



ELSEVIER

Available at www.sciencedirect.com**ScienceDirect**journal homepage: www.elsevier.com/locate/bbe

Original Research Article

A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning



Ademola Enitan Ilesanmi ^{a,*}, Utairat Chaumrattanakul ^b, Stanislav S. Makhanov ^{a,*}

^a School of ICT, Sirindhorn International Institute of Technology, Thammasat University, Thailand

^b Department of Radiology, Thammasat University, Pathum Thani, Thailand

ARTICLE INFO

Article history:

Received 14 January 2021

Received in revised form

19 May 2021

Accepted 21 May 2021

Available online 31 May 2021

Keywords:

Breast ultrasound image

Speckle noise

Deep learning

Semantic segmentation

Tumor segmentation

ABSTRACT

Background: Breast cancer is a deadly disease responsible for statistical yearly global death. Identification of cancer tumors is quite tasking, as a result, concerted efforts are thus devoted. Clinicians have used ultrasounds as a diagnostic tool for breast cancer, though, poor image quality is a major limitation when segmenting breast ultrasound. To address this problem, we present a semantic segmentation method for breast ultrasound (BUS) images.

Method: The BUS images were resized and then enhanced with the contrast limited adaptive histogram equalization method. Subsequently, the variant enhanced block was used to encode the preprocessed image. Finally, the concatenated convolutions produced the segmentation mask.

Results: The proposed method was evaluated with two datasets. The datasets contain 264 and 830 BUS images respectively. Dice measure (DM), Jaccard measure, and Hausdroff distance were used to evaluate the methods. Results indicate that the proposed method achieves high DM with 89.73% for malignant and 89.62% for benign BUSs. Moreover, the results obtained validate the capacity of the proposed method to achieve higher DM in comparison with reported methods.

Conclusion: The proposed algorithm provides a deep learning segmentation procedure that can segment tumors in BUS images effectively and efficiently.

© 2021 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

1. Introduction

Cancer is a deadly disease that starts from a specific part of the body and spreads to other organs in the body. Statistics from the World Health Organization suggest that developing coun-

tries suffer the highest death from breast cancer. Early detection, diagnosis, and treatment are preventive measures to avoid complications from breast cancer [1,2]. Breast Ultrasound is one of the numerous imaging modalities that assist clinicians in assessing and diagnosing breast cancer. BUS is

* Corresponding authors at: Sirindhorn International Institute of Technology, Thammasat University, Thailand.

E-mail addresses: ademola.eni@dome.tu.ac.th (A.E. Ilesanmi), makhanov@siit.tu.ac.th (S.S. Makhanov).

<https://doi.org/10.1016/j.bbe.2021.05.007>

0168-8227/© 2021 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

noninvasive and does not use ionizing radiation [3]. Procedures to segment BUSs can be divided into three forms: manual, semi-automatic, and fully automatic [4]. Manual segmentation solely relies on the expertise of experienced doctors. Manual segmentation is the gold standard for medical image processing. Unfortunately, it is time-consuming and laborious. Semi-automatic segmentation does not solely rely on human effort; it involves minimal human interaction for accurate and effective results. Some historic semi-automatic methods include paintbrush algorithm [5] (this method is based on watershed), active contours [6], and constraint-based region growing [7]. Automatic segmentation does not involve user interactions when segmenting tumors [8–11].

Algorithms to segment BUSs are divided into four categories: graph-based methods, deformable models (DM), learning methods, and classical methods [12]. The graph-based method uses graph analysis for segmentation. Graph methods are simple, flexible with soft constraints [13,14]. DM methods segment images using a curved surface technique. The curved surface technique moves the object boundary with a force defined on the surface of the tumor region. Essentially, DMs are classified as parametric or geometric. Thresholding, region growing, and watershed methods are popular classical methods. Watershed divides images into different segments using marker selection [15]. The region growing performs segmentation by extracting regions from a set of pixels to bigger regions. The thresholding method is used for preprocessing; they group pixels into single or multiple classes.

In this paper, a deep learning approach that segments BUS images is proposed. The proposed method is an end-to-end network inspired by the UNet architecture. Our method concatenates pooling strategy and batch normalization. A double convolution layer was added to the ReLU layer in the decoding framework. In simple terms, the proposed method (known as VEU-Net) adopts the VE block for the encoding framework and a double convolution layer for the decoding framework (see Fig. 7). Images from well-known datasets were used to test the proposed methods. VEU-Net competes with several state-of-the-art methods. The major contributions of this paper are summarized as follows:

- 1) A multi-pooling technique that merges the max and average pooling.
- 2) A double concatenated convolution technique for the decoding framework.
- 3) Low computational cost with higher accuracy.

The paper is organized as follows. A detailed literature review is discussed in Section 2. Then, the basic concepts and proposed method are reported in Section 3. The results of the experiment are reported in Section 4. Finally, a discussion is given in Section 5 and the paper is concluded in Section 6.

2. Related work

Several 2D and 3D semantic segmentation methods have been reviewed by references [16,18]. Specifically, reference [17,18] reviewed several semantic segmentation methods for

Table 1 – Summary of some recent semantic segmentation methods to segment BUS images (see Ref [17,18] for more details).

Authors	Ref	year	Number of Images	Performance Measure	Advantages	Disadvantages
Hiramatsu et al. Kumar et al.	[24] [23]	2017 2018	— 258	— Dice = 0.84	Rapid downampling of representation Use several image layers and hence produce good accuracy and fast training.	No fixed pattern of arrangement Difficult to train weight due to vanishing gradient problem.
Yang et al.	[19]	2019	454	Dice = 0.81	Can give an accurate result with a small sample dataset	Computationally expensive
Byra et al.	[30]	2020	882 and 893	Dice = 0.82	Effectively control the number of features to avoid overfitting	Do not preserve spatial and localization information effectively
Vakanski et al.	[29]	2020	510	Dice = 0.90	Focus on the target with reducing complexity and can easily be trained from scratch.	Produce long training time due to added weight
Huang et al.	[22]	2020	320	Dice = 0.90	Eliminate background noise and convergence	Depend on several parameters that give low accuracy

BUS images. Notwithstanding, we will examine some methods. Yang et al. [19] proposed the Dual-Path U-Net (DPU-Net) architecture for the segmentation of 3D cardiovascular images. The DPU-Net involves two phases: the first phase uses a small number of training samples to achieve good generalization. The second phase connects the DPU-Net to the augmentor to control the process. Precisely, the DPU-Net performs segmentation using different augmentation for cardiovascular images. The advantage of DPU-Net is that it uses a real-time augmentation technique that can give accurate segmentation from small sample datasets. However, such real-time augmentation is cumbersome and computationally expensive, and hence may not be suitable for BUS images. Li et al. [20] proposed the attention-based Nested U-Net (ANU-Net) for the segmentation of liver tumors. ANU-Net introduces the nested attention mechanism created from Densely Connected Convolutional Network (DenseNet). ANU-Net uses the attention gate mechanism to suppress irrelevant regions and highlight salient features. This suppression technique eliminates the usage of an explicit localization module that focuses on relevant regions on the segmentation task. A major drawback of ANU-Net is the increased training time. The weight added by the attention gate prolongs training time. Lian et al. [21] proposed the Attention Net (ATT-UNet) that guides the model to learn features indiscriminately. ATT-UNet has a bounding box at the base of its model with masked images used as a weighting function. The masked images, feature maps, pooling layers, and fully connected layers are combined to produce the final segmented image. ATT-UNet uses the attention mask that

forces the region to focus on the tumor. Improved retention of fine detail features and fast training time are advantages of ATT-UNet. A major drawback is the slow learning procedure in the middle of the model which causes ignored abstract features (especially for deeper models).

