RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks

Anuja Negi[1] · Alex Noel Joseph Raj[2] · Ruban Nersisson[3] · Zhemin Zhuang[2] · M Murugappan[4]

Abstract

Early-stage detection of lesions is the best possible way to fight breast cancer, a disease with the highest malignancy ratio among women. Though several methods primarily based on deep learning have been proposed for tumor segmentation, it is still a challenging problem due to false positives and the precise boundary detection required for segmentation. In this paper, we propose a Generative Adversarial Network (GAN) based algorithm for segmenting the tumor in Breast Ultrasound images. The GAN model comprises of two modules: generator and discriminator. Residual-Dilated-Attention-Gate-UNet (RDAU-NET) is used as the generator which serves as a segmentation module and a CNN classifier is employed as the discriminator. To stabilize training, Wasserstein GAN (WGAN) algorithm has been used. The proposed hybrid deep learning model is called the WGAN-RDA-UNET. The model is assessed with several quantitative metrics and is also compared with existing methods both quantitatively and qualitatively. The overall Accuracy, PR-AUC, ROC-AUC and F1-score achieved were 0.98, 0.95, 0.89 and 0.88 respectively which are better than most conventional deep net models. The results also showcase the shortcomings of CNN, RDA U-Net and other models and how they can be rectified using the WGAN-RDA-UNET model.

1 Introduction

Breast cancer is the most common cancer among women worldwide and is the second leading cause of death among women. More than two million new cases were registered in 2018 alone. The Breast Cancer Foundation estimates that over 252,710 women will be diagnosed with breast cancer in the United States and more than 40,500 will die each year. Although breast cancer is rare among men, it is approximated that 2470 men will be diagnosed with breast cancer and 460 will die each year [1].

One way of considerably improving the survival rate is by diagnosis at an early stage. Non-invasive breast cancer diagnosis modalities primarily include X-ray, Magnetic Resonance Imaging (MRI) or Ultrasound (US) imaging. Other methods include invasive diagnosis such as biopsies which can cause damage to the tissues. Ultrasound imaging is the most commonly used modality for examination as there is less exposure to ionizing radiations, combined with low cost. Figure 1 presents a few Breast Ultrasound Images (BUS) having both benign and malignant tumors. It can be seen how the malignant tumors vary in shape and complexity compared to benign tumors. It is needed that early-stage detection and accurate tumor assessment be a cost-effective and fast process. Unfortunately, diagnosing lesion is a time-consuming and manual analysis and verification require visual interpretation of the breast lesion area by an experienced professional.

✉ Alex Noel Joseph Raj
jalexnoel@stu.edu.cn

1 Research Intern, Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Department of Electronics Engineering, College of Engineering, Shantou University, Shantou 515063, China

2 Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Department of Electronics Engineering, College of Engineering, Shantou University, Shantou 515063, China

3 School of Electrical Engineering, Vellore Institute of Technology, Vellore, India

4 Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, Doha 13133, Kuwait
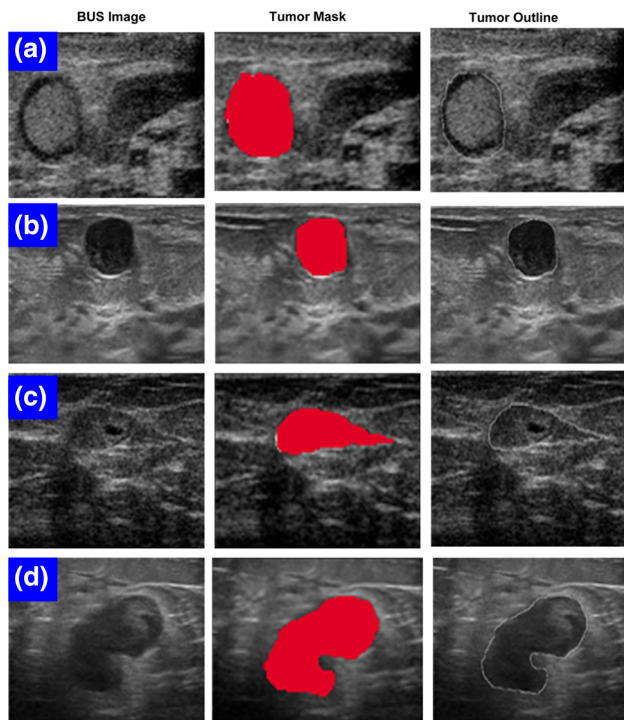
Fig. 1 Benign and malignant tumors (**a**, **b** are benign and **c**, **d** are malignant tumors. The US image, mask, and lesion outline for each have been presented for a clearer understanding and The images a and b are obtained from the Gelderse Vallei Hospital dataset, and the images c and d are obtained from First Affiliated Hospital of Shantou University and Dataset B respectively)

Generally, US images suffer from speckle noise and low contrast which makes visual analysis difficult. Due to this, diagnosis can vary widely due to the subjectivity and complexity of the task. Hence, the need for a comprehensive and automated method of lesion localization and segmentation is clearly recognized.

1.1 Literature Review

Several deep learning techniques, and algorithms have been developed for medical imaging tasks such as localization and segmentation. They have, over the years become state-of-the-art technologies providing accurate results for medical assistance. Convolutional Neural Networks (CNNs) have shown remarkable performance in segmentation tasks of medical images [2]. They have become the standard in segmentation due to high representation power, filter sharing properties and fast interference. However, CNNs rely on pixel-wise functions between the model predicted image and the standard image. Due to this, segmentation results are often blurry and suffer from false positives. To overcome this, Fully Convolutional Networks (FCNs) [3] and the U-Net [4] architectures are being used widely nowadays. Xide Xia [5] proposed a W-Net, a new architecture that ties together two FCNs into an auto encoder and presented how performance can be greatly improved compared to CNNs. The PASCALVOC2012 dataset was used for training and the Berkeley Segmentation Database (BSDS 300 and BSDS500) was used for evaluation.

