

RESEARCH

Open Access



Dunhuang murals image restoration method based on generative adversarial network

Hui Ren^{1,2,3}, Ke Sun^{1,2,3*}, Fanhua Zhao^{1,2,3} and Xian Zhu^{1,2,3}

Abstract

Murals are an important part of China's cultural heritage. After more than a 1000 years of exposure to the sun and wind, most of these ancient murals have become mottled, with damage such as cracking, mold, and even large-scale detachment. It is an urgent work to restore these damaged murals. The technique of digital restoration of mural images refers to the reconstruction of structures and textures to virtually fill in the damaged areas of the image. Existing digital restoration methods have the problems of incomplete restoration and distortion of local details. In this paper, we propose a generative adversarial network model combining a parallel dual convolutional feature extraction depth generator and a ternary heterogeneous joint discriminator. The generator network is designed with the mechanism of parallel extraction of image features by vanilla convolution and dilated convolution, capturing multi-scale features simultaneously, and reasonable parameter settings reduce the loss of image information. A pixel-level discriminator is proposed to identify the pixel-level defects of the captured image, and its joint global discriminator and local discriminator discriminate the generated image at different levels and granularities. In this paper, we create the Dunhuang murals dataset and validate our method on this dataset, and the experimental results show that the method of this paper has an overall improvement in the evaluation metrics of PSNR and SSIM compared with the comparative methods. The restored resultant image is more in line with the subjective vision of human beings, which achieves the effective restoration of mural images.

Keywords Murals restoration, Generative adversarial network, Binary convolution mechanism, Ternary joint discriminator

Introduction

According to Ji Xianlin, a master of national education, there are only four cultures in the world with a long history and a self-proclaimed system: Chinese, Indian, Greek and Islamic, and there is only one place where these four cultures converge, and that is Dunhuang. According to data records, in the fourth–fourteenth century AD, with

the development of Buddhism, ancient artists in the Dunhuang area built a large number of Buddhist caves, of which Mogao Grottoes is the largest, still preserved 735 grottoes, Dunhuang Grottoes mural is an important part of the Dunhuang Grottoes art, the existing mural paintings of about 45,000 m², the largest paintings of more than 40 m² [1]. In different historical periods, Dunhuang murals have different modeling characteristics, content subjects, and painting art styles. Mural paintings are colorful and can be roughly divided into several types, such as venerable statues, story paintings, sutra paintings, Buddhist history paintings, portraits of feeders, costume paintings, decorative patterns, and other paintings. Mural paintings are a combination of ancient art and culture, recording the life of ancient people, and showing the ancient people's ideology, culture religious

*Correspondence:

Ke Sun

sun_ke0610@163.com

¹ School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

² Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture and Tourism, Beijing 100024, China

³ Beijing Key Laboratory of Modern Entertainment Technology, Beijing 100024, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

beliefs, and other information. As the world's largest existing ancient mural resources, murals are the treasure of our country and even the world art hall. However, due to the long history of murals, the fragility of the materials used, and the impact of climate influence, water erosion, weathering, wind erosion, glacial erosion, and other natural and man-made factors for a long time, resulting in most the exquisite murals have appeared to armoring, shedding, discoloration, fading, chalking, cracking and other diseases, which seriously affects the protection and inheritance of murals [2]. Dunhuang Research Institute for Protection of experts in the investigation and analysis, Dunhuang murals used in pigments containing lead pigments easy to discolor, long-term light will also form part of the pigment discoloration. There are human factors caused by discoloration, such as the 156th cave due to the cave had been lived in, in the caves set up a stove fire and the formation of smoky discoloration. Because the Mogao Caves are located in the desert Gobi, by the influence of wind and sand, the surface layer of the murals flaking, and the content is missing, pigment shedding, pale phenomenon. As shown in Fig. 1 for the Dunhuang murals in the typical five kinds of disease example. The current status of the mural diseases shown in Fig. 1 are cracking, peeling, fading, crazing and discoloration from left to right.

As precious cultural heritage, murals carry rich historical, cultural, and artistic values, and it is an urgent task to protect and pass on these valuable cultural heritages by restoring the images of damaged murals. Manual restoration of mural images takes a lot of time and labor costs, and the restoration of large murals may take months or even years and requires well-trained professional restorers, which increases the time and cost of restoration. Secondly, in the process of manual restoration, restorers may need to make minor modifications and smears to achieve image restoration. However, such modifications may result in the loss of original information or further damage to the murals themselves. Especially for severely damaged murals, manual restoration may have irreversible effects.

As computer technology continues to advance, digital image restoration technology has ushered in a new era. Digital image restoration technology uses complex

algorithms and image processing techniques to restore, repair, and even reconstruct damaged images. Compared to traditional manual restoration, digital image restoration is safer and more reliable, avoiding further physical damage. Secondly, digital restoration uses computer technology and image processing algorithms to enable fast and accurate restoration of mural images, and the reconstruction of damaged areas through algorithms to accurately restore the original patterns and details, greatly improving the efficiency of the restoration and the accuracy of the restoration results. The key point is that the whole process is reversible, is a virtual simulation technology, the operation object is not the entity itself, and the algorithm iteration process can be back to the original image at any time to adjust and modify. This reversibility makes the restoration process more flexible and controllable. The use of digital image restoration technology to achieve virtual restoration and protection of Dunhuang murals is currently one of the important research hotspots in the field of image processing and computer vision.

The current image digital restoration methods mainly include the traditional restoration method based on partial differential equations, the traditional restoration method based on samples, and the restoration method based on deep learning.

The idea of restoration based on partial differential equations originates from imitating the process of manual image restoration by art restorers. By using partial differential equations in mathematics or physics to achieve image restoration by smoothly propagating the pixels of known regions of the image into the missing regions based on the edge information of the region to be restored, and estimating the direction of the contour lines from thick to thin. Rathish Kumar [3] proposed to use the image's Hessian matrix's L2 distance as a regular term constraint and proposed fourth-order linear PDE equations to guide grey scale image restoration, which also showed better restoration results. The image restoration method based on partial differential equations is suitable for single restoration, such as image texture or structure restoration only; the center is complete and only the boundary is restored; the broken area is curved and restoration is done along the curve; the image is the



Fig. 1 Mural diseases icon

grey scale and the color information is simple; and the area of the broken area is small and dispersed. In these cases, the reference information transition is natural and the connection is smooth, which has a better repair effect. However, the image restoration method based on partial differential equations will lead to a long restoration time and information loss caused by the propagation process, which makes some areas blurred and unclear.

Sample-based image restoration methods calculate and search for samples with the highest similarity between the missing regions of the damaged image and the known regions, and then copy and paste them into the missing regions to repair the damaged image. This type of method is representative of the Criminisi algorithm [4], repair ideas copy structure and texture information for the filling of the missing image, but its use of the sample block similarity calculation function is not stable, the priority of the region filling order is easy to confuse, inappropriate filling order makes the repaired image to form a repair error, mismatch, the image as a whole is not coordinated problem. In the repair method based on sample matching, the filling order is defined by the priority function, which combines two parameters: confidence term and data term. Subsequent researchers have proposed different improvements to the padding priority problem, Ouattara et al. [5] redefined the priority function and the updating of the confidence term to slow down the rate of decline of the confidence term and prevent it from approaching zero. Experimental results show that the Criminisi-based improved method provides better results both in terms of visual effect and PSNR and SSIM. Guo et al. [6] added the edge term, improved the prioritization algorithm and made a reasonable ordering of the image-filling sequence based on Criminisi's algorithm. overcome the shortcomings of discontinuity in the restored image and improve the quality of restoration of broken images. Li et al. [7] considered the continuity of structural and color information between the restored block and its neighboring blocks, and when searching for the matching block of the restored block, the similarity between the two blocks is calculated using the sum of squared deviations (SSD), which improves the accuracy of repair block priority calculation. Sample matching based image restoration methods can generate high quality restoration results for image restoration with simple texture structure, but it takes a lot of time to calculate the similarity between samples.

Most of the research on traditional restoration algorithms for ancient mural images is based on the idea of Criminisi sample matching to expand the research. In order to solve the matching error phenomenon, inconsistency of image structure information, and inaccurate matching criteria existing in the criminisi algorithm in

thangka image restoration, Yao [8] proposed a thangka image structure information based restoration algorithm. It introduces the correlation between the block to be restored and its neighboring blocks as a priority calculation. They improve the size selection based on paradigms and automatically adjusts the adaptive patch size based on the paradigm-based information change. To solve the problem of mismatching, the structural information of the thangka image and the color of the Euclidean distance are combined as a new matching criterion. The experimental results show that the method proposed in this paper significantly reduces the mismatching phenomenon for thangka images, and the structure of thangka images is more smooth and smoother than the methods in the comparative literature. Cao [9] proposed an adaptive sample block and local search algorithm based on Criminisi algorithm to achieve virtual restoration of broken regions for the problem of shedding murals in Kaihua Temple of Song Dynasty. The method is designed on the basis of improving the image filling order and adopting a local search strategy to improve the matching efficiency. The structure tensor is introduced, the data items are defined as the eigenvalues of the murals' compositions to form the priority computation function, while the average correlation of the structure tensor adaptively selects the sample block size. Experiments verified that the restored image by this method conforms to the image composition characteristics in terms of visual effect. Liu Yicheng [10] proposed an improved mural image restoration method based on Criminisi algorithm for the information and structural characteristics of Yunnan murals. By discarding the fixed mode of the original algorithm to select the sample block size and adopting the adaptive mode of selection, which makes the sample block able to be dynamically adjusted to the size of the sample block. The experimental results show that the improved algorithm improves the restoration efficiency while ensuring the accuracy of the restoration. Rakhi Mol et al. [11] proposed a digital reconstruction model for ancient murals based on dynamic mask generation and an extended sample region-filling algorithm. The authors implemented the broken region detection interest technique, which can do the region identification and automatically generate the corresponding mask image, the mask generation will mark the broken region pixels as 1 and the intact region pixels as 0. And then analyze the pixel values around the 1-pixel region to find the matching samples to fill the broken region. Traditional image restoration methods are challenging to repair damaged images with complex structure and texture because they are unable to obtain the high-level semantic information of the image, and

there are problems such as semantic errors and edge faults, which make it challenging to apply in practical restoration applications.

Since the birth of Generative Adversarial Networks in 2014, GANs and their network structure variants have rapidly become powerful tools for generating realistic and diverse data in various fields. Yan et al. [12] learned Peking Opera face features based on an improved style generative adversarial network StyleGan2 to generate face images with more diverse styles. In February 2018, published in the Massachusetts Science and Technology Review's "Top Ten Global Breakthrough Technologies List" was ranked among the top [13]. With the development of deep learning, many image restoration methods based on deep convolutional neural networks and generative adversarial networks have demonstrated better restoration capabilities on publicly available natural image datasets.

To solve the problem that the local and overall content of the image generated by the generator under a single discriminator is not coherent with the overall content, and there are content differences with the original image locally, Iizuka [14] proposed a globally and locally consistent image restoration method using two auxiliary contextual discriminators for training. The global discriminator is responsible for the overall consistency, while the local discriminator is only responsible for a small region centered in the restored region, to ensure the local consistency of the generated patches. But the design mechanism and network structure of the local discriminator proposed in the paper can only be used to solve the rule masking. Isola et al. [15] proposed PatchGAN as the discriminator of the CycleGAN network, and the design intention of PatchGAN and the local discriminator of the literature [14] are expected to focus on the local information. The difference is that PatchGAN is that the whole image is divided into N pieces of input, each piece is fed into the discriminator for cross-entropy loss calculation, and finally, the average loss of N pieces is calculated as the most final loss.

In about a year before and after 2017, the actual repair effect of the image complementary repair network algorithm is not ideal, and the research is relatively limited. Most of the broken regions, in reality, are irregular and randomly distributed. When the standard convolutional network is used, which uses the average of the effective pixels and the missing parts as the filler, making the filler lacks texture information. So it is prone to artifacts such as chromatic aberration and blurriness, which seriously affects the visual sensation. On the other hand, it restricts the filling repair of irregular region information. Liu [16] propose a method for image restoration, whose use of Partial Convolution can robustly deal with holes of any

shape, size, location, and distance from the image boundary. Guo et al. [17] follow the idea of restoration of Partial Convolution and propose a full-resolution residual network FRRN network model. They achieve progressive step-by-step restoration of image details and structure through a mask update mechanism. Yu et al. [18] proposed a Gated Convolution layer, which improves the mask update mechanism of partial convolutional networks, and solves the problem of traditional convolution to treat all input pixels as valid pixels by providing a learnable dynamic feature selection mechanism for each channel at each spatial location of all layers. They achieve satisfactory visual results in the repair of irregular broken regions.

In deep convolutional neural networks, shallow networks are easy to capture the low-level features of the image such as texture, and deep networks will pay more attention to the overall information, structure, and content semantics of the image. The model based on the attention mechanism can effectively combine the shallow network and the deep features to achieve the restoration results with consistent texture and reasonable structure and content. Liu et al. [19] proposed a codec for the joint restoration of structural and textural CNN network, using the deep features and shallow features extracted by the encoder to represent the structure and texture of the input image respectively, which are fed into the structure flow and texture flow of the multi-scale restoration module to fill the holes, and the feature maps processed by the structure and texture module are spliced and fed into the decoder network to obtain the restored result image.

Two-stage-based models have been popular for a long time. The two-stage model involves two encoders and decoders, where each encoder's CNN is designed to capture the structure or texture of the image. There are two types of two-stage restoration models, one in which the edge contour restoration is performed first in the first stage and the color filling is performed in the second stage, and the other in which a rough restoration result is obtained in the first stage and the fine restoration is performed in the second stage. Nazeri proposed the EC [20] method, where the inputs to the edge generator in the first stage are the grayscale image of the broken image, a masked image, and the corresponding extracted by Canny's operator. The output is predicted complete boundary information, which guides the second stage of the complete restoration of the content repair network. Yi [21] proposed the CRA model for ultra-high resolution image restoration, and the whole restoration algorithm is a two-stage coarse-to-fine process. In the first stage, the 512×512 high-resolution broken image is downsampled to 256×256 and sent to the coarse restoration network to get the low-resolution image restoration result. Image

repair results and then upsampling to generate high-resolution coarse repair results, sent to the second stage to join the high-frequency residuals for fine repair to get the final repair results map. The experimental results show that in the case of a large missing part and complex contextual environment, such as a natural landscape background mixed with people information, the restoration results will be different from the original image.

In order to explore the intrinsic connection between local spatial components and multiscale feature maps under different sensory fields, Qin et al. [22] proposed MSA-Net multiscale attention network based on the idea of characterizing the details and structure of image content with a multiscale network. It consists of a multiscale attention group and pyramidal multiscale attention units as components, to realize the breakage image from the shallow details to the high-level semantic feature extraction and thus better filling of irregular missing regions. Quan et al. [23] proposed a global and local refinement repair method considering the different sizes of peripheral information required to repair different types of missing regions. Firstly, a rough repair result is obtained from the codec structure, then the local content is refined using a network with a shallower hierarchy and a smaller receptive field, and finally, the global refinement is performed using a network with a deeper hierarchy and a larger receptive field.

Researchers focusing on the restoration of mural images have also begun to use deep learning techniques to explore the restoration effect. Cao et al. [24] proposed a consistency-enhanced Generative Adversarial Network (GAN) model to restore lost murals regions. The method first extracts the deep image features by using the convolution layer of the Fully Convolutional Network (FCN), maps the features to the size of the original image through the inverse convolution, and outputs a restored image. The method requires that the image quality around the masked region to be restored must be high. The method requires that the image quality around the masked area of the murals to be repaired must be high, and it is not effective if the repaired area is fuzzy and the texture complex is missing too much.

Li et al. [25] proposed a generative discriminator network model based on artificial intelligence algorithms for image restoration of murals in Bao'an, Shenzhen. The authors mainly improved the discriminator part by splicing real and damaged mural images together as the input to the discriminator network, obtaining a $30 \times 30 \times 1$ matrix. With emphasizing the value of this matrix by constraining it in the loss function to achieve the purpose of prompting the generator to learn to improve the texture detail information of the generated image, effectively repairing the

damaged murals with point damage and complex texture structure. Wang et al. [26] proposed a mural restoration method based on multi-scale adaptive partial convolution and stroke-like mask, using kernel-level multi-scale adaptive partial convolution for accurately distinguishing between valid pixels and invalid pixels. A parameter-configurable stroke-like mask generation method for simulating and learning stroke-like restoration patterns and implementing a two-stage learning framework based on MapConv Unet and different loss functions. Lv et al. [27] proposed the *SeparateFill* network structure, where the first stage starts with edge contour repair to fill the contour of the broken image, and the second stage uses the content fill network for color filling. Global and local discriminator networks are used to determine whether the image is repaired or not. The method performs well in restoring the line structure of damaged mural images and protecting the integrity of mural images. Li et al. [28] proposed a line-sketch-guided method for progressive restoration of damaged areas of murals, which divides the restoration process into two steps: structural reconstruction and color correction, which are performed by a structural reconstruction network (SRN) and a color correction network (CCN), respectively. In the dataset created by the authors, the images exist in pairs, with a color three-channel image of the murals corresponding to a two-valued line sketch. The input to the structural reconstruction SRN network is the broken color image along with the complete line sketch, which serves as a guarantee of large-scale content authenticity and structural stability and ensures the structural soundness of the reconstructed image in the first stage, while the input to the color correction network CCN is the four-channel image resulting from the combination of the three-channel restoration result map outputted from the first stage with the mask image, which in this stage serves to tell the network the location of the missing regions, while the color correction network only locally adjusts the color of the missing pixels to reduce the negative effects of color deviations and edge jumps. The disadvantage of this model is that the training data needs to be in pairs, and often obtaining a dataset of murals becomes the first hurdle. Deng et al. [29] proposed a structure-guided two-branch model based on generative adversarial networks to repair ancient murals from the perspective of structure-guided filling of lost content with complex structure and diverse patterns, improving the quality of texture and colour restoration of the missing regions of the murals. Aiming at the problem of fading and discolouration of murals images, Xu et al. [30] self-constructed a 1024×1024 large-size Dunhuang murals

dataset by collecting 1236 relatively intact images from “Dunhuang Architectural Studies” and “The Complete Collection of Dunhuang Murals in China”. They proposed the DC-CycleGAN model for mural image colouring, which combines the variable row convolution (DCN), ECANet, ResNet and Cycle Generative Adversarial Network (CycleGAN), using the colour style migration technique applied to colour recovery of murals images. For the restoration of murals in large damaged areas, Liu et al. [31] proposed a multilevel progressive inference network, MPR-Net, which is capable of recursive inference restoration. By restoring the structure of damaged murals globally firstly and further adding detailed textures locally, the multi-scale feature aggregation module can efficiently utilise the inferred features and dynamically select the available features, which adaptively fuses the different scale of rich information, enhancing the ability to select important features. Deep learning captures potential features in the image by self-supervised training on large-scale data, counts the distribution characteristics of pixels, and uses the learned high-dimensional feature mapping to repair the missing regions, and predicts the content of the damaged regions to be filled.

FRRN [17], DeepFillv2 [18], MEDEF [19] and EC [20] are the more classical models in the field of image restoration. FRRN has higher requirements on image quality, and local blurring problems occur in some images after murals restoration; EC consists of an edge generator and an image completion network, which repairs to get the image contour first, and then fills it with colours. But it has insufficient expressive power in the complex areas of blurred lines, and the problem of incomplete restoration occurs in some images after murals restoration. Based on the above problems, the innovations of the method in this paper include:

1. This paper proposes a parallel dual convolutional feature extraction module, which adopts the mechanism of extracting features in parallel with vanilla convolution and dilated convolution. This innovative approach uses reasonable and clever parameter settings to allow us to better capture multi-scale features during image restoration while reducing the loss of image information, which is conducive to achieving a complete restoration of the image.
2. This paper proposes a pixel-level discriminator to discriminate images pixel-by-pixel, which can accurately capture tiny artifacts or defects in images, and thus help to restore the detailed parts of images in a more detailed way. This innovative approach can solve the problem of local detail distortion in image

restoration, improve the quality of image restoration, and achieve clarity of details in restored images.

Method

Cut into the rough shape surface of the grotto walls is very rough gravel, ancient painters first mixed with crushed grass fiber material viscous clay to smooth the surface of the wall, and then a layer of thin and fine mud skin was used to support the murals paintings of the pigment, these support the murals mud layer is called the ground battle layer. Due to its unique production process, compared to natural images, mural images are rich in information and complicated texture details, so the restoration model to fill the missing content to make full use of the known content, network design considerations to reduce the loss of information, at the same time, during image generation, pixel-level prediction is performed to achieve detailed image restoration. Based on the above considerations, this paper proposes a generative adversarial network murals image restoration model consisting of a parallel dual convolutional feature extraction depth generator and a ternary heterogeneous joint discriminator. The overall architecture of the model is shown in Fig. 2, given the original Dunhuang murals image I_{gt} (ground truth), the mask image M (mask), the simulated damaged murals image I_{masked} is the object to be repaired, $I_{masked} = I_{gt} \odot (1 - M)$, \odot denotes the Hadamard product. The damaged mural image is used as the input of the parallel dual-convolution feature extraction depth generation network, and the output is the repair result map. The repair result map is fed into the ternary heterogeneous joint discriminator network composed of a global discriminator network, local discriminator network, and pixel-level discriminator network for discriminating, respectively. The generator and the discriminator fight against each other and the two are constantly iterated and optimized so that the repaired image inpainting result is infinitely close to the original image.

Parallel dual convolutional feature extraction depth generator

For the problem of incomplete restoration of murals restoration results, this paper proposes an improved generator network based on U-Net and ResNet50, which introduces a parallel double convolutional feature extraction mechanism to reduce the loss of image information. The U-Net encoder downsamples for feature extraction, and the decoder upsamples for feature reconstruction. By channel splicing the feature map obtained by downsampling with the upsampling process, the information loss caused by the reduction of

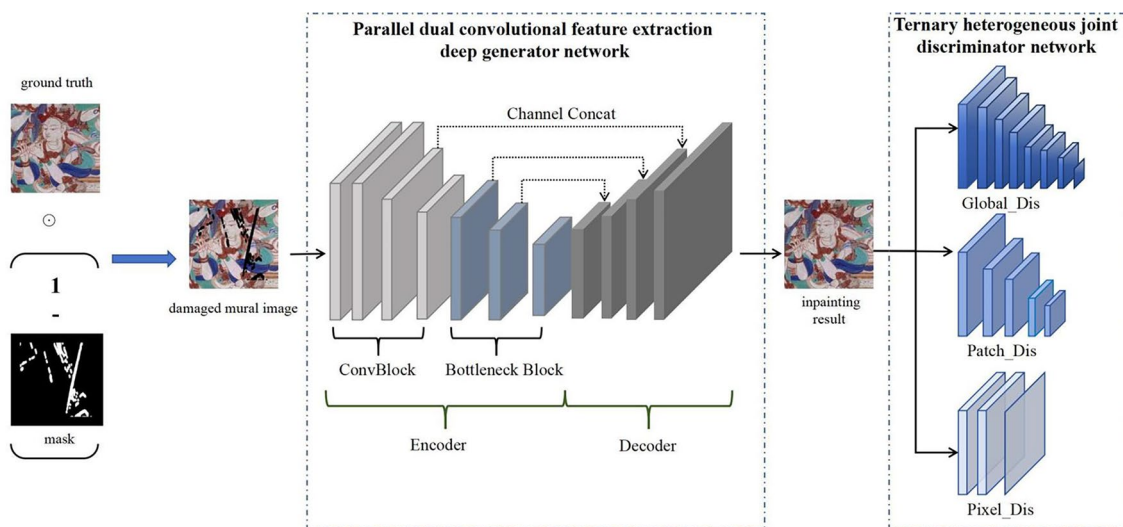


Fig. 2 Ours image restoration model general framework

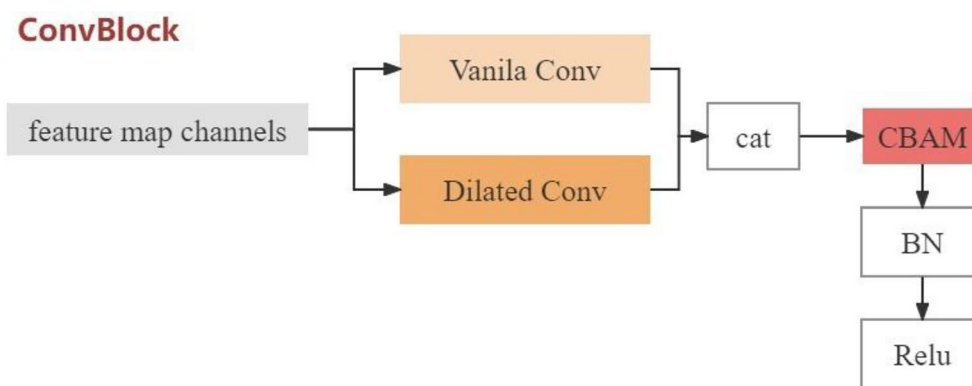


Fig. 3 ConvBlock module

image resolution in the sampling process is reduced. At the same time, Cross-layer connections enable low-level features to be expressed in high-level layers, retaining features extracted from low-level layers, allowing the network to learn more image details. ResNet50 is a deep network composed of multiple residual components Bottleneck dense stacking, with the performance of extracting complex features. The residual structure is designed to suppress gradient disappearance and explosion in deep networks. In this paper, we propose the ConvBlock parallel dual convolutional feature extraction module as shown in Fig. 3, where vanilla convolution and dilated convolution extract features in parallel for the same input feature map, we replace the original ResNet50 stem with a stack of four ConvBlocks. The network structure of the ResNet50 stem is shown in Table 1, the use of a large convolutional kernel of 7×7 increase the receptive field and capture the global

Table 1 ResNet50 stem network architecture

Input	Stage	Layer	Output
$3 \times 256 \times 256$	Stem	$7 \times 7, 64, \text{stride} = 2, p = 3$	$64 \times 128 \times 128$
		BN + RELU	BatchNorm2d (64)
		$3 \times 3, \text{max pool}, \text{stride} = 2$	$64 \times 64 \times 64$

semantic information, but the shallow use of pooling layer reduces the image resolution to a certain extent, resulting in loss of information, which leads to incomplete restoration of the restored image.

By replacing the stem stage of ResNet50 with ConvBlock, the dilated convolution retains the advantage of a large convolution kernel with a large receptive field and keeps the relative spatial position of the pixels unchanged while expanding the receptive field, without losing

resolution. In deep network design, it is customary to use a small convolutional kernel in the low-level network for local information extraction, with the downsampling feature map resolution decreases, each value of the image is mapped to a region block in the previous layer (i.e., corresponding to the region of the receptive field), and with the increase of the depth of the network, the local information is weakened, which in turn combines in the deep network to obtain a higher level of semantic features. In this paper the low-level network tries to image local information captured as much as possible, so $k=3$ and $s=1$ are set in ConvBlock, at the same time, padding is set in order to keep the image edge information intact, and the shallow network extracts the features with the same size as the original image, which effectively reduces the loss of information, and solves the problem of incomplete restoration. CBAM [32] is an attention mechanism for enhancing the performance of convolutional neural networks by combining spatial attention and channel attention. Channel attention allows the network to automatically learn and enhance important features, thus improving the representation of the feature map. Spatial attention helps to capture the importance of different regions in the image, enabling the network to better understand the location and structure of objects. Adding CBAM attention to the ConvBlock block to reconstruct the features extracted by vanilla convolution and dilated convolution helps the model to better understand the semantics of the image. Through experiments, it is observed that the performance of the model is improved by adding the CBAM attention mechanism. CBAM module is shown in Fig. 4, the parameter settings of the generator encoder are shown in Table 2, in which stem1 ~ stem4 are all ConvBlock modules, and the decoder parameter settings are shown in Table 3

Ternary heterogeneous joint discriminator

For the problem of image local detail distortion, this paper proposes the pixel-level discriminator PixelGAN_Dis, and unites the global discriminator GlobalGAN_Dis, and the local discriminator PatchGAN_Dis to form the discriminator part of the generative adversarial network. Unlike the global and local discriminators, the pixel-level discriminator is not only concerned with the general nature of the overall image, but also with the truthfulness of each pixel in the image. Although the overall image looks reasonable, zooming in on local areas will expose defects, and PixelGAN_Dis can help find these pixel-level defects, which is beneficial to generating images that are consistent with real images at the global, local and pixel levels. The GlobalGAN_Dis network parameters are shown in Table 4, using a deeper network structure to capture the overall features of the image more comprehensively, the output is a 1×1 sized feature map, which is a summary of the global information of the image, the discriminator uses a value to judge the authenticity of the whole image from the semantics of the image, and the output is 1/0 to represent that the image is the original murals image or restored image. The PatchGAN_Dis network parameters are shown in Table 5, and the output is an $N \times N$ matrix, the value 1/0 of each point in this matrix measures the truth or falsity of its corresponding region block (representing a sensory field in the original image), focusing on the local information of the image. The PixelGAN_Dis network parameters are shown in Table 6, and the image is discriminated by a small number of convolutional layers and activation functions, in each layer. The convolution kernel and step size of each layer are set to 1, so we only focuses on one pixel. The output feature map size is the same as the original image size. Each pixel is regarded as an independent discriminant unit, its discrimination results are not affected by other pixels. By judging each pixel independently, the discriminator can more finely distinguish the subtle differences between the

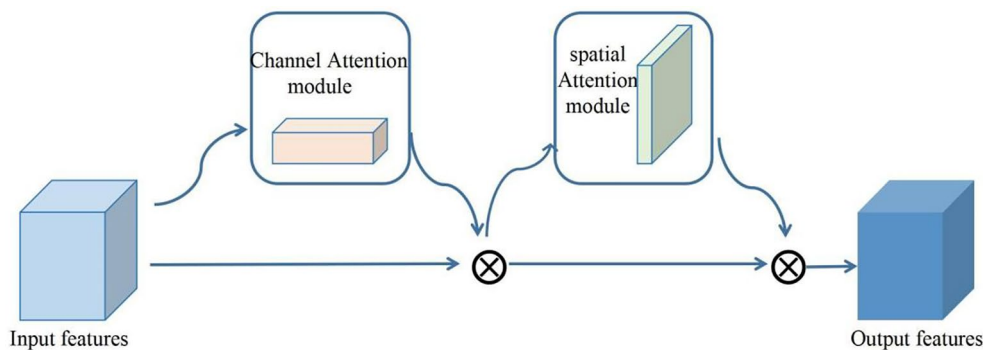


Fig. 4 CBAM module

Table 2 Generator network encoder stage parameter

	Layer	Input channel	Output channel	Input size	k	s	p	d	Output size	
Encoder										
Down	Stem1	conv1	3	8	256×256	3	1	1	1	256×256
		conv1_1	3	8	256×256	3	1	2	2	256×256
		cbam_atten1	16	16						256×256
		BN(16)+Relu()								
	Stem2	conv2	16	32	256×256	3	1	1	1	256×256
		conv2_1	16	32	256×256	3	1	2	2	256×256
		cbam_atten2	64	64						
		BN(64)+Relu()								
	Stem3	conv3	64	32	256×256	3	2	1	1	128×128
		conv3_1	64	32	256×256	3	2	2	2	128×128
		cbam_atten3	64	64						
		BN(64)+Relu()								
Stem4	conv4	64	32	128×128	3	2	1	1	64×64	
	conv4_1	64	32	128×128	3	2	2	2	64×64	
	cbam_atten4	64	64							
	BN(64)+Relu()									
Stage1	layer5	64	256		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$			64×64		
	CBAM(256)									
Stage2	layer6	256	512		$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$			32×32		
	CBAM(512)									
Stage3	layer7	512	1024		$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$			16×16		
	CBAM(1024)									

Table 3 Generator network decoder stage parameter

Decoder									
	Sampling	Input channel	Output channel	Input size	k	s	p	Output size	BN/IN
Up									
Decoder1	Pixel shuffle	1024	1024	32×32	3	1	1	32×32	IN
Decoder2		1024	512	32×32	3	1	1	64×64	IN
Decoder3		512	256	64×64	3	1	1	128×128	IN
Decoder4		256	64	128×128	3	1	1	256×256	IN
Full		64	33	256×256	3	1	1	256×256	IN
		32	3	256×256	1	1	1	256×256	IN

generated image and the real image, thereby providing a more targeted feedback signal and helping the generator model learn more accurate features and texture details, resulting in more realistic image results.

Batch Normalization and Instance Normalization are indispensable regularization methods in network training. BN solves batch information statistics and balances the differences in data distribution between samples. IN

Table 4 GlobalGAN_Dis parameter

Layer	Input channel	Output channel	Input size	k	s	p	Output size	BN/IN
Conv1	3	32	256×256	4	2	1	128×128	IN (32)
Conv2	32	64	128×128	4	2	1	64×64	IN (64)
Conv3	64	128	64×64	4	2	1	32×32	IN (128)
Conv4	128	256	32×32	4	2	1	16×16	IN (256)
Conv5	256	128	16×16	4	2	1	8×8	
Conv6	128	64	8×8	4	2	1	4×4	
Conv7	64	32	4×4	4	2	1	2×2	
Conv8	32	1	2×2	4	2	1	1×1	

Table 5 PatchGAN_Dis parameter

Layer	Input channel	Output channel	Input size	k	s	p	Output size	BN/IN
Conv1	3	64	256×256	4	2	1	128×128	
Conv2	64	128	128×128	4	2	1	64×64	BN (128)
Conv3	128	256	64×64	4	2	1	32×32	BN (256)
Conv4	256	512	32×32	4	1	1	31×31	BN (512)
Conv5	512	1	31×31	4	1	1	30×30	

Table 6 PixelGAN_Dis parameter

Layer	Input channel	Output channel	Input size	k	s	p	Output size
Net1	3	64	256×256	1	1	0	256×256
Net2	64	128	256×256	1	1	0	256×256
Net3	128	1	256×256	1	1	0	256×256

counts each sample and solves the difference between the feature map channels of each sample difference. Using normalization can speed up convergence and mitigate overfitting, and can help solve the problem of stalled gradient propagation due to improper initialization of network parameters. GAN is difficult to train, and Batch Normalization of all layers can lead to sample oscillation and model instability. Experiments have found that changing the BN of some layers can cause the network to generate noisy images. BN can handle larger receptive fields, while IN is more suitable for smaller receptive fields. Adding BN to the generator encoder stage and IN to the decoder stage, the combination of the two can better handle multi-scale features. The global discriminator

GlobalGAN_Dis uses IN for conv1–conv4, and the local discriminator PatchGAN_Dis uses BN for conv2–conv4. The combination of the two reduces the variability of data distribution and achieves consistency between training and testing modes.

Loss function

The network model training part includes generative network training and discriminative network training, the generative network tries to generate reasonable and real mural content, the discriminator tries to distinguish whether the image is the original intact image or the restored resultant image, and the two play a game, then the GAN obtains the best result when the formula (1) is satisfied.

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log(x)] + E_{z \sim P_{out}(z)}[\log(1 - D(G(z)))], \quad (1)$$

where x is the original complete murals image, $P_{data}(x)$ is the probability distribution of the input murals image, z denotes the murals image with the mask added, and $P_{out}(z)$ denotes the probability distribution of the restored murals image.

The discriminator part of the model in this paper consists of global, local and pixel-level discriminators, and the output of each discriminator is used as part of the adversarial loss, the loss of the discriminator network is denoted as.

$$\ell_{dis_total} = \alpha_{dis1}\ell_{dis1} + \alpha_{dis2}\ell_{dis2} + \alpha_{dis3}\ell_{dis3}, \quad (2)$$

where ℓ_{dis1} , ℓ_{dis2} , ℓ_{dis3} are the discriminative loss of GlobalGAN_Dis, PatchGAN_Dis and PixelGAN_Dis, α_{dis1} , α_{dis2} , α_{dis3} are their corresponding weights.

The loss function of the generated network consists of adversarial loss ℓ_{adv} , feature matching loss ℓ_{FM} , style loss ℓ_{Style} , perceptual loss ℓ_{perc} and pixel level loss ℓ_1 , expressed as Eq. (3).

$$\ell_{gen_total} = \lambda_{adv}\ell_{adv} + \lambda_{FM}\ell_{FM} + \lambda_{Style}\ell_{Style} + \lambda_{perc}\ell_{perc} + \lambda_{l1}\ell_{l1}. \quad (3)$$

Adversarial loss helps to improve the level of visual realism of the recovered image, style loss tends to correct the stylistic consistency of the high-level structure, and feature matching loss and perceptual loss help to maintain the high-level semantic features of the entire murals. Feature matching loss can be calculated from (4).

$$\ell_{FM} = E \left[\sum_{i=1}^m \frac{1}{N_i} \left\| D^{(i)}(I_{gt}) - D^{(i)}(I_{out}) \right\|_1 \right], \quad (4)$$

m is the number of convolutional layers of the discriminator, N_i is the number of feature maps in i th activation layer, $D^{(i)}$ is the feature map in the i th layer of the discriminator, I_{gt} is the original murals map, I_{out} is the

damaged Dunhuang murals restoration result map. The ℓ_{FM} proposed in this paper consists of the loss ℓ_{FM}^{dis1} computed by the global discriminator extracting features and the loss ℓ_{FM}^{dis3} computed by the pixel-level discriminator extracting features, $\ell_{FM_total} = \lambda_{FM}^{dis1}\ell_{FM}^{dis1} + \lambda_{FM}^{dis3}\ell_{FM}^{dis3}$. The computation of the perceptual loss [33] is expressed as Eq. (5), for the i th pooling layer of the real image and the generated result map for comparison, ψ_i is the feature map of the i th layer of the pre-trained VGG-19 network.

$$\ell_{perc} = E \left[\sum_{i=1}^N \left\| \psi_i(I_{out}) - \psi_i(I_{gt}) \right\|_1 \right]. \quad (5)$$

The style loss [34] ℓ_{Style} is used to calculate the L_1 distance between the Gram matrix of the generated mural image and the real image. It is assumed that the size of the i th layer feature map is $C_i \times H_i \times W_i$, $G_j^\psi(\cdot)$ is the Gram matrix of the size $C_j \times C_j$ constructed by the j th feature map. The style loss is calculated as Eq. (6).

$$\ell_{style} = E \left[\sum_{j=1}^N \left\| G_j^\psi(I_{out}) - G_j^\psi(I_{gt}) \right\|_1 \right]. \quad (6)$$

The ℓ_1 loss function is used to measure the pixel-level difference between the real murals image and the restored resultant image, which can be calculated by Eq. (7).

$$\ell_1 = \|I_{out} - I_{gt}\|_1. \quad (7)$$

It is experimentally verified that better restoration results are achieved when the weighting coefficients of the corresponding loss functions λ_{adv} , α_{dis1} , α_{dis2} , α_{dis3} , ℓ_{perc} , ℓ_{Style} , λ_{l1} are 1, 0.2, 0.4, 0.4, 2, 120 and 1, respectively.

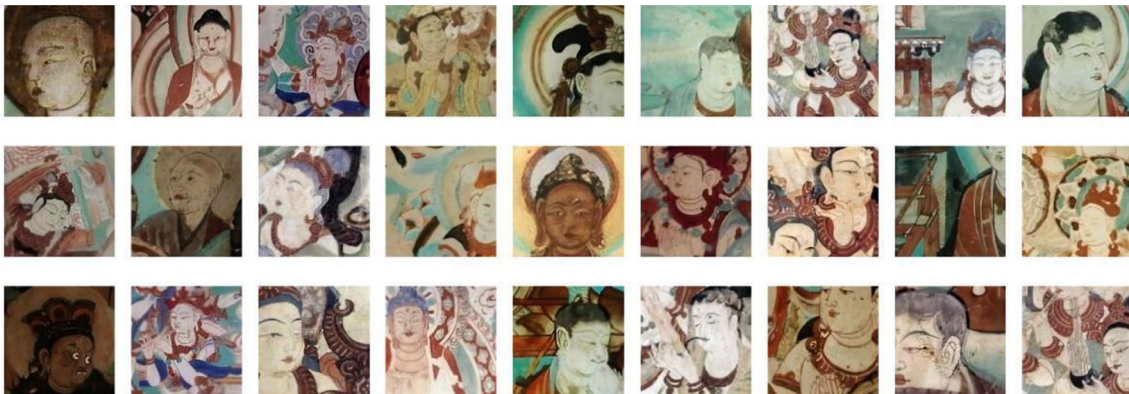


Fig. 5 Examples of images from Dunhuang murals dataset



Fig. 6 Examples of publicly available images of irregularly masked datasets

Results and discussion

Dataset description

In this paper, 21,000 Dunhuang murals images of 256×256 size with relatively complete content and clear presentation effect are collected from the "China Dunhuang murals Collection". The dataset is mainly composed of mural paintings created in the Tang Dynasty, covering caves 009, 014, 045, 054, 085, 112, 154, 420, 290, and 428, and some of the images are shown in Fig. 5. The 21,000 images of murals dataset are randomly divided into 18,000 training sets and 3,000 test sets. When the model is trained, 36,000 training samples are obtained by expanding the data enhancement with horizontal flipping of mural images in order to enhance its robustness and reduce the risk of overfitting.

Due to the irregularity of mural breakage, the simulation of mural breakage state is masked using irregular masks, and the irregular masks are adopted from the public mask dataset [16], and some of the masks are shown in Fig. 6 for example.

Experimental environment and parameter setting

In the experiments, the hardware environment was configured as an NVIDIA A800 SXM4 80 GB with an Intel(R) Xeon(R) Platinum 8358 CPU @2.60 GHz, the models in this paper were implemented with pytorch 1.10.0 and CUDA 11.3 and ran on a CentOS Linux release 7.8.2003 system, and all experiments covered in this paper were conducted in the same environment. The model was trained with an input image size of 256×256 , training epoch = 30, batch size batch_size = 32, using the Adam optimizer with the betas parameter set to 0.5 and 0.999, and the learning rate of both the generator and discriminator set to 0.0002.

Experimental results and analysis

In order to validate the effectiveness of the method proposed in this paper as well as to demonstrate the indispensability of the ConvBlock module and the pixel-level discriminator Pixel_Dis for improving the network model and increasing the absolute accuracy, restoration tests were conducted by adding random masks to a test set of 3000 murals and compared with the classical image restoration models FRRN [17], DeepFillv2 [18], MEDEF [19] and EC [20] for comparison experiments. Meanwhile, ConvBlock block and Pixel_Dis are removed from the complete model in this paper for ablation experiments, respectively. The results of the comparison and ablation experiments are analysed from both subjective and objective evaluation perspectives.

Comparison experiment

As shown in Fig. 7 is a comparison chart of the repair test results of different algorithms, where a is the ground truth, b is the masking occlusion, i.e., the simulated broken murals image, c is the repair result of the EC model, d is the repair result of the FRRN model, e is the repair result of the DeepFillv2 model, f is the repair result of the MEDEF model, and g is the repair result of the model proposed in this paper. When selecting the display legends from the 3000 test set images, the principle of sample diversity is followed, as follows, test set samples 1, 2, and 6 simulate broken areas with small area characteristics, and 3, 4, and 5 simulate broken areas with medium area characteristics, and 7 and 8 simulate relatively large broken areas. At the same time in the random mask form, select its irregular randomness close to the murals "disease" form, such as samples 3, 4 after adding the mask is similar to the mural point-like mold spot

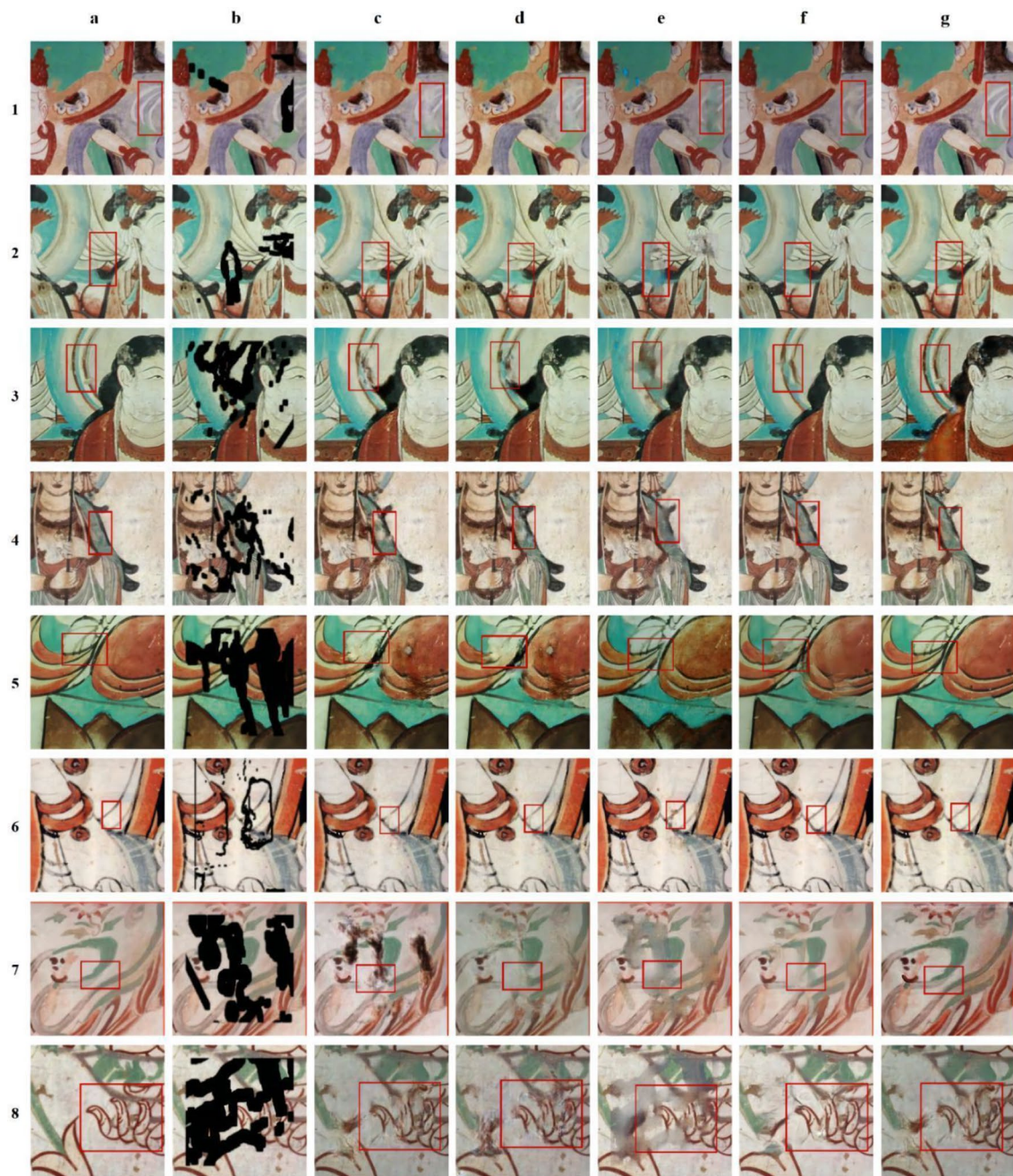


Fig. 7 Comparison results of different algorithms for murals image restoration. **a** Ground truth, **b** Damaged image, **c** EC, **d** FRRN, **e** DeepFillv2, **f** MEDEF and **g** ours

damage, samples 5, 7, 8 after adding the mask is similar to the murals shedding damage, samples 6 after adding the mask is similar to the murals cracked damage.

From the subjective visual point of view, this paper's algorithm and its comparison algorithm show excellent repair performance for small-area breakage, such as sample 2, point-like dispersed breakage, such as sample 4,

and stripe-like breakage, such as sample 6, so that broken murals can be effectively repaired. EC, which is a method based on the idea of repairing the outline first and then filling in the colors, shows blurriness in filling in the areas with a weak sense of the edge of the outline, such as samples 1 and 7. Broken, local texture details are missing phenomena, but for line-like breakages, such

Table 7 Comparison of PSNR value of each algorithm

	1	2	3	4	5	6	7	8
EC	28.89	28.39	24.04	23.15	20.67	21.91	17.58	16.89
FRRN	26.52	28.84	22.54	23.69	20.67	20.42	18.42	15.47
DeepFillv2	27.89	27.93	24.38	19.93	21.56	20.04	18.10	17.51
MEDEF	31.07	27.27	22.62	24.93	20.23	22.18	20.17	17.64
OURS	33.88	29.36	26.64	23.84	22.41	22.18	21.28	17.06

Table 8 Comparison of SSIM value of each algorithm

	1	2	3	4	5	6	7	8
EC	0.997	0.989	0.976	0.879	0.973	0.998	0.863	0.936
FRRN	0.996	0.990	0.983	0.880	0.974	0.998	0.938	0.928
DeepFillv2	0.993	0.894	0.975	0.891	0.936	0.993	0.863	0.898
MEDEF	0.997	0.988	0.991	0.999	0.993	0.999	0.998	0.997
OURS	0.999	0.992	0.998	0.996	0.998	0.998	0.980	0.997

as sample 6, EC plays its advantage. However, for sample 8 when the mask area is large and the distribution is scattered, the repair ability of EC and FRRN has a slight decrease, and the problem of incomplete repair occurs. FRRN uses residual networks to repair the image step-by-step, to achieve the gradual recovery of the details and structure of the image, but there are certain requirements on the quality of the image, and it fails to fully demonstrate its advantages during the restoration of murals, and local blurring phenomenon occurs. When using the DeepFillv2 model to repair broken images, the outline of the region of the mask can be seen in most of the repaired images with artifacts, such as samples 1, 2, and 7. No improvement was also found when the model iteration was increased from 20 to 30 epochs. In the MEDEF model, the shallow image structure information and the deep texture information were fed into the structure flow encoder and texture flow encoder for processing, it can be seen that the restoration effect map for structurally distinct mural images like samples 6 and 8 approximates the original image. Tables 7 and 8 show the results calculated using the objective evaluation metrics PSNR and SSIM, and the values obtained from the calculation of the restored images by the MEDEF method are higher. Although MEDEF is an end-to-end input in the network structure diagram, in the actual processing, it is a two-stage model, and the input data requires the mural image, the mask image, and the corresponding structure map of the mural image (which was found to be equivalent to the denoising and smoothing of the original murals image when the comparative experimental reproduction was performed, which resulted in a cumbersome data

preprocessing process and high computational overheads). The end-to-end method proposed in this paper fills in the missing regions completely during the restoration test, and the local details are clear and consistent with human subjective vision.

Ablation experiment

As shown in Fig. 8 Comparative results of ablation experiments, a Ground truth, b Damaged image,, where a is the ground truth, b is the mask occlusion, which is the simulated broken mural image, c is the restoration result of this paper's method in removing the ConvBlock, d is the restoration result of this paper's method in removing the Pixel_Dis, and e is the restoration result of the model proposed in this paper (the complete method). In the sample display, the same mural images as in the comparative experiment were selected, and the selection rules are the same as in the above section. It can be observed that, for samples 5, 7, and 8 repair test, removing the convBlock module appeared to repair the incomplete phenomenon; and removing the Pixel_Dis, in the samples shown in the overall image coordination, completed repair. However, when narrowing down to observe the local area of the local blurring, details of the stripe outline is not in place, do not protrude from the obvious problem. From Tables 9 and 10 it is noted that if any module is removed, the corresponding PSNR and SSIM values will increase, which shows the importance of ConvBlock and Pixel_Dis.

The generator designed in this paper adopts the ConvBlock module, applies the dilated convolution and vanilla convolution to extract image features in parallel,

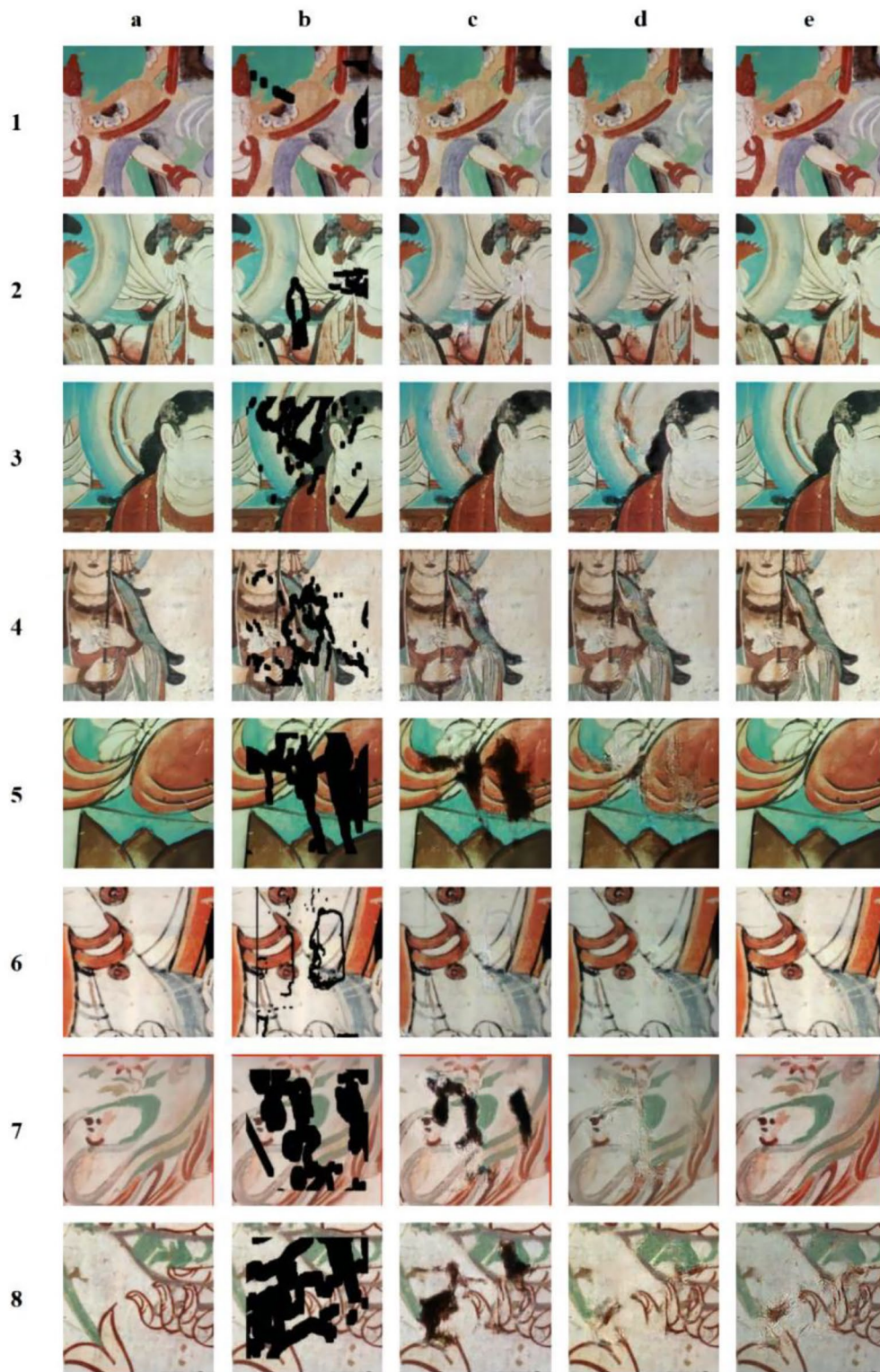


Fig. 8 Comparative results of ablation experiments, **a** Ground truth, **b** Damaged image, **c** w/oCB (without ConvBlock), **d** w/oPixel_Dis (without Pixel_Dis), **e** ours (complete method)

Table 9 Comparison of PSNR of ablation experimental restoration results

	1	2	3	4	5	6	7	8
w/oCB	22.86	24.87	22.09	19.93	17.56	20.04	14.72	12.24
w/oPixel_Dis	26.88	26.65	22.67	19.52	18.87	20.84	15.63	15.84
OURS	33.88	29.36	26.64	23.84	22.41	22.18	21.28	17.06

Table 10 Comparison of SSIM of ablation experimental restoration results

	1	2	3	4	5	6	7	8
w/oCB	0.993	0.984	0.975	0.891	0.836	0.993	0.863	0.890
w/oPixel_Dis	0.997	0.985	0.991	0.990	0.898	0.996	0.865	0.945
OURS	0.999	0.992	0.998	0.996	0.998	0.998	0.980	0.997

reduces the loss of image information, takes into account the capture of global and local information, and can realize the complete restoration of damaged murals. A pixel-level discriminator is designed, which can discriminate the truth or falsity of each pixel point, helps to find the tiny defects of the image, and realizes the details of the restored resultant image. The details of the restored image are clear and the local contents are obvious. From Fig. 8, we can see that the addition of the ConvBlock module makes the restoration result complete, and the addition of the pixel-level discriminator makes the local details obvious, and the restoration result achieved by our method has a better visual effect.

Tests of image restoration of real broken murals

In the study of image restoration using deep learning based on the use of publicly available natural images, experiments were taken in the training and testing phases by adding random masks to the original images to simulate the broken areas. For the mural image restoration study, the aim is that the model obtained from the training can be used to restore the broken mural images in the real state. As shown in Fig. 9, samples 1, 6, 7, and 9 showed different degrees of peeling; samples 2 and 3 showed crazing; samples 4 and 8 were contaminated with the content of the picture and showed a disease similar to mold; and samples 5 and 8 showed different degrees of discoloration. Mural image restoration belongs to the real application of image restoration, and this part implements the research of mural image restoration into the practical meaning and operation. We select some mural images with representative disease characteristics, and conduct restoration test to verify the real usability of this algorithm for mural image restoration by manually calibrating the damaged area to produce the corresponding mask.

For small area broken image repair, such as samples 1, and 7, the algorithm in this paper has obvious repair advantages over the comparison algorithm in the labeled area. While coping with large area peeling, such as sample 9, the missing area is relatively concentrated and continuous, less effective semantic information, the repair ability has decreased. Although it fails to achieve perfect detail sculpture, the original broken area after a repair has structural and textural information, and in the color filling and style infinitely close to the original image. Compared to the broken image, it has the ornamental and content-knowable. To cope with the fine and narrow class of micro-length scratches in samples 2 and 3, the restored area of the algorithm in this paper blends harmoniously with the surrounding image, with a smooth and natural transition, presenting a coherent and unified visual effect. Images such as 5, and 8 are contaminated resulting in color changes, the restoration algorithm in this paper pays attention to the restoration of the color, and the restored image presents a clean and neat visual effect, eliminating the visual interference caused by discoloration and damage. Overall, this paper's algorithm can better repair the disease areas of the real murals, to a certain extent, similar to the murals "original appearance", the repaired image enhances the artistic sense of the image and visual enjoyment, so that the viewer can enjoy the original charm of the murals.

Research limitations and prospects

In the 3000 test set of sample restoration tests, found that there are some samples after the restoration of the image that the original image is more vibrant colors. Fig. 10 visualizes that the restoration method in this paper fills the missing regions with complete content and has a smooth texture and reasonable structure. Although the



Fig. 9 Comparison of the results of restoration of real broken murals, **a** Ground truth, **b** Damaged image, **c** EC, **d** FRRN, **e** DeepFillv2, **f** MEDEF and **g** ours

restored images show more vivid and brighter effects in color, there are some subjective visual differences with the input data images during the training of this paper's method.

In digital image processing, colors are usually represented in the RGB (Red, Green, Blue) color mode, with each channel using an 8-bit representation, i.e. 256 different luminance levels. This means that in a single channel,

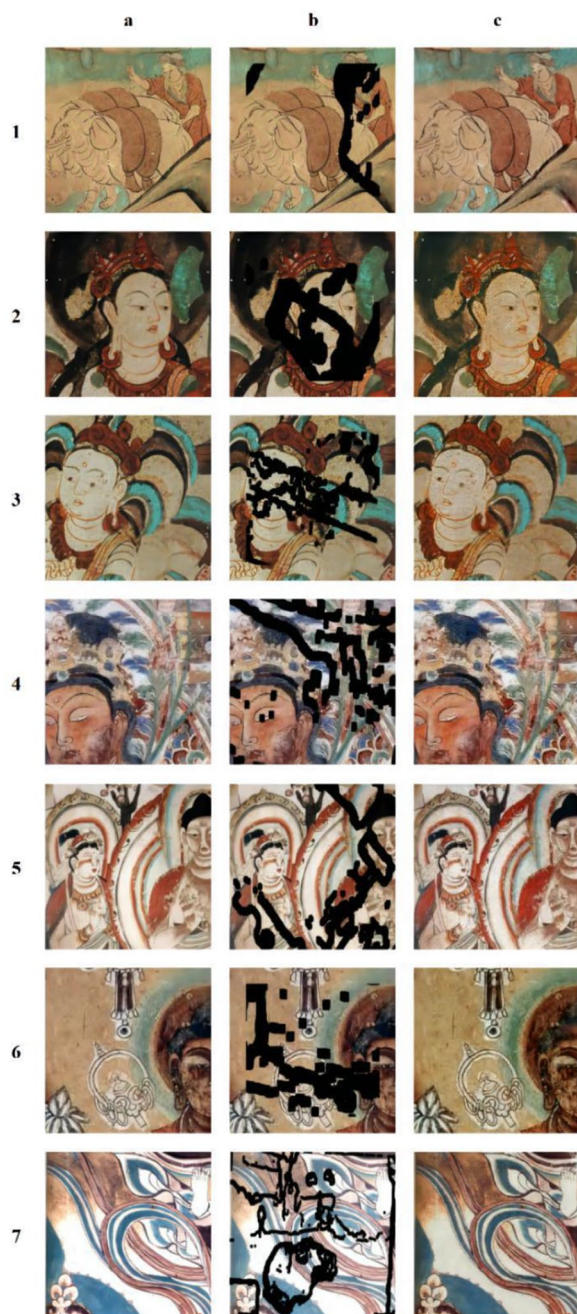


Fig. 10 Results of restoration of ours

there are 256 different reds, 256 different greens, and 256 different blues, representing a total of about 16.77 million colors. Most of the images in our self-constructed mural dataset come from the "Complete Collection of Dunhuang Murals in China". The pigments used by ancient artists in the creation of murals were diverse and colorful, unlike the color range of modern electronic devices. When computers process ancient mural images using

the standard RGB standard, the ancient pigments may contain colors and luminance levels that are not within this range will result in distortion of colors or brightness differences and other phenomena. When the color information of ancient mural images is converted into computers, the gamut (color range) of modern computer monitors may not exactly match the gamut of ancient pigments, resulting in some color deviations or approximate matches. In response to this problem, subsequent work could look to try color spaces that are better suited to ancient colors, for example, a wider range of colors could be used, such as Adobe RGB or ProPhoto RGB, which might be likely to better capture the specific colors and brightness of ancient pigments.

