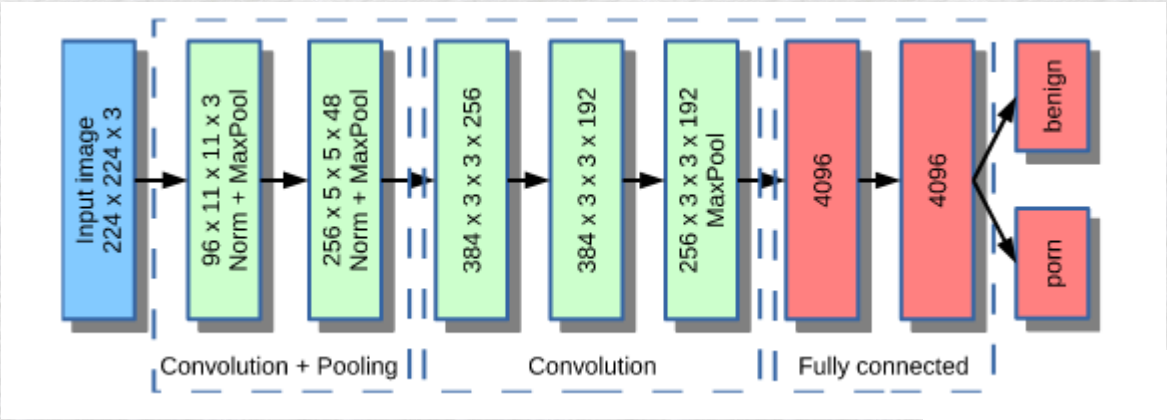
ResNet 152

ResNet152 with ResNet 34 for scale
(zoom in to slide for detail)



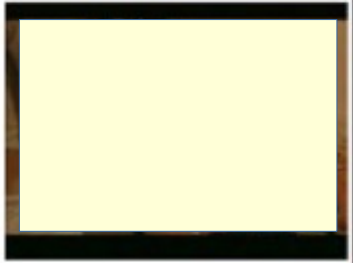
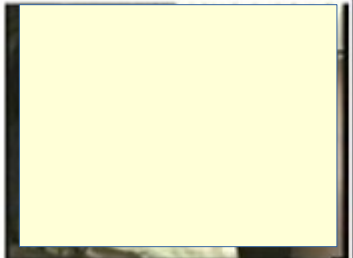
Applications

Content moderation

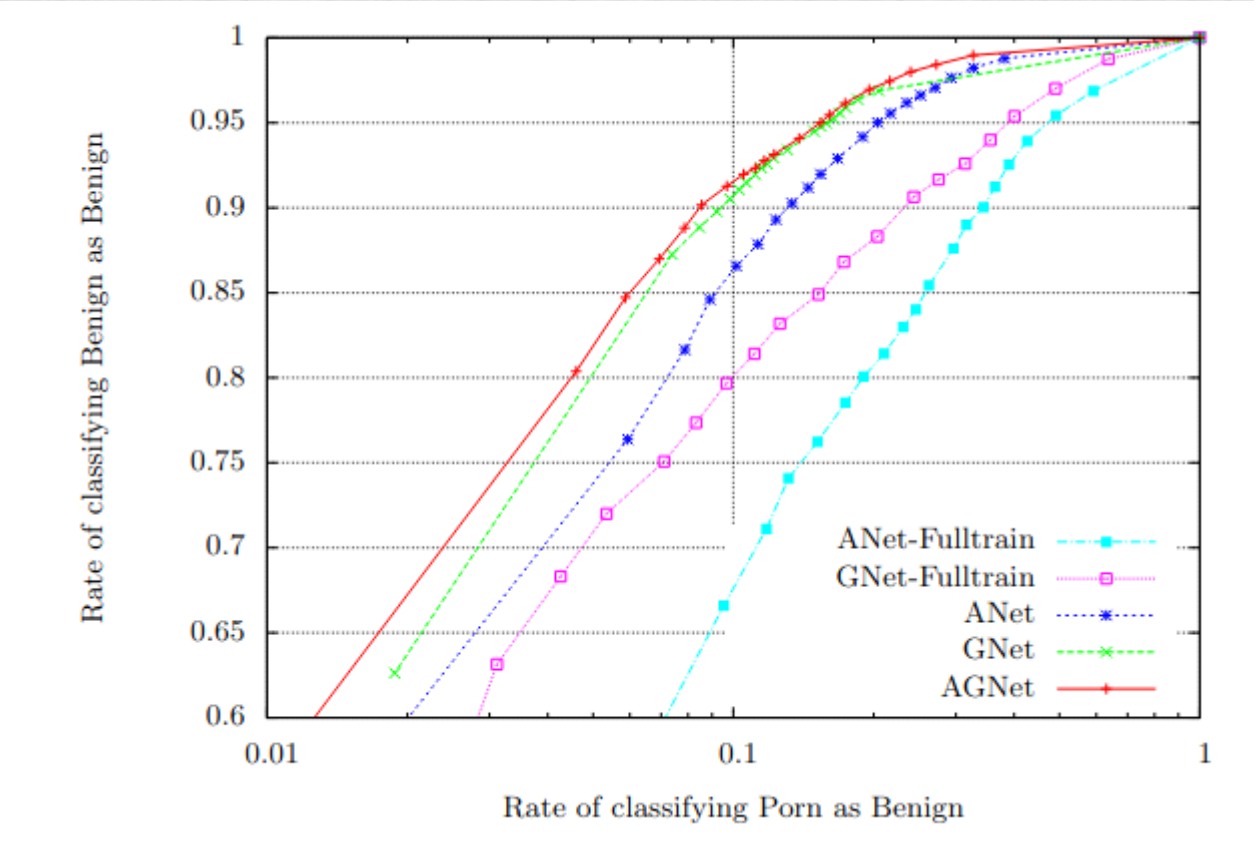


Mohammed Moustafa (2015)

Table 2. NPDI dataset samples

Porn	Non-porn ("difficult")	Non-porn ("easy")
		
		

