ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

# Research on clothing pattern generation and design optimization based on machine learning technology

*Zhihui Li*[1,*]

[1] College of Finance and Economics, Chongqing Industry & Trade Polytechnic, Fuling, Chongqing, 408000, China

**SUMMARY:** *Aiming at the diversity of clothing patterns and the complexity of design, in order to better understand the creative expression of clothing patterns, this paper utilizes the CLIP text encoder to preprocess textual information labels. Combined with the text requirements, the potential diffusion model is selected to generate simple tuples and complex patterns of clothing patterns. Stable diffusion model is selected and diffusion inversion method is used to alleviate the limitation of text shape. Based on the shape grammar theory, the samples of clothing patterns generated based on diffusion inversion are varied to realize pattern redesign. Select public datasets and multiple pattern generation models, and compare the results of FID scores, IS scores and other indexes of each modeling method. For the clothing pattern samples generated based on diffusion inversion, the amateur group and the expert group are invited to score the samples in four dimensions, namely, image quality, overall aesthetics, interpretation of the pattern to the textual content, and overall coordination between the pattern and the text, respectively. On the MSCOCO dataset, the FID scores of the diffusion inversion-based method for generating garment image samples decreased by 41.52%, 44.97%, 49.42%, and 52.87% compared to the VLMGAN model, DM-GAN model, SSA-GAN model, and AttnGAN model, respectively. On the CUB-200 dataset, this paper's method improves the IS score by 30.83% compared with the SSA-GAN model. Combining the comparison results of each index and the subjective evaluation results, the clothing pattern samples generated based on diffusion inversion fit the text requirements, have good image quality and are generally well received.*

***KEYWORDS:*** *clip text encoder; diffusion model; garment pattern generation; FID score*

## 1 Introduction

At the present time of rapid change, with the continuous improvement of people's consumption demand and the psychology of seeking differences, in addition to the rigid demand attributes of clothing products with necessities of life, its fashion, functionality and other personalized, diversified and elastic demand attributes have begun to be valued [1, 2]. Among them, pattern is a kind of decorative art of clothing, which is a kind of fine art form combining decorative and usability, and clothing pattern is the perfect embodiment of this form of fine art [3]. If clothing is a kind of culture, then clothing pattern is the carrier of this culture. Patterns can beautify the dress, can dress up the appearance of people, so as to bring out the beauty of women, men's fitness, so that people in the spirit of the enjoyment of beauty [4].

At the same time, the pattern pattern itself is an emotional symbol, which brings together the traditional culture of a nation and regional culture, contains a certain agreed cultural

*mzbaby1007@163.com
https://doi.org/10.65102/is2026037 Published: 30 April, 2026

connotation, reflects the spirit of an era, expresses the specific life sentiment and aesthetic concepts of a nation [5-7]. The use of clothing patterns is a point of focus often touched upon in clothing design. Pattern style, texture, color, content, etc. will affect the overall effect of clothing design, but it is not simply carry a few foreign letters to learn the West, nor is it hard to put together a few dragon and phoenix motifs to promote the national style has become the so-called design, appropriate use of the pattern, the overall clothing can play the role of the eye-dotting [8, 9]. Therefore, the exploration of clothing patterns, research on the entire clothing design process is very necessary.

Image recognition and generation technology based on machine learning is the key technology of visual artificial intelligence, and the rapid development of artificial intelligence technology is driving the apparel industry towards digitalization [10, 11]. How to automatically and accurately obtain relevant clothing demand information to cope with the severe market challenges has become the focus of attention of major clothing e-commerce companies. Clothing pattern is the focus of clothing information data mining, because clothing pattern contains comprehensive and clear popularity information and it is easier to transmit data and information in a visual way [12]. Color, style and fabric are the “three elements” of traditional clothing. Fabric carries a rich pattern design, and the pattern design can also be reflected by the fabric design, fabric and pattern complement each other [13]. Therefore, the key information elements to be acquired in apparel pattern are color, style and pattern, which can be used to extract customer preference information to grasp the current demand trend [14, 15]. Traditional clothing pattern design relies on experienced designers using various design software and hand-drawing skills, although the quality of the clothing pattern is high, but it takes a lot of design time and the conversion rate is low [16, 17]. The digital era requires pattern design to be popularized, efficient and diversified, and how to obtain apparel patterns with high efficiency and low cost to provide inspiration for pattern design has become a bottleneck that needs to be broken through by major enterprises [18].

Apparel pattern generation drives the digitalization process of the apparel industry, and countries are now encouraging the development of a new generation of fashionable, functional, and intelligent apparel products, and the intelligent design and generation of patterns based on artificial intelligence becomes one of the important ways to develop such products [19-21]. Jadhav et al. provided an in-depth analysis of Fashion Fusion with respect to apparel pattern generation and showed that AI technology can automatically generate apparel patterns with both creativity and individuality using deep learning and efficient innovation patterns, which is the current reference answer to drive apparel pattern generation [22]. Wang et al. launched an innovative exploration of the modernization of Chinese traditional clothing embroidery patterns, with the help of digital media technology to generate a more visually impactful traditional clothing patterns by integrating traditional patterns with modern elements, which improves the aesthetic value of clothing and effectively spreads the Chinese traditional clothing culture [23]. Peng constructed a clothing art pattern generation model based on style transfer, and the loss function value of the method used is always lower than 0.5 and the search time is only 0.28s, which has the advantages of efficient, low-cost, and personalized clothing pattern generation, and promotes the intelligent development of the clothing design industry [24]. Ye et al. generated 10 classes of human body models through the k-means clustering algorithm, and then used the batch personalized clothing pattern generation model based on biarc and ezdxf to generate personalized clothing patterns for all people of different clusters, which not only has a short generation time but also has the advantage of the pattern fitting perfectly with the human body [25].

Machine learning technology provides new ideas and references for the digital development

of the apparel industry and the intelligent design of apparel patterns [26]. Guo et al. attempted to use diffusion modeling and 3D reconstruction techniques for apparel pattern generation, where a strong diffusion model generates high-quality 2D apparel images, and then the generated apparel images are reconstructed in 3D and transformed into the corresponding apparel patterns [27]. Zhong conducted a research on clothing pattern personalization and automated generation, firstly, the key features of human body data are obtained by multi-factor LASSO regression, then the clothing feature points are extracted by spline function, and finally the personalized clothing pattern is automatically generated by the neural network fusing the two features [28]. Generative adversarial network is one of the most prominent generative methods in machine learning technology, and it has great potential for application in the field of clothing pattern generation [29]. Araújo et al. pointed out that generative adversarial networks can provide new technical support for clothing pattern design, in which the generative adversarial network consists of a generator and a discriminator, which collaborate together to accomplish the task of clothing pattern generation, and the model training process has good loss stability [30]. Tango et al. proposed an image-to-image model for automatic costume pattern generation based on generative adversarial networks and applied it to the field of cosplay costumes, while combining auxiliary techniques to connect generative adversarial networks with cosplay costume pattern features to improve the pattern generation performance of the model [31]. Wu et al. proposed a clothing pattern generation framework based on generative adversarial network and applied it to the generation of Dunhuang style patterns, which realized the automatic generation of Dunhuang style fashion clothing, and the initial score, human preference score and generation score of the generated patterns were unanimously recognized by researchers [32]. Han et al. utilized the digital intelligence technology for traditional clothing culture inheritance and protection, and the core technology is the generative adversarial network, which can identify and generate clothing patterns in response to the number of traditional clothing pattern samples, scale size, and cultural value, and confirmed the practicality of the model in pattern design applications [33]. Tirtawan et al. proposed the use of generative adversarial network derived model to generate batik patterned garments, which enriches the patterned elements in smart clothing design, the method used in the experiment scored 80.20% in F1 and the generated garment patterned textures achieved 99.992% cosine similarity [34]. Cai et al. collected clothing pattern images and constructed a style dataset, and realized clothing pattern style migration through conditional generative adversarial network and convolutional neural network, and the generated clothing patterns had better performance in color distribution and fabric texture [35]. Fan improved the deep convolutional generative adversarial network in order to realize intelligent clothing pattern design, introduced a new network structure and loss function, and embedded real-time style conversion technology on this basis to realize high-quality generation of clothing patterns [36].

This paper proposes a design process for generating clothing patterns based on understanding the variation and composition of clothing patterns. The stable diffusion model is selected as the model basis, and a method for generating clothing image samples based on diffusion inversion is designed. CLIP text encoder is utilized to convert text information in text labels. For the forward propagation process, the diffusion process of repeatedly adding noise to the implicit representation of the image is performed, and UNet is used for the estimation of noise. The experimental environment is set up, and the trained models are compared for text consistency, image consistency, FID score, IS score, and VSS index performance. For the apparel pattern samples formed by the apparel image sample generation method using diffusion inversion, multi-dimensional scoring is performed for the amateur group and expert group. Based on the scores, the feasibility of apparel image sample generation based on diffusion inversion is analyzed.

