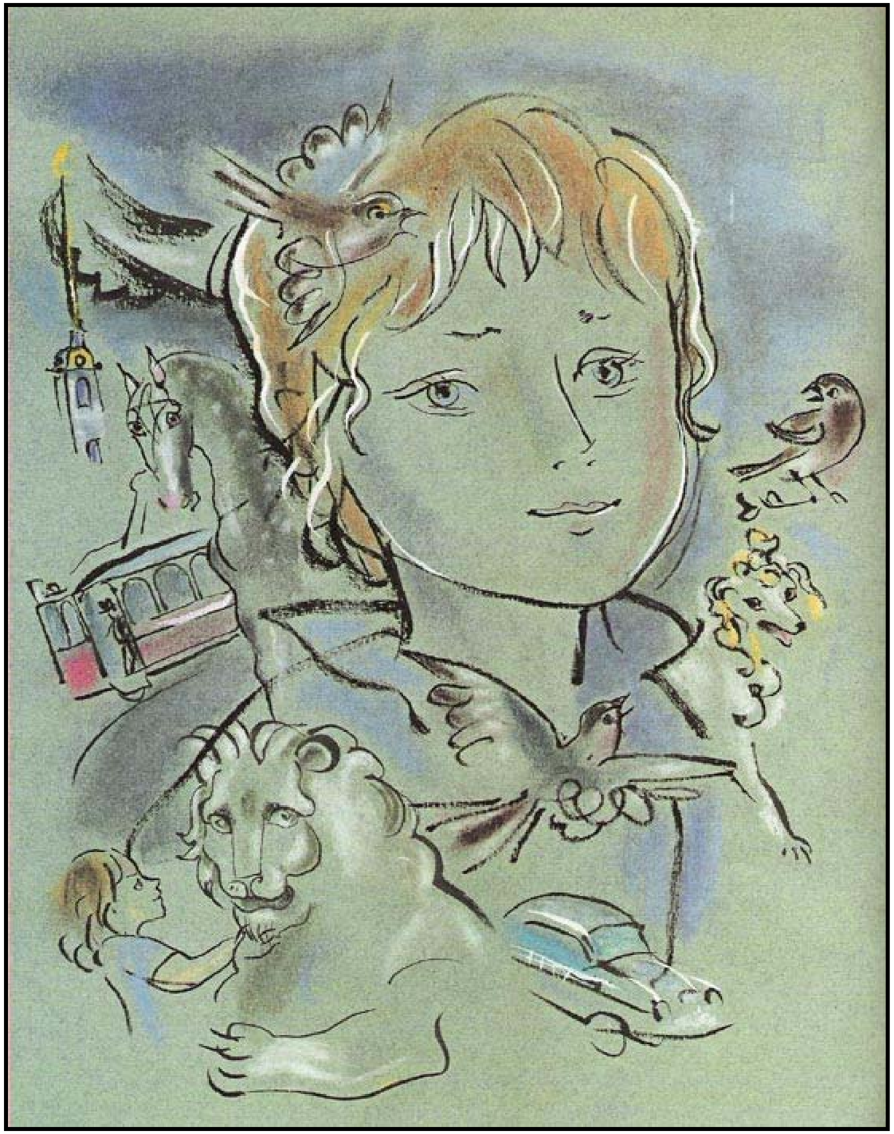
Object Recognition: Conceptual Issues



Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and K. Grauman

Issues in recognition

The statistical viewpoint

Generative vs. discriminative methods

Model representation

Generalization, bias vs. variance

Supervision

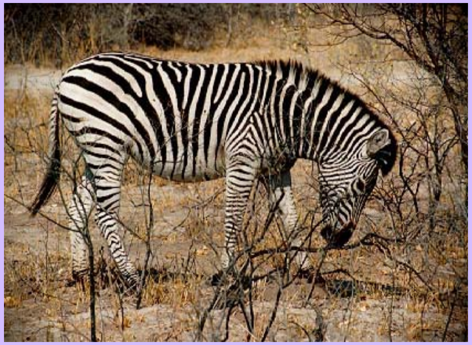
Datasets

Object categorization: the statistical viewpoint

- MAP decision: $p(zebra \mid image)$

vs.

$$p(no\ zebra \mid image)$$

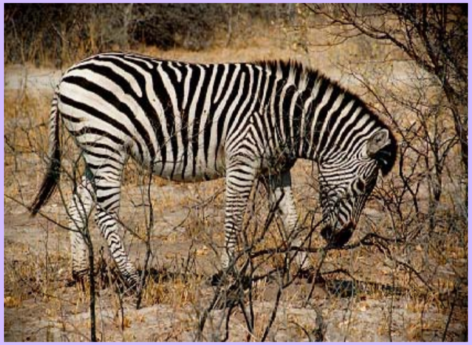


Object categorization: the statistical viewpoint

- MAP decision: $p(zebra \mid image)$

vs.

$$p(no\ zebra \mid image)$$



- Bayes rule:

$$\underbrace{p(zebra \mid image)}_{\text{posterior}} \propto \underbrace{p(image \mid zebra)}_{\text{likelihood}} \underbrace{p(zebra)}_{\text{prior}}$$

Object categorization: the statistical viewpoint

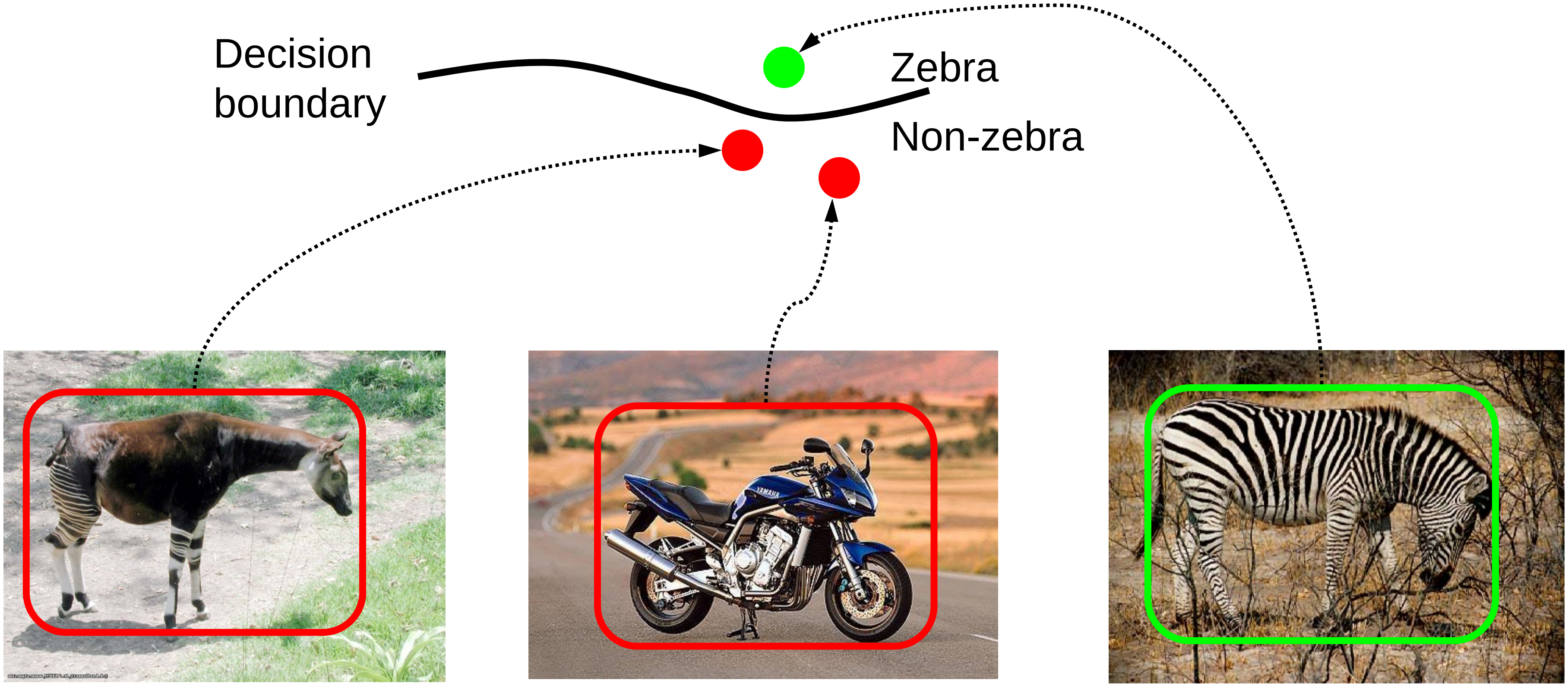
$$\underbrace{p(zebra \mid image)}_{\text{posterior}} \propto \underbrace{p(image \mid zebra)}_{\text{likelihood}}\,\underbrace{p(zebra)}_{\text{prior}}$$

