

Research Article

Fp^{roi}-GAN with Fused Regional Features for the Synthesis of High-Quality Paired Medical Images

Jiale Dong ^{1,2} Caiwei Liu ^{1,2} Panpan Man ^{1,2} Guohua Zhao,^{1,2} Yaping Wu ³
and Yusong Lin ^{2,4,5}

¹School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

²Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China

³Department of Medical Imaging, Henan Provincial People's Hospital, Zhengzhou 450003, China

⁴School of Software, Zhengzhou University, Zhengzhou 450002, China

⁵Hanwei IoT Institute, Zhengzhou University, Zhengzhou 450002, China

Correspondence should be addressed to Yaping Wu; ypwu@ha.edu.cn and Yusong Lin; yslin@ha.edu.cn

Received 31 December 2020; Revised 20 February 2021; Accepted 16 April 2021; Published 26 April 2021

Academic Editor: Dong Nie

Copyright © 2021 Jiale Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of medical image synthesis with generative adversarial networks (GAN) is effective for expanding medical samples. The structural consistency between the synthesized and actual image is a key indicator of the quality of the synthesized image, and the region of interest (ROI) of the synthesized image is related to its usability, and these parameters are the two key issues in image synthesis. In this paper, the fusion-ROI patch GAN (Fp^{roi}-GAN) model was constructed by incorporating a priori regional feature based on the two-stage cycle consistency mechanism of cycleGAN. This model has improved the tissue contrast of ROI and achieved the pairwise synthesis of high-quality medical images and their corresponding ROIs. The quantitative evaluation results in two publicly available datasets, INbreast and BRATS 2017, show that the synthesized ROI images have a DICE coefficient of 0.981 ± 0.11 and a Hausdorff distance of 4.21 ± 2.84 relative to the original images. The classification experimental results show that the synthesized images can effectively assist in the training of machine learning models, improve the generalization performance of prediction models, and improve the classification accuracy by 4% and sensitivity by 5.3% compared with the cycleGAN method. Hence, the paired medical images synthesized using Fp^{roi}-GAN have high quality and structural consistency with real medical images.

1. Introduction

Medical imaging is a clinically important noninvasive diagnostic method; imaging specialists can diagnose breast cancer or precancer through mammography images [1]. With the development of deep learning technology, medical image synthesis [2, 3], classification [4], and segmentation [5] based on deep learning have become topical issues in medical research. Deep neural networks usually require a large number of training samples, and the size of medical image data is usually small because of the high collection cost, thus limiting the application of deep learning models for medical images [6]. Generative adversarial networks [7] usually learn feature mappings

from source modality to target modality by constructing generators and discriminators that can be used to synthesize medical images and thus expand training samples [8, 9]. However, the gradient disappearance, pattern collapse, and structural consistency problems between real and synthetic images in the current GAN research process seriously affect the quality of synthetic images [3]. In addition, the region of interest (ROI) of medical images is a key factor in aiding imaging research and is often used in training medical image segmentation tasks. However, we found that the synthesis of the ROI has rarely been studied [10, 11]. Thus, in the present study, we focused on the synthesis of high-quality medical images and their ROI images.

Nie et al. [12] were the first to propose a generative adversarial model using a fully convolutional neural network as a generator that implements the conversion between MRI and CT images of brain tumor images. The 3D-based fully convolutional neural network proposed in this paper well solves the problem of discontinuity across slices in 2D neural networks, and the method improves the quality of the generated images by calculating the gradient difference of the images as a loss function. The experimental results show that the method proposed in this paper can effectively predict CT images from MRI images, which is an early research and exploration of generative adversarial networks in the field of medical image synthesis. Guibas et al. [13] propose a novel pipeline model based on generative adversarial networks for the current medical images that are not easily accessible. The model proposed in this paper consists of Stage-I GAN and Stage-II GAN, which enables the generation of higher quality images by enhancing the learning of mask image features of images. In addition, John et al. created an online synthetic medical image database called SynthMed, while again demonstrating the feasibility of GAN-based synthesis of medical images. In addition, Chartsias et al. [8] proposed a multi-input, multi-output fully convolutional neural network for MRI synthesis, which embeds all input modalities into a shared potential space and converts the shared features into target output modalities by learning the potential space mapping through a decoder. Although this method can achieve multimodal output, the generated images are adulterated with redundant information. Wolterink et al. [9] used cycleGAN to learn the mapping of source modality to target modality through adversarial loss, resulting in synthetic CT images that are similar to the real CT images. Considering the lack of direct constraints between the real CT images and the synthesized CT images, this approach still cannot guarantee the structural consistency between the synthesized and the input images. Kang et al. [14] proposed a conditional GAN to improve model estimation and quantitatively evaluate the resulting images, but this approach resulted in uneven quality across domains of the synthesized images. Huang et al. [15] synthesized glioma images by using the WEENIE model, which uses a priori information instead of noise as input to the model, but the consistency of the synthesized images with the real images needs to be improved.

In the study of GAN-based generative models, the structural consistency between the real and synthetic images usually affects the quality of the synthetic images [3]. To improve the structural inconsistency between the real and synthesized image during image synthesis and synthesize the ROI of the image, we proposed a new method for synthesizing paired medical images based on cycleGAN. The method incorporates regional a priori features on the basis of cycleGAN two-stage cycle consistency to achieve high-quality medical images and their ROI synthesis. In the medical image synthesis process, the first stage model implements feature mapping from the medical image domain to the ROI domain and targets the learning contrast features of ROI and non-ROI tissues. The second stage network reduces the ROI domain to the medical image domain to

synthesize medical images. By contrast, in the synthesis process of ROI, the input ROI image is first reduced to a medical image, and then a high-quality ROI image is synthesized based on the regional contrast of the medical image. The two-stage synthesis process is implemented through the cycle consistency function of cycleGAN [16]. In this paper, we validated the quality of the synthesized images by using two publicly available datasets, where the benign data of the INbreast dataset has no corresponding ROI images. Then, we quantitatively analyzed the synthesis results from various metrics only. The results show that our proposed method effectively improves the structural consistency between the synthesized and real image, and the quality of the synthesized image is better than several recent popular models. In addition, we have verified that the images synthesized in this paper can improve the classification performance of the prediction model in the brain glioma classification experiment. The experimental results demonstrate that the method in this paper can effectively generate high-quality paired medical images, which will bring new solutions for medical disease research where it is difficult to obtain data.

The contribution of this work is summarized as follows.

- (1) We proposed a new synthesis method for the synthesis of paired medical images on the cycle consistency mechanism of cycleGAN and called it Fp^{roi} -GAN
- (2) To improve the quality of the synthesized images, this paper assists the generative model to learn ROI and non-ROI organizational features by supplementing a priori regional features
- (3) Fp^{roi} -GAN proved its effectiveness on two experimental datasets, and the experimental results show that our method can effectively improve the structural consistency of synthesized images with real images and outperform many popular image synthesis methods

2. Materials and Methods

2.1. Dataset. INbreast [17, 18] contains 303 normal (no mass) mammograms and 107 pairs of mammograms, including mass data and corresponding ROI images. Considering that training requires paired data, only 107 pairs of images containing masses were finally selected as the experimental data and then preprocessed. The mammograms had a resolution of $3,328 \times 4,084$ pixels or $2,560 \times 3,328$ pixels, and the images were stored in dicom format. We first cropped the original images according to the provided lesion areas, and the cropped images to 256×256 were converted into PNG format, as shown in Figure 1(a). The processed paired data were divided into training and test sets in a ratio of 7:3, and the image intensity was linearly normalized to [0,1] by using maximum normalization. Subsequently, the influence of data irregularity on the experimental results was eliminated, and the network was accelerated to determine the optimal solution.

