NUS
National University
of Singapore

新加坡国立大学苏州研究院
NUS (Suzhou) Research Institute

Deep learning in Medical Ultrasound Images

Submitted by

Wu Hao

U2102068

In partial satisfaction of the requirements for the
Degree of Bachelor of Engineering in Mechanical Engineering

Supervisor：Prof. Shen Lei

Session 2021/2022

12 May 2022

Content

Acknowledgement

I thank very much for my supervisor Prof. Shen Lei to give me a lot of comments and advice on my project. I also thank the selfness people on the Internet for solving my confusions about the code.

Abstract

Overfitting often occurs when training CNN models for medical image segmentation task for the reason of lacking enough raw data. Data augmentation methods are usually applied to overcome the situation for limited data. We use generative adversarial network (GAN) to generate fake dataset. We then train the CNN models (Unet and so on) with real and fake dataset and have better results. As we know, it is the first project to use pix2pixHD image-to-image translation structure to generate fake images with masks for breast ultrasound segmentation tasks. Also, we add FID score to evaluate synthesized datasets and shift and scale data augmentation method when training pix2pixHD. At last, we make another experiment and verify the possibility of using Stylegan2-ada method for generating better fake datasets in the future.

Keywords: pix2pixHD, GAN, image-to-image translation, ultrasound, data augmentation, breast cancer, Unet, Stylegan2-ada.

List of Figures and Tables

List of figures

List of tables

1. Introduction and Objectives of the Project

Breast cancer is one of the most fatal diseases among women all over the world. Every year, there are approximately 2 million new cases and 685,000 deaths worldwide [5]. Early detection by ultrasound images of breast can reduce the possibility of death and save millions of lives. Also, ultrasound is vital in detection and diagnosis of breast cancer because it is safe, low-cost, noninvasive nature, real-time display, operator comfort, and operator experience [4]. However, manual annotations of ultrasound images are time-consuming and the results highly depend on diagnostician experience. As a result, a stable, time-saving, robust, objective and intelligent diagnosis system is needed.

Deep learning is very famous recent years, which is a part of machine learning, can solve tasks like classification, segmentation, natural language processing (NLP), speech recognition and so on. In 2013, deep learning method was selected as one of the top ten breakthrough technologies. Deep learning becomes well known since the convolutional neural network AlexNet won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [3] and AlphaGo, developed by Google based on deep learning, defeated the world's top Go player Lee Sedol with a score of 4:1 in 2016. There are two types of algorithms based on deep learning: classification and regression. Regression problems refer to predicting the output of a continuous value, while Classification problem refers to predicting the output of a discrete value. The development of deep learning has now seen six main types of branches, they are: CNN, Self-encoding neural network based on multi-layer neurons including Auto encoder and Sparse Coding, deep belief network (DBN), Restricted Boltzmann Machine (RBM), Recurrent Neural Network (RNN), generative adversarial network (GAN) [6].

In the field of ultrasound, deep learning is usually used for classification, segmentation and detection tasks for it can learn the features from raw data [1]. However, before applying deep learning in ultrasound there are some difficult problems should be solved. One is ultrasound images have low resolution caused by noise and artifacts compared to CT or X-ray images. Another concern is that medical image datasets are usually very small, which is insufficient for deep learning models to train. Without enough data for deep learning structures and it is common in medical field, it is likely to overfit during training.

This project is to explore solutions for small datasets situation. We find that using GAN to generate fake ultrasound images for data augmentation is helpful for breast ultrasound image segmentation tasks.

2. Literature Review

To address the problem of lacking data for deep learning structures to train, many scholars made large quantities of researches and many methods have been proposed. Among these methods, data augmentation is the mostly used method. Data augmentation can be separated into basic augmentation techniques and GAN-based augmentation methods.

2.1 Basic augmentation techniques

Basic augmentation techniques include geometric transformations, cropping, erasing, intensity operations, noise injection and filtering [2].

Geometric transformations are the most commonly used data augmentation methods such as scaling, translation, rotation, reflection. They are simple and efficient.

Cropping just randomly sample a portion from the original image and then resize this section to the original image size. This method is often called random cropping.

Erasing means when training the model, randomly select a rectangular area of the image, and replace the pixel value of this rectangular area with a random value or an average pixel value, resulting in the effect of partial occlusion [7].

Intensity operations for data augmentation by changing the values of pixels in an image such as modifying the brightness or contrast of the image.

Noise injection means to add noise into the images and gaussian noise is the most widely used method. Gaussian noise with zero mean basically has data points in all frequencies, effectively distorting high frequency features. It also means that lower frequency components are also distorted, but your neural network can learn to exceed it. Adding the right amount of noise can enhance learning [9].

Using convolution kernel for data augmentation is called filtering. There are many kinds of filters and each of them has its function. For instance, Gaussian filtering is used to smooth the image, Laplacian filter is used to sharpen the edge of image [8].

2.2 GAN-based augmentation methods

GAN is first published in 2014 by Goodfellow using a new framework to estimate generative models via an adversarial process. In the paper, they proposed a generative model G and a discriminative model D. G learns the characters from raw data while D evaluates the probability of a sample from training data [10]. Goodfellow took an example to explain the "adversarial", the generative model can be considered as a team of criminals trying to produce fake currency and the discriminative model is considered as the police. Both of their performance will become better in this competition so that the synthesized data are indistinguishable from the real data in the end. After that, people found that it is hard for GANs to train because the training process is not stable, Martin Arjovsky proposed the Wasserstein Distance and solved that problem [11].

Since the performance of GAN is getting better, it is then applied for medical segmentation tasks and they are divided into two main aspects. One is to directly use GANs for segmentation tasks, the other is to use GANs for data augmentation of segmentation tasks. For the former, Vivek Kumar Singh employed the cGAN for X-ray breast segmentation and shape classification [12]. Anuja Negi made a new algorithm called RDA-UNET-WGAN for breast ultrasound lesion segmentation task. By applying RDA-UNET-WGAN for the segmentation task, the performance of precision is increased by 3–4%, IOU score by 6% and the F1 score by 5% [13]. For the latter, Tomoyuki Fujioka applied the DCGAN to generate realistic breast ultrasound images but without corresponding masks, which is the main difference between this project and our project [14]. Jue Jiang augmented dataset by using image translation networks to transform CT images to MRI images and achieved SOTA solution [15]. There are many other articles about using GANs for data augmentations in classification tasks, such as [36]-[41]. Alberto Montero Agudo selected StyleGAN2-ada network in fetal brain images generation for classification task [16]. Further, also based on StyleGAN2 structure, Zong Fan adopted DATASETGAN [17] and Daiqing Li put forward semanticGAN [18], which both of them can get segmented images with few labeled images and thus effectively reduce the labor cost of labeling. Christopher Bowles used PGGAN to enlarge the dataset through combining labeled and unlabeled data and shown its ability for better segmentation performance on AD dataset [33]. Haoqi Shi proposed a style-based GAN to synthesize pulmonary nodules CT images and can lead to better performance of downstream segmentation task [34]. Christopher Bowles introduced GAN model to brain segmentation task and verify the feasibility adding synthesized data to training dataset [35].

