Introduction to Computer Vision:

Object Recognition

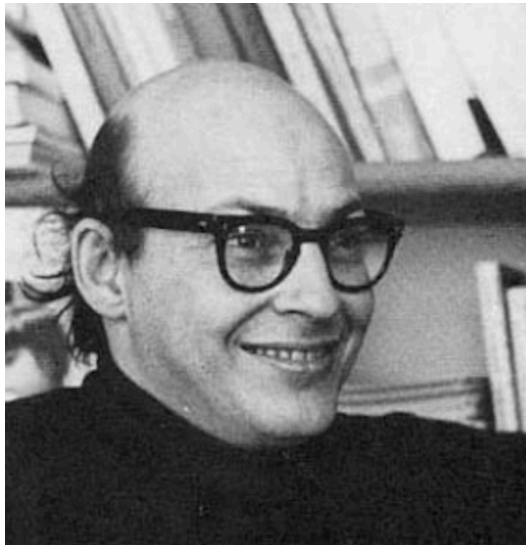
Fereshteh Sadeghi

fsadeghi@cs.washington.edu

Many slides from Larry Zitnick and Alyosha Efros

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

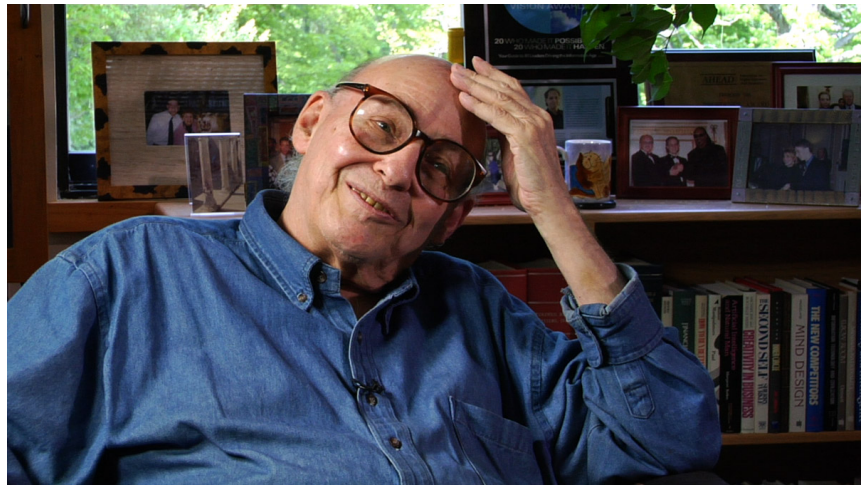
Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

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".

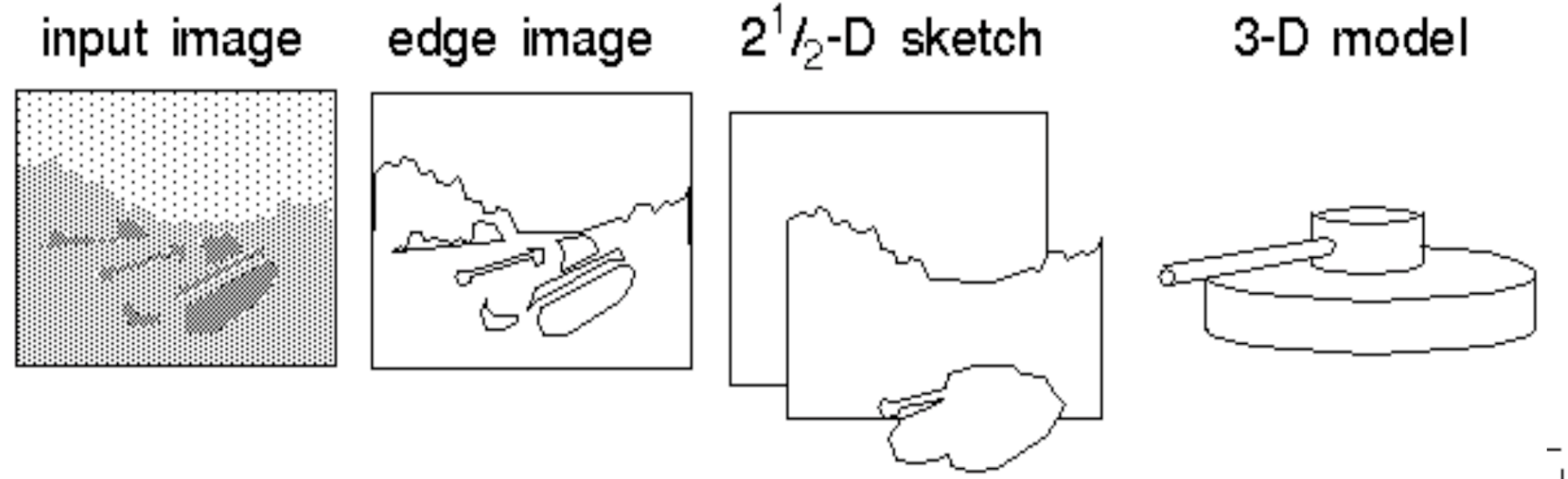
How hard is computer vision?



Marvin Minsky
Turing award, 1969

Gerald Sussman

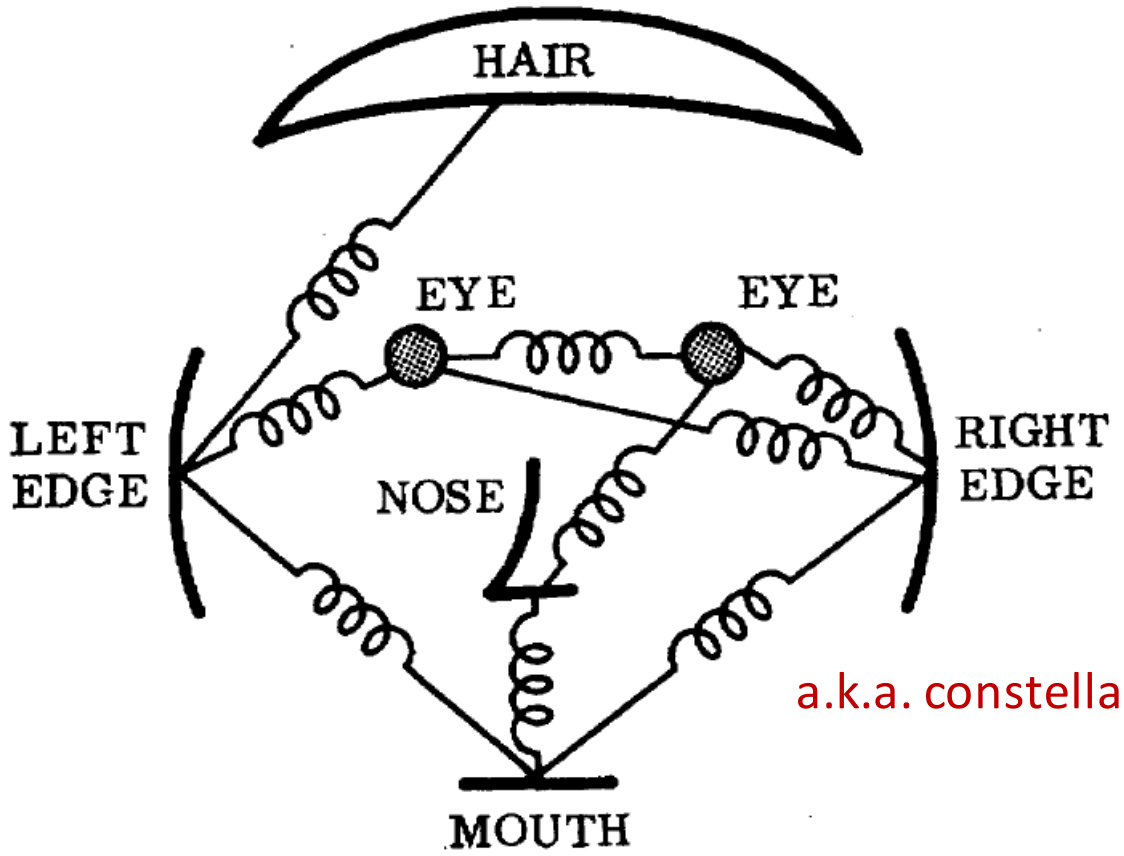
"You'll notice that Sussman never worked in vision again"
-Berthold Horn



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

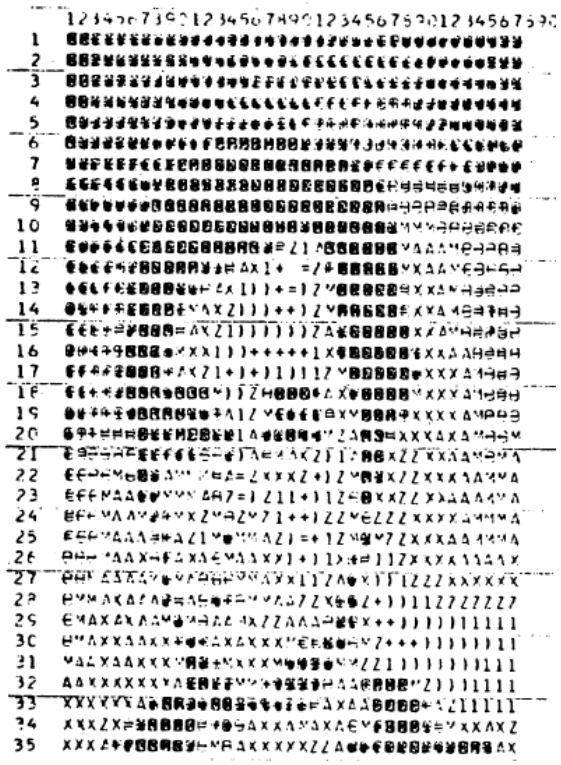
1973

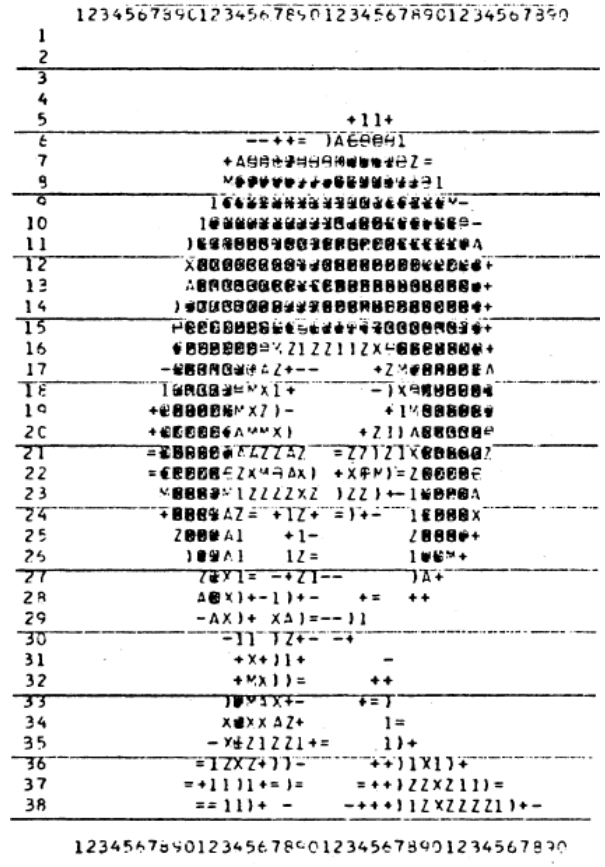


a.k.a. constellation model

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

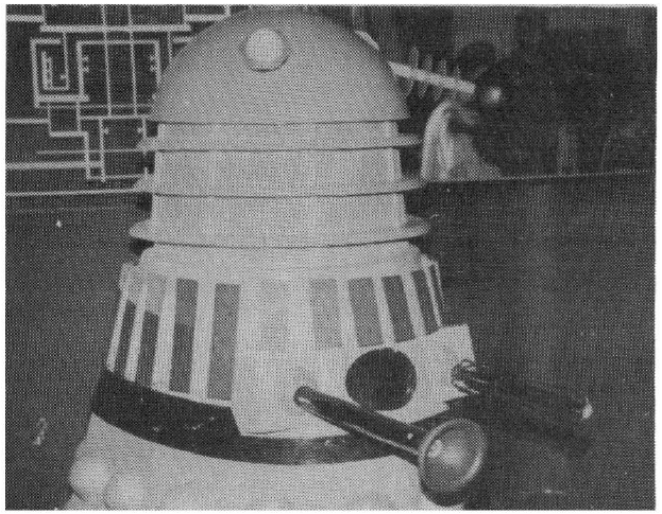
1973

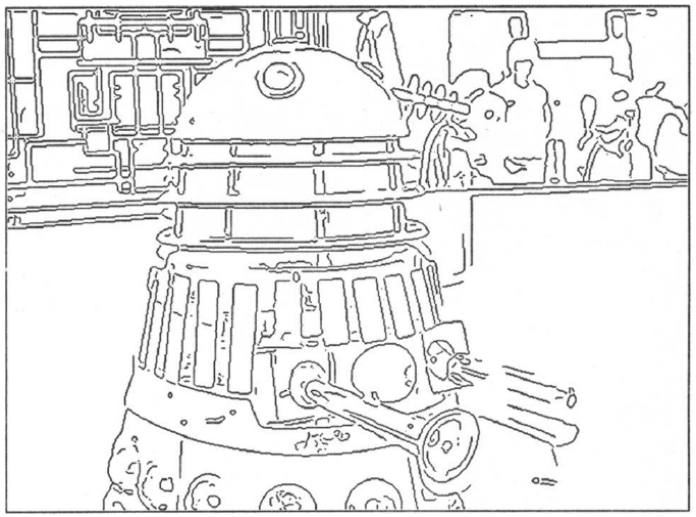




1980’s

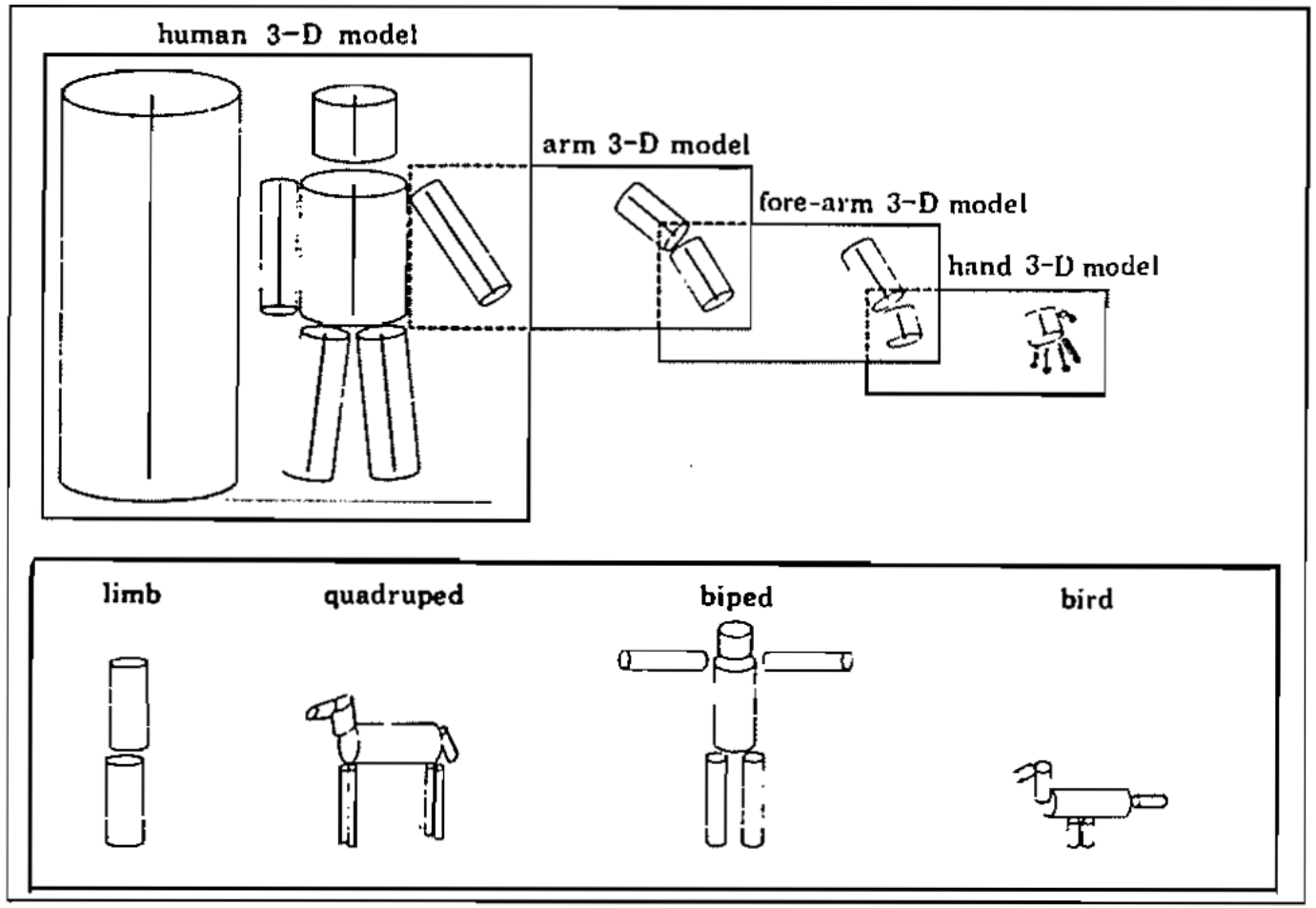
AI winter… …back to basics





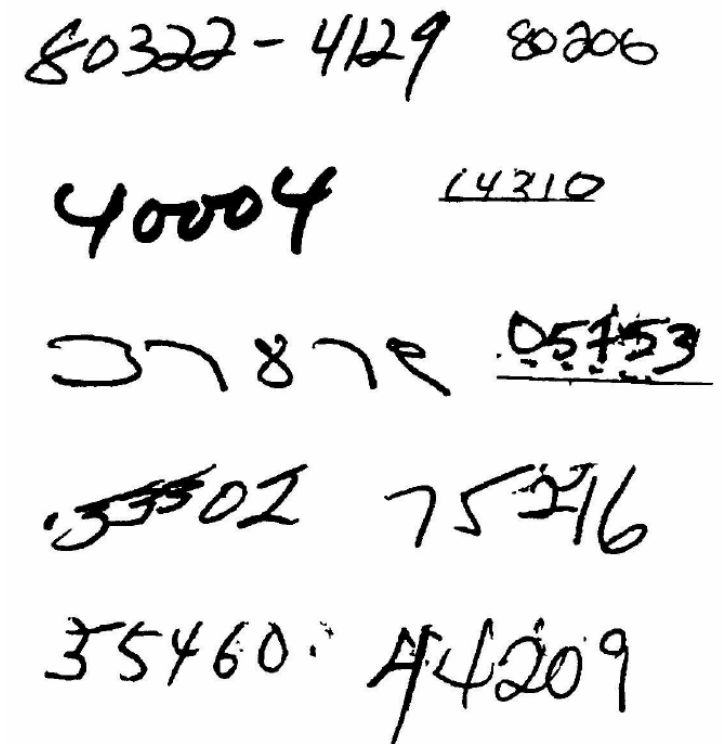
A Computational Approach to Edge Detection, Canny 1986

