Computational Visual Media
https://doi.org/10.1007/s41095-021-0264-2 Vol. 8, No. 4, December 2022, 649–663

Research Article

High fidelity virtual try-on network via semantic adaptation and distributed componentization

Chenghu Du[1], Feng Yu[1,2] (✉), Minghua Jiang[1,2], Ailing Hua[1], Yaxin Zhao[1], Xiong Wei[1], Tao Peng[1,2], and Xinrong Hu[1,2]

Abstract Image-based virtual try-on systems have significant commercial value in online garment shopping. However, prior methods fail to appropriately handle details, so are defective in maintaining the original appearance of organizational items including arms, the neck, and in-shop garments. We propose a novel high fidelity virtual try-on network to generate realistic results. Specifically, a distributed pipeline is used for simultaneous generation of organizational items. First, the in-shop garment is warped using thin plate splines (TPS) to give a coarse shape reference, and then a corresponding target semantic map is generated, which can adaptively respond to the distribution of different items triggered by different garments. Second, organizational items are componentized separately using our novel semantic map-based image adjustment network (SMIAN) to avoid interference between body parts. Finally, all components are integrated to generate the overall result by SMIAN. A priori dual-modal information is incorporated in the tail layers of SMIAN to improve the convergence rate of the network. Experiments demonstrate that the proposed method can retain better details of condition information than current methods. Our method achieves convincing quantitative and qualitative results on existing benchmark datasets.

Keywords virtual try-on; conditional image synthesis; human parsing; thin plate spline; semantic adaptation

1 School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan 430200, China. E-mail: C. Du, duceh_lzy@163.com; F. Yu, yufeng@wtu.edu.cn (✉); M. Jiang, minghuajiang@wtu.edu.cn; A. Hua, hal_wtu@163.com; Y. Zhao, zyx@mail.wtu.edu.cn; X. Wei, wx_wh@wtu.edu.cn; T. Peng, pt@wtu.edu.cn; X. Hu, hxr@wtu.edu.cn.

2 Engineering Research Center of Hubei Province for Clothing Information, Wuhan 430200, China.

Manuscript received: 2021-09-14; accepted: 2021-11-03

1 Introduction

With the rapid development of the Internet apparel industry, more and more people shop for garments online. Traditional offline garment shopping allows assessment of fit through physical try-on. However, online garment shopping only permits a visual assessment by browsing models, which cannot give first-hand experience. Therefore, more and more researchers are trying to find effective online solutions. Existing virtual try-on strategies are either 2D image-based [1–5] or 3D model reconstruction-based [6–10] methods.

3D model reconstruction-based methods use computer graphics to reconstruct a 3D human model, which can make the result more plausible by controlling model joints. However, 3D model reconstruction-based methods need intensive computation and require a high degree of precision in model construction. They are unaffordable for general users.

Therefore, 2D image-based methods are a better choice for universal online garment try-on. On the one hand, image synthesis techniques can reduce calculation costs, so are suitable for customers without a high-performance processing device. On the other hand, image processing techniques based on deep learning can produce very realistic fitting results. If the shape of an in-shop garment is the same as a garment worn a person, the in-shop garment only needs to be deformed and joined in corresponding regions of the person's image. In other cases, however, arm lengths inevitably clash with sleeve lengths, causing problems such as textural confusion. Similarly, collar type of the result can be affected by the collar of the garment worn by a person (e.g., a V-neck is changed to a crew-neck). We need to find an effective approach to solve these problems.

Other issues must also be considered in the try-on process: (i) body invariant characteristics (e.g., head, pants) need to be preserved, (ii) embroidery and textures of the in-shop garment need to be transformed accurately, and (iii) the resulting image must be seamless, clear, and free from visible defects and noise.

Early researchers conducted pioneering studies. Zheng et al. [11] proposed an image-based garment changing system, which utilizes body factor extraction and content-aware image distortion, and determines joint positions by a neural network. The shape of the model is warped to the body's shape, and head swapping is performed to produce realistic virtual results. Neuberger et al. [12] proposed an outfit try-on approach (O-VITON), which accurately synthesizes the outfit on a body by an online optimization scheme. It has the ideal effect of fitting to the outfit, and in particular, it expands the try-on application to other parts such as pants. To save time and fit all kinds of garments one by one, virtual try-on network (VITON) [2] proposed a coarse-to-fine framework to transfer the in-shop garment to the corresponding area of the human body. First, VITON composites the result by coarsely fusing the in-shop garment to the corresponding part of a person. Then, it refines unclear areas of the coarse garment by a refinement network. In contrast to VITON, characteristic-preserving image-based virtual try-on network (CP-VTON) [4] trained a special geometric matching network [13] to be used with thin plate spline (TPS) [14] to warp the in-shop garment so as to retain rich details. Then it finely integrates the warped garment with the body by a try-on network. Later, CP-VTON+ [15] improved the warping effect of CP-VTON to give better results. Adaptive content generating and preserving network (ACGPN) [16] proposed an effective architecture, which solves the preservation of image details by generative adversarial networks (GAN) [17] and produces realistic fitting results with careful alignment of the garment. However, its results may be flawed due to failure to preserve collar type, sleeve shape, and arm details.

To overcome the challenges above, we propose a novel virtual try-on framework. Its key processes are as follows: (i) Geometric matching network (GMN) [13] coarsely warps the in-shop garment and uses the warped garment to generate target semantic map, then (ii) SMIAN generates the refined warped garment and body components (arms and neck) using the target semantic map, and (iii) SMIAN integrates the generated body components to obtain the final virtual try-on result.

The main contributions of the paper are:

- a novel image-based virtual try-on framework which effectively synthesizes the in-shop garment and the reference image by componentizing the garment and person,
- a novel component generating network, SMIAN, which generates and integrates high-quality body components, which avoids texture confusion by refined generation of body parts individually, and
- to prevent the garment covering the arms, and incorrect collar type and sleeve shape, an anti-covering map and neck semantic map are introduced, which effectively increases the authenticity of generated images.

The remainder of the paper is organized as follows: Section 2 presents related work. Section 3 describes the proposed virtual try-on framework in detail. Section 4 reports experimental results. Section 5 draws brief conclusions.

2 Related work

2.1 Conditional image synthesis

The contribution of GAN [17, 19] to fashion image processing is enormous. Conditional GAN (CGAN) [20–23] makes generated images controllable according to given conditions. Cui et al. [24] proposed an end-to-end virtual garment display method to render sketches and garment fabric. Jetchev et al. [1] proposed conditional analogy GAN (CA-GAN), which defines the virtual try-on task as an image analogy problem and adds a cyclic consistency loss function. However, it can only roughly transform properties and cannot adapt to geometric deformation in generating image details. Lee et al. [3] introduced the adversarial mechanism to the warping and try-on stages. It adds GAN loss to optimize the fit of the garment, making the generated image more plausible. Our method enhances the visual effect of the generated results based on GAN.

2.2 Human parsing and understanding

Estimation of human pose [25–27] and human semantic segmentation [28–30] are widely used in human-centered image research. Long et al. [31] first proposed a CNN-based method called fully convolutional networks

(FCN) for semantic segmentation. Gong et al. [28] proposed a new benchmark, Look into Person (LIP) providing a significant advance in terms of target diversity. Moreover, they studied self-supervised structure-sensitive learning for body parsing and body estimation. In the virtual try-on task, a human semantic map segments the human image to obtain the body parts needed for the experiment. The human pose can provide a warping guide for the shape of the in-shop garment and arms. Therefore, in our method, both are necessary data for generating virtual try-on results.

