Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 6309596, 13 pages
https://doi.org/10.1155/2020/6309596

Research Article

Study on the Method of Fundus Image Generation Based on Improved GAN

Jifeng Guo,[1] Zhiqi Pang,[1] Fan Yang,[1] Jiayou Shen,[1] and Jian Zhang[2]

[1]*College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China*
[2]*School of Artificial Intelligence, Wuxi Vocational College of Science and Technology, Wuxi 214000, China*

Correspondence should be addressed to Jian Zhang; zhangjianok00@163.com

Received 16 May 2020; Accepted 12 June 2020; Published 8 July 2020

Academic Editor: Oscar Reinoso

With the continuous development of deep learning, the performance of the intelligent diagnosis system for ocular fundus diseases has been significantly improved, but during the system training process, problems like lack of fundus samples and uneven sample distribution (the number of disease samples is much smaller than the number of normal samples) have become increasingly prominent. In view of the previous issues, this paper proposes a method for generating fundus images based on "Combined GAN" (Com-GAN), which can generate both normal fundus images and fundus images with hard exudates, so that the sample distribution can be more even, while the fundus data are expanded. First, this paper uses existing images to train a Com-GAN, which consists of two subnetworks: im-WGAN and im-CGAN; then, it uses the trained model to generate fundus images, then performs qualitative and quantitative evaluation on the generated images, and adds the images to the original image set to expand the datasets; finally, based on this expanded training set, it trains the hard exudate detection system. The expanded datasets effectively improve the generalization ability of the system on the public datasets DIARETDB1 and e-ophtha EX, thereby verifying the effectiveness of the proposed method.

1. Introduction

With the continuous development of deep learning, it has been widely applied in the medical field, and the performance of corresponding medical intelligent diagnosis system has been significantly improved, but there are also many problems. For the hard exudate detection system, a large number of marked images are needed in the training process of the system, but in reality, it is difficult to obtain fundus images (obtaining fundus images requires professional medical cameras to take pictures of human eyes) and the distribution of sample data is uneven (the number of sick samples is much smaller than the number of normal samples). For the problem of uneven sample distribution [1, 2], the data-level solution strategies can be roughly divided into three types. The first type is data enhancement, which includes traditional data enhancement methods, such as flipping, scaling, cropping, and adding noise. There are also advanced data enhancement methods [3–5] such as Sample Pairing [3], which uses two images to synthesize a new sample. This type of method can indeed alleviate the problem of insufficient positive samples, but it is limited in terms of scalability and relies too much on existing datasets.

The second type is oversampling [6–8] and undersampling [9–11]. Oversampling is to expand the minority class samples (called positive samples) so as to increase the percentage to a normal value. Examples include random oversampling [6], SMOTE method [7], and integrated oversampling [8]. These methods can improve sensitivity, but due to insufficient diversity of the positive samples, it can easily cause overfitting [12], so it is usually used in combination with data enhancement. Undersampling is to discard the majority class samples (called negative samples). For example, Ng et al. [9] clustered negative samples to obtain their distribution information, so as to select representative samples and discard others. This type of method has great drawbacks: not only does it lose some of the negative sample features, but it also often leads to an

insufficient number of samples for training. For different samples, how to find the effective sampling strategy is also a challenging problem.

The third type is artificial data synthesis, such as VAE [13], which can generate low-resolution images. The Generative Adversarial Network (GAN) [14], since it was proposed by Ian Goodfellow, has shown strong generation capabilities in the field of image generation [15–17] and has been widely used to augment datasets. In theory, GAN can explore the distribution rules of data based on the existing data and then generate samples with the same distribution as the original data. The method proposed in this paper falls within the third type of methods.

Fundus image acquisition is expensive and involves patient privacy, making it a difficult subject for public research. In order to introduce private medical data into the public domain and alleviate the problems like insufficient fundus image data and uneven sample distribution, many researchers have applied GAN to expand the fundus image datasets, and there have been many successful cases, but there are also problems such as loss of details, mode collapse [18, 19], and unstable training [20, 21]. These cases can be divided into two categories. One is based on unsupervised learning GAN and its improved models [22–25]. In theory, it can generate rich picture data, but the actual training process is still very difficult, with serious mode collapse. Images can only be generated randomly and poorly controlled. Guibas et al. proposed a fundus angiography image generation method [26], which can effectively improve the quality and diversity of image generation but still suffers from loss of details and cannot generate images with corresponding labels. The other category is to modify the unsupervised learning GAN to CGAN [27] and pix2pix [28]. The most representative one is the method of generating fundus angiography images with diseased tissues proposed by Appan et al. [29]. This method significantly improves the quality of image generation but relies too much on existing datasets, so it is difficult to generate rich fundus images using this method.

In order to improve the shortcomings of the previous methods, this paper proposes a fundus images generation method based on Com-GAN: firstly, im-WGAN is used to generate a vascular tree, and then im-CGAN is used to generate a complete image. Experiments show that the model integrates the advantages of the two methods and performs better than either of the methods alone. Compared with unsupervised GAN, the proposed approach is more controllable and the quality of generated images is significantly improved; and compared with the supervised CGAN model, it takes two steps to improve the diversity of samples and generate richer images. The generated image is then added to the training set of the hard exudate detection [30, 31] model. Compared with those generated by other methods, the image generated by the proposed method can greatly improve the generalization ability of the model and effectively alleviate the problems of insufficient samples and uneven distribution.

The main contributions of this paper are as follows:

(1) A two-step method for generating fundus images based on Com-GAN is proposed, which incorporates the advantages of two adversarial networks. Unlike direct generation, this method first uses im-WGAN to generate a vascular tree [32], which can reduce the difficulty of fundus images generation, ensure the quality, and increase the diversity. (2) It improves the original CGAN network by introducing two generating conditions in the generator and the discriminator. The improved network can not only generate high-quality fundus images but also control the categories of the images generated. (3) It introduces pixel-wise mean squared error (pMSE) [33] and perception loss [34] based on the original loss function, so as to retain the characteristics of original images and improve the visual satisfaction about the generated images.

2. Related Studies

2.1. GAN. GAN is composed of two parts: generator G and discriminator D. The generator takes random noise z as input and is used to learn the distribution of training data x; and the discriminator is similar to a classifier, which is used to discriminate real data x and $G(z)$. The two networks are trained alternately. When the discriminator cannot correctly classify the sample sources, the generator and the discriminator will reach Nash equilibrium [35]. The objective function of GAN is

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))], \tag{1}$$

where $p_{\text{data}}(x)$ is the probability distribution of the real data, $p_z(z)$ is the probability distribution of random noise, and E is the mathematical expectation.

2.2. WGAN. An important reason for the difficulty in GAN training is that, due to gradient disappearance [36], that is, under the condition that the discriminator approximates optimality, when there is no nonnegligible coincidence between the generated data and the real data, optimizing the objective function is equivalent to optimizing the Jensen-Shannon [37] divergence between the generated data and the real data. At this time, the Jensen-Shannon divergence is approximately a constant and thus can no longer guide the training process. In WGAN [23], the Wasserstein distance was used instead of the original loss function to solve gradient disappearance. The objective function of WGAN is

$$\min_G \max_{f,\|f\|_L\le 1} E_{x\sim p(x)}\Big[f(x)\Big] - E_{z\sim q(z)}\Big[f(G(z))\Big], \tag{2}$$

where $f(x)$ is a discriminator function, which needs to satisfy Lipschitz constraints [38].

