

Personalized Clothing Style Generation System Based on StyleGAN3

Na Xu

School of Arts, Tiangong University, Tianjin 300384, China

Corresponding author: Na Xu (e-mail: xuxiquedan@163.com)

ABSTRACT In order to solve the core challenges of personalized clothing generation – uncontrollable style, difficult coupling of user preferences and body shape mismatch, this paper proposes a fine-grained automatic error correction system based on Bert, and uses the stylegan3 model. The system begins by parsing user-provided textual descriptions (e.g., style, color preferences) using a pretrained language model to extract structured semantic embeddings. These embeddings, along with body-shape encodings, are then mapped into a unified conditional vector. Our key contributions include a Fine-Grained Style Decoupling Encoder, which employs a frozen ResNet50 backbone with 12 parallel regression heads to independently predict continuous and categorical attributes (e.g., collar type, pattern density). An Attribute Conditional Injection Module adopts a dual-path latent modulation strategy: body conditions are mapped to deformation fields to drive pose-aware latent warping, while style/color embeddings are encoded as AdaIN parameters and injected into StyleGAN3 synthesis layers, ensuring semantic alignment between user intent and the generation process. Differentiable grid sampling enables adaptive adjustments to structural attributes such as shoulder width and waistline while preserving texture details. Experiments on DeepFashion demonstrate superior performance over mainstream methods (e.g., DragGAN, ControlNet+SD), with improved generation quality (FID=12.1), higher attribute control accuracy (AA=89.3%), and better detail preservation after body deformation (LPIPS=0.14). The system achieves high inference efficiency (54ms latency, 18.5 img/s), offering industrial-grade deployability for virtual try-on and personalized e-commerce, thereby supporting personalized e-commerce and virtual try-on. The body-aware garment model is also relevant to wearable product design in which device placement and fit depend on human morphology.

INDEX TERMS personalized clothing, stylegan3 model, style decoupling, body-shape adaptation, wearable product design

I. INTRODUCTION

WITH the rapid development of artificial intelligence and computer vision, personalized clothing generation based on user body shape, aesthetic preferences, and specific scene needs has become a core demand, not only promoting the fashion industry's transformation toward digitalization and intelligence but also critically optimizing textile production and raw material consumption across the entire supply chain [1,2]. Its industrial connotation lies in achieving a leap from standardized production of “one size fits all” to precise customization of “one version for each person” through artificial intelligence technology, aiming to break through the inherent bottlenecks of traditional design that relies on manual experience and physical proofing, such as long cycle time, high cost, and difficulty in scalable adaptation to individual differences of users [3,4]. Although generative adversarial networks [5,6] have made breakthroughs in the field of image synthesis, they still face bottlenecks in clothing generation tasks, such as fuzzy

generated styles, loss of control over local attributes, and difficulty in embedding user preferences into the latent space. Especially in commercial scenarios, users have clear control demands for fine-grained attributes such as collar shape, sleeve length, pattern, and silhouette of clothing, and at the same time need to achieve automatic adaptation based on body shape characteristics [7]. Existing generative models focus on global style transfer and lack structured decoupling and conditional injection mechanisms, resulting in a serious disconnect between generated results and user intentions [8]. Building a high-fidelity, fine-grained controllable, and low-delay personalized clothing generation system has become a key challenge in the implementation of AI-driven fashion design, as its efficacy directly impacts the ability of the textile industry to transition to on-demand, waste-reducing production workflows.

The current mainstream clothing generation methods can be divided into three categories: end-to-end generation based on GAN (Generative Adversarial Network) [9], conditional-

guided generation based on diffusion model [10], and local editing methods based on image translation [11]. The StyleGAN series of models (StyleGAN2/3) [12,13] are widely adopted due to their high-quality generation capabilities, but their latent space lacks semantic alignment, resulting in attribute control relying on posterior editing or coarse-grained label splicing, making it difficult to achieve independent adjustment. ControlNet and Stable Diffusion have made progress in spatial structure control by guiding generation through edge graphs or semantic graphs, but their high inference delay and blurred texture details make it difficult to meet the needs of real-time interaction [14]. FashionFormer applies the Transformer architecture and attribute token injection mechanism. Although it supports 12 types of attribute control, it consumes a lot of training resources, and the generated images lack the texture of real fabrics. The DragGAN series achieves interactive deformation by dragging points, but relies on manual intervention and cannot be automatically batch generated [15,16]. The DiStyleGAN adopts a dual-path style injection method, which performs well in style transfer. However, it does not consider the body adaptation problem, resulting in proportional distortion of generated clothing for users of different body shapes. The above methods either sacrifice generation quality, ignore user condition coupling [17], or lack system-level deployment capabilities. A complete solution that takes into account “high fidelity, fine-grained, low delay, and deployable” has not yet been formed.

In terms of fine-grained attribute decoupling, some studies have attempted to achieve local control through semantic segmentation graphs (such as GauGAN2 + Fashion Parser) or attribute classification heads (such as AttGAN). However, the former relies on precisely labeled segmentation masks and has poor generalization [18]; the latter only supports discrete attribute switching and has weak control capabilities for continuous variables. In terms of user preference coupling, conditional normalization techniques such as AdaIN are used for latent space modulation, but they mostly act on the global style layer and are not aligned with the semantics of the clothing structure, resulting in style drift after injection [19,20]. In terms of body adaptation, virtual fitting systems such as DressCode [21,22] achieve local replacement by preserving the source image structure, but their non-generative architecture limits design diversity. Latent Consistency Models (LCM) [23] improve inference speed, but their distillation mechanism weakens the control of local details. Pose-guided generation methods attempt to apply key points to control deformation, but fail to model the continuous deformation field of the latent space, resulting in stiff body adaptation and texture breakage [24,25]. In summary, existing research has the defects of local optimization but global imbalance in the three core issues of “uncontrollable style” [26], “difficulty in coupling preferences”, and “unsuitable body shape”, and lacks a systematic solution from latent space structure, conditional injection mechanism to deformation modeling.

This paper proposes systematic solutions to the three core shortcomings of the aforementioned research. To address the issue of “uncontrollable style”, a fine-grained style decoupling encoder is designed, using 12 parallel regression heads to independently predict attribute factors. This design is particularly well-suited to clothing, which is composed of multiple semantically independent local attributes. It effectively avoids gradient interference between attributes and enables independent and precise control of continuous and categorical attributes such as collar shape, sleeve length, and pattern density. To address the issue of “difficulty in coupling preferences”, an ACIM (Attribute Condition Injection Module) is constructed. Through a dualpath latent space modulation strategy (body $\rightarrow$ deformation field, style/color $\rightarrow$ AdaIN), user conditions at different semantic levels are decoupled and injected into different layers of the generative network, achieving strong semantic alignment between the user’s multidimensional preferences and the generative process. To address the issue of “unsuitable body shape”, a pose-aware latent warping mechanism is applied, dynamically warping the latent code in the latent space based on differentiable grid sampling. This operation achieves continuous adaptive adjustment of shoulder width, waistline, and hem with low computational overhead, while effectively preserving the integrity of clothing texture details. This research constructs a complete, controllable generation chain from semantic decoupling, conditional injection, to deformable modeling. On DeepFashion, it achieves an FID (Fréchet Inception Distance) of 12.1, an attribute accuracy of 89.3%, and an inference delay of 54ms, surpassing existing systems. The system supports industrial-grade high-concurrency deployment (18.5 img/s) and can be seamlessly integrated into platforms such as virtual fitting, online customization, and artificial intelligence design assistants. This system enables merchants to achieve personalized “what people see is what they can get” displays, significantly reducing sample production costs and inventory pressures, while improving consumer decision-making efficiency and the shopping experience. Its modular architecture provides a reusable technical paradigm for the digital transformation of the apparel industry, from design, display, to sales. It propels AI generation from “generative capability” to “precise control, on-demand generation, and system-wide application”, demonstrating clear academic leadership and commercial application value.

