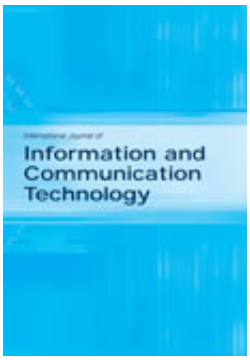
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Deep learning based automated generation method for fashion design

Chen Liang

DOI: 10.1504/IJICT.2025.10071991

Article History:

Received:	14 May 2025
Last revised:	25 May 2025
Accepted:	25 May 2025
Published online:	16 July 2025

Deep learning based automated generation method for fashion design

Chen Liang

College of Art and Design,
Zhengzhou University of Economics and Business,
Zhengzhou 451191, China
Email: 13653813369@163.com

Abstract: Intending to the issue of low fidelity of images generated by existing fashion design methods, this paper firstly segments the original garment images semantically based on VGG and Unet, encodes the target garment part images into different part features, and obtains the appearance flow field of the target garment parts. Then the target garment parts are deformed according to the appearance flow field of the garment parts, and the features of each garment part are fused. Finally, the garment images are generated in light of generative adversarial network (GAN) and diffusion model. The degrees of freedom are restricted by human posture control module and the region degree discriminator is used to enhance the local fine-grainedness of the garment images. The experimental results show that the structural similarity index (SSIM) of the proposed method is 0.895, and the generated results are clearer and more realistic.

Keywords: fashion design; automated generation; deep learning; generative adversarial network; GAN; diffusion model.

Reference to this paper should be made as follows: Liang, C. (2025) 'Deep learning based automated generation method for fashion design', *Int. J. Information and Communication Technology*, Vol. 26, No. 26, pp.51–67.

Biographical notes: Chen Liang received her Master's degree from the Southampton University in 2012. She is currently a Lecturer in the College of Art and Design at Zhengzhou University of Economics and Business. Her research interests are fashion design, intelligent fashion design and deep learning.

1 Introduction

Due to the improvement of the living standard of the common people, the demand for clothing is increasing, and online shopping is very convenient, so that the sales of online network clothing has been rapid growth, which leads to the demand for personalised clothing design is also more and more, according to a variety of information provided by the user to quickly design the finished image of the clothing becomes particularly important (Lee and Shin, 2020). Traditional design methods need to spend a lot of time to conduct market research, capture fashion trends, hand-drawn design sketches, repeated modification and improvement, the whole process is cumbersome and less efficient (Gam and Banning, 2011). Deep learning can realise the automated generation of apparel

design, quickly and accurately generate apparel style solutions to meet different needs, and greatly improve the design efficiency (Guan et al., 2019). At present, the research and application of deep learning in the automated generation of apparel design is still in the development stage, although some milestones have been achieved, but still facing many technical challenges and problems (Imtiaz et al., 2024), in-depth research on the automated generation method of apparel design based on deep learning has important theoretical significance and practical application value.

In earlier studies, researchers used CAD techniques for apparel design. Hwang and Zhang (2020) proposed CAD techniques for generating 3D garments, and they showed in more detail a framework suitable for designing garments for apparel designers, which pushed forward the progress of generating 3D garments. Jiang et al. (2020) proposed a CAD prototype of a garment assembler for the purpose of manufacturing complex shaped garments. These types of apparel generation methods produce good 3D apparel models, but require the designer to spend a lot of time learning CAD apparel modelling techniques. To cope with the above issue, the researchers placed the 3D human body model in the appropriate position, and then carried out a series of assembly to simulate the generation of three-dimensional clothing images. Later on, as sketches came into the picture, researchers focused more on how to model garments based on sketches. Yasseen et al. (2013) designed a sketch-based virtual garment generation method. Unfortunately, the results of these methods lack realism. Liu et al. (2023) proposed a context-aware image generation method for sketched garments, but the generation is not effective. Liu et al. (2019) and others proposed a method to generate garment images using only garment sketches. The above method is still similar to the 3D garment modelling method based on 2D patterns, which is still very labour-intensive and time-consuming.

The emergence of generative adversarial network (GAN) has led to further development of sketch-based garment image generation. Singh et al. (2020) utilised conditional GAN to implement a method of garment generation from sketch domain to real image domain, but the collection of fabric images was difficult. Wu et al. (2021) generated colourful images of categories such as shoes and sofas by an unsupervised method. This method does not use fabric patterns, but only uses clothing patterns that are relatively easy to collect. However, the resolution of the clothing images generated by this method is not high, and the pixel size of the image is only 128*128. Yan et al. (2022) proposed a garment generation algorithm for editing attributes such as collar and sleeve length of a blouse, which improves garment image generation. Fan (2024) used deep convolutional GAN for the style design of little black skirt and proposed a method for personalised design of clothing styles. Yan et al. (2023) enhanced the GAN family of models to optimise the quality of generated clothing pictures. Zhou et al. (2023) introduced symmetric loss to optimise the GAN model and reduced the training time.

Compared with GAN, diffusion network has higher training stability and can generate images with higher fine-grainedness, and different conditional features can be introduced into the network generation process to enhance its controllable generation ability, so as to maintain the texture and details of the garments and better produce garment images. Cao et al. (2023) by introducing parallel Unet diffusion network architecture with implicit garment deformation method. The cross-attention scheme is utilised to match the correlation among the target and real garments in implicit features of different scales, and finally the super-resolution diffusion network is utilised to generate high-resolution design results. Kang et al. (2024) introduced an autoencoder module based on diffusion networks, which utilises jump connections to embed undeformed target clothing and cue

text into implicit features, thus enhancing the generative capability of diffusion networks and controlling their generative degrees of freedom. Chen and Ma (2024) introduced a garment deformation module as a local control condition to increase the controllability of the diffusion network, and maximised the preservation of garment texture details and human body details in the generated results by establishing rough reconstruction branches and refined reconstruction branches.

