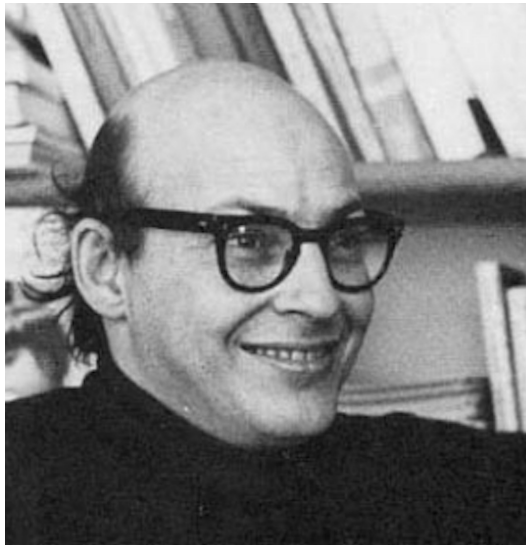
Announcements

- The final exam is in-class on Friday. It will be mainly multiple choice questions similar to the quizzes, and it will cover all of the assigned readings.
- I'm holding an extra set of office hours tomorrow from 1:30-3:30 in 3401 Walnut room 452C.
- HW5 is due on Tuesday, late days are OK to use.

- Today's lecture on Computer Vision is just for fun – it's not on the exam.

1966

"Connect a television camera to a computer and get the machine to describe what it sees."



Marvin Minsky
Turing award, 1969

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

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

THE SUMMER VISION PROJECT

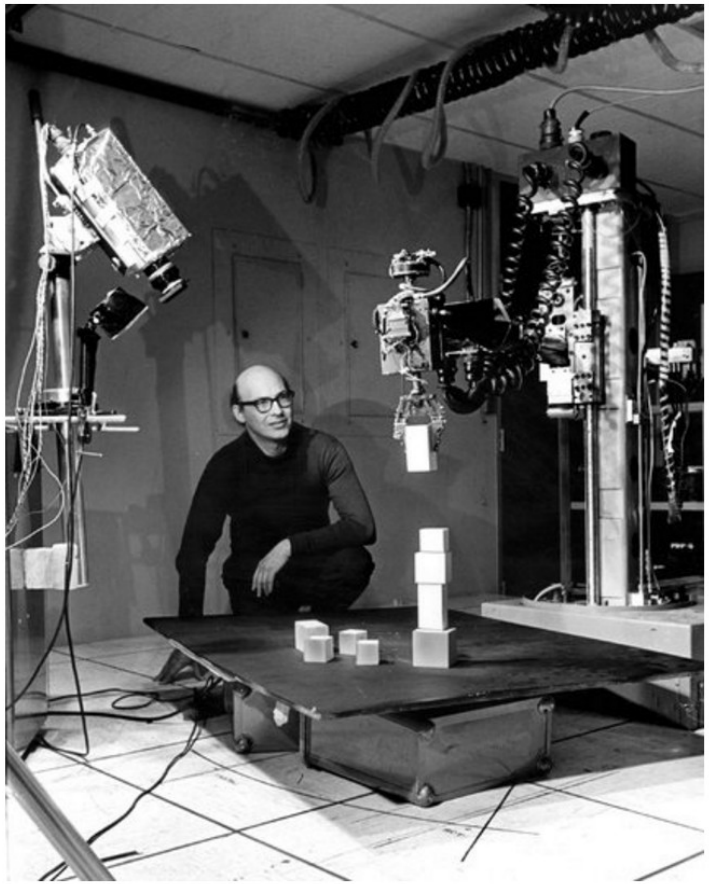
Seymour Papert

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".

The New York Times

Marvin Minsky, Pioneer in Artificial Intelligence, Dies at 88

By GLENN RIFKIN JAN. 25, 2016



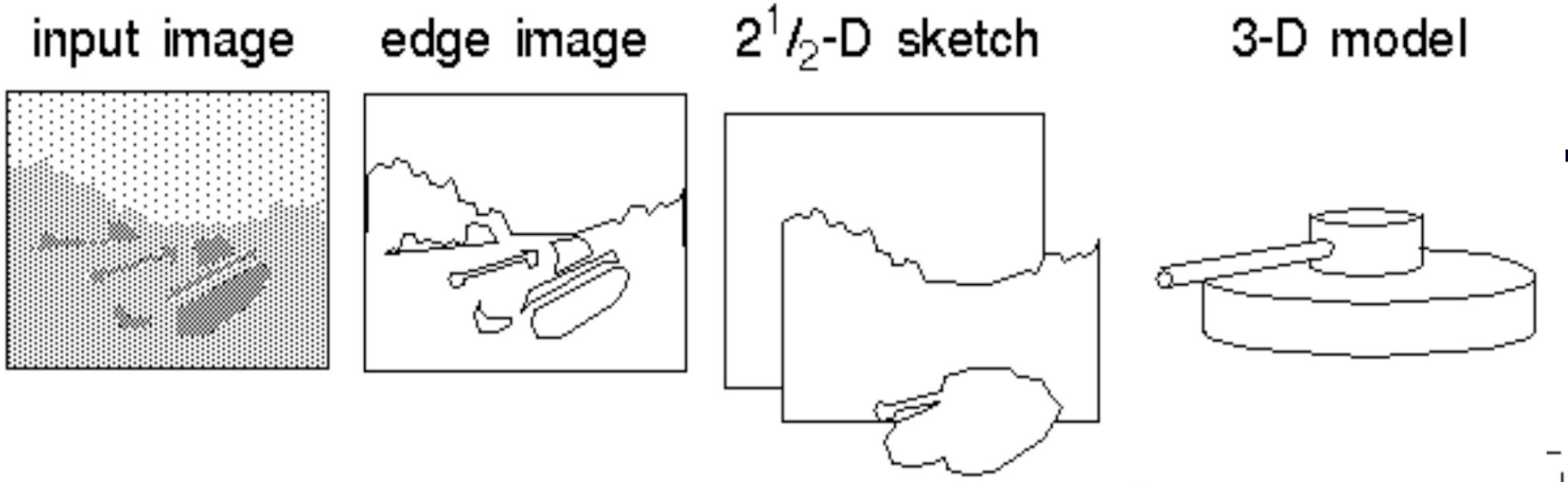
Marvin Minsky in a lab at M.I.T. in 1968. M.I.T.

Marvin Minsky, who combined a scientist's thirst for knowledge with a philosopher's quest for truth as a pioneering explorer of artificial intelligence, work that helped inspire the creation of the personal computer and the Internet, died on Sunday night in Boston. He was 88.

His family said the cause was a cerebral hemorrhage.

Well before the advent of the microprocessor and the supercomputer, Professor Minsky, a revered computer science educator at M.I.T., laid the foundation for the field of artificial intelligence by demonstrating the possibilities of imparting common-sense reasoning to computers.

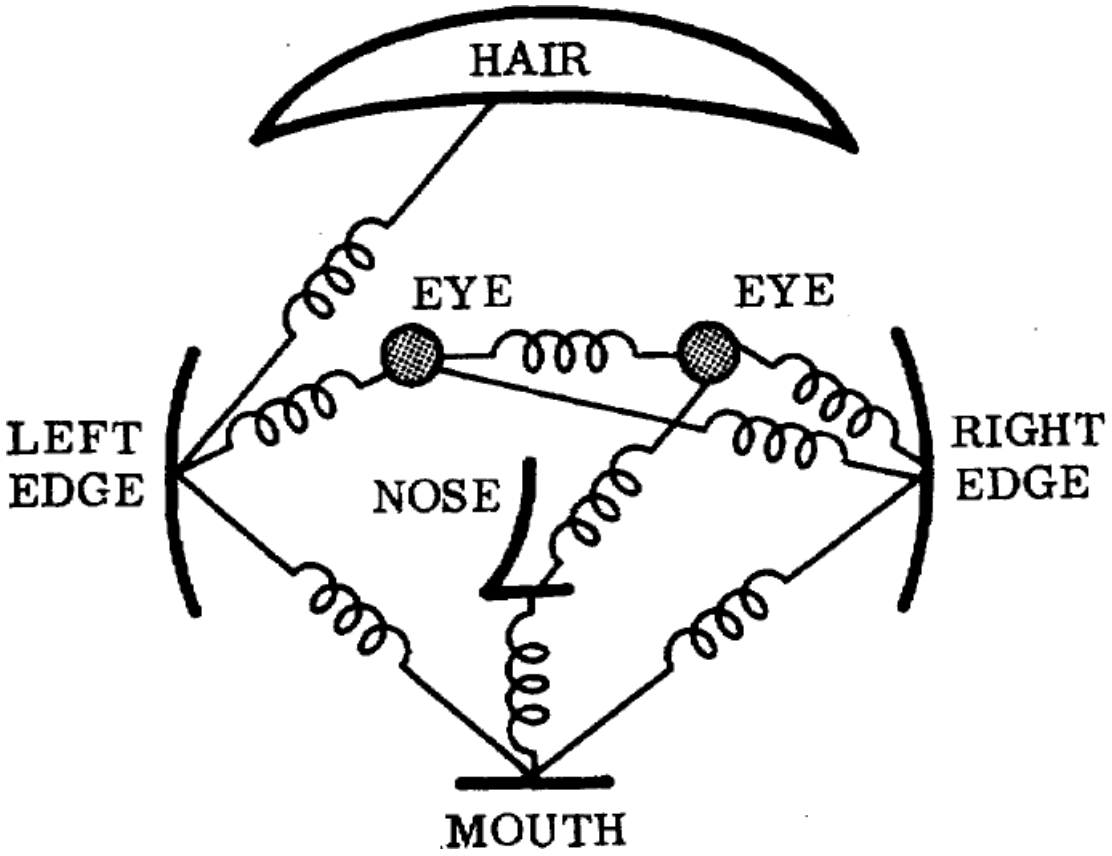
"Marvin was one of the very few people in computing whose visions and



Input Image	Primal Sketch	2 1/2-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 organised in terms of surface and volumetric primitives

Stages of Visual Representation, David Marr, 1970

1973

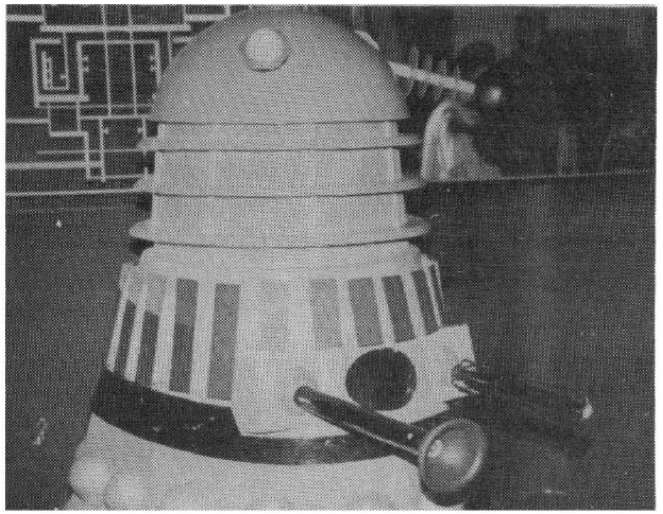


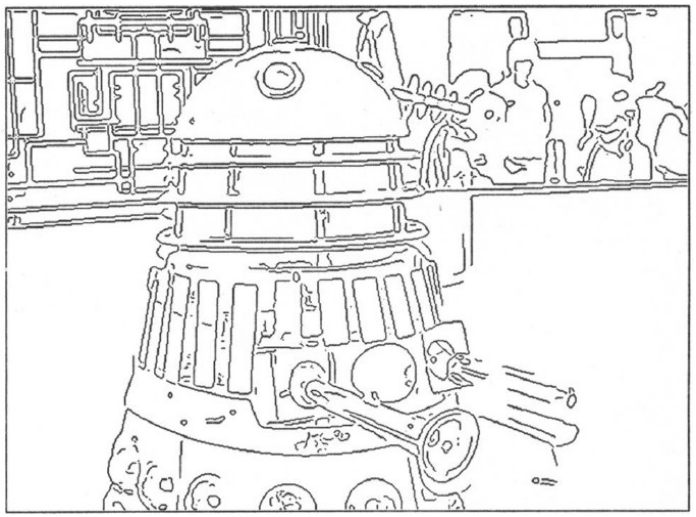
The constellation model

The representation and matching of pictorial structures,
Fischler and Elschlager, 1973