2.3. CGAN. Due to the unstable training process and poor controllability of the original GAN, CGAN came into being. CGAN adds a condition variable y to the input of the generator and the discriminator to guide the generation process. The objective function of CGAN is

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x \mid y)] + E_{z\sim p_z(z)}[\log(1 - D(G(z \mid y)))], \tag{3}$$

where $p_{\text{data}}(x)$, $p_z(z)$, and E have the same meaning as formula (1) and y represents the introduced condition variable, which can be of any form.

3. Method

In this section, the first part introduces the overall framework of Com-GAN, and the latter two parts introduce the two building blocks.

3.1. Overall Structural Framework of the Model. The fundus angiography image generation based on Com-GAN proposed in this paper consists of two networks: the im-WGAN network for generating a vascular tree and the im-CGAN network for generating a complete fundus angiography image. Both networks are improved to better adapt to fundus image generation, based on the original network. The overall framework is shown in Figure 1. The fundus image is generated in two steps, and each step of generation improves the sample diversity.

The training process can be divided into two stages, specifically described as follows:

In the first stage, an image segmentation technique [39] is used to segment a vascular tree from the existing fundus image set, and an im-WGAN is trained based on the segmented vascular tree. After the model converges, a large number of vascular trees are generated using the trained im-WGAN generator, thereby expanding the vascular tree image set.

In the second stage, based on the vascular tree segmented from the real image and the corresponding complete fundus image, a vascular tree-complete fundus image pair is formed to train the im-CGAN proposed in this paper. The network is improved based on CGAN, and the generator and the discriminator are alternately trained, until the model converges. Based on the expanded vascular tree image set, the trained im-CGAN generator is then used to generate a complete fundus image pair, which includes a normal fundus image and a fundus image containing hard exudates. The generated fundus images are added to the existing fundus image set to further expand the fundus datasets.

3.2. Im-WGAN. Compared with the original GAN, WGAN has better training stability and is suitable for the generation "from nothing" studied in this paper. The purpose of this network is to generate a vascular tree image with perfect details. However, due to the complexity of the vascular tree structure, a conventional processing method will inevitably lead to too many parameters in the network structure, which will increase the amount of calculation and also heighten the overfitting risk. Considering the previous problems, this

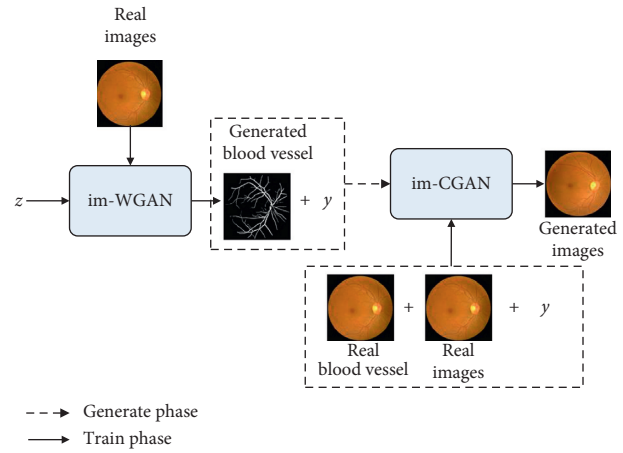


Figure 1: Com-GAN's overall structural framework, consisting of two main structures: im-WGAN and im-CGAN.

paper improves the model structure on the basis of WGAN. Im-WGAN includes two generators and two discriminators, with the overall structure shown in Figure 2.

The generator G_1 takes random noise z as the input and outputs a generated low-resolution vascular tree, and the discriminator D_1 takes a real low-resolution vascular tree or one that is generated by G_1 as the input and determines the probability of a real vascular tree. The generator G_2 takes the low-resolution vascular tree generated by G_1 as the input and outputs a reconstructed high-resolution vascular tree, and the discriminator D_2 takes a real high-resolution vascular tree or one that is generated by G_2 as the input and determines the probability of a real vascular. The size of a low-resolution image is 128×128 pixels, and that of a high-resolution image is 256×256 pixels.

The overall generation process can be divided into two stages: the first stage relies on the generator G_1, describes the basic outline of the image, and generates a low-resolution image with a simple vascular structure; the second stage fills in the details of the low-resolution image and generates a more realistic high-resolution image. Experiments show that the two-stage generation approach can enhance the stability of the training process and improve the quality and diversity of the generated images.

The generator G_1 is an improved version of the DCGAN [22] generator structure. The improvements made include increasing the number of deconvolution layers and changing the final output channel number to 1. The structure of generator G_2 is based on U-net [40]. The network structure of U-net includes downsampling encoders and upsampling decoders. Downsampling encoders are used to extract image features, and upsampling decoders combine the information of each layer of downsampling encoders and the input information of upsampling to restore detailed information and gradually restore the image accuracy. Therefore, the generator G_2 includes downsampling encoders, residual blocks [41], and upsampling decoders, and the BN layer is added after the convolutional layer [42], where the residual blocks are used to increase the network depth. The specific structure of G_2 is shown in Table 1.

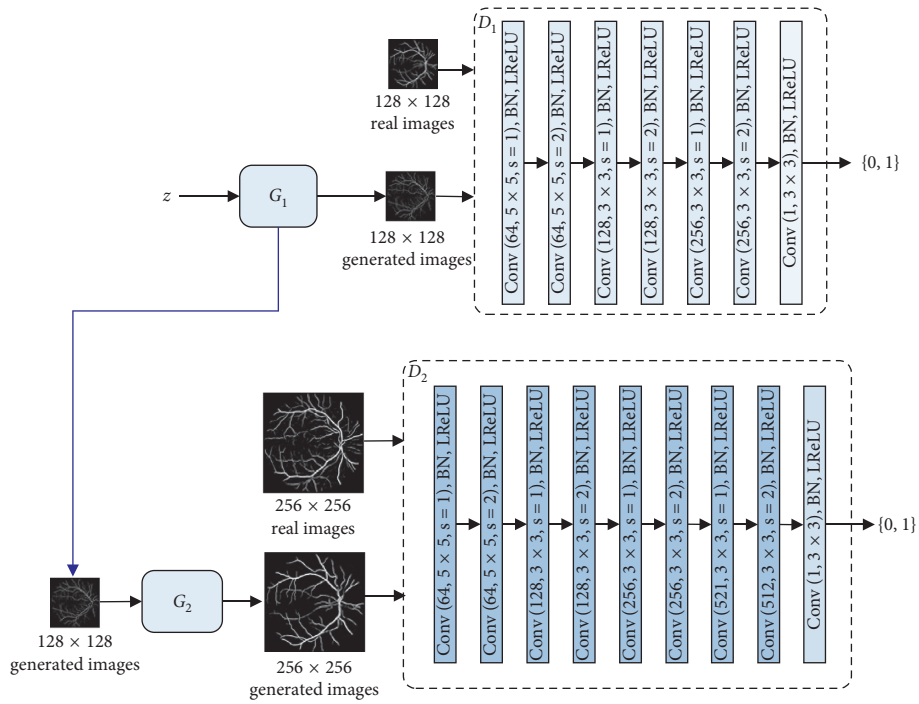


Figure 2: Overall structure of im-GAN, including two generators and two discriminators.

Table 1: Network structure of the generator G_2.

