Exploiting Saliency for Segmenting Objects from Image Level Labels

Seong Joon Oh\textsuperscript{1}, Rodrigo Benenson\textsuperscript{1}, Anna Khoreva\textsuperscript{1}, Zeynep Akata\textsuperscript{1,2}, Mario Fritz\textsuperscript{1}, Bernt Schiele\textsuperscript{1}

\textsuperscript{1}Max-Planck Institute for Informatics,\textsuperscript{2}University of Amsterdam

Task & Motivation
Learning to segment objects from image label annotations.
- Cheaper than full supervision.
- Humans can do.

1. Seed: Encode Image Labels

- **Data**: Pascal images + image labels.
- **Model**: fully convolutional network + global average pooling (GAP) [1, 2].

2. Saliency: Encode “Objectness” Prior

- **Data**: 1k MSRA single-object images with boxes [3]. Only non-Pascal classes are to obtain class-agnostic masks.
- **Model**: DeepLab [4].

Approach: Guided Segmentation

- **Image**: dense labelling from seed + saliency
- **1. Seed**: encode image labels
- **2. Saliency**: encode “objectness” prior
- **3. Guide Labels**: seed + saliency
- **4. Combination Algorithm**: [4]
- **Final task**: Semantic Segmentation

4. Segmentation Result & Comparison

- **Method**
- **Data**
- **Val** mIoU
- **Test** mIoU
- **FS%**

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>Val mIoU</th>
<th>Test mIoU</th>
<th>FS%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIL-FCN ECCV’16</td>
<td>I+P</td>
<td>25.0</td>
<td>23.6</td>
<td>36.5</td>
</tr>
<tr>
<td>DCSM ECCV’16</td>
<td>I+P</td>
<td>44.1</td>
<td>45.1</td>
<td>64.2</td>
</tr>
<tr>
<td>ECC ECCV’16</td>
<td>I+P</td>
<td>50.7</td>
<td>51.7</td>
<td>73.5</td>
</tr>
<tr>
<td>STC ECCV’16</td>
<td>I+P+S+E\textsubscript{gt}</td>
<td>49.8</td>
<td>51.2</td>
<td>71.8</td>
</tr>
<tr>
<td>CheckMask ECCV’16</td>
<td>I+P+µ</td>
<td>51.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Micro annoyed</td>
<td>I+P+µ</td>
<td>51.9</td>
<td>53.2</td>
<td>75.7</td>
</tr>
<tr>
<td>Guide Labeler</td>
<td>I+P+S</td>
<td>55.7</td>
<td>56.7</td>
<td>80.6</td>
</tr>
<tr>
<td>DeepLabv1 [4]</td>
<td></td>
<td>67.6</td>
<td>70.3</td>
<td>100</td>
</tr>
</tbody>
</table>

Qualitative results for different GAP types.

- **GAP-ROI**: GAP-DeepLab
- **GAP-LowRes**: GAP-LowRes
- **GAP-HighRes**: GAP-HighRes
- **GAP-DeepLab**: GAP-DeepLab

References

\[3\] Cheng et al. Global Context Based Salient Region Detection. TPAMI’15.

Acknowledgement: Supported by German Research Foundation (DFG CRC 1223).

- \(\text{Reach 80}\%\) of the fully supervised performance.
- Better saliency model will further improve the result; oracle saliency gives 61.8 mIoU.