Seminars in Artificial Intelligence and Robotics

Computer Vision for Intelligent Robotics

Hints on classical object detection, classification and segmentation

Dipartimento di Ingegneria informatica automatica e gestionale Antonio Ruberti

Sapienza Università di Roma

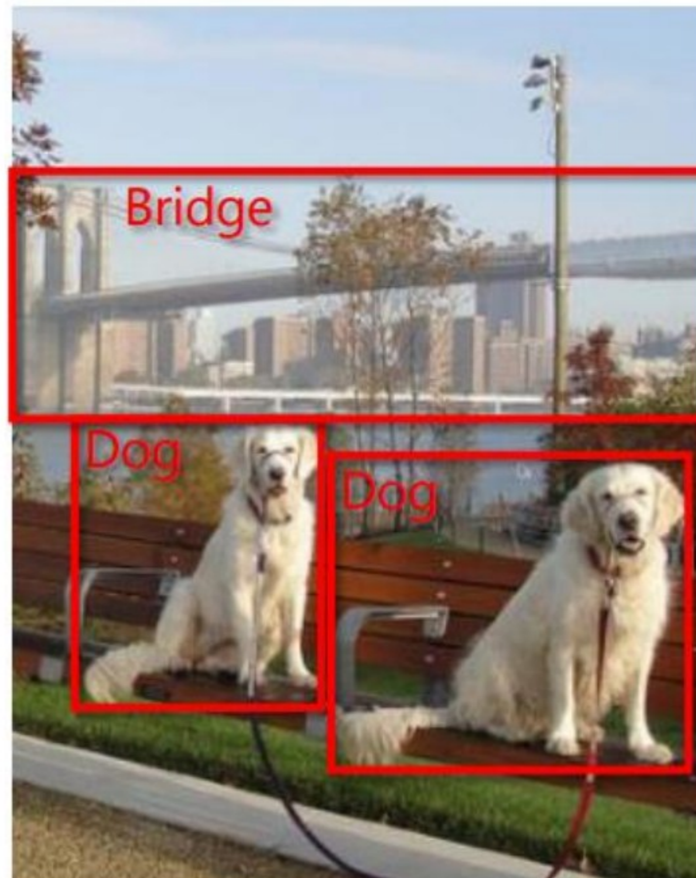
Alberto Pretto

(Mainly stripped version of the Bill Triggs ICVSS 2008 slides)

Classification Versus Detection

Classification: WHAT

Detection: WHAT and WHERE



Object detection challenges

Instances may occur anywhere in the image and at any scale

Instances may have variable geometry or internal degrees of freedom

There may be lighting variations, changes in appearance, complex backgrounds

Occlusions

Real-time constraints

A naive template matcher

Detection Phase

Scan image at all scales and locations

↓

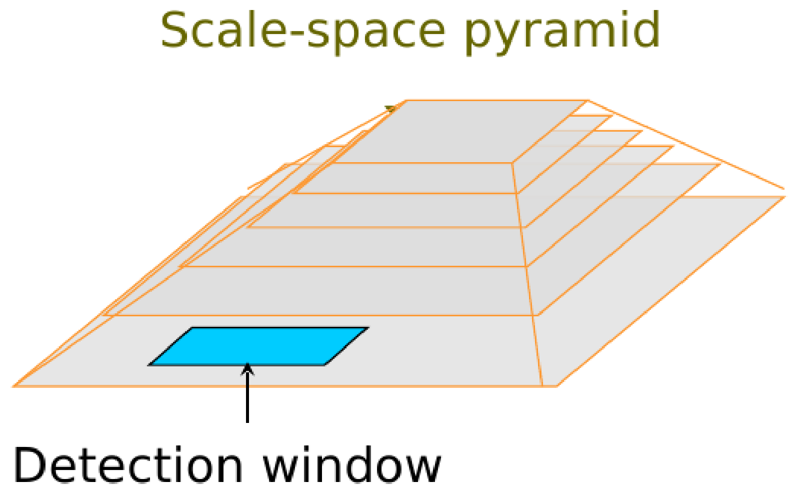
Match window against a rigid template, e.g. by correlation