	Layer	Input	Output
Downsample	Conv (64, 4 × 4), BN, LReLU	(128, 128, 1)	(128, 128, 64)
	Conv (128, 4 × 4), BN, LReLU	(128, 128, 64)	(64, 64, 128)
	Conv (256, 4 × 4), BN, LReLU	(64, 64, 128)	(32, 32, 256)
Residual block	Conv (256, 4 × 4), BN, LReLU	(32, 32, 256)	(32, 32, 256)
	Conv (256, 4 × 4), BN, LReLU	(32, 32, 256)	(32, 32, 256)
	Conv (256, 4 × 4), BN, LReLU	(32, 32, 256)	(32, 32, 256)
	Conv (256, 4 × 4), BN, LReLU	(32, 32, 256)	(32, 32, 256)
Upsample	Deconv (128, 4 × 4), BN, LReLU	(32, 32, 256)	(64, 64, 128)
	Deconv (64, 4 × 4), BN, LReLU	(64, 64, 128)	(128, 128, 64)
	Deconv (32, 4 × 4), BN, LReLU	(128, 128, 64)	(256, 256, 32)
	Conv (1, 4 × 4), BN, LReLU	(256, 256, 32)	(256, 256, 1)

This paper uses the physical significance of the matrix spectral norm [43] to make the discriminator of the im-WGAN satisfy the Lipschitz constraint in the global scope. Here, the physical significance of the matrix spectral norm means that any vector, after undergoing matrix transformation, will have a length that is less than or equal to the length of the product of this vector and the matrix spectral norm. The formula is as follows:

$$\frac{\|f(x+\delta)-f(x)\|_2}{\|\delta\|_2}=\frac{\|W\delta\|_2}{\|\delta\|_2}\le\sigma(W), \tag{4}$$

where $\sigma(W)$ represents the spectral norm of the weight matrix, x represents the input vector of the layer, and δ represents the amount of change in x.

3.3. Im-CGAN. The purpose of this network is to generate two types of complete fundus images: normal fundus images and fundus images with hard exudates.

First, a category label y is established for each real image to mark whether it contains hard exudates. Then, based on the vascular tree segmented from the real image and the corresponding complete fundus image, a vascular tree-complete fundus image pair is formed.

An advantage about CGAN is that it can use labels to control the generation, making it quite suitable for the generation process in this paper. Therefore, this paper improves the model structure based on CGAN. The overall structure of im-CGAN is shown in Figure 3. During the training process, the generator G takes the segmented

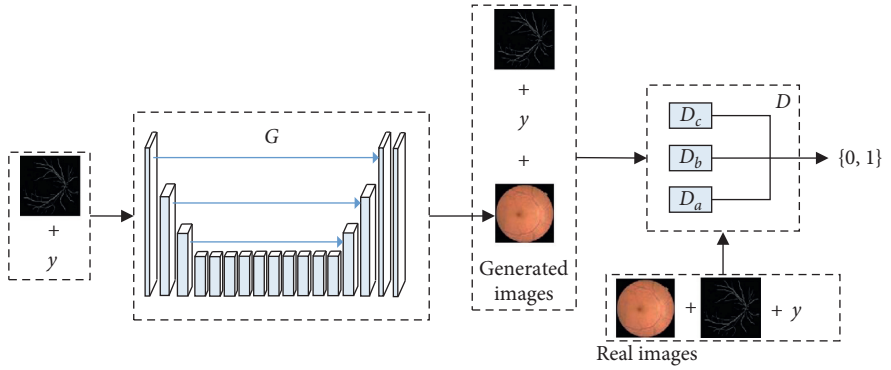


Figure 3: Overall structure of im-CGAN, including a generator and a multiscale discriminator.

vascular tree and label y as the input. The main purpose of G is to generate a complete fundus image. D takes the vascular tree and label y and the generated image or the corresponding real image as the input. Its main purpose is to guide the generation process. During the generation process, the segmented vascular tree or the one generated by im-WGAN and the category label of the image to be generated are input into the generator to obtain a complete fundus image.

The generator uses an encoder-decoder structure and introduces a U-net skip-level structure, which specifically includes 4 convolutional layers, 9 residual blocks, and 3 deconvolution layers.

In order to generate a fundus image with high resolution, a discriminator with a large receptive field is needed. For a conventional processing method, the network capacity needs to be increased. This not only consumes too much memory but also easily causes network overfitting. Therefore, a multiscale discriminator is introduced based on dilated convolution to expand the receptive field of the discriminator under the same parameters. An example of expanded convolution is shown in Figure 4.

Figure 4 shows the sizes of the receptive field when the 3×3 convolution kernel takes different expansion rates, where "*" represents the parameter point and the shaded part represents the receptive field. Figure 4(a) is a normal convolution with a corresponding expansion rate of 1; Figure 4(b) corresponds to an expansion rate of 2; and Figure 4(c) corresponds to an expansion rate of 3. As can be seen, the receptive field expands as the expansion rate increases.

In the improved discriminator model proposed in this paper, three discriminators are set, with three scales from coarse to fine. Among them, the coarsest discriminator D_a corresponds to an expansion convolution with a cyclic expansion rate of {1, 2, 7} and has the largest receptive field, and thus it is responsible for the global judgment of the fundus image. The medium-scale discriminator D_b corresponds to an expansion convolution with a cyclic expansion rate of {1, 2, 5}, which is responsible for guiding the generator to generate a smooth image; and the fine-scale discriminator D_c corresponds to a cyclic expansion rate of {1, 2, 3}. Having the smallest receptive field and being more sensitive to details, it is responsible for guiding the generator to learn more realistic details. The multiscale structure is shown in Figure 5.

In order to ensure the generation quality, retain the original image features, and improve visual satisfaction, this paper introduces pixelwise mean squared error (pMSE) and perception loss on the basis of the original loss function. pMSE is defined as

$$L_{\text{pMSE}} = \frac{1}{\text{WH}} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(I_{x,y} - G_\theta\left(I_{x,y'}\right)\right)^2, \tag{5}$$

where $I_{x,y}$ and $I_{x,y'}$, respectively, represent the pixel values of the (x, y) pixels in the complete fundus image and the vascular tree; W and H represent the height and width of the image, respectively, both of which are 256 in this paper, and θ is the generator parameter.

Because pMSE calculates the loss pixel by pixel, it will inevitably lead to too smooth texture and poor visual perception. Therefore, this paper introduces perception loss to improve visual satisfaction. Visual perception loss is defined as follows:

$$L_{\text{pl}} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I) - \phi_{i,j}(G_\theta(I'))\right)^2, \tag{6}$$

where $\phi_{i,j}$ represents the feature map before the i-th largest pooling layer and after the j-th convolutional layer in the pretrained VGG19 network [44]; I and I' represent the complete fundus image and the vessel tree, respectively; and $W_{i,j}$ and $H_{i,j}$ represent the dimensions of each feature map in the VGG network.

The overall cost function is

$$L_{\text{total}} = L_{\text{CGAN}} + \alpha L_{\text{pMSE}} + \beta L_{\text{pl}}, \tag{7}$$

where L_{CGAN} is the adversarial loss function of CGAN, L_{pMSE} is the pixelwise mean squared error, L_{pl} is the perceptual loss, and α and β are the hyperparameters for controlling the proportion, both of which are set to 0.1 in this paper.

