The PASCAL Visual Object Classes Challenge 2012 (VOC2012)

Part I – Detection Challenge

Mark Everingham, Luc Van Gool
Chris Williams, John Winn
Andrew Zisserman
Yusuf Aytar, Ali Eslami
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?
Examples

Aeroplane

Bicycle

Bird

Boat

Bottle

Bus

Car

Cat

Chair

Cow
Examples

- Dining Table
- Dog
- Horse
- Motorbike
- Person
- Potted Plant
- Sheep
- Sofa
- Train
- TV/Monitor
Annotation

- Complete annotation of objects from 20 categories

**Occluded**
Object is significantly occluded within BB

**Truncated**
Object extends beyond BB

**Difficult**
Not scored in evaluation

**Pose**
Facing left
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%
Dataset statistics

- Same size as VOC2011.

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>11,540</td>
<td>10,994</td>
</tr>
<tr>
<td>Objects</td>
<td>27,450</td>
<td>27,078</td>
</tr>
</tbody>
</table>

- Minimum ~600 training objects per category
- ~2,000 cars, 1,500 dogs, 8,500 people
- Approximately equal distribution across training and test datasets
Submitted methods

- 8 methods, 7 groups
  - VOC2011: 13 methods, 15 groups

- Common approach:
  - Deformable Part Model (Girshick, Felzenszwalb, McAllester) with variations, e.g.
    - HOG-LBP features
    - Colour features
    - Multiple kernel learning

- New approaches:
  - Selective search (UVA, NEC_STANFORD)
  - Dynamic AND-OR tree
Average precision by class

Average Precision

- aeroplane
- motorbike
- bus
- bicycle
- cat
- train
- horse
- tvmonitor
- car
- person
- dog
- sheep
- sofa
- diningtable
- cow
- bottle
- bird
- boat
- pottedplant
- chair
Improvement over VOC2011

![Graph showing average precision for various categories over years 2011 and 2012.](image-url)
### AP by class and method

<table>
<thead>
<tr>
<th>Method</th>
<th>Appliance</th>
<th>bicycle</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>car</th>
<th>chair</th>
<th>cow</th>
<th>diningtable</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>pottedplant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tvmonitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC_BOW_COLOR_HOG</td>
<td>45.4</td>
<td>49.8</td>
<td>15.7</td>
<td>16.0</td>
<td>26.3</td>
<td>54.6</td>
<td>44.8</td>
<td>35.1</td>
<td>16.8</td>
<td>31.3</td>
<td>23.6</td>
<td>26.0</td>
<td>45.6</td>
<td>49.6</td>
<td>42.2</td>
<td>14.5</td>
<td>30.5</td>
<td>28.5</td>
<td>45.7</td>
</tr>
<tr>
<td>MISSOURI_HOGLBP_MDPM CONTEXT</td>
<td>51.4</td>
<td>53.7</td>
<td>18.3</td>
<td>15.6</td>
<td>31.6</td>
<td>56.5</td>
<td>47.1</td>
<td>38.6</td>
<td>19.5</td>
<td>32.0</td>
<td>22.1</td>
<td>25.0</td>
<td>50.3</td>
<td>51.9</td>
<td>44.9</td>
<td>11.9</td>
<td>37.7</td>
<td>30.6</td>
<td>50.8</td>
</tr>
<tr>
<td>NEC_STANFORD_OCP</td>
<td>65.1</td>
<td>46.8</td>
<td>25.0</td>
<td>24.6</td>
<td>16.0</td>
<td>51.0</td>
<td>44.9</td>
<td>51.5</td>
<td>13.0</td>
<td>26.6</td>
<td>31.0</td>
<td>40.2</td>
<td>39.7</td>
<td>51.5</td>
<td>32.8</td>
<td>12.6</td>
<td>35.7</td>
<td>33.5</td>
<td>48.0</td>
</tr>
<tr>
<td>OLB_FT_DPM_R5</td>
<td>47.5</td>
<td>51.7</td>
<td>14.2</td>
<td>12.6</td>
<td>27.3</td>
<td>51.8</td>
<td>44.2</td>
<td>25.3</td>
<td>17.8</td>
<td>30.2</td>
<td>18.1</td>
<td>16.9</td>
<td>46.9</td>
<td>50.9</td>
<td>43.0</td>
<td>9.5</td>
<td>31.2</td>
<td>23.6</td>
<td>44.3</td>
</tr>
<tr>
<td>SYSU_DYNAMIC_AND_OR_TREE</td>
<td>50.1</td>
<td>47.0</td>
<td>7.9</td>
<td>3.8</td>
<td>24.8</td>
<td>47.2</td>
<td>42.8</td>
<td>31.2</td>
<td>17.5</td>
<td>24.2</td>
<td>10.0</td>
<td>21.3</td>
<td>43.5</td>
<td>46.4</td>
<td>37.5</td>
<td>7.9</td>
<td>26.4</td>
<td>21.5</td>
<td>43.1</td>
</tr>
<tr>
<td>UOC_OXFORD_DPM_MKL</td>
<td>59.6</td>
<td>54.5</td>
<td>21.9</td>
<td>21.6</td>
<td>32.1</td>
<td>52.5</td>
<td>49.3</td>
<td>40.8</td>
<td>19.1</td>
<td>35.2</td>
<td>28.9</td>
<td>37.2</td>
<td>50.9</td>
<td>49.9</td>
<td>46.1</td>
<td>15.6</td>
<td>39.3</td>
<td>35.6</td>
<td>48.9</td>
</tr>
<tr>
<td>UVA_DETECTOR_MERGING</td>
<td>47.2</td>
<td>50.2</td>
<td>18.3</td>
<td>21.4</td>
<td>25.2</td>
<td>53.3</td>
<td>46.3</td>
<td>46.3</td>
<td>17.5</td>
<td>27.8</td>
<td>30.3</td>
<td>35</td>
<td>41.6</td>
<td>52.1</td>
<td>43.2</td>
<td>18</td>
<td>35.2</td>
<td>31.1</td>
<td>45.4</td>
</tr>
<tr>
<td>UVA_HYBRID_CODING_APE</td>
<td>61.8</td>
<td>52</td>
<td>24.6</td>
<td>24.8</td>
<td>20.2</td>
<td>57.1</td>
<td>44.5</td>
<td>53.6</td>
<td>17.4</td>
<td>33</td>
<td>38.3</td>
<td>42.8</td>
<td>48.8</td>
<td>59.4</td>
<td>35.7</td>
<td>22.8</td>
<td>40.3</td>
<td>39.5</td>
<td>51.1</td>
</tr>
</tbody>
</table>
Precision/recall curves (aeroplane)
Precision/recall curves (bicycle)

