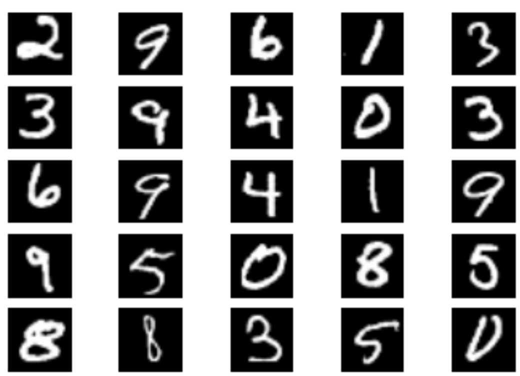
Object Detection

- Many slides are by Ross Girshick and Derek Hoiem

Image classification

- K classes
- Task: assign correct class label to the whole image



Digit classification (MNIST)

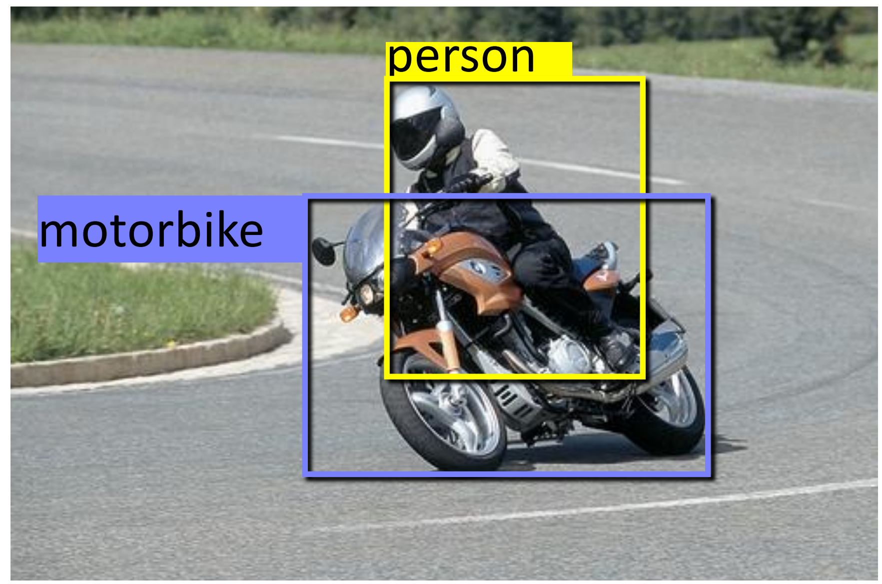
Object recognition (Caltech-101)

Classification vs. Detection

Problem formulation

{ airplane, bird, motorbike, person, sofa }

Input



Desired output

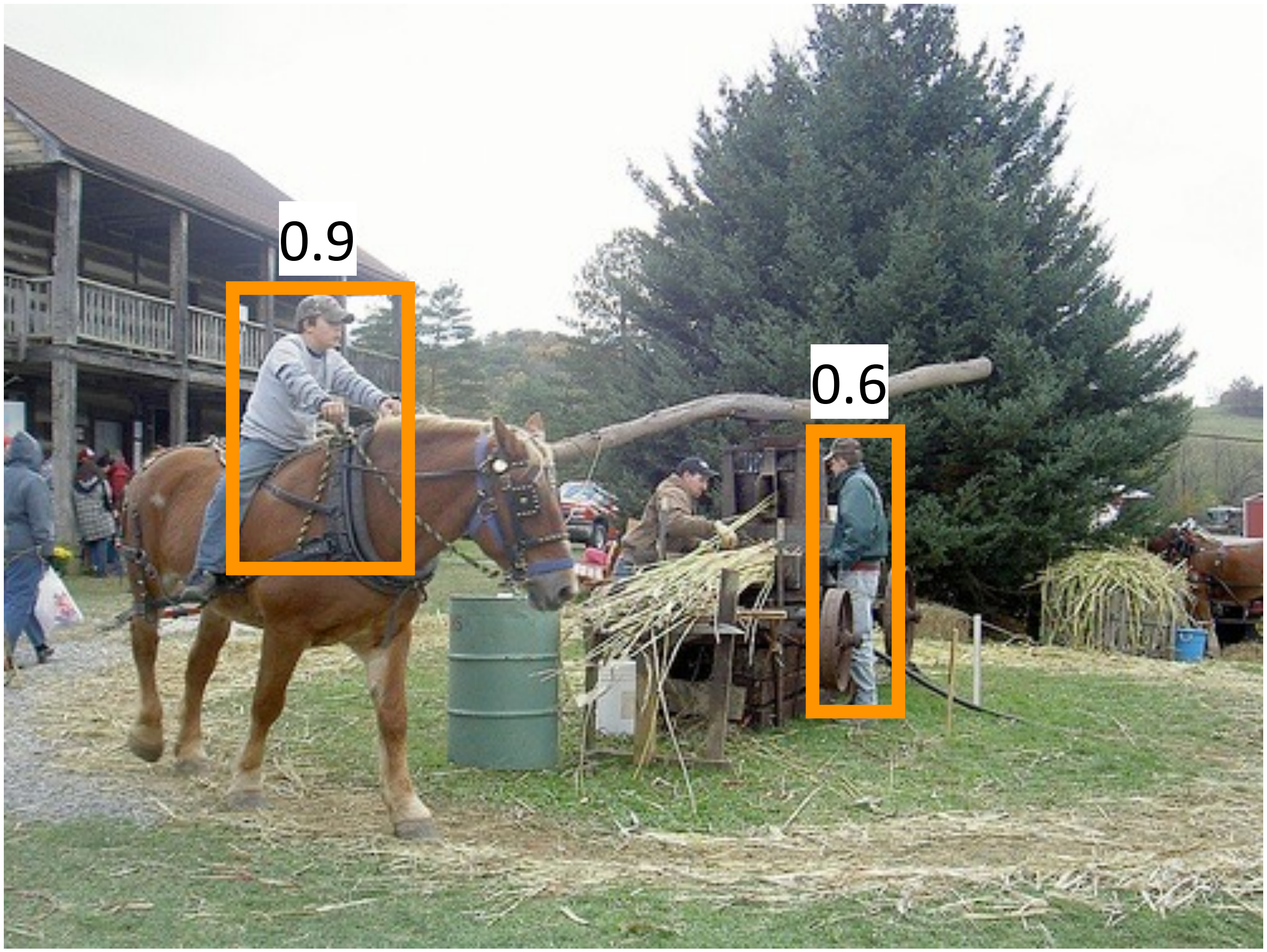
Evaluating a Detector

Test image (previously unseen)

