



Lecture 1: Introduction

Welcome to CS231n



Top row, left to right:

[Image by Augustas Didžgalvis](#); licensed under [CC BY-SA 3.0](#); changes made

[Image by Nesster](#) is licensed under [CC BY-SA 2.0](#)

[Image is CC0 1.0 public domain](#)

[Image is CC0 1.0 public domain](#)

Middle row, left to right

[Image by BGPHP Conference](#) is licensed under [CC BY 2.0](#); changes made

[Image is CC0 1.0 public domain](#)

[Image by NASA](#) is licensed under [CC BY 2.0](#); chang

[Image is CC0 1.0 public domain](#)

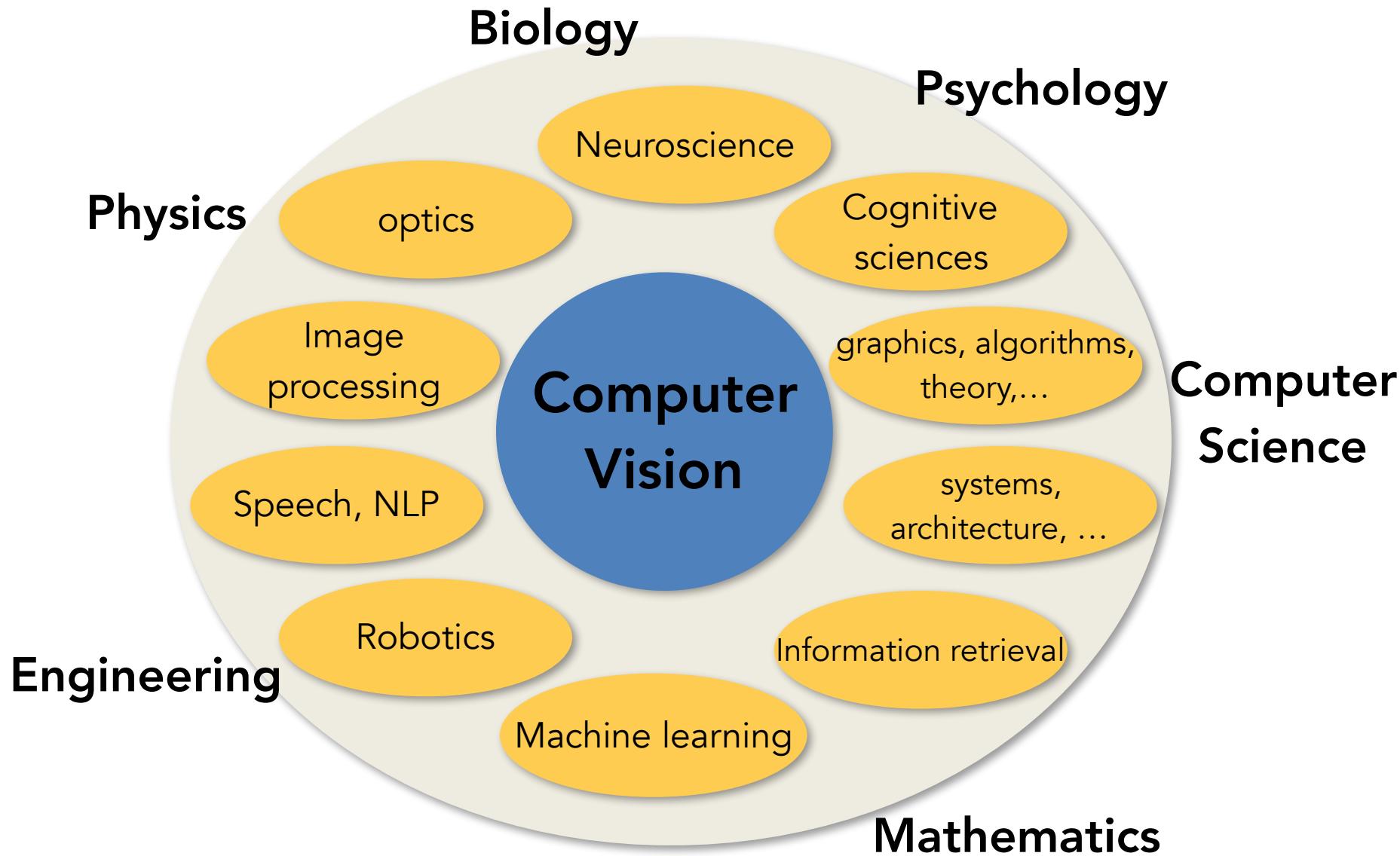
Bottom row, left to right

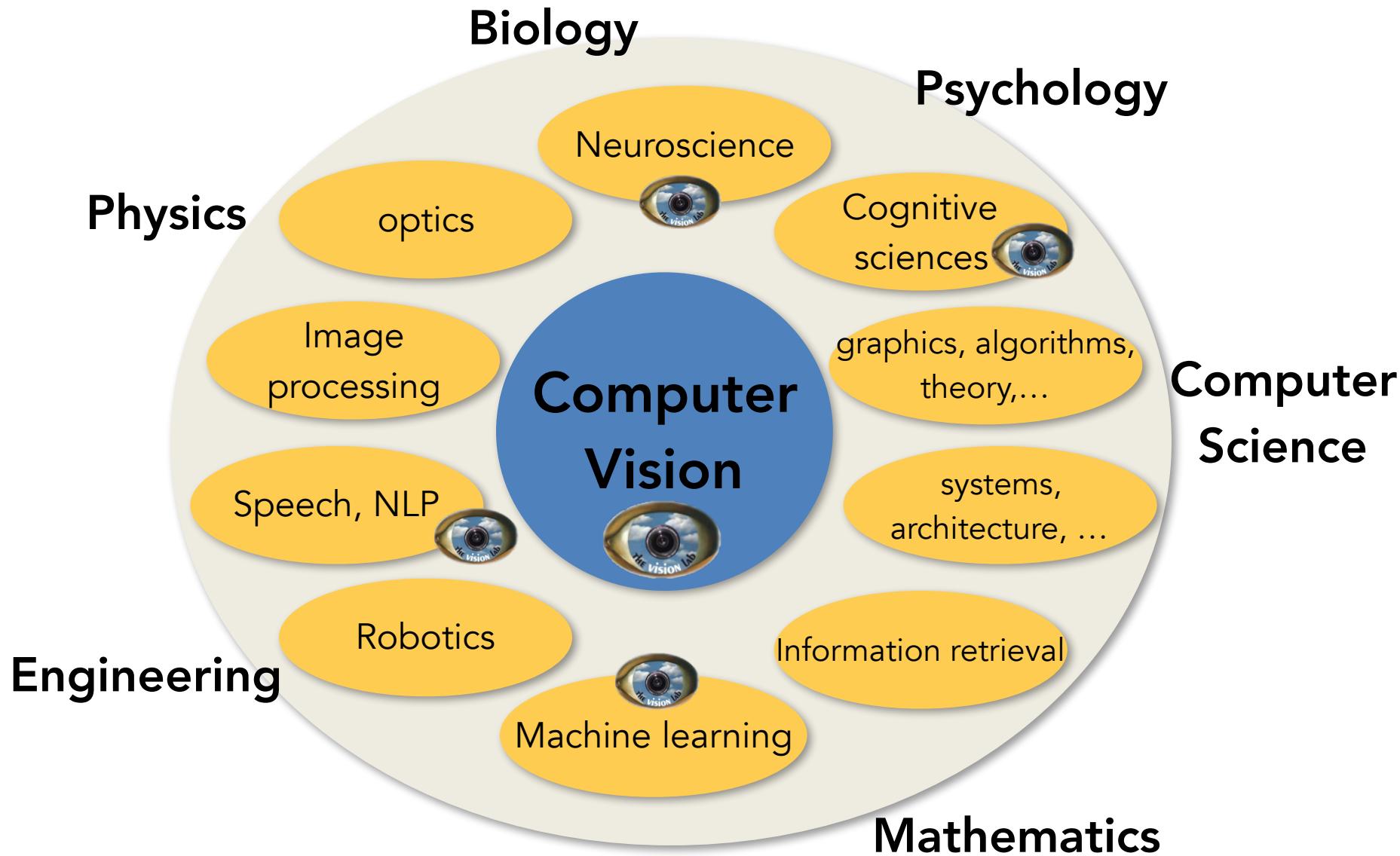
[Image is CC0 1.0 public domain](#)

