Convolutional neural networks: motivation and history

Partially based on a talk by L. Zitnick

Computer vision and “pattern recognition”

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

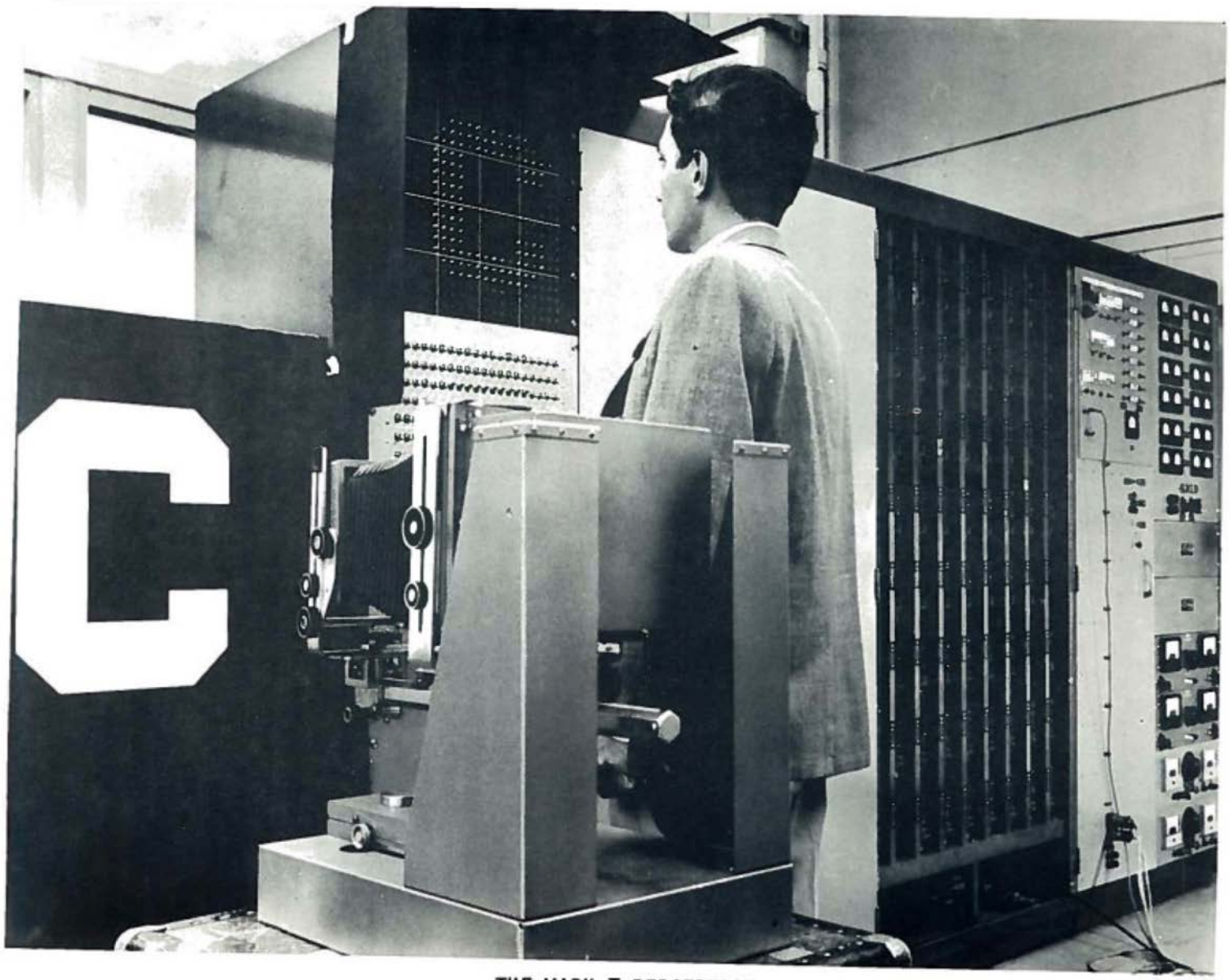
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Computer vision and “pattern recognition”

Fig. 4. Correlations in binary patterns. The center pattern, which may be considered the “unknown,” differs from each of the outside patterns (“templates”) by 9 bit positions. The effect of the correlations among the mismatch bits must thus be taken into account for correct identification. Although this is an artificially constructed example, instances of such neighborhood correlations frequently occur in practice.

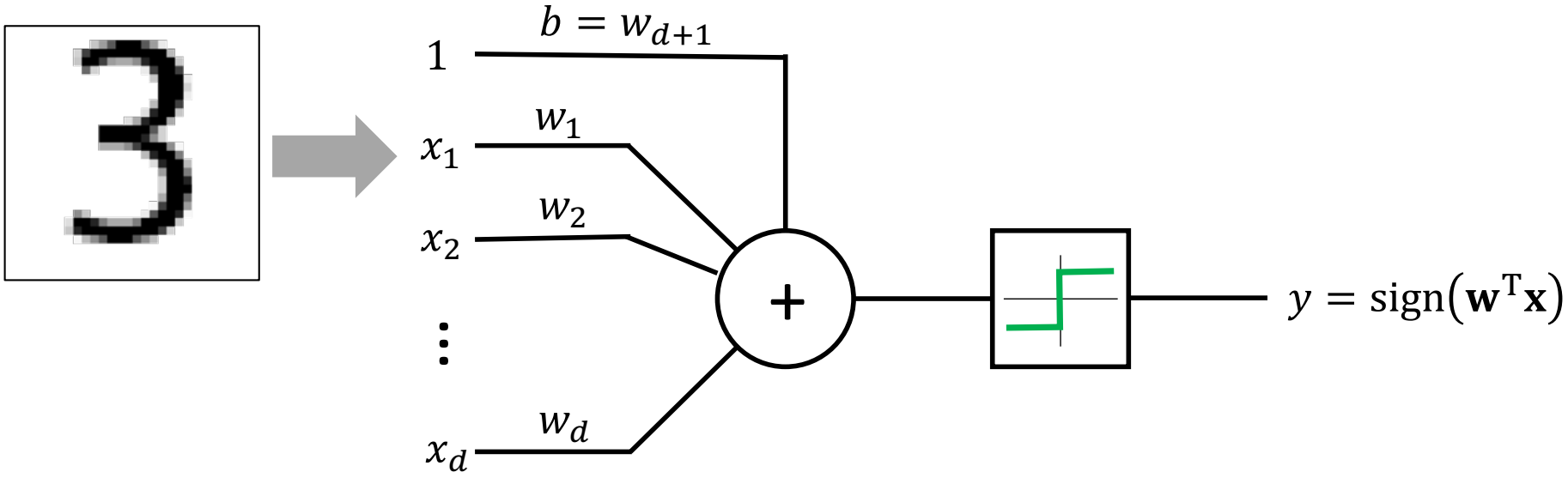
Nagy, “State of the art in pattern recognition”, 1968

Why neural networks failed in image analysis?



THE MARK I PERCEPTRON

Why neural networks failed in image analysis?



Curse of dimensionality

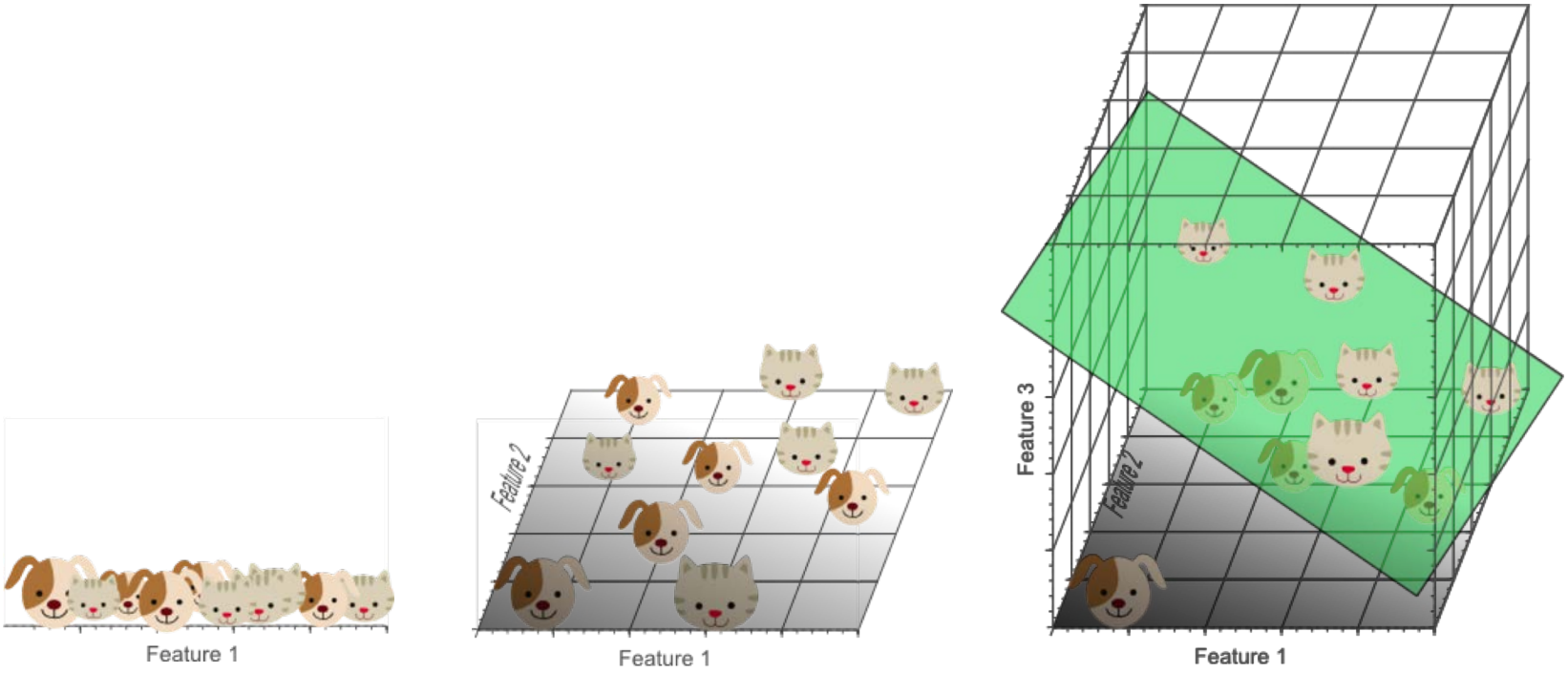
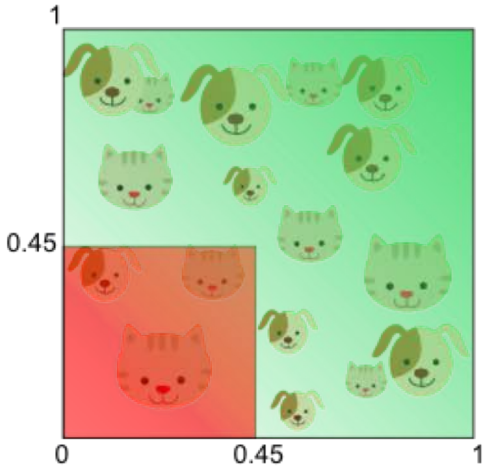


