RESEARCH ARTICLE

Data augmentation based on conditional generative adversarial networks for lesion classification in ultrasound images

Lina Cai[1], Zhenghua Zhang[2], Qingkai Li[3], Lun Zhang[1,*]

[1]Yunnan Vocational Institute of Energy Technology, Qujing, Yunnan, China. [2]Medical Imaging Department, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China. [3]Department of Radiology, First Hospital of Qujing, Qujing, Yunnan, China.

Received: February 5, 2024; **accepted:** April 10, 2024.

Ultrasound imaging is widely used in clinical diagnoses because of its nonionizing radiation, low cost, and noninvasive operation. However, making a diagnosis based on ultrasound images is a labor-intensive process. An accurate lesion classification system can thus be used to assist doctors in making diagnoses. The performance of classification algorithms typically improves when they are trained on large, labeled datasets. However, collecting labeled data is an expensive and time-consuming task. Therefore, performing lesion classification *via* ultrasound images is still challenging due to the small number of available training samples. To address this issue, a data augmentation method for ultrasound images based on a conditional generative adversarial network was proposed in this study to perform lesion classification. A real image was input into the generative adversarial network to constrain the mapping between the images. Then, the data augmentation process based on the conditional generative adversarial network generated the corresponding segmentation masks by category. Considering that the data augmentation method based on affine transformation can generate only fake ultrasound images or segmentation masks separately, this study proposed to use image-to-image translation to generate fake ultrasound images from the corresponding segmentation masks. The ResNet-50 was used to classify benign and malignant lesions to validate the effectiveness of the proposed approach. The results showed that, by comparing to the traditional data augmentation method based on affine transformation in terms of four evaluation metrics, the average performances of the proposed method increased by approximately 13.05% and 12.85% for the classification of lesions in the segmented masks and ultrasound images of lymph nodes and breasts, respectively. The results suggested that the proposed method could realize the purpose of data augmentation and greatly improve the classification performance.

Keywords: ultrasound image; data augmentation; generative adversarial network; lesion classification.

***Corresponding author:** Lun Zhang, Yunnan Vocational Institute of Energy Technology, Qujing, Yunnan 655001, China. Email: zhanglun1104@163.com.

Introduction

Ultrasound imaging is widely used in clinical diagnoses because of its nonionizing radiation, low cost, and noninvasive and comfortable operation, but the diagnostic results depend primarily on the ability of the doctor to interpret ultrasound images, and making a diagnosis based on ultrasound images is a labor-intensive process. An accurate, reliable, and effective lesion classification system can thus be used to assist doctors in making diagnoses. The conventional lesion classification system utilizes machine learning (ML) to analyze the features of ultrasound images and then uses a classifier to diagnose diseases. The key foundation of this application is the extraction of features from ultrasound images, but this process needs to be completed by medical experts in the relevant research field [1]. Accordingly, it is quite challenging for nonmedical experts to conduct relevant research in the field of ultrasound imaging *via* ML.

The rapid development of deep learning has resulted in its widespread use for classifying lesions in medical images. These lesion classification algorithms based on deep learning can extract additional medical image features through convolutional neural networks (CNNs) [2]. Nonetheless, due to the poor quality of ultrasound images and the inhomogeneity of human tissues, performing lesion classification on ultrasound images is relatively difficult. Although the signal-to-noise ratios, contrast levels, and denoising effects of ultrasound images have improved in recent years, it is still challenging to classify lesions in ultrasound images. One important reason for this challenge is that the number of available training samples is small. Due to the confidentiality of patient privacy, collecting many ultrasound images is difficult. Furthermore, it takes much time and effort for radiologists to annotate ultrasound images [3]. To solve this problem, several scholars have proposed data augmentation methods. These techniques have proven to be very effective and are widely used in the field of medical image analysis. The existing image data augmentation methods can be divided into two broad categories including traditional methods based on affine transformation and black-box methods based on deep learning [4]. Traditional data augmentation methods for medical images use a combination of affine transformations including flipping, rotation, zooming in/out, reflection, and shearing to generate fake image samples. The original images and generated fake images are subsequently used simultaneously as the model datasets. The parameters of the affine transformation process can be preset or random. Affine transformation is conducted to generate fake images with the same semantic information as that contained in the associated real images [5]. To a certain extent, performing data augmentation on medical images based on affine transformation improves the learning efficiency of CNNs and prevents overfitting. However, different types of medical images require different affine transformation parameters, which makes the selected parameters dependent on researchers' experience. Moreover, performing data augmentation on medical images based on affine transformation results in only small changes in the shapes of the images, which does not align with the disparity of medical images. Utilizing larger affine transformation parameters to create greater changes will substantially change the contents of medical images and negate the purpose of image data augmentation, thus having the opposite effect. For ultrasound images with complex imaging textures, the fake samples generated by affine transformation cannot exhibit the characteristics needed for clinical diagnoses [6]. Data augmentation technologies based on generative adversarial networks (GANs) are also widely used in the medical imaging field. GANs, inspired by game theory, are popular sophisticated data augmentation models that learn rich features and model high-dimensional data distributions. A GAN consists of a pair of competing networks. The generative network G generates fake image samples, and the discriminative network D continuously distinguishes between the real and fake images until the GAN generates high-quality synthetic images. Therefore, the use of a GAN has emerged as an effective method for addressing the challenges posed by limited medical image samples. Saman *et al*. proposed a data augmentation approach using a GAN to detect pneumonia in chest X-ray images, showing that the proposed method could be used to effectively improve the accuracy of disease diagnoses [7]. Javaid *et al*. used a GAN to extract the features of CT images to generate new fake CT images. The generated fake samples included the local and global features of the real CT images

[8]. Perez *et al*. explored and compared several common data augmentation methods and proposed a new data augmentation method based on deep learning to achieve improved classification accuracy [9]. In addition, Qin *et al*. presented a classification technique for skin lesions based on the data augmentation mechanism of a GAN [10].