We find that there are very few articles pay attention to using GANs for augmentation in segmentation tasks, especially in ultrasound segmentation area. The appearance of our project fills the gap in the field of using image-to-image translation GAN model for ultrasound images data augmentations.

3. Experiments, Materials and Methodologies

3.1 Materials

3.1.1 Dataset

In this project, we used an open dataset called BUSI [19]. It is a dataset of breast ultrasound images. Walid Al-Dhabyani and his team collected this dataset from Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt. The dataset has 780 images in total with size of about 500×500 pixels. The images are in PNG format and all the images are divided into three categories: benign, normal, and malignant. They are used for training DL models for classification, detection and segmentation of breast cancer tasks.

3.1.2 Preprocessing

Since the normal images have no cancer, the masks of normal situation are all black. We made some modifications in our project that we only used benign and malignant images without using normal images. We also resize the raw images to 256×256 pixels and all subsequent operations use the dataset with size of 256×256 pixels, as shown in figure 1.

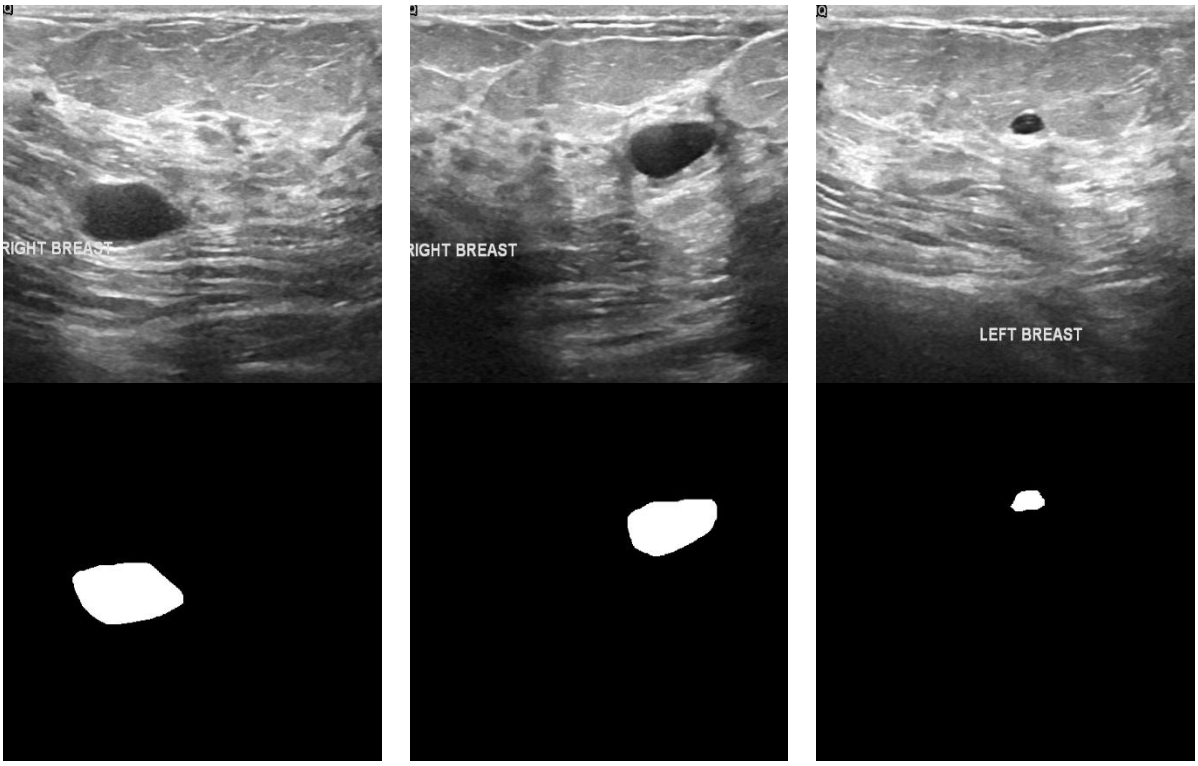


Figure 1. Three sample images of BUSI dataset

3.2 Methodologies

3.2.1 Pix2PixHD

Pix2PixHD is a cGAN model presented by Ting-Chun Wang, solving some training problems of Pix2Pix so that it can generate high resolution images [20]. Pix2PixHD and Pix2Pix both are models to solve image-to-image translation issues. As the name implies, image- to- image translation means converting images from one domain to another domain. For example, semantic labels to images, maps to aerial photos, day to night, thermal to color photos etc. Pix2Pix model first published in 2017 by Phillip Isola [21]. It used the structure that encoder-decoder network like Unet as generator and PatchGAN as discriminator, as shown in figure 2.