2.3 Virtual try-on

Current virtual try-on methods are divided into 2D image-based and 3D model reconstruction-based tasks. Mir et al. [7] proposed a method to transfer textures of garment images to 3D skinned multi-person linear (SMPL) model [32]; it is more accurate and faster than methods based on TPS. Zhao et al. [33] proposed a novel monocular-to-3D virtual try-on network, M3D-VTON, which generates a non-parametric 3D mesh model based on the generated 2D try-on result, creating a new virtual try-on mode. Yang et al. [16] proposed a content generation preservation network, ACGPN, which can adaptively determine which parts of a person's image should be preserved. It dramatically reduces artefacts and blurring in the generated results. However, it still does not entirely solve the excessive dependence of the generated result on the garment worn by a person, so the sleeve shape and collar type of the result do not match the in-shop garment. Cui et al. [34] proposed a flexible person generation framework, DiOr. It effectively performs the work of virtual try-on through a recurrent generation pipeline. Choi et al. [35] proposed a three-module framework using a high-resolution dataset and synthesizes the in-shop garment using an improved residual module. However, it is insufficient for good arm retention. Our method overcomes these challenges by body componentization.

In Table 1, we compare various state-of-the-art methods in terms of implementation and performance. Our method splits the virtual try-on problem into a multi-component problem, which has some superiority in virtual try-on tasks.

3 Methodology

3.1 Overview

In this section, we first illustrate how to estimate a semantic segmentation map from an in-shop garment, which is used to guide the generation of human components (see Section 3.2). Secondly, we explain the general structure and sub-modules of SMIAN, which is used to generate components and synthesize the result (see Section 3.3). Thirdly, we explain the generation strategy for each component and specific implementation details (see Section 3.4). Finally, we describe the training loss functions of GMN and SMIAN (see Section 3.5). The whole framework is shown in Fig. 1.

Given a reference image $I \in R^{3\times H\times W}$, an in-shop garment $c \in R^{3\times H\times W}$, a reference pose map $p_t \in R^{18\times H\times W}$, and a reference semantic map $s \in R^{1\times H\times W}$, the task of virtual try-on is to transfer the in-shop garment c to corresponding areas in the reference image I to generate an output image $\hat{I} \in R^{3\times H\times W}$, that is, the virtual try-on result. For this task, we propose a novel framework $\mathbb{T}$:

$$\hat{I} = \mathbb{T} < I, p_t, c, s > \tag{1}$$

3.2 Segmentation generation

Recently, *image-to-image translation* [36, 37] has been widely employed to generate desired images due to its

Table 1 Comparison of various state-of-the-art methods in terms of implementation and performance

	CA-GAN [1]	VITON [2]	CP-VTON [4]	VTNFP [18]	CP-VTON+ [15]	ACGPN [16]	Ours
Uses rough shape	×	√	√	√	√	×	×
Uses pose	×	√	√	√	√	√	√
Uses semantics	×	×	×	√	√	√	√
Body parts	×	×	×	√	√	√	√
Texture	×	√	√	√	√	√	√
Componentization	×	×	×	×	×	×	√
Occlusion handling	×	×	×	×	×	×	√
Garment over-warping	×	×	×	×	√	√	√
Character retention	×	√	√	√	√	√	√
Collar shape	×	×	×	×	√	×	√

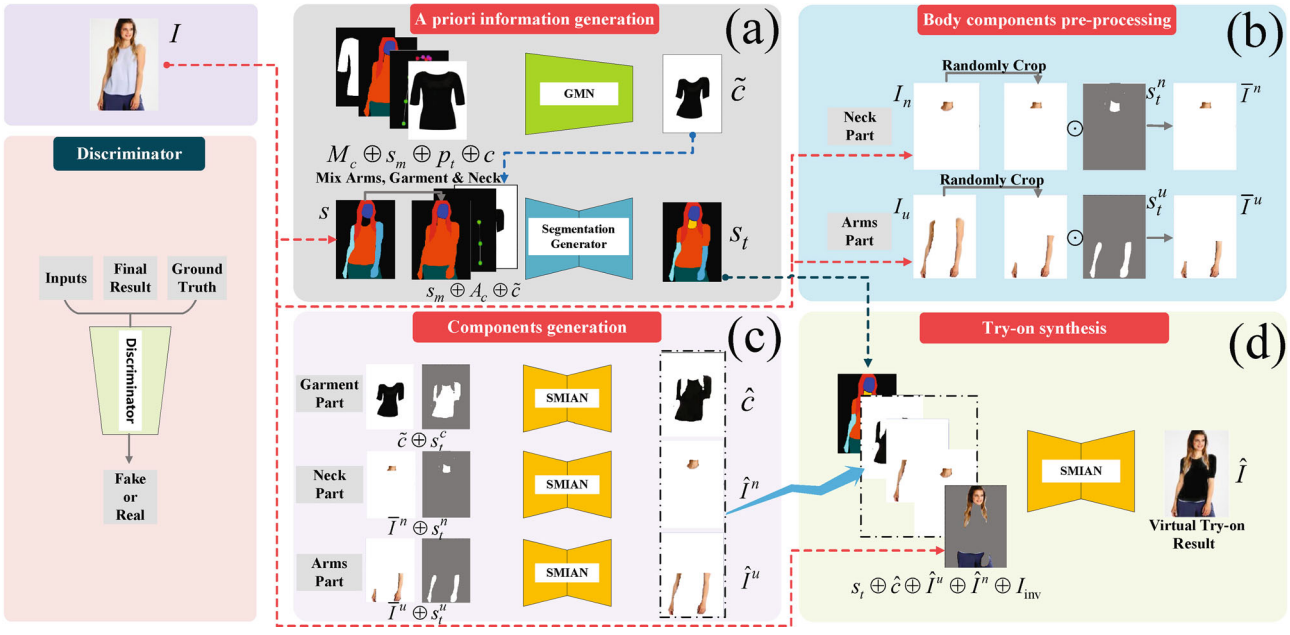

Fig. 1 Framework; the execution proceeds from (a) to (d). Module (a) coarsely warps in-shop garment c through GMN, and then predicts the target semantic map s_t using warped garment $\tilde{c}$ and the mixed semantic map s_m. Module (b) pre-processes arms and neck for generation of the corresponding components in module (c) by SMIAN to give the garment $\hat{c}$, with arms $\hat{I}^u$ and neck $\hat{I}^n$. Module (d) combines the components by SMIAN to give the final try-on result $\hat{I}$. The discriminator belongs to SMIAN. $\oplus$ denotes channel-wise concatenation and $\odot$ denotes an AND operation.

remarkable effectiveness. Inspired by this approach, we first need to generate a semantic segmentation map of a person to guide subsequent image generation.

Generating a segmentation aims to produce a target semantic map $s_t \in R^{20\times H\times W}$ containing the shape of the in-shop garment c. Specifically, in the virtual try-on task, the semantic map remains unchanged except for the garment, neck, and arm areas. Therefore, we combine the garment, neck, and arms in the reference semantic map s into one region in a reallocated semantic map s_m. This enables other parts of s_m to be used as boundary conditions for warping: the combined area is the range of warping. In the target semantic map s_t, s_m is reallocated according to the shape of the warped garment $\tilde{c}$.

