



OPEN

An image inpainting-based data augmentation method for improved sclerosed glomerular identification performance with the segmentation model EfficientNetB3-Unet

Songping He¹, Yi Zou², Bin Li¹, Fangyu Peng², Xia Lu³, Hui Guo³, Xin Tan⁴ & Yanyan Chen⁵✉

The percent global glomerulosclerosis is a key factor in determining the outcome of renal transfer surgery. At present, the rate is typically computed by pathologists, which is labour intensive and nonstandardized. With the development of Deep Learning (DL), DL-based segmentation models can be used to better identify and segment normal and sclerosed glomeruli. Based on this, we can better quantify percent global glomerulosclerosis to reduce the discard rate of donor kidneys. We used 51 whole slide images (WSIs) from different institutions that are publicly available on the internet. However, the number of sclerosed glomeruli is much smaller than that of normal glomeruli in different WSIs, which can reduce the effectiveness of Deep Learning. For better sclerosed glomerular identification and segmentation performance, we modified and trained a GAN (generative adversarial network)-based image inpainting model to obtain more synthetic sclerosed glomeruli. Our proposed inpainting method achieved an average SSIM (Structural Similarity) of 0.8086 and an average PSNR (Peak Signal-to-Noise Ratio) of 22.8943 dB in the area of generated sclerosed glomeruli. We obtained sclerosed glomerular segmentation performance improvement by adding synthetic sclerosed glomerular images and achieved the best Dice of glomerular segmentation in different test sets based on the modified Unet model.

There are a large number of patients in need of kidney transplantation waiting for kidney donors, and this demand is still growing¹. Meanwhile, many studies have shown that chronic damage to donor kidney biopsy specimens is closely related to transplant outcomes, so approximately 17–20% of collected donor kidneys need to be discarded after pathologist evaluation². Additionally, in daily practice, often due to the urgency of time, different pathologists will have subjective biases when evaluating sections, potentially resulting in unnecessary discarding of organs. We need to minimize the occurrence of this discarding due to the shortage of kidney donors.

In kidney transplant evaluation, there are many indicators to consider, among which the percent of global glomerulosclerosis is considered to be the entry point for kidney transplantation³. Due to the large number of glomeruli, the assessment of percent global glomerulosclerosis is very time-consuming and causes poor reproducibility among pathologists. The professional knowledge requirements of pathologists are high, and human error easily occurs. Therefore, automatic image processing methods that can accurately detect and classify glomeruli are needed.

¹Digital Manufacturing Equipment National Engineering Research Center, Huazhong University of Science and Technology, Wuhan, China. ²National NC System Engineering Research Center, Huazhong University of Science and Technology, Wuhan, China. ³Key Laboratory of Organ Transplantation of Ministry of Education, Institute of Organ Transplantation, Tongji Hospital, Tongji Medical College, National Health Commission and Chinese Academy of Medical Sciences, Huazhong University of Science and Technology, Wuhan, China. ⁴Wuhan Intelligent Equipment Industrial Institute Co Ltd, Wuhan, China. ⁵Department of Information Management, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ✉email: yychen@tjh.tjmu.edu.cn

Recently, due to the strong feature extraction ability of Deep Learning, an increasing number of studies have begun to use it to detect or segment objects in pathological images. In the imaging task, CNNs in particular are widely used. Unet, introduced by Ronneberger et al. based on CNN⁴, has proven to be useful in many tasks of tissue image segmentation and classification.

However, the performance of Deep Learning often depends on the quantity and quality of the training data. The acquisition of medical images involves the privacy of patients and requires the annotation of experts, so it is relatively difficult to obtain the training data of medical images. Meanwhile, in the publicly available data for glomerular studies, the number of sclerosed glomeruli is much smaller than that of normal glomeruli. Class imbalance can bring difficulties to Deep Learning.

In our study, we proposed a GAN-based image inpainting framework to generate more new sclerosed glomeruli from masks. Innovatively, newly generated sclerosed glomeruli were obtained based on the diverse shapes of the masks and the surrounding contextual information. Furthermore, we improved the segmentation network based on Unet and trained the model by combining the original data with new synthetic data. We realized the automatic segmentation and classification of normal and globally sclerosed glomeruli in digital pathological sections.

Related research

GAN and medical image generation

Since Ian Goodfellow proposed GAN in 2014, it has become possible to generate realistic images by designing the game process of the generator and discriminator⁵. Because of their powerful data generation capabilities, an increasing number of GANs have been used in the generation of pathological and medical images to perform data augmentation. In⁶, the combination of VAE and StyleGAN was proposed. The network generated the hidden code of the image through VAE as the input of StyleGAN to generate realistic cell images. In⁷, a medical image augmentation method, namely, a texture-constrained multichannel progressive generative adversarial network (TMP-GAN) was proposed. In⁸, Lei et al. proposed a lesion attention conditional generative adversarial network (LACGAN) to synthesize retinal images with realistic lesion details to improve the training of the disease detection model. Amirrajab, S. et al. proposed a novel framework consisting of image segmentation and synthesis based on mask-conditional GANs for generating high-fidelity and diverse Cardiac Magnetic Resonance (CMR) images⁹. Although different GAN-based frameworks have been applied in the generation of medical images, they still need to improve the performance in the field of pathological images, and there are also few studies on the generation of glomerular pathological images to improve segmentation performance.

Deep learning on glomerular identification and classification

In recent years, an increasing number of methods based on Deep Learning have been proposed to realize the identification and classification of glomeruli in digital pathological images. Each of these approaches has its own advantages and drawbacks. Jon N. Marsh et al. used a fully convolutional neural network based on the VGG16 architecture for glomerular segmentation and achieved 0.784 Aggregate Dice coefficients for nonglobally sclerosed glomeruli and 0.600 for globally sclerosed glomeruli¹⁰. Jaime Gallego et al. trained the Unet model on PAS-stained WSIs and H&E-stained WSIs. On the PAS-stained WSIs, normal and sclerosed glomeruli were classified with F1-scores of 97.5% and 68.8%, respectively. On H&E-stained WSIs, F1-scores of 90.8% and 78.1% were achieved¹¹. Gloria Bueno proposed the sequential CNN segmentation-classification strategy (SegNet-AlexNet) and this two-step framework achieved 98.16% accuracy in classifying normal and sclerosed glomeruli when trained on 47 PAS-stained WSIs¹². Lei Jiang et al. trained cascade mask region-based CNN architecture to detect, classify, and segment glomeruli into three categories: (i) GN, structural normal; (ii) global sclerosis; and (iii) glomerular with other lesions. They achieved F1 scores of 0.839, 0.806, and 0.753, respectively, in the whole-slide image group¹³. Tianyuan Yao et al. developed and released a holistic Glo-In-One open-source toolkit to provide holistic glomerular detection, segmentation, and lesion characterization¹⁴. Kawazoe et al. developed an automated computational pipeline for detecting glomeruli on PAS-stained WSIs, followed by segmenting Bowman's space, the glomerular tuft, the crescentic, and the sclerotic region inside of the glomeruli¹⁵. Silva et al. proposed the end-to-end network, named DS-FNet, combining the strengths of semantic segmentation and semantic boundary detection networks via an attention-aware mechanism, and it showed consistently high performance in a one-to-many-stain glomerulus segmentation¹⁶.

Most of the existing studies have not focused on the performance improvement of sclerosed glomerular segmentation. However, in fields such as kidney transplantation, the evaluation of sclerosed glomeruli is necessary and meaningful. Therefore, this paper focuses on improving the identification and segmentation performance of sclerosed glomeruli while solving the problem of automatic identification and segmentation of glomeruli.

Materials

Data source

In our study, we collected 51 WSIs from open sources, and they are introduced as follows. Thirty-one WSIs generated by the European project AIDPATH (<http://aidpath.eu>) were chosen. The tissue samples were stained using periodic acid–Schiff (PAS) and were scanned at 20× with a Leica Aperio ScanScope CS scanner¹⁷. The remaining 20 WSIs representing various human kidney pathologies came from four sources: three independent medical centres and TCGA. The data from three independent medical centres were collected by¹⁰, including 4 H&E-stained slides from the Military Institute of Medicine in Warsaw in Poland, six PAS-stained slides from Hospital Universitario Vall d'Hebron, Barcelona in Spain and five H&E-stained slides from Cedars-Sinai Medical Center in Los Angeles in the USA. Five H&E-stained slides from the publicly available TCGA repository¹⁸.