First Detection

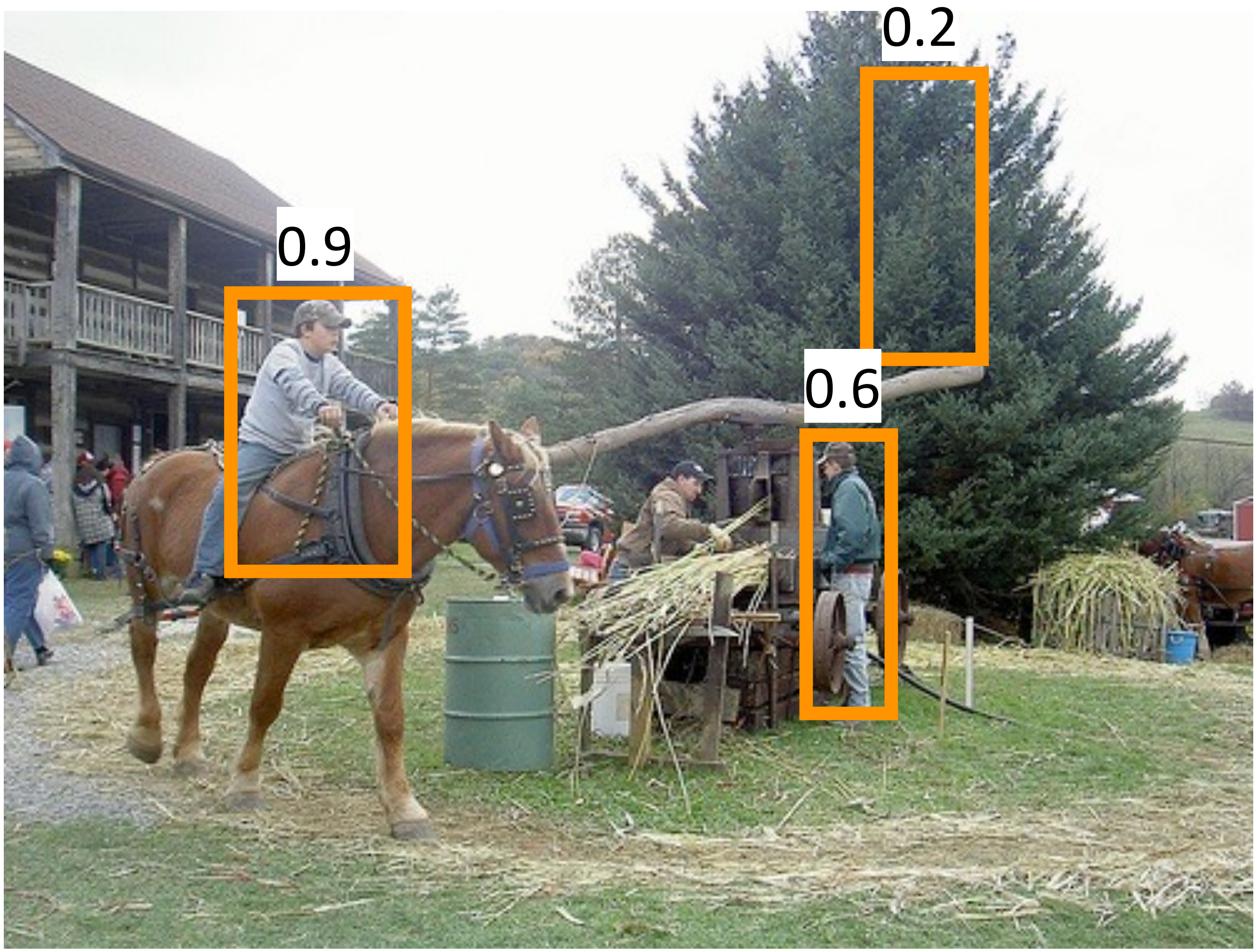
‘person’ detector predictions

Second Detection



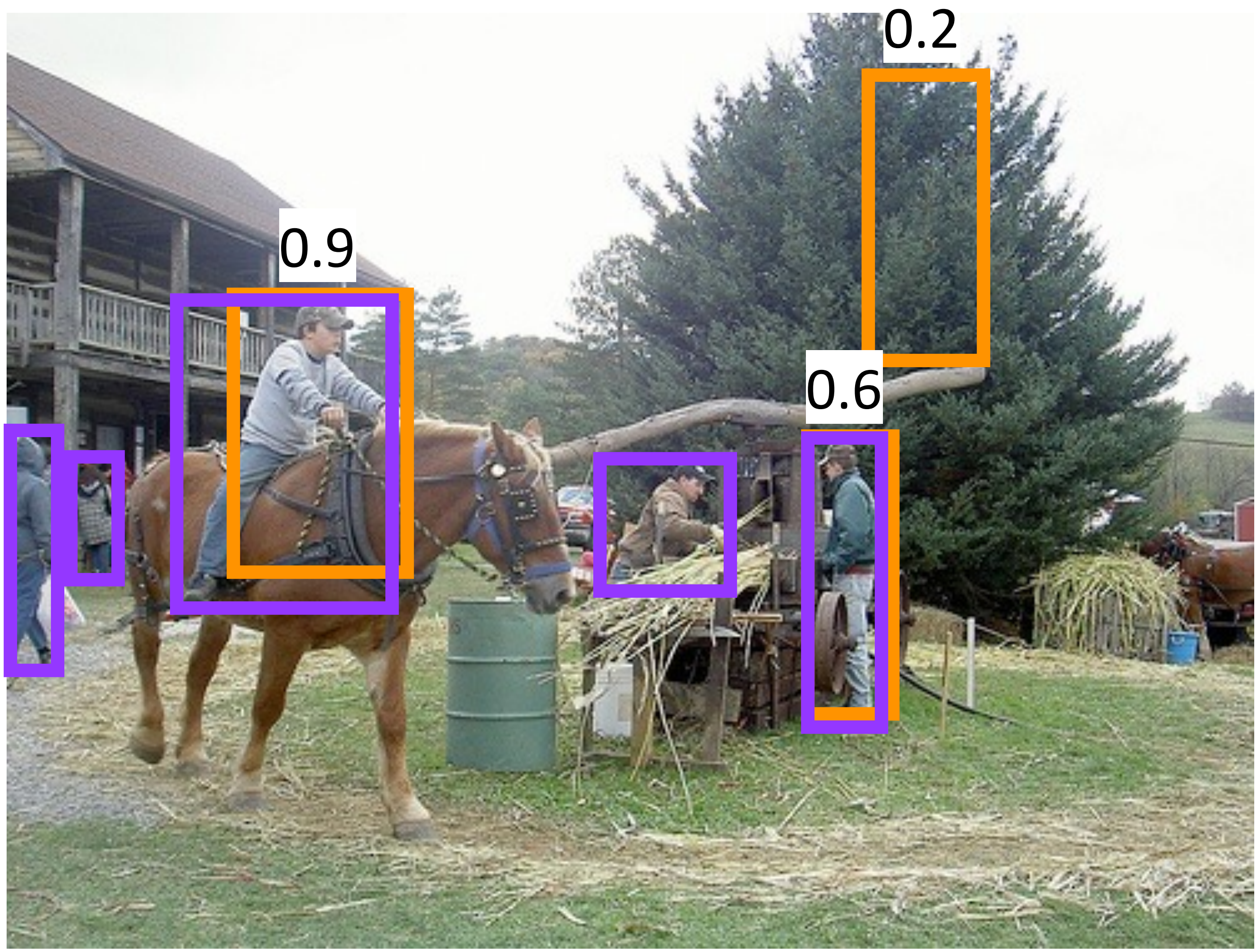
‘person’ detector predictions

Third Detection



'person' detector predictions

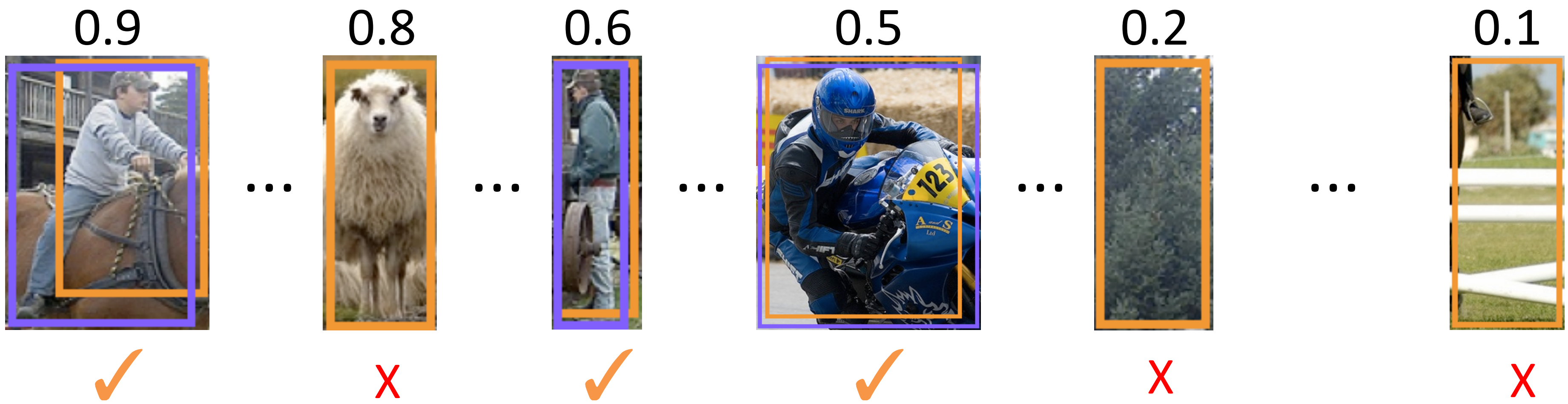
Compare to Groundtruth



‘person’ detector predictions

ground truth ‘person’ boxes

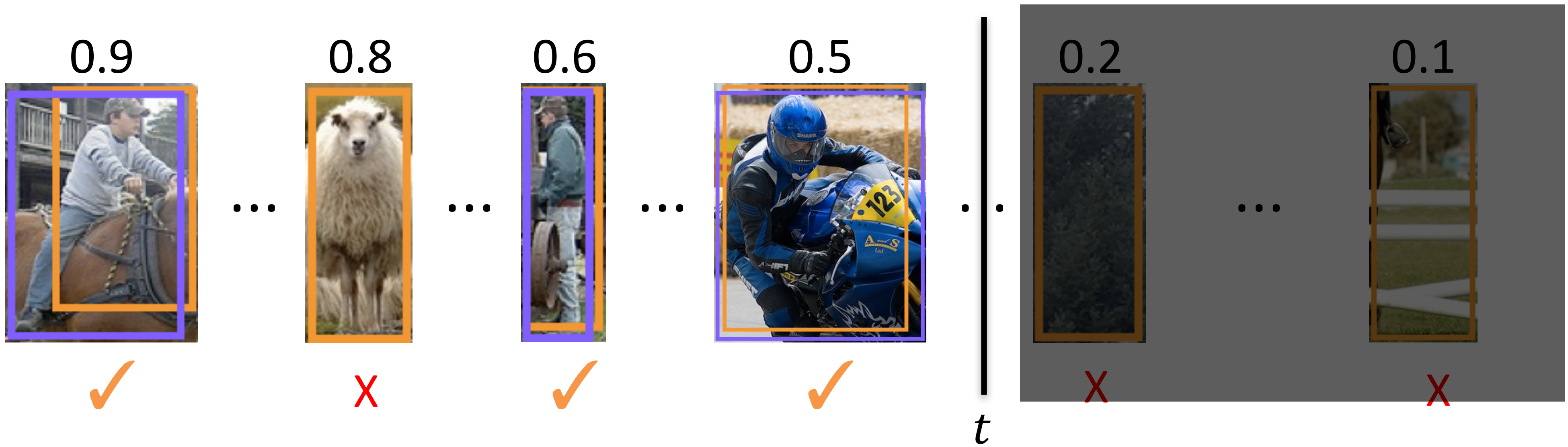
Sort by Confidence



true positive (high overlap)

false positive (no overlap, low overlap, or duplicate)

