The PASCAL Visual Object Classes Challenge 2011 (VOC2011)

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’

![Diagram showing segmentation with 5-pixel boundary]

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

Image

Object segmentation

Class segmentation
Dataset Statistics

- Contains VOC2008-10 data as subsets
- Around 20% increase in size over VOC2010

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>2,223</td>
<td>(1,928)</td>
</tr>
<tr>
<td></td>
<td>1,111</td>
<td>(964)</td>
</tr>
<tr>
<td>Objects</td>
<td>5,034</td>
<td>(4,203)</td>
</tr>
<tr>
<td></td>
<td>2,028</td>
<td>(1,663)</td>
</tr>
</tbody>
</table>

VOC2010 counts shown in brackets

- Over 2,000 training and 1,000 test images
- Over 5,000 precisely segmented objects for training
Evaluation Metric

Intersection/union of class 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)
6 direct and 8 “automatic” entries

**Methods**

- Multiple figure-ground segmentations
  - Object overlap prediction using Support Vector Regression
- Hierarchical CRFs, higher order cliques
  - Joint segmentation and detection
- Low level segmentation + region classification
  - Ultrametric contour maps
- Refinement of object detections
  - Learnt segmentation masks for part-based models
Example Segmentations

Image

Ground truth

BERKELEY_REGION_CLASSIFY

BROOKES_STRUCT_DET_CRT

BONN_SVR_SEGM

NUS_SEG_DET_MASK_CLS_CRF
Example Segmentations

Image | Ground truth | BERKELEY_REGION_CLASSIFY

BROOKES_STRUCT_DET_CRT | BONN_SVR_SEGM | NUS_SEG_DET_MASK_CLS_CRF
Example Segmentations

<table>
<thead>
<tr>
<th>Image</th>
<th>Ground truth</th>
<th>BERKELEY_REGION_CLASSIFY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROOKES_STRUCT_DET_CRT</td>
<td>BONN_SVR_SEGM</td>
<td>NUS_SEG_DET_MASK_CLS_CRF</td>
</tr>
</tbody>
</table>

[Image of a seagull and its segmentations]
Example Segmentations

Image

Ground truth

BERKELEY_REGION_CLASSIFY

BROOKES_STRUCT_DET_CRT

BONN_SVR_SEGM

NUS_SEG_DET_MASK_CLS_CRF
**Accuracy by Class/Method**

(1st, 2nd, 3rd place)

Trained on VOC2011 data

<table>
<thead>
<tr>
<th>Method</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_SEGM</td>
<td>41.4</td>
<td>83.4</td>
<td>51.7</td>
<td>23.7</td>
<td>46.0</td>
<td>33.9</td>
<td>49.4</td>
<td>66.2</td>
<td>56.2</td>
<td>41.7</td>
<td>10.4</td>
<td>41.9</td>
<td>29.6</td>
<td>24.4</td>
<td>49.1</td>
<td>50.5</td>
<td>39.6</td>
<td>19.9</td>
<td>44.9</td>
<td>26.1</td>
<td>40.0</td>
<td>41.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BONN_SVR_SEGM</td>
<td>43.3</td>
<td>84.9</td>
<td>54.3</td>
<td>23.9</td>
<td>39.5</td>
<td>35.3</td>
<td>42.6</td>
<td>65.4</td>
<td>53.3</td>
<td>46.1</td>
<td>15.0</td>
<td>47.4</td>
<td>30.1</td>
<td>33.9</td>
<td>48.8</td>
<td>54.4</td>
<td>46.4</td>
<td>28.8</td>
<td>51.3</td>
<td>26.2</td>
<td>44.9</td>
<td>37.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BROOKES_STRUCT_DET_CRT</td>
<td>31.3</td>
<td>79.4</td>
<td>36.6</td>
<td>18.6</td>
<td>9.2</td>
<td>11.0</td>
<td>29.8</td>
<td>59.0</td>
<td>50.3</td>
<td>25.5</td>
<td>11.8</td>
<td>29.0</td>
<td>24.8</td>
<td>16.0</td>
<td>29.1</td>
<td>47.9</td>
<td>41.9</td>
<td>16.1</td>
<td>34.0</td>
<td>11.6</td>
<td>43.3</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUS_CONTEXT_SVM</td>
<td>35.1</td>
<td>77.2</td>
<td>40.5</td>
<td>19.0</td>
<td>28.4</td>
<td>27.8</td>
<td>40.7</td>
<td>56.4</td>
<td>45.0</td>
<td>33.1</td>
<td>7.2</td>
<td>37.4</td>
<td>17.4</td>
<td>26.8</td>
<td>33.7</td>
<td>46.6</td>
<td>40.6</td>
<td>23.3</td>
<td>33.4</td>
<td>23.9</td>
<td>41.2</td>
<td>38.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUS_SEG_DET_MASK_CLS_CRF</td>
<td>37.7</td>
<td>79.8</td>
<td>41.5</td>
<td>20.2</td>
<td>30.4</td>
<td>29.1</td>
<td>47.4</td>
<td>61.2</td>
<td>47.7</td>
<td>35.0</td>
<td>8.5</td>
<td>38.3</td>
<td>14.5</td>
<td>28.6</td>
<td>36.5</td>
<td>47.8</td>
<td>42.5</td>
<td>28.5</td>
<td>37.8</td>
<td>26.4</td>
<td>43.5</td>
<td>45.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Trained on external data

<table>
<thead>
<tr>
<th>Method</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_REGION_CLASSIFY</td>
<td>39.1</td>
<td>83.3</td>
<td>48.9</td>
<td>20.0</td>
<td>32.8</td>
<td>28.2</td>
<td>41.1</td>
<td>53.9</td>
<td>48.3</td>
<td>48.0</td>
<td>6.0</td>
<td>34.9</td>
<td>27.5</td>
<td>35.0</td>
<td>47.2</td>
<td>47.3</td>
<td>48.4</td>
<td>20.6</td>
<td>52.7</td>
<td>25.0</td>
<td>36.6</td>
<td>35.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Best results exceed best detection-based results for all classes
- **BONN_SVR_SEGM**: 1<sup>st</sup> in 11 categories, 2<sup>nd</sup> in 9 categories
- **BERKELEY_REGION_CLASSIFY**: 1<sup>st</sup> in 4 categories using own data
Results on 2008 data improve for best methods 2009-2011 for mean and most categories

- Caveats: More training data + re-use of test data
Progress 2009-2011

- Results on 2009 data improve for best methods 2010-2011 for mean and most categories
  - Caveats: More training data + re-use of test data
Results on 2010 data improve for best 2011 methods for mean and 11/21 categories

- Caveats: More training data + re-use of test data
Prizes

- **Winner:**
  - **BONN_SVR_SEGM/BONN_FGT_SEGM**
    - João Carreira¹, Adrian Ion², Fuxin Li³, Cristian Sminchisescu¹
    - ¹University of Bonn, ²Vienna University of Technology, ³Georgia Institute of Technology

- **Honourable Mention:**
  - **BERKELEY_REGION_CLASSIFY**
    - Pablo Arbelaez, Bharath Hariharan, Saurabh Gupta, Chunhui Gu, Lubomir Bourdev and Jitendra Malik
    - University of California, Berkeley