To summarise, the existing clothing design methods do not directly take the human body posture as a control factor in the clothing generation process, resulting in low fidelity of the generated images, for this reason, this paper proposes an automated generation method for clothing design in light of deep learning. Firstly, the semantic segmentation of the original clothing image is based on VGG and Unet, and the human parsing features are embedded in the network encoding and decoding part through the cross-attention mechanism to realise the corresponding semantic information matching and improve the accuracy of clothing segmentation. After the target garment image segmentation is completed, the target garment part images are encoded into different part features to obtain the appearance flow field of the target garment parts. Then the 3D human posture information corresponding to the 2D human image is reconstructed by the 3D human posture estimation model to obtain the spatial order of each part of the human body. Using the multi-part garment deformation network, each part of the target garment is deformed according to the appearance flow field of the garment parts, and the features of each garment part are fused under the guidance of the spatial order of human body parts. Finally, the global human body features and the fused clothing features are introduced into the diffusion network, and the human body pose control module and the region degree discriminator restrict the generation degree of freedom to ensure the consistency of the human body morphology before and after the generation, so as to improve the local fine-grained clothing image and obtain high-quality clothing image generation results. The experimental outcome demonstrates that the SSIM and IS of the designed approach are 0.895 and 3.68, respectively, which are better than the benchmark method and can generate high-quality garment images.

2 Relevant technologies

2.1 Generating adversarial network

GAN is a deep learning modelling framework that generates data that is virtually indistinguishable from actual data through an adversarial process (Cheng et al., 2020). A GAN consists of two components: a generator (G) and a discriminator (D), which compete with each other during the training process to continuously improve performance. The goal of G is to generate high-quality pseudo-data that are as close as possible to the actual data distribution (Sajeeda and Hossain, 2022). It takes a random noise vector as input and transforms it into an output of the same dimension as the real data through a series of network layers. The design and complexity of G depends on the needs of the particular task, e.g., image generation, text generation, etc. The objective of D is to determine whether the input is from the real dataset or generated by the model. Typically, it is a binary classification model which generates a scalar value, representing the likelihood that the input data is genuine. Discriminator also consists of multiple

network layers whose structure can be adapted to the complexity of the task and the type of data.

The training process of a GAN can be viewed as a two-player zero-sum game of min-max, where G tries to maximise the probability that the discriminator will misclassify the generated data and D tries to minimise this probability to accurately distinguish between true and false data. This process can be represented by equation (1).

$$\min_{G}\max V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

where x is the real data, z is the input noise to the generator, $D(x)$ is the probability that D evaluates the data as real, and $G(z)$ is the fake data generated by G based on the noise.

2.2 VGG feature extraction method

As deep learning is getting more and more attention from researchers, there are many deep CNNs, such as LeNet, AlexNet, VGGNet, etc. VGGNet constructs deep networks with 16–19 layers, (e.g., VGG16, VGG19) by stacking multiple 3 × 3 small convolutional kernels, which are able to extract more complex and abstract features compared to LeNet (five-layer) and AlexNet (eight-layer). This paper uses VGGNet for feature extraction, which is able to learn more complex and abstract feature representations. Although they appeared earlier, VGG networks are still adopted by many researchers for feature extraction. Although they appeared earlier, VGG networks are still adopted by many researchers for feature extraction. The structure in the VGG network uses a convolutional kernel of size 3 × 3 and a pooling kernel of size 2 × 2, only deepening its structure improves the performance of this network. However, the total number of parameters it requires will not grow explosively with the increasing depth of the network, because most of the parameters of VGG networks are only in the fully connected layer (Kim, 2018).

The VGG is characterised as follows: it starts with several structural blocks, a block with two 3 × 3 convolutional kernels combined with a convolutional layer, which can be equivalent to a 5 × 5 convolutional level with a receptive domain. The structure consists of three convolutional levels with 3 × 3 convolutional kernels, which are similar to a convolutional layer with 7 × 7 convolutional kernels, but with half the amount of parameters of a 7 × 7 convolutional layer. Moreover, three nonlinear activation operations can be added to three 3 × 3 convolutional levels, but only one to a 7 × 7 convolutional level, so this makes the structure of the VGG network more capable of learning features.

2.3 Diffusion model theory

The diffusion model uses variational inference to train a parameterised Markov chain and is a hidden variable model (Ratcliff et al., 2016). It achieves image generation by incrementally adding Gaussian noise to an original image (diffusion process) and denoising the noise by neural network prediction. The diffusion progress is a Markov chain and the probability distribution $q(x_t|\ x_{t-1})$ of the process can be expressed as follows.

$$q\left(x_t \mid x_{t-1}\right) = N\left(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I\right) \tag{2}$$

where N is a Gaussian distribution; t is the amount of diffusion steps; x_t is the noise image after the t^{th} step of the diffusion progress; x_{t-1} is the noise image after the $t - 1^{th}$ step of the diffusion process; β_t is a preset value of the hyperparameter; I is the unit matrix and T is the total amount of diffusion steps. Define hyperparameter $\alpha_t = 1 - \beta_t$ and hyperparameter $\bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$ to obtain equation (3).

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon \tag{3}$$

where x_0 is the initial image and ε is the real Gaussian noise added at all steps of the diffusion progress.

The probability distribution $q(x_{t-1}| x_t)$ of the generation progress is approximated by the neural network, and the predicted probability distribution $p_\theta(x_{t-1}| x_t)$ is as follows.

$$p_\theta\left(x_{t-1}|x_t\right) = N\left(x_{t-1}; \mu_\theta\left(x_t, t\right), \beta_\theta\left(x_t, t\right)\right) \tag{4}$$

where θ is the model training parameter; $\mu_\theta(x_t, t)$ and $\beta_\theta(x_t, t)$ are the mean and variance of the distribution, respectively.

