

# Deep Learning for Computer Vision

## Lecture 1: Introduction

# Deep Learning for Computer Vision

# Deep Learning for Computer Vision

Building artificial systems  
that process, perceive, and  
reason about visual data

# Computer Vision is everywhere!



Left to right: [Image](#) by [Roger H Gounis](#) licensed under CC BY 2.0  
[Image](#) CCO 1.0 public domain  
[Image](#) CCO 1.0 public domain  
[Image](#) CCO 1.0 public domain



Left to right: [Image](#) free to use  
[Image](#) CCO 1.0 public domain  
[Image](#) by NASA is licensed under CC BY 2.0  
[Image](#) is CCO 1.0 public domain



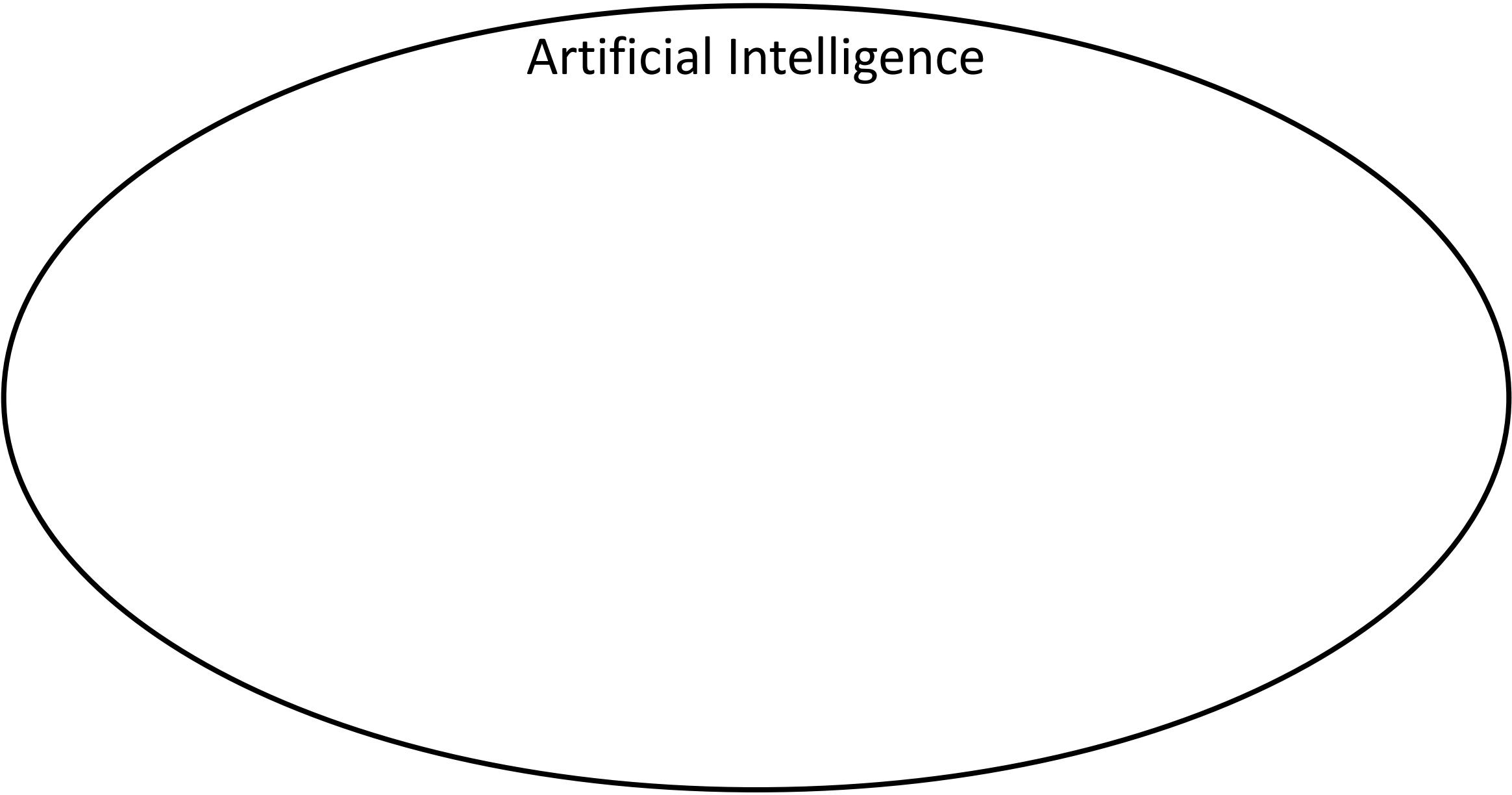
Bottom row, left to right  
[Image](#) CCO 1.0 public domain  
[Image](#) by Derek Keats is licensed under CC BY 2.0; changes made  
[Image](#) is public domain  
[Image](#) is licensed under CC-BY 2.0; changes made

