

The background of the slide features a large, faint, circular watermark of the Stanford University seal. The seal contains a redwood tree in the center, surrounded by the text 'STANFORD UNIVERSITY' at the top and 'FREIHEIT WEHT' at the bottom. The year '1891' is inscribed at the bottom of the seal.

CS231n: Convolutional Neural Network for Visual Recognition

Lecture 1: Introduction

Welcome to CS231n



Welcome to CS231n

Course Instructors



Fei-Fei Li Andrej Karpathy

Teaching Assistants



2015

Justin Johnson Yuke Zhu Brett Kuperl Ben Poole

Course Instructors



Fei-Fei Li Andrej Karpathy

Teaching Assistants



2016

Instructors



Fei-Fei Li Justin Johnson

Teaching Assistants



2017



Arman Poddada Soya (Emma) Peng Ben Poole Lida Zhao

Instructors



2018

Teaching Assistants



Arman Poddada Mike Roberts Praty Sharma Fei Xia Danfei Xu Ben Zhang

Instructors



2019

Teaching Assistants



Arman Poddada Praty Sharma William Shen Owen Wang Danfei Xu

Instructors



Course Coordinator



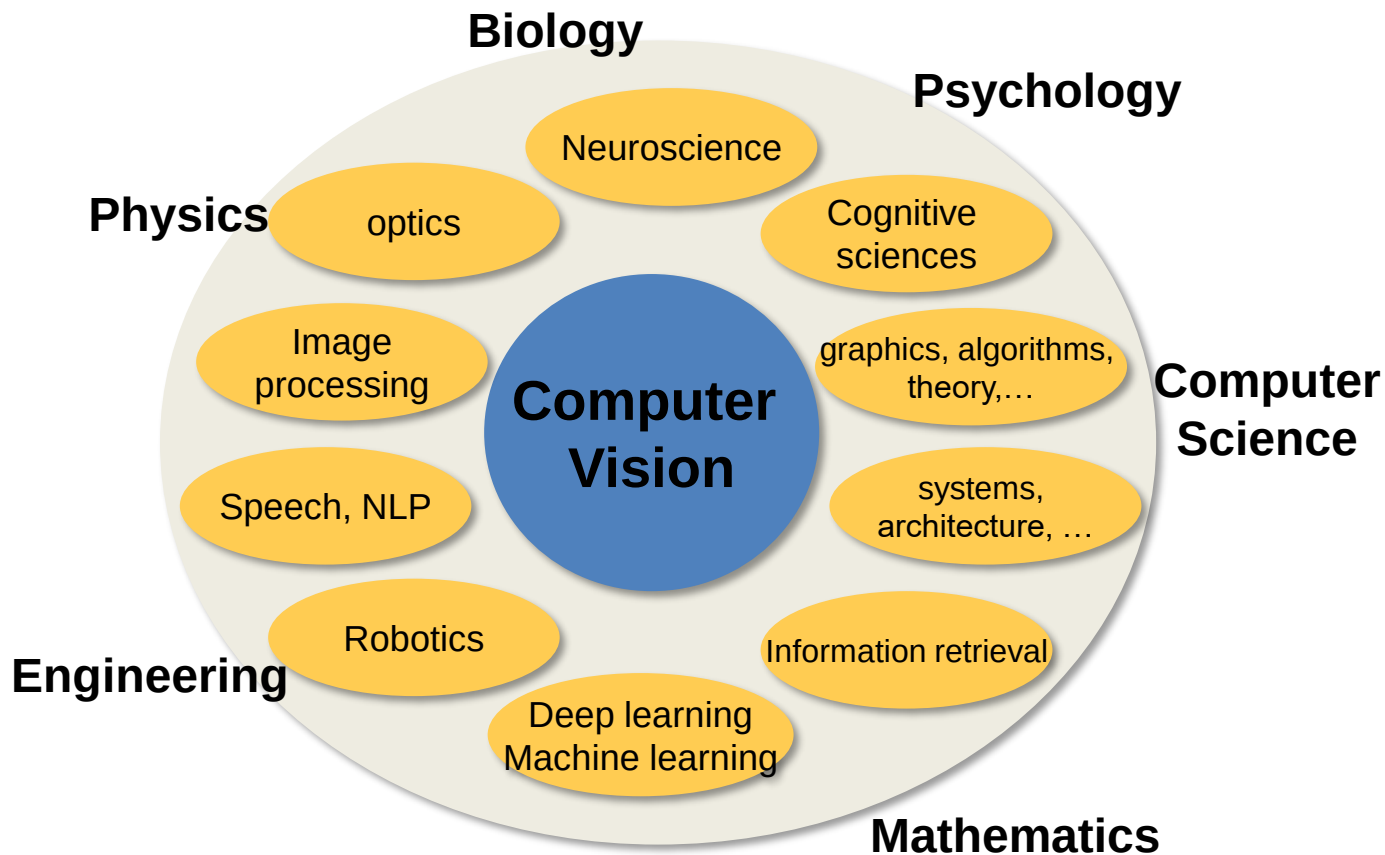
Anneli Byun

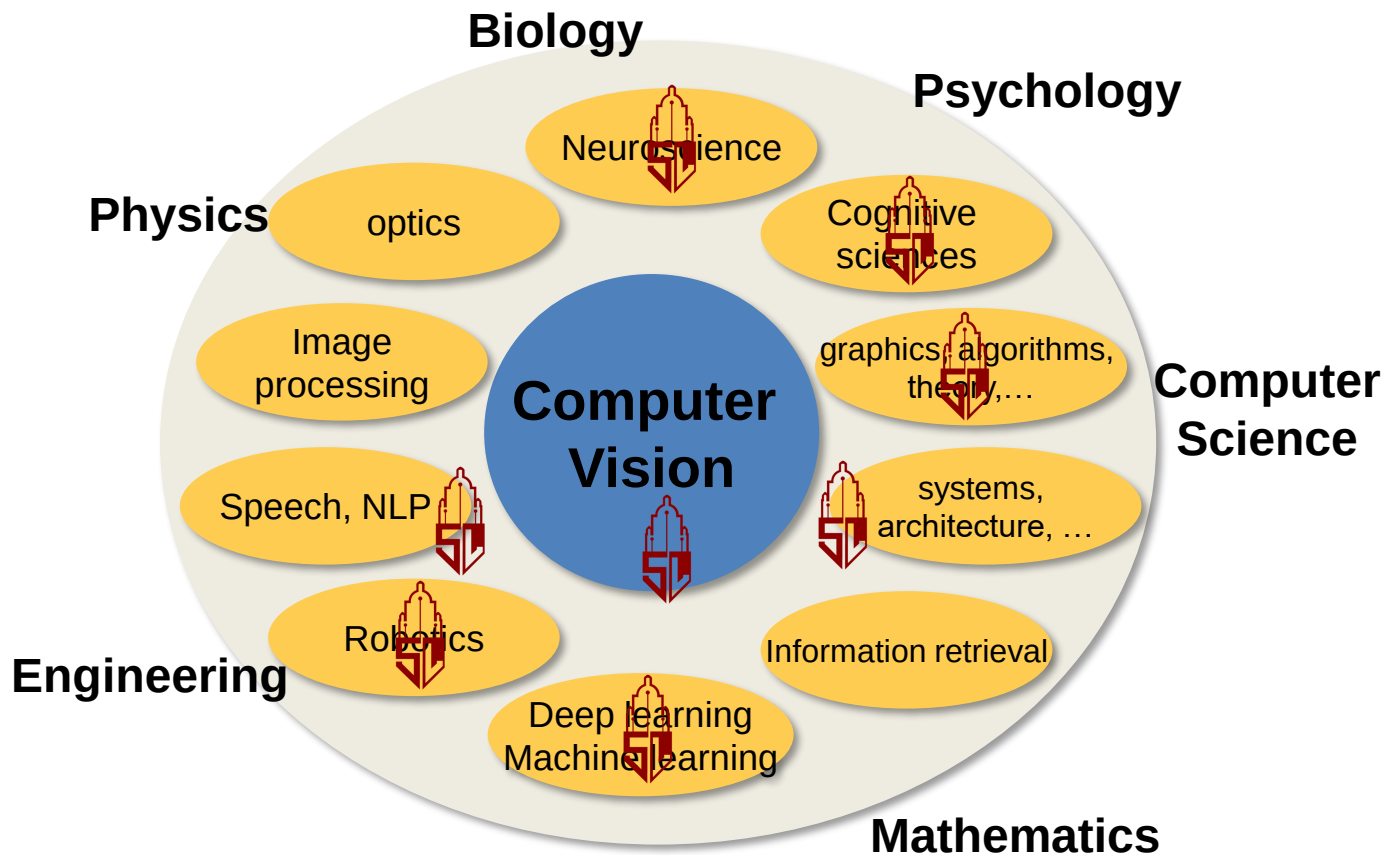