All slides were prepared from formalin-fixed paraffin-embedded (FFPE) sections with a thickness of 4 μm . The literature¹¹ specifically describes the method followed for making the slides.

The 31 WSIs from¹⁷ contain two folders. DATASET_A: Raw data with 31 whole slide images (WSIs) in SVS format. We converted these to PNG format. DATASET_B: 2340 images with a single glomerulus, 1170 normal glomeruli and 1170 sclerosed glomeruli. All of them are in PNG format and are detected from DATASET_A. As the repository only provided normal glomerular patches and sclerosed glomerular patches, with the help of professional pathologists, we should find the specific locations of these glomeruli in WSIs, and use QuPath software to label categories and draw their outlines. For 20 WSIs from three independent medical centres and TCGA, the pathologists in¹¹ first identified 78 ROIs and then delineated 1,184 glomeruli within the ROIs. ROIs were extracted for $\times 10$, which corresponded to a pixel size of $\sim 10 \mu\text{m}$. In the availability of materials and data section of this article, we provide the URL where the data can be obtained.

Data processing

We used 25 PAS-stained WSIs from AIDPATH and 20 PAS stained ROIs from three independent medical centres and TCGA as training data for glomerular identification and classification, and the remaining six WSIs from AIDPATH and 58 ROIs were used as different test sets to verify the effectiveness of the glomerular identification algorithm. As shown in Fig. 1, we called the six PAS-stained WSIs from AIDPATH Test1, the five PAS-stained ROIs from Zenodo and the 53 H&E-stained ROIs from Zenodo Test2 and Test3. Because the training data did not include H&E-stained slides, we could also test algorithm migration performances on H&E-stained WSIs. Since the resolution of a single WSI or ROI was very high, it was not easy to train. We performed overlapping cropping on WSIs or ROIs. The size of the cropping was 1024×1024 , and the step length was 512. For the test sets, we also adopted the same strategy as the processing method of the training sets. The number of patches obtained from different datasets is shown in Fig. 1. To reduce the training and testing time, we downsampled all slides two times to reduce the size of the picture before cropping. We converted all glomerular contour labelling into pixelwise mask. Specifically, each WSI corresponded to two masks, with black representing the background and white representing all the normal glomeruli and sclerosed glomeruli, respectively. Figure 2 shows the masks of normal and sclerosed glomeruli on a patch.

Methods

Framework of sclerosed glomerular generation

Cropped sclerosed glomerular masks

To realize the generation of sclerosed glomeruli considering shapes and contextual information, we must create sclerosed glomerular datasets and corresponding masks to train the image inpainting network. Inspired by^{19,20}, we synthesized our datasets using existing data. Specifically, the datasets were created as shown in Fig. 3. It is worth noting that our cropping method is designed to place the sclerosed glomeruli in the centre of the cropped image as much as possible, thus potentially assisting in the subsequent training of the generative network.

First, we need to obtain all sclerosed glomerular masks from the segmentation training set. In Fig. 3, the bottom left portion shows the mask of a certain part of sclerosed glomeruli that is shown in the full section image (Fig. 3 top half). Based on the sclerosed glomerular labels provided by the masks of the open data source, the minimum peripheral circle was made for each sclerosed glomerulus, as shown in the red circles in the bottom left of Fig. 3. The centre of the outer circle was taken as the centre of the cropped rectangle picture, and the size is 256×256 , as shown in the green rectangle box in the bottom left section. The position of the rectangular box is mapped to the position of the original slice, as shown in the bottom right of Fig. 3. The rectangular masks and the corresponding pictures were cropped. The final training data are shown in the Fig. 4.

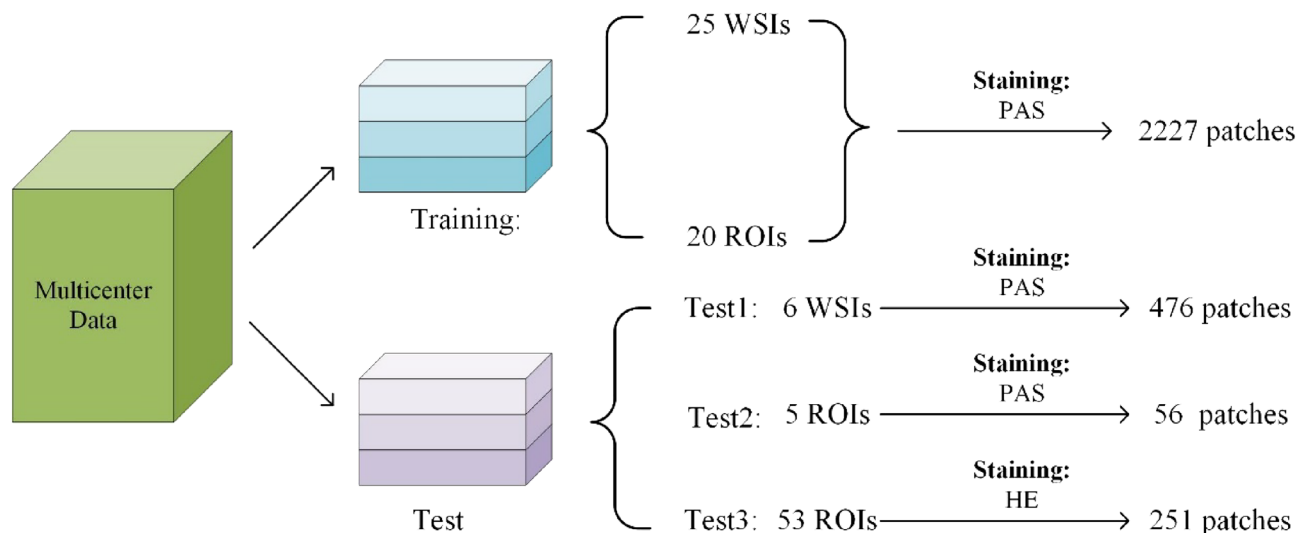


Figure 1. Data processing and settings.

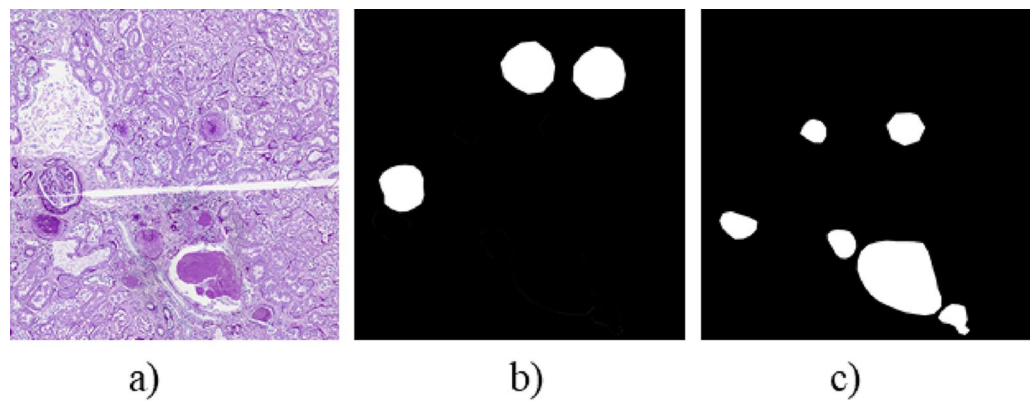


Figure 2. Label making. (a) Original patch, (b) normal glomerular mask, (c) sclerosed glomerular mask.

