


Article

Sustainable Conservation of Embroidery Cultural Heritage: An Approach to Embroidery Fabric Restoration Based on Improved U-Net and Multiscale Discriminators

Qiaoling Wang, Chengge Jiang, Zhiwen Lu , Xiaochen Liu, Ke Jiang and Feng Liu

College of Textile Engineering, Taiyuan University of Technology, Jinzhong 030600, China; 2024510869@link.tyut.edu.cn (Q.W.); 2024521308@link.tyut.edu.cn (C.J.); 2023521331@link.tyut.edu.cn (X.L.); 2023521325@link.tyut.edu.cn (K.J.); liufeng@tyut.edu.cn (F.L.)

* Correspondence: luzhiwen@tyut.edu.cn; Tel.: +86-136-0358-9211

Abstract

As a vital carrier of China's intangible cultural heritage, restoring damaged embroidery fabrics is essential for the sustainable preservation of cultural relics. However, existing methods face persistent challenges, such as mask pattern mismatches and restoration size constraints. To address these gaps, this study proposes an embroidery image restoration framework based on enhanced generative adversarial networks (GANs). Specifically, the framework integrates a U-Net generator with a multi-scale discriminator augmented by an attention mechanism and dual-path residual blocks to significantly enhance texture generation. Furthermore, fabric damage was classified into three categories (hole-shaped, crease-shaped, and block-shaped), with complex patterns simulated through dynamic randomization. Grid-based overlapping segmentation and pixel fusion techniques enable arbitrary-dimensional restoration. Quantitative evaluations demonstrated exceptional performance in complex texture restoration, achieving a structural similarity index (SSIM) of 0.969 and a peak signal-to-noise ratio (PSNR) of 32.182 dB. Complementarily, eye-tracking experiments revealed no persistent visual fixation clusters in the restored regions, confirming perceptual reliability. This approach establishes an efficient digital conservation pathway that promotes resource-efficient and sustainable heritage conservation.

Keywords: embroidery restoration; generative adversarial networks; multi-scale discriminator; image segmentation and combination; eye tracking technology



Academic Editors: Douglas O'Shaughnessy and Junseop Lee

Received: 18 July 2025

Revised: 12 September 2025

Accepted: 23 September 2025

Published: 25 September 2025

Citation: Wang, Q.; Jiang, C.; Lu, Z.; Liu, X.; Jiang, K.; Liu, F. Sustainable Conservation of Embroidery Cultural Heritage: An Approach to Embroidery Fabric Restoration Based on Improved U-Net and Multiscale Discriminators. *Appl. Sci.* **2025**, *15*, 10397. <https://doi.org/10.3390/app151910397>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Embroidery, traditionally termed huaxiu (flower embroidery), is an ancient Chinese needlework technique. Beyond its intricate craftsmanship, this art form serves as a vital material carrier of China's intangible cultural heritage [1], embodying profound cultural significance and exceptional artistic and aesthetic value. Specifically, embroidery manifests diverse cultural themes: while the four major regional styles (Suzhou, Cantonese, Sichuan, and Hunan embroidery) dominate, distinct variations such as Beijing, Shandong, Bian, Ou, and Hangzhou embroidery exhibit unique characteristics. Crucially, motifs centering on blessing prayers and ethical education permeate folk embroidery patterns, reflecting traditional societal beliefs and philosophies. However, whether preserved heirlooms or archeologically excavated artifacts, embroidery relics universally face degradation from temporal and environmental factors; the synergistic effects of natural aging, environmental

erosion, and biological deterioration inevitably cause fabric and decorative pattern damage [2]. Qing Dynasty embroideries exemplify this challenge, as extant pieces commonly demonstrate fading, dye migration, discoloration, thread fracturing, and fabric decay [3]. Such irreversible damage compromises artistic integrity, impedes historical interpretation, and hinders long-term conservation.

Currently, the restoration of textile artifacts, particularly embroideries, remains predominantly manual owing to the required high precision, intricate craftsmanship, and demanding professional expertise. However, traditional restoration methods, which are heavily reliant on practitioner experience, exhibit significant limitations: Restoration cycles often extend for several months to years; sourcing substitute materials (e.g., silk threads, dyes) that precisely match the original artifacts in terms of period, provenance, and production techniques proves highly challenging; and the physical intervention itself is inherently invasive and irreversible. Once performed, such interventions are difficult to reverse and carry the inherent risk of unpredictable secondary damage to the artifacts [4,5]. Consequently, restoring exquisite embroidered artifacts, especially those featuring complex patterns and subtle color gradations, is not only a prolonged and resource-intensive endeavor but also poses risks. Physical handling and environmental fluctuations during restoration can further compromise the fragility of artifacts [6].

Considering the prevalent damage types on embroidered artifacts, including localized deterioration, structural creases, and severe fading or discoloration, there is an urgent need for non-contact, efficient, and reversible auxiliary or alternative restoration approaches. In this context, deep learning-based image restoration technology offers a novel pathway for the restoration of artifact images. This approach enables high-fidelity digital capture and image restoration of cultural artifacts, maximizing the recovery of original colors, patterns, and intricate details without physical contact with the artifacts. This generates high-quality reference images. These outputs provide crucial visual references for physical restoration, such as accurate color matching and pattern completion, thereby assisting conservators in formulating more precise and lower-risk intervention strategies [7]. Critically, this technology circumvents the risks associated with direct physical intervention, significantly enhancing the efficiency and safety of conservation practices. Therefore, it constitutes key technological support for advancing the sustainable preservation, research, and revitalization of the embroidered cultural heritage [8].

Traditionally, image restoration techniques have been classified into conventional methods and deep learning-based approaches. Conventional methods are further subdivided into diffusion- and exemplar-based methods [9,10]. Diffusion-based methods operate on the core principle of gradually propagating (smoothing) pixel information from intact regions into damaged areas to fill in the missing content. Typically grounded in physical diffusion models, these methods effectively preserve edge and texture structures, making them suitable for repairing small-scale damage, such as scratches and noise. The BSCB model proposed by Bertalmio et al. [11] pioneered diffusion-based inpainting using partial differential equations and achieved small-defect restoration through isophote-driven propagation. However, it struggles with complex textures and large defects. Conversely, exemplar-based methods, exemplified by the work of Criminisi et al. [12], focus on searching intact image regions for the most suitable texture patches and copying or blending them into the target area. This approach proves effective for large-scale damage repair (e.g., object removal, extensive occlusions) and restoring complex textures like grass, brick walls, or hair. Although the Criminisi algorithm significantly enhances large-area restoration through its patch-matching strategy, it suffers from a critical limitation: confidence values rapidly decay to zero. This decay causes priority computation to fail, leading to erroneous filling orders and frequently resulting in semantically inconsistent outcomes. To address

this, Cheng et al. [13] optimized the algorithm by redesigning the priority function, which yielded notable performance improvements. Although conventional image restoration methods leverage both the global features and local information of the image [14] and achieve reasonable results in some scenarios, they exhibit significant limitations when confronted with complex damage patterns or extensive information loss.

Advancements in computer vision have propelled deep-learning-based image inpainting methods to the forefront of research. By learning the texture, color, and structural features from datasets, these methods enable accurate and plausible reconstruction of complex images, thereby enhancing restoration fidelity and efficiency [15]. Among the various image inpainting methods, those based on Generative Adversarial Networks (GANs) have achieved remarkable progress in this field. Generative Adversarial Nets were first proposed by Ian Goodfellow et al. [16] in 2014. The core idea involves a “game” between a generator and a discriminator to accomplish image inpainting [17]. This approach provides a novel generative model framework for image restoration, effectively avoiding the blurred results often produced by traditional methods. However, the training of the original GAN is highly unstable and prone to mode collapse. In 2014, Mehdi Mirza et al. [18] proposed the Conditional Generative Adversarial Network (CGAN), which can be applied to various image translation tasks. Although highly flexible, CGAN requires a large amount of labeled data, posing a significant challenge in scenarios with limited data availability. In 2015, Radford et al. [19] successfully incorporated convolutional neural network architectures into the GAN framework (DCGAN) for the first time, substantially improving the quality and stability of generated images. Nonetheless, this method still suffers from issues such as transposed convolution artifacts, training instability, and resolution limitations. In 2016, Pathak et al. [20] introduced the “Context Encoder,” which employs an encoder–decoder architecture combined with adversarial training to predict the content of missing image regions. Martin Arjovsky et al. [21] proposed the Wasserstein GAN, which uses the Wasserstein distance instead of the JS divergence as a measure of distribution distance. This greatly enhances training stability and alleviates mode collapse. Subsequently, Iizuka et al. [22] introduced global and local discriminators to enhance semantic consistency, significantly improving inpainting quality. However, its performance in complex scenes remains sub-optimal. Yu et al. [23] proposed a contextual attention mechanism (CA) to achieve feature patch matching, further refining the inpainting results. Still, such methods often suffer from boundary artifacts and inadequate handling of complex edges. Liu et al. [24] introduced “Partial Convolution” and a “Mask Update Mechanism,” which effectively address the issue of color contamination during the training and inference stages when dealing with irregular holes. A drawback, however, is that for large-area missing regions or complex structures, the generated results may appear blurred and lack coherent structural details. Additionally, the training process is complex and computationally expensive.

