Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Controllable Generation of Embroidery Images Method Based on Diffusion Models

Yijia Fang

How to cite: Fang Y. Controllable Generation of Embroidery Images Method Based on Diffusion Models. Textile & Leather Review. 2026; 9:3849-3878.https://doi.org/10.31881/TLR.2026.3849

How to link: https://doi.org/10.31881/TLR.2026.3849

Published: 25 April 2026

Controllable Generation of Embroidery Images Method Based on Diffusion Models

Yijia Fang

School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, China
2023210602007@mails.zstu.edu.cn

Article
https://doi.org/10.31881/TLR.2026.3849
Published 25 April 2026

ABSTRACT

Embroidery patterns, as a traditional art form, carry profound cultural significance. With the development of digital technologies, how to effectively extract and preserve embroidery patterns has become an urgent issue. This paper proposes a controllable embroidery image generation method based on diffusion models, aimed at enhancing the structural consistency and color consistency of embroidery patterns. The method uses Flux.1-dev as the core framework, combined with Low-Rank Adaptation (LoRA) for efficient fine-tuning to learn high-frequency embroidery textures. It incorporates ControlNet-Canny and ControlNet-Color to impose explicit constraints on structure and color, while Quickshift clustering is used for region segmentation and mask construction to assist in local re-painting optimization. Experimental results show that the proposed method significantly improves the stability of pattern structure and the accuracy of color, providing a feasible solution for the digital embroidery image generation and synthesis.

KEYWORDS

image generation, diffusion model, embroidery images, LoRA, ControlNet

INTRODUCTION

Embroidery, as a textile art with a long history, not only demonstrates exquisite craftsmanship in its techniques but also embodies profound cultural connotations and values, and has now been recognized as an important form of intangible cultural heritage [1]. Traditional embroidery patterns often encapsulate national history, aesthetic preferences, and spiritual beliefs. Through meticulous stitching techniques and thread routing, vivid motifs are rendered on fabric, reflecting the richness of human creativity and modes of expression [2]. However, in the digital age, how to effectively preserve, inherit, and innovate embroidery art has become an urgent issue to be addressed [3].

In the early stages of digitizing embroidery patterns, the process largely relied on manual drawing and simple image processing techniques, making it difficult to automatically generate high-quality new patterns [4]. In recent years, several methods for digital synthesis of embroidery patterns have emerged, such as using convolutional neural networks for pattern segmentation and synthesis [5]. However, these traditional approaches [6-9] exhibit notable limitations. On the one hand, the generated stitch types are overly simplistic and cannot support combinations of multiple stitching techniques. On the other hand, color and texture distortions frequently occur, with generated images often failing to accurately match the colors of the original patterns, while fine-grained textures lack realism. Even with the introduction of deep learning models for image-to-image translation, issues such as color deviation and loss of detail persist in practical applications, making it difficult to faithfully reproduce the depth and texture inherent in embroidery patterns [10,11].

To address these challenges, researchers have proposed several solutions. For example, StyleGAN2, released by Karras et al. in 2020 [12], significantly improved the realism and resolution of synthesized images. Hu et al. [13] proposed a multi-stitch embroidery generation Generative Adversarial Network (GAN) model, MSEmbGAN, which incorporates a region-aware texture generation network and a color correction network, partially alleviating issues such as limited stitch diversity and color inconsistency. Nevertheless, due to the inherent limitations of GAN models—such as training instability and the difficulty of balancing diversity with fine-detail fidelity [14]—as well as unavoidable problems like mode collapse and challenging training and debugging processes [15,16], there remains considerable room for improvement in the overall quality of embroidery image synthesis.

However, the rise of diffusion models has brought new opportunities to image synthesis. In 2020, Ho et al. [17] proposed the Denoising Diffusion Probabilistic Model (DDPM), which generates images by progressively adding and then removing noise; the quality of the generated images is comparable to that produced by GAN-based methods. Subsequently, Nichol and Dhariwal et al. [18] further optimized the model architecture and sampling strategies, not only surpassing the best-performing GAN models of that time in unconditional image synthesis, but also demonstrating that diffusion models can achieve excellent performance in both image fidelity and generation diversity simultaneously.

Following this, a series of diffusion-based text-to-image generation systems have emerged, such as OpenAI's DALL·E 2 and Google's Imagen. Among them, Stable Diffusion enables high-resolution image synthesis by performing the diffusion process in latent space and has been widely adopted in fields such as creative design

and digital art [19]. Overall, neural networks—particularly diffusion models—have achieved remarkable success in image synthesis and provide entirely new solutions for the design and generation of complex visual effects.

On the other hand, since diffusion models adopt a probabilistic generative approach, they are able to cover a more complete data distribution, thereby achieving better performance in preserving the original colors and maintaining a consistent style of patterns [20]. In particular, after appropriate fine-tuning, diffusion models can significantly alleviate the problem of color deviation, allowing the generated embroidery images to remain highly consistent with the input images in terms of color.

Recent studies [21] have combined pre-trained diffusion models with LoRA, enabling embroidery style transfer under a one-shot setting. This approach has demonstrated superior performance over traditional style transfer methods, especially in achieving fine-grained disentanglement of style and content.

Existing studies have demonstrated that diffusion models can be effectively applied to the field of embroidery image generation. They not only leverage their generative capabilities to produce realistic embroidery textures, but also utilize conditional control mechanisms to preserve the original structure and color of images as much as possible. Therefore, this paper focuses on addressing structural and color-related issues in embroidery image generation and constructs a generation framework centered on the Flux.1 diffusion model. Flux.1 is a high-performance diffusion model proposed in recent years. It adopts Flow Matching as its training objective, replacing the step-by-step denoising process used in traditional diffusion models. As a result, it offers advantages such as more stable training and higher fidelity.

In addition to the core Flux.1 model, we employ a LoRA fine-tuning network to perform low-parameter adaptation of the diffusion model, enabling it to acquire prior knowledge of embroidery textures and ensuring that the generated patterns can realistically simulate embroidery texture and appearance. Within this framework, we further introduce a ControlNet-Canny branch, which incorporates the edge contours of the original input image as structural constraints into the diffusion process, thereby ensuring that the generated embroidery images maintain structural consistency with the original design.

Beyond Canny, a ControlNet-Color branch is also adopted, allowing the model to inherit the color distribution of the original input pattern and perform color correction on local regions during the generation stage, ensuring that color consistency does not deviate from the original. In addition, we utilize the Quickshift algorithm to segment the input image into regions based on color, identifying major color blocks and pattern

areas. On the one hand, the segmentation results are used to generate corresponding local inpainting masks, facilitating local enhancement in subsequent multi-stage diffusion sampling. On the other hand, these segmented regions support color statistics, which, combined with a region-based color refilling mechanism we design, further align the generated embroidery images with the original input in terms of color. Through targeted local repainting optimization in the later stages, we effectively improve the quality of fine textures and edge transitions, as well as the fidelity of color reproduction in the generated embroidery patterns.

By comprehensively integrating the above methods, the proposed diffusion-based embroidery pattern style transfer approach achieves significant improvements in both structural fidelity and color consistency in embroidery image generation, providing a practical and effective technical solution for the preservation, inheritance, and innovative design of digital embroidery patterns.

Overall, the main contributions of this paper can be summarized as follows:

• A dedicated LoRA training method for embroidery. First, dataset preprocessing is achieved through a combination of texture coverage-based filtering and mosaic reconstruction. This approach not only mitigates the adverse effects of small-sized samples on embroidery texture learning, but also enhances the density and consistency of texture information within the training dataset. In addition, the EMBLoRA model is trained using the OneTrainer framework, enabling it to learn and preserve high-frequency texture features in embroidery images, thereby providing strong support for subsequent style transfer.