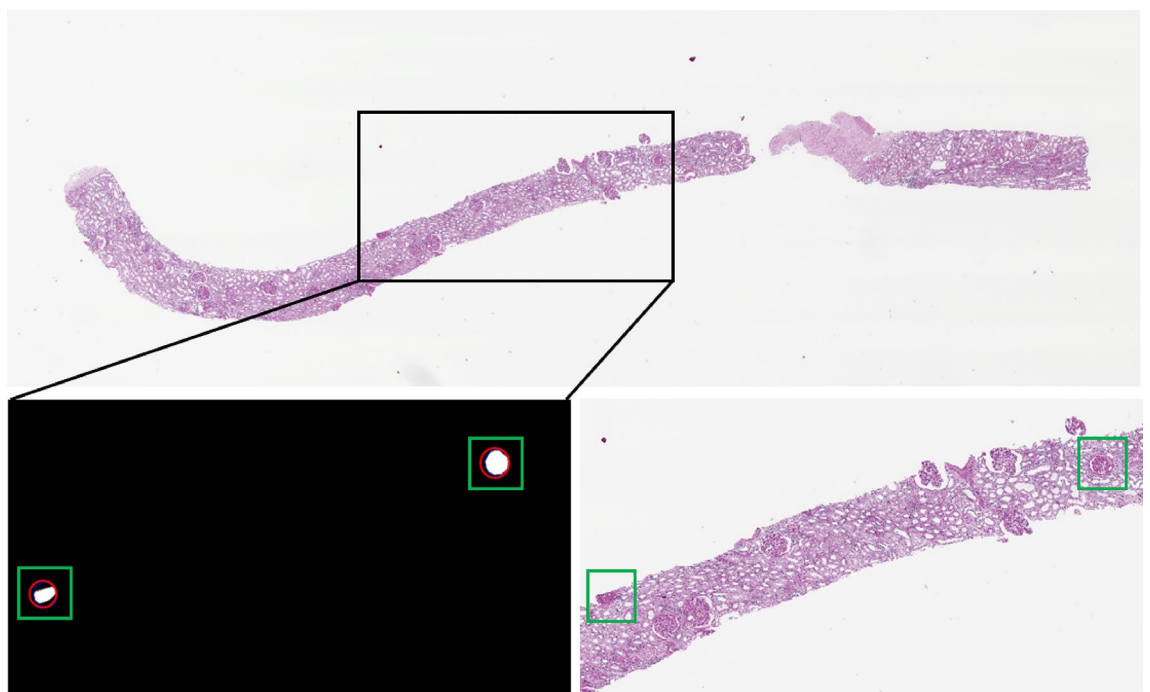


Figure 3. Explanation of the sclerosed glomerular cutting method.

The ROIs are usually synthesized in both the foreground and the background to be fair and unbiased. For example, in^{21,22}, the authors adopted this idea for synthesis. However, when considering the synthesis of ROIs in this article, we did not consider the generation both the glomeruli and the adjacent tissues (background). The reasons are as follows. In the segmentation task, the area of the sclerosed glomerular regions is relatively small compared to the area of its background. Thus, when the deep learning network segments and classifies the glomeruli in the image, the fraction of other parts is much higher than the fraction of glomerular regions. So, the diversity of other tissues can be guaranteed. Based on this, we adopted a generation way like image inpainting to make the generated glomeruli have a good fusion with the existing adjacent tissue and reduce the number of training parameters. Subsequently we can also generate the adjacent tissue and combine it with the generated glomeruli, which may make our model more robust.

Architecture of sclerosed glomerular inpainting network

The training phases of the sclerosed glomerular inpainting network are shown in Fig. 5. It is divided into four modules as follows. (1) The sclerosed glomerular mask input module controls the area of sclerosed glomerular generation. (2) The Generator module is mainly based on AutoEncoder, which consists of an encoder and a decoder. (3) The discriminator module mainly determines whether the input picture is a real picture or a generated picture and, in turn, promotes the training of the generator. (4) The sclerosed glomerular attention loss module includes the global image loss and the loss of the sclerosed glomerular foreground itself.

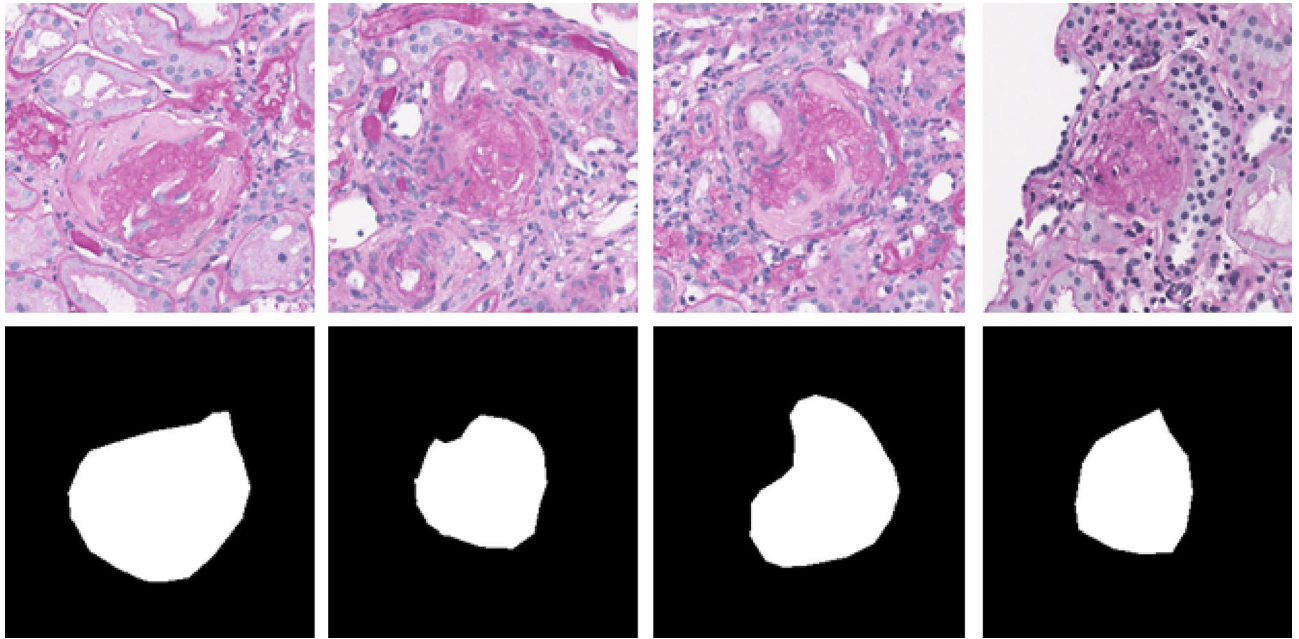


Figure 4. Illustration of cropped sclerosed glomerular images (1st row) and their corresponding images with sclerosed glomerular masks (2nd row).

A. Sclerosed glomerular mask input

A picture with only background X_{gap} is obtained by Eq. (1), where X_{ori} represents a cropped picture of the sclerosed glomerulus with background, X_{mask} represents the corresponding masks, and \odot represents pixelwise multiplication. According to Eq. (2), we can obtain the network input X_{input} by merging the images X_{gap} and the masks X_{mask} in the channel dimension, where $\text{merge}(\bullet)$ is the function realizing dimension concatenation. Through passing X_{input} into the generator, we can realize sclerosed glomerular generation at the vacancy.

$$X_{\text{gap}} = X_{\text{ori}} \odot (1 - X_{\text{mask}}) \quad (1)$$

$$X_{\text{input}} = \text{merge}(X_{\text{gap}}, X_{\text{mask}}) \quad (2)$$

B. Generator

The generator consists of an encoder, a stack of building blocks, a self-attention block and a decoder. In addition, we use skip connections between the encoder and the decoder. The generator takes the 256×256 X_{input} as the input. In the encoder section, the input first passes through a convolutional network of 7×7 convolution kernel size, with batch normalization and a LeakyReLU activation function, followed by two 4×4 convolutional layers with a stride of 2 to downsample the image. This is followed by eight AOT blocks, all with the same parameter settings to reduce the amount of computation required. The AOT block was proposed in²³, and the architecture is shown in Fig. 6a. AOT blocks adopt the split-transformation-merge strategy in three steps²⁴. During the transformation, each subkernel performs a different transformation of the input feature x_1 by using a different dilation rate. Inspired by ResNet, a gated residual connection first calculates the spatially-variant gate value β from $\times 1$ by a standard convolution and a sigmoid operation, and then the AOT block aggregates the input feature $\times 1$ and the learned residual feature $\times 2$ by a weighted sum with β . The network structure of the decoded part and the encoding part are consistent, and two deconvolution layers are used to make the size of the masked picture the same as the size of the input image. Before the first layer of the upsampling network, there is a self-attention block whose input size is 64×64 . It is proposed in²⁵. As shown in Fig. 6b, by obtaining the self-attention feature maps, we can explore the relationship between the locality of the picture and the whole to solve the problem of long-distance dependence. Finally, the tanh function is applied in the output layer.

