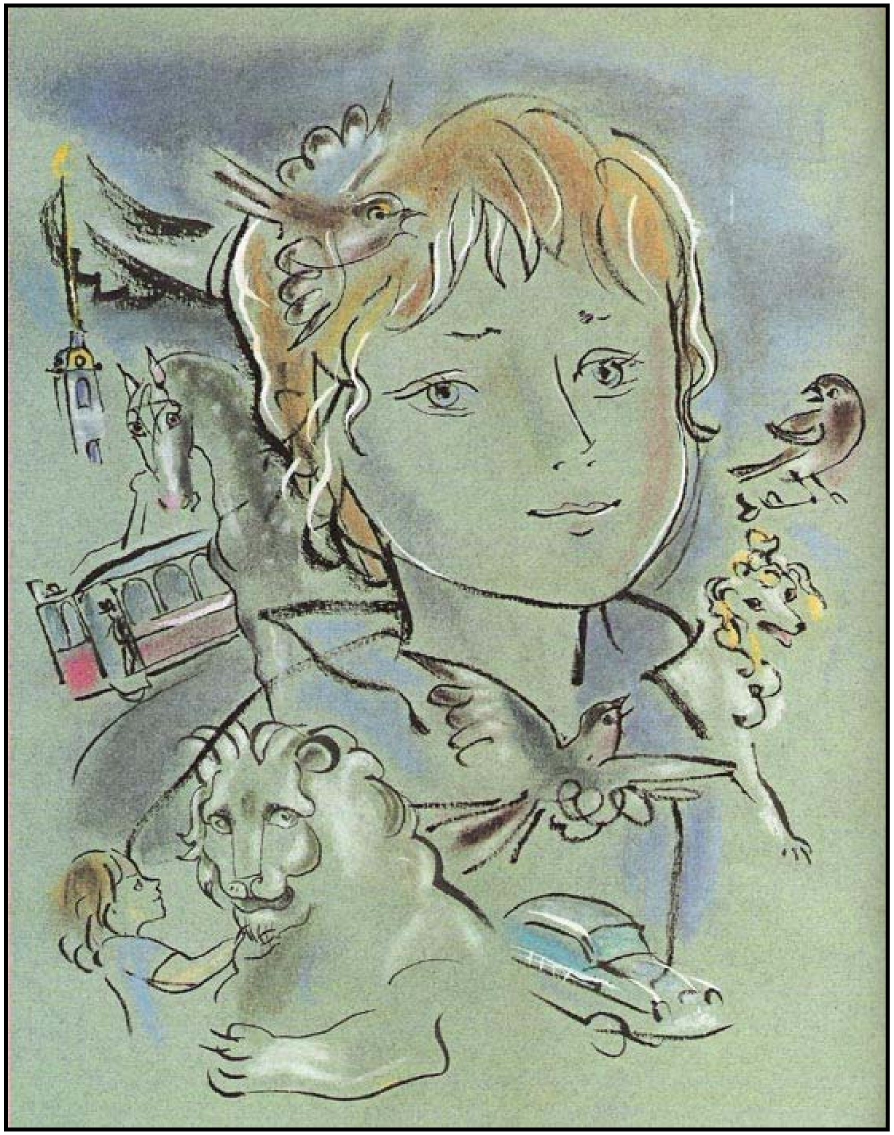
CS 558: Computer Vision 9th Set of Notes

Instructor: Philippos Mordohai
Webpage: www.cs.stevens.edu/~mordohai
E-mail: Philippos.Mordohai@stevens.edu
Office: Lieb 215

Introduction to object recognition



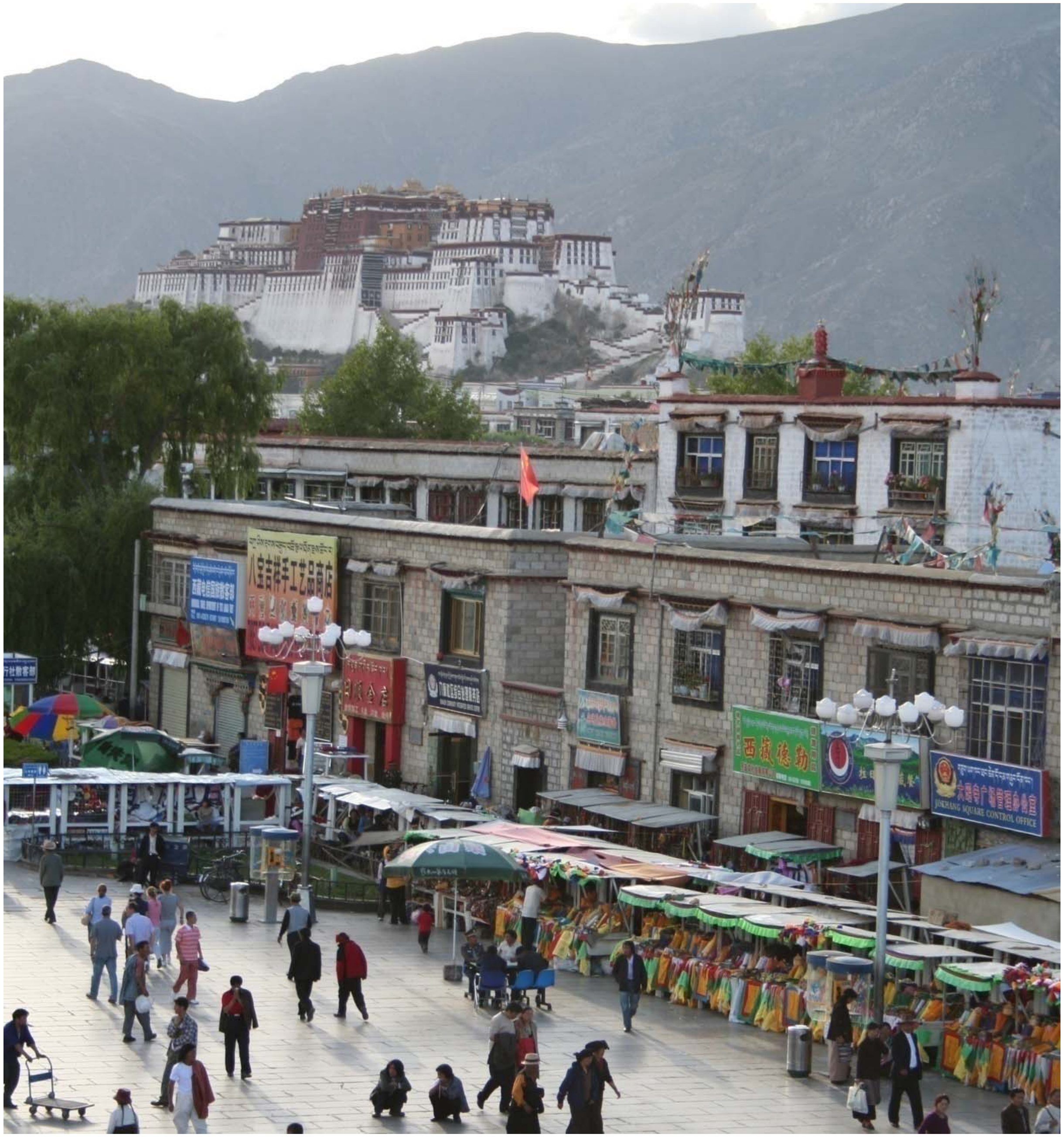
By Svetlana Lazebnik

Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce

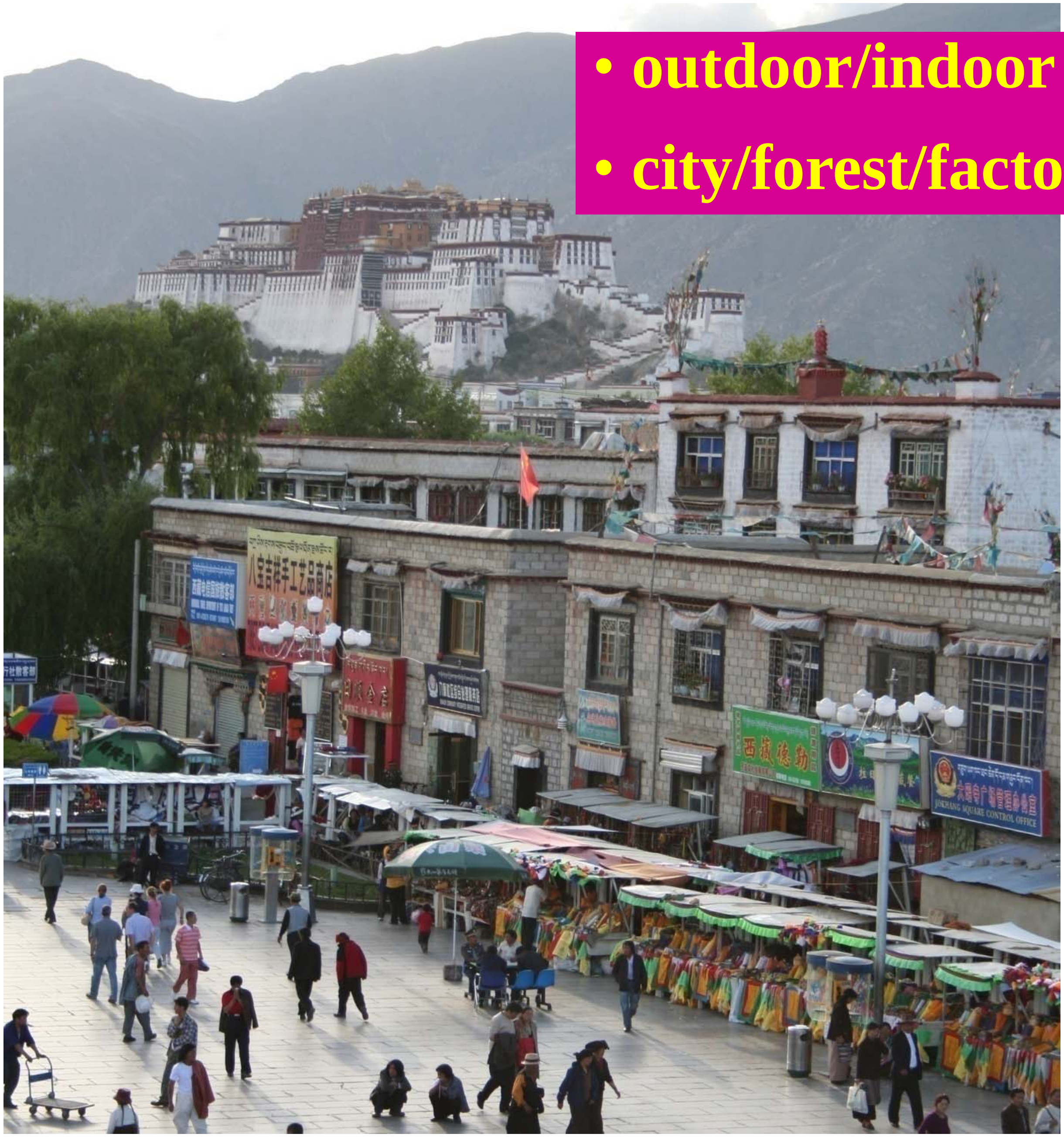
Overview

- Basic recognition tasks
- A machine learning approach
    - Example features
    - Example classifiers
    - Levels of supervision
    - Datasets
- Current trends and advanced recognition tasks