Teaching Assistants

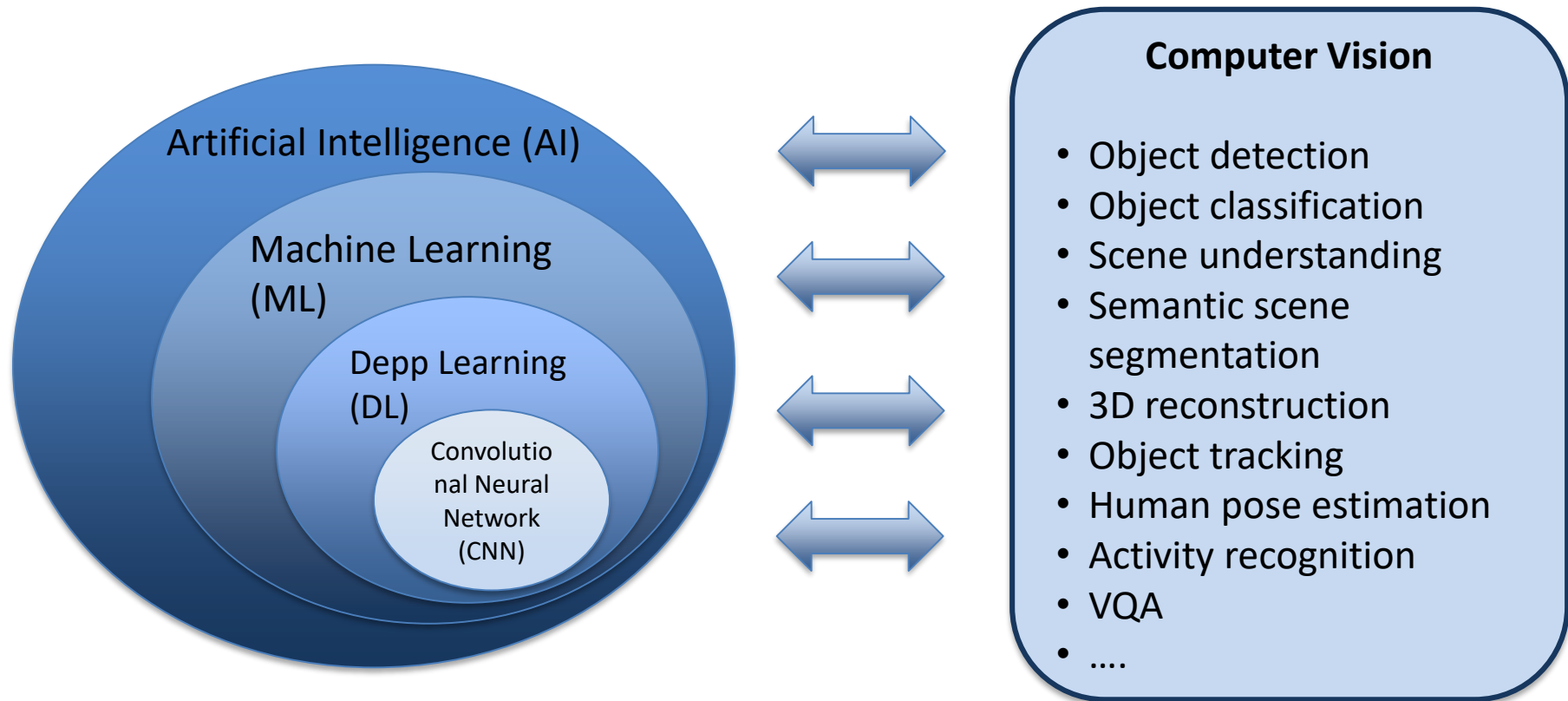


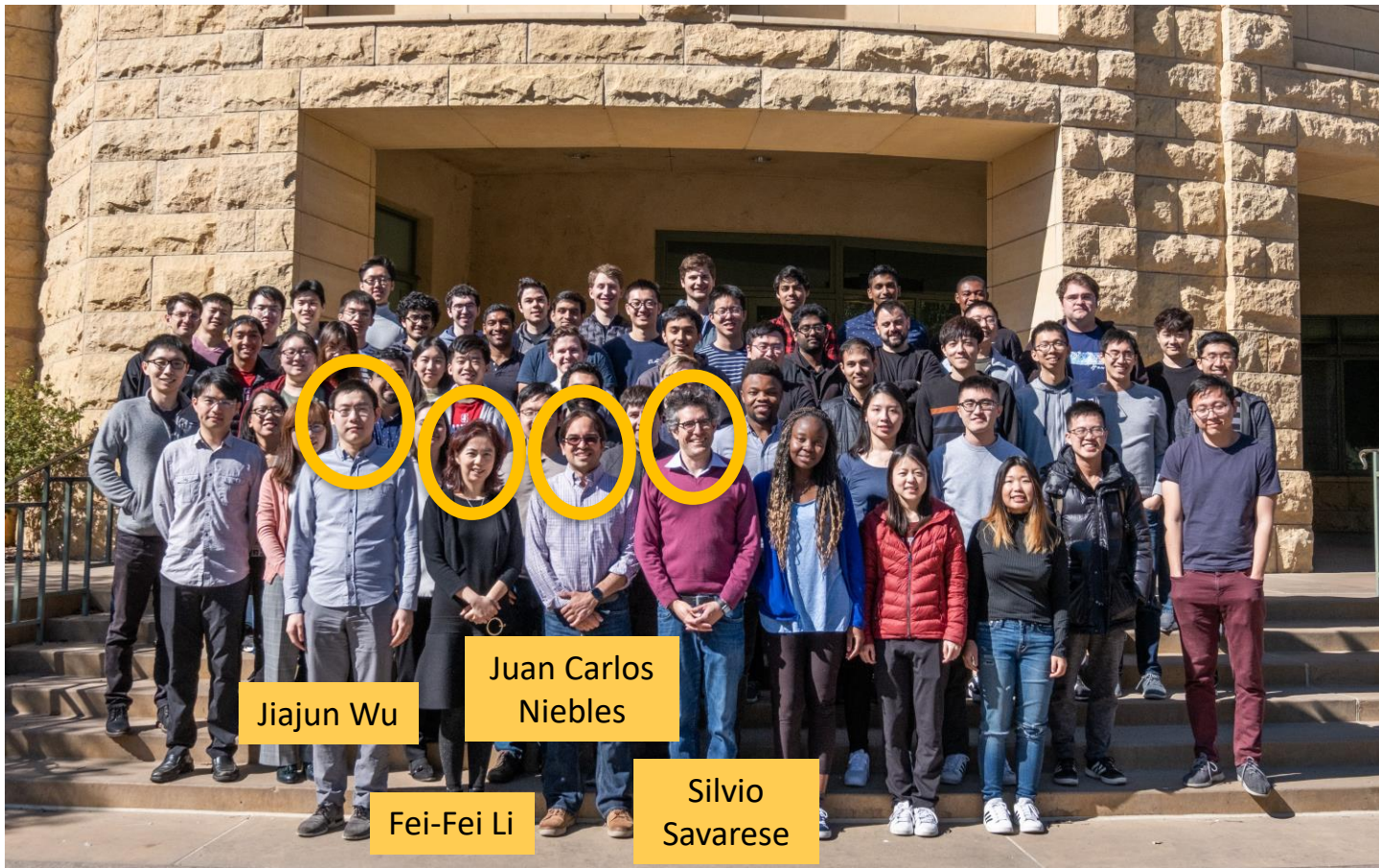
Brent Yi Christina Yuan Kevin Zakka Yiheng Zhang

2020









Today's agenda

- A brief history of computer vision
- CS231n overview

Evolution's Big Bang: Cambrian Explosion, 530-540million years, B.C.



[This image](#) is licensed under [CC-BY 2.5](#)



[This image](#) is licensed under [CC-BY 2.5](#)



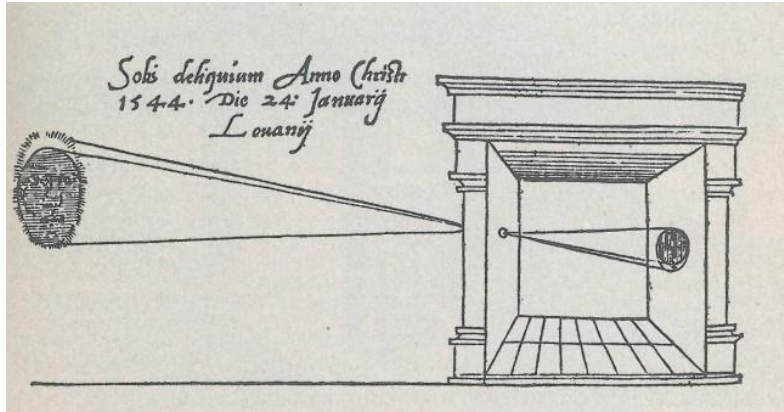
[This image](#) is licensed under [CC-BY 3.0](#)



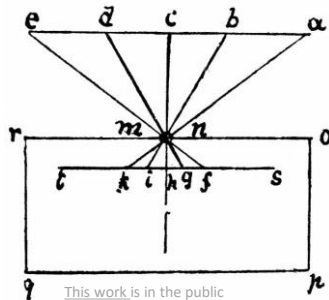
Camera Obscura

Encyclopedia, 18th Century

Gemma Frisius, 1545

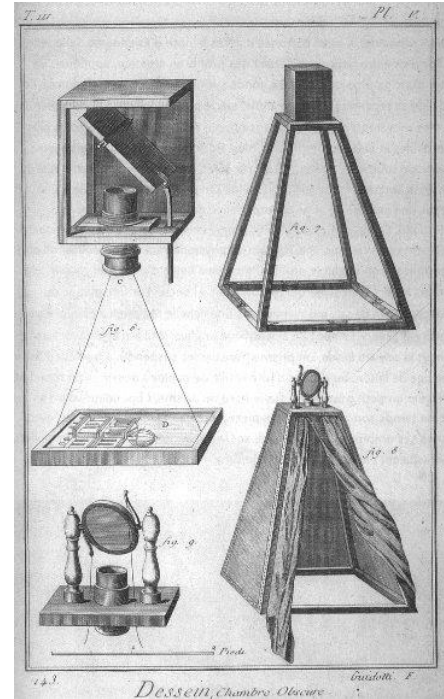


This work is in the public domain



This work is in the public domain

Leonardo da Vinci,
16th Century AD



This work is in the public domain

Where did we come from?

The known story – Neuroscience inspired AI

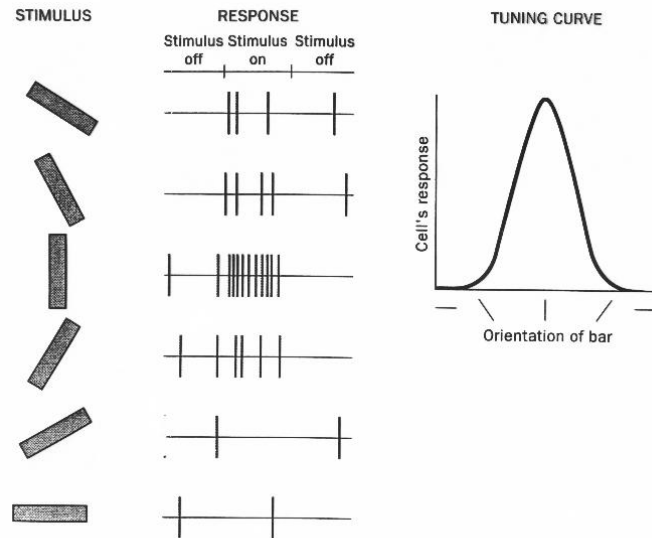
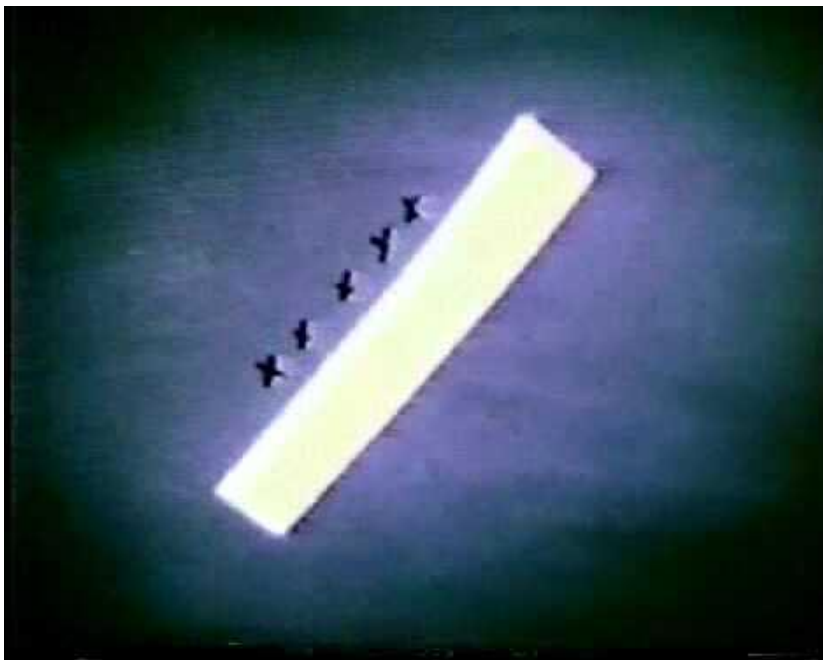


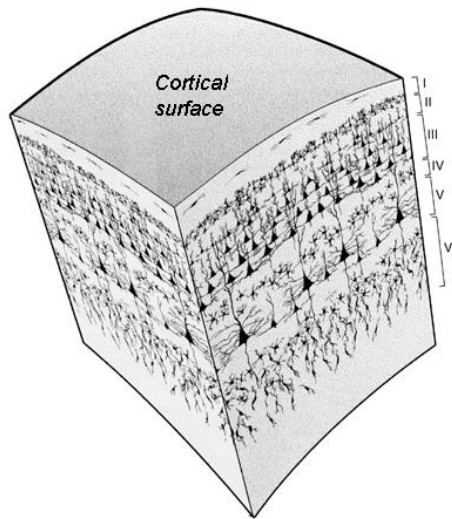
FIGURE 4.8 Response of a single cortical cell to bars presented at various orientations.