• A controllable embroidery image style transfer method based on diffusion models. Built upon the Flux.1-dev framework and combined with EMBLoRA for task-specific fine-tuning, the method further incorporates ControlNet-Canny and ControlNet-Color into the generation process to constrain structure and color, thereby improving both structural consistency and color fidelity of the generated results.

• A mask-based method for preserving structure and color fidelity in embroidery images is proposed. The Quickshift algorithm is employed to perform region segmentation and construct masks, which are then utilized in the local inpainting process during the generation stage, thereby improving the representation of fine textures and edge transitions in embroidery patterns. Meanwhile, the masks are used to conduct color statistics and color refilling within segmented regions for local color correction. This approach further ensures structural fidelity and overall color consistency in the generated embroidery images.

RELATED WORK

In industrial practice, the typical workflow for the digital production of embroidery patterns generally includes "design drafting, region segmentation, stitch type and density configuration, and embroidery path generation." Although the use of commercial computer-aided design tools can improve pattern-making efficiency to a certain extent, key steps—such as selecting stitch types, planning stitching directions, and adjusting local density parameters—still rely heavily on manual experience. As a result, handling complex patterns and fine-grained regional boundaries often requires repeated iterations [22], making it difficult to simultaneously satisfy the demands for efficiency and consistency.

Moreover, the visual appearance of embroidery images is influenced not only by colors and pattern shapes, but also by factors such as high-frequency stitch textures, material reflectance, and transitions across region boundaries. As a result, traditional generation methods based on rules or simple image processing exhibit clear limitations when handling multi-stitch combinations, maintaining stable structural boundaries, and ensuring color consistency across regions.

Application of Image Style Transfer in Pattern Generation

The objective of style transfer research is to achieve a balance between preserving content structure and injecting stylistic texture. In the field of pattern design, it is evolving from explicit feature constraints toward implicit modeling with generative models. Wang et al. [23] proposed a method that explicitly extracts content information while allowing the model to implicitly learn style information. By leveraging diffusion models, this approach enables controllable disentanglement of content and style, providing a more interpretable and controllable solution for complex texture scenarios.

Meanwhile, a number of studies have explored diffusion-based editing-style transfer methods. For instance, Hertz et al. [24] introduced Prompt-to-Prompt, which controls cross-attention maps so that modifying text prompts alone can achieve both local and global controllable edits, effectively preserving the original structural layout. Brooks et al. [25] proposed InstructPix2Pix, which further formulates instruction-driven editing into a unified diffusion-based training paradigm, allowing image editing to better align with the iterative refinement process commonly required in design workflows.

GAN-based Embroidery Image Generation Methods

GAN has developed a relatively mature technical framework in texture synthesis and image generation. Karras et al. [26] conducted a systematic study on GAN training and high-quality generation under limited data

conditions, providing valuable training strategies and empirical foundations for data-scarce pattern generation scenarios. In the domain of embroidery image generation, Hu et al. [13] proposed MSEmbGAN, which focuses on region-aware texture generation. By integrating stitch-region modeling and color consistency design, it enables multi-stitch embroidery image synthesis and achieves improved texture realism compared with earlier methods.

However, from the perspective of the underlying generative mechanism, GANs still face challenges such as training instability, limited mode coverage, and insufficient fine-grained control over spatial structure. In particular, when tasks impose stricter requirements such as precise structural alignment, non-deforming edges, and controllable local texture enhancement, additional conditional constraints or post-processing mechanisms are often required to meet the deterministic demands of practical applications [27].

Development of Diffusion Models

Diffusion models achieve stable training and high-quality sampling through a process of progressively adding noise to data and then gradually removing it. Ho et al. [17] proposed the DDPM, which established the foundational framework for this approach and laid the groundwork for subsequent research directions. Following this, Dhariwal and Nichol et al. [20] conducted further investigations, demonstrating that diffusion models can match or even surpass contemporary GANs in image synthesis quality metrics, while also maintaining strong sample diversity and high-fidelity detail preservation. This makes them particularly well-suited for generating high-frequency textures and complex patterns. To balance generation quality and computational efficiency, Rombach et al. [28] introduced Latent Diffusion Models, which perform the diffusion process in a compressed latent space. This significantly reduces computational cost while preserving high-resolution generation capability, thereby making diffusion models more practical for design-oriented tasks that require detailed texture representation.

With the advancement of diffusion-based text-to-image systems in both engineering and modeling, the Flux family of models has emerged. According to official reports, Flux demonstrates improved performance in image detail, prompt adherence, and stylistic diversity, and provides multiple variants such as pro, dev, and schnell to balance capability and accessibility.

From publicly available information, Flux.1-dev adopts a rectified flow transformer architecture, with a model scale of approximately 12B parameters, and employs guidance distillation to improve inference efficiency and prompt-following performance. MLCommons, in its updated MLPerf Training text-to-image benchmark,

selected Flux.1 as a new-generation benchmark model, noting that compared with earlier benchmark models such as Stable Diffusion XL (SDXL, approximately 3.5B parameters), the new generation exhibits significant generational shifts in both architecture and scale. Therefore, Flux more accurately reflects the current state-of-the-art capability of text-to-image models.

Based on the above considerations, Flux offers advantages in terms of model generation, scale, prompt adherence, and fine-grained detail representation. Therefore, it provides a more suitable high-quality baseline for generating high-frequency textures, such as embroidery stitches, compared with diffusion models like SDXL. Moreover, it offers a more robust foundation for the subsequent incorporation of structural conditions and local editing methods.

Controllable Image Generation Based on Diffusion Models

Design-oriented generative tasks often require more precise spatial constraints than those provided by pure text prompts. As a result, introducing conditional control into diffusion models has become an important research trend. Zhang et al. [29] proposed adding a trainable control branch alongside a pre-trained diffusion model ControlNet. By explicitly injecting structural conditions such as Canny edges, depth maps, and segmentation maps into the diffusion process, the method significantly improves the model's adherence to input structures, providing a general solution for structure-consistent pattern generation tasks.

In terms of parameter-efficient fine-tuning, Hu et al. [30] proposed LoRA, which enables lightweight customization of large models through low-rank decomposition. This approach reduces the training cost for adapting specific styles and domains, making it possible to effectively personalize models even with limited data. In generative vision tasks, LoRA is also widely used to learn texture and material characteristics of target domains at low cost, enhancing the model's ability to represent specific craft textures (such as embroidery stitch patterns) without compromising its original generative capability.

For local editing tasks, Lugmayr et al. [31] introduced RePaint, which performs high-quality inpainting of arbitrarily shaped regions by enforcing consistency constraints on unmasked regions during the diffusion sampling process. This provides a transferable technical pathway for locally enhancing or repairing textures while preserving global structural integrity.

Notably, for high-frequency texture tasks such as embroidery, Ma et al. [21] further combined contrastive learning with LoRA-based modulation to achieve one-shot embroidery customization, emphasizing improved style capture and generalization ability under complex stitch structures.

However, from the perspective of embroidery pattern generation requirements, strict structural alignment and regional color consistency often need to be satisfied jointly. The aforementioned methods typically focus more on style learning, while there remains room for further improvement in achieving joint consistency of both structure and color.

In summary, traditional digital workflows for embroidery pattern design rely heavily on manual experience and repeated parameter tuning. Although learning-based methods have improved the level of automation, they still face challenges in expressing complex multi-stitch textures, maintaining strict structural fidelity, and ensuring color consistency. While GANs demonstrate advantages in texture generation, they often require additional mechanisms to address limitations in training stability and fine-grained structural controllability.