[Image by Derek Keats](#) is licensed under [CC BY 2.0](#); changes made

[Image is public domain](#)

[Image by Ted Eytan](#) is licensed under [CC BY-SA 2.0](#); changes made





Related Courses @ Stanford

- CS131 (Fall 2016, Profs. Fei-Fei Li & Juan Carlos Niebles):
 - Undergraduate introductory class
- CS 224n (Winter 2017, Prof. Chris Manning and Richard Socher)
- CS231a (Spring 2017, Prof. Silvio Savarese)
 - Core computer vision class for seniors, masters, and PhDs
 - Topics include image processing, cameras, 3D reconstruction, segmentation, object recognition, scene understanding
- **CS231n (this term, Prof. Fei-Fei Li & Justin Johnson & Serena Yeung)**
 - **Neural network (aka “deep learning”) class on image classification**
- And an assortment of CS331 and CS431 for advanced topics in computer vision

Today's agenda

- A brief history of computer vision
- CS231n overview

Evolution's Big Bang



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



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

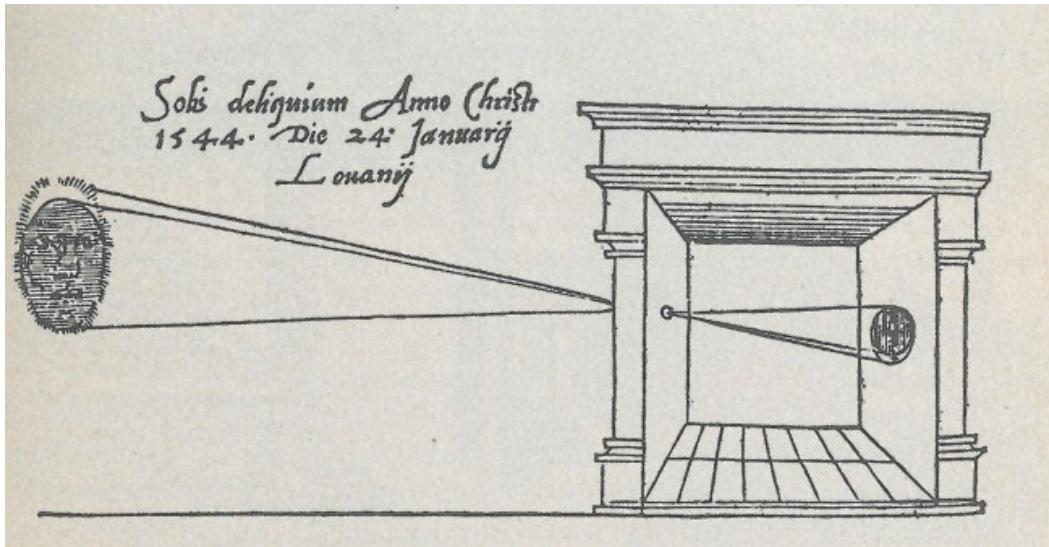


[This image is licensed under CC-BY 3.0](#)

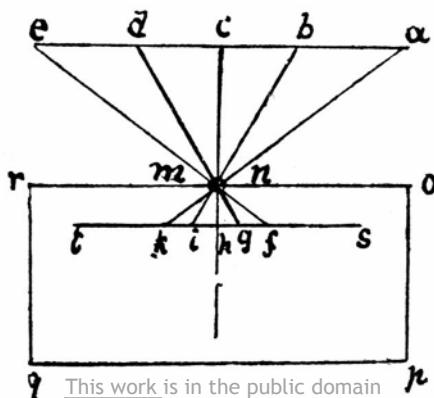
543million years, B.C.

Camera Obscura

Gemma Frisius, 1545



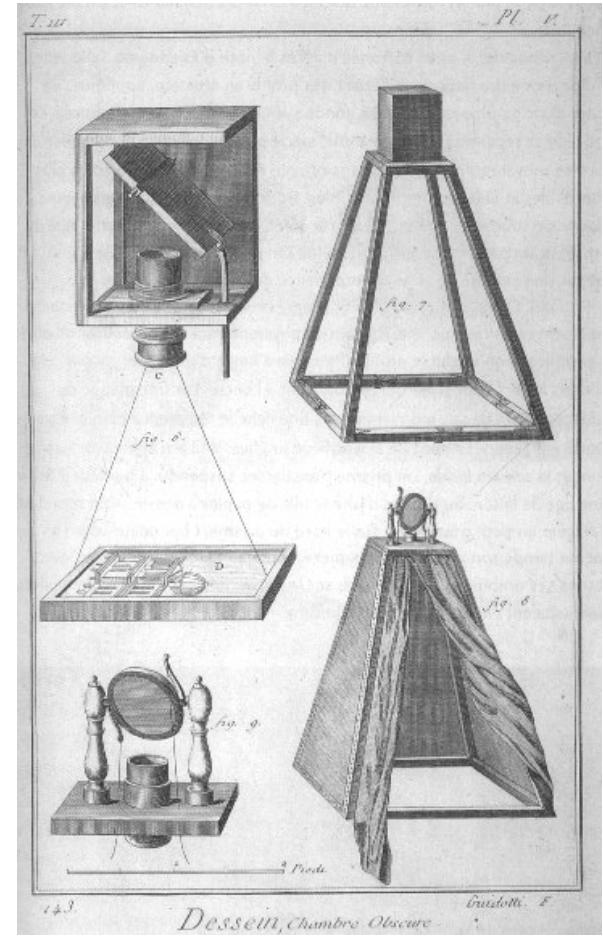
[This work is in the public domain](#)



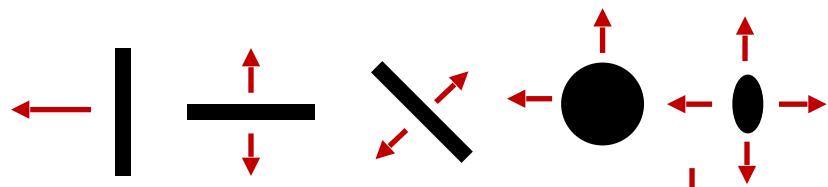
o Leonardo da Vinci,
16th Century AD

[This work is in the public domain](#)

Encyclopedie, 18th Century



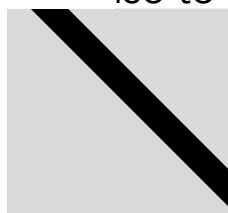
[This work is in the public domain](#)



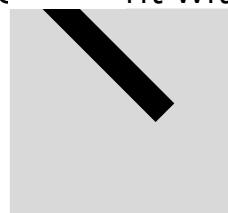
Simple cells:
Response to light orientation

