

Received September 28, 2020, accepted October 12, 2020, date of publication October 19, 2020, date of current version October 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031973

# Consecutive Context Perceive Generative Adversarial Networks for Serial Sections Inpainting

SIQI ZHANG<sup>1</sup>, LEI WANG<sup>1</sup>, JIE ZHANG<sup>2</sup>, LING GU<sup>2</sup>, XIRAN JIANG<sup>1</sup>,  
XIAOYUE ZHAI<sup>2</sup>, XIANZHENG SHA<sup>1</sup>, AND SHIJIE CHANG<sup>1</sup>

<sup>1</sup>Division of Biomedical Engineering, China Medical University, Shenyang 110122, China

<sup>2</sup>School of Basic Medicine Science, China Medical University, Shenyang 110122, China

Corresponding author: Shijie Chang (sjchang@cmu.edu.cn)

This work was supported in part by the National Science Fund of Liaoning under Grant 2018-64, in part by the National Science Fund of China (NSFC) under Grant 31971115, in part by the Big Data Research for Health Science of China Medical University under Grant Key Project 6, and in part by the Science Research Fund for Higher Education of Liaoning under Grant LQNK201744.

**ABSTRACT** Image inpainting is a hot topic in computer vision research and has been successfully applied to both traditional and digital mediums, such as oil paintings or old photos mending, image or video denoising and super-resolution. With the introduction of artificial intelligence (AI), a series of algorithms, represented by semantic inpainting, have been developed and better results were achieved. Medical image inpainting, as one of the most demanding applications, needs to meet both the visual effects and strict content correctness. 3D reconstruction of microstructures, based on serial sections, could provide more spatial information and help us understand the physiology or pathophysiology mechanism in histology study, in which extremely high-quality continuous images without any defects are required. In this article, we proposed a novel Consecutive Context Perceive Generative Adversarial Networks (CCPGAN) for serial sections inpainting. Our method can learn semantic information from its neighboring image, and restore the damaged parts of serial sectioning images to maximum extent. Validated with 2 sets of serial sectioning images of mouse kidney, qualitative and quantitative results suggested that our method could robustly restore breakage of any size and location while achieving near realtime performance.

**INDEX TERMS** Serial sectioning images, generative adversarial network, consecutive context perceive GAN.

## I. INTRODUCTION

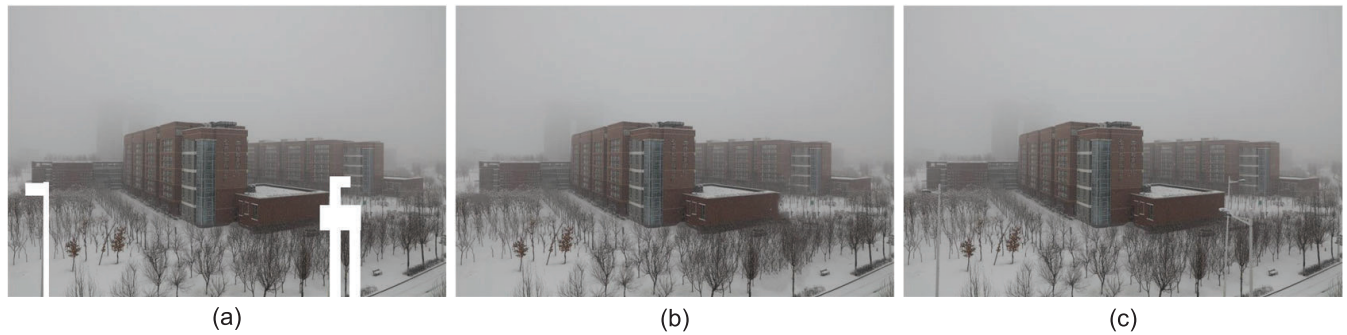
Image inpainting is a process of restoration or reconstruction of an image in the light of its background information, in which part is damaged, deteriorated, or missing with the goal of presenting the image as it is originally created. It is one of the hot topics in computer vision and artificial intelligence, and is widely used in our daily life, like conservation and restoration of old paintings or damaged photographs, elimination of red-eye, removal of watermarks from pictures, or the date-print from stamps.

This technique would also help researchers in medical study who are trying to check the images with defects that are inevitably produced, e.g., a stack of pathophysiology images for 3D reconstruction in histology research. 3D reconstruction of microstructures could provide more spatial information comparing traditional sectional images, and help us

understand the physiology or pathophysiology mechanism [1]. The histological imaging method of serial sectioning offers ultra-high-resolution images of a large tissue block with abundant staining information and nucleic acid / protein labeling information, which OPT or small animal MRI could not provide [2]. Therefore, it has been the mainstream procedure in histology study. As all the operations are done manually, especially the sectioning procedure, a small number of tissue sections are inevitably damaged in parts or in whole. These defects would seriously affect the accuracy of automatic image segmentation, object recognition and tracking of the structure, producing artifacts in virtual section rendered from the image stack, and even make a failure of 3D reconstruction. Therefore, high-quality serial sectioning images containing damaged parts are needed to be restored before we facilitate them.

Image inpainting was carried out in a series of block-removing algorithms at first. They connected the pixels with same gray-level, which obviously were not suitable

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.



**FIGURE 1.** Image inpainting with specific objects unusual to restore. (a) Image with masks representing damaged patches; (b) The inpainting result; (c) Ground Truth. Lampposts in (c) cannot be restored with existing methods.

for natural images. Kokaram *et al.* [3] tried to use motion estimation and autoregressive model to mend up the defect parts by interpolating the neighboring frame in cinema products, but this technique could not be applied to still images. Then image inpainting techniques are divided into two broad categories: non-textured and textured structure methods. (1) The non-textured structure method utilizes higher order partial differential equation (PDE) [4], or extreme value of function of prior model and data model to generate the mending patches, which is effective in small-sized breakage. Bertalmio *et al.* [5] adopted a patching algorithm based on PDE to spread the patches within boundary using the edge information of the damaged area for inpainting, in which BSCB model, simulating curvature driven diffusion (CDD) with a third-order PDE. Furthermore, improved methods, including total variation model (TV) [6] and Euler's elastica model [7] *etc.* were all designed to fill the geometric image model and improve computing speed. (2) The textured structure method, which was effective in large-sized breakage, is another way of image completion. A pixel from the boundary of patch area was selected to search a patch with the most similar textures, which is further used for replacing the hole, including exemplar-based techniques [8], Harrison's algorithm, and Criminisi's algorithm [9]. Overall, these methods without artificial intelligence filled the blanks with patches calculated from statistics features of the rest image, but they did not fully utilize semantic information.

After a series of breakthroughs of artificial intelligence (AI) in image processing, it has also been applied to image inpainting. MemNet [10], employing memory network for image restoration, was very deep convolutional neural networks (CNNs) that introduced a memory block, consisting of a recursive unit and a gate unit, to explicitly mine persistent memory through an adaptive learning process. The recursive unit learned multi-level representations of current state under different receptive fields. The representations and outputs from the previous memory blocks were concatenated and sent to the gate unit. MemNet was applied to image demising, super-resolution and JPEG deblocking. Another kind of network was the generative adversarial networks (GAN) [11]. Its generative model consisted of a feature detector or interpreter,

such as encoder-decoder [12]; while its discriminative model pre-trained by millions of images was implemented to guide the image restoration to improve the similarity of generated and original images [13]. WGAN [14] and WGAN-GP [15] were subsequently proposed to improve the stabilization in process of training. DCGAN [16], a combination of CNN and GAN, used convolution networks to generate model for unsupervised training and got excellent results of feature extraction. Recently, GAN has been introduced in different fields of image processing. Deepfill [17] employed a coarse-to-refine training process and designed a contextual attention layer to improve its spatial consistency of perception. Saliencygan [18] was proposed for semi-supervised salient object detection. TPSDicyc [19] could generate synthesized images with unpaired and unaligned data. DAGAN [20] was applied to image reconstruction. More newly designed networks based on GANs [21]–[23] have proven the ability of contextual information and perceptual information detection in daily pictures. But the existed methods were not able to restore the specific objects in the image, as shown in Fig. 1.

For medical image inpainting which seems suitable for employing AI techniques, we find that the generated image patches, always show strange results for doctors or researchers. The key reason is the correctness is more important than its visual effects and quantitative evaluation metrics. The existed methods restored the damaged images by gathering and comparing the patches features from the rest of image, but ignored the specific and unpredictable targets that existing in damaged areas. Our framework is based on contextual attention [17], a two-stage network architecture. We have designed Serial Perceive Module, 2-pathway deep learning networks with concatenation extracting features from both the damaged image and its neighboring image to acquire more valuable features. SPM consists of two encoder-decoder structures and combines the features when restoring inpainting image. For serial sectioning images, we find that the structural information and spatial completion of damage parts are also related to their neighboring images. When the neighboring image is fed into the framework, both surrounding features of damaged image and the features in neighboring image can be learned to restore damaged parts.

Therefore, the features in neighboring image could provide more detailed information of broken part, while the surroundings could provide better edge information to make the generated image more consistent. The result generated in SPM is sent to the refined generating network to generate the final image. In the refined generating network, an attention module and a dilated convolution module are employed in different networks. The attention module would focus on the most important patches for restoration. The dilated convolution module continues to focus on the features with larger receptive fields and constructs the image with deeper semantic information. In the second stage network, inpainting image would be further processed to a clear image.

Our main contributions can be summarized as following:

- In this study, we have proposed a novel framework, CCPGAN, for image inpainting in serial sections, focusing on high fidelity of tissue structure. This method can learn to recover image from the most relevant patches from both damaged image and its neighboring image.
- We design a Serial Perceive Module (SPM), including 2-pathway deep learning networks with feature maps concatenation, which could extract features from neighboring image and damaged image. The features from neighboring image could provide holistic and structured characteristics to guide the refined generating network to produce correct images.
- Our model was trained and validated on our N7 dataset, then further validated on another image dataset of N5 to assess its generalization. Furthermore, three histologists were invited in our experiment to assess correctness of our model and they valued our model performed best among image inpainting models.

