1. Introduction

GEV6135 Deep Learning for Visual Recognition and Applications

Kibok Lee

Assistant Professor of

Applied Statistics / Statistics and Data Science

Sep 1, 2022

연세대학교

YONSEI UNIVERSITY

Outline

- Course Information
- A Brief History of Computer Vision and Deep Learning

Deep Learning for Visual Recognition and Applications

Deep Learning for Visual Recognition and Applications

A fundamental and general problem in Computer Vision, that has roots in Cognitive Science

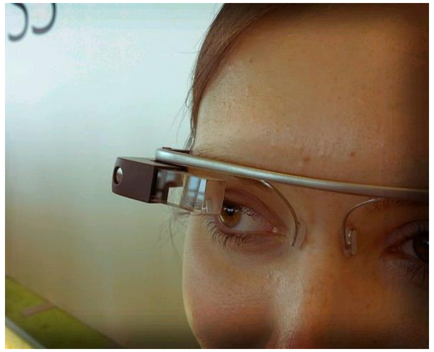
Computer Vision is Everywhere!









Deep Learning for Visual Recognition and Applications

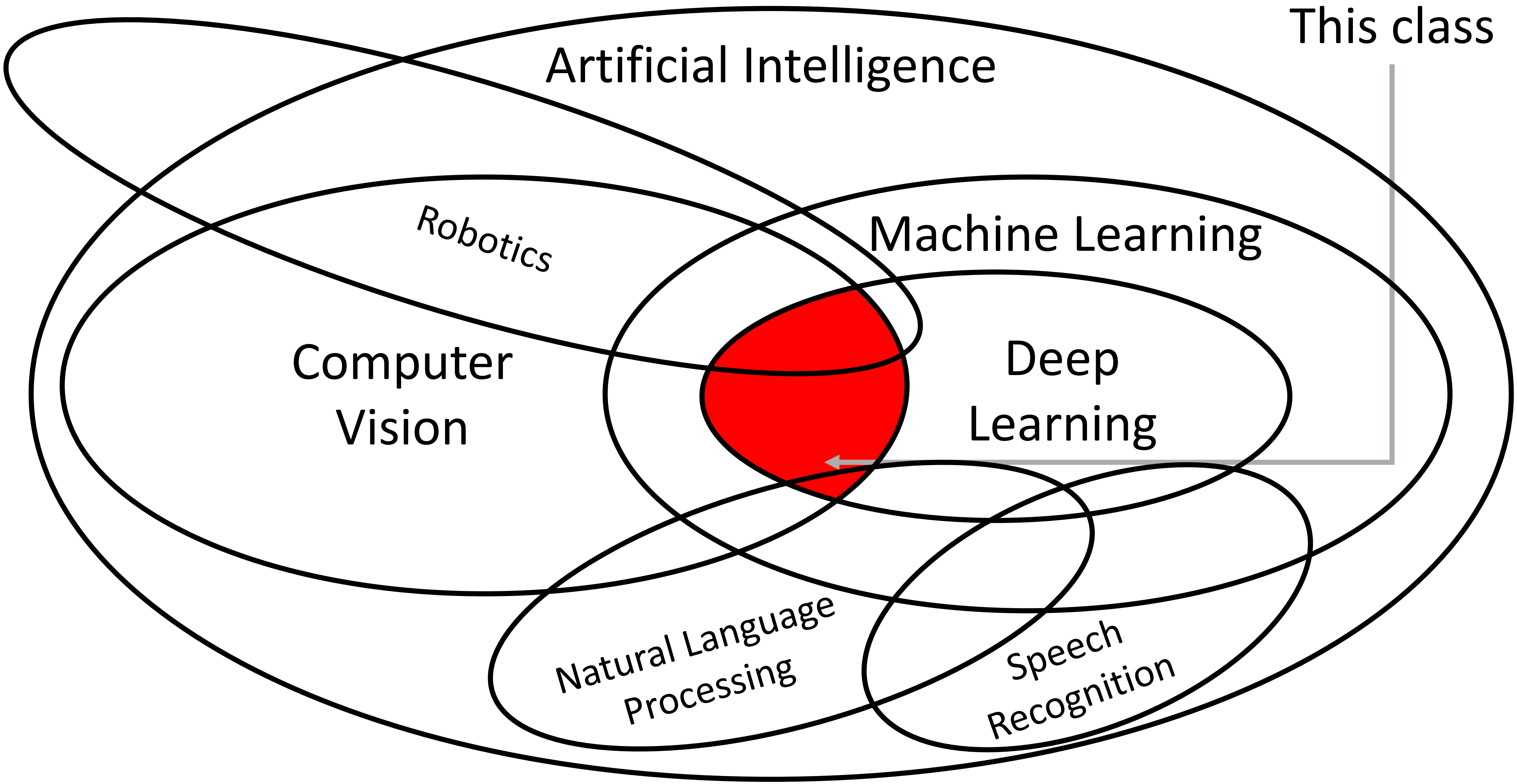
Building artificial systems that
learn from data and experience

Deep Learning for Visual Recognition and Applications

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

Deep Learning for Visual Recognition and Applications

… and we will see some interesting applications!


This class
Artificial Intelligence
Robotics
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition

Teaching Staff

- Instructor: Kibok Lee
- Email: kibok@yonsei.ac.kr
- Office: Daewoo Hall 428

- Teaching assistant: Jinu Kim
- Email: jinu_kim@yonsei.ac.kr
- Office: Daewoo Hall 401

How to Contact Us

- via CLASSUM
  - All questions about the course should go here.
  - Students are encouraged to ask and answer each other.
  - We will also use this to communicate with you.

LearnUs

- Email: Only for sensitive and/or confidential issues

Lecture Format

- Location: Engineering Hall D603
- Time: Thursday 6:50 PM – 8:20 PM
- Office hours: 24/7 @ CLASSUM
  - No regular in-person office hours

- Slides are mostly self-contained.
  - I would not write in class.
- Solution for assignments/exams will not be shared.
  - Questions are often open-ended.
  - We don’t have time to create publishable solutions.

Grading Policy

- Assignments: 60%
  - There will be 6 ~ 8 programming assignments; depending on the progress
  - Python, PyTorch, will use Google Colab
- Midterm/Final exam: 15% each
  - Written exam testing basic concepts
  - True / False, multiple choice, short answer
- Attendance: 10%
- Absolute evaluation
  - Scaling up based on a curve if the statistics is low
- Honor code (next)

Late/Absence Policy

- Assignments: -25% additive for each late day
  - No point if late for >= 4 days
  - We ignore minor stuffs, e.g., no deduction if you are late for few minutes because of some internet delays.

- Attendance:
  - **Y-attend app** at the beginning of the lecture
  - -1 for each absence
  - You have 3 free absences.
  - The score is ***min(10, 13 – N)*** if you are absent *N* times.
    - among 24 classes (except the 1st week, exams)
  - If you are absent 1/3 of the course, you will get **[F]**.

Study group

- We encourage students to form study groups.
  - Up to **four** people are allowed for each group.
- You may discuss with the study group members on assignments. But you should write your own solution **independently**.
- When submitting assignments, put the names of people with whom you have discussed.
- Please start on assignments early.
  - Warning: cramming does not work!

Honor Code

- In addition to the Yonsei University's honor code,
  - Do not give or receive unpermitted aid in exams, assignments, or any other work that is to be used by the instructor as the basis of grading.
  - Do not share solutions or program code with other students.

- Plagiarism is usually obvious and painful; we don't have many options if any plagiarism is found. Instead,
  - You can submit your work late.
  - You don't even have to solve some questions.
  - You'd better to get a low score than to ruin your academic record.
  - <u>Note: If something is difficult for you, then so are others.</u>

Assignment Options

- Option 1: Submit your own work
  - You must not copy/refer to others’ code.
  - You will get **[F]** if any plagiarism is found.

- Option 2: Refer to others’ code and add comments
  - You need to cite references clearly.
    - Website address or your study group member
  - Your comments will also be graded.
  - Your score will be downscaled to 80%.
    - e.g., if you take this option for 4 out of 6 assignments, the maximum score you can get is 2*10 + 4*10*.8 = 52
    - If you take option 2 for all assignments and get perfect scores, the expected grade is **[A-]**.

Plagiarism: Case Study

1. I don't know how to use a PyTorch function, so I googled it and carefully read the documentation.
   - This is encouraged. Please do so!

2. I found a (partial) solution on a website.
   - You are not recommended to carefully look at it, as your solution might be unintentionally affected.
   - If you referred to it, consider to go for option 2.

Plagiarism: Case Study

1. I don’t know what the steps are to implement an algorithm for Q2-2, so I discussed with my friend.
   - This is encouraged. But don’t share code directly.

2. I don’t know how to implement an algorithm for Q2-3, so I asked my friend for help.
   - If you wrote your own code, then it is likely to be fine.
   - If you referred to their code while writing yours, consider to go for option 2.

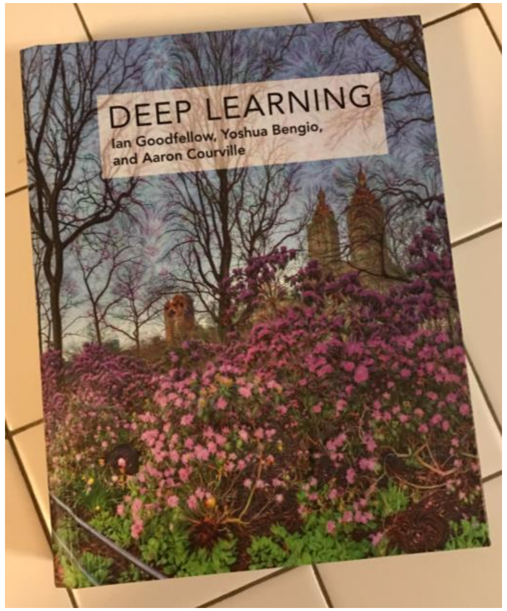
Assignment 1

- About basic PyTorch

- Will be out around this weekend
- Due Monday 9/12, 11:59pm
  - Please read the instruction carefully!
  - Make sure you do **not import** additional libraries
  - Make sure to **manually save** the .py file in Colab
  - After you download the .zip file, **check that the .py file is correct**
  - Do **not zip by yourself**, run the provided code
    - You might miss some files
    - On MAC, zip via GUI includes redundant files, violating the expected folder organization.

Reference

- EECS498/598-xxx Deep Learning for Computer Vision
  Prof. Justin Johnson @ University of Michigan
  - Most slide credits to this course
- CS231n Convolutional Neural Networks for Visual Recognition
  Prof. Fei-Fei Li *et al.* @ Stanford University



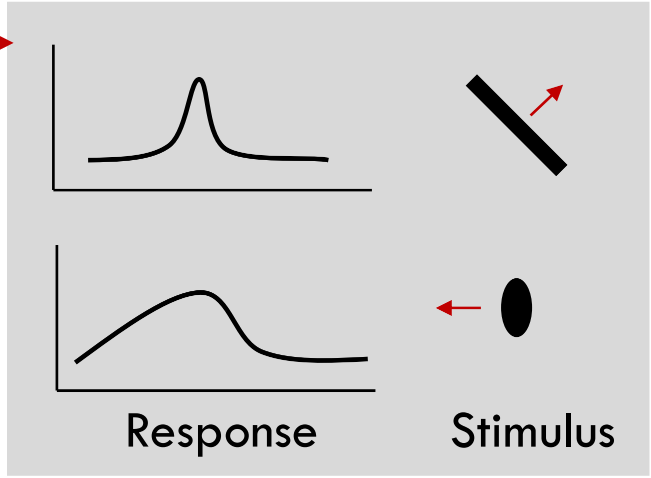
- No official textbook
- Goodfellow et al., Deep Learning. *MIT Press*, 2016.
  - https://www.deeplearningbook.org/
  - This would be a good reference.

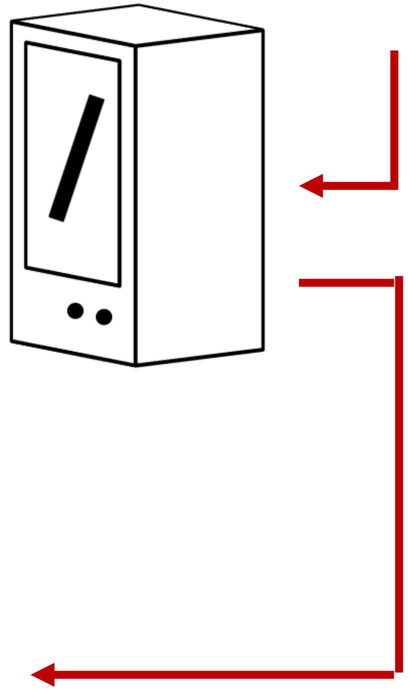
Today's Agenda

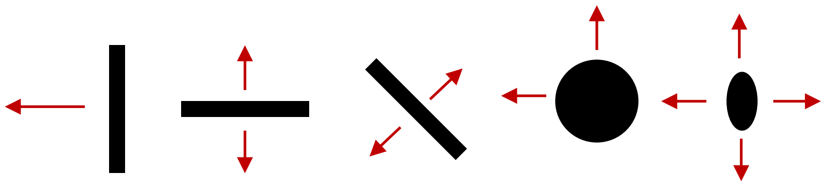
- Course overview and logistics
- A brief history of computer vision and deep learning