- **Discriminative methods:** **model posterior**

- **Generative methods:** **model likelihood and prior**

Discriminative methods

- Direct modeling of $p(zebra \mid image)$



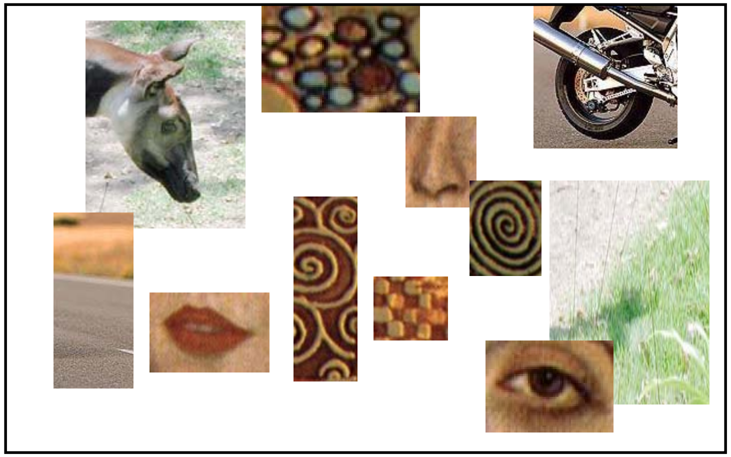
Decision boundary

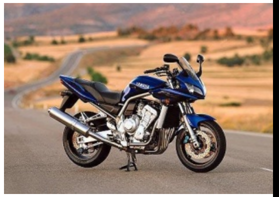
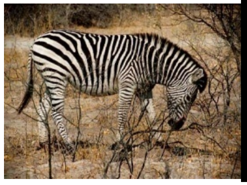
Zebra

Non-zebra

Generative methods

- Model $p(image \mid zebra)$ and $p(image \mid no\ zebra)$



	$p(image \mid zebra)$	$p(image \mid no\ zebra)$
	Low	Middle
	High	Middle→Low

Generative vs. discriminative methods

- Generative methods
  - + Interpretable
  - + Can be learned using images from just a single category
  - – Sometimes we don’t need to model the likelihood when all we want is to make a decision

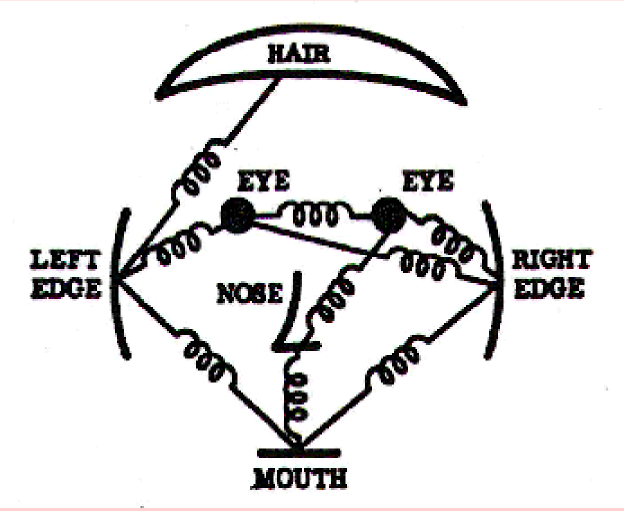
- Discriminative methods
  - + Efficient
  - + Often produce better classification rates
  - – Can be hard to interpret
  - – Require positive and negative training data

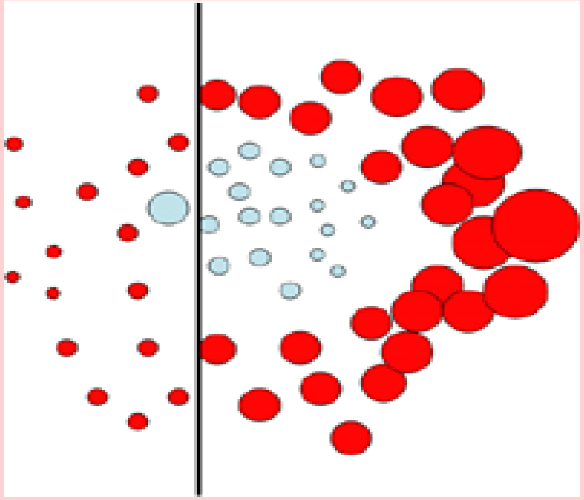
Steps for statistical recognition

- Representation
  - How to model an object category
- Learning
  - How to find the parameters of the model, given training data
- Recognition
  - How the model is to be used on novel data

Representation

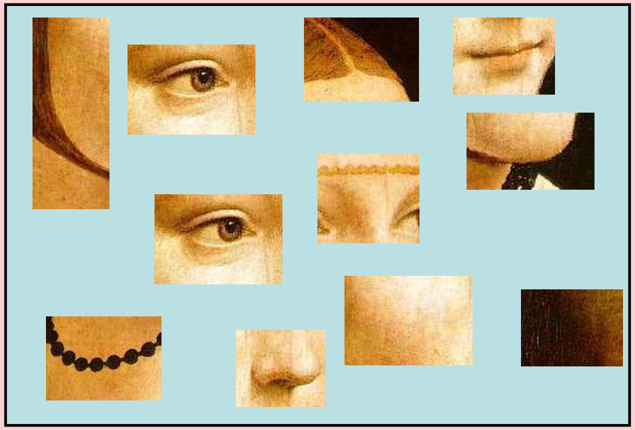
- Generative / discriminative / hybrid

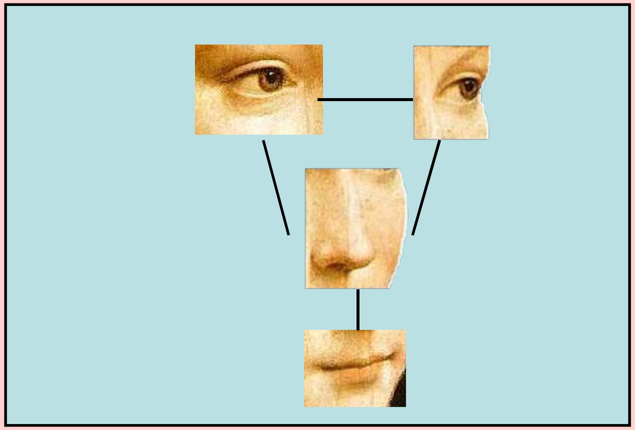




Representation

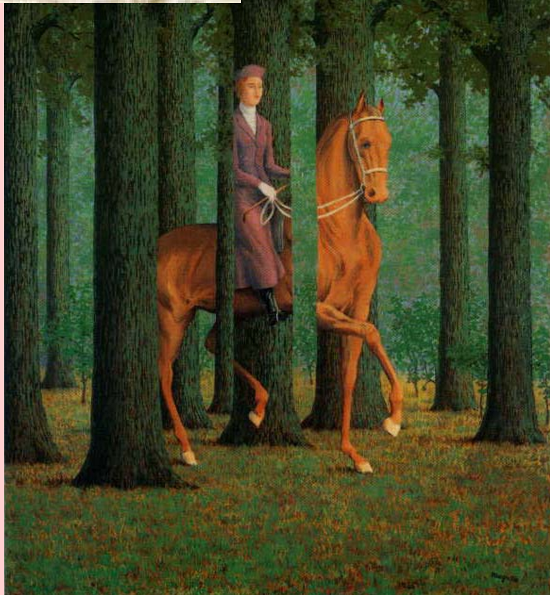
- Generative / discriminative / hybrid
- Appearance only or location and appearance





