

# Artificial Intelligence-Based Breast Nodule Segmentation Using Multi-Scale Images and Convolutional Network

**Quoc Tuan Hoang<sup>1</sup>, Xuan Hien Pham<sup>2</sup>, Anh Vu Le<sup>3</sup> and Trung Thanh Bui<sup>1\*</sup>**

<sup>1</sup>Department of Mechanical Engineering, Hung Yen University of Technology and Education, Hung Yen, Vietnam

[e-mails: hoangquoctuan@utehy.edu.vn; buitrunghanh@utehy.edu.vn]

<sup>2</sup>Department of Mechanical Engineering, University of Transport and Communications, Hanoi, Vietnam

[e-mail: hienpx@utc.edu.vn]

<sup>3</sup>Communication and Signal Processing Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[e-mail: leanhvu@tdtu.edu.vn]

\* Corresponding author: Trung Thanh Bui

*Received October 14, 2021; revised August 29, 2022; accepted February 26, 2023;  
published March 31, 2023*

---

## Abstract

Diagnosing breast diseases using ultrasound (US) images remains challenging because it is time-consuming and requires expert radiologist knowledge. As a result, the diagnostic performance is significantly biased. To assist radiologists in this process, computer-aided diagnosis (CAD) systems have been developed and used in practice. This type of system is used not only to assist radiologists in examining breast ultrasound images (BUS) but also to ensure the effectiveness of the diagnostic process. In this study, we propose a new approach for breast lesion localization and segmentation using a multi-scale pyramid of the ultrasound image of a breast organ and a convolutional semantic segmentation network. Unlike previous studies that used only a deep detection/segmentation neural network on a single breast ultrasound image, we propose to use multiple images generated from an input image at different scales for the localization and segmentation process. By combining the localization/segmentation results obtained from the input image at different scales, the system performance was enhanced compared with that of the previous studies. The experimental results with two public datasets confirmed the effectiveness of the proposed approach by producing superior localization/segmentation results compared with those obtained in previous studies.

---

**Keywords:** Breast lesion localization and segmentation, computer-aided diagnosis system, convolutional network, ultrasound imaging.

## 1. Introduction

### 1.1. Related Works

Previous studies demonstrated that breast cancer is one of the leading causes of death for women globally [1] [2] [3]. According to a report published by the World Health Organization (WHO), there were approximately 2.3 million women diagnosed with breast cancer and 685,000 deaths from breast cancer in 2020 [4]. However, breast cancer can be effectively treated, particularly if it is identified at an early stage. Therefore, early detection and treatment are critical for reducing breast cancer deaths. Traditionally, doctors can diagnose a disease based on their personal knowledge and experience gained through extensive training processes [5] [6]. Because of the training process, it is time-consuming to train a doctor, and the diagnostic process performance is highly dependent on the doctor. To reduce this limitation of the diagnostic technique and improve its performance, a double screening technique is used as an alternative, in which a disease is diagnosed by two or more doctors, and the final results are obtained by combining the individual results [5].

With developments in digital signal processing, computer-aided diagnosis (CAD) systems have been widely used to assist doctors in the diagnosis process [5]. This is a new technology that uses a computer program to detect/diagnose diseases based on captured images of human organs and prior associated expert knowledge. Accordingly, this method is used to reduce errors caused by traditional diagnosis methods, which are highly dependent on the knowledge and experience of doctors.

Many imaging methods have been studied and used in medical diagnosis systems, such as X-ray [7] [8], computed tomography scan (CT-scan) [7] [9] [10], MRI [11] [12], and ultrasound [6] [13] [14] [15]. Among these imaging techniques, ultrasound imaging uses sound waves to produce pictures of the internal structures of human organs, such as the breast [1] [2] or thyroid [15]. An important advantage of ultrasound imaging is that it does not use radiation during image acquisition process. Therefore, it is safe for patients, especially pregnant women and fetuses. In addition, it is non-invasive and easier to use than other imaging techniques. Because of these characteristics, ultrasound imaging has been widely used to capture images of thyroid or breast organs.

There are two main steps in CAD for diagnosing breast diseases: segmentation and classification [1]. In these two steps, segmentation is crucial for ensuring the performance of a diagnostic system, as its main purpose is to correctly localize the lesion region in a breast ultrasound image. Owing to its functionality, lesion segmentation can help doctors focus on only the possible lesion region, instead of examining the entire breast region during the diagnosis process, which is the case when using the conventional diagnostic method (the disease is diagnosed and treated directly by doctors without the help of the CAD system). An automatic diagnostic system helps increase the performance of the classification step by guiding feature extraction by focusing only on the possible lesion regions. Therefore, the segmentation step is key in a breast disease diagnostic system.

There have been many previous studies on breast lesion segmentation problems. Accordingly, the localization and segmentation methods can be roughly classified into two categories: non-learning-based and learning-based methods. In the first category, the breast lesion region is mainly detected and localized using a graph-based method or active contour model. Huang et al. [3] used a graph-based method to localize the breast lesion (tumor) in an ultrasound image. Aho et al. [16] used an active contour model to segment the breast lesion region. In that study, the authors propose a scheme to optimally determine the threshold value

for the segmentation process. Similar to the study conducted by Acho et al., Rampun et al. [17] used the active contour model to search for breast lesion boundaries and a Canny edge detector with contour growth to segment the pectoral region. As a result of these studies, graph-based or active contour models can be used to determine the boundary of the breast lesion region. However, the performance of these methods is highly dependent on the input image quality, which is normally associated with ultrasound images.

In the second category, learning-based methods are applied to find the boundary of the breast lesion using not only the boundary itself but also the correlation between the gray level of pixels inside and outside the lesion. Deep convolutional neural networks (DNNs), which have received considerable attention in academia owing to their success in CAD systems, are an example of such methods in this category. Previous studies have used convolutional neural networks (CNNs) for either breast lesion segmentation or classification tasks. According to the disease progression, breast lesion normally appears on captured ultrasound images as a lump of different sizes, which is different from normal breast regions. Because of this characteristic, a breast lesion can be localized by either a detection or segmentation task using a deep convolutional neural network. Yap et al. [18] used a popular object detection method, namely, Faster-RCNN, with the Inception-ResNet model to detect breast lesions region from input gray or RGB images. The results of this study indicate that the Faster-RCNN network can efficiently detect breast lesions owing to its high detection performance. Yap et al. [19] used two methods to segment lesions from breast ultrasound images: U-Net [20] and a full convolution network (FCN), namely, FCN-AlexNet. To compensate for the noise, low quality, and difference in tumor sizes in breast ultrasound images, Singh et al. [21] proposed using atrous convolutions to capture spatial and scale context and channel attention combined with a weighting mechanism to promote tumor-relevant features for breast lesion segmentation. In a recent study conducted by Zhou et al. [22], a new segmentation network architecture, namely, the UNet++ network, was proposed, which has proven to work well in medical image segmentation problems. The UNet++ network can be considered a nested network of shallow and deep U-Net networks. As a result, it can produce better segmentation performance than a conventional U-Net network. Baccouche et al. [23] recently proposed a method that combines two U-Net networks for breast mass segmentation. By using two U-Net-based networks, they not only increased the network depth but also utilized the power of nested networks for segmentation purposes. These studies confirmed that deep learning-based techniques are adaptable and sufficient for breast lesion segmentation tasks. However, it is challenge to segment small objects (lesions) using a very deep network because of the gradient vanishing problems.

To enhance the segmentation performance of a single network (U-Net-based network), Hiramatsu et al. [24] proposed using a mixture of experts (MoE) approach by integrating multiple FCN networks. Accordingly, they used several U-Net-based networks as expert networks and another U-Net-based network as a gating network to construct the segmentation networks. Expert networks were used to learn the segmentation results of an input image, and the gating network was used to learn the role (weight) of every expert network. This approach has proven to be more efficient than the conventional U-Net network in the segmentation problem. However, this approach requires a long processing time for training and inferring segmentation results obtained from multiple networks. In addition, using multiple FCN networks causes a larger segmentation model file than using a single CNN network. This is an important limitation, especially when working with a large training dataset or when segmentation systems require a fast inference time.

To overcome these problems encountered in previous studies, we propose a novel approach for a breast nodule segmentation network based on using multi-scale images and a single FCN network.

The remainder of this paper is organized as follows. In section 2, we provide a detailed explanation of our approach to the breast lesion segmentation problem. In section 3, we apply the proposed approach on two public datasets to evaluate its segmentation performance and compare it with that of previous studies, which were also evaluated using the same datasets. Finally, section 4 presents the conclusions of the study.