In contrast, diffusion models offer superior performance in terms of generation quality and training stability, and can be naturally integrated with techniques such as structural condition control, parameter-efficient fine-tuning, and local inpainting, making them more suitable for embroidery image generation tasks that are sensitive to both structure and texture. Meanwhile, studies on embroidery-specific characteristics generally indicate that relying on a single type of condition (either structure-only or style-only) is insufficient to meet practical requirements. Therefore, constructing a controllable workflow that integrates structural guidance, parameter-efficient adaptation, and local editing has become a common strategy for improving both generation stability and usability.

Accordingly, this work builds upon a diffusion-based framework by incorporating structural guidance (e.g., edge conditions), lightweight style adaptation via LoRA, and local region inpainting strategies, aiming to further enhance structural and color consistency in embroidery generation and to provide a more controllable technical pathway for digital embroidery pattern synthesis.

MATERIALS AND METHODS

Task Definition and Problem Formulation

This paper focuses on a controllable generation task for embroidery texture enhancement. The objective is to generate stylized results with embroidery stitch textures and fabric-like material appearance for the pattern regions, while preserving the structural contours and color distribution of the input image. Given an input image $x \in R^{H\times W\times 3}$, the goal is to generate an output image $y \in R^{H\times W\times 3}$, which satisfies structural consistency, color consistency, and prominent stylized texture characteristics. Specifically, structural consistency refers to the requirement that no significant deformation, repainting, or boundary expansion occurs within

the foreground pattern region in the output image. Color consistency indicates that the output maintains the same dominant color tones and region-wise color distribution as the input, avoiding noticeable color shifts introduced by the embroidery stylization process. The requirement of stylized texture ensures that the output exhibits stable and recognizable embroidery stitch patterns, such as warp-weft texture structures and high-frequency thread-like details. The overall framework of the proposed model is illustrated in the Figure 1.

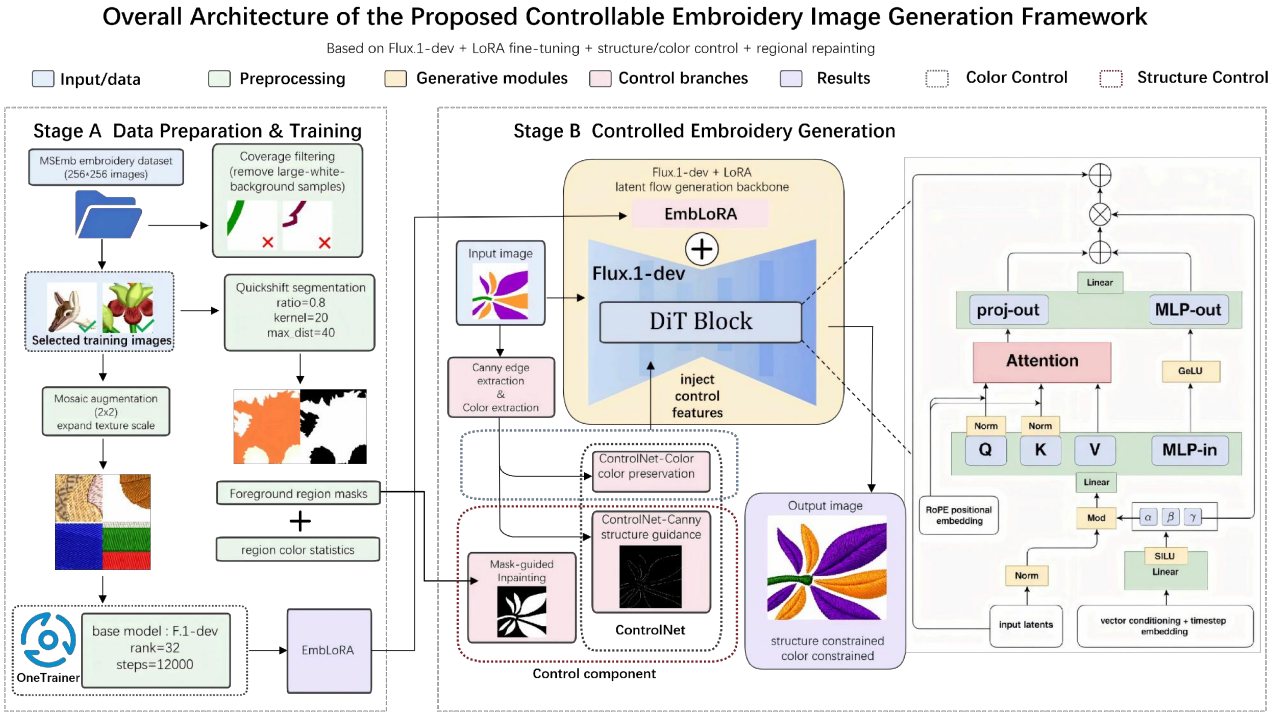

Figure 1. Model Architecture Diagram.

Dataset Preprocessing

The dataset used in this study is based on the embroidery image dataset constructed in the MSEmb. The original samples have a resolution of 256 × 256 and are subsequently split into 4,250 training images and 300 test images. To ensure that the model can sufficiently learn representative stitch textures and local structural features of embroidery images during LoRA fine-tuning, a two-stage preprocessing and reorganization procedure is applied to the original dataset prior to training.

First, to address the issue that some samples contain a low proportion of embroidery textures while background regions (e.g., large white areas) dominate, a pixel-coverage-based filtering strategy is designed. Specifically, the proportion of non-white pixels in each image is calculated, and samples with embroidery texture coverage below a predefined threshold are removed. This avoids interference from invalid or weak-

texture samples during LoRA fine-tuning. This step effectively increases the density and consistency of embroidery texture information in the training data, providing a more stable data foundation for subsequent style learning.

Second, considering that the original 256 × 256 images may lead to excessive compression of embroidery stitch details during LoRA training, thereby weakening the model's ability to learn high-frequency texture features, a mosaic-based data augmentation strategy is further introduced. Specifically, multiple embroidery images are randomly selected from the filtered dataset and concatenated in a 2 × 2 grid to construct composite samples with a resolution of 512 × 512. This strategy preserves the original texture distribution while effectively enlarging the visible scale and contextual range of embroidery patterns within a single training sample, enabling the LoRA model to learn stitch arrangement, texture orientation, and local repetitive structural characteristics within a larger receptive field.

The mosaic-based data augmentation strategy is used to combine multiple embroidery regions into a single sample to increase the proportion of effective high-frequency textures, rather than merely enlarging image resolution.

Through the combined preprocessing pipeline of texture coverage-based filtering and multi-scale mosaic reorganization, this study effectively mitigates the adverse impact of small-sized samples on embroidery texture learning. This process provides high-quality and structurally stable training data to support subsequent LoRA fine-tuning based on Flux.1-dev.

Region Mask Construction Based on Quickshift

The key to embroidery-style generation is that textures should be locally attached within the pattern or color regions, rather than spreading into background areas or crossing region boundaries. To this end, this paper first applies region segmentation to the input image to obtain a set of regions $\{R_r\}_{r=1}^{K}$, and constructs region-level masks as follows:

$$M_r(p) = \begin{cases} 1, p \in R_r \\ 0, \text{otherwise} \end{cases} \tag{1}$$

where p denotes a pixel location. In this study, Quickshift is adopted for segmentation with the parameters set to ratio = 0.8, kernel = 20, and max_dist = 40, producing multi-region results that capture color blocks and boundary structures.