## II. RELATED WORK

Image restoration based on deep learning and generative adversarial networks has developed rapidly and gained a leading position. Typical application of these methods in medical image inpainting have been done in Sogancioglu *et al.* [24], they tried some state-of-the-art methods, such as context encoder, semantic inpainting and contextual attention, to repair the Chest X-ray images, and these kinds of images.

### A. CONTEXT ENCODER

Context Encoders [12], a convolutional neural network, was trained to generate the contents of an arbitrary image region conditioned on its surroundings. It would understand the content of the entire image, as well as producing a plausible hypothesis for the missing parts. Similar to auto-encoder, it included an encoder-decoder pipeline, and a channel-wise fully connected layer. The encoder, such as AlexNet [25], took an input image with missing regions and produced a latent feature representation of that image; then the decoder, a series of five up-convolutional layers [26], took this feature representation and produced the missing image contents. Channel-wise fully connected layer helped reduce the parameters sizes from  $m^2n^4$  ( $m \times n \times n$  for both input and output

feature maps) to  $mn^4$ , which was followed by a stride 1 convolution to propagate information across channels. GAN, as one of the generative models, was used to determine the likelihood that the images were from the training set or generating set. When the generated images were consistent with ground truth in content, the discriminator could not distinguish the image from training set or not, the GAN models were optimized. Discriminators in image inpainting were designed as global and local discriminators [27] to assess the authenticity of the global image and the continuity of the local repairing region. Then, the perceptual model [17], Patch-Based GAN [28] *etc.* were proposed to improve the performance in visual effects; gated convolution [29] or partial convolution [30] were added to fix any shape defects.

### B. CONTEXTUAL ATTENTION

Contextual attention [17] was designed to extract the features similar to the damaged patch from distant parts. This method could synthesize an image utilizing a trained attention model with image features around as references, and making better forecasts. First, a simple dilated convolutional network was trained with reconstruction loss to rough out the missing contents. Then, contextual attention used the features of known patches as convolution filters to generate patches. It was designed and implemented with convolution for matching generated patches with known contextual patches and channel-wise softmax function was to weigh patches similarity. Then the inpainting result was generated by deconvolution. The method has been shown promising visual results for inpainting images of faces, building facades, and natural scenes. Earlier in the attention model improvements, spatial transformer network (STN) [31] for object classification tasks was proposed, but not suitable for patch-wise attention model. Zhou *et al.* [32] introduced an appearance flow to predict offset vectors but is not effective in predicting a flow field from the background region to the hole. Dai *et al.* [33] and Jeon and Kim [34] proposed to learn spatially attentive or active convolutional kernels, but might still be limited in exacting features from the background. Recently, Zhu *et al.* [35] introduced lesion focused SR (LFSR) and multi-scale methods to improve the perceptual quality of the super-resolved results for brain tumor MRI images.

Contextual attention, simulating the attention mechanism of the brain, picked up the critical features from a variety of information. Images were split into patches with different weights indicating the similarity of the mask. Contextual attention focused on searching the input image to find useful information with higher weights related to the damaged part, in order to improve the efficiency and accuracy of processing visual information and reduce the computing cost. The framework, such as encoder, could learn a variety of relationships between patches and represented them.

### C. GLOBAL AND LOCAL CONSISTENT

Consistency was another challenge in medical image inpainting. The globally and locally consistent image inpainting

network [27] for image restoration tried to make the missing part keep consistent with the original in part and whole. It consisted of an encoder network and two auxiliary context discriminator networks that were used as global discriminator and local discriminator merely in training stages. The global discriminator network took the whole image as input to assess similarity of the entire image; whereas the local discriminator network took only the restored area as input to assess if the generated content was similar enough.

The methods mentioned above could restore natural scene image with state-of-the-art performance, but they doesn't focus on the high fidelity restoration for medical images. In serial sectioning images, the neighboring of damaged image could provide abundant information that would be helpful in completing the missing parts. Therefore, we designed our framework to restore image utilizing both damaged image and its neighbouring image.

### III. METHOD

#### A. SERIAL PERCEIVE MODULE

To extract and learn features from damaged image and its neighboring image simultaneously, Serial Perceive Module (SPM) was proposed with two concurrent pathways. The first pathway utilizes the damaged image as input, called damaged net  $N_d$ ; the second pathway uses the neighboring image as input, called referenced net  $N_{ref}$ . Each pipeline follows the encoder-decoder mode, consisting of 17 convolutional layers, including downsampling, dilated convolution, and upsampling layers and the features could be extracted from both damaged image and its neighboring image in encoder process and reconstructed image with learned information in decoder process. Dilated convolutions [27] are adapted instead of ordinary convolutions in this module is to extend receptive fields without resolution loss and computation increase, which are helpful in multi-level restoring the border areas. It could be denoted as follows:

$$y_{u,v} = \sigma\left(b + \sum_{i=-k'_h}^{k'_h} \sum_{j=-k'_w}^{k'_w} W_{k'_h+i, k'_w+j} X_{u+\eta i, v+\eta j}\right)$$

$$k'_h = \frac{k_h - 1}{2}, \quad k'_w = \frac{k_w - 1}{2} \quad (1)$$

where  $(u, v)$  represents any position in the layer matrix, and  $x_{u,v}$  and  $y_{u,v}$  are the pixels in the input and output layer;  $\sigma(\cdot)$  is a component-wise non-linear transfer function;  $W$  is weight convolution kernel, and  $b$  is a layer bias vector,  $w$  and  $h$  are the width and height of dilation kernel (odd number),  $\eta$  is the dilation factor, when  $\eta = 1$  the equation becomes the standard convolution operation.

Our inputs of module are followed as [27]. In the damaged net, an original image with white pixels filled in the damaged part and a binary mask with the white pixels indicating the size and location of the damaged part; in the referenced net, the neighboring image will be fed as input, while the neighboring image is selected from the previous or next one, as shown in Fig. 2. Generally, most of the inpainting nets

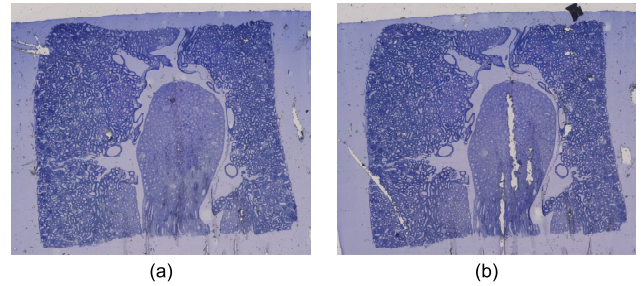


FIGURE 2. Serial sectioning images with breakage. (a) neighboring image (b) damaged image

recently will search relevant pixels from its surroundings to restore the damaged parts, but the features extracted in  $N_{ref}$  can be also leveraged to guide the restoration. During the reconstructing process in  $N_{ref}$ , the feature maps are connected to  $N_d$  after every upsampling step in  $N_d$ , thus,  $N_d$  can obtain the features from both damaged image and its neighboring image in deconvolution process. The result of 2-pathway networks would finally be fused with max operation before it outputs the coarse result.

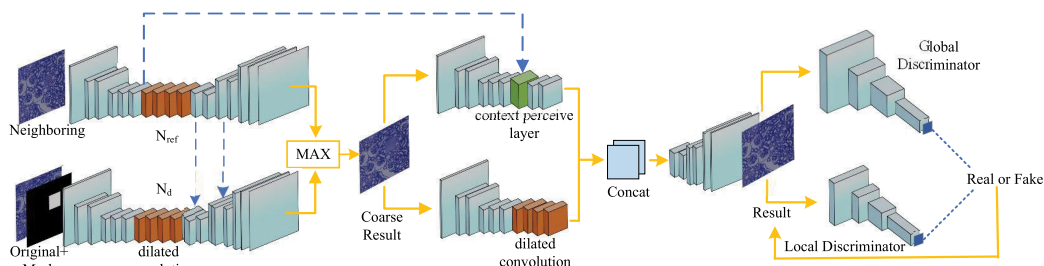
#### B. REFINED GENERATING NETWORK

The refined generating network is designed to fine-tune the roughly repaired result from SPM to be a more realistic and correct image. This network consists of 2-pathway networks. The first pathway contains context perceive layer and the second pathway contains dilated convolution layers. The context perceive layer learns to search which places are needed to pay attention for restoring in the background image (the unmasked patches) and the neighboring image; while the dilated convolution module continues to extract and learns features among the coarse result to acquire deeper semantic information. These features are then merged into one path decoder to reconstruct in the following deep network.

After 6 convolution operations, the feature maps generated from the coarse result are fed up into the context perceive layer. These feature maps multiply by mask to generate the foreground. Feature maps from undamaged part and neighboring image are concatenated as background. We extract  $3 \times 3$  patches from background and reshape them as convolution kernels. Then we measure the similarity between damaged parts  $\{f_{x,y}\}$  and its surroundings  $\{b_{x',y'}\}$  in the rest of damaged image and its neighboring image with the indicator of normalized inner product

$$S_{x,y,x',y'} = \left\langle \frac{f_{x,y}}{\|f_{x,y}\|}, \frac{b_{x',y'}}{\|b_{x',y'}\|} \right\rangle \quad (2)$$

where  $S_{x,y,x',y'}$  represents similarity of damaged part and its surroundings. To select the most similar patches, softmax function  $S'_{x,y,x',y'} = \text{softmax}(S_{x,y,x',y'})$  is employed to acquire attention scores. The patches with the highest scores will be used as the deconvolution filters to make deconvolution operation to restore the image.



**FIGURE 3.** The framework of CCPGAN. It consists of Serial Perceive Module, Refined Generating Network, Multi-scale discriminators. This framework receives both damaged image and its neighboring image as input and Serial Perceive Module generates the coarse result with the features from the two images. The dotted lines represent combination of two images and the results of  $N_d$  and  $N_{ref}$  are fused as a coarse image. The refined generating network divided two parts, coarse result and neighboring image are sent to the perceive module network to reconstruct with attention and the dilated convolution network will reconstruct image with features extracted from the coarse result. The refined generating network will generate the final inpainting image. The loss of our generator is the sum of auto encoder loss  $L_{AE}$ ,  $L_1$  loss, and  $G_{WGAN}$  loss. The loss of our discriminator ( $L_D$ ) is the sum of global WGAN-GP loss ( $L_{DG}$ ) and local WGAN-GP loss ( $L_{DL}$ ).