## 1.2. Contributions

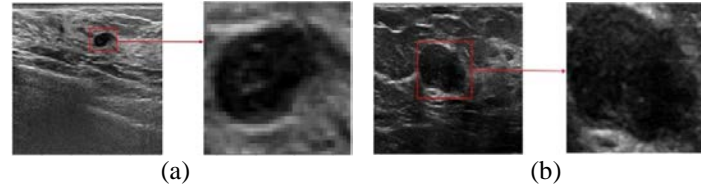
In this paper, we propose an enhanced approach for breast lesion segmentation. The main difference between our study and previous studies is that we enhance segmentation accuracy by using a single segmentation network on multi-scale pyramid images. Specifically, our approach aims that the segmentation network is able to segment breast lesions of different sizes by training the network with multi-scale pyramid images. In addition, we overcome the limitations of conventional deep segmentation networks (that use single image) and ensemble networks (that use multiple network architectures) for segmentation purposes. Our study has four novelties compared to previous studies:

- First, we propose a data augmentation method based on padding and scaling methods to compensate for the lack of volume in the image dataset as well as the large variation in the size of breast lesions.
- Second, we designed and trained a deep learning-based segmentation network based on the U-Net [20] network with additional residual connections between the network layers to deeply manipulate the image data and easily train the network.
- Third, we segmented the breast lesion from a breast image at various scales to enhance segmentation accuracy. By using multi-scale images for a single convolutional network in the segmentation task, we can not only enhance segmentation performance but also overcome the limitation of the ensemble method, in which a single image is used as the input of multiple segmentation networks.
- Fourth, we propose three combination rules concerning the resulting images at various image scales to enhance the segmentation accuracy compared to that achieved using a single-scale image. We confirmed the efficiency of these combination rules experimentally using public datasets.

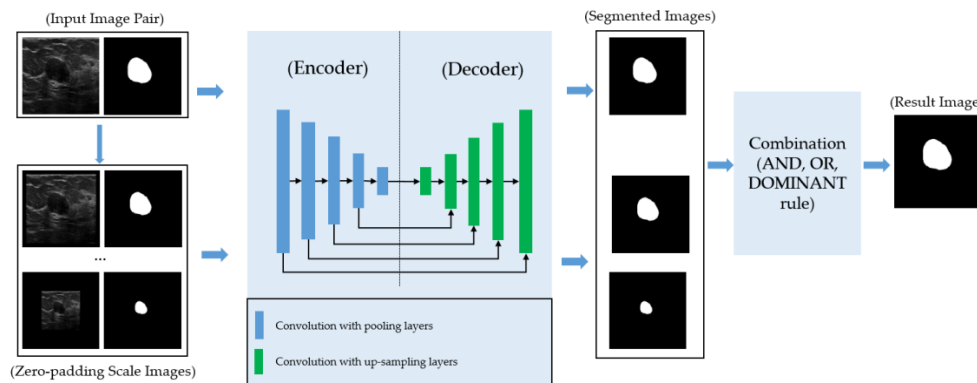
## 2. Proposed Method

### 2.1. Overview of the Proposed Method

The breast nodule region sizes are mostly different among images according to the stage of the disease. In Fig. 1, examples of breast ultrasound images with small and large lesions are depicted. As shown in this example, a small lesion region is easily recognized and localized by human perception because it appears as a spot in the dark region. However, it is challenging to localize the correct lesion region when it is grown. As shown on the right side in Figs. 1 (a) and (b), the boundary of the small lesion was clearer than that of the larger lesion. This is because larger lesions contain more noise, lower contrast, and shadow effects [2].



**Fig. 1.** Example of ultrasound breast nodule images: images (a) with a small lesion and (b) with a large lesion. (left: original image, right: lesion region)



**Fig. 2.** Overall flow chart of the proposed method.

Because of this phenomenon, conventional segmentation networks, such as the Faster-RCNN-based detection network [18] or U-Net [20], find it difficult to correctly segment the nodule region from the input images. To address this issue, we propose a multi-scale approach to breast nodule segmentation, which can not only address the problem of a large variation in nodule regions but also in nodule pixel distribution. The proposed approach is briefly depicted in Fig. 2. First, we created additional input images from a single input image using a zero-padding scaling method. Using these input images, we trained and tested using a deep learning-based segmentation network to segment the breast nodule in every image. Finally, we combine the segmentation results using various image scales to form the final segmentation image. As shown in Fig. 2, our proposed approach is still mainly based on a deep learning-based segmentation framework. However, we focused on two enhancement points to address the main problems occurring in breast nodule segmentation system. First, to solve the problem of the large variation in the nodule region that was originally large owing to the difference in disease stage, we propose a zero-padding scaling method to augment the collected image data. As a result, we obtained a larger training dataset that contained various nodule sizes (a generalized dataset). Second, we segmented the nodule region of a single input image at various scales and combined their results to enhance the segmentation accuracy. A detailed explanation of this step is provided in section 2.3. Finally, we designed a segmentation network based on U-Net [20] with the use of residual building block to enhance the segmentation accuracy compared to that achieved in previous studies.

## 2.2. Deep Learning-Based Segmentation Network

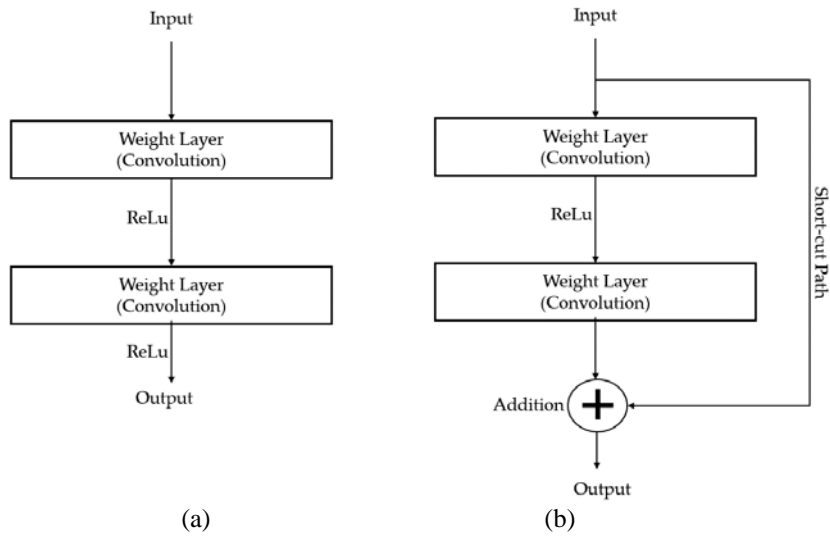
Recently, deep learning-based methods, such as convolutional neural networks (CNN), generative adversarial networks (GAN), and segmentation networks, have been widely developed and successfully applied to various computer vision tasks. For example, the CNN method has been successfully used for object classification problems [25] [26] [27] [28] [29], image generation [30] [31] [32] [33], and object detection [34] [35].

In our study, we constructed a breast lesion segmentation network based on an auto encoder-decoder network with the use of residual building blocks. Unlike conventional deep learning-based classification networks, image segmentation networks do not contain fully connected layers. Instead, they predict class labels for every pixel in an input image using a convolution network. For this purpose, an auto encoder-decoder structure is typically used [20]. The encoder part has a similar function to the convolution part of a CNN network, which is used to learn efficient abstract features from an input image, whereas the decoder performs an inverse function that uses the extracted features from the encoder and learns how to create the target label image.

In a deep learning-based framework, networks should become deeper to achieve high performance by using numerous weighted layers. Because of this formula, deep feature maps contain more abstract features than shallow ones. As a result, the deep layers tend to lose spatial information compared with the shallow layers. This is a critical problem because it can cause a vanishing gradient problem, which causes difficulties when training the deep learning-based networks. To address this problem, He et al. [27] proposed the so-called residual connection. With the residual connection, the information from the early layers can be transferred to the deeper layers not only through the next layers but also directly through the input of the deep layers by a deep-feature concatenation operation [27]. Consequently, it helps reduce the vanishing gradient problem and makes the network easier to train. Inspired by the residual connections, we constructed our segmentation network based on deep U-Net [20] with the building blocks of residual blocks, as shown in Figs. 3 and 4.