# Deep Learning 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

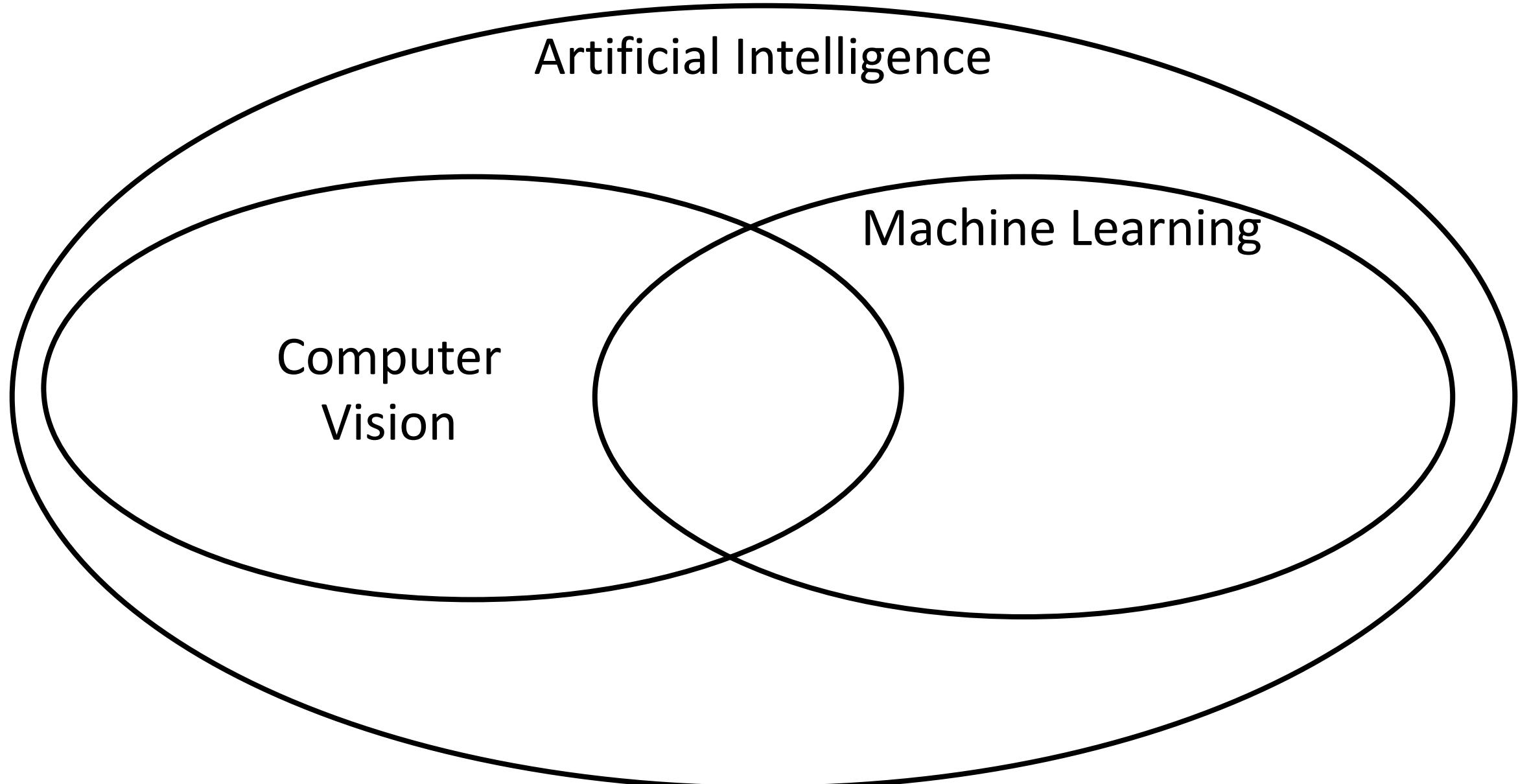


Artificial Intelligence

Artificial Intelligence

Machine Learning

Computer  
Vision



Artificial Intelligence

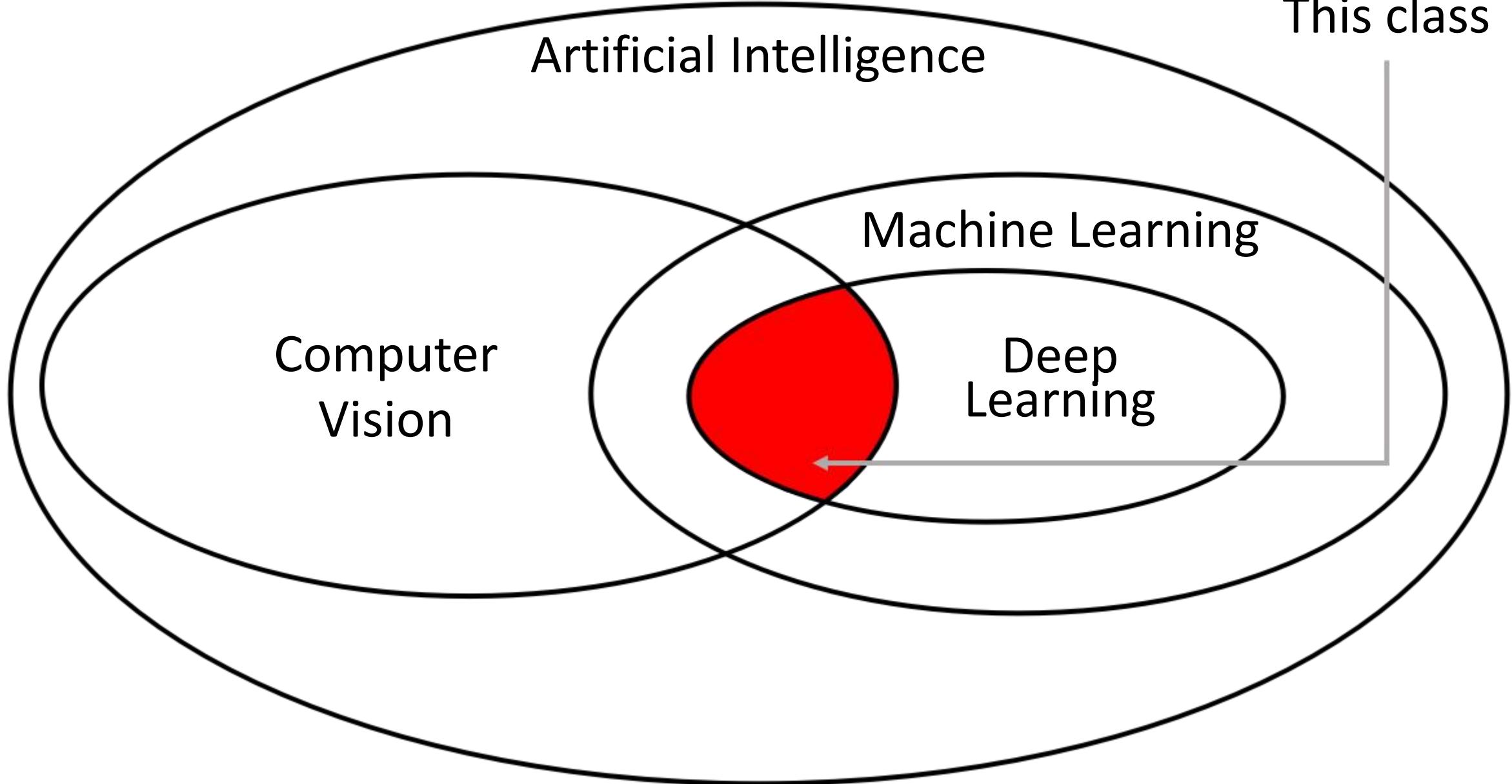
Computer  
Vision

Machine Learning

Deep  
Learning

Artificial Intelligence

This class



This class

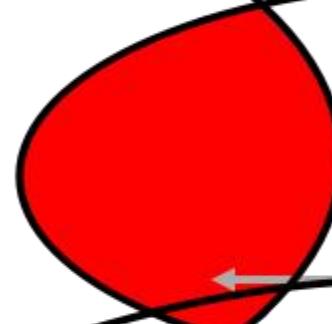
Artificial Intelligence

Machine Learning

Deep  
Learning

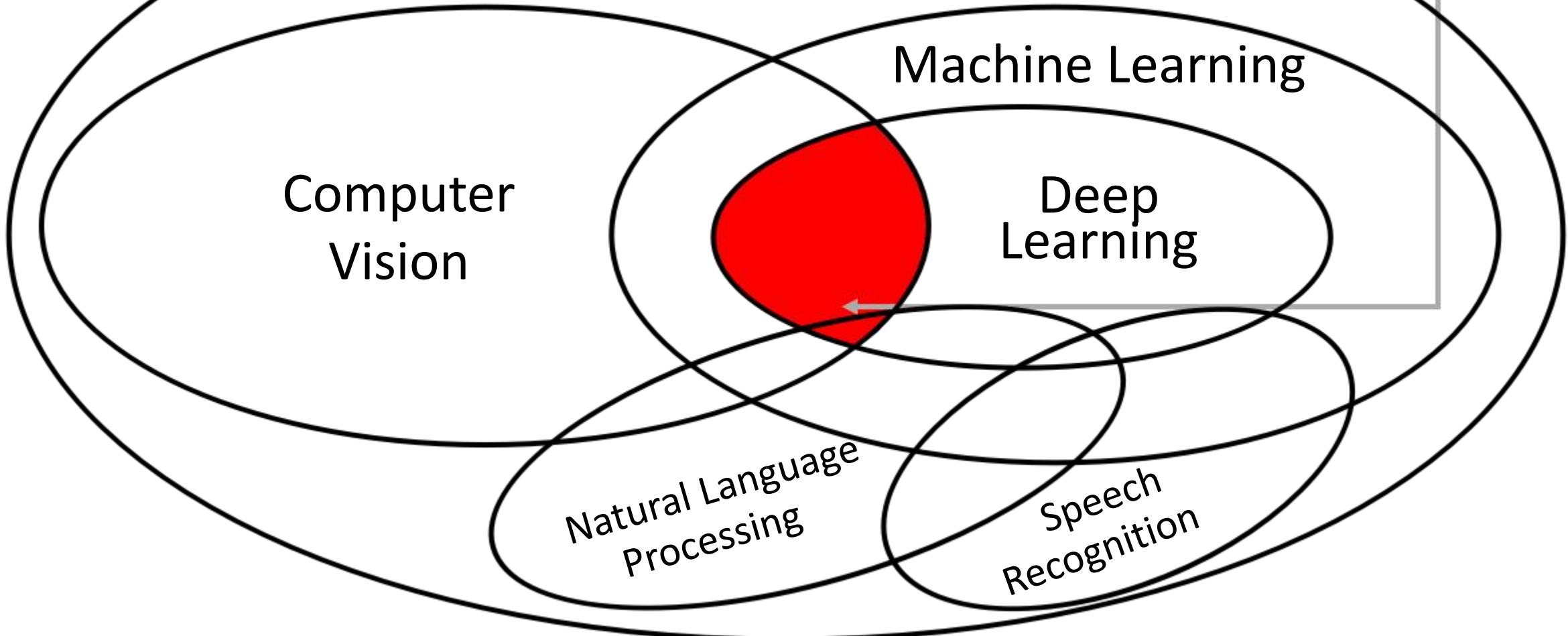
Computer  
Vision

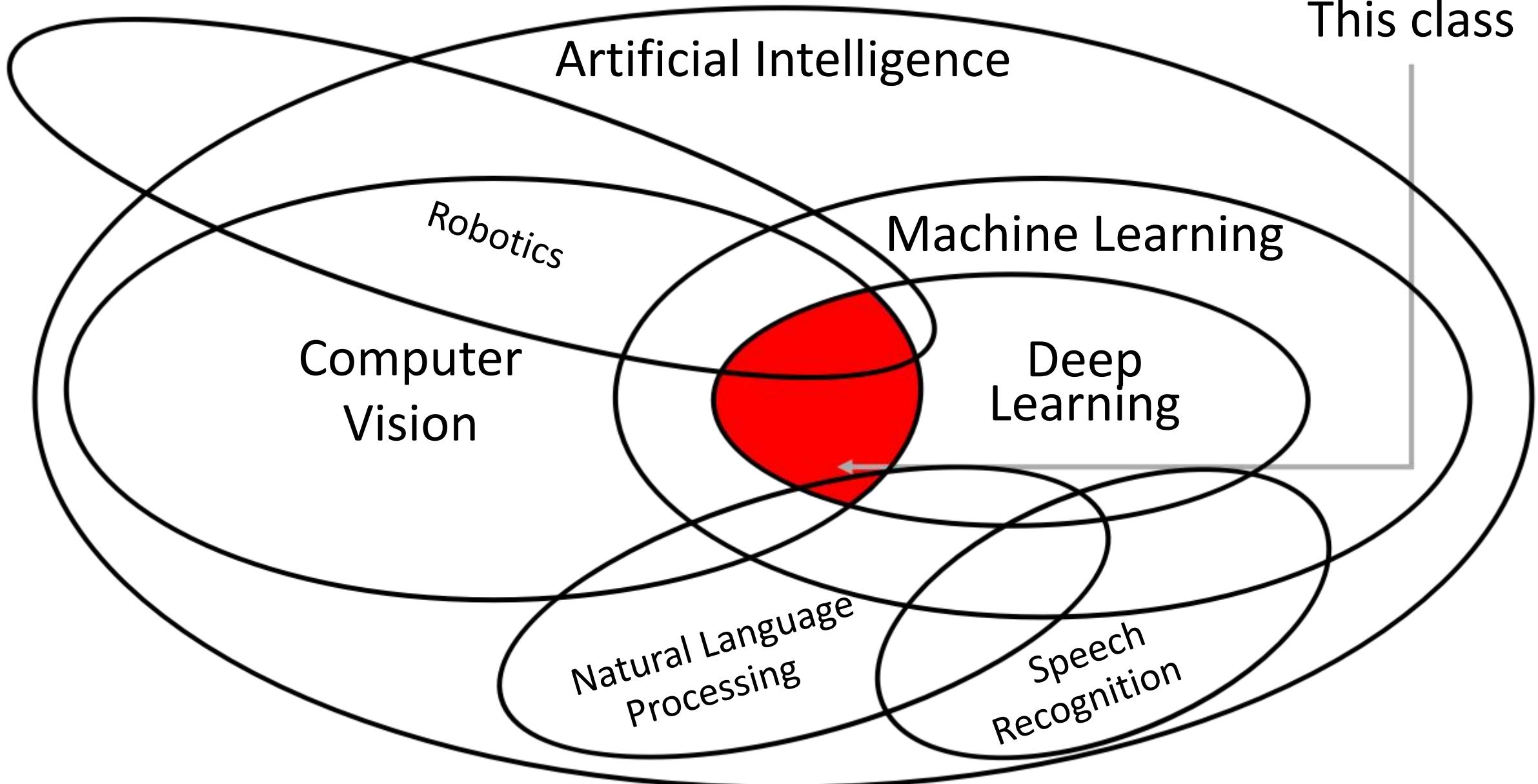
Natural Language  
Processing



Artificial Intelligence

This class



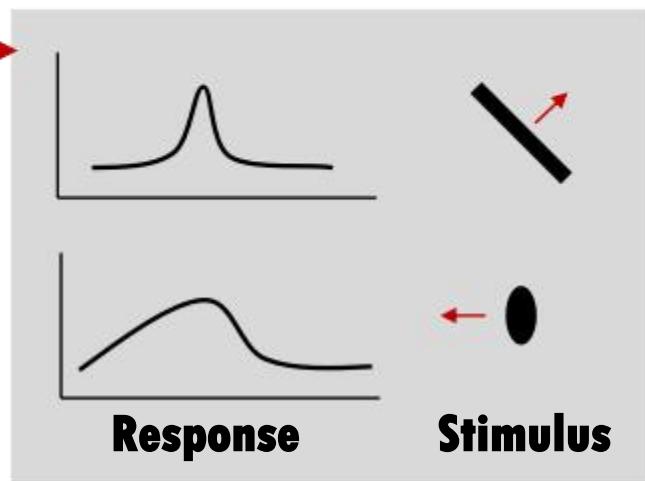
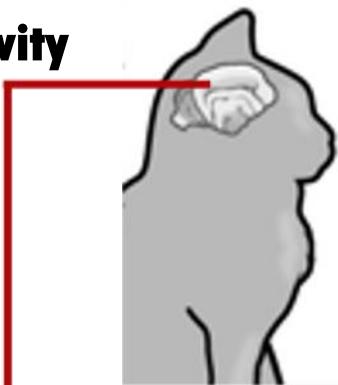


# Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics

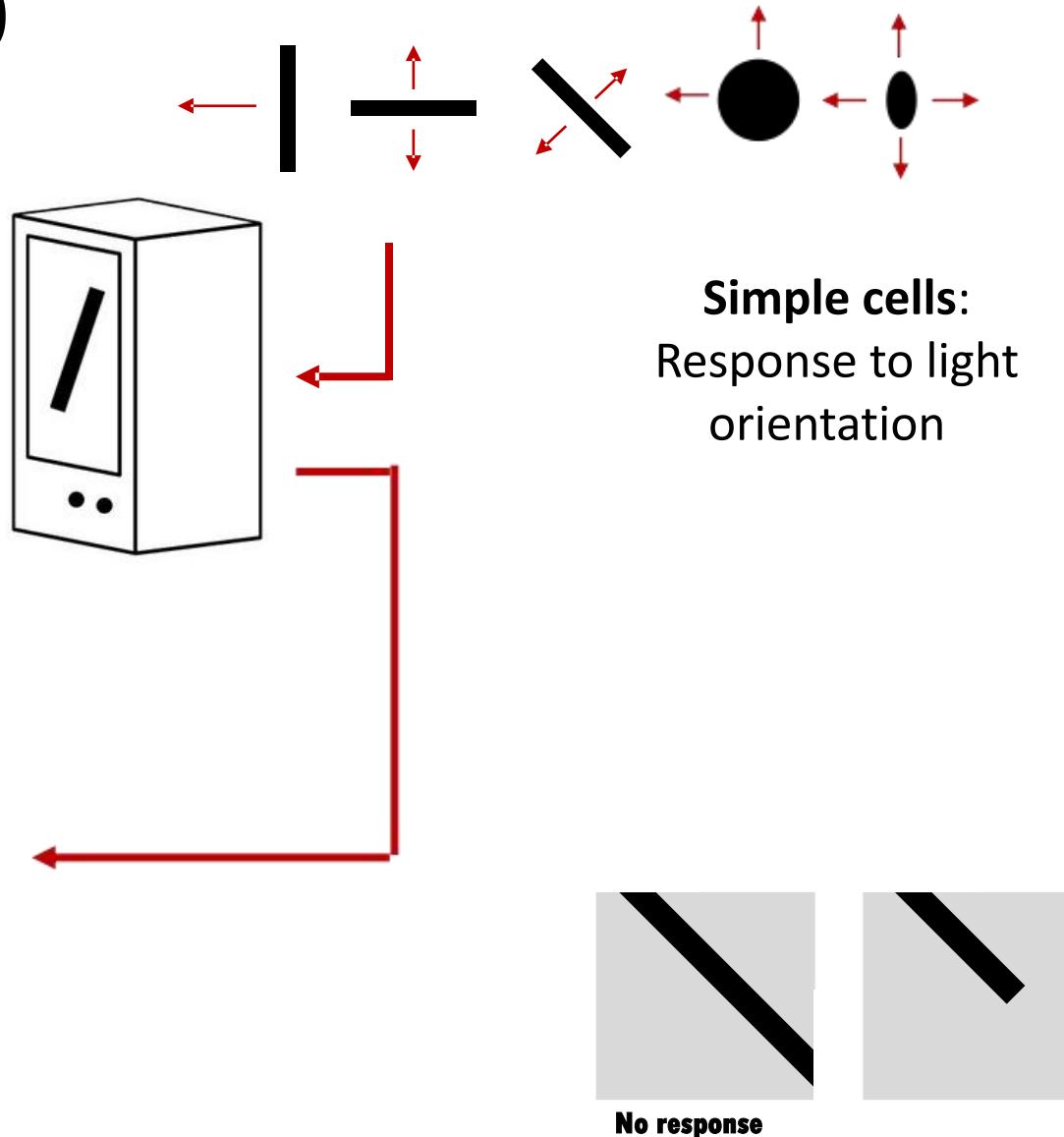
# Hubel and Wiesel, 1959

**Measure brain activity**



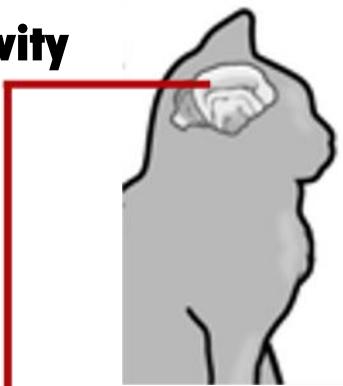
Credit image by CNX OpenStax is licensed under CC BY 4.0; changes made

1959 Hubel & Wiesel



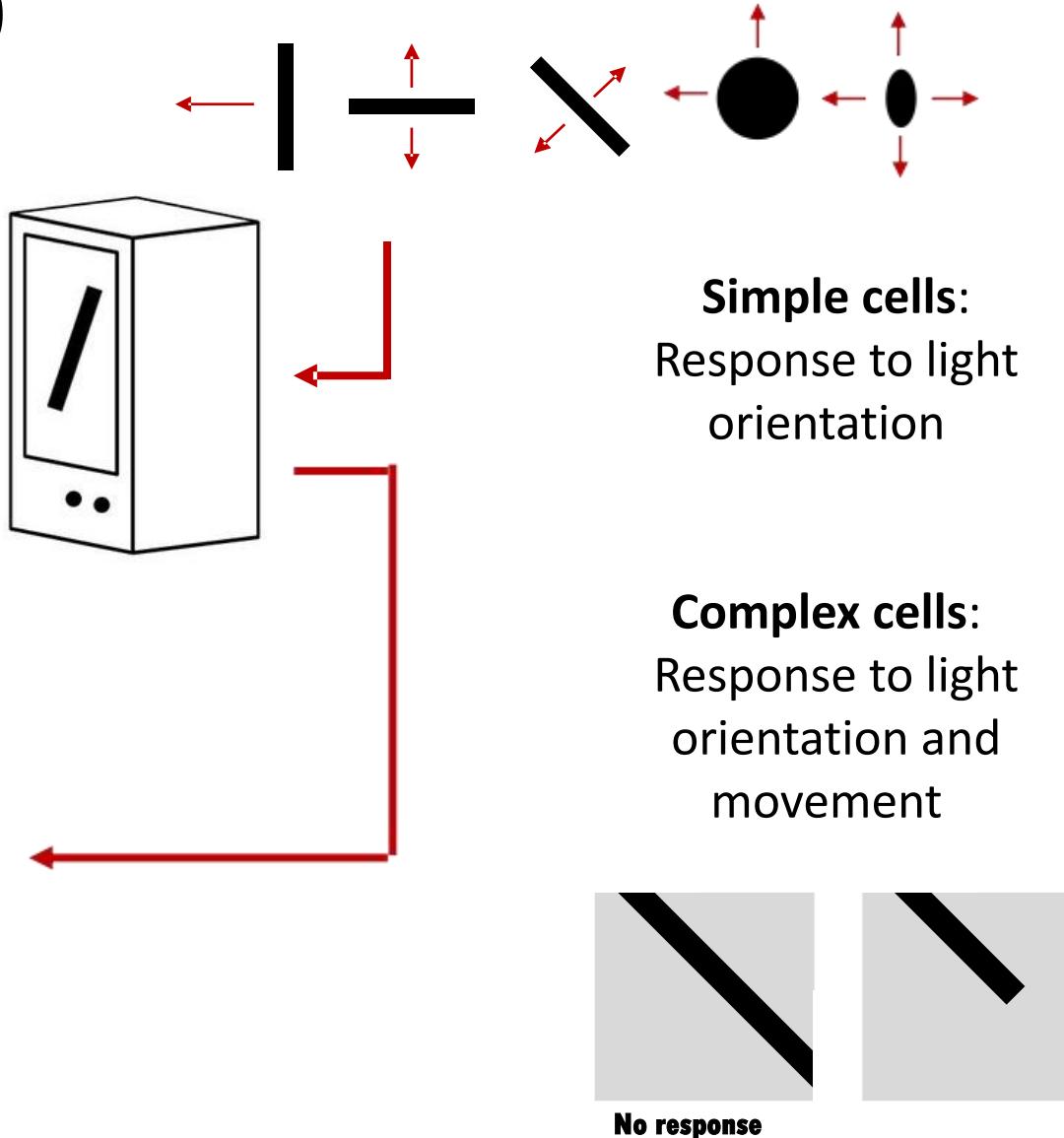
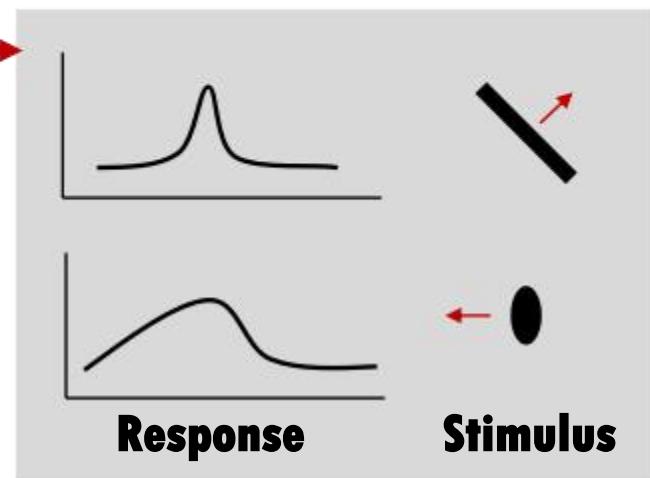
# Hubel and Wiesel, 1959

**Measure brain activity**

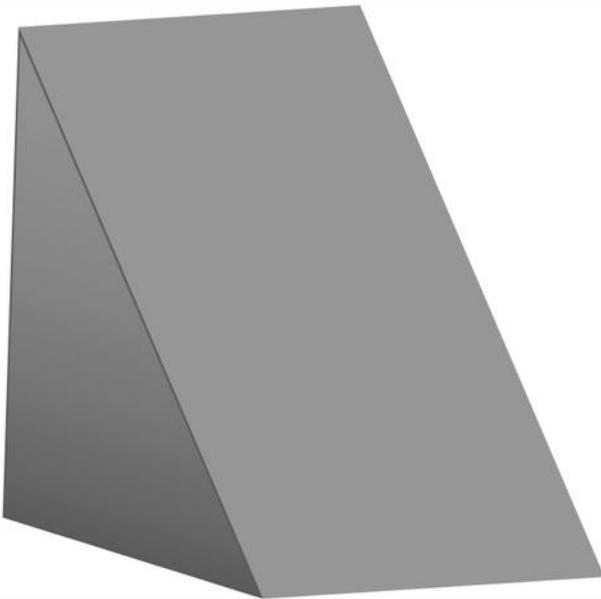


Cai image by CNX OpenStax is licensed under CC BY 4.0; changes made

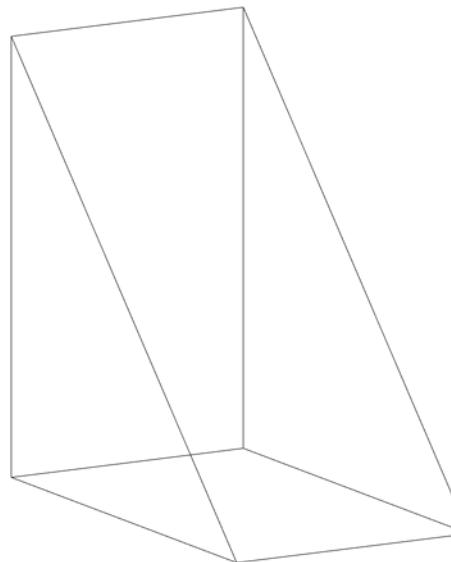
1959 Hubel & Wiesel



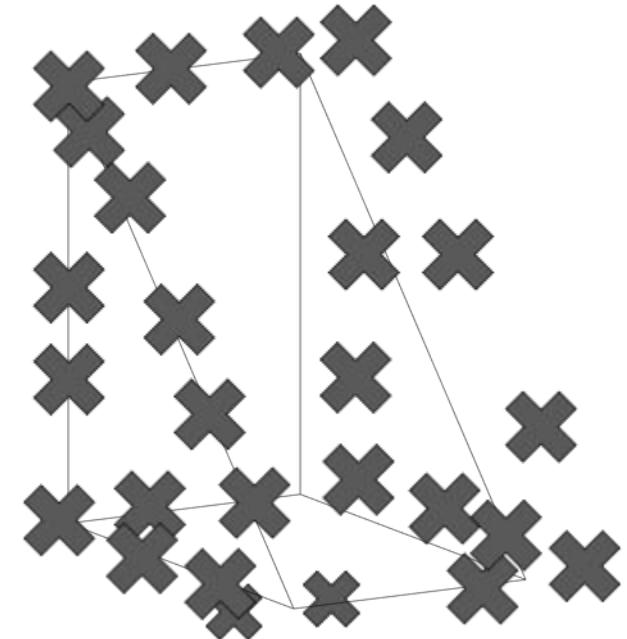
# Larry Roberts, 1963



**(a) Original picture**



**(b) Differentiated picture**



**(c) Feature points selected**

1959 Hubel &  
Wiesel

1963  
Roberts

Lawrence Gilman Roberts, "Machine Perception of Three-Dimensional Solids", 1963

Artificial Intelligence Group  
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