Evaluation Metric

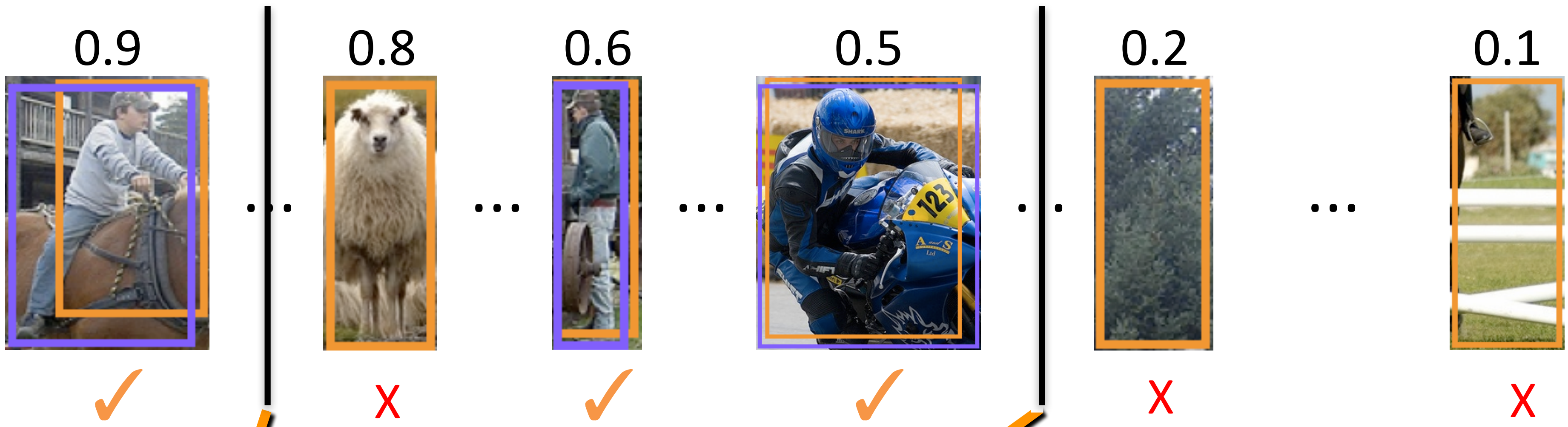


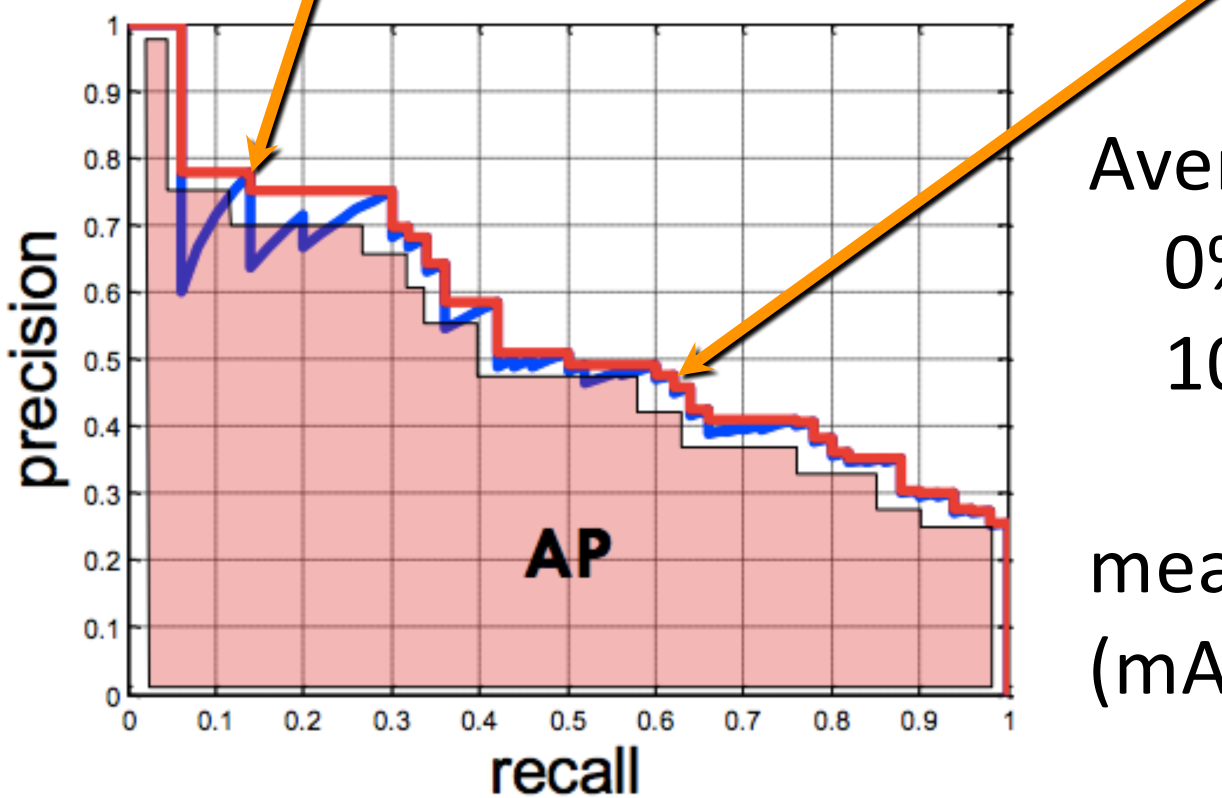
$$precision@t = \frac{\#true\ positives@t}{\#true\ positives@t + \#false\ positives@t}$$

$$\frac{\checkmark}{\checkmark + X}$$

$$recall@t = \frac{\#true\ positives@t}{\#ground\ truth\ objects}$$

Evaluation Metric





Average Precision (AP)

0% is worst

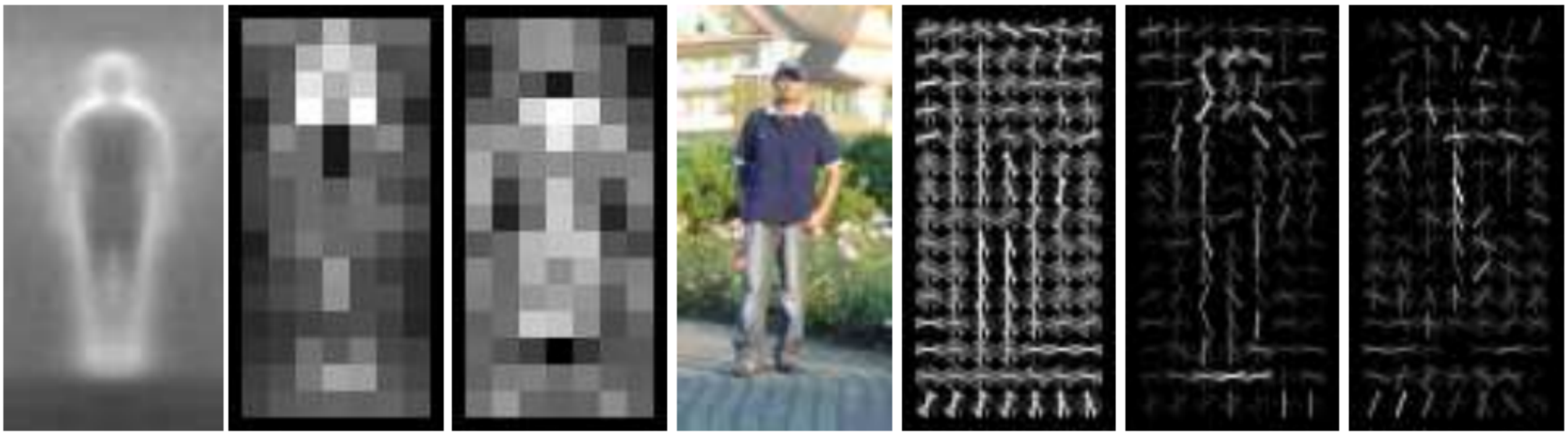
100% is best

mean AP over classes (mAP)

Pedestrians

Histograms of Oriented Gradients for Human Detection, Dalal and Triggs, CVPR 2005

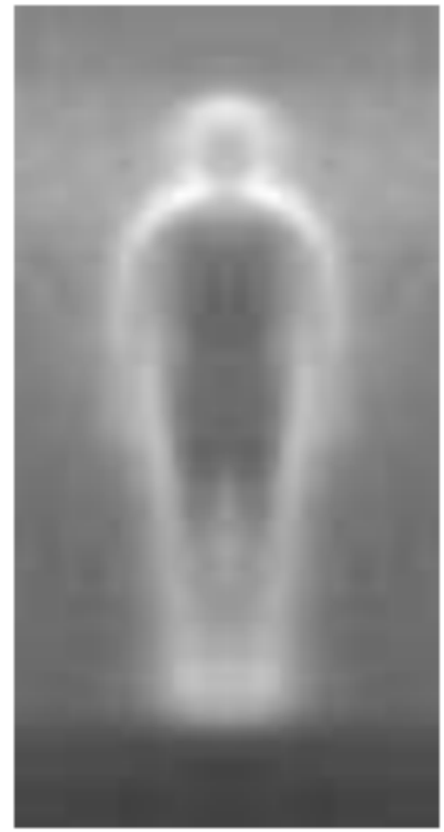
AP ~77%
More sophisticated methods: AP ~90%



(a) (b) (c) (d) (e) (f) (g)

(a) average gradient image over training examples
(b) each “pixel” shows max positive SVM weight in the block centered on that pixel
(c) same as (b) for negative SVM weights
(d) test image
(e) its R-HOG descriptor
(f) R-HOG descriptor weighted by positive SVM weights
(g) R-HOG descriptor weighted by negative SVM weights

Why did it work?



Average gradient image

Generic categories

Can we detect people, chairs, horses, cars, dogs, buses, bottles, sheep …?

PASCAL Visual Object Categories (VOC) dataset

Generic categories

- Why doesn’t this work (as well)?

Can we detect people, chairs, horses, cars, dogs, buses, bottles, sheep …?

PASCAL Visual Object Categories (VOC) dataset

Quiz time

Warm up

This is an average image of which object class?

Warm up

pedestrian

A little harder

?

A little harder

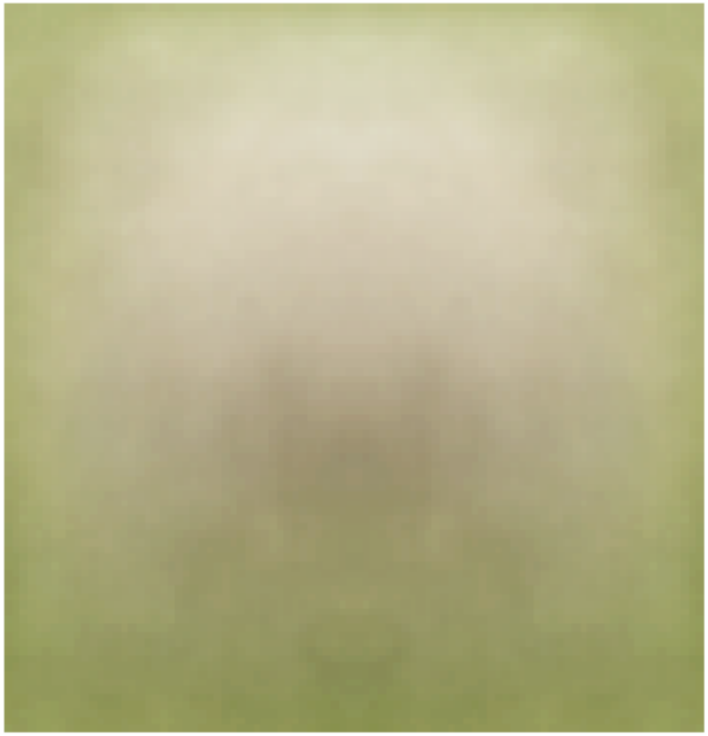
?

Hint: airplane, bicycle, bus, car, cat, chair, cow, dog, dining table

A little harder

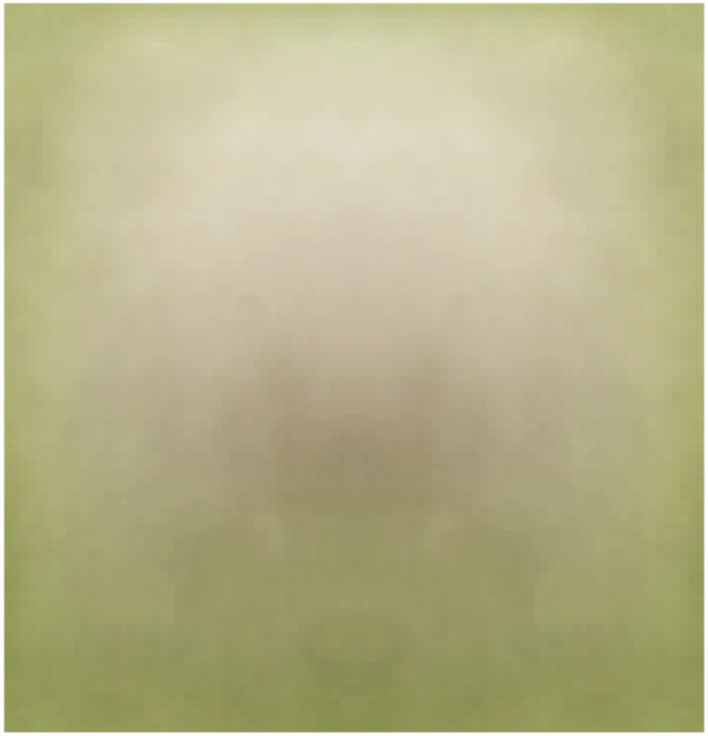
bicycle (PASCAL)

A little harder, yet



?

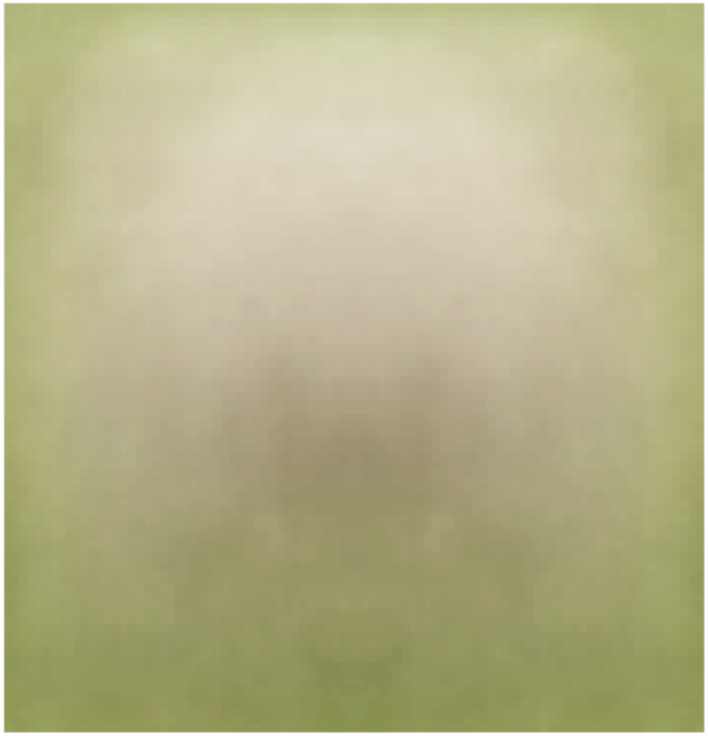
A little harder, yet



?

