Learning Feature Pyramids for Human Pose Estimation
Supplementary Material

Wei Yang[1] Shuang Li[1] Wanli Ouyang[1,2] Hongsheng Li[1] Xiaogang Wang[1]
[1] Department of Electronic Engineering, The Chinese University of Hong Kong
[2] School of Electrical and Information Engineering, The University of Sydney
{wyang, sli, wlouyang, hsli, xgwang}@ee.cuhk.edu.hk

Outline

- Figure 1: Additional qualitative results on the MPII human pose datasets [1].
- Figure 2: Additional qualitative results on the LSP datasets [5].
- Figure 3: Qualitative comparison of our method with the state-of-the-art methods on the LSP dataset.
- Figure 4: Qualitative comparison of our method with the stacked hourglass network [6] on the MPII human pose test set.
- Figure 5: Analysis of failure cases on the MPII dataset [1] and the LSP dataset [5].
- Figure 6: Demo: Human pose estimation in video.
- Figure 7: Comparison of our initialization scheme with Xavier method [3] an MSR method [4] (Training and validation curves of PCKh scores *vs*. epoch on the MPII validation set).

References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *CVPR*, 2014. 1, 2
[2] A. Bulat and G. Tzimiropoulos. Human pose estimation via convolutional part heatmap regression. In *ECCV*, 2016. 4
[3] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, volume 9, pages 249–256, 2010. 1, 7
[4] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015. 1, 7
[5] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC*, 2010. 1, 3, 5
[6] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *ECCV*. Springer, 2016. 1, 5
[7] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *CVPR*, 2016. 4