4. Experiments and Analysis

This paper evaluates the effectiveness of the Com-GAN by comparing different generation methods in terms of image

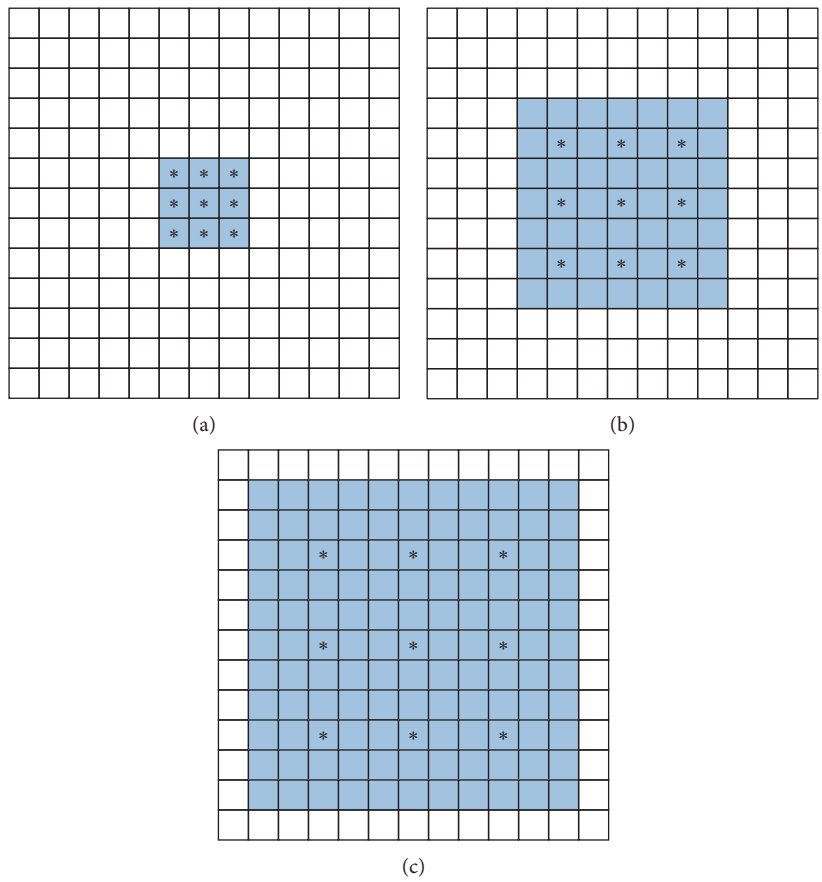


Figure 4: Example of expanded convolution.

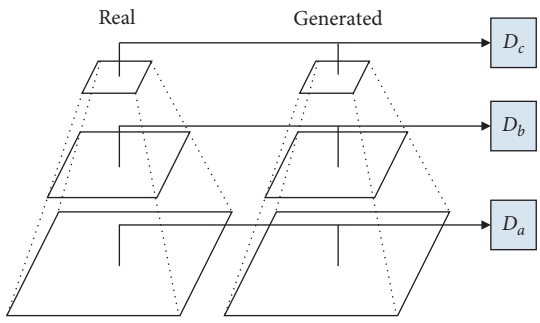


Figure 5: Multiscale discriminator, consisting of three discriminators of different scales: D_a, D_b, and D_c.

generation quality and performance of the hard exudate detection system.

4.1. Experimental Setup. Datasets: the training and testing stages of Com-GAN involve three datasets, namely, self-selected dataset, DIARETDB1 [45], and e-ophtha EX [46].

Self-selected dataset: 4000 fundus images of 3504 * 2336 pixels were selected from the training set for the 2015 Kaggle diabetic retinopathy detection competition. The ophthalmologists marked whether there were hard exudates, and of the 4000 images, 735 contained hard exudates and 3265 contained none.

DIARETDB1 : the datasets contains 89 fundus images of 1500 * 1152 pixels, including 47 images with hard exudates and 42 images without hard exudates.

E-ophtha EX : the datasets contains 82 fundus images of 3 different resolutions, including 47 images with hard exudates and 35 without hard exudates.

The self-selected dataset is used as the training set for the hard exudate detection system and also as the training set for Com-GAN. The public datasets e-ophtha EX and DIARETDB1 are used as the test sets for the hard exudate detection system to test the system performance. To facilitate training and testing, the sizes of all images were adjusted in the previous datasets to 256 * 256.

Evaluation criteria: this paper used different evaluation criteria for Com-GAN and hard exudate detection systems. For the Com-GAN, this paper conducted qualitative and quantitative evaluation on the generation quality from both subjective and objective aspects. Subjectively, three observers were asked to independently perform visual assessment during the experiment; objectively, Structural Similarity Index (SSIM) [47] and Sharpness Difference (SD) were applied to measure the similarity between the generated image and the real image at the pixel level, and Inception Score (IS) [48] and Fréchet Inception Distance (FID) [49] were used to evaluate the generated image from the perspective of high-level feature space.

SSIM models the similarity between the generated image and the real image as a combination of three different factors: brightness, contrast, and structure. The mean value is used as the brightness estimate, the standard deviation as the contrast estimate, and the covariance as the measure of structural similarity. The value can better reflect the subjective perception of human eyes, with the range being [0, 1]. The larger the value, the higher the similarity between images. The SSIM formula is

$$\mathrm{SSIM}(X,Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X{}^2 + \mu_Y{}^2 + C_1)(\sigma_X{}^2 + \sigma_Y{}^2 + C_2)}, \quad (8)$$

where μ_X and μ_Y represent the mean values of the generated image X and the real image Y, σ_X and σ_Y represent the standard deviations of the generated image X and the real image Y, and C_1 and C_2 are constants introduced to prevent the denominator from being 0.

Sharpness Difference (SD) is used to represent the difference in clarity between the generated image and the real image. The larger the SD value is, the smaller the difference in sharpness between the images is and the closer the generated image is to the real image. The formula of the SD between the generated image X and the real image Y is

$$\mathrm{SD}_{X,Y} = 10\log_{10}\left(\frac{\mathrm{MAX}_Y{}^2}{\mathrm{grads}_{X,Y}}\right), \quad (9)$$

where MAX_Y is the maximum pixel value of the image and $\mathrm{grads}_{X,Y}$ is the gradient difference between the image X and the image Y.

Inception Score (IS) evaluates the generated image from the aspects of quality and diversity. In theory, the closer the image is to the real image, the higher the IS score will be. The calculation formula is

$$\mathrm{IS} = \exp\left(\mathrm{E}_{x\sim p_g} D_{\mathrm{KL}}(p(y\mid x) \parallel p(y))\right), \quad (10)$$

where x represents the picture generated from the generator, y the predicted label of x, and D_{KL} the KL divergence between $p(y\mid x)$ and $p(y)$.

Since the ImageNet dataset does not contain labelled fundus categories, this paper does not directly use the pretrained Inception model, but the AlexNet model [50] trained on the Kaggle dataset instead for scoring.

FID assumes that the abstract features of the generated sample and the real sample in the middle layer of the classifier conform to a multivariate Gaussian distribution, and FID is the Fréchet distance between these two Gaussian distributions. The smaller the FID value is, the closer the two Gaussian distributions will be to each other and the closer the generated image will be to the real image. The FID calculation formula is

$$\mathrm{FID}(\mathrm{x},\mathrm{g}) = \left\|\mu_x - \mu_g\right\|_2^2 + T_r\left(\Sigma_x + \Sigma_g - 2\left(\Sigma_x\Sigma_g\right)^{\frac{1}{2}}\right), \quad (11)$$

where μ_g and Σ_g are the mean and variance of the generated sample Gaussian distribution and μ_x and Σ_x are the mean and variance of the true sample Gaussian distribution, respectively. T_r represents the trace of the matrix.