# 2 Key technologies

## 2.1 Diffusion and Potential Diffusion Models

The overall structure of the Noise Diffusion Probabilistic Model (DDPM) is based on a Markov chain divided into a forward diffusion process and a backward noise reduction process. The forward diffusion process adds Gaussian noise to the real data distribution $x_0 \sim q(x)$ after T times of sampling given the real data distribution 1, and obtains a series of noisy pictures $x_1, x_2, \ldots, x_T$. The modeling of the forward process based on the structure of the Markov chain can be expressed as follows:

$$q\left(x_t \mid x_{t-1}\right)=N\left(x_t ; \sqrt{1-\beta_t} x_{t-1}, \beta_t I\right), q\left(x_{1: T} \mid x_0\right)=\prod_{t=1}^{T} q\left(x_t \mid x_{t-1}\right) \tag{1}$$

where the Gaussian distribution variance is the constant $\left\{\beta_t \in(0,1)\right\}, t \in T$ given in advance. The above model indicates that when $x_0$ is known, for any $x_t$ on the Markov chain conforms to a Gaussian distribution with mean $\sqrt{1-\beta_t} x_{t-1}$ and variance $\beta_t$, then it is shown that via Eq. (1) can obtain a unified expression for $x_t$ with respect to $x_0$. Define the model to have $\alpha_t=1-\beta_t, \bar{\alpha}_t=\prod_{i=1}^{T} \alpha_i$ that randomly sampling noise z from a standard Gaussian distribution $\mathrm{N}(0,1)$ gives an expression for each $x_t$ about $x_0$ in the middle of the Markov chain, i.e., any $x_t$ on the Markov chain can be represented by $x_0$ by recursion, and the expression for $x_t$ under $x_0$ conditions is:

$$\begin{aligned} x_t &= \sqrt{\alpha_t} x_{t-1}+\sqrt{1-\alpha_t} z_{t-1} \\ &= \sqrt{\alpha_t \alpha_{t-1}} x_{t-2}+\sqrt{1-\alpha_t \alpha_{t-1}} \overline{z}_{t-2} \\ &= \ldots \\ &= \sqrt{\overline{\alpha_t}} x_0+\sqrt{1-\overline{\alpha_t}} z \end{aligned} \tag{2}$$

The $x_0$ is sampled after T times, and theoretically the $x_T$ obtained when $T \rightarrow \infty$ is a Gaussian noisy graph that does not contain any $x_0$ information.

The reverse denoising process is the opposite of the forward process, which aims at continuously denoising $x_t$ after T samples to obtain a noise-free original graph $x_0$ with complete semantic information. From the derivation of the forward diffusion process, it can be obtained that according to the Gaussian distribution theorem, if $q\left(x_t \mid x_{t-1}\right)$ satisfies the Gaussian distribution and $\beta_t$ is sufficiently small, then $q\left(x_{t-1} \mid x_t\right)$ is still Gaussian. So the relationship between $x_{t-1}$ and $x_t$ can be modeled:

$$p_\theta\left(x_{t-1} \mid x_t\right)=N\left(x_{t-1} ; \mu_\theta\left(x_t, t\right), \Sigma_\theta\left(x_t, t\right)\right) \tag{3}$$

The model training process $x_0$ is known, then it follows from Bayes' theorem:

$$
\begin{aligned}
q\left(x_{t-1} \mid x_t, x_0\right) &= q\left(x_t \mid x_{t-1}, x_0\right) \frac{q\left(x_{t-1} \mid x_0\right)}{q\left(x_t \mid x_0\right)} \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{\left(x_t-\sqrt{\alpha_t} x_{t-1}\right)^2}{\beta_t}+\frac{\left(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}} x_0\right)^2}{1-\bar{\alpha}_{t-1}}-\frac{\left(x_t-\sqrt{\bar{\alpha}_t} x_0\right)^2}{1-\bar{\alpha}_t}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t} x_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}} x_0\right) x_{t-1}+C\left(x_t, x_0\right)\right)\right)
\end{aligned}
\tag{4}
$$

Converted to the general form of the Gaussian distribution, then the variance and mean can be parameterized as:

$$
\begin{aligned}
q\left(x_{t-1} \mid x_t, x_0\right) &= N\left(x_{t-1} ; \tilde{\mu}\left(x_t, x_0\right), \tilde{\beta}_t I\right) \\
\tilde{\beta}_t &= 1 / \left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)=1 / \left(\frac{\alpha_t-\bar{\alpha}_t+\beta_t}{\beta_t\left(1-\bar{\alpha}_{t-1}\right)}\right)=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t \\
\tilde{\mu}_t\left(x_t, x_0\right) &= \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} x_t+\frac{\sqrt{\bar{\alpha}_{t-1}} \beta_t}{1-\bar{\alpha}_t} x_0=\frac{1}{\sqrt{\alpha_t}}\left(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_t\right)
\end{aligned}
\tag{5}
$$

Because of the ability to parameterize the variance and the mean, a predictive neural network can be constructed to make predictions about the noise at time t:

$$
\mu_\theta\left(x_t, t\right)=\frac{1}{\sqrt{\alpha_t}}\left(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta\left(x_t, t\right)\right) \tag{6}
$$

The objective function for DDPM training can be obtained as:

$$
L=\mathrm{E}_{x_0, z_t}\left[\left\|z_t-z_\theta\left(\sqrt{\bar{\alpha}_t} x_0+\sqrt{1-\bar{\alpha}_t} z_t, t\right)\right\|^2\right] \tag{7}
$$

A UNet network with an attention mechanism is used in DDPM to make predictions on $z_t$ to obtain $x_{t-1}$.

Although DDPM is easier to train compared to GAN, DDPM still requires a large amount of data for fitting. This is because mapping from Gaussian space to natural image domain is a complex mapping relationship. The Latent Diffusion Model (LDM) was proposed to improve the problem that diffusion models consume too much computer resources. In order to improve the performance of the model and reduce the resource consumption during training. LDM divides the generation model into two steps, Encoder-Decoder and DDPM, and the overall structure of LDM is shown in Fig. 1.

Encoder is responsible for encoding the image, Decoder reduces the encoding to a natural image, and the encoding is generated by the DDPM part. And in the process of generating the image encoding, conditional information can be added to control the generated content. Unlike DDPM, LDM first uses the encoder to compress the picture into a tensor $z$ in the forward process, and then gradually adds noise to Gaussian noise. In the reverse process, the tensor $z$

is generated by gradually introducing conditional information, or even bootstrapping information from other modalities. The decoder is then used to restore the picture $X'$. Compared with DDPM, the Encoder-Decoder part requires pre-training to encode the original picture into the hidden space, while this approach reduces the computational cost of DDPM, improves the resolution and clarity of the generated picture, and can add various information to guide the picture generation.

Due to the excellent performance of LDM, this paper uses LDM as a generative model to expand the dataset, which is used to further enhance the generalization ability of the model as well as improve the stability of the model.

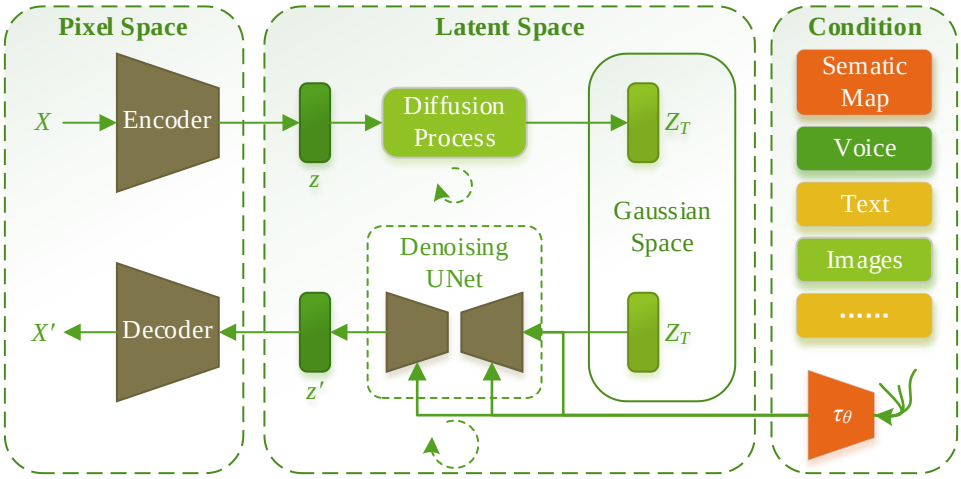


*Figure 1: LDM structure*

## 2.2 CLIP Network

Contrastive Text Image Pretraining Model (CLIP) stands for "Contrastive Language Image Pretraining," a deep learning model developed by OpenAI to improve the ability of AI systems to understand and process visual and textual information. The model is trained to recognize the relationship between natural language inputs and their corresponding visual representations, which enables the system to perform tasks such as image captioning and visual quizzing with greater accuracy and efficiency.CLIP represents a significant advancement in the field of artificial intelligence and deep learning, and has the potential to enable a wide range of new applications and services.

