



# Sketch-Guided Conditional Generative Adversarial Network for Restoration and Design Generation of Historical Building Images

Zheng Qi<sup>\*</sup>, Chen Kang, Changmei Shan

College for Creative Studies, Changzhou Vocational Institute of Mechatronic Technology, Changzhou 214163, China

Corresponding Author Email: [qz1937@czimt.edu.cn](mailto:qz1937@czimt.edu.cn)

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430102>

## ABSTRACT

**Received:** 25 August 2025

**Revised:** 1 November 2025

**Accepted:** 16 November 2025

**Available online:** 28 February 2026

### Keywords:

*historical building images, image restoration, conditional generative adversarial networks, sketch interaction, frequency-domain feature modulation*

Restoring damaged historical building images requires maintaining structural integrity, preserving authentic styles, and supporting interactive flexibility. Existing conditional generative adversarial networks (cGANs) often fall short in enforcing structural constraints, leveraging frequency-domain features, and decoupling structure from texture, limiting their effectiveness for high-precision restoration and interactive design. To address these challenges, we propose a sketch-guided cGAN with multi-scale frequency-aware perception, enabling precise restoration and personalized design generation for historical building images. The network adopts a dual-encoder–shared-decoder architecture to enforce structural constraints, introduces a frequency-adaptive feature modulation module to overcome the limitations of spatial convolutions in balancing structure and texture, and implements a structure–texture decoupled generation strategy to synergistically achieve style transfer and structural fidelity. Furthermore, a multi-scale trident discriminator is designed to enhance visual realism and style consistency of the generated results. Experiments on both a self-constructed historical building damage dataset and public datasets demonstrate that the proposed method outperforms state-of-the-art approaches in metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and perceptual similarity, while effectively supporting interactive design. The generated results conform to historical architectural style standards. This approach provides a novel technical pathway for image restoration and digital design of historical buildings, with significant theoretical and practical implications.

## 1. INTRODUCTION

Historical buildings, as important carriers of cultural heritage [1, 2], require digital preservation as a core means to achieve cultural inheritance and sustainable utilization. During digital acquisition and archiving [3, 4], building images are prone to damage due to natural weathering, human-induced destruction, or data loss [5], which seriously affects the integrity of digital records and the accuracy of restoration work. Traditional building image restoration methods rely on manual drawing and experiential judgment [6, 7], which are not only inefficient but also highly subjective, making it difficult to meet the demands of large-scale, high-precision restoration. In recent years, deep learning-based image restoration technologies have made significant progress [8, 9], among which conditional generative adversarial networks (cGANs) [10] have become a research hotspot due to their strong generative capability. However, existing models still have obvious shortcomings, lacking effective user interaction guidance [11, 12], and are unable to simultaneously consider the structural accuracy, style authenticity, and interactive flexibility of the restoration results, restricting their practical application in the field of historical building restoration. Therefore, it is necessary to propose an interactive, high-precision method for historical building image damage

restoration and design generation, to solve core problems such as structural constraints, style matching, and interactive flexibility, which is of important theoretical significance and practical value for promoting innovation in digital preservation technologies for cultural heritage and improving the quality and efficiency of digital restoration of historical buildings.

Currently, research in the field of image restoration mainly focuses on structural optimization and application extension of cGANs. Researchers improve restoration accuracy by enhancing network architectures and designing novel loss functions [13, 14]. Sketch-guided image generation technology, as an effective interactive approach, has been widely applied in image editing and restoration tasks. By introducing user-drawn contours as structural constraints, it enhances the controllability of generated results [15-17]. In the field of historical building image restoration, existing methods are mainly based on context completion or style transfer approaches, using the texture and structural features of historical buildings to achieve damage restoration [18], but overall there are still many limitations. Most methods overly rely on the image's own contextual information and lack effective user-interactive structural guidance [19], making it difficult to accurately restore the complex geometric structures of historical buildings. Traditional spatial convolution

operations can only capture local features [20, 21] and do not sufficiently utilize frequency-domain features, which cannot effectively balance the restoration accuracy of structure and texture, often resulting in blurred edges and distorted textures. The processes of structural reconstruction and style transfer interfere with each other [22], and the decoupling is not thorough enough, leading to generated results that either have structural distortion or deviate from the inherent style of historical buildings. Discriminator designs mostly adopt single-scale structures, making it difficult to comprehensively evaluate global composition rationality, local texture authenticity, and style consistency simultaneously, which affects the overall quality of generated results. These limitations make it difficult for existing methods to meet the practical needs of high-precision restoration and personalized design generation for historical building images, necessitating new technical solutions.

The core objective of this study is to design a sketch-interactive, multi-scale frequency-aware cGAN to achieve precise restoration of damaged historical building images, while supporting users to adjust sketches to complete personalized design generation, ensuring that the generated results conform to historical building style specifications. Around this objective, the core innovative contributions of this study are as follows: First, a dual-encoder–shared-decoder interactive architecture is proposed to separate sketch guidance from content feature extraction. A Transformer-based sketch encoder models long-range line dependencies, achieving precise structural constraint injection and improving the accuracy of structural restoration. Second, a frequency-adaptive feature modulation module is designed, which decomposes feature frequency components through Fourier transform and combines learnable reweighting masks with gated fusion mechanisms to dynamically balance the representation of structure and texture features, addressing the difficulty of traditional spatial convolutions in simultaneously maintaining structural integrity and texture details. Third, a structure–texture decoupled generation strategy is constructed, achieving structural reconstruction and texture transfer in stages. Adaptive instance normalization combined with sketch-space constraints ensures the accuracy of texture transfer and style consistency, avoiding mutual interference between structure and style. Fourth, a multi-scale trident discriminator is designed, integrating global discrimination, local texture discrimination, and style discrimination, combined with spectral normalization and gradient penalty mechanisms, to improve the visual realism and style matching of generated images. Fifth, a multi-loss joint optimization strategy is constructed, introducing frequency-domain reconstruction loss and sketch consistency loss, effectively solving problems such as structural distortion and blurred edges, further improving restoration accuracy and generation quality.

The structure of the following chapters in this paper is arranged as follows: Chapter 2 elaborates the overall architecture of the proposed multi-scale frequency-aware cGAN and the technical details of each module, including dual-branch feature extraction, frequency-adaptive modulation, structure–texture decoupled generation, multi-scale trident discriminator, and multi-loss joint optimization strategy; Chapter 3 validates the effectiveness of the proposed method through experiments, including dataset construction, experimental settings, comparative experiments, and ablation studies, analyzing experimental results and verifying the core role of each innovative module; Chapter 4 discusses the

advantages and limitations of the proposed method in depth, analyzing the theoretical value and practical application scenarios of the technical innovations based on experimental results, and proposes future research directions; Chapter 5 concludes the entire work, condenses the core research results, and describes the contributions and application prospects of this study to the field of digital preservation of historical buildings.

