Supplementary Material for SDAFN

Shuai Bai, Huiling Zhou, Zhikang Li, Chang Zhou, and Hongxia Yang

DAMO Academy, Alibaba Group, China
{baishuai.bs,zhule.zhl,zhikang.lzk,ericzhou.zc,yang.yhx}
@alibaba-inc.com

In this supplementary material, we provide additional visualization results and implementation details. Firstly, the overall framework of other tasks is shown in section 1. Secondly, in section 2, more detailed results about ablation study are exhibited. Then more qualitative results on four datasets are shown in section 3. Finally, some failure cases are shown.

1 Implementation Details of Other Tasks

We verify the versatility of the proposed deformable attention flow on two other image editing tasks, namely multi-view synthesis and images animation. To deal with only one image transformation (compared to paired images given in try-on task), we remove the self-MFE in DAFN from our model. Meanwhile, as shown in Fig. 1, the self-flow fields and self attention maps are also removed. Besides, the guidance Information is introduced into both of the source branch and the reference branch. Other details are the same as the virtual try-on task.

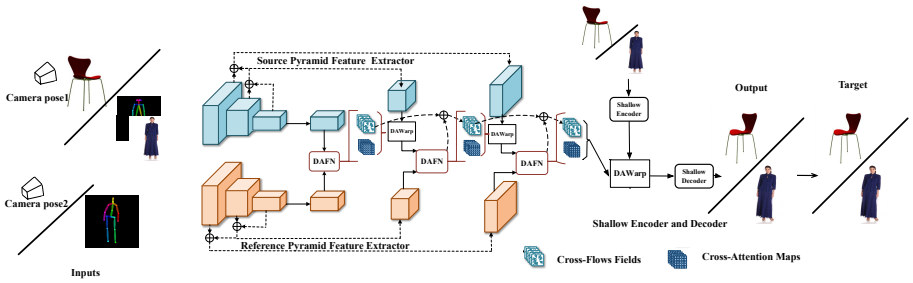

Fig. 1. The overall framework of our SDAFN in multi-view synthesis and images animation tasks.

2 Results about Ablation Study

Scalability to higher resolution. Without retraining at different resolution, as illustrated in Fig. 2, our model trained at 256×192 generates clear and photo-realistic results at 512×384.

Fig. 2. Comparisons with image interpolation and flow interpolation. flow interpolation has the advantages of scaling to higher resolution without retraining.

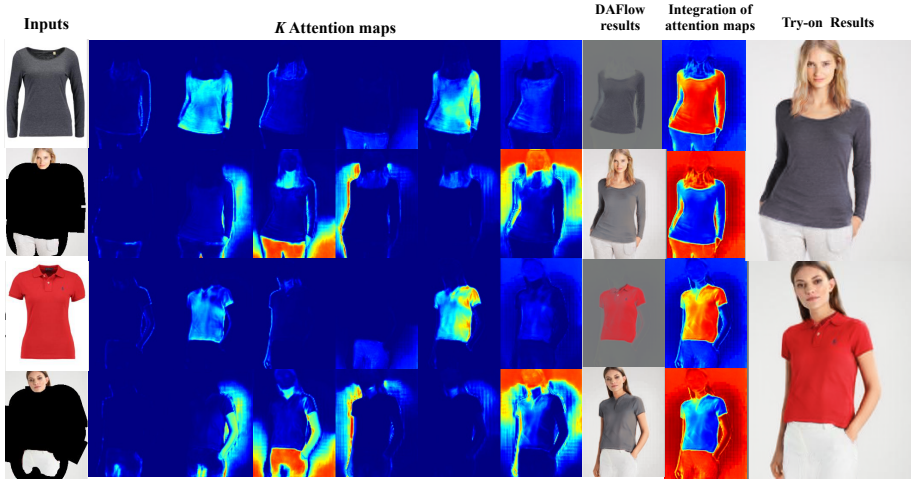

Fig. 3. Visualization of the K attention maps and results of deformable attention warping.

Fig. 4. Visualization of the sampling points in different positions. Multiple flow fields sample from different locations to extract structural or texture information. Reference points are in yellow. Sampling points are in red.

Visualization of Attention. More detailed attention maps and sampling points are visualized in Fig. 3. Different attention maps focus on different regions, such as hands, backgrounds, clothes and others. For the garment branch, the texture of the clothes is preserved with some edge shadows, such as the three-dimensional neckline. The reference person branch generates a reasonable torso as well as some folds. Our model is able to learn accurate structural information and predict the 3D priors to fit the skin and make the final try-on more realistic. As shown in Fig. 4, multiple flow fields sample from different locations to extract structural or texture information. In the first example, the generation of the contact position between the neckline and the neck needs to combine the skin color of the human body, the type of the neckline of the clothes, the color of the clothes, and even the white background to generate light and shadow effects.