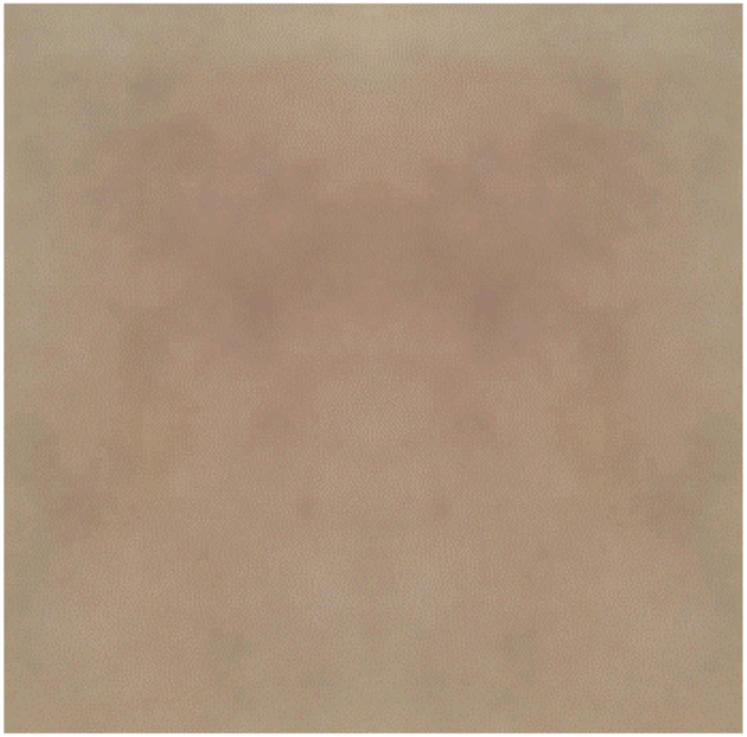
Hint: white blob on a green background

A little harder, yet



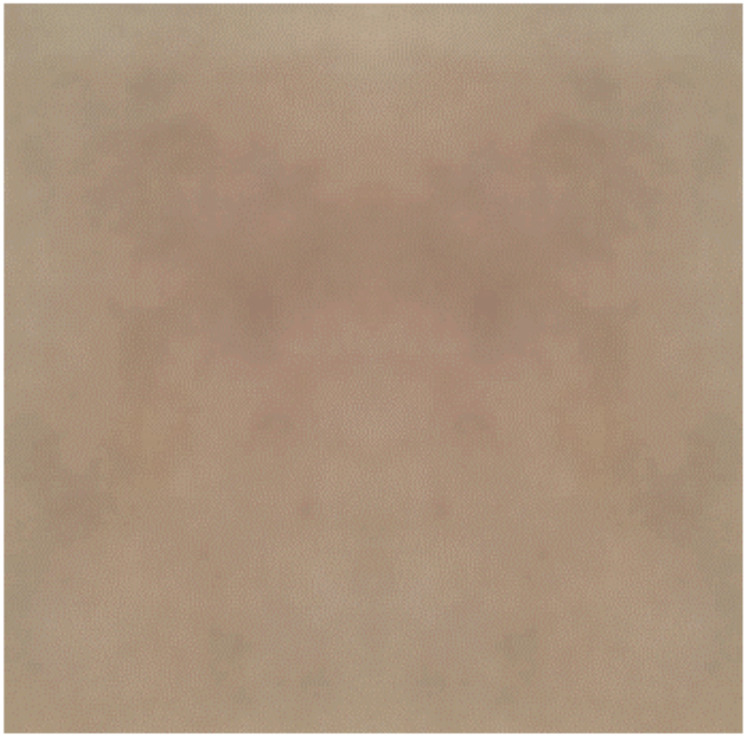
sheep (PASCAL)

Impossible?



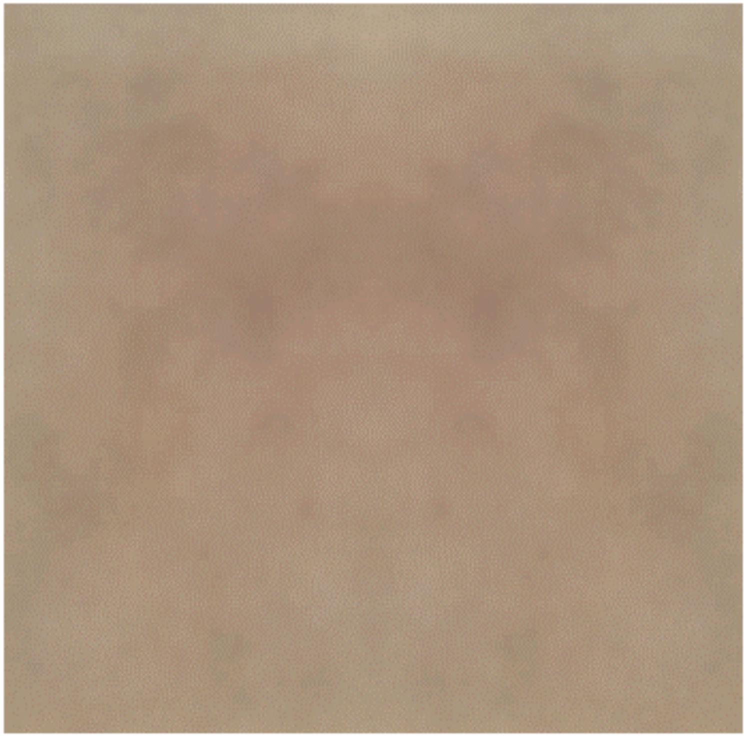
?

Impossible?



dog (PASCAL)

Impossible?



dog (PASCAL)

Why does the mean look like this?

There’s no alignment between the examples!

How do we combat this?

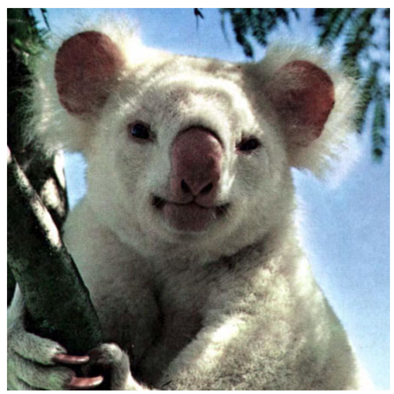
Challenges in modeling the object class

Illumination

Object pose

Clutter

Occlusions



Intra-class appearance

Viewpoint

Slide from K. Grauman, B. Leibe

Challenges in modeling the object class

True
Detections

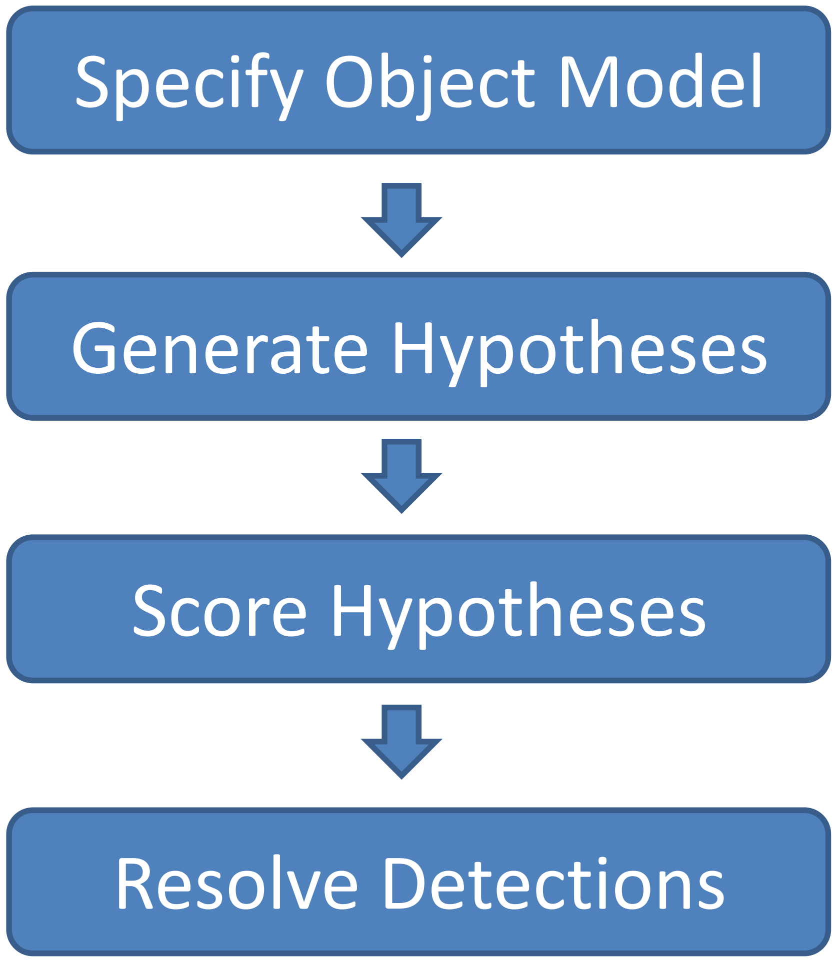
Bad
Localization

Confused with
Similar Object

Misc. Background

Confused with
Dissimilar Objects

General Process of Object Recognition

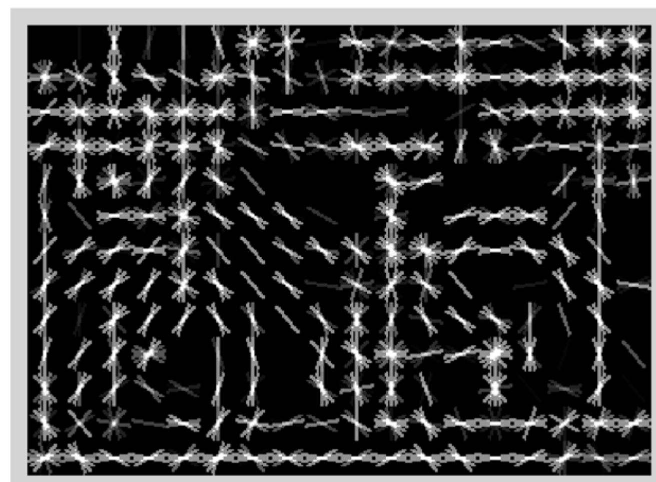


What are the object parameters?