A U-Net is based on the principle of FCNs. It is composed of an encoder to extract features and a decoder for image reconstruction. Further, skip connections are added for accurate localization by combining both low-level and high-level features. Recently, several additions and modifications have been explored with respect to U-Nets to further improve performance. Goufeng Tong [6] employed a U-Net for pulmonary nodule segmentation. Batch normalisation layer is incorporated in the U-Net to speed up training and avoid overfitting, and a residual network is also added to improve the final predictions. The author used LUNA2016 contest dataset to present the results and achieved a dice coefficient of 0.736. Ozan Oktay [7] demonstrated the implementation of Attention Gates (AG) in U-Nets for segmentation on Computed Tomography (CT) pancreas. This approach eliminated the need for using a localisation model. Evaluation is done on TCIA Pancreas CT-82 and multi-class abdominal CT-150 benchmarks. Zhuang et al. [8] proposes a modified U-Net model called Grouped-Resaunet (GRA U-Net) for nipple segmentation from Automatic Whole Breast Ultrasound (AWBUS) images with an aim to localize the nipple region from the rest of the breast region. They used coronal views of Automated Whole Breast Ultrasound Images (AWBUS) obtained from the First Affiliated Hospital of Shantou Medical College. Recently, Zhuang et al. [9] presented a RADU-Net model for accurate lesion segmentation in BUS images. The model incorporates the 'advantages' of residual networks, attention gate mechanism, and dilation modules to present segmentation results close to ground truths. An additional approach to further improve segmentation in medical imaging and obtain precise results is to employ Generative Adversarial Networks (GAN). This is a current hot topic that is still at infancy. GANs are generation models composed of two networks, generator and discriminator [10]. Whilst the generator tries to generate realistic outputs as the provided gold standard, the discriminator differentiates the generated outputs from the gold standard. Salome Kazeminia [11] and Xin Yi [12] review a variety of recent literature on the medical applications of GANs. They show how GANs have been able to not only learn existing computer vision tasks better but also synthesize images. They thereby second the benefits of adversarial training in medical image reconstruction, segmentation, detection and other such tasks. Zeju Li [13] proposes a CNN-based GAN for brain tumor segmentation. A dice score of 0.897 was achieved on the Brats2017 dataset. Weixiang Hong [14] integrated a GAN with FCN to achieve a segmentation output which represents the target ground truth more accurately. The proposed

model exceeded the mean IoU of state-of-the-art method by 12–20% on Cityscapes dataset. Zhongyi Han [15] also used GANs for automatic segmentation and classification of spinal structures from MRIs. They proposed a Recurrent Generative Adversarial Network called Spine-GAN and achieved a high pixel accuracy of 96.2%. Jaemin et. al [16] showcases how generative adversarial training produces results with less false positives with precise segmentation results. The method employed successfully generates a precise map of retinal vessels in fundoscopic images, even at the terminal ends. A Dice Coefficient of 0.829 on DRIVE dataset and 0.834 on STARE dataset was achieved. Although GANs seem very promising, they are hard to train due to issues of non-converging model parameters, model collapse diminishing gradient problem and unstable gradient. Choosing a problem specific optimization cost function is one of the most effective ways of stabilizing adversarial training. Arjovsky et al. [17] proposed an alternative way to traditional GAN training called Wasserstein GAN (WGAN). WGAN offered improved optimization stability and the new loss metric that correlates with the convergence of the generator. Enokiya et al. [18] utilized WGAN and proposed an automatic liver segmentation method using U-Nets and WGAN. The dice value was considerably improved on the two datasets using GAN.

Given the rapid advancement of GANs in medical image segmentation, this paper employs a new approach for lesion segmentation in BUS images with adversarial networks. The paper extends the work of [9] which proposes a Residual-Dilated-Attention-Gate-UNet (RDAU-NET) and combines it with a WGAN, to obtain a reliable and accurate lesion segmentation technique. The developed deep learning model is called RDA-UNET-WGAN. We show that adversarial training improves the quality of segmentation and generates outputs indistinguishable to those by professionals. The rest of the paper is organized as follows: Sect. 2 describes the proposed architecture in detail and the parameters on which the experiment results have been evaluated, Sect. 3 presents the results and comparison with existing methodologies. Finally, Sect. 4 presents the conclusion.

2 Methodology

2.1 Architecture

GANs are deep neural networks comprising of two networks called the generator and discriminator, pitting against each other. The generator captures data distribution to be able to generate new instances like the learning data. The discriminator estimates the probability of the sample being authentic i.e. from the training set or generated by the generator. Figure 2 illustrates the traditional architecture of GANs. Training of the networks is an iterative process and defined as a minmax-

type of competitive learning between the generator G and the discriminator D, as represented in Eq. (1).

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

where z is a random number, p_{data} is the real data distribution, p_z is the generated data distribution and $D(x)$ is the discriminator output value showing the probability of x being real.

The proposed RDA-UNET-WGAN architecture utilizes the above-mentioned idea and for segmenting lesions in BUS images. The trained segmentation model acts as the generator along with an adversarial network which discriminates segmentation results from the ground truths.

Here we employ the RDA-UNET [9] as the segmentation model (Generator) and a fully connected CNN network is used as the discriminator. Both these networks collectively form the RDA-UNET-WGAN. Figure 3 illustrates the outline of the RDA-UNET-WGAN architecture which enables detecting and correcting higher-order discrepancies between segmented lesion results from the generator and the ground truths provided. Due to this correction mechanism, the segmentation results generated by the RDA-UNET-WGAN are more accurate and closer to ground truths annotated by experts.

2.1.1 Generator

The generator serving as a segmentation model is a combination of Residual nets, Dilation convolution modules and an Attention Gate (AG) mechanism composed within a U-Net architecture. BUS images and their corresponding ground truths serve as input, and the predicted segmentation mask as outputs of the RDA-U-Net. The objective of a generator is to generate fake images such that they are mistaken as authentic by the discriminator. The fake images, in this case, are the lesion segmented maps of the input BUS images.

