Human Pose Estimation using Global and Local Normalization

Ke Sun, Cuiling Lan, Junliang Xing,
Wenjun Zeng, Dong Liu, Jingdong Wang

Overview of the supplementary material

In this supplementary material, we provide visualizations of our pose estimation results from Fig. 1 to Fig. 4. Fig. 1 and Fig. 3 show that the performance is apparently improved by our proposed refinement module. Fig. 4 shows the comparisons with the other two state-of-the-art methods. The proposed method is efficient on handling challenge poses, e.g., the poses with diverse orientations on the LSP dataset.

We find that even without normalization, the variation of poses is not diverse on the MPII dataset, which is similar to that on LSP after normalization. That is why the gain on the MPII dataset is smaller than that on the LSP dataset. The comparisons of the relative distributions of joints on the LSP dataset and the MPII dataset are shown in Fig. 5 and Fig. 6.

Finally, some failure cases are shown in Fig. 7. There are two main reasons: joint detector cannot always provide reasonable predictions for our normalization stages and some postures are seldom included in the training set.

Estimated pose on the LSP dataset [1]

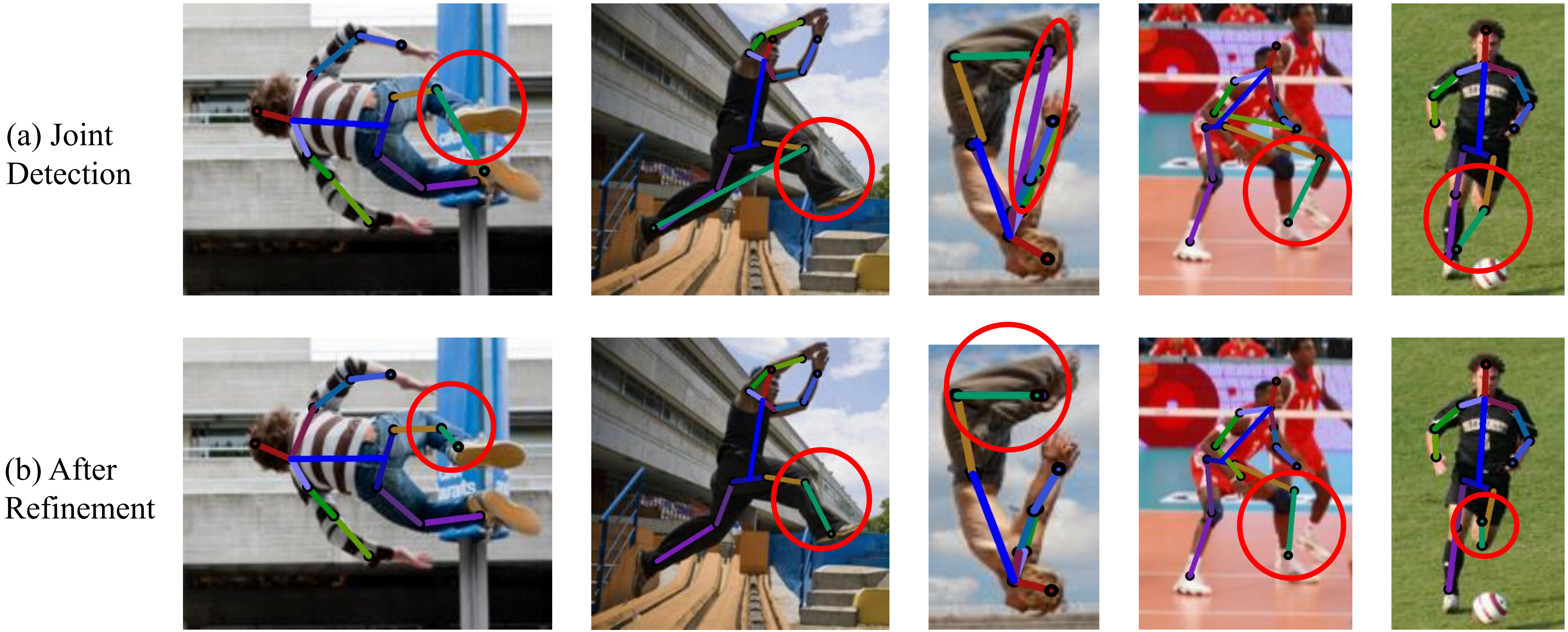

Figure 1. Pose estimation results on the LSP dataset (a) from joint detection without proposed spatial refinement, and (b) from our scheme with spatial refinement.

