**Task & Motivation**

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

**Motivation**

- Solve the task: e.g. train and rail co-occurrences.
- Need some prior knowledge on how "objects" look like.
- Supervision.

**1. Seed : Encode Image Labels**

Discriminative object locations from image-level classifiers.

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

1. Get discriminative object locations from an image-level classifier [1, 2] (seed).
2. Image labels alone do not give full object extent information (e.g. train and rail); we propose to exploit class-agnostic image-level saliency (saliency).
3. Combine the two sources of information (guide labels).
4. Refine the labeling by training a segmentor (e.g. DeepLab [4]) with the guide labels.

**2. Saliency : Encode “Objectness” Prior**

Foreground mask of generic object class.

- **Data:** MSRA single-object images with boxes [3]. Only non-Pascal classes are used for the class-genericity of the mask.
- **Model:** DeepLab [4].

**Approach : Guided Segmentation**

- **Seed + Saliency**
- **Saliency**
- **Guide Label**
- **Semantic Segmentation**

**3. Guide Label : Seed + Saliency**

Combination algorithm

i. Break seed and saliency into connected components.
ii. If seeds touch saliency: diffuse seeds inside saliency with dense CRF.
iii. If seed is alone, label as FG. If saliency is alone, label as BG.

**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</td>
<td>I+P</td>
<td>25.0</td>
<td>23.6</td>
<td>36.5</td>
</tr>
<tr>
<td>GCNet</td>
<td>I+P</td>
<td>44.1</td>
<td>41.5</td>
<td>64.2</td>
</tr>
<tr>
<td>BEC</td>
<td>I+P</td>
<td>50.7</td>
<td>51.7</td>
<td>73.5</td>
</tr>
<tr>
<td>STC</td>
<td>I+P+E</td>
<td>49.8</td>
<td>51.2</td>
<td>71.8</td>
</tr>
<tr>
<td>CheckMask</td>
<td>I+P+µ</td>
<td>51.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicroArena</td>
<td>I+P+µ</td>
<td>51.9</td>
<td>53.2</td>
<td>75.7</td>
</tr>
<tr>
<td>GuideLabel</td>
<td>I+P+S</td>
<td>55.7</td>
<td>56.7</td>
<td>80.6</td>
</tr>
<tr>
<td>DeepLabVT</td>
<td>I+P+E</td>
<td>67.4</td>
<td>70.3</td>
<td>100</td>
</tr>
</tbody>
</table>

**References**


Acknowledgement: Supported by German Research Foundation (DFG CRP 1229).