Towards Accurate Multi-person Pose Estimation in the Wild[1]
George Papandreou, Tyler Zhu, Nori Kanazawa, Alexander Toshev, Jonathan Tompson, Chris Bregler, Hartwig Adam, Kevin Murphy
Google Research

Summary
○ Top-down approach to person detection and pose estimation:
  ■ Faster-RCNN based person box detection.
  ■ Pose estimation by Hough voting on heatmaps + offsets.
  ■ Keypoint-based rescoring and non max suppression.
○ COCO 2017 keypoints competition results (single model):
  ■ 0.710 mAP on COCO test-dev.
  ■ 0.691 mAP on COCO test-challenge.

1. Person detection
○ Fixed-size image crop extraction, preserving aspect ratio.
○ Pass each crop separately to pose estimation model.

2. Pose Estimation
○ Faster-RCNN based person box detection.
○ Pose estimation by Hough voting on heatmaps + offsets.
○ Keypoint-based rescoring and non max suppression.

Evaluation on COCO Keypoints 2017 (test-dev)

<table>
<thead>
<tr>
<th></th>
<th>AP</th>
<th>AP@.5</th>
<th>AP@.75</th>
<th>AP (M)</th>
<th>AP (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU-Pose [3]</td>
<td>0.618</td>
<td>0.849</td>
<td>0.675</td>
<td>0.571</td>
<td>0.682</td>
</tr>
<tr>
<td>Mask-RCNN [4]</td>
<td>0.631</td>
<td>0.873</td>
<td>0.687</td>
<td>0.578</td>
<td>0.714</td>
</tr>
<tr>
<td>Assoc. Embed [5]</td>
<td>0.655</td>
<td>0.868</td>
<td>0.723</td>
<td>0.606</td>
<td>0.726</td>
</tr>
<tr>
<td>Ours (COCO)</td>
<td>0.669</td>
<td>0.864</td>
<td>0.736</td>
<td>0.640</td>
<td>0.720</td>
</tr>
<tr>
<td>Ours (COCO+internal)</td>
<td>0.696</td>
<td>0.872</td>
<td>0.766</td>
<td>0.670</td>
<td>0.742</td>
</tr>
<tr>
<td>Ours (COCO+internal) w. ResNet-152</td>
<td>0.710</td>
<td>0.879</td>
<td>0.777</td>
<td>0.690</td>
<td>0.752</td>
</tr>
</tbody>
</table>

● Our internal dataset is roughly 2x the size of COCO. We used it to augment the training set for the second stage pose estimator (first stage was COCO only).

References