Multimedia Tools and Applications (2023) 82:23151–23178
https://doi.org/10.1007/s11042-022-14127-w

Dress-up: deep neural framework for image-based human appearance transfer

Hajer Ghodhbani[1] · Mohamed Neji[1,2] · Abdulrahman M. Qahtani[3] · Omar Almutiry[4] · Habib Dhahri[4] · Adel M. Alimi[1,5]

Received: 30 April 2022 / Revised: 3 August 2022 / Accepted: 25 October 2022 /
Published online: 12 November 2022

Abstract

The fashion industry is at the brink of radical transformation. The emergence of Artificial Intelligence (AI) in fashion applications creates many opportunities for this industry and make fashion a better space for everyone. Interesting to this matter, we proposed a virtual try-on interface to stimulate consumers purchase intentions and facilitate their online buying decision process. Thus, we present, in this paper, our flexible person generation system for virtual try-on that aiming to treat the task of human appearance transfer across images while preserving texture details and structural coherence of the generated outfit. This challenging task has drawn increasing attention and made huge development of intelligent fashion applications. However, it requires different challenges, especially in the case of a wide divergences between the source and target images. To solve this problem, we proposed a flexible person generation framework called *Dress-up* to treat the 2D virtual try-on task. *Dress-up* is an end-to-end generation pipeline with three modules based on the task of image-to-image translation aiming to sequentially interchange garments between images, and produce dressing effects not achievable by existing works. The core idea of our solution is to explicitly encode the body pose and the target clothes by a pre-processing module based on the semantic segmentation process. Then, a conditional adversarial network is implemented to generate target segmentation feeding respectively, to the alignment and translation networks to generate the final output results. The novelty of this work lies in realizing the appearance transfer across images with high quality by reconstructing garments on a person in different orders and looks from simlpy semantic maps and 2D images without using 3D modeling. Our system can produce dressing effects and provide significant results over the state-of-the-art methods on the widely used DeepFashion dataset. Extensive evaluations show that *Dress-up* outperforms other recent methods in terms of output quality, and handles a wide range of editing functions for which there is no direct supervision. Different types of results were computed to verify the performance of our proposed framework and show that the robustness and effectiveness are high by utilizing our method.

Keywords Artificial intelligence · Outfit generation · Garment interchange · Virtual try-on · Semantic segmentation

✉ Hajer Ghodhbani
hajer.ghodhbani@regim.usf.tn

Extended author information available on the last page of the article.

1 Introduction

1.1 Background

Nowadays, technology is mostly responsible for how everything works. With the continuous development of computer technology, the wave of image processing has become popular in almost vision tasks. In this regard, our modern technology heavily relies on image, thus, image processing field is growing at an exponential rate and has a broad range of applications in various areas such as medical imaging [2, 3, 6], image security and encryption [4, 5, 8, 12, 24, 52], and industries [1, 46, 59]. These development has increased especially since 2020, the year when everything was changed around the world, after the appearance of coronavirus pandemic that sent its shockwaves in each country. These last years will go down in history as one of the most challenging and worst period on record for various sectors. However, The COVID-19 outbreak has posed an unprecedented challenge to humanity and science. Therefore, various incentives have been put in place to provide resources toward research areas strictly related to the COVID-19 emergency.

The overall picture that emerges from this situation is that there has been a profound realignment of priorities and research efforts in various fields affected by this pandemic such as the fashion industry that finding itself in mid of unprecedented adversity which marked a drop in sales and a change in customer behavior. Last year and according to McKinsey Global Fashion Index analysis, fashion companies post approximately a 90% decline of their economic profit, after an increase of 4% in 2019 [10]. In the coming years and due to the doubt of the epidemiological situation around the world, the predictions for fashion apparel performance are related to different scenarios.

Despite the expectations that the coming period will be critical for fashion apparel, it will also be a period of investment to make changes in this area. The future opportunities will be created for companies that are able to adopt new solutions for digital shopping which is the main driver of fashion industry development in next years. Thus, yet the several impacts of this pandemic, fashion companies must innovate new ways and strategies to compete.

1.2 Digital transformation of fashion industry

Prior to the pandemic, digitization and fashion were not strangers to one another. Digital transformation can be as simple as launching an e-commerce site. Covid-19 accelerated this change in the fashion industry and made a large digital transformation in this sector to create innovative solutions for new challenges. In addition, Changes in consumer demands are the reason to adapt more digital innovations. Fashion firms must pay attention to their customers' needs and respond with digital solutions. The digital opportunity in e-commerce is the most cited one by far, even in case of eradicating this pandemic. Last year, the online fashion industry growth marked higher anticipations compared to 2019 [10]. Thus, current years suggest the beginning of novel chapter for the global fashion apparel based on lessons learned from last period which have been the key of huge change. The most important idea taken from all these events is that, similar to various industries, the fashion apparel, will leave this crisis in a distinct form than that in which it entered.

Over the past year, brands have addressed digital transformation in many ways, each in their own manner. According to McKinsey's State of Fashion 2021, 45% of fashion executives identify Covid-19 as the biggest challenge these years, and 30% see going digital as the biggest opportunity. Regardless, there is one thing that can be said with certainty: people who are more tech-savvy have a clear advantage over those who are less so. For many firms,

this means that their e-commerce strategy must be adjusted to the growing digitalization and personalisation of consumers' shopping behaviors and desires. Thus, they need to find new strategies to motivate customers to purchase online. They should adapt to the faster change by adopting new working practices that have emerged from the crisis and trying to elevate the online customer experience with advanced methods. In coming years and according to fashion leaders, it will be a greatest growth in e-commerce due to the improvement of digital customer experience through the use of Artificial Intelligence (AI).

1.3 Artificial intelligence for fashion

AI is being utilized to operate businesses and boost productivity in a variety of industries, including manufacturing. Enhancing sustainability and creating a personalized consumer experience are both possible with new and developing technology. The fashion industry is utilizing AI in a variety of ways, including the usage of online fashion assistants to enhance customer experience and creating more personalized environment to assist clients in selecting clothing based on their preferences. By providing customized products, they will only select what best suits the consumer and assist them in creating exactly what they picture in their minds. As a result, fashion firms all over the world are incorporating AI into their design processes to increase customer satisfaction and decrease the amount of returns.

Thus, Making AI a part of shopping experience helps the companies bring the customer exactly what they want , and make sure them gets exactly what they are looking for. It is estimated that artificial intelligence will bring the fashion industry to 7.3 billion by 2022 and will only continue to grow [10]. A new partnership between fashion and artificial intelligence will certainly prosper for years to come by achieving numerous improvements. One of these improvements is to allow consumers purchasing fashion items after fitting them online like in real stores.

1.4 Motivation

AI has improved both in-person and online clothing shopping. One of the biggest invention made in this field is the smart mirrors used by Fashion Brands to make customers have a top-notch experience. Customers can try-on various outfits without using dressing rooms. This technology is offered for in-store purchases, but what about the online side? Online apparel shopping presents some product-related risks, especially when the consumers are notable to see and touch the products or try them on. Here, brands need to create the same experience for the e-shoppers. Therefore, fashion companies have to satisfy their preferences and engage them with customized purchasing experience to build confident purchase decisions. Following this objective, our work focuses on using AI technology to develop an image-based virtual fitting system.

Image-based system is the most aged one among above systems. In such system, users can warp a clothes' image onto a user's photo to make it look like the user is wearing the clothes. An image warping technique is applied to map the apparel image onto the individualized body. While its result is unrealistic and such techniques do not allow to see the textured images from arbitrary viewing angles. These solutions can really turn your business to a whole new level. When thinking for virtual fitting room apps, it is required to search for simpler alternatives to virtual clothing try-on techniques. A popular option here is, instead of going for fitting 3D clothing items, working with 2D clothing items and 2D person images. It is exactly what our solution does, giving users a possibility to apply several clothing types to their photo in easy way.

However, using these solutions would be a way to see the virtual effects, but they are still far from solved due to the challenge to virtually change the texture and pattern of clothes deformation, especially, when we use an image-based approach to transfer clothes. This point is raised in our survey [15] that focuses on the virtual outfits systems and it was the start to think for a solution to overcome this problem. Thus, we have proposed a flexible 2D person generation pipeline for virtual try-on based on Image-to-Image translation task (I2I). The main idea of our work is to solve the virtual fitting problem by changing outfits across images. Quality question is essential here because the outfits should be visually realistic without harmful effects of image variations. Recently, research in this task has been growing faster than ever before due to the appearance of deep generative models and their capacity to generate photorealistic images which is the case of our solution that will be detailed in coming sections.

1.5 Objectives and contributions

The aim of our research work is to create an Image-based virtual fitting system that realistically reflects the appearance and the behavior of garment. It should further adapt to specific bodies of different persons depending on their body measurements. This will be one of the main challenges since the pieces of cloth should correctly fit to the person. Our goal is to make the customer experience into a reality from their own home without the need to go out into the shops to try-on clothes items, thus, making a true virtual online shopping experience. Our objective consisting in designing a simplified framework to implement virtual try-on system to provide an advanced virtual try-on experience. The contribution of our work mainly reflects in these aspects:

- Interchanging garment appearances between pictures of persons with arbitrary pose.
- Transferring garment across images and maintaining the identity of picture while transferring clothes details to target view.
- Synthesizing an image of a person with cohesive outfit from two different view images.
- Encountering problems like degradation of details, occlusions, and physical dislocation and missing.

After an overview of related work in Section 2, we present, in the Section 3, the task of image based transfer and expose our problematic and the related challenges. Then, in Section 4, we will describe our proposed system composed of three stages which will be edited separately throughout this section. After that, we pass to the experimental part in Section 5 to show the experiments process with illustration of the implementation details. The obtained results and its analysis will be presented, in this same section, with different comparisons to show the effectivness and limitations of our framework. Then, some Future work will be propose to improve our solution. Finally, the paper will conclude with a brief summary in the last section.

2 Related work

2.1 Virtual try-on

Image based virtual try-on task focuses on interchanging desired garment to a given human body, it is considered as a challenging task since it requires to preserve details and identity

while transferring clothes from the image of such person to another in different poses. The simplest way to deal with try-on task is to replace a specific item of clothes with a target one [19, 54, 55, 64] or change the clothes of such person on different poses [11, 14, 26, 32]. These methods have emerged as an alternative for traditional fitting applications that depend on 3D reconstruction and computer graphics tools [18, 44, 45, 49, 56–58]. Such prior works [27, 34, 47], treated successfully the virtual try-on task as a visual analogy problem by generating an image across the transformation between a pair of another images which is the case of our problem.

