Human Skeleton-aware Virtual Try-on via Fashion Landmarks and Garment Deformation

Honghong He[1,2], Yanghong Zhou[1], Zhengwentai Sun[1], Jin-tu Fan[1,2] and P.Y. Mok[1,2,*]

[1] School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong.
[2] Laboratory for Artificial Intelligence in Design, Hong Kong Special Administrative Region.
* Corresponding author: tracy.mok@polyu.edu.hk

KEYWORDS:

Virtual Try-on, Fashion Landmarks, Pose Estimation, Human Parsing, Skeleton-aware, TPS.

1. INTRODUCTION

Virtual Try-On technology (VTON) uses algorithms to argument digital fashion items onto users' images, allowing them to preview how the clothing would fit and look on their bodies. The usefulness of virtual try-on is not only to provide users to preview the high-quality wearing effects, but also to allow users with more informed purchasing decisions in terms of the fitting, shape, and size of the clothing, the suitability with other fashion items (Bartle et al., 2016; Lin & Wang, 2016; Shin & Baytar, 2014). In recent years, there has been a growing number of VTON methods developed. The pioneering model of virtual try-on, VITON (Han et al., 2018), and its variants CP-VTON (Wang et al., 2018), LA-VTON (Lee et al., 2019), VTNFP (Yu et al., 2019), SP-VTON (Song et al., 2020) all focus on how to improve the visual quality of try-on results. However, these models have yet to fully address the challenge of synthesizing try-on images that faithfully represent the relationship between wearable looseness, wearing position, varying body shapes, heights, and garments. They fall short in constructing a practical VTON method that replicates the authenticity of real try-on. Our research aims to address these limitations by a novel model called Human **S**keleton-aware VTON method via Fashion **L**andmarks and Garment **D**eformation (SLD-VTON). The model is designed to generate images of garments on different-sized persons, simulating the authentic appearance and fit of the clothing. Extensive experiments on deepfashion2 sub-dataset (Ge et al., 2019) and our newly collected garment dataset (SLD dataset) from e-commerce websites, such as Adidas, SSENSE have demonstrated that SLD-VTON is faithful and rational in try-on results with attention to wearable looseness and wearing position. Compared with state-of-the-art VTON methods, the proposed method is superior in preserving the details (colors, patterns, clothing wrinkles).

© 2023 The author(s). Published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

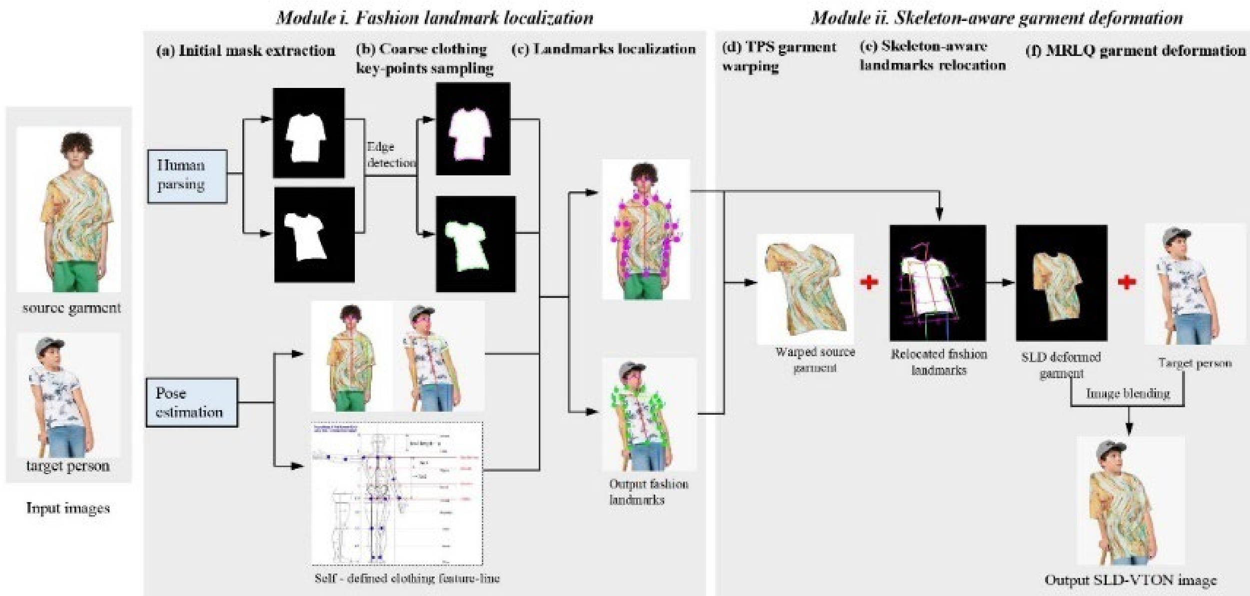

Fig. 1 The proposed SLD-VTON method framework.

2. FRAMEWORK

Fig. 1 shows the overall pipeline for the proposed SLD-VTON method, which involves (i) fashion landmarks localization module, and (ii) skeleton-aware garment deformation module. *Module (i)* determines the spatial position and correlation coefficients of clothing and human body between the input and target image. This process is achieved through a pose estimation network (Cao et al., 2017). Furthermore, it factors in proportions between the head and body, along with specific ratios such as Shoulder-to-Waist, Bust-to-Waist, and Waist-to-Hip. To obtain skeleton-aware transformed garment, *Module (ii)* incorporates prior information from fashion landmarks, human parsing, and pose estimation to guide the thin plate spline (TPS) warp (Belongie et al., 2002). Moreover, the moving regularized least squares (MRLQ) (Ma et al., 2013) is applied to refine the above wrapped results. At last, the deformed garment is synthesized with the target person to display the try-on results.

