Supplementary Materials for
"Cross-Domain Adaptation for Animal Pose Estimation"

Jinkun Cao[1] Hongyang Tang[1] Hao-Shu Fang[1] Xiaoyong Shen[2] Cewu Lu[1] Yu-Wing Tai[2]
[1]Shanghai Jiao Tong University [2]Tencent
{caojinkun, lucewu}@sjtu.edu.cn
{thutanghy, fhaoshu, goodshenxy}@gmail.com yuwingtai@tencent.com

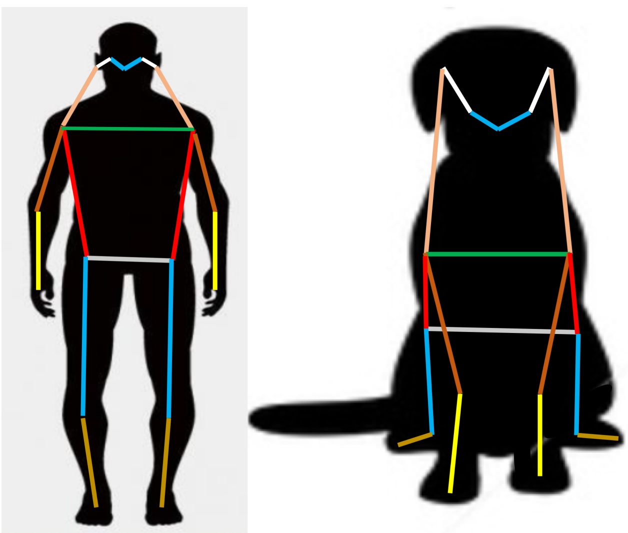

Figure 1: 18 'bones' defined based on COCO-format keypoint annotations.

1. Proposed Dataset

The built animal pose dataset has been released on www.jinkuncao.com/animalpose. We select some pose-labeled samples from our built dataset in Fig 2. Besides, to bring convenience for the attempt to do domain adaptation of animal pose estimation to novel animal categories, we also provide bounding-box-labeled images of seven novel animal categories: otter, antelope, bear, chimpanzee, rhino, bobcat, and hippopotamus. We select some samples in Fig 3.

As explained in our paper, we use 18 pre-defined 'bones' to evaluate the domain shift on keypoints between animals and humans, the definition of which is shown in Fig 1.

	mAP for each class				
N_{GT}	cat	dog	sheep	cow	horse
0	42.3	41.0	54.7	57.3	53.1
50	71.2	67.0	64.0	60.1	68.5
100	71.5	67.8	64.8	61.3	69.0
200	72.7	68.4	66.7	64.6	70.3

Table 1: Comparisons of model performance in unsupervised manner and supervised manner. N_{GT} is the number of instance with labeled ground truth label put into training. The training settings of all groups are as same as described in the paper.

2. More Results from Our Method

In Fig 4, we show more output samples by our proposed method where whether WS-CDA is used is set for comparison.

3. Failure Cases of Proposed Methods

We also sampled some representative failure cases in Fig 5. Unusual appearance may make keypoints unrecognized. For instance, our model report unrecognized withers on hedgehog with spine on back(a) and dog with clothes(b). Rhino with horn(d) makes the estimation of face-keypoints total chaos. On the other hand, bad global feature such as too low contrast ((f),(g)) or unusual gesture ((h),(i),(j)) bring difficulty to our model as well.

4. How ground truth helps model performance

To compare with the reported model performance of unsupervised domain adaptation (Table 2 in the paper), we provide more annotations on the target domain into training trying to reach an accuracy upper bound. We annotate 200 instances of each category as extra supervision (which are also contained in the released dataset). The result is shown in Table 1. It proves that even our proposed methods help a lot to do unsupervised animal pose estimation by domain adaptation, the domain shift between different animal categories still harms model performance very much. More intuitively, to introduce some ground truth on the target domain can greatly boost the model performance. To summarize, to achieve more reliable domain adaptation results and to make labor intensive labeling work less necessary, there is still much work to do. Another interesting fact found from the experiment is that the extent of the benefit model gains from introduced ground truth (performance gap between supervised and unsupervised settings) varies very much on different categories. We think it might result from the different domain shift between different domains.