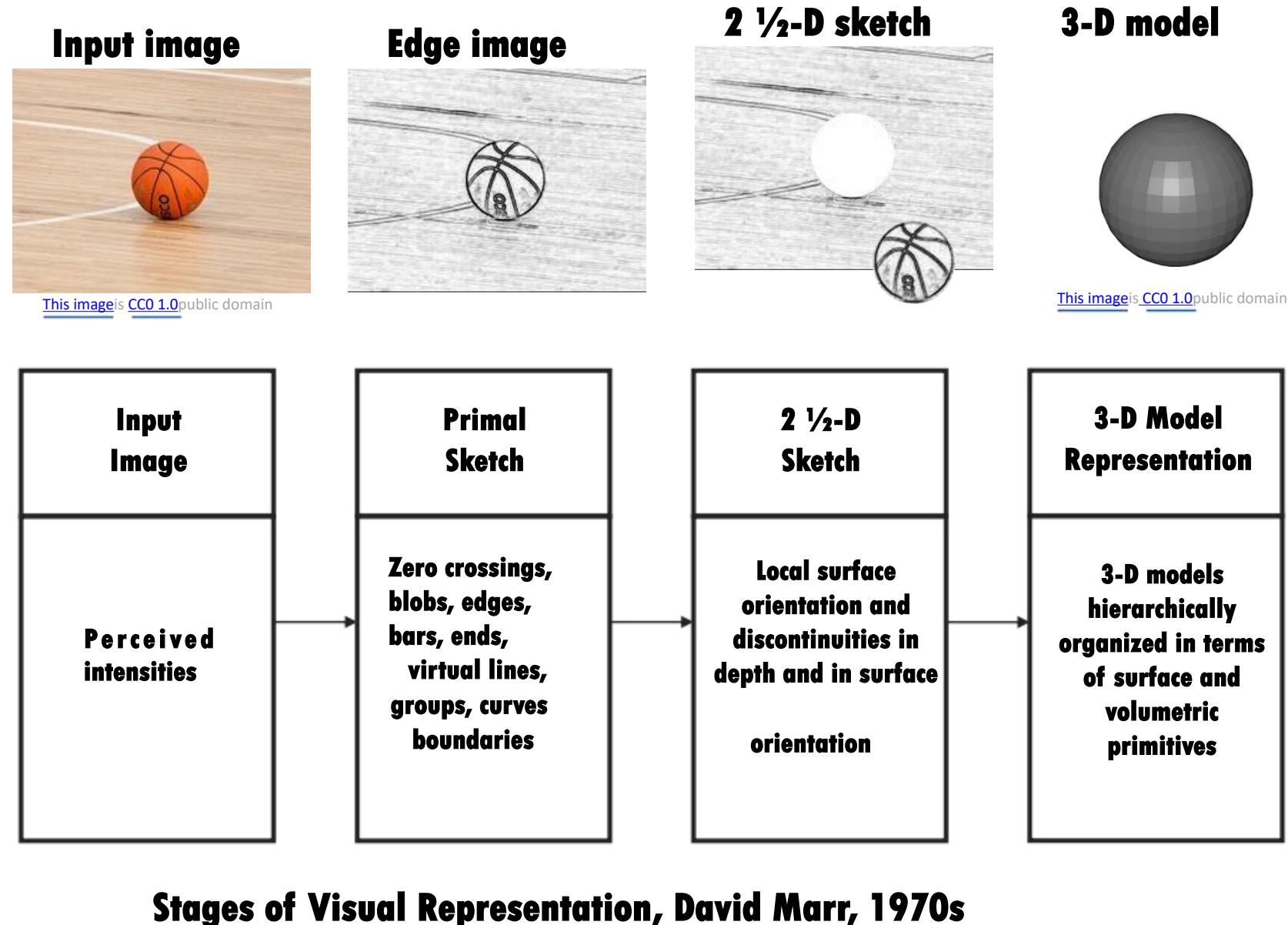
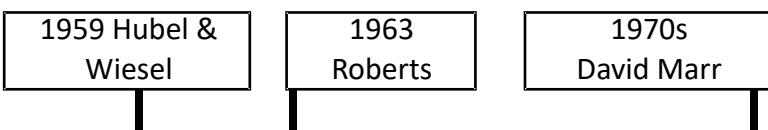
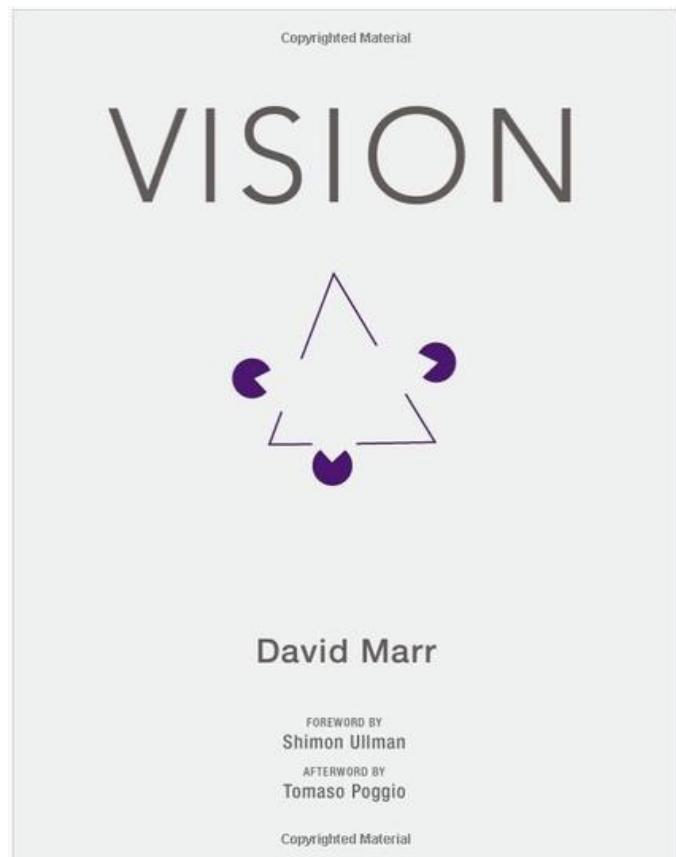
Seymour Papert

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".

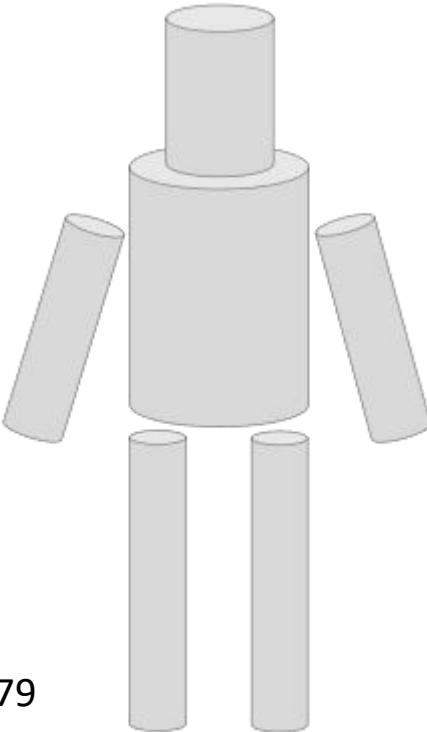
1959 Hubel &  
Wiesel

1963  
Roberts

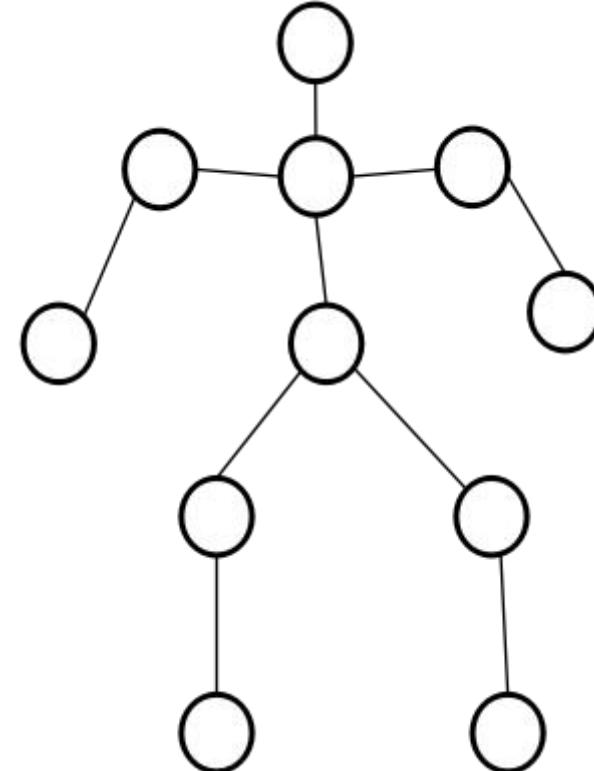
<https://dspace.mit.edu/handle/1721.1/6125>



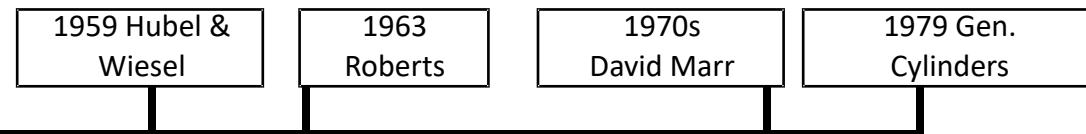
# Recognition via Parts (1970s)



Generalized Cylinders,  
Brooks and Binford, 1979



Pictorial Structures,  
Fischler and Elshlager, 1973



# Recognition via Edge Detection (1980s)



1959 Hubel &  
Wiesel

1963  
Roberts

1970s  
David Marr

1979 Gen.  
Cylinders

1986  
Canny

John Canny, 1986  
David Lowe, 1987

# Recognition via Grouping (1990s)



1959 Hubel &  
Wiesel

1963  
Roberts

1970s  
David Marr

1979 Gen.  
Cylinders

1986  
Canny

1997 Norm.  
Cuts

AI Winter

Normalized Cuts, Shi and Malik, 1997

# Recognition via Matching (2000s)



[Image](#)is public domain



[Image](#)is public domain

1959 Hubel &  
Wiesel

1963  
Roberts

1970s  
David Marr

1979 Gen.  
Cylinders

1986  
Canny

1997 Norm.  
Cuts

1999  
SIFT

AI Winter

**SIFT, David  
Lowe, 1999**

# 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

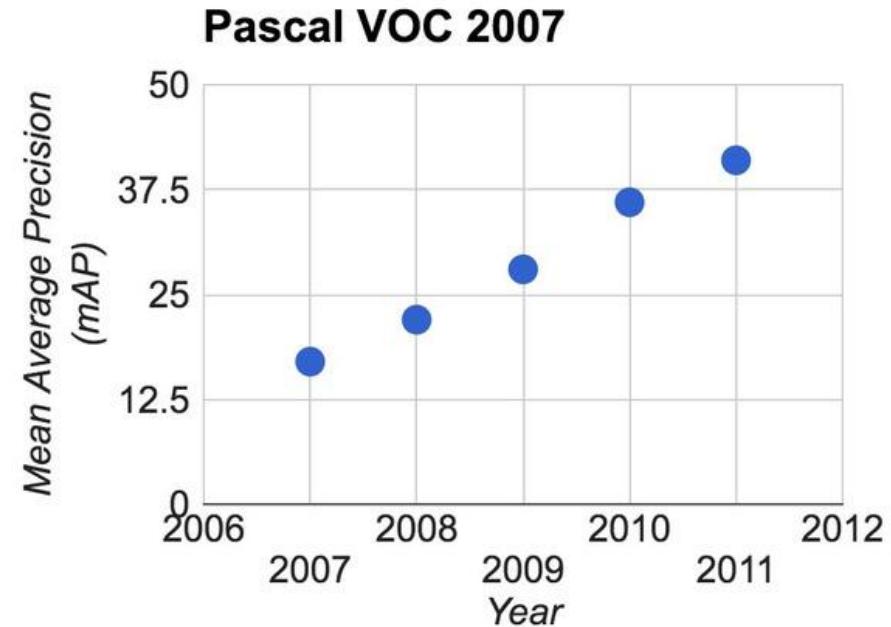
AI Winter

# PASCAL Visual Object Challenge

[Image is CC0 1.0](#) public domain



[Image is CC0 1.0](#) public domain



1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

1999 SIFT

2001 V&J

2007 PASCAL

AI Winter

# IMAGENET Large Scale Visual Recognition Challenge

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



Output:  
Scale T-shirt  
Steel drum  
Drumstick  
Mud turtle

Deng et al, 2009 Russakovsky et al. IJCV 2015

1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

1999 SIFT

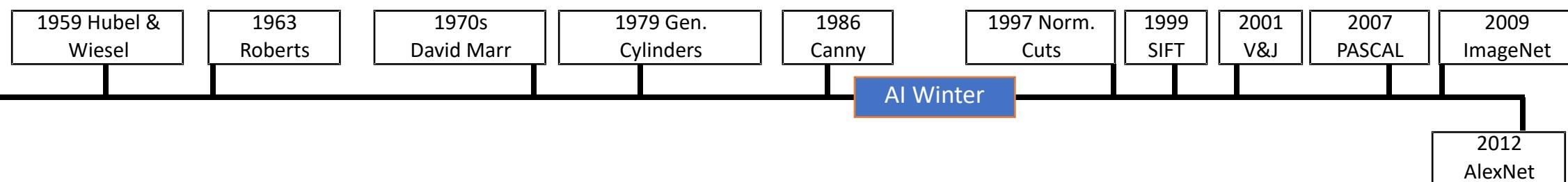
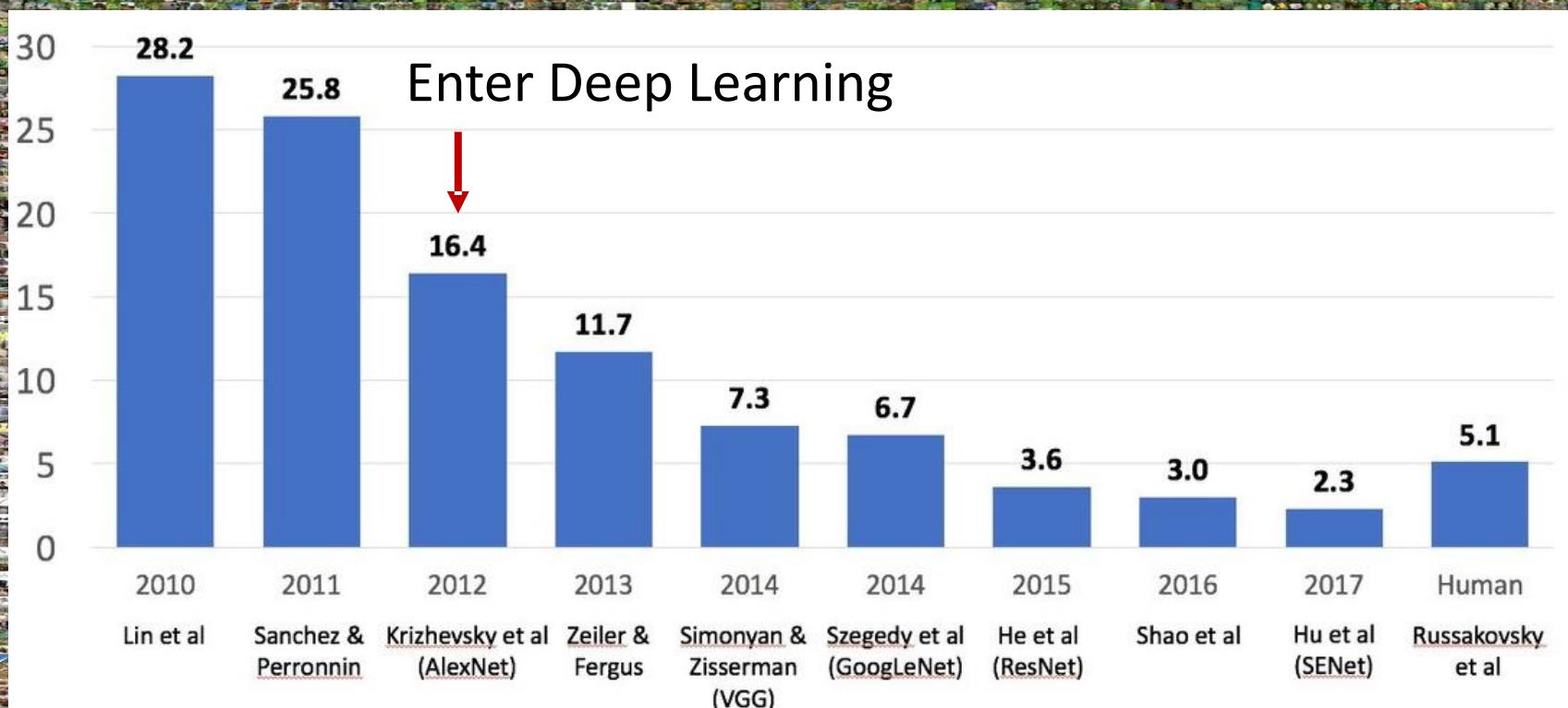
2001 V&J