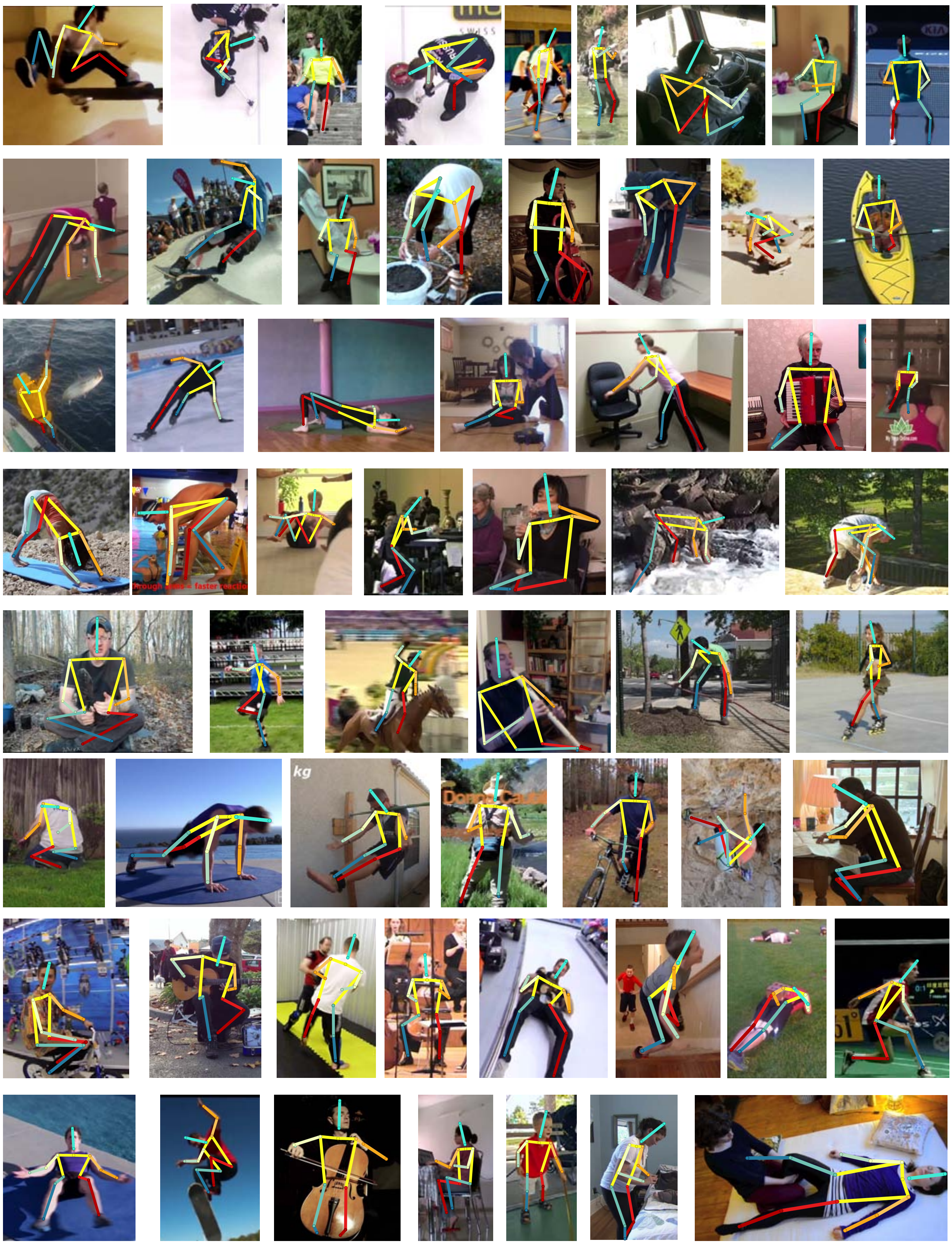

Figure 1. Qualitative results on the MPII human pose datasets [1].