3 Fashion image pre-processing and semantic segmentation

3.1 Fashion image pre-processing

Aiming at the challenge of difficult feature extraction in garment images, this paper adopts a garment region extraction method based on semantic segmentation to get the corresponding category of garment regions and perform region alignment operation to exclude the background noise interference. The semantic segmentation of garment images is performed using VGG-Unet model so as to focus on the effective garment regions.

The target clothing region is the most critical part in the input model clothing image. Therefore, for the processing and analysis of clothing images, fine detection and extraction of the target clothing region is required. The method is difficult to avoid extracting to other background regions, which makes the acquired training data of low quality. To cope with this issue, the garment region can be precisely determined by semantic segmentation of the image. Irrelevant noise regions in the image such as background, character limbs, etc. will be got rid of, thus enhancing the quality and usability of the data. On the basis of semantic segmentation, subsequent operations, such as planarisation, can be performed on the garment images to obtain more accurate and useful results.

The objective of this paper is to extract the target garment region from the human garment image. A semantic segmentation outcome map with the same size as the original picture can be obtained by semantic segmentation technique. In this graph, different clothing category regions are labelled with different colours. Filtering out non-target regions by colour labels results in a semantic segmentation graph of the target regions. Next, each pixel value in the resulting target area map is converted to a binary value so that the image can be masked in subsequent operations. By merging the semantic segmentation map with the input picture of the original approach, the final clothing

picture of the target area can be obtained. The complete process of semantic segmentation of clothing region includes the following steps. First, the input image is segmented by the semantic segmentation approach, which map of clothing is obtained. Then, the target region is selected according to the clothing category of the input image, and the semantic segmentation map of the target clothing region is obtained. Then, the target region is converted to binary value, and the original image is masked with the obtained mask image. Finally, the final image of the target area can be obtained, which has a clean background.

3.2 Semantic segmentation of clothing based on VGG and Unet

In the above pre-processing process, this paper constructs a garment semantic segmentation model based on VGG-Unet (Mei et al., 2021). First, the network encodes a semantic parsing group used to characterise 2D human body information into human parsing features at different scales through a human feature encoder, and utilises the VGG-Unet network to segment the target garment into five garment parts, including the upper left sleeve, the lower left sleeve, the front piece of the garment, the upper right sleeve, and the lower right sleeve. This VGG-Unet network embeds human parsing features in the encoding and decoding part of the network by the cross-attention scheme to realise the corresponding semantic information matching and improve the accuracy of clothing segmentation. After the target garment image segmentation is completed, the garment feature encoder encodes the target garment part image into five groups of part features, and the garment part appearance flow field estimation module captures the nonlinear correlation between the human body analytic features and each part feature to obtain the target garment part appearance flow field.

Secondly, due to the lack of complete clothing information in the target clothing component images, the multi-component clothing deformation network has a weak perception ability of clothing contours and categories, leading to a decrease in the estimation accuracy of the appearance flow field of the target clothing component. Therefore, in this paper, the missing clothing information is supplemented through the clothing contour and category coding module to enhance the low-dimensional perception ability of the network, as shown in Figure 1. According to the number of clothing parts in the target clothing parts image, the module judges the style category and carries out the category feature coding; the edge detection approach is used to extract the overall contour of the target clothing image, isometric screening of the 63 contour coordinate points and its sine and cosine coding, the coding formulas are as follows.

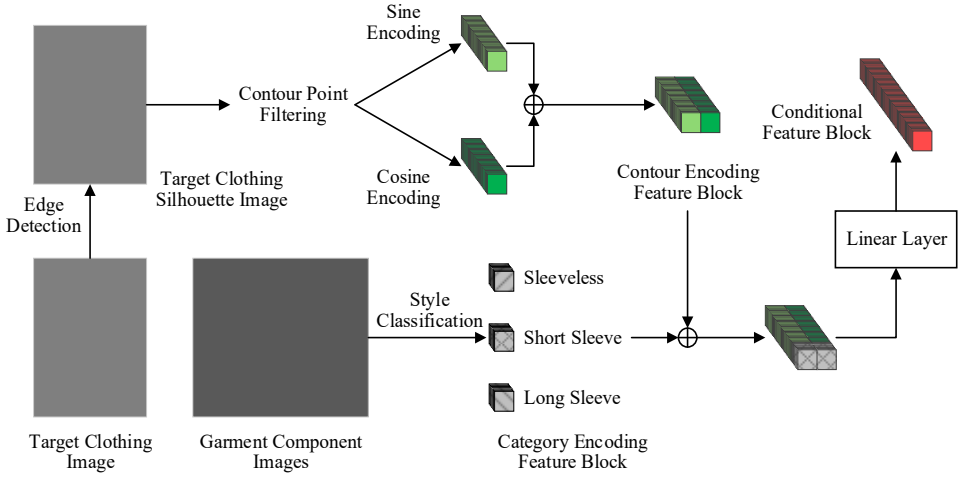
$$\left(P_x'\right)_i = \sin\left(\frac{\left(P_x\right)_i}{W}\right) \tag{5}$$

$$\left(P_y'\right)_i = \cos\left(\frac{\left(P_y\right)_i}{H}\right) \tag{6}$$

where i is the index of the outline point, $(P_x)_i$ and $(P_y)_i$ are the horizontal and vertical coordinates of the i^{th} garment outline point, $\left(P_x'\right)_i$ and $\left(P_y'\right)_i$ are the horizontal and vertical coordinates of the i^{th} outline point after sine and cosine coding, W and H are the width and height of the target garment outline image. The conditional feature block is incorporated into each garment part feature through feature embedding, enhancing the

multi-part garment deformation network's perception of garment contours and category information.