Hubel and Wiesel, 1959

Measure brain activity

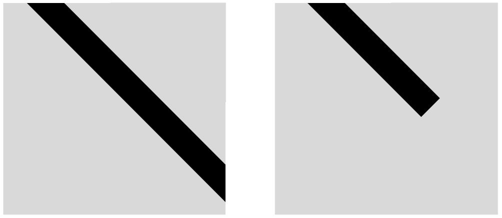






Simple cells:
Response to specific rotation and orientation

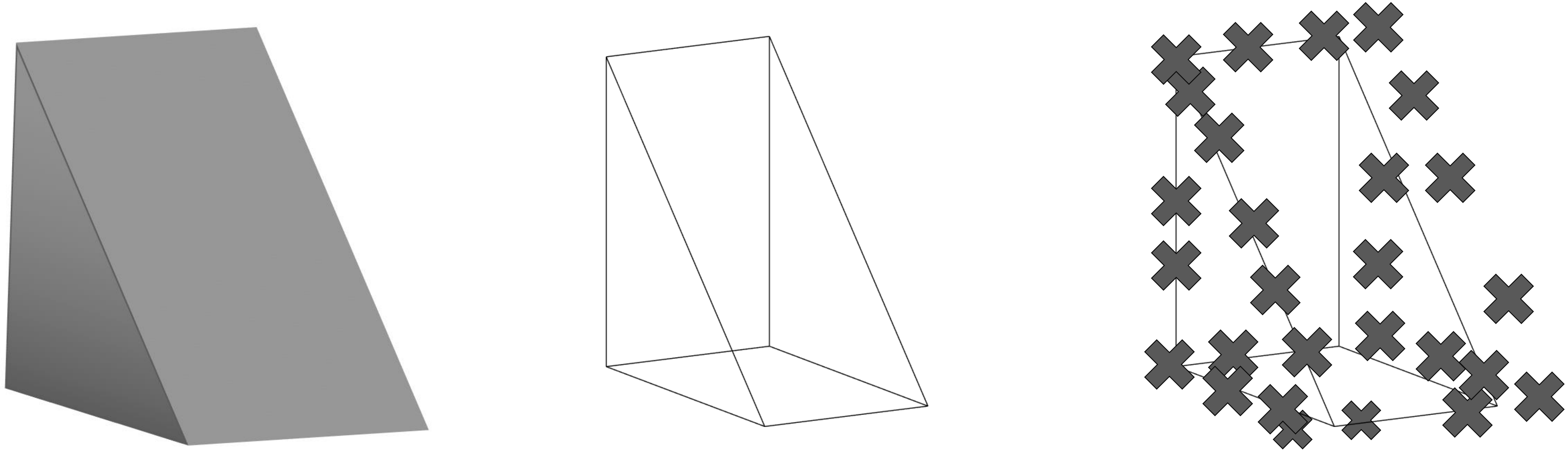
Complex cells:
Response to light orientation and movement, some translation invariance



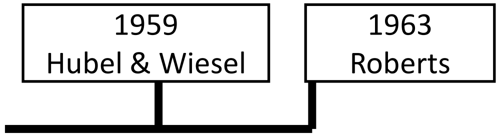
No response

1959
Hubel & Wiesel

Larry Roberts, 1963



(a) Original picture (b) Differentiated picture (c) Feature points selected



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

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

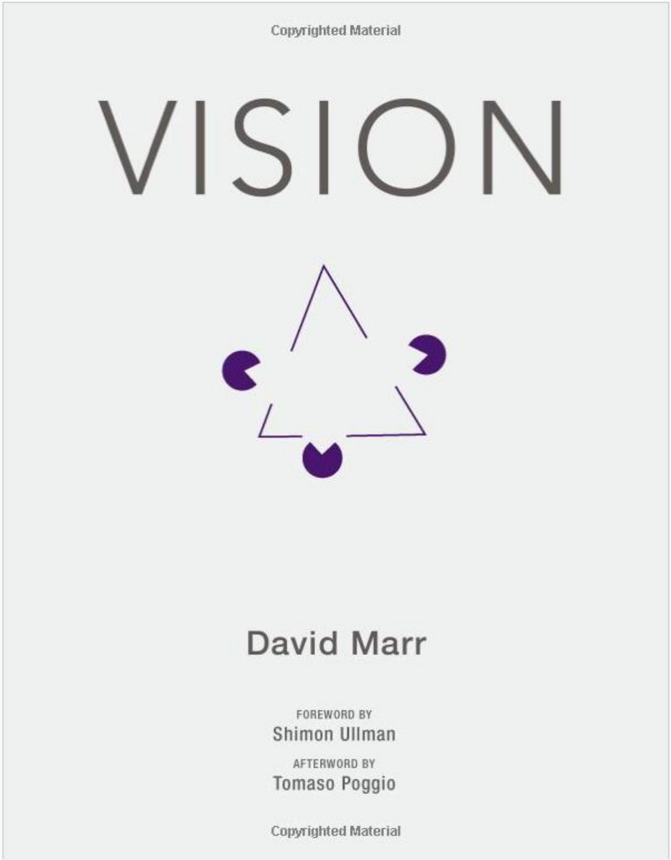
Seymour Papert

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".

1959
Hubel & Wiesel

1963
Roberts

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

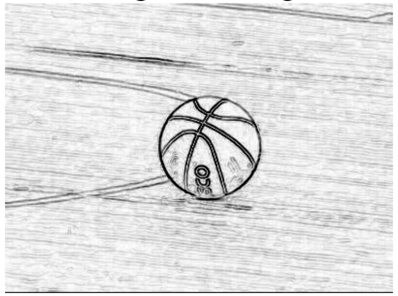


1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr

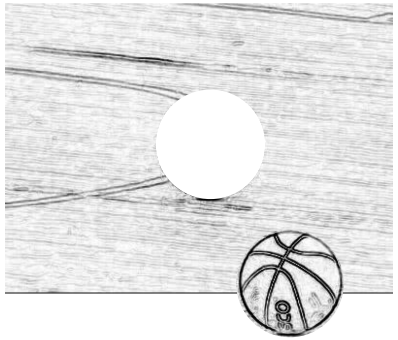
Input image

This image is CC0 1.0 public domain

Edge image



2 ½-D sketch



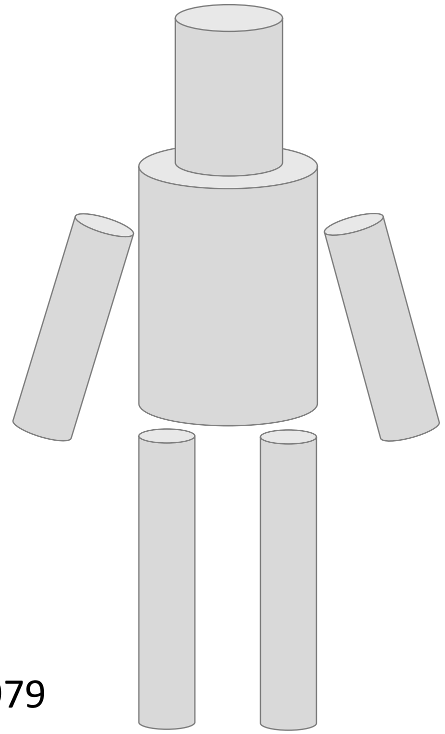
3-D model

This image is CC0 1.0 public domain

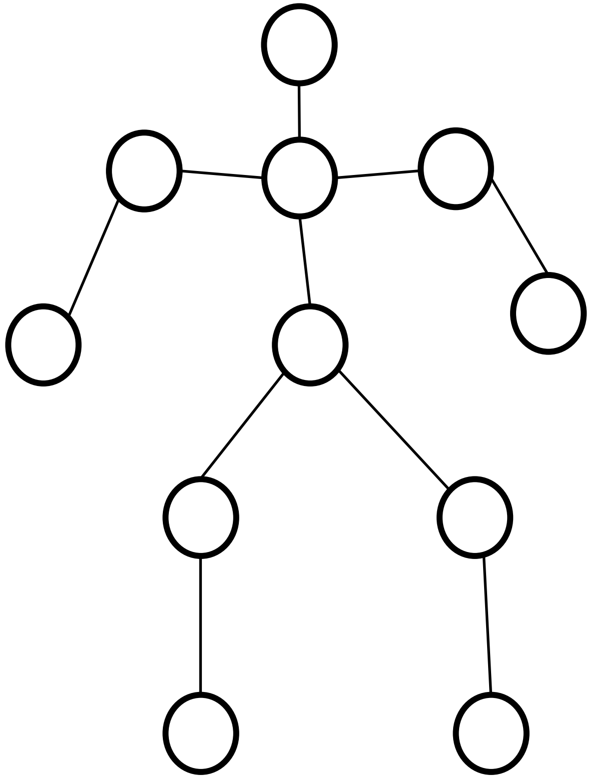
Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface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

Stages of Visual Representation, David Marr, 1970s

Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford, 1979

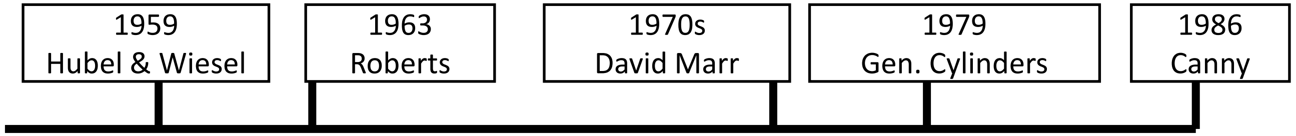


Pictorial Structures,
Fischler and Elshlager, 1973

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

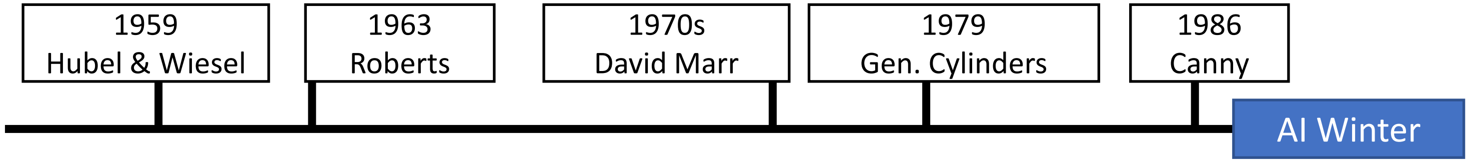
Recognition via Edge Detection (1980s)

John Canny, 1986
David Lowe, 1987



AI Winter

- Enthusiasm (and funding!) for AI research dwindled
- “Expert Systems” failed to deliver on their promises



Recognition via Grouping (1990s)

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | AI Winter | 1997 Norm. Cuts

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	AI Winter	1997 Norm. Cuts	1999 SIFT

SIFT, David Lowe, 1999

Face Detection

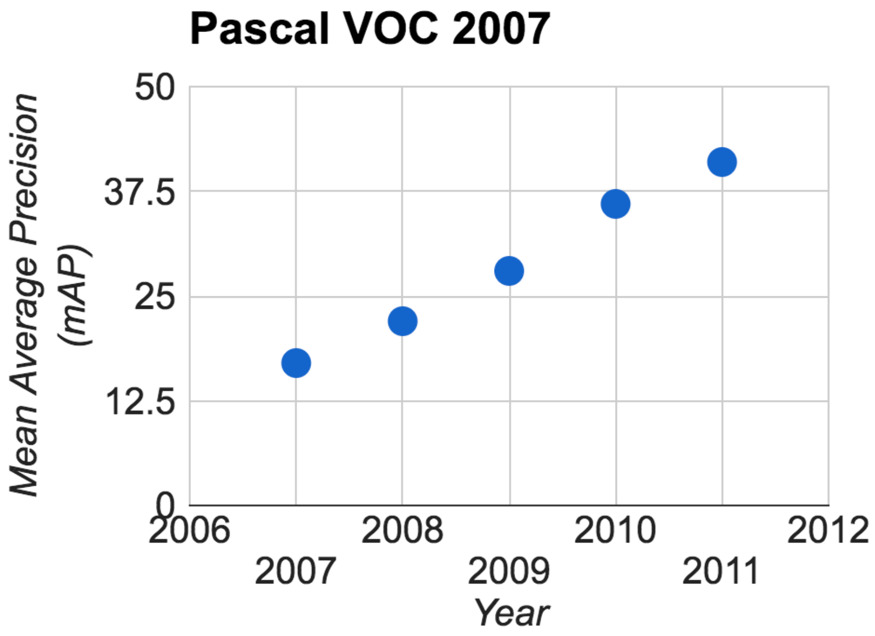
Viola and Jones, 2001

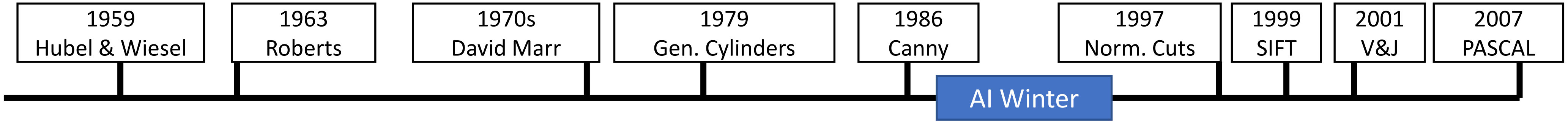
One of the first successful applications of machine learning to vision