Results on extreme poses on the LSP dataset [1]

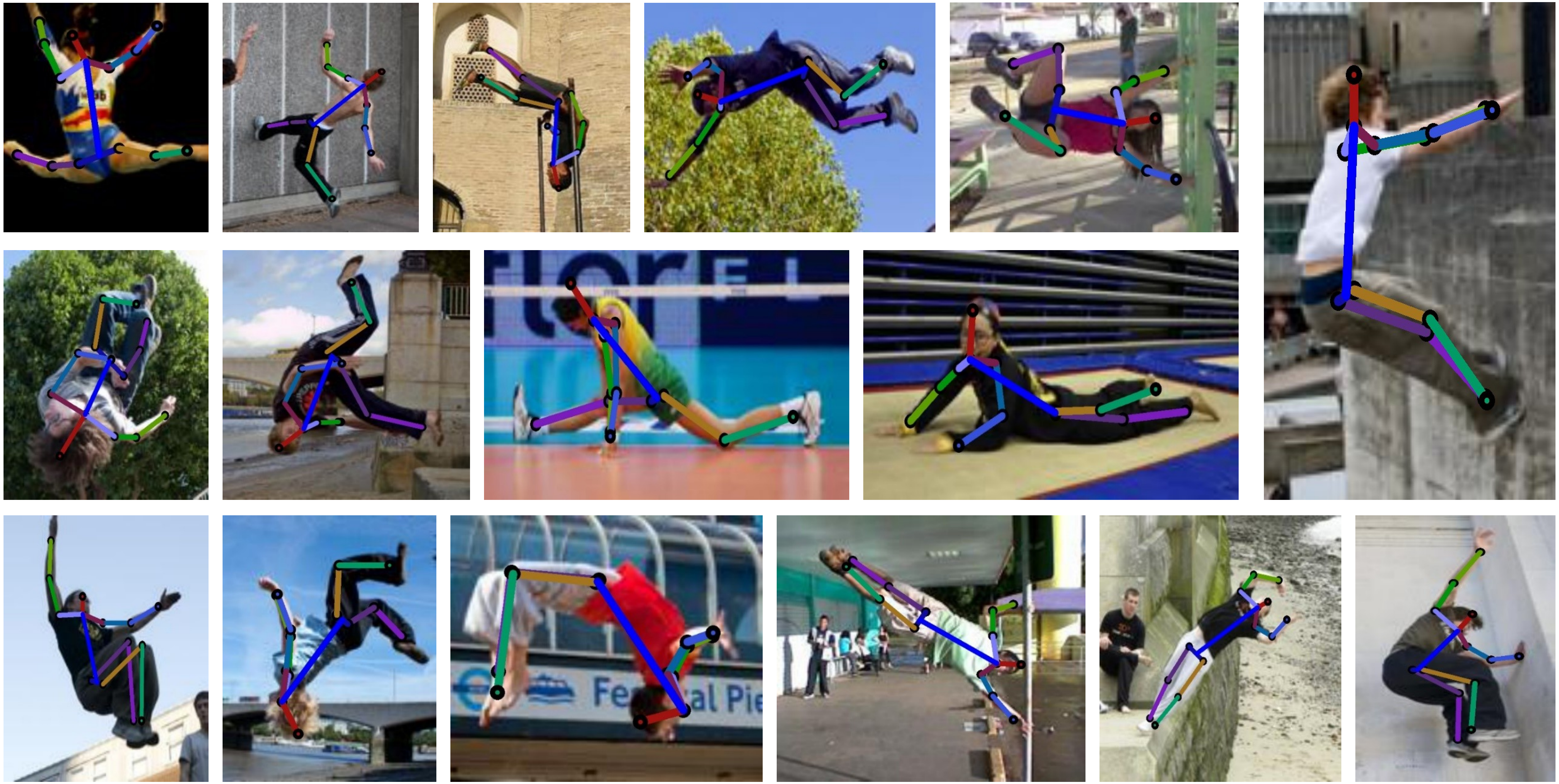

Figure 2. Pose estimation results on sample examples with extreme poses on the LSP dataset.

Estimated pose on the FLIC Dataset [2]

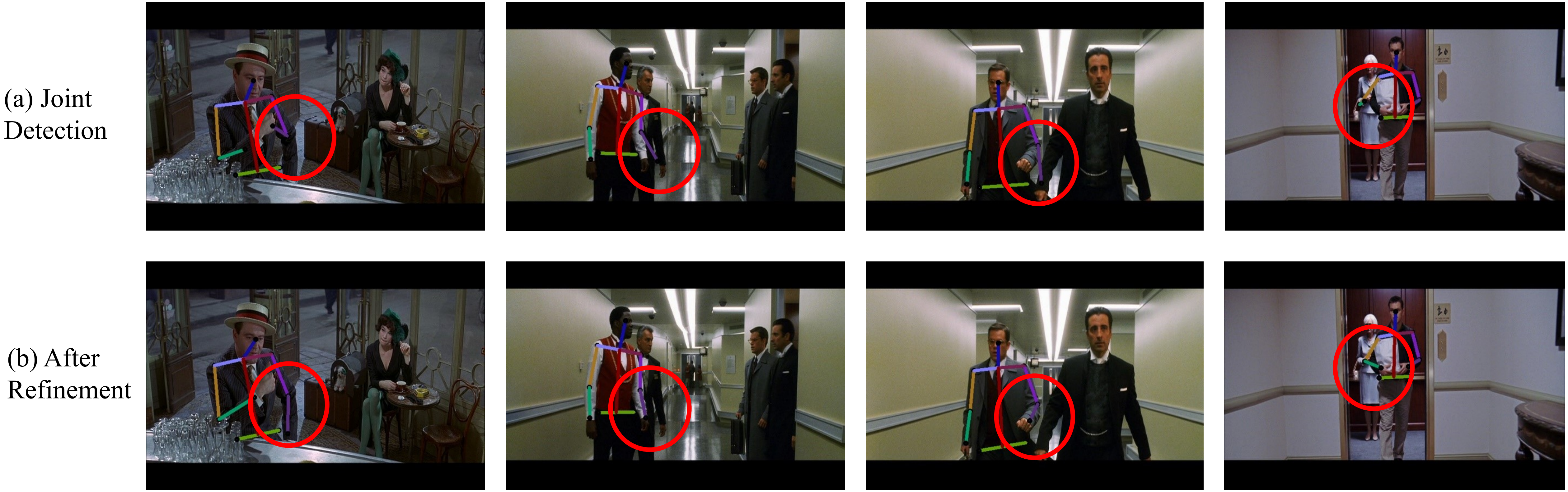

Figure 3. Pose estimation results on the FLIC dataset (a) from joint detection without proposed spatial refinement, and (b) from our scheme with spatial refinement.

Cross-method comparison on the FLIC dataset [2]

Figure 4. Pose estimation results in comparison with other state-of-the-arts for the FLIC dataset.