Figure 1 Semantic segmentation of clothing based on VGG and Unet (see online version for colours)

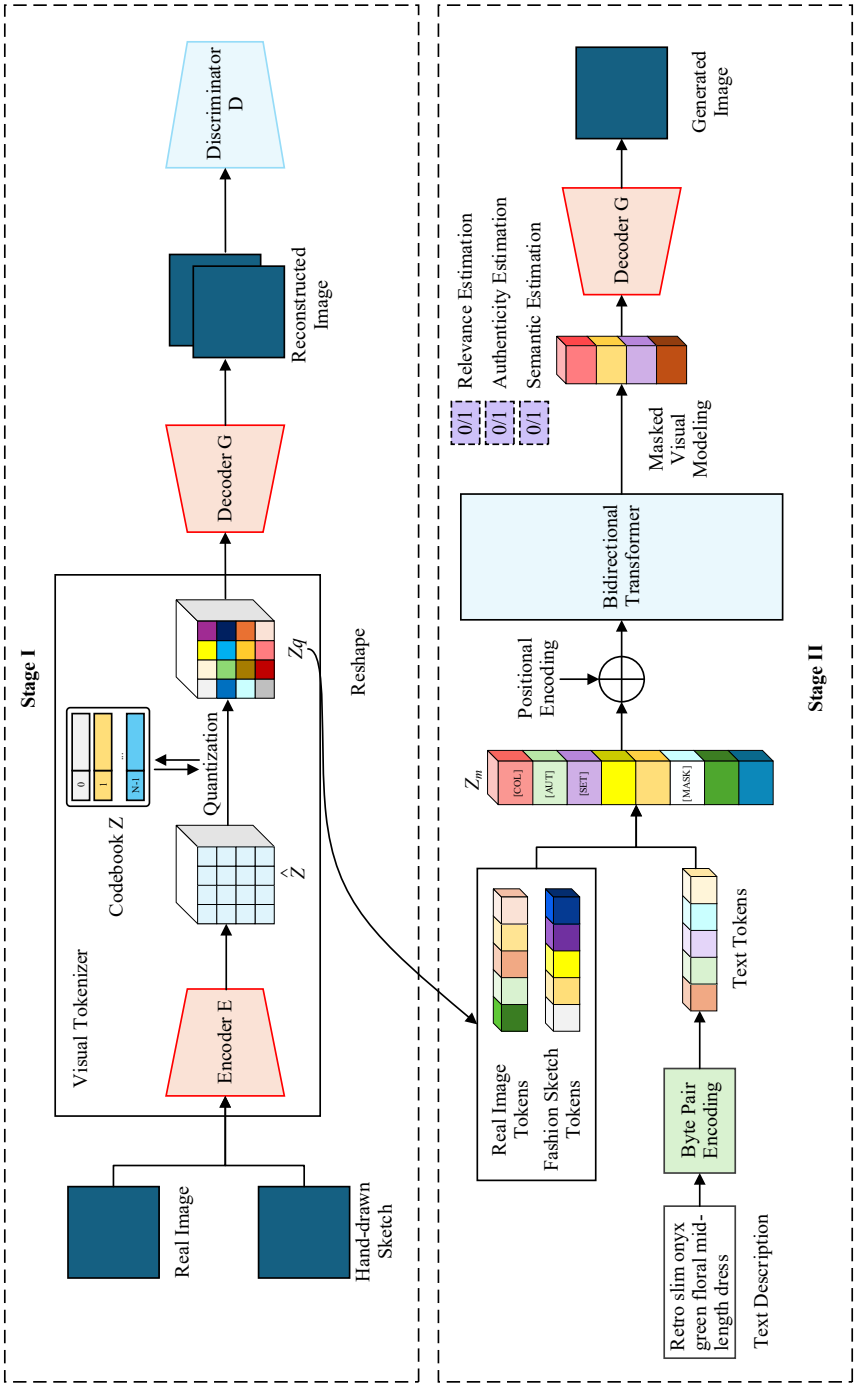


4 Automated generation of fashion designs based on deep learning

4.1 Multi-part garment deformation based on 3D human posture information networks

Aiming at the existing clothing design automatic generation method does not directly take the human body posture as a control factor to restrict the clothing generation process, resulting in low fidelity of clothing details before and after the generation of the problem, this paper proposes a clothing design automation generation method in light of deep learning, as shown in Figure 2. The network introduces a 3D human pose estimation model to obtain the spatial order of human body parts. Through the multi-part garment deformation method, the deformed garment parts are superimposed and fused based on the spatial order of various parts of the human body, so as to alleviate the distortion of garment deformation caused by joint occlusion. Extract the overall information of clothing through the clothing outline and category coding module, strengthen the low-dimensional perception ability of the network, and improve the accuracy of the overall deformation of clothing. Using GAN and diffusion model for feature fusion between human body image and deformed clothing image, the human body posture control module is introduced to improve the overall controllability of the diffusion network, to ensure the consistency of the human body morphology before and after the generation, and to obtain realistic and natural clothing generation results.