Deep learning can extract the features of ultrasound images through successive convolution. However, these algorithms based on CNNs cannot achieve good accuracy in ultrasound image lesion classification scenarios. One important cause of this issue is the small number of sample ultrasound images used for training the algorithms. To solve this problem, several scholars have proposed methods based on data augmentation for classifying lesions in ultrasound images. Breast cancer is characterized by a malignant tumor formed by the abnormal division of breast ducts or lobules. Early diagnosis and screening can effectively reduce the mortality of breast cancer [11]. Clinically, the best screening method is pathological biopsy or ultrasound imaging. Al-Dhabyani *et al*. achieved improved accuracy by integrating traditional methods with GAN-based augmentation for the classification and detection of breast cancer in ultrasound images [12]. Gheshlaghi *et al*. increased the size of a small dataset by using an auxiliary GAN that generated fake images with their class labels and evaluated the effectiveness of data augmentation by performing lesion classification on breast ultrasound images using deep convolutional neural networks [13]. Wu *et al*. trained a class-conditional GAN to conduct contextual filling and generate lesions on healthy screening mammograms, and experimentally evaluated the use of data augmentation to achieve improved breast cancer classification performance [14]. Changes in cervical lymph nodes are common surgical manifestations and include various lymph node diseases such as lymph node-reactive hyperplasia, lymphoma, and metastasis of various neoplastic lesions. Determining the nature of lymph nodes is the most important step in the clinical diagnosis process. Puncturing a lymph node is a minimally invasive diagnostic method, but a convolutional neural network allows the possibility of a painless and rapid lymph node diagnosis [15]. Tekchandani *et al*. proposed a data augmentation approach based on a GAN and used the inception network to make benign and malignant diagnoses of mediastinal lymph nodes [16]. Wang *et al*. proposed a data augmentation method to maintain the original image resolution and retain the most informative parts of images [17]. However, these data augmentation methods can only generate original images or segmentation masks separately, and the performance improvement provided by these techniques in classification tasks is limited.

To alleviate the above problems, a data augmentation method for ultrasound images based on a conditional generative adversarial network (cGAN) was proposed in this study to perform lesion classification. A cGAN was exploited to generate segmentation masks by categories under constraint mapping. Considering that data augmentation based on affine transformation can generate only fake ultrasound images or segmentation masks separately, the image-to-image translation was utilized to generate fake ultrasound images from the corresponding segmentation masks. The study expected that the proposed data augmentation method for ultrasound images could achieve better lesion classification performance than that of traditional augmentation methods for ultrasound images or segmentation masks.

Materials and methods

Data augmentation of ultrasound image segmentation masks based on a cGAN

The standard GAN uses random noise as its input to generate fake samples that are indistinguishable from real samples. It includes two neural networks. Generative network G is used to generate fake samples, and discriminative network D is used to distinguish

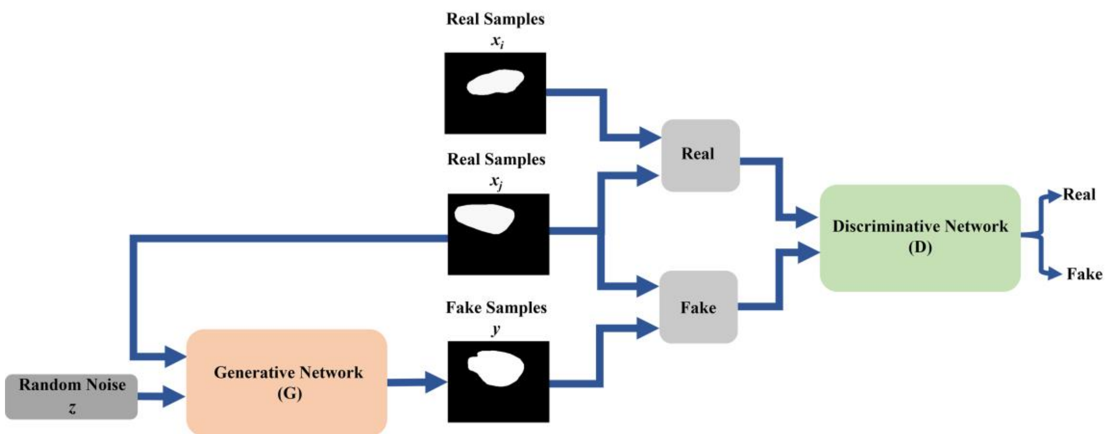

Figure 1. The architecture of ultrasound image segmentation mask data augmentation based on the cGAN.

between the fake samples and real samples. These two networks compete with each other. So that the generative network G can generate samples that the discriminative network D cannot distinguish. The standard GAN uses unsupervised learning and thus cannot generate fake samples based on their categories. In other words, a GAN cannot control the models of the samples to be generated. Rather, it can only learn mappings from random noise to the samples. To effectively solve this problem, a cGAN that added real samples as inputs was proposed in this study to enable a targeted method of generating samples based on given categories. Similarly, the cGAN was composed of a generative network G and discriminative network D. During the training process, the cGAN used the antagonistic training mechanism of the generative network and discriminative network to achieve the joint optimization of both networks until the equilibrium state was reached [18]. The goal of the generative network G was to generate new fake samples, and the discriminative network D was employed to distinguish between the generated fake samples and the real samples. The architecture of the ultrasound image and segmentation mask data augmentation method based on a cGAN was shown in Figure 1. The inputs of the cGAN included random noise z and real samples x_j. The generated fake samples were denoted as y, and the input x_j was used to constrain the mapping between the fake samples and real samples. The mapping between the input and output was G: $\{x, z\} \rightarrow y$. Its objective function was as follows:

$$\mathcal{L}_{cGAN}(G,D) = E_{x,y}[\log D(x,y)] + E_{x,z}\left[\log\left(1 - D\left(x, G(x,z)\right)\right)\right] \quad (1)$$

During the antagonistic training process of the cGAN, the generative network G tried to minimize the objective function, while the discriminative network D tried to maximize the objective function. This scheme could be represented by the following function:

$$\mathcal{L}^*_{cGAN} = arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) \quad (2)$$

The main function of the generative network was to generate corresponding samples by category under constraint mapping. Its structure was shown in Figure 2. Based on the traditional structure of the encoder-decoder network, the generative network was composed of a pair of symmetric contracting and expanding paths. The real samples x_j constrained the mapping between the generative fake samples and real samples. These samples were taken as the inputs of the contracting paths and then combined with the feature mapping of random noise as the input of the expanding paths. In the architecture of the generative network, each encoder module consisted of three convolutional modules and a

