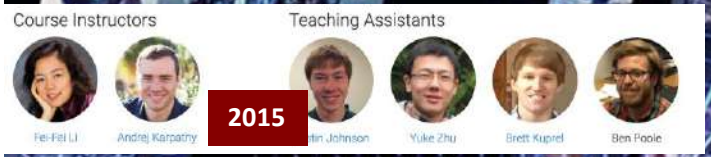
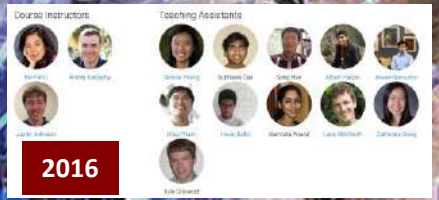
CS231n: Convolutional Neural Network for Visual Recognition

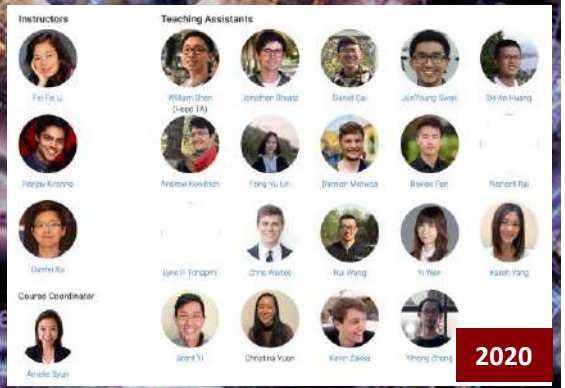
Lecture 1: Introduction

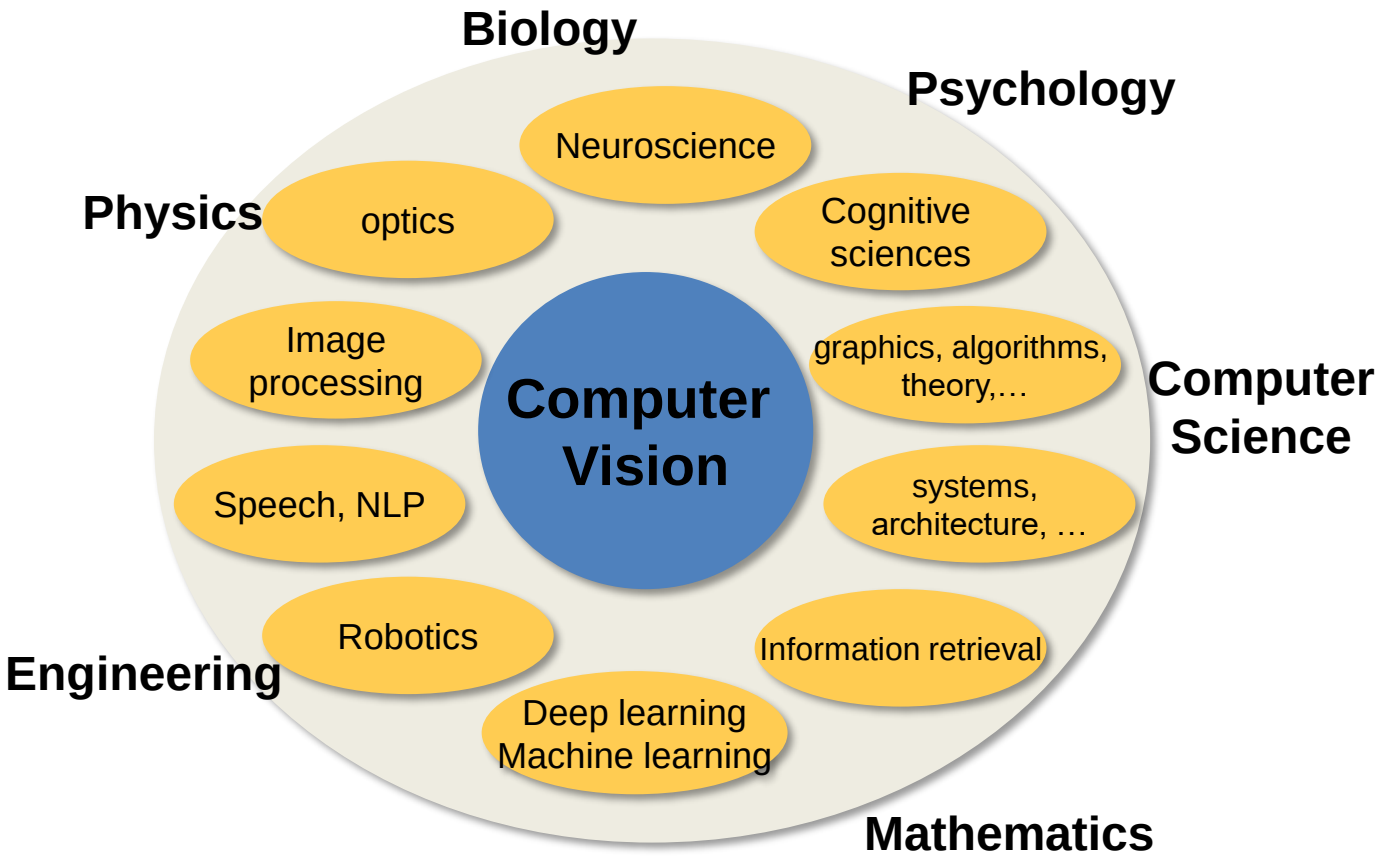
Welcome to CS231n

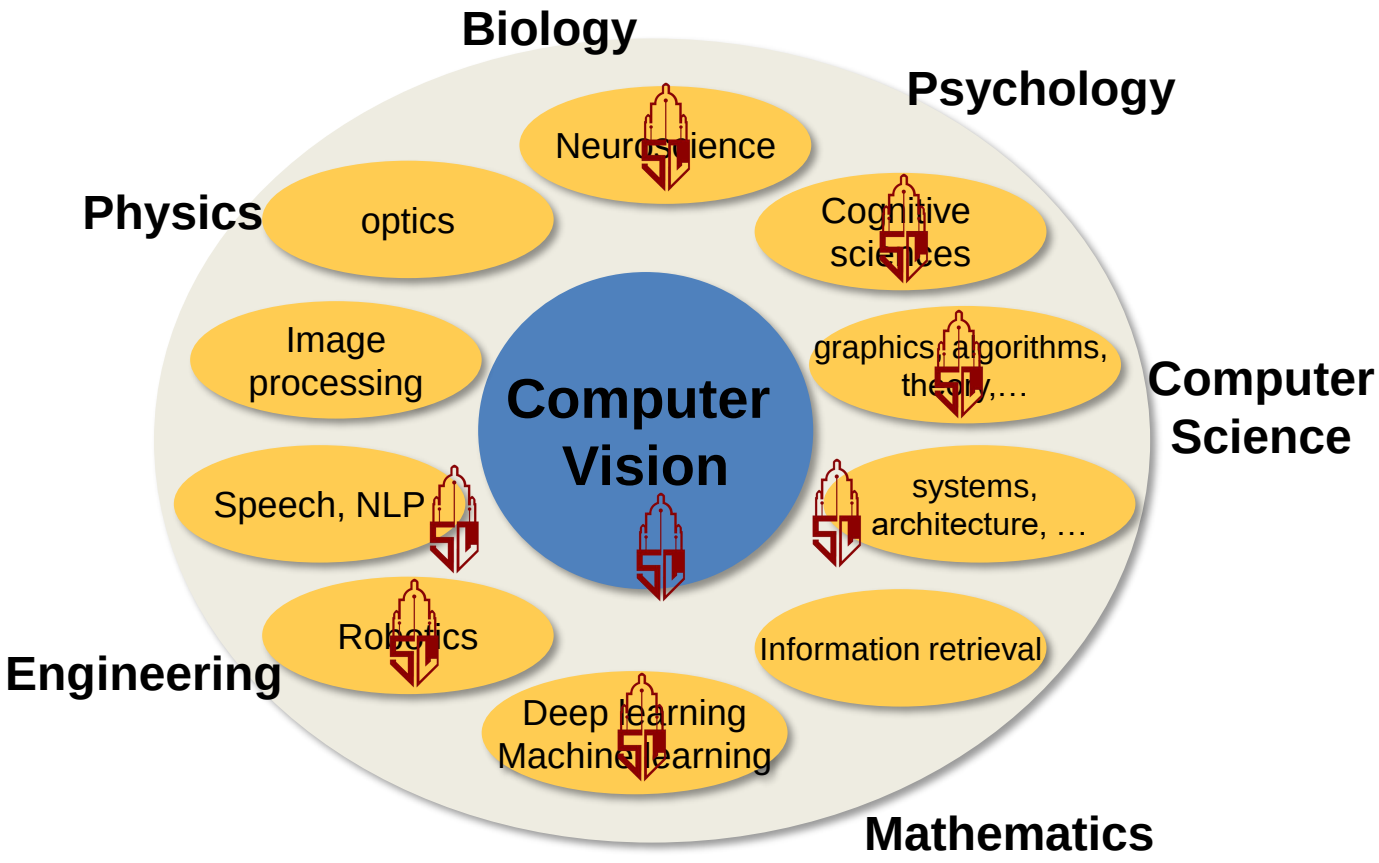
Welcome to CS231n

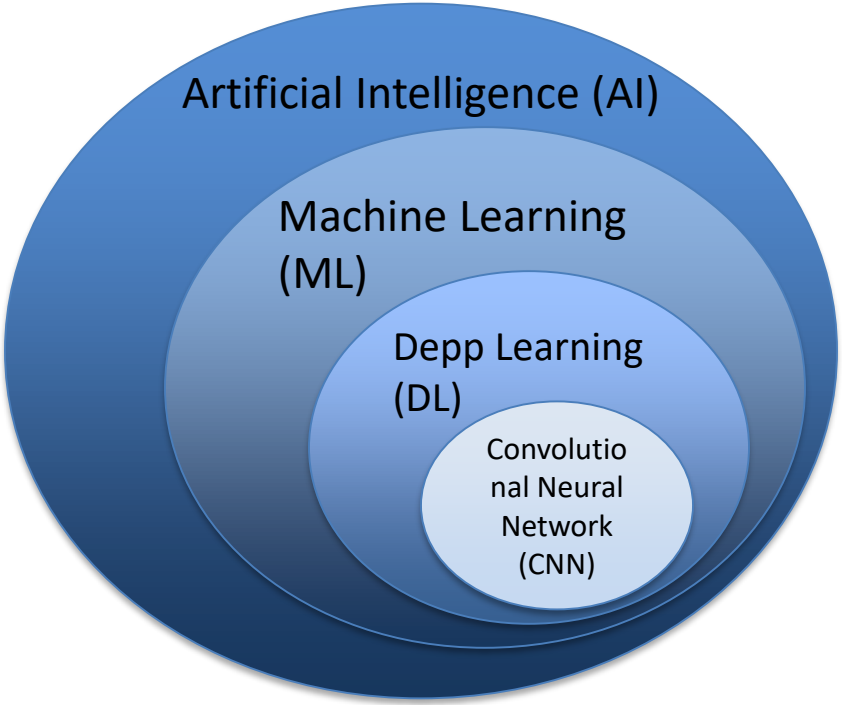


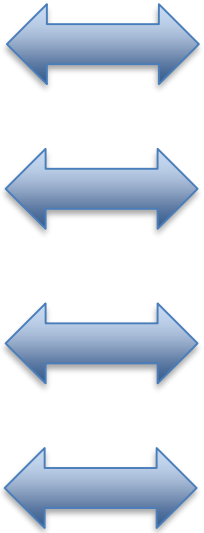





Biology
Psychology
Physics
Computer Science
Engineering
Mathematics
Computer Vision
Neuroscience
optics
Cognitive sciences
Image processing
graphics, algorithms, theory,…
Speech, NLP
systems, architecture, …
Robotics
Information retrieval
Deep learning
Machine learning