Content moderation

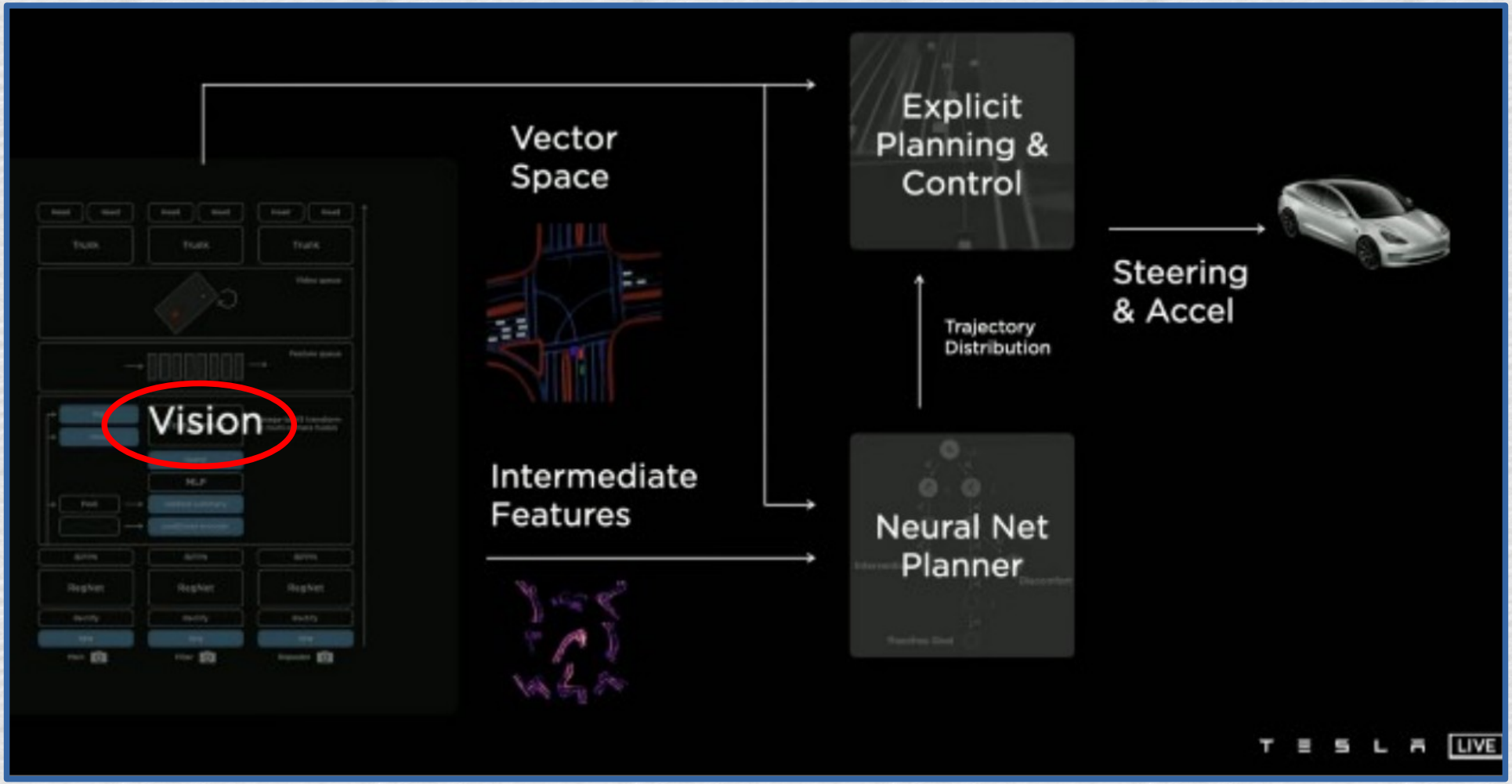


Stills

Approach	Accuracy (%)
BossaNova (HueSIFT) [2]	89.5±1
BossaNova VD (BinBoost16) [3]	90.9±1
Proposed ANet	92.01±3
Proposed GNet	93.7±3
Proposed AGNet	**93.8±2**
Proposed AGbNet	**94.1±2**

Videos

Tesla AutoPilot



Tesla AutoPilot

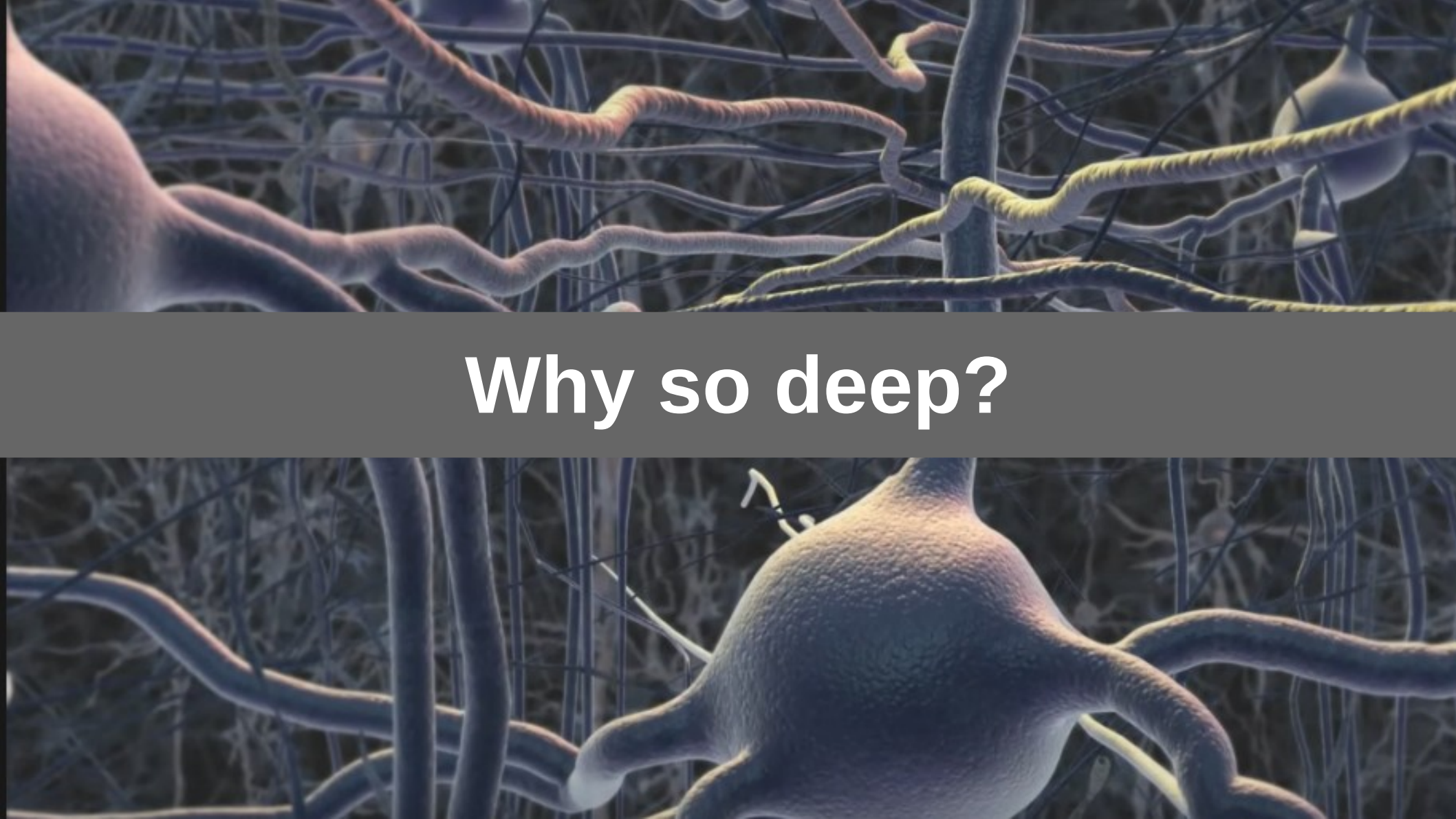
Autopilot is a complex system.

