

# The PASCAL Visual Object Classes Challenge 2011 (VOC2011)

## Part 4 – Person Layout Taster Challenge

Mark Everingham

Luc Van Gool

Chris Williams

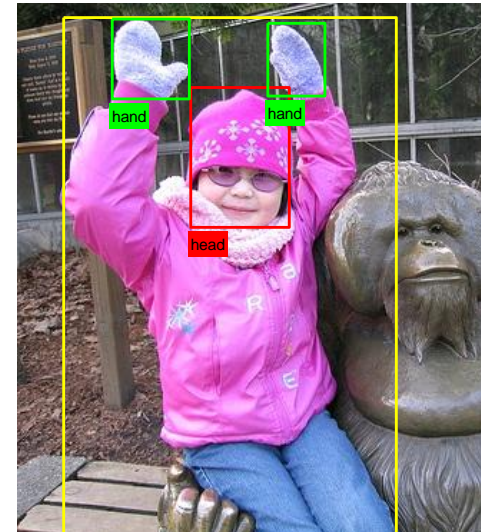
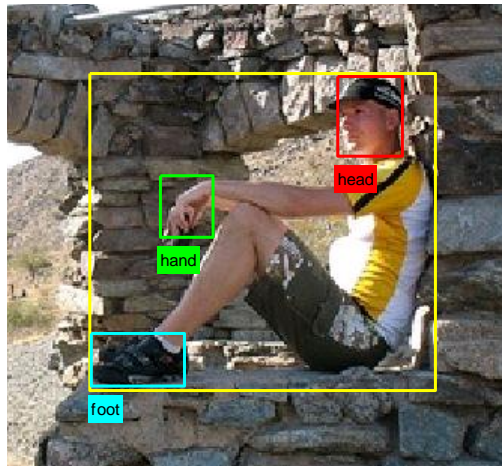
John Winn

Andrew Zisserman



# Person Layout Taster Challenge

- Given the bounding box of a person, predict the positions of head, hands and feet.



- Encourage research on more detailed image interpretation

# Dataset Statistics

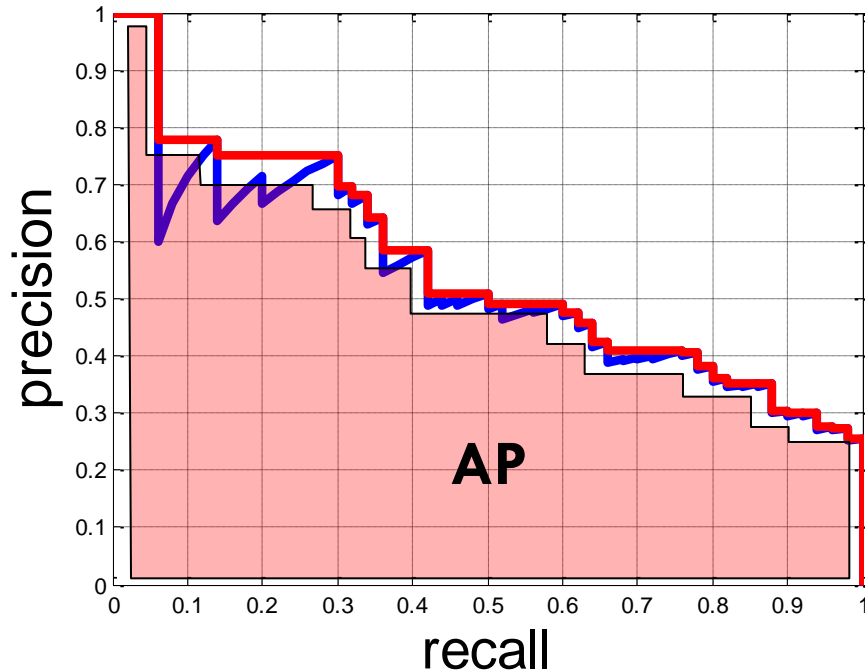
- Around 50% increase in size over VOC2010