↓

Return above-threshold matches as detections

↓

Object detections

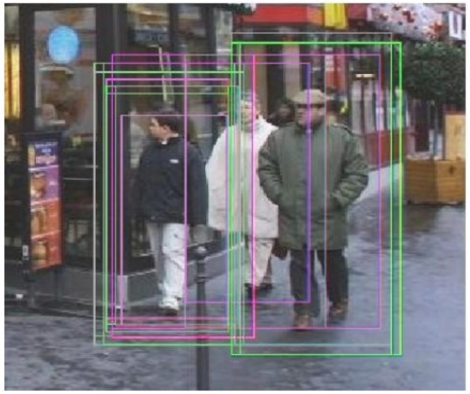


Problems with this approach

It is photometrically too rigid to resist changes in lighting and appearance variations

It is geometrically too rigid to resist shape variations

It does not have a strategy for overlapping detections



A better image detector

Detection Phase

Scan image(s) at all scales and locations

↓

Extract features over windows

↓

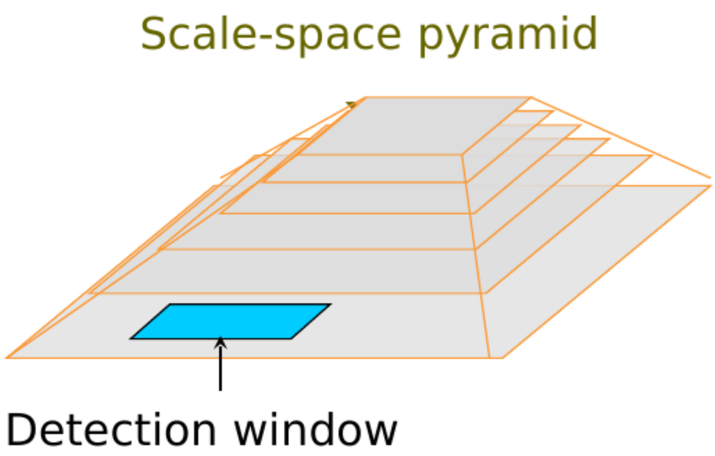
Run window classifier at all locations

↓

Fuse multiple detections in 3-D position & scale space

↓

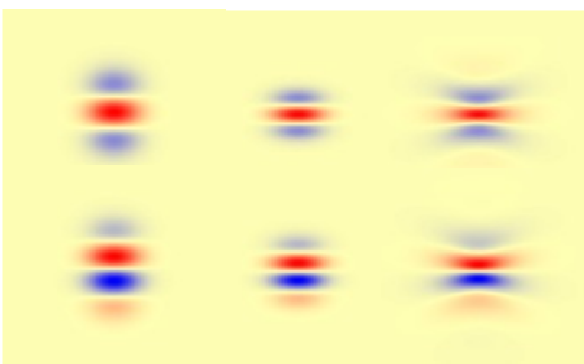
Object detections with bounding boxes



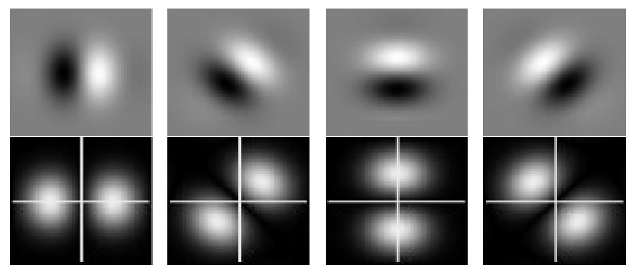
Detectors using Local Filters

Represent the image using simple convolution filters responses

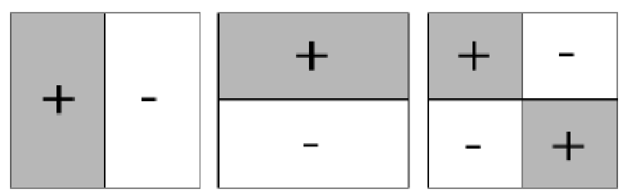
- Gaussian derivatives, Gabor filters, Haar wavelets, …



2^{nd} & 3^{rd} order Gaussian derivative, scaled Gaussian derivative and log-polar Gabor filters