Previous studies have contributed significantly to the advancement of generative adversarial network models. However, no prior work has focused specifically on model optimization or improvement tailored to the complex textures and delicate structures characteristic of embroidery images. To address this gap, this paper proposes a comprehensive workflow for the restoration of embroidered cultural heritage images, designed specifically according to the features of embroidered fabrics. The main contributions of this study are summarized as follows:

1. The proposed approach supports arbitrary combinations of three mask types—wormhole, crease, and block. During training, irregular masks are dynamically combined at random to enhance the model’s robustness in handling diverse real-world damage patterns, thereby preventing overfitting to any specific type of degradation.

2. An enhanced generative adversarial network is employed as the restoration model. The generator incorporates a U-Net architecture to mitigate issues arising from the limited sample size in embroidery image data. Through collaborative optimization with a multi-scale discriminator, the model achieves more accurate reconstruction of complex structural and textural details, which is essential for restoring intricate embroidery patterns.
3. A hybrid loss function combining adversarial loss, pixel-level L_1 loss, perceptual loss, and style loss is introduced for joint optimization. This composite loss is crucial for reconstructing the unique textural patterns and stylistic features of embroidery, encouraging the model to generate visually realistic and detailed images.
4. A complete workflow for embroidery image restoration is designed, encompassing mask generation, image segmentation, image inpainting, and image recomposition. Additionally, an eye-tracking experiment is conducted to quantitatively evaluate the visual authenticity of the restoration results. Experimental outcomes demonstrate that the proposed method exhibits notable advantages in restoring embroidery images.

The remainder of this paper is organized as follows. Section 2 details the architecture of the generative adversarial network model employed. Section 3 elaborates on the complete image inpainting pipeline, including mask generation, image segmentation, and image recomposition. Section 4 presents the experimental setup, results, and a comprehensive evaluation and analysis. Finally, Section 5 concludes the paper and suggests directions for future research in this field.

2. Methods

2.1. Generative Adversarial Network Architecture

The core mechanism of Generative Adversarial Networks involves adversarial training and continuous optimization of the generator and discriminator, ultimately enabling the generation of synthetic outputs that closely approximate real data distributions. When applied to embroidery image restoration, the GAN learns the intrinsic structural, textural, and chromatic characteristics of intact embroidery. This learned representation facilitates the generation of highly plausible content within damaged regions, thereby enabling high-fidelity reconstruction of missing structural details, texture patterns, and color information in degraded embroidery artifacts.

This study proposes an improved Generative Adversarial Network framework for deep learning-based embroidery image restoration. The framework employs a U-Net-structured generator [25] that is collaboratively optimized with a multi-scale discriminator. The generator integrates three key components: gated convolutions (GatedConv) [26], an enhanced self-attention mechanism incorporating positional encoding, and channel spatial attention residual blocks (CSARB). It reconstructs damaged regions through a four-level downsampling path followed by sub-pixel convolution (pixel shuffling) for upsampling, facilitated by cross-layer skip connections to achieve multi-scale feature fusion. The discriminator utilized a multi-scale PatchGAN architecture [27]. Equipped with instance normalization and LeakyReLU activation functions, it processes images concurrently at both the original resolution and a downsampled medium resolution (e.g., 1/2 scale) to provide collaborative discrimination. During training, a dynamic weighting strategy balances multiple loss functions: adversarial loss (MSE), pixel-level L_1 loss, VGG16-based perceptual loss, and Gram matrix style loss. This strategy was implemented alongside a progressive optimization scheme that employed learning rate decay. The architecture of the proposed GAN is illustrated in Figure 1. Experimental results demonstrate that this model achieves significantly superior performance compared to traditional methods in complex embroidery pattern restoration tasks. Importantly, this architectural design is not

only effective for embroidery image restoration but is also readily extendable to diverse textile cultural heritage digital restoration scenarios.

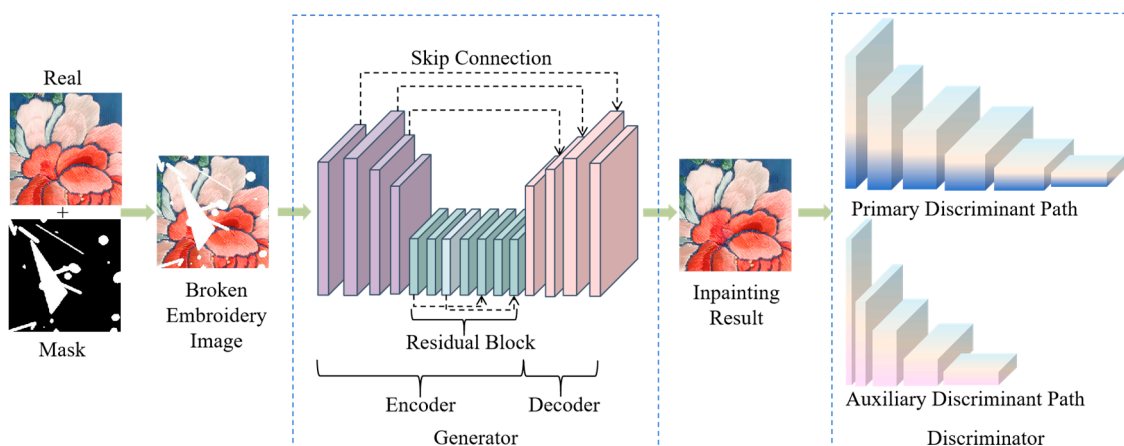


Figure 1. Improved generative adversarial network architecture.

2.2. Generator Structure

The generator constitutes the core of the generative model, and its performance directly determines the quality and diversity of the generated data. Given its advantages of being small-data-friendly, structurally simple, and easy to implement and adjust, the U-Net architecture has been widely adopted in image restoration tasks. Therefore, an improved encoder–decoder architecture was employed for the generator. Within the encoder, an initial convolutional layer performs primary feature extraction. Subsequently, three stages of downsampling were applied via convolutional layers with a stride of 2, progressively compressing the spatial dimensions and extracting the salient features of the image. Instance normalization and the rectified linear unit (ReLU) activation function were used in each layer. Six enhanced residual blocks were designed in the bottleneck layer. Feature extraction is further refined using two gated convolution structures, each combined with instance normalization. Additionally, a self-attention layer incorporating 1×1 convolutional positional encoding was inserted after the second residual block. Following further optimization of the feature representations via four gated residual blocks, a channel attention module comprising parallel global average pooling, global max pooling, and 1×1 convolutions was integrated at the end of each residual block. Within the decoder, the resolution is gradually upscaled through three stages using subpixel convolution (PixelShuffle). Skip connections link each stage to the corresponding feature map from the encoder. To specifically address large-area missing regions in the image, a context enhancement module consisting of two successive 3×3 convolutional layers is incorporated. Finally, the output layer produces a restored image matching the input size via a 3×3 convolution. Instance normalization and the ReLU activation function were applied to all layers except the output layer, which used the Tanh activation function. The structure of the generator is shown in Figure 2.