Huang et al. [22] used the simple linear iterative clustering (SLIC) algorithm and semantic segmentation to detect tumors in BUS. First, the original image was cropped by the radiologist to generate the region of interest (ROI). Then bilateral filtering and histogram equalization were used for noise removal and enhancement respectively. Next, the means shift and K-means generate initial segmentation, finally, morphological post-processing segment tumor. This method may effectively eliminate background noise and convergence. However, it depends on initial values, clustering data, outliers, and scaling dimensions that may reduce segmentation accuracy. Kumar et al. [23] segment breast tumors with the improved UNet algorithm (multiple UNet). First, a pooling technique was performed on the original image. Next, the data augmentation algorithm and weighting pixel procedure were used for segmentation. The method used multiple UNet to produce good accuracy and fast training of deep networks. A major drawback is an effect caused by a vanishing gradient that makes it difficult to train weights in the model. Hiramatsu et al. [24] use the AlexNet architecture to segment breast tumors. The original image was filtered and segmented with the thresholding and region growing algorithms. Finally, the AlexNet architecture with 5 convolution layers and 3 max-pooling layers segment the image. A major characteristic of AlexNet is the rapid downsampling of representation (with

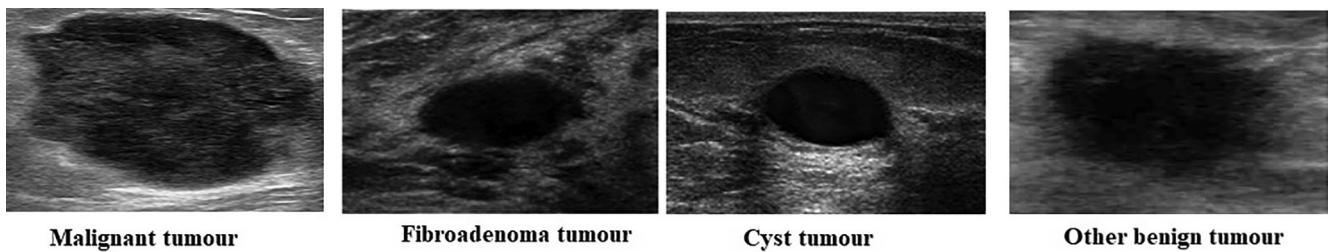


Fig. 1 – BUS tumor type [33].

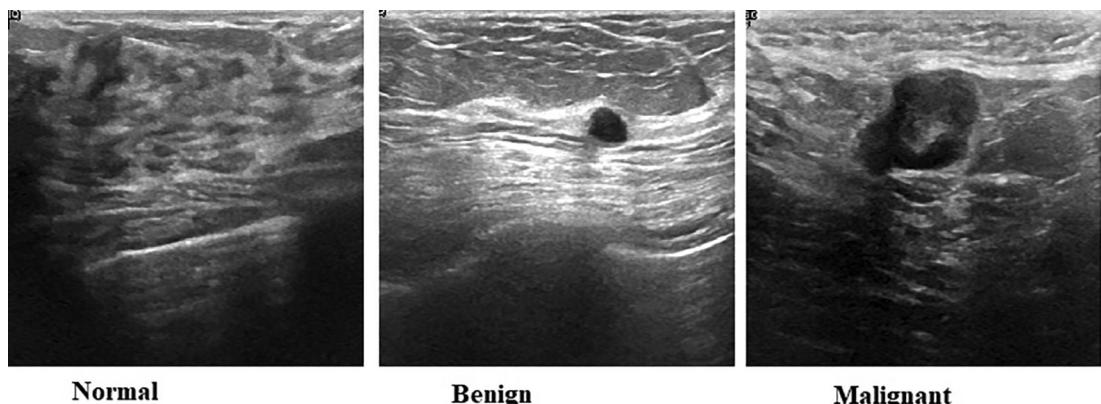


Fig. 2 – BUS type [34].

stride convolutions and max-pooling) when compared to other networks. However, AlexNet does not have fixed patterns; convolutions for each layer are decided experimentally.

Reza et al. [25] investigate the efficiency of convolutional neural networks (CNNs) on liver segmentation in animal models. The experiment used the UNet [26], V-Net [27], and the feature pyramid network (FPN) [28] for automatic liver segmentation. This method produced a multi-scale feature representation that semantically enhanced all feature levels, however, combinations of different models if not properly managed may produce low accuracy. Reference [29] added the attention gate to each encoding block of the UNet architecture. This method does not involve several convolutions, but focuses on a target, reduces complexity, and can be trained from scratch. However, because of the added weights in the model, they require lots of training time. Byra et al. [30] used the selective kernel (SK-UNet) architecture to segment breast ultrasound. The SK-UNet replaced each block in the encoding and decoding framework and consists of Batch normalization, ReLU (rectified linear unit), Fully Connected layers, and a pooling layer (SK layer has 12 convolutions). A concatenation procedure is performed on the encoding and decoding blocks to produce a segmentation mask. This method effectively controls the number of features and avoids over-fitting. However, they do not preserve spatial information and localization.

Despite the utilization of deep learning algorithms in almost every sphere of academic endeavor (including computer vision), deep learning for the segmentation of medical images is still a challenging task (see refs [31,32] for details). Although several methods have segmented BUS with deep learning approaches, we believe that the proposed method can reduce the use of several convolutions observed in some literature. The use of the VE block and concatenated convolution can give good segmentation accuracy. Finally, our method will act as an option for segmenting BUS images.

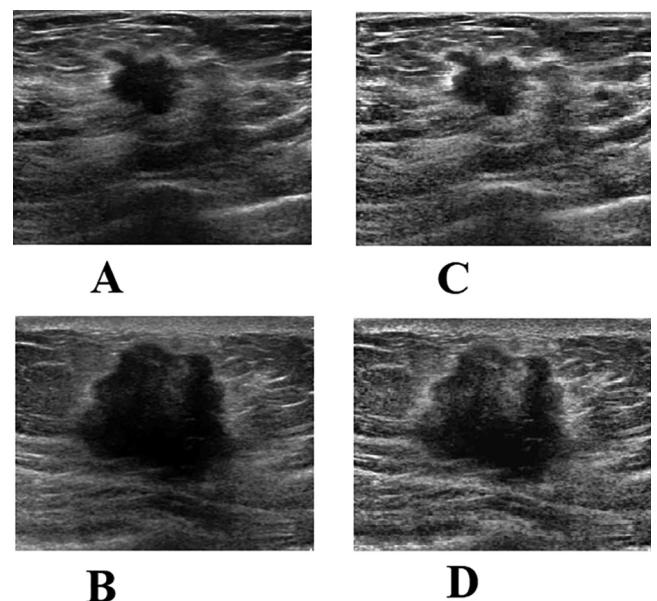


Fig. 4 – Pictorial representation of image enhancement procedure: (A) and (B) Original images; (C) and (D) results of CLAHE operation.

Table 1 is a summary of some recent semantic segmentation methods for BUS images.

3. Methods and materials

3.1. Material

The proposed algorithm was tested with two datasets; the first dataset consists of 264 images comprising 100 malignant, and 164 benign BUS [33]. The second dataset is made up of 830

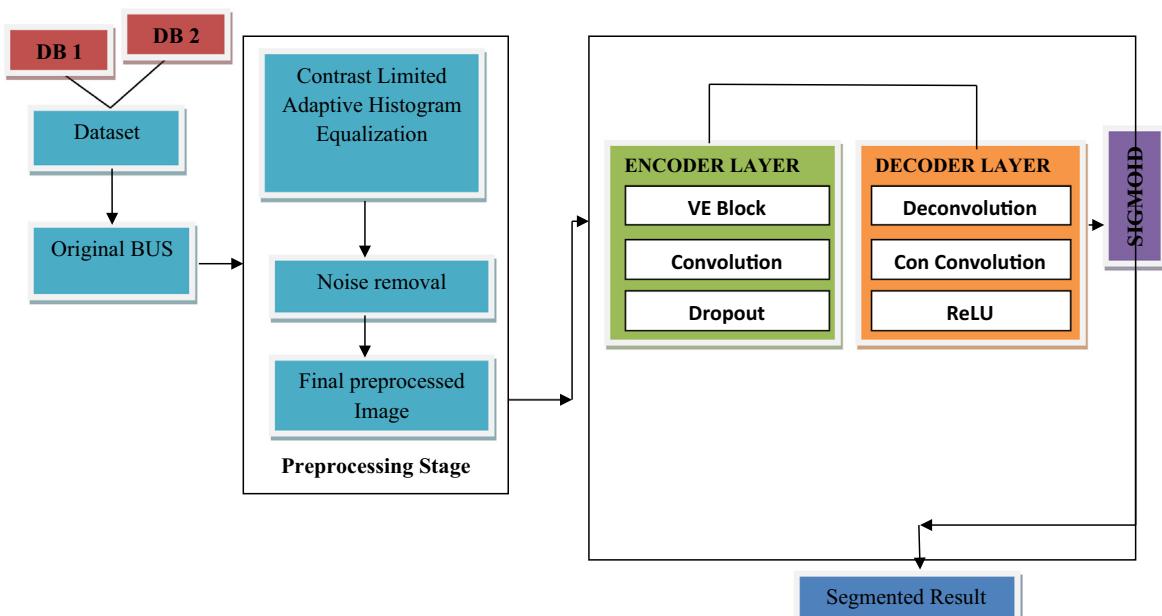


Fig. 3 – Block diagram of the proposed method.