Hubel and Wiesel, 1959

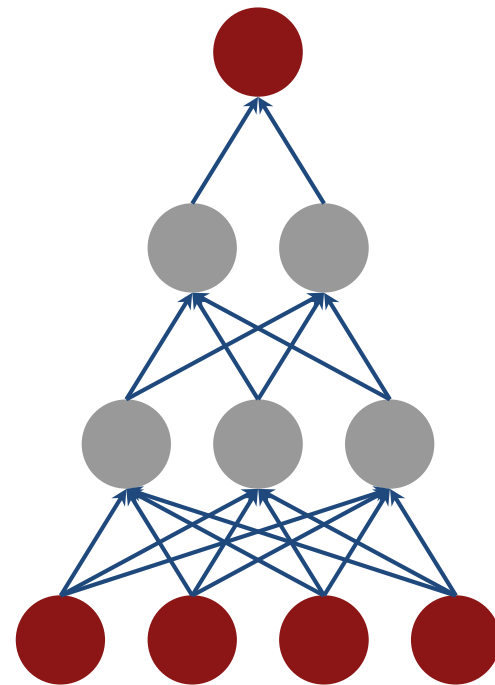
High-Level Patterns



Low-Level
Details



Cortical Column
(Biological)



Neural Networks
(Digital)

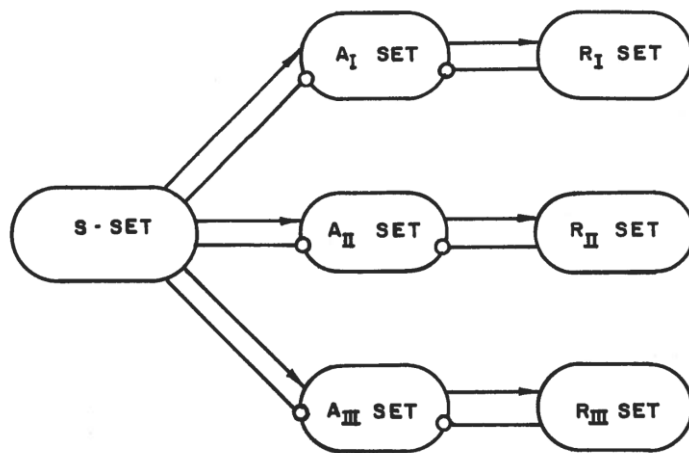


FIGURE 2
ORGANIZATION OF A PERCEPTRON WITH
THREE INDEPENDENT OUTPUT-SETS

F. Rosenblatt, 1957

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
 & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California,
 San Diego, La Jolla, California 92093, USA

† Department of Computer Science, Carnegie-Mellon University,
 Pittsburgh, Philadelphia 15213, USA

Rumelhart, Hinton & Williams, 1986



“The mere formulation of a problem is often far more essential than its solution, which [...] requires creative imagination and marks real advances in science.”

- Albert Einstein, 1921

Where did we come from?

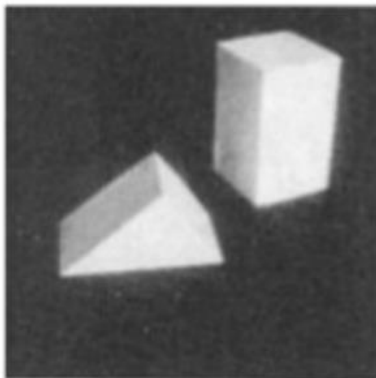
The not-so-known story – the search for computer vision's “North Star”



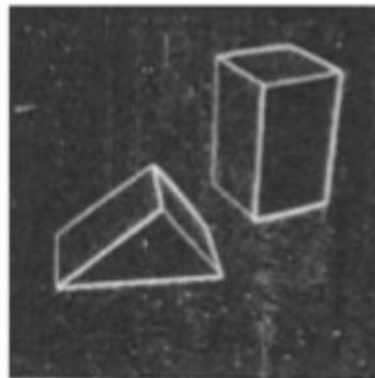
1960s: Interpretation of synthetic world



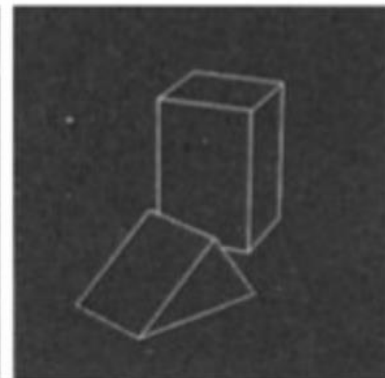
Larry Roerts
1963, 1st thesis of Computer Vision



Input image



2x2 gradient operator



computed 3D model
rendered from new viewpoint

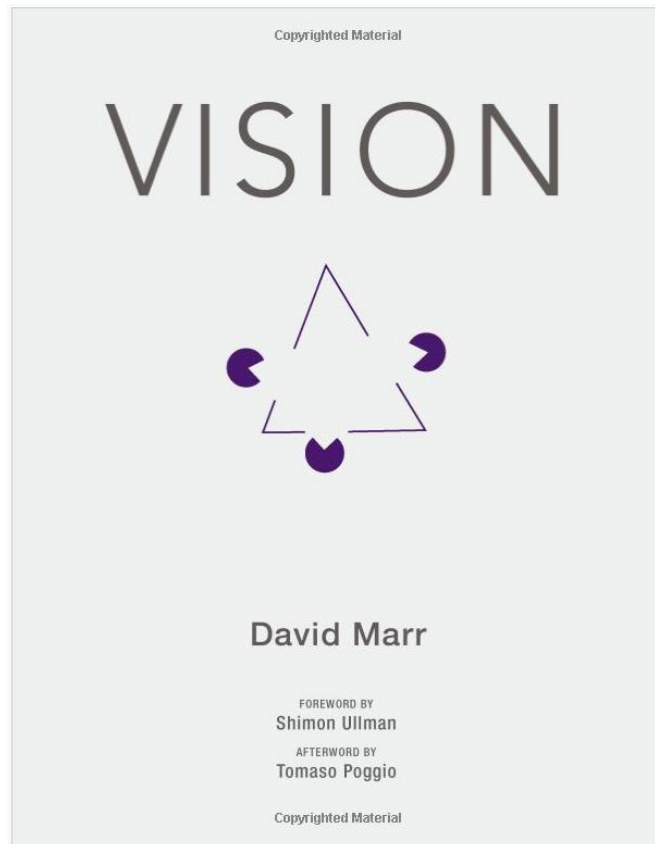
Artificial Intelligence Group
Vision Memo. No. 100.

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".



