Universal Weakly Supervised Segmentation by Pixel-to-Segment Contrastive Learning

Tsung-Wei Ke  Jyh-Jing Hwang  Stella X. Yu
Semantic Segmentation: Classify Pixels into Semantic Categories

- **Images**
- **Segmentation CNN**
- **Predictions**
- **Pixel-wise Annotations**
State-of-the-art Methods Require Pixel-wise Annotations

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Coarse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Methods</td>
<td>Class Activation Map</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Image Tags</th>
<th>Boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.jpg" alt="Image" /></td>
<td><img src="tags.jpg" alt="Tags" /></td>
<td><img src="boxes.jpg" alt="Boxes" /></td>
</tr>
</tbody>
</table>

- **Image**: A person on a motorbike.
- **Image Tags**: Person, Motorbike.
- **Boxes**: Various boxes indicating regions of interest in the image.
<table>
<thead>
<tr>
<th>Image</th>
<th>Image Tags</th>
<th>Boxes</th>
<th>Points</th>
<th>Scribbles</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image_url" alt="Image" /></td>
<td>Person, Motorbike</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Supervision
- Coarse
- Sparse

### Current Methods
- Class Activation Map
- Conditional Random Fields
<table>
<thead>
<tr>
<th>Image</th>
<th>Image Tags</th>
<th>Boxes</th>
<th>Points</th>
<th>Scribbles</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Image" /></td>
<td><img src="image.png" alt="Image Tags" /></td>
<td><img src="image.png" alt="Boxes" /></td>
<td><img src="image.png" alt="Points" /></td>
<td><img src="image.png" alt="Scribbles" /></td>
</tr>
</tbody>
</table>

**Supervision**
- Coarse
- Sparse

**Current Methods**
- Class Activation Map
- Conditional Random Fields

<table>
<thead>
<tr>
<th>Image</th>
<th>Image Tags</th>
<th>Boxes</th>
<th>Points</th>
<th>Scribbles</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Person Motorbike</td>
<td><img src="boxes" alt="Boxes" /></td>
<td><img src="points" alt="Points" /></td>
<td><img src="scribbles" alt="Scribbles" /></td>
</tr>
</tbody>
</table>

**Supervision**
- Coarse
- Sparse

**Current Methods**
- Class Activation Map
- Conditional Random Fields

**Our Method**
- single pixel-to-segment contrastive learning loss formulation
Our SPML: Contrasts Pixels with Segments on 4 Types of Relationships

Contrastive loss for pixel $i$ with positive segments $C^+$, negative segments $C^-$:

$$L(i) = \lambda_1 L_{\text{SegSort}+}(i, V^+, V^-) + \lambda_C L_{\text{SegSort}+}(i, C^+, C^-) + \lambda_O L_{\text{SegSort}+}(i, O^+, O^-) + \lambda_A L_{\text{SegSort}+}(i, \hat{C}^+, \hat{C}^-)$$

Beats All Weak Supervision SOTA’s on Pascal VOC & DensePose

VOC 2012

mIoU w.r.t Full Supervision (%)

<table>
<thead>
<tr>
<th>Labels</th>
<th>Boxes</th>
<th>Points</th>
<th>Scribbles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>Chang et al. (2020)</td>
<td>Song et al. (2019)</td>
<td>Tang et al. (2018b)</td>
</tr>
</tbody>
</table>

VOC 2012 on Varying Scribble Length

DensePose

relative mIoU (%)

<table>
<thead>
<tr>
<th>Tang et al. (2018b)</th>
<th>Ours</th>
</tr>
</thead>
</table>

VOC 2012 on Varying Scribble Length

Pascal: Varying sparsity of scribbles and point annotation
Context-Aware Segment Retrieval via Learned Pixel-wise Feature
Code available at https://github.com/twke18/SPML