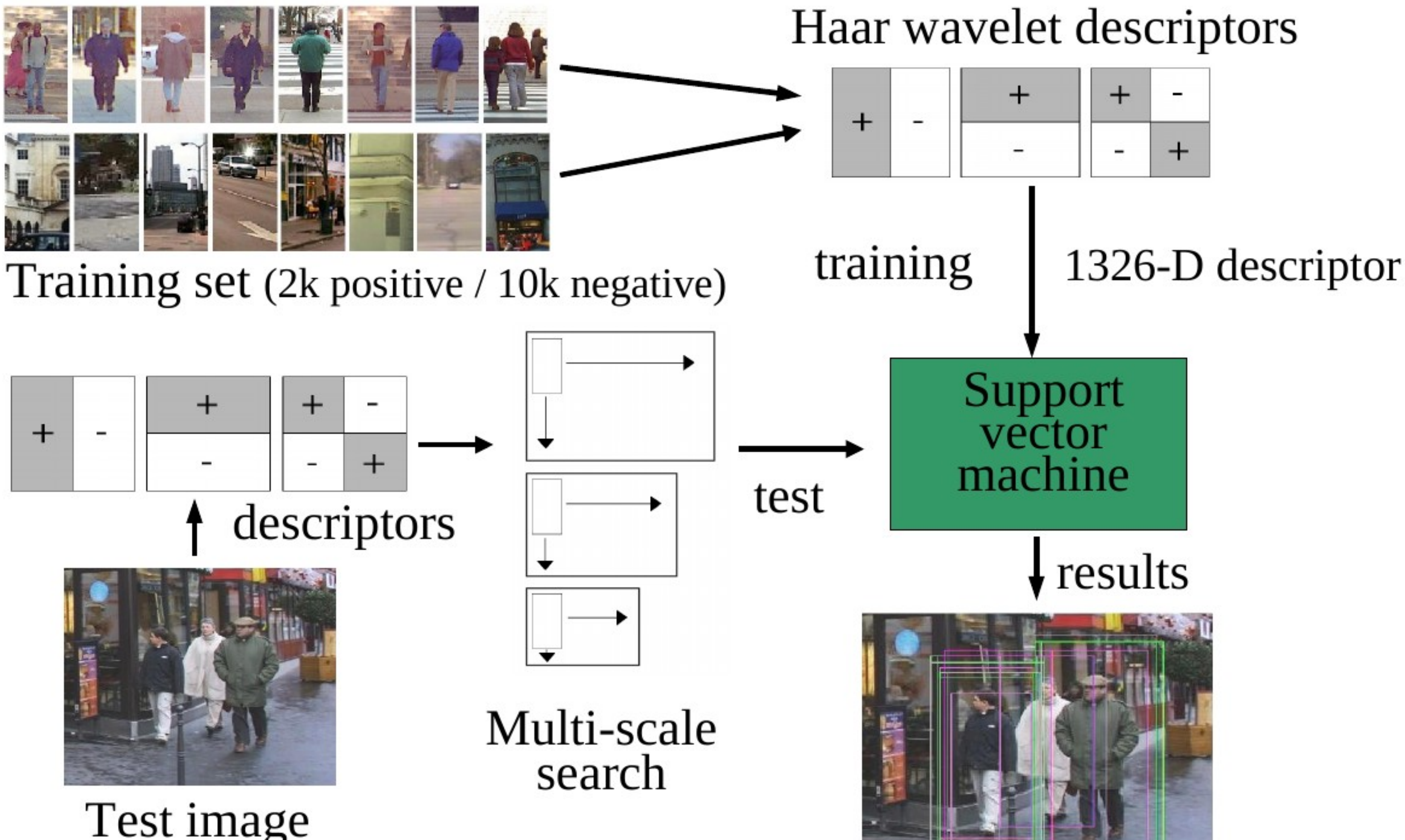


2^{nd} order steerable filter and its frequency response



Haar wavelets

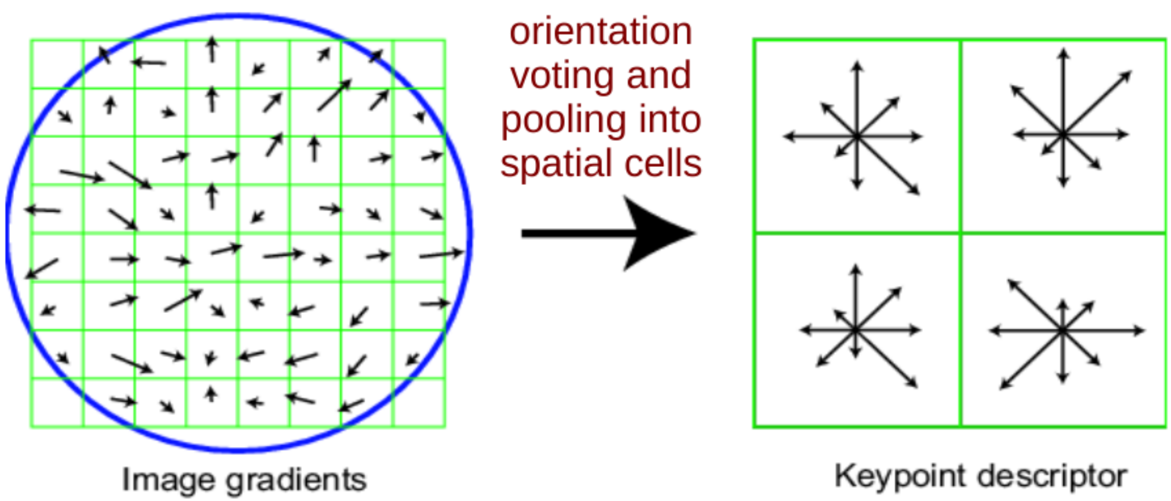
Haar Wavelet / SVM Human Detector



[Papageorgiou & Poggio, 1998]

Gradient Orientation Histograms

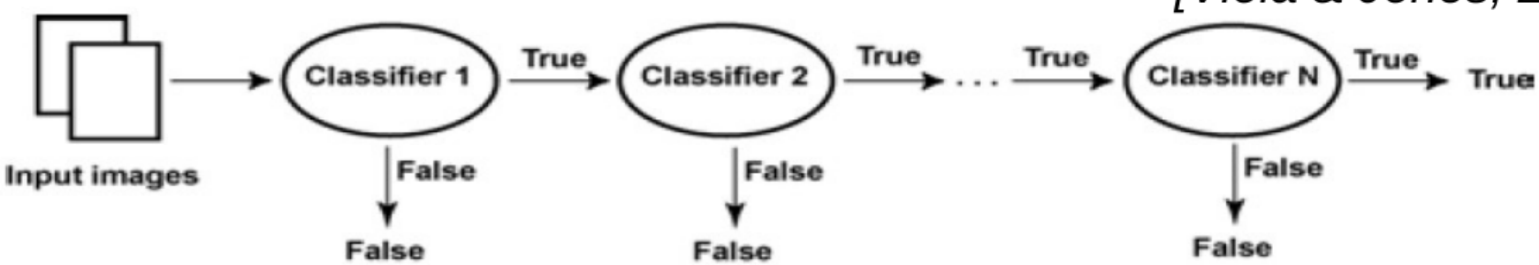
- Divide local region into spatial cells
- Calculate orientation of image gradient at each pixel
- Pool quantized orientations over each cell
  - descriptor contains an orientation histogram for each cell
  - weight votes by gradient magnitude
- Can also use edge orientations from a discrete edge detector
- Basis of the popular SIFT, HOG, Generalized Shape Context methods