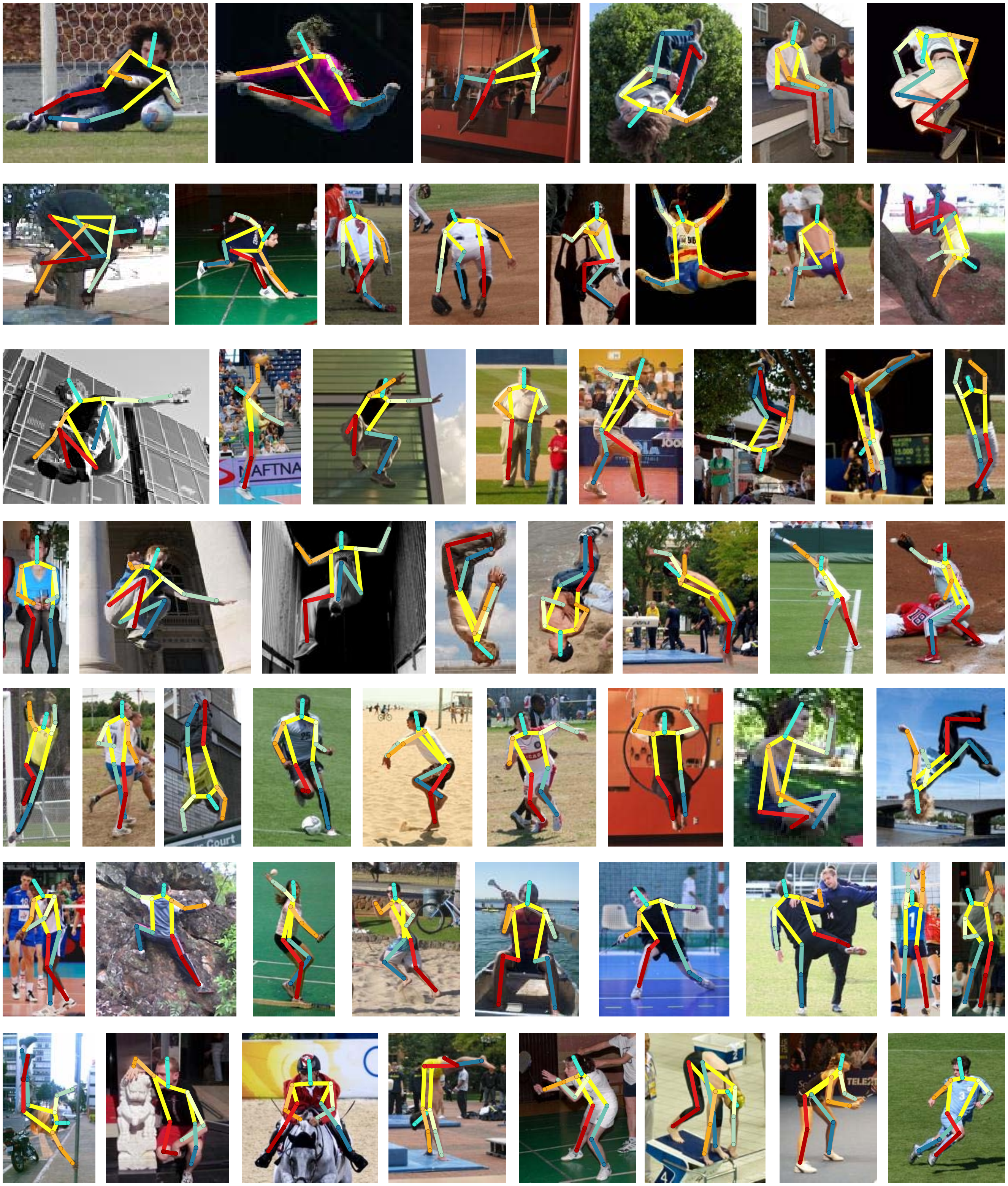

Figure 2. Qualitative results on the LSP datasets [5].

