Supplementary Material: Towards Multi-pose Guided Virtual Try-on Network

Haoye Dong[1], Xiaodan Liang[1],Xiaohui Shen[4], Bochao Wang[1],
Hanjiang Lai[1], Jia Zhu[2], Zhiting Hu[3], Jian Yin[1]*
[1]SYSU, [2]SCNU, [3]CMU, [4]ByteDance AI Lab.

1. More Ablation Study

We conduct an additional ablation study on MPV to explore the effects of each parameter, as illustrated in Table 1 and Table 2. For the Warp-GAN, the w/o adversarial loss achieves the best SSIM but obtains lower IS score. For the Refinement Render, the w/o mask loss achieves the best SSIM but obtains the lower IS score. The w/o perceptual loss achieves the best IS score but obtains the lower SSIM. Only the MG-VITON (full) can achieve the competitive score of both SSIM and IS score. The results show that the parameters are all important to enhance performance.

Table 1. Ablation Study of Warp-GAN on MPV Dataset.

Model	SSIM	IS
VITON [1]	0.640	2.394 ± 0.205
CP-VTON [3]	0.705	2.519 ± 0.107
MG-VTON (w/o adversarial loss)	**0.752**	2.969 ± 0.136
MG-VTON (w/o perceptual loss)	0.737	3.016 ± 0.123
MG-VTON (w/o feature loss)	0.741	2.934 ± 0.132
MG-VTON (w/o L1 loss)	0.744	2.970 ± 0.143
MG-VTON (full)	0.744	**3.154 ± 0.142**

Table 2. Ablation Study of Refinement Render on MPV Dataset.

Model	SSIM	IS
VITON [1]	0.640	2.394 ± 0.205
CP-VTON [3]	0.705	2.519 ± 0.107
MG-VTON (w/o perceptual loss)	0.733	**3.307 ± 0.137**
MG-VTON (w/o mask loss)	**0.756**	2.857 ± 0.120
MG-VTON (full)	0.744	3.154 ± 0.142

2. More Visual Comparison Results

VITON and CP-VTON aim at fitting new clothes into a person which are close to our MG-VTON, but they mainly focus on the fixed pose and fail to preserve the fine details. Different from DeformableGAN [2], our Warp-GAN exploits the synthesized target parsing and uses the TPS method to learn warping features from residuals is helpful to capture intrinsic geometric properties and correlations. DeformableGAN decomposes the person by using group keypoints in a coarse-level and use the affine transformation method directly. CP-VTON only use the front person image and fail to warp clothes image well with different angles. We denote DeformableGAN + CP-VTON as that first apply DeformableGAN to change the pose then try on clothes by using CP-VTON. As shown in Figure 1, DeformableGAN + CP-VTON is lack of capacity to warp clothes and preserve the details.

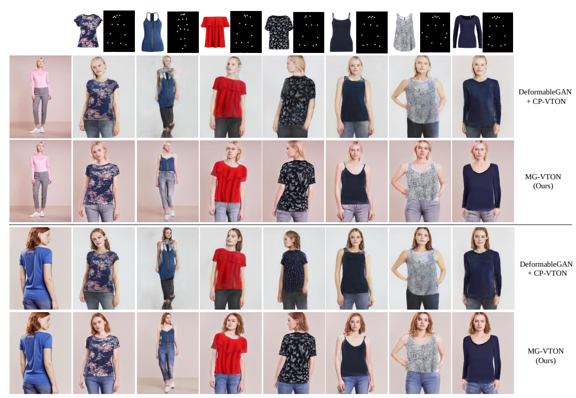

Figure 1. Visual comparison with DeformableGAN + CP-VTON. The first column is the reference images. Zoom in for details.

3. More results of our MG-VTON

Figure 2. Test results of our MG-VTON on MPV dataset.

*Corresponding author is Jian Yin

Figure 3. Test results of our MG-VTON on MPV dataset.

Figure 4. Test results of our MG-VTON, trained on MPV dataset, test on DeepFashion dataset.

References

[1] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *CVPR*, 2018. 1

[2] Aliaksandr Siarohin, Enver Sangineto, Stephane Lathuiliere, and Nicu Sebe. Deformable gans for pose-based human image generation. In *CVPR*, 2018. 1

[3] Bochao Wang, Huabin Zhang, Xiaodan Liang, Yimin Chen, and Liang Lin. Toward characteristic-preserving image-based virtual try-on network. In *ECCV*, 2018. 1