1980's

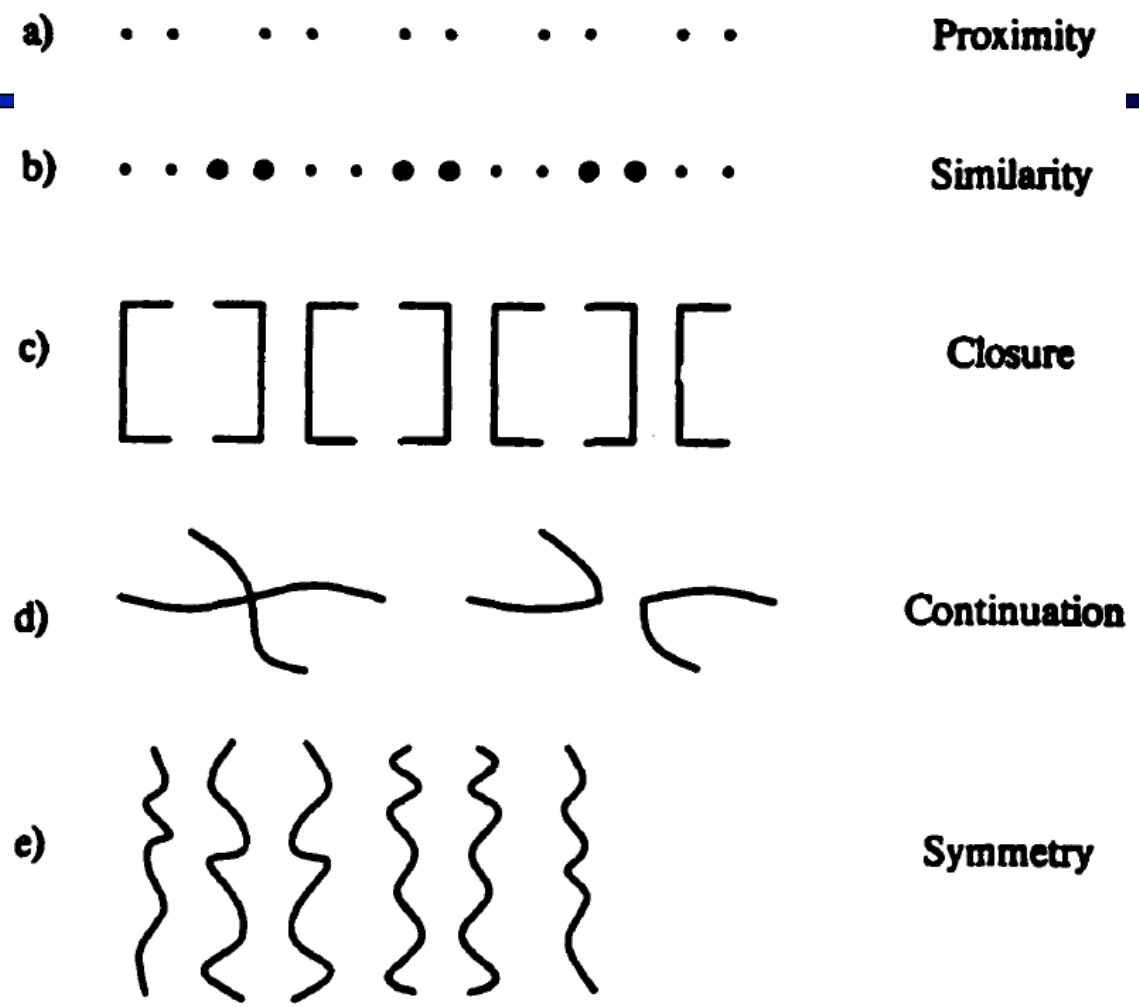
AI winter… …back to basics





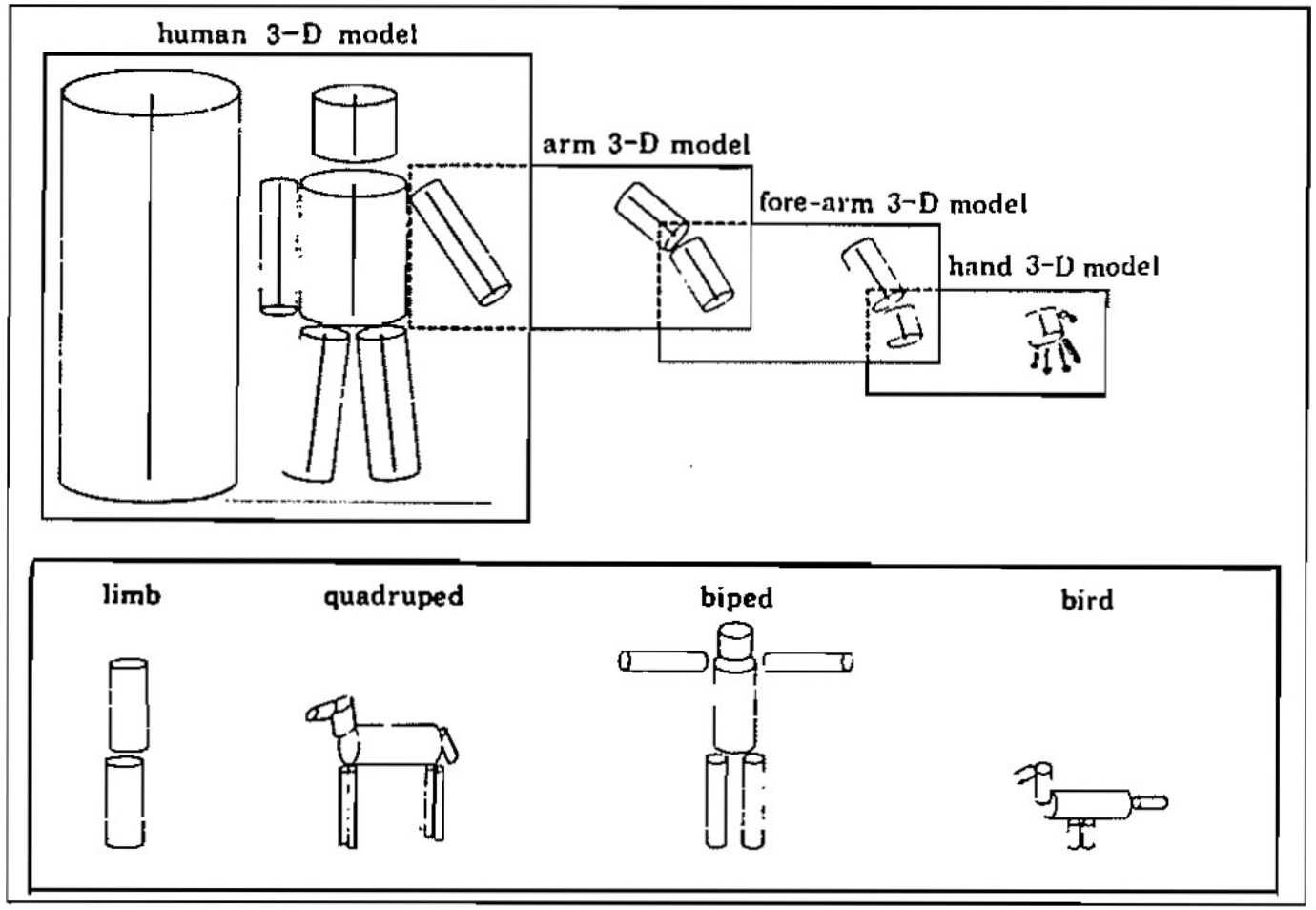
A Computational Approach to Edge Detection, Canny 1986

1984



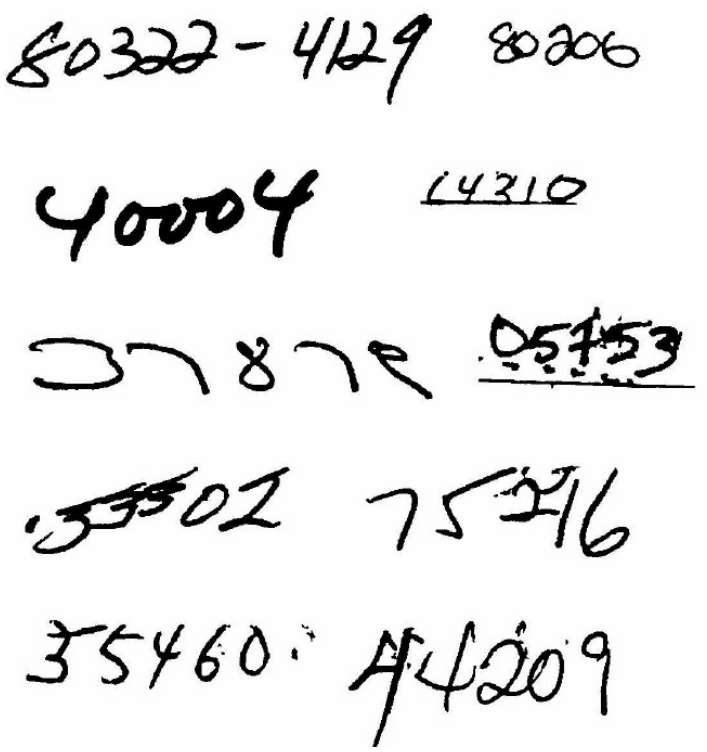
Perceptual Organization and Visual Recognition,
David Lowe, 1984

1986



Perceptual organization and the representation of natural form,
Alex Pentland, 1986