The following is an overview of the CLIP model structure:

Input coding: the CLIP network accepts two inputs: natural language queries and images. Natural language queries are encoded using a Transformer-based architecture, which is used to extract semantic and contextual information from text.

Visual coder: the images are first preprocessed using a set of standard image transforms. The preprocessed image is then passed through a convolutional neural network to extract high-level visual features.

Attention Mechanism: the encoded natural language query is combined with the encoded image features using a multi-head attention mechanism. This allows the network to selectively focus on different parts of the input data to recognize important relationships between text and images.

Contrast Loss: The CLIP network is trained using a contrast loss function that constrains the network to correctly associate a given image with its corresponding natural language query. It also minimizes the similarity between the image and any irrelevant natural language query.

This helps the network to learn meaningful and semantically rich representations of images and natural language queries.

The CLIP network is pre-trained on a large corpus of images and text, which enables it to learn to recognize and identify various objects, concepts and relationships between text and visual input. The network can then be fine-tuned for specific downstream tasks, such as image categorization or image style migration.The CLIP architecture has been shown to achieve state-of-the-art performance on a variety of natural language and visual tasks, making it a powerful tool for a wide range of applications.

## 2.3 UNet Neural Network

(1) U-Net Neural Network

U-Net neural network has consists of two main modules, an encoder which is responsible for extracting the features and a decoder which is responsible for outputting the results. This network also contains jump connections which retain more spatial information and details for the network and loss function which is responsible for measuring the gap between the predicted results and the labeling, below is the detailed description of each part.

Encoder: The left part of the U-Net is the encoder, which is responsible for extracting features from the input image. It usually includes multiple convolutional and pooling layers that gradually reduce the spatial resolution of the image while increasing the number of feature channels. This helps in capturing features of different scales and complexity in the image.

Decoder: The right part of the U-Net is the decoder, which is responsible for reducing the feature mapping extracted by the encoder to a segmentation result with the same resolution as the input image. The decoder usually consists of a convolutional transposition layer and a jump join. The jump connection connects the features of the encoder to the features of the decoder to help the network retain more spatial information and details.

Intermediate connection: there is an intermediate connection between the encoder and the decoder that is used to pass the encoder's features to the decoder. These features are used to help the decoder to restore the segmented image. Output Layer: The output layer of U-Net is usually a convolutional layer that matches the number of desired segmentation categories, using a softmax activation function to generate the probability that each pixel belongs to each category. Loss function: U-Net typically uses a cross-entropy loss function to measure the gap between the model's prediction and the true segmentation.

(2) U-Net Neural Network with SE Attention Mechanism Added

SE attention mechanism is a mechanism used to enhance the performance of convolutional neural networks (CNNs), which learns to assign different attention weights on specific channels so that the network can better capture important features.The flowchart of the SE module is shown in Fig. 2.The SE attention mechanism has been widely used in deep learning tasks.

The main idea of the SE attention mechanism is to reweight the feature map inside each channel. Specifically, the SE attention mechanism consists of the following two main steps:

Compression: in this step, a global average pooling operation is performed on the feature maps within each channel. This aggregates the information from each channel into a single value.

Elicitation: in this step, a small feed-forward neural network is used to learn the attentional weights assigned on each channel. This small network usually consists of a fully connected layer (dimensionality reduction) followed by a ReLU activation function and then another fully connected layer (dimensionality enhancement). Finally, the output is restricted to between 0 and 1 by a Sigmoid activation function to generate the attentional weights for each channel. Finally, the learned attentional weights are multiplied by the input feature map to obtain a weighted feature map.

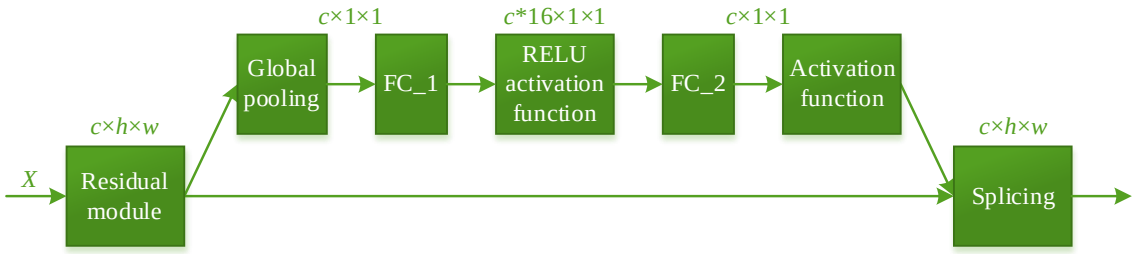


*Figure 2: SE module flow chart*

# 3 Garment pattern generation and design optimization methods

## 3.1 Design Process of Garment Patterns

Shape Grammar, also known as Shape Grammar, Shape Grammar is a design-oriented generative system, a rule-based system for describing and generating designs. The basic process of applying shape grammars is: identify basic shapes, identify spatial relationships, identify rules, identify shape language, and apply and design.

In the field of pattern design, it is crucial to understand the principles of pattern composition. Through the theory of Shape Grammar, it is found that individual patterns can be generated repeatedly through the transformation of graphic elements, while continuous patterns can be made through the skillful transformation and combination of individual patterns.

This observation is an important inspiration for the practice of pattern design. Firstly, by transforming the graphical elements, designers can easily create individual patterns with regularity and repetition. Secondly, for continuous patterns that require more complexity and variety, designers can realize the desired effects by making fine transformations and clever combinations of independent patterns. Most importantly, the patterns found to be suitable can usually be derived from the transformations of the three aforementioned basic patterns, which provides designers with more creative and design options.

The variations and compositions of garment patterns are shown in Figure 3.

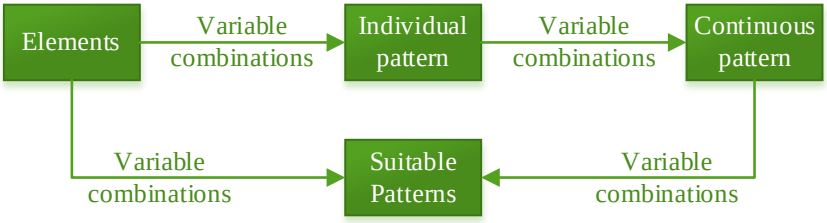


*Figure 3: Changes and composition of garment patterns*

The design process for generating garment patterns is as follows:

(1) Based on the textual requirements, the potential diffusion model is utilized to generate simple graphical elements and complex patterns of garment patterns.

(2) Based on the shape grammar, the generated pairs of garment pattern elements are varied to realize the redesign of patterns.

(3) Apply the generated garment patterns to the appearance design of simulation fabrics to realize the design of garment patterns based on simulation fabrics.

## 3.2 Generation of clothing image samples based on diffusion inversion

The stable diffusion model belongs to a kind of latent diffusion model (LDM). The stabilized diffusion model introduces a self-encoder based on the diffusion model in projecting the image into the latent space and performing two diffusion processes in that space. This modus operandi allows it to be trained on limited computational resources and can strike a good balance between preserving detail and reducing complexity.

The stabilized diffusion model is able to combine conditional information from class labels, text labels and other types of denoising objective function:

$$L_{LDM} := \mathbb{E}_{\mathbf{z}\sim\mathcal{E}(\mathbf{x}),y,\epsilon\mathcal{N}(0,1),t}\left[\left\|\boldsymbol{\epsilon}-\epsilon_\theta\left(\mathbf{z}_t,t,c_\theta(y)\right)\right\|_2^2\right] \tag{8}$$

where $\boldsymbol{\epsilon}$ denotes a noisy sample, $\epsilon_\theta$ denotes a denoising network, $c_\theta(y)$ is a text encoder model that maps the conditional input $y$ to the conditional encoding, $t$ is the time step, $\mathrm{z}_t$ is the potential encoding of the image at time $t$, and $\mathcal{E}(\cdot)$ is the encoder. In the inference phase, the potential coding $\mathbf{z}_0$ of the new image is generated by iteratively denoising the conditional vector and the random noise vector. The latent coding $z_0$ is decoded by a pre-trained self-encoder using a stable diffusion model.

