Virtual Try-On Methods: A Comprehensive Research and Analysis

Haoxuan Sun

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
guwangtu@sjtu.edu.cn

Abstract. Image-based virtual try-on, as a challenging and practical real-world task, is one of the interesting research topics in recent years. Virtual try-on will become a common way to buy fashion products in the future, however, with the development in recent years, relevant review articles are not sufficient. Therefore, this paper aims to systematically introduce the development trend and current advanced algorithms in the field of virtual try-on. This paper briefly introduces various research directions in this field at present, and gives a certain degree of analysis, including 3D virtual try-on and 2D virtual try-on. Among them, this paper uses more space to describe 2D virtual try-on as it's the mainstream direction. In conclusion, this article culminates with a comprehensive summary and projection of the complete text. The intent is to elucidate the contemporary advancements in this domain and provide a more comprehensive depiction of the potential future trajectory of development.

Keywords: Virtual try-on, deep learning, 3D,2D, diffusion model.

1 Introduction

In recent years, online shopping has become more and more developed, and the Internet and e-commerce have developed rapidly. At the same time, people's demand for fashion products is increasing, and more and more consumers choose online shopping to obtain goods. Although online fashion shopping provides the convenience of shopping and selection, consumers worry that when they buy clothes online, they will not match them when they wear them because they cannot see the effect of their own clothes. Returning or leaving clothes that you are not satisfied with will bring a bad shopping experience, and this inconvenience also reduces consumers' willingness to consume to a certain extent.

Therefore, virtual try-on technology emerged at the historic moment, aiming to provide a method of simulating the trying-on process to help users better understand the appearance and fit of clothing before purchasing. Allowing customers to digitally try on clothing can transform how people shop for clothing and improve the online buying experience. It can also be used to help shoppers chose clothing at the mall before they decide to try it on. To sum up, virtual try-on has broad application

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_35

prospects and commercial potential, and is a very practical technology that can change the consumption process and make it more convenient. The primary task of virtual try-on is to align the deformation of the provided clothes with the body posture of the tryer. According to the shape, wrinkles and occlusion may occur. In addition, the change of the skin area covered by the try-on clothes may reveal that the tryer is originally occluded body skin, so skin around the arms and neckline, for example, also needs to be generated. Finally, the changed part is fused with the part that will not change, such as the face and the non-try-on clothing area, to obtain the desired virtual try-on result.

Although virtual try-on technology has great value in practical applications, its development still faces many challenges. First of all, the complexity of clothing and human body shape, texture, lighting, etc. makes the task of virtual try-on more difficult. Secondly, real-time and interactivity put forward higher requirements on the performance of virtual try-on technology. In addition, the existing virtual try-on technology is not ideal in terms of accuracy and visual effects, which makes users have reservations about virtual try-on experience and trust.

To sum up, there are still many problems in the current virtual try-on algorithms, and the difficulties in various aspects that are not conducive to the application of technology in real life are worthy of further development. In order to provide more detailed and comprehensive systematic knowledge combing to future researchers, this paper will investigate some previous classical algorithms and latest ones, and analyze its improvement process and current shortcomings, to point out some of the direction for future research.

2 Main Body

In general, virtual try-on is divided into:

1. Use measurement data to reconstruct a 3D human body, and then perform simulation based on clothing modeling.
2. Use deep learning related technologies to achieve end-to-end try-on results generation, which is further divided into 3D The reconstruction of human body try-on results and the generation of 2D human body try-on results.

2.1 3D Virtual Try-on

Since the end of the last century, many systems for measuring the human body have been developed. Using these systems to obtain 3D anthropometric results and then reconstruct the human body model, the clothing provided is modeled in batches according to the 3D data of the clothing, and then adjusted and optimized to make the clothing model and the blending between the mannequins is more harmonious, allowing the tryer to try on multiple different garments. These are some early methods, and there are many reasons such as complex measurement operation, low measurement accuracy, and low system stability. With the development of computer graphics, people can carry out three-dimensional reconstruction through the mapping

relationship between human body images and human body models, and simulate and model clothes, simulate and render the models to achieve more realistic try-on effects [1]. However, the construction and processing of 3D models basically need to model new users and new clothing, which requires high computing resources, and the interactivity and real-time performance need to be improved. Therefore, the 3D Virtual Try-on mentioned in this paper refers to the virtual try-on to obtain 3D results.