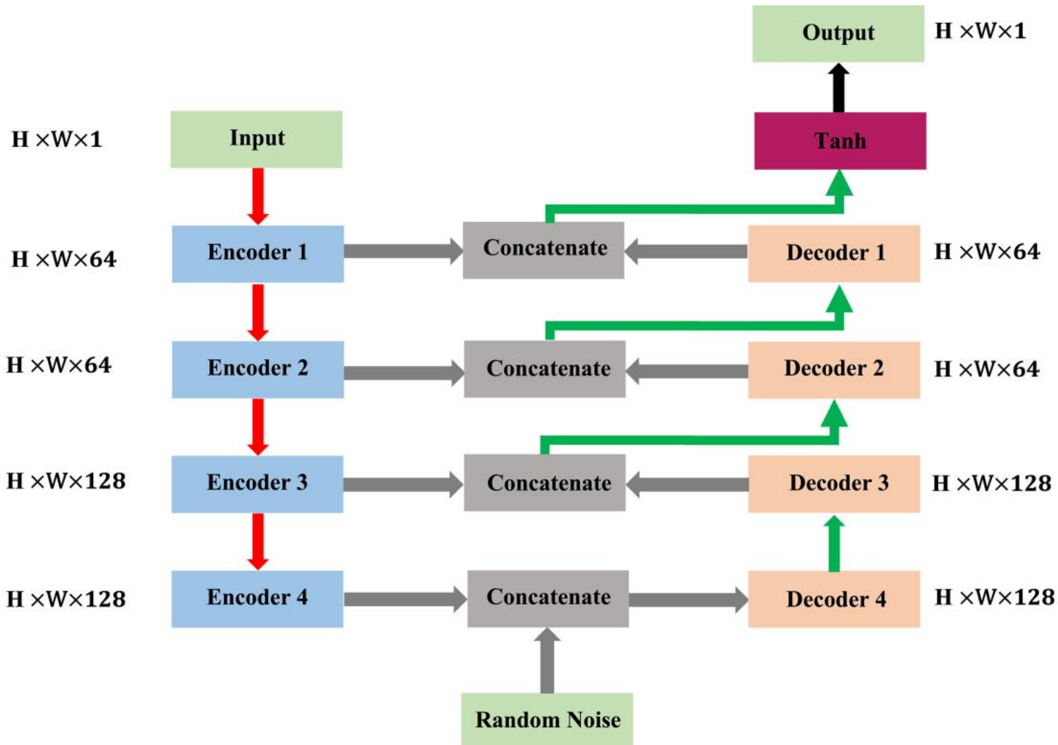

Figure 2. The architecture of the generative network.

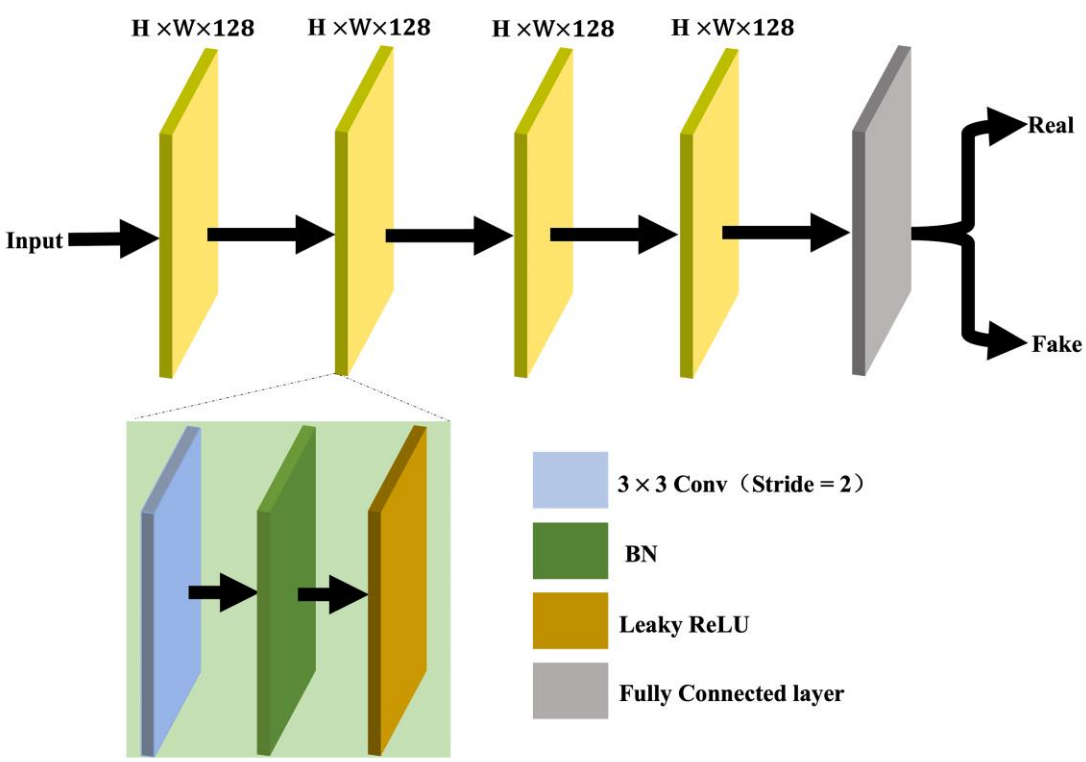

Figure 3. The architecture of the discriminative network.

downsampling module. Each convolutional module consisted of a successive 3 × 3 convolutional layer, batch normalization (BN), and a leaky ReLU function. The three convolutional modules of the encoder module used residual connections to improve the flow of information, optimize the performance of the network, and avoid gradient vanishing during training. The downsampling module consisted of successive 3 × 3 convolutional layers with strides of 2, BN, and a leaky ReLU function. The structure of the decoder module was similar to that of the encoder module. However, the difference was that the decoder module used upsampling to replace the downsampling step in the encoder module. The feature information extracted by the encoder module was simultaneously combined with the feature information extracted by the decoder module through skip connections as the inputs of the next-level decoder module. This method could effectively compensate for the loss of feature information caused by successive convolutional layers and downsampling.

The discriminative network was mainly used to distinguish between the generative fake samples and real samples. The antagonistic training process of the cGAN could prompt the generative network to generate fake samples that were closer to the real samples. Its network structure was shown in Figure 3. The discriminative network employed a feed-forward and downsampling structure, which consisted of four successive convolutional modules and a fully connected layer. Each convolutional module consisted of successive 3 × 3 convolutional layers with strides of 2, a BN operation, and a leaky ReLU function. To maintain the training stability of the discriminative network, the number of output channels in all the convolutional layers were set to 128. Finally, the fully connected layer was used to distinguish between the generative fake samples and the real samples.