	Training		Testing	
<b>Images</b>	609	(376)	569	(320)
<b>Objects</b>	850	(576)	849	(505)

VOC2010 counts shown in brackets

- Set of images taken (and removed) from main dataset
- Images contain only people (none of other 19 classes)

# Evaluation

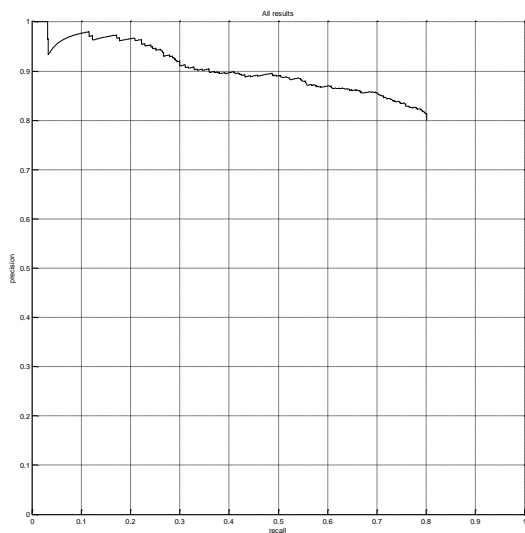


- 2010-2011: Treated as three **separate** detection tasks: Head, Hands, Feet
- Evaluation by AP as in main detection task
- 2007-9 required correct prediction of set of parts visible and bounding boxes: not sensitive enough
- **Invitation: propose a better evaluation scheme!**

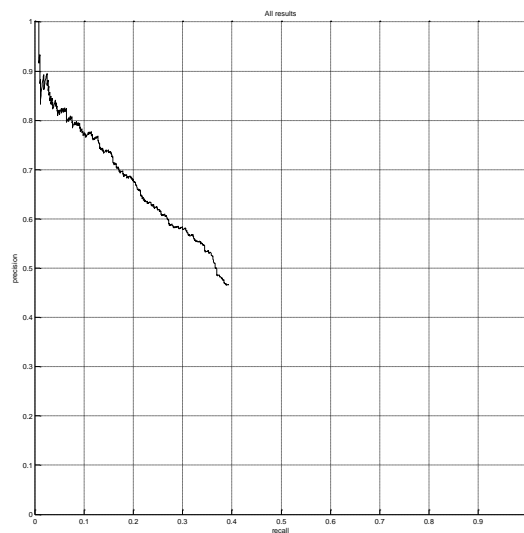
# Results

- Oxford method, using detectors trained on external data

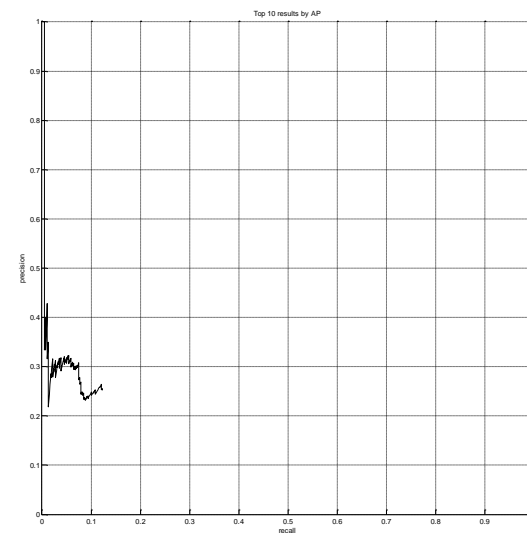
	head	hand	foot
<b>OXFORD_RANK_SLACK_RBF</b>	72.9	26.9	4.1



Head



Hands



Feet



# Structured ranking for Layout Detection

---

## VOC 2011 Human Layout Detection (competition 8) submission

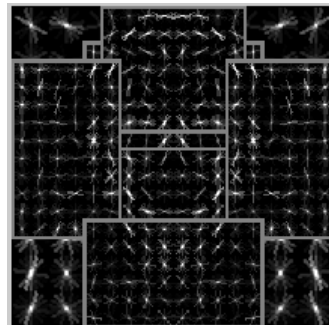
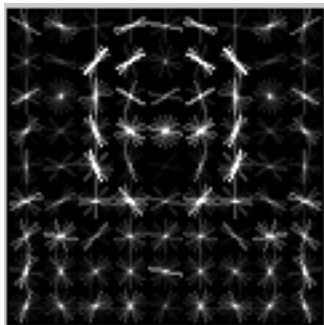
**Contributors:** Arpit Mittal, Matthew Blaschko, Andrew Zisserman, Manuel J Marin and Phil Torr.