In 2018, Alldieck proposed the first method for creating a unique 3D human body model from a single video of a moving individual [2]. Their method demonstrated resistance to noisy 3D posture estimates and can recreate the human body shape with a high degree of accuracy. But there are also several limits, such as: long hair or skirts cannot be modeled harmoniously. Sometimes improper operation is given to concave areas like the inner thighs or armpits. Fast skeletal movements will also create strong fabric movement, which will further reduce the level of detail. Xu presented an approach that automatically constructs 3D models for garments using RGB images include one front view and one rear view in 2019 [3]. They proposed a multi-task learning network, can jointly identify the landmarks of the garment and parses the garment into semantic part segments as well. The limitations of this approach are as follows. First, the representation power of the templates limits the range of cloth. Second, there is excessive local deformation caused by huge differences between source and target contours.

In 2021, Santesteban introduced a novel approach for addressing garment-body collisions [4]. They propose the use of a new canonical space for garments that extrapolates body surface properties, including skinning weights and blend shapes, to 3D point representations. Unlike physical modeling methods, data-driven approaches based on deep learning often encounter a common challenge wherein the mutual penetration of human body and clothing cannot be entirely avoided, even with the inclusion of penalty terms in the loss function during training. Consequently, additional post-processing steps are typically required to rectify such flaws during the inference stage. However, the method put forward by Santesteban aims to eliminate these imperfections without the need for additional post-processing steps, thereby greatly enhancing the overall effectiveness of the approach. Zhao proposed M3D-VTON (Monocular-to-3D Virtual Try-On Network) to combining 2D data with learning a mapping which has the ability of transforming the 2D representation into 3D [5]. This algorithm performs 2D try-on and body depth estimation respectively to achieve the purpose of the 3D try-on, such that A offers the work of monocular-to-3D virtual try-on faster and more affordably.

2.2 Image-based (2D) Virtual Try-on

The goal of virtual try-on based on 3D reconstruction is to display all aspects of dynamic try-on effects. However, due to various constraints, such as high computing consumption and high equipment costs, there are great shortcomings in terms of fidelity and practicality. But conversely, in fact, if the tryer can get some two-dimensional try-on photos, it is enough to judge whether he is suitable for the dress, so many people shift their perspective to the two-dimensional virtual try-on task.

Han has conducted groundbreaking research in the field of utilizing the VITON (Virtual Try-On Network) within an encoder-decoder architecture [6]. In the study, they input preprocessed images into the encoder-decoder generator to generate distorted clothing and produce an initial composite image. The target clothing is then superimposed onto the same pose of the individual using a refinement network in conjunction with a TPS-transformed clothing image, ultimately yielding the final try-on result. This approach provides a benchmark process for image-based virtual try-on, but if there are very rare poses or the shape of the clothes is too different from the original clothes, the results of the image will be poor.

Since then, many people have made many improvements on this [7]. For example, Wang proposed CP-VTON (Characteristic-Preserving Virtual Try-On Network), directly improved on the two main problems of VITON: due to the poor handling of deformation when the clothes and body shape are aligned, the clothes and body The degree of matching is insufficient, the appearance merging strategy of clothing images and synthetic images is not perfect. This work proposes a thin-plate spline transformation that can be learned. Compared with the previous TPS transformation, this method does not require explicit interest point correspondence. Additionally, the newly proposed try-on module has the ability to dynamically combine the outcomes of the preliminary synthesis and twisted clothing.