Figure 4 shows the network architecture of the generator. The architecture is similar to the model discussed in [9]. It has 6 residual nets that extract significant features from the BUS images and the down-sampled feature maps at the end of the forward pass are fed to a dilated convolution module. Residual units and dilation convolutions were employed to avoid accuracy saturation (vanishing gradients) during training and to improve the receptive field respectively. The output from the forward pass was fed to an up-sampler comprising five residual nets, each having an individual AG to pay scrutiny to the lesion region rather than the non-lesion regions. The output of the generator is a binary segmentation mask produced by the final convolution layer representing the classification
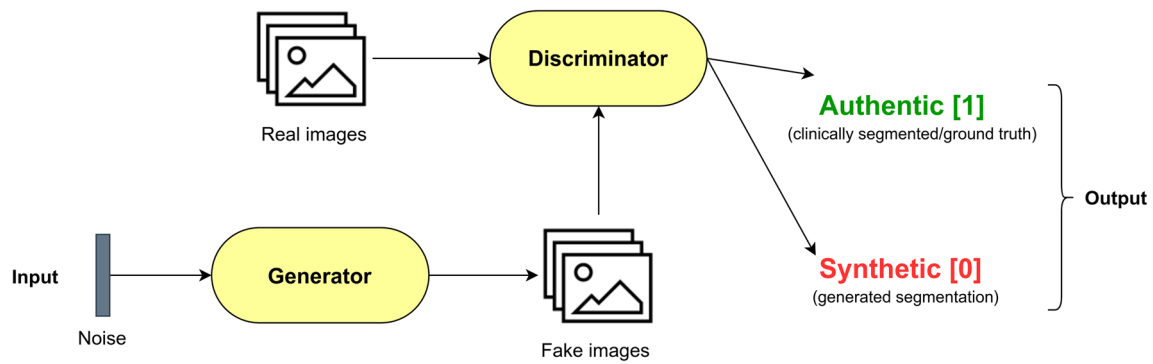
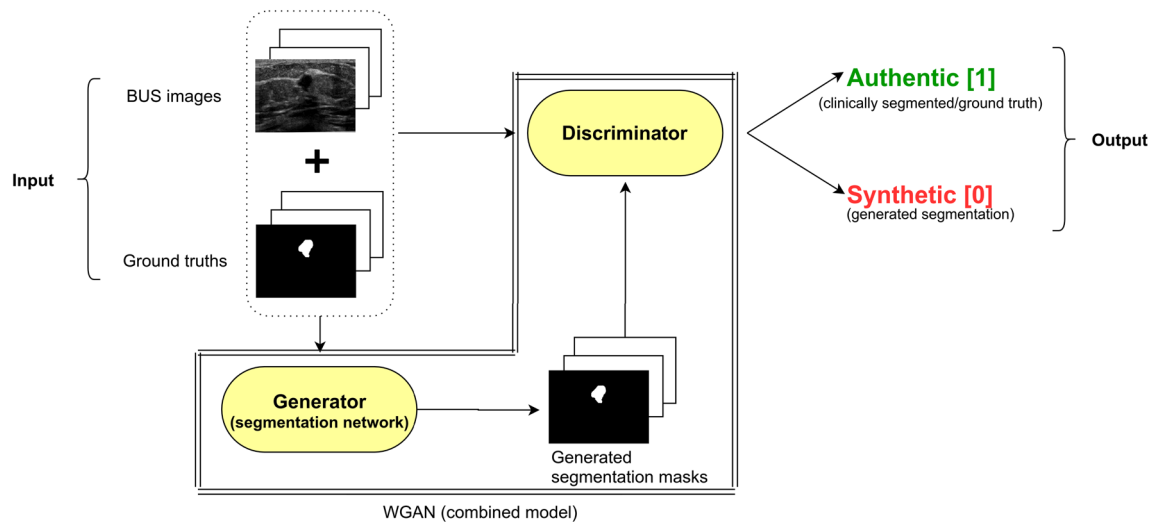**Fig. 2** Architecture overview of a traditional GAN

Fig. 3 Proposed RDA-U-Net-WGAN architecture

label for each pixel. For the generators loss function, the cost function which is stated in Eq. 11 was employed.

2.1.2 Discriminator

The discriminator is a vital component of the GAN. The goal of a discriminator is to recognize the authenticity of the instances given to it. The discriminator model in the proposed architecture is a classification network. The discriminator network is a CNN composed of ten convolution layers and 1 fully connected layer for end classification. It consists of repeated convolutions, each following a leaky rectified linear unit and a maxpooling layer for down sampling. Batch normalization is also added to regularise and speed up the training process as the training of the discriminator is directly proportional to the effectiveness of the adversarial loss.

The segmented lesion result (fake/false) from the generator and the ground truth (real/true) serve as training samples to the discriminator, in the form of one-hot encoding. This means that the binary output from the discriminator indi-

cates if the input is a generated segmentation result from the generator or a ground truth from the training set. The model classifies the legitimacy of the input at an image level that is the whole image is classified as 1 for real and 0 for fake. The discriminator is presented in Fig. 5. A learning rate of 0.0001 and Adam optimizer is used for training the discriminator. The loss function utilized is Binary Cross-Entropy (BCE) [16] which corrects the deviation of the segmentation result from the ground truth. It is calculated as follows:

$$BCE = -(y.\log(p) + (1-y).\log(1-p)) \qquad (2)$$

where p is predicted value, y is the real value.

