What Next?

Progress and Pressing Challenges in Object Recognition

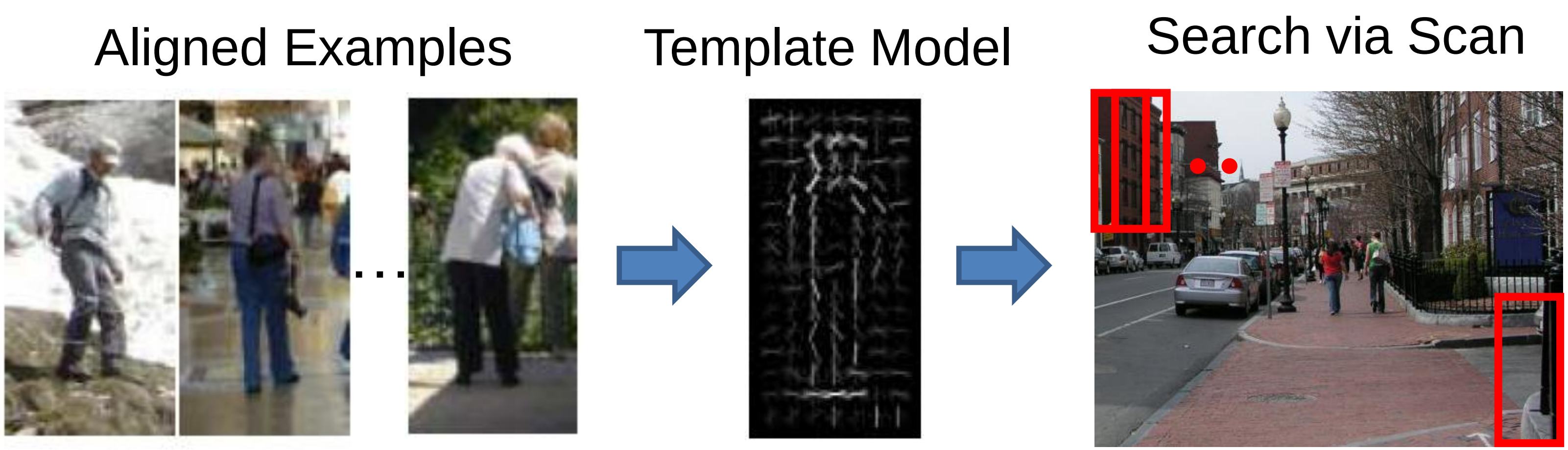
Derek Hoiem

Department of Computer Science

University of Illinois at Urbana-Champaign (UIUC)

PASCAL VOC 2012 Workshop

Object Detection, Pre-VOC



Problem statement: learn template model from aligned examples

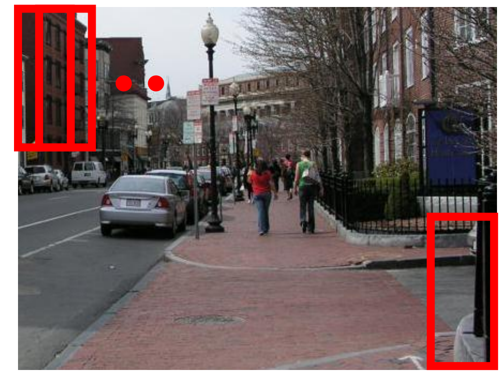
Dalal Triggs 2005

Object Detection, Pre-VOC

Aligned Examples

Template Model

Search via Scan



For multiple viewpoints, repeat for each view

Dalal Triggs 2005

Schneiderman Kanade 2000

VOC: a new crisis

- How to organize and align examples?

VOC: a new crisis

- How to organize and align examples?

Results from VOC 2006

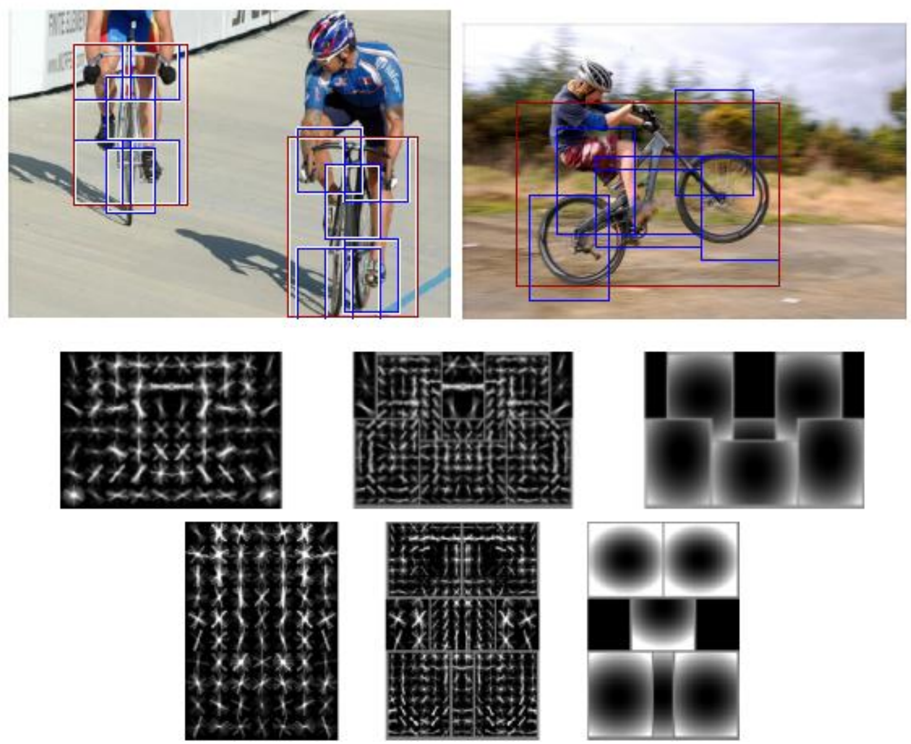
	bicycle	bus	car	cat	cow	dog	horse	motorbike	person	sheep
Cambridge	0.249	0.138	0.254	0.151	0.149	0.118	0.091	0.178	0.030	0.131
ENSMP	–	–	0.398	–	0.159	–	–	–	–	–
INRIA_Douze	0.414	0.117	0.444	–	0.212	–	–	0.390	0.164	0.251
INRIA_Laptev	0.440	–	–	–	0.224	–	0.140	0.318	0.114	–
TKK	0.303	0.169	0.222	0.160	0.252	0.113	0.137	0.265	0.039	0.227
TUD	–	–	–	–	–	–	–	0.153	0.074	–

