The PASCAL Visual Object Classes Challenge 2008 (VOC2008)

Part 3 – Segmentation Taster

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman
Segmentation Taster

- 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?

- [Competition 6: Train on any (non-test) data]
  - Not an official competition!
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 and void regions]

- 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.jpg" alt="Image 1" /></td>
<td><img src="object1.jpg" alt="Object segmentation 1" /></td>
<td><img src="class1.jpg" alt="Class segmentation 1" /></td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Image 2" /></td>
<td><img src="object2.jpg" alt="Object segmentation 2" /></td>
<td><img src="class2.jpg" alt="Class segmentation 2" /></td>
</tr>
<tr>
<td><img src="image3.jpg" alt="Image 3" /></td>
<td><img src="object3.jpg" alt="Object segmentation 3" /></td>
<td><img src="class3.jpg" alt="Class segmentation 3" /></td>
</tr>
</tbody>
</table>

Difficult objects masked
Example annotations

<table>
<thead>
<tr>
<th>Image</th>
<th>Object segmentation</th>
<th>Class segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="example_image1.jpg" alt="Image" /></td>
<td><img src="example_object1.jpg" alt="Object segmentation" /></td>
<td><img src="example_class1.jpg" alt="Class segmentation" /></td>
</tr>
<tr>
<td><img src="example_image2.jpg" alt="Image" /></td>
<td><img src="example_object2.jpg" alt="Object segmentation" /></td>
<td><img src="example_class2.jpg" alt="Class segmentation" /></td>
</tr>
<tr>
<td><img src="example_image3.jpg" alt="Image" /></td>
<td><img src="example_object3.jpg" alt="Object segmentation" /></td>
<td><img src="example_class3.jpg" alt="Class segmentation" /></td>
</tr>
</tbody>
</table>
Training/validation data sets

- 2008 data training and validation sets include and extend 2007 data set
- More than double the number of annotated images:

<table>
<thead>
<tr>
<th>Number of:</th>
<th>Training</th>
<th>Validation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>511 (209)</td>
<td>512 (213)</td>
<td>1023 (422)</td>
</tr>
<tr>
<td>Objects</td>
<td>1166 (633)</td>
<td>1203 (582)</td>
<td>2369 (1215)</td>
</tr>
</tbody>
</table>

VOC 2007 totals shown in brackets
New evaluation metric for VOC 2008

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)**
Methods

- 6 direct and 5 ‘automatic’ entries using a variety of methods

- Features:
  - SIFT, RGB, Textons, randomized forests

- Segmentation method:
  - Bottom-up oversegmentation (superpixels)
  - MRF approaches (including high order cliques)
  - Refinement of detections (also provides instance labels)