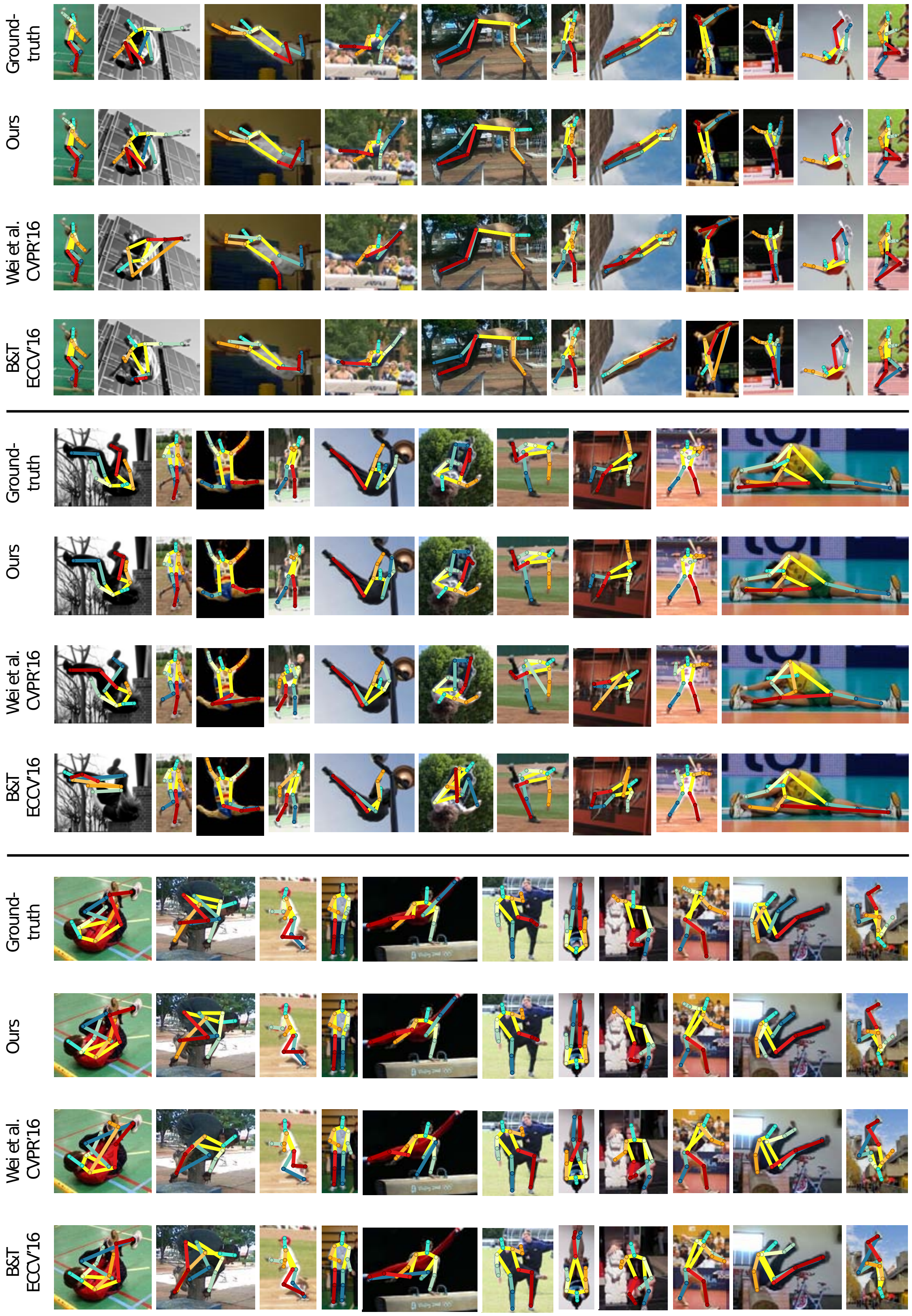

Figure 3. Qualitative comparison on the LSP test set. For each sample, we first visualize the groundtruth annotations in the 1st row. Then we compare the predictions of our method (2nd row) with those of Wei *et al*. (Convolutional Pose Machine) [7] (3rd Row) and Bulat and Tzimiropoulos [2] (4th row).

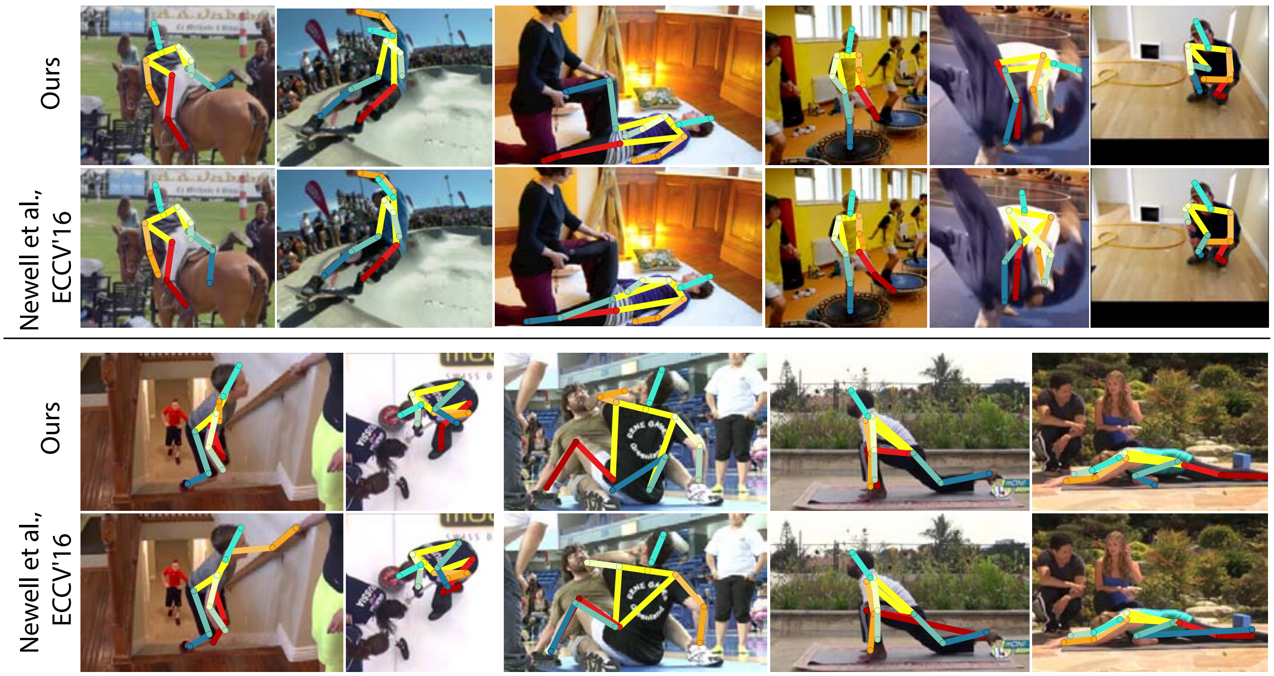

Figure 4. Qualitative comparison on the MPII human pose test set. For each sample, We compare the predictions of our approach (1st row) with the stacked hourglass network [6] (2nd row). As demonstrated, our method is more robust to the difficult camera view and the foreshortening. Note that the groundtruth of the MPII test set is held out, and the results of [6] are predicted with the code and the model released by the authors.