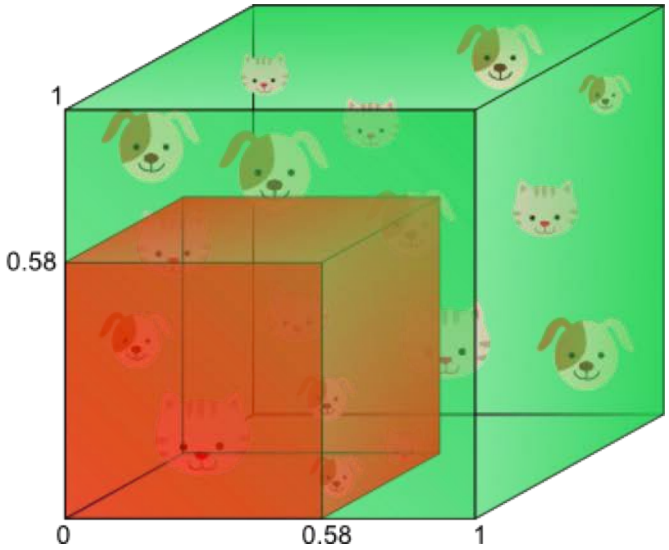
Figure: Vision Dummy

Curse of dimensionality

20%



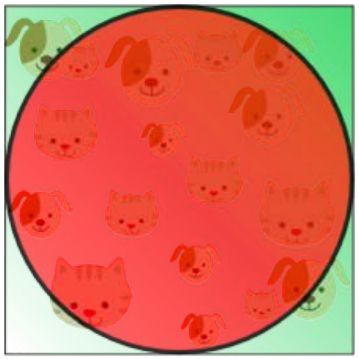
$(45\%)^2 = 20\%$

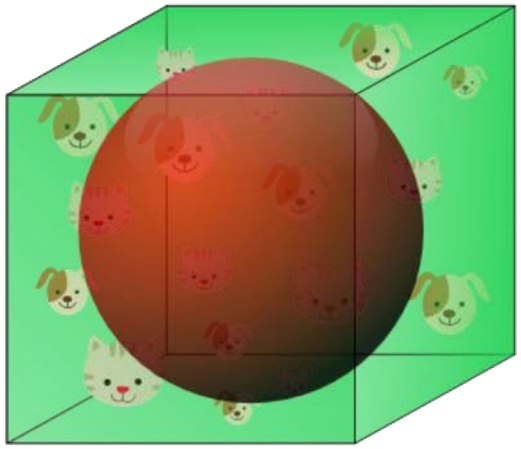


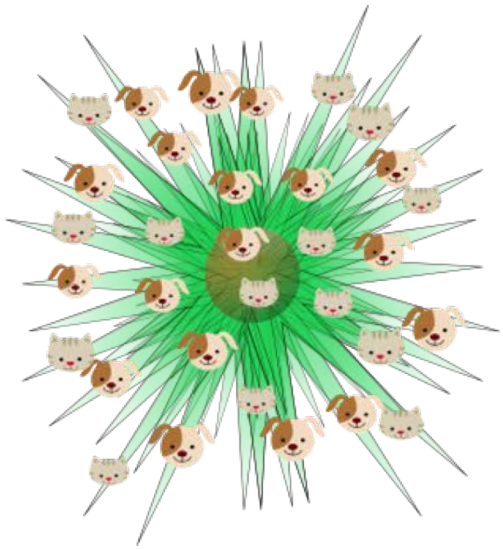
$(58\%)^3 = 20\%$

Figure: Vision Dummy

Curse of dimensionality







Volume of ball inscribed in a unit hypercube

$$V_{\text{ball}}(d) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)2^d}$$

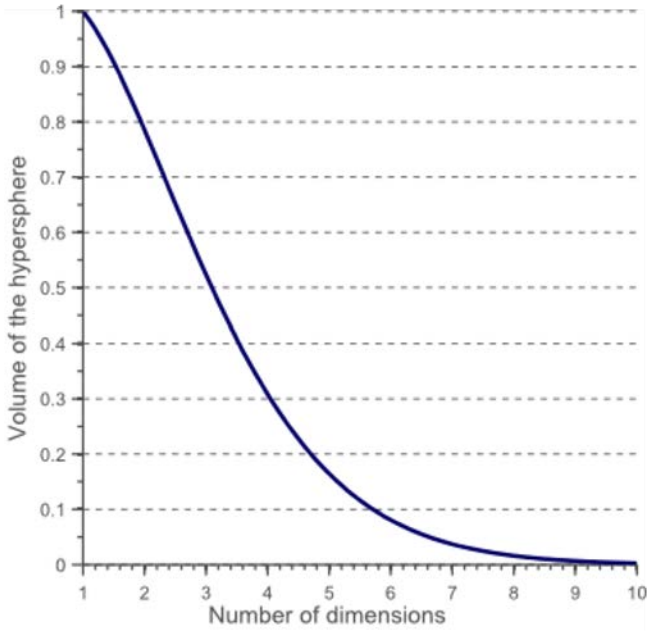
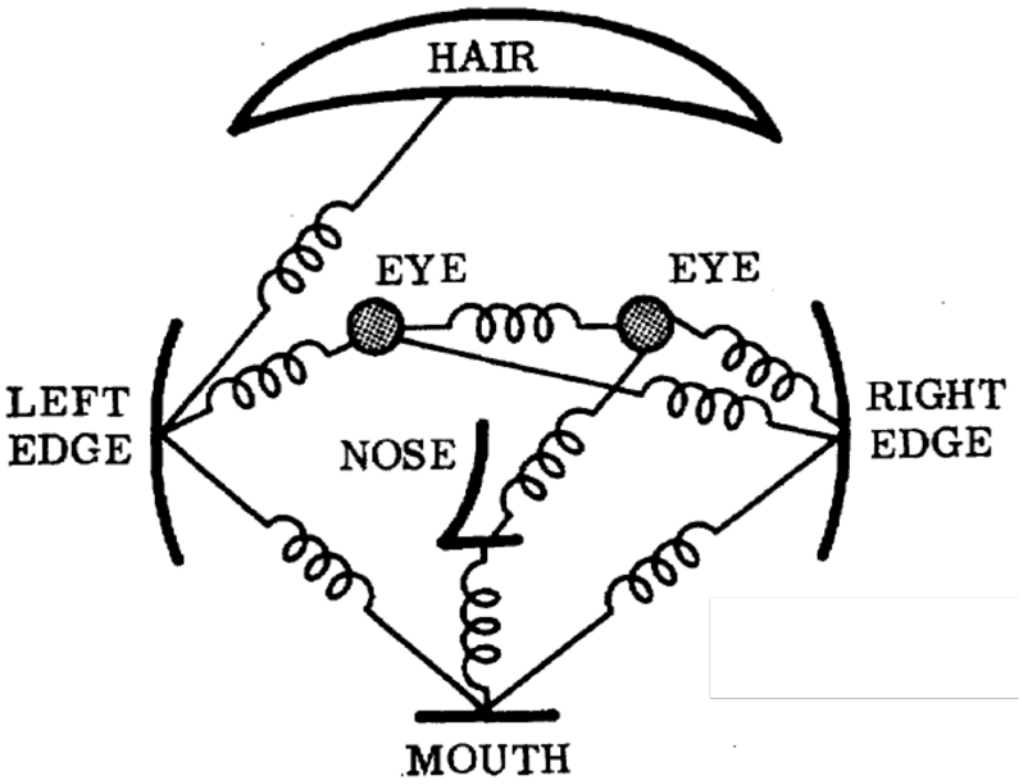


Figure: Vision Dummy

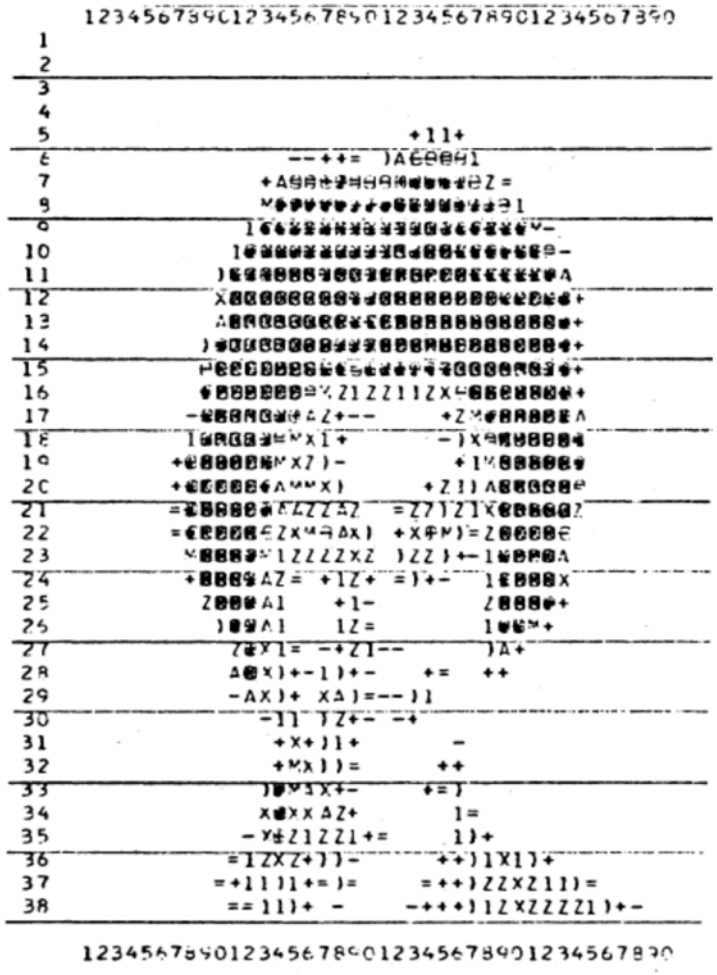
Function approximation in high-dimension

To approximate a Lipschitz function $f: \mathbb{R} \to \mathbb{R}^d$ with ϵ accuracy one needs $O\left(\epsilon^{-d}\right)$ samples