Complex cells:
Response to light orientation and movement

Hypercomplex cells:
Response to movement with end point

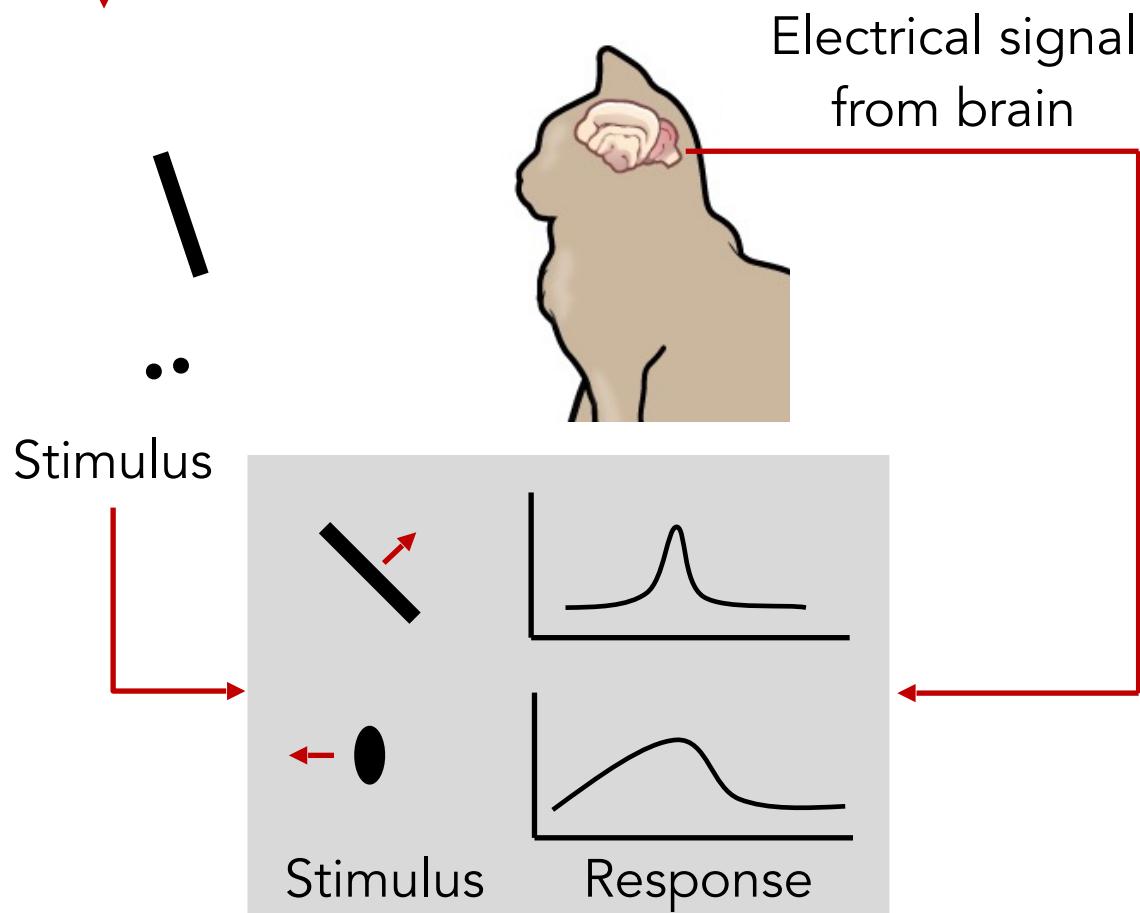


No response



Response
(end point)

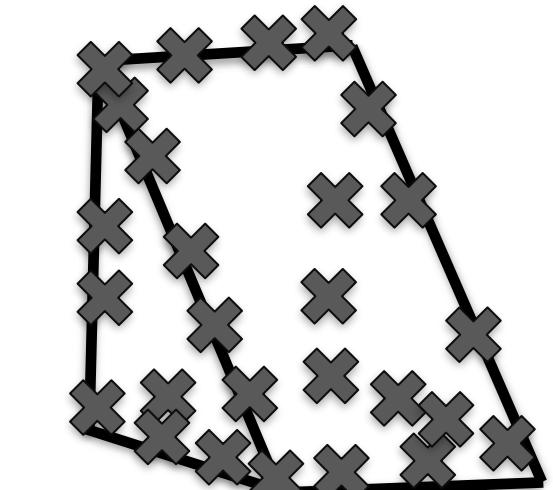
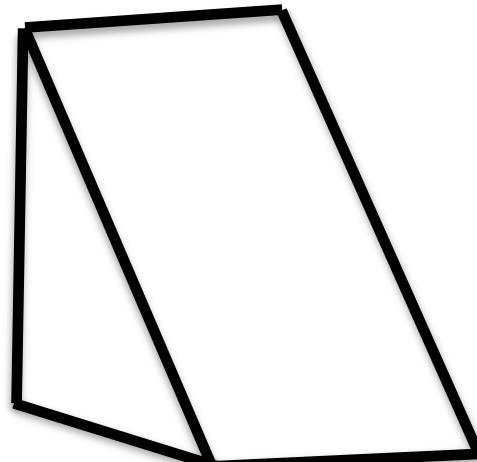
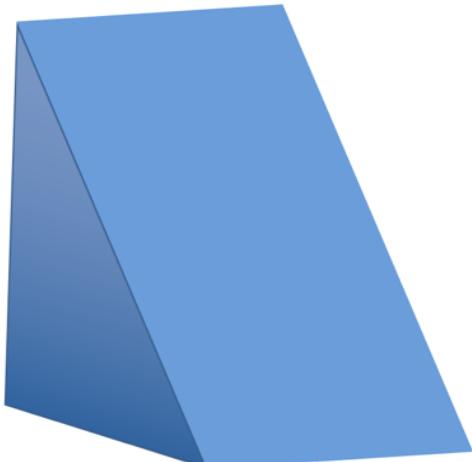
Hubel & Wiesel, 1959



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

Block world

Larry Roberts, 1963



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

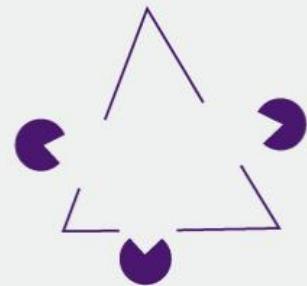
THE SUMMER VISION PROJECT

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".

Copyrighted Material

VISION



David Marr

FOREWORD BY
Shimon Ullman

AFTERWORD BY
Tomaso Poggio

Copyrighted Material

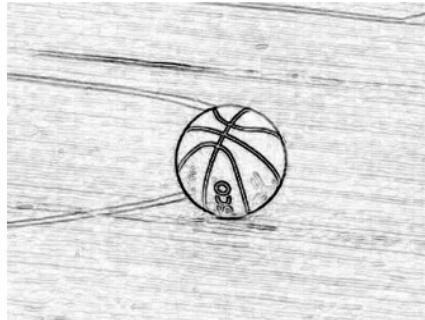
David Marr, 1970s

Input image

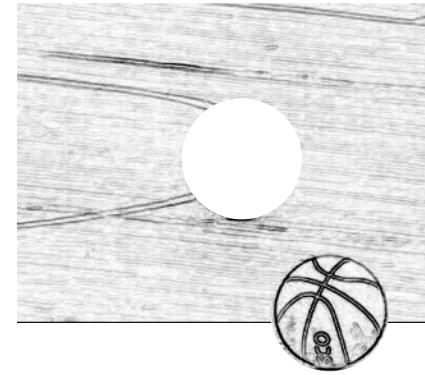


[This image is CC0 1.0 public domain](#)

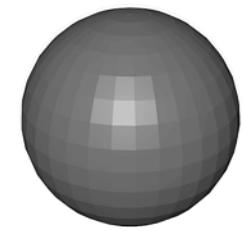
Edge image



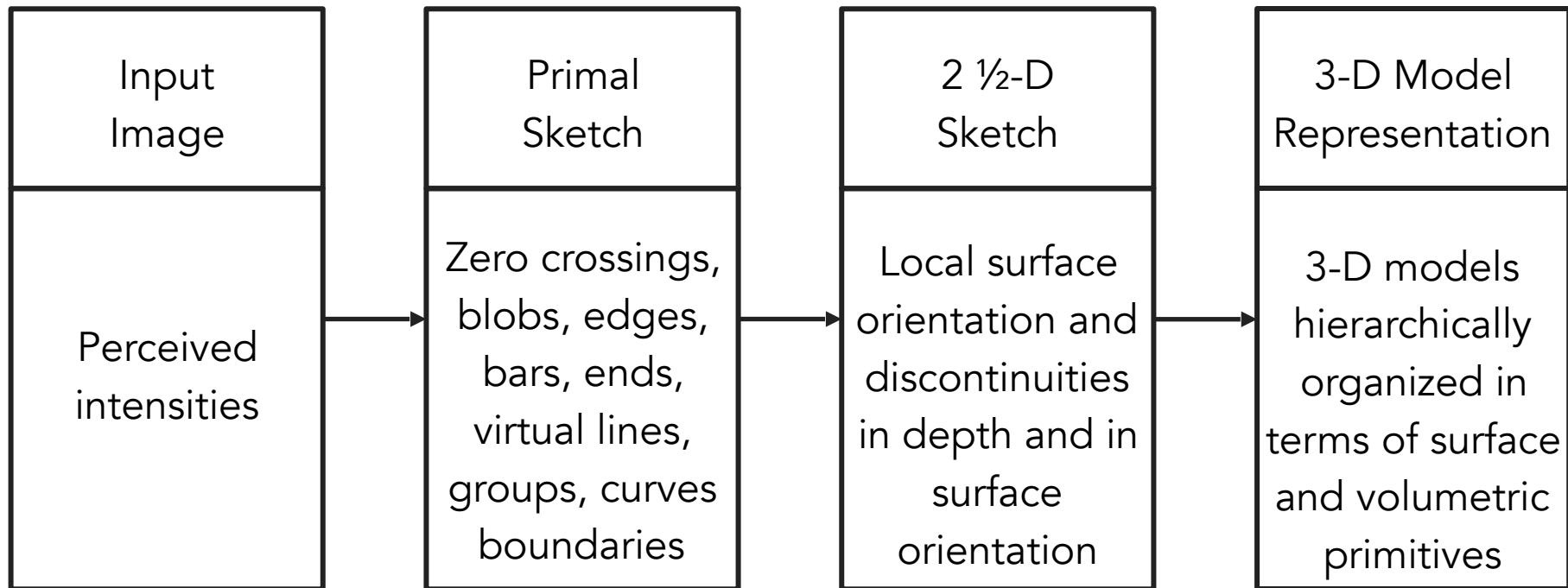
2 ½-D sketch



3-D model



[This image is CC0 1.0 public domain](#)



