Heritage Science

RESEARCH **Open Access**

A virtual restoration network of ancient murals via global–local feature extraction and structural information guidance

Hao Ge[1], Ying Yu[1*] and Le Zhang[1]

Abstract

Ancient murals are precious cultural heritages. They suffer from various damages due to man-made destruction and long-time exposure to the environment. It is urgent to protect and restore the damaged ancient murals. Virtual restoration of ancient murals aims to fill damaged mural regions by using modern computer techniques. Most existing restoration approaches fail to fill the loss mural regions with rich details and complex structures. In this paper, we propose a virtual restoration network of ancient murals based on global–local feature extraction and structural information guidance (GLSI). The proposed network consists of two major sub-networks: the structural information generator (SIG) and the image content generator (ICG). In the first sub-network, SIG can predict the structural information and the coarse contents for the missing mural regions. In the second sub-network, ICG utilizes the predicted structural information and the coarse contents to generate the refined image contents for the missing mural regions. Moreover, we design an innovative BranchBlock module that can effectively extract and integrate the local and global features. We introduce a Fast Fourier Convolution (FFC) to improve the color restoration for the missing mural regions. We conduct experiments over simulated and real damaged murals. Experimental results show that our proposed method outperforms other three comparative state-of-the-art approaches in terms of structural continuity, color harmony and visual rationality of the restored mural images. In addition, the mural restoration results of our method can achieve comparatively high quantitative evaluation metrics.

Keywords Mural restoration, Structure information, BranchBlock, FFC

Introduction

Ancient murals are important human cultural heritages, which record large amounts of contents related to the social, religious, artistic life of various ethnic groups, and some historical events [1, 2]. Due to the degradation caused by long-time environmental exposure and human activities, many ancient murals are suffering from serious diseases such as cracks, scratches, corrosion, paint loss, and even large-area falling off. These diseases may destroy the integrity of the mural contents that are of historical, cultural, religious and artistic values. Therefore, the protection and restoration of ancient murals has become an urgent work for the cultural heritage communities all over the world.

Physical restoration of the damaged murals is very difficult and time-consuming, which relies on the proficiency level of mural repair experts, and may cause irreversible damages to the mural heritages. In recent years, virtual restoration techniques of ancient murals attempt to fill the missing or deteriorated regions of the damaged murals by using intelligent computer algorithms. The restored mural images not only serve as references to the physical repair work, but also offer a permanent and replicable database for these precious cultural heritages.

*Correspondence:
Ying Yu
yuying.mail@163.com
[1] School of Information Science and Engineering, Yunnan University, Kunming 650500, China

The goal of virtual restoration of ancient murals is to fill the missing or diseased mural areas with semantically continuous and visually reasonable contents. Traditional mural restoration methods mainly include diffusion-based methods and patch-based methods. The diffusion-based methods achieve image restoration by deriving information near the damaged area [3]. Cheng et al. [4] proposed a curvature-driven diffusion model based on adaptive control and smooth fusion to repair complex shapes and irregular scratches of the murals. The diffusion-based methods can effectively restore the murals with narrow and long cracks, but are unsuitable for large-area damages. The patch-based methods fill the missing regions by matching and copying the most similar pixel patches from the known mural regions [5]. Li et al. [6] proposed a YCbCr model to analyze the brightness and color characteristics of murals, and used the patch-based method to restore the mud spot diseases. Yang et al. [7] proposed a priority algorithm based on the D–S evidence reasoning theory and data fusion that can restore the murals with large-area damages. Jiao et al. [8] repaired the damaged regions of Wutai Mountain murals by use of an improved block matching algorithm. Shen et al. [9] conducted morphological component analysis to obtain the structural part and texture part of a mural image, and used K-singular value to achieve ancient murals inpainting. Wang et al. [10] utilized line drawings and global–local patches to produce better structural continuity. The patch-based methods are suitable for restoring relatively large damaged areas, but cannot generate contents outside the undamaged mural areas.

With the technological advancement of deep neural network and intelligent information processing, a number of image inpainting methods based on the convolution neural network (CNN) [11] and the generative adversarial network (GAN) [12] have achieved outstanding performance in natural image restoration. They are superior to traditional methods because they have the ability to adaptively learn high-level features from image data.

In recent years, many researchers have attempted to use deep learning approaches to deal with the mural restoration problem. Cao et al. [13] used GAN to restore ancient murals, and introduced the dilated convolution to increase the receptive field of the network. Wang et al. [14] proposed a sparse representation model based on the so-called global and local feature consistency enhancement to predict a convincing target patch. Wang et al. [15] proposed the adaptive partial convolution to enlarge the receptive field, and used a novel mask generator to produce the matched stroke-like mask. Li et al. [16] employed manual line-drawings as auxiliary information to guide the restoration of missing mural areas. Inspired by an artist's image-making process, Ciortan et al. [17] proposed a multi-stage mural restoration network based on "lines first, color palette after, color tones at last", and used the Four random-walk masks to imitate various degradations of ancient murals. Lv et al. [18] proposed an image restoration network based on two connected generators, and obtained fine-detailed results for the Dunhuang murals. Schmidt et al. [19] utilized super-resolution and deblurring techniques to restore the deteriorated cave paintings. Yu et al. [20] adopted two types of masks that are more suitable for simulating the missing areas of deteriorated murals. In reference [21], the mural restoration network needs complete line-drawings as the auxiliary information, which were manually drawn by mural repairing experts. Zhou et al. [22] proposed a two-stage network that first generates gradient information and then fills color information for the missing mural region.

Although the above methods are capable of dealing with specific mural damages, there still exist four challenges for a mural restoration network: (1) Most ancient murals have rich colors, complex structures and abundant contents. To restore the murals with large-area damages, a mural restoration network should have the ability to capture multi-scale semantic information of a mural image. This will pose difficulties for the traditional CNN that has limited receptive fields. (2) Many murals have various diseases, and the masks for these diseases vary significantly. A mural restoration network needs suitable masks to indicate the degradation regions in the murals. (3) It is difficult for a mural research group to collect enough available mural data for training the deep neural network. This will result in poor network generalization ability. (4) Due to the scarcity of mural data, we need an efficient network suitable for mural restoration tasks.

To tackle with the challenges mentioned above, we propose a mural restoration network based on global–local feature extraction and structural information guidance. Figure 1 illustrates two examples of our model's procedure from the damaged mural images to the output restored results. The main contribution of this paper includes the following three points: First (solve challenge 1). We propose an innovative network model that has powerful ability to restore those damaged murals that contain complex structures and abundant semantic contents. The model consists of two sub-networks: the structural information generator (SIG) and the image content generator (ICG). SIG is capable of predicting the structures and the coarse contents of the missing mural regions. ICG can effectively restore the refined image contents of the missing mural regions. Moreover, we design an innovative BranchBlock module that helps SIG to effectively extract and integrate the local and

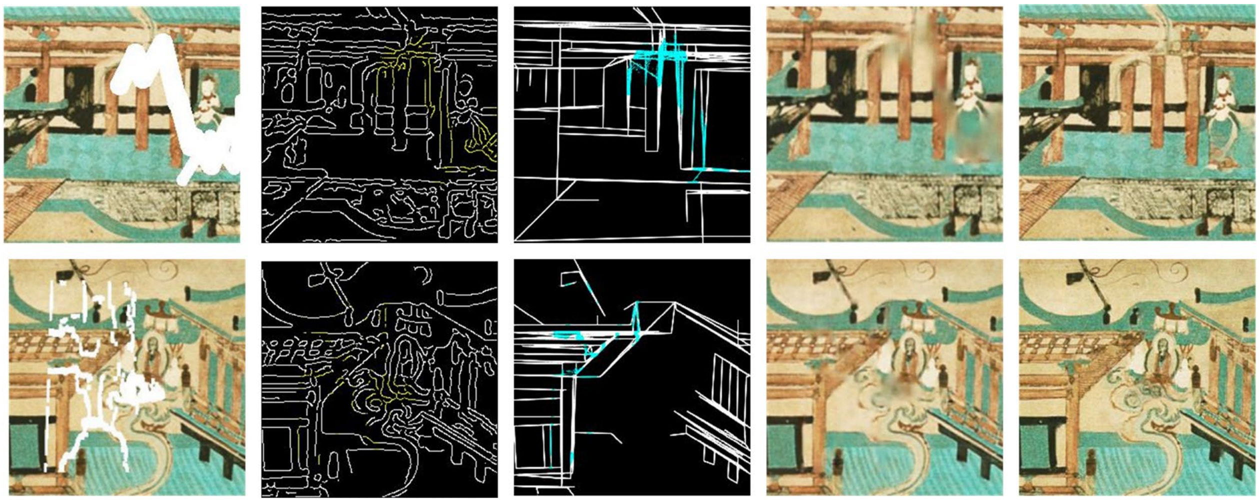

Fig. 1 Two examples of our model's procedure from the damaged mural images to the output restored results

global features. We also introduce a Fast Fourier Convolution (FFC) module that expands the receptive field of ICG. Second (solve challenge 2). We utilize stroke-like and irregular masks to simulate the cracks and falling off diseases of the murals. In this way, the proposed model can achieve better restoration results when dealing with real damaged murals. Third (solve challenge 3). We build an ancient mural image dataset by collecting 3466 high-quality ancient mural images and expanding the number of these mural images to 10,398 by use of data augmentation techniques. Fourth (solve challenge 4). To ensure the effectiveness of the network training, we employ the MobileConv module and use the partial GatedConv strategy to reduce the number of parameters in the network.

Proposed method

The damaged ancient murals usually contain complex structures and abundant semantic contents. It is a very challenging task to restore such missing regions of the damaged mural images. It has been noted that most of the image information consists in the structures, e.g., edges and lines of an image. Therefore, reliable structure information prediction for the damaged regions plays an important role in guiding the restoration of the damaged murals. As a manual procedure of mural creation, most mural experts will first sketch the line-drawings and then fill them with colors and details. Motivated by this point, we propose a mural restoration model that focuses on predicting the structure information before content restoration. Figure 2 shows the overall network architecture of the proposed model. Our model can be divided into two sub-networks: structure information generator (SIG) and image content generator (ICG). SIG aims to predict the reliable structure information and the coarse image contents of a damaged mural, and ICG performs refined content inpainting that is guided by the structure information from SIG. Each sub-network focuses on a specific task in mural restoration, i.e., the structure information restoration and image content restoration.

Given an input ground truth mural image I_{gt},we combine the mural with a binary mask M to obtain a masked mural image I_{m} by using the operation $I_{\mathrm{m}} = I_{\mathrm{gt}} \odot (1 - M)$, where the binary mask M indicates the mural region that needs to be restored, and the operation $\odot$ denotes the Hadamard product. We need to extract the structure information (lines and edges) from the masked mural image. To begin with, we attempt to extract the structure information from the mural image. To obtain an accurate line map I_{l} of the mural image, we use the line segment masking (LSM) algorithm [23] to train a deep learning-based wireframe parse HAWP [24]. This deep learning-based line extractor is referred to as LSM-HAWP in this paper. Moreover, we use the Canny detection operator to obtain the edge map I_{e} of the mural image. Therefore, the procedure of extracting structure information from the masked mural image can be formulated as $I_{\mathrm{e-mask}} = I_{\mathrm{e}} \odot (1 - M)$, $I_{\mathrm{l-mask}} = I_{\mathrm{l}} \odot (1 - M)$ on the channel dimension. We use $I_{\mathrm{e-mask}}$ and $I_{\mathrm{l-mask}}$ as the input of SIG. After that, SIG can produce the restored structure map I_{st} and the coarse content image I_{c} of the masked mural. By using I_{st} and I_{c} as input, ICG will finally restore the refined contents of the masked mural.

Structure information generator (SIG)

Large damaged areas with complex structures and rich semantic contents often lead to structural disorder and poor content consistency in the restored murals. Notice that the structure information of a mural can

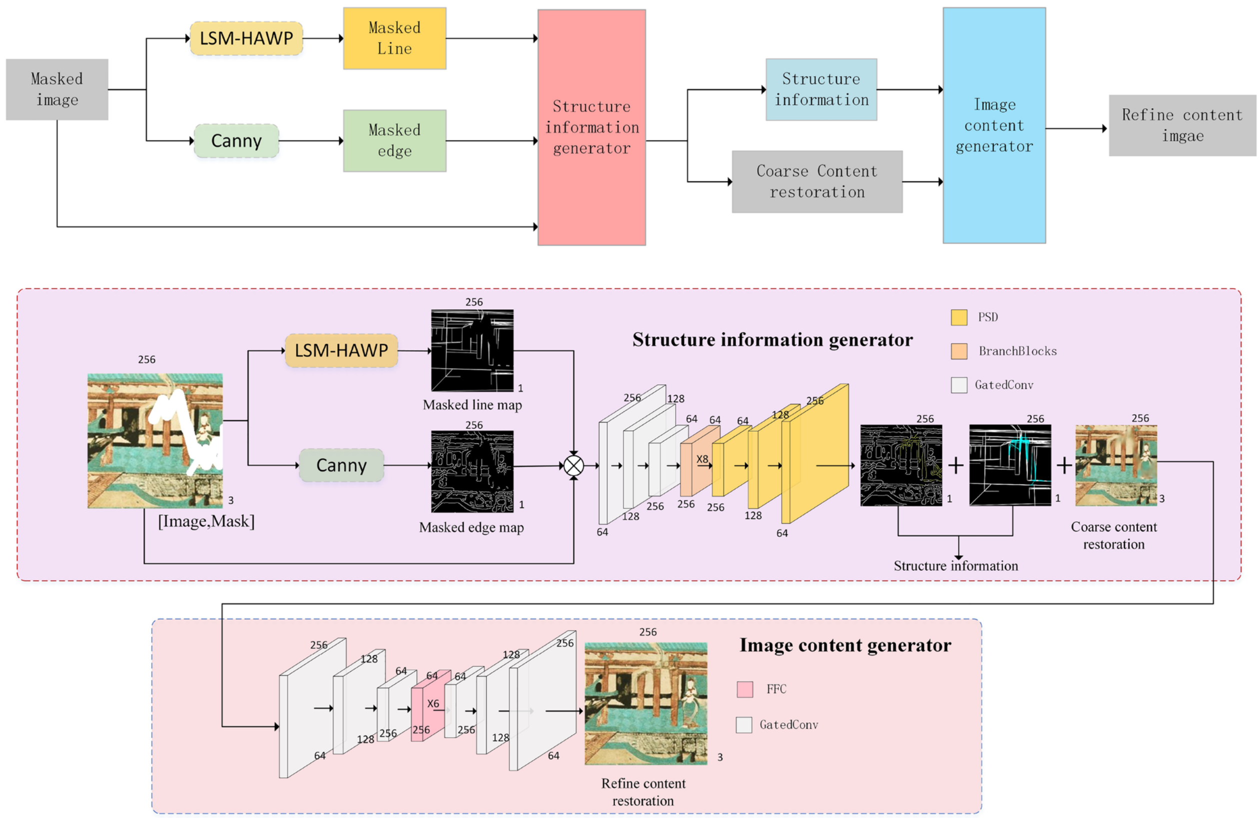

Fig. 2 The overall network architecture of the proposed mode

help to repair the missing mural areas. In this subsection, we design a structure information prediction network, also referred to as the structure information generator (SIG), to improve the restoration quality of the damaged murals. SIG aims to predict the complete structure feature of the mural. It also produces a coarse content image of the mural. The SIG network contains three kinds of modules: GatedConv [25], BranchBlock, and PSD. The input layer of SIG contains three gated convolution (GatedConv) modules that can filter out invalid pixels of the damaged region, and can enhance the ability of feature extraction. Different from vanilla convolution and partial convolution, the GatedConv uses a dynamic feature selection mechanism to adaptively select relevant features for each location and channel.

Given the input feature $F_{\text{in,GC}}$, the process of a GatedConv can be expressed as $F_{\text{out}} = \sigma(G) \odot \phi(F_e)$. The gating feature G is generated by $G = \text{Conv}_1 3 \times 3(F_{\text{in,GC}})$, and the feature F_e is obtained by $F_e = \text{Conv}_2 3 \times 3(F_{\text{in,GC}})$. The notation σ denotes the sigmoid activation function, and the notation ϕ denotes the ReLU activation function. Figure 3 shows the architecture of the input layer of SIG that is composed of three GatedConv modules.

Notice that global context information plays an important role in restoring large-area damages, whereas local context features are more suitable for predicting the detailed contents. In order to improve the quality of mural restoration, we propose a parallel module that can capture the global and local context features simultaneously. This parallel module is referred to as BranchBlock in this paper. The BranchBlock employs an adaptive learning mechanism to tune the weights of the global and local features. Since the BranchBlock adopts parallel network connections, its network size will inevitably expand. We consider adopting a lightweight design for the convolution and attention modules. To this end, we introduce the MobileConv [26] module with inverted residual blocks, which not only reduces the computation cost but also maintains relatively good performance. Figure 4 shows the detailed structure of the proposed BranckBlock.

In the branch of convolution module, taking the feature $F_{\text{in,BB}} \in \mathbb{R}^{H \times W \times C}$ as input, the MobileConv will extract features through point-wise convolution and depth-wise convolution. These operations can be expressed as

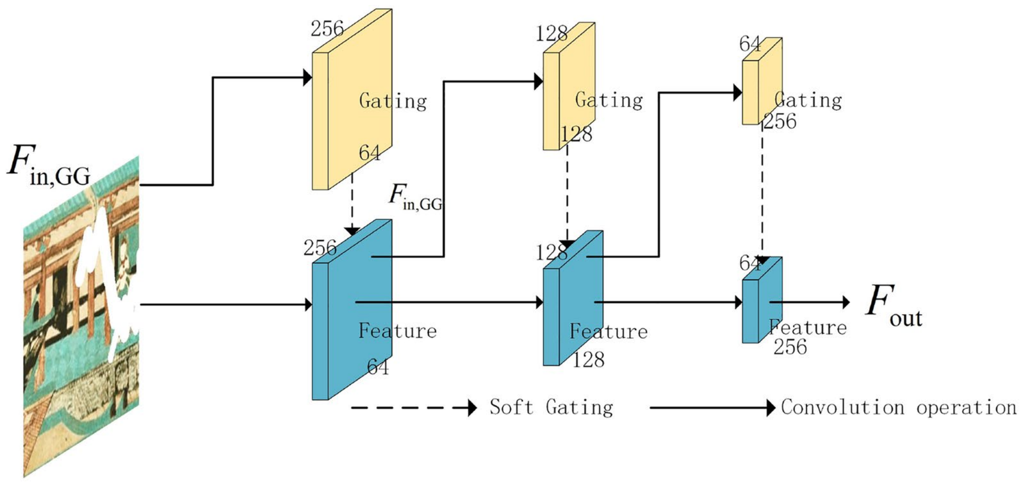

Fig. 3 The architecture of the input layer of SIG

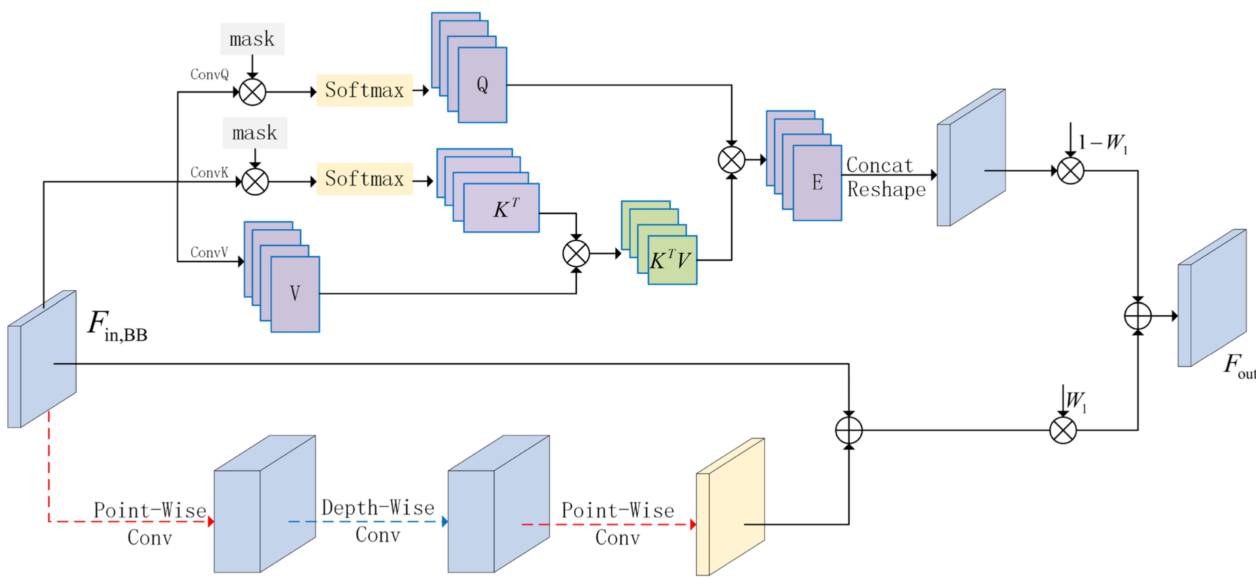

Fig. 4 The structure of the proposed BranchBlock

$$
\begin{aligned}
F_c^1 &= \sigma(\mathrm{Conv1}\times 1(F_{\mathrm{in,BB}})), \mathbb{R}^{H\times W\times C} \rightarrow \mathbb{R}^{H\times W\times \frac{3C}{2}} \\
F_c^2 &= \sigma(\mathrm{Conv3}\times 3(F_c^1)), \mathbb{R}^{H\times W\times \frac{3C}{2}} \rightarrow \mathbb{R}^{H\times W\times \frac{3C}{2}} \\
F_c^3 &= \sigma(\mathrm{Conv1}\times 1(F_c^2)), \mathbb{R}^{H\times W\times \frac{3C}{2}} \rightarrow \mathbb{R}^{H\times W\times C}
\end{aligned} \tag{1}
$$

In the branch of attention module, we reshape the input feature $F_{\mathrm{in}} \in \mathbb{R}^{H\times W\times C}$ and the binary mask $M \in \mathbb{R}^{H\times W\times 1}$ to $\mathbb{R}^{HW\times C}$ and $\mathbb{R}^{HW\times 1}$, respectively. Then we compute the *Query* and the *Key* by using the equation $Q = softmax_{row}(W_q F_{\mathrm{in,BB}} \times Mask)$ and the equation $K = softmax_{col}(W_k F_{\mathrm{in,BB}} \times Mask)$, where $softmax_{row}$ and $softmax_{col}$ denote the individual softmax operations on the row and column, and $\times Mask$ means to emphasize the masked region. The *Value* is obtained by using $V = W_v F_{\mathrm{in}}$. The parameters $W_{q,k,v}$ is learned in the attention module. To reduce the computation cost of the attention module, we modify the equation $F_a = (QK^{\mathrm{T}})V$ to $F_a = Q(K^{\mathrm{T}}V)$. After extracting features from the convolution and attention modules, the BranchBlock employs a self-learning parameter $W_1 = f_{ab}(F_{\mathrm{in,BB}})$ to assign optimal weights to each parallel module, where f_{ab} indicates the Conv → IN → Sigmoid. Finally, the BranchBlock computes the output feature by using $F_{\mathrm{out}} = W_1 \odot (F_c^3 + F_{\mathrm{in}}) + (1 - W_1) \odot F_a$.

We employ the gated convolution to adaptively filter the feature information and use the BranchBlock module to extract and integrate the local and global context information. We use a pyramid spatial decomposition (PSD) module to decouple the feature information into structure information and coarse content information. In the input layer of the PSD module, we first perform a GatedConv on the input feature $F_{\text{in,PSD}} \in \mathbb{R}^{H\times W\times C}$ to obtain a high credibility feature F_{hc} by using $F_{\text{hc}} = \text{GatedConv}(F_{\text{in,PSD}})$. Then we obtain the structure feature F_{st} by using $F_{\text{st}} = \varepsilon(\sigma(\text{Conv}(F_{\text{hc}})))$, where σ denotes the ReLU activation function, and ε denotes the Instance Norm operation. Next, we split F_{st} into the edge feature F_{edge} and the line feature F_{line} by using $\{F_{\text{line}}, F_{\text{line}}\}$ =split(F_{st}). Afterwards, the edge map S_{edge} and the line map S_{line} are computed by using $S_{\text{edge}} = \phi(\text{Conv}(F_{\text{edge}}))$ and $S_{\text{line}} = \phi(\text{Conv}(F_{\text{line}}))$ respectively, where ϕ denotes the sigmoid activation function.

We compose the edge feature F_{edge} and the line feature F_{line} to predict a coarse content feature F_{cc} by using $F_{\text{cc}} = W_2 \odot F_{\text{edge}} + (1 - W_2)$, the weight W_2 is computed by $W_2 = f_{\text{cd}}(F_{\text{in.PSD}})$, where f_{cd} denotes a series of operations Conv $\rightarrow$ *IN* $\rightarrow$ *ReLU* $\rightarrow$ Conv2D $\rightarrow$ Sigmoid. The coarse content image S_{img} is visualized by using $S_{\text{img}} = \phi(\text{Conv}(F_{\text{cc}}))$. The final output of the PSD module can be computed as $F_{\text{out}} = F_{\text{hc}} + F_{\text{cc}}$. Figure 5 shows the detailed structure of the PSD module.

The proposed SIG network contains three PSD modules. The first PSD module takes the input feature $F \in \mathbb{R}^{64\times 64\times 64}$ And this module would process this input

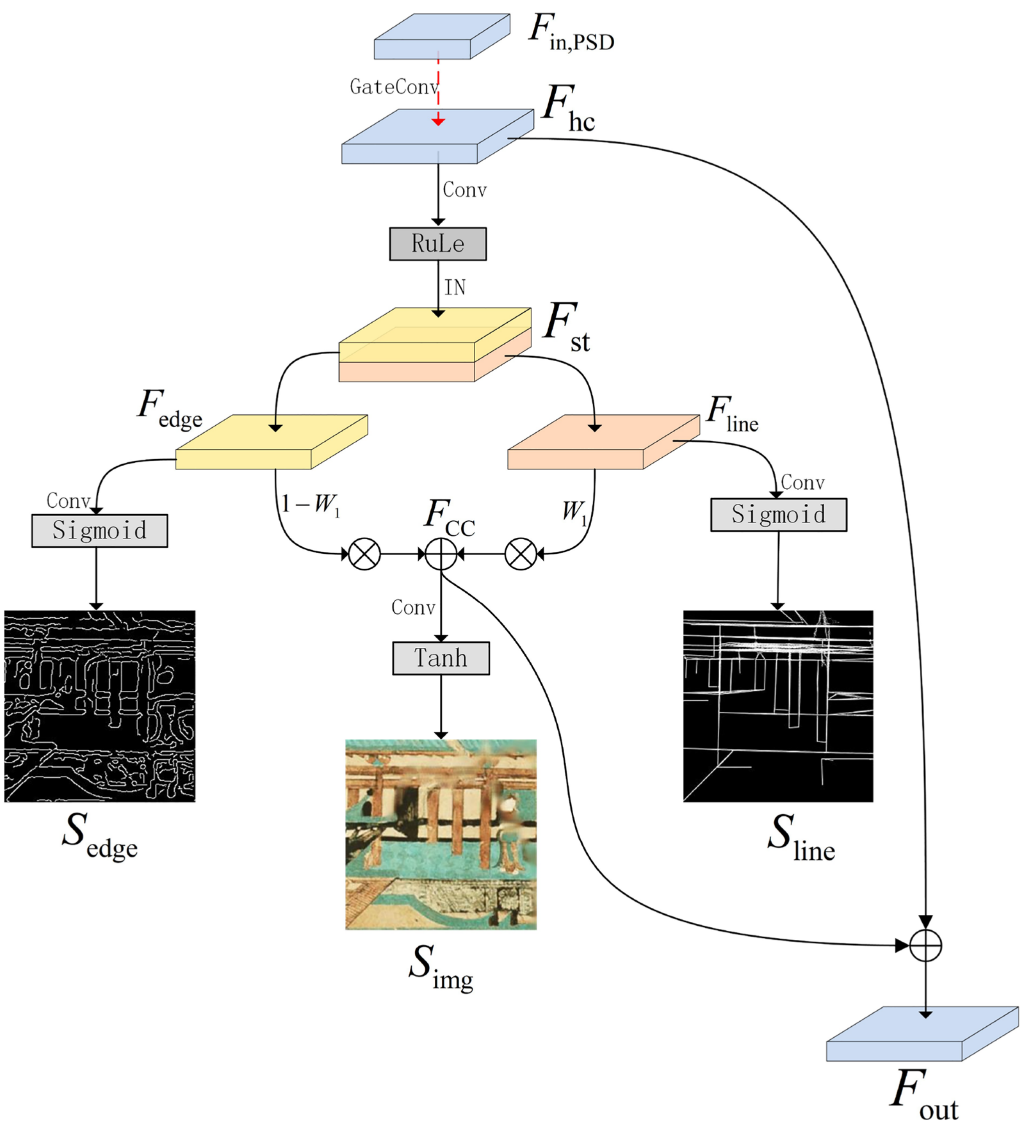

Fig. 5 The structure of the PSD module

feature to obtain $[F^1_{\text{out}} \in \mathbb{R}^{128\times128\times128}, S^1_{\text{img}} \in \mathbb{R}^{64\times64\times3}, S^1_{\text{edge}} \in \mathbb{R}^{64\times64\times1}, S^1_{\text{line}} \in \mathbb{R}^{64\times64\times1}]$. Through the iterative process of three PSD modules, SIG can predict a coarse content image $I_c = S^3_{\text{img}} \in \mathbb{R}^{256\times256\times256}$, a line map $S^3_{\text{line}} \in \mathbb{R}^{256\times256\times1}$ and a edge map $S^1_{\text{edge}} \in \mathbb{R}^{256\times256\times1}$. These three predicted results have the same size with the original 256 × 256px mural image. The structure information I_{st} is combined by using $I_{st} = \left\{S^3_{\text{edge}}, S^3_{\text{line}}\right\}$.

Image content generator (ICG)

The goal of the second sub-network ICG is to restore the refined contents of the mural based on the predicted information $\left\{S^3_{\text{img}}, S^3_{\text{line}}, S^3_{\text{edge}}\right\}$ from the first sub-network SIG. The ICG can restore the detailed textures and enhanced colors of the damaged mural. The backbone of ICG consists of six FFC [27] modules which can fully utilize the hardware devices and expand the receptive field of the ICG network. Figure 6 shows the detailed structure of a FFC module. In order to effectively leverage the predicted information $\left\{S^3_{\text{img}}, S^3_{\text{line}}, S^3_{\text{edge}}\right\}$, we employ the GatedConv modules in the rest part of ICG.

We use the FFC module to capture global and local context information. The FFC splits all the input channels into global and local branches with the ratio of 3:1. The local branch updates the feature through vanilla convolution with 3 × 3 kernel size. The global branch employs spectral transform to update the feature, which can effectively obtain the global context information of the mural image. The implementation steps of the FFC module are as follows:

(1) applies Real *FFT2d* to the input feature map, and concatenates real and imaginary parts across channel dimension:

$$\mathbb{R}^{H\times W\times C} \xrightarrow{FFT2d} \mathbb{C}^{H\times\frac{W}{2}\times C} \xrightarrow{concat} \mathbb{R}^{H\times\frac{W}{2}\times 2C}$$

(2) applies convolution block in frequency domain:

$$\mathbb{R}^{H\times\frac{W}{2}\times 2C} \xrightarrow{Conv1\times1\to BN\to ReLU} \mathbb{R}^{H\times\frac{W}{2}\times 2C}$$

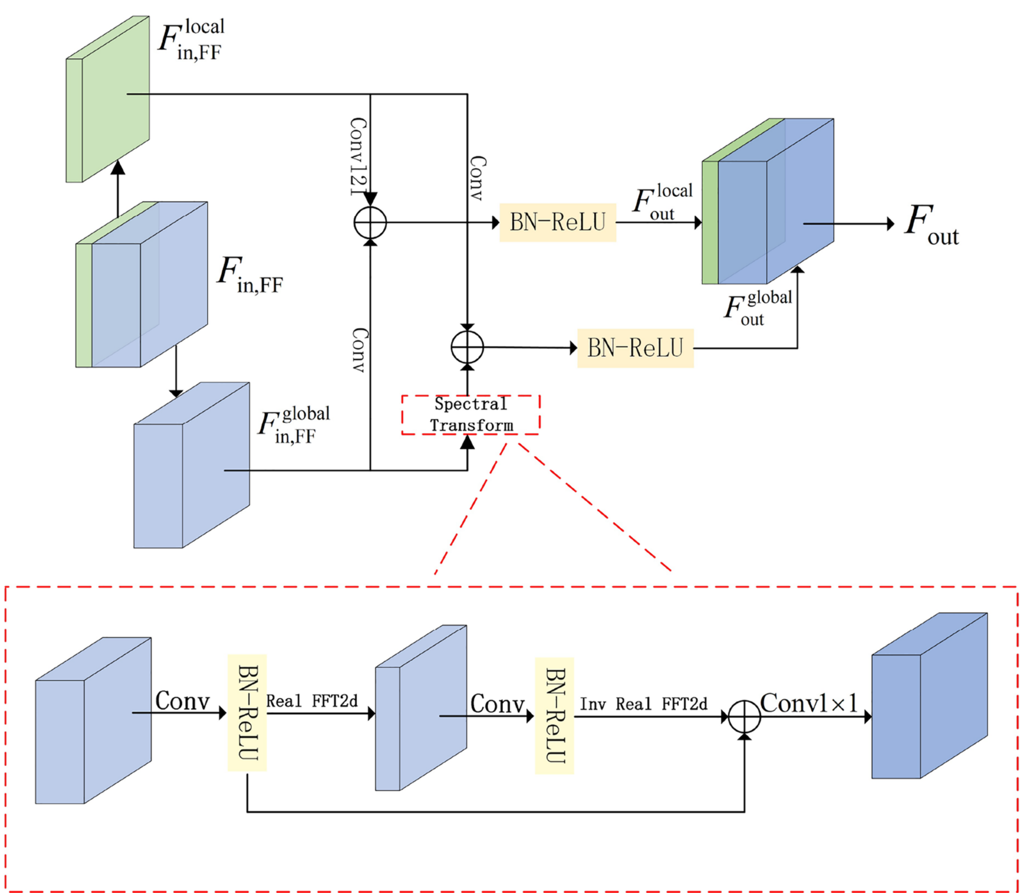

Fig. 6 The detailed structure of a FFC module

(3) applies inverse Fourier transform to recover a spatial structure:

$$\mathbb{R}^{H\times\frac{W}{2}\times 2C}\xrightarrow{concat}\mathbb{C}^{H\times\frac{W}{2}\times C}\xrightarrow{iFFT2d}\mathbb{R}^{H\times W\times C}$$

Finally, we fuse the updated global and local features to obtain the output features. With the outstanding performance of FFC in color restoration and under the guidance of the structural information from SIG, our proposed network can significantly improve the quality of mural restoration with consistent structures and natural colors.

Loss function

Since the goal of the SIG network is to predict reasonable structural information, we employ a line discriminator and an edge discriminator based on SN-PathGAN [25] in the training process. The two discriminators utilize a loss function by comparing the ground truth edge e_{gt} and line l_{gt} with the predicted edge S^3_{edge} and line S^3_{line}. The loss function of the discriminator of SIG can be expressed as

$$\begin{aligned} L_D^{SIG} &= L_{D_l} + L_{D_e}, where \\ L_{D_l} &= -E[\log D_{line}(l_{gt})] - E[1-\log D_{line}(S^3_{img})] \\ L_{D_e} &= -E[\log D_{edge}(e_{gt})] - E[1-\log D_{edge}(S^3_{edge})] \end{aligned} \tag{2}$$

The loss function of the generator of SIG is formulated as

$$L_G^{SIG} = \lambda_\alpha L_{adv}^{SIG} + \lambda_f L_{fm} + \sum_i^3 [S^i_{img} - I^i_{gt}] \tag{3}$$

$$L_{adv}^{SIG} = -E[\log D_{edge}(S^3_{edge})] - E[\log D_{line}(S^3_{line})] \tag{4}$$

where the predicted coarse content S^i_{img} computed by the i-th PSD module, and the L_{fn} is the feature matching loss [28]. The size of the predicted coarse content S^i_{img} and the ground truth image I_{gt} is 64 × 64px when i = 1, and 128 × 128px when i = 2, and 256 × 256px when i = 3. In this work, we set $\lambda_\alpha = 0.1$, $\lambda_f = 10$.

In the ICG network, the loss function is computed as

$$L_D^{ICG} = -E[\log D(I_{gt})] - E[1-\log D(S^3_{img})] \tag{5a}$$

$$L_G^{ICG} = L_{l_1} + \lambda_\alpha L_{adv}^{ICG} + \lambda_s L_{style}^{ICG} + \lambda_p L_{per}^{ICG} \tag{5b}$$

$$L_{adv}^{ICG} = -E[\log D(S^3_{img})] \tag{5c}$$

where L^{ICG}_{style} and L^{ICG}_{per} are respectively the style loss [29] and the perceptual loss [30] based on VGG-19. In this work, we set $\lambda_\alpha = 0.1$, $\lambda_p = 0.1$, $\lambda_s = 25$.

Experimental results and analysis

To verify the performance of our proposed network, we conduct experiments on both simulated and real damage of ancient murals. We compare our network with three state-of-the-art approaches: EC [28], RFR [31], DS-Net [32]. All these approaches use the same datasets for training. In the experiment of simulated damage restoration, we employ peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as evaluation metrics. In the experiment of real damage restoration, we conduct visual comparison on the restored murals. We also conduct the ablation experiment for each module and loss function of our proposed network. All tests are run on a Windows platform. The computer is equipped with an Intel 3.5 GHz CPU and a NVIDIA GeForce RTX 3090 GPU. We implement our model in the PyTorch framework with the Adam optimizer.

We build a mural dataset by manually collecting 3466 high-quality images of different sizes from ancient mural album. We crop the original images into small sub-images as minimal overlap as possible. These sub-images can be roughly categorized into human figures, buildings, rich textures, animals, as shown in Fig. 7. To alleviate the over-fitting problem during the training process, we expand the training data through data augmentation techniques such as random rotation, cropping, mirror flipping, etc. Finally, we choose 10,398 mural sub-images for training, and use 180 deteriorated murals for the experiment of real damage restoration. In the experiment of simulated damage, we employ the irregular masks from the MST dataset [23] and choose the stroke-like masks from the Thangka inpainting dataset [15]. These synthetic masks are very similar to the degradation areas of mural cracks and falling offs. Especially, we choose 10–20% mask rate for the

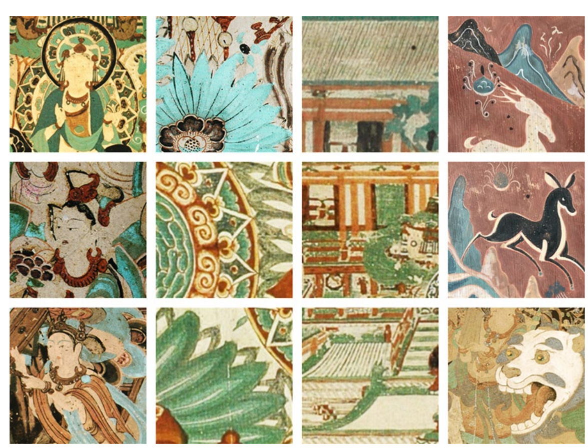

Fig. 7 Examples of different mural data type

irregular masks and 5–10% mask rate for the stroke-like masks in the process of model training and testing.

Experiment on simulated damage

In this subsection, we conduct experiment over the murals with simulated damages to demonstrate the restoration ability of the proposed model. We select several intact (non-damaged) mural images, and employ irregular and stroke-like masks to imitate the deterioration regions.

Figure 8 illustrates three example murals with irregular masks and the restoration results of four comparative methods. It can be seen that the DS-net produces color confusion and semantic discontinuity in the skirt and the pillar in 1st image. RFR causes structure disorders and cannot predict reasonable textures (e.g. 1st and 3rd images). EC fails to restore the missing mural regions, and produces blurred and ghosting effects. Although these damaged murals have lost most of their structure and color information, our model can successfully predict structural information and color details of the missing regions. We also quantitatively evaluate the restoration results by use of the PSNR and SSIM. Table 1 shows the evaluation metrics in this mural restoration test with simulated irregular damages. As can be seen, our network outperforms other three approaches in both PSNR and SSIM metrics. Especially, our model can achieve considerably greater values of SSIM than other comparative approaches. This indicates that our network can restore the missing mural regions with better semantic and visual continuity.

Figure 9 shows five example murals with stroke-like masks and the restoration results of four comparative methods. It can be seen that our proposed model and the DS-net can generate better structural continuity of the missing regions (e.g., 1st image) than RFR and EC. The 2nd mural image is a comparatively difficult test because the missing regions contain complex structures

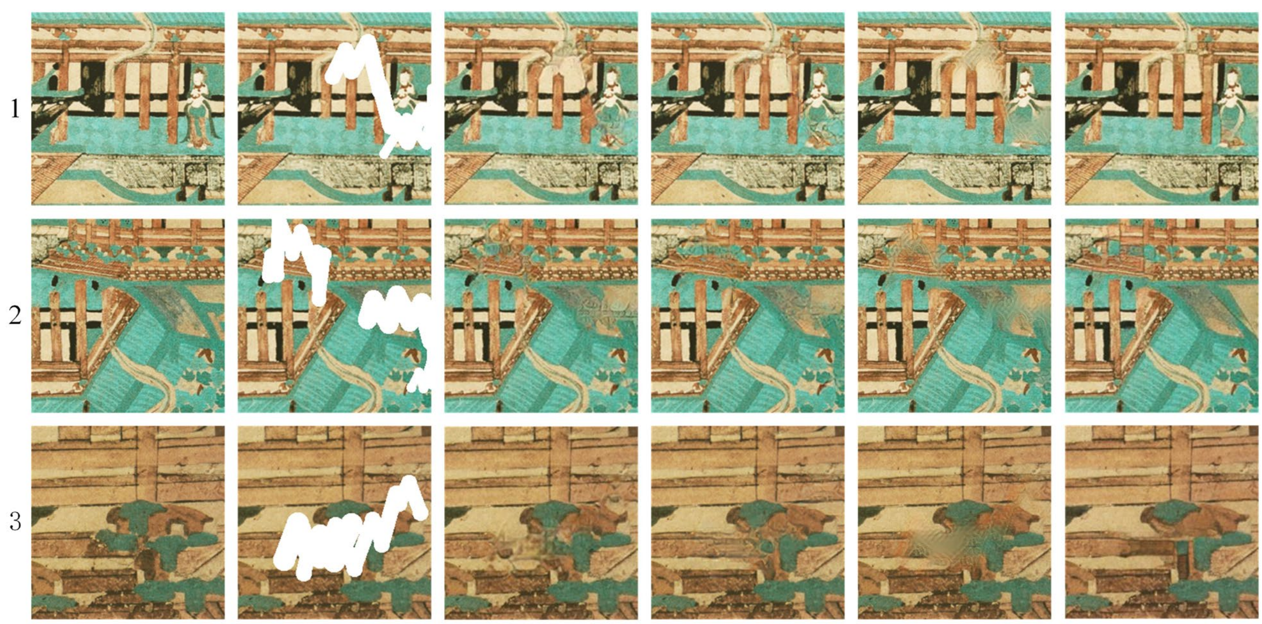

Fig. 8 Restoration results of three murals with irregular masks

Table 1 The PSNR and SSIM values for the mrual restoration of irregular masks

Murals	RFR		DS-net		EC		ours	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
1	23.9058	0.8861	25.0925	0.9085	25.0017	0.9021	**26.6548**	**0.9417**
2	25.0520	0.8179	26.8556	0.8911	26.2696	0.8631	**27.65613**	**0.9286**
3	24.4107	0.8291	26.0728	0.8799	25.5625	0.8710	**26.9239**	**0.9188**

Note: The best result in each column is boldfaced

Fig. 9 Restoration results of eight murals with stroke-like mask

and rich textures (particularly for the pedestal of the little person). It can be observed that RFR produces discontinuous and sharp edges. DS-net generates confused textures and colors. EC cannot recover any textures. By comparison, our model can fill the missing regions with clear structures, textures and colors. The last three images also demonstrate our model's outstanding performance in repairing the structures

Table 2 The PSNR and SSIM values for the mrual restoration of stroke-like masks

Murals	RFR		DS-net		EC		ours	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
1	27.8648	0.8773	29.8823	0.9457	29.3457	0.9230	**33.4027**	**0.9606**
2	28.8939	0.8851	32.3775	0.9637	28.4945	0.9465	**33.0886**	**0.9738**
3	29.1493	0.8707	31.7588	0.9442	29.6621	0.9406	**31.7678**	**0.9608**
4	28.1710	0.8715	31.1147	0.9469	31.3706	0.9582	**34.8384**	**0.9776**
5	27.3813	0.8978	30.5143	0.9596	30.7133	0.9277	**32.8415**	**0.9588**

Note: The best result in each column is boldfaced

and textures of the missing mural regions. Table 2 gives the objective evaluation metrics in the mural restoration test over simulated stroke-like damages. As can be seen, our model achieves higher PSNR and SSIM values than other comparative approaches. It is worthy of stating that our model can obtain considerably better SSIM values. This is just because our model attempts to predict the structure information of the missing regions before restoring the mural contents.

Experiment on real damage

In the experiment of real damaged mural restoration, we choose 180 murals with real damaged or deteriorated regions. We manually mark and label the damaged regions to obtain the masks. Figure 10 shows some damaged murals and their restoration results of four comparative approaches. It can be seen that the DS-net produces some obvious artifacts and unnatural textures when dealing with large damaged areas (e. g., 1st image). DS-net also generates disharmonious colors with its surrounding areas (e. g., 6th, 7th, and 8th images). EC produces discontinuous structures and blurred contents for the missing mural regions (e. g., 1st and 3rd images). RFR performs well in crack restoration, but fails to restore reasonable structures and colors for the large area damages (e. g., 1st and 7th images). As compared to other approaches, our model can generate clear structures, plausible textures, and vivid colors.

It should be noted that real damaged murals have no ground truth images for quantitative evaluation. We invite 20 volunteers to rate the mural restoration results of four comparative approaches based on structural continuity, color consistency, texture clarity and overall effect. We choose 8 mural images as test cases. We assign scores of 1, 2, 3, 4, 5 to five levels of user ratings. Higher scores indicate better evaluation. Figure 11 shows the average scores of 20 volunteers on the restoration results of four approaches. It can be seen that our model obtains all the highest scores in this test.

Ablation study

In this subsection, we conduct the ablation experiment on the proposed network to verify the effects of the core modules and loss functions.

Ablation study of the BranchBlock and the FFC modules

To begin with, we conduct the ablation experiment to study the effects of the BranchBlock and the FFC modules. The SIG network aims to predict plausible structure information of the missing regions. As a core module of SIG, the BranchBlock plays an important role in the extraction and integration of the structure features. We remove the BranchBlock from SIG and keep other modules unchanged. Figure 12 shows the test results of predicting the edge information by the SIG network with/without the BranchBlock. We also provide the ground-truth edge maps that are generated by using the Canny edge detector from the original mural images. As can be seen, the SIG network with the BranchBlock can predict reasonable and consistent structure information (edge map) of the missing areas. By comparison, the network without the BranchBlock produces some implausible and disordered structure information for the damaged murals.

The ICG network attempts to restore the missing contents of the damaged murals by utilizing the predicted structure information from SIG. As a core module of ICG, the FFC module can significantly improve the quality of the restored mural contents. In this test, we remove the FFC from ICG and keep other modules unchanged. Figure 13 shows the test results of restoring the missing contents by the ICG network with/without the FFC module. We also provide the ablation results by removing both the FFC and the BranchBlock from our network. It can be seen that the ICG network without the FFC module produces some ghosting artifacts and inconsistent colors. When we remove the FFC and the BranckBlock simultaneously, the network will produce worse restoration results for the missing mural regions. By comparison, our complete network with the FFC and BranchBlock modules can generate plausible structures,

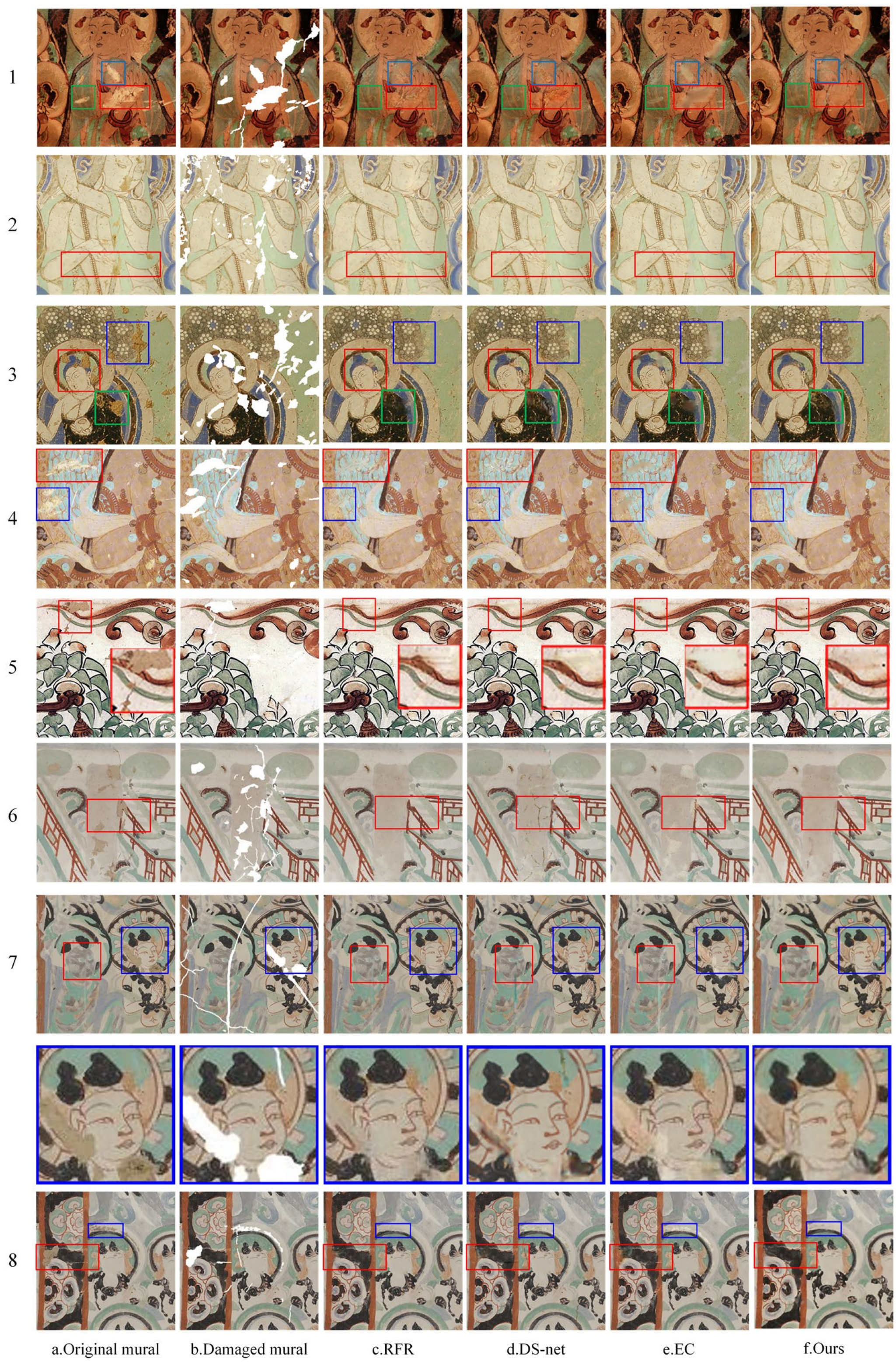

Fig. 10 Comparison of the restoration results from real damage

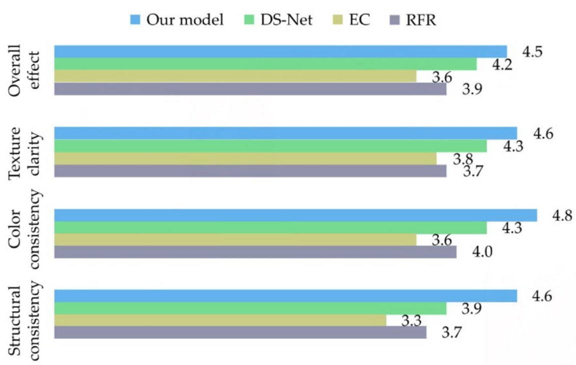

Fig. 11 The average scores of 20 volunteers on the restoration results in real damage experiment

vivid colors, consistent textures for the missing mural regions.

We also use the PSNR and SSIM metrics to evaluate the results of the ablation experiment. Table 3 gives the PSNR and SSIM values averaged over 151 simulated damaged murals. As can be seen, compared with other three ablation models, our complete network shows considerable improvement in the PSNR and SSIM metrics. This demonstrates that the BranchBlock and the FFC modules play an important role in the restoration of the damaged murals.

Ablation study of the loss functions

In the following test, we conduct an ablation study on the loss functions so as to analyze the effects of them. We remove the loss functions one by one, and obtain five different ablation strategies (Ablation1, 2, 3, 4, 5) that are given in Table 4. The symbol "–" denotes the "remove" opration.

Figure 14 provides the visual comparison of all five ablation strategies and our proposed model. In each group of comparison, our model is compared to an ablation strategy that removes a certain loss function. It can be seen that each loss function has obvious improvement on the quality of the restored murals. When a certain loss function is removed from the model, the restored murals will appear obvious degradation such as disordered structures, blurred textures, and implausible colors.

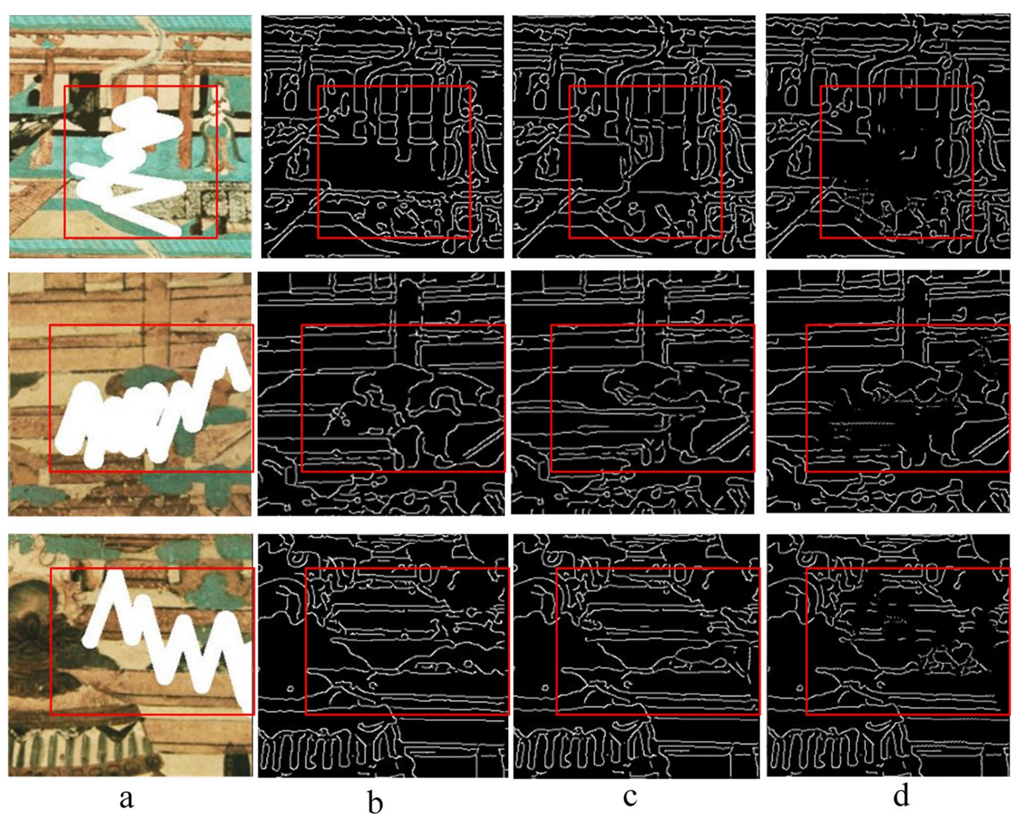

Fig. 12 The ablation test of predicting the edge information by using the BranchBlock and removing the BranchBlock. Note that the red panels indicate the areas of focus for comparison. **a** Masked mural. **b** Ground-truth edge map. **c** Network with Branchblock. **d** Network without Branchblock

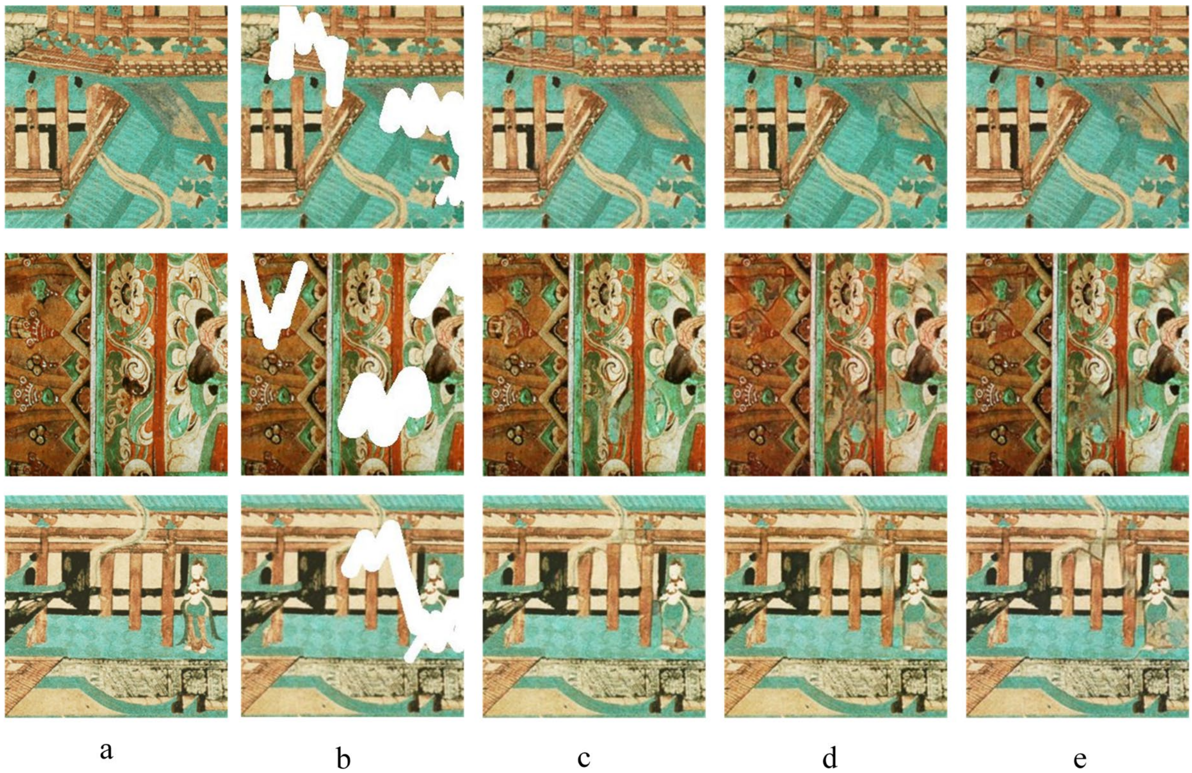

Fig. 13 The ablation test of restoring the missing contents. **a** Original mural. **b** Maked mural. **c** Complete network. **d** Network without FFC. **e** Network without FFC and BranchBlock

Table 3 The objective evaluation metrics of our complete network and three ablation models

Network combination	PSNR/dB	SSIM
Complete network	27.02132	0.931
Network without FFC	25.3842	0.8905
Network without BranchBlock	25.6012	0.8931
Network without BranchBlock and FFC	25.1611	0.8623

Table 4 Different loss strategy in the ablation study

Loss strategy	ELMG LOSS	ICG loss
Ablation1	$L_G^{ELMG} - L_{fm}$	L_G^{ICG}
Ablation2	$L_G^{ELMG} - \sum_i^3 [S_{img}^i - I_{gt}]$	L_G^{ICG}
Ablation3	L_G^{ELMG}	$L_G^{ICG} - L_{l_1}^{ICG}$
Ablation4	L_G^{ELMG}	$L_G^{ICG} - L_{per}^{ICG}$
Ablation5	L_G^{ELMG}	$L_G^{ICG} - L_{sytle}^{ICG}$

Conclusion

In this paper, we proposed a two-stage ancient mural restoration network, which consists of the structure information generator (SIG) and the image content generator (ICG). Our inspiration for designing this network comes from the process of creating handmade murals. SIG is capable of predicting the structure information of the missing mural regions. ICG can effectively restore the missing mural contents under the guidance of the predicted structure information from SIG. In order to

