

The PASCAL Visual Object Classes (VOC) Dataset and Challenge

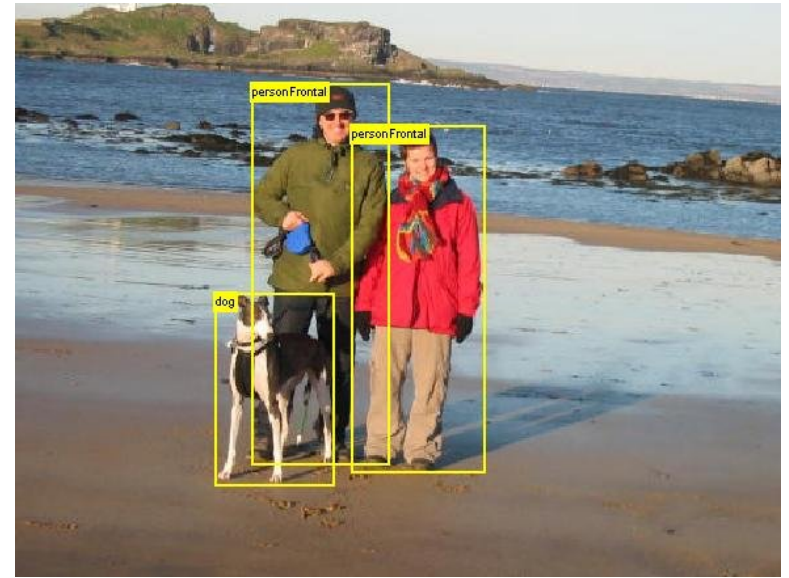
Andrew Zisserman

Visual Geometry Group
University of Oxford



The PASCAL VOC Challenge

- Challenge in visual object recognition funded by PASCAL network of excellence
- Publicly available dataset of annotated images



- Main competitions in **classification** (is there an X in this image), **detection** (where are the X's), and **segmentation** (which pixels belong to X)
- Competition has run each year since 2006.
- Organized by Mark Everingham, Luc Van Gool, Chris Williams, John Winn and Andrew Zisserman

Annotation

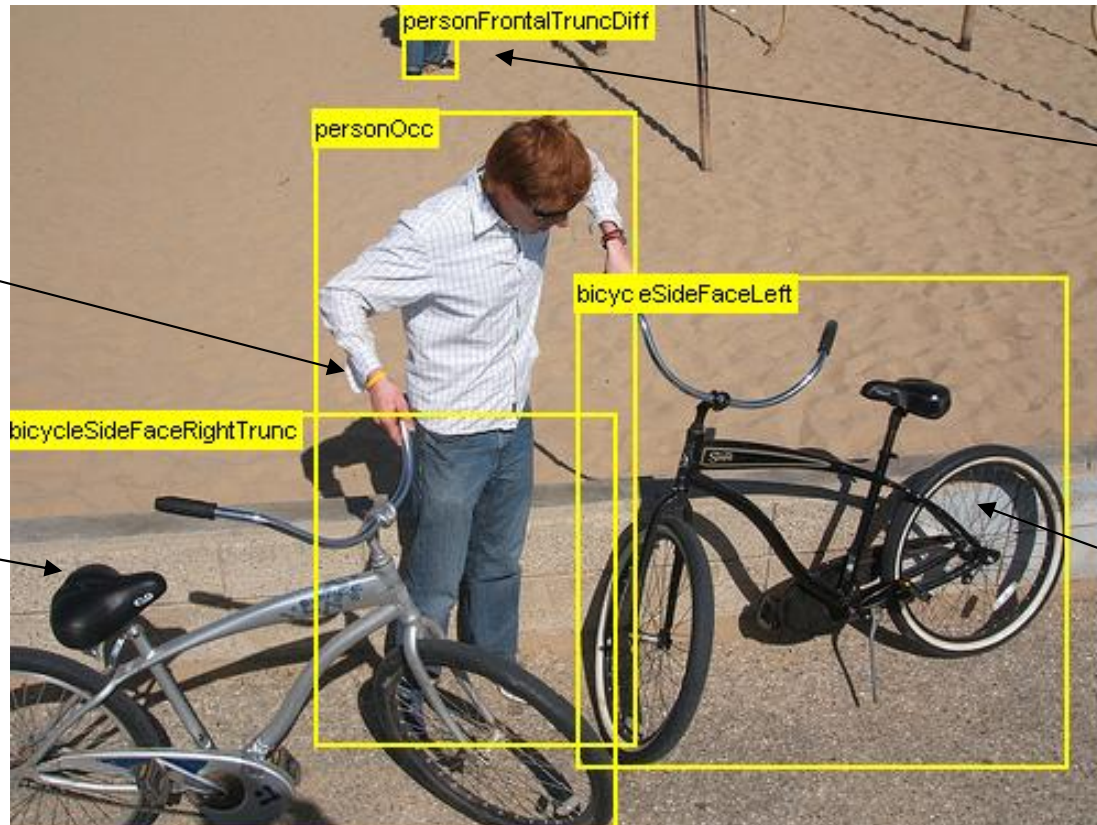
- Complete annotation of all objects
- Annotated in one session with written guidelines