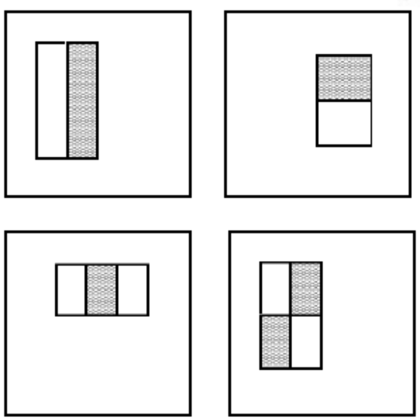


AdaBoost Cascade Face Detector

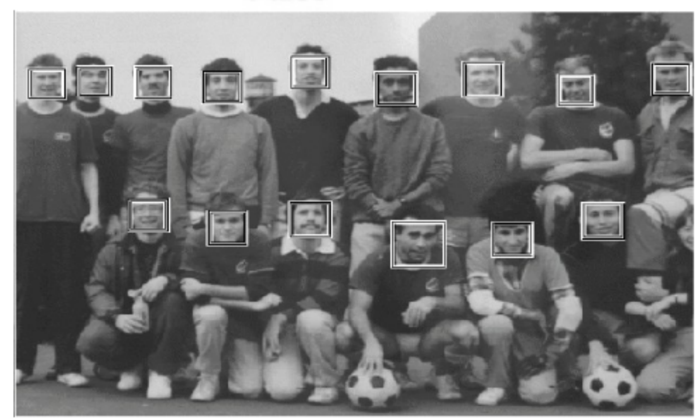
- A computationally efficient architecture that rapidly rejects unpromising windows
  - A chain of classifiers that each reject some fraction of the negative training samples while keeping almost all positive ones
- Each classifier is an AdaBoost ensemble of rectangular Harr-like features sampled from a large pool

[Viola & Jones, 2001]