Specifying an object model

1. Statistical Template in Bounding Box
   - Object is some (x,y,w,h) in image
   - Features defined wrt bounding box coordinates

Image

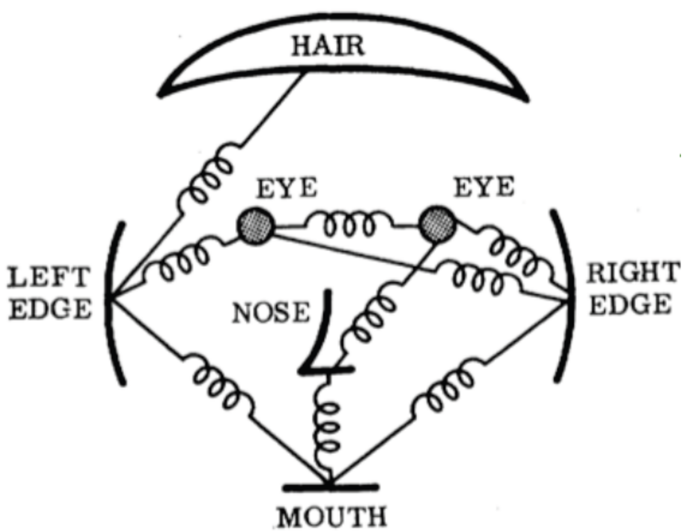


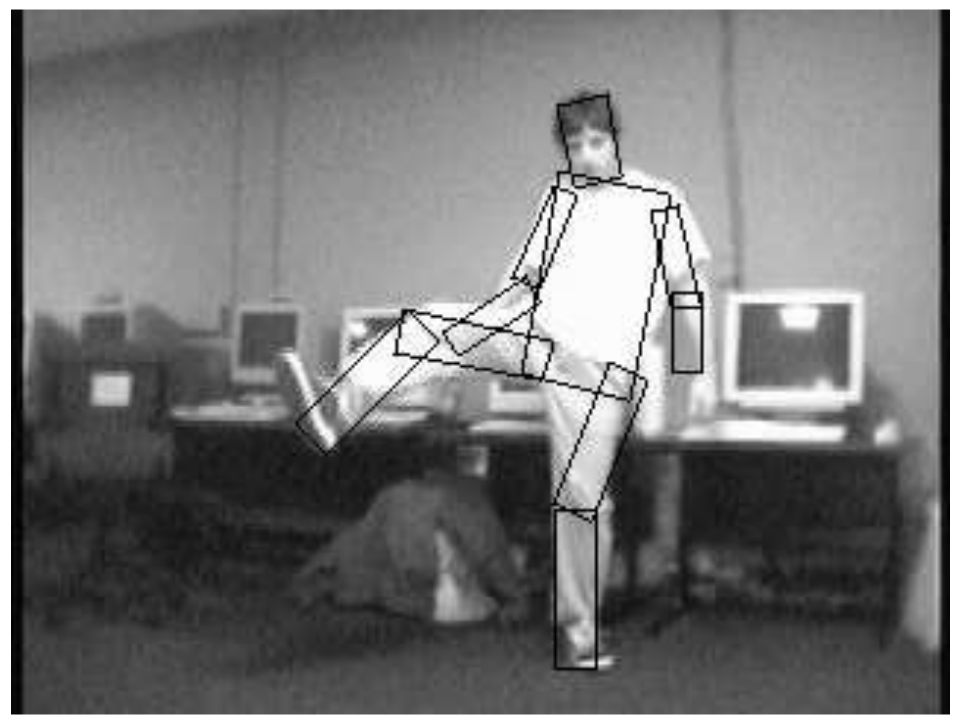
Template Visualization

Images from Felzenszwalb

Specifying an object model

2. Articulated parts model
   – Object is configuration of parts
   – Each part is detectable



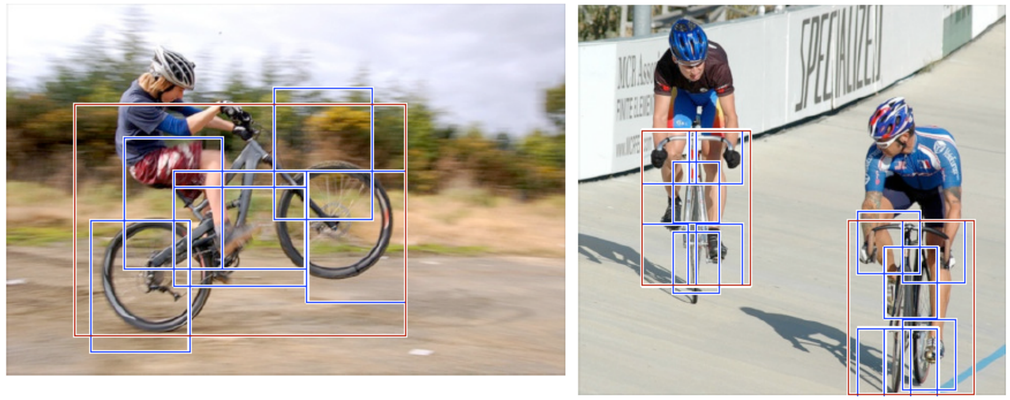


Images from Felzenszwalb

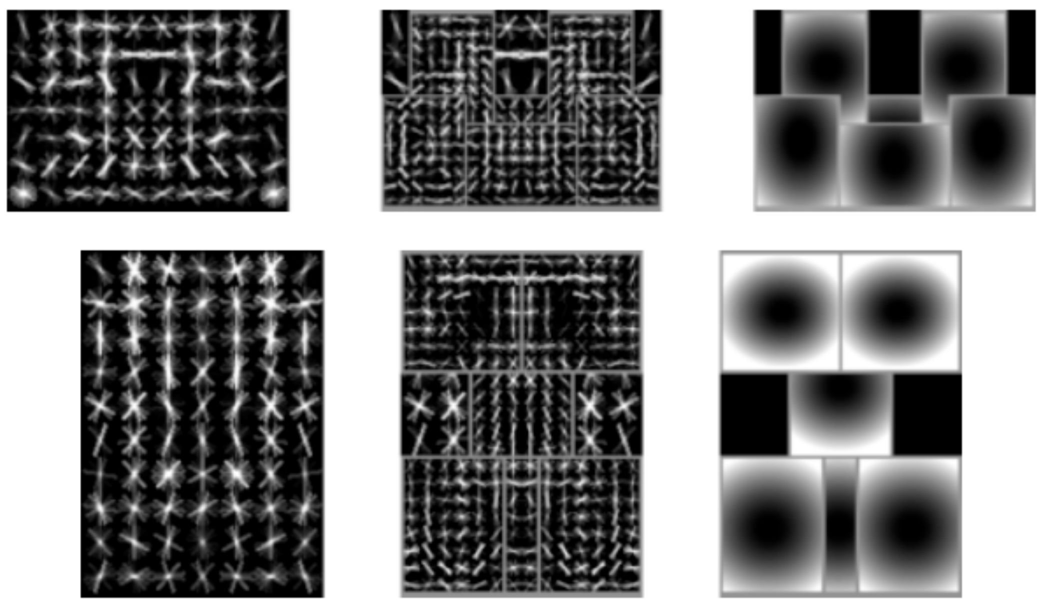
Specifying an object model

3. Hybrid template/parts model

Detections



Template Visualization



root filters
coarse resolution

part filters
finer resolution

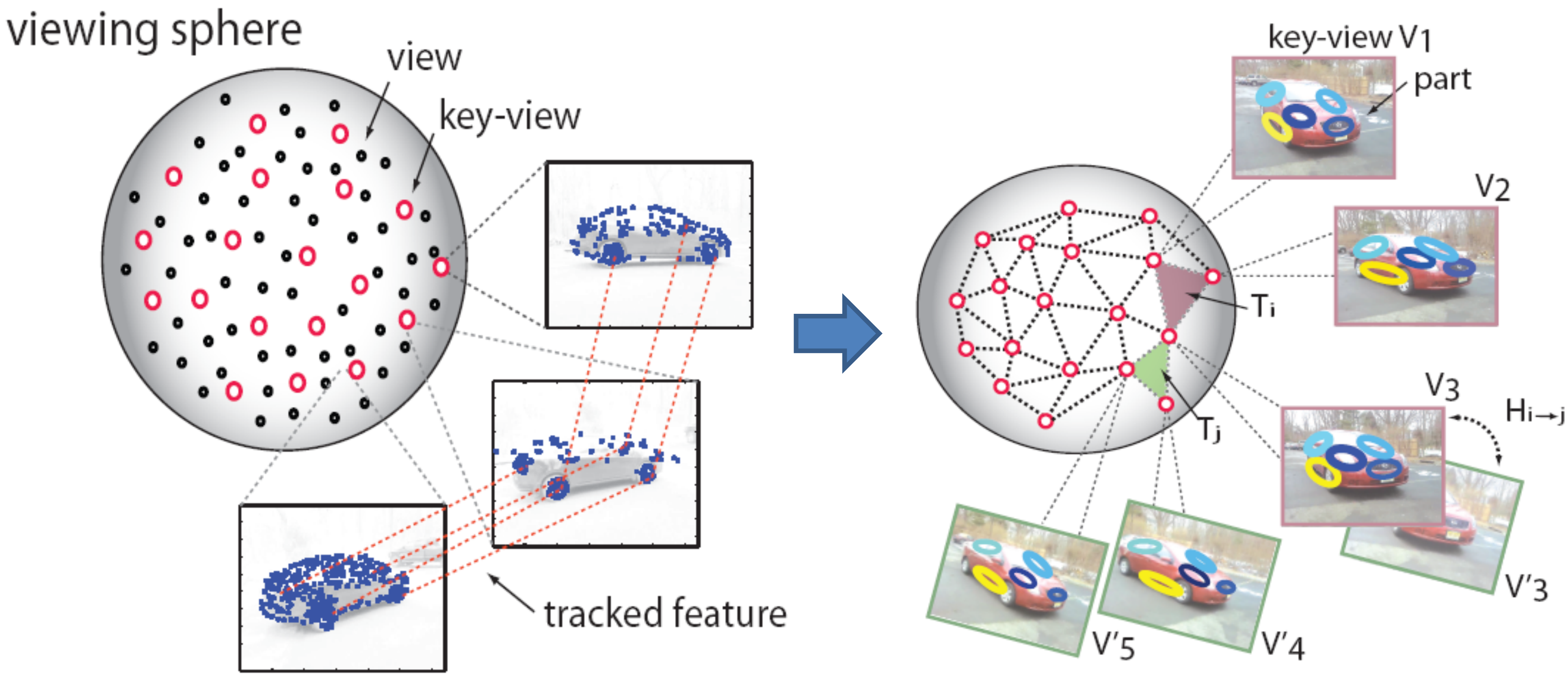
deformation