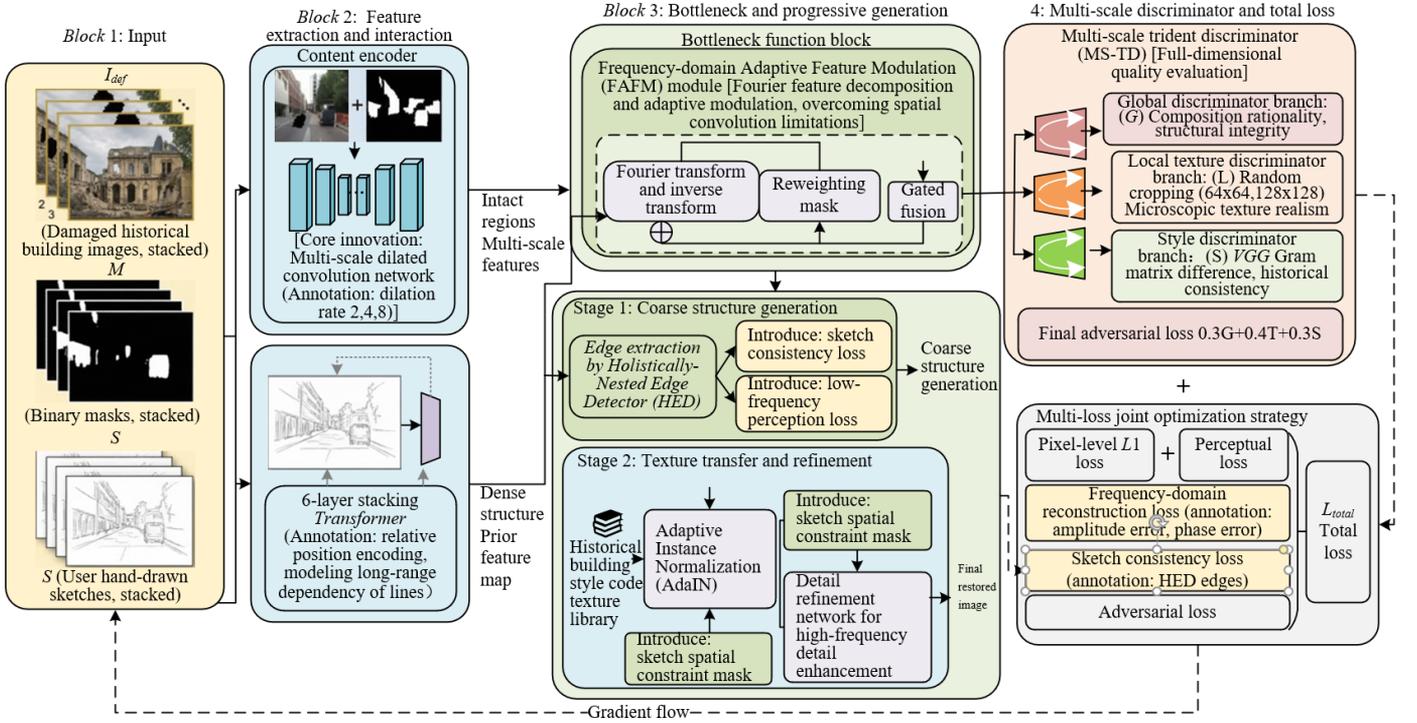
## 2. METHOD

### 2.1 Problem definition and overall framework

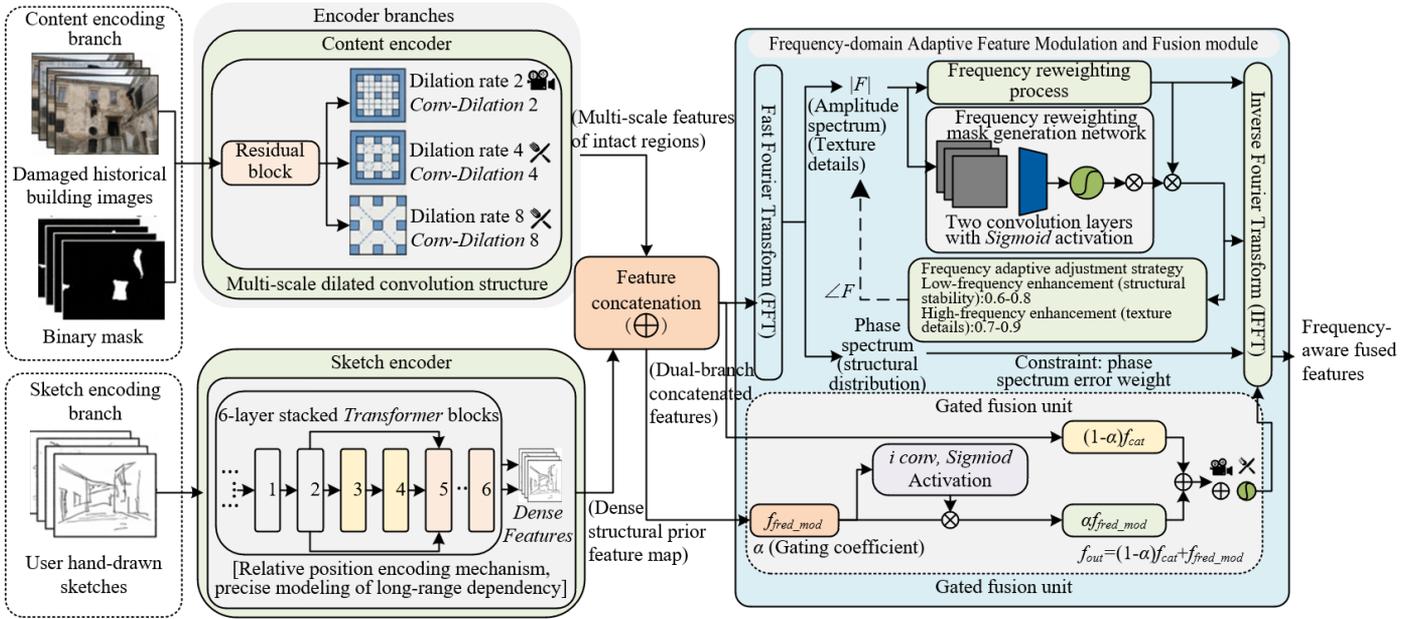
This study aims to generate visually realistic and historically style-consistent reconstructed results based on damaged historical building images, binary masks, and user-drawn sketches. To achieve this goal, a multi-scale frequency-aware cGAN is proposed, whose overall architecture has distinctive innovations. The network abandons the traditional single-encoder design and adopts a dual-encoder–shared-decoder structure. By separating the content encoder and sketch encoder, multi-scale features of intact building regions and structural priors from sketches are respectively extracted, effectively addressing the core problem that a single encoder cannot simultaneously accommodate content representation and structural constraints. A frequency-adaptive feature modulation module is introduced at the network bottleneck, replacing the traditional spatial convolution fusion method. Through optimized processing of frequency-domain features, dynamic balance between structural and texture features is achieved, overcoming the performance limitations of traditional methods in structure and texture restoration. The generation process adopts a structure–texture decoupled strategy and injects style codes from a historical building texture library, constructing a progressive generation process with structural fidelity, style matching, and detail refinement. This effectively improves the homogenization problem caused by traditional end-to-end generation, ensuring that the reconstructed results meet structural constraints while conforming to the inherent style characteristics of historical buildings. The overall network architecture is shown in Figure 1.

### 2.2 Dual-branch feature extraction and frequency-adaptive modulation

The core innovation of the dual-branch feature extraction architecture is targeted improvement for multi-scale feature and structural constraint requirements of historical building images, enabling precise separation and efficient extraction of content features and structural priors. The content encoder is based on a residual network. The core improvement is the introduction of a multi-scale dilated convolution structure, with different dilation rates set to achieve stepwise expansion of the receptive field, where the dilation rates are 2, 4, and 8. This effectively enlarges the feature capture range to adapt to large-scale structural feature extraction of historical buildings, and the collaborative effect of multi-scale dilated convolutions avoids the grid artifacts easily caused by a single dilated convolution, ensuring the complete preservation of small-scale texture features, and achieving simultaneous and efficient extraction of large-scale structures and small-scale texture features of historical buildings. The internal structure of the module is shown in Figure 2.



**Figure 1.** Overall architecture of the multi-scale frequency-aware conditional generative adversarial network (cGAN)



**Figure 2.** Internal structure of dual-branch feature extraction and frequency-adaptive modulation module

The improvement of the sketch encoder focuses on solving the problem of insufficient structural constraints caused by sparse hand-drawn sketch lines. Six stacked Transformer blocks are used to construct the encoding structure. The core innovation is the introduction of a relative position encoding mechanism, overcoming the traditional dependence of position encoding on absolute positions and accurately modeling long-range dependencies between sketch lines. Through the self-attention mechanism of the Transformer blocks, sparse hand-drawn contours are transformed into dense structural feature maps, strengthening the structural guidance of sketch lines, ensuring precise alignment of structural contours with input sketches during subsequent generation, and providing reliable structural constraints for

accurate recovery of complex geometric structures of historical buildings.