Stages of Visual Representation, David Marr, 1970s

- Generalized Cylinder Brooks & Binford, 1979
- Pictorial Structure Fischler and Elschlager, 1973

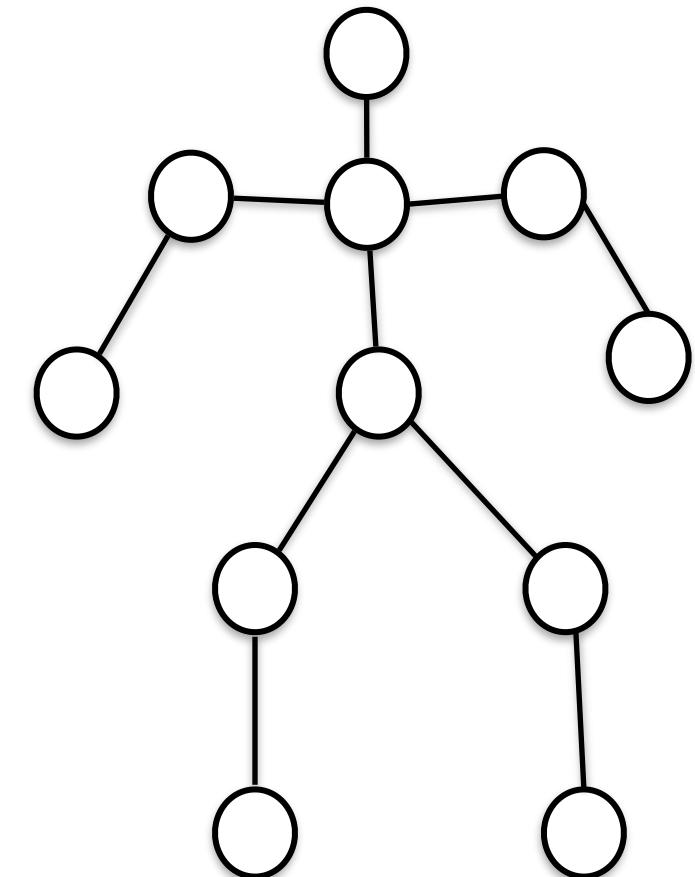
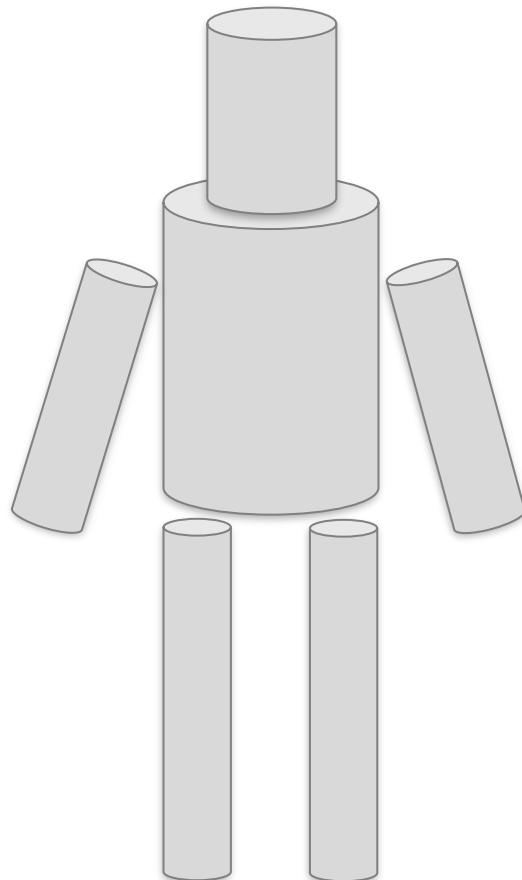




Image is CC BY-SA 4.0

David Lowe, 1987

Normalized Cut (Shi & Malik, 1997)

[Image](#) is CC BY 3.0



[Image](#) is public domain



[Image](#) is CC-BY SA 3.0



Face Detection, Viola & Jones,
2001



Image is public
domain



[Image](#) is public domain

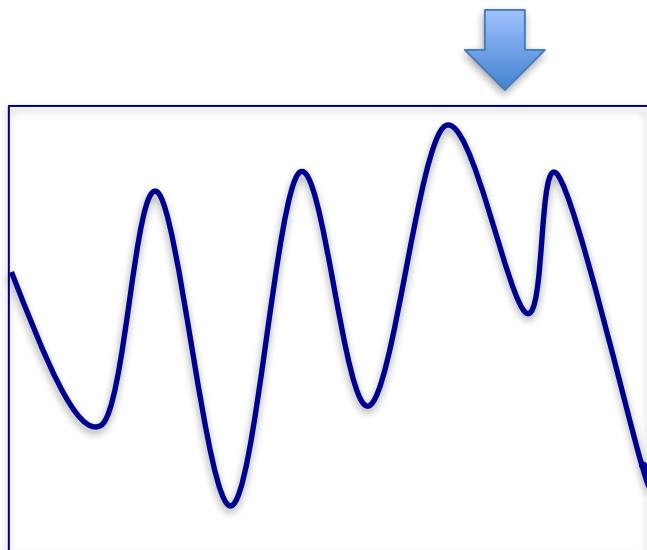


[Image](#) is CC BY-SA 2.0

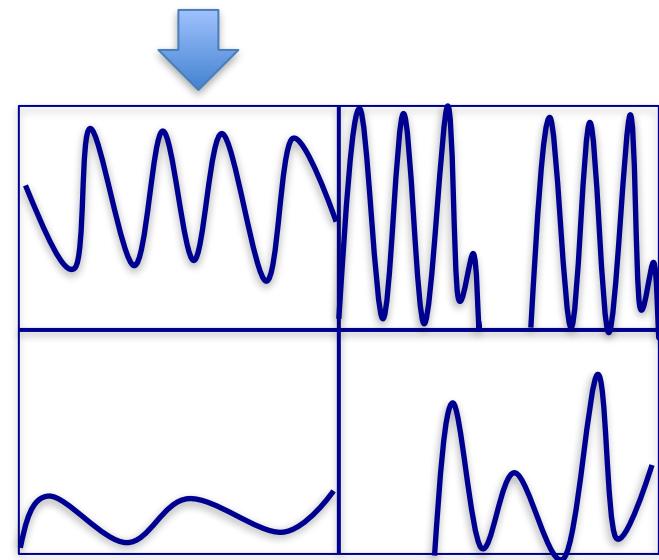
“SIFT” & Object Recognition, David Lowe, 1999



Image is CC0 1.0 public domain

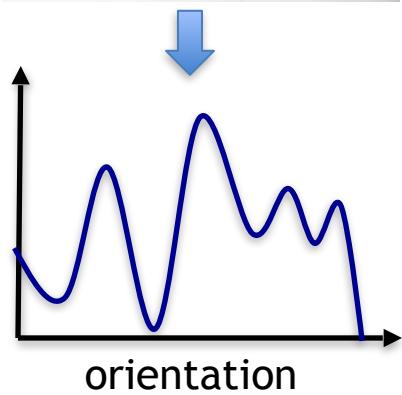


Level 0

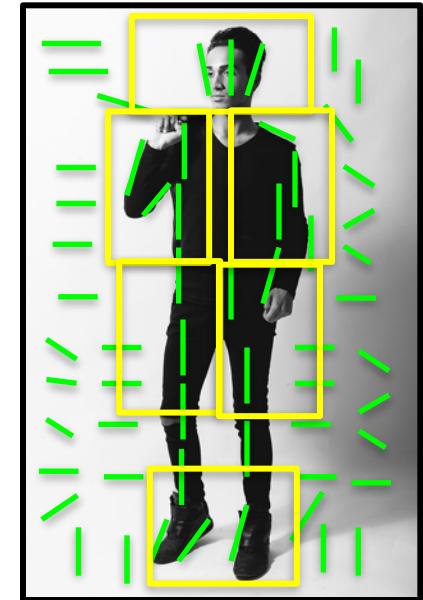


Level 1

Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006



Histogram of Gradients (HoG)
Dalal & Triggs, 2005



Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009

PASCAL Visual Object Challenge (20 object categories)

