Vision AI

Introduction

Notes based on
CS231n, Stanford University, and
EECS 498-007 / 598-005, University of Michigan
with permission from Justin Johnson

Deep Learning for Computer Vision

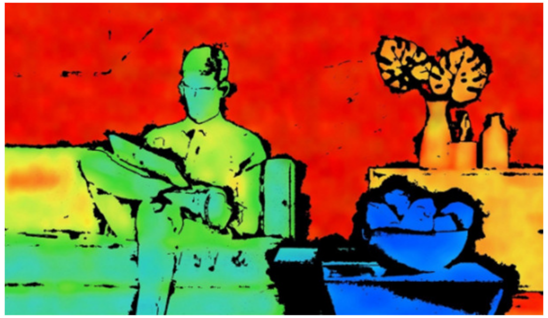
Building artificial systems that

process,

perceive, and

reason about visual data

Computer Vision is <u>everywhere</u>

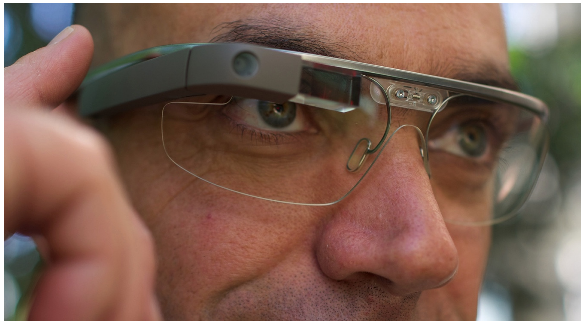


Left to right:
Image by Intel is copyrighted
Image from Unsplash
Image from Unsplash
Image by Nasa is public



Left to right:
Image by Nasa is public
Image from Unsplash
Image from Unsplash
Image from Unsplash



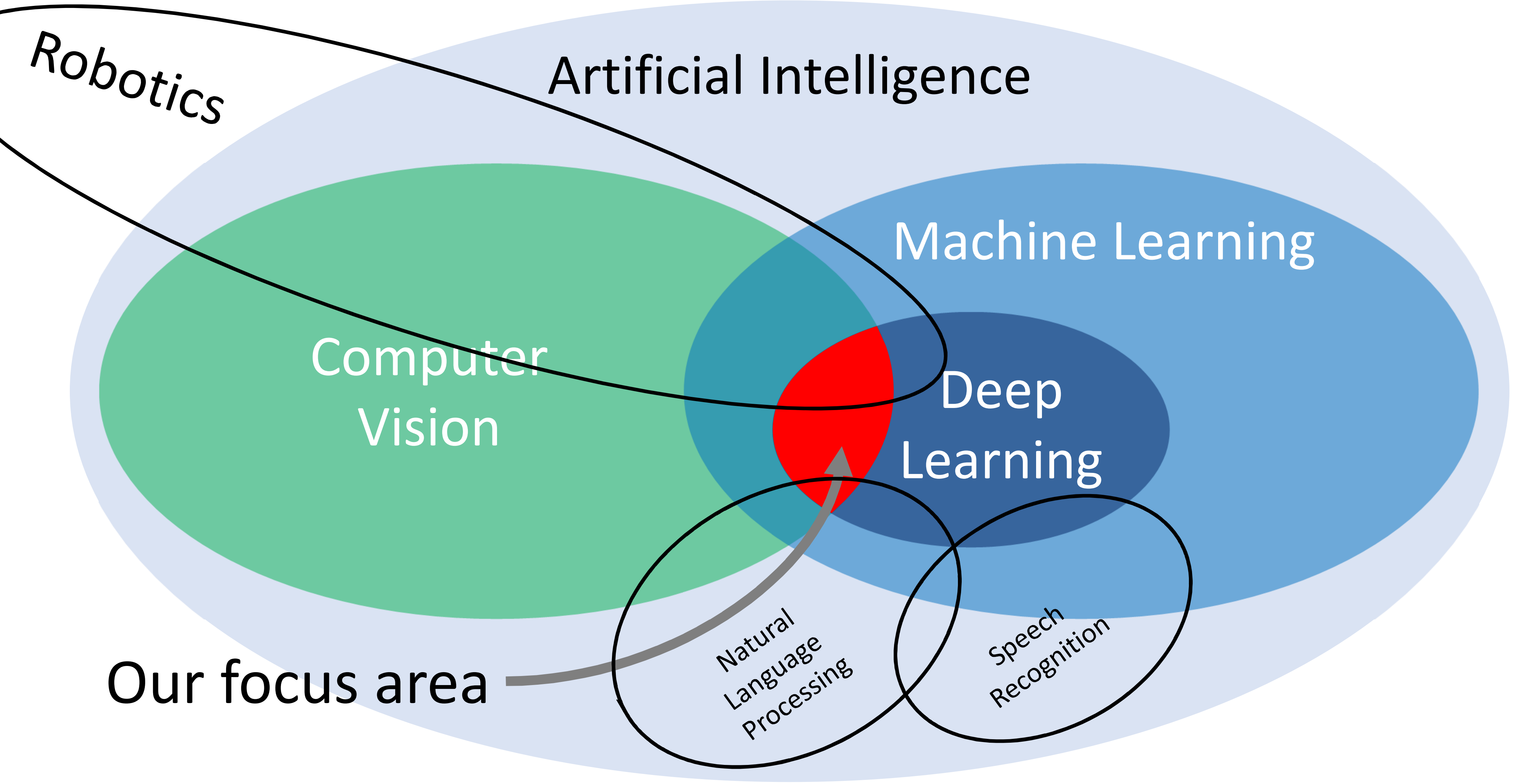
Left to right:
Image from Unsplash
Image from Unsplash
Image from Unsplash
Image copyrighted by ExtremeTech

Deep <u>Learning</u> for Computer Vision

Building artificial systems that
learn from data and experience

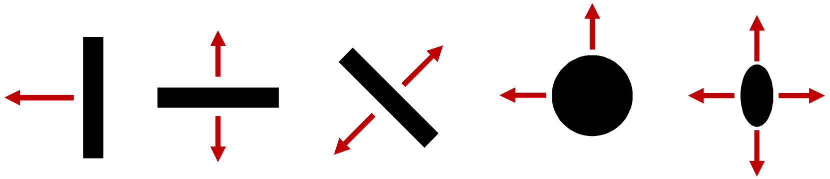
Deep Learning for Computer Vision

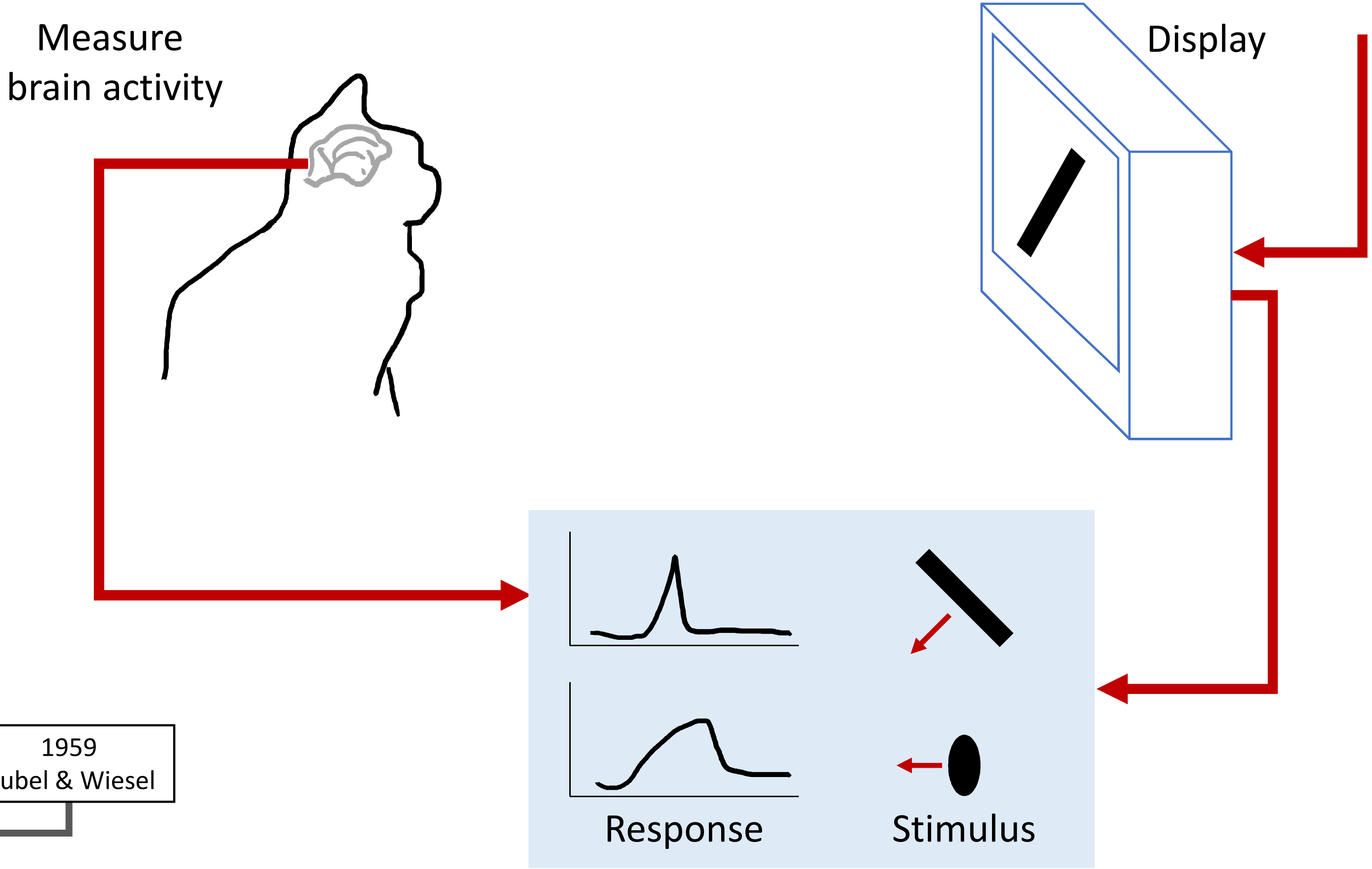
Hierarchical learning algorithms
with many “layers”,
(very) loosely inspired by the brain


