# The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

# 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
  - New for VOC 2009
  - 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'



**Background    Void    Object**

5 pixels
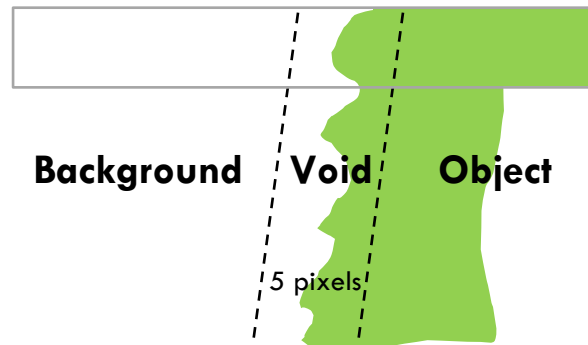
  - 1-pixel wide structures (whiskers, wires) can be ignored

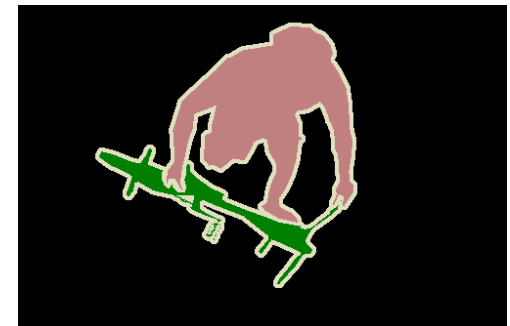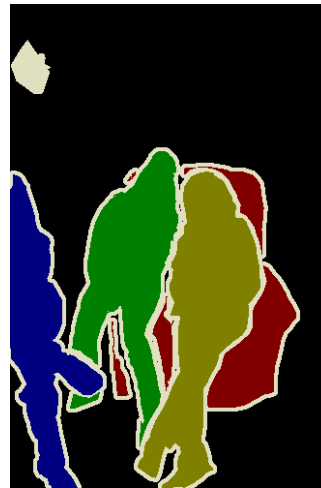  - Surface objects considered part of the object (e.g. items on a table)

# Example annotations
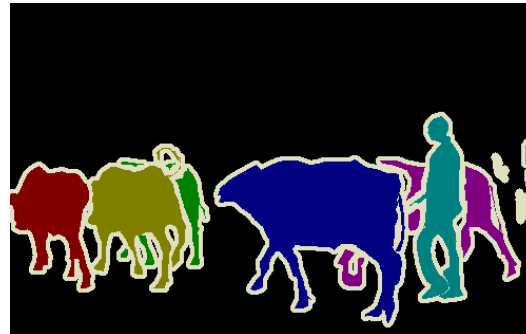
**Image**  **Object segmentation**  **Class segmentation**
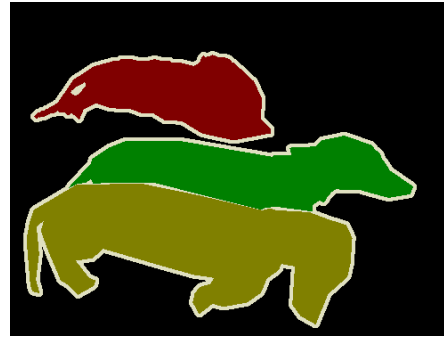


Difficult objects masked

# Example annotations

**Image**          **Object segmentation**          **Class segmentation**
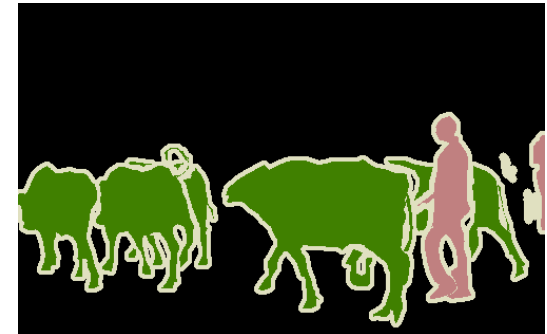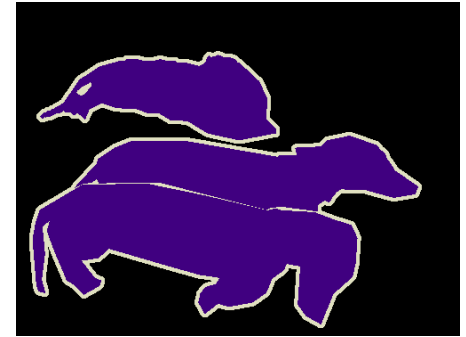
# Training/validation data sets