2.3. Discriminator Structure

The discriminator constitutes an essential component in generative adversarial networks, with its core task being to distinguish between real data and synthetic samples generated by the generator. Through adversarial training with the generator, the discriminator drives the generator to progressively enhance its output quality. The proposed discriminator employs a dual-branch, multi-scale architecture that achieves collaborative discrimination via parallel processing of features at different resolutions. The primary

branch incorporates four downsampling stages. Each stage utilizes a 4×4 convolutional layer with a stride of 2, ultimately producing a 16×16 patch-based output map. Here, each point corresponds to a receptive field region within the input image for authenticity assessment. Meanwhile, the auxiliary branch first applies bilinear downsampling to the input image, followed by feature extraction through three convolutional stages, and finally outputs its discrimination result. To preserve the statistical properties of the original input, instance normalization (except after the first layer) and LeakyReLU activation functions are applied throughout both branches to stabilize training and mitigate gradient vanishing. Subsequently, the discrimination matrices (probability maps) from both branches are fused using arithmetic averaging. This design leverages the primary branch’s retention of high-resolution detail (output size: $H/16 \times W/16$, where H and W denote the input image height and width) and the auxiliary branch’s enhanced global perception (output size: $H/8 \times W/8$). Consequently, the discriminator combines implicitly aligned features from dual-scale training to strengthen global semantic understanding while maintaining local texture discrimination capabilities. The structure of the double-branch discriminator is shown in Figure 3.

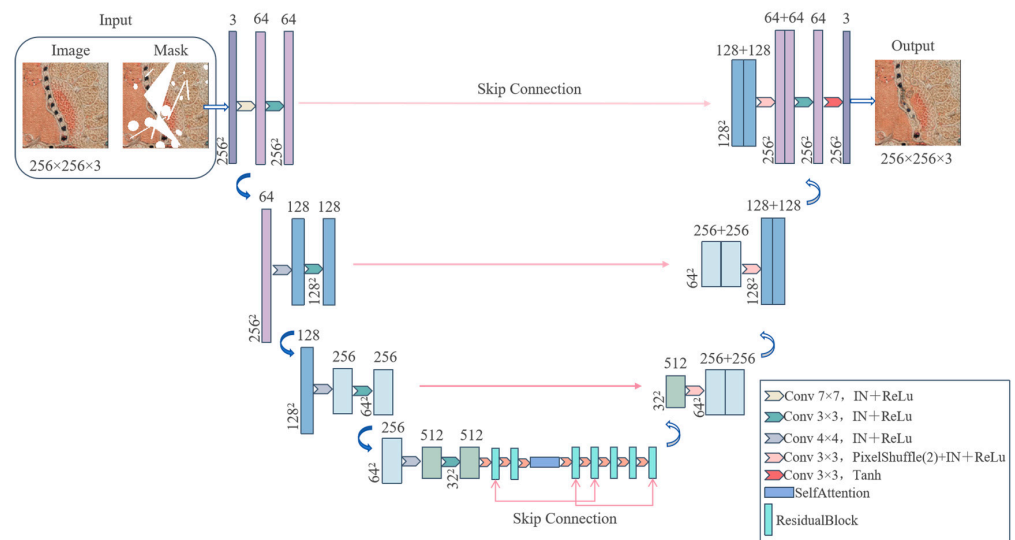


Figure 2. Improved generator structure diagram.

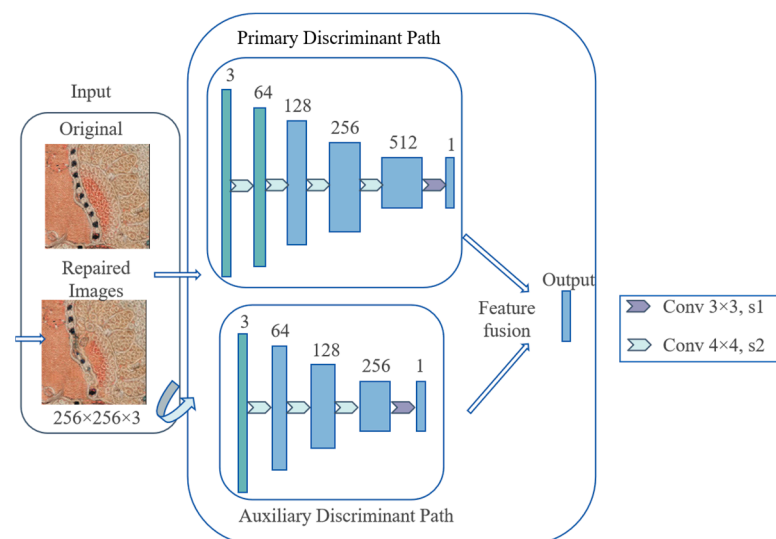


Figure 3. Double-branch discriminator structure diagram.

2.4. Loss Function

The loss function serves as a core component in machine learning and deep learning, quantifying the discrepancy between model predictions and ground-truth values to guide the updating of model parameters via optimization algorithms [28]. In this work, we employ a composite loss function for generator optimization. Specifically, the generator loss comprises the following key components: adversarial loss, pixel-level L1 loss, perceptual loss, and style loss.

Within the adversarial loss component, the Mean Squared Error (MSE) loss replaces the traditional Binary Cross-Entropy (BCE) loss commonly used in GANs. This component calculates the mean squared error between the discriminator's output probability map and the target label (indicating "real"). Its objective is to encourage the discriminator to classify the generated images as authentic. By minimizing this loss, the generator is driven to synthesize images with photorealistic texture details that reduce discernible differences from real images. The adversarial loss based on MSE is formally defined as follows:

$$L_{adv} = E_{\hat{x} \sim p_{data}} \left[(D(G(\hat{x})) - 1)^2 \right] \quad (1)$$

In the formula, $G(\hat{x})$ denotes a restored embroidery image sample produced by the generator. $D(G(\hat{x}))$ represents the discriminator's output probability for the restored image, ranging from 0 to 1, where 1 signifies a real image. This adversarial loss quantitatively measures the generator's ability to deceive the discriminator.

The pixel-level L1 loss calculates the Mean Absolute Error (MAE) between the generated embroidery image and the corresponding ground-truth image. Specifically, this complementary loss function enhances generative tasks by ensuring pixel-wise consistency between the generated output and real data, thereby preserving fine-grained details. The L1 loss is formally defined as follows:

$$L_{pixel} = \|G(\hat{x}) \odot M - x \odot M\|_1 \quad (2)$$

In the formula, M denotes the binary mask matrix. $G(\hat{x}) \odot M$ represents the restored image region within the mask area. Correspondingly, $x \odot M$ denotes the ground-truth image region within the mask area. This pixel-level L1 loss enforces consistency between the restored region and the ground-truth region, thus directly constraining pixel-level reconstruction accuracy.

The perceptual loss leverages feature maps extracted from the first 16 layers of a pre-trained VGG convolutional network. This loss function enforces consistency between the restored region and the original image within a deep feature space, thereby ensuring semantic-level coherence beyond mere pixel-level similarity. Specifically, it is defined as follows:

$$L_{perceptual} = \|\phi_{vgg}(G(\hat{x})) - \phi_{vgg}(x)\|_1 \quad (3)$$

In the formula, ϕ_{vgg} denotes the feature extraction function of the VGG16 network. $\phi_{vgg}(x)$ represents the feature representation of the ground-truth image, $\phi_{vgg}(G(\hat{x}))$ while corresponds to the feature representation of the restored image. The hierarchical convolutional operations in VGG16 inherently capture local texture patterns, thus making this network particularly effective at preserving the intricate structural details characteristic of traditional embroidery.

The style loss is critical for transferring textural style attributes and maintaining global coherence in traditional embroidery image restoration. This loss quantifies the discrepancy

between the Gram matrices of the restored image and the ground-truth image within the VGG feature space. Specifically, it is defined as follows:

$$L_{\text{style}} = \|\mathcal{G}(\phi_{\text{vgg}}(G(\hat{x}))) - \mathcal{G}(\phi_{\text{vgg}}(x))\|_2^2 \quad (4)$$

The Gram matrix is computed as follows:

$$\mathcal{G}(F) = \frac{1}{H \times W} F \cdot F^T \quad (5)$$

In the formula, \mathcal{G} denotes the Gram matrix computation function, and F represents the feature map matrix with dimensions $C \times (H \times W)$. The Gram matrix captures the relevance between feature channels, thereby characterizing the textural style of the image. Thus, this loss function ensures consistency between the textural style of the repaired area and that of the surrounding original image.