In order to alleviate the constraints of the textual morphology, the method in this paper first employs a diffusion inversion method to train a conditional mapping network whose purpose is to invert the image into its generative conditional encoding in the diffusion model. The text encoder ($c_\theta$) in the stabilized diffusion model is considered as a constant mapping, and the conditional mapping network is used to project the image into the conditional coding space. The optimal conditional encoding corresponding to each image is found by minimizing the $L_{LDM}$ loss. The above equation can thus be transformed into an optimization problem for the conditional coding $c$:

$$\mathbf{c}_* = \arg\min_{\mathbf{c}} \mathbb{E}_{\mathbf{z}\sim\mathcal{E}(\mathbf{x}),y,\epsilon\mathcal{N}(0,1),t}\left[\left\|\boldsymbol{\epsilon}-\epsilon_\theta\left(\mathbf{z}_t,t,\mathbf{c}\right)\right\|_2^2\right] \tag{9}$$

where each obtained vector $\mathbf{c}_*$ is regarded as the unique conditional vector of the corresponding image.

Inspired by the sample-guided and appearance-guided feature imagination methods, a transform-guided conditional generation network based on transform-guidance is proposed to fully exploit the relocatable information in the underlying dataset and apply it to small sample categories. The condition vectors generated by the network are represented as follows:

$$\mathbf{c}_{i,j} = G\left(\mathbf{r}_i,\mathbf{a}_j\right) \tag{10}$$

$\mathbf{r}_i$ is the category-related condition for category $i$, and $\mathbf{a}_j$ is the transformed related condition for the $j$ th sample in the other categories. Since the category sample mean removes the differences between the samples in the category and retains the common parts, the category-related condition for the category can be constructed by calculating the mean of the conditional distribution of the category in the latent space:

$$\mathbf{r}_i = \frac{1}{|S_i|}\sum_{\mathbf{c}_k \in \mathcal{S}_i} E(\mathbf{c}_k) \tag{11}$$

$\mathcal{S}_i$ is the set of condition vectors corresponding to the samples of the category $i$. The methods in this chapter use the difference vectors of the samples to the category prototypes as an intraclass transformation, and the condition vectors $\mathbf{c}_k$ in the category $i$ with respect to the transformation can be expressed as follows:

$$\mathbf{a}_k = \mathbf{c}_k - \mathbf{r}_i \tag{12}$$

Construct a $N$-way $K$-shot task dataset by sampling from the dataset. First sample $N$ categories from the base dataset and sample $K$ samples for each category to obtain a support set for the task $S_{sup} = \left\{\left\{\left(\mathbf{c}_{n,k}, y_n'\right)\right\}_{k=1}^{K}\right\}_{n=1}^{N}$, where $c_{n,k}$ denotes the $k$th conditional vector of the $i$th category, $y_n'$ is the $N$-way label, and $y_n' \in \{1,2,\ldots,N\}$. To generate the new conditional vectors, a reference set $\mathcal{S}_{ref}$ is formed by sampling $N$ base categories and a sample from each category. Finally, using Equation (10), $M$ condition vectors are generated for each category in $\mathcal{S}_{ref}$ to generate the corresponding images for the subsequent model. Where the generated condition vectors are:

$$\mathbf{c}_{n,k} = G\left(\mathbf{r}_n, \mathbf{a}_r\right) \tag{13}$$

Reconstruct the condition vector for the category-related and transformation-related conditions from the same sample $c_{n,k} = G\left(\mathbf{r}_n, \mathbf{a}_{n,k}\right)$. Since the reconstructed condition vector should be similar to the original condition vector, the constraints can be expressed as:

$$L_c = \min_{\mathbf{c}_{n,k}} \mathbb{E}\left[\left\|\mathbf{c}_{n,k} - \mathbf{c}_{i,k}\right\|\right] \tag{14}$$

The prototype of the category to which the small-sample correlation condition belongs is updated using an additional condition vector, which can be expressed as:

$$\mathbf{r}_n' = \frac{K\mathbf{c}_{n,k} + \sum_{k=1}^{M}\mathbf{c}_{n,k}}{K+M} \tag{15}$$

When transform-related conditions are obtained from other categories, the generated condition vectors should belong to the category to which the category-related conditions belong. To preserve the category information of the generated vectors, a meta-classifier is jointly trained. Since the meta-learning classification loss encourages the generation conditions that improve the classification performance, the constraint can be achieved by optimizing the classification loss. For the query set $\mathcal{S}_{query}$, the classification loss can be expressed as:

$$L_m = \min \mathbb{E}_{\left(\mathbf{c}_q, y_q'\right)\in \mathcal{S}_{query}}\left[-\sum_{n=1}^{N}\mathbb{I}\left(y_q' = n\right)\log P\left(n \mid \mathbf{c}_q\right)\right] \tag{16}$$

where $\mathbb{I}(true)=1$. The probability that the conditional vector $\mathbf{c}_q$ is predicted by the model to be in category $n$ can be expressed as:

$$P\left(n \mid \mathbf{c}_q\right)=\frac{\exp-\left\|\mathbf{c}_q-\mathbf{r}_n'\right\|}{\sum_i \exp-\left\|\mathbf{c}_q-\mathbf{r}_i'\right\|} \tag{17}$$

To ensure that the generated conditions can cover more sample distributions, the method in this paper introduces consistency loss. On the one hand, the transform-related conditions used by a condition vector should be consistent with the transform-related conditions separated from that generated condition vector. This constraint can be expressed as:

$$L_{con}=\min \mathbb{E}\left[\left\|\left(\mathbf{c}_{n,k}-\mathbf{r}_n'\right)-\mathbf{a}_{r,k}\right\|\right] \tag{18}$$

On the other hand, the condition vector generated using the isolated transformation-related condition and the category-related condition of that condition should be consistent with the original condition. This constraint can be expressed as:

$$L_{recon}=\min \mathbb{E}\left[\left\|G\left(\mathbf{r}_r,\left(\mathbf{c}_{n,k}-\mathbf{c}_n'\right)\right)-\mathbf{a}_{r,k}\right\|\right] \tag{19}$$

Summarizing the above constraints, the loss function of the proposed embedding generation model can be expressed as follows:

$$L=\lambda_m L_m+\lambda_c L_c+\lambda_{con} L_{con}+\lambda_{recon} L_{recon} \tag{20}$$

## 3.3 Model Training

The training method of the model in this paper is shown in Fig. 4. The diffusion inversion based model training method for garment image sample generation follows the following steps:

Step1: Images are mapped from pixel space to latent space using pre-trained AutoEncoderKL self-encoder to learn the implicit representation of the image.

Step2: Text labels are encoded by the pre-trained CLIP text encoder to generate the embedded representation.

Step3: The diffusion process of adding noise to the implicit representation of the image is performed iteratively during the forward propagation process. For the image after adding noise, the noise is estimated using UNet. It is worth noting that UNet also receives both the implicit representation of the image as well as the text embedding. During the training process, contextual information is introduced as a condition, and the model uses the attention mechanism to better learn the matching relationship between text and image. The trained UNet is able to generate the corresponding garment pattern from the textprompt.

The core of this approach is to combine image and textual information for better image generation. This integration and the use of the attention mechanism help the model to better understand the connection between the textual description and the image content, thus generating images that better match the textual description.

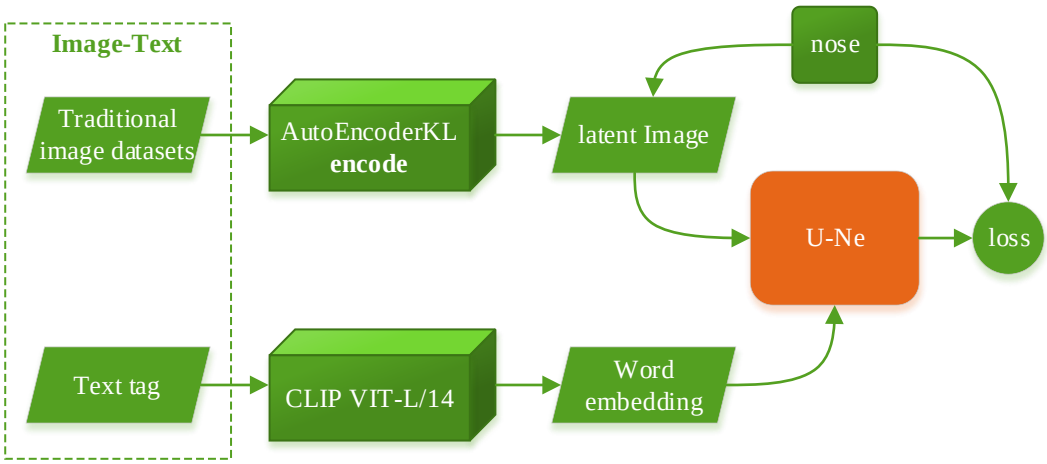


*Figure 4: The training method of this article model*

# 4 Garment pattern generation and design effects

## 4.1 Experimental design

Experiments were conducted on the MSCOCO and CUB-200 datasets, and the specifics of the datasets are shown in Table 1.