C. Discriminator

The discriminator in this task was set to two, namely, the local discriminator and the global discriminator. The local discriminator only discriminated the generated sclerosed glomeruli, and the global discriminator discriminated the complete generated images, including the sclerosed glomeruli and the background. When we used the local discriminator, the region of the nonglomerulus can be filled in white so that the size of the local glomeruli image is consistent with the input whole image. In this way, the local discriminator has the same network structure as the global discriminator to reduce the amount of calculation. The input size of each discriminator is 256×256 pixels. There are a total of six convolutional layers, and each convolutional layer uses a 4×4 kernel with a stride of 2 (Convolution + LeakyReLU + Batch normalization) to decrease the size of the

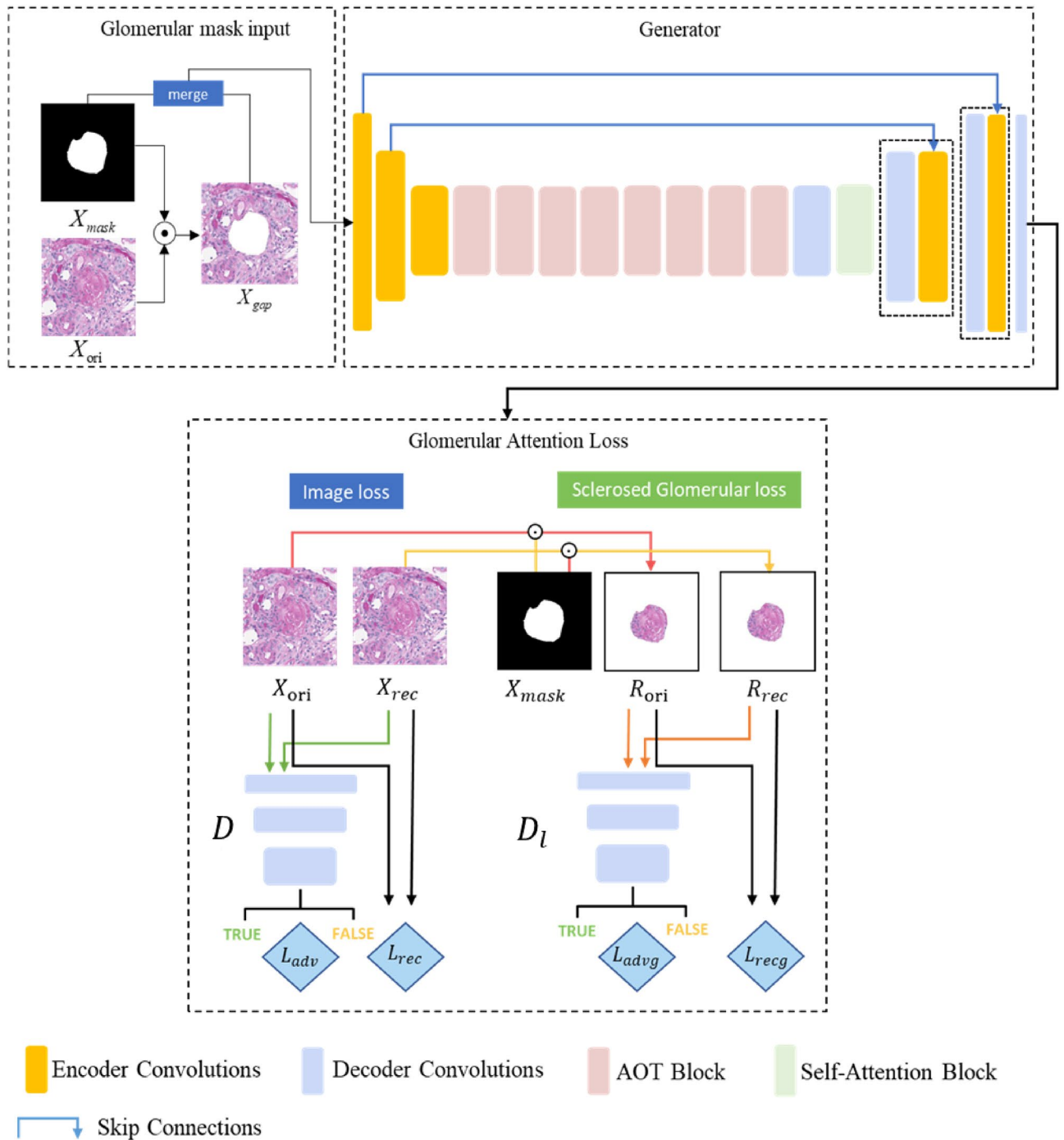


Figure 5. Architecture of sclerosed glomerular inpainting network.

feature representations. The number of channels in the discriminator is set to 64, 128, 256, 512 and 1. The last layer of both discriminators produces $N \times N$ output patches representing classification scores ('real' or 'fake').

D. Sclerosed glomerular attention loss module

As shown in Fig. 5, the image loss and sclerosed glomerular loss are set up to ensure that the whole pictures remain consistent and that the sclerosed glomeruli show a sense of clear texture and staining. Based on the designed loss module with the nature of attention, the generated network achieved a balance between the generation of glomeruli itself and the inpainting of the complete image.

The adversarial loss of the global image and local sclerosed glomeruli are shown in Eqs. (3) and (4), respectively, where D is the global discriminator and D_l is the local discriminator. To reduce the amount of computation required, we set the network parameters of D and D_l to be the same. X_{rec} is the generated global image, which is obtained by the generator as shown in Eq. (5), where G represents the generator. R_{ori} and R_{rec} are the original