Computer vision: back to basics

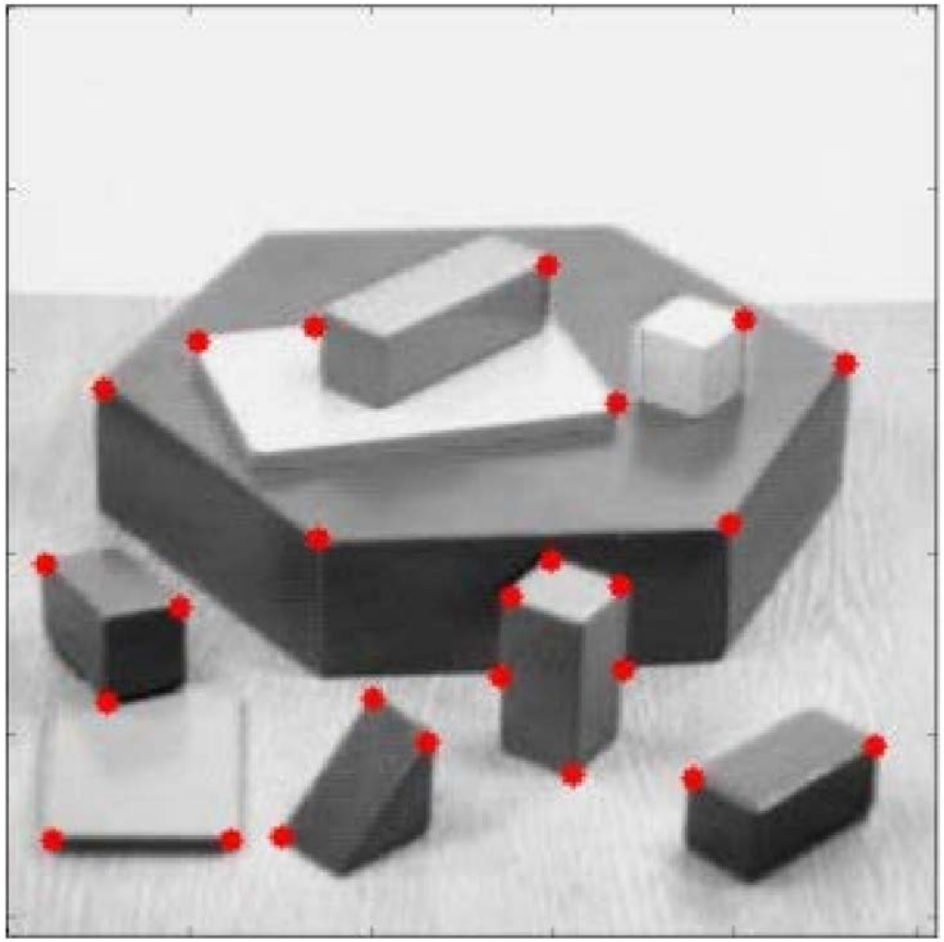


Constellation model



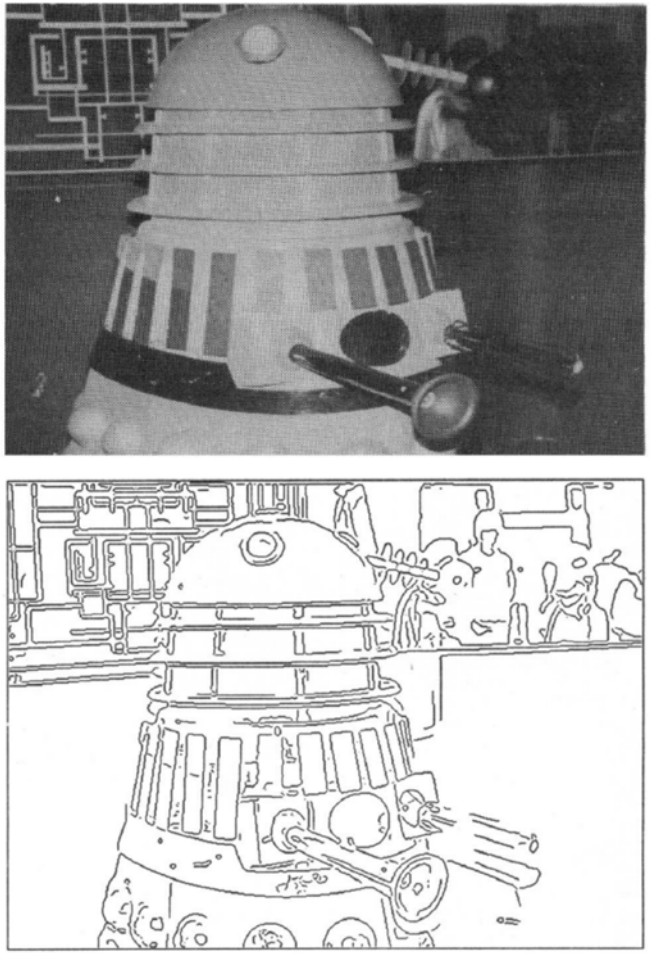
Fischler and Elschlager 1973

Computer vision: back to basics



Corners

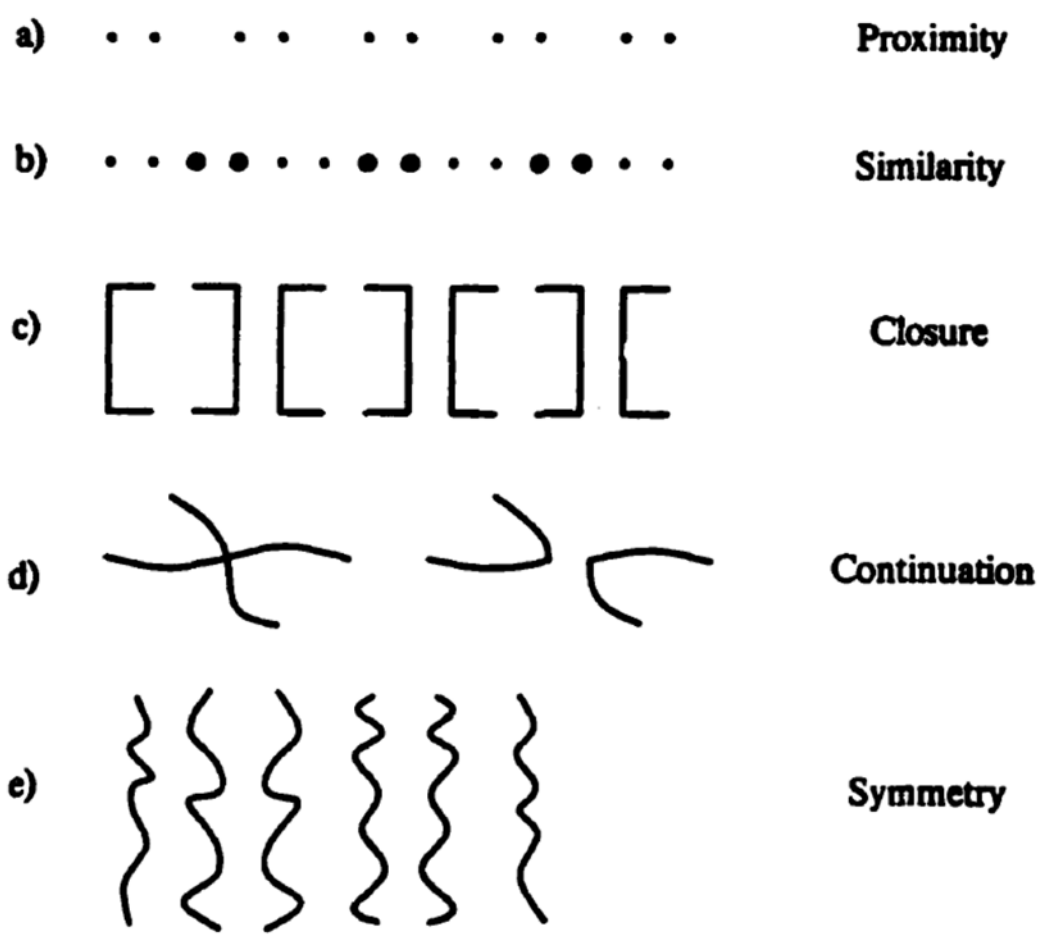
Harris, Stephens 1980



Edge

Canny 1986

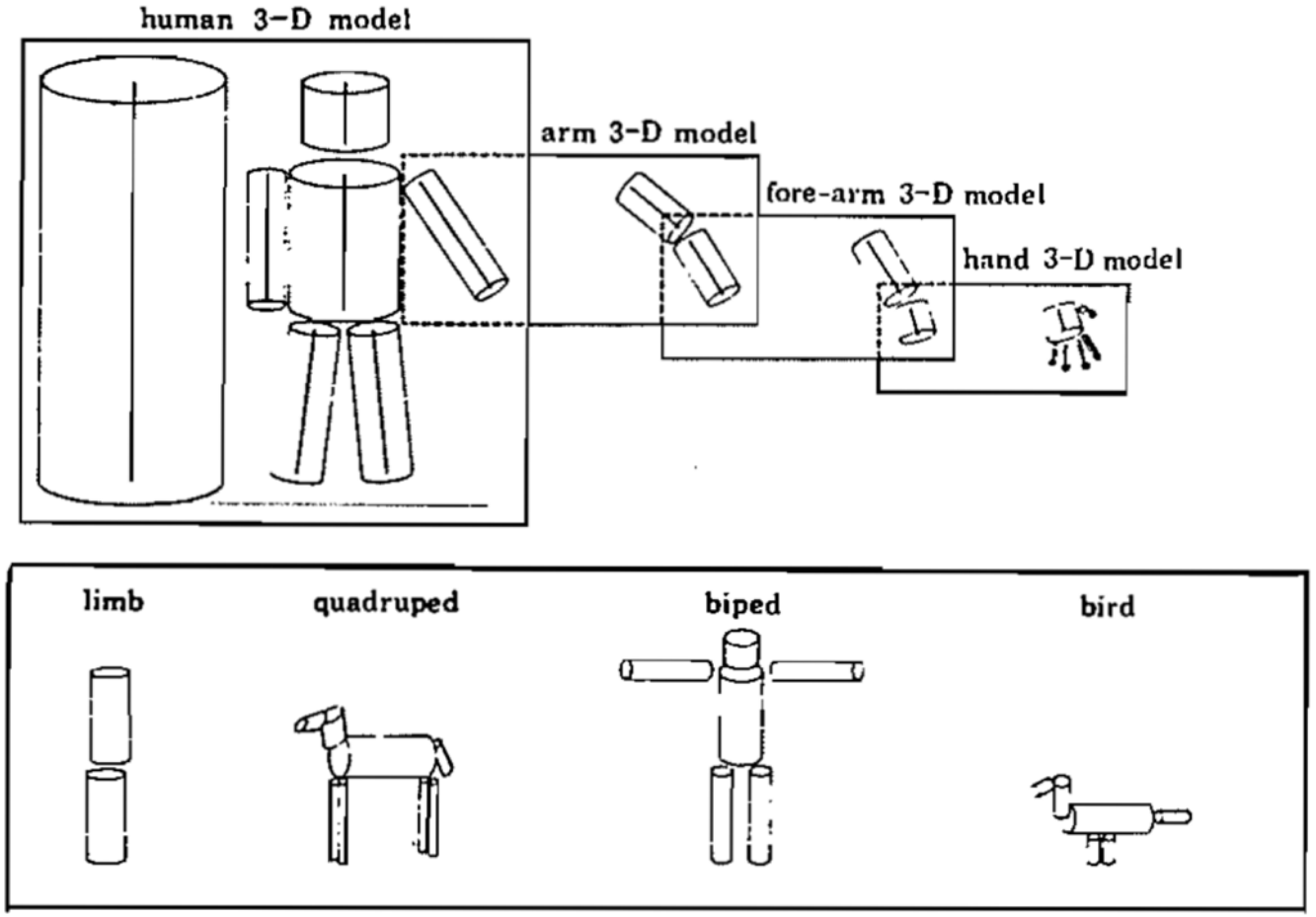
Computer vision: back to basics



Perceptual organization

Lowe 1984

Computer vision: back to basics



Perceptual organization

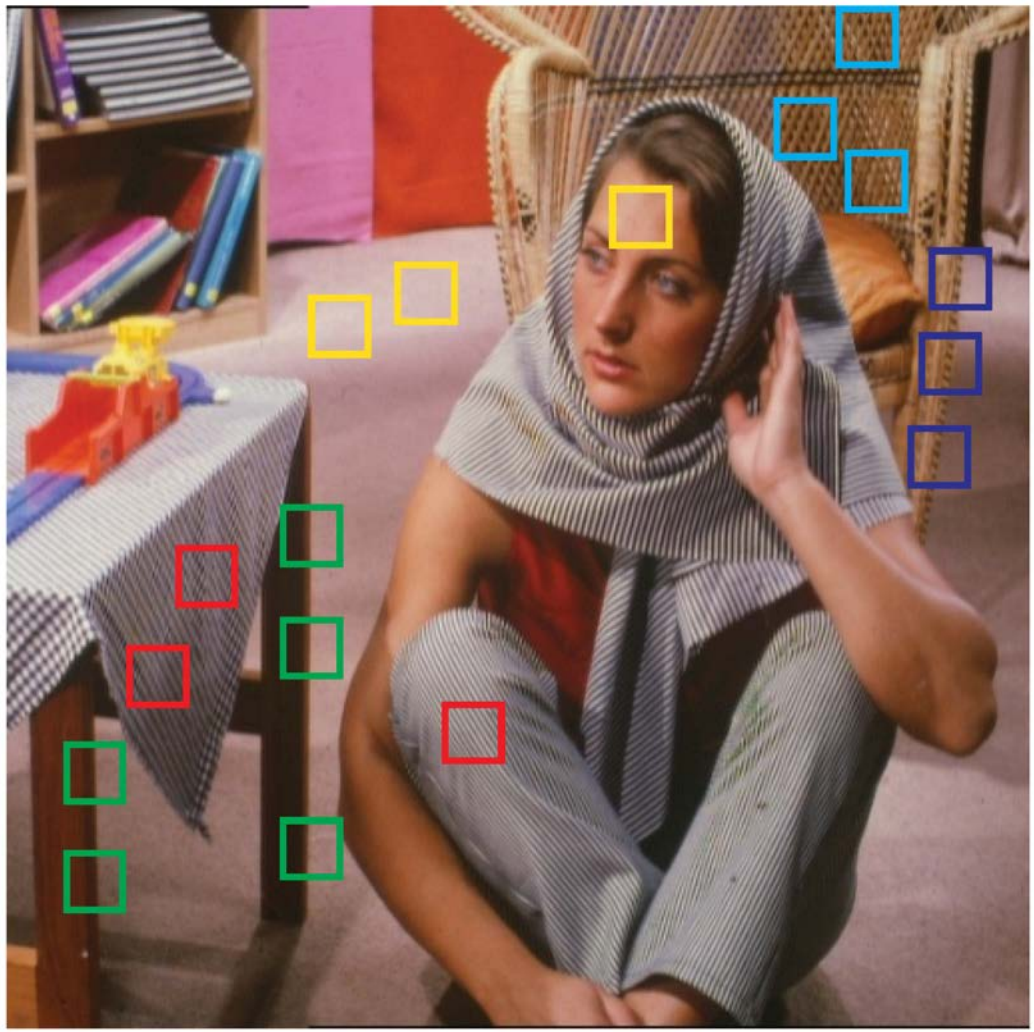
Pentland 1986

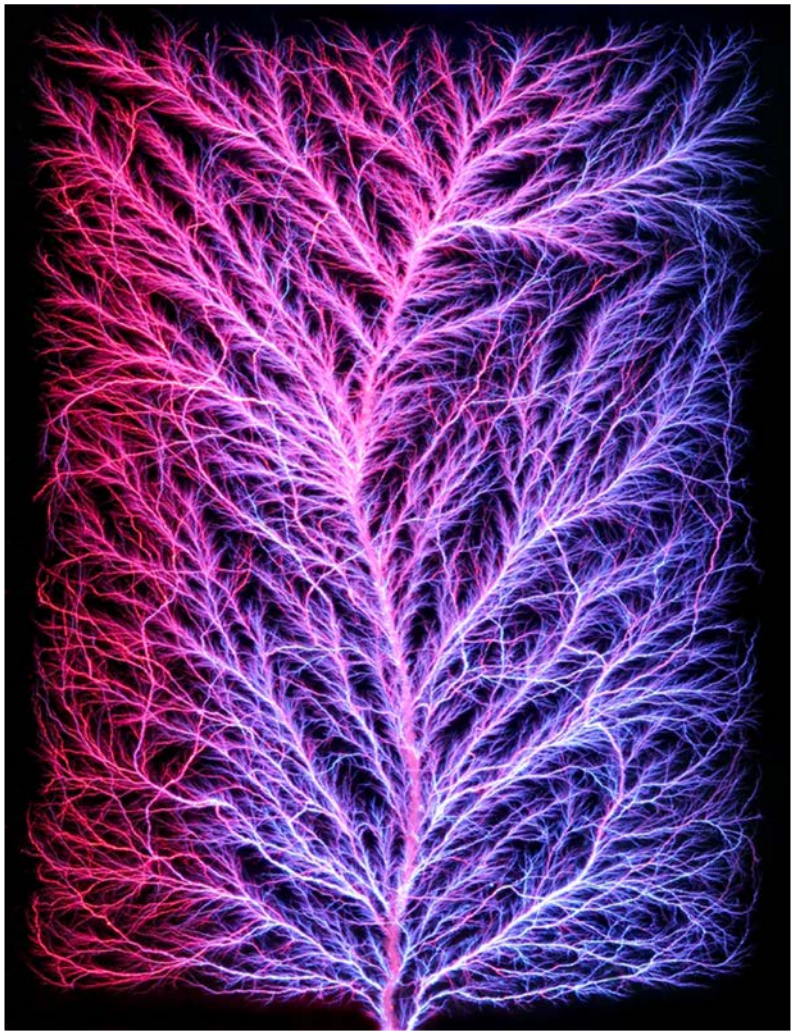
Dealing with the curse of dimensionality

To approximate a Lipschitz function $f: \mathbb{R} \to \mathbb{R}^d$ with ϵ accuracy one needs $O\left(\epsilon^{-d}\right)$ samples

Need priors about data!