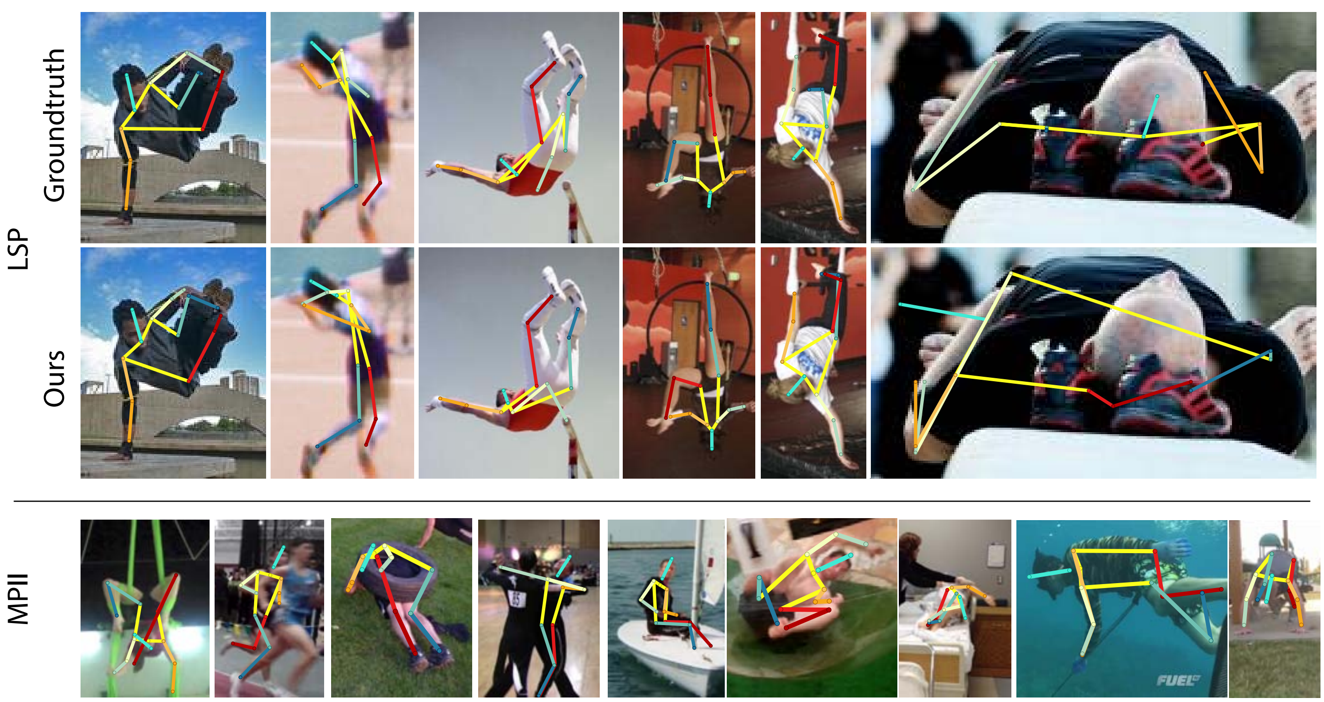

Figure 5. Failure cases on the LSP dataset [5] and the MPII human pose test set [6]. We only show the groundtruth of the LSP test set here because that the groundtruth of the MPII test set is held out. Our method may generate wrong estimations due to heavy occlusion, low resolution, limb-like objects, and extreme foreshortening.

