XingGAN for Person Image Generation –Supplementary Material–

Hao Tang[1,2], Song Bai[2], Li Zhang[2], Philip H.S. Torr[2], and Nicu Sebe[1,3]

[1]University of Trento (hao.tang@unitn.it)
[2]University of Oxford [3]Huawei Research Ireland

This document provides additional experimental results on the person image generation task. First, we compare the proposed XingGAN with the most state-of-the-art method Pose-Transfer [3] (Sec. 1). Additionally, we show more ablation results of the proposed XingGAN (Sec. 2). Lastly, we also provide the visualization results of the generated co-attention maps (Sec. 3).

1 State-of-the-Art Comparisons

In Fig. 1 and 2, we provide more generation results of the proposed XingGAN and Pose-Transfer [3] on both the Market-1501 [2] and DeepFashion [1] datasets. Note that we generated the results of Pose-Transfer [3] using the well-trained models provided by the authors[1] for fair comparisons. We observe that the proposed XingGAN consistently achieves photo-realistic results with fewer visual artifacts than Pose-Transfer on both challenging datasets.

2 More Ablation Results

In Fig. 3, we provide more qualitative ablation comparisons of the proposed XingGAN on Market-1501. These results further demonstrate the advantage of each component of the proposed XingGAN. Moreover, we observe that our full model consistently generates more coherent and natural person images.

In Fig. 4, we show the results of varying the number of the proposed Xing blocks on Market-1501. We observe that adopting about 9 Xing blocks makes the generated person images more natural and realistic, revealing the benefits of our progressive generation strategy.

3 Visualization of Co-Attention Maps

We also provide more visualization results of the generated co-attention maps and intermediate results in Fig. 5. We show 10 randomly chosen intermediate results, their corresponding 10 co-attention maps, and the input attention map. It is clear that these co-attention maps have learned different activated content between the generated intermediate results and the input image for generating the final person images, revealing the effectiveness of the proposed co-attention fusion module.