The BRATS 2017 [19, 20] dataset contains 285 medical images and their corresponding ROI images from four

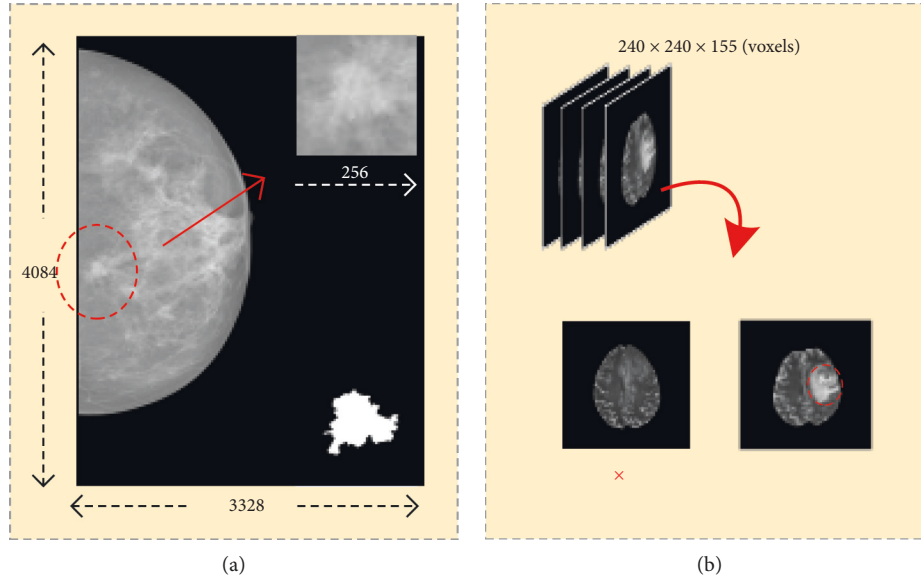


FIGURE 1: (a) Cropping of the INbreast. (b) Slices containing tumor regions were extracted from the 3D images of glioma; × indicates that images that do not contain tumor domains were excluded.

sequences, T1-weighted (T1), T1-weighted and contrast-enhanced (T1ce), T2-weighted (T2), and FLAIR, including 210 high-grade gliomas (HGG) and 75 low-grade gliomas (LGG), with image sizes of $240 \times 240 \times 155$ voxels. T2 sequences were selected as the experimental data, and the 90th and 100th layer slices of the HGG images (the middle layer contains more brain image information relative to the edge of the voxel images) and the corresponding ROI slices were extracted first. Similarly, the 90th, 95th, 100th, and 105th layer slices of LGG images and the corresponding ROI slices were extracted, and each of the 272 pairs of HGG and LGG images were collected, as shown in Figure 1(b). Finally, all images were adjusted to 256×256 pixels. The two small datasets, HGG and LGG, were normalized according to INbreast's partitioning and processing method.

2.2. Methods. To enable the network learn the contrast information of ROI and non-ROI tissues, we improved the cycleGAN model and proposed a pairwise image synthesis method that incorporates regional features. Figure 2 shows the flowchart of the model, where the input of the network is the medical image and its corresponding ROI. Before the network started training, the medical image matrix was first multiplied with its ROI image matrix to obtain the regional image containing only the tumor. Then, we designed a regional feature extraction block (RFB) to extract the semantic features of regional images and fuse the extracted regional features with medical images as the input of source domain X and ROI as the input of target domain Y . During network training, the model discriminates between ROI and non-ROI organizational features by learning the mapping of domain X to domain Y . The a priori regional features enhance the learning process and then reduces domain Y to domain X to synthesize medical images. ROI synthesis first reduces the mapping of domain Y to domain X and

synthesizes high-quality ROI images based on the mapping of domain X to domain Y . Figure 2(c) shows the synthesis of medical images, and the process can be represented as: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ as shown in (i); similarly, the synthesis process of ROI can be represented as: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$, as shown in (ii). The model proposed in this paper is composed of two generators, namely, G and F , and two discriminators, namely, D_x and D_y .

2.2.1. Regional Feature Extraction. First, the medical image was multiplied with the ROI matrix to obtain the regional image, and the operation steps are shown in Figure 2(a). We designed the RFB for extracting high-level semantic features of the region image, and its structure is shown in Figure 3. The feature extraction block is a simple convolutional neural network consisting of two mirror fill layers, three convolutional layers, and one deconvolutional layer. In the network, the operational details of the three convolutional layers are zoomed into the right side of Figure 3, where the convolutional details include convolution, instance normalization, and activation operations. The feature map output after the RFB is fused with the medical image as the input of domain X .

2.2.2. Network Architecture. The Fp^{roi} -GAN model consists of two generators and two discriminators, where the structures of the generators G and F are shown in Figure 4. The generator consists of four convolutional layers, two fusion layers, and two deconvolutional layers, and the operation details of each convolutional layer include convolution, instance normalization, and activation operations. To extract each pixel in the fused image, the generator first performs a 3×3 mirror fill of the image, and the feature map size is filled from 256×256 pixels to 262×262 pixels after filling. After three convolution processes, a 128-dimensional

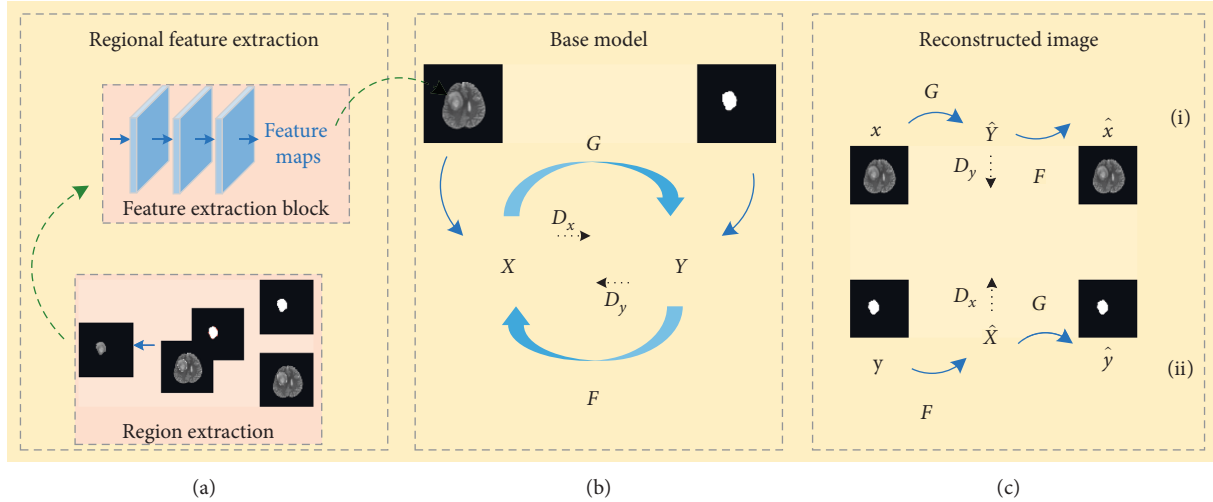


FIGURE 2: (a) Regional feature extraction method. (b) The base model is a like-cycleGAN model consisting of two generators and two discriminators. (c) Synthesis of paired images. (i) Process synthesis of medical images. (ii) Process synthesis of ROI images.