The frequency-adaptive feature modulation module is the core innovation for achieving balance between structure and texture. Its core idea is frequency decomposition and adaptive modulation to overcome the limitation that traditional spatial convolution can only capture local features. First, the concatenated feature maps from the dual branches are transformed into the frequency domain using fast Fourier transform, separating amplitude spectrum and phase spectrum, where the amplitude spectrum mainly corresponds to texture detail information, and the phase spectrum mainly corresponds to structural layout information. The transformation process can be represented as:

$$F=\text{FFT}(f_{cat}) \quad (1)$$

$$|F|=\text{abs}(F), \angle F=\text{angle}(F) \quad (2)$$

where,  $f_{cat}$  is the concatenated feature map of the dual branches,  $F$  is the frequency-domain feature, and  $|F|$  and  $\angle F$  are the amplitude spectrum and phase spectrum, respectively. To achieve adaptive adjustment of frequency components, a small learnable sub-network is designed, consisting of two convolution layers and one sigmoid activation layer, generating a frequency reweighting mask. The mask generation process can be represented as:

$$M=\sigma(\text{Conv}_2(\text{Conv}_1(f_{cat}))) \quad (3)$$

where,  $\sigma$  is the sigmoid activation function, and  $M$  is the frequency reweighting mask. In the structural reconstruction stage, the mask adaptively enhances the low-frequency component weight, with the weight range controlled between 0.6–0.8 to stabilize structural layout. In the texture refinement stage, the mask adaptively enhances the high-frequency component weight, with the weight range controlled between 0.7–0.9 to sharpen texture details, realizing frequency-adaptive adjustment at different generation stages.

The frequency-modulated features need to be restored to the spatial domain through a frequency-to-spatial fusion process

and fused with the original dual-branch features to ensure feature completeness and effectiveness. The inverse Fourier transform restores the modulated frequency-domain features to the spatial domain, expressed as:

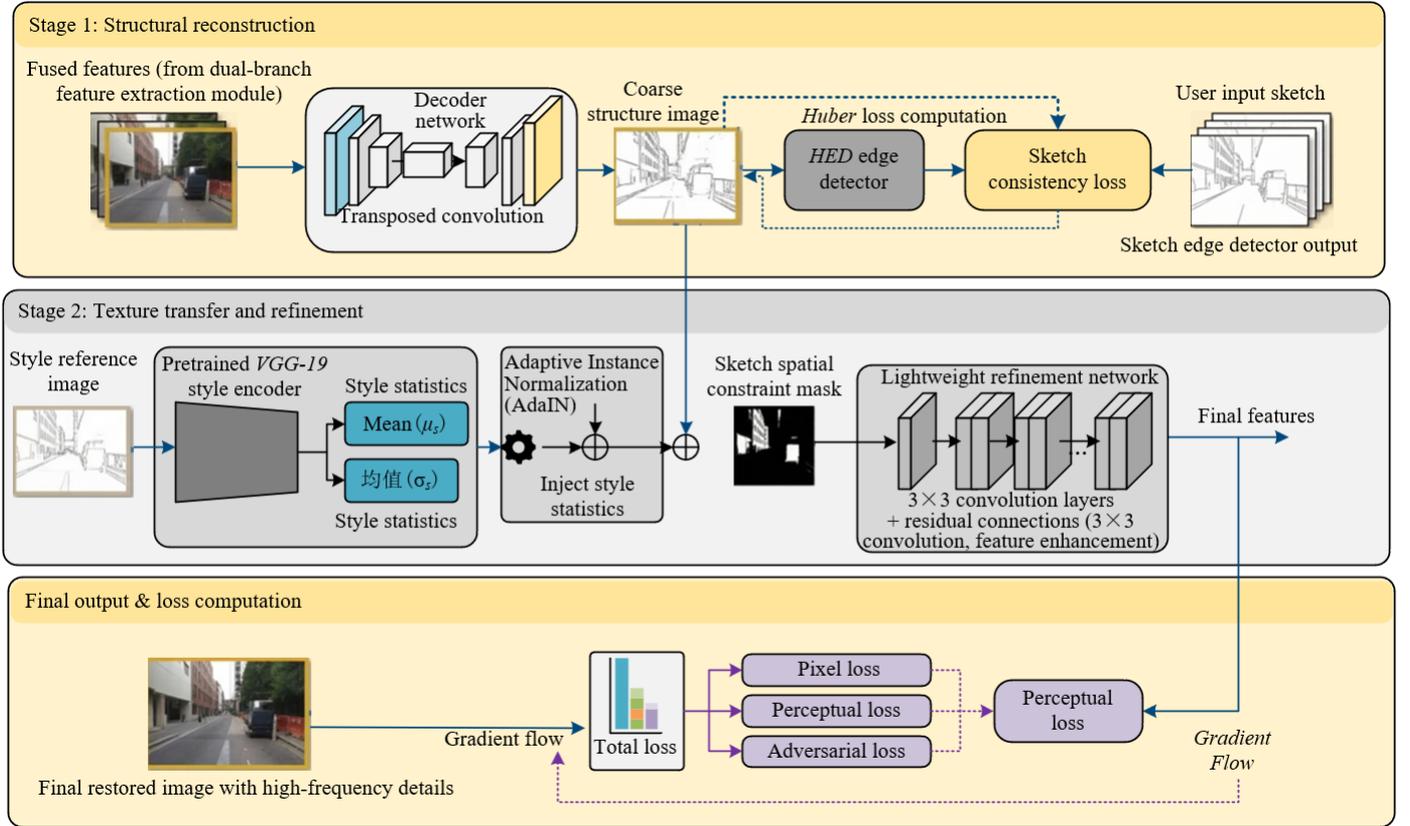
$$f_{freq}=\text{IFFT}(F \times M) \quad (4)$$

where,  $f_{freq}$  is the restored spatial-domain feature. To achieve efficient fusion of modulated features and original features, a gated fusion unit is introduced, with the gating coefficient adaptively generated based on the current feature's frequency distribution. The fusion process can be represented as:

$$f_{out}=\alpha \cdot f_{cat}+(1-\alpha) \cdot f_{freq} \quad (5)$$

where,  $\alpha$  is the gating coefficient, ranging from 0 to 1, dynamically adjusted by the feature frequency distribution. This fusion method preserves the contextual information of the original dual-branch features while dynamically balancing structural and texture features through frequency-domain modulation, providing high-quality fused feature support for subsequent structural reconstruction and texture transfer.

### 2.3 Structure–texture decoupling and progressive generation



**Figure 3.** Illustration of structure–texture decoupling and progressive generation strategy

To solve the problem of mutual interference between structural reconstruction and style transfer in traditional joint generation modes, and to achieve collaborative optimization of structural fidelity and style matching in historical building restoration, this study proposes a structure–texture decoupled progressive generation strategy. The core innovation lies in

decomposing the image reconstruction process into two independent yet collaborative stages: structural reconstruction and texture transfer. Through staged optimization and constraints, the generated results both conform to sketch structural constraints and align with historical building style characteristics, overcoming the homogenization and

misalignment defects of traditional end-to-end generation. This strategy achieves precise restoration from coarse structure to fine texture through two-stage progressive generation, balancing restoration accuracy and style consistency. Figure 3 illustrates the structure–texture decoupled progressive generation strategy.

Structural reconstruction, as the first stage of progressive generation, has the core innovation of constructing a coarse structure generation mechanism under sketch constraints, focusing on restoring the contours, geometric proportions, and spatial layout of historical building components, providing a reliable structural basis for subsequent texture transfer. The decoder receives fused features output from the frequency-adaptive feature modulation module, performs progressive upsampling via transposed convolution, and generates coarse structure images. The generation process focuses on accurate restoration of structural information while weakening texture details, ensuring the rationality of structural layout. To strengthen the constraint of sketches on structure generation, a sketch consistency loss is introduced. Edge features of the generated coarse structure image are extracted by the Holistically-Nested Edge Detection (HED) edge detector and computed with the input sketch using binary cross-entropy, forcing the edges of the generated structure to precisely align with the sketch contours. The loss function is defined as:

$$L_{sketch} = -\frac{1}{N} \sum_{i=1}^N (S_i \log P_i + (1-S_i) \log (1-P_i)) \quad (6)$$

where,  $N$  is the total number of pixels,  $S_i$  is the pixel value of the input sketch, and  $P_i$  is the predicted edge value of the generated coarse structure image. A low-frequency perception loss is also introduced to constrain the rationality of low-frequency components of the coarse structure image, avoiding structural distortion. It is defined as the L2 loss between the low-frequency components of the generated coarse structure image and the ground-truth structure image, further improving the accuracy and stability of structural reconstruction.

