LEARNING OBJECT SEGMENTATION USING A MULTI NETWORK SEGMENT CLASSIFICATION APPROACH

Simone Albertini
Ignazio Gallo, Angelo Nodari, Marco Vanetti
albertini.simone@gmail.com

February 25th, 2012
The problem

Object of interest segmentation

Prior Knowledge

Object Segmentation

The class the object belongs to

Useful for various high level tasks
The context

Commercial products images:

Low resolution

Artifacts
The proposed solution

Multi-net for Object Segmentation (MNOS)

- Inner node $C$: MLP network with its configuration $P$
- Leaf $F$: feature node
The proposed solution

Each inner node C takes as input several map-images and predicts an output map. The Network predicts the probability each pixel belongs to the Object of Interest.
The proposed solution

A leaf node only applies operators and transformations to the original image:

- Extracts color channels (RGB or HSV)
The proposed solution

A leaf node only applies operators and transformations to the original image:

- Haar, Hog, High Frequencies
Multi-Net for Object Detection

Feed-forward and Multi-Neural Networks model
Two possible internal nodes

- Sliding window approach:
  
  *Pattern generated from the raw intensity values of the input images*

- Segment-based approach:
  
  *Input images are partitioned and features are extracted from each segment*

---

Two possible internal nodes

Nodes with sliding windows operate at pixel level
Two possible internal nodes

Nodes with sliding windows operate at pixel level
Two possible internal nodes

Nodes with sliding windows operate at pixel level

Good for detection

Objective:

*sharp and neat edges to segment the object of interest*
Two possible internal nodes

Nodes with sliding windows operate at pixel level
Segments

- Original Image
- Input maps
- Output map

- Features
  - Geometry based
  - Histogram based
  - Moments

For each segment

- Build Output map
- Output predictions
- MLP

Internal node
Segment Features

• Geometric:
  Area, Perimeter
  Perimeter over area
  Bounding box locations and dimensions

• Intensity histogram
• Hu Moments
The MNOS tree

- We should feed the majority of the node with at least some leaf

- High frequency filters seems to be very effective

... Usage of sliding window nodes in the low levels of the tree
Experiments

Custom Dataset: Drezzy

8 classes:

- Bags (285)
- Shoes (400)
- Hats (158)
- Ties (203)
- Man Clothing (150)
- Man Underwear (278)
- Woman Clothing (355)
- Woman Underwear (239)

Jpeg images
Resolution: 100 x 100 or 200 x 200 pixels

VOC Challenge metrics adopted
First experiment

Only nodes based on segments classification

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Obj Acc Train</th>
<th>Obj Acc Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bags</td>
<td>83,92</td>
<td>79,13</td>
</tr>
<tr>
<td>Shoes</td>
<td>80,05</td>
<td>77,76</td>
</tr>
<tr>
<td>Hats</td>
<td>68,63</td>
<td>64,25</td>
</tr>
<tr>
<td>Ties</td>
<td>90,42</td>
<td>77,76</td>
</tr>
<tr>
<td>Man Clothing</td>
<td>73,09</td>
<td>68,06</td>
</tr>
<tr>
<td>Man Underwear</td>
<td>40,97</td>
<td>38,75</td>
</tr>
<tr>
<td>Woman Clothing</td>
<td>52,53</td>
<td>57,84</td>
</tr>
<tr>
<td>Woman Underwear</td>
<td>36,33</td>
<td>35,13</td>
</tr>
</tbody>
</table>
First experiment

The network cannot solve the problem:
• with complex backgrounds
• when we have other objects with high contrast
Hybrid Model

**Hybrid model**: Sliding window and segment classification

- **Output Map**
- **Features**
- **Segments**
- **Sliding Window**
Hybrid Model

**Hybrid model**: Sliding window and segment classification

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Acc test</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bags</td>
<td>79,00</td>
<td>- 0,13</td>
</tr>
<tr>
<td>Shoes</td>
<td>88,39</td>
<td>+ 10,63</td>
</tr>
<tr>
<td>Hats</td>
<td>62,55</td>
<td>- 1,70</td>
</tr>
<tr>
<td>Ties</td>
<td>81,52</td>
<td>+ 3,76</td>
</tr>
<tr>
<td>Man Clothing</td>
<td>73,40</td>
<td>+ 5,34</td>
</tr>
<tr>
<td>Man Underwear</td>
<td>65,25</td>
<td>+ 26,50</td>
</tr>
<tr>
<td>Woman Clothing</td>
<td>62,64</td>
<td>+ 4,80</td>
</tr>
<tr>
<td>Woman Underwear</td>
<td>66,68</td>
<td>+ 31,55</td>
</tr>
</tbody>
</table>

![Graph showing hybrid model results](image-url)
Hybrid Model

**Hybrid model**: Sliding window and segment classification
Post Processing with GrabCut

Two approaches: Bounding Box vs Region Mask

The MNOS segmentation mask is labeled in order to initialize the GrabCut
### Post Processing with GrabCut

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GrabCut</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bags</td>
<td>89,29</td>
<td>+10,29</td>
</tr>
<tr>
<td>Shoes</td>
<td>92,70</td>
<td>+4,31</td>
</tr>
<tr>
<td>Hats</td>
<td>82,19</td>
<td>+19,64</td>
</tr>
<tr>
<td>Ties</td>
<td>92,39</td>
<td>+10,87</td>
</tr>
<tr>
<td>Man Clothing</td>
<td>81,50</td>
<td>+8,10</td>
</tr>
<tr>
<td>Man Underwear</td>
<td>80,10</td>
<td>+14,85</td>
</tr>
<tr>
<td>Woman Clothing</td>
<td>66,97</td>
<td>+4,33</td>
</tr>
<tr>
<td>Woman Underwear</td>
<td>61,22</td>
<td>-5,46</td>
</tr>
</tbody>
</table>