models

Felzenszwalb et al. 2008

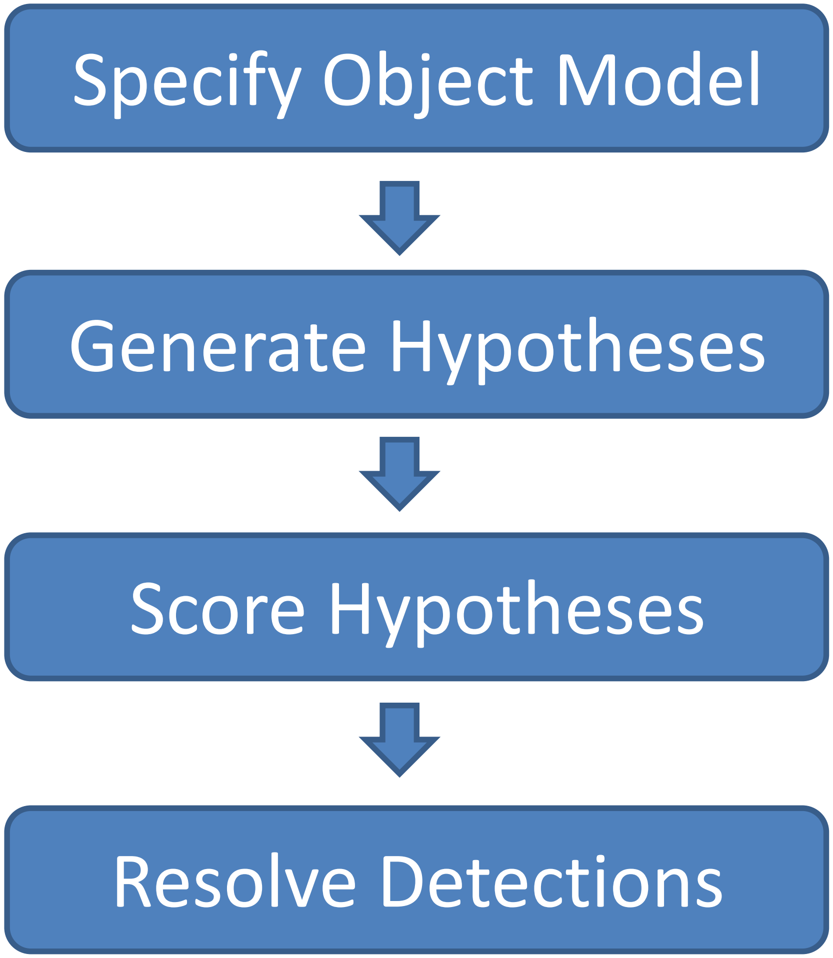
Specifying an object model

4. 3D-ish model

- Object is collection of 3D planar patches under affine transformation



General Process of Object Recognition



Propose an alignment of the model to the image

Generating hypotheses

1. Sliding window
   – Test patch at each location and scale

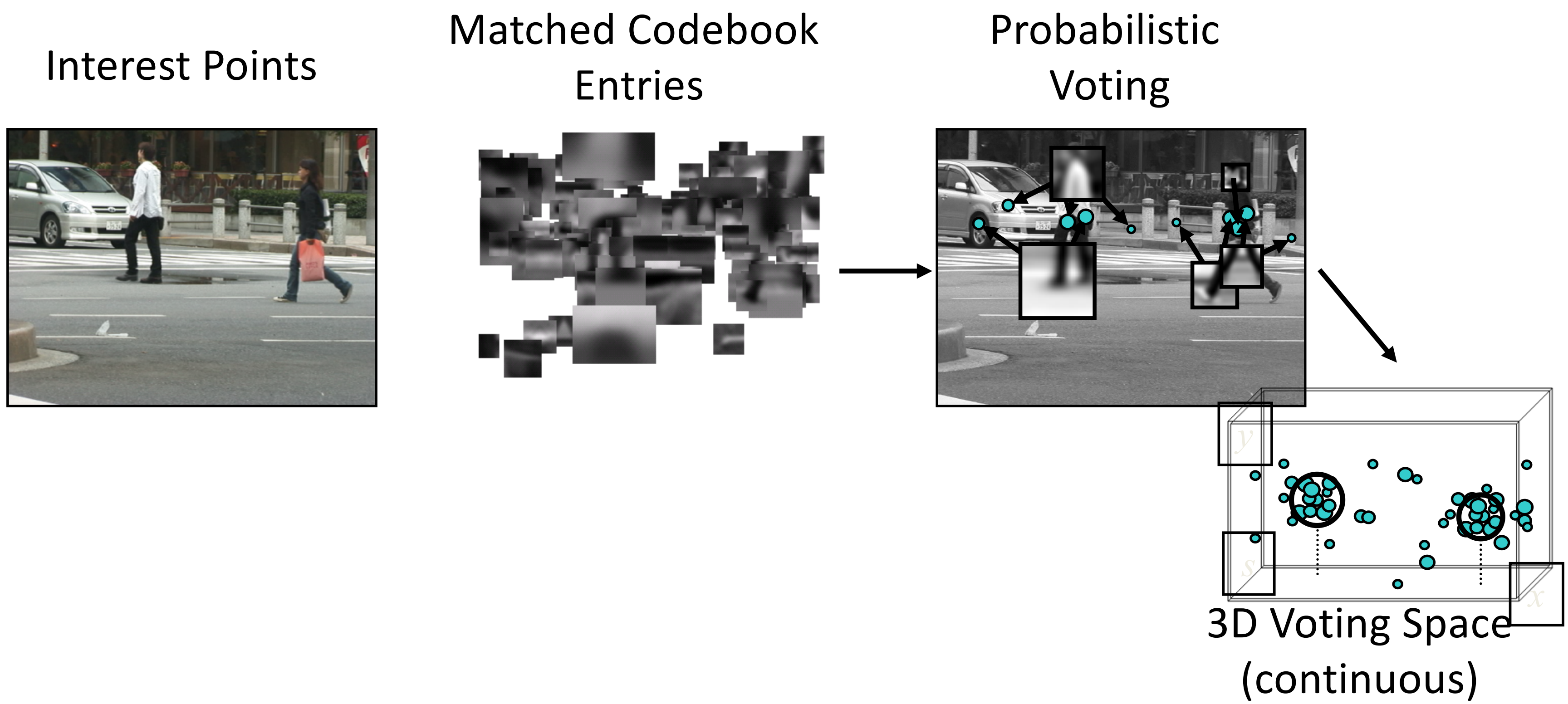
Generating hypotheses

1. Sliding window
   – Test patch at each location and scale

Note – Template did not change size

Generating hypotheses

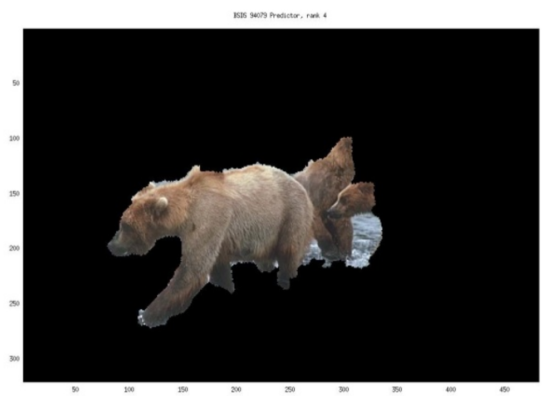
2. Voting from patches/keypoints

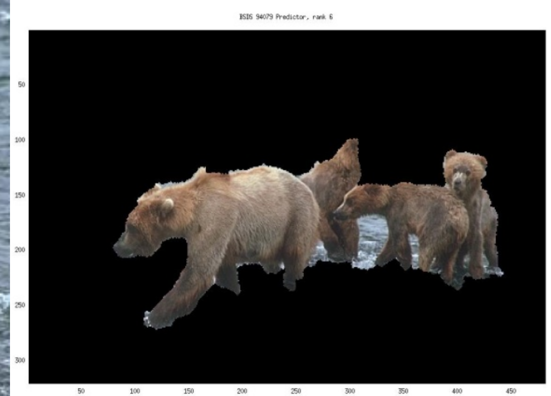


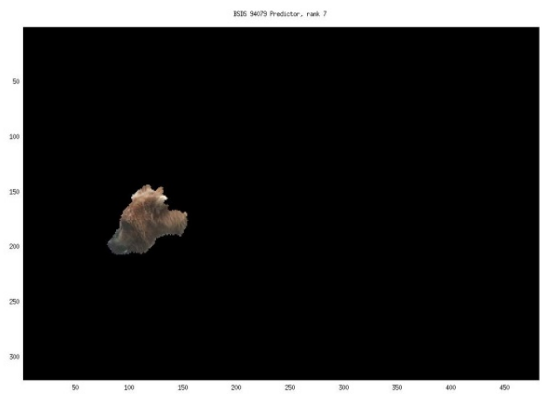
ISM model by Leibe et al.

