The PASCAL Visual Object Classes Challenge 2010 (VOC2010)

Part 3 – Segmentation Challenge

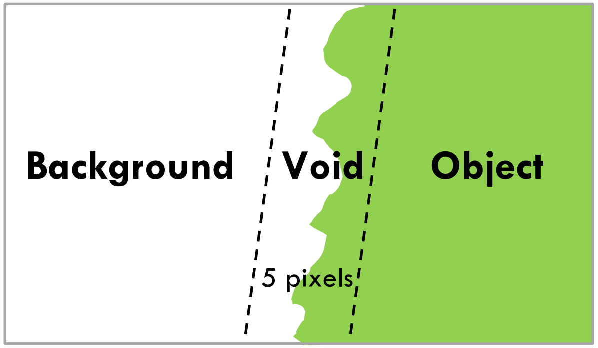
Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

Segmentation Challenge

- For each pixel, predict the class of the object containing that pixel or ‘background’.
- Competition 5: Train on the supplied data
  - Which methods perform best given specified training data?
  - Can use bounding box data as well as seg. data
- Competition 6: Train on any (non-test) data
  - Available since VOC2009
  - Allows for use of own data

Annotation

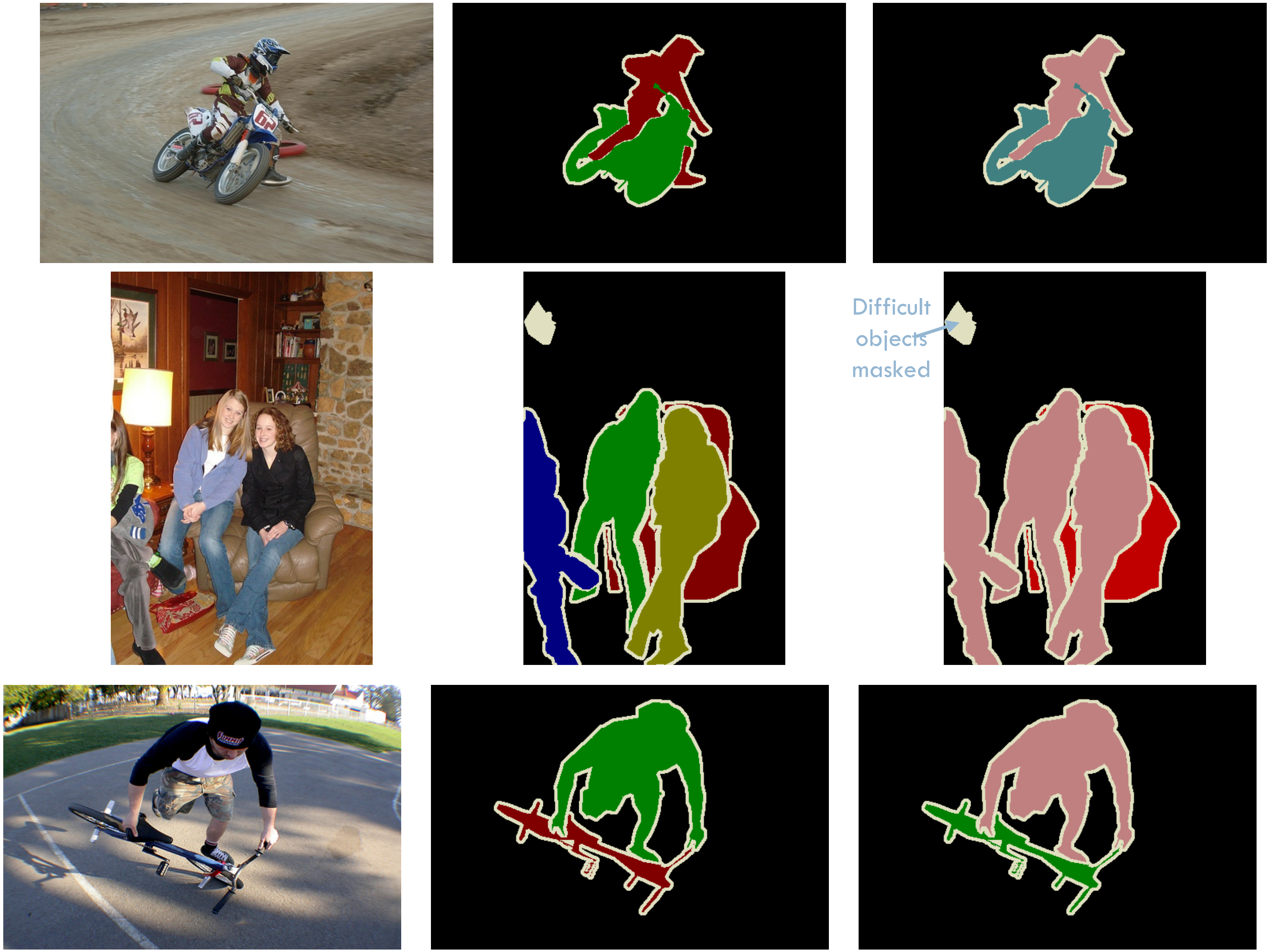
- Annotation in one session with <u>written guidelines</u>
  - Segmentation is ‘refinement’ of bounding box (but may go outside it)
  - Segmentation accurate to within 5-pixel boundary region which is marked ‘void’



