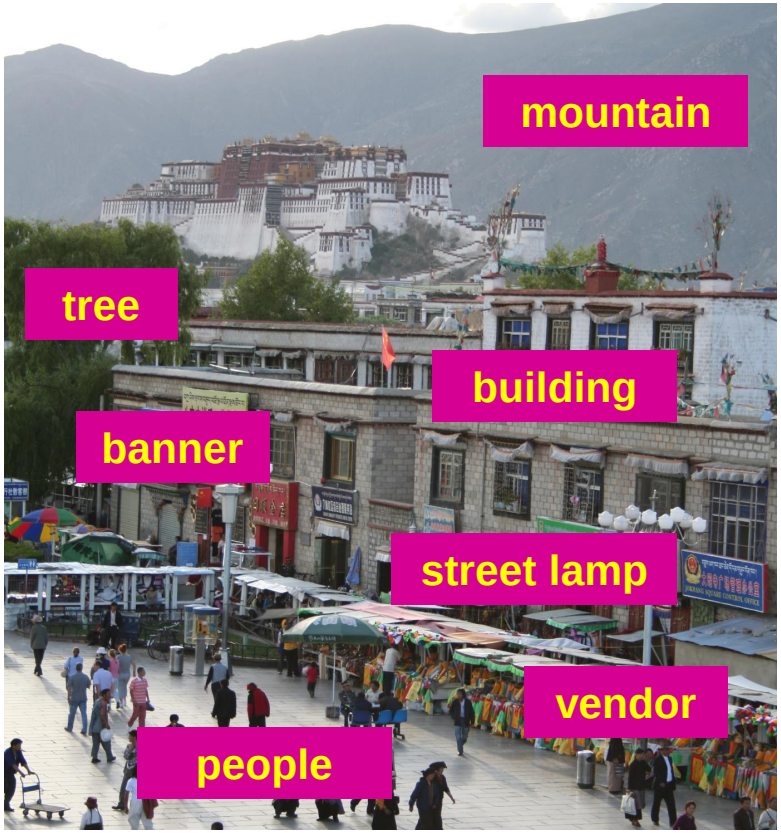
CS5670: Intro to Computer Vision

Noah Snavely

Introduction to Recognition



Announcements

- Final exam, in-class, last day of lecture (5/9/2018, 12:30 – 1:45pm)
- Voting for Project 3 artifacts
- Project 4 (Stereo) to be released soon

Where we go from here

- What we know: Geometry
  - What is the shape of the world?
  - How does that shape appear in images?
  - How can we infer that shape from one or more images?

- What's next: Recognition
  - What are we looking at?

What do we mean by “object recognition”?

Next 15 slides adapted from Li, Fergus, & Torralba’s excellent short course on category and object recognition

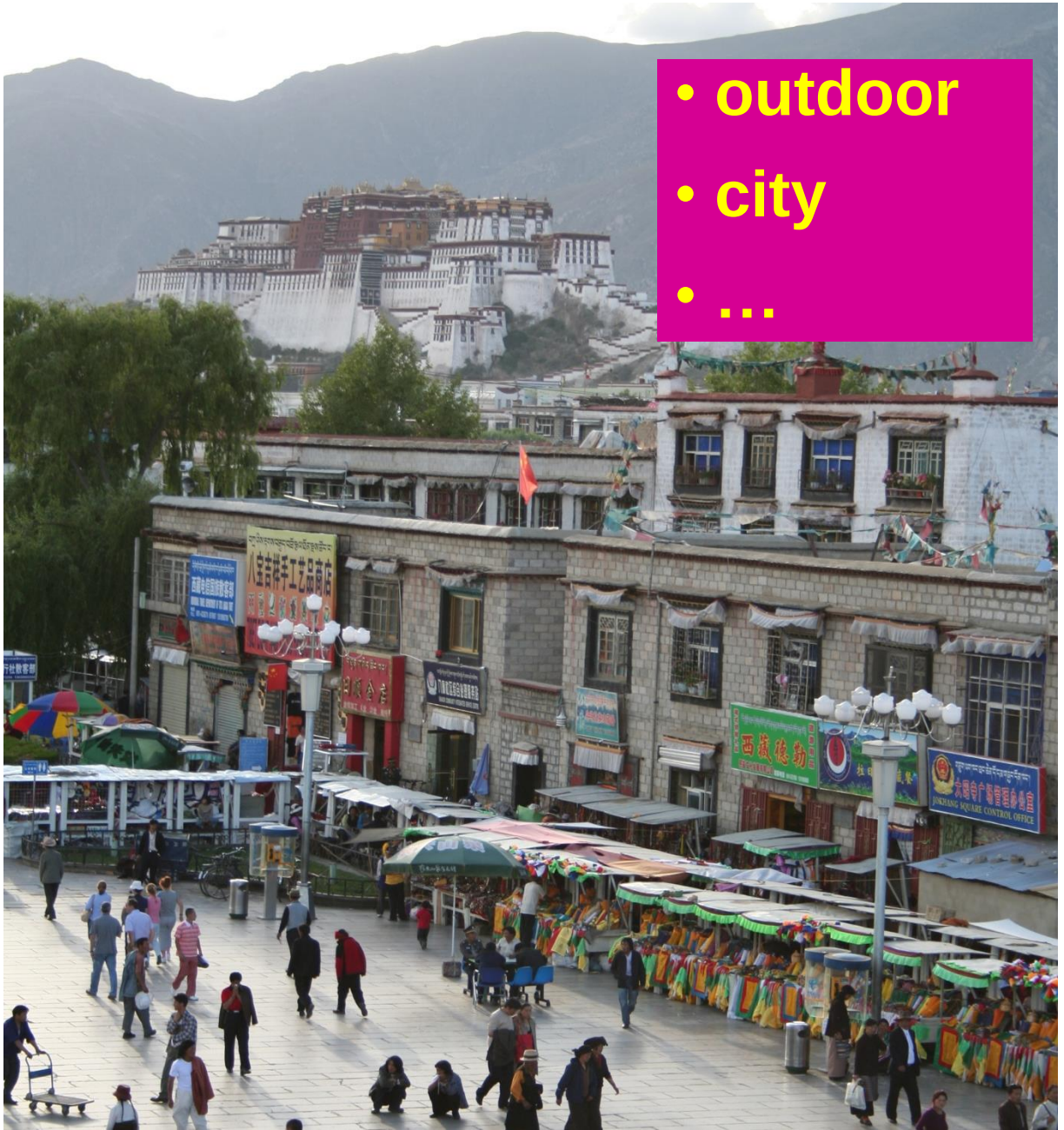
Verification: is that a lamp?

Detection: are there people?

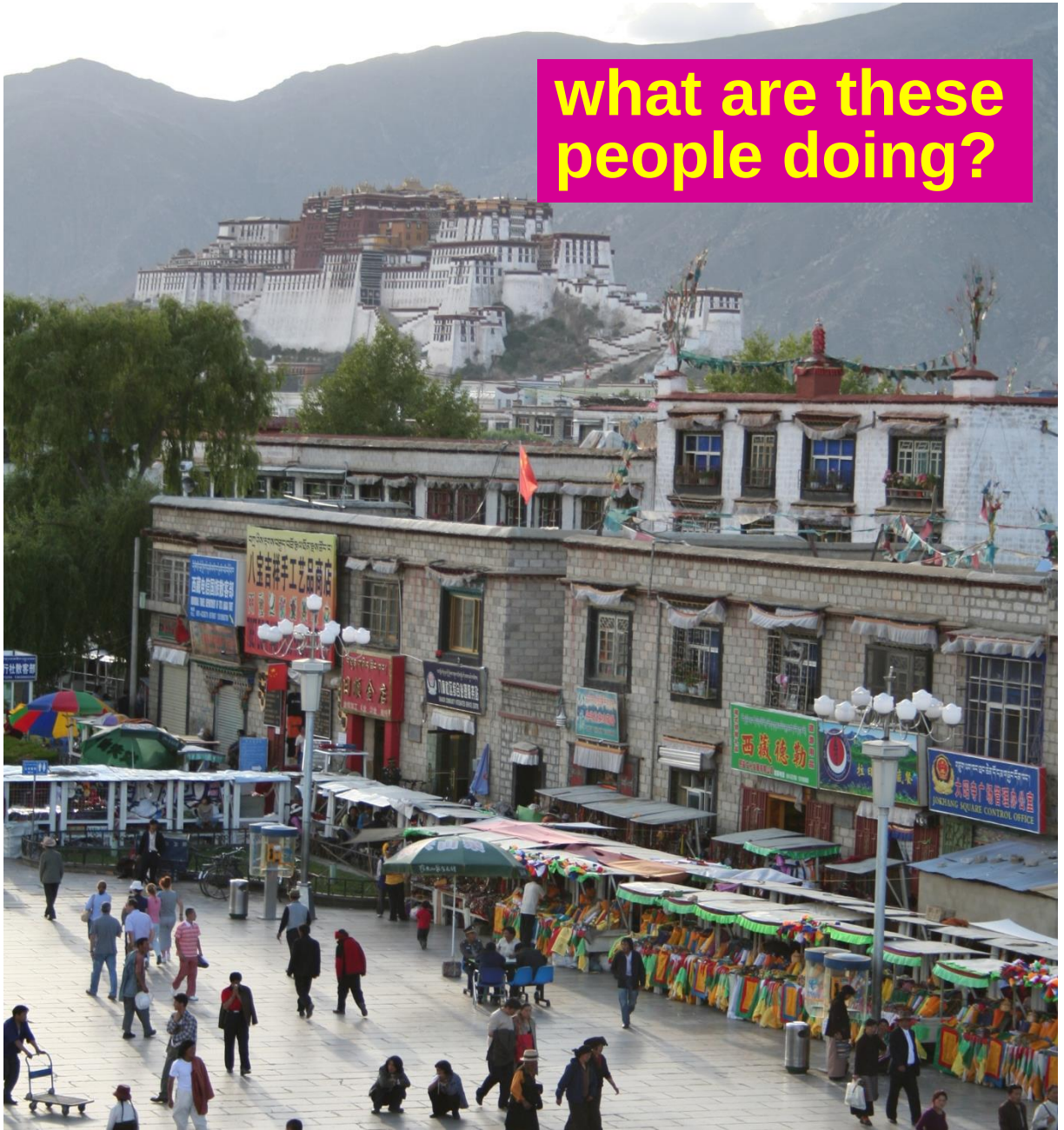
Identification: is that Potala Palace?

Object categorization

Scene and context categorization

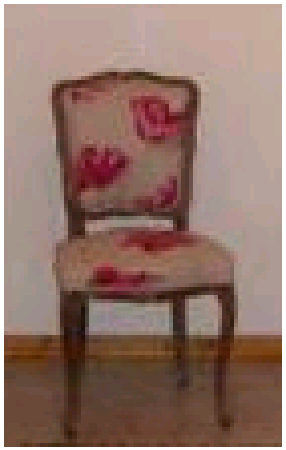


Activity / Event Recognition



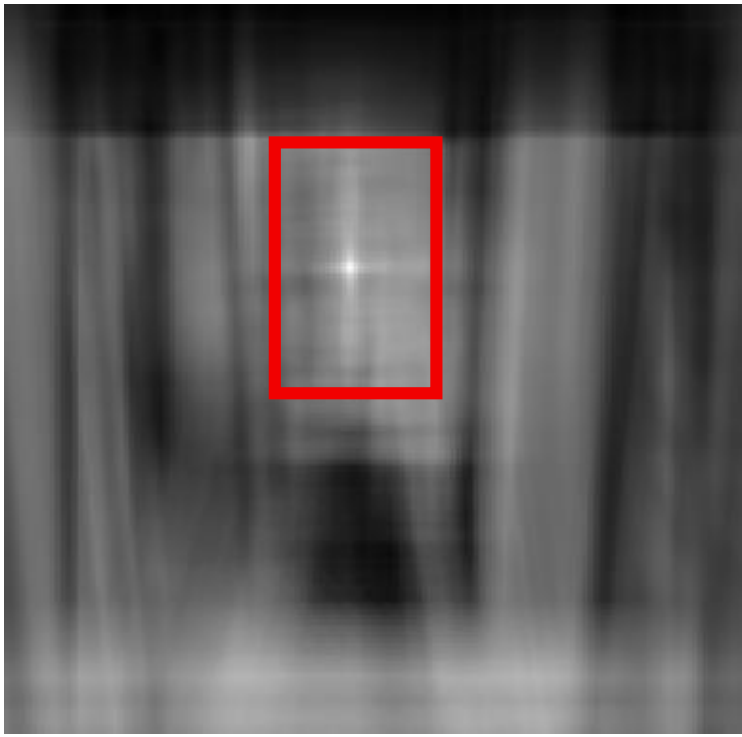
Object recognition
Is it really so hard?