However, the human pose is flexible, and there are many poses in which the arms and garment overlap. During training with these poses, it is difficult for the network to learn which part represents the arms in the semantic map, causes the arms' semantic map to be occupied by the garment's semantic map. To enable the network to easily locate the arms in general, we create an anti-cover map $A_c \in R^{1\times H\times W}$ (connecting the shoulder, elbow, and wrist with a one-pixel wide short line) as input conditioning information to highlight the arms' positions for more accurate segmentation.

As shown in Fig. 1(a)(below), we adopt U-Net [38] as the generation network. Furthermore, s_m, A_c, and the coarsely warped in-shop garment $\tilde{c}$ (see later) are used as inputs, and the weighted cross-entropy loss $\mathcal{L}_s$ [28] is used to optimize the network.

3.3 Semantic map-based image adjustment network (SMIAN)

3.3.1 Basics

We propose a novel *semantic map-based image adjustment network* (SMIAN), which aggregates dual-modality features from the source image to reconstruct body components through the semantic map. As Fig. 2 shows, SMIAN consists of three modules: a parsing module, a content module, and an integration module. Unlike the generator in Ref. [39], which adds several style blocks between the encoder and decoder, the decoder of SMIAN consists of several AdaIN ResNet blocks (ARBs).

3.3.2 Parsing module and content module

The content and parsing modules have the same structure, and both consist of five down-sampled convolutional layers. However, their roles are different. The parsing module is used to obtain the feature map of the varying semantic map of components for

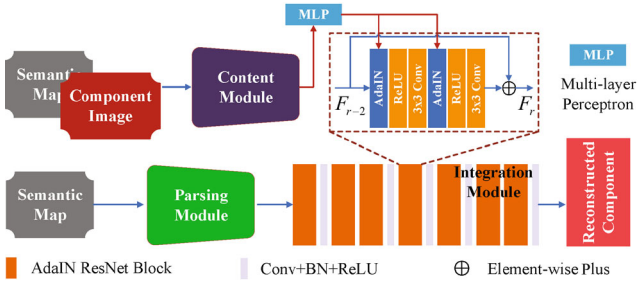

Fig. 2 Overview of SMIAN structure.

input to the integration module. The content module is used to obtain content information (e.g., colour, texture, embroidery) of the components and spatial information from the semantic map. It provides the basis for the ARB in the integration module.

3.3.3 Integration module

The integration module analyses the feature map extracted by the parsing module to reconstruct component. It consists of seven ARBs and five up-sampled convolutional layers. StyleGAN [19] successfully applied adaptive instance normalization (AdaIN) to the progressive generative model, which distributes the features to the latent variables. Inspired by this, we introduce several ARBs to finely reconstruct body components, using AdaIN to restructure the spatial distribution information of the semantic map and the content information of the component image. The AdaIN calculation is

$$\mathrm{AdaIN}(x,y)=\gamma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\mu(y) \tag{2}$$

where x is the feature map produced by the previous ARB or the previous convolution layer, and y is the feature map produced by the multi-layer perceptron (MLP). σ and μ denote mean and standard deviation, respectively. The formula adjusts the mean and standard deviation of the semantic map to the component image.

We define the input feature map of the ARB as F_{r-2} and F_m (from the MLP). Note that the input signals from the content module are all the same in the ARB. As shown in Fig. 2, feature processing operations in the overall ARB can be defined as

$$F_{r-1}^{\mathrm{mid}}=\phi_{r-2}\sigma_{r-2}\mathrm{AdaIN}(F_{r-2},F_m) \tag{3}$$

$$F_r^{*}=\phi_{r-1}\sigma_{r-1}\mathrm{AdaIN}(F_{r-1}^{\mathrm{mid}},F_m) \tag{4}$$

where $r=3,\ldots,R$, and R is the number of execution steps. ϕ_{r-1} denotes the convolution operation in step $r-1$, and σ_{r-1} denotes the Relu activation function in step $r-1$. Following the standard ResNet block [40], a skip connection structure was added to fuse the input and output feature maps using:

$$F_r=F_{r-2}+F_r^{*} \tag{5}$$

where F_r is the output of the ARB.

3.4 Body component generation

We now describe how we create the human's arms, garment, and neck as body components by SMIAN to minimize mutual interference in the resulting image distribution.

3.4.1 Garment component (GC)

As shown in Fig. 1(a)(above), the in-shop garment mask M_c (see Fig. 3, additional shape information), the reference pose map p_t (see Fig. 3), and the reallocation semantic map s_m are used as the inputs to the GMN [13], to warp the in-shop garment to generate the coarsely warped garment $\tilde{c}$.

Because there are only a few controlled parameters θ in GMN that manipulate the warping of the garment, an error in one of the parameters θ can lead to over-warping (unnatural excessive partial distortion of the garment) or under-warping (garment not fully aligned with the garment semantic map).

To overcome over-warping, we introduce a sampling interval consistency loss $\mathcal{L}_{\mathrm{sic}}$ [3, 15] into the GMN to limit the spacing between sampling points; it is given by

$$\begin{aligned}\mathcal{L}_{\mathrm{sic}}(\hat{G}_x,\hat{G}_y)=&\sum_{i=-1,1}\sum_x\sum_y|\hat{G}_x(x+i,y)-\hat{G}_x(x,y)|\\&+\sum_{j=-1,1}\sum_x\sum_y|\hat{G}_y(x,y+j)-\hat{G}_y(x,y)|\end{aligned} \tag{6}$$

where $\hat{G}_x$ and $\hat{G}_y$ are the x and y coordinates of the sampled grid, respectively, and the absolute difference $|a-b|$ measures the distance between two adjacent nodes a and b. $\mathcal{L}_{\mathrm{sic}}$ totals the distances of all points to adjacent points in the sampled grid.

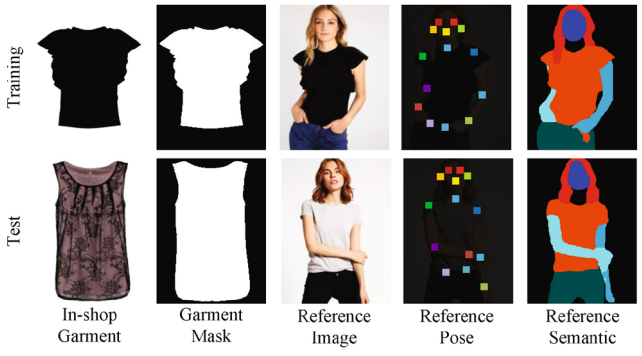

Fig. 3 Training and test set samples from VITON.

As shown in the garment part of Fig. 1(c), to overcome under-warping, the predicted garment semantic map s_t^c and the warped garment $\tilde{c}$ are processed by SMIAN to repair missing areas and remove redundant areas. In this way, the garment component $\hat{c}$ is obtained via

$$\hat{c} = \text{SMIAN} < \text{concat}(\tilde{c}, s_t^c), s_t^c > \tag{7}$$

where concat$(\cdot)$ denotes channel-wise concatenation.

3.4.2 Arms component (AC)

Different in-shop garments have different sleeve shapes. Therefore, the arms I_u in the reference image cannot be directly reused as input. The simplest way to handle this is as follows: (i) reusing the correct arms directly, (ii) repairing missing areas in the arms using the generator, and (iii) removing unnecessary areas in the arms using the generator.