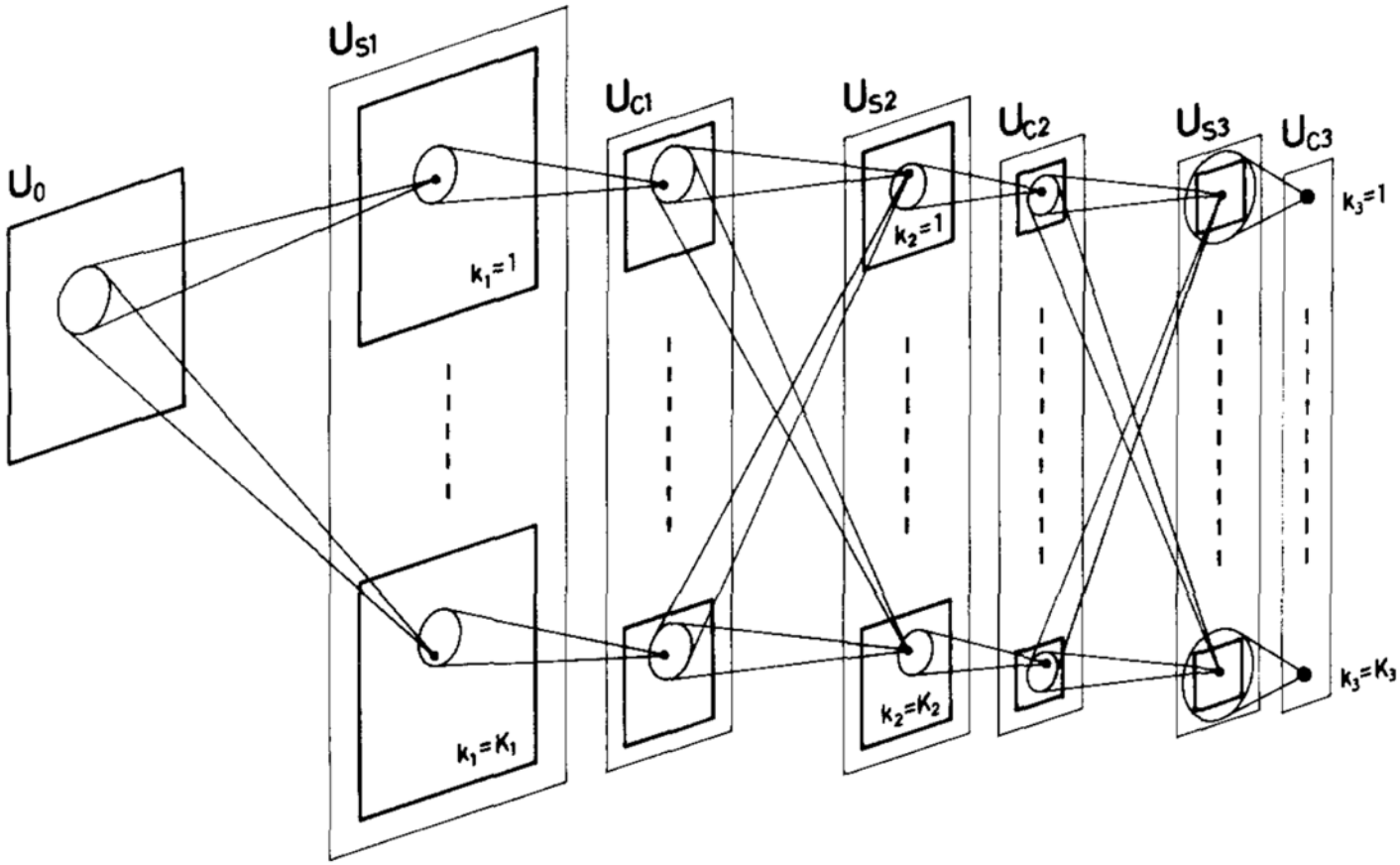
Self-similarity





Neocognitron

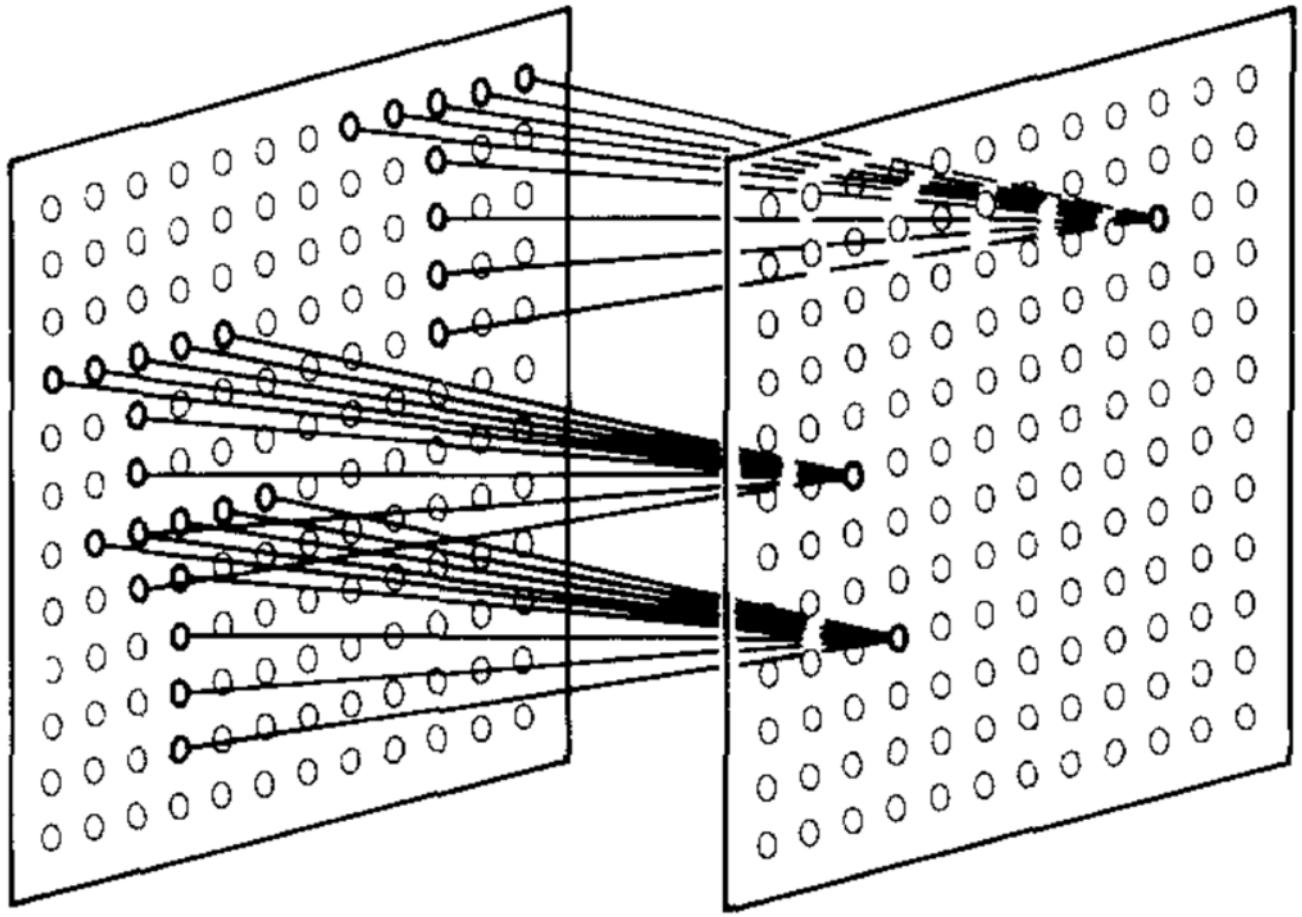
Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position



K. Fukushima

Fukushima 1980

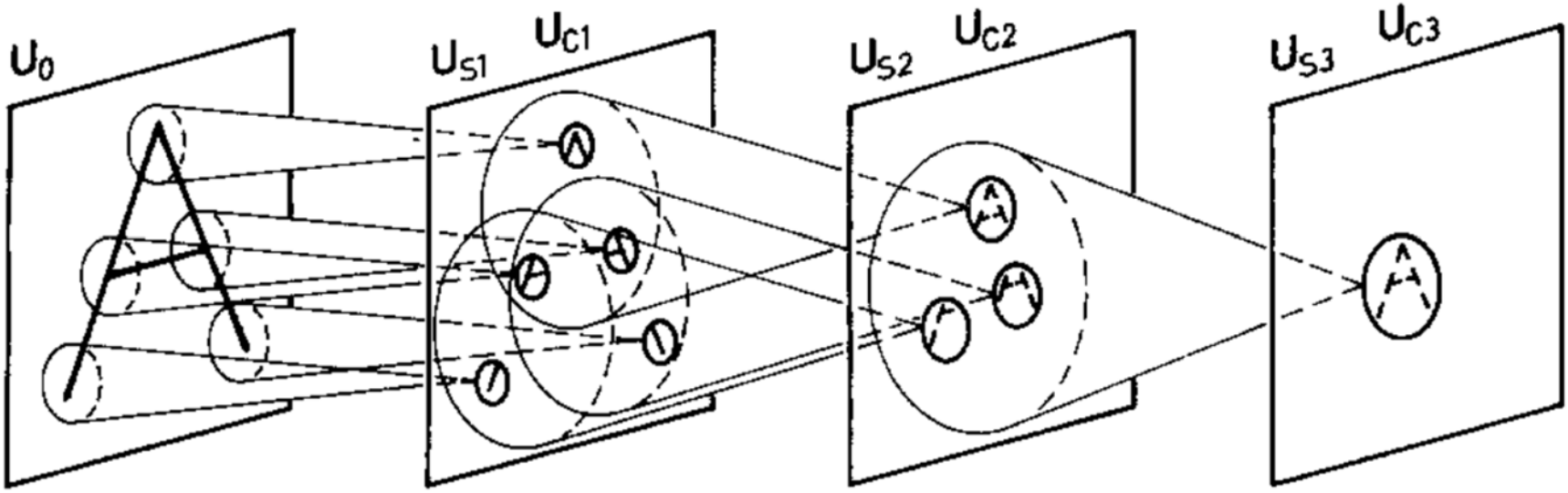
Neurocognitron: key features



Local connectivity

Fukushima 1980

Neurocognitron: key features



"Pooling"

Fukushima 1980

Neurocognitron: key features

"The response of [Perceptron and similar models…] was severely affected by the shift in position and/or by the distortion in shape of the input patterns. Hence, their ability for pattern recognition was not so high."

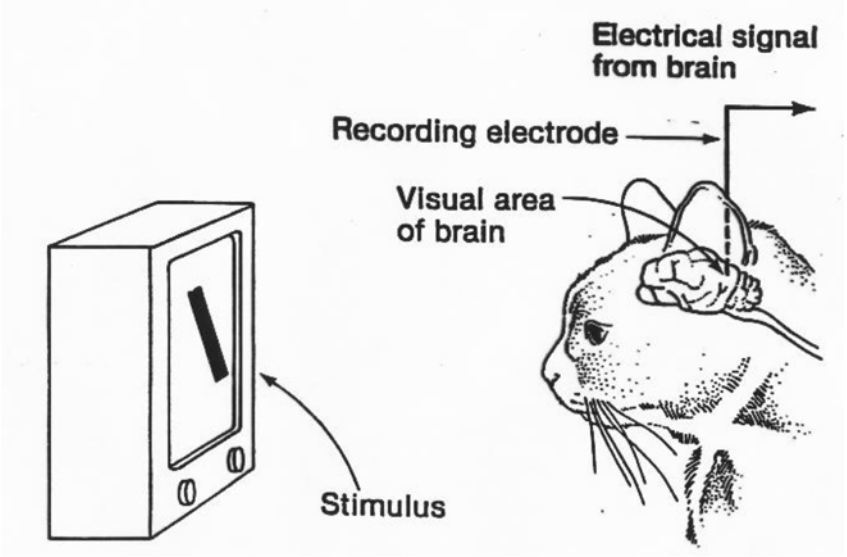
"The network is self-organized by "learning without a teacher", and acquires an ability to recognize stimulus patterns based on the geometrical similarity (Gestalt) of their shapes without affected by their positions."

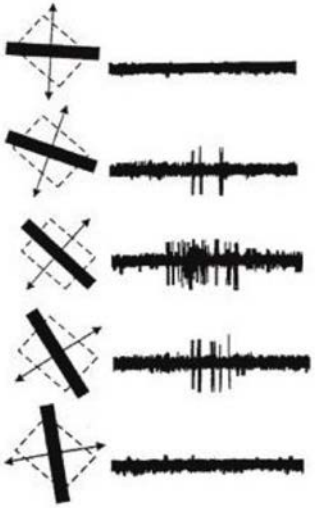
$$\varphi[x] = \begin{cases} x & x \geqq 0 \\ 0 & x < 0 \end{cases}$$

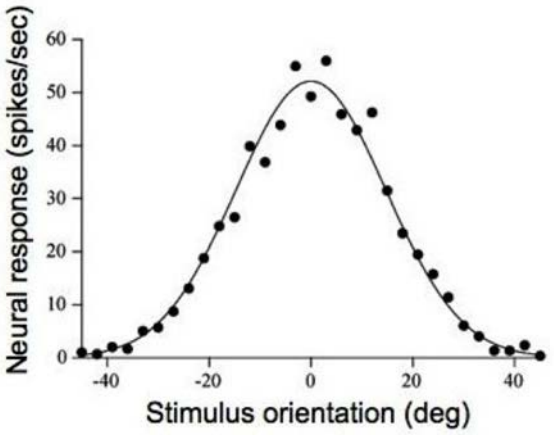
ReLU in 1980!

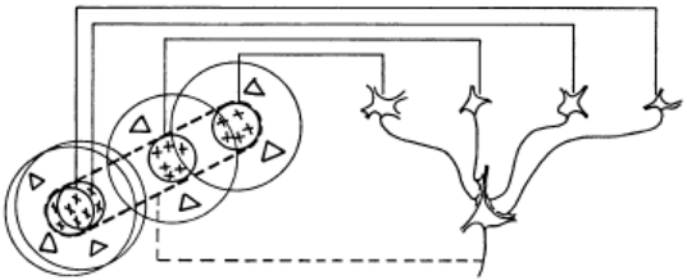
"The structure of this network has been suggested by that of the visual nervous system of the vertebrate"

Biological inspiration

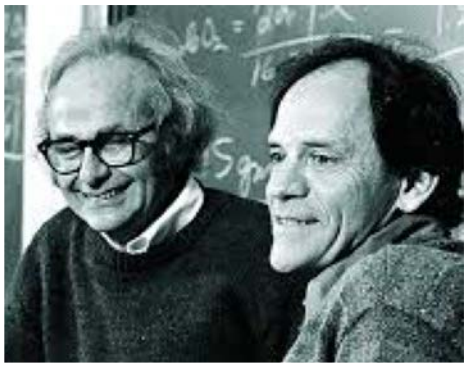








Text-fig. 19. Possible scheme for explaining the organization of simple receptive fields. A large number of lateral geniculate cells, of which four are illustrated in the upper right in the figure, have receptive fields with 'on' centres arranged along a straight line on the retina. All of these project upon a single cortical cell, and the synapses are supposed to be excitatory. The receptive field of the cortical cell will then have an elongated 'on' centre indicated by the interrupted lines in the receptive-field diagram to the left of the figure.

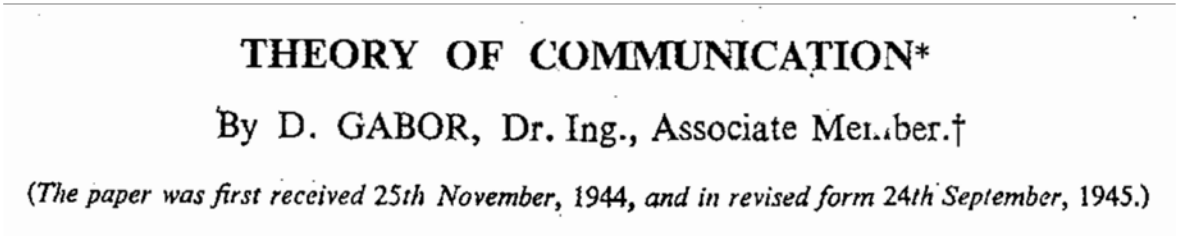


Hubel & Wiesel

Nobel Prize
1981

Hubel, Wiesel 1962, 1965

Signal processing realization: Gabor transform



THEORY OF COMMUNICATION*

By D. GABOR, Dr. Ing., Associate Member.†

(The paper was first received 25th November, 1944, and in revised form 24th September, 1945.)

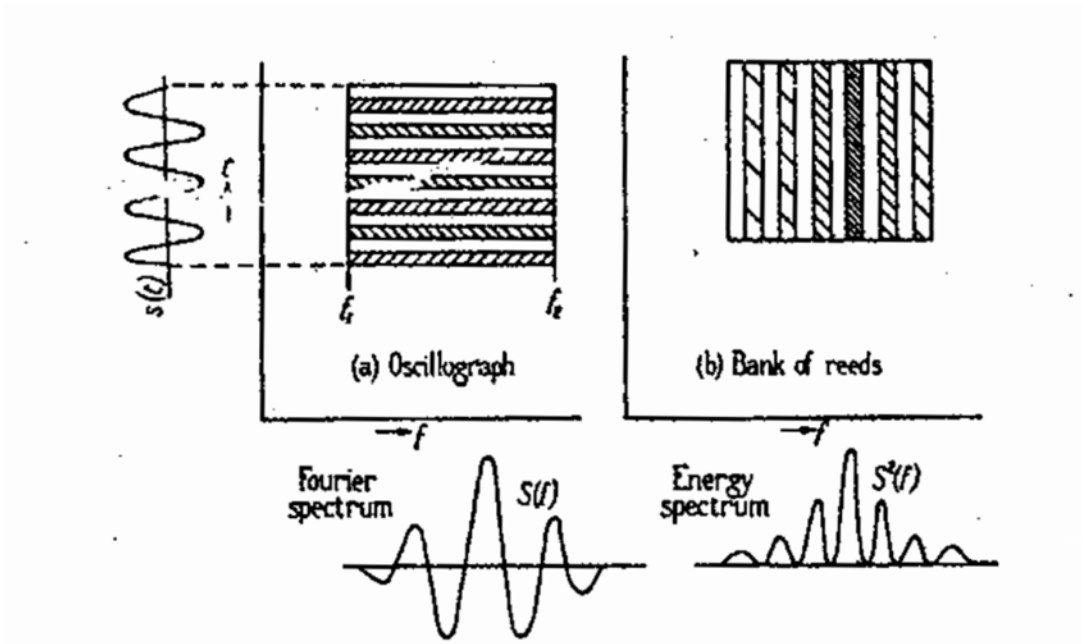
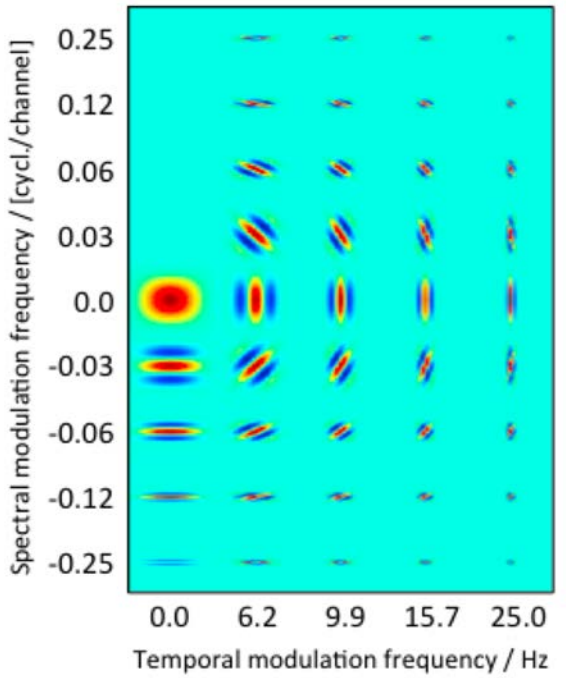
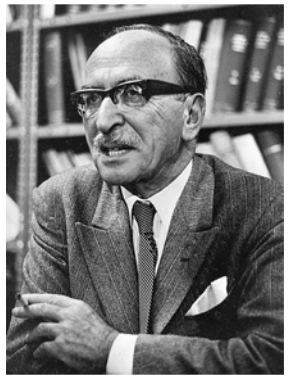


Fig. 1.4.—Time/frequency diagram of the response of physical instruments to a finite sine wave.