Occluded

Object is significantly occluded within BB

Truncated

Object extends beyond BB



Difficult

Not scored in evaluation

Pose

Facing left

Examples

Aeroplane



Bicycle



Bird



Boat



Bottle



Bus



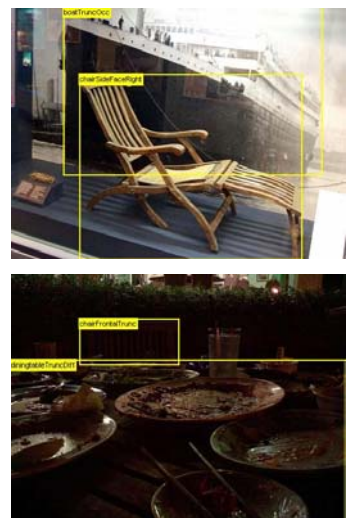
Car



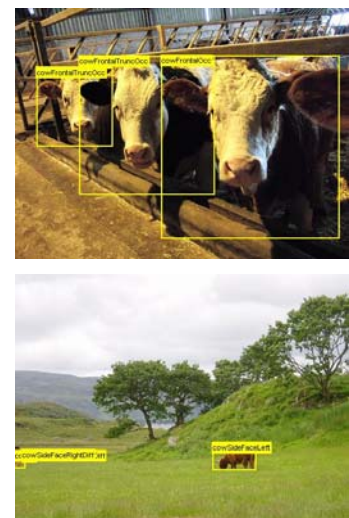
Cat



Chair

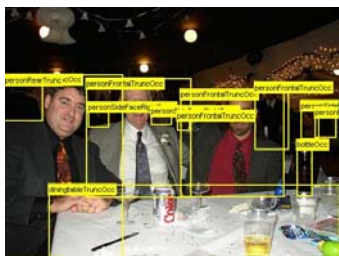


Cow

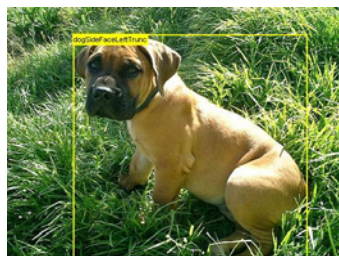


Examples

Dining Table



Dog



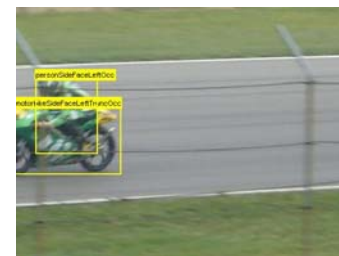
Horse



Motorbike



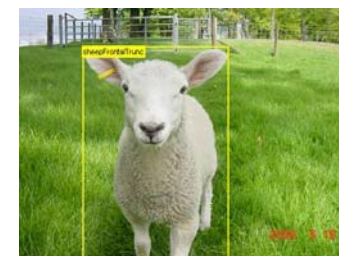
Person



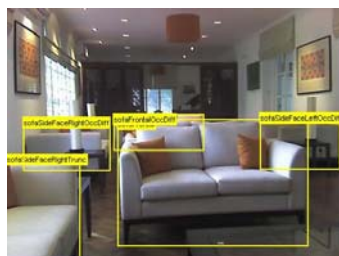
Potted Plant



Sheep



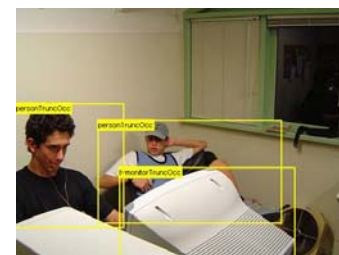
Sofa



Train



TV/Monitor



Dataset Statistics – VOC 2010

	Training	Testing
Images	10,103	9,637
Objects	23,374	22,992

- Minimum ~500 training objects per category
 - ~1700 cars, 1500 dogs, 7000 people
- Approximately equal distribution across training and test sets
- Data augmented each year

Design choices: Things we got right ...

1. Standard method of assessment

