

Article

Chinese Ancient Paintings Inpainting Based on Edge Guidance and Multi-Scale Residual Blocks

Zengguo Sun ^{1,2,*} , Yanyan Lei ² and Xiaojun Wu ^{1,2}

¹ Key Laboratory of Intelligent Computing and Service Technology for Folk Song, Ministry of Culture and Tourism, Xi'an 710119, China; xjwu@snnu.edu.cn

² School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; leiyy@snnu.edu.cn

* Correspondence: sunzg@snnu.edu.cn

Abstract: Chinese paintings have great cultural and artistic significance and are known for their delicate lines and rich textures. Unfortunately, many ancient paintings have been damaged due to historical and natural factors. The deep learning methods that are successful in restoring natural images cannot be applied to the inpainting of ancient paintings. Thus, we propose a model named Edge-MSGAN for inpainting Chinese ancient paintings based on edge guidance and multi-scale residual blocks. The Edge-MSGAN utilizes edge images to direct the completion network in order to generate entire ancient paintings. It then applies the multi-branch color correction network to adjust the colors. Furthermore, the model uses multi-scale channel attention residual blocks to learn the semantic features of ancient paintings at various levels. At the same time, by using polarized self-attention, the model can improve its concentration on significant structures, edges, and details, which leads to paintings that possess clear lines and intricate details. Finally, we have created a dataset for ancient paintings inpainting, and have conducted experiments in order to evaluate the model's performance. After comparing the proposed model with state-of-the-art models from qualitative and quantitative aspects, it was found that our model is better at inpainting the texture, edge, and color of ancient paintings. Therefore, our model achieved maximum PSNR and SSIM values of 34.7127 and 0.9280 respectively, and minimum MSE and LPIPS values of 0.0006 and 0.0495, respectively.

Keywords: ancient paintings inpainting; edge image; polarized self-attention mechanism; multi-scale residual blocks



Citation: Sun, Z.; Lei, Y.; Wu, X.

Chinese Ancient Paintings Inpainting Based on Edge Guidance and Multi-Scale Residual Blocks.

Electronics **2024**, *13*, 1212. <https://doi.org/10.3390/electronics13071212>

Academic Editor: Xenophon Zabulis

Received: 28 February 2024

Revised: 21 March 2024

Accepted: 21 March 2024

Published: 26 March 2024



Copyright: © 2024 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

Chinese ancient paintings have high cultural and artistic value. However, with the passage of time and interference from external factors, many ancient paintings have suffered damage and color fading. It is crucial to protect the damaged paintings and restore their original state. Traditional manual inpainting methods are limited by manpower and technology, which makes it difficult to cope with serious damage. Additionally, physical and chemical methods may cause further damage to the ancient paintings [1,2]. In recent years, digital inpainting technology has emerged as a viable solution. By using computer technology and image processing algorithms, digital inpainting can accurately restore ancient paintings, reduce inpainting time, and minimize the risk of further damage [3].

In recent years, deep convolutional neural networks have made significant progress in natural image inpainting [4–11]. Pathak et al. [4] first used a generative adversarial network (GAN) with a contextual encoder to restore images from learned image features. Then, Liu et al. [5] and Yu et al. [6] proposed partial convolution and gated convolution respectively. These convolutions use pixel diffusion to achieve progressive inpainting, which solves the problem of the limited receptive field range of ordinary convolution. Yu et al. [7] designed a multi-stage image inpainting network to overcome the limitations of single-stage networks in terms of inpainting capability. In addition, the inpainting process of the network is divided into two stages: coarse inpainting and fine inpainting, which

simplifies the training process at each stage and thus improves the inpainting efficiency. Based on this idea, a series of multi-stage inpainting networks were proposed in [8–10] to restore images progressively from small to large scales using a pyramid structure. These networks are composed of multiple sets of GAN at different scales. Li et al. [11] designed a recurrent feature reasoning network, where a feature reasoning module can iteratively derive the mask boundary feature map and use it as a clue for further inference. In addition, the author also designed knowledge consistency attention to adaptively fuse attention scores and gradually refine the feature map. Although current deep learning-based methods can restore natural images well, there are still challenges in terms of inpainting ancient paintings with varied features and dense textures. These challenges include insufficient feature extraction and loss of detail. Cao et al. [12] proposed a method that uses a fully convolutional network (FCN) to extract deep image features. They also used global and local discriminators to assess the authenticity of the restored mural images. Wan [13] designed a two-stage model for restoring cultural relic images. In the first stage, a coarse restoration model is used, which constructs an encoder and decoder using gated convolution. A fine restoration model is applied in the second stage, which uses a semantic attention mechanism to guide the restoration process based on semantic information. Zhang et al. [14] proposed a painting inpainting method by embedding multiple attention dilation convolutions to obtain dense multi-scale context information. Zhang [15] designed an ancient painting model that uses an edge inpainting network to guide the generation of missing texture areas. Although the above methods can restore information from ancient paintings, blurring and lack of semantics may arise when filling in large, damaged areas. Moreover, these art and painting inpainting methods struggle with precise color control, which can result in color bias in the output. Ultimately, the goal is to restore ancient paintings to their original state as accurately as possible. However, relying solely on damaged paintings as a reference for the network may result in restored paintings that look visually acceptable but do not truly reflect the original appearance.

To address these issues, this paper proposes a Chinese ancient painting inpainting network based on edge guidance and multi-scale residual blocks (Edge-MSGAN). The following are the specific details of the network:

- An ancient painting completion network is designed to deal with ancient paintings' complex structure and rich texture. The network uses the edge image of the original painting as a reference to ensure that the inpainting content matches the original edges.
- Multi-scale residual block (MSRB) is designed to extract detailed features by expanding the perceptual range, which provides a more effective method by which to preserve the rich textures and features of ancient paintings.
- To effectively adjust the color of the inpainting area and reduce color deviation, a multi-branch color correction network is constructed, which makes full use of feature information at different levels, resulting in rich colors similar to those of original ancient paintings.

The other parts of this paper are organized as follows: Section 2 introduces related works in areas of the traditional Chinese paintings inpainting methods and the deep learning-based Chinese paintings inpainting methods. Next, Section 3 gives a detailed introduction to the proposed model. Then, Section 4 presents the inpainting results of Chinese paintings and Dunhuang murals. Finally, Section 5 concludes our work.

2. Related Work

Digital image inpainting relies on the correlation between the known and damaged areas, using known pixels to restore missing pixels. It can be divided into two categories: traditional methods and deep learning-based methods.

2.1. Traditional Chinese Paintings Inpainting Methods

Traditional methods can be further divided into diffusion-based methods and exemplar-based methods. Diffusion-based methods utilize partial differential equations (PDE) to

diffuse pixels into the missing areas gradually. Zhao [16] proposed a method for restoring the color of faded areas using PDE. Xu et al. [17] proposed a Bertalmio-Sapiro-Caselles-Bellester (BSCB) model based on the autologous theory segmentation system, in order to address the problems of blurring and isoline crossing. However, these diffusion-based methods can suffer from error propagation, causing over-smoothing and making them unsuitable for larger areas. Exemplar-based methods [18] match the similarity between the missing area and the known area and then copy the most similar block to the corresponding missing part. Zhou et al. [19] proposed an improved Criminisi algorithm by inserting a dispersion term for restoring the ancient painting named Yan Yun Shuang Lu. Ma et al. [20] used an interactive inpainting method based on the decomposition of drawing curves. This method parses a painting into contents and canvases, and then merges the inpainting results of contents and canvases. Wang et al. [21] proposed a sparse model that selects candidate patches based on texture similarity and structural continuity, while using the corresponding line drawing of the mural image to add line structures in the missing areas. Bhele et al. [22] proposed a texture-structure conserving patch matching algorithm (TSCPMA). The algorithm improves the capabilities of the Criminisi algorithm in repairing large, damaged areas and small gaps by redefining the minimum similarity distance criterion to select the best-matching patches. Although the above, exemplar-based, methods solve the over-smoothing problem of diffusion-based methods, they have low retrieval speed and matching efficiency. In addition to the above two methods, Yan [23] used a multispectral method for ancient painting inpainting, which helped avoid the metamerism phenomenon. Zhou et al. [24] used classified linear regression of hyperspectral images to remove stains. Hou et al. [25] proposed a virtual inpainting method, the method uses the maximum noise fraction transformation on the hyperspectral imaging to reduce the impact of stains on the ancient paintings. Zhang et al. [26] improved the image processing algorithms in OpenCV, making them more suitable for the processing of ancient paintings.

2.2. Deep Learning-Based Chinese Paintings Inpainting Methods

Deep learning-based methods, such as convolutional neural networks (CNN) and generative adversarial networks (GAN), have greatly progressed in recent years. These models can learn the features and semantic information of images, enabling them to produce natural and realistic results even in complex scenes. Ancient painting inpainting is a difficult task, but several approaches have been proposed by which to address this challenge. Zhao et al. [27] designed a multi-channel encoder to learn the semantic features at different scales, and then used the learned macro-, meso-, and micro-level semantic features for ancient painting inpainting. Xue [28] decomposed the ancient painting inpainting task into content and line, allowing the inpainting network to focus on the details of ancient paintings. Yang [29] proposed an ancient paintings inpainting network based on GAN, by introducing gated convolution and attention mechanisms, and mixing the coding of image texture and structure. Zhao et al. [30] proposed a progressive multi-level feature inpainting model, which accomplished the inpainting from high-level features to low-level semantic features. Liu et al. [31] used a global attention mechanism to carry out coarse-grained inpainting on the structure and color of Chinese ancient paintings. They also used local attention mechanisms and residual blocks for fine-grained inpainting, focusing on the small-scale structures and intricate textures of Chinese ancient paintings. Zhou et al. [32] used color information extracted from deep features to improve the inpainting quality of color in missing regions. Additionally, they proposed a multi-step feature refinement model to effectively transfer feature information from undamaged areas to damaged areas. Lv et al. [33] believe that introducing prior information can improve the quality of the inpainting results. To achieve this, they designed a GAN network model with two generators. The model first uses an edge inpainting network to repair the contour edges and then uses the edges to guide the content completion network to complete the image inpainting. Peng et al. [34] proposed a model that uses dual-domain partial convolution to process valid pixels and combines frequency conversion to promote effective fusion of

multi-scale features. Deng et al. [35] believe that most existing mural inpainting models neglect the importance of structural guidance, making it impossible to fill in complex and diverse damaged content with structures. Thus, the authors proposed a structure-guided model based on GAN for the inpainting of ancient murals.