1986



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

1989



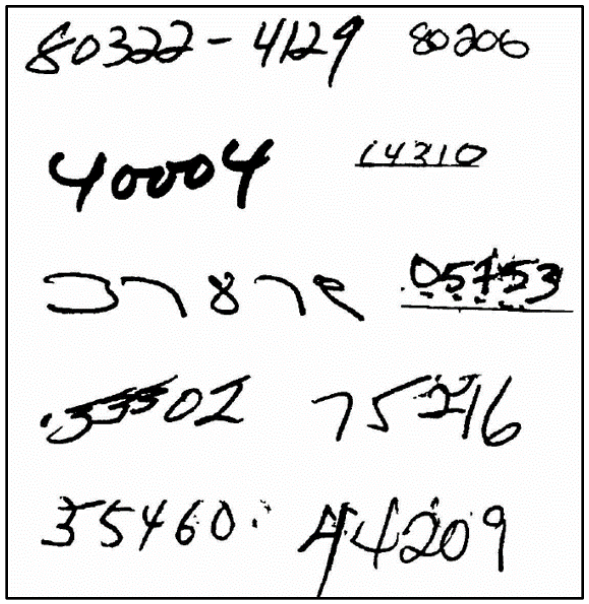
Zip codes

MNIST

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

Filters

Input

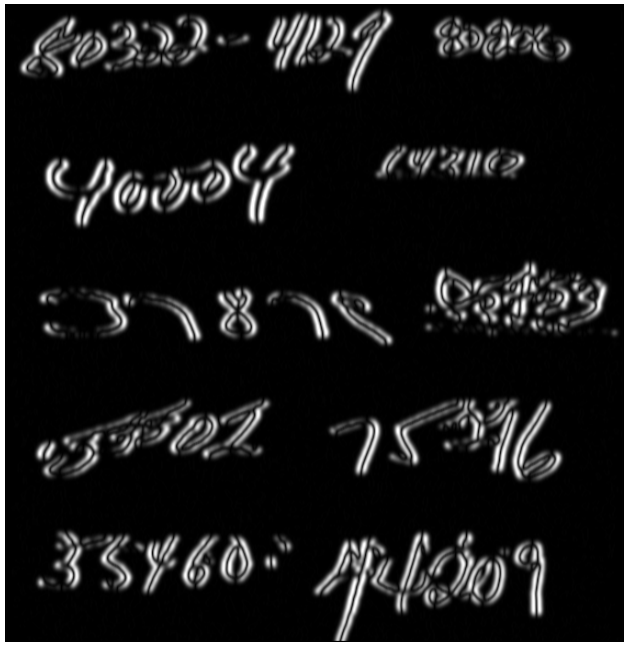


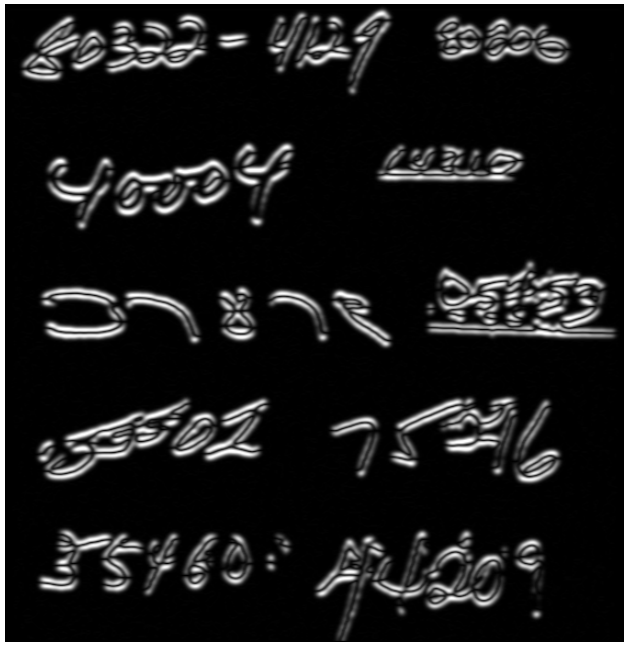
-1	0	+1
-2	0	+2
-1	0	+1

x filter

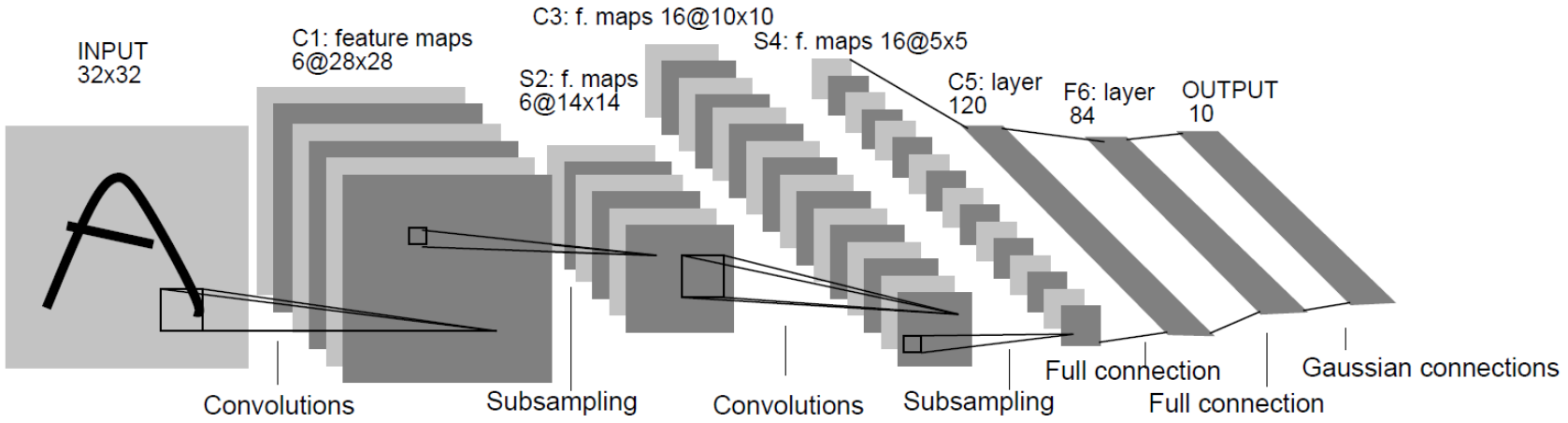
+1	+2	+1
0	0	0
-1	-2	-1

y filter





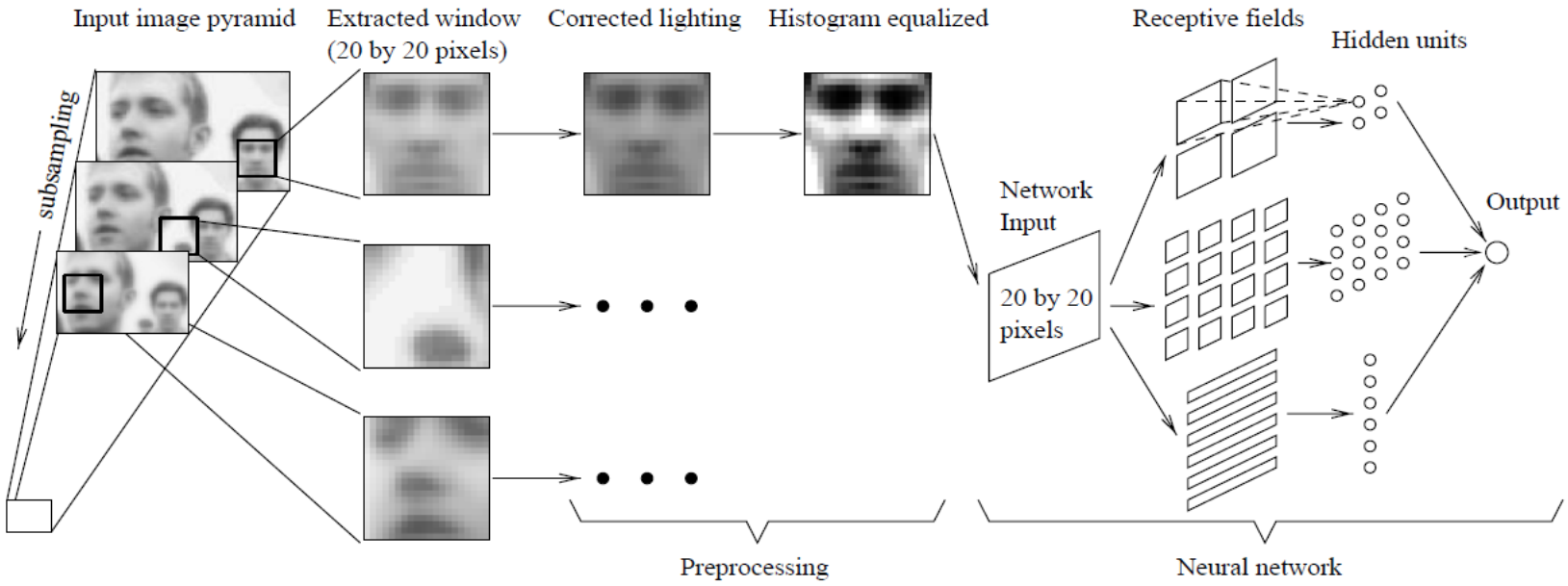
1989



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

1998

Faces

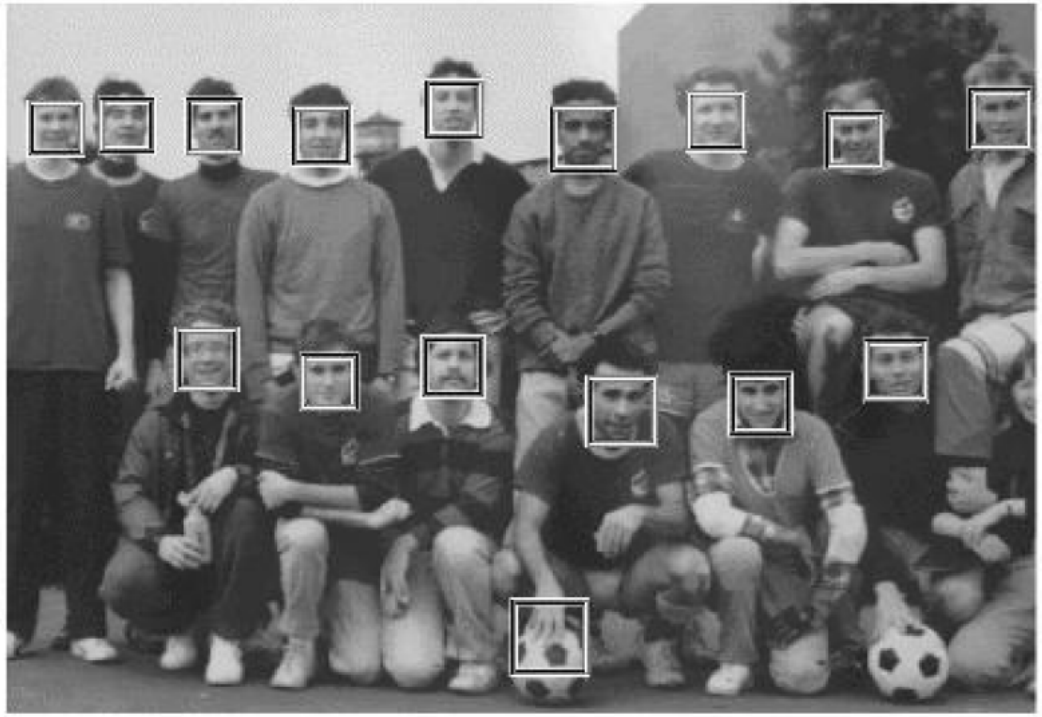


Neural Network-Based Face Detection, Rowley at al., PAMI 1998

2001

Sliding window in real time!

Boosting + Cascade = Speed



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

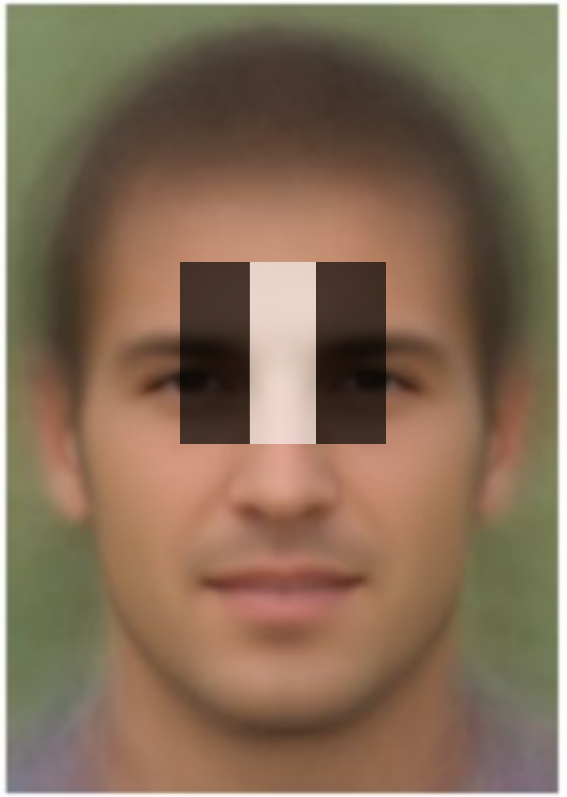
Why did it work?

- Simple features (Haar wavelets)

Integral images + Haar wavelets = fast

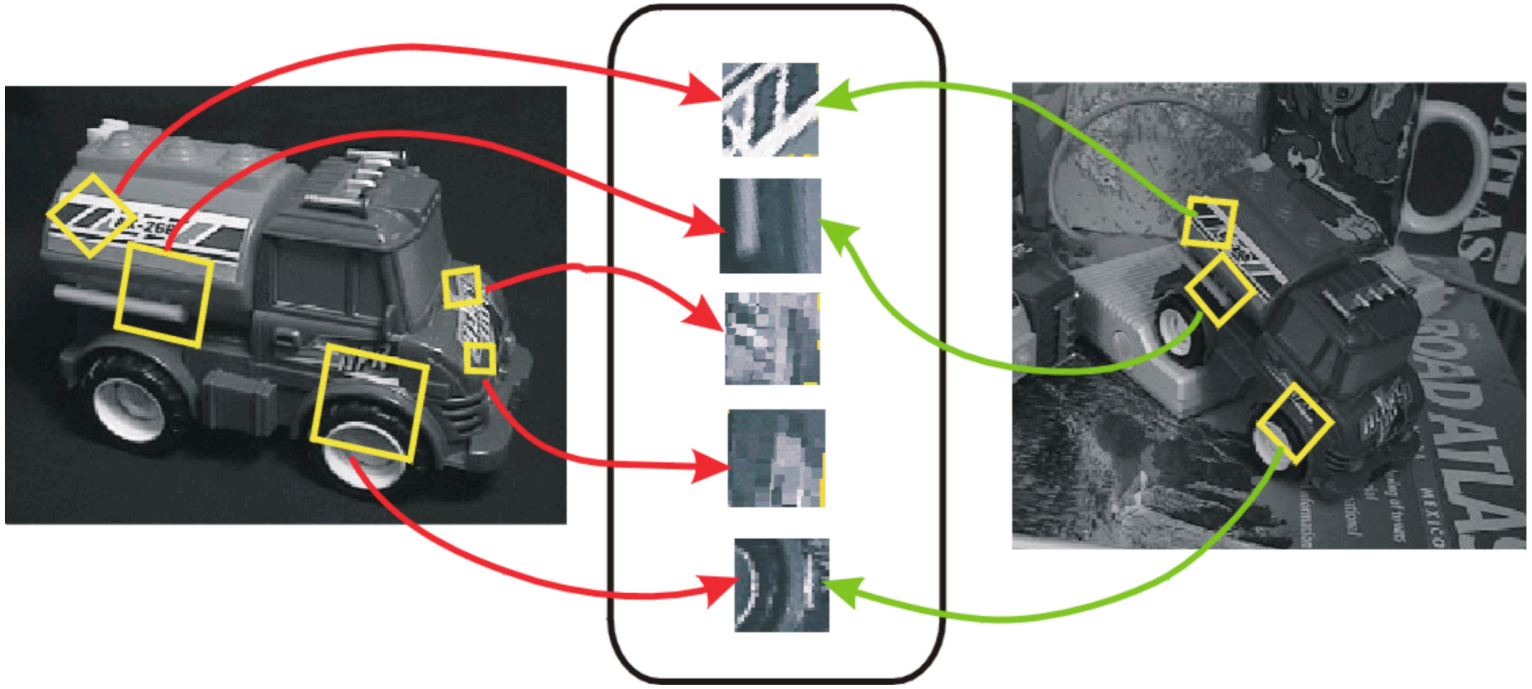
Face Detection, Viola & Jones, 2001

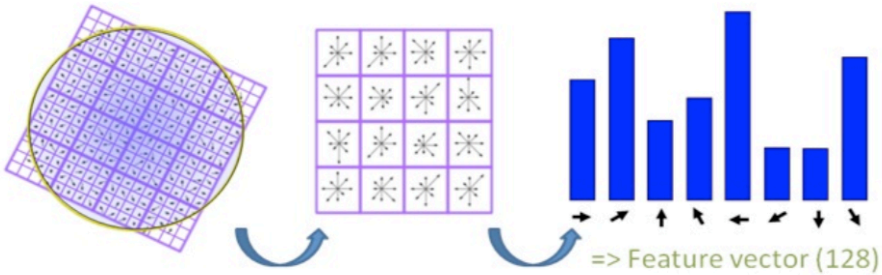
Why did it work?



Why did it fail?

