The PASCAL Visual Object Classes Challenge 2008 (VOC2008)

Part 2 – Detection Task

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

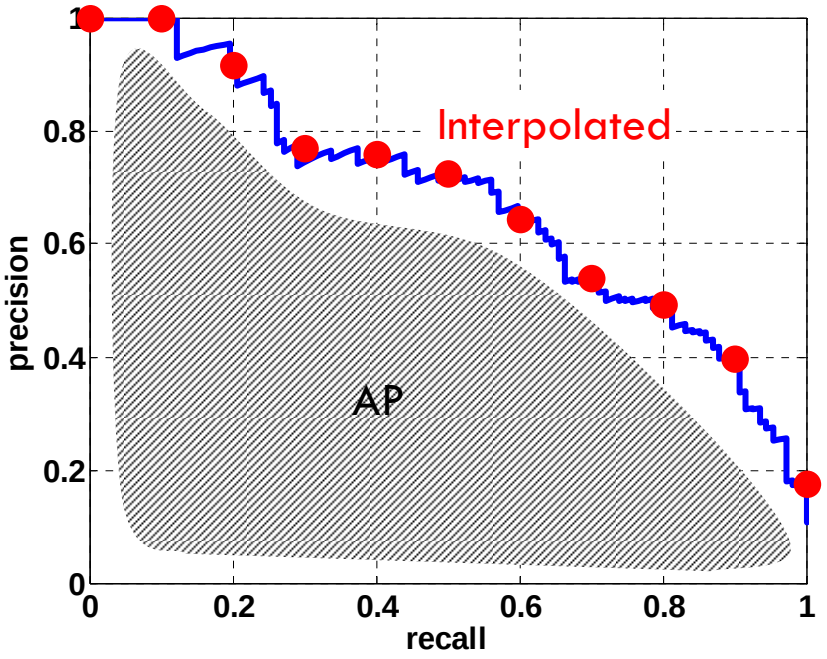
Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)

- Competition 3: Train on the supplied data
  - Which methods perform best given specified training data?
- Competition 4: Train on any (non-test) data
  - How well do state-of-the-art methods perform on these problems?

Evaluation

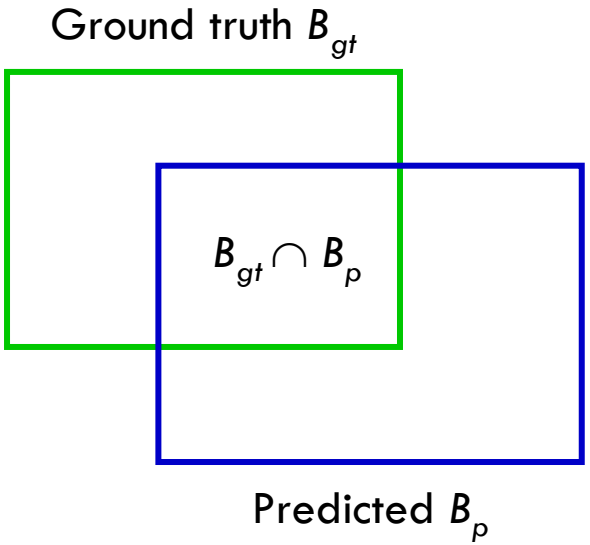
- Average Precision [TREC] averages precision over the entire range of recall
  - Curve interpolated to reduce influence of "outliers"



- A good score requires both high recall and high precision
- Application-independent
- Penalizes methods giving high precision but low recall

Evaluating Bounding Boxes

- Area of Overlap (AO) Measure



$$AO(B_{gt}, B_p) = \frac{|B_{gt} \bigcap B_p|}{|B_{gt} \bigcup B_p|}$$

- Need to define a threshold *t* such that $AO(B_{gt},B_p)$ implies a correct detection: 50%

Methods

- **"Sliding window classifier" predominant**
- Features
  - HoG, pyramid of HOG
  - Bag of words (SIFT)
  - Spatial pyramids (SIFT, Color SIFT)
  - Self-similarity
- Classifiers
  - Linear SVM
  - SVM with χ^2 kernel
  - Mixture of "star models" of linear SVM's
  - Structured output regression

Methods

- Other aspects
  - Efficient search for window (MPI)
  - Windows derived from “superpixels” (CASIA)
  - Heterogeneous multiple-stage classifiers e.g. HOG then bag of words (LEAR/Oxford)
  - Combining confidence of detection with whole-image context (LEAR/Chicago-UCI)
  - Regression to improve accuracy of bounding box (Chicago-UCI)
- Other
  - “Kitchen sink”

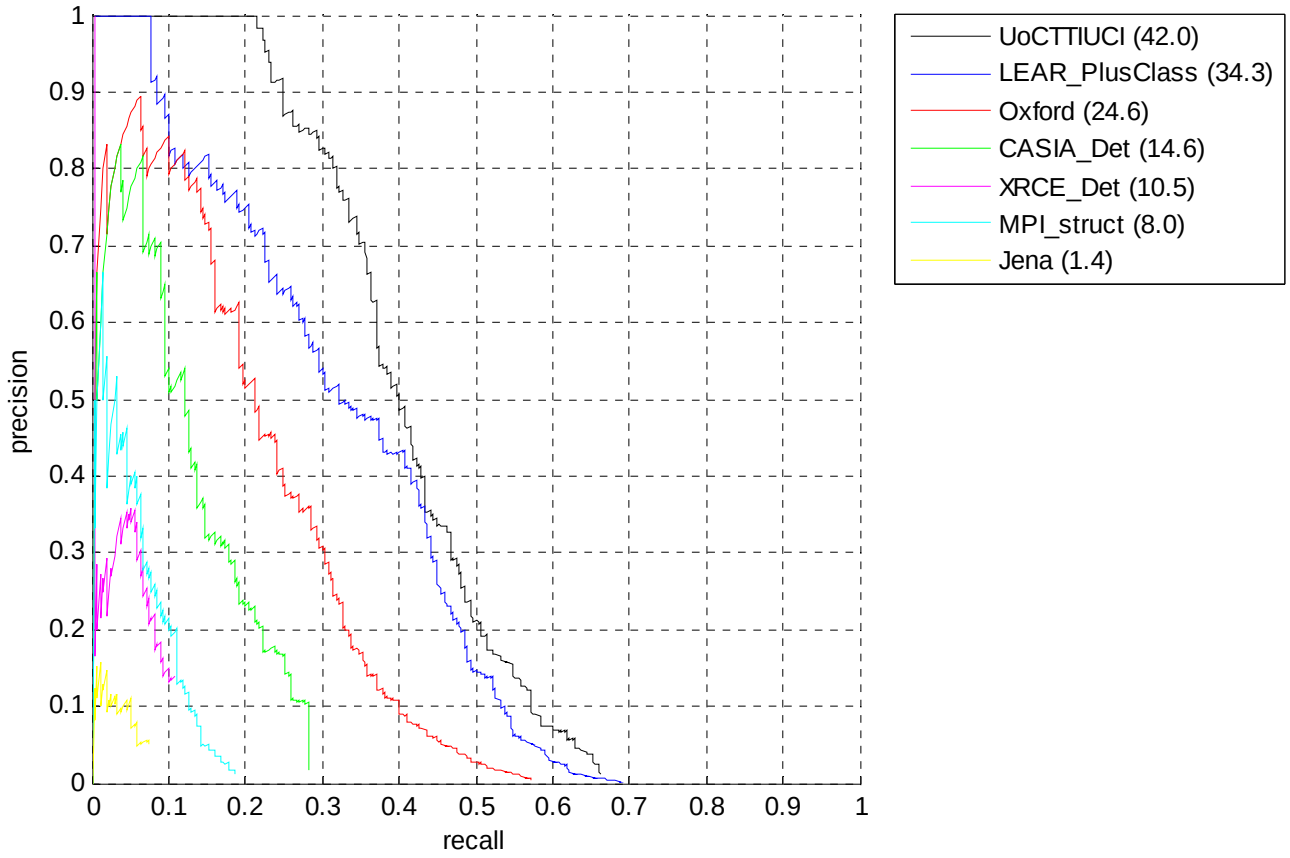
AP by Method and Class

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	pers	plant	sheep	sofa	train	tv
CASIA_Det	25.2	14.6	9.8	10.5	6.3	23.2	17.6	9.0	9.6	10.0	13.0	5.5	14.0	24.1	11.2	3.0	2.8	3.0	28.2	14.6
Jena	4.8	1.4	0.3	0.2	0.1	1.0	1.3	-	0.1	4.7	0.4	1.9	0.3	3.1	2.0	0.3	0.4	2.2	6.4	13.7
LEAR_PlusClass	36.5	34.3	10.7	11.4	22.1	23.8	36.6	16.6	11.1	17.7	15.1	9.0	36.1	40.3	19.7	11.5	19.4	17.3	29.6	34.0
MPI_struct	25.9	8.0	10.1	5.6	0.1	11.3	10.6	21.3	0.3	4.5	10.1	14.9	16.6	20.0	2.5	0.2	9.3	12.3	23.6	1.5
Oxford	33.3	24.6	-	-	-	-	29.1	-	-	12.5	-	-	32.5	34.9	-	-	-	-	-	-
UoCTTIUCI	32.6	42.0	11.3	11.0	28.2	23.2	32.0	17.9	14.6	11.1	6.6	10.2	32.7	38.6	42.0	12.6	16.1	13.6	24.4	37.1
XRCE_Det	26.4	10.5	1.4	4.5	0.0	10.8	4.0	7.6	2.0	1.8	4.5	10.5	11.8	13.6	9.0	1.5	6.1	1.8	7.3	6.8