*Table 1: Details of the data set*

| Data set | The number of images of the training set | The number of images of the test set | Text description/image | Categories |
|---|---|---|---|---|
| MSCOCO | 8000 | 30000 | 6 | 60 |
| CUB-200 | 9000 | 3000 | 12 | 180 |

In this paper, experiments were conducted using Python language in Pytorch platform and NVIDIA A100 GPU was used to train and test the network. During the training process, the momentum-based Adam optimizer is used to train the model of this paper, setting $\beta_1$=0.3 and $\beta_2$ =0.6, respectively. The initial learning rate of the generator is $10^{-4}$, and that of the discriminator is $3\times10^{-4}$, and an exponential decay function with a rate of 0.999 is used to adjust the learning rate, with a text-image data pair batch size of 64 and an iteration number of 10,000.

In this paper, FID score, Initial Score (IS) is used to measure the performance of the model in this paper.FID score is a metric used to calculate the distance between the features of the generated image and the features of the real image.Lower value of FID means that the features of the two are closer to each other, which indicates that the generated image is closer to the real image i.e., the generated image is more vivid and graphic.

The features of the generated image and the real image are extracted by CLIP image encoder and the FID score is derived based on their feature mean and feature covariance, which is calculated by the formula:

$$\mathrm{FID}=\left\|\mu_r-\mu_g\right\|+\mathrm{Tr}\left(\left(\sum_r+\sum_g-2\left(\sum_r\sum_g\right)^2\right)^{\frac{1}{2}}\right) \tag{21}$$

where $\mu_r, \mu_*, \sum_r, \sum_g$ represent the real image mean, the generated image mean, the real image feature covariance, and the generated image feature covariance, respectively. IS is an important index for evaluating the performance of GAN network, and IS is calculated by the formula:

$$\mathrm{IS} = \mathrm{e}^{\mathbb{E}x\sim p_g D_{\mathrm{KL}}(p(d|x)\| p(y))} \tag{22}$$

where $p(d \mid x), p(y)$ denote the conditional probability and edge probability of the label $y$ predicted by the pre-trained image encoder model, respectively. If the conditional probability is lower and the edge probability is higher, the greater the KL scattering is, which represents the higher quality of the image. Therefore, a larger IS value indicates that the generated image is of higher quality with richer details and diversity.

Visual-Semantic Similarity (VSS) is used to measure the degree of match between textual description and image semantics, and the larger the result of its computation indicates better semantic consistency. The overall idea of the method is to train a visual-semantic embedding model to measure the semantic match between the generated image and the text description, which is defined as follows:

$$VSS = \frac{f_t(t) \bullet f_I(I)}{\left\| f_t(t) \right\|_2 \bullet \left\| f_I(I) \right\|_2} \tag{23}$$

where $f_t(\bullet)$ and $f_t(\bullet)$ denote the text encoder and the image encoder, respectively, which are trained and supplied by the HDGAN model.The VSS uses the two encoders to map the textual descriptions and the generated images into a common semantic space, which leads to semantic coherence matching estimation.

## 4.2 Results of the evaluation of indicators

### 4.2.1 Image generation similarity analysis

The image generation model that renders the subject content of an image is selected for performance comparison, and the comparison model is as follows:

Textual Inversion renders the subject in the image creatively according to the text without changing the basic attributes of the image subject, firstly learns new concepts through the text encoder in the hidden vector space, and then realizes the fine control of the image according to the specific concepts contained in the text.

DreamBooth model: the core idea lies in associating the subjects in the image with the corresponding logos by means of a pre-trained text generation model and mapping them to the output domain.

DreamArtist model: Based on the text encoder and denoising network to learn the expressed hidden vectors from both sides, i.e., the learning strategy is utilized to balance the feature retention of the reference image and the controllability of the generation, in order to improve the quality of details and diversity of the final generated image.

Custom Diffusion model: accomplishes the task of image editing based on a given sample and text by fine-tuning the key- and value-related parameters in the cross-attention layer of the pre-trained textual image generation model.

ELIFE model: The image encoder in CLIP is used to extract hierarchical features, and then the image is mapped into text vectors and feature vectors using global mapping and local mapping, respectively, and finally these two vectors are introduced into the denoising network

of the diffusion model to generate images.

Cones model: a part of the network parameters in the diffusion model will control the generation of specific objects, these parameters are called conceptual parameters, if these parameters are frozen, the model can generate images of different scenes according to different texts. When connecting networks corresponding to different objects, the Cones model is able to generate images containing multiple objects.

The SVDiff model: introduces a compact and efficient parameter space with 1/2 the number of parameters of DreamBooth.200 In addition, the model employs data augmentation techniques to improve the model's ability to learn multiple contents.

In order to have a clearer understanding of the ability of the above methods to deduce the image subject with the diffusion inversion-based garment image sample generation methods, the correlation of the model's generated content with the text and the original image subject is summarized.

A comparison of the text consistency of the image generation models is shown in Fig. 5, where the text consistency indicates the relevance of the generated content of the model to the given text.

A total of ten image generation results are counted in the figure, and the distribution of the models in the figure can be synthesized to show that the text consistency of the generated content of this paper's method is better than that of the other compared methods.

The following is the ranking according to the text consistency results of each model, and the mean values of the ten text consistency results of each comparison model are as follows: this paper's method (0.3712) > SVDiff model (0.356) > DreamArtist model (0.3165) > ELIFE model (0.2939) > Cones model (0.2596) > DreamBooth model (0.2595) > Custom Diffusion model (0.2515) > Textual Inversion model (0.1881).

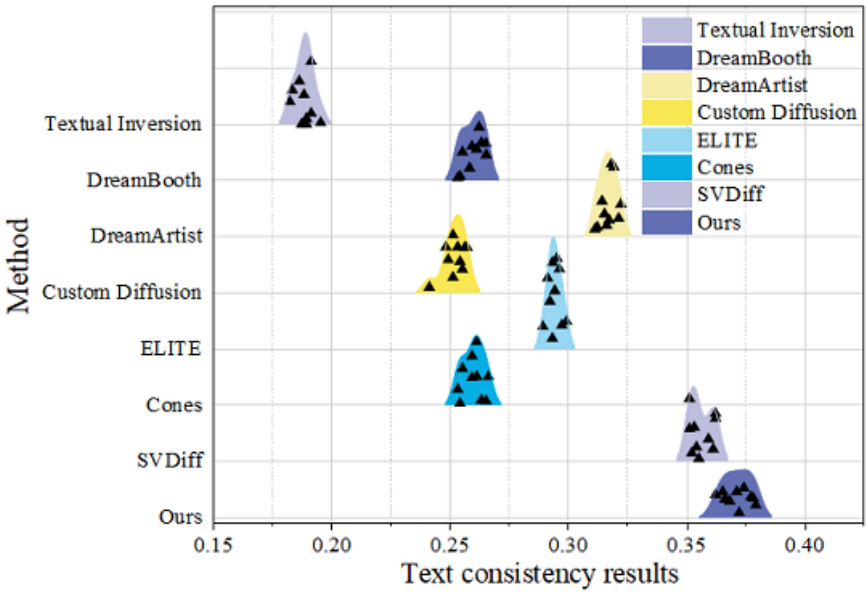


*Figure 5: The text consistency of the image generation model is compared*

Comparison of image consistency of each image generation model is shown in Fig. 6, image consistency indicates the degree of similarity between the generated content of the model and the content of the original image.

The image consistency result between the generated content and the original image of this paper's method is 0.9264 under the multiple testing results, followed by Custom Diffusion model with the average value of image consistency of 0.9175 under multiple testing. The Cones

model with the average value of 0.9041 is in the third place.

The Textual Inversion model with a mean value of 0.7261 for image consistency performed the worst among all models. However, the image generation consistency of all the compared models is more than 0.5%.