2.1.3 RDA-UNET-WGAN Network

The generator and the discriminator incorporated together form the combined model, the RDA-UNET-W-GAN. In this combined model, the BUS images and their corresponding ground truths serve as the input and labels respectively. A batch of BUS images is provided as input to the generator
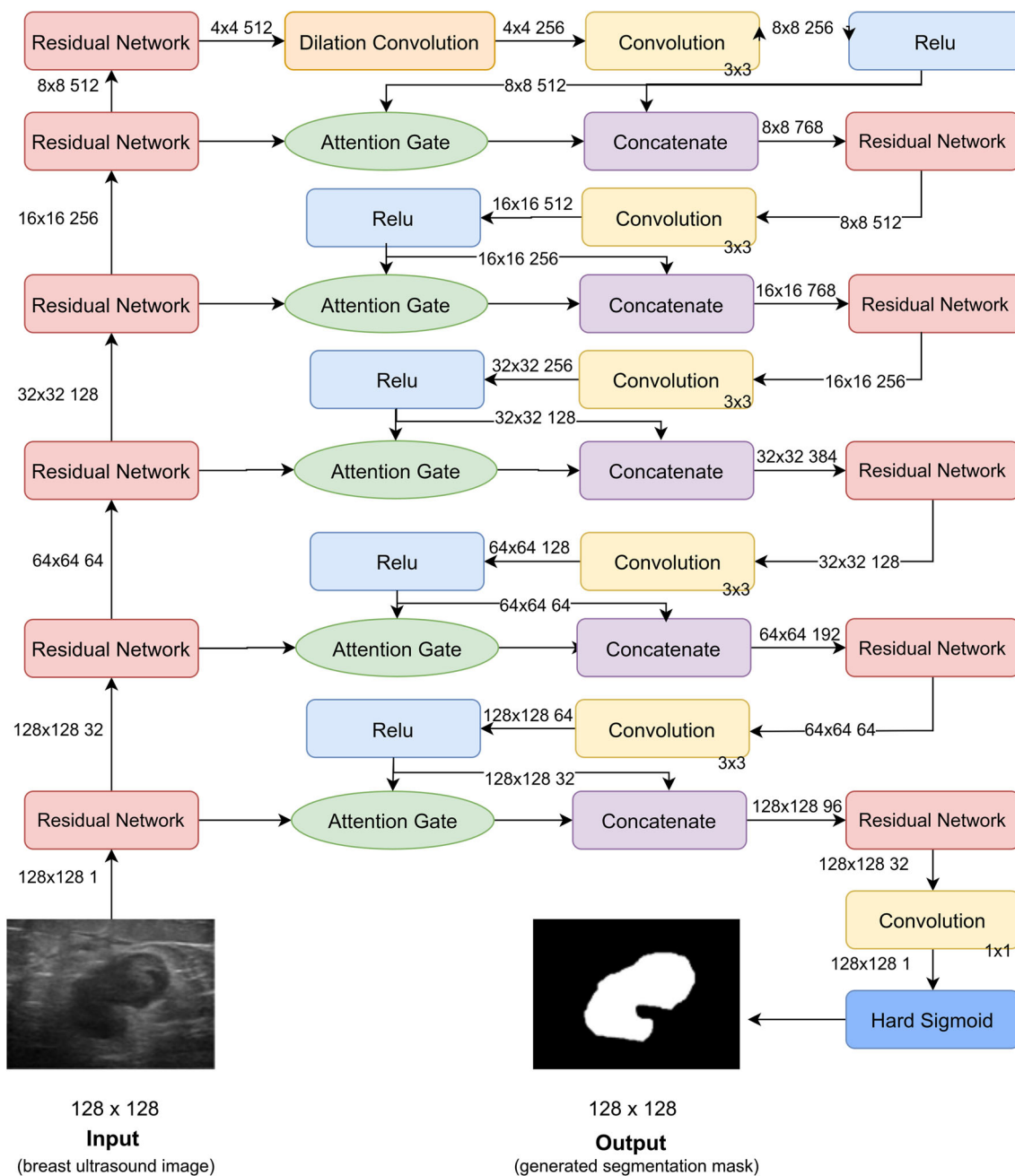
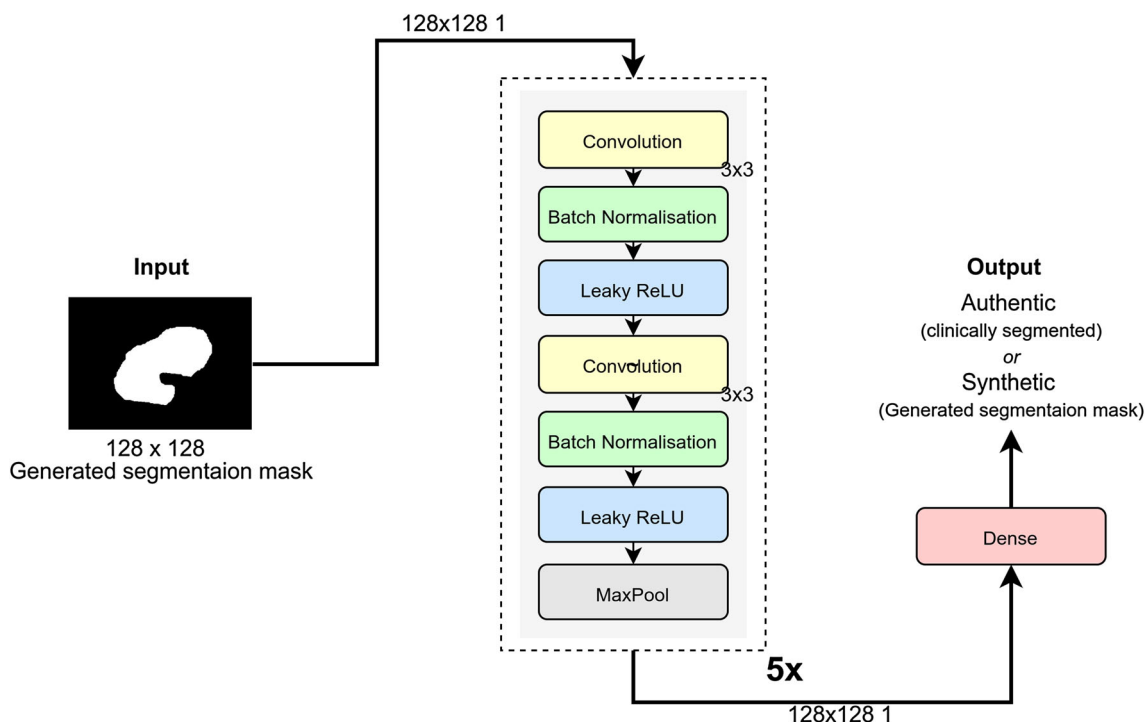**Fig. 4** Generator architecture overview

to create segmentation maps. Later the segmentation maps along with their respective ground truth labels are given as input to the discriminator. The discriminator presents the overall output of the combined model. The data flow is presented in Fig. 3.

In adversarial training, the objective function plays a key role. Here WGAN is used which differs from conventional GANs in regards to its objective function. WGAN finds the Wasserstein distance between the segmentation result and the ground truth. As a result, learning is more stable than that of a conventional GAN [18]. In contrast to Eq. 1, the

minmax-type of competitive learning of the generator G and the discriminator D in WGAN can be expressed as:

$$\min_{G} \max_{D} V(G, D) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[D(x)] \\ - \mathbb{E}_{z \sim p_z(z)}[D(G(z))] \tag{3}$$

where z is a random number, p_{data} is the real data distribution, p_z is the generated data distribution and $D(x)$ is the discriminator output value showing the probability of x being real. In our case, z will be the input BUS image and x will be the seg-

128x128 1

Fig. 5 Discriminator architecture overview

mented result from the generator. The loss function penalizes the dissimilarities between the segmented results compared to the ground truth, and the discriminator for incorrect classification. This ensures that both the networks are stable and do not overpower the other.

2.2 Training Scheme

The primary goal of the proposed model is to produce state-of-the-art segmentation results. If the training of both the generator and the discriminator is started from scratch, the effective adjustment of the parameters will be poor as the networks by themselves will not be stable initially. To overcome this, the generator model is partially trained separately. For the combined model, the RDA-UNET-WGAN is adversarial trained. The previously trained generator model is loaded and the discriminator is built. Then, the combined model is trained. The training scheme is an iterative process with several rounds of alternated generator and discriminator training. First, the discriminator is trained for one step and the parameters updated. The gradient is propagated and forms the adversarial loss. The discriminator is then frozen and the combined network is trained for one step. Finally, both models are evaluated against the validation set. This cycle executes for several rounds and depicts the min-max of the proposed model as stated in Eq. 3. The adversarial loss participates in training the generator network. With each cycle, the segmentation becomes more accurate and reliable

with respect to the ground truth provided. This improvement with each iterative cycle is illustrated in Fig. 6.

The updation of the generator and discriminator parameters is dependent on the discriminators classification result. When the discriminator fails in discriminating real data from generated data, its own parameters are updated. The generator is updated if the discriminator succeeds. The generators parameters are optimized based on the discriminator and its own loss, thus enabling the generator to segment with a high accuracy rate such that the discriminator is fooled.

2.3 Evaluation Metrics

Evaluation metrics are mandatory to measure the success of an experiment. The image segmentation performance is quantitively evaluated using the following indices: Accuracy, Sensitivity/Recall, Specificity, Precision, F1score, Mean-Intersection-Over-Union (M-IOU), Dice Similarity Coefficient (DSC), Precision-Recall (PR) Area-Under-Curve (AUC) and Receiver Operating Characteristic (ROC) AUC. These metrics evaluate region-based segmentation performance from varied aspects and have values ranging between 0 and 1. False Negatives (FN), True Positives (TP), True Negatives (TN), False Positives (FP) are basic values used to compute the metrics.

