

Towards 3D Human Pose Estimation in the Wild: a Weakly-supervised Approach

Supplementary Material

A. Detail of geometry constraint loss

Forward

We consider four groups of bones, defined by $G = \{\text{arm, leg, shoulder, hip}\}$.

$R_{\text{arm}} = \{\text{left lower arm, left upper arm, right lower arm, right upper arm}\}$, $R_{\text{leg}} = \{\text{left lower leg, left upper leg, right lower leg, right upper leg}\}$, $R_{\text{shoulder}} = \{\text{left shoulder bone, left shoulder bone}\}$, $R_{\text{hip}} = \{\text{left hip bone, left hip bone}\}$.

A bone (e.g. left lower arm) is represented by the index of its two end-points, $e = (j_L, j_R)$, i.e., left lower arm = (left wrist, left elbow). Let $Y_{2D}^{(j)} = (u^{(j)}, v^{(j)})$, we have

$$l_e = \|Y_{3D}^{(j_L)}, Y_{3D}^{(j_R)}\| \\ = \sqrt{(u^{(j_L)} - u^{(j_R)})^2 + (v^{(j_L)} - v^{(j_R)})^2 + (\hat{Y}_{\text{dep}}^{(j_L)} - \hat{Y}_{\text{dep}}^{(j_R)})^2}$$

Reminder that \bar{l}_e is a pre-defined constant, we have

$$L_{\text{geo}}(\hat{Y}_{\text{dep}} | Y_{2D}) = \sum_{i \in G} \frac{1}{|R_i|} \sum_{e \in R_i} \left(\frac{l_e}{\bar{l}_e} - \bar{r}_i \right)^2,$$

where

$$\bar{r}_i = \frac{1}{|R_i|} \sum_{e \in R_i} \frac{l_e}{\bar{l}_e}.$$

Backward

$$\frac{\partial L_{\text{geo}}}{\partial l_e} = \frac{\partial}{\partial l_e} \sum_{i \in G} \frac{1}{|R_i|} \sum_{e \in R_i} \left(\frac{l_e}{\bar{l}_e} - \bar{r}_i \right)^2 \\ = \frac{\partial}{\partial l_e} \sum_{i \in G} \left(\frac{1}{|R_i|} \sum_{e \in R_i} \left(\frac{l_e}{\bar{l}_e} \right)^2 - \left(\frac{1}{|R_i|} \sum_{e \in R_i} \frac{l_e}{\bar{l}_e} \right)^2 \right) \\ = \sum_{i \in G} \frac{2}{|R_i|} \sum_{e \in R_i} \frac{1}{\bar{l}_e} \left(\frac{l_e}{\bar{l}_e} - \bar{r}_i \right)$$