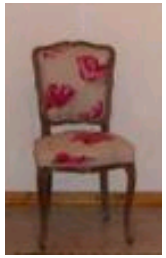
This is a chair



Find the chair in this image

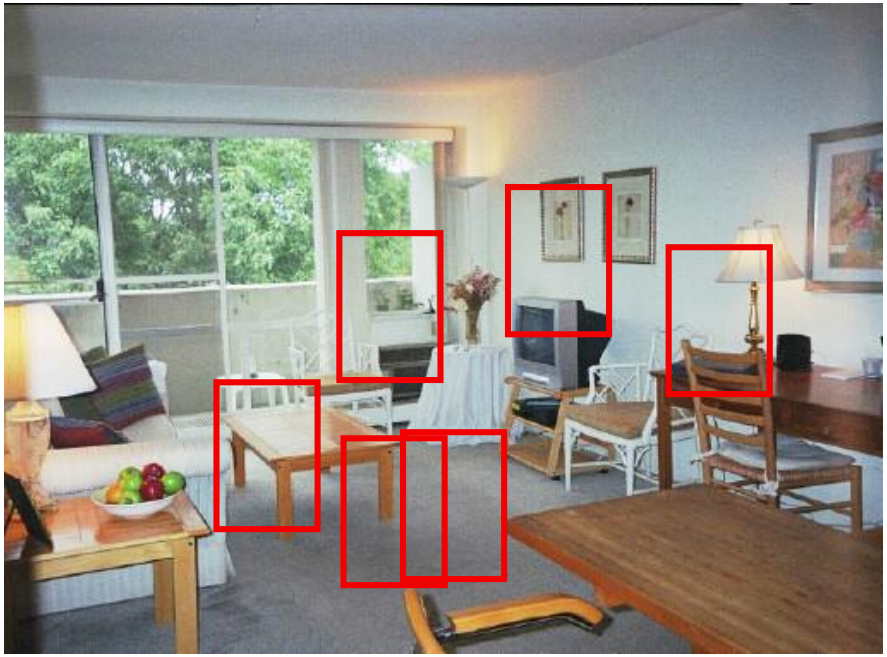
Output of normalized correlation

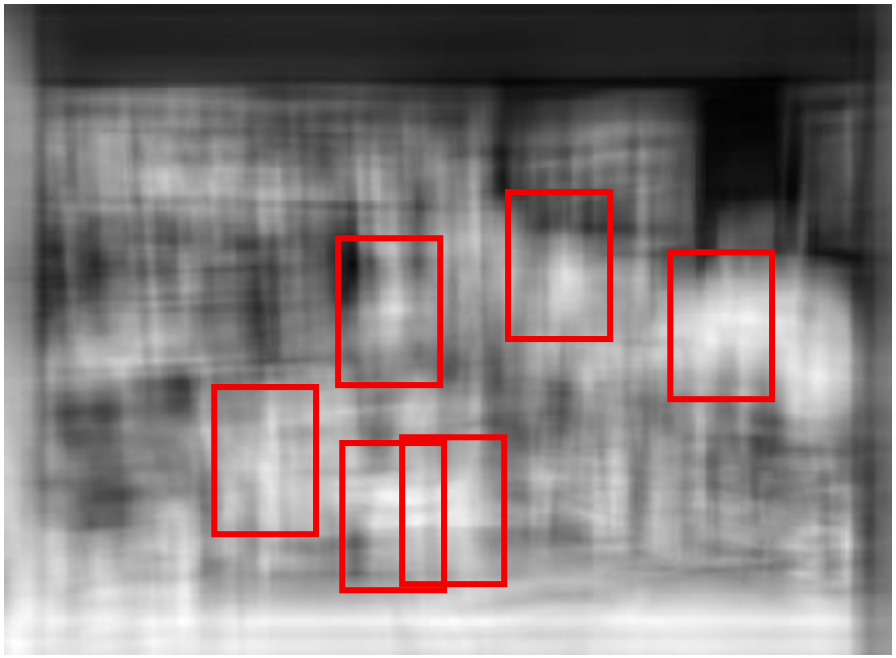




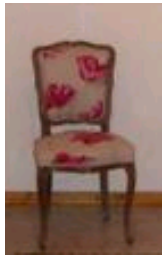
Object recognition
Is it really so hard?

Find the chair in this image





Pretty much garbage
Simple template matching is not going to do the trick



Object recognition
Is it really so hard?

Find the chair in this image

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

Why not use SIFT matching for everything?

- Works well for object *instances* (or distinctive images such as logos)

- Not great for generic object *categories*



And it can get a lot harder

Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422

How do human do recognition?

- We don’t completely know yet
- But we have some experimental observations.

Observation 1

- We can recognize familiar faces even in low-resolution images

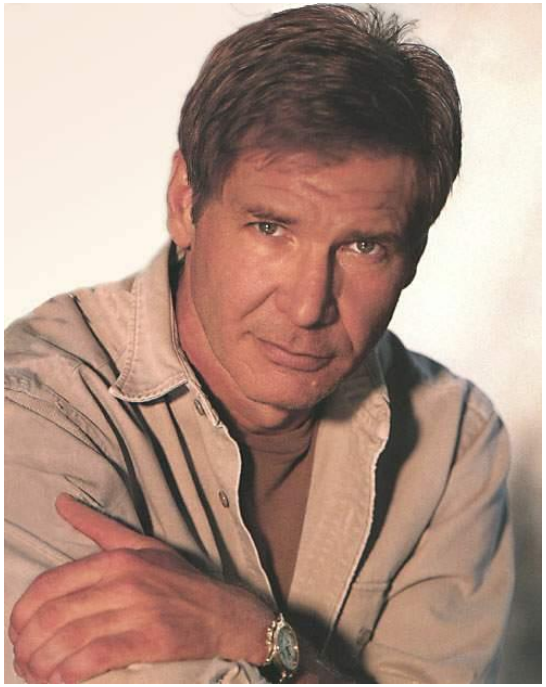
Observation 2:

Jim Carrey

Kevin Costner

- High frequency information is not enough

What is the single most important facial features for recognition?



Observation 4:

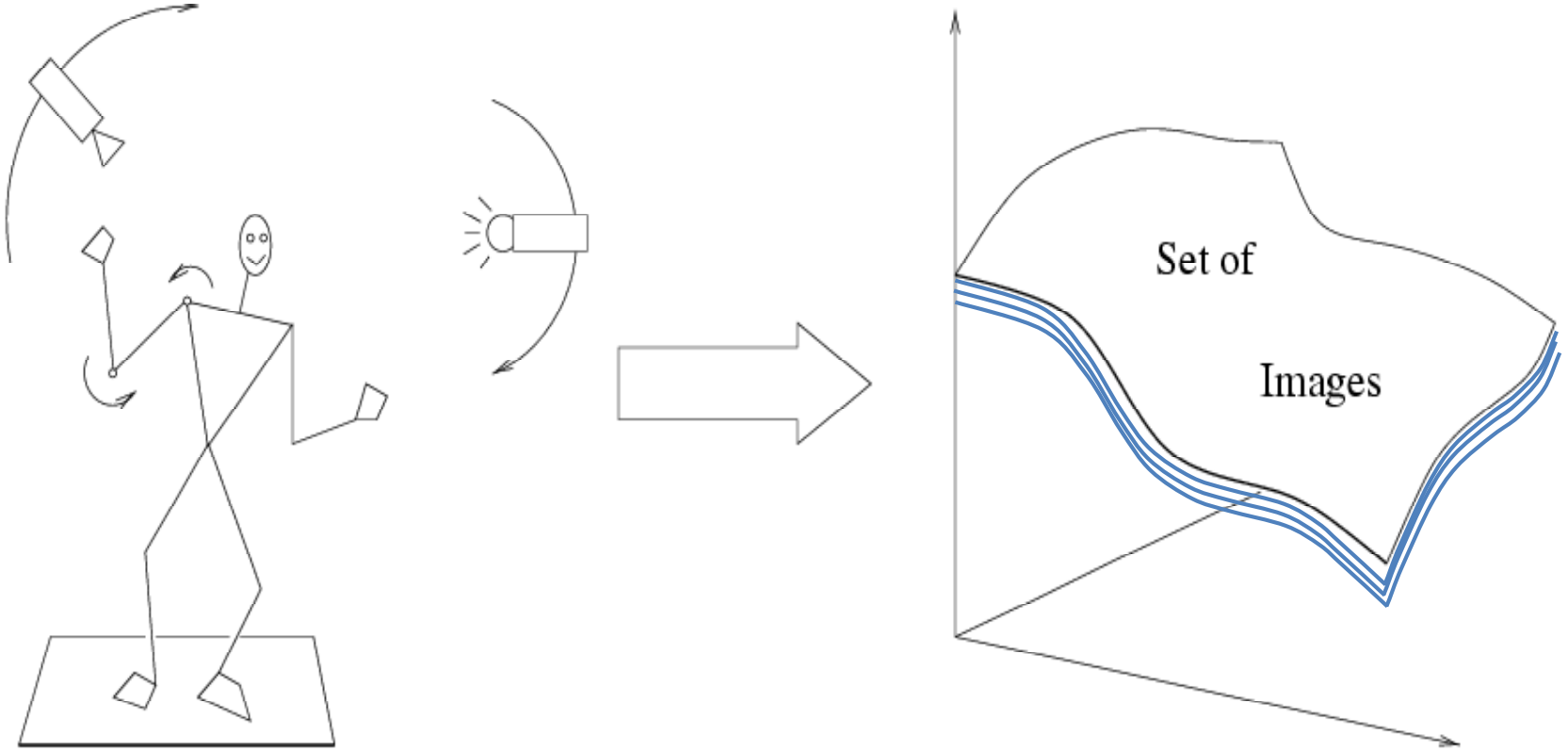
- Image Warping is OK

The list goes on

Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About

- http://web.mit.edu/bcs/sinha/papers/19results_sinha_etal.pdf

Why is this hard?



Variability: Camera position
Illumination
Shape parameters

How many object categories are there?

Challenge: variable viewpoint

Michelangelo 1475-1564

Challenge: variable illumination

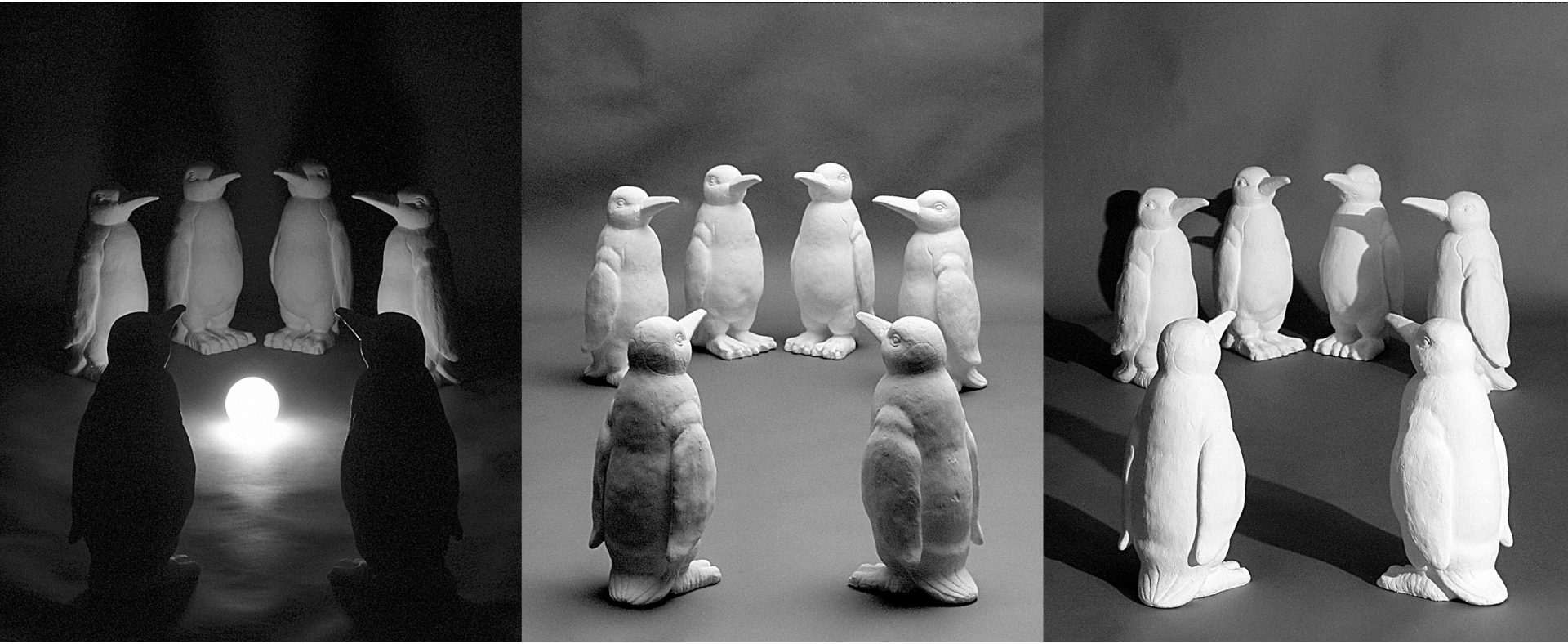
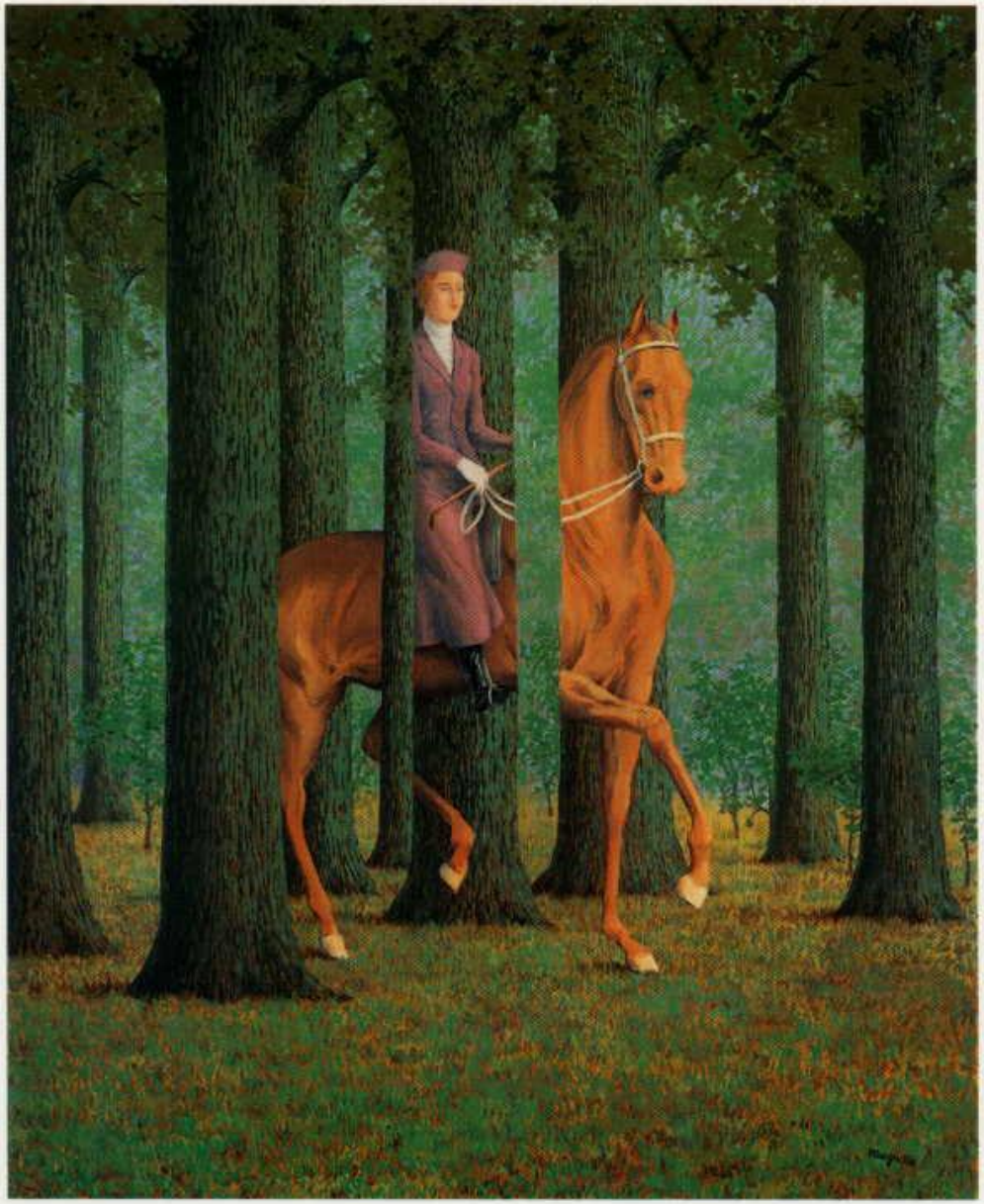