– *Accuracy* is the quality of being correct and the most common evaluation index. It is the measurement of closeness

Fig. 6 Segmentation maps generated at different cycles. (The image has been taken from Dataset B. Block **a**, **b**, **c**, **d**, **e**, **f**, **g**, **h** and **i** showcase the BUS image, the ground truth mask, ground truth outline and the predicted segmentation maps at cycle 0, 5, 10, 15, 20 and 25 respectively. Regions marked in purple represent the intersection of predicted segmentation and the ground truth, red represents the lesion area not predicted as lesion and blue represents the non-lesion regions classified as lesion by the model)
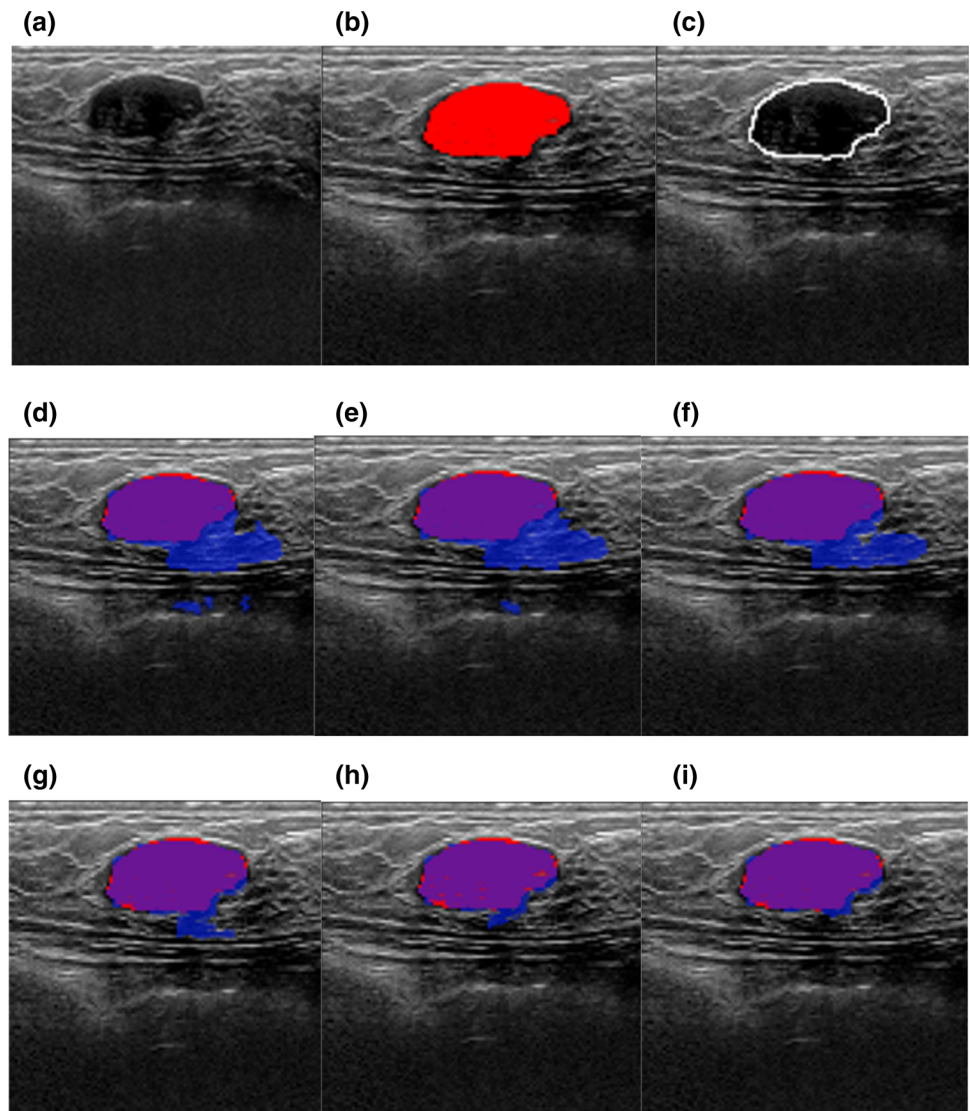
of the predicted segmentation to the ground truth. It is calculated as;

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

– *Sensitivity/recall* or True Positive Rate (TPR) is the proportion of actual positives that are identified as such. It is calculated as;

$$Sensitivity/Recall = \frac{TP}{TP + FN} \quad (5)$$

– *Specificity* or True Negative Rate (TNR) is the proportion of actual negatives that are identified as such. It is calculated as;

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

– *Precision* is the ratio of true positive values to all positive values. It is calculated as,

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

– *F1 score* is one of the most important evaluation metrics. It especially serves as an evaluation measure for uneven class distributions. It is the harmonic mean of precision and sensitivity and calculated as;

$$F1score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \quad (8)$$

– *Mean-intersection-over-union (M-IOU)* is a measure coincidence between the ground truth and the predicted segmentation result. It is the ratio between the intersection and union of the ground truth (G) and the predicted segmentation result (P). Higher the M-IOU, greater is the

Table 1 Dataset details . (Source of the dataset along with number of images, usage and additional details are specified)

	Number of imagesPurpose		Details
Gelderse Vallei Hospital in Ede, Netherlands	857	Training and Validation	For training, the original 730 images were expanded to 2919 through data augmentation techniques[1] as mentioned in [9]. Validation: 127 images
First Affiliated Hospital of Shantou University, Guangdong Province, China	42	Testing	Unseen during the training phase, contains both malignant and benign tumours
Breast Ultrasound Lesions Dataset (Dataset B)	163	Testing	Unseen during the training phase, contains both malignant and benign tumours

[1]The BUS images are first fused with their ground truth followed by the application of 4 affine transform operations: symmetrical shifting along the vertical and horizontal axis, shear transformation with an angle of 10 degrees and mirroring along the horizontal plane

segmentation accuracy. It is calculated as;

$$M - IOU = \frac{G \cap P}{G \cup P} = \frac{TP}{TP + FP + FN} \qquad (9)$$

– *Dice similarity coefficient (DSC)* is the measure of overlap between the two samples. It is the most frequently used metric for evaluating segmentation tasks. It represents the similarity between the ground truth (G) and the predicted segmentation (P). Higher DSC signifies that the segmentation result and the ground truth are identical. DSC is calculated as;

$$DSC = \frac{2 \mid G \cap P \mid}{\mid G \mid \cup \mid P \mid} = \frac{2 * TP}{TP + TN + FP + FN} \qquad (10)$$

$$DSCloss = 1 - DSC \qquad (11)$$

– *Precision-recall (PR) area-under-curve (AUC)* is the area under the PR curve. The PR curve showcases the precision-recall trade-offs at different thresholds. A high PR-AUC indicates both low false-positive rate and low false-negative rate. An ideal network has an PR-AUC of 1 that is 0 error probability.