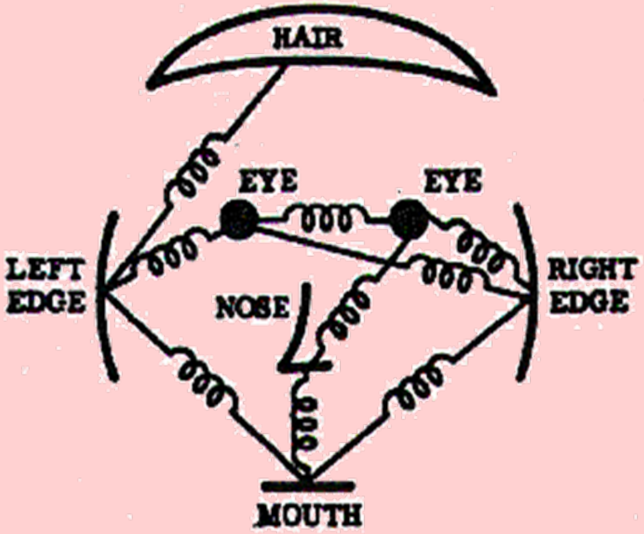
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
  - Viewpoint
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



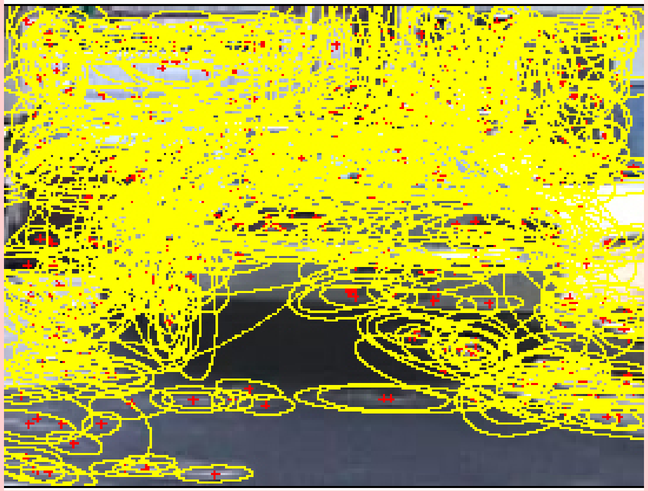
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
- Global, sliding window, part-based



Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
- Global, sliding window, part-based
- If part-based, what is the spatial support for parts?

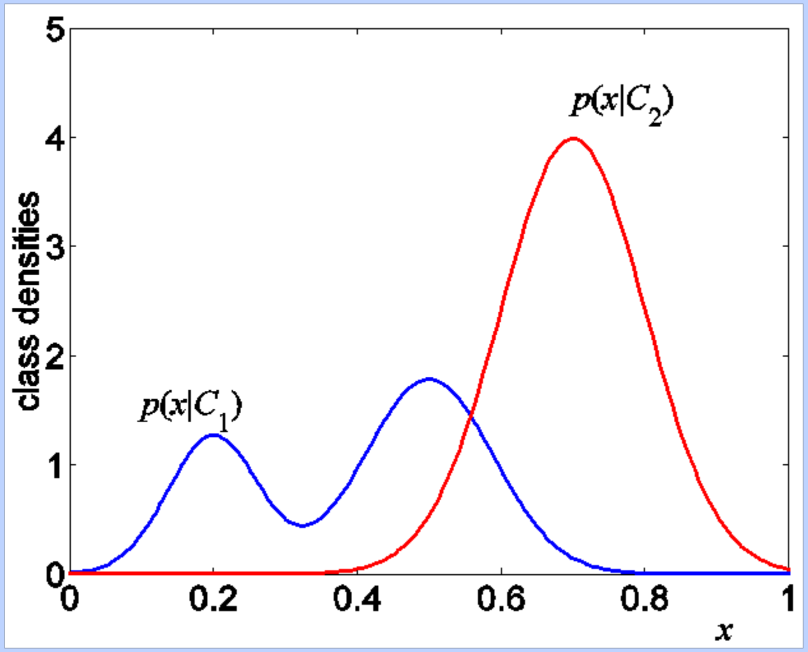


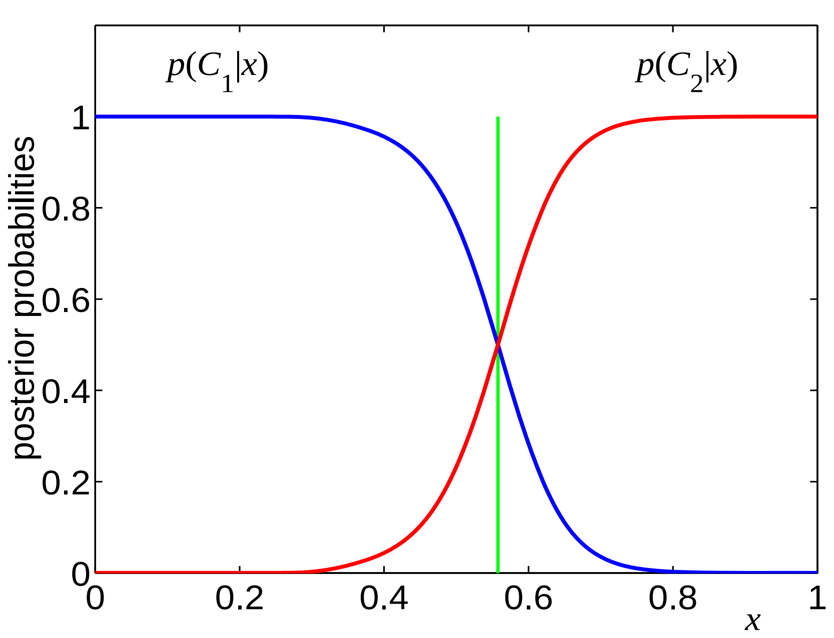
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative





Learning

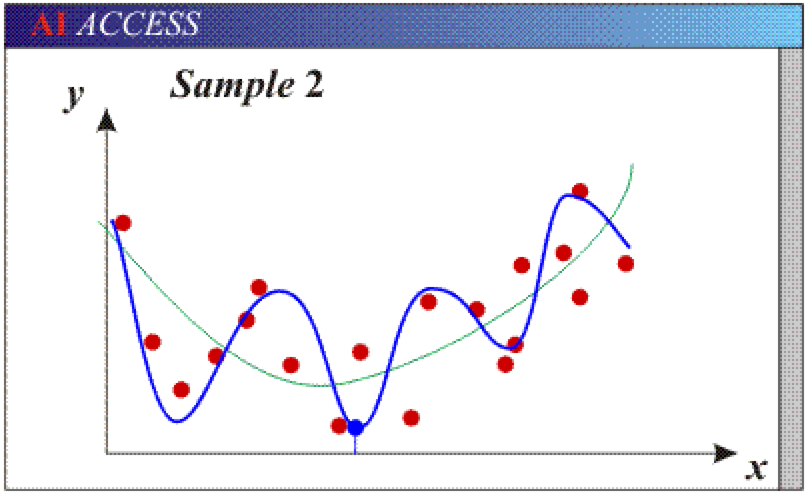
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance

Generalization

- How well does a learned model generalize from the data it was trained on to a new test set?