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





IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

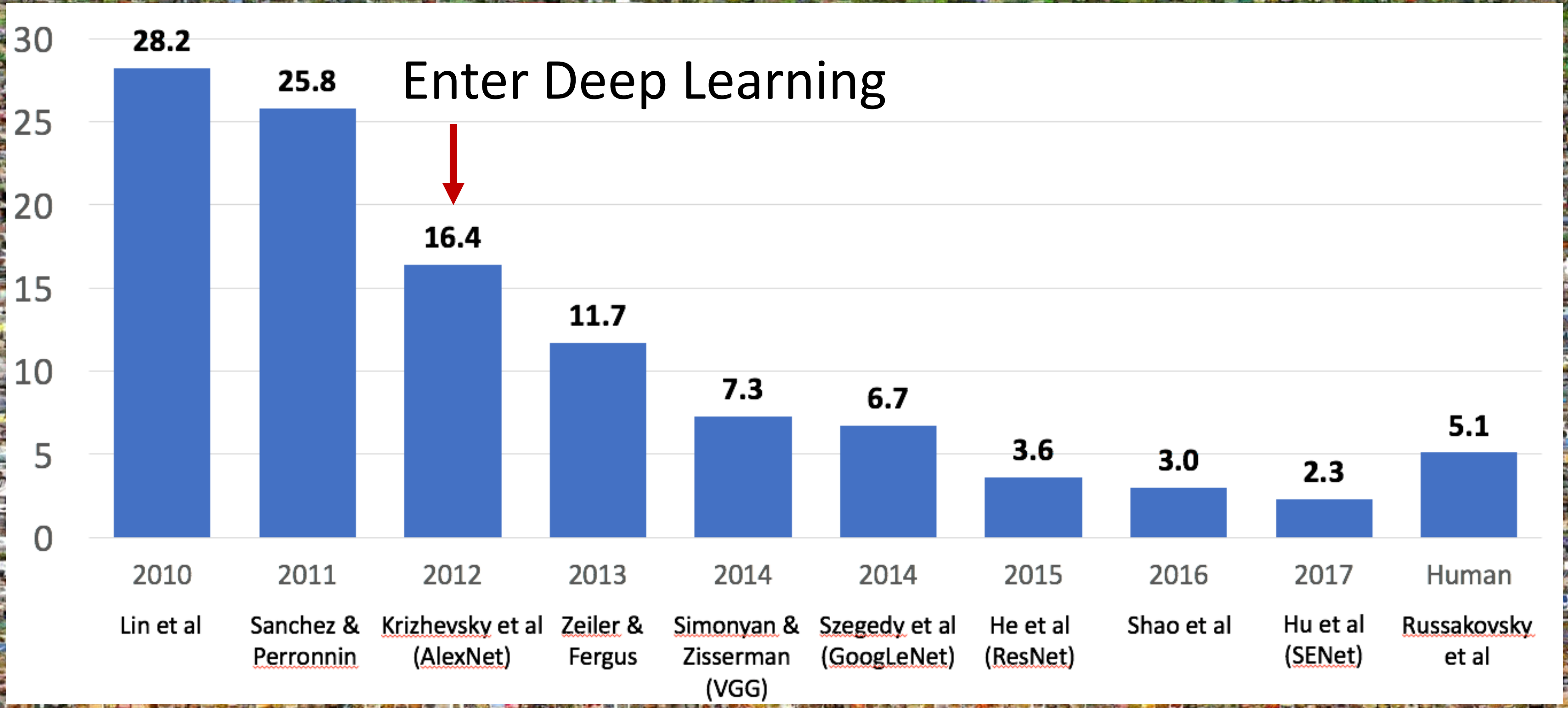
1,431,167 images

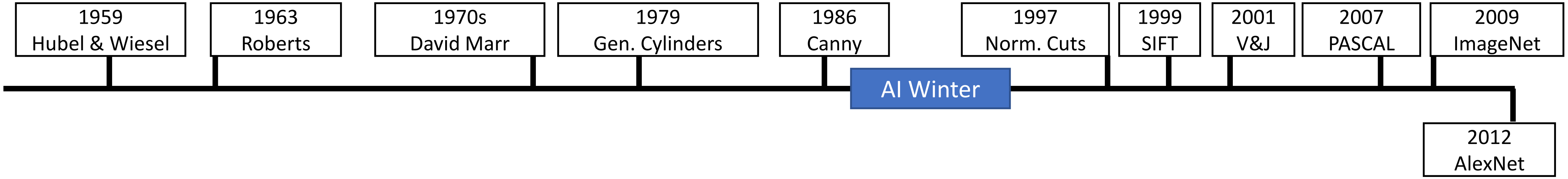
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Deng et al, 2009
Russakovsky et al. IJCV 2015

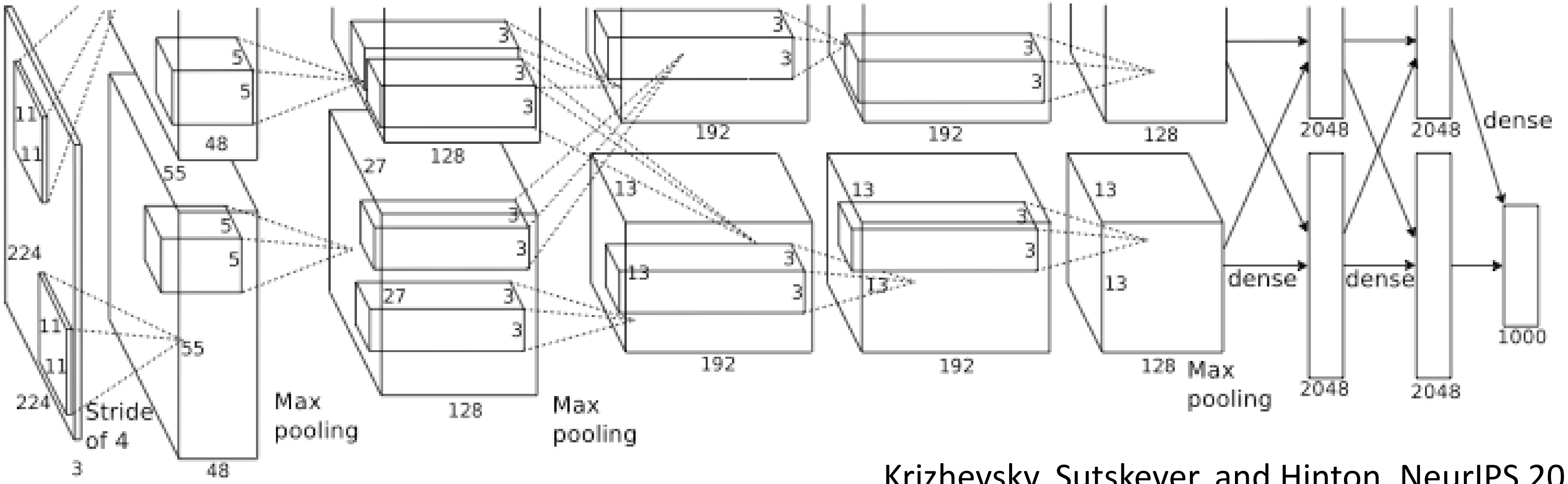
1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | AI Winter | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2007 PASCAL | 2009 ImageNet

IMAGENET Large Scale Visual Recognition Challenge





AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT
2001 V&J
2007 PASCAL
2009 ImageNet
2012 AlexNet

Perceptron

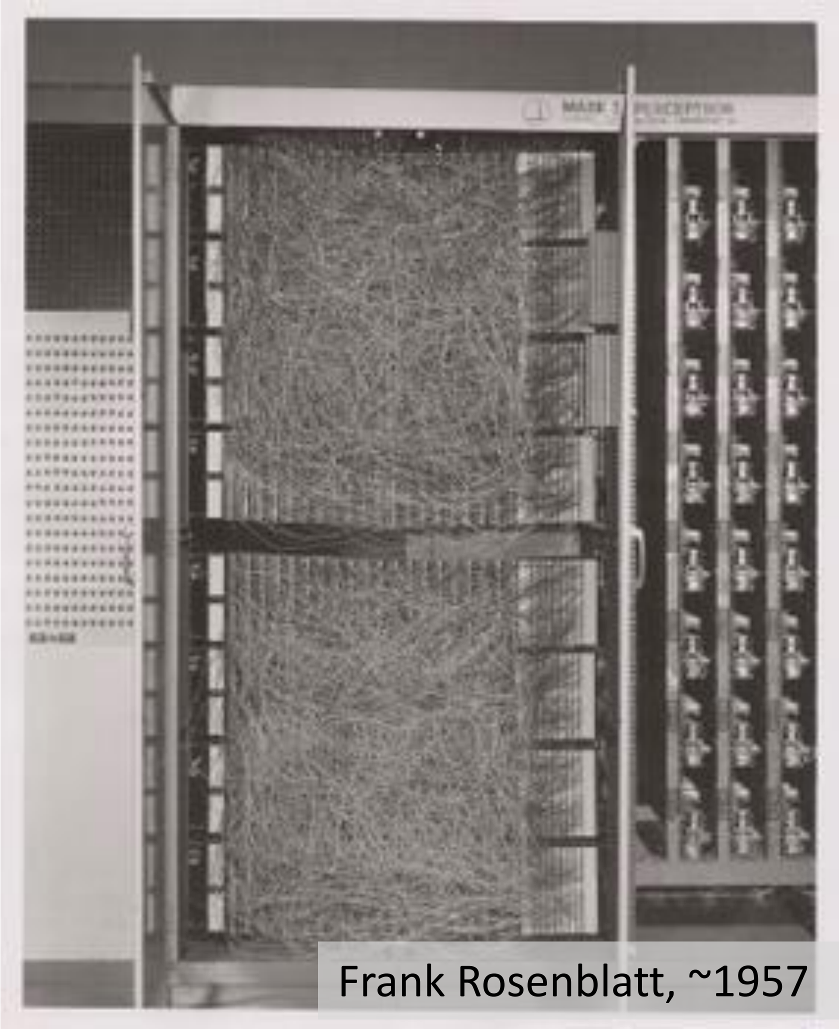
One of the earliest algorithms that could learn from data

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**



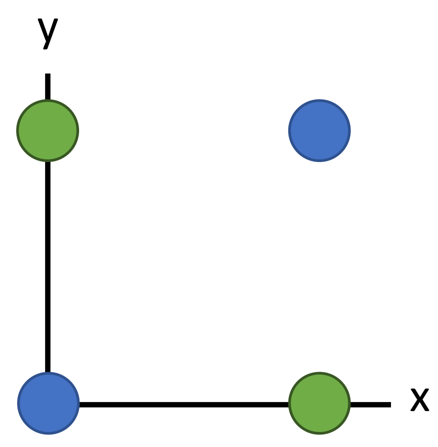
Frank Rosenblatt, ~1957

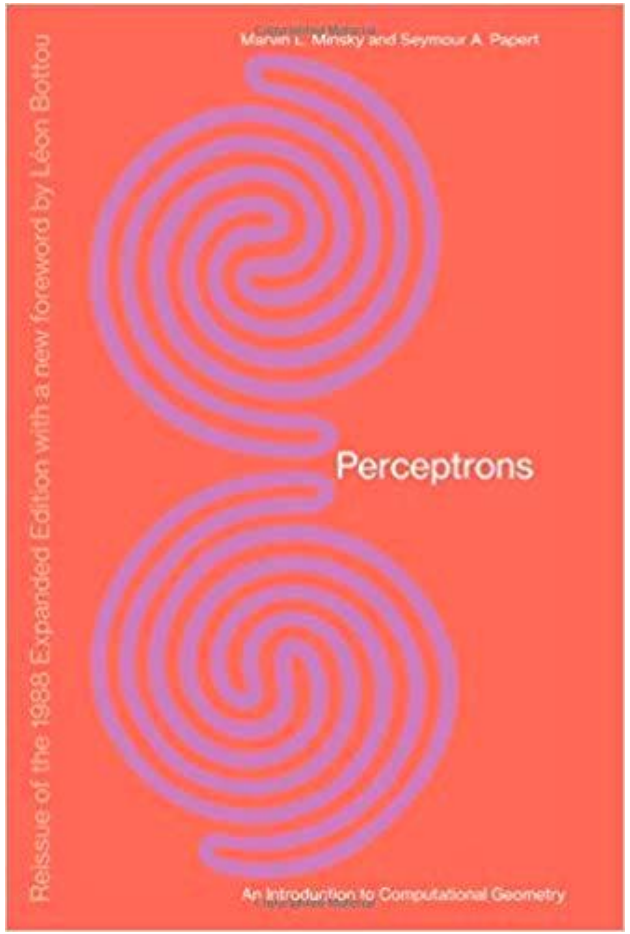
1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | AI Winter | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J | 2007 PASCAL | 2009 ImageNet