2007 PASCAL

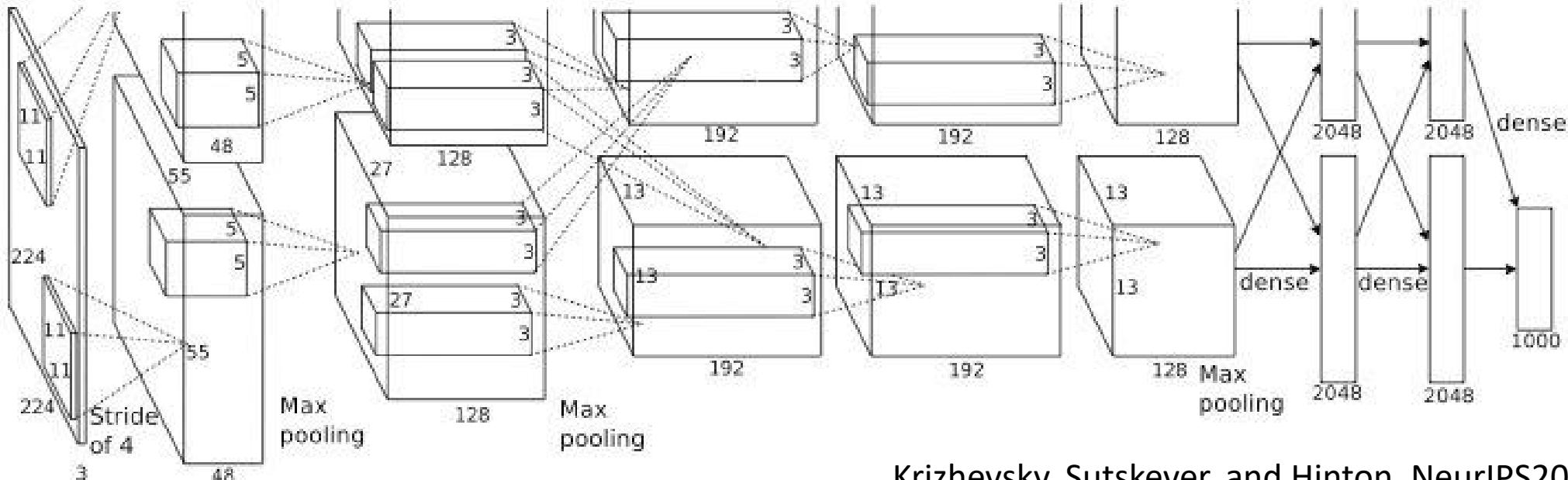
2009 ImageNet

AI Winter

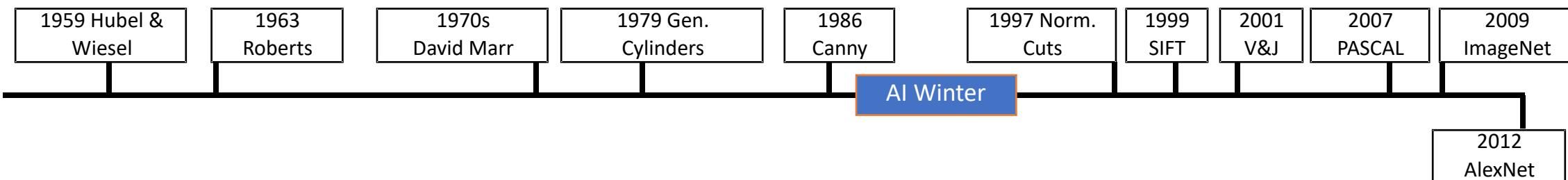
# IMAGENET Large Scale Visual Recognition Challenge



# AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS2012



# Perceptron

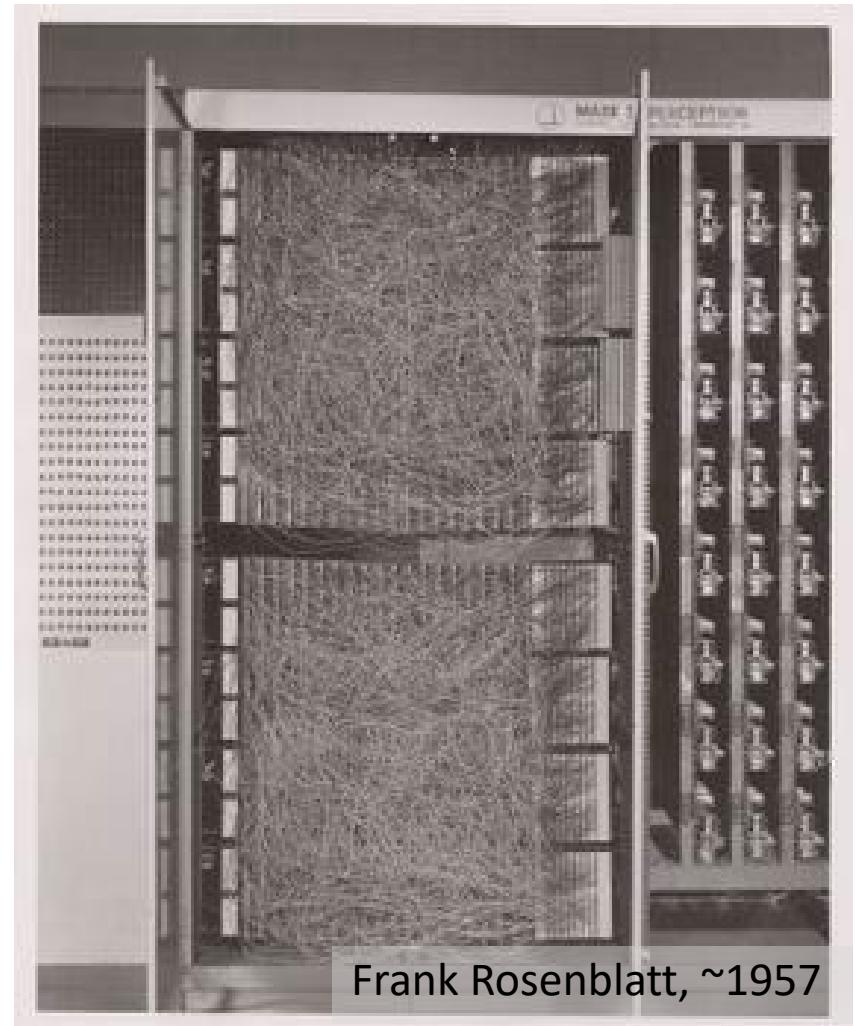
One of the earliest algorithms that could learn from data

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

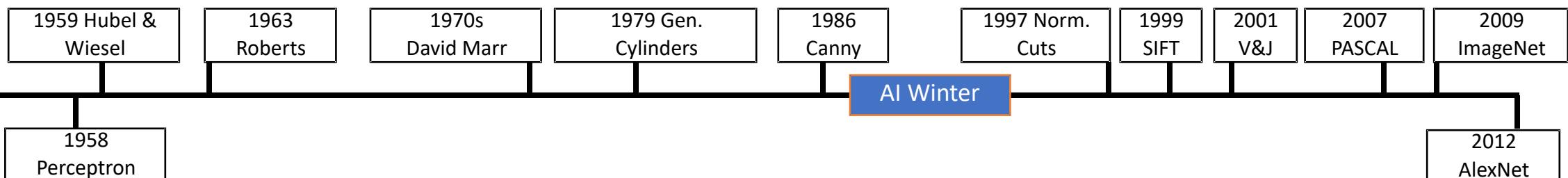
Connected to a camera that used 20x20 cadmium sulfide photocells to make a 400-pixel image

Could learn to recognize letters of the alphabet Today we

would recognize it as a **linear classifier**

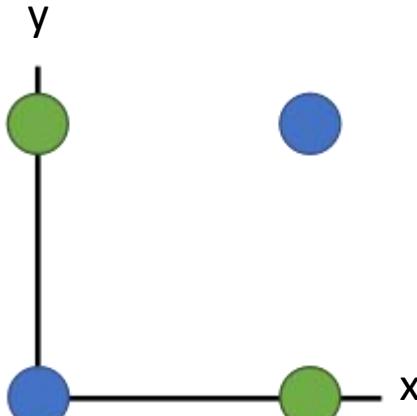


Frank Rosenblatt, ~1957

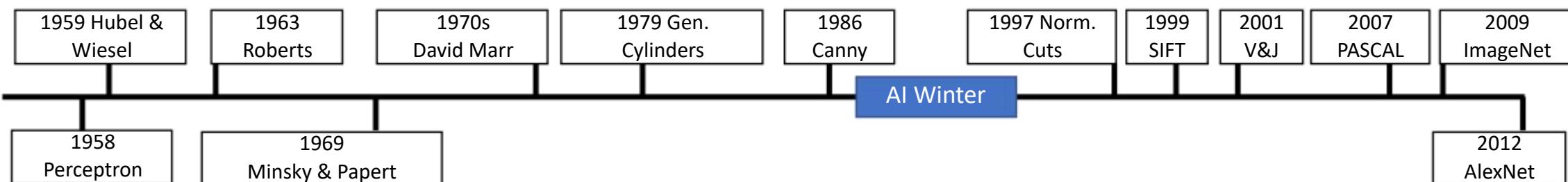
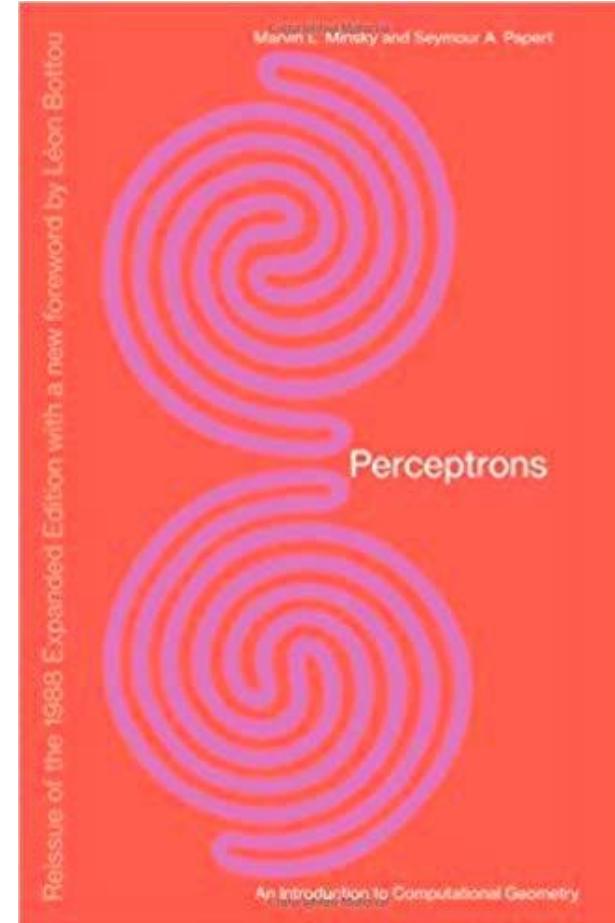


# Minsky and Papert, 1969

X	Y	F(x,y)
0	0	0 1 1
0	1	0
1	0	
1	1	



Showed that Perceptrons could not learn the XOR function  
Caused a lot of disillusionment in the field

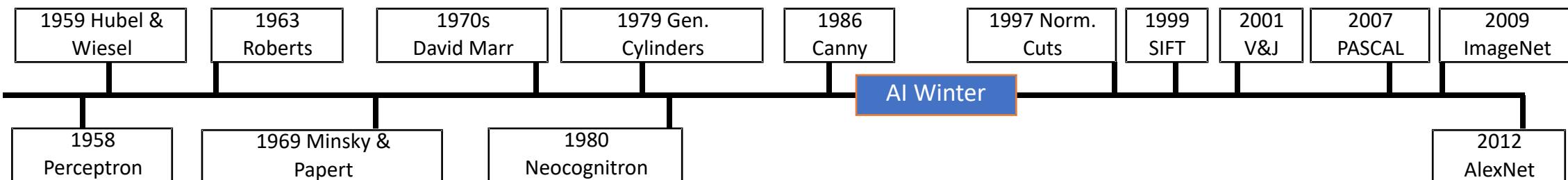
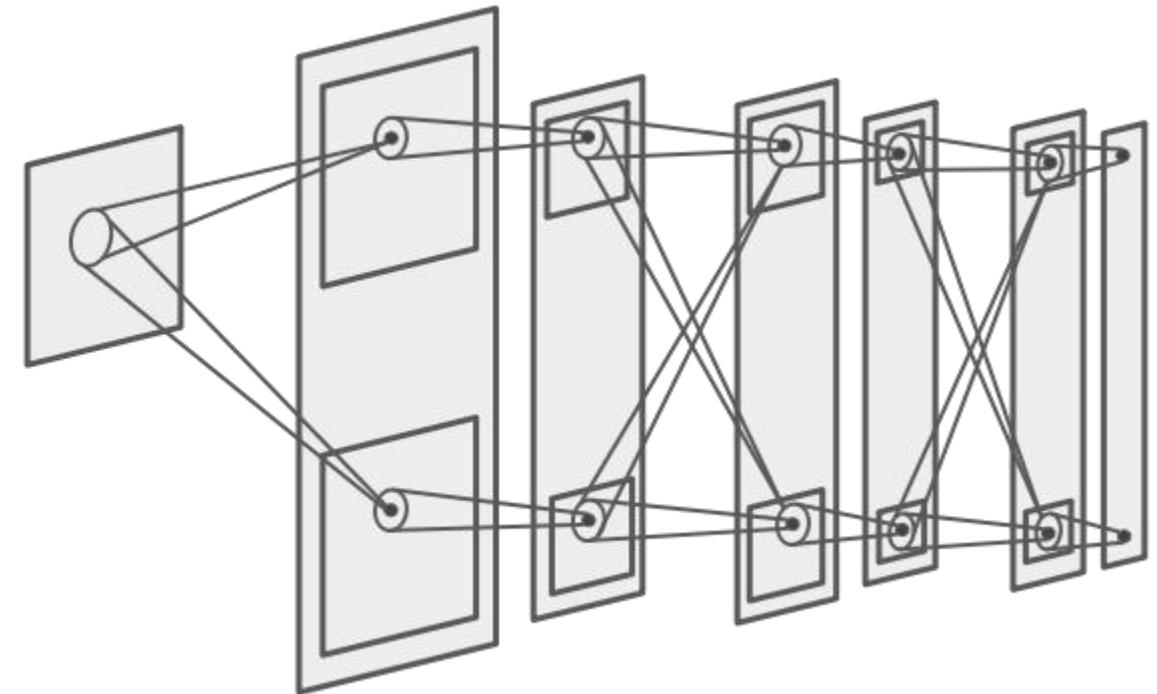


# Neocognitron: Fukushima, 1980

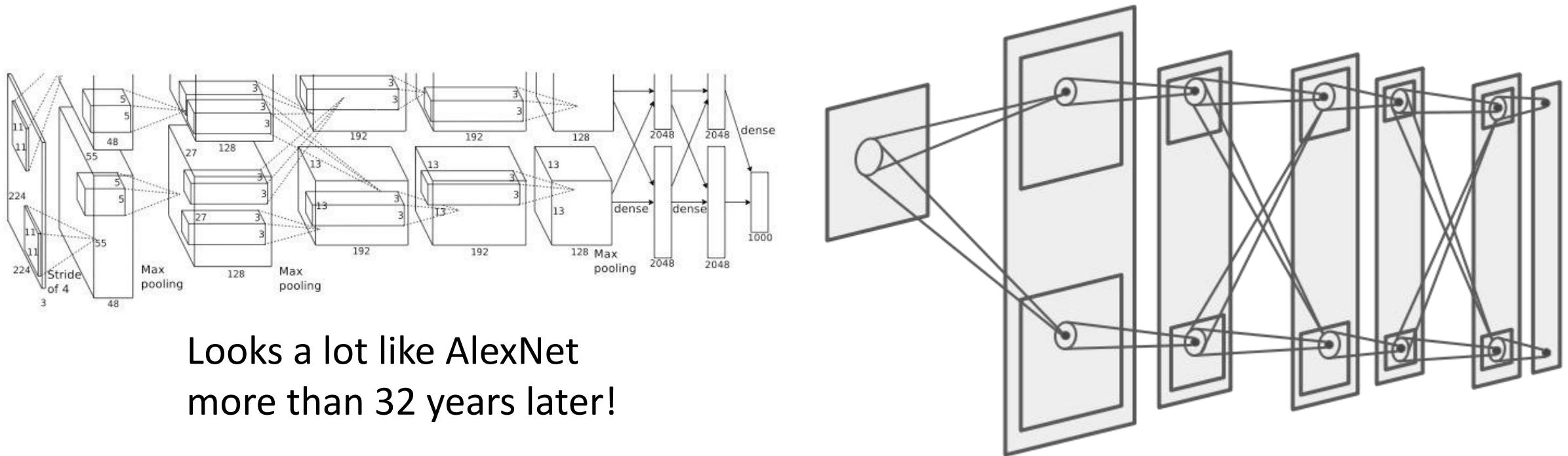
Computational model the visual system,  
directly inspired by Hubel and Wiesel's  
hierarchy of complex and simple cells

