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 *written guidelines*
  - 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

<table>
<thead>
<tr>
<th>Image</th>
<th>Object segmentation</th>
<th>Class segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="object1.png" alt="Object segmentation 1" /></td>
<td><img src="class1.png" alt="Class segmentation 1" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="object2.png" alt="Object segmentation 2" /></td>
<td><img src="class2.png" alt="Class segmentation 2" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image 3" /></td>
<td><img src="object3.png" alt="Object segmentation 3" /></td>
<td><img src="class3.png" alt="Class segmentation 3" /></td>
</tr>
</tbody>
</table>
Example Annotations

<table>
<thead>
<tr>
<th>Image</th>
<th>Object segmentation</th>
<th>Class segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="object_segmentation1.png" alt="Object segmentation 1" /></td>
<td><img src="class_segmentation1.png" alt="Class segmentation 1" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="object_segmentation2.png" alt="Object segmentation 2" /></td>
<td><img src="class_segmentation2.png" alt="Class segmentation 2" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image 3" /></td>
<td><img src="object_segmentation3.png" alt="Object segmentation 3" /></td>
<td><img src="class_segmentation3.png" alt="Class segmentation 3" /></td>
</tr>
</tbody>
</table>
Dataset Statistics

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

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>1,928</td>
<td>964</td>
</tr>
<tr>
<td>Objects</td>
<td>4,203</td>
<td>1,663</td>
</tr>
</tbody>
</table>

VOC2009 counts shown in brackets

- Almost 2,000 training and 1,000 test images
- Over 4,000 precisely segmented objects for training
Evaluation Metric

Intersection/union of class labels

\[ \text{Intersection/union} = \frac{\text{true pos. class}}{\text{true pos.} + \text{false pos.} + \text{false neg.}} \]

- 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

<table>
<thead>
<tr>
<th>Image</th>
<th>Ground truth</th>
<th>BERKELEY_POSELETS_ALIGN_PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROOKES_AHCRF</td>
<td>BONN_FGT_SEGM</td>
<td>CVC_HARMONY_DET</td>
</tr>
</tbody>
</table>
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_LSV_MDPM
Example Segmentations
Example Segmentations
Example Segmentations

Image

Ground truth

STANFORD_REGLABEL

BERKELEY_POSELETS_ALIGN_PB

CVC_HARMONY

UOCTTI_L SVM_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**

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

**Trained on external data**

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

- 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?
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