Breast Tumor Segmentation in Ultrasound Images using Contextual-Information-Aware Deep Adversarial Learning Framework

Vivek Kumar Singh[a,*], Mohamed Abdel-Nasser[a,c], Farhan Akram[b], Hatem A.Rashwan[a], Md. Mostafa Kamal Sarker[a], Nidhi Pandey[d], Santiago Romani[a], Domenec Puig[a]

[a]*Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, Spain.*
[b]*Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, Abu Dhabi, 127788, UAE*
[c]*Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt*
[d]*Department of Medicine and Health Sciences, University Rovira i Virgili, Reus 43204, Spain*

Abstract

Automatic tumor segmentation in breast ultrasound (BUS) images is still a challenging task because of many sources of uncertainty, such as speckle noise, very low signal-to-noise ratio, shadows that make the anatomical boundaries of tumors ambiguous, as well as the highly variable tumor sizes and shapes. This article proposes an efficient automated method for tumor segmentation in BUS images based on a contextual information-aware conditional generative adversarial learning framework. Specifically, we exploit several enhancements on a deep adversarial learning framework to capture both texture features and contextual dependencies in the BUS images that facilitate beating the challenges mentioned above. First, we adopt atrous convolution (AC) to capture spatial and scale context (i.e., position and size of tumors) to handle very different tumor sizes and shapes. Second, we propose the use of channel attention along with channel weighting (CAW) mechanisms to promote the tumor-relevant features

*The Spanish Government partly supported this research through project PID2019-105789RB-I00.

*Corresponding author

Email addresses: vivekkumar.singh@urv.cat (Vivek Kumar Singh), mohamed.abdelnasser@urv.cat (Mohamed Abdel-Nasser), farhan.akram@ku.ac.ae (Farhan Akram), hatem.abdellatif@urv.cat (Hatem A.Rashwan), mdmostafakamal.sarker@urv.cat (Md. Mostafa Kamal Sarker), nidhi.pandey3790@gmail.com (Nidhi Pandey), santiago.romani@urv.cat (Santiago Romani), domenec.puig@urv.cat (Domenec Puig)

(without extra supervision) and mitigate the effects of artifacts. Third, we propose to integrate the structural similarity index metric (SSIM) and L1-norm in the loss function of the adversarial learning framework to capture the local context information derived from the area surrounding the tumors. We used two BUS image datasets to assess the efficiency of the proposed model. The experimental results show that the proposed model achieves competitive results compared with state-of-the-art segmentation models in terms of Dice and IoU metrics. The source code of the proposed model will be publicly available at https://github.com/vivek231/Breast-US-project.

Keywords: Breast Cancer, CAD System, Deep Adversarial Learning, Ultrasound Image Segmentation

1. Introduction

Breast cancer is one of the most common causes of women's death worldwide (Siegel et al., 2017). Screening with mammography can reveal tumors at early stages. However, it is difficult to detect breast cancer using mammograms when the breast tissue has a high density. Ultrasound scan is a cheap imaging alternative and a powerful adjunctive screening tool: it is recommended by the physicians to analyze and successfully detect the occluded breast tumors in the presence of high-density breast tissue (Lauby-Secretan et al., 2015). Although ultrasound scans can reveal some tumors unnoticeable in mammograms, some scans are hard to read due to their noisy nature and the high similarity between healthy and unhealthy tissue. An expert sonographer is needed to retrieve tumor information from the noisy ultrasound scans. Considering the numbers of ultrasound scans a doctor has to view, it is a quite cumbersome and time-consuming task. In this context, a computer-aided diagnosis (CAD) system can reduce the workload of sonographers (Doi, 2007), by providing valuable cues for diagnosis, such as the possible location of tumors, their plausible boundaries and a prediction of the tumor type. One of the main tasks of BUS CAD systems is tumor segmentation, which is the focus of this article.

Various segmentation algorithms based on traditional computer vision have been introduced in the literature to segment breast tumors in ultrasound images. These ap-

proaches include active contours (Rodtook and Makhanov, 2013), (Huang and Chen, 2006), thresholding (Xian et al., 2015), clustering (Lo et al., 2014), (Shan et al., 2012), graph-based (Huang et al., 2012), and watershed-based automated segmentation method (Gu et al., 2016). A computationally efficient segmentation algorithm was proposed in (Horsch et al., 2001), which maximizes a utility function over partition margins defined through gray-value thresholding of a preprocessed image. Even though this threshold technique is computationally competent, parameters such as center, height, and width of initial lesion margins should be set manually for more reliable segmentation results, which makes it less robust if constant parameters are set for different variety of images. An active contours model that incorporates the watershed algorithm in the objective function was proposed in (Huang and Chen, 2006). This method tries to minimize the energy function to segregate the boundaries of the tumor region properly.

With technological advancement and the great success of *deep learning paradigm*, several deep learning-based segmentation methods have been proposed recently in the context of image segmentation (Lai, 2015). Many researchers have used semantic segmentation methods for different medical imaging modalities, such as brain lesion segmentation in magnetic resonance imaging (MRI) (Litjens et al., 2017), skin lesion segmentation in dermoscopy images (Singh et al., 2019), breast tumor segmentation in mammograms (Singh et al., 2020, 2018b) and optic disc segmentation in retinal fundus images (Singh et al., 2018a).

Over the last two decades, several BUS image segmentation methods have been proposed, which can be categorized into semi-automated and fully automated according to the degree of human intervention. In (Abdel-Nasser et al., 2017), a region growing algorithm used to automatically extract the region of interest (ROI) of the tumor. Image super-resolution and texture analysis methods were both incorporated in their approach to discriminate benign tumors from the malignant ones. Recently, some deep learning-based models proposed to enhance the performance of breast tumor segmentation. In (Xu et al., 2019), two convolutional neural networks (CNN) architectures have been employed to segment BUS images into the skin, mass, fibro-glandular, and fatty tissues achieving an accuracy of 90%. In (Hu et al., 2019) joined a dilated fully convolutional network with a phase-based active contour model to segment breast tumors,

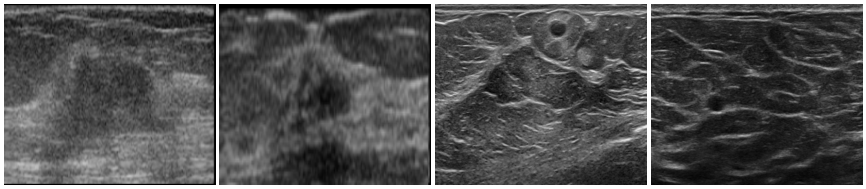

Figure 1: Examples of BUS images with presence of ambiguous anatomical boundaries due to speckle noise, low contrast and low signal-to-noise ratio.

achieving a dice score of 88.97%.