64×64 feature map is obtained. The convolution aims to downsample the image and extracts its structural features, where the details of the three convolution layer operations are zoomed into the corresponding color boxes on both sides. In addition, we added two fusion layers to the generator to preserve the low-level image information. Finally, two deconvolution layers restore the image to its initial size and complete the image synthesis.

The inputs of the discriminator D_x include real and synthetic medical images, while the inputs of the discriminator D_y include synthetic and real ROI images. The discriminator consists of four convolutional layers, flatten layer, dense layer, and sigmoid activation layer. The

convolved image is flattened by the flatten layer, and the dense layer reduces the features to one dimension. Finally, the sigmoid function determines whether the image is synthetic or real, and the details of the discriminator layers are depicted in Figure 5. The discriminator is executed immediately after the output of the generator.

2.2.3. Training Loss. The loss functions used in the synthesis of the images include the traditional adversarial [7] and cycle consistency loss [16]. The model uses adversarial loss as the mapping function. The mapping function $G: X \rightarrow Y$ and its discriminator D_Y are expressed in (1) as follows:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim P_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim P_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

where $G(x)$ generates an image similar to the Y domain, and D_Y distinguishes between the synthesized sample and the real sample. In this process, $G(x)$ aims to distinguish between information from ROI and non-ROI tissue, resulting in subsequent $F(G(x))$ restoration process. G aims to minimize this objective against an adversary D_Y that tries to maximize it, in which $\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$. Similarly, in the restoration process of $F(G(x))$, a similar mapping function $F: Y \rightarrow X$ learns the mapping from the ROI image to the medical image, in which $\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$, where D_X represents its discriminator.

Traditional adversarial losses can only intermittently learn the mapping function from domain X to domain Y or vice versa. To constrain the consistency of the real image with the synthetic image, we used a cycle consistency loss function in the model to enhance the reduction process. In Figure 2(c), $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ constrains the synthesis process of the medical image, while $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ constrains the synthesis process of the ROI image. These two components constitute the cycle consistency loss, as shown in the following:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim P_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim P_{data}(y)}[\|G(F(y)) - y\|_1] \quad (2)$$

2.3. Evaluation Measures. The peak signal-to-noise ratio (PSNR) [21], structural similarity (SSIM) [22], and

multiscale structural similarity (MS-SSIM) [23] were used for the quantitative evaluation of the synthesized medical

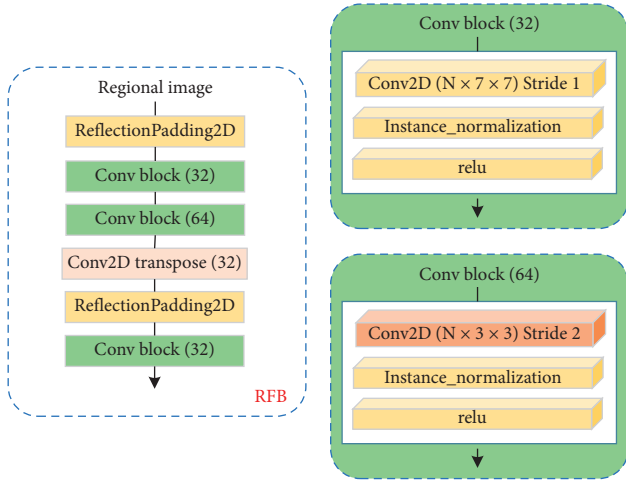


FIGURE 3: RFB architecture; the convolution process zoomed into the box on the right side corresponding to the dimension.

images. Dice coefficient and Hausdorff distance were used for the quantitative evaluation of the synthesized ROI images. Given the original input and synthetic images, the PSNR can be defined as follows:

$$\text{PSNR}(x, F(G(x))) = 10 \log_{10} \frac{\text{MAX}_{\text{range}}^2(x, F(G(x)))}{N_{\text{voxel}}^{-1} \|x - F(G(x))\|_2^2}, \quad (3)$$

$$\text{Hausdorff}(y, G(F(y))) = \max(\max_{y \in y} (\min(d(y, G(F(y))))), \max_{G(F(y)) \in G(F(y))} (\min(d(y, G(F(y)))))), \quad (6)$$

where d represents the Euclidean distance.

3. Results and Discussion

Our network implementation was based on the PyTorch framework. All experiments were performed on a 12-core Intel Xeon 3.7 GHz CPU and GeForce RTX 2080 (8 GB) by using the Ubuntu 18.04 operating system. All figures were plotted on a computer with Windows 10 (8 GB) operating system. During the synthesis task, all models were trained for 300 epochs, where the trained models used the Adam optimizer [28] with default parameters, and the learning rate was set to 0.0002.

3.1. Results of the INbreast Dataset. This subsection provides a comparison of three commonly used synthesis models, namely, DCGAN [11], Pix2Pix [15], and cycleGAN [16]. Table 1 evaluates the whole and tumor domains of the synthesized images, and Table 2 compares the synthesis results of ROI images. Tables 1 and 2 compare the differences between Fp^{roi}-GAN and the other methods using paired-samples T -tests [29], and the underline indicates a significant difference between Fp^{roi}-GAN and the other methods at a significance level of 0.05. Based on the experimental results in Table 1, the Fp^{roi}-GAN image synthesis method achieved

where $\text{MAX}_{\text{range}}(x, F(G(x)))$ represents the maximum number of pixels for x and $F(G(x))$ images, and N_{voxel} represents the total number of pixels for x or $F(G(x))$. The higher the PSNR value, the better the synthesis performance. SSIM was used to measure three metrics of image brightness, contrast, and structure, which can be expressed as follows:

$$\text{SSIM}(x, F(G(x))) = \frac{(2\mu_x\mu_{F(G(x))} + c_1)(2\sigma_{xF(G(x))} + c_2)}{(\mu_x^2 + \mu_{F(G(x))}^2 + c_1)(\sigma_x^2 + \sigma_{F(G(x))}^2 + c_2)}, \quad (4)$$

where μ and σ^2 denote the mean and variance of the image, respectively, and $\sigma_{xF(G(x))}$ denotes the covariance of x and $F(G(x))$. The closer the SSIM is to 1, the higher the structural similarity is. The larger MS-SSIM values represent a better synthesis performance [24]. Dice coefficients [25, 26] are often used to represent the performance of the synthesized ROI image based on the ROI image y and the synthesized ROI images $G(F(y))$ as follows:

$$\text{Dice}(y, G(F(y))) = \frac{2|y \cap G(F(y))|}{|y| + |G(F(y))|} \quad (5)$$

The Hausdorff distance [27], a complement to the Dice evaluation metric, can be expressed as follows:

the highest results for the three evaluation metrics, whereas the DCGAN synthesized image results were the lowest. Based on the quantitative analysis results of the whole image domain, the Fp^{roi}-GAN values were 0.832, 0.053, and 0.016 higher than those of the cycleGAN method in the three evaluation metrics of PSNR, SSIM, and MS-SSIM, respectively, and 1.813, 0.113, and 0.056 higher than the DCGAN, respectively. In the tumor domain, the Fp^{roi}-GAN values were 3.657, 0.085, and 0.042 higher than those of the cycleGAN method in the three evaluation metrics of PSNR, SSIM, and MS-SSIM, respectively, and 4.911, 0.095, and 0.052 higher than the DCGAN method, respectively. Fp^{roi}-GAN method was significantly improved relative to other synthesis methods in Table 1. Based on the experimental results in Table 2, Fp^{roi}-GAN obtained the highest DICE coefficient, which is 0.154 higher than DCGAN, and the lowest evaluated value in Hausdorff Distance, which is 3.10 lower than DCGAN. Figure 6 shows the visual performance of the four synthesis methods, and Fp^{roi}-GAN performs closer to the original image in some detail positions.