Specific recognition tasks

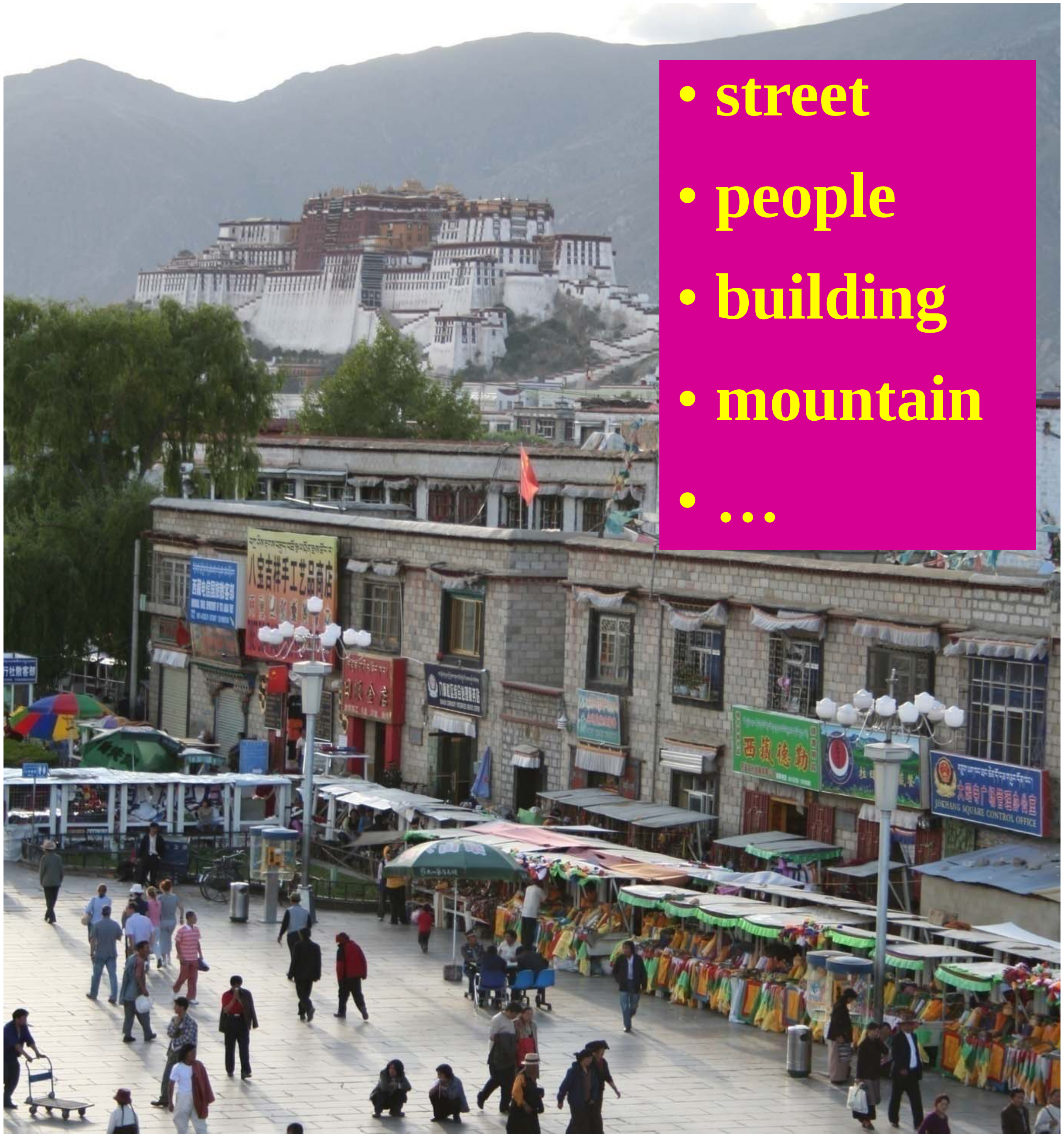


Scene categorization



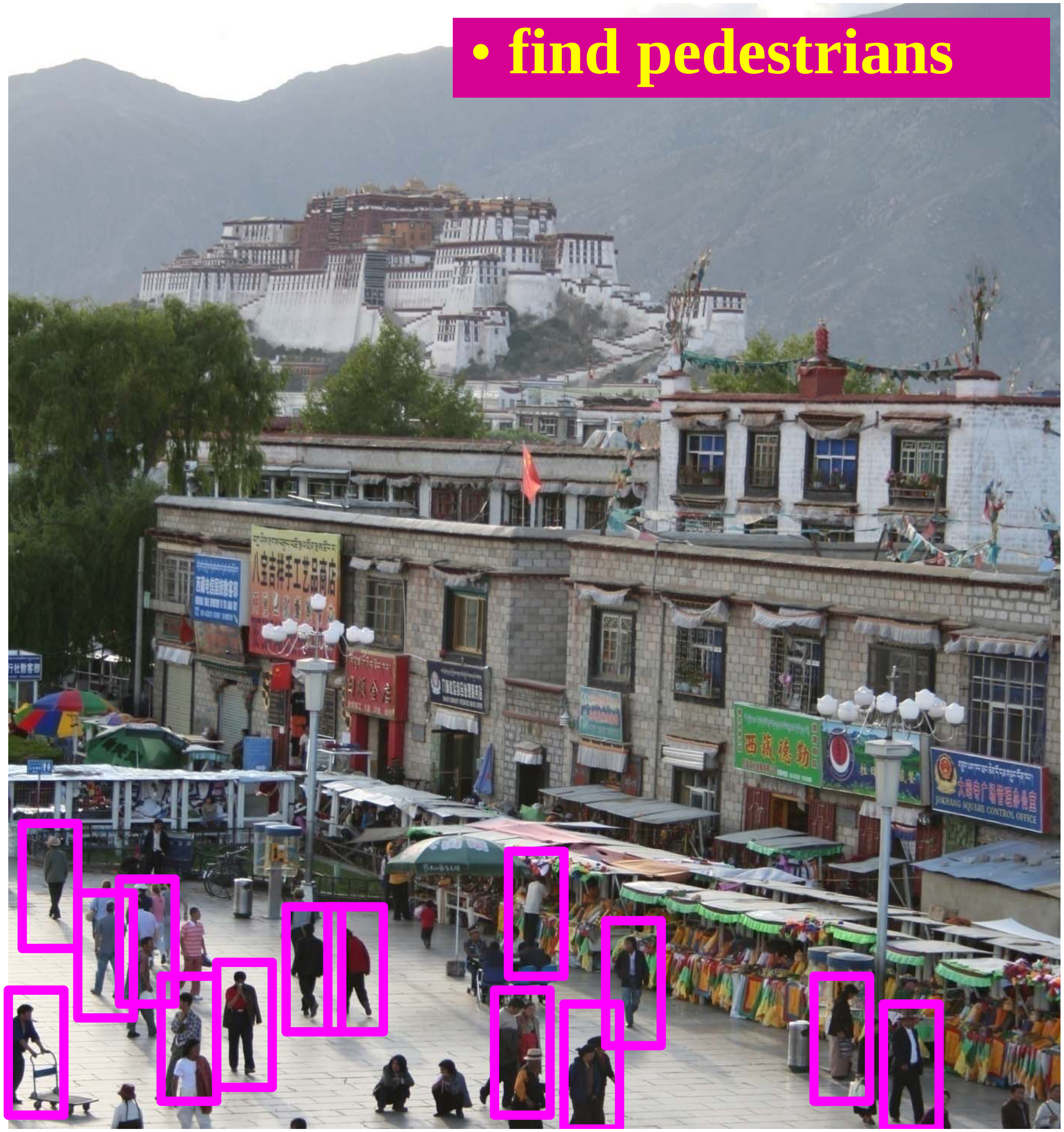
- **outdoor/indoor**
- **city/forest/factory/etc.**

Image annotation/tagging



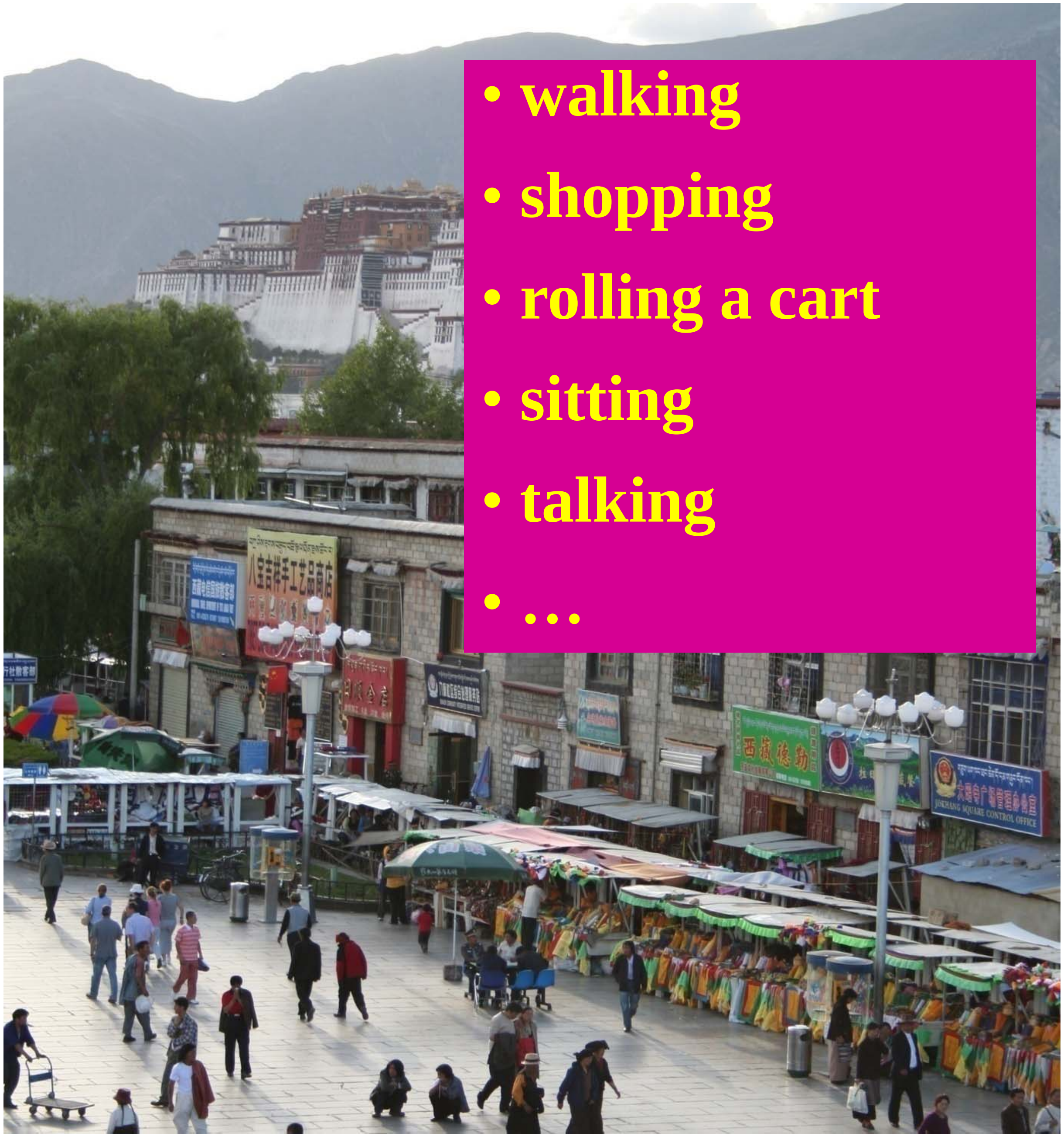
- **street**
- **people**
- **building**
- **mountain**
- **…**