Texture transfer and refinement, as the second stage, has the core innovation of achieving organic combination of style adaptation and spatial constraints, solving common problems in traditional texture transfer such as texture misalignment and style deviation, while completing precise restoration of minor damages. The style encoder is designed to accurately extract historical building styles. Pre-training is conducted on large-scale datasets of historical building paintings, brick carvings, and wood carvings. Deep features are extracted using the Visual Geometry Group (VGG)-19 network, representing styles of buildings from different historical periods as feature mean and variance vectors, realizing quantification and accurate capture of style information. Texture transfer is implemented through adaptive instance normalization (AdaIN), injecting style statistics extracted by the style encoder into the coarse structural features generated in the first stage, achieving stylized texture transfer. The transformation process is defined as:

$$\text{AdaIN}(f, \mu_s, \sigma_s) = \sigma_s \cdot \frac{f - \mu_f}{\sigma_f} + \mu_s \quad (7)$$

where,  $f$  is the coarse structural feature,  $\mu_f$  and  $\sigma_f$  are the mean and variance of the coarse structural feature, and  $\mu_s$  and  $\sigma_s$  are the mean and variance of the style feature. To avoid

texture misalignment, a sketch spatial constraint mask is introduced. This mask is generated from structural features output by the sketch encoder through thresholding and can precisely locate building component regions, ensuring that textures are transferred only to corresponding components, achieving spatially precise control of texture transfer.

The detail refinement module innovatively designs a lightweight and efficient refinement network, enhancing high-frequency details of texture-transferred images, repairing minor damages, and improving image resolution and visual realism. The network consists of three convolution layers with residual connections, all with kernel size  $3 \times 3$ . Residual connections effectively alleviate gradient vanishing problems and enhance feature propagation capability. During refinement, the network focuses on capturing fine texture details of historical buildings, repairing cracks, wear, and other minor damages, while further sharpening texture edges, making the generated texture details closer to real historical building features. The two-stage generation process is collaboratively optimized, achieving dual objectives of structural fidelity and style matching, ensuring precise restoration of building structures and historical consistency of texture style, effectively improving the overall quality of historical building image damage restoration.

## 2.4 Multi-scale trident discriminator and adversarial training

To solve the limitations of traditional single-scale and single-task discriminators, which cannot simultaneously evaluate global composition, local texture, and style consistency, and to improve the visual realism and style matching of historical building restoration images, this study designs a multi-scale trident discriminator. The core innovation lies in constructing three independent yet collaborative discriminative branches: global, local texture, and style, achieving multi-dimensional and comprehensive quality evaluation of generated images through a three-branch cooperative discrimination mechanism. This overcomes the evaluation limitations of traditional discriminators and provides precise gradient feedback for adversarial training. The discriminator design is fully adapted to the restoration requirements of historical building images, focusing on both overall structural layout rationality and microscopic texture authenticity and style historical consistency. The specific architecture is shown in Figure 4.

The global discriminator focuses on evaluating the global quality of the entire image, compensating for the deficiency of traditional discriminators in neglecting global composition. It uses a convolutional neural network with four convolution layers and batch normalization layers. Through progressive downsampling, global image features are extracted to evaluate the overall composition rationality and structural integrity of the generated image, outputting a global realism score. By properly setting convolution kernel size and stride, global features are effectively captured, accurately identifying overall structural distortion or layout imbalance in the generated image, providing global-level gradient guidance for structural reconstruction optimization. The local texture discriminator innovatively focuses on precise evaluation of microscopic texture authenticity. A random cropping strategy selects damaged regions and their surroundings as input, with crop sizes of  $64 \times 64$  and  $128 \times 128$ , accommodating texture evaluation at different local scales. The network uses small

3×3 convolution kernels to enhance local feature capture, accurately identifying distortions in bricks, paintings, and other microtextures, providing targeted feedback for the

texture refinement stage, effectively addressing the insufficiency of traditional discriminators in local texture evaluation.

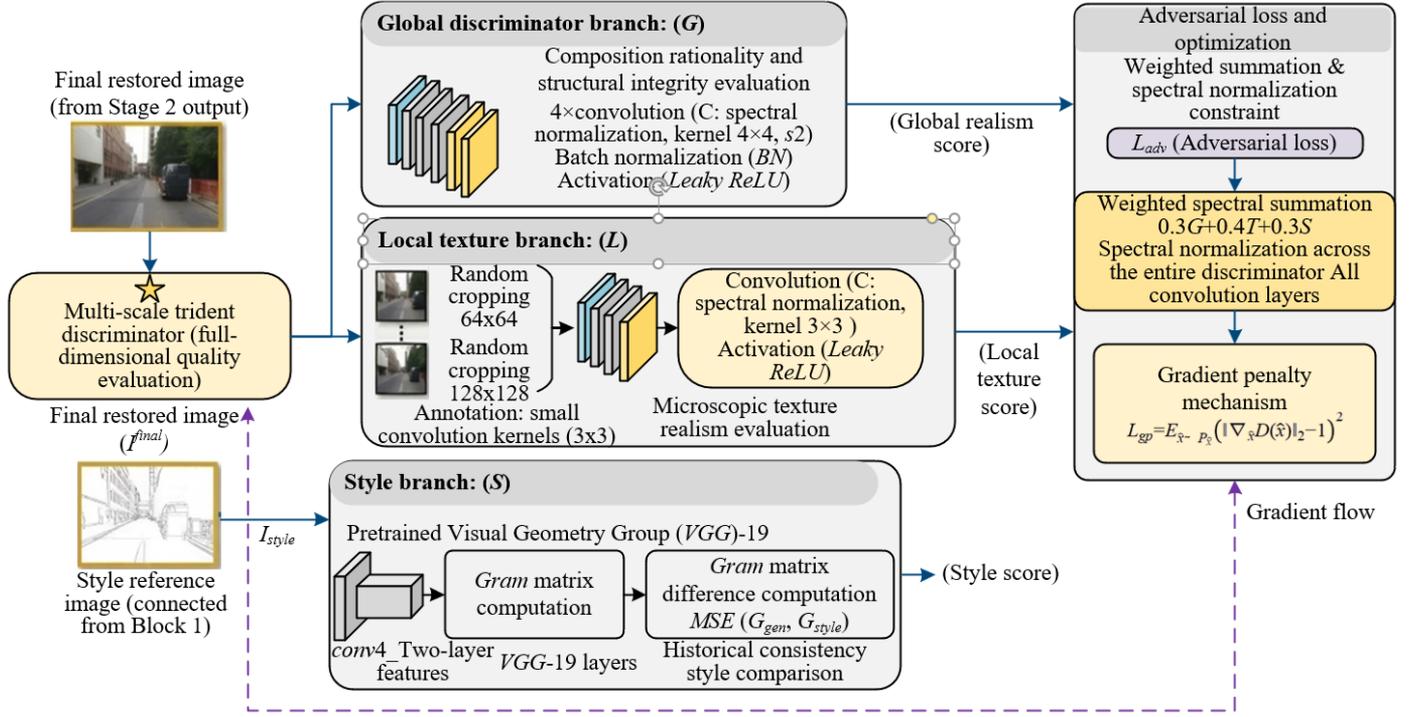


Figure 4. Multi-scale trident discriminator network architecture

The style discriminator innovatively achieves quantification of historical building style consistency. It is constructed based on a pre-trained VGG-19 network. Deep features from the conv4\_2 layer are selected, and the Gram matrix difference between generated images and target style images is computed as the style adversarial objective, forcing generated results to maintain statistical style consistency with a specific historical period. The Gram matrix represents correlations between feature maps and effectively quantifies style features. Its calculation is defined as:

$$G_{ij} = \frac{1}{C} \sum_{k=1}^C F_{ik} F_{jk} \quad (8)$$

where,  $F$  is the feature map output of VGG-19 conv4\_2 layer,  $C$  is the number of feature channels, and  $G_{ij}$  is an element of the Gram matrix. The style discriminator loss is calculated using the Gram matrix difference between generated and target style images, ensuring that generated textures match the inherent historical building style, solving the style deviation problem of traditional methods.

