

The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

Discussion

Mark Everingham

Luc Van Gool

Chris Williams

John Winn

Andrew Zisserman



Dataset

- Known Bias
 - Some bias due to keyword-based image collection
 - Images with only many small objects are discarded
 - Segmentation/person layout data is biased towards simple scenes with larger objects
- Small Objects/Context
 - Objects unrecognizable in isolation are ignored in the evaluation but **are** included in the annotation



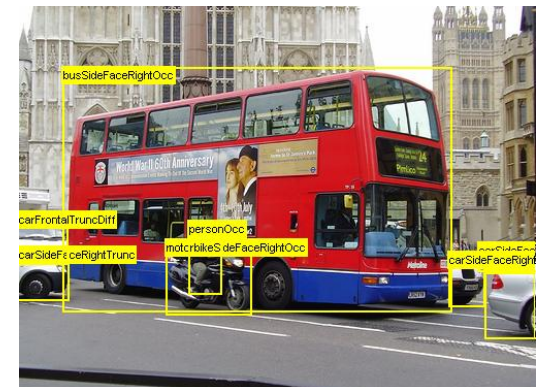
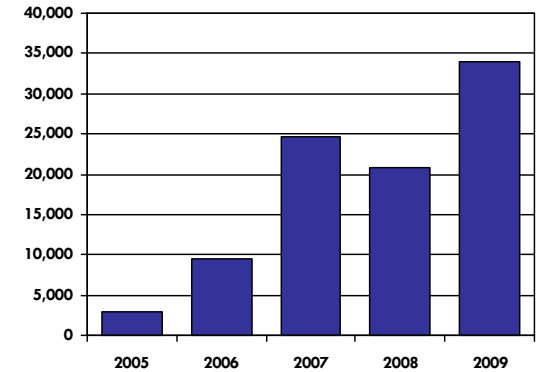
Sustainability

■ Cost & Difficulty

- Annotation is expensive: ~700 person hours for 2009
- Difficult to maintain high quality annotation with increased number of object classes (“cognitive load”)
- Withholding test annotation from 2008 and combining datasets to reduce cost
- 2010 will use Mechanical Turk

■ Availability of Data

- Becoming difficult to find examples of certain categories on flickr



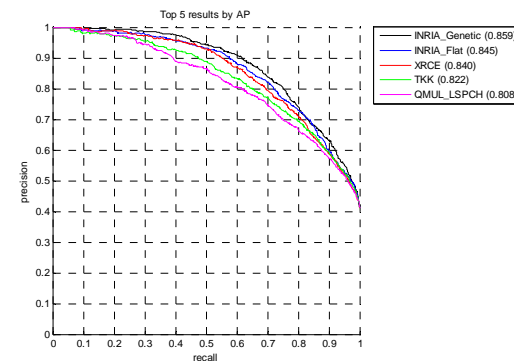
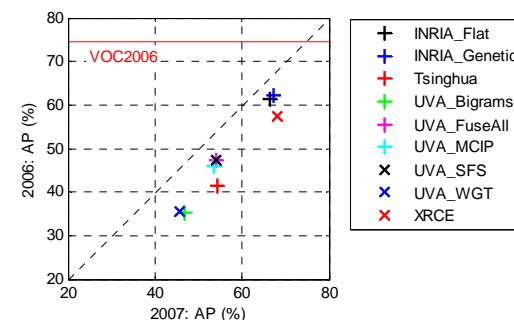
Challenge

■ “Longitudinal” Data

- New test set every year makes measuring improvement difficult
- 2009 allows test on 2008 but with more training data
- Stop collecting more (test) data?

■ Classification: “Pushing the curve”?

- Are we encouraging incremental research?
- Is “bag of words” the last word?
- Can we (please) drop this task or make more “realistic”?



Annotation

- Bounding Boxes?

- More suitable for some objects than others...



- Alternatives?

- Should we be annotating less data in more detail?
 - Polygons, “sketches”, parts, pixels, ...?
- Should we be annotating more data in less detail?
 - Weak supervision e.g. keywords at image level?
- Are we annotating the right data?
 - Video?

Evaluation

- Useful to the community?
 - Are we measuring the right thing?
 - How to provide useful diagnostic information to guide research?

- “Taster” Challenges
 - Layout evaluation too punitive...
 - Should we continue with this task?
 - What other tasks should be introduced to stimulate research?