...but nearly the first thing that happens to the data from each of the 8 cameras is that it's passed through a **RegNet.** This is a development of the ResNet system.

Tesla AutoPilot

Autopilot is a complex system.

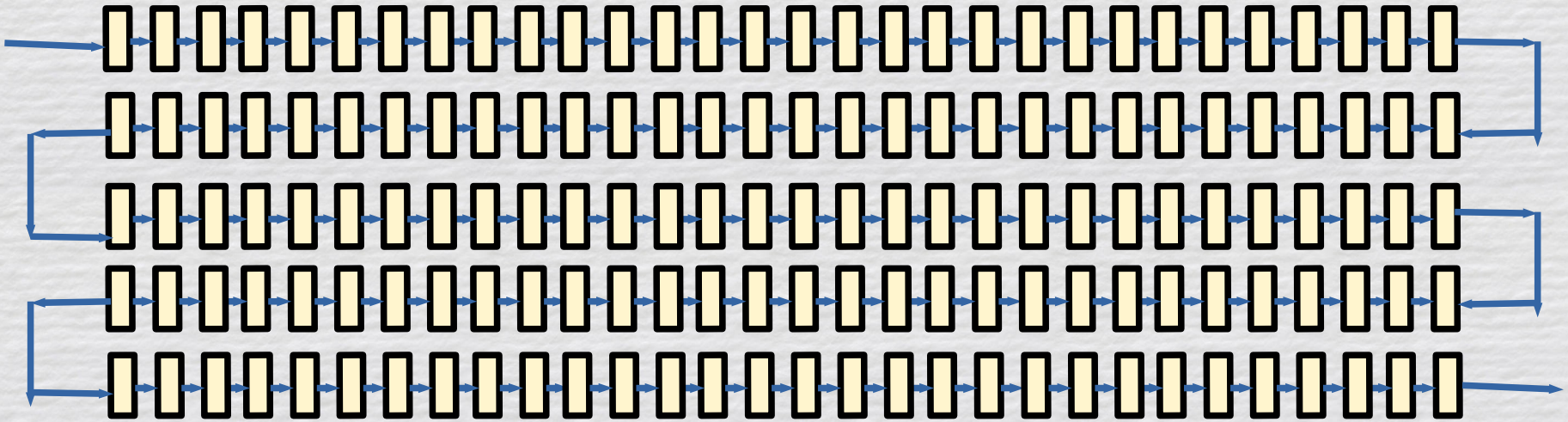
...but nearly the first thing that happens to the data from each of the 8 cameras is that it's passed through a **RegNet.** This is a development of the ResNet system.



Why so deep?

The paradox

A multilayer feedforward network, with sufficient hidden units can represent any deterministic mapping between its inputs and its outputs - Hornik, Stinchcombe & White (1989)



ResNet152 (!)