Object detection



- find pedestrians

Activity recognition



- walking
- shopping
- rolling a cart
- sitting
- talking
- …

Image parsing

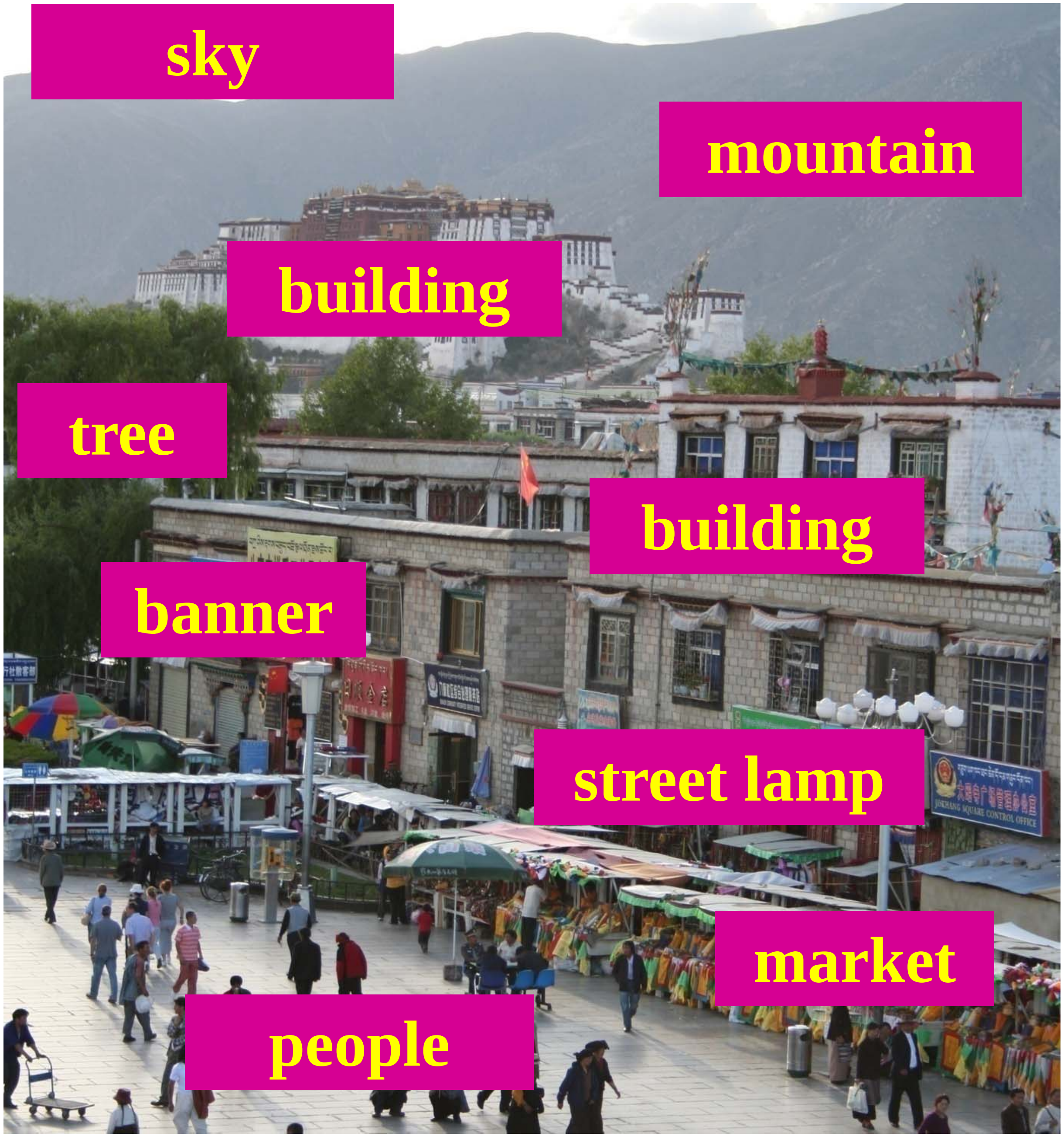
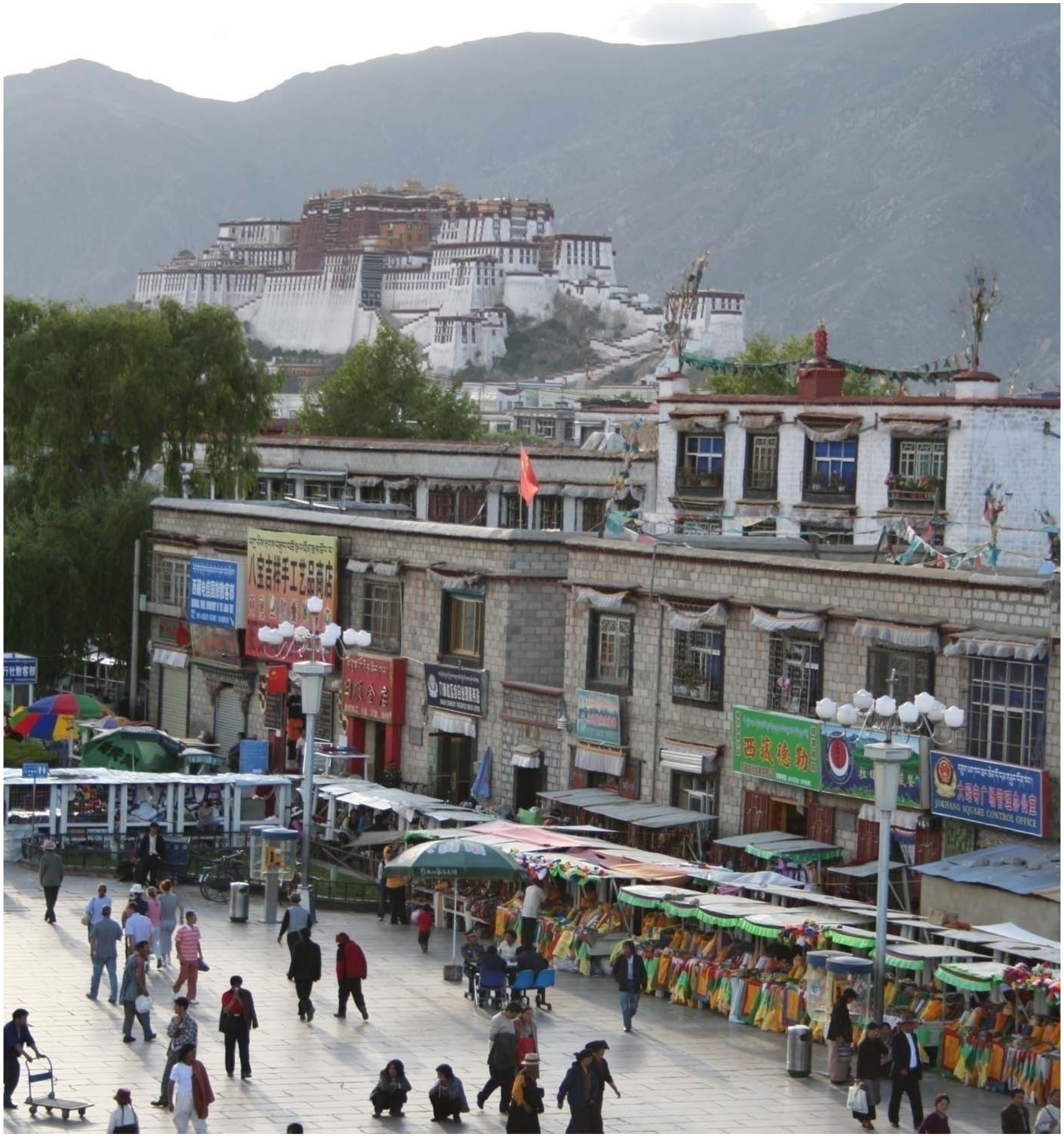
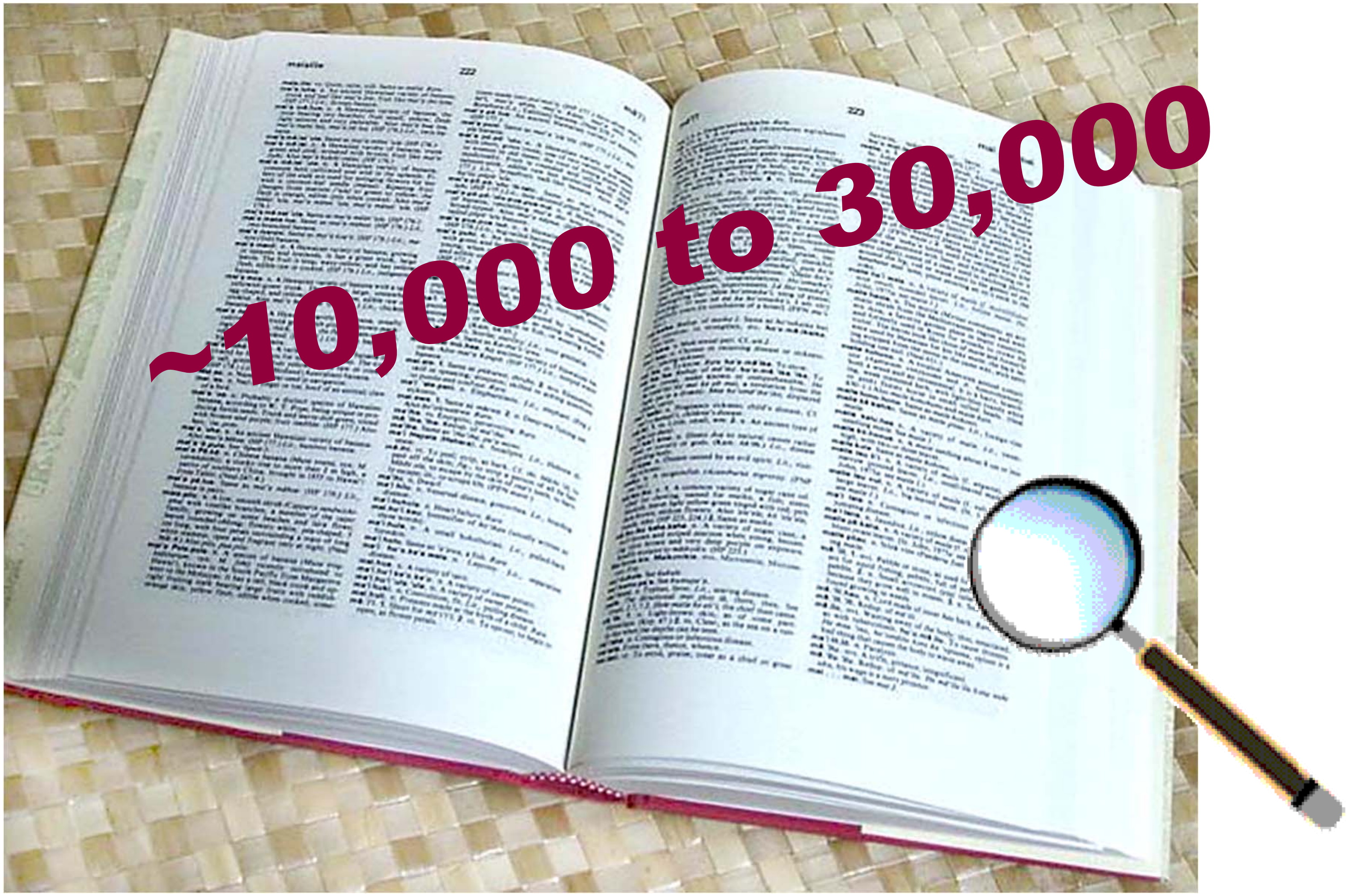


Image understanding?



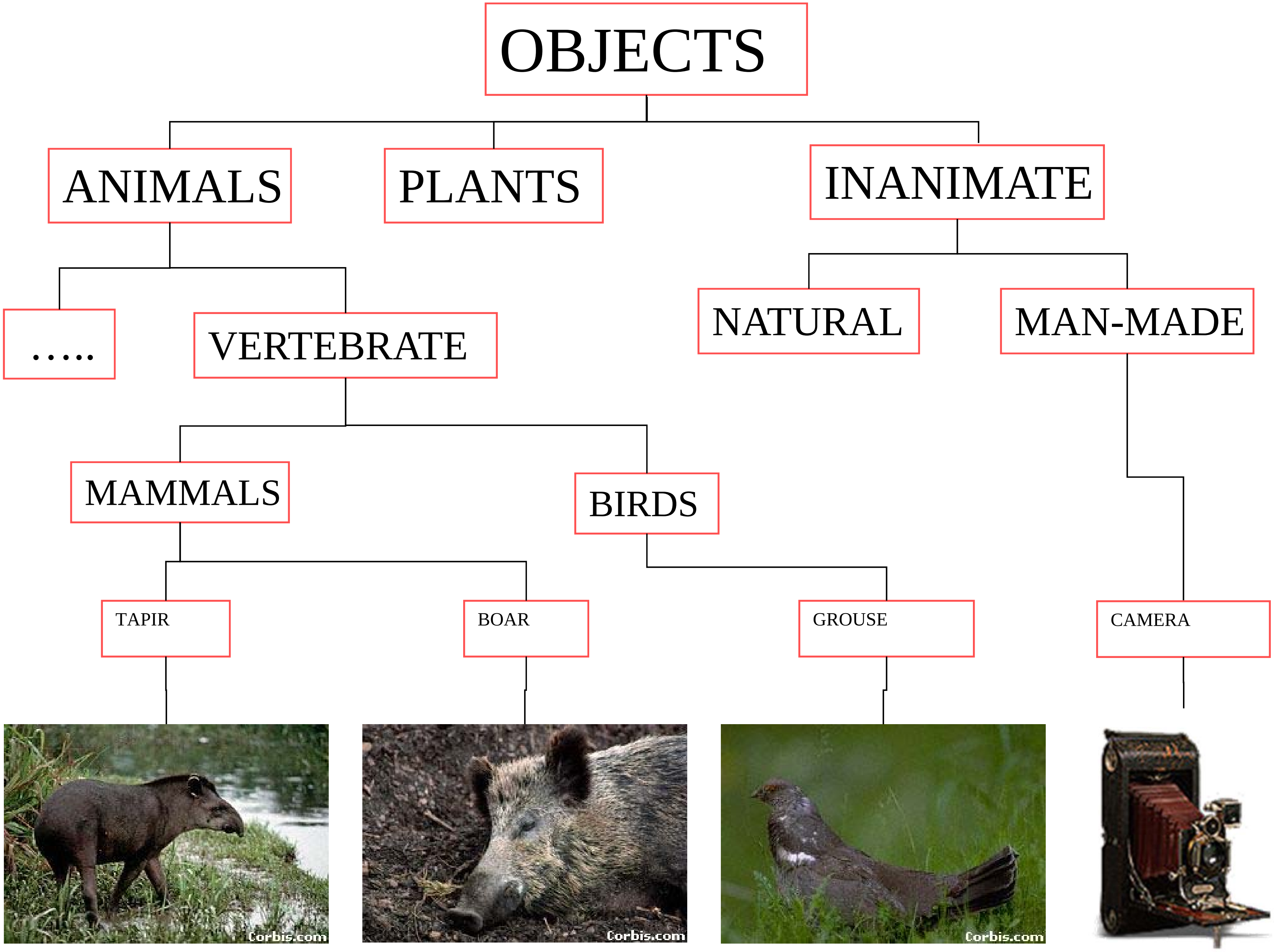
How many visual object categories are there?



~10,000 to 30,000

Biederman 1987

~10,000 to 30,000



OBJECTS

ANIMALS

PLANTS

INANIMATE

…..

VERTEBRATE

NATURAL

MAN-MADE

MAMMALS

BIRDS

TAPIR

BOAR

GROUSE

CAMERA

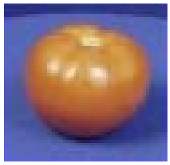
Recognition: A machine learning approach

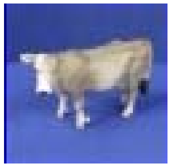
Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, Kristen Grauman, and Derek Hoiem

The machine learning framework

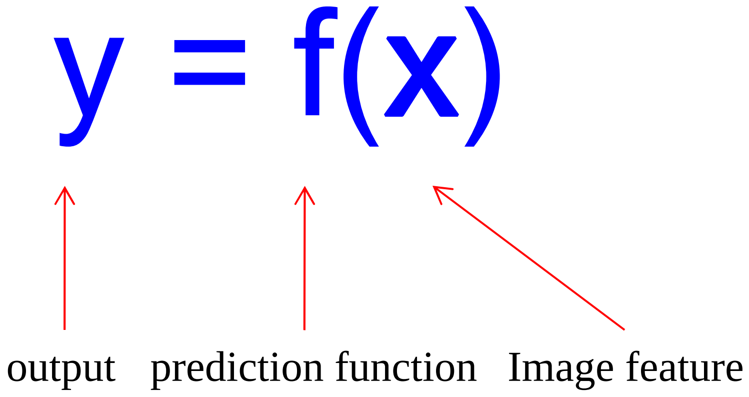
- Apply a prediction function to a feature representation of the image to get the desired output:

f( ) = “apple”

f() = “tomato”

f() = “cow”