However, as Fig. 3 shows, while the arms and in-shop garment correspond in the training set, in practice, they do not always correspond due to the different garment shape. As shown in Fig. 4 and the arms part of Fig. 1(b), to overcome this problem, during training, we randomly crop the arms I_u to provide input to enable the generator to learn an inpainting capability. Furthermore, we perform an AND operation between the randomly cropped I_u and the arms semantic map s_t^u to remove possible background.

During testing, only the AND operation is performed between s_t^u and the arms I_u to remove unnecessary areas of arms I_u. This maximally retains $\bar{I}^u$ of the original arms, which can be expressed as

$$\bar{I}^u = \begin{cases} \text{rand}(I_u) \odot s_t^u, & \text{training} \\ I_u \odot s_t^u, & \text{testing} \end{cases} \tag{8}$$

where $\odot$ denotes an AND operation, and rand$(\cdot)$ denotes random cropping.

Finally, the arms component $\hat{I}^u$ is obtained by SMIAN. It can be formulated as

$$\hat{I}^u = \text{SMIAN} < \text{concat}(s_t^u, \bar{I}^u), s_t^u > \tag{9}$$

3.4.3 Neck component (NC)

As Fig. 3 shows, the reference semantic map s of the VITON dataset does not contain a neck semantic map, so the neck's shape is unchanged before and after try-on. Therefore, we add a neck semantic map to s to remove limitations due to neck shape.

The neck component is handled similarly to the arms component. As shown in Fig. 5 and the neck part of Fig. 1(b), during training, the neck I_n in the reference image is randomly cropped to learn an inpainting capability. During testing, only the AND operation is performed between s_t^n and the randomly cropped neck to remove unnecessary areas of the neck I_n. This maximally retains $\bar{I}^n$ of the original neck, which can be expressed as

$$\bar{I}^n = \begin{cases} \text{rand}(I_n) \odot s_t^n, & \text{training} \\ I_n \odot s_t^n, & \text{testing} \end{cases} \tag{10}$$

In this way, the collar shape of the in-shop garment determines the neck shape in the result. Finally, the neck component $\hat{I}^n$ is obtained by SMIAN. It can be formulated as

$$\hat{I}^n = \text{SMIAN} < \text{concat}(s_t^n, \bar{I}^n), s_t^n > \tag{11}$$

3.4.4 Component synthesizer (CS)

All the components ($\hat{c}, \hat{I}^n$, and $\hat{I}^u$) and unchanged parts I_{inv} in the reference image I are spliced to generate the try-on result $\hat{I}$. However, in the actual splicing operation, cracks occur between boundaries

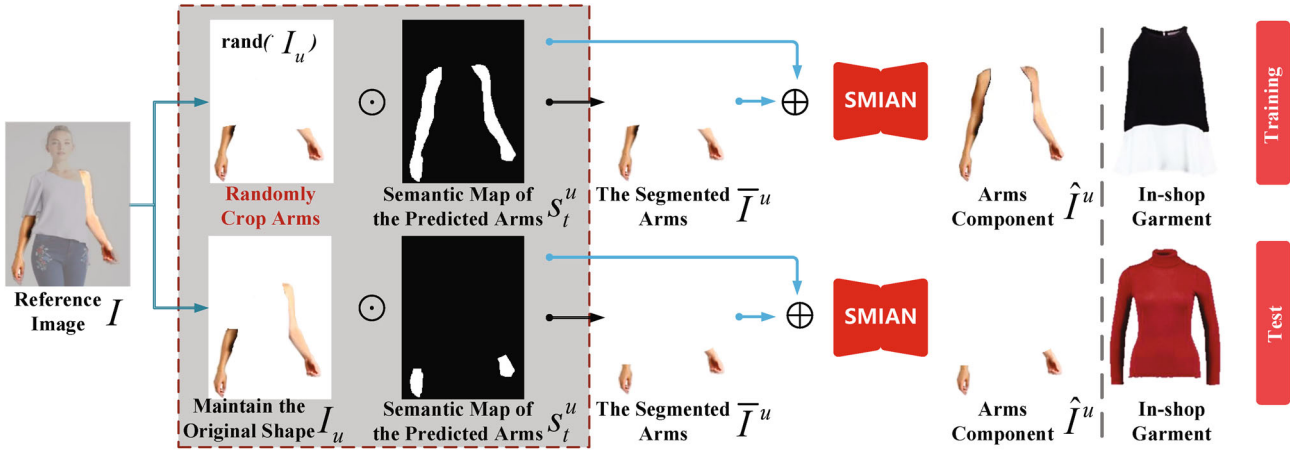

Fig. 4 Overview of the arms component generation process. $\oplus$ denotes channel-wise concatenation and $\odot$ denotes an AND operation.

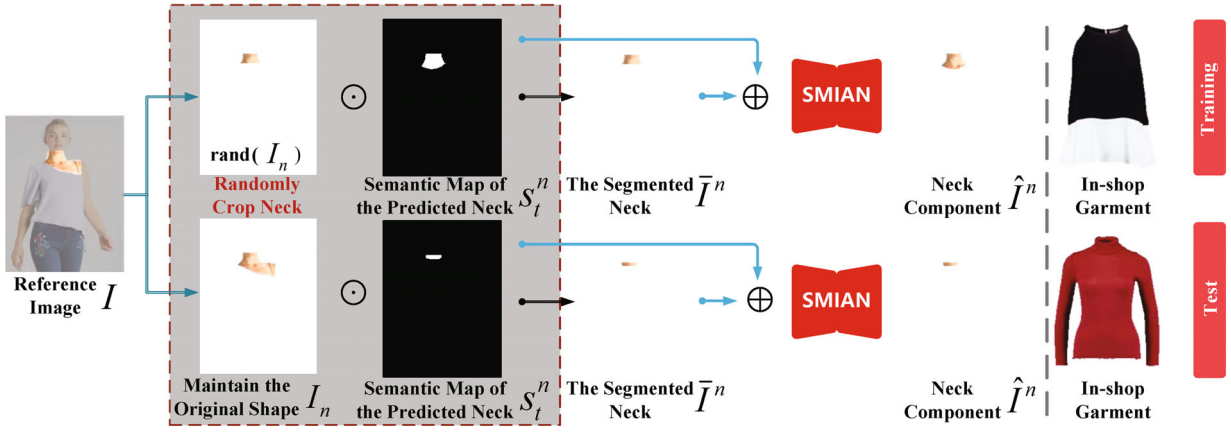

Fig. 5 Overview of neck component generation. $\oplus$ denotes channel-wise concatenation and $\odot$ denotes an AND operation.

of the components due to edge errors in the semantic map. As Fig. 1(d) shows, we use SMIAN to repair the cracks with guidance from the target semantic map s_t. This allows a natural transition between components to be realised, and noise in the image is also reduced. This can be described as

$$\hat{I} = \text{SMIAN} < \text{concat}(s_t, \hat{c}, \hat{I}^u, \hat{I}^n, I_{\text{inv}}), s_t > \quad (12)$$

3.5 Training

During training, pixel-wise $\mathcal{L}_1$, perceptual loss $\mathcal{L}_{\text{per}}$ [41], and $\mathcal{L}_{\text{sic}}$ are used to optimize the GMN, which can be expressed as

$$\mathcal{L}_{\text{cw}} = \lambda_1 \mathcal{L}_1 + \lambda_{\text{per}} \mathcal{L}_{\text{per}} + \lambda_{\text{sic}} \mathcal{L}_{\text{sic}} \quad (13)$$

where λ are trade-off hyper-parameters for the corresponding loss functions.