Distribution comparison between LSP and MPII[5]

Original
Distribution
on the LSP

After
normalization
on the LSP

Without
normalization
on the MPII

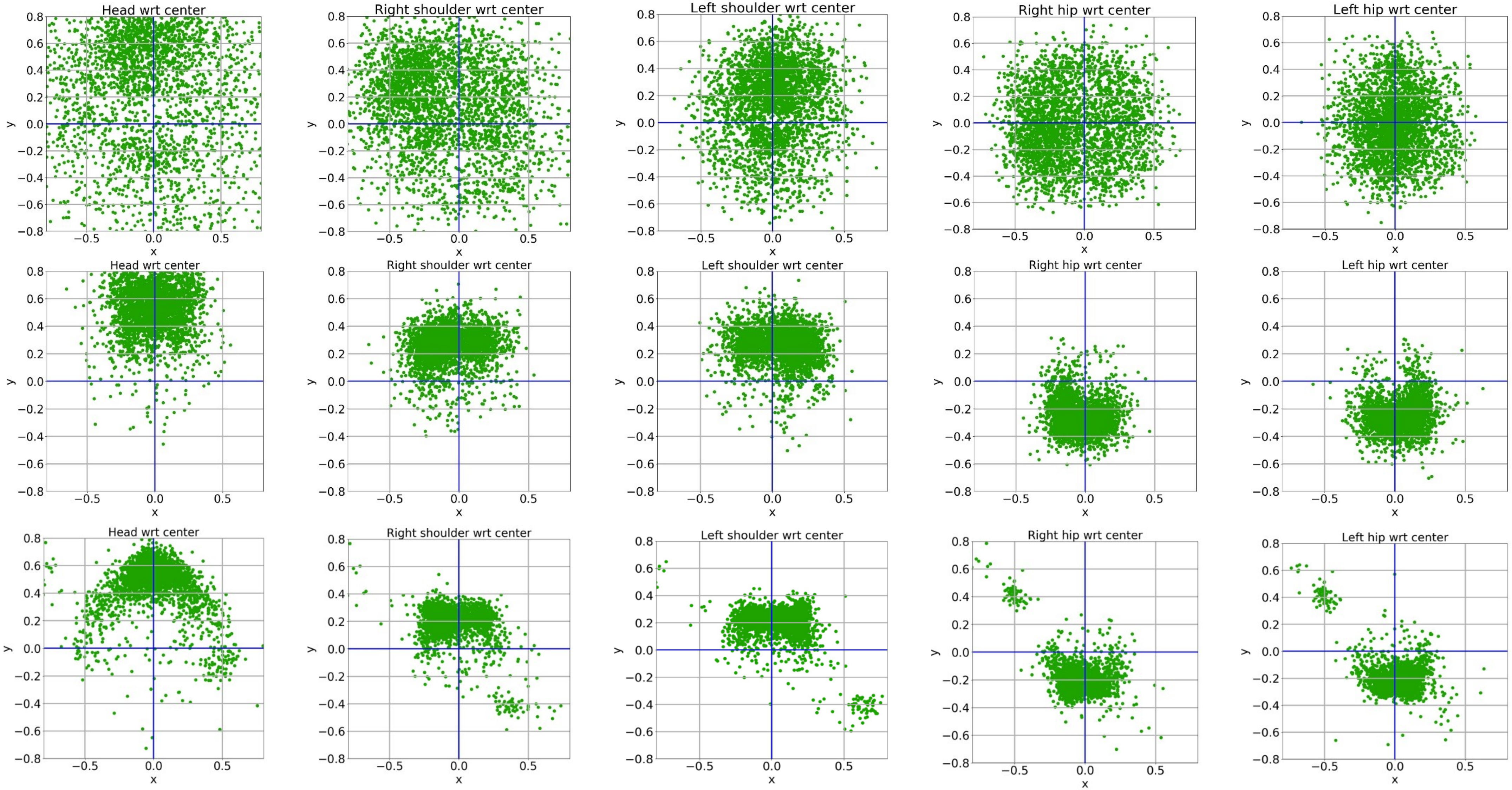

Figure 5. Relative position distributions of the body joints on the LSP and the MPII dataset (PC). The first row is the original distribution. The second row is the distribution after body normalization on the LSP training dataset. The third row is the distribution without any normalization on the MPII validation dataset. The distributions on the MPII dataset are much compact than that of the LSP dataset.