- 2009 data training and validation sets include and extend 2008 data set – allows comparison

- We now have over 3,200 precisely segmented objects available for training (~50% more than last year):

| Number of: | Training | Validation | Total |
|---|---|---|---|
| **Images** | 749 (511) | 750 (512) | **1499** (1023) |
| **Objects** | 1601 (1166) | 1610 (1203) | **3211** (2369) |

VOC 2008 counts shown in brackets

# Evaluation metric

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

Ground truth

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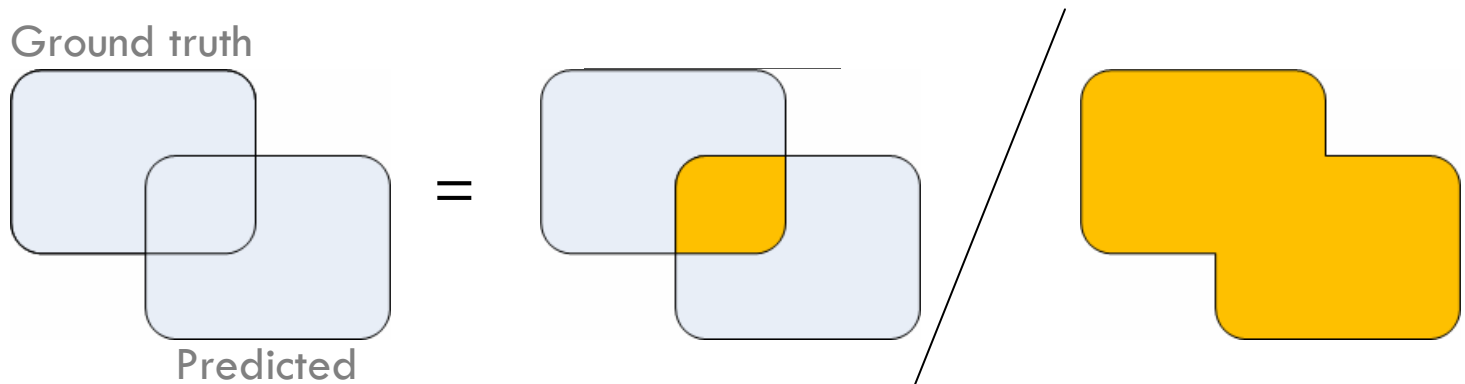
- Metric chosen because:
  - Allows per-class participation
  - Penalises both over- and under-estimates
- Overall evaluation metric is average over all classes (including background)

# Methods

- 12 direct and 10 'automatic' entries – double last year!! (6 direct, 5 automatic)
- Features:
  - SIFT, RGB, Textons, randomized forests
- Methods:
  - Multiple segmentations, superpixels
  - Hierarchical CRFs, high order cliques
  - Combined with classification/detection entries
  - Instance level part models and masks
  - 3D information

# Example segmentations



Image      Ground truth      Bonn

CVC      UCI      UoCTTI

# Example segmentations



Image     Ground truth     Bonn
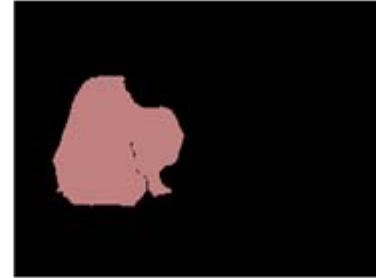
CVC     UCI     UoCTTI