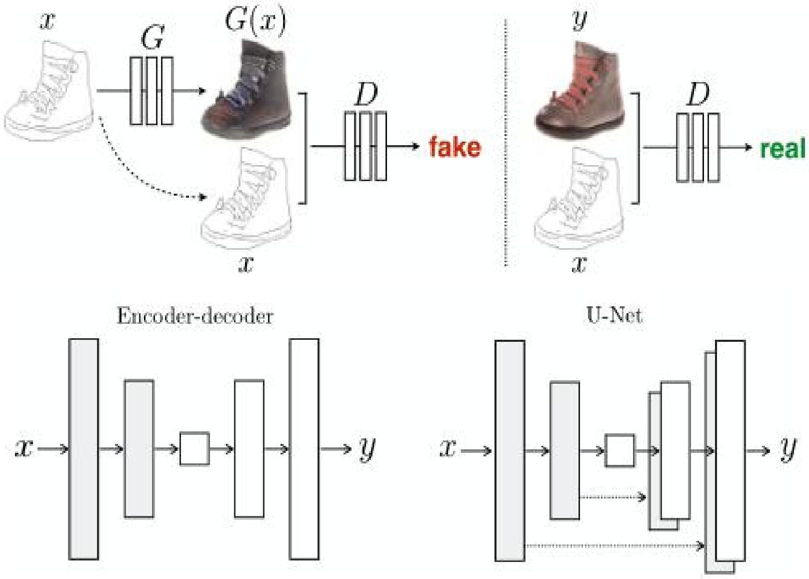


Figure 2. The structure of Pix2Pix model

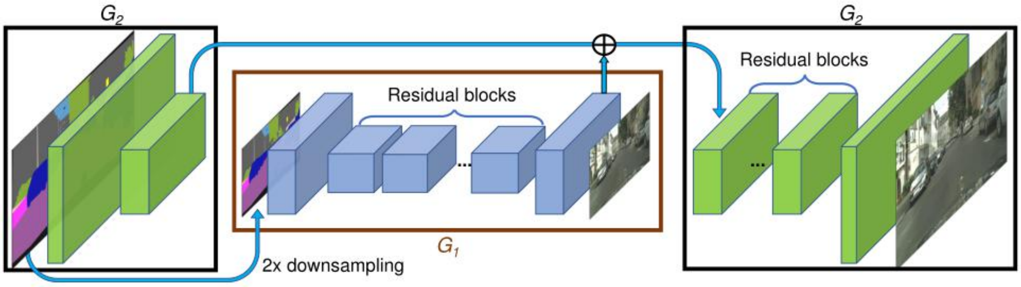


Figure 3. The structure of Pix2PixHD model

Although Pix2Pix achieved good results and made great contributions in image-to-image translation tasks, many researches find it hard to train cGAN while the results look strange in details and lack realistic textures especially on the edge. The structure of Pix2PixHD is shown in figure 3. There are two main improvements of Pix2PixHD. One is using coarse-to-fine generator, which is made up of two parts called the local G and global G, global G for basic character and local G to enhance it. The other is using improved adversarial loss by adding the feature matching loss and content loss. The objective of Pix2Pix as follows:

$$G^* = argmin_{G}\, max_{D}\, \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)\ \#(1)$$

While, the objective of Pix2PixHD is:

$$G^* = arg \min_{G} \left(\left(\max_{D1,D2,D3} \sum_{k=1,2,3} \mathcal{L}_{cGAN}(G, D_k) \right) + \lambda \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) \right) \#(2)$$

After these modifications, Pix2PixHD can generate images with high resolution and less detail errors. In our project, we use Pix2PixHD to generate fake images from masks to build a synthesized dataset, as shown in figure 4.

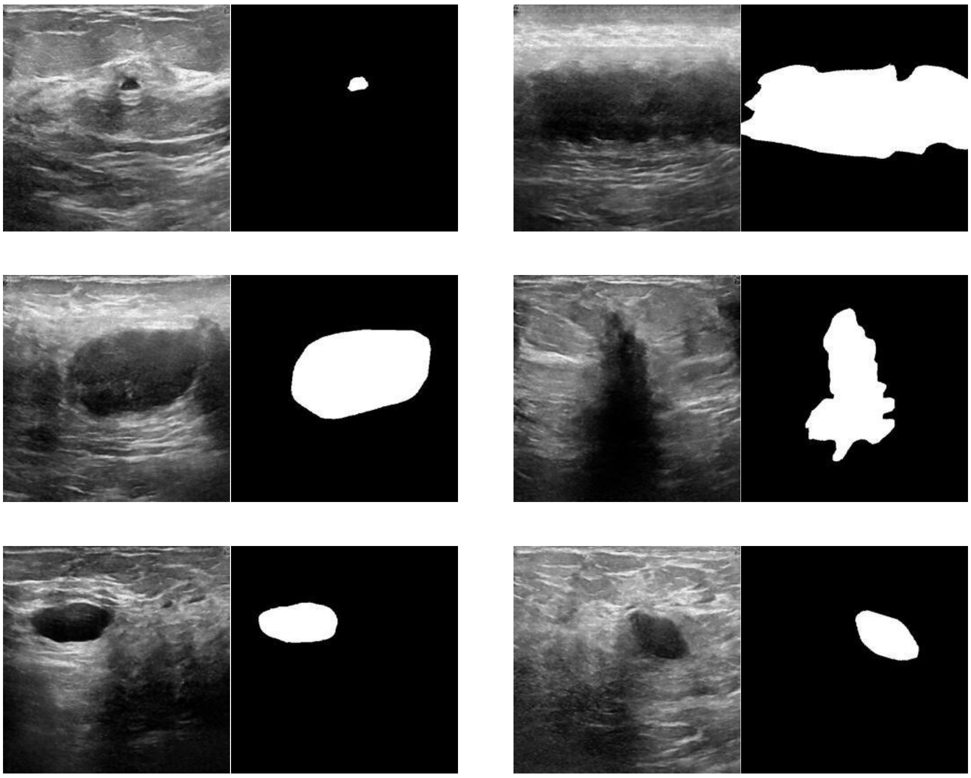


Figure 4. Synthesized dataset by Pix2PixHD

3.2.2 Unet

Unet was published by Olaf Ronneberger in 2015 and won the ISBI challenge [22]. It can effectively reduce the amount of data required for training and get better segmentation results. The architecture of Unet, as shown in figure 5, can be regarded as an encoder-decoder structure. The left part is encoder part, consisting of repeated two 3×3 convolutional layer, one ReLU function and one 2×2 max pooling layer. The number of feature channels will double after one downsampling step. The decoder part is just the opposite operation of encoder part but use stitching as a fusion method for feature maps.