image credit: J. Koenderink

Challenge: scale

Challenge: deformation

Challenge: Occlusion

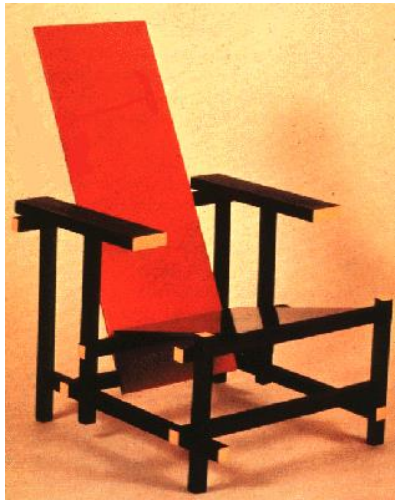


Magritte, 1957

Challenge: background clutter

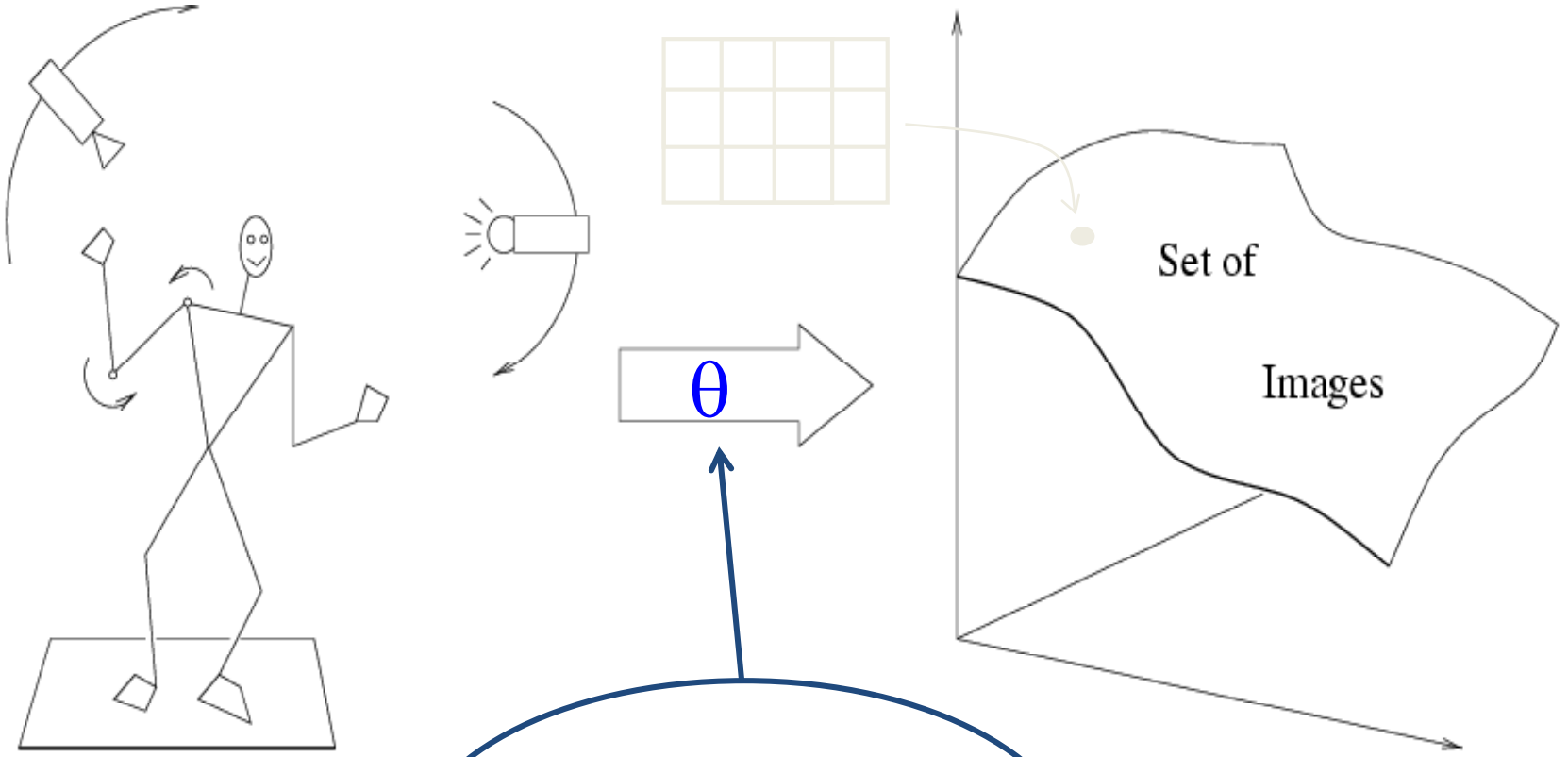
Kilmeny Niland. 1995

Challenge: intra-class variations



History of ideas in recognition

- 1960s – early 1990s: the geometric era



Variability:

Camera position
Illumination

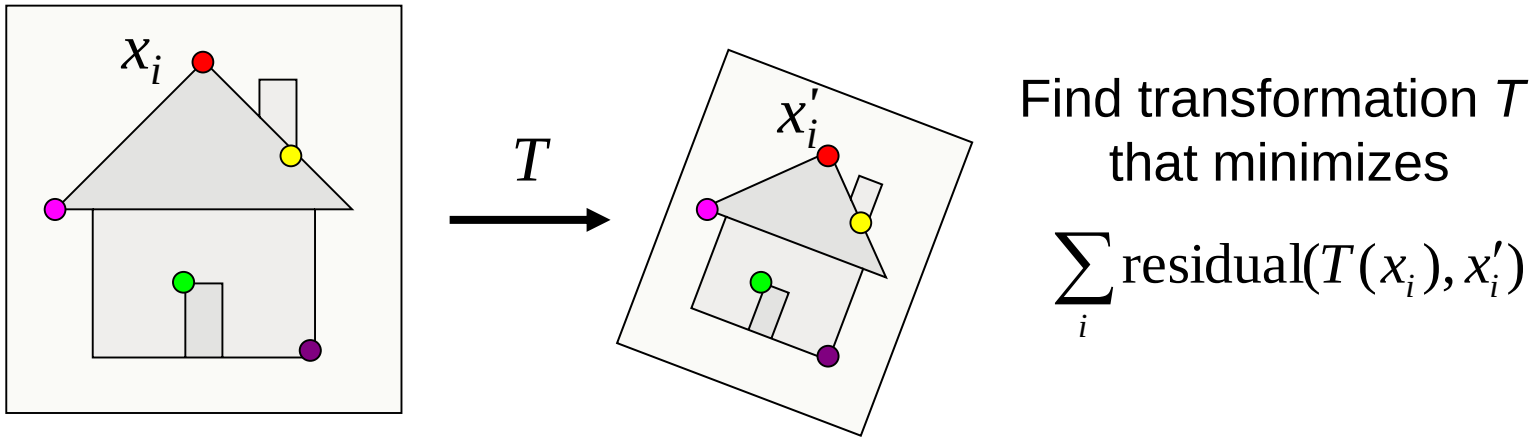
Alignment

Shape: assumed known

Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

Instance Recognition

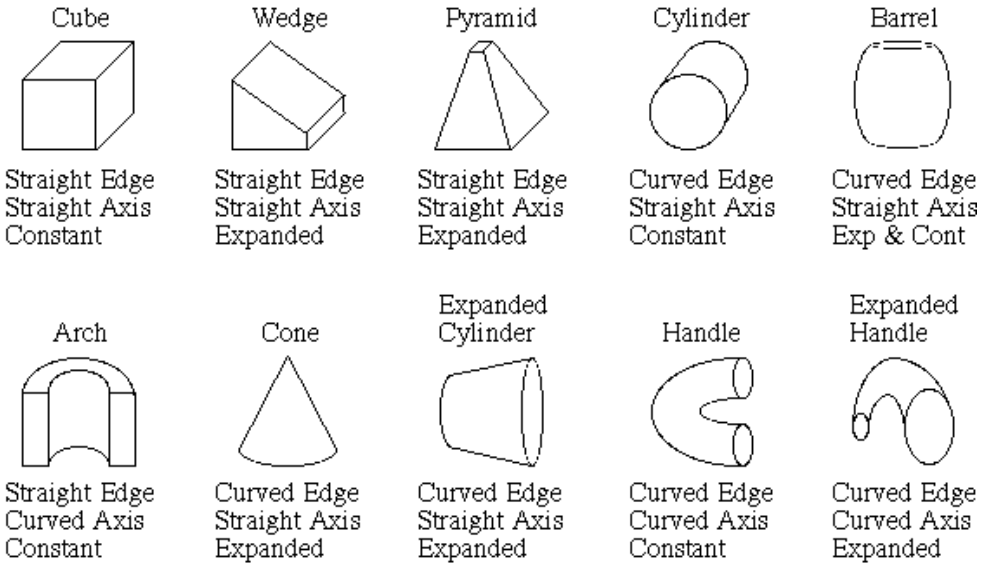
- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



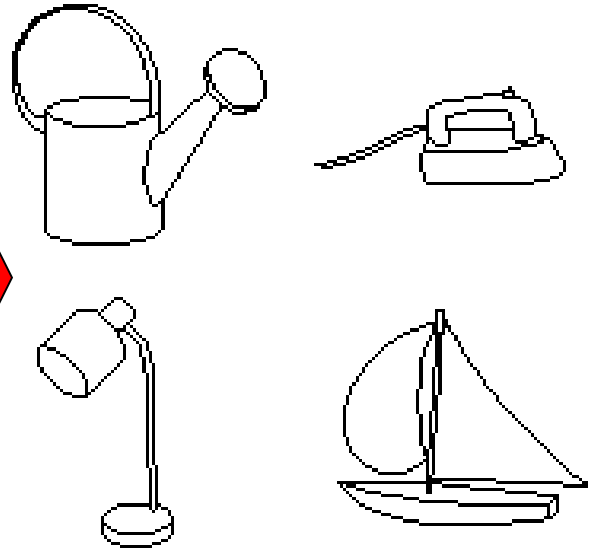
Recognition by components

Biederman (1987)

Primitives (geons)



Objects



http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models

Eigenfaces (Turk & Pentland, 1991)

Basis faces

Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

Svetlana Lazebnik

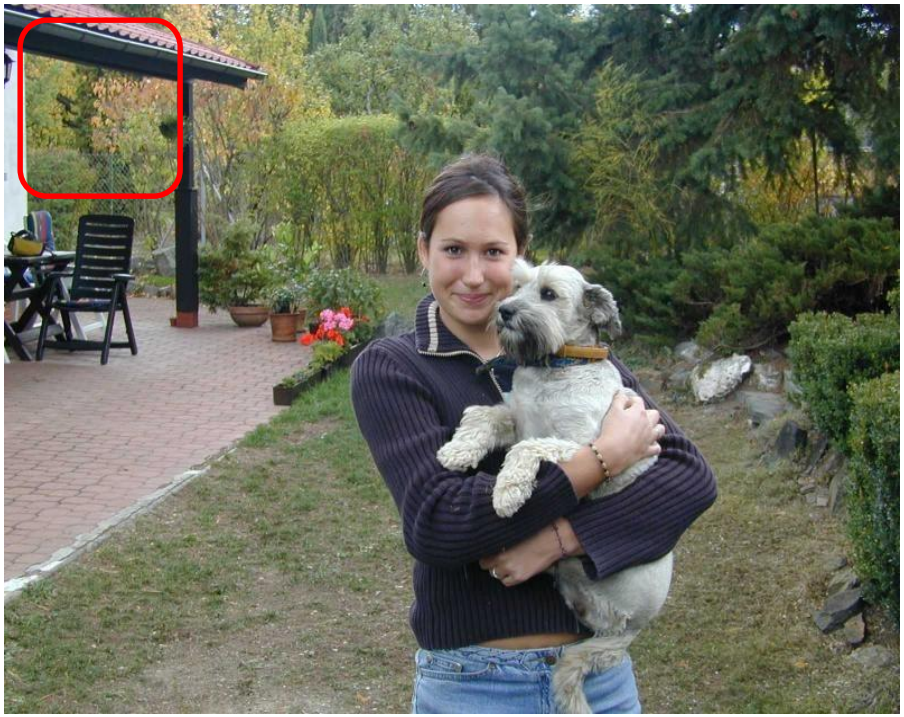
Limitations of global appearance models

- Requires global registration of patterns
- Not robust to clutter, occlusion, geometric transformations

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s – present: sliding window approaches

Sliding window approaches



Sliding window approaches

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

- Schneiderman & Kanade, 2004
- Agrawal and Roth, 2002
- Poggio et al. 1993

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features

Local features for object instance recognition

D. Lowe (1999, 2004)

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models

Parts-and-shape models

- Model:
  - Object as a set of parts
  - Relative locations between parts
  - Appearance of part