- No clear "winner"
  - LEAR_PlusClass 1st on 11 classes, 2nd on 7
  - UoCTTIUCI 1st on 7 classes, 2nd on 8

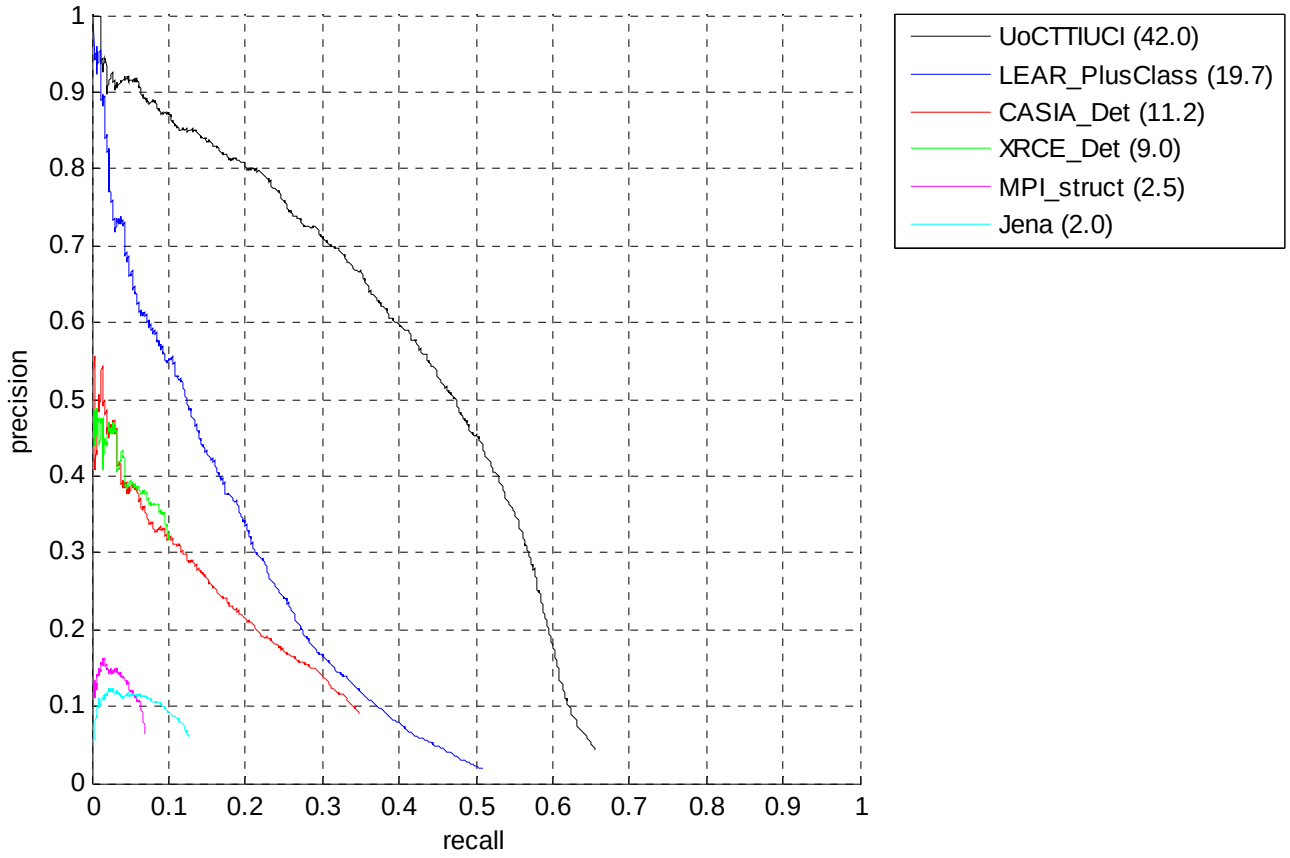
Example Precision/Recall: Bicycle

- Bicycle



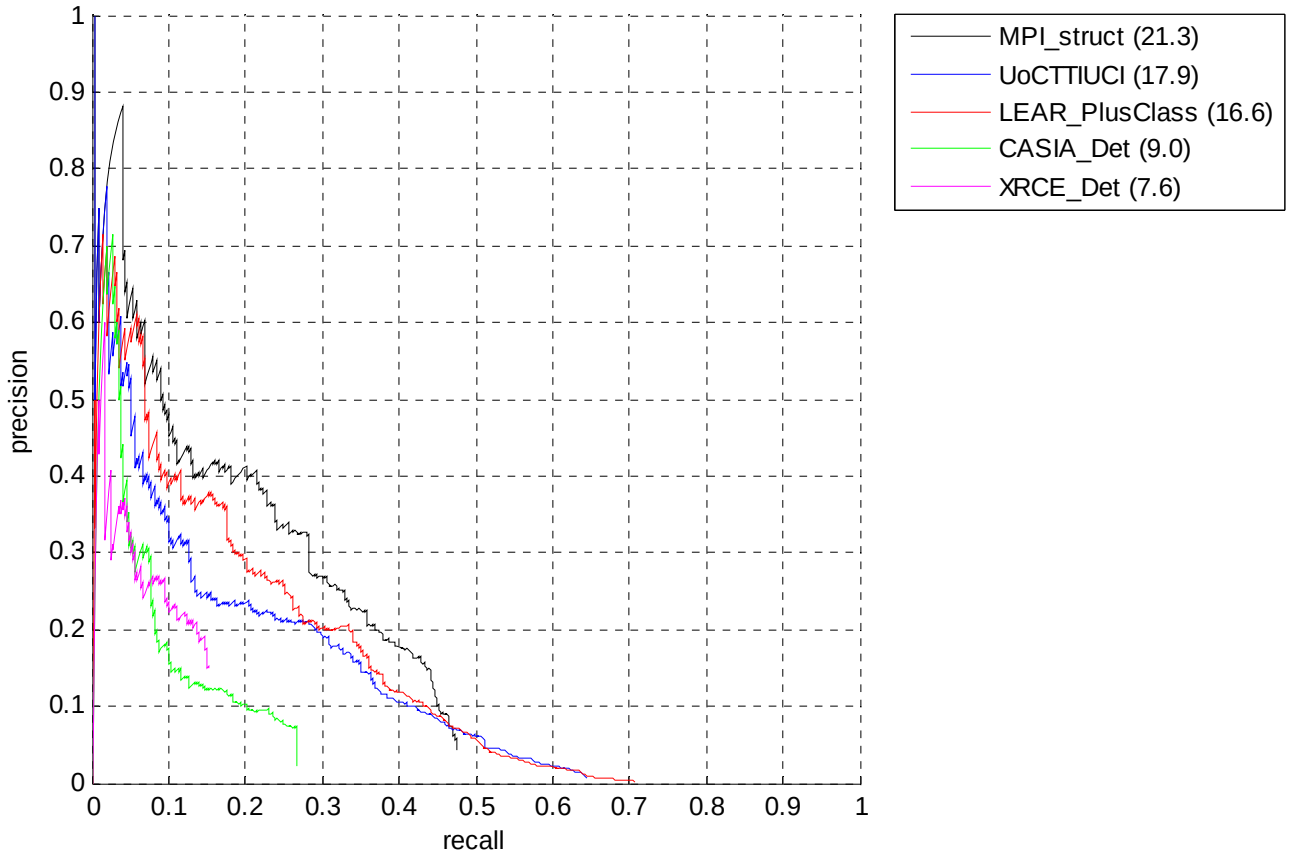
Example Precision/Recall: Person

- Person