### C. LOSS FUNCTION

The original image is the ground truth for validating the generative image. In our model, both generative loss and discriminative loss are designed. The generative loss is the indicator of similarity between generated image and ground truth image in generator, which is the sum of auto encoder loss  $L_{AE}$ ,  $L_1$  loss, and  $G_{WGAN}$  loss. The discriminator loss ( $L_D$ ) is the sum of global WGAN-GP loss ( $L_{DG}$ ) and local WGAN-GP loss ( $L_{DL}$ ), which is designed to ensure global and local consistency of inpainting results. The  $L_{DG}$  is used to discriminate the generated images and ground truth; while  $L_{DL}$  is used to ensure the local correctness of generated images. When the sum of  $L_{DG}$  and  $L_{DL}$  reaches a global minimum, the generated images would be the results of good visual effects and correct microstructures content.

$$L_G = E_{y \sim P_g} \log[1 - D(y)] \quad (3)$$

Auto Encoder loss consists of the mean distance between coarse result and ground truth image and the mean distance between generated image and ground truth image.

$L_1$  loss is the L1 regulation in reconstruction progress, which is defined as

$$L_1 = \|M \odot y - M \odot x\| \quad (4)$$

where,  $M$  denotes the undamaged area (unmasked) in images,  $x$  is original image,  $y$  is inpainting image,  $\odot$  is pixelwise multiplication [27],  $\|\cdot\|$  is the Euclidean distance.  $L_1$  loss is used in both generator and discriminator.

Both global WGAN-GP loss and local WGAN-GP loss are defined as in [15]. Wasserstein distance  $W(P_r, P_g)$  defined as in (5) is adopted for comparing the distributions between the generated images and ground truth images.

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y) \sim \gamma} [\|x-y\|] \quad (5)$$

where  $P_r$  and  $P_g$  are the distribution of generated images and ground truth images.  $\Pi(P_r, P_g)$  is the set of all possible joint distributions,  $\gamma$  is any possible distribution.  $x$  and  $y$  are pixels in the ground truth and the inpainting results. Here, we use the improved gradient penalty of WGAN-GP instead of WGAN, avoiding the strict constraint of Lipschitz continuity. In the

damaged region to restore,

$$L_d = -E_{x \sim P_r} [D(x)] + E_{y \sim P_g} [D(y)] + \lambda E_{\hat{x} \sim P_{\hat{x}}} (\|\nabla_{\hat{x}} D(\hat{x}) \odot (1 - m)\|_2 - 1)^2 \quad (6)$$

where  $D$  is a set of 1-Lipschitz functions,  $\hat{x}$  is sampled from the straight line between points sampled from distribution  $P_r$  and  $P_g$ ,  $\hat{x} = tx + (1-t)y$ ,  $t \sim \text{Uniform}[0,1]$ , the mask value is 0 for missing pixels and 1 for elsewhere.

### D. NETWORK ARCHITECTURE

The overview of our framework is presented as Fig. 3, consisting of Serial Perceive Module, refined generating network, and global and local discriminators. Popular image inpainting methods usually try to detect similar or relevant features within the damaged image and restore the missing parts, such as face repairing *etc.*; but it is not effective in processing medical images as these textures are too similar to predict the original patches with various contents. Here, we assume that the neighboring images after registering are similar enough in tissue structure to restore the corresponding part in damaged image and can provide specific features. In order to restore the correct content with tiny differences, SPM utilizes the features from the neighboring image; while typical GAN-based methods detect features information only within the damaged image.

The SPM proposed in this study utilizes information from both damaged image and corresponding damaged part in the neighboring image, and detects the features in a large scale, merges the features after upsampling processes and fuses the results to generate a coarse repaired image, which is used to guide the following generating procedures. The coarse result is compared with original image with loss  $L_{AE}$  and  $L_1$  in training process. Then, in refined generating network, context perceive layer extracts patches from its surroundings and assesses the similarity between missing part in the coarse result and the patches, then reconstruct image with the filters reshaped by extracted patches. The dilated convolution layers in refined network further extract features with larger receptive fields and obtain deeper semantic information for restoring image with more details. These 2-pathway encoder

networks combine the features into one pathway to reconstruct a realistic result. The generator is guided by optimizing the loss function  $L_{AE}$  and  $L_1$  and penalizing the difference between generated and original patches. Comparing with  $L_2$  loss, our model with  $L_1$  loss could provide an image that is more similar to the ground truth, and cost less computing consumption in training process [36].

The discriminators used in our model are binary classifiers [37]. The discriminators are used to distinguish a real image from a generated one in order to improve the generated image quality. Both global discriminator  $D_G$  and local discriminator  $D_L$  have the same structure of CNN, consisting of 4 convolution layers and a fully connected layer. The global discriminator distinguishes a genuine image from a generated one in the scale of whole image to make the generated image more consistent with surrounding parts; while the local discriminator distinguishes a genuine patch from a generated one in the scale of the masked part of image to make the generated contents correct. Both of the loss functions of discriminators are added up together avoiding one of which are over processed.

Before our model is trained, the neighboring image is selected from previous or next image in the stack according to peak signal to noise ratio (PSNR). Then, the mask  $m$  is generated randomly and the damaged image  $x_d$  is generated with the mask by  $x_d = x \odot m$ . Next, the damaged image, neighboring image and the mask are fed into our SPM module to obtain coarse result  $y_s$ .  $y_s$  and the feature maps  $f_{nei}$  from  $N_{ref}$  in SPM are sent into the attention module pathway and  $y_s$  is sent into another feature extractor containing dilated convolution layer to generate the inpainting result  $y_o$  together. All the input images are normalized in  $[-1, 1]$ . The detail of whole procedure is shown in Algorithm 1.

## IV. EXPERIMENTAL RESULTS

### A. DATASETS

In this experiment, 2 sets of serial sectioning images were used to validate the algorithm we proposed. These serial sections were originally used for kidney research via a method of 3-D visualization of microstructures, therefore there were no animals sacrificed intended for this experiment [38]. The mice were provided by Animal Experimental Department of China Medical University. Serial sectioning images were carried out following a standard procedure: (1) tissue blocks were removed a few minutes after animal anesthesia, and were further fixed for 1h in fixation buffer; (2) the fixed tissue blocks were embedded in Epon 812 (for electron microscopy) or paraffin (for light microscopy); (3) after shape correction, it was cut into serial semi-thin or ordinary thickness (2.5 - 30 $\mu$ m approx.) using a microtome; (4) These serial sectioning images were stained and scanned by using a pathology slide scanner or a microscope. The experiments were approved by the Medical Ethics Committee of China Medical University.

We assessed our model and its generalization ability on the datasets with different sectioning or staining protocols:

### Algorithm 1 Procedure of Our Proposed Framework

**Input:**  $X(x_1 \dots x_n)$ : consecutive sample images;

$m$ : mask indicating damaged part;

**Output:**  $y_o$ : generated image;

- 1: **while** inpainting network has not converged **do**
- 2:   Select neighboring image  $x_{nei}$  from previous or next of  $x$ ;
- 3:   Generated damaged image  $x_d = x \odot m$ ;
- 4:   **for** Iterations **do**
- 5:     In the SPM module,  $(x_d, m) \xrightarrow{N_d} y_m, (x_{nei}, m) \xrightarrow{N_{ref}}$   
 $y_a, y_s = \max(y_m, y_a), f_{nei}$  represents feature maps before dilated convolution operation in  $N_{ref}$ ;
- 6:     Calculate  $l_1, l_{AE}$  losses between  $y_s$  and  $x$ ;
- 7:     Feed  $y_s$  and  $f_{nei}$  into context perceive layer network,  $y_s$  into dilated convolution module network and output  $y_o$ ;
- 8:     Calculate adversarial losses between  $y_o$  and  $x$ ;
- 9:     Update the inpainting network weights with pan-  
 elty;
- 10:   **end for**
- 11: **end while**

in N7 dataset, there were 1145 images (embedded with Epon 812 and stained with toluidine blue); and in N5 dataset there were 413 images (embedded with paraffin and stained with hematoxylin-eosin). Their volume resolutions were  $1.84 \times 1.84 \times 2.5 \mu\text{m}/\text{pixel}$  for N7 dataset, and  $1.84 \times 1.84 \times 5 \mu\text{m}/\text{pixel}$  for N5 dataset. Masks were manually labeled on the damaged parts. These images were aligned as in [39] and were cropped into the size of  $512 \times 512$  pixels in these experiments due to the limitation of GPU memory size. In order to train our model, 275 images were selected from N7 dataset only, masks were randomly placed in representation of partial damage in sections. Then the model was assessed with 164 images from N7 dataset in Experiment 1, and with 116 images from N5 dataset in Experiment 2. The algorithm above was implemented in Python 3.7 and ran on a workstation (Dell Precision T7920, CPU Intel Xeon Silver 4110  $\times$  2, RAM 32GB, GPU NVidia Titan RTX).

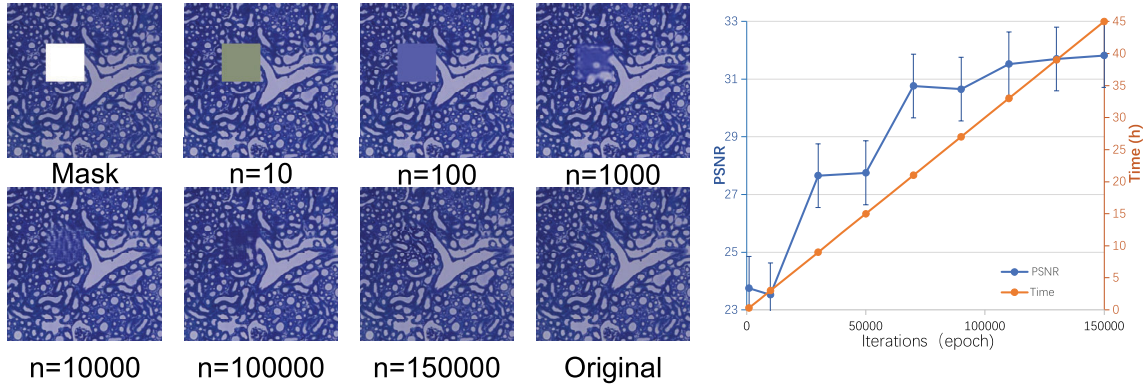
### B. EVALUATION METRICS

#### 1) QUANTITATIVE EVALUATION

We performed the quantitative evaluation using 5 metrics on testing datasets: Structural similarity index (SSIM) [40], Multi-scale structural similarity (MS-SSIM) [41], Feature Similarity Index (FSIM) [42], Peak signal-to-noise ratio (PSNR) [43] and Learned perceptual image patch similarity (LPIPS) [44].