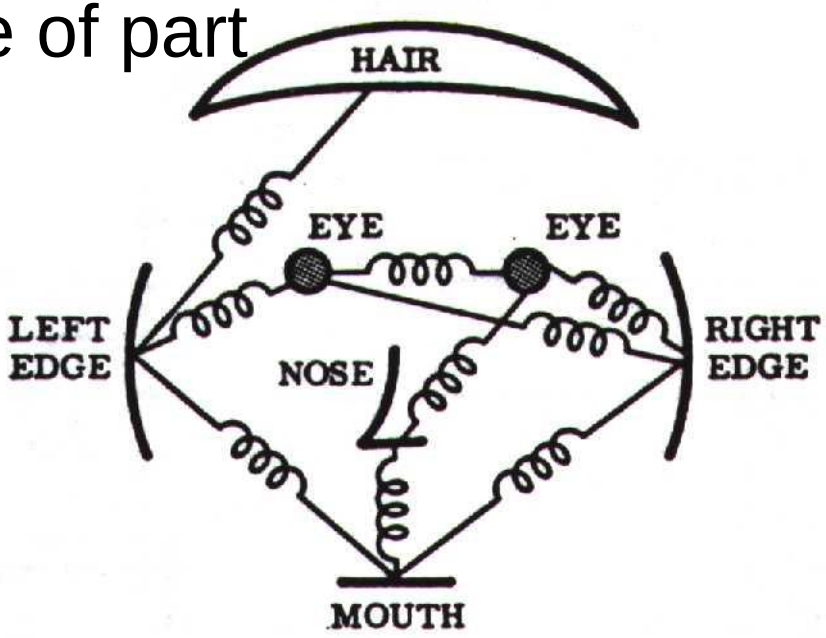
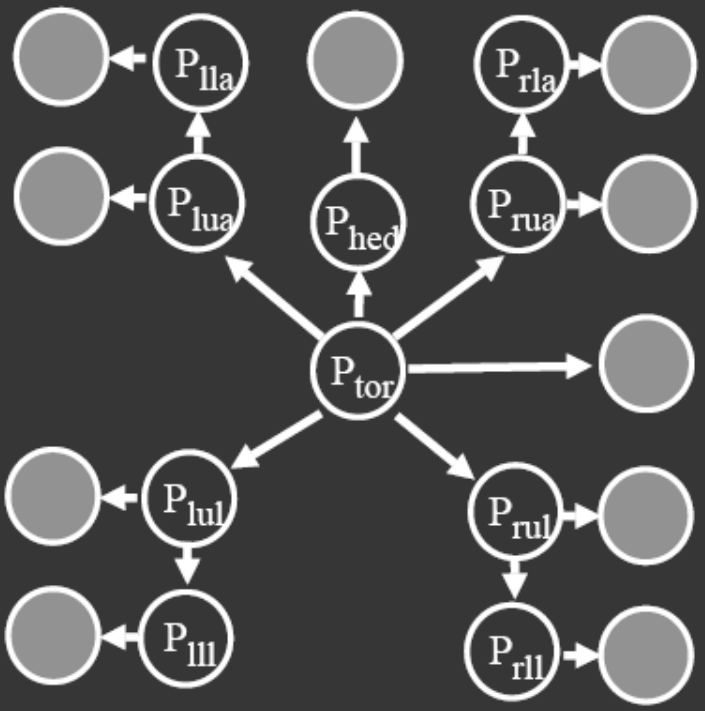
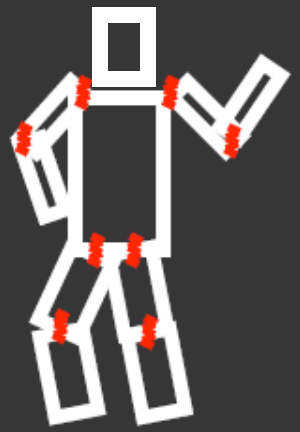


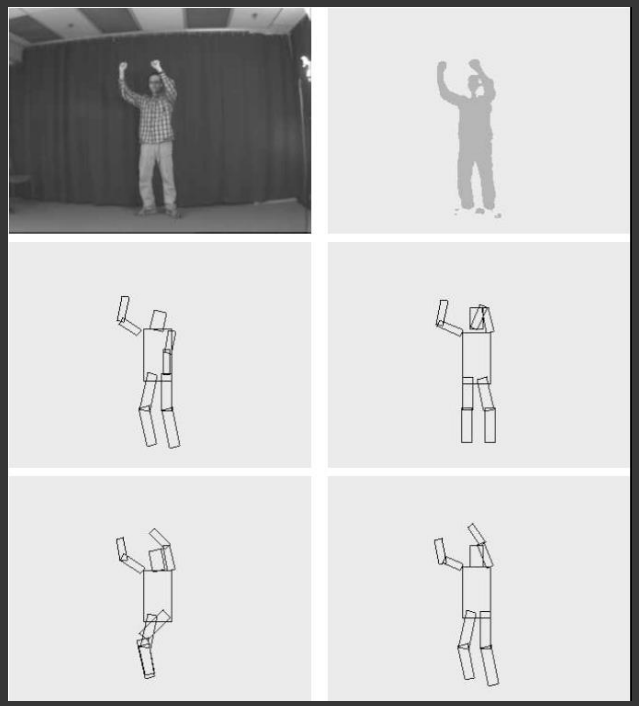
Figure from [Fischler & Elschlager 73]

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)





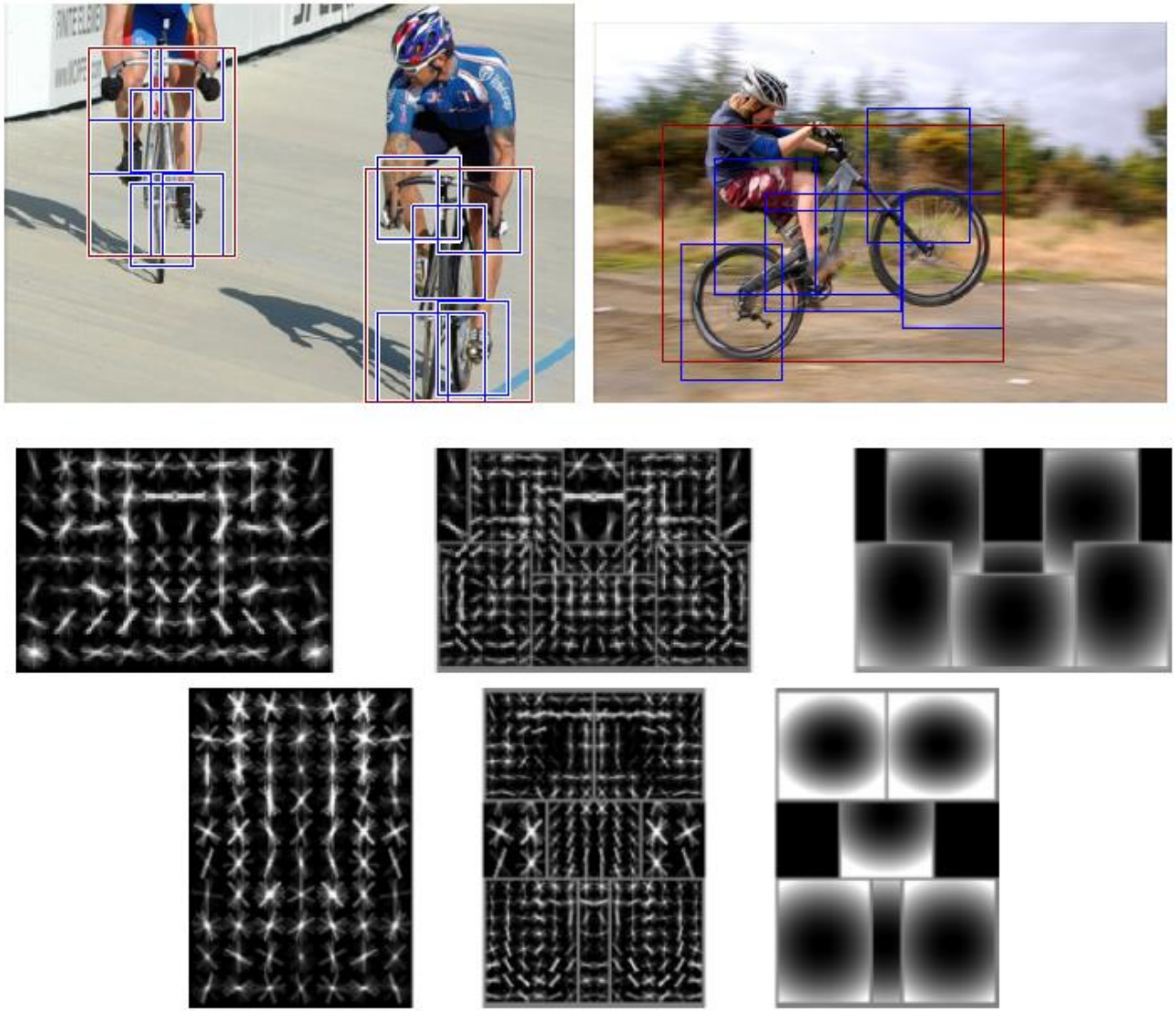


$$\Pr(P_{tor}, P_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \Pr(P_i | P_j) \prod_i \Pr(\mathrm{Im}(P_i))$$

part geometry

part appearance

Discriminatively trained part-based models



P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, **"Object Detection with Discriminatively Trained Part-Based Models,"** PAMI 2009

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features

Bag-of-features models

Bag-of-features models

Object → **Bag of ‘words’**

Svetlana Lazebnik

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: data-driven methods, **deep learning**

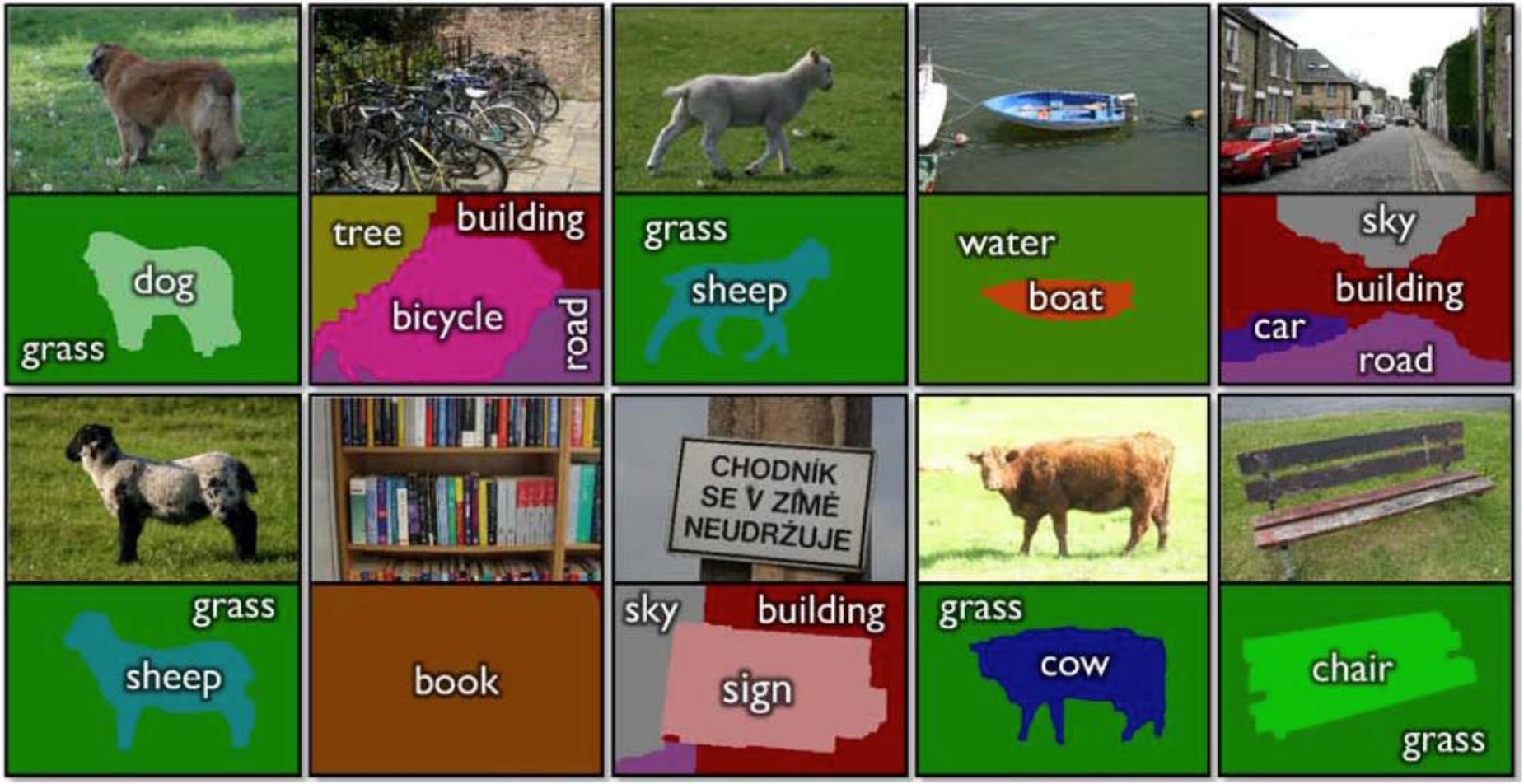
What Matters in Recognition?

