Convolutional Neural Networks – Lecture 1

Sagie Benaim and Tomer Galanti

Tel Aviv University

Slides based on cs231n lecture: http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture01.pdf and Deep Learning book: https://www.deeplearningbook.org/contents/intro.html

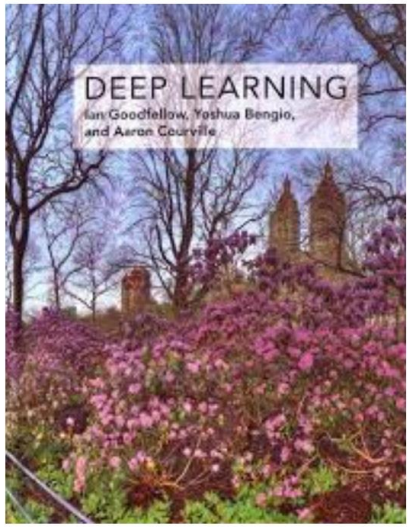
- Sagie Benaim: sagieb@mail.tau.ac.il

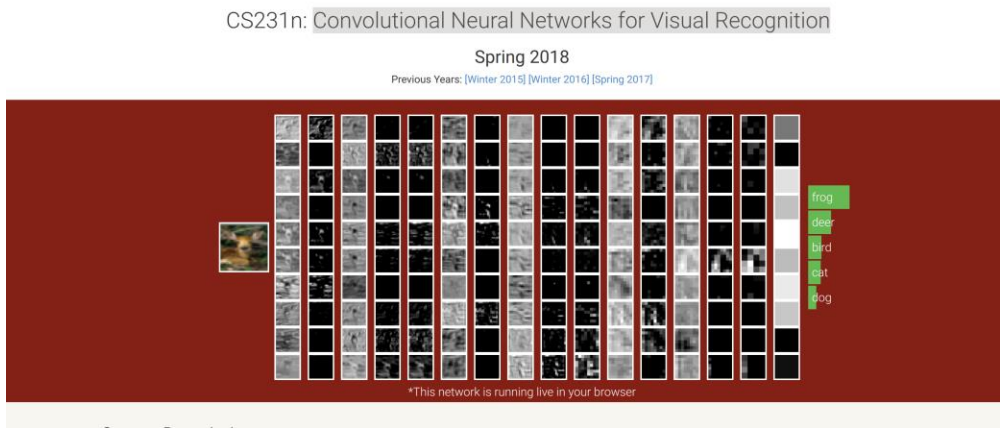
  Tomer Galanti: tomer22g@gmail.com
- Reception: Sundays 10am to 11am (please coordinate in advance via email)
- Lectures, exercises and announcements are on moodle.
- Class Exam: 70%, Homework: 30%
- 4 Homework Sheets in group of up to 3 students (Submit on moodle)

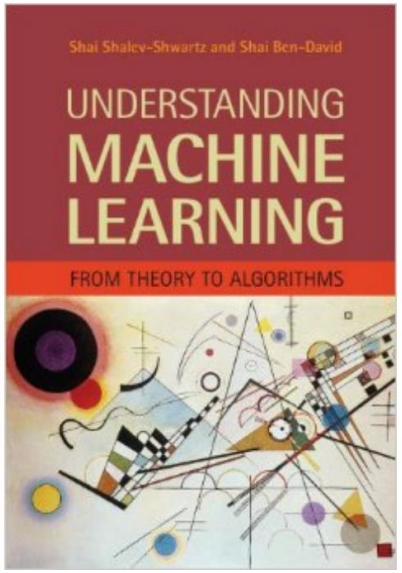
Prerequisites

- Exercises in Python: Good programming knowledge required
- Linear Algebra, College Calculus
- Useful but not required: Introduction to Machine Learning

- Course is based on:
  - Standford's CS231n: Convolutional Neural Networks for Visual Recognition, http://cs231n.stanford.edu/.
  - Deep Learning Book: https://www.deeplearningbook.org/
  - Understanding Machine Learning: From Theory to Algorithms. Useful for further theoretical understanding.
  - Many slides and material are taken from these sources







What is Deep Learning?

Representations Matter

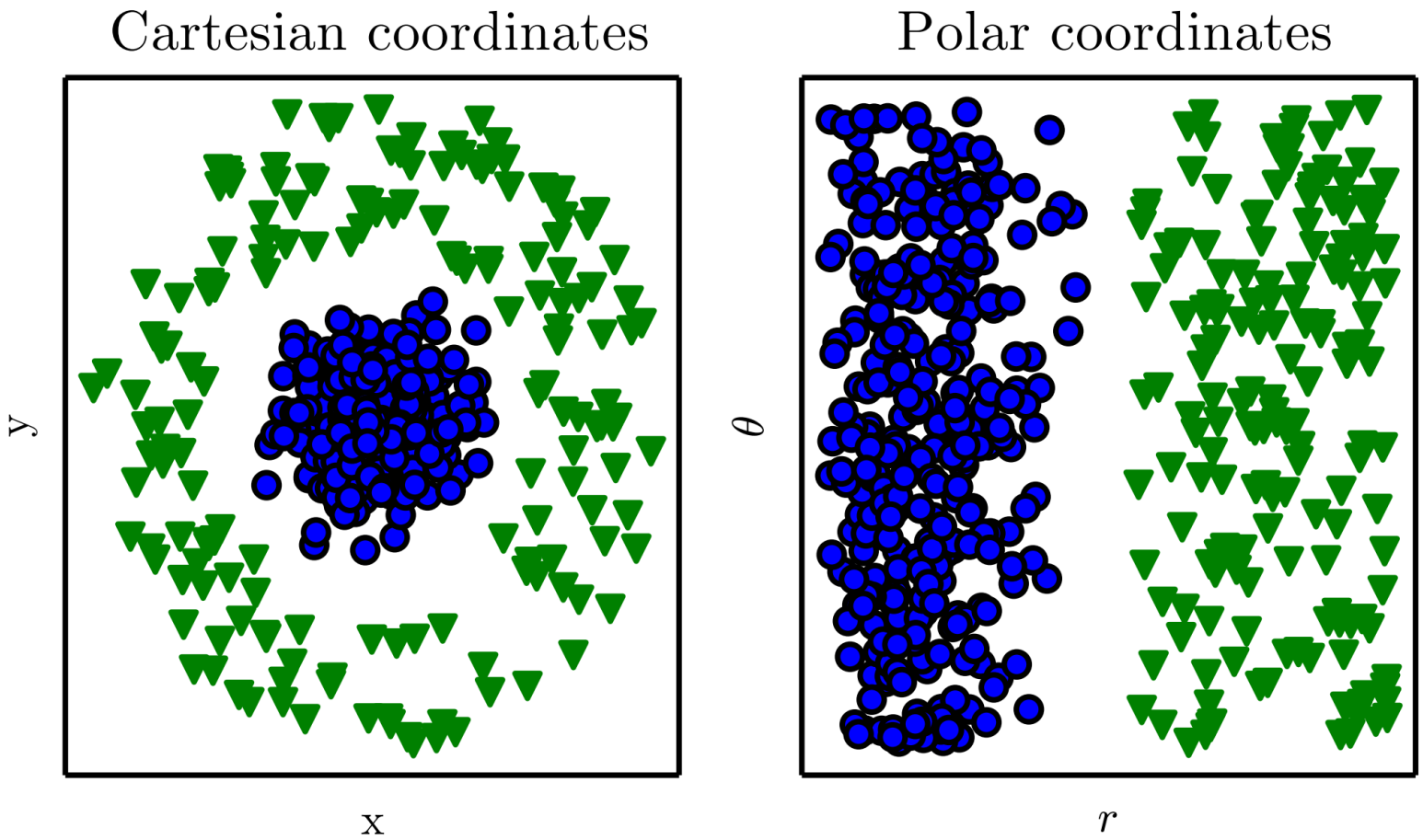
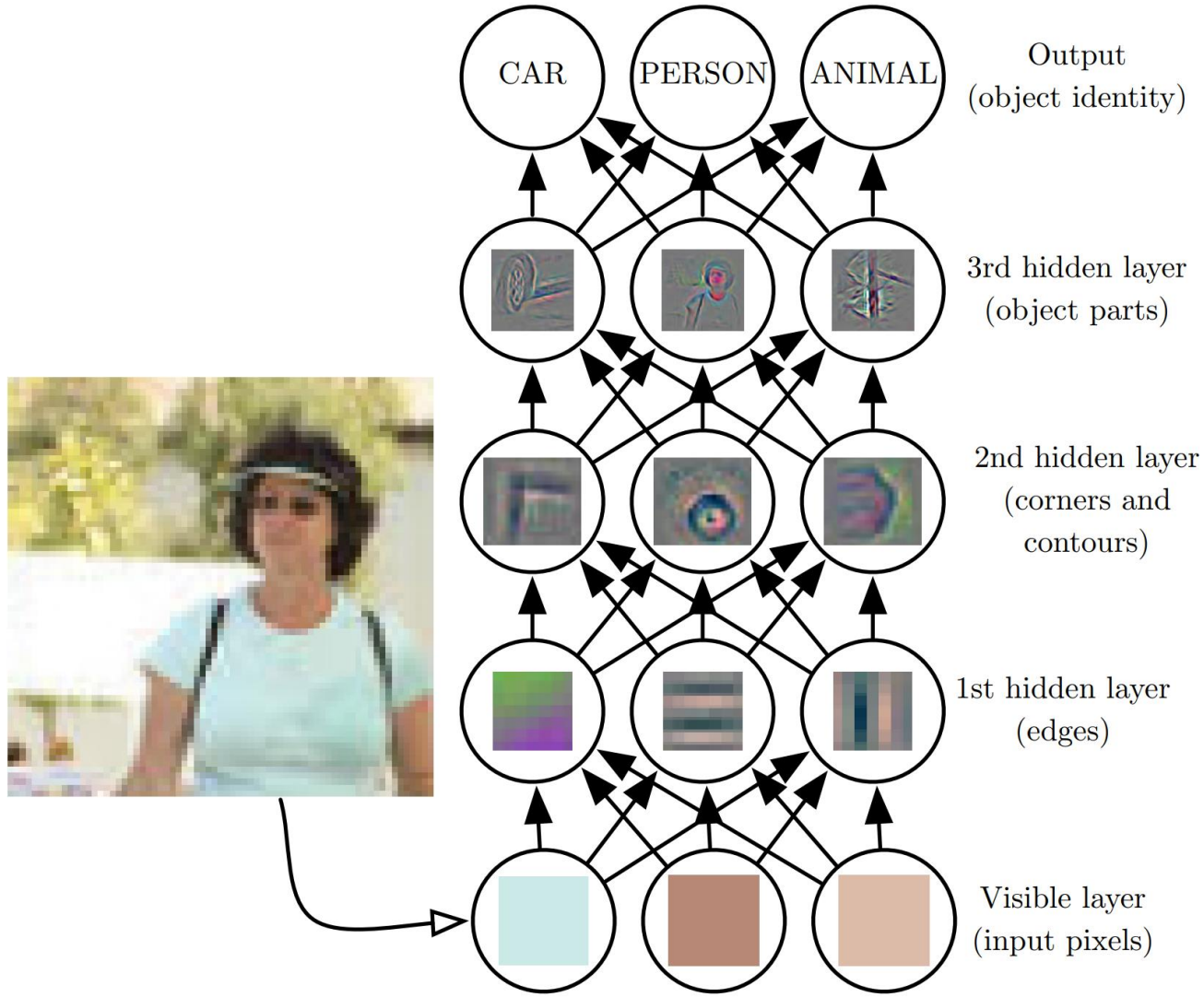