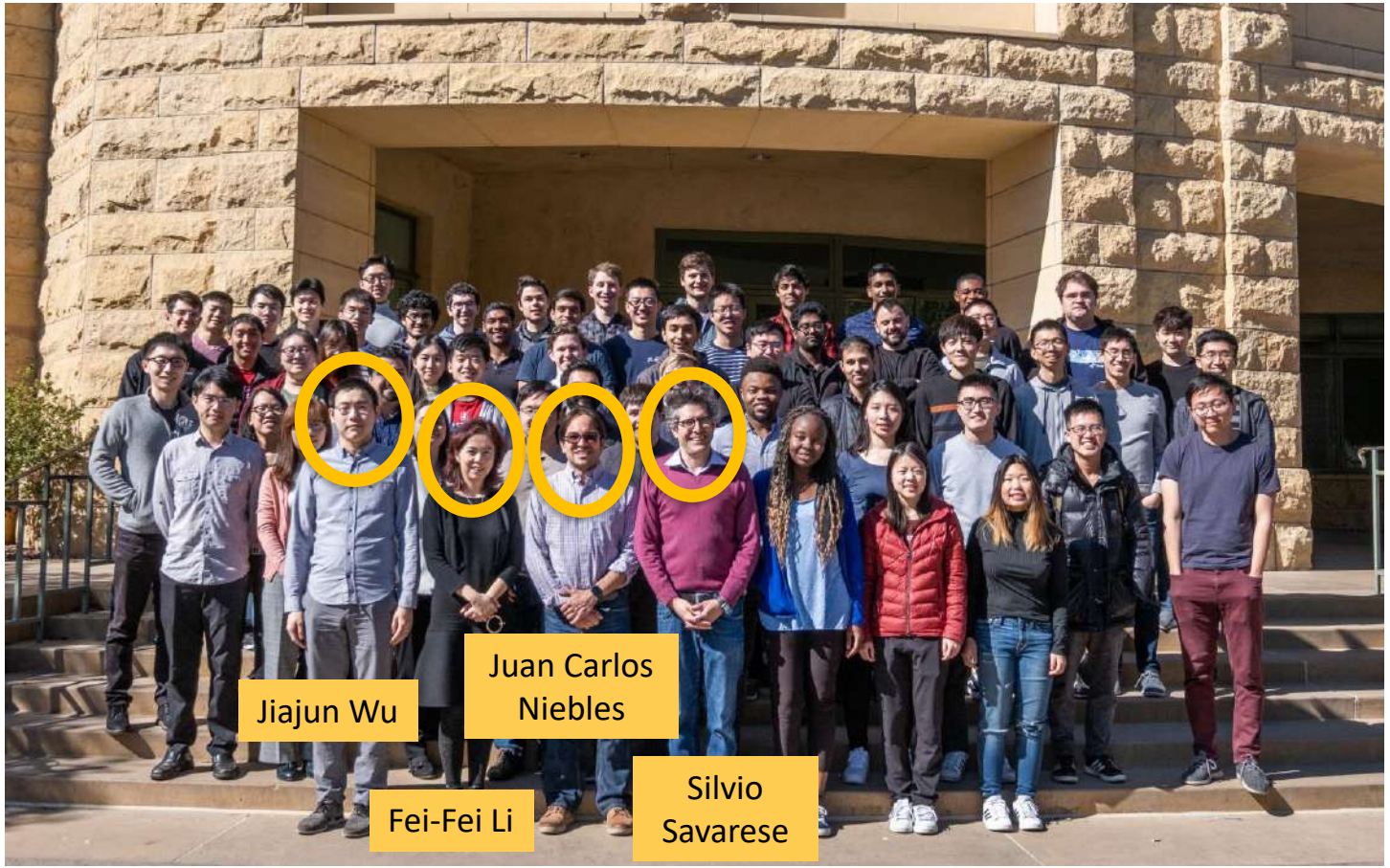

Biology
Psychology
Physics
Computer Science
Engineering
Mathematics
Neuroscience
optics
Cognitive sciences
Image processing
Computer Vision
graphics, algorithms, theory,…
Speech, NLP
systems, architecture, …
Robotics
Information retrieval
Deep learning
Machine learning





Computer Vision

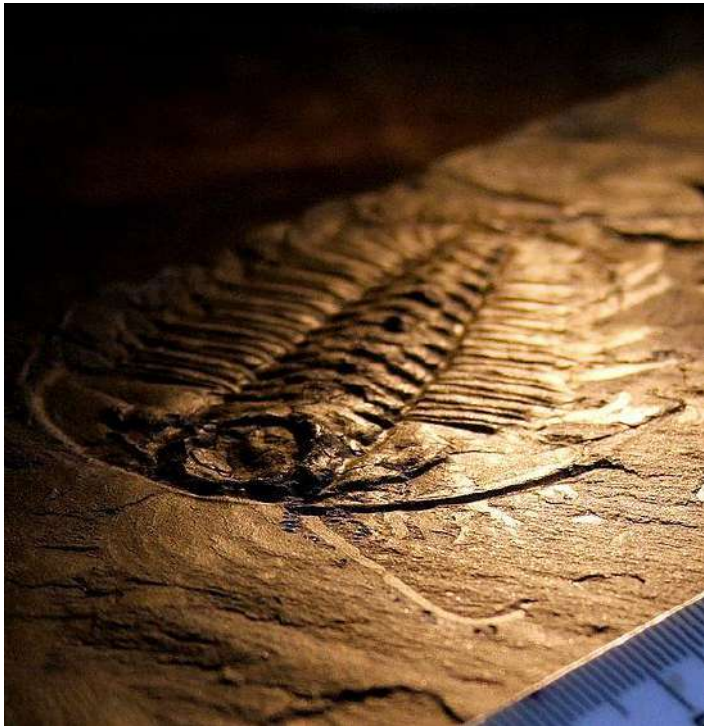
- Object detection
- Object classification
- Scene understanding
- Semantic scene segmentation
- 3D reconstruction
- Object tracking
- Human pose estimation
- Activity recognition
- VQA
- ….


Jiajun Wu
Juan Carlos Niebles
Fei-Fei Li
Silvio Savarese

Today's agenda

- A brief history of computer vision
- CS231n overview

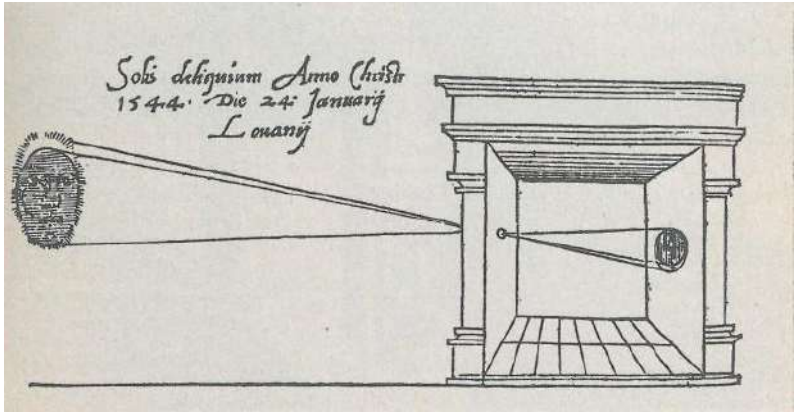
Evolution's Big Bang:
Cambrian Explosion, 530-540million years, B.C.

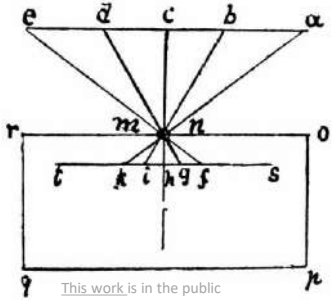





Camera Obscura

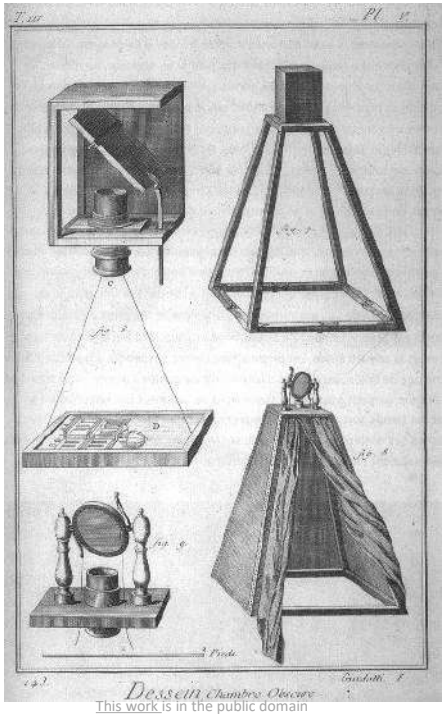
Gemma Frisius, 1545





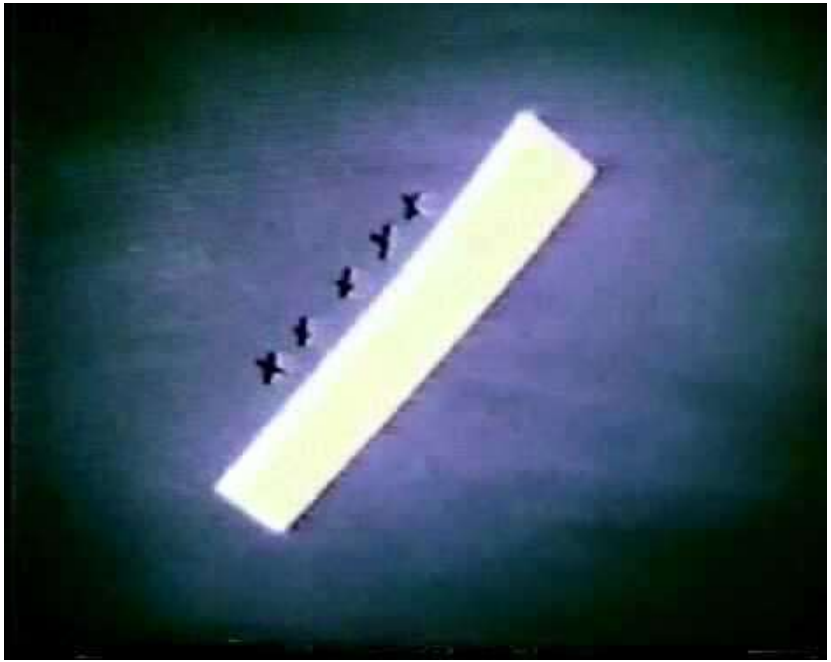
Leonardo da Vinci, 16th Century AD

Encyclopedia, 18th Century



Where did we come from?