The machine learning framework

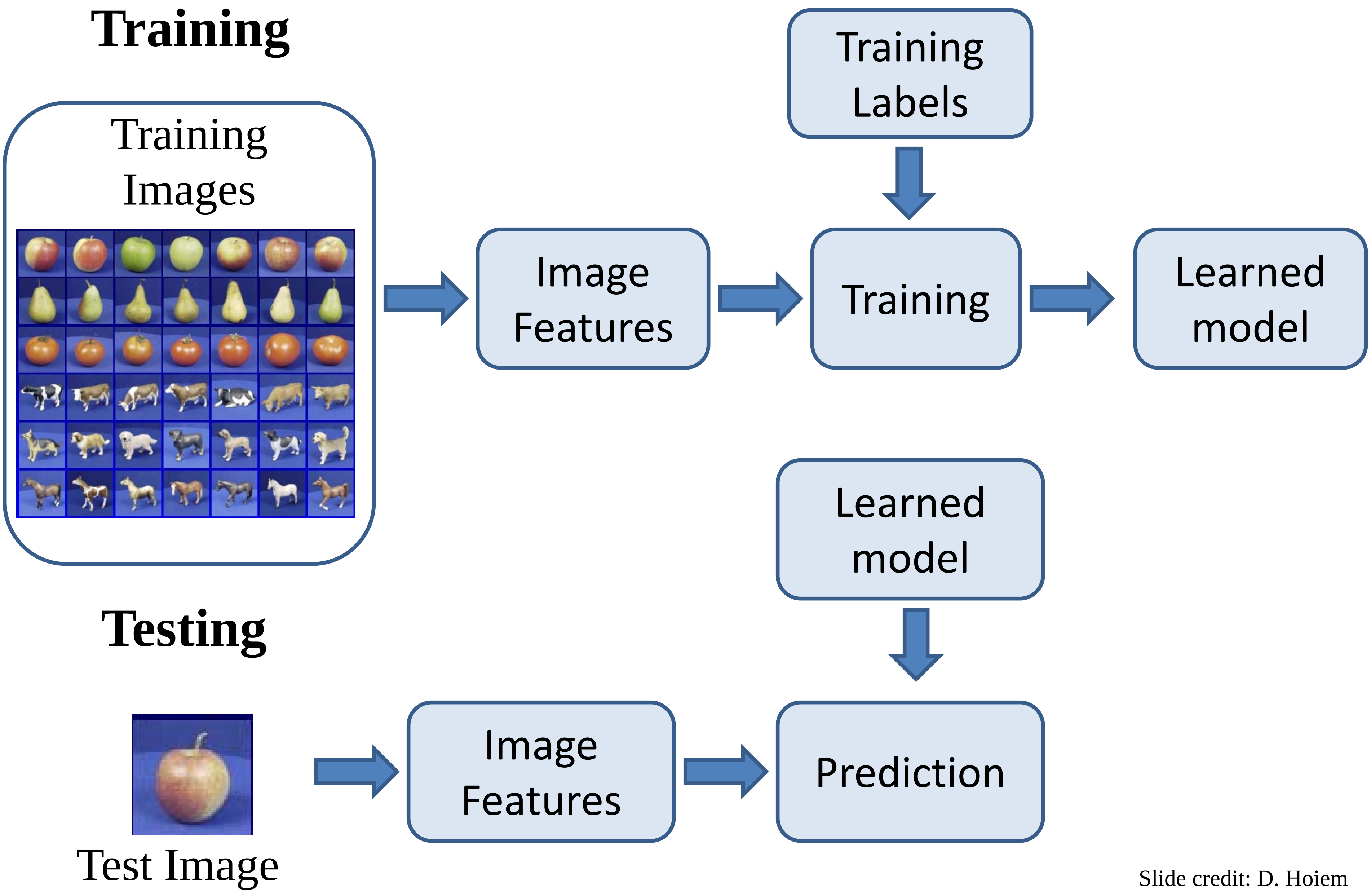


$$y = f(\mathbf{x})$$

output prediction function Image feature

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Steps



Training

Training Images

Image Features

Training Labels

Training

Learned model

Testing

Test Image

Image Features

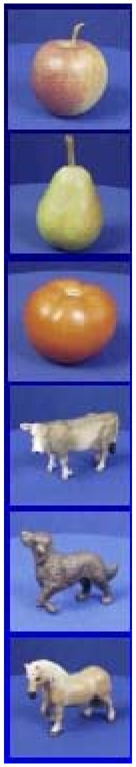
Learned model

Prediction

Slide credit: D. Hoiem

Generalization

Training set (labels known)

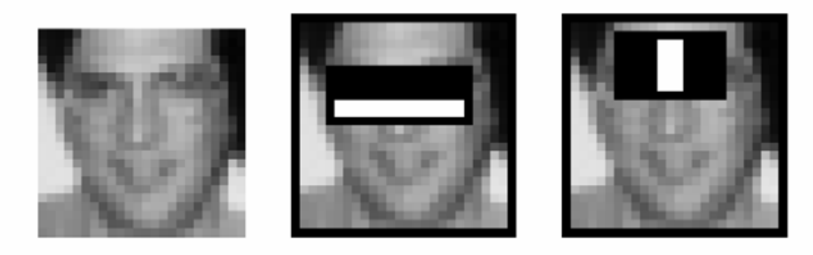


Test set (labels unknown)

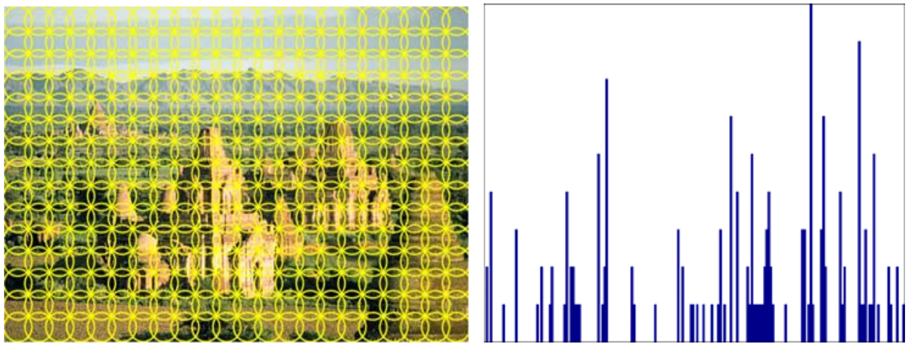
- How well does a learned model *generalize* from the data it was trained on to a new test set?

Popular global image features

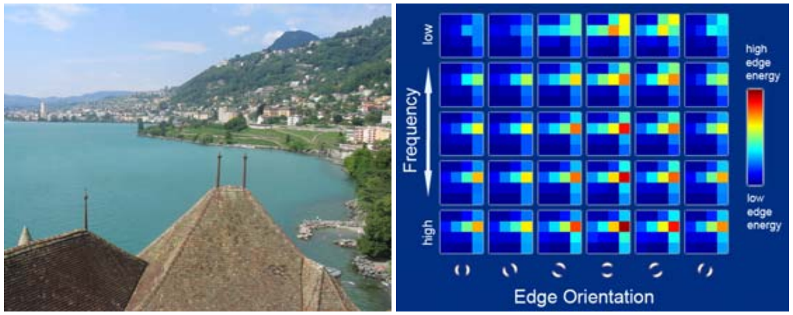
- Raw pixels (and simple functions of raw pixels)



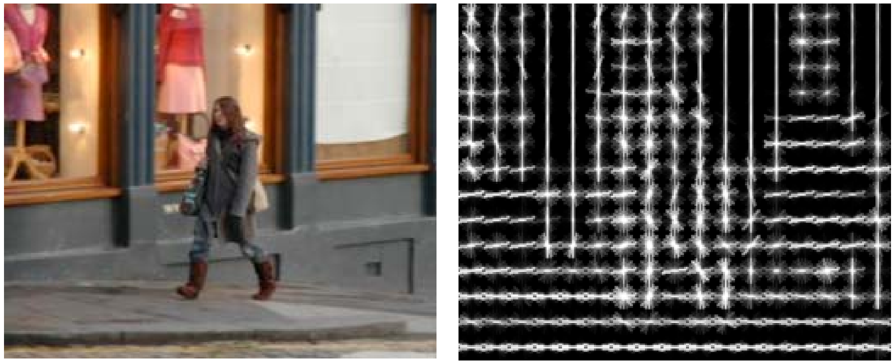
- Histograms, bags of features



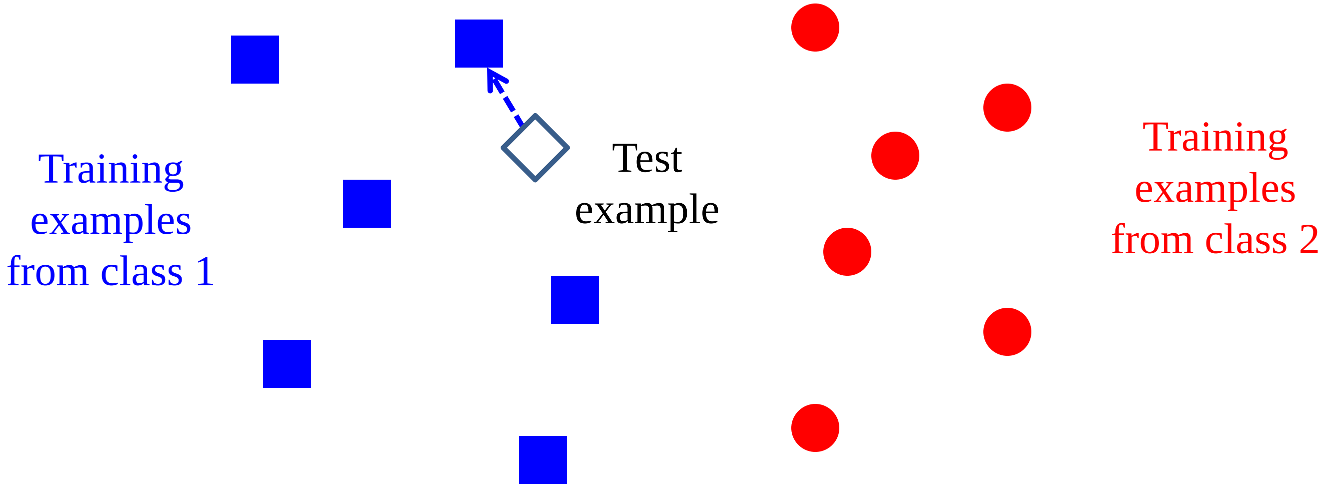
- GIST descriptors [Oliva and Torralba, 2001]



- Histograms of oriented gradients (HOG) [Dalal and Triggs, 2005]



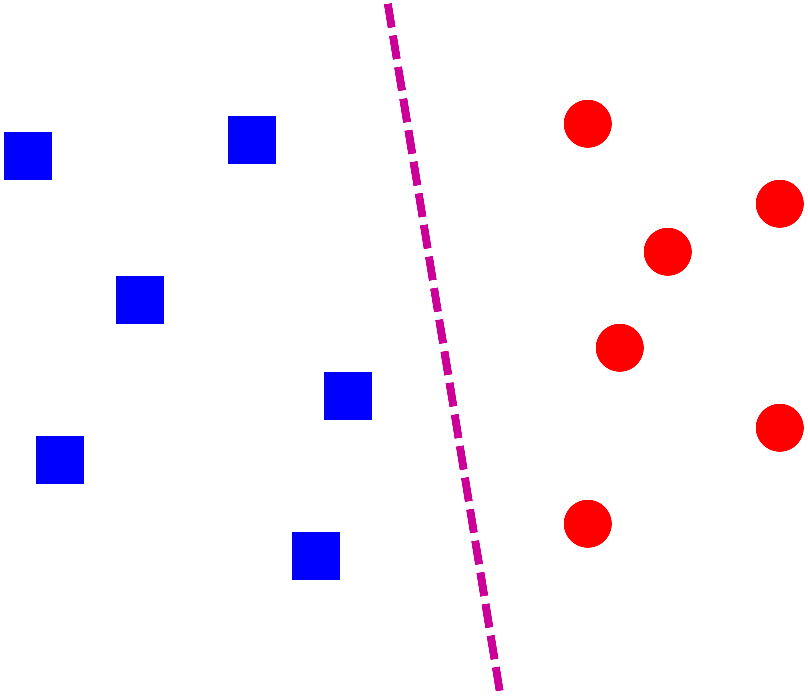
Classifiers: Nearest neighbor


Training examples from class 1
Test example
Training examples from class 2

f(x) = label of the training example nearest to x

- All we need is a distance function for our inputs
- No training required!

Classifiers: Linear



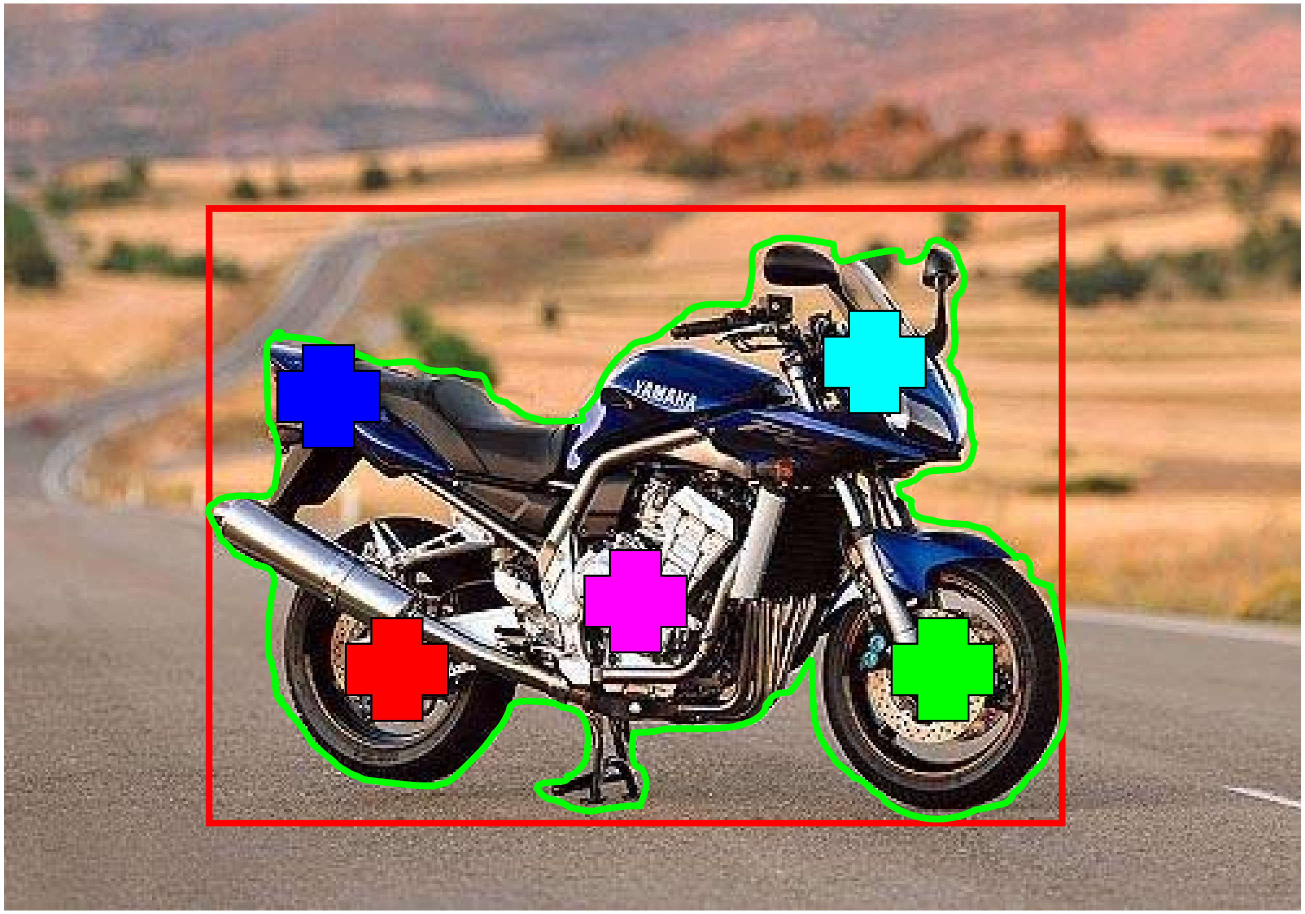
- Find a *linear function* to separate the classes:

$$f(\mathbf{x}) = \mathrm{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$$