1958 Perceptron | 2012 AlexNet

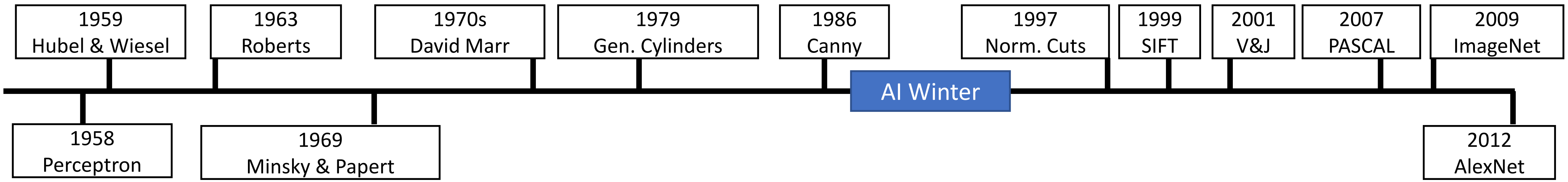
Minsky and Papert, 1969

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





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

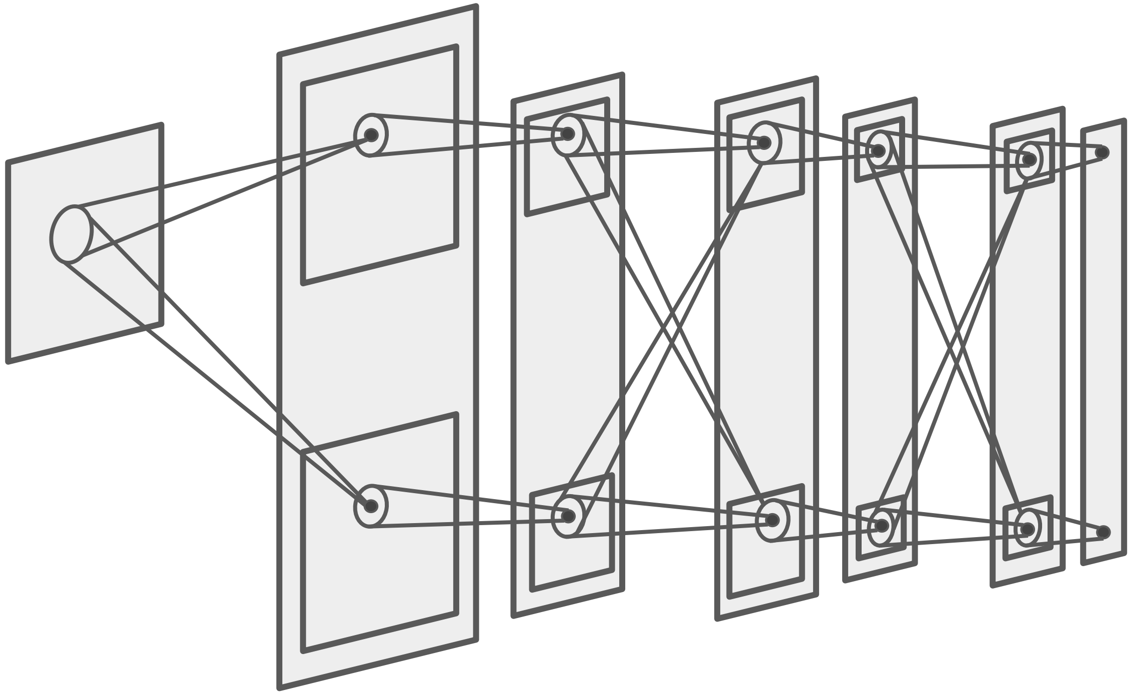


Neocognitron: Fukushima, 1980

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

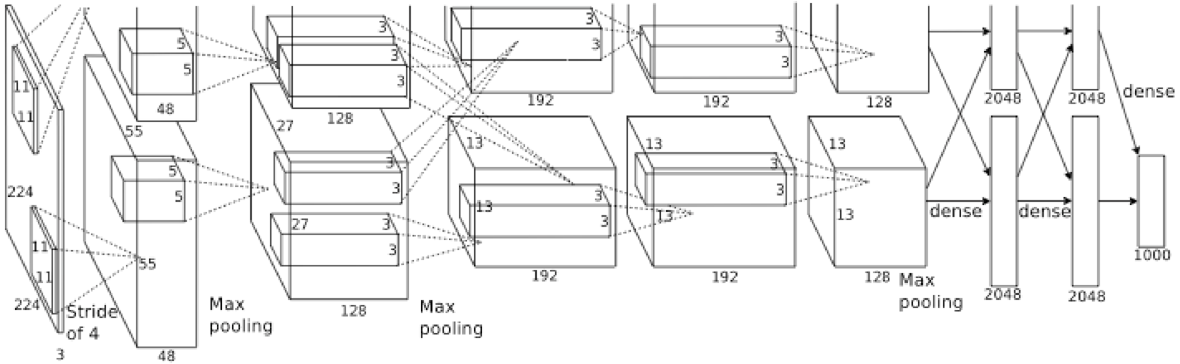


1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1997 Norm. Cuts
1999 SIFT
2001 V&J
2007 PASCAL
2009 ImageNet

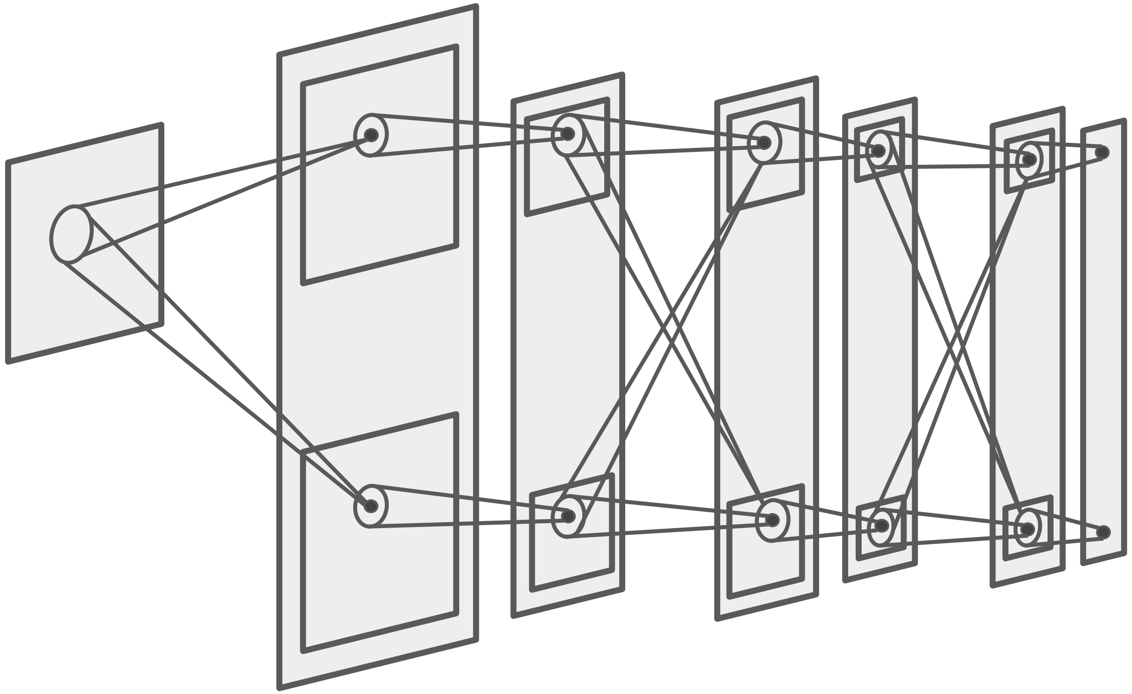
AI Winter

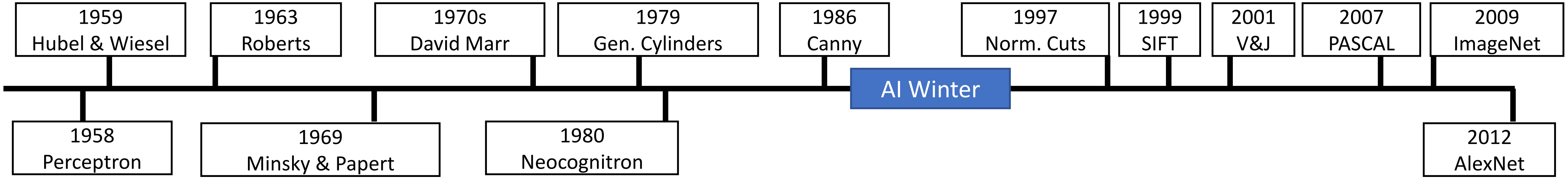
1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
2012 AlexNet

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

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

recognizable math

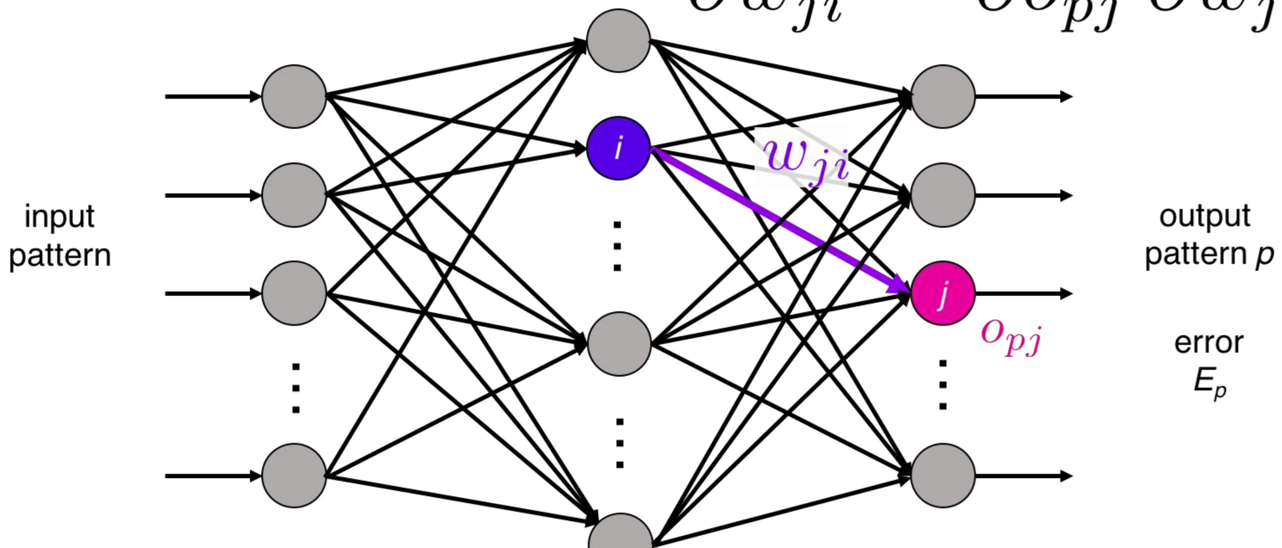
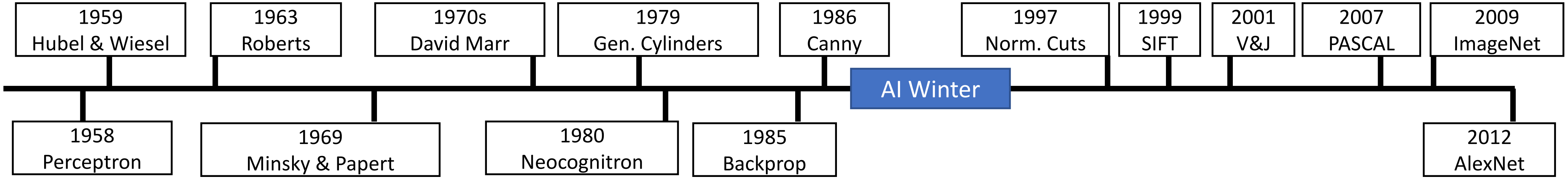
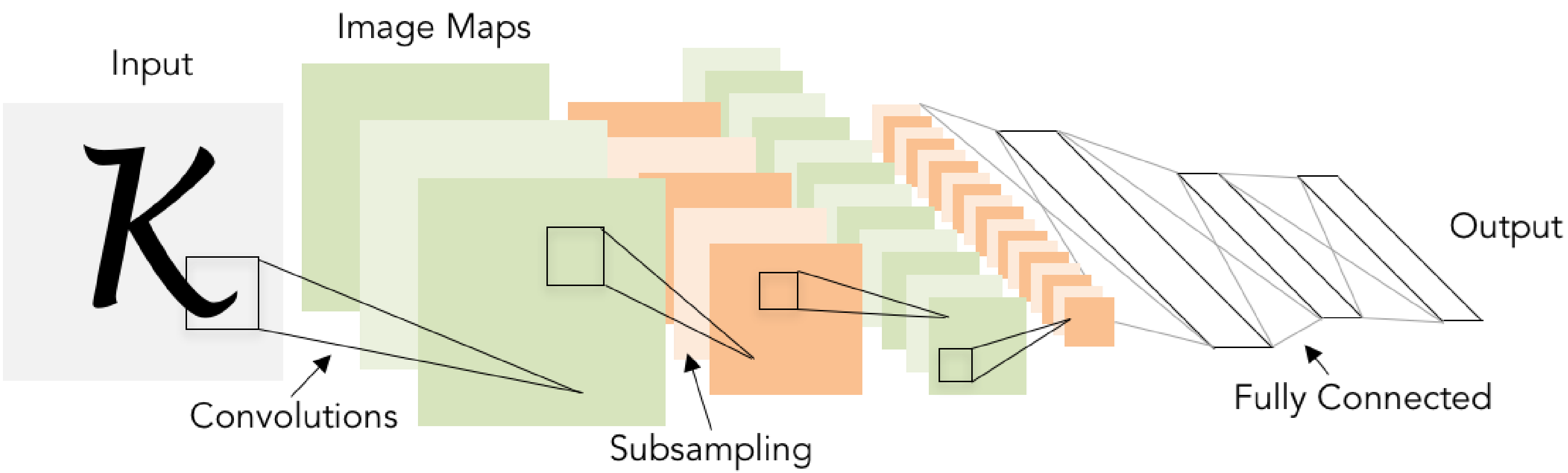


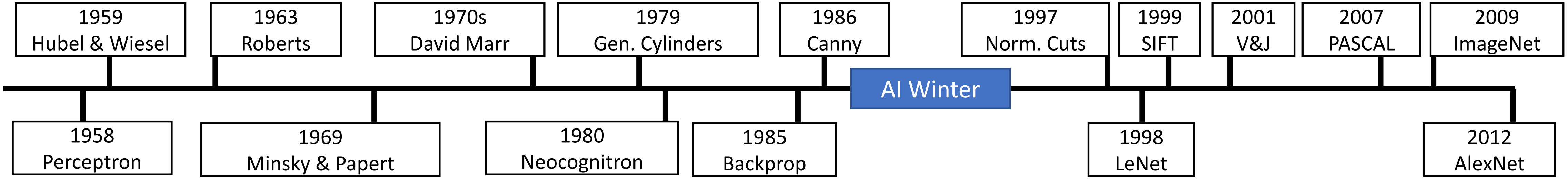
Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017



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!