Notes from submission using Dalal-Triggs HOG method

- “*The results on the classes cat, dog and horse were too bad to be significant*.”
- “comp4 det *test person.txt was trained on our own person dataset*. On the validation dataset *it performed better* than the corresponding comp3 result, presumably *thanks to more appropriate annotations*.”

How to organize and align examples?

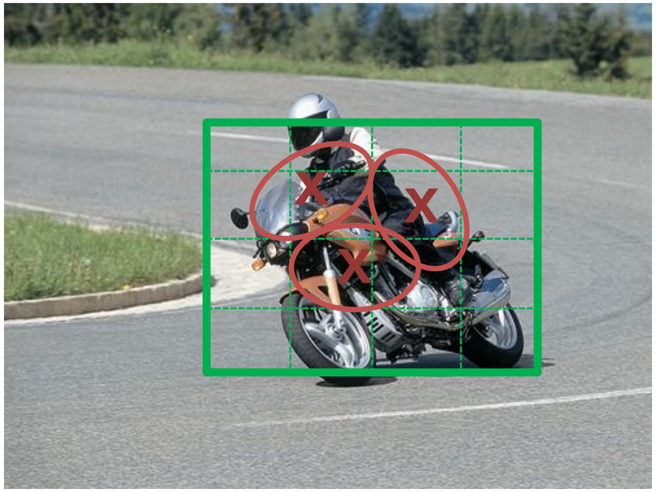
- Deformable Parts Model
  - Automatic clustering via latent components
  - Automatic alignment via latent position of whole object and of part positions



Felzenszwalb Girshick McAllester Ramanan 2010

How to organize and align examples?

- Spatial Pyramid Bag of Words Models
  - Organize by clustering mini-parts (visual words)
  - Align through loose spatial constraints via spatial pyramid