It should be noted that the embroidery dataset employed in this study consists of pre-segmented patches and does not provide full semantic annotations for all pattern elements. Consequently, semantic segmentation methods are not feasible in this context. Quickshift is chosen because it can efficiently generate region-level masks based on color and local density in an unsupervised manner. These masks are then used to guide local repainting and color refilling, allowing the model to optimize fine-grained textures within each region while preserving overall structural consistency. This choice ensures that high-frequency embroidery textures are effectively captured, even without access to complete semantic labels.

Since the dataset background is pure white, to prevent background regions from being mistakenly treated as foreground textured areas, the mean color μ_r of each region is computed. If the distance between μ_r and [255, 255, 255] is below a predefined threshold τ, the region is regarded as background and removed. Finally, a set of foreground masks $\{M_r\}$ is obtained, and the unified foreground mask is defined as:

$$M = \bigcup_{r=1}^{K} M_r \tag{2}$$

where M is used for global structural and color constraint computation, while $\{M_r\}$ is used for regional color statistics and region-level consistency evaluation.

Embroidery Image Generation Model Construction

Flux.1-dev Model Architecture

In this study, Flux.1-dev is selected as the base generative model for embroidery image synthesis. Its core is a rectified flow transformer operating in the latent space, which can be regarded as a flow-matching implementation of the diffusion-based generative paradigm. It improves training stability and sampling efficiency by learning an approximately linear denoising trajectory from noise to the data distribution.

From an overall architectural perspective, the execution pipeline of Flux.1-dev can be summarized as text condition encoding, latent space representation, flow field prediction with iterative sampling, and latent space decoding. First, the input text prompt is encoded into a semantic condition vector. In publicly available baseline implementations, Text-to-Text Transfer Transformer–Extra Extra Large (T5-XXL) and Contrastive Language–Image Pre-training (CLIP)-based vision–language encoders are commonly used together to provide both robust semantic understanding and effective cross-modal alignment constraints, thereby improving long-text comprehension, object attribute binding, and fine-grained controllability.

Meanwhile, the target image is not modeled directly in the pixel space. Instead, a Variational Autoencoder (VAE) is used to compress the image into a low-dimensional latent space, producing a more compact latent representation. This significantly reduces the computational and memory cost of high-resolution generation, while providing a more suitable representation space for constructing multi-scale features of "texture, structure, and color" in subsequent modeling stages.

In the generation stage, the Transformer backbone takes the "text condition and current noise latent variable" as input, and outputs a velocity field that describes the evolution direction of the latent variable. This can be viewed as the flow-matching formulation of the denoising prediction objective in diffusion models. Through multi-step iterative updates, the initial random noise is gradually transported toward a target latent variable that satisfies the conditional distribution. Compared with the stepwise denoising process in conventional discrete diffusion models, rectified flow tends to model a more continuous generation trajectory, thereby achieving better inference efficiency and scalability while maintaining high generation quality.

After sampling is completed, the resulting target latent variable is decoded back into the pixel space via the VAE decoder to produce the final output image, forming an end-to-end conditional generation pipeline. This combination of "latent space generation and Transformer-based modeling" not only inherits the advantages of diffusion-style generation in terms of detail fidelity and mode coverage, but also leverages the Transformer's capability in modeling long-range dependencies and fusing multimodal conditions. As a result, it provides a higher-quality generative foundation for subsequent tasks in this study, particularly for embroidery-related high-frequency texture generation that requires strong structural consistency and fine-grained detail representation.

LoRA Fine-Tuning Network

Although Flux.1-dev has strong general-purpose image generation capabilities, embroidery images contain high-frequency stitch textures, thread orientations, and craft-like material appearances that are significantly different from natural images. Directly using the original pre-trained model often fails to stably reproduce embroidery-style visual characteristics in the target domain.

To inject embroidery-domain features efficiently without compromising the general generative capability of the base model, this paper adopts LoRA to fine-tune key linear projection layers of Flux.1-dev in a parameter-efficient manner. Specifically, the original weights are frozen while only a low-rank residual branch is trained,

which significantly reduces the number of trainable parameters and memory consumption, and improves convergence efficiency and controllability under small- to medium-scale training data.

The core idea of LoRA is that, during fine-tuning of large models, the original weight matrix W_0 is not directly updated. Instead, a low-rank weight update ΔW is learned, such that the adapted mapping can be expressed as:

$$W = W_0 + \Delta W \tag{3}$$

Given a linear layer with input feature $x \in R^{d_{in}}$ and output $y \in R^{d_{out}}$, the original mapping is defined as $y = W_0 x$. LoRA constrains the weight increment $\Delta W \in R^{d_{out} \times d_{in}}$ to a low-rank decomposition form:

$$\Delta W = s \cdot BA, A \in R^{r \times d_{in}}, B \in R^{d_{out} \times r}, r \ll \min(d_{in}, d_{out}) \tag{4}$$

where r denotes the rank. Matrix A projects the input from d_{in} to a low-rank subspace, and B maps the low-rank representation back to the d_{out} -dimensional output space. The scaling factor s is typically set in practice as:

$$s = \frac{\alpha}{r} \tag{5}$$

where α is a LoRA scaling hyperparameter used to balance the contribution of the incremental branch relative to the original mapping. Accordingly, the forward propagation after fine-tuning can be expressed as:

$$y = (W_0 + sBA)x = W_0 x + sB(Ax) \tag{6}$$

During training, only $\{A, B\}$ and optionally bias terms are optimized, while W_0 remains frozen. This enables rapid adaptation to the target domain's style distribution with a minimal number of trainable parameters, while reducing interference with the base model's semantic capabilities.

For the Transformer backbone of Flux.1-dev, this paper injects LoRA modules into the linear projection layers that have the most significant impact on generation quality and style representation. Typical insertion points include the Q, K, V, and O projection layers in the Self-Attention module, as well as the up-projection and

down-projection linear layers in the feed-forward network, in this context, it specifically refers to the Multi-Layer Perceptron (MLP).

The intuition behind this design is that attention projection layers determine global dependency modeling and semantic alignment across tokens, while MLP projection layers more directly affect the nonlinear reconstruction of local textures and fine-grained details. In embroidery generation tasks, stitch patterns and thread-level details are high-frequency local attributes that rely heavily on these representations. Therefore, introducing low-rank adaptations at these key positions enables improved learnability and transferability of embroidery-domain textures at a relatively low computational cost.

In terms of the training objective, Flux.1-dev follows a flow matching (continuous generative) paradigm, which can be formulated as learning a conditional velocity field (or an equivalent denoising direction prediction). Let the latent variable be z , the text condition be c, and time be t . The model predicts a velocity field $v_\theta(z_t, t, c)$, and the supervision signal is $v\backslash^*$ (constructed during training). The commonly used mean squared error (MSE) objective can then be written as:

$$L(\theta) = E_{z,t,c}[\| v_\theta(z_t, t, c) - v\backslash^* \|_2^2] \tag{7}$$

where θ denotes the set of trainable parameters. In the LoRA fine-tuning setting proposed in this study, θ consists only of the injected low-rank matrices $\{A, B\}$ and optional bias terms. This design concentrates the capacity for learning the embroidery style distribution into the low-rank incremental branch.

Through the above LoRA fine-tuning mechanism, the model is able to maintain its foundational generative capability and semantic consistency while more stably learning embroidery textures, stitch-like material properties, and craft-specific details. This provides essential support for achieving stylistic consistency in the subsequent generated results.

ControlNet Conditional Control Network

Although the base generative model has strong texture synthesis capability, text-only conditions are often insufficient to strictly constrain the generation process to follow the input structure in pattern-related tasks, which may lead to issues such as contour drift and local deformation. Meanwhile, at the color level, diffusion- or flow-matching-based generation may also introduce local color shifts and cross-region color bleeding during iterative sampling.