Data augmentation of ultrasound images based on image-to-image translation

The data augmentation method applied to ultrasound images was more complex than that used for ultrasound image segmentation masks. Two types of data augmentation methods were available for ultrasound images. The first type included traditional methods based on several affine transformations including flipping, rotation, scaling, and translation. The other type included black-box methods based on GANs. To augment ultrasound image data based on affine transformation, fake ultrasound image samples were generated by setting fixed parameters, which involved only a tiny geometric transformation of the ultrasound images, and the images could not be transformed based on their categories. Therefore, the traditional data augmentation methods based on affine transformation had relatively narrow application fields. Due to the large variations in the lesion structures of ultrasound images, a GAN was used for augmenting ultrasound images data, which might produce inconsistent results. Therefore, image-to-image translation was used to transform ultrasound image segmentation masks into ultrasound images. The image-to-image translation process was based on the pix2pix model, which used U-Net as the generative network and PatchGAN as the discriminative network [19]. In the U-Net structure, the feature information of the symmetric encoder module and decoder module was combined to form the inputs of the next decoder module through skip connections. These connections could compensate for the loss of feature information caused by successive convolutional layers and downsampling. This method effectively improved the translatability of images. In PatchGAN, each image was divided into $N \times N$ patches. PatchGAN then tried to determine whether each patch was real or fake. The final classification result was determined by averaging the results of all the patches. The structure of the proposed data augmentation method for ultrasound images based on image-to-image translation was shown in Figure 4. The developed data augmentation method for ultrasound images based on image-to-image translation employed a type of cGAN. During the antagonistic training process of a

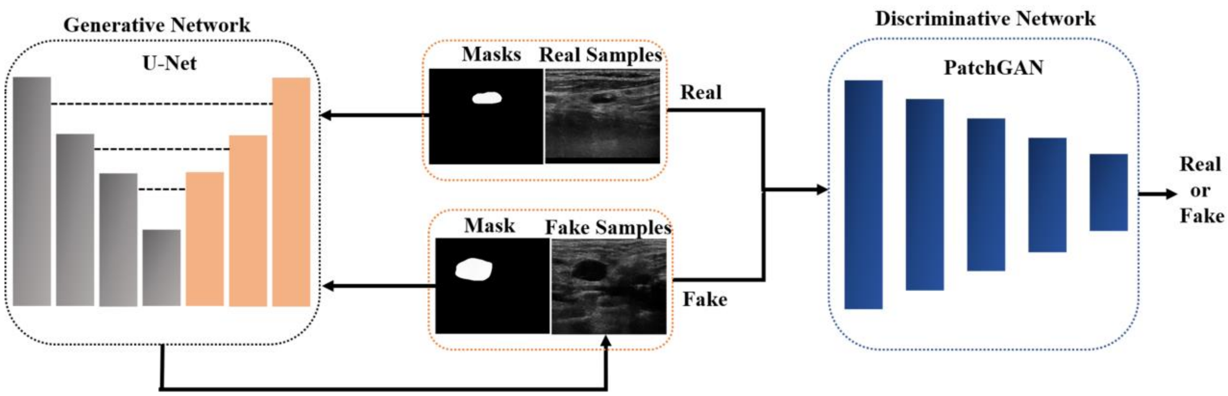

Figure 4. The architecture of image-to-image translation.

cGAN, the generative network constantly adjusted its parameters so that the discriminative network could not distinguish between the generative fake samples and real samples. The generative network was then prompted to generate fake ultrasound image samples that were closer to the real samples. The training dataset of this algorithm consisted of pairs of image samples, including ultrasound images and the corresponding ultrasound image segmentation masks. The inputs of image-to-image translation included random noise *z* and real samples *x*. The outputs of the generative network were denoted as *y*. The input *x* was used to constrain the mapping between the generative fake samples and real samples. The mapping between the input and output was defined as *G*: {*x*, *z*}→*y*. To form the objective function, the L1 loss function was added to the objective function of the cGAN to optimize the algorithm. The objective function could be expressed as follows:

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[\|y - G(x,z)\|_1] \quad (3)$$

During antagonistic training process of the cGAN, generative network G tried to minimize the objective function, while discriminative network D tried to maximize the objective function. This scheme could be expressed by the following function:

$$\mathcal{L}_c^* = arg\min_G \max_D(\mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G)) \quad (4)$$

where λ represented the weight value, and the value of λ was set to 100.

Lesion classification for ultrasound images based on ResNet-50

ResNet-50 was used for the classification of lesions in ultrasound images and the segmentation masks of ultrasound images. ResNet-50 consisted of five convolutional modules, a pooling layer, and a fully connected layer. The structure of ResNet-50 was shown in Figure 5. The convolutional module was composed of successive 7 × 7 convolutional layers with strides of 2, BN, a ReLU function, and a max-pooling layer. Among the residual convolutional modules, different modules contained different numbers of residual units. As the depth of the network increases, its performance may improve. When the depth of the network exceeded 20, the performance decreased rather than improved, which was called gradient vanishing or gradient explosion [20]. To mitigate this problem, ResNet-50 used different numbers of residual connection units. Each residual connection unit was composed of three successive 1 × 1, 3 × 3, and 1 × 1 convolutional layer and each convolutional layer was followed by BN and ReLU activation. Moreover, each residual connection unit adopted skip connections, which could degrade the deep network to a shallow network during training and optimize the network performance.