Minaret al proposed that CP-VTON+ has made further improvements to the warping part of CP-VTON [8]. Minar revealed the challenges and root causes of the image-based virtual try-on method and redesigned the pipeline by improving the input representation and improving the training cost function (Fig.1). The incorrect labeling of the chest area in the body parsing map and the absence of garments in the reserved area are the first errors that CP-VTON+ corrects. Second, the authors observed problems in the clothing deformation network: unbalanced geometric matching input and training loss function. Finally, it uses the input clothing mask and a specific loss function to improve the synthetic mask.

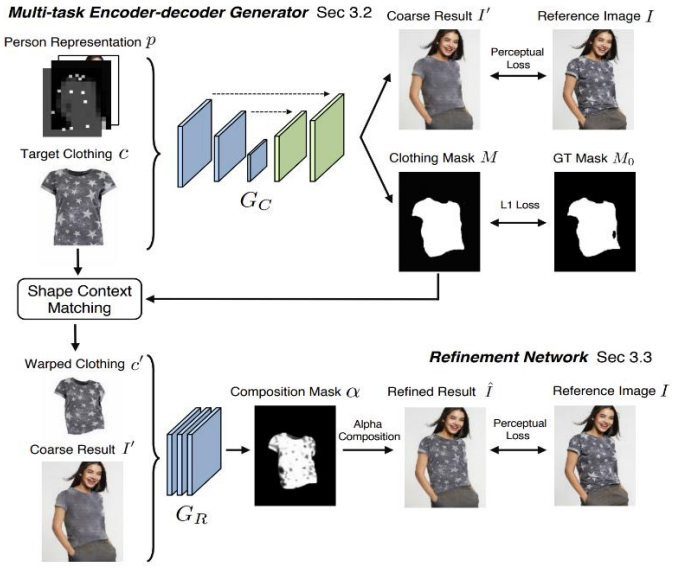

Fig. 1. Overall of VITON [6]

In order to transfer specific garment regions to the target even in challenging pose configurations and with self-occlusions, Fele introduced the C-VTON (Context-driven Virtual Try-on Network) [9]. C-VTON includes two key steps, one is geometric matching part, BPGM (Body-Part Geometric Matcher) can align the target clothing with the pose of the input person, and the other is the virtual fitting synthesis part, CAG (Context-Aware Generator), which leverages contextual information when synthesizing results. The limitations of C-VTON includes some unconvincing results when trying on loose clothes because the model cannot distinguish the front and back of the clothes, which will lead to unrealistic and soft edges of cloth and improperly rendered neck areas.

Xie proposed GP-VTON (General Purpose Virtual Try-On) in order to solve two problems [10]. First, it is impossible to maintain the semantic information of various elements when faced with difficult inputs (such as complex human poses or ornate clothing). Second, directly distorting the input clothing to match the preserved area is difficult. The process of boundary alignment uses texture extrusion to satisfy the boundary shape constraints which leads to texture distortion problems. The core method is warping module called LFGP (Local-Flow Global-Parsing) and training strategy called DGT (Dynamic Gradient Truncation). Since GP-VTON conducts local warping for different garment parts individually, it would fail to obtain accurate warped result when the input in-shop garment is incomplete. Besides, GP-VTON is unable to address the parsing error. To alleviate the influence of the parsing error, the author mentioned to the knowledge distillation mechanism, which is commonly used to obtain a parsing-free model.

In addition to the VITON branch, there is also work on the use of GAN's generation confrontation network architecture. Jetchev proposed a conditional

analogy confrontation generation network called CAGAN (The Conditional Analogy GAN), which first made the chore of exchanging clothing items on pictures of people [11]. This U-Net-based GAN method, but because this network cannot handle large spatial deformations, the results produced are not ideal. Ge presented DCTON (Disentangled Cycle-consistency Try-On Network), which introduces an approach to virtual try-on by breaking it down into multiple distinct steps [12]. Specifically, these steps involve clothing deformation, skin synthesis, and image synthesis in order to produce realistic try-on images. Furthermore, the network can be trained using self-supervised methods in addition to cycle-consistency learning, further enhancing its versatility and effectiveness.