Recently, other solutions [11, 33] have also been proposed to improve the results of image synthesis task by using secondary neural networks that generate clothing segmentations matching the target garment and utilizing these as additional sources of information for the generator. Further works [7, 13] followed recent developments in image synthesis [43] and introduced generators with conditional normalization layers to help with the quality and realism of the synthesized try-on results. Similarly to these techniques, our work also uses an advanced image generator to realize conditional synthesis based on certain input data, and specifically, we intersted in converting a semantic segmentation mask to a photo-realistic image and improving the realism of the generated results.

With further works, the challenge gets more difficult, and the suggested virtual fitting system should be more closely by generating images with all the target garments [9, 38, 41, 47, 50, 51]. Men et al. [38] implemented generative model called Attribute-decomposed GAN (ADGAN) to control image synthesis. The idea is to encode garment attribute of the input person image into a series of decomposed component codes and embed them into the latent space to build the full style code by feeding a concatenation of these codes into StyleGAN generator [28]. Then, another framework called Outfit-VITON [41] is proposed as an image-based virtual try-on approach dependent on splitting segmentation mask of garment image into various regions to be associated with encoded garments. This work is not able to change the pose of the target person which is ensured by our framework. In last year, Sarkar et al. [50, 51] proposed an efficient photo-realistic re-rendering system with high-quality results by adopting an explicit control over both style and pose for image synthesis. Before that, SwapNet [47] was proposed as one of the important virtual try-on framework based on conditioning network such as U-net [48] used to generate a mutually exclusive segmentation mask of the desired garment. Last year, a person generation framework called Dressing in Order (DiOr) [9] is proposed to support two main tasks, 2D pose transfer and virtual try-on, by explicitly encoding the shape and texture of each garment and editing them separately.

Our work falls within this objective by proposing a system able to realize the garment transfer between different images of persons. The most advantages of our work compared with the above methods are the possibility to allow transfer of all the garment item and not only such one garment in each class because our system is based on segmentation masks and not on garment classes; and also the high quality of generated results by preserving details and identity of target images. We are adopted the same process shown on SwapNet architecture [47], an end-to-end pipeline with specific sub-networks for each level combined to ensure the garment transfer from one person's image onto different target pose.

2.1.1 Image-to-image translation

Image-to-image translation (I2I) task focus on transferring appearances or other specific information from source image to target image while maintaining the content represen-

tations. Recently, I2I has drawn great interest and made huge development due to the appearance of its wide range of applications in many computer vision and image processing problems. Generative adversarial networks (GAN) [17] is the most used method for I2I next to variational auto-encoders [30]. In our work, we are intersted in using the GAN architectures due to their powerful abilities to ensure the image generation task. They have achieved a high-grade visual quality in different image processing problems such as conditional image generation that has broad several applications such as style transfer [22], and image synthesis [25, 43, 53, 60, 62, 63, 65].

To achieve the image synthesis task, Isola et al. [25] introduced a method based on conditional GANs [39] that is effective for synthesizing images from label maps. Zhu et al. [63] presented a network for learning to translate an image from a source image to a target one in the absence of paired examples. Therefore, In our work, we have interested in the same process to deal with the conditional image synthesis to generate images conditioning on specific constraint task by adopting a U-Net network [48] on the warp module of our architecture, which is based on two conditioning images where the first provides the desired garment and the second presents the target body. Additionally, we have focused on another form of conditional image synthesis by using conditional semantic labels to synthesize images following the given category labels [43, 53]. This method has a great advantage to add more diversity and control ability at the semantic segmentation to generate photo-realistic images. Other works [62, 65], adopted the exemplar-based image translation to synthesize photo-realistic image from the input in a distinct domain to generate the output that should have the style (e.g., color, texture) in consistency with the semantically corresponding objects in the exemplar. In our texture module, we have used a combination between these conditioned networks [43, 60, 62] to achieve realistic results.

3 Image-based appearance transfer

3.1 Problem statement

Garment transfer task presents a crucial challenge because successful transfer requires significant structural changes to source images, and as introduced previously, directly interchange texture details of desired clothes to target body presents overburden results due to the poor quality of transfer process. Given the Fig. 1 where I_C presented image of person with desired garment, and I_B , corresponding to person on target body, the large differences in body structure and clothes outlines between the input images make the task of directly transferring clothes items from I_C to I_B very difficult. Thus and instead of this process, we suggest to transfer, firstly, the clothing segmentation C_{seg} of I_C , based on B_{seg} and the body segmentation of I_B , to generate the desired clothes segmentation C'_{seg} . When C'_{seg} is generated, we pass to the step of transferring the texture details from I_C to I_B conditioned on C'_{seg} to obtain final result I_G .

The proposed solution must treat the image translation task to generate photo-realistic images conditioning on specific inputs such as semantic segmentation. However, reaching high fidelity image translation while maintaining high quality of generated images remains a big challenge due to the difficulty to control image styles and shapes. Thus, a typical system to ensure the image-based transfer task will be needed.

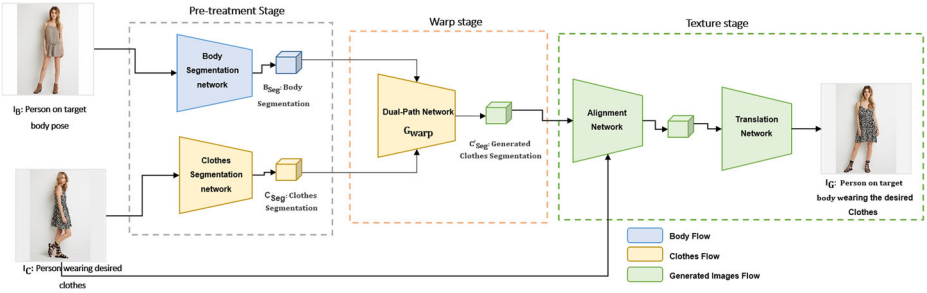

Fig. 1 An illustration of our framework consists of three stages. (1) The pre-treatment stage implemented to pre-process the input data. (2) The warp stage aims to generate intermediary clothes representation consistent with the target body. (3) The texture stage synthesizes detailed according to the output of warp stage

3.2 Challenge

Our work requires treating the challenge of jointly inferring the body pose and garment of person by solving three sub-problems. In our case, the adopted approach is to deal with the image content by transferring clothes information between input images. Firstly, the garment pieces must be identified from target clothes images. Then, the shape will be transferred across different poses. Finally, the clothes textures need to be synthesized in the target bodies with realistic details. A demonstration of clothing transfer results obtained with our work is presented in the following Fig. 2.

Therefore, our work focuses on garment interchange between pictures of different persons with no constraints on identity in the input and the output images. The clothing transfer is a challenging task requiring to disentangle the desired clothes from the corresponding person and retargeting it to a different body representations. To deal with this challenge, we

Fig. 2 Demonstration of clothing transfer task: *Dress-up* allowing the transfer of garment appearance between two different images

proposed our framework called *Dress-up* and we have presented a detailed explanation of our solution in the next section.

4 Dress-up

4.1 Overview

This section describes our garment transfer system called *Dress-up* illustrated in the following figure (Fig. 3), which is developed to interchange clothes between different views of images while maintaining details of the target garment and identity of the target person. This challenging task is achieved by implementing three stages, to change either the person or the clothes and recombine them as we desire. The scenario is as follow, taking an image representing the person with target clothes and another one concerned the person in target body, we generate a final output showing a person wearing the desired clothes from the first input image but on the pose of the second one.

The proposed pipeline aims to treat the pose and clothes synthesis, separately, by providing both the garment and the body segmentations obtained from based representations of desired clothes and target body. Therefore, we used, firstly, these segmentations to perform the target pose of I_B but with the clothes in I_C . The clothes segmentation of I_C and the body segmentation of I_B are obtained in the pre-treatment stage by using two different methods [33, 42]. In the third module, we presented a texturing network that takes the synthesized clothing segmentation and the desired clothing images as inputs to generate the final result. The architecture of the proposed framework is presented in Fig. 2, and the detailed representations of each stage are illustrated in Fig. 3 with a description in the following sections.

4.2 Functionalities

The process of our work is presented as an end-to-end pipeline to achieve three main functionalities (Fig. 3). Firstly, we used the Deep-Fashion dataset [36] to generate target segmentations by training two main networks [33, 42], the first [33] for the clothing segmentation and the second [42] for body segmentation. These representations obtained from the pretreatment stage will be an inputs to the warp stage of our framework, one image, showing the desired clothing segmentation C_{seg} of I_C , and other presented the body segmentation B_{seg} of the target body I_B . Then, we proceed to transfer the clothing segmentation C_{seg}, in

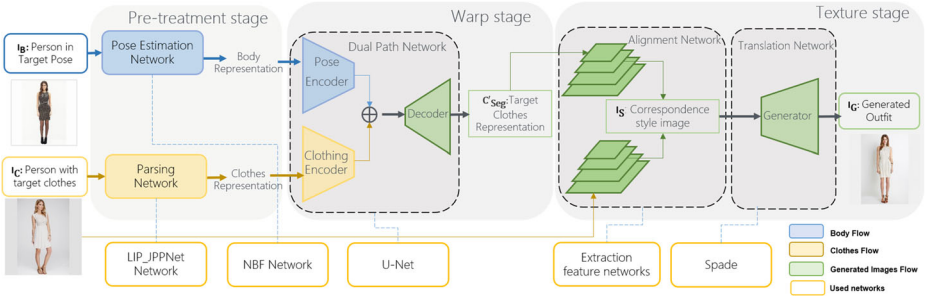

Fig. 3 Detailed illustration of the proposed pipeline

consitence with the body segmentation of B_{seg} to generate the target clothes segmentation C'_{seg}. The results from the pre-treatment stage and warp stage are presented respectively in Figs. 4 and 5.

Furthermore, to maintain the identity of target person before feeding it into texture stage, the face and hair segments must be replaced with appropriate segments in the clothes representation. Without this step, the target clothes will transferred in the same person instead of different persons. Finally and when the target clothes segmentation C'_{seg} is obtained, we pass to transfer the garment details from I_C to I_B conditioned on C'_{seg} for final output. Fig. 5 shows some obtained results.