To introduce explicit prior constraints without modifying the core parameters of the base model, this paper adopts the ControlNet conditional control framework. During the generation process, structural conditions (Canny edges) and color conditions (color guidance) are injected separately, thereby enabling controllable generation with both structural consistency and color consistency.

Specifically, ControlNet takes a conditional image s as input and produces a control feature Δh , which is then fused with the backbone features to continuously guide the generation process during sampling, ensuring that the output satisfies the specified constraints:

$$h \leftarrow h + \lambda \cdot \Delta h(s) \tag{8}$$

where h denotes the intermediate features of the backbone network, $\Delta h(s)$ represents the control signal derived from the conditional map through the control branch, and λ is a control strength coefficient used to regulate the influence of the constraints. In this study, two types of conditions are used: a Canny-based structural condition and a color-preserving condition. For structural consistency, the Canny edges extracted from the input reference image x are used as structural priors:

$$s_E = C(x) \tag{9}$$

This condition is used to constrain the generated results to align with the contours and key boundaries, thereby reducing geometric distortions during embroidery texture overlay. For color consistency, a color condition is introduced to provide the color layout of the input pattern, enabling the generation process to preserve the original color distribution as much as possible. When region-wise statistics are used to enhance stability, a color hint map can be constructed using the segmentation label $\pi(u, v)$ and the region-wise mean color μ_k :

$$s_C(u, v) = \mu_{\pi(u,v)} \tag{10}$$

During inference, both structural and color controls can be enabled simultaneously, with their respective strength coefficients set independently:

$$h \leftarrow h + \lambda_E \Delta h_E(s_E) + \lambda_C \Delta h_C(s_C) \tag{11}$$

where λ_E emphasizes structural constraints and λ_C focuses on color preservation constraints. With the above formulation, the model is able to more stably preserve the original image's contour structure during embroidery texture generation, while reducing the risk of color deviation and cross-region color bleeding during the sampling process, thereby providing a controllable foundation for subsequent embroidery-style synthesis.

It is important to clarify the functional distinction between the region-based color refilling mechanism and the ControlNet-Color branch. While ControlNet-Color provides global guidance for overall color distribution during generation, the region-based color refilling serves as a local correction step. It adjusts colors within segmented regions to address local deviations, color drifting, or under-saturation. The refilling process is applied with controlled weighting to avoid conflicts or over-saturation, ensuring stable and accurate local color reproduction without altering the global color trend.

During the generation process, ControlNet-Color and the region-based color refilling mechanism work in a complementary manner. The ControlNet-Color branch enforces the global color tendency of the input pattern, maintaining the overall color distribution consistency. In contrast, the region-based color refilling performs local adjustments based on the segmentation masks derived from Quickshift, correcting small deviations in color saturation or hue within each region. This division of responsibilities ensures that global guidance and local refinement work together harmoniously, preventing color conflicts and over-saturation while improving overall color fidelity.

Structural Loss and Color Loss

To transform "structure preservation" and "color consistency" into quantifiable objectives, this study designs a structural loss L_{struct} and a color loss L_{color} , which are also used as quantitative evaluation metrics to assess the quality of outputs from different methods. In addition, these losses are employed for candidate selection: multiple candidate results $\left\{y^{(n)}\right\}$ are generated for the same input, and the one with the minimum joint loss is selected as the final output.

Structural Loss L_{struct}

Structure Preservation Emphasis: The output aligns with the input's contour boundaries within the foreground region. This paper employs a combination of intra-mask structural similarity and edge consistency:

$$L_{struct} = (1 - \text{SSIM } (x \odot M, y \odot M)) + \lambda_e \cdot \frac{1}{\mid M \mid} \parallel (E(x) - E(y)) \odot M \parallel_1 \quad (12)$$

where SSIM denotes the structural similarity index, $E(\cdot)$ represents the Canny edge operator, $\odot$ denotes element-wise multiplication, $\mid M \mid$ is the number of pixels within the mask, and λ_e is the weight of the edge term.

Color Loss L_{color}

Embroidery stylization often introduces color bias or color drift across regions. To better align with human visual perception, this study measures color differences in the CIELab color space and introduces region-wise statistical consistency:

$$L_{color} = \frac{1}{\mid M \mid} \sum_{p \in M} \Delta E(\text{Lab } (x_p), \text{Lab } (y_p)) + \lambda_s \sum_{r=1}^{K} (\parallel \mu_r(x) - \mu_r(y) \parallel_2 + \parallel \sigma_r(x) - \sigma_r(y) \parallel_2) \quad (13)$$

where $\Delta E(\cdot)$ denotes the color difference in the Lab space, $\mu_r(\cdot)$ and $\sigma_r(\cdot)$ represent the mean and standard deviation within the region mask M_r , respectively, and λ_s is the weighting factor for the region-wise statistical term.

Joint Criterion and Output Selection

By integrating both structural and color objectives, the joint criterion is defined as:

$$L = \lambda_{struct} L_{struct} + \lambda_{color} L_{color} \quad (14)$$

For the candidate set $\{y^{(n)}\}$, the final output is selected as:

$$y^* = \arg\min_n L(y^{(n)}) \quad (15)$$

thereby improving structural stability and color consistency in an interpretable manner without modifying the training procedure.

EXPERIMENTS AND RESULTS ANALYSIS

Experimental Setup

The experiments in this paper were conducted on a Windows 11 (64-bit) platform, equipped with a 13th Gen Intel Core i7-13700KF (3.40 GHz) processor, 64 GB of RAM, and an NVIDIA GeForce RTX 4070 Ti SUPER (16 GB) GPU. The deep learning framework used was PyTorch 2.8.0, with CUDA 12.8 for GPU acceleration.

The dataset used in this study is a preprocessed version of the MSEmb embroidery image dataset, consisting of approximately 4,250 training samples and 300 testing samples. The training data are organized in the form of 256 × 256 image patches. During the preprocessing stage, the Quickshift algorithm is applied for color segmentation to generate region masks, and the mean color of each region is computed as a color prior (ratio = 0.8, kernel = 20, max_dist = 40).

The basic LoRA training parameters are shown in Table 1. The detailed configuration is as follows: the base model is Flux.1-dev; the LoRA rank is set to 32, α is 32, and the LoRA dropout is 0.05; the weight data type is bfloat16. The optimizer is AdamW with a learning rate of 5×10^{-5} , using a constant learning rate schedule with 500 warmup steps. The model is trained for 14 epochs, with a local batch size of 2 and accumulation steps of 1. The gradient clipping threshold is set to 1.0, and gradient checkpointing is enabled. Training stops at 12,000 steps. The final trained LoRA model is named EmbLoRA, which is subsequently used for controllable embroidery image generation and evaluation.

Table 1. Basic LoRA training parameter settings

Training parameters	values
Base Model	Flux.1-dev
LoRA α	32
LoRA dropout	0.05
Weight data type	Bfloat16
Dropout optimizer	AdamW
Epoch	14
Learn rate	5e-5
Batch size	2
resolution	512

Evaluation of Embroidery Generation Results

To comprehensively evaluate the effectiveness of the proposed embroidery controllable generation method, this section analyzes the generated results from both subjective visual assessment and objective metric evaluation perspectives.

Specifically, the subjective visual evaluation is conducted by presenting representative samples generated during the experiments along with their local details, followed by visual comparisons of these results. The assessment focuses on whether the contour structure is preserved, the strength and realism of the embroidery texture, and whether the overall color scheme remains consistent with the content image.