Lower complexity

2-layer dense network

150,000 inputs
75,000 hidden
1,000 outputs

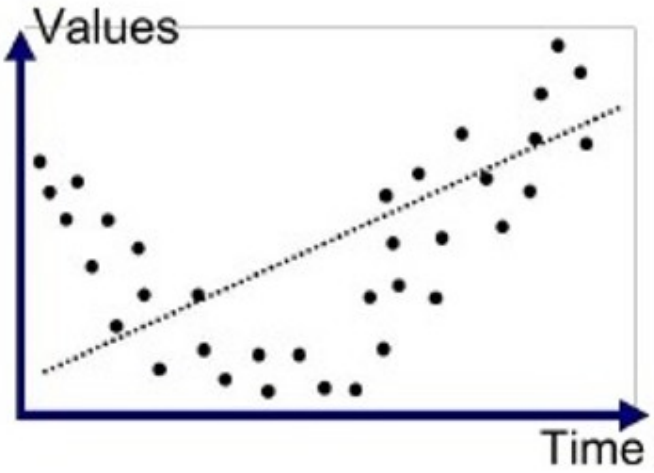
= 11 billion connections

50-layer ResNet

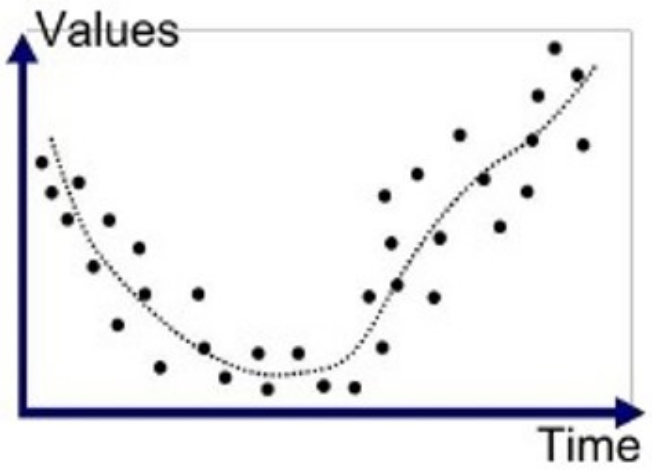
150,000 inputs
1,000 outputs

24 million connections
400 times simpler

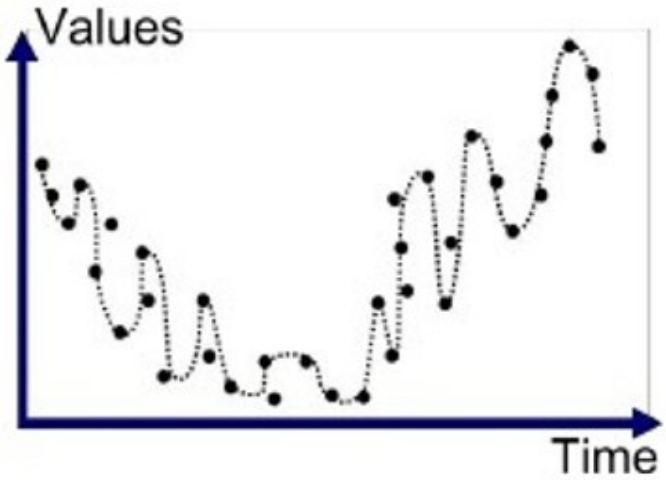
Overfitting



Underfitted

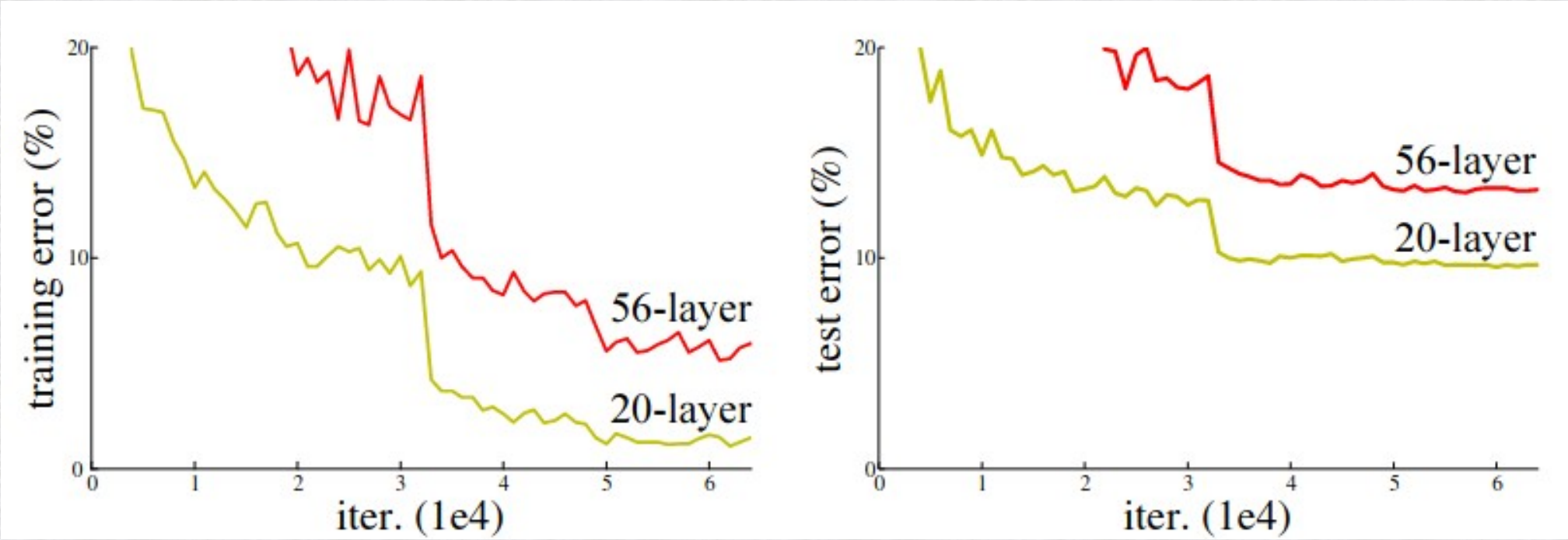


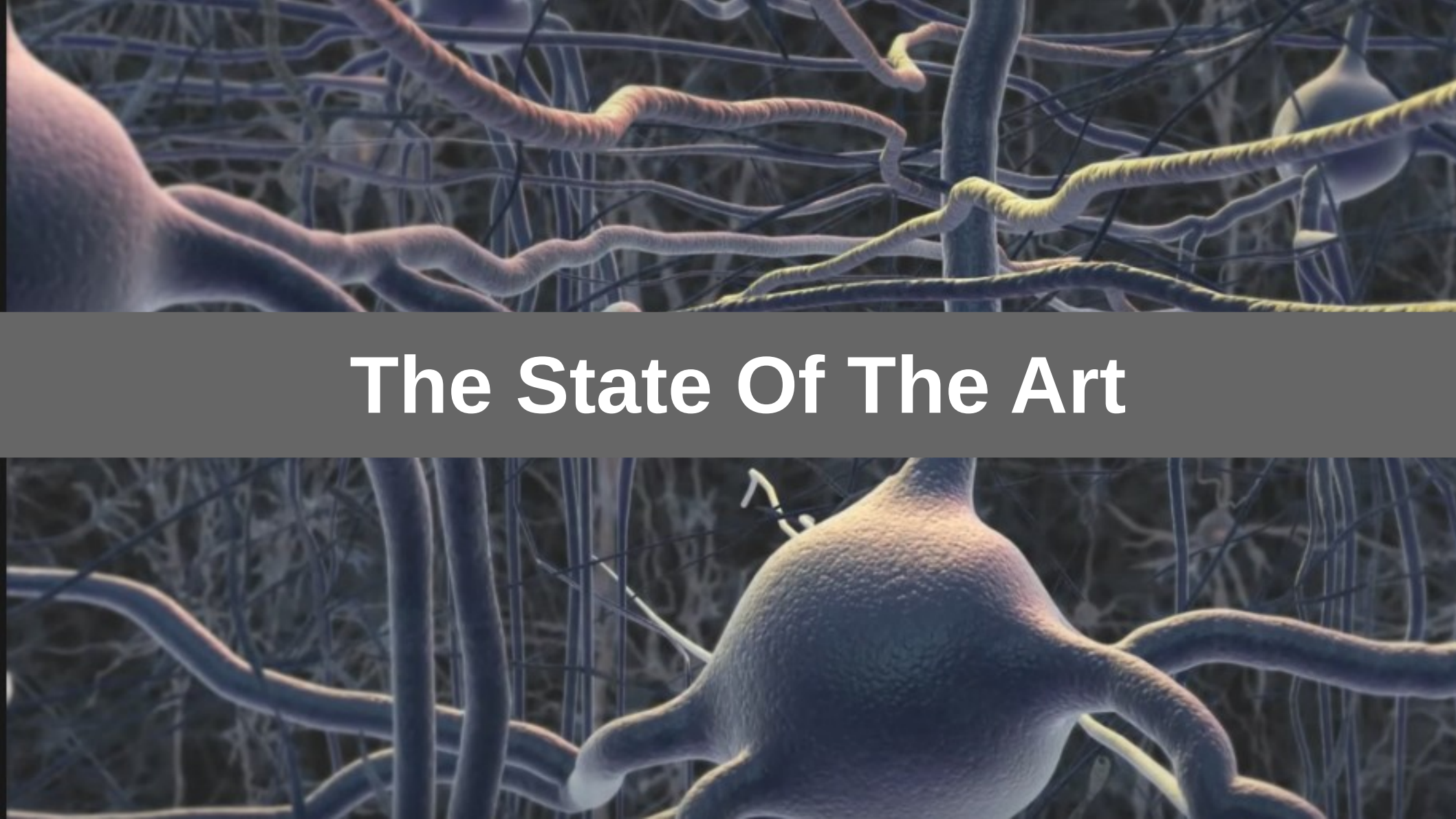
Good Fit/Robust



Overfitted

Overfitting





The State Of The Art

How good are they?