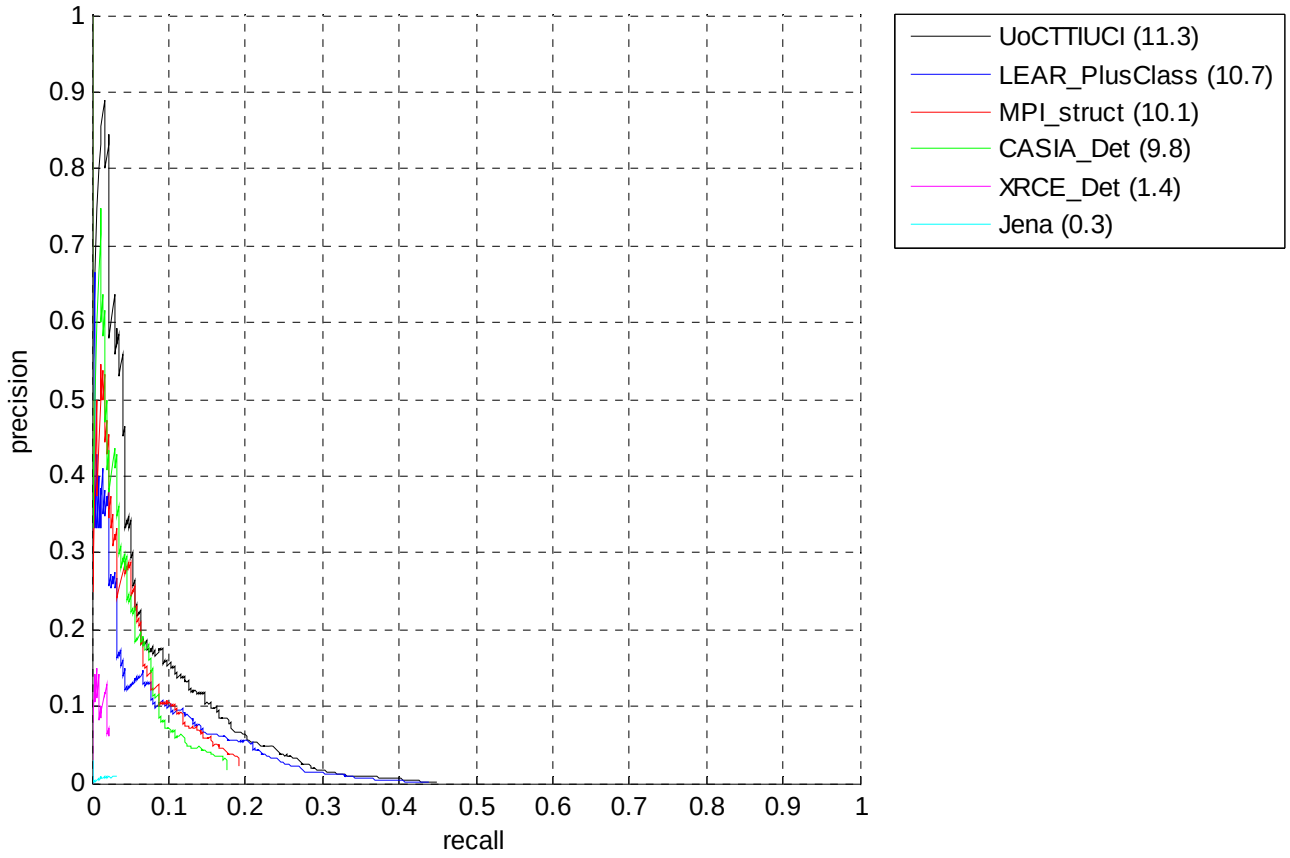
Example Precision/Recall: Cat

- Cat

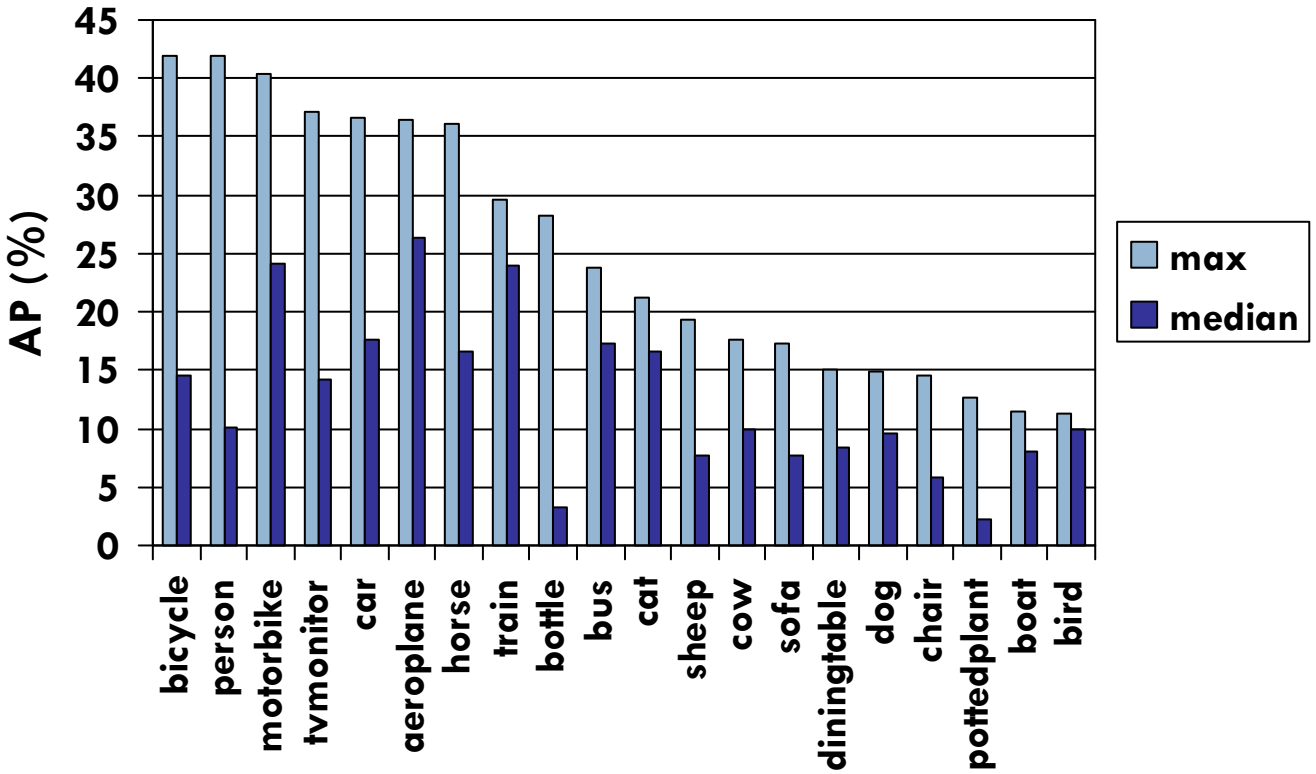


Example Precision/Recall: Bird

- Bird



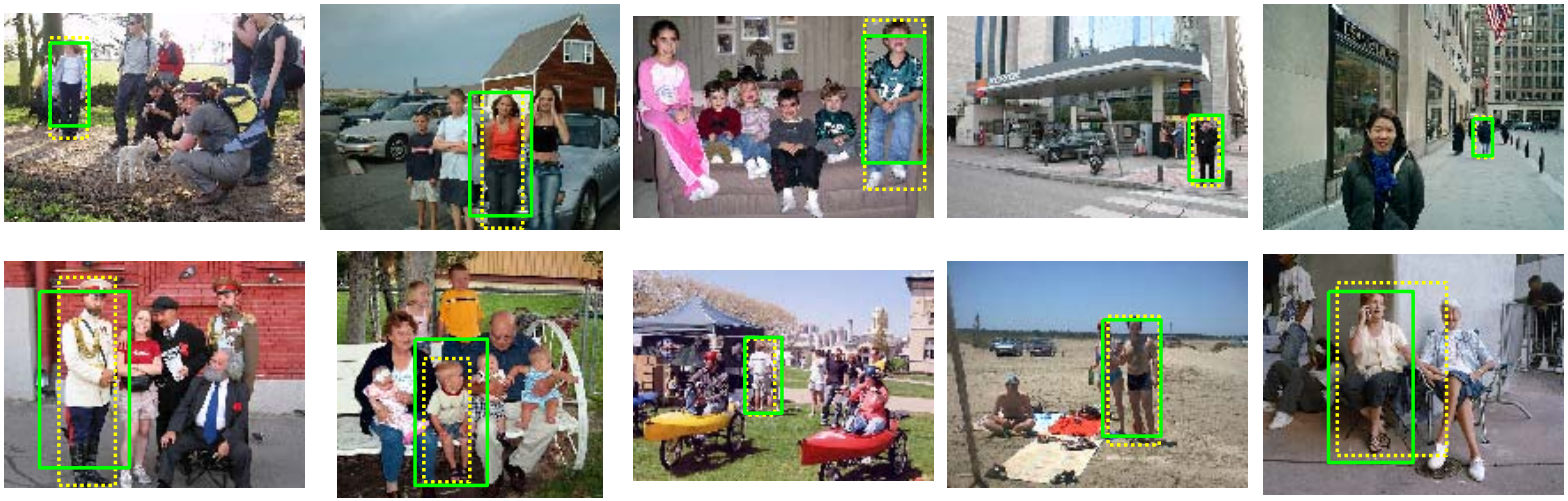
AP by Class



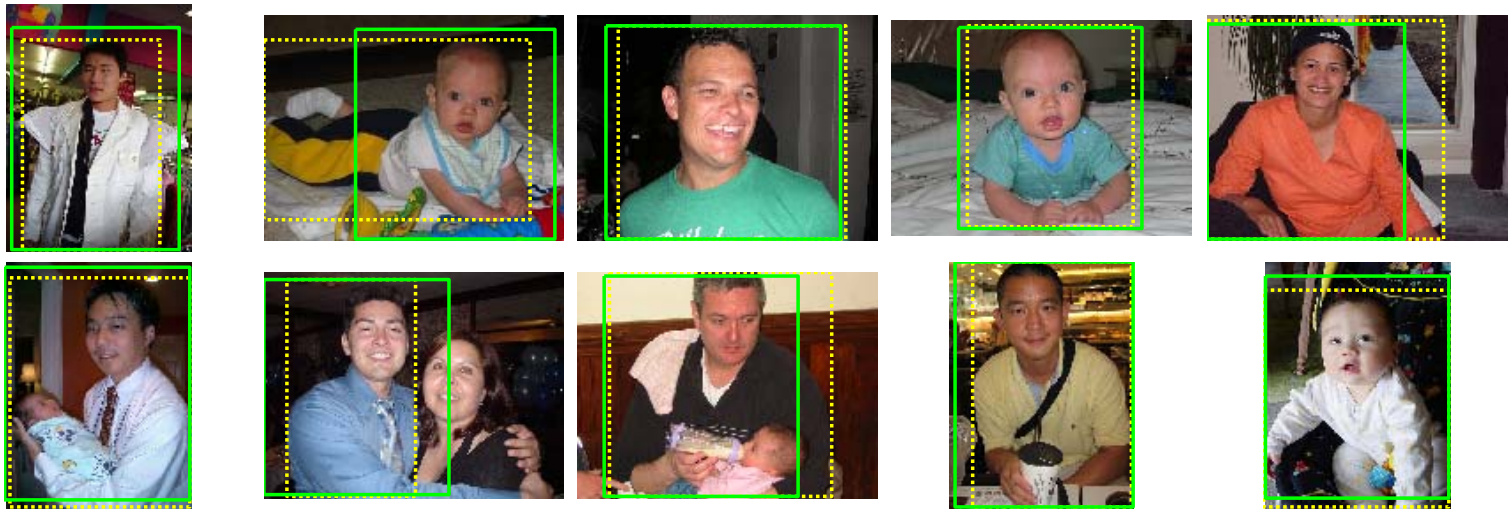
- Substantial difference between median and maximum results