The function of the hard exudate detection system is to determine whether the image contains hard exudates. In this paper, the image containing hard exudates is marked as a positive sample, and accuracy (AC), sensitivity (SE), and specificity (SP) are used as performance evaluation indices. The calculation formulas are as follows:

$$AC = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \quad (12)$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad (13)$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \quad (14)$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

Experimental parameters and environment: both the generator model and the discriminator model used the Adam optimizer [51]. The parameter $\beta 1$ was set to 0.9, the parameter $\beta 2$ to 0.99, and the learning rate to 0.001. This paper used the Pytorch platform for coding as it can dynamically create new calculation charts to facilitate experiment debugging. The server configuration used is CPU E5-2620 v4 @ 2.10 GHz, NVIDIA Tesla V100 16 G.

4.2. Experimental Results of the Com-GAN. This paper used the trained im-WGAN to generate vascular trees and then denoised the resulting images, with the results shown in Figure 6.

Vascular trees segmented from a real fundus image are in the first row in Figure 6. During the training process, they were input to the discriminator as a training set to guide the generator to generate images. The generated vascular trees are in the second row.

The vascular trees generated by im-WGAN and the labels of the images to be generated were input into the trained im-CGAN generator to obtain the complete generated images. The specific results are shown in Figure 7. The generated fundus images without hard exudates are in the first row; generated fundus images with hard exudates are in the second.

The segmented vascular trees are in the first column in Figure 7, which were input into the generator together with the labels during the training process. The real fundus images are in the second column. The real fundus images in the training process were input into the discriminator together with the corresponding vascular trees and the labels. The vascular trees generated by im-WGAN are in the third column. During the test, the vascular trees and the label of the images to be generated were input into the generator to generate the complete fundus images shown in the fourth column.

4.3. Generation Quality Evaluation. This section compares the proposed method with the current mainstream generation models for evaluation of generation quality:

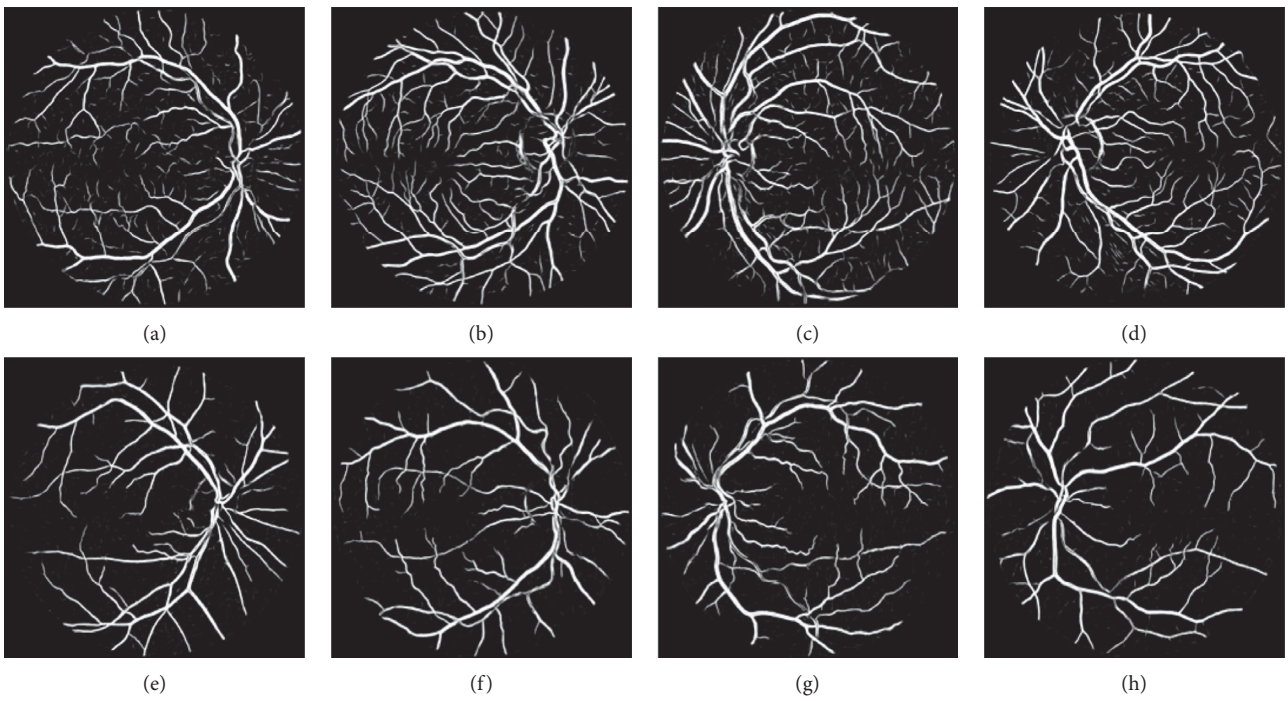


Figure 6: Results of vessel tree generation.

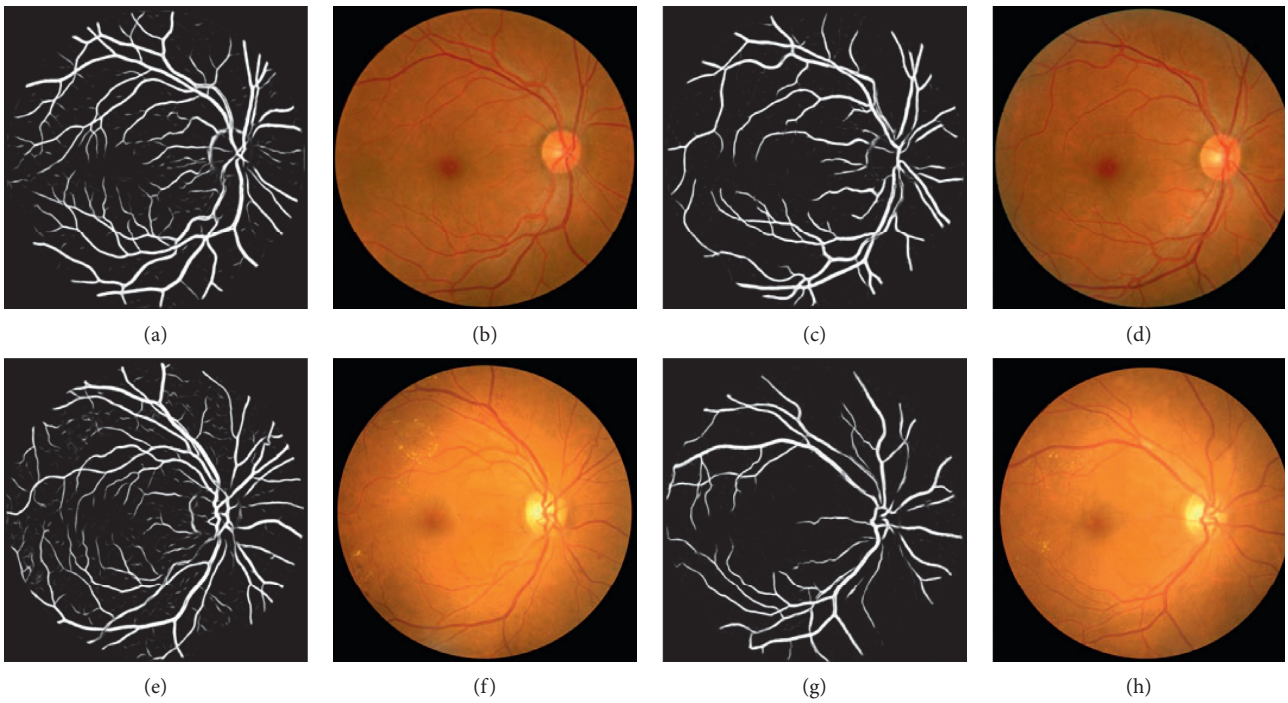


Figure 7: Results of complete image generation.