In Fig. 3, we depict the concept of the residual learning approach compared with a conventional convolution learning. Fig. 3a presents the conventional convolution operation used to manipulate information on conventional CNNs, such as AlexNet [25] or VGG-Net [26], while Fig. 3b presents the residual block concept. As shown in Fig. 3a, conventional CNNs process the input information through a linear stack of convolution layers. Because of this working style, the deeper layers contain more abstract information than the shallower ones. As a result, the deeper layers lose the detailed information in the shallow layers. Owing to this problem, the conventional CNN network is difficult to train because of the vanishing gradient problem. In Fig. 3b, the output tensor comprises two components: a manipulated tensor after several weight layers that contain more abstract information and short-cut tensor that contains low abstract information. By combining the two tensors, the output of the residual block contains richer information than the conventional convolution blocks, which can help enhance the network performance. In addition, the residual connection helps reduce the vanishing gradient problem and alleviates the network training difficulties.



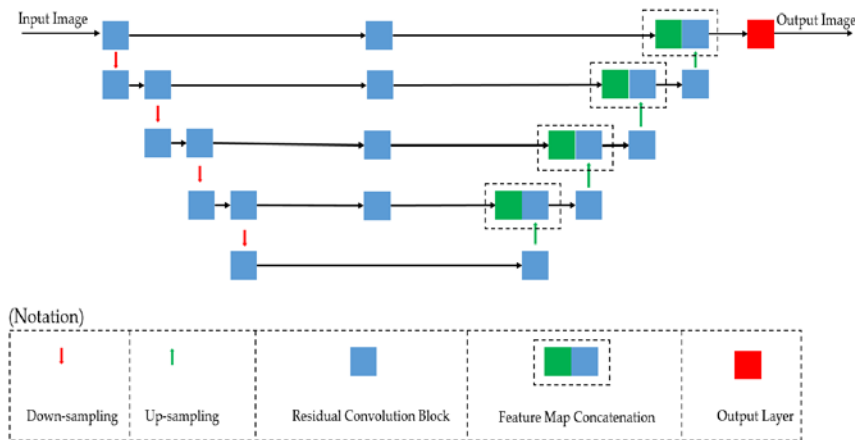


**Fig. 3.** Concepts of the (a) conventional convolution blocks and (b) residual block used in deep learning networks

Using the residual connections shown in **Fig. 3**, we constructed our segmentation network based on U-Net [20], as shown in **Fig. 4**. We constructed our segmentation network for breast lesion segmentation in a manner similar to the conventional U-Net network [20]. However, we used a residual block to manipulate the image information provided to the network instead of using conventional convolution layers. In addition, we further processed the image tensor in the short-cut path using the residual block. To train the segmentation network shown in **Fig. 4**, we used the Dice loss function instead of the conventional cross-entropy loss function. As proven by previous studies [36], the Dice loss function is more suitable compared with the conventional cross-entropy loss function for a deep learning-based segmentation network, especially for unbalanced segmentation tasks. This is a statistical measurement of the correction done by the segmentation technique. The Dice score measurement (DSC) assesses the similarity between two objects in terms of the overlapped region concerning the ground-truth object, as formulated in equation (1). In this equation, the “ $\cap$ ” indicates the intersection operator between two sets, such as X and Y, and  $|X|$  and  $|Y|$  indicate the size of these two sets. As explained by this equation, when the two sets, that is, X and Y, completely overlap, DSC is 1. However, when these two sets are completely separated, the DSC is 0. In the other cases, the DSC ranges from 0 to 1. Therefore, the DSC measures the quality of the segmentation method. Based on the meaning of the Dice score measurement, the Dice loss function is formed by taking the complement of the DSC, stated in equation (2). As a result, the loss ( $L_{DSC}$ ) is large (approaching 1) if there is no correct prediction (segmentation) by the segmentation method, and it is small (approaching 0) if there exists a large correct prediction region by the segmentation method. By using the Dice loss function, we aim to maximize the overlapped region between the predicted and ground-truth sets rather than the detailed pixel prediction.

$$DSC(X, Y) = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (1)$$

$$L_{DSC} = 1.0 - DSC(X, Y) \quad (2)$$



**Fig. 4.** Segmentation network architecture based on U-Net and convolutional residual connection in our proposed approach

### 2.3. Data Augmentation

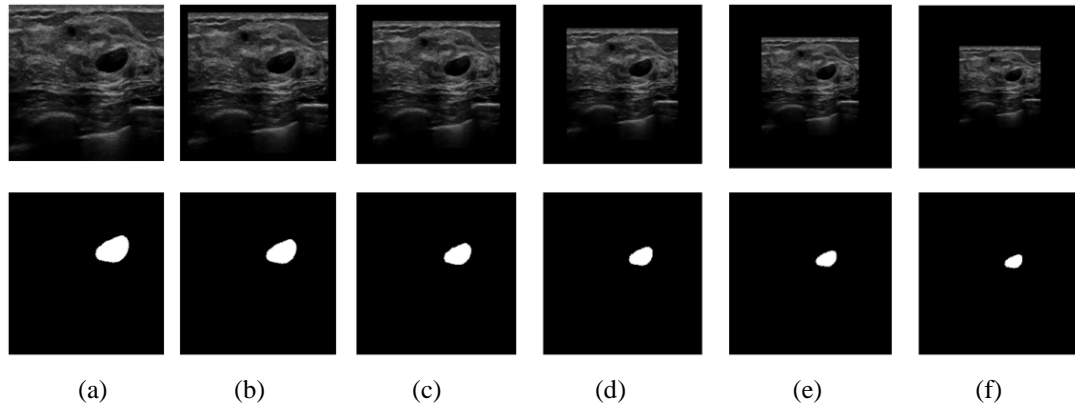
One of the most significant issues in medical image processing systems is the lack of image data for training and evaluating system performance. This is a common problem because medical image data have the characteristics of personal information. Consequently, it is normal to ask for patients' consent to use their private data. In addition, it requires strong efforts to collect data, expensive data collection devices, and help of radiologists (experts) to correctly label the collected images. Data augmentation methods are typically used to solve this problem [8] [18] [25] [26] [27]. Data augmentation is a popular technique that has been widely used in many previous deep learning studies to enlarge training datasets [8] [18] [25] [26] [27]. This is a group of methods for generating additional data from original data by using methods, such as cropping and scaling, mirroring, rotation, and artificial generation by deep learning networks. The size of breast nodules varies according to the disease stage. Knowing this phenomenon, we employed various data augmentation methods, including conventional augmentation methods such as image mirroring (flipping up-down, left-right) and the zero-padding scaling method.

As conventional data augmentation methods, we employ two common methods, namely, left-right and up-down direction flipping. This method has been widely used in previous studies on computer vision problems [25] [26]. Because breast nodules are non-directed lesions, the direction of the nodule is not important for the segmentation problem. As a result, this method can help to not only enlarge but also generalize the training dataset.

In this study, we adopted a new data augmentation method, namely, the zero-pad scaling method. As explained in the above sections, when the breast nodule is too large, the detailed texture inside the nodule becomes unusual with associated noise, which causes a very large variation among the lesion's pixels even if they belong to the breast lesion (same label). This problem can cause errors in the segmentation network or make it difficult to be successfully trained. To reduce these effects, we apply a scaling method to the input image. Consequently, we created a new sequence of scale images in which the breast lesion size ranged from large to small, as shown in Fig. 5. In addition, we pad the boundaries of the scaled image with zero-based pixels to create scaled images of a uniform size. By using the zero-padding scaling method, we can not only enlarge the training dataset by creating more small-size breast lesions (creating small-size breast lesion images from the large-size breast lesion images) but also detect breast lesions at various image scales to enhance segmentation accuracy. In our



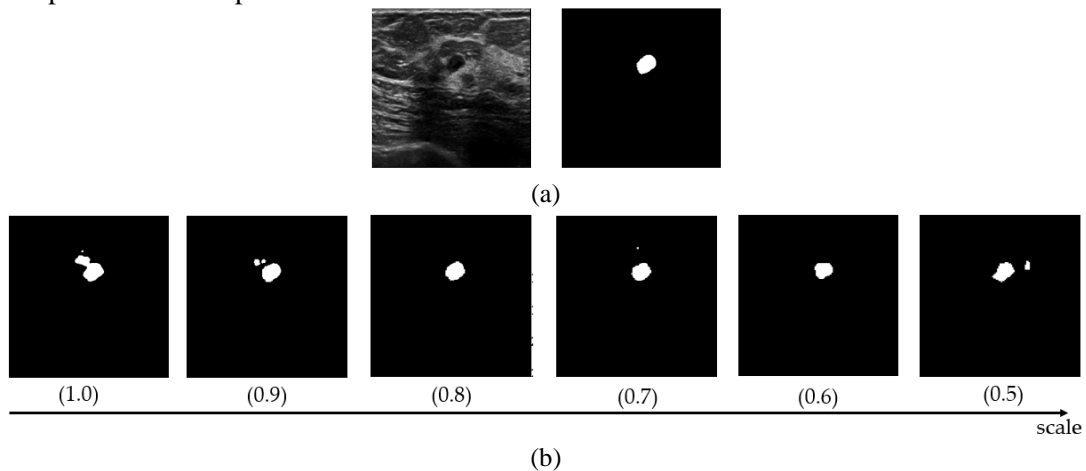
experiments, we used five different scale factors to create a sequence of images from an original input breast image (i.e., 0.9, 0.8, 0.7, 0.6, and 0.5). Consequently, we obtained six images for each original image, including one original image and five additional scale images, as shown in Fig. 5. Observe from Fig. 5, the breast lesion size decreases according to the scale factor.



