

# Supplementary Materials: GPD-VVTO: Preserving Garment Details in Video Virtual Try-On

Anonymous Authors

## 1 OUR DATASET

Visual comparison between the VVT [2] dataset and our collected dataset is illustrated in Figure 1. In the left panel, we can remark that the VVT dataset has a low spatial resolution ( $256 \times 192$ ), the video background is uniform, the person actions are repetitive, and it only includes upper-body garments, which deviates significantly from real scenes. In the right panel, our collected dataset has higher resolution (from  $960 \times 720$  to  $1280 \times 960$ ), diverse video backgrounds and person actions, and includes three types of garment categories: upper-body, lower-body, and dresses. Hence, our dataset can compensate for the shortcomings of current video virtual try-on datasets.

## 2 MORE QUALITATIVE RESULTS

### 2.1 Video Virtual Try-On

**More visualization results of our method.** We illustrate more video virtual try-on results generated by our method on the VVT [2] dataset and our dataset at a resolution of  $512 \times 384$  in the file named *1565\_video.mp4*. Our method ensures both the consistency of the garments and the continuity of the video under complex circumstances.

**Qualitative comparison with other methods.** We conduct qualitative comparison on video virtual try-on task with StableVITON [3] as shown in Figure 2. For fair comparison, we adapt StableVITON to a video virtual try-on model (namely StableVITON<sup>‡</sup>) by inserting temporal attention blocks. In the first row, StableVITON<sup>‡</sup> fails to preserve the style of the garment, introduces additional sleeves, and causes jittering along the edges of the garment. In the second row, the results generated by StableVITON<sup>‡</sup> exhibit significant pattern jittering on the garment. On the contrary, our method can generate more stable results that are more consistent with the target garment. Please refer to the file *1565\_video.mp4* for a more intuitive comparison.

**The video results shown in both the main text and supplementary materials can be found in the file named *1565\_video.mp4*.**

### 2.2 Image-based Virtual Try-On

Figure 3 shows the image-based virtual try-on results generated by our method on VITON-HD [1] dataset at a resolution of  $1024 \times 768$ . We demonstrate the results of different persons wearing the same garment and the same person wearing different garments. It can be observed that we successfully preserve the identity of persons and the details of garments across various situations.

Figure 4, 5 and 6 illustrate the image-based virtual try-on results generated by our method on Dress Code [4] dataset at a resolution of  $1024 \times 768$  given garments of different categories. Our method can generate high-quality try-on images on different categories of garments, further validating the generalization and robustness of our model.

## REFERENCES

- [1] Seunghwan Choi, Sunghyun Park, Minsoo Lee, and Jaegul Choo. 2021. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In *CVPR*.
- [2] Haoye Dong, Xiaodan Liang, Xiaohui Shen, Bowen Wu, Bing-Cheng Chen, and Jian Yin. 2019. Fw-gan: Flow-navigated warping gan for video virtual try-on. In *ICCV*.
- [3] Jeongho Kim, Gyojung Gu, Minho Park, Sunghyun Park, and Jaegul Choo. 2023. StableVITON: Learning Semantic Correspondence with Latent Diffusion Model for Virtual Try-On. *arXiv preprint arXiv:2312.01725* (2023).
- [4] Davide Morelli, Matteo Fincato, Marcella Cornia, Federico Landi, Fabio Cesari, and Rita Cucchiara. 2022. Dress code: high-resolution multi-category virtual try-on. In *CVPR*.