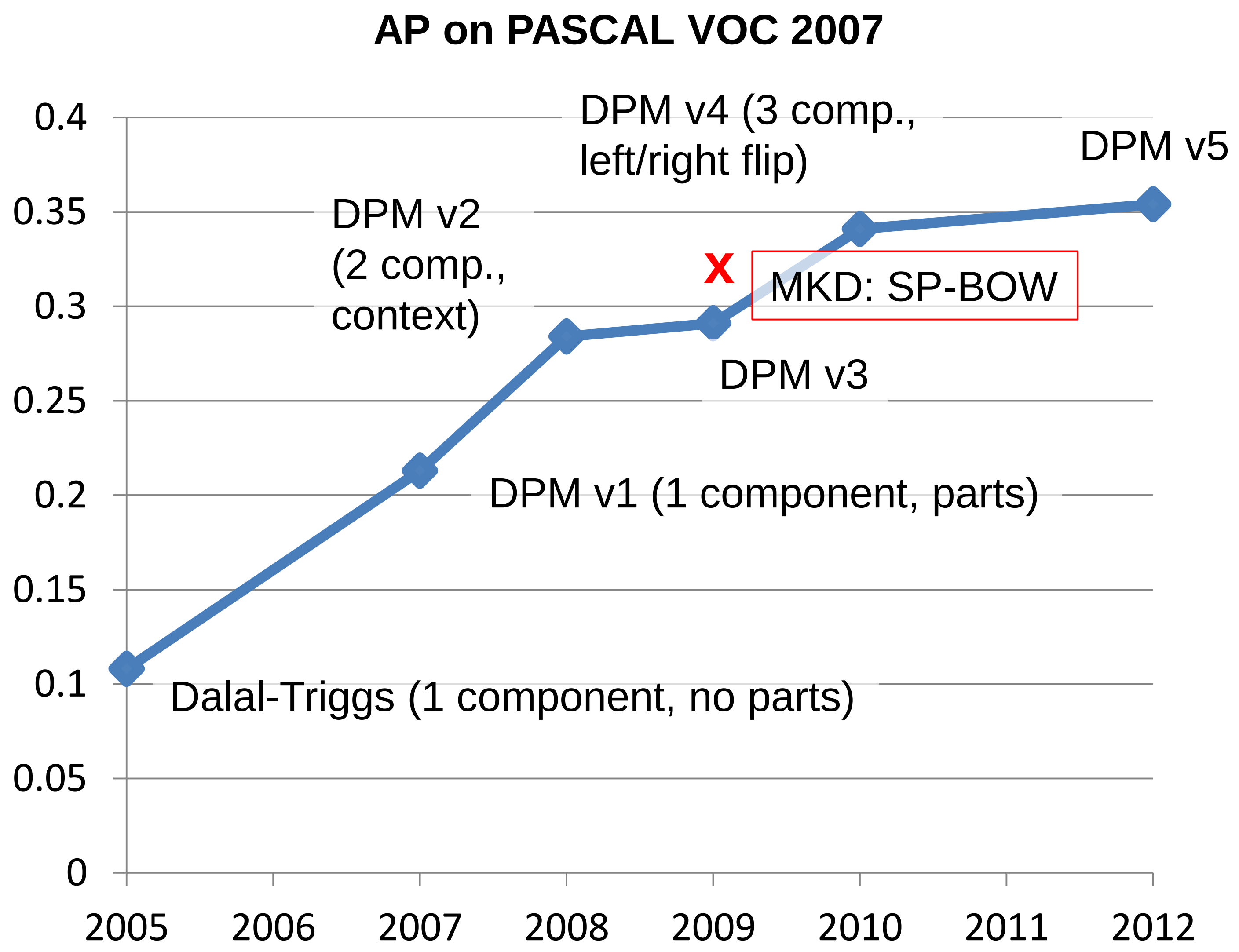
Vedaldi Gulshan Varma Zisserman 2009

How to organize and align examples?

- Poselets Model
  - Alignment in training via hand-annotated poselets
  - Organize via clustering of pose annotations