On the other hand, the objective evaluation introduces quantitative metrics to measure the structural consistency and color consistency between the stylized images and the content images, yielding direct and interpretable numerical results. In addition, comparisons with other generation methods are conducted to intuitively demonstrate the overall performance of the proposed method in terms of stability and controllability.

Subjective Visual Evaluation

For the subjective evaluation of embroidery controllable image generation, this study primarily analyzes four aspects: structural consistency, color consistency, embroidery texture quality, and artifacts and stability.

In the literature, structural consistency generally refers to whether the contour, spatial position, and key boundary information of the generated result remain consistent with the input pattern. Color consistency emphasizes the preservation of hue and brightness, with the aim of avoiding color bleeding between adjacent regions. Embroidery texture quality focuses on the naturalness of stitch patterns, the rationality of stitch direction and arrangement, and whether the result exhibits the fiber-like and thread-level layering characteristic of embroidered works. Artifacts and stability refer to whether the generated images contain redundant or irrelevant structures, noise, local blurring, or any semantic inconsistencies with the input.

The comparison of embroidery generation results is shown in Figure 2, where (a) is the input image, and (b)–(g) are the results generated by Qwen_Image, Stable Diffusion XL (SDXL), OpenAI's DALL·E 3, Midjourney, NANO BANANA, and the proposed model, respectively. As can be observed, the image generated by Qwen_Image in (b) exhibits relatively realistic stitch structures; however, it performs poorly in terms of content preservation and color reproduction, leading to weak structural consistency and noticeable color deviations.

As shown in (c), the images generated by SDXL demonstrate a certain degree of stability in both structural and color consistency. However, it fails to effectively simulate the characteristic texture of embroidery, making it difficult to produce a recognizable embroidery style. The result from DALL·E 3 in (d) maintains a reasonable level of consistency in overall content structure, and the color reproduction is also relatively faithful. Nevertheless, the generated image lacks clear thread-like stitch characteristics typical of real embroidery, instead exhibiting a blurred, wool-like texture, which does not reflect the distinct linear texture of traditional embroidery.

The results from Midjourney in (e) are stable in both color and structural consistency, but the overall embroidery appearance is overly flat, lacking depth and layered texture, with insufficient texture clarity. In (f), the NANO BANANA results show structural disorder, with irregular and unstructured texture directions, leading to a coarse and chaotic visual appearance. However, this method demonstrates relatively strong color preservation, and the overall color scheme remains consistent with the input image.

In contrast, the proposed method in (g) achieves the best overall performance. It consistently outperforms the other compared methods in terms of color matching accuracy, fine-grained embroidery texture representation, and preservation of the main structural content of the input image.

Figure 2. Embroidery patterns generated by different models.

Objective Metric Evaluation

To further quantitatively evaluate the embroidery image generation performance, this study adopts Structural Similarity (SSIM), color histogram distance (CHD), and Learned Perceptual Image Patch Similarity (LPIPS) as the primary objective metrics, measuring performance from three aspects: structural consistency, color consistency, and texture fidelity, respectively.

SSIM is widely used to measure the similarity between two images in terms of structure, texture, and luminance. In this study, SSIM is adopted to quantitatively evaluate the structural consistency between the generated embroidery images and the content images, particularly in terms of contour preservation, boundary alignment, and fine-grained details. In general, a higher SSIM value (closer to 1) indicates greater structural similarity between the two images.

To effectively evaluate the color distribution consistency between the generated images and the original content images, this study employs the CIEDE2000 color difference formula to compute color discrepancies. CIEDE2000 is one of the most commonly used metrics for color evaluation, with the advantage of better aligning with human visual perception of color differences. When calculating the color difference between two images, this metric jointly considers variations in lightness, chroma, and hue, and applies corresponding perceptual corrections across different color regions. A smaller value indicates greater similarity in color between the two images.

For evaluating the texture quality of the generated images, this study adopts the LPIPS metric for quantitative assessment. LPIPS is computed by comparing the embroidery images generated by the proposed method with the style images provided in the dataset. LPIPS is a perceptual similarity measure based on deep learning networks, which evaluates image differences in a manner consistent with human visual perception. It is particularly effective in capturing differences in texture features, fine-grained details, and image blur. Specifically, a lower LPIPS value indicates higher similarity in texture characteristics between the compared images.

By jointly employing these three metrics for comprehensive evaluation and comparison, we can systematically analyze the performance of the generated embroidery images in terms of structural preservation, color fidelity, and texture quality. This further enables a clear comparison of the advantages and limitations of the proposed method against other approaches. The comparative results of different methods under these metrics are presented in Table 2.

Table 2. Quantitative Metrics of Various Image Generation Models.

Method	SSIM ↑	CIEDE2000 ↓	LPIPS↓
Our method	0.3328	2.8211	0.2495
SDXL	0.3019	5.3244	0.3495
DALL E3	0.2803	4.8596	0.5465
NANO BANANA	0.2495	4.1636	0.4230
Midjourney	0.3274	3.1693	0.3676
Qwen-Image	0.2432	3.9286	0.5125

Ablation Study

To verify the effectiveness of each component in the proposed method, three ablation experiments are designed. These experiments focus on three key modules—namely the LoRA module, the Canny edge control module, and the color control module—to evaluate their respective impacts and contributions to the final generation performance.

For each ablation setting, comparisons with the baseline method are conducted to analyze the specific impact of each module on texture consistency, structural consistency, and color consistency. The following sections will describe the detailed configurations of each experiment and present as well as analyze the corresponding ablation results.

Contribution of LoRA Fine-tuning

To verify the contribution of LoRA fine-tuning to the final generation performance, we remove the LoRA module from the generation model and observe its impact on texture realism and overall image generation quality.

In the experimental setup, two groups are defined: the control group uses only the base model Flux.1-dev without any LoRA fine-tuning, while the experimental group builds upon the same base model Flux.1-dev with the additionally trained EmbLoRA fine-tuning module incorporated.

The results of the ablation study are shown in the Figure 3. It can be observed that the LoRA fine-tuning module plays a significant role in the experimental group, substantially improving the texture details and embroidery-like appearance of the generated images. In the control group without LoRA fine-tuning, the generated embroidery images exhibit significant deficiencies in texture hierarchy and fine-detail preservation, especially in high-frequency texture regions and fine structural areas. In contrast, with LoRA fine-tuning, the generated textures become more natural and coherent, effectively reducing local blurring and distortion.

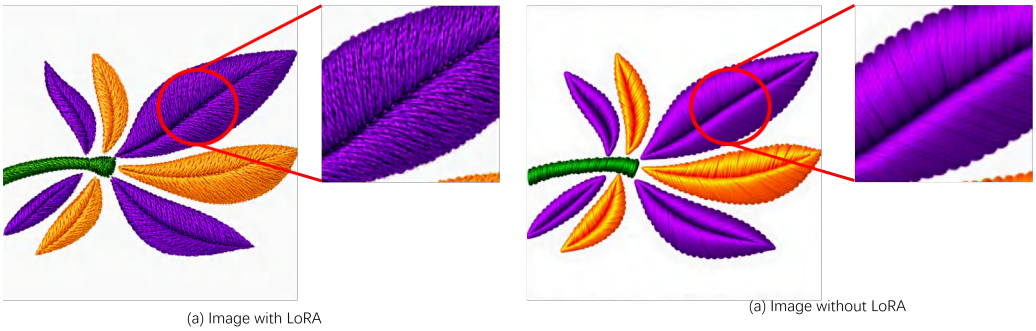

Figure 3. Comparison of LoRA Ablation Experiments.

Contribution of the Canny Edge Control Module

To verify the contribution of the Canny edge control module to structural consistency, we define a control group and an experimental group. The control group uses the base model Flux.1-dev without Canny structural control and without mask-guided local inpainting. The experimental group is also based on Flux.1-dev, but incorporates Canny edge control and mask-guided local inpainting.