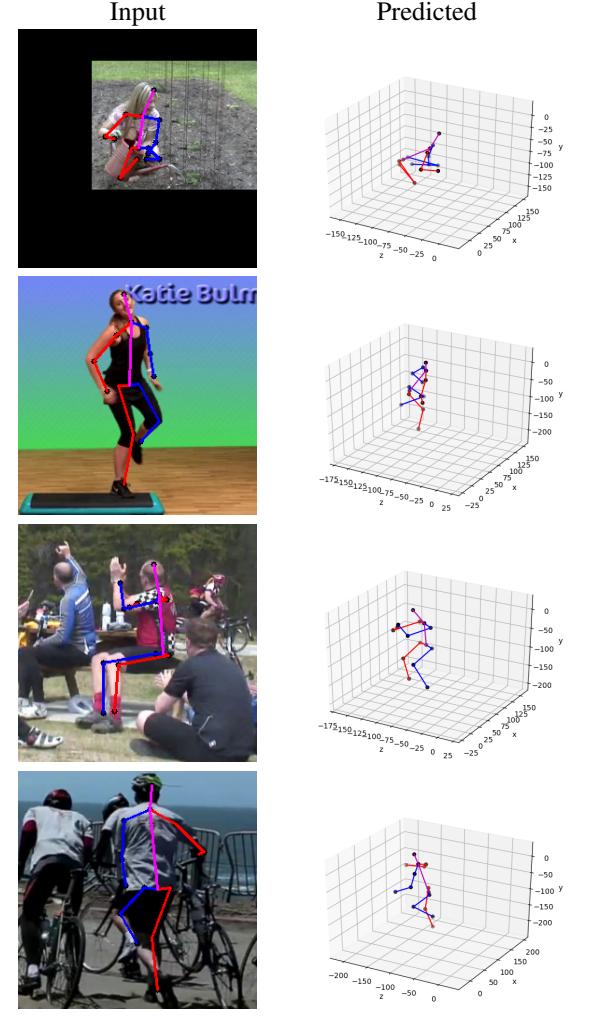
Let $e = (j, j')$

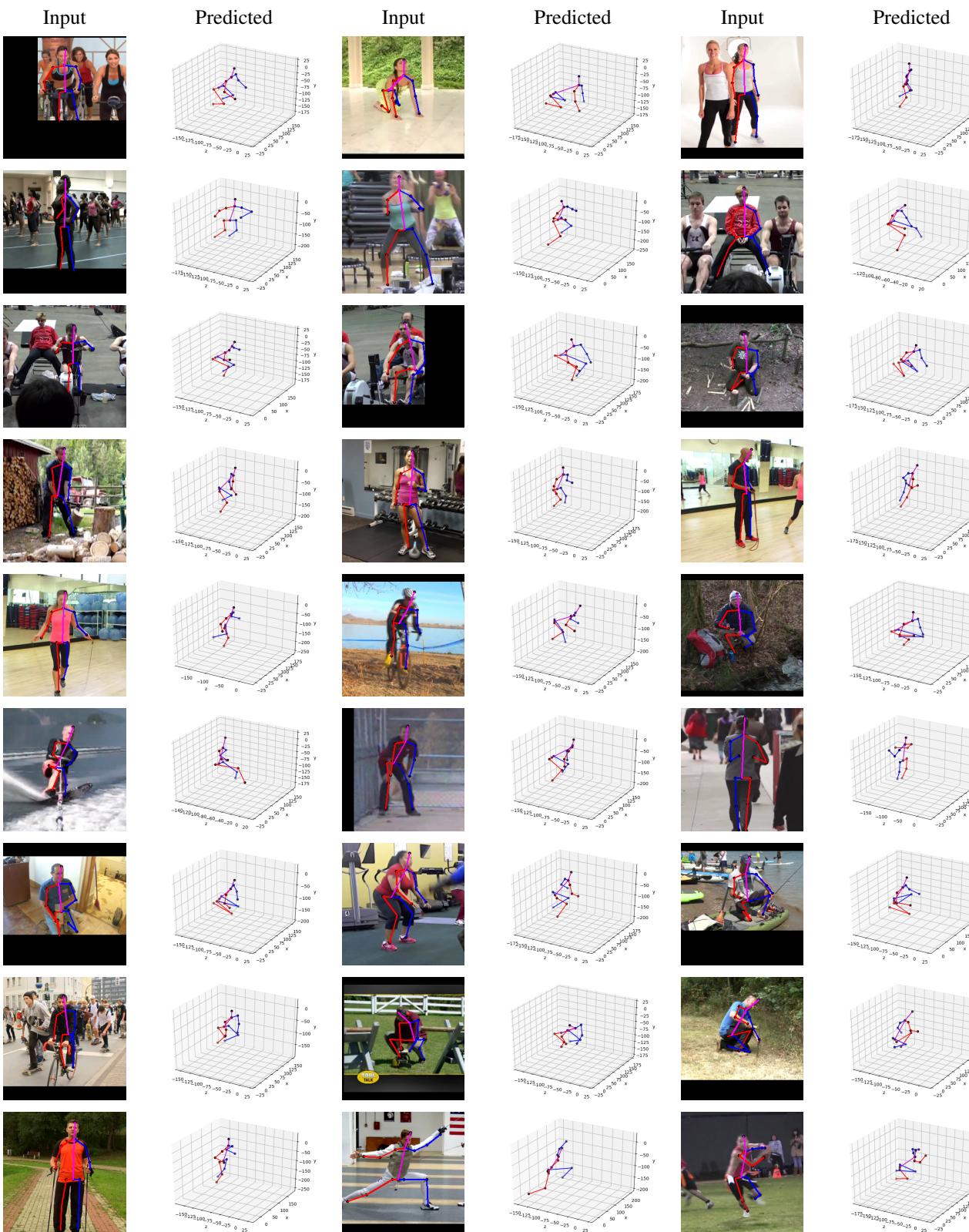
$$\frac{\partial l_e}{\hat{Y}_{\text{dep}}^{(j)}} = \frac{1}{l_e} (\hat{Y}_{\text{dep}}^{(j)} - \hat{Y}_{\text{dep}}^{(j')})$$