Robotics
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition
Our focus area

A brief history of computer vision and deep learning

Hubel and Wiesel, 1959 [source]

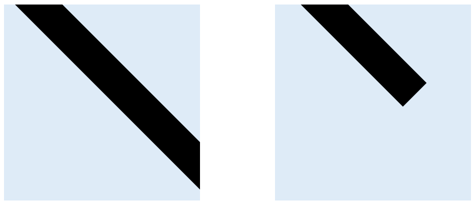




Simple cells:
Response to light
orientation

Complex cells:
Response to light
orientation and movement

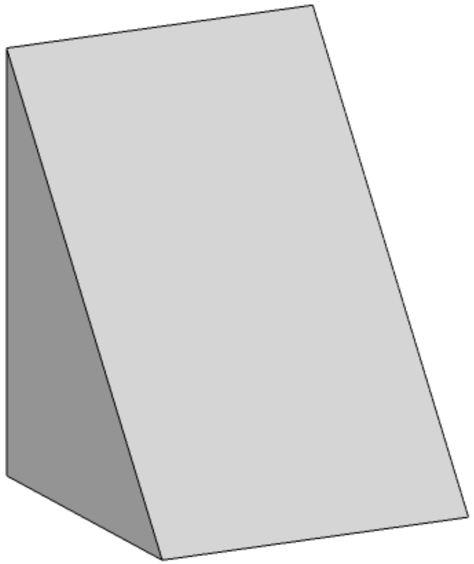
Hypercomplex cells:
response to movement
with an end point



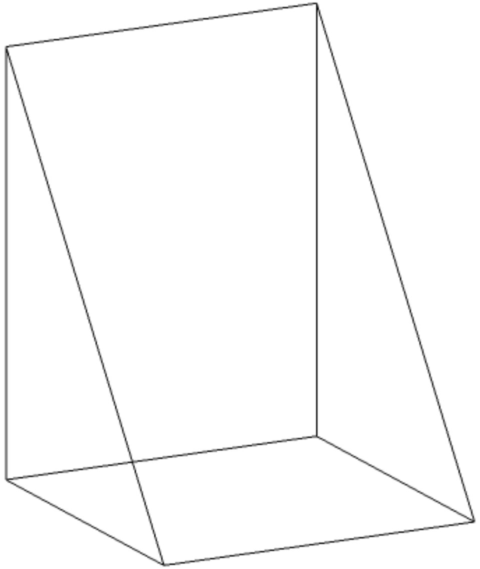
No response

1959
Hubel & Wiesel

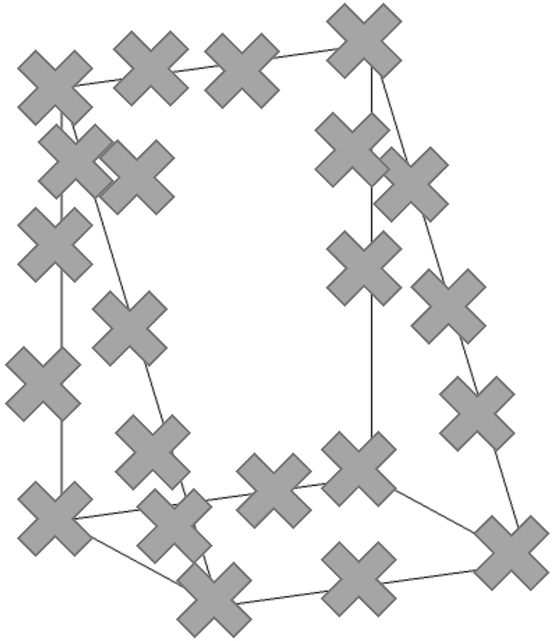
Larry Roberts, 1963



(a) Original picture



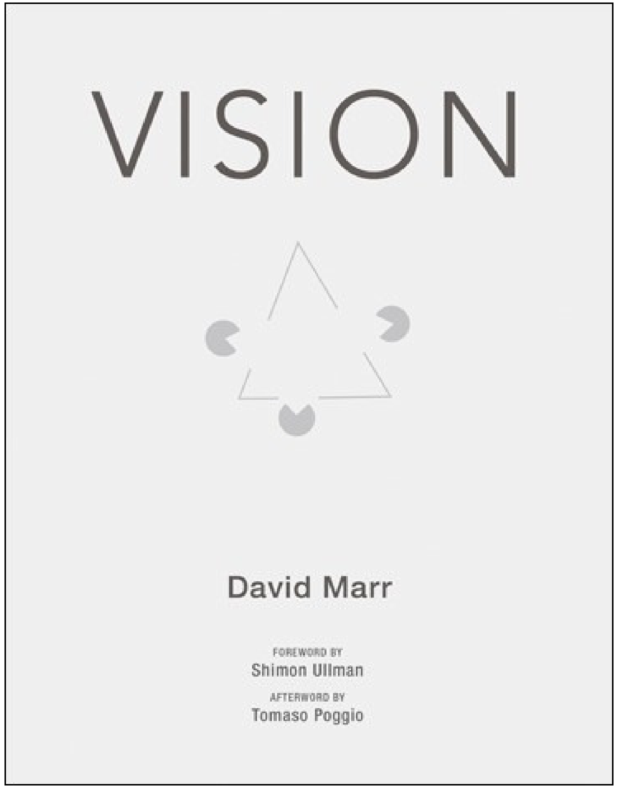
(b) Differentiated picture



(c) Feature points selected

1959
Hubel & Wiesel

1963
Roberts



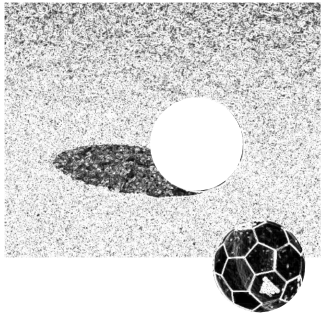
1959	1963	1970s
Hubel & Wiesel	Roberts	David Marr

Input image

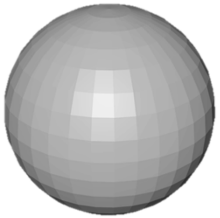
This image from Unsplash

Edge image

2 ½-D Sketch



3-D Model



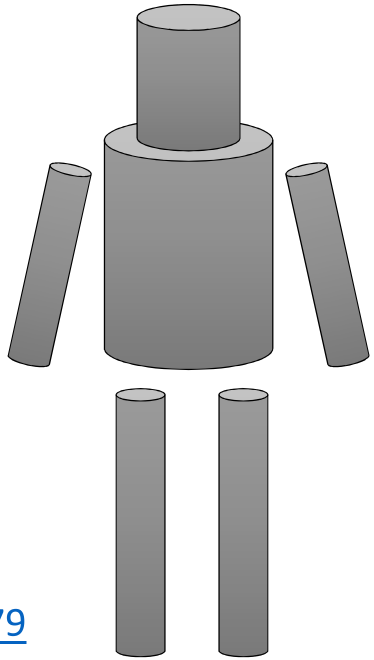
This image from pixabay

Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

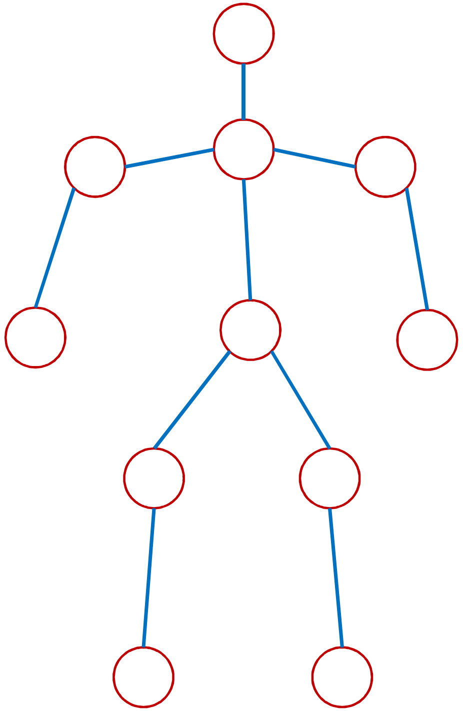
Stages of Visual Representation, David Marr, 1970s