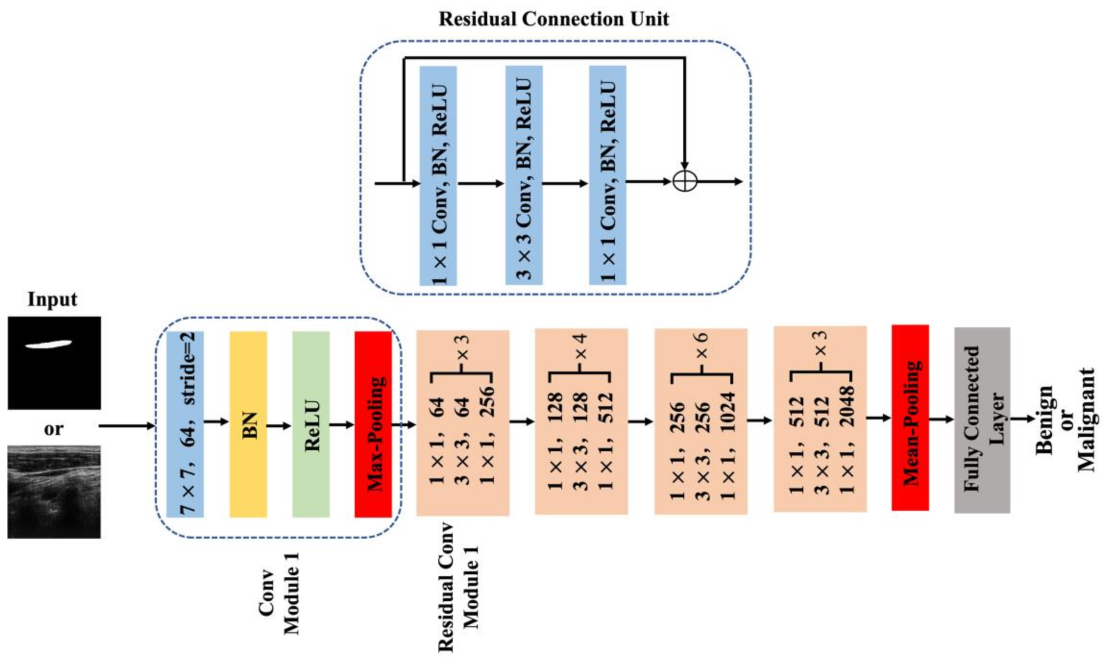

Figure 5. The architecture of ResNet-50.

Datasets selection and model construction

The datasets used in this study included segmentation masks and corresponding ultrasound images of lymph nodes and breast lesions. The lymph node dataset consisted of 143 samples acquired from Hua Shan Hospital (Shanghai, China) [21]. Among the 143 lymph node samples, 99 were benign samples and 44 were malignant samples. Twenty-four benign samples and 24 malignant samples were randomly selected to form the test datasets, and the remaining samples were used to produce the training datasets. The breast lesion ultrasound image datasets consisted of 647 samples acquired from the Kaggle challenge [22]. Of the 647 breast lesion ultrasound image samples, 437 were benign, and 210 were malignant. One hundred benign samples and 100 malignant samples were randomly selected to compose the test datasets, and the remaining samples were used to construct the training datasets. All the samples were cropped to 320 × 256. PyTorch (version 1.7.0) (https://pytorch.org/) was employed as the framework for implementing network. The training process was conducted on a workstation equipped with an NVIDIA 2080Ti graphics card possessing 11 GB of memory. The accuracy, precision, recall, and F1-score were used as evaluation indices and were defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

where TP, FP, TN, and FN denoted true positives, false positives, true negatives, and false negatives, respectively. Malignant samples were defined as positive, so the recall value was more worthy of attention than the accuracy achieved in the study. Notably, networks based on deep learning that employed the same training datasets and seed points yielded different results.

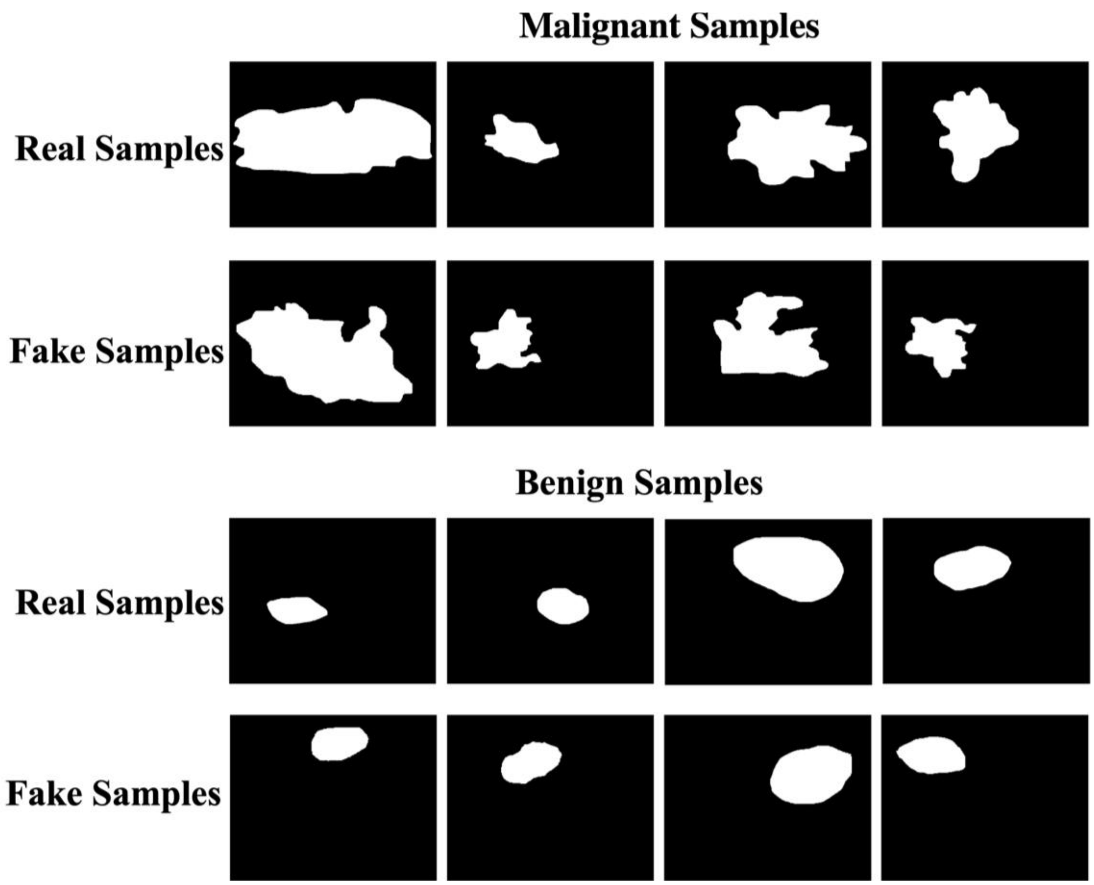

Figure 6. The samples of lymph node segmentation masks.

Therefore, the study for classifying benign and malignant lesions was repeated three times and the average values were calculated according to the test results [23]. The data augmentation process based on affine transformation included primarily horizontal and vertical flipping, 90% to 110% random scaling, and 0 to 10 degrees of random rotation.

Statistical analysis

SPSS (version 26.0) (IBM, Armonk, NY, USA) was employed for statistical analysis of this study to verify the superiority of proposed data augmentation method. Because the evaluation indices did not follow Gaussian distributions, the nonparametric Friedman test was used to evaluate the performance of proposed algorithms. A P value that was less than 0.05 indicated a significant difference.