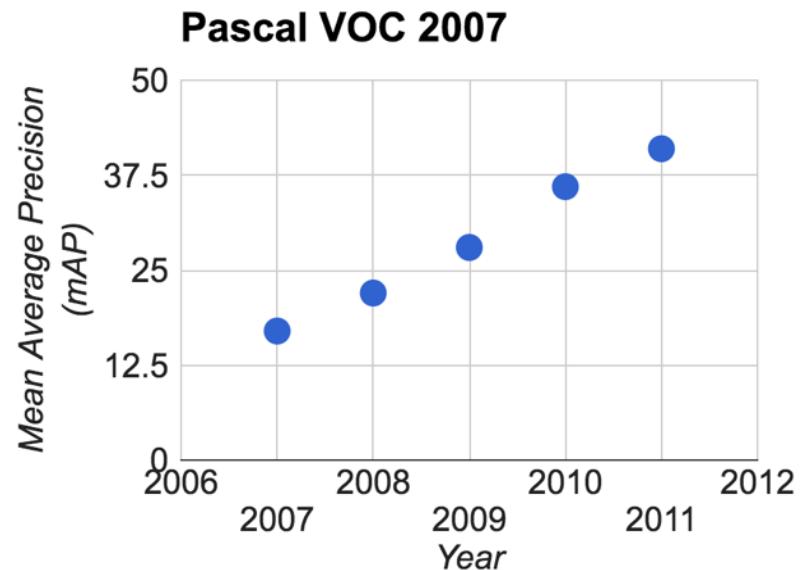
[Everingham et al. 2006-2012]

[Image](#) is CC BY-SA 3.0



[Image](#) is CC0 1.0 public domain

This image is licensed under
CC BY-SA 2.0; changes made





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

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle



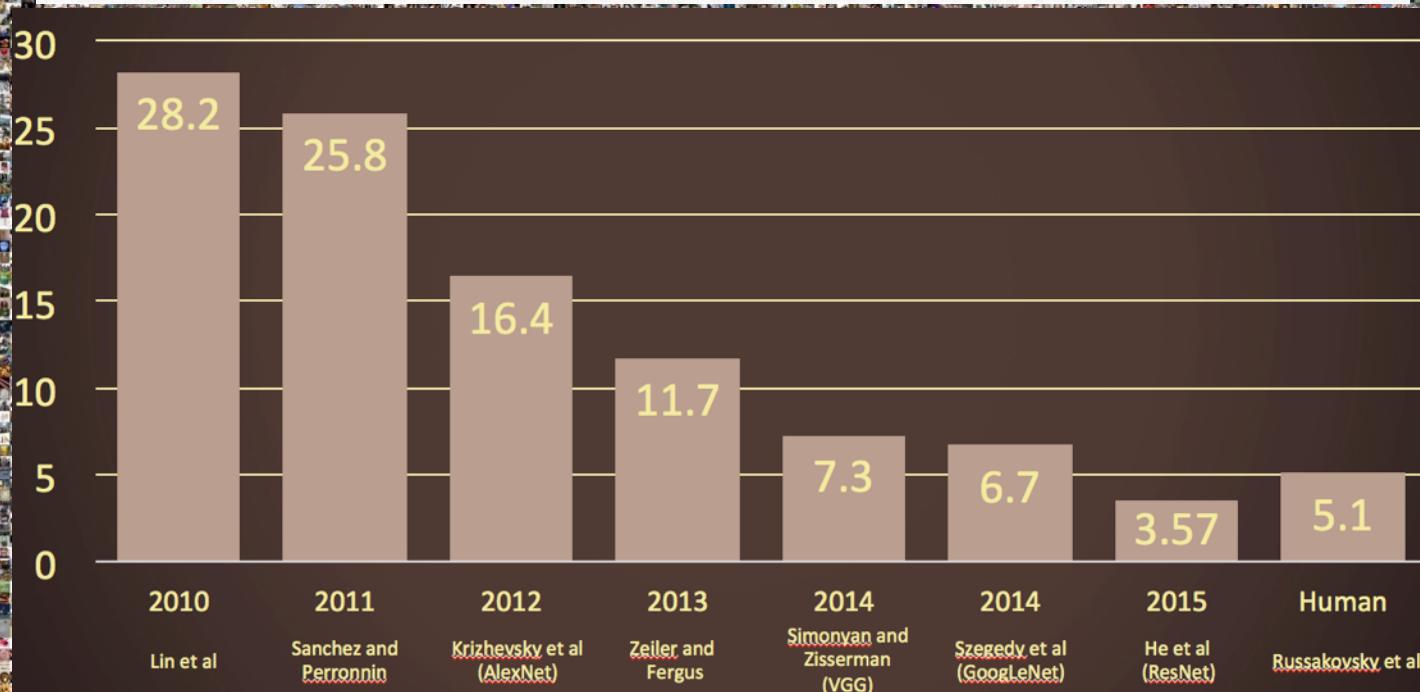
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle



Russakovsky et al. arXiv, 2014

IMAGENET Large Scale Visual Recognition Challenge

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



Russakovsky et al. arXiv, 2014

Today's agenda

- A brief history of computer vision
- CS231n overview

CS231n focuses on one of the most important
problems of visual recognition –
image classification



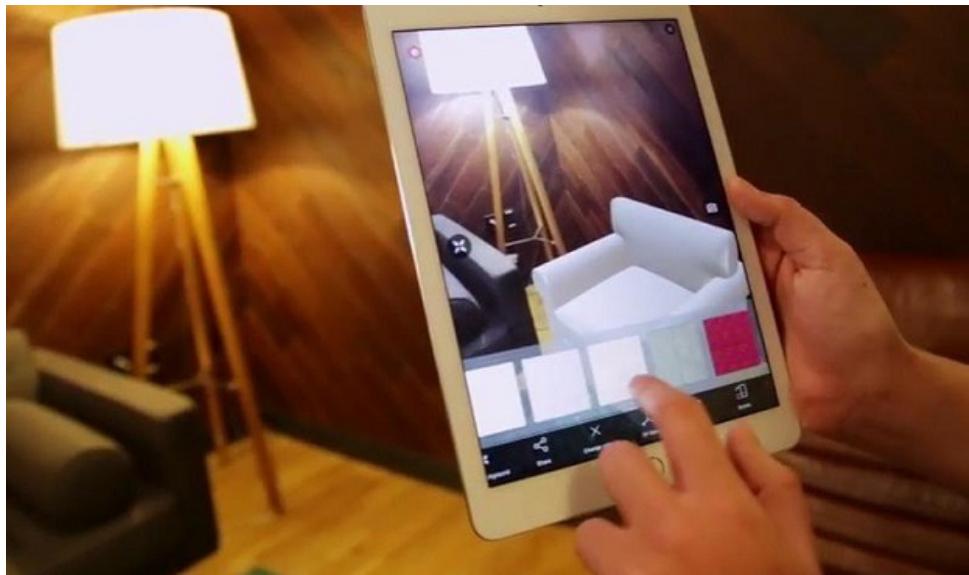
[Image by US Army](#) is licensed under CC BY 2.0