- Train/validation/test splits given
- Standard evaluation protocol – AP per class
- Software supplied
 - Includes baseline classifier/detector/segmenter
 - Runs from training to generating PR curve and AP on validation or test data out of the box
- Has meant that results on VOC can be consistently compared in publications
- “Best Practice” webpage on how the training/validation/test data should be used for the challenge

Design choices: Things we got right ...

2. Evaluation of test data

Three possibilities

1. Release test data and annotation (most liberal) and participants can assess performance
 - Cons: open to abuse
2. Release test data, but test annotation withheld - participants submit results and organizers assess performance (use an evaluation server)
3. No release of test data - participants have to submit software and organizers run this and assess performance

Design choices: Things we got right ...

3. Augmentation of dataset each year

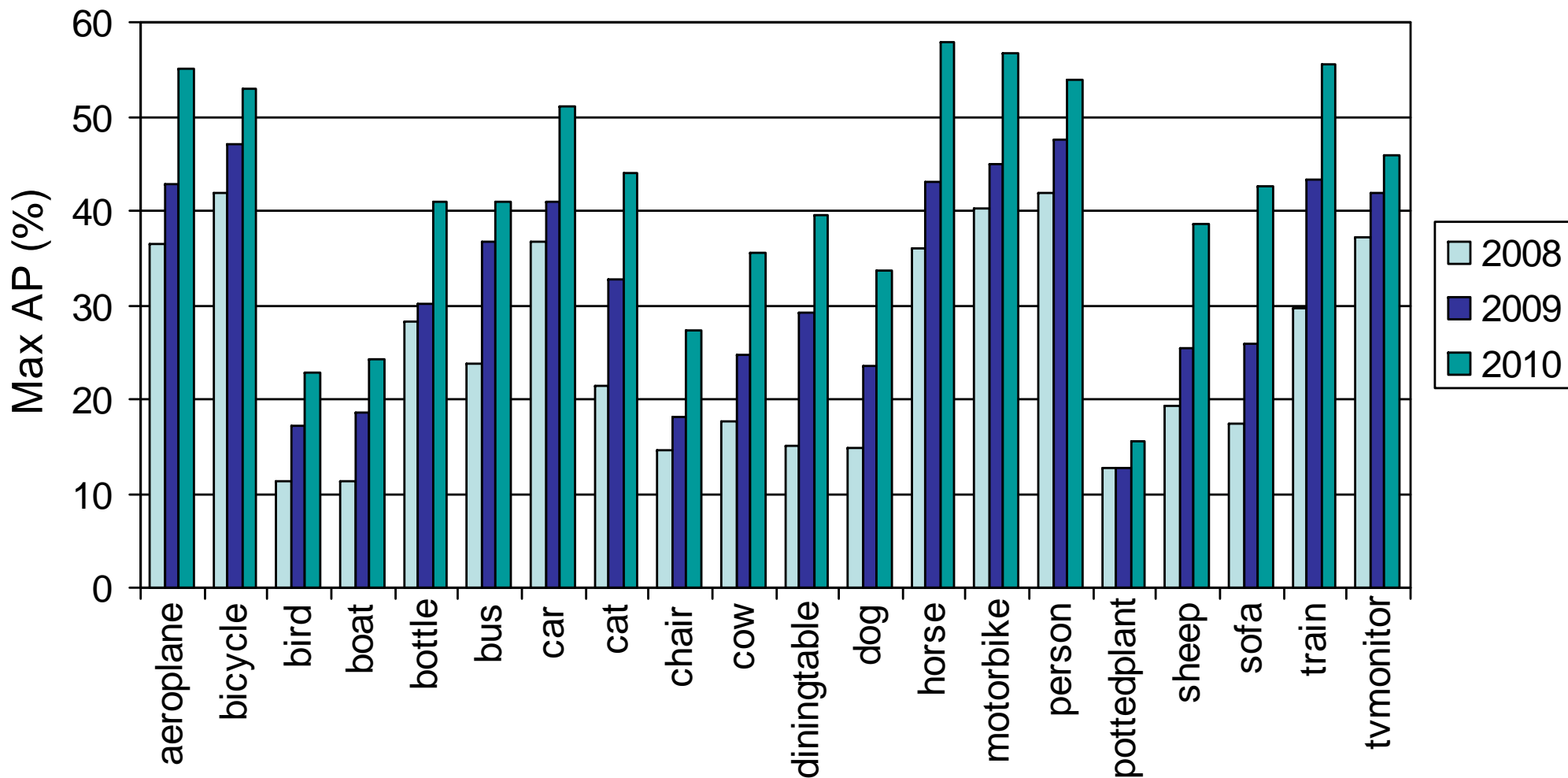
- VOC2010 around 40% increase in size over VOC2009

