End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation

Supplementary Material

Wei Yang Wanli Ouyang Hongsheng Li Xiaogang Wang
Department of Electronic Engineering, The Chinese University of Hong Kong
{wyang, wlouyang, hsli, xgwang}@ee.cuhk.edu.hk

This supplementary file presents qualitative human pose estimation results on LSP [2], FLIC [5], and Image Parse [4] datasets in Figure 2-5. Approaches in comparison include Pishchulin *et al*. [3],Tompson *et al*. [6], Chen and Yuille [1], and the proposed approach. PDJ curve of each body joint for the Image Parse dataset [4] is also shown in Figure 1. Please refer to the captions for the detailed discussion.

Additionally, the heat-maps generated by the first two message passing layers are visualized in Figure 6. Four examples are shown in Figure 6. For each example, the first row visualizes the predicted pose and the heat-maps generated by the first message passing layer, and the second row shows the predicted pose and the heat-maps generated by the second message passing layer.

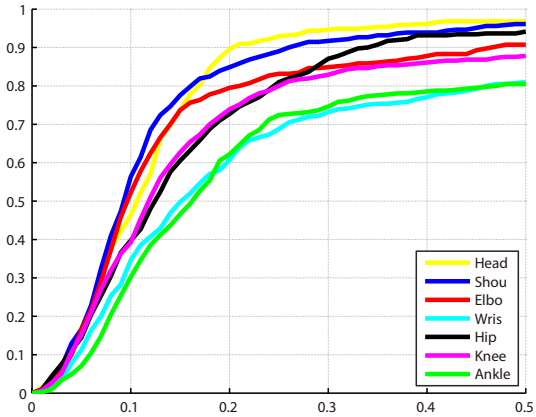

Figure 1. PDJ curve of each body joint on the Image Parse dataset.

References

[1] X. Chen and A. L. Yuille. Articulated pose estimation by a graphical model with image dependent pairwise relations. In *NIPS*, 2014.

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

[3] L. Pishchulin, M. Andriluka, P. Gehler, and B. Schiele. Strong appearance and expressive spatial models for human pose estimation. In *ICCV*, 2013.

[4] D. Ramanan. Learning to parse images of articulated objects. In *NIPS*, 2006.

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

[6] J. Tompson, A. Jain, Y. LeCun, and C. Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *NIPS*, 2014.