1999* SIFT (Scale Invariant Feature Transform)



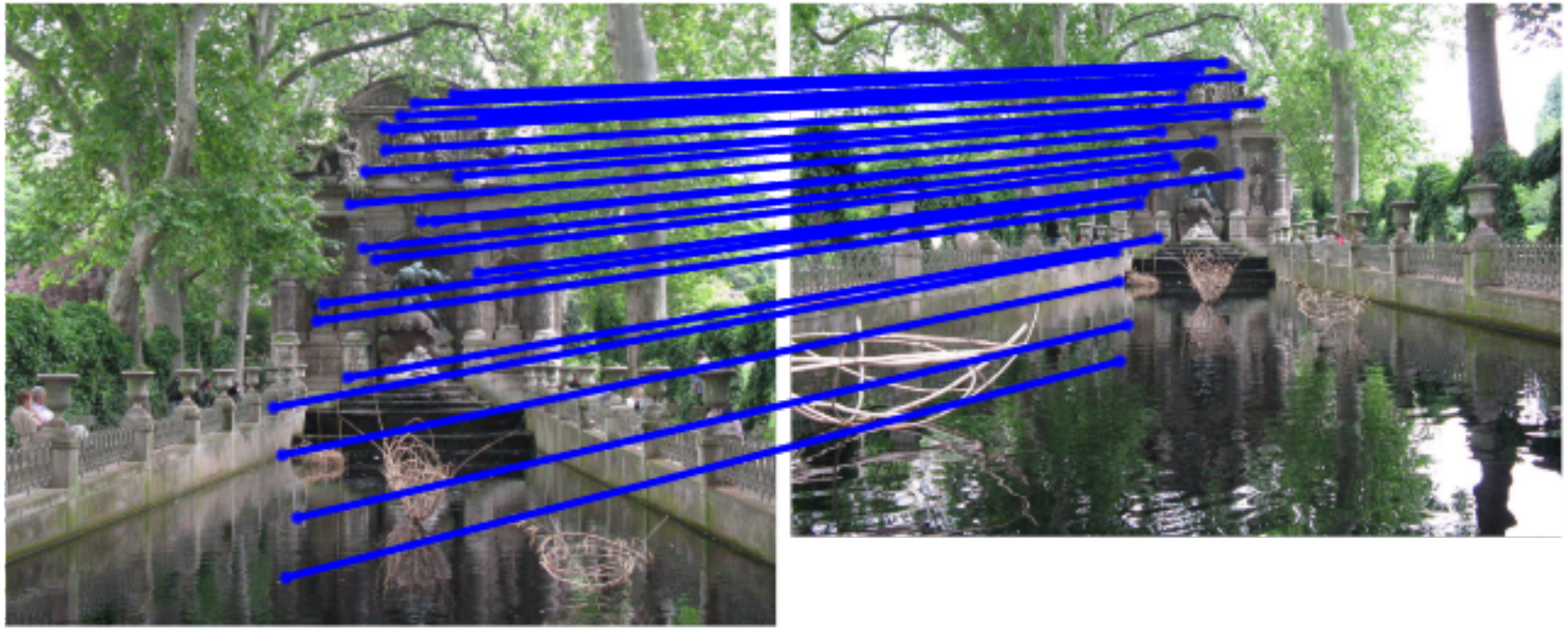


No more sliding windows (interest points)

Better features (use more computation)

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

SIFT Matching



[SIFT: Lowe, 2004]

What worked

Panorama stitching

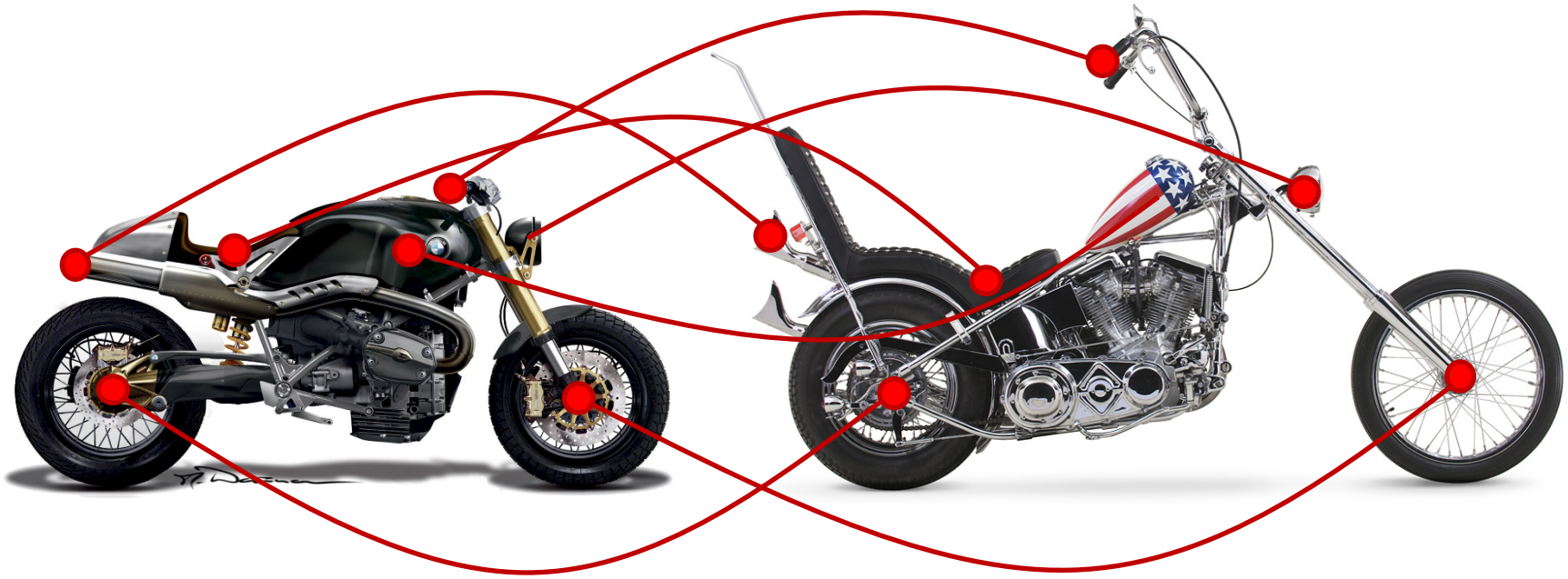
Recognizing panoramas, Brown and Lowe, *ICCV* 2003

SIFT Matching

[SIFT: Lowe, 2004]

2003

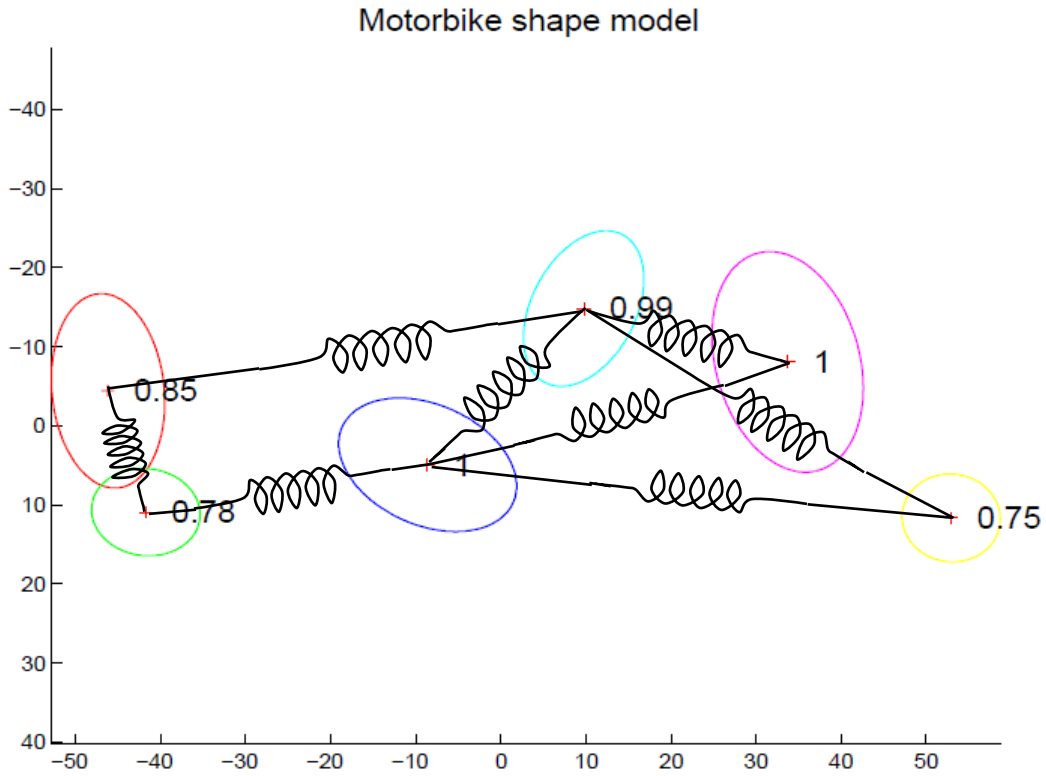
Constellation model (redux)



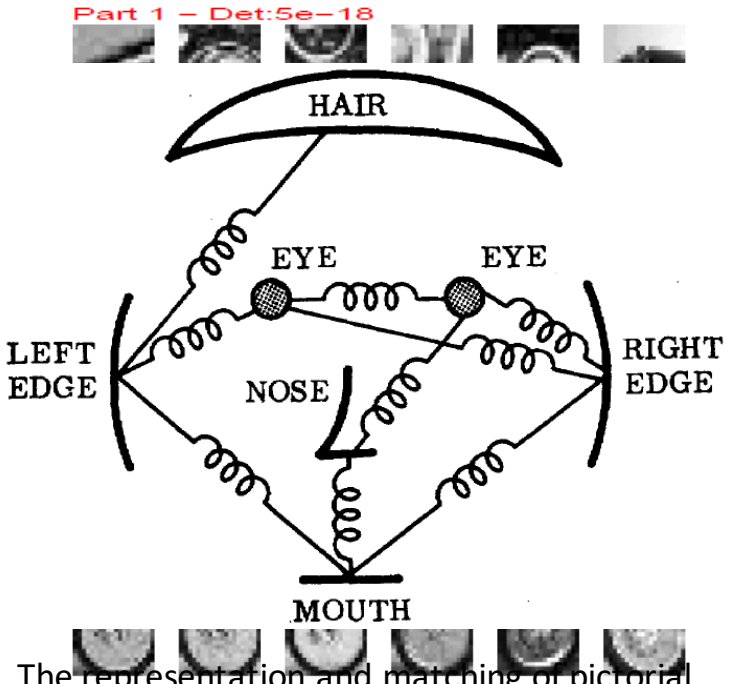
Object Class Recognition by Unsupervised Scale-Invariant Learning, Fergus et al., *CVPR* 2003.

2003

Constellation model (redux)

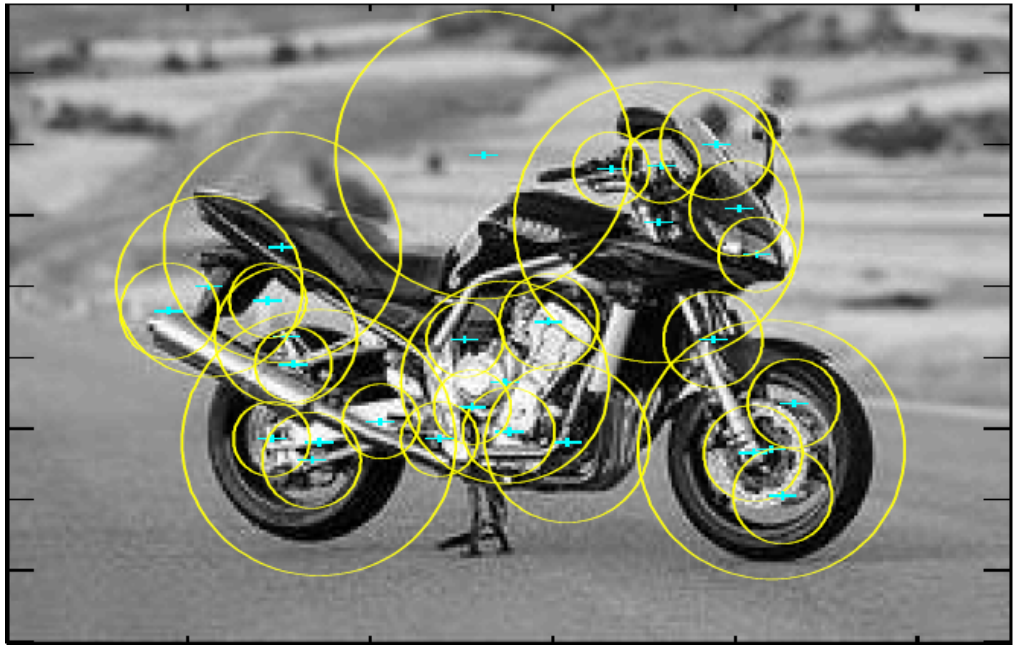


Joint Gaussian density



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

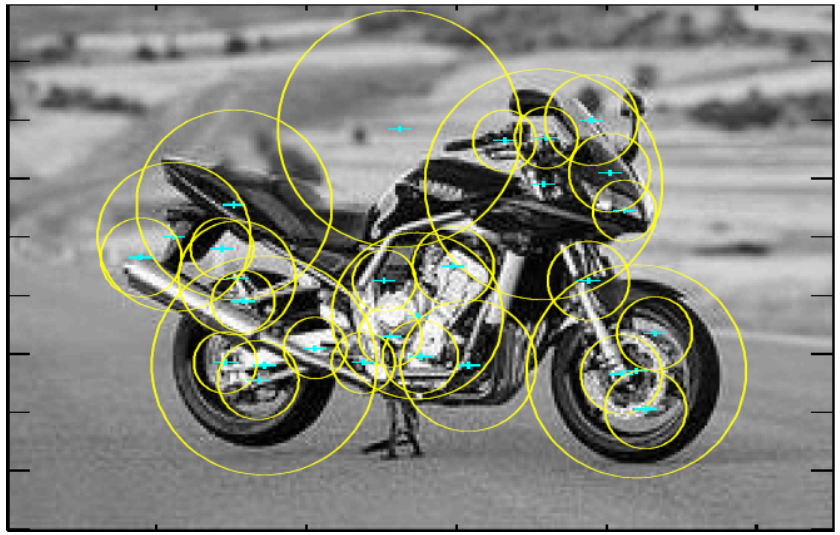
Interest points used to find parts:

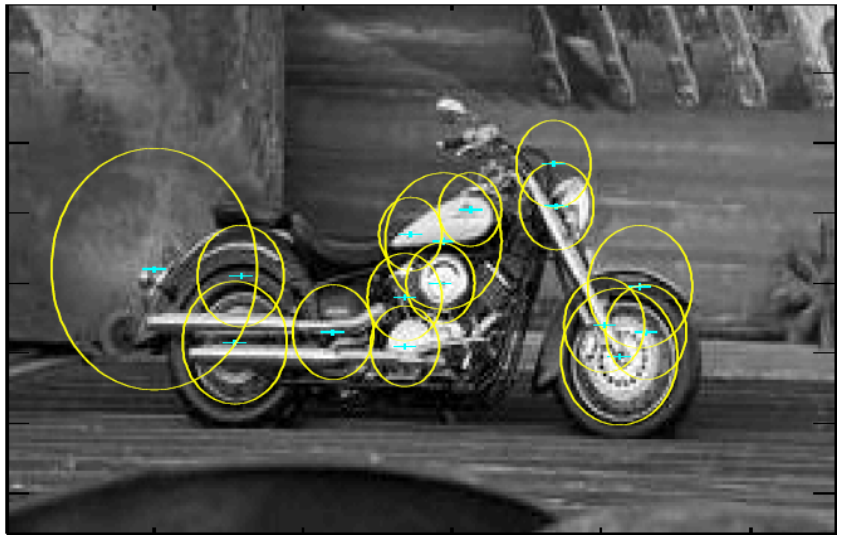


Smaller number of candidate parts allows for more complex spatial models.

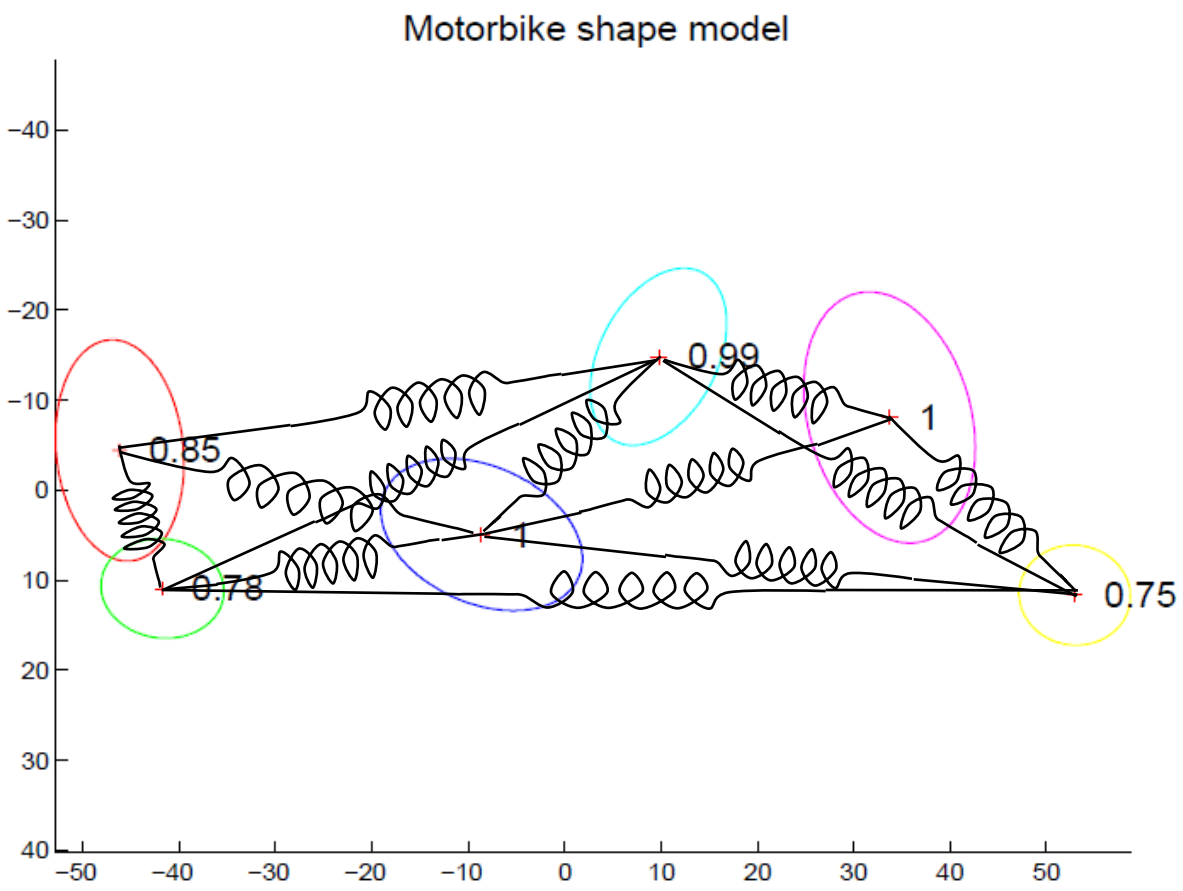
Why it fails

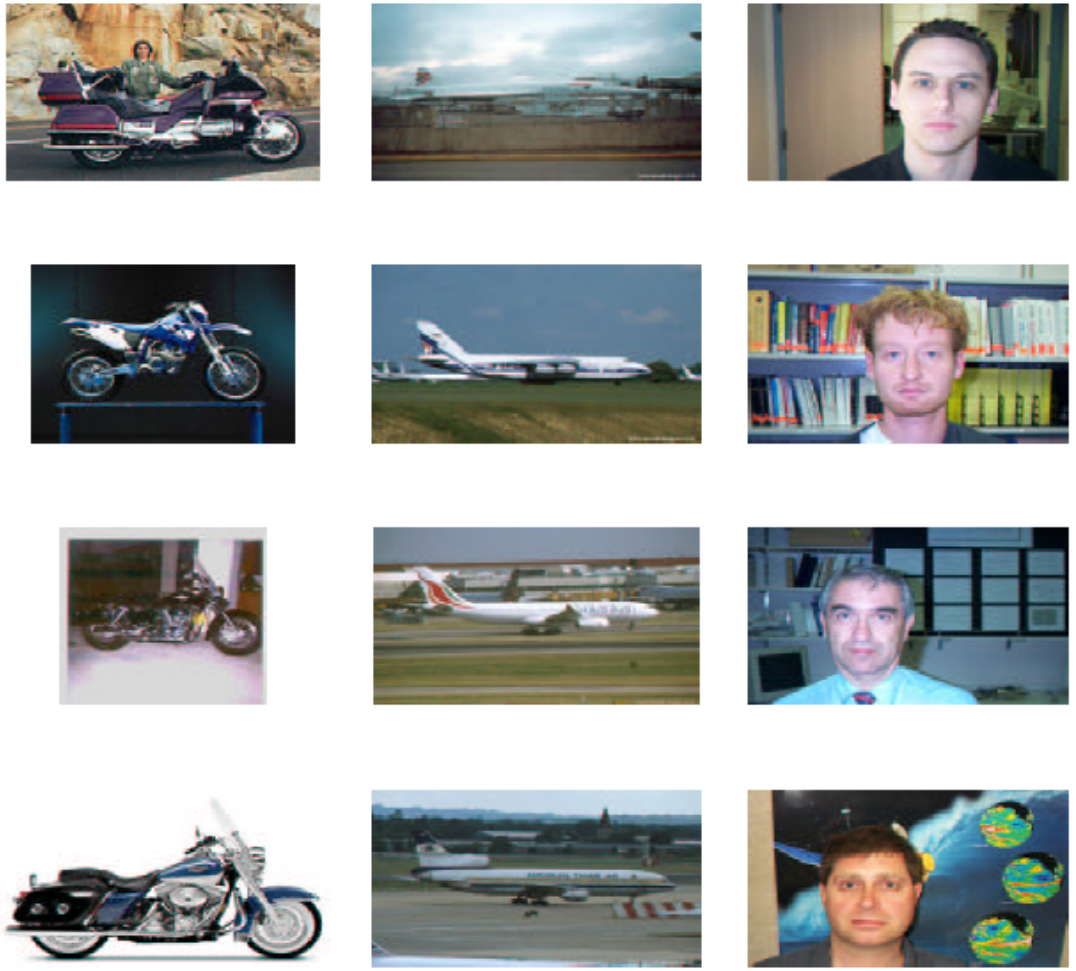
Interest points don’t work for category recognition