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

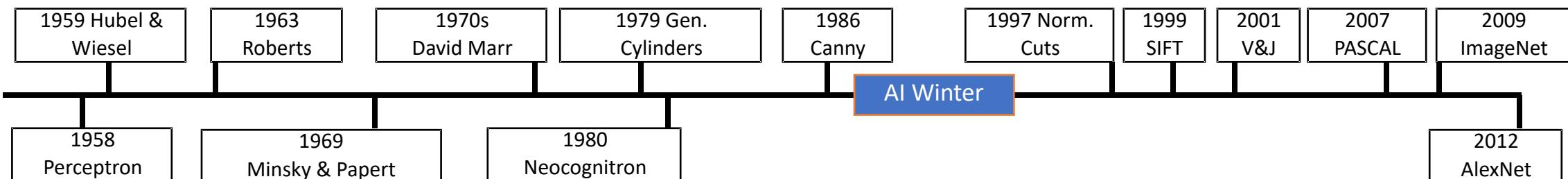
No practical training algorithm



# Neocognitron: Fukushima, 1980



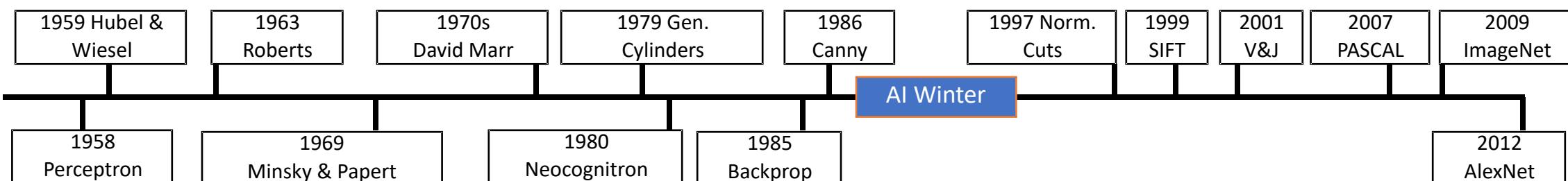
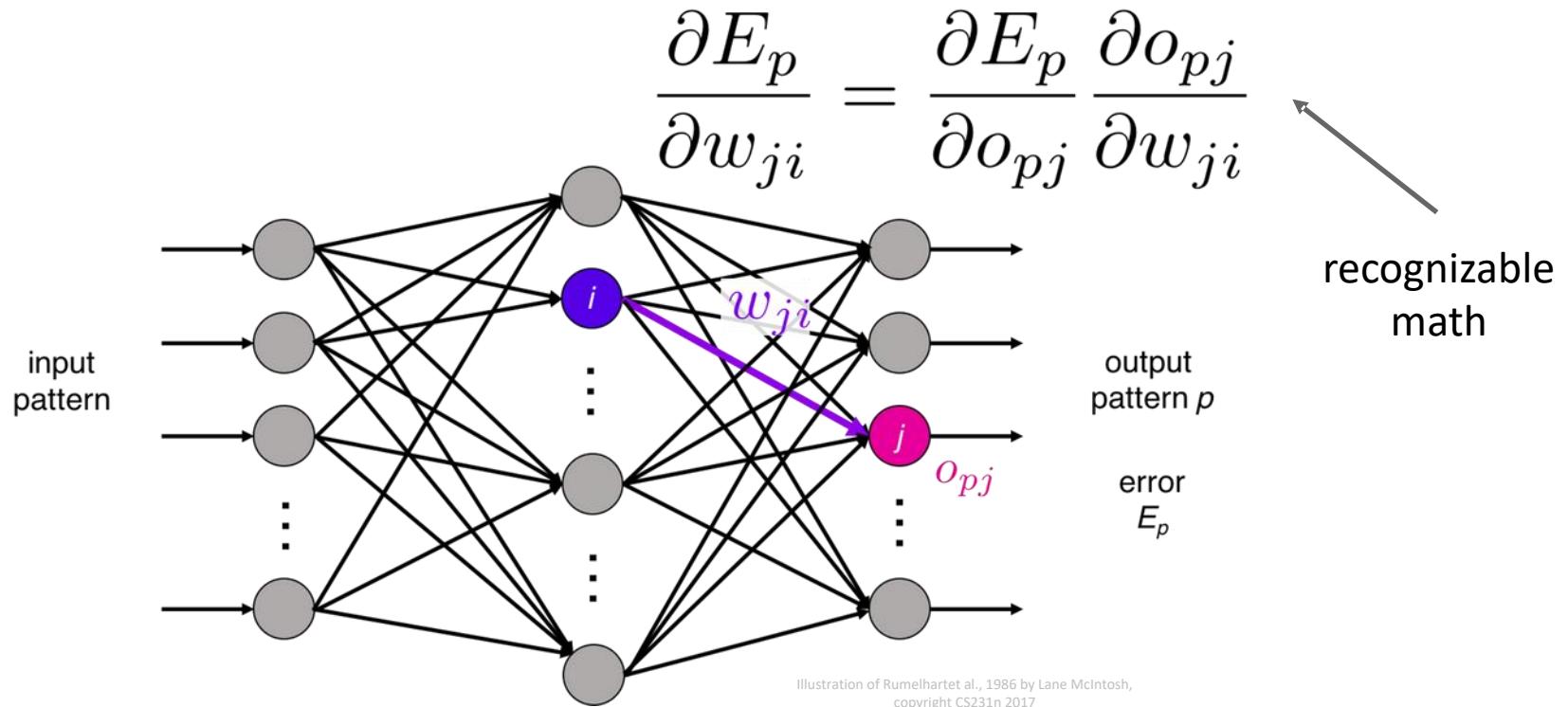
Looks a lot like AlexNet  
more than 32 years later!



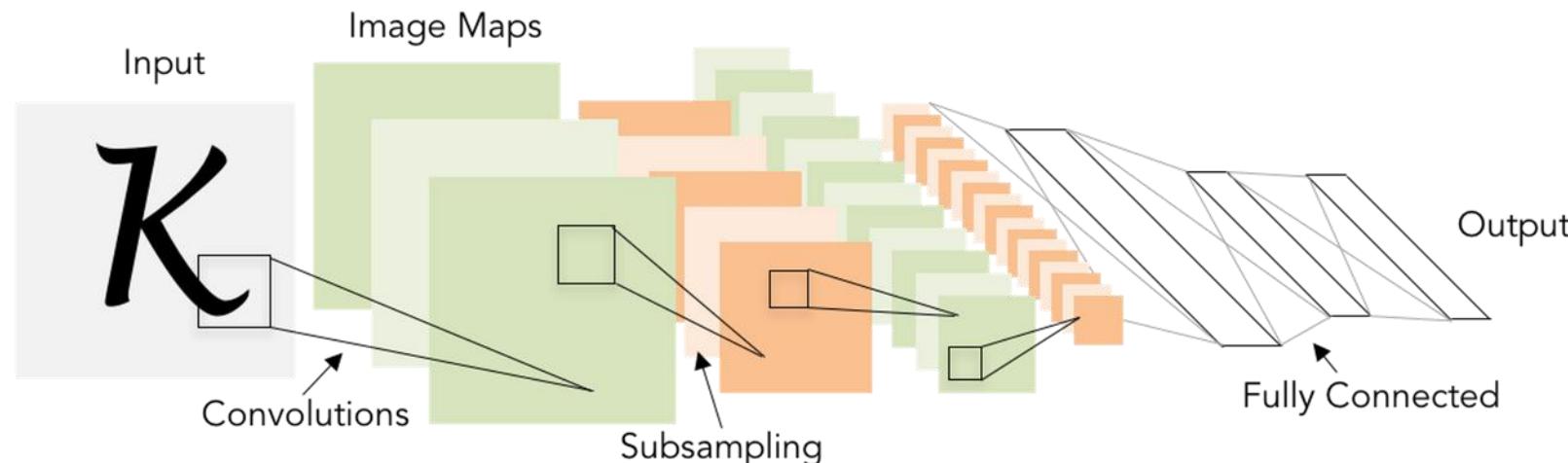
# Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

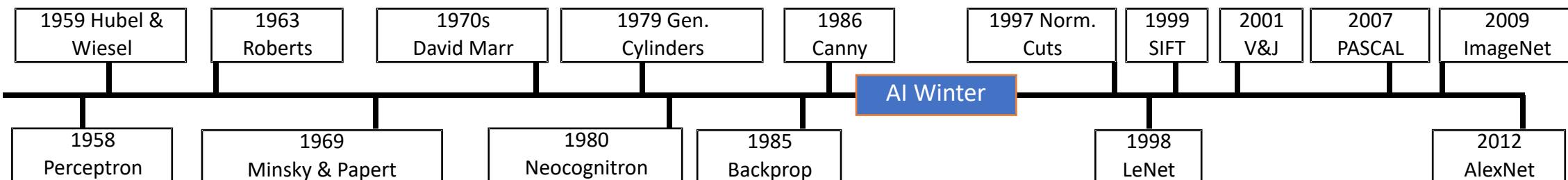
Successfully trained perceptrons with multiple layers



# Convolutional Networks: LeCun et al, 1998



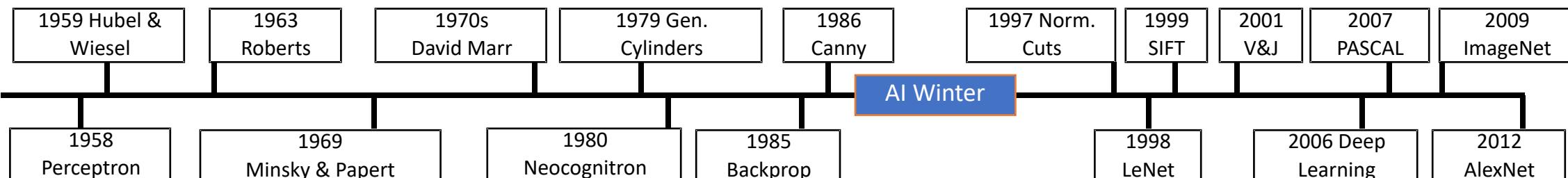
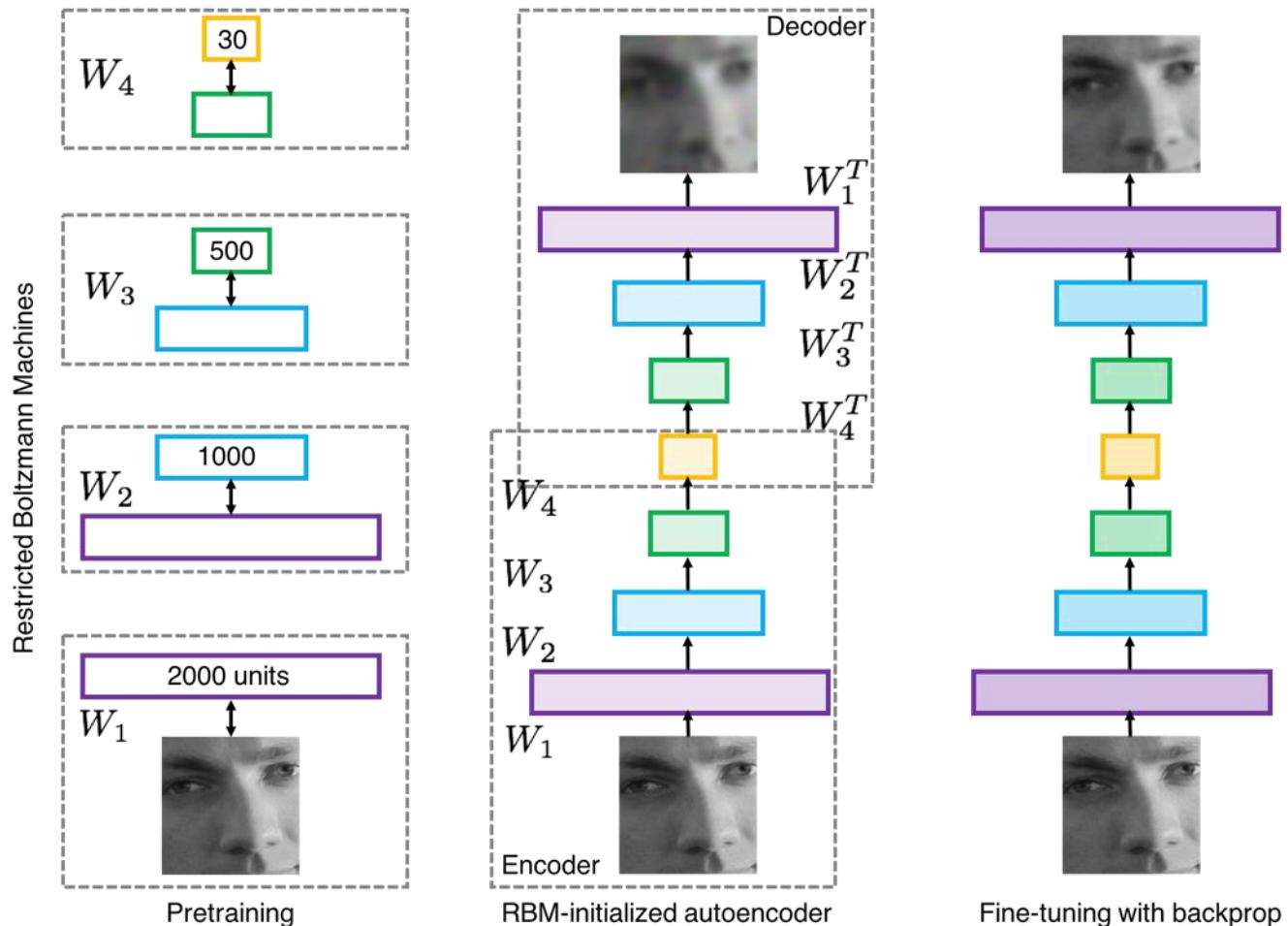
Applied backprop algorithm to a Neocognitron-like architecture Learned to recognize handwritten digits Was deployed in a commercial system by NEC, processed handwritten checks Very similar to our modern convolutional networks!



# 2000s: “Deep Learning”

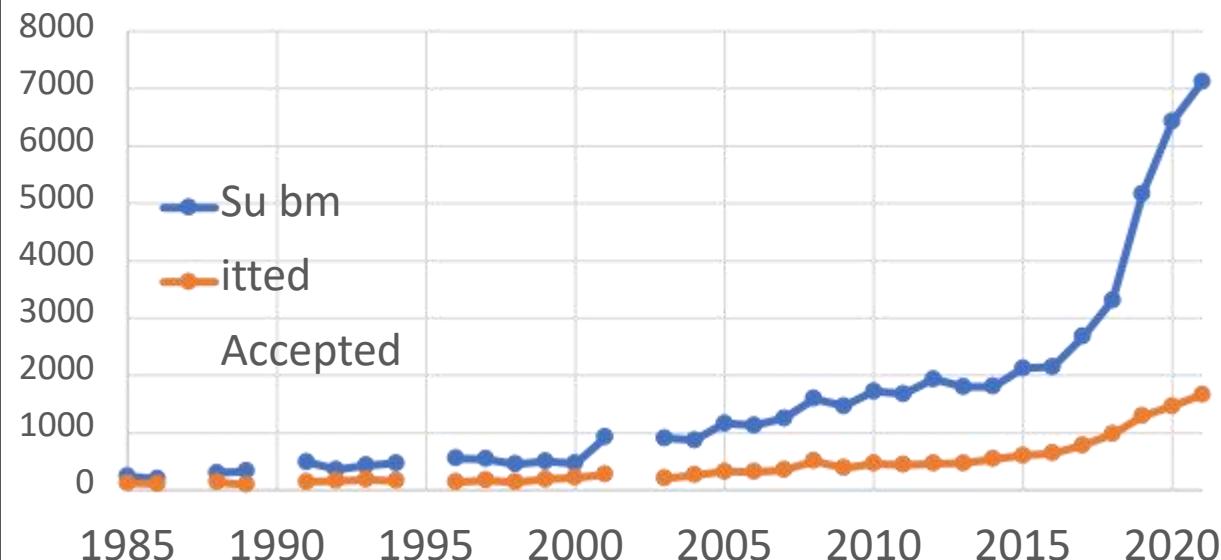
People tried to train neural networks that were deeper and deeper Not a mainstream research topic at this time