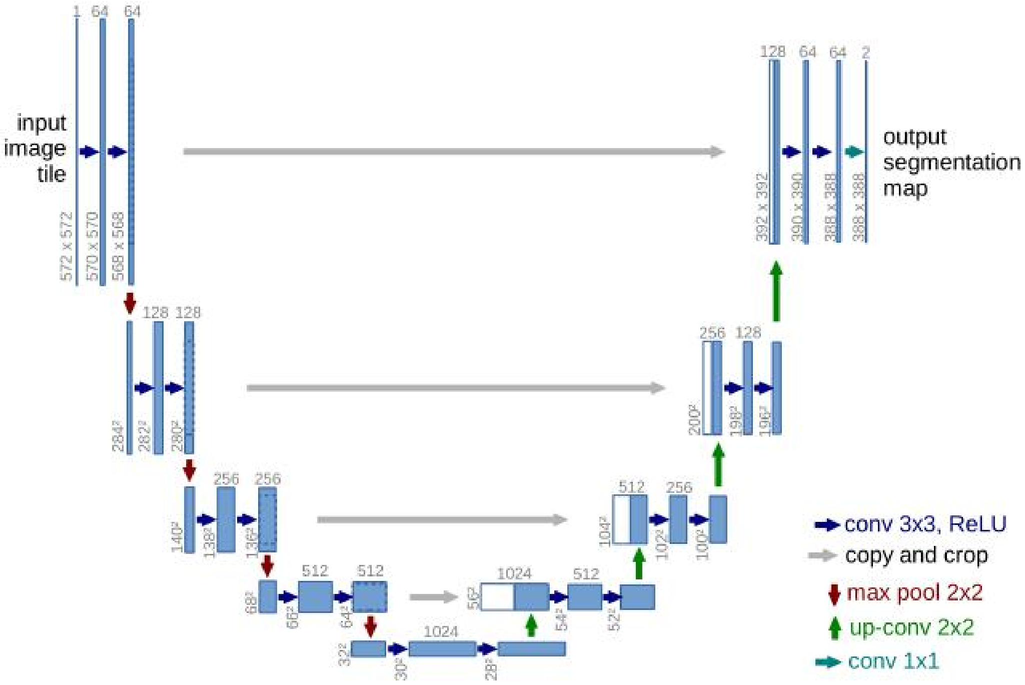


Figure 5. Unet architecture (224×224 images as input)

The Unet model achieved very great performance so that it becomes the baseline for many articles in medical image segmentation area. There are several reasons for extensive applications of Unet. One is because of the characteristics of Unet that the depth of network provides larger field of view and the losing edge features in the downsampling can be retrieved by applying concatenation with feature map. The other reason is because of the characteristics of medical images. All the features of medical images are really important so the skip connection structure of Unet really works. The medical datasets are usually very small so that Unet, which is light and simple, can reduce calculation and give researches more chances to modify the Unet for better adapting to a specific task. Therefore, there are large quantities of Variants of Unet, such as UNet++ [23] and W-Net [24].

In our project, we used the Unet for our segmentation task and visualize the results, as shown in figure 6.

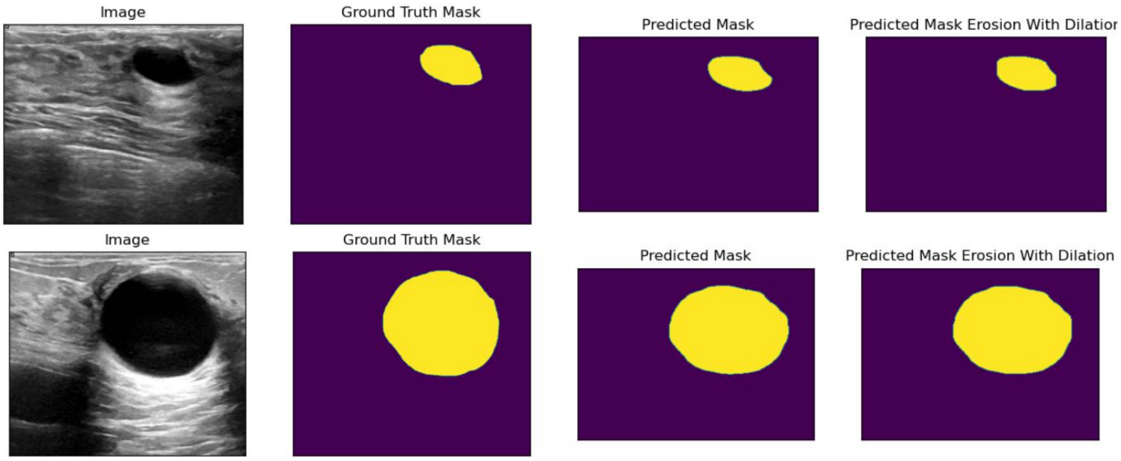


Image
Ground Truth Mask
Predicted Mask
Predicted Mask Erosion With Dilatior
Image
Ground Truth Mask
Predicted Mask
Predicted Mask Erosion With Dilation

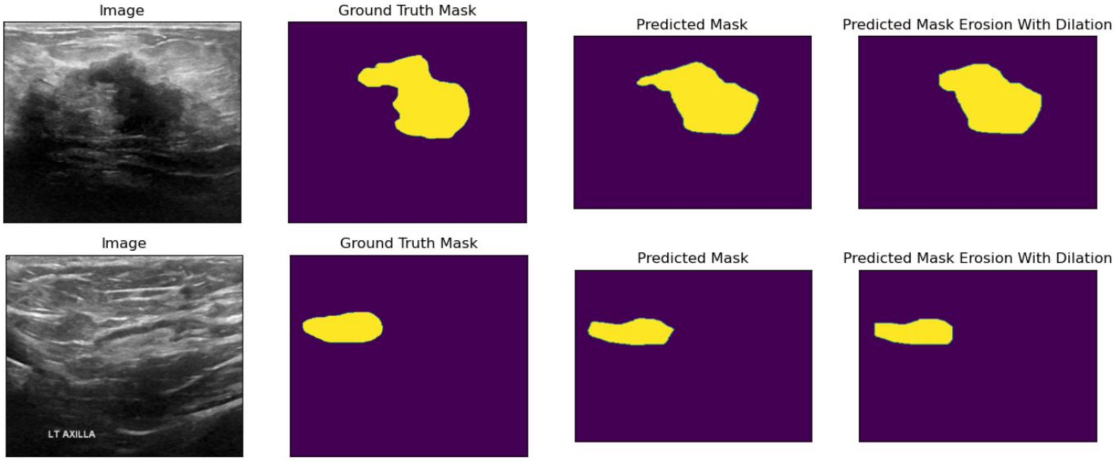


Figure 6. Visualization of Unet model's results on BUSI dataset

3.2.3 Evaluation Metrics

When using Unet for segmentation tasks, there are several evaluation metrics to evaluate the performance of segmentation results. For a binary classification task, the confusion matrix can be introduced to measure performance, which includes true positive (TP), false positive (FP), true negative (TN), false negative (FN) [25]. Segmentation task of our project is similar to binary classification situation and the relationship between confusion metrics and real situation, prediction results is shown in table 1 below. And we can obviously know that TP+FN+FP+TN = total population