II. METHODS

A. OVERALL SYSTEM ARCHITECTURE

To achieve high-fidelity, fine-grained, and controllable personalized clothing generation [27,28], this paper constructs an end-to-end system architecture that integrates user-condition encoding, latent space attribute decoupling, multi-path conditional injection, and a deformation adaptation mechanism. Based on StyleGAN3, the system precisely responds to multi-dimensional inputs such as body shape, style, and color, maintaining generation quality while en-

abling independent control of local attributes and real-time inference capabilities. The overall system architecture is shown in Figure 1.

The input layer receives a user condition vector, consisting of a 3D body shape encoding, a 5D style embedding, a 3D color preference, and a 4D occasion label. These are then normalized and used as a global control signal. The encoder layer deploys a fine-grained style decoupling encoder, whose backbone is a frozen ResNet50 pre-trained on ImageNet. Its output is connected to 12 parallel fully-connected regression heads, which predict scores for 12 attribute categories, such as collar shape, sleeve length, and pattern density, to generate decoupling factors. This design ensures independent modeling of local attributes, avoiding semantic coupling. All tensor dimensions are strictly aligned across modules, and gradients are back-propagated only in the FGSDE (fine-grained style decoupling encoder) head and the ACIM module. The backbone feature extractor parameters are frozen, improving training stability and efficiency.

The injection layer is implemented by an attribute-conditional injection module, centered around a two-branch MLP (Multi-Layer Perceptron) architecture. The figure branch maps the body into a latent space displacement field, driving pose-aware latent warping. The style and color branches output AdaIN parameters, which are injected into Layers 4–12 of the StyleGAN3 synthesis network, achieving fine-grained style modulation. The generation layer, based on the official StyleGAN3 architecture, integrates a differentiable spatial transformer to apply a deformation field to the latent code before the mapping network. Finally, after 18 layers of style modulation and upsampling, a high-fidelity image is output. Key injection locations are allocated based on the attribute semantic hierarchy to ensure semantic alignment between control signals and the generation stage. The system supports end-to-end training, and its inference delay meets industrial-grade real-time interactive requirements.

B. FINE-GRAINED STYLE DECOUPLING ENCODER

To solve the problem of “uncontrollable style” in the generation process, that is, local attributes are affected by the global latent code coupling and are difficult to adjust independently, this paper designs FGSDE, whose core goal is to decompose the visual semantics of the input image into 12 structurally independent and semantically clear continuous/categorical attribute factors to achieve precise directional control in the generation stage.

FGSDE adopts a “feature extraction + multi-task regression” architecture. The input is a standardized 512×512 RGB (Red, Green, Blue) clothing image $I \in \mathbb{R}^{512 \times 512 \times 3}$, and the high-level semantic features are extracted through the backbone network:

$$F = \text{ResNet50}_{\text{backbone}}(I) \in \mathbb{R}^{2048} \quad (1)$$

ResNet50 is an ImageNet pre-trained model with frozen

parameters to avoid overfitting and accelerate training. The feature vector F is input in parallel to 12 independent fully connected regression heads, each head corresponding to a class of clothing attributes, with a unified structure:

$$s_i = \sigma_i(W_i \cdot F + b_i), \quad i = 1, 2, \dots, 12 \quad (2)$$

In Formula 2, $W_i \in \mathbb{R}^{d_i \times 2048}$ and b_i are the learnable parameters of the i -th attribute head; σ_i is the attribute-specific activation function.

To achieve independent adjustment of local attributes in the “uncontrollable style” problem during the generation process, this paper designs a fine-grained style decoupling encoder. Its core architecture is a “feature extraction+multi task regression” structure: using ResNet50 pre-trained on ImageNet as the backbone network, and always keeping its weights frozen. This design is based on two considerations: firstly, utilizing the powerful general visual feature extraction ability of pre-trained models to avoid overfitting on small-scale fashion datasets; Secondly, freezing the backbone can significantly reduce the number of training parameters, improve training stability and efficiency, and enable gradient backpropagation through only 12 lightweight regression heads, which is beneficial for the model to converge quickly.

This system selected 12 visual attributes that are most representative, semantically independent, and often of interest to users in fashion design for decoupling modeling. These attributes cover multiple dimensions of clothing such as structure, pattern, fit, and material. Each attribute is predicted by an independent regression head, achieving gradient isolation and semantic decoupling at the feature level. Although clothing styles in the real world may be influenced by more continuous and subtle attributes, the 12 attributes selected in this article have high control priority and interpretability in personalized generation tasks, and have low correlation with each other. Therefore, they are suitable as the basic set for decoupling control.

The decoupled attribute information is shown in Table 1.

TABLE 1. Definition and quantification range of 12 types of decoupling attributes

Attribute	Type	Quantitative approach	Control target layer
Collar type	Classification	3 types of one hot	Layer 6
Sleeve length	Classification	3 types of one hot	Layer 8
Pattern density	Continuous	[0.0, 1.0]	Layer 4
Color saturation	Continuous	[0.0, 1.0]	Layer 5
Loose hem	Continuous	[0.0, 1.0]	Layer 10
Shoulder line inclination	Continuous	[-0.5, +0.5]	Layer 7
Waist line position	Continuous	[0.3, 0.7]	Layer 9
Collar size	Continuous	[0.1, 0.9]	Layer 6
Tightness of cuffs	Continuous	[0.0, 1.0]	Layer 8
Pattern complexity	Continuous	[0.0, 1.0]	Layer 4
Material reflectivity	Continuous	[0.0, 1.0]	Layer 14
Overall profile	Classification	4 types of one hot	Layer 12

The 12 parallel regression heads proposed in this article mainly target key visual attributes in clothing that have clear

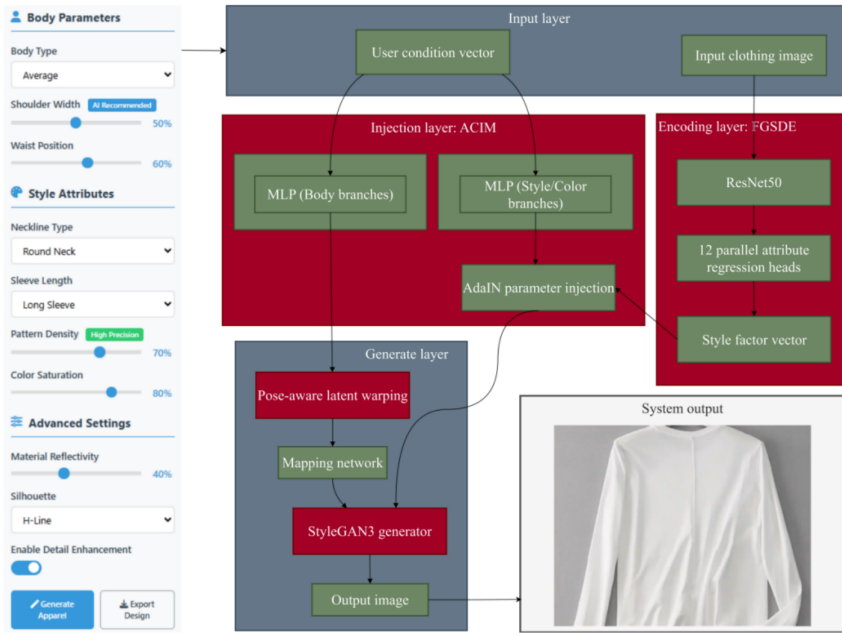


FIGURE 1. Overall system architecture

semantic boundaries and can be structurally described, such as collar shape, sleeve length, pattern density, etc. These attributes have high user attention and control requirements in personalized generation tasks. Each head is independently modeled, avoiding gradient interference between attributes and achieving precise decoupling control in semantically explicit dimensions. However, clothing styles also contain many continuous and fine-grained physical attributes, such as the dynamic drape of fabrics, local elasticity adaptation, microscopic texture mapping, etc., which are difficult to fully capture through discrete or low dimensional continuous variables. Future work may consider introducing finer grained physical perception modules (such as fabric simulation priors, high-dimensional material representation learning) or hierarchical decoupling strategies to further enhance the system's performance and realism in real complex styles. For continuous attributes $\sigma_i(x) = \text{Sigmoid}(x)$, the output range is $[0, 1]$ or linearly scaled to the target interval; for categorical attributes $\sigma_i(x) = \text{Softmax}(x)$, the output is a one-hot probability distribution, and cross entropy loss is used during training. The final output style factor vector is:

$$S = [s_1, s_2, \dots, s_{12}]^T \in \mathbb{R}^{12} \quad (3)$$

This design uses a frozen ResNet50 backbone network during the training phase, updating only the parameters of 12 parallel regression heads, greatly reducing training costs and improving stability. In the inference stage, the backbone network only performs forward propagation once, and feature extraction and attribute prediction can be performed in parallel. The computational overhead is mainly concentrated in the lightweight fully connected layer, so the overall latency remains at a low level. In addition, the system supports TensorRT acceleration and layer fusion

optimization, which can further improve throughput through model quantization and dynamic batch processing in actual deployment.

C. ATTRIBUTE CONDITION INJECTION MODULE

To solve the problem of “difficult coupling” of user multi-dimensional preferences (body type, style, color, occasion) in the latent space, that is, the semantic drift and generation distortion caused by traditional splicing or global modulation, this paper proposes ACIM. Its core idea is to decouple the user condition vector C branch according to the principle of aligning the attribute semantic level with the generation stage, and inject it into the StyleGAN3 middle layer to achieve precise semantic modulation of the latent space.

ACIM receives the user condition vector $C \in \mathbb{R}^{15}$, and its structure is defined as:

$$C = \left[\underbrace{c_{\text{body}}}_{\in \mathbb{R}^3}; \underbrace{c_{\text{style}}}_{\in \mathbb{R}^5}; \underbrace{c_{\text{color}}}_{\in \mathbb{R}^3}; \underbrace{c_{\text{occasion}}}_{\in \mathbb{R}^4} \right]^T \quad (4)$$

In Formula 4, c_{body} is the one-hot encoding of body type; c_{style} is the style embedding; c_{color} is the color preference vector; c_{occasion} is the occasion label embedding.

The body condition c_{body} inputs a lightweight MLP and outputs a latent space displacement field:

$$\Delta Z = \text{MLP}_{\text{body}}(c_{\text{body}}) \in \mathbb{R}^{512} \quad (5)$$

This displacement field is used to drive the Pose-Aware Latent Warping module to perform spatial deformation on the original latent code Z :

$$Z' = W(Z, \Delta Z) \quad (6)$$

In Formula 6, $W(\cdot)$ is a differentiable spatial transformation function to ensure that the gradient of the deformation process is transferable.

The style and color conditions $[c_{\text{style}}; c_{\text{color}}] \in \mathbb{R}^8$ are encoded into an intermediate semantic vector through a shared MLP:

$$h_{sc} = \text{MLP}_{\text{style}}([c_{\text{style}}; c_{\text{color}}]) \in \mathbb{R}^{256} \quad (7)$$

For each target layer $l \in L = \{4, 5, 6, \dots, 12\}$ of the StyleGAN3 [29,30] synthesis network, affine transformation parameters are generated independently:

$$\gamma^{(l)} = W_{\gamma}^{(l)} \cdot h_{sc} + b_{\gamma}^{(l)} \in \mathbb{R}^{512} \quad (8)$$

$$\beta^{(l)} = W_{\beta}^{(l)} \cdot h_{sc} + b_{\beta}^{(l)} \in \mathbb{R}^{512} \quad (9)$$

During the generation process of layer l , adaptive instance normalization (AdaIN) is applied to the intermediate feature map $F^{(l)} \in \mathbb{R}^{H_l \times W_l \times 512}$:

$$\text{AdaIN}(F^{(l)}, \gamma^{(l)}, \beta^{(l)}) = \gamma^{(l)} \cdot \frac{F^{(l)} - \mu(F^{(l)})}{\sigma(F^{(l)})} + \beta^{(l)} \quad (10)$$

In Formula 10, μ and σ are the channel dimension mean and standard deviation. This operation realizes the local modulation of style and color semantics in the latent space.

D. POSE-AWARE LATENT WARPING MECHANISM

This paper proposes a pose-aware latent warping mechanism. The core idea is to apply a semantically aware continuous deformation field to the original latent code based on the body shape code in the latent space, so that the generation process can adapt to different body shapes while maintaining the structure. This mechanism is placed before the StyleGAN3 Mapping Network. The input is the original latent code and the body shape conditional code c_{body} , and the output is the deformed latent code.

The body shape code c_{body} is mapped to the latent space displacement vector by a lightweight multi-layer perceptron:

$$\Delta Z = \text{MLP}_{\text{pose}}(c_{\text{body}}) \in \mathbb{R}^{512} \quad (11)$$

In Formula 11, MLP_{pose} consists of two layers of linear transformation and activation function. This paper uses a differentiable space transformer to deform the latent code. Since Z is a vector (not an image), it is considered as a pseudo tensor of “ $1 \times 1 \times 512$ ” and the affine transformation matrix Θ is constructed in the latent space:

$$\Theta = \begin{bmatrix} 1 + \alpha_x & 0 & \delta_x \\ 0 & 1 + \alpha_y & \delta_y \end{bmatrix} \quad (12)$$

In Formula 12, the scaling factors α_x and α_y and the translation factors δ_x and δ_y are obtained by the linear projection of the first 4 dimensions of ΔZ :

$$[\alpha_x, \alpha_y, \delta_x, \delta_y]^T = W_{\text{affine}} \cdot \Delta Z_{1:4} + b_{\text{affine}} \in \mathbb{R}^4 \quad (13)$$

The grid sampling function $G(\cdot)$ is defined to deform the latent code:

$$Z' = \mathcal{G}(Z, \Theta) \quad (14)$$

$G(\cdot)$ uses bilinear interpolation to resample on the latent space grid, and its gradient can be transferred through standard back propagation:

$$\frac{\partial L}{\partial Z} = \sum_i \frac{\partial L}{\partial Z'_i} \cdot \frac{\partial Z'_i}{\partial Z} \quad (15)$$

$$\frac{\partial L}{\partial \Delta Z} = \sum_i \frac{\partial L}{\partial Z'_i} \cdot \frac{\partial Z'_i}{\partial \Theta} \cdot \frac{\partial \Theta}{\partial \Delta Z} \quad (16)$$

The body encoding definition and deformation target are shown in Table 2.

TABLE 2. Body encoding definition and deformation target

Body type	Encoding values	Target deformation direction	Corresponding clothing adjustments
Slim and tall	[1,0,0]	Stretch shoulder width and extend hem	Optimization ratio
Well-proportioned	[0,1,0]	No significant deformation	Maintain the original design
Full-figured	[0,0,1]	Relax the waistline and increase the sleeve cage	Enhance inclusivity

E. LOSS FUNCTION

To jointly optimize the generation quality, attribute control accuracy, and body deformation fidelity, this paper defines a multi-objective loss function:

$$L = L_{\text{GAN}} + \lambda_1 L_{\text{attr}} + \lambda_2 L_{\text{pose}} \quad (17)$$

In Formula 17, λ_1 controls the attribute control strength, and λ_2 controls the deformation fidelity strength. The StyleGAN3 native non-saturated GAN loss is used:

$$L_{\text{GAN}} = \mathbb{E}_{z \sim p_z} [-\log D(G(z))] \quad (18)$$

In Formula 18, D is the discriminator; G is the generator; z is the input latent code. This loss drives the generated image distribution to approach the real data distribution, ensuring visual fidelity.