- UOC_OXFORD_DPM_MKL (54.5)
- MISSOURI_HOGLBP_MDPM_CONTEXT (53.6)
- UVA_HYBRID_CODING_APE (52.0)
- OLB_FT_DPM_R5 (51.6)
- UVA_DETECTOR_MERGING (50.2)
- CVC_BOW_COLOR_HOG (49.8)
- SYSU_DYNAMIC_AND_OR_TREE (47.0)
- NEC_STANFORD_OCP (46.8)
Precision/recall curves (person)

- UOC_OXFORD_DPM_MKL (46.1)
- MISSOURI_HOGLBP_MDPM_CONTEXT (44.9)
- UVA_DETECTOR_MERGING (43.2)
- OLB_FT_DPM_R5 (43.0)
- CVC_BOW_COLOR_HOG (42.2)
- SYSU_DYNAMIC_AND_OR_TREE (37.5)
- UVA_HYBRID_CODING_APE (35.7)
- NEC_STANFORD_OCP (32.8)
Precision/recall curves (bottle)
Median average precision by method

![Bar chart showing median average precision by method]
Prizes

- **Winner**
  - `UVA_HYBRID_CODING_APE`
  - Koen E. A. van de Sande,
  - Jasper R. R. Uijlings,
  - Cees G. M. Snoek,
  - Arnold W. M. Smeulders
  - *University of Amsterdam*

- **Honourable mention**
  - `OXFORD_DPM_MKL`
  - Ross Girshick, Andrea Vedaldi,
  - Karen Simonyan
  - *University of Oxford*
The PASCAL Visual Object Classes Challenge 2012 (VOC2012)

Part I – Detection Ranking Uncertainty

Mark Everingham, Luc Van Gool
Chris Williams, John Winn
Andrew Zisserman
Yusuf Aytar, Ali Eslami
Ranking uncertainty

- Only one AP curve per class and method
- However, we can use bootstrap resampling to obtain multiple AP curves (see e.g. blog post by Brendan O’Connor, 2010)
- Compare AP or rank of two methods \( A \) and \( B \)
- Can obtain a confidence interval for AP
- If \( \text{rank}(A) < \text{rank}(B) \) with high probability then \( A \) is significantly different from \( B \)
Ranking uncertainty

for each replication

1. sample a subset of the test images
2. compute AP of each submission on sample
3. compute rank of each submission based on APs

for each pair $m^1$ and $m^2$

1. $m^1$ and $m^2$ equivalent if rank of one method is not higher than the rank of the other in at least in 95% of replications
## Equivalencies by class and method