Predict results Real situation	True	False
True	TP	FN

False	FP	TN

Table 1. Confusion matrix

After having the confusion matrix, the evaluation metrics of our project are based on it. IoU score, which is very commonly used in medical segmentation area, is defined as the ratio of the intersection and union between the predicted results and the true labels. So, the equation as follows:

$$IoU = \frac{X \cap Y}{X \cup Y} = \frac{TP}{TP + FP + FN} \#(3)$$

F1 score is also a very useful evaluation metrics in medical segmentation area, and the equation can be derived by its definition as follows:

$$F1 = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \#(4)$$

Other evaluation metrics are accuracy, precision and recall. They can be calculated by following equations:

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

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

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

3.2.4 FID Score

In the original Pix2PixHD paper, there is no metric to assess the quality of synthesized images. In our project, we introduce the FID score into the Pix2PixHD model. This part will present the basic knowledge of FID score and how it works.

In 2018, Martin Heusel published FID score and is considered as the improvement method of IS score, which was published before [26]. FID score and IS score both base on Inception Net-V3. FID is short for Frechet Inception Distance, solving the problem that IS score evaluates the synthesized images directly without considering the real images. So, FID score considers both of them by calculating the distance between real and synthesized images. The formula is as shown below:

$$d^2\big((m,C),(m_w,C_w)\big) = \|m - m_w\|_2^2 + Tr\big(C + C_w - 2\sqrt{CC_w}\big)\#(8)$$

In the formula, $\|X\|$ means the norm of matrix X, m represents the mean of real images, C stands for the covariance of real images. While, the m, C with subscript w means for synthesized images. Tr means the trace of a matrix. From the formula, it is obvious that the smaller the FID score is, the higher images quality and diversity will be.

3.3 Experiments

In our project, we design our experiments into three procedures. The first procedure is to use some geometric transformations augmentation techniques to generate some synthesized masks. This method is proposed by us and it is very easy and time-saving. The second procedure is to use Pix2PixHD model to generate synthesized images and evaluate the performance of results by FID score to find out if the scale and shift data augmentation works and which epoch is the best during training. Then combine the synthesized images and masks so that it will be a synthesized dataset. The last procedure is to combine real dataset and synthesized dataset and then use the Unet to train the combined dataset and test its performance using evaluation metrics which has mentioned before. However, it is probable happen that the synthesized dataset has the features of validation set, which cause the features leaking. It is not rigorous so that before the first procedure, it is necessary to split the dataset carefully. We split the dataset into training set and validation set. In order to avoid leaking, the images in the training set and those in the validation dataset are isolated. It means only use the training set to generate fake masks and images, then fake dataset. The validation set is only used for evaluation of Unet model. It will prevent the leak problem from happening and thus make the results convincing.

The computational environment of our project is that we use the NVIDIA 1060 with MAX-Q design for GPU training. The framework of deep learning we select PyTorch since it is friendly to researchers and newcomers. We also select Visual Studio Code as our coding editor. VS code is free and lightweight but can provide users plenty of extensions and includes git to bring convenience control of coding for users. Also, VS code can support Jupyter notebook in the software. We use Jupyter notebook for Unet segmentation task because it can run the code directly through the browser and display the running results below the code block. Anyhow, the selections above make it easier to write and test the code. The Unet structure, which with resnet34 for encoder and ImageNet for encoder weights, is supported by Segmentation Models library [28]. During training Unet for segmentation tasks, there are several settings. If used, the basic data augmentation method will be one of vertical flip, horizontal flip, shift, random rotate 90 and one of optional distortion, grid distortion, elastic transform, which is supported by Albumentations library [27]. The loss of Unet is the sum of Dice loss and BCE loss. The optimizer is Adam with initial learning rate equals to 0.0003, while the learning rate scheduler is step learning rate. The learning will multiple gamma (0.2) every 20 epochs. In our project, we trained our Unet model for 50 epochs. The settings of training as shown in table 2 below:

Dataset	BUSI
GPU	NVIDIA 1060 with MAX-Q design
Data augmentation	One of vertical flip, horizontal flip, shift, random rotate 90 and one of optional distortion, grid distortion, elastic transform
Loss	Dice Loss + BCE Loss
Model	Unet
Encoder	Resnet34
Encoder_weight	ImageNet
Activation function	sigmoid
Optimizer	Adam, lr=0.0003
Lr_scheduler	StepLR (gamma=0.2 every 20 epochs)
Evaluation metrics	IOU score, F1score (Dice score), accuracy, recall, precision
Epochs	50
Framework	PyTorch
Coding editor	VS code, Jupyter notebook

Table2. The basic settings of training

4. Results and Discussion

4.1 Pix2PixHD Generation Results

We random split the dataset and use the same training set for Pix2PixHD training. We train the model in three conditions. One is using the global type, which means only to use global G during training. Another is using the local type that means to use both global G and local G. The other is adding the shift and scale data augmentation pipeline into local type. Our test verifies the change in FID score during the GANs training process, which is observed by Tero Karras [29]. FID score will decrease in the beginning and increase later because of the "overfitting" will happen after training some epochs. As a result of the change during training, it is possible to find the best epoch with its FID score in the training process. The results are shown in table 3.

Condition	Best epoch	FID score
Global	80	102.42
Local	100	90.25
Local with augmentations	90	76.13

Table 3. Best epoch and FID score of three conditions

From the table, the result of our data augmentation pipeline is better than the original Pix2PixHD. We successfully achieve 15.6% reduction in FID score by using shift and scale data augmentation method. In the next part, we will test if the higher FID score will lead to better Unet segmentation results.

4.2 Unet Segmentation Results

After getting the synthesized dataset, we test the performance of Unet model to verify the effectiveness of Pix2PixHD augmentation method. So, we compare the results between raw data, data with basic data augmentation, data with Pix2PixHD data augmentation, data with both data augmentations, as shown in table 4. The condition of Pix2PixHD data augmentation is global type and its FID score is 102.42. In the table 4, we use aug1 to represent basic data augmentation and aug2 to represent global type Pix2PixHD data augmentation. In addition, we show the epoch with best IoU score and its other evaluation metrics. From the table, using either data augmentation can increase IoU score by 3%, while using both of them can increase IoU score by 4%.

