The PASCAL Visual Object Classes Challenge 2011 (VOC2011)

Part 2 – Detection Task

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning
Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)

- Competition 3: Train on the supplied data
  - Which methods perform best given specified training data?

- Competition 4: Train on any (non-test) data
  - How well do state-of-the-art methods perform on these problems?
Average Precision

- Interpolate curve to create version for which the precision is monotonically non-increasing
- Measure area under interpolated curve

- Sawtooth shape is ignored
- Area is measured with maximum accuracy
- Good score requires both high precision and recall
Evaluating Bounding Boxes

- **Area of Overlap (AO) Measure**

$$ AO(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt} \cup B_p|} $$

- Need to define a threshold $t$ such that $AO(B_{gt}, B_p)$ implies a correct detection: 50%
Methods

- 13 Methods, 10 Groups
  - VOC2010: 22 Methods, 15 Groups

Methods

- Sliding window, SVM, multiple features e.g. HOG, LBP, bag of words
- Classification-like representations: LLC, spatial pyramids, max pooling
- Parts-based models, hierarchical models
- Segmentation-based object hypotheses ("jumping window")
- Combination with whole-image classification
Methods

• Novelty
  • Use of explicit modelling of 3D structure
  • CRF for joint detection and segmentation inference
  • Representing part locations using GMM
  • Grammar model for person including occluder part
### AP by Class/Method

(1st, 2nd, 3rd place)