So we have

$$\frac{\partial L_{\text{geo}}}{\partial \hat{Y}_{\text{dep}}^{(j)}} = \frac{\partial L_{\text{geo}}}{\partial l_e} \frac{\partial l_e}{\partial \hat{Y}_{\text{dep}}^{(j)}} \\ = \sum_{i \in G} \frac{2}{|R_i|} \sum_{e \in R_i} \frac{1}{\bar{l}_e} \left(\frac{l_e}{\bar{l}_e} - \bar{r}_i \right) (Y_{\text{dep}}^{(j)} - Y_{\text{dep}}^{(j')})$$

B. More Qualitative results





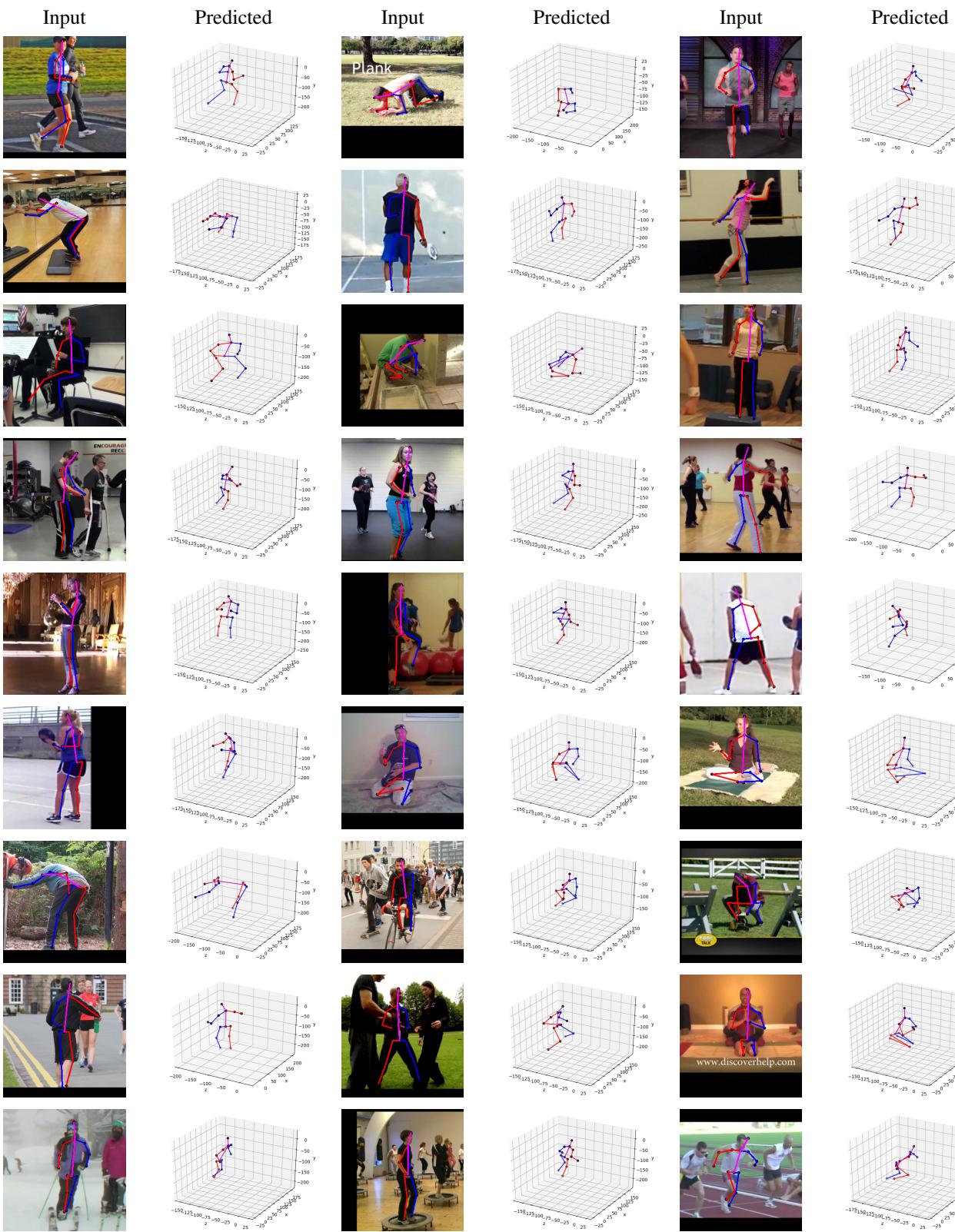




Table 1. Qualitative results from MPII validation sets. We show our predicted poses. Our framework performs well in a wide range of poses and situations. Typical failure cases (the last two lines) include inaccurate 2D prediction and ambiguous depth.