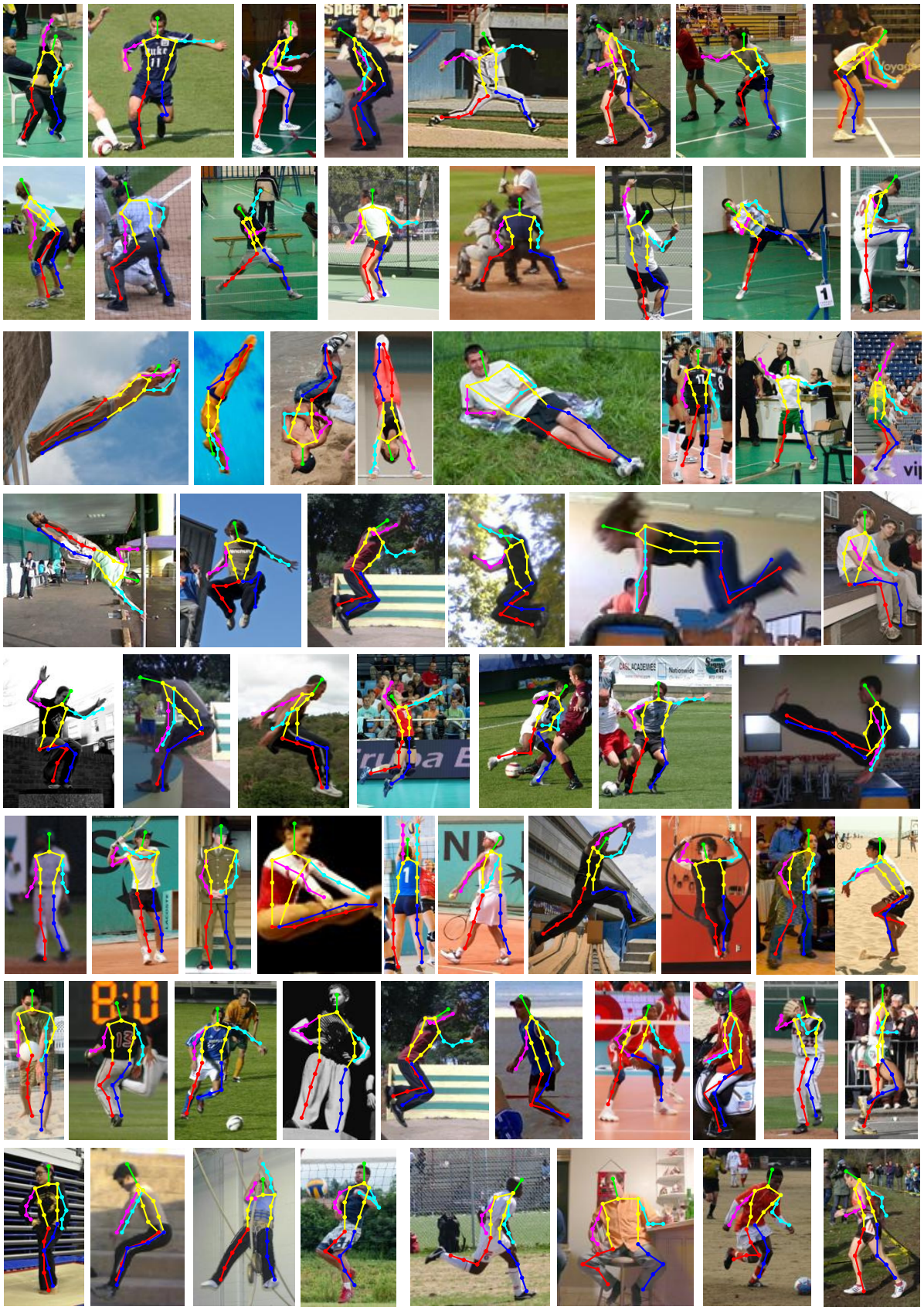

Figure 2. Additional results on LSP dataset.

