# Supplementary Material for
# Does Learning Specific Features for Related Parts Help Human Pose Estimation?

Wei Tang and Ying Wu
Northwestern University
2145 Sheridan Road, Evanston, IL 60208
{wtt450, yingwu}@eecs.northwestern.edu

## 1. More Qualitative Results

Fig. 1 and Fig. 2 respectively display some pose estimation results obtained by an eight-stack PBN on the MPII [1] and LSP [2] datasets.

We collect a few short videos from YouTube to test whether our approach can generalize beyond the MPII and LSP datasets. They contain different styles of dance performances, *e.g*., Latin, hip hop, ballet and contemporary. A Faster RCNN [3, 4] is utilized to locate the dancers. Then we apply an eight-stack PBN trained on the MPII dataset for frame-by-frame pose estimation (single-scale testing with flipping). No pose tracking was performed. Using one NVIDIA TITAN X GPU (12 GB memory), it can infer the poses of 23.74 persons per second. Fig. 3 shows our results on some frames. The file 'demo_with_sound.mp4'[1] demonstrates the pose estimates obtained by our approach on some video clips. Most of them have been sped up by a factor of 1.5 or 2 to shorten the demo.

## 2. Failure Cases

We use the MPII validation set to analyze the failure cases of our approach. Most of them are inaccurate part localizations due to occlusion. Some examples are shown in Fig. 4 (a)(b)(c)(d). We can see most of these wrong estimates are plausible. A few other failures are caused by motion blur, overlapping persons, rare part appearance and incorrect ground truth, respectively illustrated in Fig. 4 (e)(f)(g)(h).

## 3. Detailed Part Grouping Results

In Sec. 4.3 of the paper, we have compared the human pose estimation performances when there are 2, 5, 8 or 16 groups of related parts. Their corresponding part grouping results are detailed in Tab. 1.

[1] http://www.ece.northwestern.edu/~wtt450/project/CVPR19_PBN

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

[3] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NIPS*, 2015. 1

[4] J. Yang, J. Lu, D. Batra, and D. Parikh. A faster pytorch implementation of faster r-cnn. *https://github.com/jwyang/faster-rcnn.pytorch*, 2017. 1