The total generator loss comprises:

$$L_G = \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{pixel}} L_{\text{pixel}} + \lambda_{\text{perceptual}} L_{\text{perceptual}} + \lambda_{\text{style}} L_{\text{style}} \quad (6)$$

In this formulation, λ_{adv} , λ_{pixel} , $\lambda_{\text{perceptual}}$, and λ_{style} denote the weighting coefficients for the adversarial, pixel-level, perceptual, and style losses, respectively. Collectively, these components enable the generator to: Achieve pixel-accurate reconstruction. Preserve semantic and stylistic fidelity to the source image. Generate adversarially plausible restoration results.

The discriminator employs a dual-branch multi-scale architecture. This design integrates adversarial losses at two distinct scales, enabling joint optimization for fine-grained local authenticity assessment. The composite loss is formulated as:

$$L_D = \frac{1}{2} \left[\mathbb{E}_{x \sim P_{\text{data}}} [(D(x) - 1)^2] + \mathbb{E}_{\hat{x} \sim P_{\text{masked}}} [D(G(\hat{x}))] \right] \quad (7)$$

In the formula, P_{masked} represents the probability distribution of the masked image, and this loss trains the discriminator to distinguish between real images and restored images. The first objective enables the discriminator to recognize real images, while the second term trains it to identify generated images. Through progressive weighting, the generator and discriminator adapt to stage-specific requirements during alternating optimization. This process continuously enhances restoration quality, producing realistic images with optimal classification performance to form a comprehensive image restoration objective. The loss function weight parameters are shown in Table 1.

Table 1. Loss function training weight parameters.

Loss Function Weight	Default Weight Value	Epoch \geq 300
adv_weight	1.0	remain unchanged
pixel_weight	10.0	remain unchanged
perceptual_weight	0.1	0.15
style_weight	100.0	remain unchanged
optimizer_D.lr	0.0001	Reduced to 5% of the initial value (0.000005)

3. Repair Framework

3.1. Repair Process

This paper proposes an enhanced image restoration method based on generative adversarial networks for the high-precision restoration of traditional embroidery patterns. The restoration process comprises two sequential phases:

Phase 1: Dataset construction and model training. Embroidery images are acquired through systematic data collection and preprocessing to establish a dedicated embroidery dataset. Subsequently, a restoration network model is constructed and trained, undergoing iterative optimization with early stopping during training to derive the final restoration model.

Phase 2: Image processing and evaluation. Following the establishment of a complete restoration pipeline, embroidery images are inputted and subjected to randomly generated irregular masking. The masked images are partitioned into local patches, which are then restored by loading the pre-trained model weights. These restored patches are reintegrated into complete images. Consequently, eye-tracking experiments and restoration efficacy assessments are conducted on the reconstructed images, ultimately outputting the optimally restored complete images.

3.2. Mask Generation

Damage to embroidered artifacts often exhibits diverse forms due to environmental factors. Conventional embroidery image restoration methods typically employ rectangular or arbitrary masks, which are insufficient to accurately replicate the actual damage conditions of cultural heritage objects [29]. To better simulate realistic degradation in embroidered images, our mask generation strategy incorporates varied morphological structures to mimic common damage types. Using a segment of the “Longevity Embroidery” on crimson silk from the Hunan Provincial Museum as an example (Figure 4), we categorize actual damage patterns into three mask types: wormhole-shaped, crease-shaped, and block-shaped masks. Wormhole-shaped masks simulate insect bites or punctiform defects through randomly generated irregular circles; crease-shaped masks imitate linear folds using randomized segments and their branches; block-shaped masks represent large-area damage via irregular polygons. These masks can be applied individually, in pairs, or simultaneously, with controllable coverage rates and positional modes (center, edge, or random distribution). The three irregular mask combinations are illustrated in Figure 5.

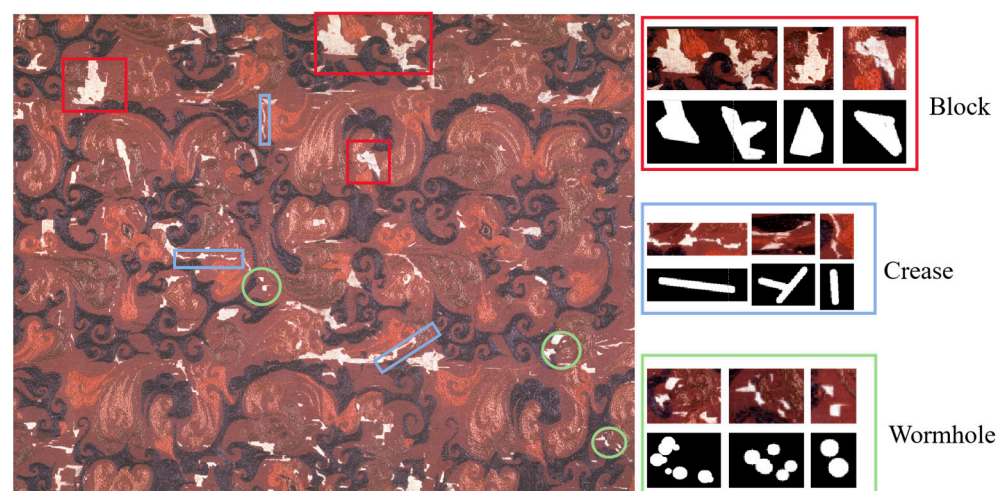


Figure 4. Example of a mask physical object.

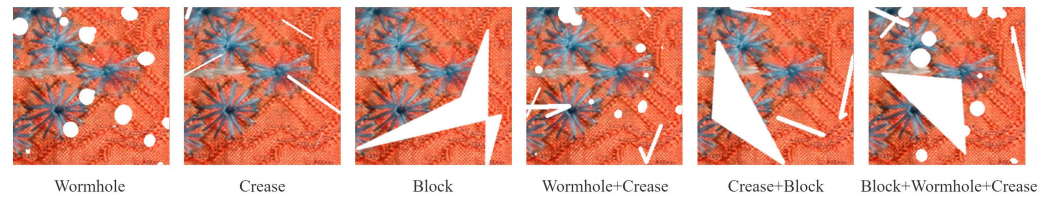


Figure 5. Stochastic Mask Patterns.

During mask generation, iterative refinements are applied to ensure morphological authenticity, with all processing implemented with OpenCV 4.5+ morphological operations. The resulting binary masks guide both the training and inference phases of the inpainting model. Furthermore, a dynamic randomization strategy is adopted during training—each iteration generates masks with independent variations in morphology, scale, and spatial distribution. This approach prevents pattern fixation and enables the generator to learn diverse damage-repair priors rather than memorizing specific mask patterns. The dynamic mask generation mechanism significantly enhances the model’s generalization capability for repairing complex real-world damage in embroidered images.

3.3. Image Segmentation and Combination

3.3.1. Image Segmentation

Upon completion of the image masking stage, the segmentation phase commences utilizing a grid-based tiling method with overlap to achieve precise segmentation of large-scale embroidery images. Specifically, an adaptive grid computation determines the optimal partitioning scheme based on preset tile dimensions and overlap width. Given an original image size of $H \times W$ pixels, the row count R and column count C are derived as defined in Equations (8) and (9):

$$R = \text{ceil}\left(\frac{H - 16}{256 - 16}\right) \quad (8)$$

$$C = \text{ceil}\left(\frac{W - 16}{256 - 16}\right) \quad (9)$$

When the image dimensions are not integer multiples of the stride length, the terminal block is adaptively resized to ensure complete image coverage without exceeding the original boundaries:

$$\begin{cases} (R - 1) \times (\text{patch_size-overlap}) + \text{patch_size} \geq H \\ (C - 1) \times (\text{patch_size-overlap}) + \text{patch_size} \geq W \end{cases} \quad (10)$$

This ensures complete coverage of the original image without altering its resolution. Image segmentation is performed by moving in steps of 240 pixels. This approach ensures that adjacent blocks maintain an overlap region of 16 pixels in both horizontal and vertical directions. In the practical application stage, the size of the image overlap area can be adjusted according to the actual image size. For boundary areas, adaptive cropping dynamically adjusts the block size according to the image’s actual boundaries, preserving the original content integrity within all blocks. The process precisely divides the image into blocks with fixed overlapping pixel areas. Each block forms a continuous overlap region with its adjacent blocks in all four directions (up, down, left, right). The actual size and spatial information of each block are recorded, generating a metadata matrix that contains spatial coordinates, mask status, and file paths. As illustrated in Figure 6, this division is based on an image measuring 1024×1024 pixels. Consequently, this segmentation method guarantees contextual continuity between blocks, avoids the boundary artifacts inherent in traditional grid segmentation, and significantly improves local inpainting

results. Furthermore, it provides accurate spatial correspondence for subsequent image reassembly and weighted fusion, thereby establishing a structurally sound and information-rich preprocessing foundation for high-quality image restoration tasks.