CGAN [27]: CGAN increases controllability by introducing labels in the original GAN, thereby generating images corresponding to the labels.

Pix2pix [28]: this model introduces the image x in the generator and the discriminator to guide the generator to generate the image y, where x and y, respectively, represent images in different domains X and Y, so the function of pix2pix can also be understood as completing image translation from domain X to domain Y. In the experiment in this paper, the segmented vascular tree was used as the image x, which was then mapped to the complete image y using the pix2pix model.

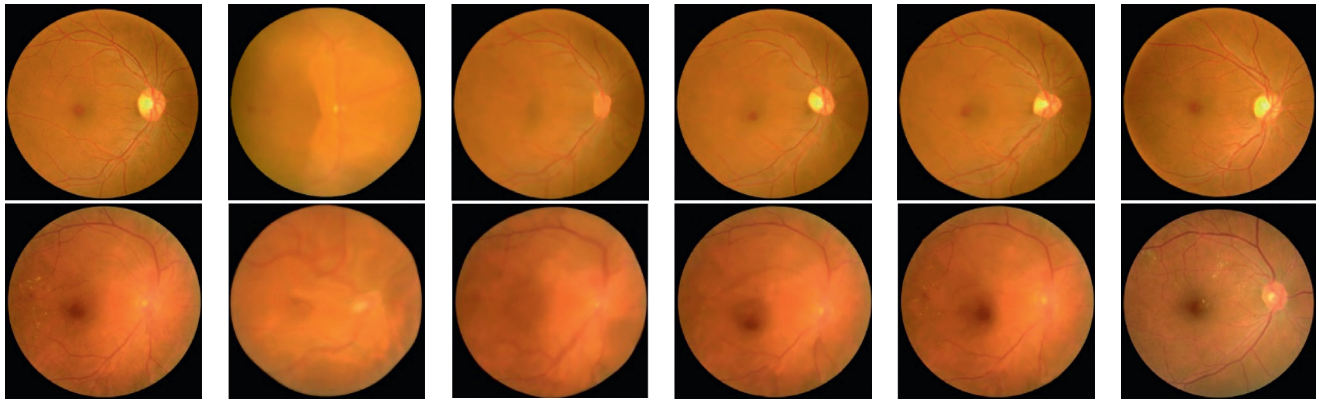


Figure 8: Samples generated by each model. Real image (a), CGAN (b), pix2pix (c), BiGAN (d), IntroVAE (e), and Com-GAN (f).

BiGAN [24]: the discriminator in the original GAN only receives samples as input and cannot learn meaningful intermediate representations. BiGAN inputs the intermediate results and samples from the generator into the discriminator, so that the model has the ability to learn meaningful feature representations.

IntroVAE [25]: IntroVAE can improve itself by self-evaluating the generated samples. By integrating the advantages of both GAN and VAE, not only can it generate high-quality images, but also it also maintains VAE stability.

4.3.1. Qualitative Evaluation. From the visual point of view, this paper conducted a qualitative evaluation on the images generated by the previous methods, as shown in Figure 8.

Normal image samples are in the first row of Figure 8 , and image samples with hard exudates are in the second row. From the visual point of view, the evaluations of the three observers are summarized as follows: the images generated by the CGAN model only show the fundus outline and part of the main blood vessels, with almost no clear details of the blood vessels and hard exudates; those generated by pix2pix exhibit clear details of blood vessels, but the optic disc and macular area are blurred, and no obvious hard exudate area is observed.

Compared with those generated by the previous two methods, the images generated by BiGAN, IntroVAE, and the proposed method are more realistic. The images generated by Com-GAN have more semantic details and the sharpness is the closest to that of real images. As shown in Figure 9, BiGAN and IntroVAE fabricated some vascular details that do not conform to medical principles to deceive the discriminator, while the image generated by Com-GAN is based on a vascular tree, so the vascular details are more realistic.

All the other methods generate complete fundus images in one step. In order to control whether the generated image contains hard exudates, pix2pix, BiGAN, and IntroVAE, all use the reconstruction method to generate an image with the same label as the input image, so it is a one-to-one relationship between the generated image and the reconstructed one. CGAN and Com-GAN can control the type of image by the label y, so it is a one-to-many relationship between the label and the generated images. In this way, the diversity of the images generated is better than that by other methods.

4.3.2. Quantitative Evaluation. This section conducted a comparative analysis of the normal images, images with hard exudates, and vascular trees generated by the previous models. The vascular trees under evaluation were segmented from the images generated by each model using the segmentation model. Here, FID was used to measure the similarity between these vascular trees and the real segmented ones. The experimental results are shown in Table 2. As can be seen, the larger the SSIM, SD, and IS and the smaller the FID, the higher the similarity between the generated image and the real one, the better the generation quality, and the richer the diversity.

It can be seen from the experimental results that, in terms of complete image generation, Com-GAN and IntroVAE were very close in SSIM, SD, and FID and better than the other three methods, but the IS score of Com-GAN was higher than those of the other four models, 22.20% higher than the average value of IntroVAE, the best among the other methods, indicating that Com-GAN is superior to other models in terms of both generation quality and diversity. The evaluation results of vascular trees show that the vascular trees generated by Com-GAN were the closest to the real ones. Therefore, judging from the three aspects, the images generated by Com-GAN are more realistic than those generated by the other four models.

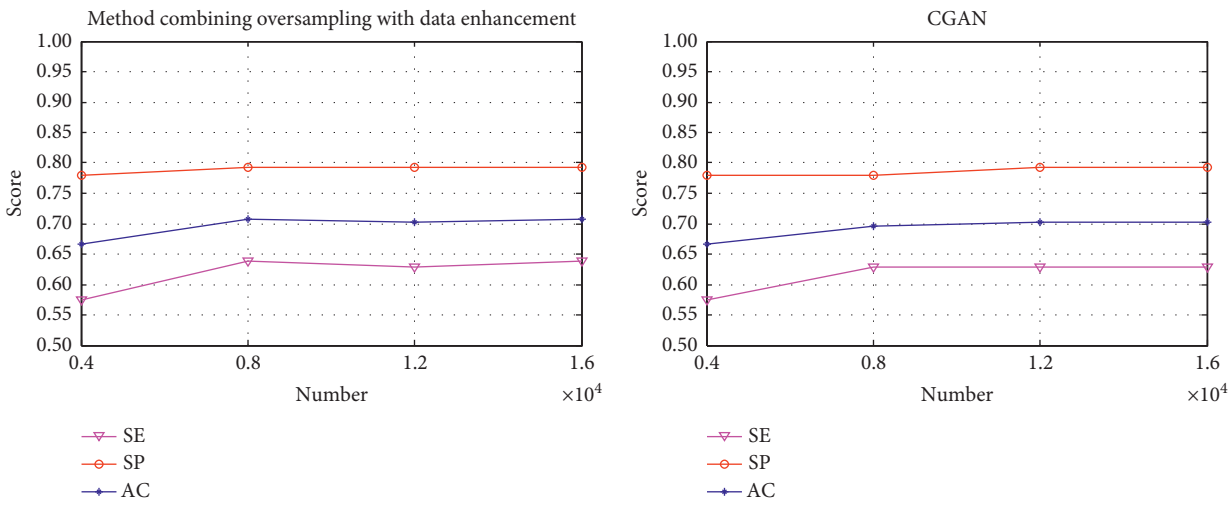
4.4. Performance Comparison of Intelligent Diagnostic Systems. In order to verify the effectiveness of this method in practical applications, this paper applied the generated images to the training of the hard exudate detection system and tested the performance of the detection system on the data-enhanced test set. The original test set was a mixed dataset of DIARETDB1 and e-ophtha EX. The enhanced

(a) (b) (c)

Figure 9: BiGAN (a), IntroVAE (b), and Com-GAN (c) generation sample.