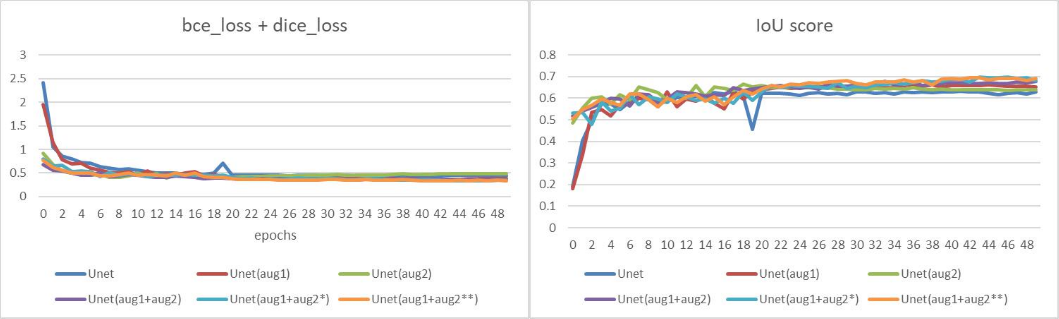
Model	Unet	Unet(aug1)	Unet(aug2)	Unet(aug1+aug2)
Loss	0.4419	0.3623	0.4166	0.3546
IoU score	0.6325	0.6641	0.6662	0.6787
F1 score	0.7544	0.7749	0.7770	0.7899
Accuracy	0.9623	0.9665	0.9610	0.9649
Precision	0.7856	0.8367	0.7805	0.7945
Recall	0.7685	0.7699	0.8205	0.8088

Table 4. BUSI best segmentation results on different augmentations

We also test the extent Pix2PixHD model can influence Unet segmentation result by comparing the segmentation results on different conditions of Pix2PixHD model with different FID scores, as shown in table 5. In the test, all conditions are using Unet with basic augmentations and Pix2PixHD augmentation (aug1+aug2). And the detailed validation data for each model is shown in figure 7, aug2* represents for Local condition and aug2** represents for Local with augmentations condition. We can find that by changing global type to local type of training, the IoU score increases by 2%. However, adding shift and scale data augmentation does not optimize the performance even though the decline of FID score is similar to the former.

Conditions	Global	Local	Local with augmentations
FID score	102.42	90.25	76.13
Loss	0.3546	0.3345	0.3312
IoU score	0.6787	0.6988	0.6954
F1 score	0.7899	0.8028	0.8042
Accuracy	0.9649	0.9678	0.9671
Precision	0.7945	0.8212	0.8267
Recall	0.8088	0.8198	0.8094

Table 5. segmentation results on different Pix2PixHD conditions



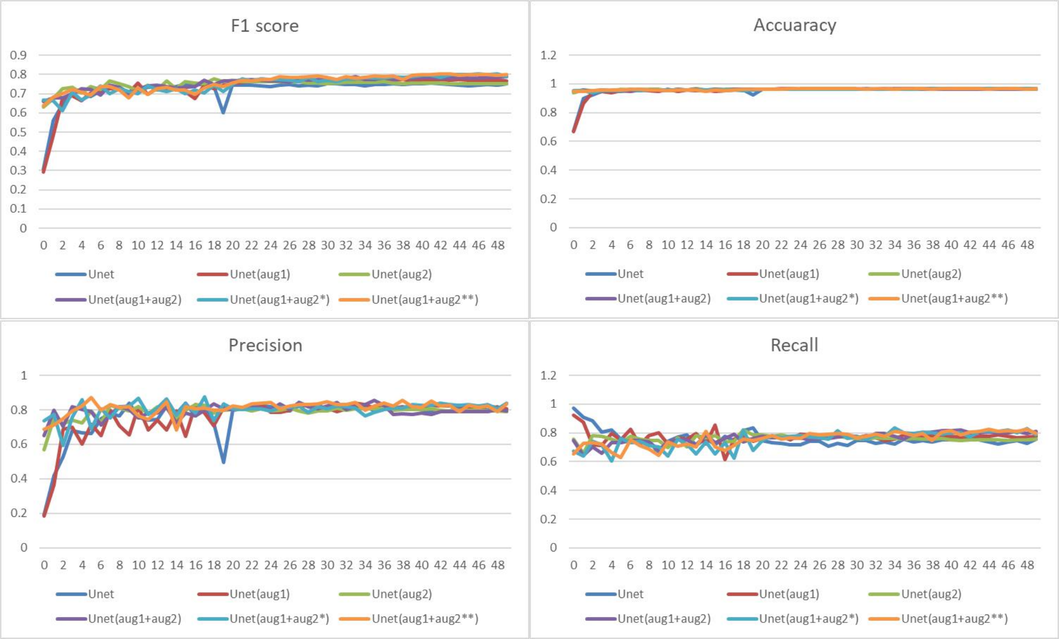


Figure 7. Detailed validation data for each model

4.3 Discussion

From the experiments, we can draw the conclusion that the usage of data augmentation will increase the value of evaluation metrics, which means it can improve the segmentation results. Also, combine basic data augmentation methods with Pix2PixHD data augmentation method will continually improve the performance. Applying some data augmentation methods to Pix2PixHD model to obtain synthesized dataset with higher FID score is helpful to segmentation task as well. We also need to point out that although FID score is associated with segmentation results, there may not be a linear relationship between them. They should be nonlinear since the larger the value of FID score, the greater the impact on segmentation results of the change of value.

Our project is meaningful in ultrasound segmentation area, which can increase the usage efficiency of raw data. The method to generate synthesized dataset is considered time saving and the improvement of segmentation result will help diagnostician reduce misdiagnosis. The introduction of FID score into Pix2PixHD can help researchers distinguish good and bad effects after training is completed.

Meanwhile, we need to point out some limitations of our project. First, our project is for a certain target and we did not Verify its generalization ability. Second, using our data augmentation method cannot generate synthesized normal images even we put normal part of dataset into training. Third, the images and masks were resized to 256 × 256 pixels png format. Features may be lost in the process and have bad effects on Pix2PixHD generation model. Finally, some data augmentation methods may not correspond to reality since disease has its own features.