4.3 Method

4.3.1 Pre-treatment stage

This stage preprocess data to obtain the two inputs required for training the model during the next stage (warp stage). These inputs (C_{seg} , B_{seg}) are considered respectively, as a concise representation of desired clothes and target body used to ensure the desired shape change by generating the synthesized segmentation C'_{seg} in the target body pose conditioning on the target clothes. This representation is considered as the input of the texture stage and used to obtain final output I_G .

For the pre-processing phase, we have trained existing networks [33, 42] on DeepFashion dataset [36], in particular, the In-shop Clothes Retrieval Benchmark, and extracting body and clothing representations from each input. In this step, we have followed the SwapNet work [47] which used the DeepFashion dataset to train the preprocessing networks. The authors resorted to LIP_SSL [16] and Unite-the People [31] for clothes and body segmentation respectively, but in our case, we replaced these networks with other more suitable models used for the same purposes. These choices are demonstrated in the following sections.

Clothes segmentation For clothes segmentation, we employed *LIP-JppNet* network [33] which is an updated version of *LIP_SSL* [16] with more accuracy on segmentation task. This network aims to associate into a single network different contexts such as the image-level context, the body joint and the part context and refined context (Fig. 6). It treats different

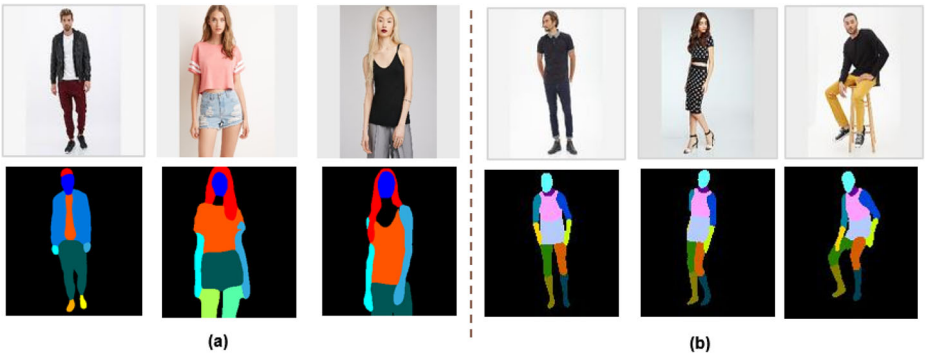

Fig. 4 Results of Pre-treatment stage consisting of two target segmentations: (a) body segmentation results obtained by using NBF network [42] and (b) clothes segmentation results generated by LIP-JppNet network [33]

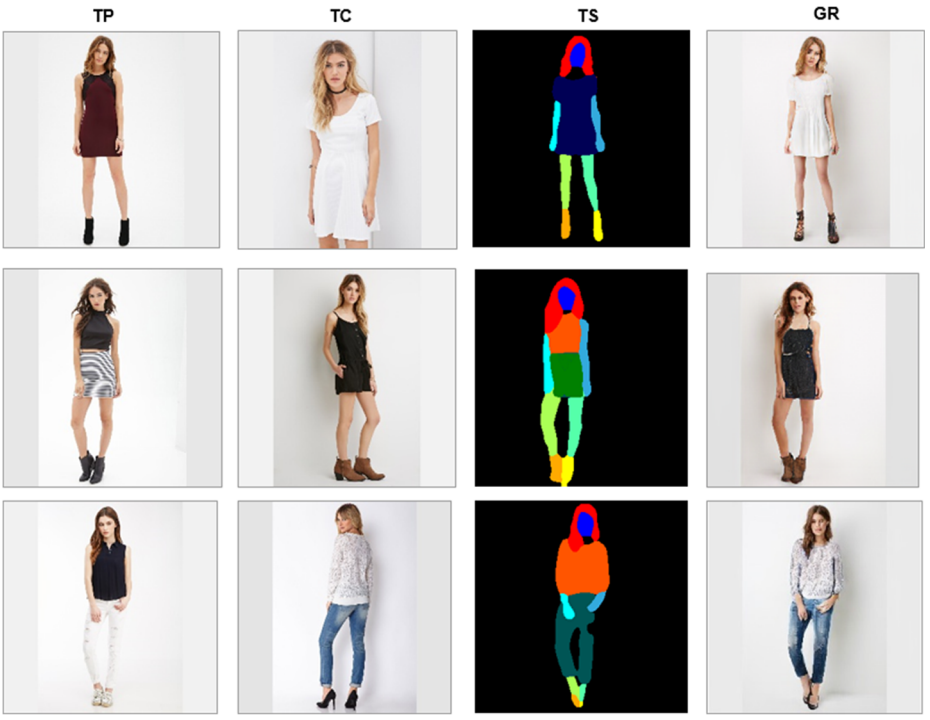

Fig. 5 Example of generated results with guidance of body segmentation. (TP: Target Pose, TC: Target Clothes, TS: Target Clothes segmentation; GR: Generated Result)

various challenging problems especially, shared feature extraction using *ResNet-101* [20], then, pixel-wise label prediction and keypoint heatmap prediction with a part module and a joint module to capture the part context and keypoint context while generating parsing score maps and pose heatmaps. Finally, an iterative refinement is used to predict maps and generate context to obtain better results.

The *SwapNet* work [47] represents the clothing segmentation labels as 18-channel maps which means that this representation is consisting on eighteen different clothes labels (e.g. dress, pants, and denim). This representation allowing the model to warp each individual segment separately therefore we need high capacity to work with this network. In the other hand, by using *LIP-JppNet* network, we have represented the desired clothing segmentation simply with 3-channel color-coded segmentation which can give sufficient information to generate the desired representation.

Body segmentation The second network is *Neural Body Fitting (NBF)* [42] with a standard semantic segmentation CNN into 12 semantic parts (Fig. 7). An encoding CNN processes the semantic part probability maps to predict a Skinned Multi-person Linear body model parameters (SMPL) [37]. Then an SMPL implementation is used to obtain a projection of the pose-defining points to 2D. With these points, a loss on 2D vertex positions can be back propagated through the entire model.

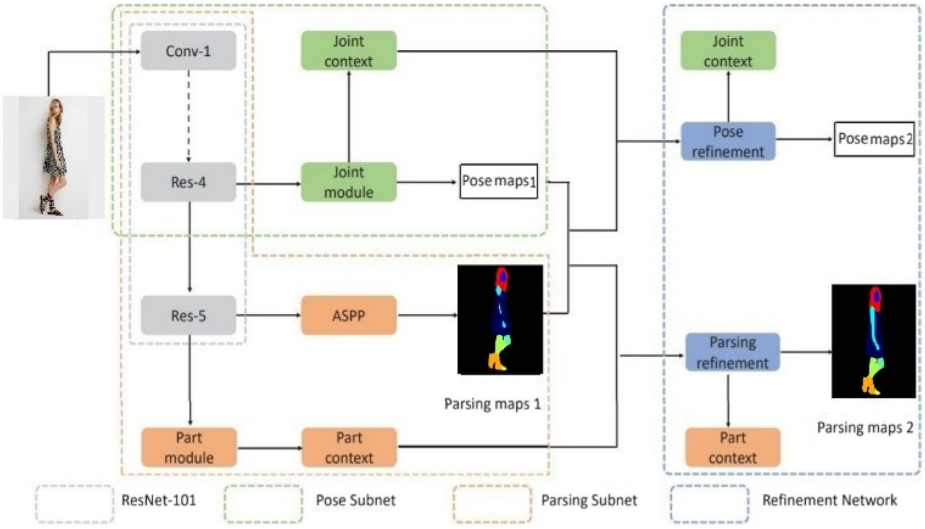

Fig. 6 Architecture of LIP-JppNet network [33]

In addition, *NBF* can treat 3d pose by regressing the parameters of *SMPL* from a single image consisting of an intermediate 2D body part segmentation but in our case, we simply ignore this part and interesting only on obtaining the 2D representation. Therefore,it is easier to deal with *NBF* network and run it with a simple implementation which is not the case with *Unite-the People* used by *SwapNet* that working in real-time. Also, the inaccuracy of *NBF* on such results cannot cause much concern on generated results because it can be obtained with high quality even with noisy representation from the preprocessing models.

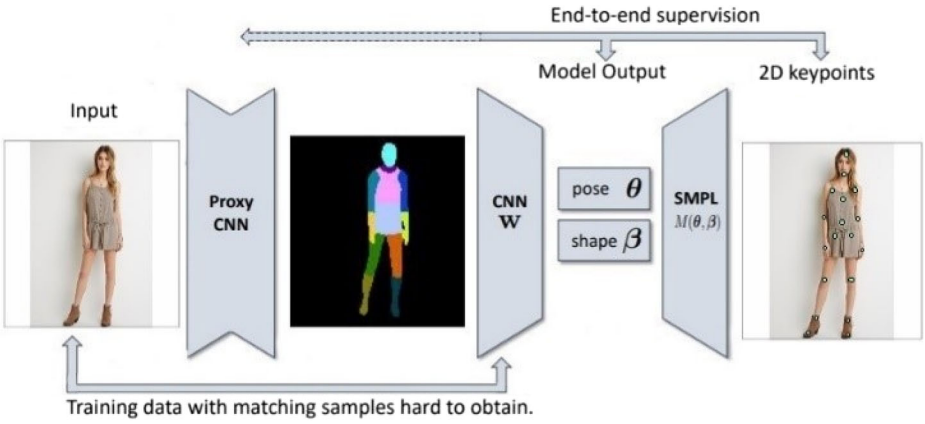

Fig. 7 Architecture of NBF network [42]

4.3.2 Warp stage

The warp stage is the second part of our pipeline, which takes as inputs the outputs of previous stage (C_{seg} , B_{seg}), respectively, the clothes segmentation and the body segmentation, to generate as output a target garment segmentation C'_{seg} consistent with C_{seg} , and strictly following the B_{seg} . This process is illustrated on Fig. 8. This process is based on a conditioned generative method, and precisely, on a dual path U-net network [48] to deal with the dual conditioning process consistent with the clothes and body segmentation. This network is composed of two encoders, one for the clothes and the other for the body, and a decoder to generate the final output by merging the two encoded representations. The generated segmentation C'_{seg} is robustly conditioned on the body representation and weakly conditioned on the clothes representation.