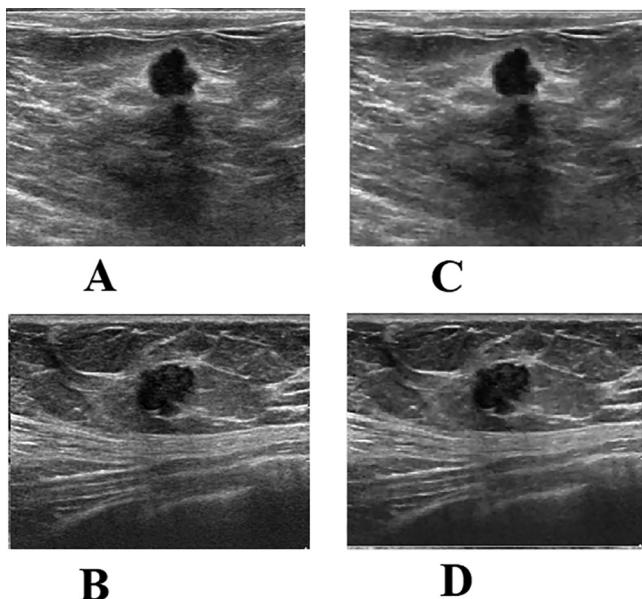


Fig. 5 – Pictorial representation of noise removal procedure: (A) and (B) Original images; (C) and (D) results of bilateral filtering.

images, with 487 malignant, 210 benign, and 133 normal BUS [34]. The image composition of BUSs tumor can be seen in Figs. 1 and 2. Images from the first dataset were obtained from the Thammasat University Hospital (<http://onlinemedicalimages.com/index.php/en/>). Philips iU22 ultrasound machine was used to obtain the BUS. A total of 3 leading expert radiologists participated in drawing the ground truth. The experts are renowned members of the Department of

Radiology, in the University. An electronic pen and Samsung Galaxy Tablet computer were used in drawing the ground truth. To arrive at the final ground truth image, a majority voting mechanism was adopted. The image resolution ranges from 540×720 to 2345×3549 pixels. The reason for resizing images for the CNN application is because smaller images train faster with CNN (see Refs [35–38]). Hence, images in the two datasets were resized to 256×256 respectively.

Images from the second dataset were obtained from the Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt (<https://scholar.cu.edu.eg/?q%afahmy-pages/dataset>). The LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound machine were used to obtain the images. Transducers used are 1–5 MHz on ML6-15-D matrix linear probe. It is important to note that for all data used in this research, patient and hospital were adequately informed about ethical considerations and all data stay unknown. Written consent documentation was signed by the hospital and patient for the use of the data for experimental purposes only.

BUS images have unique and specific features that present extreme difficulty when segmenting. For example, the malignant tumor gives an irregular shape making it difficult to segment. The fibroadenoma is characterized by low contrast, while the Cyst gives a regular shape but is attached to irregular shadows (see Fig. 1). Overall, the benign BUS is characterized by a regular shape while the malignant BUS produces an irregular shape pattern.

3.2. Methods

The proposed method involves the preprocessing stage, and the end-to-end deep learning segmentation stage. The block diagram of the proposed framework is shown in Fig. 3.

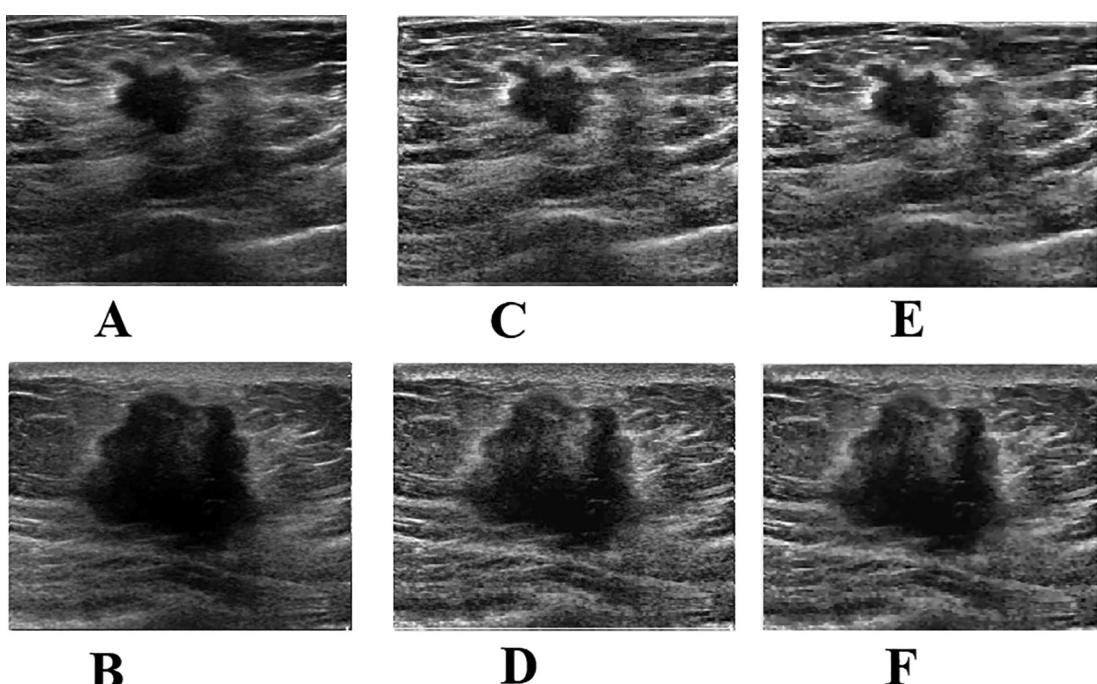


Fig. 6 – Pictorial representation of the preprocessing stage: (A), (B) Original images; (C), (D) CLAHE enhanced images; (E), (F) Bilateral filtering after enhancement.

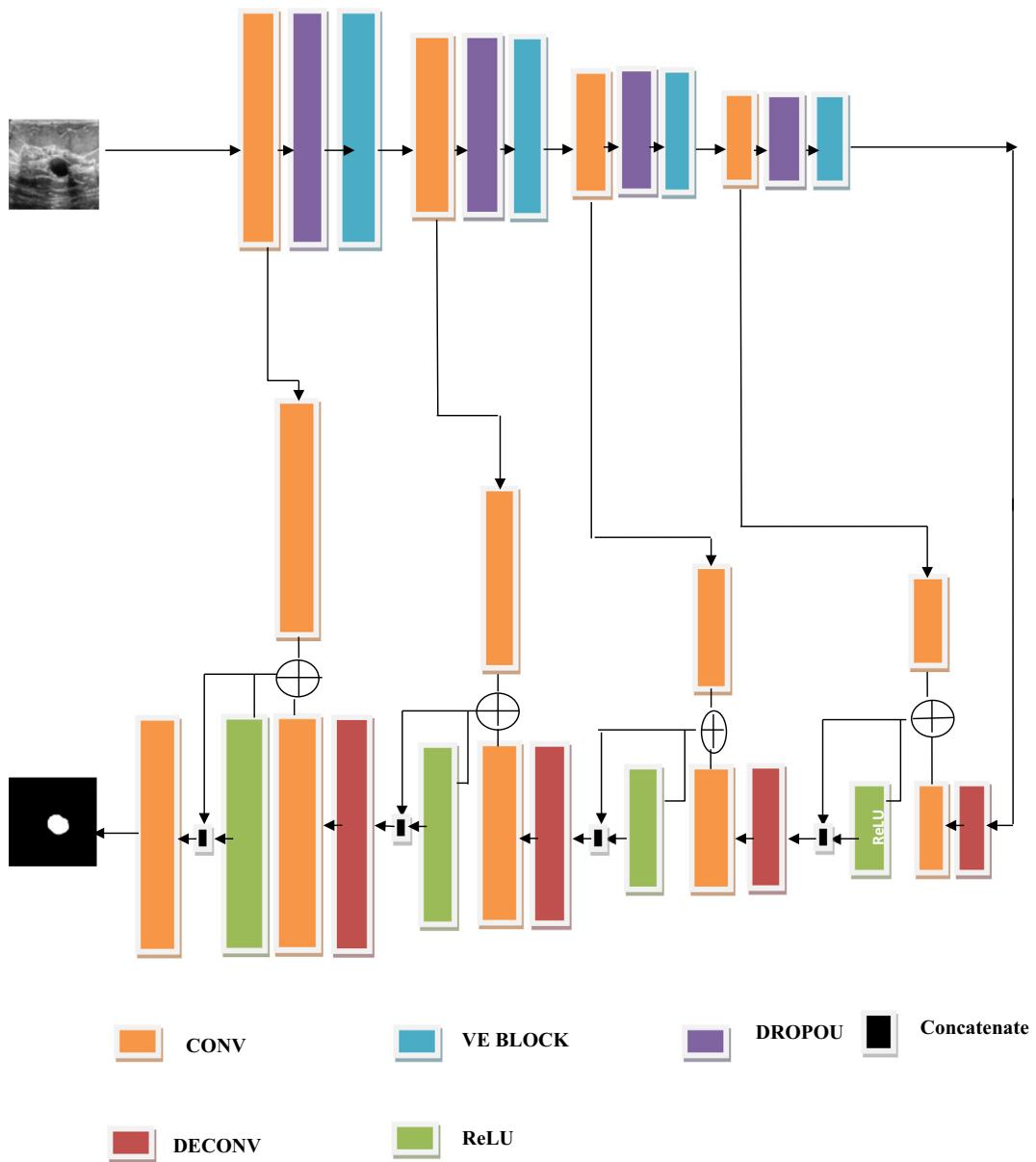


Fig. 7 – Diagram of the VEU-Net segmentation.