- Learning Techniques
  - E.g. choice of classifier or inference method
- Representation
  - Low level: SIFT, HoG, GIST, edges
  - Mid level: Bag of words, sliding window, deformable model
  - High level: Contextual dependence
  - Deep features
- Data
  - More is always better (as long as it is good data)
  - Annotation is the hard part

Types of Recognition

- Instance recognition
  - Recognizing a known object but in a new viewpoint, with clutter and occlusion
  - Location/Landmark Recognition
    - Recognize Paris, Rome, ... in photographs
    - Ideas from information retrieval
- Category recognition
  - Harder problem, even for humans
  - Bag of words, part-based, recognition and segmentation

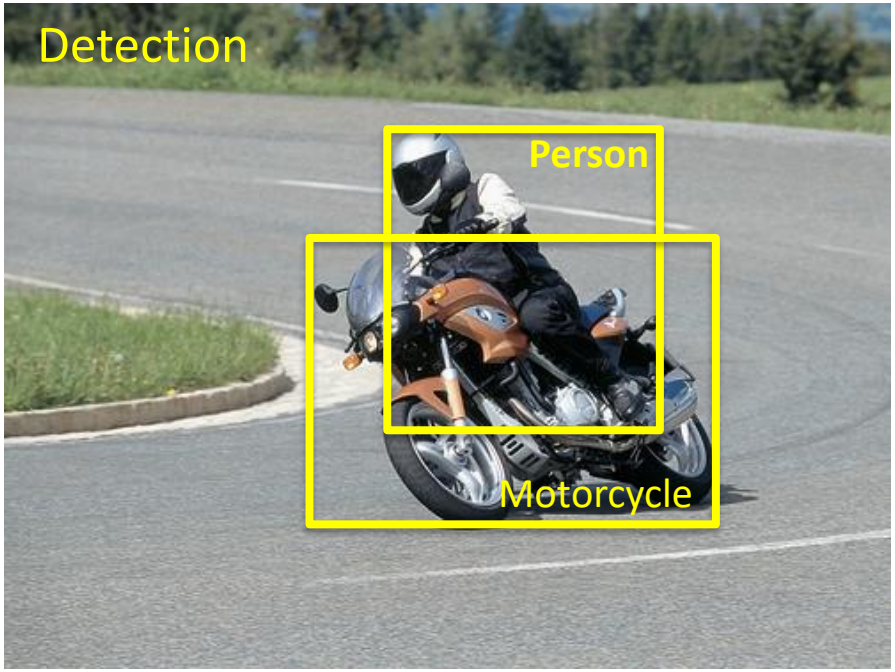
Simultaneous recognition, detection, and segmentation

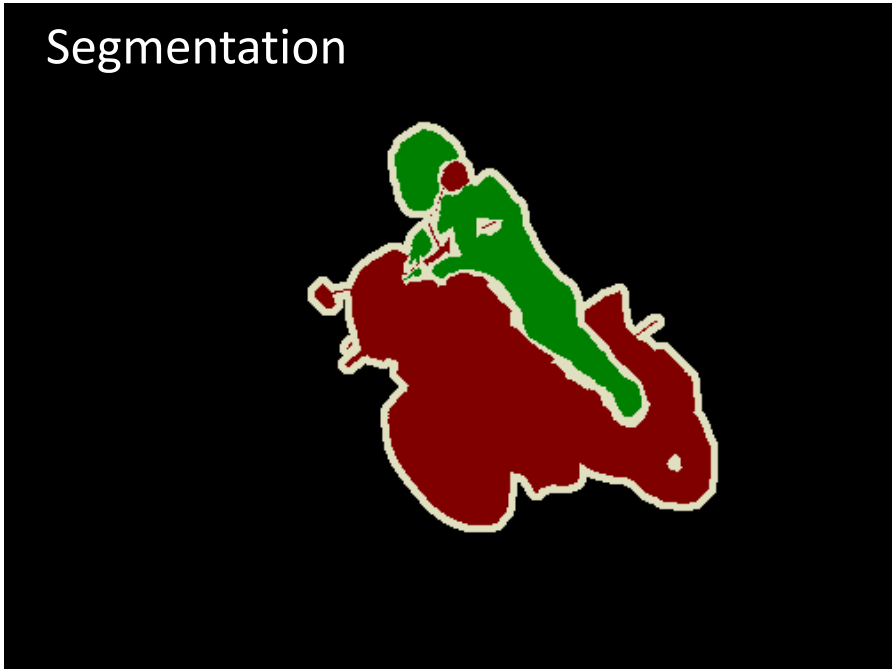


PASCAL VOC 2005-2012

20 object classes **22,591 images**

Classification: person, motorcycle



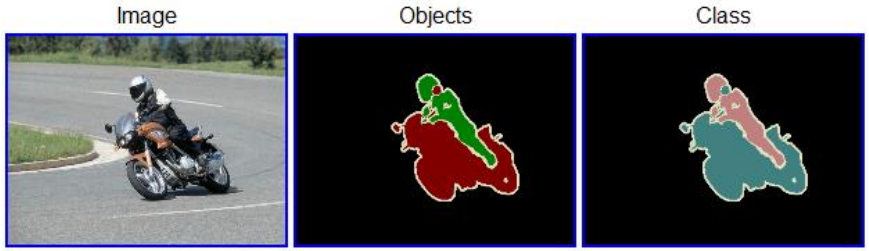


Action: riding bicycle

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

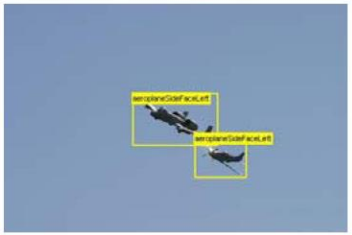
The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

- 20 object categories (aeroplane to TV/monitor)
- Three (+2) challenges:
  - Classification challenge (is there an X in this image?)
  - Detection challenge (draw a box around every X)
  - Segmentation challenge (which class is each pixel?)



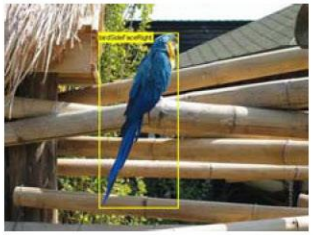
Examples

Aeroplane

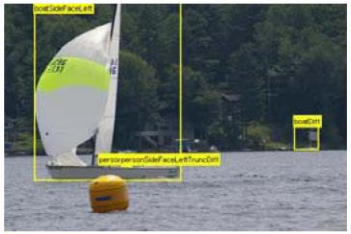


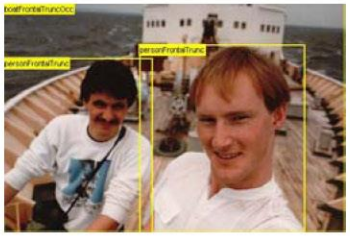
Bicycle

Bird



Boat





Bottle

Bus

Car

Cat

Chair

Cow

Classification Challenge

- Predict whether at least one object of a given class is present in an image

is there a cat?

Precision / Recall for a Category X

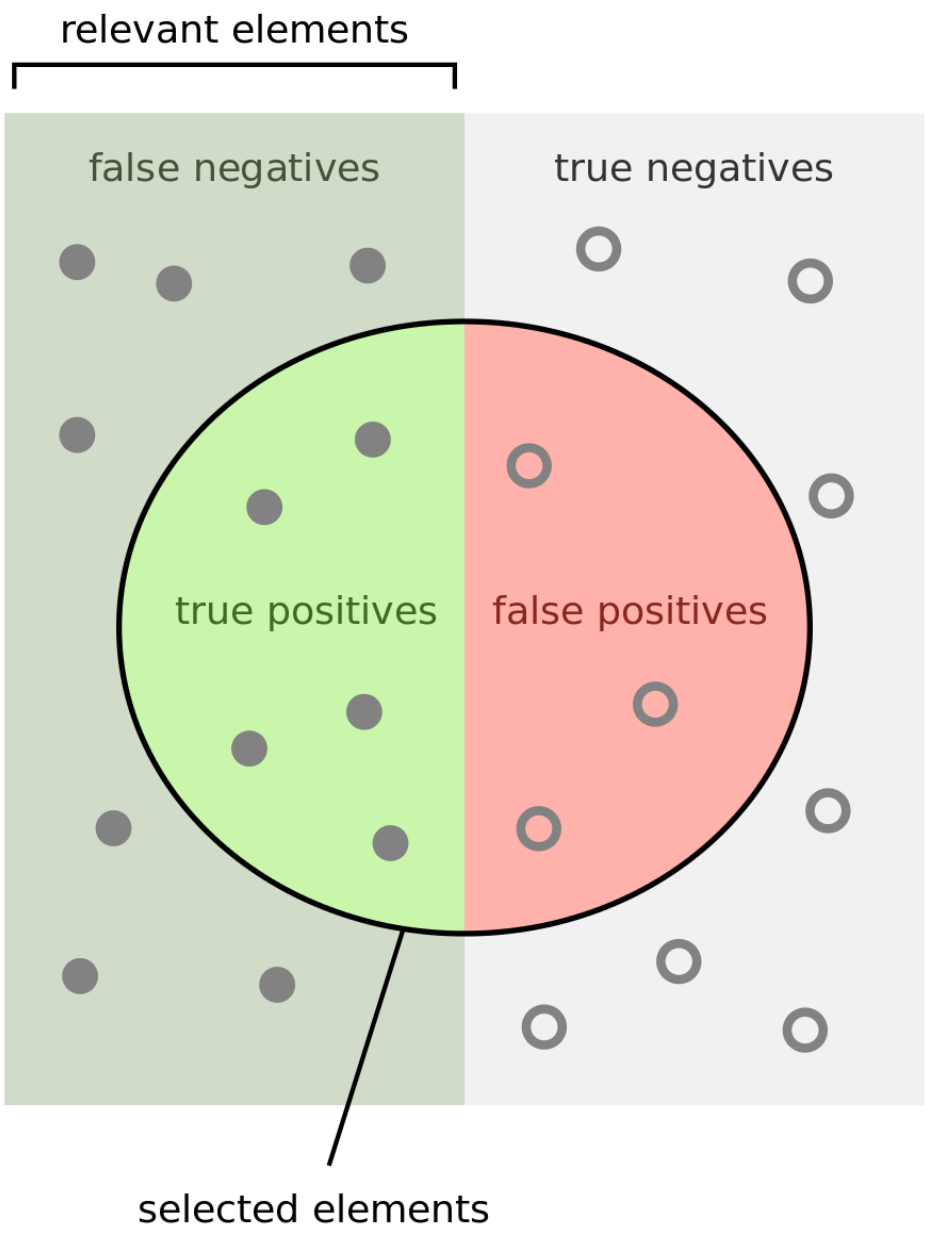
- Precision:

$$\frac{|\{\text{images that contain an X}\}| \cap |\{\text{images classified as X}\}|}{|\{\text{images classified as X}\}|}$$

- Recall:

$$\frac{|\{\text{images that contain an X}\}| \cap |\{\text{images classified as X}\}|}{|\{\text{images that contain an X}\}|}$$

- In reality, methods give a continuous-valued score for each image / category ➔ PR curve



How many selected items are relevant?

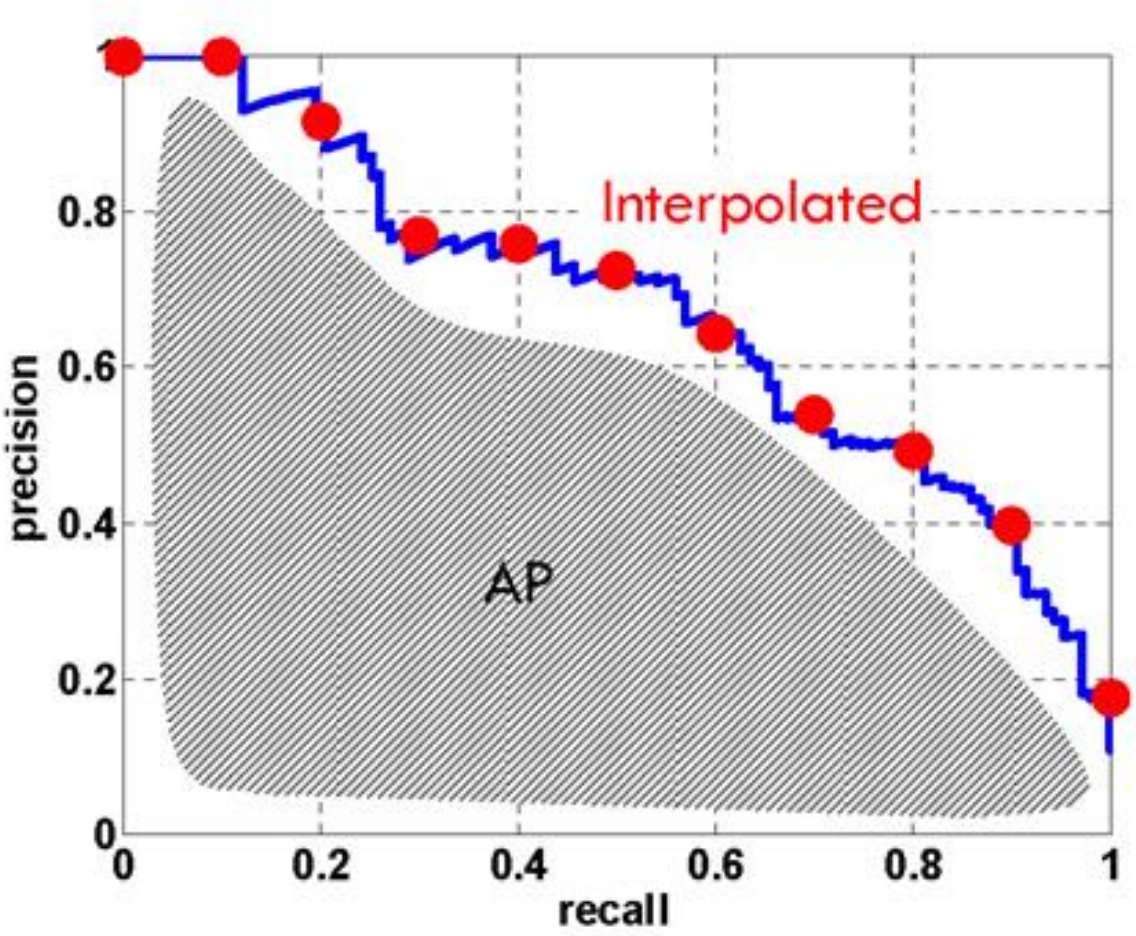
$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