3. Method

An inpainting model for Chinese ancient paintings is proposed based on edge-guided and multi-scale residual blocks, namely Edge-MSGAN, for ancient paintings' rich line structure characteristics.

3.1. Overall Structure of Edge-MSGAN

The Edge-MSGAN mainly consists of three parts, as shown in Figure 1, which are the ancient painting completion network, the multi-branch color correction network, and the Markov discriminator. The model first uses the Canny edge detection algorithm [36] to detect the outline of the original ancient painting, and then uses the obtained edge image to guide the ancient painting completion network to reconstruct the missing parts. Finally, the inpainting result is input into a multi-branch color correction network for local color adjustment.

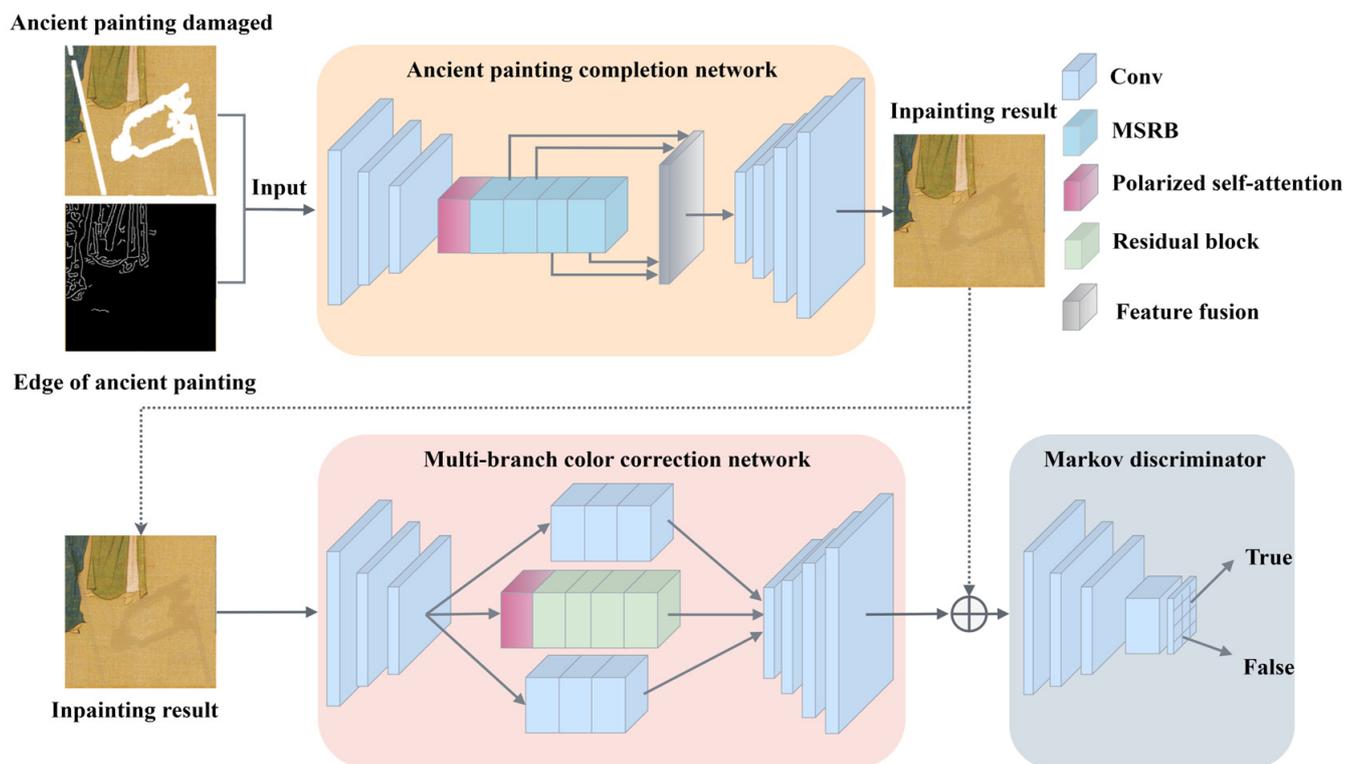


Figure 1. Overall structure diagram of Edge-MSGAN.

The U-Net [37] with skip connections [38] is employed in the ancient painting completion network and multi-branch color correction network to achieve multi-scale feature fusion. The ancient painting completion network also uses a multi-scale residual block (MSRB) between the encoder and decoder. MSRB combines dilated convolution [39] and channel attention mechanisms in order to expand the perceptual field and extract vital features. Additionally, the polarized self-attention (PSA) mechanism can focus on key areas of ancient paintings, which preserves the essential details during encoding. As a result, the network generates high-quality ancient paintings with consistent content and complete semantics during decoding.

3.2. Polarized Self-Attention

The attention mechanism is a crucial technique that finds extensive use in deep learning-based methods. It allocates distinct weights to various features, allowing the model to automatically identify and concentrate on the most relevant information. This results in an improved ability of the model to process essential information. There are three main types of attention mechanisms: the spatial attention mechanism [40], channel attention mechanism [41,42], and channel–spatial hybrid attention mechanism [43–45]. The spatial attention mechanism evaluates the importance of spatial location information. It enhances useful features and reduces useless ones. The channel attention mechanism models the relationship between channels. It ensures that important channels receive more attention. The channel–spatial hybrid attention mechanism combines the spatial and channel attention mechanisms. It can be used in parallel or series to capture critical features.

For large, damaged ancient paintings inpainting, the model must be able to capture the relationship between distant pixels. To address this issue, the Edge-MSGAN model incorporates the polarized self-attention (PSA) mechanism [46]. PSA operates on two branches, separately processing self-attention in the channel and spatial dimensions, and finally fusing the results. This approach enables the model to better capture the feature correlation of distant pixels, making it highly effective when dealing with large-scale damage. The structure of PSA is shown in Figure 2.

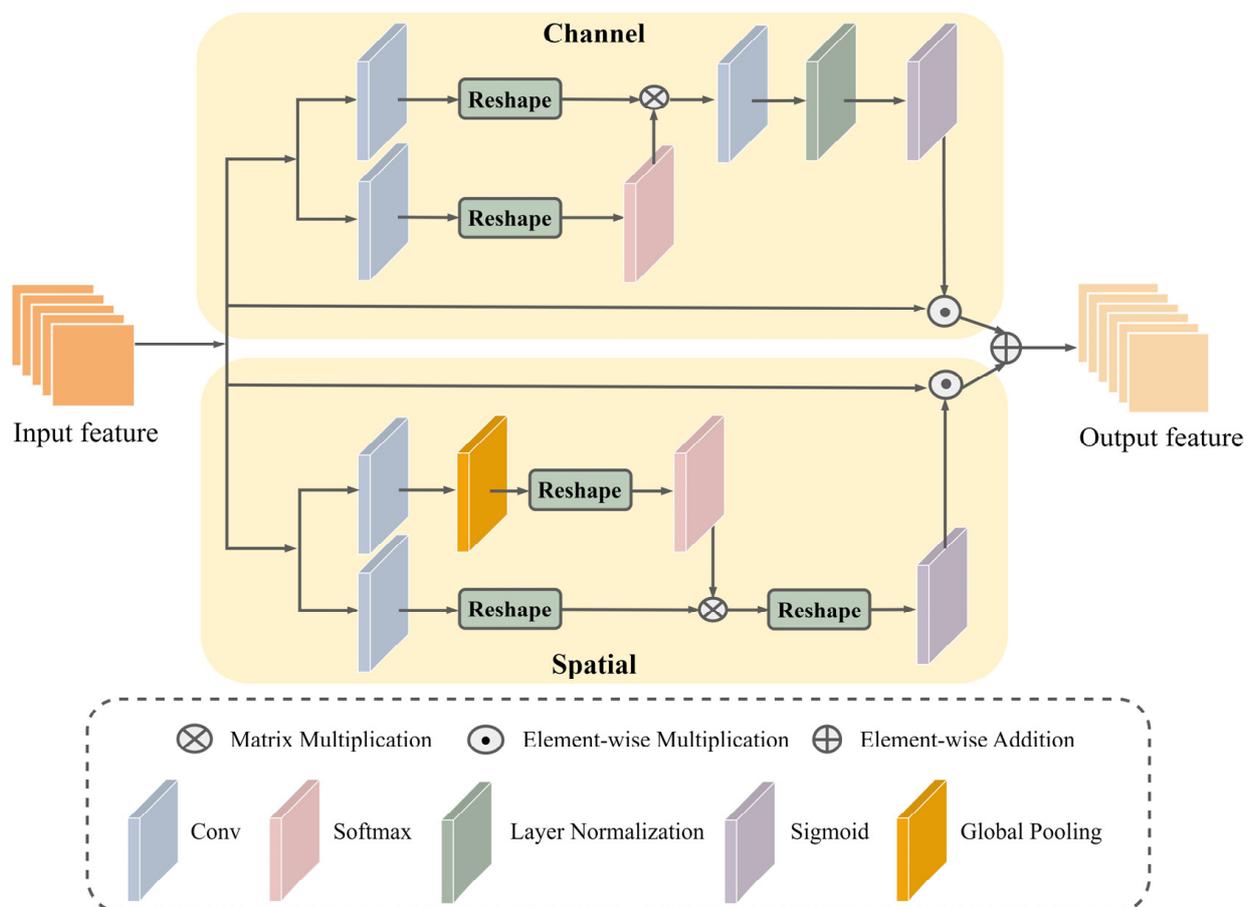


Figure 2. Structure of the polarized self-attention mechanism.