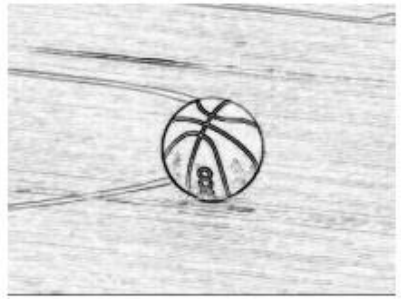
Figure 1.1

Depth: Repeated Composition

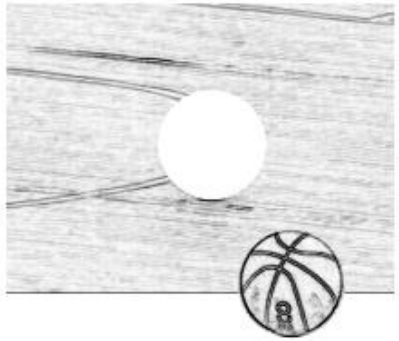


Input image

Edge image



2 ½-D sketch



3-D model

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

Machine Learning and AI

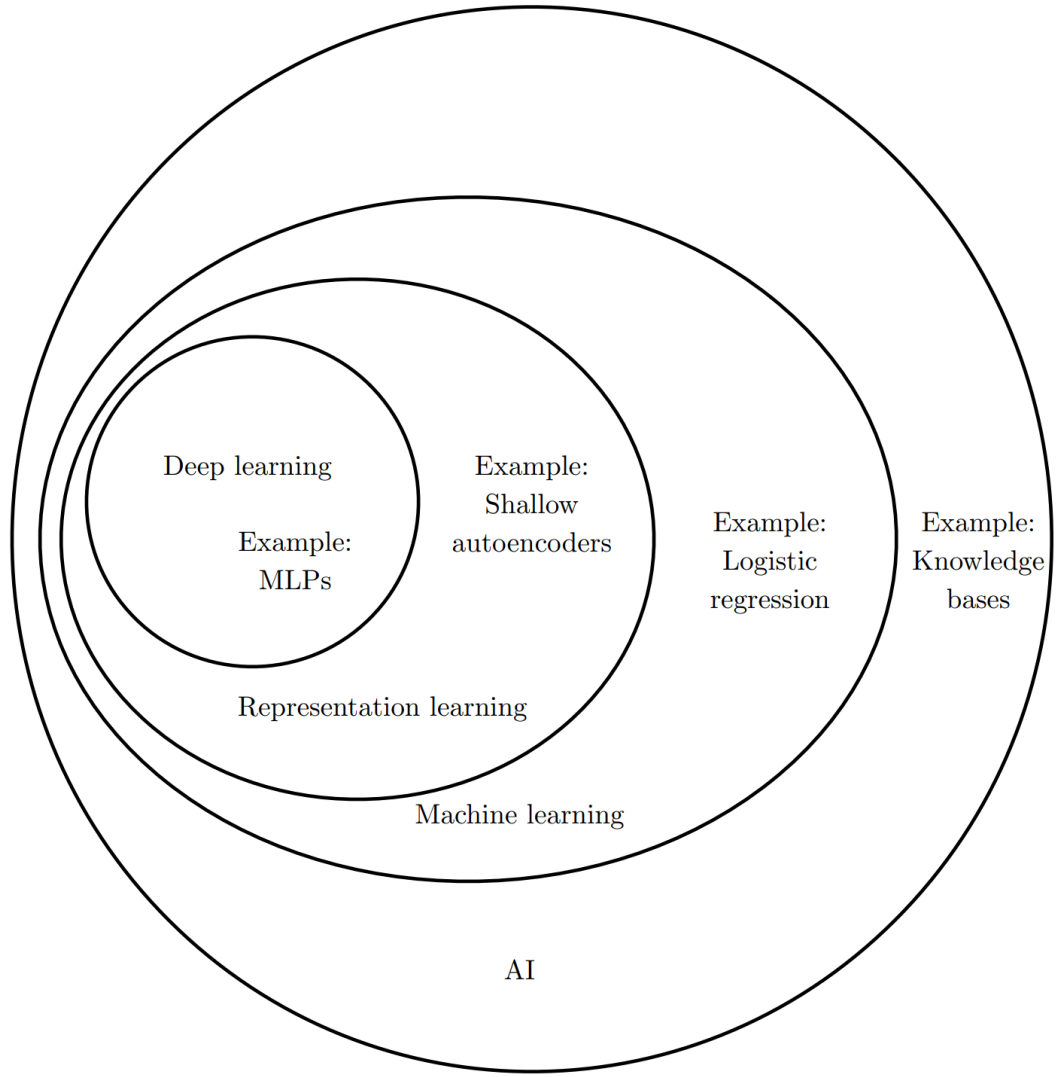
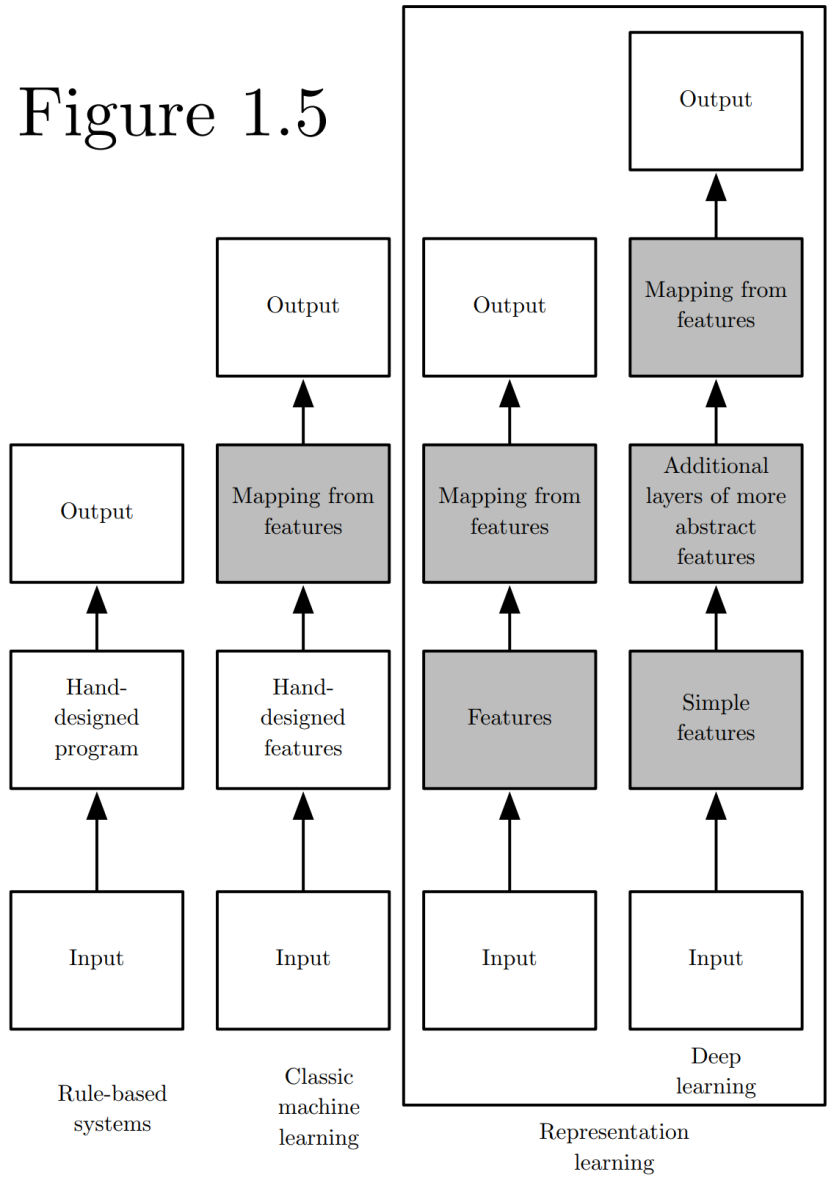