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

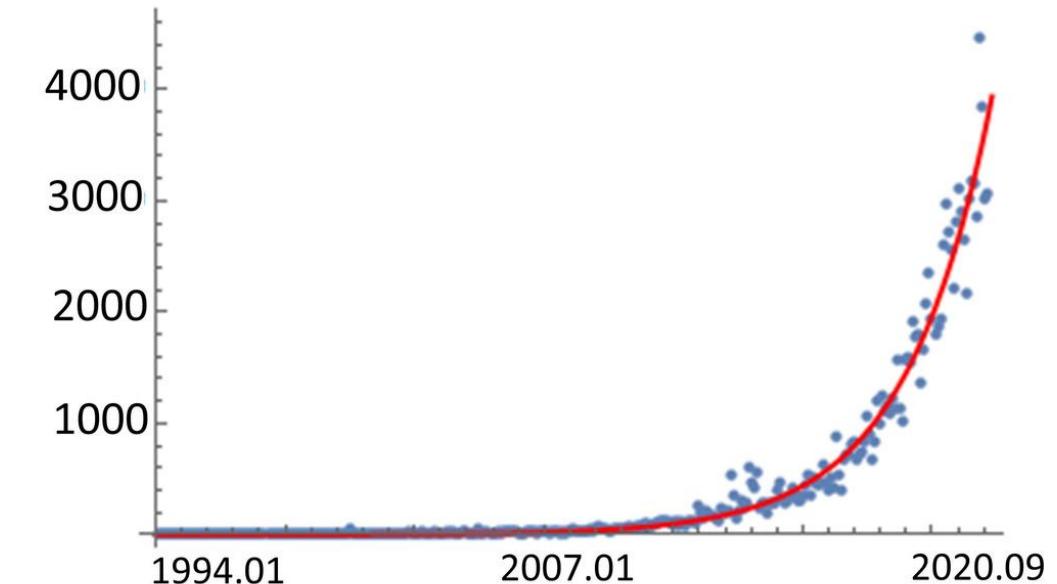


# 2012 to Present: Deep Learning Explosion

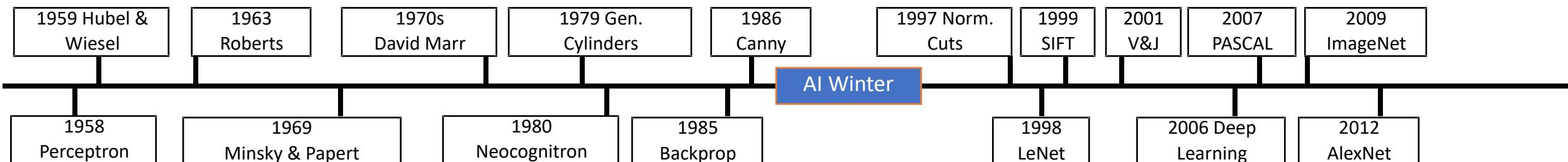
CVPR Papers



ML+AI arXiv papers per month



Publications at top Computer Vision conference



# 2012 to Present: Deep Learning is Everywhere

Image Classification

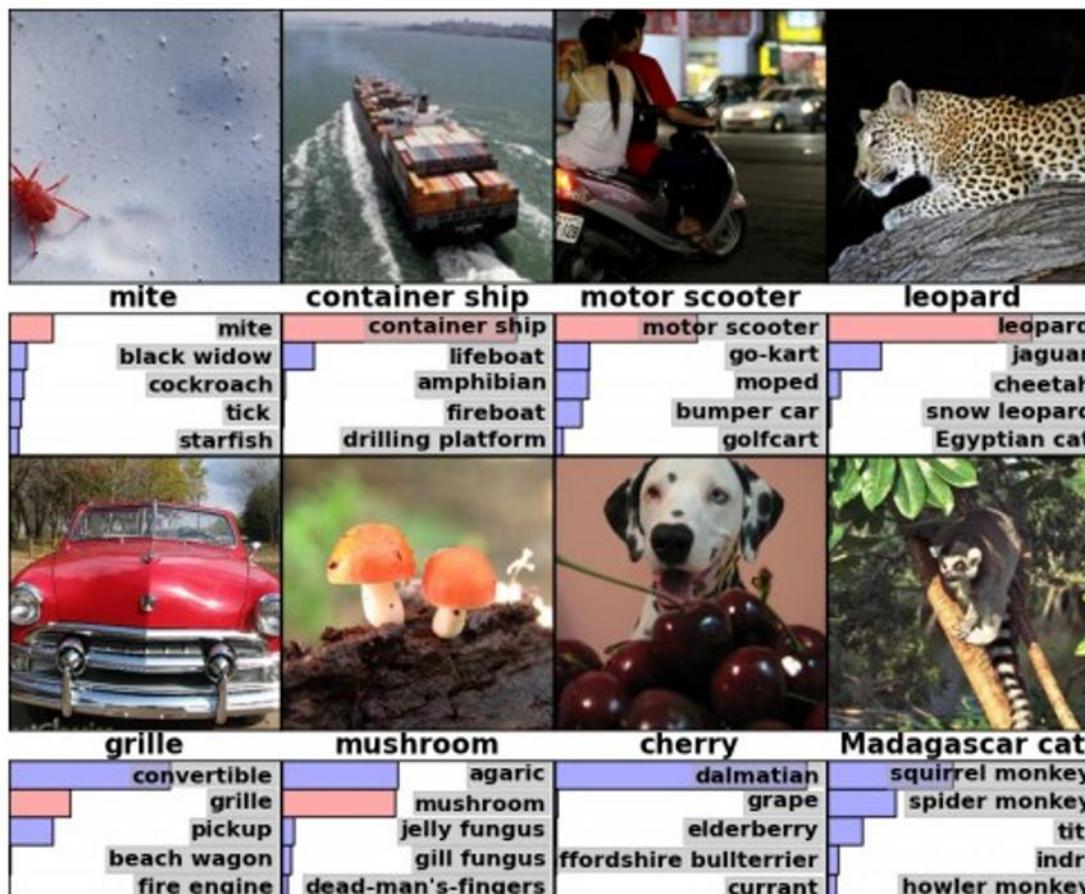
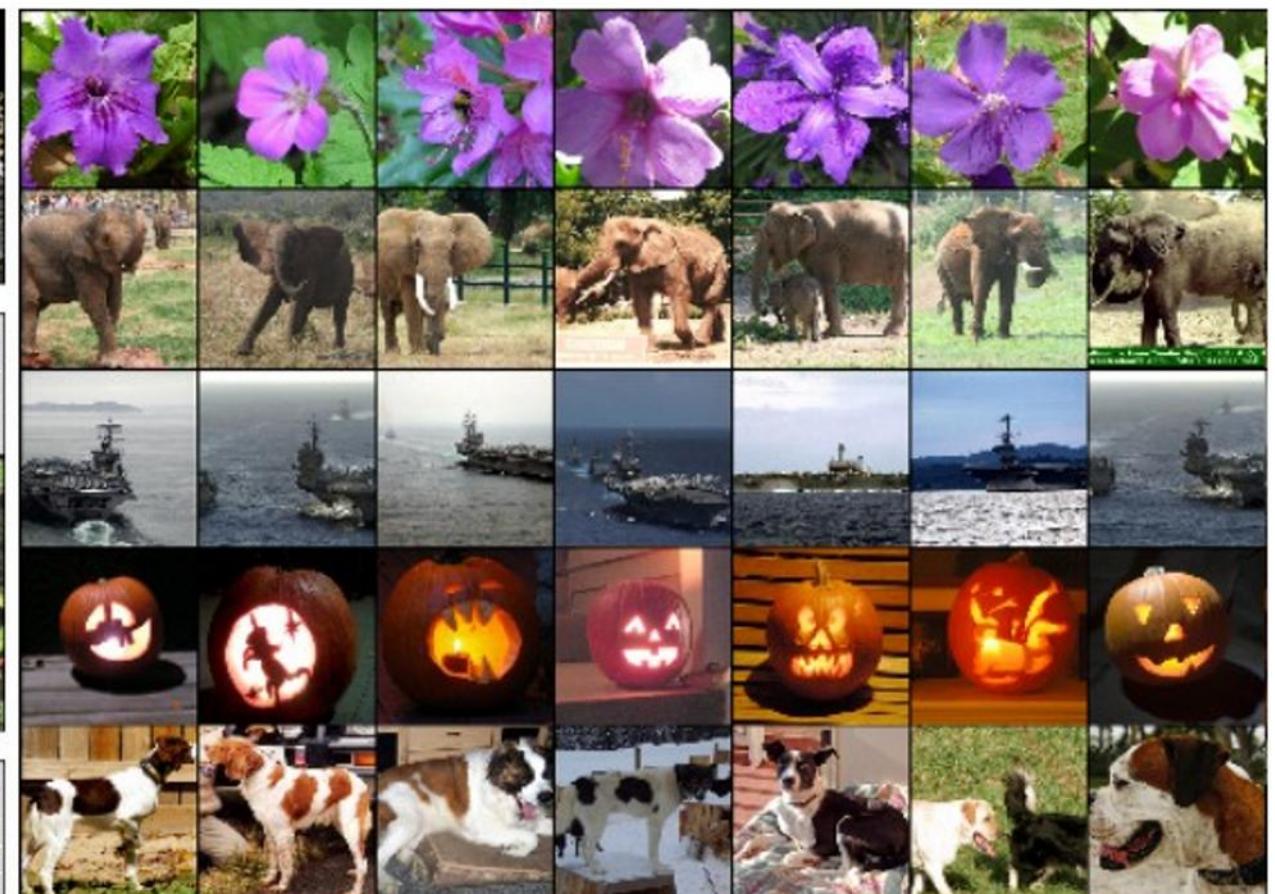


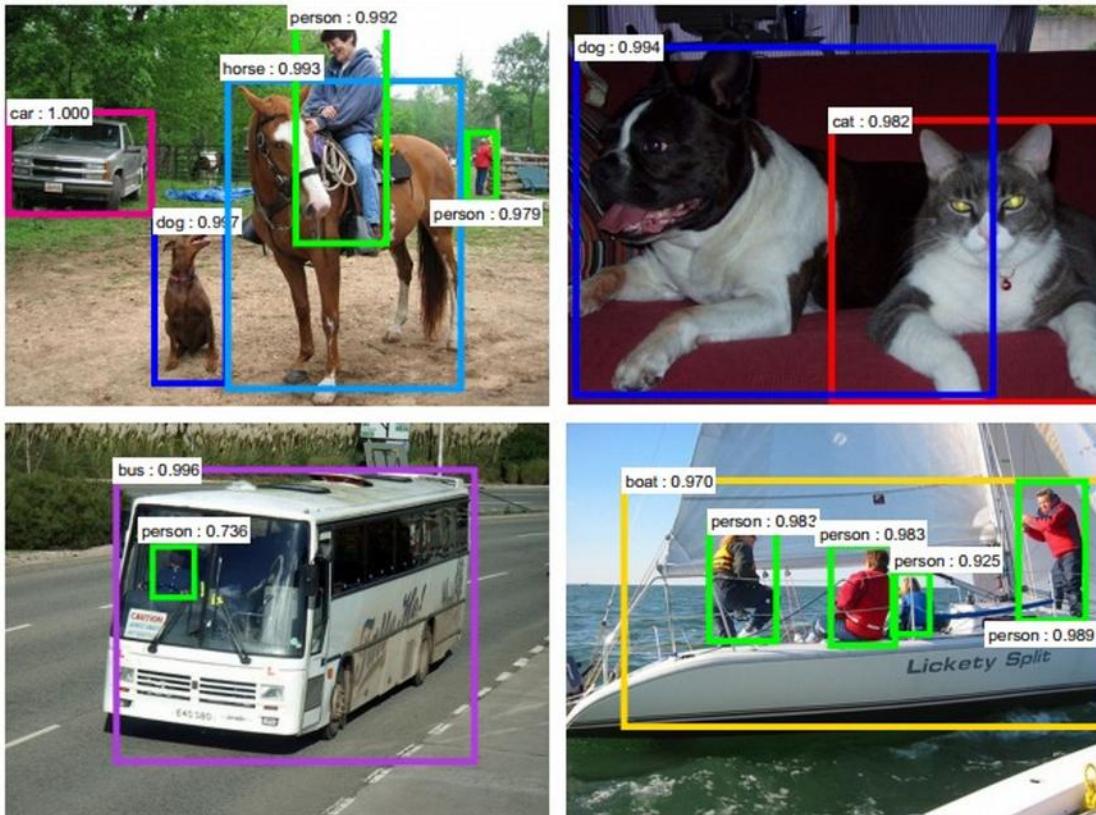
Image Retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

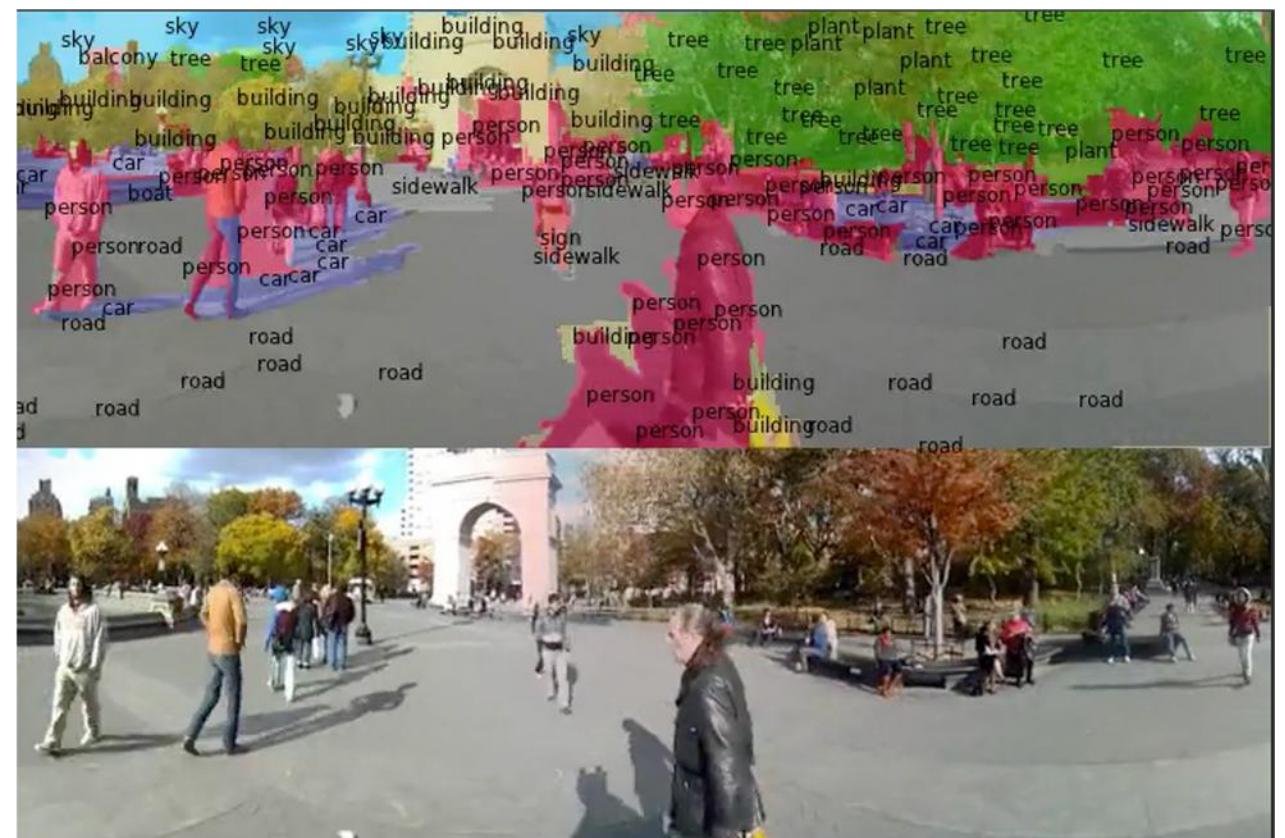
# 2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

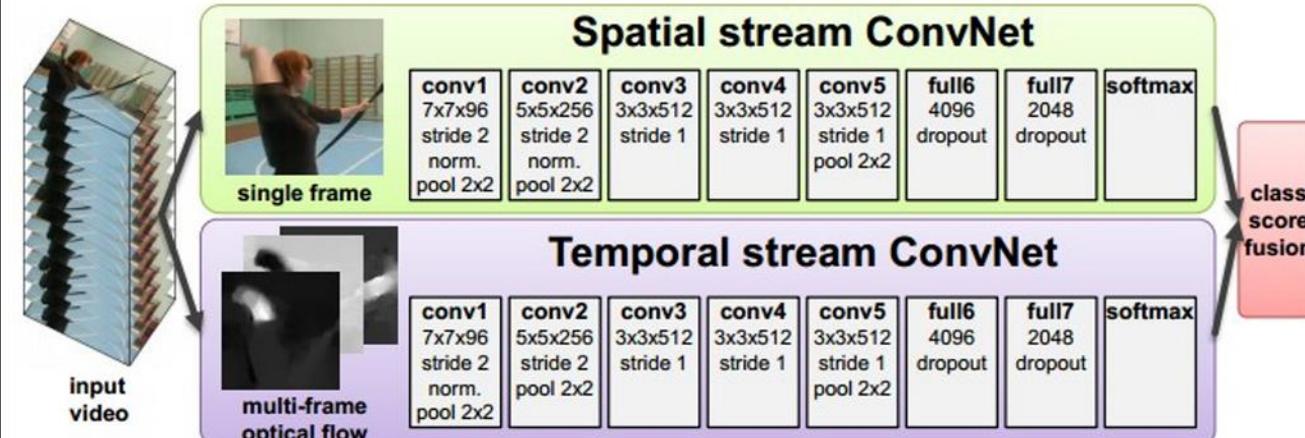
Image Segmentation



Fabaretet al, 2012

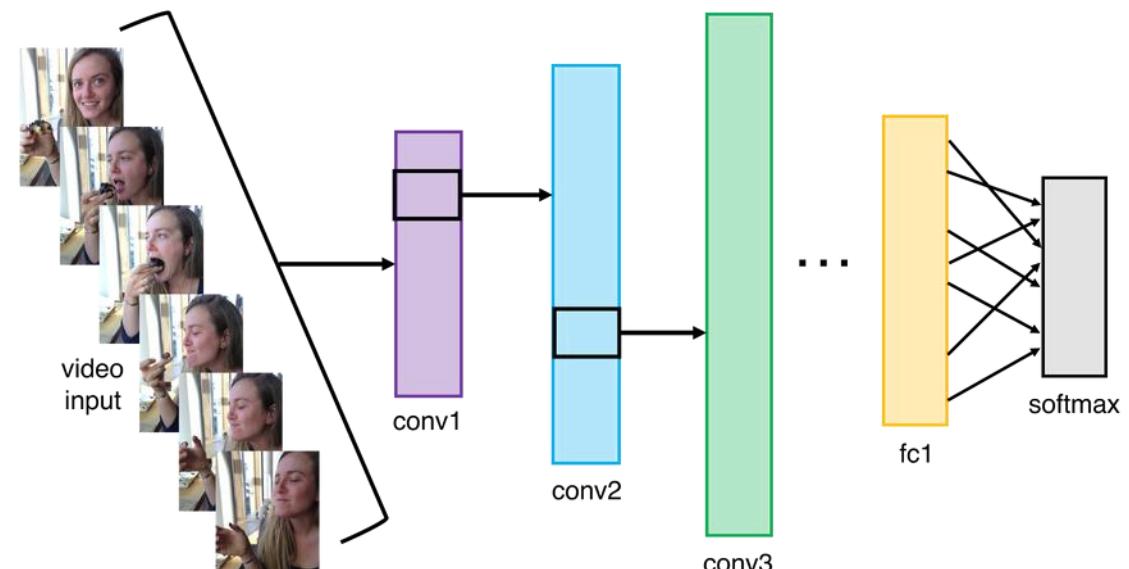
# 2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