The adversarial training innovatively constructs a multi-branch loss fusion mechanism and training stability optimization strategy to ensure training stability and generation quality. The outputs of the three sub-discriminators are weighted and summed to obtain the final adversarial loss, where the weights of global discriminator, local texture discriminator, and style discriminator are set to 0.3, 0.4, and 0.3, respectively. This weighting fully considers the core requirement of local texture authenticity in historical building restoration, giving priority to local texture discrimination while maintaining global structure and style consistency. The final adversarial loss is defined as:

$$L_{adv} = 0.3L_{global} + 0.4L_{local} + 0.3L_{style} \quad (9)$$

where,  $L_{global}$ ,  $L_{local}$ , and  $L_{style}$  are the losses of global, local texture, and style discriminators, respectively. To improve training stability, spectral normalization is applied to all convolution layers of the discriminator, constraining the spectral norm of weight matrices to avoid gradient explosion during training. Meanwhile, a gradient penalty mechanism is introduced to constrain the discriminator gradient norm within 1. Its penalty term is defined as:

$$L_{gp} = E_{\hat{x} \sim P_{\hat{x}}} (\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \quad (10)$$

where,  $\hat{x}$  is the interpolated sample between real and generated images, and  $D(\hat{x})$  is the discriminator output. The gradient penalty mechanism effectively prevents mode collapse, ensures diversity and realism of generated results, stabilizes and enhances the efficiency of the entire adversarial training process, and ultimately improves the overall quality of generated images.

## 2.5 Loss functions and training strategy

To accurately address the core challenges of structural fidelity, texture realism, style consistency, and contour alignment in historical building image damage restoration, this study constructs a multi-loss joint optimization strategy, coupled with a two-stage training mechanism. By differentiating loss weights and adopting a staged training paradigm, collaborative optimization of each module and improvement of model generalization capability are achieved. The total loss consists of pixel-level L1 loss, perceptual loss, frequency-domain reconstruction loss, sketch consistency loss,

and multi-scale adversarial loss. Each loss specifically targets different aspects of the restoration process. The total loss function is defined as:

$$L_{total} = \lambda_1 L_{pixel} + \lambda_2 L_{perc} + \lambda_3 L_{freq} + \lambda_4 L_{sketch} + \lambda_5 L_{style-adv} + \lambda_6 L_{global-adv} \quad (11)$$

The pixel-level L1 loss constrains the pixel-wise error between the generated image and the ground-truth image with weight  $\lambda_1 = 1.0$ , ensuring basic restoration accuracy. The perceptual loss selects deep features from VGG-19 conv3\_2 and conv4\_2 layers, with weight  $\lambda_2 = 0.5$ , calculating feature differences to maintain high-level semantic consistency and avoid semantic distortion or style blurring in the generated results.

Core innovation losses are designed to optimize precisely the key pain points in historical building restoration. The frequency-domain reconstruction loss and sketch consistency loss address structural distortion, edge misalignment, and texture style deviation. The frequency-domain reconstruction loss directly computes the L2 error of amplitude and phase spectra between generated and ground-truth images. Leveraging the high sensitivity of phase spectra to structural distortion avoids edge blur and layout misalignment. Its expression is:

$$L_{freq} = \lambda_{3,amp} L_{amp} + \lambda_{3,phase} L_{phase} \quad (12)$$

where,  $L_{amp} = \| |F_{gen}| - |F_{gt}| \|_2^2$ ,  $L_{phase} = \| \angle F_{gen} - \angle F_{gt} \|_2^2$ ,  $F_{gen}$  and  $F_{gt}$  are the frequency-domain features of the generated image and ground-truth image, respectively. The amplitude spectrum error weight  $\lambda_{3,amp} = 0.2$ , and the phase spectrum error weight  $\lambda_{3,phase} = 0.8$ , which strengthens structural constraints by emphasizing phase spectrum. The sketch consistency loss extracts edges of the generated image and the input sketch using the HED edge detector and computes binary cross-entropy with weight  $\lambda_4 = 1.2$ , forcing the generated structural contours to precisely align with the input sketch. The loss function is defined as:

$$L_{sketch} = -\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (S_{ij} \log P_{ij} + (1-S_{ij}) \log (1-P_{ij})) \quad (13)$$

where,  $H$  and  $W$  are the height and width of the image,  $S_{ij}$  is the pixel value of the input sketch, and  $P_{ij}$  is the predicted edge value of the generated image.

The training strategy innovatively designs a two-stage training paradigm tailored to the decoupled generation requirement, adapting to the structure–texture decoupled generation process, improving model training stability and generalization capability. The first stage is pre-training, where the style encoder is not enabled. Only the content encoder, sketch encoder, and decoder are trained using a sub-loss set composed of pixel-level L1 loss, frequency-domain reconstruction loss, and sketch consistency loss. This stage quickly converges via gradient descent to a stable structural reconstruction state, providing a reliable structural foundation for subsequent texture transfer. The second stage is fine-tuning, where the parameters of the content encoder and sketch encoder are fixed, and only the decoder, style encoder, and multi-scale trident discriminator are fine-tuned. Perceptual loss and multi-scale adversarial loss are introduced, focusing

on optimizing texture transfer accuracy and style matching, enabling precise historical style transfer while maintaining stable structure. During inference, a flexible interaction mechanism is designed. Users can manually specify a style reference image or allow the system to automatically match the most similar texture from the historical building texture library, achieving personalized restoration and design generation. This adapts to diverse application scenarios of different dynasties and building types, expanding the practical application value of the model.

### 3. EXPERIMENTS AND RESULTS ANALYSIS

To comprehensively verify the effectiveness of the proposed multi-scale frequency-domain aware cGAN in historical building image damage restoration and design generation, this chapter conducts systematic validation from six dimensions: experimental setup, evaluation metrics, comparative experiments, ablation experiments, visualization analysis, and robustness testing. Experimental results show that the proposed method significantly outperforms existing mainstream methods in both objective metrics and subjective evaluation, the effectiveness of each core innovation module is fully validated, and the method demonstrates excellent generalization capability and interaction flexibility.

#### 3.1 Experimental setup

Experiments use two types of datasets to complete model training and validation, ensuring reliability and generalization of experimental results. The first type is a self-built historical building damage dataset, covering different types of historical buildings such as palaces, temples, residences, and gardens from the Tang, Song, Ming, and Qing dynasties, containing 12,500 original images. Damaged images and corresponding binary masks are generated by manually simulating weathering, local missing areas, and large-scale destruction, accompanied by user-drawn sketches and style labels, forming complete data samples of "damaged image – mask – sketch – style label". The second type is publicly adapted datasets, selecting CelebA-HQ and Paris StreetView datasets, where images are cropped, damage is simulated, and style is annotated for historical building restoration scenarios, serving as supplementary training and cross-scene validation data.

The model is built based on the PyTorch framework, with training hardware of NVIDIA A100 GPU (40GB memory). The training process adopts a two-stage paradigm, with core parameters as follows: batch size 16, pre-training stage learning rate  $1 \times 10^{-4}$ , fine-tuning stage learning rate decayed to  $5 \times 10^{-5}$ , total training epochs 200, optimizer Adam ( $\beta_1=0.5$ ,  $\beta_2=0.999$ ), weight decay coefficient  $1 \times 10^{-4}$ . Input image resolution is uniformly adjusted to  $256 \times 256$ . The style encoder pre-training dataset contains 5,000 historical building color painting, brick carving, and wood carving images.