Figure 1.4

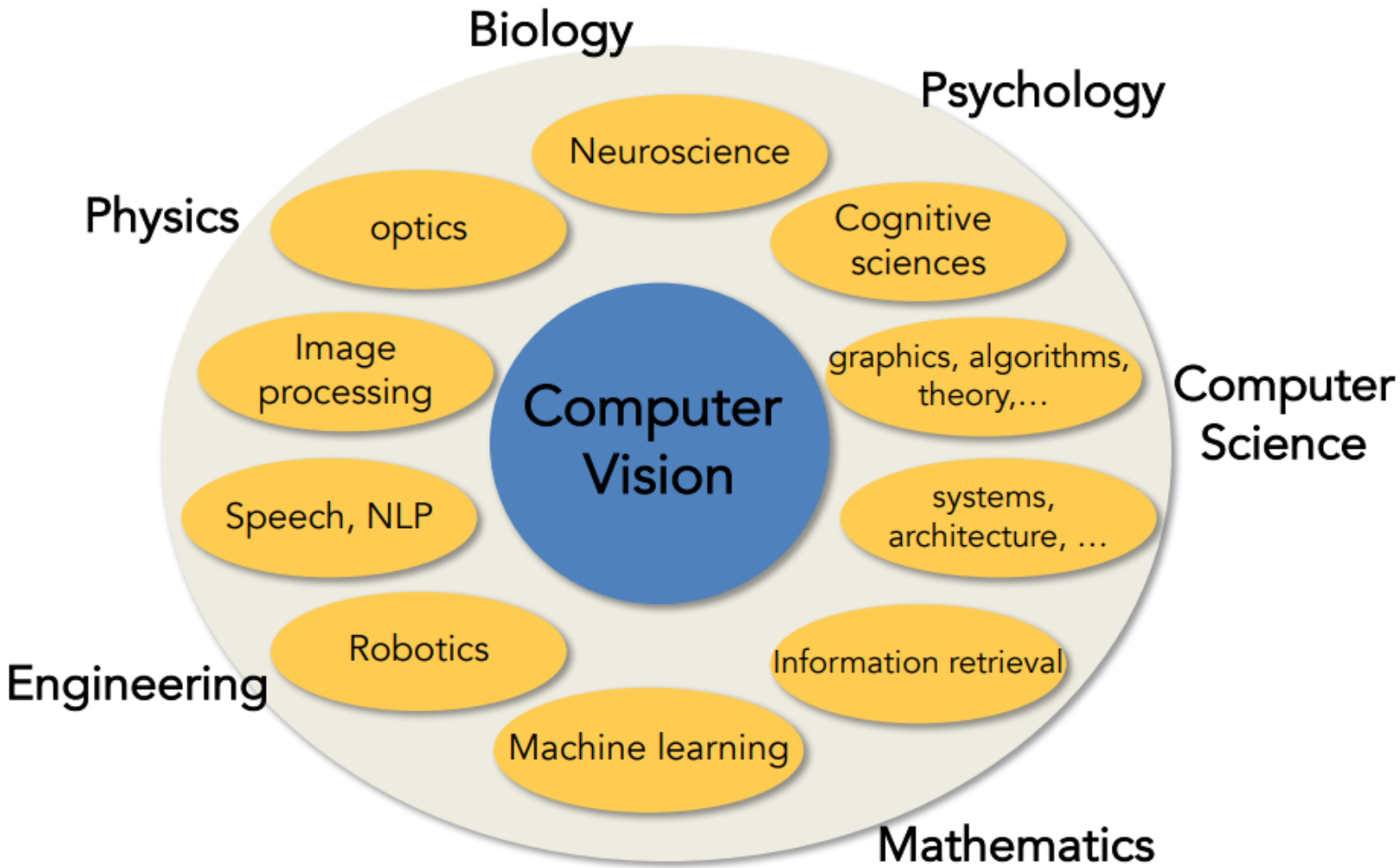
(Goodfellow 2016)

Learning Multiple Components

Figure 1.5



(Goodfellow 2016)


Biology
Psychology
Physics
Neuroscience
optics
Cognitive sciences
Image processing
Computer Vision
graphics, algorithms, theory,…
Computer Science
Speech, NLP
systems, architecture, …
Robotics
Information retrieval
Engineering
Machine learning
Mathematics

Why Do We Need It?

Machine learning tasks:

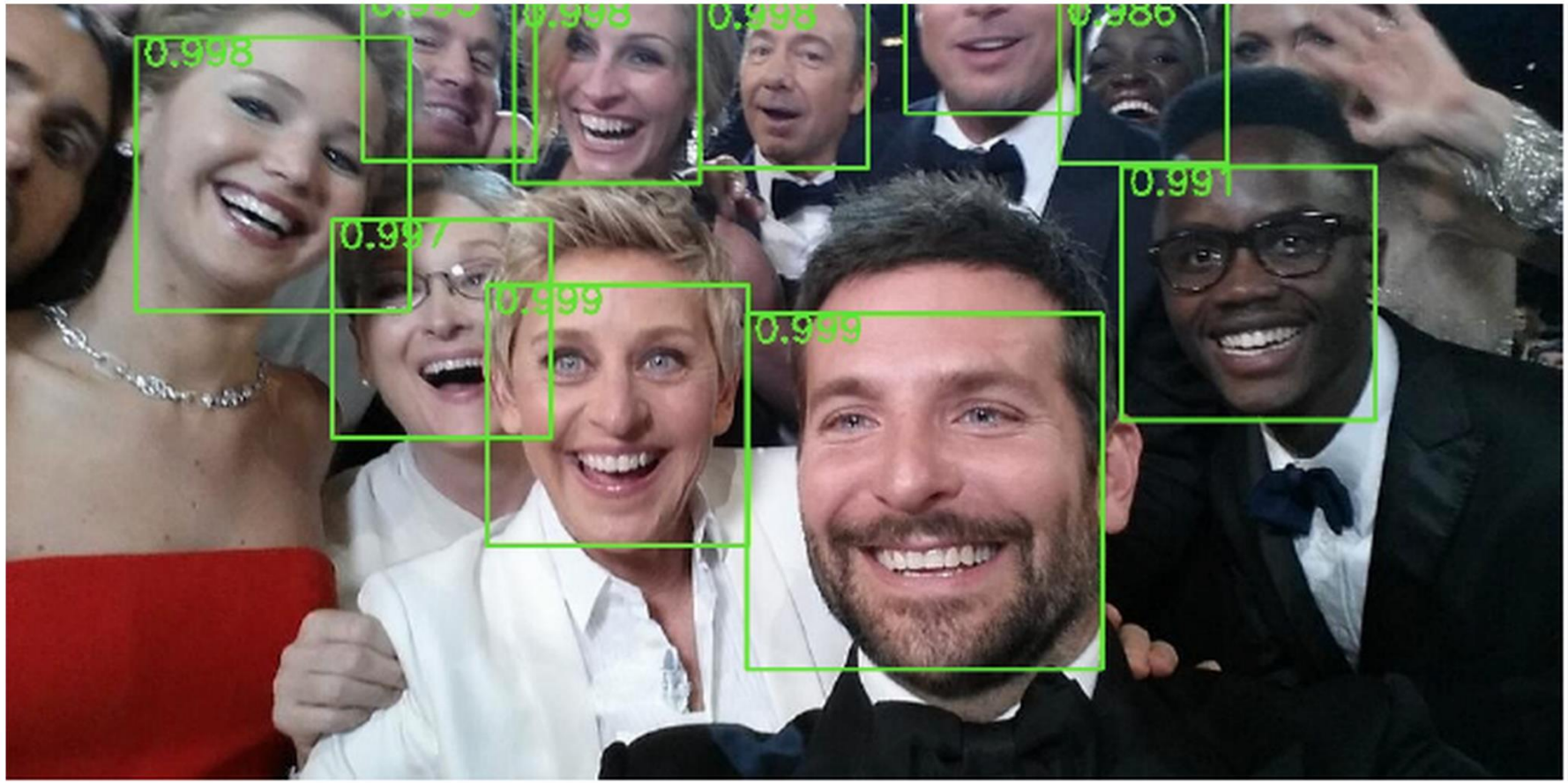
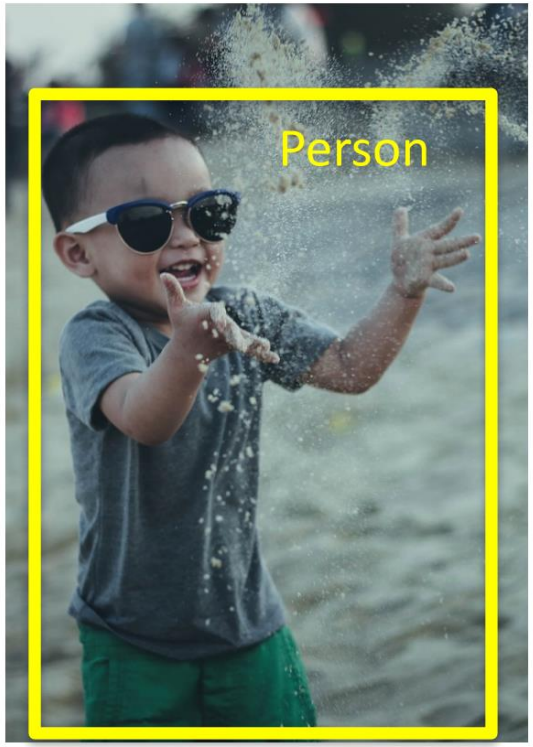
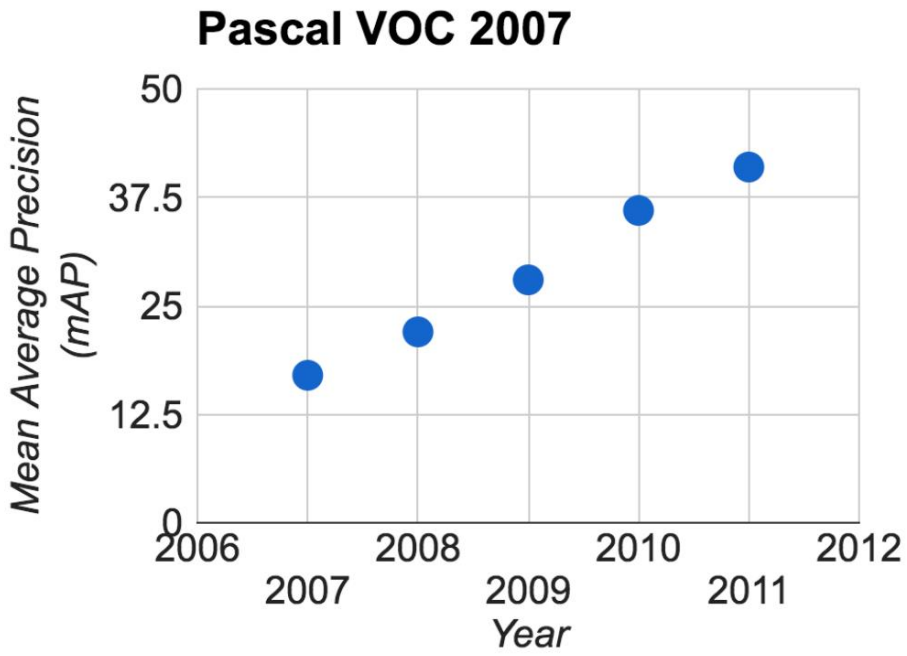