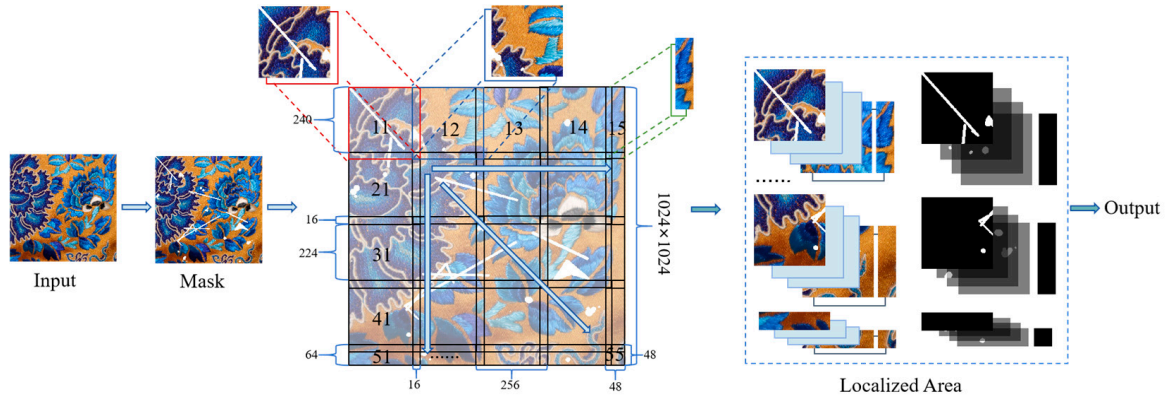


Figure 6. Schematic diagram of segmentation based on grid overlap.

3.3.2. Image Combination

A metadata-driven reverse grid reconstruction method was employed during the image combination stage. First, the pre-saved metadata matrix was loaded. This matrix accurately records the boundary positions of each sub-block within the original image coordinate system, along with their corresponding inpainted file paths. Utilizing the location indices provided by the metadata, each inpainted sub-block was precisely positioned and output to its corresponding area on the canvas, ensuring strict geometric alignment with its initial segmentation position. A distance-weighted pixel fusion algorithm was applied to the predefined overlapping regions between adjacent blocks. Specifically, based on the geometric positions of these overlapping areas, a weight map was generated by normalizing the distance from each pixel to the internal boundaries of its respective sub-block. This process achieves a spatially continuous and adaptive transition of the inpainting results at the seams between adjacent blocks. Consequently, the contributions from overlapping areas are equally weighted, facilitating smooth color transitions proportional to the size of each block and thereby eliminating harsh seams. The non-overlapping areas are directly populated with their corresponding sub-block pixel values. By iterating through all sub-blocks and performing this positioning fusion operation, a globally consistent and seamless output image is synthesized. This approach effectively eliminates the seam artifacts inherent to block-based processing while maintaining texture continuity and structural integrity. The complete inpainting process is shown in Figure 7.

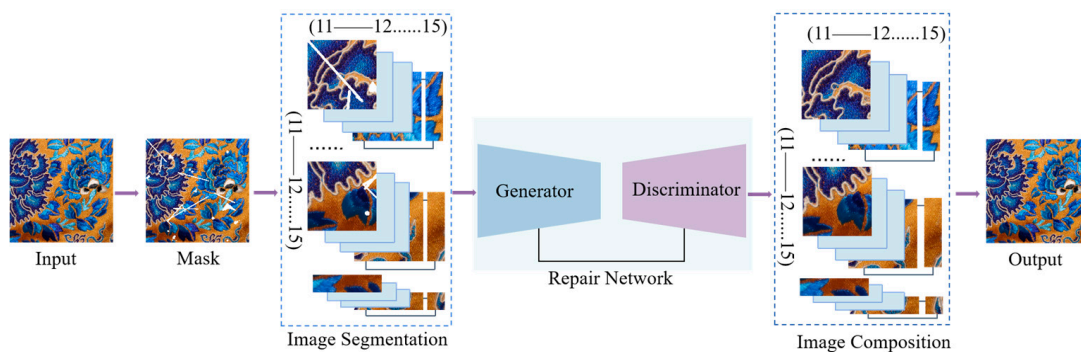


Figure 7. Complete repair process.

4. Experiments

4.1. Data Collection and Experimental Setup

The embroidery image data were captured using a Canon EOS M6 II digital camera, Tokyo, Japan (4256 × 3200 pixels) equipped with an 18–45 mm standard lens. All images were acquired under natural lighting conditions. A total of 480 raw images of embroidered fabrics were initially collected. After removing duplicates and low-quality samples, the remaining images were uniformly resized to a high-resolution format of 1024 pixels. These images were further cropped to produce a final dataset comprising 3143 embroidery image patches, each sized 256 × 256 pixels. This curated dataset was used to train the generative adversarial network (GAN) model. The samples exhibit clear embroidery texture details, with representative examples provided in Figure 8. The proposed model was evaluated on this self-constructed embroidery dataset. Detailed training parameters and experimental configurations are summarized in Table 2.

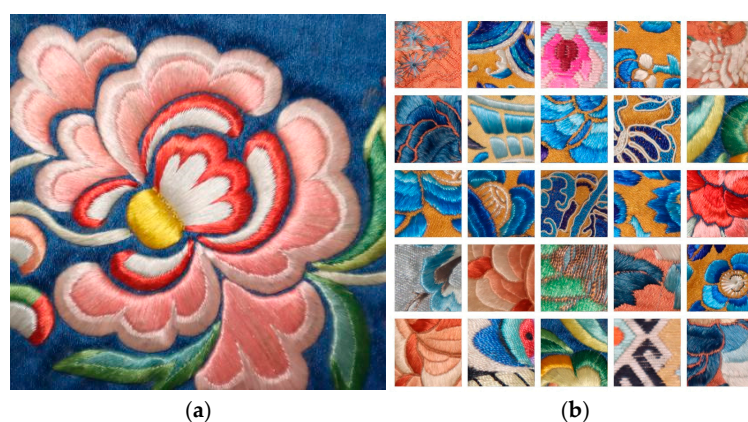


Figure 8. Sample images from the dataset. (a) 1024 × 1024 pixel sample image; (b) 256 × 256 pixel sample image.

Table 2. Training Parameters and Configuration.

Parameter/Configuration	Details
Dataset split	Training: 90%, Testing: 10%
Resolution	256 × 256
Optimizer	Adam
β_1	0.5
β_2	0.999
Learning rate	Initial: 2×10^{-4} , Discriminator: 1×10^{-4}
Learning rate adjustment	ReduceLROnPlateau (factor = 0.5, patience = 25)
Batch size	16
Experimental Server	
CPU	16 vCPU Intel(R) Xeon(R) Gold 6430
GPU	RTX 4090 (24 GB)
Operating System	ubuntu20.04
Python Version	3.8
CUDA Version	11.4

4.2. Qualitative Comparison

Based on preliminary data collection and parameter configuration, the image restoration results obtained through model training are presented in Figure 9, illustrating the model's performance across different types of masks. To further evaluate the effectiveness of the proposed method, comparative experiments were conducted with several state-of-the-art approaches, namely CE [20], PCNOV [24], and GCNOV [26]. The restoration

outcomes of each algorithm on the same test set of embroidery images are compared in Figure 10. As observed, the proposed algorithm achieves visually superior restoration results compared to other methods. Specifically, the CE algorithm is able to recover approximate color information of the embroidery images, but the restored regions appear noticeably blurred. Both GCNOV and PCNOV methods exhibit semantic inconsistencies and inadequate texture reconstruction, failing to reproduce fine textural details. For quantitative assessment, Peak Signal-to-Noise Ratio (PSNR) [30] and Structural Similarity Index Measure (SSIM) [31] were adopted as evaluation metrics. As summarized in Table 3, the proposed method achieves higher scores in both PSNR and SSIM, demonstrating its effectiveness in recovering pixel-level accuracy and structural coherence in embroidered image restoration.