SSIM is a full reference metric to estimate the global similarity between two images from 3 aspects of brightness, contrast and structure, as in (7). The value of SSIM is between 0 and 1, and SSIM=1, if the two images are identical.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$



**FIGURE 4.** Evaluation for iterations. (a) Qualitative evaluation for visual effect, it could be considered that our model could provide a restored image after 150000 epochs of training. (b) Quantitative evaluation, iterations with PSNR metric and training computing consumption.

where,

$$\mu_x = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N X(i, j)$$

$$\mu_y = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N Y(i, j)$$

$$\sigma_x^2 = \frac{1}{MN-1} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - \mu_x)^2$$

$$\sigma_y^2 = \frac{1}{MN-1} \sum_{i=1}^M \sum_{j=1}^N (Y(i, j) - \mu_y)^2$$

$$\sigma_{xy} = \frac{1}{MN-1} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - \mu_x)(Y(i, j) - \mu_y)$$

$$c_1 = (K_1 \cdot L)^2, c_2 = (K_2 \cdot L)^2$$

where,  $X(i, j)$  denotes the original image, and  $Y(i, j)$  denotes the inpainting image. Usually, set  $K_1 = 0.01$ ,  $K_2 = 0.03$ ,  $L = 100$  as in [40].

MS-SSIM is an improved version of SSIM for incorporating the variations of viewing conditions, which is more effective in multi-scale structure similarity evaluation, as in (8). The value of MS-SSIM is also between 0 and 1.

$$MS-SSIM = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (8)$$

where,

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

$$c_1 = (K_1 \cdot L)^2, c_2 = (K_2 \cdot L)^2, c_3 = \frac{c_2}{2}$$

We set  $\beta_1 = \gamma_1 = 0.0448$ ,  $\beta_2 = \gamma_2 = 0.2856$ ,  $\beta_3 = \gamma_3 = 0.3001$ ,  $\beta_4 = \gamma_4 = 0.2363$  and  $\alpha_5 = \beta_2 = \gamma_2 = 0.1333$  as in [41].

FSIM metric considers the phase congruency (PC) and the image gradient magnitude (G) as the significance features for gray level images. The FSIM index can be extended to FSIMc by incorporating the chromatic information in a straightforward manner. FSIMc index is used in this study to evaluate feature similarity. Its range is between 0 and 1.

$$FSIM_c = \frac{\sum_{x \in \Omega} S_L(x) \cdot [S_C(x)]^\lambda \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (9)$$

where,

$$S_L(x) = S_{PC}(x) \cdot S_G(x)$$

$$S_C(x) = S_I(x) \cdot S_Q(x)$$

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) \cdot PC_2^2(x) + T_1}$$

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) \cdot G_2^2(x) + T_2}$$

$$S_I(x) = \frac{2I_1(x) \cdot I_2(x) + T_3}{I_1^2(x) \cdot I_2^2(x) + T_3}$$

$$S_Q(x) = \frac{2Q_1(x) \cdot Q_2(x) + T_4}{Q_1^2(x) \cdot Q_2^2(x) + T_4}$$

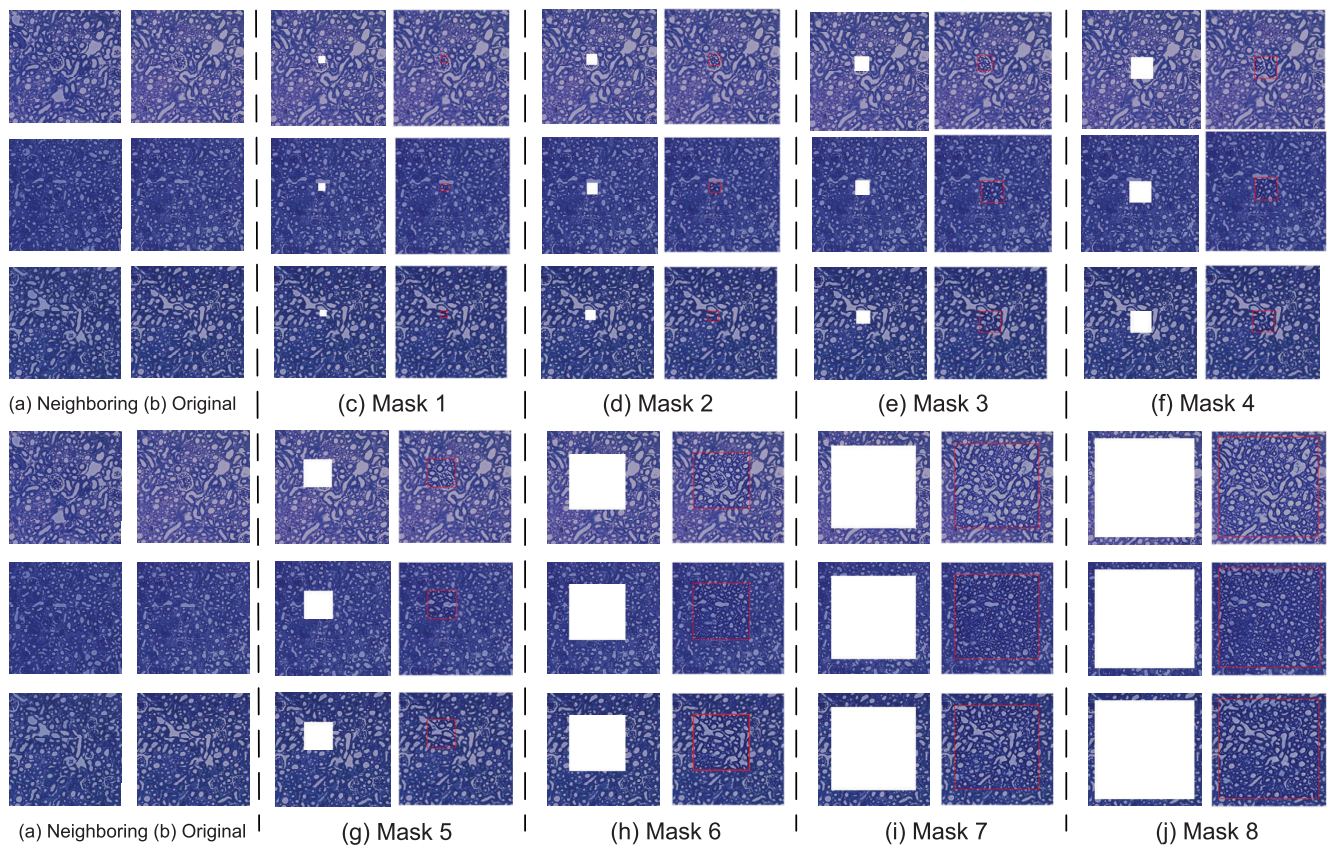
I and Q are In-phase and Quadrature-phase in YIQ color space,  $\Omega$  represents the whole image; We set  $\lambda = 0.03$ ,  $T_1 = 0.85$ ,  $T_2 = 160$ ,  $T_3 = 200$  and  $T_4 = 200$  as in [42].

PSNR is used as consistent quality metric, as in (10), which directly measures the difference in pixel levels. The higher value means the better image quality. Usually, it could be considered as same as the original image, if  $PSNR > 40$ . Due to the ignorance of the character of human visual frequency response, PSNR evaluation differs from subjective perception of human sometimes.

$$PSNR = 10 \lg \left\{ \frac{255^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - Y(i, j))^2} \right\} \quad (10)$$

where,  $X(i, j)$  denotes the original image, and  $Y(i, j)$  denotes the inpainting image.

LPIPS metric is based on deep convolution neural network to calculate visual similarity and has been proved to correlate



**FIGURE 5.** Completion results under different mask sizes. (a) neighboring images. (b) original images (Ground Truth). (c)-(j) random masks with different sizes (Left) and inpainting results (Right, red boxes highlight the restoration patch). (c) mask size of  $32 \times 32$  pixels. (d) mask size of  $48 \times 48$  pixels. (e) mask size of  $64 \times 64$  pixels. (f) mask size of  $96 \times 96$  pixels. (g) mask size of  $128 \times 128$  pixels. (h) mask size of  $256 \times 256$  pixels. (i) mask size of  $384 \times 384$  pixels. (j) mask size of  $448 \times 448$  pixels. The restoration areas are in red boxes.

well with perceptual judgments. The lower value of LPIPS mean the better quality and the higher value of other metrics mean better. [45]

## 2) QUALITATIVE EVALUATION

Medical images are usually acquired under standardized procedures, therefore, variations are very little between images. Researchers pay more attention to the accuracy of image contents rather than visual results or quantitative scores in medical image inpainting comparing daily life image inpainting. In this study, 3 experienced histologists are invited to evaluate the visual quality in aspects of color consistency, structure and texture accuracy, border continuity of patch, and correctness of morphology in different sizes or places of masks.

In the following, we present both quantitative and qualitative evaluation of our model, and the comparison with recent popular image inpainting methods with Contextual Attention (CA) [17], Globally and Locally Consistent Image Completion (GLC) [27] and Pyramid-Context Encoder Network (PEN) [46].