Moreover, a residual-dilated-attention-gate-UNet (RDAU-NET) model was suggested in (Zhuang et al., 2019) to segment the tumors in BUS images. This model was based on the baseline U-Net segmentation model. However, the conventional neural units were replaced with residual units to increase the boundary information and mitigate the deep network performance degradation issue. In that network, the dilated convolutions process the feature maps obtained from the encoder layers to maximize the receptive field and extract more characteristic features. The conventional cropping and copying between the encoder-decoder pipelines were substituted with the *attention gate* modules, which enhanced the learning capabilities through suppression of background information.

Although the proposed methods in the literature provide useful solutions, there are still difficulties in accurately segmenting tumors due to speckle noise and shadows that exist in BUS images, as well as the high variability of tumors in shape, size, appearance, texture, and location. As shown in Fig. 1, the four BUS images have speckle noise, shadows, and different appearance and texture. Fig. 1(a) and Fig. 1(b) contain big masses in the middle of the images. These masses have ill-defined and irregular boundaries. In turn, the masses in Fig. 1(c) and Fig. 1(d) are at different locations and have clear boundaries. In this article, we propose an automatic method for breast tumor segmentation in BUS images based on the adversarial learning framework. Specifically, the main contributions of this article are as follows:

1. We propose an efficient model for breast tumor segmentation in BUS images using a contextual-information-aware deep adversarial learning framework, in

which we adopt a deep learning paradigm to capture both the texture features and the contextual dependencies in the BUS images.

2. We adopt atrous convolution (AC) block (Chen et al., 2017) to capture both spatial context and scale context, i.e., position and size of breast tumors. The proposed AC block solves this issue by increasing the receptive field with different atrous rates while maintaining the spatial dimension of feature maps at a very low computational burden.
3. Furthermore, we propose the use of a channel attention (Fu et al., 2018) along with channel weighting (Hu et al., 2018) mechanisms (hereinafter, we refer to it as CAW block), to promote the tumor-relevant features (without extra supervision), thus making the segmentation framework more robust against the mentioned artifacts common in BUS images.
4. Finally, we propose a combination of structural similarity index metric (SSIM) and L1-norm in the loss function of the adversarial learning framework. Such combination can address the challenges induced by the mentioned artifacts (e.g., speckle noise, irregular and ambiguous boundaries of tumors) and facilitate capture of the local context information derived from the area that surrounds breast tumors.

The rest of this article is structured as follows. Section 2 presents the proposed model. The results are shown in section 3. The article is concluded, and some possible future works are discussed in section 4.

2. Proposed Methodology

In this article, we aim at segmenting tumors from a given 2D raw input BUS image I of size $W \times H$, where W and H are width and height of the image, respectively. We introduce a deep learning model, which outputs a label map of size $W \times H$ whose pixels $p(x,y)$ hold a label 1 for lesion and 0 otherwise. As depicted in Fig. 5, the proposed segmentation model is based on a conditional generative adversarial network (cGAN), which comprises of two interdependent networks: a generator (G) and a discriminator (D) (Goodfellow et al., 2014). The generator produces a generated label map (i.e.,

predicted mask) from an input BUS image, while the discriminator is trained to discern between ground-truth (real) and generated label maps (predicted).

In the cGAN model, the training of both the generator and the discriminator is under the observation of the input BUS images as a condition. Below, we explain the model in detail and how we use AC and CAW techniques to capture texture features and contextual dependencies in BUS images for detecting and segmenting breast tumors accurately.

2.1. Atrous convolution (AC) block

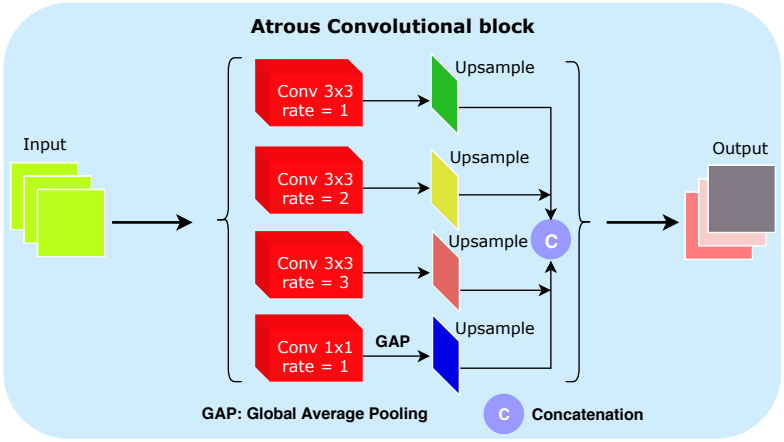

Figure 2: The architecture of the AC block with different rates of AC (r = 1, 2 and 3).

AC helps to manage the resolution of the feature responses computed from CNN. They can incorporate local as well as global contextual interactions (pixel, region, tumor, and tumor-background interactions), without increasing the number of trainable parameters, by enlarging the field-of-view of filters. Fig. 2 presents the architecture of the utilized AC block. The first three convolutional layers have a kernel size of 3×3 and rates of 1, 6, and 9, respectively. The fourth convolutional layer has a kernel size of 1×1 followed by a global average pooling (GAP). This layer is employed to assure that high-level feature maps of small size have features from the region of breast tumors. An up-sampling layer is employed after each branch and then all features are concatenated. As shown in Fig. 3, AC can increase the receptive field, and thus it ac-

commodates to variable size and shapes of breast tumors. The employed AC block can mitigate the problem of loss of small tumor-relevant features due to the consecutive downsampling layers.

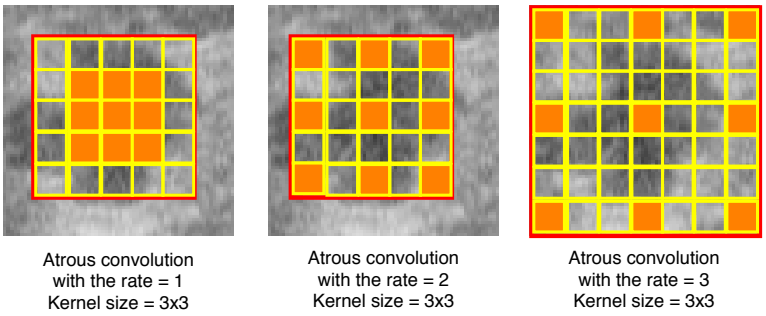

Figure 3: Use of AC to increase the receptive field in order to accommodate variable size and shapes of breast tumors.

2.2. CAW block

Fig. 4 describes the architecture of the proposed CAW block. It has two branches: the channel attention process (top branch) and the channel weighting process (bottom branch). Since we have placed the CAW block after the last encoder layer, the processed activation map has spatial dimensions (H x W) of 1×1: indeed, it is a vector of C= 512 scalars. Hence, the method works only on the channel feature space.

The attention mechanism (Fu et al., 2018) computes a feature correlation matrix of $C \times C$ elements, as the multiplication of the input features vector by its transpose. Then, the input vector is multiplied by the transpose of this matrix to get the tumor relevant features. Afterward, this vector is then multiplied by a learnable scalar parameter and summed to the original vector.