3 Qualitative Results with SDAFN

We present some qualitative results of SDAFN on four datasets, including VITON, MPV, FashionVideo and ShapeNet in Fig. 5 to Fig. 8.

More Try-on Results. As illustrated in Fig. 5 and Fig. 6, our model generates satisfactory results faced with different kinds of clothes and various postures. It verifies the robustness of our method.

More Results on Other Image Editing Tasks. Example animations produced by our SDAFN on five actions are shown in Fig. 7, accurate reconstruction of the input pose is generated even in the case of complex motion like the back of the body. As shown in Fig. 8, our SDAFN is able to predict the one-shot input chair under different views with accurate textures. It successfully reconstructs the unseen parts which are under occlusion from the input image. These results demonstrate our DAFlow not only has the ability of image spatial transformation but also can be used for image generation.

4 Failure Cases

Fig. 9 shows some failure cases of our method. Although our model matches the clothes to the pose of the human body, there are also unreasonable estimates for some rare styles and poses. In future work, we will further strengthen the learning of clothing structure to avoid these situations.

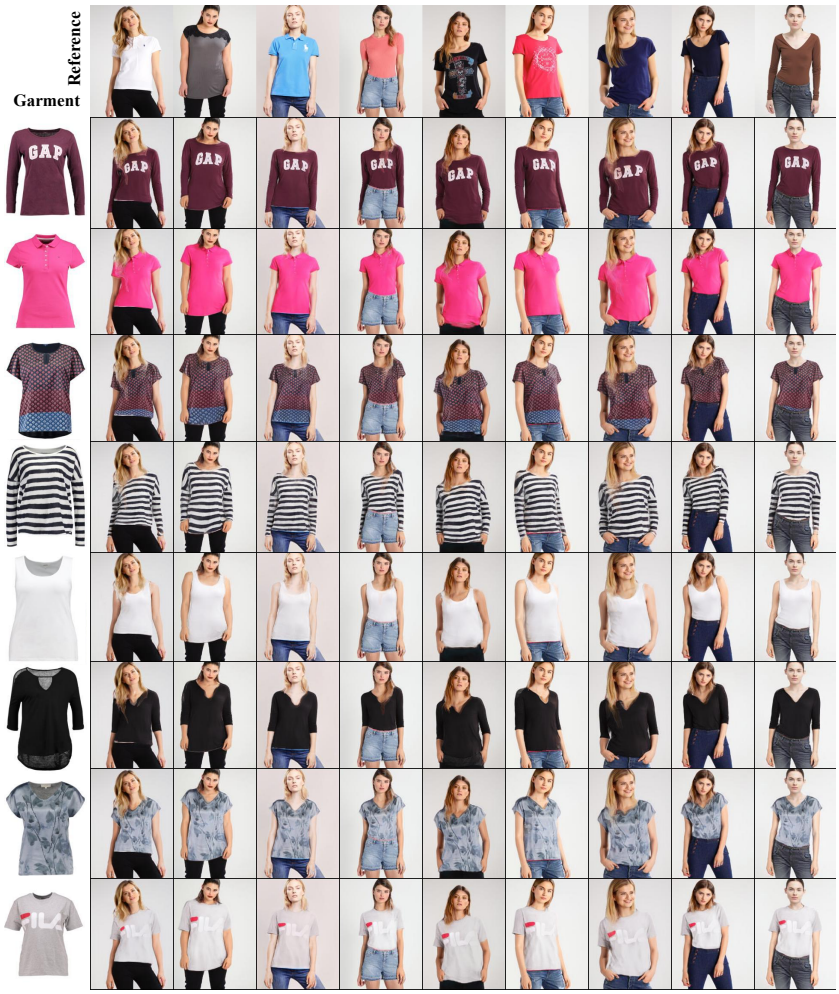

Fig. 5. Qualitative results on VITON dataset.

Fig. 6. Qualitative results on MPV dataset.