D. Gabor

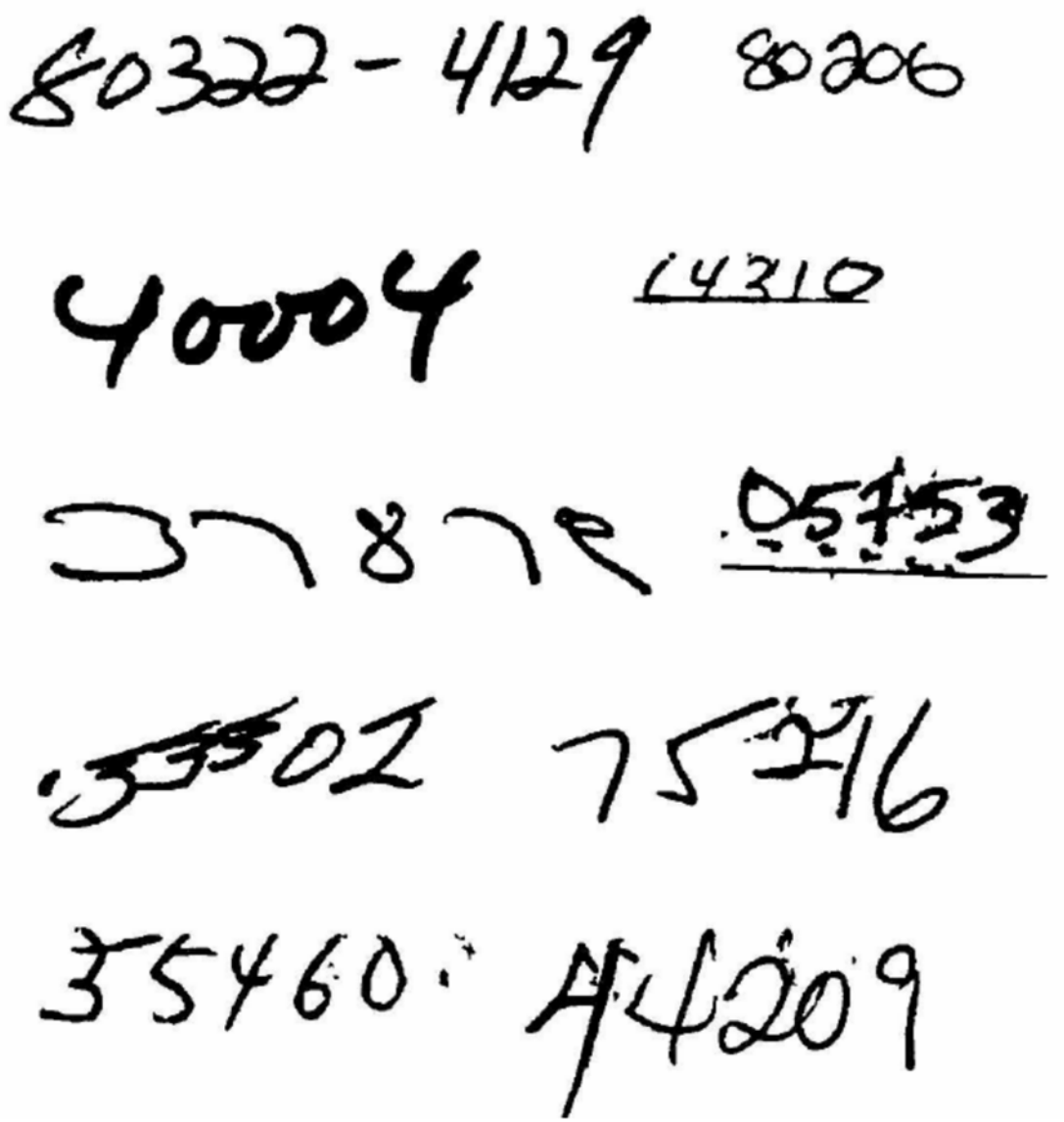
Nobel Prize 1971

Gabor 1946

Key insights

- Neurons arranged into planes preserving image pixel arrangement
- Pooling (“complex cells”)
- Local connectivity (“receptive field”)
- Weight sharing
- Hierarchical organization

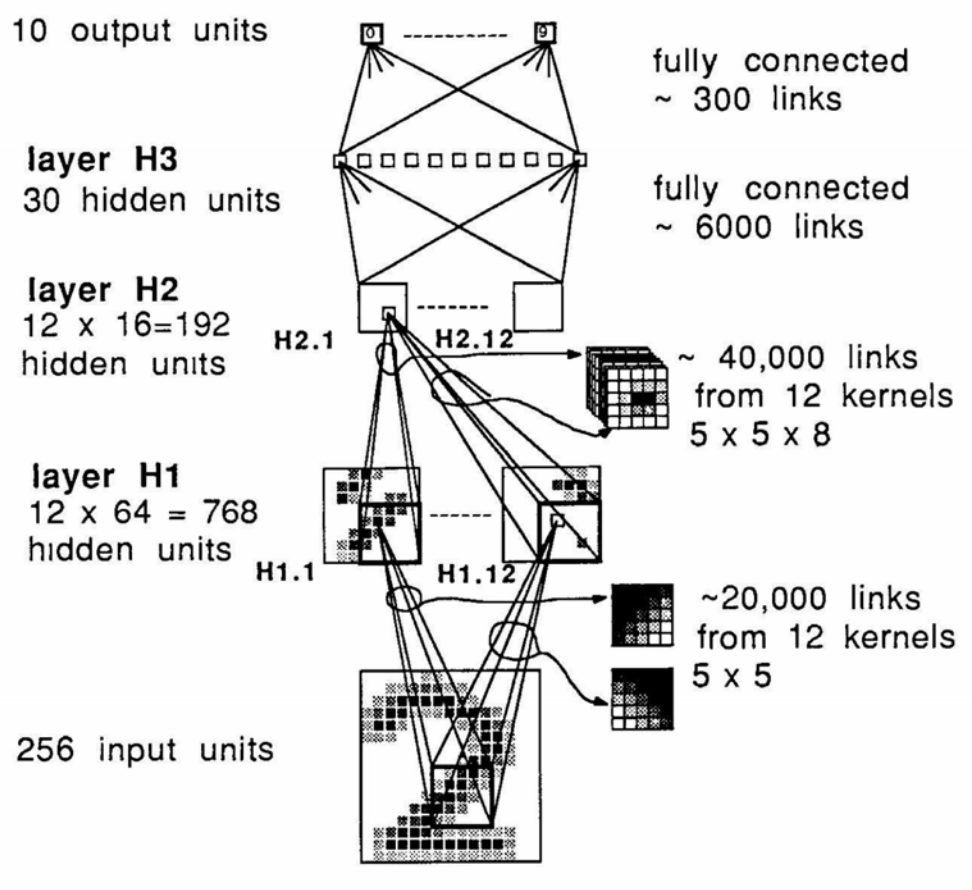
Handwritten digit recognition



Y. LeCun

LeCun et al. 1989

Convolutional neural networks



Backpropagation Applied to Handwritten Zip Code Recognition

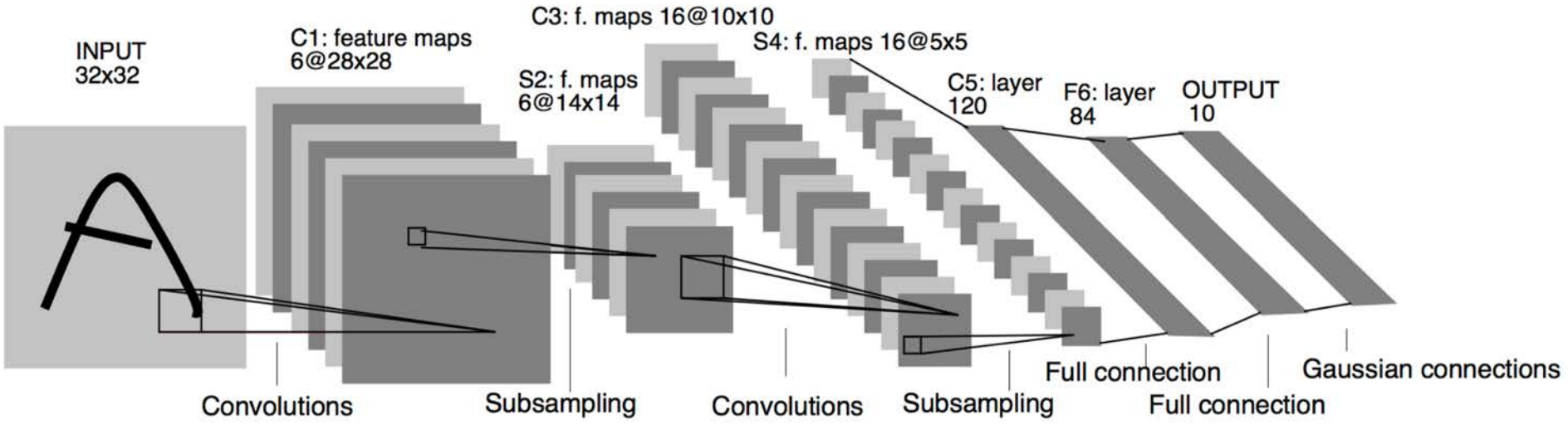
Y. LeCun
B. Boser
J. S. Denker
D. Henderson
R. E. Howard
W. Hubbard
L. D. Jackel
AT&T Bell Laboratories Holmdel, NJ 07733 USA

LeCun et al. 1989

Convolutional neural networks

Gradient-Based Learning Applied to Document Recognition

YANN LECUN, MEMBER, IEEE, LÉON BOTTOU, YOSHUA BENGIO, AND PATRICK HAFFNER



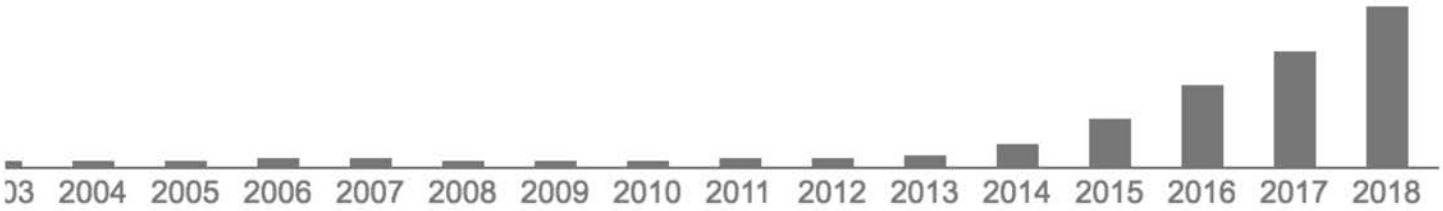
LeCun et al. 1998

Backpropagation applied to handwritten zip code recognition

Authors: Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, Lawrence D Jackel

Publication date: 1989/12

Total citations: Cited by 4208

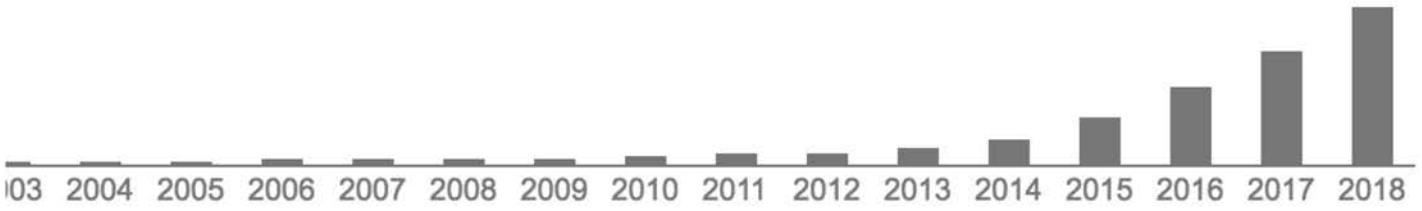


Gradient-based learning applied to document recognition

Authors: Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner

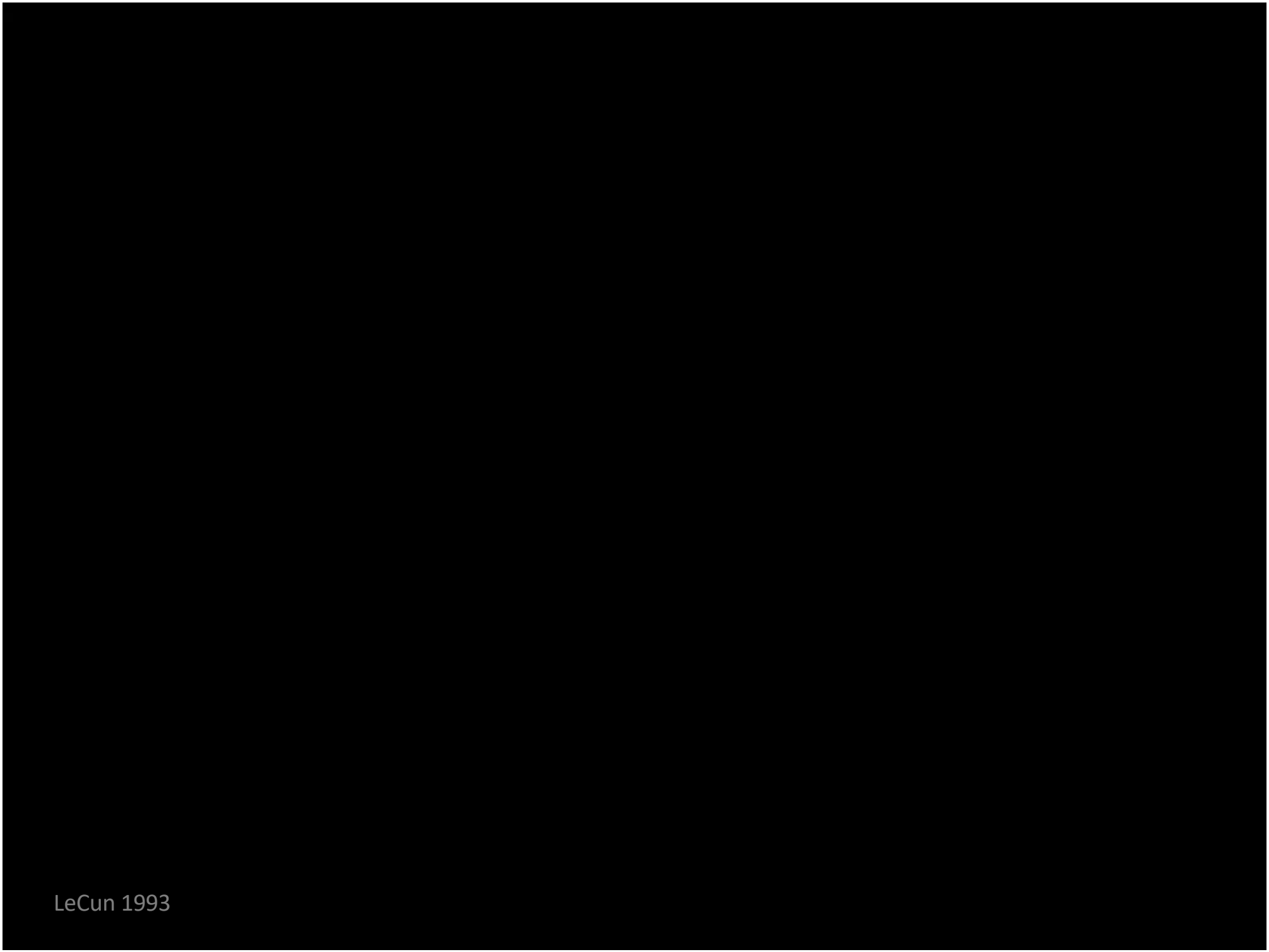
Publication date: 1998/11

Total citations: Cited by 16339



Key insights

- Trained with backpropagation
- Convolutional filters
- 5 layers
- Average pooling
- Sparse connections (to reduce complexity)
- tanh activation
- Dataset: MNIST (60K training images)



LeCun 1993

What's Hidden in the Hidden Layers?