  - 1-pixel wide structures (whiskers, wires) can be ignored
  - Surface objects considered part of the object (e.g. items on a table)

Example Annotations

Image | **Object segmentation** | **Class segmentation**



Example Annotations

Image | **Object segmentation** | **Class segmentation**

Dataset Statistics

- Contains VOC2008/9 data as subsets
- Around 30% increase in size over VOC2009

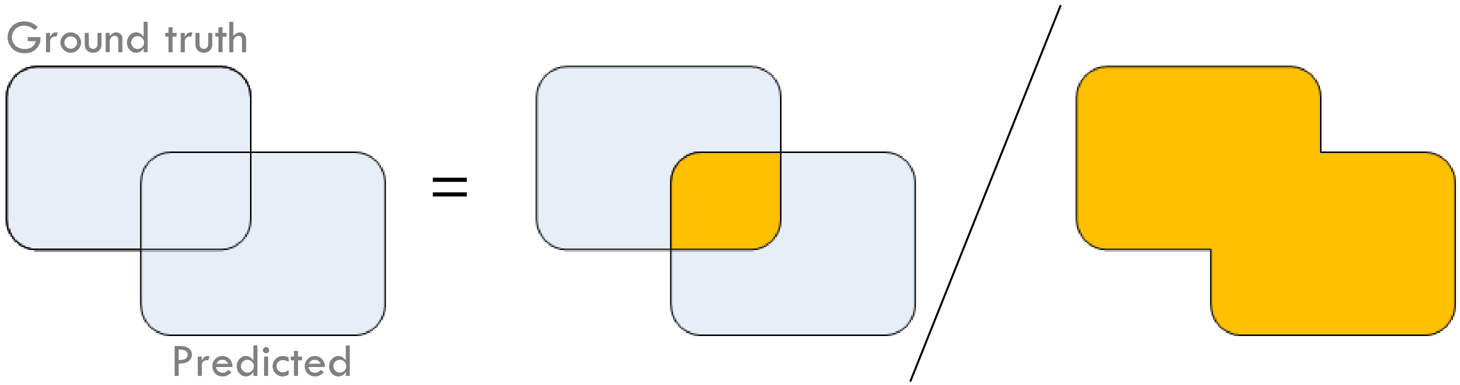
	Training		Testing	
Images	1,928	(1,499)	964	(750)
Objects	4,203	(3,211)	1,663	(1,202)

VOC2009 counts shown in brackets

- Almost 2,000 training and1,000 test images
- Over 4,000 precisely segmented objects for training

Evaluation Metric

$$\text{Intersection/union of } \textbf{class} \text{ labels} = \frac{\text{true pos. class}}{\text{true pos.} + \text{false pos.} + \text{false neg.}}$$



- Metric chosen because:
  - Allows per-class participation
  - Penalises both over- and under-estimates
- Overall evaluation metric is average over all classes (including background)