3.2. Results of the BraTS 2017 Dataset. This subsection provides comparison with three commonly used synthetic models, such as DCGAN [11], Pix2Pix [15], and cycleGAN

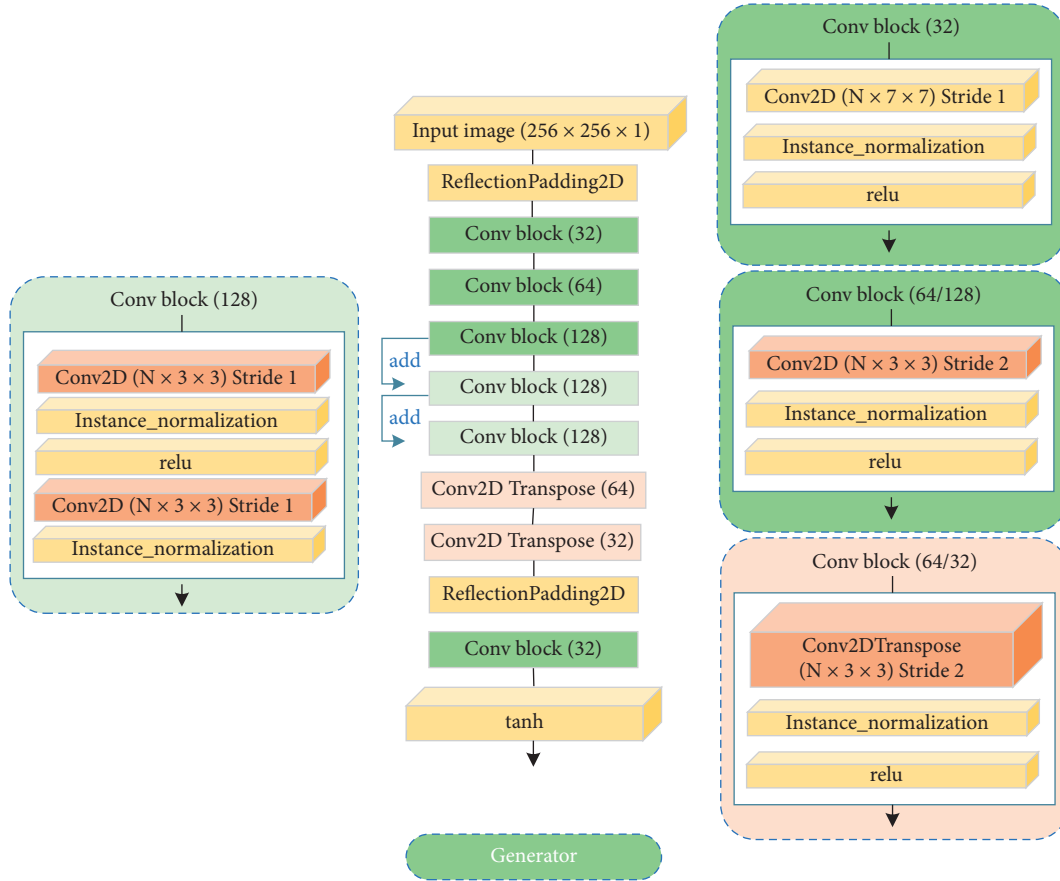


FIGURE 4: Generator architecture; the convolution details of the generator are zoomed into the boxes on both sides.

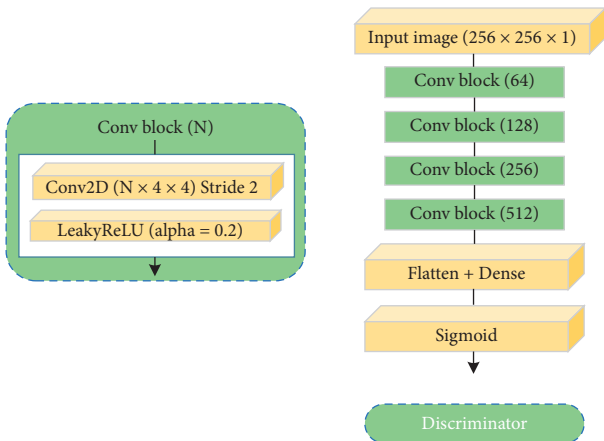


FIGURE 5: Discriminator architecture; Conv2D and LeakyReLU layers were applied to all Conv blocks.

[16]. Tables 3 and 4 compare the differences between Fp^{roi} -GAN and other methods by using paired-sample t -test [29], and the underline indicates that Fp^{roi} -GAN is statistically significantly different from other methods at a significance level of 0.05. Based on the experimental results in Table 3, the quantitative analysis results of Fp^{roi} -GAN in the HGG data for the whole image domain are higher than those of cycleGAN in PSNR, SSIM, and MS-SSIM by 0.604, 0.002,

and 0.003, respectively, and by 9.135, 0.104, and 0.097, compared with DCGAN, respectively. In the tumor domain, the Fp^{roi} -GAN values were higher than cycleGAN in PSNR, SSIM, and MS-SSIM by 6.236, 0.02, and 0.023, respectively, and 12.349, 0.094, and 0.083 higher than DCGAN method, respectively. The quantitative analysis results of Fp^{roi} -GAN in LGG data in the whole image domain are 1.999, 0.006, and 0.008 higher than cycleGAN in the three evaluation metrics of PSNR, SSIM, and MS-SSIM, respectively, and 6.951, 0.069, and 0.066 higher than DCGAN, respectively. In the tumor domain, the Fp^{roi} -GAN values were 11.248, 0.004, and 0.007 higher than cycleGAN and 14.631, 0.105, and 0.079 higher than DCGAN.

Based on the experimental results in Table 4, Fp^{roi} -GAN in HGG data achieved the highest DICE coefficient, which is 0.128 higher than DCGAN, and the lowest evaluated value in Hausdorff distance, which is 3.44 lower than DCGAN. The DICE coefficient of Fp^{roi} -GAN in LGG data part was higher than DCGAN by 0.101, and the Hausdorff distance was lower than DCGAN by 3.75. Figure 7 shows the visual results of the four synthesis methods, in which the synthesis results of the tumor domain, as well as the results of the non-tumor domain, are compared, as shown in the medical images of LGG. III shows the results of the synthesized paired images in ITK-SNAP [30], and the results show that Fp^{roi} -GAN method has less noise points than the other synthesis

TABLE 1: Quantitative evaluation of the INbreast dataset (mean \pm standard deviation). We compared the measurements of the different synthesis methods over the whole image domain and the tumor domain at a significance level of 0.05, and the underline indicates that Fp^{roi} -GAN is statistically significantly different from other methods.