RDM reliable.deliver.manage

Recognition via Parts (1970s)



Generalized Cylinders,
[Brooks and Binford, 1979]

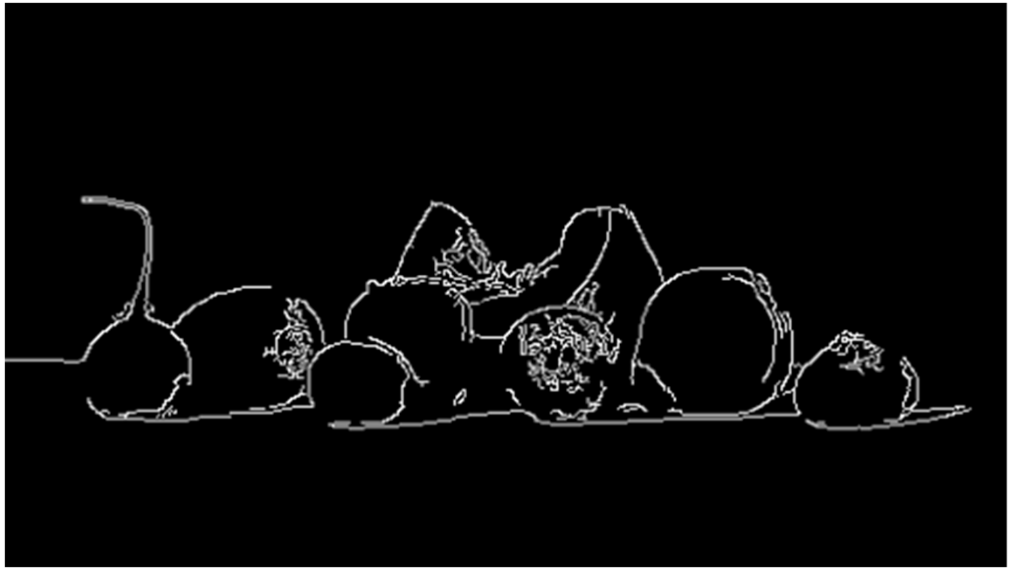


Pictorial Structures,
[Fischler and Elshlager, 1973]

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders

Recognition via Edge Detection (1980s)

Image from Unsplash



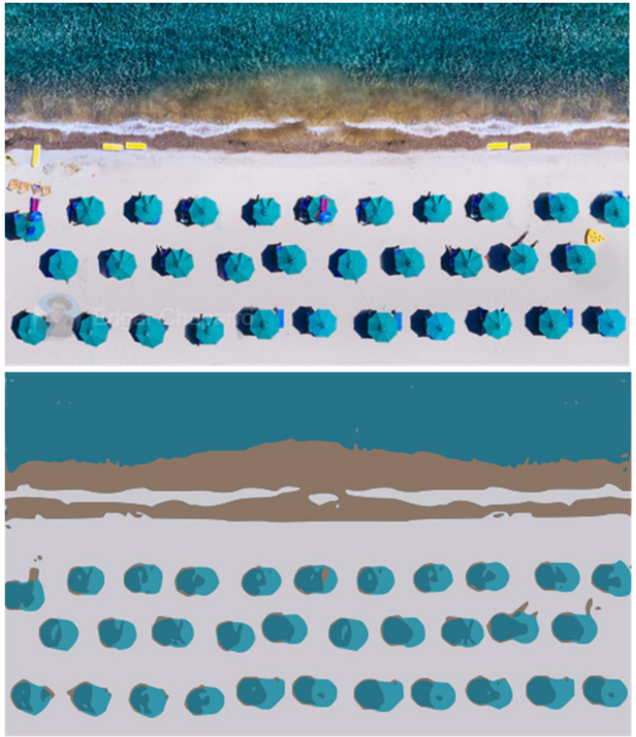
Apply Canny edge detection in OpenCV

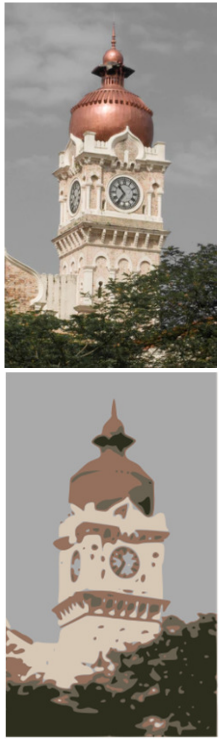
1959 Hubel & Wiesel	1963 Roberts	1970s David Marr	1979 Gen. Cylinders	1986 Canny

John Canny, 1986
David Lowe, 1987

reliable.deliver.manage

Recognition via Grouping (1990s)





1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

Normalized Cuts, Shi and Malik, 1997

Image from Unsplash

Image from Unsplash

Image from Unsplash

Recognition via Matching (2000s)

Image is Open Source   Image is Open Source   Implemented code

1959 Hubel & Wiesel — 1963 Roberts — 1970s David Marr — 1979 Gen. Cylinders — 1986 Canny — 1997 Norm. Cuts — 1999 SIFT

SIFT, David Lowe, 1999

reliable.deliver.manage

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision

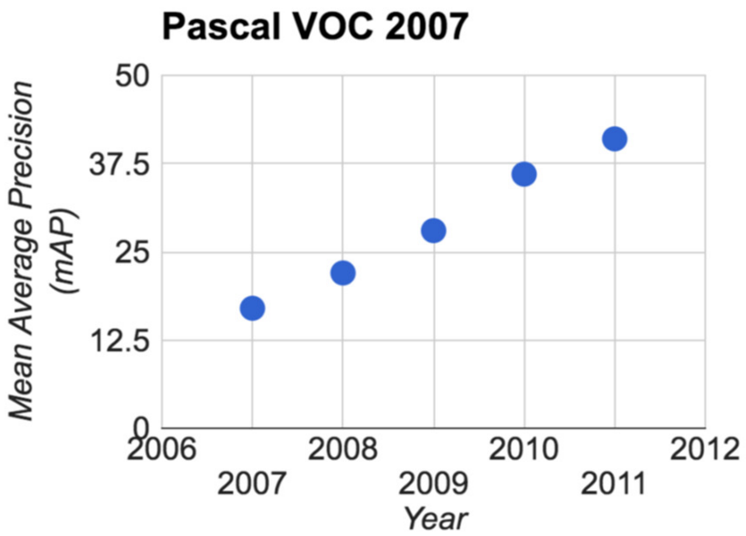
1959 Hubel & Wiesel	1963 Roberts	1970s David Marr	1979 Gen. Cylinders	1986 Canny	1997 Norm. Cuts	1999 SIFT	2001 V&J

PASCAL Visual Object Challenge [source]

Image from Unsplash

Image from Unsplash

Image from Unsplash



1959 Hubel & Wiesel	1963 Roberts	1970s David Marr	1979 Gen. Cylinders	1986 Canny	1997 Norm. Cuts	1999 SIFT	2001 V&J	2007 PASCAL

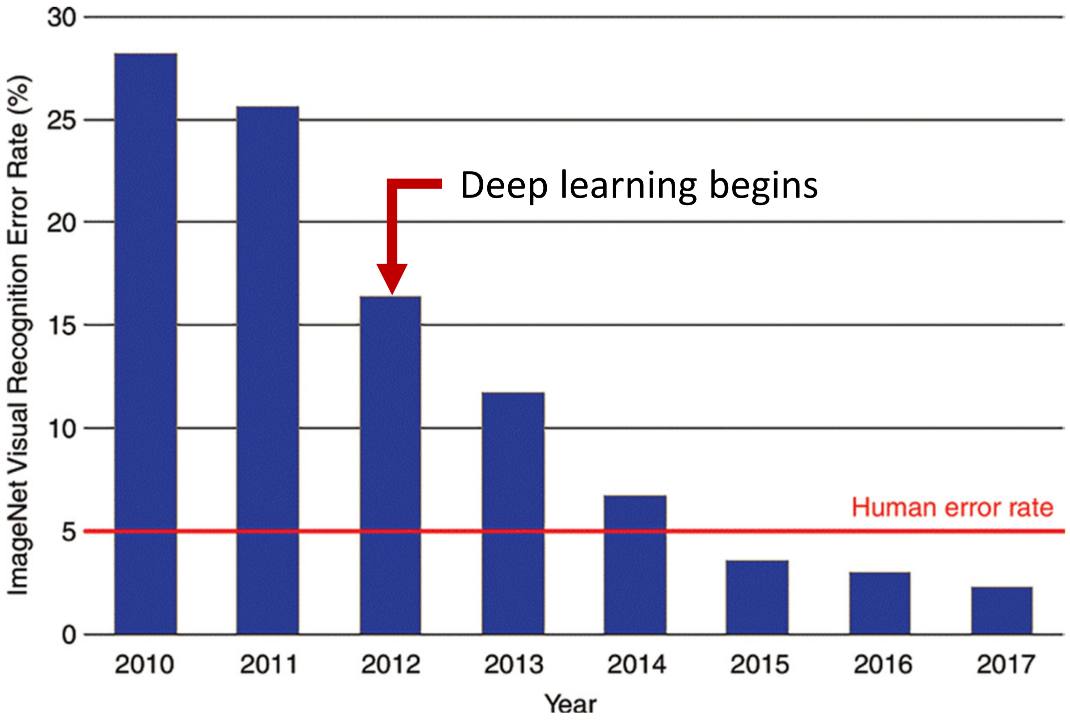
IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Deng et al, 2009
Russakovsky et al. IJCV 2015