“recent advances from machine learning led to the discovery of hierarchical neural network models that achieved near-human-level performance level on challenging object categorization tasks”

- Yamins & DiCarlo (2016)

Can you be more specific?

PNASNet:

- 96.2% Top-5 accuracy on ImageNet (Liu et al., 2018)

(1) Hatstand

(2) Orange

(3) Battleship

(4) Dandelion

(5) Cat

How about a sensible answer?

PNASNet:

~73% Top-1 accuracy on ~300 ImageNet categories (Barbu et al., 2019)

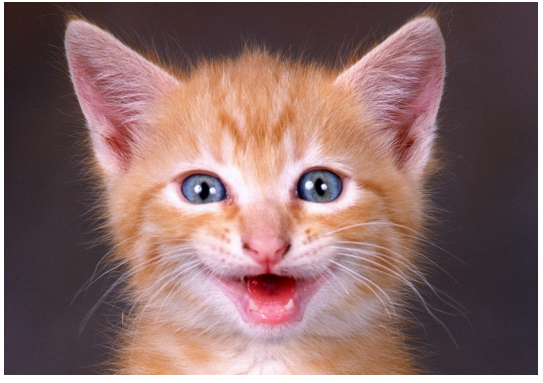
(1) Cat

(2) Orange

(3) Battleship

(4) Dandelion

(5) Hatstand



Barbu et al. (2019)