3. EXPERIMENTAL RESULTS

We compared our method with the state-of-the-art VTON methods (see Fig. 2) within the last two years: PFAFN (Ge et al., 2021), DiOr (Cui et al., 2021), and Flow-style VTON (He et al., 2022) corresponding to columns (c), (d) and (e) of Fig. 2, respectively. All methods were tested on the sub-deepfashion2 and SLD dataset to ensure fairness. We perform 1000 samples on each method and observe that these SOTA methods are not effective in preserving the texture of the customer's

© 2023 The author(s). Published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

non-garment areas (e.g., arms, face, and background, etc.). Particularly, when handling target images with complex poses and backgrounds, PFAFN and Flow-style VTON are struggling to synthesize realistic fitting garments for users. This is especially noticeable in areas such as the sleeves and the shoulder, where poor matching can occur. Although DiOr VTON is satisfactory in retaining the style and colors of the clothes, it fails to generate details of the source garment patterns (e.g., logo, stripes, see column (a)) and the background of the target image (column (b)) are barely satisfactory. In addition, artifacts are evident in areas such as the target's face and hair. Compared with these three methods, our proposed SLD-VTON is superior in preserving the texture details (see column (f) of Fig. 2).

Fig. 2 Qualitative comparison for SOAT VTON methods (along different columns) from results of (c) PFAFN 's, (d) DiOr 's, (e) Flow-style VTON's, and (f) the proposed SLD-VTON, respectively. Columns (a) and (b) are input source garment and target person.

© 2023 The author(s). Published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), *which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

4. CONCLUSION

This paper proposes a human skeleton-aware virtual try-on method via fashion landmarks garment deformation. Compared to the previous works, this paper mainly has four contributions. 1). The proposed SLD-VTON method can transfer the wearable looseness and wearing position between the source garment and the target person, benefiting from geometric characteristics of human and garment by fashion landmarks localization module and skeleton-aware garment deformation module. 2). The fashion landmarks localization module is a pose-aware segmentation-based localization method that sufficiently takes garment-body correspondence into account. 3). Skeleton-aware garment deformation module is designed to deform, scale the clothing adaptably to different target persons based human skeleton. 4). Extensive experiments demonstrate that SLD-VTON is not only faithful to users of different body shapes and heights but is also superior in terms of the textural details of the synthetic virtual try-on results.

ACKNOWLEDGEMENTS:

The work described in this paper was supported by the Laboratory for Artificial Intelligence in Design (Project Code: RP1-1) under the InnoHK Research Clusters, Hong Kong.

REFERENCES:

Bartle, A., Sheffer, A., Kim, V. G., Kaufman, D. M., Vining, N., & Berthouzoz, F. (2016). Physics-driven pattern adjustment for direct 3D garment editing. *ACM Trans. Graph.*, *35*(4), 50:51-50:11.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, *24*(4), 509-522.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. Proceedings of the IEEE conference on computer vision and pattern recognition, (pp. 7291-7299).

Cui, A., McKee, D., & Lazebnik, S. (2021). Dressing in order: Recurrent person image generation for pose transfer, virtual try-on and outfit editing. Proceedings of the IEEE/CVF International Conference on Computer Vision, (pp. 14638-14647).

Ge, Y., Song, Y., Zhang, R., Ge, C., Liu, W., & Luo, P. (2021). Parser-free virtual try-on via distilling appearance flows. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, (pp. 8485-8493).

Ge, Y., Zhang, R., Wang, X., Tang, X., & Luo, P. (2019). Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, (pp. 5337-5345).

© 2023 The author(s). Published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2018). Viton: An image-based virtual try-on network. Proceedings of the IEEE conference on computer vision and pattern recognition,

He, S., Song, Y.-Z., & Xiang, T. (2022). Style-based global appearance flow for virtual try-on. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (pp. 3470-3479).

Lee, Rokkyu, Kang, M., Cho, M., & Park, G. (2019). LA-VITON: A network for looking-attractive virtual try-on. Proceedings of the IEEE/CVF international conference on computer vision workshops, (pp. 0-0).

Lin, Y.-L., & Wang, M.-J. J. (2016). The development of a clothing fit evaluation system under virtual environment. *Multimedia Tools and Applications*, *75*, 7575-7587.

Ma, J., Zhao, J., & Tian, J. (2013). Nonrigid image deformation using moving regularized least squares. *IEEE Signal Processing Letters*, *20*(10), 988-991.

Shin, E., & Baytar, F. (2014). Apparel fit and size concerns and intentions to use virtual try-on: Impacts of body satisfaction and images of models' bodies. *Clothing and Textiles Research Journal*, *32*(1), 20-33.

Song, D., Li, T., Mao, Z., & Liu, A.-A. (2020). SP-VITON: shape-preserving image-based virtual try-on network. *Multimedia Tools and Applications*, *79*, 33757-33769.

Wang, B., Zheng, H., Liang, X., Chen, Y., Lin, L., & Yang, M. (2018). Toward characteristic-preserving image-based virtual try-on network. Proceedings of the European conference on computer vision (ECCV), (pp. 589-604).

Yu, R., Wang, X., & Xie, X. (2019). Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. Proceedings of the IEEE/CVF international conference on computer vision, (pp. 10511-10520).

© 2023 The author(s). Published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), *which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*