1989



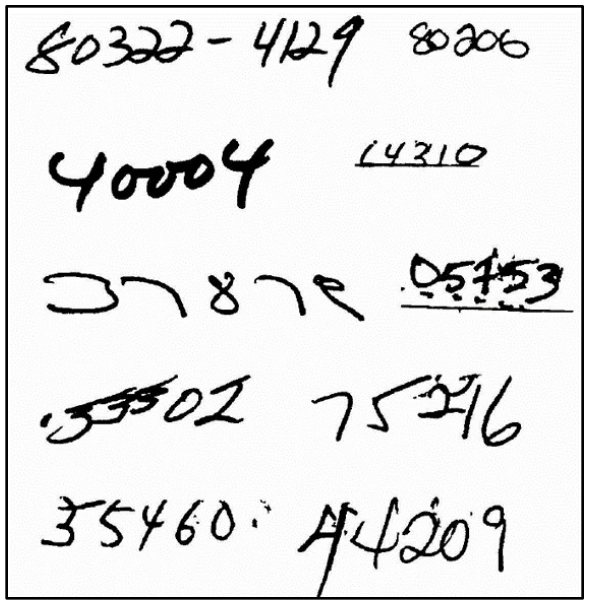
Zip codes

MNIST

Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989

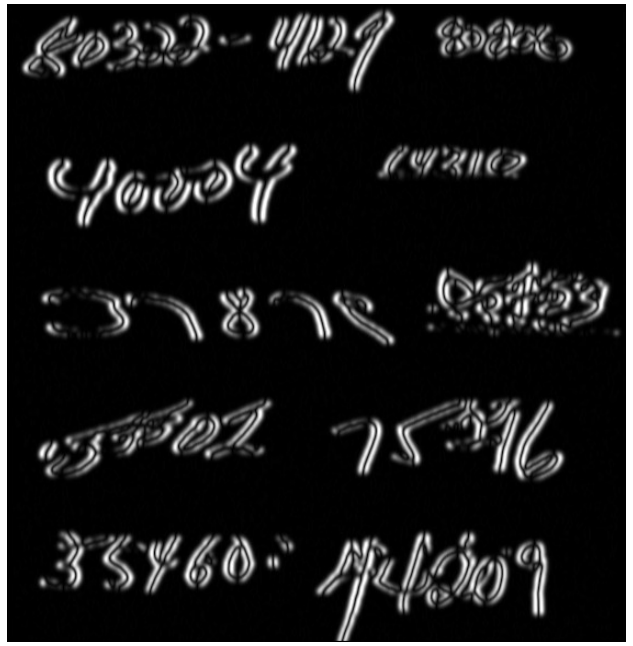
Filters

Input



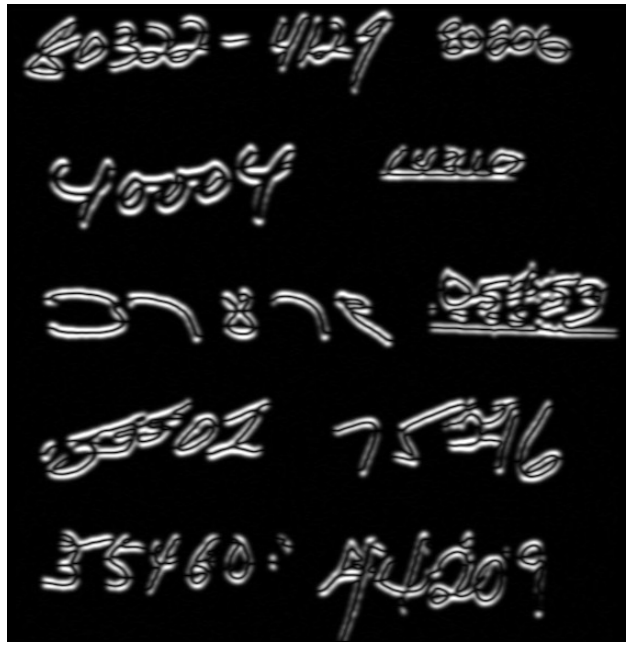
-1	0	+1
-2	0	+2
-1	0	+1

x filter

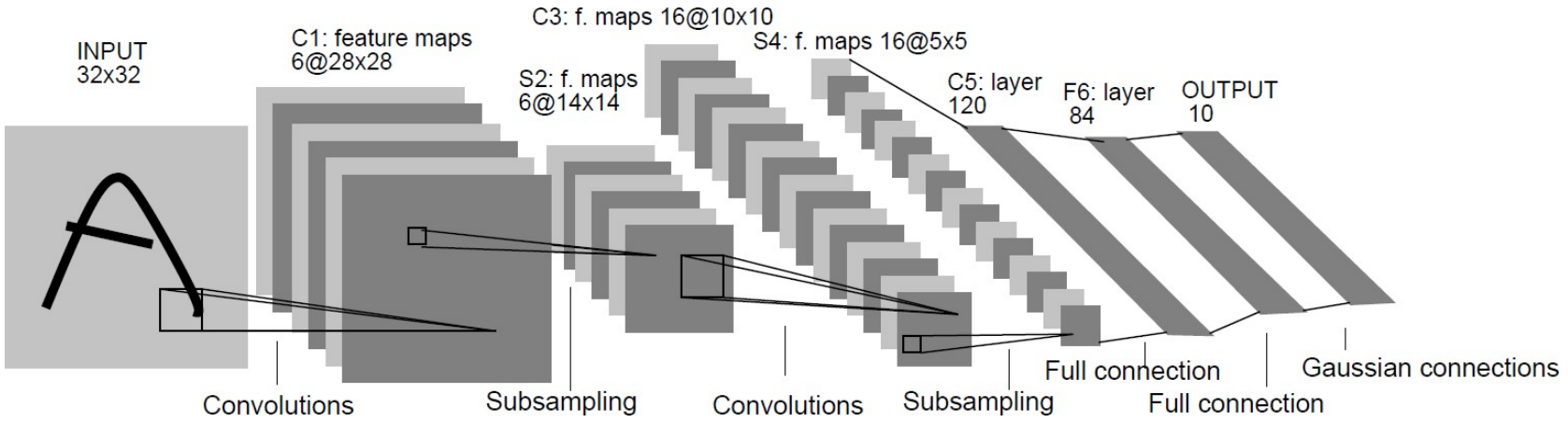


+1	+2	+1
0	0	0
-1	-2	-1

y filter

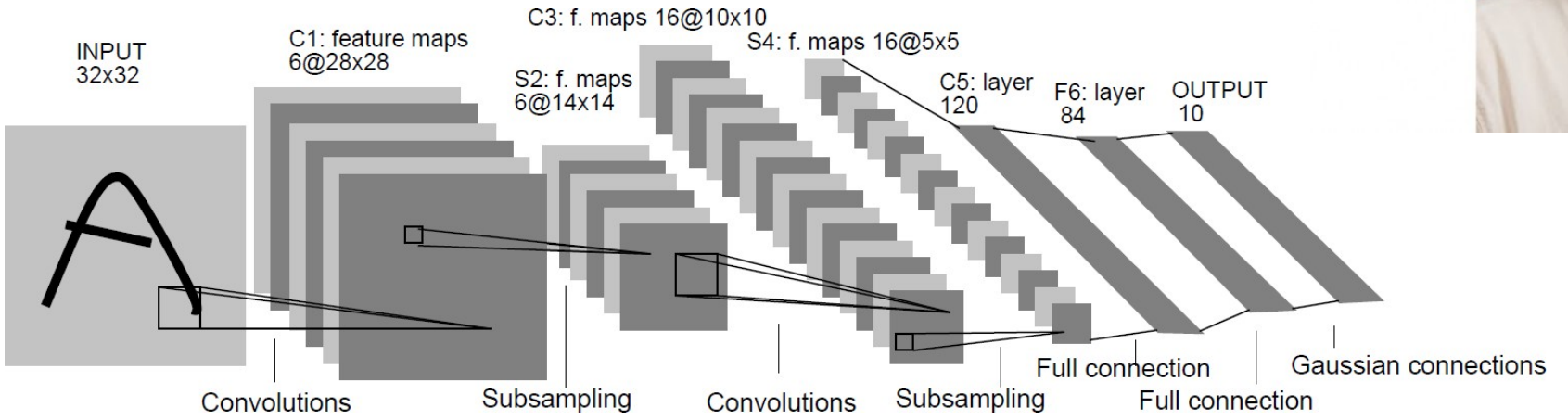


1989



Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989

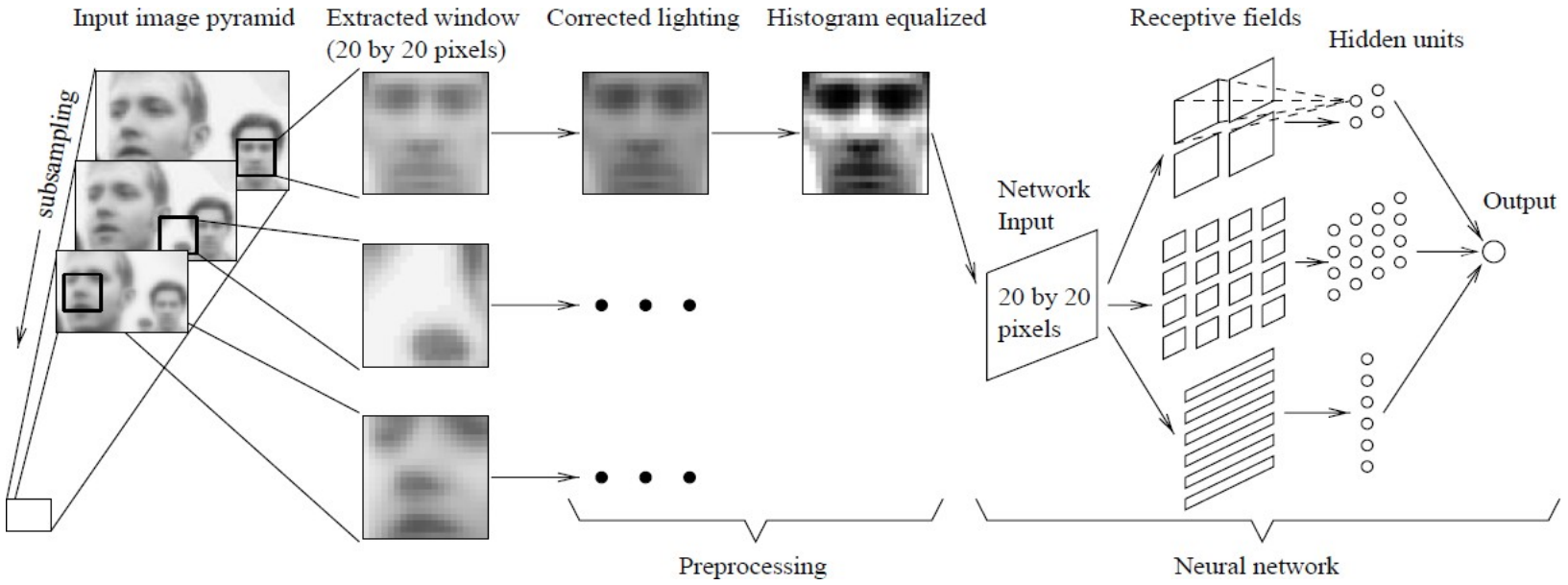
1989



Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989

1998

Faces

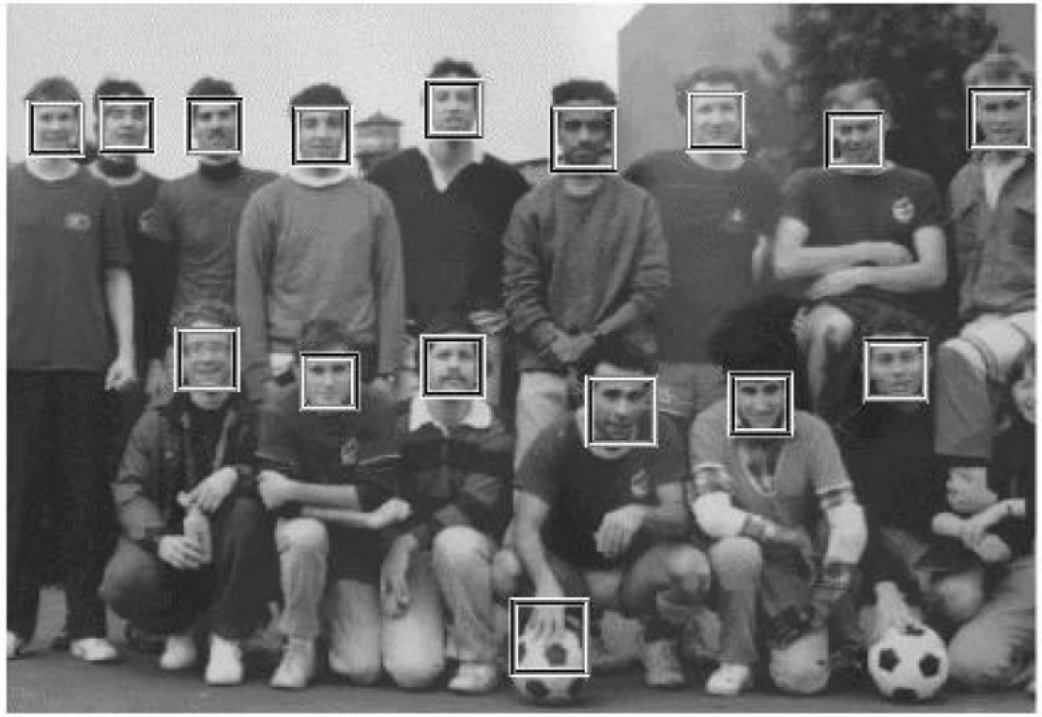


Neural Network--Based Face Detection, Rowley at al., PAMI 1998

2001

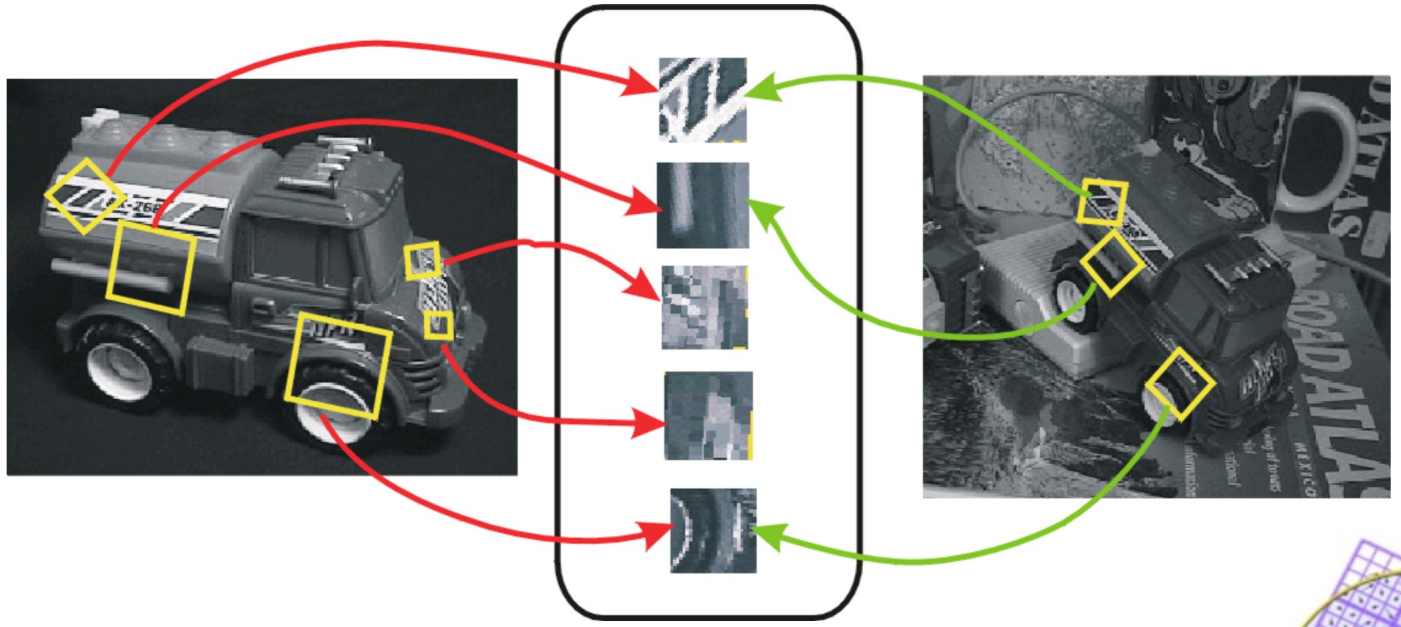
Sliding window in real time!

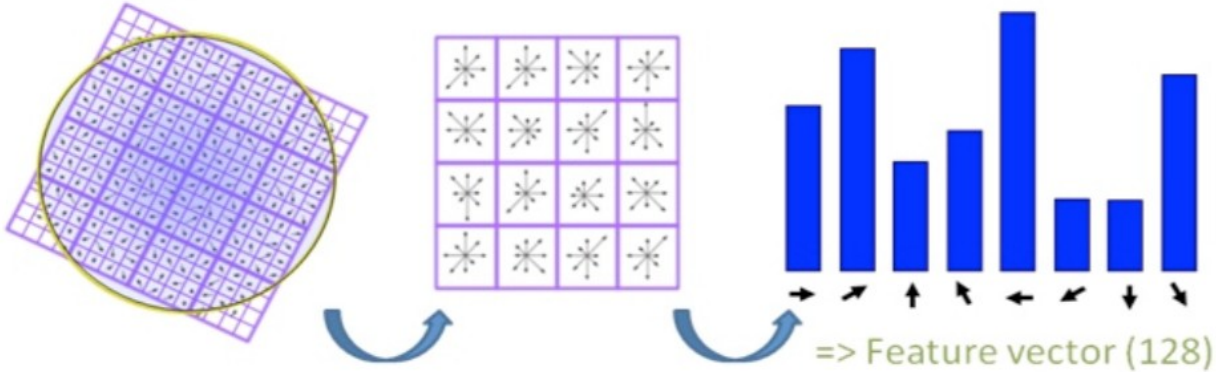
Boosting + Cascade =Speed



Rapid Object Detection using a Boosted Cascade of Simple Features,
Viola and Jones, CVPR 2001

1999 SIFT (Scale Invariant Feature Transform)

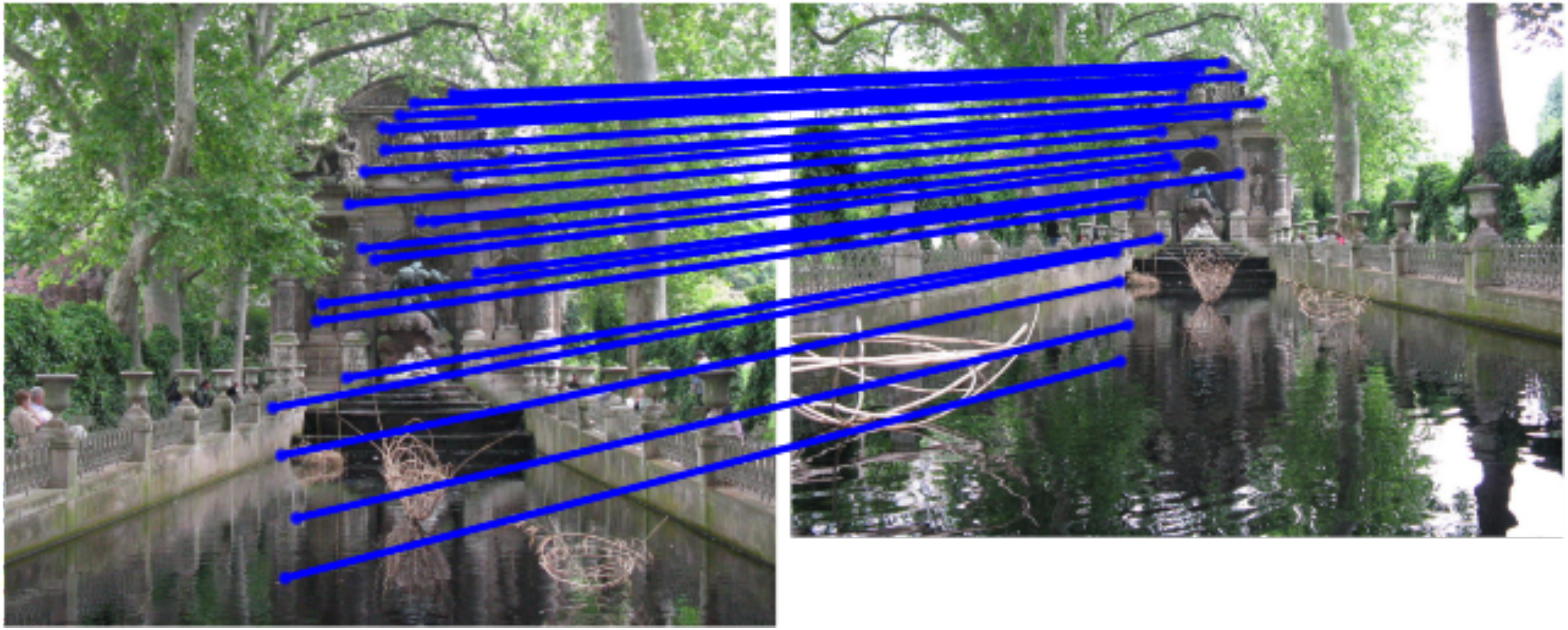




No more sliding windows (interest points)
Better features (use more computation)

Object Recognition from Local Scale--Invariant Features, Lowe, ICCV 1999.

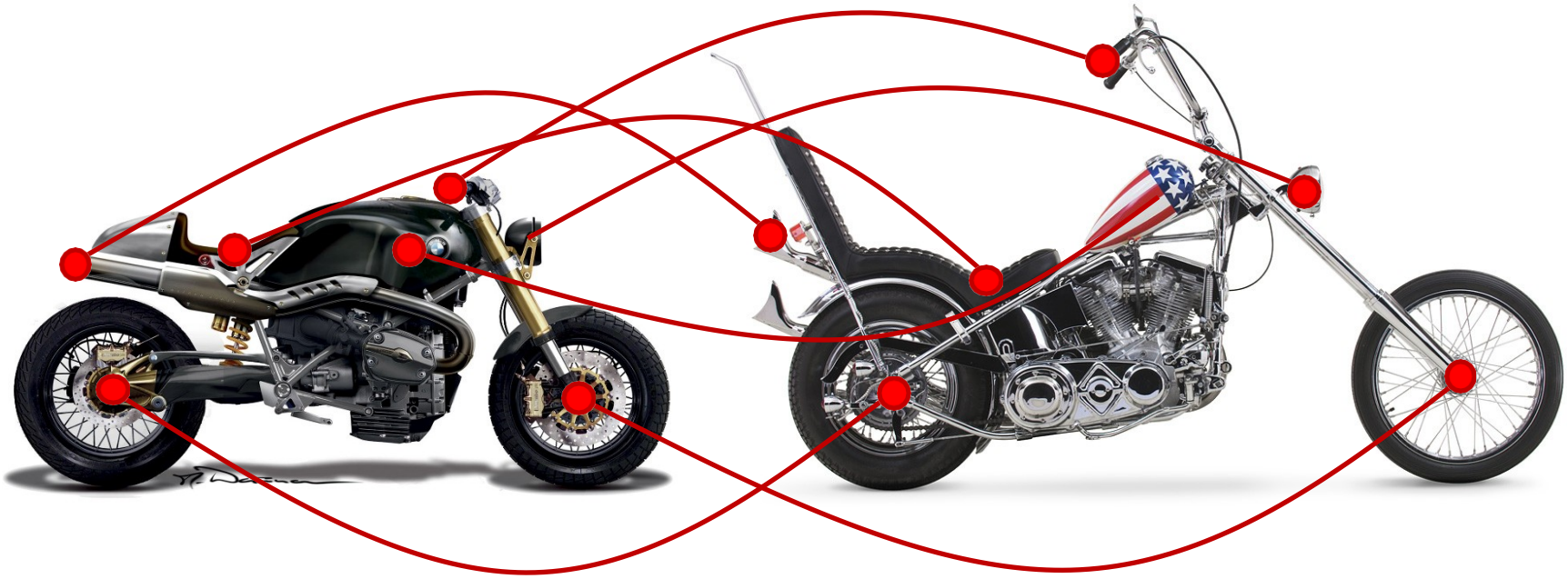
SIFT Matching



[SIFT: Lowe, 2004]

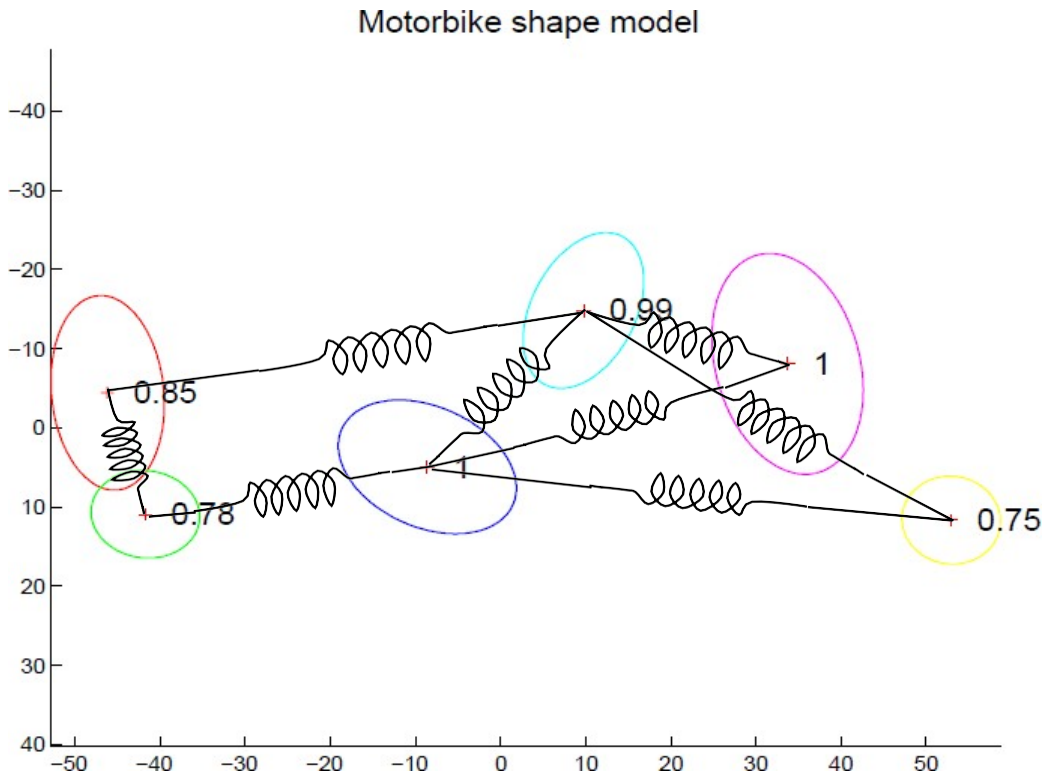
Interest points

2003 Constellation model (redux)

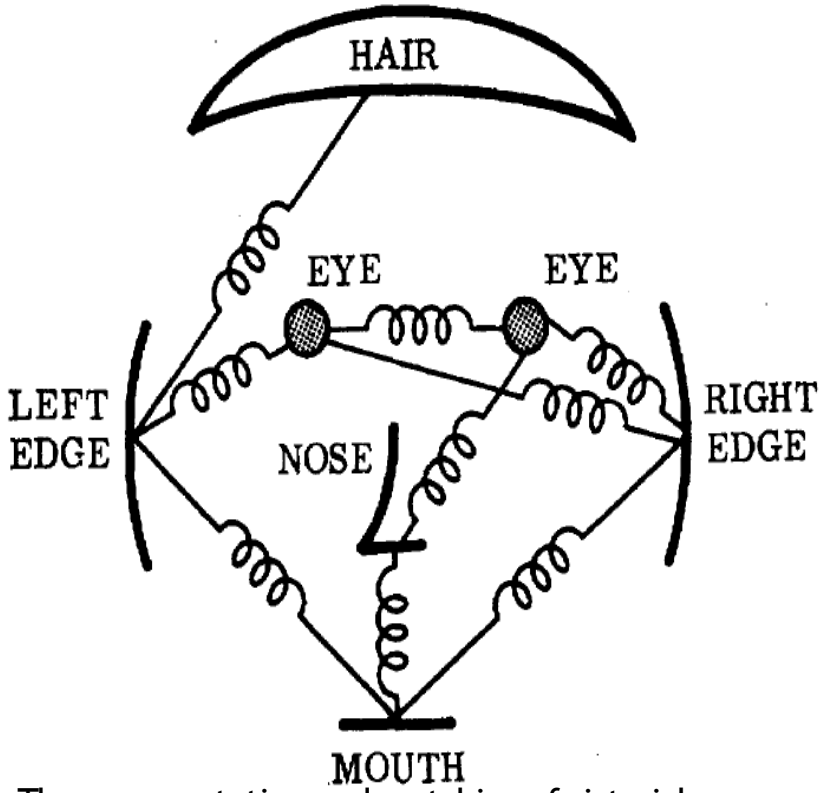


Object Class Recognition by Unsupervised Scale--Invariant Learning,
Fergus et al., *CVPR* 2003.