Table 2: Quantitative evaluation results.

Models	Normal images				Images with hard exudates				Vascular trees
	SSIM	SD	IS	FID	SSIM	SD	IS	FID	FID
CGAN	0.52	14.45	3.14	19.66	0.50	14.41	3.00	20.01	24.32
pix2pix	0.64	17.98	3.92	16.03	0.58	17.32	3.78	17.77	16.43
BiGAN	0.71	19.51	4.55	15.42	0.67	18.99	4.41	15.98	17.56
IntroVAE	0.78	20.18	4.92	15.13	0.72	19.08	4.81	15.44	17.03
Com-GAN	0.77	20.42	6.10	14.89	0.74	19.11	5.89	15.37	15.87



(a) (b)

Figure 10: Continued.

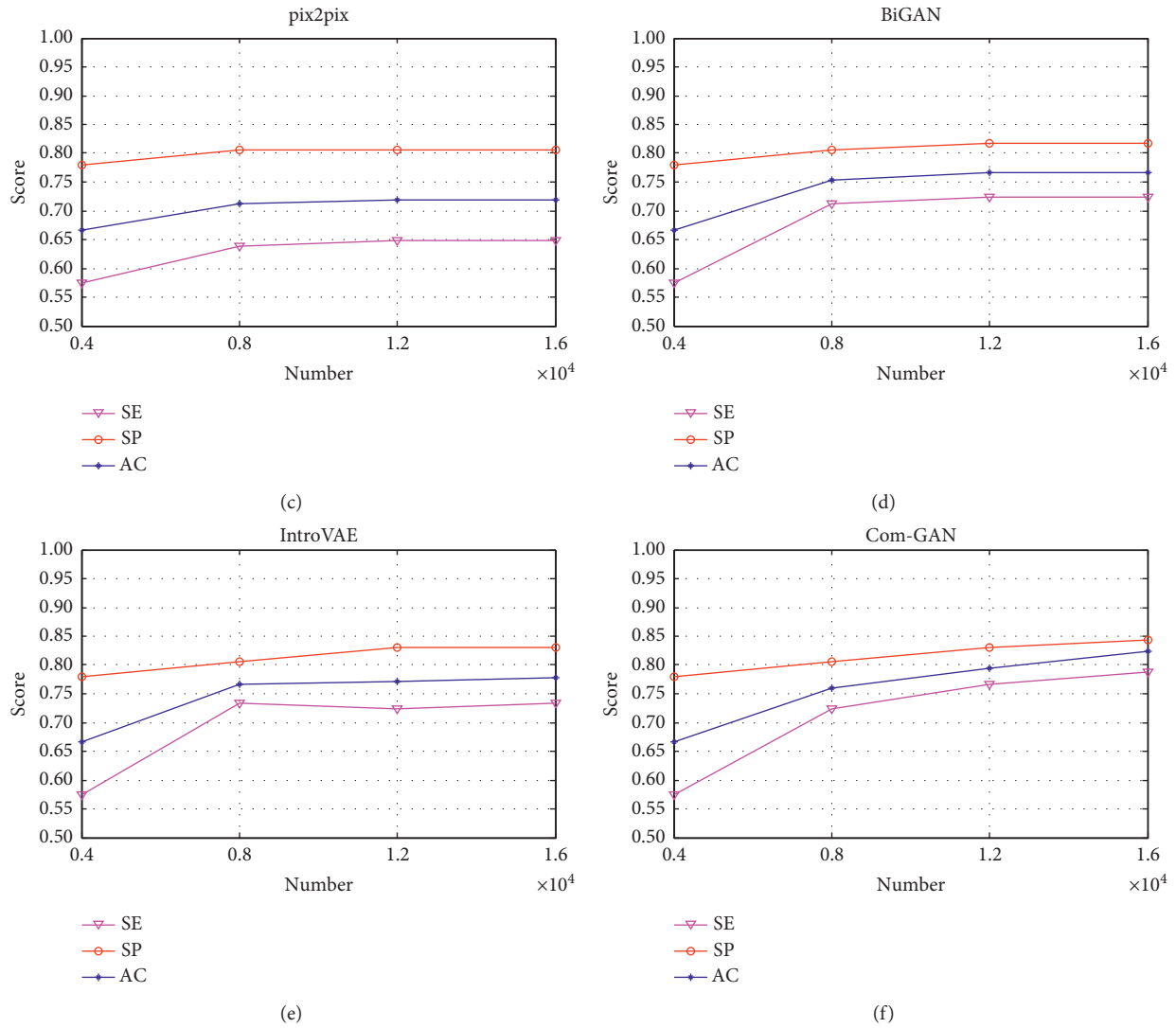


Figure 10: Evaluation results of various methods. (a) Method combining oversampling with data enhancement. (b) CGAN. (c) pix2pix. (d) BiGAN. (e) IntroVAE. (f) Com-GAN.

dataset contained 470 fundus images with hard exudates and 385 fundus images without hard exudates.

The hard exudate detection system was implemented using the AlexNet model, and the final classification results were changed into two categories, that is, input images with hard exudates and those without hard exudates.

This paper used the image sets generated by the previous models to train the hard exudate detection system, with the test results shown in Figure 10. In the figure, the horizontal axis represents the size of the training set. The initial set was a self-selected dataset containing 4000 real images. Then, 4000 images were added each time, and finally it was increased to a size of 16000 images. The ratio of positive to negative samples was adjusted to 1 : 1 from the first expansion of data. The vertical axis represents the evaluation index score. Figures 10(a)–10(f), respectively, show the evaluation results of the method combining oversampling with data enhancement, CGAN, pix2pix, BiGAN, IntroVAE, and Com-GAN.

It can be drawn from Figure 10 that the method combining oversampling with data enhancement and the methods that directly generate images showed significant improvements in the first expansion of the dataset. However, when the training dataset was expanded to 12000 and 16000, no significant improvement was observed in the performance. On the other hand, with the expanded dataset of Com-GAN, the system performance improved in all of the last three evaluations. What is more, after the third expansion of data, the final SE, SP, and AC reached 0.787, 0.844, and 0.824, respectively, which were higher than the final results of the other models. Compared with those of the initial dataset, the indices were increased by 0.213, 0.065, and 0.157, respectively, and compared with those of IntroVAE, the best among the other methods, the indices were higher by 0.053, 0.013, and 0.046, respectively. This verifies that the proposed method is superior to other models in practical applications.