There is another thing worth noting that the generation method is the diffusion model, which has received widespread attention because of its excellent performance. Inspired by it, Zhu proposed TryOnDiffusion to preserve clothing details and warp clothing for dramatic pose and body changes in a single network [13]. The main contribution of this method is wo solve the two tasks of clothing deformation and character blending using the cross-attention mechanism in a unified process. This method achieves an excellent performance both qualitatively and quantitatively. Besides it also embodies the potential of diffusion models. However, this article still has many deficiencies. The impact of restricted preprocessing errors is relatively heavy. This method uses an image independent RGB map to represent identity, which has room for improvement. What's more, TryOnDiffusion has not been tested on more complex datasets, such as full-body try-on datasets and real-world images.

3 Conclusion

This paper first introduces the importance of virtual try-on technology research and its wide application prospects and points out the purpose of writing this paper. In the main part, this paper analyzes many methods of virtual try-on at the present stage, including 3D method, 2D method and virtual try-on based on other conditions. Among them, this paper focuses on the analysis and comparison of 2D virtual try-on algorithms and discusses the pros and cons of different methods and possible improvement directions.

Although with the gradual development and maturity of virtual try-on technology based on deep learning, people can achieve better results using existing virtual try-on algorithms. However, in the face of more complex scenes and unpredictable body posture and with a wide variety of clothing styles, there are still many problems in the current virtual try-on task. This article aims to focus on several branches of the mainstream virtual try-on field in the artificial intelligence field and analyze the improvement process and shortcomings of existing algorithms in order to predict the direction of future development and improvement. In future research, the focus will be on the following aspects:

1. For generating deformations in clothing regions, the future direction is still the robustness of generation in occluded regions.

2. For the development of applicability, we still need to work on the simplicity of the model and the good results of the lower dataset.

It is believed that with the development and innovation of technology, virtual try-on will become an indispensable part of people's life and bring revolutionary changes to the fashion industry.

Reference

1. Pons-Moll G, Pujades S, Hu S, ClothCap: Seamless 4D clothing capture and retargeting [J]. ACM Transactions on Graphics (ToG), 2017, 36(4): 1-15.
2. Alldieck T, Magnor M, Xu W, Video based reconstruction of 3d people models[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8387-8397.
3. Xu Y, Yang S, Sun W, et al. 3d virtual garment modeling from rgb images[C] 2019 IEEE international symposium on mixed and augmented reality, 2019: 37-45.
4. Santesteban I, Thuerey N, Otaduy M A, et al. Self-supervised collision handling via generative 3d garment models for virtual try-on[C] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11763-11773.
5. Zhao F, Xie Z, Kampffmeyer M, et al. M3d-vton: A monocular-to-3d virtual try-on network[C] Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 13239-13249.
6. Han X, Wu Z, Wu Z, Viton: An image-based virtual try-on network[C] Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7543-7552.
7. Wang B, Zheng H, Liang X, Toward characteristic-preserving image-based virtual try-on network[C] Proceedings of the European conference on computer vision. 2018: 589-604.
8. Minar M R, Tuan T T, Ahn H, Cp-vton+: Clothing shape and texture preserving image-based virtual try-on [C] Conference on Computer Vision and Pattern Recognition Workshops. 2020, 3: 10-14.
9. Fele B, Lampe A, Peer P, et al. C-VTON: Context-driven image-based virtual try-on network[C] Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 3144-3153.
10. Xie Z, Huang Z, Dong X, et al. GP-VTON: Towards General Purpose Virtual Try-on via Collaborative Local-Flow Global-Parsing Learning[C] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 23550-23559.
11. Jetchev N, Bergmann U. The conditional analogy gan: Swapping fashion articles on people images[C] Proceedings of the IEEE international conference on computer vision workshops. 2017: 2287-2292.
12. Ge C, Song Y, Ge Y, Disentangled cycle consistency for highly-realistic virtual try-on[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 16928-16937.
13. Zhu L, Yang D, Zhu T, et al. TryOnDiffusion: A Tale of Two UNets[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 4606-4615.