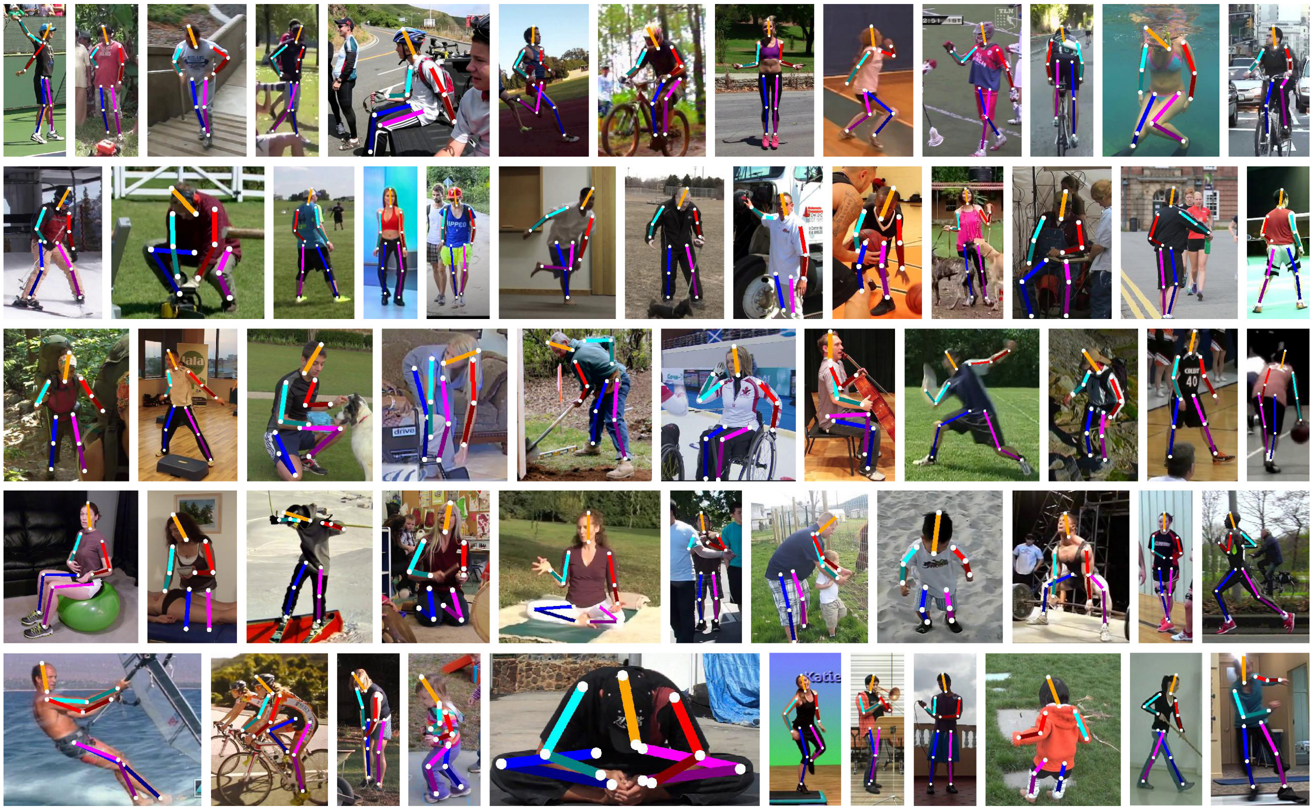

Figure 1. More human pose estimation results obtained by our approach on the MPII dataset.