The results are shown in the Figure 4. Structural consistency is evaluated using the SSIM metric, and the results indicate that the Canny edge control module significantly improves the stability of generated image contours. In particular, it effectively reduces boundary drift and structural distortion in complex edges and fine-detail regions. When the Canny control is removed, the generated images commonly exhibit contour shifts and blurred boundaries. This issue is especially evident in texture-rich regions, where boundary ambiguity degrades the overall generation quality.

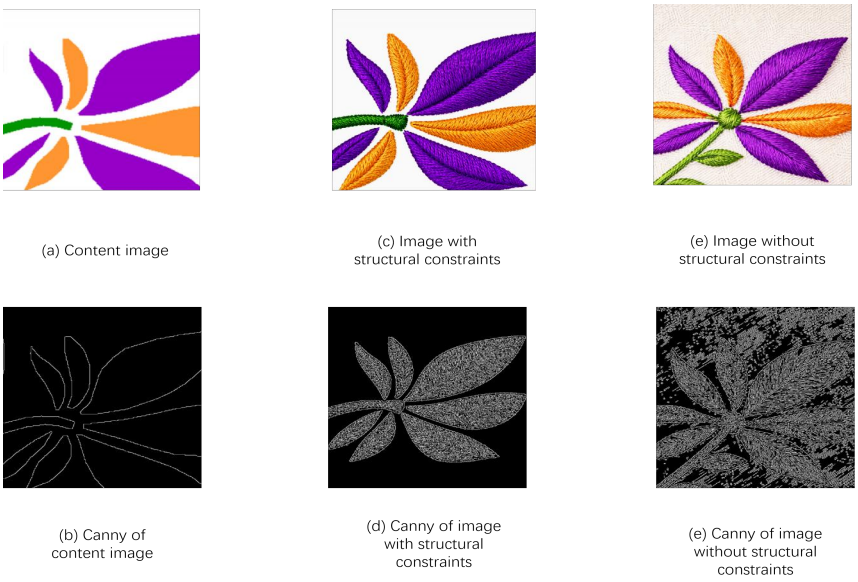

Figure 4. Comparison of Structural Constraint Ablation Experiments.

Contribution of the Color Control Module

To verify the contribution of the color control module to color consistency, we define a control group and an experimental group. The control group uses the base model Flux.1-dev without any color control, while the experimental group is also based on Flux.1-dev but incorporates ControlNet-Color for color preservation guidance.

The results are shown in the Figure 5. Color consistency is evaluated using the CIEDE2000 color difference metric. The results demonstrate that the color control module effectively reduces color shifts and luminance drift in the generated images. In particular, it significantly improves color fidelity in regions with adjacent multi-color patches and large uniform color areas. In the absence of color control, the generated images commonly exhibit color bleeding between regions and uneven color block distribution.

Through the ablation studies, we validate the importance of the three modules—LoRA fine-tuning, Canny edge control, and color control—in the overall generation performance. The experimental results demonstrate that LoRA fine-tuning effectively improves texture fidelity and enhances fine-detail quality, particularly in high-frequency textures and small structural regions. The Canny edge control module significantly improves structural consistency by reducing boundary drift and ensuring contour stability during image generation. The color control module effectively maintains color consistency, alleviating color bleeding issues in multi-region compositions and ensuring overall color stability.

(a) Content image (b) Image with color constraints (c) Image without color constraints

Figure 5. Comparison of Color Constraint Ablation Experiments.

DISCUSSION

This study presents a novel approach for generating embroidery images using a diffusion model, Flux-1, which effectively improves both the structural consistency and color accuracy of the generated designs. The results demonstrate significant advancements over traditional methods, especially in addressing common issues like texture fidelity and color mismatch.

Our approach, which integrates LoRA fine-tuning with ControlNet's edge detection and color control, has proven effective in enhancing the realism of embroidery patterns. The LoRA fine-tuning process contributes to the preservation of high-frequency texture details, particularly in intricate areas such as stitch directions, which were challenging for prior methods, including GAN-based models. This improvement is especially noticeable in high-frequency stitching areas, where the texture appears more natural and less prone to distortion. These findings suggest that LoRA is a powerful tool for learning domain-specific features like embroidery stitch patterns without compromising the overall model's generalizability.

The addition of ControlNet, particularly the Canny edge control and color consistency mechanisms, further stabilizes the generation process. These enhancements ensure that the generated images maintain their structural integrity, which is vital for tasks requiring precise shape and boundary preservation. This aspect is crucial for embroidery, where the accuracy of contours is integral to the final output. Moreover, the color control mechanism effectively mitigates color shifts, ensuring that the generated images retain the same tonal and color block integrity as the original designs, as verified through objective color difference metrics.

One notable feature of this work is the introduction of Quickshift for image segmentation, which allows for accurate region-based color distribution analysis. This segmentation not only improves color consistency but also aids in creating localized masks for texture enhancement during the multi-stage diffusion process. By focusing on specific regions, this technique enhances the details and edge transitions, which are often challenging in complex embroidery patterns.

The combination of these techniques positions our method as a reliable and efficient tool for digital embroidery design, overcoming many limitations faced by traditional methods. However, there are still areas that could benefit from further exploration. For example, while the structure and color consistency have significantly improved, the challenge remains in modeling more complex stitching styles and their interaction with underlying fabrics. Future work could focus on refining the model to handle various stitch types and material properties, enhancing its versatility and applicability in real-world scenarios.

In conclusion, the proposed approach represents a step forward in embroidery image generation by combining cutting-edge diffusion models with targeted fine-tuning and control mechanisms. It opens new possibilities for digital design in the textile industry, particularly in preserving cultural heritage through digital means. Further research will focus on improving model scalability and introducing new embroidery techniques to enrich the generated outputs.

CONCLUSIONS

This paper addresses common issues in embroidery image generation, including structural instability, color drift, and insufficient representation of embroidery textures, and proposes a controllable embroidery image generation method based on diffusion models. The proposed approach adopts Flux.1-dev as the base generative model and leverages LoRA for efficient parameter fine-tuning, enabling the model to learn characteristic high-frequency stitch texture features inherent in embroidery images. In addition, ControlNet-Canny and ControlNet-Color modules are incorporated to explicitly constrain structural contours and color distributions, respectively. This design not only preserves generation diversity but also significantly enhances structural consistency and color fidelity in the generated embroidery images.

To address the characteristics of embroidery images in terms of color composition and texture distribution, this paper introduces a Quickshift-based color segmentation strategy in both the data preparation and inference stages. During training, Quickshift is first applied to segment the image into distinct color regions, followed by statistical analysis of color feature information within each region, enabling the model to learn more stable prior knowledge of color composition. During inference, the segmented region masks are further utilized as guidance for a local inpainting process. This allows the generation pipeline to perform localized texture enhancement and embroidery-style reconstruction without altering the overall structure and global color distribution. As a result, the proposed method effectively suppresses common issues in diffusion-based free generation, such as structural misalignment and local color mixing.

In the experimental section, we conduct a comprehensive evaluation of the proposed method from both subjective visual assessment and objective metric analysis, comparing it with several mainstream generative models. The results demonstrate that our approach achieves more stable and superior performance in terms of structural preservation, color consistency, and embroidery texture quality. Furthermore, ablation studies are carried out to analyze the contributions of three key components: LoRA-based fine-tuning, Canny-based structural control, and color-based control. The findings confirm that these modules play complementary roles

across texture, structure, and color dimensions, and are all essential to the overall performance. In addition, the local inpainting constraint guided by color-region segmentation is shown to have a significant and positive effect on improving the structural stability of generated embroidery images.