	Training		Testing	
Images	10,103	(7,054)	9,637	(6,650)
Objects	23,374	(17,218)	22,992	(16,829)

VOC2009 counts shown in brackets

- Has prevented over fitting on data
- 2008/9 datasets retained as subset of 2010
 - Assignments to training/test sets maintained
 - So can measure progress from 2008 to 2010

Detection Challenge: Progress 2008-2010



- Results on 2008 data improve for best 2009 and 2010 methods for all categories, by over 100% for some categories
 - Caveat: Better methods or more training data?

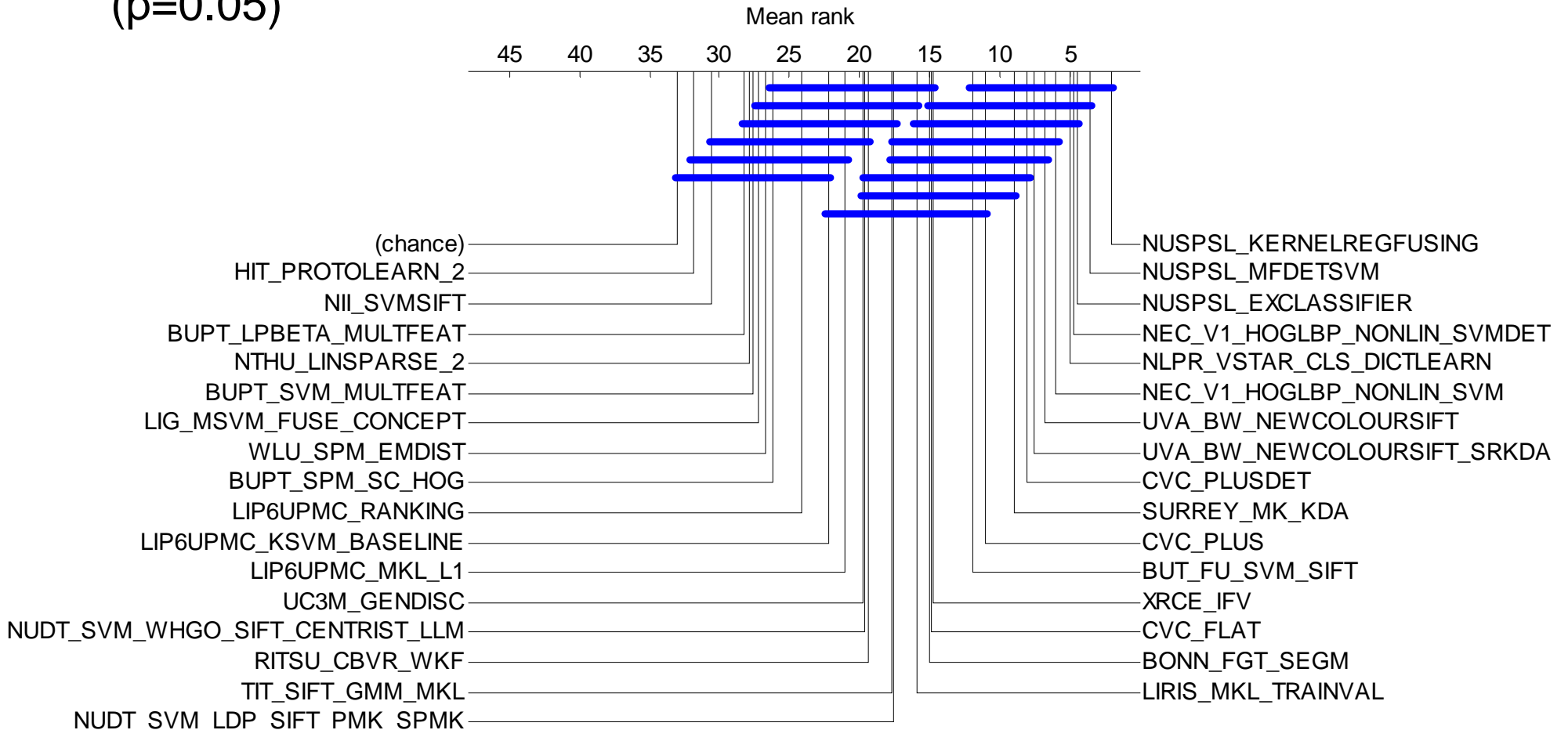
Design choices: Things we got right ...