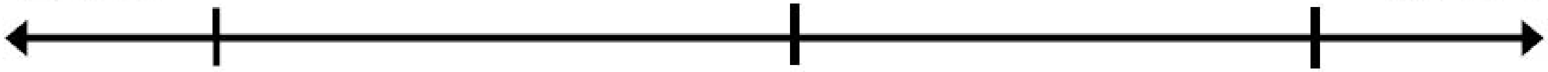
Recognition task and supervision

- Images in the training set must be annotated with the "correct answer" that the model is expected to produce

Contains a motorbike



Spectrum of supervision

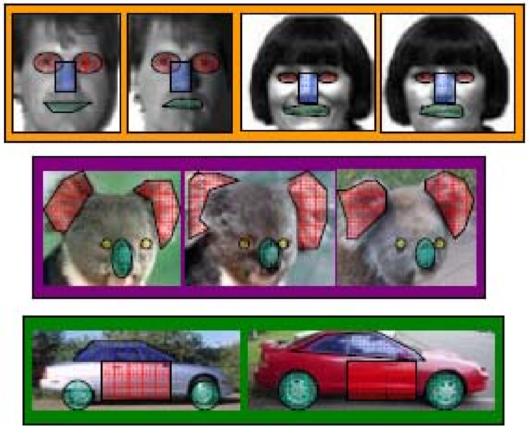


Less

More

Unsupervised

“Weakly” supervised



Fully supervised

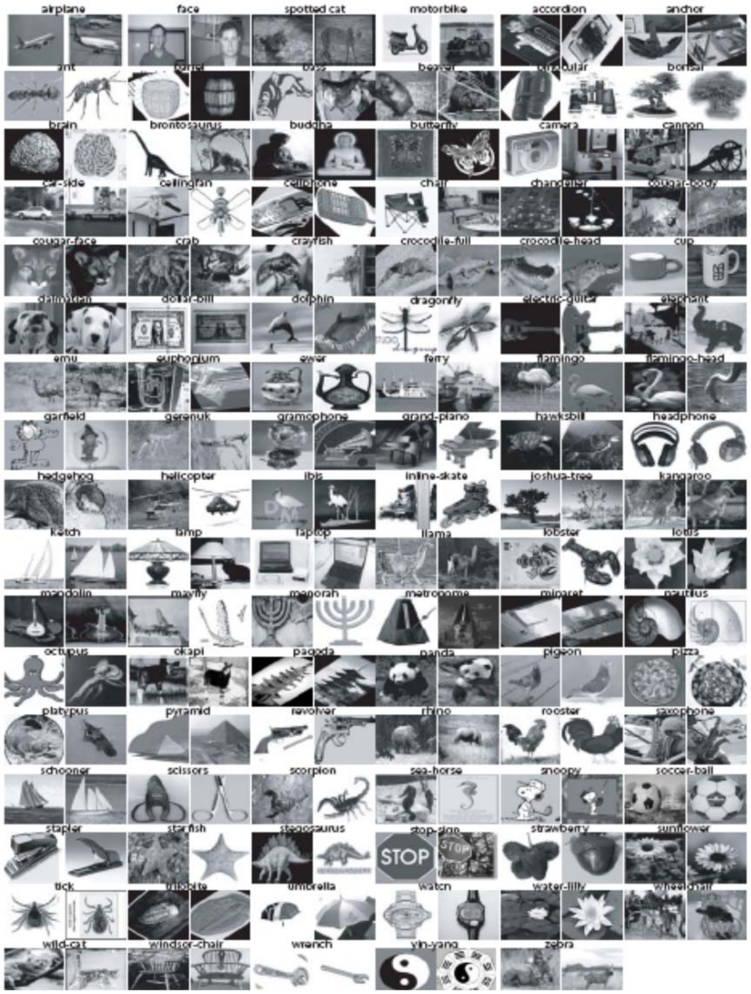
Definition depends on task

Datasets

- **Circa 2001:** five categories, hundreds of images per category
- **Circa 2004:** 101 categories
- **Today:** tens of thousands of categories, millions of images

Caltech 101 & 256

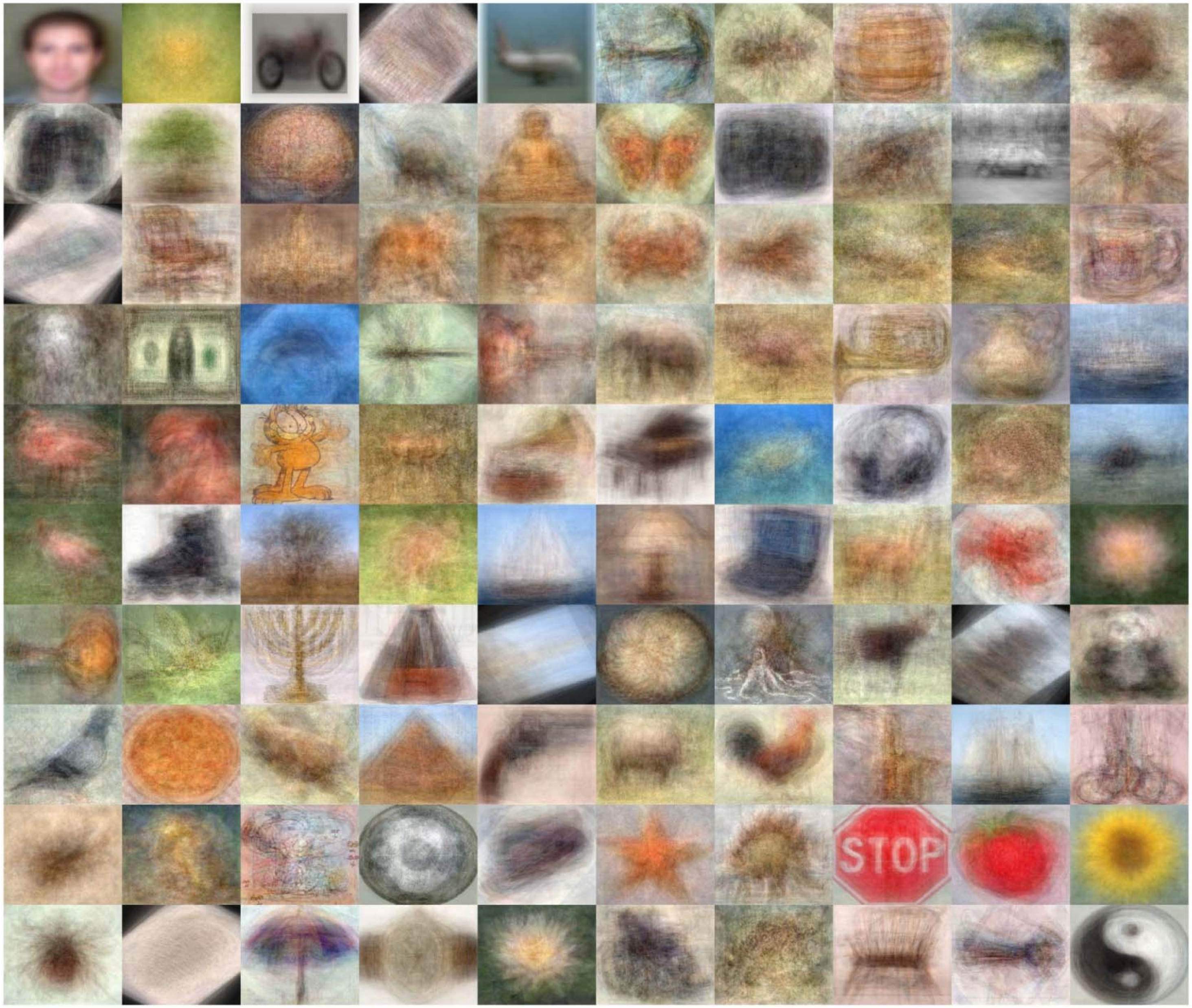
http://www.vision.caltech.edu/Image_Datasets/Caltech101/
http://www.vision.caltech.edu/Image_Datasets/Caltech256/



Fei-Fei, Fergus, Perona, 2004

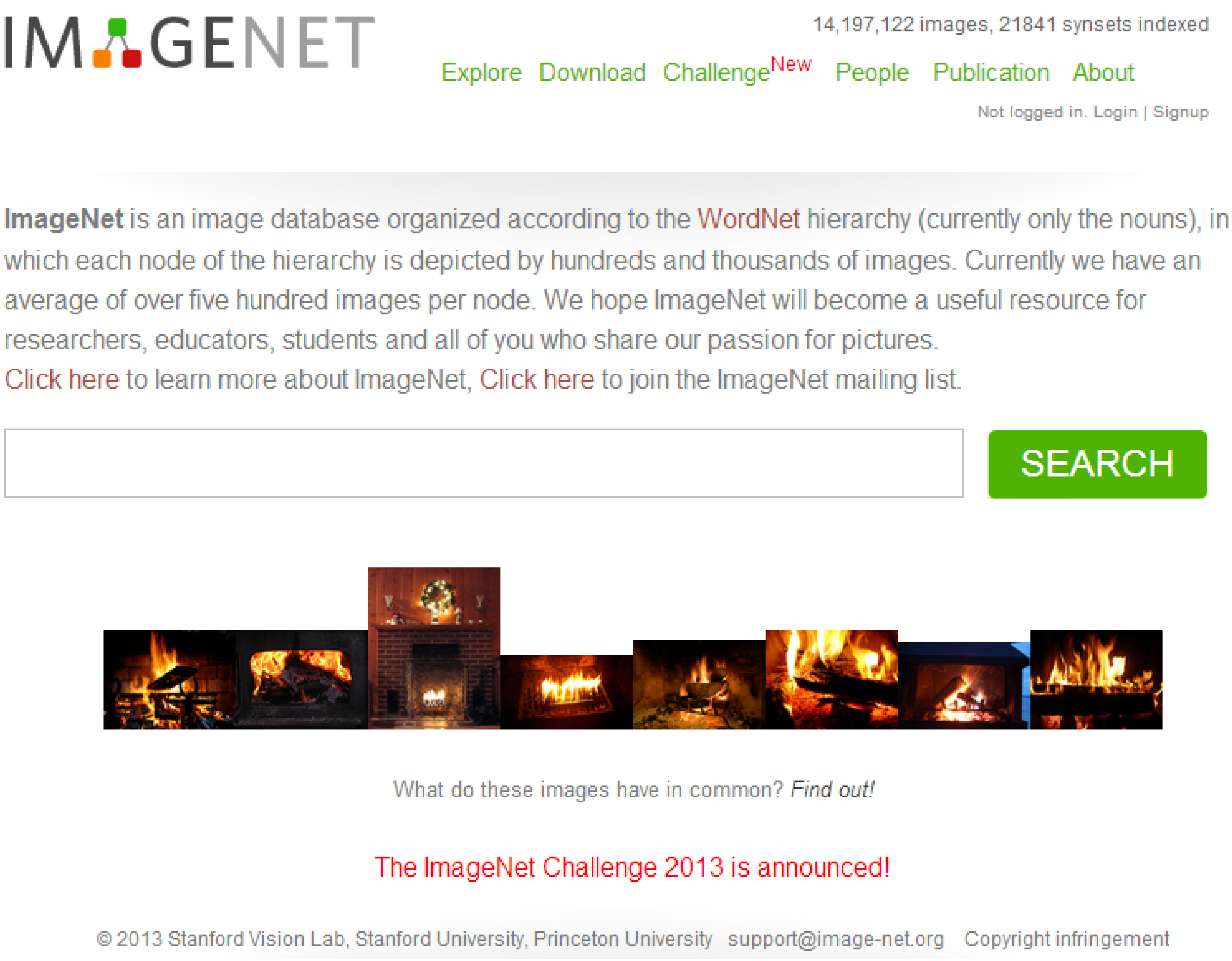
Griffin, Holub, Perona, 2007

Caltech-101: Intra-class variability



ImageNet

http://www.image-net.org/



14,197,122 images, 21841 synsets indexed

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

What do these images have in common? *Find out!*

The ImageNet Challenge 2013 is announced!