3.3. Multi-Scale Residual Blocks

Capturing the color and detailed texture information of ancient paintings is crucial for inpainting models. We propose a multi-scale residual block (MSRB) to achieve this. The MSRB consists of a 3×3 convolutional branch and a 5×5 convolutional branch, as shown in Figure 3. This enables the model to integrate features of different scales, capturing

multi-scale information in ancient paintings. In addition, the ReLU activation function is introduced after each convolution to enhance the non-linear capability of the residual block. MSRB can reuse low-level features from the input and retain the original style of the painting through residual connections. This also helps to alleviate problems of gradient disappearance and explosion, leading to improved stability and convergence of the model. The MSRB is formulated by (1):

$$M_n = F_{CA} \{ W_{1 \times 1}^l \times [Q_1, Q_2] + b^l \} + M_{n-1} \tag{1}$$

where F_{CA} is the attention mechanism operation, $W_{1 \times 1}^l$ and b^l are the weight tensor and offset tensor of the l -th layer, $[Q_1, Q_2]$ is the feature fusion of the two convolutional layers, and M_{n-1} is the output of the previous residual block.

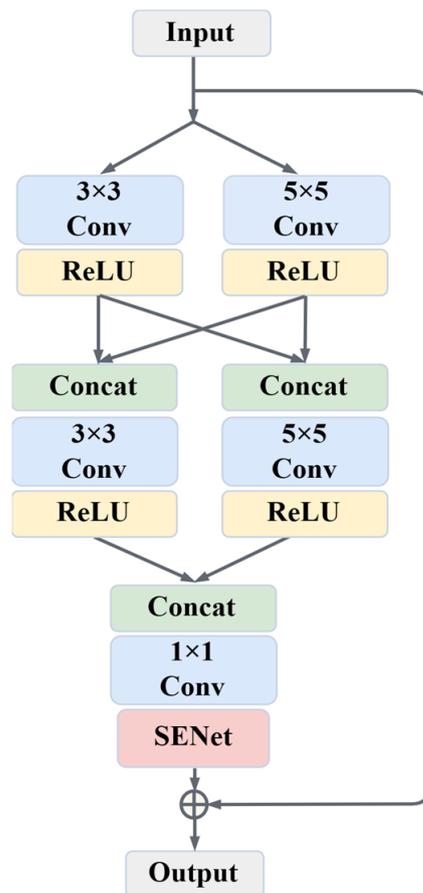


Figure 3. Multi-scale channel attention residual block.

The MSRB introduces the attention mechanism network (SENet) [47] after feature fusion. SENet consists of two parts, squeeze and excitation. The core idea is to adaptively weight the channels of the feature map after convolution, obtaining the weights of each channel, so that the network pays more attention to the most useful channel features.

In the squeeze operation, a feature map $C \times H \times W$ is compressed into $C \times 1 \times 1$ to represent global information. The squeeze is formulated by (2):

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j) \tag{2}$$

where $z \in \mathbb{R}^C$ is the channel descriptor, z_c is the c -th element of z , F_{sq} is the squeeze operation, and $u_c(i, j)$ is the feature map at (i, j) .

Excitation first performs full connection on the result of squeeze to obtain a vector of the C/r dimension, and, after ReLU activation, the vector of the C/r dimension is changed back to a vector of the C dimension. The weight is finally obtained by sigmoid activation. The excitation is formulated by (3):

$$S = \text{Sigmoid}(W_2 \cdot \text{ReLU}(W_1 z)) \quad (3)$$

where S is a statistic, $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ is the up-sampling weight for dimension increment, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ is the down-sampling weight for dimension decrement, the reduction ratio r is a hyperparameter, $z \in \mathbb{R}^C$ is the result obtained by the squeeze, and $W_1 z$ is a fully connected operation.

3.4. Multi-Branch Color Correction Network

The multi-branch color correction network is an improved version of the ancient painting completion network. It owns two additional branches, as shown in Figure 4. These branches use three layers of 5×5 and 7×7 kernel sizes to gather a wider range of contextual information. With the multi-branch color correction network, it is possible to extract features from various levels and scales, and then merge them to more precisely adjust the color of the inpainting area.

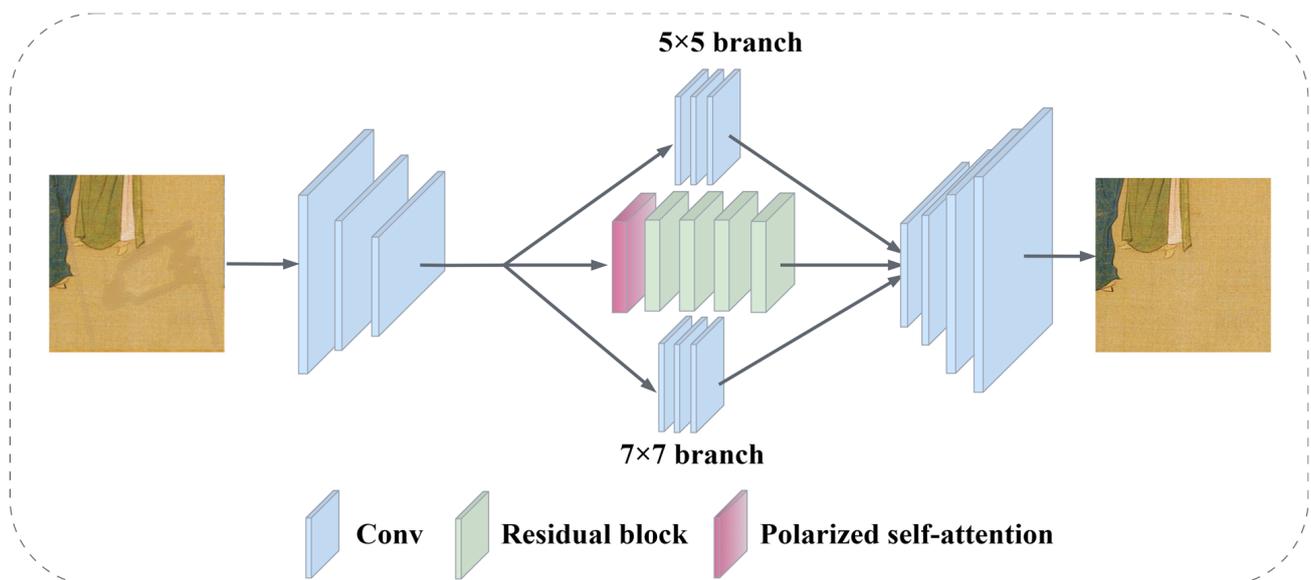


Figure 4. Structure diagram of multi-branch color correction network.

3.5. Markov Discriminator

For inpainting models of ancient paintings, an ordinary discriminator that outputs a vector for the whole image cannot capture all of the intricate details. To address this issue, the Markov discriminator [48] was introduced. It divides the input image into multiple blocks of 70×70 and performs binary classification discrimination on each of these blocks. This enables the discriminator to evaluate the localization and details more accurately than an ordinary discriminator. As a result, the feedback information provided by the Markov discriminator is more informative for the ancient painting completion and multi-branch color correction networks. This improves the generation of more realistic and intricate ancient paintings. The Markov discriminator consists of five convolutional layers, as shown in Table 1.

Table 1. Parameters of convolutional layers in the Markovian discriminator.

Layer	Kernel	Stride	Output (H × W)	Output Channels	Activation Function
Conv-1	4 × 4	2	128 × 128	64	Leaky-ReLU
Conv-2	4 × 4	2	64 × 64	128	Leaky-ReLU
Conv-3	4 × 4	2	32 × 32	256	Leaky-ReLU
Conv-4	4 × 4	1	31 × 31	512	Leaky-ReLU
Conv-5	4 × 4	1	30 × 30	1	Sigmoid

3.6. Loss Function

Given the unique color, texture and stylistic characteristics of ancient paintings, this paper combines the following loss functions to constitute the overall loss function of the model. The total loss function is formulated by (4):

$$L_{Total} = L_{Perceptual} + L_1 + L_{Histogram} + L_{TV} \quad (4)$$

where $L_{Perceptual}$ is the perceptual loss [49], which aims to minimize the feature space distance between the inpainting result and the original ancient painting, the L_1 loss [50] is used to measure the absolute distance between the inpainting result and the original ancient painting, $L_{Histogram}$ is the histogram loss [51], which calculates the loss by comparing the color histogram of the inpainting result and the original ancient painting, and L_{TV} is the TV loss [52], which is used to measure the difference between adjacent pixels and thus constrain pixel changes.

3.6.1. Perceptual Loss

We use a pre-trained VGG19 network to extract the features of the ancient paintings. By comparing the differences between the inpainting result and the original ancient painting in the feature space, the style and structural features of the ancient painting are preserved. The perceptual loss is formulated by (5):

$$L_{Perceptual} = \sum_j \|\phi_j(y) - \phi_j(\bar{y})\|_2^2 \quad (5)$$

where y is the original ancient painting, \bar{y} is the inpainting result, ϕ is the feature extraction layer of VGG19, and ϕ_j is the output features of the j -th layer.

3.6.2. L_1 Loss

L_1 loss, also known as mean absolute error (MAE), aims to calculate the average of the absolute errors between the inpainting results and the original ancient paintings. This helps to preserve the details, textures, and overall characteristics of the ancient painting. The L_1 loss is formulated by (6):

$$L_1 = \frac{\sum_{i=1}^n |f(x_i) - y_i|}{n} \quad (6)$$

where $f(x_i)$ and y_i are the inpainting results and the corresponding ancient paintings of the i -th respectively, and n is the number of images.

3.6.3. Histogram Loss

Histogram loss is used to measure the difference in color distribution. The proposed model adopts the histogram loss to adjust the color distribution of the inpainting results. The histogram loss is formulated by (7):

$$L_{Histogram} = \sum_{l=1}^L w_l \|O_l - R(O_l)\|_F^2 \quad (7)$$

where w_l is the weight of the l -th layer, O_l is the activation function layer, R is histogram matching, $R(O_l)$ is the activation function layer after histogram matching, and $\|\cdot\|_F$ is the Frobenius norm.

3.6.4. Total Variation Loss

Total variation loss measures the smoothness by calculating the pixel gradient value. Introducing the TV loss can avoid excessive noise and flaws. The TV loss is formulated by (8):

$$L_{TV} = \sum_{i,j} \sqrt{(x_{i,j-1} - x_{i,j})^2 + (x_{i+1,j} - x_{i,j})^2} \quad (8)$$

where i and j are the pixel coordinates, and $x_{i,j}$ is a pixel point in the input image.

4. Experiments

This section will introduce the dataset, training process, comparison experiments, and ablation study.

4.1. Construction of Dataset

A dataset containing 12,509 images of Chinese ancient paintings has been established for the purpose of the inpainting of ancient paintings. These images are obtained by cropping four famous paintings from different times and styles. One of these paintings, *Along the River During the Qingming Festival*, is a copy of an original work, painted by Zhang Zeduan during the Song Dynasty, which was painted by the Ming Dynasty painter Qiu Ying. The construction process of the dataset is detailed in Figure 5. Initially, original ancient paintings were cropped with a size of 256×256 using Photoshop. Blank samples were excluded to ensure the diversity and quality of the dataset. Subsequently, the dataset was divided into a training set, validation set, and test set, with quantities of 10,500, 1373, and 624 respectively, as shown in Table 2. To accommodate different damage rates, a mask dataset publicly available in [5] was adopted, which provides mask images with damage rates ranging from 1% to 60%.

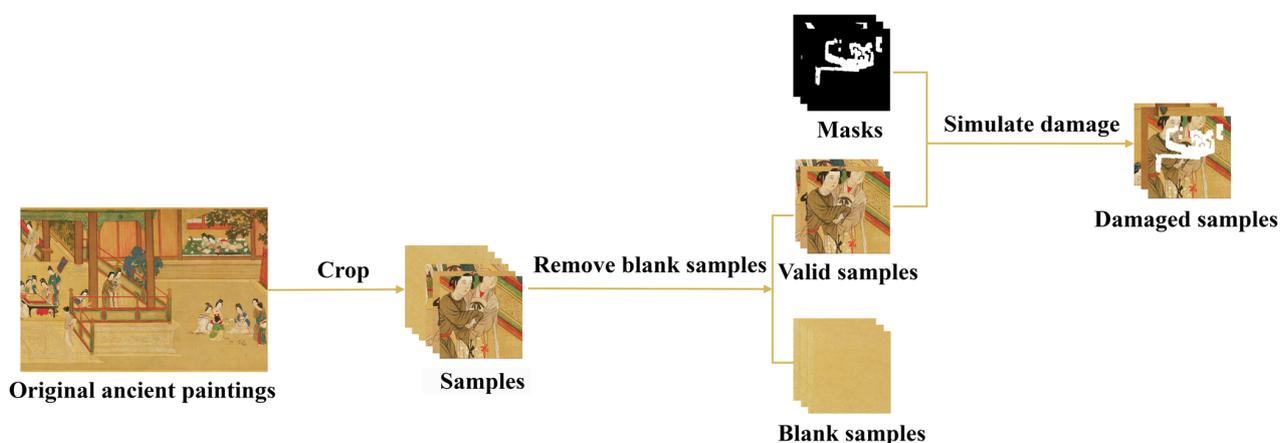


Figure 5. Dataset construction process.

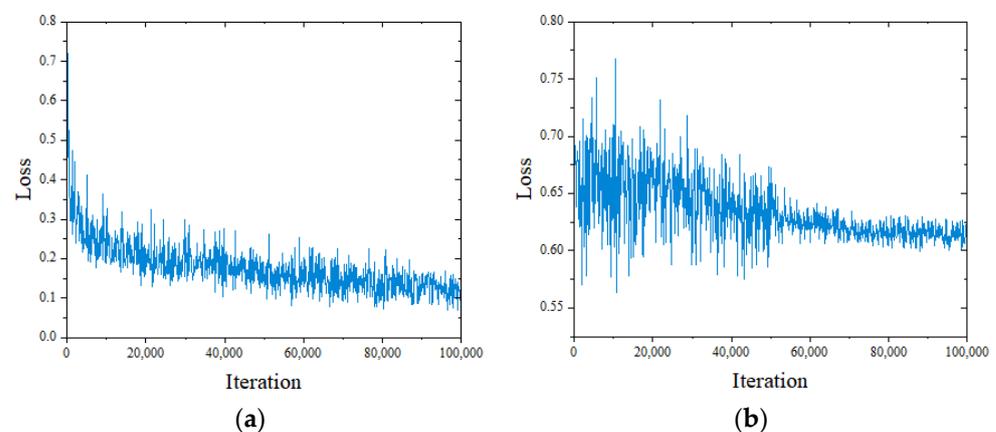
Table 2. Detailed information about the ancient painting inpainting dataset.

Chinese Ancient Paintings	Dynasty	Painter	Original Size	Training Set	Validation Set	Test Set	Total
<i>Along the River During the Qingming Festival</i>	Ming	Qiu Ying	159,010 × 4339	978	128	96	1202
<i>Spring Morning in the Han Palace</i>	Ming	Qiu Ying	102,680 × 5299	641	157	96	894
<i>Dwelling in the Fuchun Mountains</i>	Yuan	Huang Gongwang	196,301 × 5197	4722	464	202	5388
<i>A Thousand Li of Rivers and Mountains</i>	Song	Wang Ximeng	152,089 × 6083	4159	624	230	5025

4.2. Training Process

The Edge-MSGAN is trained by using the ancient painting inpainting dataset that we constructed, with a batch size of 8. The learning rate of the discriminator and generator are set to 1.0×10^{-4} and 1.0×10^{-5} , respectively. The weights of the loss function for the L_1 loss, perceptual loss, histogram loss, TV loss, and adversarial loss are 1, 0.1, 0.0005, 0.01, and 0.1, respectively. The Adam optimizer is used to optimize the parameters of the generator and discriminator.

The curve of the loss function value during the Edge-MSGAN training process is displayed in Figure 6. At the beginning, when the number of iterations is small, the generator produces poor quality ancient paintings that are easily distinguishable by the discriminator. This leads to a rapid decline in generator loss (G_loss) and large fluctuations in discriminator loss (D_loss). However, as the number of iterations increases, the generator gradually improves and produces better quality ancient paintings, resulting in a steady decline in G_loss and eventual stabilization. The fluctuation amplitude of D_loss also decreases with the training of the generator, making the difference between the generated ancient paintings and original ancient paintings gradually smaller. A plot of G_loss over iteration is shown in Figure 6a. Initially, the value of G_loss is approximately 0.73. As the iterations progress to 50,000, it is greatly decreased to almost 0.2. At 100,000 iterations, G_loss further decreases to around 0.1. Similarly, the amplitude of D_loss in Figure 6b decreases with each iteration and reaches around 0.615 by 100,000 iterations. Based on the above analysis, at the end of training, both G_loss and D_loss stabilize, with values of around 0.1 and 0.615, respectively, indicating that the generator can generate high-quality ancient paintings and that the discriminator can correctly distinguish between original ancient paintings and generated paintings. This proves that the proposed Edge-MSGAN model can achieve a stable state during training and complete the task of ancient painting inpainting.

**Figure 6.** Plot of the loss function value for Edge-MSGAN. (a) Plot of G_loss and (b) plot of D_loss.

4.3. Experimental Results

The inpainting results based on the proposed dataset are shown in Figure 7. The first row depicts the inpainting results of *A Thousand Li of Rivers and Mountains*, while the second and third rows showcase the inpainting results of *Dwelling in the Fuchun Mountains*. The fourth and fifth rows demonstrate the inpainting results of *Spring Morning in the Han Palace*. Edge-MSGAN is able to successfully restore the elements of mountains, rivers, trees, and buildings in landscape paintings, as shown in the first three rows of Figure 7. The mountains and rivers retain their original lines and flow, while the trees regain their natural forms and branches. In figure paintings, the last two rows of Figure 7 show the recovery of facial features, clothing textures, and gestures. The details of the face are preserved, and the contours of the eyes, nose, and mouth are clear. In summary, our proposed model owns significant inpainting effects, whether it is applied to landscape paintings, such as *A Thousand Li of Rivers and Mountains* and *Dwelling in the Fuchun Mountains*, or figure paintings, such as *Spring Morning in the Han Palace*.

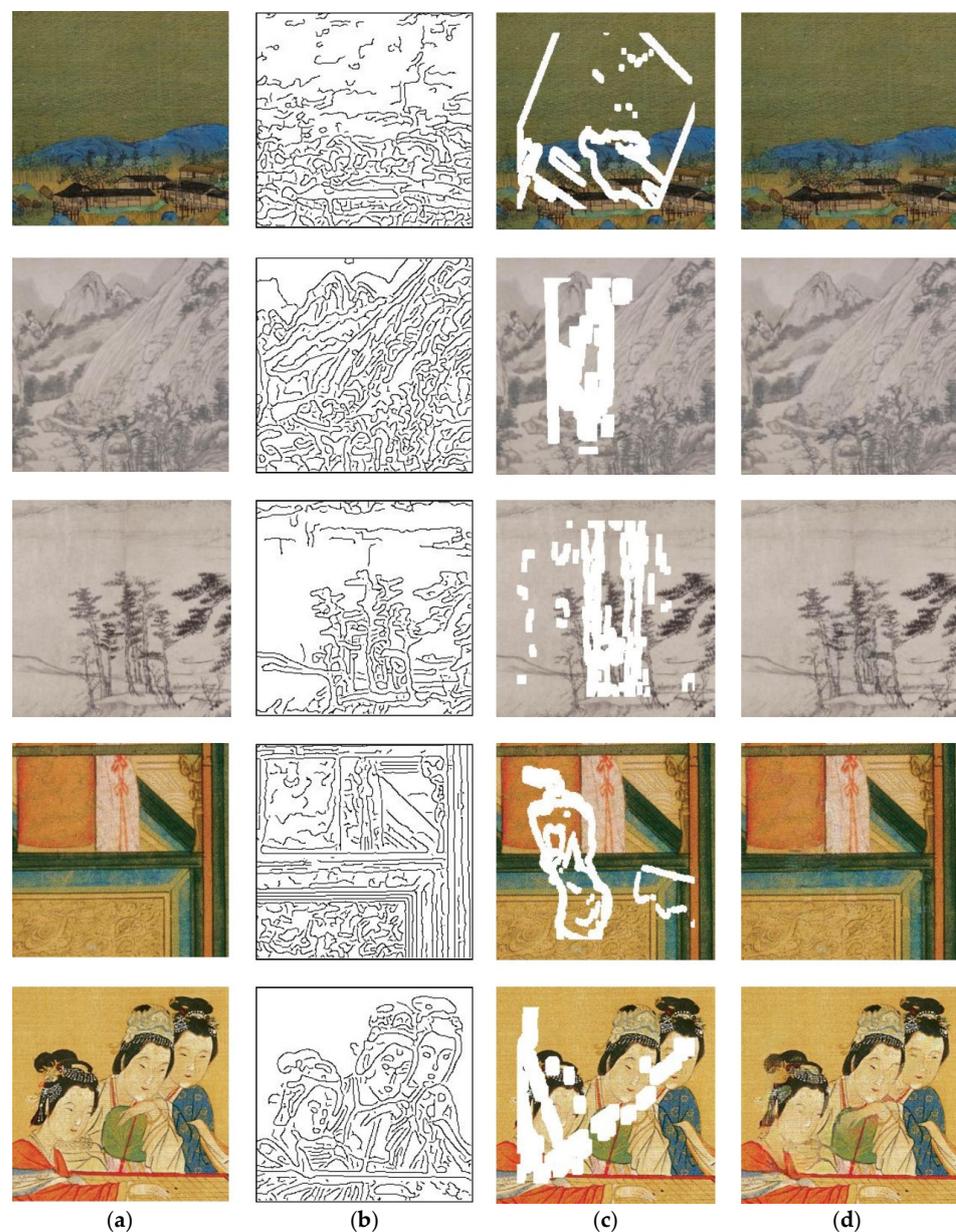


Figure 7. Inpainting results of the proposed model. (a) Original ancient paintings, (b) edge of ancient paintings, (c) damaged ancient paintings, and (d) inpainting results.

4.4. Comparison Experiments

To verify the effectiveness of the proposed method, this paper selects four typical models in recent years for comparison on our dataset. These models are as follows. (1) PI [53]: This model introduces two parallel training paths, namely the reconstruction path and generation path based on a variational autoencoder (VAE), to achieve diverse inpainting results. (2) LG [54]: This model uses local networks with different receptive fields and global networks to perform image inpainting. (3) EC [55]: This model adopts a two-stage adversarial model, where the edge generator is responsible for inpainting the edges of missing areas, and the image completion network uses the edge as prior information to fill in the missing image areas. (4) RFR [11]: The recurrent feature reasoning module can utilize the correlation between pixels, strengthening the constraints of deeper pixels and thus improving the effect of inpainting large broken areas.

4.4.1. Qualitative Comparison

The inpainting results using the comparison models and the proposed model are presented in Figure 8. The first two images in Figure 8c display successful inpainting of most of the content and texture. However, the last three images show poor results, with large areas of missing content. Figure 8d shows improved texture structures and details over Figure 8c. However, in the third ancient painting of Figure 8d, there is noise and structural distortion in the inpainting areas. Figure 8e shows reduced texture distortion and noise compared with Figure 8d, but color deviation and missing facial contours remain. Figure 8f improves the structural distortion compared with Figure 8e. However, some minor issues still exist, such as color deviation, incomplete facial contours, and irregular blocky textures.

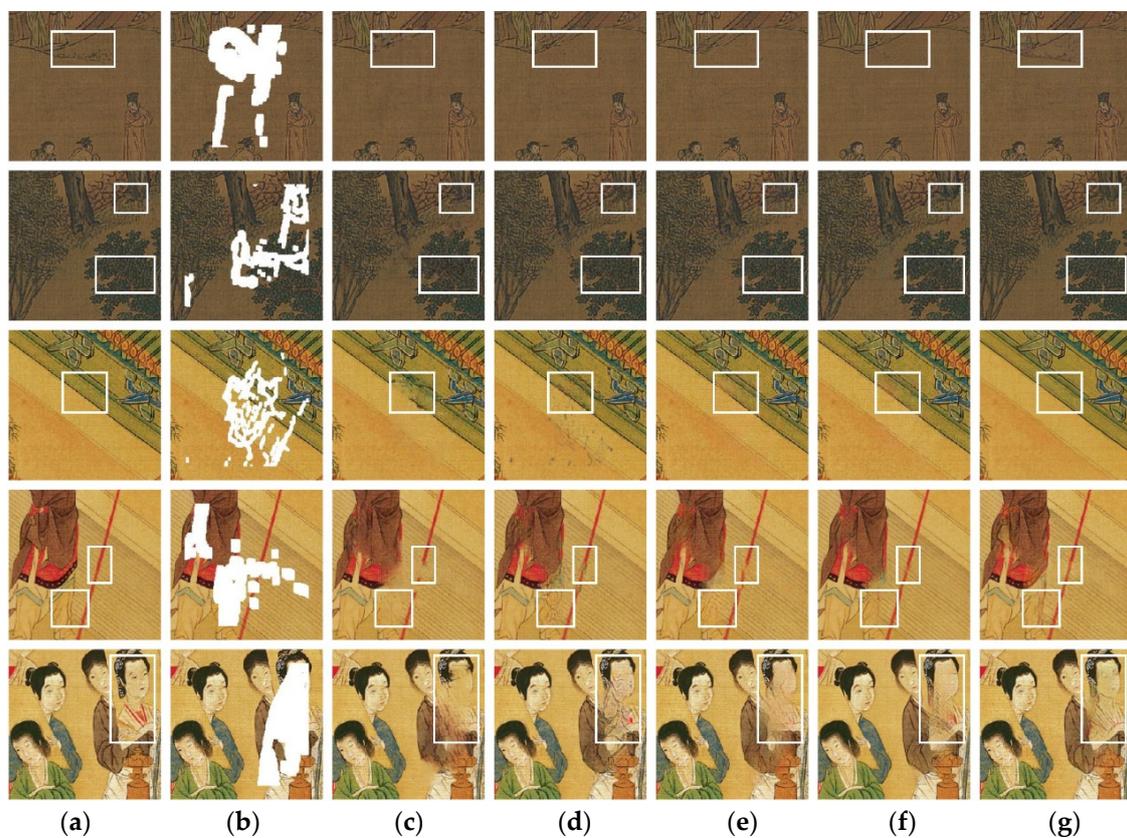


Figure 8. Comparison of inpainting results using different models (White boxes are used for comparison of details). (a) Original ancient paintings, (b) damaged ancient paintings, (c) PI, (d) RFR, (e) EC, (f) LG, and (g) Edge-MSGAN (ours).

In contrast, the proposed Edge-MSGAN model restores texture structure that other models cannot. The second ancient painting in Figure 8g solves the blurring problem that appeared in the previous four models and restores clear details. The third painting shows minimal loss of structure and deviation in color. In the fourth painting, the proposed model performs the inpainting without any distortion in the structure. In the fifth painting, the proposed model successfully restores the eyes, eyebrows, and other crucial areas. This qualitative comparison clearly demonstrates that the proposed model's inpainting effect is outstanding.

4.4.2. Quantitative Comparison

In this experiment, we take mean square error (MSE), structure similarity index measure (SSIM), peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS) as evaluation metrics on our test set.

MSE measures the pixel difference, and the smaller its value, the better the inpainting result. MSE is formulated by (9):

$$\text{MSE} = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - K(i, j)]^2 \quad (9)$$

where M and N are the length and width of the image, respectively, and $I(i, j)$ and $K(i, j)$ are the pixels of the different images at the same position.

SSIM measures the similarity between two images. It is more consistent with human visual perception, and the larger its value is, the more similar the images are. SSIM is formulated by (10):

$$\text{SSIM} = l(x, y) \cdot c(x, y) \cdot s(x, y) \quad (10)$$

where $l(x, y)$, $c(x, y)$, and $s(x, y)$ are the brightness comparison, contrast comparison and structure comparison of the image, respectively.

PSNR is obtained by calculating the mean square error (MSE) between the pixels at the corresponding positions, with a larger value indicating a less distorted image. PSNR is formulated by (11):

$$\text{PSNR} = 10 \log_{10} \frac{2^{bits} - 1}{\text{MSE}} \quad (11)$$

where, $bits$ represents the number of bits occupied by each pixel.

LPIPS evaluates similarity close to human perception. It is more consistent with human perception than PSNR and SSIM. The lower the value, the more similar the inpainting result is to the original ancient painting. LPIPS is formulated by (12):

$$\text{LPIPS} = d(x, x_o) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{hw}^l - \hat{y}_{ohw}^l)\|_2^2 \quad (12)$$

where x is the original ancient painting, x_o is the inpainting result, and d is the distance between x_o and x . LPIPS first extracts the features of the l -th layer, activates the output of each layer and normalizes it, denoted as $\hat{y}_{hw}^l, \hat{y}_{ohw}^l \in \mathbb{R}^{H_l \times W_l \times C_l}$, and then calculates the \downarrow_2 distance after multiplying \odot the weights $w_l \in \mathbb{R}^{C_l}$.