Distribution comparison between LSP and MPII[5]

Original Distribution on the LSP

After normalization on the LSP

Without normalization on the MPII

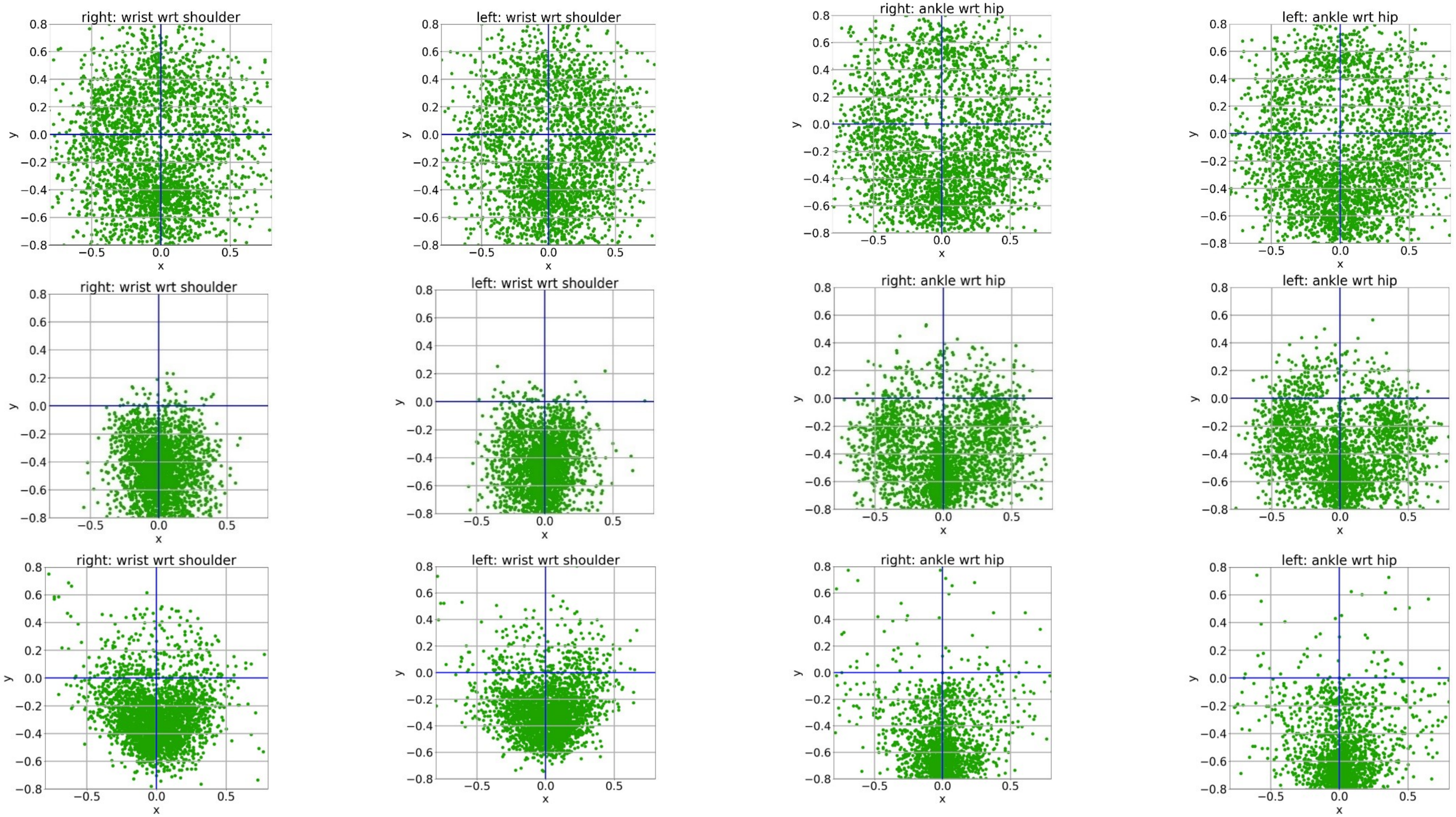

Figure 6. Relative position distributions of the flexible joints on the LSP and MPII dataset (PC). The first row is the original distribution. The second row is the distribution after limb normalization on the LSP training dataset. The third row is the distribution without any normalization on the MPII validation dataset. **Even without any normalization, the distributions of joints on the MPII dataset are very similar to the distributions after normalization on the LSP dataset.**