True Positives: Person

LEAR_PlusClass

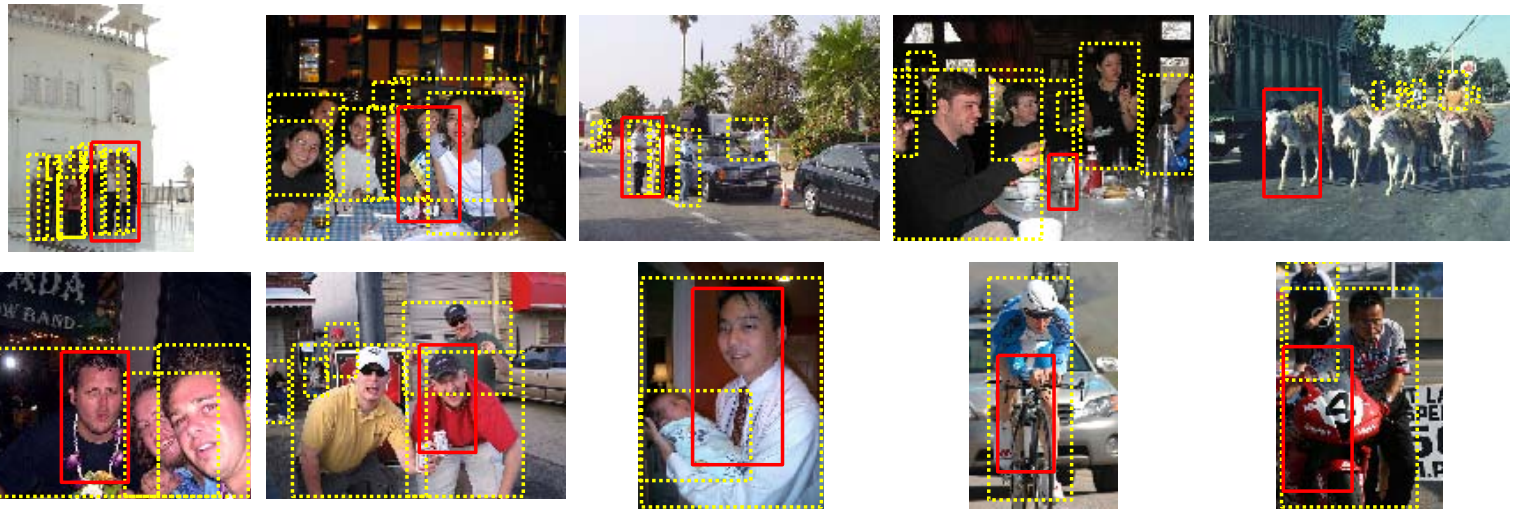


UoCTTIUCI

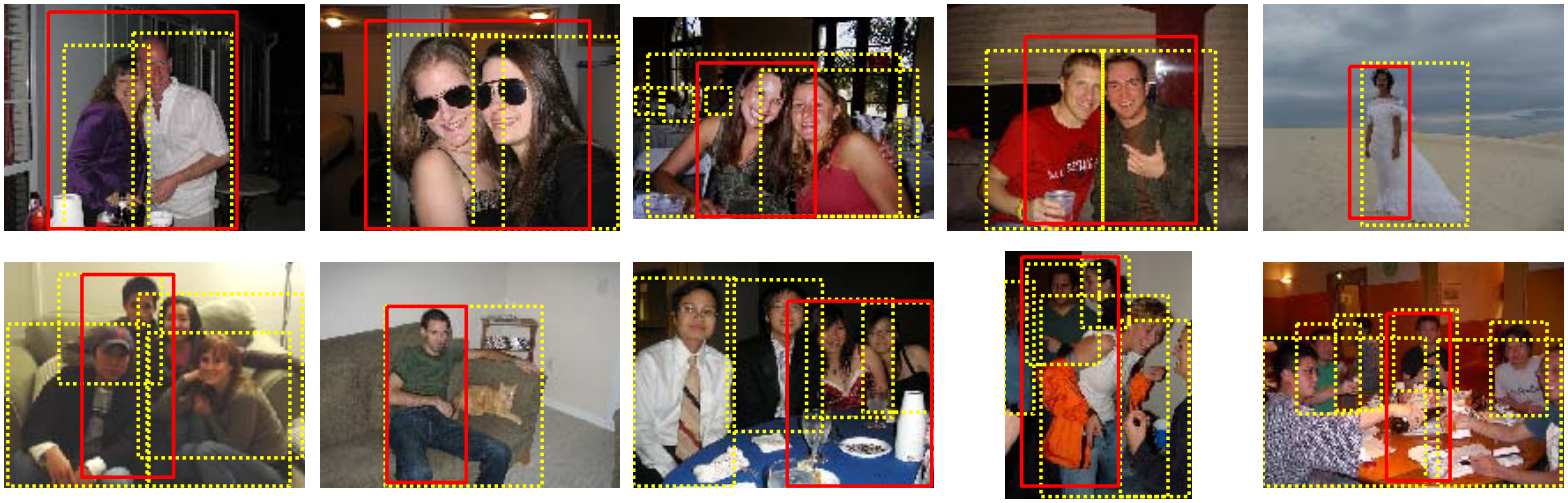


“Near Misses”: Person

LEAR_PlusClass



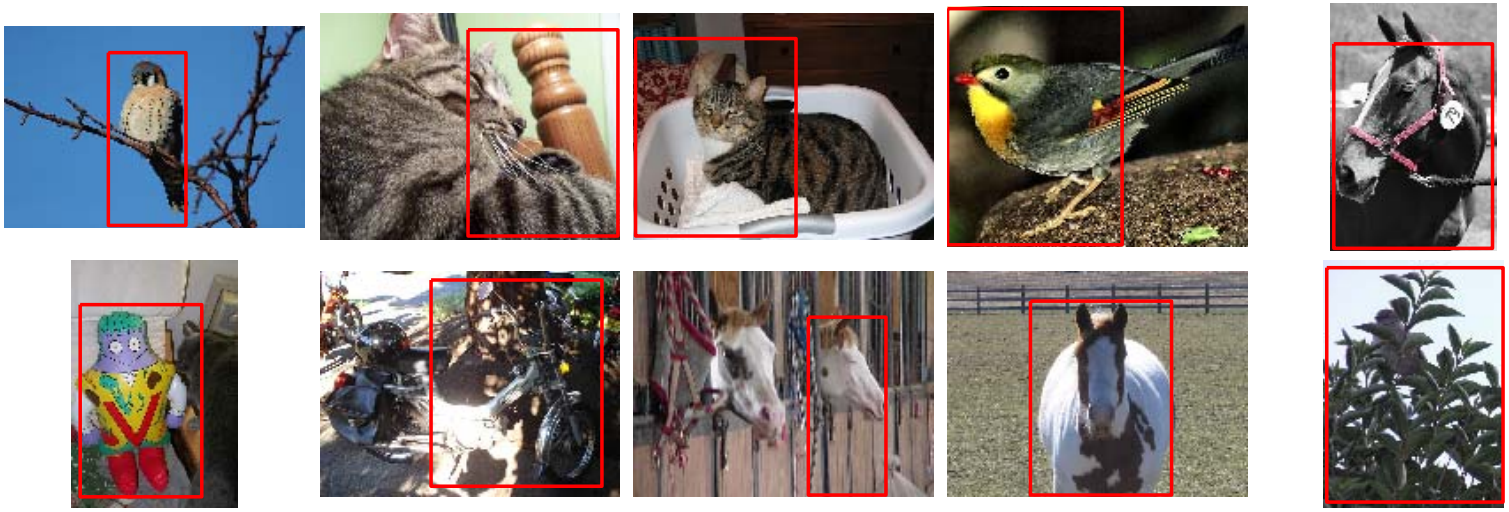
UoCTTIUCI



False Positives: Person

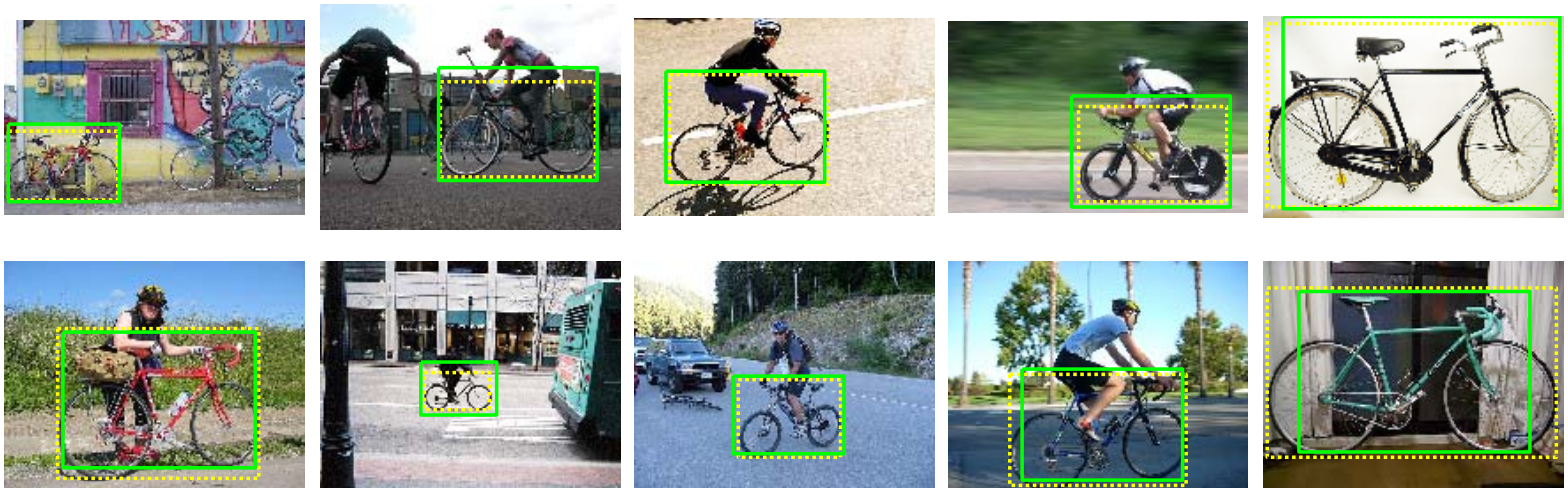
LEAR_PlusClass

UoCTTIUCI

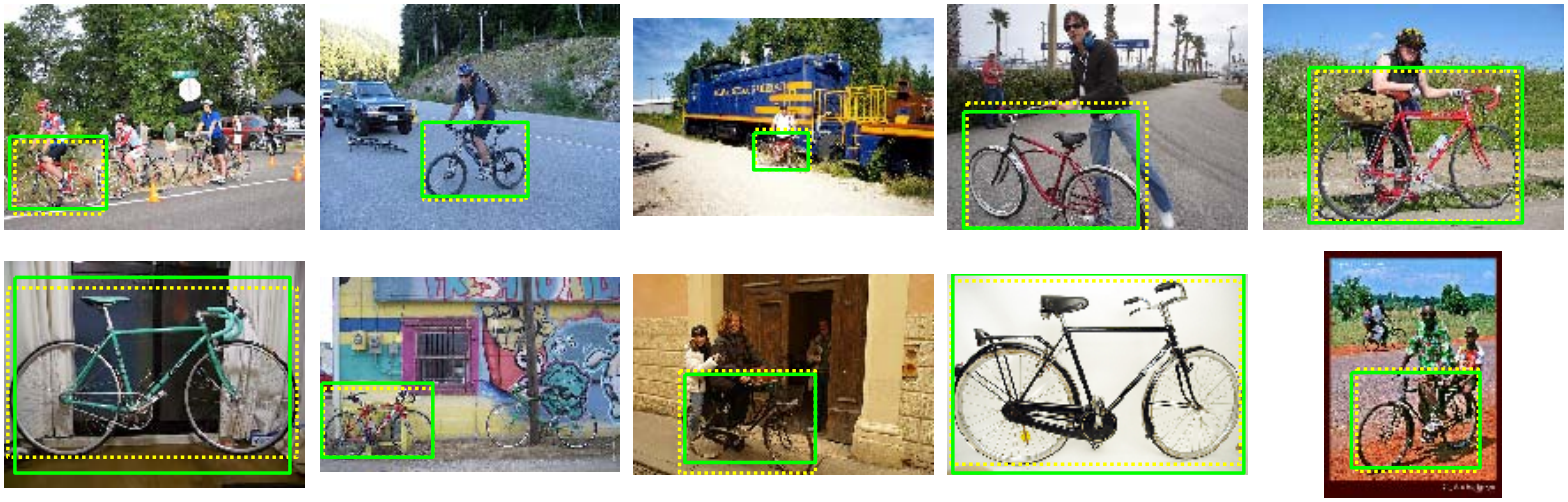


True Positives: Bicycle

LEAR_PlusClass

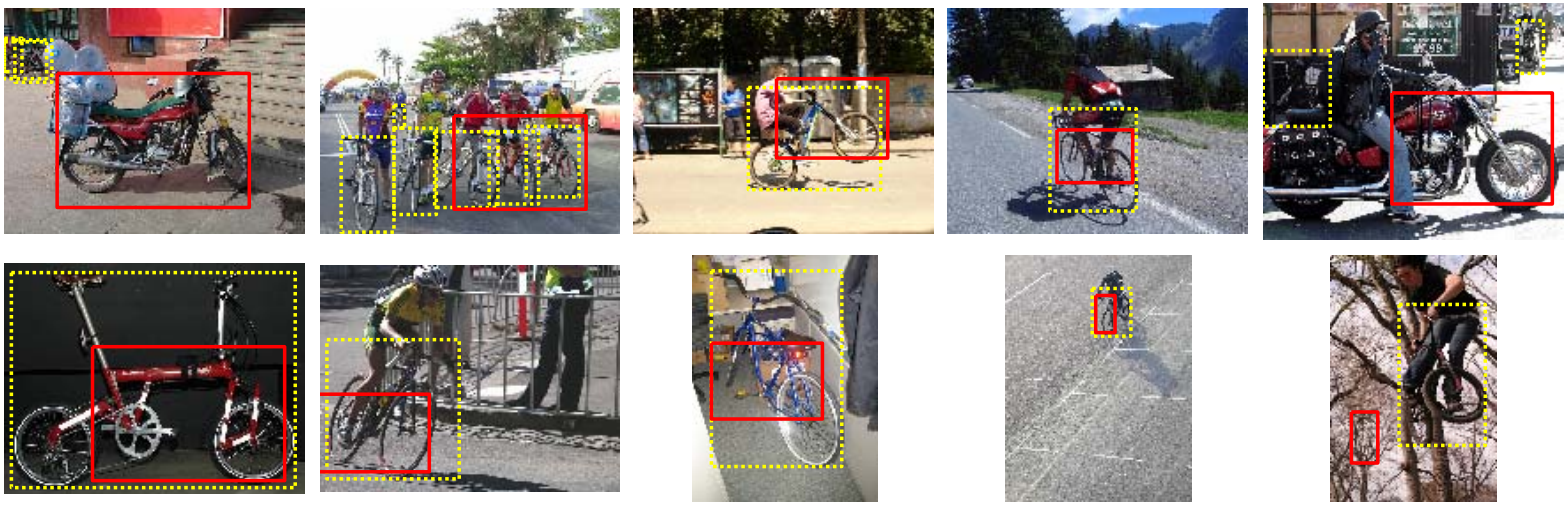


UoCTTIUCI

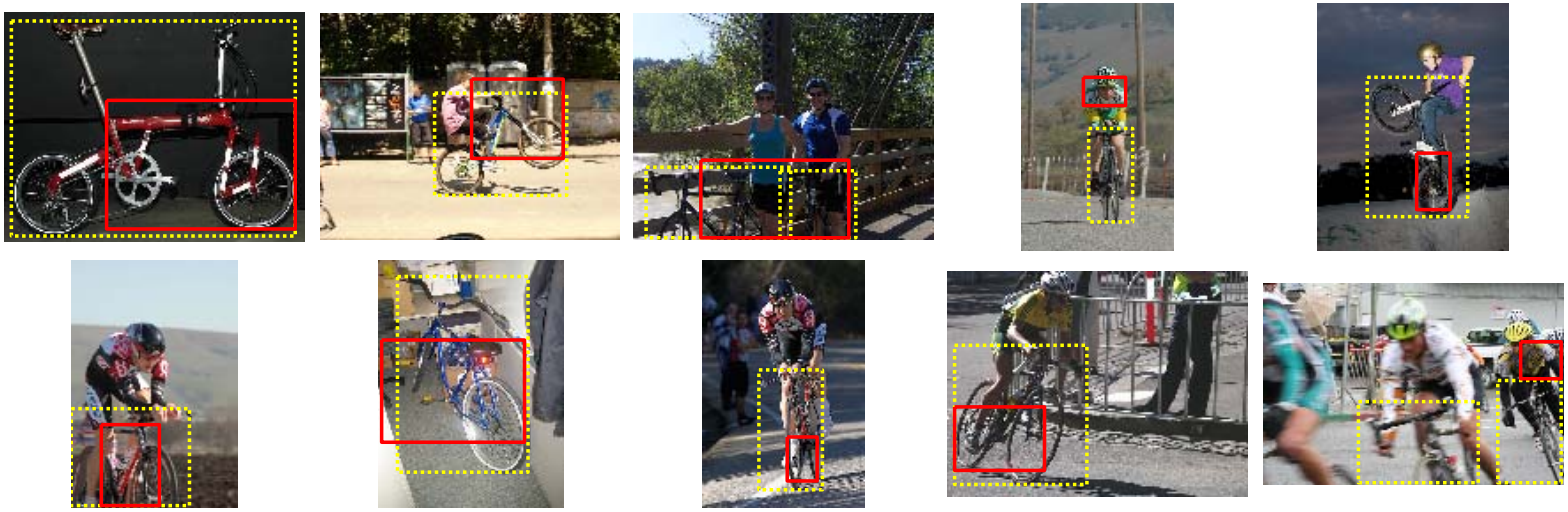


“Near Misses”: Bicycle

LEAR_PlusClass

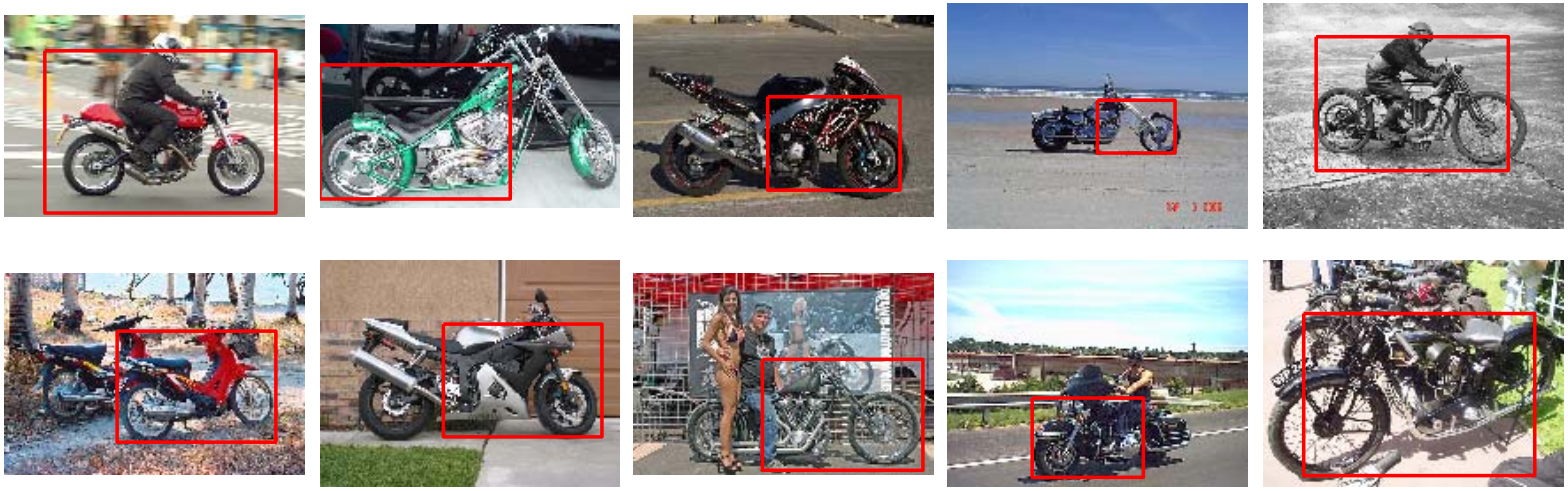