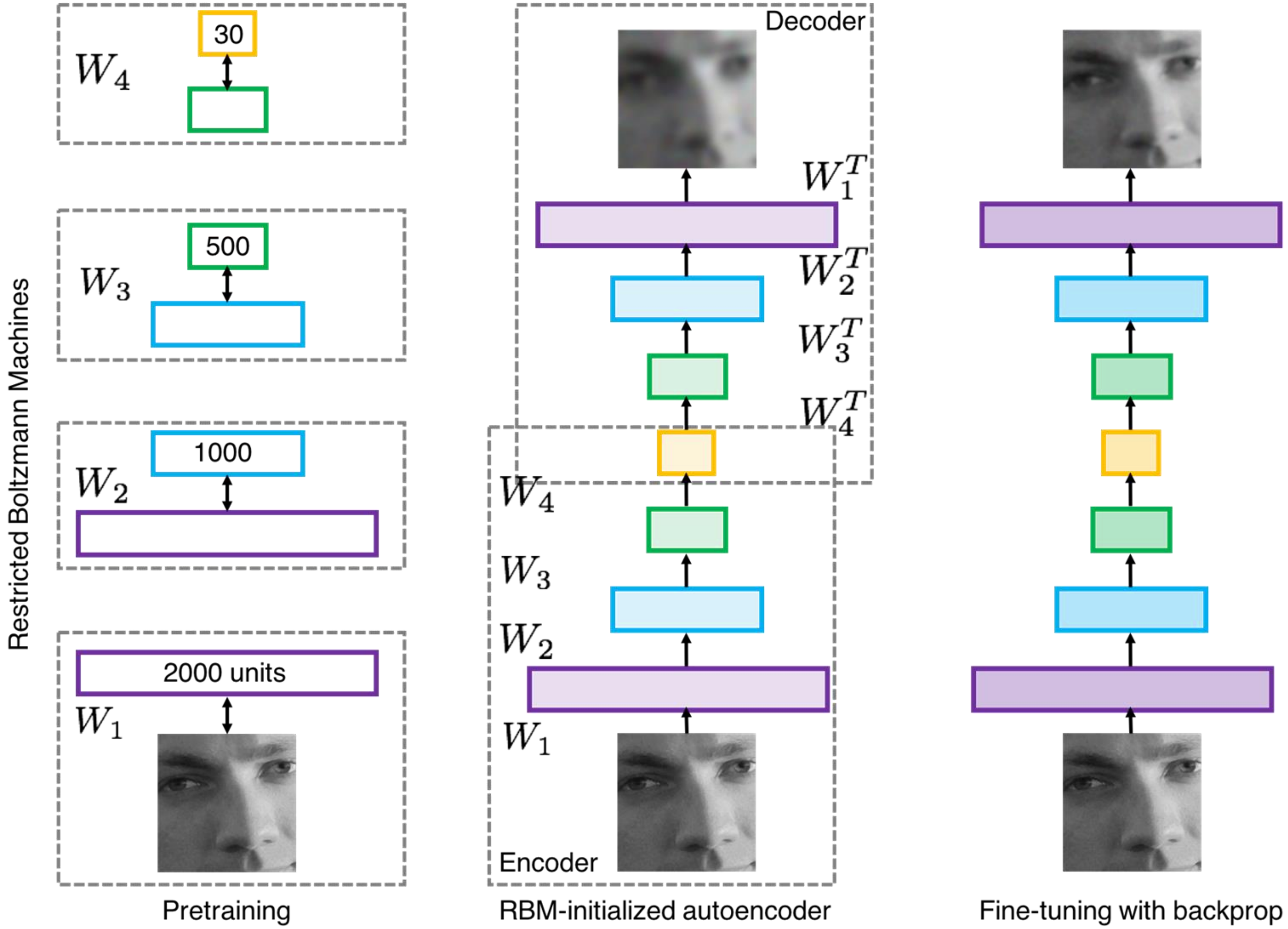


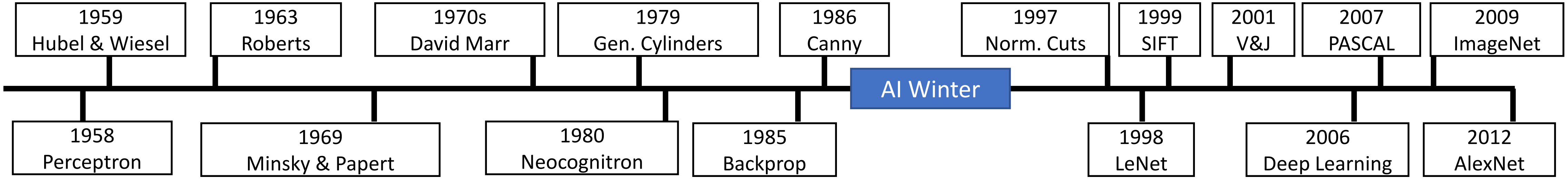
2000s: "Deep Learning"

People tried to train neural networks that were deeper and deeper

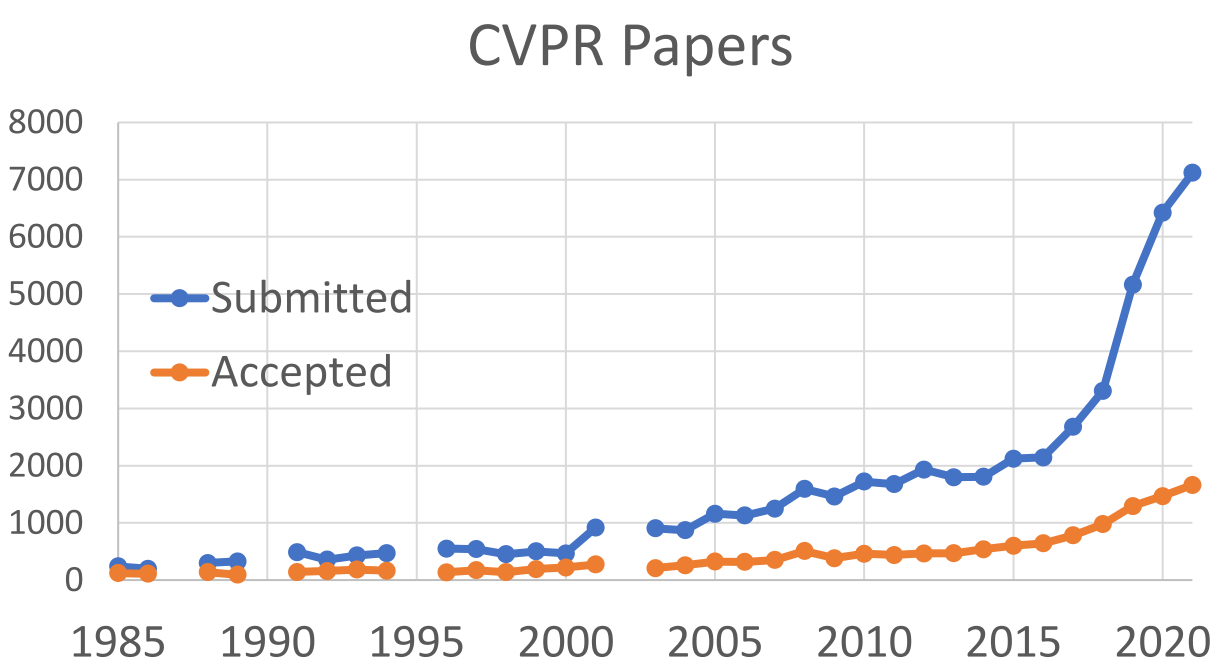
Not a mainstream research topic at this time

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

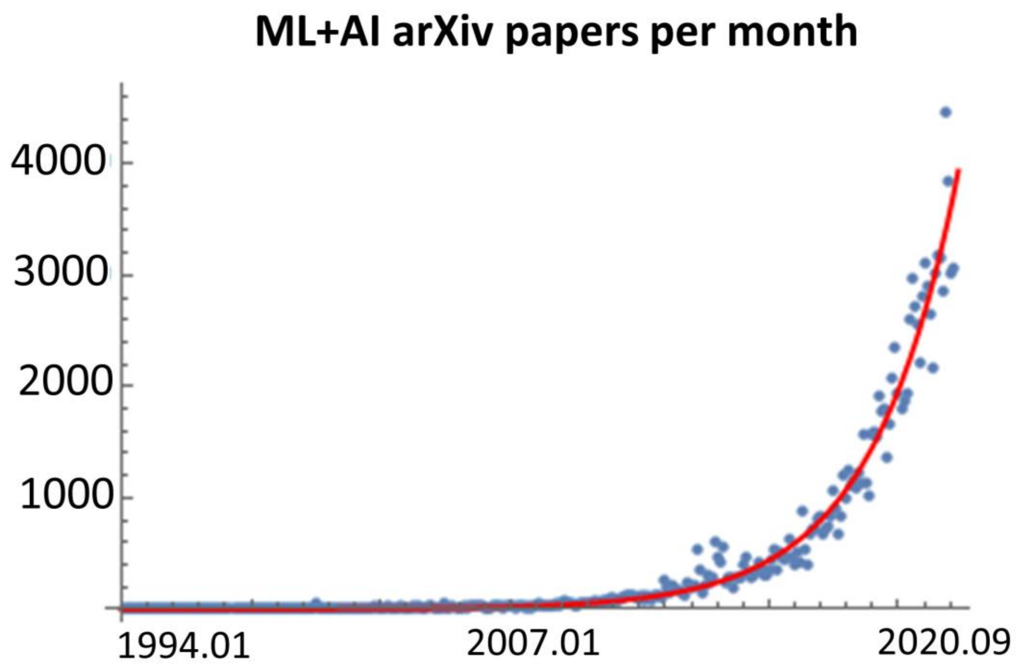




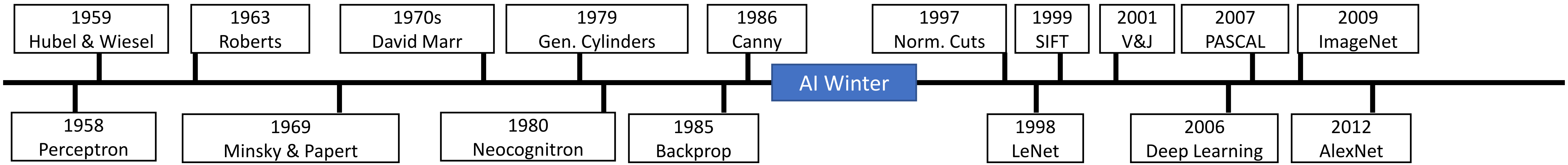
2012 to Present: Deep Learning Explosion



Publications at top Computer Vision conference



arXiv papers per month [(source)]



2012 to Present: Deep Learning is Everywhere

Image Classification

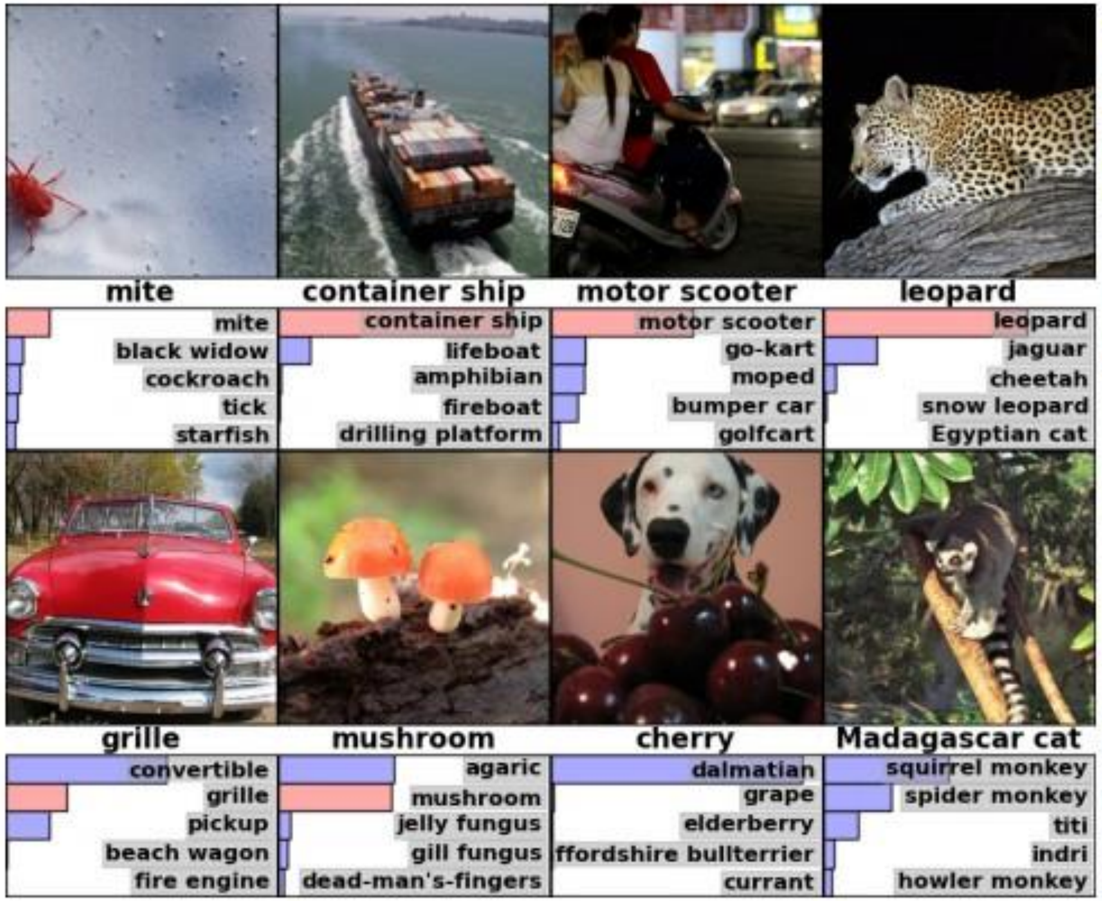
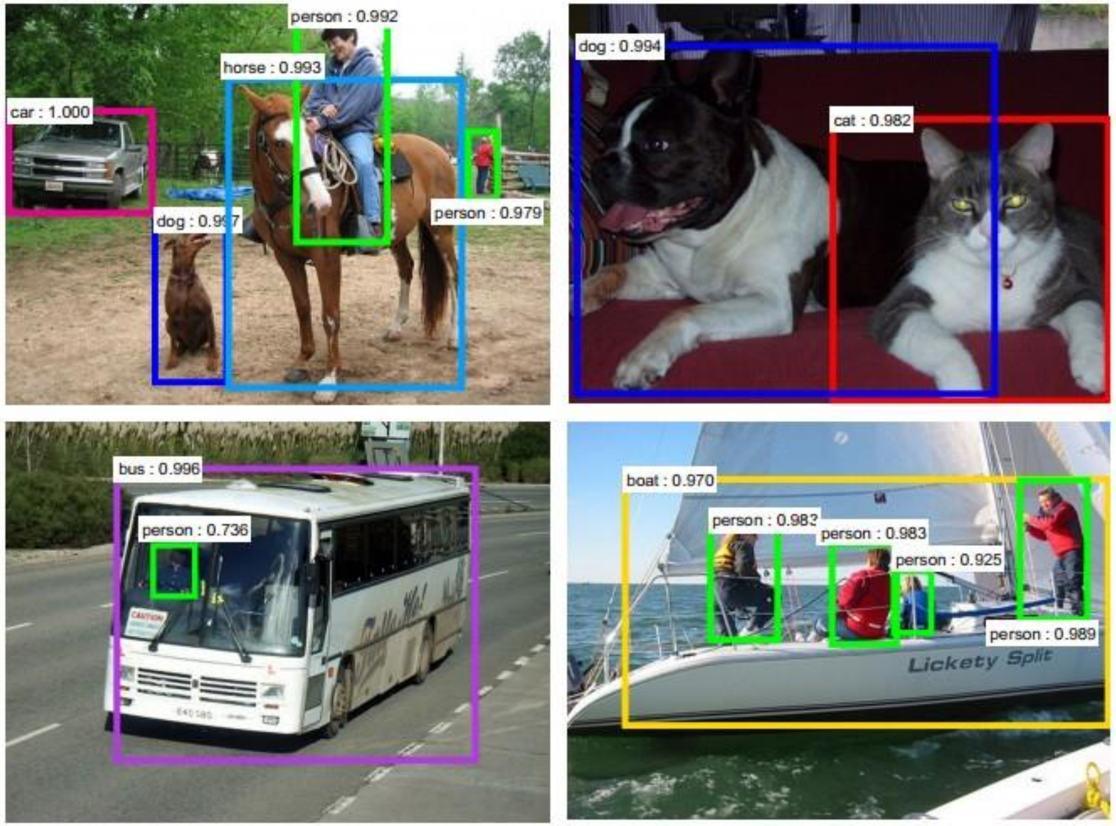