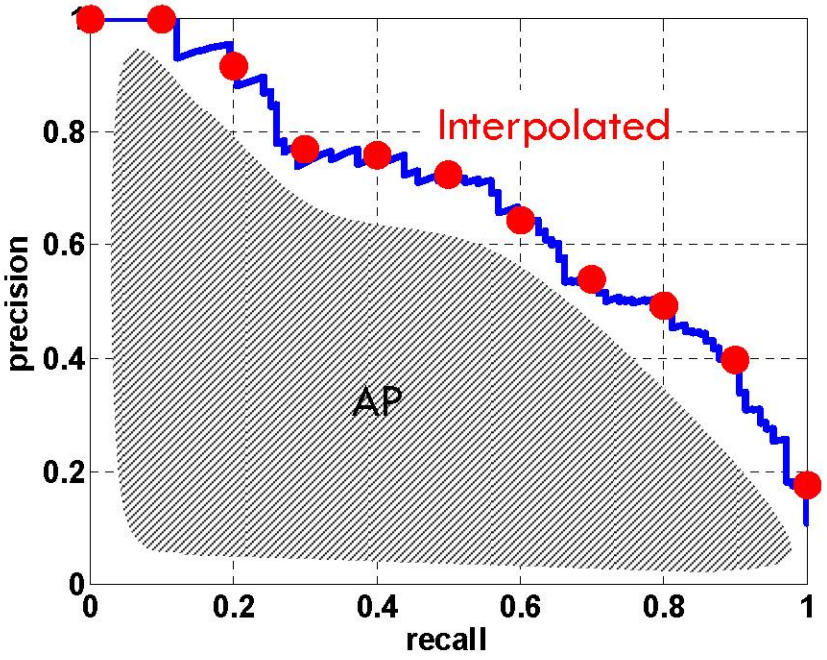
Precision / Recall Curve

- Similar to the ROC curves you saw in Project 2



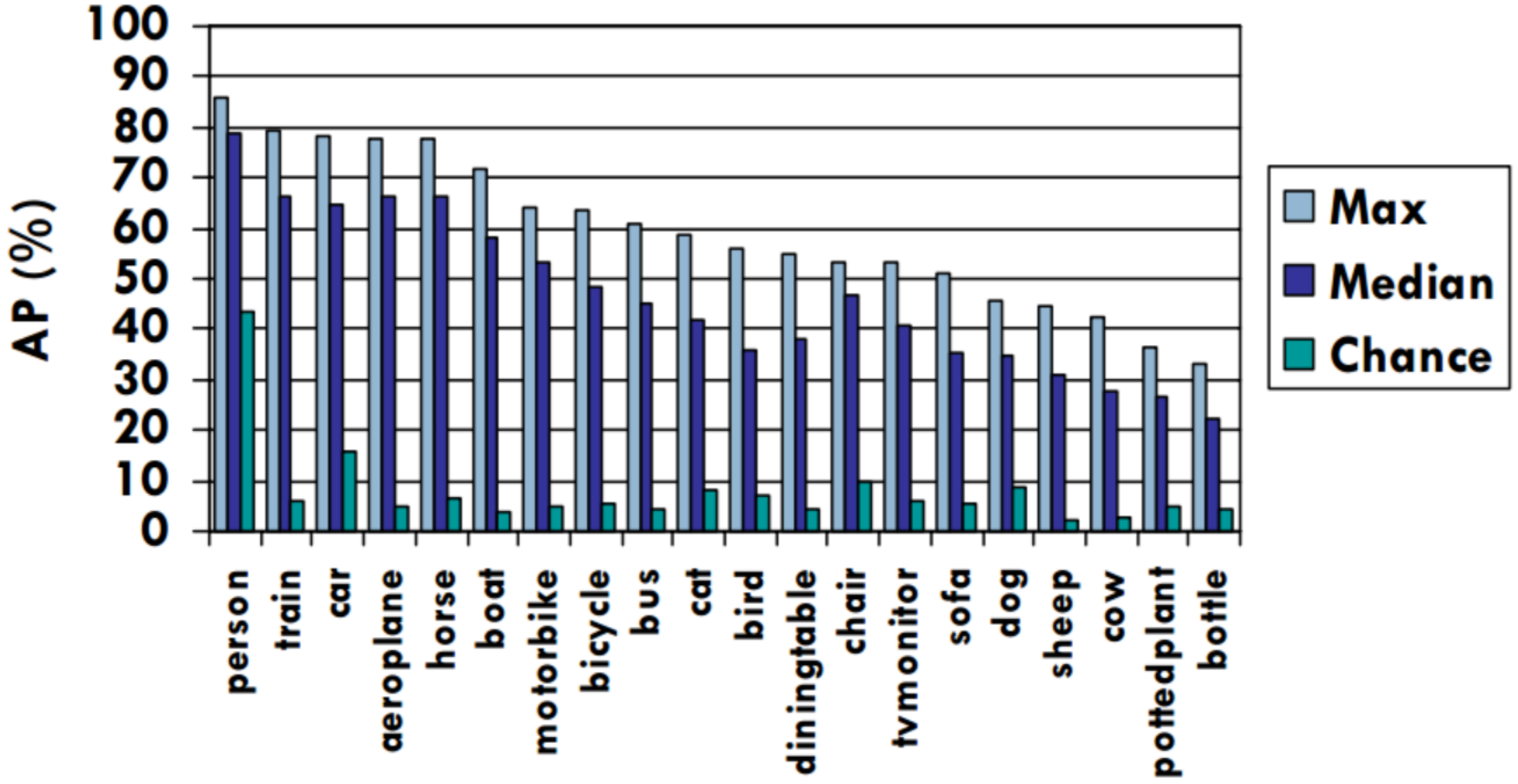
Evaluation

- Average Precision [TREC] averages precision over the entire range of recall
  - Curve interpolated to reduce influence of "outliers"

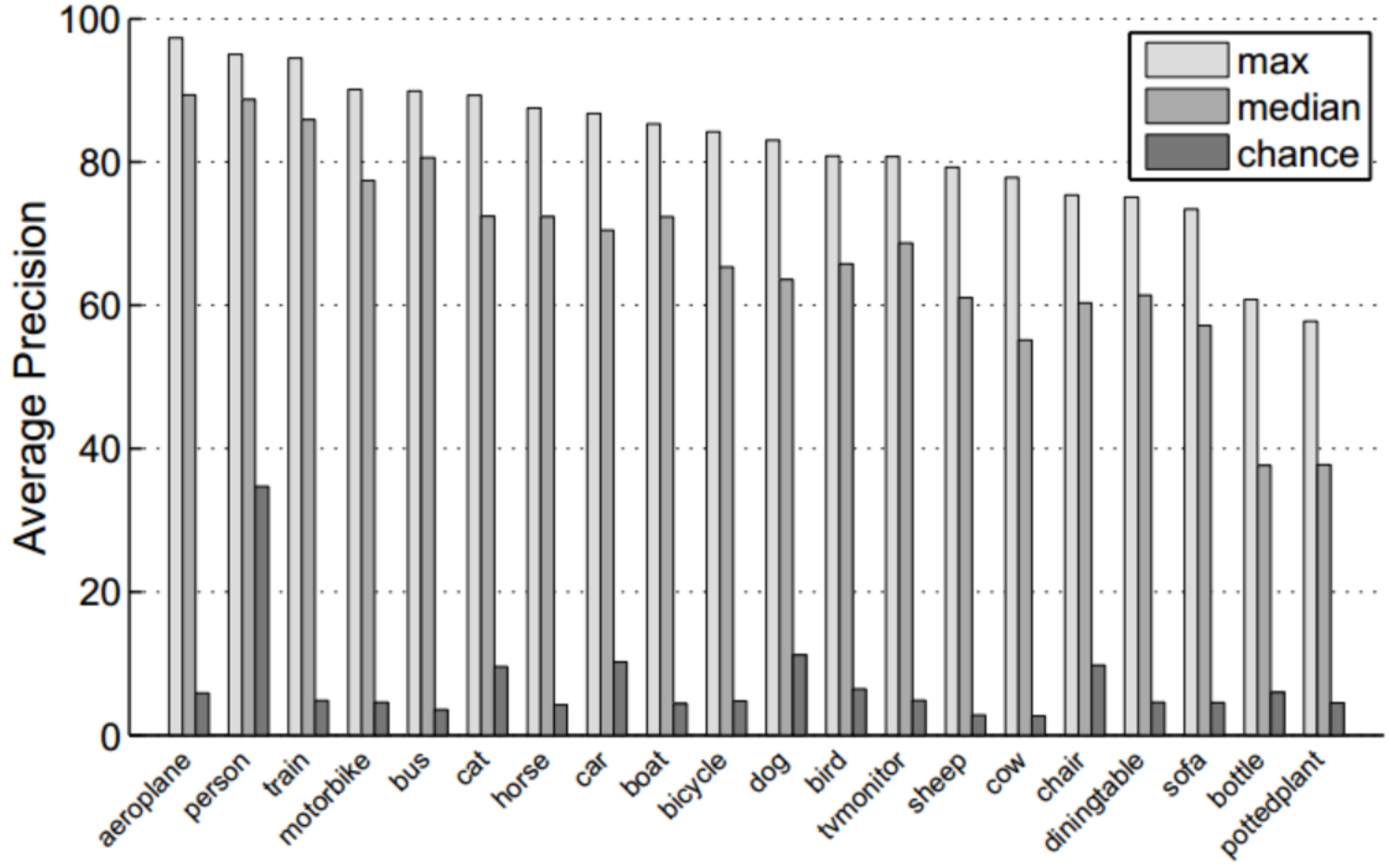


- A good score requires both high recall and high precision
- Application-independent
- Penalizes methods giving high precision but low recall

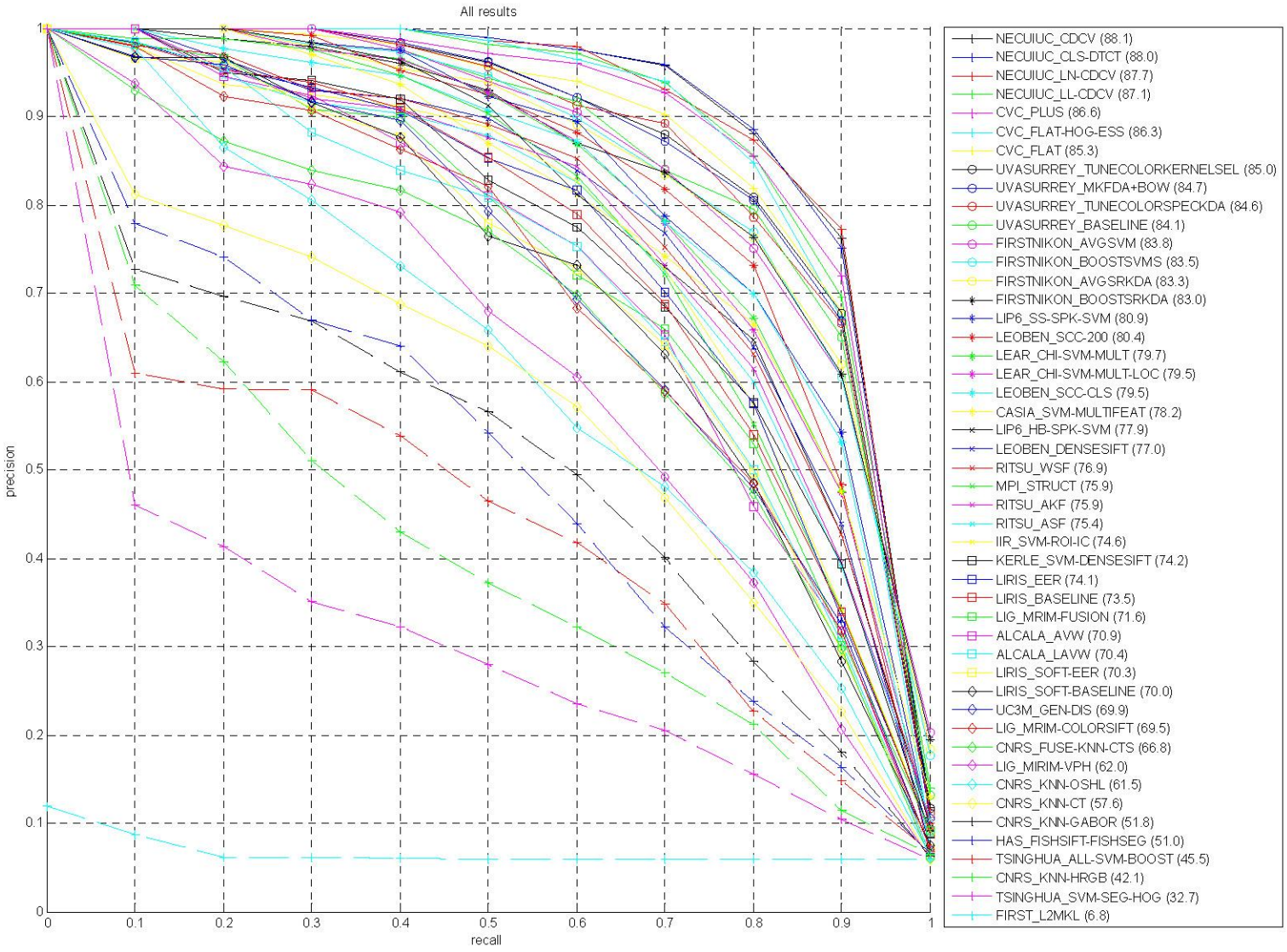
Pascal VOC 2007 Average Precision



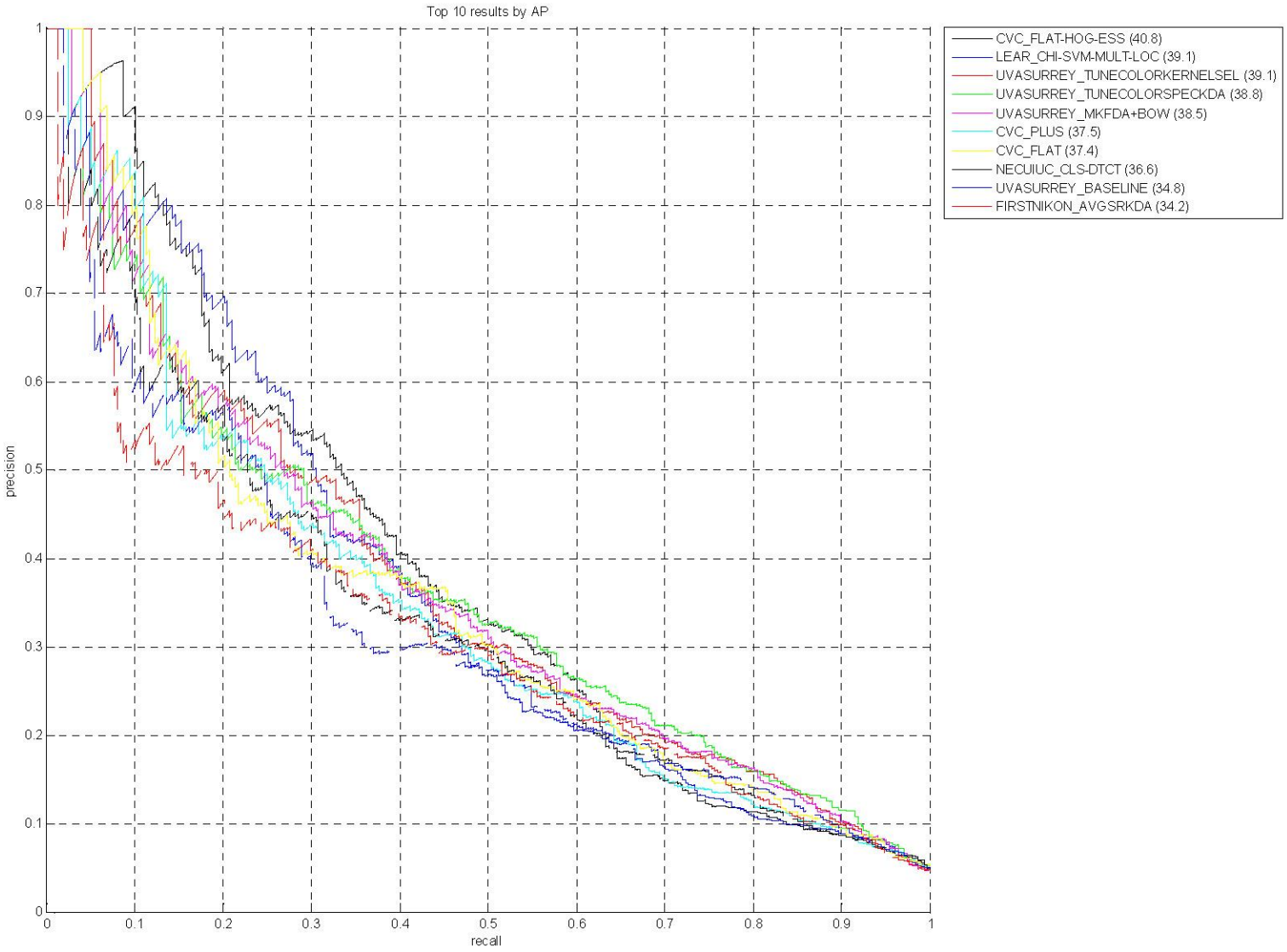
Pascal VOC 2012 Average Precision



Precision/Recall: Aeroplane (All)