The SMIAN in the framework need to be trained separately. The total loss $\mathcal{L}_{\text{SMIAN}}$ consists of $\mathcal{L}_1$, $\mathcal{L}_{\text{per}}$, $\mathcal{L}_{\text{adv}}$ [37], and $\mathcal{L}_{\text{fm}}$ [37]. It can be expressed as

$$\mathcal{L}_{\text{SMIAN}} = \lambda_1 \mathcal{L}_1 + \lambda_{\text{per}} \mathcal{L}_{\text{per}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{fm}} \mathcal{L}_{\text{fm}} \quad (14)$$

4 Experiments and analysis

In this section, we first introduce the experimental dataset, VITON [2], and describe the implementation details of the experiment. We then verify the execution performance of SMIAN, and qualitatively and quantitatively compare our results with those of other state-of-the-art networks. We also conduct ablation experiments to demonstrate the effectiveness of each submodule in the framework. Finally, we conduct a user study to demonstrate the practicality of the proposed method.

4.1 Dataset

The experimental dataset is from the VITON [2] dataset. It contains 16,253 image groups. Each group consists of a front-view female image I, an in-shop garment image c and its mask M_c, a reference semantic map s, and a reference pose p_t. The size of each image is 256 pixels × 192 pixels. The dataset contains 14,221 groups in the training set and 2032 groups in the test set. Figure 3 shows a sample from the training set and another from the test set in VITON. It can be seen that the in-shop garment and the garment worn on paper are the same in the training set, while in the test set they are different.

4.2 Implementation details

Our experiments were carried out on 2 Tesla V100 GPUs with 32 G RAM. By default, the learning rate for the generator and the discriminator were 0.0001, reduced linearly to 0 over half of the epochs with a batch size of 4. The experiment adopted the ADAM optimizer [42], with parameters set to $\beta_1 = 0.5, \beta_2 = 0.999$. In the loss function, $\lambda_1 = \lambda_{\text{per}} = 1, \lambda_{\text{sic}} = 40$ in $\mathcal{L}_{\text{cw}}$. $\lambda_1 = 1$, $\lambda_{\text{per}} = 10, \lambda_{\text{adv}} = 1$, and $\lambda_{\text{fm}} = 10$ in $\mathcal{L}_{\text{SMIAN}}$.

The discriminator in SMIAN is the one from pix2pixHD [37], as shown in the discriminator in Fig. 1.

4.3 Performance

To assess the execution performance of SMIAN, we compared the convergence rate during training and differences of test results, using $\mathcal{L}_1$, $\mathcal{L}_{\text{per}}$, $\mathcal{L}_{\text{adv}}$, $\mathcal{L}_{\text{fm}}$, and $\mathcal{L}_{\text{SMIAN}}$ (vertical axis) as an indirect representation to compare performance between SMIAN and the U-Net (used in CP-VTON, CP-VTON+, and ACGPN).

In Fig. 6, to show the effect clearly, we test the result once every four images (horizontal axis). The total loss $\mathcal{L}_{\text{SMIAN}}$ of the proposed method stands

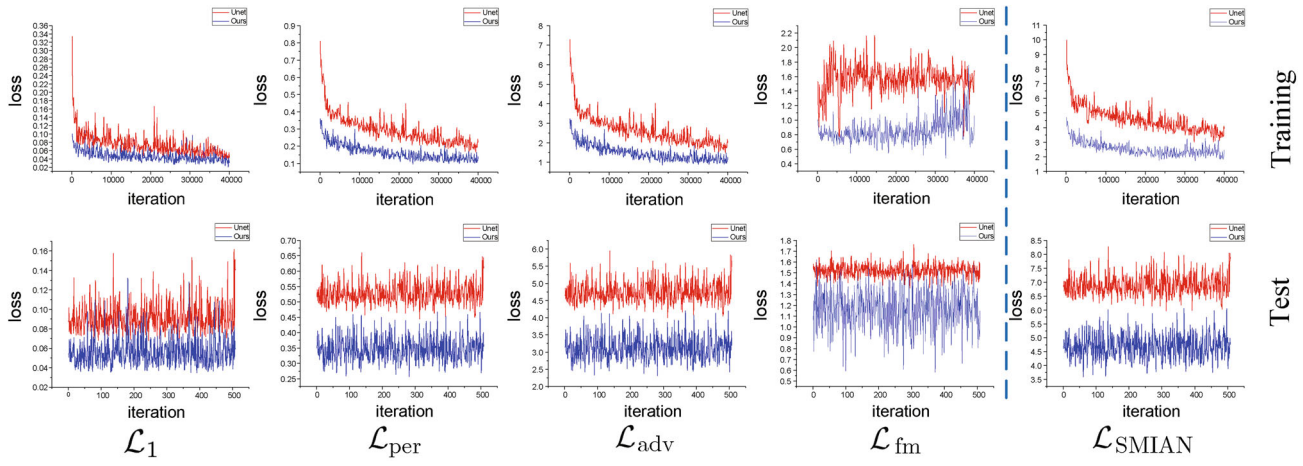

Fig. 6 Losses in SMIAN and U-Net during training and testing.

at about 2.0 after iterating, while that for U-Net remains at about 3.5. The convergence rate of the other sub-losses is also better than for U-Net. During testing, the proposed method also has lower total loss: $\mathcal{L}_{\text{SMIAN}}$ of each image from the proposed method stands at about 4.65, and for U-Net, about 6.75. In the other sub-losses, the loss differences in each batch of images show a remarkable distance. The pixel-wise $\mathcal{L}_1$ loss represents a slight difference between the image generated by SMIAN and ground truth. $\mathcal{L}_{\text{per}}$, $\mathcal{L}_{\text{adv}}$, and $\mathcal{L}_{\text{fm}}$ indicate more realistic images are generated by SMIAN.

The above results demonstrate that our method has significantly improved network performance and better generated image quality.

4.4 Qualitative results

We next provide qualitative results of the proposed method and compare visual outputs with those from three state-of-the-art models, CP-VTON [4], CP-VTON+ [15], and ACGPN [16].

4.4.1 Semantic map correctness

Figure 7(above) shows the effect after adding the neck semantic map. Column 5 (for CP-VTON) is the effect without the neck semantic map, where the reference image limits collar type. The neck semantic map in column 6 (for ACGPN) is in error. The last column is the effect of the proposed method, where the collar type dependence from the garment worn on the person is eliminated. The collar type, which changes to the shape of the in-shop garment, has a more natural appearance in the result.

Fig. 7(below) shows the effect after adding the anti-cover map A_c (column 3). Column 5 (for CP-VTON) is the effect without the anti-cover map; the garment covers the arms. The arms semantic map in column 6 (for ACGPN) is in error. The last column shows

Fig. 7 Influence of semantic map correctness. Column 3 is the anti-cover map A_c, while column 7 shows parts in the target semantic map s_t.

the arms part is generated clearly to prove that the anti-cover map works as intended.