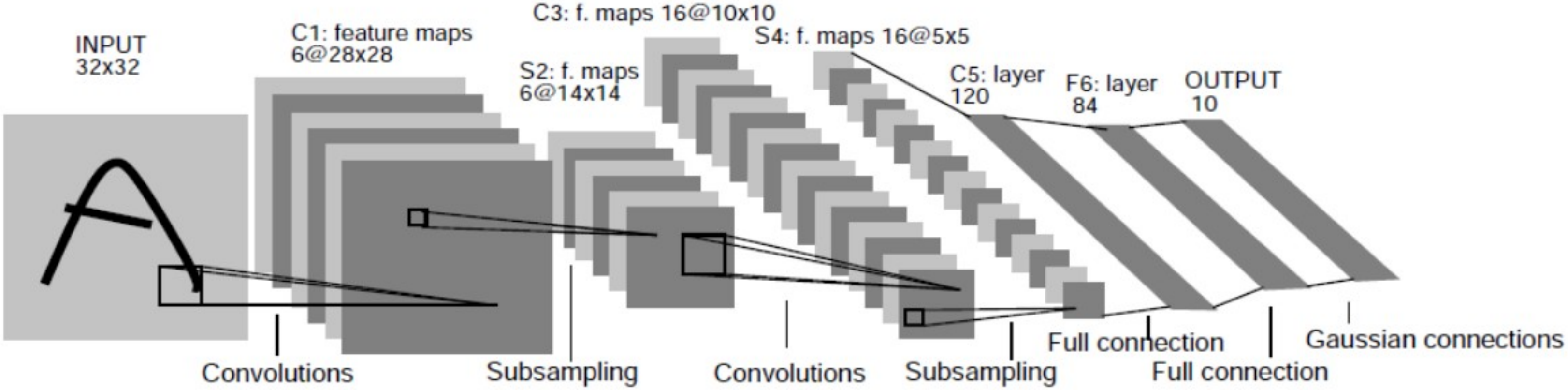


Rectangular Haar features and the first two features chosen by AdaBoost

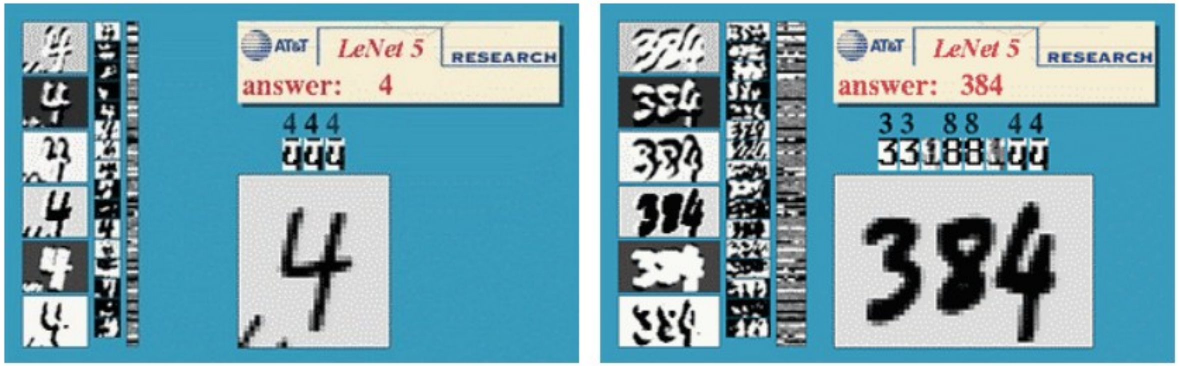


Convolutional Neural Nets

- A series of banks of convolution filters that alternately analyse the output images of the previous bank (“simple cells”) and spatially pool the resulting rectified responses (“complex cells”)
- Trained by gradient descent on large training sets



AT&T system – reads ~10% of U.S. cheques



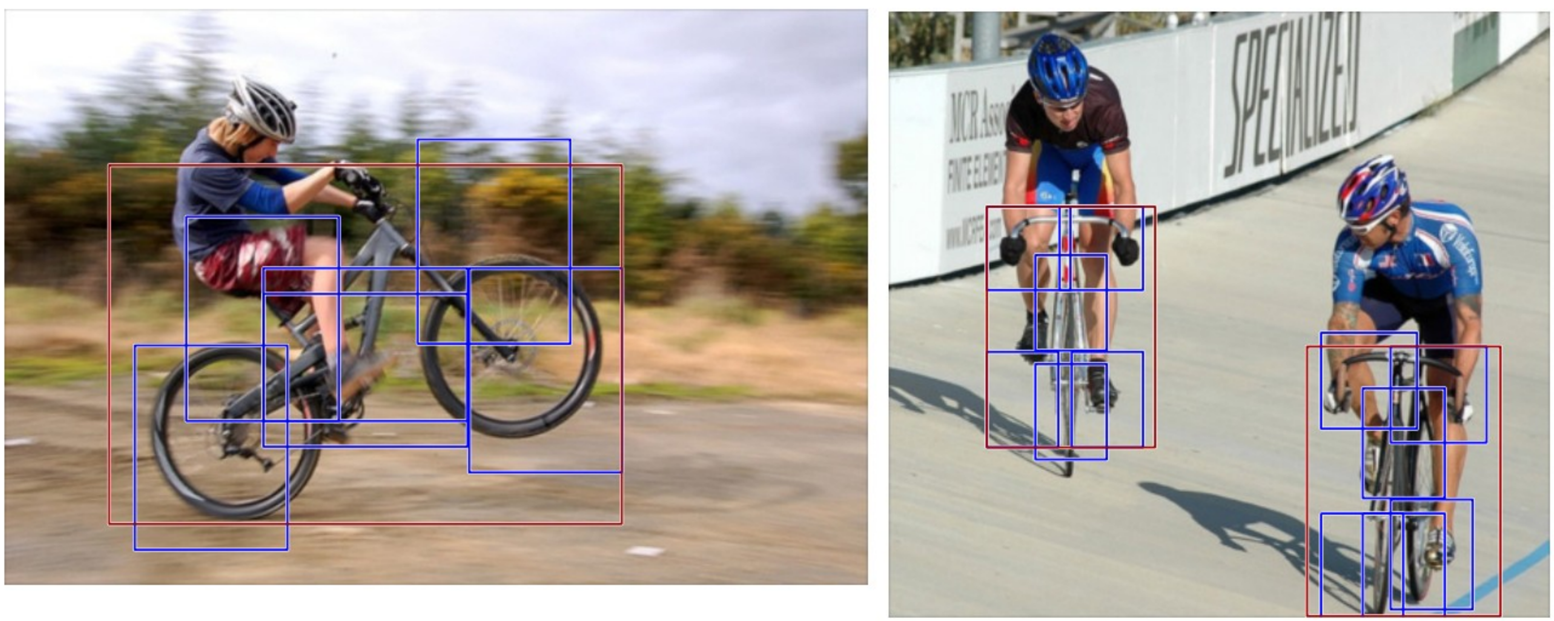
[Lecun 1992-8]