Assume that $X \in R^{C \times C}$ is the channel attention map and $\gamma \in R^{C \times H \times W}$ are the original features. Individually, we reshape γ to $\mathbb{R}^{C \times N}$, we then perform a matrix multiplication among γ and the transpose of γ. Finally, we apply a softmax layer to make the channel attention map $X \in \mathbb{R}^{C \times C}$:

$$x_{ji} = \frac{exp(\gamma_i . \gamma_j)}{\sum_{i=1}^{C} exp(\gamma_i . \gamma_j)} \tag{1}$$

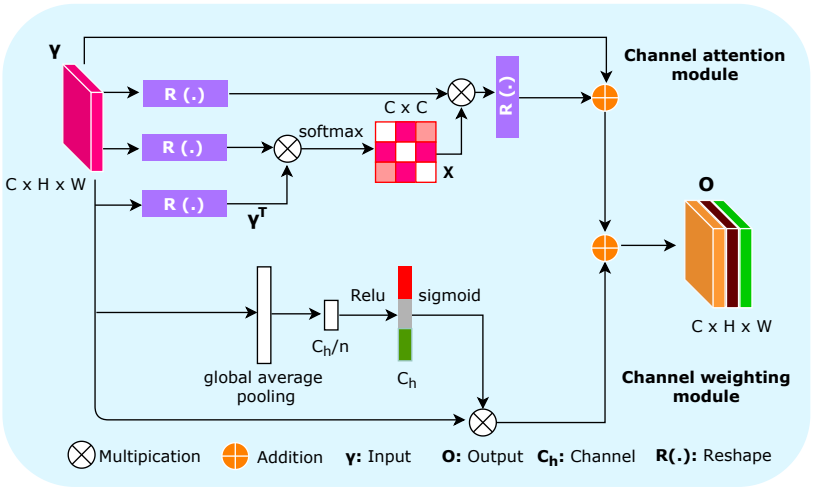

Figure 4: The architecture of the CAW block.

where x_{ji} estimate the i^{th} channel's impact on the j^{th} channel. Moreover, we implement a matrix multiplication among the transpose of X and γ and reshape their result to $\mathbb{R}^{C\times H\times W}$. Later we multiply the result by a scale parameter β and do an element-wise sum operation with γ to get the final output $E \in \mathbb{R}^{C\times H\times W}$:

$$E = \beta \sum_{i=1}^{C} (x_{ji}\gamma_i) + \gamma_j \tag{2}$$

where β is the weight factor. This mechanism can enhance the segmentation accuracy because relevant patterns similar to breast tumor regions will produce high activation values in several feature channels (i.e., it will enhance their importance in the output feature maps).

In channel weighting (Hu et al., 2018), the weighting mechanism begins with a global average pooling to transform each channel map into a single value (squeeze), but we can omit this step since we already have one value per channel. The next step involves two fully connected layers, with C/n where n is dimensional reduction ratio (n=4) and C neurons, respectively, that outputs C weights at each channel, which multiply the original vector values to dynamically promote breast tumors relevant features in each sample (excitation).

Finally, the output vectors of the two branches are summed. In the training phase,

the back-propagated gradient can flow through both branches that promote the tumor-relevant features for the final segmentation. In the network, the attention method promotes the high-level features that repeat their values for given input image patterns, which indicate that these shared features are very critical for the target classes (tumor and non-tumor pixels). In turn, the excitation method looks for the optimal non-linear re-calibration of the high-level features that tend to provide better inferences of the final output. In this manner, we can enrich the representational power of the highest-level features of the generator network, which turns out in a clear improvement compared to the baseline (BL) model.

2.3. Model architecture

Fig. 5 presents the network architecture of the proposed model. It consists of a generator network that extracts breast tumor relevant features, and a discriminator network that predicts if a label mask is a real or fake segmentation of the input BUS image. More details about these networks are given below.

Generator network: The generator network incorporates an encoder composed of seven convolutional layers and a decoder composed of seven deconvolutional layers (note that transposed Conv is named as Dconv). We have modified the plain encoder-decoder structure by inserting an AC block (Yu and Koltun, 2015) between *Conv3* and *Conv4*, in addition to a CAW block inserted between *Conv7* and *Dconv1*. It worth noting that the locations of AC and CAW blocks have been experimentally tuned. Indeed, the AC block helps the generator network to recognize breast tumors relevant features at different scales and to expand the actual receptive field of the filters. Consequently, the network is more aware of contextual information without increasing the number of parameters nor the amount of computation. In this study, we use 1, 6, and 9 dilation rates with kernel size 3×3 and a stride of 2. In turn, The CAW block is an aggregation of a channel attention (Fu et al., 2018) with a channel weighting module (Hu et al., 2018). In turn, the CAW block increases the representational power of the highest-level features of the generator network, which turns out in a clear improvement of the accuracy of the breast tumor segmentation in ultrasound images.

Each layer in the encoder section is followed by batch normalization (BN) (ex-

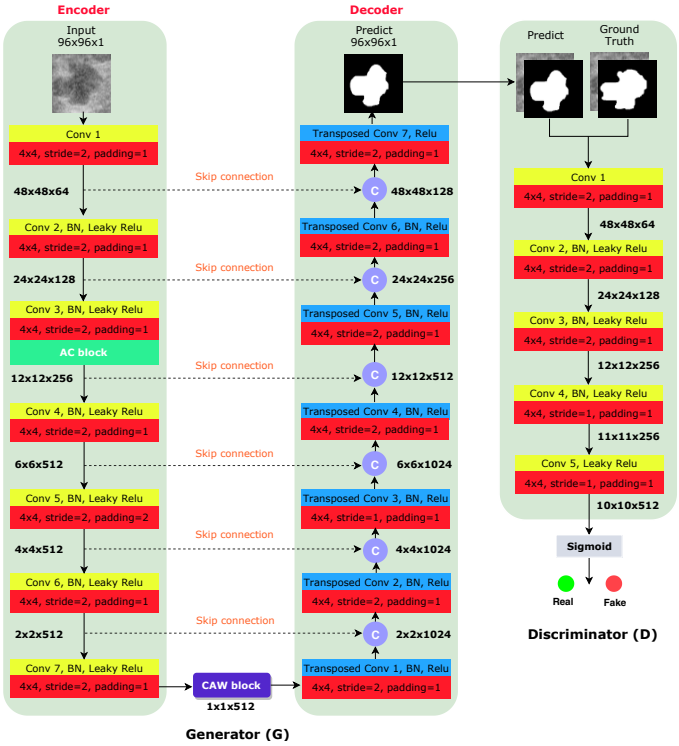

Figure 5: The architecture of the proposed model that consists of Generator (G) and Discriminator (D) networks.

cept for *Conv1*) and *LeakyReLU* with slope 0.2. The decoder section is a sequence of transposed-convolutional layers followed by batch normalization, and dropout with rate 0.5 (only in *DConv1, DConv 2* and *DConv3*) with *ReLU*. The filters of the convolutional and deconvolutional layers are defined by a kernel of 4×4, and they are shifted with a stride of 2. Skip connections are employed between the corresponding layers in the encoder and decoder sections, which improve the features in the output image by merging deep, coarse, semantic information and simple, fine, appearance information. After the last decoding layer (*Dconv7*), the *tanh* activation function is utilized as a non-linear output of the generator, which is trained to create a binary mask of the breast tumor.