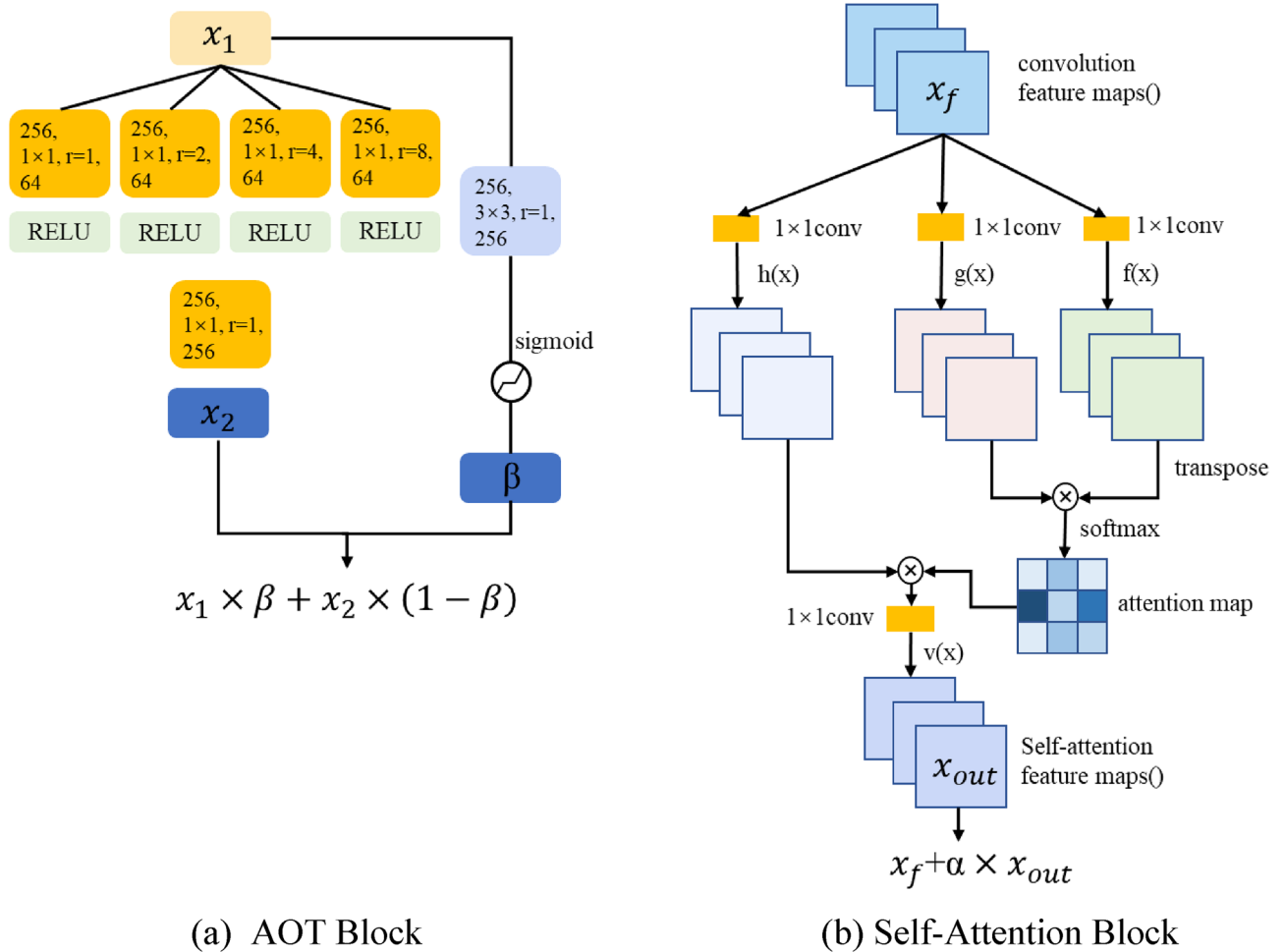


Figure 6. Key modules in the inpainting network. (a) Architecture of the AOT block. (b) Architecture of the self-attention block.

local sclerosed glomeruli and the generated local sclerosed glomeruli, respectively. They are obtained as shown in Eqs. (6) and (7).

$$L_{adv_g} = E[D(X_{rec})^2] + E[(1 - D(X_{ori}))^2] \tag{3}$$

$$L_{adv_l} = E[D_l(R_{rec})^2] + E[(1 - D_l(R_{ori}))^2] \tag{4}$$

$$X_{rec} = G(X_{input}) \tag{5}$$

$$R_{ori} = X_{ori} \odot X_{mask} + (1 - X_{mask}) \tag{6}$$

$$R_{rec} = X_{rec} \odot X_{mask} + (1 - X_{mask}) \tag{7}$$

Usually, in the field of image generation, we use pixel reconstruction loss (L_1) to describe the pixel difference between images. As shown in Eqs. (8) and (9), L_{1g} and L_{1l} represent global L_1 and local L_1 .

$$L_{1g} = \| X_{ori} - X_{rec} \|_1 \tag{8}$$

$$L_{1l} = \| R_{ori} - R_{rec} \|_1 \tag{9}$$

With the good effect of the generative algorithm in the field of image style transformation, the image features extracted by the convolutional network have been widely used as part of the objective function. We use the perceptual loss and style loss of the global image, which are shown in Eqs. (10) and (11), respectively.

$$L_{per} = \sum_{i=1}^{N_i} \frac{\| \phi_i(X_{ori}) - \phi_i(X_{rec}) \|_1}{N_i} \tag{10}$$

where ϕ_i is the activation map from the i -th layers of a pretrained network (e.g., VGG19²⁶) and N_i is the number of elements in ϕ_i . Similarly, the style loss is defined as the L1 distance between the Gram matrices of deep features of inpainting and real images.

$$L_{sty} = \mathbb{E}_i [\| \phi_i(X_{ori})^T \phi_i(X_{ori}) - \phi_i(X_{rec})^T \phi_i(X_{rec}) \|_1] \tag{11}$$

The loss values of each part are added by a certain weight to obtain the final loss function, as shown in Eq. (12).

$$L_{total} = \lambda_{adv}(L_{adv_g} + L_{adv_l}) + \lambda_{1g}L_{1g} + \lambda_{1l}L_{1l} + \lambda_{per}L_{per} + \lambda_{sty}L_{sty} \tag{12}$$

where L_{total} is the total loss and $\lambda_{adv}=0.02$, $\lambda_{1g}=1$, $\lambda_{per}=0.1$, and $\lambda_{sty}=150$.

Process of synthesizing datasets

As shown in Fig. 7, in the stage of sclerosed glomerular synthesis, we used Deep Convolutional Generative Adversarial Network (DCGAN)²⁷ to generate masks of different shapes and sizes based on the existing masks. Since colourful pixel values are likely to appear during mask generation, it is necessary to grayscale the generated masks and set a threshold at the same time to eliminate isolated regions in the masks whose area was smaller than the threshold. The glomerular contours in the masks are scaled so that the number of contours of different sizes are evenly distributed. Based on the pathologist’s recommendation, we locate the potential positions for sclerosed glomeruli and cropped out squares of 256×256 in these positions. Similar to the operation during training, masked images (X_{gap_t}) are obtained by pixelwise multiplication, as shown in Eq. (13) based on randomly selected generated masks (X_{gmask}) and cropped images (X_{ori_t}). X_{gmask} and X_{gap_t} are concatenated as the input of the inpainting model. Finally, the generated images were merged into the original cropped area, and a new ROI with several sclerosed glomeruli in different positions was obtained.

$$X_{gap_t} = X_{ori_t} \odot (1 - X_{gmask}) \tag{13}$$

Glomerular segmentation network

For the design of the glomerular image segmentation network, we generally adopt an encoder-decoder architecture, within which the decoder structure is Unet. The skip connection in Unet is used to fuse multiscale features from the encoder with upsampled features. Here, shallow features and deep features are connected together to reduce the spatial information loss caused by downsampling. In the encoder part, we select EfficientNet as our encoder backbone. The reason why the more advanced transformer structure is not adopted here is that its performance heavily relies on pretraining and requires a large amount of computation. Thus, its training time and computation time will be higher than those of the CNN model under the same parameters. Meanwhile, its prediction time will be longer. We hope that EfficientNet can obtain the result faster while ensuring the effect, which is very important for slice evaluation. EfficientNet was proposed in²⁸ and takes into account both the depth and width of the network. There are currently several versions of EfficientNet including B0-B7. To meet

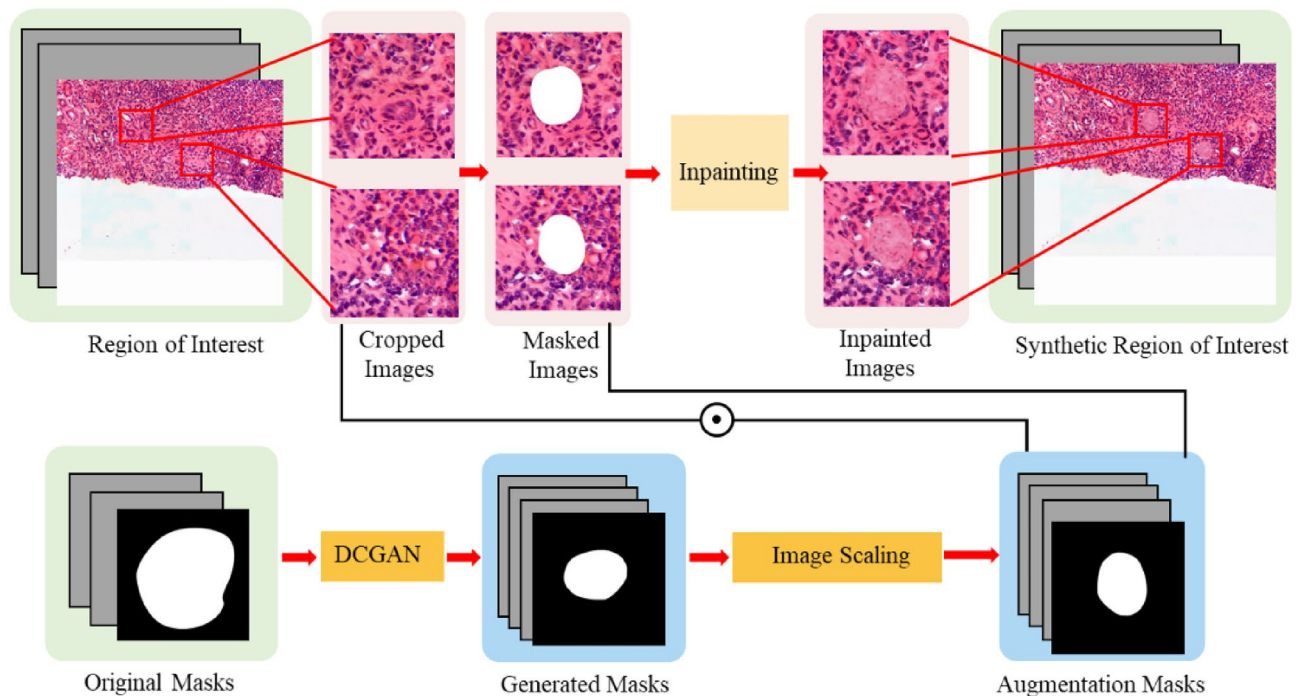


Figure 7. Testing phases of glomerular synthesis.

the speed and accuracy requirements of network training, we use EfficientNetB3 as the backbone of our encoder. The input size of the network is 1024×1024 . Before training the model, the inputs are normalized. The network architecture is shown in Fig. 8.