4.4.2 Garment alignment

Figure 8 compares our method with state-of-the-art methods in terms of garment alignment, using a grid image (column 2) as a visual representation of the degree of warping. CP-VTON (column 4) exhibits over-warping because it does not impose constraints. In addition, it uses the rough shape as condition information causing the absence of warp boundaries. Although CP-VTON+ (column 6) incorporates constraints, its input contains a rough shape, and the network lacks perceptual loss as an optimisation function, leading to local over-warping. ACGPN (column 8) excessively limits the spacing by second-order-difference constraints, which results in an almost uniform degree of warping throughout the garment. It enhances the sleeve area in the garment, which as a result is extremely unnatural. Specifically, we show the result without $\mathcal{L}_{\text{sic}}$ in column 10, where the garment shows over-warping. In contrast, our proposed method shows a more natural warping effect without over-warping by using $\mathcal{L}_{\text{sic}}$, s_m, and perceptual loss in column 12.

4.4.3 Comparison of try-on results

As Fig. 9 shows, CP-VTON, CP-VTON+, and ACGPN have defective garment alignment which affects the final result. The way the garment and body are integrated in CP-VTON leads to problems such as texture confusion. Although CP-VTON+ works well on restoration of collar type, the lack of a strategy for retaining original details causes problems such as occlusion and loss of detail. There are many errors in the ACGPN semantic map, which produce unsatisfactory results.

Our method uses interval consistency loss and perceptual loss to overcome over-warping, making the results more realistic. Under-warping and mismatches of sleeve shape and collar type between the result and in-shop garment are avoided by semantic adaptive componentization. Finally, our method preserves the most original content in the reference image and in-shop garment. Compared to CP-VTON, CP-VTON+, and ACGPN, our method works better.

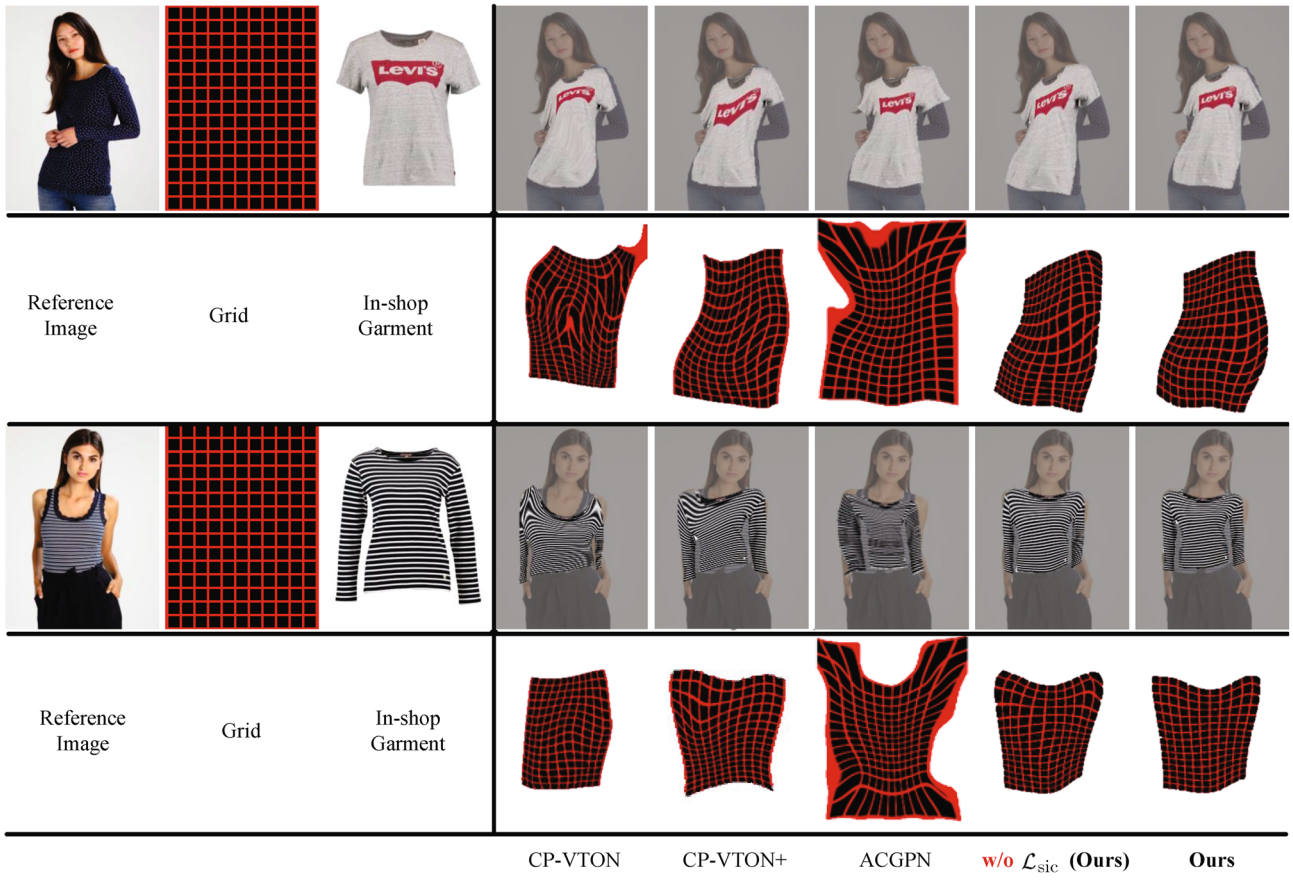

Fig. 8 Garment warping effects from state-of-the-art methods and the proposed method.

Fig. 9 Qualitative comparison of the baselines on VITON, the dashed box highlighting improvements from our proposed method.

4.4.4 Coupling study

The application goal of the virtual try-on task is to try on the desired garment online, so we chose four different garments for a coupling study. As shown in Fig. 10, the proposed method works well.

4.5 Quantitative results

We further analyzed performance using benchmark metrics for image quality, adopting *structural similarity* (SSIM) [43] defined in Eq. (15) and *Fréchet inception distance* (FID) [44] defined in Eq. (16) to measure the similarity between the try-on result and ground truth. The *inception score* (IS) [45] defined in Eq. (17) and *peak signal to noise ratio* (PSNR) defined in Eq. (18) are adopted to measure the image quality between the try-on result and ground truth.

Fig. 10 Coupling performance in multi garment try-on.

$$\mathrm{SSIM} = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (15)$$

where μ_X is the mean of X, μ_Y is the mean of Y, σ_X^2 is the variance of X, σ_Y^2 is the variance of Y, σ_{XY} is the covariance of X and Y, and C_1 and C_2 are two variables to ensure stability.

$$\mathrm{FID} = \| \mu_r - \mu_g \|^2 + T_r[\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}] \quad (16)$$

where T_r denotes matrix trace , μ is the mean, and Σ is covariance.

$$\mathrm{IS}(G) = \exp(\mathbb{E}_{x\sim p_g} D_{KL}(p(y \mid x)) \parallel p(y))) \quad (17)$$

where $p(y \mid x)$ is a particular classification obtained from the generated data x, $p(y)$ is the edge distribution of the obtained classification, and D_{KL} denotes relative entropy.

$$\text{PSNR} = 10\log_{10}\frac{255^2}{\varepsilon} \tag{18}$$

where ε denotes mean square error (MSE) between the ground truth and generated image.

Table 2 summarizes the performance of state-of-the-art methods. Our method improves IS from 2.85 (for DCTON [47]) to 2.86, which reflects the fact that the distribution of the generated image is closer to the ground truth distribution. Original detail retention strategy for arms and garment component improves SSIM from 0.83 (for DCTON) to 0.87. Also, PSNR increased by around 10% from 23.067 (for ACGPN) to 25.423. The component synthesizer contributes to defect-free fusion between components, reducing noise and improving FID (for DCTON) from 14.82 to 12.63. The experimental data shows that our method provides convincing virtual try-on results.