In this stage, we are followed the *SwapNet* work [47] which aims to represent the generated clothes with segmentation mask. The process is as follow, the clothes encoder produces a feature map of size $512 \times 16 \times 16$ for each clothes representation C_{seg}. In the other side, the body encoder produces also a feature map with the same size as C_{seg} (16×16 features of size 512) to represent the desired body segmentation B_{seg} . The generated image is strongly conditioned on B_{seg} and weakly conditioned on C_{seg} by encoding it into a narrow representation of $2 \times 2 \times 1024$, and then up-sampling it to a feature map of demanded size. After obtaining these encoded feature maps, the step of concatenation is needed and ensured by using 4 residual blocks to get a target feature map that had to be up-sampled to generate the desired clothing segmentation C'_{seg} . This representation makes the model more flexible to warp each individual segment separately.

As mentioned above, we implemented the warp module as a dual-path U-net [48]. The warping generator G_{warp} has an encoder-decoder architecture that synthesizes a new shape map C'_{seg} conditioned on body representation B_{seg} and clothes representation C_{seg} .

$$C'_{seg} = G_{warp}(B_{seg}, C_{seg}) \tag{1}$$

The warp stage is trained with the combination of cross entropy loss and GAN loss. Specifically, it has the following learning objectives:

$$L_{CE} = -\sum_{c=1}^{3} 1(C_{seg}(i,j) = c)(log(G_{warp}(i,j))) \tag{2}$$

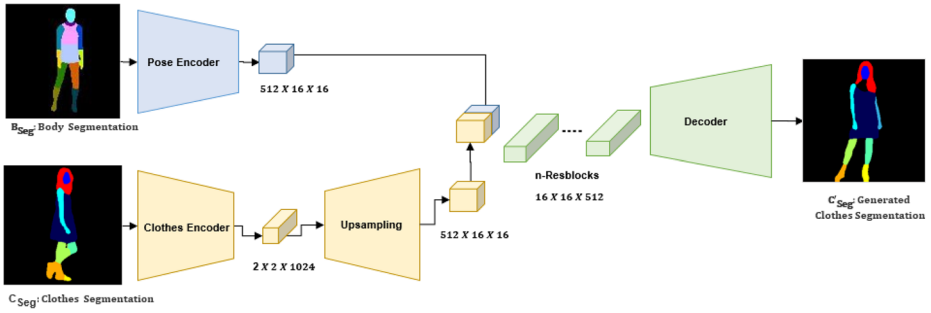

Fig. 8 Architecture of warp stage based on the dual-path U-net network with two conditioning representations, one on the body segmentation and the other on clothes segmentation, inspired from *SwapNet* work [47]

$$L_{adv} = E_{x\sim p_{(C_{seg})}}[D(x)] + E_{z\sim p_{(f1_{enc}(C_{seg},B_{seg}))}}[1 - D(f1_{dec}(z))] \quad (3)$$

$$L_{warp} = L_{CE} + \lambda_{adv} L_{adv} \quad (4)$$

where, $\lambda_{adv} L_{adv}$ refers to the adversarial component of the loss and $f1_{enc}$ and $f1_{dec}$ are the encoder and decoder components of the warp module.

4.3.3 Texture stage

The third stage in our pipeline is, the texture module, aiming to generate texture details. The followed process consisting of using the target clothes segmentation C'_{seg} to control the whole structure, and the image of person on target clothes to control texture generation. This module takes the generated clothes segmentation C'_{seg} conditioned on body information I_B and the reference clothes I_C , as inputs and synthesize an output I_G of person on target body I_B wearing the desired clothing on I_C . In other words, we aim to convert a segmentation mask to photorealistic image by extracting the high-level information of clothes representation and target style.

Similar to prior work [60], our texture stage consists of two essential parts (Fig. 9) working with an end-to-end manner: an alignment network and a translation network which are interconnected to synthesize realistic image from the semantic segmentation mask, and a reference style image. The alignment network aligns the features of conditional inputs and the translation network produces the final synthesis. The generated image contains the style in correspondence with the semantically objects in the reference image. Thus, the use of both the segmentation and image reference facilitate the task of translation with a weak supervised manner. The images are aligned to an intermediate representation to create the correspondence between the inputs and synthesize images according to the style image.

Alignment network The feature transport network aims to transport the feature of exemplars to be aligned with that of conditional inputs, thus providing accurate style guidance for the image translation. This network is inspired from the work of Zhang et al. [62] that used semantic correspondence to map the input images and create consistency representation of target style able to represent the semantics for both input images. As shown in Fig. 9, we first adapt the target segmentation body C'_{seg} and the image of target clothes I_C to a shared style image. In another words, C'_{seg} and I_C are entered to the feature pyramid networks (F_x,

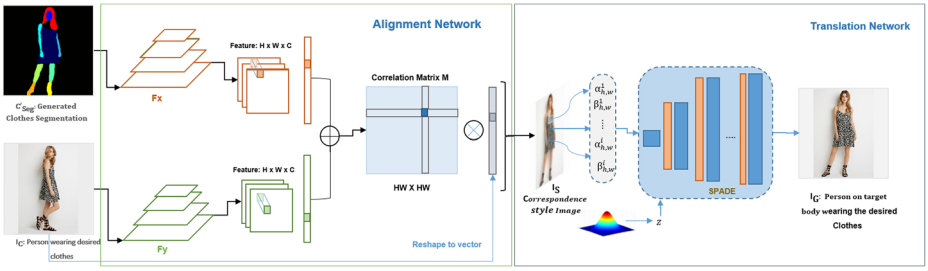

Fig. 9 Texture stage architecture: The Alignment network takes the generated segmentation maps and target clothes as inputs, and generates a correspondence style that will be entered to the Translation network to obtain the image of person on target clothes

F_y) that extract feature maps which will be converted to the following adapted representations: $T_B \in R^{HW*C}$ and $T_C \in R^{HW*C}$ (HW are feature spatial size; C is the channel-wise dimension).

These representations constitute the discriminative features that characterize the semantics of inputs. Thus, the alignment task presents an important step to realize the correspondence between different images. The features of T_B and T_C are matched with the correspondence layer proposed by the based work [60] in this task. Then, a correlation matrix M $\in R^{HW*HW}$ is computed:

$$M(u, v) = \frac{\widehat{T}_B(u)^T \widehat{T}_C(v)}{\left\|\widehat{T}_B(u)\right\| \left\|\widehat{T}_C(v)\right\|} \tag{5}$$

Where T_B(u) and T_C(v) $\in R^C$ represent the channel-wise centralized feature of T_B and T_C in position u and v. M (u, v) indicates a higher semantic similarity between T_B(u) and T_C(v).

The importance of this step resides on providing the easier way to the translation network to generate realistic results by referring to the correct regions in consistence with the reference clothes image I_C , which implicitly drives the model to learn the accurate correspondence. The correspondence style I_S is generated according to M by selecting the most correlated pixels in T_C and calculating their weighted average,

$$I_S(u) = \sum_v \underset{v}{softmax}(\alpha\ M(u, v))\ .\ T_C(v) \tag{6}$$

Where, α is the coefficient that controls the sharpness of Softmax.

Translation network According to correspondent style image I_S generated from the Alignment network, the translation network converts the constant code z to the final output I_G . With the aim to maintain the structural information of I_S , we used SPADE network [43] which has the ability to project the spatially adaptive style to different activation locations (Fig. 9), and adopts a specific strategy of details generation by avoiding the normalization and feeding the input image through spatially adaptive modulation. This network consisting of an adversarial trained encoder-decoder generator inspired by the idea of VAE [30] aiming to learn the mapping from a semantic representation to a reference image, and encoding its style into a latent style vector, from which the network generates output with desired style correspondent to reference image.

The specificity of the spatially-adaptive normalization is its simple architecture with a simple layer used to synthesize photorealistic images by using reference representation as input. SPADE overcomes the lack of poor quality by taking the input layout through this layer and learned transformation to modulate the activations in normalization layers. The use of this network in texture stage demonstrate its advantage by allowing the control over style details.

Despite the great advantage of SPADE network to generate directly the desired appearance, there is an important point to mention about the style code which can only characterize the global style of the target style image, whatever the relevant spatial information, and produces some local style "wash away" in generated image. For this reason, the built of alignment network as an intermediate step is required before using this network.

This step consisting in creating a correspondent style to the target image to guide the image translation by proposing an alignment network that transforms the input images to an intermediate representation to create the correspondence between them. During training, the correspondent style image I_S feeds into the SPADE layers, projected onto an embedding

space, convolved to produce the modulation parameters α and β which characterize the style of clothes image, and mapped from I_C. Given the activation $F^i \in R^{C_i X H_i X W_i}$ before the i^{th} normalization layer, the reference style is injected through:

$$\alpha^i_{h,w}(I_S) X \frac{F^i_{c,h,w} - \mu^i_{h,w}}{\sigma^i_{h,w}} + \beta^i_{h,w}(I_S) \tag{7}$$

where $\mu^i_{h,w}$ and $\sigma^i_{h,w}$ are the mean and standard deviation of activations and the symbols $\alpha^i_{h,w}$, $\beta^i_{h,w}$ to denote the functions that convert m to the scaling and bias values at the site (h, w) in the i^{th} activation map.

At this stage, we used the adversarial loss by training a discriminator that discriminates the generated outputs and the real samples of target clothes. The translation network G and the discriminator D are trained until the generation of images looking very similar to the real ones. Therefore, the adversarial objectives of D and G are respectively defined as:

$$L^D_{adv} = -E[h(D(T_C))] - E[h(-D(G(T_B, T_C))] \tag{8}$$

$$L^G_{adv} = -E[D(G(T_B, T_C))] \tag{9}$$

As a summary, the texture stage consisting of two related networks: i) Alignment Network aiming to convert the inputs from two different images to an intermediate feature images useful for obtaining correspondent style. ii) The translation network is based on the SPADE network [43] to generate the final result, employing the details of desired style from a generated style aligned semantically to the mask consisting on the estimated correspondence. These two networks work in a complementary way to ensure the objective of the texture stage that aims to generate photo-realistic images.