Activity Recognition

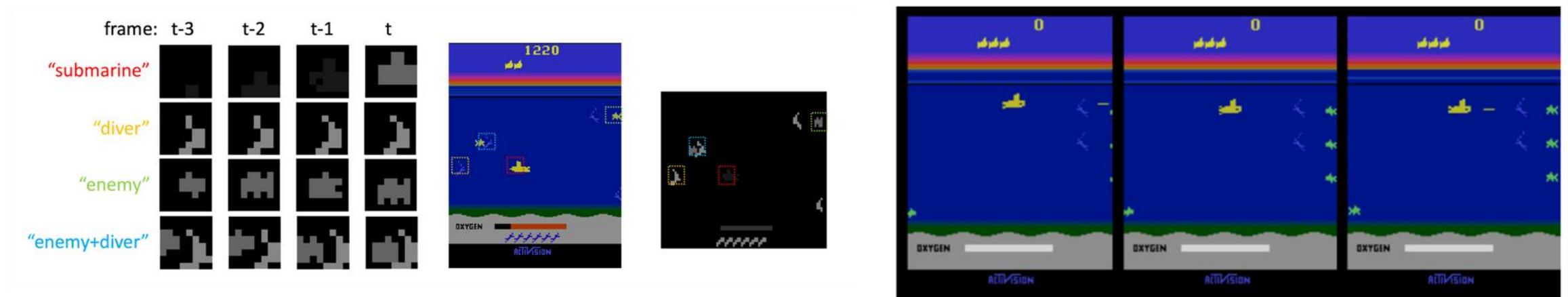


# 2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

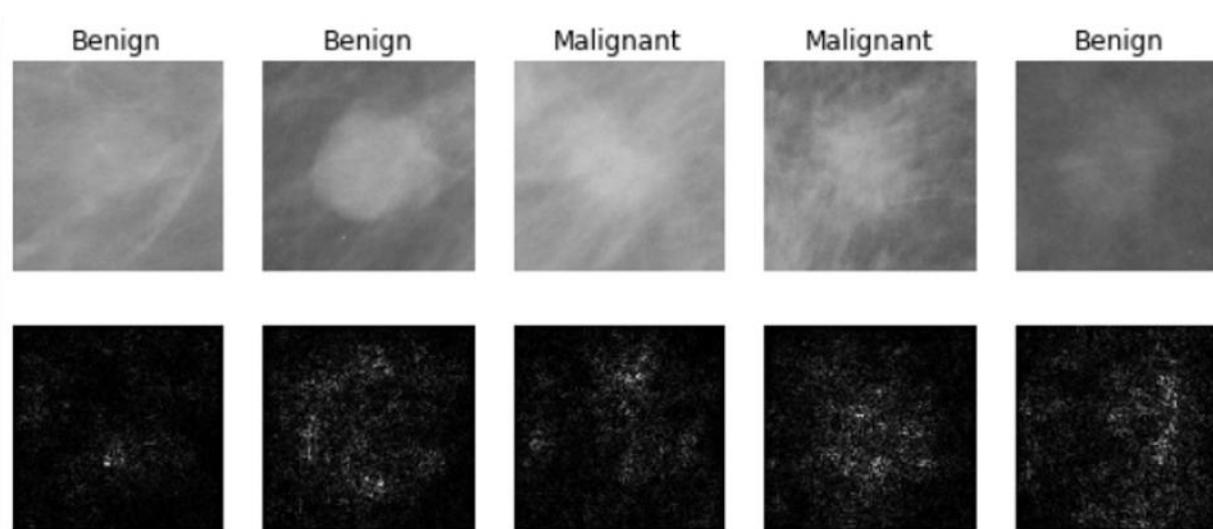


Playing Atari games (Guo et al, 2014)



# 2012 to Present: Deep Learning is Everywhere

Medical Imaging



Levy et al, 2016 Figure

reproduced with Galaxy Classification



Dieleman et al, 2014

From left to right: [public domain by NASA](#), usage [permitted](#) by  
ESA/Hubble, [public domain by NASA](#), and [public domain](#)

Whale recognition



[Kaggle Challenge](#)

This image by Christin Khan is in the public domain and originally came from the U.S. NOAA.

# 2012 to Present: Deep Learning is Everywhere



*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*

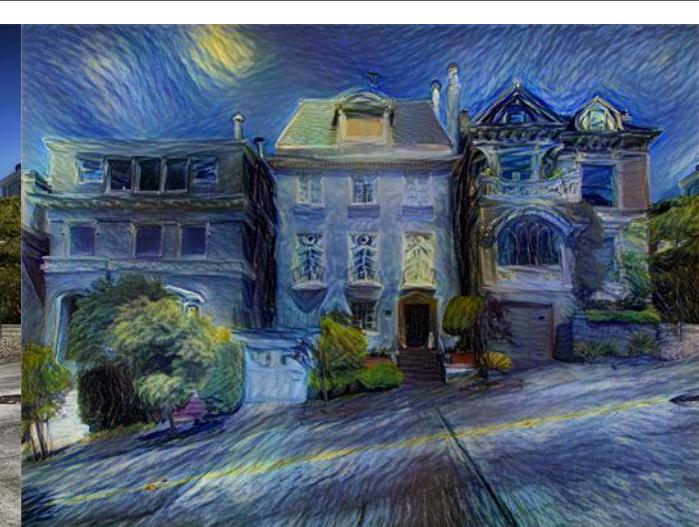
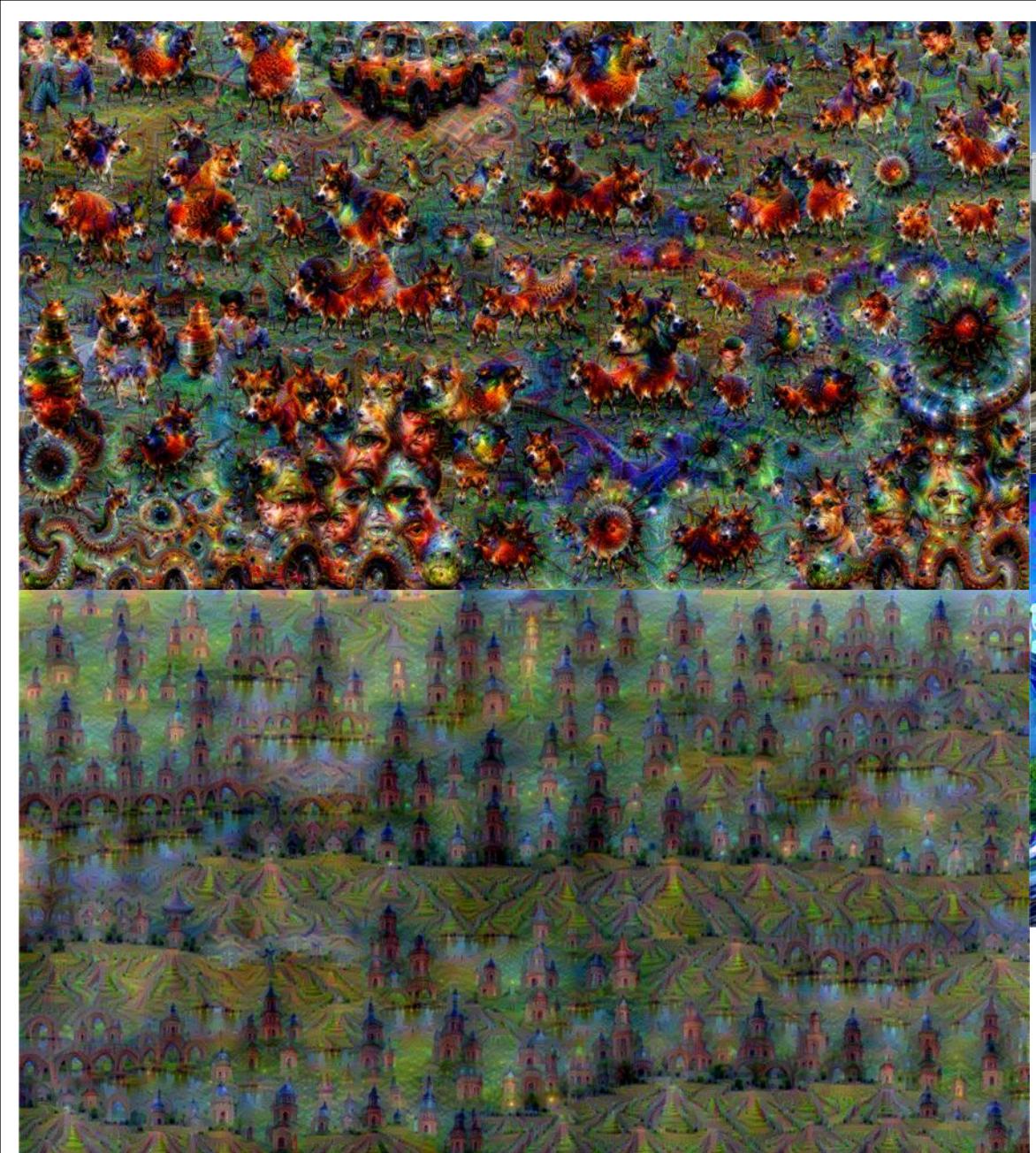
## Image Captioning

Vinyalsetal, 2015

Karpathyand Fei-Fei, 2015

All images are CC0 Public domain: <https://pixabay.com/en/luggage-antique-cat-1643010/> <https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-1623436/> <https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/> <https://pixabay.com/en/woman-female-model-portrait-adult-983967/> <https://pixabay.com/en/handstand-lake-meditation-496008/> <https://pixabay.com/en/baseball-player-shortstop-infield-1045263/>

Captions generated by using [Neuraltalk2](#)



Original image is CC0 public domain [Starry Night](#) and [Tree Roots](#) by Van Gogh are in the public domain  
Bokeh image is in the public domain  
Stylized images copyright , 2017;  
reproduced with permission

Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a [blog post](#) by Google Research.

# 2012 to Present: Deep Learning is Everywhere

## TEXT PROMPT

an armchair in the shape of an avocado. an armchair imitating an avocado.

## AI-GENERATED IMAGES



# 2012 to Present: Deep Learning is Everywhere

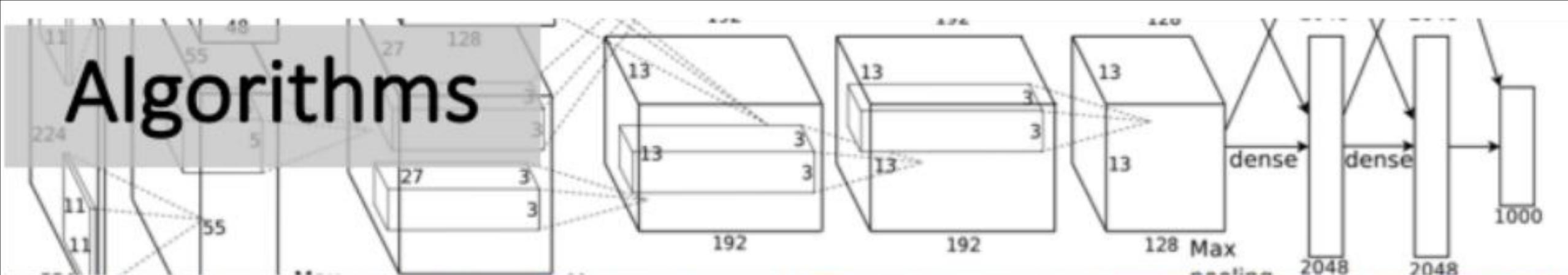
## TEXT PROMPT

an armchair in the shape of a peach. an armchair imitating a peach.