UoCTTIUCI

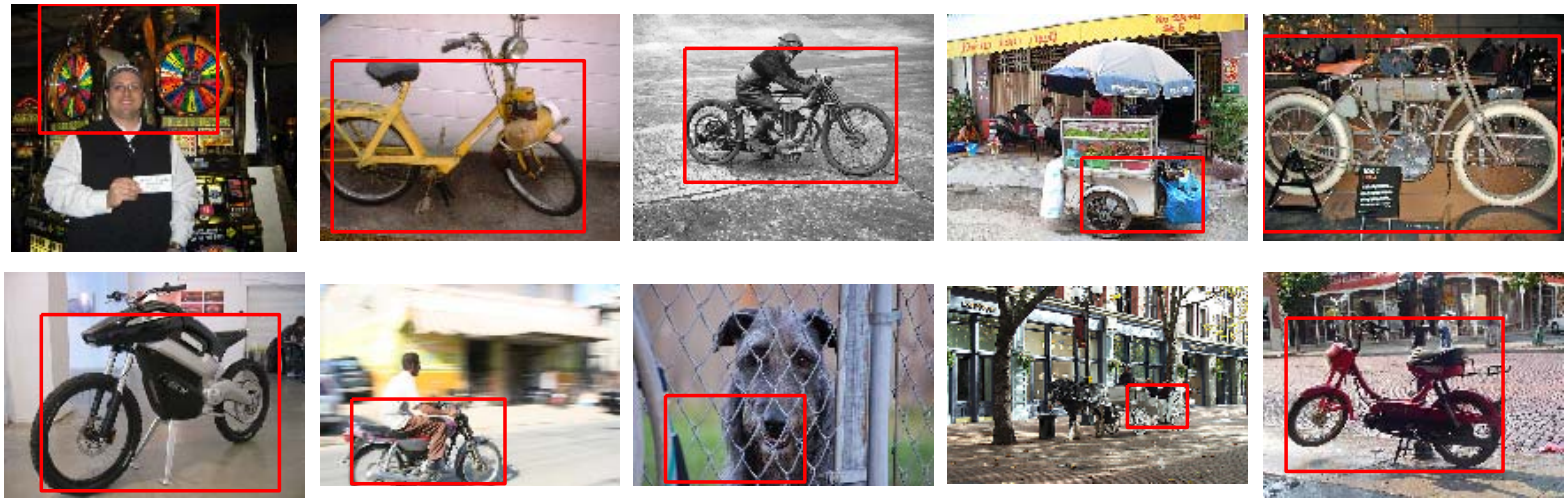


False Positives: Bicycle

LEAR_PlusClass

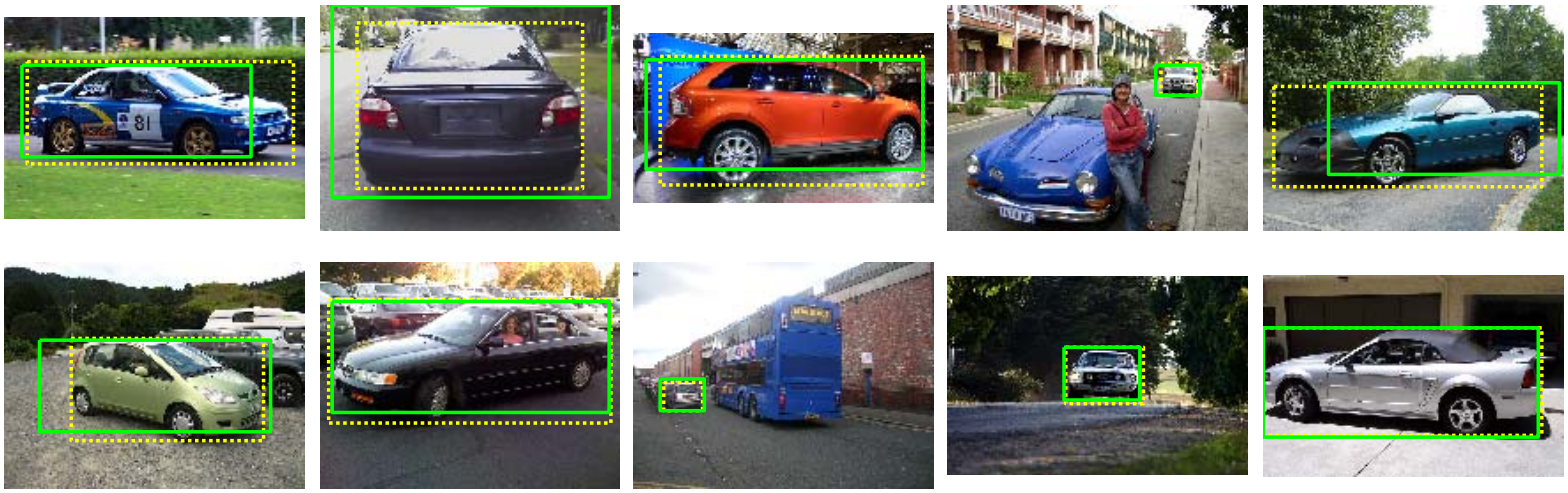


UoCTTIUCI

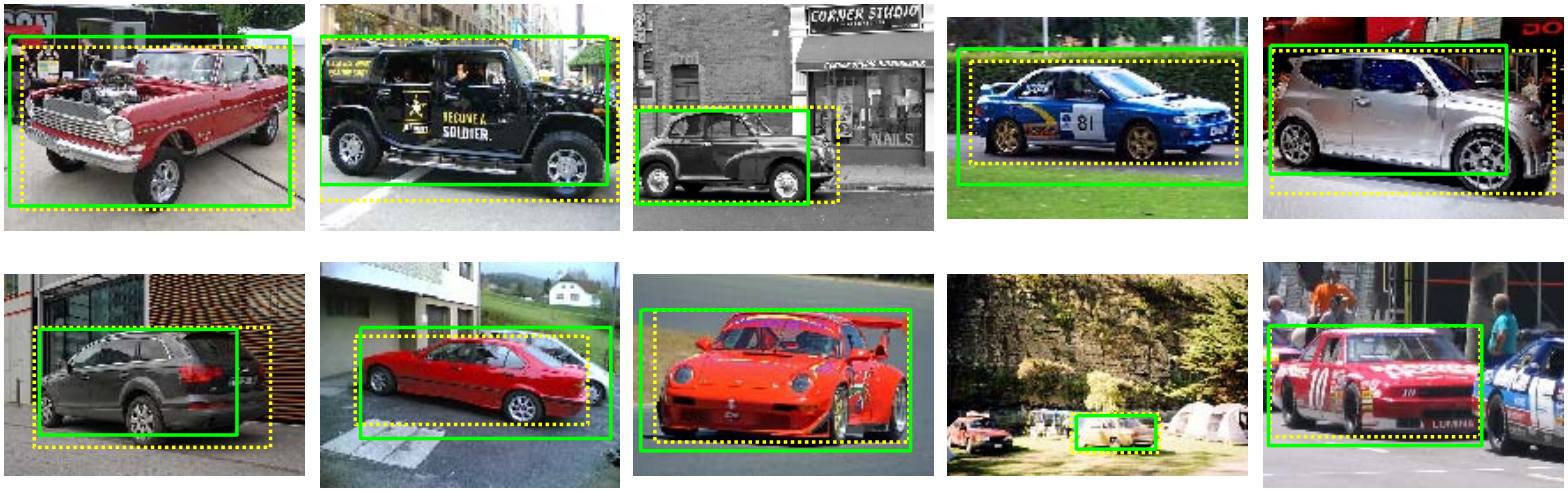


True Positives: Car

LEAR_PlusClass

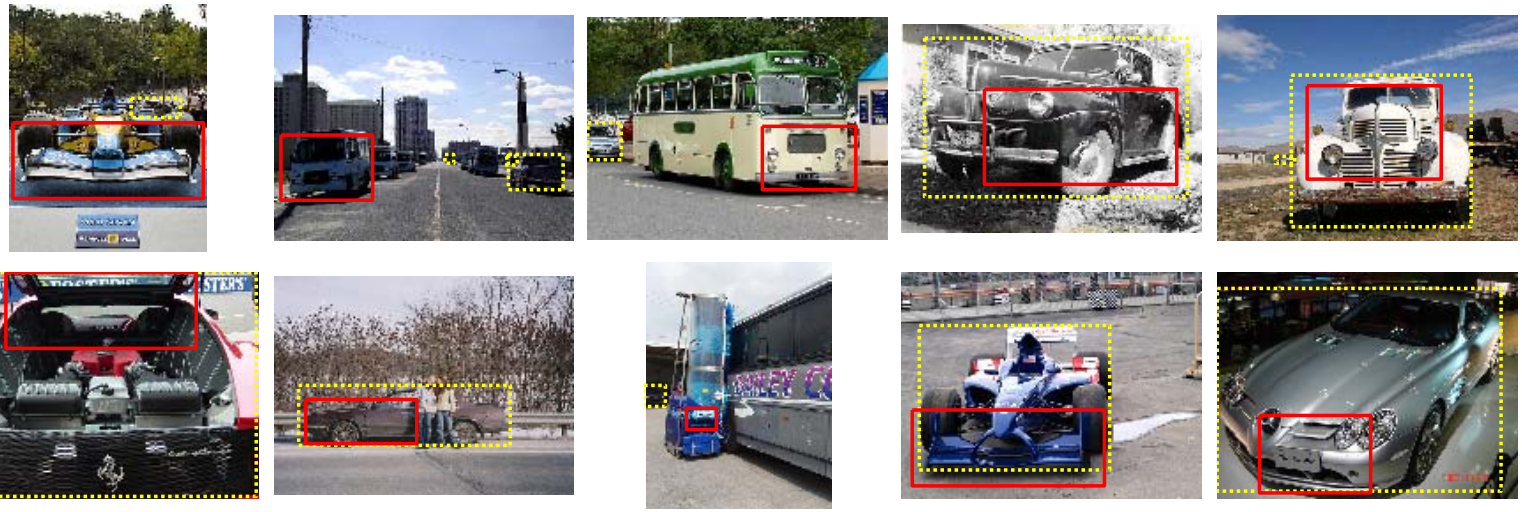


UoCTTIUCI

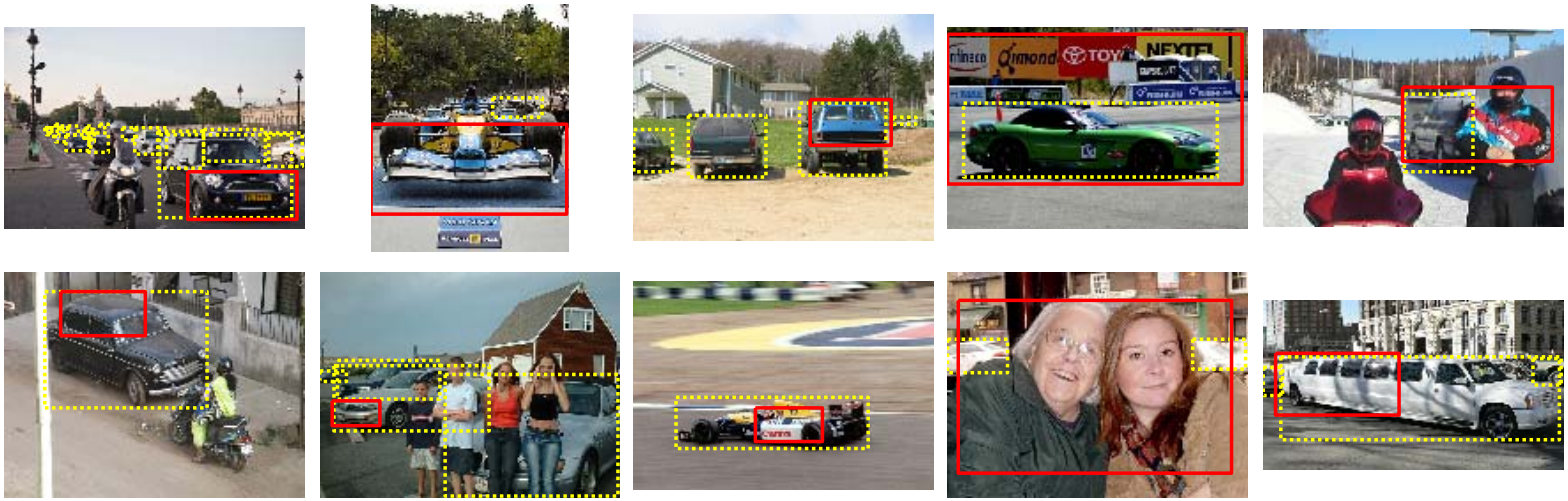


“Near Misses”: Car

LEAR_PlusClass



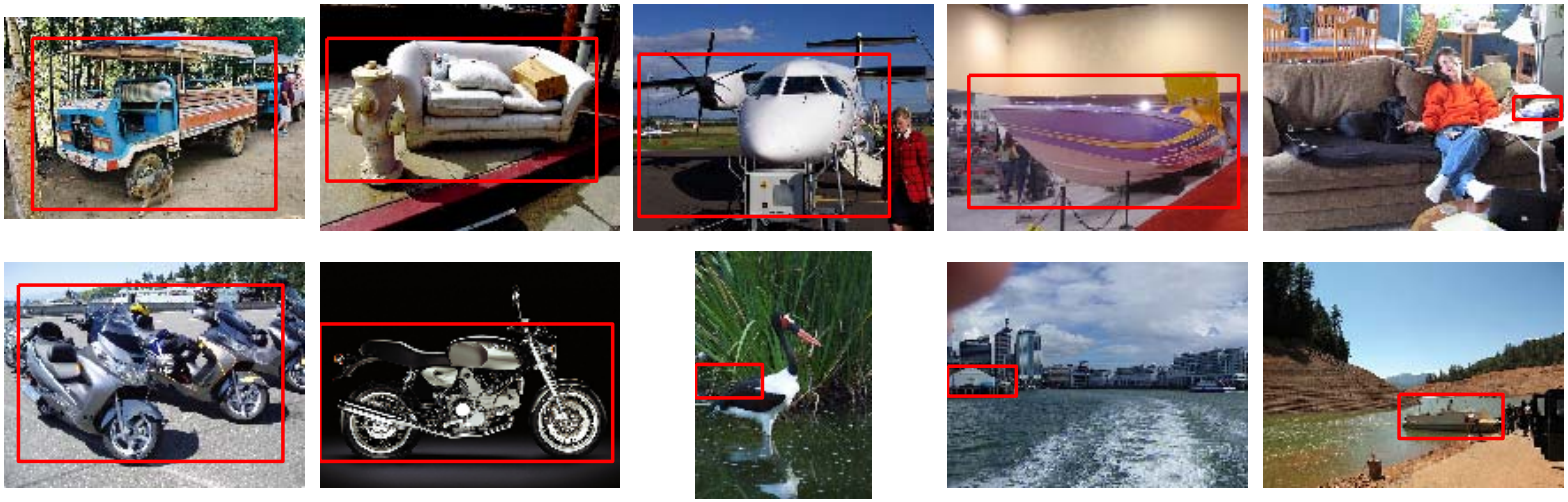
UoCTTIUCI



False Positives: Car

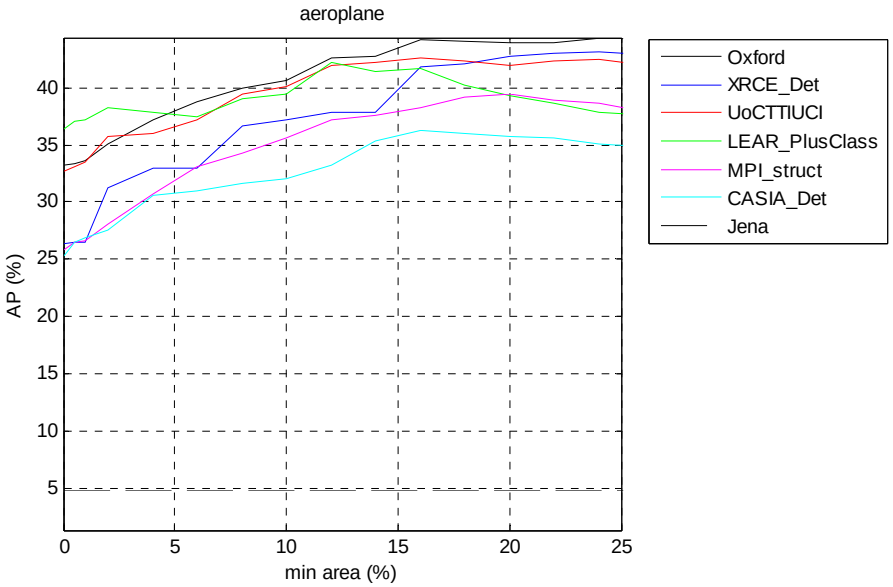
LEAR_PlusClass

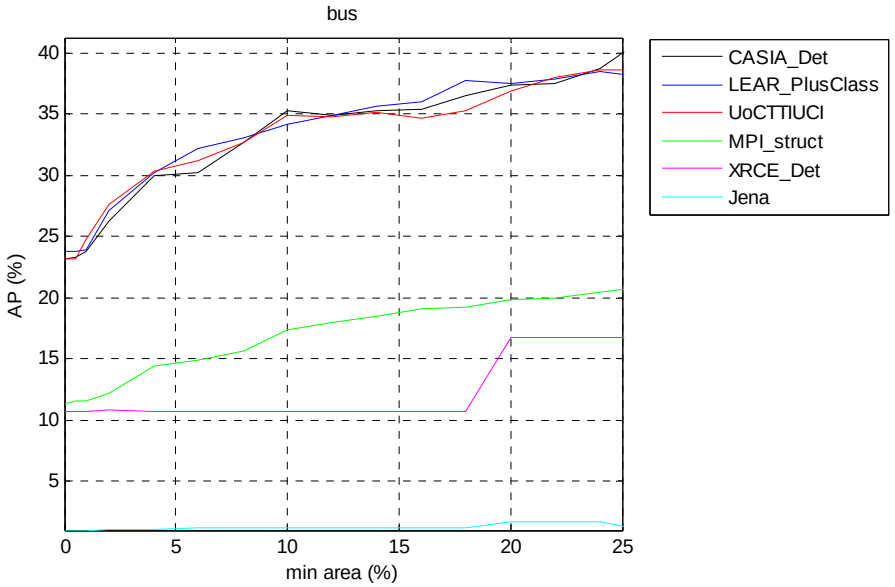
UoCTTIUCI



AP vs. Object Area