Results and discussion

Data augmentation for segmentation masks based on the cGAN

The fake segmentation masks that were indistinguishable from real segmentation masks was generated. For both the generative network and the discriminative network, Adam was selected as the optimizer. The initial learning rate was set to 0.0001, and the batch size was set to 4 during training. The studies were conducted on the ultrasound image segmentation masks of the lymph nodes and breast lesions, during which the benign samples and malignant samples were iterated 300 times. Finally, 9,600 fake benign and malignant segmentation masks of lymph node and breast lesion ultrasound images were generated. The results showed that data augmentation process based on the cGAN could generate fake lymph node segmentation masks

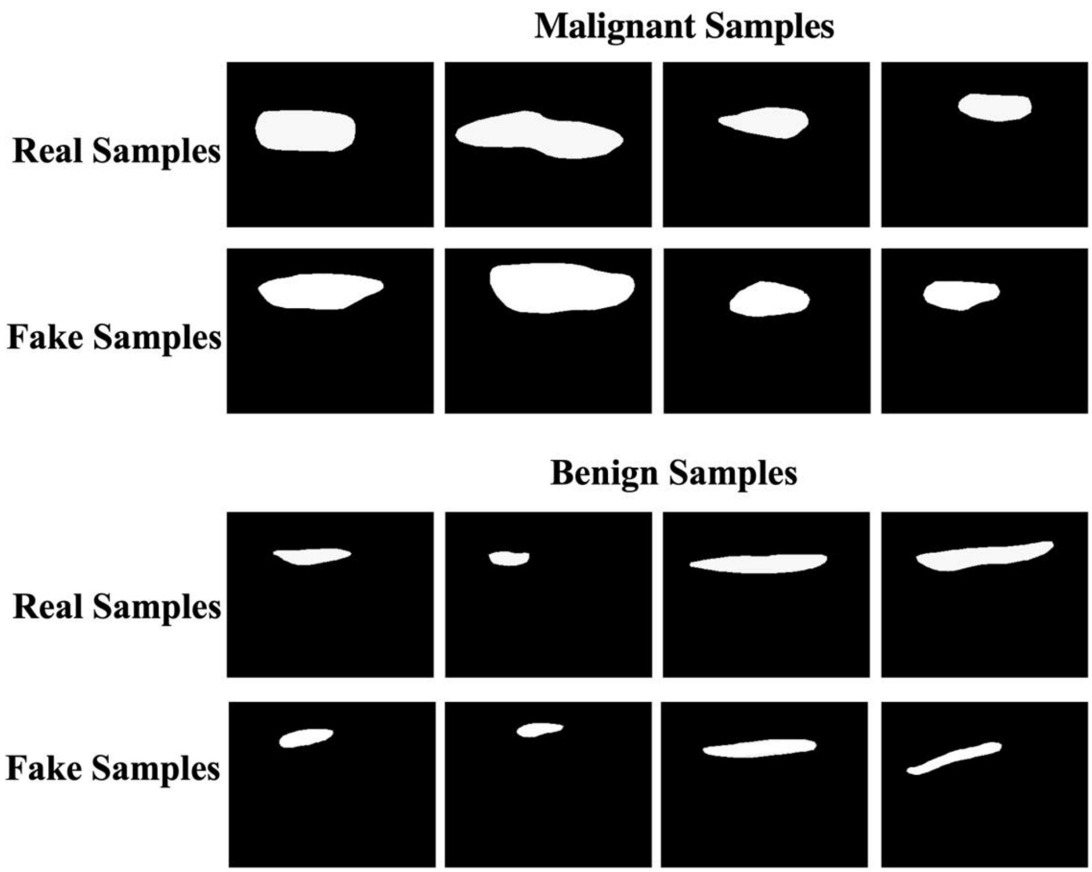

Figure7. Breast lesion segmentation mask samples.

(Figure 6) and fake breast lesion segmentation masks (Figure 7) according to the given categories under imposed the constraints.

Data augmentation for ultrasound images based on image-to-image translation

The fake ultrasound images from the corresponding segmentation masks *via* image-to-image translation was generated. The Adam optimizer was used in both the generative network and the discriminative network for image-to-image translation purposes. The learning rate was set to 0.0002, and the batch size was set to 1. The ultrasound image datasets consisting of lymph nodes and breast lesions were applied for this study, in which the benign and malignant samples were iterated 2,000 times. A total of 9,600 fake benign and malignant ultrasound images of lymph nodes and breast lesions were generated. The results showed that the network continuously refined and filled in the texture information of the ultrasound images centered on the segmentation masks and ultimately generated fake ultrasound images of lymph nodes (Figure 8). However, distinguishing real ultrasound images of lymph nodes from fake ultrasound images of lymph nodes generated by image-to-image translation using the subjective judgment mechanism of the human eyes was impossible. Moreover, the proposed method continuously refined and filled in the texture information of the ultrasound images centered on the segmentation masks and ultimately generated fake ultrasound images of breast lesions (Figure 9). The results demonstrated that it was also impossible to distinguish real breast lesion ultrasound images from fake breast lesion ultrasound images generated by image-to-image translation using the subjective judgment mechanism of the human eyes.