<table>
<thead>
<tr>
<th>Method</th>
<th>aeroplane</th>
<th>bicycle</th>
<th>boat</th>
<th>bottle</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>dining table</th>
<th>dog</th>
<th>horse</th>
<th>motor bike</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_STRUCT_DET_CRF</td>
<td>37.1</td>
<td>42.6</td>
<td>2.0</td>
<td>0.0</td>
<td>16.0</td>
<td>43.8</td>
<td>38.6</td>
<td>17.0</td>
<td>10.3</td>
<td>7.7</td>
<td>2.4</td>
<td>1.5</td>
<td>34.3</td>
<td>41.1</td>
<td>38.4</td>
<td>1.5</td>
<td>14.7</td>
<td>5.3</td>
</tr>
<tr>
<td>CMIC_GS_DPM</td>
<td>13.3</td>
<td>26.4</td>
<td>41.5</td>
<td>12.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41.6</td>
<td>8.3</td>
<td>31.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMIC_SYNTHDPM</td>
<td>40.4</td>
<td>47.8</td>
<td>11.4</td>
<td>23.7</td>
<td>48.9</td>
<td>40.9</td>
<td>23.5</td>
<td>11.9</td>
<td>25.5</td>
<td>10.9</td>
<td>42.0</td>
<td>38.6</td>
<td>40.7</td>
<td>7.5</td>
<td>30.4</td>
<td>38.4</td>
<td>34.8</td>
<td></td>
</tr>
<tr>
<td>CORNELL_ISVM_VIEWPOINT</td>
<td>42.5</td>
<td>43.7</td>
<td>5.4</td>
<td>4.8</td>
<td>18.1</td>
<td>28.6</td>
<td>36.6</td>
<td>24.2</td>
<td>12.6</td>
<td>20.5</td>
<td>4.4</td>
<td>17.5</td>
<td>15.2</td>
<td>38.2</td>
<td>7.9</td>
<td>1.7</td>
<td>23.2</td>
<td>7.1</td>
</tr>
<tr>
<td>MISSOURI_LCC_TREE_CODING</td>
<td>41.1</td>
<td>51.7</td>
<td>13.7</td>
<td>11.9</td>
<td>27.3</td>
<td>52.1</td>
<td>41.7</td>
<td>32.9</td>
<td>17.6</td>
<td>27.3</td>
<td>18.5</td>
<td>23.1</td>
<td>45.2</td>
<td>48.6</td>
<td>41.9</td>
<td>11.6</td>
<td>32.4</td>
<td>27.5</td>
</tr>
<tr>
<td>MISSOURI_TREE_MAX_POOLING</td>
<td>43.8</td>
<td>51.7</td>
<td>13.7</td>
<td>12.7</td>
<td>27.3</td>
<td>51.5</td>
<td>43.7</td>
<td>32.9</td>
<td>18.3</td>
<td>27.3</td>
<td>18.5</td>
<td>23.1</td>
<td>45.2</td>
<td>48.6</td>
<td>42.9</td>
<td>11.6</td>
<td>32.4</td>
<td>27.5</td>
</tr>
<tr>
<td>NLP_DD_DC</td>
<td>55.0</td>
<td>58.1</td>
<td>22.5</td>
<td>18.8</td>
<td>33.9</td>
<td>57.6</td>
<td>54.5</td>
<td>42.6</td>
<td>20.2</td>
<td>40.3</td>
<td>29.3</td>
<td>37.1</td>
<td>54.6</td>
<td>58.3</td>
<td>51.6</td>
<td>14.7</td>
<td>44.8</td>
<td>32.1</td>
</tr>
<tr>
<td>NUS_CONTEXT_SVM</td>
<td>51.4</td>
<td>52.9</td>
<td>20.1</td>
<td>15.7</td>
<td>26.9</td>
<td>53.0</td>
<td>45.6</td>
<td>37.6</td>
<td>15.2</td>
<td>36.0</td>
<td>25.1</td>
<td>32.6</td>
<td>50.4</td>
<td>55.8</td>
<td>36.8</td>
<td>12.3</td>
<td>37.6</td>
<td>30.5</td>
</tr>
<tr>
<td>NYU_UCLA_HIERARCHY</td>
<td>56.3</td>
<td>55.9</td>
<td>23.4</td>
<td>20.3</td>
<td>27.2</td>
<td>56.6</td>
<td>48.1</td>
<td>53.8</td>
<td>23.2</td>
<td>32.9</td>
<td>33.3</td>
<td>39.2</td>
<td>53.0</td>
<td>56.9</td>
<td>43.6</td>
<td>14.3</td>
<td>37.9</td>
<td>39.4</td>
</tr>
<tr>
<td>OXFORD_DPM_MK</td>
<td>56.0</td>
<td>53.3</td>
<td>19.2</td>
<td>17.2</td>
<td>25.8</td>
<td>53.1</td>
<td>45.4</td>
<td>44.5</td>
<td>20.1</td>
<td>32.1</td>
<td>37.2</td>
<td>52.3</td>
<td>56.6</td>
<td>43.3</td>
<td>34.3</td>
<td>12.1</td>
<td>37.6</td>
<td>51.8</td>
</tr>
<tr>
<td>UOCTTI_LSVM_MDPM</td>
<td>53.2</td>
<td>53.9</td>
<td>13.1</td>
<td>13.5</td>
<td>30.5</td>
<td>55.5</td>
<td>51.2</td>
<td>31.7</td>
<td>14.5</td>
<td>29.0</td>
<td>16.0</td>
<td>22.1</td>
<td>43.1</td>
<td>50.3</td>
<td>46.3</td>
<td>8.8</td>
<td>33.0</td>
<td>22.9</td>
</tr>
<tr>
<td>UOCTTI_WL-SSVM_GRAMMAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UVA_SELSEARCH</td>
<td>56.9</td>
<td>43.4</td>
<td>16.6</td>
<td>15.8</td>
<td>18.0</td>
<td>52.3</td>
<td>38.3</td>
<td>48.9</td>
<td>12.2</td>
<td>29.7</td>
<td>32.8</td>
<td>36.7</td>
<td>45.7</td>
<td>54.4</td>
<td>30.4</td>
<td>16.2</td>
<td>37.2</td>
<td>34.7</td>
</tr>
</tbody>
</table>

**Classes:**
- aeroplane
- bicycle
- boat
- bottle
- car
- cat
- chair
- cow
- dining table
- dog
- horse
- motor bike
- person
- potted plant
- sheep
- sofa
- train
- tv/monitor

**Methods:**
- BROOKES_STRUCT_DET_CRF
- CMIC_GS_DPM
- CMIC_SYNTHDPM
- CORNELL_ISVM_VIEWPOINT
- MISSOURI_LCC_TREE_CODING
- MISSOURI_TREE_MAX_POOLING
- NLPR_DD_DC
- NUS_CONTEXT_SVM
- NYU_UCLA_HIERARCHY
- OXFORD_DPM_MK
- UOCTTI_LSVM_MDPM
- UOCTTI_WL-SSVM_GRAMMAR
- UVA_SELSEARCH
Precision/Recall - Motorbike

- NLPR_DD_DC (58.3)
- NYUUCLA_HIERARCHY (56.9)
- OXFORD_DPM_MK (56.6)
- NUS_CONTEXT_SVM (55.8)
- UVA_SELSEARCH (54.4)
- UOCTTI_LSVMDPM (50.3)
- MISSOURI_LCC_TREE_CODING (48.6)
- MISSOURI_TREE_MAX_POOLING (48.6)
- CMIC_GS_DPM (41.6)
- BROOKES_STRUCT_DET_CRF (41.1)
- CMIC_SYNTHDPM (38.6)
- CORNELL_ISVM_VIEWPOINT (38.2)
Precision/Recall - Person