### Overview of the method

- Individual body parts (head, hand and foot) are detected using separate detectors.
- Candidate bounding boxes are filtered and scored using local position and scale cues.
- Remaining boxes are combined and scored for human layout confidence using structured SVM ranking

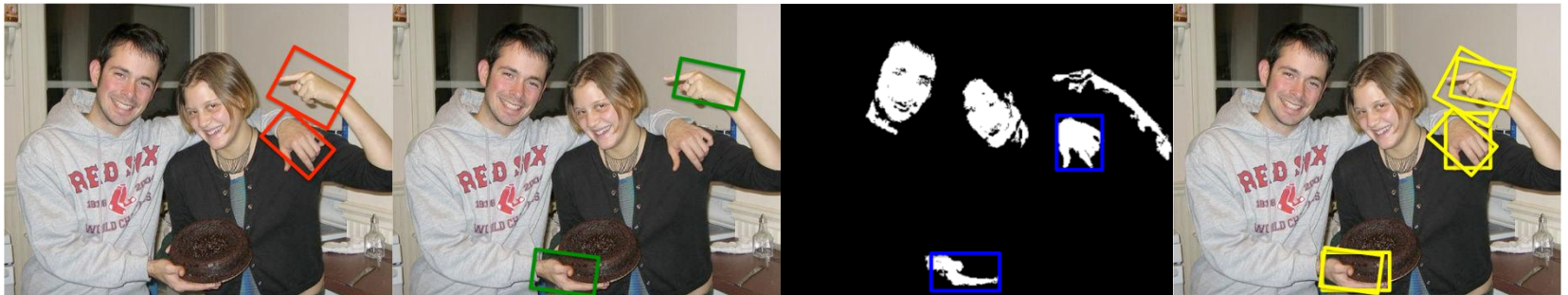
# Head Detector

- Felzenszwalb et. al's part-based model trained on upper bodies and heads
- The model is capable of detecting heads in different poses (frontal, profile and backward)
- Training of the model is done on the Hollywood movie dataset
- Project page: <http://www.robots.ox.ac.uk/~vgg/research/laeo/>



# Hand Detector

- Hands are detected using Mittal et. al's hand detector (BMVC, 2011).
- This uses a hypothesis and verification framework.
- Hands are proposed using shape, context and skin cues.