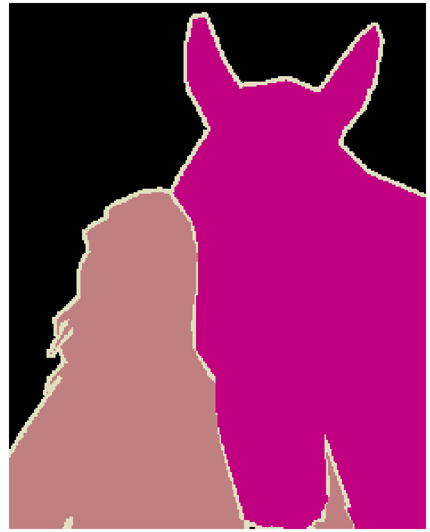
Methods

- 9 direct and 11 "automatic" entries
  – VOC2009: 12 direct, 10 "automatic"
- Methods
  - Multiple figure-ground segmentations
  - Hierarchical CRFs, higher order cliques
    - Co-occurrence of object class labels
    - Incorporation of object detectors as CRF potentials
  - Topic models for joint classification & segmentation
  - Refinement of object detections
    - Learnt segmentation masks for part-based models
    - Alignment of detections to bottom-up segmentation

Example Segmentations

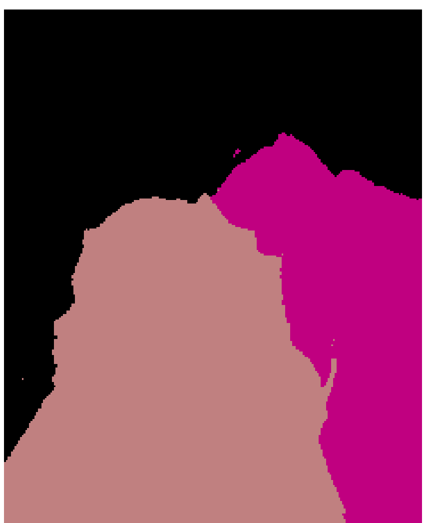
Image

Ground truth



BERKELEY_POSELETS_ALIGN_PB

BROOKES_AHCRF



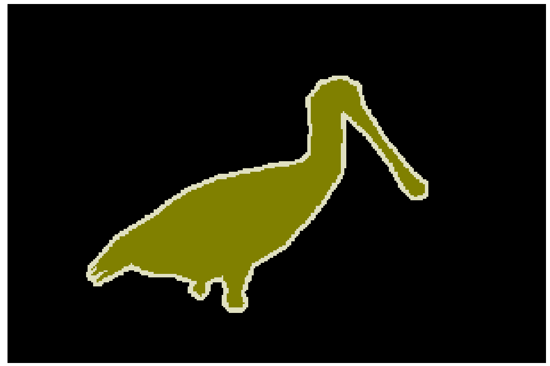
BONN_FGT_SEGM

CVC_HARMONY_DET

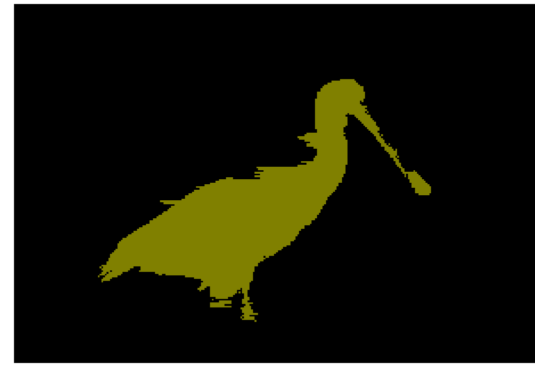
Example Segmentations

Image