- NLPR_DD_DC (51.6)
- UOCTTI_WL-SSVM_GRAMMAR (49.2)
- UOCTTI_LSVM_MDPM (46.3)
- NYUUCLA_HIERARCHY (43.6)
- OXFORD_DPM_MK (43.3)
- MISSOURI_TREE_MAX_POOLING (42.9)
- MISSOURI_LCC_TREE_CODING (41.9)
- CMIC_SYNTHDPM (40.7)
- BROOKES_STRUCT_DET_CRF (38.4)
- NUS_CONTEXT_SVM (36.8)
- UVA_SELSEARCH (30.4)
- CORNELL_ISVM_VIEWPOINT (7.9)
Max AP: 58.3% (motorbike) ... 16.2% (potted plant)
Median AP by Method

<table>
<thead>
<tr>
<th>Method</th>
<th>AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLPR_DD_DC</td>
<td>41.8</td>
</tr>
<tr>
<td>NYU_UCLA_HIERARCHY</td>
<td>41.5</td>
</tr>
<tr>
<td>OXFORD_DPM_MK</td>
<td>40.5</td>
</tr>
<tr>
<td>NUS_CONTEXT_SVM</td>
<td>37.2</td>
</tr>
<tr>
<td>UVA_SELSEARCH</td>
<td>37.0</td>
</tr>
<tr>
<td>CMIC_SYNTHDPM</td>
<td>34.8</td>
</tr>
<tr>
<td>MISSOURI_LCC_TREE_CODING</td>
<td>32.7</td>
</tr>
<tr>
<td>MISSOURI.Tree_max_POOL1..</td>
<td>32.7</td>
</tr>
<tr>
<td>UOCTTI_LSVT_MDP</td>
<td>32.4</td>
</tr>
<tr>
<td>CMIC_GS_DPM</td>
<td>26.4</td>
</tr>
<tr>
<td>CORNELL_SVM_VIEWPOINT</td>
<td>19.3</td>
</tr>
<tr>
<td>BROOKES_STRUCT_DET_CRF</td>
<td>16.5</td>
</tr>
</tbody>
</table>
“True Positives” - Person

NLPR_DD_DC

UOCTTI_WL-SSVM_GRAMMAR

UOCTTI_LSVVM_MDPM
“False Positives” - Person

NLPR_DD_DC

UOCTTI_WL-SSVM_GRAMMAR

UOCTTI_LSVRM_MDPM
“Near Misses” - Person

NLPR_DD_DC

UOCTTI_WL-SSVM_GRAMMAR

UOCTTI_LSBM_MDPM
“True Positives” - Motorbike

NLPR_DD_DC

NYU_UCLA_HIERARCHY

OXFORD_DPM_MK
“False Positives” - Motorbike

NLPR_DD_DC

NYUUCLA_HIERARCHY

OXFORD_DPM_MK
“True Positives” - Cat

NYU UCLA HIERARCHY

OXFORD DPM MK

UVA SELSEARCH
“False Positives” - Cat

NYUUCLA_HIERARCHY

OXFORD_DPM_MK

UVA_SELSEARCH
“Near Misses” - Cat

NYUUCLA_HIERARCHY

OXFORD_DPM_MK

UVA_SELSEARCH
- Results on 2008 data improve for best methods 2009-2011 for almost all categories
  - Caveats: More training data + re-use of test data
Results on 2009 data improve for best methods 2010-2011 for almost all categories
- Caveats: More training data + re-use of test data
- Results on 2010 data improve for best 2011 methods for all but one category (aeroplane)
  - Caveats: More training data + re-use of test data
Prizes

- **Joint Winners:**
  - **NLPR_DD_DC**
    Junge Zhang, Yinan Yu, Yongzhen Huang, Chong Wang, Weiqiang Ren, Jinchen Wu, Kaiqi Huang, Tieniu Tan
    National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
  - **NYUUCLA_HIERARCHY**
    Yuanhao Chen, Li Wan, Long Zhu, Rob Fergus, Alan Yuille
    NYU, UCLA

- **Honourable Mention (Novelty):**
  - **CORNELL_ISVM_VIEWPOINT**
    Joshua Schwartz, Noah Snavely, Daniel Huttenlocher
    Cornell University