Bias-variance tradeoff

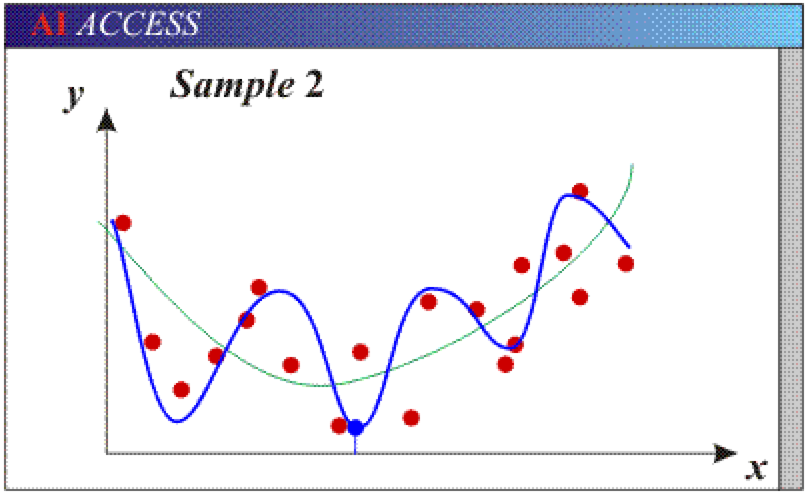
- Models with too many parameters may fit a given sample better, but have high *variance*
- Generalization error is due to *overfitting*

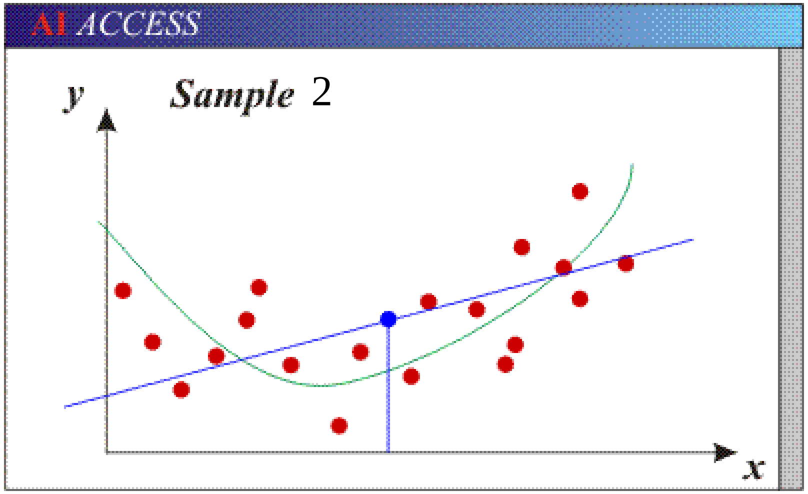


Bias-variance tradeoff

- Models with too many parameters may fit a given sample better, but have high *variance*
- Generalization error is due to *overfitting*

- Models with too few parameters may not fit a given sample well because of high *bias*
- Generalization error is due to *underfitting*





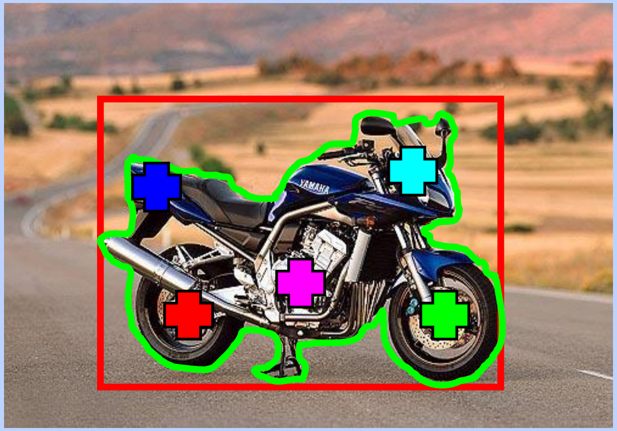
Occam’s razor

- Given several models that describe the data equally well, the simpler one should be preferred
- There should be some tradeoff between error and model complexity
  - This is rarely done rigorously, but is a powerful “rule of thumb”
  - Simpler models are often preferred because of their robustness (= low variance)

Learning

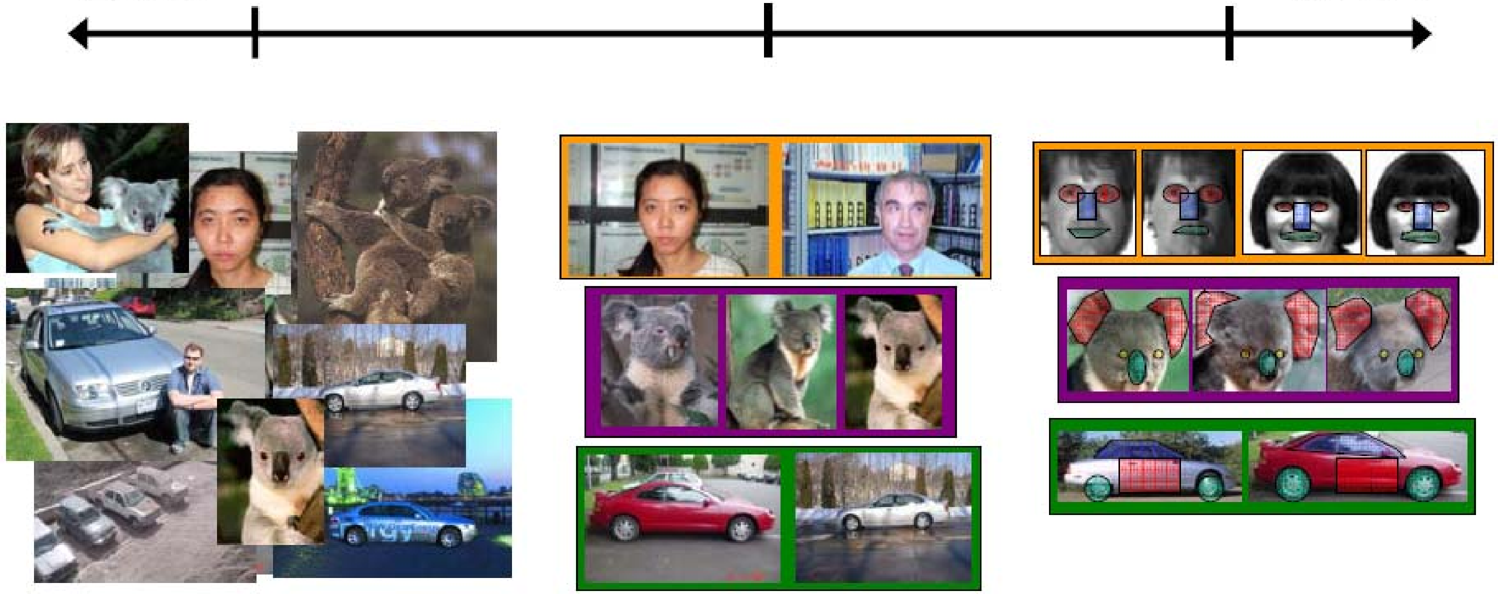
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