3.2.1. Preprocessing stage

Images used in this research were obtained from two datasets. Thereafter, all the images were resized to a minimum dimension for the experimental trial. The minimum experimental dimension obtained in our work is 256×256 . Hence, all the images in the two datasets were resized to 256×256

respectively. Other pre-processing procedures are discussed in subsequent sub-sections.

3.2.1.1. Contrast limited adaptive histogram equalization (CLAHE). Contrast limited adaptive histogram equalization (CLAHE) is used to improve the low quality of images. The

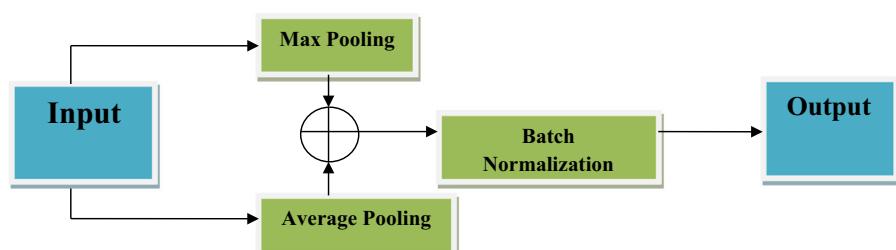


Fig. 8 – Diagram of the Variant Enhanced Block.

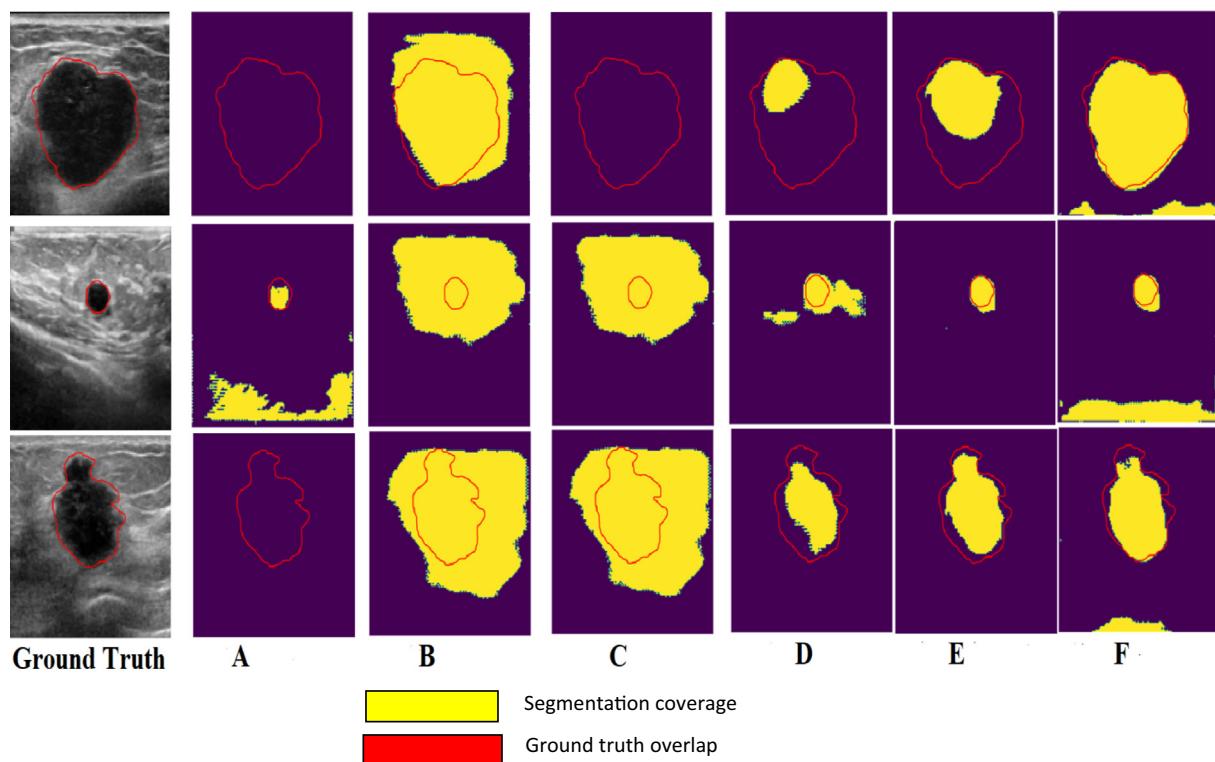


Fig. 9 – BUS segmentation (Philips iU22 machine): (A) SegNet; (B) UNet; (C) MultiNet; (D) Dual UNet; (E) VEU-Net (without VE block); (F) Proposed method.

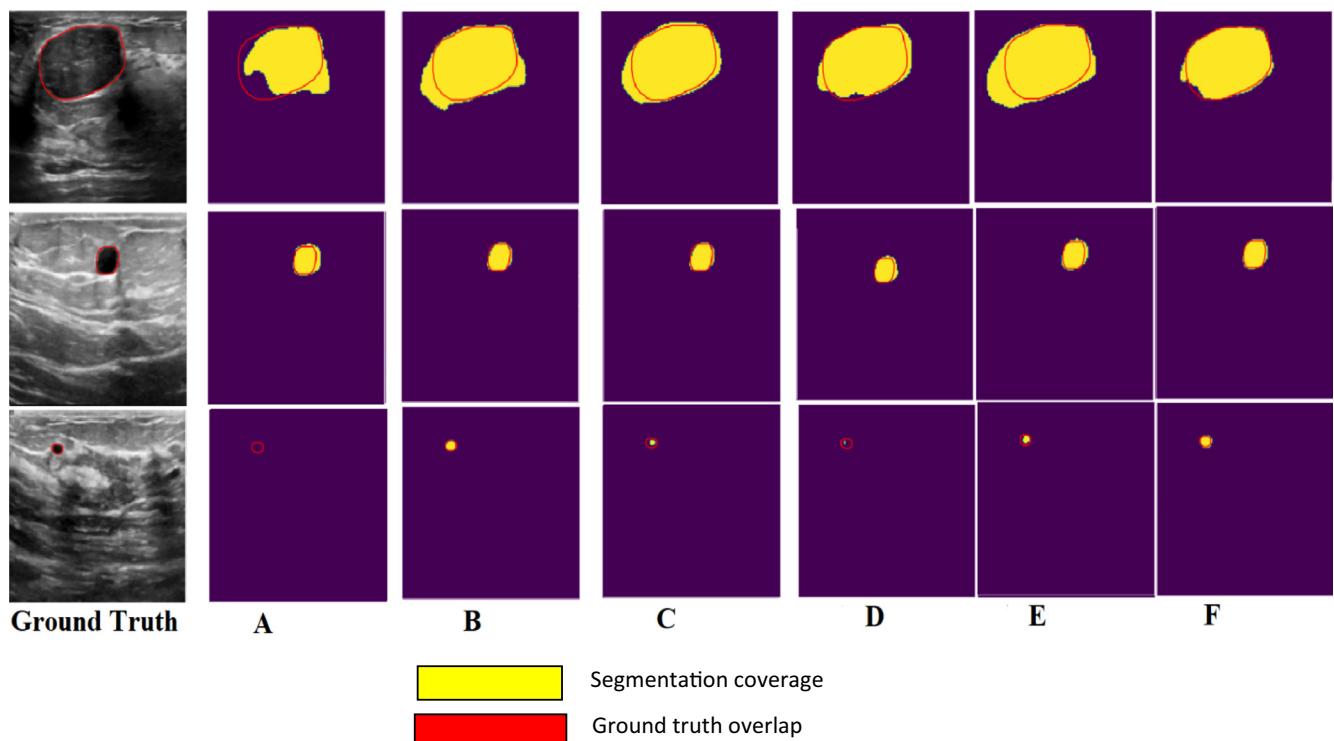


Fig. 10 – BUS segmentation (LOGIQ E9 and LOGIQ E9 Agile): (A) SegNet (B) UNet; (C) MultiUNet; (D) Dual UNet; (E) VEU-Net (without VE block); (F) Proposed method.

Table 2 – Performance result for benign images obtained with Philips iU22 machine.