Bourdev Malik 2009

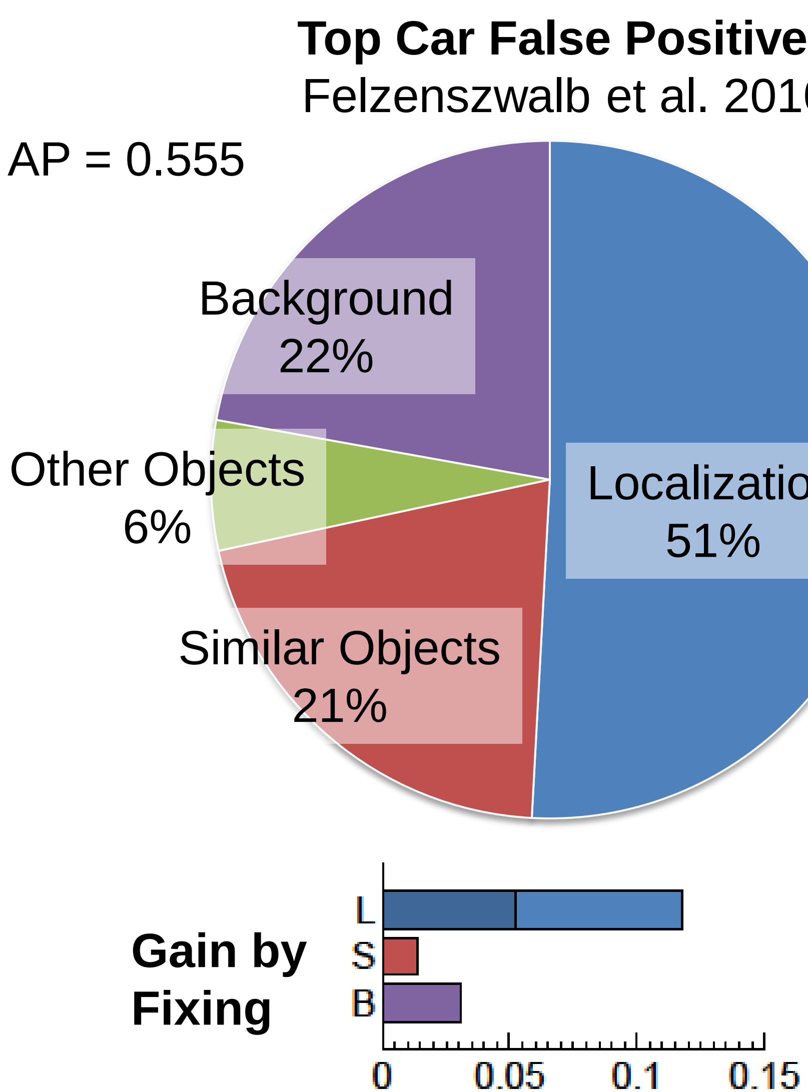
Improvement over time

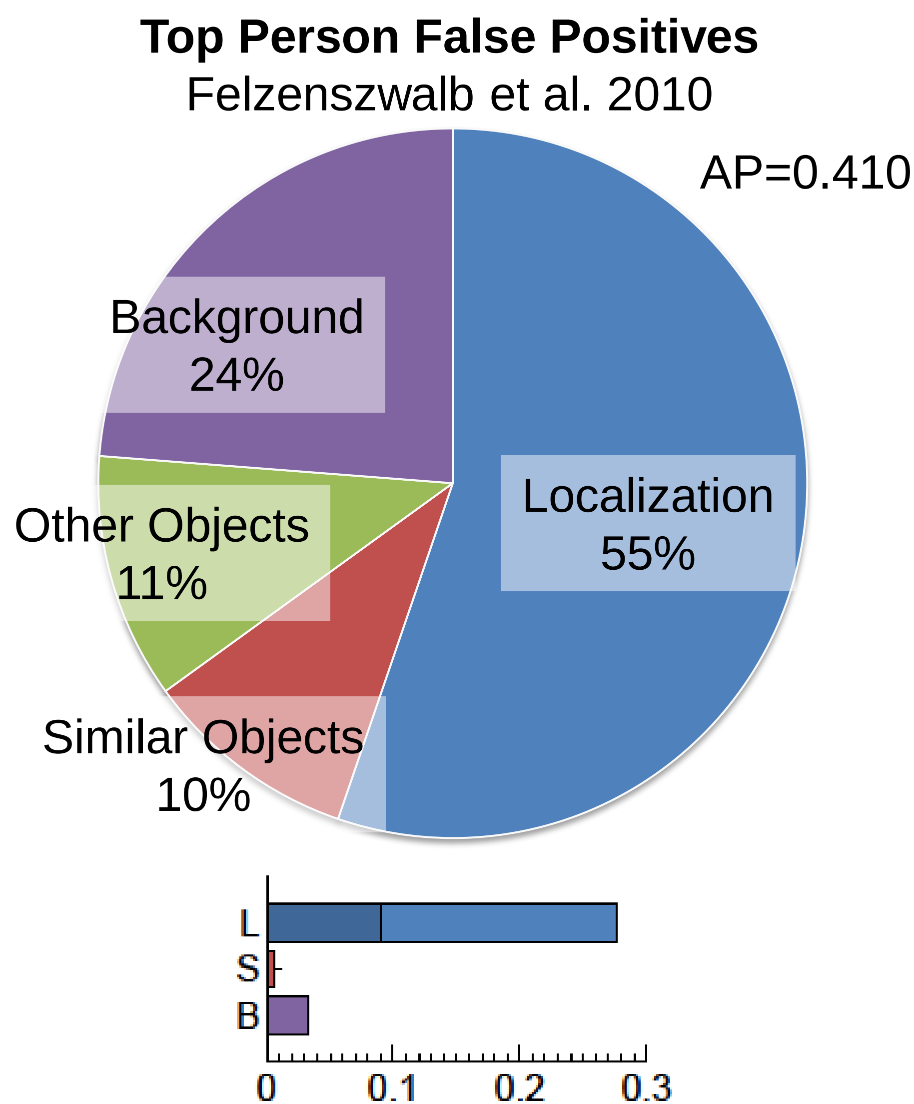


Short-term Challenges within VOC Detection

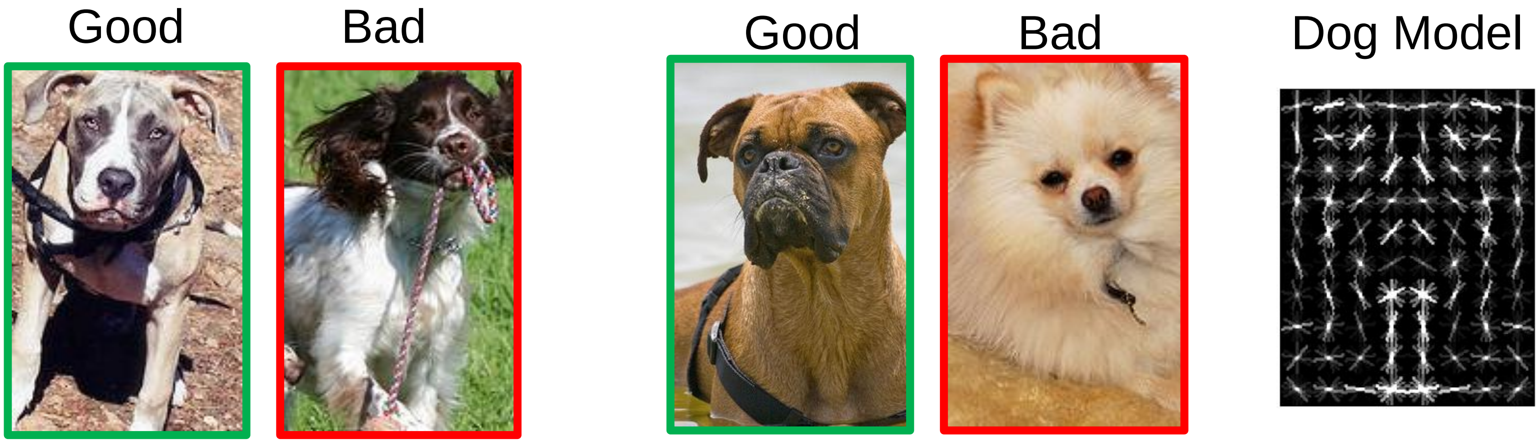
Diagnosing Error in Object Detectors
Derek Hoiem, Yodsawalai Chodpathumwan, and Qieyun Dai
ECCV, 2012.