L_{attr} is used to supervise the alignment of the FGSDE output attribute factor with the real label to improve the local control accuracy:

$$L_{\text{attr}} = \frac{1}{12} \sum_{i=1}^{12} \text{MSE}(\hat{s}_i, s_i) \quad (19)$$

In Formula 19, $\hat{s}_i$ is the FGSDE prediction value, and s_i is the manually labeled real value (continuous attribute) or one-hot label (categorical attribute). For categorical attributes, CrossEntropyLoss is actually used, but for the sake of consistency, its contribution strength is expressed in the form of equivalent mean square error in the total loss.

To prevent image details from being lost after body deformation, the LPIPS (Learned Perceptual Image Patch Similarity) metric is used to measure the perceptual distance between the generated image I_{gen} and the reference image $I_{\text{body_ref}}$:

$$L_{\text{pose}} = \text{LPIPS}(I_{\text{gen}}, I_{\text{body_ref}}) \quad (20)$$

In Formula 20, $I_{\text{body_ref}}$ is the reference image generated under the same style conditions and well-proportioned body, which is used to measure the structural fidelity when “only the body changes”. This loss prevents deformation from causing texture blur or structural collapse.

The loss function parameters are determined as shown in Figure 2.

Comprehensive score = $\text{FID} \times (100 - \text{AA}) + 10 \times \text{LPIPS}$ (the lower the better); AA: (Attribute Accuracy). The comprehensive loss reaches the global minimum (112.3) at $\lambda_1=1.0$ and $\lambda_2=0.5$, verifying that this combination is the optimal hyperparameter configuration. When $\lambda_1 < 1.0$, the attribute control signal is too weak, causing the generated style to deviate from the user’s intent. When $\lambda_1 > 1.0$, the attribute constraints are too strong, sacrificing image naturalness and diversity. When $\lambda_2 < 0.5$, body deformation lacks perceptual constraints, easily leading to texture breakage. When $\lambda_2 > 0.5$, deformation freedom is overly suppressed, reducing adaptation flexibility. Selecting $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ achieves a Pareto optimal between generation quality, attribute accuracy, and deformation fidelity, supporting high-performance and stable operation of the system and meeting the requirements of industrial-grade controllable generation.

F. FINE-GRAINED PERSONALIZED GENERATION SYSTEM

The fine-grained personalized generation system constructed in this paper adopts a modular and scalable microservice architecture, supporting high-concurrency, low-delay, end-to-end inference. The frontend receives structured user input (JSON format: body type, style keywords, color preferences) and routes it to the core inference service via an API (Application Programming Interface) gateway. The system backbone consists of four connected modules: the FGSDE attribute encoder, the ACIM condition injector, the Pose-Aware Latent Warper, and the StyleGAN3 generator. Modules communicate via shared-memory tensors to avoid serialization overhead. The control flow utilizes an asynchronous pipeline design, supporting a maximum batch size of 16 and improving GPU (Graphics Processing Unit) utilization. The system features a built-in attribute-to-layer mapping routing table, dynamically binding user

input to generated layer modulation parameters, ensuring independent control of 12 attribute types. All modules are hot-swappable, facilitating the subsequent expansion of new attributes or adaptation to new body models.

The system is deployed in a Kubernetes cluster, supporting automatic scaling and load balancing. The inference service is packaged as a Docker container, leveraging the NVIDIA Triton Inference Server for multi-model collaborative scheduling and stable memory usage on the A100 GPU. TensorRT is used to perform layer fusion and kernel optimization for the ACIM and StyleGAN3 backbones. The system integrates a Prometheus monitoring module, collecting real-time metrics such as FID drift and attribute accuracy, and automatically rolling back to a stable version in the event of anomalies. To adapt to edge devices, a lightweight deployment package is provided to meet the needs of real-time interactive scenarios such as e-commerce and virtual fitting. The system interface of this paper is shown in Figure 3.

The system performs deep semantic analysis of user commands, decomposing the global representation of the clothing into 12 independent and semantically distinct attribute factors. Each attribute is predicted and encoded by a dedicated regression head, ensuring precise and independent control of details such as whether the neckline is V-shaped, round, or square, and whether the sleeve length is short, medium, or long. This structured decoupling allows users or designers to adjust individual attributes, much like adjusting a music mixer, without causing cascading changes in other features, thus enabling fine-grained, high-fidelity personalized design.

To precisely respond to users’ multi-dimensional personalized preferences, a dual-path modulation strategy is used: body shape is encoded into a latent space displacement field through an MLP network, driving the subsequent deformation module to adaptively adjust the clothing contour; style and color preferences are encoded into AdaIN parameters and precisely injected into specific intermediate layers of the StyleGAN3 generator. This “divide and conquer” injection mechanism ensures that control signals are applied at the semantically relevant network level, ensuring that the generated results are highly consistent with user intent. The system applies a Pose-Aware Latent Warping mechanism to perform differentiable spatial deformation on the input latent code. Based on the user’s selected body type, the system generates corresponding affine transformation parameters through a lightweight network, dynamically warping the latent code in latent space. This results in physical deformation effects such as stretched shoulders, loosened waistlines, and extended hems in the final image. This process achieves intelligent adaptation of clothing patterns to different body shapes while maintaining the texture details and style characteristics of the clothing. The generated clothing is natural and fits well, solving the problems of texture breakage and proportion distortion common in previous methods.

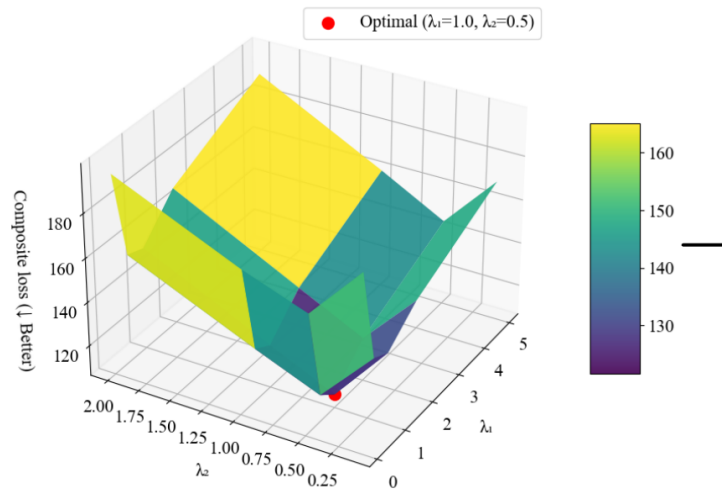


FIGURE 2. Loss function parameter determination



FIGURE 3. System interface

G. BERT BASED SEMANTIC ERROR CORRECTION AND INTENT UNDERSTANDING MODULE

In order to accurately convert user natural language descriptions into structured control signals and avoid semantic conflicts or ambiguous expressions that may interfere with the generation process, this system introduces a semantic error correction and intent understanding module based on a pre-trained language model BERT. This module does not directly participate in visual generation or latent spatial operations, but serves as a front-end semantic processor to ensure the clarity, consistency, and feasibility of input conditions. Its core functions include:

1. Semantic standardization and ambiguity resolution:

User input may contain ambiguous, polysemous, or unconventional expressions. The BERT module performs intent parsing and semantic clustering on input text, mapping it to a predefined attribute vocabulary in the system to eliminate ambiguity and generate standardized attribute description vectors.