© 2013 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

The PASCAL Visual Object Classes Challenge (2005-2012)

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

- **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

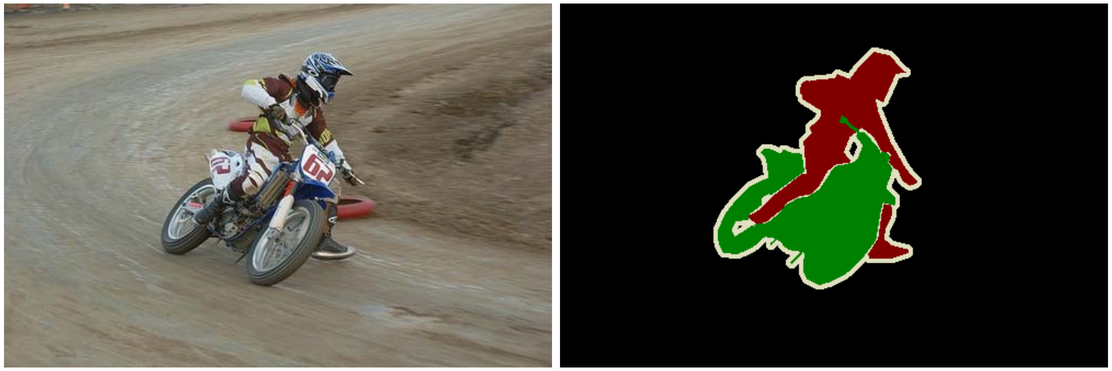
- **Dataset size (by 2012):**
  11.5K training/validation images, 27K bounding boxes, 7K segmentations

PASCAL competitions

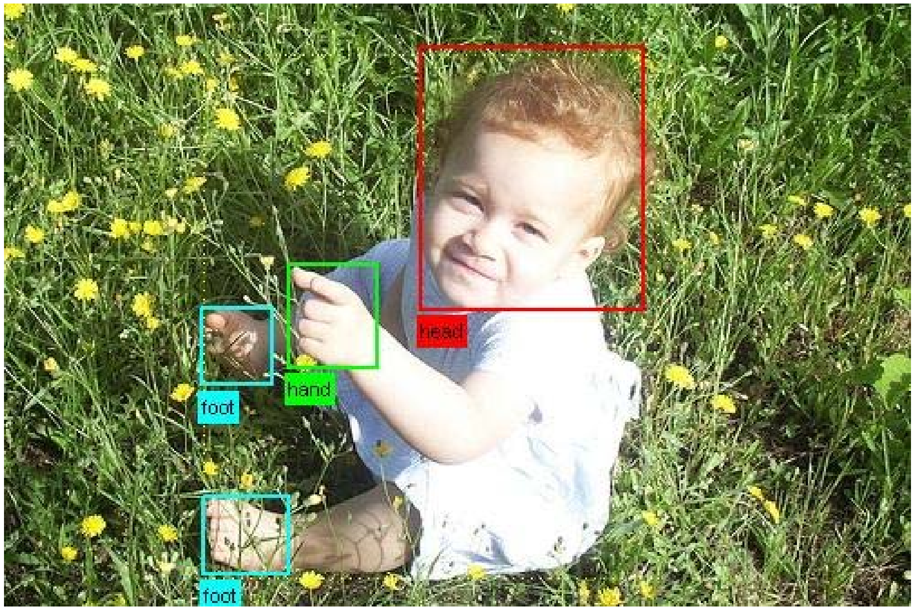
- **Classification:** For each of the twenty classes, predicting presence/absence of an example of that class in the test image
- **Detection:** Predicting the bounding box and label of each object from the twenty target classes in the test image

PASCAL competitions

- **Segmentation:** Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise

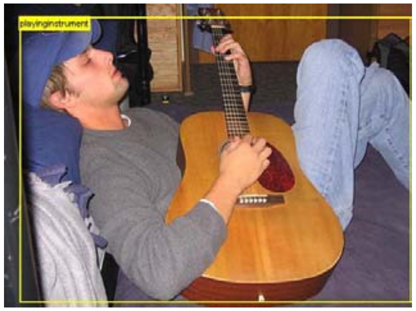


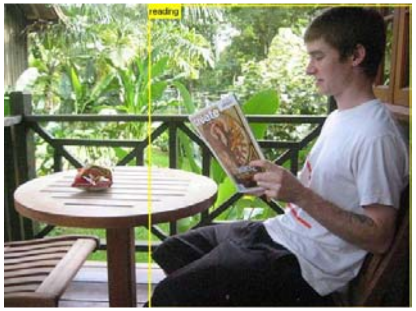
- **Person layout:** Predicting the bounding box and label of each part of a person (head, hands, feet)

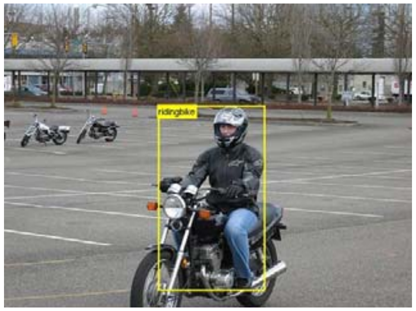


PASCAL competitions

- Action classification (10 action classes)

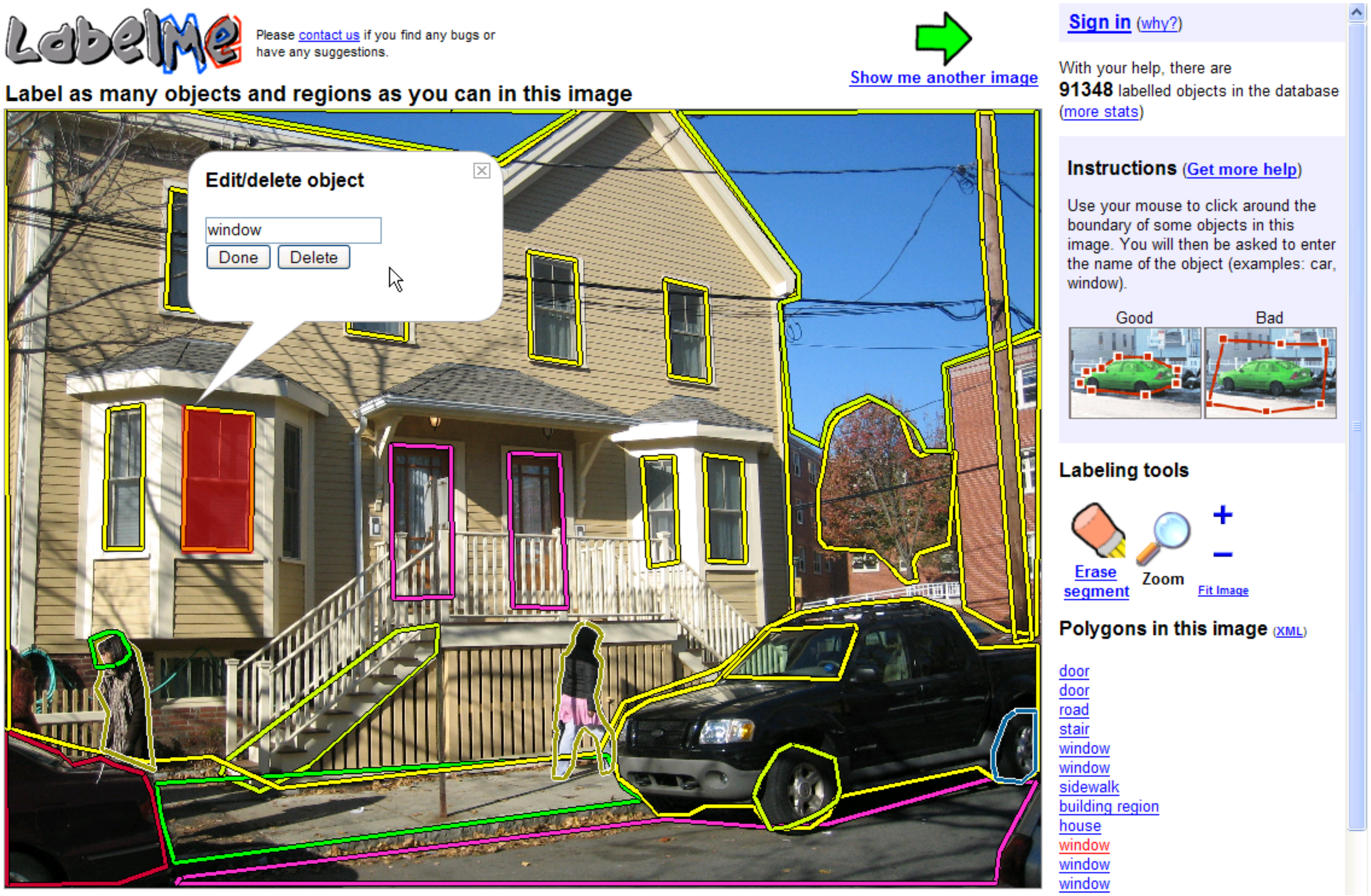






LabelMe Dataset

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

window

Sign in (why?)

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).

Labeling tools

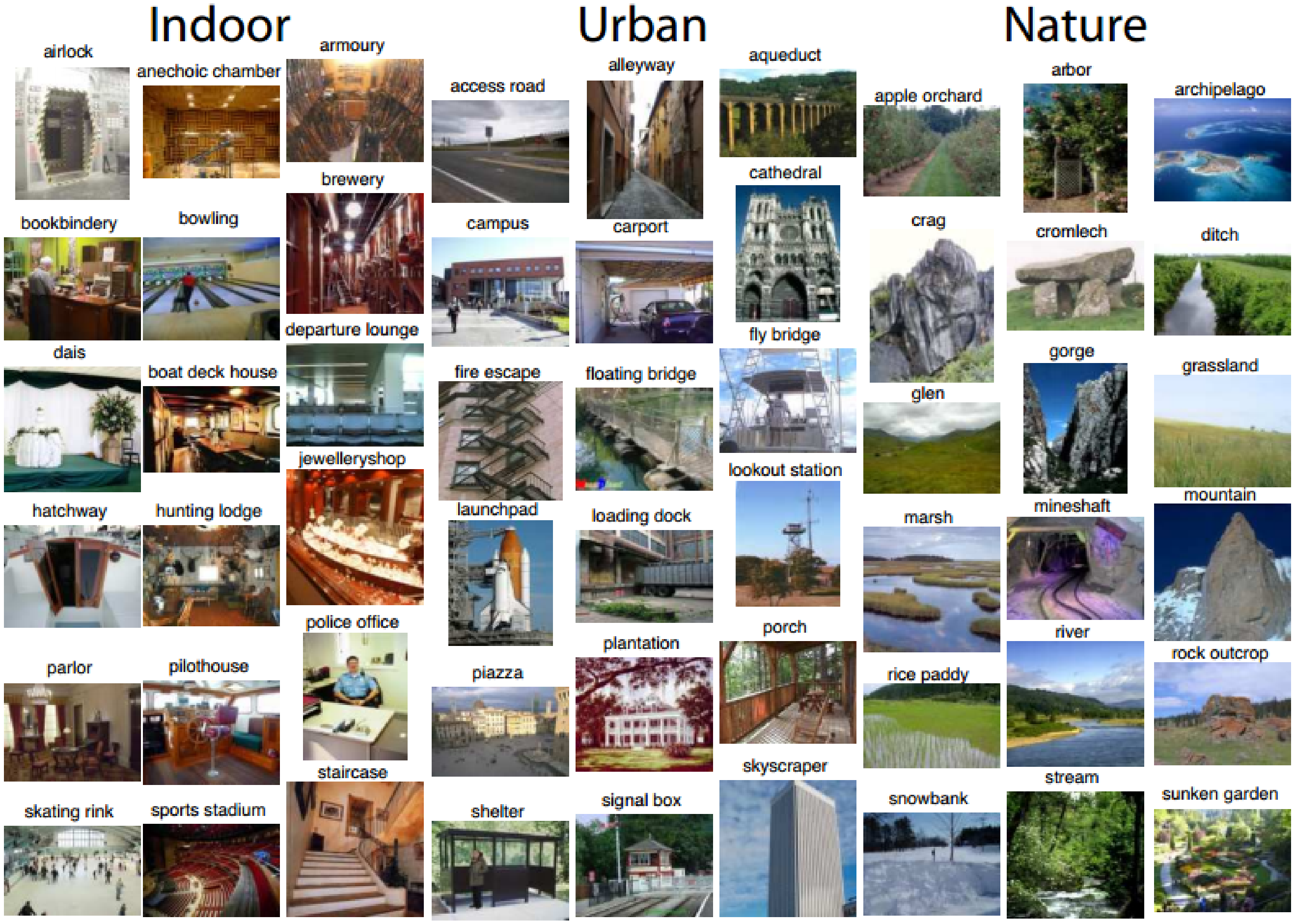
Polygons in this image (XML)

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

Russell, Torralba, Murphy, Freeman, 2008

SUN dataset

~900 scene categories (~400 well-sampled), 130K images



J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "SUN Database: Large-scale Scene Recognition from Abbey to Zoo," CVPR 2010

http://groups.csail.mit.edu/vision/SUN/

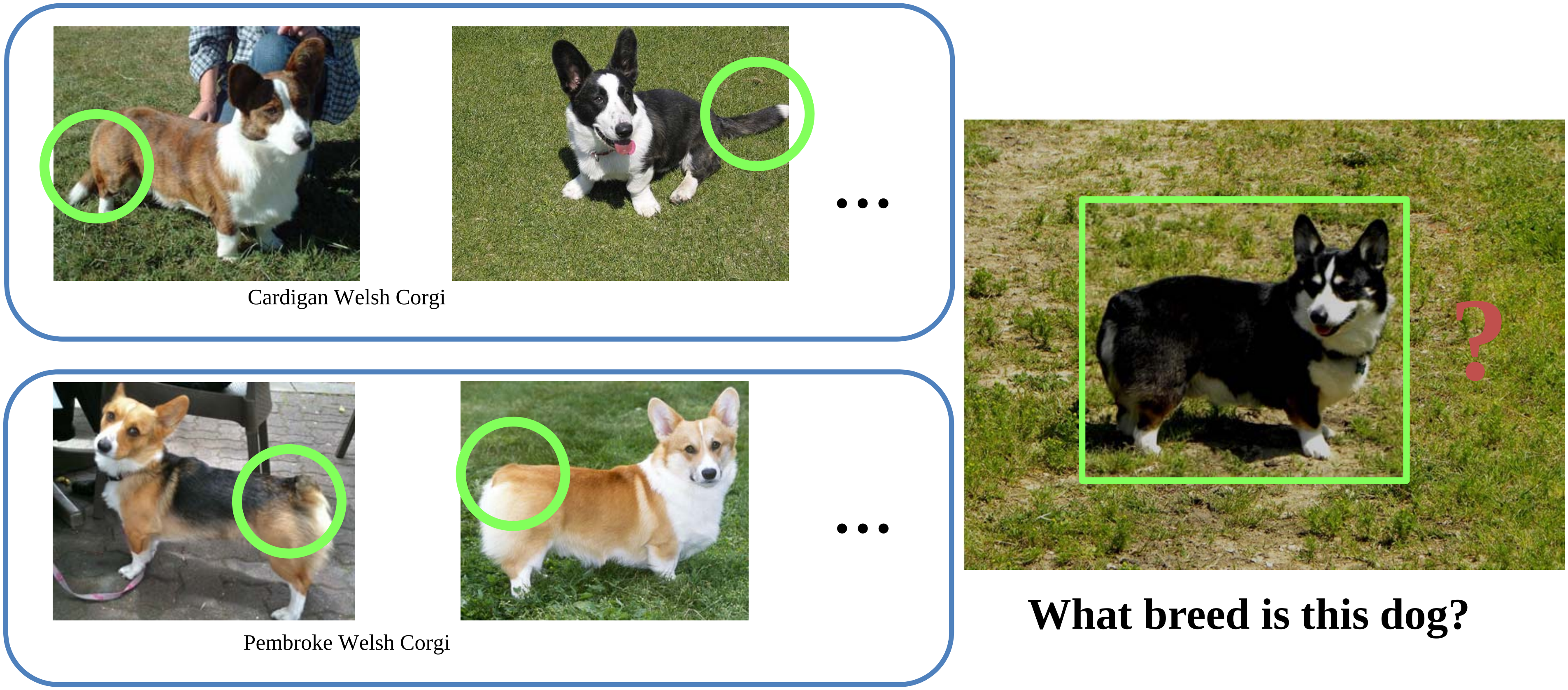
Fine-grained recognition

Airedale
200
Basset Hound
201
Beagle
202
Bernese
203
Bichon Frise
204
Blue Heeler
205
Border Collie
206
Boston
207
Boxer
208
Boxer
209
Briard
210
Bulldog
211
Cairn Terrier
212
Cavalier
213
Chihuahua
214
Chihuahua
215
Cocker Spaniel
216
Cocker Spaniel
217
Corgi
218
Dachshund
219
Dachshund
220
Dalmatian
221
Doberman
222
Doberman
223
German Shorthair
224
German Shepherd
225
Golden Retriever
226
Great Dane
227
Great Dane
228
Greyhound
229
Greyhound
230
Havanese
231