The known story – Neuroscience inspired AI



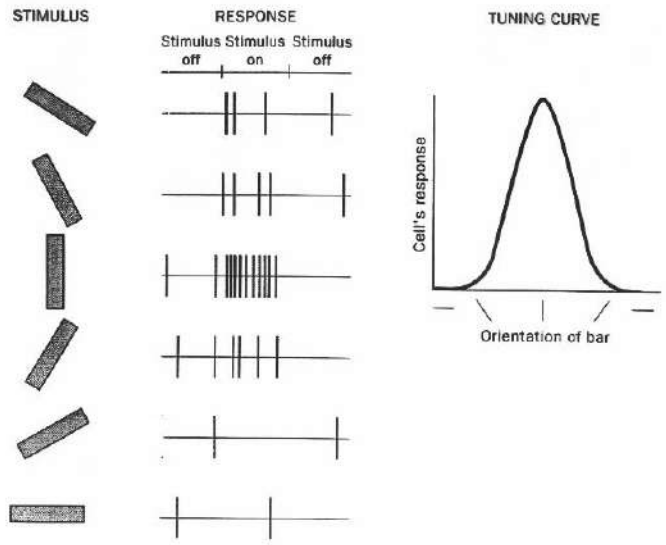
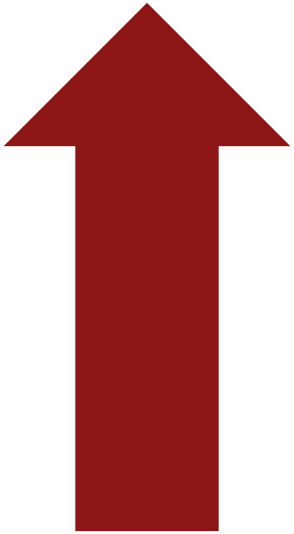


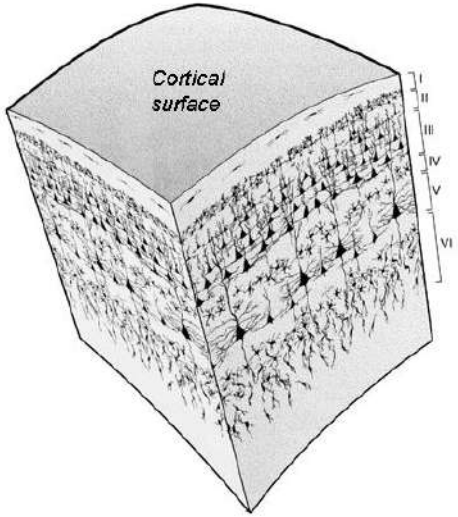
FIGURE 4.8 Response of a single cortical cell to bars presented at various orientations.

Hubel and Wiesel, 1959

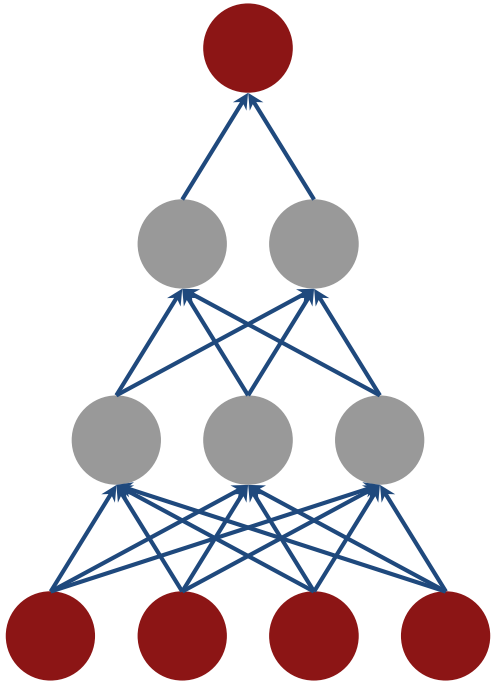
High-Level Patterns



Low-Level Details



Cortical Column (Biological)



Neural Networks (Digital)

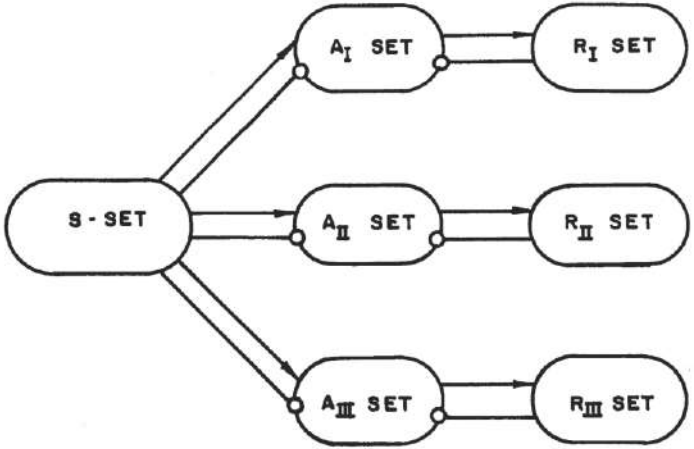


FIGURE 2

ORGANIZATION OF A PERCEPTRON WITH THREE INDEPENDENT OUTPUT-SETS

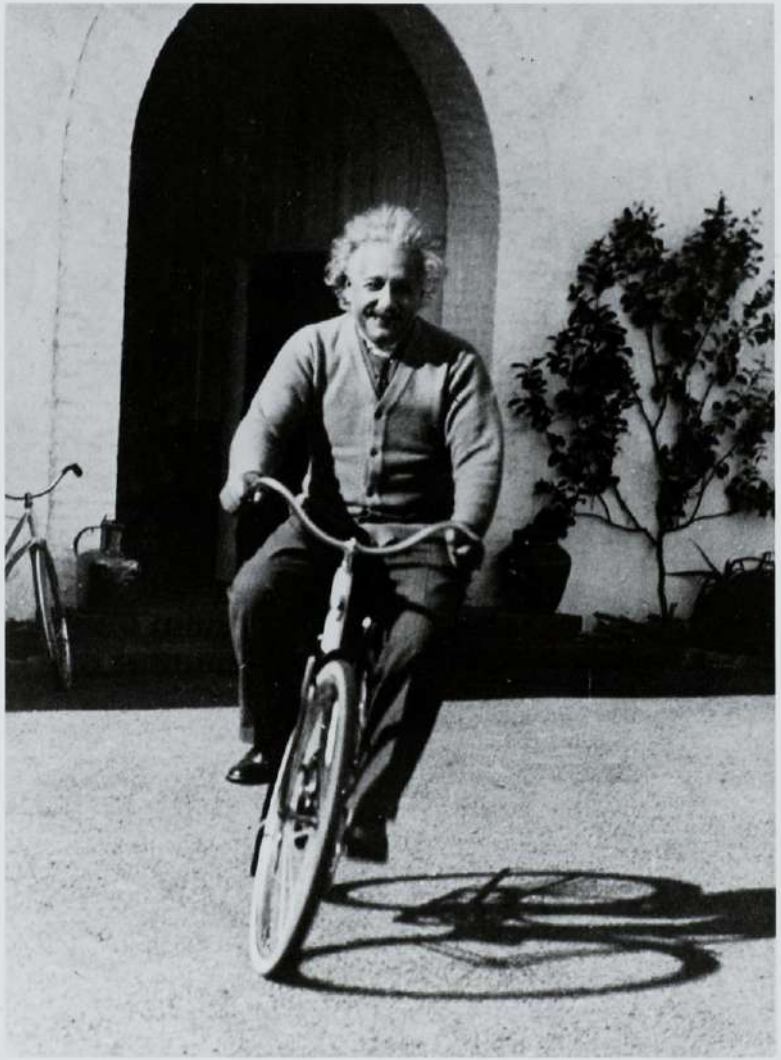
F. Rosenblatt, 1957

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."

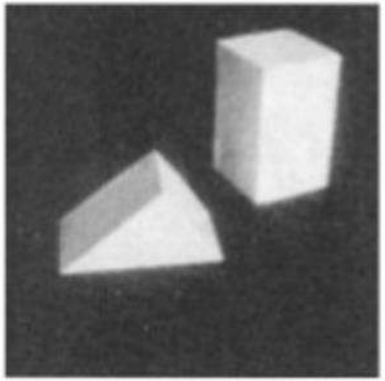
- Albert Einstein, 1921

Where did we come from?

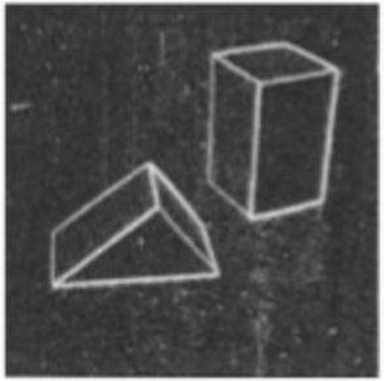
The not-so-known story – the search for computer vision's "North Star"

1960s: Interpretation of synthetic world

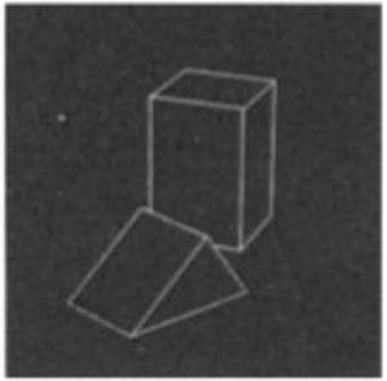
Larry Roerts
1963, 1st thesis of Computer Vision



Input image



2x2 gradient operator



computed 3D model
rendered from new viewpoint

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

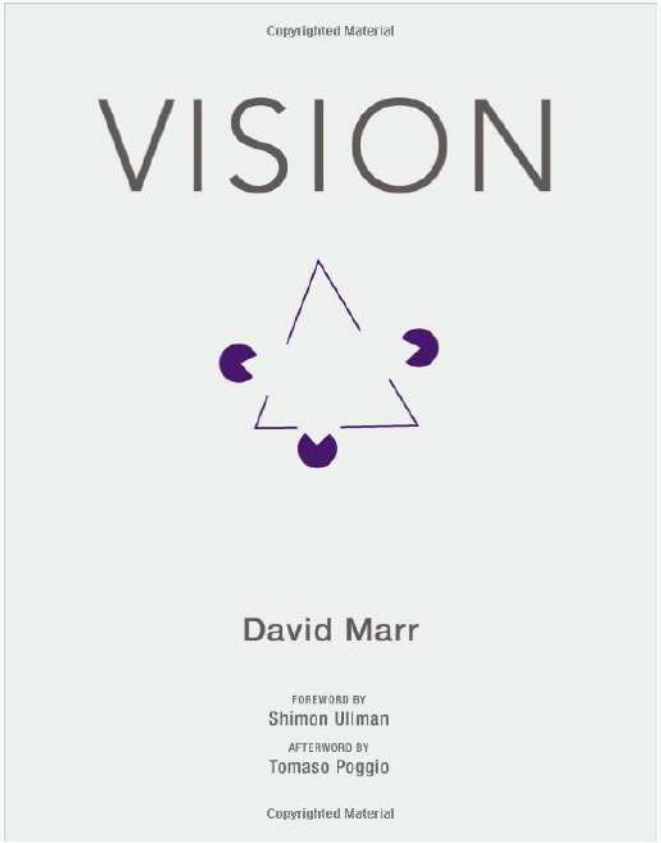
PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

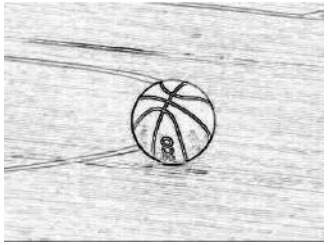
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".