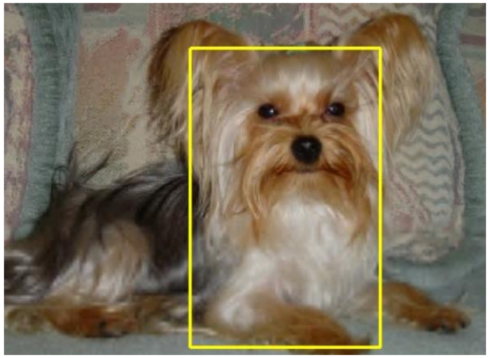
Localizing detected objects





Localizing detected objects

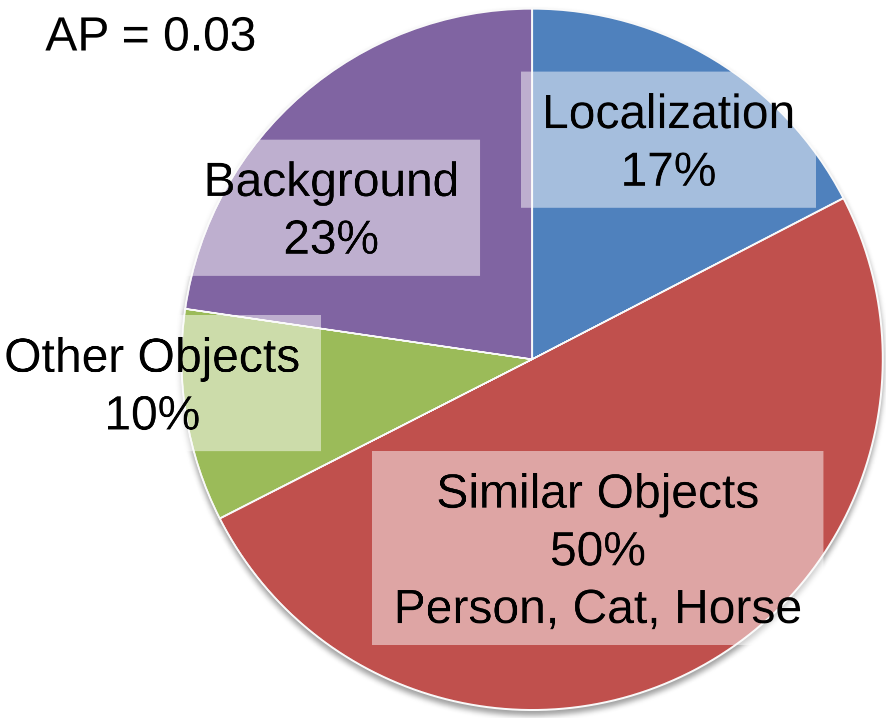




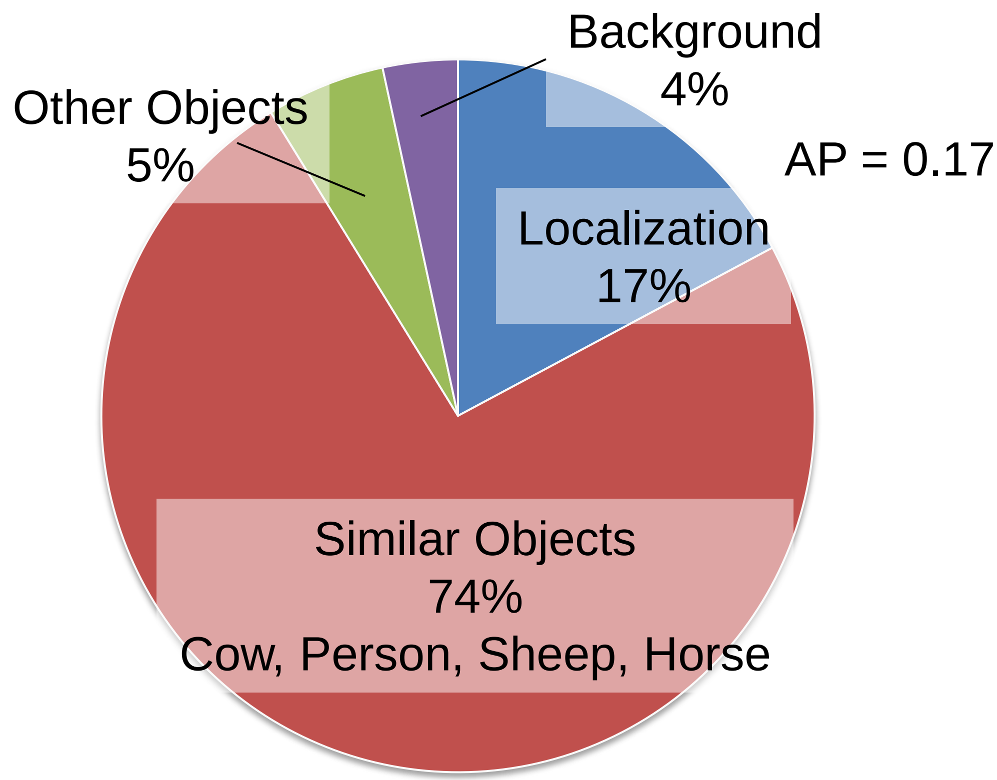
Need good category-sensitive segmentation methods