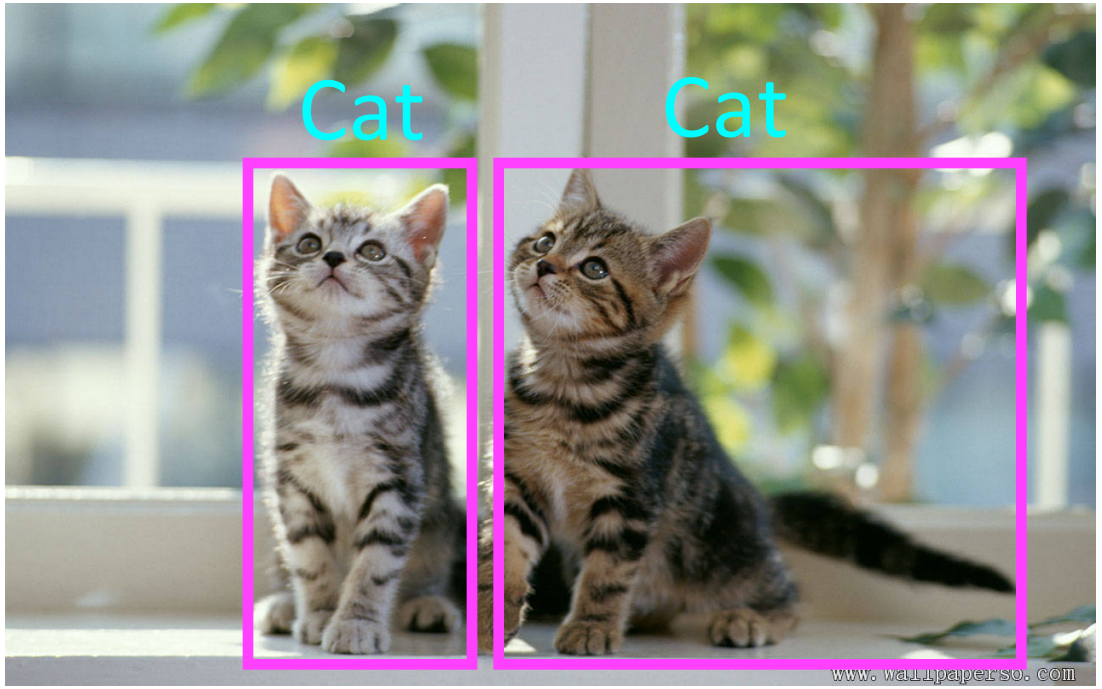
Too many springs…



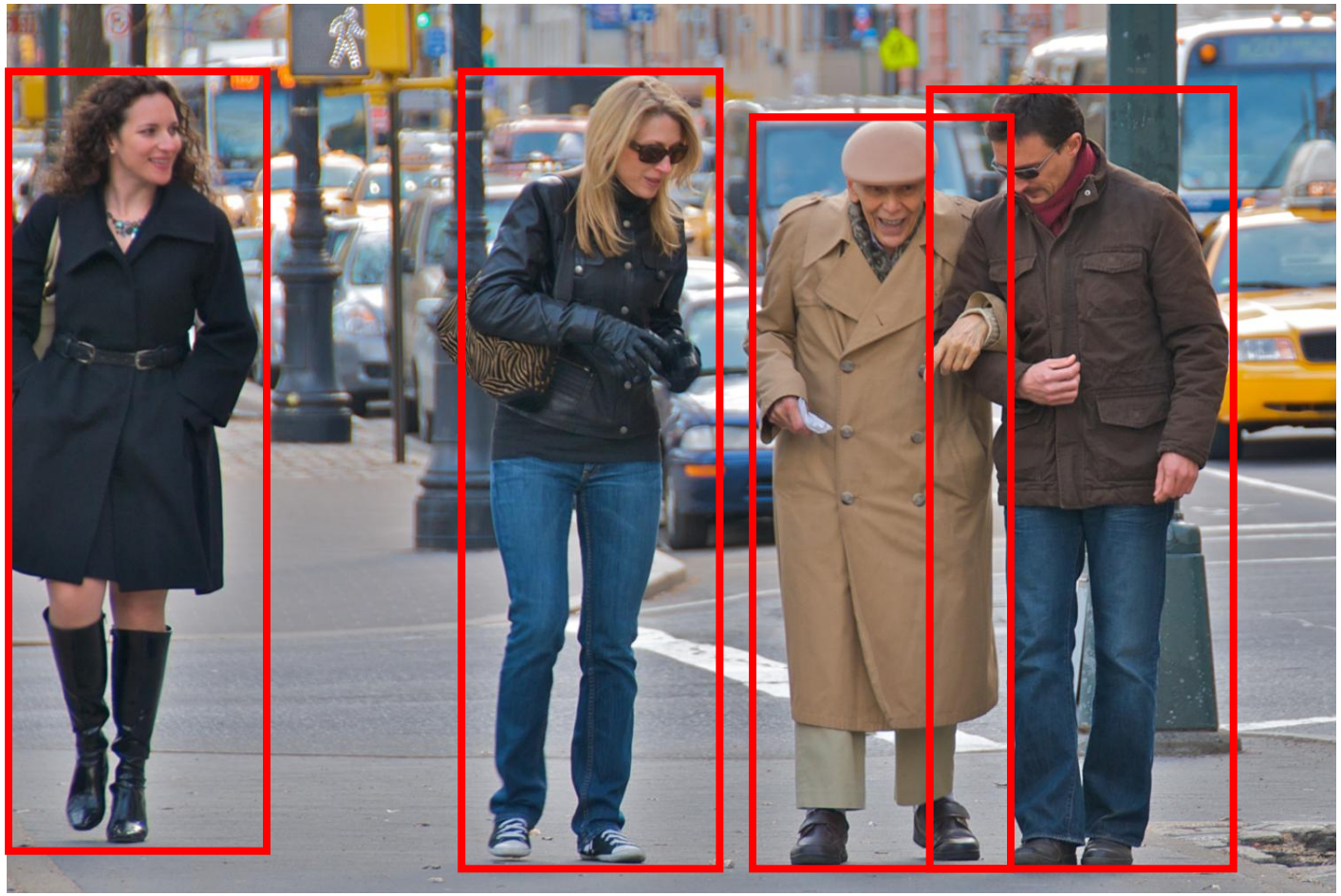


Cat?

Classification Vs. Detection



2005 HOG (histograms of oriented gradients)



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

Pedestrians

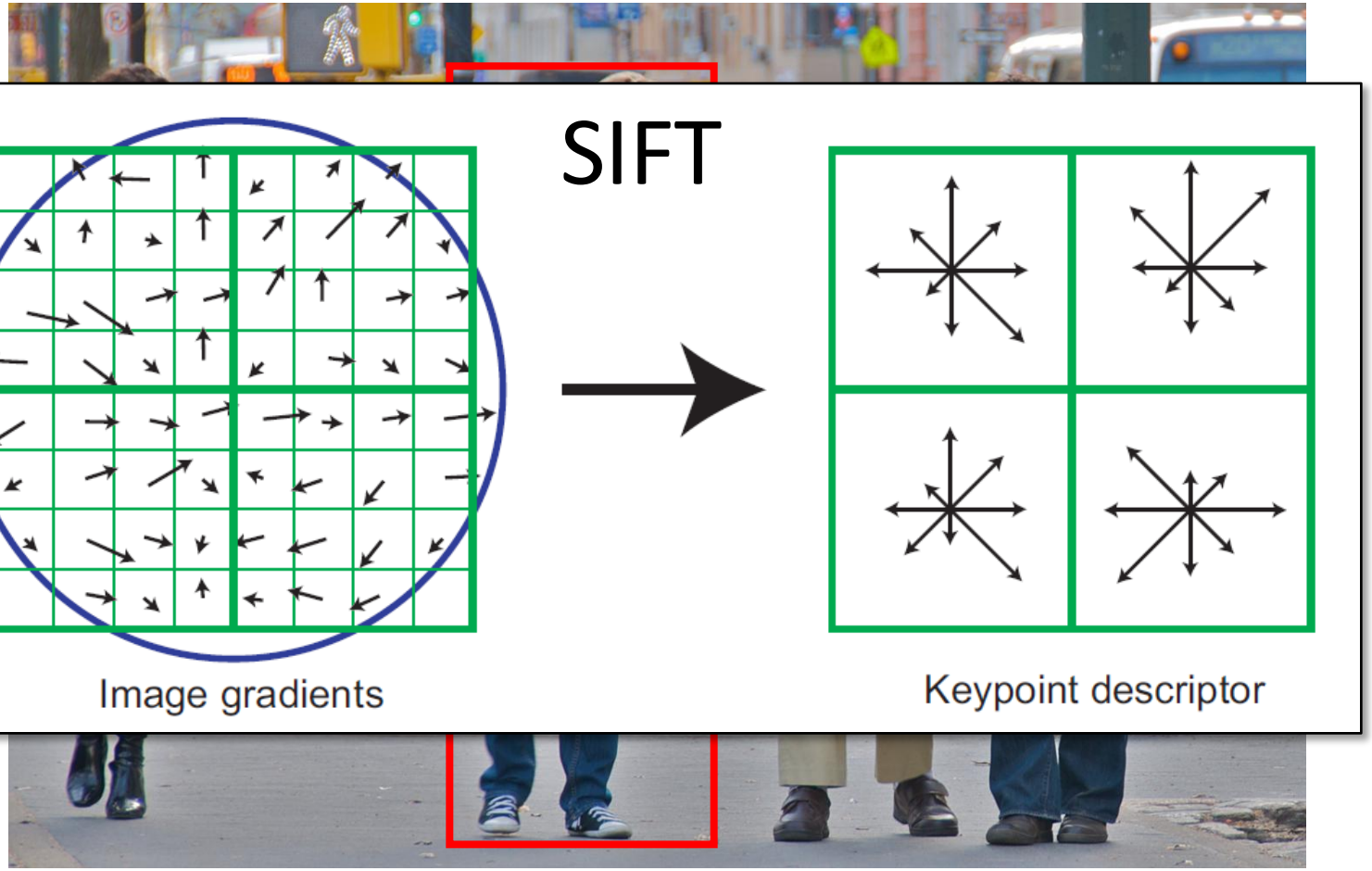
- Defined by their contours
- Cluttered backgrounds
- Significant variance in texture

Interest points won’t work…

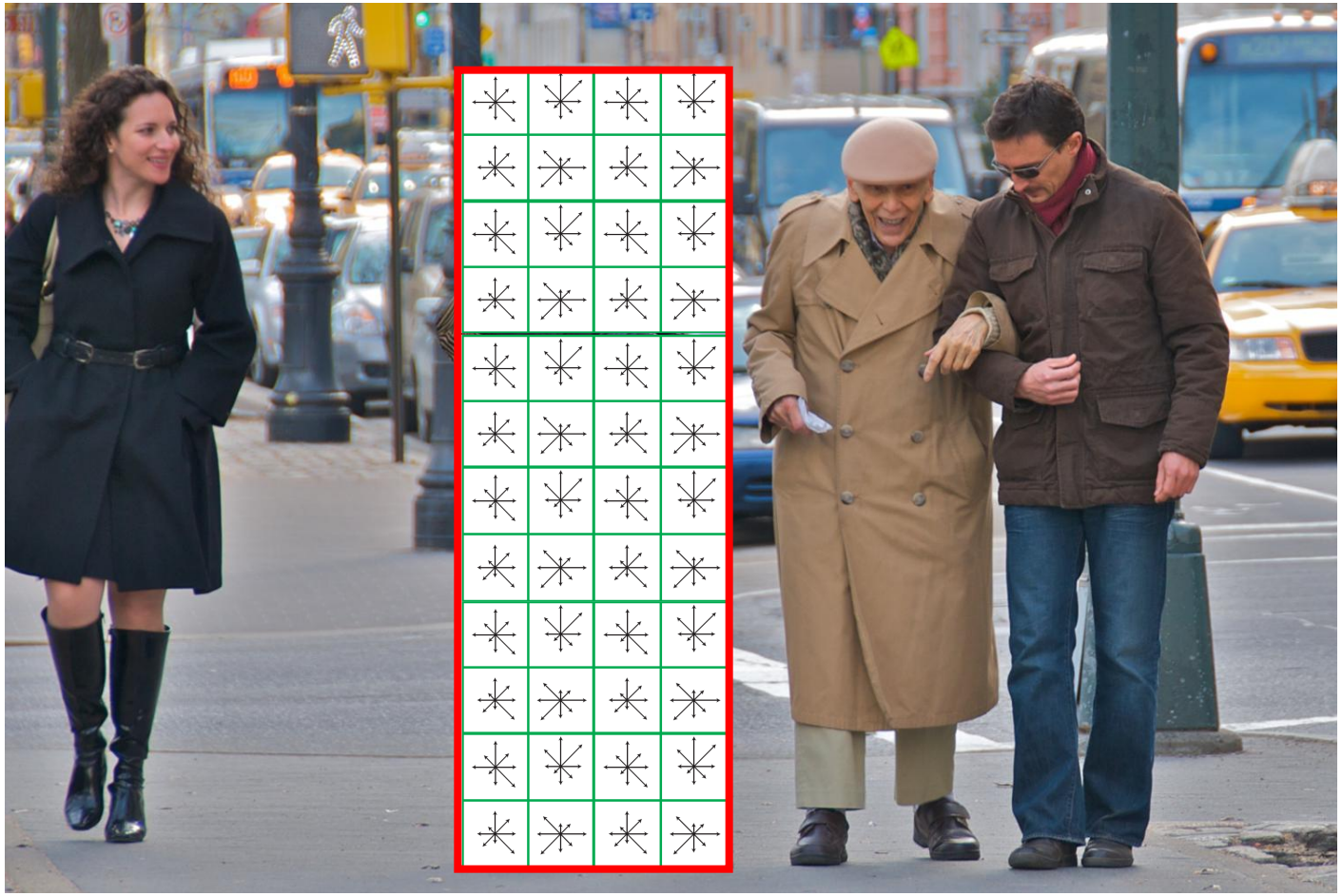
…back to sliding window.

2005 HOG (histograms of oriented gradients)

2005 HOG (histograms of oriented gradients)



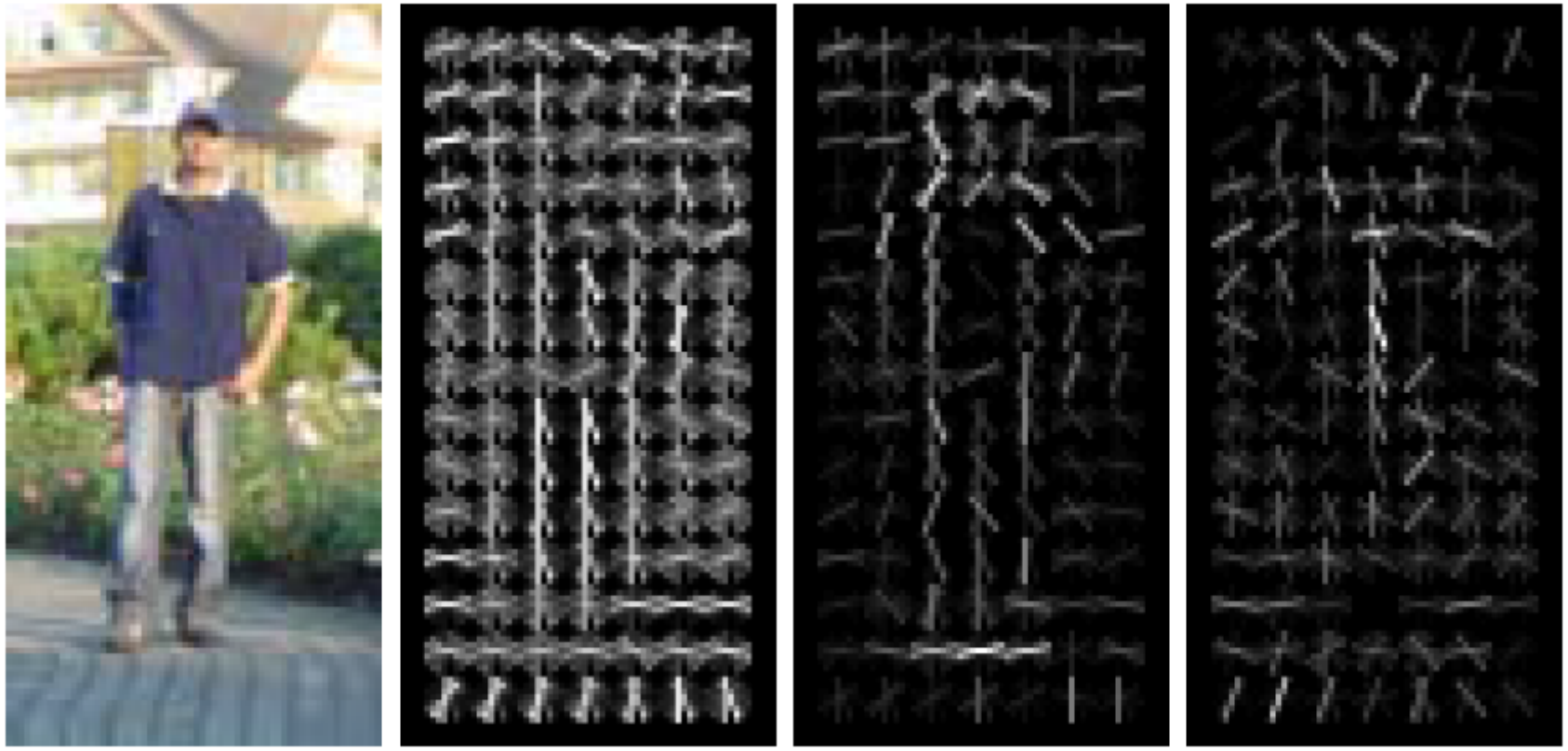
2005 HOG (histograms of oriented gradients)



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

Why it worked

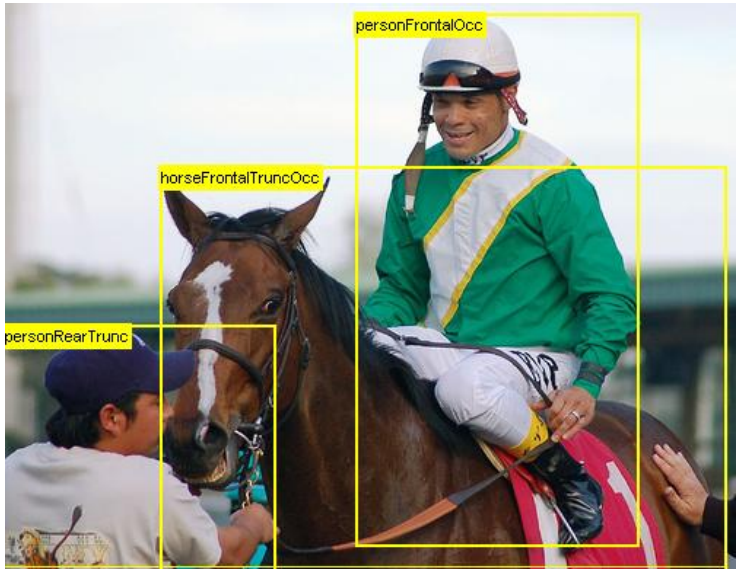
We can finally detect object boundaries in a reliable manner!