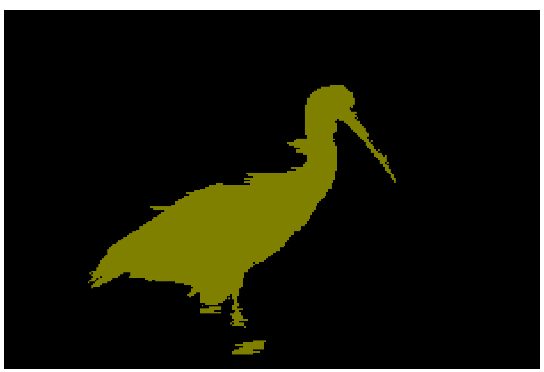
Ground truth



CVC_HARMONY



CVC_HARMONY_DET



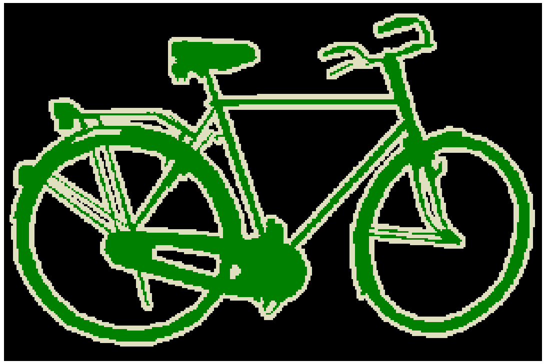
BERKELEY_POSELETS_ALIGN_PB

BROOKES_AHCRF

Example Segmentations

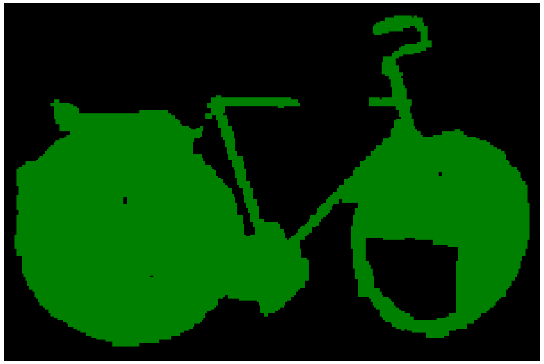
Image

Ground truth



BONN_SVR_SEGM

BONN_FGT_SEGM



BERKELEY_POSELETS_ALIGN_PB

UOCTTI_LSVM_MDPM

Example Segmentations

Image

Ground truth

BROOKES_AHCRF

CVC_HARMONY_DET

BONN_FGT_SEGM

BONN_SVR_SEGM

Example Segmentations

Image

Ground truth

BERKELEY_POSELETS_ALIGN_PB

UOCTTI_LSVM_MDPM

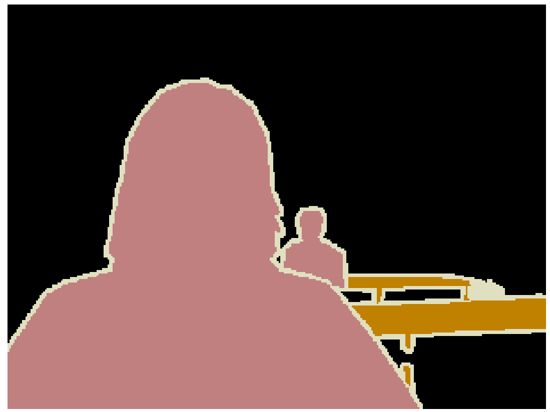
STANFORD_REGLABEL

CVC_HARMONY_DET

Example Segmentations

Image

Ground truth



STANFORD_REGLABEL

BERKELEY_POSELETS_ALIGN_PB

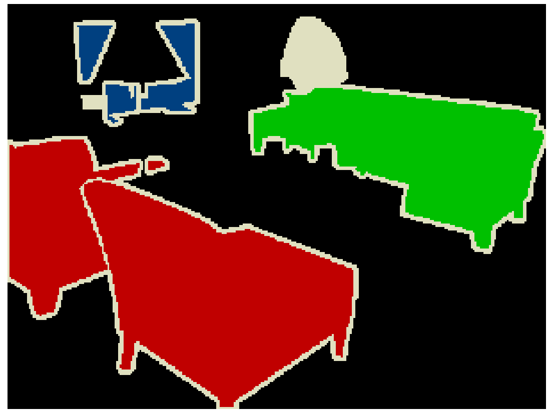
CVC_HARMONY

UOCTTI_LSVM_MDPM

Example Segmentations