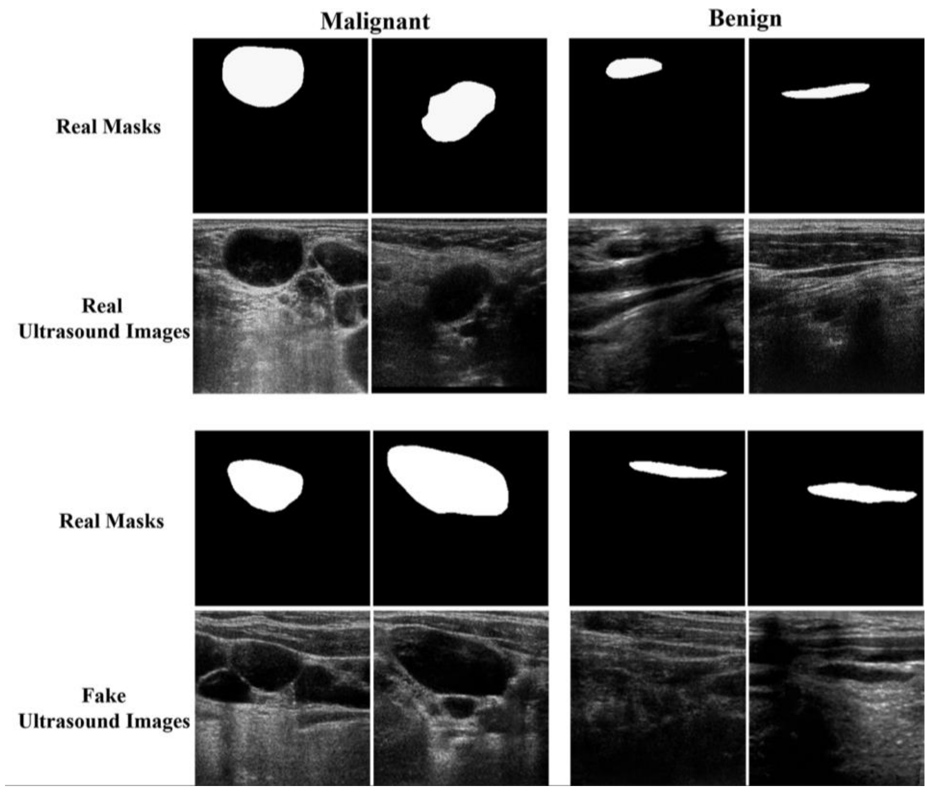

Figure 8. The samples of generative lymph node ultrasound images.

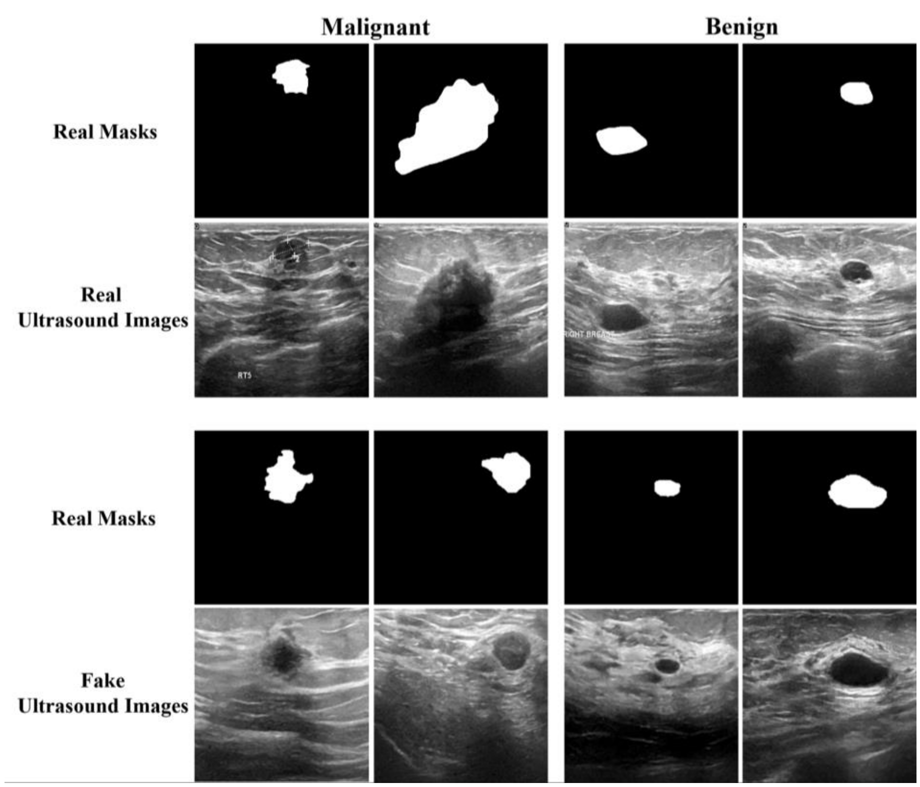

Figure 9. The samples of generative ultrasound images of breast lesion.

Table 1. The lesion classification results of segmentation masks.

Datasets	Lymph node				Breast lesion			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
TDA	79.30%	79.90%	75.35%	77.56%	92.00%	97.19%	86.50%	91.53%
DA_cGAN	95.31%	95.78%	94.79%	95.28%	93.25%	98.06%	88.25%	92.89%
O_DA_cGAN	96.88%	97.91%	95.83%	96.86%	93.58%	98.10%	89.05%	93.35%

Classifying lesions in segmentation masks based on ResNet-50

The classification of benign and malignant lesions was conducted using the segmentation mask datasets constructed in three ways including traditional data augmentation based on affine transformation (TDA), data augmentation based on a cGAN (DA_cGAN), and adding the original segmentation masks to the datasets of DA_cGAN (O_DA_cGAN). The results showed that the O_DA_cGAN dataset yielded the best lesion classification results in terms of four evaluation indices (Table 1). Compared with those obtained on the TDA lymph node dataset, the four evaluation indices of the O_DA_cGAN improved by 24.19% on average. Compared with those obtained on the TDA breast lesion dataset, the four evaluation indices achieved on O_DA_cGAN improved by 1.90% on average. With respect to the data augmentation method for segmentation masks based on affine transformation, the classification performance improved little on the lymph node dataset, while it greatly improved on the breast lesions dataset. Data augmentation based on affine transformation yielded inconsistent improvement effects on the lesion classification results obtained for different types of segmentation masks. Therefore, the data augmentation approach based on affine transformation lacked generalizability. The data augmentation method based on a cGAN could achieve data augmentation and improve the lesion classification performance achieved on segmentation masks.

Classifying lesions in ultrasound images based on ResNet-50

The benign and malignant lesion classification experiments were conducted on ultrasound images of lymph nodes and breast lesions. Distinguishing benign lesions from malignant lesions was more difficult in ultrasound images than in segmentation masks. The ultrasound image datasets were constructed in three ways including traditional data augmentation based on affine transformation (TDA), data augmentation based on a cGAN (DA_cGAN), and the addition of the original ultrasound images to the dataset generated by DA_cGAN (O_DA_cGAN). The results showed that, in the lesion classification task conducted on the lymph node ultrasound images, O_DA_cGAN yielded the best classification performance in terms of the four evaluation indices. Compared with those of the TDA lymph node dataset, the four indices of O_DA_cGAN improved by 25.04% on average. For the lymph node lesion classification task conducted on ultrasound images based on affine transformation, little improvement in the classification performance was observed, while data augmentation based on image-to-image translation yielded a great improvement. In the lesion classification task conducted on the breast ultrasound images, although the O_DA_cGAN dataset did not yield optimal results for the four evaluation indices, the recall was optimal on this dataset. Compared with those of the TDA breast lesion dataset, the four evaluation indices of O_DA_cGAN improved by 0.65% on average. For the lesion classification task involving breast ultrasound images, the O_DA_cGAN dataset was superior to the TDA and DA_cGAN datasets (Table 2).