Contains a motorbike



Spectrum of supervision

Less **More**



Unsupervised

“Weakly” supervised

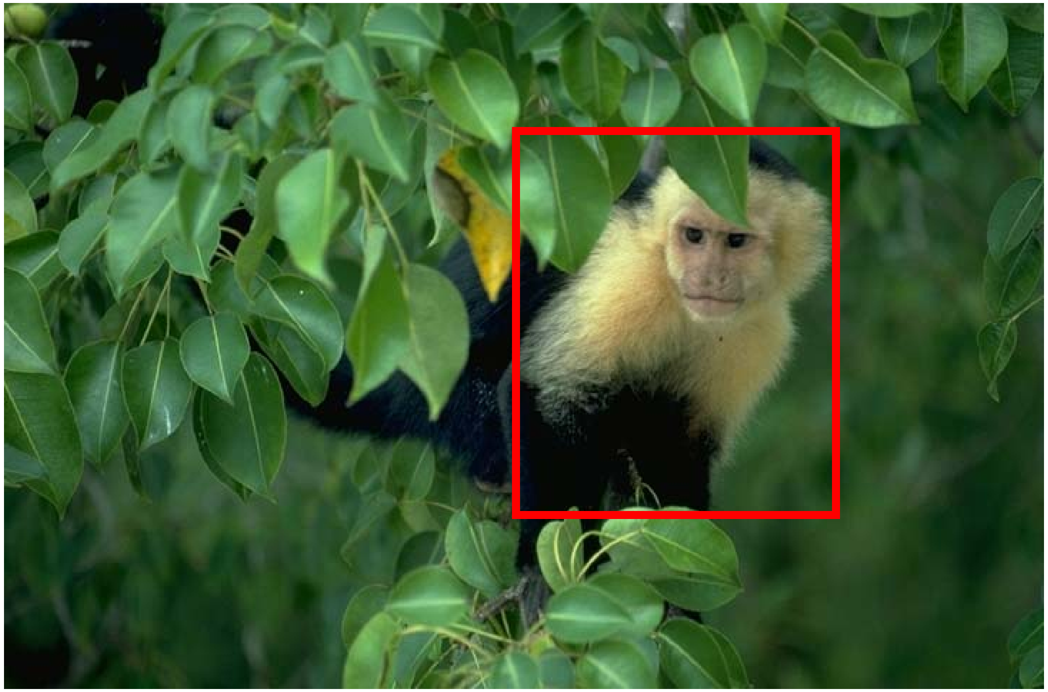
Supervised

Definition depends on task

What task?

- Classification
  - Object present/absent in image
  - Background may be correlated with object

- Localization / Detection
  - Localize object within the frame
  - Bounding box or pixel-level segmentation

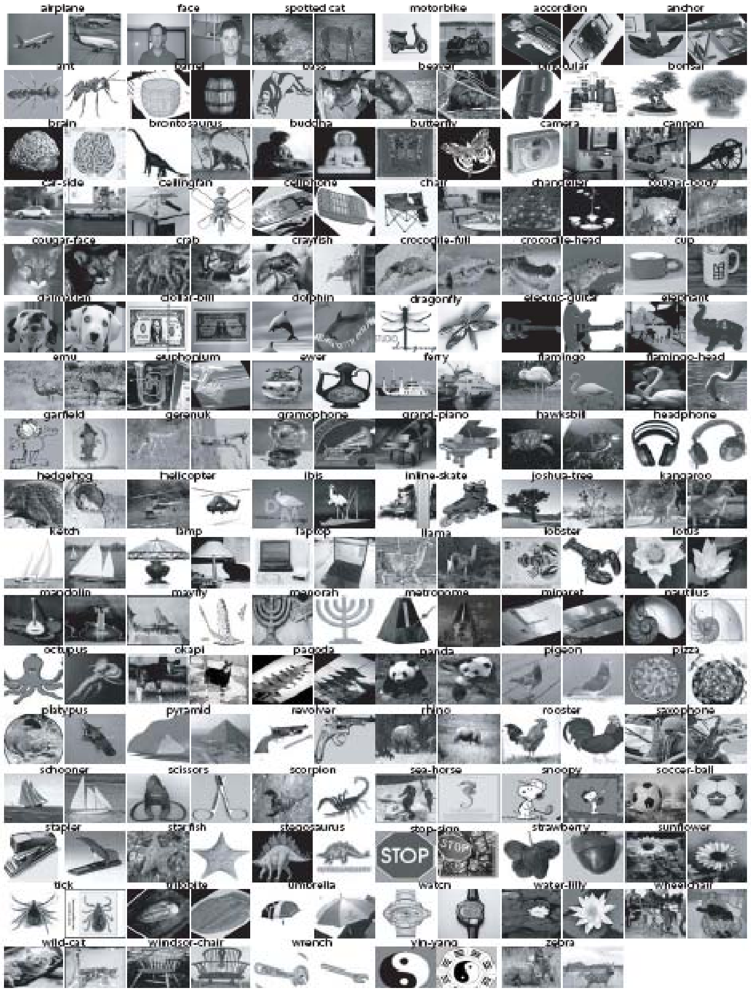


Datasets

- Circa 2001: 5 categories, 100s of images per category
- Circa 2004: 101 categories
- Today: thousands of categories, tens of thousands of images

Caltech 101 & 256

http://www.vision.caltech.edu/Image_Datasets/Caltech101/
http://www.vision.caltech.edu/Image_Datasets/Caltech256/



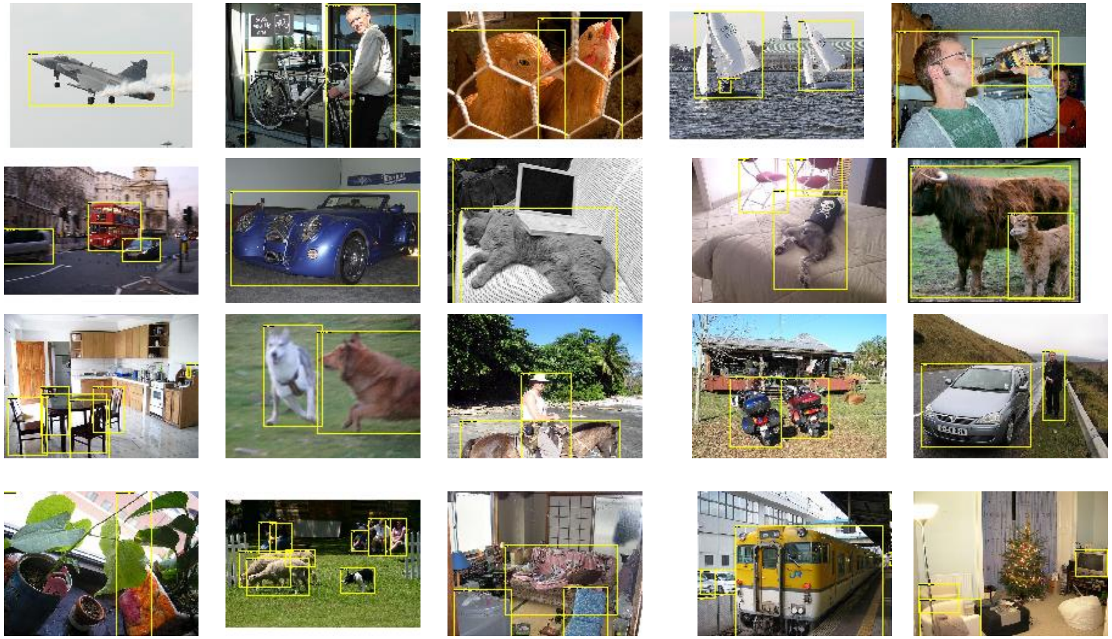
Fei-Fei, Fergus, Perona, 2004

Griffin, Holub, Perona, 2007

The PASCAL Visual Object Classes Challenge (2005-2009)

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

2008 Challenge classes:
Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor



The PASCAL Visual Object Classes Challenge (2005-2009)