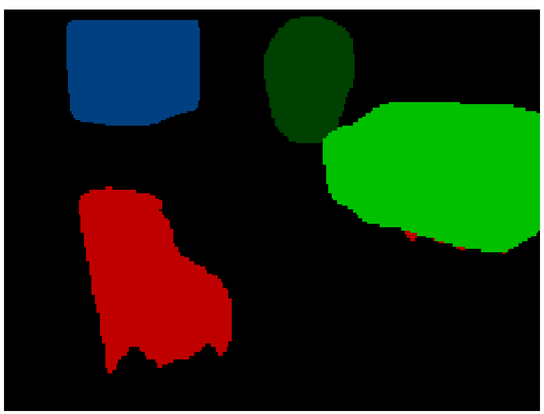
Image

Ground truth

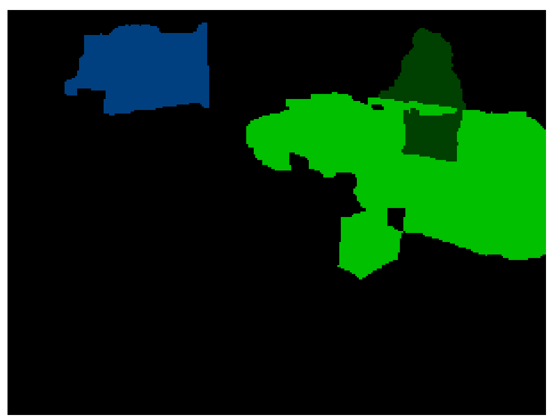


BROOKES_AHCRF

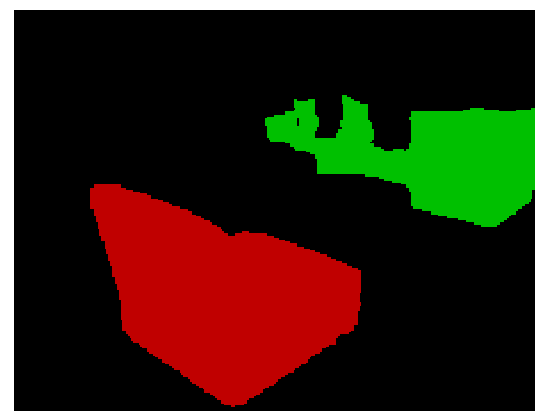
UOCTTI_LSVM_MDPM



BERKELEY_POSELETS_ALIGN_PB



BONN_SVR_SEGM



Accuracy by Class/Method

Trained on VOC2010 data

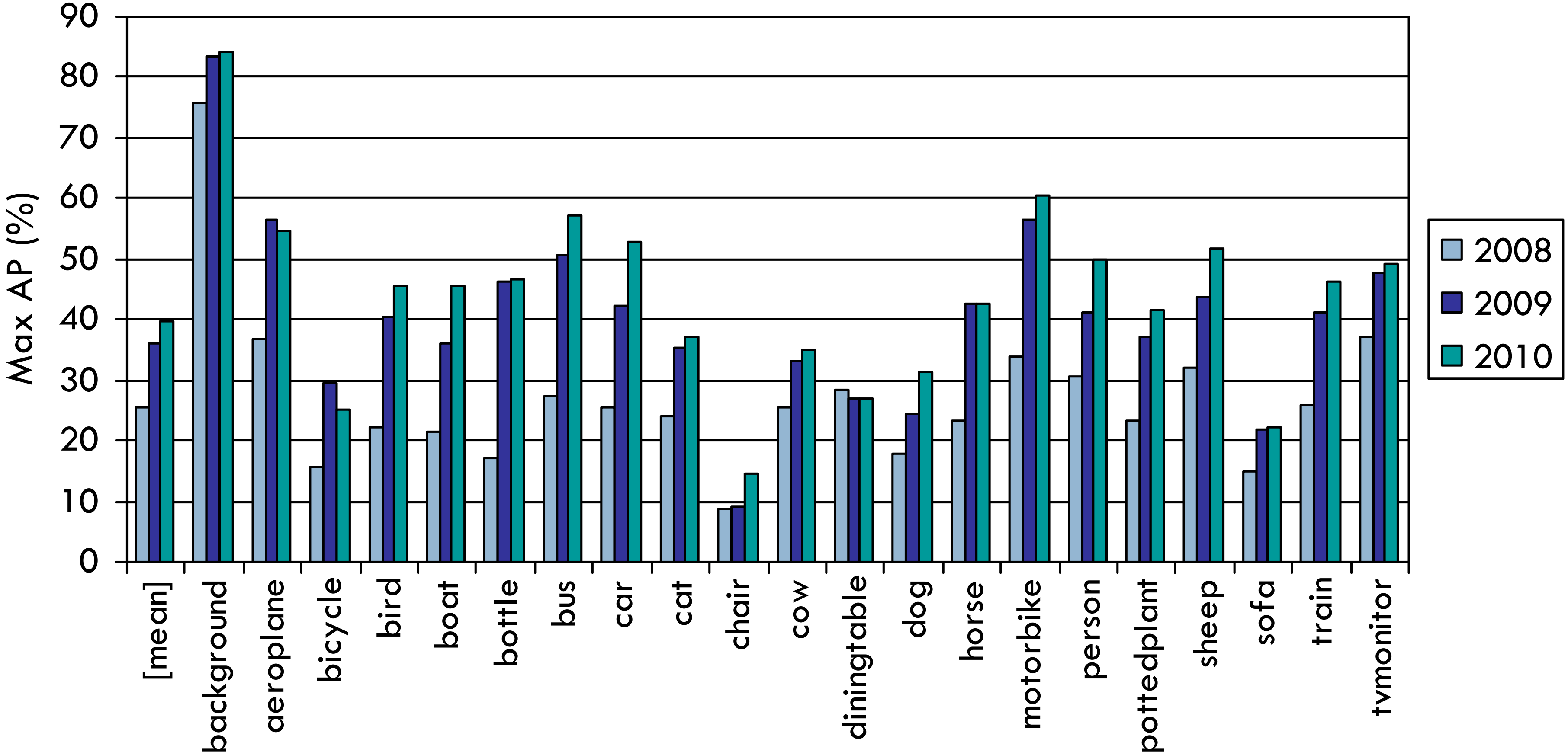
	[mean]	back ground	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
BONN_FGT_SEGM	36.5	82.5	54.6	22.5	25.1	27.6	40.0	60.2	48.3	39.4	7.3	30.8	21.3	25.3	34.9	54.1	36.6	22.5	45	17.6	33.5	37.0
BONN_SVR_SEGM	39.7	84.2	52.5	27.4	32.3	34.5	47.4	60.6	54.8	42.6	9.0	32.9	25.2	27.1	32.4	47.1	38.3	36.8	50.3	21.9	35.2	40.9
BROOKES_AHCRF	30.3	70.1	31.0	18.8	19.5	23.9	31.3	53.5	45.3	24.4	8.2	31.0	16.4	15.8	27.3	48.1	31.1	31.0	27.5	19.8	34.8	26.4