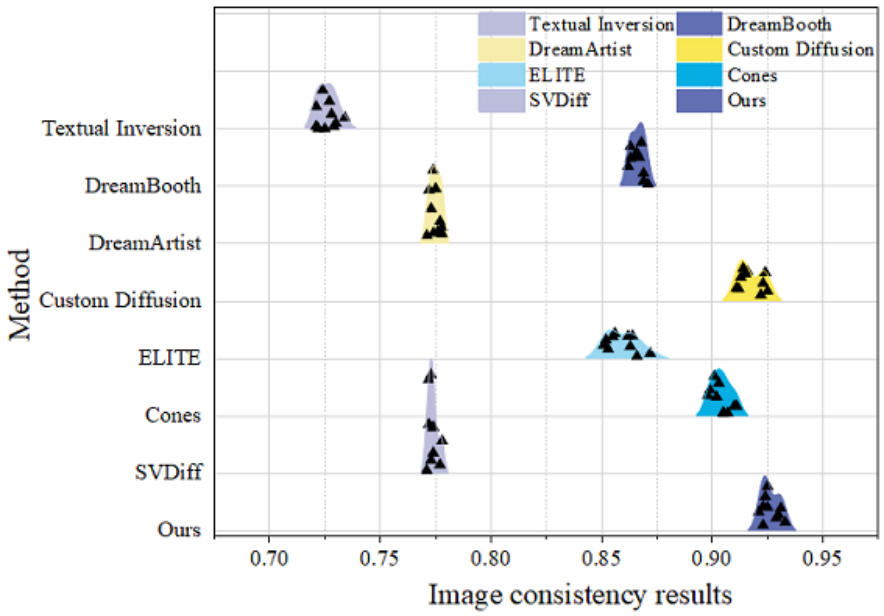


*Figure 6: Image consistency comparison of each image generation model*

### 4.2.2 Comparison of FID scores

In this paper, 10,000 images were randomly generated in the test set of MSCOCO and CUB-200 datasets to compute the FID scores and compared with the state-of-the-art of current GANs such as Stacked Generative Adversarial Networks (StackGAN), Efficient Neural Networks-based Text Generative Image Model (EFF-T2I), Attention Generative Adversarial Networks (AttnGAN), Dynamic Memory Generative Adversarial Network (DM-GAN), Deep Fusion Generative Adversarial Network (DF-GAN), Visual Linguistic Matching Score based Generative Adversarial Network (VLMGAN), Semantic Spatial Sensing based Generative Adversarial Network (SSSA-GAN) for quantitative comparison.

The comparison of FID scores of different models in the two datasets is shown in Figure 7.

The FID scores of this paper's method on the MSCOCO and CUB-200 datasets are 11.76 and 18.48, respectively.

On the MSCOCO dataset, the FID scores of this paper's model decreased by 41.52%, 44.97%, 49.42%, and 52.87% compared to the VLMGAN model (20.11), DM-GAN model (21.37), SSA-GAN model (23.25), and AttnGAN model (24.95) respectively.

On the CUB-200 dataset, the FID score of this paper's model is lower than most of the models, and it is 3.88, or 17.35%, lower than the state-of-the-art EEF-T2I model. It shows that the generated images of this paper's model on different datasets are closer to the real images, which improves the fidelity of the generated images.

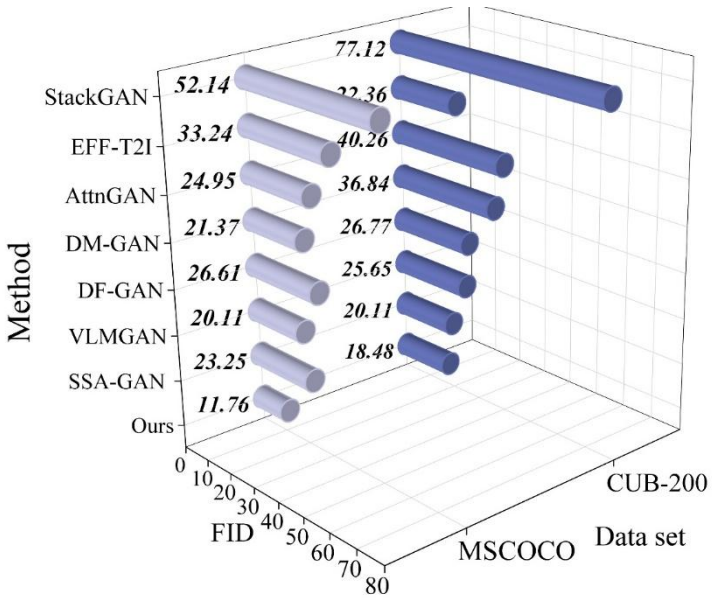


*Figure 7: The comparison of FID fractions of different models in two data sets*

### 4.2.3 Comparison of IS scores

The comparison of IS index scores of different models is shown in Fig. 8.

On the MSCOCO dataset, compared with the classical DM-GAN model, the IS score of the diffusion inversion-based sample generation model for apparel images designed in this paper is improved by 13.40%, indicating that the model in this paper has improved the clarity and diversity of the generated images.

On the CUB-200 dataset, compared with the current SSA-GAN model, which is more advanced in this field, the IS score grows from 6.26 to 8.19, which is an improvement of 30.83%, indicating that the model in this paper generates apparel images with a significant improvement in clarity, more vivid content, and a significant improvement in image quality.

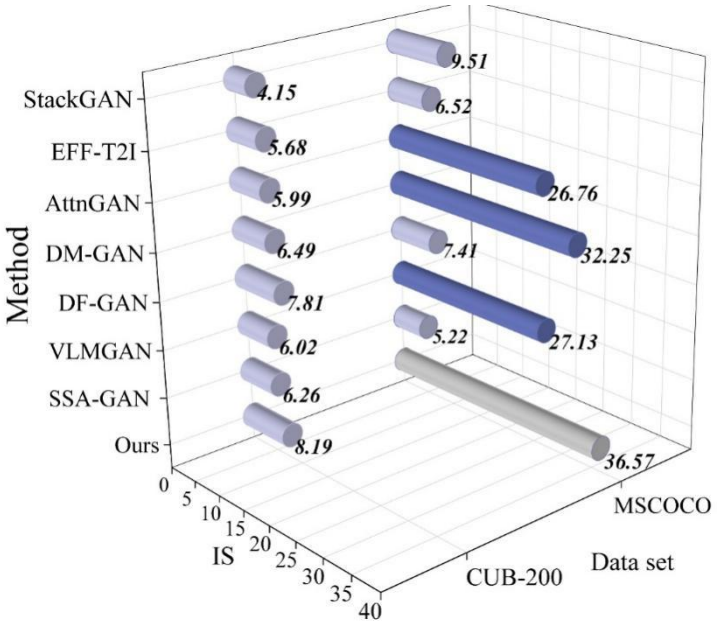


*Figure 8: The IS index fraction of different models is compared*

### 4.2.4 Comparison of VSS indicators

The results of the VSS metrics of this paper's method on the CUB dataset are shown in Figure 9. The output images of all models are used to calculate the visual-semantic similarity with 256×256 resolution images.

Compared with VLMGAN and SSA-GAN models, the network model of this paper is better in terms of visual-semantic consistency. The StackGAN model is 0.235 and also has a large improvement compared with the StackGAN model. All these aspects show that the images generated by this paper's model have a better semantic consistency, and it is proved that the matching-minimized $L_{LDM}$ loss that can effectively improve the semantic consistency of text images.

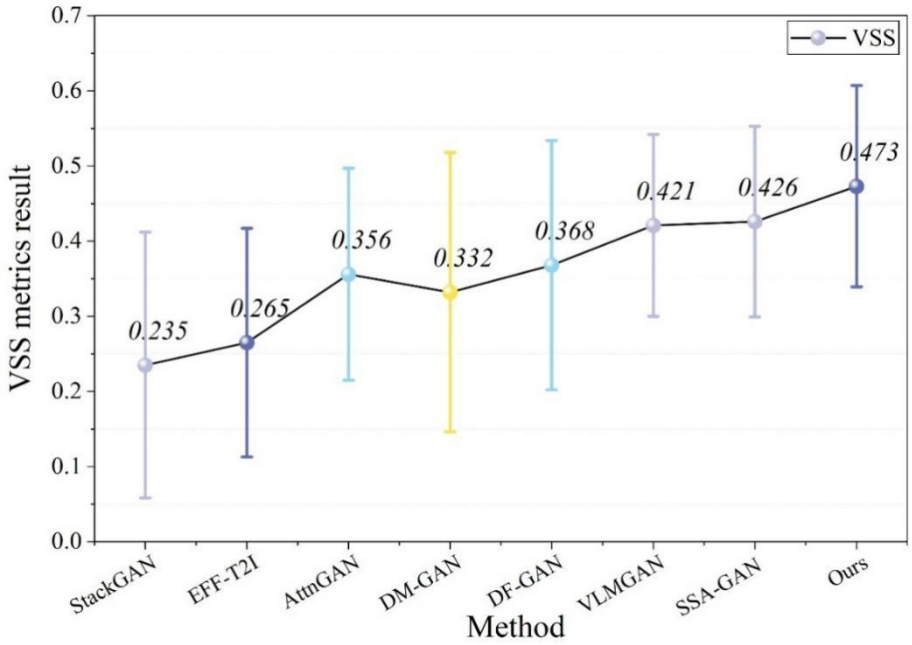


*Figure 9: The results of the VSS index in the CUB data set*

## 4.3 Evaluation of Garment Pattern Generation

In order to further make an accurate evaluation of the clothing patterns generated by the clothing image sample generation model based on diffusion inversion, this paper designs a corresponding subjective evaluation method, which evaluates four dimensions, namely, image quality, overall aesthetics, interpretation of the pattern to the textual content, and the overall coordination between the pattern and the text. The evaluation details are shown in Table 2. 160 generated images were selected for subjective evaluation. A total of 80 observers were invited to score the images to be tested in this paper, divided into an amateur group and an expert group, with the number of amateur group being 60 and the number of expert group being 20.

