Technical Report:
Articulated Part-based Model for Joint Object Detection and Pose Estimation

Min Sun   Silvio Savarese

Dept. of Electrical and Computer Engineering, University of Michigan at Ann Arbor, USA

{sunmin,silvio}@umich.edu

1. Percentage of Correct Parts (PCP)

As mentioned the main paper, Percentage of Correct Parts (PCP) is a common performance measure used by researchers on datasets such as Stickmen [1]. Since the PCP for each part is calculated as (#of detected parts)/(#of detected objects), the set of detected objects need to be fix/predefined for this performance measure. For example, on the stickmen dataset, only 360 our of 1283 people are used as the pre-defined set for PCP evaluation. As also pointed out in [3], it is not an ideal way to measure the joint object detection and pose estimation performance.

2. Matching Criteria

PCP uses the following matching criteria to measure the performance. It is a criteria based on both endpoints of each part (e.g., matching the elbow and the wrist correctly): A part prediction is correct if the part's endpoints are on average within r of the corresponding ground truth segments, where r is a fraction of the ground truth part length. Researchers refer to the matching criteria as $PCPr$.

3. Recall v.s. False Positive Per Image(FPPI)

In our paper, we use recall v.s. False Positive Per Image (FPPI) curves to show the joint object detection and pose estimation (i.e.,parts localization) performance. The criteria for a correct object detection is the same as [2], where the intersection divided by the union between a candidate bounding box and the closest ground truth bounding box needs to be bigger than 50%. A part detection is considered to be correct if both the object corresponding to the part is detected and the PCP matching criteria for the part is satisfied. For person category, we used the same PCP0.5 criteria as [1]. For cats and dogs categories, we use PCP0.7 criteria.

When plotting the recall v.s. FPPI curves, we first collect all the candidate object instances (including object bounding boxes and part locations), then we sort the candidate instances in a descend order according to the overall matching score (Eq.[4] in the main paper). Finally, we calculate the recalls and FPPIs using different thresholds of the matching score to generate the curves.

References

[1] M. Eichner and V. Ferrari. Better appearance models for pictorial structures. In *BMVC*, 2009. 1

[2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL VOC2010 Results. 1

[3] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. 2011. 1