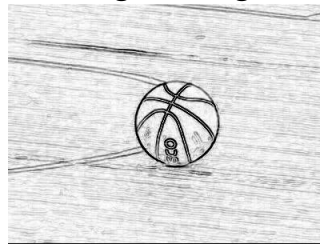
David Marr, 1970s

Input image

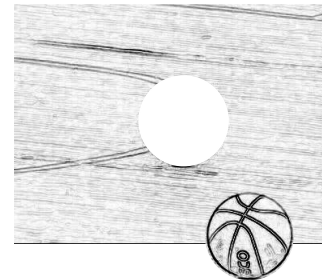


This image is CC0 1.0 public domain

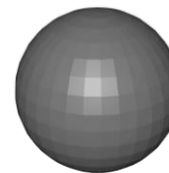
Edge image



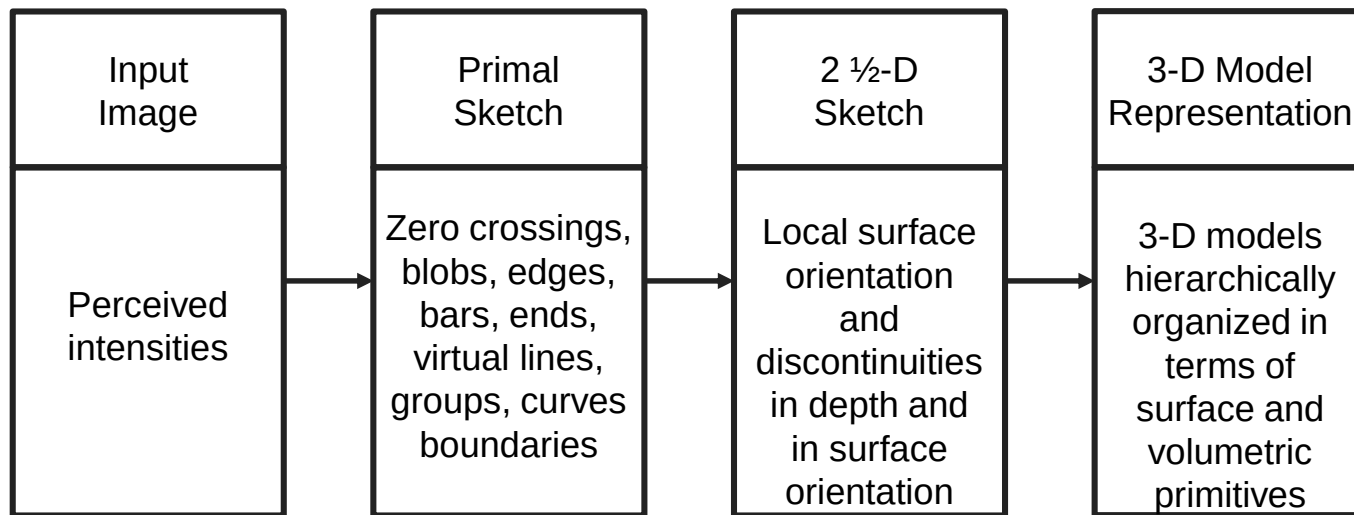
2 ½-D sketch



3-D model

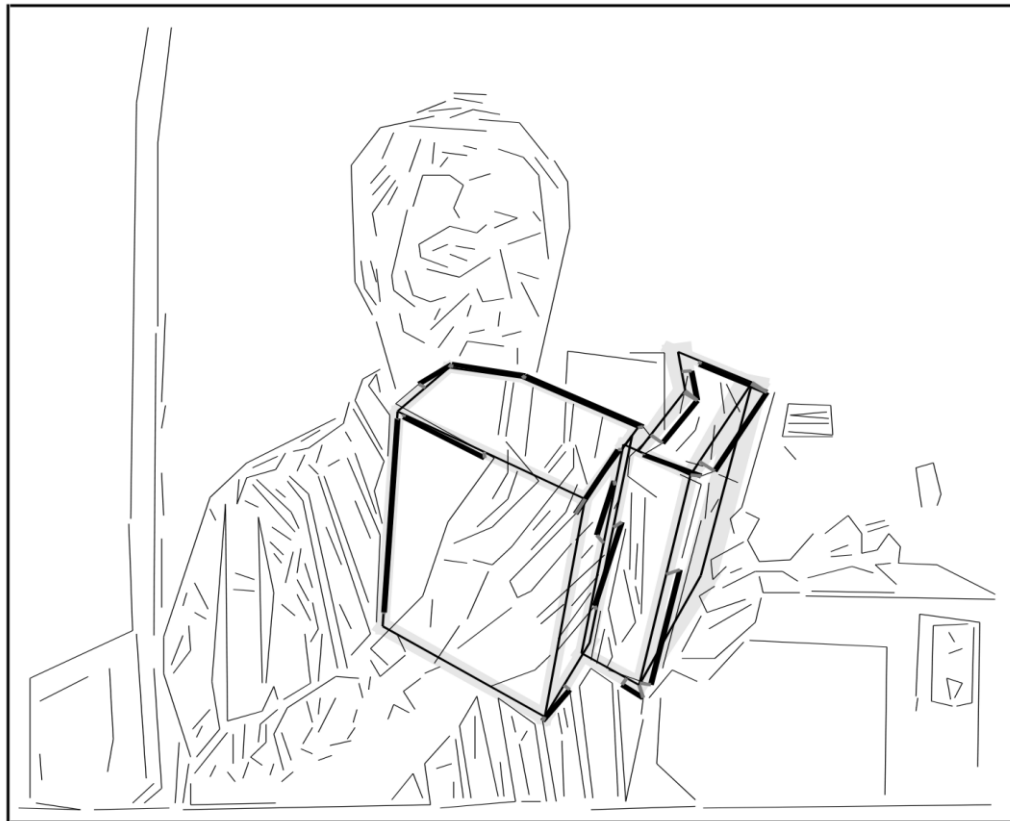


This image is CC0 1.0 public domain



Stages of Visual Representation, David Marr, 1970s

Edges, segmentation, and perception



D. Lowe. *IJCV*, 1992

Normalized Cut (Shi & Malik, 1997)

Image is CC BY 3.0



Image is public domain

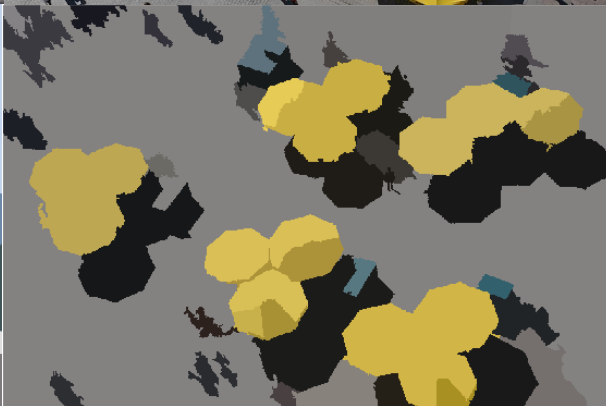
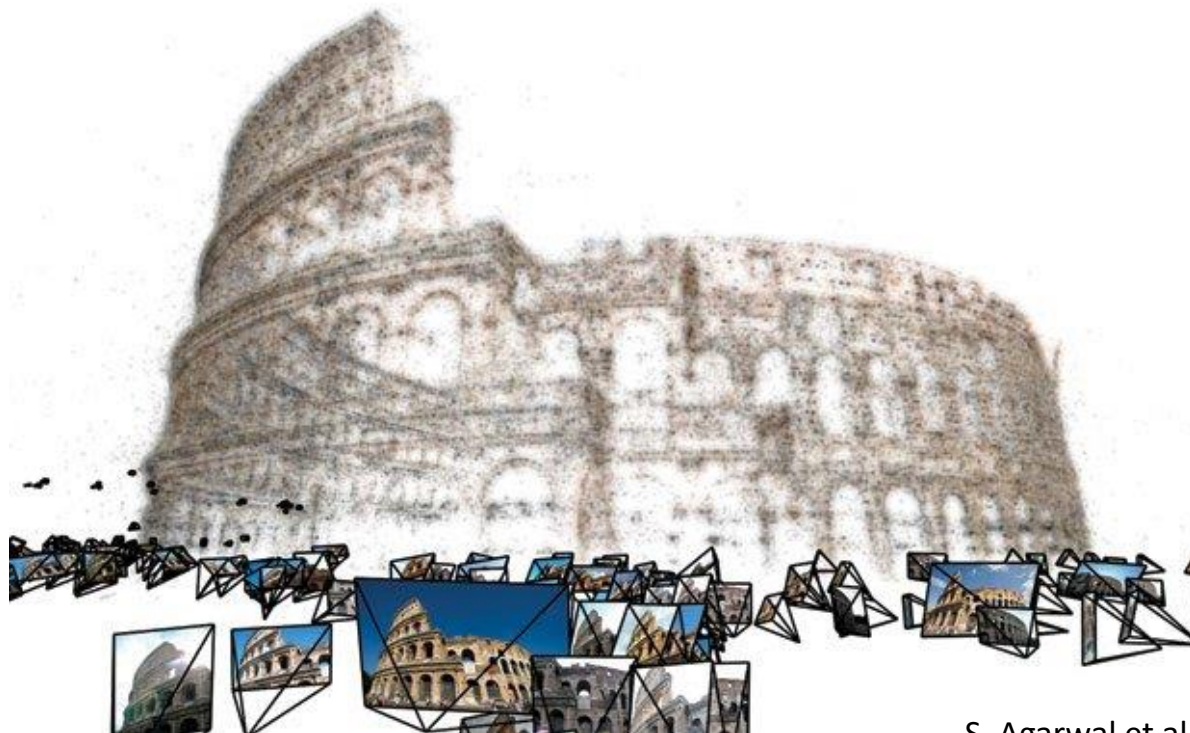