5. Future Recommendations

In the process of our project, we also tried other methods in order to achieve our goals. However, since the reasons of time and technical problems, we did not use them in the experiment part. Here, we will introduce some methods may be possible to the improvement of our project.

5.1 Results of StyleGAN2-ada

In 2019, NVIDIA corporation proposed a GAN model called StyleGAN, which can generate images with high quality and variety [30]. One year later, NVIDIA improve the StyleGAN and update it to StyleGAN2 [31]. Later, the same group also apply adaptive discriminator augmentation (ADA) into StyleGAN2 and can effectively reduce the number of images needed for dataset. So far, StyleGAN3 has released which is more suitable for video and animation [32]. This part we mainly discuss StyleGAN2 and ADA method.

StyleGAN2-ada model has been released online, which is available at https://github.com/NVlabs/stylegan2-ada. We use StyleGAN2-ada method on BUSI dataset and achieved 62.10 FID score at best. We did not use it because it is unable to generate corresponding masks. However, the feasibility of using StyleGAN2-ada in the future can be seen from the results, as shown in figure 8. If the problem of generating masks can be solved by adding the masks generation part, it will be a great contribution. Our project may put energy into this part in the future.

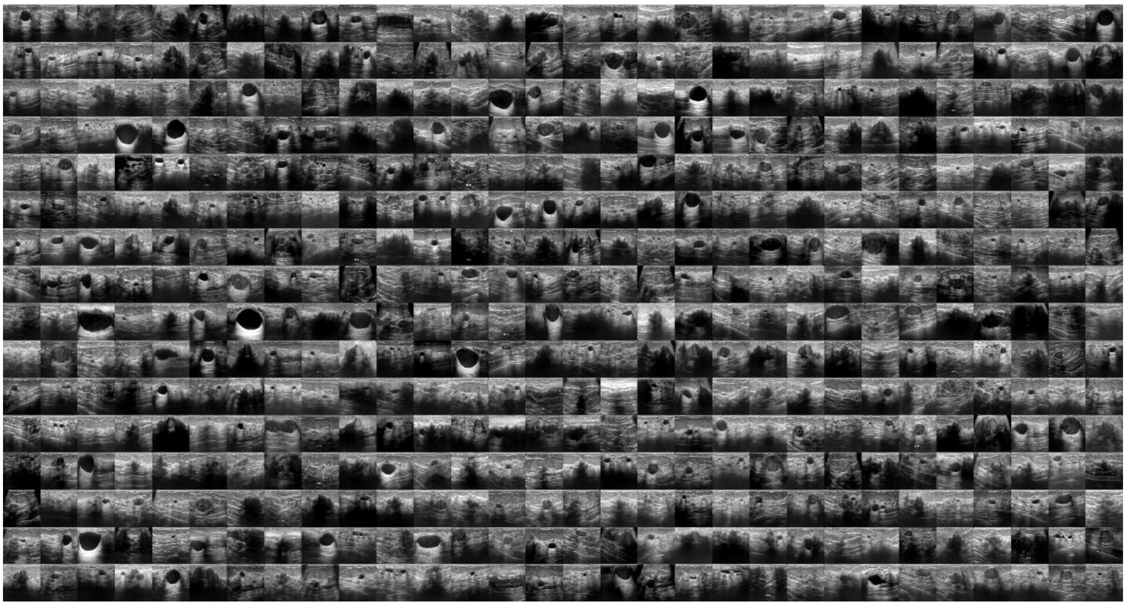


Figure 8. Synthesized images by StyleGAN2-ada

5.2 K-fold Cross Validation

Cross validation is a commonly used method for analysis [25]. In the past, we only separated the dataset one time, so the error may be large. 5-fold cross validation, 10-fold cross validation and 20-fold cross validation are the most common choices in cross validation. k-fold cross validation means to separate k mutually exclusive subsets with similar size, the separated results need to satisfy the equation below.

$$D = D_1 \cup D_2 \cup \ldots \cup D_k, D_i \cap D_j = \emptyset \ (i \neq j)$$

After separating the dataset, using the k-1 subsets as training set and the remaining 1 subset as the validation set just as the right picture shows. This way can perform k times training, which means k times training and return the mean of k results.

The usage of k-fold cross validation can effectively reduce the random error in experiment. Our procedures of experiment are complicated, so if we take the time and efficiency into considerations, 5-fold cross validation is more suitable for our project.

Reference

[1] Liu, Shengfeng, et al. “Deep Learning in Medical Ultrasound Analysis: A Review.” *Engineering*, vol. 5, no. 2, Apr. 2019, pp. 261–75. *DOI.org (Crossref)*, https://doi.org/10.1016/j.eng.2018.11.020.

[2] Chlap, Phillip, et al. “A Review of Medical Image Data Augmentation Techniques for Deep Learning Applications.” *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, Aug. 2021, pp. 545–63. *DOI.org (Crossref)*, https://doi.org/10.1111/1754-9485.13261.

[3] Wang, Ge. "A perspective on deep imaging." IEEE access 4 (2016): 8914-8924.

[4] Reddy, Uma M., Roy A. Filly, and Joshua A. Copel. "Prenatal imaging: ultrasonography and magnetic resonance imaging." Obstetrics and gynecology 112.1 (2008): 145.

[5] Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. CA Cancer J. Clin. 2020, 70, 7–30.

[6] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

[7] Zhong, Zhun, et al. “Random Erasing Data Augmentation.” *ArXiv:1708.04896 [Cs]*, 2, Nov. 2017. *arXiv.org*, http://arxiv.org/abs/1708.04896.

[8] Serra, Jean, and Luc Vincent. "An overview of morphological filtering." Circuits, Systems and Signal Processing 11.1 (1992): 47-108.

[9] Hussain, Zeshan, et al. "Differential data augmentation techniques for medical imaging classification tasks." AMIA annual symposium proceedings. Vol. 2017. American Medical Informatics Association, 2017.

[10] Goodfellow, Ian J., et al. "Generative Adversarial Networks." *ArXiv:1406.2661 [Cs, Stat]*, 1, June 2014. *arXiv.org*, http://arxiv.org/abs/1406.2661.