The contents can be easy to find with a geometrical problem, but the hidden layers have yet to give up all their secrets

David S. Touretzky and Dean A. Pomerleau

tions, we fed the network road images taken under a wide variety of viewing angles and lighting conditions. It would be impractical to try to collect thousands of real road images for such a data set. Instead, we developed a synthetic road-image generator that can create as many training examples as we need.

To train the network, 1200 simulated road images are presented 40 times each, while the weights are adjusted using the back-propagation learning algorithm. This takes about 30 minutes on Carnegie Mellon's Warp systolic-array supercomputer. (This machine was designed at Carnegie Mellon and is built by General Electric. It has a peak rate of 100 million floating-point operations per second and can compute weight adjustments for back-propagation networks at a rate of 20 million connections per second.)

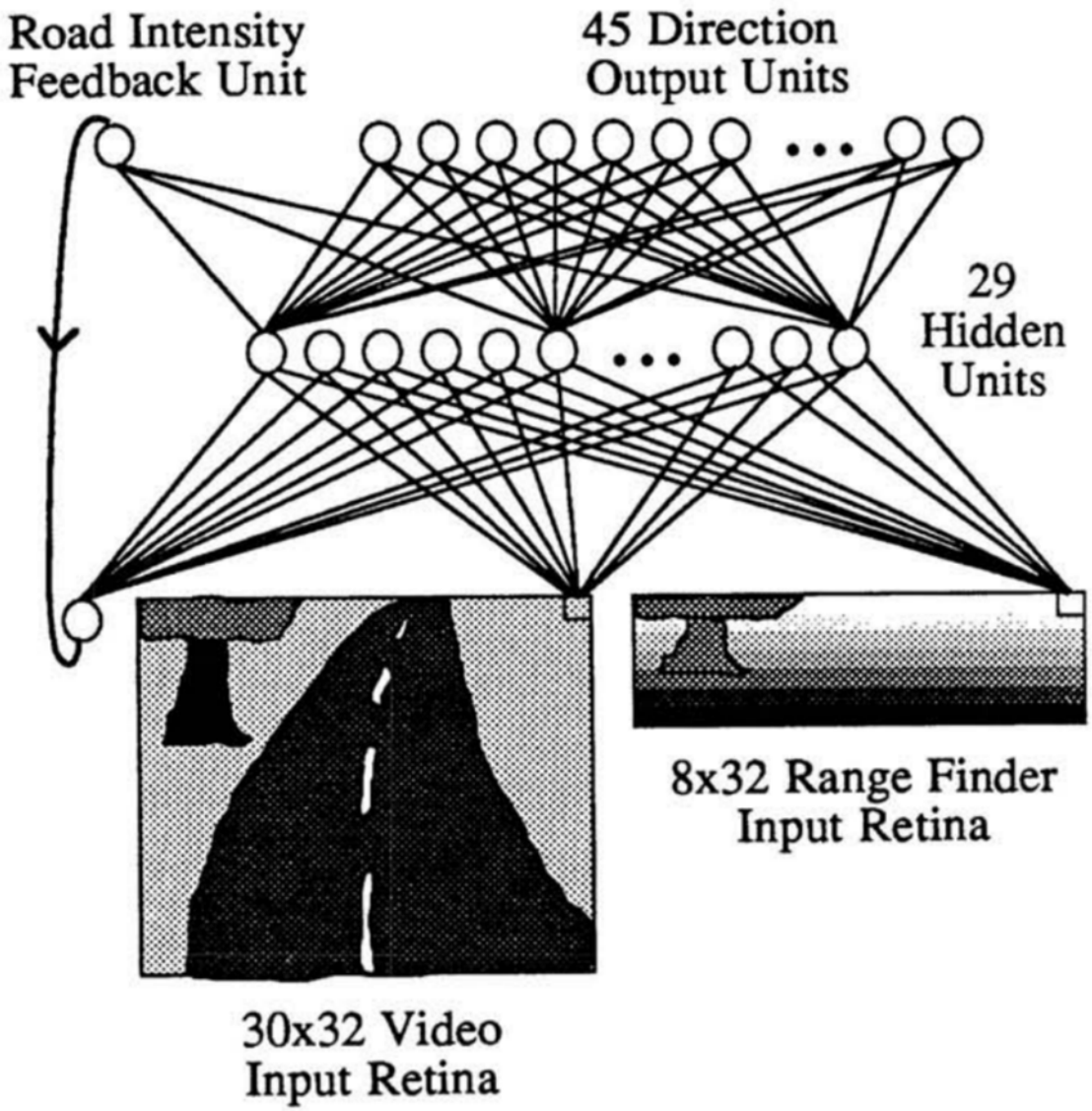
Once it is trained, ALVINN can accurately drive the NAVLAB vehicle at about 3½ miles per hour along a path through a wooded area adjoining the Carnegie Mellon campus, under a variety of weather and lighting conditions. This speed is nearly twice as fast as that achieved by non-neural-network algorithms running on the same vehicle. Part of the reason for this is that the forward pass of a back-propagation network can be computed quickly. It takes about 200 milliseconds on the Sun-3/160 workstation installed on the NAVLAB.

The hidden-layer representations ALVINN develops are interesting. When trained on roads of a fixed width, the network chooses a representation in which hidden units act as detectors for complete roads at various positions and orientations. When trained on roads of variable

continued

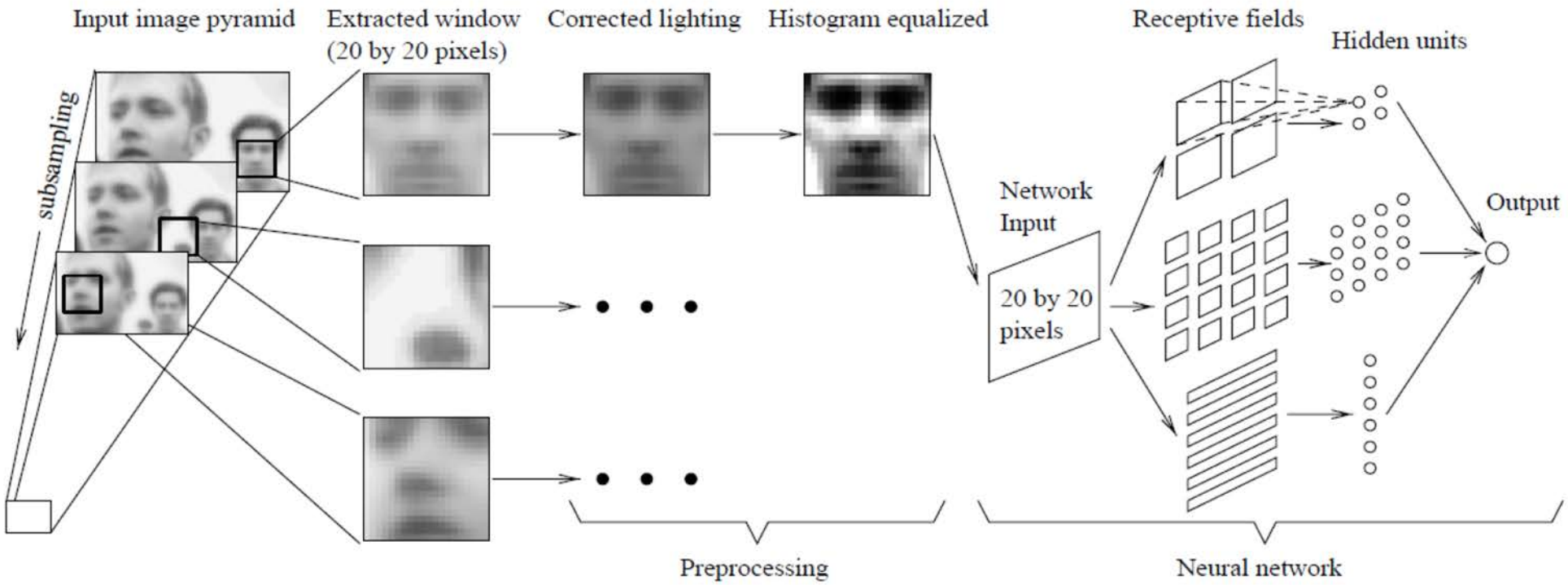
Photo 1: *The NAVLAB autonomous navigation test-bed vehicle and the road used for trial runs.*

Beyond digits: autonomous driving



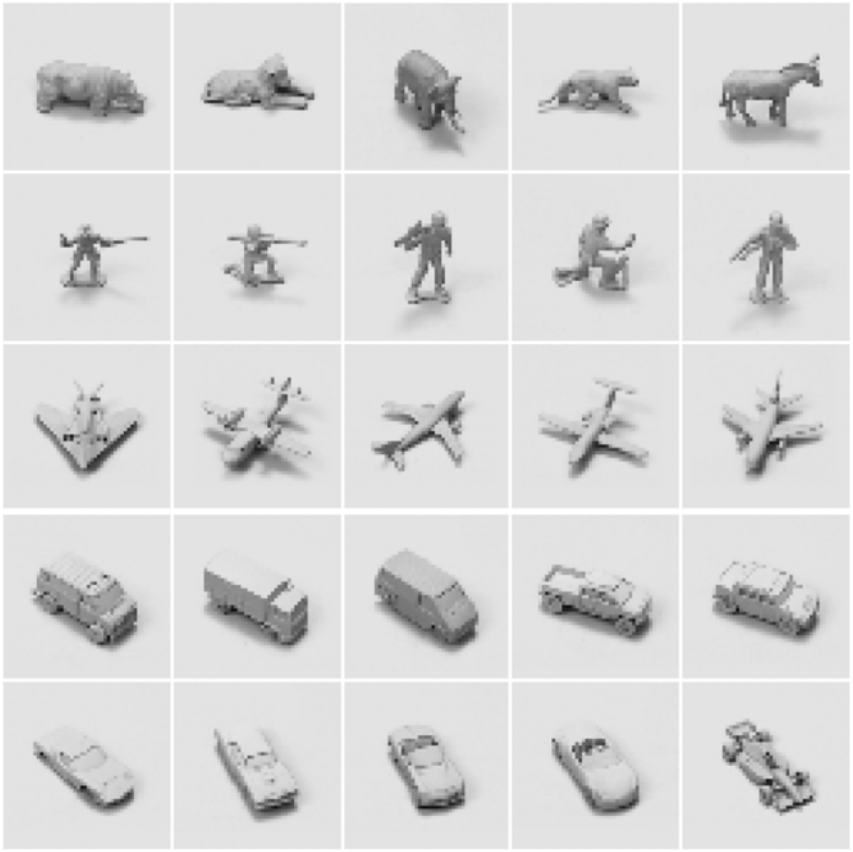
Pomerleau 1995

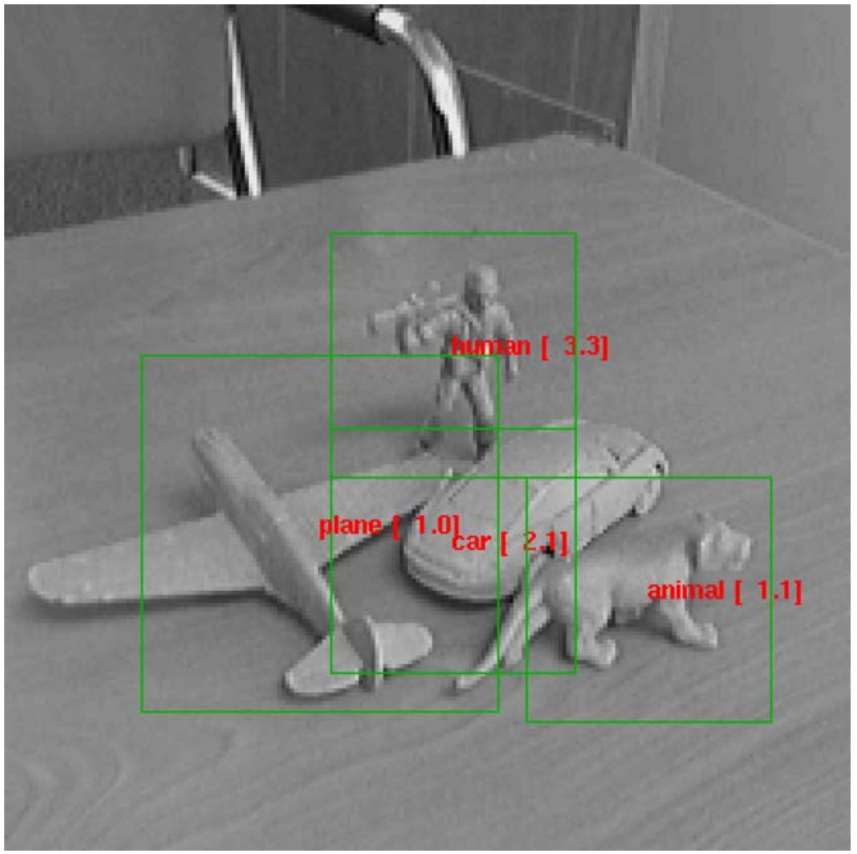
CNNs beyond digits: faces



Rowley 1998

CNNs beyond digits: objects





LeCun et al. 2004

Datasets: MNIST (1998)