Image is CC-BY 2.0;
changes made



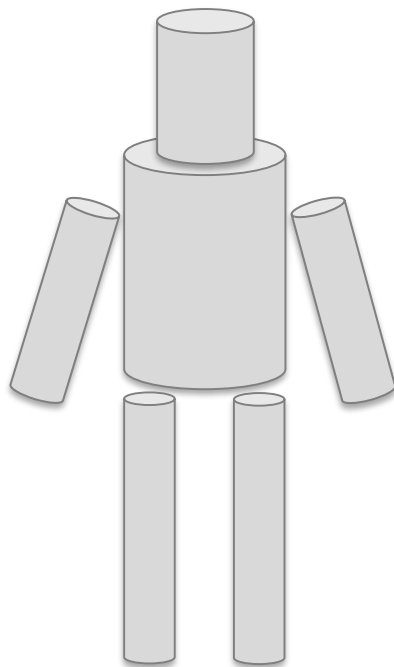
3D reconstruction



S. Agarwal et al. ICCV, 2009

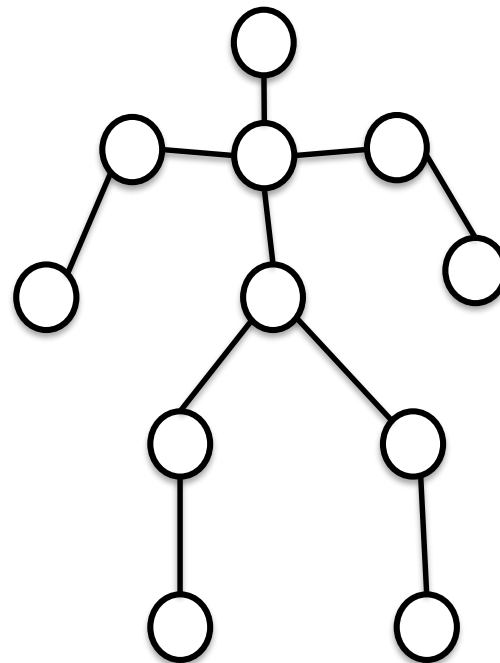
- Generalized Cylinder

Brooks & Binford, 1979



- Pictorial Structure

Fischler and Elschlager, 1973



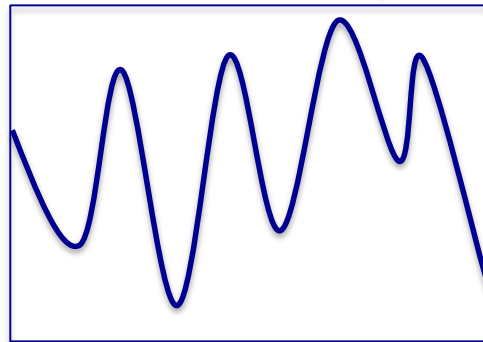
Single Object Recognition



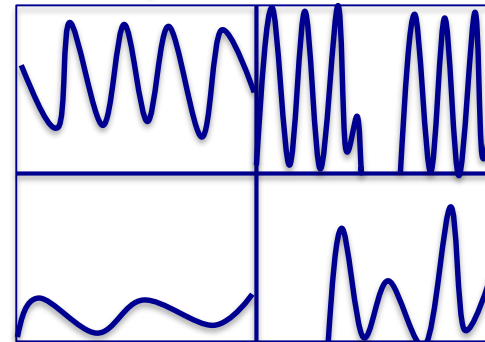
D. Lowe. ICCV, 1999



Image is [CC0 1.0](#) public domain

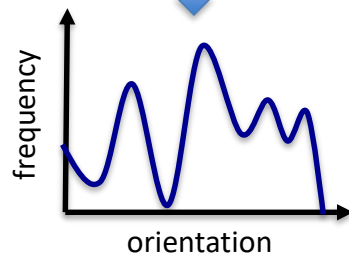


Level 0

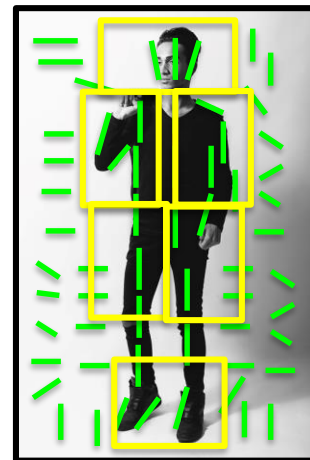


Level 1

Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006

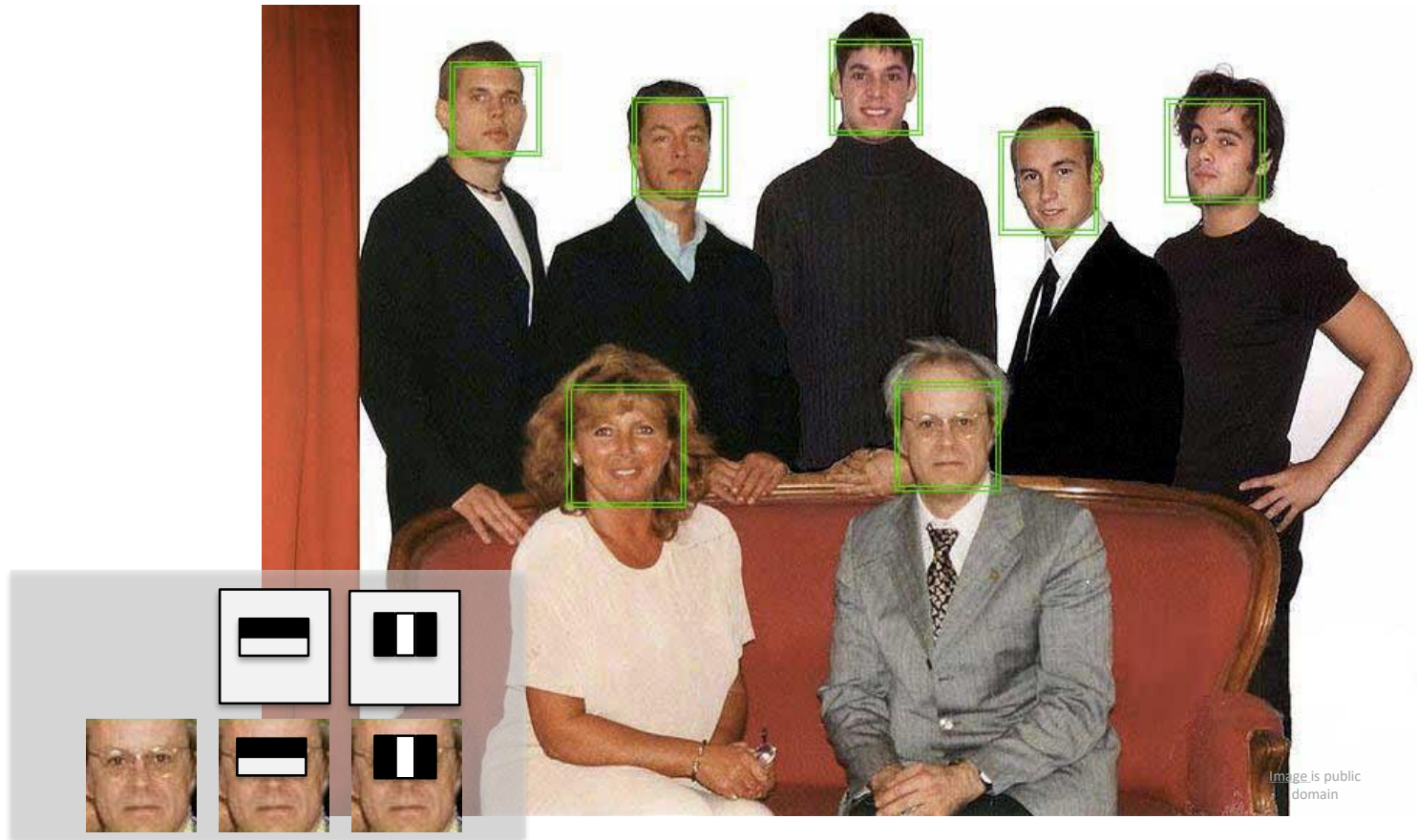


Histogram of Gradients (HoG)
Dalal & Triggs, 2005

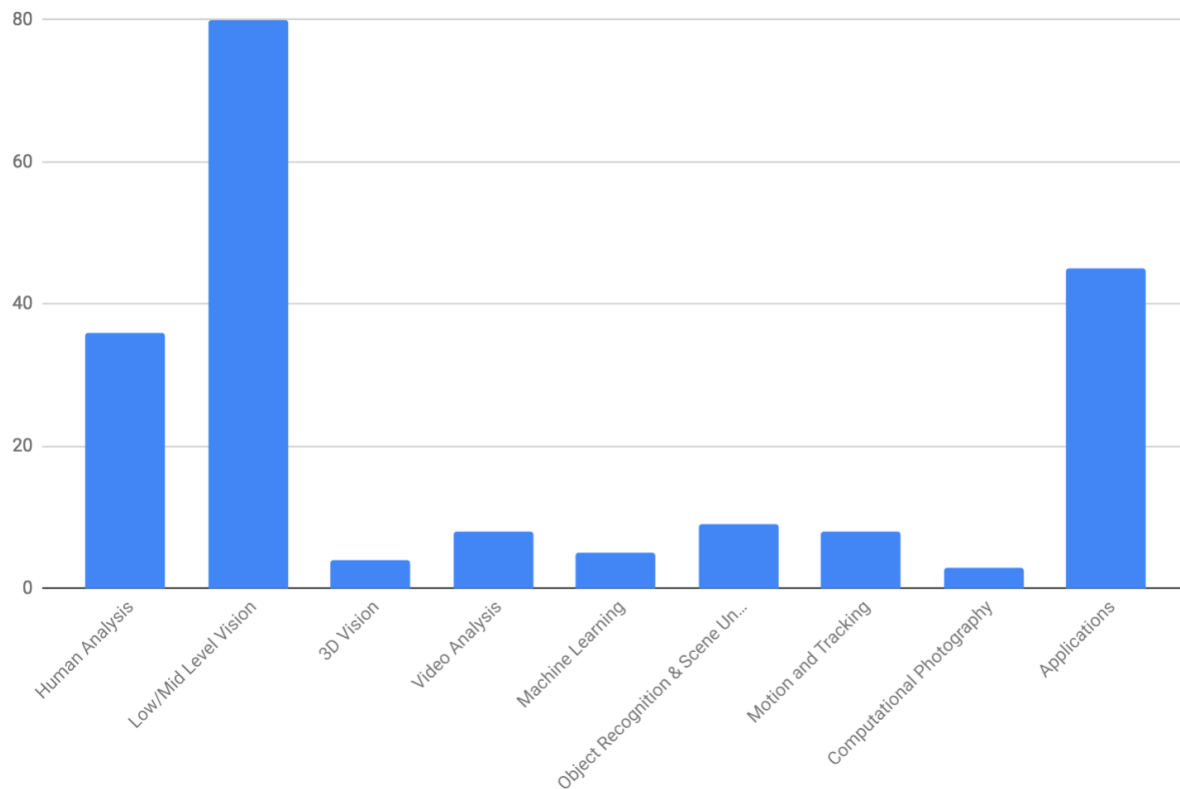


Deformable Part Model
Felzenswalb, McAllester, Ramanan,
2009

Face Detection, Viola & Jones, 2001



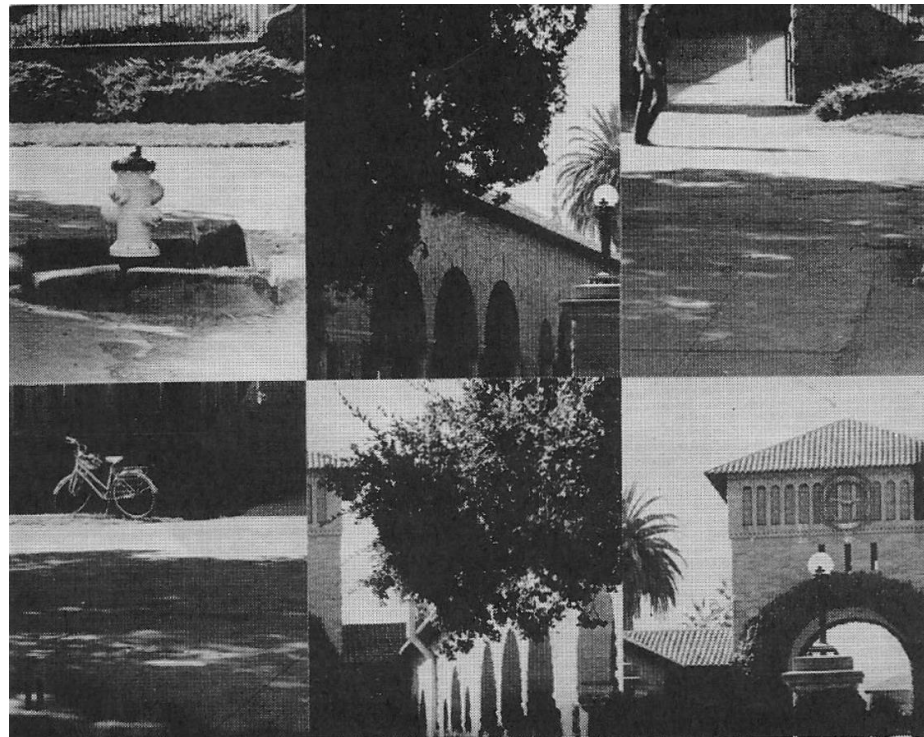
CVPR topic distribution: 2000



In the mean time...