[Image](#) is CC0 1.0 public domain

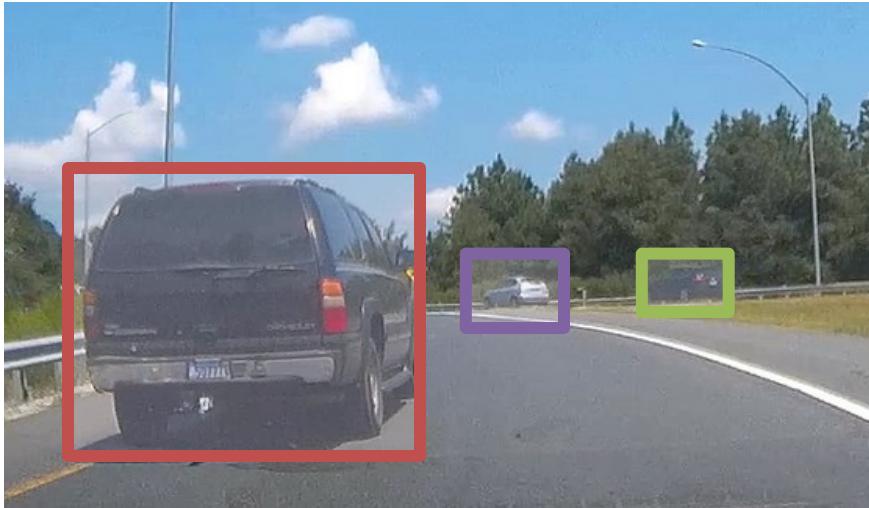


[Image by Kippelboy](#) is licensed under CC BY-SA 3.0



[Image by Christina C.](#) is licensed under CC BY-SA 4.0

There is a number of visual recognition problems that are related to image classification, such as object detection, image captioning



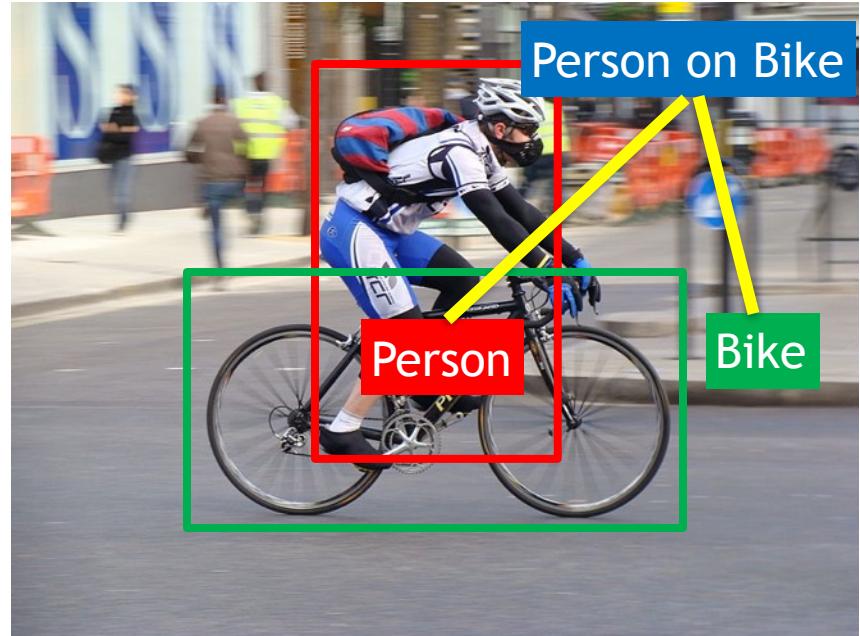
This image is licensed under CC BY-NC-SA 2.0; changes made



Person
Hammer

This image is licensed under CC BY-SA 2.0; changes made

- Object detection
- Action classification
- Image captioning
- ...



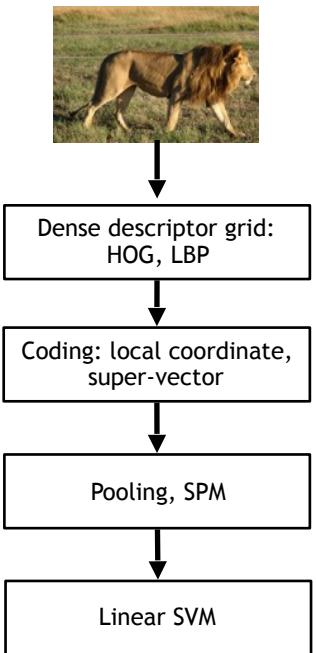
This image is licensed under CC BY-SA 3.0; changes made

Convolutional Neural Networks (CNN) have become an important tool for object recognition

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

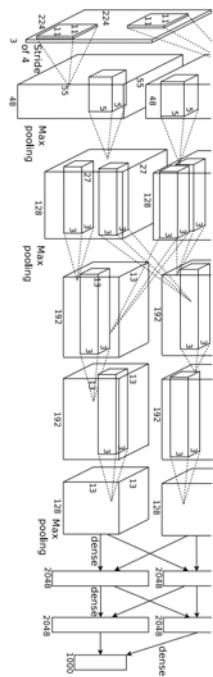


[Lin CVPR 2011]

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

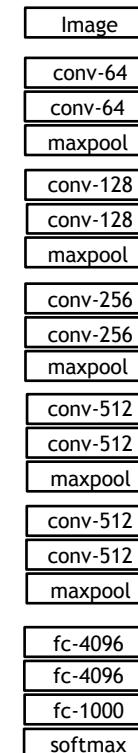
GoogLeNet

- Pooling
- Convolution
- Softmax
- Other



[Szegedy arxiv 2014]

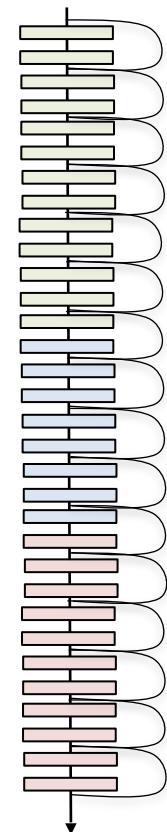
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA

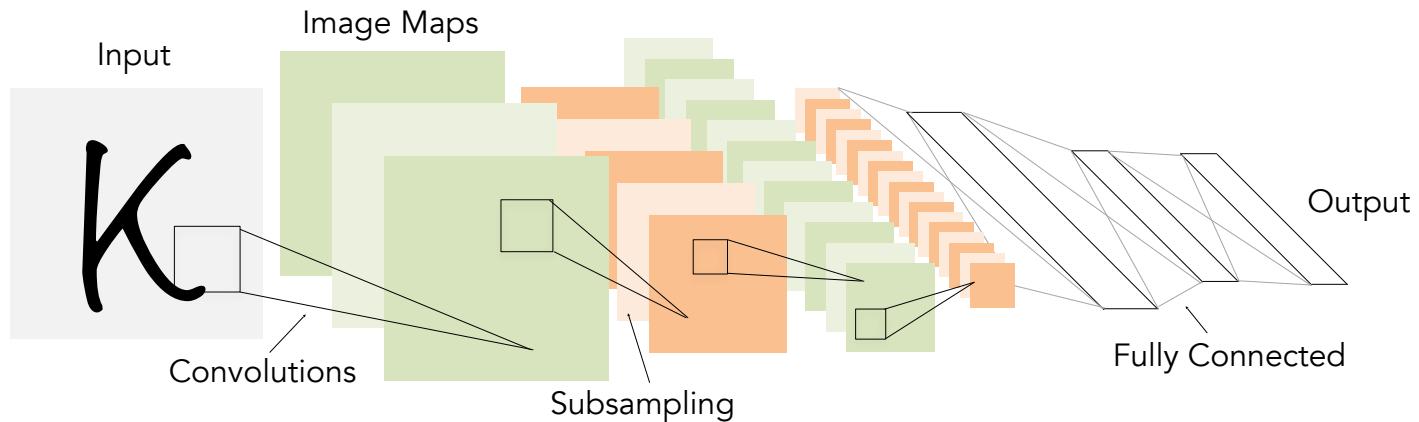


[He ICCV 2015]

Convolutional Neural Networks (CNN)
were not invented overnight

1998

LeCun et al.



of transistors



10^6

pentium® II

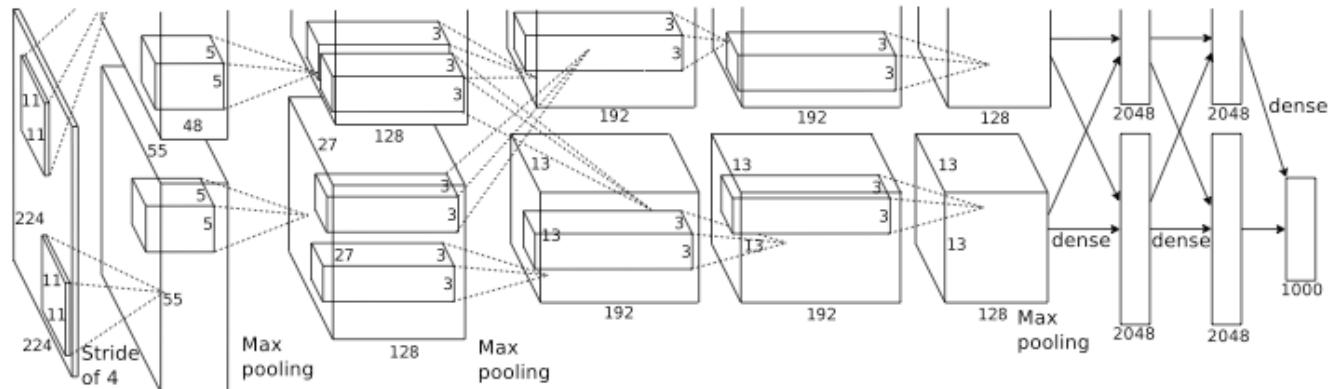
of pixels used in training

10^7



2012

Krizhevsky et al.



of transistors



10^9

GPUs



of pixels used in training

10^{14}



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

The quest for visual intelligence
goes far beyond object recognition...