The inpainting results in Figure 8 correspond to the quantitative metrics presented in Table 3. Based on the analysis of the metrics used to evaluate the performance of the different models, it is evident that the PI model has the lowest PSNR and SSIM values, while the MSE and LPIPS values are the highest. The reason behind this is that the PI model tends to cause a loss of information, which in turn leads to missing edges. After comparing the PI model with the RFR model, it can be observed that the RFR model has shown an improvement in both the PSNR and SSIM, while the MSE and LPIPS have decreased. The introduction of a progressive mechanism on the feature map level during the inpainting process in the RFR model is responsible for this improvement. However, the downside of this model is that it produces some noise. Compared with the RFR model, the EC model

shows improvement in PSNR and SSIM while that for LPIPS has decreased. This is because the EC model introduces edge structures. However, it is restricted by ordinary residual networks, which results in an incapability to achieve shared and reused feature information and leading to the failure of filling local details. Compared with the other models, the LG model produces better PSNR and SSIM, due to its use of a large receptive field coarse network to repair the overall structure and some texture details, as well as a small receptive field local refinement network to eliminate artifacts. However, the local refinement network has a smaller receptive field, which means that it cannot capture global information when the damaged area is large. This leads to a more severe patchiness effect, resulting in higher LPIPS value.

Table 3. Comparison of evaluation metrics of different models on the test set.

Model	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	LPIPS \downarrow
PI	31.9015	0.8870	0.0010	0.0790
RFR	33.4395	0.9115	0.0007	0.0710
EC	33.6492	0.9145	0.0007	0.0546
LG	33.7284	0.9198	0.0007	0.0562
Edge-MSGAN (Ours)	34.7127	0.9280	0.0006	0.0495

Compared with the comparison model, the proposed model in Figure 8g extends the receptive field by utilizing the multi-scale residual block (MSRB), and also incorporates edge guidance to provide additional constraints, which helps to reconstruct the edge structure of the ancient paintings. Furthermore, the proposed model improves the accuracy of inpainting content by introducing the PSA mechanism, which in turn helps the model to learn the importance of different areas. To utilize features at different levels and scales, multi-branch structures are introduced in the color correction stage. This enables the model to capture more comprehensive feature information. The proposed model achieves the best inpainting results by incorporating edge prior information, MSRB, PSA, and multi-branch color correction networks. It has the highest values of PSNR and SSIM and the lowest values of MSE and LPIPS. In conclusion, both qualitatively and quantitatively, the proposed model outperforms the comparison models.

The PSNR and SSIM values of the proposed model and comparison models are presented in Figure 9, for different mask rates ranging from 10% to 50%. As depicted in Figure 9a, the PSNR of all inpainting models decrease with an increase in mask rate, but the decline rate varies significantly. At lower mask rates, only the LG model has a higher PSNR than the proposed model. However, as the mask rate increases, the proposed model gradually outperforms other models, as its PSNR values are higher than those of the comparison models.

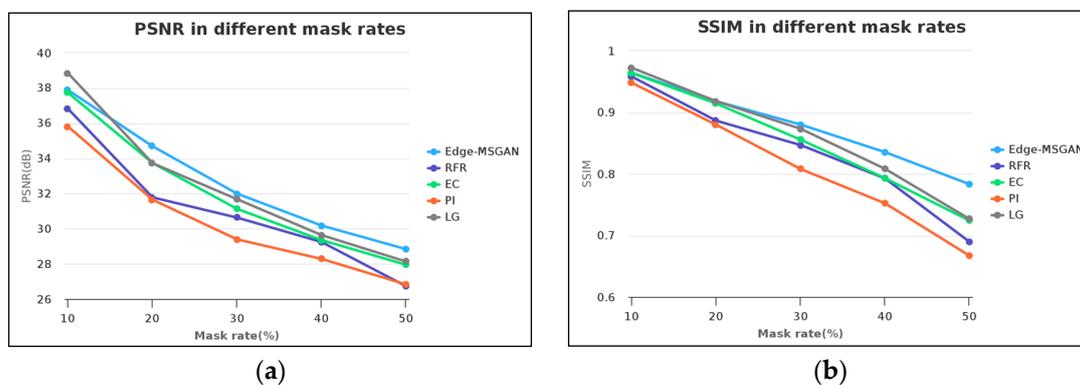


Figure 9. Line charts of PSNR and SSIM in different mask rates. (a) PSNR in different mask rates and (b) SSIM in different mask rates.

As shown in Figure 9b, the SSIM values decrease as the mask rate increases. Although the proposed model is similar to the EC model and slightly inferior to the LG model at low mask rates, its benefits become more evident as the mask rate increases. The SSIM values of the proposed model are higher than other models, particularly when the mask rate is high. This indicates that the proposed model performs better under high mask rates.

In conclusion, the proposed model outperforms other models in terms of PSNR and SSIM values. It exhibits superior performance in ancient paintings inpainting, particularly at higher mask rates.

4.5. Ablation Study

An ablation experiment was conducted on the ancient painting inpainting dataset to verify the effect of TV loss, multi-scale residual blocks (MSRB), additional branch structures added to the color correction network, and polarization self-attention (PSA) on the inpainting ability of the proposed model Edge-MSGAN.

The results of the ablation experiments are presented in Table 4 for quantitative comparison. TV loss plays a crucial role in optimizing the smoothness of the inpainting process, thereby reducing noise and discontinuity in the inpainting area. The removal of TV loss leads to an increase in noise and discontinuity, causing a decrease in the PSNR and SSIM values, and an increase in MSE and LPIPS values.

Table 4. Ablation study.

	Without TV Loss	Without MSRB	Without Branch	Without PSA	Edge-MSGAN (Ours)
PSNR ↑	33.3866	33.1841	33.8397	32.9173	34.7127
SSIM ↑	0.9135	0.9182	0.9182	0.9087	0.9280
MSE ↓	0.0007	0.0007	0.0007	0.0008	0.0006
LPIPS ↓	0.0580	0.0595	0.0553	0.0644	0.0495

The MSRB model uses a multi-scale feature fusion and channel attention mechanism to handle relationships between different scales and channels. This results in improvements in texture inpainting and handling detail loss. When MSRB is not used, there is severe detail loss, which causes a decrease in consistency of texture and color. This is reflected in lower PSNR and SSIM values and higher MSE and LPIPS values.

The multi-branch color correction network that was built can correct the color deviation, while still preserving the edge details. If the additional branch is removed, various issues, like color deviation and loss of detail, may occur in the inpainting results, which leads to lower PSNR and SSIM, and higher MSE and LPIPS.

PSA allows the system to learn the correlation between the inpainting area and the surrounding pixels. By dynamically adjusting the weights between them, PSA focuses more on pixels that are similar to the inpainting area. This results in a more consistent blend between the inpainting area and the surrounding area. In the absence of PSA, the fusion between the inpainting area and the surrounding area may not be as effective, leading to lower PSNR and SSIM, and higher MSE and LPIPS.

In conclusion, after removing the above modules, the PSNR and SSIM values of the inpainting results decreased, while the MSE and LPIPS values increased, indicating their importance in Edge-MSGAN.

The difference between the inpainting results before and after using the corresponding modules is shown in Figure 10. MSRB can capture more details of ancient paintings through multi-scale feature extraction, thereby improving the quality and clarity of the inpainting results. By comparing Figure 10a with Figure 10b, it can be observed that, after adding MSRB, the phenomena of detail loss and blurred inpainting areas are reduced. The proposed model uses TV loss to adjust the difference between adjacent pixels, resulting in the elimination of incoherent textures. A comparison between Figure 10c,d shows

an improvement in the smoothness of the inpainting results after adding TV loss. By comparing Figure 10e,f, it can be observed that, when the color correction network does not add branches, there is a problem of color deviation in the inpainting results, with missing clothing and facial contour lines in the characters. After adding two branches, the color and edge clarity in the inpainting results are improved. Comparing Figure 10g,h, it is evident that PSA prevents issues such as over-smoothing and color inconsistency. Its introduction reduces errors in filling.