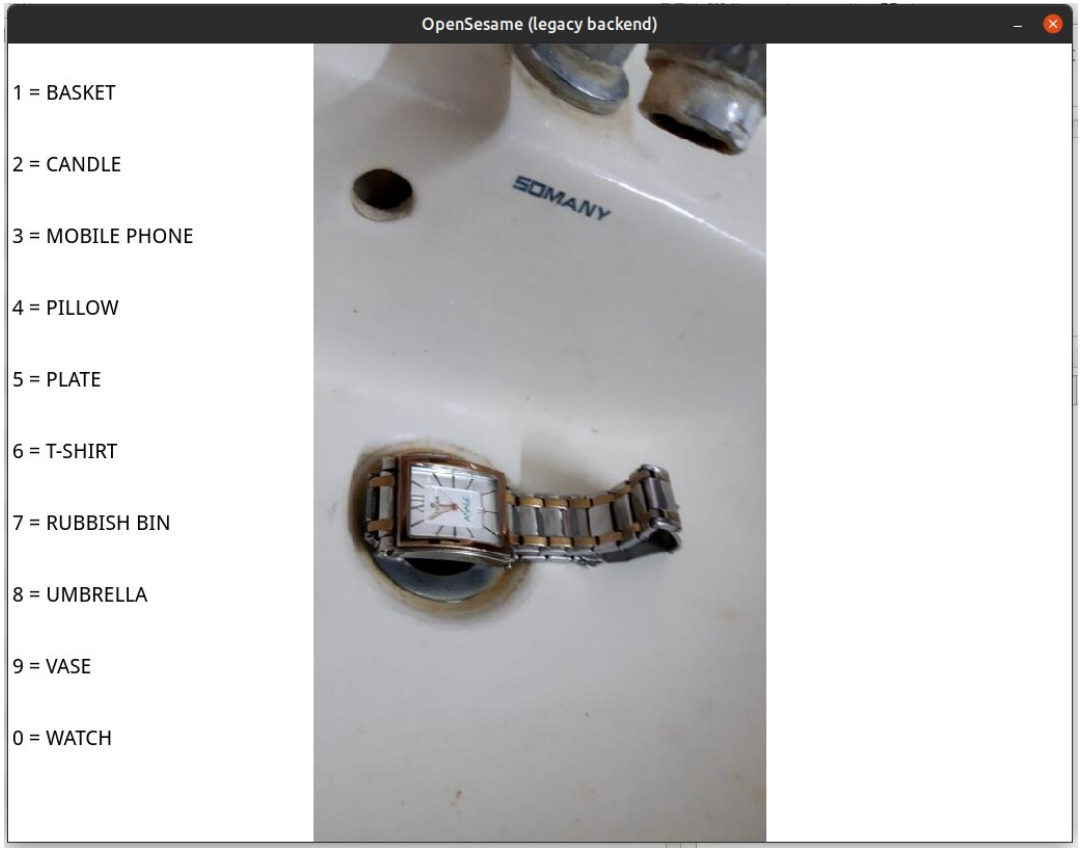
Internet objects

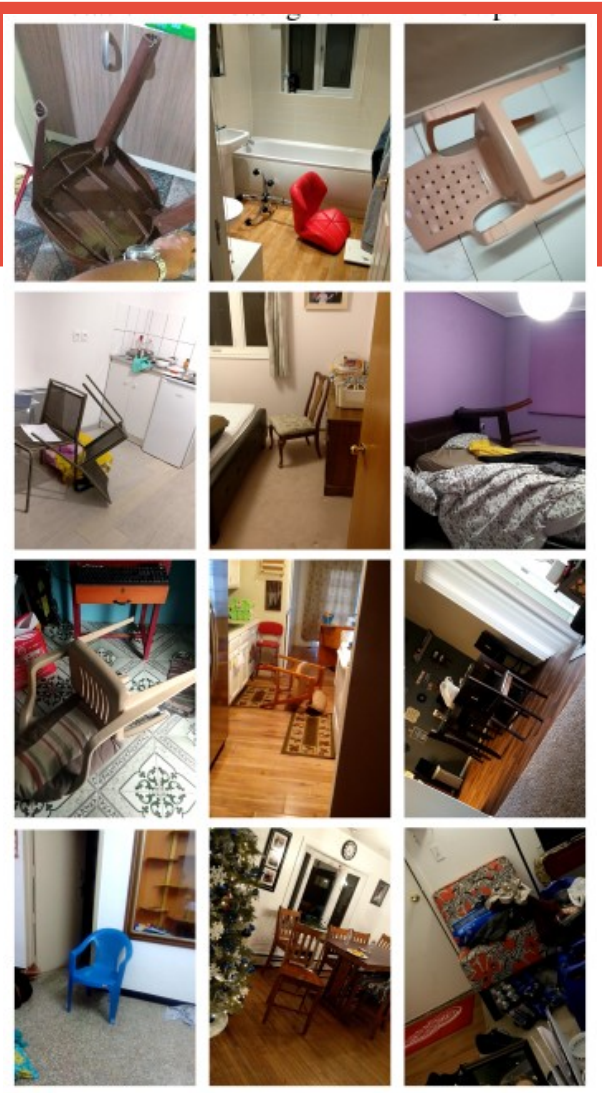
72% correct

Objects in the real world

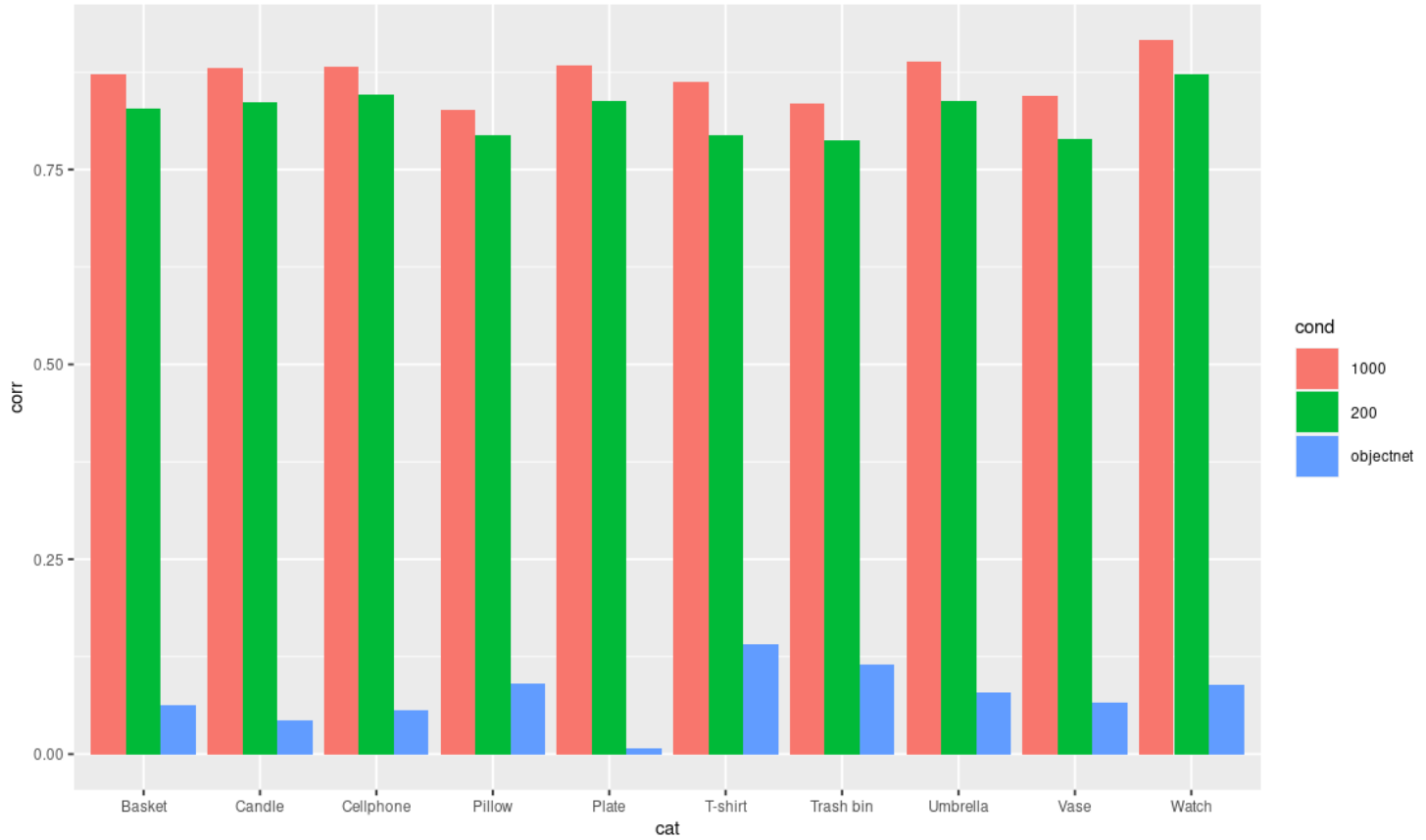
30 % correct

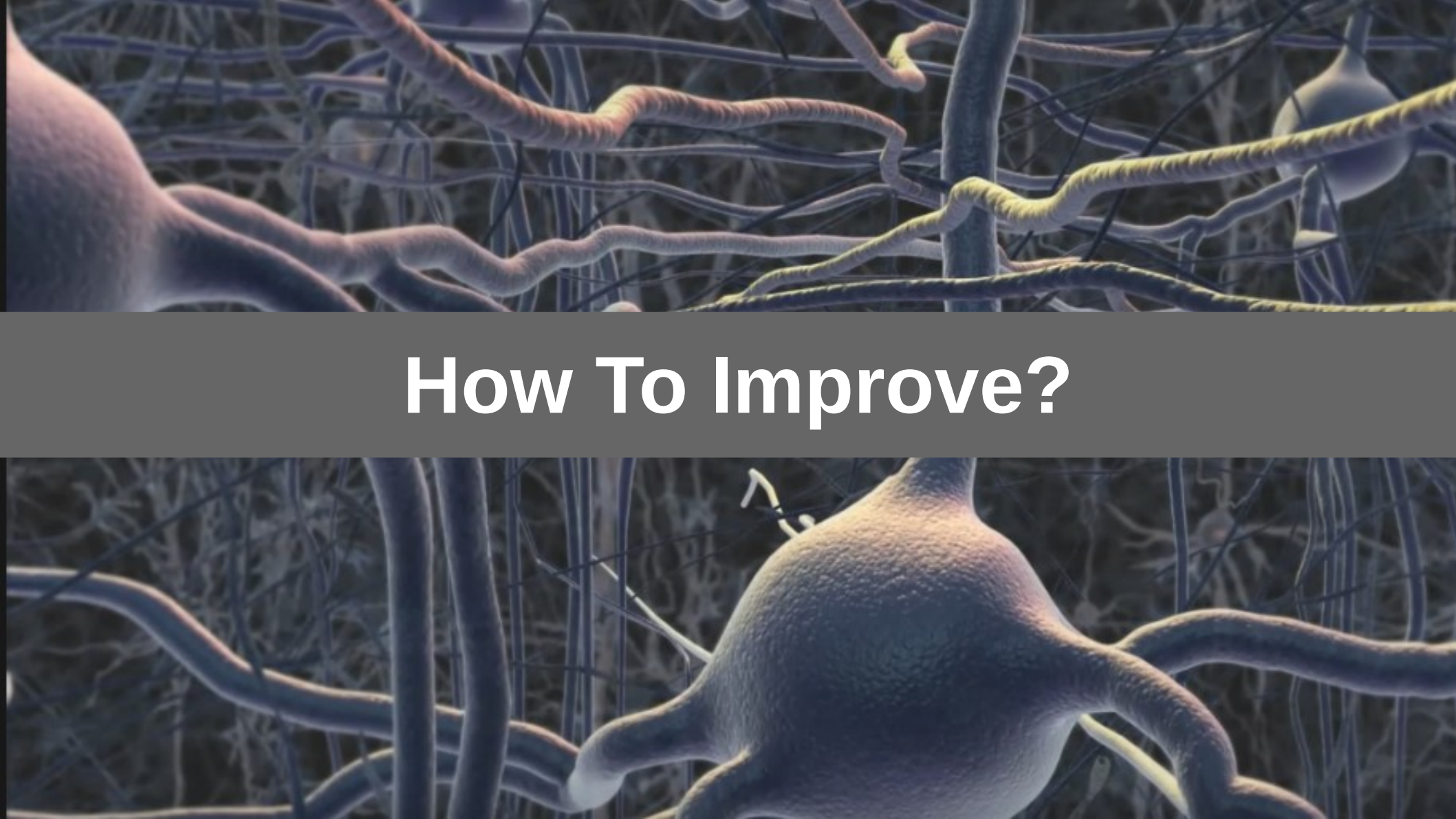
How good are people?