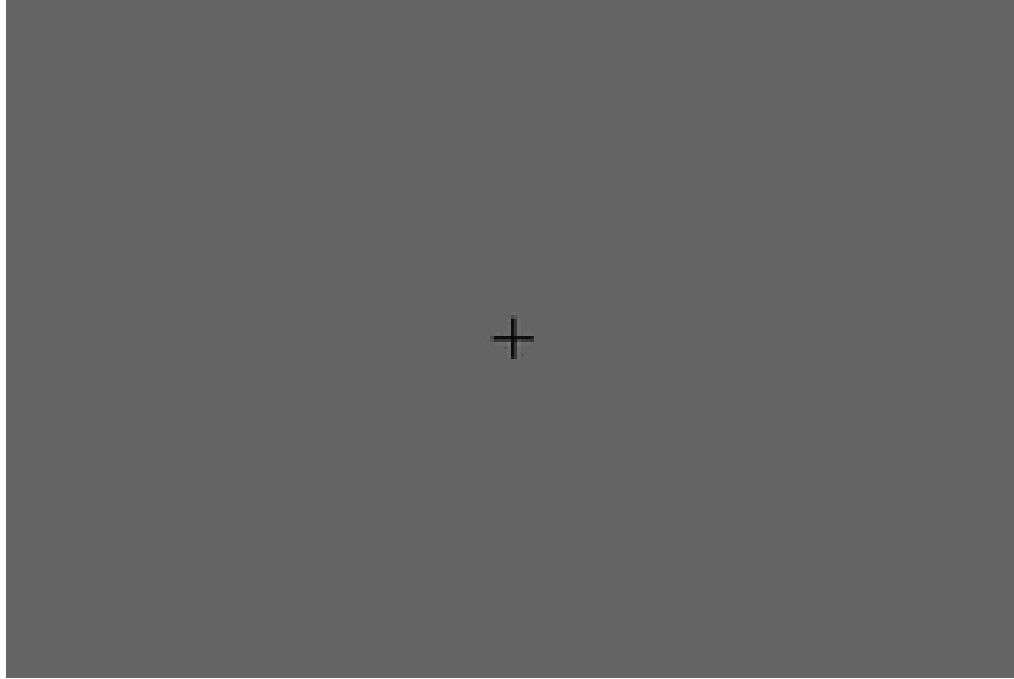
Perceiving Real-World Scenes

Irving Biederman



I. Biederman, *Science*, 1972

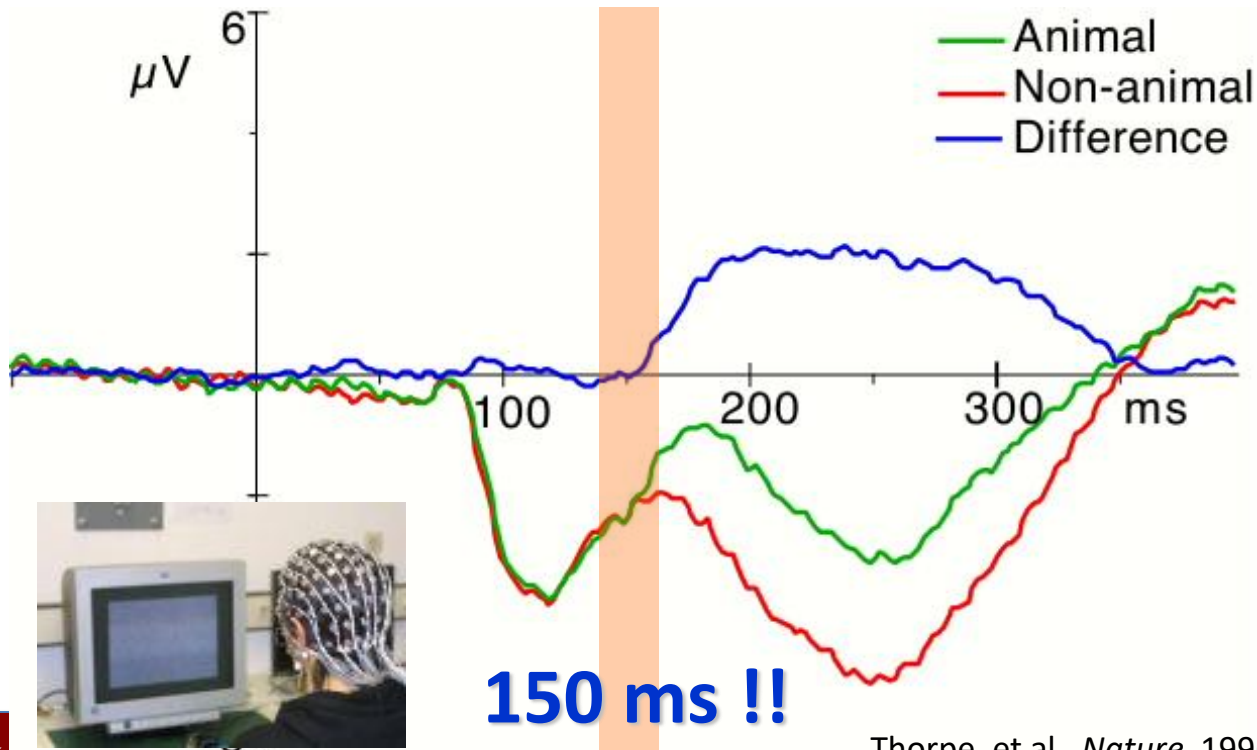
Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

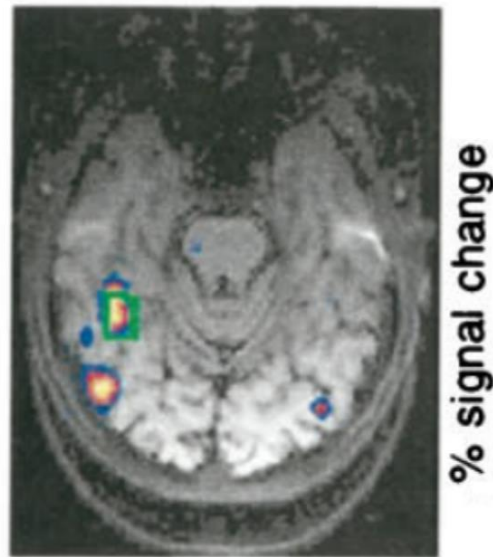
Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

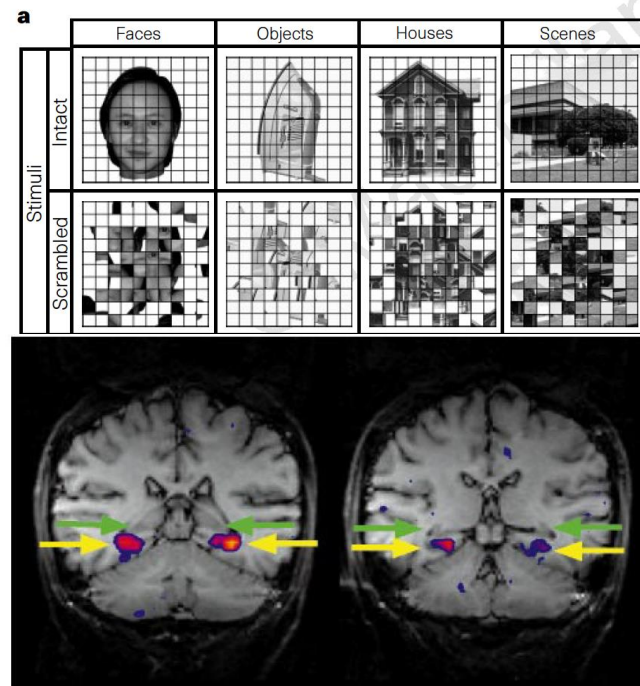


Neural correlates of object & scene recognition

Faces > Houses

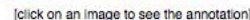


Kanwisher et al. J. Neuro. 1997



Epstein & Kanwisher, Nature, 1998

Caltech 101 images

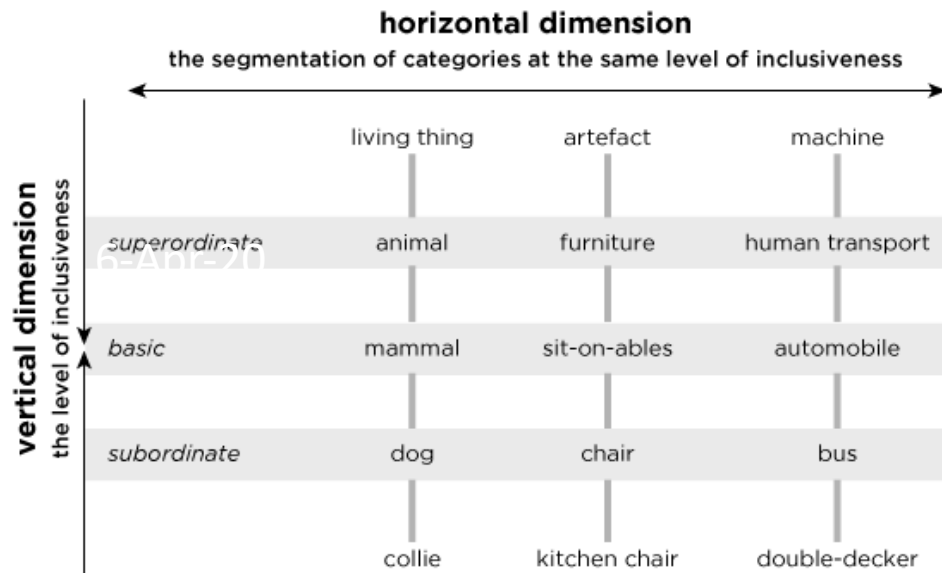


Everingham et al. 2006-2012

There are **MANY** objects; organized **HIERARCHICALLY**

(in the case of cups) to perhaps 15 or more (in the case of lamps) readily discernible exemplars.¹¹ Let us assume (liberally) that the mean number of types is 10. This would yield an estimate of 30,000 readily discriminable objects (3,000 categories $\times$ 10 types/category).

A second source for the estimate derives from considering plausible rates for learning new objects. Thirty thousand objects would require learning an average of 4.5 objects per day, every day for 18 years, the modal age of the subjects in the experiments described below.



Eleanor Rosch: Principles of Categorization, 1978

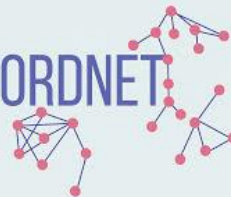
- Biederman: Recognition by Component, 1987



George A. Miller

Psychology, Cognitive Science
Princeton University

WORDNET





IMAGENET

22,000 categories

⋮

15,000,000 images



IMAGENET Large Scale Visual Recognition Challenge

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



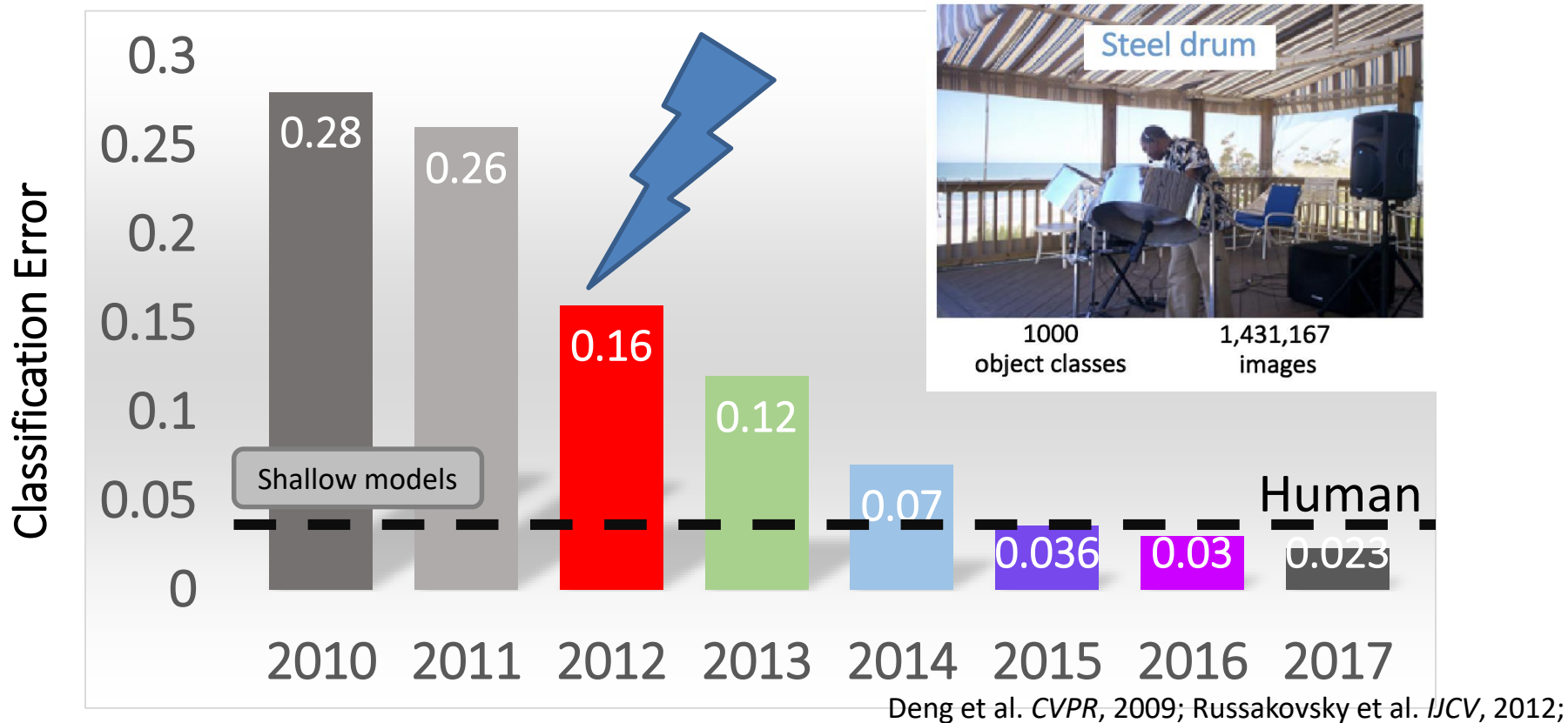
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle



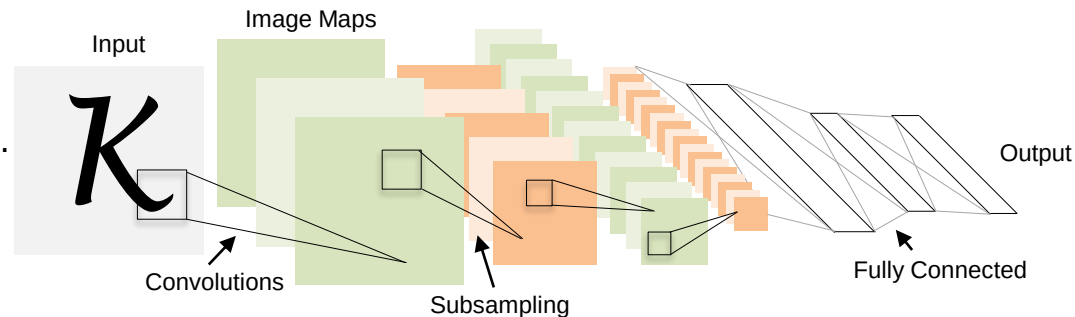
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle



Russakovsky et al. IJCV 2015



1998
LeCun et al.



of transistors

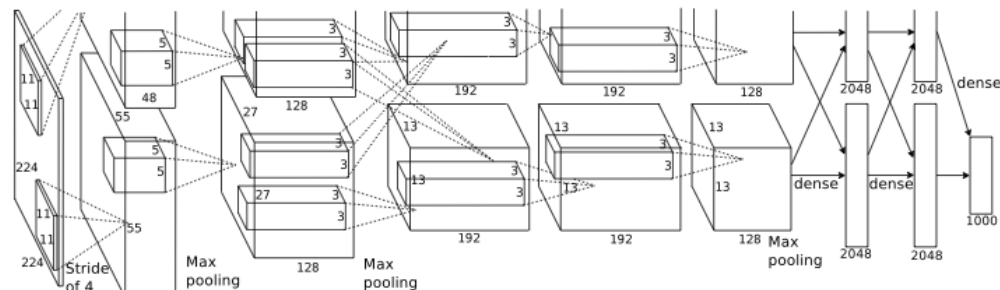


10^6

of pixels used in training

10^7 **NIST**

2012
Krizhevsky et al.



of transistors



10^9

GPUs



of pixels used in training

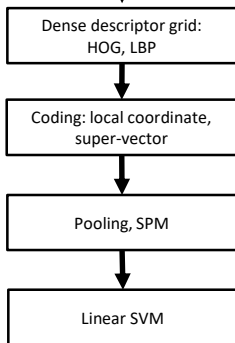
10^{14} **IMAGENET**

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

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

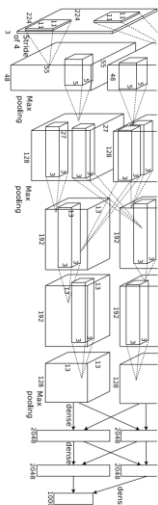


[Lin CVPR 2011]

[Lion image](#) by Swissfrog
is
licensed under [CC BY 3.0](#)

Year 2012

SuperVision

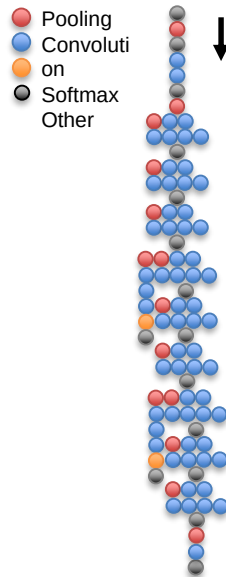


[Krizhevsky NIPS 2012]

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

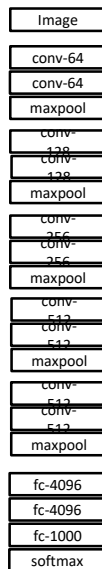
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

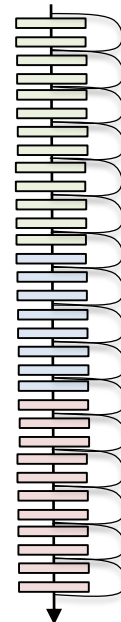
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

Image Captioning: Richer Descriptions



Image
Captioning



A man riding a horse drawn carriage down a street

Prior
Work

Dense
Captioning



JKF, CVPR 2016

Horse pulling a cart. A wheel on a cart. A window on a building. A horse in a picture. A large white umbrella. A woman sitting on a bench. Man sitting on a motorcycle. A cart with a cart.

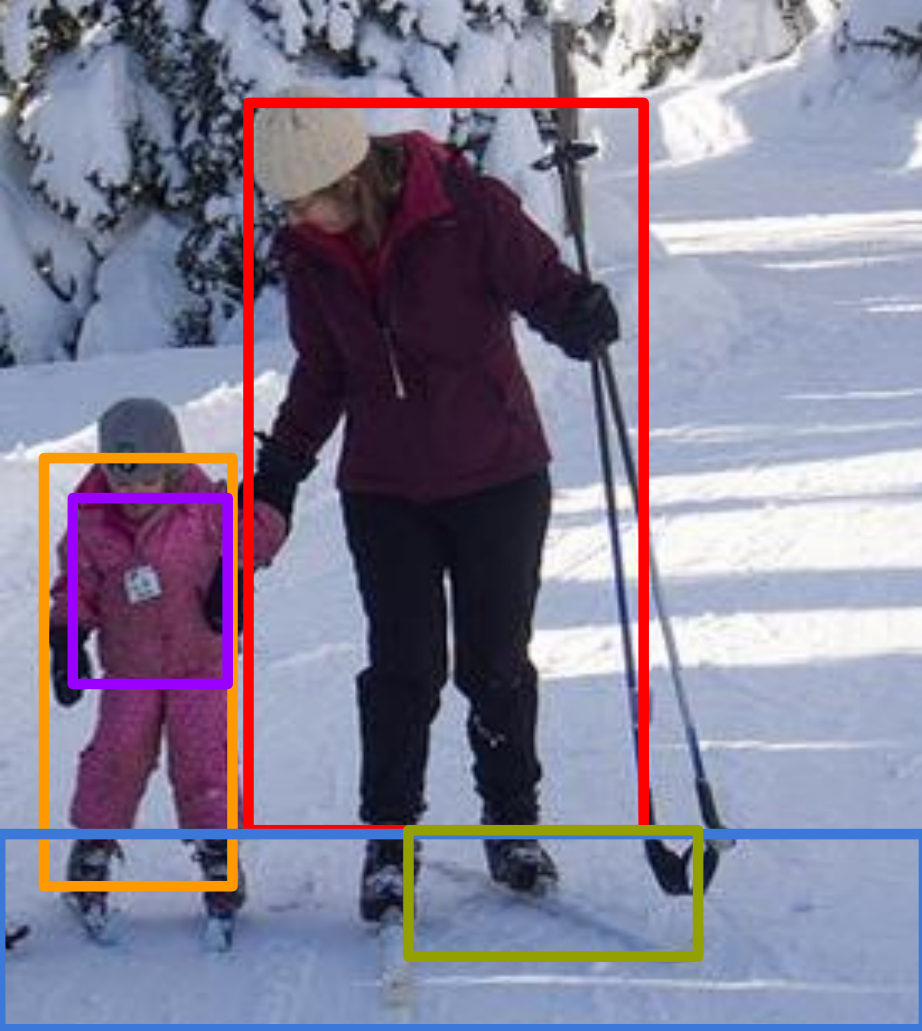
Our Recent Work

Paragraph
Captioning



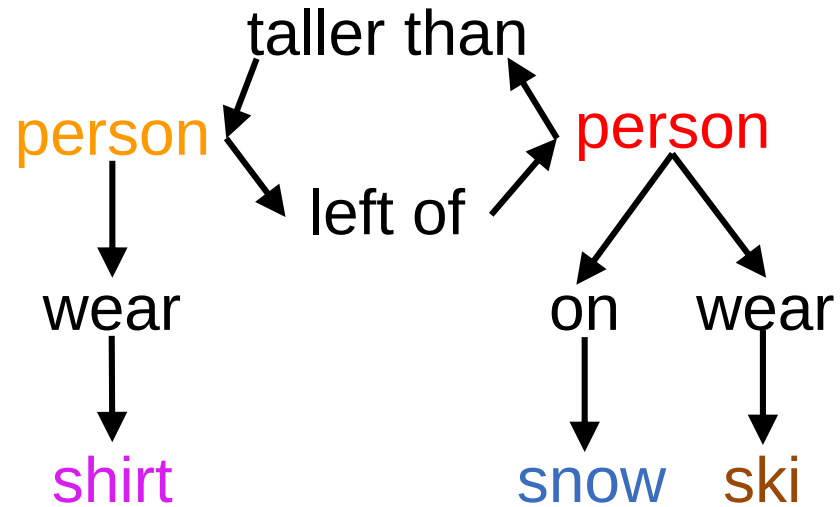
KJKF, CVPR
2017

A man is riding a carriage on a street. Two people are sitting on top of the horses. The carriage is made of wood. The carriage is black. The carriage has a white stripe down the side. The building in the background is a tan color.