*Table 2: Evaluation details*

| Dimension | Evaluation scale | Evaluation | Scoring |
|---|---|---|---|
| Image quality evaluation | It can be very good embodiment of costume design style. | Very good | 5 |
| | It is better to embody the style of dress design. | Good | 4 |
| | Clearly the quality of the design falls, the style of the dress is general. | General | 3 |
| | The pattern is poor and the style of the dress is unknown | Difference | 2 |
| | The design quality is extremely poor, do not have the costume design style. | Very bad | 1 |
| Overall aesthetics | The pattern is very beautiful. | Very good | 5 |
| | The whole pattern is better, but it still needs to be promoted. | Good | 4 |
| | The pattern is general, no obvious defect. | General | 3 |
| | The overall aesthetics of the pattern is poor, the existence is obviously insufficient. | Difference | 2 |
| | The overall beauty of the pattern is very low and difficult to accept. | Very bad | 1 |
| The interpretation of the text content | The pattern accurately interprets the text content. | Very good | 5 |
| | The pattern interprets the text content better. | Good | 4 |
| | The pattern has a certain interpretation of the content of the text. | General | 3 |
| | The design of the text is more vague and difficult to express. | Difference | 2 |
| | The pattern is almost impossible to interpret the text and is out of touch with the text. | Very bad | 1 |
| Overall coordination of patterns and text | Patterns and text are perfectly integrated. | Very good | 5 |
| | The pattern and text have better coordination. | Good | 4 |
| | Patterns have a certain correlation with text. | General | 3 |
| | Patterns and text are relatively weak. | Difference | 2 |
| | The pattern is almost irrelevant to the text. | Very bad | 1 |

### 4.3.1 Amateur group

The results of the subjective evaluation of the amateur group are shown in Fig. 10. The average subjective score of the image quality dimension is 4.02, which is more than 4. This indicates that the patterns leave a better impression on the evaluators in terms of quality, and their aesthetics are recognized by most of the evaluators.

The average subjective scores for pattern quality, image aesthetics, the degree of the pattern's interpretation of the textual content, and the overall coordination between the pattern and the text were 4.02, 3.97, 3.62, and 3.66. This indicates that in the eyes of most of the evaluators who were not majoring in apparel arts, the design and quality of the pattern itself was quite excellent and had a high degree of ornamental appeal.

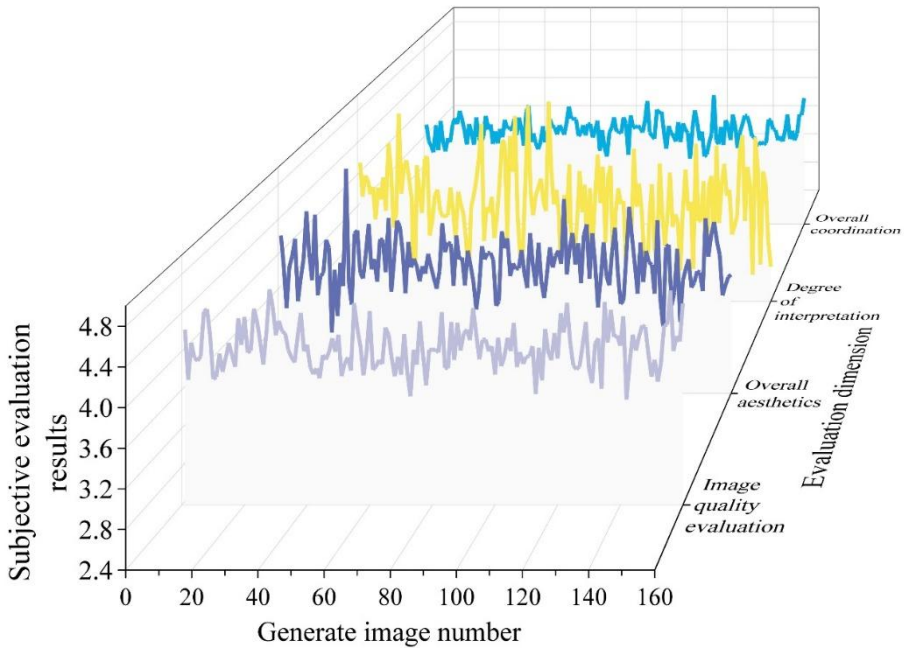


*Figure 10: Subjective evaluation results of the industry*

### 4.3.2 Expert groups

The results of the subjective evaluation by the Expert Group are shown in Figure 11.

In the case of subjective evaluation scores, the average score of pattern quality is 4.12, which indicates that the patterns are widely recognized in terms of features and styles, and that their design details and production level have reached a high standard. The quality evaluation of the model-generated samples by the expert group is higher than that of the amateur group, which indicates that the garment patterns generated using the model of this paper have a certain degree of professionalism. The average score of pattern aesthetics is 3.91, which is slightly lower than the score of pattern quality, but still shows the advantage of patterns in visual presentation. This indicates that the patterns have a certain degree of aesthetics in terms of color matching and composition, which can attract the attention of the evaluators. The score of the pattern's interpretation of the text content is 3.75, which indicates that the pattern is able to reflect or explain the text content to a certain extent, but there is a pattern design that is not deep enough, and the connection between the pattern and the text is not direct or clear enough. The overall coordination between the pattern and the text scored 3.69, and this score reflects that the visual combination of the pattern and the text needs to be improved.

By synthesizing the four dimensions of the expert group's scores, it can be seen that the garment patterns generated by this paper's method are highly evaluated in terms of quality, aesthetics, and overall coordination.

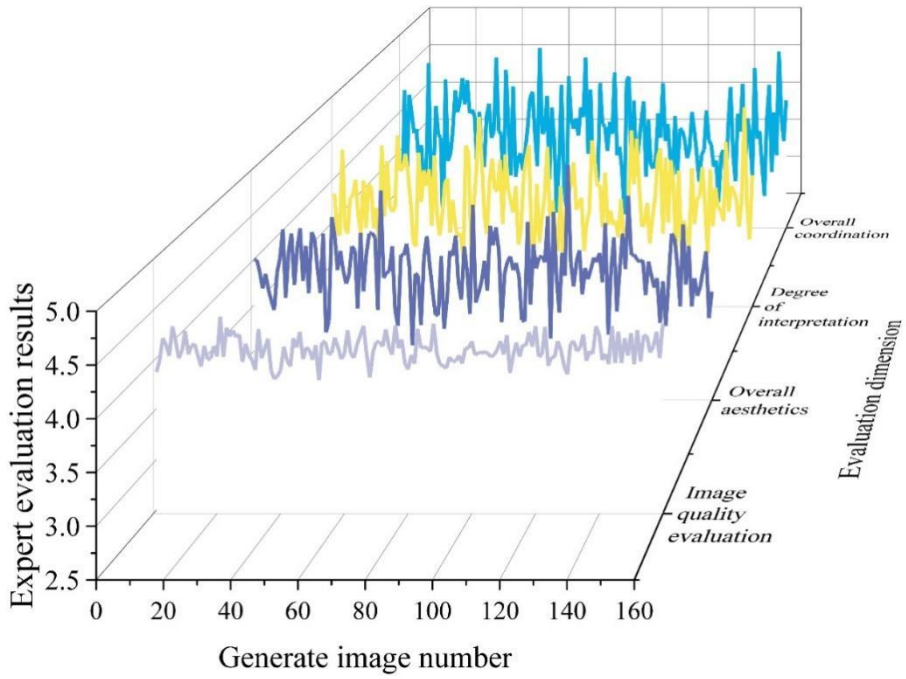


*Figure 11: The results of the panel's subjective evaluation*

# 5 Conclusion

## 5.1 Conclusion

In this paper, the basic steps of garment pattern generation are designed, and the method of garment pattern sample generation based on diffusion inversion is proposed in combination with the stable diffusion model. FID, IS, and VSS indicators are selected to test the garment pattern sample generation method, and the multi-dimensional evaluation of the garment pattern generation sample is carried out. The text consistency and image consistency of the garment pattern samples generated based on the stable diffusion model are 0.3712 and 0.9264, respectively, indicating that the garment pattern sample generation method using diffusion inversion can make the content of the garment pattern generation fit the given text and be similar to the content of the original image.On the MSCOCO and CUB-200 datasets, the FID score of the garment pattern sample generation method based on diffusion inversion is 11.76%. The FID scores are 11.76 and 18.48, and the IS scores are 36.57 and 8.19 for the MSCOCO and CUB-200 datasets. The ratings of the amateur group and the expert group for the model-generated clothing pattern samples in the four dimensions of image quality, overall aesthetics, interpretation of the pattern to the text content, and overall coordination between the pattern and the text are all above 3.5, indicating that the clothing patterns generated by the diffusion inversion-based clothing pattern sample generation method have received a common favorable opinion.