5 Experiments

In the experiments section, we present the details of implementation process realized to develop our framework, then, we expose the results of different modules of our architecture before presenting quantitative and qualitative comparisons with baselines and discussing our findings.

5.1 Implementation details

Dataset We conduct experiments on the *DeepFashion* dataset [36] and precisely, the *Inshop Clothes Retrieval* Benchmark which containing 52,712 training images captured from fashion images of men and women models. We choose this dataset because it is the largest existing fashion dataset and it shows lots of diversities of various fashion images. Due to the difficulty to show different views of input images with similar clothes on distinct bodies, we have turned to the data augmentation technique.

Network implementation and training details *Dress-up* is implemented in Python using PyTorch. Basically, encoder-decoder structures are employed to design our pipeline. The residual block is used as the basic component of these models. We train our model using 256 $\times$ 256 images for the DeepFashion dataset [36]. Architectural details for the main *Dress-up* components are given below.

The pre-tratment module It consists of two main task: (i) Clothing segmentation: based on ResNet-101 [20] as the basic architecture and employs atrous convolution, multiscale inputs with max-pooling to merge the results from all scales, and atrous spatial pyramid pooling. ResNet-101 is trained on the human parsing task and we have applied data augmentation, including randomly scaling, cropping and leftright flipping to perform the training process. (ii) Body segmentation: The network produces a part segmentation from cropped images, a RefineNet [35] model based on ResNet-101 [20] is used. This part segmentation is color-coded to fed as an RGB image to a regression network (ResNet-50) that outputs the SMPL parameters (shape and pose).

The warp module In this stage, the input images and their corresponding segmentation maps are used to train the network. It is based on U-Net network [48] that learns segmentation in an end-to-end setting. We input a raw image and get a segmentation map as the output. We used data augmentation to reduce the number of annotated images required for training. To supervise the training, we constructed a ground-truth triplets, and we used a self-supervised approach to generate this required triplets. We admitted data augmentations and we performed random affine transformations to guide the network to discard locational cues from clothing segmentation and pick up only high-level cues regarding the clothing segments. The warping module is trained with the combination of cross entropy loss and GAN loss.

The texture module We use Adam solver [29] with $\beta 1 = 0$, $\beta 2 = 0.999$. We set imbalanced learning rates, 1e$-$4 for the generator. Spectral normalization [40] is applied to all the layers for both networks to stabilize the adversarial training. Also, a synchronized batch normalization are used to synchronize the mean and variance statistics across multiple GPUs. Experiments are conducted on at least 4 P40 GPUs.

5.2 Comparisons

To evaluate the effectiveness of our proposed framework, we choose to deal with different tasks treated by our solution, the first one is the garment transfer task in which we elaborate the qualitative results in comparison with popular virtual try-on systems. Then, to show the performance of our results in quantitative criteria, we used the pose transfer task to compare outputs with ground-truth images, and finally, another comparison is realised to evaluate the main task of our framework that is the label-to-image translation.

5.2.1 Garment transfer

In garment transfer task, we conduct a comparison with leading image based transfer methods: the *SwapNet* [47] system, the *Deep Re-rendering* system [51]. *SwapNet* is a GAN-based conditional image synthesis framework proposed to interchange clothes between two images of persons. The *Deep Re-rendering* uses a neural image-translation network to ensure the human image synthesis task by allowing the change of pose and clothes of the person in the source image. For comparison, we used the results provided by the authors on the DeepFashion dataset [36]. The visual comparisons of *Dress-up* and these methods are shown in Figs. 10 and 11 respectively. Additionally, extra examples of our work can be found in Fig. 12.

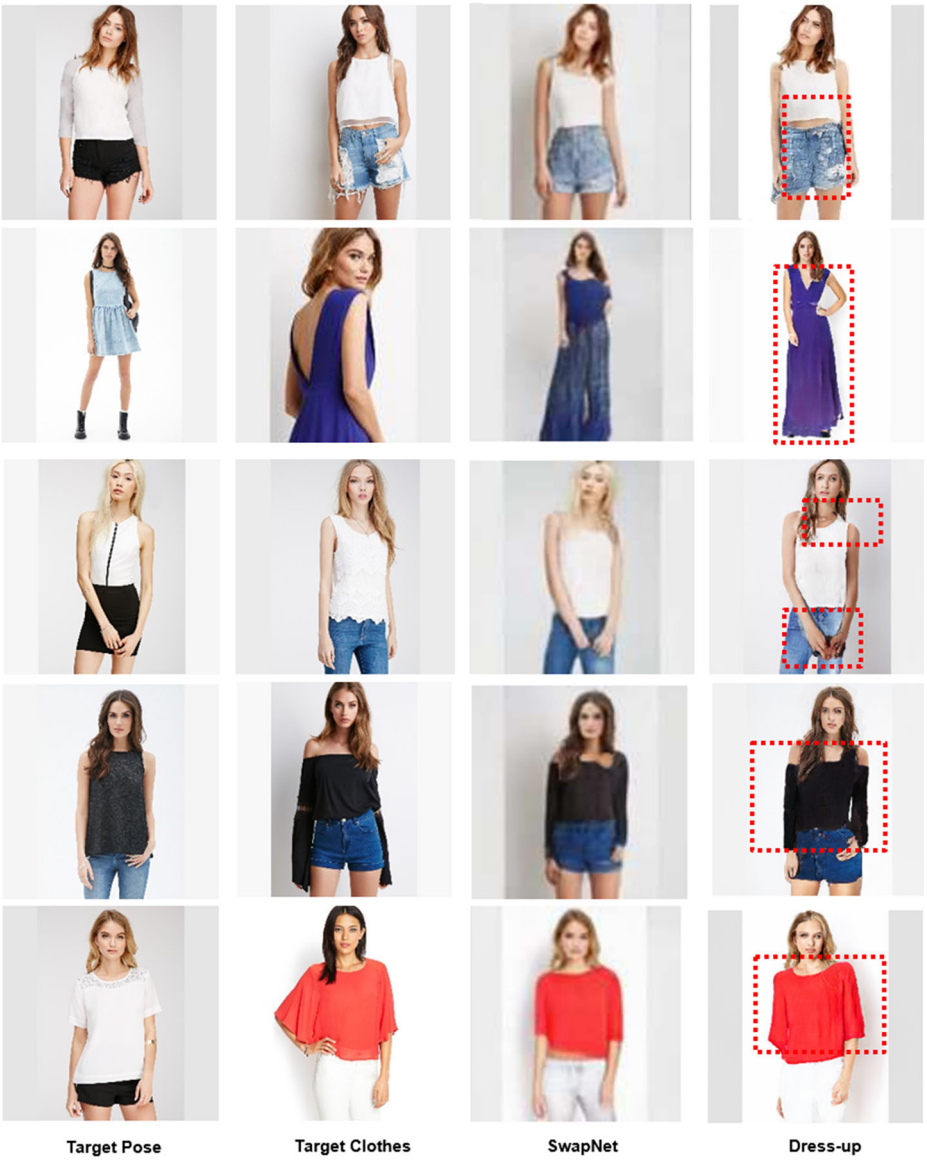

Fig. 10 Comparison between Dress-up and SwapNet [47] on the DeepFashion dataset

These different comparisons with these important baselines are treated to improve the effectivness of our proposed framework for virtually try-on the clothes. According to these findings presented in Figs. 10 and 11, our method presents realistic results and overcomes different issues not resolved by SwapNet and Re-rendering that have difficulty handling various pose changes and cannot hallucinate details for occluded regions. Further analysis of these comparisons are detailed in the discussion section.

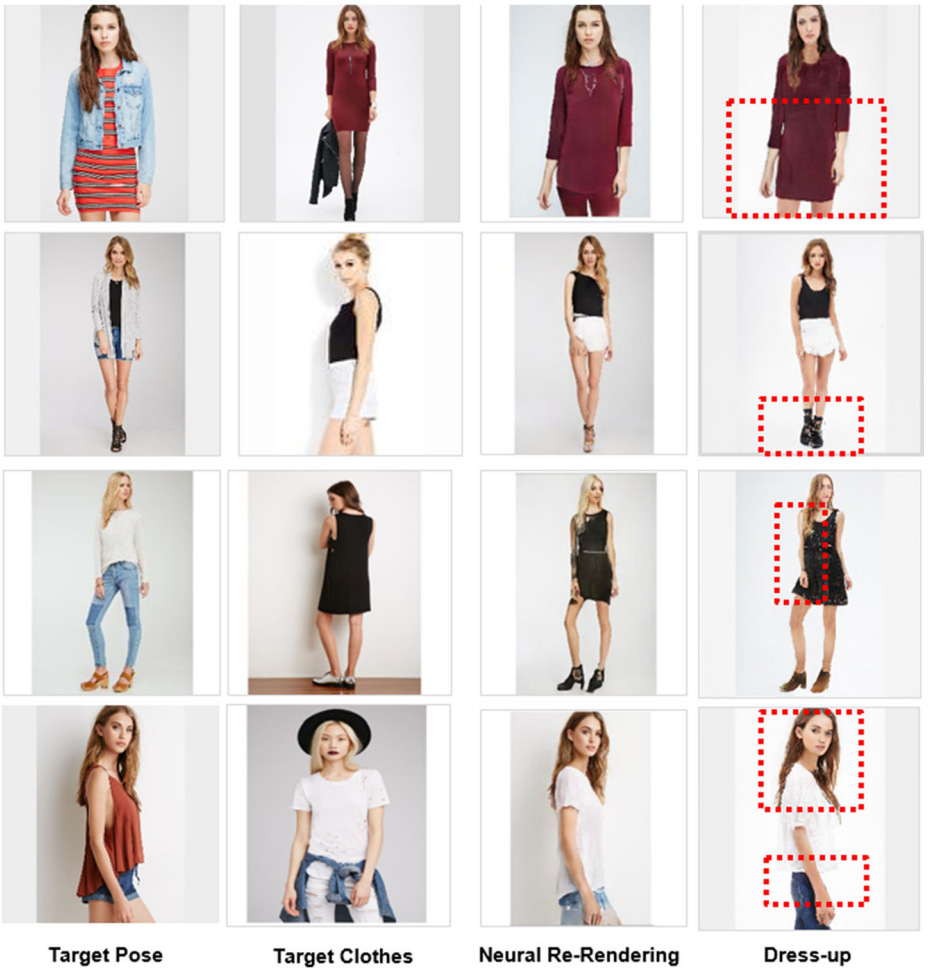

Fig. 11 Comparison between Dress-up and Deep-Re-rendering [51] on DeepFashion dataset

5.2.2 Pose transfer

We run automatic evaluation on the pose transfer task, which is the only one that has reference images available. Figure 12 shows a qualitative comparison of our method with baselines treated the pose transfer task. Table 1 shows a comparison of our results with Ivan et al. [14], ADGAN [38], and DiOr [9], these methods have codes and models publicly available, and we have used the DeepFashio dataset to proceed this comparison.