Differentiating similar categories

Top Dog False Positives
Felzenszwalb et al. 2010



Top Dog False Positives
Vedaldi et al. 2009



Differentiating similar categories



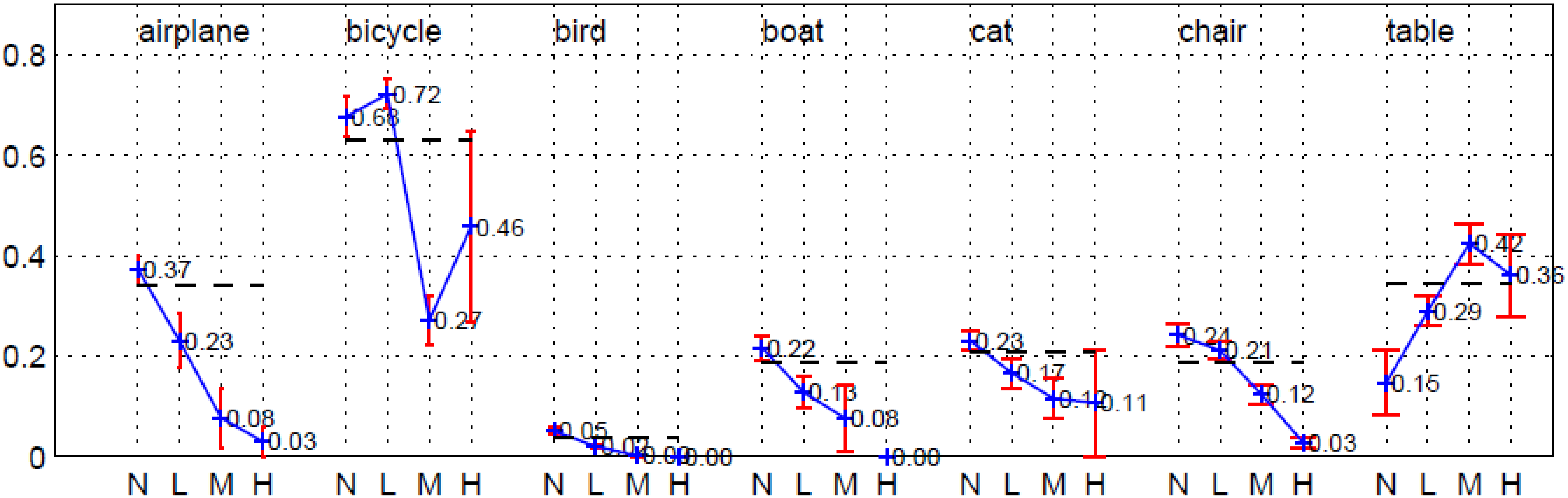


Compare details, rather than holistic appearance

Dog Model

Detecting occluded objects

Felzenszwalb et al. (v4) Sensitivity to Occlusion



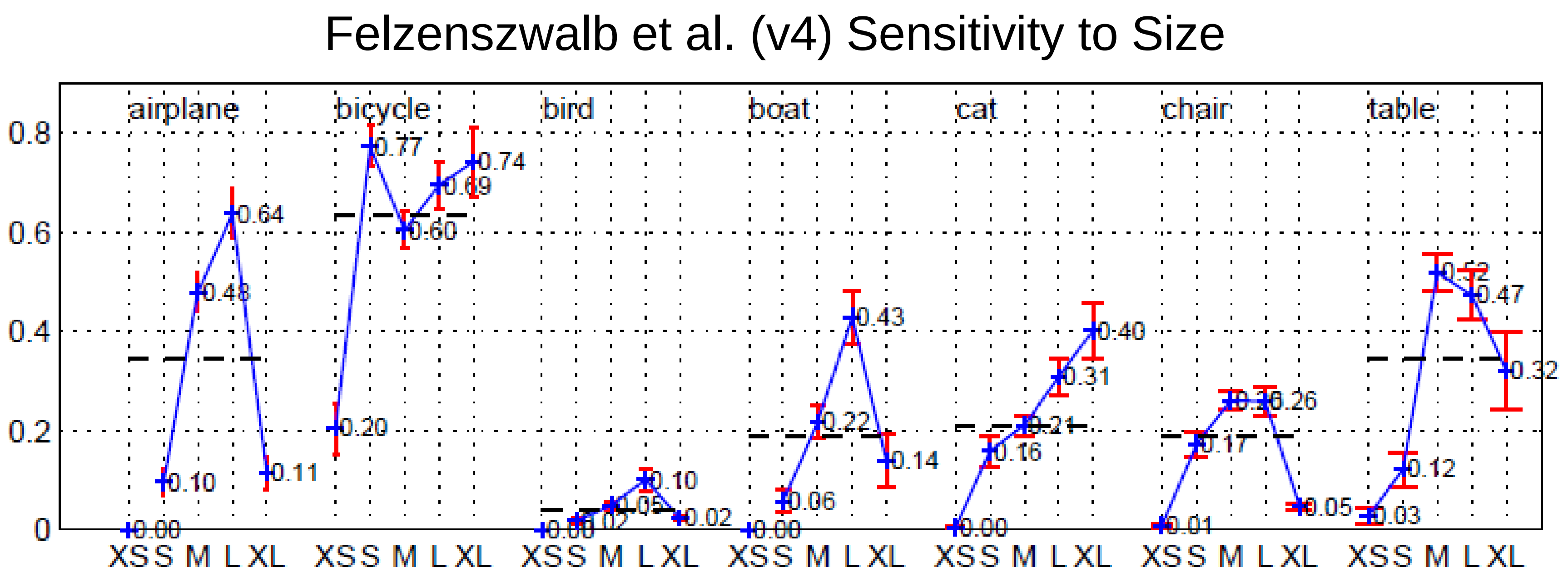
Example efforts:

Wu Nevatia ICCV 2005

Wang Han Yan ICCV 2009

Yang et al. CVPR 2010

Detecting small or very near objects



- Benefit from high resolution of large objects
- Robustness to perspective effects
- Context for better detection of small objects

Example efforts:

Park Ramanan Fowlkes ECCV 2010

Proposal: standardized sub-challenges with leader boards

Detection tasks

- AP ignoring specific types of error
  - Localization error (e.g., place a ‘+’ on each object)
  - Confusion with similar objects
- Targeted subset challenges
  - Performance on occluded objects
  - Performance on smallest 25% of objects

Other tasks

- Category-based segmentation
  - Average overlap when provided with bounding box (perhaps with random perturbations)
- Categorization given bounding box