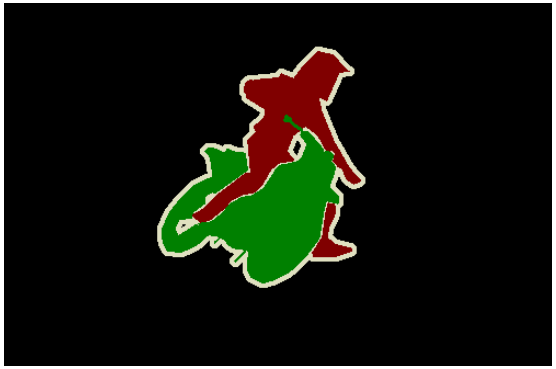
http://pascallin.ecs.soton.ac.uk/challenges/VOC/

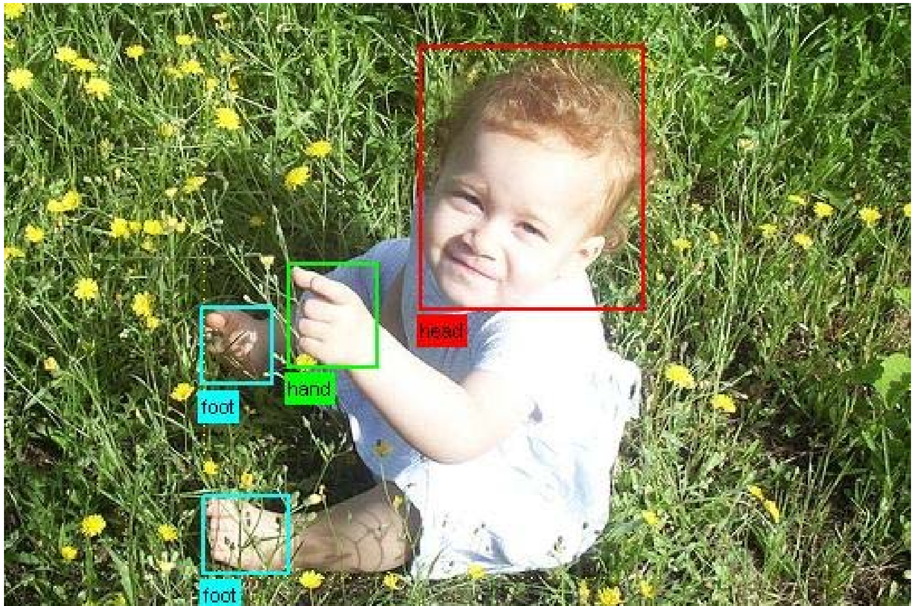
- Main competitions
  - **Classification:** For each of the twenty classes, predicting presence/absence of an example of that class in the test image
  - **Detection:** Predicting the bounding box and label of each object from the twenty target classes in the test image

The PASCAL Visual Object Classes Challenge (2005-2009)

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

- “Taster” challenges
  - **Segmentation:** Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise
  - **Person layout:** Predicting the bounding box and label of each part of a person (head, hands, feet)





Lotus Hill Research Institute image corpus

http://www.imageparsing.com/

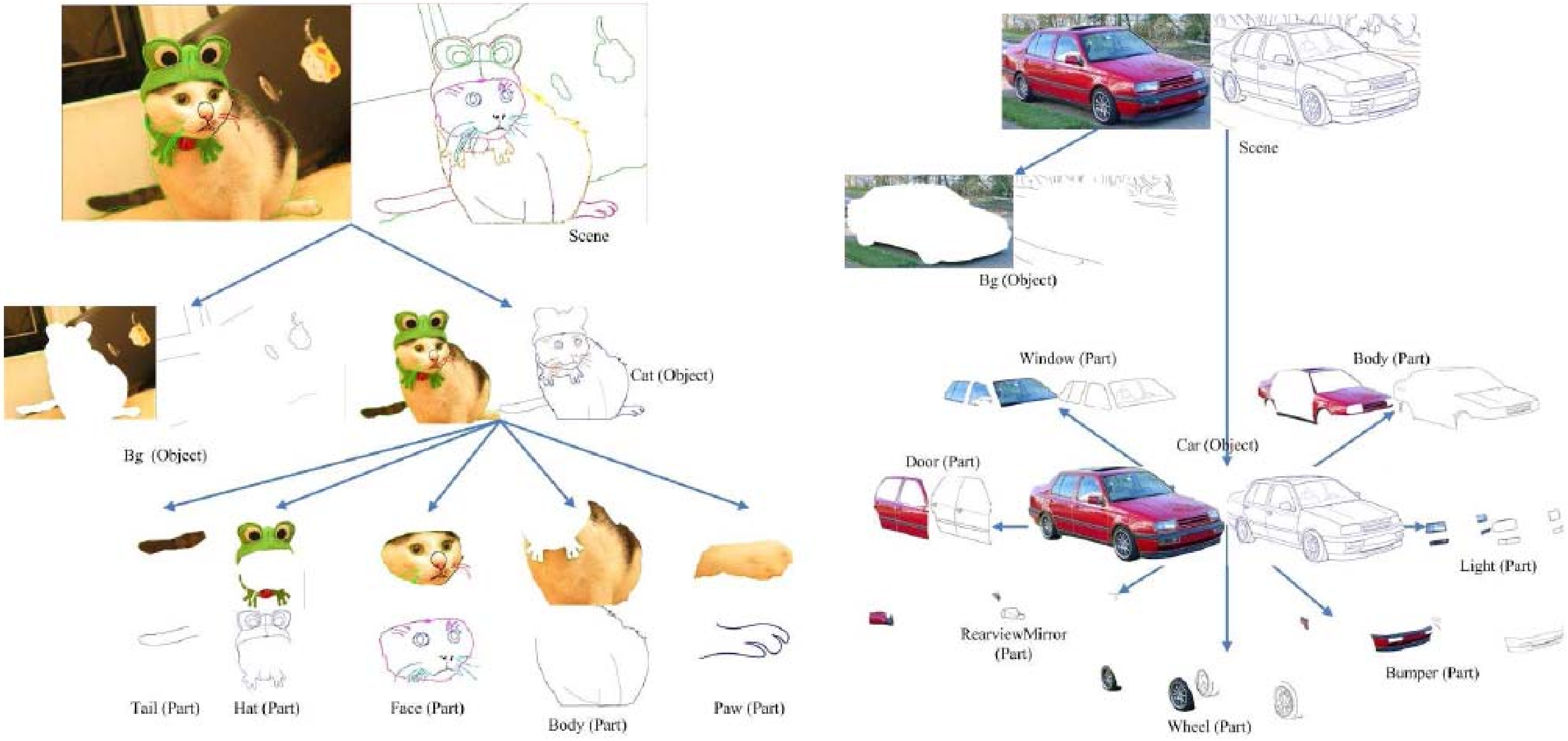
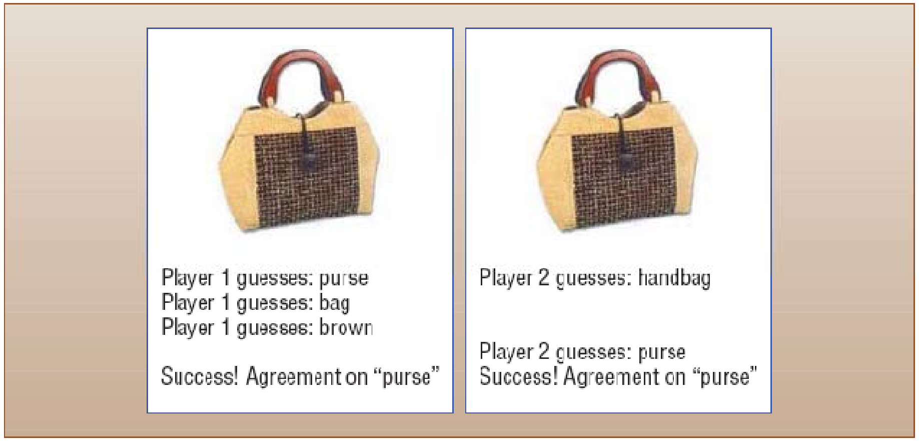


Figure 5: Two examples of the parse trees (cat and car) in the Lotus Hill Research Institute image corpus. From [87].