[11] Arjovsky, Martin, et al. "Wasserstein GAN." *ArXiv:1701.07875 [Cs, Stat]*, Dec. 2017. *arXiv.org*, http://arxiv.org/abs/1701.07875.

[12] Singh, Vivek Kumar, et al. "Conditional Generative Adversarial and Convolutional Networks for X-Ray Breast Mass Segmentation and Shape Classification." *ArXiv:1805.10207 [Cs]*, June 2018. *arXiv.org*, http://arxiv.org/abs/1805.10207.

[13] Negi, Anuja, et al. "RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks." *Arabian Journal for Science and Engineering*, vol. 45, no. 8, Aug. 2020, pp. 6399–410. *DOI.org (Crossref)*, https://doi.org/10.1007/s13369-020-04480-z.

[14] Fujioka, Tomoyuki, et al. "Breast Ultrasound Image Synthesis Using Deep Convolutional Generative Adversarial Networks." *Diagnostics*, vol.

9, no. 4, Nov. 2019, p. 176. *DOI.org (Crossref)*, https://doi.org/10.3390/diagnostics9040176.

[15] Jiang, Jue, et al. “Cross-modality (CT-MRI) Prior Augmented Deep Learning for Robust Lung Tumor Segmentation from Small MR Datasets.” *Medical Physics*, vol. 46, no. 10, Oct. 2019, pp. 4392–404. *DOI.org (Crossref)*, https://doi.org/10.1002/mp.13695.

[16] Agudo, Alberto Montero. Generative Adversarial Networks Based Data Augmentation for Ultrasound Fetal Brain Planes Classification. p. 86.

[17] Fan, Zong, et al. “Application of DatasetGAN in Medical Imaging: Preliminary Studies.” *ArXiv:2202.13463 [Cs]*, Feb. 2022. *arXiv.org*, http://arxiv.org/abs/2202.13463.

[18] Li, Daiqing, et al. “Semantic Segmentation with Generative Models: Semi-Supervised Learning and Strong Out-of-Domain Generalization.” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2021, pp. 8296–307. *DOI.org (Crossref)*, https://doi.org/10.1109/CVPR46437.2021.00820.

[19] Al-Dhabyani, Walid, et al. “Dataset of Breast Ultrasound Images.” *Data in Brief*, vol. 28, Feb. 2020, p. 104863. *DOI.org (Crossref)*, https://doi.org/10.1016/j.dib.2019.104863.

[20] Wang, Ting-Chun, et al. “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs.” *ArXiv:1711.11585 [Cs]*, Aug. 2018. *arXiv.org*, http://arxiv.org/abs/1711.11585.

[21] Isola, Phillip, et al. “Image-to-Image Translation with Conditional Adversarial Networks.” *ArXiv:1611.07004 [Cs]*, Nov. 2018. *arXiv.org*, http://arxiv.org/abs/1611.07004.

[22] Ronneberger, Olaf, et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *ArXiv:1505.04597 [Cs]*, May 2015. *arXiv.org*, http://arxiv.org/abs/1505.04597.

[23] Zhou, Zongwei, et al. “UNet++: A Nested U-Net Architecture for Medical Image Segmentation.” *ArXiv:1807.10165 [Cs, Eess, Stat]*, July 2018. *arXiv.org*, http://arxiv.org/abs/1807.10165.

[24] Xia, Xide, and Brian Kulis. “W-Net: A Deep Model for Fully Unsupervised Image Segmentation.” *ArXiv:1711.08506 [Cs]*, Nov. 2017. *arXiv.org*, http://arxiv.org/abs/1711.08506.

[25] Zhou, Zhihua "Machine Learning". Tsinghua University Press. 2016

[26] Heusel, Martin, et al. “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.” *ArXiv:1706.08500 [Cs, Stat]*, Jan. 2018. *arXiv.org*, http://arxiv.org/abs/1706.08500.

[27] Buslaev, Alexander et al. "Albumentations: Fast and Flexible Image Augmentations". Information 11. 2(2020).

[28] Pavel Yakubovskiy. "Segmentation Models Pytorch. " https://github.com/qubvel/segmentation_models.pytorch. (2020).

[29] Karras, Tero, et al. “Training Generative Adversarial Networks with Limited Data.” *ArXiv:2006.06676 [Cs, Stat]*, Oct. 2020. *arXiv.org*, http://arxiv.org/abs/2006.06676.

[30] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[31] Karras, Tero, et al. "Analyzing and improving the image quality of stylegan." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[32] Karras, Tero, et al. "Alias-free generative adversarial networks." Advances in Neural Information Processing Systems 34 (2021).

[33] Bowles, Christopher, et al. “GANsfer Learning: Combining Labelled and Unlabelled Data for GAN Based Data Augmentation.” *ArXiv:1811.10669 [Cs]*, Nov. 2018. *arXiv.org*, http://arxiv.org/abs/1811.10669.

[34] Shi, Haoqi, Junguo Lu, and Qianjun Zhou. "A novel data augmentation method using style-based GAN for robust pulmonary

nodule segmentation." 2020 Chinese Control and Decision Conference (CCDC). IEEE, 2020.

[35] Bowles, Christopher, et al. "Gan augmentation: Augmenting training data using generative adversarial networks." arXiv preprint arXiv:1810.10863 (2018).

[36] Rashid, Haroon, M. Asjid Tanveer, and Hassan Aqeel Khan. "Skin lesion classification using GAN based data augmentation." 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.

[37] Madani, Ali, et al. "Chest x-ray generation and data augmentation for cardiovascular abnormality classification." Medical Imaging 2018: Image Processing. Vol. 10574. International Society for Optics and Photonics, 2018.

[38] Pang, Ting, et al. "Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification." Computer Methods and Programs in Biomedicine 203 (2021): 106018.

[39] Chaudhari, Poonam, Himanshu Agrawal, and Ketan Kotecha. "Data augmentation using MG-GAN for improved cancer classification on gene expression data." Soft Computing 24.15 (2020): 11381-11391.

[40] Yu, Suxiang, et al. "Generative adversarial network based data augmentation to improve cervical cell classification model." Math. Biosci. Eng 18 (2021): 1740-1752.

[41] Bhagat, Vedant, and Swapnil Bhaumik. "Data augmentation using generative adversarial networks for pneumonia classification in chest xrays." 2019 Fifth International Conference on Image Information Processing (ICIIP). IEEE, 2019.