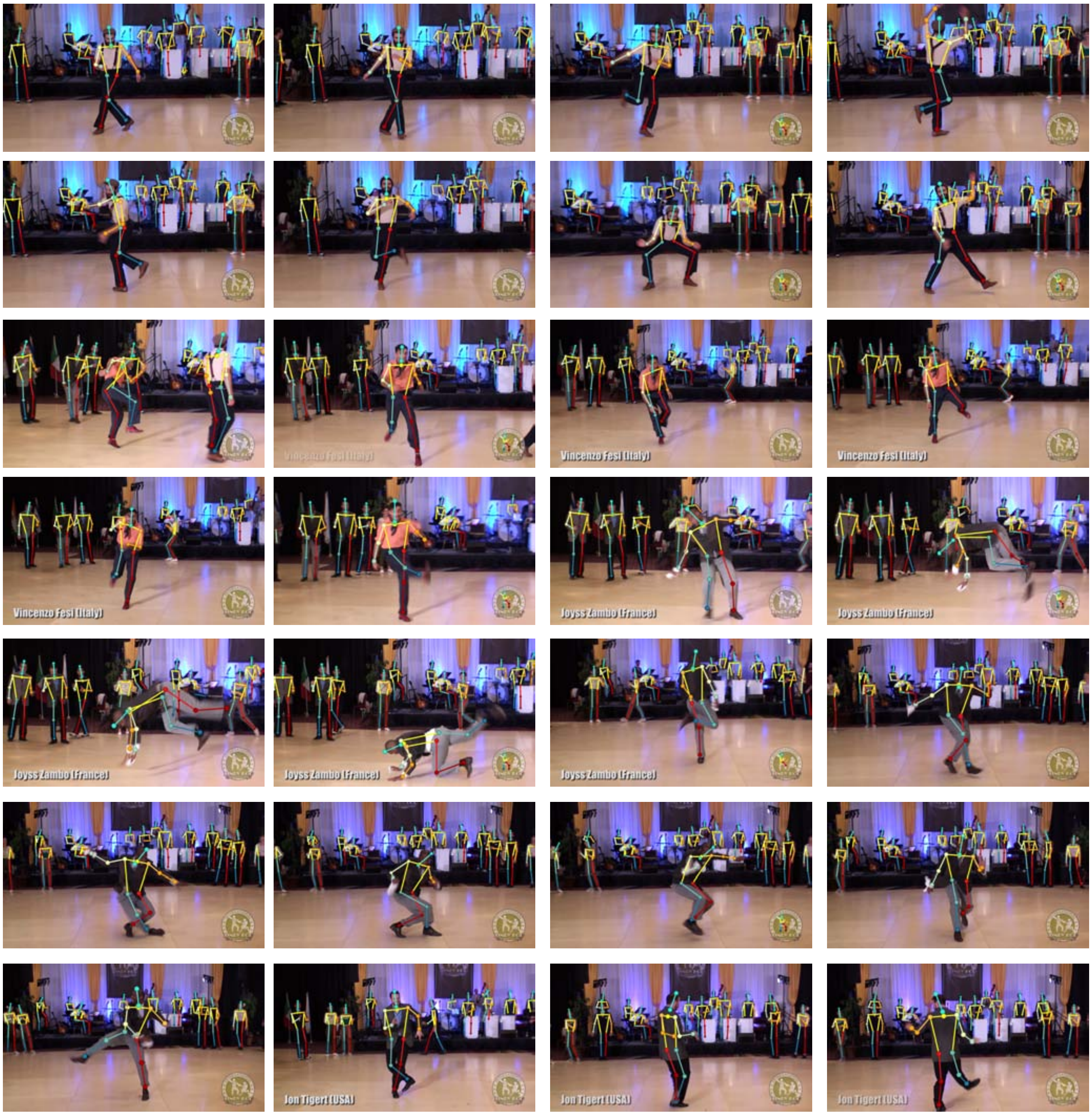

Figure 6. Our results on a video sequence. The rough locations and the scales of the people is first estimated per frame by a human detector. Then the human poses are obtained by applying our method on each possible location. Note that no temporal information is used in this example. Please refer to https://youtu.be/17atJhXb-Eg to watch our video demo.