2003 Constellation model (redux)

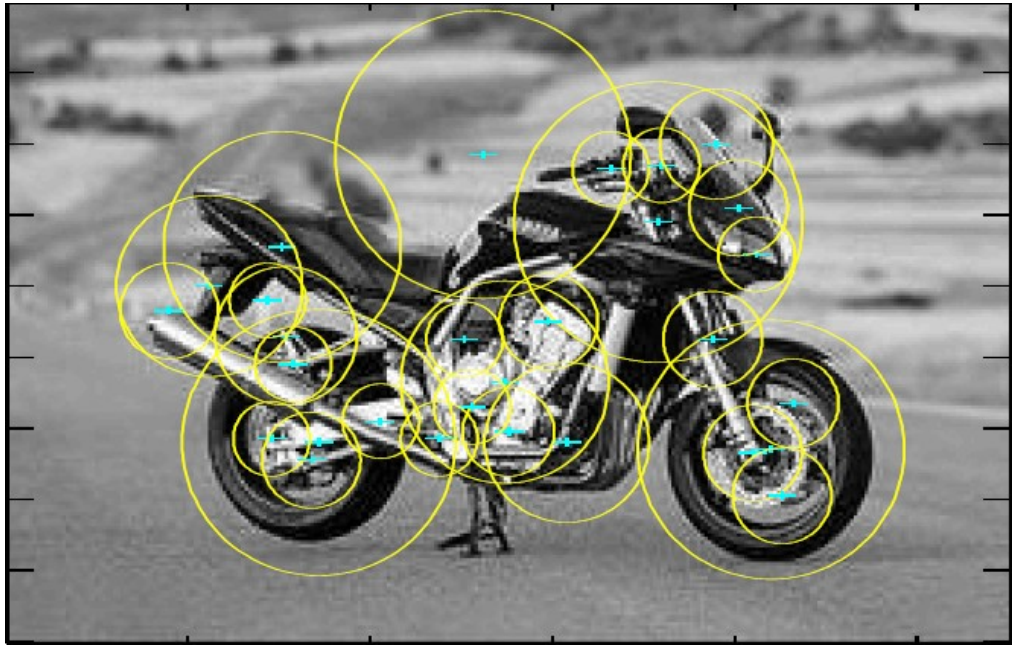


Joint Gaussian density



The representation and matching of pictorial structures, Fischler and Elschlager, 1973

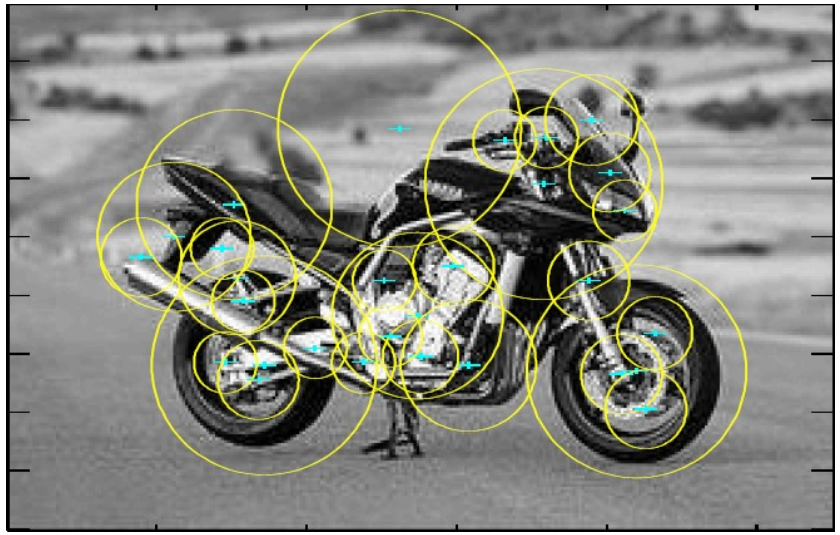
Interest points used to find parts:

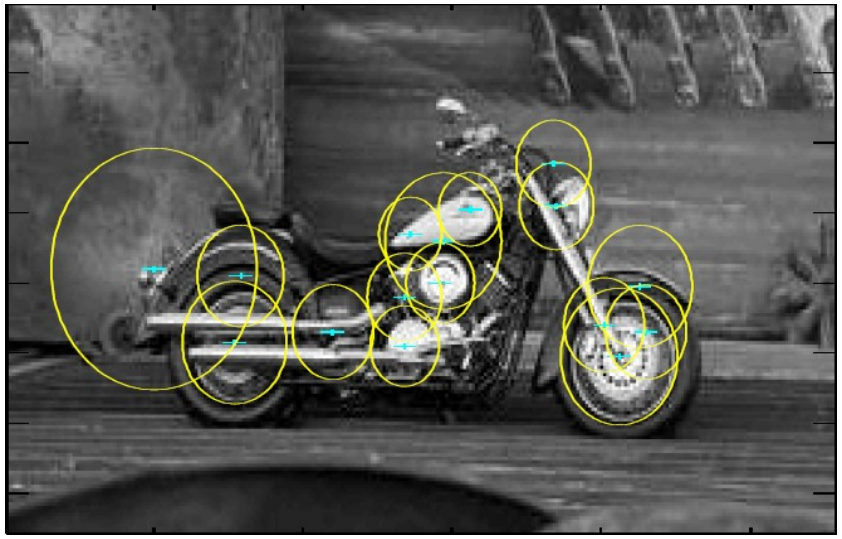


Smaller number of candidate parts allows for more complex spatial models.

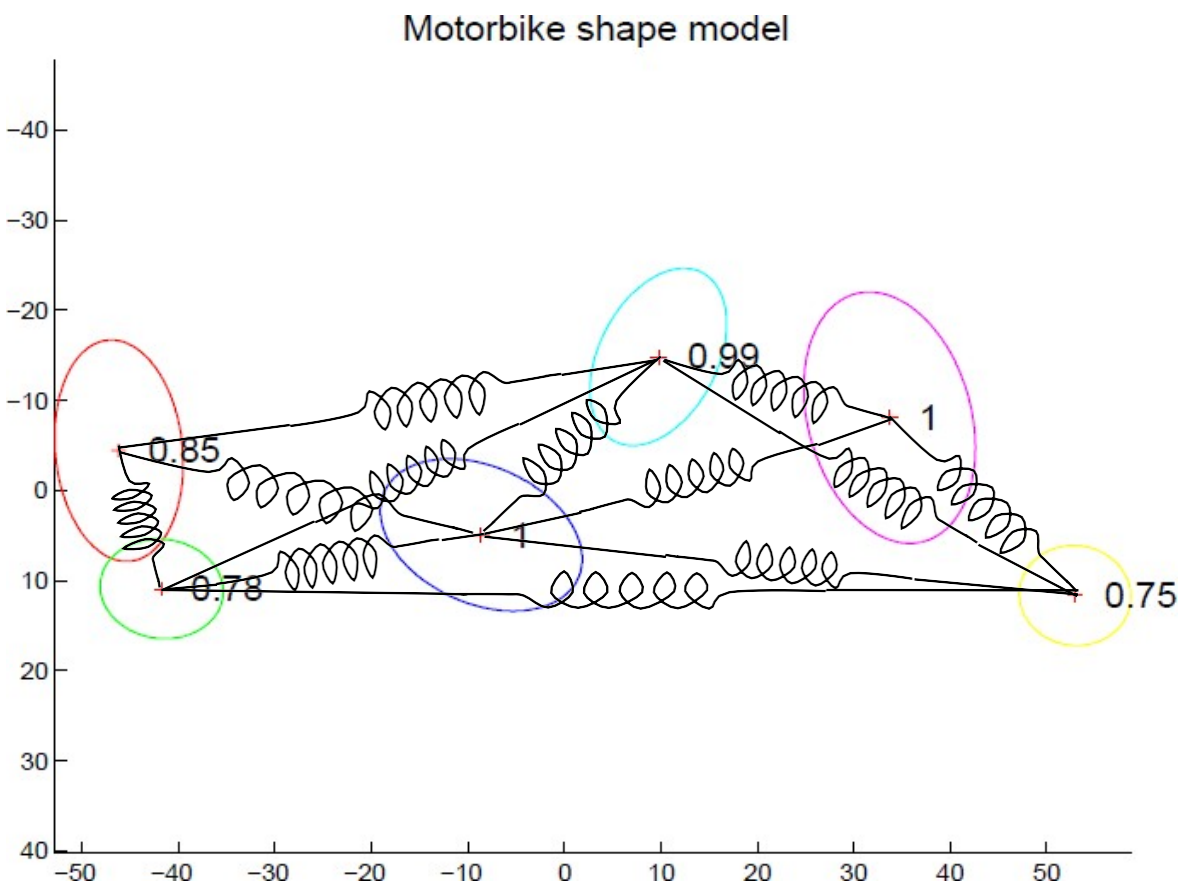
Why it fails

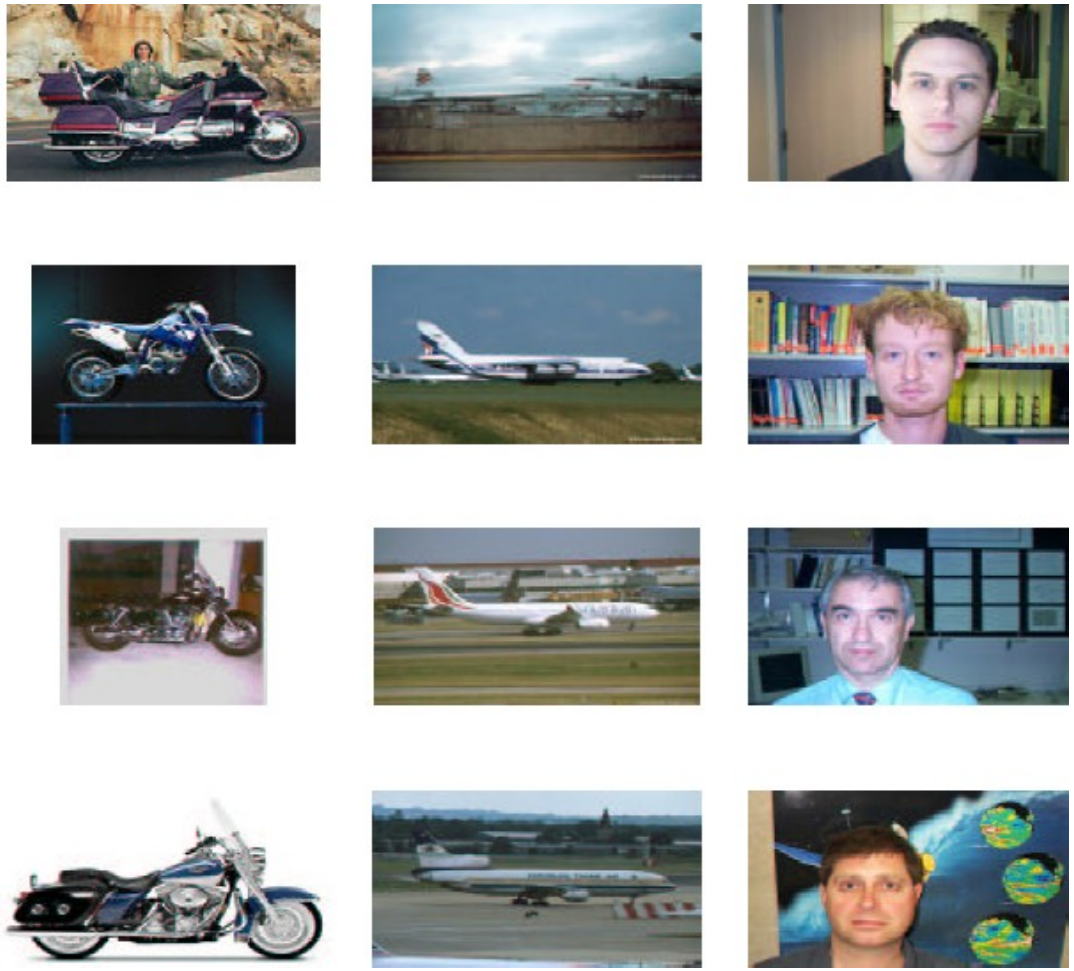
Interest points don't work for category recognition



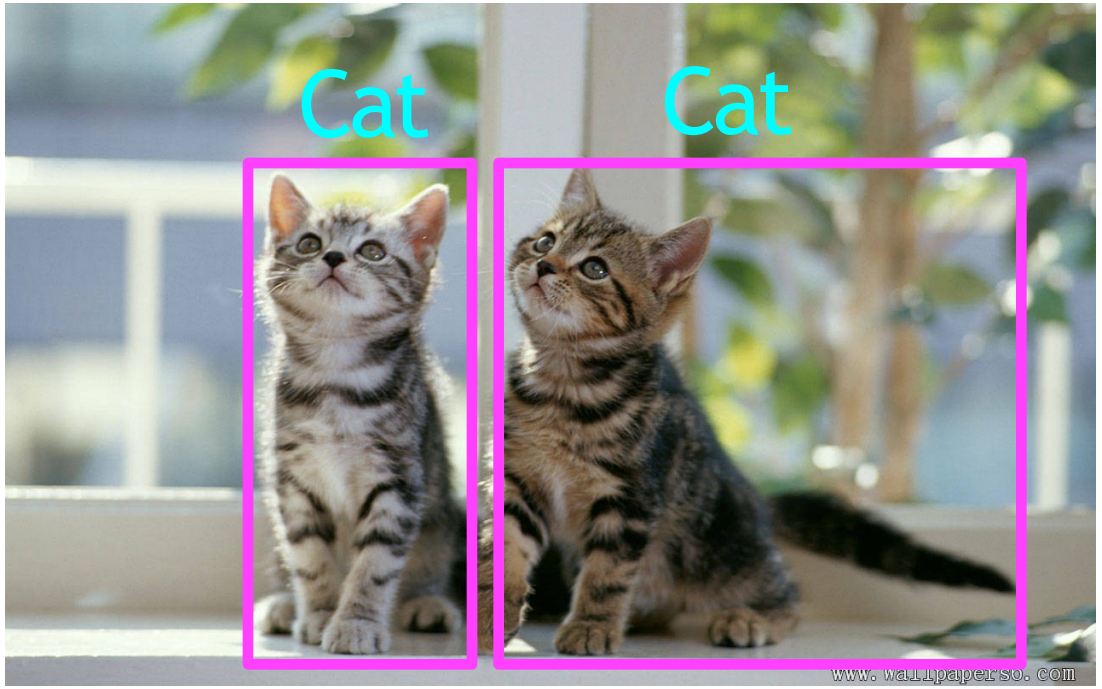


Too many springs…

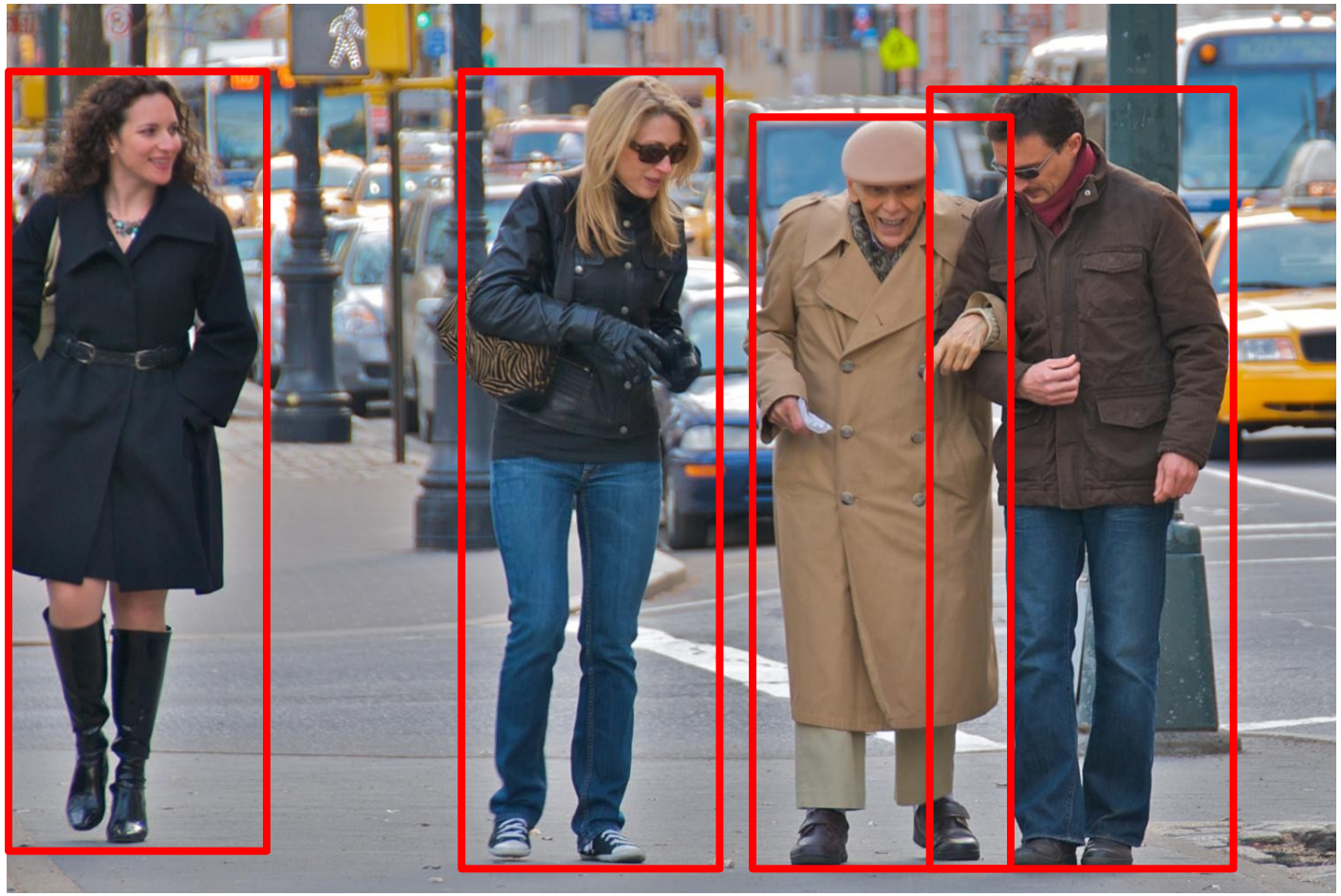




Classification Vs. Detection



2005 HOG (histograms of oriented gradients)



Histograms of oriented gradients for human detection,
Dalal and Triggs, CVPR 2005.