Figure: Image Recognition.

PASCAL Visual Object Challenge (20 object categories)

[Everingham et al. 2006-2012]





IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

✔

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

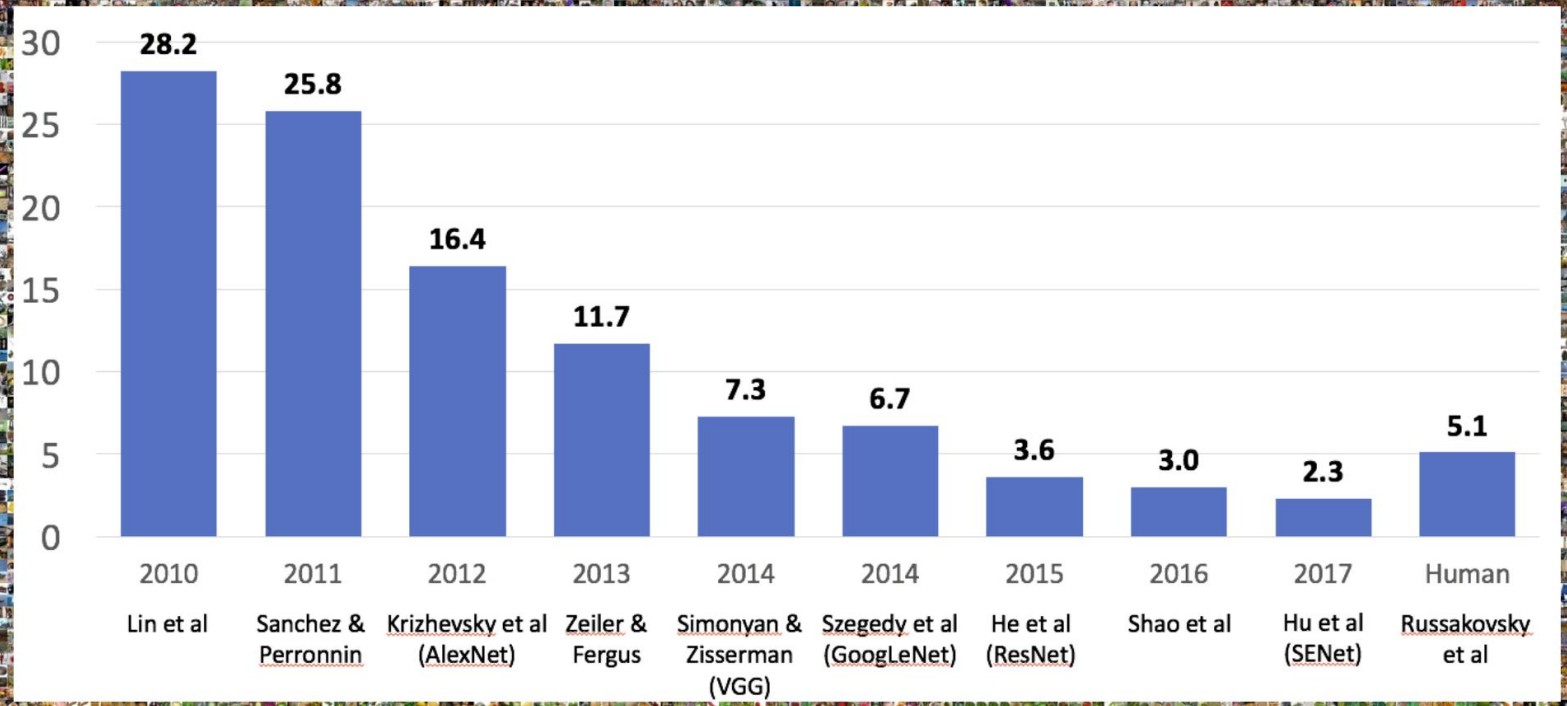
✗

IMAGENET Large Scale Visual Recognition Challenge

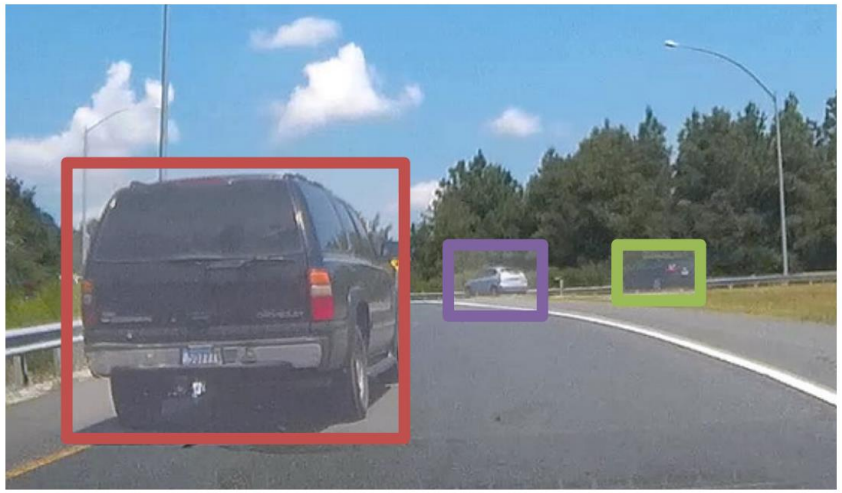
The Image Classification Challenge:

1,000 object classes

1,431,167 images



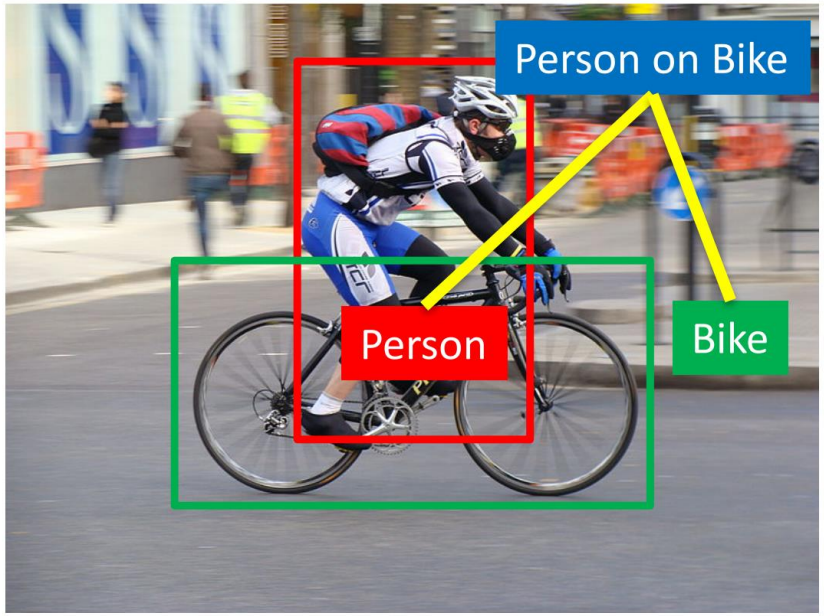
Russakovsky et al. IJCV 2015

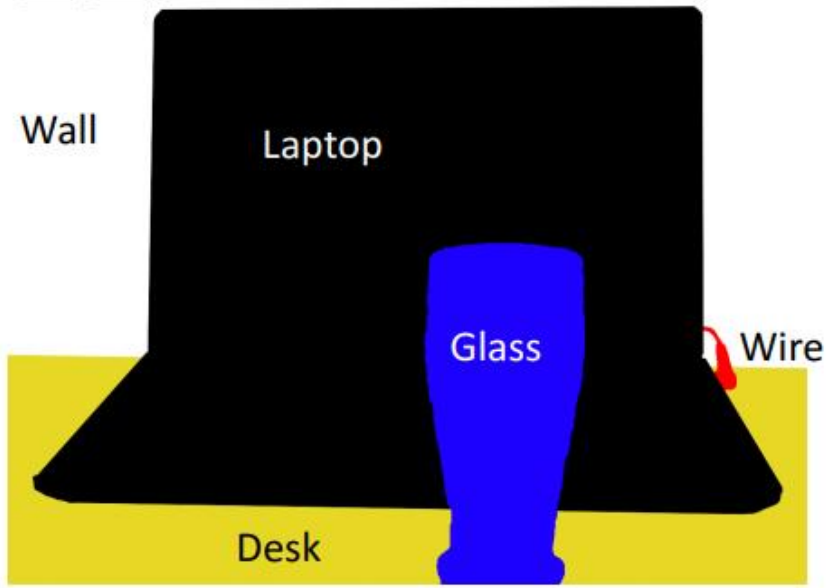


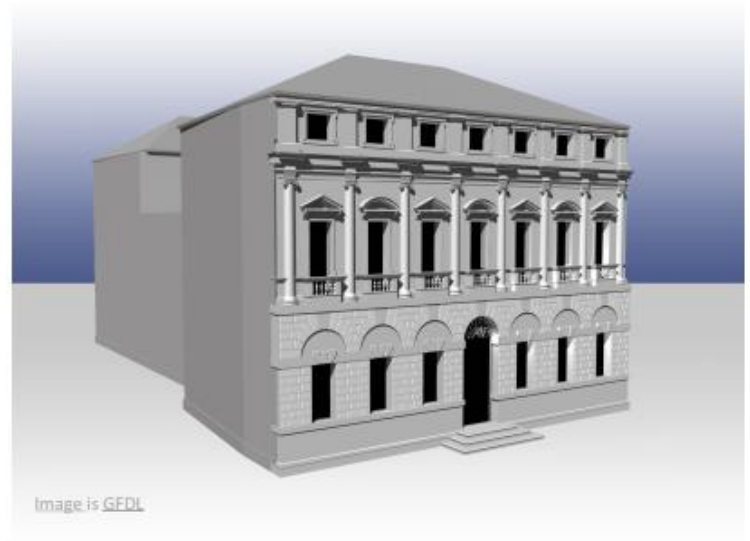
- Object detection
- Action classification
- Image captioning
- …

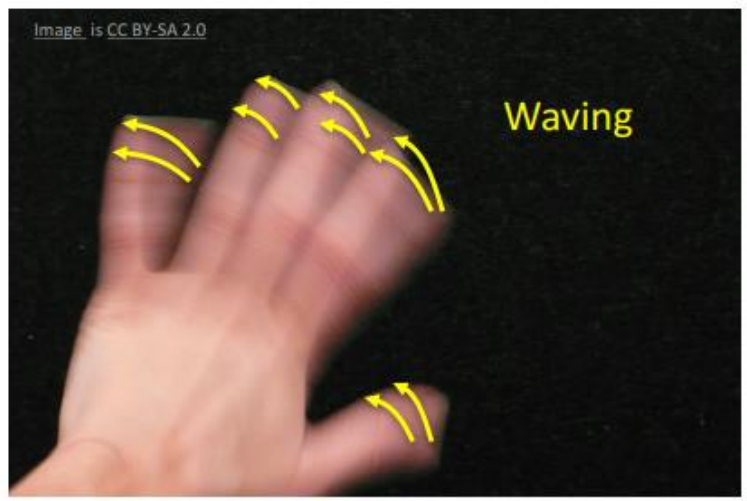
Person

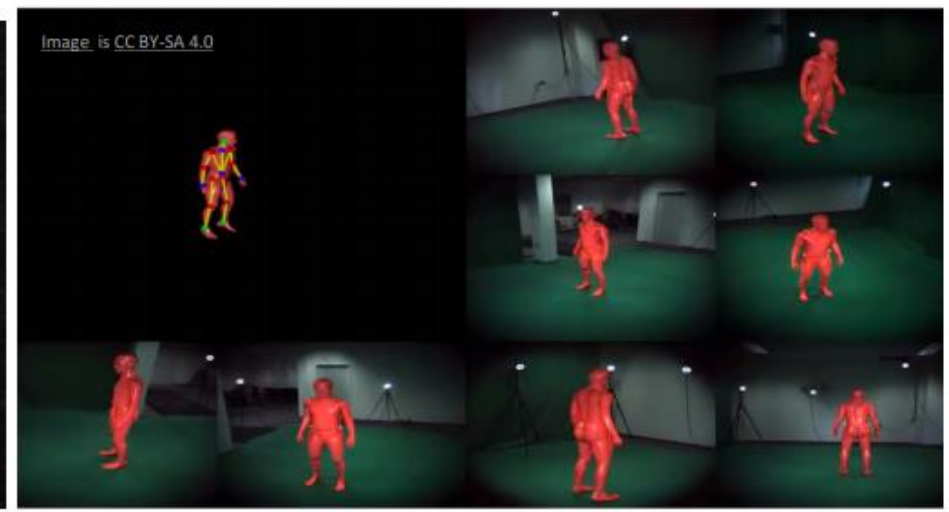
Hammer











Machine learning tasks:

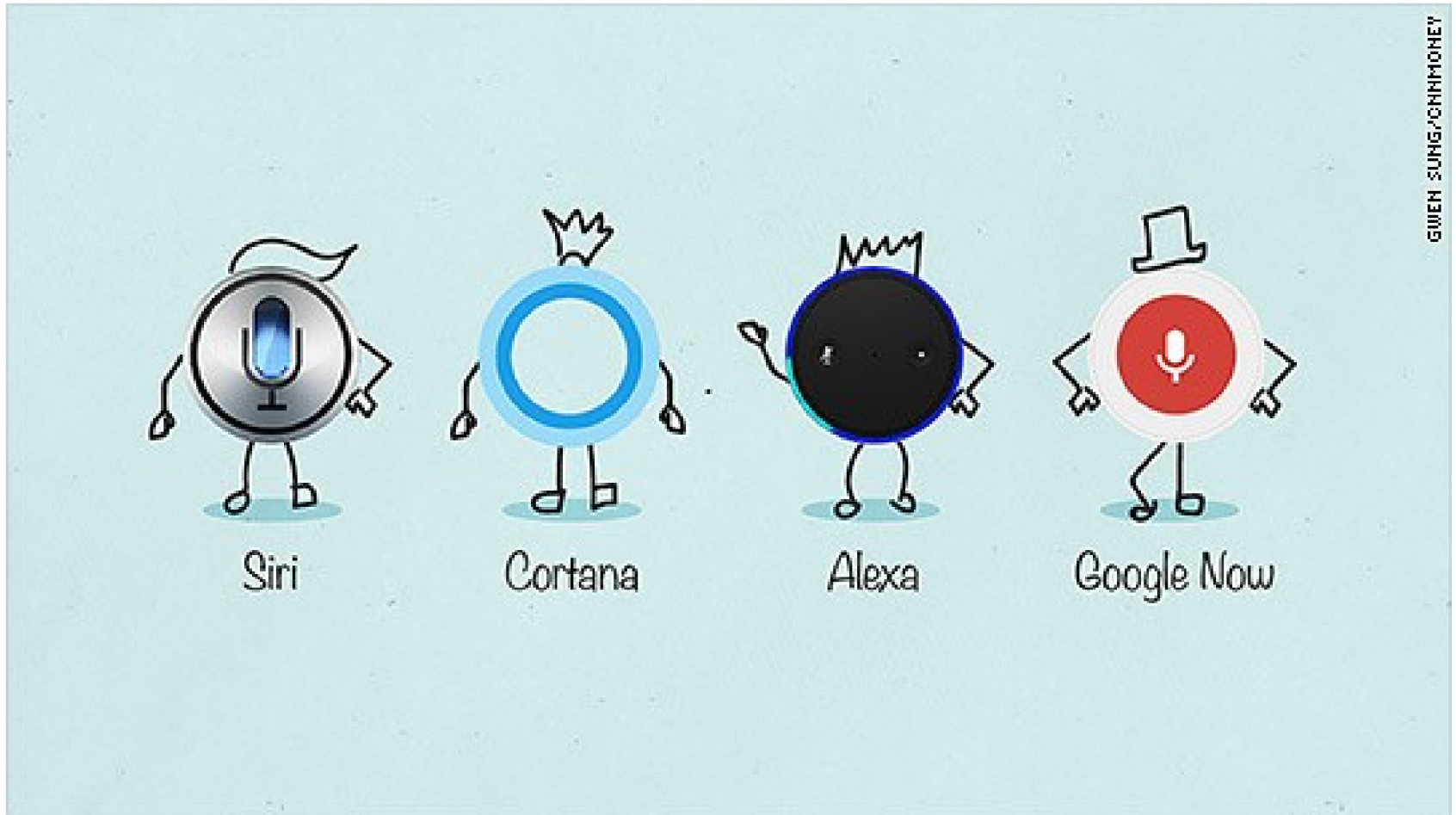


Figure: Speech Recognition.

Machine learning tasks:

Figure: Language Translation.

Machine learning tasks:

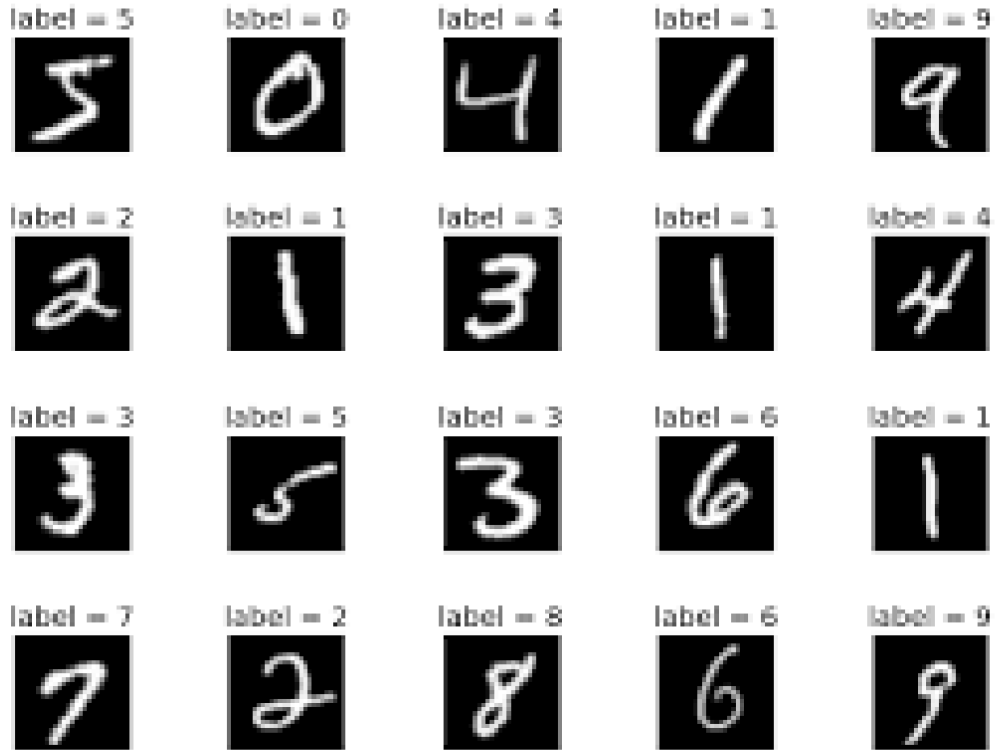


Figure: Handwriting Detection.

Image Generation

- StyleGAN: https://www.youtube.com/watch?v=o46fcRl2yxE

Image to Image Translation (CycleGAN, DistanceGAN...)

Male to Female (First row is input and Second row is output):

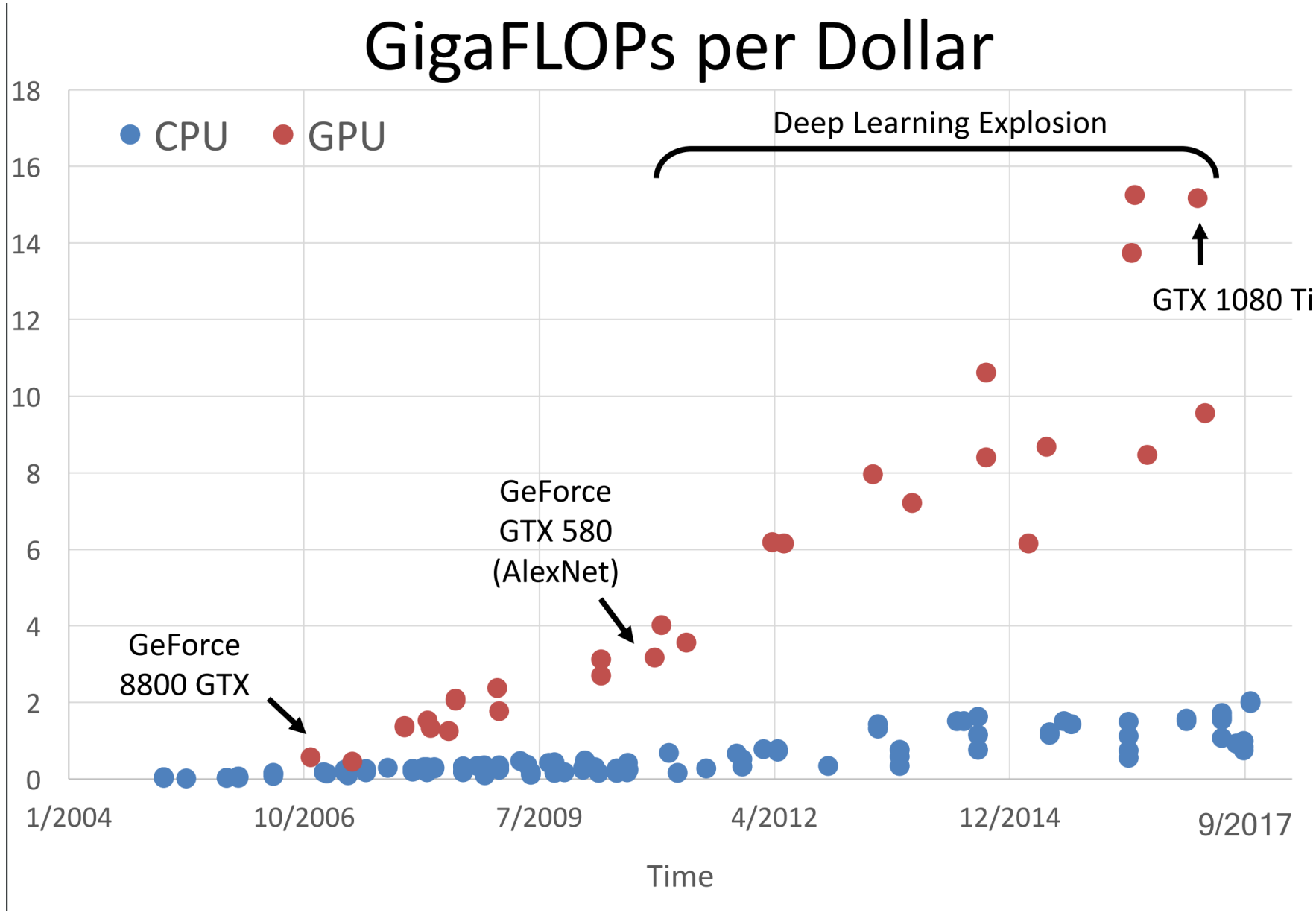
Blond to Black Hair:

With to Without Eyeglasses:

Still an long way to go

Example credit: Andrej Karpathy

Why Now?