## 5.2 Shortcomings and prospects

This paper achieves the expected goal in the design of clothing pattern generation, but there are still some shortcomings that can be further improved in the subsequent research, mainly including:

(1) The diffusion model requires content-rich training datasets, while many training data are unscreened. In the subsequent analysis, the sample model is generated for the garment

pattern of the stable diffusion model, and the sample dataset shall be preprocessed.

(2) The clothing image sample generation method based on diffusion inversion is capable of stable training and generating high-quality images. However, how to further improve the diversity of generated images on this basis is still a problem to be studied.

(3) Expanding and enriching the existing multimodal dataset of apparel patterns, adding more diverse text descriptions and pattern samples, and improving the generalization ability and applicability range of the diffusion model.

# References

[1] Zhang, X. H., & Xu, X. B. (2011). Study on the Main Factors of Marketable Costume. Advanced Materials Research, 175, 987-992.

[2] Makryniotis, T. (2018). Fashion and costume design in electronic entertainment—bridging the gap between character and fashion design. Fashion Practice, 10(1), 99-118.

[3] Inui, S., Mesuda, Y., & Horiba, Y. (2020). Making a dart for a clothing pattern in virtual space. International Journal of Clothing Science and Technology, 32(4), 589-600.

[4] Al-Majed, R., & Hussain, M. (2024). Entropy-Based Ensemble of Convolutional Neural Networks for Clothes Texture Pattern Recognition. Applied Sciences, 14(22), 10730.

[5] Gao, F., & Ji, D. (2024). Design of metadata scheme for traditional Chinese clothing patterns for cultural digitization. Academic Journal of Humanities & Social Sciences, 7(2), 162-168.

[6] Wang, Z., & Li, L. (2024, November). Exploration of Cultural Products Based on Design Semiotics: A Case Study of Traditional Yugu Clothing Patterns. In 2024 4th International Conference on Public Relations and Social Sciences (ICPRSS 2024) (pp. 252-264). Atlantis Press.

[7] Zhu, Y., Wang, W., & Jiang, C. (2022). The application of clothing patterns based on computer-aided technology in clothing culture teaching. Computer-aided design and applications, 145-155.

[8] Sun, W., & Guo, D. (2022). Design of Three-Dimensional Pleated Clothing Pattern Based on Computer Animation Technology. Mathematical Problems in Engineering, 2022(1), 9907865.

[9] Ji, Z., Huang, W. H., & Lin, M. (2020). Design mode innovation of local color cultures: A case study of the traditional female costume of Yi nationality. Designs, 4(4), 56.

[10] Myagila, K., & Kilavo, H. (2022). A comparative study on performance of SVM and CNN in Tanzania sign language translation using image recognition. Applied Artificial Intelligence, 36(1), 2005297.

[11] Schraml, D. (2019). Physically based synthetic image generation for machine learning: a review of pertinent literature. Photonics and Education in Measurement Science 2019, 11144, 108-120.

[12] Chen, J. C., & Liu, C. F. (2015). Visual-based deep learning for clothing from large database. In Proceedings of the ASE BigData & SocialInformatics 2015 (pp. 1-10).

[13] Choi, S., Kwon, S., Kim, H., Kim, W., Kwon, J. H., Lim, M. S., ... & Choi, K. C. (2017). Highly flexible and efficient fabric-based organic light-emitting devices for clothing-shaped wearable displays. Scientific reports, 7(1), 6424.

[14] Wu, Y. (2017, October). A Method of Pattern Feature Extraction for Clothing Texture. In 2017 International Conference on Robots & Intelligent System (ICRIS) (pp. 8-11). IEEE.

[15] Ng, S. Y., & Mok, P. Y. (2023). An empirical study on consumer preferences for online customised clothing platform. IADIS International Journal on Computer Science and Information Systems, 18(2), 1446-160.

[16] Qi, X., & Li, H. (2025). Digital Protection and Display of Cantonese Embroidery Patterns Based on CAD and Human-Computer Interaction. Theory and Practice of Science and Technology, 6(1), 68-78.

[17] Yaoyuan, G., & Hong, X. (2012, July). Research on parameters reasoning of size for blouse in custonmization system. In 2012 7th International Conference on Computer Science & Education (ICCSE) (pp. 101-104). IEEE.

[18] Zhao, P., Yu, D., & Liu, Y. (2024, August). Design of Clothing Pattern Digitization System Based on Artificial Intelligence Algorithms. In 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON) (pp. 1-5). IEEE.

[19] Danner, M., Brake, E., Kosel, G., Kyosev, Y., Rose, K., Rätsch, M., & Cebulla, H. (2024). AI-assisted pattern generator for garment design. Communications in development and assembling of textile products, 5(2), 195-206.

[20] Yu, Z. Y., & Luo, T. J. (2021). Research on clothing patterns generation based on multi-scales self-attention improved generative adversarial network. International Journal of Intelligent Computing and Cybernetics, 14(4), 647-663.

[21] Mohiuddin Babu, M., Akter, S., Rahman, M., Billah, M. M., & Hack-Polay, D. (2024). The role of artificial intelligence in shaping the future of Agile fashion industry. Production Planning & Control, 35(15), 2084-2098.

[22] Jadhav, V., Bhavsar, S. S., Jadhav, S., Saraf, H., & Patil, V. (2025). Fashion Fusion: AI-Based Clothing Pattern Creator. Innovative Journal of Applied Science, 22-22.

[23] Wang, Y., Ramli, M. F., Song, H., & Li, X. (2024). Exploring the path of cultural sustainability for traditional costume embroidery patterns based on digital generative Art. Cultura: International Journal of Philosophy of Culture and Axiology, 21(4), 271-286.

[24] Peng, X. (2025). Research on the Generation of Artistic Patterns in Fashion Design Based on Style Transfer. International Journal of High Speed Electronics and Systems, 2540460.

[25] Ye, Q., Huang, R., Liu, H., & Wang, Z. (2023). Individualized garment pattern generation in batches based on biarc and ezdxf. AATCC Journal of Research, 10(4), 250-262.

[26] Sun, H., Yao, J., Zhang, H., Li, Z., & Cai, X. (2023). A Digital Simulation and Re-Editing Method for Clothing Patterns Based on Deep Learning and Somatosensory Interaction. International Journal of Pattern Recognition and Artificial Intelligence, 37(11), 2352016.

[27] Guo, Y., & Sun, L. (2025). Garment pattern generation method based on Diffusion Model and 3D reconstruction technology. Textile Research Journal, 00405175251328296.

[28] Zhong, X. (2025, February). Personalized Clothing Pattern Generation Driven by Body Shape Data. In 2025 5th International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 213-216). IEEE.

[29] Pan, J., Liao, Y., & Liang, D. (2024, September). A Method for Automatic Generation of Decorative Patterns for Volleyball Training Uniforms Based on Generative Adversarial Networks. In International Conference on Advanced Hybrid Information Processing (pp. 412-426). Cham: Springer Nature Switzerland.

[30] Araújo, D., Romero, L., & Faria, P. M. TEXTILE PATTERN DESIGN GENERATION USING GENERATIVE ADVERSARIAL NETWORKS. MCCSIS 2024, 131.

[31] Tango, K., Katsurai, M., Maki, H., & Goto, R. (2022). Anime-to-real clothing: Cosplay costume generation via image-to-image translation. Multimedia Tools and Applications, 81(20), 29505-29523.

[32] Wu, Q., Zhu, B., Yong, B., Wei, Y., Jiang, X., Zhou, R., & Zhou, Q. (2021). ClothGAN: generation of fashionable Dunhuang clothes using generative adversarial networks. Connection Science, 33(2), 341-358.

[33] Han, C., Lei, S., Mingming, W., Xiangfang, R., & Xiying, Z. (2021, February). Innovative design of traditional calligraphy costume patterns based on deep learning. In Journal of Physics: Conference Series (Vol. 1790, No. 1, p. 012029). IOP Publishing.

[34] Tirtawan, T., Susanto, E. K., Zaman, P. L., & Kristian, Y. (2021, April). Batik clothes auto-fashion using conditional generative adversarial network and U-Net. In 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) (pp. 145-150). IEEE.

[35] Cai, X., Li, Z., Xi, M., & Sun, H. (2023). Costume Pattern Sketch Colorization and Style Transfer Based on Neural Network. Journal of System Simulation, 35(3), 604-615.

[36] 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, 1-14.