SeparaFill：Two generators connected mural image restoration based on generative adversarial network with skip connect

Yinghua Shen
Communication University of China
Zilu Li (zilulee@cuc.edu.cn)
Communication University of China
Jinghua Li
Communication University of China
Jin Zheng
Communication University of China
Chaohui Lv
Communication University of China

Research Article

Keywords: SeparaFill, contour restoration generator, the content completion network, mural restoration

Posted Date: May 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1638201/v1

SeparaFill：Two generators connected mural image restoration based on generative adversarial network with skip connect

Yinghua Shen[1], Zilu Li[1]*, Jinghua Li[1], Jin Zheng[1] and Chaohui Lv[1]

1Communication University of China, School of Information and Communication Engineering, Digital Media Technology, Chaoyang District, Beijing, China, 100000

***Correspondence:**Zilu Li**,** E-mail: zilulee@cuc.edu.cn

Abstract

Murals are the important components of culture and arts of Dunhuang. Unhappily, these murals have been ruined or are being ruined by some diseases such as cracking, hollowing, falling off, getting mildewed, dirt, and so on. Due to a lack of a standard mural datasets, Dunhuang mural datasets are created by ourselves. Meanwhile, our proposed network architecture SeparaFill which is connected two generators based on U-Net. First, the contour restoration generator network is used to repair contour lines. Then, the color mural image is repaired by the content completion network with help of the repaired contour. Next, global and local discriminant networks are applied to determine whether the repaired mural image is authentic in terms of both the modified and unmodified areas. Compared with existing mural restoration algorithms, the proposed method increases the peak signal-to-noise ratio (PNSR) and increases the structural similarity (SSIM). SeparaFill shows good performance in restoring the line structure of the damaged mural images and retaining the contour information of the mural images.

Keywords: SeparaFill, contour restoration generator, the content completion network, mural restoration

Introduction

Dunhuang murals retain the authentic works of famous artists over thousands of years, which have profound artistic, historical and cultural value. Due to the long-term impact of natural weathering and human factors, these murals have undergone different kinds of disease, such as cracking, falling off, hollowing, pulverization, fading, color changing, getting mildewed, smudging and scratches, and so on. Therefore, there is an urgent need to restore the murals combined with the environment and painting materials. Meanwhile, the manual restoration work of Dunhuang murals is arduous and complex, requiring the joint efforts and support of multiple disciplines[1]. The development of image processing and deep learning technology have allowed virtual restoration of mural images to become a research hot-spot.

The research for virtual restoration of mural images are divided into two main categories, including the digital image restoration based on traditional algorithms and the restoration method based on deep learning. Cao et al.[2], Liu et al.[3], and Yin et al.[4] improved the traditional algorithm to restore the mural image. While Wang et al.[5], Zhang et al.[6], and Wen et al.[7] proposed the improvement algorithms of the Generative Adversarial Networks (GAN) [8] based on the deep learning research. Currently, most of the researches are based on the public datasets of natural images in the field of deep learning images restoration, which can capture the potential statistical laws in natural images through the learning of massive data. Dunhuang murals are meticulous heavy color painting with extremely and complexly structures and are obviously different from the natural images restored by self-similarity characteristics, while the structure of mural images are mainly determined by the line drawing. Due to the complex content of murals, it is also difficult to distinguish the foreground

and background. So, it is very difficult to restore the mural image well using the original GAN networks. As there is lack of public mural image datasets so far, it is essential to establish the mural datasets based on the Dunhuang mural album.

The mural images are mainly subject to damage from a wide variety of causes, resulting in some small scale of mural diseases and some large areas missing. These small missing areas such as cracking, small falling off blocks and pulverization of the mural, tend to appear in clusters and the areas intertwined with the complete areas. These large missing areas such as scratch and large peeling, account for a significant portion of the information of the mural image, which seriously affects the integrity of the mural. In order to take advantage of effective information of the image, good results can also be obtained in large-scale image restoration, a two generators connected image restoration networks (SeparaFill) is proposed according to the characteristics of mural painting. This algorithm based on U-Net network[9] restore the contour lines of the image first and then restore the color areas inside the contour lines of the image. The improvements achieved by the proposed method are mainly reflected in the following aspects: (1) skip connections are added to the contour restoration network for feature channel fusion to realize the reuse of low-level features and extract the high-level semantic features of the image by using Hierarchical Residual Networks(Res2Net)[10]; (2) an accumulation feature extraction mechanism is proposed to realize multi-level feature fusion under different resolutions in the content completion restoration network, and self-attention mechanism is introduced to restore image details. (3) Because of the small datasets, the Siamese Network idea of Meta learning is introduced to the discriminator network, and the contrast loss function commonly used in the Siamese Network is introduced into the discriminator optimization. Compared with other algorithms, this method has achieved better performance.

The traditional Image restoration algorithms are the ill-posed problems of filling in the missing pixels in digital images by interpolating from the prior information. Rudin et al.[11] proposed the Total Variation (TV), which used the partial differential equation based on the principle of thermal diffusion to carry out anisotropic diffusion to repair the damaged areas. Zhang et al.[12] replaced the integer order differential in the TV model with fractional order and further considered the image texture structure information to improve the accuracy of image restoration. The TV method achieves good effects on small missing areas, but it is prone to the problems of error diffusion and poor visual connectivity. The algorithm based on texture synthesis proposed by Criminisi et al. [13] considered the continuity of texture and can solve the problem of poor visual connectivity of partial differential equations to a certain extent. On the basis of the above algorithm, Liu et al. [14] and Cao et al.[2] improved the confidence of priority and the adaptive selection of template window size to optimize the effect of image restoration and avoided the diffusion phenomenon caused by the restoration.

With the continuous development of deep learning, deep neural networks have shown excellent ability in the prior knowledge learning of massive data. GAN networks can generate non-existent images through the learning of image features and become a research hotspot in the field of image restoration. Yeh et al.[15] proposed a semantic image restoration model with depth generation mode by adopting the Deep Convolutional GAN (DCGAN) structure and using context-weighted loss to search for the closest encoding of the corrupted image, which has good performance on restoring images with the simple structure. Zhang et al.[16] used a four step incremental generation networks to restore the image under the square mask, but this restoration method cannot deal with the irregular mask. Liu et al.[17] proposed an image restoration method based on partial convolution (Pconv) by using partial convolution with mask instead of full convolution filling and it can repair

89 the irregular holes of images. Guo et al.[18]designed the network generator as a network block with
90 two parallel branches of low resolution and full resolution. Through the superposition of network
91 blocks, the mask area was gradually cleaned and reduce to zero. Yu et al.[19], Nazeri et al.[20] and
92 Zamir et al.[21] adopted two connected generators to restore the image. Yu et al.[22] further proposed
93 gate convolution on the previously model and replaced the ordinary convolution in the generator
94 to make the mask update learnable. Xiong et al.[23] proposed a three-stage image restoration model
95 to distinguish the foreground and background, which was suitable for a single object or
96 architectural image with clear outline and was not operable for images with complex lines and rich
97 content for murals. Ronneberger et al.[24], Li et al.[25], Yang et al.[26], Jo et al.[27] and Liu et al.[28]
98 introduced skip connections into the corresponding convolution layer of down sampling and up
99 sampling based on the U-Net networks model to transfer the feature information extracted by the
100 networks, so as to improve the utilization efficiency of the networks for low-level features and
101 refine the texture of image restoration.

102 **Methods**

103 It is observed that the murals are dominated by contour lines through the analysis of the painting
104 characteristics of murals. Therefore, its pixels can be divided into contour line parts and color
105 block parts separated by contour line for a mural image. The contour line parts of the mural image
106 are composed of rich thin, narrow and continuous lines. Meanwhile, the color is relatively single,
107 which usually contains the dark colors such as black, brown and red. However the color block
108 parts contain rich color information. The size and shape depend on the contour line, but the gray
109 scale inside the color block is continuous and the texture information is simple. With the mural
110 image characteristics above, a two generators connected mural image restoration network based
111 on U-Net network architecture is proposed. This method restores the mural image contour and its

internal color blocks separately and reduces the difficulty of restoration. After two days of training, the network can obtain the better restoration result compared with other algorithms.

The mural restoration network mainly consists of three parts: contour restoration generator network, content completion restoration generator network, global and local discriminator network.

The damaged image is obtained by multiplying the image with the mask,Let I_{gt} be ground truth images, M be mask images, $x_{masked}^{(i)} = I_{gt}$ e M be the damaged image, $sketchgray^{(i)}$ is obtained by the Holistically-nested Edge Detection (HED) algorithm[29], and the contour damaged image $sketch_{masked}^{(i)} = x_{masked}^{(i)}$ e $sketchgray^{(i)}$ e M . Feed the damaged contour image and damaged image into the contour restoration generator network, at the same time, Let $sobel^{(i)}$ as the edge image of the damaged image via Sobel edge detection processing. Using the $x_{masked}^{(i)}$, $sketch_{masked}^{(i)}$ and $sobel^{(i)}$ as input of contour restoration generator. Fill the contour recovery map $sketch_{g}^{(i)}$ into the image's missing area, $x_{masked1}^{(i)} = sketch_{g}^{(i)}$ e $(1-M)+x_{masked}^{(i)}$, the large missing block is further divided into several small areas. The damaged contour image obtained by multiplying with the mask is input into the contour repair generator network. At the same time, the sobel edge detection processing is conducted on the damaged image to acquire the edge image, which is input to the generator input end together with the damaged color image to assist the contour restoration. The recovered contour is filled into the image to be repaired to further split the damaged areas of the image. Meanwhile, the contour recovery map is also sent to the content completion restoration generator in the second stage to guide the restoration of the image. The discriminator network of contour restoration is as same as content completion restoration, which is composed of local discriminator and global discriminator. The network framework is shown in Fig.1.

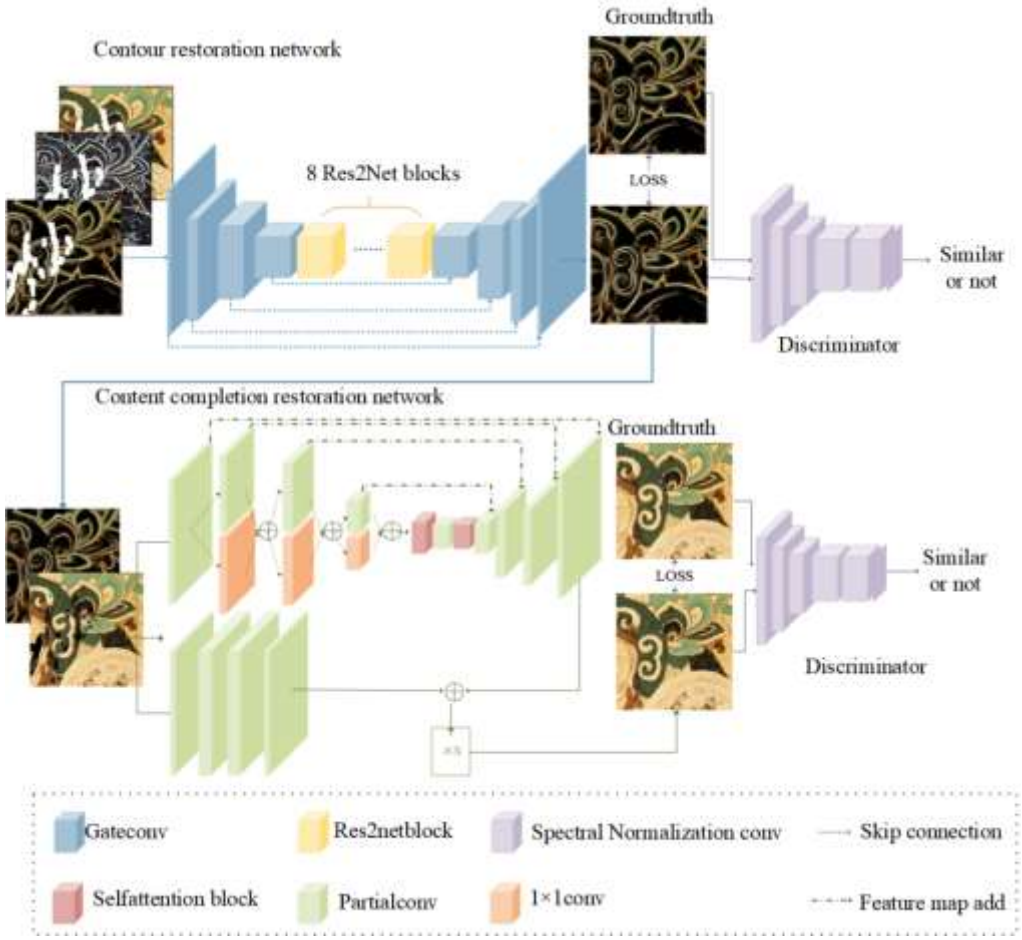

Fig.1 Our network framework

Contour restoration network

The generator of the contour restoration network is improved on the Edgeconnect model[20]. 9 channels are put into the network, which include the damaged contour image of RGB (red, green, blue) three channels, the damaged image and the edge feature map extracted by Sobel edge detection. The input of the damaged image with the edge image together provide rich information for network and guide the contour restoration. The network consists of a down sampling convolution block contained a 5×5 convolution layer and three 3×3 convolution layers, 8 Res2Net blocks and a up sampling recovery resolution convolution block. Added skip connections between the corresponding layers of 4-layer down sampling convolution layers and up sampling

convolution layers of the contour restoration generator, which, reuses the low-layer edge features, and retains more dimensional information. So, the up sampling network of the generator can select between shallow features and deep features to enhance the robustness of the network. For the convolution in the shallow feature extraction network and the up-sampling convolution network, we use gate convolution to replace the ordinary convolution.

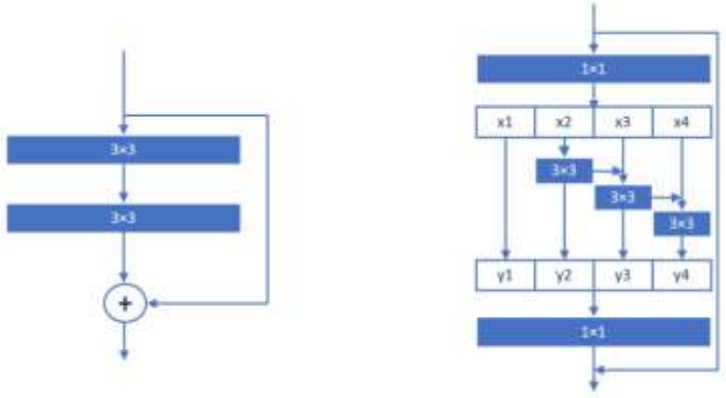

Fig. 2 (a) Residual block of Edgeconnect and (b) Residual block of Res2Net

The block of Res2Net used to replace the residual block structure in contour restoration network. Res2Net block modifies the structure of Residual Networks (ResNet) block[30] which is shown in Fig.2. Firstly, the input features are passed through a layer of 1×1 convolution, further divide the output features equally according to the number of channels, and fuse the segmented features in different channel blocks. The expression formula is as follows:

$$y_i = \begin{cases} x_i & i = 1; \\ K_i(x_i + y_{i-1}) & 1 < i \le s \end{cases} \quad (1)$$

Where x_i represents the number of equally divided blocks, s represents the number of equally divided blocks, y represents the output of convolution, and $K_i(\,)$ represents 3×3 convolution. The residual structure of Res2Net retains the function of ResNet to avoid gradient disappearance and

gradient explosion, and realizes channel block multiplexing of the 3×3 convolution layer in the ResNet block. This multi-scale channel fusion of input features can make the ability of feature extraction stronger without add network parameters.

Content completion restoration network

The aim of the second stage is to complete the color blocks between the contour lines. The network inputs are the contour map generated in the first stage and the corresponding damaged image after repairing contour lines. The inpainting blocks include a U-shaped restore network branch and an original resolution branch without down sampling, filling the missing area gradually and finely by superimposing multiple blocks. The U-shaped restore branch consists of the 4 down sampling convolution layers that the kernel is 3×3 with 2 step sizes, the feature extraction network with two self-attention blocks, and up sampling layers of 3×3 convolution corresponding to the down sampling layers. In order to reduce the network parameters and the memory occupation, so as to increase the number of superimposed inpainting blocks for restoring delicately. The feature map is directly added so that the feature map of each dimension will contains more characteristics. Furthermore, an accumulation feature extraction mechanism is proposed, the feature map output through each convolution layer is superimposed with the feature map output through the layer in front, so as to realize multi-level feature fusion under different resolutions. The implementation formula is as follows:

$$y^l = \frac{y^l}{2^l} + \frac{1}{2^l}\sum_{i=1}^{l} 2^{i-1} y^i, \tag{2}$$

Where y^l represents the convolution fusion output of layer *l*, and y^i represents the fusion output of convolution layer *i*. It can be seen from the formula that the superposition fusion mechanism makes the features extracted by the fusion convolution layer accumulate the characteristics output

of all the convolution layers in front. Since the feature fusion of direct addition requires the same size of the input feature map, the feature map with different resolution is matched by reducing the resolution through 1×1 convolution layers.

As the convolution kernel operates from the local region of the image and represents the local features, the influence of the global features on the current region becomes very small with the deepening of the convolution network. Self-attention mechanism[31] can capture long-distance dependencies, namely paying attention to the global characteristics, so as to enlarge receptive fields of the network. After the feature accumulation layer, the self-attention mechanism is introduced to capture the overall features and detail features of the mural image, making the generated image more detailed.

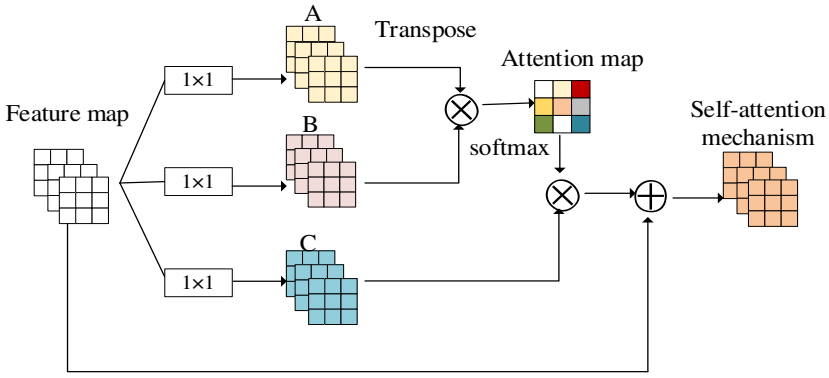

Fig. 3 Structure of self-attention mechanism

The original resolution branch does not use down sampling in the process of transmitting and processing the input image information, so as to maintain the resolution of the original input information, assist in thinning the texture of the image restoration, and reuse the input information. Dilated convolutions are used in the content completion restoration network. The dilation is set as a loop of 1, 2 and 5 to increase the perception domain of the convolution. Since the large damaged area has divided into small pieces after the contour lines have been repaired in the first stage, the

difficulty of restoration becomes easier. Therefore, the partial convolutions with fewer parameters are used to update the mask and perform detailed restore through the superposition of modules.

Loss function

The loss function in the contour inpainting phase is expressed as:

$$L_{s\_G} = \lambda_{adv} L_{adv} + \lambda_{rec} L_{rec} + \lambda_{FM} L_{FM} \,, \tag{3}$$

Where L_{adv} is the adversarial loss based on the discriminator, L_{rec} is the L_1 reconstruction loss, L_{FM} is the feature matching loss, and λ_{adv}, λ_{rec} and λ_{FM} are the weights of each loss respectively. Generally, GAN obtains the optimal solution by optimizing the value function. The value function is expressed as:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log D(x)\right] + E_{i\sim p_{out}(i)}\left[\log\left(1-D\left(G(i)\right)\right)\right], \tag{4}$$

Where x represents the input data, $P_{data}(x)$ represents the distribution of the real data, $P_{out}(i)$ represents the distribution of the image generated by the generator, D represents the discriminator, the probability that the output input is the real data, and G represents the generator, which outputs the generated image.

The goal of the discriminator is to maximize the value function. In view of the small mural datasets and the characteristics of the discriminator structure, the contrast loss function commonly used in the Siamese Network[32] is introduced to transform the process of maximizing the value function of the discriminator into the process of reducing the Euclidean distance between the truth image and the generated image. The comparison loss function is expressed as follows:

$$d\_loss = \frac{1}{2N}\sum_{n=1}^{N} y * d^2 + (1-y)\max(m\text{-}d,0)^2 \,, \tag{5}$$

Where y is the label of whether the input two samples (truth image and generator output image) match or not, y=1 represents the similarity or matching of the two samples, y=0 represents the

mismatch, which $d = \|I_{gt} - I_{recover}\|_2$ represents the Euclidean distance between the ground truth image and the generated image, where $I_{recover}$ is the recovered image generated by the generator, I_{gt} is the input real image sample, *m* is the set threshold, which is set to 3 here.

The reconstruction loss is used to constrain the image pixel level restoration, so as to optimize the detail restore ability of the contour.

$$L_{rec} = \left\| I_{recover} - I_{gt} \right\|_1 * \lambda_{rec} + \left\| I_{recover} \odot masks - I_{gt} \odot masks \right\|_1 * \lambda_{rec} \ , \tag{6}$$

Where *masks* is the binary mask image, and $\odot$ is the Hadamard product, used to calculate the global and local reconstruction losses for the generated image and the hole area under mask constraints respectively, λ_{rec} represents the weight value of the loss function.

The feature-matching loss is used to compare the feature maps in the intermediate layers of the discriminator. The feature-matching loss is expressed as follows:

$$L_{FM} = \mathrm{E}\left[\sum_{i=1}^{L} \frac{1}{N_i} \left\| D_1^{(i)}\left(I_{gt}\right) - D_1^{(i)}\left(I_{recover}\right) \right\|_1 \right] * \lambda_{FM} \ , \tag{7}$$

Where *L* is the number of convolution layers of the discriminator, N_i is the number of characteristic diagrams of the activation layer of layer i, and $D_1{}^{(i)}$ is the activation number of layer of the discriminator. λ_{FM} is the regularization parameter.

The content restoration network needs to restore the texture of the image and maintain the semantic consistency between the restored image and the ground truth image. The loss function consists of confrontation loss, reconstruction loss, perception loss and structural similarity loss. The loss function is expressed as follows:

$$L_G = \lambda_{adv} L_{adv} + \lambda_{rec} L_{rec} + \lambda_{SSIM} L_{MS-SSIM} + \lambda_{style} L_{style} , \tag{8}$$

The loss function and weight of the reconstruction loss L_{rec} and adversarial loss L_{adv} are the same as first part of contour restore. In order to better ensure that the texture and color of the image

restoration area fit the original mural, and make the style of the whole restored image consistent, the perceptual loss function[33] is introduced. The perceptual function is divided into content loss and style loss, compares the high-level abstract features through the VGG 19 pre-training model, the formula is as follows:

$$1_{feat}^{\varphi,j}\left(\hat{y},y\right)=\frac{1}{C_jH_jW_j}\left\|\varphi_j\left(\hat{y}\right)-\varphi_j\left(y\right)\right\|_2^2, \tag{9}$$

Where C_j, H_j and W_j represent the channel numbers, height and width of the characteristic graph respectively, j represents the j-th layer of the network, and φ represents the output after convolution network processing. Content loss let the generated image obtain better visual effect, but large loss weight will produce texture to image that does not conform to the original image, so it is necessary to reduce the weight of content loss in the later stage of training.

The multi-scale structure similarity loss function[34] is introduced. The combination of structural similarity loss and L1loss can balance the brightness and color of the image, thus making the restored image more detailed. The function expression is as follows:

$$L_{MS\text{-}SSIM}\left(P\right)=1-MS\text{-}SSIM\left(\tilde{p}\right), \tag{12}$$

Where $MS\text{-}SSIM(\tilde{p})$ is SSIM calculation for images with different resolutions after scaling, which can obtain better results than simple SSIM loss.

Training and testing procedures

Limited by the small mural image datasets, when the parameter of batch size of training is set too large, which will affect the training results. So the parameter of batch size is set as 5, each of which batch has 3000 data, and the parameter of num_workers is set as 16, which is used to preload the batch data of the next iteration into memory. The specific algorithm steps are as follows:

Table 1 Pseudocode of algorithm

Initialize the parameters of generator G, discriminators D1 and D2, iterations=n, batch_size=m, and the super parameter k=1 representing the steps of discriminator

for number of training iterations do

for k steps do

- The m real images$\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ were taken randomly from $P_{data}(x)$
- The m mask maps$\{M_1, M_2, \dots, M_m\}$ were taken in M_{mask} paired with the images
- via the HED networks to obtain contour $sketchgray^{(i)}$ from $x^{(i)}$
- $sketch^{(i)} = x^{(i)} \odot sketchgray^{(i)}$
- $x_{masked}^{(i)} = x^{(i)} \odot M$, $sketch_{masked}^{(i)} = sketch^{(i)} \odot M$
- $sketch_g^{(i)} = sketch^{(i)} \odot M + G_1(sketch^{(i)}, M) \odot (1 - M)$
- $x_{masked1}^{(i)} = sketch_g^{(i)} \odot (1 - M) + x_{masked}^{(i)}$
- Image inpainting $x_g^{(i)} = x_{masked1}^{(i)} \odot M + G_2(x_{masked1}^{(i)}, M) \odot (1 - M)$
- Calculate the restoration loss L_{dS}, L_d based on the input and restore samples
- Update the parameters of discriminators D1 and D2 through Adam optimization algorithm

end for

- Get m samples randomly from $P_{data}(x)$ and M_{mask} , constructing damaged images $x_{masked}^{(i)}$, $sketch_{masked}^{(i)}$, calculate the loss L_{dS}, L_d based on the input samples and restored samples generated by the network
- Update the parameters of discriminators D1 and D2 through Adam optimization algorithm

end for

Results and discussion

Data source

In this study, some clear and well preserved murals of characters in the Tang Dynasty from the electronic scanning edition of *Dunhuang Mural Art* (all 10 volumes), *Chinese Dunhuang murals* (all 11 volumes) and *Dunhuang Grottoes* (all 26 volumes) were used. After cleaning the image data, eliminating duplicate mural images, 2175 original data images, consist of 172 images of the early Tang Dynasty, 271 images of the prosperous Tang Dynasty, 859 images of the middle Tang Dynasty, 743 images of the late Tang Dynasty and 130 images of the Five Dynasties were obtained, and all of them were 512×512 in size. A total of 300 images were randomly selected from the images of each dynasty as the test sets, and the remaining 1875 images as training sets. The datasets

is expanded by mirror operation, and then divided each original mural image into four small sub-images (256×256) by horizontal and vertical segmentation. After that, a 15000 split training datasets and a 2400 split test datasets were obtained and the size of each image was 256×256×3, which contained rich mural feature elements such as the clothing, the texture, the face and the decoration. The same number of images was randomly selected from the mask datasets to correspond to the mural images one by one. The size of each mask image was 256×256×1. Mask datasets imitated irregular damages such as cracking, falling off, pulverization, getting mildewed, smudging and scratches in mural diseases. The damaged RGB image was obtained by multiplying the mural image and the mask image.

Experimental environment

To verify the effectiveness of the proposed method, tests on mural image restoration were conducted. The hardware environment in this experiment mainly consists of an Intel Xeon e5-2620 V4 @2.1GHZ with 128GB memory and four Nvidia GeForce GTX 1080 Ti graphics cards with 11GB memory. The software environment includes the JetBrains Pycharm compiler, running on a Windows 10 system. The software was written in python 3.8, and Pytorch was used as the framework for complete mural image restoration.

Restoration of the randomly damaged murals

In order to verify the effectiveness of the mural image restoration model proposed in this paper, our network model are compared with Pconv[17], Edgeconnect[20], FRRN[18], RN[35] and RFR[36] networks on the test image datasets established in this paper. For the image restoration results of different network models, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as the quantitative evaluation indices. The objective comparison of repair results is shown in

Table 2 and 3.Getting random mask from test mask datasets covers the test images to obtain the artificially damaged mural image, which is sent to the trained network to compare with the mural restore consequence. Otherwise, the running environment of all networks above is the same as our algorithm.

As can be seen from Fig.4, the restoration part of the algorithms in RN[35], Edgeconnect[20] and RFR[36] is relatively fuzzy, and when the damaged area contains structural information, the original contour of the image cannot be well restored in algorithms Pconv[17] and FRRN[18], either. In this case, our network can not only obtain the prediction map with high PSNR and SSIM values, but also express the visual connectivity and structural consistency of the image well. With the increase of the degree of damage, the PSNR and SSIM of all 5 compaired algorithms decrease significantly, problems such as incomplete inpainting, and image pixel error diffusion have appeared, however our algorithm can still maintain the contour integrity and continuity of mural image, such as ribbon, sphere, hat wing, shoulder of the characters in the painting, restoring its complete structure with less effective information of the image.

Table 2 Comparison of SSIM value of each algorithm

Sample	Pconv	Edgeconnect	FRRN	RN	RFR	SeparaFill(ours)
1	0.9556	0.9187	0.9769	0.9322	0.9354	**0.9788**
2	0.9347	0.9113	**0.9699**	0.9434	0.9483	0.9677
3	0.7801	0.7904	**0.9310**	0.8774	0.8453	0.9234
4	0.7982	0.7509	**0.8999**	0.8108	0.8355	0.8826
5	0.8166	0.8135	**0.9118**	0.8291	0.8623	0.9067

Table 3 Comparison of PSNR value of each algorithm

Sample	Pconv	Edgeconnect	FRRN	RN	RFR	SeparaFill(ours)
1	31.73	24.86	33.78	26.92	24.57	**33.86**
2	31.54	26.16	33.69	30.57	30.45	**34.25**
3	30.54	24.58	32.71	28.66	27.52	**33.36**
4	22.36	19.69	**25.96**	22.39	22.86	25.33

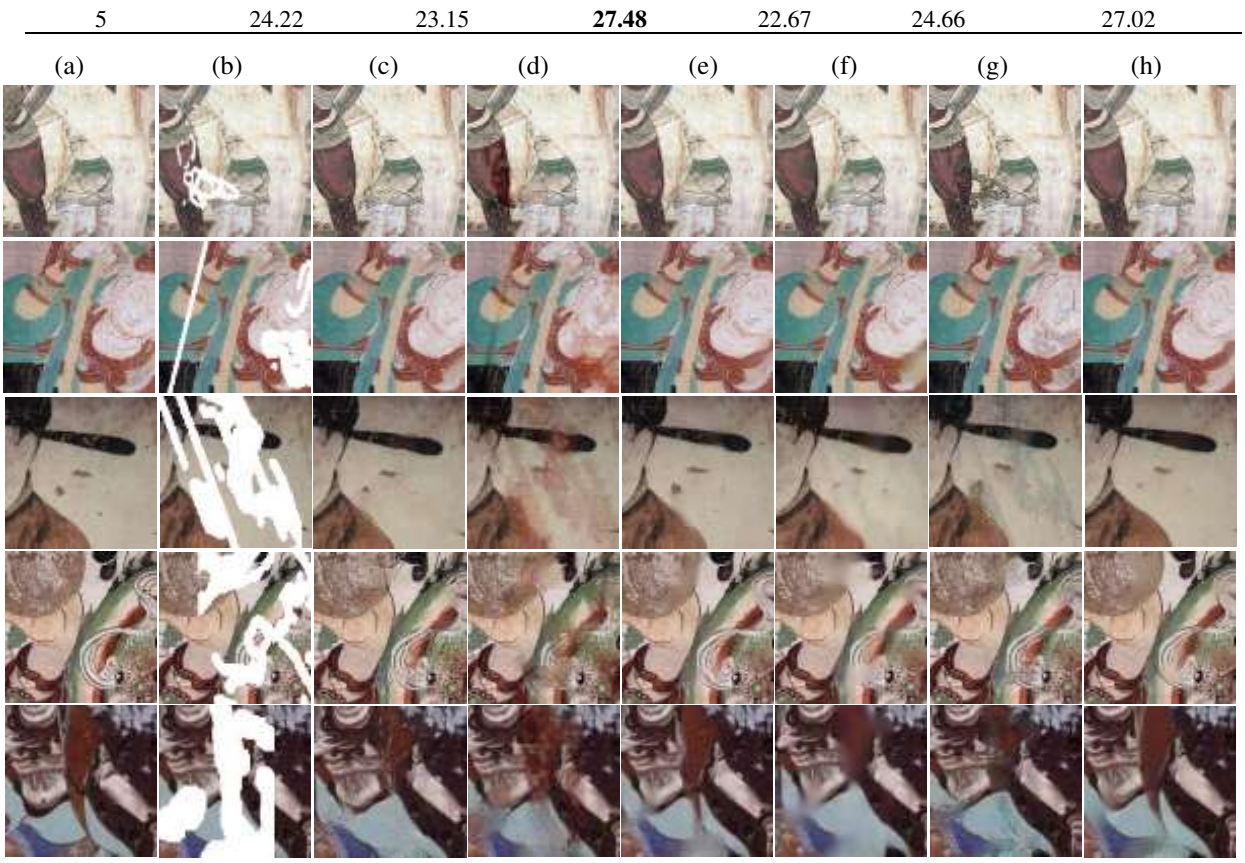

Fig.4 Comparison of large area damage restore results, (a) Ground truth, (b) Damaged image, (c) Pconv, (d) Edgeconnect, (e) FRRN, (f) RN, (g) RFR and (h) SeparaFill(ours)

Restoration of the authentic damaged murals

For the real damaged Dunhuang mural images, need to label and mark the real damaged area artificially, and the ground truth for these damaged historical relics disappeared many years ago.Therefore, we select some mural images and label the real damages with mask manually and compare our method with the algorithm in Ref [20] and [18], the objective comparison of repair results is shown in Table 4. It can be seen that the proposed method has significant advantages in objective evaluation in PSNR and SSIM indexes. As can be seen from Fig.5, our algorithm has a good and stable restore effect on distributed discrete damages, small pulverization, scratches,

cracks and other diseases, and keep the fill area consistent with the style and texture of the original image.

Table 4 Comparison of SSIM value and PSNR value of each algorithm with large area damage restoration

Sample	Edgeconnect		FRRN		SeparaFill(ours)	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
1	25.19	0.8874	29.86	0.9457	**30.65**	**0.9512**
2	30.86	0.9540	33.72	0.9678	**34.53**	**0.9714**
3	25.69	0.8946	29.77	0.9566	**31.76**	**0.9613**
4	23.03	0.8327	26.98	0.9059	**28.51**	**0.9206**
5	26.11	0.8687	30.23	0.9342	**32.36**	**0.9533**

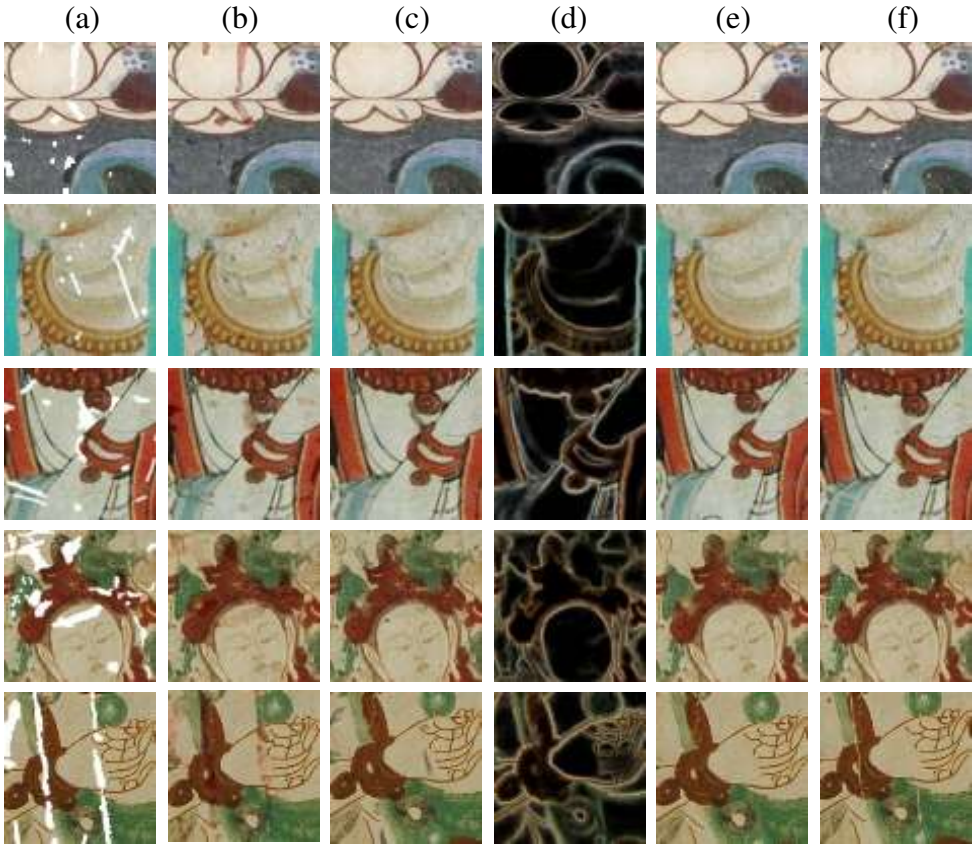

Fig.5 Comparison of small area real damage inpainting, (a) masked image, (b) Edgeconnect, (c) FRRN, (d) restored contour(ours), (e) SeparaFill(ours), (f) ground truth

For the large-area pulverization, falling off, fading and other diseases of murals, it is difficult to obtain realistic inpainting results because there is less effective information and it is very difficult to label the mask completely. As can be seen from Fig.6, our algorithm are more realistic and

natural than the results of the model Edgeconnect[20] and FRRN[18], it can reduce the impact of the damaged area and restore the "original appearance" of murals as much as possible when the mask completely covers the damaged area.

Table 5 Comparison of SSIM value and PSNR value of each algorithm with large area damage restoration

Sample	Edgeconnect		FRRN		SeparaFill(ours)	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
1	24.14	0.8524	**26.51**	0.9128	26.49	**0.9558**
2	18.32	0.6138	**20.49**	0.7185	18.92	0.7010
3	19.44	0.5834	**20.62**	0.6989	20.13	**0.7027**
4	14.72	0.5490	**18.69**	**0.6601**	15.13	0.6295

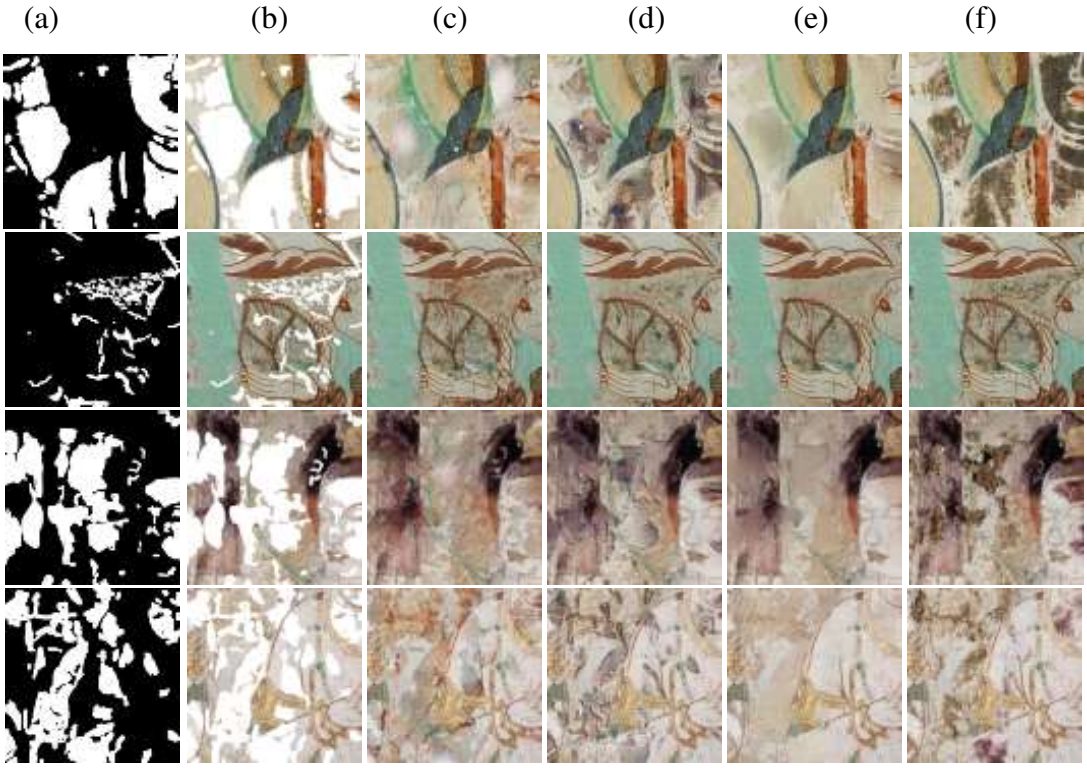

Fig.6 Comparison of large area real damage inpainting, (a) manual mask,(b) masked image, (c) Edgeconnect, (d) FRRN, (e) SeparaFill(ours), (f) ground truth

Conclusions

According to the painting characteristics of Dunhuang murals, we proposed a two generators connected image restoration networks (SeparaFill) based on U-Net and an accumulation feature

extraction mechanism to reuse the low-level features of images effectively. Firstly, the contour of the image is completed, and then the completed image contour is used to guide the restoration of color mural image. Experimental results indicate that our algorithm is very effective and possesses outstanding inpainting performance for both the objective comparisons and the visual performance. Compared with recent algorithms, our algorithm is also performed well in the large-scale damaged image restoration and complex texture structures.

However, the contour line extracted from mural image is not very distinct because the mural image is extracted directly on the HED network trained by the mural datasets. When extracting the contour line, there is no improvement of network and neither modify the parameters of network. At the same time, the image source comes from the electronic scanning version of the mural album, which has greatly influence on the quality of the mural image in the process of printing and scanning. The pixel blocks at the edge of the damaged area are badly jagged, so it is very difficult to cover them one by one during the manual mask labels. Due to the above problems, the inpainting of some local details of line structures is not very ideal. In the future, we will search for high-quality mural images to create datasets, and the contour extraction algorithm will be improved to obtain the distinctly contour of mural images. For large damaged mural images, the interactive manual assistance will be provided to restore line details and labeled more accurately to reduce the adverse effects caused by inaccurate calibration.

Abbreviations

GAN: Generative adversarial network; ResNet: Residual Networks; Res2Net:: Hierarchical Residual Networks; VGG :Visual Geometry Group Network.

Availability of data and materials

All data for analysis in this study are included within the article.

Competing interests

The authors declare that they have no competing interests

Funding

None.

Authors' contributions

All the authors contributed to the current work. S.YH and L.ZL devised the study plan and led the writing of the article. L.JH and Z.J arranged the data of experiment. S.YH and L.CH supervised the entire process and provided constructive advice. All authors read and approved the final manuscript.

Acknowledgments

None.

Authors' information

1School of Information and Communication Engineering, Communication University of China, Beijing 100000, China

References

1. Fan JS. For the long-term existence of Dunhuang -Dunhuang Grotto protection exploration. Dunhuang research.2004;111:35-9.
2. Cao JF et al. Repair of ancient temple view murals by local search algorithms for adaptive sample blocks. Journal of Computer-aided Design and Graphs. 2019;31:112030-2037.

3. Liu YC. Study on Yunnan mural restoration based on the improved Criminisi algorithm. Disseration. Yunnan University, Yunnan, 2017.
4. Yin Y. Application of digital image repair technology in barley field rock painting repair. Disseration. Ningxia University, Ningxia, 2019.
5. Wang N, Wang W, Hu W, et al. Thanka Mural Inpainting Based on Multi-scale Adaptive Partial Convolution and Stroke-like Mask. IEEE Transactions on Image Processing, 2021;30:3720-3733.
6. Zhang MF. Computer-aided Mural Restoration Study Based on generative Adversarial Network. Disseration. Northwest University, Shannxi,2019.
7. Wen LL, Xu D, Zhang X, et al. Some irregular partial repair method of ancient mural based on generative model. Graphic Journal. 2019;5:925-931.
8. Goodfellow I. J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. Advances in Neural Information Processing Systems. 2014;3:2672-2680.
9. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer, Cham. 2015; doi:10.1007/978-3-319-24574-4_28.
10. Gao S, Cheng M. M, Zhao K, et al. Res2Net: A New Multi-scale Backbone Architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019;43:2652-662.
11. Rudin L. I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms.Physica D Nonlinear Phenomena. 1992; doi:10.1016/0167-2789(92)90242-F.
12. Zhang G,Yan Y. Image repair of a fractional TV model combining the texture structure. Chinese graphic Journal. 2019;24:5700-713.
13. Criminisi A,Pérez P,Toyama K. Object Removal by Exemplar-Based Inpainting. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003;doi: 10.1109/CVPR.2003.1211538.
14. Liu X. Image and video repair based on the hybrid Criminisi algorithm. Disseration. Shandong University, Shandong,2020.

15. Yeh R. A, Chen C, Lim T. Y, et al. Semantic Image Inpainting with Deep Generative Models. IEEE Conference on Computer Vision and Pattern Recognition. 2017;1:6882-6890.
16. Zhang H, Hu Z, Luo C, et al. Semantic Image Inpainting with Progressive Generative Networks. acm multimedia. 2018;1939-1947.
17. Liu G, Reda F. A, Shih K. J, et al. Image Inpainting for Irregular Holes Using Partial Convolutions. European Conference on Computer Vision. 2018;doi:10.1145/3240508.3240625.
18. Guo Z, Chen Z, Yu T, et al. Progressive Image Inpainting with Full-Resolution Residual Network. in Proceedings of the 27th ACM International Conference on Multimedia (MM '19). 2019;doi:10.1145/3343031.3351022.
19. Yu J, Lin Z. Yang J, et al. Generative Image Inpainting with Contextual Attention. CVF Conference on Computer Vision and Pattern Recognition IEEE. 2018;2:5505-5514.
20. Nazeri K, Ng E, Joseph T, et al. EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. ICCV Workshops. 2019; arXiv.1901.00212.
21. Zamir S. W, Arora A, Khan S, et al. Multi-Stage Progressive Image Restoration. Computer Vision and Pattern Recognition IEEE. 2021;14821--14831.
22. Yu J, Lin Z, Yang J, et al. Free-Form Image Inpainting with Gated Convolution.International Conference on Computer Vision. IEEE; 2019;4470-4479.
23. Xiong W, Yu J, Lin Z, et al. Foreground-Aware Image Inpainting. Computer Vision and Pattern Recognition. IEEE;2019;1:5833-5841.
24. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer Vieweg. Berlin, Heidelberg;2015;doi:10.1007/978-3-662-54345-0_3.
25. Li H, Wang W, Guo L, Zhou L. Deep feature rearrangement image repair algorithm based on a dual-transfer network. Journal of Huazhong University of Science and Technology (Natural Science Edition).2021;49:774-781.

26. Yang H, Yu Y. Image repair using channel attention with hierarchical residual networks. Journal of Computer-aided Design and Graphs.2021;33:5671-681.
27. Jo Y, Park J. SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color. International Conference on Computer Vision: ETRI;2019;1:1745-1753.
28. Liu H, Jiang B, Song Y, et al. Rethinking Image Inpainting via a Mutual Encoder-Decoder with Feature Equalizations. European Conference on Computer Vision Lecture Notes in Computer Science.Springer. Cham. 2020;12347:725-741.
29. Xie S, Tu Z. Holistically-Nested Edge Detection. IEEE International Conference on Computer Vision (ICCV). 2015;doi:10.1109/ICCV.2015.164.
30. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE;2016;770-778.
31. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. Advances in Neural Information Processing Systems.2017;6000-6010.
32. Chopra S, Hadsell R, Lecun Y. Learning a similarity metric discriminatively, with application to face verification. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Piscataway: IEEE;2005;1:539-546.
33. Johnson J, Alahi A, Fei-Fei L.Perceptual Losses for Real-Time Style Transfer and Super-Resolution. European Conference on Computer Vision, Springer, Cham.2016;doi: 10.1007/978-3-319-46475-6_43.
34. Zhao H, Gallo O, Frosio I, et al. Loss Functions for Neural Networks for Image Processing. Computer Science. 2015.
35. Yu T, Guo Z, Jin X, et al. Region Normalization for Image Inpainting. In: the 34th AAAI Conference.2019.p.12733-12740.
36. Li J, Wang N, Zhang L, et al. Recurrent Feature Reasoning for Image Inpainting. Conference on Computer Vision and Pattern Recognition (CVPR).2020.p. 7757-7765.