4.6 Discussion and ablation study

We performed an ablation study to verify the utility of each part of our framework. The ablation study considered in turn: removing the ARB, removing the AC, removing the GC, and removing the CS. We observe from the ablation experiment results in Fig. 11 that: (i) the ARB preserves rich component details by correlating the spatial distribution between component and semantic map, (ii) the AC preserves arm details from the reference image very well, with fingers and arms distinguished, (iii) the GC effectively prevents any mismatch between arm length and sleeve length of the in-shop garment, and details of the garment are well preserved, and (iv) the CS fuses all components through semantic map guidance, gaps between components are repaired, and noise is reduced, making the results more natural and realistic.

Table 3 provides the same metrics as before for these four cases, the data in the table once again confirming the points above.

Table 2 Quantitative evaluation of different methods

Method	IS↑	SSIM↑	PSNR↑	FID↓
CA-GAN [1]	2.56±0.09	0.74	—	47.34
VITON [2]	2.29±0.07	0.74	—	55.71
CP-VTON [4]	2.59±0.13	0.72	16.956	24.45
CP-VTON+ [15]	2.75±0.14	0.75	16.956	21.08
SieveNet [46]	2.82±0.09	0.77	16.98	14.65
VTNFP [18]	2.78±0.10	0.80	—	—
ACGPN [16]	2.69±0.12	0.81	23.067	15.67
DCTON [47]	2.85±0.15	0.83	—	14.82
Ours	**2.86±0.07**	**0.87**	**25.423**	**12.63**
Real	2.88±0.12	1	N/A	0

4.7 User study

To further evaluate the effectiveness of our approach, we designed a user study using a questionnaire. First, the results obtained from three virtual try-on methods on the test set were mixed. Then, we invited two volunteers from fashion design and computer vision to score all test images (in the range 0 to 1). Finally, we

Table 3 Ablation study results

Method	IS↑	SSIM↑	PSNR↑	FID↓
w/o ARB	2.78±0.11	0.83	23.812	25.78
w/o AC	2.84±0.11	0.86	23.845	15.46
w/o GC	2.82±0.12	0.84	23.649	18.56
w/o CS	**2.86±0.13**	0.87	23.963	14.85
Ours	2.86±0.07	**0.87**	**25.423**	**12.63**
Real	2.88±0.12	1	N/A	0

Fig. 11 Ablation study: visual results obtained by different methods, dashed boxes indicating areas with poorer results.

calculated the average score as the satisfaction for each image and plotted the scores as a statistical graph.

The results in Fig. 12 indicate that the image quality obtained by the proposed method provide a better sensory experience than CP-VTON and ACGPN, intuitively indicating that our method is a superior method for virtual try-on field.

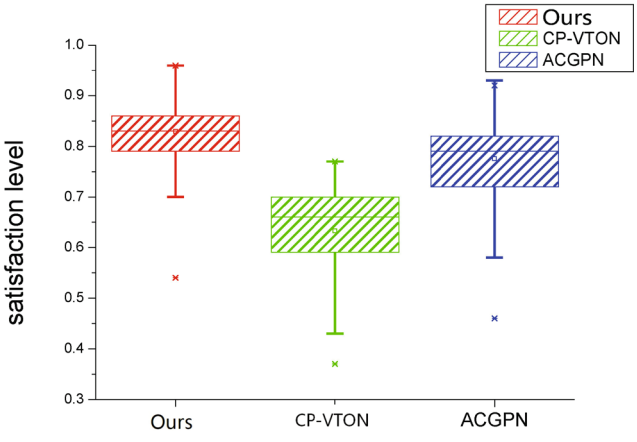

Fig. 12 User satisfaction results for three methods.

5 Conclusions

In this work, we have proposed a novel virtual try-on framework. To preserve details of the human body and garment, SMIAN is proposed to accelerate network convergence rate and optimize the generation effect. It improves the performance of the virtual try-on framework. Moreover, the body parts to be synthesized are componentized for local-to-global generation, solving existing problems such as occlusion and loss of detail. The componentization of the body area also reduces coupling in the result, which helps the network pay more attention to local details. Compared to the state-of-the-art works, our pipeline provides quantitatively better results and visual effects. User satisfaction is increased to 83.5%. In future, we plan to expand our framework to deal with image-based pose transfer with complex appearance-aware information.

Acknowledgements

This manuscript is an extended version of our previous work which appeared at the IEEE International Conference on Tools with Artificial Intelligence (C. Du et al. VTON-HF: High fidelity virtual try-on network via semantic adaptation. ICTAI 2021, 224–231, doi: 10.1109/ICTAI52525.2021.00038). We declare that we submit this manuscript to *Computational Visual Media* with permission.

We would like to thank the anonymous reviewers for their constructive comments. The findings and observations in this paper are those of the authors and do not necessarily reflect the views of the supporters.

Funding

This work was supported by Young Talents Programme of Scientific Research Program of Hubei Education Department (Project No. Q20201709), Research on the Key Technology of Flexible Intelligent Manufacturing of Clothing based on Digital Twin of Hubei Key Research and Development Program (Project No. 2021BAA042), and Open Topic of Engineering Research Center of Hubei Province for Clothing Information (Project No. 900204).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Jetchev, N.; Bergmann, U. The conditional analogy GAN: Swapping fashion articles on people images. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2287–2292, 2017.

[2] Han, X. T.; Wu, Z. X.; Wu, Z.; Yu, R. C.; Davis, L. S. VITON: An image-based virtual try-on network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7543–7552, 2018.

[3] Lee, H. J.; Lee, R.; Kang, M.; Cho, M.; Park, G. LA-VITON: A network for looking-attractive virtual try-on. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop, 3129–3132, 2019.

[4] Wang, B.; Zheng, H.; Liang, X.; Chen, Y.; Lin, L.; Yang, M. Toward characteristic-preserving image-based virtual try-on network. In: *Computer Vision – ECCV 2018. Lecture Notes in Computer Science, Vol. 11217*. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Eds. Springer Cham, 607–623, 2018.

[5] Han, X. T.; Huang, W. L.; Hu, X. J.; Scott, M. ClothFlow: A flow-based model for clothed person generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 10470–10479, 2019.

[6] Ma, Q. L.; Yang, J. L.; Ranjan, A.; Pujades, S.; Pons-Moll, G.; Tang, S. Y.; Black, M. J. Learning to dress 3D people in generative clothing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6468–6477, 2020.

[7] Mir, A.; Alldieck, T.; Pons-Moll, G. Learning to transfer texture from clothing images to 3D humans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7021–7032, 2020.

[8] Zhu, H. M.; Cao, Y.; Jin, H.; Chen, W. K.; Du, D.; Wang, Z. Y.; Cui, S.; Han, X. Deep Fashion3D: A dataset and benchmark for 3D garment reconstruction from single images. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science, Vol. 12346*. Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 512–530, 2020.

[9] Lähner, Z.; Cremers, D.; Tung, T. DeepWrinkles: Accurate and realistic clothing modeling. In: *Computer Vision – ECCV 2018. Lecture Notes in Computer Science, Vol. 11208*. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Eds. Springer Cham, 698–715, 2018.

[10] Liang, J. B.; Lin, M. C. Machine learning for digital try-on: Challenges and progress. *Computational Visual Media* Vol. 7, No. 2, 159–167, 2021.