**Fig. 5.** Example of zero-pad scaling image sequence generated from a single input breast lesion image. (a) input breast lesion image and (b)–(f) scale images with the scale factor of 0.9, 0.8, 0.7, 0.6, and 0.5, respectively.

#### 2.4. Fusion of Network Outputs

Unlike existing conventional segmentation networks [20] [22] [23], our proposed method performs segmentation on multiple zero-padding scale image sequences to produce a sequence of segmented images at the output of the segmentation network. At the output, we combine the segmented images to obtain the final segmentation image. Fig. 6 shows an example of the output images obtained using our proposed method. In this figure, the original image was segmented with noise near and inside the region of a ground-truth nodule, whereas the scale images were segmented with a finer structure of the nodule. These results show that, although the original image can be used to segment breast nodules, the large variation of pixels inside the nodule can create an imperfect segmented image, whereas the scale image can be used to compensate for this problem.



**Fig. 6.** Example of predicted breast lesions at various image scales using our proposed network: (a) input image (left) and ground-truth lesion region (right) and (b) outputs of our proposed network at various image scales.

To combine the segmentation result sequences, we propose to use three combination rules, including the “AND,” “OR,” and “DOMINANT” rules, as defined in equations (3) – (5). The “AND” rule is performed by taking the overlapped results of these output images using the and-logical operator. Consequently, the final output image contains the smallest overlapped regions of all images in the output sequence of the segmentation network. The OR rule is performed by considering the largest cover region of the images in the output sequence based on the or-logical operator. Finally, the DOMINANT rule is applied based on the most dominant result in the output sequence. This implies that a pixel in the final output image is considered a lesion pixel if most of the output image in the sequence depicts it as a lesion pixel. These combination rules are performed by excluding the output images that do not contain any lesion pixels, as the purpose of this study is to segment (detect) lesion regions from the input lesion image.

$$\text{AND\_RULE} = \text{AND}(O_i) \quad (3)$$

$$\text{OR\_RULE} = \text{OR}(O_i) \quad (4)$$

$$\text{DOMINANT} = \frac{1}{N} \sum_{i=1}^N O_i > 0.5 \quad (5)$$

### 3. Experimental Results

#### 3.1. Datasets

To validate the performance of our proposed method, we used two public datasets, including the BUS [19] and BUSI datasets [37], which have been used in previous studies on the segmentation problem. The statistical characteristics of the two datasets are presented in Table 1. The BUS dataset [19] contained 163 images collected in 2012 from the UDIAT Diagnosis Center of the Parc Tauli Corporation, Spain, using a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz). All images in the BUS dataset [19] contained lesions with an image resolution of 760-by-570 pixels. The BUSI dataset is newly released by Al-Dhabyani et al. [37] and contains 780 ultrasound breast images of women between 25 and 75 years of age. Among the 780 images, 647 presented nodules. Therefore, we used only 647 images that contained breast nodules in our experiments.

To measure the performance of our proposed method, we performed a five-fold cross-validation approach. For this purpose, we randomly divided the entire dataset (BUS or BUSI) into five different equal parts. Among these five parts, four are used for training, and the remaining part is used for testing. This procedure was repeated five times by exchanging the training and testing datasets. Based on the experimental results obtained with these five parts, the final performance of the proposed system was calculated by taking the average performance of all five parts.

**Table 1.** Brief description of the BUS and BUSI datasets used in our experiments

Dataset	With Lesion Image	Without Lesion Image	Image Resolution	Total
BUS [19]	163	0	760-by-570	163
BUSI [37]	647	133	500-by-500	780

### 3.2 Performance Measurement Metrics

For a detection/segmentation system, multiple metrics were used in performance evaluations. Inspired by previous studies conducted on BUS and BUSI datasets, we used six measurements, including precision, recall (sensitivity), F1-score, Dice score, specificity, and pixel classification accuracy. Based on the system use case, a segmentation system can be used as a detection system to localize the rough localization of the lesion or as a segmentation system to correctly localize pixels of the lesion region.

First, a segmentation method can be considered a detection problem. In this setup, we used three popular measurement metrics to measure the performance of the proposed method: precision, recall, and F1-score. Using these measurements helps evaluate the performance of our proposed method in terms of the detection problem, as suggested by Yap et al. [18]. In detail, the precision, recall, and F1-score were measured using equations (6) – (8) as follows:

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

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

$$F_1 = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (8)$$

There are two possible evaluation methods where detection results can be evaluated, namely, the intersection over union (IOU) and detected point (DP) methods [18]. Therefore, the definitions of TP, FP, and FN in equations (6) and (7) are different. When using IOU, the definition of TP, FP, or FN is based on the measurement of IOU between the detected lesion box and ground truth box. If the IOU measurement is equal to or greater than 0.5, the detection result is regarded as a TP case; if the IOU measurement is smaller than 0.5, the detection results are regarded as FP. The FN case occurs when we fail to locate/segment any object from the input image that is with the ground-truth object. When using the detected point, a detection result is considered a TP if the detected center of the bounding box is placed within the ground-truth bounding box of a lesion (the ground-truth region is provided by an expert radiologist). Otherwise, it is considered as an FP. Similar to the case of using IOU, a detection result is considered as an FN if we fail to segment/locate any object in an input image.

For the segmentation problem, previous studies [18] [38] [39] [40] used the measurements of Dice score (DSC), sensitivity, specificity, and overall accuracy for performance evaluation. These metrics are defined in equations (9) – (12) as follows:

$$DSC(X, Y) = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (9)$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (10)$$

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

In these equations, the TP, TN indicate the number of pixels those are correctly classified as foreground (object pixels) or background, respectively. The FN indicates the number of foreground pixels those are falsely classified as background pixels; and the FP indicates the number of background pixels those are falsely classified as foreground pixels. As indicated by equation (9), the Dice score measures the overlapped region resulting from the union of two regions. The higher the value of DSC, the higher the performance of the segmentation system. In addition to the Dice score, the sensitivity and specificity, as indicated in equations (10) and (11), are used to evaluate the performance of the segmentation system regarding the detectability of background (negative pixels) or foreground (positive pixels). As indicated in equation (10), the sensitivity measures whether a segmentation system correctly detects the foreground region, that is, correctly classifying positive pixels as positive pixels. This indicates how good the test is for disease detection. Specificity measures whether a segmentation system correctly detects background regions, that is, correctly classifying negative pixels as negative pixels. As a result, the high values of these two measurements indicate the high performance of the segmentation system. Finally, the overall detection performance is measured by the accuracy, as shown in equation (12). In our experiments, we used the Python programming language to implement the source code with the help of the TensorFlow [41] library when implementing the deep convolutional networks.

### 3.3 Results

#### 3.3.1. Performance of the Proposed Method as a Detection System

In our first experiment, we measured the performance of the proposed method as a detection system, as suggested by Yap et al. [18]. As explained in section 3.2, we used three measurement metrics, namely, precision, recall, and F1-score, to measure the performance of the proposed method in this experiment.

##### 3.3.1.1 Experimental Results Using the BUS Dataset

In the first experiment, we measured the performance of the proposed approach using the BUS dataset. For comparison, we additionally measured the detection performance using the conventional cross-entropy (CE) loss function to demonstrate the efficiency of the Dice loss method. The detailed experimental results with a five-fold cross-validation scheme are listed in Tables 2 and 3 for the cross-entropy and Dice loss functions, respectively. Table 2 reports the experimental results using the cross-entropy loss function with two network configurations: the residual U-Net and proposed approach. As shown in this table, we obtained an average precision of 83.64%, recall of 97.92%, and F1-score of 89.98% when using residual U-Net with the IOU-based evaluation method, and a precision of 92.5%, recall of 98.12%, and F1-score of 95.12% when using residual U-Net with the DP-based evaluation method. These experimental results imply that the residual U-Net network is sufficient for detecting breast lesions.