## AI-GENERATED IMAGES



# Algorithms

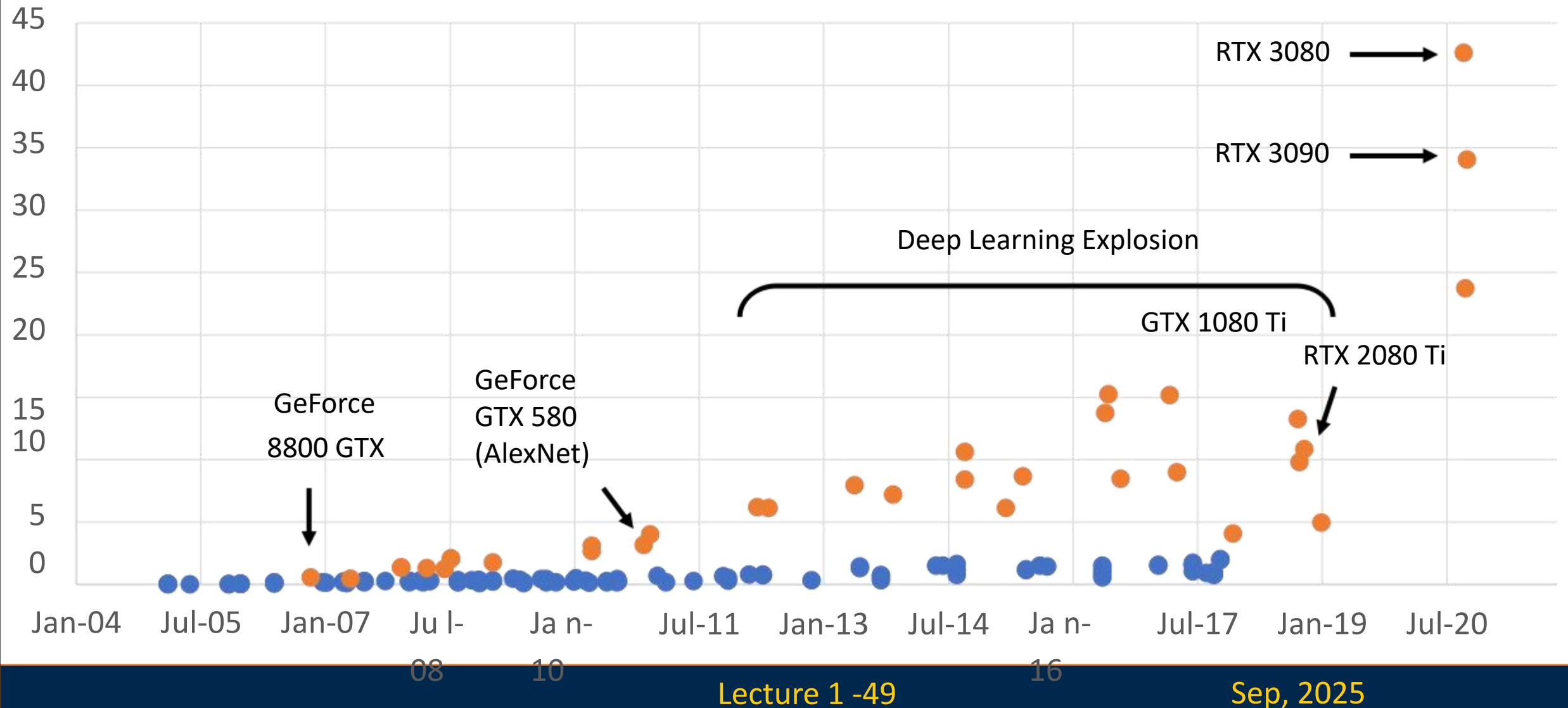


# Data

# Computation

# GFLOP per Dollar

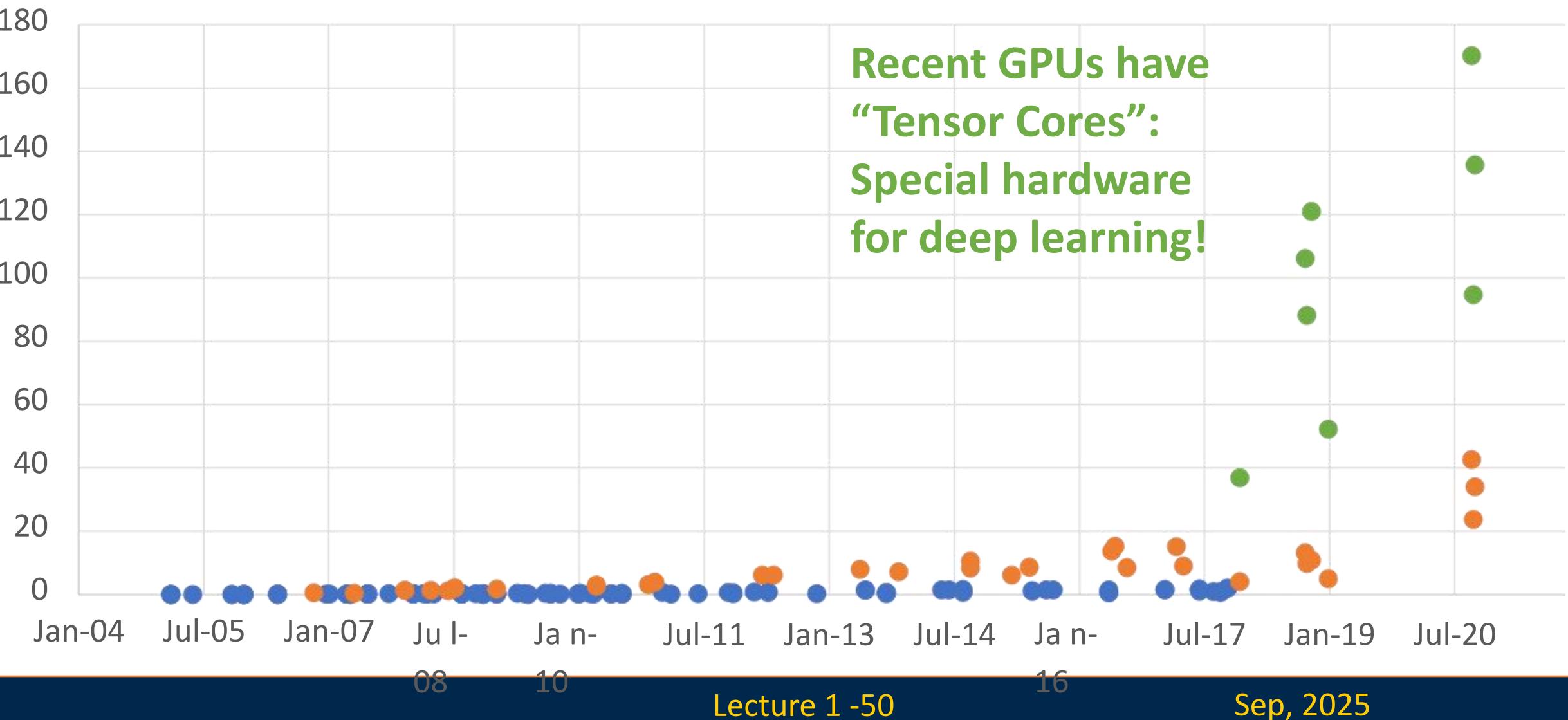
● CPU   ● GPU(FP32)



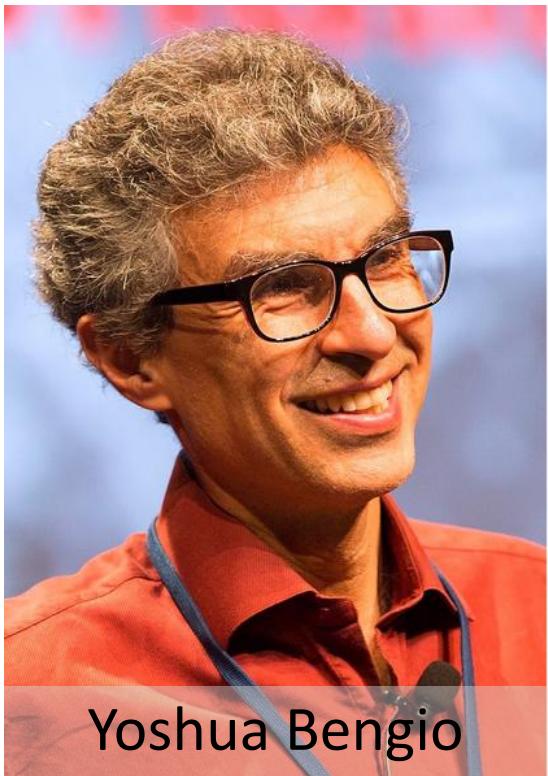
# GFLOP per Dollar

● CPU    ● GPU(FP32)    ● GPU(Tensor Core)

Recent GPUs have  
“Tensor Cores”:  
Special hardware  
for deep learning!



# 2018 Turing Award



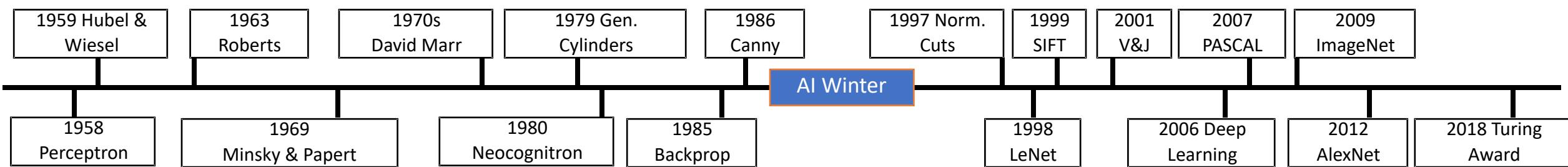
Yoshua Bengio



Geoffrey Hinton



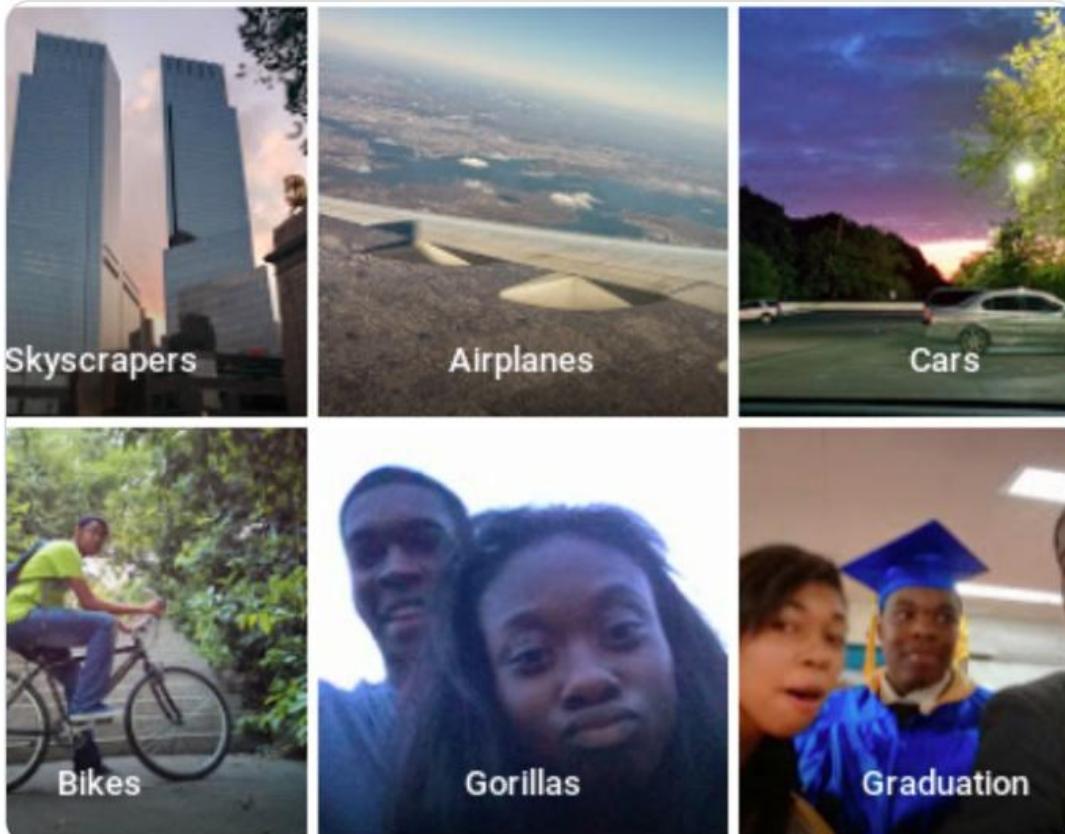
Yann LeCun



Despite our success, computer vision still has a long way to go...

# Computer Vision can cause harm

Harmful Stereotypes



Barocas et al, "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGKDD 2017

Kate Crawford, "The Trouble with Bias", NeurIPS2017 Keynote Source:

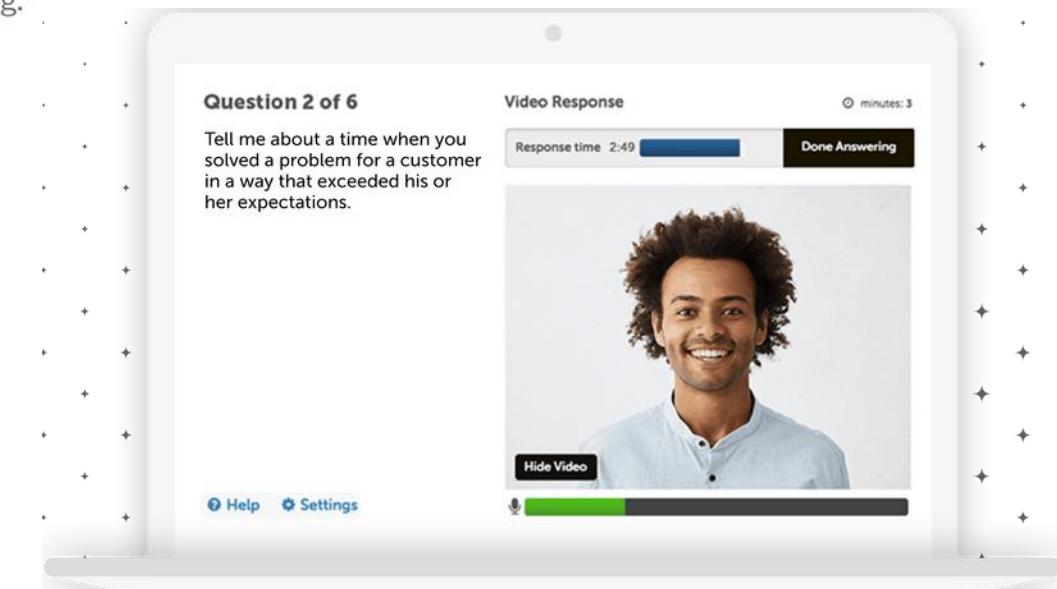
<https://twitter.com/jackyalcine/status/615329515909156865>(2015)

Affect people's lives

Technology

## A face-scanning algorithm increasingly decides whether you deserve the job

HireVue claims it uses artificial intelligence to decide who's best for a job. Outside experts call it 'profoundly disturbing.'



Source: <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>

<https://www.hirevue.com/platform/online-video-interviewing-software> Example Credit: TimnitGebru



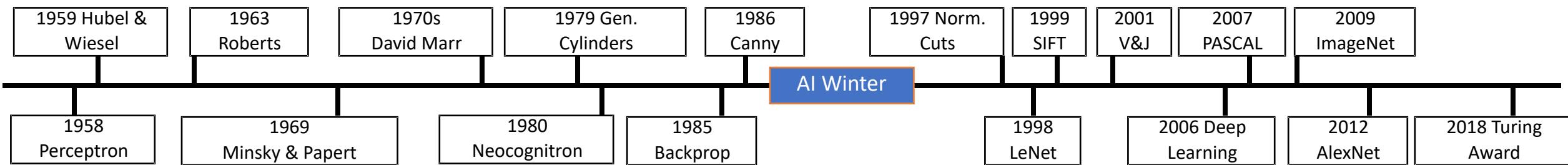
This image is copyright-free United States government work

# Computer Vision Technology Can Better Our Lives



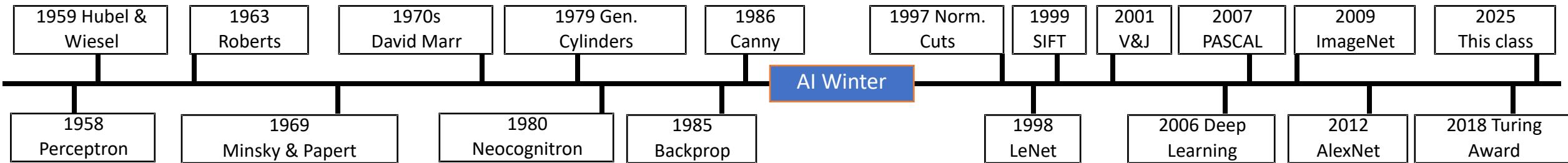
# Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics



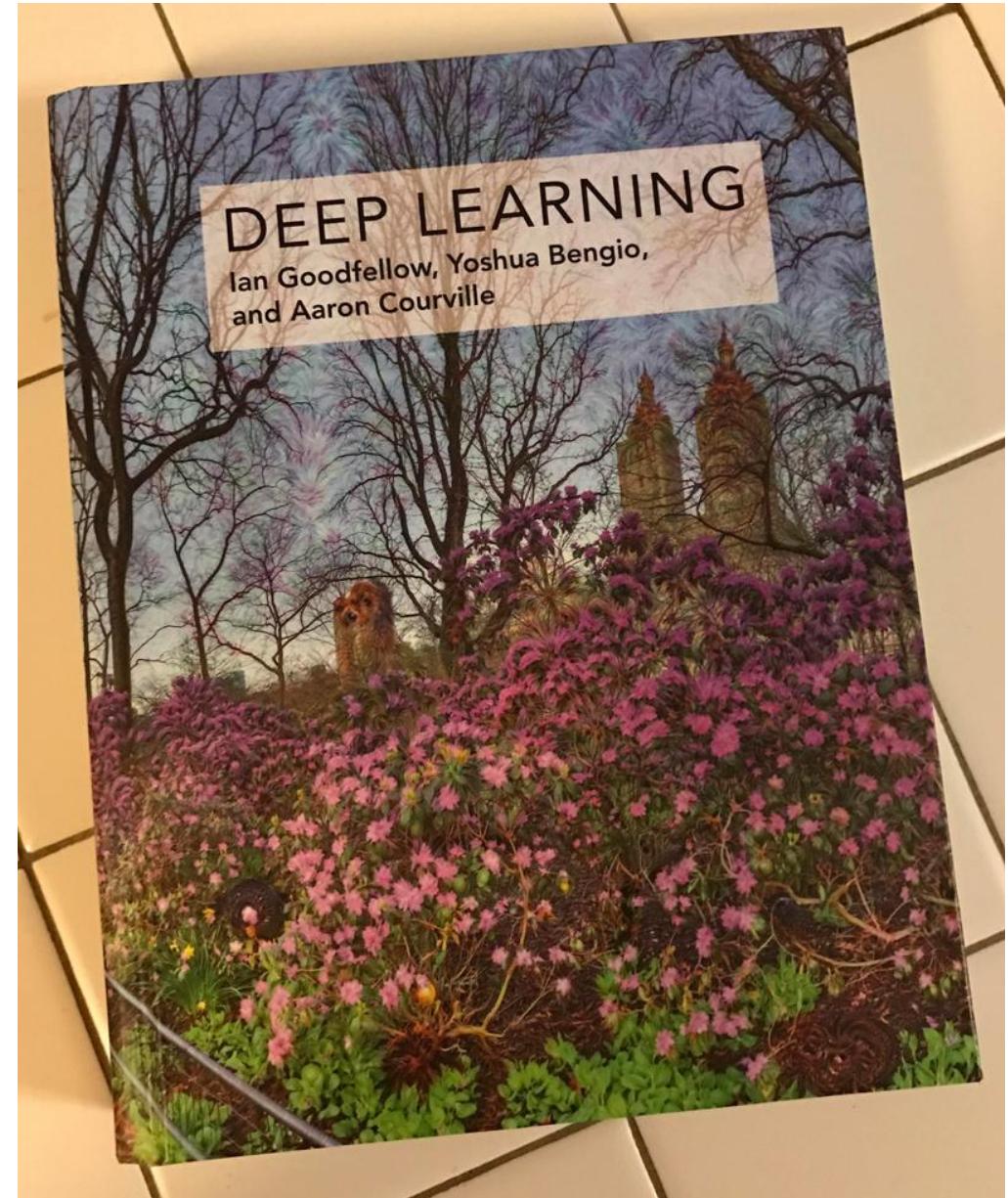
# Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics



# Optional Textbook

- [Deep Learning](#) by Goodfellow, Bengio, and Courville
- [Free online](#)



# Course Content and Grading

<https://github.com/naziherrahel/Deep-Learning-for-Computer-Vision>

# First homework assignment

- Will be released by today or tomorrow
- Next lecture will be enough to complete it

# Next time: Image Classification