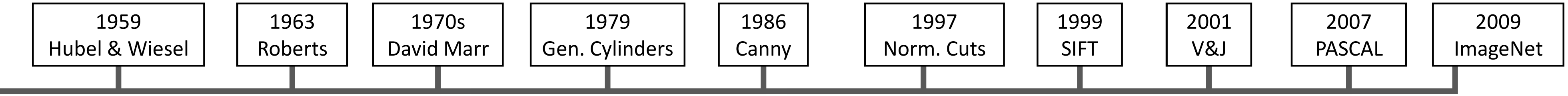
1959	1963	1970s	1979	1986	1997	1999	2001	2007	2009
Hubel & Wiesel	Roberts	David Marr	Gen. Cylinders	Canny	Norm. Cuts	SIFT	V&J	PASCAL	ImageNet

IMAGENET Large Scale Visual Recognition Challenge



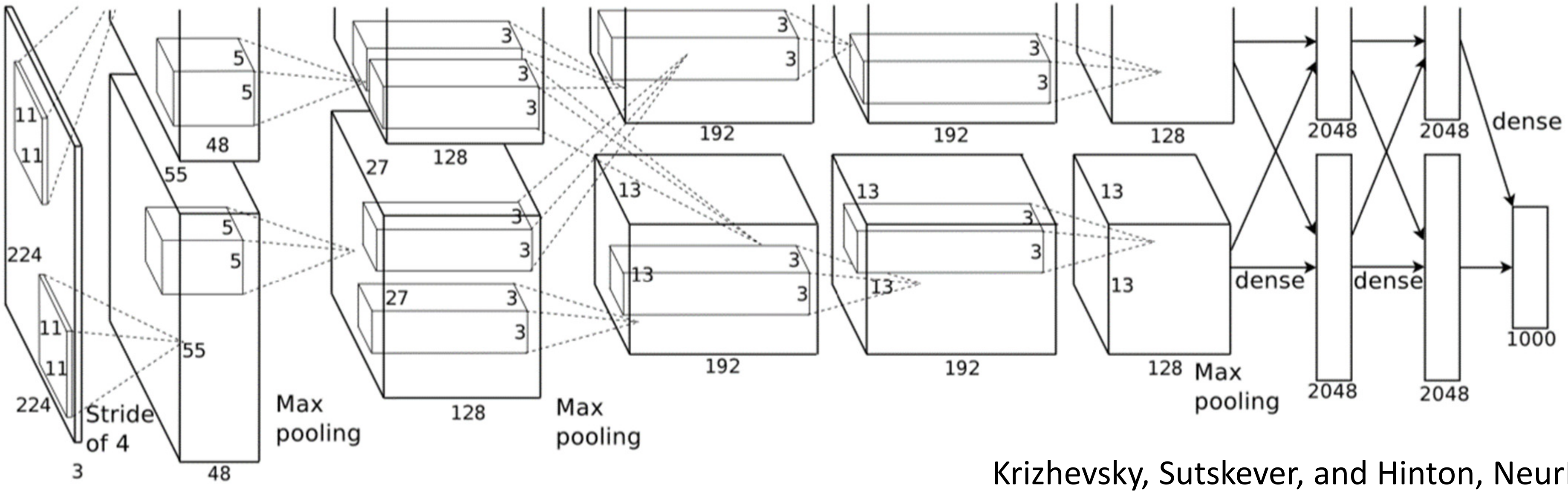
[Interactive Graph]()

Deng et al, 2009
Russakovsky et al. IJCV 2015



RDM reliable.deliver.manage

Deep Learning now famous: AlexNet



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012
[paper]

1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

2007
PASCAL

2009
ImageNet

2012
AlexNet

reliable.deliver.manage

Perceptron

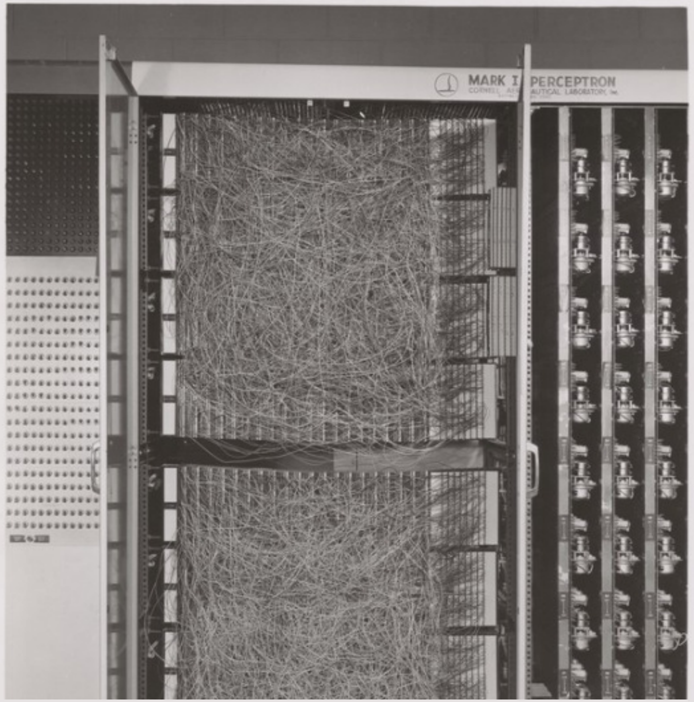
One of the earliest algorithms that could learn from data

Implemented in hardware! Weights stored in potentiometers, updated with electric motors during learning

Connected to a camera that used 20x20 cadmium sulfide photocells to make a 400-pixel image

Could learn to recognize letters of the alphabet

Today we would recognize it as a **linear classifier** [try]



Mark I Perceptron

at the Cornell Aeronautical Laboratory
Cornell University News Service records, #4-3-15. Division of Rare and Manuscript Collections, Cornell University Library.

1958 Perceptron

1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

1999 SIFT

2001 V&J

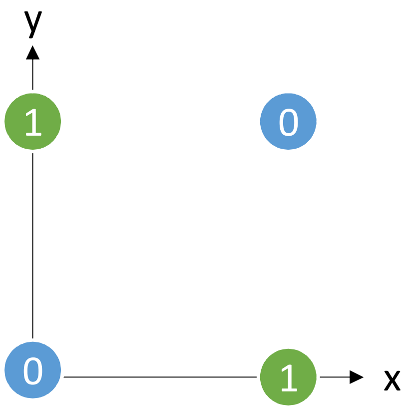
2007 PASCAL

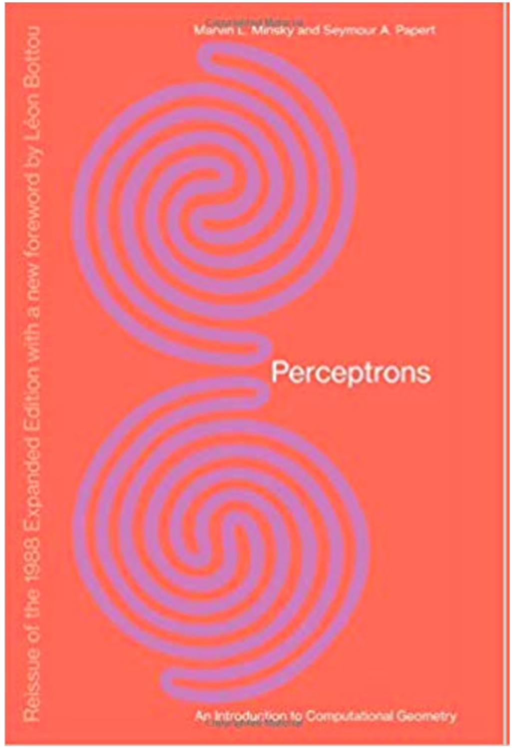
2009 ImageNet

2012 AlexNet

Minsky and Papert, 1969

X	y	XOR(x,y)
0	0	0
0	1	1
1	0	1
1	1	0

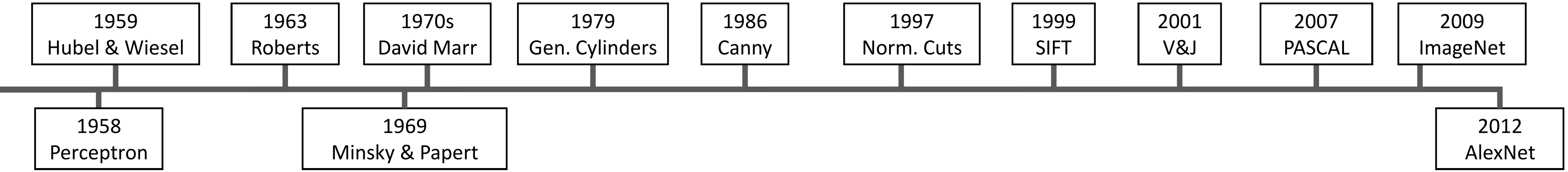




MIT Press

Showed that Perceptrons could not learn the XOR function.