Algorithm	HD	JM	DM
UNet [26]	11.19 ± 5.06	50.24 ± 3.76	63.46 ± 4.12
Dual U-Net [57]	10.83 ± 4.92	55.31 ± 4.43	66.23 ± 5.71
SegNet [58]	11.51 ± 5.19	50.85 ± 4.76	63.75 ± 4.77
MultiUNet [59]	12.92 ± 5.73	43.89 ± 6.79	54.32 ± 6.42
VIU-Net (Without VE block)	10.31 ± 4.80	60.91 ± 5.81	71.89 ± 5.94
Proposed method	8.10 ± 4.24	63.21 ± 5.76	74.46 ± 5.12

Table 3 – Performance result for malignant images obtained with Philips iU22 machine.

Algorithm	HD	JM	DM
UNet [26]	12.43 ± 6.06	52.72 ± 5.73	64.75 ± 5.69
Dual U-Net [57]	11.78 ± 4.73	56.83 ± 4.76	67.31 ± 5.44
SegNet [58]	12.45 ± 6.19	50.77 ± 4.86	62.85 ± 4.84
MultiUNet [59]	12.72 ± 6.73	47.94 ± 6.99	57.49 ± 6.31
VIU-Net (Without VE block)	10.71 ± 4.22	61.42 ± 4.23	72.89 ± 4.22
Proposed method	8.66 ± 4.46	65.26 ± 4.01	76.07 ± 4.07

Table 4 – Performance result for benign images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine.

Algorithm	HD	JM	DM
UNet [26]	9.42 ± 4.31	75.34 ± 5.60	86.21 ± 5.80
Dual U-Net [57]	11.26 ± 5.82	70.44 ± 5.89	82.46 ± 5.61
SegNet [58]	11.89 ± 5.93	72.43 ± 5.41	85.31 ± 5.90
MultiUNet [59]	13.43 ± 6.24	69.43 ± 6.62	82.89 ± 6.29
VIU-Net (Without VE block)	9.78 ± 4.63	74.24 ± 5.92	86.43 ± 5.88
Proposed method	8.80 ± 4.20	78.43 ± 5.06	89.62 ± 5.62

Table 5 – Performance result for malignant images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine.

Algorithm	HD	JM	DM
UNet [26]	9.20 ± 4.41	75.41 ± 5.12	86.94 ± 5.02
Dual U-Net [57]	12.06 ± 6.02	70.86 ± 5.93	82.66 ± 5.13
SegNet [58]	10.42 ± 5.46	73.20 ± 5.60	85.61 ± 5.60
MultiUNet [59]	13.08 ± 6.13	70.31 ± 6.01	82.77 ± 5.93
VIU-Net (Without VE block)	9.60 ± 4.76	76.31 ± 5.74	88.04 ± 5.21
Proposed method	8.40 ± 4.10	78.87 ± 5.12	89.73 ± 5.42

Table 6 – Ablation experiment setting.

Criteria	VE Block	Concatenated convolutions	Image enhancement and noise removal	Augmentation
Criteria 1	NO	NO	NO	YES
Criteria 2	NO	YES	YES	YES
Criteria 3	YES	YES	NO	YES
Proposed method	YES	YES	YES	YES

CLAHE algorithm enhances image quality by limiting the amplification and clipping the predefined histogram levels. Precisely, CLAHE is used to enhance the inverse of image intensity. The tremendous result from CLAHE has made it a better

option for image enhancement when compared to adaptive histogram equalization (AHE) or other histogram equalization methods [39]. In this paper, image quality enhancement was performed with the CLAHE algorithm (see Fig. 4).

Table 7 – Combined ablation result for benign and malignant images obtained with Philips iU22 machine (with augmentation).

Algorithm	HD	JM	DM
Criteria 1	8.96 ± 4.42	77.81 ± 5.46	88.63 ± 5.23
Criteria 2	8.73 ± 4.22	78.90 ± 5.32	88.68 ± 5.64
Criteria 3	8.32 ± 4.21	79.16 ± 5.21	90.09 ± 5.12
Proposed method	7.81 ± 3.89	80.07 ± 5.16	91.82 ± 5.11

Table 9 – Combined ablation result for benign and malignant images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine (with augmentation).

Algorithm	HD	JM	DM
Criteria 1	8.36 ± 4.12	77.21 ± 5.16	88.26 ± 5.13
Criteria 2	8.23 ± 4.08	78.89 ± 5.42	89.36 ± 5.04
Criteria 3	8.01 ± 4.00	79.42 ± 5.26	90.14 ± 5.33
Proposed method	7.71 ± 4.00	80.20 ± 4.83	91.88 ± 5.68

Table 8 – Combined ablation result for benign and malignant images obtained with Philips iU22 machine (without augmentation).

Algorithm	HD	JM	DM
Criteria 1	11.81 ± 5.56	51.48 ± 4.74	64.11 ± 4.90
Criteria 2	10.51 ± 4.51	61.16 ± 4.23	72.39 ± 4.22
Criteria 3	9.62 ± 4.91	62.12 ± 4.81	74.12 ± 5.12
Proposed method	8.38 ± 4.10	64.23 ± 4.88	75.27 ± 4.59

Table 10 – Combined ablation result for benign and malignant images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine (without augmentation).

Algorithm	HD	JM	DM
Criteria 1	9.31 ± 4.36	75.37 ± 5.36	86.57 ± 5.41
Criteria 2	9.69 ± 4.69	76.27 ± 5.83	87.23 ± 5.54
Criteria 3	9.43 ± 4.21	77.48 ± 5.43	88.43 ± 5.31
Proposed method	8.60 ± 4.15	78.65 ± 5.09	89.67 ± 5.52

3.2.1.2. Noise removal. We use the conventional bilateral filter to reduce noise (see Fig. 5). Basic mathematical computation and further explanation about the conventional bilateral filter are available in [40]. A flow of the pre-processing phase is seen in Fig. 6.

3.2.2. Segmentation stage

Deep learning methods have achieved higher accuracies than most traditional methods in tasks such as image classification, semantic segmentation, object detection, and simultaneous localization mapping [41–44]. However, that is not to

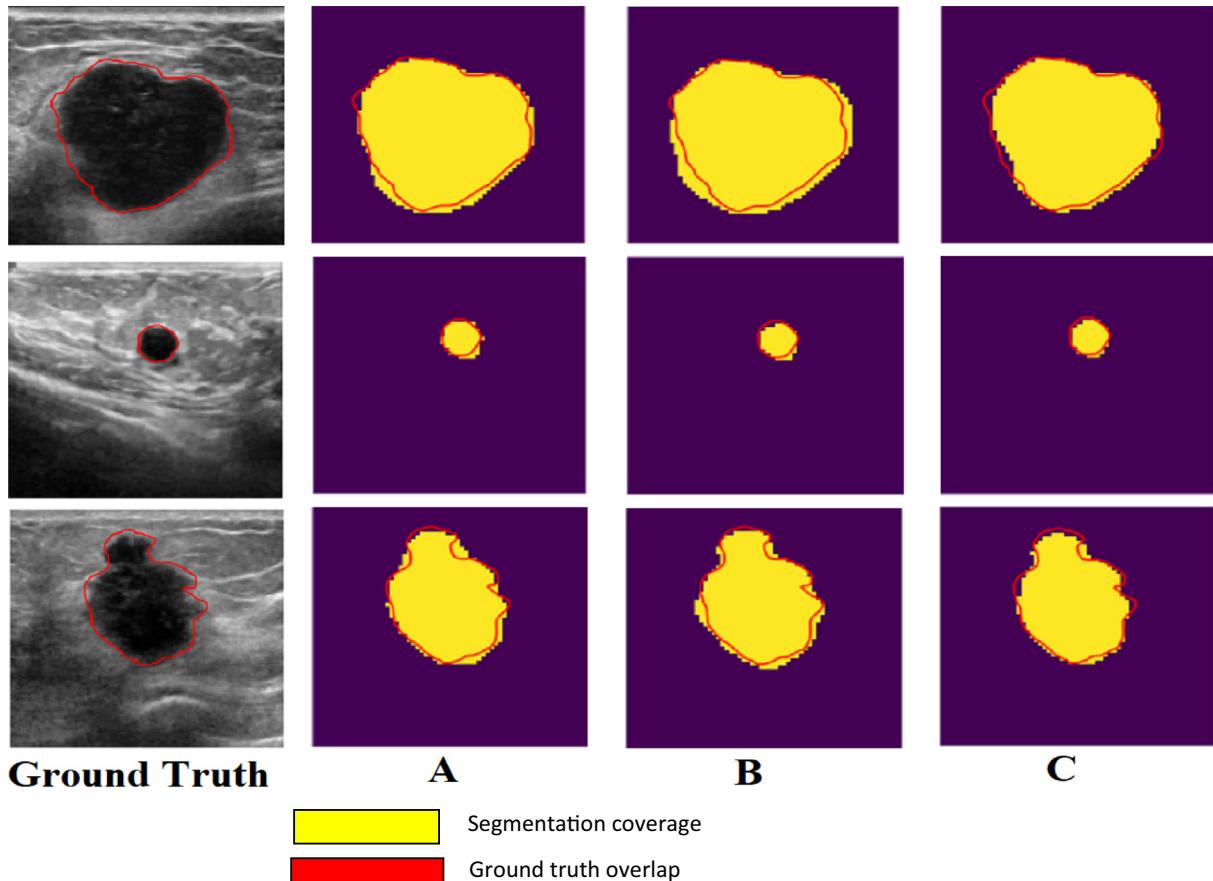


Fig. 11 – Results of augmentation and ablation (Philips iU22 machine): (A) Results of Criteria 1; (B) Results of Criteria 2; (C) Results of the proposed method.

say that traditional methods have become obsolete. We propose an end-to-end semantic segmentation method for BUS images. The proposed framework used the VE block and concatenated convolution. The added convolutions are the input to the ReLU layer.