Wall

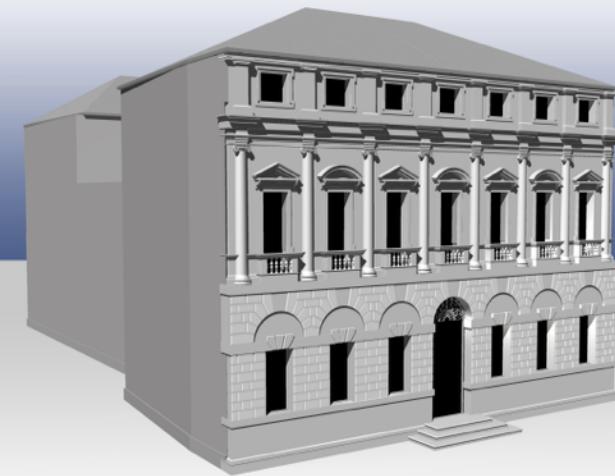
Laptop

Glass

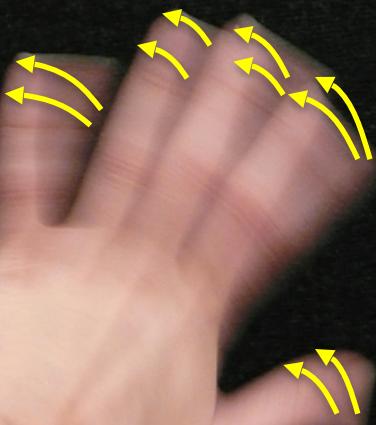
Wire

Desk

[Image](#) is GFDL



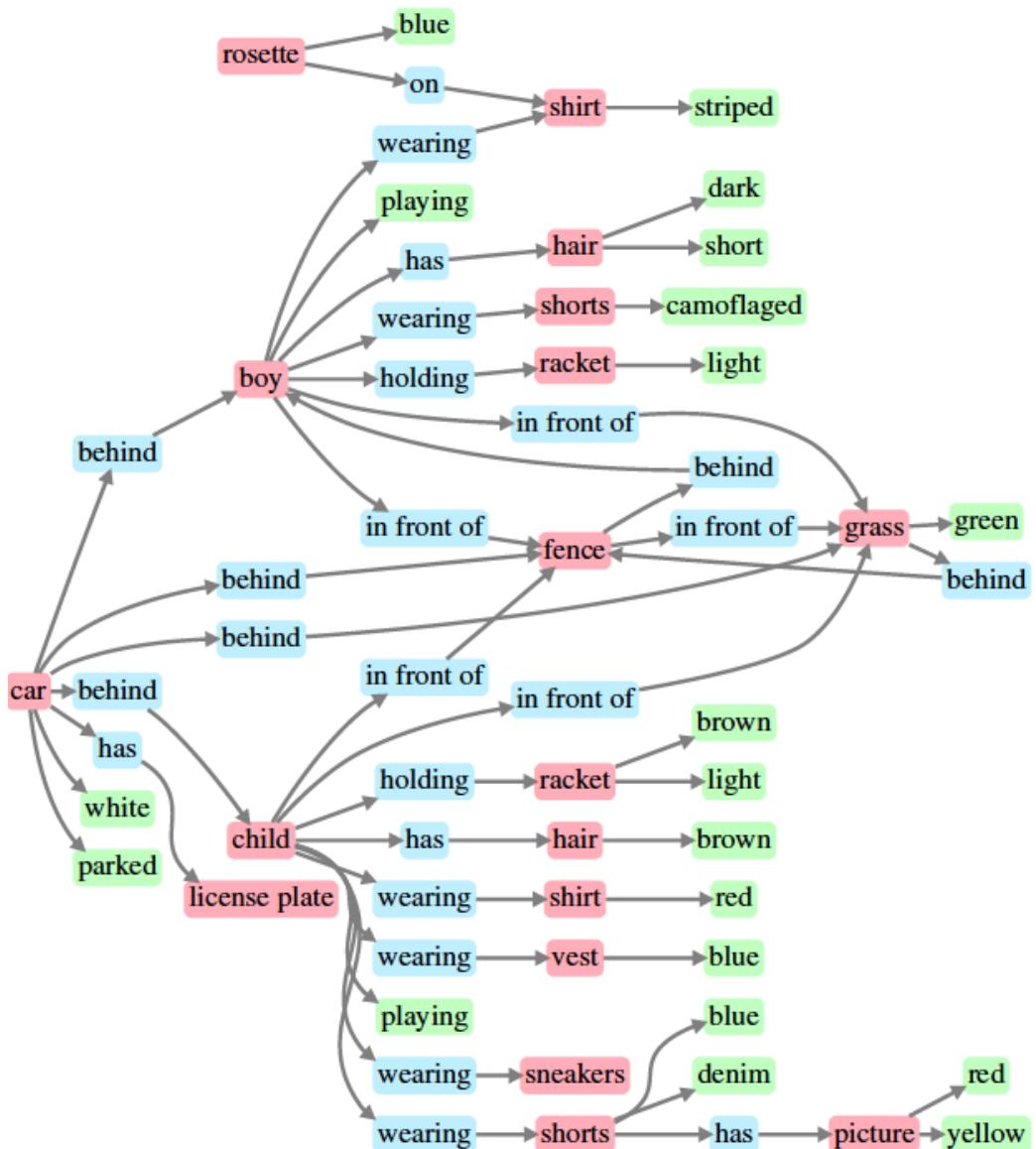
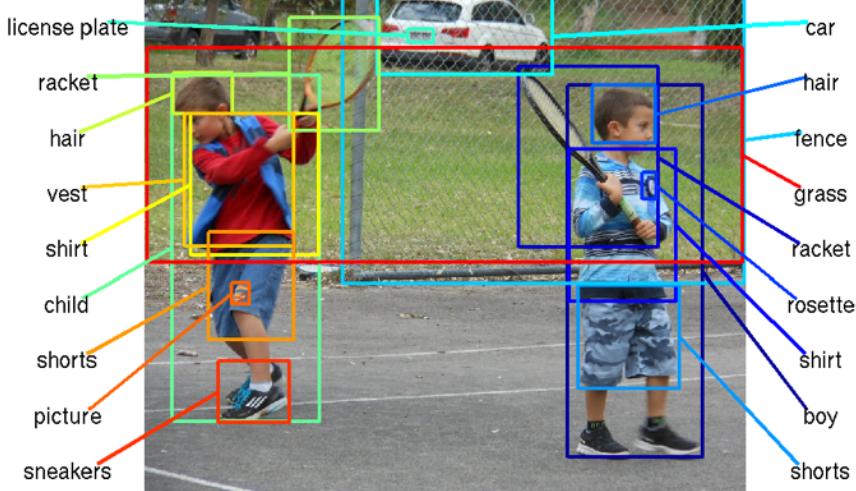
[Image](#) is CC BY-SA 2.0



Waving

[Image](#) is CC BY-SA 4.0





Johnson *et al.*, “Image Retrieval using Scene Graphs”, CVPR 2015

Figures copyright IEEE, 2015. Reproduced for educational purposes

PT = 500ms



Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM)

Fei-Fei, Iyer, Koch, Perona, JoV, 2007

Image is licensed under [CC BY-SA 3.0](#); changes made

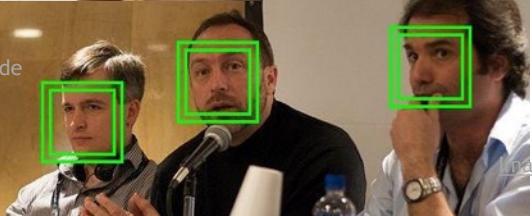


This image is copyright-free [United States government work](#)