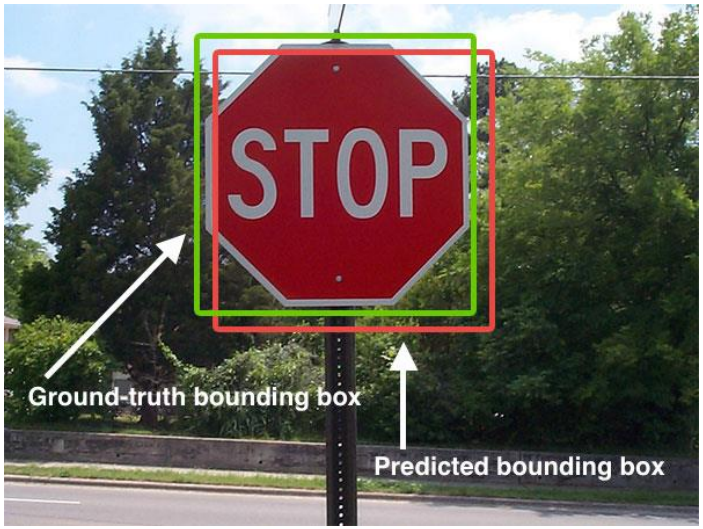
Precision/Recall: Potted plant (Top 10 by AP)



Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)

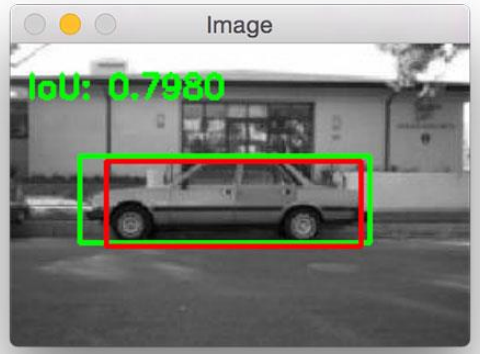
“Intersection over union” (IoU) score



$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Source: https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

“Intersection over union” (IoU) score



Source: https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

True Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT

False Positives - Person

UoCTTI_LSVM-MDPM



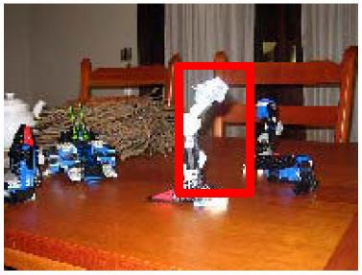


MIZZOU_DEF-HOG-LBP





NECUIUC_CLS-DTCT


“Near Misses” - Person

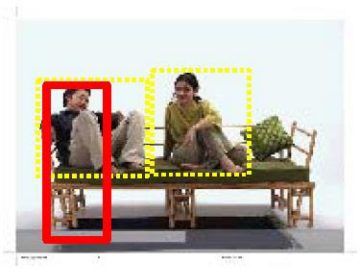
UoCTTI_LSVM-MDPM





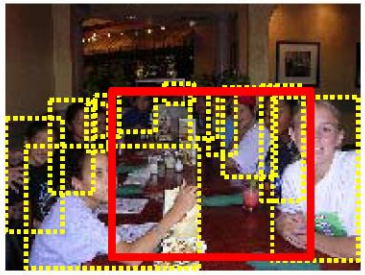
MIZZOU_DEF-HOG-LBP




NECUIUC_CLS-DTCT





True Positives - Bicycle

UoCTTI_LSVM-MDPM





OXFORD_MKL





NECUIUC_CLS-DTCT





False Positives - Bicycle

UoCTTI_LSVM-MDPM



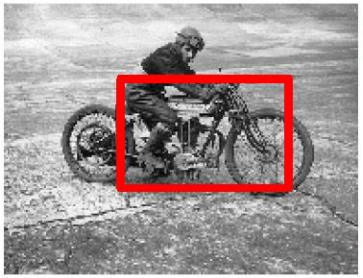


OXFORD_MKL





NECUIUC_CLS-DTCT

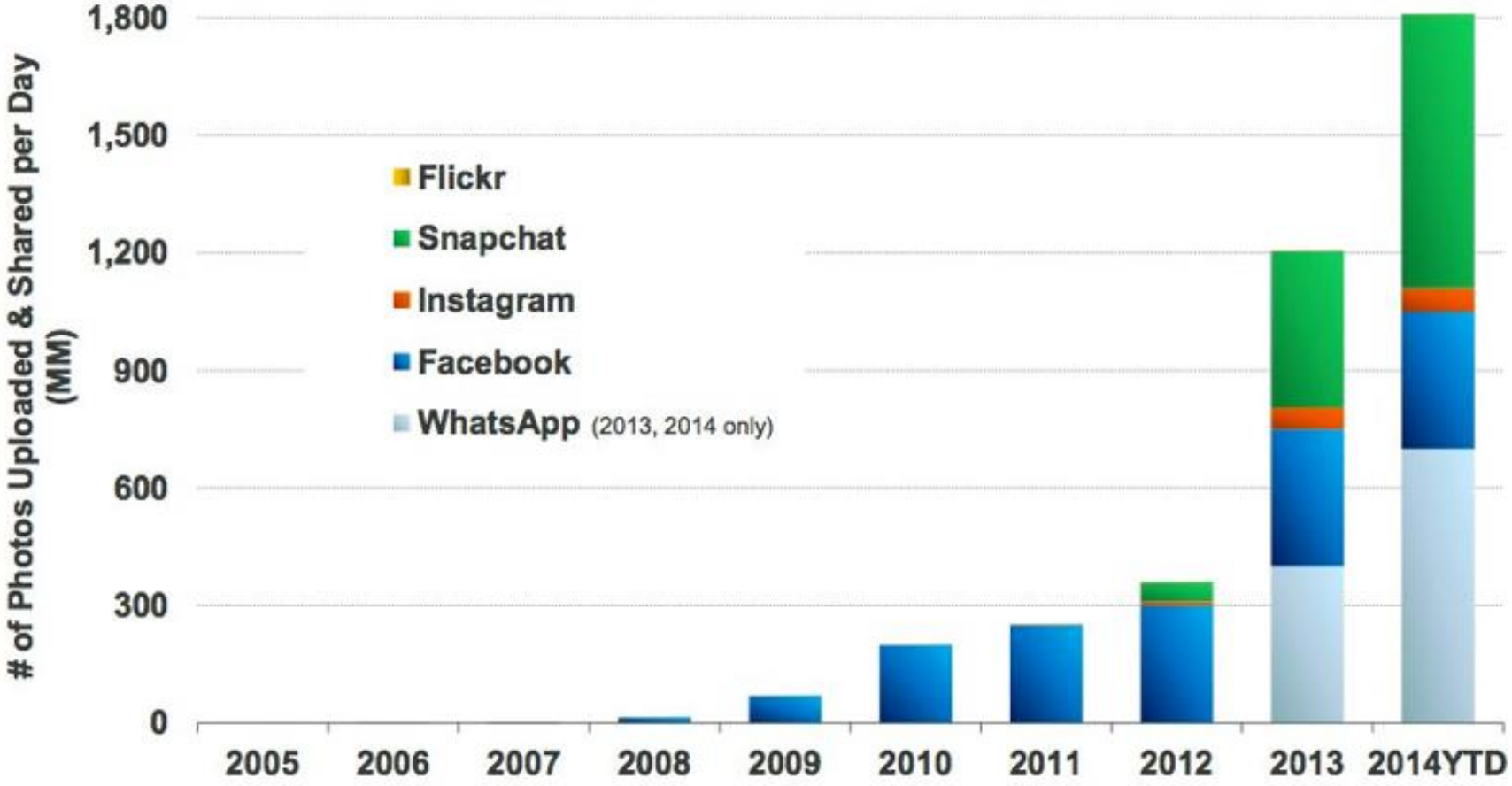

Where to from here?

- Scene Understanding
  - Big data – lots of images
  - Crowd-sourcing – lots of people
  - Deep Learning – lots of compute

24 Hrs in Photos

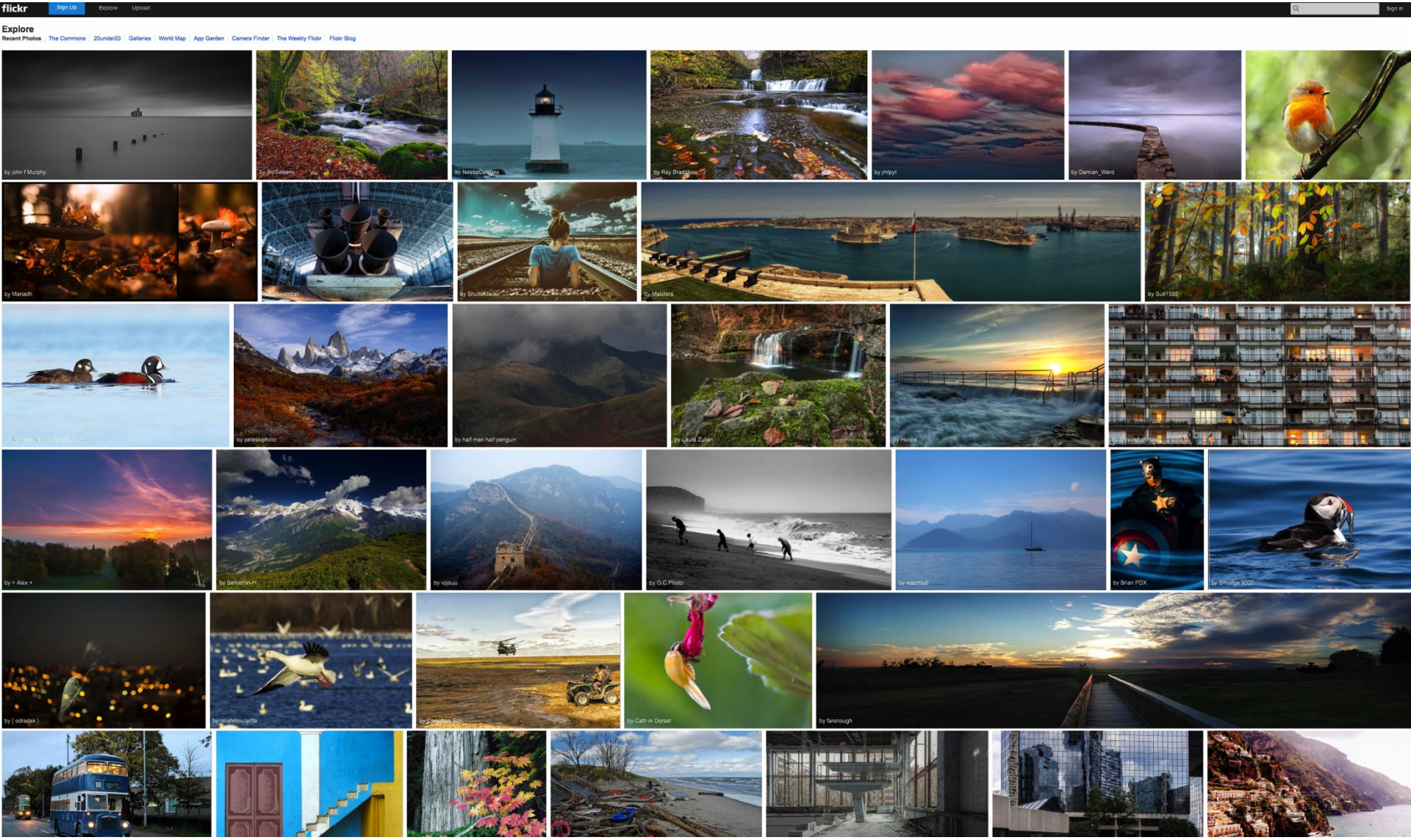
http://www.kesselskramer.com/exhibitions/24-hrs-of-photos installation by Erik Kessels

Daily Number of Photos Uploaded & Shared on Select Platforms, 2005 – 2014YTD



@KPCB

Source: KPCB estimates based on publicly disclosed company data, 2014 YTD data per latest as of 5/14.


flickr
Sign Up
Explore
Upload
Sign In
Explore
Recent Photos | The Commons | 20under20 | Galleries | World Map | App Garden | Camera Finder | The Weekly Flickr | Flickr Blog
by john f Murphy
by BySweeny
by NestorDesigns
by Ray Bradshaw
by jmfpyt
by Damian_Ward
by Manadh
by Wombat
by ShutterJack
by Maistora
by Sue1585
by petesiophoto
by half man half penguin
by Laura Zulian
by Hoops
by + Alex +
by Benjamin-H
by vpjkuu
by O.C Photo
by wazimu0
by Brian PDX
by Smudge 9000
by (odradek)
by reishmountaffe
by Caith in Dorset
by farenough