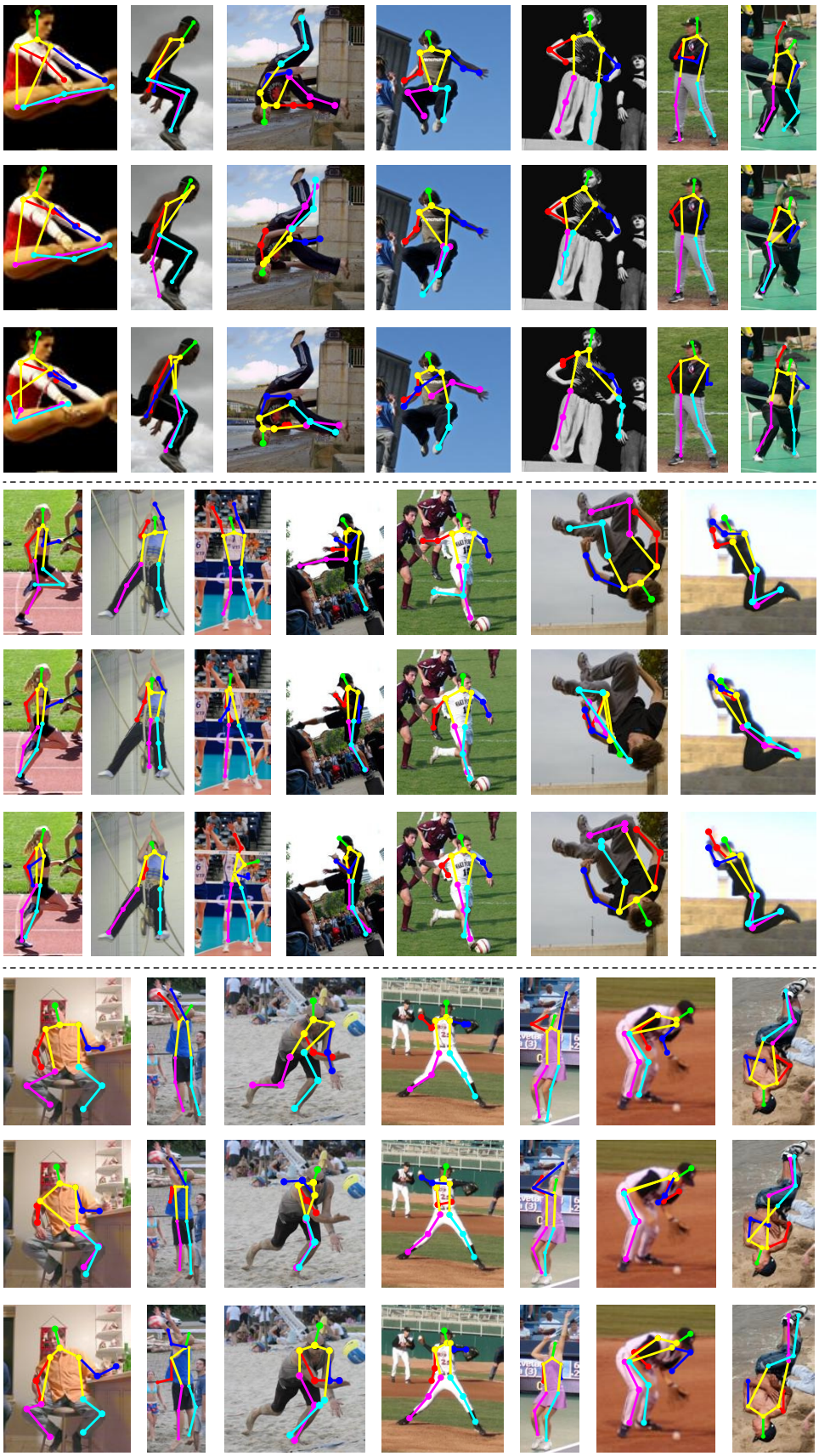

Figure 3. Qualitative comparison of estimated poses on the LSP dataset. We compare our approach (1st row) with Chen and Yuille [1] (2nd row) and Pishchulin *et al*. [3] (3rd row).

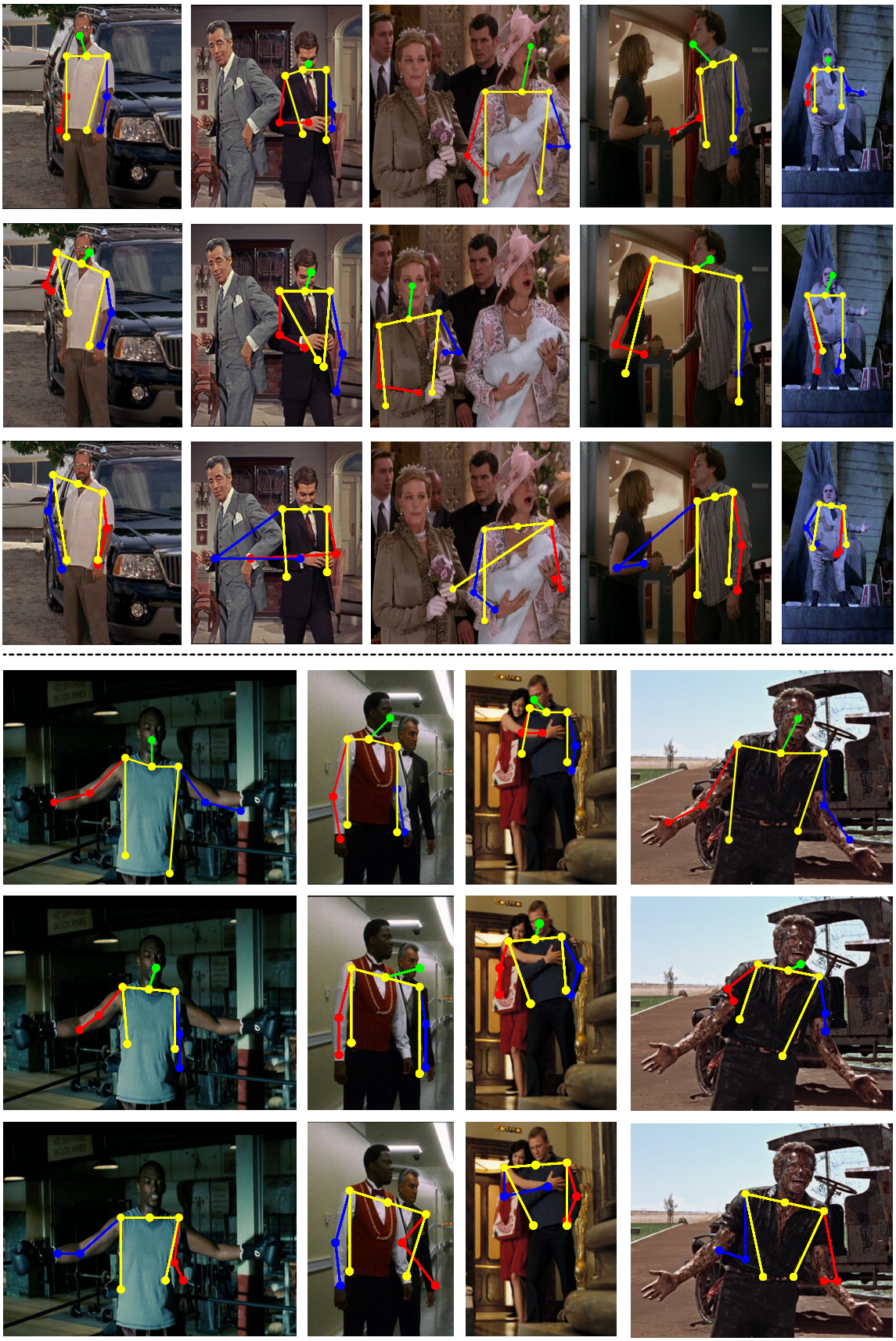

Figure 4. Qualitative comparison of estimated poses on the FLIC dataset. We compare our approach (1st row) with Chen and Yuille [1] (2nd row) and Tompson *et al*. [6] (3rd row). Note that Tompson *et al*. [6] used additional data (FLIC-plus dataset [6]) for training, and there is no *nose* in their published results.