Caused a lot of disillusionment in the field.

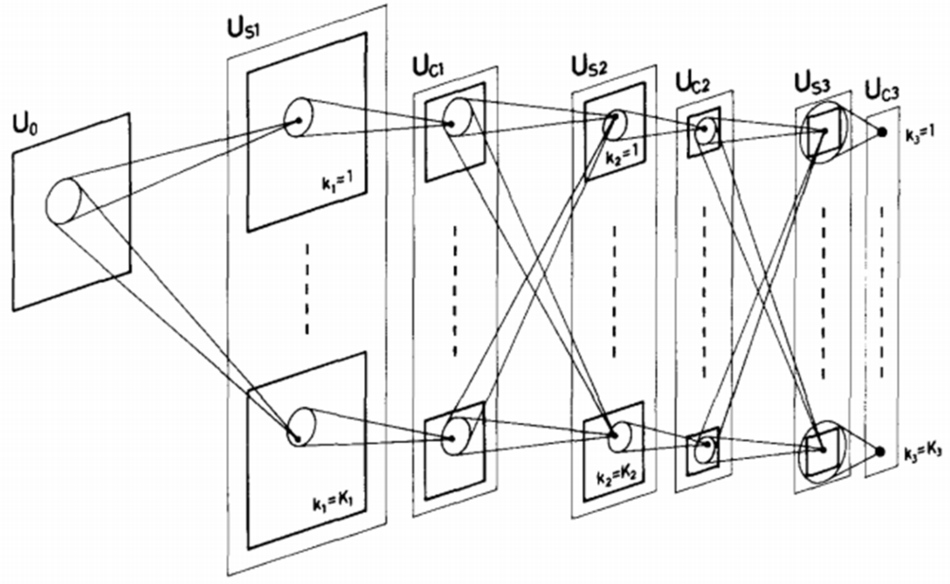


Neocognitron: Fukushima, 1980 [paper]

Computational model the visual system, directly inspired by Hubel and Wiesel's hierarchy of complex and simple cells

Interleaved simple cells (convolution) and complex cells (pooling)

No practical training algorithm



- 1958 Perceptron
- 1959 Hubel & Wiesel
- 1963 Roberts
- 1969 Minsky & Papert
- 1970s David Marr
- 1979 Gen. Cylinders
- 1980 Neocognitron
- 1986 Canny
- 1997 Norm. Cuts
- 1999 SIFT
- 2001 V&J
- 2007 PASCAL
- 2009 ImageNet
- 2012 AlexNet

reliable.deliver.manage

Backprop: Rumelhart, Hinton, and Williams, 1986 [paper]

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$

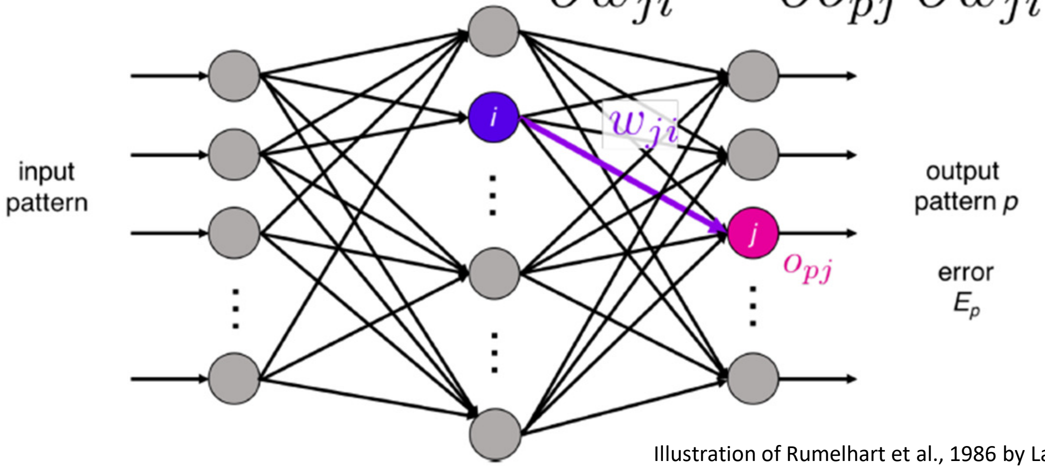
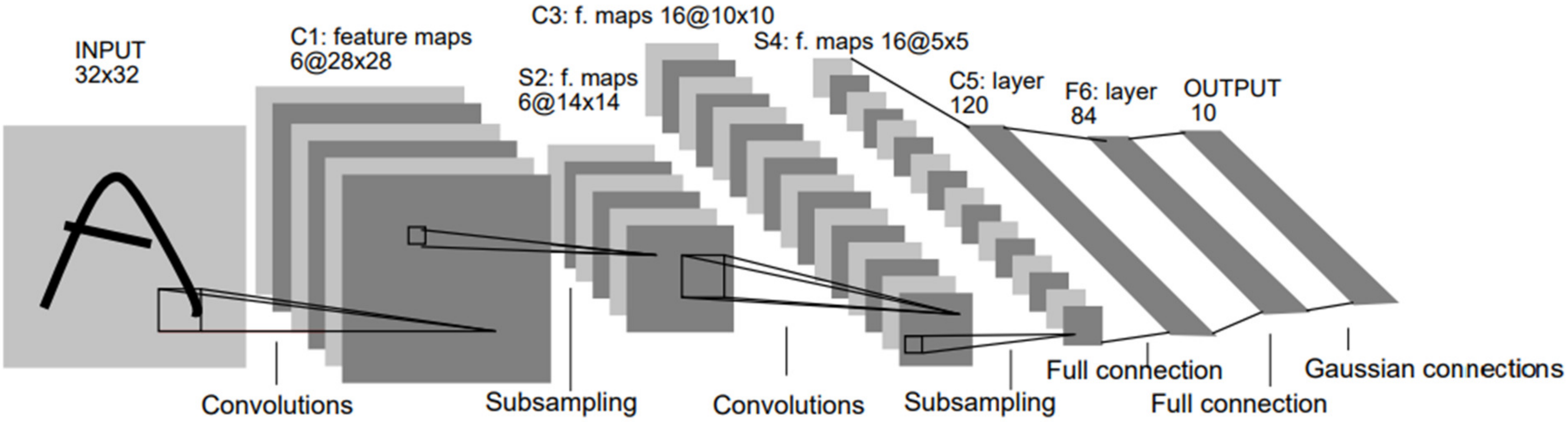


Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2007 PASCAL | 2009 ImageNet

1958 Perceptron | 1969 Minsky & Papert | 1980 Neocognitron | 1985 Backprop. | 2012 AlexNet

Convolutional Networks: LeCun et al, 1998 [Lecun's Paper]



- Applied backprop algorithm to a Neocognitron-like architecture
- Learned to recognize handwritten digits
- Was deployed in a commercial system by NEC, processed handwritten checks
- Very similar to our modern convolutional networks!