Figure 9. Model training for image restoration.

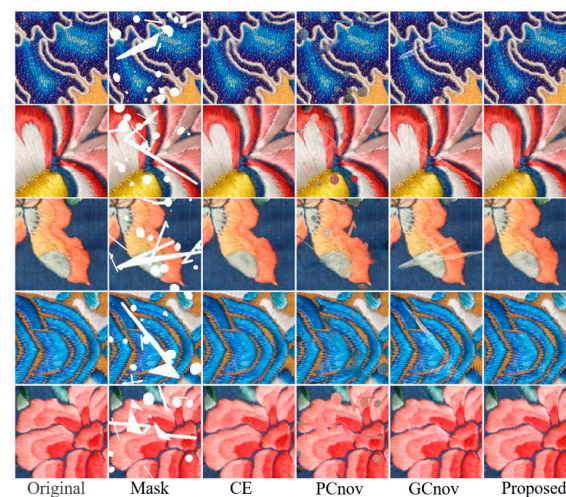


Figure 10. Comparison of repair effects of various methods.

Table 3. PSNR and SSIM Performance of Different Methods.

Method	PSNR	SSIM
CE	28.33 dB	0.9001
PCnov	22.74 dB	0.8495
GCnov	25.11 dB	0.8822
OURS	28.64 dB	0.9074

4.3. Ablation Experiment

To evaluate the effectiveness of the proposed method and assess the contribution of each module to the restoration quality, we conducted a series of ablation studies. Using the original model as the baseline, we designed three variant models under identical experimental conditions: the first variant replaced the gated convolutional layers with standard convolutional layers; the second removed the attention mechanism; and the third omitted the context enhancement module. The performance of each variant was compared against

the full model on the same dataset, with quantitative results presented in Table 4. Compared to the ablated versions, the baseline model achieved the highest scores in pixel-level metrics (PSNR and SSIM), demonstrating that the incorporation of gated convolution, the attention mechanism, and the context enhancement module significantly improves performance. Visual examples of the restoration results after each module modification are provided in Figure 11. The outcomes indicate that each component contributes positively to the model's overall efficacy, validating the rationality and effectiveness of the proposed architecture.

Table 4. PSNR and SSIM Performance Results of the Ablation Experiment.

Method	PSNR	SSIM
Group 1	27.72 dB	0.9055
Group 2	27.76 dB	0.9038
Group 3	27.94 dB	0.9043
Complete model	28.18 dB	0.9066

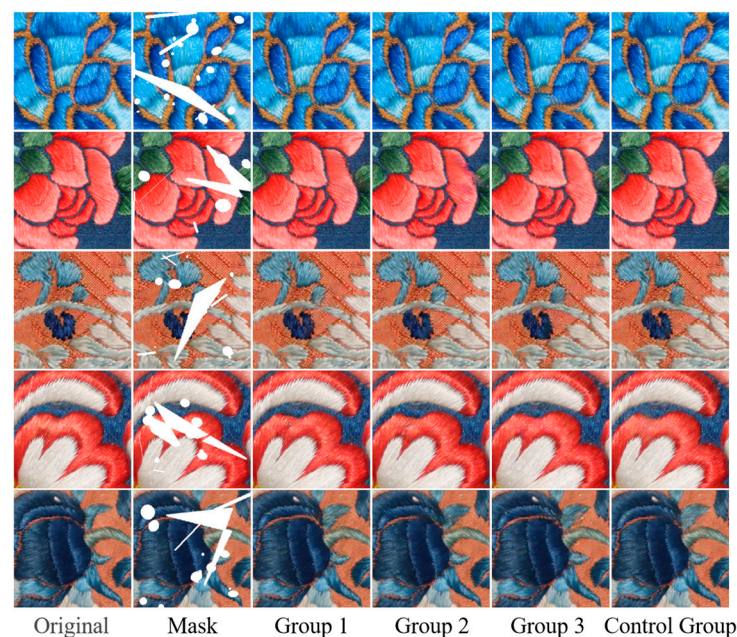


Figure 11. Visual examples of images after module changes.

4.4. Overall Restoration Effect Evaluation

To objectively evaluate the visual authenticity of the restoration results, this study designed a rigorous eye-tracking experiment based on the core hypothesis that high-quality inpainted regions should not attract abnormal visual attention. Thirty participants with normal or corrected-to-normal vision were recruited and tested using a Tobii Pro Lab eye tracker under controlled lighting conditions. The experiment consisted of a calibration phase followed by a formal task phase. Prior to data collection, the eye tracker was calibrated for each participant to ensure measurement accuracy. During the formal session, each participant was presented with four restored embroidery images in a standardized setting, with each image displayed for 15 s followed by a 2 s blank interval to reduce carryover effects. Participants were instructed to freely inspect each image and actively identify any regions that appeared artificially restored or visually flawed, while gaze data were recorded throughout the session. After all experiments were completed, the gaze patterns were aggregated across participants to generate composite heatmaps, as shown in Figure 12.

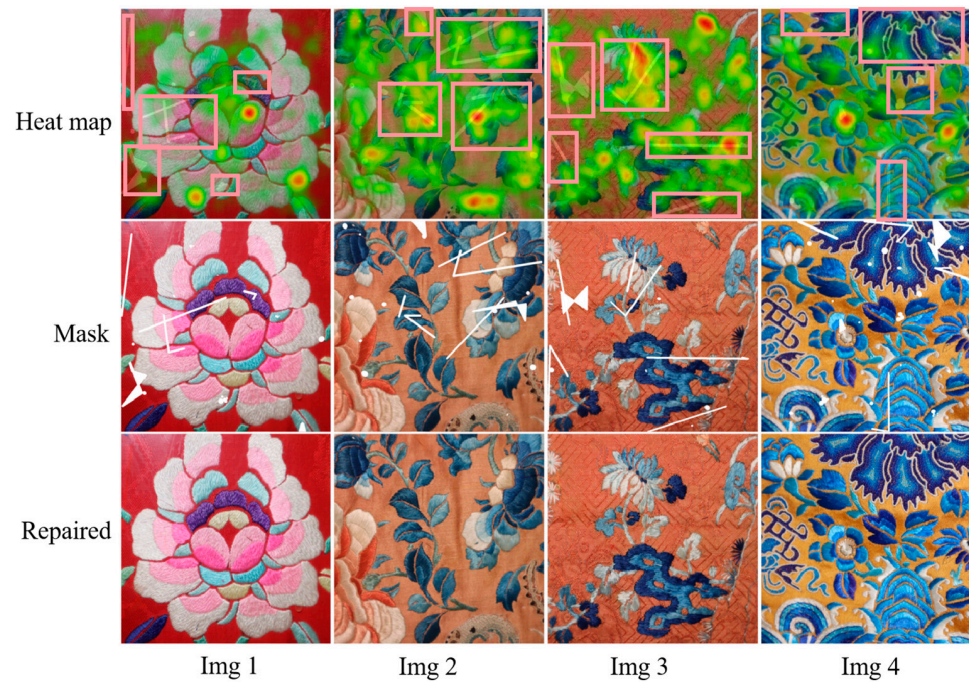


Figure 12. Eye tracking heat map.

Subsequently, based on the gaze aggregation heatmap, an eye-tracking spatial quantification analysis was conducted by comparing the embroidery mask image with the heatmap, with the results shown in Table 5. Eye-tracking-based spatial alignment analysis confirmed that, among the four embroidery samples: the spatial overlap between gaze hotspots and the repaired area was low ($\text{IoU} = 0.040 \pm 0.015$), the repair boundaries did not form visual aggregation ($\text{BAI} = 0.086 \pm 0.015$), and visual attention was evenly distributed within the repaired area ($\text{VSD} = 0.889 \pm 0.007$). The results indicate that the proposed method achieves visually natural restoration effects, with the restoration area not triggering significant visual attention and restoration traces not causing significant interference at the visual perception level, thereby achieving the technical goal of “visually seamless restoration. This confirms the validity of the hypothesis. Based on the experimental results, it can be inferred that the current restoration method generates embroidery images with good visual naturalness, effectively avoiding the explicit exposure of artificial restoration traces. The eye-tracking experiment method also effectively quantifies the visual authenticity and naturalness of the restored images.