CVC_HARMONY	35.4	80.8	56.7	20.6	31.0	33.9	20.8	57.6	51.4	35.8	7.1	28.1	22.6	24.3	29.3	49.4	37.8	23.3	37.6	18.1	45.6	30.7
CVC_HARMONY_DET	40.1	81.1	58.3	23.1	39.0	37.8	36.4	63.2	62.4	31.9	9.1	36.8	24.6	29.4	37.5	60.6	44.9	30.1	36.8	19.4	44.1	35.9
STANFORD_REGLABEL	29.1	80.0	38.8	21.5	13.6	9.2	31.1	51.8	44.4	25.7	6.7	26.0	12.5	12.8	31.0	41.9	44.4	5.7	37.5	10.0	33.2	32.3
UC3M_GENDISC	27.8	73.4	45.9	12.3	14.5	22.3	9.3	46.8	38.3	41.7	0.0	35.9	20.7	34.1	34.8	33.5	24.6	4.7	25.6	13.0	26.8	26.1
UOCTTI_LSVM_MDPM	31.8	80.0	36.7	23.9	20.9	18.8	41.0	62.7	49.0	21.5	8.3	21.1	7.0	16.4	28.2	42.5	40.5	19.6	33.6	13.3	34.1	48.5

Trained on external data

	[mean]	back ground	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
BERKELEY_POSELETS	34.7	82.0	49.7	23.3	20.6	19.0	47.1	58.1	53.6	32.5	0.0	31.1	0.0	29.5	42.9	41.9	43.8	16.6	39.0	18.4	38.0	41.5