Discriminator network: It comprises a set of five convolutional layers with kernels of size 4×4 with a stride of 2, except for *Conv4* and *Conv5* where the stride is 1.

The batch normalization is used after *Conv2* to *Conv4*. *LeakyReLU* with slope 0.2 is the non-linear activation function used after *Conv2* to *Conv5*, while the sigmoid function is used after *Conv5*. The input of the discriminator is the concatenation of the BUS image and a binary mask identifying the tumor area. Note that this mask can either be the ground truth or the one predicted by the generator network. The output of the discriminator is a 10×10 matrix having values varying from 0.0 (completely fake) to 1.0 (real).

Loss Functions: Assume x is a BUS image comprising a breast tumor, y is the ground truth mask of that tumor within the image, $G(x,z)$ and $D(x,G(x,z))$ are the outputs of the generator and the discriminator, respectively. The loss function of the generator G involves three terms: adversarial loss (binary cross-entropy loss), L1-norm to promote similarity between the generated mask and ground-truth mask, and SSIM loss (Wang et al., 2004) to enhance the shape of the boundaries of segmented masks:

$$\begin{aligned} \ell_{Gen}(G,D) = \mathbb{E}_{x,y,z}(-\log(D(x,G(x,z)))) \\ + \lambda \mathbb{E}_{x,y,z}(\ell_{L1}(y,G(x,z))) + \alpha \mathbb{E}_{x,y,z}(\ell_{SSIM}(y,G(x,z))) \end{aligned} \tag{3}$$

Here z is a random variable, and λ and α are empirical weighting factors. The variable z is introduced as a dropout in the decoding layers *Dconv1* to *Dconv3* at both training and testing phases, which helps generalize the learning process and avoid overfitting.

Structural Similarity Index Metric (SSIM) Loss: It can be defined as (Wang et al., 2004):

$$SSIM(a,b) = \frac{2(\mu_a \mu_b + c_1)(2\sigma_{ab} + c_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)} \tag{4}$$

where a and b are two 2D images to be compared (in our case, two binary masks), and μ_a, μ_b, σ_a and σ_b are the corresponding mean and standard deviations, while σ_{ab} is the covariance between a and b. The value of c_1 and c_2 were experimentally set to 0.012 and 0.032, respectively.

Notably, the values of $D(x,G(x,z))$ should approach 1.0 if the generator network is well optimized, indicating that discriminator cannot differentiate between produced tumor masks from ground truth masks. In turn, L1 and SSIM losses should approach 0.0, meaning that every generated mask matches the corresponding ground truth both

in overall pixel-to-pixel distances (L1-norm) and in basic statistic descriptors (SSIM). The loss function of the discriminator D can be expressed as follows:

$$\ell_{Dis}(G,D) = \mathbb{E}_{x,y,z}(-\log(D(x,y))) + \mathbb{E}_{x,y,z}(-\log(1-D(x,G(x,z)))) \quad (5)$$

During the training phase, the optimizer tries to fit D in order to maximize the loss values for ground truth masks (by minimizing $-\log(D(x,y))$) and minimize the loss values for created masks (by minimizing $-\log(1-D(x,G(x,z))$). The two terms calculate BCE loss using both masks, assuming that the expected class for ground truth and generated masks are 1 and 0, respectively. Notably, the G and D networks are concurrently optimized: an optimization step for both networks at each iteration, where G tries to generate a correct tumor segmentation, and D learns how to distinguish between generated and real masks.

Model training: In the pre-processing step, we resized each BUS image to 96×96 pixels, and normalized pixel values between [0, 1]. The hyperparameters of the model were experimentally tuned. We explored different optimizers, such as SGD, AdaGrad, Adadelta, RMSProp, and Adam, with different learning rates, obtaining the best results with Adam optimizer (β_1= 0.5, β_2= 0.999) and learning rate = 0.0002 with a batch size of 8. The SSIM loss and L1-norm loss weighting factors λ and α were set to 10 and 5, respectively. We have trained the model from scratch for 200 epochs, getting the best results for both generator and discriminator on 40 epochs. In a post-processing step, morphological operations (3×3 closing and 2×2 erosion) are applied to the generated masks in order to remove mislabeled pixels.

3. Experiments and Discussion

In this study, we use two different BUS image datasets to assess the performance of the proposed method. Both datasets are obtained with different BUS imaging device specifications at different acquisition times. In this article, we refer to the two datasets used as dataset A and dataset B.

Dataset A: It includes BUS images from Mendeley dataset (Rodrigues, 2017), which contains 150 malignant and 100 benign tumors cases. This dataset has no ground

truth for tumor segmentation. Therefore, a cooperative expert radiologist has manually segmented the tumors in the BUS images, which are used as ground truth to train our proposed segmentation model.

Dataset B: It is provided by the UDIAT Diagnostic Centre of Sabadell (Spain) (Yap et al., 2017), (Abdel-Nasser et al., 2015). The images were collected from 267 patients using a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz), where BUS images of 163 patients include masses and the others are of healthy patients. The ground truths of the tumor regions in the BUS images are already available in this dataset.

With both BUS image datasets, we have randomly split the datasets into training (70%), validation (10%), and testing (20%) sets. It is important noting that the proposed model is trained, validated, and tested using the train, validation, and test data of both datasets, individually.

Data augmentation: We implemented data augmentation techniques to improve the network performance, invariance and robustness. To augment each dataset, we performed the following operations: 1) we scaled the images by varying scaling variable from 0.5 to 2 with a step size of 0.25, 2) we applied gamma correction on the images by varying gamma scaling constant from 0.5 to 2.5 with a step size of 0.5, and 3) we horizontally and vertically flipped the images and rotated them with different angles. This data augmentation procedure increases the number of training images of dataset A and B from 175 and 118 to 8,225 and 5,546, respectively.

Evaluation Metrics: Five metrics are used to assess the performance of the segmentation models, namely the accuracy, Dice Coefficient, Jaccard index (IoU), sensitivity, and specificity. To compute these metrics we firstly define true positive (TP), false positive (FP) and false-negative (FN) rates. Let A and B are the ground-truth mask and the corresponding mask obtained by a segmentation model, respectively. TP rate, which is the area of the segmented part common in A and B, can be defined as $TP = A \cap B$. FP rate, which is the segmented region not belonging to A, can be expressed as $\overline{A} \cap B$. FN rate, which is the true area missed by the segmentation model, can be expressed as $A \cap \overline{B}$. The five metrics can be formulated as follows: (Lalande

Table 1: Analyzing different configurations of the proposed method with Dataset A and Dataset B.