Using the proposed approach, precision was enhanced from 83.64% to 85.76%, recall from 97.92% to 99.34%, and F1-score from 89.98% to 91.92% using the “DOMINANT” rule with the IOU-based evaluation method. Meanwhile, precision is from 92.5% to 93.82%, recall from 98.12% to 99.36%, and F1-score from 95.12% to 96.50% using the “DOMINANT” rule with the DP-based evaluation method. Table 2 shows that the performance of our proposed approach with “AND” and “OR” rules is worse than the residual U-Net and the “DOMINANT” rule. The reason for this result is because of the methodology of the “AND” and “OR” rules.

As explained in section 2.4, the “AND” rule is performed by taking the and-logical combination of all outputs at various scales. As a result, the output of our approach using the “AND” rule only contains the highest probability of lesion pixels. In contrast to the “AND” rule, the “OR” rule is performed by taking the or-logical combination of all outputs at various scales. Therefore, the output of our approach using the “OR” rule contains all possible lesion pixels predicted by input images at various scales. Because of this reason, the prediction result obtained using the “OR” rule contains more noise than the prediction result obtained using the residual U-Net network. From these experimental results, we observe that our proposed approach with the “DOMINANT” combination rule yields better detection/segmentation accuracy than the conventional residual U-Net network.

**Table 2.** Performance evaluation as a detection framework of the segmentation network with and without our proposed approach with the cross-entropy loss and three combination rules (unit: %)

Using CE Loss	Using IOU-based evaluation method			Using DP-based evaluation method		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Residual U-Net	83.64	97.92	89.98	92.5	98.12	95.12
AND rule	80.82	99.26	89.02	91.36	99.34	95.14
OR rule	82.50	94.02	87.82	96.26	94.92	95.48
DOMINANT rule	85.76	99.34	91.92	93.82	99.36	96.50

**Table 3** shows the experimental results obtained using the Dice loss function. Using the IOU-based evaluation method, a precision of 87.46%, recall of 98.66%, and F1-score of 92.6% were obtained using residual U-Net. These results are higher than those (precision of 83.64%, recall of 97.92%, and F1-score of 89.98% in **Table 2**) when using the cross-entropy loss function. Using the DP-based evaluation method, we obtained a precision of 94.36%, recall of 98.76%, and F1-score of 96.44%. Again, these experimental results are higher than those of the system with cross-entropy loss reported in **Table 2** (precision of 92.5%, recall of 98.12%, and F1-score of 95.42% in **Table 2**). These experimental results indicate that the Dice loss function is more effective than the conventional cross-entropy loss function in the training segmentation system. In addition, we obtained the best detection accuracy using the “DOMINANT” combination rule for both IOU- and DP-based evaluation method. Specifically, we obtained the best precision measurement of 89.96%, recall of 99.34%, and F1-score of 94.28% in this experiment using the IOU-based evaluation method. Using the DP-based evaluation method, we obtained the highest precision measurement of 94.36%, recall of 99.36%, and F1-score of 96.74%. Based on these experimental results, we conclude that our approach using the “DOMINANT” combination rule and Dice loss function outperforms conventional residual U-Net and the “AND” and “OR” combination rules when evaluated in the detection framework.

**Table 3.** Performance evaluation as a detection framework using the Dice loss (unit: %)

Using Dice Loss	Using IOU-based Evaluation Method			Using DP-based Evaluation Method		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Residual U-Net	87.46	98.66	92.60	94.36	98.76	96.44
AND rule	85.66	99.32	91.84	92.54	99.36	95.74
OR rule	82.90	97.70	89.44	95.54	98.04	96.74
DOMINANT rule	89.96	99.34	94.28	94.36	99.36	96.74

### 3.3.1.2 Experimental Results Using the BUSI Dataset

We assess the performance of the segmentation networks in the form of a detection framework using the BUSI dataset; the corresponding experimental results are presented in [Table 4](#). As shown in section 3.3.1.1, the “DOMINANT” rule outperforms the “AND” and “OR” rules. Therefore, we report only the performance using the “DOMINANT” rule in [Table 4](#). First, for the IOU-based evaluation method, we obtained a precision value of 81.380% using our proposed method that is higher than 79.520% that is produced by the conventional segmentation method using cross-entropy loss. When using the Dice loss, the precision results slightly increased from 81.380% to 81.800%. Similarly, we obtained higher values for both recall and F1-score using our proposed approach compared with the conventional segmentation network. We obtained the highest recall of 99.620% and F1-score of 89.600% in our experiment.

Second, for the DP-based method, we obtained the highest precision value of 93.300% with our proposed approach combined with the Dice loss function, the highest value of recall of 99.660% with our proposed approach combined with the cross-entropy loss function, and the highest F1-score of 96.140% with our proposed approach combined with the Dice loss function. All the measured performance measurements using our proposed method were higher than those of the conventional method. These results indicate that the proposed approach outperforms the conventional segmentation method using the BUSI dataset.

**Table 4.** Performance evaluation as a detection framework of our proposed approach with the BUSI dataset (unit: %)

Loss Method		Using IOU-based evaluation method			Using DP-based evaluation method		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
CE Loss	Without Multi-scale	79.520	99.420	88.340	89.900	99.480	94.460
	With Multi-scale	81.380	99.620	89.600	92.240	99.660	95.800
Dice Loss	Without Multi-scale	79.960	97.300	87.740	90.540	97.600	93.920
	With Multi-scale	81.800	99.040	89.600	93.300	99.180	96.140

### 3.3.2. Performance of the Proposed Method as a Segmentation System

In the second experiment, we evaluated the performance of the proposed approach in a segmentation framework. For this purpose, we used our proposed approach to detect pixel-wise lesion regions and measure the detection/segmentation performance based on four measurements, namely, Dice, sensitivity, specificity, and overall accuracy as mentioned in Section 3.2

#### 3.3.2.1. Experimental Results Using the BUS Dataset

[Tables 5](#) and [6](#) report the experimental results for using the BUS dataset [19] with two loss functions of cross-entropy and the Dice function, respectively. When using the cross-entropy loss function, we obtained a Dice score of 77.511% with a sensitivity of 80.049%, specificity of 99.311%, and overall pixel classification accuracy of 98.22% using residual U-Net network. By using our proposed approach with the “DOMINANT” rule, we enhance the Dice score to 79.589% with a sensitivity of 82.892%, specificity of 99.318%, and overall pixel classification accuracy of 98.360%. As presented in [Table 5](#), the “DOMINANT” combination rule again produces the highest segmentation accuracy compared with the conventional residual U-Net



and other two combination rules (“AND” and “OR” rule). By comparing the performance measurements in **Table 5**, we observe that there is minimal enhancement in the specificity (99.311% vs. 99.318%) and overall accuracy (98.220% vs. 98.360%). However, the Dice score is enhanced by more than 2% (from 77.511% to 79.589%), and sensitivity is enhanced by almost 2.9% (from 80.049% to 82.892%). This result indicates that our proposed approach is sufficient to enhance the detection ability of the foreground (breast lesion region), which is the main purpose of medical image processing systems, while keeping the detection ability of background.

**Table 5.** Performance measurement as a segmentation framework using cross-entropy loss and three combination rules (unit: %)

Using CE Loss	Dice	Sensitivity	Specificity	Accuracy
Residual U-Net	77.511	80.049	99.311	98.220
AND rule	74.913	87.780	98.413	97.940
OR rule	74.655	73.526	99.652	98.240
DOMINANT rule	79.589	82.892	99.318	98.360

The performance measurements for the case of using the Dice loss function are presented in **Table 6**. From this table, the performance obtained using the Dice loss for the conventional residual U-Net is better than that obtained using the cross-entropy loss function (mentioned in **Table 5**). Specifically, the Dice score is enhanced from 77.511% to 81.794%, sensitivity is enhanced from 80.049% to 80.682%, specificity is enhanced from 99.311% to 99.536%, and overall pixel classification accuracy is enhanced from 98.220% to 98.480%. These results confirm that the Dice loss function is more efficient than the cross-entropy loss function in the segmentation problem. Using our proposed approach, we obtained the best segmentation performance with a Dice score of 83.623%, sensitivity of 82.606%, specificity of 99.562%, and overall pixel classification accuracy of 98.480%. Although the best sensitivity in **Table 6** is slightly lower than that reported in **Table 5**, the difference is very small (approximately 0.3%). In addition, we obtained the highest Dice score of 83.623%, which is much higher than 79.589% in **Table 5**. Based on these experimental results, we again confirm that our proposed approach with the Dice loss and “DOMINANT” combination rule outperforms conventional residual U-Net and two combination rules (“AND” and “OR”) in breast lesion segmentation problems using the BUS dataset.

