Scaling up Instance Segmentation using **Approximately Localized Phrases**

Karan Desai | Ishan Misra | Justin Johnson | Laurens van der Maaten
---|---|---|---
1 University of Michigan | 2 Meta AI

### Motivation and Goal

Labeled masks and boxes are expensive to collect.

Learn to segment **novel object classes** without using any training masks or labeled boxes for them.

### Main Idea

Training instance segmentation models with:

- **Mask annotations** (base classes)
- **Approximately Localized Phrases** (novel classes)

### Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Novel class supervision</th>
<th>AP$_{50}$-base</th>
<th>AP$_{50}$-novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-vocab Mask R-CNN*</td>
<td>none</td>
<td>62.8</td>
<td>21.4</td>
</tr>
<tr>
<td>ALP-supervised Mask R-CNN*</td>
<td>ALPs</td>
<td>56.7</td>
<td>27.8 (+6.4)</td>
</tr>
<tr>
<td>ALP-supervised FCOS-MO</td>
<td>ALPs</td>
<td>60.0</td>
<td>30.5 (+9.1)</td>
</tr>
<tr>
<td>Box-supervised Mask R-CNN* (oracle)</td>
<td>Labeled boxes</td>
<td>61.8</td>
<td>50.5</td>
</tr>
</tbody>
</table>

**Observation 1:** ALP-supervised models outperform an open-vocab baseline.

**Observation 2:** When trained with ALPs, FCOS-MO outperforms Mask R-CNN* indicating its effectiveness in handling noisy supervision of ALPs.

... see our paper for evaluations on COCO subsets and ablation studies.