Failure cases on the LSP dataset [1]

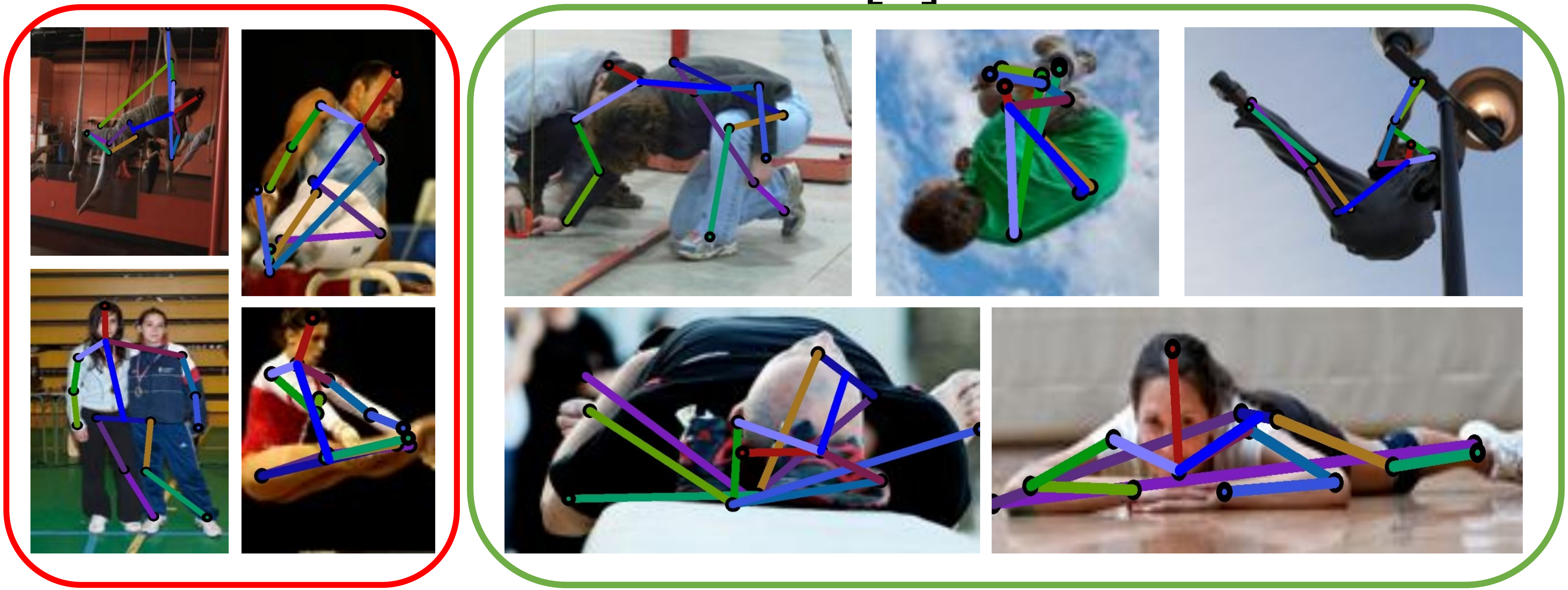

Figure 7. Failure cases on the LSP dataset. As the cases in red box, the joint detector produces unreliable prediction for our refinement. The postures in green box are rare in the training dataset.

References

[1]. S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC*, 2010.

[2]. B. Sapp and B. Taskar. Modec. Multimodal decomposable models for human pose estimation. In *CVPR*, 2013.

[3]. A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *ECCV*, 2016.

[4]. S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *CVPR*, 2016.

[5]. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In CVPR, 2014.