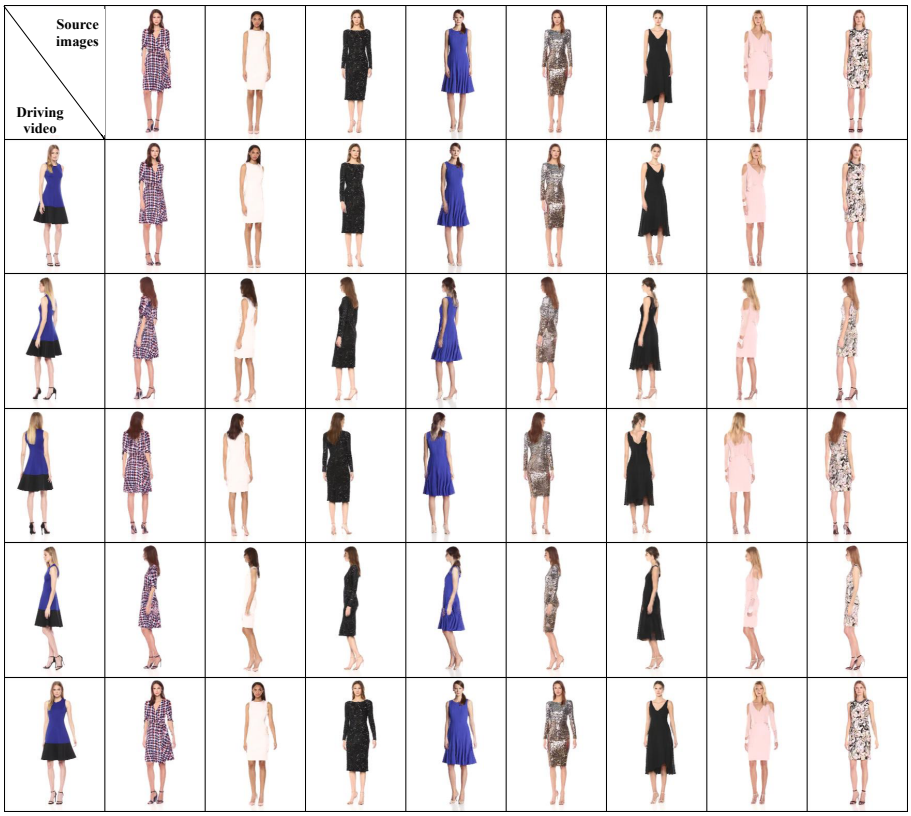

Fig. 7. Qualitative results on FashionVideo dataset.

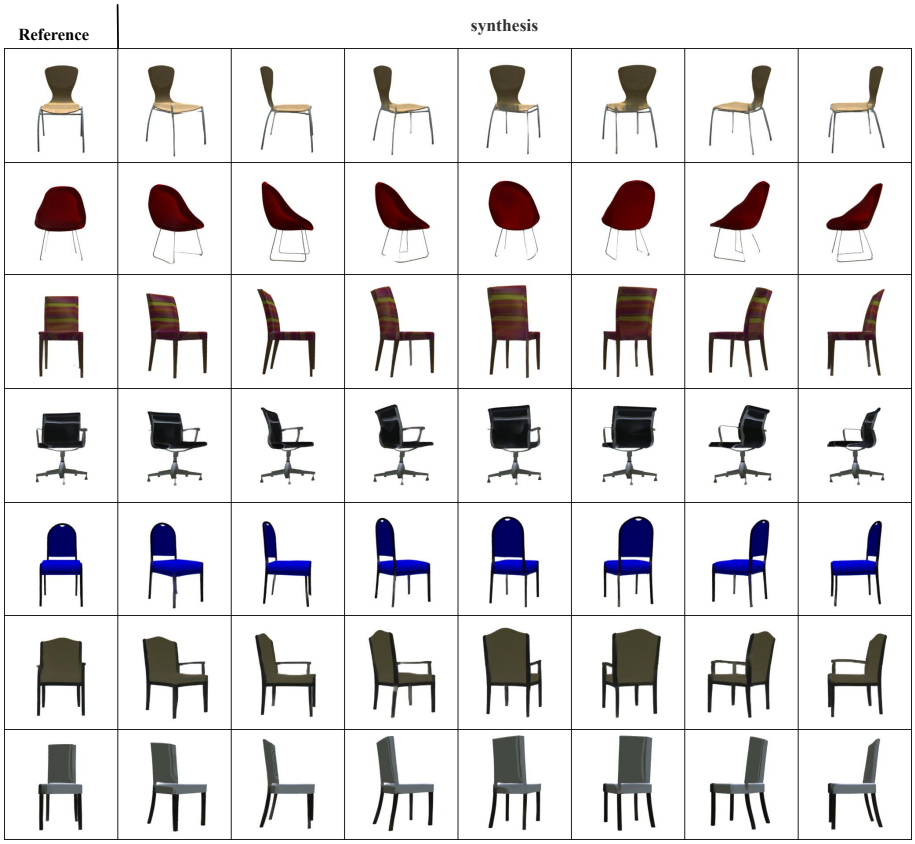

Fig. 8. Qualitative results on ShapeNet dataset.

Fig. 9. Failure cases.