Table 2 Parameter summaries of models

Model	Trainable parameters	Non trainable parameters	Total parameters
Generator	24,079,536	18,874	24,098,410
Discriminator	4,714,433	3968	4,718,401
RDA-UNET-WGAN (Combined Model)	24,079,536	4,737,275	28,816,811

– *Receiver operating characteristic (ROC)* is the area under the curve of Recall/TPR and FPR. It represents the measure of separability and distinguishing capability of lesion regions from non-lesion regions. Higher the ROC-AUC, better is the predicted lesion segmentation. The ideal ROC-AUC is one.

3 Experiment and Results

BUS images were segmented for lesions using the proposed architecture. The dataset has a total of 1062 images obtained from three different sources namely: Gelderse Vallei Hospital in Ede, Netherlands [19], First Affiliated Hospital of Shantou University, Guangdong Province, China, and BUS images obtained from the Breast Ultrasound Lesions Dataset (Dataset B) [20]. Information about the number of images used for training, validation and testing from each source is presented in Table. 1. For training and testing the images of size 128×128 with a batch size of 32, an Adam optimizer with a learning rate of 0.0001 was used to train the WGAN. These were chosen in par to the system and memory constraints. The experiments were performed on a workstation with $2\times$ Intel Xeon E2620 v4 CPU, 64GB RAM, and Nvidia Tesla K40 GPU. The summary of model parameters of the generator, discriminator and RDA-UNET-WGAN are presented in Table 2.

Figure 7 presents the segmentation results on the test data set. For more intuitiveness, various regions are differentiated by colour on the BUS images to analyze the results as compared to the ground truths. It can be seen that the segmentation results are remarkably close to the ground standards. To further qualitatively appreciate the performance of the model, it has been compared with the segmentation results of FCN8s [3], Segnet [21], Unet [4] and RDA-U-net [9]. Figure 8 presents ground truths and the comparison between the segmentation results of these models with the proposed model. It can be clearly seen that the proposed models results are close to the gold standards. The over-segmentation compared to RDAU-net is reduced. Also, the results are smoother with

Fig. 7 Segmentation results of the RDA-UNET-WGAN (The first three images are from the First Affiliated Hospital of Shantou University and the rest are from Dataset B). Column **a**, **b** and **c** denote the BUS image, the ground truth mask and predicted segmentation mask with various regions marked respectively. Regions marked in purple represent the intersection of predicted segmentation and the ground truth, red represents the lesion area not predicted as lesion and blue represents the non-lesion regions classified as lesion by the model)
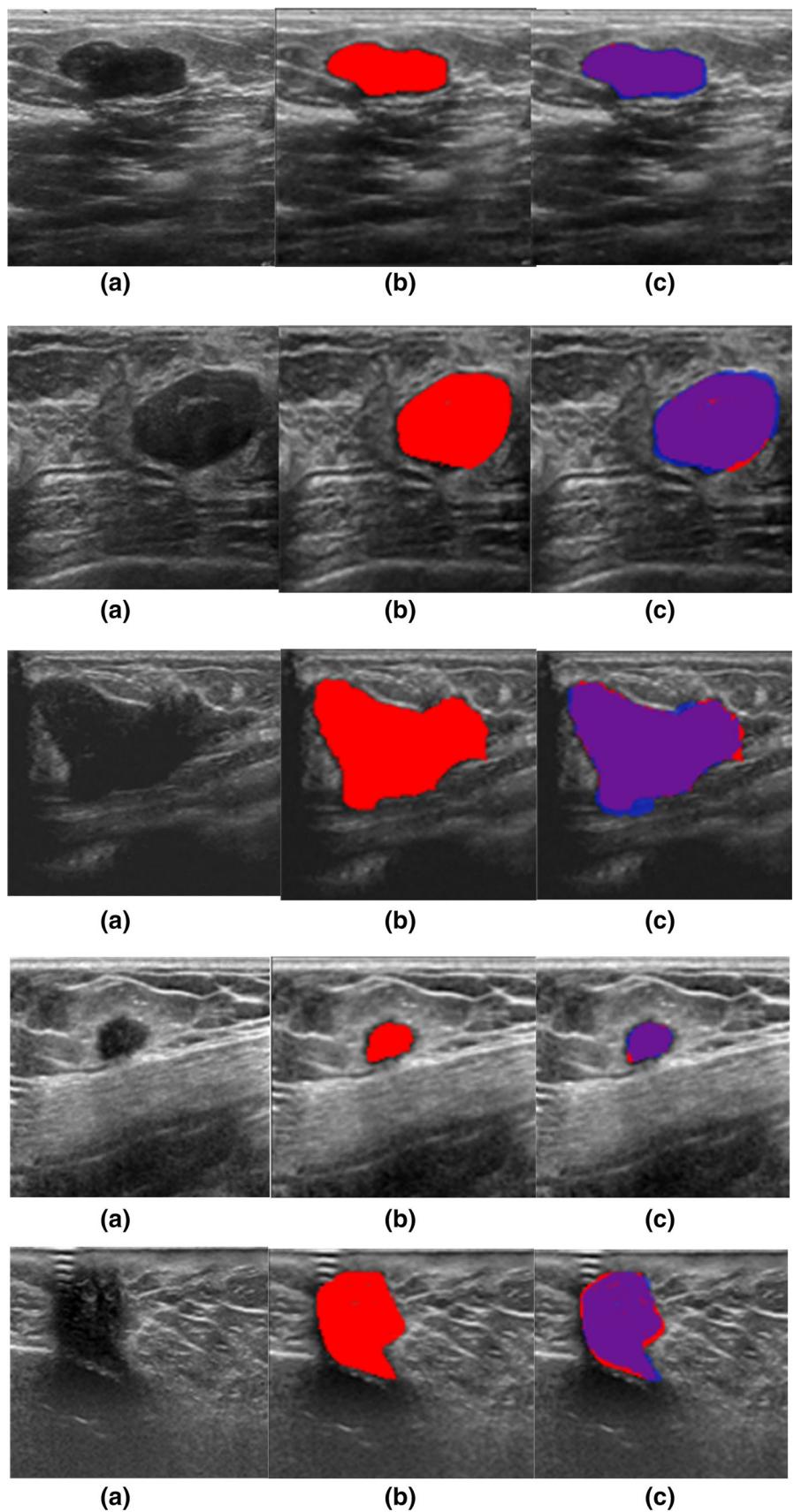
Fig. 8 Comparison with segmentation results of state-of-the-art models (images **a**, **b** and Malignant are from the First Affiliated Hospital of Shantou University, and images **d**, **e**, **f** and **g** are obtained from dataset B. Row **a**, **b**, **c**, **d**, **e**, **f**, and **g** presents the BUS image, the ground truth, the RDA-UNET-WGAN, RDAU-Net, Unet, Segnet, and FCN8s predicted segmentation maps respectively. Regions marked in purple represent the intersection of predicted segmentation and the ground truth, red represents the lesion area not predicted as lesion and blue represents the non-lesion regions classified as lesion by the model)
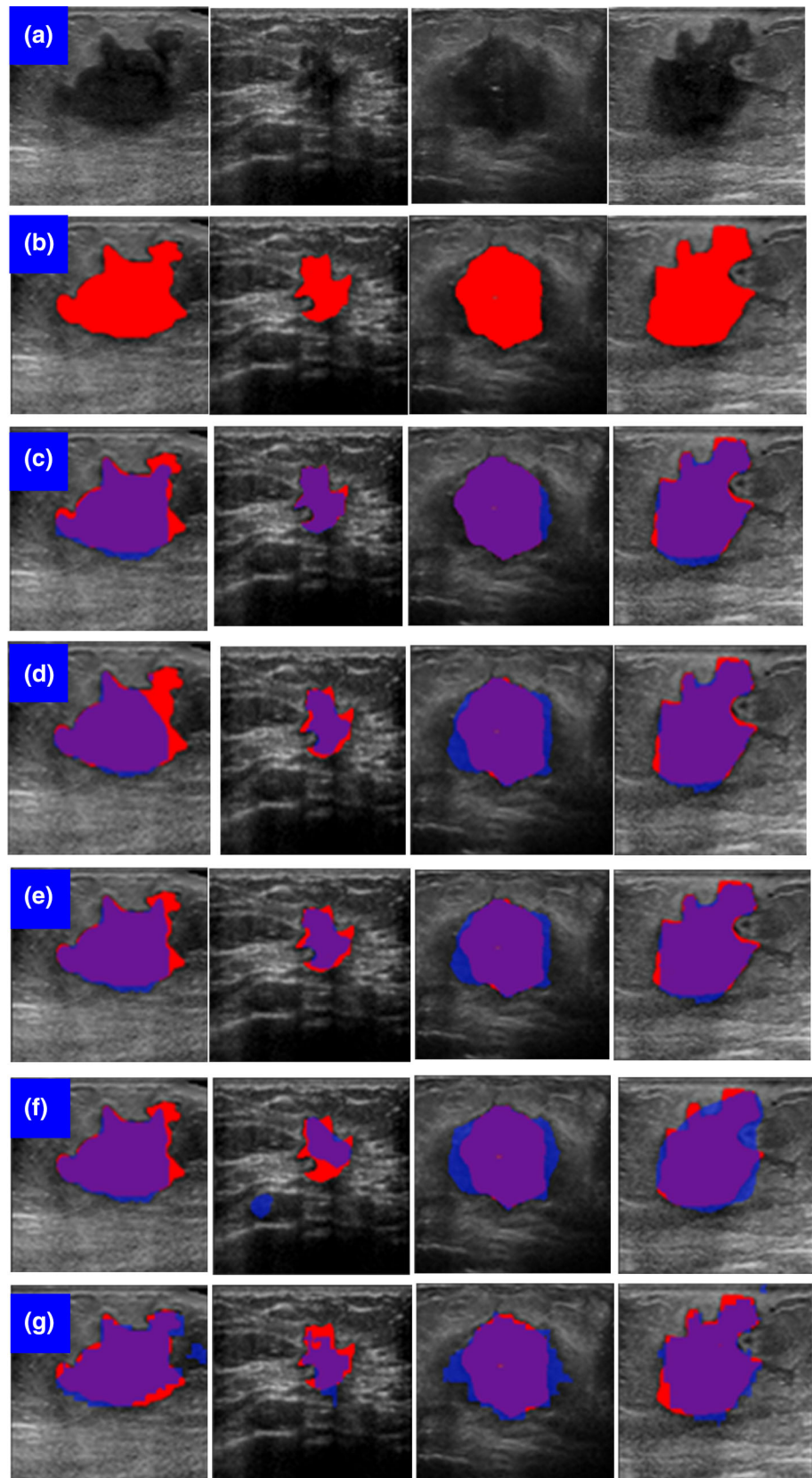
Table 3 Evaluation metric values for segmentation results from different models on the testing dataset (Combining the dataset from First Affiliated Hospital of Shantou University and Dataset B. Abbreviations used are: *M-IOU* Mean-intersection-over-union, *DSC* dice similarity coefficient, *AUC* area under the curve, *PR-AUC* Precision Recall AUC, *ROC* AUC-receiver operating characteristic AUC)