Region	Methods	PSNR	SSIM	MS-SSIM
Whole image	DCGAN [11]	16.834 \pm 3.28	0.769 \pm 0.15	0.879 \pm 0.21
	Pix2Pix [15]	17.398 \pm 3.81	0.843 \pm 0.13	0.923 \pm 0.19
	cycleGAN [16]	17.815 \pm 5.18	0.829 \pm 0.17	0.919 \pm 0.18
	Fp^{roi}-GAN	18.647 \pm 3.25	0.882 \pm 0.16	0.935 \pm 0.15
Tumor region	DCGAN [11]	19.231 \pm 7.43	0.872 \pm 0.15	0.894 \pm 0.23
	Pix2Pix [15]	21.811 \pm 6.98	0.915 \pm 0.11	0.902 \pm 0.22
	cycleGAN [16]	20.485 \pm 6.15	0.882 \pm 0.07	0.904 \pm 0.18
	Fp^{roi}-GAN	24.142 \pm 6.70	0.967 \pm 0.08	0.946 \pm 0.18

TABLE 2: Results of the quantitative evaluation of the ROI images of the INbreast dataset (mean \pm standard deviation) with a significance level of 0.05; the underline indicates that the Fp^{roi} -GAN is statistically significantly different from other methods.

Methods	Dice coefficient	Hausdorff distance
DCGAN [11]	0.827 \pm 0.25	7.31 \pm 4.95
Pix2Pix [15]	0.945 \pm 0.17	7.27 \pm 4.18
cycleGAN [16]	0.952 \pm 0.13	6.83 \pm 3.38
Fp^{roi}-GAN	0.981 \pm 0.11	4.21 \pm 2.84

methods. The results of the image distribution of the four synthesis methods are compared in Figure 8, where the histogram indicates the distribution of the image grayscale and the trend of the image grayscale. The Fp^{roi} -GAN method is always closer to the original image than the three other methods, both in terms of image distribution and trend of image grayscale.

In addition to the quantitative evaluation of the synthesized MR images, this paper supplements a glioma HGG and LGG classification experiment to verify the auxiliary effect of the synthesized MR images for the classification experiment. Considering that the INbreast dataset without mass data lacks corresponding ROI images, our synthesis method is not applicable, and the auxiliary effect on its dataset could not be verified in the classification. In the image synthesis experiments, the training set included 380 images, consisting of 190 HGG and LGG data, and the test set included 164 images, consisting of 82 HGG and LGG data. In the classification experiments, the data used for testing in the synthesis method were used as the training set with 164 images, the data used for training in the synthesis method were used as the test set with 380 images, and the data from each of the four groups synthesized images were added as comparison experiments, as shown in Table 5. Referring to article [31, 32] for the classification method, the first 500 features were extracted for each group of images by using the Resnet [33] network, followed by 30 features selected by the recursive feature elimination [34] with fivefold cross-validation, and the filtered features were classified using the kernel-based SVM algorithm [35]. The metrics used to assess the classification results include AUC, accuracy (Acc), sensitivity (Sen), and specificity (Spe), where AUC represents the area of the ROC curve and the other three metrics can be defined as (7)–(9):

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (7)$$

$$Sen = \frac{TP}{TP + FN}, \quad (8)$$

$$Spe = \frac{TN}{FP + TN}, \quad (9)$$

where TP represents the number of samples, in which HGG was correctly predicted, TN represents the number of samples, in which LGG is correctly predicted, FN represents the number of samples, in which HGG is predicted as LGG, and FP represents the number of samples, in which LGG is predicted as HGG. The experimental results in Table 5 show that, by adding the images synthesized by our method for training the machine learning model, the prediction ability of the model was effectively improved, in which Fp^{roi} -GAN achieved the best results in the four metrics, and our method achieved a high classification sensitivity of 0.913. The ROC of the classification experiments is shown in Figure 9.

3.3. Discussion. Currently, most image synthesis methods are in single-input, single-output mode, and the synthesis of ROI images is rarely studied. Our work utilizes the cycleGAN's cyclic consistency mechanism to solve the problem of structural inconsistency between real and synthetic images and improves the contrast information between ROI and non-ROI domains by incorporating regional features a priori, resulting in the synthesis of high-quality medical images as well as the corresponding ROI images. To evaluate the quality of the synthesized images, we compared several currently popular synthesis methods, such as DCGAN, Pix2Pix, and cycleGAN, and evaluated the synthesis results in terms of the whole image domain of the images and the tumor domain. The results show that the Fp^{roi} -GAN method synthesized high-quality medical images on both datasets and achieved the best results in PSNR, SSIM, MS-SSIM, dice, and Hausdorff distance metrics. The poor quality of the DCGAN synthesized images may be due to the collapse of the model during training, and we found that the synthesized images of Pix2Pix and cycleGAN are not of high quality due to the low structural consistency of the model. In addition, the comparison results from the whole and tumor domain of

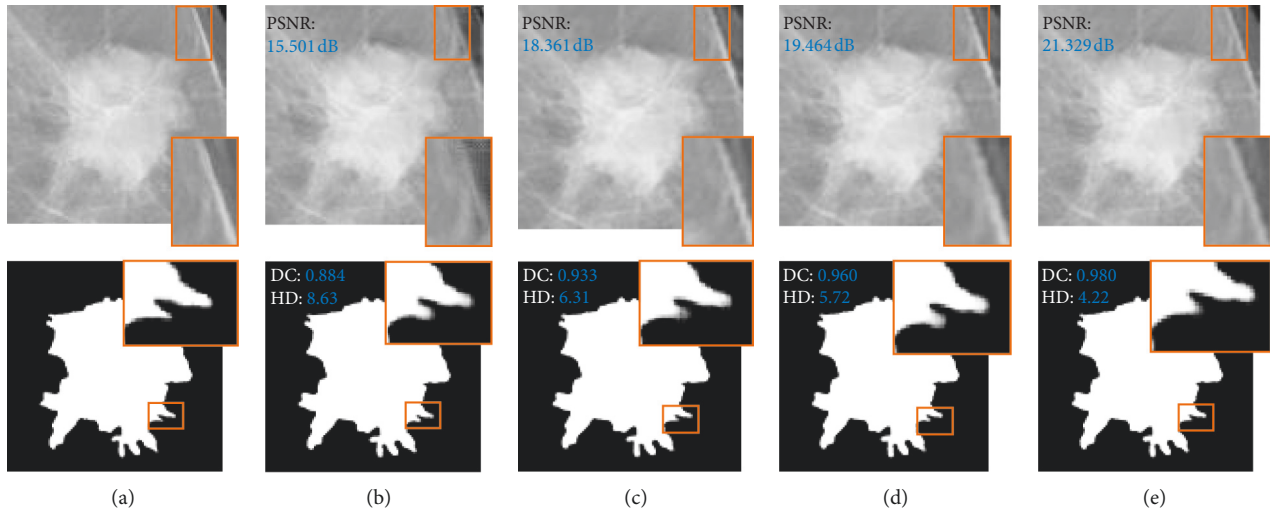


FIGURE 6: Comparison of Fp^{roi} -GAN with the three other synthesis methods on the INbreast dataset. (a) Input image. (b) DCGAN. (c) Pix2Pix. (d) cycleGAN. (e) Fp^{roi} -GAN.