2. Conflict detection and logical error correction: There may be physical or stylistic conflicts between clothing attributes. The module detects such conflicts through predefined attribute compatibility rules and semantic constraint libraries, and automatically adjusts them based on priority or user historical preferences to avoid generating unreasonable designs.

3. Intent completion and condition enhancement: When the user input is too brief, the module can automatically complete the missing but necessary attributes based on the fashion knowledge graph and common matching rules, forming a complete condition vector and improving the completeness and practicality of the generated results.

The output of this module is a structured, conflict free semantic control vector with dimensions and format consistent with the 15 dimensional user condition vector defined by the system, which is directly used as input for subsequent FGSDE and ACIM modules. Through this preprocessing stage, the system can respond more robustly to user language instructions, improve the controllability and user experience of personalized generation, while avoiding generation failures or semantic drift caused by input noise. In deployment, this module can independently serve multiple interactive scenarios, achieving a closed-loop personalized process of "description error correction preview adjustment".

III. EXPERIMENTS

A. DATASET AND PREPROCESSING

The dataset used in this study is DeepFashion [31,32], which contains over 800,000 diverse fashion images. To ensure generation quality and control consistency, this paper strictly selects samples that meet the following conditions: only frontal, full-body, standing, and unobstructed images of models wearing clothing are retained, excluding side profiles, sitting poses, close-ups, or images with multiple people; clothing integrity: samples with accessories obscuring the subject (such as large scarves and backpacks), severe wrinkles, or perspective distortion are excluded; all images are uniformly cropped and resized to 512×512 pixels, keeping the subject vertically centered and shoulders aligned in the frame to ensure latent space deformation consistency; balanced sampling is performed according to clothing categories (tops, dresses, jackets, and pants) to avoid style distribution bias.

To support the 12-dimensional attribute decoupling training of the FGSDE module, this paper designs a structured annotation protocol. Three professional fashion annotators complete the annotation in a unified tool, taking approximately 45 seconds per image. The annotation consistency is verified using Krippendorff's α coefficient ($\alpha = 0.87$, indicating high consistency).

To simulate real user input, this paper synthesizes user condition vectors based on labeled attributes and automatically maps them based on the model's labeled height-weight ratio (HWR):

$HWR > 1.8 \rightarrow$ slim and tall [1,0,0];

$1.6 \leq HWR \leq 1.8 \rightarrow$ well-proportioned [0,1,0];

$HWR < 1.6 \rightarrow$ full-figured [0,0,1];

Style Embedding c_{style} : the clothing style keywords are input into the CLIP (Contrastive Language–Image Pretraining) text encoder, and its 512-dimensional output is reduced

to 5 dimensions through PCA (Principal Component Analysis);

Color Preference c_{color} : the HSV (Hue, Saturation, Value) mean of the clothing's primary colors (the top three colors from K-means clustering) is extracted, normalized, and the top three dimensions are taken. Occasion Label $c_{occasion}$: it is mapped to one-hot categories based on product metadata: {Commuting, Party, Sports, Leisure}.

This synthesis strategy ensures a strong correlation between the conditional vector and image semantics while preserving noise and diversity, improving model robustness.

The acquired image data is divided into training, validation, and test sets in a 3:1:1 ratio. The experimental hyperparameter configuration is shown in Table 3.

TABLE 3. Experimental hyperparameter configuration

Parameter category	Parameter	Set value	Instructions
Hardware	GPU model	4×NVIDIA A100 80GB	NVLink interconnection, enable Tensor Core
Hardware	CUDA version	12.1	–
Framework	Deep learning framework	PyTorch 2.1 + TorchScript	–
Framework	StyleGAN3 code base	NVIDIA Official v1.2	Enable R1 regularization and path length regularization
Optimizer	Optimizer type	Adam	–
Optimizer	Learning rate	2×10^{-3}	No Warmup, no attenuation
Optimizer	Batch size	32	Gradient accumulation steps=1
Optimizer	Loss weight λ_1 / λ_2	1.0 / 0.5	Determined by grid search in Figure 2

Note: CUDA (Compute Unified Device Architecture).

B. COMPARISON SYSTEMS

To comprehensively evaluate the performance of the proposed system, advanced generative models are selected as baselines, including GAN, Diffusion, Flow, and interactive generation. All models are trained uniformly on the DeepFashion dataset to ensure fair evaluation. Model comparison information is shown in Table 4.

TABLE 4. Model comparison information

System	Model	Conditional control method
S1	StyleGAN3	Unconditional/global label concatenation to input layer
S2	StyleGAN-XL	Category tags+layered latent spatial modulation
S3	DiStyleGAN	Dual path style injection (content+style hidden code separation)
S4	ControlNet + SD (Stable Diffusion)	Edge map/semantic map+text cross attention guidance
S5	FashionFormer	Transformer encoding attribute Token+injection of cross attention layer
S6	GauGAN2 + Fashion Parser	Semantic segmentation map guided generation+regional style control
S7	LCM	Consistency distillation in latent space+text condition guidance
S8	DragGAN	Interactive point drag+latent space optimization deformation
S9	StyleGAN3 + FGSDE + ACIM + Pose-Aware Latent Warping	12 dimensional decoupling factor+body shape encoding+AdaIN layer injection+latent space deformation

C. EVALUATION METRICS

FID [33,34] measures the distance between the distribution of generated images and the distribution of real images in the feature space. The lower the value, the higher the quality of the generated image. The calculation formula is:

$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}) \quad (21)$$

In Formula 21, μ_r and Σ_r are the mean and covariance matrices of the real image features; μ_g and Σ_g are the mean and covariance matrices of the generated image features; Tr is the matrix trace operation. KID (Kernel Inception Distance) [35] is an unbiased estimate of FID, calculated using a polynomial kernel:

$$KID = \frac{1}{n(n-1)} \sum_{i \neq j} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i \neq j} k(y_i, y_j) - \frac{2}{nm} \sum_{i,j} k(x_i, y_j) \quad (22)$$

In Formula 22, x_i and y_j are the real image and the generated image, respectively.

AA measures the consistency of the generated image with the user input conditions in 12 predefined attributes. For each generated image, an independently trained attribute classifier is used to predict the attribute score, and the accuracy is calculated with the true label:

$$AA = \frac{1}{N \times 12} \sum_{i=1}^N \sum_{j=1}^{12} \mathbb{I}(\hat{a}_{ij} = a_{ij}^{\text{gt}}) \quad (23)$$

In Formula 23, N is the number of test images, and $\hat{a}_{ij}$ is the predicted value of the j-th attribute of the i-th image (the categorical attribute takes argmax, and the continuous attribute is binarized with a threshold of 0.5). a_{ij}^{gt} is the true

conditional label input by the user, and $\mathbb{I}$ is the indicator function (1 if the prediction is correct, otherwise 0).

LPIPS measures the difference between two images at the human visual perception level and is used to evaluate the ability to preserve details after body deformation. This paper calculates the perceptual distance between the generated image I_{gen} and the reference image I_{ref} (same style, well-proportioned body):

$$LPIPS(I_{\text{gen}}, I_{\text{ref}}) = \sum_{l=1}^L w_l \cdot \left\| \frac{\phi_l(I_{\text{gen}}) - \phi_l(I_{\text{ref}})}{\sigma_l} \right\|_2^2 \quad (24)$$

In Formula 24, ϕ_l is the feature map of the l-th layer of the pre-trained AlexNet; σ_l is the standard deviation of the feature channels of each layer; w_l is the layer weight.

SSIM (Structural Similarity) measures the local similarity between two images in three dimensions: brightness, contrast, and structure, which is used to evaluate the consistency of the generated image with the reference image in terms of macrostructure:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (25)$$

In Formula 25, x, y are local image patches; μ_x and μ_y are local means; σ_x^2 and σ_y^2 are local variances; σ_{xy} is local covariance.