Image Retrieval

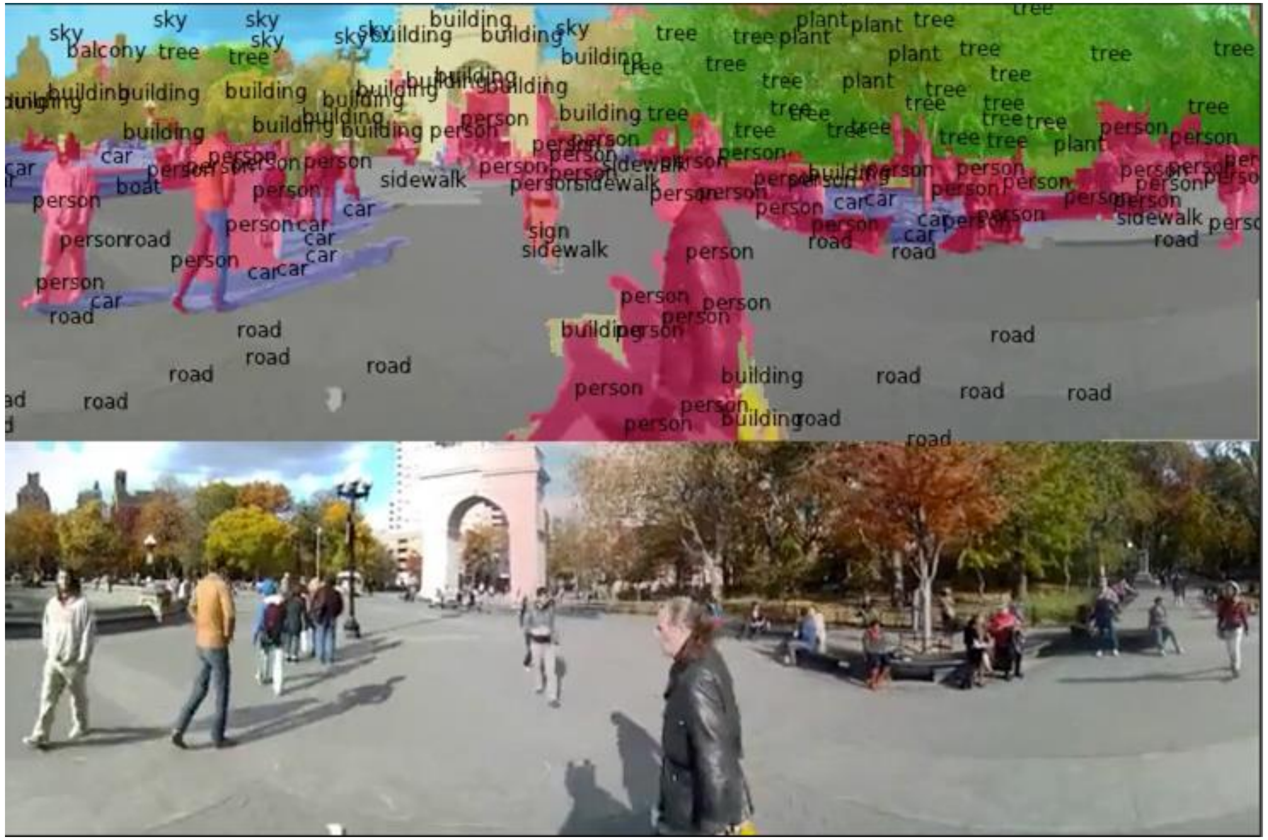
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

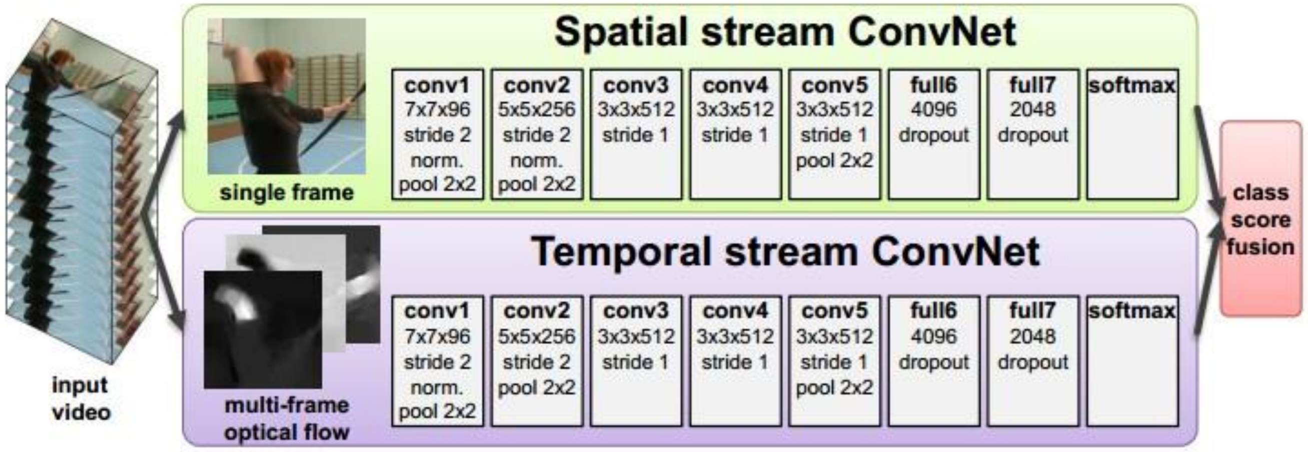
Image Segmentation



Fabaret et al, 2012

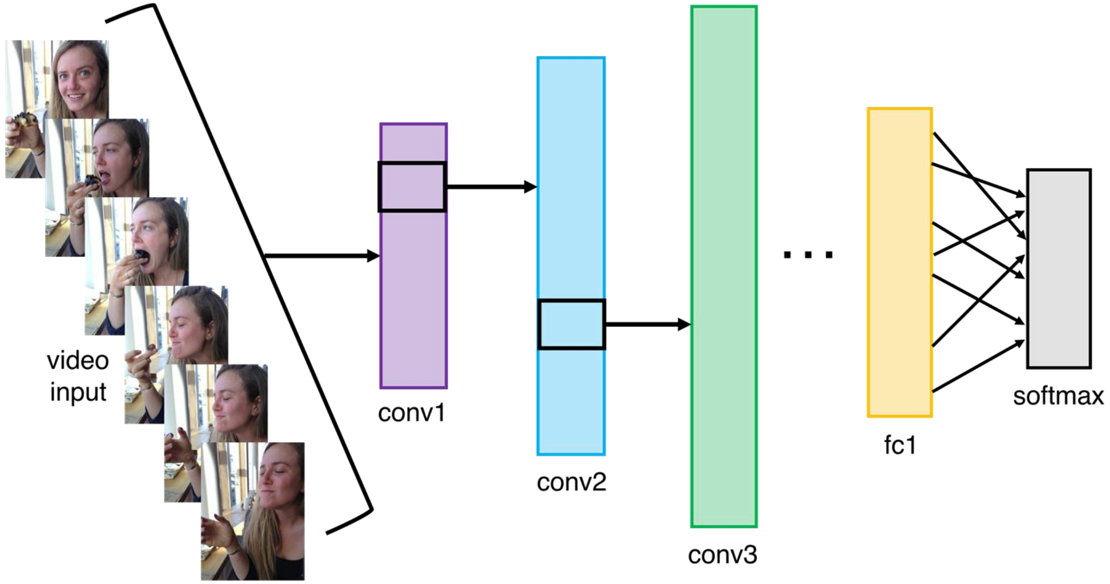
2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

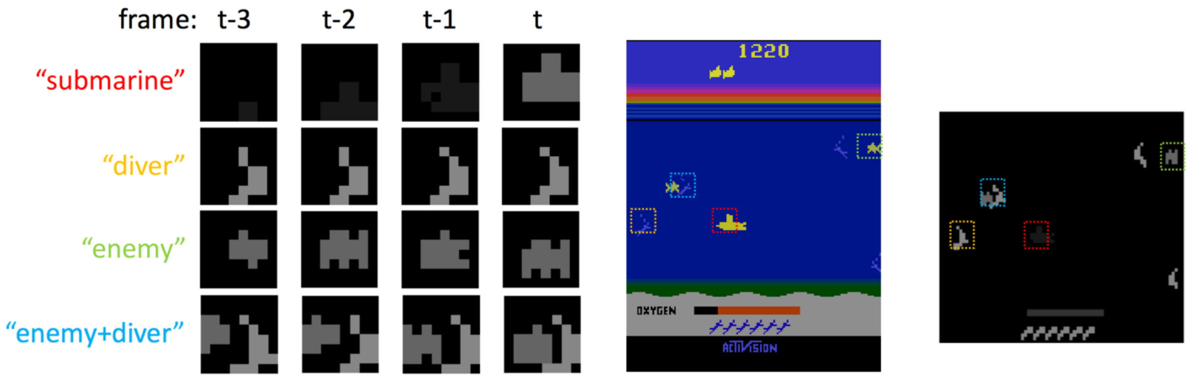
Activity Recognition

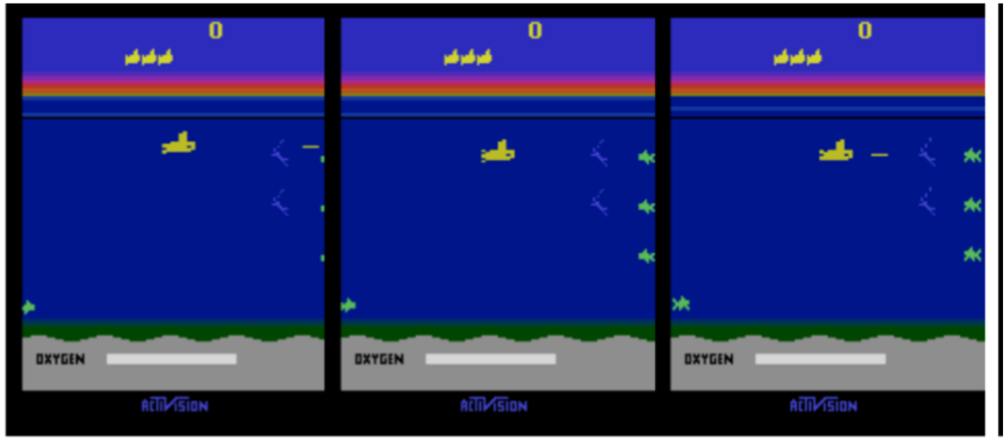


2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

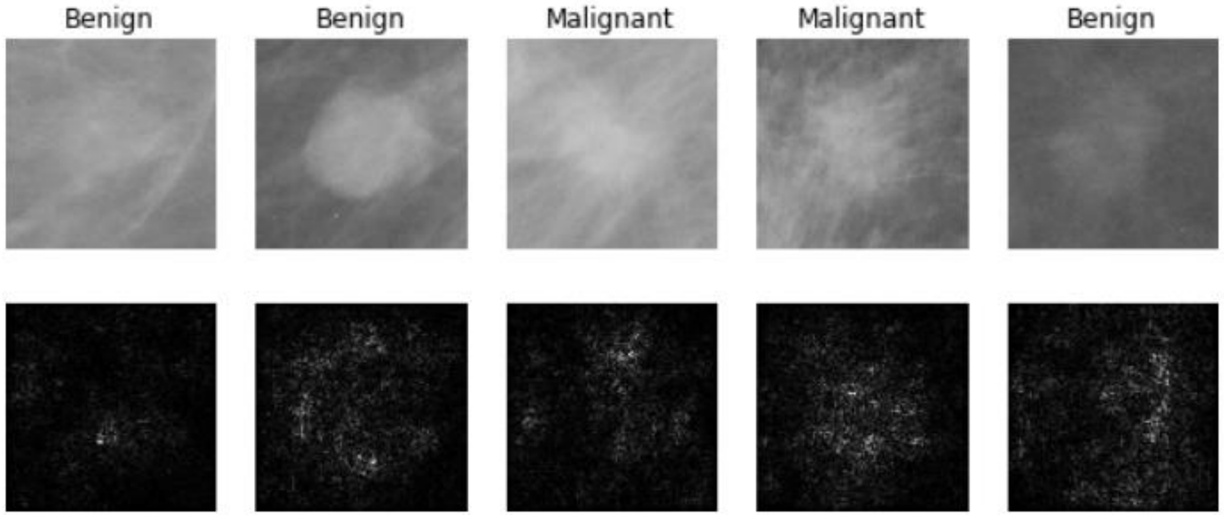
Playing Atari games (Guo et al, 2014)