Methods	Dataset A					Dataset B				
	Accuracy	Dice	IoU	Sensitivity	Specificity	Accuracy	Dice	IoU	Sensitivity	Specificity
cGAN	0.9811	0.8604	0.7756	0.8385	0.9827	0.9756	0.8241	0.7562	0.8572	0.9807
cGAN+AC	0.9762	0.8714	0.8077	0.8550	0.9839	0.9803	0.8354	0.7712	0.8846	0.9869
cGAN+CAW	0.9818	0.9065	0.8340	0.8772	0.9911	0.9817	0.8479	0.7802	0.8892	0.9884
Proposed	**0.9926**	**0.9376**	**0.8882**	**0.9011**	**0.9973**	**0.9822**	**0.8682**	**0.8037**	**0.9155**	**0.9949**

et al., 2015).

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \tag{6}$$

$$Dice = 2TP/(2TP+FP+FN) \tag{7}$$

$$IoU = TP/(TP+FP+FN) \tag{8}$$

$$Sensitivity = TP/(TP+FN) \tag{9}$$

$$Specificity = TN/(TN+FP) \tag{10}$$

Implementation platform: The proposed BUS image segmentation method is implemented using Python 3.6, CUDA 8.0, cuDNN 7.0, PyTorch 0.4.1 running on a 64-bit Ubuntu operating system using a 3.4 GHz Intel Core-i7 with 16 GB of RAM and NVIDIA GTX 1070 GPU with a memory of 8 GB.

3.1. Ablation study

We performed an ablation study to show the effects of different blocks in the baseline cGAN model. Firstly, we trained a baseline cGAN model without any AC and CAW (channel attention and weighting) blocks. Then we added an AC block after the third encoder layer of the baseline cGAN model and called this model cGAN+AC. In the next step, we added the CAW block to the encoder layers of the generator network in the baseline cGAN model and called this model cGAN+CAW. In the last step,

we formulated the proposed model (cGAN+AC+CAW) by adding both AC and CAW blocks to the baseline cGAN model.

Table 1 presents a quantitative comparison between the cGAN, cGAN+AC, cGAN+CAW, and proposed (cGAN+AC+CAW) model, in terms of accuracy, Dice, Intersection over Union (IoU), sensitivity, and specificity, using both datasets. The results show that the baseline cGAN model yields Dice and IoU scores of 86.41% and 77.56%, respectively, for dataset A, whereas, 82.41% and 75.62% for dataset B, respectively. The baseline cGAN utilizes the basic network capability to segment the breast lesion from the noisy ultrasound images. Therefore, it cannot extract the low-resolution image features from the BUS images to properly segment the tumor. We added an AC block in the encoder part of the generator network of the baseline cGAN, which manages the resolution of the feature responses from the previous encoder layers. The AC block helps the encoding layers to extract the relevant features by ignoring the echogenic or shadows surroundings the lesion pixels. It provides a slight improvement in the Dice, IoU, sensitivity, and specificity on both datasets.

Furthermore, we evaluated the effects of the CAW block on baseline cGAN, which was added in the encoder part of the generator network. After checking various configurations, we combined both AC and CAW blocks in the baseline cGAN to formulate the proposed model. The proposed model provided an improvement of 8% to 11% compared to the baseline cGAN with Dataset A and around 2% with Dataset B in terms of Dice and IoU, respectively. Additionally, the added blocks provided more powerful pixel discriminability between the tumor and non-tumor areas. Fig. 6 presents examples of channel attention maps (CAMs) of BUS images with tumors. The CAMs shown in the figure is taken from the third encoding layer that shows the maps correlation with tumor and healthy tissue regions. The red color indicates a higher probability of the presence of a breast tumor while the blue color indicates a lower probability of the existence of a breast tumor. As shown, the CAW block help to focus on the regions of breast tumors while giving less importance to other regions.

Fig. 7 presents the performance of the proposed model with different combinations of loss functions: BCE, BCE+L1-norm, BCE+SSIM, BCE+Lovasz-Hinge and BCE+L1-norm+SSIM loss. As shown all loss functions gives a dice score higher

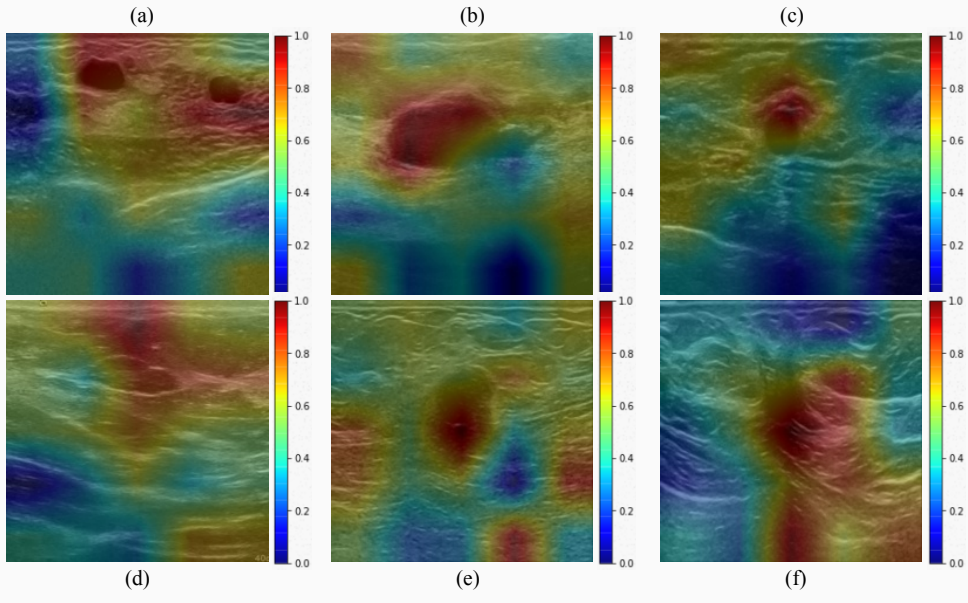

Figure 6: Six examples of Channel attention maps generated from the proposed segmentation model. The red color indicates a higher probability of the presence of a breast tumor while the blue color indicates a lower probability.

than 80%, however, BCE, BCE+L1-norm and BCE+Lovasz-Hinge losses achieves an IoU score lower than 80%. The best dice and IoU scores are archived with BCE+L1-norm+SSIM loss, and therefore it has been utilized with the proposed model.