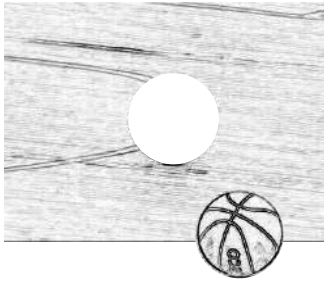
David Marr, 1970s

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

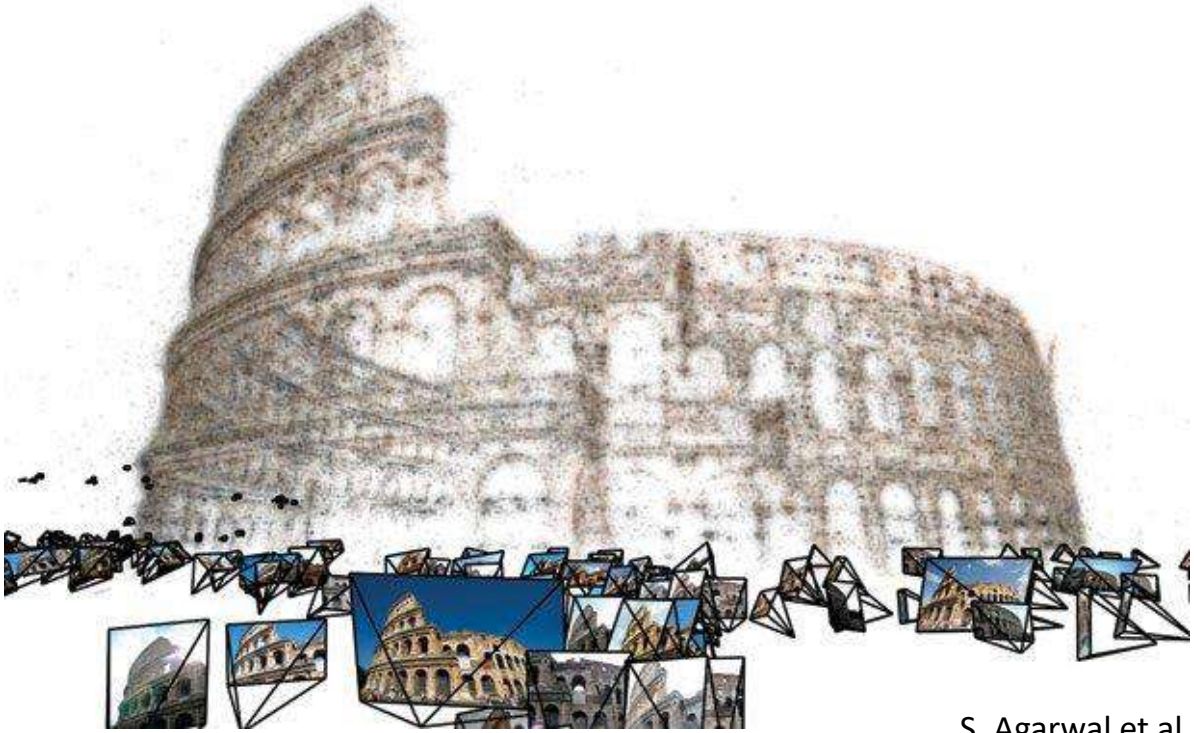
Stages of Visual Representation, David Marr, 1970s

Edges, segmentation, and perception

D. Lowe. *IJCV*, 1992

Normalized Cut (Shi & Malik, 1997)

