APPENDIX

In this document, we provide additional materials to supplement our main text. In Appendix A, we show more visualization results, including a qualitative comparison of image try-on on DressCode dataset [30], a qualitative comparison of video try-on on VVT dataset [8] and additional video try-on results on TikTok dataset [21] generated by our WildVidFit framework. Then we show failure cases and discuss the limitations of our method in Appendix B.

A More Visualization Results

A.1 Image Try-on Results on DressCode

Fig. 11 provides qualitative results to demonstrate the superiority of our one-stage try-on network. Our network can effectively handle the garments with complex designs, accurately reproducing their features for impressive try-on results. This includes, for example, the multi-layered wrinkled sleeves in row 1 and the lace garment in row 2. Such designs are challenging to restore through warping, hence other methods fail to make reasonable predictions. Moreover, on striped garments, as in rows 3 and 4, other methods result in blurred outcomes, while our approach successfully maintains the textures. As demonstrated in Tab. **??**, the performance of our method significantly surpasses that of others.

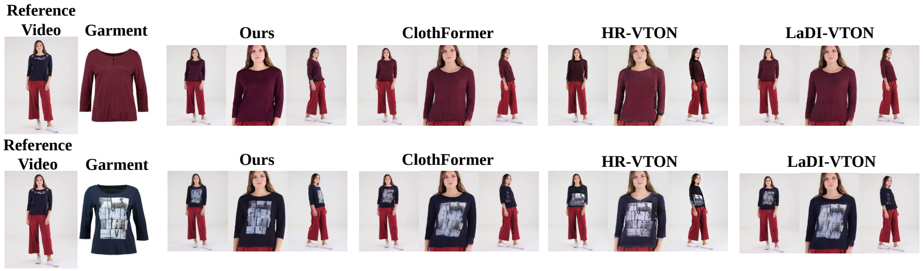

Fig. 8: Qualitative comparison with state-of-the-art methods ClothFormer [22], HR-VOTN [25] and LaDI-VTON [29] on the VVT dataset. Our model achieved robust results under various poses, thereby forming a coherent video sequence.

A.2 Qualitative Comparison on VVT Dataset

In the video try-on task, we compare our method on VVT dataset [8] against HR-VTON [25], LaDI-VTON [29] and ClothFormer [22]. As shown in Fig. 8, although our method has some difficulties with very fine details like small text on clothes due to implicit warping, it outperforms others in robustness when generating with various poses. In comparison, HR-VTON and ClothFormer tend to produce blurred textures in side views. In terms of temporal consistency, HR-VTON and LaDI-VTON naturally lag behind due to their lack of a temporal module, whereas our method performs on par with ClothFormer.

A.3 More Video Try-on Results on TikTok Dataset

Fig. 12 showcases more of our video virtual try-on results on the TikTok dataset [21]. Our method maintains a good appearance of the garments and is able to produce continuous, smooth videos. (**Notes**: Recommend to see the supplementary videos.) These examples further illustrate the effectiveness of our method in processing wild videos.

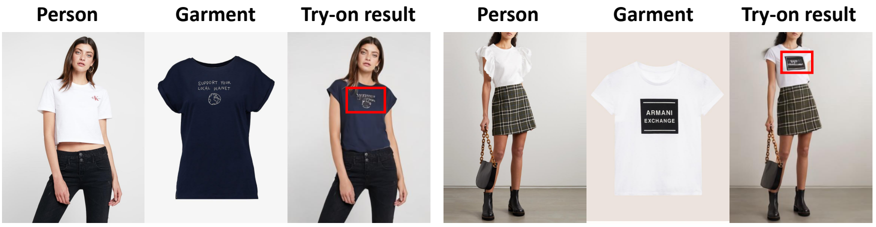

Fig. 9: Failure cases of image try-on results on DressCode (1st row) and VITON-HD (2nd row). Please zoom in to see differences highlighted by red boxes.