Results

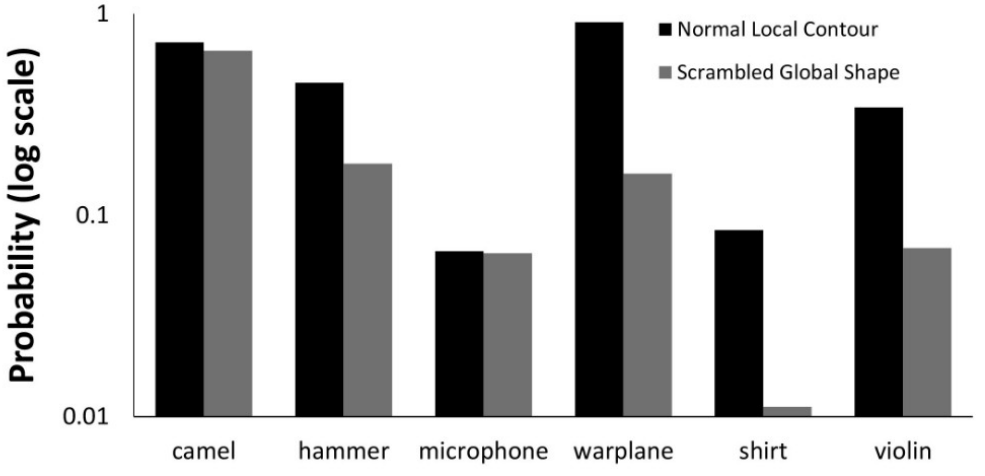

0.75
0.50
0.25
0.00
corr
Basket
Candle
Cellphone
Pillow
Plate
T-shirt
Trash bin
Umbrella
Vase
Watch
cat
cond
1000
200
objectnet



How To Improve?

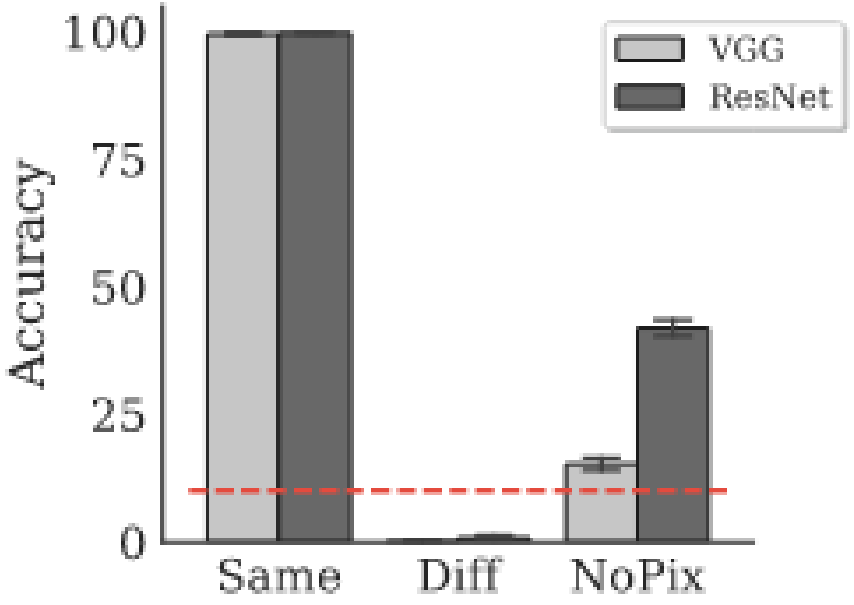
Baker et al. (2018)