Quantitative results are presented on different models using several common metrics purporting to measure the similarity between generated and real reference images to evaluate the performance of our work and compare the quality of obtained results with baselines methods: (i) structural similarity (SSIM) [23], computes the similarity between input and output images ranging from zero (dissimilarity) to one (similarity). (ii) Frechet Inception Score (FID) [21] to measure the distance between the distributions of synthesized images and real images, and (iii) LPIPS [61] is used to assess the generation diversity of a model by computing the weighted distance between deep features of image pairs.

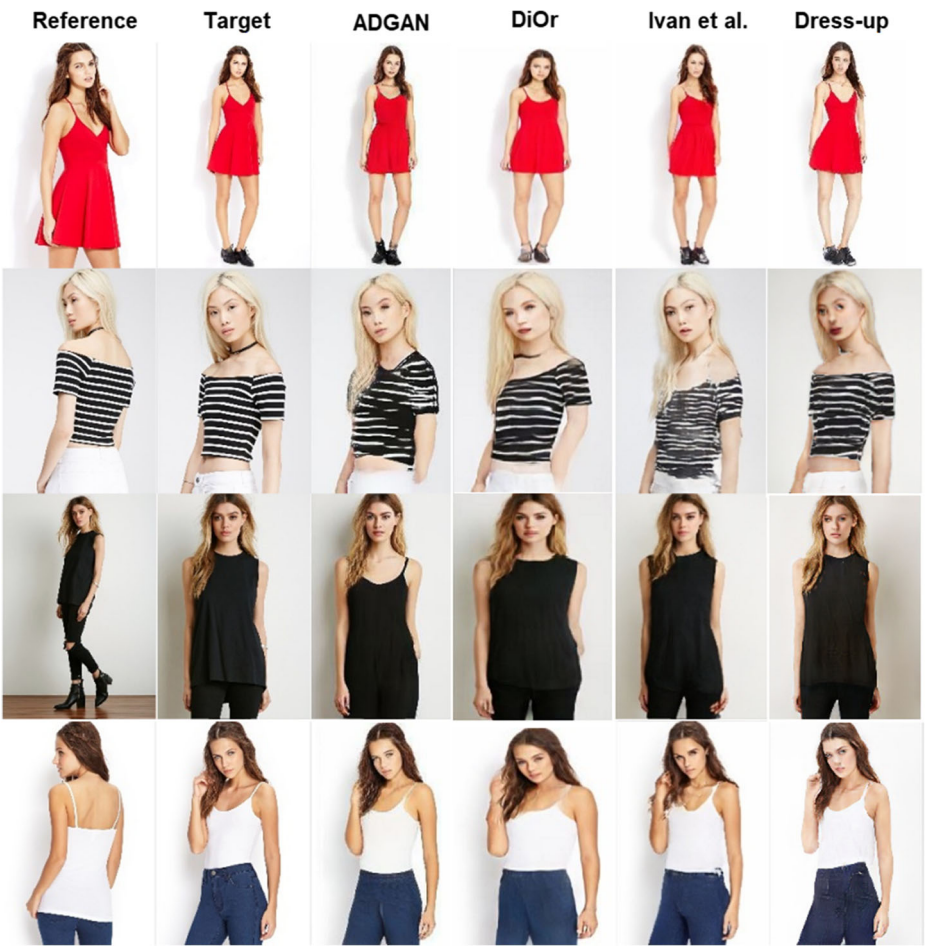

Fig. 12 Qualitative comparison of our model with several pose transfer methods on the DeepFashion dataset

According to Table 1, we observe that SSIM, FID and LPIPS metrics provide a clear proxy to measure performance and we can say that *Dress-up* performs the best in comparison with baselines models. There, our output is qualitatively produce a high quality, and consistently better than baselines methods. We detailed these results in the discussion section.

Table 1 Quantitative comparison of Pose Transfer models on DeepFashion Dataset. Bold font indicates best results

Method	SSIM ↑	LPIPS ↑	FID ↓
ADGAN [38]	0.77	0.226	16.69
DioR [9]	0.80	0.176	14.34
Ivan et al. [14]	0.71	0.175	12.74
Ours (Dress-up)	**0.95**	**0.228**	**8.60**

5.2.3 Label-to-image translation

In this section, We compare some label-to-image algorithms using the FID, Intersection-over-Union (mIoU) and pixel accuracy (Acc) metrics to mesure the accuracy of generated images and evaluate the quality of our method. We choose these state-of-the-art methods based on semantic labels: pix2pixHD [53], SPADE [43] and SMIS [65] as baselines for our comparison that is performed across the DeepFashion dataset. Quantitative results are shown in Table 2, and we can conclude from the obtained values that our network generally retains approximately the same performance as SPADE because a part of our work is based on this method, the same case for SMIS becaus this last is also based on Spade network. With other criteria such as FID, the superiority of *Dress-up* is proved and our generated result are similar to the ground truth images. This interpretation is demonstrated also with Qualitative comparisons illustrated in Fig. 13. In general, the images generated by *Dress-up* are more realistic and plausible than others. These visual results consistently show the high image quality of Dress-up's generated images, verifying its efficacy on DeepFashion dataset. In addition, these results presented a powerful characteristic of our proposed solution consisting in its ability to preserve the original texture according to the given example provided as reference image and this is the advantage of our method that is based on a correspondence approach using a reference image to generate the desired style similar to it.

5.3 Discussion

As presented above, to demonstrate the performance of our method, we realized comparisons in different tasks and with competing methods. We find that our method produces results with much better visual quality and fewer visible artifacts, especially for diverse scenes in the *DeepFashion* dataset, and the quantitative results in different tasks are very promissing. In this section, we demonstrate some of Dress-up's strengths and weaknesses by: (i) presenting virtual try-on results for multiple target garments and different subjects, (ii) high-lighting the performance of pose transfer, and (iii) illustrating some of the model's limitations.

5.3.1 Effectivness of dress-up

Garment transfer Despite the interesting results obtained with the baselines methods in the garment transfer task, our model shows more high details quality in the appearance of the clothes item (e.g. texture, color, shoes) which have been properly transferred onto the target body. Figures 10 and 11 proved this affirmation by showing visual try-on results for garment transfer task which demonstrate that our method generates more realistic outfit and better preserve the shape and texture of the target clothing in the final try-on result. As illustrated by the example in the top row, the better alignment ensured by *Dress-up* leads to a correctly

Table 2 Quantitative comparison with label-to-image models on DeepFashion Dataset. Bold font indicates best results

Method	mIoU ↑	Acc ↑	FID ↓
Pix2pixHD [53]	85.2	98.8	17.76
SMIS [65]	87.3	**98.9**	9.50
SPADE [43]	87.1	**98.9**	10.02
Ours (Dress-up)	**87.6**	**98.9**	**8.60**

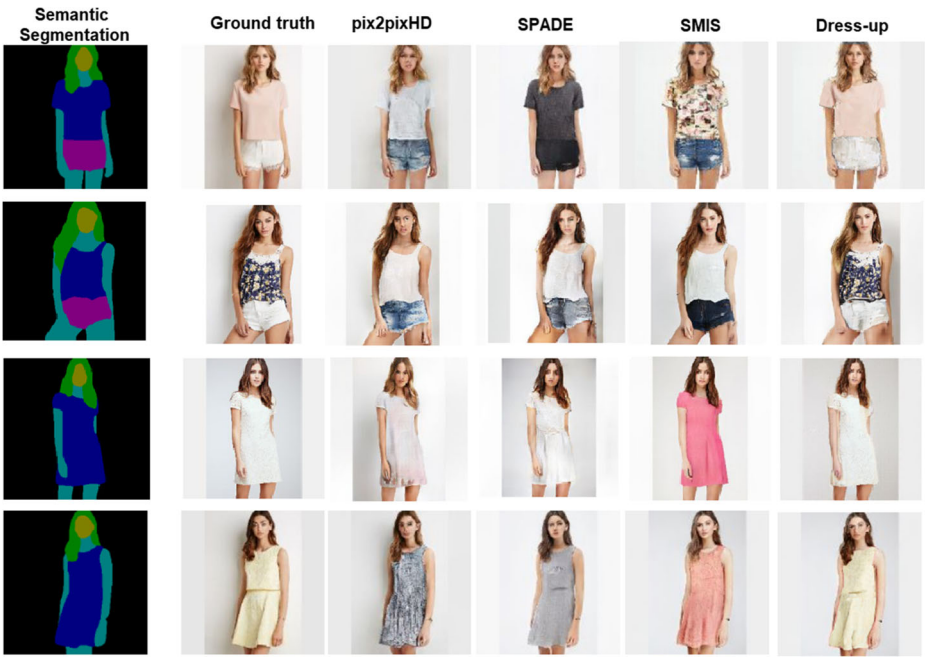

Fig. 13 Qualitative comparison of our model with several label-to-image methods on the DeepFashion dataset

transfer the appropriate texture. Also, the second row of Fig. 10 and the first row of Fig. 11 show that our model is very close for generation details even in case of large pose deviation (e.g. interchange garment from a person in full body to a person in half body). The shoes, the hair, and varying sleeve lengths are better preserved with *Dress-up* (surrounded by a red rectangle in Figs. 10 and 11). More results of the garment transfer task are presented in Fig. 14 in which we have showed more examples in different body and garment variations.

Pose Changes *Dress-up* has successfully solving wide pose changes between source and target images (Fig. 12). For example, when an input images contain a truncated body and other contains a full body, our model is able to generate high quality details. Furthermore, our framework is capable to manage differents poses variations according to the body segmentation which provides better guidance to the reference body. It can be seen from the quantitative and qualitative results for all the examples that the proposed method leads the performance in SSIM, FID and LPIPS metrics. Furthermore, Fig. 12 shows that our method presents a much richer and detailed visual appearance of the target outfit. More results in pose changes are presented in the following Fig. 15.