Detection with deformable part models

Collection of templates arranged in a deformable configuration

Each model has global template + part templates

Fully trained from bounding boxes alone



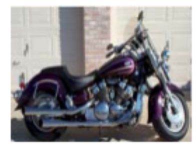
[Felzenszwalb et al. "Object Detection with Discriminatively Trained Part-Based Models", 2009]

Image classification

Given:

- positive training images containing an object class, and





- negative training images that don’t



Classify:

- a test image as to whether it contains the object class or not

?

Classification with bags of visual words

Represent each training image by a bag of visual words representation

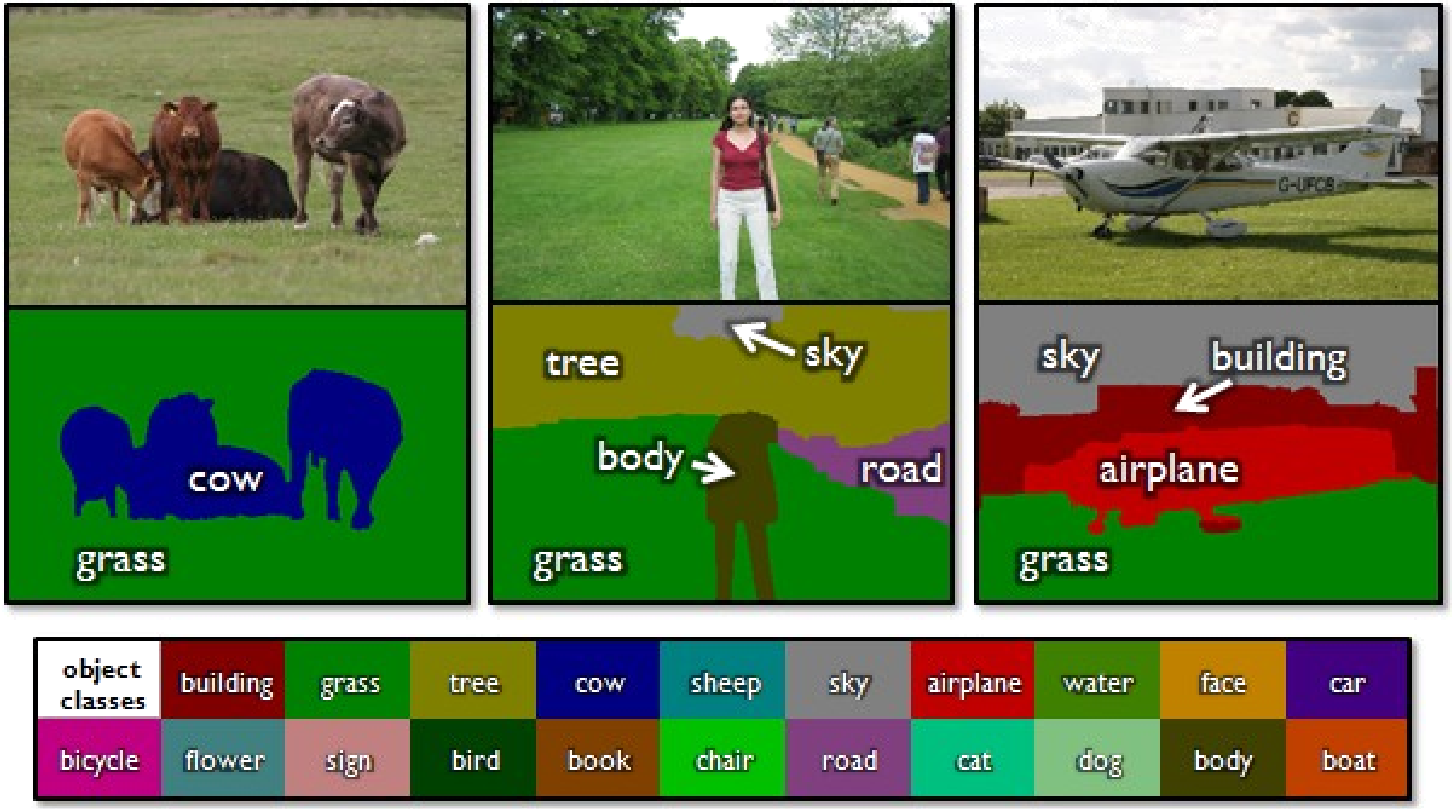
Train a classifier to discriminate vectors corresponding to positive and negative training images