2012 to Present: Deep Learning is Everywhere

Medical Imaging



Levy et al, 2016 Figure reproduced with permission

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

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

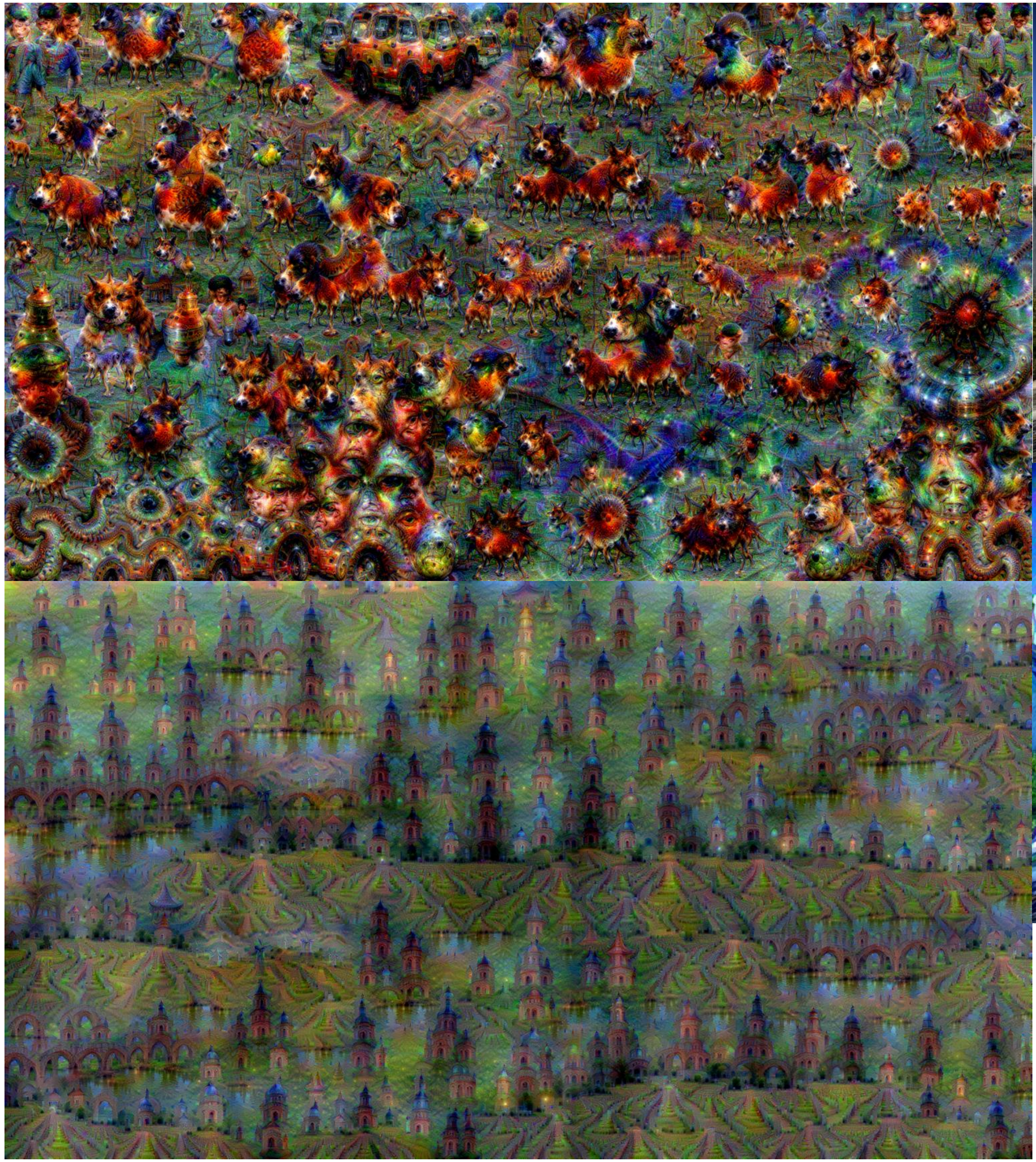
A woman is holding a cat in her hand

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



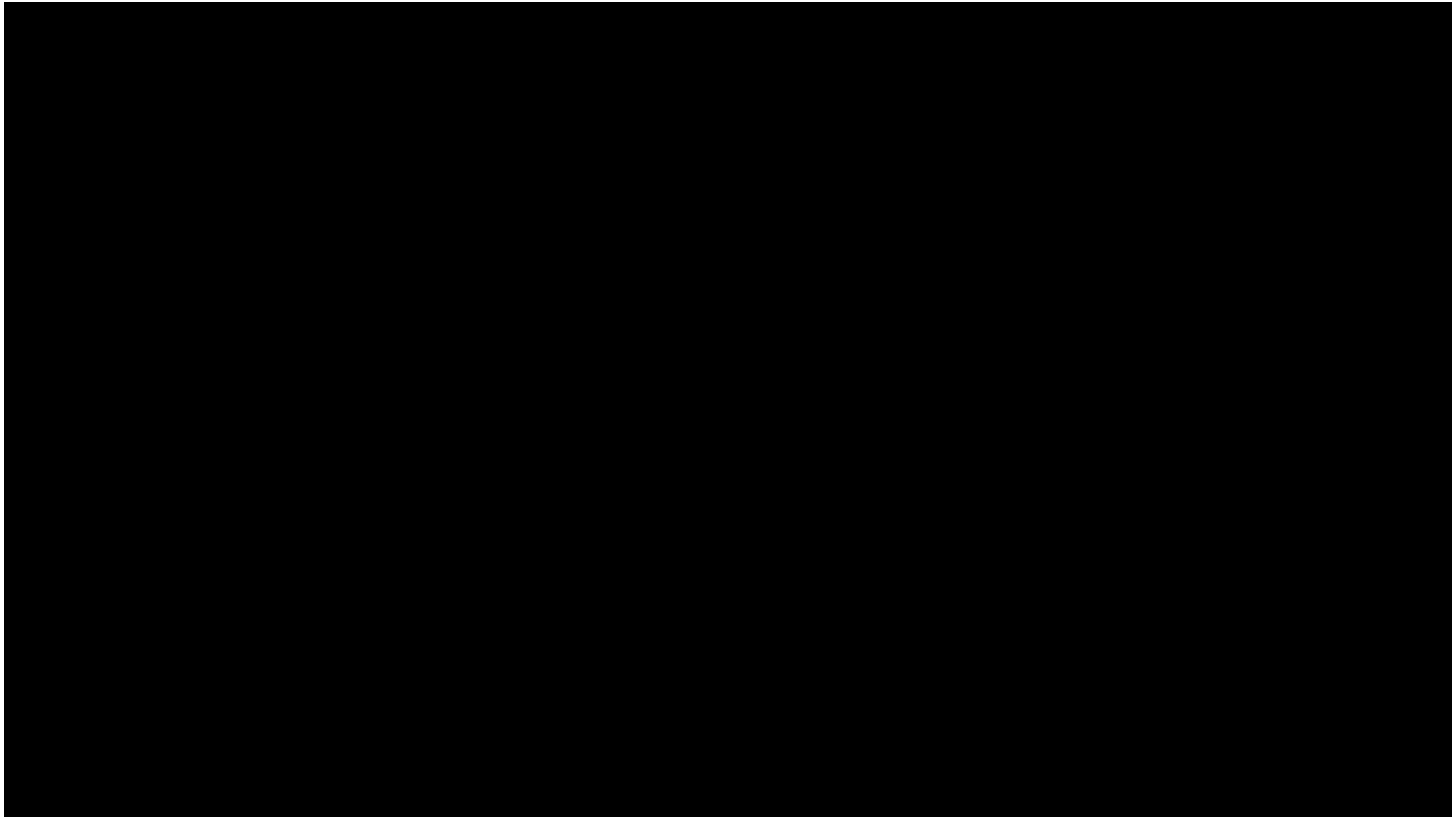
Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a blog post by Google Research.

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 Justin Johnson, 2017;
reproduced with permission

Mordvinsev et al, 2015

Gatys et al, 2016

2012 to Present: Deep Learning is Everywhere

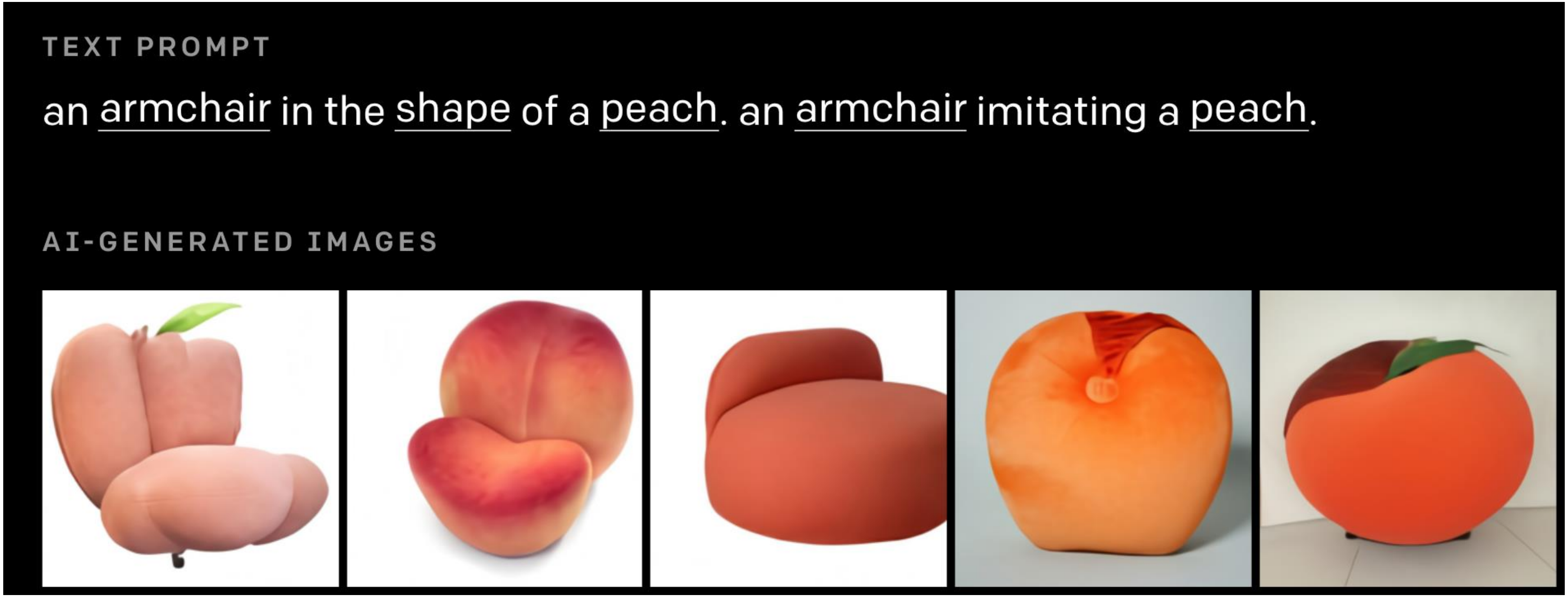


Karras et al, “Progressive Growing of GANs for Improved Quality, Stability, and Variation”, ICLR 2018

2012 to Present: Deep Learning is Everywhere

Ramesh et al, "DALL·E: Creating Images from Text", 2021. https://openai.com/blog/dall-e/

2012 to Present: Deep Learning is Everywhere



Ramesh et al, “DALL·E: Creating Images from Text”, 2021. https://openai.com/blog/dall-e/

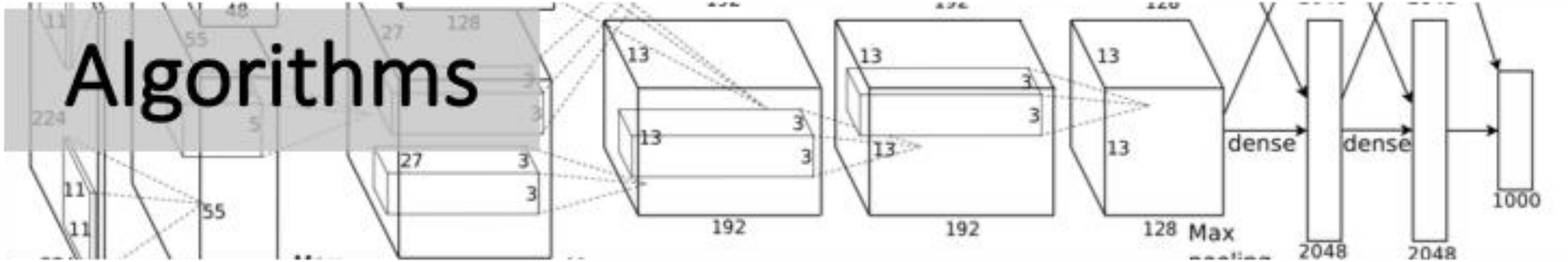
2012 to Present: Deep Learning is Everywhere

A rabbit detective sitting on a park bench and reading a newspaper in a victorian setting

A shark and a dolphin cruise hand-in-hand with an undersea city in the background