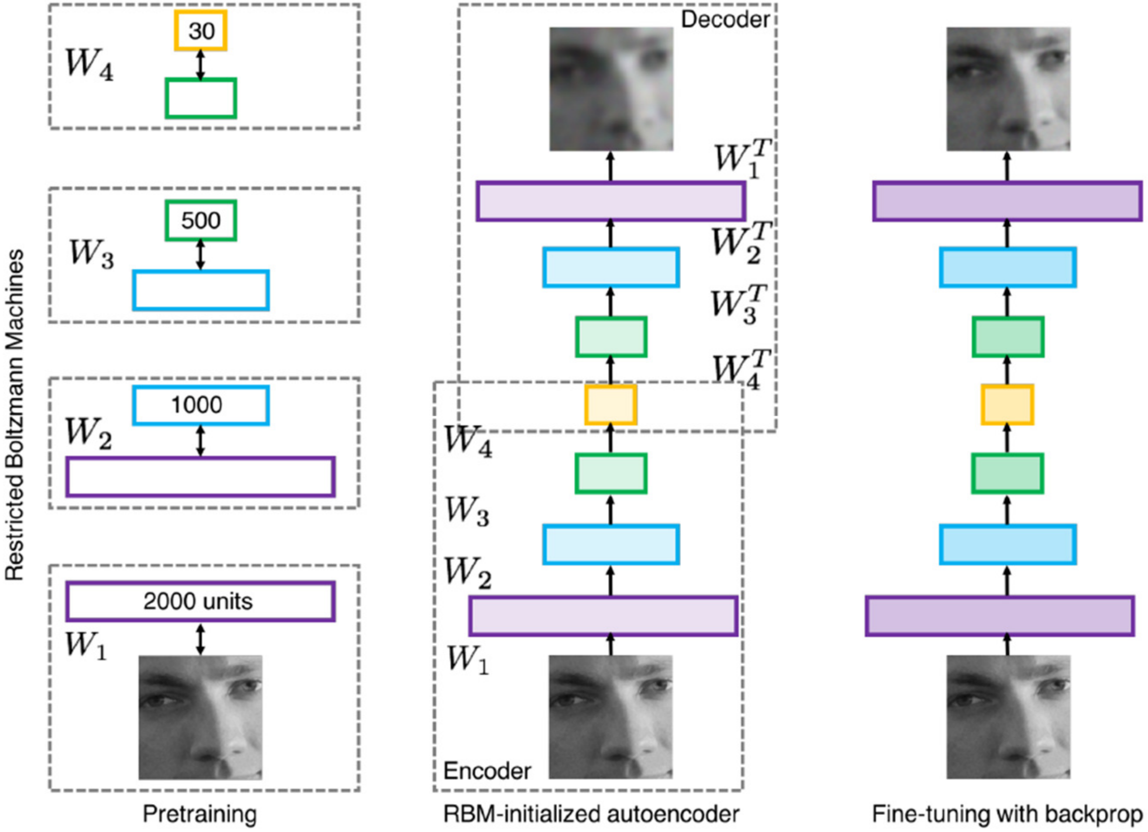
1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1997 Norm. Cuts
1999 SIFT
2001 V&J
2007 PASCAL
2009 ImageNet

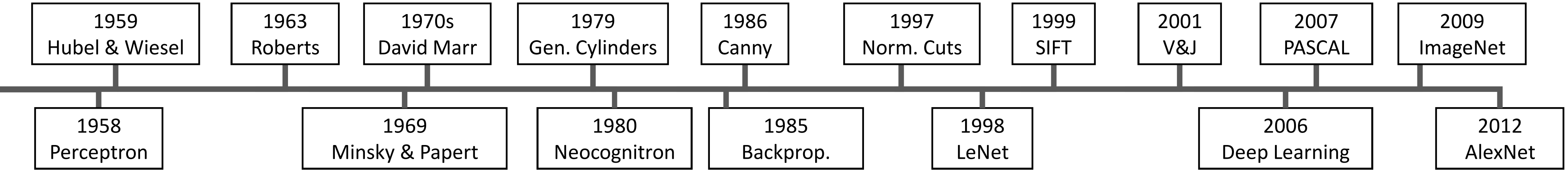
1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
1985 Backprop.
1998 LeNet
2012 AlexNet

2000s: "Deep Learning"

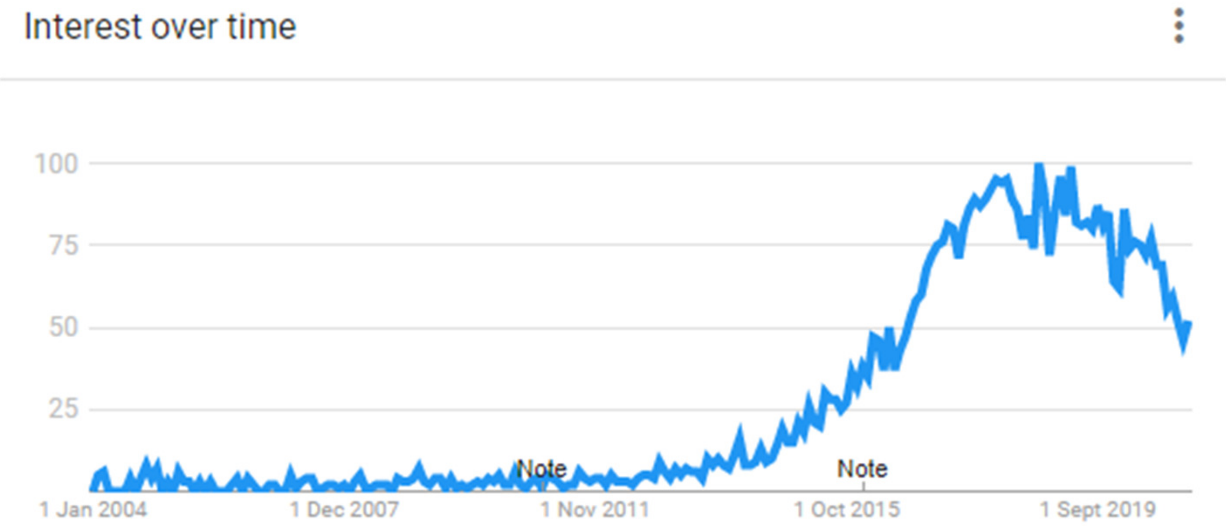
- People tried to train neural networks that were deeper and deeper
- Not a mainstream research topic at this time

Hinton and Salakhutdinov, 2006
Bengio et al, 2007
Lee et al, 2009
Glorot and Bengio, 2010



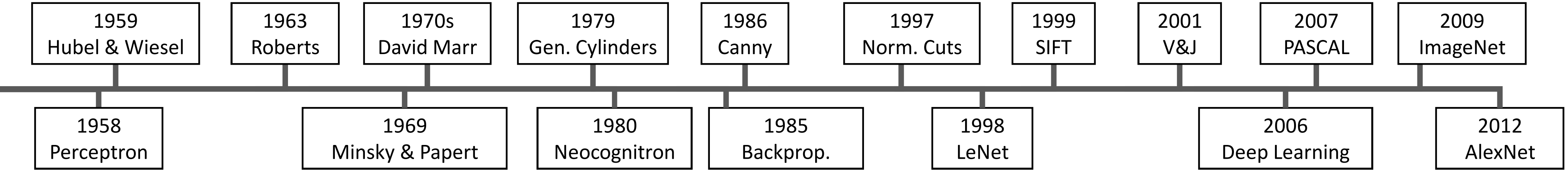


2012 to Present: Deep Learning Explosion



Google Trends: "Deep Learning"