Hard negative mining

Computers are fast enough.

2007 PASCAL VOC

20 classes



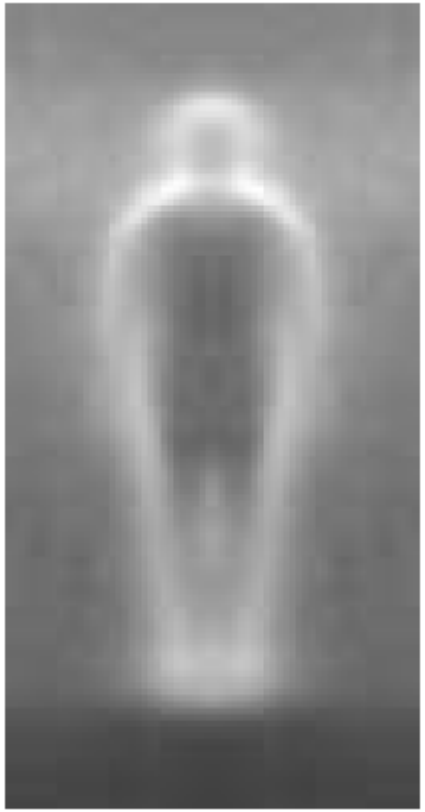
The PASCAL Visual Object Classes (VOC) Challenge, Everingham, Van Gool, Williams, Winn and Zisserman, *IJCV*, 2010

2009 ImageNet

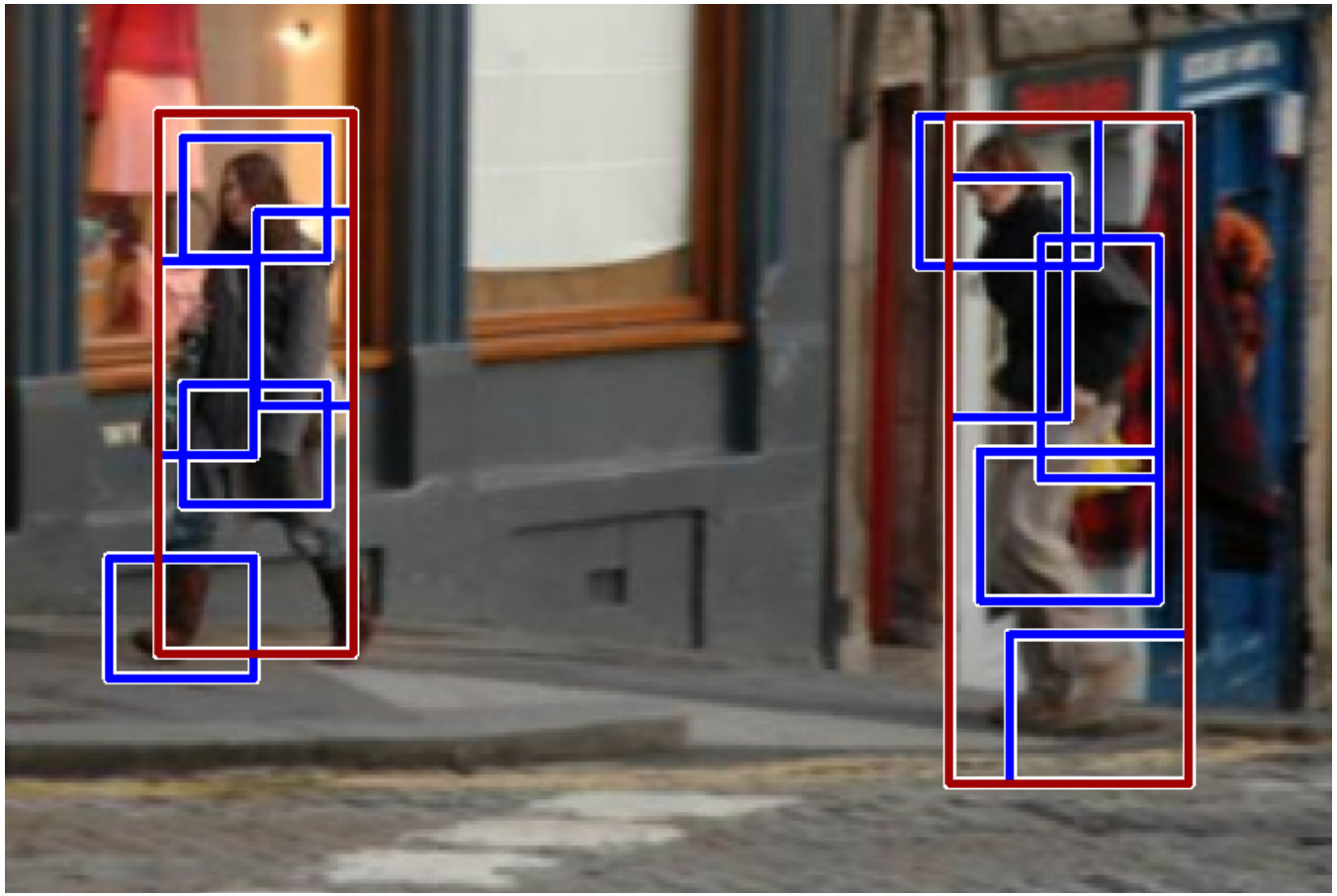
22K categories, 14M images

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

Why it failed

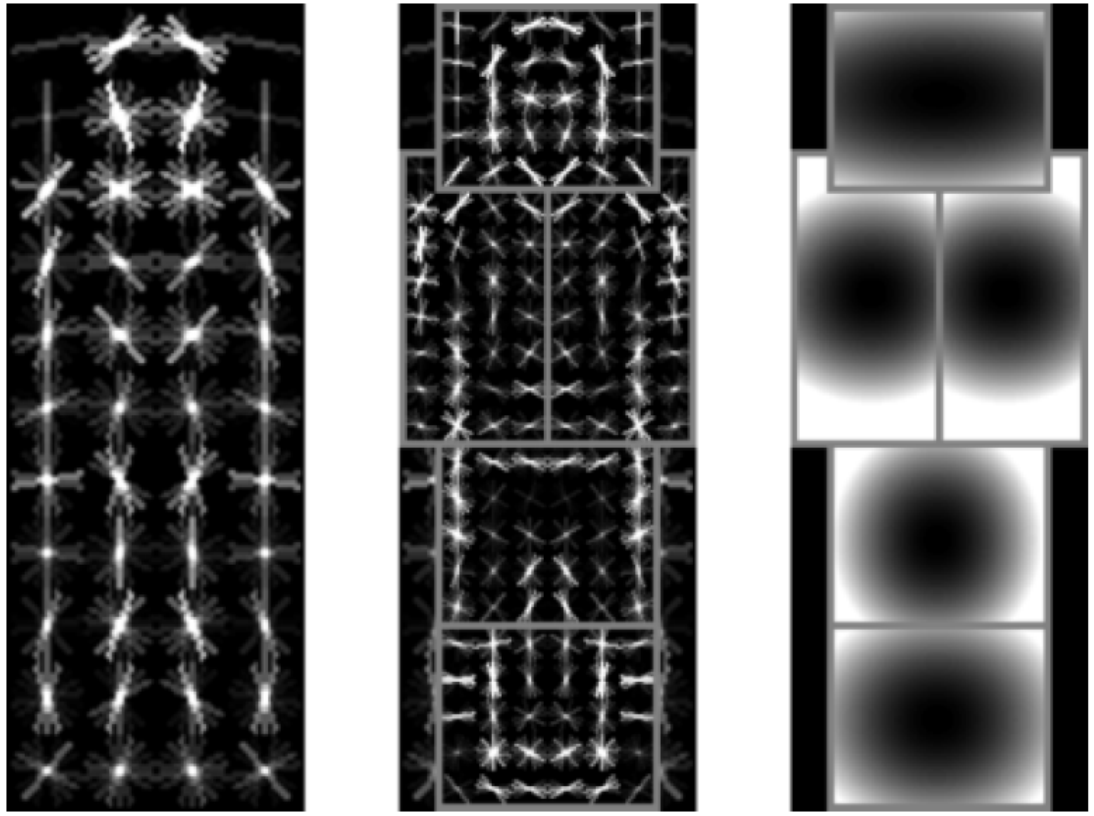


2008 DPM (Deformable parts model)



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

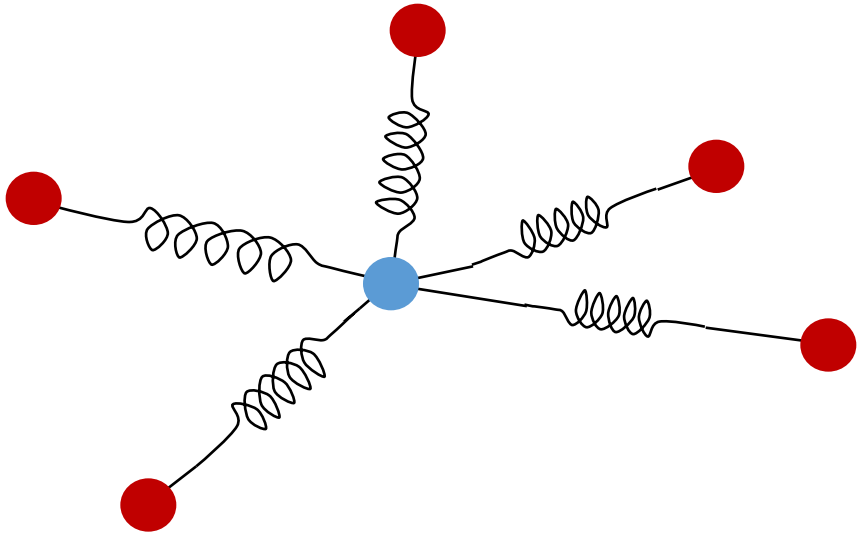
2008 DPM (Deformable parts model)



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

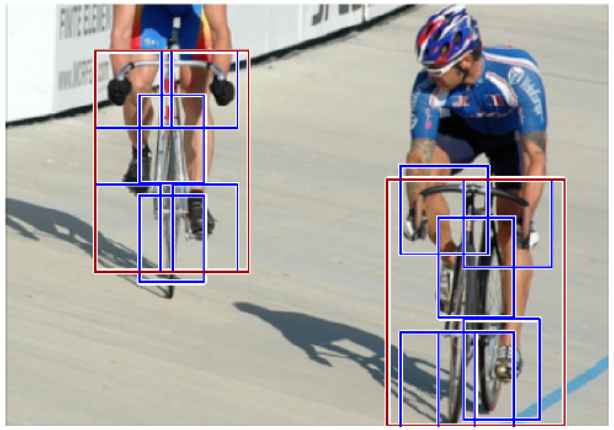
Star-structure

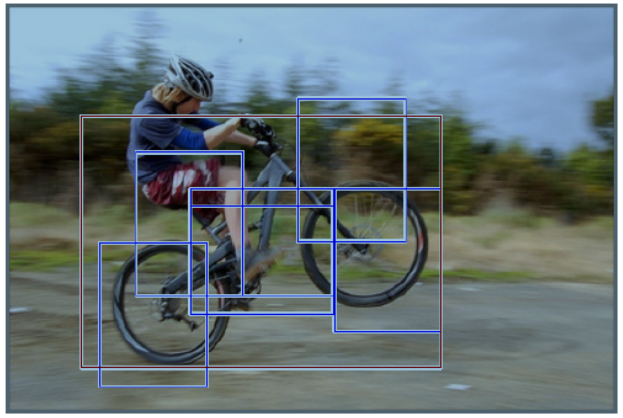
- Computationally efficient (distance transform)

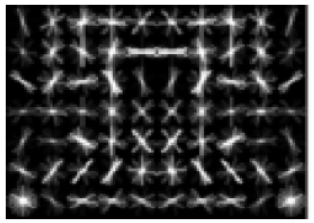


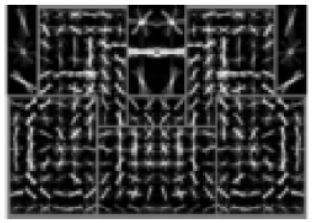
Distance transforms of sampled functions, Felzenszwalb and Huttenlocher, Cornell University CIS, Tech. Rep. 2004.

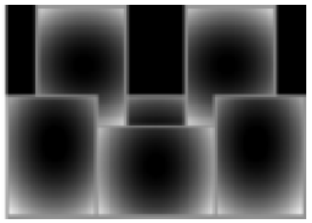
Multiple components

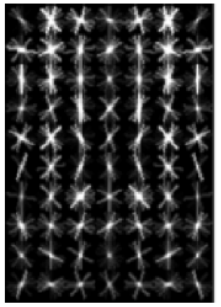


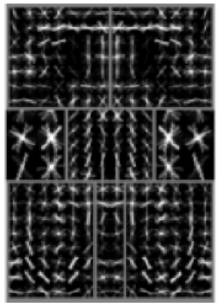


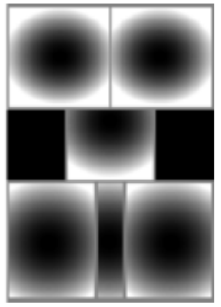








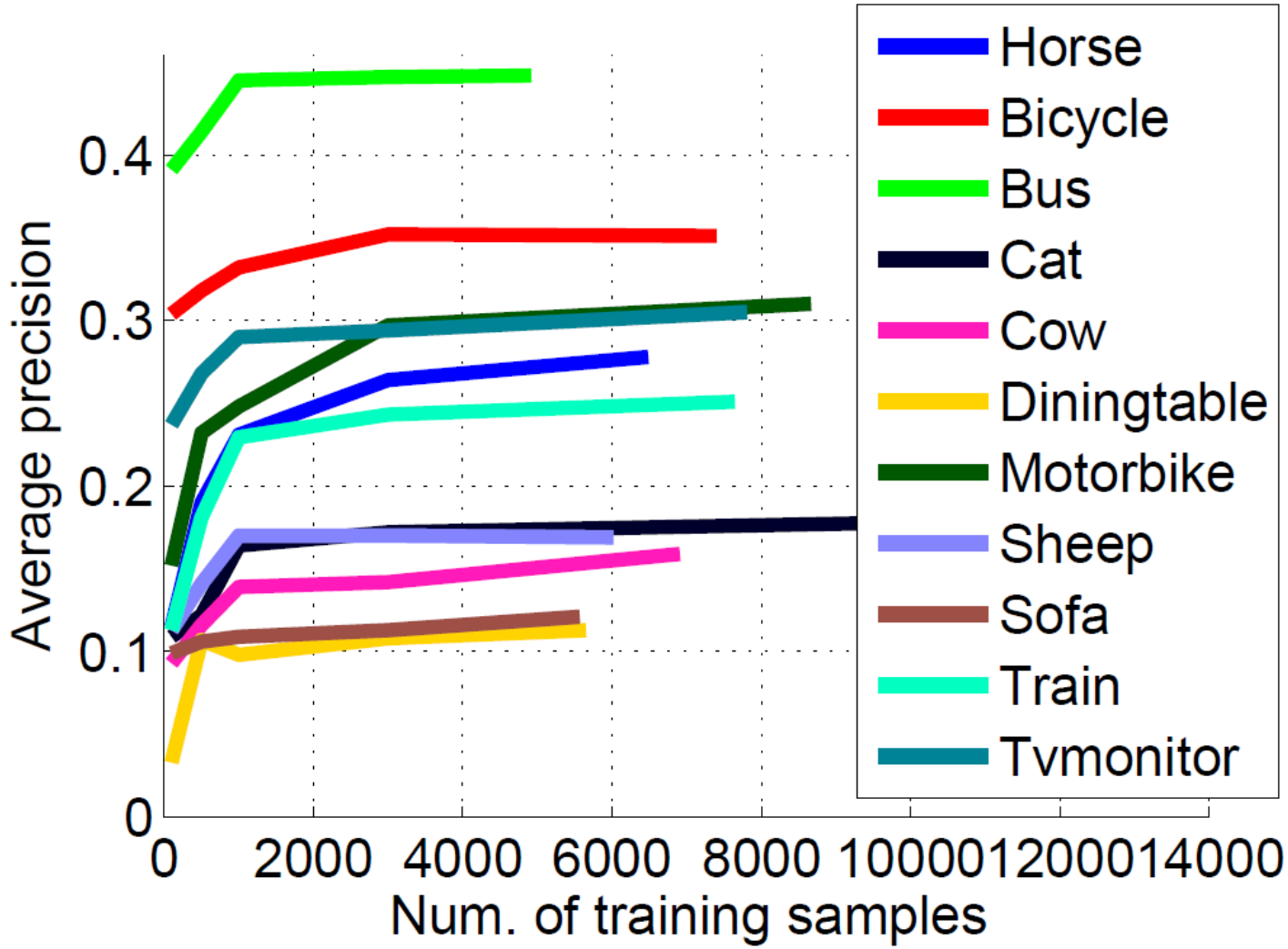




Why it worked

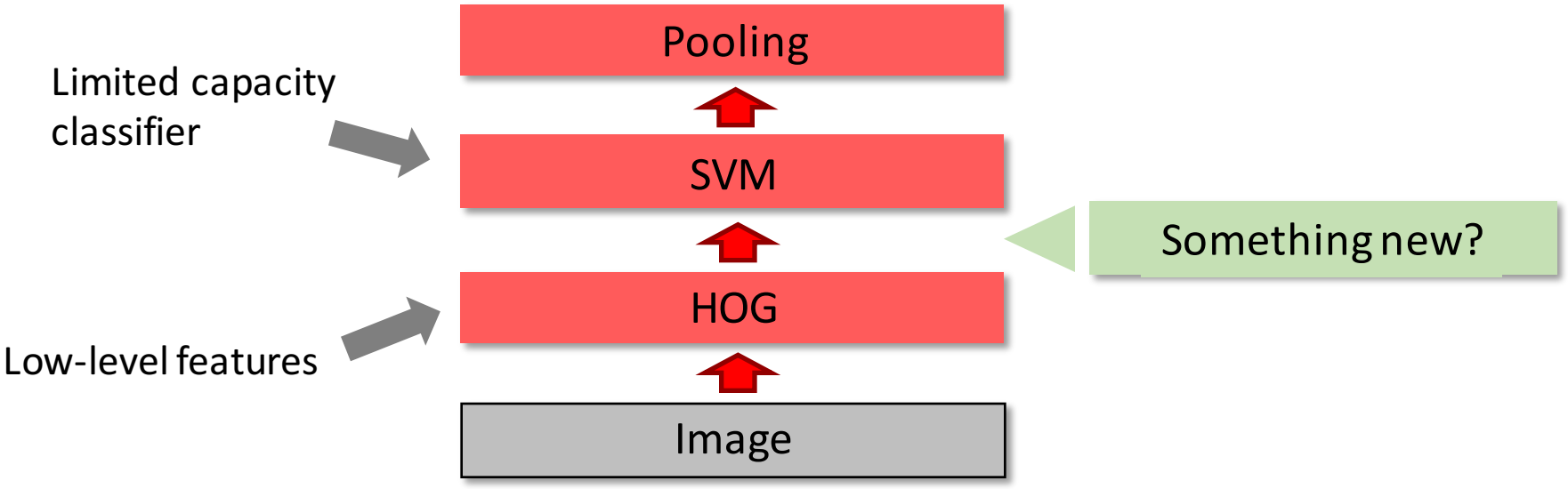
- Multiple components
- Deformable parts?
- Hard negative mining
- Good balance

"How important are 'Deformable Parts' in the Deformable Parts Model?“,
Divvala, Efros, and Hebert, *Parts and Attributes Workshop, ECCV,* 2012



Do We Need More Training Data or Better Models for Object Detection?
Zhu, Vondrick, Ramanan, Fowlkes, *BMVC* 2012.