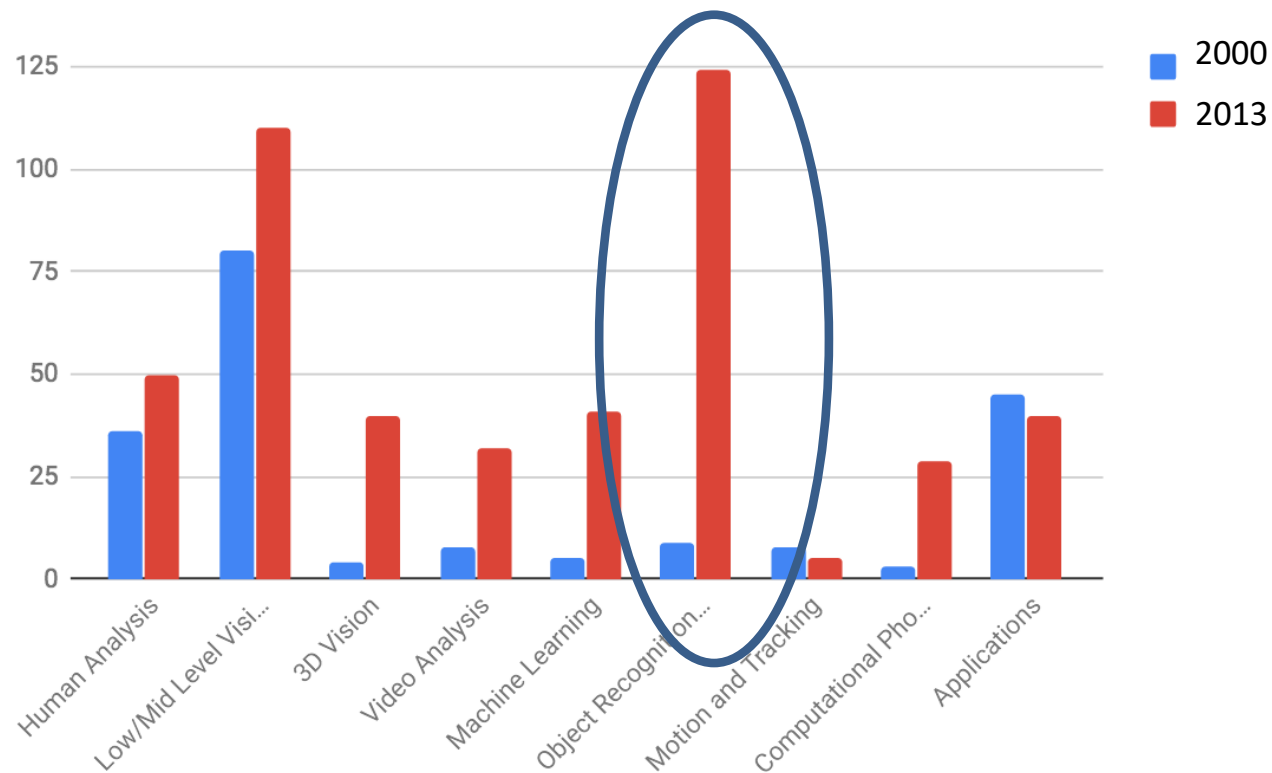


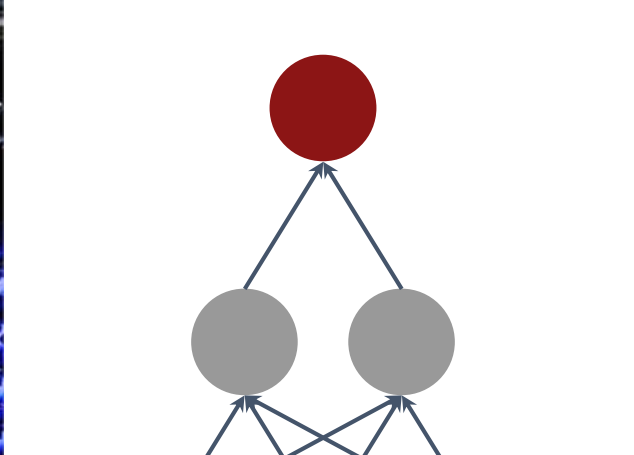
Results:

spatial, comparative, asymmetrical,
verb, prepositional

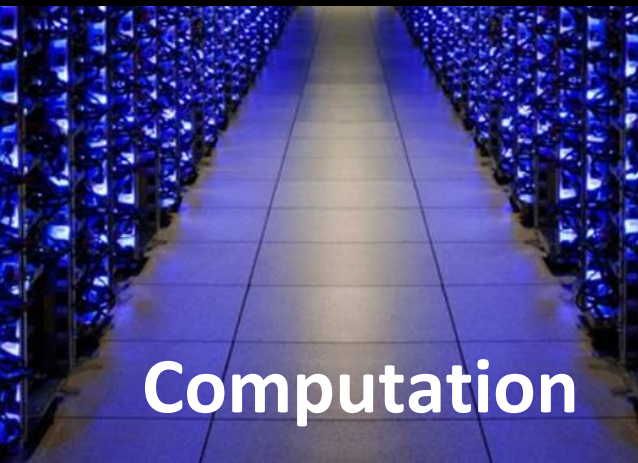


CVPR topic distribution: 2000 vs. 2013

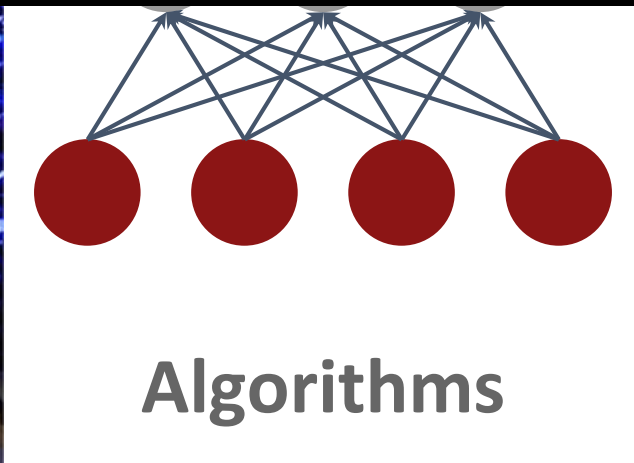




The Deep Learning Revolution



Computation



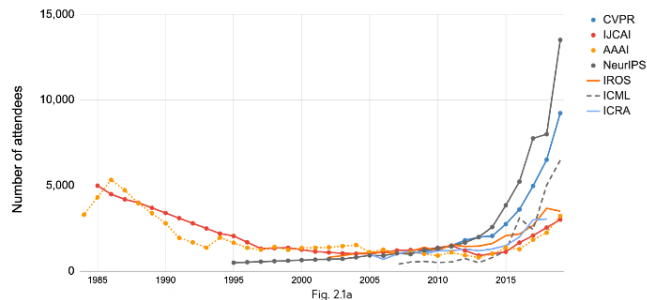
Algorithms



Data

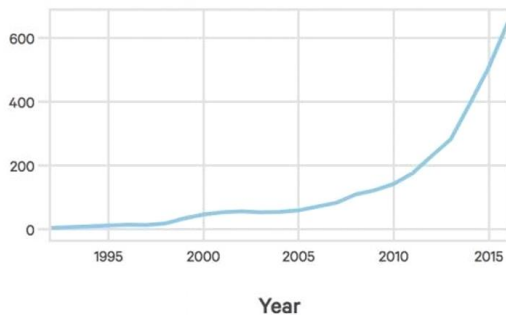
AI's Explosive Growth & Impact

Attendance at large conferences (1984-2019)
Source: Conference provided data.



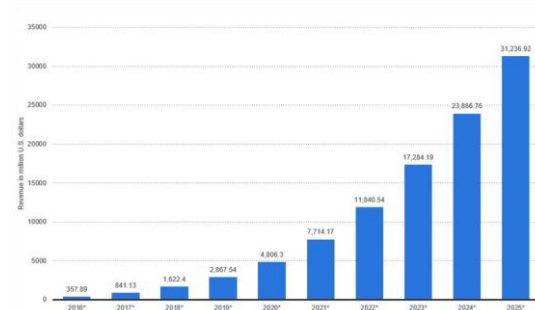
**Number of attendance
At AI conferences**

Source: The Gradient



**Startups Developing AI
Systems**

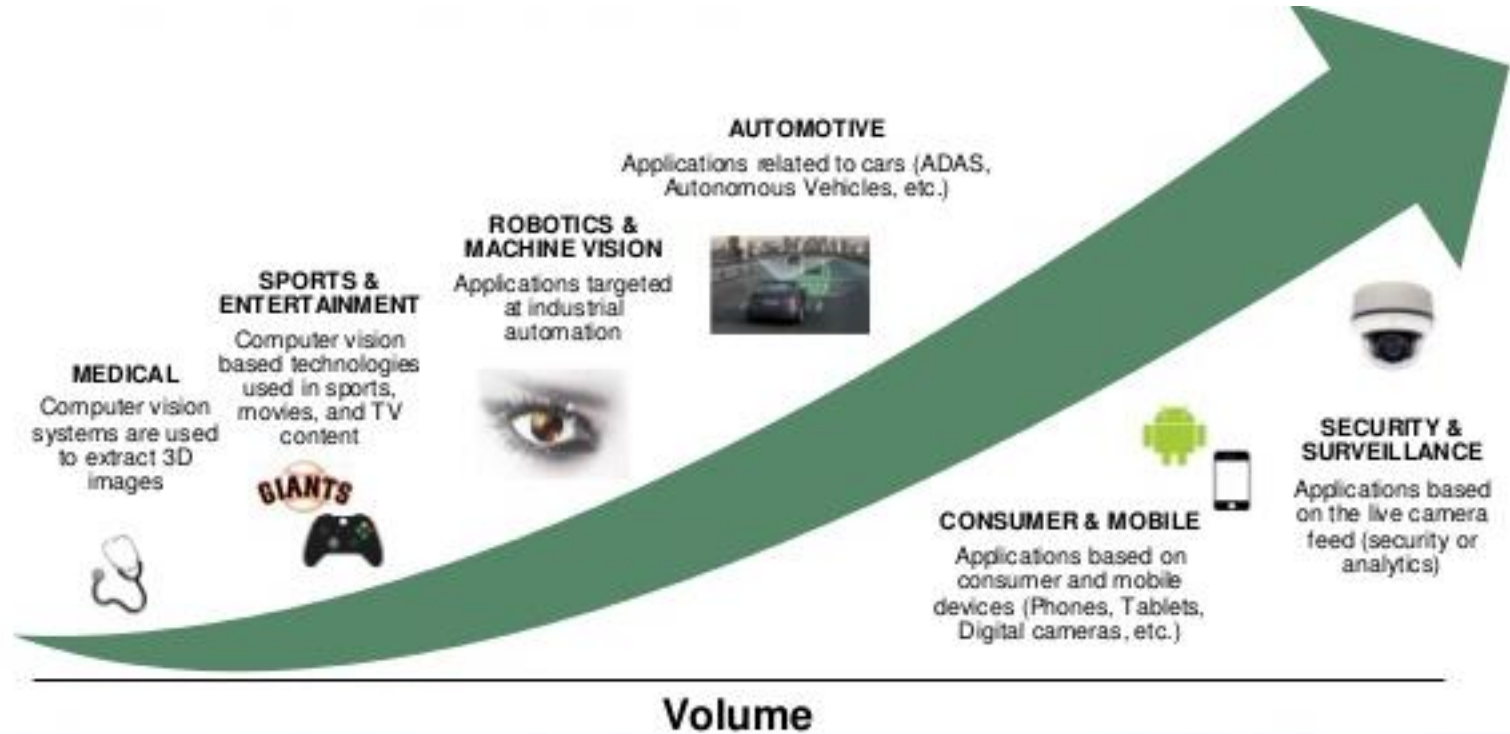
Source: Crunchbase, VentureSource, Sand
Hill Econometrics



**Enterprise Application AI
Revenue**

Source: Statista

Many Applications of computer vision



Slide source: World Capital Partners, 2017

How to take care of seniors while keeping them safe?



Early Symptom Detection
of COVID-19



Monitor Patients with
Mild Symptoms



Manage Chronic Conditions



Versatile



Mobility



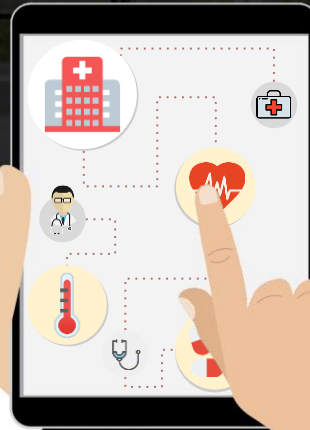
Infection



Sleep



Diet



Scalable



Low-cost



Burden-free



Today's agenda

- A brief history of computer vision
- CS231n overview