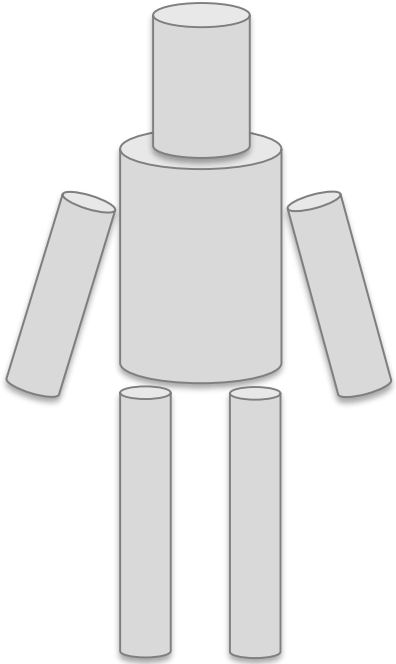
3D reconstruction



S. Agarwal et al. ICCV, 2009

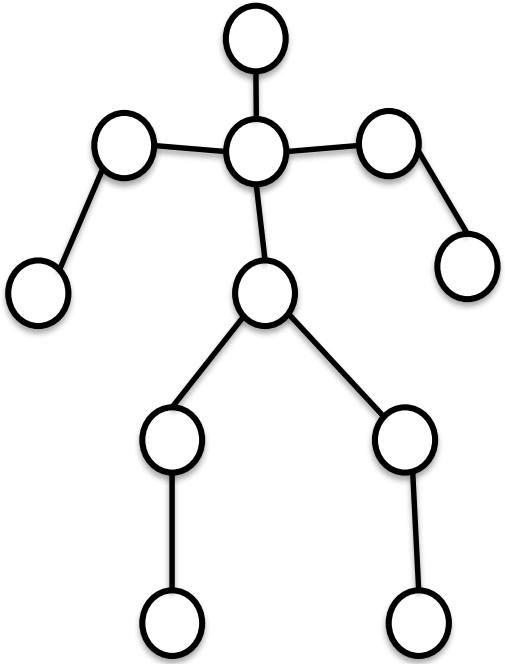
- Generalized Cylinder

Brooks & Binford, 1979



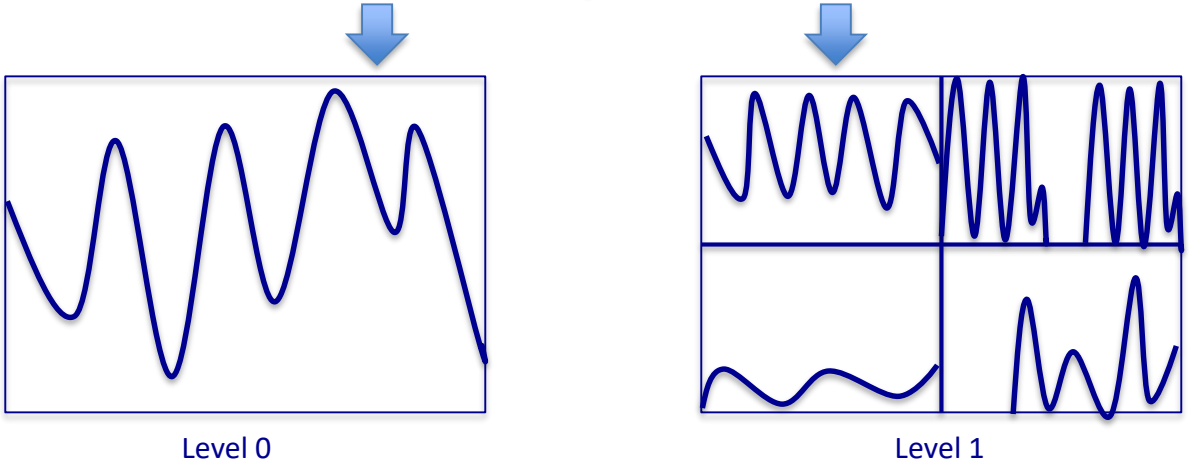
- Pictorial Structure

Fischler and Elschlager, 1973

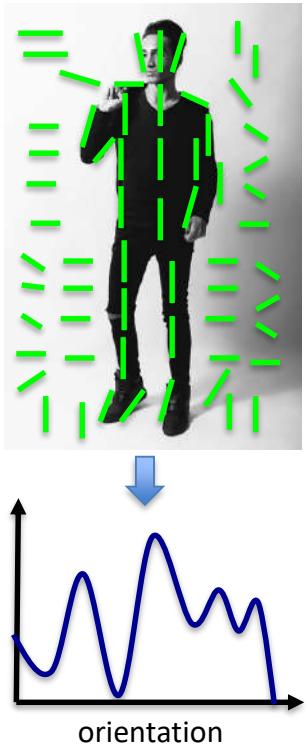


Single Object Recognition

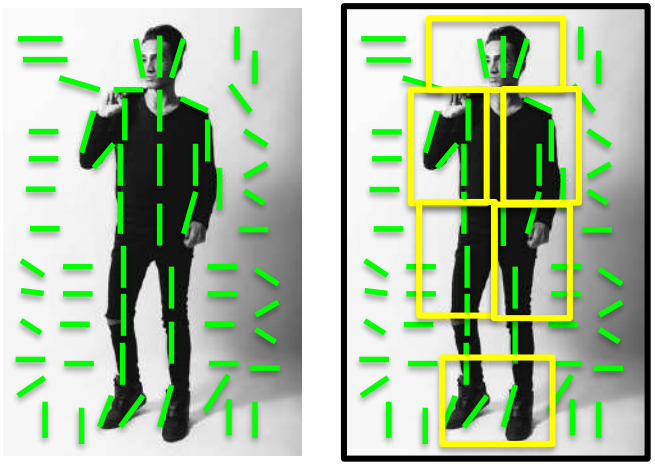
D. Lowe. *ICCV*, 1999



Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006

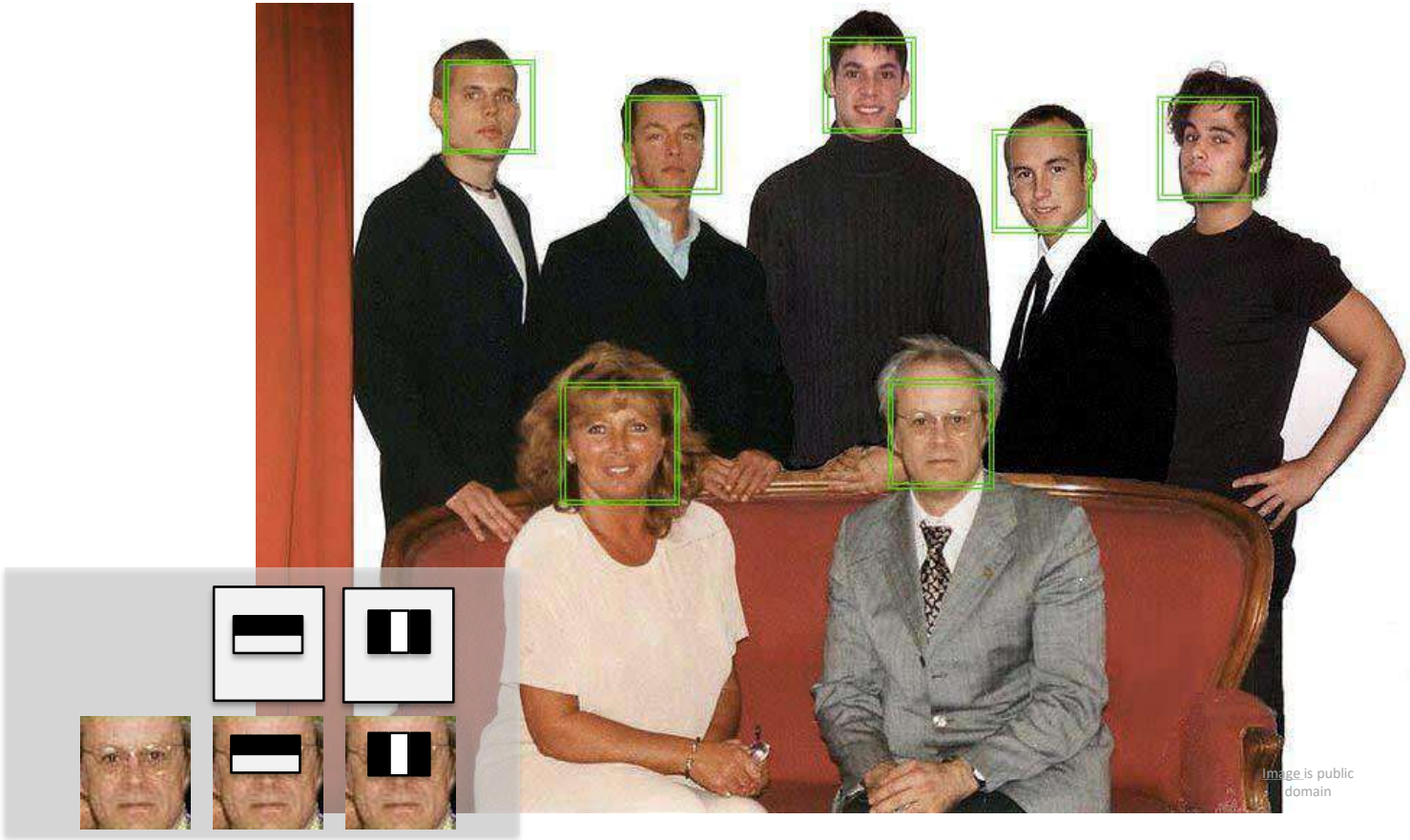


Histogram of Gradients (HoG)
Dalal & Triggs, 2005

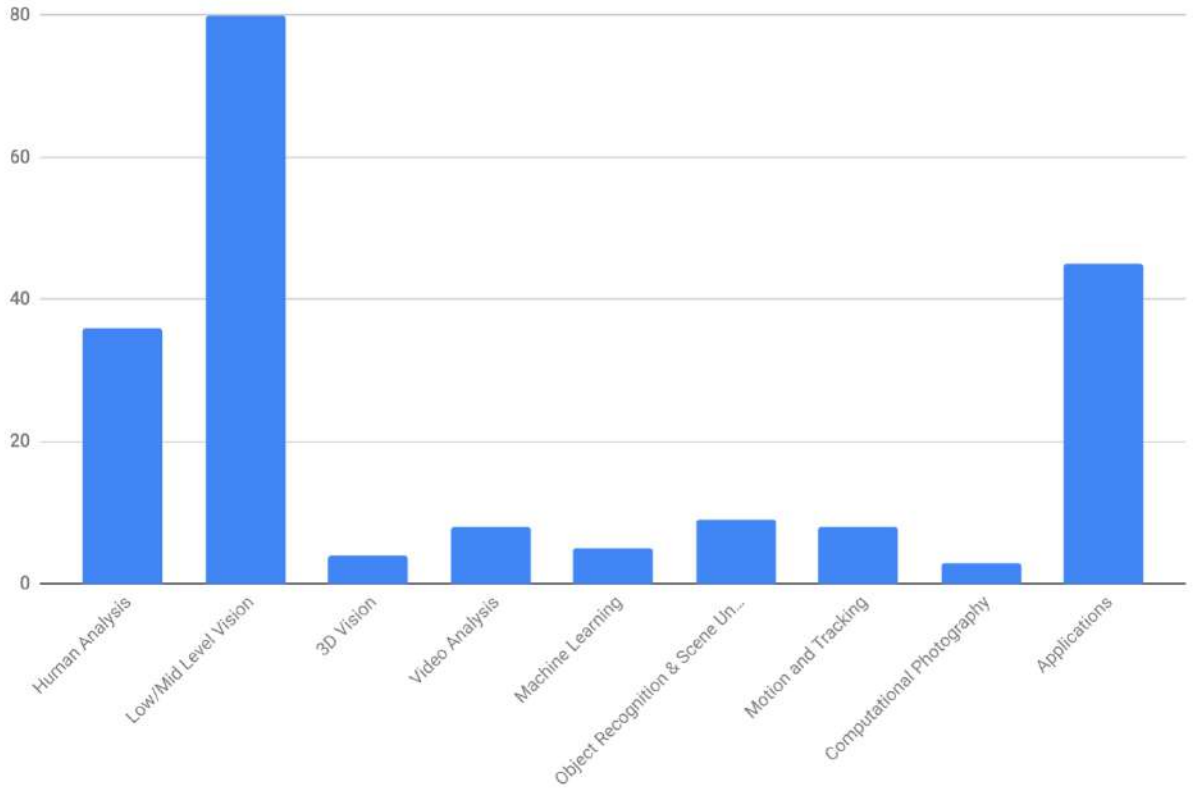


Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009

Face Detection, Viola & Jones, 2001



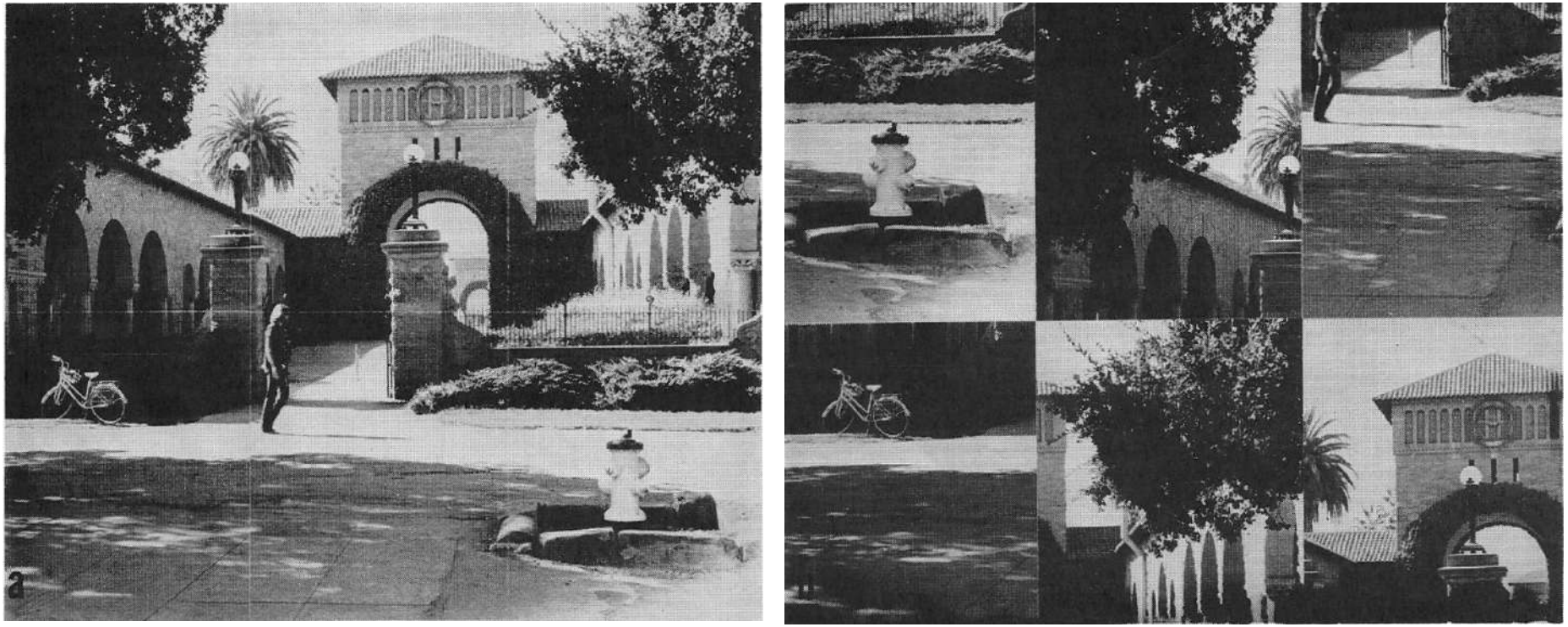
CVPR topic distribution: 2000



In the mean time…

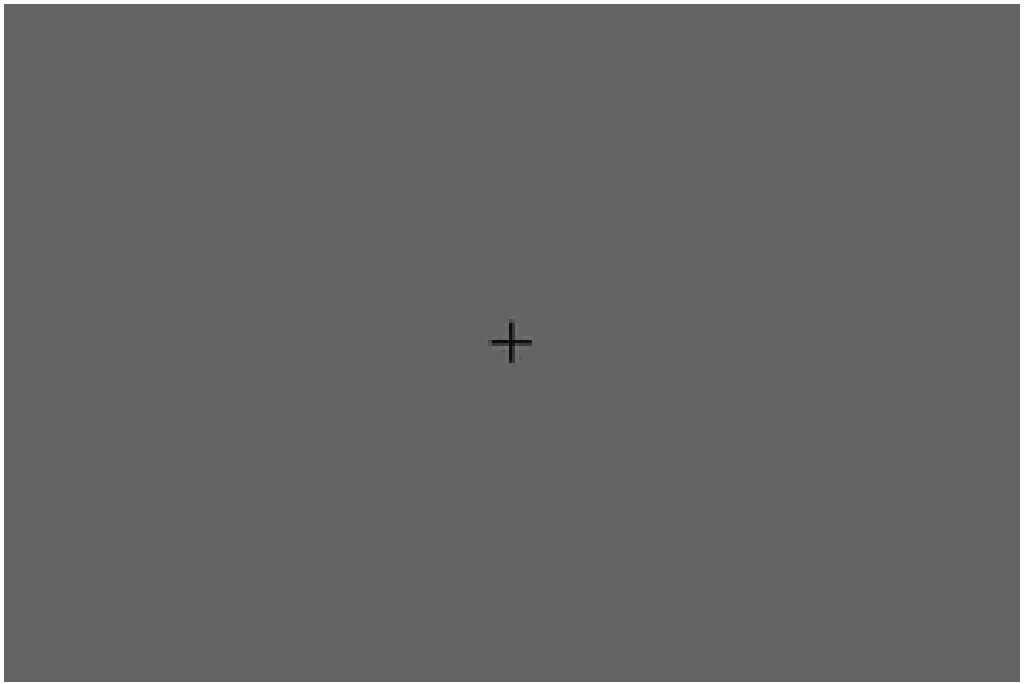
Perceiving Real-World Scenes

Irving Biederman



I. Biederman, *Science*, 1972

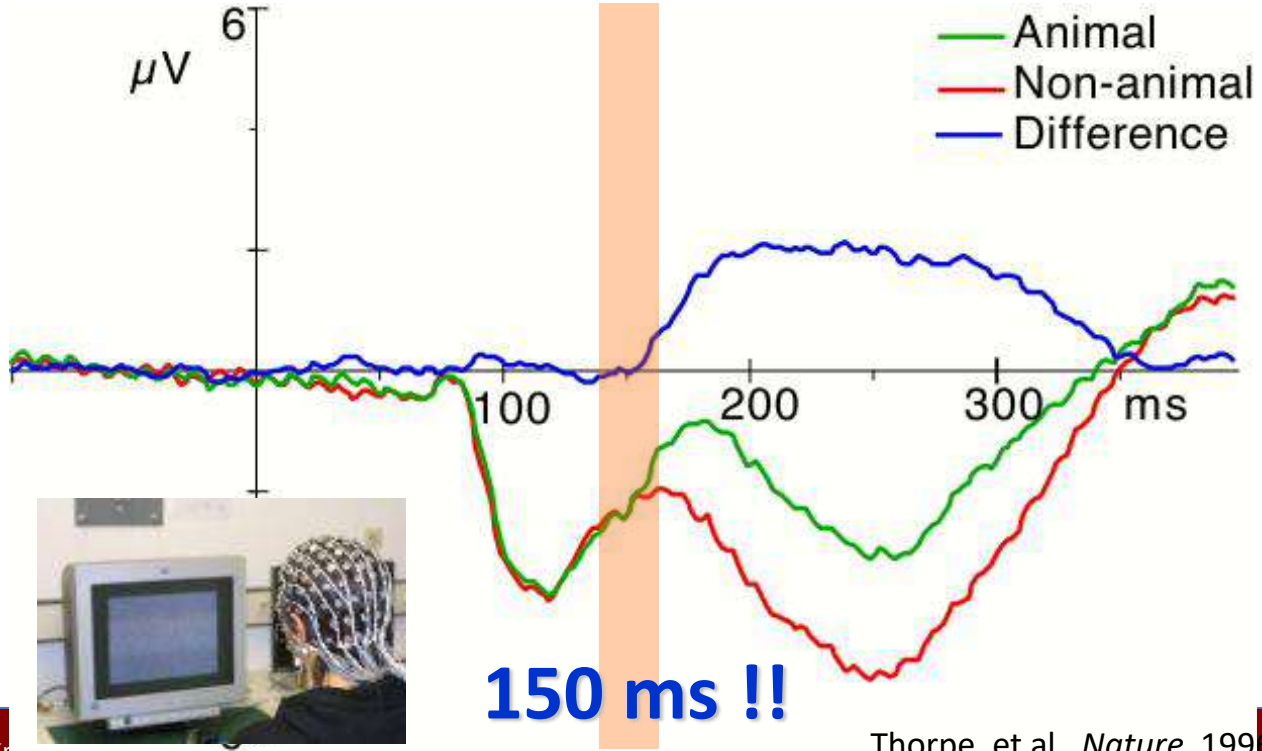
Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

Speed of processing in the human visual system