**Table 6.** Performance measurement as a segmentation framework using the Dice loss and three combination rules (unit: %)

Using Dice Loss	Dice	Sensitivity	Specificity	Accuracy
Residual U-Net	81.794	80.682	99.536	98.480
AND rule	80.170	87.411	98.772	98.220
OR rule	78.695	73.422	99.828	98.240
DOMINANT rule	83.623	82.606	99.562	98.480

### 3.3.2.2. Experimental Result Using the BUSI Dataset

In this section, we evaluate the performance of our proposed approach using the BUSI dataset. From our experiment with the BUS dataset, we observe that the “DOMINANT” combination rule yields the best performance compared to conventional residual U-Net as well as the “AND” and “OR” rules. Therefore, we performed experiments using only the “DOMINANT” combination rule in this section. The detailed experimental results are presented in **Table 7**.

In this experiment, we measured the Dice score with additional sensitivity, specificity, and overall pixel classification accuracy. As shown in **Table 7**, the conventional residual U-Net segmentation system with the cross-entropy loss function produced an average Dice score of 74.567%, sensitivity of 78.128%, specificity of 98.008%, and overall pixel classification accuracy of 95.760%. By using our proposed approach, we enhanced the Dice score to 76.509%, which is about 2% higher than the conventional system. Sensitivity, specificity, and overall pixel classification accuracy were also enhanced to 80.640%, 98.173%, and 96.060%. These results indicate that the multi-scale approach is sufficient to enhance the segmentation performance with the BUSI dataset.

A similar situation occurred when using the Dice loss function. As reported in the latter part of **Table 7**, using the Dice loss function yielded similar segmentation results. In detail, the conventional segmentation network yielded 74.223% Dice, 72.579% sensitivity, 98.721% specificity, and 95.760% overall pixel classification values. Using our proposed approach, we enhanced the Dice score to 77.768%, which was the highest Dice score in the experiment with the BUSI dataset. In addition, the measurement results of sensitivity, specificity, and overall pixel classification accuracy were also enhanced when applying the multi-scale approach compared to the results obtained by the conventional segmentation network. Through these experimental results, we confirmed that the multi-scale approach is more efficient and accurate than the conventional segmentation network with the BUSI dataset.

**Table 7.** Performance evaluation as a segmentation framework using the BUSI dataset (unit: %)

Loss Method		Dice	Sensitivity	Specificity	Accuracy
CE Loss	Residual U-Net	74.567	78.128	98.008	95.760
	Proposed approach	76.509	80.640	98.173	96.060
Dice Loss	Residual U-Net	74.223	72.579	98.721	95.760
	Proposed approach	77.768	75.445	98.953	96.100

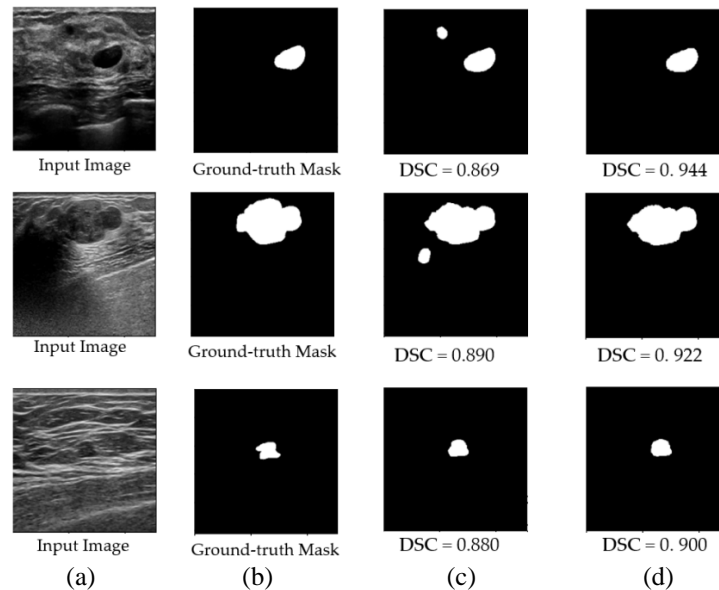
### 3.4. Comparison with Previous Study and Discussion

As discussed in section 3.3, our proposed approach is sufficient for segmenting breast nodule lesions captured in ultrasound images. In **Fig. 7**, we show some examples of the experimental results of the proposed approach on testing images. In this figure, we show the results of a segmentation system with and without our proposed approach (**Figs. 7d** and **7c**), along with the input captured grey-level ultrasound image (**Fig. 7a**) and ground-truth lesion region (**Fig. 7b**), which was given by an expert radiologist. For easy observation, we provide the measured Dice score (DSC) along with the prediction results. As observed from this figure, the conventional segmentation system (without our proposed approach) can efficiently segment breast lesions from ultrasound images. However, it can also produce incorrect lesion regions because of the appearance of noise, as depicted in the top and middle rows of **Fig. 7**. As a result, the Dice score was affected. Using our proposed approach, we efficiently removed the incorrectly segmented region while maintaining the correct region. Consequently, the Dice score was significantly enhanced. In the last row of **Fig. 7**, the two results of the segmentation systems (with and without the proposed approach) are similar, although it is difficult to segment the breast lesion by human perception. In addition, the segmentation result of the proposed approach was slightly higher than that of the conventional method.

**Fig. 8** shows some example cases in which the proposed approach is worse than the conventional segmentation system. Similar to **Fig. 7**, we provide the measured Dice score for each prediction result image. From this figure, the case when our proposed approach is worse than the conventional segmentation system occurs when the input images contain clear breast

lesion regions. As a result, there is little noise in this type of images, which yields the high performance of conventional segmentation systems. However, we observe that the difference in the Dice scores between the prediction results of our proposed approach and the conventional system is very small (0.960 vs. 0.954 in the first row of Fig. 8, and 0.908 vs. 0.906 in the second row of Fig. 8). In addition, in section 3.3, we demonstrated that our proposed approach outperforms the conventional segmentation system on average. Therefore, the proposed method is more efficient than the conventional system for breast lesion segmentation.

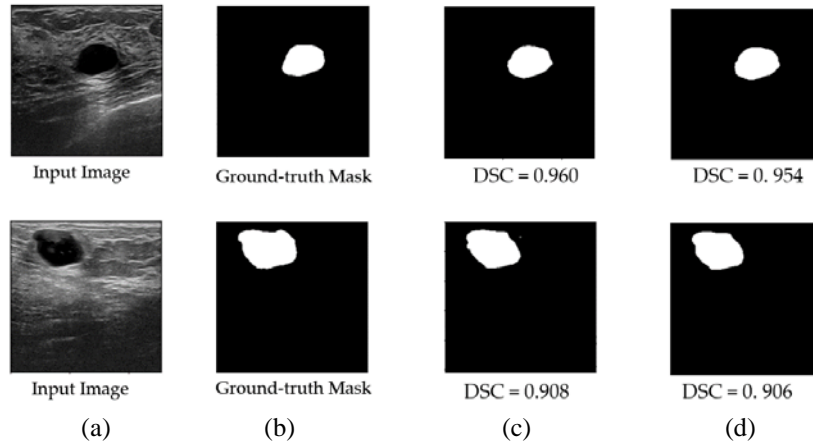
To validate the performance of our proposed approach, we conducted a final experiment to compare its performance with that obtained in previous studies. In this experiment, we compared the segmentation performance for the two datasets, as shown in Tables 8 and 9 for the BUS and in Table 10 for the BUSI datasets. Recently, Zhou et al. [22] proposed a novel network architecture called the UNet++ network, which has been proven to work well in medical image segmentation problems. The UNet++ network can be considered as a nested network of shallow and deep U-Net networks. As a result, it can produce better segmentation performance than a conventional U-Net network. For comparison, we performed experiments using the UNet++ network with the BUS and BUSI datasets, as reported in Tables 8, 9, and 10.