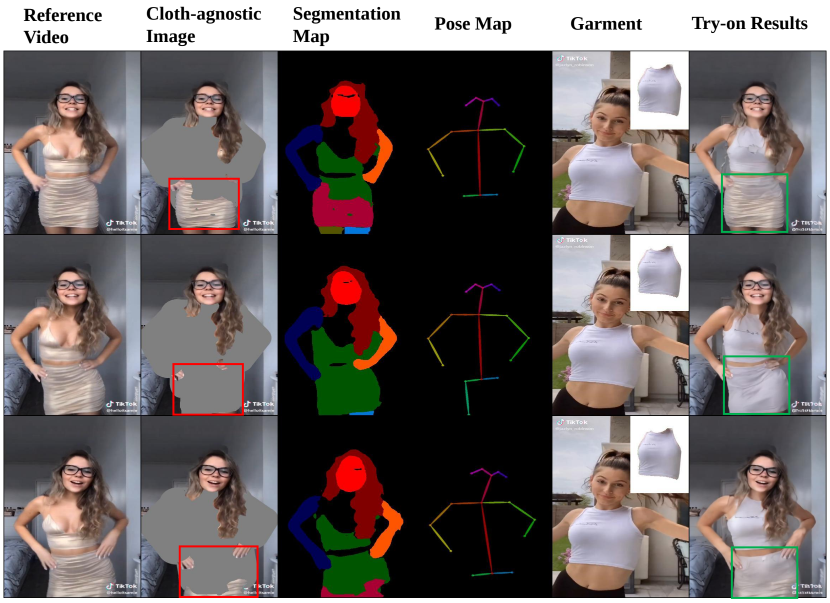

Fig. 10: Failure cases of inaccurate segmentation results. The inappropriate masked regions are highlighted by red boxes and the imperfect restoration of the lower garment caused by this is indicated by green boxes. For optimal viewing, please zoom in or see the supplementary videos.

B Failure Cases and Limitations

Despite some satisfying results, our method exhibits certain limitations. As illustrated in Fig. 9, our one-stage try-on network fails to reproduce tiny and complex patterns, such as little writing on the garment. This might stem from the implicit warping processed in the latent space, where overly detailed textures are lost during compression. Another limitation arises from inaccurate parsing results. As shown in Fig. 10, the segmentation algorithm fails to distinguish between upper and lower garment, resulting in unreasonable cloth-agnostic region where the lower garment is inappropriately masked (highlighted by red boxes). Such erroneous and discontinuous segmentation maps not only lead to imperfect restoration of the lower garment (highlighted by green boxes) but also adversely affect the virtual try-on result on the upper garment. For a clearer understanding, please refer to the supplementary video.