Table 5. Quantitative analysis results of eye movement spatial alignment.

Indicator	Average	Standard Deviation	Minimum Value	Maximum Value	Threshold
IoU	0.040	0.020	0.015	0.062	<0.30
BAI	0.086	0.027	0.057	0.116	<0.20
VSD	0.889	0.038	0.832	0.930	<0.10

To objectively evaluate the effectiveness of the embroidery restoration method, a dual-metric validation approach was adopted using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR measures pixel-level accuracy between the restored and original images, with higher values indicating better fidelity—results above 30 dB are generally considered acceptable. SSIM assesses perceptual quality, focusing on structural and edge consistency, where values closer to 1 reflect stronger visual coherence. On a set of 30 test samples, each sized 1024×1024 pixels, the proposed method

achieved an average PSNR of 32.18 dB and an average SSIM of 0.969. This discrepancy in metric performance highlights the model's emphasis on perceptual naturalness: it prioritizes structural continuity and seamless edge blending between restored and original regions, effectively avoiding visual attention anomalies—as corroborated by eye-tracking experiments. Nevertheless, due to the high-frequency complexity inherent in embroidery textures, there remains room for improvement in terms of local pixel-level fidelity, as captured by PSNR. The high SSIM scores confirm the method's success in achieving visually coherent and imperceptible repairs, validating its capability for “seamless restoration” of embroidered heritage.

Although the proposed method has demonstrated promising results in restoring embroidery images, it is not without limitations. For instance, the model still struggles with large-area defects, as illustrated in Figure 13. When addressing extensive damaged regions, the restored embroidery images may exhibit issues such as semantic inconsistency, texture degradation, and boundary artifacts. In particular, unnatural texture transitions and visual artifacts are more likely to occur near the junctions between restored and original regions.

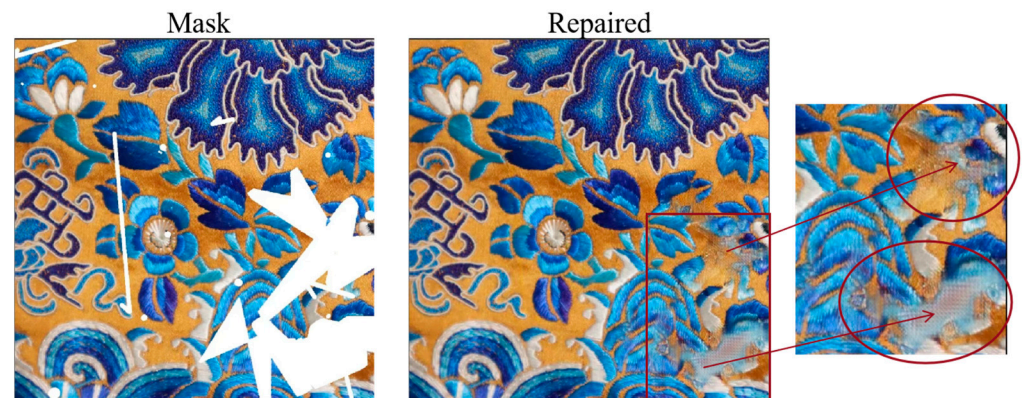


Figure 13. Large-Area Defect Repair Diagram.

4.5. Practical Application

To validate the effectiveness of the proposed method in the practical restoration of embroidered cultural relics, a digital fragment of the “Longevity Embroidery” on vermillion silk (1401×1080 pixels), from the collection of the Hunan Provincial Museum, was selected for experimental inpainting. A digital detection method based on multi-feature fusion was first applied. Specifically, the damaged regions of the embroidery were preliminarily extracted using adaptive thresholding in the HSV color space combined with Canny edge detection. Since the initially identified mask regions exhibited irregular boundaries—which could adversely affect the inpainting results—local refinements were manually performed to ensure precision. This process yielded a binary mask image that accurately delineates the damaged areas and meets the input requirements of the inpainting algorithm. The proposed algorithm was then employed to perform the inpainting task. During processing, the overlap size for image segmentation was adaptively set to 32 pixels based on the image dimensions to optimize the restoration quality. The practical application workflow is illustrated in Figure 14, which includes the original embroidery fragment, the mask identification procedure, the initially detected mask, the refined mask after adjustment, and the final inpainted result. The proposed method successfully restored the deep red chromaticity of the silk substrate while achieving high fidelity in texture detail recovery and semantic coherence in the damaged regions.

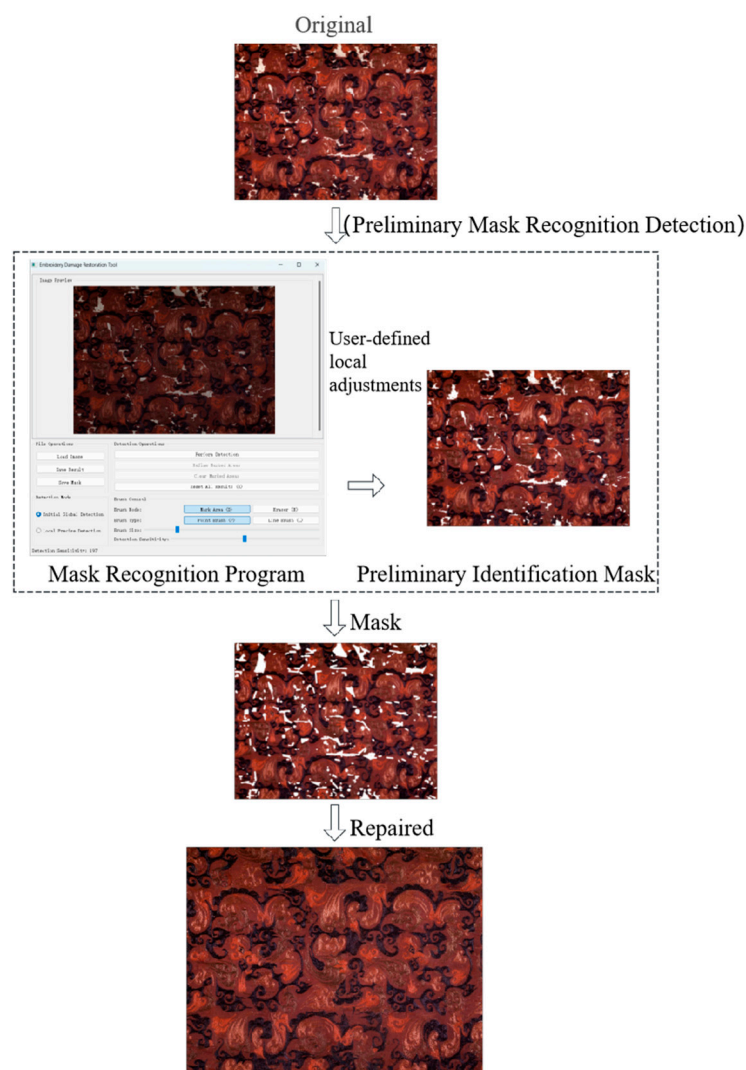


Figure 14. Actual application diagram.

This case study demonstrates not only the practical applicability and robustness of the proposed approach for restoring complex real-world embroidered cultural heritage images, but also its significant value in digital cultural preservation. By providing a high-quality digital surrogate of the fragile artifact, this method contributes to the irreversible documentation and preservation of historical embroidery patterns. Moreover, the convincing restoration outcome offers museum professionals a reliable reference to inform physical conservation decisions, thereby supporting sustainable heritage protection and enhancing public accessibility to restored cultural visual materials.

5. Conclusions

The restoration of embroidered cultural relics presents multiple challenges, including the inherent fragility of the artifacts, the uniqueness of embroidery techniques, and the high complexity and risks associated with physical restoration processes. Consequently, non-invasive digital restoration techniques have become a crucial means for preserving valuable cultural heritage. This study proposes an improved generative adversarial network (GAN)-based framework for end-to-end digital restoration of embroidery images. The approach innovatively incorporates three mask training strategies, significantly enhancing the model's generalization capability and restoration robustness for complex patterns. Furthermore, by introducing an adaptive image segmentation and fusion mechanism, the

method overcomes the limitations of traditional approaches regarding input size flexibility. Experimental results demonstrate that the proposed model achieves superior performance in both texture recovery and structural coherence. The framework shows great potential for application in the digital conservation of embroidered cultural relics.