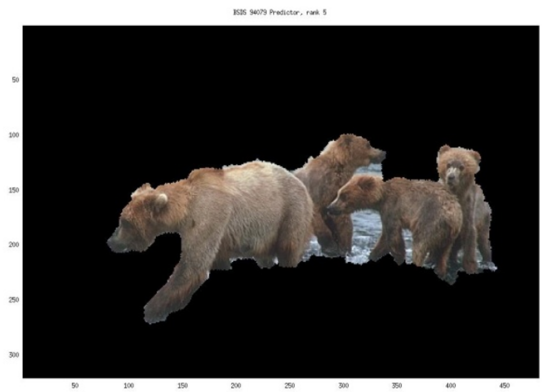
Generating hypotheses

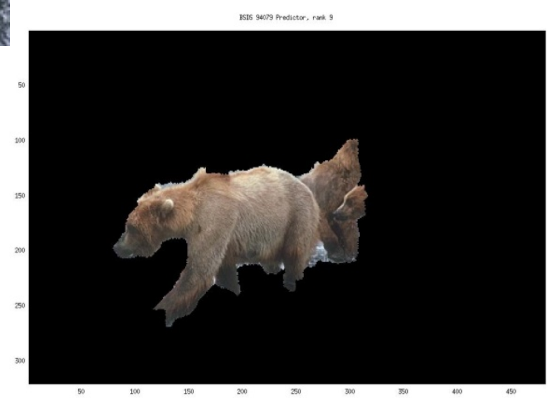
3. Region-based proposal





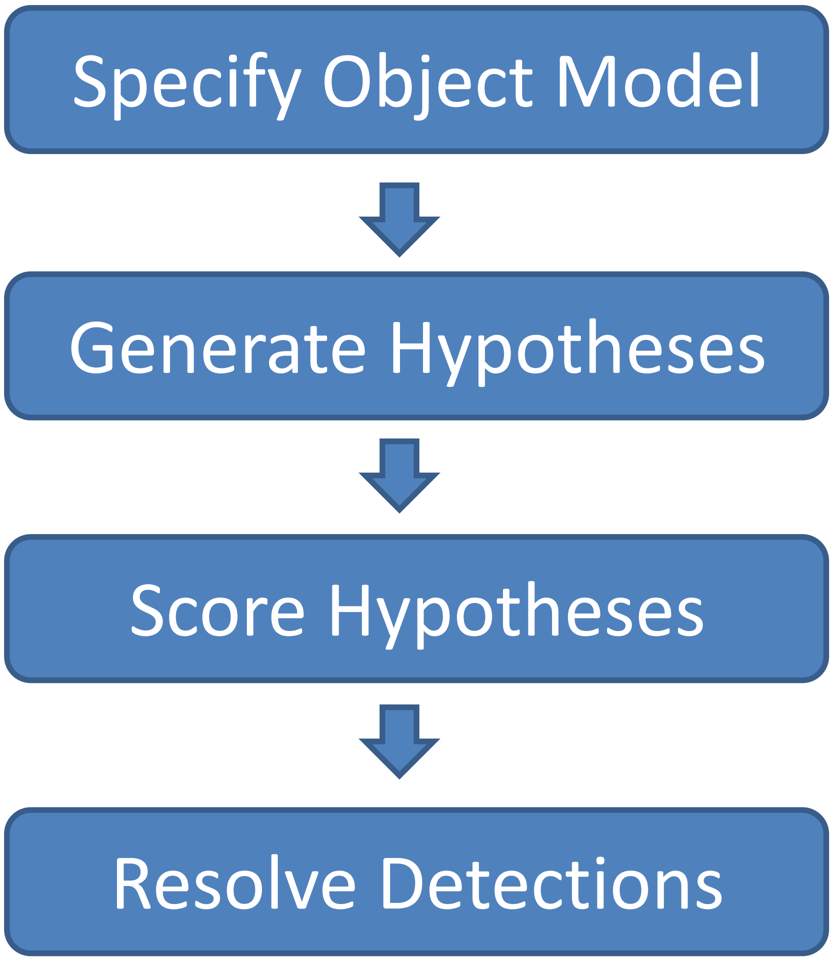






Endres Hoiem 2010

General Process of Object Recognition

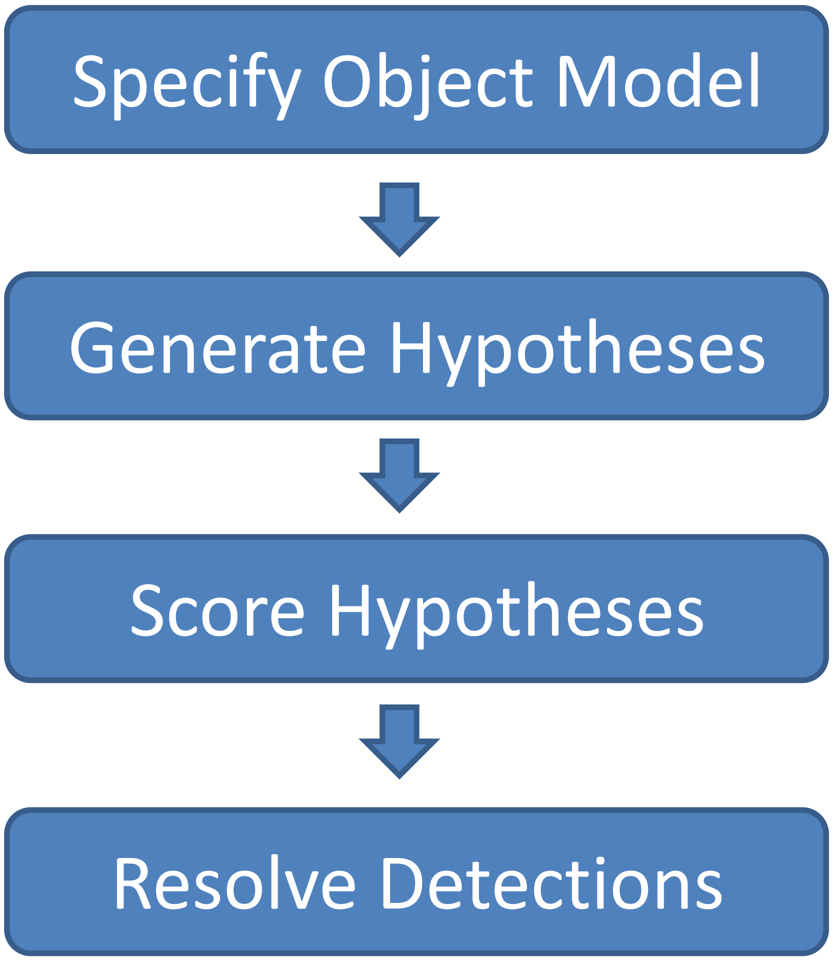


Typical types of features:

- gradient based, e.g., HOG, SIFT
- CNN features

Many classifiers: Adaboost, SVM, NN

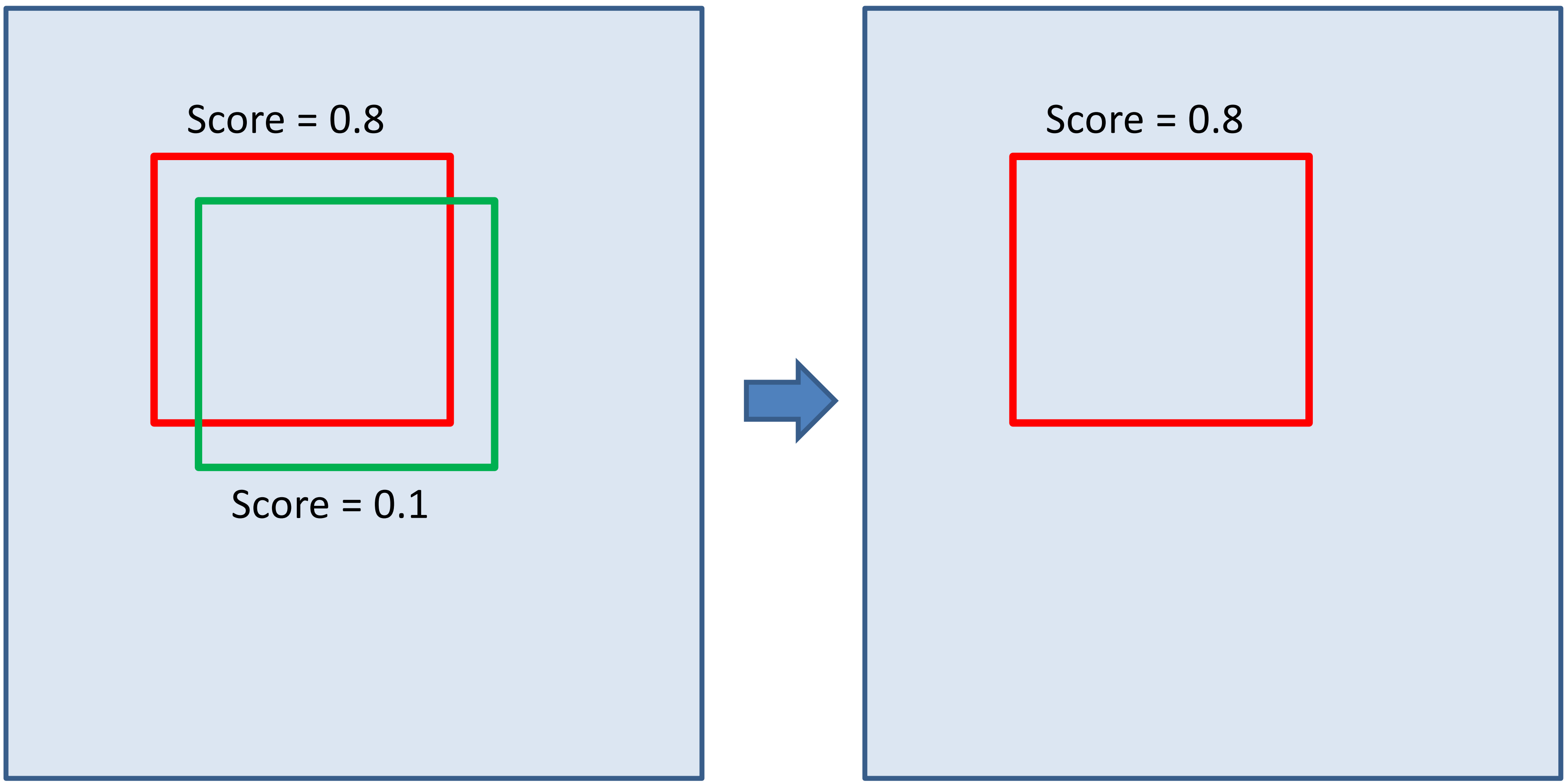
General Process of Object Recognition



Rescore each proposed object based on whole set

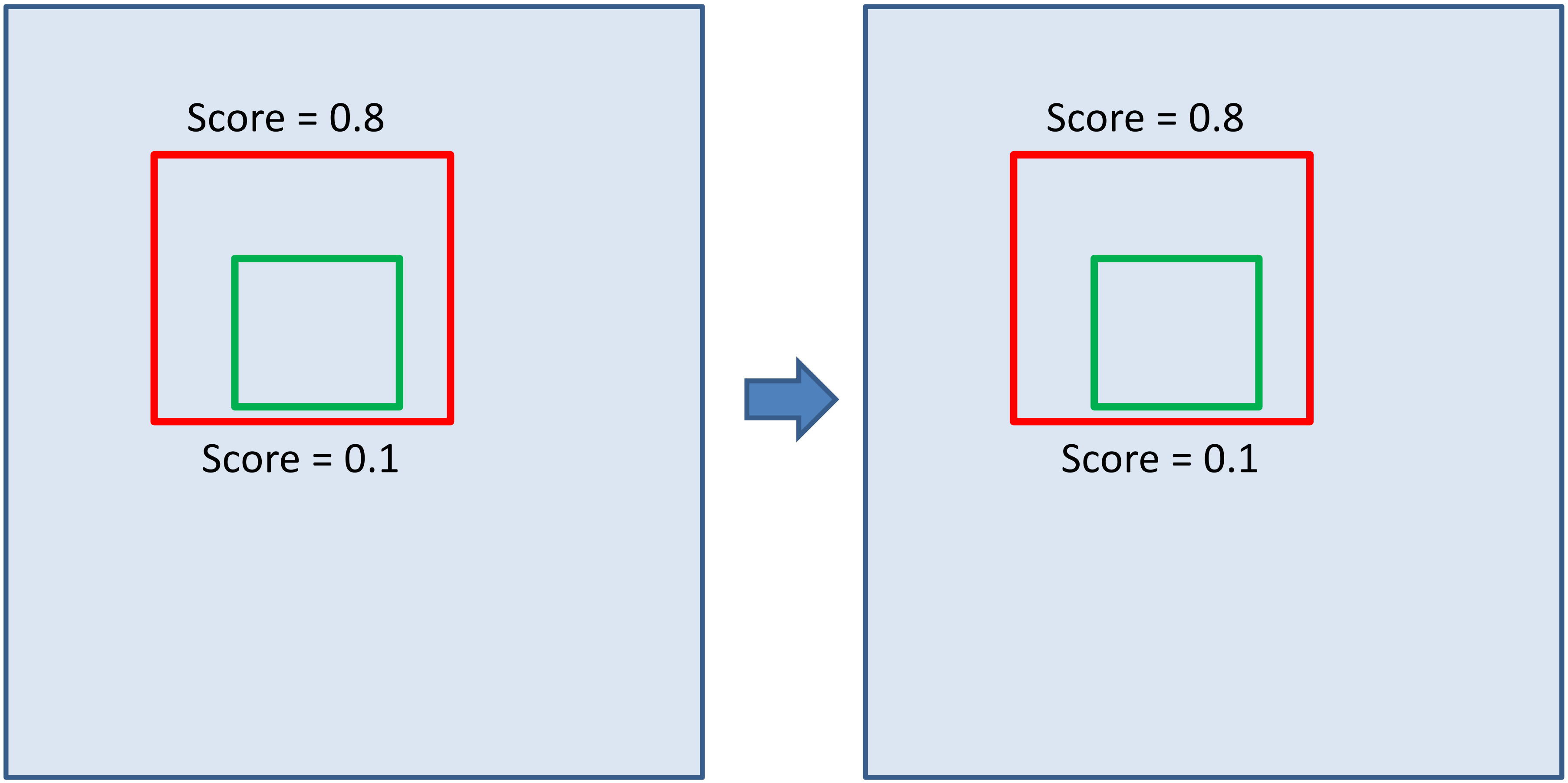
Resolving detection scores

1. Non-max suppression



Resolving detection scores

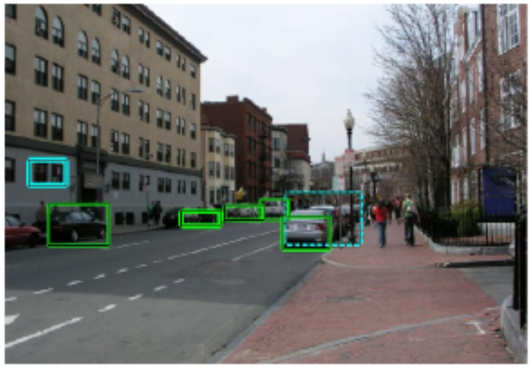
1. Non-max suppression



"Overlap" score is below some threshold

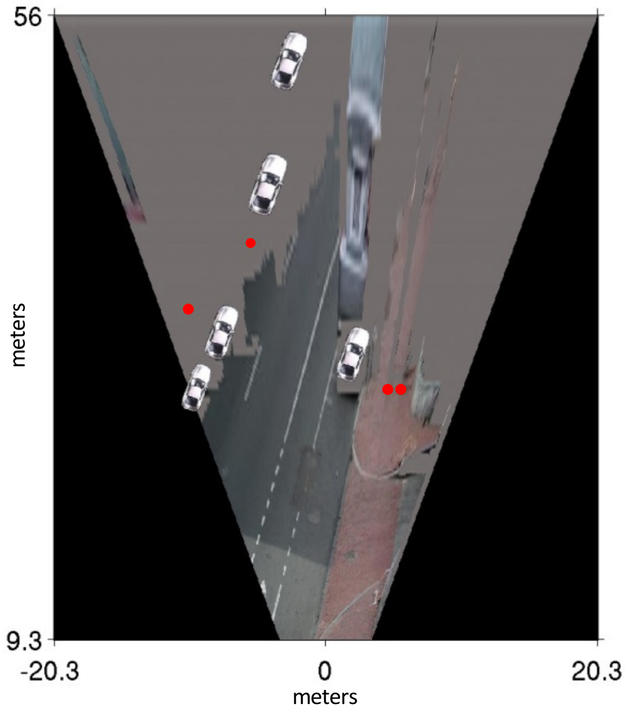
Resolving detection scores

2. Context/reasoning



(g) Car Detections: Local

(h) Ped Detections: Local



Hoiem et al. 2006

Design challenges

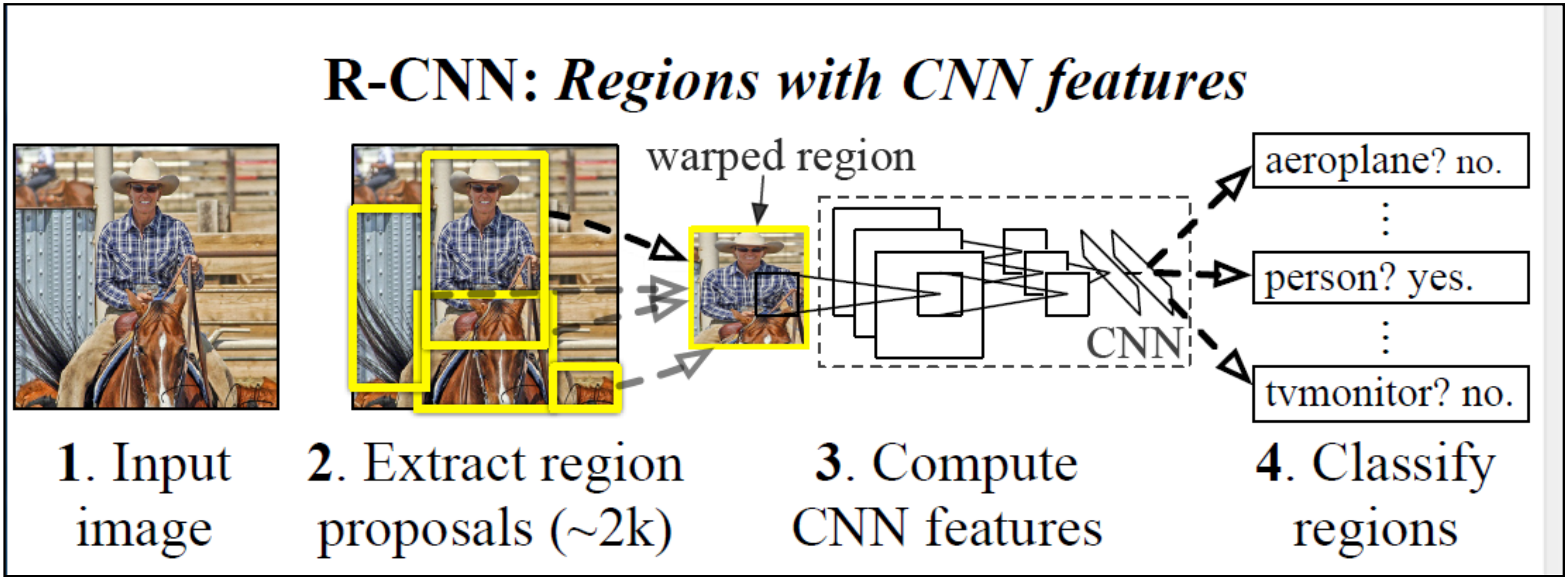
- How to efficiently search for likely objects
  - Even simple models require searching hundreds of thousands of positions and scales
- Feature design and scoring
  - How should appearance be modeled? What features correspond to the object?
- How to deal with different viewpoints?
  - Often train different models for a few different viewpoints
- Implementation details
  - Window size
  - Aspect ratio
  - Translation/scale step size
  - Non-maxima suppression

Example: Dalal-Triggs pedestrian detector

1. Extract fixed-sized (64x128 pixel) window at each position and scale
2. Compute HOG (histogram of gradient) features within each window
3. Score the window with a linear SVM classifier
4. Perform non-maxima suppression to remove overlapping detections with lower scores

Navneet Dalal and Bill Triggs, Histograms of Oriented Gradients for Human Detection, CVPR05

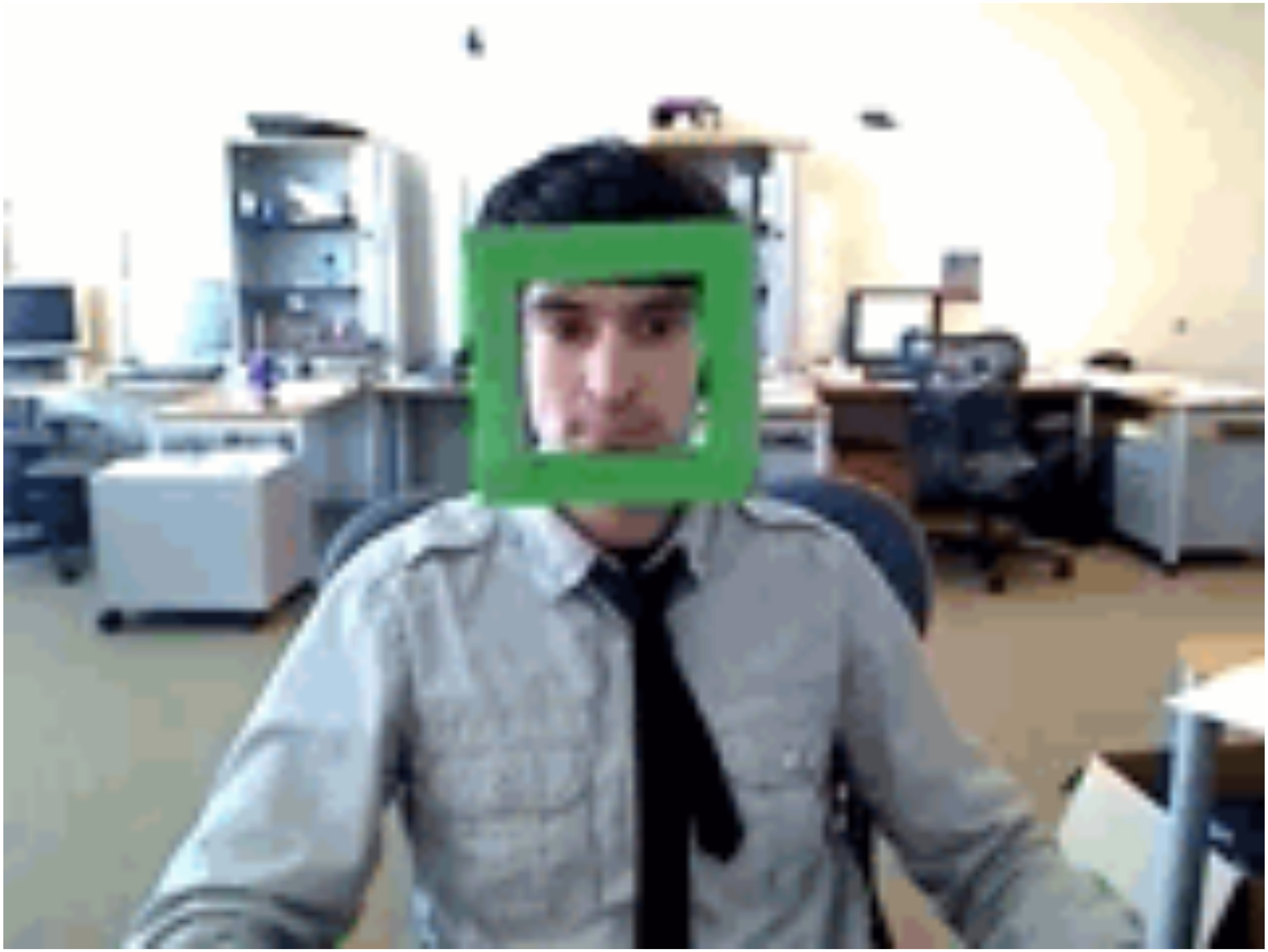
Example: Region-CNN (R-CNN)



What you need to know

- Object detection:
  - Learn an object model
  - Generate hypotheses
  - Score hypotheses
  - Resolve detections
- Evaluation metric:
  - Precision-recall curve
  - Average Precision (AP)
- Sliding window approach
- Non-maximum suppression

State-of-the-art Face detection demo



(Courtesy [Boris Babenko])