HOLISTIC, INSTANCE-LEVEL HUMAN PARISING
Qizhu Li*, Anurag Arnab*, Philip H.S. Torr

INTRODUCTION

A holistic solution to instance-level part segmentation is one which readily produces:

i. Instance-level human segmentation
ii. Semantic part segmentation

In contrast to existing instance segmentation methods [1], our approach segments humans at multiple granularities in a single forward pass through our network.

NETWORK OVERVIEW

INSTANCE SEG. MODULE

The category-level seg. module assigns each pixel to one of the P body parts. Each of the D detections defines a possible human instance, resulting in a label space of \((1, 2, \ldots, D) \times \{(1, 2, \ldots, P) \cup \{(0,0)\}\}\), including background (0,0). A label of \((i, j)\) denotes part \(j\) of human \(i\).

The box consistency term \(\psi_{box}\) encourages pixels inside a human bounding box \(B_i\) to associate with the \(i\)-th human detection:

\[
\psi_{box}(V_k = (i, j)) = \begin{cases} s_i Q_k(j), & k \in B_i \\ 0, & \text{otherwise} \end{cases}
\]

The global term \(\psi_{global}\) handles poor detection localisation by assuming equal likelihood for a pixel to belong to any of the detected humans:

\[
\psi_{global}(V_k = (i, j)) = Q_k(i)
\]

We formulate a Dense CRF [2] over these \(V\) variables:

\[
E(V = v) = -\frac{N}{K} \sum_{k=1}^{K} \log(w_k \psi_{box}(v_k) + w_2 \psi_{global}(v_k) + x) + \sum_{k, j \neq k} \psi_{crf}(v_k, v_j)
\]

RESULTS

We evaluate our instance-level part segmentation method on the Pascal Person-Parts (PPP) Dataset and obtain state-of-the-art results using Multi-task Network Cascades (MNC) [1] as a strong baseline.

The AP\(_r\) metric [3] is used to compare to other methods. A prediction is only considered correct if it has an intersection over union (IoU) with a ground truth instance above a certain threshold.

\[
AP_{r_{val}} = \sum_{r=0.5}^{0.9} AP_{r_{val}}
\]

Converting our output to instance-level human segmentation only involves mapping the predicted label \((i, j)\) to \(i\). We compare to other instance segmentation methods on the human category of the VOC12 val. set, and achieve state-of-the-art performance.

<table>
<thead>
<tr>
<th>Method</th>
<th>IoU threshold</th>
<th>AP(<em>r</em>{val})</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNC [1]</td>
<td>38.8 28.1 19.3</td>
<td>36.7</td>
</tr>
<tr>
<td>Ours, piecewise, box term only</td>
<td>38.7 28.9 17.5</td>
<td>36.7</td>
</tr>
<tr>
<td>Ours, piecewise</td>
<td>39.7 29.7 18.7</td>
<td>37.4</td>
</tr>
<tr>
<td>Ours, end-to-end</td>
<td>40.6 30.4 19.1</td>
<td>38.4</td>
</tr>
</tbody>
</table>

Table 1. Comparison of AP\(_r\) for instance-level part segmentation on PPP val. set

<table>
<thead>
<tr>
<th>Method</th>
<th>IoU threshold</th>
<th>AP(<em>r</em>{vol})</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS [3]</td>
<td>47.8 31.8 15.7 3.3 0.1</td>
<td>-</td>
</tr>
<tr>
<td>Chen et al. [4]</td>
<td>48.3 35.6 22.6 6.5 0.6</td>
<td>-</td>
</tr>
<tr>
<td>PFPN [5]</td>
<td>48.4 38.0 26.5 16.5 5.9</td>
<td>41.3</td>
</tr>
<tr>
<td>Arnab et al. [6]</td>
<td>58.6 52.6 41.1 30.4 10.7</td>
<td>51.8</td>
</tr>
<tr>
<td>R2-IOS [7]</td>
<td>60.4 51.2 33.2 - -</td>
<td>-</td>
</tr>
<tr>
<td>Arnab et al. [8]</td>
<td>65.6 58.0 46.7 33.0 14.6</td>
<td>57.4</td>
</tr>
<tr>
<td>Ours, piecewise</td>
<td>64.0 59.8 51.0 38.3 20.1</td>
<td>57.2</td>
</tr>
<tr>
<td>Ours, end-to-end</td>
<td>70.2 63.1 54.1 41.0 19.6</td>
<td>61.0</td>
</tr>
</tbody>
</table>

Table 2. Comparison of AP\(_r\) for instance-level human segmentation on VOC12 val. set

---


* Indicates equal contribution by the authors.

robots.ox.ac.uk/~aarnab/hihp