Virginia Rail
Magnificent Frigatebird
Double-crested Cormorant
Common Loon
Brown Pelican
Pied-billed Grebe
King Rail
Common Moorhen
Northern Gannet
Anhinga
Black Rail
Clapper Rail
Horned Grebe
Greater Shearwater
Wilson's Storm-petrel
Limpkin
Sora
Audubon's Shearwater
American White Pelican
Red-throated Loon
American Coot

Source: J. Deng

Fine-grained recognition



Cardigan Welsh Corgi

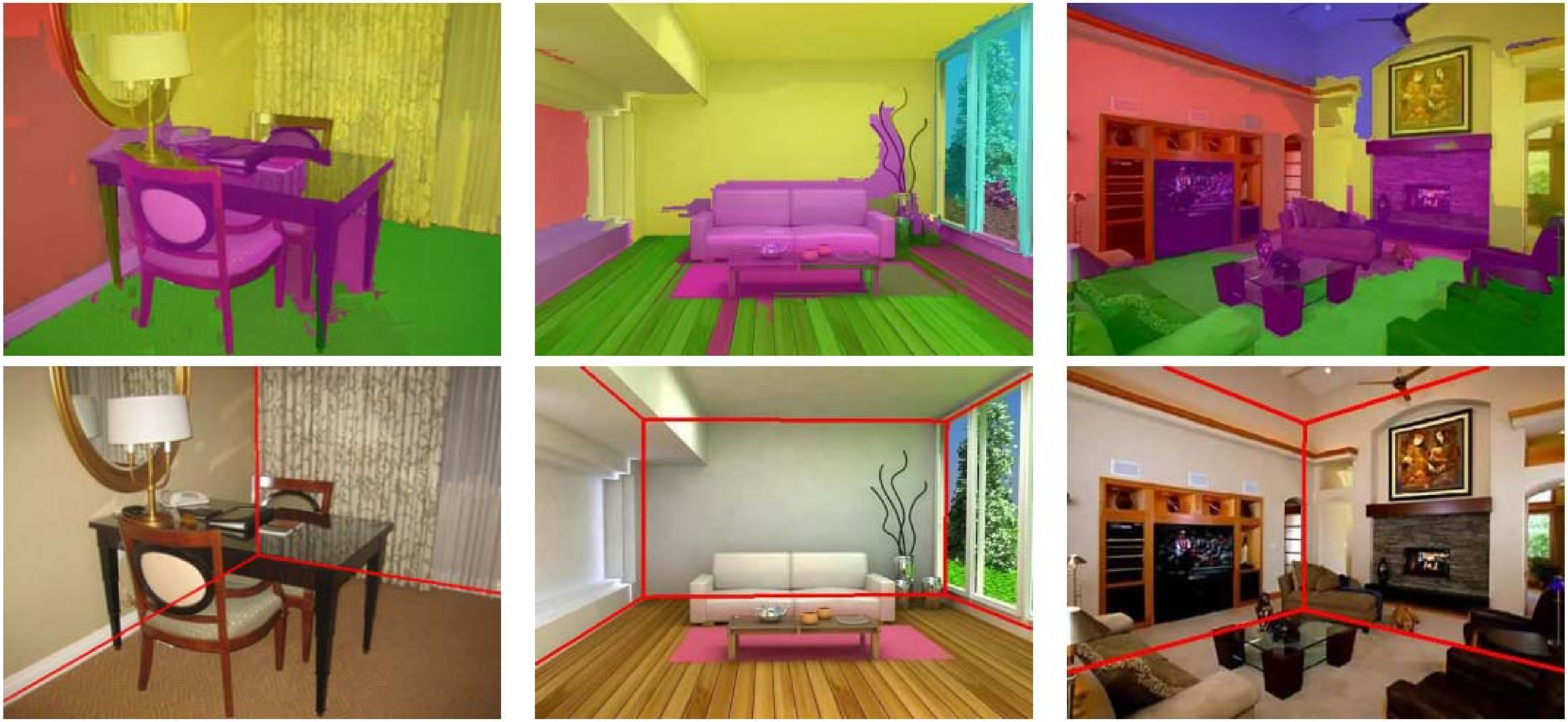
Pembroke Welsh Corgi

What breed is this dog?

Key: Find the right features.

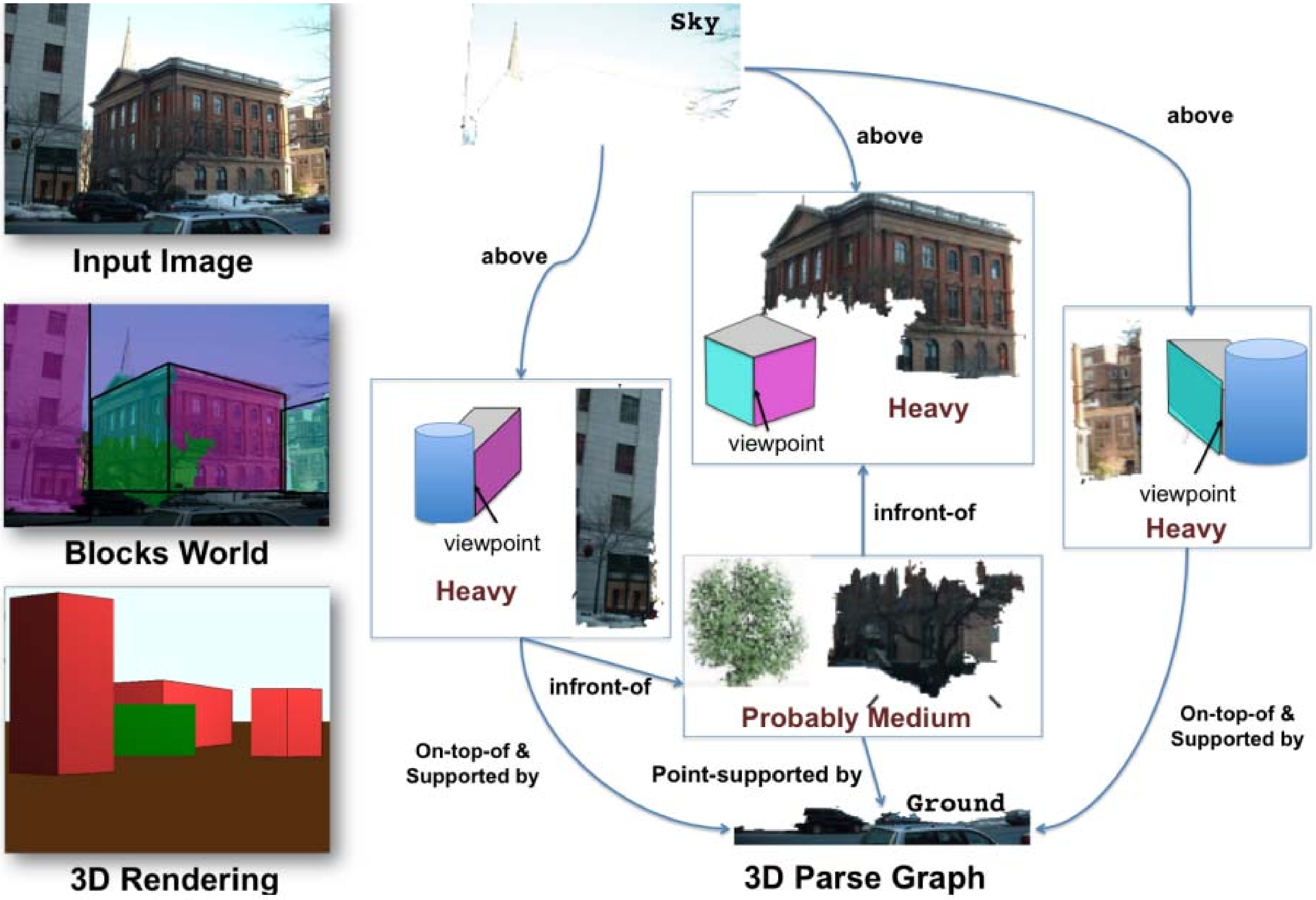
Source: J. Deng

Geometric image interpretation



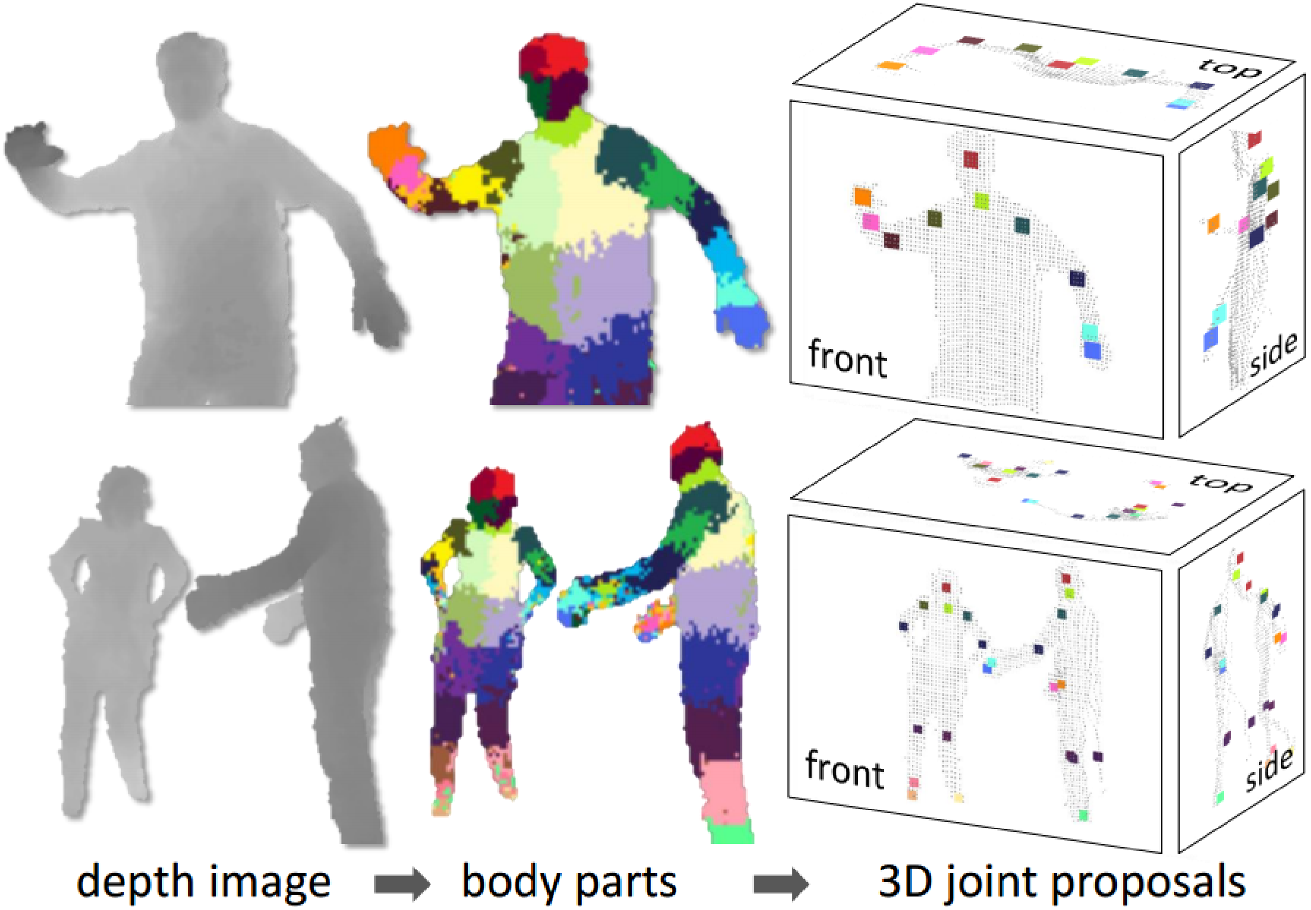
V. Hedau, D. Hoiem, and D. Forsyth, Recovering the Spatial Layout of Cluttered Rooms, ICCV 2009.