Figure 2 Automated generation of garment designs based on deep learning (see online version for colours)



To improve the local estimation accuracy of the appearance flow field and solve the problem of low accuracy of garment deformation at the joints, this paper introduces three-dimensional human body posture information, proposes a multi-component garment deformation method, and constructs a multi-component garment deformation network based on three-dimensional human body posture information. The 3D human posture information corresponding to the image to be generated is obtained through the 3D human posture estimation model, and the spatial order of each part of the human body is extracted, based on which the deformed images of each garment part are fused. Using hybrid neural analysis inverse motion network to reconstruct the 3D human pose point cloud model of a single image to be generated, the 3D point cloud coordinates are mapped to the 2D plane by dimensional regression method to obtain the preliminary 2D human depth image, as shown below.

$$C = F(R(I)) \tag{7}$$

$$I_{\mathrm{D}}(x, y) = (C_i)_z \tag{8}$$

where I is the image to be generated; R is a single image to restore the coordinates of the 3D human point cloud; F is the point cloud filtering operation, the point cloud filtering operation returns all the point clouds to the two-dimensional plane and filters the overlapping point clouds to obtain the one-sided human point cloud information; C is the set of filtered point clouds; i is the index of the point cloud; $(C_i)_z$ is the z-axis coordinates of the i^{th} point cloud; and $I_{\mathrm{D}}(x, y)$ is the pixel value of corresponding pixel point in the two-dimensional depth image of the human body.

To obtain the spatial order of each part of the human body, this paper assigns depth values to each part of the human body through 2D human body depth image and human body dense pose semantic image. This paper constructs a human body shape deformation network to reduce the contour variability and improve the accuracy of depth assignment through the principle of clothing deformation method. The network utilises human joint point images, human dense pose semantic images to deform the initially obtained 2D human depth image up to its contour as shown below.

$$L_1^D = \left| I_{Dm} - I'_{Dm} \right| \tag{9}$$

where L_1^D is the L_1 paradigm loss for the mask of the deformed 2D human depth image, I_{Dm} is the mask of the dense pose semantic image of the human body, and I'_{Dm} is the mask of the deformed 2D human depth image output by the human body shape deformation network.

In this paper, the pixel-level average depth of each body part is calculated and ranked based on the dense pose semantic image of the human body and the deformed 2D body depth image as follows.

$$\sigma_i = \frac{I'_{\mathrm{D}} \cdot I_i^{Dm}}{s_i^{Dm}} \tag{10}$$

where I'_{D} is the deformed two-dimensional human body depth image, I_i^{Dm} is the mask image of the i^{th} human body part in the dense pose semantic image of the human body; s_i^{Dm} is the area of this mask image; σ_i is the average pixel depth value of the corresponding human body part; and $i \in \{1, 2, \ldots, 5\}$ corresponds to the left shoulder, the

left arm, the upper limb torso, the right shoulder, and the right arm of the human body, respectively.

After obtaining the three-dimensional spatial order of each part of the human body, the deformed garment parts are superimposed and fused, and the corresponding relationships between the human body parts and the garment parts are as follows: left shoulder – left upper sleeve, left arm – left lower sleeve, upper limb torso – front piece of the garment, right shoulder – right upper sleeve, right arm – right lower sleeve. The superposition equation is as below.

$$w_i^c = \begin{cases} w_i^p, i = 1; \\ w_{i-1}^c + \left(w_{i-1}^c \cdot w_i^p \right), i > 1 \end{cases} \tag{11}$$

where w_i^p is the i^{th} deformed clothing part after sorting the spatial order of clothing parts from far to near, and w_i^c and w_{i-1}^c are part of the clothing images superimposed on i and $i - 1$ clothing parts. When i is 5, the garment superposition process ends and the final deformed target garment image is obtained.

To train the multi-part garment deformation network proposed in this paper, this paper uses L_1-paradigm losses L_1^c, L_1^{cm} and image-perception loss L_{VGG}^c on the final deformed target garment images and their masks to enhance the consistency of the texture details and shape distribution of the target garments before and after deformation. The equations for calculating L_1^c, L_1^{cm}, and L_{VGG}^c are as follows.

$$L_1^c = \left| w_c \cdot w_{cm} - w_c' \cdot w_{cm}' \right| \tag{12}$$

$$L_1^{cm} = \left| w_{cm}' - w' \right| \tag{13}$$

$$L_{VGG}^c = \sum \left| \varnothing_i \left(w_c \cdot w_{cm} \right) - \varnothing_i \left(w_c' \cdot w_{cm}' \right) \right| \tag{14}$$

where w_c and w_{cm} are real human attire images and their masks, w_c' and w_{cm}' are deformed clothing images and their masks, and $\varnothing_i$ is the i^{th} level of the pre-trained VGG network.

4.2 Automated generation of apparel designs based on GAN and diffusion models

In this paper, the clothing design generation phase is realised based on GAN and diffusion model training, and its training objective function is as follows.

$$E_{x_0,t,\varepsilon} \left(\left\| \varepsilon - \varepsilon_\theta \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon, t \right) \right\| \right) \tag{15}$$

where x_0 is the real data distribution, time $t \in \{1, 2, 3, \ldots T\}$ follows a uniform distribution, ε is the randomly generated Gaussian noise, α_t is the training variance, $\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon$ is the noise-added image at time t with the addition of a random Gaussian noise ε to x_0 and ε_θ is the generative diffusion network. To limit the generative degrees of freedom of generative diffusion networks in the automated generation phase of clothing design and to improve its feature retention ability, this paper proposes a

diffusion network based on human posture control module and region degree discriminator.

First of all, the network processes the image to be tried on into the image to be filled with only the head, the lower limbs, the deformed target garment, and part of the background according to the 2D human body resolution information, and takes the real noise-added image and the mask to be filled as the input of the diffusion network together. Unlike other diffusion network-based 2D fitting, this paper does not choose the mask filling method based on human body contour, but utilises a rectangular region on its basis to completely remove the dress and some body information of the person, in order to prevent the generation process from being affected by the shape of the filling mask and the residual dress of the human body.