[1] https://github.com/tengteng95/Pose-Transfer

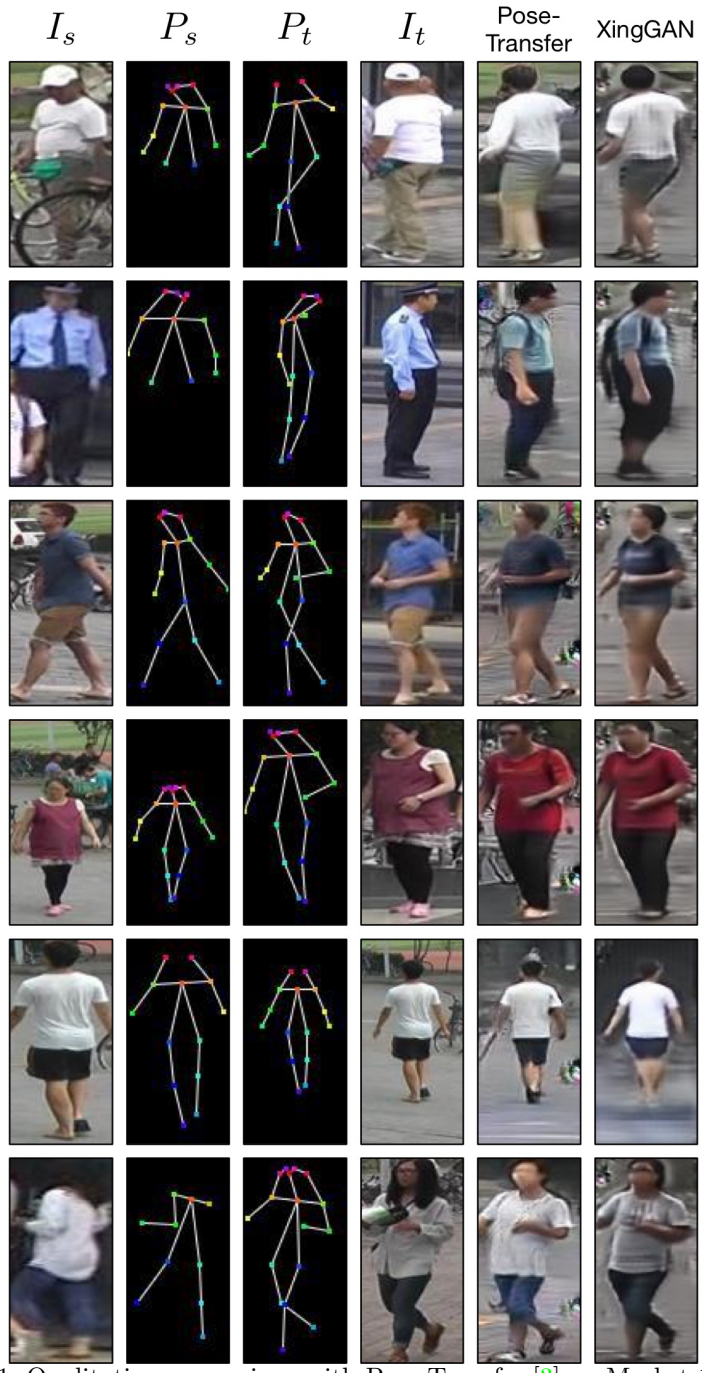

Fig. 1: Qualitative comparison with Pose-Transfer [3] on Market-1501.

Fig. 2: Qualitative comparison with Pose-Transfer [3] on DeepFashion.

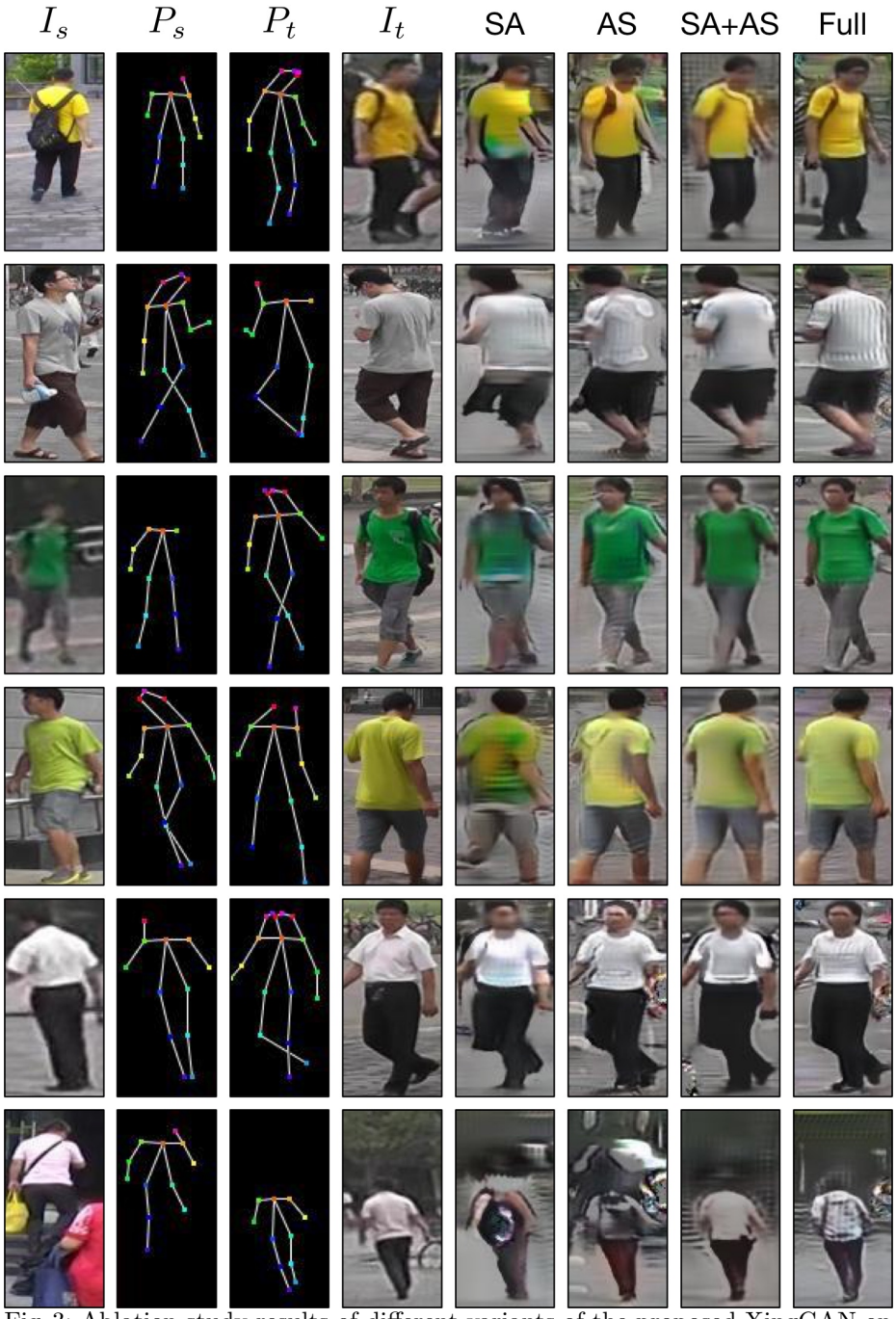

Fig. 3: Ablation study results of different variants of the proposed XingGAN on Market-1501.