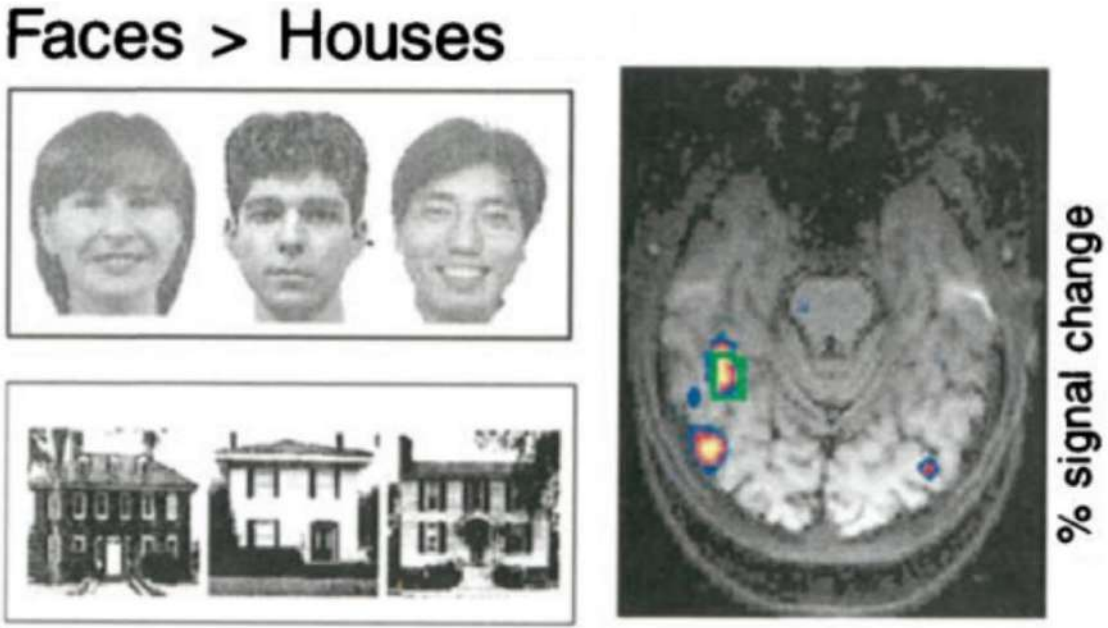
Simon Thorpe, Denis Fize & Catherine Marlot



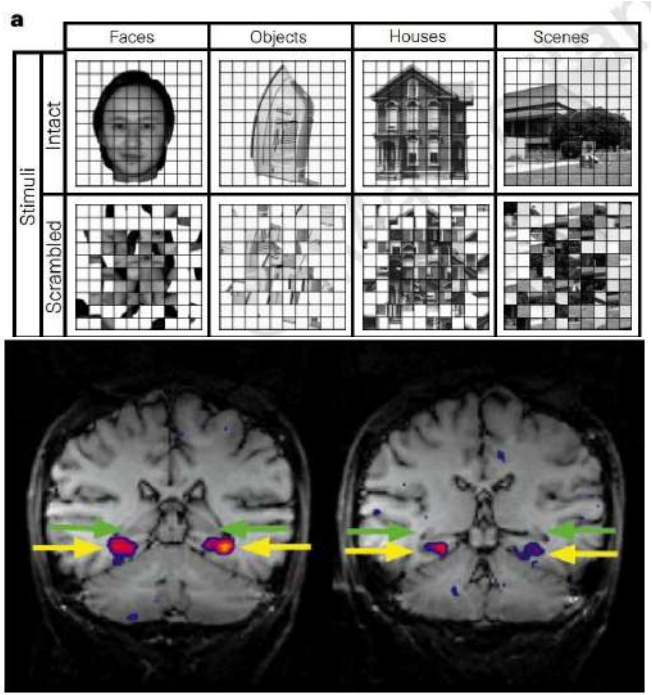
150 ms !!

Thorpe, et al. *Nature*, 1996

Neural correlates of object & scene recognition



Kanwisher et al. J. Neuro. 1997



Epstein & Kanwisher, Nature, 1998

A Computer Vision/AI ”holy grail” – Object Recognition

Caltech 101 images

Fei-Fei et al. 2004

Visual Object Classes Challenge 2009 (VOC2009)

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

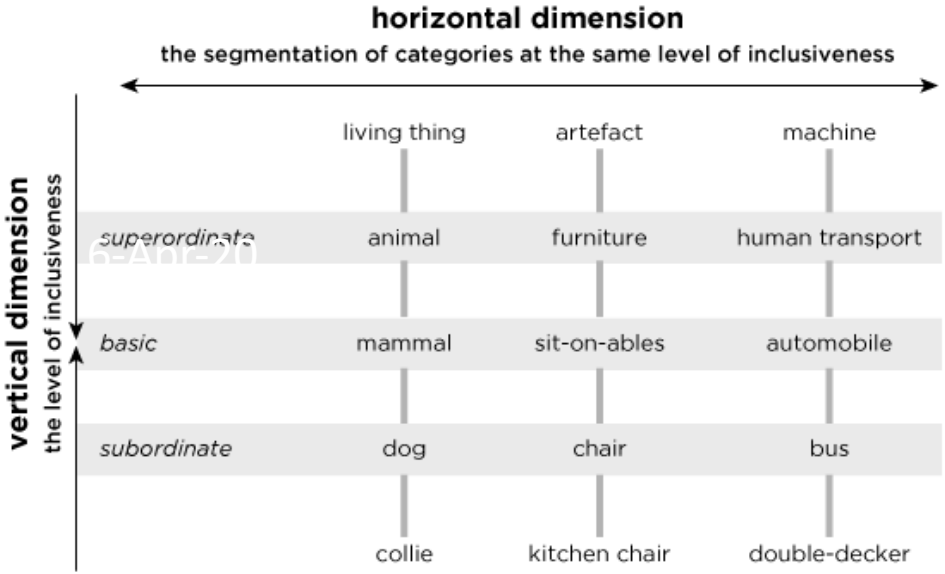
[click on an image to see the annotation]

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 × 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.



- Biederman: Recognition by Component, 1987

Eleanor Rosch: Principles of Categorization, 1978

George A. Miller

Psychology, Cognitive Science
Princeton University

WORDNET

G. A. Miller, *Communications of the ACM*, 1995

IMAGENET

22,000 categories ⋮ 15,000,000 images

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li & L. Fei-Fei. *CVPR*, 2009.