GigaFLOPs per Dollar
CPU
GPU
Deep Learning Explosion
GTX 1080 Ti
GeForce
GTX 580
(AlexNet)
GeForce
8800 GTX
18
16
14
12
10
8
6
4
2
0
1/2004
10/2006
7/2009
4/2012
12/2014
9/2017
Time

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

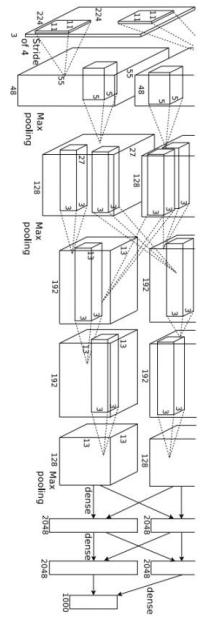
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

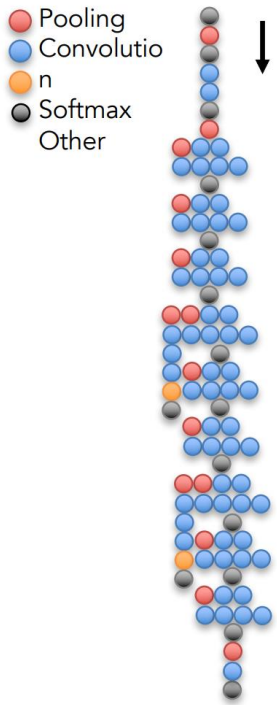
SuperVision



[Krizhevsky NIPS 2012]

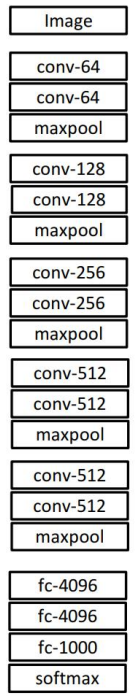
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

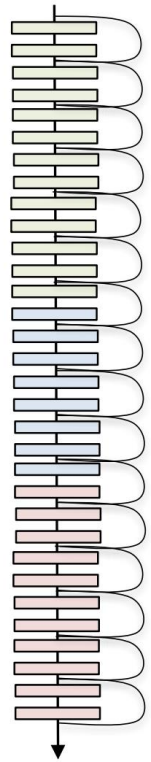
VGG



[Simonyan arxiv 2014]

Year 2015

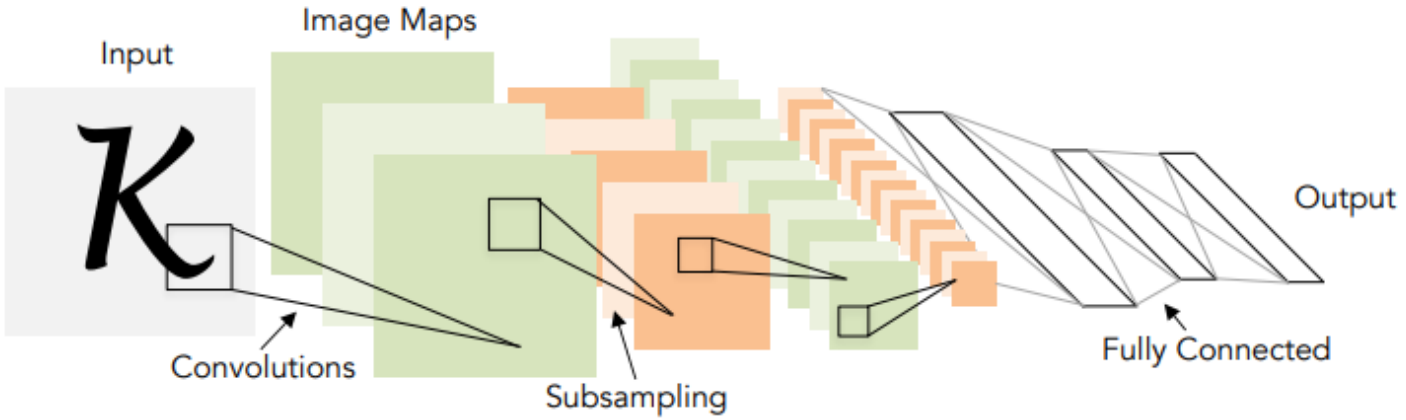
MSRA



[He ICCV 2015]