Insufficiently sensitive to shape



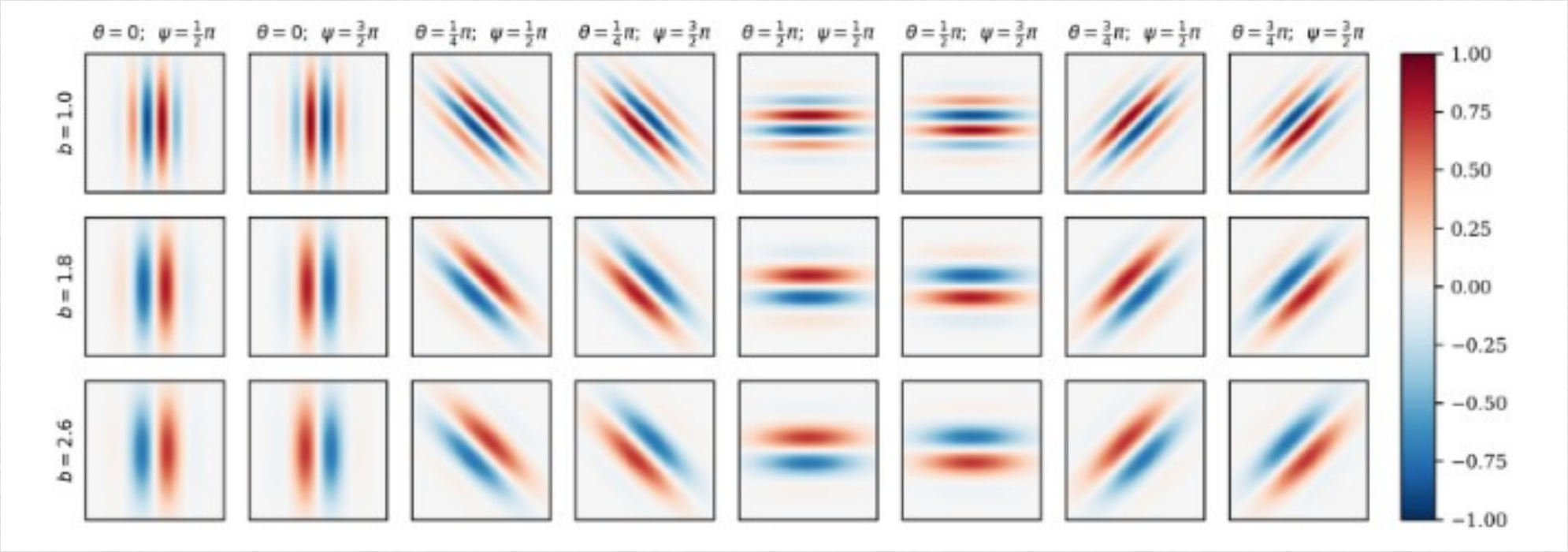
Malhorta et al. (2020)

Overly sensitive to tiny local features

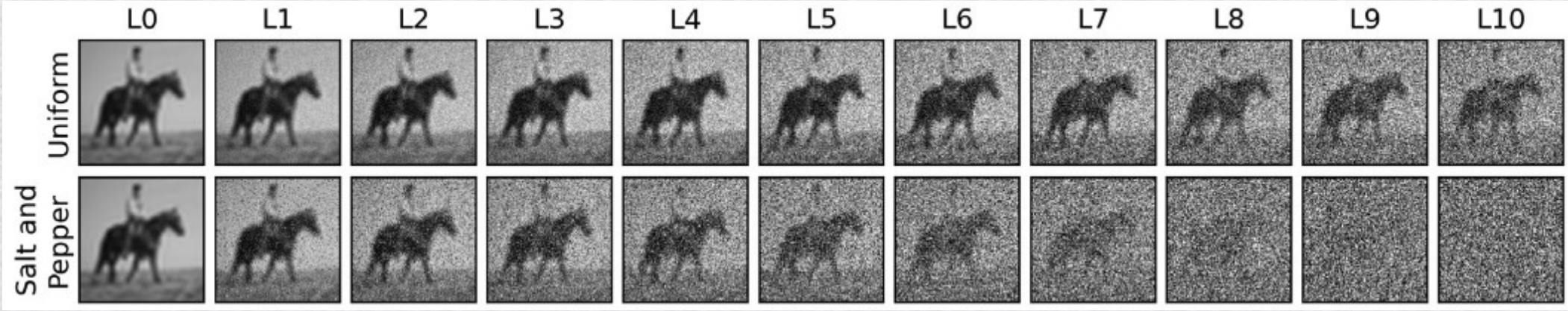


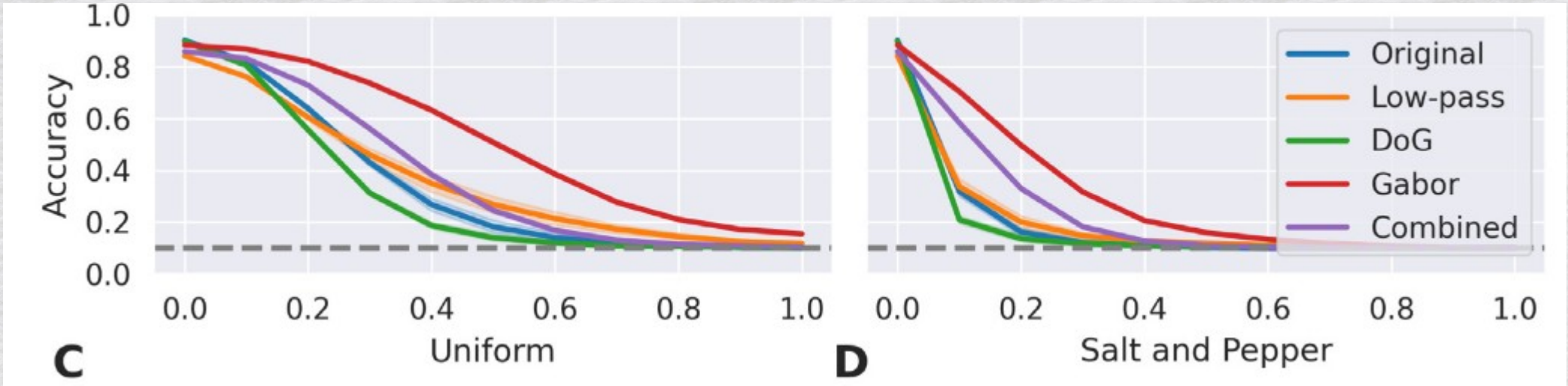
(c) Single diagnostic pixel

Evans et al. (2022)



Evans et al. (2022)





Summary