TABLE 3: Results of the quantitative evaluation of the BRATS 2017 dataset (mean \pm standard deviation), where we compare the measurements of the different synthesis methods over the whole image domain and the tumor domain at a significance level of 0.05, and the underline indicates that Fp^{roi} -GAN is statistically significantly different from the other methods.

Data	Region	Methods	PSNR	SSIM	MS-SSIM
HGG	Whole image	DCGAN [11]	25.749 \pm 3.49	0.882 \pm 0.04	0.890 \pm 0.05
		Pix2Pix [15]	28.938 \pm 4.68	0.952 \pm 0.03	0.956 \pm 0.05
		cycleGAN [16]	34.280 \pm 4.85	0.984 \pm 0.02	0.984 \pm 0.05
		Fp^{roi}-GAN	34.884 \pm 5.18	0.986 \pm 0.02	0.987 \pm 0.04
	Tumor region	DCGAN [11]	29.539 \pm 5.05	0.903 \pm 0.02	0.910 \pm 0.05
		Pix2Pix [15]	33.031 \pm 5.99	0.951 \pm 0.02	0.952 \pm 0.04
		Fp^{roi}-GAN	41.888 \pm 6.06	0.997 \pm 0.004	0.993 \pm 0.03
LGG	Whole image	DCGAN [11]	23.093 \pm 4.71	0.895 \pm 0.11	0.908 \pm 0.06
		Pix2Pix [15]	25.912 \pm 4.95	0.933 \pm 0.09	0.945 \pm 0.07
		cycleGAN [16]	28.045 \pm 4.47	0.958 \pm 0.08	0.966 \pm 0.03
		Fp^{roi}-GAN	30.044 \pm 4.21	0.964 \pm 0.08	0.974 \pm 0.03
	Tumor region	DCGAN [11]	25.809 \pm 4.39	0.892 \pm 0.09	0.911 \pm 0.07
		Pix2Pix [15]	30.228 \pm 5.28	0.939 \pm 0.08	0.948 \pm 0.07
		Fp^{roi}-GAN	40.440 \pm 7.51	0.997 \pm 0.02	0.990 \pm 0.03

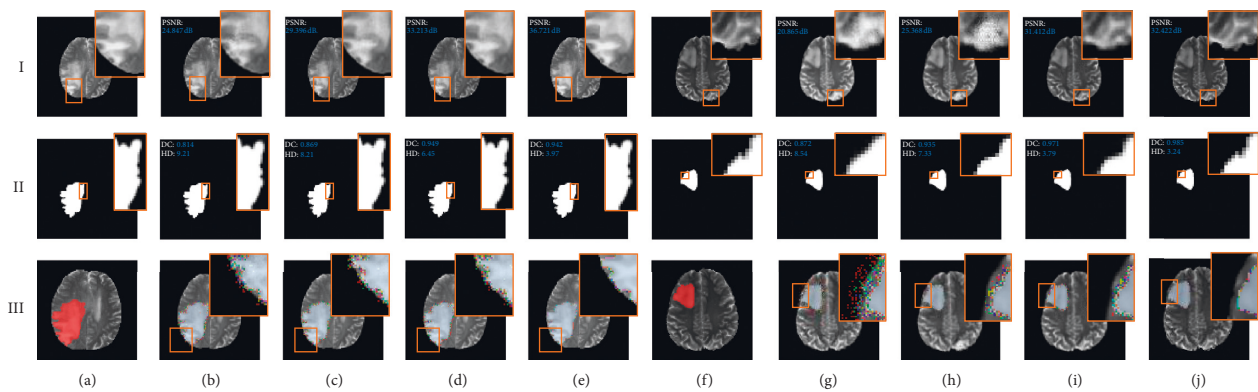


FIGURE 7: Comparison of Fp^{roi} -GAN with the three other synthesis methods on the BRATS 2017 dataset, where III is the visual performance in ITK-SNAP. (a) Input image (HGG). (b) DCGAN. (c) Pix2Pix. (d) cycleGAN. (e) Fp^{roi} -GAN. (f) Input image (LGG). (g) DCGAN. (h) Pix2Pix. (i) cycleGAN. (j) Fp^{roi} -GAN.

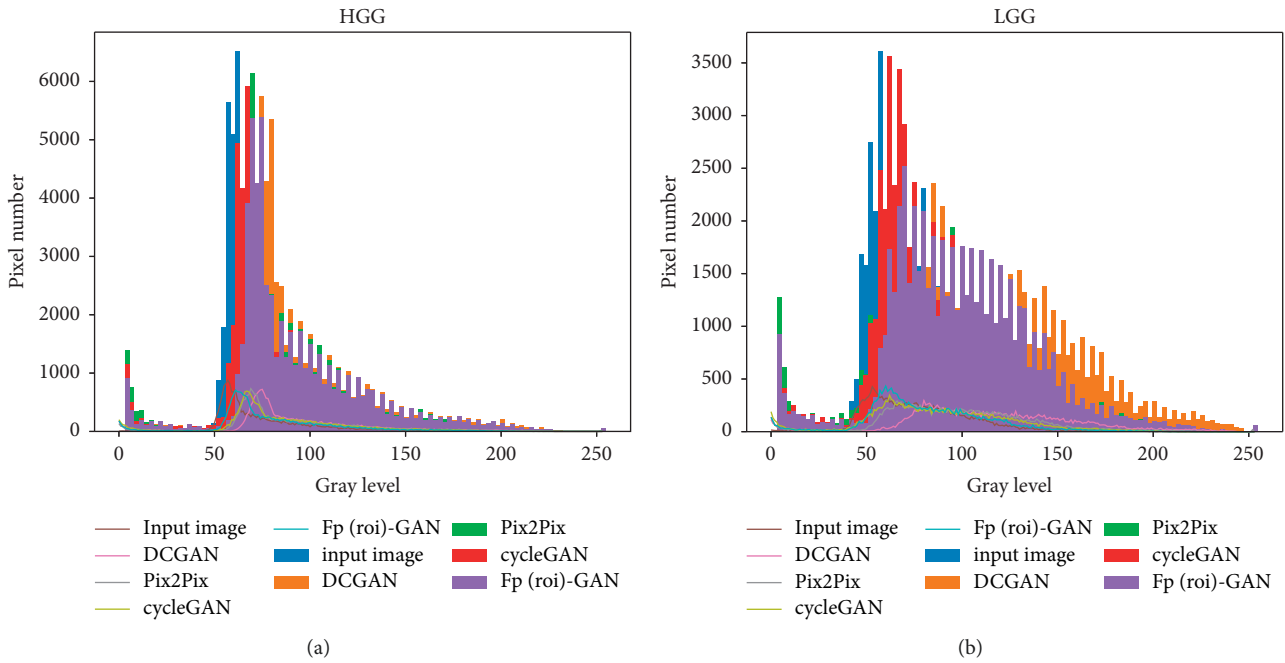


FIGURE 8: Image distribution results and grayscale trends of the four synthesis methods under HGG and LGG, where Fp(roi)-GAN represents Fp^{roi}-GAN. (a) HGG. (b) LGG.