**Fig. 7.** Example results for segmentation network with and without our proposed method: (a) input image, (b) ground-truth lesion region given by expert radiologists, (c) result obtained by the segmentation network without our proposed method, and (d) result obtained by our proposed method

For the experiment with the BUS dataset, we compared the performance using two evaluation methods, namely the DP-based and IOU-based methods [18], and conventional residual U-Net [20]. Yap et al. [18] used three types of segmentation and detection networks to localize breast lesions, including a fully connected network with FCN-AlexNet, FRCNN, and FRCNN with RGB images. They reported their segmentation results based on the detected points, as listed in Table 8. Using FCN-AlexNet, they obtained a recall of 90.80%, precision of 86.05%, and F1-score of 88.36%. Similarly, they obtained values for recall, precision, and F1-score of 91.41%, 93.71%, and 92.55%, respectively, using FRCNN, and 85.89%, 88.61%, and 87.23%, respectively, using FRCNN with RGB images. Using the UNet++ network, we obtained a recall of 98.68%, precision of 92.50%, and F1-score of 95.42%. Using our proposed

method, we obtained a recall of 99.36%, precision of 94.36% and F1-score of 96.50% when using cross-entropy loss; and a recall of 99.36%, precision of 94.36%, and F1-score of 96.74% when using Dice loss. This result indicates that our performance measurements were much better than those obtained in previous studies [18] [20] [22].



**Fig. 8.** Example errors of segmentation network with and without our proposed method: (a) input image, (b) ground-truth lesion region, (c) result obtained by the segmentation network without our proposed method, and (d) result obtained by our proposed method

In **Table 9**, the comparison between our obtained performance and those produced by Yap et al. [18] using the IOU-based method is reported. As shown in this table, our proposed method with the Dice loss function produced the best segmentation result, with a recall of 99.34%, precision of 89.96%, and F1-score of 94.28%. These performance measurements were much higher than those reported in previous studies [18] [20] [22].

**Table 8.** Performance comparison of our proposed approach with previous studies based on DP-based evaluation method using the BUS [19] dataset (unit: %)

Method	Recall	Precision	F1-Score
Residual U-Net [20] with CE loss	98.12	92.50	95.12
Residual U-Net [20] with Dice loss	98.76	94.36	96.44
FCN-AlexNet [18]	90.80	86.05	88.36
FRCNN [18]	91.41	93.71	92.55
FRCNN(RGB) [18]	85.89	88.61	87.23
UNet++ [22]	98.68	92.50	95.42
Ours (Using CE loss)	99.36	93.82	96.50
Ours (Using Dice loss)	99.36	94.36	96.74

**Table 9.** Performance comparison of our proposed method with previous studies based on the IOU-based evaluation method using the BUS dataset (unit: %)

Method	Recall	Precision	F1-Score
Residual U-Net [20] with CE loss	97.92	83.64	89.98
Residual U-Net [20] with Dice loss	98.66	87.46	92.60
FCN-AlexNet [18]	67.49	63.95	65.67
FRCNN [18]	87.73	89.94	88.82
FRCNN(RGB) [18]	82.21	84.81	83.49
UNet++ [22]	98.60	83.12	90.08
Ours (Using CE loss)	99.34	85.76	91.92
Ours (Using Dice loss)	99.34	89.96	94.28

For the case of using the BUSI dataset, the previous study by Byra et al. [38] proposed an SK-UNet with or without fine-tuning for segmentation purposes. In their study, they reported a Dice score of 63.70% using their proposed network when training from scratch and 68.90% when fine-tuning a pre-trained network. As reported in section 3.3 and Table 7, our proposed method produced a Dice score of 76.509% when training with the cross-entropy loss function and 77.768% when training with the Dice loss function. Byra et al. [38] reported the highest image pixel classification accuracy of 93.0%, whereas our best pixel classification accuracy was approximately 96.1%. Our results are also better than the performance yielded by the UNet++ network, which produced an average Dice score of 68.046% and overall accuracy of 95.072%. Consequently, we state that our proposed approach is more efficient than previous studies for breast lesion localization/segmentation tasks using the BUS [19] and BUSI [37] datasets.

**Table 10.** Performance comparison of our proposed method with previous studies using the BUSI dataset (unit: %)

Method	Dice	Accuracy
Residual U-Net [20] with CE loss	74.567	95.760
Residual U-Net [20] with Dice loss	74.223	95.760
SK-U-Net without Fine-tuning [38]	63.700	91.900
SK-U-Net with Fine-tuning [38]	68.900	93.000
UNet++ [22]	68.046	95.072
Ours (Using CE loss)	76.509	96.060
Ours (Using Dice loss)	77.768	96.100

In Table 11, we compare the processing time of our proposed network with that of the conventional network. In this table, the conventional network indicates a segmentation network with a single original scale (segmentation using the original input image). As we can observe from this table, the conventional network takes 90.260 ms to produce the final result, whereas our proposed method takes 402.95 ms. This result indicates that our proposed method takes almost four times longer than a conventional segmentation network. However, this long processing time is acceptable in medical image processing applications, where accuracy, not processing time, is a mandatory requirement.

**Table 11.** Processing time of the segmentation network with and without the proposed approach (unit: ms)

Segmentation without the proposed approach (conventional network)	Segmentation with the proposed approach
90.260	402.95

### 3.5. Implication and Limitation of the Proposed Method

Based on the experimental results presented in the above sections, we conclude that our proposed method based on the use of multi-scale images is sufficient for enhancing the breast nodule segmentation performance compared to previous studies. In addition, by using a single segmentation network for segmentation purposes, we reduced the training time, system complexity and model size of the segmentation model compared to the ensemble of multiple network approaches.

However, because our proposed method uses multi-scale images for segmentation purposes, it requires a longer processing time than the conventional U-Net-based segmentation network that uses a single input image for breast nodule segmentation, as presented in Table 11. In

addition, our proposed method also requires additional steps for creating pyramid multi-scale images and combining the segmentation results of multi-scale images.

## 4. Conclusion

In this paper, we proposed a new approach for the breast nodule segmentation problem based on the use of multi-scale pyramid ultrasound images. The size of nodules varies according to the stage of breast disease. As a result, it reduces the performance of the segmentation methods, especially when detecting nodules in the early stage of breast cancer and/or when captured images are associated with much noise. To overcome this problem, we segmented breast lesions using multiple images at various image scales. By training the segmentation network with images at various scales, we developed the segmentation model invariant with nodule size. Finally, we enhanced the segmentation performance by combining the segmentation results across multiple image scales, as compared to previous studies. Through experiments with two public datasets, we confirmed the efficiency of our proposed approach. We made our implementation in public via our official website [42]. Therefore, other researchers can refer to our study for reference and comparison.

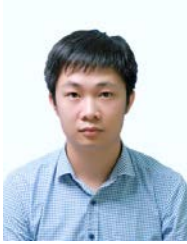
## References

- [1] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognit.*, vol. 43, no. 1, pp. 299–317, Jan. 2010. [Article \(CrossRef Link\)](#)
- [2] J. Shan, H. D. Cheng, and Y. Wang, "Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 262–275, Feb. 2012. [Article \(CrossRef Link\)](#)
- [3] Q.-H. Huang, S.-Y. Lee, L.-Z. Liu, M.-H. Lu, L.-W. Jin, and A.-H. Li, "A robust graph-based segmentation method for breast tumors in ultrasound images," *Ultrasonics*, vol. 52, no. 2, pp. 266–275, Feb. 2012. [Article \(CrossRef Link\)](#)
- [4] World Health Organization (WHO). [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on March. 02, 2023).
- [5] N.Y. Jung, B. J. Kang, H.S. Kim, et al., "Who could benefit the most from using a computer-aided detection system in full-field digital mammography?," *World J. Surg. Oncol.*, vol. 12, no. 1, 2014, Article no. 168. [Article \(CrossRef Link\)](#)
- [6] D. T. Nguyen, T. D. Pham, G. Batchuluun, H. S. Yoon, and K. R. Park, "Artificial Intelligence-Based Thyroid Nodule Classification Using Information from Spatial and Frequency Domains," *J. Clin. Med.*, vol. 8, no. 11, Nov. 2019. [Article \(CrossRef Link\)](#)
- [7] M. Imani, "Automatic diagnosis of coronavirus (COVID-19) using shape and texture characteristics extracted from X-Ray and CT-Scan images," *Biomed. Signal Process. Control*, vol. 68, p. 102602, Jul. 2021. [Article \(CrossRef Link\)](#)
- [8] J. De Moura et al., "Deep convolutional approaches for the analysis of Covid-19 using chest X-Ray images from portable devices," *IEEE Access*, vol. 8, pp. 195594–195607, 2020. [Article \(CrossRef Link\)](#)
- [9] J. Wang et al., "Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images," *IEEE Trans. Med. Imaging*, vol. 39, no. 8, pp. 2572–2583, Aug. 2020. [Article \(CrossRef Link\)](#)
- [10] M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model," *Data*, vol. 5, no. 1, 14, Feb. 2020. [Article \(CrossRef Link\)](#)