Robot dinosaurs versus monster trucks in the colosseum

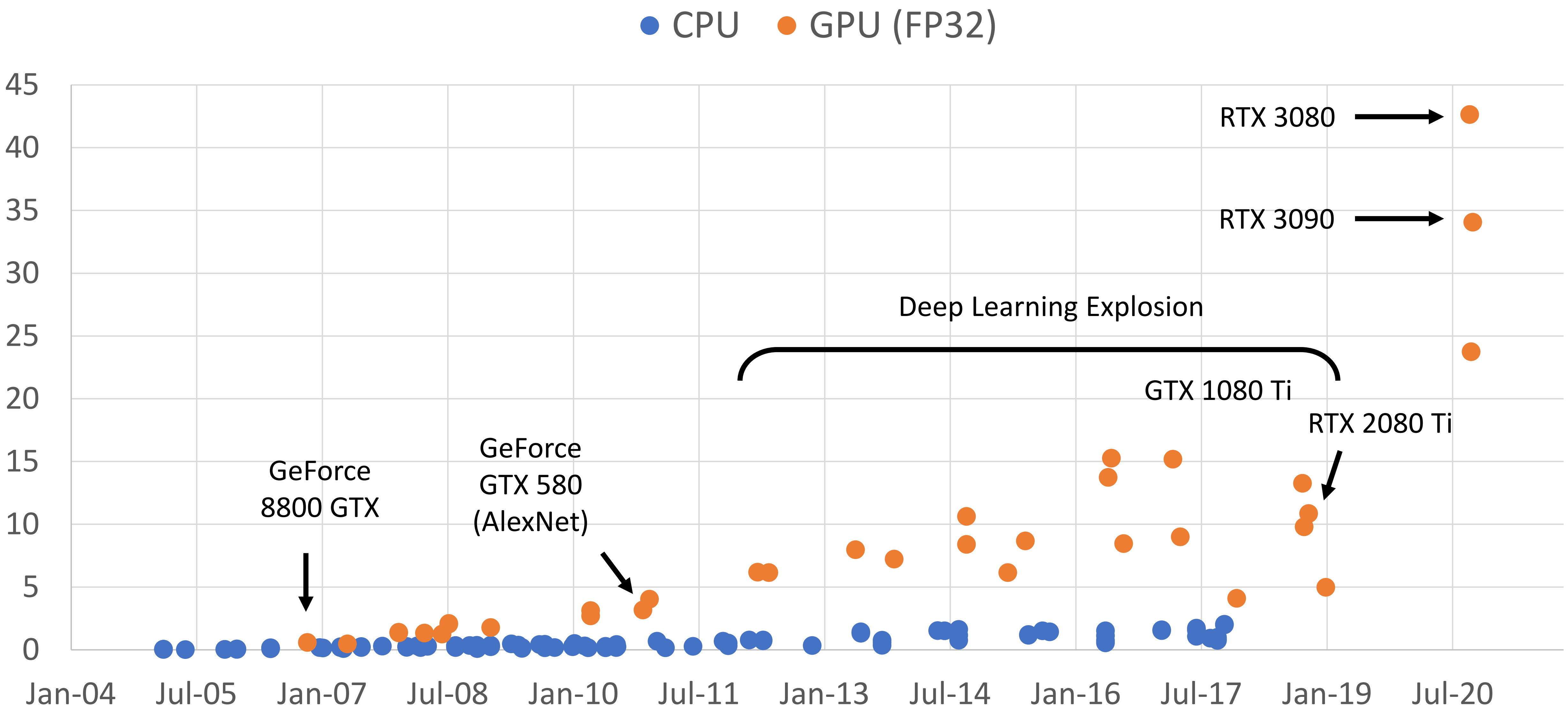
Ramesh et al, "DALL·E 2", 2022. https://openai.com/dall-e-2/
Source: https://twitter.com/sama/status/1511724264629678084


Algorithms
13
3
27
11
55
192
128
Max
dense
2048
1000

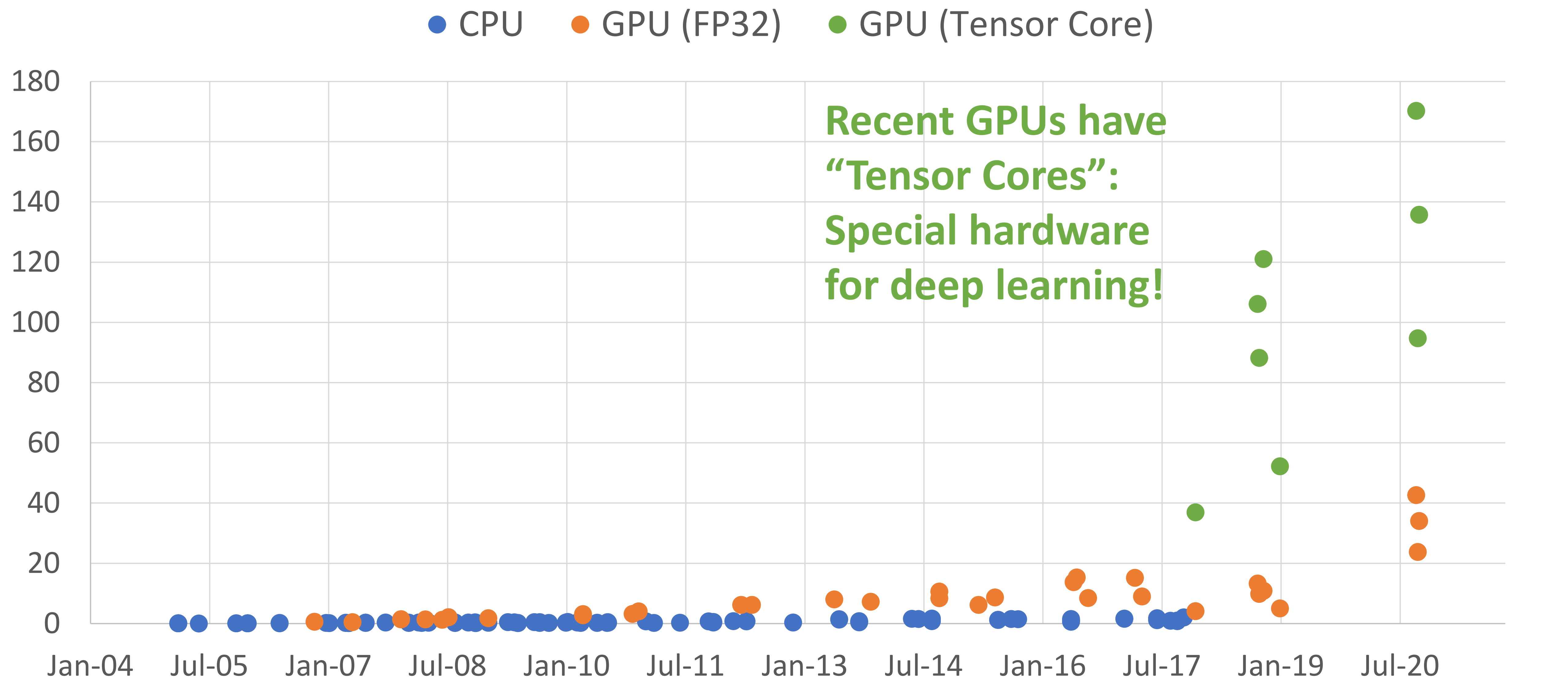

Data


Computation
GTX
GTX

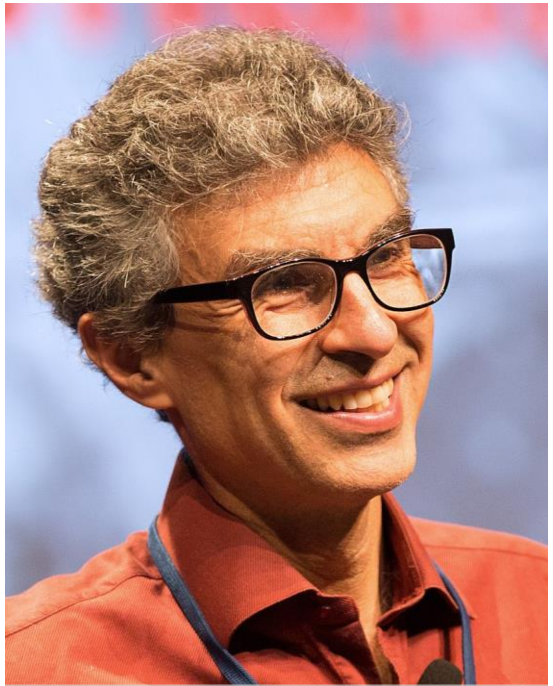
GFLOP per Dollar



GFLOP per Dollar



2018 Turing Award


Yoshua Bengio


Geoffrey Hinton

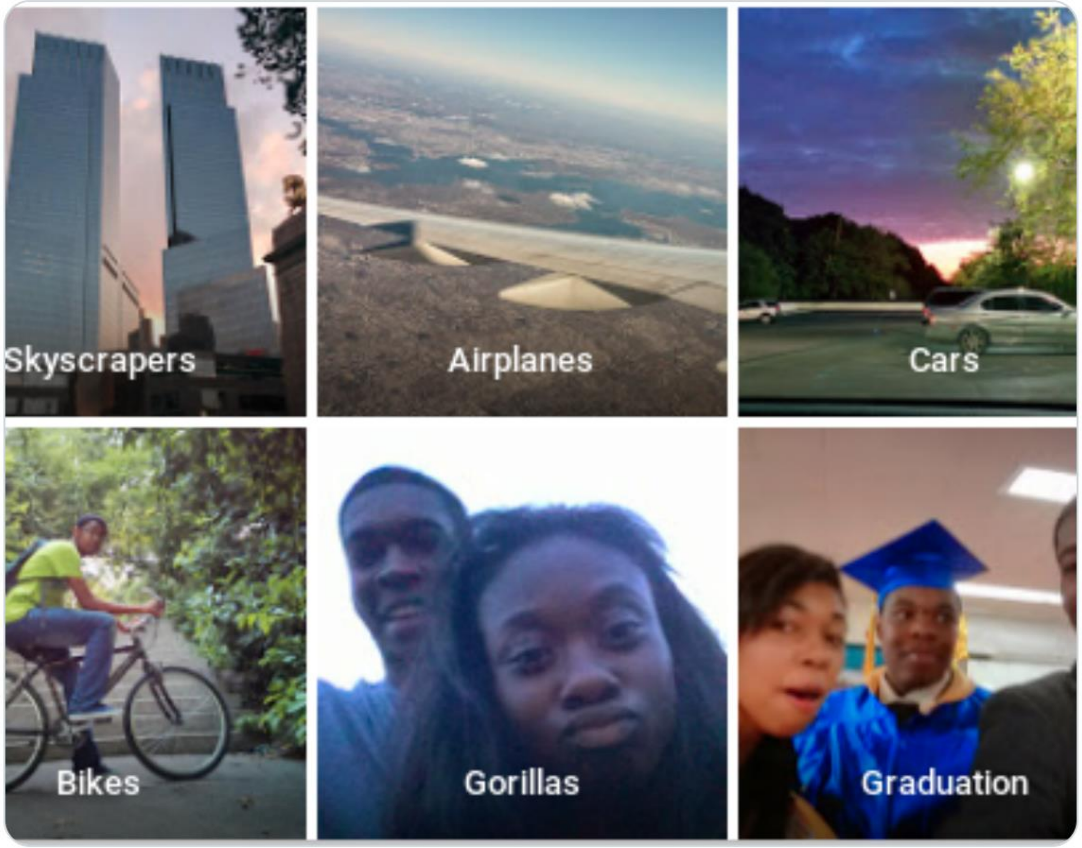

Yann LeCun

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

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

Computer Vision Can Cause Harm

Harmful Stereotypes



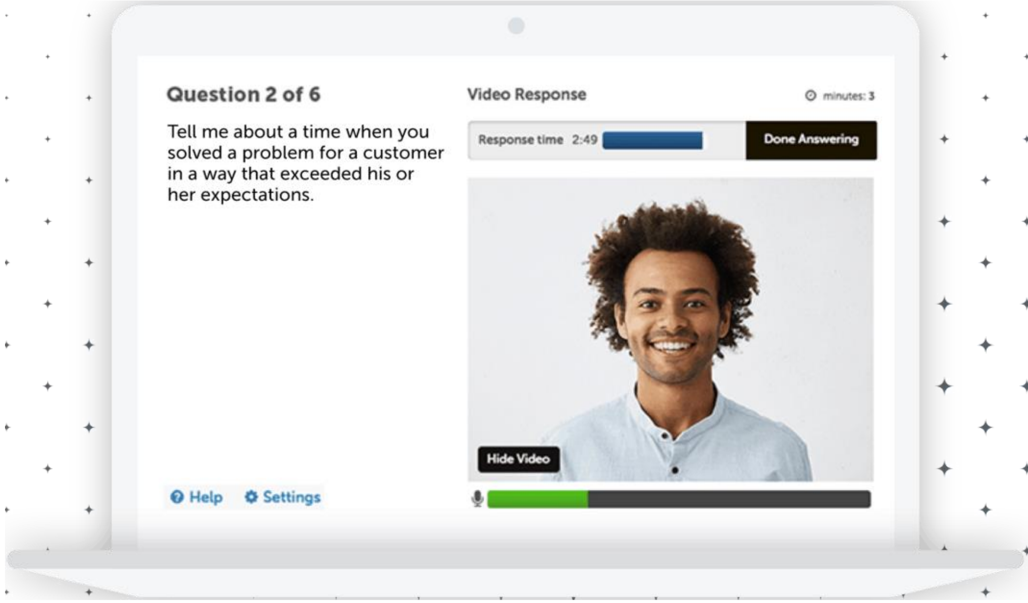
Barocas et al, “The Problem With Bias: Allocative Versus Representational Harms in Machine Learning”, SIGCIS 2017
Kate Crawford, “The Trouble with Bias”, NeurIPS 2017 Keynote
Source: https://twitter.com/jackyalcine/status/615329515909156865 (2015)

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: Timnit Gebru

This image is copyright-free United States government work

Example credit: Andrej Karpathy

Computer Vision Technology Can Better Our Lives