DPM


Pooling
Limited capacity classifier
SVM
Something new?
HOG
Low-level features
Image

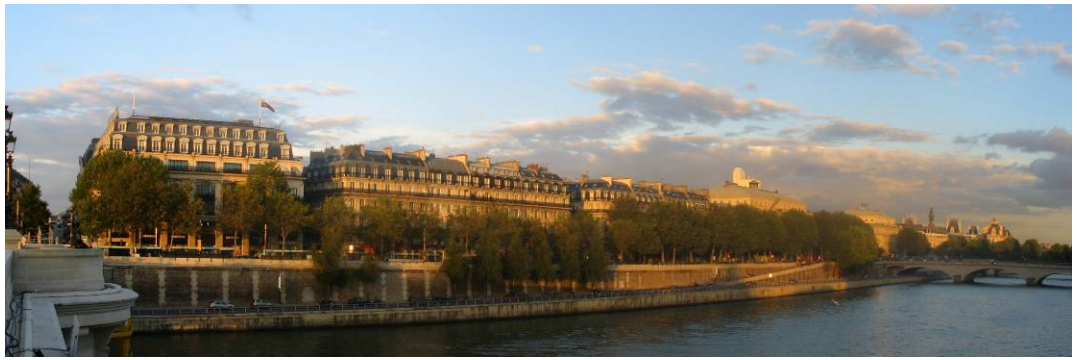
Problems with Visual Categories

- A lot of categories are functional

Char



- World is too varied

- Categories are 3D, but images are 2D

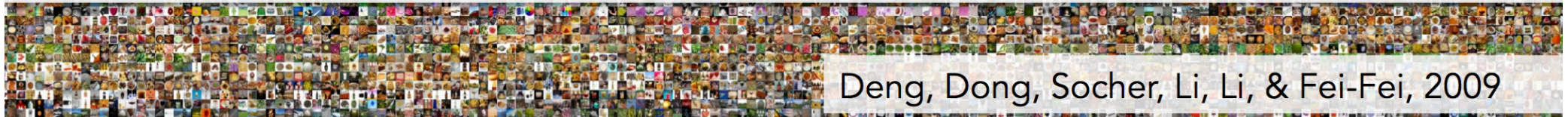
car


IMAGENET

www.image-net.org

22K categories and **14M** images

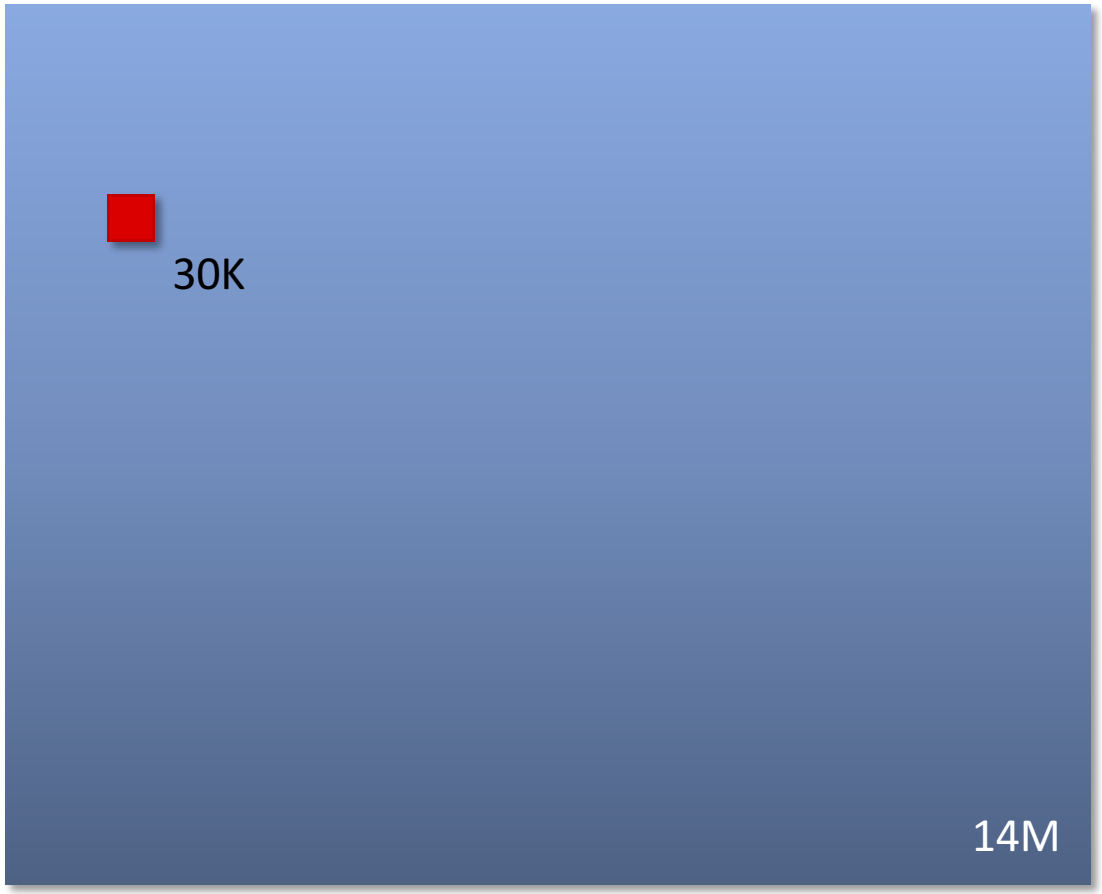
- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities



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

Images

2009 2012

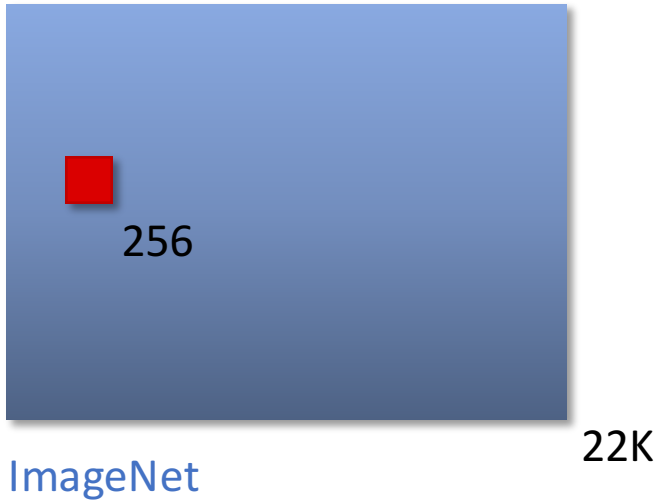


ImageNet

Categories

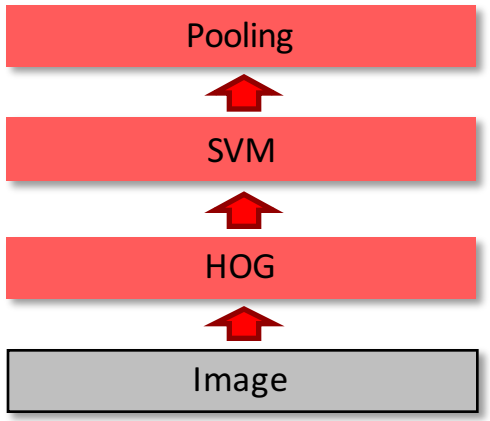
2009

2012

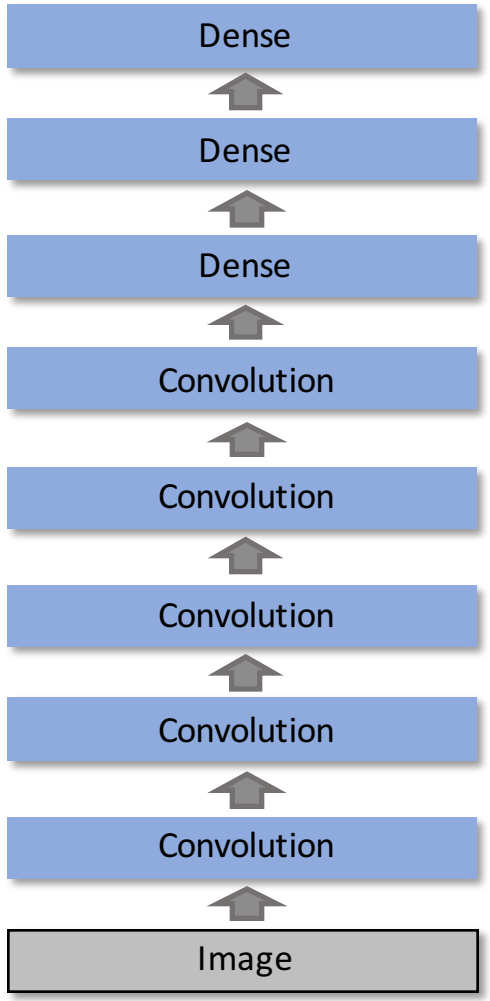


Algorithms

2009

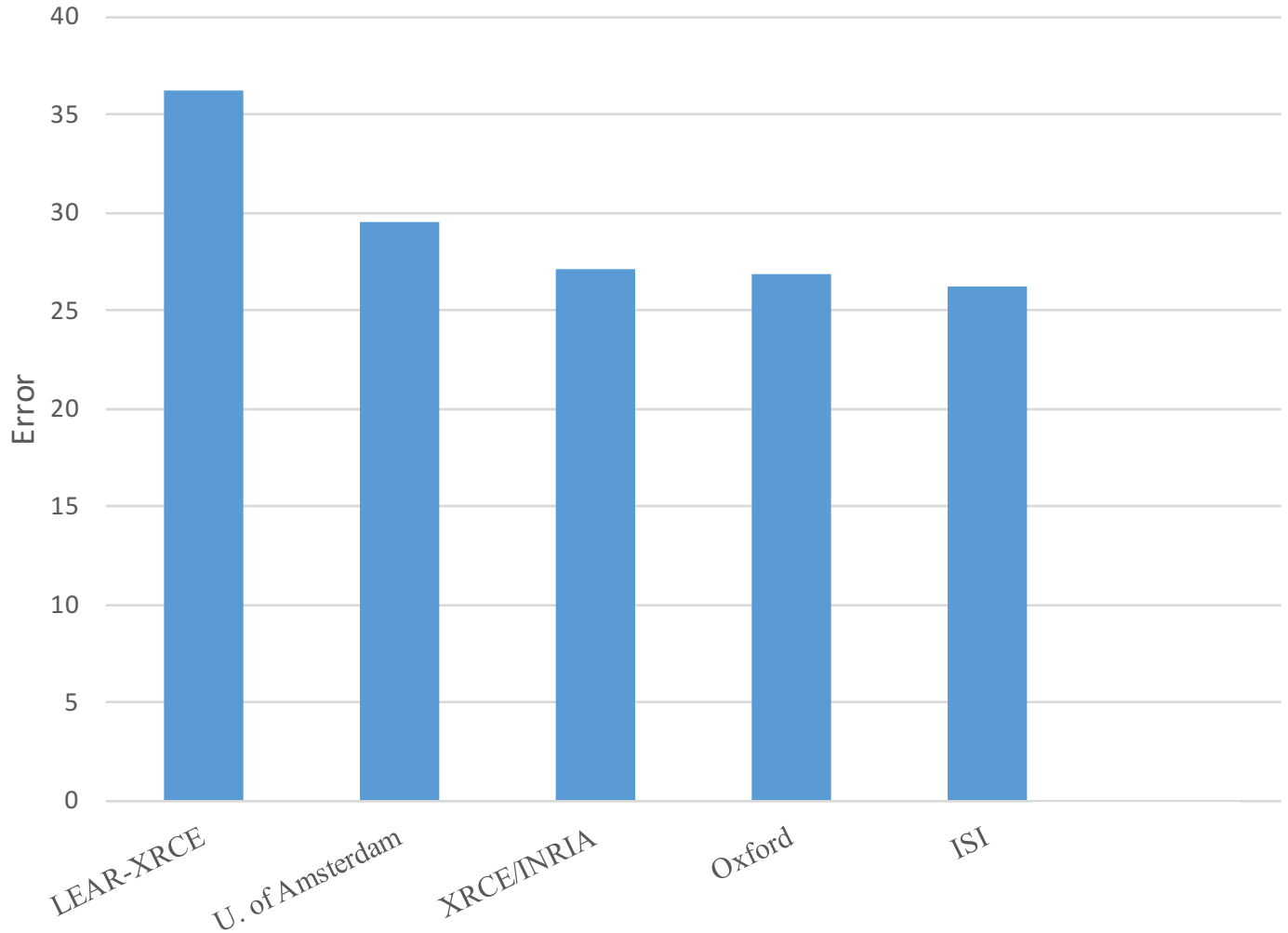


2012



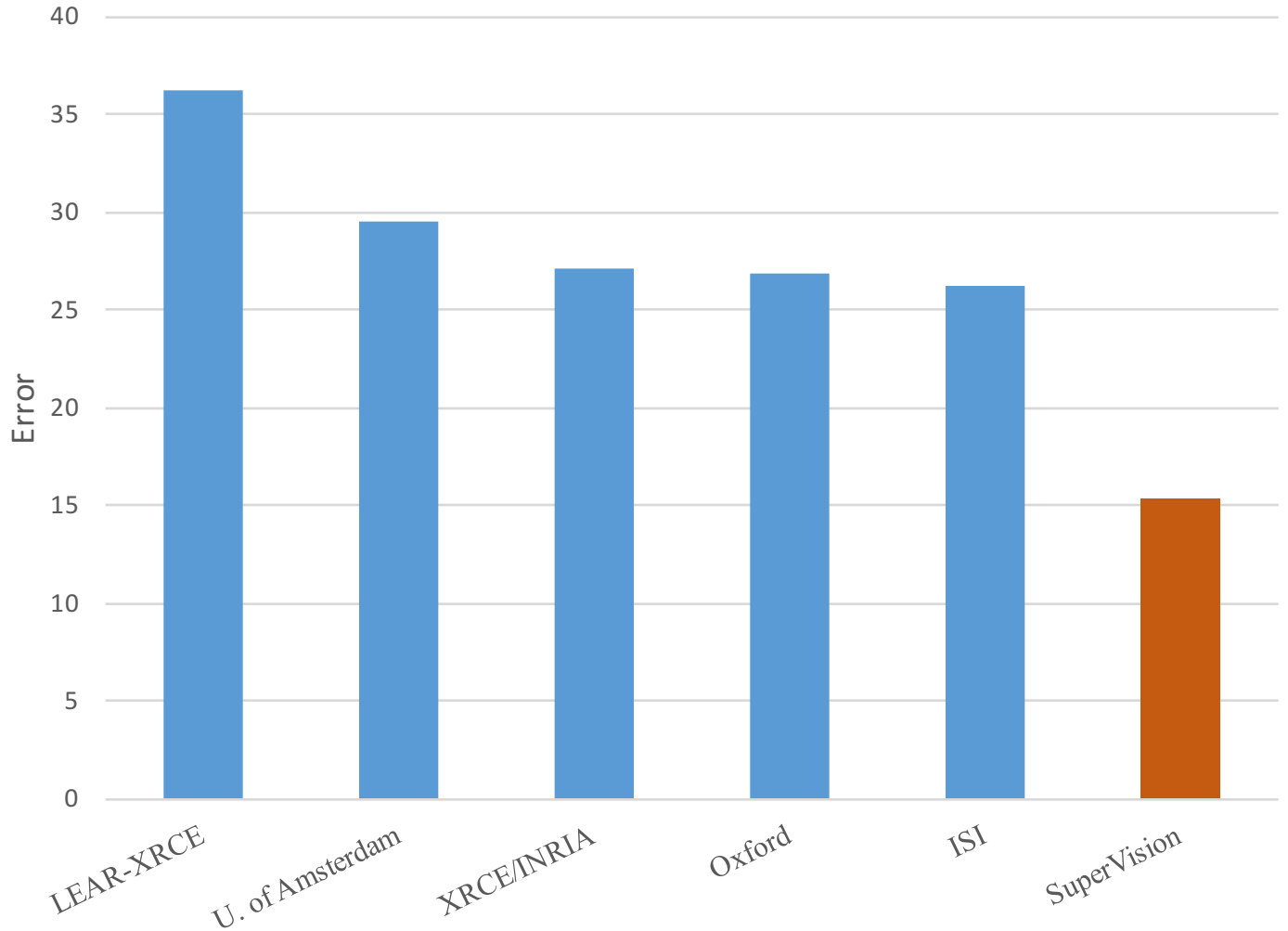
2012 ImageNet 1K

(Fall 2012)



2012 ImageNet 1K

(Fall 2012)