Example credit: [Andrej Karpathy](#)



Outside border images, clockwise, starting from top left:
 Image by Pop Culture Geek is licensed under CC BY 2.0; changes made
 Image by the US Government is in the public domain
 Image by the US Government is in the public domain
 Image by Glogger is licensed under CC BY-SA 3.0; changes made
 Image by Sylenius is licensed under CC BY 3.0; changes made
 Image by US Government is in the public domain



Inside four images, clockwise, starting from top left:
 Image is CC0 1.0 public domain
 Image by Tucania is licensed under CC BY-SA 3.0; changes made
 Image by Intuitive Surgical, Inc. is licensed under CC BY-SA 3.0; changes made
 Image by Oyundari Zorigtbaatar is licensed under CC BY-SA 4.0

Who we are

Instructors



Fei-Fei Li



Justin Johnson



Serena Yeung

Emeritus



Andrej Karpathy

Teaching Assistants



De-An Huang



Russell Kaplan



Leo Keselman



Nishith
Khandwala



Shayne Longpre



Zelun Luo



Lane McIntosh



Oliver Moindrot



Amani Peddada



Emma Peng



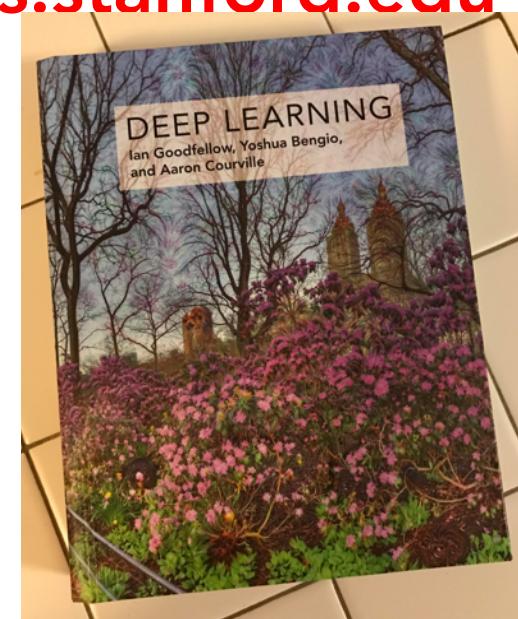
Ben Poole



Luda Zhao

Course Logistics

- Keeping in touch:
 - Piazza
 - **cs231n-spring1617-staff@lists.stanford.edu**
- Optional textbook:
 - Deep Learning by Goodfellow, Bengio, and Courville
 - Free online

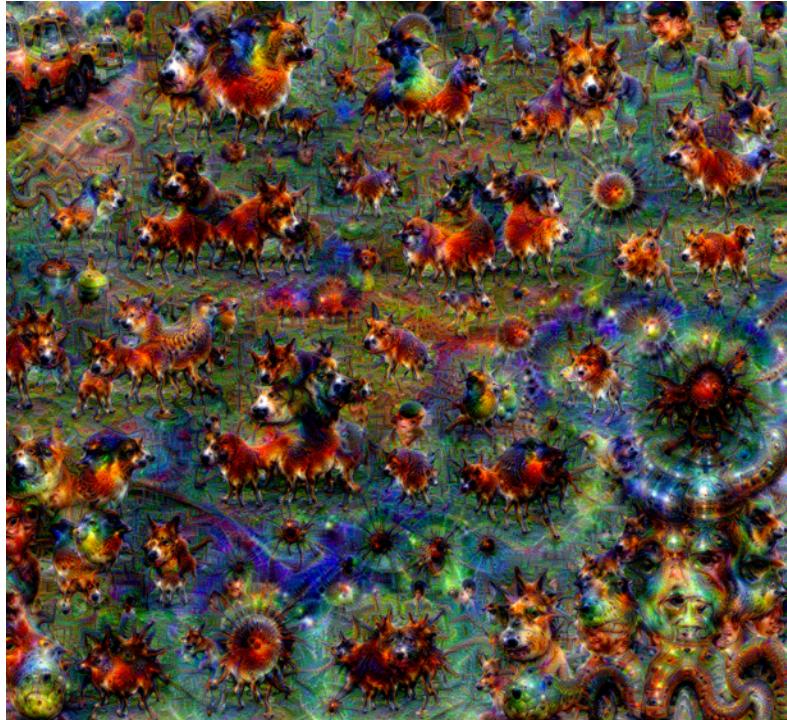


Our philosophy

- Thorough and Detailed.
 - Understand how to write from scratch, debug and train convolutional neural networks.
- Practical.
 - Focus on practical techniques for training these networks at scale, and on GPUs (e.g. will touch on distributed optimization, differences between CPU vs. GPU, etc.) Also look at state of the art software tools such as Caffe, TensorFlow, and (Py)Torch
- State of the art.
 - Most materials are new from research world in the past 1-3 years. Very exciting stuff!
- Fun.
 - Some fun topics such as Image Captioning (using RNN)
 - Also DeepDream, NeuralStyle, etc.

Our philosophy (cont'd)

- Fun.
 - Some fun topics such as Image Captioning (using RNN)
 - Also DeepDream, NeuralStyle, etc.



Grading policy

- 3 Problem Sets: 15% \times 3 = 45%
- Midterm Exam: 15%
- Final Course Project: 40%
 - Milestone: 5%
 - Final write-up: 35%
 - Bonus points for exceptional poster presentation
- Late policy
 - 7 free late days – use them in your ways
 - Afterwards, 25% off per day late
 - Not accepted after 3 late days per PS
 - Does not apply to Final Course Project
- Collaboration policy
 - Read the student code book, understand what is 'collaboration' and what is 'academic infraction'

Pre-requisite

- Proficiency in Python, some high-level familiarity with C/C++
 - All class assignments will be in Python (and use numpy), but some of the deep learning libraries we may look at later in the class are written in C++.
 - A Python tutorial available on course website
- College Calculus, Linear Algebra
- Equivalent knowledge of CS229 (Machine Learning)
 - We will be formulating cost functions, taking derivatives and performing optimization with gradient descent.

Syllabus

- Go to website...

<http://cs231n.stanford.edu/>

References

- Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *The Journal of physiology* 160.1 (1962): 106. [\[PDF\]](#)
- Roberts, Lawrence Gilman. "Machine Perception of Three-dimensional Solids." Diss. Massachusetts Institute of Technology, 1963. [\[PDF\]](#)
- Marr, David. "Vision." The MIT Press, 1982. [\[PDF\]](#)
- Brooks, Rodney A., and Creiner, Russell and Binford, Thomas O. "The ACRONYM model-based vision system. " In *Proceedings of the 6th International Joint Conference on Artificial Intelligence* (1979): 105-113. [\[PDF\]](#)
- Fischler, Martin A., and Robert A. Elschlager. "The representation and matching of pictorial structures." *IEEE Transactions on Computers* 22.1 (1973): 67-92. [\[PDF\]](#)
- Lowe, David G., "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, 31, 3 (1987), pp. 355-395. [\[PDF\]](#)
- Shi, Jianbo, and Jitendra Malik. "Normalized cuts and image segmentation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000): 888-905. [\[PDF\]](#)
- Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.* Vol. 1. IEEE, 2001. [\[PDF\]](#)
- Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International Journal of Computer Vision* 60.2 (2004): 91-110. [\[PDF\]](#)
- Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Vol. 2. IEEE, 2006. [\[PDF\]](#)

- Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005. [\[PDF\]](#)
- Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008 [\[PDF\]](#)
- Everingham, Mark, et al. "The pascal visual object classes (VOC) challenge." International Journal of Computer Vision 88.2 (2010): 303-338. [\[PDF\]](#)
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. [\[PDF\]](#)
- Russakovsky, Olga, et al. "Imagenet Large Scale Visual Recognition Challenge." arXiv:1409.0575. [\[PDF\]](#)
- Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and SVM training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. [\[PDF\]](#)
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. [\[PDF\]](#)
- Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014). [\[PDF\]](#)
- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [\[PDF\]](#)
- He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." arXiv preprint arXiv:1406.4729 (2014). [\[PDF\]](#)
- LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. [\[PDF\]](#)
- Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." Journal of vision 7.1 (2007): 10. [\[PDF\]](#)