### C. EXPERIMENT 1

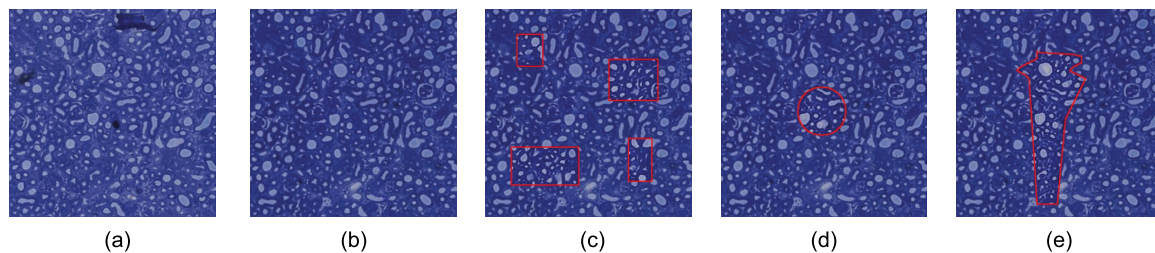
In this experiment, we first performed and demonstrated the assessment of our method in training iteration, mask sizes and then compared with the popular methods. A total of 275 images from N7 dataset, which were compact in

microstructure, were selected randomly as training data, and 164 images as testing data.

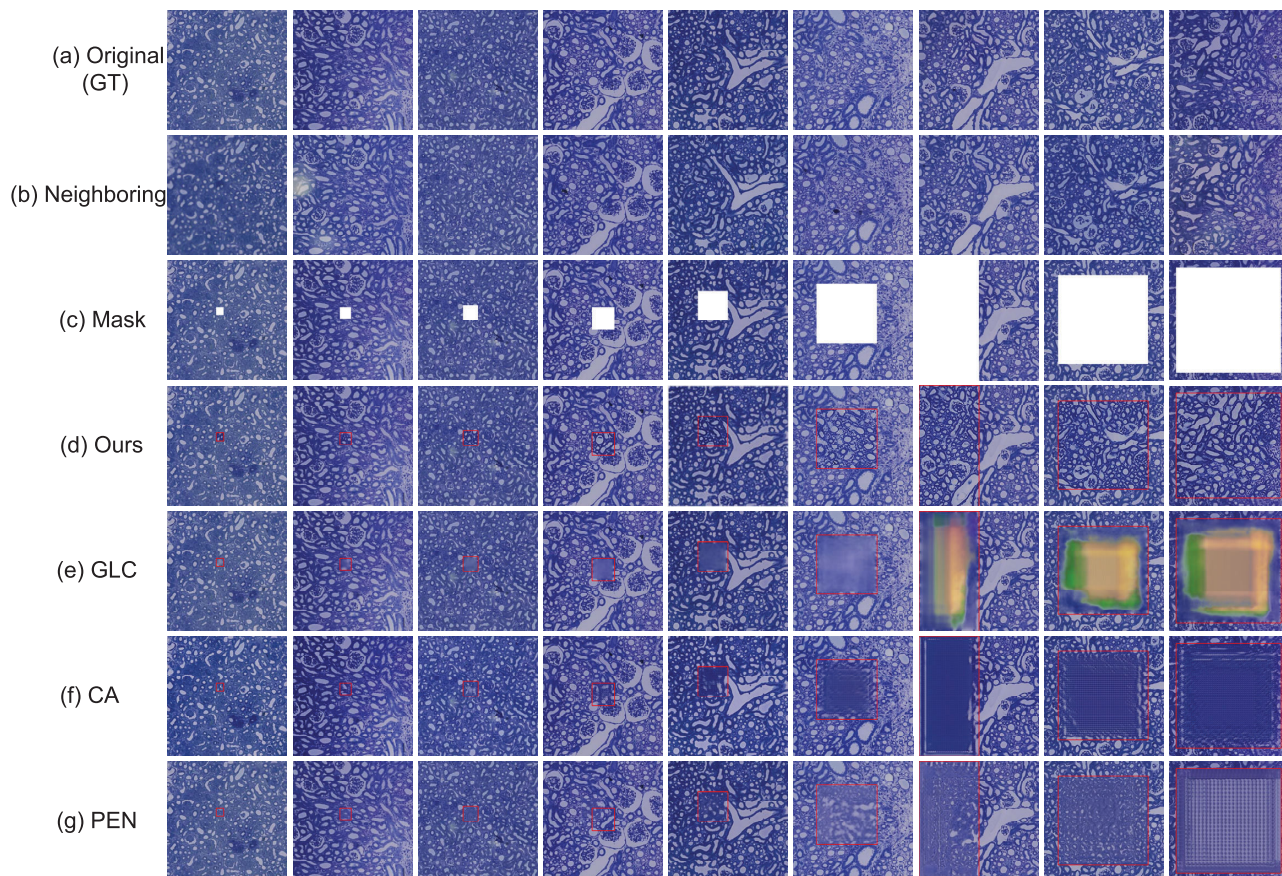
We have conducted an experiment to evaluate the performance when the number of iterations increased, the result suggested that the performance enhancements of our model almost stopped after 130000 iterations. Therefore, iterations of 150000 was a balance between training cost and the model performance, as shown in Fig. 4. We also tested the performance with 100 images ( $512 \times 512$ ) and got the computing consumption of  $52.3 \pm 1.9$ ms per image.

The sizes and locations of masks were further validated effects of the performance in our model and we wished to identify the reliability of our model. In this study, 8 kinds of square masks were used, the side length of which were from 32 pixels to 448 pixels, equivalent to coverage area of 0.39% to 76.56%. As shown in Fig. 5, our model could repair the masked images with information learned from the corresponding parts of neighboring images. In the situation of small-sized missing patch (Masks 1-4, missing area  $\leq 5\%$  image area), the shapes and details of glomeruli, vessels and renal tubules were well restored comparing with the original images; in the situation of median-sized missing patch (Masks 5-6,  $5\% \text{ image area} < \text{missing area} \leq 25\% \text{ image area}$ ), similar results were shown; as the missing area increased (Masks 7-8, missing area  $> 25\% \text{ image area}$ ), most of structures were also well restored, but at the boundary of





**FIGURE 6.** Multiple damaged validation: (a) neighboring image. (b) original image (Ground Truth). (c) result with multi-mask. (d)-(e) results with arbitrary shape. The restoration areas are in red boxes.



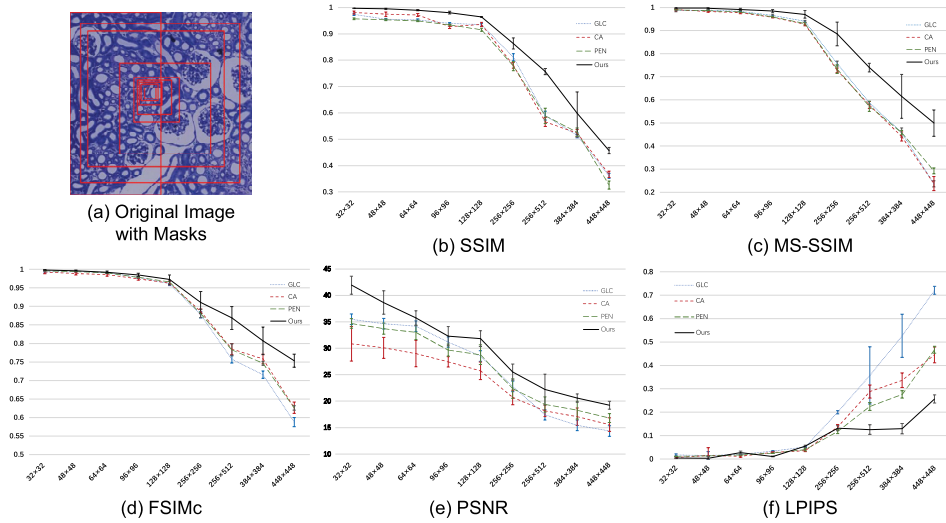
**FIGURE 7.** Section completion results of different mask types on the N7 dataset. From the fourth line to the seventh line are the results of different methods of repairing. These images, from left to right, are random block (side length 32, 48, 64, 96, 128, 256 pixels) and left block and center block (side length 384,448pixels) masked images. As the mask becomes larger, our result is more effective. The restoration areas are in red boxes.

the mask, some of contours of the microstructures did not have good continuity.

The location and count of masks was another issue that we concerned about. The images with 4 masks, 20% coverage area in total, were tested, which showed a similar result to the images with one single mask as shown in Fig. 6. Furthermore, different shapes of masks, such as polygon and circle, were also tested in our study. The experiment on real defect image could not be evaluated due to the lack of ground truth, so we draw a mask similar to real condition as shown in Fig. 6 (e). We could conclude that our model was suitable for the restoration of serial sectioning images and irrelevant to the number, size or shape of missing areas.

Then we compared our model with CA, GLC and PEN, as shown in Fig. 7. Qualitative evaluation suggested that:

- (1) In the situation of small-sized missing patch (Masks 1-4, missing area  $\leq 5\%$  image area), CA, PEN and ours could provide pretty good visual results, results from GLC were acceptable in overall appearance.
- (2) All of the 4 methods provided shapes and details of glomeruli, vessels and renal tubules in visual effect, but only the results from ours were correct in content comparing with the original images.
- (3) In the situation of median-sized missing patch (Masks 5-6,  $5\% \text{ image area} < \text{missing area} \leq 25\% \text{ image area}$ ), restorations from CA and GLC were blurred or distorted as there was less effective information available around the masks, PEN was only effective in visual effects; but our performance was still robust.
- (4) in the situation of large-sized missing patch (Masks 7-9, missing area  $> 25\% \text{ image area}$ ), our method could still provide valuable information. We could conclude



**FIGURE 8.** Quantitative evaluation between GLC, CA, PEN and our model on the N7 dataset. (a) Masks. (b) SSIM evaluation. (c) MS-SSIM evaluation. (d) FSIMc evaluation. (e) PSNR evaluation. (f) LPIPS evaluation.