Convolutional Neural Networks (CNN)
were not invented overnight

1998
LeCun et al.



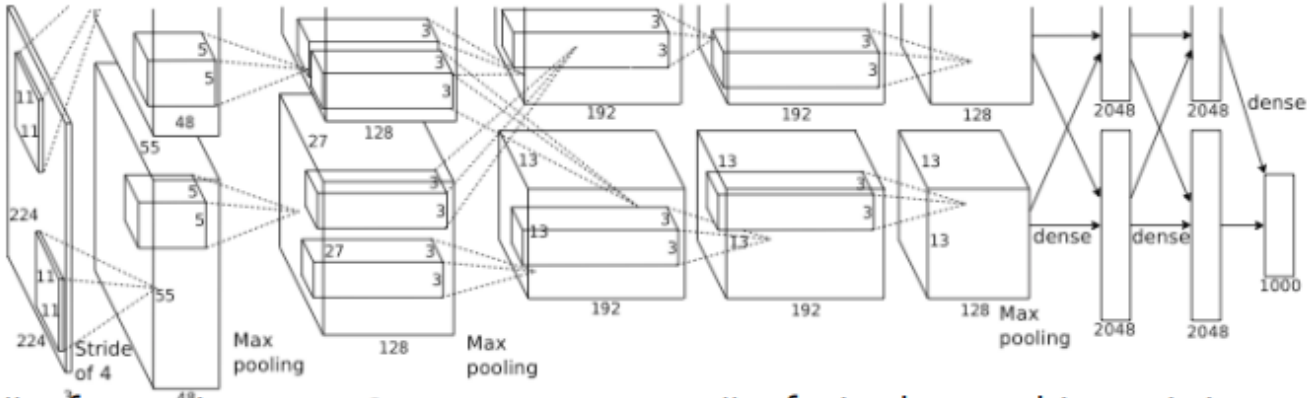
# of transistors

10^6

# of pixels used in training

10^7 NIST

2012
Krizhevsky et al.



# of transistors

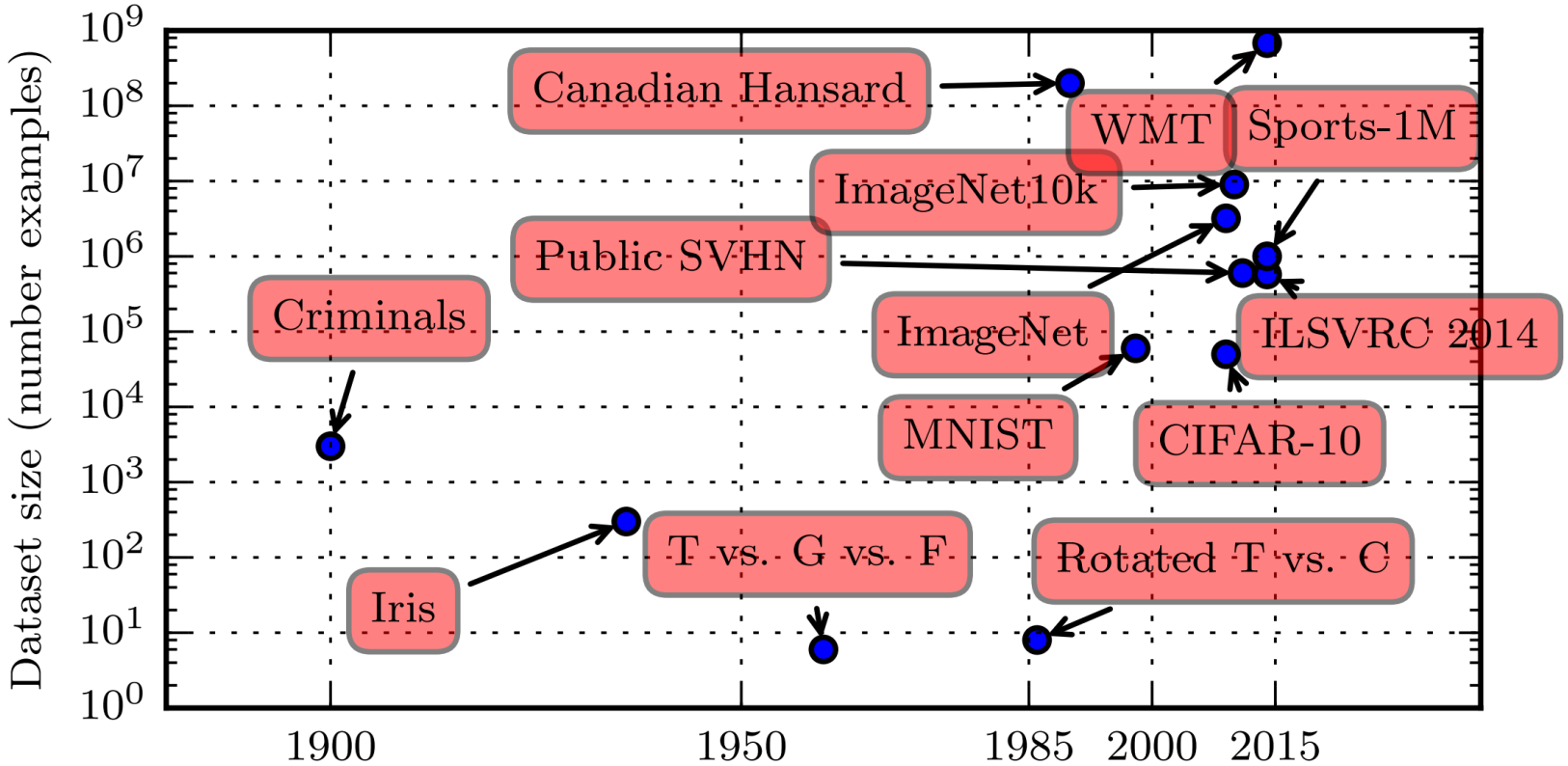
10^9

GPUs

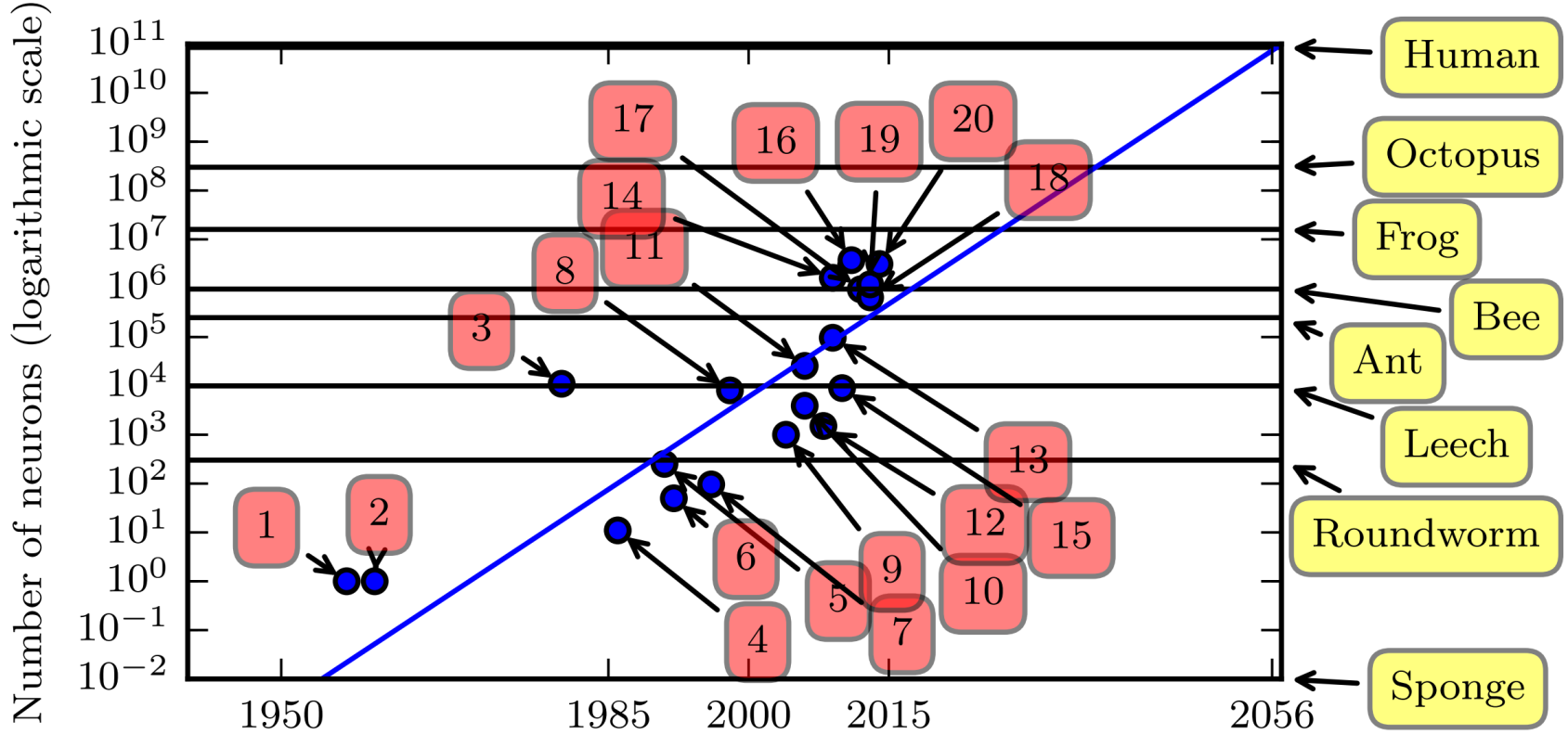
# of pixels used in training

10^{14} IMAGENET

Historical Trends: Growing Datasets



Number of Neurons



Connections per Neuron

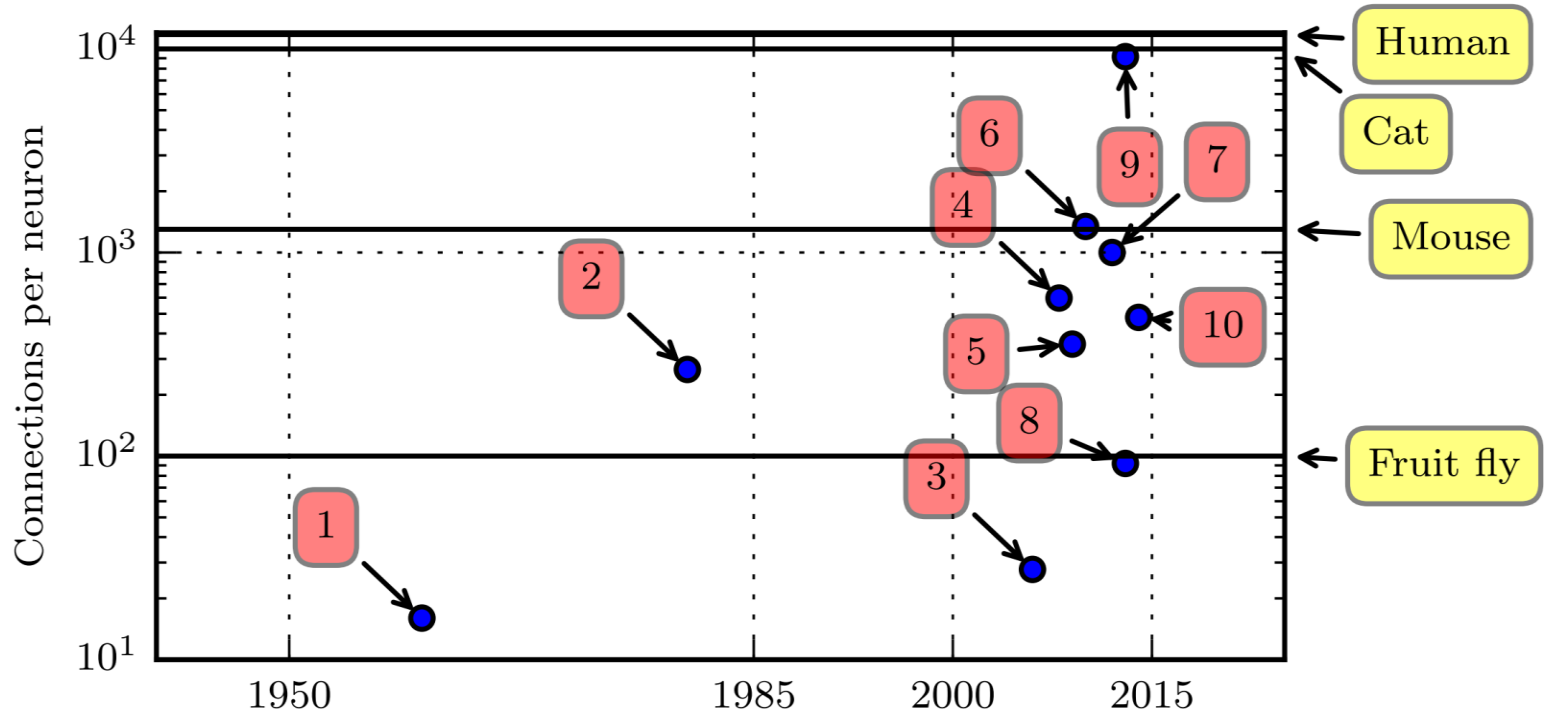


Figure 1.10

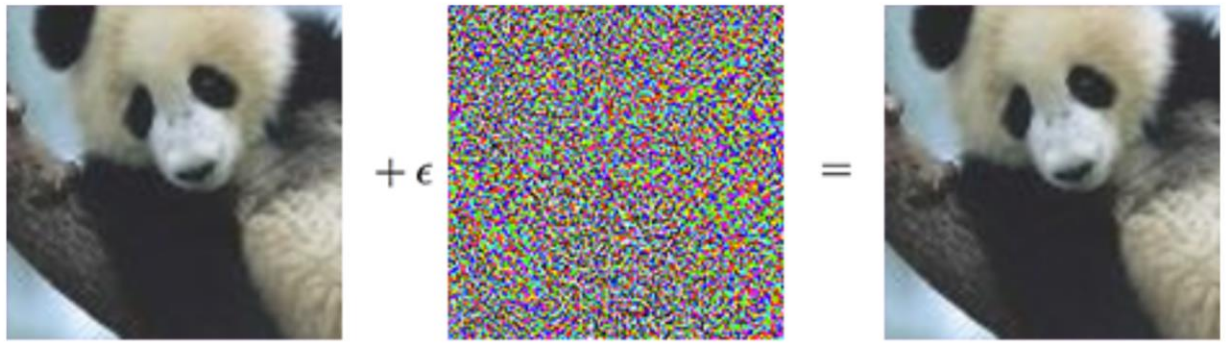
- Will Computers Take Over?
- Can we build General AI?

THE TERMINATOR™

SKYNET

Are computers imitating humans?

- Not really, consider adversarial examples:



- Community Culture: Open Source, Share Everything
- Can one compete with Google or FB: YES
  - Face Recognition
  - Action Recognition
  - Image Annotation
  - Image to Image Translation
  - Central NLP Tasks
  - Many other tasks produced state of the art results in the deep learning lab

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

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".