IV. RESULTS

A. QUANTITATIVE RESULTS OF THE MAIN EXPERIMENT

The performance comparison of different models is shown in Figure 4.

The proposed model comprehensively outperforms baseline models in five core metrics, validating its ability to collaboratively optimize generation quality, attribute control, and structural fidelity. In terms of generation quality, its FID=12.1 and KID=1.73 demonstrate that the generated image distribution closely matches real-world data and exhibits no pattern collapse. In terms of controllability, its AA=89.3% demonstrates that the FGSDE+ACIM architecture effectively decouples 12 attribute classes, enabling precise directional control. In terms of deformation fidelity, LPIPS=0.14 and SSIM=0.82 demonstrate that the Pose-Aware Warping mechanism maintains texture and structural integrity during body adaptation. In comparison, ControlNet+SD suffers from poor FID/KID performance due to the inherent ambiguity of the diffusion model. While FashionFormer achieves high AA, its LPIPS is only 0.31, exposing the Transformer architecture's limitations in detail reconstruction. This paper achieves quality, control, and fidelity in a fully automatic, end-to-end framework, establishing a new benchmark for controllable generation.

The FID-AA Pareto plot clearly shows that the proposed model is at the optimal equilibrium point in the upper right corner (FID=12.1, AA=89.3%), achieving the

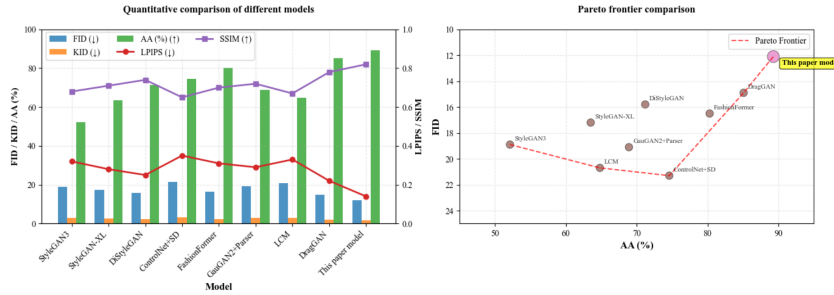


FIGURE 4. Performance comparison of different models

Pareto optimal between generation quality and attribute control accuracy. Traditional GAN-based models (such as StyleGAN3/XL) cluster in the lower left region ($AA < 65\%$, $FID > 17$), reflecting their inherent flaw of prioritizing quality over control. Some conditional models (such as FashionFormer and DiStyleGAN) shift to the right ($AA > 70\%$), but their FID does not fall below 15.8, indicating that attribute injection sacrifices image naturalness. While DragGAN approaches the proposed FID, its $AA = 85.1\%$ still lags behind, and its interactive mechanism is not scalable. This paper uses FGSDE to decouple attribute factors from structured data, avoiding control interference generation. ACIM precisely injects user preferences into the semantic alignment layer to suppress style drift. Pose-Aware Latent Warping uses latent space deformation, rather than pixel-level distortion, to ensure detail fidelity. These results demonstrate that fine-grained controllable generation does not have to sacrifice quality. This architecture provides a reusable technical paradigm for industrial-grade AI design systems.

B. ATTRIBUTE CONTROL ACCURACY

The independent control capabilities of 12 clothing attributes are analyzed, and the prediction accuracy of each attribute is quantified, demonstrating high-precision decoupled control of local details such as collar shape, sleeve length, and pattern. The results of the attribute control accuracy analysis are shown in Figure 5.

The results in Figure 5 show that the proposed model outperforms the baseline model on all 12 attribute categories, validating the effectiveness of the FGSDE architecture for fine-grained decoupled control. In particular, the proposed model achieves over 91% accuracy for structurally sensitive attributes (such as collar style and sleeve length). This is due to the semantic alignment between the attribute head and the generated layer (such as collar style control layer 6), which avoids the semantic confusion caused by global modulation in traditional methods. In comparison, DiStyleGAN, which only models style latents, has weak control over structural attributes (shoulder line, waist position); ControlNet+SD relies on spatial guidance, resulting in insufficient quantization accuracy for continuous variables (pattern density, cuff tightness); while FashionFormer applies attribute tokens, the Transformer's attention mechanism blurs local features,

resulting in low accuracy for complex attributes; DragGAN, which relies on point dragging, has limited control over non-geometric attributes. The proposed method, through a parallel regression head, frozen backbone, and layer-map injection, achieves gradient isolation and semantic focus between attributes, making it the only end-to-end system currently capable of high-precision independent control of all attributes.

The experimental results show that the system achieves high control accuracy on all 12 preset structured attributes, verifying the effectiveness of parallel header design in semantically explicit attributes. However, for unstructured and continuously changing style details such as drape and local wrinkle distribution, the current system still relies on StyleGAN3 prior distribution for generation, without explicitly modeling its mapping relationship with control signals. This also explains to some extent the limitations of generating results in highly dynamic detail restoration, which can be further improved in the future by introducing physically guided loss functions or hierarchical latent variable decomposition.

To explore the effect of switching classification attributes, in this paper, other conditions are fixed, and only the collar type code (V-neck, square neck, round neck) is switched. The visualization results of switching classification attributes are shown in Figure 6.

The system in this paper only switches the collar type code (V-neck, square collar, round collar) and generates results that precisely match the target collar type, while the other attributes remain stable and unchanged, verifying the strong decoupling capability of the FGSDE+ACIM architecture. The core of this method lies in: collar attributes are modeled using independent regression heads, whose output factors are directed into the sixth layer of StyleGAN3 (corresponding to the upper body structure generation stage); local modulation of AdaIN parameters prevents global style interference; simultaneously, a frozen ResNet50 backbone extracts high-level semantics, ensuring that collar features are not contaminated by gradients of other attributes. This paper achieves fully automatic, high-fidelity, and semantically aligned fine-grained control, providing a reliable technical path for personalized style generation.

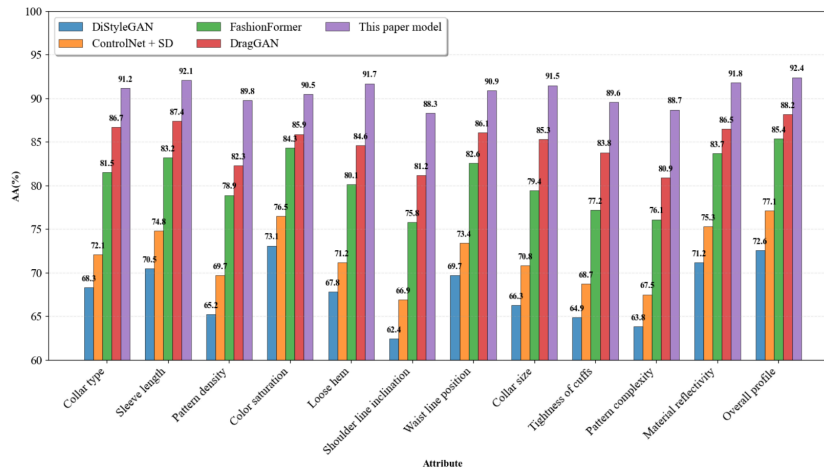


FIGURE 5. Attribute control accuracy



FIGURE 6. Visualization of categorical attribute switching

C. BODY FITTING

The personalized clothing styles for different body types (slim and tall/well-proportioned/full-figured) are shown in Figure 7.