5. Conclusion

This paper proposes a new type of GAN : Com-GAN, used to generate fundus images. Com-GAN divides the fundus image generation process into two stages. First, im-WGAN is used to generate a vascular tree, and then im-CGAN is used to generate a complete fundus image on the basis of the vascular tree. The proposed method alleviates the problem of uneven distribution of samples in the fundus image training set and at the same time mitigates the problem of insufficient training samples. After qualitative and quantitative evaluation and application in the detection system, it is proved that Com-GAN can generate high-quality fundus images compared with current mainstream generation models, and the generated images are highly diversified, rather than simple repeats of the images in the training set. In addition, the proposed two-step generation method can be flexibly applied to expand other datasets. In the future, more research will be carried out to explore the application of this method in fields like image style transfer and image translation.

Data Availability

All data included in this study are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61300098 and the Natural Science Foundation of Heilongjiang Province under Grant LH2019C003.

References

[1] S. Subbotin, A. Oliinyk, V. Levashenko, and E. Zaitseva, "Diagnostic rule mining based on artificial immune system for a case of uneven distribution of classes in sample," *Communications-Scientific Letters of the University of Zilina*, vol. 18, no. 3, pp. 3–11, 2016.

[2] C. Savu-Krohn, G. Rantitsch, P. Auer, F. Melcher, and T. Graupner, "Geochemical fingerprinting of coltan ores by machine learning on uneven datasets," *Natural Resources Research*, vol. 20, no. 3, pp. 177–191, 2011.

[3] H. Inoue, "Data augmentation by pairing samples for images classification," 2018, https://arxiv.org/abs/1801.02929.

[4] P. D. Faris, W. A. Ghali, R. Brant, C. M. Norris, P. D. Galbraith, and M. L. Knudtson, "Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses," *Journal of Clinical Epidemiology*, vol. 55, no. 2, pp. 184–191, 2002.

[5] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," in *Proceedings of the European Conference on Computer Vision*, pp. 3–17, Zurich, Switzerland, September 2014.

[6] A. Estabrooks, T. Jo, and N. Japkowicz, "A multiple resampling method for learning from imbalanced data sets," *Computational Intelligence*, vol. 20, no. 1, pp. 18–36, 2004.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.

[8] H. Cao, X.-L. Li, D. Y.-K. Woon, and S.-K. Ng, "Integrated oversampling for imbalanced time series classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 12, pp. 2809–2822, 2013.

[9] W. W. Y. Ng, J. Hu, D. S. Yeung, S. Yin, and F. Roli, "Diversified sensitivity-based undersampling for imbalance classification problems," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2402–2412, 2014.

[10] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, 2009.

[11] S. García and F. Herrera, "Evolutionary undersampling for classification with imbalanced datasets: proposals and taxonomy," *Evolutionary Computation*, vol. 17, no. 3, pp. 275–306, 2009.

[12] M. Gao, X. Hong, S. Chen, and C. J. Harris, "A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems," *Neurocomputing*, vol. 74, no. 17, pp. 3456–3466, 2011.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, https://arxiv.org/abs/1312.6114.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," in *Proceedings of the Advances in Neural Information Processing Systems*, Morgan Kaufmann Publishers, Burlington, MA, USA, pp. 2672–2680, June 2014.

[15] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1155–1164, Long Beach, CA, USA, February 2019.

[16] R. Liu, Y. Liu, X. Gong, X. Wang, and H. Li, "Conditional adversarial generative flow for controllable image synthesis," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7992–8001, Long Beach, CA, USA, April 2019.

[17] M. Qi, H.-Y Lee, H.-Y Tseng, S. Ma, and M.-H. Yang, "Mode seeking generative adversarial networks for diverse image synthesis," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1429–1437, Long Beach, CA, USA, March 2019.

[18] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton, "Veegan: reducing mode collapse in gans using implicit variational learning," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3308–3318, Long Beach, CA, USA, May 2017.

[19] J. Y. Zhu, R. Zhang, D. Pathak et al., "Toward multimodal image-to-image translation," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 465–476, Long Beach, CA, USA, November 2017.

[20] X. Chen, C. Xu, X. Yang, L. Song, and D. Tao, "Gated-gan: adversarial gated networks for multi-collection style transfer," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 546–560, 2019.

[21] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: fine-grained image generation through asymmetric training," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2745–2754, Venice, Italy, March 2017.

[22] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, https://arxiv.org/abs/1511.06434.

[23] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 214–223, Sydney, NSW, Australia, August 2017.

[24] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, https://arxiv.org/abs/1605.09782.

[25] H. Huang, Z. Li, R. He, Z. Sun, and T. Tan, "IntroVAE: introspective variational autoencoders for photographic image synthesis," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 52–63, Montreal, Canada, July 2018.

[26] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2017, https://arxiv.org/abs/1709.01872.

[27] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, https://arxiv.org/abs/411.1784.

[28] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125–1134, IEEE, Honolulu, HI, USA, July 2017.

[29] P. Appan and J. Sivaswamy, "Retinal image synthesis for CAD development," in *Proceedings of the International Conference On Image Analysis and Recognition*, pp. 613–621, Póvoa de Varzim, Portugal, June 2018.

[30] M. García, C. I. Sánchez, M. I. López, D. Abásolo, and R. Hornero, "Neural network based detection of hard exudates in retinal images," *Computer Methods and Programs in Biomedicine*, vol. 93, no. 1, pp. 9–19, 2009.

[31] A. Sopharak, M. N. Dailey, B. Uyyanonvara et al., "Machine learning approach to automatic exudate detection in retinal images from diabetic patients," *Journal of Modern Optics*, vol. 57, no. 2, pp. 124–135, 2010.

[32] B. R. Masters, "Fractal analysis of the vascular tree in the human retina," *Annual Review of Biomedical Engineering*, vol. 6, no. 1, pp. 427–452, 2004.

[33] W. Xue, X. Mou, L. Zhang, and X. Feng, "Perceptual fidelity aware mean squared error," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, pp. 705–712, IEEE, Sydney, NSW, Australia, December 2013.

[34] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, IEEE, Honolulu, HI, USA, July 2017.

[35] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local nash equilibria in continuous games," in *Proceedings of the 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 917–924, IEEE, Monticello, IL, USA, October 2013.

[36] Q. Jin, X. Luo, Y. Shi, and K. Kita, "Image generation method based on improved condition GAN," in *Proceedings of the 2019 6th International Conference On Systems and Informatics (ICSAI)*, pp. 1290–1294, IEEE, Shanghai, China, November 2019.

[37] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. Eugene Stanley, "Analysis of symbolic sequences using the Jensen-Shannon divergence," *Physical Review E*, vol. 65, no. 4, 2002.

[38] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 5767–5777, Long Beach, CA, USA, December 2017.

[39] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, October 2015.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, July 2016.

[42] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, Lille, France, July 2015.

[43] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, https://arxiv.org/abs/1802.05957.

[44] M. Mateen, J. Wen, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, 2019.

[45] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen et al., "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proceedings of the British Machine Vision Conference 2007*, vol. 1, pp. 1–10, Warwick, UK, September 2007.

[46] X. Zhang, G. Thibault, E. Decencière et al., "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical Image Analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[47] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, IEEE, Istanbul, Turkey, August 2010.

[48] T. Salimans, I. Goodfellow, W. Zaremba et al., "Improved techniques for training GANs," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2234–2242, Barcelona, Spain, December 2016.

[49] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 6626–6637, Long Beach, CA, USA, December 2017.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.

[51] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.