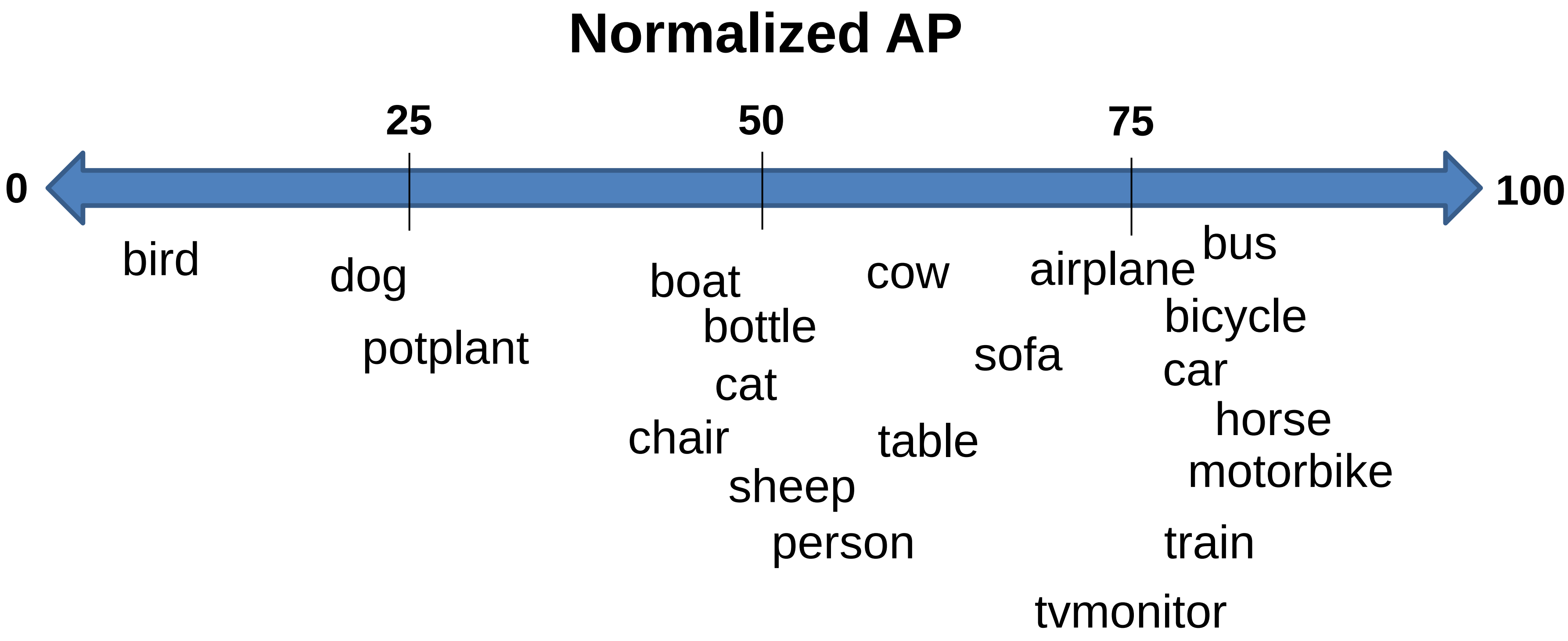
Long Term Challenge:
Beyond Recognition as Visual Pattern Matching

An unsolved crisis: heavy-tailed distribution of objects

- Many modes of object appearance
  - Pose, view, shape, distance, texture/color
- Some modes are common (prototypical views), many others are not
- Much progress due to division of categories into visual subcategories
  - Often high performance for common modes, poor for others
- Learning less common modes is important for dealing with variation within and across categories

Performance on “Easy” Examples

- Ignore truncated, smallest 30% of objects
- allow moderate localization errors (ov>=0.2)



VOC 2007, FGMR (v4)

Bus (avg = 83)

Poor (0-10)

OK (10-50)

Good (50-90)

Excellent (90-100)

Bird (avg = 11)

Poor (0-10)






OK (10-50)




Good (50-90)





Excellent (90-100)

Learning about objects from vision

Not what does it look like, but what is it?

- What is the 3D shape?
- How big is it?
- What are the functional parts?
- What are distinctive markings?
- What can it do?
- In what kind of settings is it likely to appear?

Training: Aye-Aye

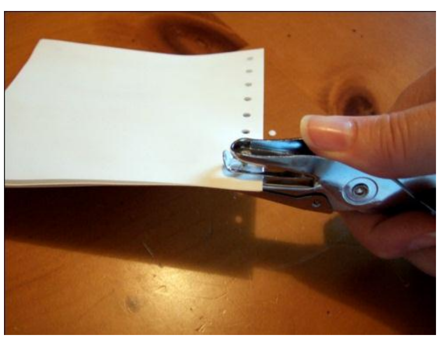
Can you learn to recognize this category from two examples?

Testing: detect aye-aye

Training: hole-puncher

Can you learn to recognize this category from two examples?