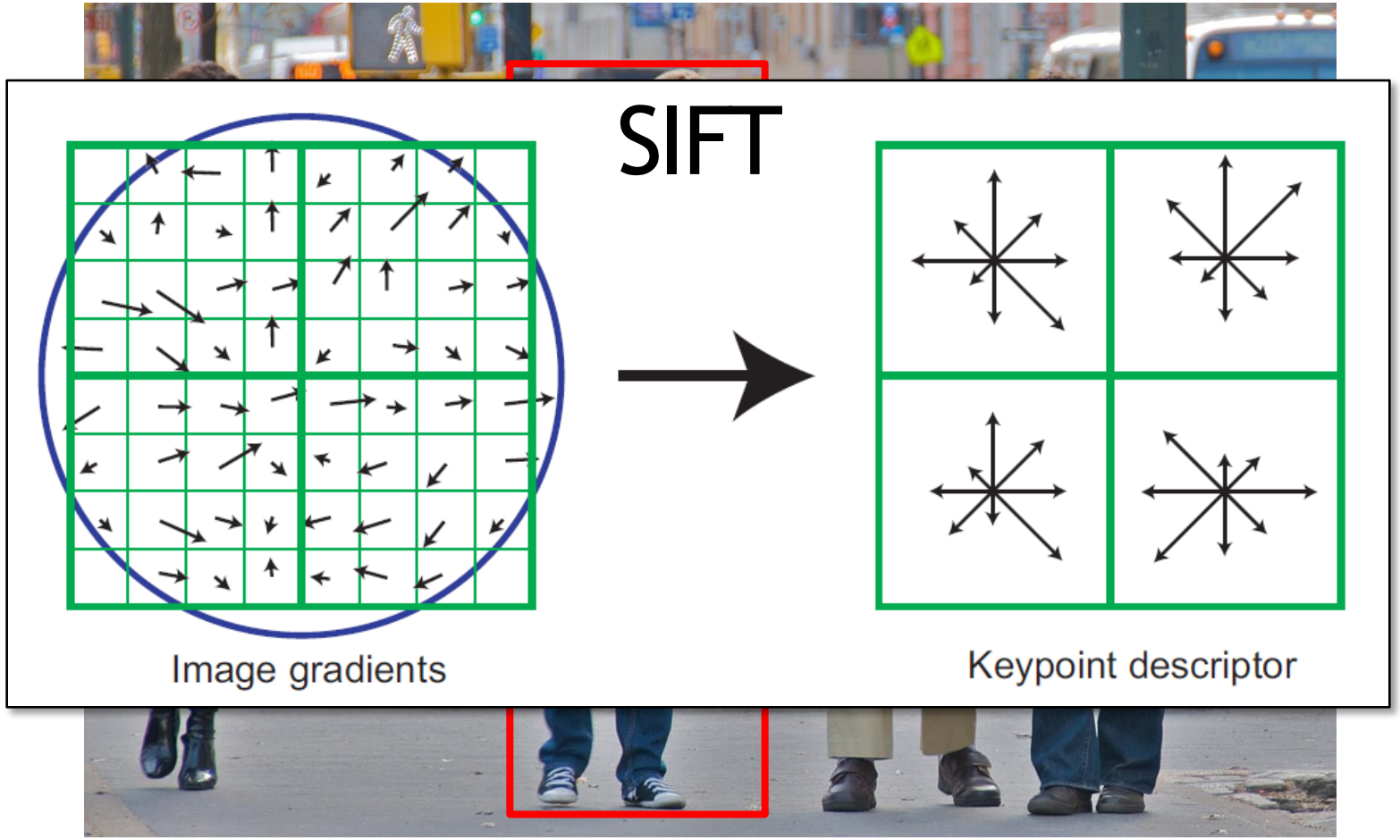
Pedestrians

- Defined by their contours
- Cluttered backgrounds
- Significant variance in texture

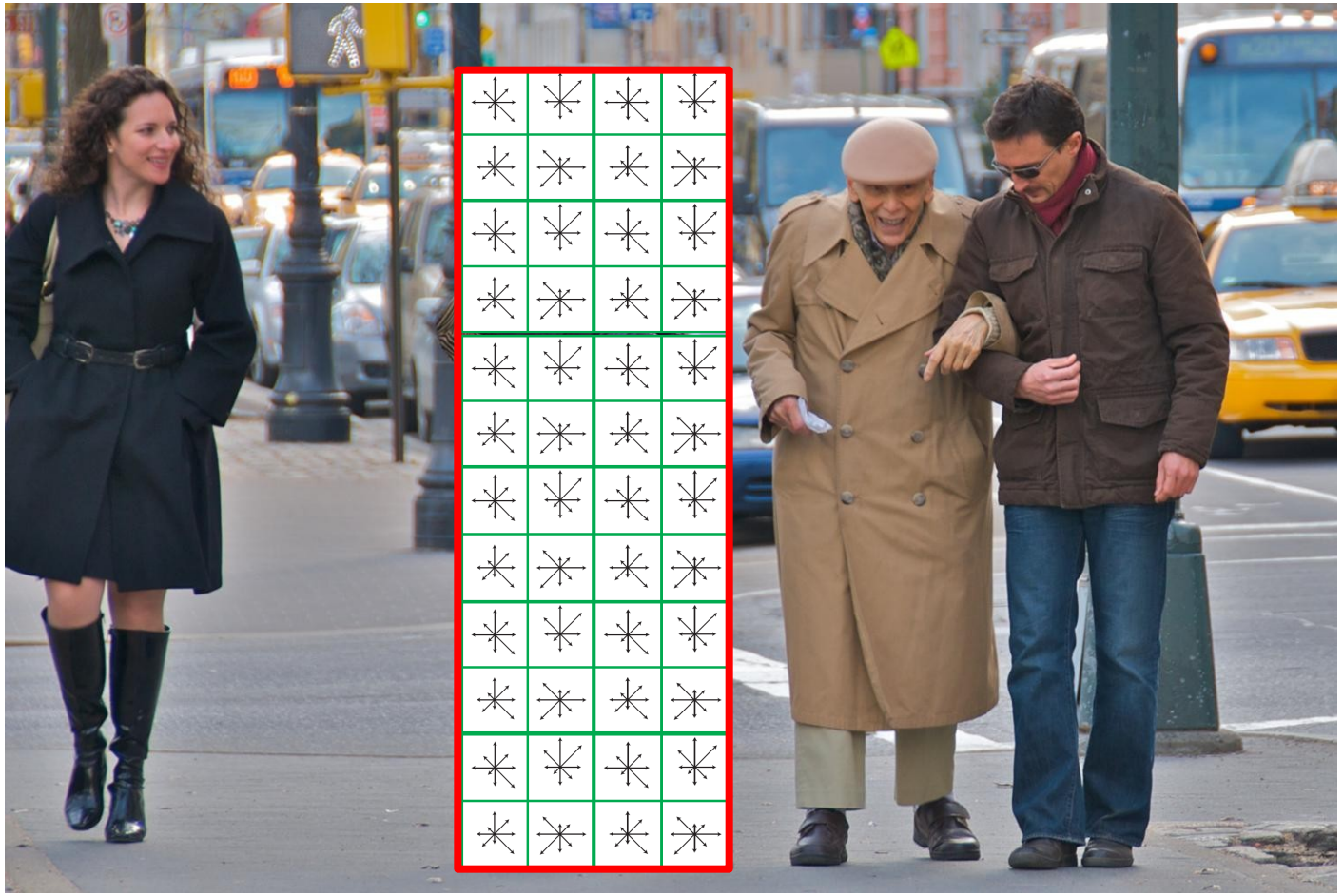
Interest points won't work...

...back to sliding window.

2005 HOG (histograms of oriented gradients)



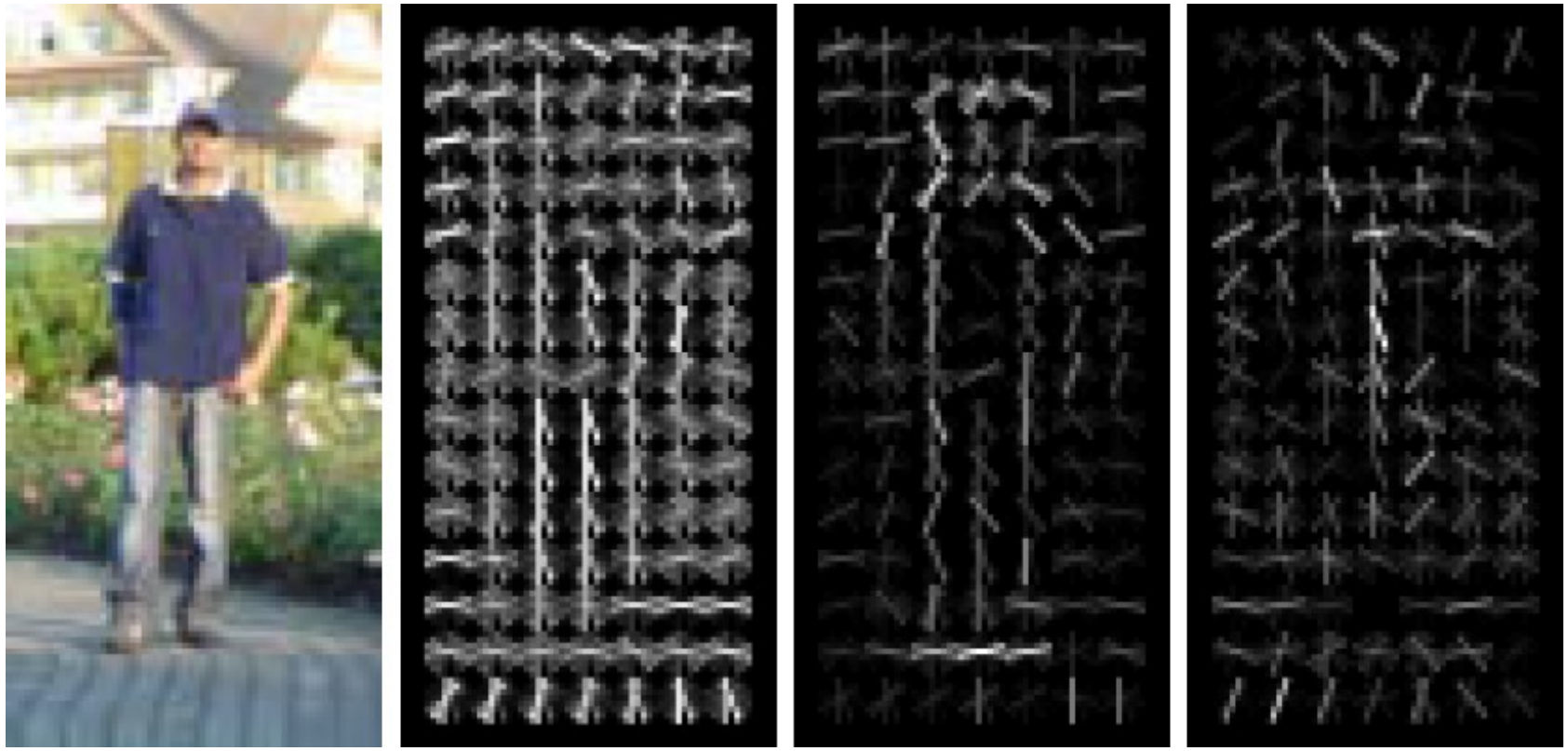
2005 HOG (histograms of oriented gradients)



Histograms of oriented gradients for human detection,
Dalal and Triggs, *CVPR* 2005.

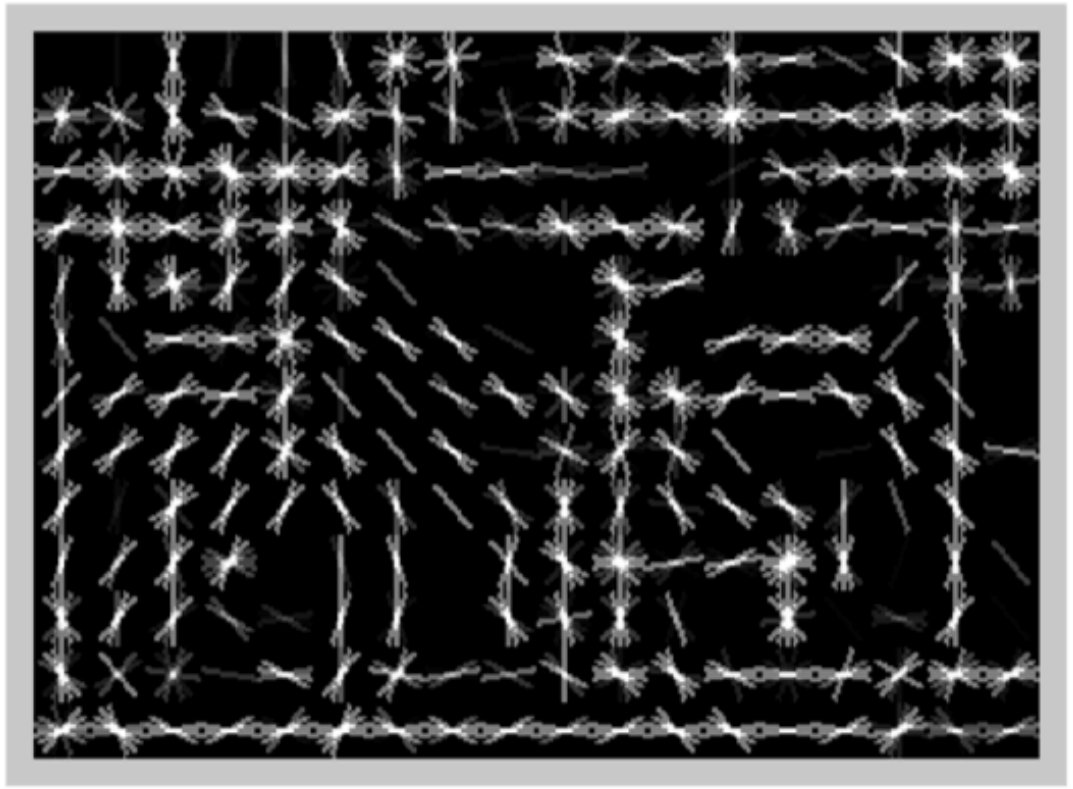
2005 HOG (histograms of oriented gradients)

Presence > Magnitude



✓✓ Normalization by a local window

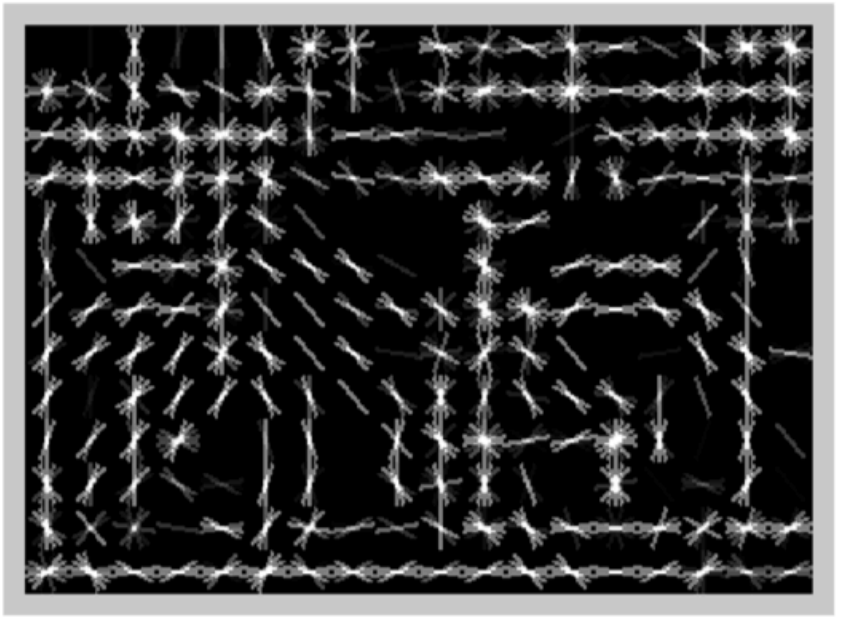
What's this?