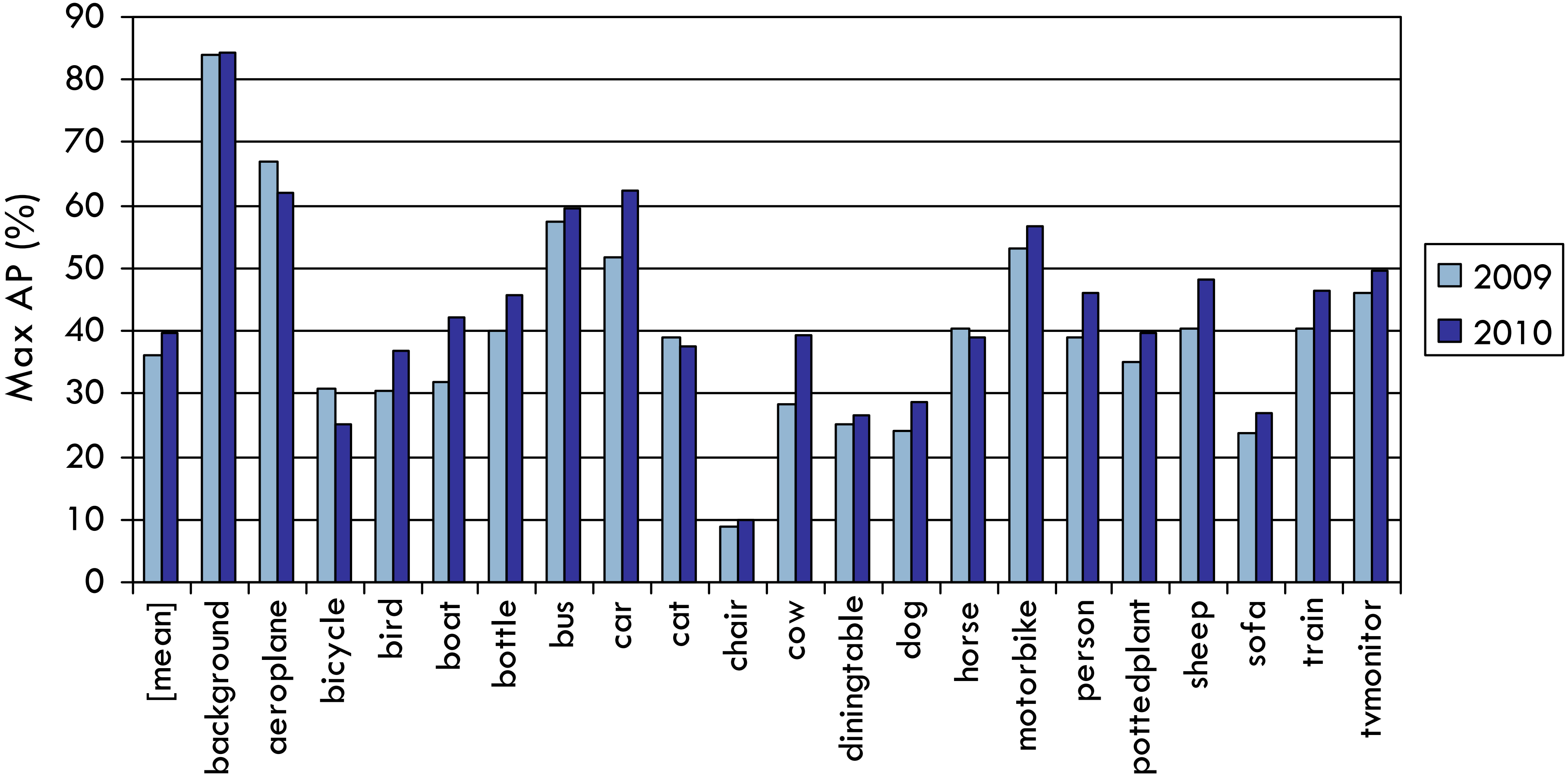
- Best results exceed best detection-based results for all classes
- BERKELEY_POSELETS method uses additional training annotation for object detection: improves on "horse"

Progress 2008-2010



- Results on 2008 data improve for best 2009 and 2010 methods for mean and 17/21 classes
  - Caveat: Better methods or more training data?

Progress 2009-2010



- Best 2010 methods improve on 2009 mean and for 16/21 categories
  - Caveat: Better methods or more training data?

Prizes

- Joint Winners:
  - **CVC_HARMONY_DET**
    Josep Maria Gonfaus, Xavier Boix,
    Fahad Kahn, Joost van de Weijer,
    Andrew Bagdanov, Marco Pedersoli,
    Joan Serrat, Xavier Roca, Jordi Gonzàlez
    Computer Vision Center, Universitat Autònoma de Barcelona
  - **BONN_SVR_SEGM**
    João Carreira, Fuxin Li, Cristian Sminchisescu
    University of Bonn
- Honourable Mention:
  - **BERKELEY_POSELETS_ALIGN_PB**
    Thomas Brox, Lubomir Bourdev, Subhransu Maji,
    Jitendra Malik
    University of California, Berkeley