Figure 7 shows that when the proposed system inputs different body codes (slim and tall/well-proportioned/full-figured), it can adaptively adjust the clothing structure: the slim and tall model automatically stretches the shoulder width and hem ratio; the full-figured model intelligently relaxes the waistline and armholes; the well-proportioned model maintains the original cut, and non-structural attributes such as texture, pattern, and collar type also remain consistent, verifying the effectiveness of the Pose-Aware Latent Warping mechanism. The core of this mechanism is that the body code is generated using an MLP to generate a latent space displacement field. Semantic deformation is applied through differentiable grid sampling, rather than pixel-level

distortion, ensuring structural coherence and lossless detail during the deformation process. By operating in the latent space, this paper decouples geometric adaptation from style fidelity, enabling end-to-end, high-fidelity, and automated body customization to meet the personalized needs of real users.

D. ABLATION EXPERIMENT RESULTS

The ablation experiment results are shown in Table 5.



FIGURE 7. Generation results for different body codes

TABLE 5. Ablation experiment results

Model	FID	KID	AA (%)	LPIPS	SSIM
This paper model	12.1	1.73	89.3	0.14	0.82
w/o FGSDE	16.9	2.41	62.4	0.15	0.80
w/o ACIM	15.7	2.28	75.2	0.16	0.79
w/o Pose-Aware Latent Warping	13.8	1.92	88.7	0.29	0.73
StyleGAN3	18.9	2.81	52.1	0.32	0.68

Note: w/o denotes without.

Ablation experiments validate the necessity and synergistic effects of the three core modules in this paper. Removing FGSDE causes a sharp drop in AA to 62.4%, while FID increases to 16.9, demonstrating that its 12-way parallel head is key to achieving attribute decoupling and precise control. Without it, the system degenerates into coarse-grained label control, resulting in severe local attribute drift. Removing ACIM reduces AA to 75.2% and FID to 15.7, reflecting the strong coupling between its dual-path injection mechanism (body $\rightarrow$ deformation, style $\rightarrow$ AdaIN) and user preferences and the generation process. Without it, the conditional signals fail to semantically align the latent space layers. Removing Pose-Aware Latent Warping, while having a minimal impact on AA/FID (AA = 88.7%, FID = 13.8), degrades LPIPS to 0.29 and SSIM to 0.73, demonstrating that warping effectively maintains texture and structural consistency in the latent space. Without it, the system relies

on pixel-level adaptation, resulting in fragmented details. All three modules are essential: FGSDE ensures precise control; ACIM ensures tight coupling; Pose-Aware Latent Warping ensures realism. Together, they support the system's Pareto optimality in terms of quality, control, and adaptation.

E. SYSTEM PERFORMANCE COMPARISON

The proposed system is compared with other systems, and the system performance comparison is shown in Figure 8.

Notes: DragGAN requires manual selection of control points and has no standardized delay; delay refers to the end-to-end inference time (including preprocessing).

While maintaining generation quality (FID = 12.1) and control accuracy (AA = 89.3%), the proposed system (S9) achieves industrial-grade real-time performance: delay of 54ms (only 9ms higher than the lightweight S1) and throughput of 18.5 img/s, outperforming the high-precision but low-efficiency ControlNet+SD (S4 = 420ms), FashionFormer (S5 = 185ms), and LCM (S7 = 310ms). The system uses 11.2GB of video memory, slightly higher than the S1 (8.1GB), due to the parameters of the FGSDE and ACIM modules. Compared to the S8 (DragGAN), this system supports fully automatic batch generation and boasts 37 times the throughput, meeting the high-concurrency requirements of e-commerce and design platforms. This performance advantage stems from three factors: FGSDE uses a frozen ResNet50 to avoid training overhead; ACIM uses a lightweight MLP+AdaIN, eliminating the need for complex attention; Pose-Aware Latent Warping operates in the latent

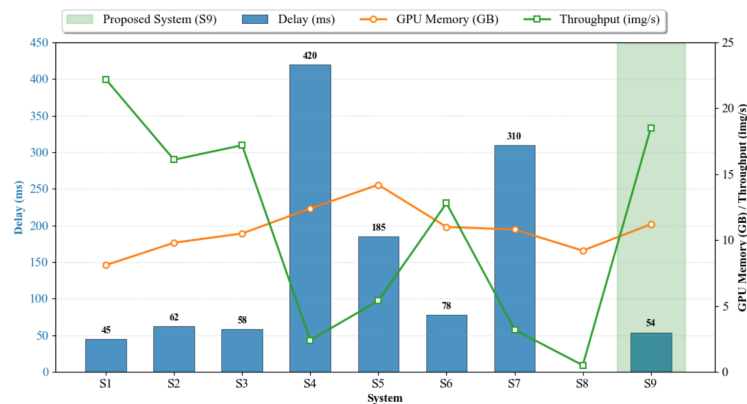


FIGURE 8. System performance comparison

space, making its computational complexity far lower than pixel-level deformation. The system achieves an optimal balance between quality, efficiency, and controllability.

Despite the introduction of FGSDE and ACIM modules, the system's graphics memory usage on A100 GPU is still within an acceptable range of 11.2GB. By combining asynchronous inference pipeline, memory reuse, and model lightweight technology, the system can maintain high-precision control while still meeting the real-time interaction requirements of industrial grade high concurrency and low latency. For example, in virtual fitting and e-commerce personalized recommendation scenarios, it can achieve second level response and batch generation.

V. CONCLUSION

Based on StyleGAN3, this study constructs an end-to-end generation framework integrating FGSDE, ACIM, and Pose-Aware Latent Warping, systematically addressing the challenges of style control, preference coupling, and body adaptation in clothing generation to enable a fully digital, personalized production workflow in the textile and garment manufacturing sector. The FGSDE module achieves precise decoupling and independent control of 12 categories of clothing attributes. The ACIM module ensures strong semantic alignment between user conditions and the latent space through a dual-path injection mechanism. The latent space deformation mechanism enables adaptive adjustment of clothing patterns while preserving texture details. The study proposes a clothing generation system that combines high generation quality (FID=12.1) with finegrained control (AA=89.3%), achieves industrial-grade deployment capabilities with low delay (54ms) and high throughput (18.5 img/s), and builds a scalable modular architecture, providing a technical paradigm for controllable generation across categories. This paper advances the evolution of AI generation technology from "capable generation" to the practical stage of "precise control, on-demand generation, and system application," which holds significant academic value and commercial application prospects, particularly in enabling efficient, highly customized textile and garment

production.

AUTHOR CONTRIBUTIONS

Na Xu designed, collected and analyzed the data, and drafted the manuscript. Na Xu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Na Xu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Q. Shi and R. Luo, "Research progress on clothing image generation based on generative adversarial networks," *Advanced Textile Technology*, vol. 31, no. 2, pp. 36, 2023.
- [2] J. Zhang, Y. Yu, and H. Yuan, "Construction of clothing and apparel knowledge graph and personalized recommendation," *Computer and Digital Engineering*, vol. 52, no. 5, pp. 1366-1372, 2024.
- [3] T. Cheng, "Establishment and optimization of personalized bra pattern," *Advanced Textile Technology*, vol. 30, no. 2, pp. 216, 2022.
- [4] J. Lin, M. Chen, Y. Shi, J. Lei, S. Guo, X. Gao, et al., "Personalized virtual clothing display for high-end customization," *Journal of Zhejiang University (Science Edition)*, vol. 48, no. 4, pp. 418, 2021.
- [5] H. Yan, H. Zhang, J. Shi, J. Ma, and X. Xu, "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, 2023, doi: 10.1145/3567596.
- [6] Q. Wu, B. Zhu, B. Yong, Y. Wei, X. Jiang, R. Zhou, et al., "ClothGAN: Generation of fashionable Dunhuang clothes using generative adversarial networks," *Connection Science*, vol. 33, no. 2, pp. 341-358, 2021, doi: 10.1080/09540091.2020.1822780.
- [7] H. Yan, H. Zhang, J. Shi, J. Ma, and X. Xu, "Inspiration transfer for intelligent design: A generative adversarial network with fashion attributes disentanglement," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 4, pp. 1152-1163, 2023, doi: 10.1109/TCE.2023.3255831.