[Dalal and Triggs, 2005]

Yup.

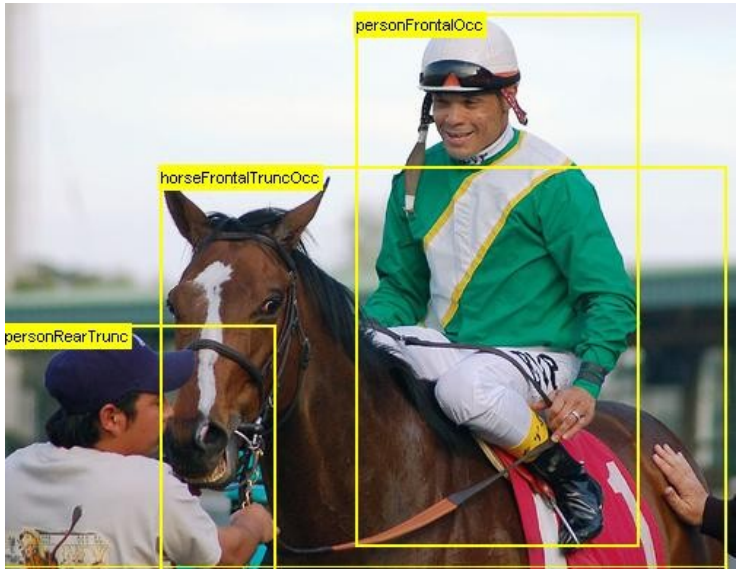
Image



HoG

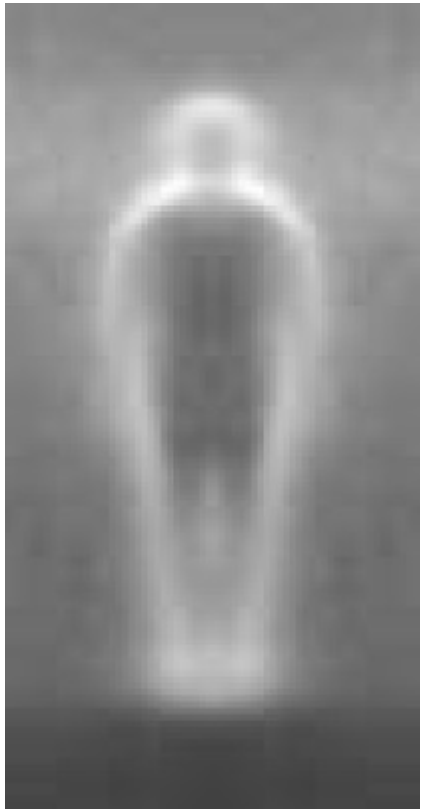
2007 PASCALVOC

20 classes

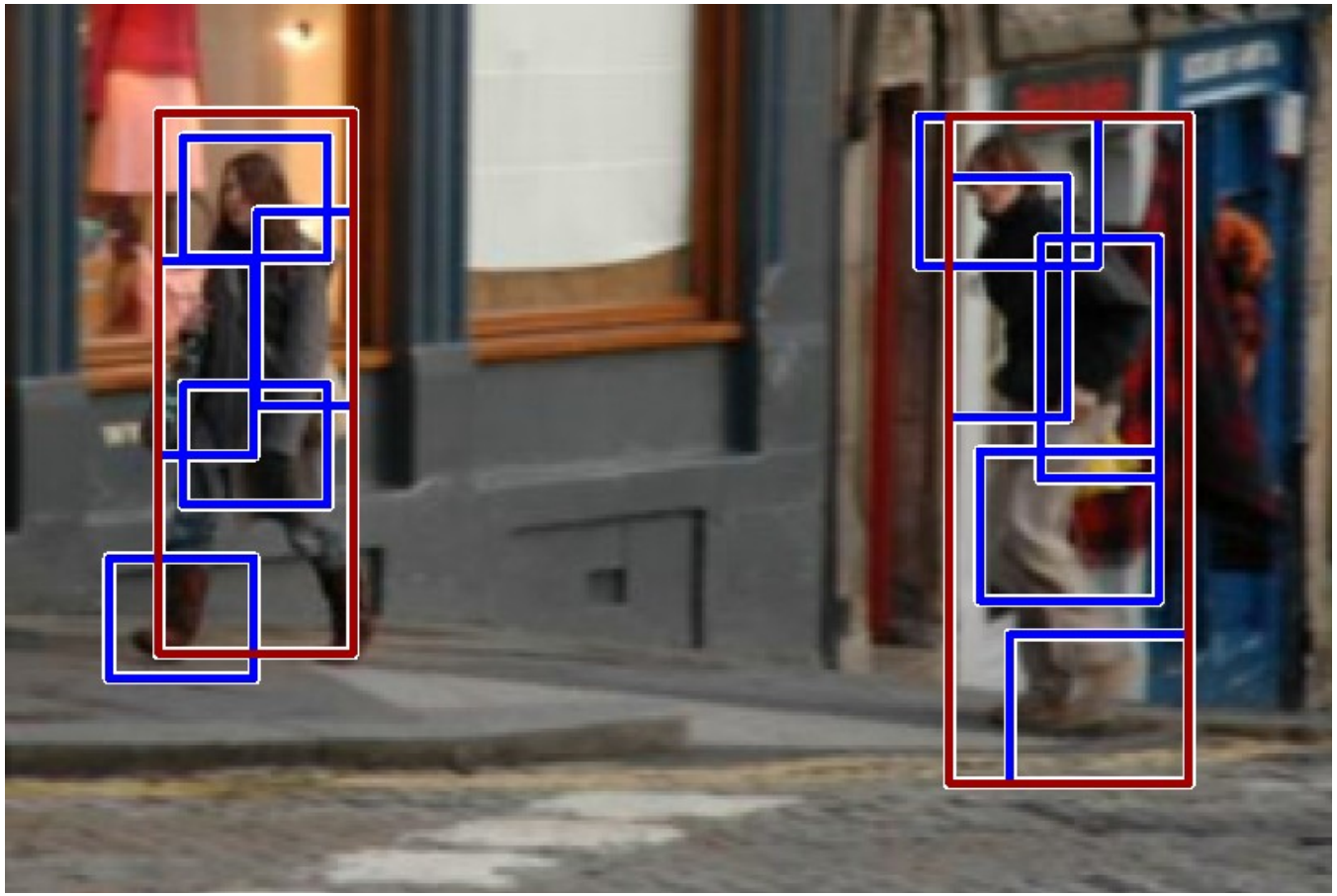


The PASCAL Visual Object Classes (VOC) Challenge, Everingham, Van Gool, Williams, Winn and Zisserman, *IJCV*, 2010

Why it is hard

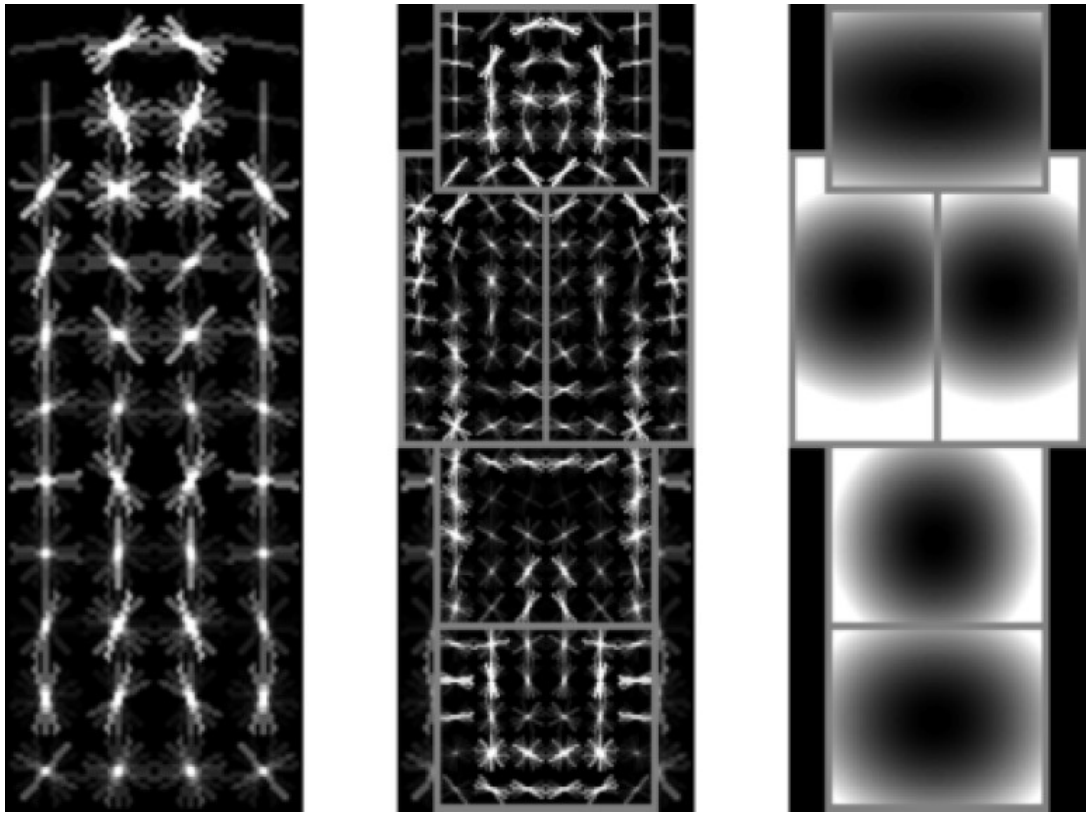


2008 DPM (Deformable parts model)



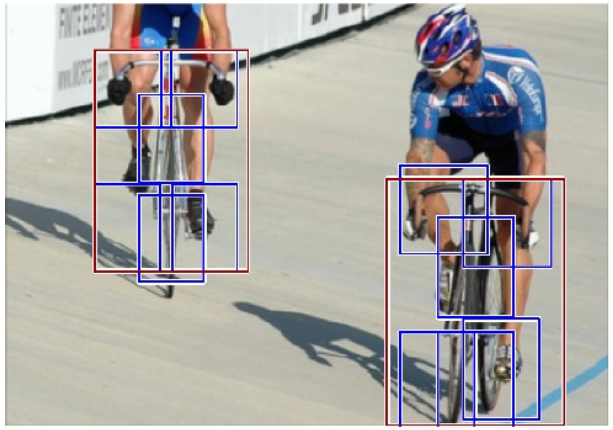
Object Detection with Discriminatively Trained Part Based Model,
Felzenszwalb, Girshick, McAllester and Ramanan, *PAMI*, 2010

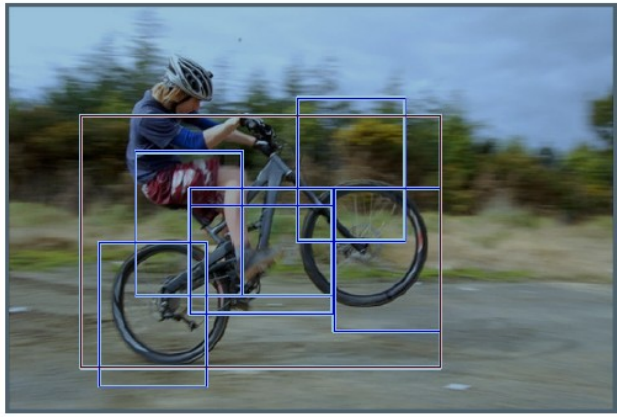
2008 DPM (Deformable parts model)

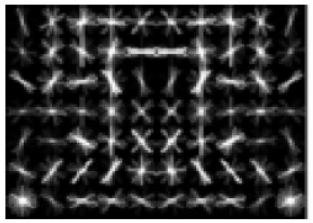


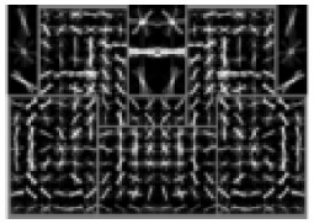
Object Detection with Discriminatively Trained Part Based Model,
Felzenszwalb, Girshick, McAllester and Ramanan, *PAMI*, 2010

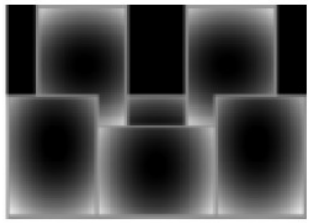
Multiple components

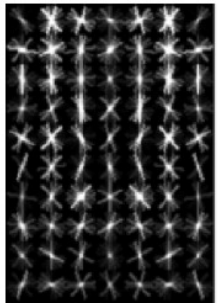


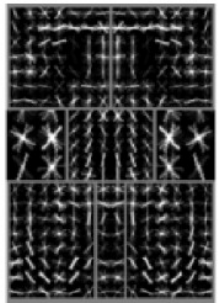


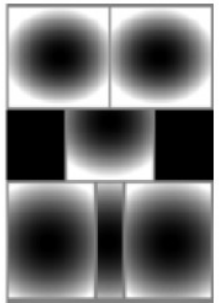












Problems with Visual Categories

- A lot of categories are functional

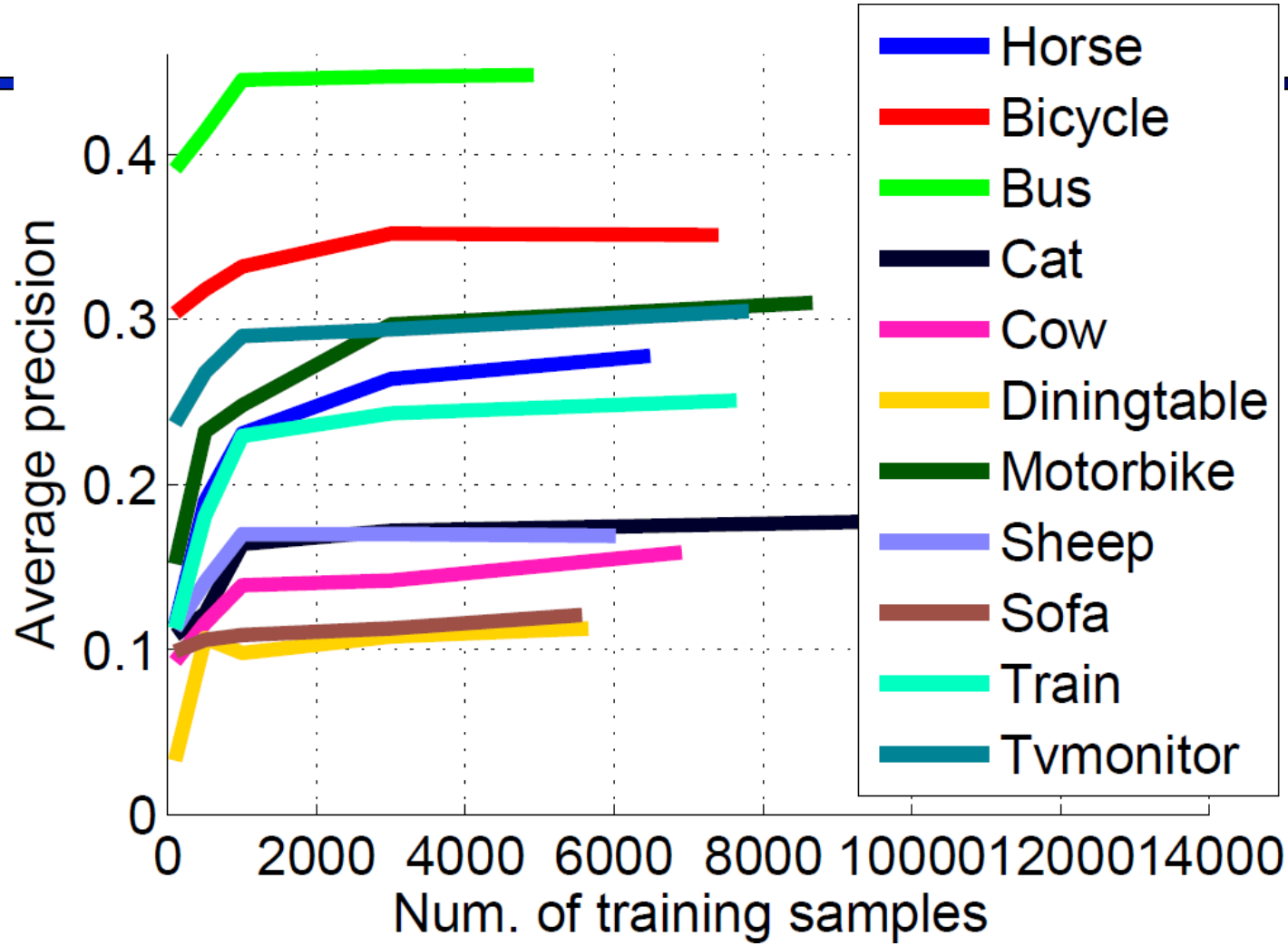
Char



- World is too varied


- Categories are 3D, but images are 2D

car

Do We Need More Training Data or Better Models for Object Detection?
Zhu, Vondrick, Ramanan, Fowlkes, *BMVC* 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