Secondly, in this paper, the proposed human posture control module downsamples and encodes the dense posture semantic image of human body, obtains the feature blocks with different scale sizes, and sums them with the output of the sampling and decoding stage on the generative diffusion network to realise the dense matching between the feature information, so as to ensure the consistency of the human body before and after trying on the clothes. Meanwhile, to obtain the global information of the target garment, this paper encodes the conditional features of the target garment image through the linguistic image comparison pre-training network CLIP (Tu et al., 2023), and uses the cross-attention mechanism to incorporate this encoded feature into the backbone network to control the generation of garment texture and colour. The CLIP network and the encoder part of the generative diffusion network do not update the network parameters during the training process to enhance the convergence speed and training stability of the diffusion network.

Finally, in order to make the generative diffusion network can accurately restore the local details of the image, and enhance the network's capability to capture and maintain the local features, this paper introduces the region degree discriminator in PatchGAN (Chen et al., 2023) and the diffusion network for adversarial training, so as to strengthen the network's learning ability, and its training objective function is as follows.

$$L_P^D = E\left(\log(1 - D(I_n'))\right) + E\left(\log\left(D(I_n)\right)\right) \tag{16}$$

where L_P^D is the loss of region degree discriminator, I_n' is the pseudo-noise added image by adding the predicted noise to the image to be tried on, I_n is the real noise added image by adding the input random noise to the image to be tried on, and D is the region degree discriminator. In addition, the region degree discriminator can work with the human posture control module to guarantee the authenticity and integrity of the results produced by the diffusion network in the localised garment image region.

In this paper, the above training losses are numerically weighted to obtain the total training losses L_W and L_G of the multi-component garment deformation network based on 3D human posture information, the diffusion network based on the human posture control module and the region degree discriminator.

$$L_W = \lambda_c \cdot L_1^c + \lambda_{cm} \cdot L_1^{cm} + \lambda_{VGG} \cdot L_{VGG}^c + \lambda_S \cdot L_S \tag{17}$$

$$L_G = \lambda_\varepsilon \cdot L_2^{DM} + \lambda_D \cdot L_P^D \tag{18}$$

where λ_c, λ_{cm}, λ_{VGG}, λ_S, λ_ε, λ_D and are the training hyperparameters.

5 Experimental results and analyses

This paper selects the pictures of women's fashion shows in Paris, Milan, London, and New York Fashion Weeks from 2018–2023 as the source of training data, and for the goal of ensuring the unity of the human body characteristics of the generated pictures, only the pictures of models in walking postures in the shows are selected as the training set, and the pictures of standing postures, studio shots, and art blockbusters are manually removed to ensure that the training data set has a more unified model posture. Before training, all the dataset images were batch pre-processed to obtain 42,287 show images with a resolution of 445 × 445, and the generated images had to be uniformly adjusted back to the normal scale due to the vertical compression of the image scale after pre-processing.

The deep learning framework used for the experiments is PyTorch. During training, this model sets the batch size to 8 and the amount of training iterations to 200. The initial learning rate is also set to 0.0001, and after 100 iterations, the studying rate decreases linearly, and the whole experiment is optimised by Adam's optimiser. The selection of hyperparameters in the experiments was made after several experiments to choose the parameters with the best results so far, and the running environment is Linux operating system and the central core processor of i5 is used.

To assess the practical performance of the designed network, Frechette distance (FID), perceived image block similarity (LPIPS), structural similarity index (SSIM), and inception score (IS) are selected as the quantitative evaluation metrics of the generated images for garment design. GAN-DN is compared and analysed with three methods, namely, TID-GAN (Yan et al., 2022), DRB-DM (Cao et al., 2023), and GM-SDN (Chen and Ma, 2024), and the comparative metrics of the effectiveness of different methods in generating garment images are shown in Figure 3. GAN-DN outperforms other methods on FID, LPIPS, KID, and IS. The improvement of GAN-DN over TID-GAN, DRB-DM, and GM-SDN on SSIM indicates that the method is capable of generating higher-quality garment images. The reduction of FID indicates that the garment pictures produced by GAN-DN are better than other networks in representing low-level features. The decrease in LPIPS indicates that the garment images generated by GAN-DN are more in line with human visual perception. The increase in the IS score indicates that GAN-DN is able to generate more realistic and high-quality garment images.

Comparison of the loss rate of different methods is shown in Figure 3, because the generalisation of the features extracted from the GAN-DN backbone feature extraction part is high, and the performance of the benchmark methods selected in this paper are all very good, so this paper adopts the method of freezing the training to improve the training efficiency. The epoch of the freezing phase is set to 50. In the freezing phase, the weights of the backbone feature extraction network are frozen, and only the network is fine-tuned. After 50 epochs the network unfreezes and starts to adjust the weights of the individual methods, thus leading to an abrupt change in the loss rate. In terms of the loss rate, all four methods perform well, as shown in Figure 4, after thawing, the loss rate decreases rapidly and eventually stabilises, indicating that the results have been stabilised, and the final value of GAN-DN is less than 0.16 but not zero, indicating that GAN-DN is not overfitted.