5.3.2 Limitations

While our results are promising, there remain a number of limitations and failure modes. Some of these are illustrated in Fig. 16: complex or rarely seen poses are not always transfered correctly, some artifacts are present, and such garments are not always filled in properly. More generally, the shading, texture warping, and garment detail preservation of

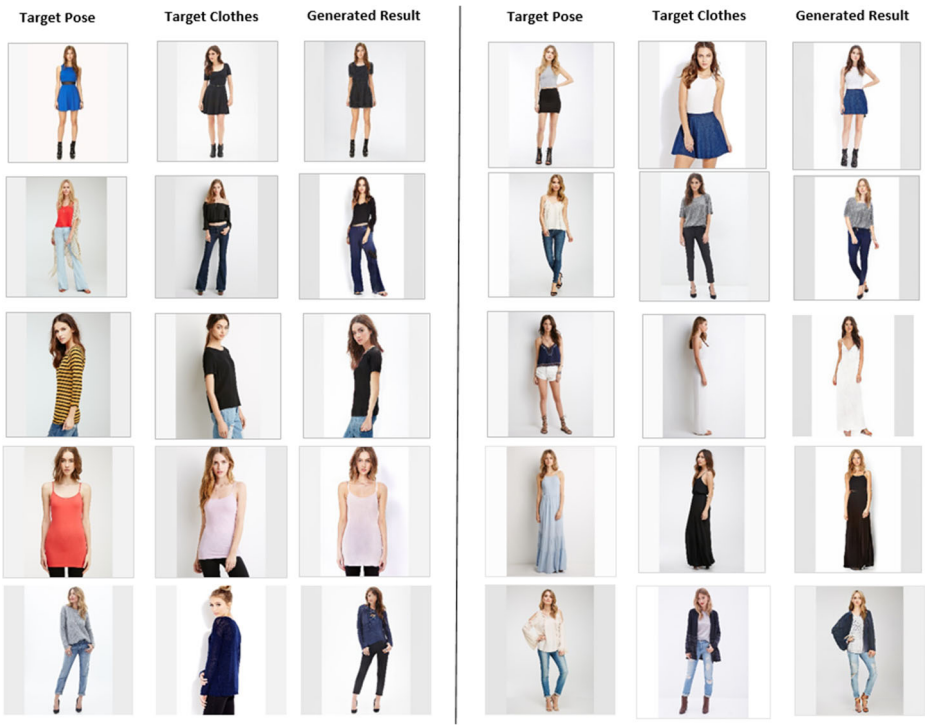

Fig. 14 Examples of obtained results demonstrate the effectiveness of dress-up for clothing transfer

our method, while better than those of other recent methods, are still not entirely realistic in some cases. Since we use an intermediate representation, the try-on results are depended on it and present some failures in special cases.

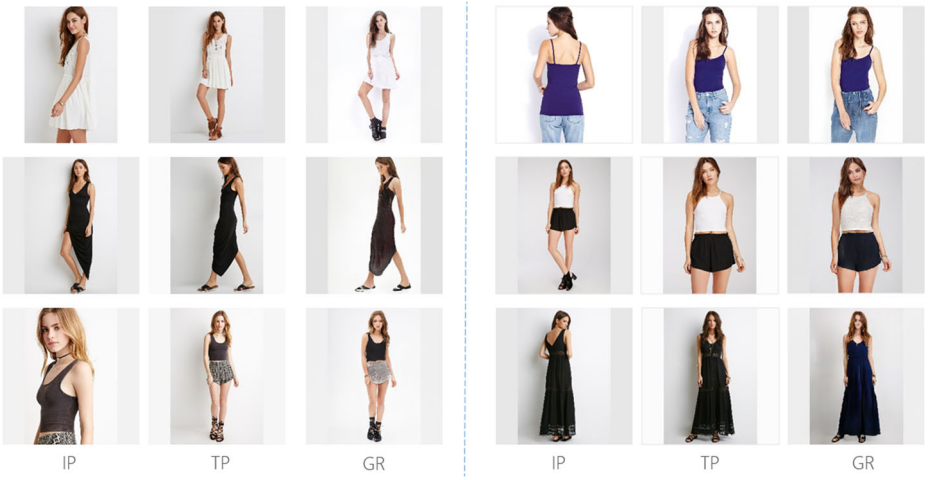

Fig. 15 Pose transfer results on DeepFashion dataset: IP: Initial Pose, TP: Target Pose, GR: Generated Result

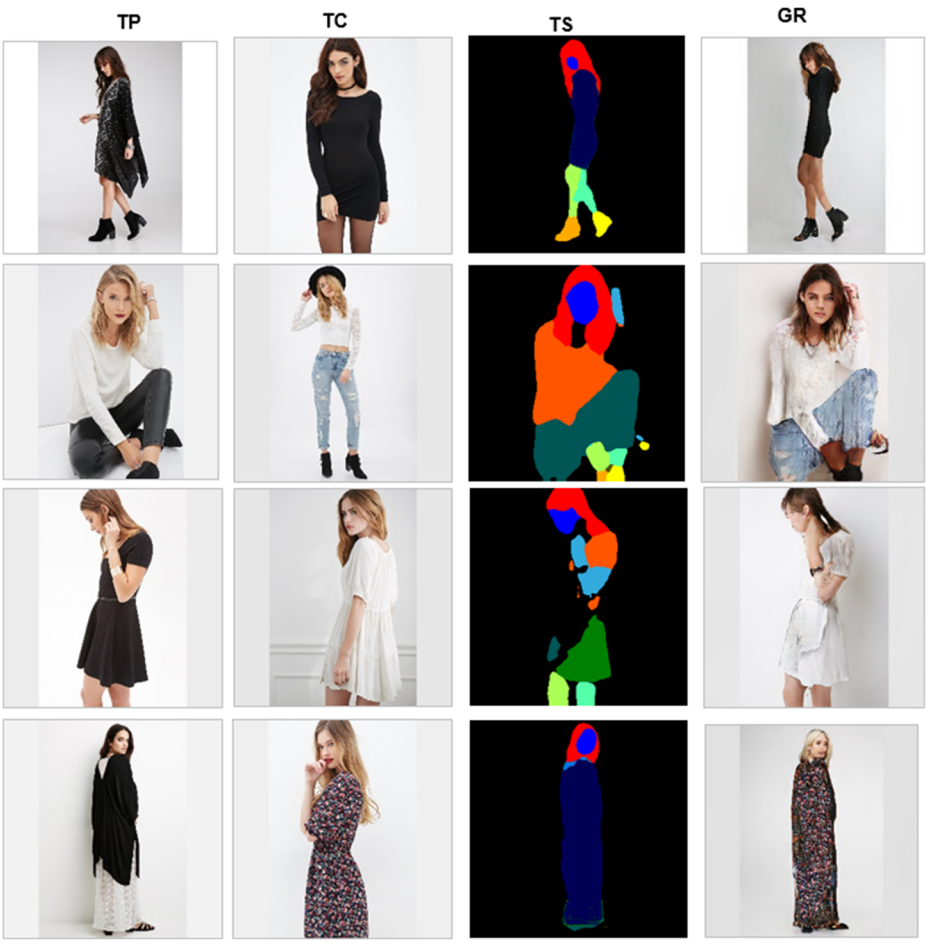

Fig. 16 Limitations of Dress-up for clothing transfer: Results demonstrate extreme pose changes (TP: Target Pose, TC: Target Clothes, TS: Target segmentation; GR: Generated image)

In the previous section, we demonstrate the convenience of adopting clothes segmentation as an intermediate representation (Fig. 5). In cases where the generated clothes is inappropriate, we can edit this representation to better fit the clothes. The following Fig. 16 shows the limitation related to some obtained results from our framework in case of unsuitable clothes segmentation. The solution to deal with this issue is to edit the warp stage to generate the right segmentation even in large pose variations between source and target images.

5.3.3 Future work

With our solution, we have demonstrated promising results in high resolution on a challenging task of try-on. While that, our method still fails in cases of extreme poses and complicated garments. The try-on application is designed to visualize fashion on any person, including different body shapes, height, weight, in the highest quality. However, any

deployment of our method in a real-world setting would need more precision to ensure the design decisions.

For this aim, we plan to improve our results and we will pay more attention to the following directions: (i) Improve the intermediate representation generated by the warp module. We will examine in detail how much good segmentations of human model images can improve the overall results and improve the fitting performance without being limited by the original reference person. (ii) Being able to change all fashion item categories (fats, accessories, bags) on a human picture to ensure full virtual try-on experience. (iii) Improving the quality of our output through more advanced warping and higher-resolution training and generation. (iv) Improving the results when swapping clothing with all kind of textures – currently the *Dress-up* is inaccurate with complex textures. (v) Another interesting topic would be treated to improve the scene compositing for a more realistic appearance transfer.

6 Conclusion

Virtual garments try-on system is a system in which users can wear different garments without actually wearing the clothes. Thus, these solutions could improve the shopping experience by assisting users to make purchase decisions and giving them the possibility to experiment the virtual fitting without the integration of human interaction. To achieve that, we proposed a personalized virtual try-on system called *Dress-up* capable of synthesizing high-quality try-on results by considering a wide range of input-image characteristics. The main idea of our framework is to incorporate different strategies to extract and to transfer human appearance between two real person views. Differently from classic retargeting methods that directly transfer both the shape and the texture details of the desired clothing to a target body which gives the network too much burden resulting in poor transfer quality, our solution uses an intermediate representations for the body and the texture style to realize garment transfer with photorealistic outputs, and learning the semantic correspondence in full-resolution. Experimental results on DeepFashion dataset demonstrated that this strategy enables more realistic appearance transfer across images and flexible control over pose. Furthermore, our method is not only well suited to generate person images but also can be potentially adapted to other image synthesis tasks especially those with insufficient data annotation.