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

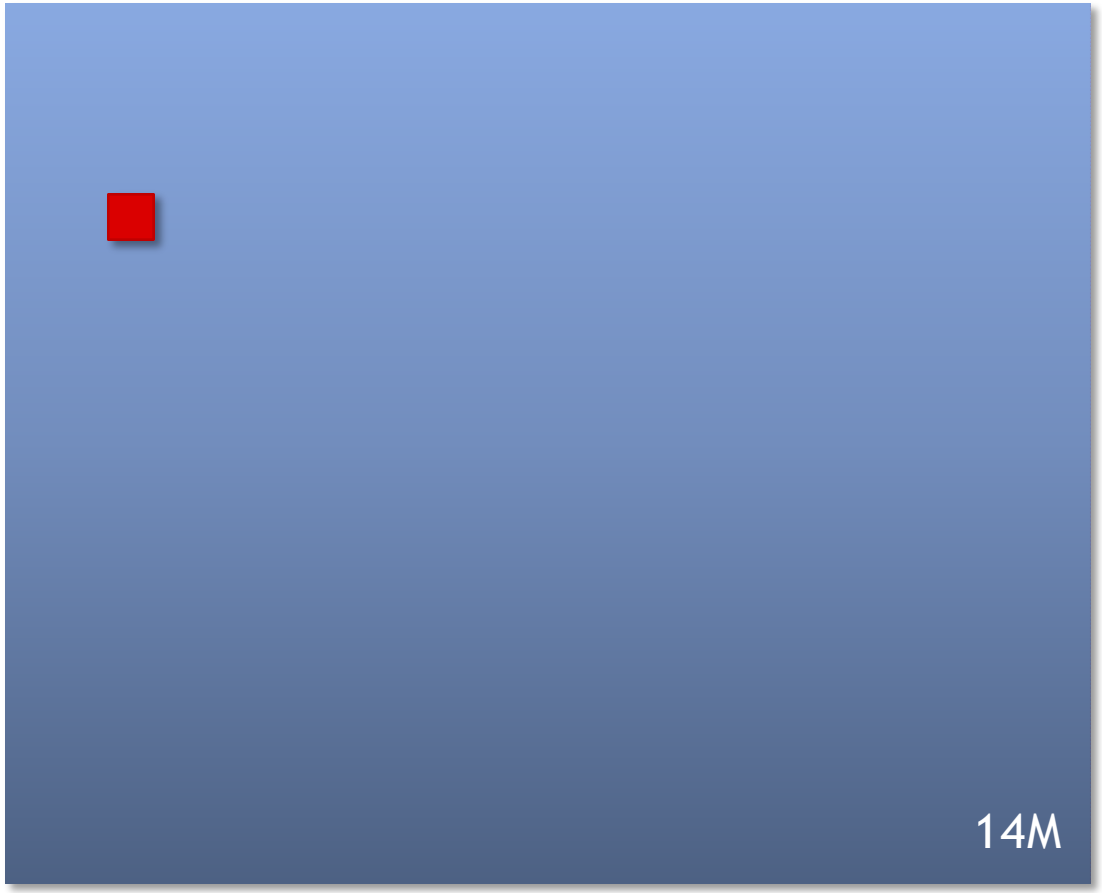
2009 ImageNet

22K categories, 14M images

ImageNet: A Large–Scale Hierarchical Image Database,
Deng, Dong, Socher, Li, Li and Fei-Fei, *CVPR*, 2009

Images

2009 2012



ImageNet

Categories

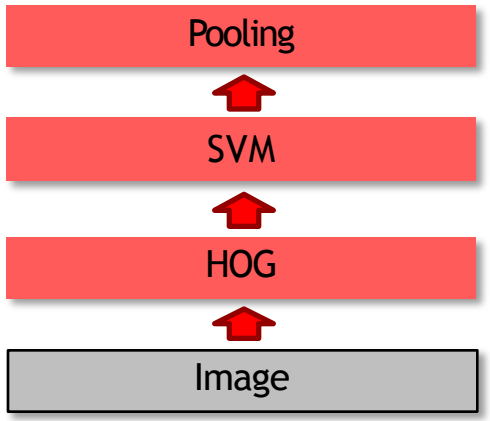
2009 2012

ImageNet

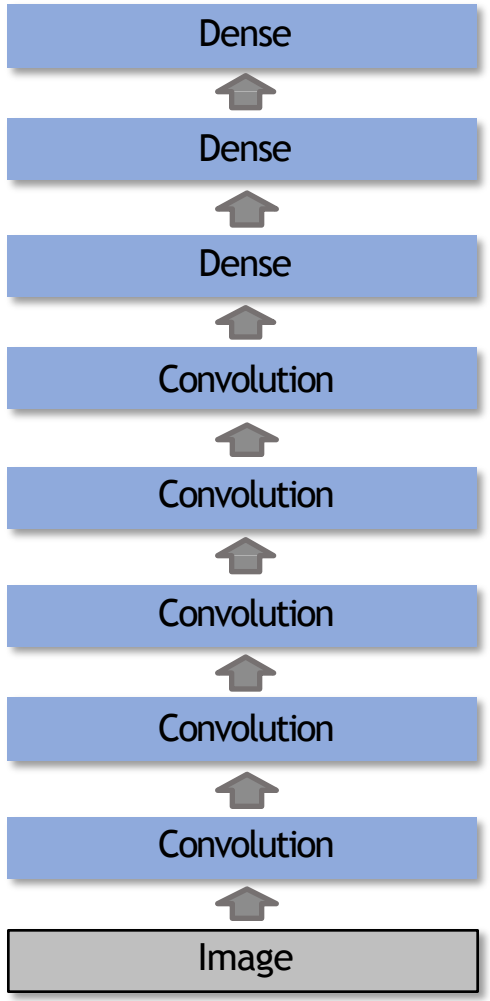
22K

Algorithms

2009



2012

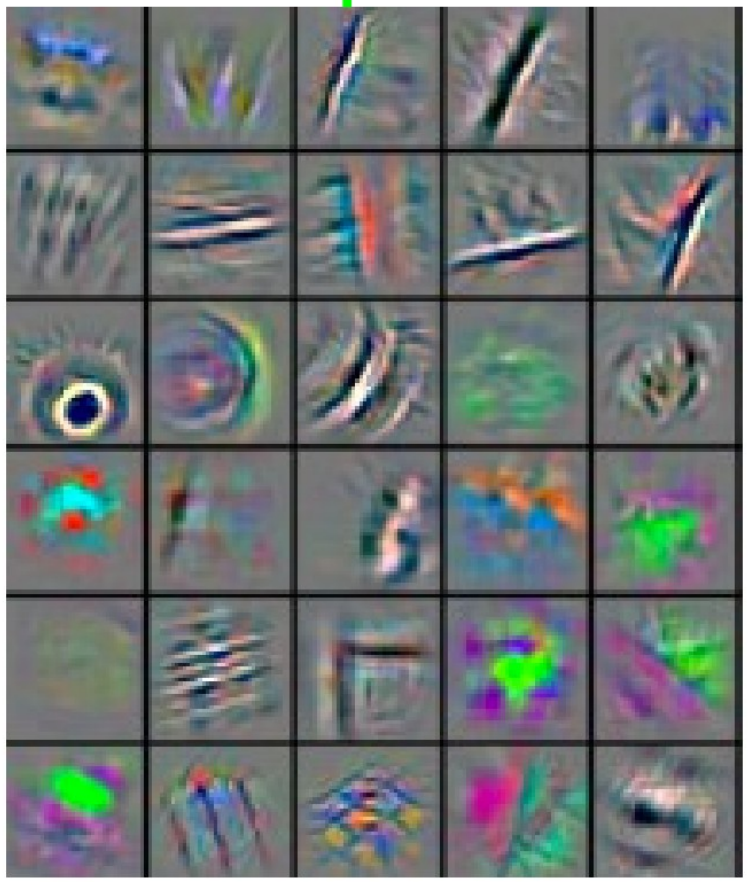


Low-Level Feature

Mid-Level Feature

High-Level Feature

Trainable Classifier



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense grid descriptor: HOG, LBP

Coding: local coordinate, super-vector

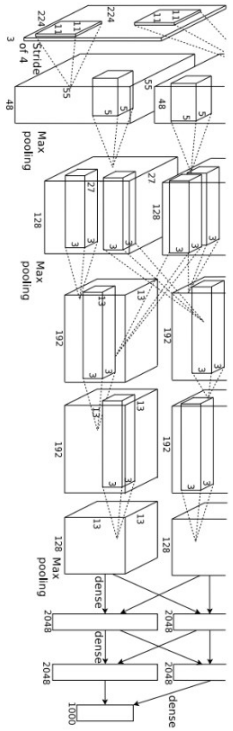
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

Year 2014

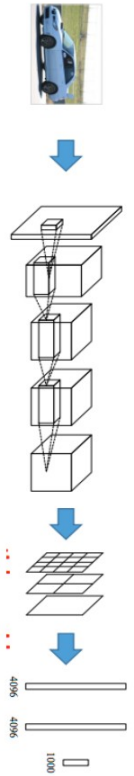
GoogLeNet

[Szegedy arxiv 2014]

VGG

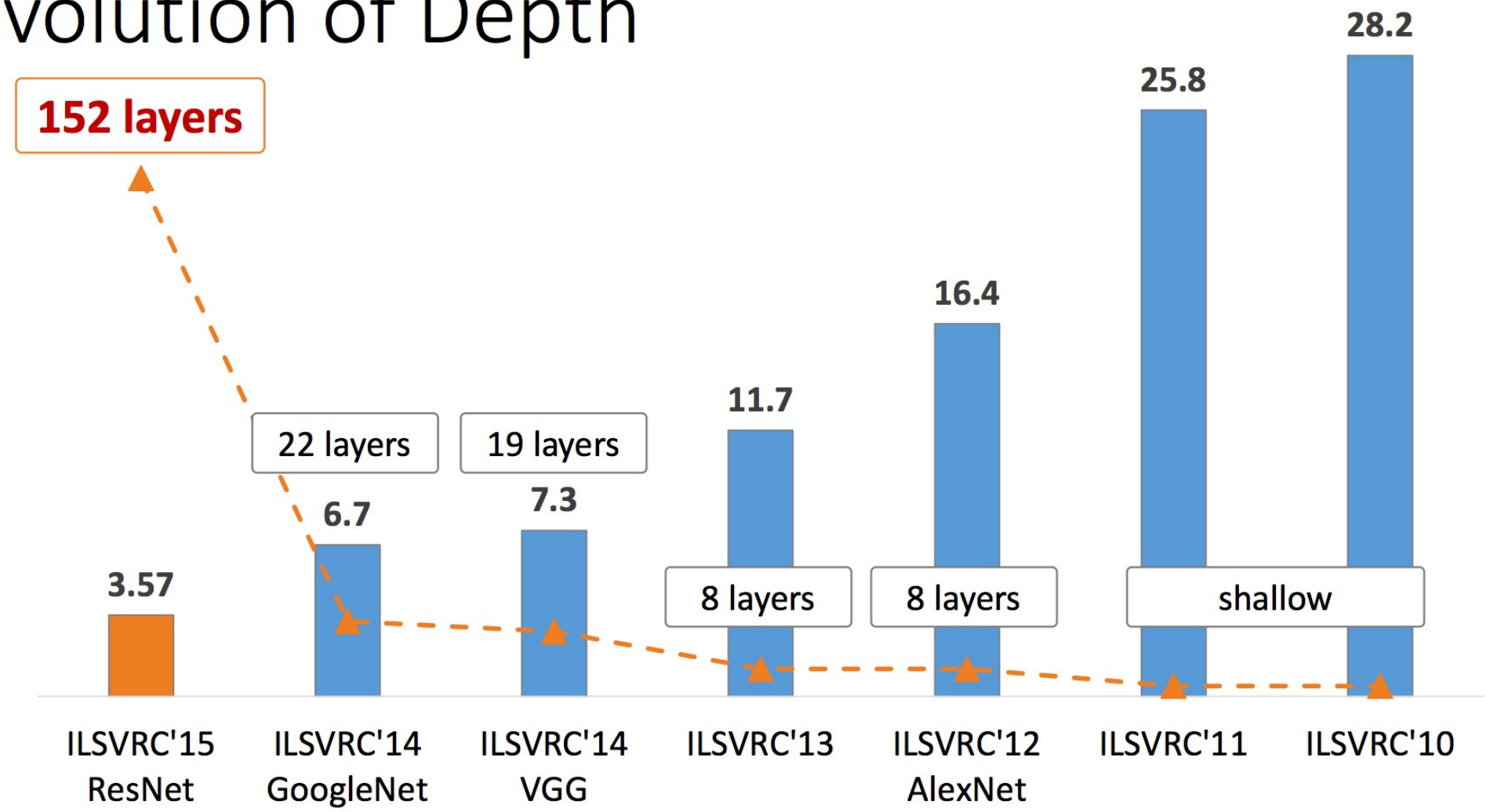
[Simonyan arxiv 2014]

MSRA



[He arxiv 2014]

Revolution of Depth

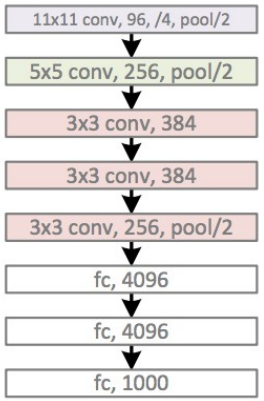


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

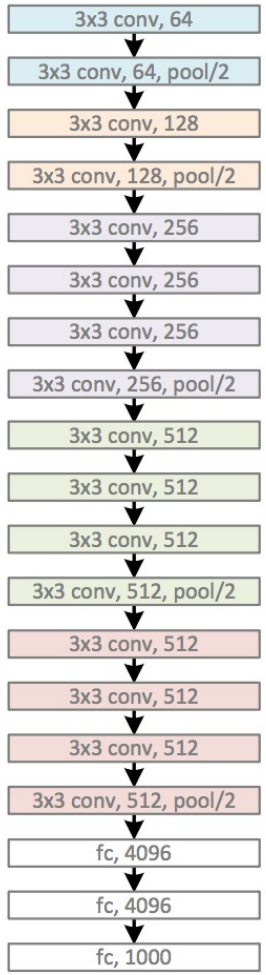
Microsoft Research

Revolution of Depth

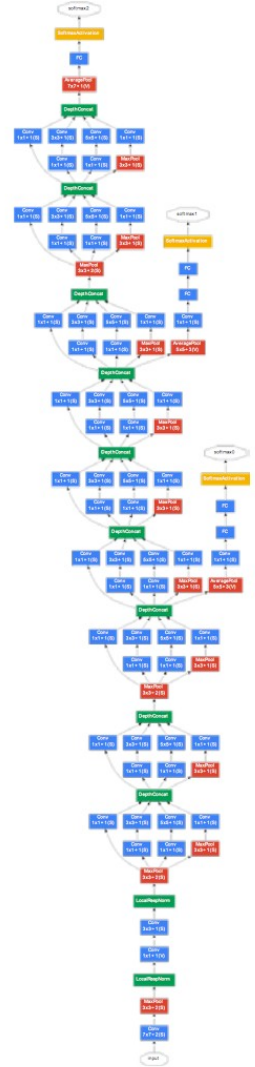
AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



GoogleNet, 22 layers
(ILSVRC 2014)



