Virtual Try-On: Real-Time Interactive Hybrid Network with High-Fidelity

Umer Waqas, Yunwan Jeon, Donghun Lee
AItheNutrigene
umer@aithenutrigene.com, jyw@aithenutrigene.com, donghun.lee@aithenutrigene.com

Abstract

A significant upsurge in the fashion e-commerce industry in recent years has brought considerable attention to image-based virtual fitting. This image-based technology allows users to try on clothes virtually without physically touching them. However, the current techniques have notable limitations in terms of real-world scenarios, noisy results, partial clothing categories and computational cost, thus limiting the real-world applications. To address these critical limitations, we propose a hybrid interactive network that allows actual users to interact with the system to try on clothes virtually. The network is composed of state of art keypoint extraction, appearance flow alteration and wrapping modules. The proposed network facilitates real-time application with high-quality noise-free results, a variety of clothing categories and efficient computational cost.

Introduction

In recent years, image-based virtual try-on which aims to fit clothes virtually on the human body have received significant attention due to the rise of the fashion e-commerce industry. Various image-based synthesis methods have been introduced on several widely used benchmarks. However critical deficiencies prevent them from real-world applications. The critical challenges include artifacts, partial clothing categories and computational cost.

In the past, (Han et al. 2018) proposed VITON which employs GAN based approach to virtually fit clothes but lacked high-fidelity fitting results, contained artifacts in the final generated output and used upper body clothes only such as shirts, etc. To address the artifact issues, (Yang et al. 2020) introduced ACGPN to generate photo-realistic tryon images to improve fitting results, but noticeable artifacts remained visible. Similarly (Ge et al. 2021), (Choi et al. 2021), (He et al. 2022) and (Fele et al. 2022) proposed different methods that improved the results over time, but critical challenges persisted. This is because of the fact that most of the previous work was limited to data benchmarks that considered upper-body clothes thus restricting them to a partial clothing category, while in real-world scenarios there are many clothing categories to choose from. Additionally, the artifacts issue continued due to only relying on generative models, therefore strictly limiting their use for practical real-world applications.

To overcome the critical challenges and shortcomings of previous work, we propose a hybrid network that addresses three major issues. 1) The artifact reduction issue is addressed by constraining the use of a sole generative model and adding a new hybrid module that generates noise-free and high-quality results. 2) To solve the partial clothing categories issue, the network is developed in a way that works on several new clothing categories which include the upper body (e.g., half-sleeve shirts, full-sleeve shirts, diverse number of neck style options, etc.), lower body clothing (e.g., cargo pants, leggings, jeans, mini skirt, long skirt, etc.). The clothing categories are not only limited to general categories, but the proposed model also generates high-fidelity results on subcategories and complex styles e.g., flared skirts, asymmetrical skirts, etc. 3) The computational cost is significantly reduced, enabling real-time service to users.

System Overview

Figure 1 illustrates the overview of our proposed system. The uploaded user image is processed by a validation check, segmentation component and then passed to three major modules of the network that perform operations on the input image as follows:

Posture Detection and Keypoint Extraction: This module consists of three steps. In the first step, OpenPose is applied to get the basic body keypoints. In the second step, to detect the full body in the image and its posture, a new function θ is defined to get the angles of different body points $P_1 \; to \; P_n$, where P contains the x and y coordinates. E.g., verifying if the body has an A-style posture is defined as follows:

$$\theta = \arctan\left(\frac{P_f y - P_m y}{P_f x - P_m x}\right) - \arctan\left(\frac{P_l y - P_m y}{P_l x - P_m x}\right) \qquad (1)$$

where P_m stands for middle point, P_f and P_l extends in the opposite direction from P_m. θ is calculated by using each set

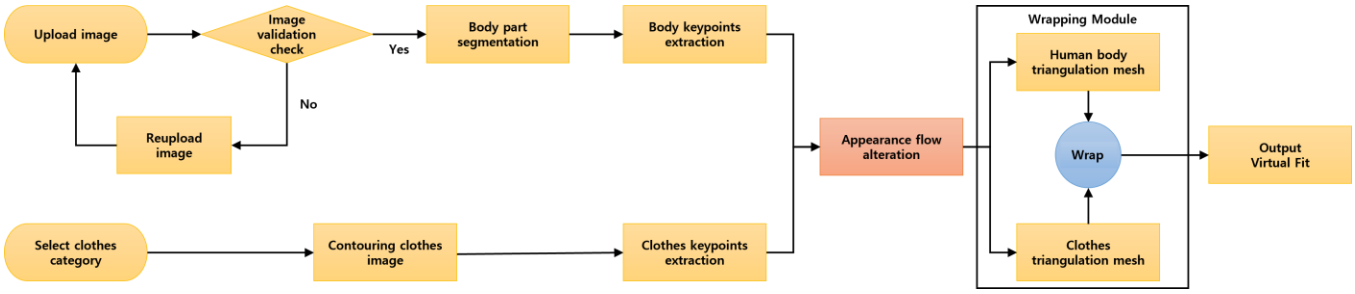

Figure 1: The system overview of interactive hybrid network

of three points and through a given set of criteria for different body keypoints for posture.