# Example segmentations

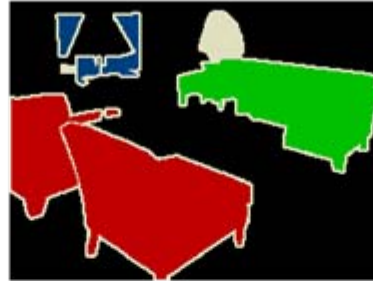# Example segmentations



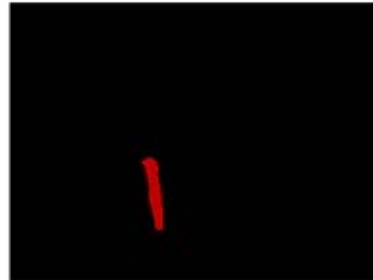Image     Ground truth     Bonn

CVC     UCI     UoCTTI

# Example segmentations
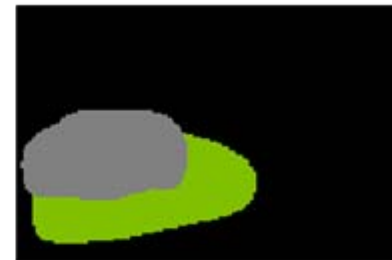


Image     Ground truth     Bonn

CVC     UCI     UoCTTI

# Results

**Comp 5:** All 12 segmentation entries beat all 11 automatic detection entries (not shown)

| | 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* | 36.3 | 83.9 | 64.3 | 21.8 | 21.7 | 32.0 | 40.2 | 57.3 | 49.4 | 38.8 | 5.2 | 28.5 | 22.0 | 19.6 | 33.6 | 45.5 | 33.6 | 27.3 | 40.4 | 18.1 | 33.6 | 46.1 |
| *BrookesMSRC* | 24.8 | 79.6 | 48.3 | 6.7 | 19.1 | 10.0 | 16.6 | 32.7 | 38.1 | 25.3 | 5.5 | 9.4 | 25.1 | 13.3 | 12.3 | 35.5 | 20.7 | 13.4 | 17.1 | 18.4 | 37.5 | 36.4 |
| *CVC* | 34.5 | 80.2 | 67.1 | 26.6 | 30.3 | 31.6 | 30.0 | 44.5 | 41.6 | 25.2 | 5.9 | 27.8 | 11.0 | 23.1 | 40.5 | 53.2 | 32.0 | 22.2 | 37.4 | 23.6 | 40.3 | 30.2 |
| *LEAR* | 25.7 | 79.1 | 44.6 | 15.5 | 20.5 | 13.3 | 28.8 | 29.3 | 35.8 | 25.4 | 4.4 | 20.3 | 1.3 | 16.4 | 28.2 | 30.0 | 24.5 | 12.2 | 31.5 | 18.3 | 28.8 | 31.9 |
| *MPI* | 15.0 | 70.9 | 16.4 | 8.7 | 8.6 | 8.3 | 20.8 | 21.6 | 14.4 | 10.5 | 0.0 | 14.2 | 17.2 | 7.3 | 9.3 | 20.3 | 18.2 | 6.9 | 14.1 | 0.0 | 13.2 | 13.2 |
| *NEC-UIUC* | 29.7 | 81.8 | 41.9 | 23.1 | 22.4 | 22.0 | 27.8 | 43.2 | 51.8 | 25.9 | 4.5 | 18.5 | 18.0 | 23.5 | 26.9 | 36.6 | 34.8 | 8.8 | 28.3 | 14.0 | 35.5 | 34.7 |
| *NEC-UIUC2* | 28.3 | 81.5 | 39.3 | 20.9 | 22.6 | 21.7 | 26.1 | 37.1 | 51.5 | 25.2 | 5.7 | 17.5 | 15.7 | 24.2 | 27.4 | 35.3 | 33.0 | 7.9 | 23.4 | 12.5 | 32.1 | 33.3 |
| *UC3M* | 14.5 | 69.8 | 20.8 | 9.7 | 6.3 | 4.3 | 7.9 | 19.7 | 21.8 | 7.7 | 3.8 | 7.5 | 9.6 | 9.5 | 12.3 | 16.5 | 16.4 | 1.5 | 14.2 | 11.0 | 14.1 | 20.3 |
| *UCI* | 24.7 | 80.7 | 38.3 | 30.9 | 3.4 | 4.4 | 31.7 | 45.5 | 47.3 | 10.4 | 4.8 | 14.3 | 8.8 | 6.1 | 21.5 | 25.0 | 38.9 | 14.8 | 14.4 | 3.0 | 29.1 | 45.5 |
| *UCLA* | 13.8 | 51.2 | 13.9 | 7.0 | 3.9 | 6.4 | 8.1 | 14.4 | 24.3 | 12.1 | 6.4 | 10.3 | 14.5 | 6.7 | 9.7 | 23.6 | 20.0 | 2.3 | 12.6 | 12.3 | 17.0 | 13.2 |
| *UoCTTI* | 29.0 | 78.9 | 35.3 | 22.5 | 19.1 | 23.5 | 36.2 | 41.2 | 50.1 | 11.7 | 8.9 | 28.5 | 1.4 | 5.9 | 24.0 | 35.3 | 33.4 | 35.1 | 27.7 | 14.2 | 34.1 | 41.8 |