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

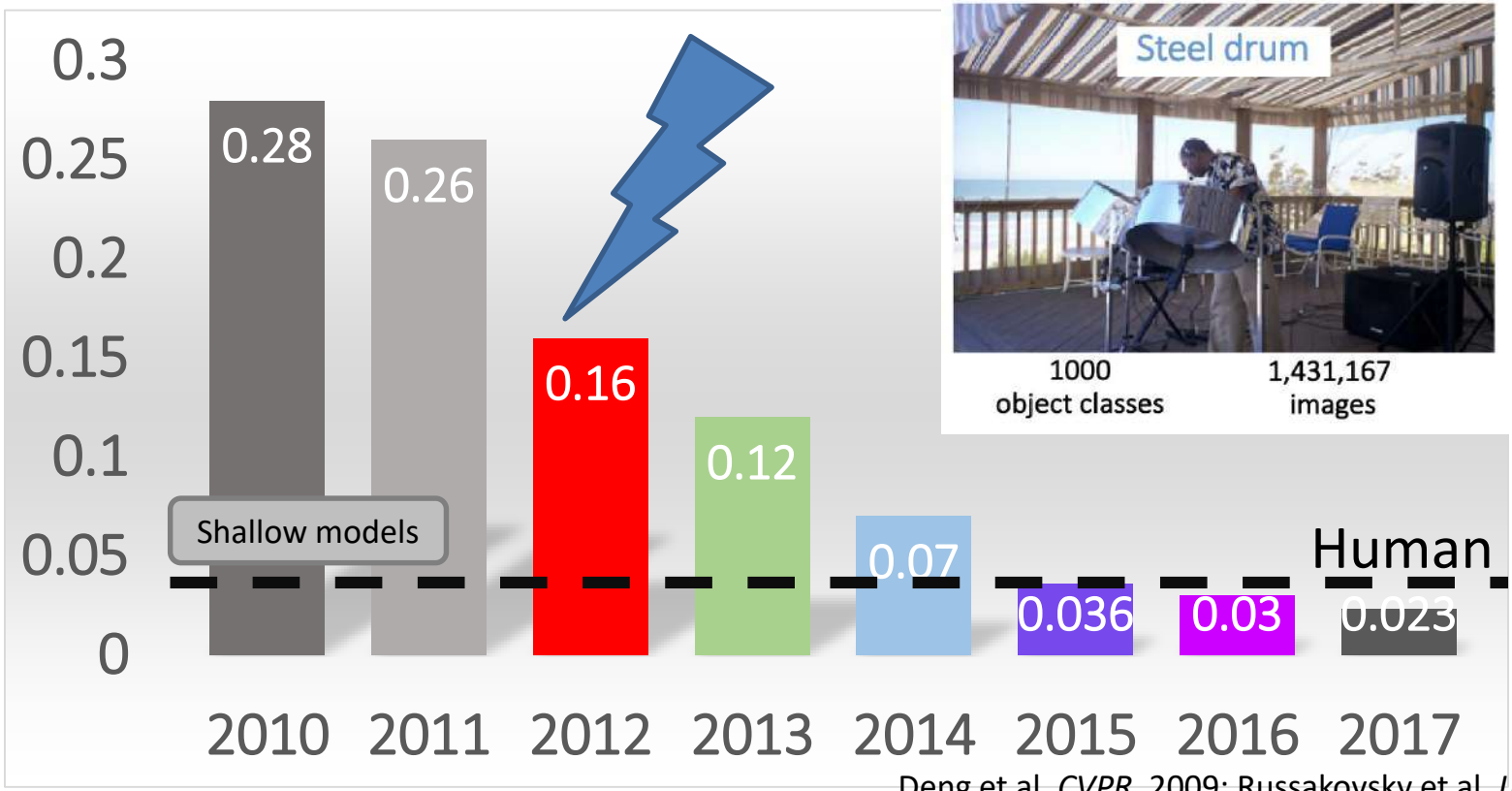
✓

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

✗

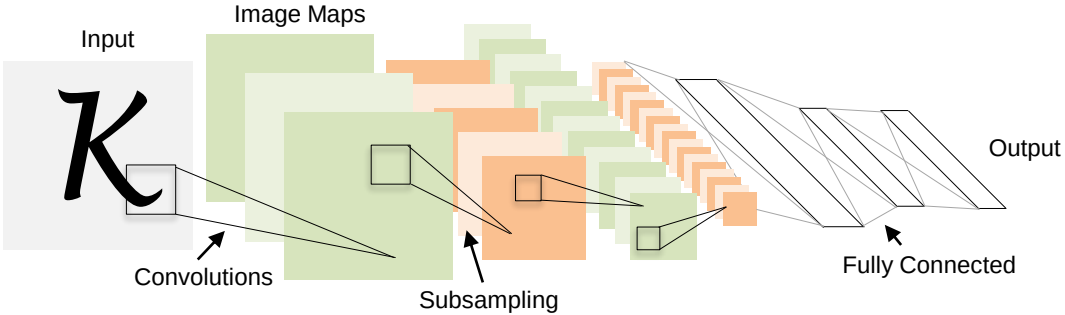
Russakovsky et al. IJCV 2015

IMAGENET Classification **Task**



Deng et al. *CVPR*, 2009; Russakovsky et al. *IJCV*, 2012;

1998
LeCun et al.



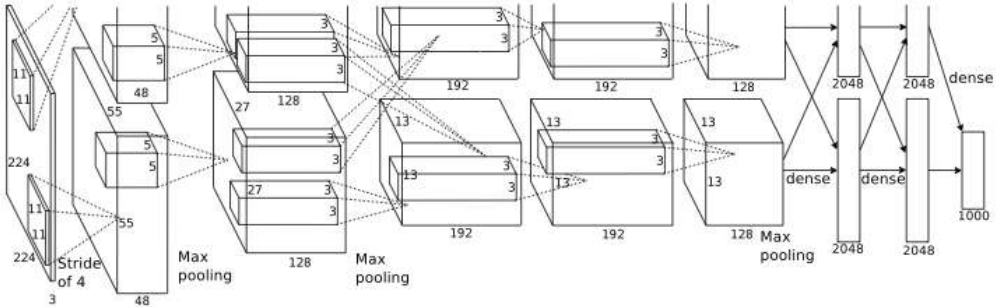
# of transistors

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

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

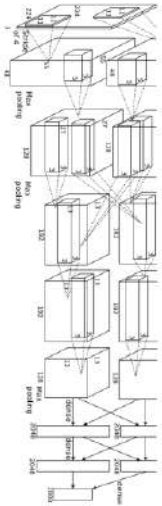
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

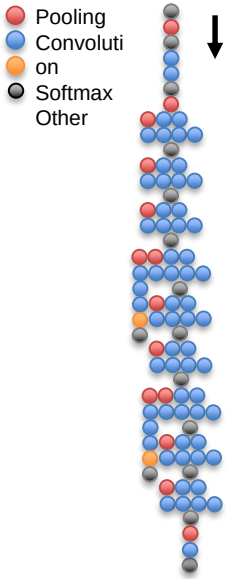
SuperVision



[Krizhevsky NIPS 2012]

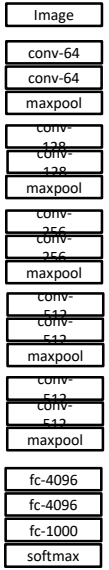
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

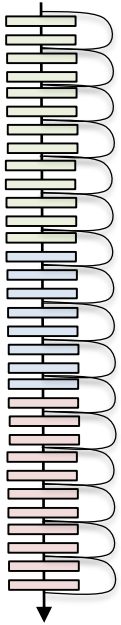
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

Image Captioning: Richer Descriptions

Prior Work

Image Captioning →

A man riding a horse drawn carriage down a street

Our Recent Work

Dense Captioning →

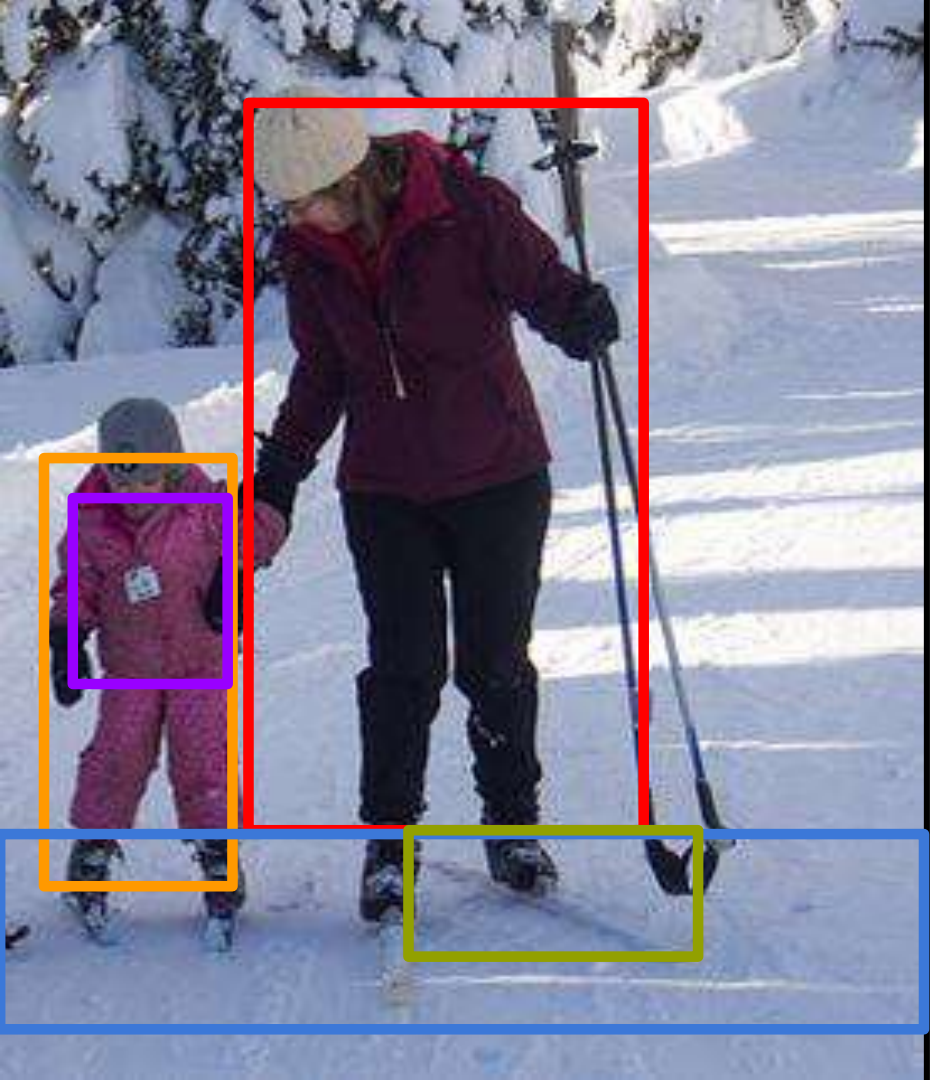
JKF, CVPR 2016

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.

Paragraph Captioning →

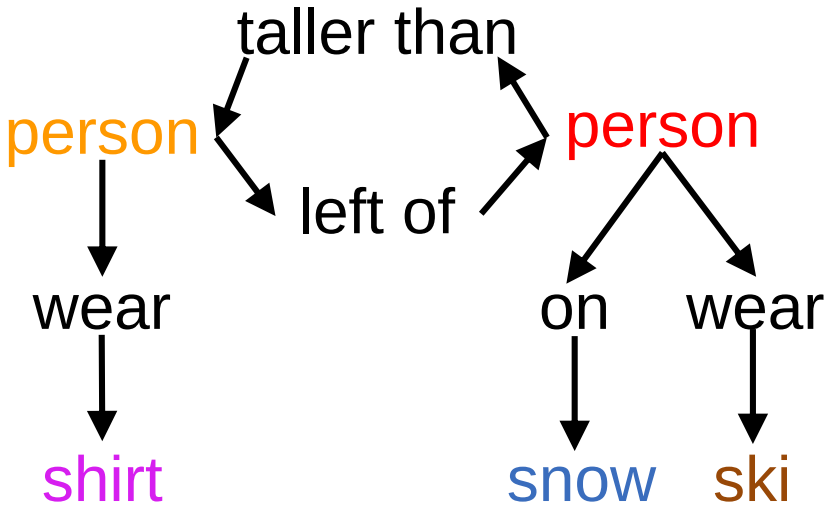
KJKF, CVPR 2017

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.