Figure 3 Comparison of garment image generation results (see online version for colours)

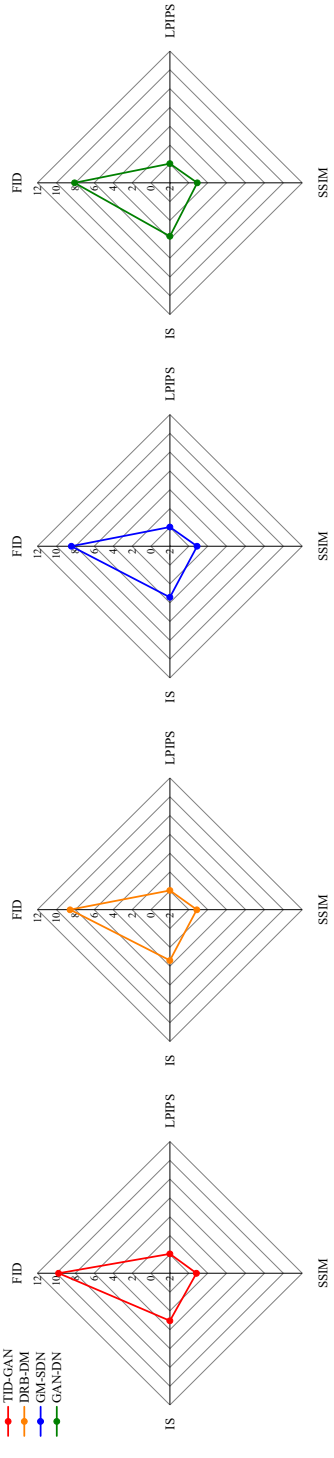
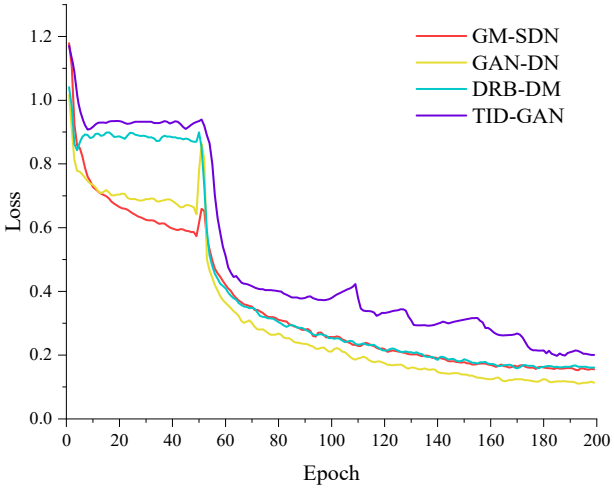


Figure 4 Comparison of loss ratios for different models (see online version for colours)



For the goal of further validating the effectiveness of each component in GAN-DN, this paper conducts ablation experiments on multi-part garment deformation network and generative diffusion network, respectively. First of all, to verify the effectiveness of different modules in the multi-component garment deformation network in improving the overall performance of the network, the garment outline module, the spatial order embedding module of each part of the human body, and the region discriminator module were used as the control variables of the elimination experiments, and were tested and analysed in the GAN-DN, as shown in Table 1.

Table 1 Results of ablation experiments for each component

Method	*Silhouette*	*Spatial embedding*	*Regionality discriminator*	*FID*	*LPIPS*	*SSIM*	*IS*
M1	×	√	√	8.46	0.0453	0.819	3.48
M2	√	×	√	8.57	0.0482	0.861	3.09
M3	√	√	×	8.29	0.0447	0.842	3.31
GAN-DN	√	√	√	8.08	0.0426	0.895	3.68

When shielding the clothing outline module (M1), the scores of each image evaluation index have decreased, and the decrease is larger, the consistency of the human body pose before and after the generation of clothing images is poor, and there is a loss of clothing details, distortion phenomenon, and the overall quality of the generation of a serious decline. When adding the spatial order embedding (M2) module for each part of the human body, the global feature information used to characterise the clothing features can be extracted and interacted and matched with the generative diffusion network to realise the feature information, and the generative freedom of the diffusion network is constrained to make the details of the generative results complete and clear. When the regional degree discriminator (M3) was not added during the network training process, the variant network was reduced in all image evaluation index scores, its ability to maintain local garment details was poor, and fabric fold patterns unrelated to the target garments appeared in the generated results, so it was verified that the regional degree

discriminator could enhance the network's ability to capture local features and generation stability. Combining the above analysis, the GAN-DN incorporating all the components shows the best generation results.

6 Conclusions

Fashion design is a technique that applies design aesthetics and natural beauty to clothing and its accessories. In this paper, to address the problem that existing clothing design methods do not directly take the human body posture as a control factor in the process of clothing generation, which leads to low fidelity of the generated images, we first semantically segment the original clothing images based on VGG and Unet, and then embed the human body parsing features in the network coding and decoding part through the cross-attention mechanism in order to realise the matching of the corresponding semantic information. Then the images of the target garment parts are encoded into different part features to obtain the appearance flow field of the target garment parts. Secondly, the 3D human posture estimation model is used to estimate the 3D human posture, obtain the spatial order of human body parts, constrain and guide the multi-part garment deformation process, and enhance the spatial sensing ability of the garment deformation network, and add the garment contour and category coding module, which enhances the network's data sensitivity and low-dimensional sensing ability. Finally, based on the diffusion network, the global and local information capturing ability of the diffusion network is improved by adding the human body pose control module and the regional degree GAN discriminator, which limits the network generation degree of freedom and obtains high-quality image generation results. The experimental results show that the FID, LPIPS, SSIM and IS of the designed approach are better than those of the benchmark approach, and it can generate clear and natural garment images, which can help to enhance consumers' satisfaction in purchasing garments online and promote the development of the garment e-commerce industry.

There are still some deficiencies in the approach designed in this paper, such as: in the garment deformation stage, for garments with complex textures, there are texture distortion and distortion problems after deformation; in the image generation stage, the texture of the garments can not be kept intact. To cope with these issues, it is essential to further enhance the estimation accuracy of the appearance flow field and limit the degree of freedom of the image generation network to meet the user's demand for the generation of garments with complex textures.

Declarations

All authors declare that they have no conflicts of interest.

References

Cao, S., Chai, W., Hao, S., Zhang, Y., Chen, H. and Wang, G. (2023) 'Difffashion: reference-based fashion design with structure-aware transfer by diffusion models', *IEEE Transactions on Multimedia*, Vol. 26, pp.3962–3975.