**Comp 6:** 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* | | 78.5 | | | | | | | | | | | | | | | 36.3 | | | | | |
| *BrookesMSRC* | 24.5 | 79.6 | 40.1 | 9.0 | 17.6 | 1.5 | 20.6 | 34.9 | 29.4 | 24.1 | 6.1 | 13.8 | 28.3 | 13.3 | 9.3 | 31.1 | 23.0 | 17.1 | 18.0 | 24.7 | 36.1 | 37.5 |

# Comparison on VOC 2008 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* | **36.2** | **83.5** | 52.8 | 22.4 | 20.7 | 35.8 | **46.1** | **50.5** | 39.9 | **35.3** | 6.1 | **33.1** | 25.2 | 19.7 | **42.7** | 50.6 | 36.8 | 23.5 | **43.6** | 16.7 | 26.8 | **47.8** |
| *CVC* | 34.8 | 79.4 | **56.3** | 26.6 | **40.6** | **36.1** | 27.3 | 48.4 | 37.9 | 23.4 | **9.1** | 21.4 | 10.1 | **24.5** | 41.2 | **56.4** | 32.8 | 26.8 | 39.2 | **21.9** | **41.0** | 31.1 |
| *XRCE 2008* | 25.4 | 75.9 | 25.8 | 15.7 | 19.2 | 21.6 | 17.2 | 27.3 | 25.5 | 24.2 | 7.9 | 25.4 | 9.9 | 17.8 | 23.3 | 34.0 | 28.8 | 23.2 | 32.1 | 14.9 | 25.9 | 37.3 |
| *Brookes 2008* | 20.1 | 75.0 | 36.9 | 4.8 | 22.2 | 11.2 | 13.7 | 13.8 | 20.4 | 10.0 | 8.7 | 3.6 | 28.3 | 6.6 | 17.1 | 22.6 | 30.6 | 13.5 | 26.8 | 12.1 | 20.1 | 24.8 |

- New methods improved significantly on 2008 performance *but* had access to ~50% more training data
- Some new method beat every 2008 method on each class
- Bonn beat XRCE (2008 winner) on 20/21 classes

# Prizes

- ## Winner: Bonn
  João Carreira, Fuxin Li, Cristian Sminchisescu
  *University of Bonn*

- ## Runner up: CVC
  Xavier Boix, Josep Maria Gonfaus, Fahad Kahn,
  Joost van de Weijer, Andrew Bagdanov, Marco Pedersoli,
  Jordi González, Joan Serrat
  *Computer Vision Center Barcelona*

**Congratulations!!**

Many thanks to all teams for their hard work in taking part. The segmentation competition is now firmly established as a VOC challenge.