- 10 classes
- 70K images

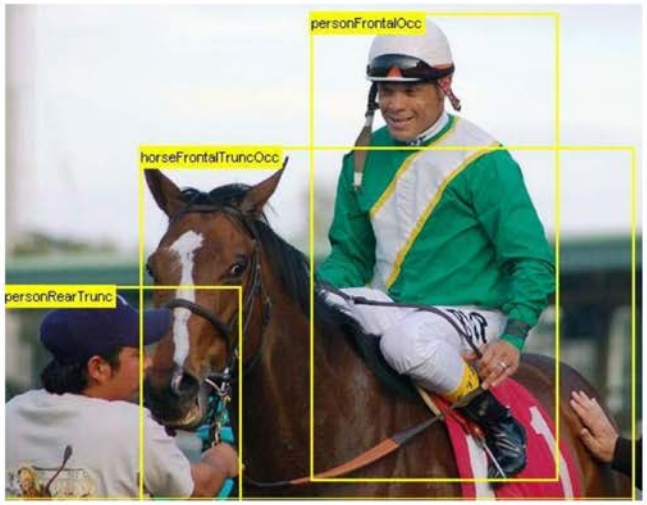
LeCun, Cortes 1998

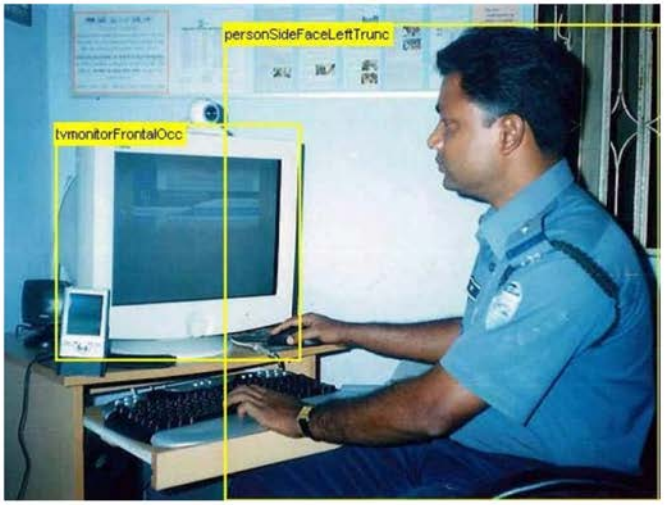
Datasets: Caltech (2004)

- 10 classes
- 10K images

Datasets: PASCAL VOC (2007)

- 20 classes
- 20K images





Van Gool et al. 2010

Datasets: Imagenet

- 22K classes
- 14M images

Deng et al. 2009

Datasets

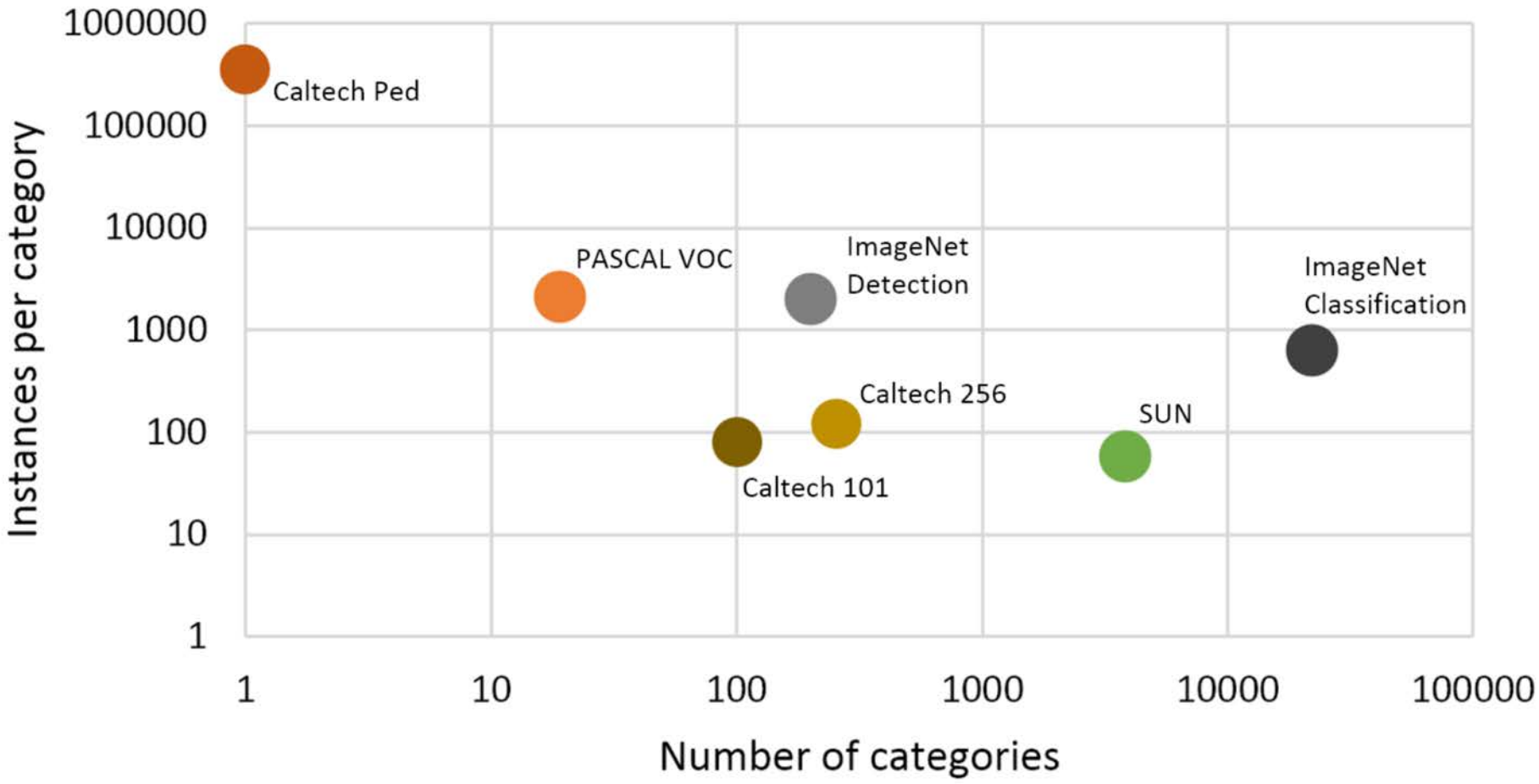


Figure: L. Zitnick

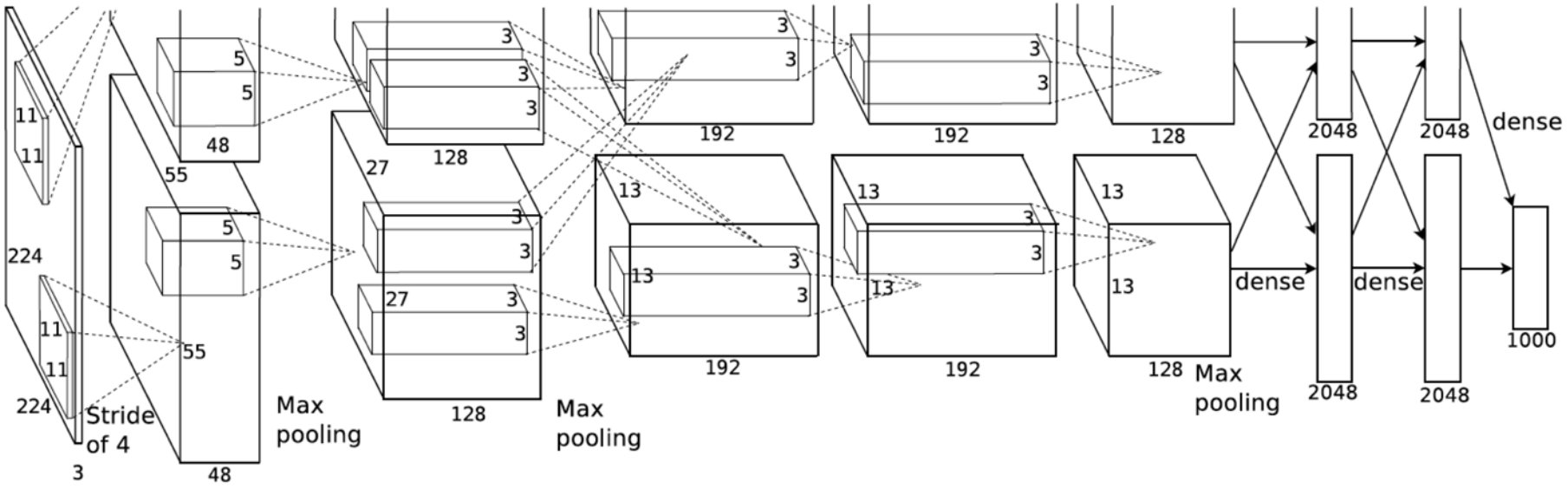
Winning ImageNet

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



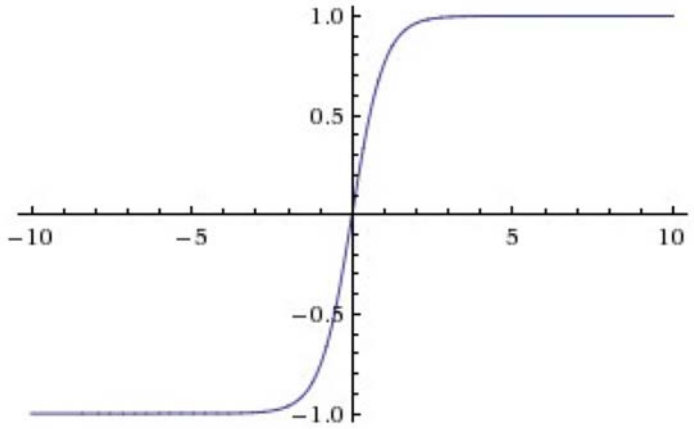
Krizhevsky et al. 2012

Key insights

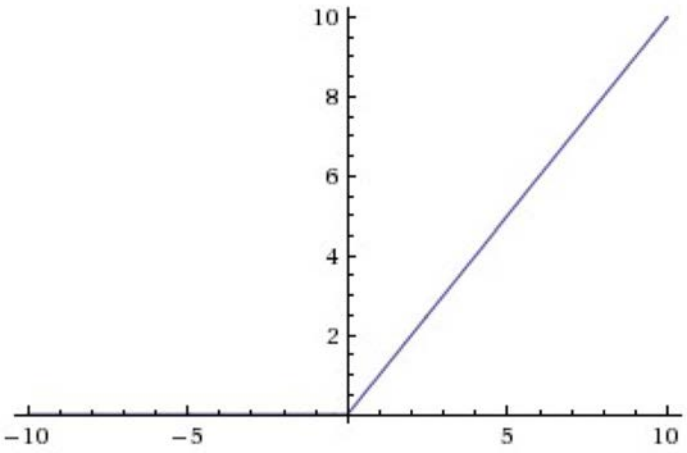
- 11 layers
- Large convolutional filters 11x11
- Max pooling
- ReLU activation
- Dataset: ImageNet (1.2M training images)
- Data augmentation
- Trained on GPU (GTX 580, took 5-6 days!)