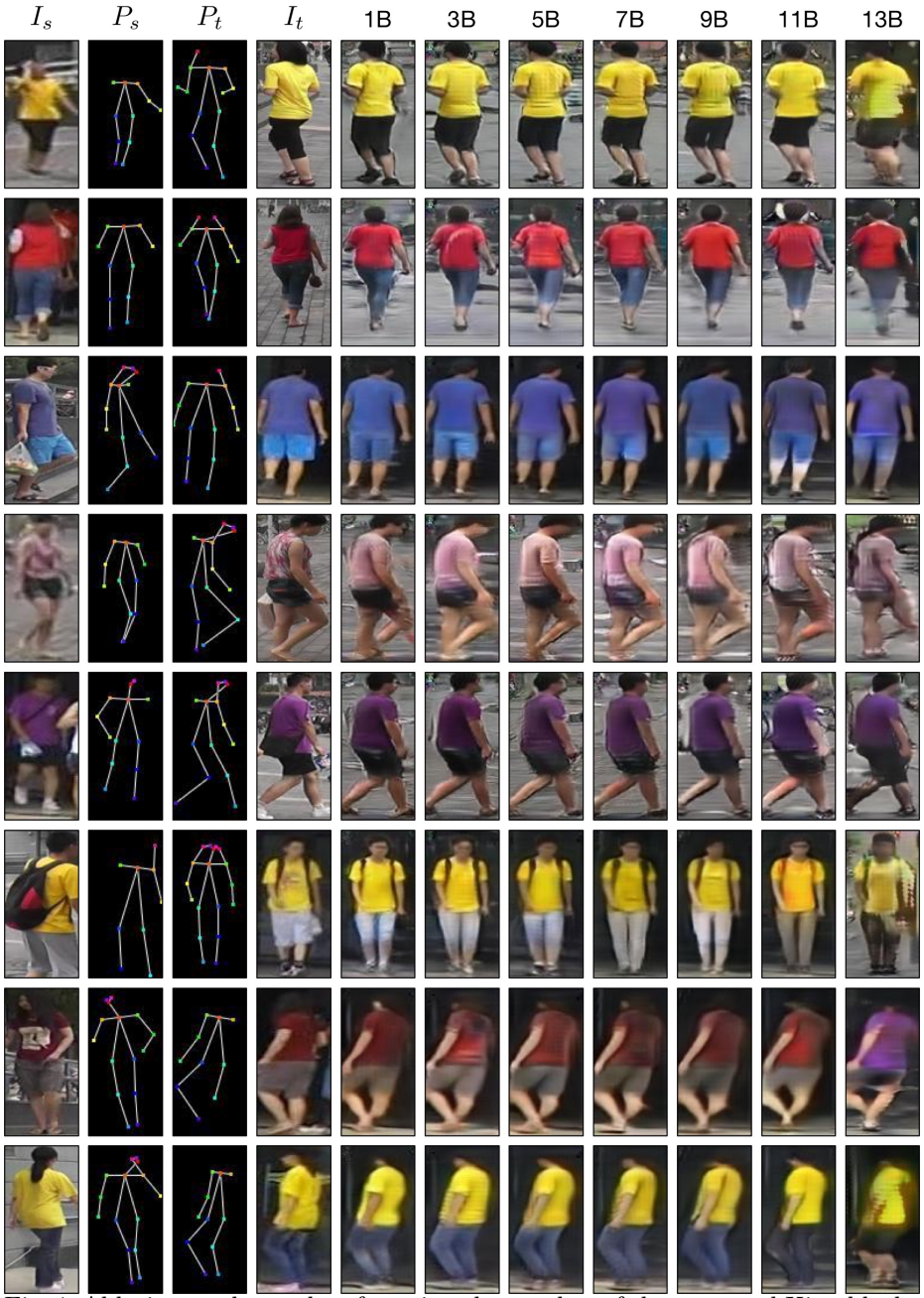

Fig. 4: Ablation study results of varying the number of the proposed Xing blocks on Market-1501. 'B' stands for the proposed Xing Blocks.