Acknowledgements We deeply acknowledge Taif University for Supporting this study through Taif University Researchers Supporting Project number (TURSP-2020/327), Taif University, Taif, Saudi Arabia. The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Arashpour M, Ngo T, Li H (2021) Scene understanding in construction and buildings using image processing methods: a comprehensive review and a case study. J Build Eng 33:101672

2. Bhatti UA, Huang M, Wang H, Zhang Y, Mehmood A, Di W (2018) Recommendation system for immunization coverage and monitoring. Human Vaccines Immunotherap 14(1):165–171
3. Bhatti UA, Huang M, Wu D, Zhang Y, Mehmood A, Han H (2019) Recommendation system using feature extraction and pattern recognition in clinical care systems. Enterprise Inf Syst 13(3):329–351
4. Bhatti UA, Yu Z, Chanussot J, Zeeshan Z, Yuan L, Luo W, Nawaz SA, Mehmood A (2021) Local Similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and Gabor filtering. IEEE Trans Geosci Remote Sensing 60:1–15
5. Bhatti UA, Yuan L, Yu Z, Li J, Nawaz SA, Mehmood A, Zhang K (2021) New watermarking algorithm utilizing quaternion Fourier transform with advanced scrambling and secure encryption. Multimed Tools Appl 80(9):13367–13387
6. Bhatti UA, Zeeshan Z, Nizamani M, Bazai S, Yu Z, Yuan L (2022) Assessing the change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19. Chemosphere 288:132569
7. 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, pp 14131–14140
8. Chowdhary CL, Patel PV, Kathrotia KJ, Attique M, Perumal K, Ijaz MF (2020) Analytical study of hybrid techniques for image encryption and decryption. Sensors 20(18):5162
9. Cui A, McKee D, Lazebnik S (2021) Dressing in order: recurrent person image generation for pose transfer, virtual try-on and outfit editing. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 14638–14647
10. De Bogotá CDC (2021) The state of fashion 2021
11. Dong H, Liang X, Shen X et al (2019) Towards multi-pose guided virtual try-on network. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 9026–9035
12. Etoundi CML, Nkapkop JD, Tsafack N, Ngono JM, Ele P, Wozniak M, Shafi J, Ijaz MF (2022) A novel compound-coupled hyperchaotic map for image encryption. Symmetry 14(3):493
13. Fele B, Lampe A, Peer P, 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, pp 3144–3153
14. Fincato M, Cornia M, Landi F, Cesari F, Cucchiara R (2022) Transform, warp, and dress: a new transformation-guided model for virtual try-on. ACM Trans Multimed Comput Commun Appl (TOMM) 18(2):1–24
15. Ghodhbani H, Neji M, Razzak I, Alimi AM (2022) You can try without visiting: a comprehensive survey on virtually try-on outfits. Multimed Tools Appl:1–32
16. Gong K, Liang X, Zhang D, Shen X, Lin L (2017) Look into person: self-supervised structure-sensitive learning and a new benchmark for human parsing. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 932–940
17. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inf Process Syst, vol 27
18. Guan P, Reiss L, Hirshberg DA, Weiss A, Black MJ (2012) Drape: dressing any person. ACM Trans Graph (ToG) 31(4):1–10
19. Han X, Wu Z, Wu Z, Yu R, Davis LS (2018) Viton: an image-based virtual try-on network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7543–7552
20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
21. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. Adv Neural Inf Process Syst, vol 30
22. Hobley MA, Prisacariu VA (2018) Say yes to the dress: shape and style transfer using conditional GANs. In: Asian conference on computer vision. Springer, Cham, (pp 135-149)
23. Hore A, Ziou D (2010) Image quality metrics: PSNR vs. SSIM. In: 2010 20th International conference on pattern recognition. IEEE, pp 2366-2369
24. Hussain R, Karbhari Y, Ijaz MF, Woźniak M, Singh PK, Sarkar R (2021) Revise-net: exploiting reverse attention mechanism for salient object detection. Remote Sensing 13(23):4941
25. Isola P, Zhu JY, Zhou T, Efros A (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
26. Ivan VA, Mistreanu I, Leica A, Yoon SJ, Cheon M, Lee J, Oh J (2021) Improving key human features for pose transfer. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 1963–1972
27. Jetchev N, Bergmann U (2017) The conditional analogy gan: swapping fashion articles on people images. In: Proceedings of the IEEE international conference on computer vision workshops, pp 2287–2292

28. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4401–4410
29. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980
30. Kingma DP, Welling M (2014) Auto-encoding variational bayes
31. Lassner C, Romero J, Kiefel M, Bogo F, Black MJ, Gehler PV (2017) Unite the people: closing the loop between 3d and 2d human representations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6050–6059
32. Lewis KM, Varadharajan S, Kemelmacher-Shlizerman I (2021) Tryon-gan: body-aware try-on via layered interpolation. ACM Trans Graph (TOG) 40(4):1–10
33. Liang X, Gong K, Shen X, Lin L (2018) Look into person: joint body parsing & pose estimation network and a new benchmark. IEEE Trans Pattern Anal Mach Intell 41(4):871–885
34. Liao J, Yao Y, Yuan L, Hua G, Kang SB (2017) Visual attribute transfer through deep image analogy. SIGGRAPH
35. Lin G, Milan A, Shen C, Reid I (2017) Refinenet: multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1925–1934
36. Liu Z, Luo P, Qiu S, Wang X, Tang X (2016) Deepfashion: powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1096–1104
37. Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ (2015) SMPL: a skinned multi-person linear model. ACM Trans Graph (TOG) 34(6):1–16
38. Men Y, Mao Y, Jiang Y, Ma WY, Lian Z (2020) Controllable person image synthesis with attribute-decomposed gan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5084–5093
39. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:1411.1784
40. Miyato T, Kataoka T, Koyama M, Yoshida Y (2018) Spectral normalization for generative adversarial networks. arXiv:1802.05957
41. Neuberger A, Borenstein E, Hilleli B, Oks E, Alpert S (2020) Image based virtual try-on network from unpaired data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5184–5193
42. Omran M, Lassner C, Pons-Moll G, Gehler P, Schiele B (2018) Neural body fitting: unifying deep learning and model based human pose and shape estimation. In: 2018 International conference on 3D vision (3DV). IEEE, pp 484-494
43. Park T, Liu MY, Wang TC, Zhu JY (2019) Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2337–2346
44. Patel C, Liao Z, Pons-Moll G (2020) Tailornet: predicting clothing in 3d as a function of human pose, shape and garment style. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7365–7375
45. Pons-Moll G, Pujades S, Hu S, Black MJ (2017) ClothCap: seamless 4D clothing capture and retargeting. ACM Trans Graph (ToG) 36(4):1–15
46. Rahman M (2021) Applications of the digital technologies in textile and fashion manufacturing industry. Technium: Romanian J Appl Sci Technol 3(1):114–127
47. Raj A, Sangkloy P, Chang H, Lu J, Ceylan D, Hays J (2018) Swapnet: garment transfer in single view images. In: Proceedings of the european conference on computer vision (ECCV), pp 666–682
48. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 234–241
49. Santesteban I, Otaduy MA, Casas D (2019) Learning-based animation of clothing for virtual try-on. Comput Graph Forum 38(2):355–366
50. Sarkar K, Mehta D, Xu W, Golyanik V, Theobalt C (2020) Neural re-rendering of humans from a single image. In: European conference on computer vision. Springer, Cham, pp 596-613
51. Sarkar K, Golyanik V, Liu L, Theobalt C (2021) Style and pose control for image synthesis of humans from a single monocular view. arXiv:2102.11263
52. Tamang J, Nkapkop JD, Ijaz MF, Prasad PK, Tsafack N, Saha A, Kengne J, Son Y (2021) Dynamical properties of ion-acoustic waves in space plasma and its application to image encryption. IEEE Access 9:18762–18782

53. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B (2018) High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8798–8807
54. Wang B, Zheng H, Liang X, Chen Y, Lin L, Yang M (2018) Toward characteristic-preserving image-based virtual try-on network. In: Proceedings of the european conference on computer vision (ECCV), pp 589–604
55. Yang H, Zhang R, Guo X, Liu W, Zuo W, 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, pp 7850–7859
56. Yuan Y, Huh JH (2018) Customized CAD modeling and design of production process for one-person one-clothing mass production system. Electronics 7(11):270
57. Yuan Y, Huh JH (2019) Automatic pattern setting system reacting to customer design. J Inf Process Syst 15(6):1277–1295
58. Yuan M, Khan IR, Farbiz F, Yao S, Niswar A, Foo MH (2013) A mixed reality virtual clothes try-on system. IEEE Trans Multimed 15(8):1958–1968
59. Yuan Y, Park MJ, Huh JH (2021) A proposal for clothing size recommendation system using chinese online shopping malls: the new era of data. Appl Sci 11(23):11215
60. Zhang B, He M, Liao J, Sander PV, Yuan L, Bermak A, Chen D (2019) Deep exemplar-based video colorization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8052–8061
61. Zhang R, Isola P, Efros A, Shechtman E, Wang O, The unreasonable effectiveness of deep features as a perceptual metric (2018). In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 586–595
62. Zhang P, Zhang B, Chen D, Yuan L, Wen F (2020) Cross-domain correspondence learning for exemplar-based image translation
63. Zhu JY, Park T, Isola P, Efros A (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232
64. Zhu S, Urtasun R, Fidler S, Lin D, Change Loy C (2017) Be your own prada: fashion synthesis with structural coherence. In: Proceedings of the IEEE international conference on computer vision, pp 1680–1688
65. Zhu Z, Xu Z, You A, Bai X (2020) Semantically multi-modal image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5467–5476

Affiliations

Hajer Ghodhbani[1] · Mohamed Neji[1,2] · Abdulrahman M. Qahtani[3] · Omar Almutiry[4] · Habib Dhahri[4] · Adel M. Alimi[1,5]

Mohamed Neji
mohamed.neji@ieee.org

Abdulrahman M. Qahtani
amqahtani@tu.edu.sa

Omar Almutiry
oalmutiry@ksu.edu.sa

Habib Dhahri
hdhahri@ksu.edu.sa

Adel M. Alimi
adel.alimi@regim.usf.tn

1 REsearch Groups in Intelligent Machines (REGIM Lab), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia

2 National School of Electronics and Telecommunications of Sfax Technopark, BP 1163, CP 3018 Sfax, Tunisia

3 Department of Computer Science, College of Computers and Information Technology, Taif University, P.O.Box. 11099, Taif, 21944, Saudi Arabia

4 College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia

5 Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa