The Visual Computer
https://doi.org/10.1007/s00371-022-02592-1

ORIGINAL ARTICLE

Multi-scale self-attention generative adversarial network for pathology image restoration

Meiyan Liang[1] · Qiannan Zhang[1] · Guogang Wang[1] · Na Xu[1] · Lin Wang[2,3] · Haishun Liu[4] · Cunlin Zhang[4]

Accepted: 9 June 2022

Abstract
High-quality histopathology images are significant for accurate diagnosis and symptomatic treatment. However, local cross-contamination or missing data are common phenomena due to many factors, such as the superposition of foreign bodies and improper operations in obtaining and processing pathological digital images. The interpretation of such images is time-consuming, laborious, and inaccurate. Thus, it is necessary to improve diagnosis accuracy by reconstructing pathological images. However, corrupted image restoration is a challenging task, especially for pathological images. Therefore, we propose a multi-scale self-attention generative adversarial network (MSSA GAN) to restore colon tissue pathological images. The MSSA GAN uses a self-attention mechanism in the generator to efficiently learn the correlations between the corrupted and uncorrupted areas at multiple scales. After jointly optimizing the loss function and understanding the semantic features of pathology images, the network guides the generator in these scales to generate restored pathological images with precise details. The results demonstrated that the proposed method could obtain pixel-level photorealism for histopathology images. Parameters such as RMSE, PSNR, and SSIM of the restored image reached 2.094, 41.96 dB, and 0.9979, respectively. Qualitative and quantitative comparisons with other restoration approaches illustrate the superior performance of the improved algorithm for pathological image restoration.

Keywords Multi-scale · Self-attention · Generative adversarial network · Pathological image restoration

1 Introduction

Histopathology examination is the gold standard of disease diagnosis. Pathologists' workflow is to observe the macroscopic characteristics and microstructure of the tissue through microscopy at different magnifications, and conduct a comprehensive assessment of the disease. Therefore, high-quality pathology images are of great significance for accurate diagnosis and symptomatic treatment. However, local cross-contamination or partial missing information are common phenomena due to many factors, such as the superposition of foreign bodies and improper operations in obtaining and processing pathological digital images. In these scenarios, corrupted histopathology images will have severe impacts on diagnosis accuracy. Therefore, it is necessary to restore corrupted pathological images to predict diseases.

Image inpainting reconstructs the missing part of the corrupted image so that it has good consistency with the surrounding features, ensuring that the reconstructed image is indistinguishable from the original image. However, the feasibility and effectiveness of restoration operations depend on the image type or application domain. Recently, many classical image restoration algorithms have been proposed to reconstruct natural images [1–6]. These algorithms are sequentially based and classified into two categories:

✉ Meiyan Liang
meiyanliang@sxu.edu.cn

✉ Guogang Wang
kingguogang@sxu.edu.cn

1 School of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China

2 Shanxi Bethune Hospital, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Shanxi Academy of Medical Sciences, Taiyuan 030032, China

3 Tongji Medical College, Tongji Hospital, Huazhong University of Science and Technology, Wuhan 430030, China

4 Beijing Key Laboratory for Terahertz Spectroscopy and Imaging, Key Laboratory of Terahertz, Optoelectronics, Capital Normal University, Ministry of Education, Beijing 100048, China

Published online: 17 August 2022

diffusion-based image inpainting and patch matching. In 2000, Bertalmio et al.[7] proposed an inpainting algorithm that does not require any user intervention when restoring the selected region. The algorithm uses the Laplacian operator to obtain the gradient of the repaired boundary and performs inpainting iterations in the isophote direction. This method can restore images by gradually extending the known information from the border to the missing region, which has already achieved good performance on small missing areas, such as restoring old photographs and damaged films. However, it is inapplicable for repairing images with large missing regions. In 2009, Barnes et al.[8] presented interactive image completion tools using a new randomized algorithm to search candidate patches in corrupted images and propagate such matches quickly to surrounding areas according to the coherence of natural images. The result is better when the complement region has structural consistency with surrounding regions, but it cannot generate new content. Huang et al.[9] proposed an improved image inpainting algorithm, which can deal with the mismatched texture blocks in the repairing process. However, this method is not suitable for repairing images with apparent textures.

2 Literature review

Classical image restoration algorithms are mainly based on mathematical and physical models. They require appropriate information in the input image, such as similar pixels, structures, or patches. The deep generative model solves the technical bottleneck of image inpainting by understanding the deep structure of the image. The deep generative model can intelligently control the image generation process by defining the corresponding loss functions. The restored image can be more realistic and reasonable using deep-learning based image inpainting [10]. Image restoration based on deep learning mainly includes CNN-based models[11–17] and GANs [18–24]. In 2016, Pathak et al.[25] proposed a context encoder (CE) to reconstruct images by semantic inpainting, which uses reconstruction loss and adversarial loss to guide the image generation process. The peak signal-to-noise ratio (PSNR) of the restored image is 17.59 dB on the Paris StreetView dataset. CNN-based model also includes the image inpainting frameworks based on partial convolution and gated convolution. In 2018–2019, Liu et al.[26] and Yu et al.[27] proposed to use partial convolution and gated convolution to replace standard convolution in the image inpainting process. But these methods cannot handle images with subtle features, such as pathological images.

In 2017, Iizuka et al.[28] introduced a globally and locally consistent image completion (GLCIC) network architecture to restore a wide variety of image scenes and ensure the consistency of the generated images. The GLCIC network uses two context discriminator networks to control the fully convolutional network (FCN) to generate restored images. Compared with CE, the PSNR of the restored image can be improved by more than 2 dB. In 2019, Nazeri et al. [29] developed a new approach called EdgeConnect, which first generates the edge of the missing region and then uses the edge information as a prior to restore the corrupted image. The approach not only solves the problems of blurriness and distortion in natural image restoration but also improves the PSNR and structural similarity index measure (SSIM) score of the restored image to 24.92 dB and 0.861, respectively. However, the restoration effect is limited when the input image has highly textured regions or large areas of image corruption. In 2020, Zhao et al.[30] proposed a UCT-GAN for image inpainting, which significantly reducing the possibility of mode collapse. But it requires a priori information of the data distribution to constraint the inpainting process. Zheng et al.[31] present a dense multi-scale fusion network (DMFN) for fine-grained image inpainting, which focuses on uncertain regions and enhances semantic details by designing a self-guided regression loss. This method can significantly improve the quality of produced images. Li et al. [32] proposed a network called GLAGAN for image completion. The network adds a cross-attention mechanism in GLCIC to obtain the correlation between the restored area and the known area. Compared with GLCIC, the PSNR of image can be improved by ~ 3.7 dB. In 2020, a coarse network and a refinement network (CNRN) is developed by Uddin et al.[23], which provide local-level refinement and global-level structural consistency using mask pruned attention mechanism. The most relevant work is classified and summarized in Table 1.

At present, almost all the existing image restoration methods are for natural images. However, the urgent demand for medical image restoration promotes the development of deep learning in the medical field[33–40]. In 2020, Armanious et al. [41, 42] proposed an ip-MedGAN and ipA-MedGAN for computed tomography (CT) and magnetic resonance imaging (MRI) image inpainting. The results showed that the PSNR and SSIM scores were enhanced to 30.62 and 0.9606 using ipA-MedGAN without prior localization of the regions of interest. However, no study has been reported on corrupted pathological image reconstruction. This is because pathological images have vital differences from natural images, CT images, and MRI in image distribution. Pathology images have more complex structures and more subtle features, which poses a great challenge for reconstruction methods. Classical image restoration algorithms such as diffusion-based image restoration and patch matching are inapplicable for pathological images. Moreover, some image restoration methods based on deep learning are not highly effective for

Table 1 A summary of the related works presented in this study

References	Methodology	Performance evaluation metrics	Quantitative index	Limitations
Bertalmio et al.[7]	Diffusion- based image inpainting and patch matching	Visual effect	/	It is inapplicable for repairing images with large missing regions
Barnes et al.[8]		Visual effect	/	This algorithm can complete large missing regions when the complement region has structural consistency with surrounding regions, but it cannot generate new content
Huang et al.[9]		PSNR	46.49 ~ 54.59 dB	This method is not suitable for repairing images with apparent textures
Pathak et al.[25]	CNN-based image inpainting	Mean L1 Loss, Mean L2 Loss, PSNR	10.33% 2.35% 17.59 dB	This method lacks fine texture and detail information when repairing high-resolution images, and repairing artifacts are generated around the corrupted area
Liu et al.[26]		L1 error, PSNR, SSIM, Inception Score	0.49% ~ 5.72% 19.74 ~ 33.75 dB 0.484 ~ 0.946 0.051 ~ 1.588	Image inpainting using partial convolution will treat the corrupted areas with different pixel values equally, causing the network to be less sensitive to missing areas
Yu et al.[27]		Mean L1 error, Mean L2 error	8.6% ~ 9.1% 1.6% ~ 2.0%	The result depends on the selection of gate function used in gated convolution
Iizuka et al.[28]	GAN-based image inpainting	Naturalness	77.0%	This method lacks consideration of image texture details, and cannot reconstruct images with complex textures
Zhao et al.[30]		PSNR, SSIM, IS, MIS, L1 error, LPIPS	26.38 dB 0.8862 3.0127 0.0178 1.51 0.030	This method requires a priori information of the data distribution
Zheng et al.[31]		PSNR, SSIM, LPIPS	26.50 dB 0.8932 0.0460	DMFN adopts self-guided regression loss to focus on correcting semantic structure errors of the corrupted area, while ignoring the uncorrupted area of the image. The restored image shows repaired artifacts in the boundary
Li et al. [32]		PSNR, SSIM	31.84 ~ 31.54 dB 0.9472 ~ 0.9502	GLAGAN ignores the relationship between pixels of the missing area, resulting in lower PSNR and SSIM scores in this area
Ours	GAN-based image inpainting	PSNR, SSIM, RMSE	41.96 dB 0.9979 2.094	Image completion can only be performed for rectangular missing areas

pathology image restoration, especially when a large portion of the pathology image is corrupted. Here, we propose a multi-scale self-attention generative adversarial network (MSSA GAN) to restore two types of contaminated colonic pathological images. The main contributions of the paper are summarized as follows:

(1) We propose a regional self-attention module to restore the corrupted image. This module focuses on the restoration of the corrupted area by establishing the correlation between the corrupted and uncorrupted area, as well as the correlation within the corrupted area through two parallel weighted branches.
(2) The MSSA GAN uses a regional self-attention mechanism in the generator to efficiently learn the correlations between the corrupted region and all other positions at

multiple scales, which enables the model to accurately reconstruct image details without inpainting artifacts around the corrupted area.

3 Methodology

3.1 Proposed multi-scale self-attention generative adversarial network (MSSA GAN)

Histopathology image is composed of a large number of cells with different shapes. it shows more texture features than natural images in a macroscopic view. Therefore, compared with natural image inpainting, pathological images inpainting is a more challenging task in deep learning. Therefore, a multi-scale self-attention generative adversarial network is proposed to restore pathological images. The block diagram of the network is shown in Fig. 1.

The framework of MSSA GAN contains two parts: generator and discriminator. The backbone of the generator is an AutoEncoder, which can reconstruct corrupted images based on the gradient of the discriminator. The discriminator determines whether the input is the actual pathological image or generated, and provides the gradient for the generator through the backpropagation algorithm. The MSSA GAN network achieves convergence through the game between the generator and the discriminator.

3.1.1 Generator

In our MSSA GAN, the backbone of the generator is an AutoEncoder, which consists of two parts: an encoder and a decoder. The encoder part consists of 6 convolutional layers, followed by batch normalization and a LeakyReLU activation function with the slope of 0.2. After the encoding operation, the input image is encoded by different convolutional layers into feature maps, the size of which are 128 $\times$ 128 $\times$ 64, 64 $\times$ 64 $\times$ 128 and 32 $\times$ 32 $\times$ 256, respectively. The output of the encoder is treated as a latent feature representation. Then, four dilated convolution layers (kernel size 3 $\times$ 3) with dilated rates 2, 4, 8, 16 are applied to the latent feature representation (32 $\times$ 32 $\times$ 256) to extract semantic information of the image, which are treated as the bottleneck of the AutoEncoder[43]. This technique can effectively capture the high-level features of a larger area without increasing the computational complexity. After semantic feature extraction and representation, the network decodes the semantic feature representation to reconstruct the corrupted image. The architecture of the decoder network and encoder network is almost symmetrical to each other. Decoder uses a deconvolution layer to upsample the obtained semantic feature representations. Finally, the Tanh function is applied in the output layer. Typically, we added a multi-scale image restoration module to extract features from different decoder layers to form output images of different sizes. This operation not only captures more detailed information of the image from different scales but also ensures the stability of the network generation process. In our model, the three scales are selected, which are128 $\times$ 128, 64 $\times$ 64, and 32 $\times$ 32. This is also why we call our model multi-scale self-attention generative adversarial network (MSSA GAN). In our framework, regional self-attention module and skip-connection technique are added on each scale of the generator to achieve better image restoration results and prevent mode collapse.

3.1.1.1 Regional self-attention module (RSA module)

Self-attention, also known as intra-attention, is an attention mechanism relating to different parts of an image to restore the corrupted image. It has been proven to be remarkably effective in GANs, which enable both the generator and the discriminator to better model relationships between spatial regions of an image [44]. However, the drawback of using self-attention is that all pixel values in the feature map have dependencies. It is insufficient to integrate the original self-attention mechanism into global attention due to the presence of mask values. Therefore, we propose a regional self-attention module to restore the corrupted image. This module focuses on the restoration of the corrupted area by establishing the correlation between the corrupted and uncorrupted area, as well as the correlation within the corrupted area through two parallel weighted branches, and the weights are adjusted adaptively according to the inpainted image. The architecture of the module is shown in Fig. 2., which is added to the generator(decoder) for efficient image restoration.

As shown in Fig. 2., the corrupted area is treated as the foreground and the uncorrupted area as the background. In the regional self-attention module, two branches are included. When restoring foreground features, we obtain reference information from the background. For instance, the texture, color, and style information of the foreground should be consistent with the background. This is the function of branch 1. Furthermore, when the corrupted area is large and not located in the center of the image, the relations between different parts of the foreground can also be considered as a complementary to ensure the consistency of the statistical features in the corrupted area. For instance, the similarity of the cells in the foreground can be utilized as a supplementary information in image restoration. This can be achieved with the help of branch 2.

The feature map $\mathbf{F}(\boldsymbol{x})$ ($M \times N \times C$) after convolution is divided into foreground $\mathbf{F}_f$ ($m \times n \times C$) and background $\mathbf{F}_b$ according to the size of the mask. Then, the foreground tensor $\mathbf{F}_f$ and background tensor $\mathbf{F}_b$ are reshaped into $P \times C$ and $P' \times C$, respectively. Where, P and P' are the number

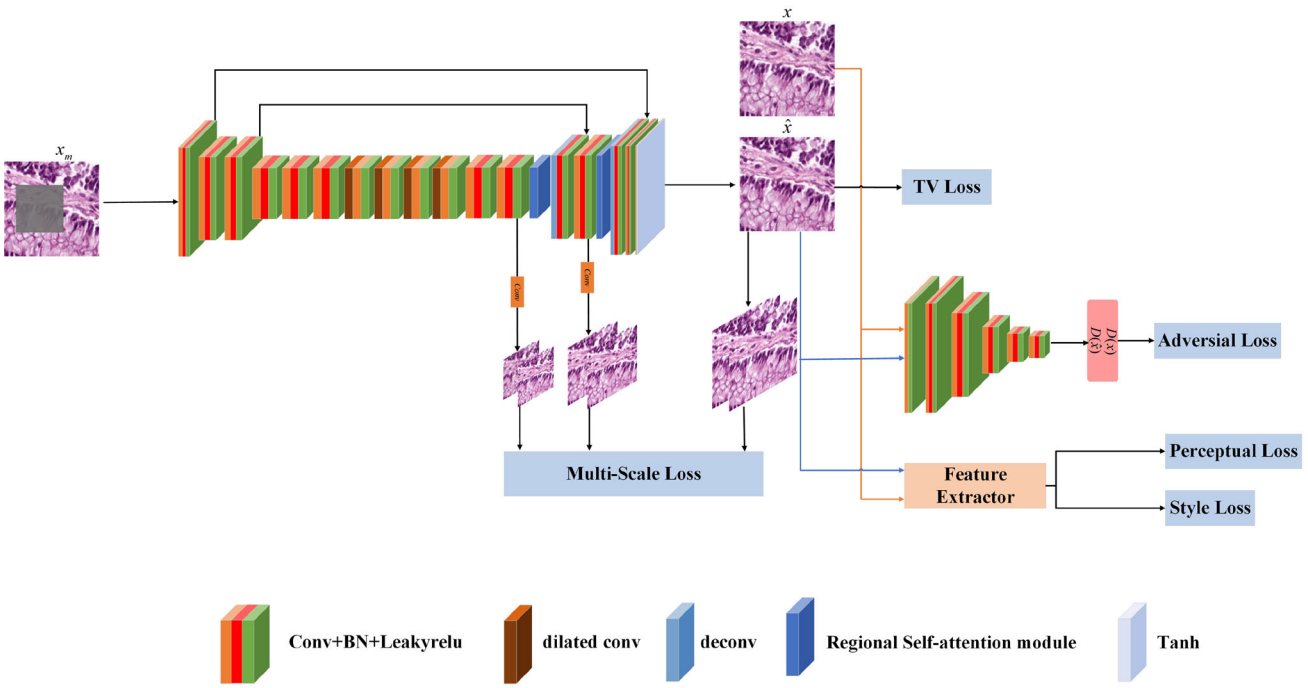

Fig. 1 Block diagram of MSSA GAN

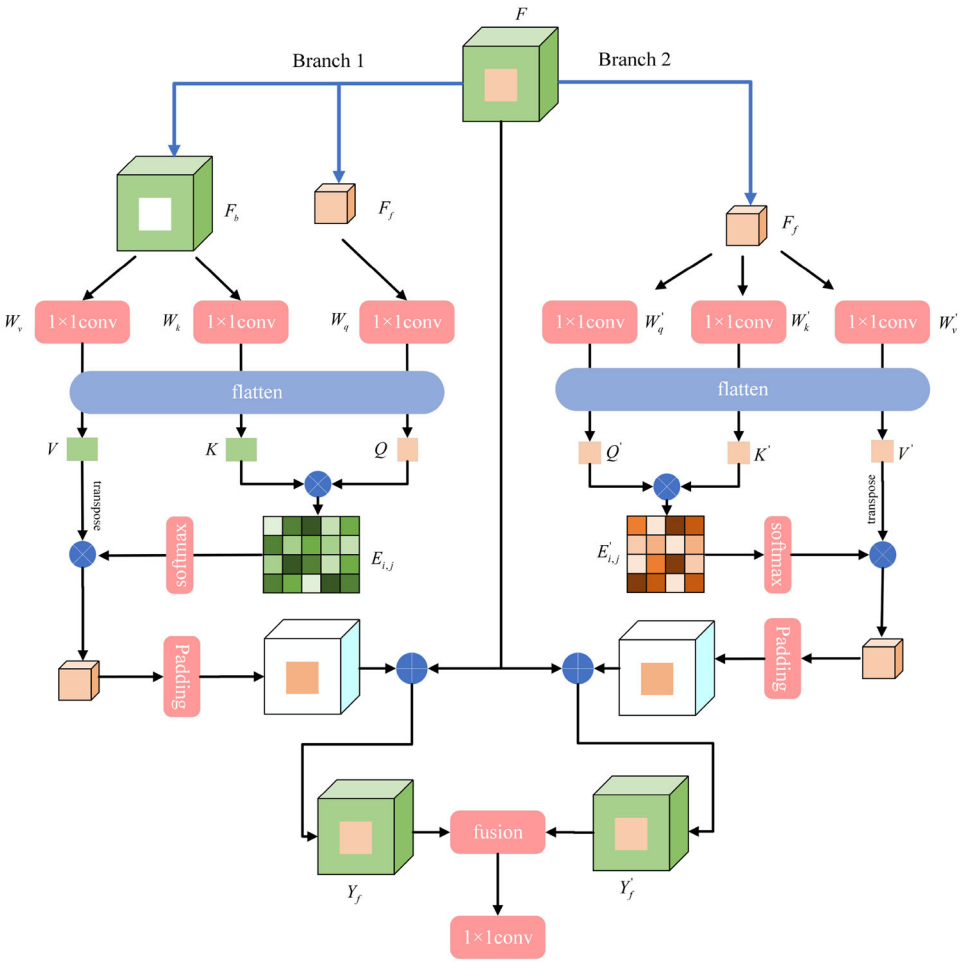

Fig. 2 Regional self-attention module

of foreground pixels and background pixels in $\mathbf{F}_f$ and $\mathbf{F}_b$, respectively. Here, C is the number of channels.

$$P = m \times n \tag{1}$$

$$P' = (M \times N) - (m \times n) \tag{2}$$

In branch 1, the foreground $\mathbf{F}_f$ and background $\mathbf{F}_b$ are transformed into three representations in the feature space, corresponding to the concepts of the query($\boldsymbol{Q}$), key ($\boldsymbol{K}$), and value ($\boldsymbol{V}$).

$$Q = W_q F_f \tag{3}$$

$$K = W_k F_b \tag{4}$$

$$V = W_v F_b \tag{5}$$

where, W_q, W_k and W_v are three 1×1 convolution operations for feature transformation of foreground and background in branch 1, which are the learnable parameters of the regional self-attention module;

The attention map of the foreground and background can be obtained by

$$E_{i,j} = \frac{exp(Q_i^T, K_j)}{\sum_{j=1}^{P'} exp(Q_i^T, K_j)} \tag{6}$$

Here, $E_{i,j}$ denotes the attention score of feature representation of foreground ($\boldsymbol{Q}$) and background ($\boldsymbol{K}$) in branch 1, which depicts the importance of the $\boldsymbol{V}$. The superscript T represents the transpose operation.

The output of branch 1 can be written as

$$Y_f = \beta_1 pad(V^T \otimes E) + F \tag{7}$$

where $pad(\bullet)$ denotes zero-padding operation. $\otimes$ is the matrix multiplication operation. Y_f represents an enhanced feature map using the reference information from the background.

In branch 2, only the foreground $\mathbf{F}_f$ is transformed and branched out into three copies—Q' K' and V'.

$$Q' = W_q^{'} F_f \tag{8}$$

$$K' = W_k^{'} F_f \tag{9}$$

$$V' = W_v^{'} F_f \tag{10}$$

where, $W_q^{'}$, $W_k^{'}$ and $W_v^{'}$ are three 1×1 convolution operations for feature transformation of foreground in branch 2, which are also the learnable parameters of the regional self-attention module;

The foreground self-attention map is obtained by

$$E\prime_{i,j} = \frac{\exp\left(Q\prime_i^T, K\prime_j\right)}{\sum_{j=1}^{P}\left(\exp(Q\prime_i^T, K\prime_j\right)} \tag{11}$$

Here, $E\prime_{i,j}$ denotes the attention score of feature representation $\boldsymbol{Q'}$ and $\boldsymbol{K'}$ in branch 2, which depicts the importance of the $\boldsymbol{V'}$.

The output of branch 2 can be written as

$$Y\prime_f = \beta_2 pad(V\prime^T \otimes E\prime) + F \tag{12}$$

Here, $Y\prime_f$ represents an enhanced feature map using the information from the foreground itself.

In Eqs. (7) and (12), β_1 and β_2 are the weights of the two branches, which are learnable parameters in the regional self-attention module. Scenario 1) when the corrupted area is relatively small, the pathological image can be reconstructed mainly by using the background information. In this case, the weight of branch 2 (β_2) is smaller than the weight of branch 1 (β_1). Scenario 2) when the corrupted area is increase or not located in the center of the image, the decrease in the valid pixel ratio in the background will result in an imbalance restoration of the corrupted image. Thus, the information in the background area is insufficient to provide a reference for pathological image reconstruction. In this case, the relations between different parts of the foreground should be considered as supplementary information to restore the pathological image. The supplementary information can be the similarity of cells, textures, and glands in the corrupted area. Therefore, the weight of branch 2 will become larger in this scenario.

Finally, the refined feature map can be obtained by

$$Y = \text{conv}\ (Y_f + Y\prime_f) \tag{13}$$

Here, conv($\bullet$) represents a 1×1 convolution. This operation can fuse the features of the two branches in the regional self-attention module to obtain a refined feature map Y, which can be regarded as fine feature enhancement.

3.1.1.2 Skip-connection Skip Connection is an effective technique to accelerate convergence and prevent gradient vanishing of the deep neural networks, which is proposed by Kaiming He [45, 46]. For GANs, mode collapse is a common phenomenon in the training process [47, 48], especially for pathological image inpainting tasks, because more similar patterns are contained in the pathological image database. Therefore, skip connections are added in the generator to connect the layers in the encoder network directly with the corresponding layers deep in the decoder network. This

technique allows the low-level features in the encoder to be propagated directly to the decoder. The technique can increase the sensitivity and effectiveness of the self-attention mechanism in the generator, thereby improving the quality of the restored images and avoiding model collapse.

3.1.2 Discriminator

The discriminator is treated as a binary classifier, which attempts to discriminate whether the restored image $\hat{x}$ is actual or generated. The discriminator of MSSA GAN is designed with six convolutional layers. All convolution operations use a 4×4 kernel with the stride of 2×2 to decrease the size of feature representations. Davide et al.[49] demonstrate that the inpainting results can be improved by increasing the capacity of the discriminator network. Therefore, the depth of every convolutional layer is doubled compared with the configuration of the decoder (in generator). The number of channels in the discriminator is set to 128, 128, 256, 512,1024, and 1, respectively. The output is transformed into a probability score after the sigmoid function, which indicates whether the corrupted image is restored.

3.2 The joint loss function

As shown in Fig. 1., the network optimizes the five joint loss functions, including multi-scale losses, adversarial loss, perceptual loss, style loss, and TV loss, to obtain the MSSA GAN model.

3.2.1 multi-scale losses (M-S losses)

For the multi-scale losses, we extract features from different decoder layers to form outputs in different scales. By adopting this, we intend to capture more detailed information in these scales.

We define multi-scale loss function as:

$$L_m(\{S\}, \{T\}) = \sum_{i=1}^{m} \lambda_i L_1(S_i, T_i) \tag{14}$$

Here,

$$L_1 = \|G(x_M) - x\|_1 \tag{15}$$

$$x_M = M \odot x \tag{16}$$

where, x_M and x denote the masked pathological image and original image, respectively. M is a squared mask that represents a corrupted region.

Therefore,

$$L_m = \sum_{i=1}^{m} \lambda_i [\|S_i(G(x_M)) - T_i(x)\|_1] \tag{17}$$

where, S_i represents the i th scale output image extracted from the decoder. And T_i represents the ground truth of the same image in the i th scale. $\{\lambda_i\}_{i=1}^{m}$ are the weights of the scales. To obtain more detailed restored images, we set a higher weight on a larger scale. Specifically, the outputs of the last 1st, 3rd, and 5th layers are used, whose sizes are 1/16, 1/4, and 1 of the original size, respectively. Thus, we set $\{\lambda_i\}_{i=1}^{m}$ to $\{0.6, 0.8, 1.0\}$. L_m is a multi-scale loss function in image restoration. Pathological images have more high-frequency features than natural images. So we prefer L_1-norm rather than L_2 to restore the images as it is not affected by the outliers. Multi-scale loss function can also ensure the stability of the network generation process, thereby obtaining more refined image details.

3.2.2 Adversarial loss

Adversarial loss is a crucial part of our MSSA GAN network. The generator uses the gradient of the discriminator for training, and the discriminator is updated by distinguishing the input image is actual or generated. Both the generator and discriminator are evolving in the adversarial training process until the Nash equilibrium. Therefore, every participant can optimize its outcome based on the decision of a rival in the min–max game.

The adversarial loss can be written as follows:

$$\min_G \max_D L_{adv} = E_{x \in \chi}[\log D(x) + \log(1 - D(G(M \odot x)))] \tag{18}$$

where, $G(\bullet)$ and $D(\bullet)$ denote the generator and discriminator network, respectively. $\odot$ is an element-wise multiplication operator.

3.2.3 Perceptual loss

Pathological images have more detailed features, which are significant in diagnosis. Thus, the perceptual loss is taken into the consideration in the generator to provide more accurate texture restoration results. In the implementation, the generated image is input to a VGG-16 feature extractor and the feature maps of each layer is compared with the ones corresponding to the ground truth image.

$$L_{\text{perceptual}}(x, \hat{x}) = \sum_{j=1}^{N} \left(1/H_j W_j C_j\right) \left\| \phi_j^{\text{gt}}(x) - \phi_j^{\text{pred}}(\hat{x}) \right\|_1 \tag{19}$$

In this paper, $\hat{x}$ and x denote the restored pathological image and ground truth, respectively. ϕ represents the VGG-16 network. Thus, $\phi_j^{\text{pred}}(\hat{x})$ and $\phi_j^{\text{gt}}(x)$ are the feature maps

extracted from the j-th layer of the generated image and corresponding ground truth, respectively. Where H_j, W_j, C_j refer to the height, width, and channels number of the feature map extracted from the j-th layer. N is the number of layers in the VGG-16 feature extractor.

3.2.4 Style loss

Perceptual loss helps to capture high-level structure, but it still cannot preserve style consistency. Style differences in color, texture, and common patterns should be penalized. Hence, style loss is employed as a part of our total loss function. In our work, style loss guided our model to learn the stylistic features of the target image.

$$\begin{aligned} L_{\text{style}}(x, \hat{x}) &= \sum_{j=1}^{N} \left\| \text{Gr}_j^{\phi}(\phi_j^{\text{gt}}(x)) - Gr_j^{\phi}(\phi_j^{\text{pred}}(\hat{x})) \right\|_1 \\ &= \sum_{j=1}^{N} \left(1/H_j W_j C_j\right) \left\| \phi_j^{\text{gt}}(x) - \phi_j^{\text{pred}}(\hat{x}) \right\|_1 \end{aligned} \quad (20)$$

where Gr_j^{ϕ} denotes the Gram matrix of the input image in the j-th layer using the VGG-16 network. Here, we use the L_1-norm in pathological image restoration.

3.2.5 Total variation loss (TV loss)

Total variation loss (L_{TV}) is applied to constrain the spatial smoothness of the generated image [50].

Therefore, the joint loss function of the MSSA GAN network can be given by:

$$L = \alpha_1 L_m + \alpha_2 L_{\text{adv}} + \alpha_3 L_{\text{perceptual}} + \alpha_4 L_{\text{style}} + \alpha_5 L_{\text{TV}} \quad (21)$$

where $\{\alpha_1, ..., \alpha_5\}$ is a group of tradeoff hyperparameters for weighing the joint loss function. Here, we set $\{\alpha\}$ to {100, 10, 1, 1, 1} in the implementation.

The parameters of the network can be optimized as follows:

$$\theta_g, \theta_d = \arg\min(L) \quad (22)$$

The restored image can be expressed as:

$$\hat{x} = G(x_M; \theta_g) \quad (23)$$

where θ_g represent the parameters of the optimized generator, and $\hat{x}$ is the restored pathology image using the proposed network.

Table 2 shows the corresponding pseudocode of the proposed method in the paper. The input of the network is corrupted histopathological image dataset x_M, which is 128 × 128 with a 64 × 64 square-shaped contaminated region. The output of the network is the restored images $\hat{x}$, the size of which is also 128 × 128.

4 Experiment

4.1 Database and implementation details

The histopathology images are two types of hematoxylin- and eosin-stained human colon tissues obtained from the National Cancer Center of Heidelberg and the Medical Center of Heidelberg University in Germany. These images were manually extracted from $N_0 = 86$ H&E stained human cancer tissue slides from formalin-fixed paraffin-embedded (FFPE) samples from the NCT Biobank. All images are 128 × 128 pixels at ~ 1 microns per pixel (MPP). All images are color-normalized using Macenko's method. The dataset contains 25,054 non-overlapping colon tissue image patches, including 9,504 normal colon mucosa (NORM) and 15,550 colorectal adenocarcinoma epithelium (TUM). Here, all the pathology images are applied for training and testing according to the ratio of 11:1. Therefore, 23,080 images are used in the training process, and 1,974 images are used to test the performance of MSSA GAN. The input of the MSSA GAN is locally contaminated pathological images, and the output is the corresponding restored images with high quality. In our case, the input of the network was a degraded 2D histological pathology image patch $x_M = M \odot x$, the size of which was 128 × 128 with a 64 × 64 square-shaped contaminated region (25% of the total image area). Where x is the original histological pathology image, and M represents the contaminated region mask. Here, M is set as 0.1 to simulate the situation that is seriously corrupted by foreign objects.$\odot$ is Hadamard's product. In the training process, The MSSA GAN uses a regional self-attention mechanism and multi-scale image restoration framework in the generator to efficiently learn the nonlinear mapping function between the corrupted region and the uncorrupted region at multiple scales. After jointly optimizing the loss function and understanding the semantic features of pathology images, the discriminator guides the generator to generate high-quality pathological images in different scales. Here, we use ADAM as the optimizer with an initialized learning rate of 0.0002 and betas {0.5; 0.999}. On a single Tesla V100 (16 GB), we train our model for 60 epochs with a batch size of 64. The size of the network output is also 128 × 128.

4.2 Results

Figure 3 shows the restoration results of locally contaminated histopathology images using MSSA GAN. The images

Table 2 The corresponding pseudocode of the MSSA GAN

1. Algorithm: Training procedure of the MSSA GAN
2. INPUT: $x_M = x \odot M$,/* x_M is the corrupted pathological images, M is the mask, x is ground truth images. */
3. OUTPUT: $\hat{x}$, /* the restored pathological image */
4. Procedure:
5. for number of training iterations **do**
6. Sample minibatch of k input samples $\{x_M^{(1)},\ldots,x_M^{(k)}\}$ from corrupted images x_M .
7. Sample minibatch of k ground truth samples $\{x^{(1)},\ldots,x^{(k)}\}$ from ground truth images x .
8. Update the generator :
9. $\theta_g \leftarrow \arg\min(10L_{adv} + 100L_m + L_{perceptual} + L_{style} + L_{TV})$
10. Update the discriminator:
11. $\theta_d \leftarrow \arg\max(10L_{adv})$
12. end for
13. $\hat{x} \leftarrow G(x_M;\theta_g)$

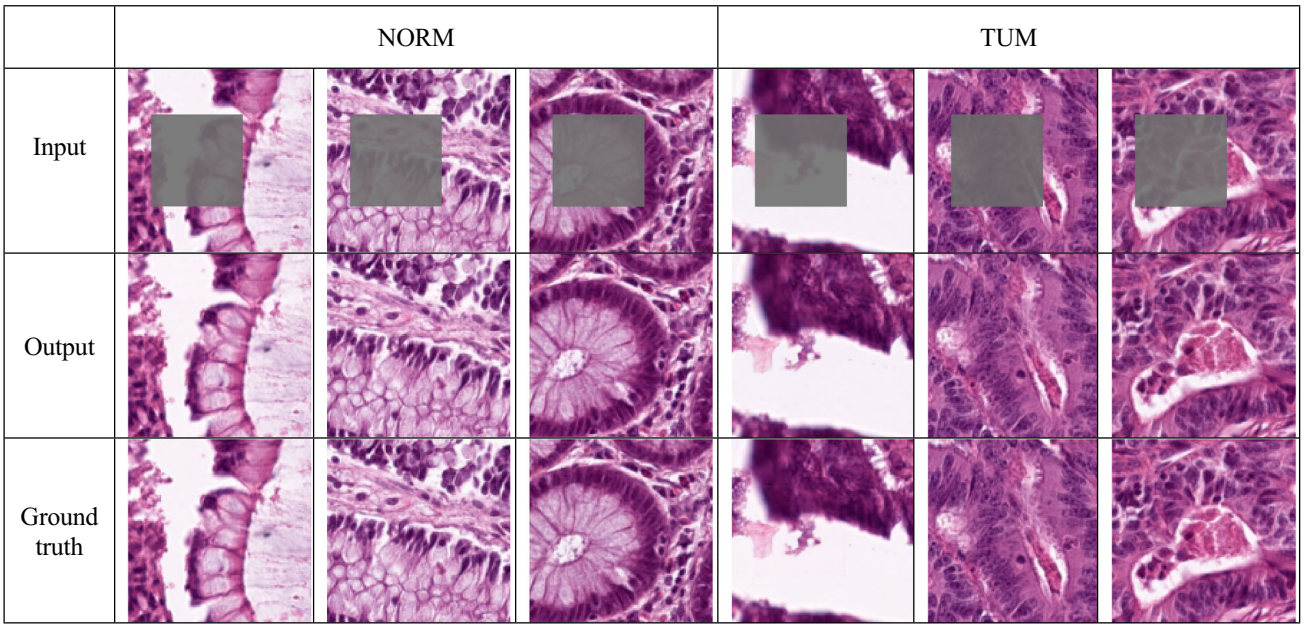

Fig. 3 Restoration results for colon pathological images (The top row is corrupted image patches, the second row is the output of the MSSA GAN, the bottom row is ground truth)

in the top row, second row, and bottom row are the locally degraded pathology images, restored images, and the ground truth, respectively. The results showed that the reconstructed images have good consistency with the ground truth in texture, color, and style. Particularly, there are almost no artifacts appeared in the resorted image. Indicators such as average PSNR and SSIM indexes after image restoration have reached 41.96 dB and 0.9979, respectively. The results indicate that the MSSA GAN can not only restore the.

pathological image in structure but also maintain the precise high-frequency textures. To better explain the image reconstruction process, the restoration results at different scales are also visualized in Fig. 4. The image of columns (a)–(c) represents the generated pathological images of different scales, the size of which is 32 × 32, 64 × 64, and 128 × 128, respectively. It is shown that the pathological images are reconstructed sequentially from coarse to fine as the scale increases. The results demonstrated that low-frequency components such as structure and context are recovered at small-scale (here is 32 × 32), and high-frequency features.

such as texture and image details are restored at large-scale (here is128 × 128). After image restoration using MSSA GAN, the results were peer-reviewed by pathologists from different medical centers. The pathology experts have reached a consensus on the effectiveness of the proposed method. We also used the image inpainting models such as CE, GLGIC, DMFN, EdgeConnect and CNRN to reconstruct the corrupted pathological image for comparisons. Qualitative and quantitative comparisons are conducted in Fig. 5. and Table 3.

5 Discussion

Currently, almost all the existing image restoration algorithms based on deep learning are practical for natural

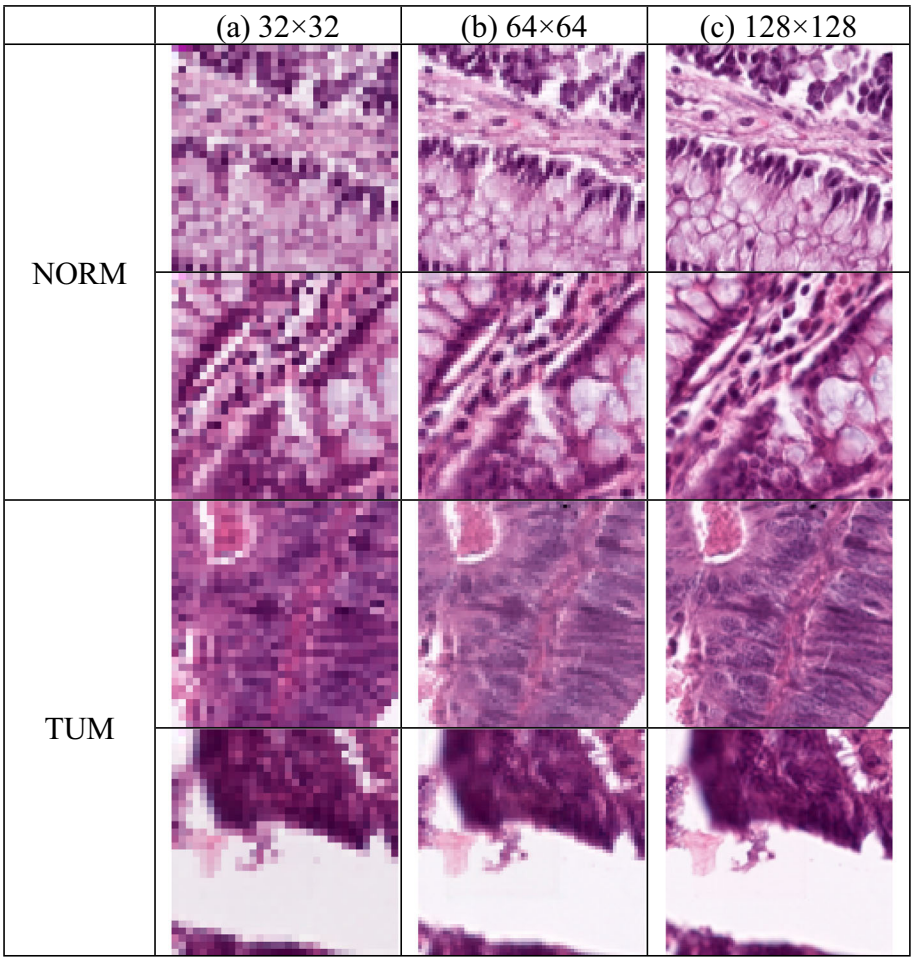

Fig. 4 Pathological image restoration results at different scales

images. Several frameworks have also been proposed for restoration, such as CT and MRI. However, no studies have reported on corrupted pathological image reconstruction because pathological images have many differences from natural images, CT images and MRI in image distribution. Natural images include relatively obvious semantic information, which can be inpainted by the canonical orientation, semantic integrity, and structural consistency features. Medical images such as CT and MRI have symmetrical properties and semantic information simultaneously, which can be inpainted by medical image distribution and symmetry as a generative prior. In contrast to these images, pathological images have no canonical orientation, structure consistency, or balance as a prior. The semantic information of pathological slides is rather challenging to obtain, meaning it takes a longer time for deep learning algorithms to find their way for image analysis. Furthermore, pathological images have more singularities and relatively detailed information, which increases the difficulty of image restoration. Therefore, MSSA GAN is proposed to restore pathological images with precise details. Figure 5. shows the qualitative restoration results of pathological images using the existing image inpainting models such as CE, GLCIC, DMFN, EdgeConnect and CNRN. These models are trained on the same pathological image dataset for comparisons.

For image inpainting, texture details of the filled pixels are significantly important, especially for pathological images. The pathological images restored by CE and GLCIC were partially blurred or lost coherence in corrupted areas, which will lead to diagnostic errors. When DMFN is applied, the restored image shows repaired artifacts in the boundary. This is because DMFN uses self-guided regression loss and a geometrical alignment constraint to focus on correcting semantic structure errors of the corrupted area, while ignoring the uncorrupted area of the image. Therefore, these methods are not effective for pathological image restoration, whereas for CNRN, since the undamaged regions are replaced by the corresponding regions in the original image, the restored image also shows artifacts at the mask boundary due to style inconsistency of the two images. The performance of MSSA GAN can restore the texture details with high precision, which is remarkably better than that of the CE, GLCIC, EdgeConnect, and DMFN do. Our framework mainly focuses on learning the exact relations between the corrupted and

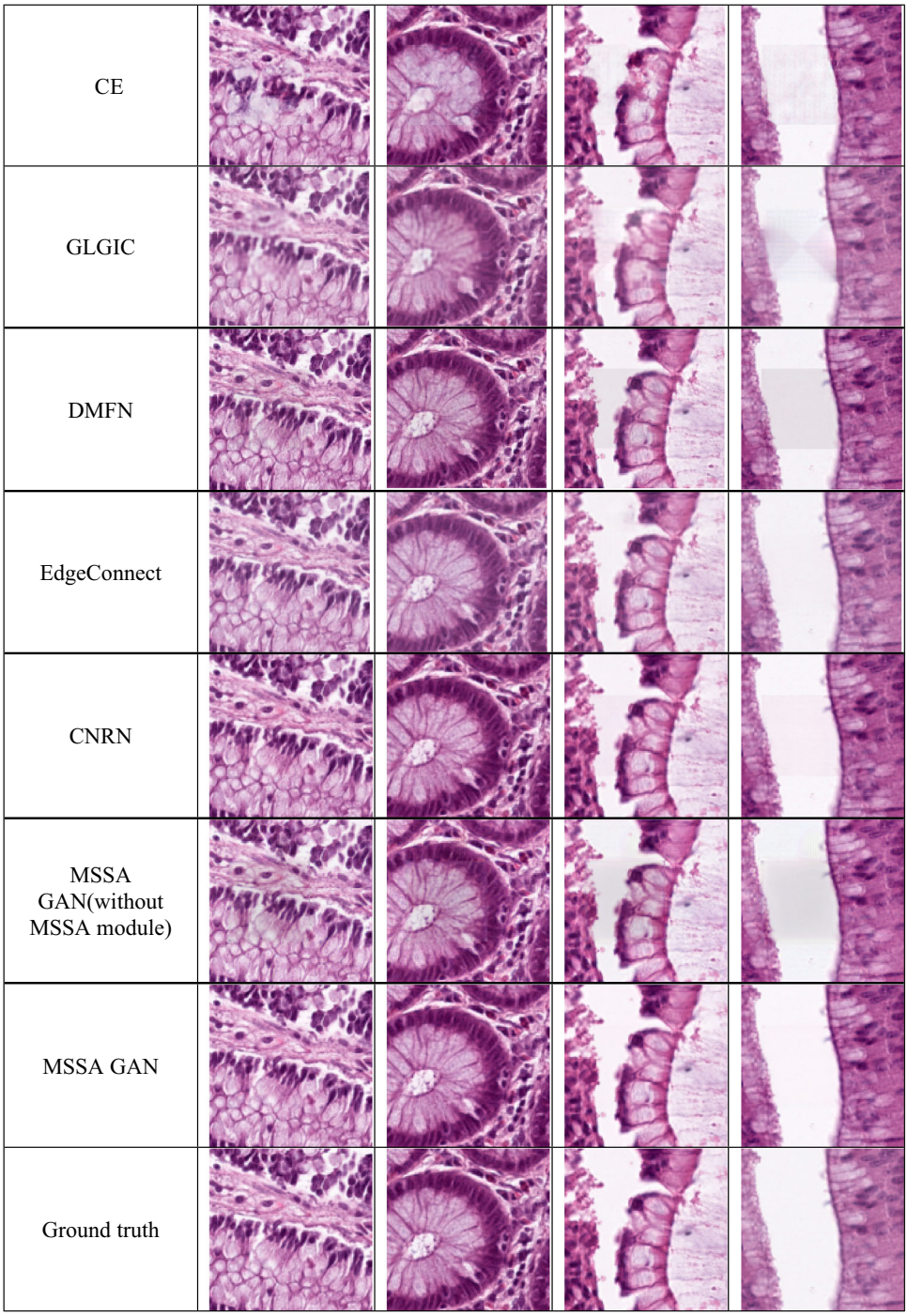

Fig. 5 Qualitative comparison of image restoration using the proposed method

Table 3 Comparison of image restoration results using different methods

Parameters	CE [25]	GLCIC [28]	DMFN [31]	EdgeConnect [29]	CNRN[23]	Without MSSA	MSSA GAN
PSNR	26.37	26.57	34.96	38.72	41.15	23.46	**41.96**
SSIM	0.8908	0.8932	0.9901	0.9938	0.9967	0.8305	**0.9979**
RMSE	12.41	12.24	4.650	3.012	2.34	17.30	**2.094**

Bold values represent the best model

uncorrupted areas to restore the structure and texture details of the pathological images step by step at multiple scales. Quantitative results of the four models are shown in Table 3. Among the five models, it is shown that the average PSNR SSIM and RMSE of the restored pathological images using the proposed method reached 2.094, 41.96 dB, and 0.9979, respectively. Compared to other inpainting frameworks such as CE, GLGIC, EdgeConnect, and DMFN, the PSNR and SSIM scores of the proposed method increased by more than 3 dB and 0.004, respectively. And for CNRN, the performance is almost similar to MSSA GAN. But the PSNR and SSIM of the corrupted area are 35.97 dB and 0.9909, respectively, which are 4 dB and 0.005 lower than MSSA GAN.

Here, the MSSA GAN have variety of loss terms in the cost function, but it does not lead to non-convergence or overfitting in the training process. The reasons are as follows: (1) Each term of the cost function has assigned a weight with different orders of magnitude. In MSSA GAN, multi-scale loss and adversarial loss have larger weights than other losses, which are regarded as the main losses to guide the network in image inpainting process. While other losses are used as auxiliary losses to ensure the consistency of the restored image and the real image in terms of style and structure, especially in the corrupted area. Therefore, it will not lead to non-converge or overfitting for the model. (2) These loss functions are focus on different aspects of inpainted image based on the statistical characteristics. So it does not lead to non-convergence. Here, perceptual loss, style loss and TV loss are also used to increase the feature matching degree between the generated image and the real image. It changes the cost function for the generator to minimize the statistical difference between the features of the real images and the generated images. Here, we measure the L_1-distance between the means of their feature vectors. Therefore, these losses expand the goal of MSSA GAN from beating the opponent to matching features in real images, thereby avoiding overfitting. (3) The dataset contains 23,080 and 1,974 non-overlapping pathological image patches for training and testing. The large database enables the model to learn as many special cases as possible, thereby effectively preventing overfitting by iteratively updating the network parameters. The results on the test set demonstrated that the PSNR and SSIM indicator can reach 41.96 and 0.9979, which also proves the generalization ability of the model.

To demonstrate the effectiveness of the MSSA GAN, we build our network without RSA module, multi-scale part and loss terms to perform an ablation experiment separately.

5.1 Without RSA module

Table 4 shows that the PSNR and SSIM of the pathological image using our network without the RSA module is only 37.33 and 0.9931, respectively. This is because the RSA module uses background information as a benchmark to improve the PSNR and SSIM of missing regions. The model loses contextual information as a reference after ablation.

5.2 Without multi-scale part

As shown in Table 4, the PSNR and SSIM of the restored image is decreased by ~ 14 dB and ~ 0.06 without multi-scale part. This means that only using RSA module is not sufficient for accurate image restoration. The fine-grained texture of pathological images should be restored by fine-tuning the RSA module at multiple scales.

5.3 Ablation studies of the loss terms

There are many loss terms in the loss function, so ablation studies are performed on each term of loss function to analyze the impacts on the final results. From Table 4, it can be seen that the multi-scale loss affects the results more than other losses. Therefore, we extend our study to analyze the impact of the number of multi-scale losses on the restored results. Parameters in columm 2 to columm 4 are the average PSNR and SSIM score after ablation of 3, 2, and 1 scale losses of the M-S losses, respectively. The result showed that the inpainted images can be improved as the number of multi-scales increases. Meanwhile, ablation experiments also demonstrate that the multi-scale loss at scale 128 is the most important one among the three scales, as it can provide the exact details for the image. Table 4 also demonstrated the impact of the other four losses on the restoration results. The results showed that the overall index such as PSNR and SSIM of the restored image do not decrease significantly after ablation. Nevertheless, in the corrupted region, the average PSNR index of the restored image is not more than 39.76 dB. It means that the restored image would lose stylistic consistency and continuity between corrupted and uncorrupted areas without these four losses.

Nevertheless, for image inpainting tasks, the restored result in the corrupted area can reflect the effectiveness of the proposed method. Therefore, PSNR and SSIM score of the corrupted areas are calculated to evaluate the restoration performance of MSSA GAN. Figure 6. shows the quantitative comparison of the four methods using the two indicators in the corrupted area. It showed that the average PSNR and SSIM score of the corrupted area could reach 40.23 dB and 0.9964 when the proposed method is applied. While the PSNR using SOTA can only 36.94 dB. It also illustrates the denoising capability of the proposed method. In summary, the results demonstrate that the proposed model can not only restore pathological images with detailed textures effectively but also suppress noise and

Table 4 Ablation experiment results

Parameters	Without RSA	Without M-S(3)	Without M-S(2)	Without M-S(1)	Without $L_{perceptual}$	Without L_{style}	Without L_{adv}	Without L_{TV}	MSSA GAN
PSNR (global)	37.33	27.59	30.55	36.33	40.58	41.19	40.85	41.63	**41.96**
SSIM (global)	0.9931	0.9377	0.9632	0.9813	0.9976	0.9975	0.9966	0.9973	**0.9979**
PSNR (corrupted)	35.65	25.26	29.21	34.09	37.82	39.76	37.99	38.18	**40.23**
SSIM (corrupted)	0.9899	0.8987	0.9491	0.9737	0.9963	0.9960	0.9930	0.9934	**0.9964**

Bold values represent the best model

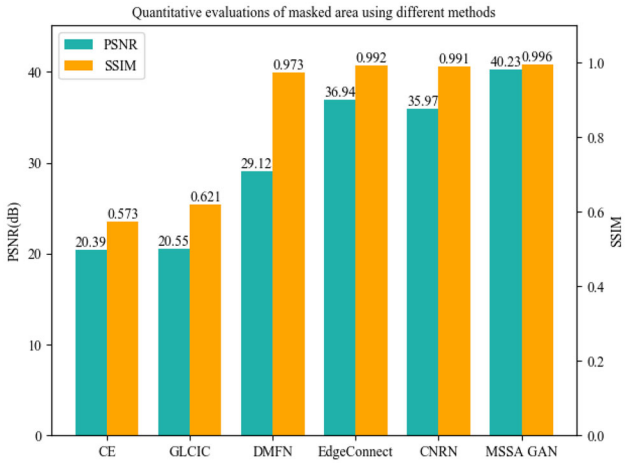

Fig. 6 PSNR and SSIM score of the six models in the corrupted region

image distortion. We also evaluate the generalization ability of MSSA GAN using different mask sizes from 64 to 96 with stride 8. In Fig. 7a and b, we averaged PSNR and SSIM score 6 times under each mask after removing the highest and lowest evaluation indicator. It is shown that the PSNR and SSIM indicators can be maintained above 40 dB and 0.99, respectively. The PSNR and SSIM score will not decrease too much as the mask size increases. This means that the performance of MSSA GAN is not getting worse when the size of the mask is within a specific range.

In the paper, the regional self-attention mechanism shares some similarities with the global and local attention (GLA) modules in ref [23], as both of them contain two branches with similar structures. However, they have some essential differences in algorithm. *Firstly,* in ref [23], the correlation maps in the global and local attention modules are pruned by the input masks. While for regional self-attention module, corrupted area and uncorrupted area are treated as foreground and background, which are separated before input to the RSA module. This can be considered as a mask pruning

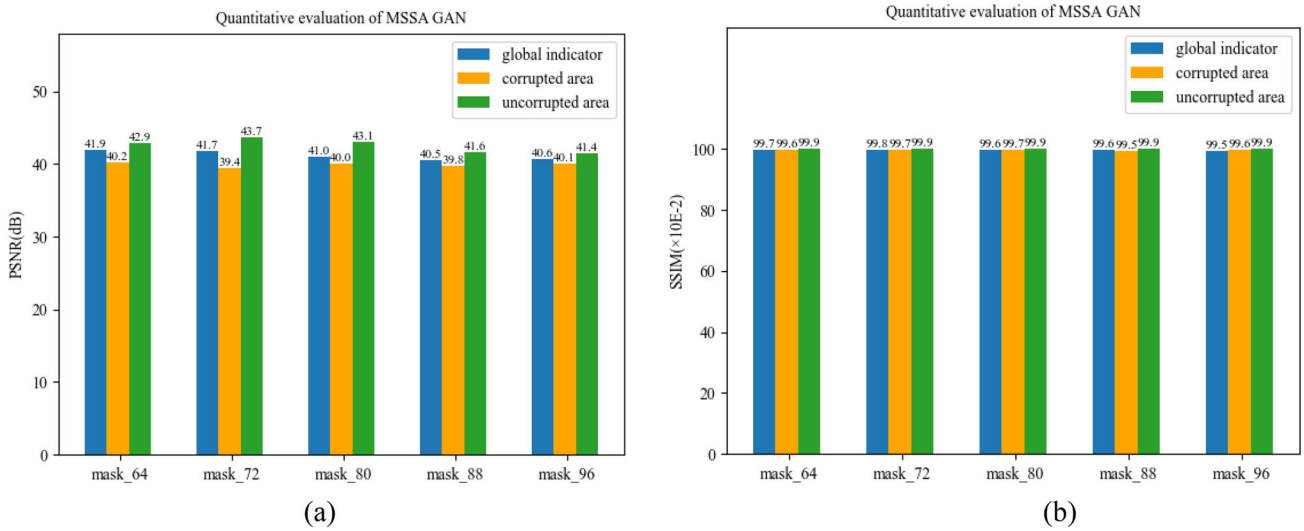

Fig. 7 PSNR and SSIM score of the restored image using MSSA GAN at different mask size

operation in the rectangular area. *Secondly*, mask pruning-based global and local attention module calculates both the global dependencies and local similarities of the image at the global (i.e., feature map) or local level (i.e., image patches), respectively. Thus, GLA module can efficiently calculate the features that contribute the most to reconstructing the missing regions and select the best candidate patches from patch-based feature map in local attention branch. For RSA module, the correlation between corrupted and the uncorrupted area, as well as the correlation within the corrupted area are assigned in two parallel branches to minimize the effect of mask values on valid pixels, which enables the model to adaptively adjust the weights of the two branches according to the current restoration result for further image inpainting. *Thirdly,* in ref [23], local attention branch can generate refined image features locally. So the global attention branch is to eliminate the discontinuities between the missing and background regions using image semantic information, whereas in our network, the two branches of RSA module are focuses on forground restoration. The structural consistency and coherence are achieved by image restoration at multiple scales.

6 Conclusion and future works

High-quality and high-resolution histopathological images are of great significance for precision medicine. However, local image contamination or missing data are common phenomena due to a multitude of factors, such as the superposition of foreign bodies and improper operations in the process of obtaining and processing pathological digital images. Almost all the algorithms based on deep learning are practical for natural images, CT images, and MRI, which are not effective for pathological images. Therefore, MSSA GAN is proposed to restore two colonic histopathological images with a large square contaminated region. The network uses a multi-scale image reconstruction architecture and regional self-attention technique to restore pathological images. The results indicate that MSSA GAN can maintain the precise textures in the original image and improve the efficiency of image restoration, simultaneously. After adaptively adjusting the learning rate and optimizing the network with joint loss function, the model can generate restored images that are extremely similar to the original pathological images. Parameters such as RMSE, PSNR and SSIM reached 2.094, 41.96 dB, and 0.9979, respectively. The superior performance of the network opens up a new perspective for image restoration with tiny details.

In this paper, MSSA GAN is proposed for pathological image restoration tasks with complex structural information corruption. The main contribution of our work is to use the regional self-attention module at multiple scales to reconstruct square-shaped corrupted areas. The next step is to extend our work to deal with the reconstruction of irregular-shaped corrupted areas in pathological images. This enables medical image restoration technology to be applied in practical medical scenarios.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant 11804209, Natural Science Foundation of Shanxi Province under Grant 201901D211173, Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi under Grant 2019 L0064, and Natural Science Foundation of Shanxi Province under Grant 201901D111031.

Funding This work was supported by the National Natural Science Foundation of China under Grant 11804209, Natural Science Foundation of Shanxi Province under Grant 201901D211173, Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi under Grant 2019 L0064, and Natural Science Foundation of Shanxi Province under Grant 201901D111031.

References

1. Tfc, A., Js, B.: Non-texture Inpainting by curvature-driven diffusions. J. Vis. Commun. Image Represent. **12**(4), 436–449 (2001)
2. Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. **3**(9), 1200–1212 (2004)
3. Ruzic, T., Pizurica, A.: Context-aware patch-based image inpainting using markov random field modeling. IEEE Trans. Image Process. **24**(1), 444–456 (2015)
4. Jin, K.H., Ye, J.C.: Annihilating filter-based low-rank hankel matrix approach for image inpainting. IEEE Trans. Image Process. **24**(11), 3498–3511 (2015)
5. Xue, H., Zhang, S., Cai, D.: Depth image inpainting: improving low rank matrix completion with low gradient regularization. IEEE Trans. Image Process. **26**(9), 4311–4320 (2017)
6. Wei, Y., Liu, S.: Domain-based structure-aware image inpainting. SIViP **10**(5), 911–919 (2016)
7. Bertalmio, M., Sapiro, G., Caselles, V.: Image inpainting. Siggraph **4**(9), 417–424 (2000)
8. Barnes, C.: Patchmatch: a randomized correspon-dence algorithm for structural image editing. ACM Trans. Graph. (2009). https://doi.org/10.1145/1531326.1531330
9. Ying, H., Kai, L., Ming, Y.: An improved image inpainting algorithm based on image segmentation. Procedia Comput. Sci. **107**, 796–801 (2017)
10. Qin, Z., Zeng, Q., Zong, Y.: Image inpainting based on deep learning: a review. Displays **69**(2), 102028 (2021)
11. Chang Y L, Liu Z Y, Hsu W: Vornet: Spatio-temporally consistent video inpainting for object removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. (2019)
12. Zeng Y, Fu J, Chao H: Learning pyramid-context encoder network for high-quality image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Pp.1486–1494. (2019)
13. Liu H, Jiang B, Xiao Y: Coherent semantic attention for image inpainting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4170–4179 (2019)