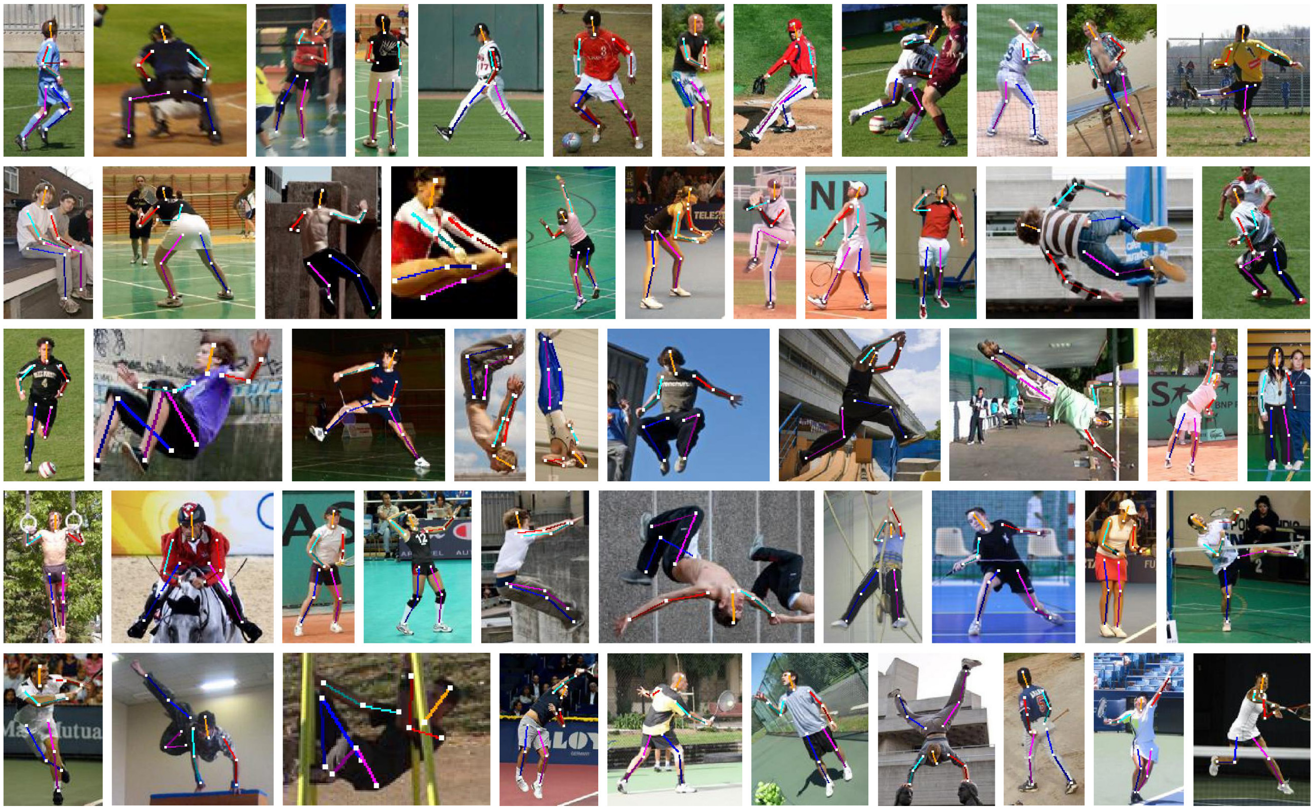

Figure 2. More human pose estimation results obtained by our approach on the LSP dataset.

