A. Derivation of Bias

As defined in the main paper, we obtain the normalized heatmap by soft-argmax function as follows:

\[
\tilde{H}(p) = \frac{\exp(\beta \cdot H(p))}{\sum_{p' \in \Omega} \exp(\beta \cdot H(p'))}, \quad \beta > 0, \tag{1}
\]

where \(H(p)\) is the heatmap output of the network and indexed by pixel \(p\) over the range of pixels \(\Omega\). For convenience, we further define a variable \(C\) as the denominator of Eq. (1):

\[
C = \sum_{p' \in \Omega} \exp(\beta \cdot H(p')). \tag{2}
\]

We can further partition the heatmap’s pixels \(\Omega\) into four sections \(\{\Omega_1, \Omega_2, \Omega_3, \Omega_4\}\) as visualized in Fig. 1. \(\Omega_1\) is defined such that the true joint location \((x_o, y_o)\) is the expected value and center of the section. The key assumption that we make in our work is that the heatmap support for the true joint location \((x_o, y_o)\) is well localized and fully contained within \(\Omega_1\) in \(H\). As such, the sections \(\Omega_2\) to \(\Omega_4\) contain only zero or near-zero elements so we can approximate Eq. (1) for the four sections as follows:

\[
\tilde{H}(p) \approx \begin{cases} 
\frac{1}{C} \cdot \exp(\beta \cdot H(p)) & \text{for } p \in \Omega_1 \\
\frac{1}{C} & \text{for } p \in \{\Omega_2, \Omega_3, \Omega_4\}
\end{cases} \tag{3}
\]

where the normalized heatmap value approximates to \(1/C\) for \(p \in \{\Omega_2, \Omega_3, \Omega_4\}\), since the exponential of a zero in the numerator is simply 1.

The (biased) joint location \(J'(x_r, y_r)\) is defined as the expected value of the entire heatmap, which can be further decomposed into the four sections:

\[
J' = \sum_{p \in \Omega} \tilde{H}(p) \cdot p
= \sum_{p \in \Omega_1} \tilde{H}(p) \cdot p + \sum_{p \in \Omega_2, \Omega_3, \Omega_4} \tilde{H}(p) \cdot p. \tag{4}
\]

We can also view \(J' = (x_r, y_r)\) as a weighted sum of the expected location of each section:

\[
J' = w_1 J_1 + w_2 J_2 + w_3 J_3 + w_4 J_4, \quad \text{where } w_k = \sum_{p \in \Omega_k} \tilde{H}(p), \quad \text{for } k = 1, 2, 3, 4 \tag{5}
\]
where $J_1 = (x_o, y_o)$, $J_2 = (x_o, y_o + \frac{w}{2})$, $J_3 = (x_o + \frac{h}{2}, y_o)$, and $J_4 = (x_o + \frac{h}{2}, y_o + \frac{w}{2})$. Due to the symmetry of each region, we can also represent the weights $w_2$ to $w_4$ as an expression of $w$, $h$ and $C$.

\[
\begin{align*}
    w_2 &= \frac{1}{C} \cdot 2x_o(w - y_o), \\
    w_3 &= \frac{1}{C} \cdot 2(h - x_o)y_o, \\
    w_4 &= \frac{1}{C} \cdot (h - 2x_o)(w - 2y_o).
\end{align*}
\]

We can reformulate Eq. (6) in matrix format:

\[
\begin{bmatrix}
    x_r \\
    y_r
\end{bmatrix} = \begin{bmatrix}
    w_1x_o + w_2x_o + w_3(x_o + \frac{h}{2}) + w_4(x_o + \frac{h}{2}) \\
    w_1y_o + w_2(y_o + \frac{w}{2}) + w_3y_o + w_4(y_o + \frac{w}{2})
\end{bmatrix}.
\]

Substituting the weights from Eq. (7) into Eq. (8) and with the knowledge that $w_1 = 1 - w_2 - w_3 - w_4$, we arrive at the following linear equation:

\[
J^r = \begin{bmatrix}
    x_r \\
    y_r
\end{bmatrix} = \begin{bmatrix}
    (1 - \frac{hw}{2C})x_o + \frac{hw}{2C} & \frac{hw}{C} \\
    (1 - \frac{hw}{2C})y_o + \frac{hw}{C} & \frac{hw}{C}
\end{bmatrix}.
\]

Even though we began our derivation with $(x_o, y_o)$ being located in $\Omega_1$ which is in the upper left quadrant, Eq. (9) is equally applicable when $(x_o, y_o)$ is located in the other three quadrants. If we look at the Eq. (9), if $x_o < \frac{h}{2}$, then $x_r > x_o$ which pushes the coordinate to move towards the center. If $x_o > \frac{h}{2}$, then $y_r < y_o$ which also make the prediction to be closer to the center. $y_o$ is same as $x_o$. Therefore, this equation is applicable to all quadrants.