TABLE 5: Classification results.

Data	Methods	AUC	Acc	Sen	Spe
BRATS2017		0.872	0.789	0.823	0.778
BRATS2017 + DCGAN		0.881	0.803	0.720	0.831
BRATS2017 + Pix2Pix	Resnet + SVM	0.894	0.815	0.857	0.855
BRATS2017 + cycleGAN		0.928	0.855	0.910	0.843
BRATS2017 + Fp ^{roi} -GAN		0.943	0.882	0.913	0.868

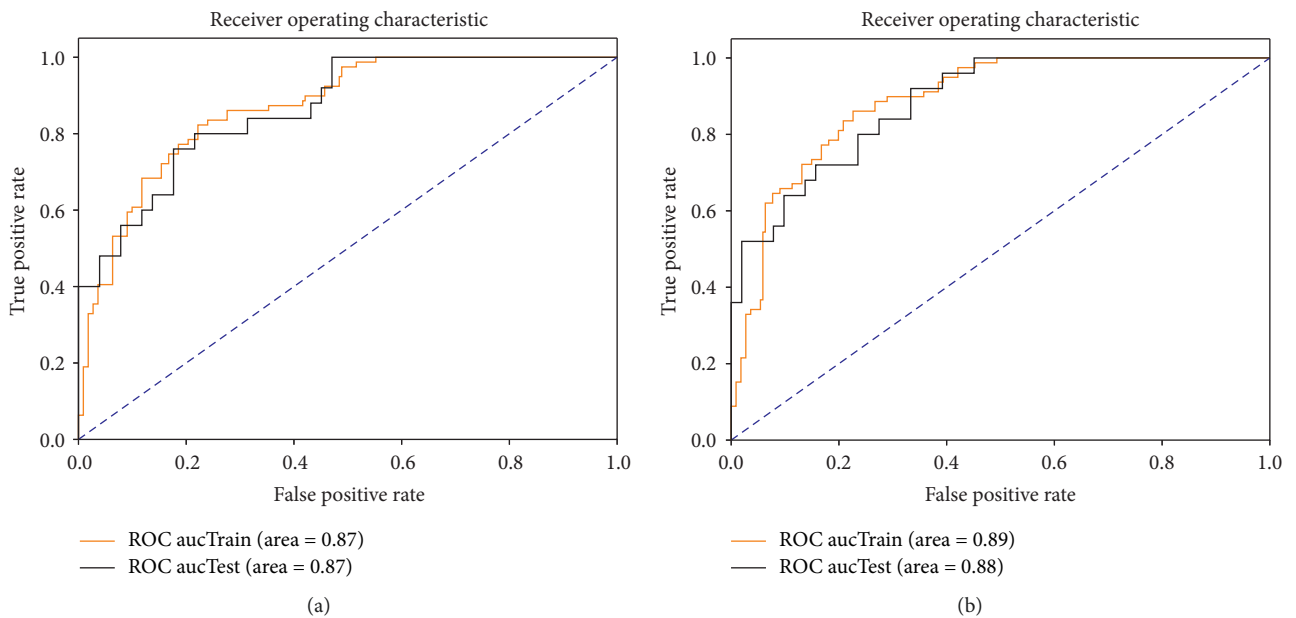


FIGURE 9: Continued.

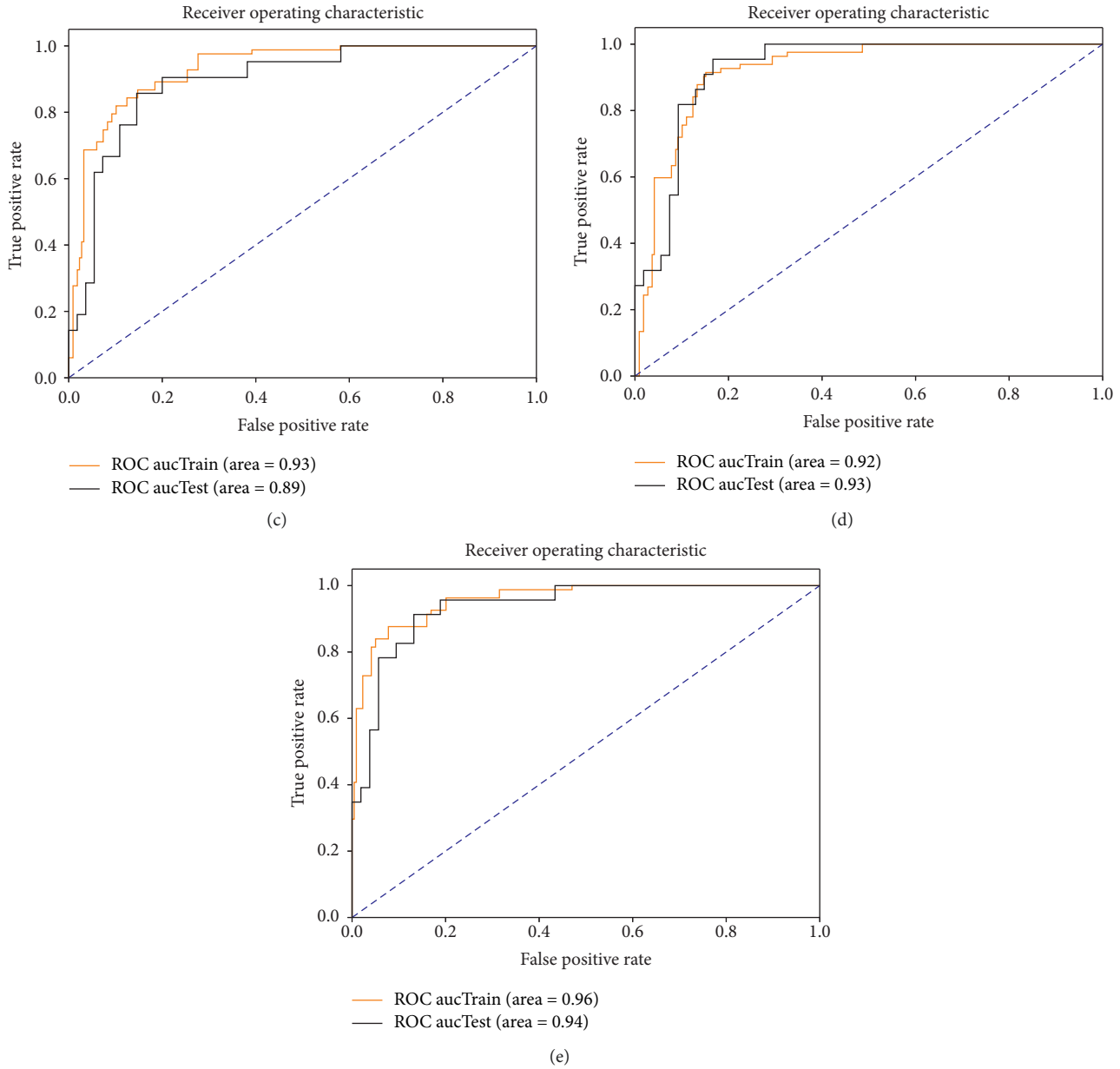


FIGURE 9: ROC plot of the classification experiment. (a) BRATS 2017. (b) BRATS 2017 + DCGAN. (c) BRATS 2017 + Pix2Pix. (d) BRATS 2017 + cycleGAN. (e) BRATS 2017 + Fp^{roi} -GAN.

the images showed that the tumor domain is more informative than the whole domain by incorporating regional features.