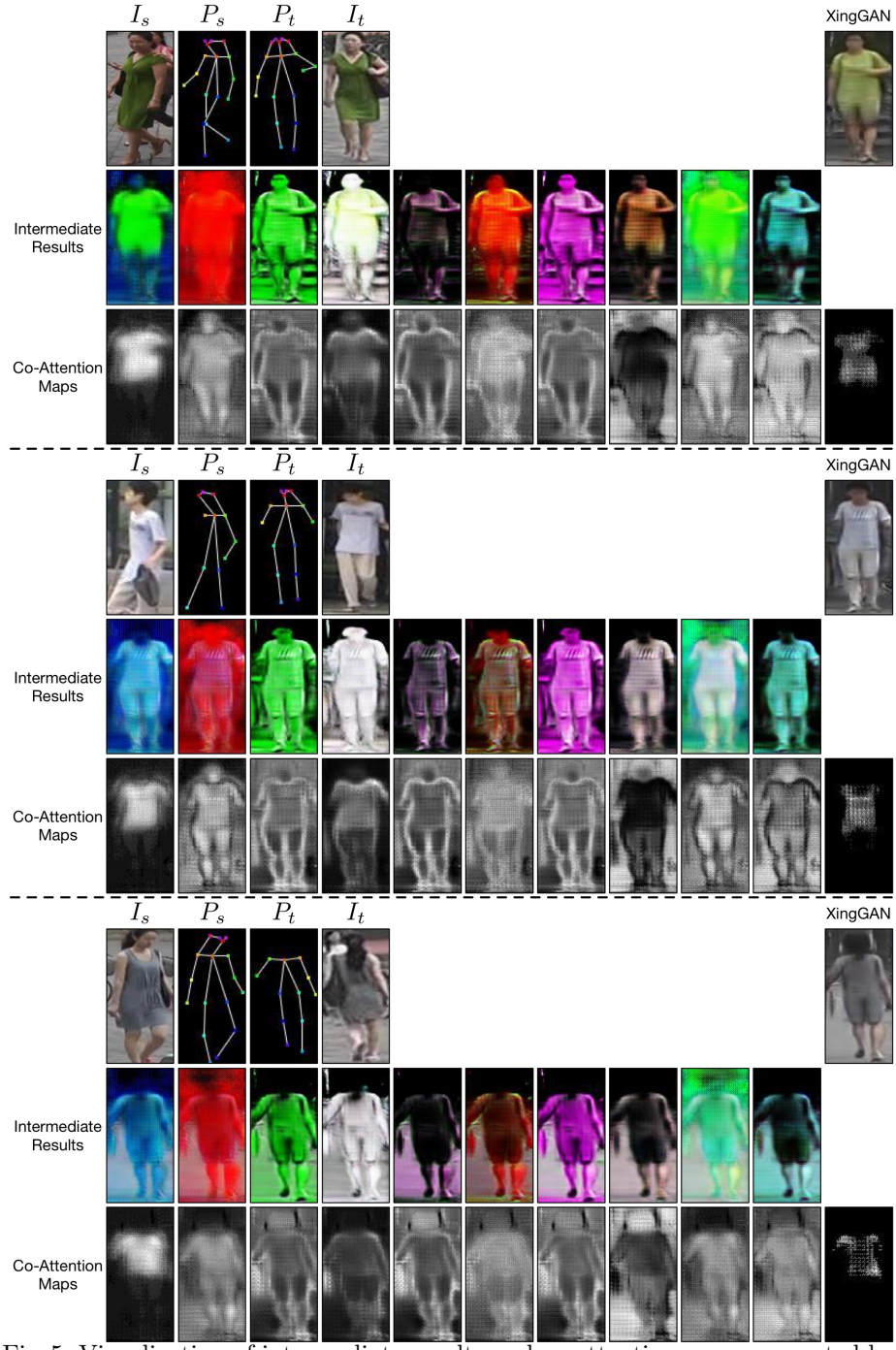

Fig. 5: Visualization of intermediate results and co-attention maps generated by the proposed XingGAN on Market-1501.

References

1. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: CVPR (2016) 1
2. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV (2015) 1
3. Zhu, Z., Huang, T., Shi, B., Yu, M., Wang, B., Bai, X.: Progressive pose attention transfer for person image generation. In: CVPR (2019) 1, 2, 3