![Graph showing performance comparison between MNOS and GrabCut for different datasets]
Post Processing with GrabCut
## Results with VOC 2011 dataset

<table>
<thead>
<tr>
<th>Class</th>
<th>MNOS</th>
<th>GC</th>
<th>Voc Best</th>
<th>Class</th>
<th>MNOS</th>
<th>GC</th>
<th>Voc Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeroplane</td>
<td>36.04</td>
<td>55.60</td>
<td>54.3</td>
<td>DiningTable</td>
<td>26.01</td>
<td>26.42</td>
<td>30.1</td>
</tr>
<tr>
<td>Bicycle</td>
<td>14.99</td>
<td>13.43</td>
<td>23.9</td>
<td>Dog</td>
<td>34.56</td>
<td>38.16</td>
<td>33.9</td>
</tr>
<tr>
<td>Bird</td>
<td>24.18</td>
<td>37.06</td>
<td>46.0</td>
<td>Horse</td>
<td>31.11</td>
<td>42.15</td>
<td>49.1</td>
</tr>
<tr>
<td>Boat</td>
<td>31.05</td>
<td>36.11</td>
<td>35.3</td>
<td>Motorbike</td>
<td>51.13</td>
<td>49.44</td>
<td>54.4</td>
</tr>
<tr>
<td>Bottle</td>
<td>20.80</td>
<td>19.11</td>
<td>49.4</td>
<td>Person</td>
<td>32.73</td>
<td>35.82</td>
<td>46.4</td>
</tr>
<tr>
<td>Bus</td>
<td>50.93</td>
<td>57.14</td>
<td>66.2</td>
<td>Pottedplant</td>
<td>19.38</td>
<td>24.44</td>
<td>28.8</td>
</tr>
<tr>
<td>Car</td>
<td>37.27</td>
<td>37.52</td>
<td>56.2</td>
<td>Sheep</td>
<td>31.26</td>
<td>31.58</td>
<td>51.3</td>
</tr>
<tr>
<td>Cat</td>
<td>36.82</td>
<td>39.33</td>
<td>46.1</td>
<td>Sofa</td>
<td>26.45</td>
<td>28.02</td>
<td>26.4</td>
</tr>
<tr>
<td>Chair</td>
<td>4.78</td>
<td>9.13</td>
<td>15.0</td>
<td>Train</td>
<td>50.26</td>
<td>52.37</td>
<td>44.9</td>
</tr>
<tr>
<td>Cow</td>
<td>40.65</td>
<td>54.07</td>
<td>47.4</td>
<td>Tvmonitor</td>
<td>18.18</td>
<td>26.54</td>
<td>45.8</td>
</tr>
</tbody>
</table>

Mean MNOS: 30.93  
Mean GC: 35.67  
Best Voc Mean: 43.3
# Results with VOC 2011 dataset

<table>
<thead>
<tr>
<th>Class</th>
<th>MNOS</th>
<th>GC</th>
<th>Voc Best</th>
<th>Class</th>
<th>MNOS</th>
<th>GC</th>
<th>Voc Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeroplane</td>
<td>36,04</td>
<td>55,60</td>
<td>54,3</td>
<td>DiningTable</td>
<td>26,01</td>
<td>26,42</td>
<td>30,1</td>
</tr>
<tr>
<td>Bicycle</td>
<td>14,99</td>
<td>13,43</td>
<td>23,9</td>
<td>Dog</td>
<td>34,56</td>
<td>38,16</td>
<td>33,9</td>
</tr>
<tr>
<td>Bird</td>
<td>24,18</td>
<td>37,06</td>
<td>46,0</td>
<td>Horse</td>
<td>31,11</td>
<td>42,15</td>
<td>49,1</td>
</tr>
<tr>
<td>Boat</td>
<td>31,05</td>
<td>36,11</td>
<td>35,3</td>
<td>Motorbike</td>
<td>51,13</td>
<td>49,44</td>
<td>54,4</td>
</tr>
<tr>
<td>Bottle</td>
<td>20,80</td>
<td>19,11</td>
<td>49,4</td>
<td>Person</td>
<td>32,73</td>
<td>35,82</td>
<td>46,4</td>
</tr>
<tr>
<td>Bus</td>
<td>50,93</td>
<td>57,14</td>
<td>66,2</td>
<td>Pottedplant</td>
<td>19,38</td>
<td>24,44</td>
<td>28,8</td>
</tr>
<tr>
<td>Car</td>
<td>37,27</td>
<td>37,52</td>
<td>56,2</td>
<td>Sheep</td>
<td>31,26</td>
<td>31,58</td>
<td>51,3</td>
</tr>
<tr>
<td>Cat</td>
<td>36,82</td>
<td>39,33</td>
<td>46,1</td>
<td>Sofa</td>
<td>26,45</td>
<td>28,02</td>
<td>26,4</td>
</tr>
<tr>
<td>Chair</td>
<td>4,78</td>
<td>9,13</td>
<td>15,0</td>
<td>Train</td>
<td>50,26</td>
<td>52,37</td>
<td>44,9</td>
</tr>
<tr>
<td>Cow</td>
<td>40,65</td>
<td>54,07</td>
<td>47,4</td>
<td>Tvmonitor</td>
<td>18,18</td>
<td>26,54</td>
<td>45,8</td>
</tr>
</tbody>
</table>

Mean MNOS: 30,93
Mean GC: 35,67
Best Voc Mean: 43,3
Commercial application

www.drezzy.it

 Offers export

Images of the products

Segmentator

Segmentation masks

Visual Information Extraction

Color

Texture
Commercial application

www.drezzy.it