<table>
<thead>
<tr>
<th>Method</th>
<th>aeroplane</th>
<th>bicycle</th>
<th>boat</th>
<th>bull</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>diningtable</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>pottedplant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tvmonitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC_BOW_COLOR_HOG</td>
<td>45.4</td>
<td>49.8</td>
<td>15.7</td>
<td>16.0</td>
<td>26.3</td>
<td>54.6</td>
<td>44.8</td>
<td>35.1</td>
<td>31.3</td>
<td>23.6</td>
<td>26.0</td>
<td>45.6</td>
<td>49.6</td>
<td>42.2</td>
<td>14.5</td>
<td>30.5</td>
<td>28.5</td>
<td>45.7</td>
</tr>
<tr>
<td>MISSOURI_HOGGBP_MDPM_CONTEXT</td>
<td>51.4</td>
<td>53.7</td>
<td>18.3</td>
<td>15.6</td>
<td>31.6</td>
<td>56.5</td>
<td>47.1</td>
<td>38.6</td>
<td>19.5</td>
<td>32.0</td>
<td>22.1</td>
<td>25.0</td>
<td>50.3</td>
<td>51.9</td>
<td>44.9</td>
<td>11.9</td>
<td>37.7</td>
<td>30.6</td>
</tr>
<tr>
<td>NEC_STANFORD_OCP</td>
<td>65.1</td>
<td>46.8</td>
<td>25.0</td>
<td>24.6</td>
<td>16.0</td>
<td>51.0</td>
<td>44.9</td>
<td>51.5</td>
<td>13.0</td>
<td>26.6</td>
<td>31.0</td>
<td>40.2</td>
<td>39.7</td>
<td>51.5</td>
<td>32.8</td>
<td>12.6</td>
<td>35.7</td>
<td>33.5</td>
</tr>
<tr>
<td>OLB_FT_DPM_R5</td>
<td>47.5</td>
<td>51.7</td>
<td>14.2</td>
<td>12.6</td>
<td>27.3</td>
<td>51.8</td>
<td>44.2</td>
<td>25.3</td>
<td>17.8</td>
<td>30.2</td>
<td>18.1</td>
<td>16.9</td>
<td>46.9</td>
<td>50.9</td>
<td>43.0</td>
<td>9.5</td>
<td>31.2</td>
<td>23.6</td>
</tr>
<tr>
<td>SYSU_DYNAMIC_AND_OR_TREE</td>
<td>50.1</td>
<td>47.0</td>
<td>7.9</td>
<td>3.8</td>
<td>24.8</td>
<td>47.2</td>
<td>42.8</td>
<td>31.2</td>
<td>17.5</td>
<td>24.2</td>
<td>10.0</td>
<td>21.3</td>
<td>43.5</td>
<td>46.4</td>
<td>37.5</td>
<td>7.9</td>
<td>26.4</td>
<td>21.5</td>
</tr>
<tr>
<td>UOC_OXFORD_DPM_MKL</td>
<td>59.6</td>
<td>54.5</td>
<td>21.9</td>
<td>21.6</td>
<td>32.1</td>
<td>52.5</td>
<td>49.3</td>
<td>40.8</td>
<td>19.1</td>
<td>35.2</td>
<td>28.9</td>
<td>37.2</td>
<td>50.9</td>
<td>49.9</td>
<td>46.1</td>
<td>15.6</td>
<td>39.3</td>
<td>35.6</td>
</tr>
<tr>
<td>UVA_DETECTOR_MERGING</td>
<td>47.2</td>
<td>50.2</td>
<td>18.3</td>
<td>21.4</td>
<td>25.2</td>
<td>53.3</td>
<td>46.3</td>
<td>46.3</td>
<td>17.5</td>
<td>27.8</td>
<td>30.3</td>
<td>35</td>
<td>41.6</td>
<td>52.1</td>
<td>43.2</td>
<td>18.0</td>
<td>35.2</td>
<td>31.1</td>
</tr>
<tr>
<td>UVA_HYBRID_CODING_APE</td>
<td>61.8</td>
<td>52.0</td>
<td>24.6</td>
<td>24.8</td>
<td>20.2</td>
<td>57.1</td>
<td>44.5</td>
<td>53.6</td>
<td>17.4</td>
<td>33</td>
<td>38.3</td>
<td>42.8</td>
<td>48.8</td>
<td>59.4</td>
<td>35.7</td>
<td>22.8</td>
<td>40.3</td>
<td>39.5</td>
</tr>
</tbody>
</table>
Equivalencies by class and method

Difference is statistically significant
Equivalencies by class and method

Difference is not statistically significant
### Equivalencies by class and method