	RDA-UNET-WGAN	RDA-UNET [9]	Segnet [21]	UNET [4]	FCN8s
Loss	0.1159	0.1530	0.1830	0.1768	0.3527
Accuracy	0.9808	0.9791	0.9756	0.9779	0.9551
Sensitivity	0.8837	0.8363	0.8371	0.8445	0.7088
Specificity	0.9926	0.9926	0.9801	0.9907	0.9688
Precision	0.9117	0.8863	0.8143	0.8262	0.6127
F1Score	0.8975	0.8605	0.8255	0.8352	0.6572
M-IOU	0.8844	0.8053	0.8015	0.7983	0.7013
DSC	0.8841	0.8461	0.8120	0.8031	0.6472
PR-AUC	0.9542	0.9227	0.9538	0.9261	0.9148
ROC-AUC	0.8911	0.8551	0.8803	0.8892	0.8061
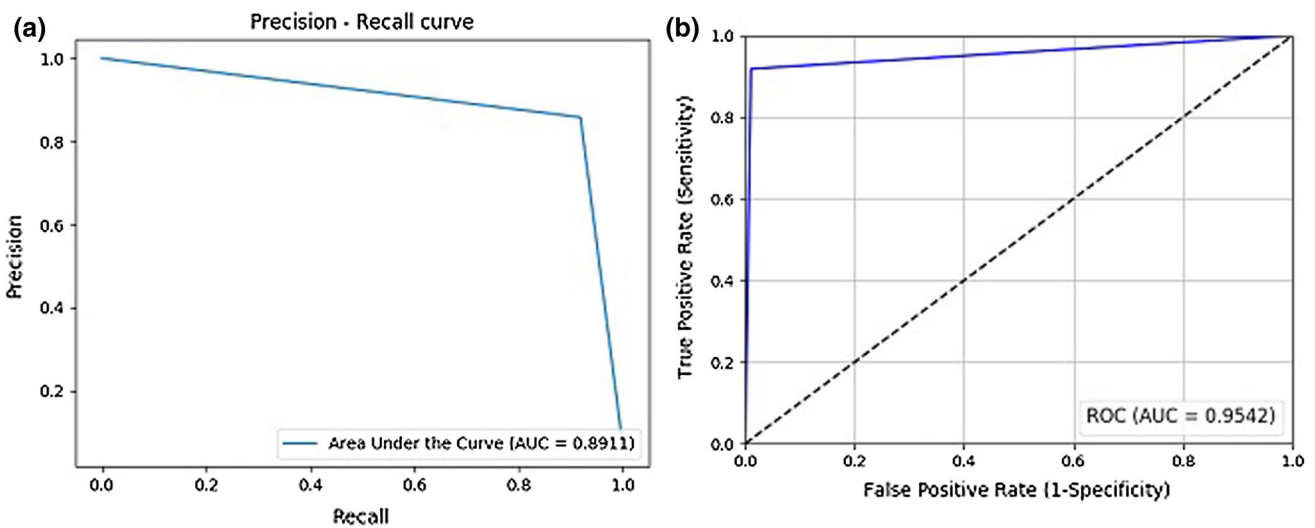
Fig. 9 PR-curve and ROC (**a** and **b** are the PR-curve and ROC of the RDA-UNET-WGAN on the testing dataset. The AUC values for each are also mentioned. Abbreviations used are: *PR* Precision-Recall, *ROC* receiver operating characteristic and *AUC* area under the curve)