**TABLE 1.** The quantitative evaluation on N7 dataset.

Mask Size	Method	SSIM	MS-SSIM	FSIMc	PSNR	LPIPS
0.39% 32 × 32 pixels	GLC	0.9740±0.0038*	0.9866±0.0051*	0.9932±0.0031*	35.4763±0.8371*	0.0176±0.0049*
	CA	0.9809±0.0061*	0.9902±0.0051*	0.9927±0.0044*	30.8410±3.2786*	0.0068±0.0039*
	PEN	0.9568±0.0033*	0.9897±0.0011*	0.9959±0.0003*	34.6827±0.9099*	0.0117±0.0014*
	Ours	<b>0.9971±0.0009</b>	<b>0.9978±0.0010</b>	<b>0.9983±0.0009</b>	<b>41.9384±1.7050</b>	<b>0.0046±0.0020</b>
0.879% 48 × 48 pixels	GLC	0.9552±0.0032*	0.9883±0.0009*	0.9946±0.0004*	34.6416±0.7361*	0.0151±0.0167*
	CA	0.9748±0.0085*	0.9824±0.0034*	0.9886±0.0037*	30.0894±1.9819*	0.0152±0.0340#
	PEN	0.9533±0.0035*	0.9858±0.0018*	0.9938±0.0005*	33.7042±1.0609*	0.0138±0.0016*
	Ours	<b>0.9946±0.0020</b>	<b>0.9960±0.0024</b>	<b>0.9963±0.0016</b>	<b>38.6418±2.2218</b>	<b>0.0037±0.0075</b>
1.56% 64 × 64 pixels	GLC	0.9528±0.0034*	0.9830±0.0017*	0.9910±0.0012*	34.1662±0.7595*	0.0200±0.0015*
	CA	0.9714±0.0057*	0.9777±0.0049*	0.9859±0.0045*	28.9851±2.4947*	<b>0.0119±0.0042*</b>
	PEN	0.9491±0.0032*	0.9802±0.0017*	0.9905±0.0013*	33.0027±1.3602*	0.0177±0.0036*
	Ours	<b>0.9896±0.0025</b>	<b>0.9912±0.0047</b>	<b>0.9924±0.0033</b>	<b>35.7344±1.3770</b>	0.0276±0.0058
3.52% 96 × 96 pixels	GLC	0.9397±0.0032*	0.9655±0.0025*	0.9783±0.0028*	31.1678±1.0124*	0.0338±0.0020*
	CA	0.9273±0.0072*	0.9568±0.0021*	0.9742±0.0034*	27.4344±0.9696*	0.0284±0.0033*
	PEN	0.9336±0.0038*	0.9588±0.0021*	0.9784±0.0025*	29.6383±1.1367*	0.0256±0.0013*
	Ours	<b>0.9805±0.0054</b>	<b>0.9842±0.0072</b>	<b>0.9850±0.0045</b>	<b>32.2914±1.8177</b>	<b>0.0113±0.0017</b>
6.25% 128 × 128 pixels	GLC	0.9349±0.0051*	0.9402±0.0033*	0.9614±0.0055*	28.5373±0.9897*	0.0521±0.0022*
	CA	0.9356±0.0077*	0.9276±0.0072*	0.9632±0.0049*	25.7342±1.6500*	<b>0.0358±0.0042*</b>
	PEN	0.9154±0.0059*	0.9303±0.0055*	0.9662±0.0053*	28.7789±1.8669*	0.0402±0.0049*
	Ours	<b>0.9644±0.0011</b>	<b>0.9701±0.0163</b>	<b>0.9726±0.0121</b>	<b>31.8220±1.5000</b>	0.0546±0.0046
25% 256 × 256 pixels	GLC	0.8111±0.0136*	0.7571±0.0102*	0.8775±0.0084*	22.8053±1.3765*	0.1999±0.0067*
	CA	0.7805±0.0111*	0.7279±0.0104*	0.8866±0.0066*	20.7046±1.3765*	0.1368±0.0111*
	PEN	0.7775±0.0171*	0.7329±0.0189*	0.8807±0.0113*	22.3591±1.8408*	<b>0.1182±0.0093*</b>
	Ours	<b>0.8635±0.0206</b>	<b>0.8852±0.0509</b>	<b>0.9107±0.0293</b>	<b>25.5060±1.5100</b>	0.1322±0.0106
50% 256 × 512 pixels	GLC	0.5896±0.0171*	0.5873±0.0063*	0.7573±0.0104*	17.4100±1.0332*	0.3574±0.1227*
	CA	0.5659±0.0191*	0.5777±0.0154*	0.7849±0.0150*	18.1650±1.0963*	0.2885±0.0280#
	PEN	0.5887±0.0292*	0.5719±0.0216*	0.7827±0.0161*	19.3713±1.4419*	0.2250±0.0168*
	Ours	<b>0.7568±0.1113</b>	<b>0.7399±0.0186</b>	<b>0.8687±0.0310</b>	<b>22.1998±2.9034</b>	<b>0.1261±0.2041</b>
56.25% 384 × 384 pixels	GLC	0.5144±0.0093*	0.4584±0.0199*	0.7155±0.0100*	15.4223±0.2696*	0.5268±0.0926*
	CA	0.5233±0.0155*	0.4413±0.0200*	0.7582±0.0129*	17.0571±1.6673*	0.3368±0.0310*
	PEN	0.5266±0.0167*	0.4584±0.0096*	0.7465±0.0052*	18.2866±0.6788*	0.2764±0.0158*
	Ours	<b>0.5981±0.0816</b>	<b>0.6155±0.0951</b>	<b>0.8072±0.0372</b>	<b>20.5657±0.7871</b>	<b>0.1298±0.0226</b>
76.56% 448 × 448 pixels	GLC	0.3617±0.0094*	0.2366±0.0133*	0.5877±0.0123*	14.3551±0.2650*	0.7210±0.0164*
	CA	0.3670±0.0123*	0.2375±0.0303*	0.6264±0.0152*	15.5508±1.2736*	0.4450±0.0337*
	PEN	0.3265±0.0154*	0.2923±0.0130*	0.6259±0.0060*	16.7891±0.7770*	0.4668±0.0148*
	Ours	<b>0.4571±0.1164</b>	<b>0.4993±0.0567</b>	<b>0.7533±0.0175</b>	<b>19.2338±0.7384</b>	<b>0.2568±0.0173</b>

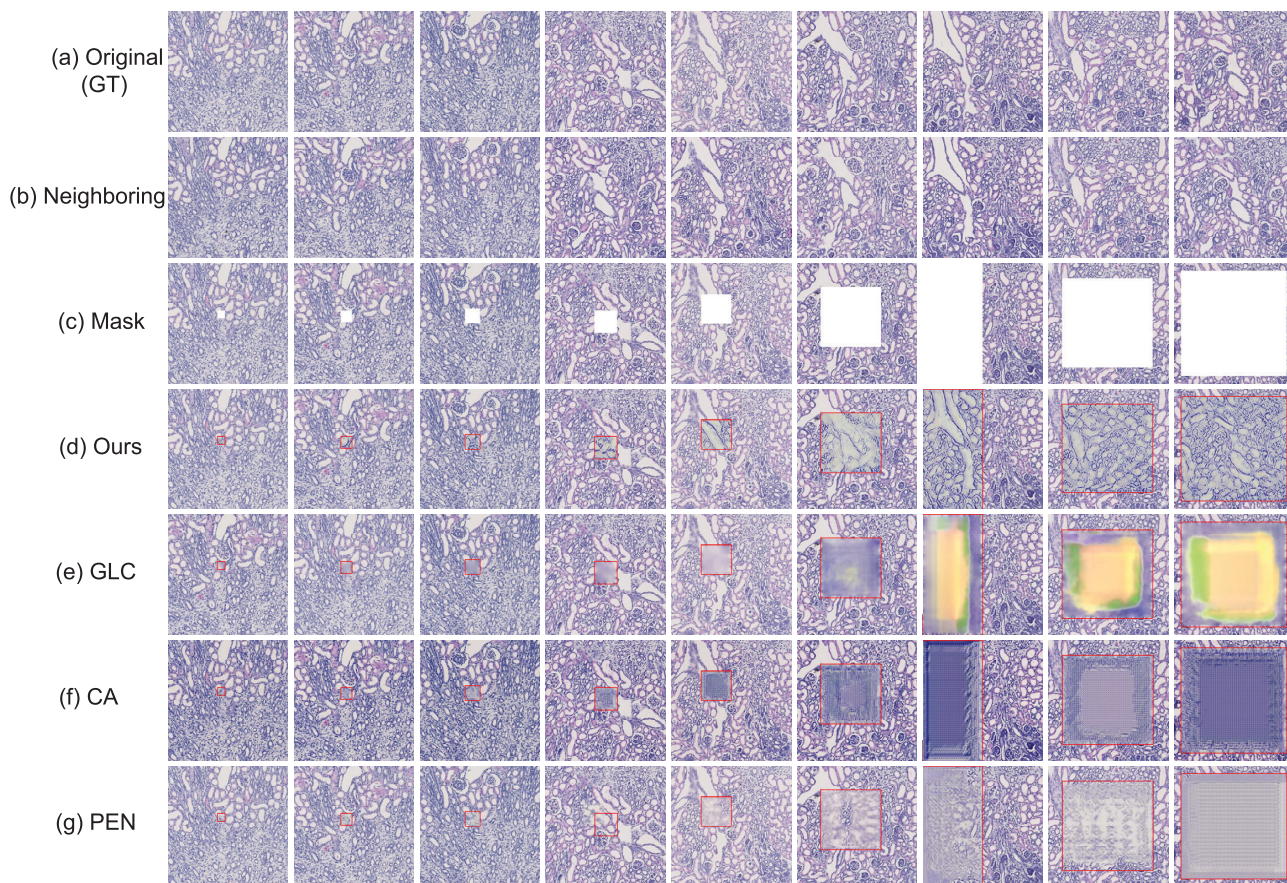
\* P < 0.01

# P < 0.05

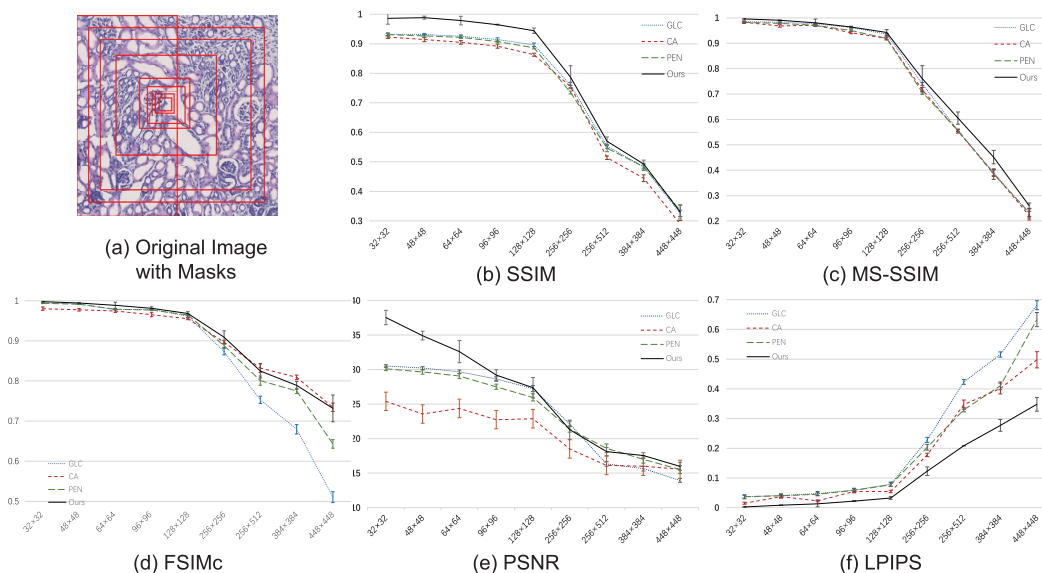
that our model was more effective in both visual effects and providing correct details and microstructures of complicated tissue with any size of damaged areas.