3.2. Breast tumor segmentation results

In Table 2, we compared the proposed model with six state-of-the-art image segmentation methods: FCN (Long et al., 2015), UNet (Ronneberger et al., 2015) SegNet (Badrinarayanan et al., 2017), ERFNet (Romera et al., 2018), and DCGAN (Kim et al., 2017). All methods are evaluated both quantitatively and qualitatively. For the quantitative analysis, we have calculated accuracy, Dice, IoU, sensitivity, and specificity metrics. In Table 2, using dataset A, the proposed method yielded Dice and IoU scores of 93.76% and 88.82%, respectively. The proposed method outperformed the second-best segmentation method in Table 2, i.e., cGAN in terms of Dice and IoU by 8% and 11%, respectively, which is quite significant considering their proximity to the maximum value. The SegNet and ERFNet models yielded the worst segmen-

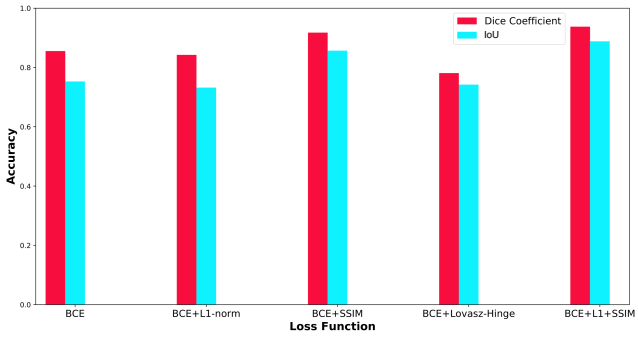

Figure 7: The performance of the proposed model with different combinations of loss functions.

Table 2: Comparison between the proposed model and six state-of-the-art methods in terms of accuracy, Dice, IoU, sensitivity and specificity, using datasets A and B.

Methods	Dataset A					Dataset B				
	Accuracy	Dice	IoU	Sensitivity	Specificity	Accuracy	Dice	IoU	Sensitivity	Specificity
FCN	0.9702	0.7973	0.6629	0.7649	0.9811	0.9791	0.7166	0.6227	0.7723	0.9879
SegNet	0.8369	0.5035	0.4165	0.7086	0.8698	0.8594	0.5144	0.4112	0.7235	0.8744
UNet	0.9710	0.8528	0.7706	0.7977	0.9794	0.9819	0.7521	0.6617	0.7925	0.9891
ERFNet	0.8849	0.6702	0.5332	0.7283	0.9116	0.9371	0.6010	0.4953	0.6954	0.9482
DCGAN	0.9758	0.8555	0.7728	0.8128	0.9803	0.9742	0.8007	0.7344	0.8323	0.9791
cGAN	0.9811	0.8604	0.7756	0.8385	0.9827	0.9756	0.8241	0.7562	0.8572	0.9807
Proposed	**0.9926**	**0.9376**	**0.8882**	**0.9011**	**0.9973**	**0.9822**	**0.8682**	**0.8037**	**0.9155**	**0.9949**

tation results. In turn, UNet and DCGAN yielded comparable segmentation results to cGAN, the second-best segmentation model, in terms of Dice and IoU. The proposed model also yielded the accuracy, sensitivity, and specificity of 99.26%, 90.11%, and 99.73%, respectively, that achieves an improvement of 1%, 6%, and 1.5% compared to the second-best segmentation model, i.e., cGAN.

Similarly, using dataset B, the proposed method yielded Dice and IoU scores of 86.82% and 80.37%, respectively, which is 4% and 5% more than Dice and IoU score of the second-best segmentation method, i.e., cGAN. For dataset B, only cGAN and DCGAN yielded acceptable segmentation results compared to the proposed method, with Dice and IoU scores over 80% and 70%, respectively. All remaining models: FCN, SegNet, UNet, and ERFNet, yielded Dice and IoU lower than 80% and 70%,

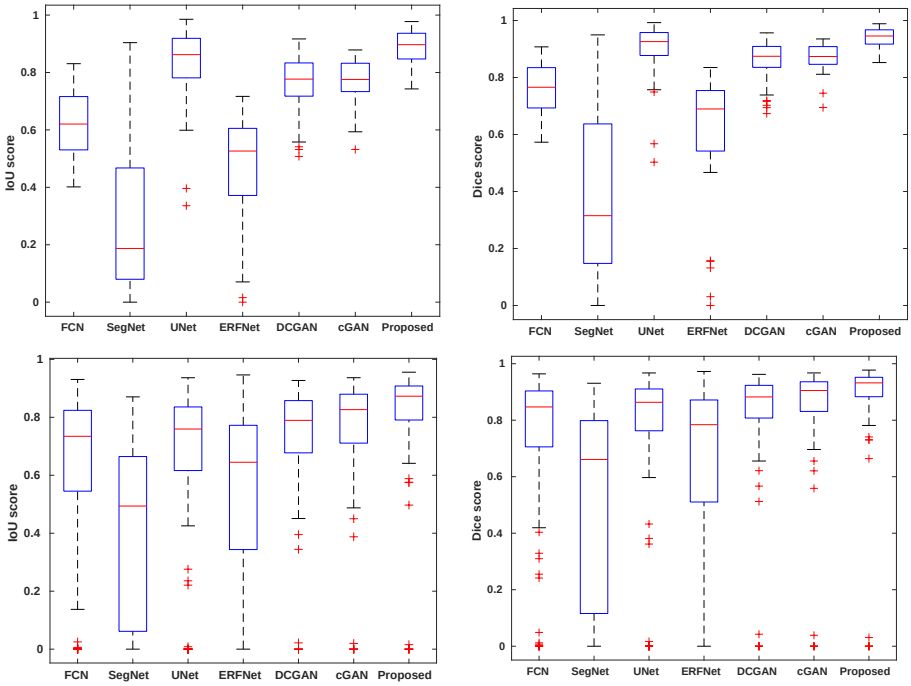

Figure 8: Boxplots of Dice and IoU scores for all test samples in Dataset A (plots in top) and Dataset B (plots in bottom). Different color boxes indicate the score range of several methods, the red line inside each box represents the median value, box limits include interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are computed as 1.5 times the distance of upper and lower limits of the box, and all values outside the whiskers are considered as outliers, which are marked with the (+) symbol.

respectively. The proposed model yielded the accuracy, sensitivity, and specificity of 98.22%, 91.55%, and 99.49%, respectively, outperforming the second-best cGAN model by 6% in terms of sensitivity. In turn, our model outperformed the second-best UNet model in terms of accuracy and specificity by very small values of 0.03% and 0.6%, respectively. Notably, the sensitivity and specificity of the proposed method, which is essential to tumor detection tasks, are higher than 90.0%. These results reveal that the proposed model not only achieves accurate segmentation results but also high tumor detection accuracy.