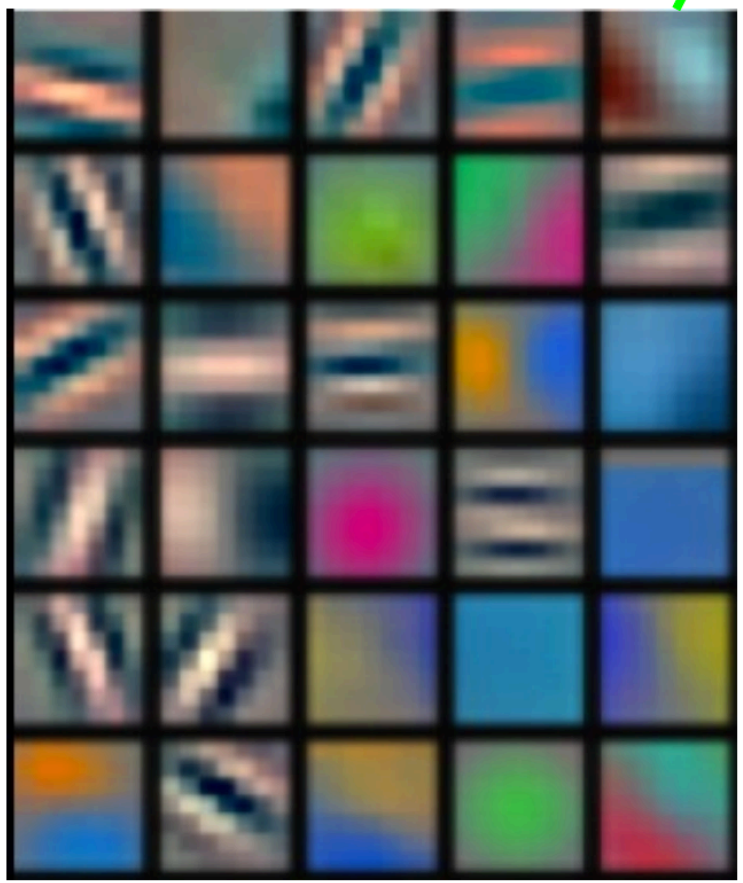


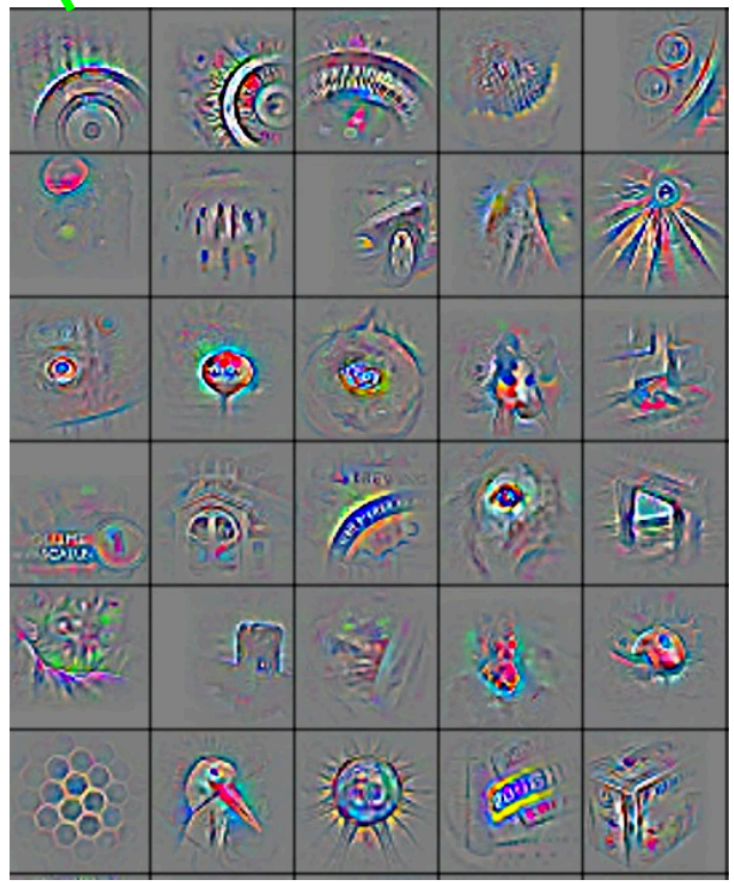
Low-Level Feature

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

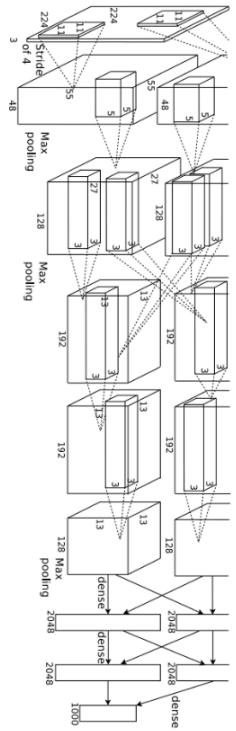
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

Year 2014

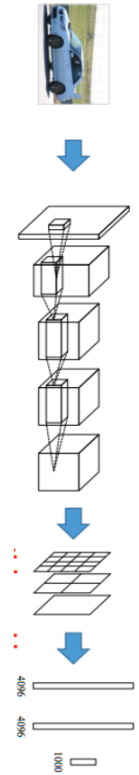
GoogLeNet

[Szegedy arxiv 2014]

VGG

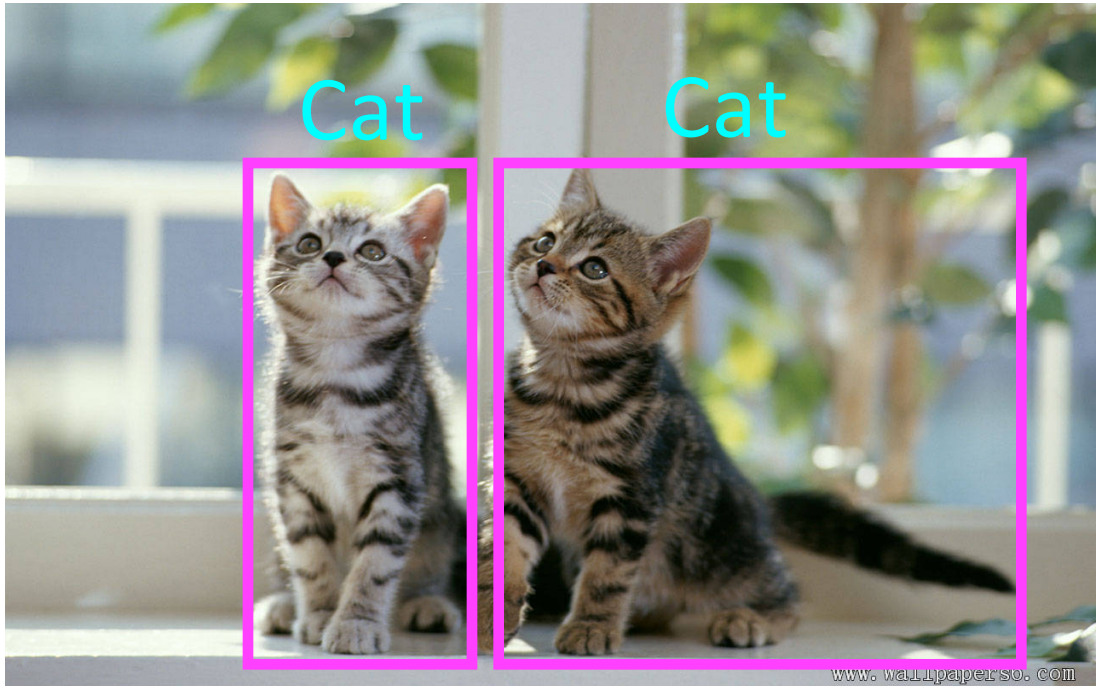
[Simonyan arxiv 2014]

MSRA

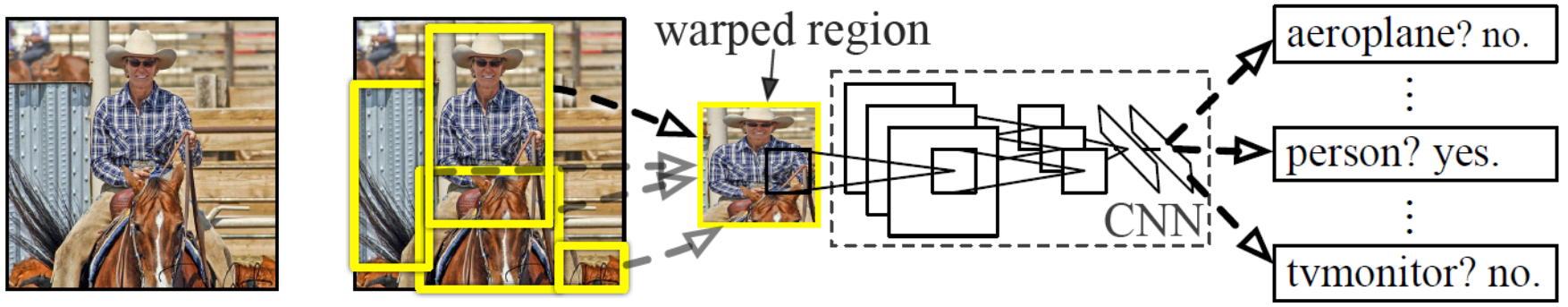


[He arxiv 2014]

Classification Vs. Detection



Object detection



1. Input image

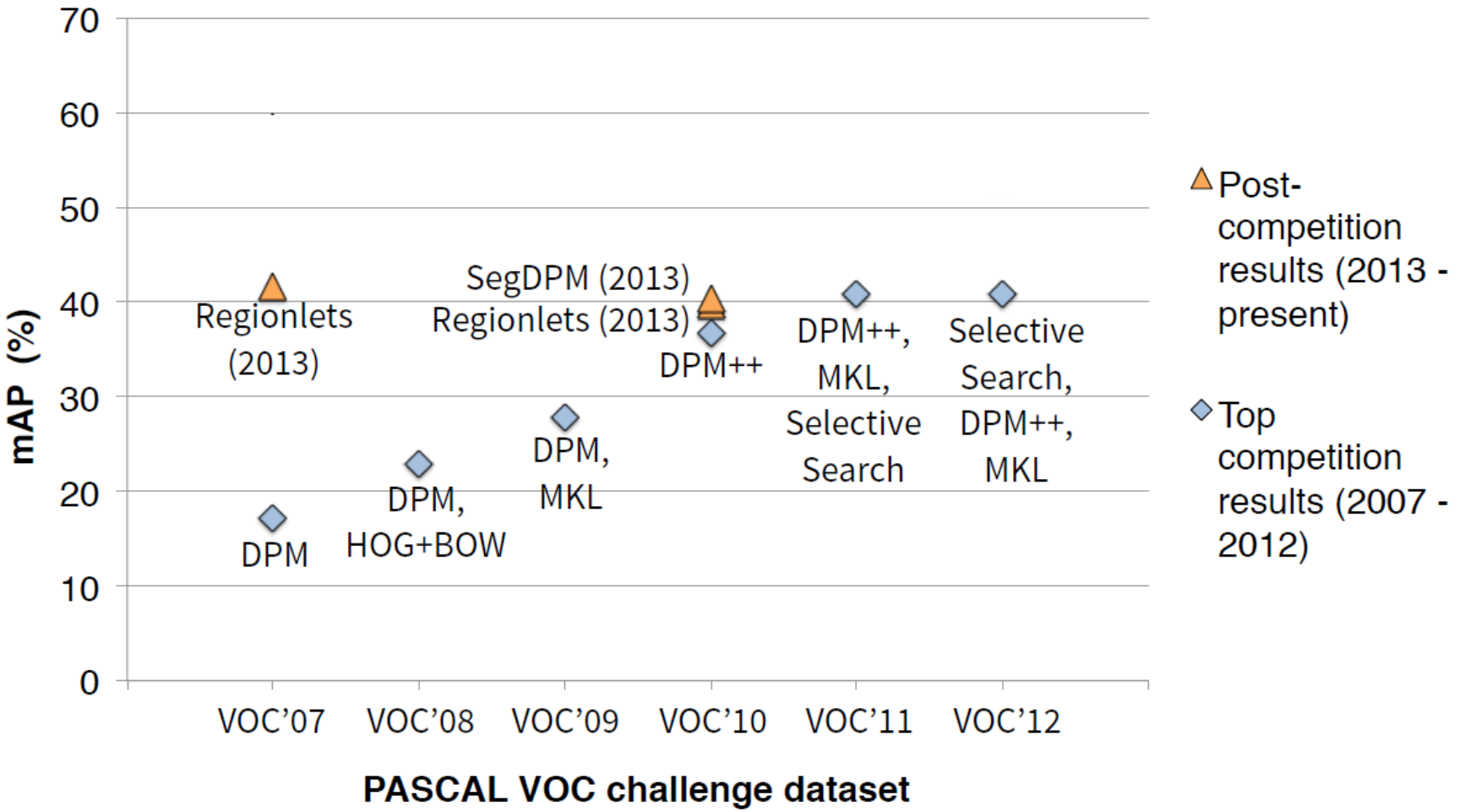
2. Extract region proposals (~2k)

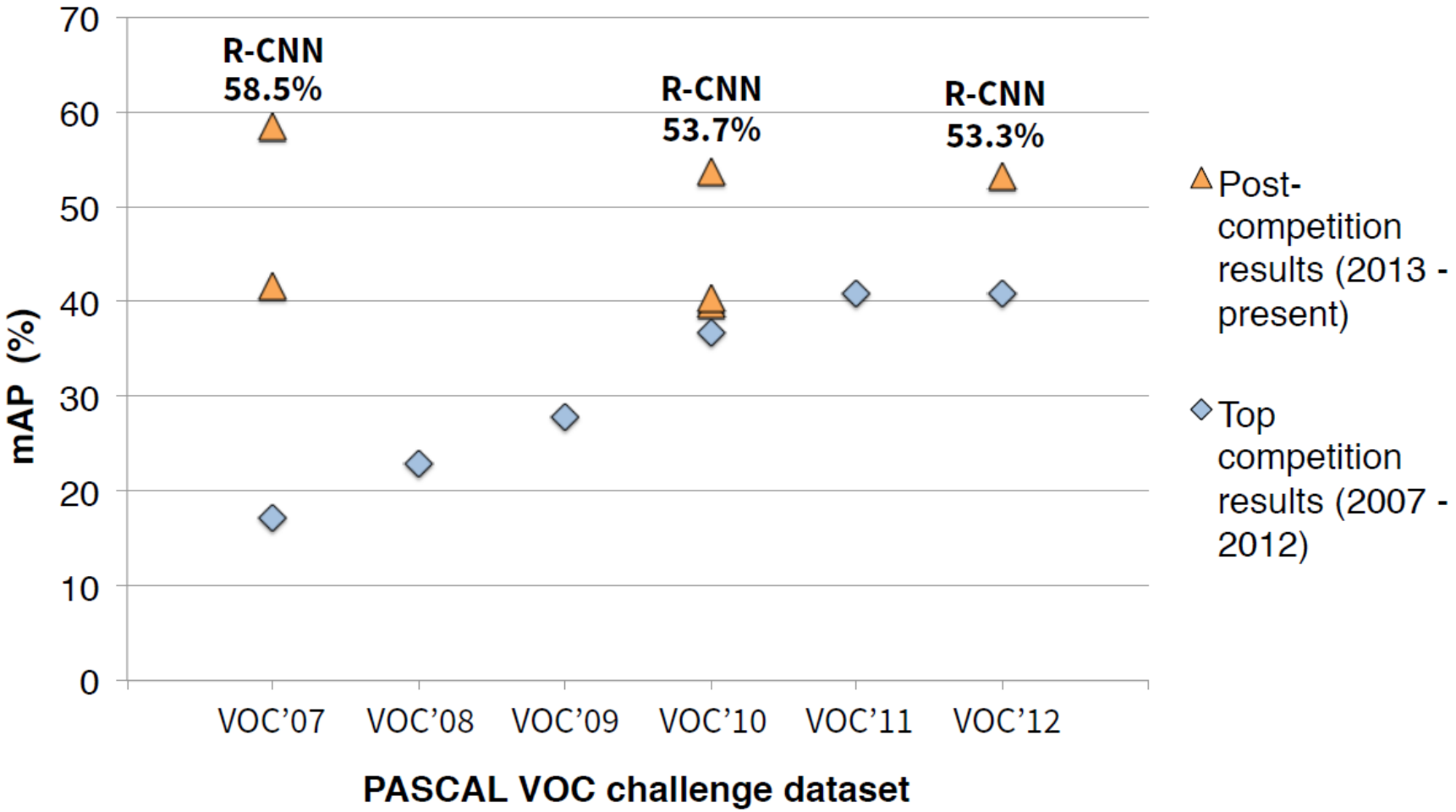
3. Compute CNN features

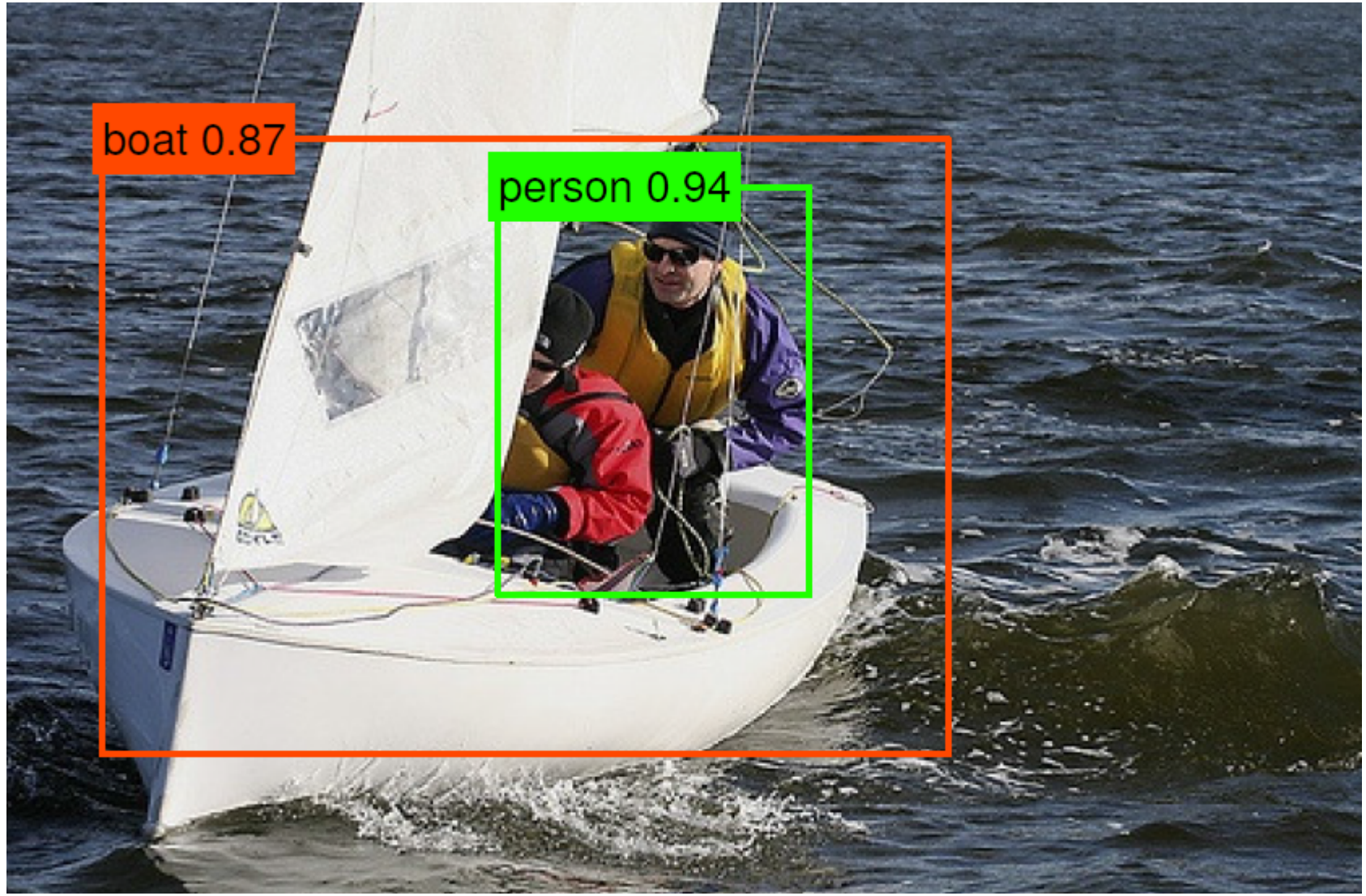
4. Classify regions

Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Girshick, Donahue, Darrell, Malik, *CVPR* 2014.