Z.Y. Yao, X. Yang, and S.C. Zhu, 2007

Labeling with games

http://www.gwap.com/gwap/



Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on “purse”

Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on “purse”

Figure 1. Partners agreeing on an image in the ESP Game. *Neither player can see the other’s guesses.*

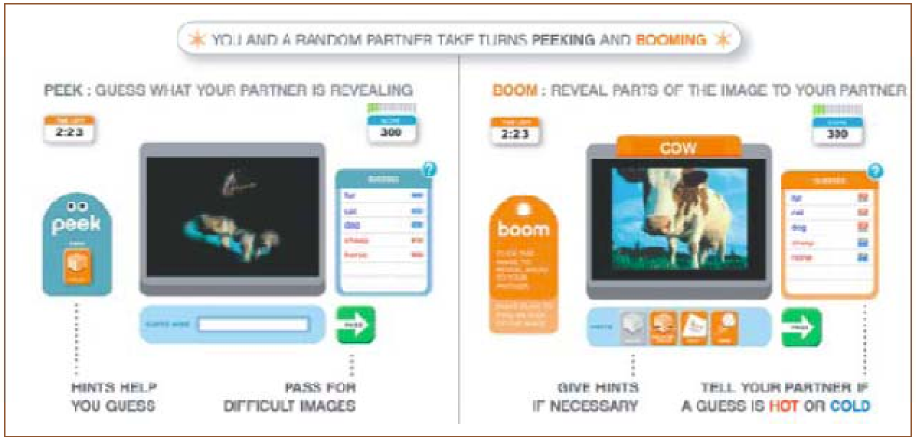
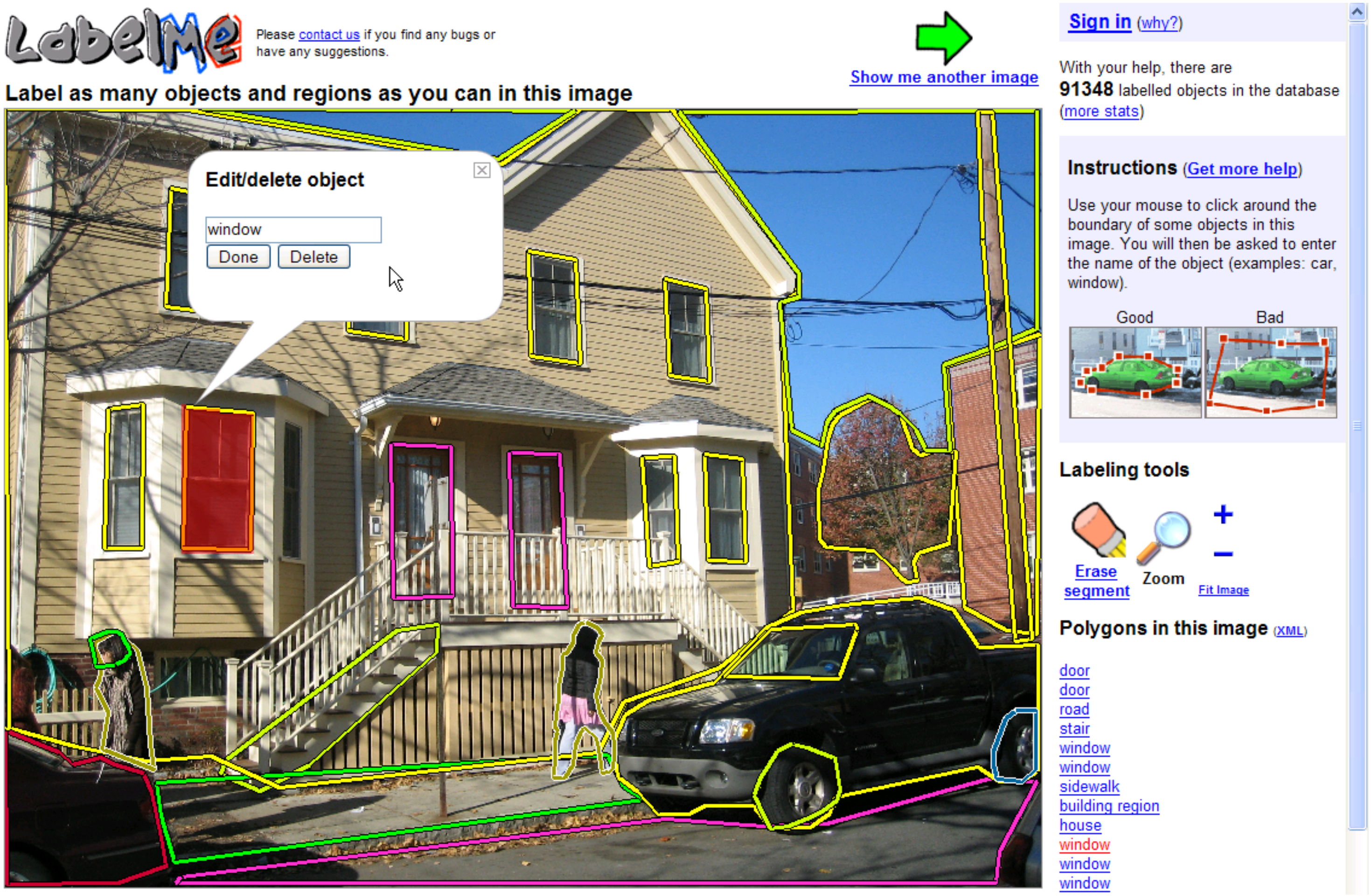

YOU AND A RANDOM PARTNER TAKE TURNS PEEKING AND BOOMING
PEEK : GUESS WHAT YOUR PARTNER IS REVEALING
BOOM : REVEAL PARTS OF THE IMAGE TO YOUR PARTNER
2:23
300
2:23
330
peek
boom
COW
HINTS HELP YOU GUESS
PASS FOR DIFFICULT IMAGES
GIVE HINTS IF NECESSARY
TELL YOUR PARTNER IF A GUESS IS HOT OR COLD

Figure 2. Peekaboom. *“Peek” tries to guess the word associated with an image slowly revealed by “Boom.”*

L. von Ahn, L. Dabbish, 2004; L. von Ahn, R. Liu and M. Blum, 2006

LabelMe

http://labelme.csail.mit.edu/



Please contact us if you find any bugs or have any suggestions.

Label as many objects and regions as you can in this image

With your help, there are 91348 labelled objects in the database (more stats)

Instructions (Get more help)

Use your mouse to click around the boundary of some objects in this image. You will then be asked to enter the name of the object (examples: car, window).