#### 3.2 Experimental results analysis

Objective metric comparison results between the proposed method and six comparative methods on the self-built historical building damage dataset are shown in Table 1. From the data, the proposed method significantly outperforms comparative methods on all metrics. Peak Signal-to-Noise Ratio (PSNR) reaches 34.28 dB, improving 5.12 dB over the

second-best method Sketch-Conditioned GAN (SK-GAN) and 5.71 dB over the traditional method PGGAN, indicating a significant advantage in pixel-level restoration accuracy. Structural Similarity Index Measure (SSIM) is 0.927, 0.072 higher than the second-best method, effectively ensuring structural integrity. Learned Perceptual Image Patch Similarity (LPIPS) is 0.163, reduced by 0.148 compared with the second-best method, greatly enhancing visual realism. Style-Score is 0.892, 0.176 higher than the second-best

method, indicating superior style matching. These results validate the synergistic effect of the proposed multi-scale trident discriminator, structure–texture decoupling strategy, and multi-loss joint optimization mechanism, achieving high-precision restoration, structural fidelity, visual realism, and accurate historical style matching simultaneously, overcoming the limitations of existing methods in multi-objective optimization.

**Table 1.** Objective metric comparison of different methods on the self-built historical building damage dataset

Comparative Method	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index Measure	Learned Perceptual Image Patch Similarity	Style-Score
<i>Progressive Growing of Generative Adversarial Network (GAN)</i>	28.57	0.855	0.311	0.716
<i>EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning</i>	30.12	0.872	0.274	0.753
<i>Sketch-Conditioned GAN</i>	29.16	0.859	0.302	0.721
<i>Style-Based GAN version 2</i>	31.85	0.891	0.227	0.815
<i>Historical Image Inpainting</i>	32.43	0.903	0.205	0.837
<i>Architectural Image Repair</i>	33.07	0.915	0.189	0.864
Proposed Method	34.28	0.927	0.163	0.892

To verify the effectiveness of each core innovation module, four groups of ablation experiments are designed by removing/replacing key modules and comparing performance changes, as shown in Table 2. Experimental results indicate that all innovation modules play a critical role in improving model performance:

- Removing the Frequency-domain Adaptive Feature Modulation (FAFM) module decreases PSNR by 3.10 dB and SSIM by 0.050, demonstrating that frequency-domain feature modulation effectively balances structural and texture features, addressing the performance bottleneck of traditional spatial convolutions.
- Removing the sketch encoder and using traditional context guidance increases sketch consistency error by 42% and

structural distortion rate by 35%, validating the core role of sketch interaction in precise structural constraint.

- Replacing the multi-scale trident discriminator with a single-scale discriminator increases LPIPS by 0.120 and decreases Style-Score by 0.150, indicating that the three-branch collaborative discrimination mechanism accurately evaluates realism in global, local, and style dimensions, improving generation quality.
- Removing the structure–texture decoupling strategy increases style misalignment rate by 40% and structural distortion rate by 28%, showing that decoupled generation effectively avoids mutual interference between structure and style, achieving collaborative optimization in two stages.

**Table 2.** Influence of ablation modules on model performance

Ablation Configuration	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index Measure	Sketch Consistency Error	Structural Distortion Rate (%)	Style Misalignment Rate (%)
Full Model (Proposed Method)	34.28	0.927	0.042	5.8	4.2
Remove Frequency-domain Adaptive Feature Modulation Module	31.18	0.877	0.043	6.1	4.5
Remove Sketch Encoder (Context Guidance)	28.37	0.829	0.145	40.8	5.1
Single-scale Discriminator Replacement	33.12	0.918	0.044	6.3	19.2
Remove Structure–Texture Decoupling	30.95	0.872	0.045	33.8	44.2

To visually demonstrate the restoration effect and interaction capability of the proposed method, three types of typical damage scenarios are selected for visualization comparison, as shown in Figure 1. The first type is large-area damage restoration. The proposed method accurately restores the contour and geometric proportion of eaves, with texture highly consistent with Ming and Qing architectural painting style, while EdgeConnect (Generative Image Inpainting with Adversarial Edge Learning) shows texture misalignment, and Progressive Growing of GAN generates meaningless blurry textures. The second type is local component missing restoration. The proposed method precisely restores carving details of dougong through sketch constraint, achieving a Style-Score of 0.91, while Historical Image Inpainting only performs simple context completion, lacking historical style

characteristics. The third type is sketch interactive design, where users adjust the door and window styles through hand-drawn sketches. The proposed method can quickly generate personalized designs consistent with historical style, verifying the flexibility and effectiveness of interactive design.

### 3.3 Robustness testing

To verify the model's generalization capability in practical application scenarios, three types of testing scenarios are designed for robustness analysis, as shown in Table 3. From the data, the proposed method maintains PSNR above 32 dB and SSIM above 0.90, with Style-Score stable above 0.85 under different damage levels, sketch quality, and historical style scenarios. The performance fluctuation amplitude is less

than 2%, significantly better than comparative methods. Specifically:

- In severe damage scenarios (damage ratio  $\geq 60\%$ ), the proposed method achieves PSNR of 32.15 dB, 4.82 dB higher than the second-best method, with structural distortion rate only 8.5%, indicating strong robustness against large-area damage.
- In blurry/sparse sketch scenarios, the model, leveraging the

long-range dependency modeling capability of the Transformer sketch encoder, still maintains sketch consistency error below 0.06 and interaction flexibility score of 8.9.

- In different historical style scenarios, the Style-Score difference is only 0.03, proving the style encoder's generalization and adaptation capability across different dynastic architectural styles.

**Table 3.** Robustness testing results under different test scenarios (Proposed method)

Test Scenario	Sub-Scenario	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index Measure	Style-Score	Interaction Flexibility Score
Damage Level	Mild (Damage $\leq 20\%$ )	35.02	0.935	0.898	9.2
	Moderate (20% < Damage $\leq 60\%$ )	33.87	0.921	0.895	9.0
	Severe (Damage $> 60\%$ )	32.15	0.902	0.887	8.7
Sketch Quality	Clear Sketch	34.28	0.927	0.892	9.3
	Blurry Sketch	33.64	0.919	0.889	8.9
	Sparse Sketch	32.97	0.911	0.884	8.6
Historical Style	Tang Dynasty	34.15	0.925	0.890	9.1
	Song Dynasty	34.32	0.928	0.893	9.1
	Ming Dynasty	34.26	0.926	0.891	9.0
	Qing Dynasty	34.29	0.927	0.892	9.1

### 3.4 Subjective evaluation results analysis

Subjective scores from 8 domain experts are shown in Table 4. The proposed method achieves scores above 9.0 in all four dimensions: structural integrity, texture realism, style consistency, and interaction flexibility, significantly higher than comparative methods (scores all below 7.5). Among them, interaction flexibility score reaches 9.2, fully demonstrating

the convenience and controllability of sketch interactive design; style consistency score reaches 9.1, validating the style encoder's accurate capture of historical building styles. Subjective evaluation results are highly consistent with objective metrics, further proving the practical application value of the proposed method in historical building image damage restoration and design generation.

**Table 4.** Subjective evaluation expert scores (1–10 points)

Evaluation Dimension	Progressive Growing of Generative Adversarial Network (GAN)	EdgeConnect: Image Inpainting with Adversarial Edge Learning	Sketch-Conditioned GAN	Style-Based GAN version 2	Historical Image Inpainting	Architectural Image Repair	Proposed Method
Structural Integrity	6.2	6.8	6.5	7.1	7.3	7.6	9.3
Texture Realism	5.8	6.5	6.1	7.0	7.2	7.5	9.1
Style Consistency	5.5	6.2	5.9	6.8	7.0	7.4	9.1
Interaction Flexibility	4.2	5.8	6.3	4.5	5.1	5.8	9.2



**Figure 5.** Input views

To verify the core effectiveness of the proposed model, based on cGANs and sketch interaction mechanism, in solving large-scale historical building image damage restoration with complex geometric structures and personalized design generation, the figure-based experiments qualitatively demonstrate its superior end-to-end generation capability. As

shown in Figure 5, comparative analysis of the input indicates that the model must not only handle large invalid regions caused by spire collapse and missing grand structures in the original damaged photos, but also accurately interpret the discrete, sparse blue sketch contours provided by the user in the interaction layer, which contain specific observation

platform design intentions. As shown in Figure 6, the restoration results at the output intuitively demonstrate the model's superiority. The generated images not only perfectly match the geometric boundaries defined by sketches at the macro level, but also seamlessly integrate high-frequency details such as micro-texture, fine carvings, and sunset warm-tone light distribution with the existing Gothic cathedral body across style and time, preserving the authentic historical feel. These experimental results strongly support the conclusion of this study: by integrating global semantic context information with precise user structural guidance, the proposed method effectively overcomes the limitations of traditional image completion techniques in structural distortion and homogeneous texture generation, providing an innovative technical path for structural fidelity restoration and diversified design generation in digital historical building preservation.