Comparison of different datasets construction

The comparison statistical analysis of the lesion classification results based on different datasets construction methods was shown in Table 3. The results demonstrated that the higher the mean rank was, the better the performance of the

Table 2. The lesion classification results of ultrasound images.

Datasets	Lymph Node				Breast Lesion			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
TDA	76.95%	75.68%	75.31%	75.49%	70.50%	66.72%	82.00%	73.58%
DA_cGAN	93.75%	93.83%	93.75%	93.79%	70.63%	**67.53%**	80.25%	73.34%
O_DA_cGAN	94.79%	93.91%	95.83%	94.86%	70.67%	66.48%	83.67%	74.09%

Table 3. The lesion classification results mean rank.

Datasets	Mean Rank			*P* Value
	TDA	DA_cGAN	O_DA_cGAN	
Lymph Node	1.06	2.08	2.99	3.35E-6
Breast Lesion	1.38	1.88	2.75	0.021

corresponding algorithm. The proposed data augmentation method significantly improved the resulting lesion classification performance.

Conclusion

In this study, we proposed a data augmentation method for ultrasound images based on a conditional generative adversarial network for lesion classification. A cGAN was used to generate fake segmentation masks of ultrasound images by category. Then, image-to-image translation was applied to generate fake ultrasound images from the corresponding segmentation masks for ultrasound image classification. Three types of tests were performed including data augmentation involving segmentation masks based on a cGAN, data augmentation involving ultrasound images based on image-to-image translation, and benign/malignant classification based on ResNet-50. Compared with the traditional data augmentation method based on affine transformation, the proposed method exhibited average improvements of 13.05% and 12.85% in terms of four evaluation metrics, the precision, recall, accuracy, and F1-score, in lymph node and breast lesion classification tasks conducted on segmentation masks and ultrasound images, respectively. The results showed that the data augmentation method based on the cGAN generated fake segmentation masks according to the given categories under the imposed constraints. The proposed data augmentation method was based on image-to-image translation, which changed the traditional data augmentation process and improved the ultrasound images generated by segmentation masks. Moreover, this data augmentation method could be applied to other ultrasound image analysis tasks. The lesion classification study based on ResNet-50 proved that the proposed data augmentation was superior to the traditional data augmentation methods based on affine transformation. The results of this study confirmed that combining the fake samples generated by a cGAN or image-to-image translation with the original samples could yield optimal classification performance. The data augmentation method based on a cGAN exhibited better generalization and classification performance on different types of ultrasound images or segmentation masks. When attempting to achieve improved lesion classification performance for benign and malignant samples, segmentation masks outperformed ultrasound images.

Acknowledgement

The research was supported by the Science Foundation of Education Department of Yunnan

Province (Grant No. 2024J1691). The authors would like to thank Li Y for the constructive discussion during the manuscript revision.

References

1. Schmidhuber J. 2015. Deep learning in neural networks: An overview. Neural Netw. 61:85-117.
2. LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature. 521(7553):436-444.
3. Komatsu M, Sakai A, Komatsu R, Matsuoka R, Yasutomi S, Shozu K, *et al*. 2021. Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Appl Sci. 11(1):371.
4. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. 2021. A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol. 65(5):545-563.
5. Shorten C, Khoshgoftaar TM. 2019. A survey on image data augmentation for deep learning. J Big Data. 6(1):1-48.
6. Hussain Z, Gimenez F, Yi D, Rubin D. 2018. Differential data augmentation techniques for medical imaging classification tasks. AMIA Annu Symp Proc. 2017:979-984.
7. Motamed S, Rogalla P, Khalvati F. 2021. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of pneumonia and COVID-19 in chest X-ray images. Inform Med Unlocked. 27:100779.
8. Javaid U, Lee JA. 2018. Capturing variabilities from computed tomography images with generative adversarial networks. arXiv preprint arXiv:1805.11504.
9. Wang J, Perez L. 2017. The effectiveness of data augmentation in image classification using deep learning. Recognit. 11(2017):1-8.
10. Qin Z, Liu Z, Zhu P, Xue Y. 2020. A GAN-based image synthesis method for skin lesion classification. Comput Methods Programs Biomed. 195:105568.
11. Debelee TG, Schwenker F, Ibenthal A, Yohannes D. 2020. Survey of deep learning in breast cancer image analysis. Evol Syst. 11(1):143-163.
12. Al-Dhabyani W, Gomaa W, Khaled H, Aly F. 2019. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int J Adv Comput Sci Appl. 10(5):1-11.
13. Gheshlaghi SH, Nok Enoch Kan C, Ye DH. 2021. Breast cancer histopathological image classification with adversarial image synthesis. In Annu Int Conf IEEE Eng Med Biol Soc. 2021:3387-3390.
14. Wu E, Wu K, Cox D, Lotter W. 2018. Conditional infilling GANs for data augmentation in mammogram classification. Lect Notes Comput Sci. 11040:98-106.
15. Egert M, Steward JE, Sundaram CP. 2020. Machine learning and artificial intelligence in surgical fields. Indian J Surg Oncol. 11(4):573-577.
16. Tekchandani H, Verma S, Londhe N. 2020. Performance improvement of mediastinal lymph node severity detection using GAN and Inception network. Comput Methods Programs Biomed. 194:105478.
17. Wang J, Liu Q, Xie H, Yang Z, Zhou H. 2021. Boosted efficientnet: Detection of lymph node metastases in breast cancer using convolutional neural networks. Cancers. 13(4):661.
18. Antoniou A, Storkey A, Edwards H. 2017. Data augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340.
19. Isola P, Zhu JY, Zhou T, Efros AA. 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:1125-1134.
20. Wen L, Li X, Gao L. 2020. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Comput Appl. 32(10):6111-6124.
21. Zhang J, Wang Y, Shi X. 2009. An improved graph cut segmentation method for cervical lymph nodes on sonograms and its relationship with node's shape assessment. Comput Med Imaging Graph. 33 (8):602-607.
22. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. 2020. Dataset of breast ultrasound images. Data Brief. 28:104863.
23. Zhang L, Zhang J. 2022. Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss. PeerJ Comput Sci. 8:e873.