https://trends.google.com/trends/explore?date=all&q=Deep%20learning



2012 to Present: ConvNets are everywhere

Image Classification

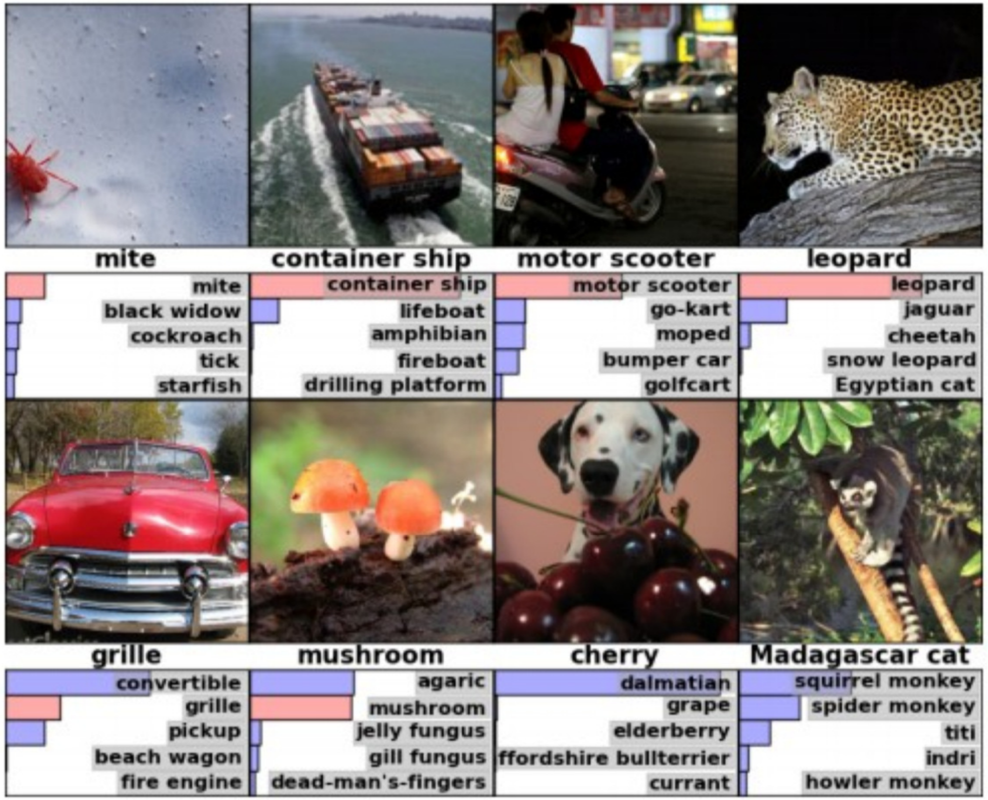


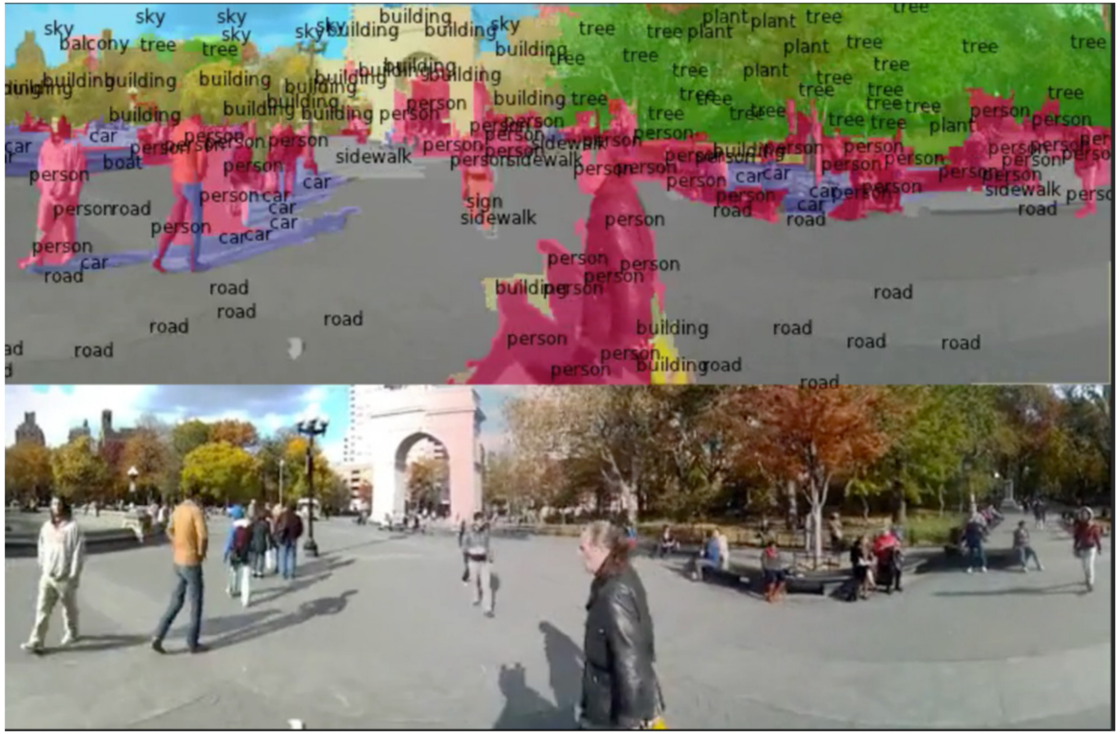
Image Retrieval

2012 to Present: ConvNets are everywhere

Object Detection

Ren, He, Girshick, and Sun, 2015

Image Segmentation

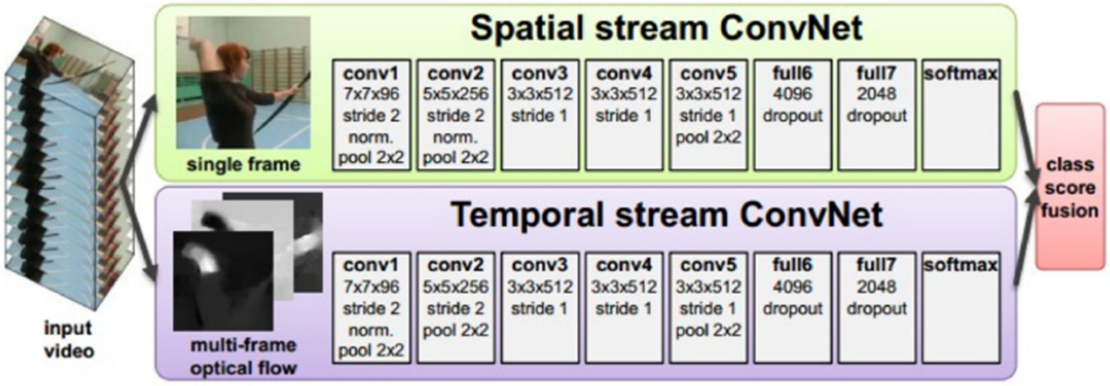


Fabaret et al, 2012

reliable.deliver.manage

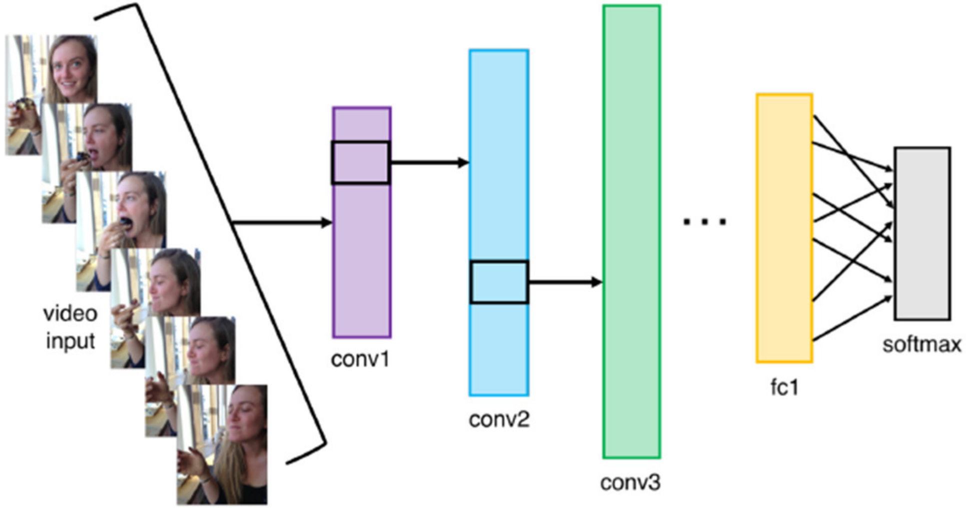
2012 to Present: ConvNets are everywhere

Video Classification



Simonyan et al, 2014

Activity Recognition

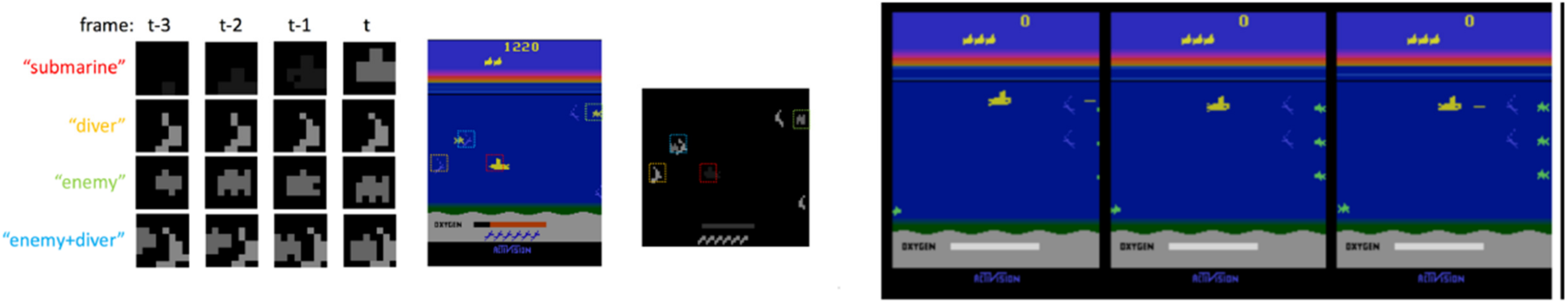


RDM reliable.deliver.manage

2012 to Present: ConvNets are everywhere

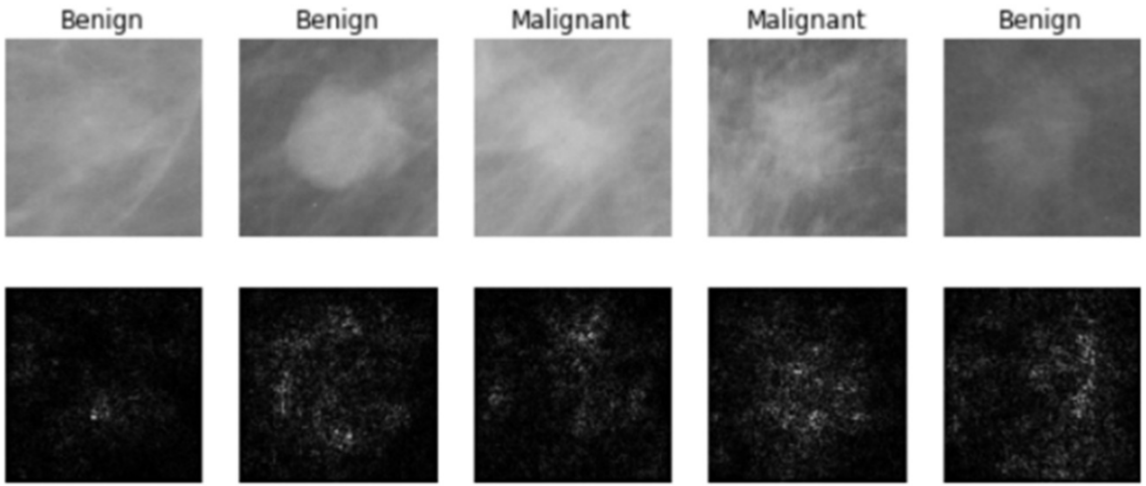
Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