Due to the unique nature of ancient pigments, matching ancient colors exactly and precisely can be a complex task. During the image generation process, colors that are visually flattering and relatively close to each other may be selected to give a more natural and flattering effect to the atmosphere of the original artwork. In the restoration result map shown in Fig. 10 the colors presented in the generated image may be relatively darker, presenting an image with more depth and texture. In the subjective evaluation experiment, the five students from different majors agreed that the restored images of the murals brought a better sense of visual experience. The reviewers believed that in striving to maintain the original appearance of the ancient murals and upholding the principle of restoring the structural semantics of the main content with respect, the overall texture and attractiveness of the restored effect was improved, which is permissible and acceptable. And the overall view of the result was Optimised. The reviewer believes that the restored image better restores the unique colors of the murals.

On the other hand, considering from the dataset point of view, the training data has a lot of example data with bright colors. And the imbalance of colors has an impact on the learning of the model, and the model tends to generate more brightly colored results. Most of the images used for training and testing of restoration algorithms for the restoration of ancient murals have been contaminated and damaged, and it is difficult to find clear and complete high-definition images, and the colors initially set by the ancient painters are no longer available. Since there is no publicly available standard dataset, researchers use self-constructed datasets for model training and testing. In my next work, the structure of the dataset composition can be adjusted from the perspective of the training data to ensure that images with various color saturation and brightness are covered. Establishing publicly available good-quality datasets of ancient mural images will be a key part of the performance improvement of the mural image restoration algorithms. The pigments

used in ancient murals have changed over time, fading and discoloration by light, oxidation, chemical reactions, and other environmental factors. In the work of the last 2 years, research scholars have taken an alternative approach to restoring and protecting murals by using X-rays and other methods to study the materials used in murals and to determine the chemical composition and physical structure [35].

Conclusion

In this paper, for the problem of incomplete restoration and blurring of local details in Dunhuang murals restoration, a generative adversarial network model composed of a parallel dual convolutional feature extraction depth generator and a ternary heterogeneous joint discriminator is proposed. The ConvBlock module designed in the generator reduces the loss of feature information in the lower layers while realizing the capture of global information by the lower layer network. The residual structure and skip connection mechanism integrate image features and improve the expression of the consistency of the overall content of the image in the restoration algorithm. The ternary heterogeneous joint discriminator judges the image in terms of global, local, and independent pixel values, which enhances the repair capability of the generator. By comparing with the representative restoration methods mentioned in the paper, it is intuitively found that the restored resultant image is closer to the original image, while the objective evaluation PSNR value has been significantly improved.

Abbreviations

GAN	Generative adversarial network
ResNet	Residual networks
VGG	Visual geometry group network
BN	Batch normalization
IN	Instance normalization

Acknowledgements

None.

Author contributions

All authors contributed to the current work. RH proposed the research plan and supervised the whole process to provide constructive comments, SK completed the method design and model construction, and ZFH and ZX completed the dataset production and organized the experimental data. All authors read and approved the final manuscript.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 13 October 2023 Accepted: 27 January 2024
Published online: 02 February 2024

References

- Chen P, Wang Y. Research on the mural art of Dunhuang Mogao Grottoes. *Art Criticism*. 2017;20:45–6 (in Chinese).
- Zhang Y. Research on the preservation status of murals in Cave 196 of Mogao Grottoes. Northwest University, 2018 (in Chinese with an English abstract).
- Rathish Kumar BV, Halim A. A linear fourth-order PDE-based gray-scale image inpainting model. *Comput Appl Math*. 2019;38:1–21. <https://doi.org/10.1007/s40314-019-0768-x>.
- Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans Image Process*. 2004;13(9):1200–12.
- Ouattara N, Loum GL, Pandry GK, Atiampo AK. A new image inpainting approach based on Criminisi algorithm. *Int J Adv Comput Sci Appl*. 2019;10(6):423–33. <https://doi.org/10.14569/IJACSA.2019.0100655>.
- Guo Q, Li J. Damaged image restoration based on improved Criminisi algorithm. In: International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2019, pp. 31–35. <https://doi.org/10.1109/ICCNEA.2019.00016>.
- Li C, Chen H, Han X, Pan X, Niu D. An improved Criminisi method for image inpainting. *J Phys Conf Ser*. 2022;2253(1):012023. <https://doi.org/10.1088/1742-6596/2253/1/012023>.
- Yao F. Damaged region filling by improved Criminisi image inpainting algorithm for thangka. *Cluster Comput*. 2019;22:13683–91. <https://doi.org/10.1007/s10586-018-2068-4>.
- Cao J, Li Y, Zhang Q, Cui H. Restoration of an ancient temple mural by a local search algorithm of an adaptive sample block. *Herit Sci*. 2019;7(1):39. <https://doi.org/10.1186/s40494-019-0281-y>.
- Liu Y. Research on Yunnan mural restoration based on improved Criminisi algorithm. Yunnan University, 2017 (in Chinese with an English abstract).
- Mol VR, Maheswari PU. The digital reconstruction of degraded ancient temple murals using dynamic mask generation and an extended exemplar-based region-filling algorithm. *Herit Sci*. 2021;9(1):137. <https://doi.org/10.1186/s40494-021-00604-2>.
- Yan M, Xiong R, Shen Y, Jin C, Wang Y. Intelligent generation of Peking opera facial masks with deep learning frameworks. *Herit Sci*. 2023;11(1):20. <https://doi.org/10.1186/s40494-023-00865-z>.
- Chakraborty T, Reddy UKS, Naik SM, Panja M, Manvitha B. Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Mach Learn Sci Technol*. 2023. <https://doi.org/10.1088/2632-2153/ad1f77>.
- Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Trans Graph*. 2017;36(4):1–14. <https://doi.org/10.1145/3072959.3073659>.
- Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 1125–1134.
- Liu G, Reda FA, Shih KJ, Wang TC, Tao A, Catanzaro B. Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European conference on computer vision (ECCV). 2018, pp. 85–100.
- Guo Z, Chen Z, Yu T, Chen J, Liu S. Progressive image inpainting with full-resolution residual network. In: Proceedings of the 27th ACM international conference on multimedia. 2019, pp. 2496–2504.
- Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS. Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, pp. 4471–4480.
- Liu H, Jiang B, Song Y, Huang W, Yang C. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In: Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020, pp. 725–741.
- Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M. Edgeconnect: generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019. <https://doi.org/10.48550/arXiv.1901.00212>.

21. Yi Z, Tang Q, Azizi S, Jang D, Xu Z. Contextual residual aggregation for ultra high-resolution image inpainting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, pp. 7508–7517.
22. Qin J, Bai H, Zhao Y. Multi-scale attention network for image inpainting. *Comput Vis Image Underst*. 2021;204: 103155. <https://doi.org/10.1016/j.cviu.2020.103155>.
23. Quan W, Zhang R, Zhang Y, Li Z, Wang J, Yan DM. Image inpainting with local and global refinement. *IEEE Trans Image Process*. 2022;31:2405–20. <https://doi.org/10.1109/TIP.2022.3152624>.
24. Cao J, Zhang Z, Zhao A, Cui H, Zhang Q. Ancient mural restoration based on a modified generative adversarial network. *Herit Sci*. 2020;8(1):7. <https://doi.org/10.1186/s40494-020-0355-x>.
25. Li J, Wang H, Deng Z, Pan M, Chen H. Restoration of non-structural damaged murals in Shenzhen Bao'an based on a generator–discriminator network. *Herit Sci*. 2021;9(1):6. <https://doi.org/10.1186/s40494-020-00478-w>.
26. Wang N, Wang W, Hu W, Fenster A, Li S. Thangka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. In: IEEE Transactions on Image Processing, 2021, vol. 30, pp:3720–3733. <https://doi.org/10.1109/TIP.2021.3064268>.
27. 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. <https://doi.org/10.1186/s40494-022-00771-w>.
28. Li L, Zou Q, Zhang F, Yu H, Chen L, Song C, Wang X. Line drawing guided progressive inpainting of mural damages. arXiv preprint [arXiv:2211.06649](https://arxiv.org/abs/2211.06649), 2022. <https://doi.org/10.48550/arXiv.2211.06649>.
29. Deng X, Yu Y. Ancient mural inpainting via structure information guided two-branch model. *Herit Sci*. 2023;11(1):131. <https://doi.org/10.1186/s40494-023-00972-x>.
30. Xu Z, Zhang C, Wu Y. Digital inpainting of mural images based on DC-CycleGAN. *Herit Sci*. 2023;11(1):169. <https://doi.org/10.1186/s40494-023-01015-1>.
31. Liu W, Shi Y, Li J, Wang J, Du S. Multi-stage progressive reasoning for Dunhuang murals inpainting. In: 2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML). IEEE, 2023, pp:211–217.
32. Woo S, Park J, Lee JY, Kweon IS, Cbam. Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). 2018, pp. 3–19.
33. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016, pp. 694–711. https://doi.org/10.1007/978-3-319-46475-6_43.
34. 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. 2016, pp. 2414–2423.
35. Wang Y, Wu X. Current progress on murals: distribution, conservation and utilization. *Herit Sci*. 2023;11(1):61. <https://doi.org/10.1186/s40494-023-00904-9>.

Publisher's Note

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