Future work will focus on further optimization of the current generative adversarial network framework, with particular interest in exploring the potential of Vision Transformers (ViTs) for long-range texture modeling in embroidery restoration. The global self-attention mechanism of ViTs could enhance structural coherence when reconstructing complex pattern continuations. Additionally, we plan to integrate the semantic understanding capabilities of multimodal generative models—such as Stable Diffusion and DALL·E—into a hybrid restoration pipeline to better address large-area damage reconstruction. Ultimately, we aim to develop a cross-modal validation system that bridges digital restoration technology and art historical interpretation, establishing a new paradigm in cultural heritage preservation that harmonizes technical innovation with contextual cultural understanding.

Author Contributions: Conceptualization, Q.W., Z.L.; formal analysis, Q.W.; methodology, Q.W.; software, Q.W., C.J.; validation, C.J., K.J. and Z.L.; writing—original draft, Q.W.; writing—review and editing, Z.L., X.L. and F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Exploration of Graduate Student Talent Cultivation Mode in Textile and Garment Direction of “Civic and Political Leadership + Digital Intelligence Empowerment” in Shanxi Province (No. 2024JG042), Research on the Protection Strategy of Mural Painting Cultural Relics in Shanxi under the Background of Digital Intelligence Technology of Shanxi Province (No. 202404030401110).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The relevant data are included in the paper.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their helpful suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GAN	Generative Adversarial Networks
SSIM	Structure Similarity Index Measure
PSNR	Peak Signal-to-Noise Ratio
BSCB	Bertalmio-Sapiro-Caselles-Ballester Model
AE	Autoencoder
CNN	Convolutional Neural Network
CA	Contextual Attention
LeNet-5	LeNet-5 Convolutional Neural Network
U-Net	U-Shaped Network
GatedConv	Gated Convolution
PatchGAN	Patch-based Generative Adversarial Network
LeakyReLU	Leaky Rectified Linear Unit
VGG16	Visual Geometry Group 16-layer Convolutional Neural Network

References

1. Yan, W.-J.; Chiou, S.-C. The Safeguarding of Intangible Cultural Heritage from the Perspective of Civic Participation: The Informal Education of Chinese Embroidery Handicrafts. *Sustainability* **2021**, *13*, 4958. [\[CrossRef\]](#)
2. Sha, S.; Li, Y.; Wei, W.; Liu, Y.; Chi, C.; Jiang, X.; Deng, Z.; Luo, L. Image Classification and Restoration of Ancient Textiles Based on Convolutional Neural Network. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 11. [\[CrossRef\]](#)
3. Zhang, Y.; Chen, Z.; Chen, J.; Duan, Z. Research on the Color Characteristics of Dragon Patterns in the Embroidery Collection of the Palace Museum During the Qing Dynasty Based on K-Means Clustering Analysis and Its Application in the Color. In *Advances in Printing, Packaging and Communication Technologies*; Song, H., Xu, M., Yang, L., Zhang, L., Eds.; Springer Nature: Singapore, 2024; pp. 43–57.
4. Li, H.; Zhang, J.; Peng, W.; Tian, X.; Shi, J. Digital Restoration of Ancient Loom: Evaluation of the Digital Restoration Process and Display Effect of Yunjin Dahualou Loom. *npj Herit. Sci.* **2025**, *13*, 3. [\[CrossRef\]](#)
5. Yang, J.; Cao, J.; Yang, H.; Li, Y.; Wang, J. Digitally Assisted Preservation and Restoration of a Fragmented Mural in a Tang Tomb. *Sens. Imaging* **2021**, *22*, 32. [\[CrossRef\]](#)
6. Liu, K.; Zhao, J.; Zhu, C. Research on Digital Restoration of Plain Unlined Silk Gauze Gown of Mawangdui Han Dynasty Tomb Based on AHP and Human–Computer Interaction Technology. *Sustainability* **2022**, *14*, 8713. [\[CrossRef\]](#)
7. Liu, K.; Wu, H.; Ji, Y.; Zhu, C. Archaeology and Restoration of Costumes in Tang Tomb Murals Based on Reverse Engineering and Human-Computer Interaction Technology. *Sustainability* **2022**, *14*, 6232. [\[CrossRef\]](#)
8. Lozano, J.S.; Pagán, E.A.; Roig, E.M.; Salvatella, M.G.; Muñoz, A.L.; Peris, J.S.; Vernus, P.; Puren, M.; Rei, L.; Mladenich, D. Open Access to Data about Silk Heritage: A Case Study in Digital Information. *Sustainability* **2023**, *15*, 14340. [\[CrossRef\]](#)
9. Rastogi, R.; Rawat, V.; Kaushal, S. Advancements in Image Restoration Techniques: A Comprehensive Review and Analysis through GAN. In *Generative Artificial Intelligence and Ethics: Standards, Guidelines, and Best Practices*; IGI Global: Hershey, PA, USA, 2025; pp. 53–90.
10. Li, H.; Luo, W.; Huang, J. Localization of Diffusion-Based Inpainting in Digital Images. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 3050–3064. [\[CrossRef\]](#)
11. Bertalmio, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, LA, USA, 23–28 July 2000; pp. 417–424.
12. Criminisi, A.; Perez, P.; Toyama, K. Region Filling and Object Removal by Exemplar-Based Image Inpainting. *IEEE Trans. Image Process.* **2004**, *13*, 1200–1212. [\[CrossRef\]](#)
13. Cheng, W.; Hsieh, C.; Lin, S.; Wang, C.; Wu, J. Robust Algorithm for Exemplar-Based Image Inpainting. In Proceedings of the International Conference on Computer Graphics, Imaging and Visualization, Beijing, China, 26–29 July 2005; pp. 64–69.
14. Zhang, X.; Zhai, D.; Li, T.; Zhou, Y.; Lin, Y. Image Inpainting Based on Deep Learning: A Review. *Inform. Fusion* **2023**, *90*, 74–94. [\[CrossRef\]](#)
15. Xu, Z.S.; Zhang, X.F.; Chen, W.; Yao, M.D.; Liu, J.T.; Xu, T.T.; Wang, Z.H. A Review of Image Inpainting Methods Based on Deep Learning. *Appl. Sci.* **2023**, *13*, 11189. [\[CrossRef\]](#)
16. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems 27, Montreal, QC, Canada, 8–13 December 2014.
17. Moghadam, M.M.; Boroomand, B.; Jalali, M.; Zareian, A.; Daeijavad, A.; Manshaei, M.H.; Krunk, M. Games of GANs: Game-Theoretical Models for Generative Adversarial Networks. *Artif. Intell. Rev.* **2023**, *56*, 9771–9807. [\[CrossRef\]](#)
18. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784. [\[CrossRef\]](#)
19. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
20. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
21. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. *Int. Conf. Mach. Learn. PMLR* **2017**, *70*, 214–223.
22. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* **2017**, *36*, 1–14. [\[CrossRef\]](#)
23. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative Image Inpainting with Contextual Attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514.
24. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.-C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8 September 2018; pp. 85–100.
25. Zhong, C.; Yu, X.; Xia, H.; Xu, Q. Restoring intricate Miao embroidery patterns: A GAN-based U-Net with spatial-channel attention. *Vis. Comput.* **2025**, *41*, 7521–7533. [\[CrossRef\]](#)
26. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4471–4480.

27. Chen, G.; Zhang, G.; Yang, Z.; Liu, W. Multi-Scale Patch-GAN with Edge Detection for Image Inpainting. *Appl. Intell.* **2023**, *53*, 3917–3932. [[CrossRef](#)]
28. Barron, J.T. A General and Adaptive Robust Loss Function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4331–4339.
29. Wang, N.; Wang, W.; Hu, W.; Fenster, A.; Li, S. Thangka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. *IEEE Trans. Image Process.* **2021**, *30*, 3720–3733. [[CrossRef](#)]
30. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
31. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.