- [8] Z. Yu and T. Luo, "Research on clothing patterns generation based on multi-scales self-attention improved generative adversarial network," *International Journal of Intelligent Computing and Cybernetics*, vol. 14, no. 4, pp. 647-663, 2021, doi: 10.1108/IJICC-04-2021-0065.
- [9] M. Singh, U. Bajpai, V. V. and S. Prasath, "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, 2020, doi: 10.1108/IJCST-12-2018-0148.
- [10] Y. Chen and J. Ma, "An intelligent generative method of fashion design combining attribute knowledge and Stable Diffusion Model," *Textile Research Journal*, vol. 95, no. 11-12, pp. 1231-1254, 2025, doi: 10.1177/00405175241289578.
- [11] J. Jung, H. Kim, and J. Park, "Deep Fashion Designer: Generative Adversarial Networks for Fashion Item Generation Based on Many-to-One Image Translation," *Electronics*, vol. 14, no. 2, pp. 220, 2025, doi: 10.3390/electronics14020220.
- [12] W. Shi, W. Wong, and X. Zou, "Generative AI in Fashion: Overview," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 4, pp. 1-73, 2025, doi: 10.1145/3718098.
- [13] A. Imtiaz, N. Pathirana, S. Saheel, K. Karunanayaka, and C. Trenado, "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, 2024, doi: 10.62411/faith.3048-3719-29.
- [14] H. Y. Cheng, C. C. Su, C. L. Jiang, and C. C. Yu, "Pose Transfer with Multi-Scale Features Combined with Latent Diffusion Model and ControlNet," *Electronics*, vol. 14, no. 6, pp. 1179, 2025, doi: 10.3390/electronics14061179.
- [15] H. Quan, S. Li, C. Zeng, H. Wei, and J. Hu, "Big data and AI-driven product design: a survey," *Applied Sciences*, vol. 13, no. 16, pp. 9433, 2023, doi: 10.3390/app13169433.
- [16] A. El Saddik, J. Ahmad, M. Khan, S. Abouzahir, and W. Gueaieb, "Unleashing creativity in the metaverse: Generative ai and multimodal content," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, no. 7, pp. 1-43, 2025, doi: 10.1145/3713075.
- [17] W. Z. Wang, H. M. Xiao, and Y. Fang, "Clothing image attribute editing based on generative adversarial network, with reference to an upper garment," *International Journal of Clothing Science and Technology*, vol. 36, no. 2, pp. 268-286, 2024, doi: 10.1108/IJCST-09-2023-0129.
- [18] P. Bhattacharya, D. Saraswat, D. Savaliya, S. Sanghavi, A. Verma, V. Sakariya, et al., "Towards future internet: The metaverse perspective for diverse industrial applications," *Mathematics*, vol. 11, no. 4, pp. 941, 2023, doi: 10.3390/math11040941.
- [19] H. Yan, H. Zhang, J. Shi, and J. Ma, "Mixing and matching elements for intelligent fashion design: A generative adversarial network with structure and texture disentanglement," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 358-370, 2023, doi: 10.1109/TCE.2023.3329574.
- [20] T. Fontanini and C. Ferrari, "Would your clothes look good on me? towards transferring clothing styles with adaptive instance normalization," *Sensors*, vol. 22, no. 13, pp. 5002, 2022, doi: 10.3390/s22135002.
- [21] H. Heim and C. Hopper, "Dress code: the digital transformation of the circular fashion supply chain," *International journal of fashion design, technology and education*, vol. 15, no. 2, pp. 233-244, 2022, doi: 10.1080/17543266.2021.2013956.
- [22] K. He, K. Yao, Q. Zhang, J. Yu, L. Liu, L. Xu, et al., "Dresscode: Autoregressively sewing and generating garments from text guidance," *ACM Transactions on Graphics (TOG)*, vol. 43, no. 4, pp. 1-13, 2024, doi: 10.1145/3658147.
- [23] Q. Hu, Y. Peng, J. Xu, Z. Shao, Z. Tian, J. Chen, et al., "Adaptive Stylized Image Generation for Traditional Miao Batik Using Style-Conditioned LCM-LoRA Enhanced Diffusion Models," *Mathematics*, vol. 13, no. 12, pp. 1947-1947, 2025, doi: 10.3390/math13121947.
- [24] T. Sha, W. Zhang, T. Shen, Z. Li, and T. Mei, "Deep person generation: A survey from the perspective of face, pose, and cloth synthesis," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1-37, 2023, doi: 10.1145/3575656.
- [25] B. AlBahar, J. Lu, J. Yang, Z. Shu, E. Shechtman, J. B. Huang, et al., "Pose with style: Detail-preserving poseguided image synthesis with conditional stylegan," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1-11, 2021, doi: 10.1145/3478513.3480559.
- [26] D. Zhou, H. Zhang, J. Ma, and J. Shi, "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, 2023, doi: 10.1109/TCSVT.2023.3318216.
- [27] M. Gu, Y. Chu, and X. Dong, "Fashion-tile: A tiled clothing image generation model based on improved conditional generative adversarial network," *International Journal of Clothing Science and Technology*, vol. 37, no. 3, pp. 495-516, 2025, doi: 10.1108/IJCST-05-2024-0115.
- [28] D. Zhou, H. Zhang, K. Yang, L. Liu, H. Yan, X. Xu, et al., "Learning to synthesize compatible fashion items using semantic alignment and collocation classification: An outfit generation framework," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5226-5240, 2022, doi: 10.1109/TNNLS.2022.3202842.
- [29] A. Melnik, M. Miasayedzenkau, D. Makaravets, D. Pirshtuk, E. Akbulut, D. Holzmam, et al., "Face generation and editing with stylegan: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 3557-3576, 2024, doi: 10.1109/TPAMI.2024.3350004.
- [30] T. Oike, H. Yamashita, and R. Shimizu, "Generation of Images with Many Likes Using Conditional StyleGAN2-ADA Considering User Attribute Information," *Industrial Engineering & Management Systems*, vol. 24, no. 1, pp. 21-28, 2025, doi: 10.7232/iems.2025.24.1.021.
- [31] A. Kalinin, A. A. Jafari, E. Avots, C. Ozcinar, and G. Anbarjafari, "Generative AI-based style recommendation using fashion item detection and classification," *Signal, Image and Video Processing*, vol. 18, no. 12, pp. 9179-9189, 2024, doi: 10.1007/s11760-024-03538-x.
- [32] J. Lin, X. Song, T. Gan, Y. Yao, W. Liu, L. Nie, et al., "PaintNet: A shape-constrained generative framework for generating clothing from fashion model," *Multimedia Tools and Applications*, vol. 80, no. 11, pp. 17183-17203, 2021, doi: 10.1007/s11042-020-09009-y.
- [33] J. Lee, J. Kim, and H. Lee, "Diffusedesigner: Sketch-based controllable clothing image generation," *Fashion and Textiles*, vol. 12, no. 1, pp. 26, 2025, doi: 10.1186/s40691-025-00426-x.
- [34] M. Kang, J. Kim, and S. Kim, "Unsupervised generation of fashion editorials using deep generative model," *Fashion and Textiles*, vol. 11, no. 1, pp. 4, 2024, doi: 10.1186/s40691-023-00367-3.
- [35] M. Fukuda, S. Kotaki, M. Nozawa, K. Tsuji, M. Watanabe, H. Akiyama, et al., "An attempt to generate panoramic radiographs including jaw cysts using StyleGAN3," *Dentomaxillofacial Radiology*, vol. 53, no. 8, pp. 535-541, 2024, doi: 10.1093/dmfr/twae044.