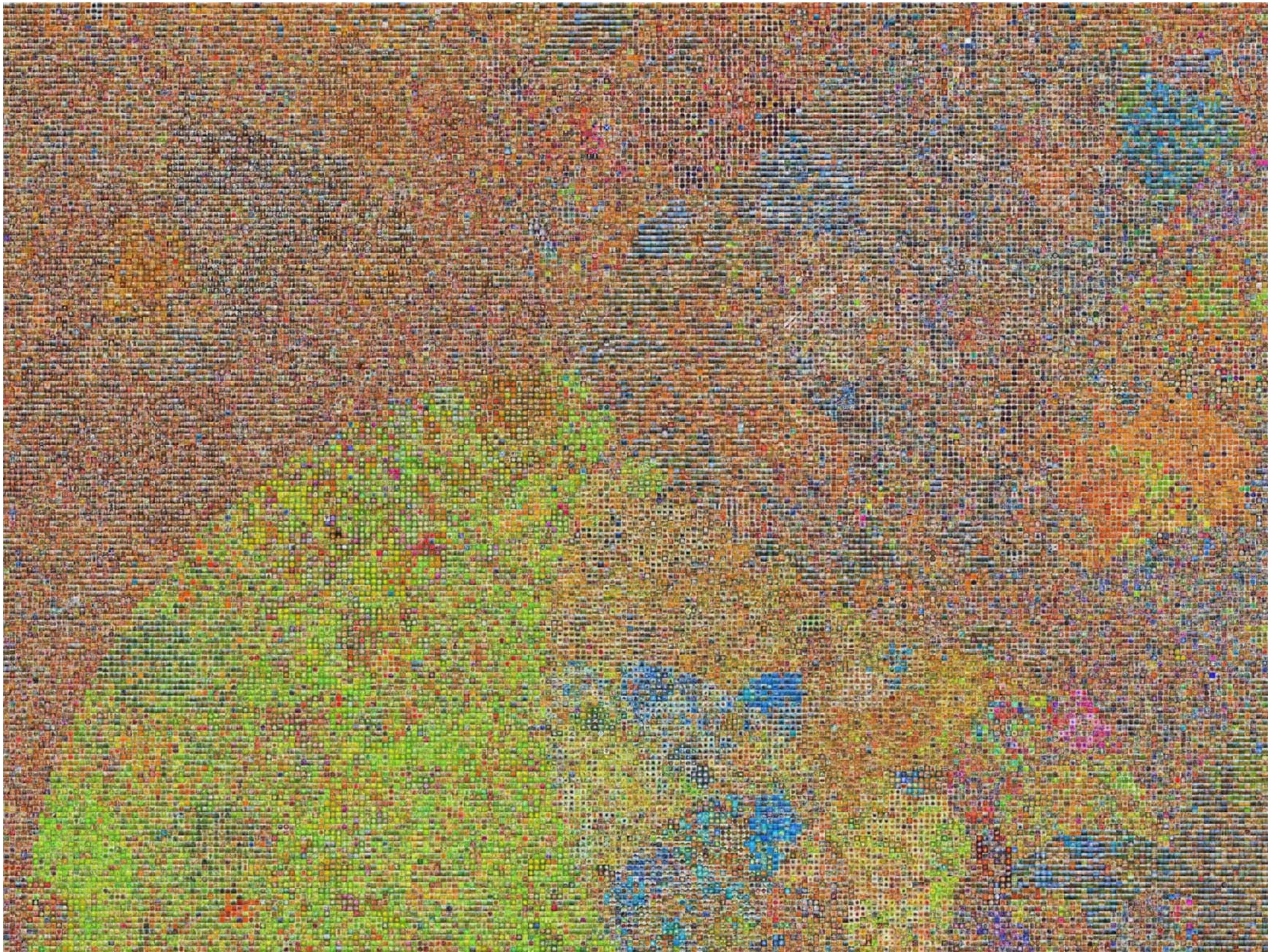
Polygons in this image (XML)

door
door
road
stair
window
window
sidewalk
building region
house
window
window
window

Russell, Torralba, Murphy, Freeman, 2008

80 Million Tiny Images

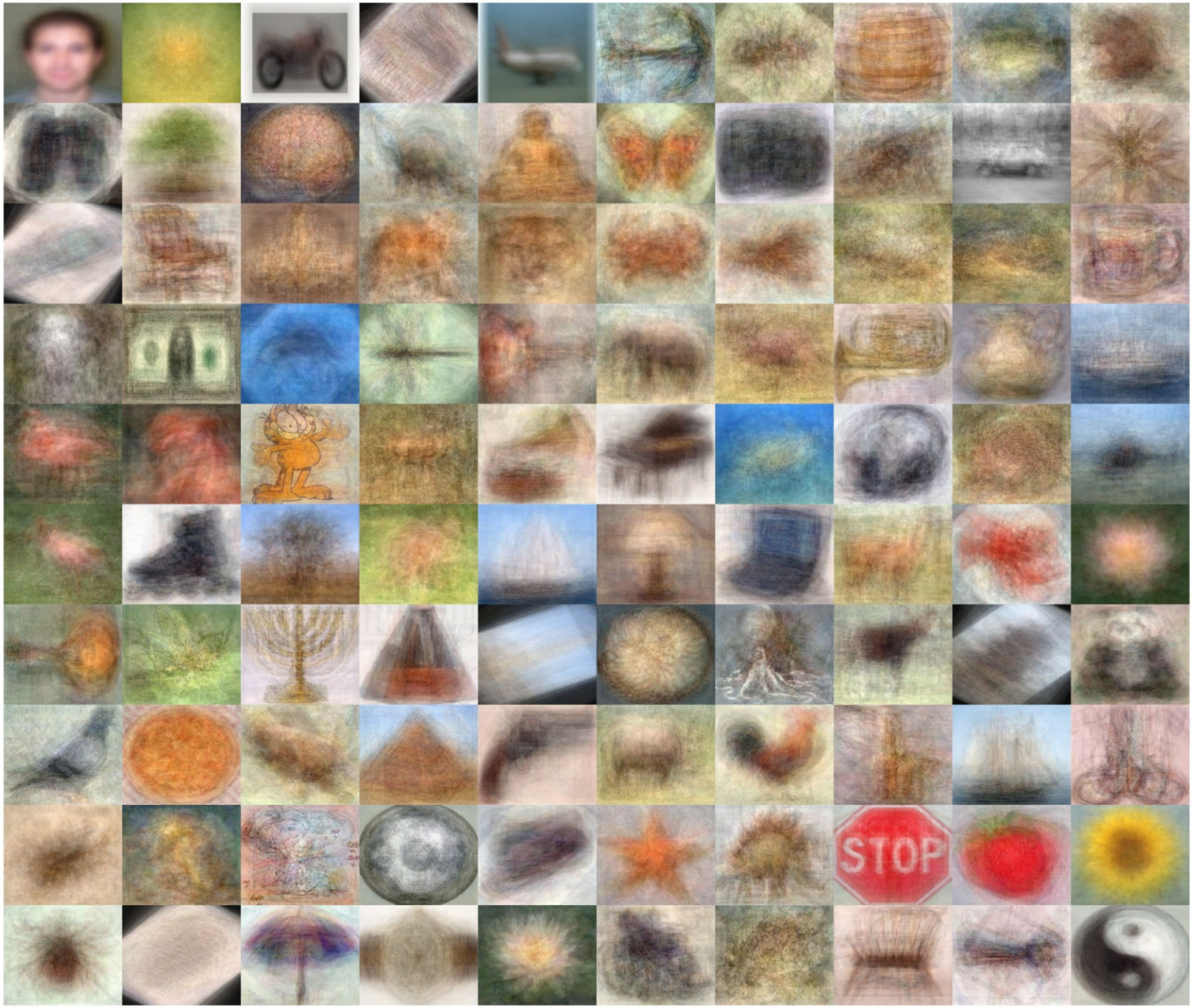
http://people.csail.mit.edu/torralba/tinyimages/



Dataset issues

- How large is the degree of intra-class variability?
- How “confusable” are the classes?
- Is there bias introduced by the background? I.e., can we “cheat” just by looking at the background and not the object?

Caltech-101



Summary

- Recognition is the “grand challenge” of computer vision
- History
  - Geometric methods
  - Appearance-based methods
  - Sliding window approaches
  - Local features
  - Parts-and-shape approaches
  - Bag-of-features approaches
- Issues
  - Generative vs. discriminative models
  - Supervised vs. unsupervised methods
  - Tasks, datasets