Overall, this paper constructs a complete embroidery image generation pipeline based on a diffusion-based image generation framework, integrating data preprocessing, LoRA-based efficient parameter fine-tuning, structural and color condition control, and region-level inference constraints. This pipeline provides a highly controllable, relatively low-cost, and extensible technical solution for the re-digital design and intelligent generation of traditional embroidery patterns.

Although embroidery textures are inherently 3D, our method operates in 2D pixel space. The LoRA fine-tuning allows the model to capture directional and high-frequency embroidery-like patterns, but physically accurate stitch simulation is left for future work.

Building upon this work, future research will further explore more fine-grained modeling of embroidery stitch types and craft semantics, as well as investigate how to achieve stable generation under higher-resolution settings. In addition, we will explore multi-level constraint strategies based on regional semantics, with the aim of promoting broader practical applications of this approach in digital design and cultural creative industries.

Author Contributions

Conceptualization – Yijia Fang; methodology – Yijia Fang; formal analysis – Yijia Fang; investigation – Yijia Fang; resources – Yijia Fang; writing-original draft preparation – Yijia Fang; writing-review and editing – Yijia Fang; visualization – Yijia Fang; supervision – Yijia Fang. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

REFERENCES

[1] Yan WJ, Chiou SC. The safeguarding of intangible cultural heritage from the perspective of civic participation: The informal education of Chinese embroidery handicrafts. Sustainability. 2021; 13: 4958. doi: 10.3390/su13094958

[2] Guan Y. Digital extraction and modern translation of embroidery patterns from an archaeological perspective. Mediterr. Archaeol. Archaeom. 2025. doi: 10.1234/maa.2025.1695

[3] Liang Y, Xie B, Tan W, Zhang Q. Ontology-based construction of embroidery intangible cultural heritage knowledge graph: A case study of Qingyang sachets. PLoS ONE. 2025; 20: e0317447. doi: 10.1371/journal.pone.0317447

[4] Adiji BE, Ibiwoye TI. Effects of graphics and computer aided design software on the production of embroidered clothing in south Western Nigeria. Art. Des. Rev. 2017; 5: 230-240. doi: 10.4236/adr.2017.54019

[5] Chen X, McCool M, Kitamoto A, Mann S. Embroidery modeling and rendering. In Proceedings of the Graphics Interface 2012, Toronto, ON, Canada, 28-30 May 2012; pp. 131-139. doi: 10.5555/2305276.2305299

[6] Cui D, Sheng Y, Zhang G. Image-based embroidery modeling and rendering. Comput. Animat. Virtual Worlds. 2017; 28: 1-12. doi: 10.1002/cav.1725

[7] Baeva D. Using lindenmayer systems for generative modeling of graphic concepts, set in elements of Bulgarian folklore embroidery. In Proceedings of the International Conference on Computer Systems and Technologies (CompSysTech), Ruse, Bulgaria, 21-22 June 2019; pp. 234-239. doi: 10.1145/3345252.3345295

[8] Ma D, Cheng M, Zheng D, Fan X, Wang W, Fang J. Development of Sichuan brocade with imitating embroidery effect based on free-floats interlacing weave. J. Text. Sci. Technol. 2019; 6: 11-18. doi: 10.4236/jtst.2020.61002

[9] Guan X, Luo L, Li H, Wang H, Liu C, Wang S, Jin X. Automatic embroidery texture synthesis for garment design and online display. Vis. Comput. 2021; 37: 2553-2565. doi: 10.1007/s00371-021-02216-0

[10] Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016; pp. 2414-2423. doi: 10.1109/CVPR.2016.265

[11] Shu Y, Yi R, Xia M, Ye Z, Zhao W, Chen Y, Lai YK, Liu YJ. GAN-based multi-style photo cartoonization. IEEE Trans. Vis. Comput. Graph. 2022; 28: 3376-3390. doi: 10.1109/TVCG.2021.3067201

[12] Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13-19 June 2020; pp. 8110-8119. doi: 10.1109/CVPR42600.2020.00813

[13] Hu X, Yang C, Fang F, Huang J, Li P, Sheng B. MSEmbGAN: Multi-stitch embroidery synthesis via region-aware texture generation. IEEE Trans. Vis. Comput. Graph. 2024; 31: 5334-5347. doi: 10.1109/TVCG.2024.3447351

[14] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. IEEE Signal Process. Mag. 2018; 35: 53-65. doi: 10.1109/MSP.2017.2765202

[15] Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. arXiv 2018, arXiv:1802.05957. doi: 10.48550/arXiv.1802.05957

[16] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. arXiv 2018, arXiv:1809.11096. doi: 10.48550/arXiv.1809.11096

[17] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 2020; 33: 6840-6851. doi: 10.48550/arXiv.2006.11239

[18] Nichol AQ, Dhariwal P. Improved denoising diffusion probabilistic models. In Proceedings of the 38th International Conference on Machine Learning (ICML), Virtual Event, 18-24 July 2021; pp. 8162-8171. doi: 10.48550/arXiv.2102.09672

[19] Patricio A, Dehban A, Ventura R. FLORA: Efficient synthetic data generation for object detection in low-data regimes via fine-tuning Flux LoRA. arXiv 2025, arXiv:2508.21712. doi: 10.48550/arXiv.2508.21712

[20] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. Adv. Neural Inf. Process. Syst. 2021; 34: 8780-8794. doi: 10.48550/arXiv.2105.05233

[21] Ma J, He Q, He G, Chen H, Liu C, Jin X, Wang H. One-shot embroidery customization via contrastive LoRA modulation. ACM Trans. Graph. 2025; 44: 1-15. doi: 10.1145/3763290

[22] Zhang L, Li M, Zhang L, Liu X, Tang Z, Wang Y. Mastersu: The sustainable development of Su embroidery based on digital technology. Sustainability. 2022; 14: 7094. doi: 10.3390/su14127094

[23] Wang Z, Zhao L, Xing W. StyleDiffusion: Controllable disentangled style transfer via diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2-6 October 2023; pp. 7677-7689. doi: 10.1109/ICCV51070.2023.00706

[24] Hertz A, Mokady R, Tenenbaum J, Aberman K, Pritch Y, Cohen-Or D. Prompt-to-prompt image editing with cross attention control. arXiv 2022, arXiv:2208.01626. doi: 10.48550/arXiv.2208.01626

[25] Brooks T, Holynski A, Efros AA. InstructPix2Pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18-22 June 2023; pp. 18392-18402. doi: 10.48550/arXiv.2211.09800

[26] Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training generative adversarial networks with limited data. Adv. Neural Inf. Process. Syst. 2020; 33: 12104-12114. doi: 10.48550/arXiv.2006.06676

[27] Iglesias G, Talavera E, Díaz-Álvarez A. A survey on GANs for computer vision: Recent research, analysis and taxonomy. Comput. Sci. Rev. 2023; 48: 100553. doi: 10.1016/j.cosrev.2023.100553

[28] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18-24 June 2022; pp. 10684-10695. doi: 10.1109/CVPR52688.2022.01042

[29] Zhang L, Rao A, Agrawala M. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2-6 October 2023; pp. 3836-3847. doi: 10.1109/ICCV51070.2023.00355

[30] Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, Chen W. LoRA: Low-rank adaptation of large language models. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual Event, 2022. doi: 10.48550/arXiv.2106.09685

[31] Lugmayr A, Danelljan M, Romero A, Yu F, Timofte R, Van Gool L. RePaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18-24 June 2022; pp. 11461-11471. doi: 10.1109/CVPR52688.2022.01117