Key insights: ReLU

tanh



ReLU



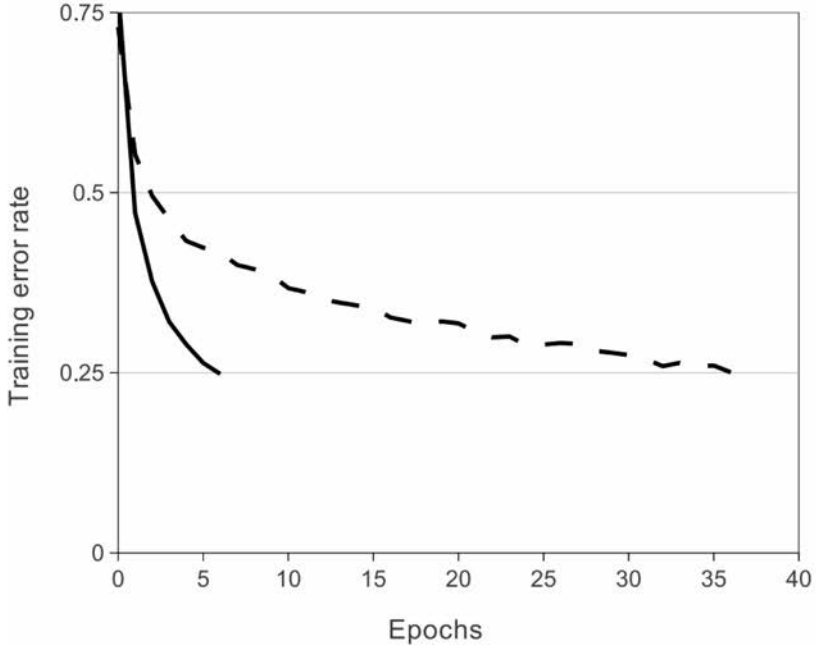
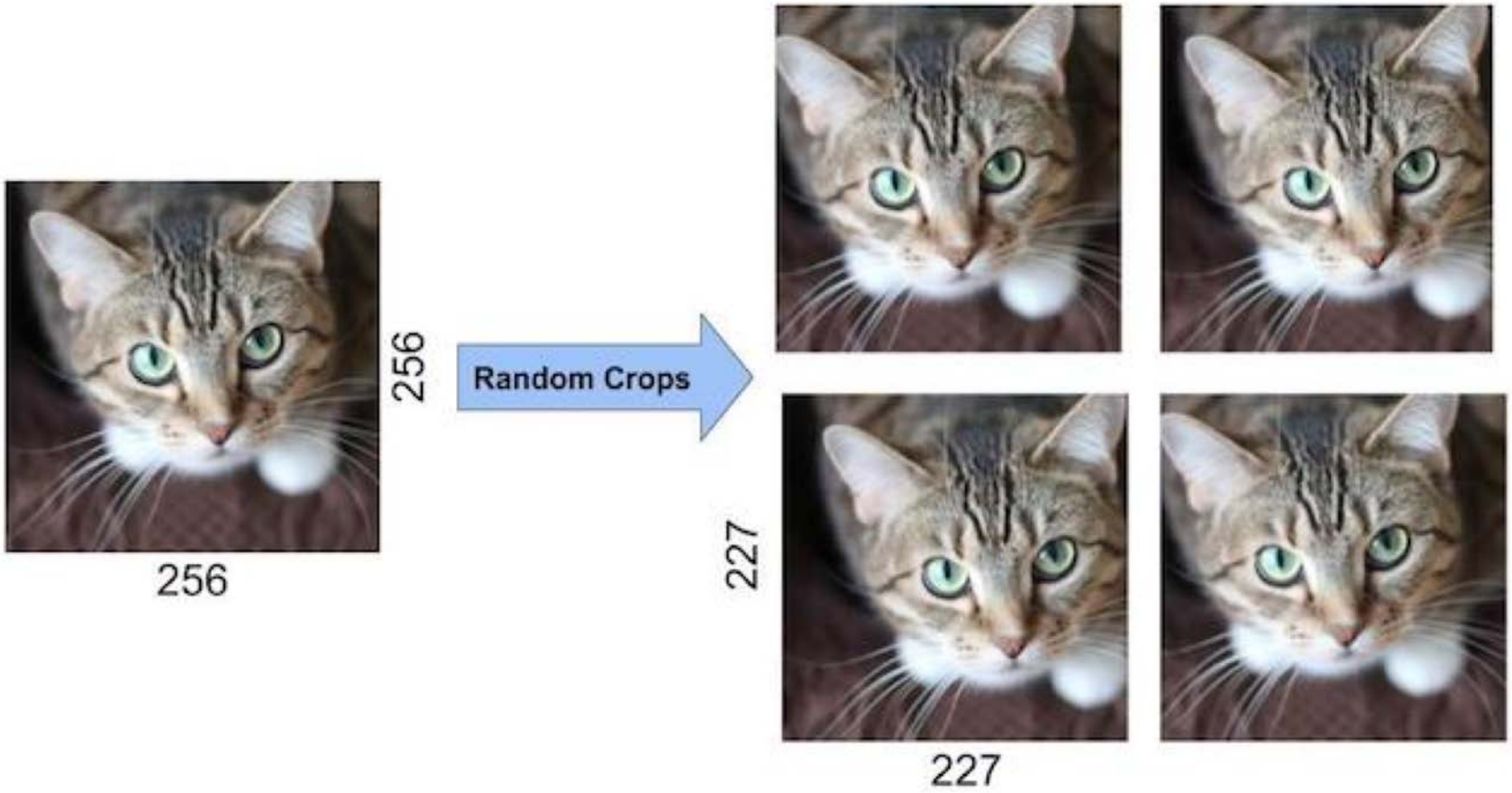
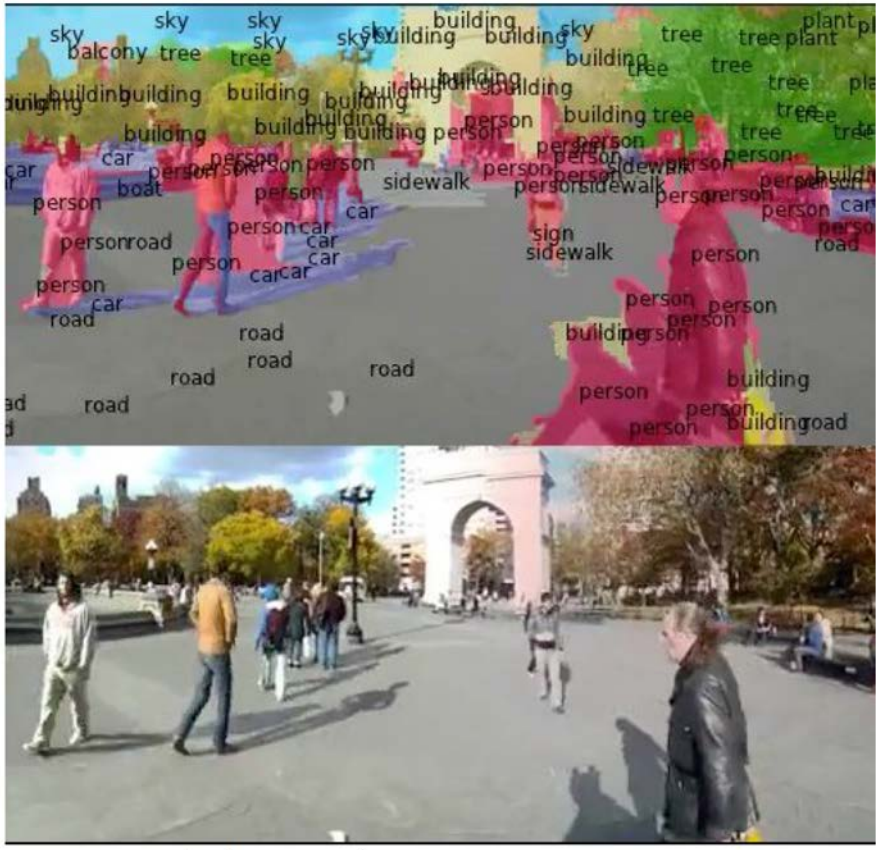


Figure 1: A four-layer convolutional neural network with ReLUs **(solid line)** reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons **(dashed line)**. The learning rates for each network were chosen independently to make training as fast as possible. No regularization of any kind was employed. The magnitude of the effect demonstrated here varies with network architecture, but networks with ReLUs consistently learn several times faster than equivalents with saturating neurons.

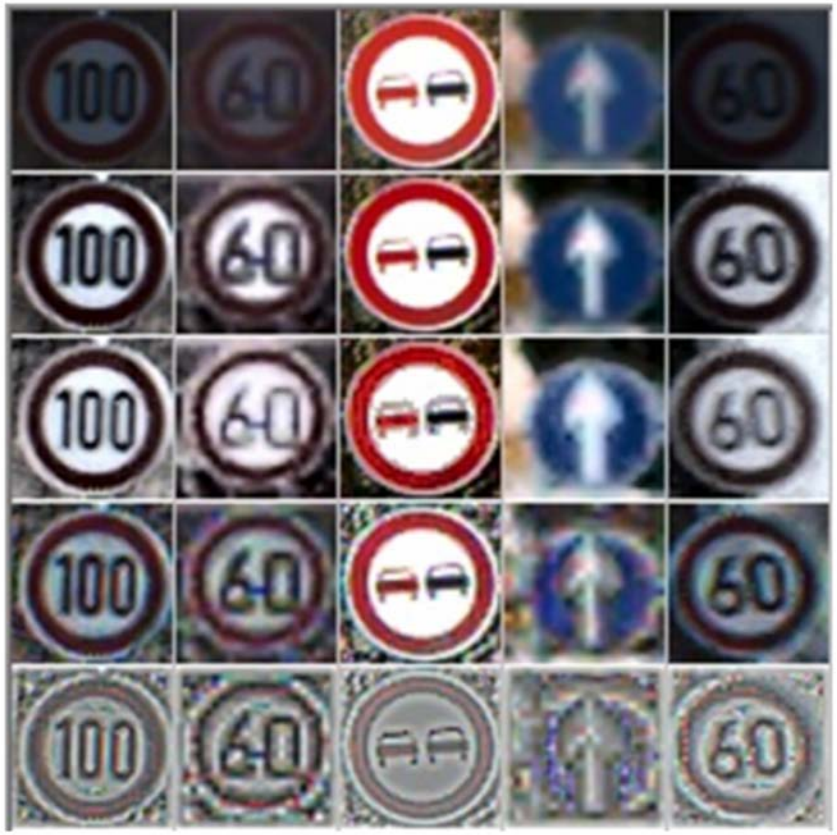
Key insights: data augmentation



Parallel works



Farabet et al. 2012



Ciresan et al. 2012

Today CNNs are everywhere

Applications in computer vision

Toshev 2014

Pose detection

Krizhevsky et al. 2012

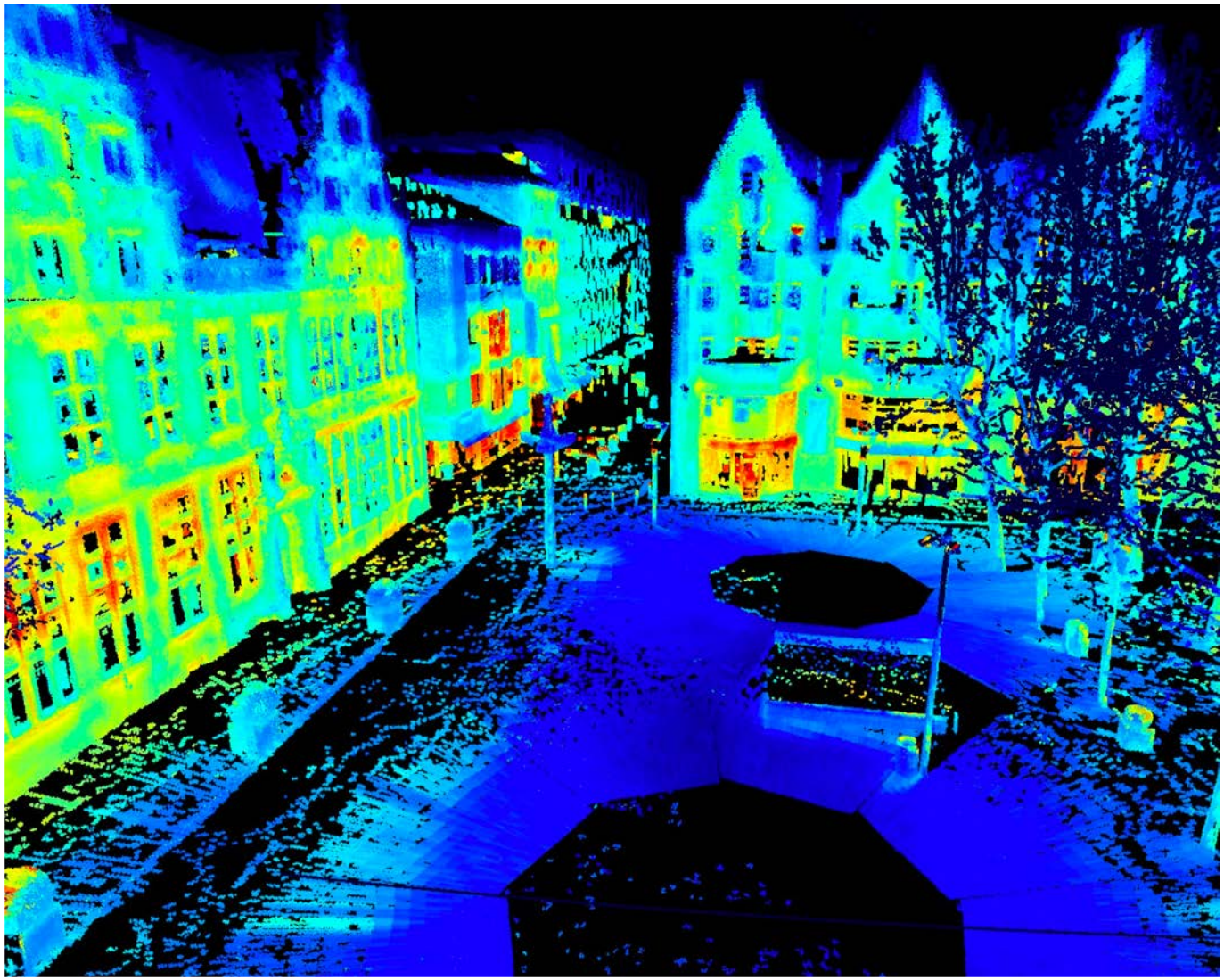
Image retrieval

Style transfer

Artistic style transfer

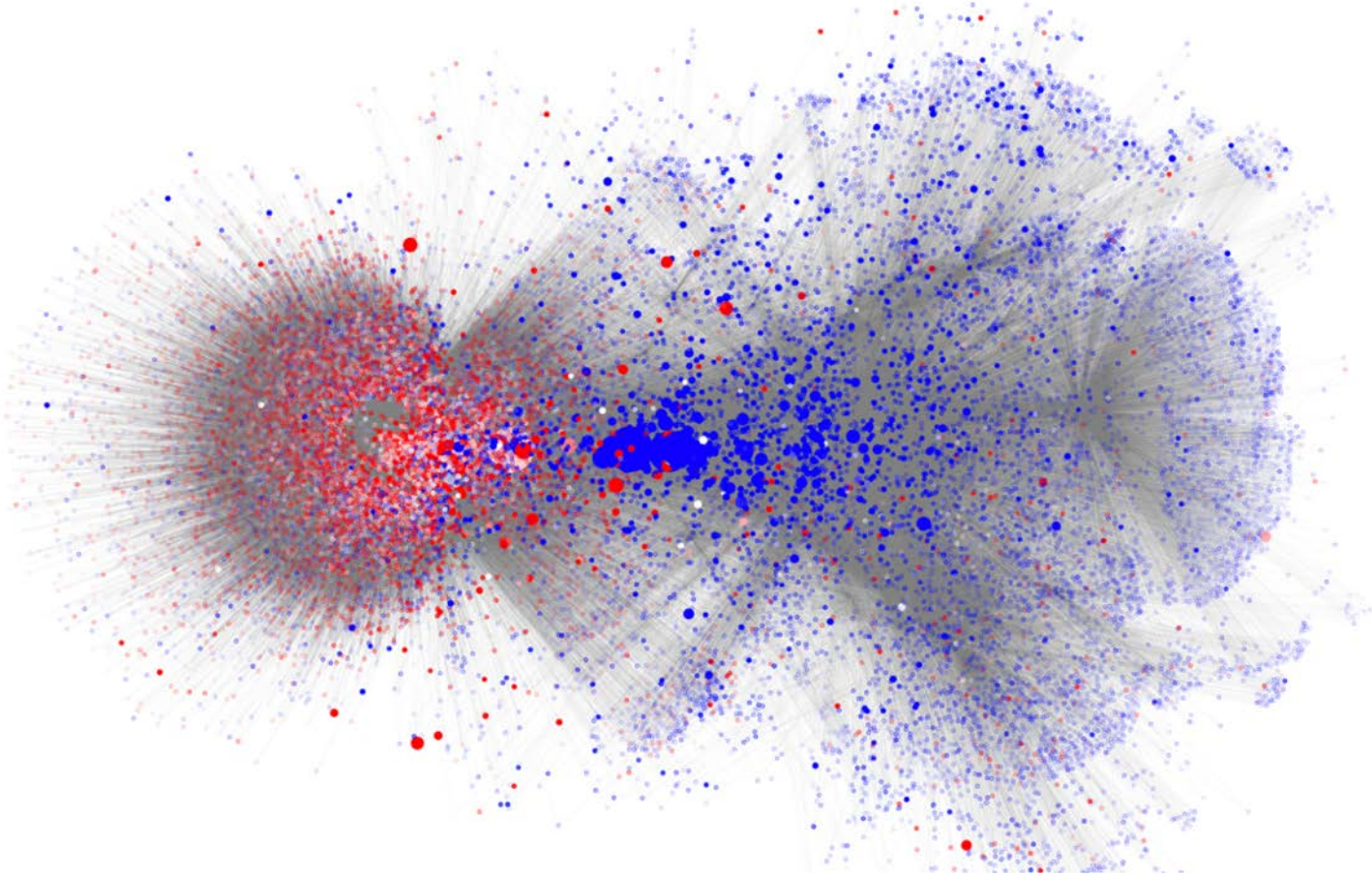
Gatys 2015

3D CNNs



CNNs on point clouds

Graph CNNs



Fake news detection using “Graph CNNs”