The quantitative evaluation was carried out using 5 indicators: SSIM, MS-SSIM, FSIMc, PSNR and LPIPS. We chose 9 different-sized and placed masks, as shown in Fig. 8. Table I showed the inpainting results and the paired t-test. T-test suggested significant difference between our method and

popular inpainting methods. SSIM/MS-SSIM are sensitive to structural similarity [40] and the results showed our model performed higher than the best of other methods. SSIM: 2.59% (missing size ≤ 128 × 128), 17.41% (256 × 256 ≤ missing size ≤ 256 × 512), 19.07% (384 × 384 ≤ missing size ≤ 448 × 448); MS-SSIM: 1.50% (missing size ≤ 128 × 128), 21.45% (256 × 256 ≤ missing size ≤ 256 × 512), 52.54% (384 × 384 ≤ missing size ≤ 448 × 448). FSIMc is sensitive



**FIGURE 9.** Section completion results of different mask types on the N5 dataset. From the fourth line to the seventh line are the results of different methods of repair. These images, from left to right of, are random block (side length 32,48,64,96,128,256 pixels) and left block and center block (side length 384,448pixels) masked images. As the mask becomes larger, our result is more effective. The restoration areas are in red boxes.



**FIGURE 10.** Quantitative evaluation between GLC, CA, PEN and our model on the N5 dataset. (a) Masks. (b) SSIM evaluation. (c) MS-SSIM evaluation. (d) FSIMc evaluation. (e) PSNR evaluation. (f) LPIPS evaluation.

to low-level boundary features [42] and its results showed the best among the methods: 0.38% (missing size  $\leq 128 \times 128$ ), 6.70% ( $256 \times 256 \leq$  missing size  $\leq 256 \times 512$ ), 13.36%

( $384 \times 384 \leq$  missing size  $\leq 448 \times 448$ ). PSNR only calculates pixel values but not structural features and the results showed: 9.71% (missing size  $\leq 128 \times 128$ ), 13.20%

TABLE 2. The quantitative evaluation on N5 dataset.

Mask Size	Method	SSIM	MS-SSIM	FSIMc	PSNR	LPIPS
0.39% 32 × 32 pixels	GLC	0.9323±0.0047*	0.9851±0.0014*	0.9947±0.0005*	30.5129±0.1728*	0.0368±0.0068*
	CA	0.9227±0.0048*	0.9828±0.0048*	0.9801±0.0042*	25.4007±1.6286*	0.0151±0.0032*
	PEN	0.9308±0.0045*	0.9825±0.0014*	0.9935±0.0005*	30.1072±0.2468*	0.0372±0.0049*
	Ours	<b>0.9857±0.0196</b>	<b>0.9958±0.0009</b>	<b>0.9976±0.0006</b>	<b>37.5294±1.0396</b>	<b>0.0026±0.0012</b>
0.879% 48 × 48 pixels	GLC	0.9312±0.0039*	0.9826±0.0020*	0.9918±0.0009*	30.2128±0.2013*	0.0408±0.0061*
	CA	0.9141±0.0067*	0.9682±0.0048*	0.9774±0.0031*	23.5627±0.9933*	0.0371±0.0049*
	PEN	0.9270±0.0044*	0.9781±0.0016*	0.9912±0.0007*	29.6518±0.3077*	0.0403±0.0049*
	Ours	<b>0.9883±0.0050</b>	<b>0.9904±0.0015</b>	<b>0.9943±0.0007</b>	<b>34.9147±0.6288</b>	<b>0.0085±0.0009</b>
1.56% 64 × 64 pixels	GLC	0.9256±0.0046*	0.9769±0.0023	0.9781±0.0014*	29.6707±0.2682*	0.0475±0.0071*
	CA	0.9048±0.0060*	0.9709±0.0039*	0.9742±0.0034*	24.3636±1.3117*	0.0231±0.0033*
	PEN	0.9221±0.0044*	0.9711±0.0025*	0.9787±0.0013*	29.0672±0.3735*	0.0456±0.0049*
	Ours	<b>0.9784±0.0143</b>	<b>0.9806±0.0146</b>	<b>0.9886±0.0078</b>	<b>32.6054±1.6246</b>	<b>0.0133±0.0102</b>
3.52% 96 × 96 pixels	GLC	0.9141±0.0065*	0.9627±0.0031*	0.9792±0.0013*	28.6424±0.3158*	0.0591±0.0067*
	CA	0.8919±0.0067*	0.9414±0.0040*	0.9652±0.0051*	22.7519±0.9464*	0.0547±0.0044*
	PEN	0.9075±0.0042*	0.9500±0.0031*	0.9766±0.0015*	27.5055±0.3776*	0.0585±0.0055*
	Ours	<b>0.9646±0.0014</b>	<b>0.9641±0.0030</b>	<b>0.9811±0.0037</b>	<b>29.2070±0.7709</b>	<b>0.0224±0.0017</b>
6.25% 128 × 128 pixels	GLC	0.8963±0.0058*	0.9358±0.0032*	0.9610±0.0037*	27.2427±0.4380	0.0788±0.0073*
	CA	0.8630±0.0049*	0.9191±0.0061*	0.9551±0.0024*	22.8809±0.7338*	0.0546±0.0041*
	PEN	0.8865±0.0042*	0.9189±0.0050*	0.9645±0.0024*	25.9734±0.5235*	0.0777±0.0059*
	Ours	<b>0.9443±0.0091</b>	<b>0.9423±0.0115</b>	<b>0.9682±0.0045</b>	<b>27.3740±1.4813</b>	<b>0.0327±0.0045</b>
25% 256 × 256 pixels	GLC	0.7594±0.0104*	0.7399±0.0054*	0.8728±0.0078*	<b>22.1332±0.4126*</b>	0.2283±0.0086*
	CA	0.7551±0.0045*	0.7171±0.0090*	0.8964±0.0047*	18.5037±1.1269*	0.1778±0.0060*
	PEN	0.7350±0.0048*	0.7072±0.0075*	0.8881±0.0046*	21.3863±0.4549*	0.2035±0.0094*
	Ours	<b>0.7867±0.0388</b>	<b>0.7608±0.0508</b>	<b>0.9092±0.0159</b>	21.2944±1.3836	<b>0.1230±0.0145</b>
50% 256 × 512 pixels	GLC	0.5526±0.0067*	0.5566±0.0056*	0.7530±0.0090*	16.3336±0.3445*	0.4231±0.0081*
	CA	0.5140±0.0059*	0.5533±0.0072*	<b>0.8323±0.0110*</b>	16.1083±0.6962*	0.3470±0.0157*
	PEN	0.5461±0.0117*	0.5566±0.0035*	0.8009±0.0114*	<b>18.5924±0.6365*</b>	0.3297±0.0068*
	Ours	<b>0.5696±0.0149</b>	<b>0.6053±0.0242</b>	0.8250±0.0173	18.1240±0.5921	<b>0.2084±0.0120</b>
56.25% 384 × 384 pixels	GLC	0.4829±0.0068*	0.3900±0.0112*	0.6797±0.0116*	15.6527±0.3486*	0.5158±0.0090*
	CA	0.4447±0.0114*	0.3889±0.0132*	<b>0.8091±0.0056*</b>	15.9854±0.7301*	0.4010±0.0183*
	PEN	0.4830±0.0117*	0.3846±0.0208*	0.7751±0.0063*	17.0010±0.5585*	0.4101±0.0132*
	Ours	<b>0.4926±0.0134</b>	<b>0.4515±0.0266</b>	0.7889±0.0098	<b>17.5396±0.4218</b>	<b>0.2772±0.0198</b>
76.56% 448 × 448 pixels	GLC	0.3251±0.0113	0.2298±0.0177*	0.5106±0.0133*	13.9467±0.2822*	0.6817±0.0151*
	CA	0.2896±0.0117*	0.2212±0.0171*	<b>0.7345±0.0107</b>	15.5665±1.1480#	0.4978±0.0274*
	PEN	<b>0.3351±0.0196</b>	0.2332±0.0153*	0.6432±0.0107*	15.5100±0.5731*	0.6333±0.0239*
	Ours	0.3300±0.0235	<b>0.2603±0.0105</b>	0.7317±0.0333	<b>15.9828±0.5674</b>	<b>0.3480±0.0230</b>

\* P &lt; 0.01

# P &lt; 0.05

(256 × 256 ≤ missing size ≤ 256 × 512), 13.50% (384 × 384 ≤ missing size ≤ 448 × 448). LPIPS calculates visual similarity by comparing the deep structure and deep features between images, which is more consistent with human visual effect and sensitive to nuances of images in human perception. Our model generated better results and obtained lower values among all the methods: 19.94% (missing size ≤ 256 × 512), 46.43% (384 × 384 ≤ missing size ≤ 448 × 448). The indicators above suggested our model performed best among all tested models. The reason was that our method could utilize features from its neighboring image. All the methods performed well in small and even median size condition. As the mask size became larger, our method showed a greater advantage than other methods. When the mask size became extremely large, other methods invalidated.