In the third step, we extract the contours of the body, this is because Openpose lacks information of exact x and y coordinates on the body edges which is crucial to contain wrapping clothes in the next modules. Therefore, another mathematical method is built on top of the OpenPose output layer to measure and associate each contour with a different body part. It involves measuring unique and fixed x and y coordinates by using the distance function associated with key body points on edges.

Appearance Flow Alteration: This module performs a unique operation as compared to previous works where mostly clothes are deformed and fit on the body. This is a critical problem because sticking clothes on the body does not truly represent the fitting and clothes lose their original shape or features. E.g., a flare skirt should fit on the waist but needs to be a bit loose on the lower bottom sides, which represents one of the real-world scenarios when a person tries a flare skirt in reality. Keeping the skirt's original shape or design requires extra space and boundary that falls out of the body contours. To address this critical challenge and a variety of clothing categories a new mathematical approach is proposed which includes keypoint movement and aligning function that considers clothing keypoints and then the same shape and size of the area is created on the body including space if required outside of the body contours. E.g., generating a new area for a flared skirt that requires extra space beyond the body contours is defined as follows:

$$\theta = arctan\left(\frac{y_{P_s} - y_{P_e}}{x_{P_s} - x_{P_e}}\right) \tag{2}$$

$$\theta_i = arctan\left(\frac{y_{PC_s} - y_{PC_i}}{x_{PC_s} - x_{PC_i}}\right) - \theta_c \tag{3}$$

$$d_i = \sqrt{((x_{PC_s} - x_{PC_i})^2 + (y_{PC_s} - y_{PC_i})^2)} * r \tag{4}$$

$$\begin{aligned} \acute{x}_i &= x_{PB_s} + d_i * \cos(\theta_i + \theta_b), \\ \acute{y}_i &= y_{PB_s} + d_i * \sin(\theta_i + \theta_b) \end{aligned} \tag{5}$$

$$\acute{P}_b = (\acute{x}, \acute{y}) \tag{6}$$

where Eq. 2 initially takes a set of two keypoints that connect the start point P_s and end point P_e to calculate section angles of clothe θ_c and body θ_b. Eq. 3 and Eq. 4 perform further detailed angles estimation, slope operation and distance calculation on clothe keypoints whereas Eq. 5 and Eq. 6 compute alteration of keypoints on body in relation to the clothe shape.

Wrapping Module: To wrap clothe images on the area generated by the appearance flow alteration module, a structured triangulation mesh method is proposed. In d dimensional space, a set of a fixed number of points P extracted through the previous module is used to form a simplex given that no point P is inside the circum-hypersphere which forms a unique convex hull. This provides a raw mesh but lacks an accurate representation of the spatial relationships between points. At this point, we introduce another function δ which allows us to insert a collection of P_e points that represent the structured edges from the output of the appearance flow alteration module, hence the final output supersedes the raw mesh simplices criterion and provides a structural triangulation mesh to wrap clothe segments on the body.

Demonstration

The demonstration shows three sections. The first section shows a variety of clothes categories. The second section takes the actual user input image where the user uploads his/her image. The third section shows the virtual try-on results. The overall demonstration shows a complete operation of uploading different user images, choosing different clothes from diverse categories and displaying virtual try-on results.

Conclusion

In this work, we present a real-time application of virtual try-on by leveraging a hybrid approach that shows a high-fidelity fitting result and is practically useable in real-world scenarios. The proposed method addresses artifact issues and limitations of clothing categories while maintaining computational efficiency to generate noise-free and high-quality results. In future work, a diverse number of complex human poses will be addressed with easy integration available for current fashion e-commerce systems.

References

Choi, S., Park, S., Lee, M., & Choo, J. 2021. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14131-14140.

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

Fele, B.; Lampe, A.; Peer, P.; and Struc, V. 2022. C-VTON: Context-driven image-based virtual try-on network. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 3144-3153.

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

Han, X.; Wu, Z.; Wu, Z.; Yu, R.; and Davis, L. S. 2018. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7543-7552.

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

Yang, H.; Zhang, R.; Guo, X.; Liu, W.; Zuo, W.; and Luo, P. 2020. Towards photo-realistic virtual try-on by adaptively generating-preserving image content. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7850-7859.