- [11] J. Mitra, P. Bourgeat, J. Fripp, S. Ghose, S. Rose, O. Salvado, A. Connelly, B. Campbell, S. Palmer, G. Sharma, S. Christensen, and L. Carey, "Lesion segmentation from multimodal MRI using random forest following ischemic stroke," *Neuroimage*, vol. 98, pp. 324–335, Sep. 2014. [Article \(CrossRef Link\)](#)
- [12] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: a dimension fusion U-shape network for chronic stroke lesion segmentation," *arXiv:1908.05104*, 2019. [Article \(CrossRef Link\)](#)
- [13] D. T. Nguyen, J. K. Kang, T. D. Pham, G. Batchuluun, and K. R. Park, "Ultrasound Image-Based Diagnosis of Malignant Thyroid Nodule Using Artificial Intelligence," *Sensors*, vol. 20, no. 7, 1822, Mar. 2020. [Article \(CrossRef Link\)](#)
- [14] T. Pang, J. H. D. Wong, W. L. Ng, and C. S. Chan, "Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification," *Comput. Meth. Programs Biomed.*, vol. 203, 106018, May 2021. [Article \(CrossRef Link\)](#)
- [15] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, "An open access thyroid ultrasound image database," in *Proc. of the 10th International Symposium on Medical Information Processing and Analysis*, Cartagena de Indias, Colombia, Vol. 9287, pp. 1–6, 28 January 2015. [Article \(CrossRef Link\)](#)
- [16] S. N. Acho and W. I. D. Rae, "Interactive breast mass segmentation using a convex active contour model with optimal threshold values," *Phys. Medica*, vol. 32, no. 10, pp. 1352–1359, Oct. 2016. [Article \(CrossRef Link\)](#)
- [17] A. Rampun, P. J. Morrow, B. W. Scotney, and J. Winder, "Fully automated breast boundary and pectoral muscle segmentation in mammograms," *Artif. Intell. Med.*, vol. 79, pp. 28–41, Jun. 2017. [Article \(CrossRef Link\)](#)
- [18] M. H. Yap, M. Goyal, F. Osman, R. Martí, E. Denton, A. Juette, and R. Zwigelaar, "Breast ultrasound region of interest detection and lesion localisation," *Artif. Intell. Med.*, vol. 107, 101880, Jul. 2020. [Article \(CrossRef Link\)](#)
- [19] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentis, R. Zwigelaar, A. K. Davison, and R. Martí, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018. [Article \(CrossRef Link\)](#)
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *arXiv 2015*, *arXiv:1505.04597*, 2015. [Article \(CrossRef Link\)](#)
- [21] V. K. Singh, M. Abdel-Nasser, F. Akram, H. A. Rashwan, M. M. K. Sarker, N. Pandey, S. Romani, and D. Puig, "Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework," *Expert Syst. Appl.*, vol. 162, 113870, Dec. 2020. [Article \(CrossRef Link\)](#)
- [22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *arXiv 2018*, *arXiv:1807.10165*, 2018. [Article \(CrossRef Link\)](#)
- [23] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, et al., "Connected-UNets: a deep learning architecture for breast mass segmentation," *npj Breast Cancer*, 7, 2021, Article No.151. [Article \(CrossRef Link\)](#)
- [24] Y. Hiramatsu, K. Hotta, A. Imanishi, M. Matsuda, and K. Terai, "Cell image segmentation by integrating multiple CNNs," in *Proc. of Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, pp. 2286–22866, 18–22 June 2018. [Article \(CrossRef Link\)](#)
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Article \(CrossRef Link\)](#)
- [26] K. Simonyan, and A. Zisserman, "Very deep convolutional neural networks for large-scale image recognition," *arXiv 2014*, *arXiv:1409.1556*, 2014. [Article \(CrossRef Link\)](#)
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv 2015*, *arXiv:1512.03385*, 2015. [Article \(CrossRef Link\)](#)
- [28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv 2014*, *arXiv:1409.4842v1*, 2014. [Article \(CrossRef Link\)](#)

- [29] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *arXiv 2016*, *arXiv:1608.06993*, 2016. [Article \(CrossRef Link\)](#)
- [30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv 2014*, *arXiv:1406.2661*, 2014. [Article \(CrossRef Link\)](#)
- [31] P. M. Chu, Y. Sung, and K. Cho, "Generative Adversarial Network-Based Method for Transforming Single RGB Image into 3D Point Cloud," *IEEE Access*, vol. 7, pp. 1021–1029, 2019. [Article \(CrossRef Link\)](#)
- [32] J-Y. Zhu, T. Park, P. Isola, and A-A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv 2017*, *arXiv:1703.10593v6*, 2017. [Article \(CrossRef Link\)](#)
- [33] Y. Choi, M. Choi, M. Kim, J-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," *arXiv 2017*, *arXiv:1711.09020*, 2017. [Article \(CrossRef Link\)](#)
- [34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *arXiv 2015*, *arXiv:1506.01497*, 2015. [Article \(CrossRef Link\)](#)
- [35] J. Redmon, and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv 2018*, *arXiv:1804.02767*, 2018. [Article \(CrossRef Link\)](#)
- [36] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *arXiv 2017*, *arXiv:1707.03237*, 2017. [Article \(CrossRef Link\)](#)
- [37] W. Al-Dhabyani, M. Goma, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief.*, vol. 28, 104863, Feb. 2020. [Article \(CrossRef Link\)](#)
- [38] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, and M. Andre, "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," *Biomed. Signal Process. Control*, vol. 61, 102027, Aug. 2020. [Article \(CrossRef Link\)](#)
- [39] A. W. Setiawan, "Image Segmentation Metrics in Skin Lesion: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index, and Matthews Correlation Coefficient," in *Proc. of the International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, Surabaya, Indonesia, pp. 97-102, 2020. [Article \(CrossRef Link\)](#)
- [40] K. Devanathan and E. S. D, "Lesion Segmentation in Dermoscopic Images using Superpixel based Fast Fuzzy C-means Clustering," in *Proc. of the IEEE Congreso Biental de Argentina (ARGENCON)*, Resistencia, Argentina, pp. 1-6, 2020. [Article \(CrossRef Link\)](#)
- [41] Tensorflow library for deep learning platform. [Online]. Available: <https://www.tensorflow.org/> (accessed on March. 02, 2023)
- [42] HYU-BREAST-SEGNET Model. [Online]. Available: [http://fme.utehy.edu.vn/AI\\_lab](http://fme.utehy.edu.vn/AI_lab) (accessed on March. 02, 2023)



**Quoc Tuan Hoang** received a B. S. degree in automation engineering from Hung Yen University of Technology and Education, Vietnam, in 2008, and his Master's degree in automation engineering at Le Quy Don Technical University, Vietnam, in 2012. Currently, He is a Ph.D. candidate at Hung Yen University of Technology and Education, Vietnam. His research interests include automation, embedded systems, and digital signal processing.



**Xuan Hien Pham** received a B.S degree in electrical engineering from Hung Yen University of Technology and Education, Vietnam, in 2008, and the Ph.D. degree in electrical engineering from Dongguk University, Republic of Korea in 2016. He has been a lecturer at the Department of Mechanical Engineering of University of Transport and Communications Technology and Education, Vietnam, since 2019. His research interests include gaseous electronic, digital signal processing, control and automation.



**Anh Vu Le** received the B.S. degree in electronics and telecommunications from the Hanoi University of Technology, Vietnam, in 2007, and the Ph.D. degree in electronics and electrical from Dongguk University, South Korea, in 2015. Currently, He is with the Communication and Signal Processing Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. He is also working as a Postdoctoral Research Fellow with the ROAR Laboratory, Singapore University of Technology and Design. His current research interests include robotics vision, robot navigation, human detection, action recognition, feature matching, and 3D video processing.



**Trung Thanh Bui** received B. S. degree in Telecommunication, University of Communication and Transport, Hanoi, Vietnam, in 2000. He received his Master's and Ph.D. degrees in Mechatronics, Asian Institute of Technology, Thailand, in 2005 and 2008, respectively. He was dean of the Mechanical engineering faculty at Hung Yen University of Technology and Education since April 2013. From April 2019, he has been Rector of Hung Yen University of Technology and Education. His research interests include Robotics (Mobile robots, Human-robot interaction, sensor fusion) and embedded systems.