As shown in Figure 7, our synthesis method resulted in the least noise points in the medical image processing tool ITK-SNAP, but the images generated by DCGAN contain more noise points. Based on the image distribution and grayscale change trend in Figure 8, the proposed method is closest to the distribution of the original image. From the previously mentioned experimental evaluation results, the sets of experiments show that the method proposed in this paper is more likely to be applied in the near future research of medical images. At last, in the BRATS 2017 classification experiments, we supplemented the synthesized data into the

training set to effectively assist the training of the machine learning model and improve the classification effect of the model, in which the highest classification accuracy was achieved by adding the data synthesized by the Fp^{roi} -GAN method. Although many adversarial generation models have been proposed, the quality of the generated images has been an important goal for researchers to pay attention to, and in addition, whether the generated images can be used in recent studies is also a key concern for research. In this paper, our proposed method generates high-quality images and is validated in brain glioma classification experiments, which proximately illustrates the feasibility and superiority of our proposed generation method in the process of medical imaging research.

4. Conclusions

GAN is widely studied in the field of medical imaging, including cross-modal synthesis, super-resolution reconstruction, and medical image denoising. In this paper, we proposed the Fp^{roi}-GAN method to synthesize paired medical images. Moreover, we validated the results of the synthesized images via quantitative analysis, image distribution comparison, and visual evaluation. In the BRATS experiment, we added a classification experiment to verify the effect of synthesized data on the classification experiment. The results show that the addition of synthetic images effectively assisted the training of the machine learning model and improved the classification performance of the prediction model. Although this paper does not further validate the impact of the synthesized ROI images on the segmentation problem, the quantitative analysis indicated that our method has higher quantitative evaluation results than the other synthesis methods. In the future, we will further determine the effect of synthetic images on tasks, such as medical image classification and segmentation.

Data Availability

The datasets used in this paper are public dataset BRATS2017 and public dataset INbreast. BRATS2017 can be obtained through the following URL: <https://www.med.upenn.edu/sbia/brats2017/data.html>, and INbreast can be obtained through the following URL: <http://medicalresearch.inescporto.pt/breastresearch>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant no. 81772009 and Collaborative Innovation Major Project of Zhengzhou under grant no. 20XTZX06013.

References

- [1] Z. Zhou, Y. Wang, Y. Guo, Y. Qi, and J. Yu, "Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 298–311, 2020.
- [2] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: data augmentation using auxiliary classifier GAN for improved covid-19 detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020.
- [3] B. Yu, Y. Wang, L. Wang, D. Shen, and L. Zhou, "Medical image synthesis via deep learning," *Advances in Experimental Medicine and Biology*, vol. 1213, pp. 23–44, 2020.
- [4] C. Galván Casas, A. Català, G. Carretero Hernández et al., "Classification of the cutaneous manifestations of COVID-19: a rapid prospective nationwide consensus study in Spain with 375 cases," *British Journal of Dermatology*, vol. 183, no. 1, pp. 71–77, 2020.
- [5] N. Wang, S. H. Ma, J. Y. Li, Y. P. Zhang, and L. F. Zhang, "Multistage attention network for image inpainting," *Pattern Recognition*, vol. 106, p. 107448, 2020.
- [6] L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan, and J. Paisley, "An adversarial learning approach to medical image synthesis for lesion detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2303–2314, 2020.
- [7] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.
- [8] A. Chatsias, T. Joyce, M. V. Giuffrida, and S. A. Tsiftaris, "Multimodal MR synthesis via modality-invariant latent representation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 803–814, 2018.
- [9] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije et al., "Deep MR to CT synthesis using unpaired data," in *Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 14–23, Québec City, Canada, September 2017.
- [10] X. Xiao, J. Zhao, Y. Qiang et al., "Radiomics-guided GAN for segmentation of liver tumor without contrast agents," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention-Miccai*, pp. 237–245, Shenzhen, China, October 2019.
- [11] T. Fujioka, M. Mori, K. Kubota et al., "Breast ultrasound image synthesis using deep convolutional generative adversarial networks," *Diagnostics*, vol. 9, no. 4, p. 176, 2019.
- [12] D. Nie, R. Trullo, J. Lian et al., "Medical image synthesis with context-aware generative adversarial networks," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 417–425, Québec City, Canada, September 2017.
- [13] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2017, <http://arxiv.org/abs/1709.01872>.
- [14] H. Kang, J.-S. Park, K. Cho, and D.-Y. Kang, "Visual and quantitative evaluation of amyloid brain PET image synthesis with generative adversarial network," *Applied Sciences-Basel*, vol. 10, no. 7, p. 2628, 2020.
- [15] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5787–5796, Honolulu, HI, USA, July 2017.
- [16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "UNPAIRED image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision*, pp. 2242–2251, Venice, Italy, October 2017.
- [17] I. Domingues, P. H. Abreu, and J. Santos, "BI-RADS classification of breast cancer: a new pre-processing pipeline for deep models training," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing*, pp. 1378–1382, Athens, Greece, October 2018.
- [18] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012.
- [19] N. K. Batmanghelich, B. Taskar, and C. Davatzikos, "Generative-discriminative basis learning for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 31, no. 1, pp. 51–69, 2012.

- [20] B. H. Menze, A. Jakab, S. Bauer et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *Ieee Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [21] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.
- [22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [23] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1398–1402, Pacific Grove, CA, USA, November 2003.
- [24] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.
- [25] N. Ibtehaz and M. S. Rahman, "MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [26] Y. Kong, J. Wu, G. Yang et al., "Iterative spatial fuzzy clustering for 3D brain magnetic resonance image supervoxel segmentation," *Journal of Neuroscience Methods*, vol. 311, pp. 17–27, 2019.
- [27] H. Berhane, M. Scott, M. Elbaz et al., "Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning," *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 2204–2218, 2020.
- [28] D. Kingma and J. J. C. E. Ba, "Adam: a method for stochastic optimization," 2014, <http://arxiv.org/abs/1412.6980v9>.
- [29] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "EaGANs: edge-aware generative adversarial networks for cross-modality MR image synthesis," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.
- [30] P. A. Yushkevich, Y. Gao, and G. Gerig, "ITK-SNAP: an interactive tool for semi-automatic segmentation of multimodality biomedical images," in *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3342–3345, Orlando, FL, USA, August 2016.
- [31] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Systems with Applications*, vol. 164, p. 114054, 2021.
- [32] X. Yu, N. Zeng, S. Liu, and Y.-D. Zhang, "Utilization of DenseNet201 for diagnosis of breast abnormality," *Machine Vision and Applications*, vol. 30, no. 7-8, pp. 1135–1144, 2019.
- [33] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.
- [34] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [35] Z. Liu, Z. Li, J. Qu et al., "Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study," *Clinical Cancer Research*, vol. 25, no. 12, pp. 3538–3547, 2019.