When training the segmentation model, the batch size was 8, and the optimizer was Adam. We adopted CosineAnnealingLR in the Pytorch framework for the learning strategy and the minimum learning rate was 0.00001. We used the weighted sum of Binary CrossEntropy Loss (L_{BCE}) and Dice Loss (L_{Dice})²⁹ as the total loss (L_{seg}). The specific formula is shown in Eq. (14), where $\lambda = 0.5$. We trained our segmentation models on a single Tesla P100 (16 GB). A total of 200 epochs were trained, and the ratio of training sets and validation sets was 8:2. When evaluating the test set, the model with the lowest Dice coefficient of the validation sets was selected for evaluation.

$$L_{seg} = \lambda L_{BCE} + L_{Dice} \tag{14}$$

Experiments and results

Details and evaluation methods

Based on the images of sclerosed glomeruli and the corresponding masks obtained in the previous chapter, we trained and tested the inpainting network. The ratio of the training set to the test set was 8:2. We trained the models on a single Tesla P100 (16 GB). Here, we used ADAM as the optimizer with an initialized learning rate of 0.0001 and betas of {0.5; 0.999}. We trained our model for 100 epochs with a batch size of 16. The size of the network output was 256×256 .

To characterize the model's glomerular segmentation ability, especially the performance on sclerosed glomeruli, quantitative evaluation is needed. In the segmentation process, for each pixel in the image, there are two categories: positive and negative. If the prediction of positive or negative is correct, it is TP or TN. Conversely, it is FP or FN. Based on these four values, we can also obtain other commonly used metrics, as shown in Table 1.

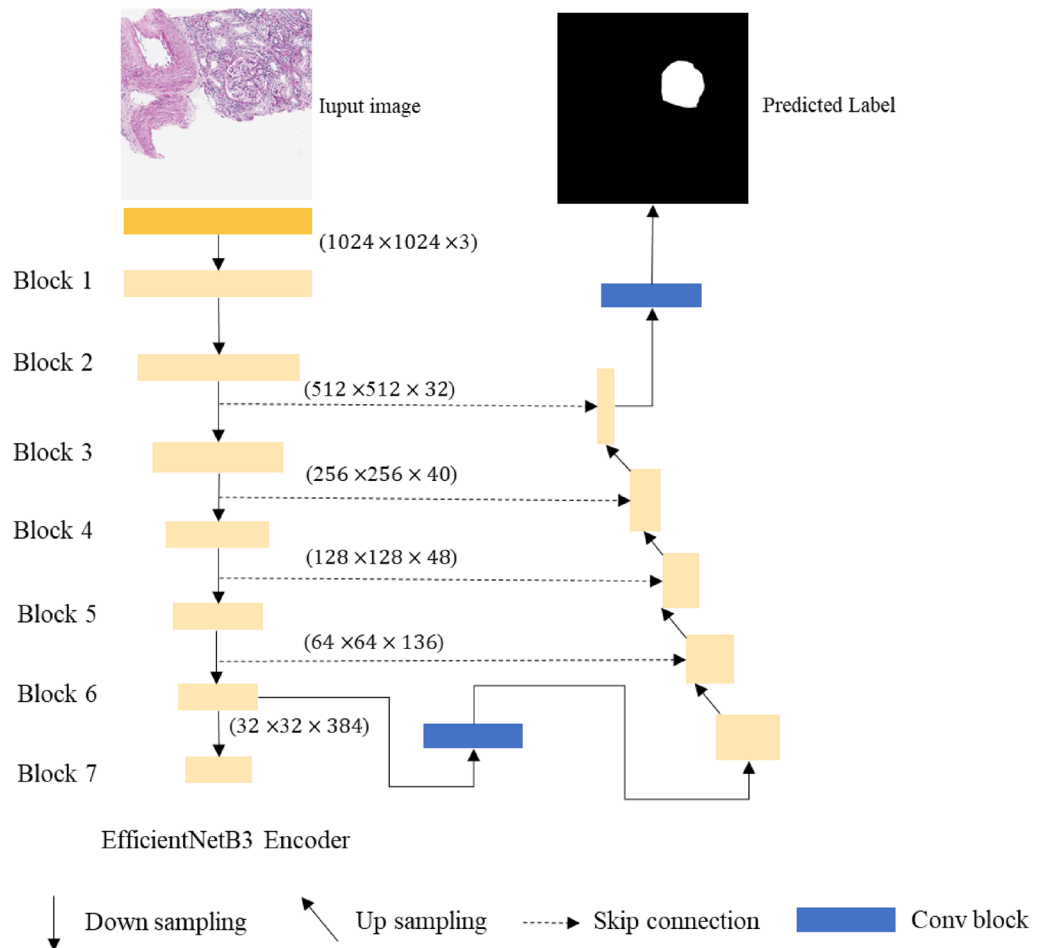


Figure 8. Architecture of the glomerular segmentation network.

Metric	Equation
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
F1-score	$2 \times \frac{Precision \times Recall}{Precision+Recall}$
Dice	$\frac{2 \times TP}{FP+2 \times TP+FN}$

Table 1. Equations of metrics performance.

Generation results and analysis

To evaluate the quality of generated sclerosed glomeruli quantitatively, we used the mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) to compare the differences between generated glomeruli and the original. Table 2 shows the values of the three metrics.

As shown in Fig. 9, we obtained the synthetic images by combining the cropped glomeruli from the generated images with the background of the original images.

Glomerular segmentation and analysis

Before adding generated sclerosed glomeruli to the training data, we verified the effect when using traditional data augmentation, including random flipping and rotating. We evaluated the performance of glomerular segmentation with traditional data augmentation and without traditional data augmentation for three test sets. The data augmentation strategy we adopted is an online method and each input training picture has a certain chance to be flipped or rotated. We performed ten experiments, with each experimental training set and validation set randomly divided. As shown in Table 3, we obtained the performance of glomerular segmentation based on our segmentation model under traditional data augmentation. “√” represents the use of traditional data augmentation.

As seen from the Table 3, for different test sets, applying traditional data augmentation to training data can improve the overall effect of glomerular segmentation to a certain extent, but there may be a decline in precision. The experiments show that the segmentation performance of normal glomeruli is much better than that of sclerosed glomeruli, which is consistent with most of the current studies and validates the necessity of our data generation. At the same time, we see that the performance on test 3 is lower than that of test 1 and test 2, and we can conclude that the migration ability of the algorithm in recognition of renal pathological images with different staining needs to be improved because of the characteristics of different staining methods.

Metric	Value
MAE	0.0152
PSNR	22.6943
SSIM	0.8086

Table 2. Values of inpainting performance.