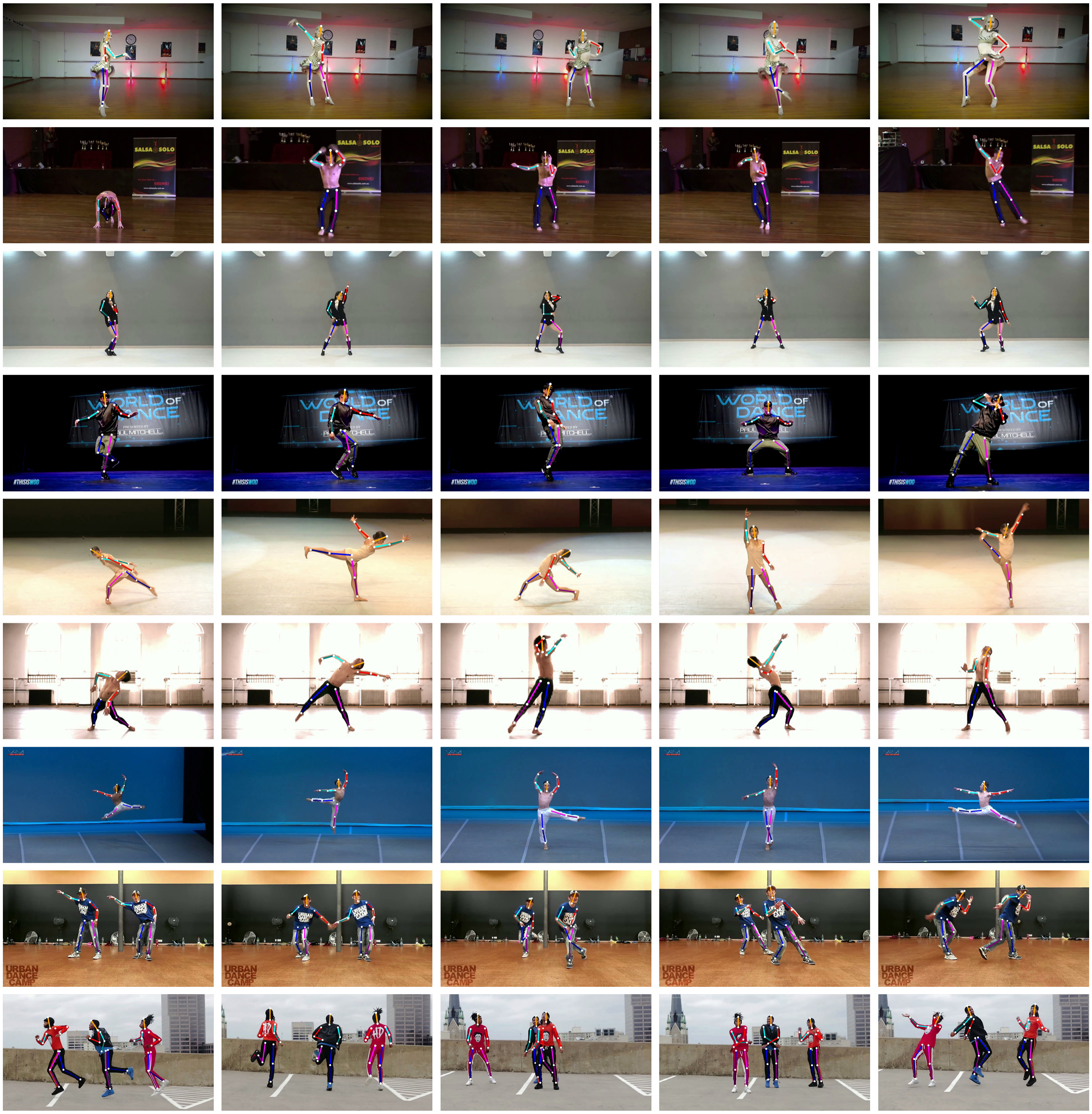

Figure 3. Human pose estimation results obtained by our approach on some dance video frames.

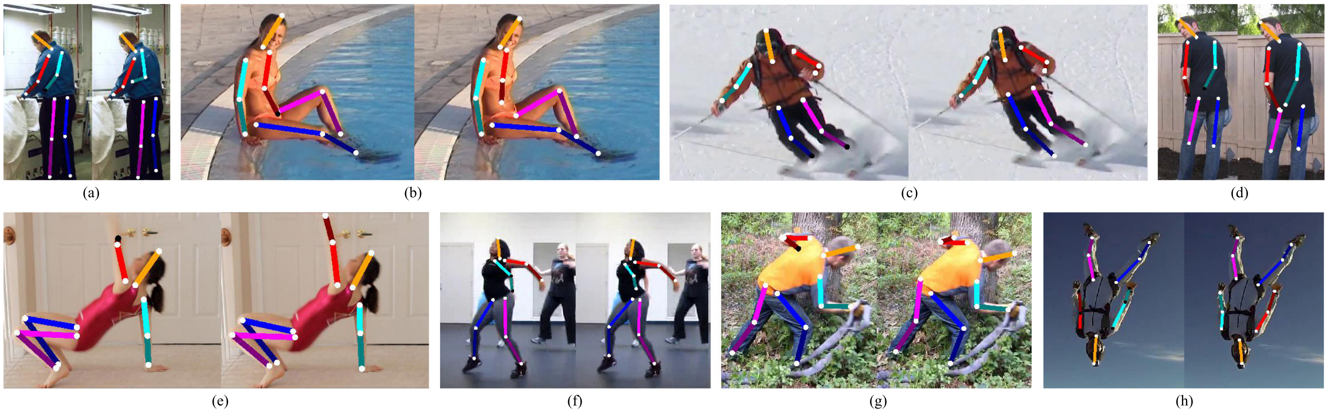

Figure 4. Failure cases of our approach on the MPII validation set. For each pair of images, the left and the right are respectively our result and the ground truth. Incorrect part localizations are denoted by black dots. Most of them are due to occlusion, *i.e.*, (a)(b)(c)(d). A few other failures are caused by (e) motion blur, (f) overlapping persons, (g) rare part appearance and (h) wrong ground truth.

| #Groups | Part grouping result |
|---|---|
| 2 | (1) Right and left ankles, right and left knees, right and left hips, and pelvis<br>(2) Thorax, upper neck, head top, right and left wrists, right and left elbows, and right and left shoulders |
| 5 | (1) Right and left ankles, and right and left knees<br>(2) Right and left hips, and pelvis<br>(3) Thorax, upper neck, head top, and right and left shoulders<br>(4) Right wrist and right elbow<br>(5) Left wrist and left elbow |
| 8 | (1) Right ankle and right knee<br>(2) Left ankle and left knee<br>(3) Right and left hips, and pelvis<br>(4) Thorax and left shoulder<br>(5) Upper neck and head top<br>(6) Right wrist and right elbow<br>(7) Left wrist and left elbow<br>(8) Right shoulder |
| 16 | Each body part is a group. |

Table 1. Part grouping results corresponding to different group numbers.