4. Equivalence bands

- Uses Friedman/Nemenyi ANOVA significance test
- Makes explicit that many methods are just slight perturbations of each other
- Need more research on this area for obtaining equivalence classes for ranking (rather than classification)
- Plan to include banner/header results on evaluation server (to aid comparisons for reviewing etc)

Statistical Significance

- Friedman/Nemenyi analysis
 - Compare differences in **mean rank** of methods over classes using non-parametric version of ANOVA
 - Mean rank must differ by at least 5.4 to be considered significant ($p=0.05$)



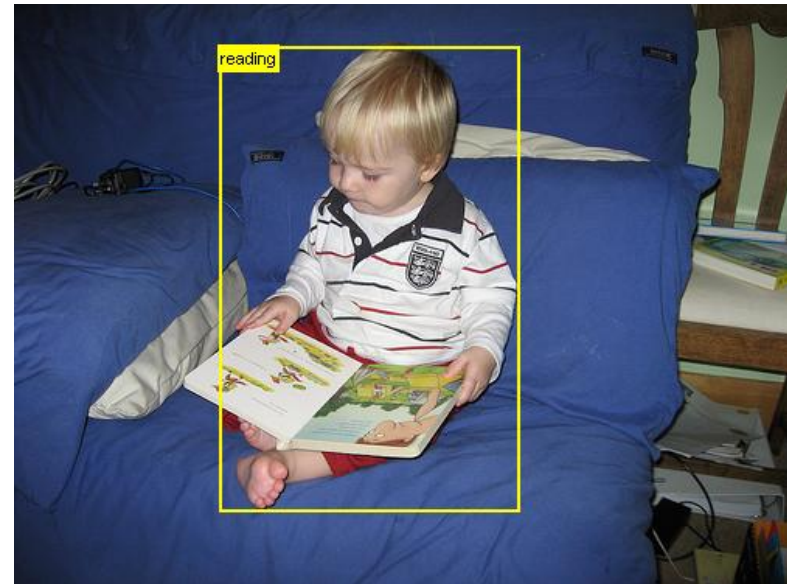
The future ...

Action Classification Taster Challenge – 2010 on

- Given the bounding box of a person, predict whether they are performing a given action



Playing Instrument?



Reading?