Testing: detect hole-punchers



Learning to recognize, beyond gradients



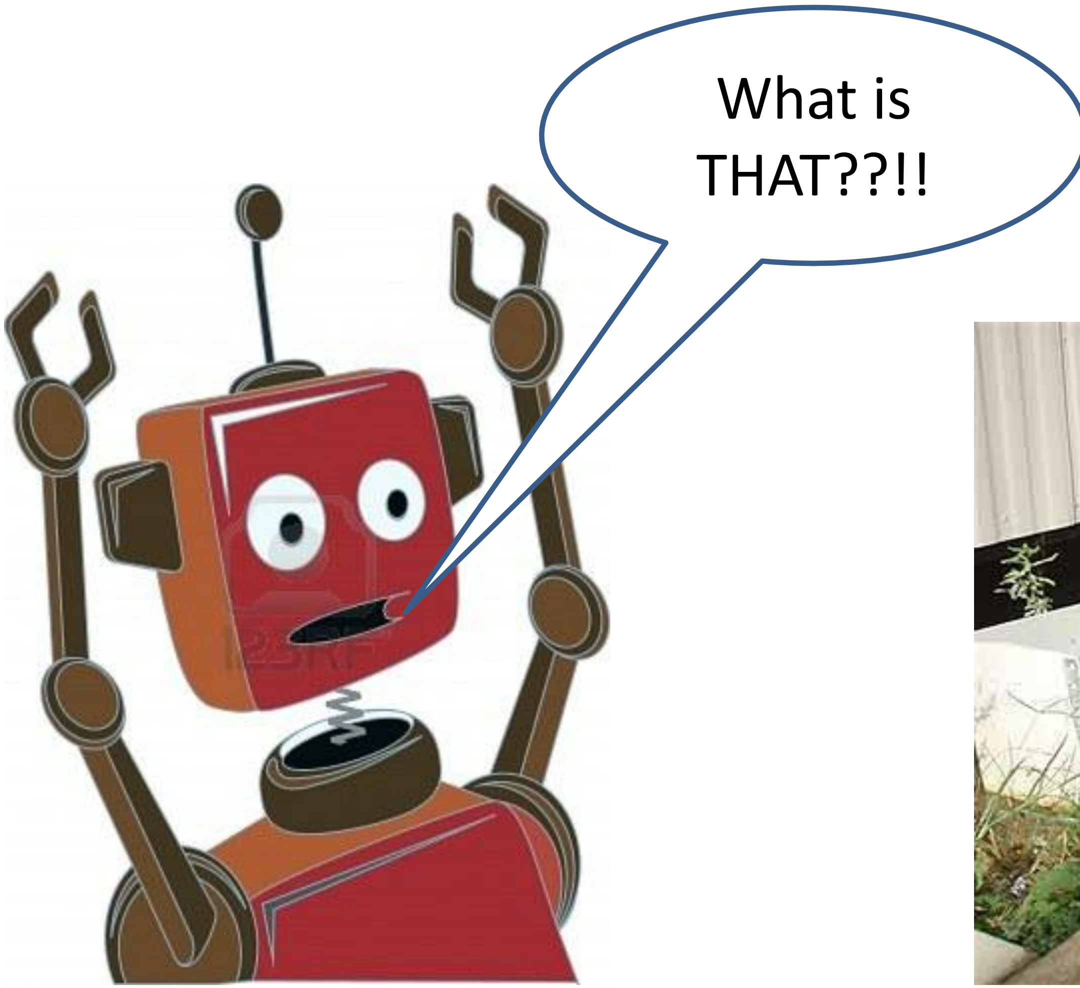

- More explicit shape representations
- Applying domain knowledge (based on similar categories)
  - Which parts are important for function?
  - Which parts will have stable appearance?
  - Which features are distinctive?
  - What kinds of deformations are likely/possible?

Long-term Challenges in Object Representation

Recognition as search

Figure credit: http://www.123rf.com/photo_10431638_screaming-businessman-and-desktop-computer.html/

Recognition as interpretation



http://www.123rf.com/photo_9362895_cartoon-robot-surprised-expression-raised-claw-arms.html

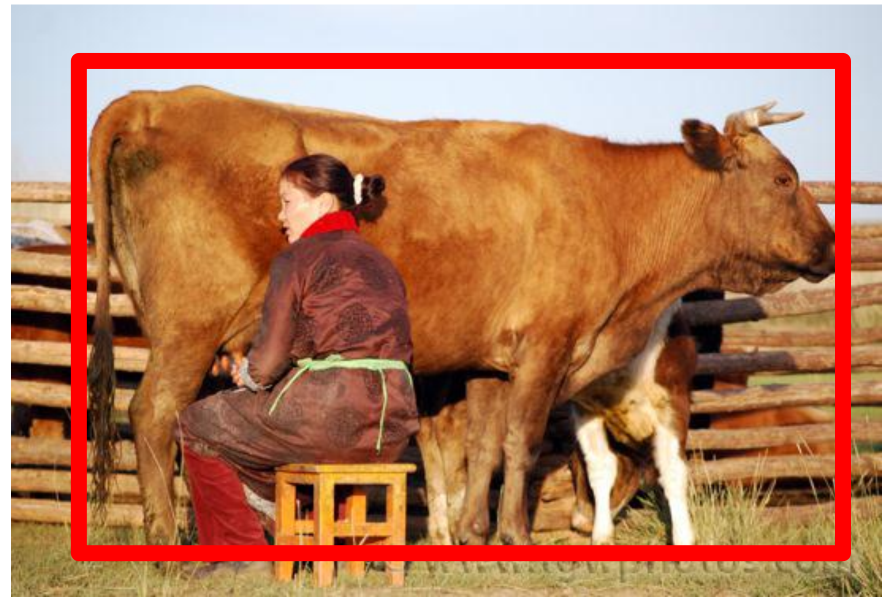
How to localize without categorization?

Efforts:

Carreira Sminchisescu 2010

Endres Hoiem 2010

Task-dependent representations



Big animal ahead, moving left

Cow

Which objects are relevant, and how are they relevant?

Physical context-dependent representations

How to infer physical relations (contact, engagement, etc.)?

How to interpret an object's role in the scene?

Interesting upcoming challenge: Visual Entailment

Which statements can be inferred from the image?

Correct entailments:
1) Exactly one bird is visible.
2) There is a white bird.
3) The bird is touching the shopping cart.
4) The bird is on a wooden surface.

Incorrect entailments:
1) The image contains a blue bird.
2) The scene is a grocery store.
3) The scene contains a cat.
4) The cat is eating the bird.

Current effort by Julia Hockenmaier and Tamara Berg

Final comments

- Detection has many subproblems: may accelerate research to create specialized challenges
- Important Major Problems
  - Object segmentation or boundary labeling: important for inferring shape
  - Representing 3D shape: important for viewpoint robustness, function/affordance analysis
  - Representing function: more robust recognition, broader recognition applications
- Need for datasets to evaluate shape, function, task-centric recognition

Thank you