6. Pseudocode

Our method contains a parser-based network PB-AFN and a parser-free network PF-AFN. We first introduce the training process of the parser-based network PB-AFN in Alg. 1 following the existing training pipeline [30, 8, 18, 32], which masks the clothing region of the person image and reconstructs the person image with the corresponding in-shop clothes and the person representations. After training PB-AFN, we use the generated fake images as the input of PF-AFN, which is supervised by real images, and further distill the appearance flows to find accurate correspondences between the target clothes and the person image. The complete training process of PF-AFN is presented in Alg. 2. During inference, only a target clothes image and a reference person image will be given to PF-AFN to generate the try-on results.

7. More Try-on Results

Results of VITON-HD Different from VITON [9] dataset with low resolution 256×192, VITON-HD with the image resolution 512×384 poses a severe challenge for the model to retain the details of the target clothes and generate high-resolution try-on images with satisfactory visual quality. The results on VITON-HD is shown in Fig. 8. Since VITON-HD hasn’t been tackled before by previous methods [30, 8, 18, 32], we up-sample their low-resolution results to 512×384. Compared with the parser-based methods, our PF-AFN is able to warp the target clothes to the reference person seamlessly where the logo and the embroidery are retained without being distorted, preserve non-target clothes such as trousers and skirt, and keep photo-realistic body details like hands and fingers.

Results of MPV We show more try-on results on MPV [5] dataset in Fig. 9. Compared with WUTON [13] which is also a parser-free method, our PF-AFN generates try-on images with much better visual quality. PF-AFN yields accurate warping for the target clothes, preserves the details of body parts (i.e., hands) even in complex postures and retains the characteristics of the target clothes (i.e., color, collar and sleeve).

Results of MPV We here show extensive try-on results produced by our PF-AFN on VITON [9] dataset. Fig. 10 shows four reference persons with different target clothes and Fig. 11 shows four target clothes to different persons. Our PF-AFN can adapt to different kinds of clothes with satisfactory performance. It also performs robustly with various poses and generates high-quality results.
Figure 8. Visual comparison on VITON-HD dataset with the image resolution 512 × 384. Our model is capable of generating high-resolution try-on images, where details of the target clothes (i.e., collar, logo and embroidery), non-target clothes (i.e., trousers and skirt) and body parts (i.e., hands and fingers) are very well retained, compared with the recent proposed parser-based methods [30, 8, 18, 32].
Figure 9. Extensive visual comparison on MPV dataset with parser-free inputs. Compared with WUTON [13], our PF-AFN generates try-on images with much better visual quality, which warps the target clothes to the corresponding region on the person accurately, preserves the details of body parts (i.e. hands) even in complicated poses and retains the characteristics of the target clothes (i.e. color, collar and sleeve).
Figure 10. Extensive try-on results of four reference persons with different target clothes on VITON dataset. Our PF-AFN achieves satisfactory performance with various clothes such as short sleeve top, long sleeve top, vest, sling, high-necked top and polo top, where the characteristics of the target clothes are retained without being interfered by the original clothes on the reference person. Also, the texture, logo and embroidery on the target clothes are preserved without being distorted.
Figure 11. Extensive try-on results of four target clothes to different reference persons on VITON dataset. Our PF-AFN performs robustly with various poses including arms akimbo, two hands blocking in front of the body, cross-arms and one arm behind the back, where large misalignment and deformation occurs between the target clothes and the reference person. Our PF-AFN warps the target clothes to different persons accurately and generates high-quality results.