Fig. 8 has shown a comparison between the FCN, SegNet, ERFNet, UNet, DCGAN, cGAN and the proposed model by using the boxplots of Dice and IoU values

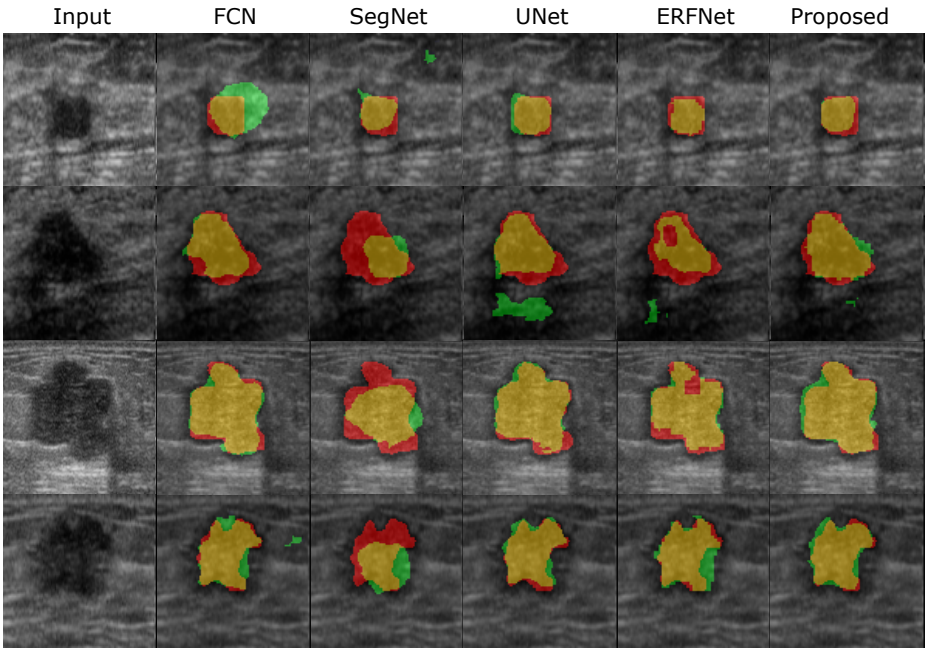

Figure 9: Segmentation results of five models using Dataset A. (Col 1) original input images, (Col 2) FCN, (Col 3) SegNet, (Col 4) UNet, (Col 5) ERFNet, and (Col 6) the proposed model. For each experiment, the original BUS image and the segmentation results of compared segmentation methods are shown. The segmentation results are color-coded in yellow (TP), red (FN), and green (FP), and background (TN).

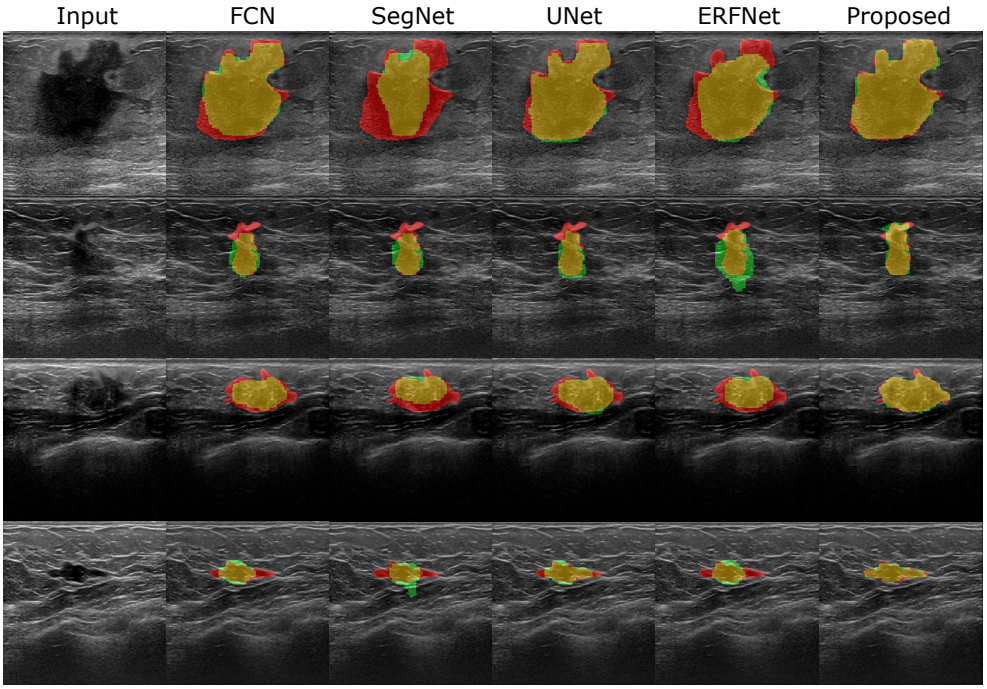

Figure 10: Segmentation results of five models with the Dataset B. (Col. 1) original input images, (Col. 2) FCN, (Col. 3) SegNet, (Col. 4) UNet, (Col. 5) ERFNet, and (Col. 6) the proposed model. For each experiment, the original BUS image and the segmentation results of compared segmentation methods are shown. The segmentation results are color-coded in yellow (TP), red (FN), green (FP), and background (TN).

obtained by using 50 and 32 testing samples of Dataset A and Dataset B, respectively. The two models based on GAN provided small ranges of Dice and IoU values, while other deep segmentation methods, FCN, SegNet, ERFNet, and UNet showed a more extensive range of values. There are many outliers in the segmentation results of compared state-of-the-art methods. In turn, the proposed method yielded no outliers for dataset A and only five outliers with Dataset B. Figure 8 have shown an apparent deviation between the proposed and the compared methods.

Fig. 9 and 10 presented a qualitative comparison between the segmentation results of the proposed and state-of-the-art methods using datasets A and B. It is evident from both figures that the SegNet and ERFNet yielded the worst segmentation results as the predicted regions of both methods contain high false negatives (in red) and false positive (in green) pixels than the other tested methods. FCN also shows rather significant inaccurate segmented areas, although it has fairly segmented the example in the second row. In turn, UNet provided proper segmentation, but it has a less accurate boundary around the tumor region. In turn, our model yielded the best segmentation with the highest TP and smallest FP and FN pixels among the five tested methods. Further visualizations of segmentation results of the proposed method can be found at https://youtu.be/jcNk7D8nyzs.

Moreover, we computed the execution time (inference time) of each segmentation model. FCN, UNet, SegNet, ERFNet, DCGAN, and the proposed model achieve 35.15, 20.33, 17.71, 78.78, 18.27, and 19.62 frames per second (FPS), respectively. Although FCN achieves 35.15 FPS, its IoU and Dice values with both BUS image datasets are much lower than the ones of the proposed model.

Although the proposed method achieves accurate segmentation results, it achieves limited tumor detection segmentation accuracy with some BUS images, such as the three examples shown in Fig 11. Indeed, such cases represent a challenge for any segmentation method since tissues surrounding the tumors have poor contrast as well as ambiguous boundaries. Even it is difficult to recognize the tumor in these BUS images with the naked eye. To further improve the performance of the proposed model, we will integrate it with compressed representations in the frequency domain that contain rich patterns generated by a discrete cosine transform (Xu et al., 2020).

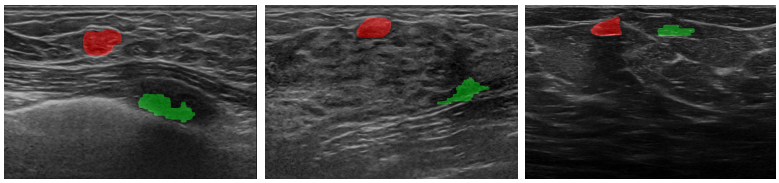

Figure 11: Examples of incorrect tumor segmentation and localization results.