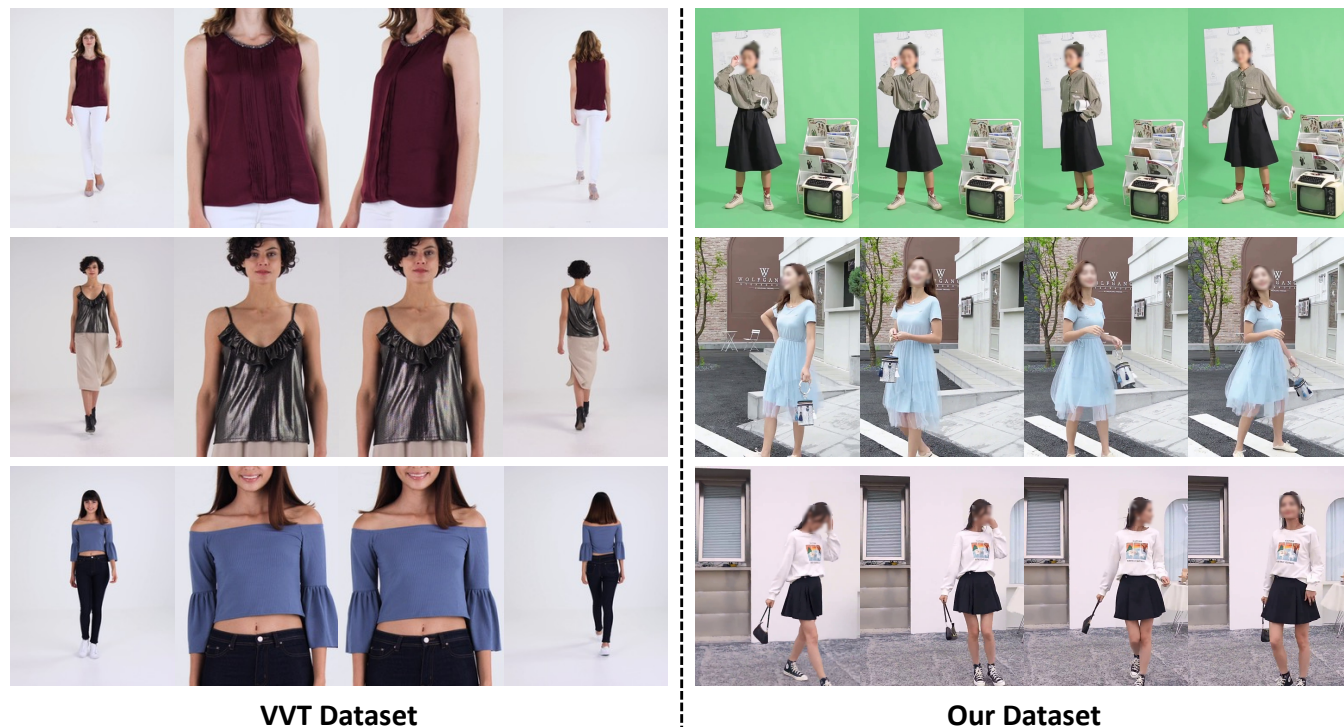


Figure 1: Visual comparison between VVT and our dataset. The faces are blurred due to privacy concerns.

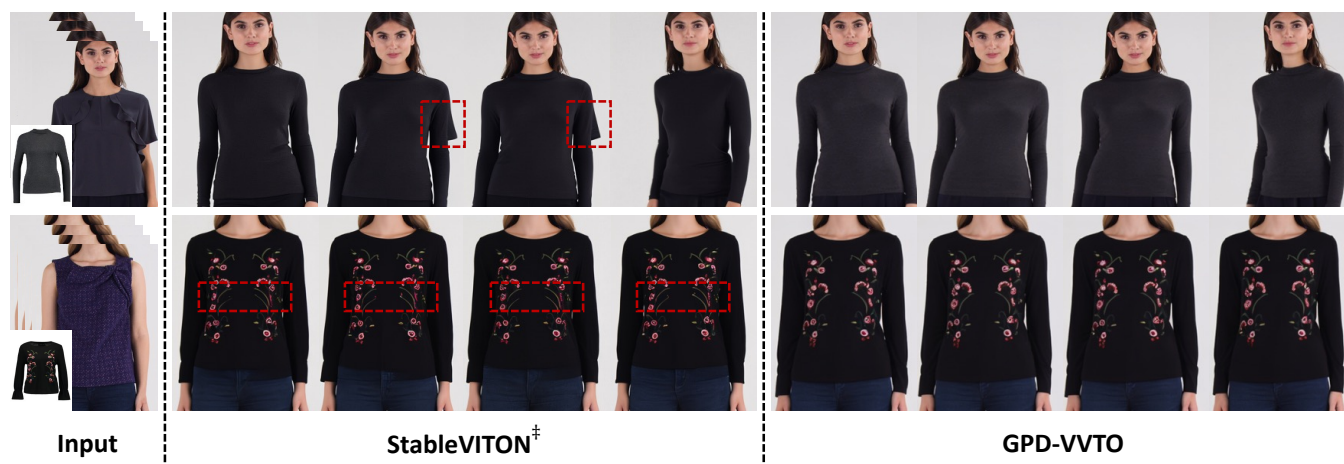


Figure 2: Qualitative comparison of video virtual try-on on VVT dataset. Shortcomings of previous methods are highlighted in red dashed boxes. Better viewed on a zoomed, color monitor.



Figure 3: Image-based virtual try-on results of multiple persons (rows) wearing multiple garments (columns) on VITON-HD dataset. Best viewed on a zoomed, color monitor.





Figure 4: Image-based virtual try-on results of multiple persons (rows) wearing multiple upper-body garments (columns) on Dress Code dataset. Best viewed on a zoomed, color monitor.



Figure 5: Image-based virtual try-on results of multiple persons (rows) wearing multiple lower-body garments (columns) on Dress Code dataset. Best viewed on a zoomed, color monitor.





Figure 6: Image-based virtual try-on results of multiple persons (rows) wearing multiple dresses (columns) on Dress Code dataset. Best viewed on a zoomed, color monitor.