3.2.2.1. UNet architecture. In medical semantic segmentation, the UNet architecture is used for tumor identification. UNet is an end-to-end architecture that minimizes cost in the encoding part and constructs an image mask at the decoding part (the diagram of UNet is available in Ref [26]). The UNet architecture consists of two parts: encoding and decoding. The encoding part has a convolution layer followed by a ReLU activation method and a pooling layer. The max-pooling layer increases the robustness of the input image and can help to reduce over-fitting in semantic segmentation. The decoding part collects input from the last layer of the encoding part and expands the feature map. Between the encoding and decoding layers, there is a skip connection that preserves the full content of the input image. Overall, the UNet segmentation is used for biomedical segmentation tasks and can preserve the full context of the input image [27].

3.2.2.2. Proposed VEU-Net. The proposed segmentation (VEU-Net) method involves the encoding and decoding frameworks. The model has 4 encoding and 4 decoding blocks. A skip connection that transmits information from the encoder

to the decoder is adopted. The skip connection helps to reinforce the gradient that tends to vanish when the feature map moves from the encoding to the decoding network [45]. VEU-Net is an end-to-end segmentation method that uses the VE block and the concatenated convolutions for segmentation. The VE block consists of a concatenated max and average pooling with batch normalization. The input of the convolution layer is added to the input of the deconvolution layer. The added convolutions are the input of the ReLu layer. The diagram of the VEU-Net segmentation is depicted in Fig. 7.

The pooling technique is used to transform important features into a more usable form. In CNN, max and average pooling are used to reduce computational complexity, variance, and to extract low features [46]. Analysts believe that no particular pooling method is better than the other. The particular data determine the choice of a pooling method. For example, the addition of max-pooling layer to several popular architecture (such as: AlexNet [47], [GoogleNet [48], ResNet [49]]) has shown its effectiveness for building deep neural networks. The average pooling layer was added to reference [50] producing an excellent result. Average pooling calculates the arithmetic mean of an element to smoothen the features. However, features associated with shapes may be ignored with average pooling. The mathematical formula of average pooling is given by:

$$Vw_{nm} = \frac{1}{|U_{mn}|} \sum_{(pq) \in U_{mn}} Zw_{pq} \quad (1)$$

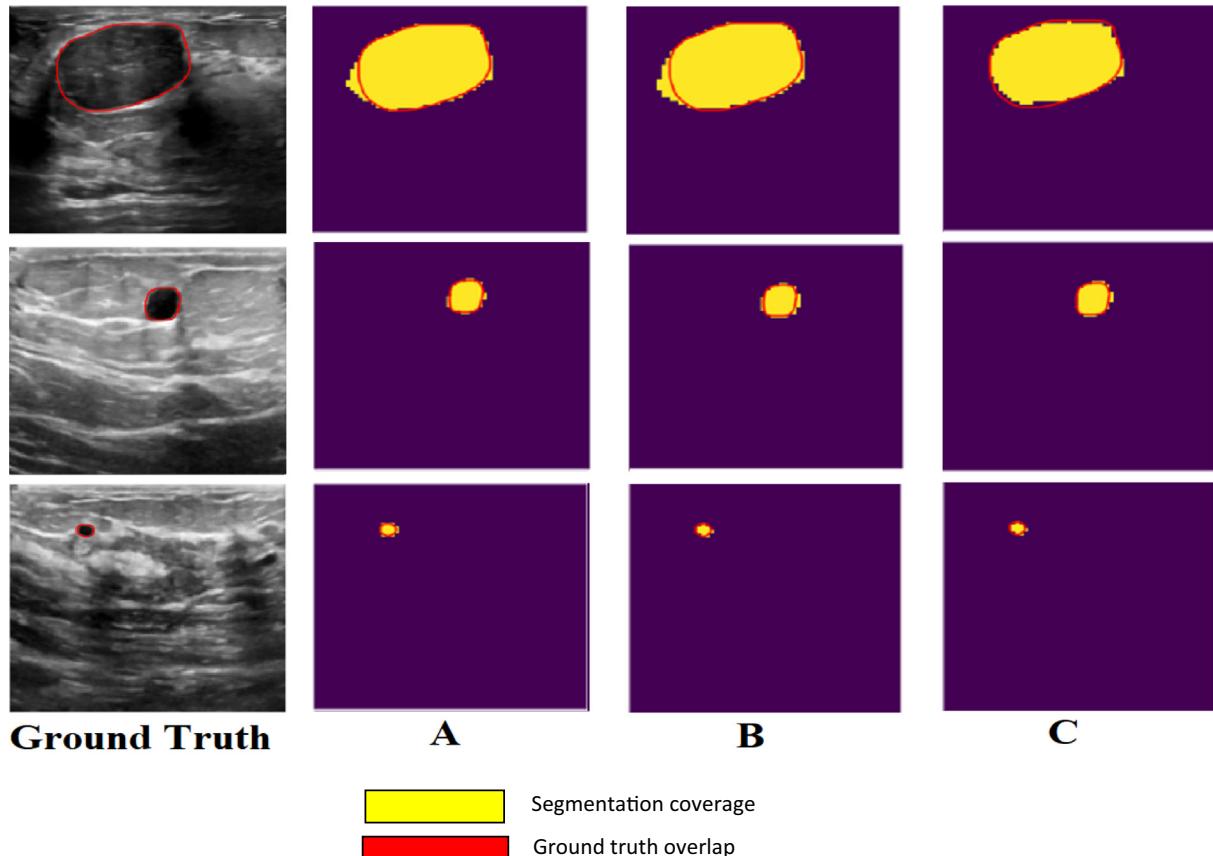


Fig. 12 – Results of augmentation and ablation (LOGIQ E9 and LOGIQ E9 Agile): (A) Results of Criteria 1; (B) Results of Criteria 2; (C) Results of the proposed method.

where U_{mn} is the pooling region, $|U_{mn}|$ is the size of the pooling region, Vw_{nm} is the output of the pooling operator, and Zw_{pq} is the element of the pooling region.

The max-pooling calculates maximum values in the feature map. When performing a max-pooling operation, the largest element in each pooling region is selected. The mathematical formula of max-pooling is given by:

$$Kw_{nm} = \text{Max}_{(pq) \in U_{mn}} Zw_{pq} \quad (2)$$

The max and average pooling can perform well on the same datasets; however, it is not clear which pooling method can perform better on a new project. Bearing this in mind, we have adopted the max and average pooling for the proposed method. The reason for concatenating the max and average pooling is to combine the advantages of both pooling methods for segmentation. Research has shown that the combination of pooling techniques provides effective results (see [46,51,52]). Leveraging on these experiments, we proposed a VE block that concatenates pooling methods. The VE block consists of max pooling, average pooling, and batch normalization (see Fig. 8). A contraction procedure was performed on the feature map in the encoding framework before each

convolution layer. In the decoding framework, two convolutions were concatenated to produce the input of the next block.

Overall, the VEU-Net is a semantic segmentation method that uses the VE Block and a concatenated convolution. The flow of the proposed VEU-Net is described below.