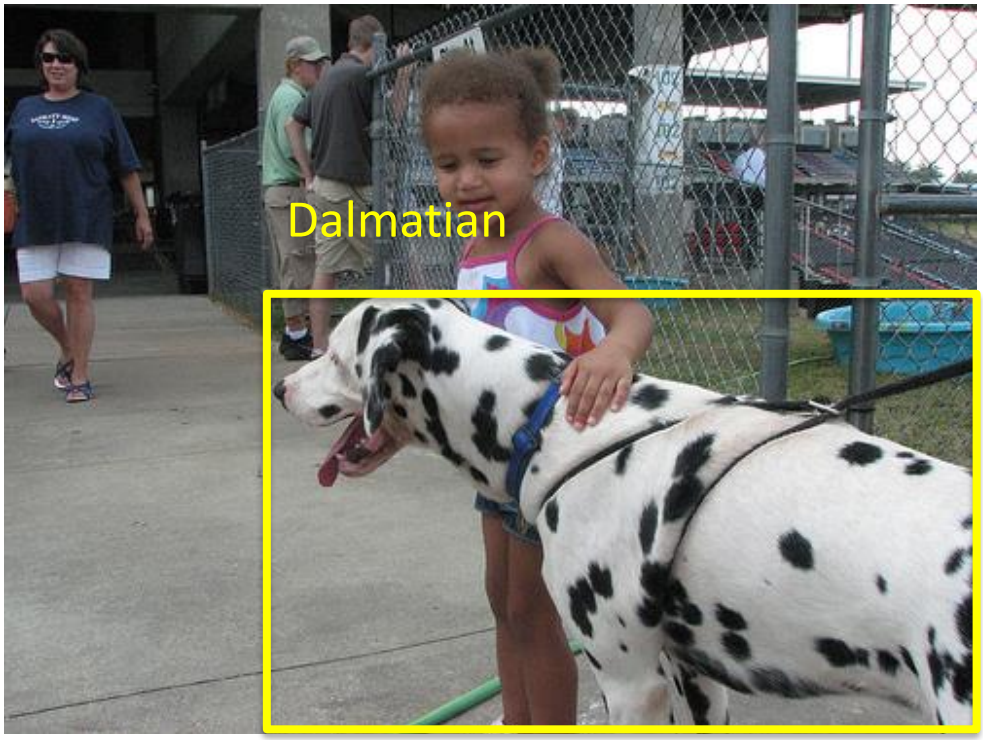
Data Sets

- ImageNet
  - Huge, Crowdsourced, Hierarchical, *Iconic* objects
- PASCAL VOC
  - *Not* Crowdsourced, bounding boxes, 20 categories
- SUN Scene Database, Places
  - *Not* Crowdsourced, 397 (or 720) scene categories
- LabelMe (Overlaps with SUN)
  - Sort of Crowdsourced, Segmentations, Open ended
- SUN *Attribute* database (Overlaps with SUN)
  - Crowdsourced, 102 attributes for every scene
- OpenSurfaces
  - Crowdsourced, materials
- Microsoft COCO
  - Crowdsourced, large-scale objects

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2010-2012

~~20 object classes~~ ~~22,591 images~~

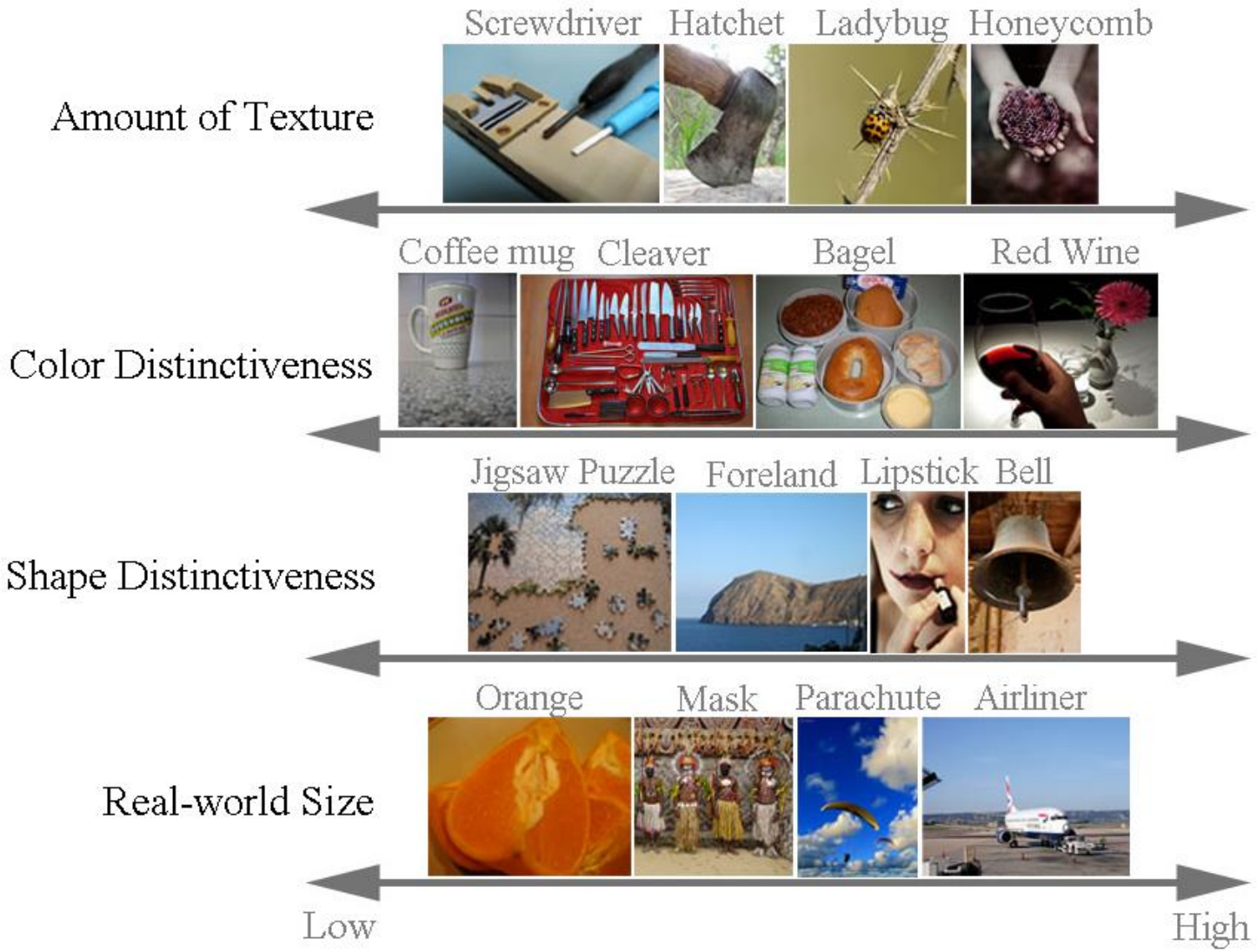
1000 object classes **1,431,167 images**

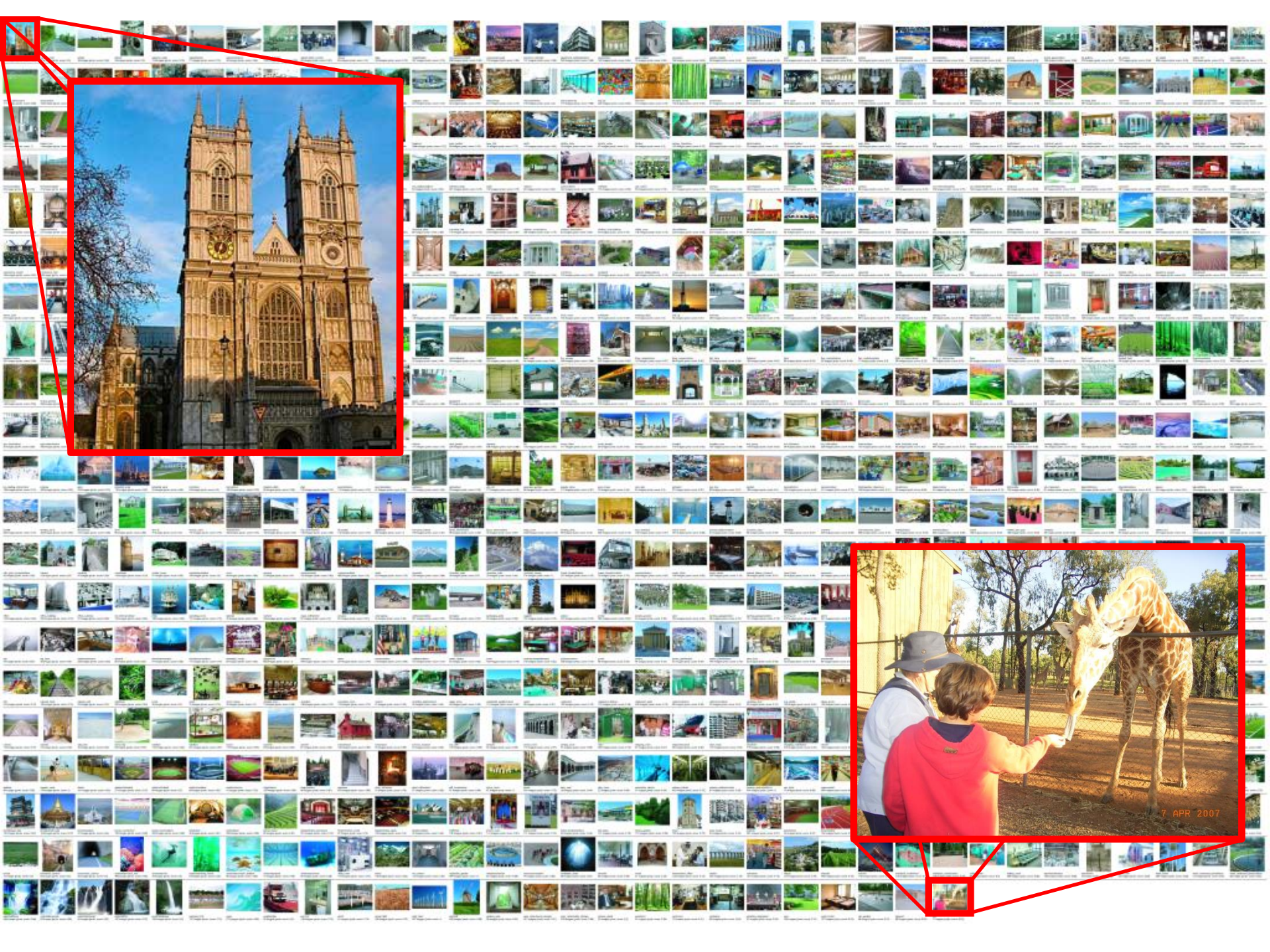


http://image-net.org/challenges/LSVRC/{2010,2011,2012}

Variety of object classes in ILSVRC

Variety of object classes in ILSVRC




7 APR 2007

What are attributes?

What do we want to know about this object?

Object recognition expert: “Dog”

Next step: Infer object properties

Can I **poke with it**?

Is it **alive**?

What **shape** is it?

Does it have a **tail**?

Can I **put stuff in it**?

Is it **soft**?

Will it **blend**?

What are attributes?

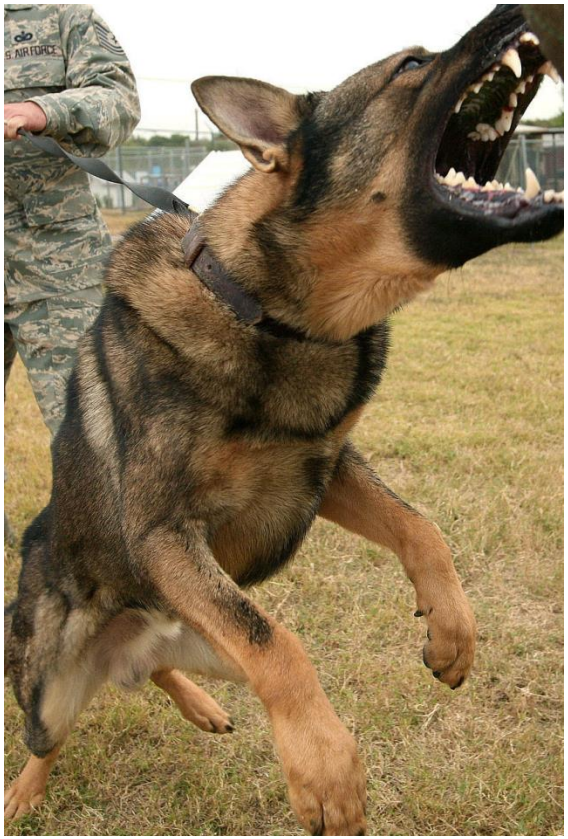
What do we want to know about this object?

Object recognition expert:
“Dog”

Person in the Scene:
“Big pointy teeth”, “Can move fast”, “Looks angry”

Why infer properties

1. We want detailed information about objects



"Dog"
vs.
"Large, angry animal with pointy teeth"

Why infer properties

2. We want to be able to infer something about unfamiliar objects

Familiar Objects



New Object

Why infer properties

2. We want to be able to infer something about unfamiliar objects

If we can infer properties…

Familiar Objects

Has Stripes
Has Ears
Has Eyes
….

Has Four Legs
Has Mane
Has Tail
Has Snout
….

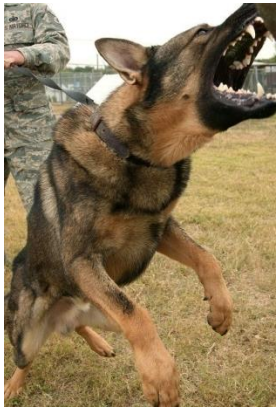
Brown
Muscular
Has Snout
….

New Object

Has Stripes (like cat)
Has Mane and Tail (like horse)
Has Snout (like horse and dog)

Why infer properties

3. We want to make comparisons between objects or categories



What is unusual about this dog?

What is the difference between horses and zebras?

Questions?