Use e.g. a support Vector Machine (SVM) classifier

Apply the trained classifier to the test image

[Csurka et al. "Visual Categorization with Bags of Keypoints", 2004]

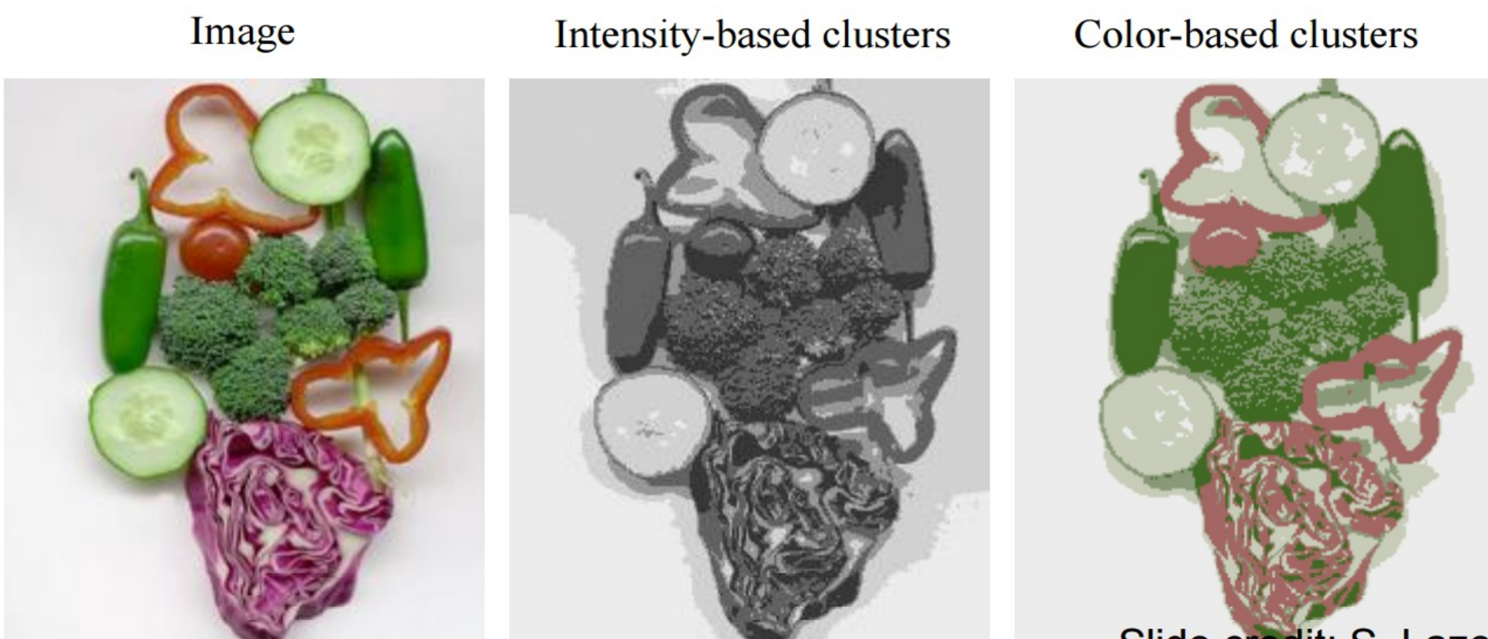
Image semantic segmentation



Segmentation as clustering

K-means clustering based on intensity or color or other local image statistics is essentially vector quantization of the image attributes

– Clusters don't have to be spatially coherent



Slide credit: S. Lazebnik

Using texture features for segmentation

- Convolve image with a bank of filters
- Find textons by clustering vectors of filter bank outputs
- The final texture feature is a texton histogram computed over image windows at some "local scale"