Figure 5. Additional results on the Image Parse dataset.

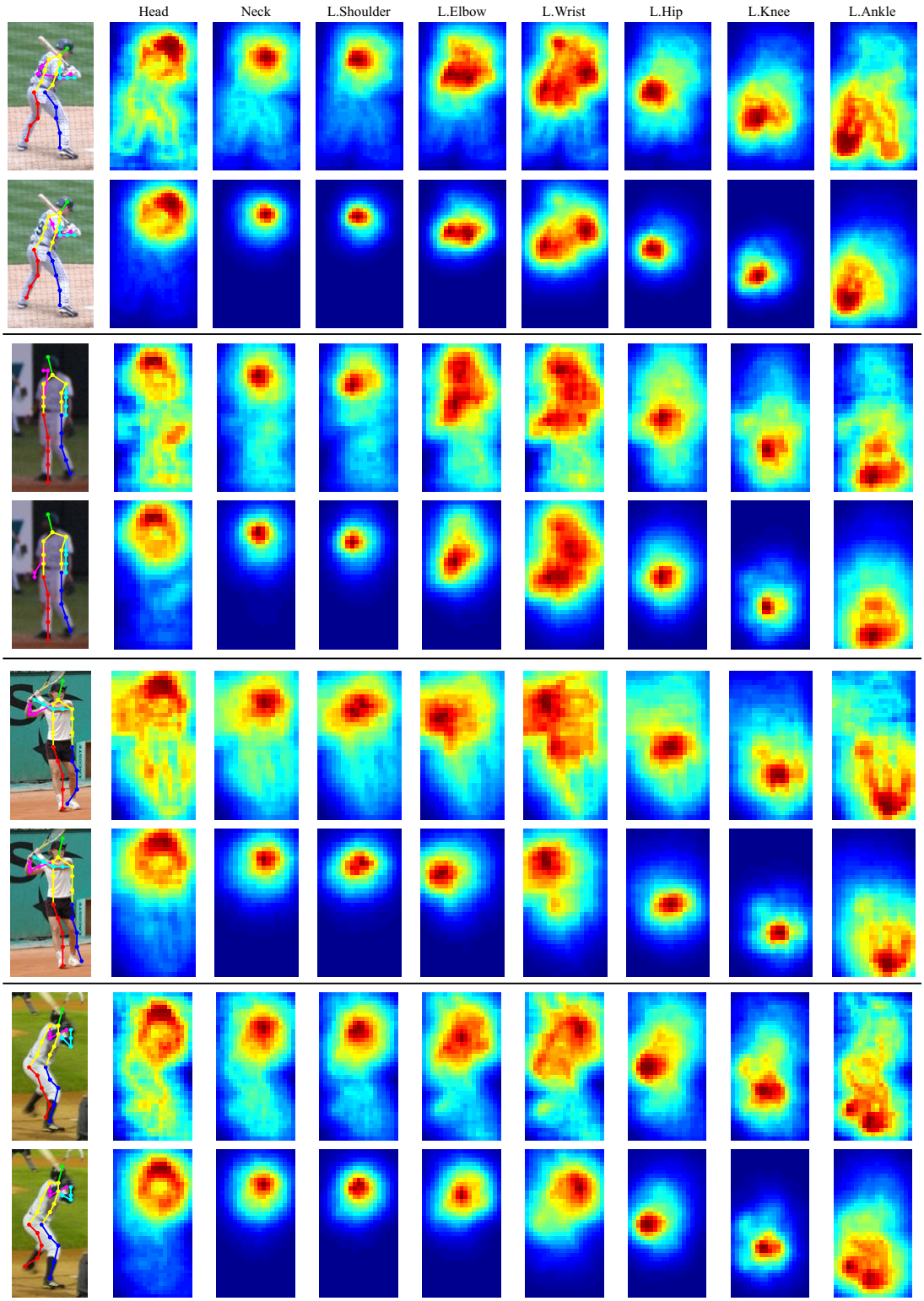

Figure 6. From left to right: predicted pose and the correspondent heat-maps for some exemplar joints. For each example, the first row visualizes the predicted pose and the heat-maps generated by the first message passing layer, and the second row shows the predicted pose and the heat-maps generated by the second message passing layer.