Algorithm 1: Proposed VEU-Net

```

Input: Preprocessed Image  $I_{pre}$ 
1:  $I_{con} \leftarrow$  Convolution of ( $I_{pre}$ )
2:  $I_{dro} \leftarrow$  Dropout of ( $I_{con}$ )
3:  $I_{max} \leftarrow$  Max pooling ( $I_{dro}$ )
4:  $I_{avr} \leftarrow$  Average pooling ( $I_{dro}$ )
5:  $I_{conca} \leftarrow$  Perform concatenation on ( $I_{avr}$ ) and ( $I_{max}$ )
6:  $I_{bat} \leftarrow$  Batch Normalization ( $I_{conca}$ )
7. Repeat this for every block in the encoder framework
8. Skip connection  $I_{ski} \leftarrow$ 
9:  $I_{con2} \leftarrow$  Convolution of ( $I_{up}$ )
10:  $(I_{con3} \leftarrow I_{con2} \leftarrow$  Convolution of ( $I_{con}$ )
11:  $I_{conca2} \leftarrow$  Perform concatenation
12:  $I_{con3} \leftarrow$  ReLU of ( $I_{conca2}$ )
Output: Segmented mask ( $I_{mas}$ )

```

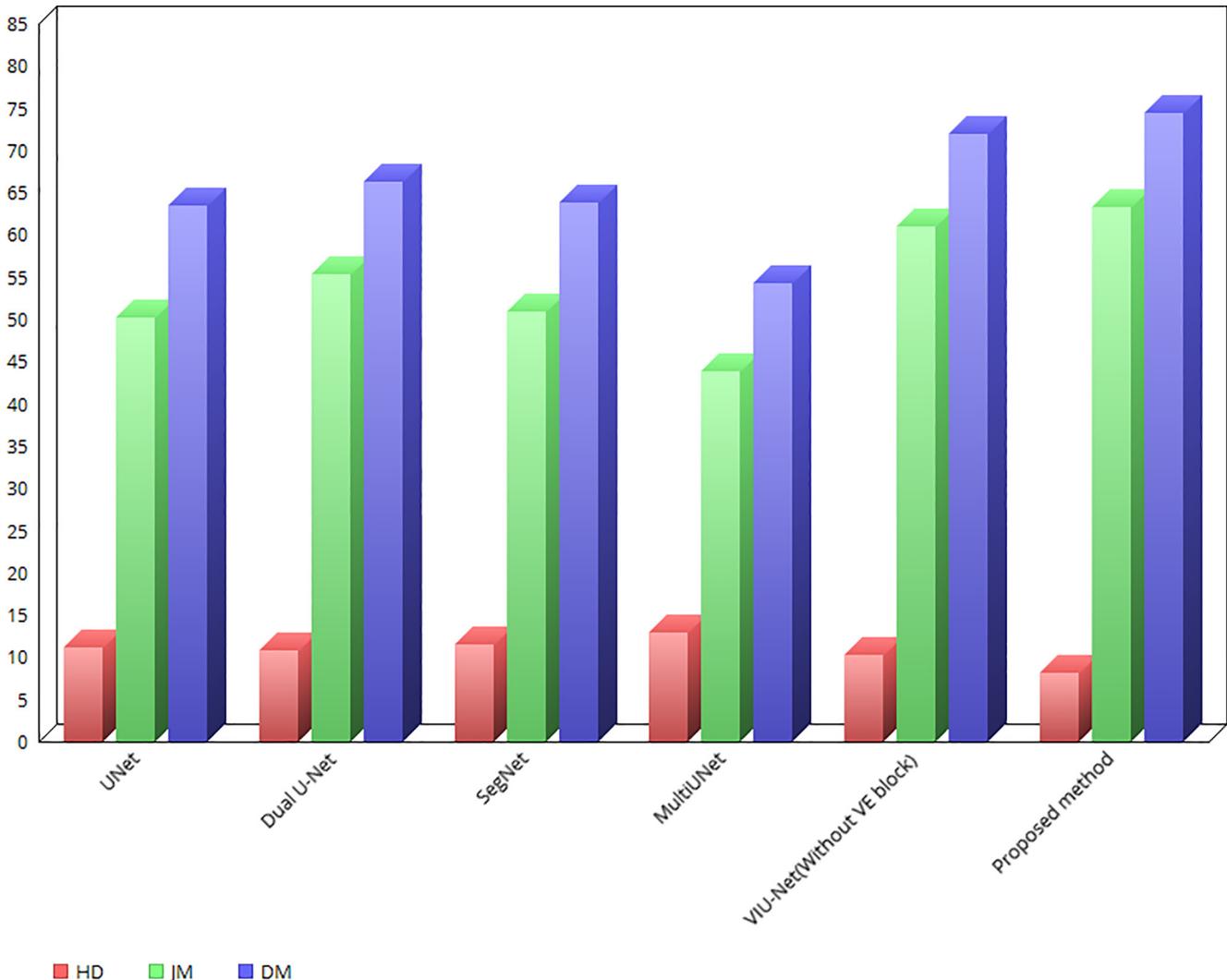


Fig. 13 – Results of benign images obtained with Philips iU22 machine.

4. Experiments and results analysis

4.1. Experiments setup

The performance of the methods was evaluated with metrics such as Jaccard Measure (JM), Dice Measure (DM), and the Hausdorff Distance (HD). The equations of the evaluation metrics are given below:

A. **Jaccard Measure (JM).** JM is the ratio of intersection against the union of the output and actual segments [53]. The JM can be a value between 0 and 100%, with 0 indicating no overlap and 100 indicating complete overlap. A similarity of 100% means that the segmented image is a perfect match with the ground truth. JM is computed by:

$$JSC(B_U, B_{GT}) = \frac{|B_U \cap B_{GT}|}{|B_U \cup B_{GT}|} \quad (3)$$

where B_U , and B_{GT} stands for segmented image and ground truth, respectively.

B. **Dice Measure (DM).** DM is the measure of overlap between two sets of segmented results. The higher the DM numbers the better the image. DM is computed by:

$$DSC(B_U, B_{GT}) = 2X \frac{|B_U \cap B_{GT}|}{|B_U| + |B_{GT}|} \quad (4)$$

C. **Hausdorff Distance (HD).** HD is the measure of how far the segmented image and the ground truth are from each other [54]. The lower the HD number the better the image distance, however, when the HD value is close to zero a better image is indicated. HD is computed by the following:

$$dist_{H1}(X, Y) = \max\{\max_{a \in X} \min_{b \in Y} \|a - b\|, \max_{b \in Y} \min_{a \in X} \|a - b\|\} \quad (5)$$

where $\|$ is the Euclidian distance, X the ground truth contour, and Y the resulting contour.

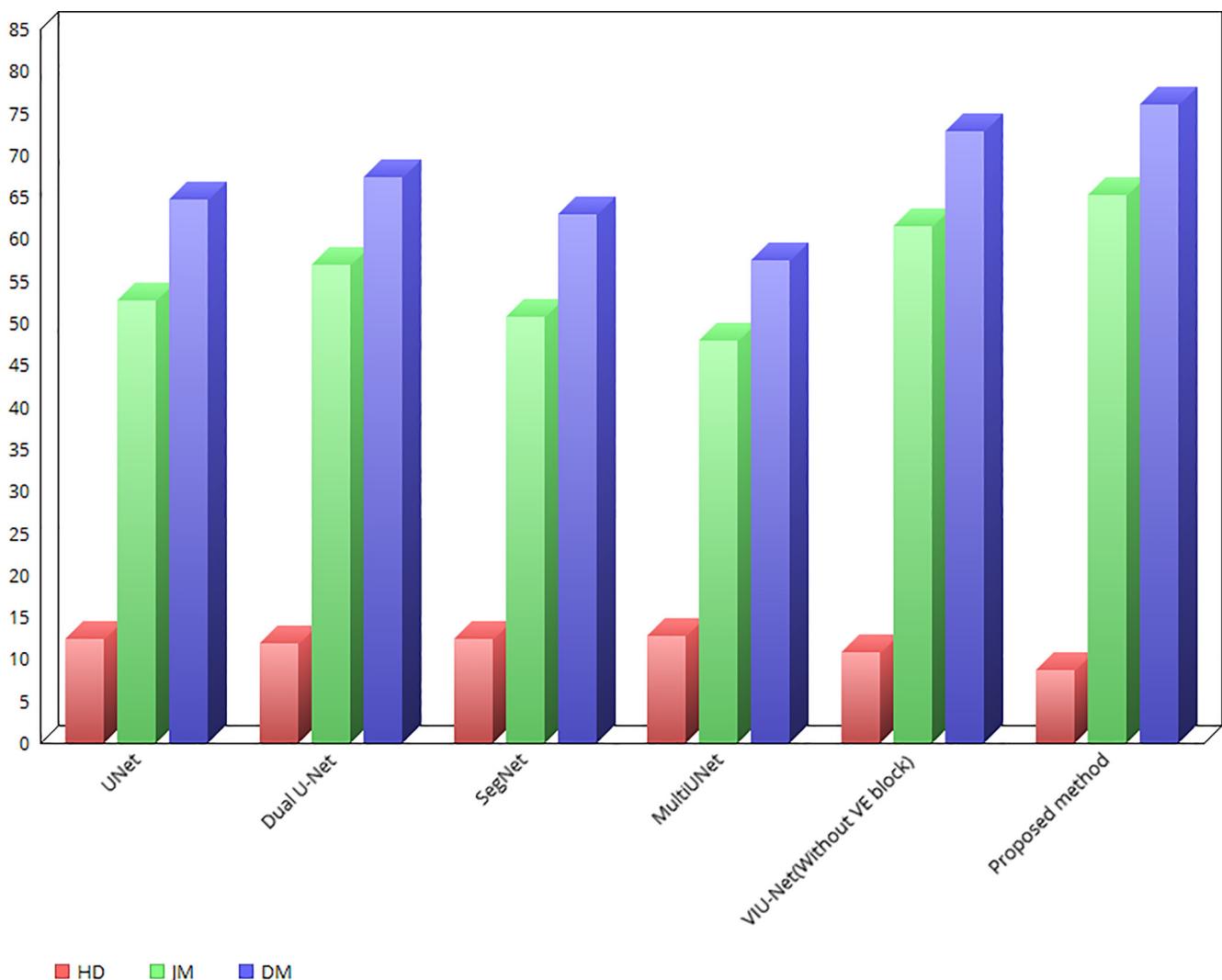


Fig. 14 – Results of malignant images obtained with Philips iU22 machine.

