Multistage Adversarial Losses for Pose-Based Human Image Synthesis

Wenjie Niu

June 28,2018

Abstract

The paper proposed a pose-based human image synthesis method which can keep the human posture unchanged in novel viewpoints. And they adopt multistage adversarial losses seperately for the foreground and background generation, which fully exploits the multi-modal characteristics of generative loss to generate more realistic looking images.

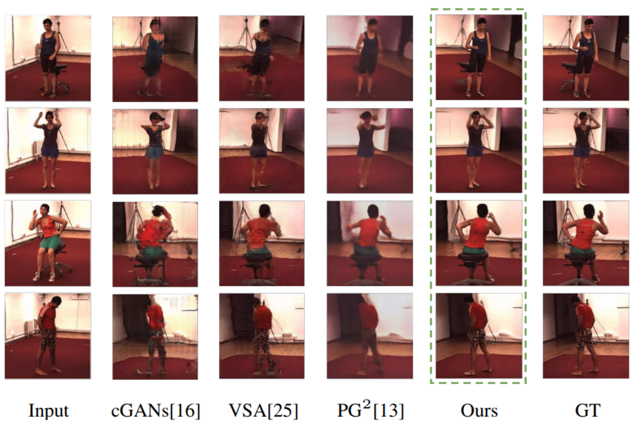

Figure 2. Human image synthesis by our proposed method and three state-of-the-art methods: cGANs [2], VSA [5], PG2 [1]. [4]

1. Introduction

The method(shown in Fig. 1) keeps human posture unchanged during generating novel view images from a single image. Fig. 1 shows a procedure of human image synthesis, which contains three thansformer networks for three stages. (1) In the first stage, the paper propose a pose transformer network which can synthesis 2D target pose P_t^* of other perspectives from the condition pose P_s corresponding to the condition image I_s. (2) In the second stage, the authors extract human transformer network to synthesis the target human foreground F_s from condition image with the segmentation method CFR-RNN [6]. (3) In the third stage, a bachground transformer network is proposed to generated target full image I_t^* with the condition image I_s and the generated foreground image F_t^* as the input.

Table 1. The comparison results between our method and the other state-of-the-art methods.[4]

Methods	SSIM	PSNR
Mirza*et al.* [2](cGANs)	0.52	17.05
Villegas*et al.* [5](VSA)	0.54	17.52
Ma*et al.* [1](PG2)	0.60	19.19
Ours	0.72	20.62

Fig. 2 shows the comparision results between the method and three state-of-the-art approches [2],[5],[1]. The image show much better foreground and bach ground images than the other methods.

2. Conclusions

As can be seen, in this paper it focus on human image synthesis and do not apply the results on other visual tasks. They will further improve the image quality and apply the generated images on various visual tasks, *e.g.*cross-view gait recognition and person re-identification.

References

[1] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. V. Gool. Pose guided person image generation. In *Conference on Neural Information Processing Systems*, 2017. 1

[2] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 1

[3] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, 2016. 2

[4] C. Si, W. Wang, L. Wang, and T. Tan. Multistage adversarial losses for pose-based human image synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2

[5] R. Villegas, J. Yang, Y. Zou, S. Sohn, X. Lin, and H. Lee. Learning to generate long-term future via hierarchical prediction. In *International Conference on Machine Learning*, 2017. 1

[6] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr. Conditional ran-

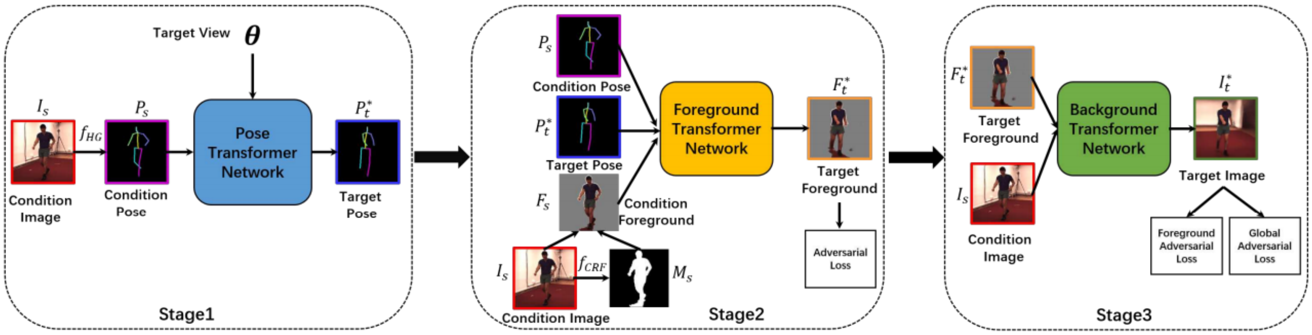

Figure 1. The overall pipeline of our multistage approach which contains three transformer networks for three stages. In the first stage, the pose transformer network synthesizes a novel view 2D pose. Then, the foreground transformer network synthesizes the target foreground image in the second stage. Finally, the background transformer network generates the target image. f_{HG} and f_{VRF} donate the stacked hourglass networks [3] and the CRF-RNN [6] for pose estimation from image and foreground segmentation, respectively.[4]

dom fields as recurrent neural networks. In *IEEE International Conference on Computer Vision*, 2016. 1, 2