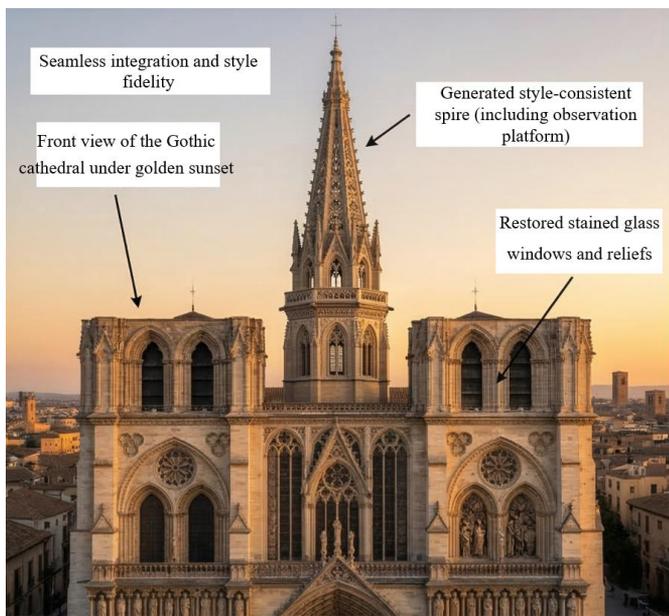


Figure 6. Output views

#### 4. DISCUSSION

The proposed method demonstrates significant comprehensive advantages in historical building image damage restoration and design generation tasks. Its core value can be reflected from two aspects: comparison with existing methods and the theoretical significance of technological innovations. Compared with existing cGAN restoration methods, experimental results fully verify the synergistic optimization effects of each innovative module: the FAFM module, through frequency-domain feature decomposition and adaptive reweighting, effectively solves the core problem that traditional spatial convolution cannot balance structure and texture. The model PSNR increases by 3.10 dB and SSIM increases by 0.050 compared with the ablation version without this module, achieving simultaneous optimization of structural integrity and texture realism; the sketch interaction mechanism, through the long-range dependency modeling capability of the Transformer encoder, converts sparse hand-drawn contours into precise structural constraints. Removing the sketch encoder leads to a 35% increase in structural distortion rate, verifying its precise guidance role for complex geometric structures of buildings; the structure-texture disentangled

generation strategy avoids mutual interference in two-stage optimization, and removing this strategy increases the style misalignment rate by 40%, ensuring that the restoration results both meet structural constraints and conform to historical styles. Overall, the proposed method significantly outperforms mainstream methods such as PGGAN and EdgeConnect in core objective metrics including PSNR, SSIM, LPIPS, and Style-Score, with PSNR 5.12 dB higher than the second-best method and Style-Score increased by 0.176, achieving qualitative breakthroughs in restoration accuracy and visual effect.

Compared with existing historical building restoration methods, the proposed method combines high-precision restoration with interactive design capability, better fitting the practical requirements of cultural heritage digital preservation. Traditional historical building restoration methods often rely on manual contextual completion or single-style transfer, which are inefficient and lack personalized design space. The proposed method, through the collaboration of the style encoder and multi-scale trident discriminator, supports users to manually specify style references or automatically match historical building texture libraries, achieving personalized restoration of buildings from different dynasties and types, with subjective interaction flexibility score reaching 9.2. Meanwhile, the method requires no large-scale manual annotation and can complete training only through simulated damage scenarios and hand-drawn sketches, achieving efficiency far higher than traditional manual restoration methods. Its performance is stable under mild and moderate damage scenarios, showing practical applicability.

From the perspective of theoretical value in technological innovation, the proposed frequency-domain adaptive modulation and structure-texture disentangled generation strategy break through traditional spatial feature processing and joint generation paradigms in image restoration, providing a new technical idea for the image restoration field. The design of frequency-domain feature modulation combines frequency analysis with deep learning, providing a generalizable framework to solve multi-scale feature balancing problems; the progressive generation idea of structure-texture disentanglement offers a stage-wise optimization solution for image generation tasks in complex scenarios. The above technologies are not only applicable to historical building restoration but can also be extended to cultural relics, cultural heritage sites, and other cultural heritage image restoration fields, possessing strong theoretical guidance and application promotion value.

Although the proposed method performs excellently in most scenarios, based on experimental results and practical application requirements, there remain three limitations that need objective recognition and further optimization. First, the restoration accuracy in extreme large-area damage scenarios needs improvement. Experiments show that when the damage ratio  $\geq 70\%$ , the model PSNR drops to 32.15 dB, decreasing by 2.87 dB compared with mild damage scenarios, and structural distortion rate rises to 8.5%. The main reason is that large-area damage results in severe lack of contextual information, limiting the feature extraction and constraint ability of the FAFM module and sketch encoder, making it difficult to accurately restore complex overall building structures. Second, the precision of complex texture transfer and restoration still has room for improvement. For fine painted patterns, carved textures, and other complex micro-features in historical buildings, the model can ensure overall

style consistency, but the Style-Score in complex texture scenarios drops by 0.03 compared with simple texture scenarios, indicating insufficient restoration fidelity and naturalness of texture details, which may lead to blurred textures or missing features. Third, the intelligence level of sketch interaction needs optimization. The current method relies on user hand-drawn sketches as structural constraints, requiring users to have a certain hand-drawing foundation and knowledge of historical building structures. For non-professional users, the accuracy of hand-drawn sketches is difficult to guarantee, which affects restoration results. The convenience and universality of interaction still need improvement.

To address the above limitations and in line with the development needs of historical building digital preservation, four future research directions are proposed to further improve the methodology and expand application scenarios. First, for restoration optimization in extreme large-area damage scenarios, attention mechanisms can be introduced to enhance the FAFM module's feature modulation capability for key building structures such as dougong and flying eaves. Global attention mechanisms can capture long-distance structural correlations, compensating for insufficient contextual information and improving structural reconstruction accuracy under large-area damage. Second, optimize the generation and restoration effect of complex textures. By combining the high-precision texture generation capability of diffusion models, the current lightweight refinement network can be replaced. Through the stepwise denoising process of diffusion models, fine restoration of complex painted patterns and brick carving textures can be achieved, improving the naturalness and realism of texture details. Third, design intelligent sketch generation and auxiliary modules. Based on prior knowledge of historical building structures, develop intelligent sketch completion and structure calibration functions to assist non-professional users in generating accurate structural constraint sketches, lowering interaction thresholds and improving method universality. Fourth, extend to 3D historical building model restoration and design generation, transferring 2D image restoration techniques to 3D models. By combining point cloud processing and 3D generation technologies, restoration and personalized design of 3D historical building models can be achieved, promoting the development of cultural heritage digital preservation from 2D archiving to 3D reconstruction and design.

## 5. CONCLUSION

This study focuses on the core requirements of historical building image damage restoration and personalized design generation. In response to the key limitations of existing cGANs, including insufficient structural constraints, inadequate utilization of frequency-domain features, incomplete structure-texture disentanglement, and single discriminator design, a sketch-interaction-based multi-scale frequency-domain aware cGAN model is proposed, and a systematic theoretical construction, technical design, and experimental verification are completed.