good boundaries as compared to the other models. Segnet produces more generalized and rounded boundaries which are not very reliable. The U-Net fails at segmenting the least prominent tumor regions. FCN8s produces poor results are spiky and uneven edges.

For quantitative evaluation, the proposed model has been compared with results produced by FCN8s, Segnet, Unet and RDAU-net. The metrics specified in Sect. 2.3 have been used for evaluation. Table 3 presents the evaluation metric results on the test dataset (combining the dataset from First Affiliated Hospital of Shantou University and Dataset B) by the above models. The size of the testing dataset used is approximately 30% of the training dataset. In almost all the metrics, the proposed model outperforms the others. It can also be clearly seen that U-Net based architecture perform better than FCNs in segmentation tasks. Figure 9 presents the PR and ROC curve for the WGAN on the testing dataset.

From the segmentation results obtained and the comparisons, it is affirmed that the WGAN improves the segmentation results. This is perceived from both the qualitative and quantitative analysis.

4 Conclusion

GANs have been used extensively to generate fake data for several applications such as dataset creation and image editing. Recently, GANs have been utilized for image segmentation problems. Although GANs introduce the problem of non-convergence and diminishing gradient, they significantly improve the performance of segmentation tasks. With adversarial training and discriminator inputs, more clear and accurate segmentation results can be achieved. In this paper, we address tumor segmentation from Breast Ultrasound images. We propose a novel WGAN-based approach to this problem, and adversarial train the segmentation model to generate tumor masks close to the ground truth. The proposed model requires further optimisation to combat converging

parameters after training. The proposed method outperforms other state of the art approaches in both qualitative and quantitative analysis. Using GANs improved the precision by 3–4%, Mean IOU by 6% and the dice score by 5%. The proposed model is highly sensitive to hyperparameter selections and can be further optimised to get even improved results. Experimenting the model with other medical imaging datasets and making the performance robust still remains as future work.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. National breast cancer foundation inc: Breast cancer facts. http://www.nationalbreastcancer.org/breast-cancer-facts Accessed 01 Nov 2019
2. Xu, Y.; Wang, Y.; Yuan, J.; Cheng, Q.; Wang, X.; Carson, P.L.: Medical breast ultrasound image segmentation by machine learning. Ultrasonics **91**, 1–9 (2019)
3. Long, J.; Shelhamer, E.; Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440 (2015)
4. Ronneberger, O.; Fischer, P.; Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241. Springer, Berlin (2015).
5. Xia, X.; Kulis, B.: W-net: a deep model for fully unsupervised image segmentation. *arXiv preprint* arXiv:1711.08506 (2017)
6. Tong, G.; Li, Y.; Chen, H.; Zhang, Q.; Jiang, H.: Improved u-net network for pulmonary nodules segmentation. Optik **174**, 460–469 (2018)
7. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B. et al.: Attention u-net: learning where to look for the pancreas. *arXiv preprint* arXiv:1804.03999 (2018)
8. Zhuang, Z.; Raj, A.N.J.; Jain, A.; Ruban, N.; Chaurasia, S.; Li, N.; Lakshmanan, M.; Murugappan, M.: Nipple segmentation and localization using modified u-net on breast ultrasound images. J. Med. Imag. Health Inf. **9**(9), 1827–1837 (2019)
9. Zhuang, Z.; Li, N.; Raj, A.N.J.; Mahesh, V.G.V.; Qiu, S.: An RDAU-NET model for lesion segmentation in breast ultrasound images. PloS One **14**(8), e0221535 (2019)
10. Goodfellow, I.P.-A.; Mirza, J.; Xu, M.; Warde-Farley; Bing, O.; David, C.; Sherjil, B.; Aaron, Y.: Generative adversarial nets. *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
11. Salome, K.; Christoph, B.; Arjan, K.; Bram, van G.; Nassir Navab, S.; Albarqouni, A.; Mukhopadhyay: GANs for medical image analysis. *arXiv preprint* arXiv:1809.06222 (2018)
12. Xin, Y.; Ekta, W.; Paul, B.: Generative adversarial network in medical imaging: a review. Med Image Anal. **58** (2019)
13. Li, Z.; Wang, Y.; Yu, J.: Brain tumor segmentation using an adversarial network. In: International MICCAI brainlesion workshop, pp. 123–132. Springer, Berlin (2017)
14. Hong, W.; Wang, Z.; Yang, M.; Yuan, J.: Conditional generative adversarial network for structured domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1335–1344 (2018)
15. Zhongyi, H.; Benzheng, W.; Ashley, M.; Stephanie, L.; Shuo, L.: Spine-GAN: semantic segmentation of multiple spinal structures. *Med. Image Anal.*, pp. 23–35, 50 (2018)
16. Son, J.; Park, S.J.; Jung, K.-H.: Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint* arXiv:1706.09318 (2017)
17. Arjovsky, M.; Chintala, S.; Bottou, L.: Wasserstein gan. *arXiv preprint* arXiv:1701.07875 (2017)
18. Enokiya, Y.; Iwamoto, Y.; Chen, Y.-W.; Han, X.-H.: Automatic liver segmentation using u-net with wasserstein GANs. J Image Graph **6**(2) (2018)
19. Ultrasound cases info. https://www.ultrasoundcases.info/cases/breast-and-axilla/malignant-breast-lesions/ Accessed 15 Aug 2019
20. Yap, M.H.; Pons, G.; Martí, J.; Ganau, S.; Sentís, M.; Zwiggelaar, R.; Davison, A.K.; Martí, R.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inf. **22**(4), 1218–1226 (2017)
21. Badrinarayanan, V.; Kendall, A.; Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)