<table>
<thead>
<tr>
<th>CVC_BOW_COLOR_HOG</th>
<th>aeroplane</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>diningtable</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>pottedplant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tvmonitor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45.4</td>
<td>49.8</td>
<td>15.7</td>
<td>16.0</td>
<td>26.3</td>
<td>54.6</td>
<td>44.8</td>
<td>35.1</td>
<td>16.8</td>
<td>31.3</td>
<td>23.6</td>
<td>26.0</td>
<td>45.6</td>
<td>49.6</td>
<td>42.2</td>
<td>14.5</td>
<td>30.5</td>
<td>28.5</td>
<td>25.0</td>
<td>40.0</td>
</tr>
<tr>
<td>MISSOURI_HOGLBP_MDPM_CONTEXT</td>
<td>51.4</td>
<td>53.7</td>
<td>18.3</td>
<td>15.6</td>
<td>31.6</td>
<td>56.5</td>
<td>47.1</td>
<td>38.6</td>
<td>19.5</td>
<td>32.0</td>
<td>22.1</td>
<td>25.0</td>
<td>50.3</td>
<td>51.9</td>
<td>44.9</td>
<td>11.9</td>
<td>37.7</td>
<td>30.6</td>
<td>50.8</td>
<td>39.3</td>
</tr>
<tr>
<td>NEC_STANFORD_OCP</td>
<td>65.1</td>
<td>46.8</td>
<td>25.0</td>
<td>24.6</td>
<td>16.0</td>
<td>51.0</td>
<td>44.9</td>
<td>51.5</td>
<td>13.0</td>
<td>26.6</td>
<td>31.0</td>
<td>40.2</td>
<td>39.7</td>
<td>51.5</td>
<td>32.8</td>
<td>12.6</td>
<td>35.7</td>
<td>33.5</td>
<td>48.0</td>
<td>44.8</td>
</tr>
<tr>
<td>OLB_FT_DPM_R5</td>
<td>47.5</td>
<td>51.7</td>
<td>14.2</td>
<td>12.6</td>
<td>27.3</td>
<td>51.8</td>
<td>44.2</td>
<td>25.3</td>
<td>17.8</td>
<td>30.2</td>
<td>18.1</td>
<td>16.9</td>
<td>46.9</td>
<td>50.9</td>
<td>43.0</td>
<td>9.5</td>
<td>31.2</td>
<td>23.6</td>
<td>44.3</td>
<td>22.1</td>
</tr>
<tr>
<td>SYSU_DYNAMIC_AND_OR_TREE</td>
<td>50.1</td>
<td>47.0</td>
<td>7.9</td>
<td>3.8</td>
<td>24.8</td>
<td>47.2</td>
<td>42.8</td>
<td>31.2</td>
<td>17.5</td>
<td>24.2</td>
<td>10.0</td>
<td>21.3</td>
<td>43.5</td>
<td>46.4</td>
<td>37.5</td>
<td>7.9</td>
<td>26.4</td>
<td>21.5</td>
<td>43.1</td>
<td>36.7</td>
</tr>
<tr>
<td>UOC_OXFORD_DPM_MKL</td>
<td>59.6</td>
<td>54.5</td>
<td>21.9</td>
<td>21.6</td>
<td>32.1</td>
<td>52.5</td>
<td>49.3</td>
<td>40.8</td>
<td>19.1</td>
<td>35.2</td>
<td>28.9</td>
<td>37.2</td>
<td>50.9</td>
<td>49.9</td>
<td>46.1</td>
<td>15.6</td>
<td>39.3</td>
<td>35.6</td>
<td>48.9</td>
<td>42.8</td>
</tr>
<tr>
<td>UVA_DETECTOR_MERGING</td>
<td>47.2</td>
<td>50.2</td>
<td>18.3</td>
<td>21.4</td>
<td>25.2</td>
<td>53.3</td>
<td>46.3</td>
<td>46.3</td>
<td>17.5</td>
<td>27.8</td>
<td>30.3</td>
<td>35</td>
<td>41.6</td>
<td>52.1</td>
<td>43.2</td>
<td>18</td>
<td>35.2</td>
<td>31.1</td>
<td>45.4</td>
<td>44.4</td>
</tr>
<tr>
<td>UVA_HYBRID_CODING_APE</td>
<td>61.8</td>
<td>52</td>
<td>24.6</td>
<td>24.8</td>
<td>20.2</td>
<td>57.1</td>
<td>44.5</td>
<td>53.6</td>
<td>17.4</td>
<td>33</td>
<td>38.3</td>
<td>42.8</td>
<td>48.8</td>
<td>59.4</td>
<td>35.7</td>
<td>22.8</td>
<td>40.3</td>
<td>39.5</td>
<td>51.1</td>
<td>49.5</td>
</tr>
</tbody>
</table>