14. Liao, L., Hu, R., Xiao, J.: Artist-net: decorating the inferred content with unified style for image inpainting. IEEE Access. **7**, 36921–36933 (2019)
15. Hertz A, Fogel S, Hanocka R: Blind visual motif removal from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6858–6867 (2019).
16. Yang C, Lu X, Lin Z: igh-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6721–6729 (2017)
17. Liu P, Zhang H, Zhang K: Multi-level wavelet-CNN for image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 773–782 (2018)
18. Dolhansky B, Ferrer C C: Eye Inpainting with Exemplar Generative Adversarisal Networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7902–7911 (2018)
19. Li, H., Li, G., Lin. Li,: Context-aware semantic inpainting. IEEE Trans. Cybern. **14**(8), 4398–4411 (2015)
20. Zheng C, Cham T J, J Cai: Pluralistic image completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019)
21. Yu J, Lin Z, Yang J: Generative Image Inpainting with Contextual Attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition IEEE. pp. 5505–5514 (2018)
22. Chen, Y., Hu, H.: An improved method for semantic image inpainting with gans: progressive inpainting. Neural Process. Lett. **49**(3), 1355–1367 (2018)
23. Uddin, S.M., Jung, Y.J.: Global and local attention-based free-form image inpainting. Sensors **20**(11), 3204 (2020)
24. Yang, Y., Cheng, Z., Yu, H.: MSE-Net: generative image inpainting with multi-scale encoder. Vis. Comput. (2021). https://doi.org/10.1007/s00371-021-02143-0
25. Pathak D, Krahenbuhl P, Donahue J: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016)
26. Liu G, Reda F A, Shih K J: Image inpainting for irregular holes using partial convolutions. In: European Conference on Computer Vision. (2018).
27. Yu J, Lin Z, Yang J: Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019)
28. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. **36**(4), 1–14 (2017)
29. Nazeri K, Ng E, Joseph T: EdgeConnect: Generative image inpainting with adversarial edge learning. https://arxiv.org/abs/1901.00212 (2019).
30. Zhao L, Mo Q, Lin S: Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5741–5750 (2020)
31. Hui Z, Li J, X Wang: Image fine-grained inpainting. https://arxiv.org/abs/2002.02609 (2020)
32. Li X, Zhou S: GLAGAN image inpainting algorithm based on global and local consistency.In: International Information Technology and Artificial Intelligence Conference (ITAIC). (2020)
33. Yang, G., Yu, S., Dong, H.: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. IEEE Trans. Med. Imaging **37**(6), 1310–1321 (2017)
34. Quan, T.M., Nguyen-Duc, T., Jeong, W.K.: Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE Trans. Med. Imaging **37**(6), 1488–1497 (2018)
35. Lei, B., Kim, J., Kumar, A.: Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). Lect. Notes Comput. Sci. **1055**, 43–51 (2017)
36. Kaushik, H., Singh, D., Kaur, M.: Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models. IEEE Access. **9**, 108276–108292 (2021)
37. Liang, M., Ren, Z., Yang, J.: Identification of colon cancer using multi-scale feature fusion convolutional neural network based on shearlet transform. IEEE Access. **8**, 208969–208977 (2020). https://doi.org/10.1109/ACCESS.2020.3038764
38. Pimkin A, Samoylenko A, Antipina N: Multidomain CT metal artifacts reduction using partial convolution based inpainting. In: International Joint Conference on Neural Networks (IJCNN). (2020)
39. Deng K, Sun C, Liu Y: Real-time limited-view CT inpainting and reconstruction with dual domain based on spatial information. https://arxiv.org/abs/2101.07594 (2021).
40. Liu, X., Xing, F., Yang, C.: Symmetric-constrained irregular structure inpainting for brain MRI registration with tumor pathology. Int. MICCAI Brainlesion Workshop (2020). https://doi.org/10.1007/978-3-030-72084-1_8
41. Armanious K, Kumar V, Abdulatif S: ipA-MedGAN: inpainting of arbitrary regions in medical imaging. In: Proceedings of the IEEE International Conference on Image Processing. pp. 3005–3009 (2020)
42. Armanious K, Mecky Y, Gatidis S: Adversarial inpainting of medical image modalities. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3267–3271 (2019). https://doi.org/10.1109/ICASSP.2019.8682677
43. F Yu, Koltun V: Multi-scale context aggregation by dilated convolutions. https://arxiv.org/abs/1511.07122 (2015).
44. Zhang H, Goodfellow I, Metaxas D. Self-attention generative adversarial networks. In: Proceedings of International conference on machine learning. pp. 7354–7363 (2019)
45. He K, Zhang X, Ren S: Deep residual learning for image recognition. In: Proceedings of IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
46. Liu F, X Ren, Zhang Z: Rethinking skip connection with layer normalization. In: Proceedings of International Conference on Computational Linguistics. pp. 3586–3598 (2020)
47. Li J, Madry A, Peebles J: On the limitations of first-order approximation in GAN dynamics. In: Proceedings of International Conference on Machine Learning. pp. 3005–3013 (2018)
48. Lei, N., An, D., Guo, Y.: A geometric understanding of deep learning. Engineering **6**(3), 361–374 (2020)
49. Belli D, Hu S, Sogancioglu E: Context encoding chest X-rays. https://arxiv.org/abs/1812.00964 (2018)
50. Mahendran A, Vedaldi A: Understanding deep image representations by inverting them. In: Proceedings of IEEE conference on computer vision and pattern recognition. pp. 5188–5196 (2015)

Meiyan Liang was born in Xinzhou, Shanxi, China, in 1984. She earned her PhD degree at the School of Opto-Electronics, Beijing Institute of Technology, Beijing, China, 2015. From 2019 to 2020, she was a visiting scholar with the Department of Computer Science and Electrical Engineering at the University of Tennessee, Knoxville. She is currently an associate professor at the School of Physics and Electronic Engineering, Shanxi University, Taiyuan, China. Her current research interests focus on machine learning, deep learning and medical image processing.

Qiannan Zhang was born in Yuncheng, Shanxi, China, in 1997. She received her B.S. degree from Shanxi Datong University, Datong, China, in 2019, and she is pursuing an M.S. degree at Shanxi University. Taiyuan, Shanxi, China. Her research interests include medical image processing and pathological image reconstruction.

Guogang Wang was born in Taian, Shan Dong, China in 1977. He received his B.S. degree in applied mathematics from Qufu Normal University, Qufu, Shandong, China, in 2000. He received his M.S. degree in applied mathematics from Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China, in 2011. He received his PhD degree in signal and information processing from Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China, in 2016. He is currently an Associate Professor in Shanxi University. His main research interests focus on computer vision, image processing and pattern recognition.

Na Xu received the bachelor's degree in Electronic Information Science and Technology from Baoji University of Arts and Sciences, Baoji, Shanxi, China, in 2009, the Master's Degree in Engineering from Chongqing University, Chongqing, China, in 2013. She is a lecturer of Department of Electronic Information Engineering with Shanxi University,Her research interests include motion video processing and computed tomography (CT) image reconstruction.

Lin Wang was born in Jiexiu, Shanxi, China, in 1982. She earned her master's degree from Shanxi Medical University in 2011, majoring in pathophysiology. She completed postgraduate scientific research and clinical study in the General Hospital of Beijing military region from 2008 to 2011, and completed the advanced research course on early digestive tract cancer in Nanjing Gulou Hospital in 2017. She is currently the deputy chief physician of the Department of Pathology of Shanxi Bethune Hospital (Shanxi Academy of Medical Sciences, the third Clinical Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Tongji Medical College of Huazhong Medical University). her main research interests are pathological diagnosis of respiratory and digestive system tumors, molecular pathological diagnosis, and remote pathological diagnosis.

Haishun Liu was born in Tianjin, China, in 1991. He received his M.S. degree in physics from Capital Normal University, Beijing, China, in 2018. He is pursuing his PhD degree at Capital Normal University. His research interest focuses on the biomedical image applications.

Cunlin Zhang was born in 1961. He earned his PhD degree at the School of Opto-Electronics, Beijing Institute of Technology, Beijing, China. He is currently a professor with the Key Laboratory of Terahertz Optoelectronics, Ministry of Education, Capital Normal University. His research interests include nondestructive testing of THz wave and THz spectroscopy, Deep learning and its applications