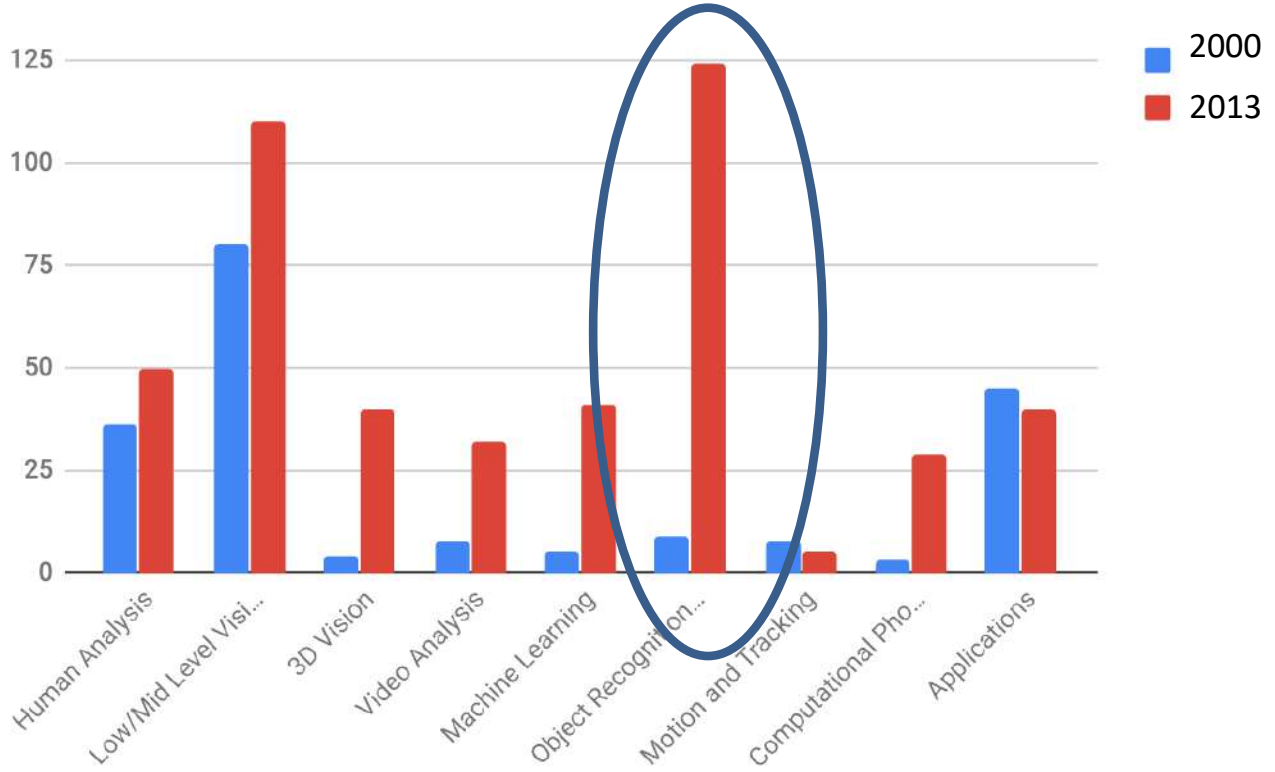
Results:

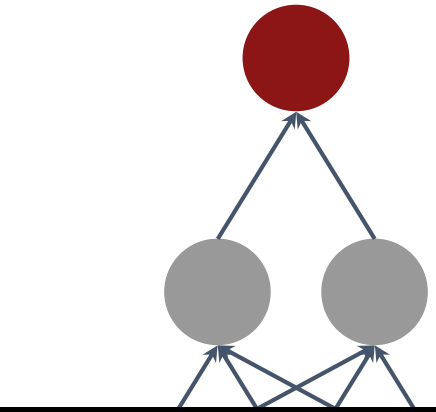
spatial, comparative, asymmetrical, verb, prepositional



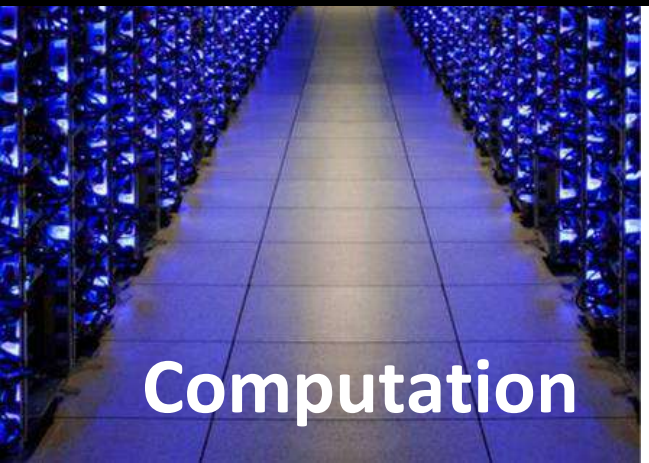
Krishna*, Lu*, Bernstein, Fei-Fei, *ECCV* 2016

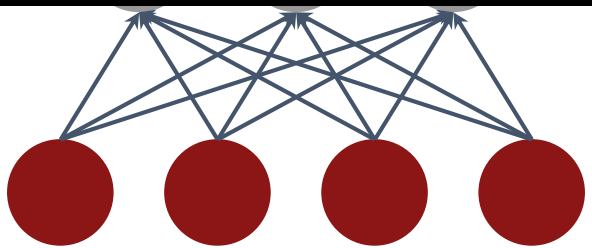
CVPR topic distribution: 2000 vs. 2013





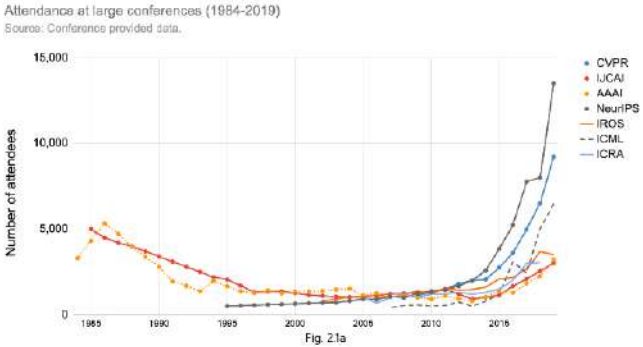
The Deep Learning Revolution





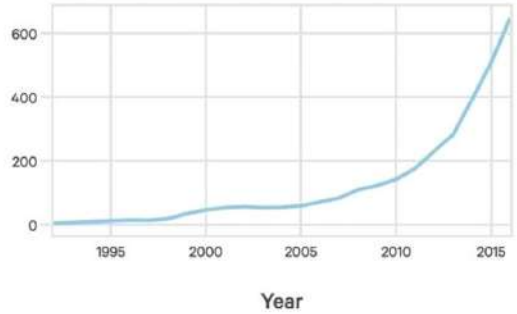
Algorithms

AI's Explosive Growth & Impact



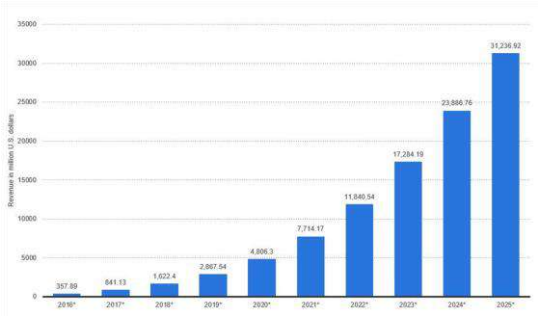
Number of attendance At AI conferences

Source: The Gradient



Startups Developing AI Systems

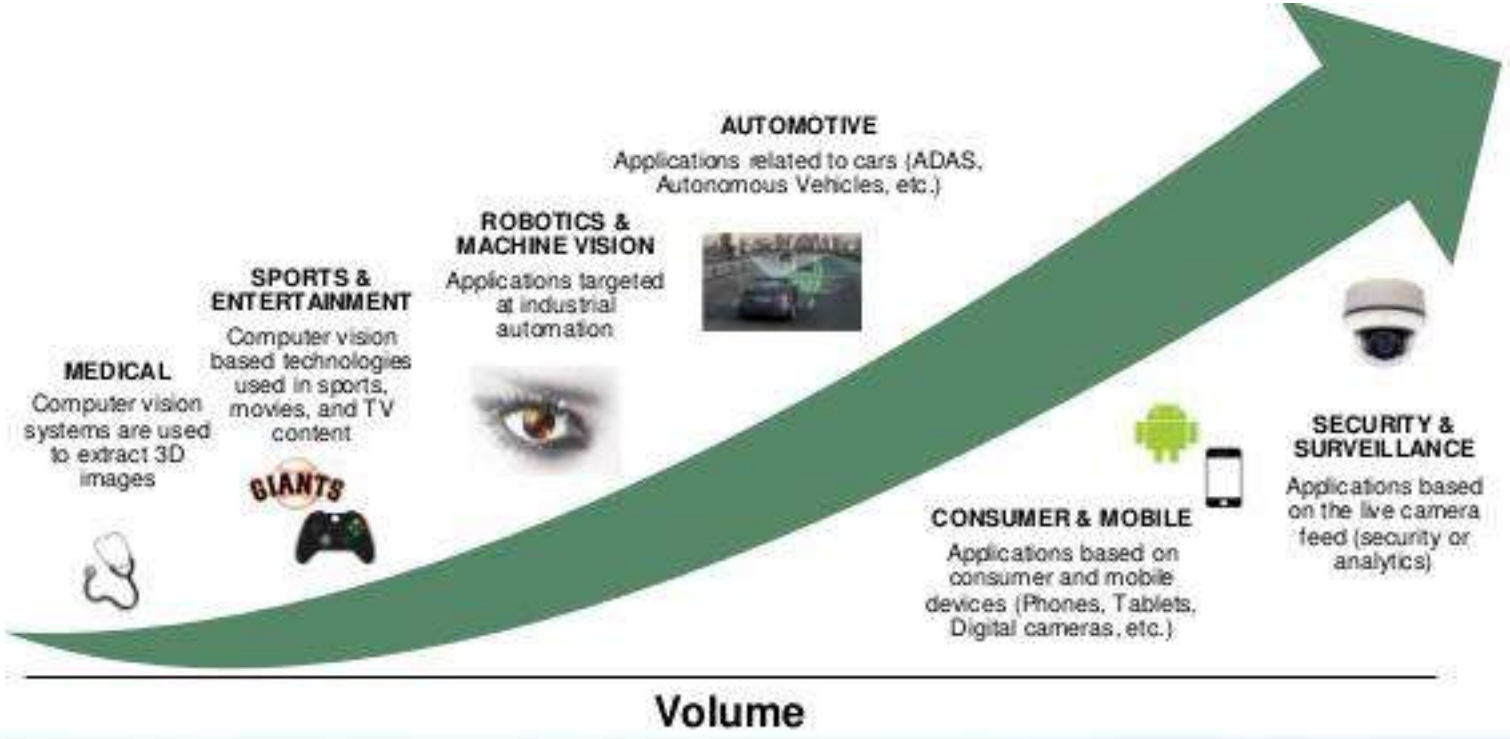
Source: Crunchbase, VentureSource, Sand Hill Econometrics



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