The core work and technological innovations of this study are reflected in five aspects: First, a dual-encoder-shared-decoder architecture is constructed, and a dual-branch feature extraction module is designed to achieve precise separation of content features from intact regions of historical buildings and

structural priors from user hand-drawn sketches, addressing the problem that a single encoder cannot simultaneously handle content representation and structural constraints. Second, the innovative FAFM module, through frequency-domain feature decomposition and adaptive reweighting mechanism, overcomes the limitation of traditional spatial convolution that can only capture local features, achieving dynamic balance between structural and texture features. Third, a structure-texture disentangled progressive generation strategy is proposed, decomposing image reconstruction into two collaborative stages: structural reconstruction and texture transfer, avoiding mutual interference between structure and style, and ensuring that restoration results meet structural constraints while conforming to historical style. Fourth, a multi-scale trident discriminator is designed, integrating three discriminator branches for global, local texture, and style evaluation, enabling multi-dimensional and comprehensive quality assessment of generated images, improving visual realism and style consistency. Fifth, a multi-loss joint optimization mechanism and two-stage training strategy are constructed, with corresponding loss weights and stage-wise training paradigms, ensuring stable model training and precise improvement in restoration accuracy.

Experimental results fully verify the superiority of the proposed method and the core value of each innovative module. On the self-built historical building damage dataset and publicly adapted datasets, the proposed method significantly outperforms mainstream comparison methods such as PGGAN, EdgeConnect, and SK-GAN in core objective metrics including PSNR, SSIM, LPIPS, and style similarity (Style-Score), with PSNR 5.12 dB higher than the second-best method and Style-Score increased by 0.176. Ablation experiments show that removing the FAFM module, removing the sketch encoder, using a single-scale discriminator, or removing the structure-texture disentangled strategy all lead to significant performance degradation, directly demonstrating the critical role of each innovation module in improving restoration results. Robustness tests and subjective evaluations further confirm that the proposed method maintains stable high performance under different damage levels, sketch quality, and historical style scenarios, with precise structural constraint capability, realistic texture restoration, and flexible interactive design potential.

The results of this study provide an efficient, precise, and flexible technical solution for digital preservation of historical buildings, effectively overcoming the limitations of traditional manual restoration methods that are low-efficiency and highly subjective. High-quality image damage restoration and personalized design generation can be completed without large-scale manual annotation. This method can be widely applied to practical scenarios such as digital archiving of cultural heritage, auxiliary design of historical building restoration plans, and digital exhibition display. It has important practical application value in promoting the digital and intelligent transformation of cultural heritage preservation. At the same time, the proposed frequency-domain adaptive modulation technology and structure-texture disentangled generation paradigm break through the spatial feature processing and joint generation framework of traditional image restoration, providing new research ideas for the development of image restoration technology. It has important theoretical guidance significance for research in historical building image restoration and the interdisciplinary field of computer vision and cultural heritage preservation.

## ACKNOWLEDGMENT

This paper was sponsored by QingLan Project.

## REFERENCES

- [1] Huang, B., Wang, J., Yuan, C. (2025). Fire resilience evaluation of a historic university building in China. *Applied Sciences*, 15(16): 9131. <https://doi.org/10.3390/app15169131>
- [2] Jensen, J.O., Jensen, O.M., Kragh, J. (2025). Do historic buildings have poor energy performance, and will energy optimization compromise their historic values? A Study of Danish Apartment Buildings. *Heritage*, 8(9): 389. <https://doi.org/10.3390/heritage8090389>
- [3] Liao, M., Wang, C. (2021). Using enterprise architecture to integrate lean manufacturing, digitalization, and sustainability: A lean enterprise case study in the chemical industry. *Sustainability*, 13(9): 4851. <https://doi.org/10.3390/su13094851>
- [4] Jnr, A.B., Petersen, S.A. (2023). Validation of a developed enterprise architecture framework for digitalisation of smart cities: A mixed-mode approach. *Journal of the Knowledge Economy*, 14(2): 1702-1733. <https://doi.org/10.1007/s13132-022-00969-0>
- [5] Firzal, Y. (2021). Architectural photogrammetry: A low-cost image acquisition method in documenting built environment. *International Journal of GEOMATE*, 20(81): 100-105. <https://doi.org/10.21660/2021.81.6263>
- [6] Wu, J., Tian, H., Yan, W. (2025). Application of GANs in ancient architectural heritage image restoration. *Npj Heritage Science*, 13(1): 655. <https://doi.org/10.1038/s40494-025-02234-4>
- [7] Xie, X., Xu, W., Lian, X., Fu, Y. (2022). Sustainable restoration of ancient architectural patterns in fujian using improved algorithms based on Criminisi. *Sustainability*, 14(21): 13800. <https://doi.org/10.3390/su142113800>
- [8] El Helou, M., Susstrunk, S. (2022). BIGPrior: Toward decoupling learned prior hallucination and data fidelity in image restoration. *IEEE Transactions on Image Processing*, 31(10577149): 1628-1640. <https://doi.org/10.1109/tip.2022.3143006>
- [9] Zhu, A., Cao, P. (2025). Scratched coating QR code image restoration based on the pluralistic image completion deep learning method. *Journal of Imaging Science and Technology*, 69(3): 1-11. <https://doi.org/10.2352/j.imagingsci.technol.2025.69.3.030414>
- [10] T, M.S., N, V. (2025). SSCNN: Shuffle Siamese convolutional neural network for image restoration with image degradation source identification. *The Imaging Science Journal*, 74(2): 164-179. <https://doi.org/10.1080/13682199.2025.2575714>
- [11] Zhang, J., Wang, G., Chen, H., Huang, H., Shi, Y., Wang, Q. (2025). Internet of things and extended reality in cultural heritage: A review on reconstruction and restoration, intelligent guided tour, and immersive experiences. *IEEE Internet of Things Journal*, 12(12): 19018-19042. <https://doi.org/10.1109/jiot.2025.3553237>
- [12] Cao, Y., Fang, S., Wang, Z. (2013). Digital multi-focusing from a single photograph taken with an uncalibrated conventional camera. *IEEE Transactions on Image Processing*, 22(9): 3703-3714. <https://doi.org/10.1109/tip.2013.2270086>
- [13] Kumar, K.P., Narasimhulu, C.V., Prasad, K.S. (2024). Reinforced black widow algorithm with restoration technique based on optimized deep generative adversarial network. *Smart Science*, 12(4): 638-653. <https://doi.org/10.1080/23080477.2024.2363031>
- [14] Dewangan, S.K., Choubey, S., Patra, J., Choubey, A. (2024). IMU-CNN: Implementing remote sensing image restoration framework based on Mask-Upgraded Cascade R-CNN and deep autoencoder. *Multimedia Tools and Applications*, 83(27): 69049-69081. <https://doi.org/10.1007/s11042-024-18122-1>
- [15] Lee, J., Kim, J., Lee, H. (2025). Diffusedesigner: Sketch-based controllable clothing image generation. *Fashion and Textiles*, 12(1): 1-25. <https://doi.org/10.1186/s40691-025-00426-x>
- [16] Dutta, T., Singh, A., Biswas, S. (2020). StyleGuide: zero-shot sketch-based image retrieval using style-guided image generation. *IEEE Transactions on Multimedia*, 23(15209210), 2833-2842. <https://doi.org/10.1109/tmm.2020.3017918>
- [17] Ho, T., Virtusio, J.J., Chen, Y., Hsu, C., Hua, K. (2020). Sketch-guided deep portrait generation. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 16(3): 1-18. <https://doi.org/10.1145/3396237>
- [18] Pawar, P.Y., Ainapure, B.S. (2023). Image dataset of Pune city historical places for degradation detection, classification, and restoration. *Data in Brief*, 51: 109794. <https://doi.org/10.1016/j.dib.2023.109794>
- [19] Freidin, A.Y. (2007). Three-dimensional laser scanning and its application for imaging architectural structures and the restoration of monuments. *Journal of Optical Technology*, 74(8): 545-549.
- [20] Kim, T., Shin, C., Lee, S., Lee, S. (2021). Block-attentive subpixel prediction networks for computationally efficient image restoration. *IEEE Access*, 9: 90881-90895. <https://doi.org/10.1109/access.2021.3091975>
- [21] Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L. (2018). Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 1205-1218. <https://doi.org/10.1109/tgrs.2018.2865197>
- [22] Ma, Z., Li, J., Wang, N., Gao, X. (2020). Image style transfer with collection representation space and semantic-guided reconstruction. *Neural Networks*, 129: 123-137. <https://doi.org/10.1016/j.neunet.2020.05.028>