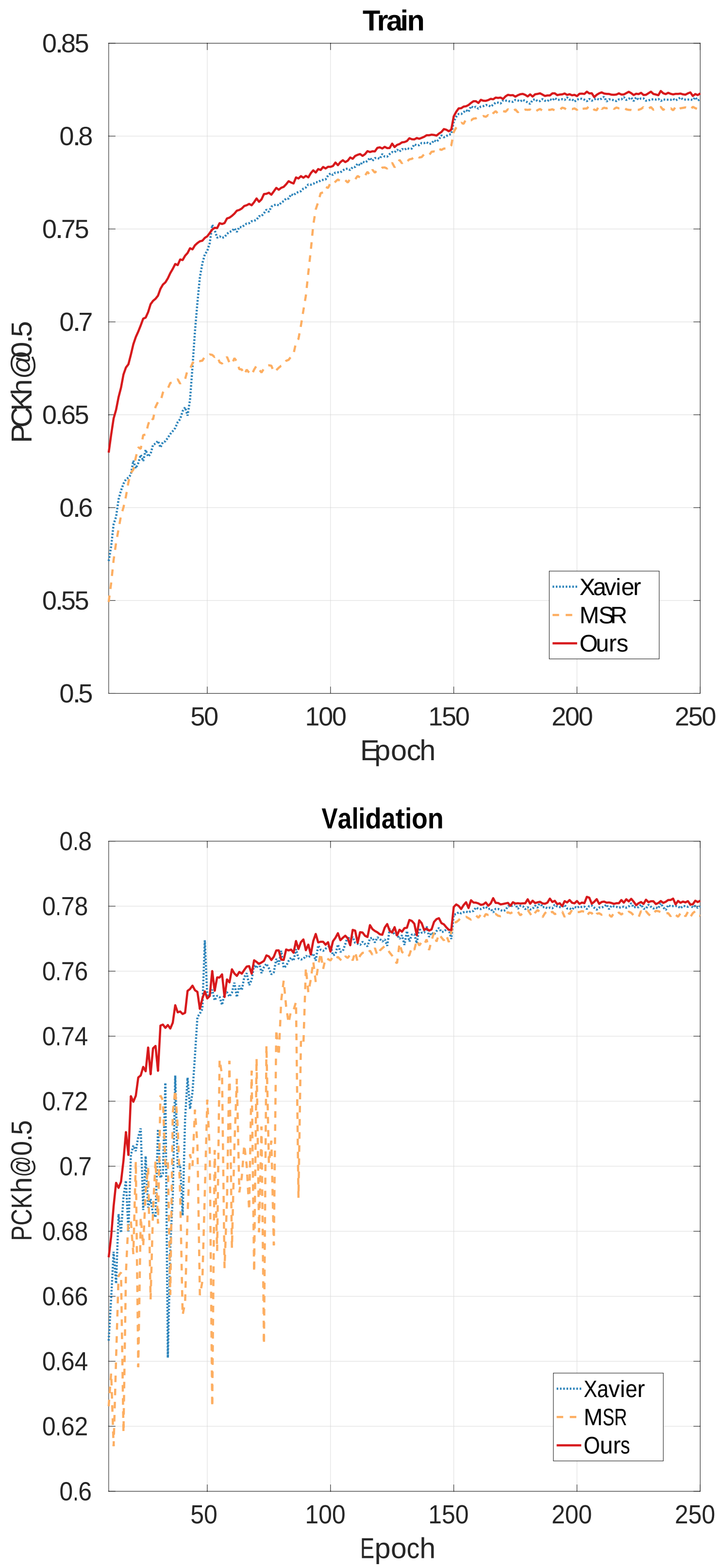

Figure 7. Comparison of our initialization scheme with Xavier method [3] an MSR method [4] (Training and validation curves of PCKh scores *vs*. epoch on the MPII validation set). We observe that both the Xavier method and the MSR method have difficulties in optimization at the beginning of the training. Moreover, the validation accuracies of the Xavier and the MSR methods oscillate significantly before reaching a stable point.