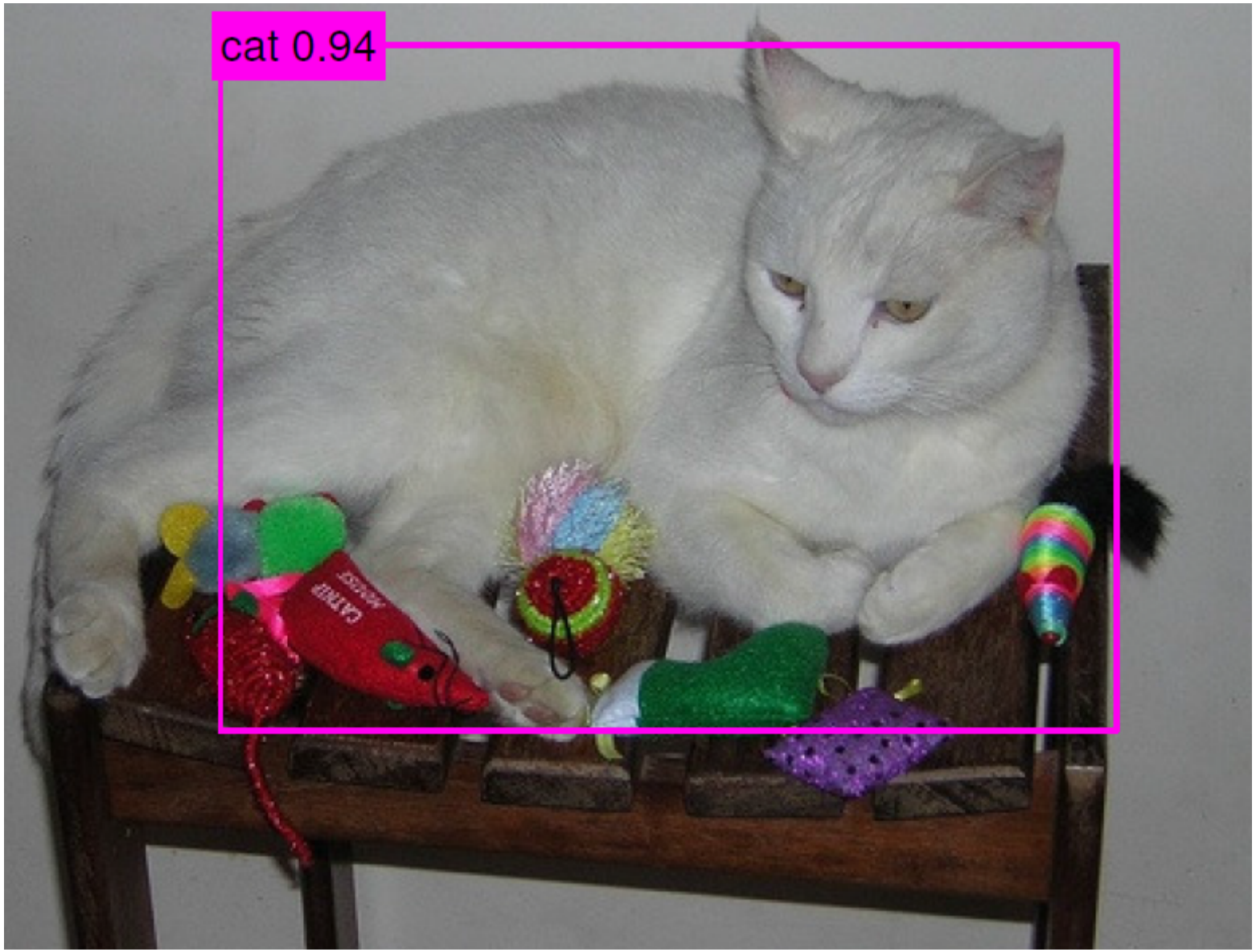
Online classification demo:
http://decaf.berkeleyvision.org/


70
60
50
40
30
20
10
0
mAP (%)
Regionlets (2013)
SegDPM (2013)
Regionlets (2013)
DPM++
DPM++, MKL, Selective Search
Selective Search, DPM++, MKL
DPM
DPM, HOG+BOW
DPM, MKL
VOC'07
VOC'08
VOC'09
VOC'10
VOC'11
VOC'12
PASCAL VOC challenge dataset
Post-competition results (2013 - present)
Top competition results (2007 - 2012)


70
60
50
40
30
20
10
0
mAP (%)
R-CNN
58.5%
R-CNN
53.7%
R-CNN
53.3%
VOC'07
VOC'08
VOC'09
VOC'10
VOC'11
VOC'12
PASCAL VOC challenge dataset
Post-competition results (2013 - present)
Top competition results (2007 - 2012)


boat 0.87
person 0.94


bottle 0.98
bottle 0.86
bottle 0.82

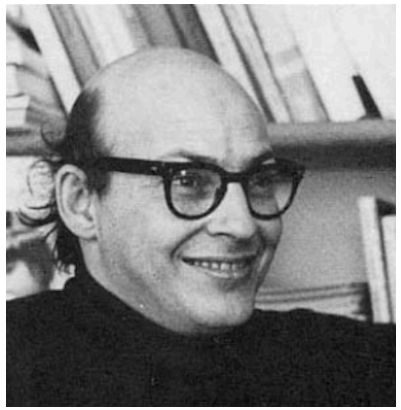

cat 0.94


cat 0.95


cat 0.82
http://phyllischan.blogspot.com

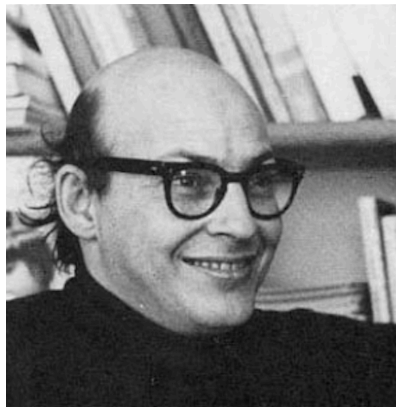
Going beyond categorization…

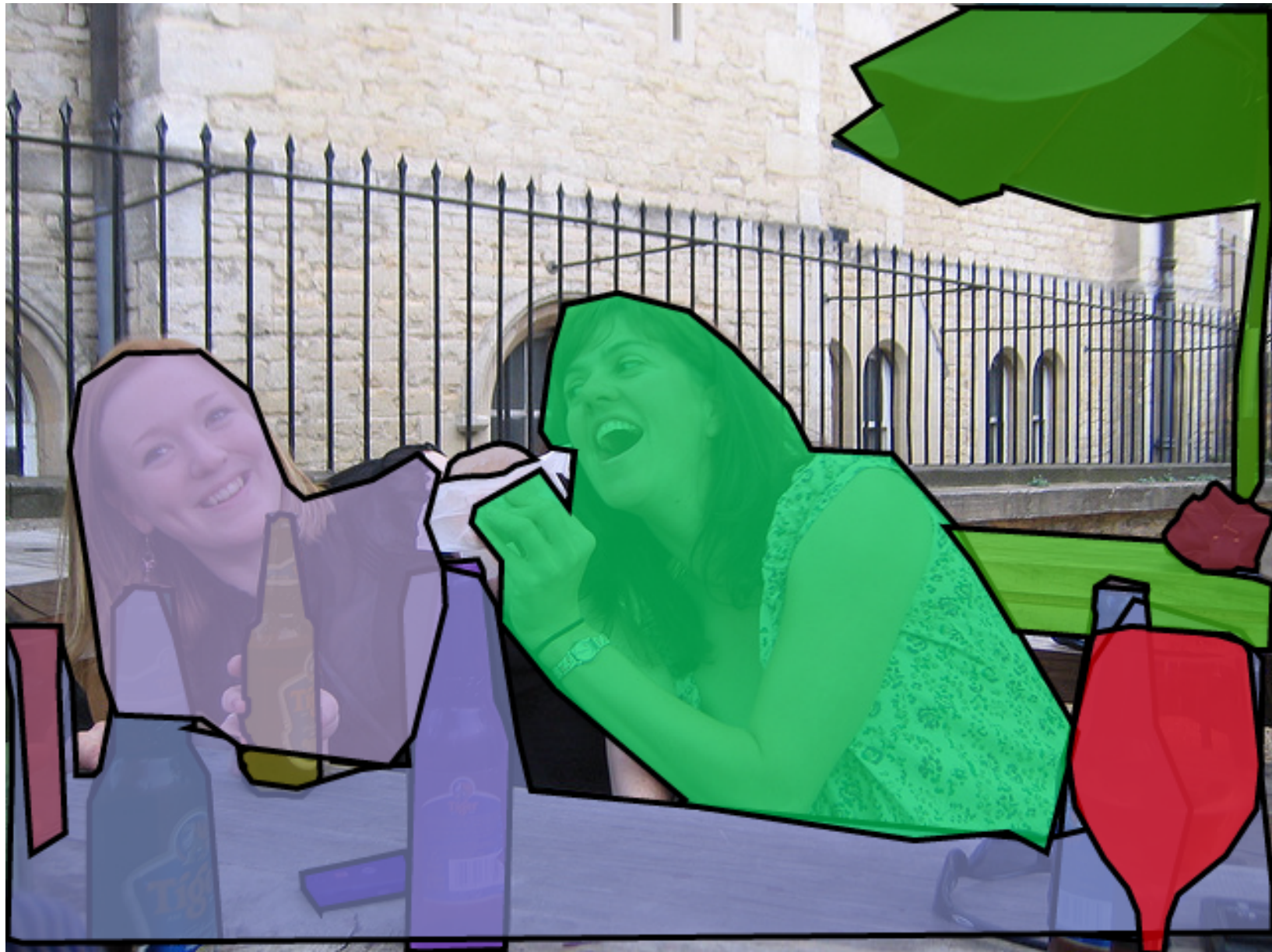
"Connect a television camera to a computer and **get the machine to describe what it sees."**



Going beyond categorization…

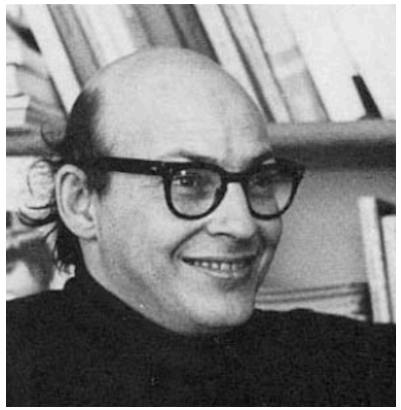
"Connect a television camera to a computer and **get the machine to describe what it sees."**





Going beyond categorization…

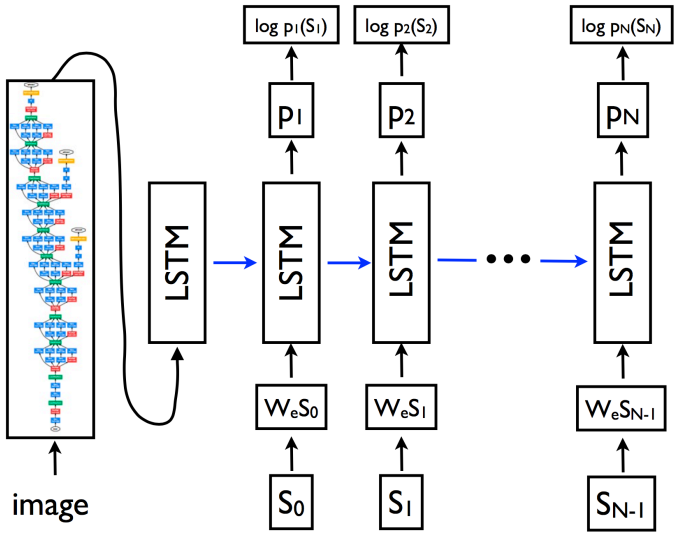
"Connect a television camera to a computer and **get the machine to describe what it sees.**"



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.

MS COCO

Going beyond categorization…



Show and Tell: A Neural Image Caption Generator
Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan, CVPR 2015

…the “giraffe-tree” problem ☹

a giraffe next to a tree

Big Visual Data

flickr

6 billion images

3.5 trillion photographs

the simple image sharer imgur

1 billion images served daily

YouTube

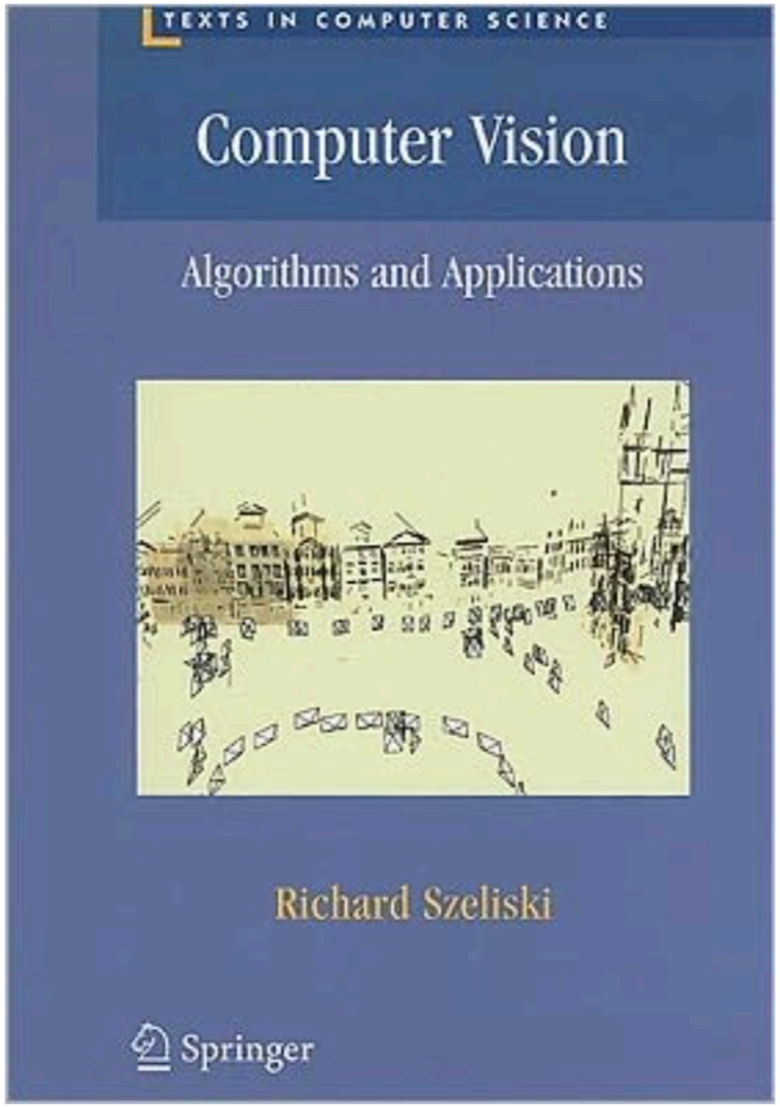
100 hours uploaded per minute

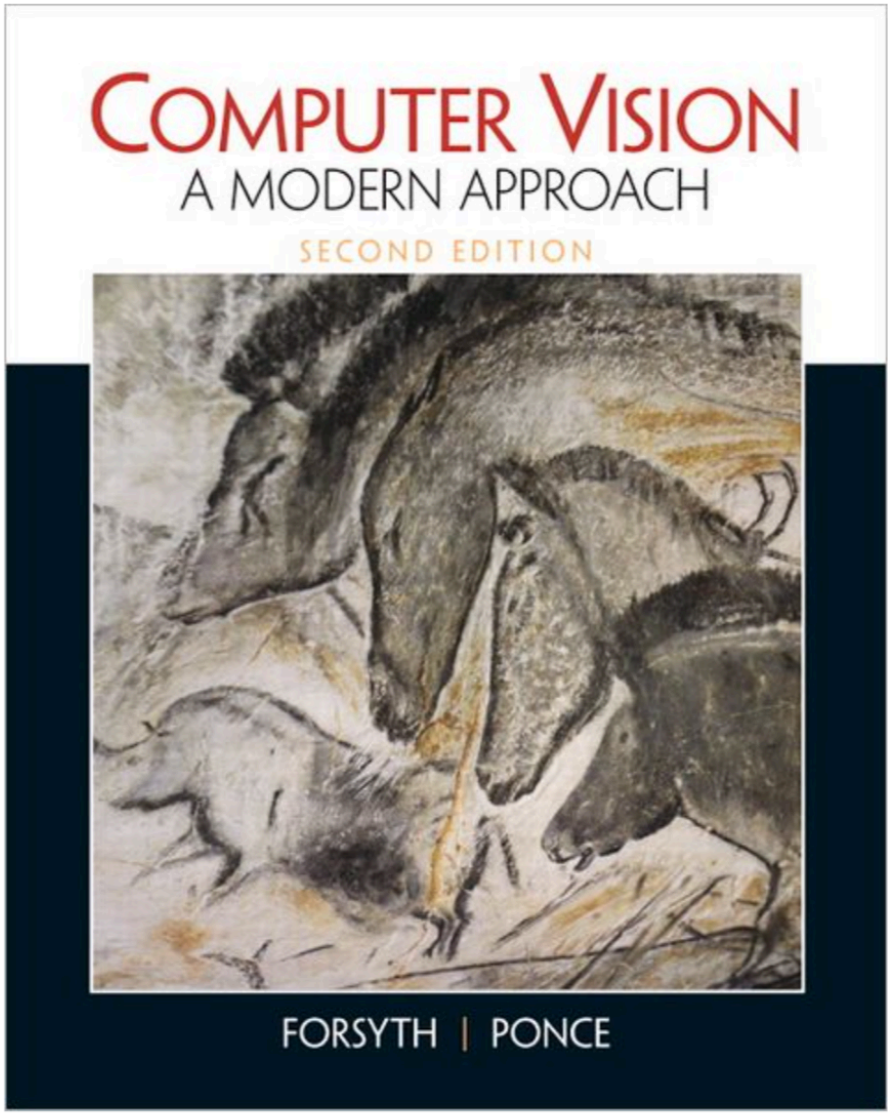
facebook

70 billion images


Too Big for Humans
Digital Dark Matter
[Perona 2010]

Books


TEXTS IN COMPUTER SCIENCE
Computer Vision
Algorithms and Applications
Richard Szeliski
Springer


COMPUTER VISION
A MODERN APPROACH
SECOND EDITION
FORSYTH | PONCE