Therefore, we can predict $J^p$ from $J^r$ in closed form as follows:

\[
\begin{bmatrix}
    x_0 \\
    y_0
\end{bmatrix} = \begin{bmatrix}
    \frac{C}{C-hw}x_r - \frac{hw^2}{2(C-hw)} \\
    \frac{C}{C-hw}y_r - \frac{hw^2}{2(C-hw)}
\end{bmatrix},
\]

which is the result in the main paper.

### B. Experiment Details

We report the experiment results of HRNet [2]'s performance on different sub-benchmarks in Table 1. We report the number of person instances in each sub benchmarks divided by the proposed factors on COCO [1] validation set in Table 2 and Table 3.

We also report the detailed EPE of our method on the divided sub benchmarks in Table 4 to support the Fig. 5 in the main paper.

### References


---

**Table 1:** HRNet - Comparisons about EPE on COCO validation set. De and Re refers to detection and regression method respectively.

<table>
<thead>
<tr>
<th></th>
<th>Few kpts</th>
<th>some kpts</th>
<th>many kpts</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many occ</td>
<td>28.5/26.6</td>
<td>14.1/13.3</td>
<td>14.2/14.7</td>
<td>16.8/16.6</td>
</tr>
<tr>
<td>some occ</td>
<td>22.2/20.5</td>
<td>6.72/6.93</td>
<td>7.18/7.20</td>
<td>8.27/8.20</td>
</tr>
<tr>
<td>few occ</td>
<td>26.9/26.4</td>
<td>7.19/7.38</td>
<td>5.12/5.29</td>
<td>5.78/5.95</td>
</tr>
<tr>
<td>all</td>
<td>25.8/24.1</td>
<td>7.98/8.10</td>
<td>6.74/6.93</td>
<td>8.09/8.14</td>
</tr>
</tbody>
</table>

**Table 2:** number of person instances when separating the benchmarks according to number of present joints and input size.

<table>
<thead>
<tr>
<th></th>
<th>Few kpts</th>
<th>some kpts</th>
<th>many kpts</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>401</td>
<td>794</td>
<td>1218</td>
<td>2413</td>
</tr>
<tr>
<td>L</td>
<td>174</td>
<td>327</td>
<td>747</td>
<td>1248</td>
</tr>
<tr>
<td>XL</td>
<td>103</td>
<td>187</td>
<td>492</td>
<td>782</td>
</tr>
<tr>
<td>XXL</td>
<td>256</td>
<td>484</td>
<td>1169</td>
<td>1909</td>
</tr>
<tr>
<td>all</td>
<td>934</td>
<td>1792</td>
<td>3626</td>
<td>6352</td>
</tr>
</tbody>
</table>

**Table 3:** number of person instances when separating the benchmarks according to number of present joints and percentage of occlusions.

<table>
<thead>
<tr>
<th></th>
<th>Few kpts</th>
<th>some kpts</th>
<th>many kpts</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many occ</td>
<td>167</td>
<td>149</td>
<td>52</td>
<td>368</td>
</tr>
<tr>
<td>some occ</td>
<td>182</td>
<td>584</td>
<td>602</td>
<td>1368</td>
</tr>
<tr>
<td>few occ</td>
<td>585</td>
<td>1059</td>
<td>2972</td>
<td>4616</td>
</tr>
<tr>
<td>all</td>
<td>934</td>
<td>1792</td>
<td>3626</td>
<td>6352</td>
</tr>
</tbody>
</table>

**Table 4:** Comparison of EPE of our method with detection and regression based method on sub benchmarks divided by our proposed method on COCO validation set.

\[ x_o > \frac{h}{2}, \text{ then } x_r < x_o \text{ which also make the prediction to be closer to the center. } y_o \text{ same as } x_o. \] Therefore, this equation is applicable to all quadrants. Therefore, we can predict $J^p$ from $J^r$ in closed form as follows:

\[
\begin{bmatrix}
    x_0 \\
    y_0
\end{bmatrix} = \begin{bmatrix}
    \frac{C}{C-hw}x_r - \frac{hw^2}{2(C-hw)} \\
    \frac{C}{C-hw}y_r - \frac{hw^2}{2(C-hw)}
\end{bmatrix},
\]

which is the result in the main paper.