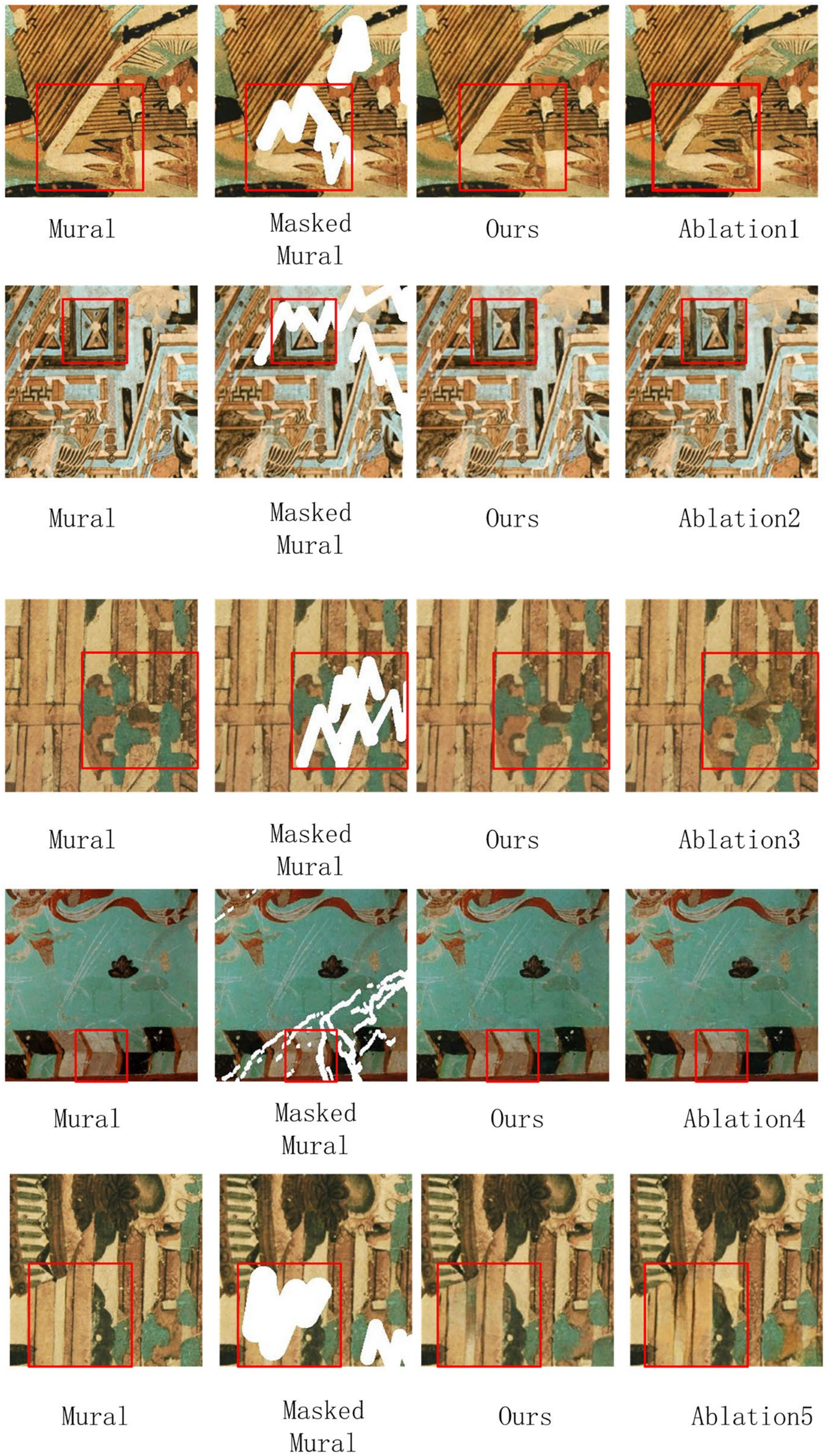

Fig. 14 The ablation test of the loss functions. Note that the red panels indicate the areas of focus for comparison.

predict the structure information for the missing mural regions, we designed an innovative BranchBlock module as the core component of SIG. In order to extract the local and global features from the mural contexts, we introduced a Fast Fourier Convolution (FFC) module as the core component of ICG. The proposed network is performed over both simulated and real damaged murals. The experimental results demonstrate that our model can effectively restore the ancient murals with various damaged regions. As compared with three state-of-the-art approaches, our model can generate more satisfactory results when evaluated by use of visual comparison and objective metrics.

It is worth stating that deep neural network-based image restoration requires a large amount of training data. The network performance is inevitably influenced by the quality of training data. Most of the remaining ancient Chinese murals suffer from varying degrees of diseases such as erosion, flaking, cracks, scratches, sootiness, microorganism corrosion, etc. It is very difficult for us to collect sufficient high-quality (non-diseased) ancient Chinese mural images for the network training. While we expand the training dataset through data augmentation techniques such as rotation, cropping, flipping, etc., the data augmentation may cause information redundancy of the mural image dataset. This will probably affect the generalization ability of the deep neural network. Although our proposed model is superior to existing approaches, it still suffers from the lack of high-quality training mural data. In our future work, we will collect more ancient mural images through field visits across the country. Moreover, we will consider utilizing some intelligent algorithms (e. g., image super-resolution or style transfer based on deep neural networks) to build a large-scale synthetic mural training dataset.

Abbreviations

BBC	Branchblock
FFC	Fast Fourier convolution residual block
GLSI	Global–local features extraction and structural information guidance

Acknowledgements

None.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This research was supported by the National Natural Science Foundation of China (Grant Nos. 62166048, 61263048), by the Applied Basic Research Project of Yunnan Province (Grant No. 2018FB102).

Availability of data and materials

The datasets used and/or analyzed in the current study are available from the corresponding author by reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 4 September 2023 Accepted: 2 December 2023
Published online: 15 December 2023

References

1. Guo D, Liang Y. Research on modeling characteristics and composition forms of Dunhuang mural art in Tang Dynasty: Research Institute of Management Science and Industrial Engineering. In: Proceedings of 2017 2nd international conference on education, sports, arts and management engineering (ICESAME 2017). Atlantis Press; 2017. 4. (**in Chinese with an English abstract**).
2. Liang Y, Guo D. Research on the color representation of Dunhuang mural art. In: Proceedings of the 2017 2nd international conference on education, sports, arts and management engineering. 2017 (**in Chinese with an English abstract**).
3. Bertalmio M, Sapiro G, Caselles V, Ballester C. Image inpainting. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques; 2000. p. 417–24.
4. Cheng Y, Ai Y, Guo H. Inpainting algorithm for Dunhuang mural based on improved curvature-driven diffusion model. J Comput-Aid Des Comput Graph. 2020;32(05):787–96 (**in Chinese with an English abstract**).
5. Criminisi A, Perez P, Toyama K. Object removal by exemplar-based inpainting. In: 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. Proceedings, vol 2. p. II–II.
6. Li C, Wang H, Wu M, Pan S. Automatic recognition and virtual restoration of mud spot disease of Tang dynasty tomb murals image. Comput Eng Appl. 2016;52(15):233–6 (**in Chinese with an English abstract**).
7. Yang X, Wang S. Dunhuang mural inpainting in intricate disrepaired region based on improvement of priority algorithm. J Comput-Aid Des Comput Graph. 2011;23(2):284–9 (**in Chinese with an English abstract**).
8. Jiao L, Wang W, Li B, Zhao Q. Wutai mountain mural inpainting based on improved block matching algorithm. Comput Aid Design Comput Graph. 2019;31(01):118–25 (**in Chinese with an English abstract**).
9. Shen J, Wang H, Wu M, Yang W. Tang Dynasty tomb murals inpainting algorithm of MCA decomposition. J Front Comput Sci Technol. 2017;11(11):1826–36.
10. Wang H, Li Q, Jia S. A global and local feature weighted method for ancient murals inpainting. Int J Mach Learn Cybern. 2020;11:1197–216.
11. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
12. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Communications of the ACM 63.11 (2020): 139–144
13. Cao J, Zhang Z, Zhao A, Cui Y, Zhang Q. Ancient mural restoration based on a modified generative adversarial network. Herit Sci. 2020;8:1–14.
14. Wang H, Li Q, Jia S. A global and local feature weighted method for ancient murals inpainting. Int J Mach Learn Cybern. 2020;11:1197–216.
15. Wang N, Wang W, Hu W, Fenster A, Li S. Thanka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. IEEE Trans Image Process. 2021;30:3720–33.
16. Li L, Zou Q, Zhang F, Chen L, Song C, Wang X. Line drawing guided progressive inpainting of mural damages. In: 2022IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 2788–97. 1, 2, 3.1. arXiv preprint arXiv:2211.06649.
17. Ciortan IM, George S, Hardeberg JY. Colour-balanced edge-guided digital inpainting: applications on artworks. Sensors. 2021;21(6):2091.
18. Lv C, Li Z, Shen Y, Li J, Zheng J. SeparaFill: two generators connected mural image restoration based on generative adversarial network with skip connect. Herit Sci. 2022;10(1):135.
19. Schmidt A, Madhu P, Maier A, Christlein V, Kosti R. ARIN: adaptive resampling and instance normalization for robust blind inpainting of Dunhuang cave paintings. In: 2022 Eleventh international conference on image processing theory, tools and applications (IPTA); 2022. IEEE. p. 1–6.

20. Yu T, Lin C, Zhang S, You S, Ding X, Wu J, Zhang J. End-to-end partial convolutions neural networks for Dunhuang grottoes wall-painting restoration. In: Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019; p. 1447–55.
21. Wang H, Li Q, Zou Q. Inpainting of Dunhuang murals by sparsely modeling the texture similarity and structure continuity. J Comput Cult Herit. 2019;12(3):1–21.
22. Zhou Z, Liu X, Shang J, Huang J, Li Z, Jia H. Inpainting digital Dunhuang murals with structure-guided deep network. J Comput Cult Herit. 2022;15(4):1–25.
23. Huang K, Wang Y, Zhou Z, Ding T, Gao S, Ma Y. Learning to parse wireframes in images of man-made environments. In: 2018 IEEE/CVF conference on computer vision and pattern recognition; 2018. p. 626-35.
24. Xue N, Wu T, Bai S, Wang F, Xia G, Zhang L, Torr PH. Holistically-attracted wireframe parsing. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2020. p. 2788-97.
25. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS. Free-form image inpainting with gated convolution. In: 2019 IEEE/CVF international conference on computer vision (ICCV); 2019. p. 4471–80.
26. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018; p. 4510–20.
27. Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, Silvestrov A. Resolution-robust large mask inpainting with Fourier convolutions. In: 2022 IEEE/CVF winter conference on applications of computer vision (WACV) 2022; p. 2149–59.
28. Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M. EdgeConnect: structure guided image inpainting using edge prediction. IEEE/CVF international conference on computer vision workshop (ICCVW). 2019;2019:3265–74.
29. Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition 3.4; 2016. p. 2414–23
30. Johnson J, Alahi A, Li F. Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. 3.4. Springer; 2016. p. 694–711
31. Li J, Wang N, Zhang L, Du B, Tao D. Recurrent feature reasoning for image inpainting. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2020. p. 7760–68.
32. Wang N, Wang N, Zhang Y, Zhang L. Dynamic selection network for image inpainting. IEEE Trans Image Process. 2021;30:1784–98.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.