Modeling Image Context using Object Centered Grid

Sobhan Naderi, CVAP, KTH (sobhannp@kth.se)
Ivan Laptev, INRIA, Paris, (ivan.laptev@inria.fr)
Alireza Tavakoli, CVAP, KTH (att@kth.se)
Outline

• Introduction to contextual priming
• Bag of Features (BoF) framework
• Spatial-aware BoFs
• Object Centered Grid (OCG)
• Results
• Conclusion
Contextual Priming

• Context conveys significantly large amount of information:
  ▫ Location prediction
  ▫ Size prediction
  ▫ Scene discrimination
Holistic View of Scene
Bag of Features (BoF) Framework

- Visual vocabularies are created using K-Means
  - Shape Features (we use HOG)
  - Color Features (we use distribution of RGB textons)

- Dense feature sampling
  - Starting from 12x12 patches
  - 5 different scales

- Features are labeled according to the visual vocabs

- Each image is represented by concatenation of
  - 1000 bin shape-label histogram
  - 1000 bin color-label histogram
Spatial-aware BoFs
Our Motivation

Fixed-grid approach

Mean Average Precision on PASCAL 2007 = 55.52

Our approach
(Object Centered Grid)

Mean Average Precision on PASCAL 2007 = 70.27
Object Centered Grid (OCG)
Results

<table>
<thead>
<tr>
<th>Object Category</th>
<th>Uijlings All</th>
<th>Fixed-Grid</th>
<th>Uijlings Context-Only</th>
<th>OCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>46.2</td>
<td>50.4</td>
<td>17.8</td>
<td>66.45</td>
</tr>
<tr>
<td>Car</td>
<td>69.0</td>
<td>67.2</td>
<td>43.1</td>
<td>83.79</td>
</tr>
<tr>
<td>Cat</td>
<td>43.7</td>
<td>48.8</td>
<td>15.5</td>
<td>68.11</td>
</tr>
<tr>
<td>Chair</td>
<td>44.9</td>
<td>47.7</td>
<td>39.0</td>
<td>60.81</td>
</tr>
<tr>
<td>Horse</td>
<td>69.2</td>
<td>72.9</td>
<td>56.4</td>
<td>81.94</td>
</tr>
<tr>
<td>Motorbike</td>
<td>49.1</td>
<td>56.2</td>
<td>25.3</td>
<td>75.01</td>
</tr>
<tr>
<td>Person</td>
<td>79.2</td>
<td>80.4</td>
<td>61.6</td>
<td>89.89</td>
</tr>
<tr>
<td>Sheep</td>
<td>28.4</td>
<td>34.0</td>
<td>15.0</td>
<td>54.26</td>
</tr>
<tr>
<td>TV monitor</td>
<td>40.9</td>
<td>42.1</td>
<td>33.7</td>
<td>52.17</td>
</tr>
<tr>
<td>Mean AP</td>
<td>52.29</td>
<td>55.52</td>
<td>34.16</td>
<td>70.27</td>
</tr>
</tbody>
</table>

- Fixed-grid outperforms spatial-blind BoFs (cf. 1st and 2nd columns)
- False positives of Car are mostly Bus (same context)
- Ideal OCG (last column): shows integral importance of smart localization of grid cells
Results

- Simulated performance of using object detectors
- Dashed line shows performance of fixed-grid
- If we have a car(motorbike) detector whose recall is more than 30%(40%), OCG is worthwhile using
Conclusion

- Incorporating spatial information into the BoF framework helps image classification.

- Object Centered Grids are superior to fixed-grids given rough location of dominant objects.

- Smart localization of grid cells crucially impacts coherency and robustness of BoF representation.