Figure 2: Samples from proposed animal-pose dataset.

Figure 3: Samples of seven novel animal categories with bounding box provided in our dataset.

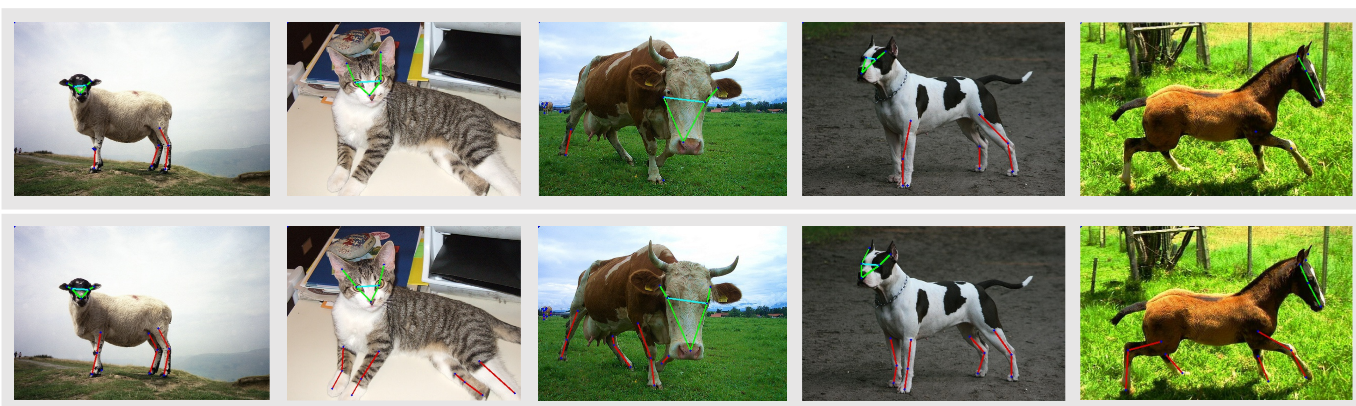

Figure 4: Upper images are estimated pose by model trained without WS-CDA. Lower ones are obtained after model being trained with WS-CDA. Common 17 keypoints between animals and humans are selected for visualization.

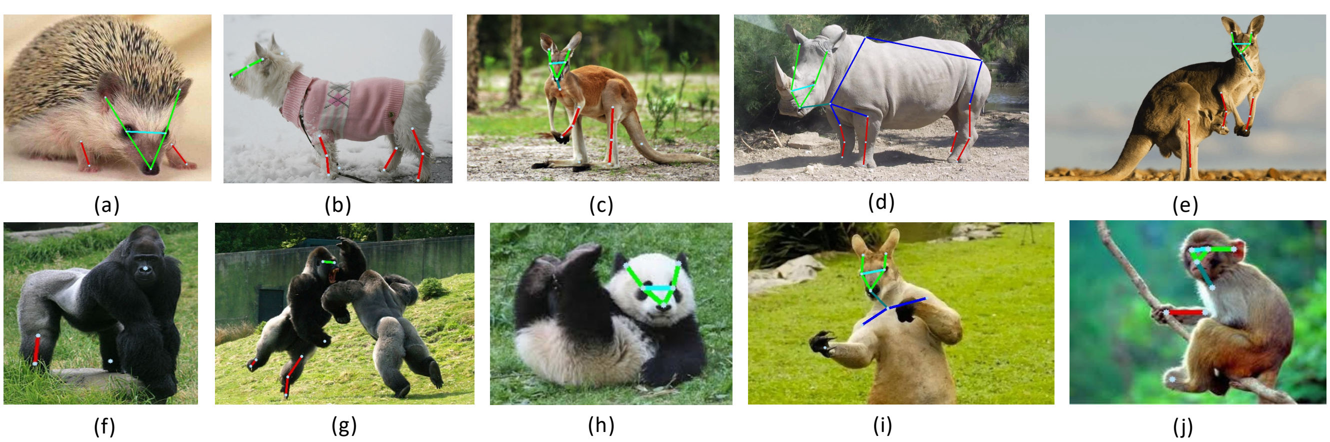

Figure 5: Samples of failure cases generated by our proposed methods on unseen animal categories.