- Use of image-level classifiers
## Results

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>background</th>
<th>aereoplane</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 table</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>potted plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv/monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brookes/MSRC</td>
<td>20.1</td>
<td>75.0</td>
<td><strong>36.9</strong></td>
<td>4.8</td>
<td>11.2</td>
<td>13.7</td>
<td>13.8</td>
<td>10.0</td>
<td>8.7</td>
<td>3.6</td>
<td><strong>28.3</strong></td>
<td>6.6</td>
<td>17.1</td>
<td>22.6</td>
<td><strong>30.6</strong></td>
<td></td>
<td>13.5</td>
<td>26.8</td>
<td>12.1</td>
<td>20.1</td>
<td>24.8</td>
<td></td>
</tr>
<tr>
<td>CASIA_det*</td>
<td>13.5</td>
<td>44.7</td>
<td>8.5</td>
<td>9.4</td>
<td>10.7</td>
<td>7.5</td>
<td><strong>13.3</strong></td>
<td>37.3</td>
<td>10.3</td>
<td>14.4</td>
<td>2.0</td>
<td>5.4</td>
<td>4.0</td>
<td>8.0</td>
<td>9.0</td>
<td>15.9</td>
<td>10.5</td>
<td>8.2</td>
<td>16.8</td>
<td>4.1</td>
<td>21.6</td>
<td>22.1</td>
</tr>
<tr>
<td>Jena</td>
<td>8.0</td>
<td>47.8</td>
<td>7.2</td>
<td>3.1</td>
<td>4.6</td>
<td>5.6</td>
<td>2.2</td>
<td>0.6</td>
<td>13.4</td>
<td>0.0</td>
<td>0.7</td>
<td>7.5</td>
<td>0.7</td>
<td>5.7</td>
<td>4.4</td>
<td>8.9</td>
<td>8.7</td>
<td>5.0</td>
<td>9.2</td>
<td>3.4</td>
<td>12.2</td>
<td>17.8</td>
</tr>
<tr>
<td>LEAR PlusClass*</td>
<td>3.7</td>
<td>5.5</td>
<td>4.1</td>
<td>2.2</td>
<td>0.0</td>
<td>3.5</td>
<td>4.9</td>
<td>3.9</td>
<td>9.6</td>
<td>0.8</td>
<td>1.5</td>
<td>0.1</td>
<td>0.4</td>
<td>0.9</td>
<td>1.2</td>
<td>2.5</td>
<td>7.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>4.0</td>
<td>24.7</td>
</tr>
<tr>
<td>MPI_norank</td>
<td>7.0</td>
<td>66.3</td>
<td>6.7</td>
<td>1.2</td>
<td>2.1</td>
<td>3.1</td>
<td>2.5</td>
<td>5.8</td>
<td>2.6</td>
<td>2.9</td>
<td>1.1</td>
<td>1.7</td>
<td>4.0</td>
<td>2.6</td>
<td>3.7</td>
<td>5.1</td>
<td>10.5</td>
<td>0.8</td>
<td>5.8</td>
<td>2.1</td>
<td>8.4</td>
<td>8.0</td>
</tr>
<tr>
<td>MPI_single</td>
<td>12.9</td>
<td>75.4</td>
<td>19.1</td>
<td>7.7</td>
<td>6.1</td>
<td>9.4</td>
<td>3.8</td>
<td>11.0</td>
<td>12.1</td>
<td>5.6</td>
<td>0.7</td>
<td>3.7</td>
<td>15.9</td>
<td>3.6</td>
<td>12.2</td>
<td>16.1</td>
<td>15.9</td>
<td>0.6</td>
<td>19.7</td>
<td>5.9</td>
<td>14.7</td>
<td>12.5</td>
</tr>
<tr>
<td>MPI_struct*</td>
<td>12.9</td>
<td>59.2</td>
<td>15.1</td>
<td>5.5</td>
<td>8.2</td>
<td>19.1</td>
<td>8.6</td>
<td>13.0</td>
<td>11.0</td>
<td>11.9</td>
<td>6.8</td>
<td>0.0</td>
<td>3.8</td>
<td>11.0</td>
<td>14.3</td>
<td>14.9</td>
<td>12.1</td>
<td>0.1</td>
<td>7.2</td>
<td>5.1</td>
<td>19.6</td>
<td>23.5</td>
</tr>
<tr>
<td>UoCTTIUCI*</td>
<td>11.6</td>
<td>0.9</td>
<td>9.1</td>
<td>12.8</td>
<td>1.0</td>
<td>6.7</td>
<td>11.8</td>
<td>24.7</td>
<td>11.4</td>
<td>8.9</td>
<td>1.6</td>
<td>9.0</td>
<td>1.3</td>
<td>7.9</td>
<td>14.4</td>
<td>18.8</td>
<td>13.1</td>
<td>9.5</td>
<td>24.3</td>
<td>11.0</td>
<td>17.1</td>
<td>27.8</td>
</tr>
<tr>
<td>XRCE_Det*</td>
<td>18.9</td>
<td>64.8</td>
<td>11.1</td>
<td>12.7</td>
<td>15.4</td>
<td>10.9</td>
<td>7.8</td>
<td>17.7</td>
<td>18.4</td>
<td>19.4</td>
<td>5.6</td>
<td>14.0</td>
<td>8.5</td>
<td>16.8</td>
<td>16.0</td>
<td>30.9</td>
<td>25.3</td>
<td>19.1</td>
<td>29.7</td>
<td>5.5</td>
<td>18.0</td>
<td>29.0</td>
</tr>
<tr>
<td>XRCE_Seg</td>
<td><strong>25.4</strong></td>
<td><strong>75.9</strong></td>
<td><strong>25.8</strong></td>
<td><strong>15.7</strong></td>
<td><strong>19.2</strong></td>
<td><strong>21.6</strong></td>
<td><strong>17.2</strong></td>
<td><strong>27.3</strong></td>
<td><strong>25.5</strong></td>
<td><strong>24.2</strong></td>
<td><strong>7.9</strong></td>
<td><strong>25.4</strong></td>
<td><strong>9.9</strong></td>
<td><strong>17.8</strong></td>
<td><strong>23.3</strong></td>
<td><strong>34.0</strong></td>
<td><strong>28.8</strong></td>
<td><strong>23.2</strong></td>
<td><strong>32.1</strong></td>
<td><strong>14.9</strong></td>
<td><strong>25.9</strong></td>
<td><strong>37.3</strong></td>
</tr>
</tbody>
</table>

*Automatic entry from detection competition entry

### Trained on external data:

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>background</th>
<th>aereoplane</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 table</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>potted plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv/monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>UIUC_CMU</td>
<td>19.5</td>
<td><strong>79.3</strong></td>
<td><strong>31.9</strong></td>
<td><strong>21.0</strong></td>
<td>8.3</td>
<td>6.5</td>
<td><strong>34.3</strong></td>
<td>15.8</td>
<td>22.7</td>
<td>10.4</td>
<td>1.2</td>
<td>6.8</td>
<td>8.0</td>
<td>10.2</td>
<td>22.7</td>
<td>24.9</td>
<td>27.7</td>
<td>15.9</td>
<td>4.3</td>
<td>5.5</td>
<td>19.0</td>
<td>32.1</td>
</tr>
</tbody>
</table>
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

- Image
- GroundTruth
- Brookes MSRC
- MPI single
- UIUC CMU
- XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Example segmentations

Image

GroundTruth

Brookes MSRC

MPI single

UIUC CMU

XRCE
Layout Taster

- Given the bounding box of a person, predict the visibility and positions of head, hands and feet.
  - ~300 training examples
  - Please give it a try!
Prizes

- **Winner: XRCE**
  Gabriela Czurka, Florent Perronnin, Yan Liu
  Xerox Research Centre Europe (XRCE), Textual and Visual Pattern Analysis Group

- **Runner up: Brookes/MSRC**
  Lubor Ladicky¹, Phil Torr¹, Pushmeet Kohli²
  ¹Oxford Brookes University, ²Microsoft Research Cambridge

- **Honorable Mention: UIUC_CMU**
  Derek Hoiem¹, Santosh Divvala², James H. Hays²
  ¹University of Illinois Urbana-Champaign; ²Carnegie Mellon University