Shape proposals

Context proposals

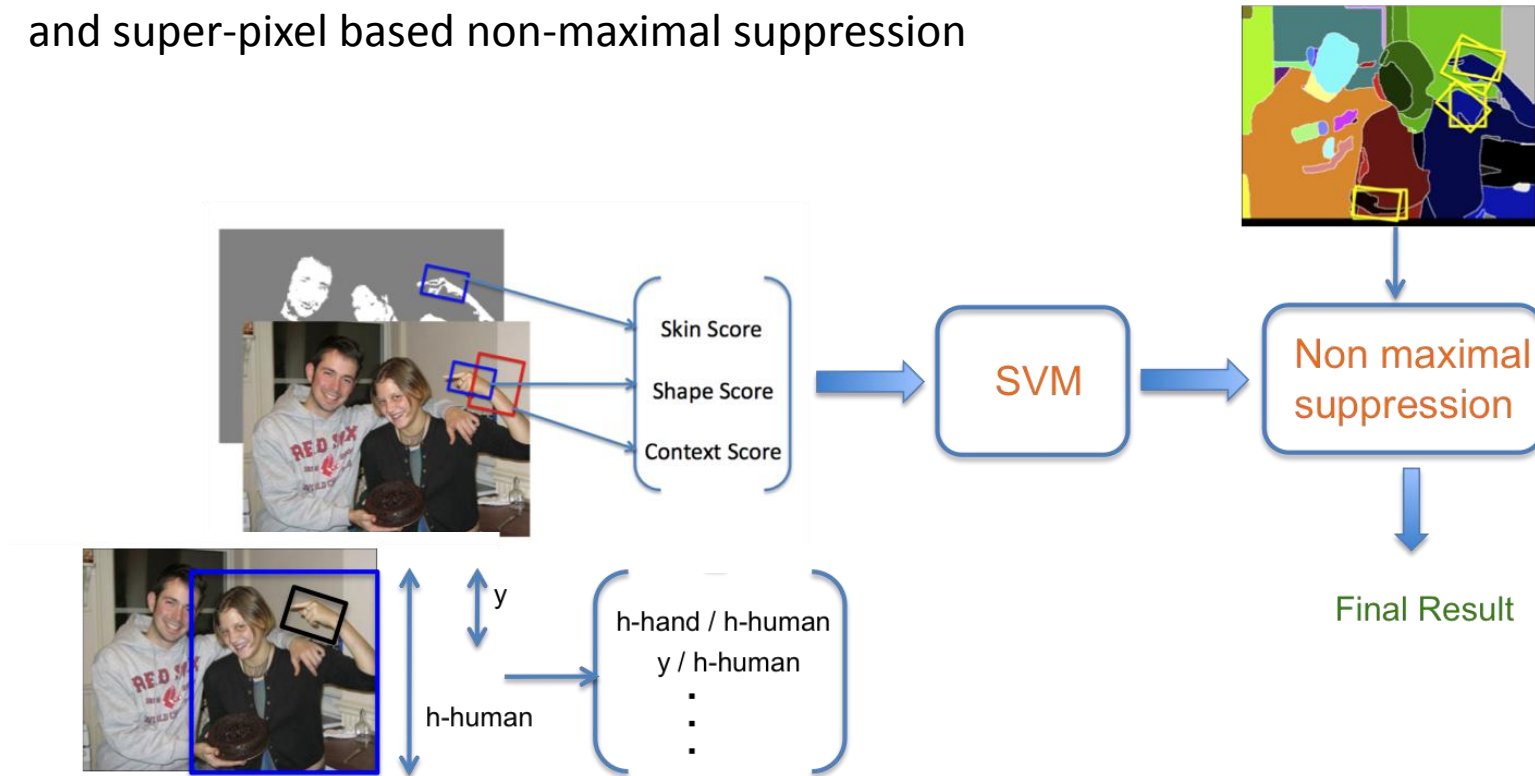
Skin Proposals

All Proposals



# Classification of hand candidates

- Proposed hand boxes are filtered using a linear SVM classifier, and super-pixel based non-maximal suppression



- Project page: <http://www.robots.ox.ac.uk/~vgg/research/hands/index.html>
- Head candidates are classified in a similar manner using positional and scale cues.

## Increase in AP after filtering

For each provided ROI:

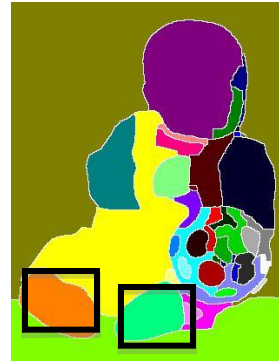
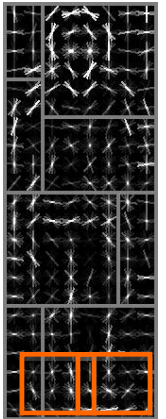
- Retain top scoring head bounding box
- Retain a maximum of 2 hand bounding boxes above threshold
- This increased the performance score significantly.