It worth noting that the proposed model has no assumption for a specific image modality, and so it is a promising segmentation model for breast tumor segmentation in mammograms, vessel segmentation in fundus images as well as skin lesion segmentation in dermoscopic images. In general, it could be used with any binary segmentation task with minimal modification (the model should be trained using data of each task). Besides, the proposed segmentation model could be integrated in the segmentation step of existing breast cancer CAD systems, such as breast cancer subtype classification in mammograms (Singh et al., 2018b), breast cancer detection in thermograms (Abdel-Nasser et al., 2016), and tumor classification in ultrasound images (Abdel-Nasser et al., 2017).

4. Conclusion

In this article, we have proposed an efficient method for tumor segmentation in BUS images based on a contextual-information-aware GAN architecture. To promote tumor-relevant features, we have exploited an AC block in the generator network, allowing for the use of spatial and scale contextual information (i.e., position and size of tumors). We also have adopted a channel attention and channel weighting (CAW) mechanism in the generator network to boost the tumor-relevant features without additional supervision and mitigate the effect of artifacts. Lastly, we have employed a combination of the SSIM index and L1-norm in the loss function of cGAN to model the local contextual information derived from the area surrounding tumors. Our model outperforms the FCN, SegNet, ERFNet, UNet, DCGAN, and cGAN segmentation models in terms of Dice and IoU metrics, achieving the top scores of 93.76% and 88.82%, respectively, with Dataset A. With Dataset B, the proposed model obtained promising

results of 86.82% and 80.37% in terms of Dice and IoU score. The proposed segmentation model can help physicians in detecting and diagnosing breast tumors from BUS images. It could be employed with diverse binary segmentation tasks with minimal modification, such as breast tumor segmentation in mammograms and thermograms, as well as skin lesion segmentation in dermoscopic images. In the future, learning in the frequency domain approaches, e.g. (Xu et al., 2020), will be incorporated to further improve the provenance of the proposed method. Besides, we will integrate our proposed model with an automated BUS image CAD system to make benefits for physicians in clinical practice for making an accurate diagnosis.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Abdel-Nasser, M., Melendez, J., Moreno, A., Omer, O.A., Puig, D., 2017. Breast tumor classification in ultrasound images using texture analysis and super-resolution methods. Engineering Applications of Artificial Intelligence 59, 84–92.

Abdel-Nasser, M., Puig, D., Moreno, A., Saleh, A., Martí, J., Martin, L., Magarolas, A., 2015. Breast tissue characterization in x-ray and ultrasound images using fuzzy local directional patterns and support vector machines., in: VISAPP (1), pp. 387–394.

Abdel-Nasser, M., Saleh, A., Moreno, A., Puig, D., 2016. Automatic nipple detection in breast thermograms. Expert Systems with Applications 64, 365–374.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39, 2481–2495.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 .

Doi, K., 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics 31, 198–211.

Fu, J., Liu, J., Tian, H., Fang, Z., Lu, H., 2018. Dual attention network for scene segmentation. arXiv preprint arXiv:1809.02983 .

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Advances in neural information processing systems, pp. 2672–2680.

Gu, P., Lee, W.M., Roubidoux, M.A., Yuan, J., Wang, X., Carson, P.L., 2016. Automated 3d ultrasound image segmentation to aid breast cancer image interpretation. Ultrasonics 65, 51–58.

Horsch, K., Giger, M.L., Venta, L.A., Vyborny, C.J., 2001. Automatic segmentation of breast lesions on ultrasound. Medical Physics 28, 1652–1659.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

Hu, Y., Guo, Y., Wang, Y., Yu, J., Zhou, S., Chang, C., 2019. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Medical physics 46, 215–228.

Huang, Q.H., Lee, S.Y., Liu, L.Z., Lu, M.H., Jin, L.W., Li, A.H., 2012. A robust graph-based segmentation method for breast tumors in ultrasound images. Ultrasonics 52, 266–275.

Huang, Y.L., Chen, D.R., 2006. Automatic contouring for breast tumors in 2-d sonography, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE. pp. 3225–3228.

Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 1857–1865.

Lai, M., 2015. Deep learning for medical image segmentation. arXiv preprint arXiv:1505.02000 .

Lalande, A., Garreau, M., Frouin, F., 2015. Evaluation of cardiac structure segmentation in cine magnetic resonance imaging. Multi-Modality Cardiac Imaging: Processing and Analysis , 169–215.

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., Straif, K., 2015. Breast-cancer screening—viewpoint of the iarc working group. New England journal of medicine 372, 2353–2358.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88.

Lo, C., Shen, Y.W., Huang, C.S., Chang, R.F., 2014. Computer-aided multiview tumor detection for automated whole breast ultrasound. Ultrasonic imaging 36, 3–17.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.

Rodrigues, P.S., 2017. Breast ultrasound image. Mendeley Data doi:http://dx.doi.org/10.17632/wmy84gzngw.1.

Rodtook, A., Makhanov, S.S., 2013. Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer. Journal of Visual Communication and Image Representation 24, 1414–1430.

Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R., 2018. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems 19, 263–272.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Shan, J., Cheng, H., Wang, Y., 2012. A novel segmentation method for breast ultrasound images based on neutrosophic l-means clustering. Medical physics 39, 5669–5682.

Siegel, R.L., Miller, K.D., Jemal, A., 2017. Cancer statistics, 2017. CA: a cancer journal for clinicians 67, 7–30.

Singh, V.K., Abdel-Nasser, M., Rashwan, H.A., Akram, F., Pandey, N., Lalande, A., Presles, B., Romani, S., Puig, D., 2019. Fca-net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. IEEE Access 7, 130552–130565. URL: https://doi.org/10.1109/ACCESS.2019.2940418, doi:10.1109/ACCESS.2019.2940418.

Singh, V.K., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Abdulwahab, S., Maaroof, N., Torrents-Barrena, J., Romani, S., et al., 2018a. Retinal optic disc segmentation using conditional generative adversarial network., in: CCIA, pp. 373–380.

Singh, V.K., Rashwan, H.A., Romani, S., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D., et al., 2020. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Expert Systems with Applications 139, 112855.

Singh, V.K., Romani, S., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M.M.K., Abdulwahab, S., Torrents-Barrena, J., Saleh, A., Arquez, M., Arenas, M., Puig, D., 2018b. Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification, in: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, pp. 833–840. URL: https://doi.org/10.1007/978-3-030-00934-2_92, doi:10.1007/978-3-030-00934-2_92.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 600–612.

Xian, M., Zhang, Y., Cheng, H.D., 2015. Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. Pattern Recognition 48, 485–497.

Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.K., Ren, F., 2020. Learning in the frequency domain, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1740–1749.

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L., 2019. Medical breast ultrasound image segmentation by machine learning. Ultrasonics 91, 1–9.

Yap, M.H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A.K., Martí, R., 2017. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics 22, 1218–1226.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 .

Zhuang, Z., Li, N., Joseph Raj, A.N., Mahesh, V.G., Qiu, S., 2019. An rdau-net model for lesion segmentation in breast ultrasound images. PloS one 14, e0221535.