2012 to Present: ConvNets are everywhere

Medical Imaging, Levy et al, 2016



Galaxy Classification, Dieleman et al, 2014

Whale recognition

[Kaggle Challenge]

2012 to Present: ConvNets are everywhere

Image Captioning

Vinyals et al, 2015

Karpathy and Fei-Fei, 2015

A white teddy bear sitting in the grass

A man in a baseball uniform throwing a ball

A woman is holding a cat in her hand

A man riding a wave on top of a surfboard

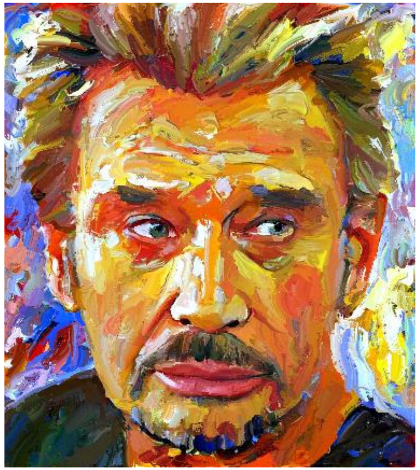
A cat sitting on a suitcase on the floor

A woman standing on a beach holding a surfboard

Captions generated by Justin Johnson using Neuraltalk2

Style transfer

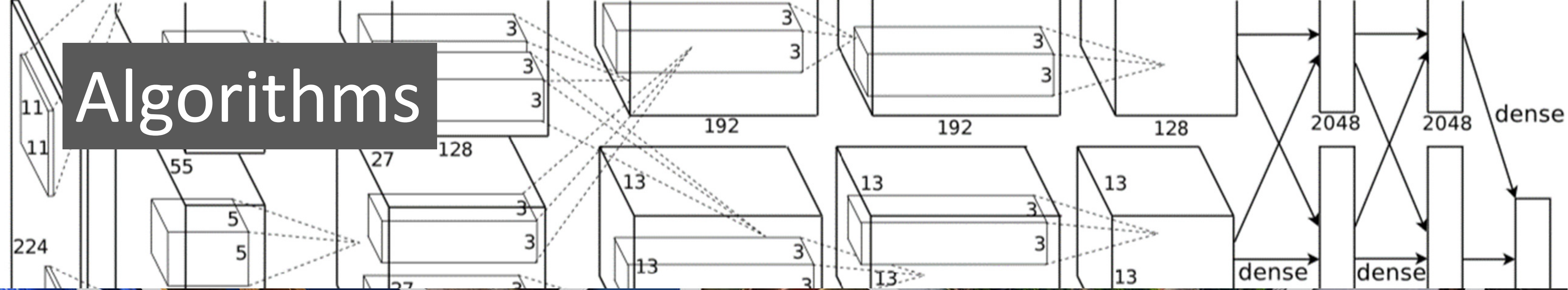
https://deepart.io



Style transfer

https://deepart.io

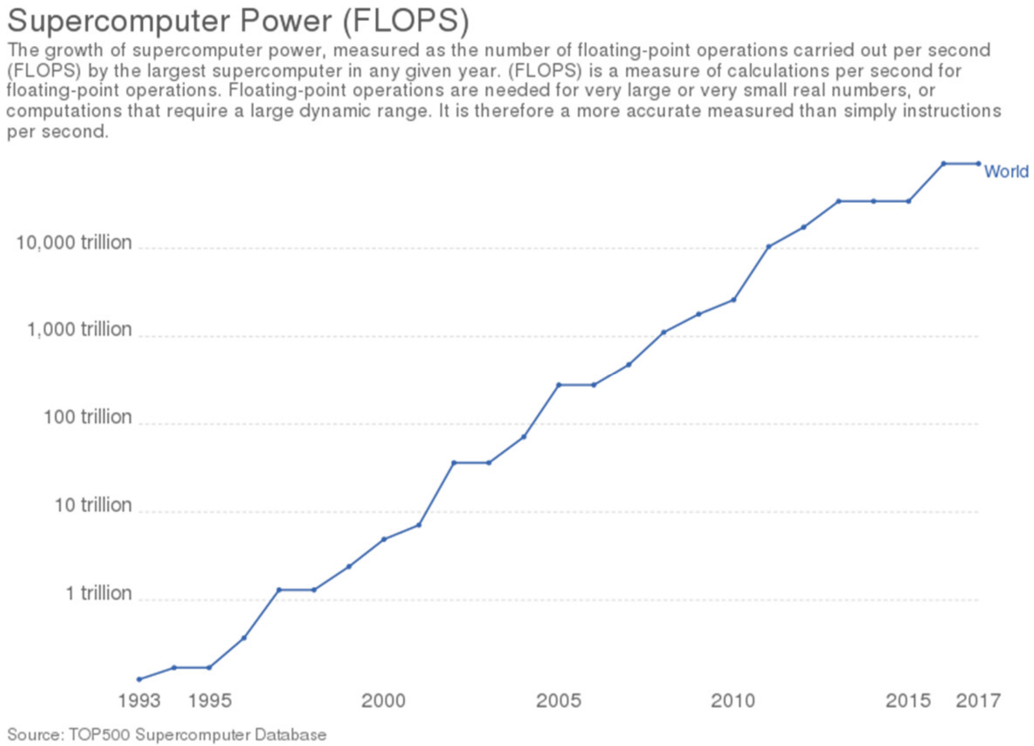
Algorithms

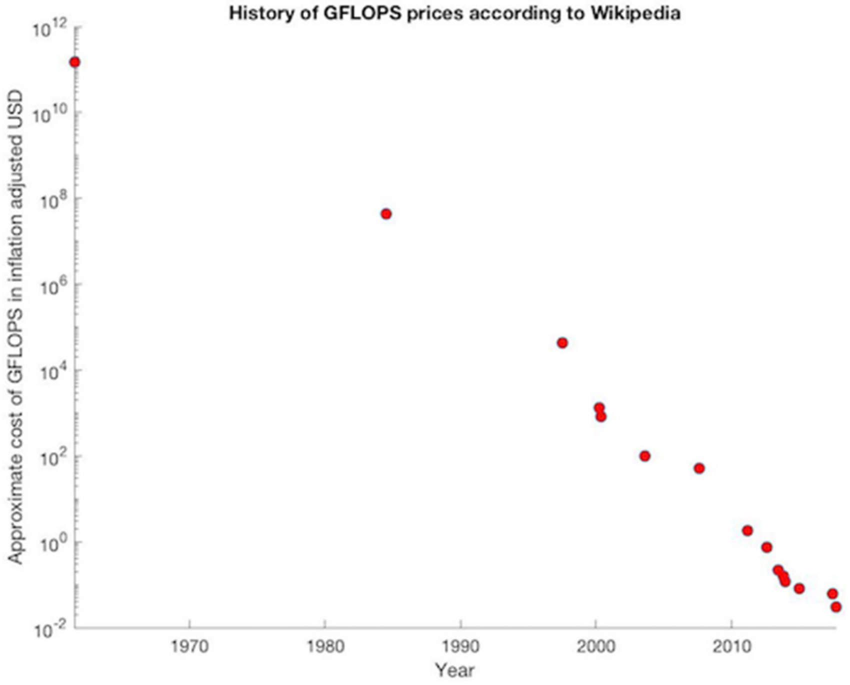


Data

Computation

Higher GigaFLOPs at a lower cost





http://en.wikipedia.org/wiki/FLOPS#Hardware_costs

2018 Turing Award

For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

Yoshua Bengio

Geoffrey Hinton

Yann LeCun

1958 Perceptron
1959 Hubel & Wiesel
1963 Roberts
1969 Minsky & Papert
1970s David Marr
1979 Gen. Cylinders
1980 Neocognitron
1985 Backprop.
1986 Canny
1997 Norm. Cuts
1998 LeNet
1999 SIFT
2001 V&J
2006 Deep Learning
2007 PASCAL
2009 ImageNet
2012 AlexNet
2018 Turing Award

Artificial Intelligence Technology Can Better Our Lives