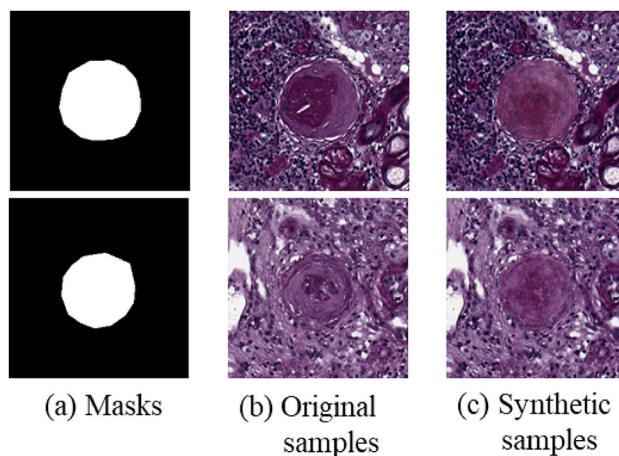


Figure 9. Examples of synthetic sclerosed glomeruli. (a) Sclerosed glomerular masks, (b) original samples, (c) synthetic samples.

	Trad aug	Class	Precision	Recall	F1	Dice
Test1 (PAS)	-	NG	0.8952 ± 0.0018	0.9169 ± 0.0031	0.9059 ± 0.0010	0.9036 ± 0.0005
		SG	0.8492 ± 0.0020	0.4089 ± 0.0025	0.5520 ± 0.0014	0.5145 ± 0.0006
	✓	NG	0.9028 ± 0.0013	0.9228 ± 0.0017	0.9127 ± 0.0006	0.9134 ± 0.0003
		SG	0.8451 ± 0.0021	0.5353 ± 0.0016	0.6534 ± 0.0009	0.6218 ± 0.0004
Test2 (PAS)	-	NG	0.9328 ± 0.0023	0.7727 ± 0.0020	0.8452 ± 0.0009	0.8445 ± 0.0004
		SG	0.7531 ± 0.0019	0.5577 ± 0.0028	0.6408 ± 0.0014	0.6355 ± 0.0007
	✓	NG	0.9202 ± 0.0030	0.7768 ± 0.0021	0.8424 ± 0.0011	0.8406 ± 0.0005
		SG	0.8409 ± 0.0032	0.5820 ± 0.0019	0.6879 ± 0.0012	0.6828 ± 0.0006
Test3 (HE)	-	NG	0.9725 ± 0.0028	0.6770 ± 0.0022	0.7983 ± 0.0011	0.7956 ± 0.0005
		SG	0.7496 ± 0.0029	0.2585 ± 0.0030	0.3844 ± 0.0023	0.3637 ± 0.0008
	✓	NG	0.9779 ± 0.0027	0.6871 ± 0.0021	0.8071 ± 0.0010	0.8052 ± 0.0005
		SG	0.7353 ± 0.0041	0.2334 ± 0.0032	0.3543 ± 0.0027	0.3439 ± 0.0010

Table 3. Performance of glomerular segmentation based on our segmentation model under traditional data augmentation or not, where NG-normal glomeruli and SG-sclerosed glomeruli. The values represent the mean and standard deviation of the ten training times.

Therefore, in order to better evaluate the effect of the sclerosed glomerulus we generated on image segmentation, we analyse the influence of different amounts of synthetic data on the model identification ability for sclerosed glomeruli on the basis of traditional data augmentation, as shown in Table 4.

Table 4 shows that by adding different amounts of synthetic data based on our algorithm, the segmentation performance of scleral glomerulus is greatly improved. This shows that the ability to recognize sclerosed glomeruli is improved, as our generated sclerosed glomeruli have different shapes and sizes distributed in different locations. However, the performance of adding more synthetic data is not always better than that of others. The reason for our analysis is that although the diversity of the generated data shapes is greatly improved, the mechanism features inside the sclerosed glomeruli are still generated based on the existing data, and the distribution of features is still consistent with the original ones.

Additionally, we compared our segmentation model with other classical models to verify the advantages of our method in the task of glomerular segmentation, as shown in Table 5.

Compared with other medical semantic segmentation algorithms including Unet and Unet++³⁰ which are all trained with 100% generated data, we calculated the mean value of each metric of the two classes on the different test sets. We see that our algorithm performs better than other algorithms on different test sets. In addition, on the test set stained by H&E, our algorithm has greater advantages than the others, such as better generalization and migration ability.

Figure 10 shows the visualization of our final model output, with the annotations of the data in the left column and the output of the model in the right column. Blue represents normal glomeruli, and red represents sclerosed glomeruli. We can see that our model can label sclerosed glomeruli that missed the mark in the original data label, especially in the marginal part, which shows the excellent identification ability of our model. However, at the same time, there are smaller sclerosed glomeruli that are missing and need to be improved.

Conclusion

In the task of glomerular identification and classification, it is difficult and costly to obtain large amounts of data for training the model, and there is a problem of class imbalance because the number of sclerosed glomeruli is much larger than that of normal glomeruli in the available data. Therefore, we proposed a sclerosed glomerular

	Synthetic data	Precision	Recall	F1-score	Dice
Test1 (PAS)	+ 30%	0.8813 ± 0.0014	0.6025 ± 0.0038	0.7157 ± 0.0016	0.7010 ± 0.0007
	+ 60%	0.8924 ± 0.0016	0.6426 ± 0.0032	0.7472 ± 0.0013	0.7328 ± 0.0006
	+ 100%	0.8947 ± 0.0021	0.6408 ± 0.0029	0.7468 ± 0.0012	0.7491 ± 0.0006
Test2 (PAS)	+ 30%	0.8825 ± 0.0023	0.6023 ± 0.0012	0.7160 ± 0.0008	0.7313 ± 0.0004
	+ 60%	0.9022 ± 0.0019	0.6219 ± 0.0027	0.7363 ± 0.0012	0.7641 ± 0.0006
	+ 100%	0.9068 ± 0.0025	0.6211 ± 0.0031	0.7372 ± 0.0014	0.7685 ± 0.0007
Test3 (HE)	+ 30%	0.7729 ± 0.0032	0.4017 ± 0.0042	0.5286 ± 0.0025	0.4278 ± 0.0009
	+ 60%	0.7761 ± 0.0029	0.4165 ± 0.0037	0.5421 ± 0.0022	0.4871 ± 0.0009
	+ 100%	0.7801 ± 0.0042	0.4303 ± 0.0025	0.5547 ± 0.0018	0.4916 ± 0.0007

Table 4. Performance comparison of sclerosed glomerular segmentation based on our segmentation model when adding different amounts of synthetic data. The values represent the mean and standard deviation of the ten training times.

	Model	Mean Precision	Mean Recall	Mean F1-score	Mean Dice
Test1 (PAS)	Unet	0.7223 ± 0.0012	0.6630 ± 0.0010	0.6914 ± 0.0006	0.7690 ± 0.0003
	Unet++	0.7816 ± 0.0014	0.6719 ± 0.0011	0.7226 ± 0.0006	0.7828 ± 0.0003
	Swin-Unet	0.8545 ± 0.0010	0.7730 ± 0.0014	0.8117 ± 0.0005	0.8025 ± 0.0003
	EfficientNetB3-Unet	0.9002 ± 0.0012	0.7918 ± 0.0009	0.8425 ± 0.0004	0.8325 ± 0.0002
Test2 (PAS)	Unet	0.8628 ± 0.0027	0.6423 ± 0.0012	0.7364 ± 0.0009	0.7310 ± 0.0004
	Unet++	0.8901 ± 0.0022	0.6310 ± 0.0027	0.7385 ± 0.0012	0.7524 ± 0.0006
	Swin-Unet	0.8804 ± 0.0008	0.6621 ± 0.0015	0.7556 ± 0.0006	0.7730 ± 0.0003
	EfficientNetB3-Unet	0.9137 ± 0.0020	0.6713 ± 0.0024	0.7740 ± 0.0010	0.8045 ± 0.0005
Test3 (HE)	Unet	0.7909 ± 0.0011	0.4871 ± 0.0020	0.6029 ± 0.0010	0.5521 ± 0.0004
	Unet++	0.8054 ± 0.0020	0.4993 ± 0.0017	0.6164 ± 0.0010	0.5564 ± 0.0004
	Swin-Unet	0.8621 ± 0.0029	0.5335 ± 0.0009	0.6591 ± 0.0009	0.6102 ± 0.0004
	EfficientNetB3-Unet	0.8790 ± 0.0018	0.5626 ± 0.0012	0.6861 ± 0.0007	0.6882 ± 0.0003

Table 5. Performance comparison of glomerular segmentation based on three modes, including our proposed method. The values represent the mean and standard deviation of the ten training times.