[11] Zheng, Z. H.; Zhang, H. T.; Zhang, F. L.; Mu, T. J. Image-based clothes changing system. *Computational Visual Media* Vol. 3, No. 4, 337–347, 2017.

[12] Neuberger, A.; Borenstein, E.; Hilleli, B.; Oks, E.; Alpert, S. Image based virtual try-on network from unpaired data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5183–5192, 2020.

[13] Rocco, I.; Arandjelovi? R.; Sivic, J. Convolutional neural network architecture for geometric matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 41, No. 11, 2553–2567, 2019.

[14] Duchon, J. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: *Constructive Theory of Functions of Several Variables. Lecture Notes in Mathematics, Vol. 571*. Schempp, W.; Zeller, K. Eds. Springer Berlin Heidelberg, 85–100, 1977.

[15] Minar, M. R.; Tuan, T. T.; Ahn, H.; Rosin, P.; Lai. Y.-K. CP-VTON+: Clothing shape and texture preserving image-based virtual try-on. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[16] Yang, H.; Zhang, R. M.; Guo, X. B.; Liu, W.; Zuo, W. M.; Luo, P. 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, 7847–7856, 2020.

[17] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Communications of the ACM* Vol. 63, No. 11, 139–144, 2020.

[18] Yu, R. Y.; Wang, X. Q.; Xie, X. H. VTNFP: An image-based virtual try-on network with body and clothing feature preservation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 10510–10519, 2019.

[19] Karras, T.; Laine, S.; Aila, T. M. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4396–4405, 2019.

[20] Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv preprint* arXiv:1411.1784, 2014.

[21] Jo, Y.; Park, J. SC-FEGAN: Face editing generative adversarial network with user's sketch and color. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 1745–1753, 2019.

[22] Choi, Y.; Choi, M.; Kim, M.; Ha, J. W.; Kim, S.; Choo, J. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8789–8797, 2018.

[23] Honda, S. VITON-GAN: Virtual try-on image generator trained with adversarial loss. In: Proceedings of the Eurographics 2019 - Posters, 2019.

[24] Cui, Y. R.; Liu, Q.; Gao, C. Y.; Su, Z. FashionGAN: Display your fashion design using Conditional Generative Adversarial Nets. *Computer Graphics Forum* Vol. 37, No. 7, 109–119, 2018.

[25] Zhang, F.; Zhu, X. T.; Dai, H. B.; Ye, M.; Zhu, C. Distribution-aware coordinate representation for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7091–7100, 2020.

[26] Cheng, B. W.; Xiao, B.; Wang, J. D.; Shi, H. H.; Huang, T. S.; Zhang, L. HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5385–5394, 2020.

[27] Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S. H.; Sheikh, Y. OpenPose: Realtime multi-person 2D pose estimation

using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 43, No. 1, 172–186, 2021.

[28] Gong, K.; Liang, X. D.; Zhang, D. Y.; Shen, X. H.; Lin, L. 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, 6757–6765, 2017.

[29] Wang, W.; Yu, K. C.; Hugonot, J.; Fua, P.; Salzmann, M. Recurrent U-net for resource-constrained segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2142–2151, 2019.

[30] Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. CCNet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 603–612, 2019.

[31] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440, 2015.

[32] Osman, A. A. A.; Bolkart, T.; Black, M. J. STAR: Sparse trained articulated human body regressor. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science, Vol. 12351*. Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 598–613, 2020.

[33] Zhao, F. W.; Xie, Z. Y.; Kampffmeyer, M.; Dong, H. Y.; Han, S. F.; Zheng, T. X.; Zhang, T.; Liang, X. M3D-VTON: A monocular-to-3D virtual try-on network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 13219–13229, 2021.

[34] Cui, A.; McKee, D.; Lazebnik, S. 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, 14638–14647, 2021.

[35] Choi, S.; Park, S.; Lee, M.; Choo, J. VITON-HD: High-resolution virtual try-on via misalignment-aware normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14126–14135, 2021.

[36] Isola, P.; Zhu, J. Y.; Zhou, T. H.; Efros, A. A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5967–5976, 2017.

[37] Wang, T. C.; Liu, M. Y.; Zhu, J. Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8798–8807, 2018.

[38] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9351*. Navab, N.; Hornegger, J.; Wells, W.; Frangi, A. Eds. Springer Cham, 234–241, 2015.

[39] Men, Y. F.; Mao, Y. M.; Jiang, Y. N.; Ma, W. Y.; Lian, Z. H. Controllable person image synthesis with attribute-decomposed GAN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5083–5092, 2020.

[40] He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778, 2016.

[41] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556, 2014.

[42] Kingma, D. P.; Ba, J. L. Adam: A method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations, 2015.

[43] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* Vol. 13, No. 4, 600–612, 2004.

[44] Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, 6629–6640, 2017.

[45] Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2234–2242, 2016.

[46] Jandial, S.; Chopra, A.; Ayush, K.; Hemani, M.; Kumar, A.; Krishnamurthy, B. SieveNet: A unified framework for robust image-based virtual try-on. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2171–2179, 2020.

[47] Ge, C. J.; Song, Y. B.; Ge, Y. Y.; Yang, H.; Liu, W.; Luo, P. Disentangled cycle consistency for highly-realistic virtual try-on. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16923–16932, 2021.

Chenghu Du is currently a master student in the School of Computer Science and Artificial Intelligence, Wuhan Textile University, where he received his B.S. degree in computer science and technology in 2019. His research interests include image processing and computer vision.

Feng Yu is currently a lecturer with the School of Computer Science and Artificial Intelligence, Wuhan Textile University. He received his Ph.D. degree from the School of Computer Science and Technology, Huazhong University of Science and Technology. His research interests include machine vision algorithms, artificial intelligence applications, and clothing intelligent manufacturing.

Minghua Jiang is currently the vice-chancellor of Wuhan Textile University, where he is also a professor with the School of Computer Science and Artificial Intelligence. He received his Ph.D. degree from the School of Computer Science and Technology, Huazhong University of Science and Technology. His research interests include computer system architecture, artificial intelligence applications, and clothing intelligent manufacturing.

Ailing Hua is currently a master student in the School of Computer Science and Artificial Intelligence, Wuhan Textile University, where she received her B.S. degree in software engineering in 2019. Her research interests include image processing and machine learning.

Yaxin Zhao is currently a master student in the School of Computer Science and Artificial Intelligence, Wuhan Textile University, where she received her B.S. degree in information management and information systems in 2019. Her research interests include deep learning and image processing.

Xiong Wei received his Ph.D. degree in computer architecture in 2011 and carried out postdoctoral research in computer architecture in 2011 in the National University of Defense Technology. He is currently an associate professor and a vice dean in the School of Computer Science and Artificial Intelligence at Wuhan Textile University. His research interests include storage architecture, GPUs, and parallel algorithms.

Tao Peng received his M.Sc. and Ph.D. degrees in computer science from Huazhong University of Science and Technology in 2006 and 2011, respectively. He is currently an associate professor in the School of Computer Science and Artificial Intelligence, Wuhan Textile University. His research interests include data mining, pattern recognition, and network security.

Xinron Hu earned her Ph.D. degree at the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology in 2008. Now she is a professor and dean in the School of Computer Science and Artificial Intelligence, Wuhan Textile University. Her research interests include image processing, virtual reality technology, and computer vision.