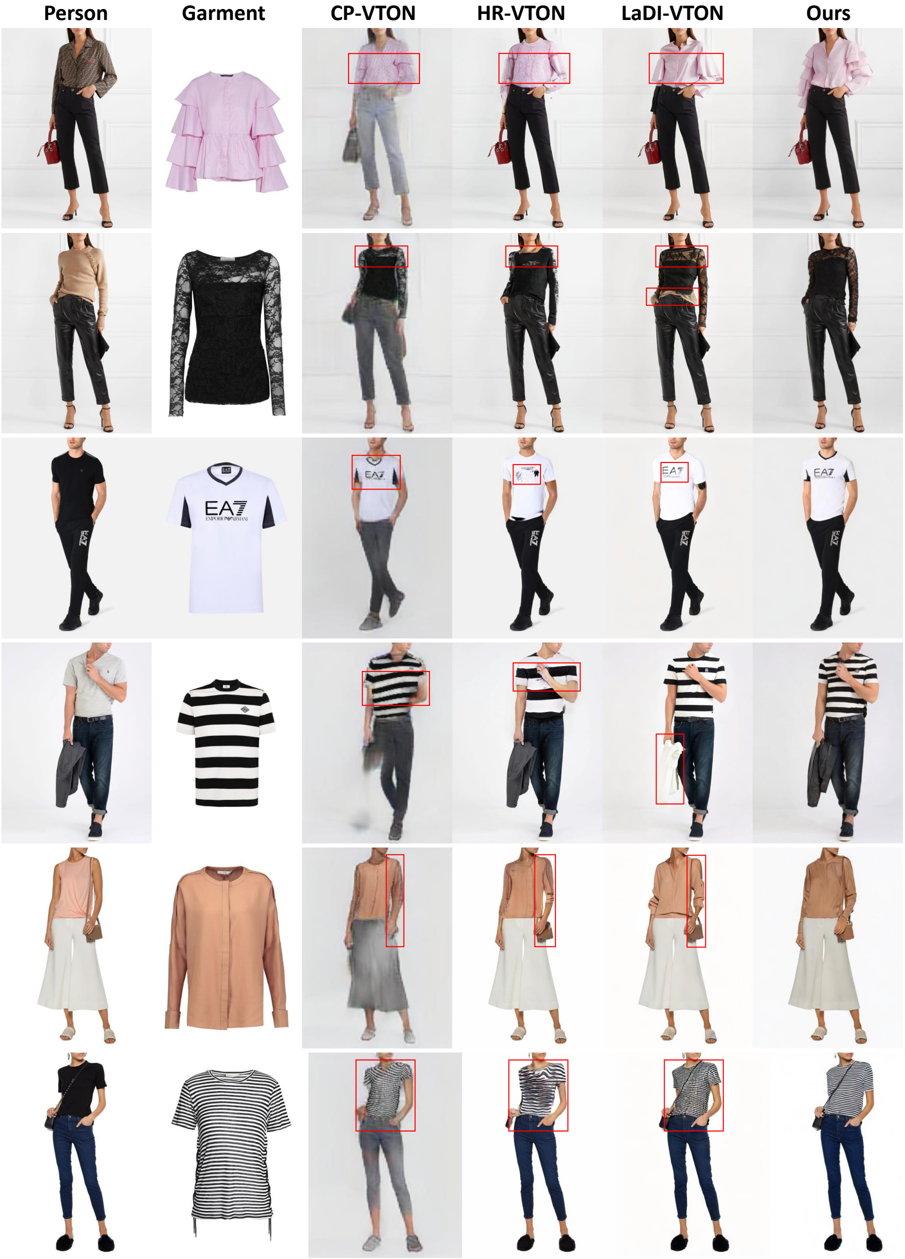

Fig. 11: Qualitative comparison on DressCode dataset. Please zoom in for best view.

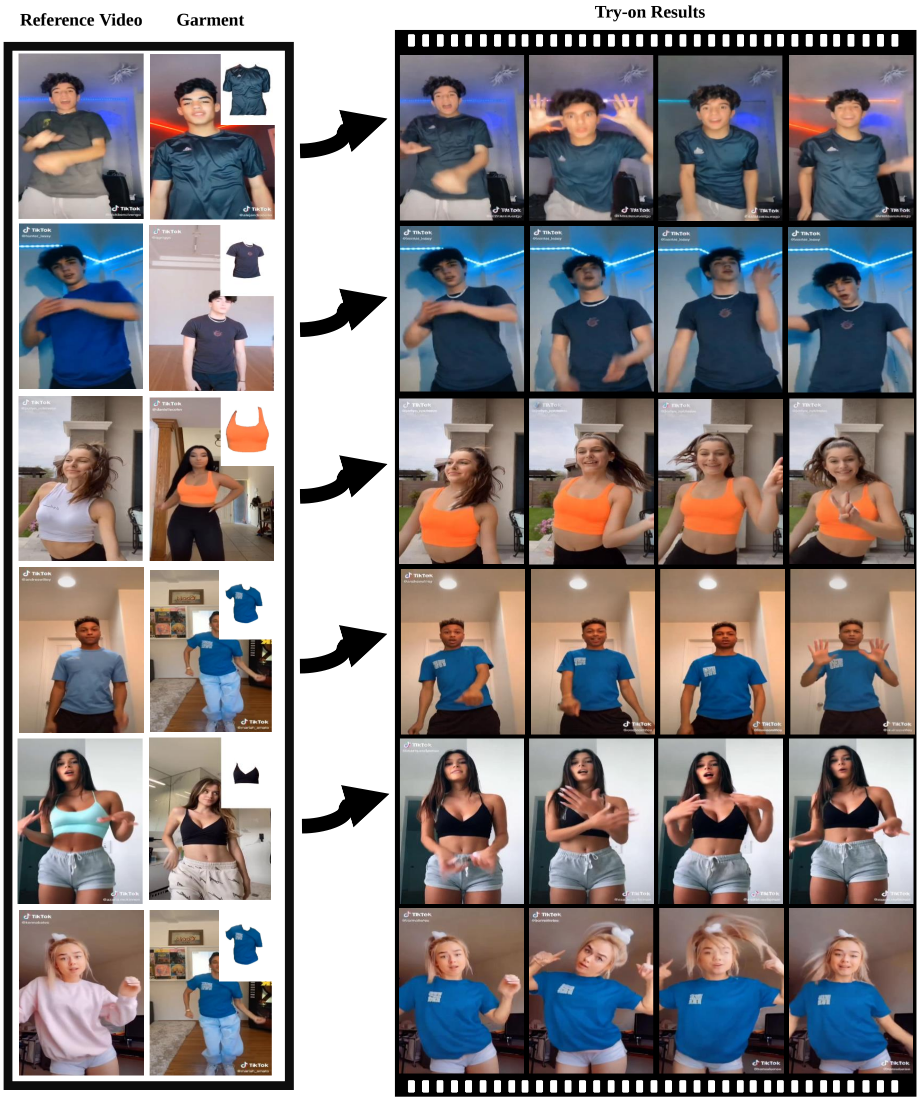

Fig. 12: More examples of our virtual try-on results on real-life TikTok videos. For optimal viewing, please zoom in or see the supplementary videos.