Our model was further applied to 230 damaged images in N7 dataset. Three histologists invited in our experiment assessed correctness of inpainting region individually and they valued our model performed best among image inpainting models.

## D. EXPERIMENT 2

The performance of generalization of our model has also been assessed in this experiment. Notably, no new model was trained in this section, the model in previous experiment was directly verified with another dataset N5. As shown in Fig. 9, Fig. 10, and Table II, similar quantitative conclusions could

be drawn from these results. Most of the results showed significant difference, but results of images with 448 × 448 size mask in SSIM suggested P > 0.05 because much information of neighboring image was referred and the inpainting image would be more similar to neighboring image than the ground truth. The model was further applied to 75 damaged images in N5 dataset and experts gave the same evaluation as in Experiment 1.

Experiment 2 also suggested that our model was still effective in the dataset of N5, although there was no further training.

## V. CONCLUSION

In this study, we proposed a novel Consecutive Context Perceive Generative Adversarial Networks (CCPGAN) for serial sections inpainting in medical study. The framework is composed of an SPM, a refined generating network, and two discriminators. To our knowledge, it is the first time to introduce 2-pathway deep learning networks in image inpainting field, which could detect histological and structured characteristics from both the damaged images and its neighboring images. Concatenation is employed during the upsampling operation to combine the features to reconstruct more details. Qualitative comparison and quantitative comparison suggest that our method could complete the damaged area to acquire desirable measurements and visual effects, keeping both global and local consistency, as well as generating correct objects,

such as an individual glomerulus sporadic distributed in the sectioning. The quality of inpainting image is irrelevant to mask location or the staining method. For the medical image analysis, the correctness of the result is much more important than consistent with visual effects and our method utilizes neighboring image as a reference so that more information and details can be learned for restoration. There is a limitation when the damaged patch became extremely large. The restored region will be more similar to the neighboring image as little surrounding and much neighboring information is used; although this situation scarcely occurs.

## ACKNOWLEDGMENT

(Siqi Zhang and Lei Wang contributed equally to this work.)

## REFERENCES

- X. Y. Zhai, "Digital three-dimensional reconstruction and ultrastructure of the mouse proximal tubule," *J. Amer. Soc. Nephrol.*, vol. 14, no. 3, pp. 611–619, Mar. 2003.
- J. F. Bertram, L. A. Cullen-McEwen, G. F. Egan, N. Gretz, E. Balde-lomar, S. C. Beeman, and K. M. Bennett, "Why and how we determine nephron number," *Pediatric Nephrol.*, vol. 29, no. 4, pp. 575–580, Apr. 2014.
- A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Detection of missing data in image sequences," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1496–1508, 1995.
- A. Halim and B. V. R. Kumar, "An anisotropic PDE model for image inpainting," *Comput. Math. Appl.*, vol. 79, no. 9, pp. 2701–2721, May 2020.
- M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. 27th Annu. Conf. Comput. Graph. Interact. Techn.*, 2000, pp. 417–424.
- T. F. Chan, J. Shen, and H.-M. Zhou, "Total variation wavelet inpainting," *J. Math. Imag. Vis.*, vol. 25, no. 1, pp. 107–125, Jul. 2006.
- Y. Duan, Y. Wang, and J. Hahn, "A fast augmented lagrangian method for Euler's elastica models," *Numer. Math. Theory Methods Appl.*, vol. 6, no. 1, pp. 47–71, 2013.
- J. Liu, S. Yang, Y. Fang, and Z. Guo, "Structure-guided image inpainting using homography transformation," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3252–3265, Dec. 2018.
- A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
- Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4539–4547.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.
- D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.
- A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2634–2645, Oct. 2010.
- M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: <https://arxiv.org/abs/1701.07875>
- I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
- Y. Yu, Z. Gong, P. Zhong, and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *Proc. Int. Conf. Image Graph.* Springer, 2017, pp. 97–108.
- J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.
- C. Wang, S. Dong, X. Zhao, G. Papanastasiou, H. Zhang, and G. Yang, "SaliencyGAN: Deep learning semisupervised salient object detection in the fog of IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2667–2676, Apr. 2020.
- C. Wang, G. Papanastasiou, S. Tsafaris, G. Yang, C. Gray, D. Newby, G. Macnaught, and T. MacGillivray, "Tpsdicyc: Improved deformation in-variant cross-domain medical image synthesis," in *Proc. Int. Workshop Mach. Learn. Med. Image Reconstruction*. Springer, 2019, pp. 245–254.
- G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.
- L. Jiao, H. Wu, H. Wang, and R. Bie, "Multi-scale semantic image inpainting with residual learning and GAN," *Neurocomputing*, vol. 331, pp. 199–212, Feb. 2019.
- W. Cai and Z. Wei, "PiiGAN: Generative adversarial networks for pluralistic image inpainting," *IEEE Access*, vol. 8, no. 1, pp. 48451–48463, 2020.
- Z. Li, H. Zhu, L. Cao, L. Jiao, Y. Zhong, and A. Ma, "Face inpainting via nested generative adversarial networks," *IEEE Access*, vol. 7, pp. 155462–155471, 2019.
- E. Sogancioglu, S. Hu, D. Belli, and B. van Ginneken, "Chest X-ray inpainting with deep generative models," 2018, *arXiv:1809.01471*. [Online]. Available: <http://arxiv.org/abs/1809.01471>
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, 2017.
- U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018, *arXiv:1803.07422*. [Online]. Available: <https://arxiv.org/abs/1803.07422>
- J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 4471–4480.
- G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 85–100.
- M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.
- T. Zhou, S. Tulsiani, W. Sun, J. Malik, and A. A. Efros, "View synthesis by appearance flow," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 286–301.
- J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 764–773.
- Y. Jeon and J. Kim, "Active convolution: Learning the shape of convolution for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4201–4209.
- J. Zhu, G. Yang, and P. Lio, "How can we make gan perform better in single medical image super-resolution? A lesion focused multi-scale approach," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, 2019, pp. 1669–1673.
- R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5485–5493.
- J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
- X.-Y. Zhai, J. S. Thomsen, H. Birn, I. B. Kristoffersen, A. Andreassen, and E. I. Christensen, "Three-dimensional reconstruction of the mouse nephron," *J. Amer. Soc. Nephrol.*, vol. 17, no. 1, pp. 77–88, Jan. 2006.
- Y.-L. Zhang, S.-J. Chang, X.-Y. Zhai, J. S. Thomsen, E. I. Christensen, and A. Andreassen, "Non-rigid landmark-based large-scale image registration in 3-D reconstruction of mouse and rat kidney nephrons," *Micron*, vol. 68, pp. 122–129, Jan. 2015.
- Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Rec. Signals, Syst. Comput.*, 2003.

- [42] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [43] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.
- [44] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.
- [45] L. Yuan, C. Ruan, H. Hu, and D. Chen, "Image inpainting based on patch-GANs," *IEEE Access*, vol. 7, pp. 46411–46421, 2019.
- [46] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 1486–1494.



**XIRAN JIANG** received the Ph.D. degree from Fudan University. He is currently working as a Professor of biomedical engineering with China Medical University. His research interests include artificial intelligence in clinical imaging and microfluidic technologies.



**SIQI ZHANG** received the bachelor's degree in biomedical engineering from Jilin University, Changchun, China, in 2017. She is currently pursuing the degree with China Medical University, Shenyang, China. Her research interests include deep learning, image inpainting of mouse kidney section, small target detection for glomerulus, and the three-dimensional reconstruction of developing and adult mouse kidney.



**XIAOYUE ZHAI** received the bachelor's degree in clinical medicine from Jinzhou Medical College, Jinzhou, Liaoning, China, in 1985, the master's degree in histology and embryology from China Medical University, Shenyang, Liaoning, China, in 1992, and the Ph.D. degree in medicine from Aarhus University, Aarhus, Denmark, in 2006. From 1998 to 2006, she studied mainly in Denmark, two years as a Visiting Student, followed by three years Ph.D. education and Postdoctoral training. Her research interests include morphogenesis of kidney, cellular and molecular mechanism, the spatial arrangement of renal tubules, and vessels using computer assisted 3D visualization technique. Meanwhile, she focuses the mechanism related to kidney development in recovery of acute kidney injury.



**LEI WANG** received the bachelor's degree in electronic information engineering from Shandong University, Jinan, China, in 2014. He is currently pursuing the degree with China Medical University, Shenyang, China. His research interests include the region of deep learning and image processing.



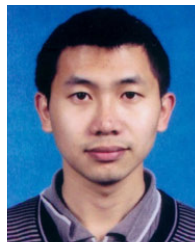
**JIE ZHANG** received the bachelor's degree in medicine in medical information management and the Ph.D. degree in human anatomy, histology, and embryology from China Medical University, Shenyang, China, in 2011 and 2018, respectively. Since 2018, she has been a Lecturer with the Department of Histology and Embryology, College of Basic Medicine, China Medical University. Her research interests include the three-dimensional reconstruction of developing and adult mouse kidney, development of renal vasculature, and vascular dysfunction in chronic kidney disease.



**XIANZHENG SHA** received the B.S. degree in radio electronics from Shandong University, in 1983, and the M.S. degree in biomedical engineering from China Medical University, in 1991. He is currently a Professor and the Dean of the Department of Biomedical Engineering, China Medical University. He is also a member of Chinese Society of Biomedical Engineering and the Chairman of sensor branch of Chinese Society of Biomedical Engineering. His research interests include sensors and machine learning.



**LING GU** received the Ph.D. degree in histology and embryology from China Medical University, in 2017. Her research interests include the morphogenesis of developing kidneys and based on the 3D reconstruction of developing renal tubules.



**SHIJIE CHANG** received the B.S. degree from Northeastern University, in 2004, and the M.S. and Ph.D. degrees from China Medical University, in 2007 and 2013, respectively. He is currently the Vice-Professor with the Department of Biomedical Engineering, China Medical University. His research interests include artificial intelligence, machine learning, and data analysis.

...