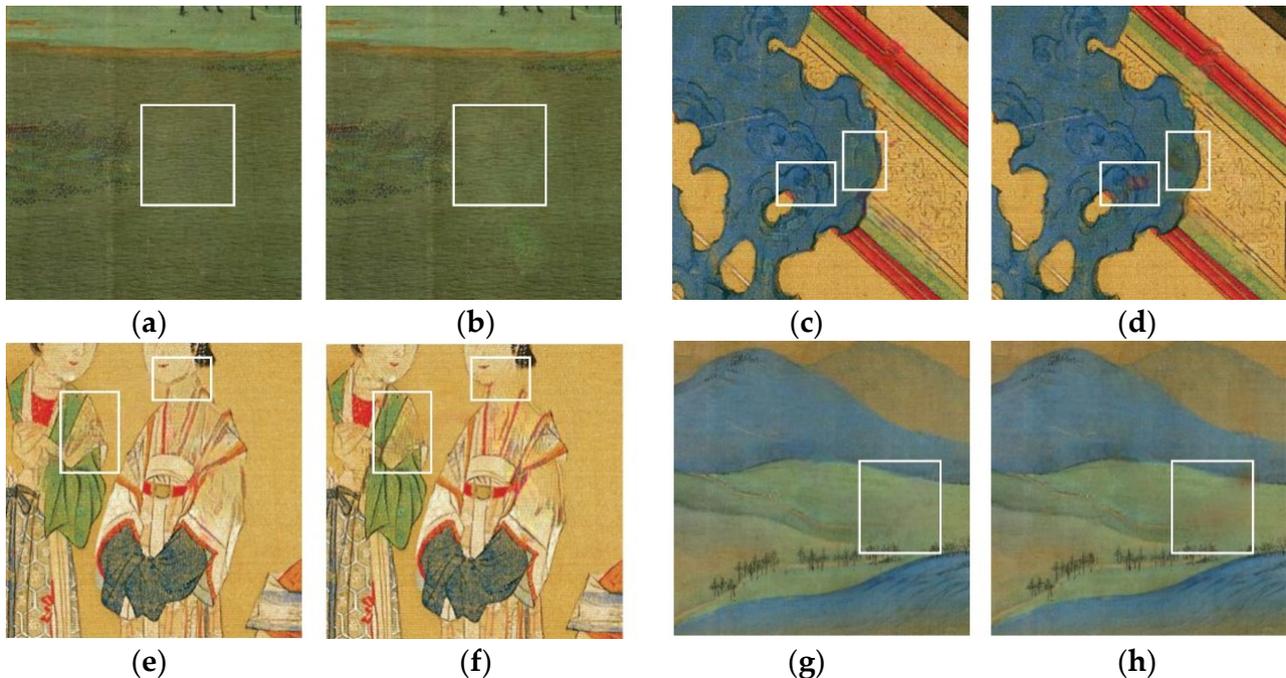


Figure 10. Results comparison before and after using the module (White boxes are used for comparison of details). (a) With MSRB, (b) without MSRB, (c) with TV loss, (d) without TV loss, (e) with additional branch, (f) without additional branch, (g) with PSA, and (h) without PSA.

Based on the experimental results shown in Figure 10 and Table 4, it can be concluded that introducing TV loss, MSRB, branches to the multi-branch color correction network, and polarization self-attention mechanism PSA addressed the issues of detail loss, texture blurring, over-smoothing, and color deviation. These modules play an important role in ancient painting inpainting.

4.6. Experiment on Murals

In order to evaluate the effectiveness of the Edge-MSGAN model in various scenarios, we conducted inpainting experiments on mural dataset [50]. These experiments are then compared with the contrast models discussed in Section 4.4. The mural dataset consists of a total of 1714 murals, including 1564 training images, 50 testing images, and 100 validation images. The murals are obtained from the Mogao Grottoes in Dunhuang, and include both real murals and artist reproductions of Dunhuang murals. The dataset contains 525 real murals that were obtained through scanning, along with 1189 reproduction murals that were obtained through the scanning of art books.

The experimental results and corresponding quantitative metrics for each model are presented in Figure 11 and Table 5. Due to the PI causing information loss, missing edges and distorted contours, the PSNR and SSIM values are the lowest, while the MSE value is high and the LPIPS value is the highest. Compared with the PI model, RFR's results exhibit improvement, as evidenced by the increased values of PSNR and SSIM, and decreased LPIPS. However, the inpainting results of RFR still suffer from edge distortion, leading to a slight increase in MSE. The results of the EC model do not show improvement compared

with RFR. Additionally, the EC model has a severe color deviation problem, which causes a decrease in PSNR and SSIM, and an increase in MSE and LPIPS. The LG model exhibits improvement over the EC model. As a result, PSNR and SSIM increase while MSE decreases. However, facial contour completeness is poorer, resulting in increased LPIPS. Compared with the previous four models, the proposed Edge-MSGAN improves color deviation and accurately restores the details, colors, and textures of murals, and all the indexes are optimal.

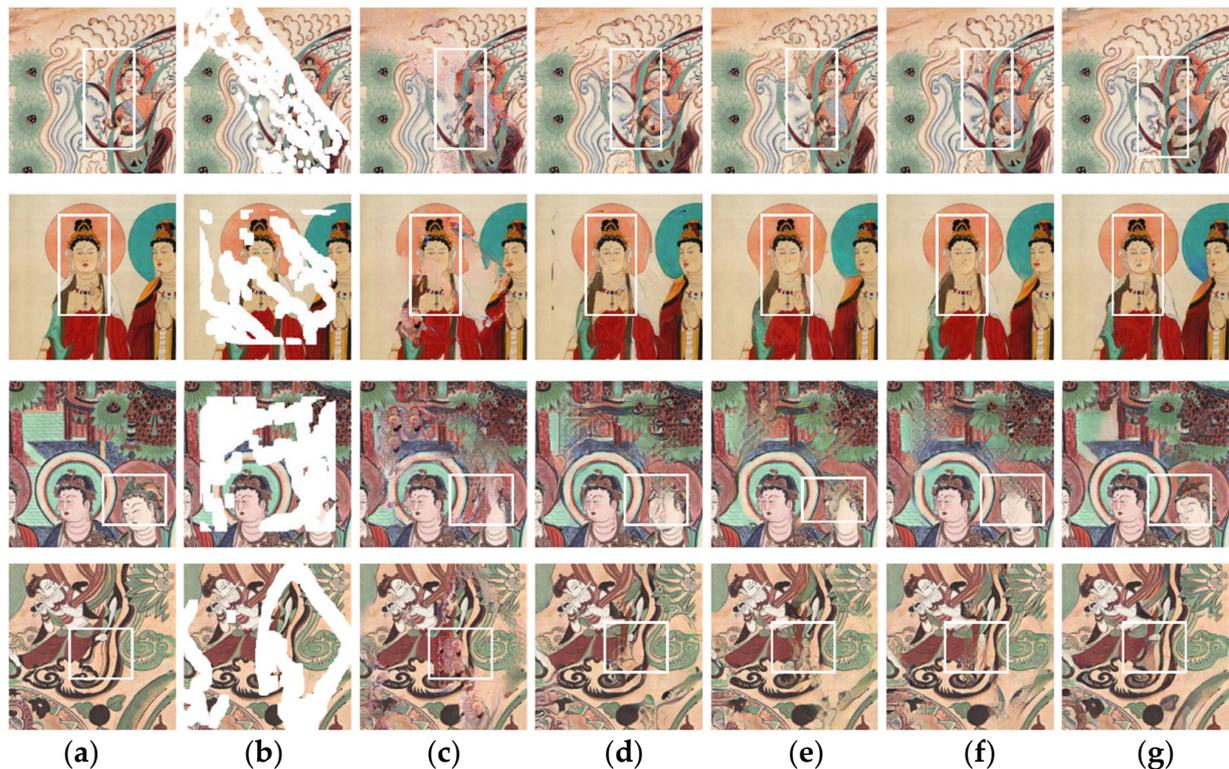


Figure 11. Inpainting results of different models using the mural dataset (White boxes are used for comparison of details). (a) Original mural images, (b) damaged mural images, (c) PI, (d) RFR, (e) EC, (f) LG, and (g) Edge-MSGAN (ours).

Table 5. Comparison of evaluation metrics of different models on the mural test set.

Model	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	LPIPS \downarrow
PI	19.1456	0.7180	0.0131	0.2235
RFR	20.1886	0.7457	0.0152	0.1747
EC	19.9326	0.7341	0.0156	0.1828
LG	21.1360	0.7894	0.0086	0.1969
Edge-MSGAN (Ours)	23.8079	0.8794	0.0047	0.0979

Based on the inpainting results obtained from the constructed ancient painting inpainting dataset and existing mural dataset, the proposed model has demonstrated excellent performance in the digital inpainting of historical and cultural heritage such as ancient paintings. This indicates that the model has a wide range of applications in the inpainting of ancient paintings and other cultural heritages.

5. Conclusions

A Chinese ancient painting inpainting model is proposed in this paper based on edge guidance and multi-scale residual blocks. The model effectively restores the texture,

contour, and color by utilizing the edge images of original ancient paintings, an ancient painting completion network, a multi-branch color correction network, and a Markov discriminator. During the ancient painting completion stage, multi-scale residual blocks are used to better handle the details and texture information of different scales in ancient paintings. Additionally, a polarized self-attention mechanism is employed to increase the model's focus on important structures, edges, and details of ancient paintings. Furthermore, a multi-branch color correction network is designed to achieve feature transmission and fusion at different levels and scales, which helps in restoring the color consistency of ancient paintings. Through testing on the ancient painting inpainting dataset, the proposed model demonstrates excellent performance in various types of ancient paintings inpainting, whether it is landscape or figure painting. The inpainting results exhibit clear lines, rich details, and harmonious colors. Moreover, inpainting experiments are conducted on mural datasets to validate the model under different scenarios, proving its effectiveness and applicability. This research aims to fully embody the integration of culture and technology and provide technological support for the digital inpainting of cultural heritage such as ancient paintings.