Chen, G., Zhang, G., Yang, Z. and Liu, W. (2023) 'Multi-scale patch-GAN with edge detection for image inpainting', *Applied Intelligence*, Vol. 53, No. 4, pp.3917–3932.

Chen, Y. and Ma, J. (2024) 'An intelligent generative method of fashion design combining attribute knowledge and stable diffusion model', *Textile Research Journal*, Vol. 12, pp.11–23.

Cheng, J., Yang, Y., Tang, X., Xiong, N., Zhang, Y. and Lei, F. (2020) 'Generative adversarial networks: a literature review', *KSII Transactions on Internet and Information Systems (TIIS)*, Vol. 14, No. 12, pp.4625–4647.

Fan, L. (2024) 'Design and application of an intelligent generation model for fashion clothing images based on improved generative adversarial networks', *Service Oriented Computing and Applications*, Vol. 21, No. 4, pp.1–14.

Gam, H.J. and Banning, J. (2011) 'Addressing sustainable apparel design challenges with problem-based learning', *Clothing and Textiles Research Journal*, Vol. 29, No. 3, pp.202–215.

Guan, C., Qin, S. and Long, Y. (2019) 'Apparel-based deep learning system design for apparel style recommendation', *International Journal of Clothing Science and Technology*, Vol. 31, No. 3, pp.376–389.

Hwang, C. and Zhang, L. (2020) 'Innovative sustainable apparel design: application of CAD and redesign process', *Sustainability in the Textile and Apparel Industries: Sustainable Textiles, Clothing Design and Repurposing*, Vol. 12, pp.87–107, Centro de Estudios para el Lujo Sustentable, Buenos Aires, Argentina.

Imtiaz, A., Pathirana, N., Saheel, S., Karunanayaka, K. and Trenado, C. (2024) 'A review on the influence of deep learning and generative AI in the fashion industry', *Journal of Future Artificial Intelligence and Technologies*, Vol. 1, No. 3, pp.201–216.

Jiang, Z., Guo, J. and Zhang, X. (2020) 'Fast custom apparel design and simulation for future demand-driven manufacturing', *International Journal of Clothing Science and Technology*, Vol. 32, No. 2, pp.255–270.

Kang, M., Kim, J. and Kim, S. (2024) 'Unsupervised generation of fashion editorials using deep generative model', *Fashion and Textiles*, Vol. 11, No. 1, pp.21–33.

Kim, M-K. (2018) 'Contactless palmprint identification using the pretrained VGGNet model', *Journal of Korea Multimedia Society*, Vol. 21, No. 12, pp.1439–1447.

Lee, J.E. and Shin, E. (2020) 'The effects of apparel names and visual complexity of apparel design on consumers' apparel product attitudes: a mental imagery perspective', *Journal of Business Research*, Vol. 120, pp.407–417.

Liu, K., Zeng, X., Tao, X. and Bruniaux, P. (2019) 'Associate design of fashion sketch and pattern', *IEEE Access*, Vol. 7, pp.48830–48837.

Liu, X., Li, J. and Lu, G. (2023) 'Wrinkles realistic clothing reconstruction by combining implicit and explicit method', *Computer-Aided Design*, Vol. 16, pp.35–43.

Mei, Y., Jin, H., Yu, B., Wu, E. and Yang, K. (2021) 'Visual geometry group-UNet: deep learning ultrasonic image reconstruction for curved parts', *The Journal of the Acoustical Society of America*, Vol. 149, No. 5, pp.2997–3009.

Ratcliff, R., Smith, P.L., Brown, S.D. and McKoon, G. (2016) 'Diffusion decision model: current issues and history', *Trends in Cognitive Sciences*, Vol. 20, No. 4, pp.260–281.

Sajeeda, A. and Hossain, B.M. (2022) 'Exploring generative adversarial networks and adversarial training', *International Journal of Cognitive Computing in Engineering*, Vol. 3, pp.78–89.

Singh, M., Bajpai, U., Vijayarajan, V. and Prasath, S. (2020) 'Generation of fashionable clothes using generative adversarial networks: a preliminary feasibility study', *International Journal of Clothing Science and Technology*, Vol. 32, No. 2, pp.177–187.

Tu, W., Deng, W. and Gedeon, T. (2023) 'A closer look at the robustness of contrastive language-image pre-training (clip)', *Advances in Neural Information Processing Systems*, Vol. 36, pp.13678–13691.

Wu, Q., Zhu, B., Yong, B., Wei, Y., Jiang, X., Zhou, R. and Zhou, Q. (2021) 'ClothGAN: generation of fashionable Dunhuang clothes using generative adversarial networks', *Connection Science*, Vol. 33, No. 2, pp.341–358.

Yan, H., Zhang, H., Liu, L., Zhou, D., Xu, X., Zhang, Z. and Yan, S. (2022) 'Toward intelligent design: an AI-based fashion designer using generative adversarial networks aided by sketch and rendering generators', *IEEE Transactions on Multimedia*, Vol. 25, pp.2323–2338.

Yan, H., Zhang, H., Shi, J., Ma, J. and Xu, X. (2023) 'Toward intelligent fashion design: a texture and shape disentangled generative adversarial network', *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 19, No. 3, pp.1–23.

Yasseen, Z., Nasri, A., Boukaram, W., Volino, P. and Magnenat-Thalmann, N. (2013) 'Sketch-based garment design with quad meshes', *Computer-Aided Design*, Vol. 45, No. 2, pp.562–567.

Zhou, D., Zhang, H., Ma, J. and Shi, J. (2023) 'BC-GAN: a generative adversarial network for synthesizing a batch of collocated clothing', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 34, No. 5, pp.3245–3259.