- Developed with Ivan Laptev
- Strong participation from the start (11 methods, 8 groups)

Ten Action Classes

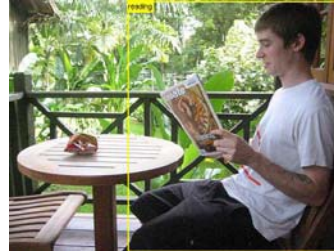
Phoning



Playing Instrument



Reading



Riding Bike



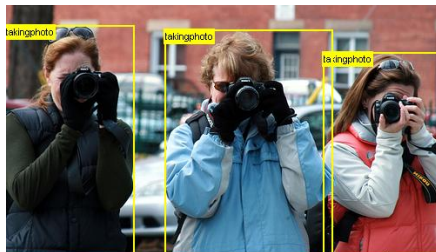
Riding Horse



Running



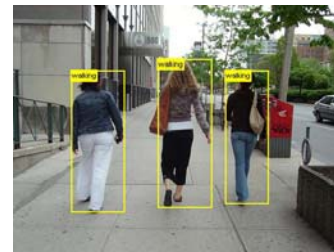
Taking Photo



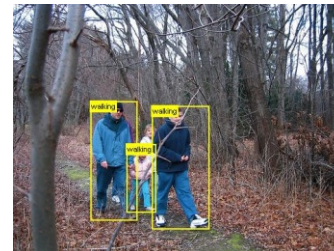
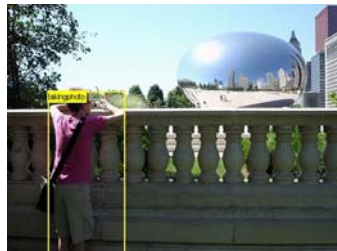
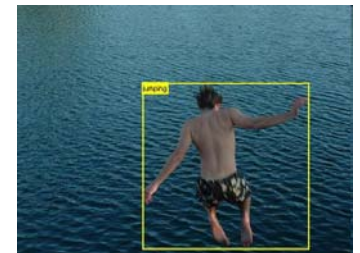
Using Computer



Walking



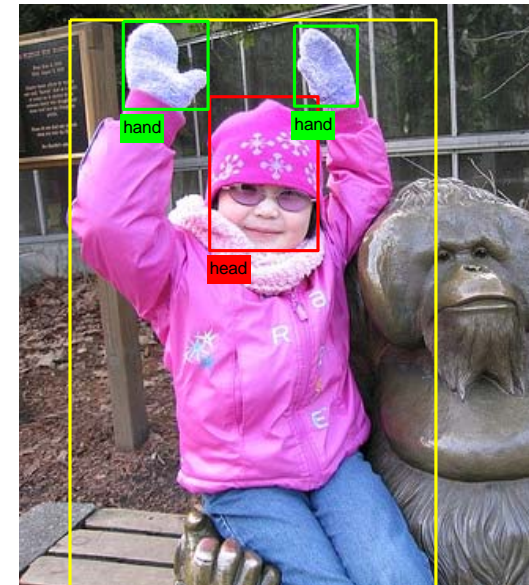
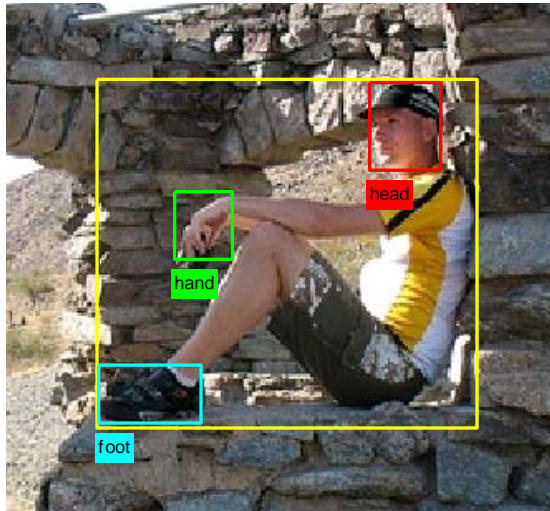
Jumping



- 2011: added jumping + `other' class

Person Layout Taster Challenge – 2009 on

- Given the bounding box of a person, predict the positions of head, hands and feet.
- aim: to encourage research on more detailed image interpretation



- 2009: no participation
- 2010: relaxed success criteria, 2 participants
- Assessment method still not satisfactory?

Futures

2012

- Last official year of PASCAL2
- May introduce a new taster competition, e.g.
 - Materials
 - Actions in video
- Open to suggestions