Geometric image interpretation



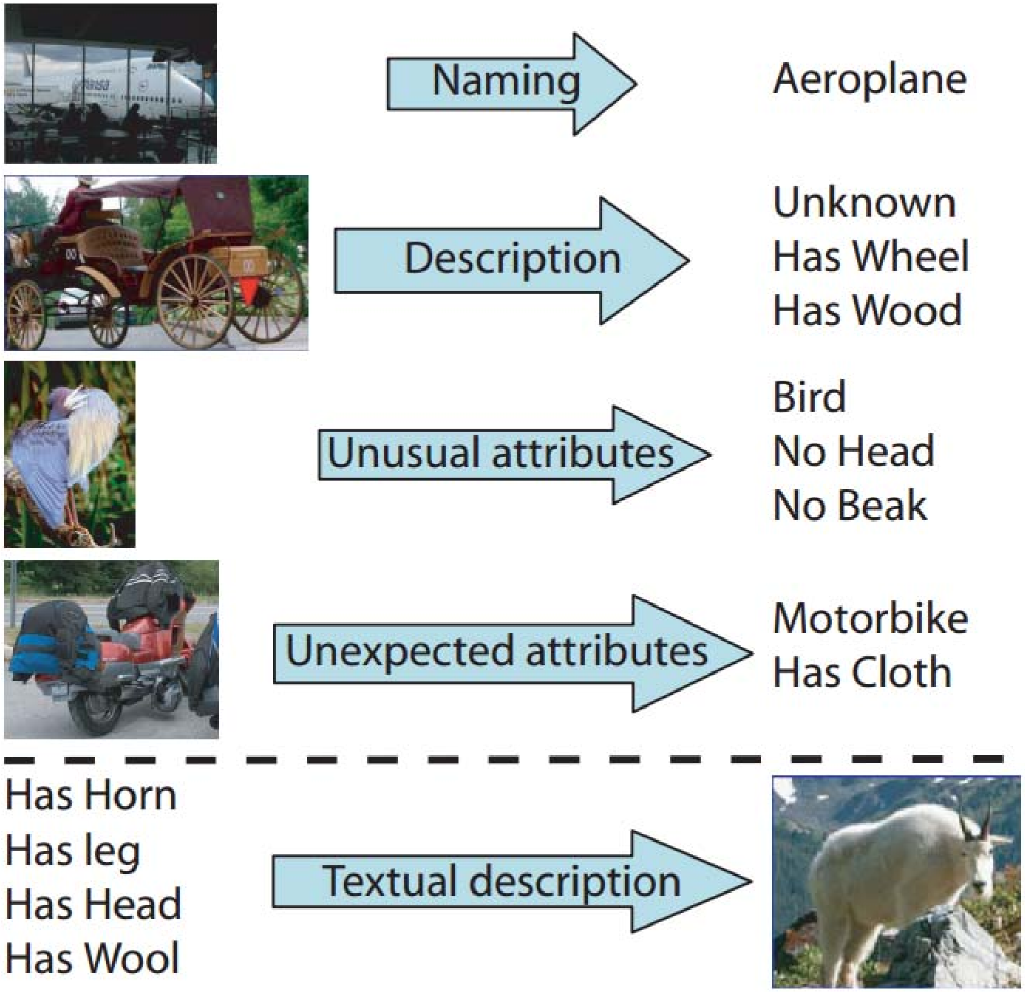
A. Gupta, A. Efros and M. Hebert, **Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics**, ECCV 2010

Recognition from RGBD Images



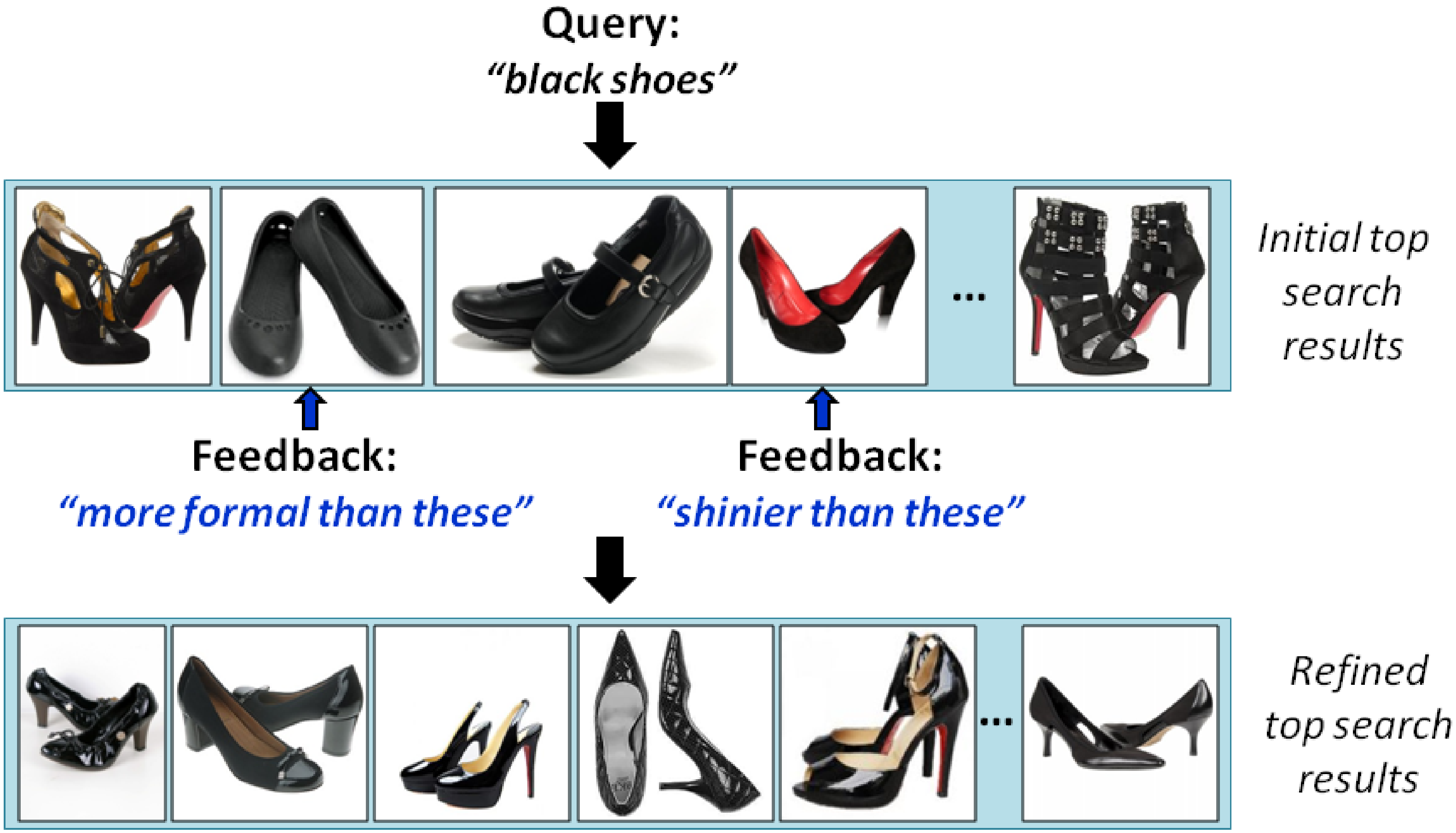
J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, **Real-Time Human Pose Recognition in Parts from a Single Depth Image**, CVPR 2011

Attribute-based recognition



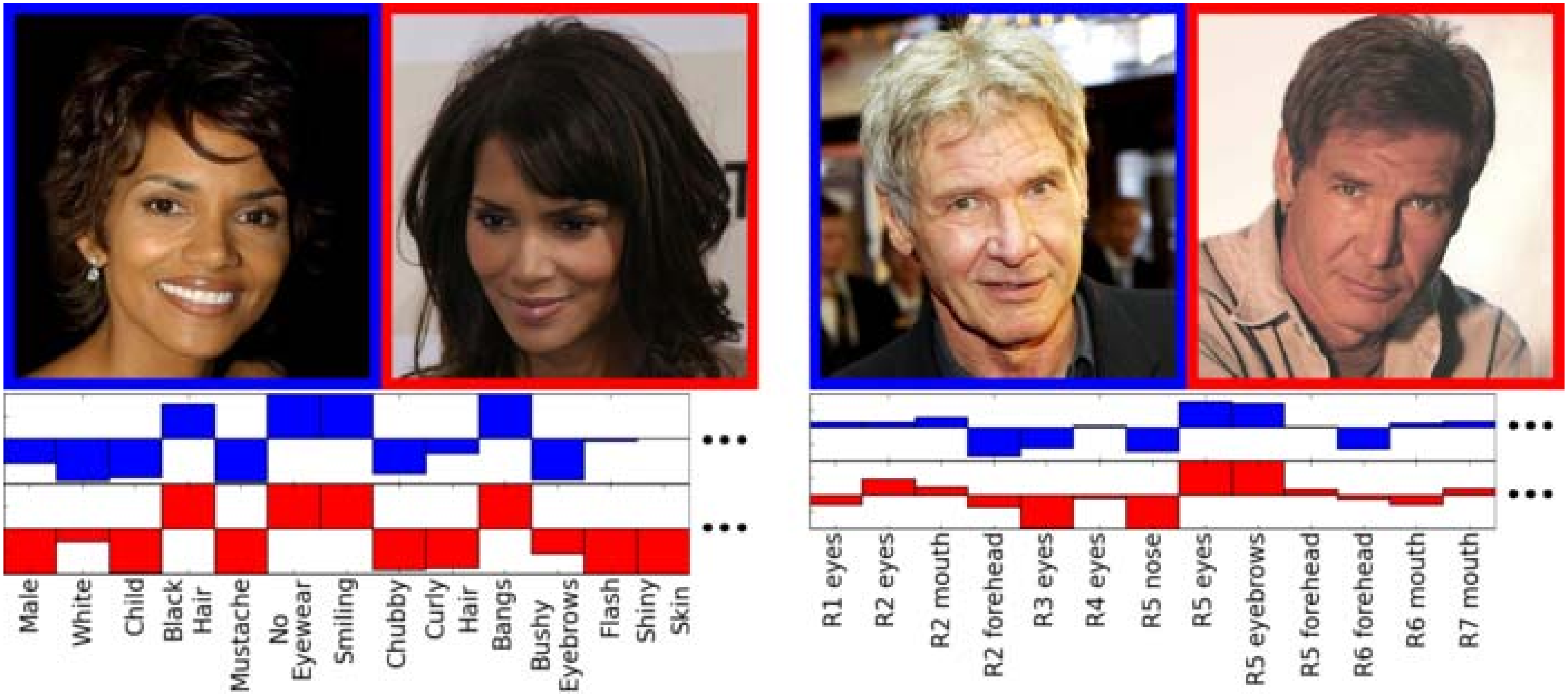
A. Farhadi, I. Endres, D. Hoiem, and D Forsyth, **Describing Objects by their Attributes**, CVPR 2009

Attribute-based search



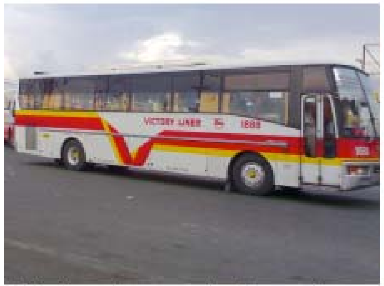
A. Kovashka, D. Parikh and K. Grauman, **WhittleSearch: Image Search with Relative Attribute Feedback**, CVPR 2012

Face verification

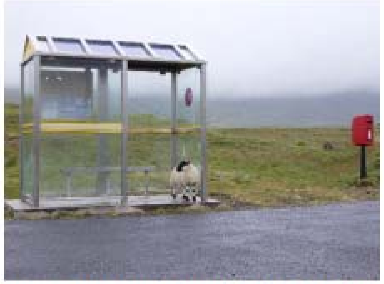


N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, Attribute and Simile Classifiers for Face Verification, ICCV 2009

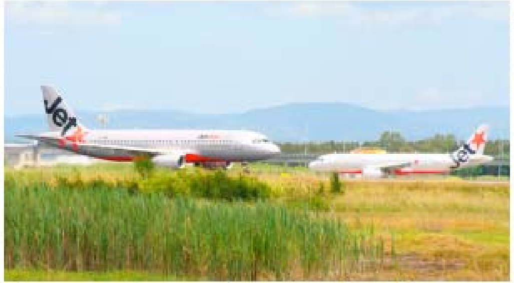
Sentence generation from images



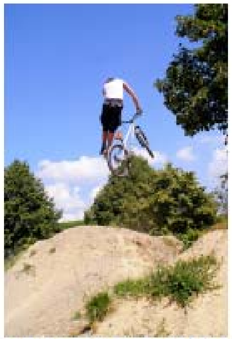
This is a photograph of one sky, one road and one bus. The blue sky is above the gray road. The gray road is near the shiny bus. The shiny bus is near the blue sky.



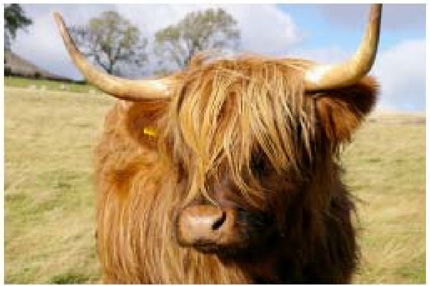
This is a picture of one sky, one road and one sheep. The gray sky is over the gray road. The gray sheep is by the gray road.



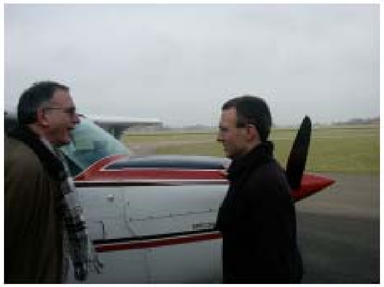
There are two aeroplanes. The first shiny aeroplane is near the second shiny aeroplane.



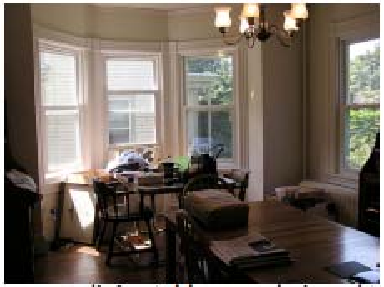
Here we see one road, one sky and one bicycle. The road is near the blue sky, and near the colorful bicycle. The colorful bicycle is within the blue sky.



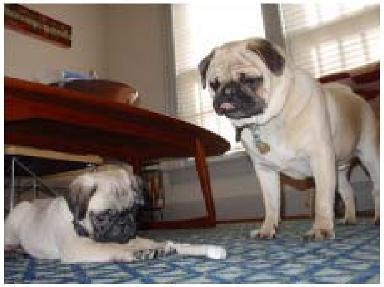
There are one cow and one sky. The golden cow is by the blue sky.



Here we see two persons, one sky and one aeroplane. The first black person is by the blue sky. The blue sky is near the shiny aeroplane. The second black person is by the blue sky. The shiny aeroplane is by the first black person, and by the second black person.



There are one dining table, one chair and two windows. The wooden dining table is by the wooden chair, and against the first window, and against the second white window. The wooden chair is by the first window, and by the second white window. The first window is by the second white window.



This is a picture of two dogs. The first dog is near the second furry dog.

G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. Berg, T. Berg, **Baby Talk: Understanding and Generating Simple Image Descriptions**, CVPR 2011