- Do these methods have a bias toward larger objects?





- Most methods show moderate preference for larger objects – use of bag of words stages and whole-image classifiers

External Training Data

- UIUC_CMU used pre-built detectors, im2gps, etc.

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	pers	plant	sheep	sofa	train	tv
Best 2008	36.5	42.0	11.3	11.4	28.2	23.8	36.3	21.3	14.6	17.7	15.1	14.9	36.1	40.3	42.0	12.6	19.4	17.3	29.6	34.0
UIUC_CMU	34.5	32.7	12.3	11.0	22.4	18.5	27.8	21.6	8.8	14.1	15.2	17.8	27.4	40.9	37.4	11.2	7.0	13.5	28.2	38.5

- Greater AP than best 2008-only method on 6 classes

VOC2007 vs. VOC2008 Test Data

		aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	pers	plant	sheep	sofa	train	tv
Test on 2008	LEAR_PlusClass	36.5	34.3	10.7	11.4	22.1	23.8	36.6	16.6	11.1	17.7	15.1	9.0	36.1	40.3	19.7	11.5	19.4	17.3	29.6	34.0
	Oxford	33.3	24.6	-	-	-	-	29.1	-	-	12.5	-	-	32.5	34.9	-	-	-	-	-	-
Test on 2007	LEAR_PlusClass	28.5	39.0	10.7	11.2	20.2	41.0	48.4	15.2	16.1	25.7	10.1	11.5	34.9	39.7	16.8	10.3	21.8	22.8	37.0	36.3
	Oxford	27.7	29.1	-	-	-	-	41.5	-	-	16.3	-	-	31.9	33.8	-	-	-	-	-	-
2007	Best	26.2	40.9	9.8	9.4	21.4	39.3	43.2	24.0	12.8	14.0	9.8	16.2	33.5	37.5	22.1	12.0	17.5	14.7	33.4	28.9

- High correlation between results on 2008 and 2007 test data
- For 14/20 classes, 2008 methods did better than the best 2007 method
  - Caveat: 2008 training data helped or hindered?

Prizes

- Joint Winners:
  - LEAR_PlusClass

    Hedi Harzallah, Cordelia Schmid, Frederic Jurie, Adrien Gaidon

    INRIA Rhone-Alpes
  - UOCTTIUCI

    Pedro Felzenszwalb[1], Ross Girshick[1], David McAllester[2], Deva Ramanan[3]

    [1]University of Chicago; [2]TTI Chicago;

    [3]University of California, Irvine