Technical Report: Conditional Regression Forests for Human Pose Estimation

Min Sun[†⋆*] Pushmeet Kohli[†] Jamie Shotton[†]

[†]Microsoft Research Cambridge [⋆]University of Michigan at Ann Arbor

1. Score Re-weighting

For the model condition on player height, we observe that the scales of $S(X^*(a), a)$ in Eq. 5 of the main paper are not comparable for different states of the global variable (i.e., short or tall person). This is because the examples in the training data are not evenly distributed in four different height bins (0.14 : 0.47 : 1 : 0.17). We propose to learn a weight $w(a)$ for each height bin a so that $w(a) \cdot S(X^*(a), a)$ among different as become comparable. Assuming the ground truth global states of all examples $\{a_n\}$ are given, we learn the weights w following the max-margin formulation below,

$$\begin{aligned} \min_{w} \quad & \frac{1}{2} w^T w + C \sum_n \xi_n \\ \text{s.t.} \quad & w(a_n) \cdot S(X^*(a_n), a_n) - w(a) \cdot S(\mathcal{X}^*(a), a) \\ & \geq 1 - \xi_n ; \forall a \neq a_n, \forall n \ . \end{aligned} \tag{1}$$

where the constraints encourage the re-weighted score $w(a_n) \cdot S(X^*(a_n), a_n)$ of the ground truth state a_n is bigger than the re-weighted scores of other states. After the weights are learned, we linearly scale the learned weights so that $w(3)$ (i.e., the weight corresponding to the bin with the largest number of examples) is one. In our experiment, we used $1.5433; 1.0230; 1; 2.8497$ which follow the trend of the inversion of the training proportion (0.14 : 0.47 : 1 : 0.17) in different bins.

2. Votes Re-scaling

According to Eq. 8 in the main paper, we enable votes from other states of the global variable to contribute to the scoring function. However, we notice that the voting direction of a short player $p(x|c_l, a = 1)$ is quite different from the voting direction of a tall player $p(x|c_l, a = 4)$. We mitigate this problem by defining the re-scale factor $r(a)$ for each state, where $r(3)$ of the states corresponding to the bin with the largest number of examples is set to one, and $r(a)$ is set to be the median height of bin 3 divided by the median height of bin a. Hence, Eq. 7 in the main paper is modified as below,

$$p(\hat{x}|c_l, a) = \sum_b p(\hat{x}|c_l, b) p(b|a) \ . \tag{2}$$

where $\hat{x} = x \cdot r(b)/r(a)$.

[*]This work was conducted while Min Sun was an intern in Microsoft Research Cambridge.