Author Contributions: Conceptualization, Z.S.; data curation, Y.L.; funding acquisition, Z.S.; investigation, Y.L.; methodology, Z.S.; supervision, Z.S.; validation, Y.L. and X.W.; writing—original draft, Z.S.; writing—review and editing, Z.S., Y.L. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No. 2017YFB1402102), the National Natural Science Foundation of China (No. 62377033), the Shaanxi Key Science and Technology Innovation Team Project (No. 2022TD-26), the Xi'an Science and Technology Plan Project (No. 23ZDCYJSGG0010-2022), and the Fundamental Research Funds for the Central Universities (No. GK202205036, GK202101004).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Mu, T.Q. Research on Intelligent Virtual Recovery Technology and Application. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2022.
2. You, Y. An analysis of ancient calligraphy and painting restoration processes and conservation methods. *Appreciation* **2020**, *36*, 32–33.
3. Wu, Y.F. The Application of Image Restoration Algorithms to Chinese Paintings. Master's Thesis, Zhejiang University, Hangzhou, China, 2008.
4. Pathak, D.; Krähenbühl, 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.
5. Liu, G.L.; 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, Munich, Germany, 8–14 September 2018; pp. 85–100.
6. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4471–4480.
7. 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.
8. Qu, S.Y.; Niu, Z.X.; Zhu, J.K.; Dong, B.; Huang, K.Z. Structure first detail next: Image inpainting with pyramid generator. In Proceedings of the IEEE Conference on Multimedia and Expo, Brisbane, Australia, 10–14 July 2023; pp. 1265–1270.
9. Wang, T.F.; Ouyang, H.; Chen, Q.F. Image inpainting with external internal learning and monochromic bottleneck. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 5120–5129.
10. Zeng, Y.H.; Fu, J.L.; Chao, H.Y.; Guo, B.N. Learning pyramid-context encoder network for high-quality image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 1486–1494.
11. Li, J.Y.; Wang, N.; Zhang, L.F.; Du, B.; Tao, D.C. Recurrent feature reasoning for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7760–7768.
12. Cao, J.F.; Zhang, Z.B.; Zhao, A.D.; Cui, H.Y.; Zhang, Q. Ancient mural restoration based on a modified generative adversarial network. *Herit. Sci.* **2020**, *8*, 7. [[CrossRef](#)]

13. Wan, G.L. Research on the Restoration Algorithm of Cultural Relics Based on Gated Convolution and Coherent Semantic Attention Mechanism. Master's Thesis, Ningxia University, Yinchuan, China, 2022.
14. Zhang, Y.H.; Ding, J.H.; Yu, J.Y. Art image inpainting via embedding multiple attention dilated convolutions. In *Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1–14.
15. Zhang, Y.H. Research on Art Image Inpainting Method. Master's Thesis, Hangzhou Dianzi University, Hangzhou, China, 2023.
16. Zhao, N. Design of Chinese paintings' color restoration system based on image processing technology. *Mod. Electron. Tech.* **2020**, *43*, 60–63, 68.
17. Xu, G.B.; Yu, Y.M.; Li, J.; Wang, X.J.; Chen, Q.; Wang, Q. Defective Chinese painting digital image restoration using improved BSCB algorithm. *Laser Optoelectron Prog* **2022**, *59*, 81–89.
18. Yao, F. Damaged region filling by improved criminisi image inpainting algorithm for thangka. *Clust. Comput.* **2019**, *22*, 13683–13691. [[CrossRef](#)]
19. Zhou, J.; Duan, L.R. Study of digital restoration of ancient paintings by Yan Yun Shuang Lu. *Identif. Apprec. Cult. Relics* **2021**, *198*, 78–80.
20. Ma, W.; Long, Q.Q.; Qin, Y.; Xu, S.B.; Zhang, X.P. Repairing high-definition ancient paintings based on decomposition of curves. *J. Comput.-Aided Des. Comput. Graph.* **2018**, *30*, 1652–1661. [[CrossRef](#)]
21. Wang, H.; Li, Q.Q.; Jia, S. A global and local feature weighted method for ancient murals inpainting. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1197–1216. [[CrossRef](#)]
22. Bhele, S.; Shriramwar, S.; Agarkar, P. An efficient texture-structure conserving patch matching algorithm for inpainting mural images. In *Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1–22.
23. Yan, L.X. Researches on Virtual Recovery of Chinese Paintings Based on Multispectral Technology. Master's Thesis, Tianjin University, Tianjin, China, 2012.
24. Zhou, P.P.; Hou, M.L.; Zhao, X.S.; Lv, S.Q.; Hu, Y.G.; Zhang, X.D.; Zhao, H.Q. Virtual restoration of ancient painting stains based on classified linear regression of hyper-spectral image. *J. Spatio-Temporal Inf.* **2017**, *24*, 113–118.
25. Hou, M.L.; Zhou, P.P.; Lv, S.Q.; Hu, Y.G.; Zhao, X.S.; Wu, W.T.; He, H.P.; Li, S.N.; Tan, L. Virtual restoration of stains on ancient paintings with maximum noise fraction transformation based on the hyperspectral imaging. *J. Cult. Herit.* **2018**, *34*, 136–144. [[CrossRef](#)]
26. Zhang, R.; Jiao, X.Q. Application of image processing technology based on OpenCV in traditional Chinese painting. *Autom. Instrum.* **2019**, *9*, 226–229.
27. Zhao, L.; Ji, B.Y.; Xing, W.; Lin, H.Z.; Lin, Z.J. Ancient painting inpainting algorithm based on multi-channel encoder and dual attention. *J. Comput. Res. Dev.* **2023**, *50*, 334–341.
28. Xue, J.T. Research on Artificial Image Completion and Translation Model Based on Wasserstein Generative Adversarial Network. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2021.
29. Yang, L.H. Research on Image Inpainting of Ancient Paintings Based on Optimized Generative Adversarial Networks. Master's Thesis, Northwest University, Xi'an, China, 2022.
30. Zhao, L.; Lin, S.H.; Lin, Z.J.; Ding, J.H.; Huang, J.; Xing, W.; Lin, H.Z.; Lu, D.M. Progressive multilevel feature inpainting algorithm for Chinese ancient paintings. *J. Comput.-Aided Des. Comput. Graph.* **2023**, *35*, 1–13.
31. Liu, H.W.; Yao, J.C.; Liu, B.; Bi, X.L.; Xiao, B. Two-stage method for restoration of heritage images based on multi-scale attention mechanism. *Comput. Sci.* **2023**, *50*, 334–341.
32. Zhou, Z.H.; Liu, X.R.; Shang, J.Y.; Huang, J.C.; Li, Z.H.; Jia, H.P. Inpainting digital Dunhuang murals with structure-guided deep network. *ACM J. Comput. Cult. Herit.* **2022**, *15*, 77. [[CrossRef](#)]
33. Lv, C.H.; Li, Z.L.; Shen, Y.H.; Li, J.H.; Zheng, J. SeparaFill: Two generators connected mural image restoration based on generative adversarial network with skip connect. *Herit. Sci.* **2022**, *10*, 135. [[CrossRef](#)]
34. Peng, X.L.; Zhao, H.Y.; Wang, X.Y.; Zhang, Y.Q.; Li, Z.; Zhang, Q.X.; Wang, J.; Peng, J.Y.; Liang, H.D. C3N: Content-constrained convolutional network for mural image completion. *Neural Comput. Appl.* **2023**, *35*, 1959–1970. [[CrossRef](#)]
35. Deng, X.C.; Yu, Y. Ancient mural inpainting via structure information guided two-branch model. *Herit. Sci.* **2023**, *11*, 131. [[CrossRef](#)]
36. Horvath, M.; Bowers, M.; Alawneh, S. Canny edge detection on GPU using CUDA. In Proceedings of the Computing and Communication Workshop and Conference, Virtual Conference, 8–11 March 2023; pp. 419–425.
37. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
38. Wu, D.X.; Wang, Y.S.; Xia, S.T.; Bailey, J.; Ma, X.J. Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv* **2020**, arXiv:2002.05990.
39. Li, H.Y.; Wu, Z.Y.; Guo, L.; Chen, J.H. Multi-discriminator image inpainting algorithm based on hybrid dilated convolution network. *J. Huazhong Univ. Sci. Technol.* **2021**, *49*, 40–45.
40. Li, Y.W.; Fan, Y.C.; Xiang, X.Y.; Demandolx, D.; Ranjan, R.; Timofte, R.; Gool, L.V. Efficient and explicit modelling of image hierarchies for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–27 June 2023; 18278–18289.
41. Hou, Q.B.; Zhou, D.Q.; Feng, J.S. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13713–13722.

42. Zhang, H.; Zu, K.K.; Lu, J.; Zou, Y.R.; Meng, D.Y. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. In Proceedings of the Asian Conference on Computer Vision, Macao, China, 4–8 December 2022; pp. 1161–1177.
43. Fu, J.; Liu, J.; Tian, H.J.; Li, Y.; Bao, Y.J.; Fang, Z.W.; Lu, H.Q. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
44. Huang, Z.L.; Wang, X.G.; Huang, L.C.; Huang, C.; Wei, Y.C.; Liu, W.Y. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 603–612.
45. Liu, Y.C.; Shao, Z.R.; Teng, Y.Y.; Hoffmann, N. NAM: Normalization-based attention module. *arXiv* **2021**, arXiv:2111.12419.
46. Liu, H.J.; Liu, F.Q.; Fan, X.Y.; Huang, D. Polarized self-attention: Towards high-quality pixel-wise regression. *arXiv* **2022**, arXiv:2111.12419.
47. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
48. Chen, G.; Zhang, G.P.; Yang, Z.G.; Liu, W.Y. Multi-scale patch-GAN with edge detection for image inpainting. *Appl. Intell.* **2023**, *53*, 3917–3932. [[CrossRef](#)]
49. Zhuang, X.Q.; Li, C.X.; Li, P.X. Style transfer based on cross-layer correlation perceptual loss. *Acta Sci. Nat. Univ. Sunyatseni* **2020**, *59*, 126–135.
50. Li, L.X.; Zou, Q.; Zhang, F.; Yu, H.K.; Chen, L.; Song, C.F.; Huang, X.F.; Wang, X.G. Line drawing guided progressive inpainting of mural damages. *arXiv* **2022**, arXiv:2211.06649.
51. Mor, A.-A.; Arblle, A.; Raviv, T.R. Differentiable histogram loss functions for intensity-based image-to-image translation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *48*, 11642–11653.
52. Chen, M.; Pu, Y.F.; Bai, Y.C. Low-dose CT image denoising using residual convolutional network with fractional TV loss. *Neurocomputing* **2021**, *452*, 510–520. [[CrossRef](#)]
53. Zheng, C.X.; Cham, T.J.; Cai, J.F. Pluralistic image completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 1438–1447.
54. Quan, W.Z.; Zhang, R.S.; Zhang, Y.; Li, Z.F.; Wang, J.; Yan, D.M. Image inpainting with local and global refinement. *IEEE Trans. Image Process.* **2022**, *31*, 2405–2420. [[CrossRef](#)]
55. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.Z.; Ebrahimi, M. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv* **2019**, arXiv:1901.00212.

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.