ICCV15 International Conference on Computer Vision

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

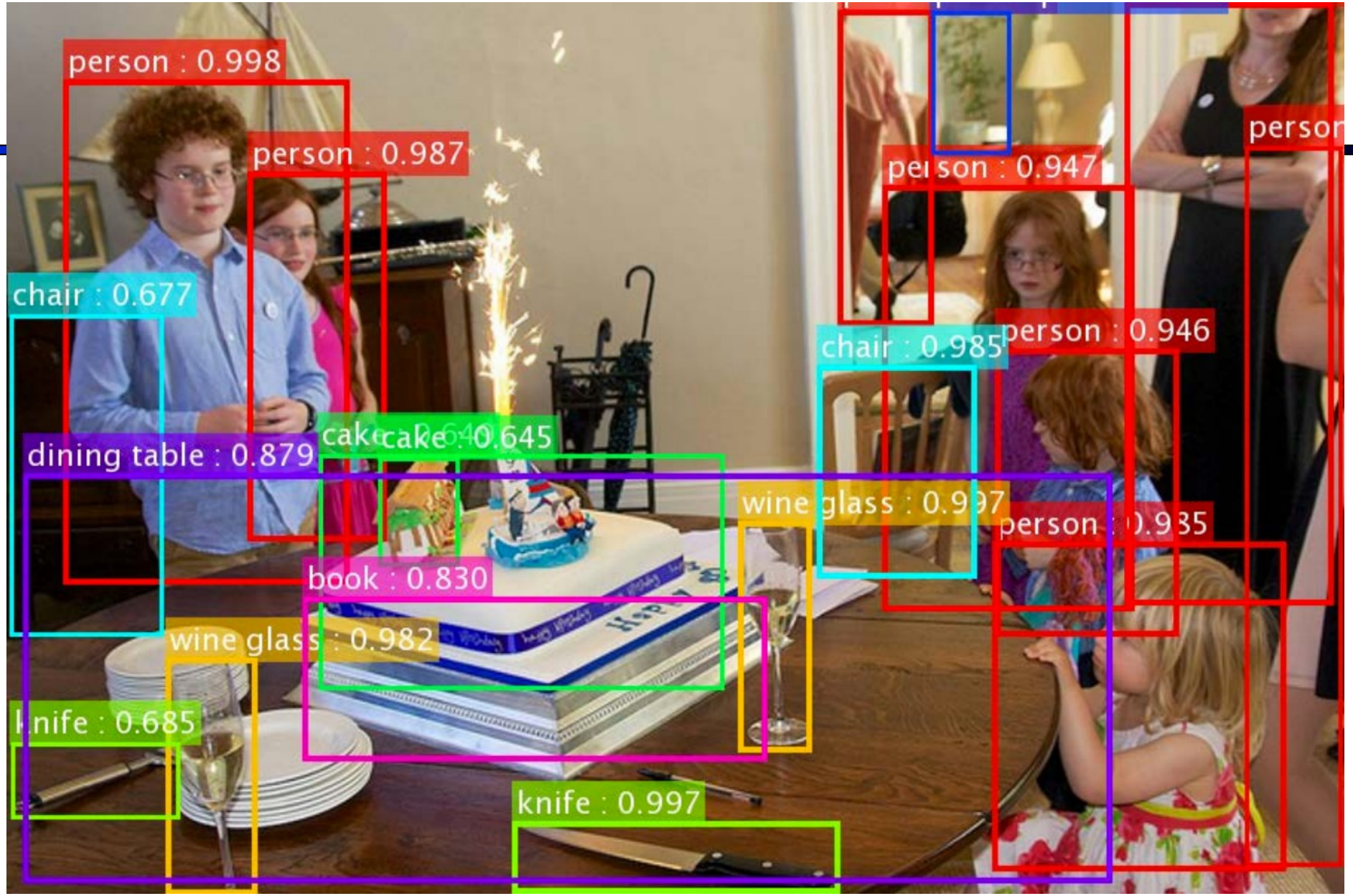
Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)

VGG, 19 layers
(ILSVRC 2014)

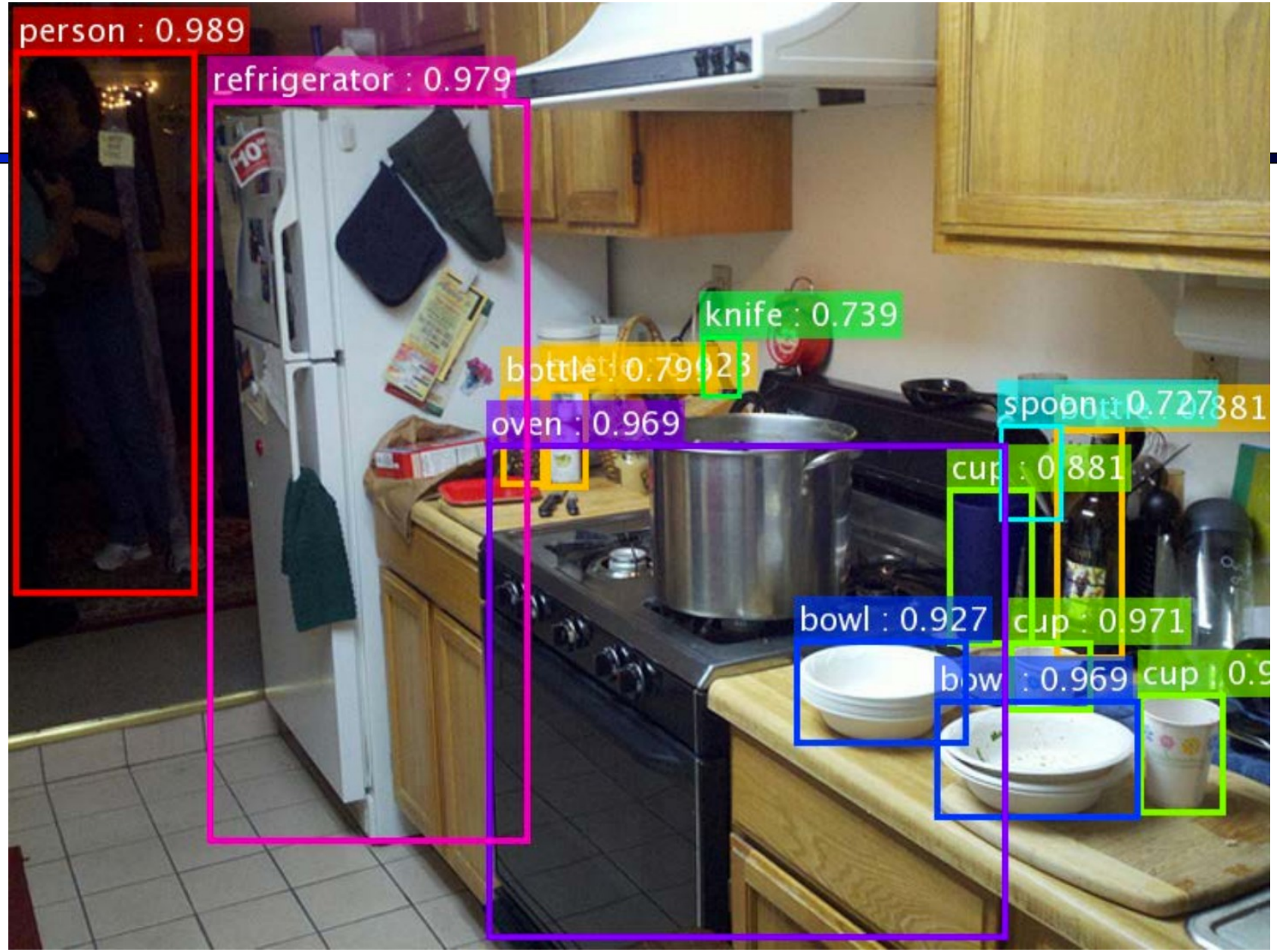
ResNet, 152 layers
(ILSVRC 2015)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.



*the original image is from the COCO dataset

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.
Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. NIPS 2015.



*the original image is from the COCO dataset

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.
Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. NIPS 2015.

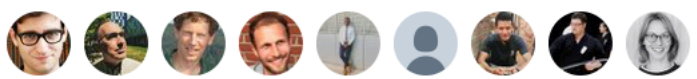
Oren Etzioni
@etzioni

Following

The winner of the 2014 ImageNet competition had 4 million parameters, while the winner of the 2017 challenge had 145.8 million parameters - a 36X increase in three years. source: @jackclarkSF Shall we increase parameters by another 36X, or solve more interesting problems?1

9:21 PM - 27 Nov 2018

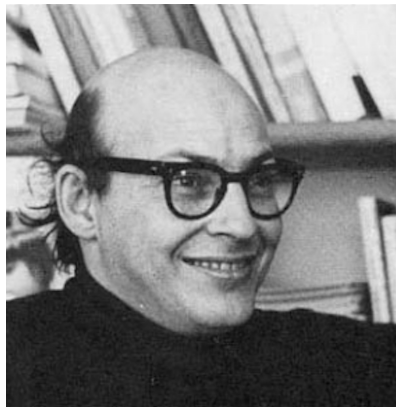
63 Retweets **222** Likes



12 63 222

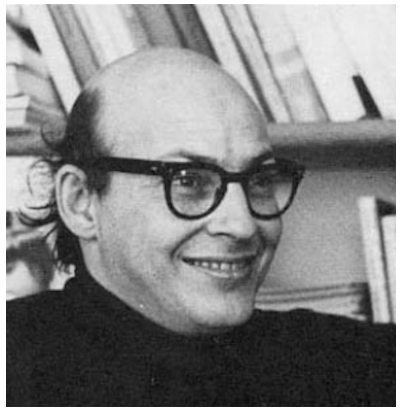
Going beyond categorization…

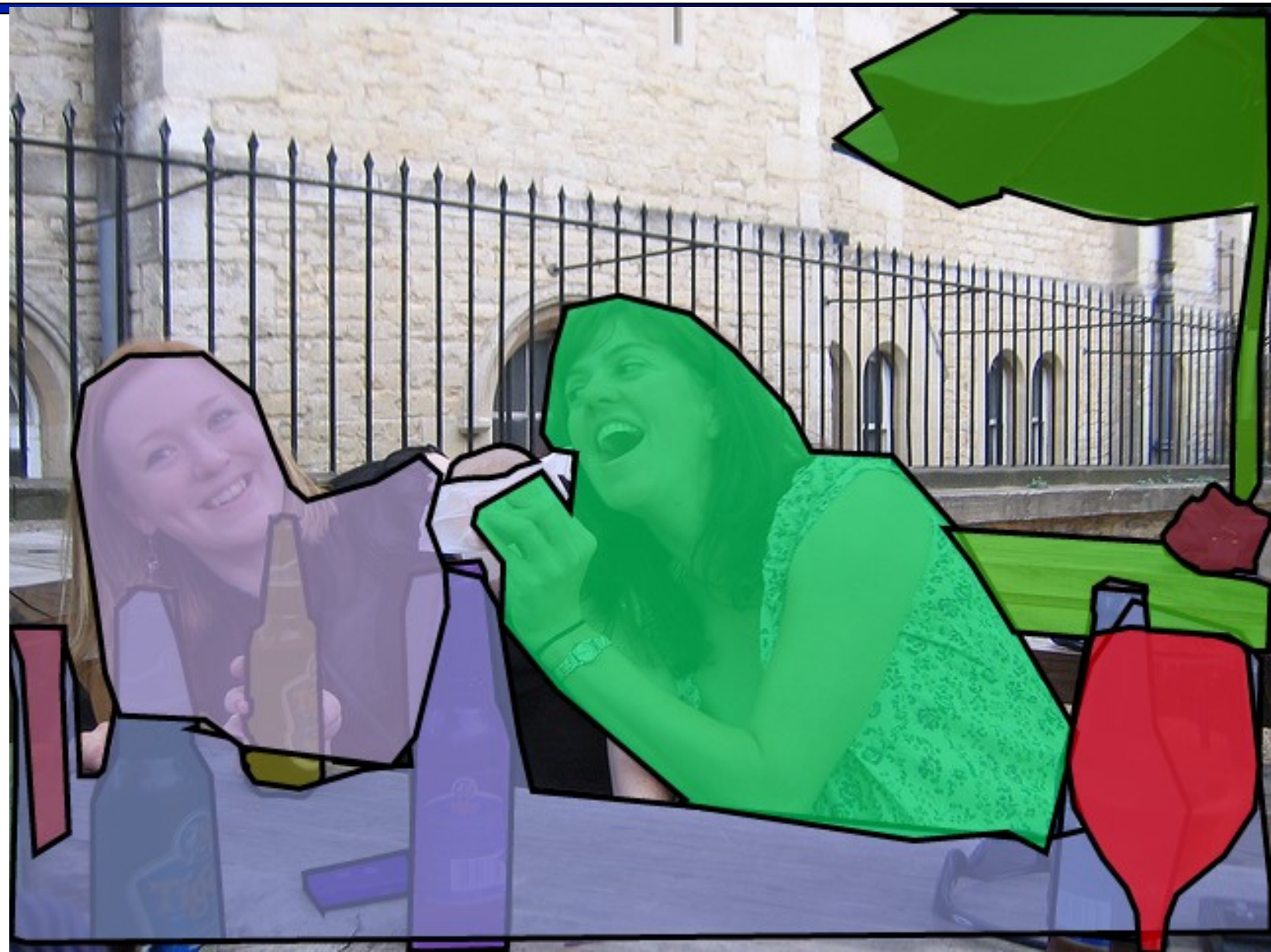
"Connect a television camera to a computer and **get the machine to describe what it sees.**"



Going beyond categorization…

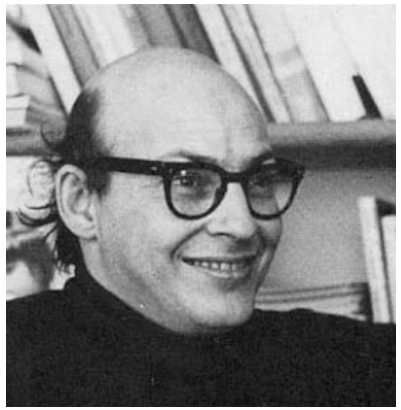
"Connect a television camera to a computer and **get the machine to describe what it sees."**





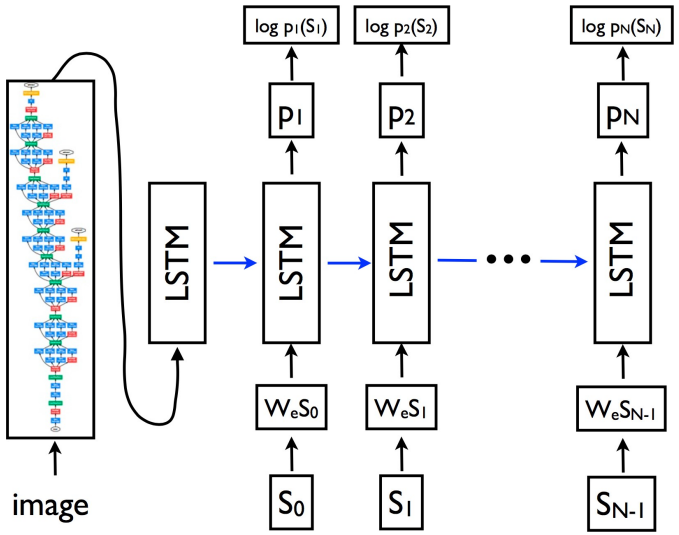
Going beyond categorization…

"Connect a television camera to a computer and **get the machine to describe what it sees.**"



two girls sitting at a table smiling and eating and drinking.
a woman is eating a doughnut and drinking beer.
there are two woman drinking beers and eating food
a woman leaning into another woman as she holds a sandwich towards her.
two ladies are enjoying beer and treats at the table.

Going beyond categorization…



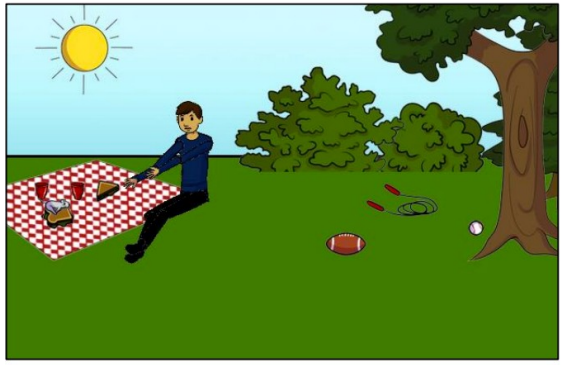
VQA: Visual Question Answering

www.visualqa.org

Stanislaw Antol*, Aishwarya Agrawal*, Jiasen Lu, Margaret Mitchell,
Dhruv Batra, C. Lawrence Zitnick, Devi Parikh

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

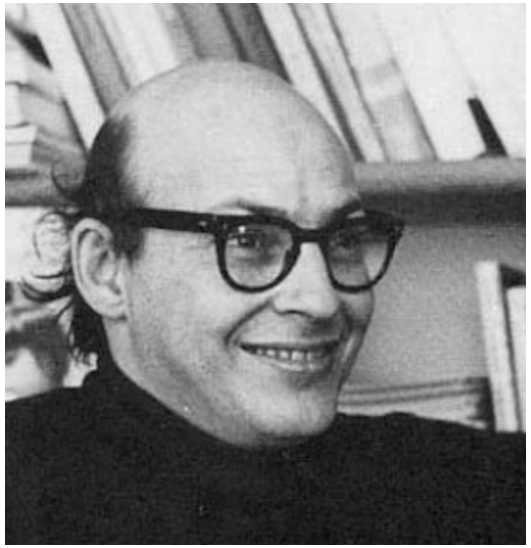


Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

1966

"Connect a television camera to a computer and get the machine to describe what it sees."



Marvin Minsky
Turing award, 1969

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".