	Candidates AP {recall}	After classification AP {recall}
Head	65.7 {74.6}	78.9 {82.6}
Hand	21.3 {29.5}	26.6 {37.6}

Training data and test data: VOC 2011 layout data train and validation sets

## Foot Detector

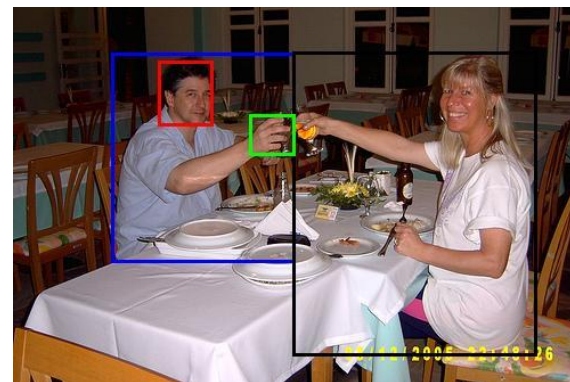
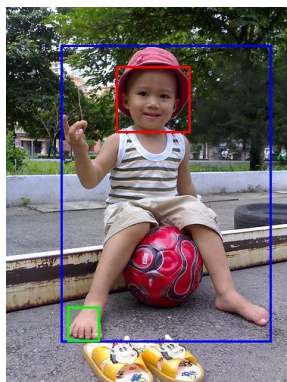
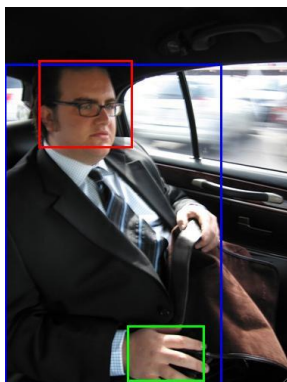
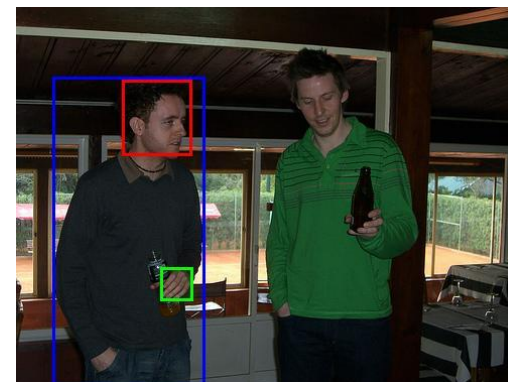
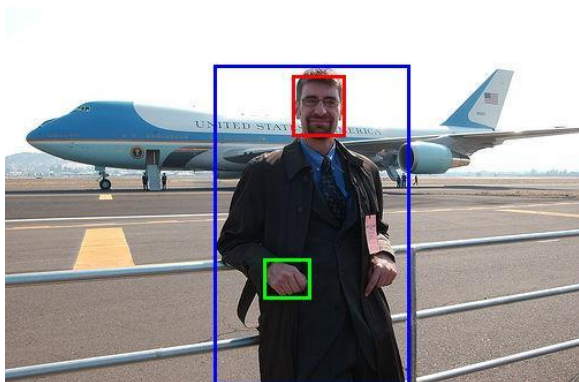
- Foot detection is done in an ad-hoc manner.
- It is detected as the foot part of Felzenszwalb et. al's pedestrian model.
- Also as the bounding box around the super-pixel resembling a human foot in the lower portion of the image.



## Structured Output SVM ranking

- The task is now to combine the part boxes to get a single score for the human ROI.
- This is done using structured output SVM ranking
- The features are those used previously for classifying the hands and head, but now combined into a joint vector
- The loss function used for training is 1-precision, i.e. encouraging configurations where all the parts returned are correct
- The feet do not contribute to the ranking
- The score obtained from the ranking function provides the confidence score for the human ROI
- The slack rescaled variant of SVM structured ranking is used with a RBF kernel.

## Top ranked results on VOC 2011 test dataset



# Prizes



- Winner (!)
  - **OXFORD\_RANK\_SLACK\_BRF**  
Arpit Mittal, Matthew Blaschko,  
Andrew Zisserman, Manuel J Marin, Phil Torr  
*University of Oxford*  
*Oxford Brookes University*