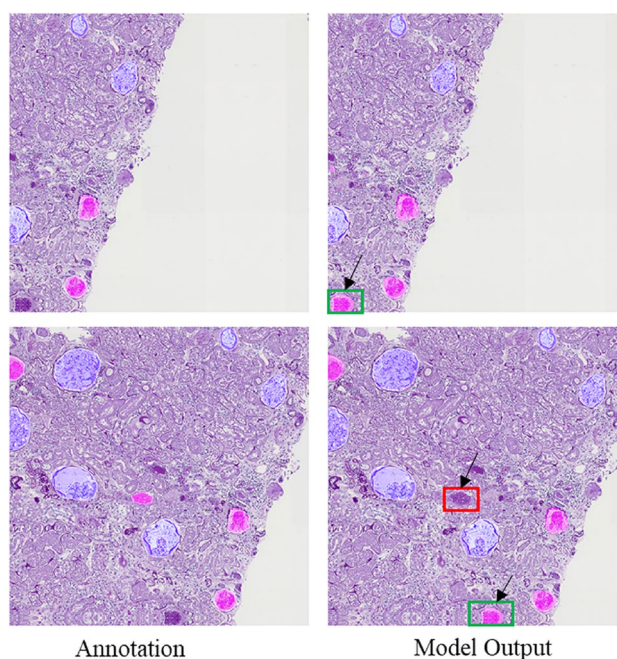


Figure 10. Segmentation results obtained by our model.

generation method based on image inpainting. With the existing masks, we generated diverse masks and scaled them to obtain more small sclerosed glomeruli. By using the proposed image inpainting-based method and generated masks, we synthesized multiple images of sclerosed glomeruli with good fusion with the background and ensured that the texture of sclerosed glomeruli was clear and true. We combined the synthesized data with the original data and passed them into the segmentation network. The glomerular segmentation network was based on Unet where we used EfficientNetB3 as the backbone of the encoder. When we incorporated synthetic sclerosed glomeruli, we achieved better sclerosed glomerular identification under traditional data augmentation. Compared with other segmentation models, our model achieved the best mean F1 and Dice coefficients containing 2 classes by using EfficientNetB3-Unet.

Since our identification algorithm is mainly targeted at the training of globally sclerosed and normal glomeruli, the identification performance of other classes using our algorithm needs to be improved. Our generated glomeruli are also globally sclerosed, and we can further explore the controlled generation of different degrees of sclerosed glomeruli. In this way, we can reduce the number of missed glomerular tests.

Data availability

The 31 WSIs generated in the AIDPATH are hosted by Mendeley at: <https://data.mendeley.com/datasets/k7nvtgn2x6/3> and 78 ROIs from 21 WSIs are hosted by Mendeley at: <https://zenodo.org/record/4299694>.

Received: 23 July 2023; Accepted: 8 January 2024

Published online: 10 January 2024

References

1. Tullius, S. G. & Rabb, H. Improving the supply and quality of deceased-donor organs for transplantation. *N. Engl. J. Med.* **378**(20), 1920–1929 (2018).
2. Moeckli, B. *et al.* Evaluation of donor kidneys prior to transplantation: An update of current and emerging methods. *Transpl. Int.* **32**(5), 459–469 (2019).
3. Stewart, D. E. & Klassen, D. K. Early experience with the new kidney allocation system: A perspective from UNOS. *Clin. J. Am. Soc. Nephrol. CJASN* **12**(12), 2063 (2017).
4. Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, 234–241. (Springer, 2015).
5. Goodfellow, I. J. *et al.* Generative adversarial networks. *Proc. Adv. Neural Inf. Process. Syst.* **3**, 2672–2680 (2014).
6. Liu, K., Shuai, R. & Ma, L. Cells image generation method based on VAE-SGAN. *Proc. Comput. Sci.* **183**, 589–595 (2021).
7. Guan, Q. *et al.* Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN. *Comput. Biol. Med.* **145**, 105444 (2022).
8. Lei, H. *et al.* LAC-GAN: Lesion attention conditional GAN for ultra-widefield image synthesis. *Neural Netw.* **158**, 89–98 (2023).
9. Amirrajab, S. *et al.* Label-informed cardiac magnetic resonance image synthesis through conditional generative adversarial networks. *Comput. Med. Imaging Graph.* **101**, 102123 (2022).
10. Marsh, J. N. *et al.* Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Trans. Med. Imaging* **37**(12), 2718–2728 (2018).
11. Gallego, J. *et al.* A U-Net based framework to quantify glomerulosclerosis in digitized PAS and H&E stained human tissues. *Comput. Med. Imaging Graph.* **89**, 101865 (2021).
12. Bueno, G., Fernandez-Carrobles, M. M., Gonzalez-Lopez, L. & Deniz, O. Glomerulosclerosis identification in whole slide images using semantic segmentation. *Comput. Methods Programs Biomed.* **184**, 105273 (2020).
13. Jiang, L., Chen, W., Dong, B., Mei, K., Zhu, C., Liu, J., & Shi, H. A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy (2021).
14. Yao, T. *et al.* Glo-In-One: Holistic glomerular detection, segmentation, and lesion characterization with large-scale web image mining. *J. Med. Imaging* **9**(5), 052408–052408 (2022).
15. Kawazoe, Y. *et al.* Computational pipeline for glomerular segmentation and association of the quantified regions with prognosis of kidney function in IgA nephropathy. *Diagnostics* **12**(12), 2955 (2022).
16. Silva, J. *et al.* Boundary-aware glomerulus segmentation: toward one-to-many stain generalization. *Comput. Med. Imaging Graph.* **100**, 102104 (2022).
17. Bueno, G., Gonzalez-Lopez, L., Garcia-Rojo, M., Laurinavicius, A. & Deniz, O. Data for glomeruli characterization in histopathological images. *Data Brief* **29**, 105314 (2020).
18. Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**(10), 1113–1120 (2013).
19. Shin, Y., Qadir, H. A. & Balasingham, I. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access* **6**, 56007–56017 (2018).
20. Qadir, H. A., Balasingham, I. & Shin, Y. Simple U-net based synthetic polyp image generation: Polyp to negative and negative to polyp. *Biomed. Signal Process. Control* **74**, 103491 (2022).
21. Efros, A. A., & Freeman, W. T. Image quilting for texture synthesis and transfer. In *Seminal Graphics Papers: Pushing the Boundaries*. Vol. 2. 571–576 (2023).
22. Glotsos, D., Kostopoulos, S., Ravazoula, P. & Cavouras, D. Image quilting and wavelet fusion for creation of synthetic microscopy nuclei images. *Comput. Methods Programs Biomed.* **162**, 177–186 (2018).
23. Zeng, Y., Fu, J., Chao, H. & Guo, B. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Trans. Vis. Comput. Graph.* **29**, 3266 (2022).
24. Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1492–1500 (2017).
25. Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. Self-attention generative adversarial networks. In *International Conference on Machine Learning*. 7354–7363 (PMLR, 2019).
26. Sengupta, A., Ye, Y., Wang, R., Liu, C. & Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* **13**, 95 (2019).
27. Radford, A., Metz, L., & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv Preprint arXiv:1511.06434* (2015).
28. Tan, M., & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*. 6105–6114. (PMLR, 2019).
29. Milletari, F., Navab, N., & Ahmadi, S. A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. 565–571 (IEEE, 2016).
30. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMI 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*. Vol. 4. 3–11. (Springer, 2018).

Acknowledgements

This research was supported by the National Natural Science Foundation of China in 2021 (Num. 72104085).

Author contributions

H.S. and Y.S. wrote the main manuscript text. Y.Z. prepared all figures .B.L. , F.Y. ,X.L. , H.G. ,X.T. and Y.C. provided guidance on the article. All authors reviewed the manuscript.

Competing interests

Authors declare no conflict of interest of the manuscript entitled: ‘An Image Inpainting-based Data Augmentation Method for Improved Sclerosed Glomerular Identification Performance with the Segmentation Model- EfficientNetB3-Unet’.

Additional information

Correspondence and requests for materials should be addressed to Y.C.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2024