4.2. Model training

The models were trained and tested on NVIDIA GeForce GTX 1080Ti GPU. Random subsets of images were used as the validation set during training. The first dataset has 264 images for training while 40 images were used for testing. Similarly, the second dataset contains 592 training images and 105 testing images. In the second dataset, normal BUS images were not used in the experiment, hence a total of 697 images were used in the second dataset. Overall, 15% of total images were used for testing and 85% for training. K-fold validation was used to estimate the unseen data. The dataset was shuffled randomly with a split of 5 fold. Each unique group was used as the test data and the remaining as a training dataset. We fit the model on the training set before evaluating the test set. The cross-validation has 5 error estimates that were averaged to obtain a more robust estimate of the test error (see details in [60]). It took approximately 40 min or less to complete model training. The proposed system was implemented with TensorFlow (version 1.15.0) [55]. The model was trained with the Adam optimizer (with a learning rate of 1×10^{-4}) [56], 96 epochs, a batch size of 6, and 144 iterations. A learning

rate of 0.0001 and a callback mechanism were used in the training process. All experiments in this paper adopt the same training strategy. The output activation uses the sigmoid function given by:

$$\beta_{(y)} = \frac{1}{1 - e^{-y}} \quad (6)$$

Loss function uses the binary cross-entropy method given by:

$$LF_{BCE}(R, R') = \frac{1}{P} \sum_{i,j} -(R_{ij} \log(R'_{ij}) + (1 - R_{ij}) \log(1 - R'_{ij})) \quad (7)$$

where P represents the total number of pixels, R_{ij} are the predicted probabilities of BUS, and R'_{ij} is ground truth.

4.3. Data augmentation

We augmented our data by flipping the images vertically and horizontally. The reason for this is to avoid overfitting during training. The random rotation procedure was used to adjust the orientations of the images. A wrapper for computer vision was generated and used to fit and predict the model.

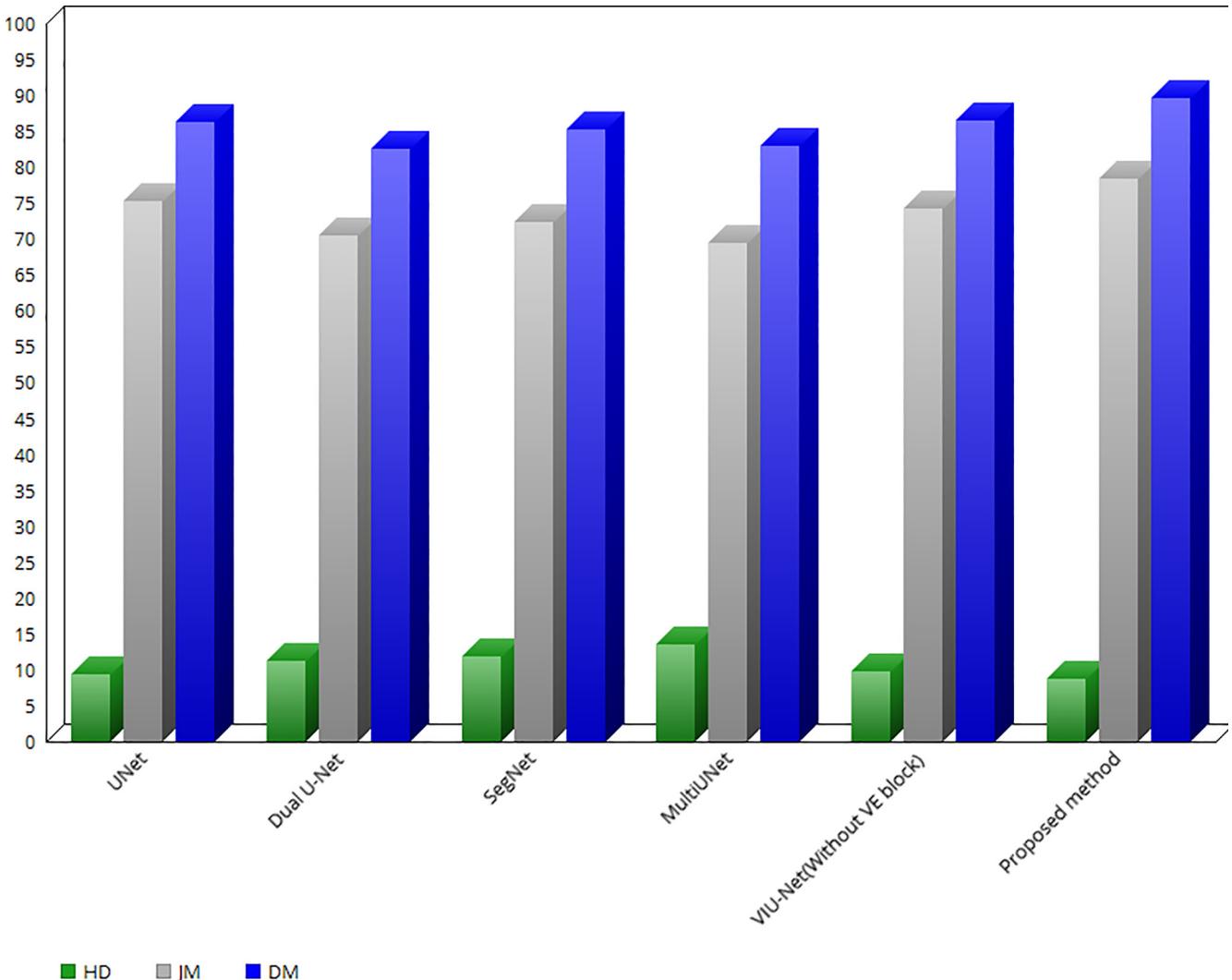


Fig. 15 – Results of benign images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine.

The wrapper provides the augments with the test data for horizontal and vertical flips.

4.4. Results of different methods

The proposed method was benchmarked against four different methods (UNet [26], Dual U-Net [57], SegNet [58], MultiU-Net [59]). The segmentation results section is divided into two parts: (1) results obtained from Philips iU22 ultrasound machine, and (2) results obtained from LOGIQ E9 and LOGIQ E9 Agile ultrasound machine. Segmentation results of these methods are depicted in Figs. 9 and 10. Furthermore, Tables 2 and 3 represent quantitative results from the first dataset, while Tables 4 and 5 represent results from the second dataset.

In Figs. 9 and 10, the first column is the ground truth image, the second to fourth columns are the results of notable CNN methods, and the last column is the result of the proposed method. In the first part of the experiment (Fig. 9, images from the first dataset), the results of competing CNN methods were very poor. They could not segment tumor regions correctly (for example, the first row has only

three methods that could segment tumors). However, the proposed method was closer to the ground truth. In the second part (Fig. 10, images from the second dataset), most CNN methods performed well, however, the proposed method performed better. The introduction of the concatenated pooling mechanism made the proposed method capable of locating tumor regions in BUS. Quantitative results showing values of DM, HD, and JM are available in Tables 2–5.

4.5. Ablation and augmentation results

The ablation experiment was performed to assess the effectiveness of the VE block, concatenated convolutions, and the noise removal/image enhancement methods. The ablation experiment involves 3 criteria. Since the experiment is proposed based on the UNet method, we use the UNet model as the first criteria (criteria 1). The second criterion is based on criteria 1 with a concatenated convolution at the decoding layer (criteria 2). The final criterion is criteria 2 with the VE block, but without the image removal/enhancement (criteria 3) (see Table 6 for details).

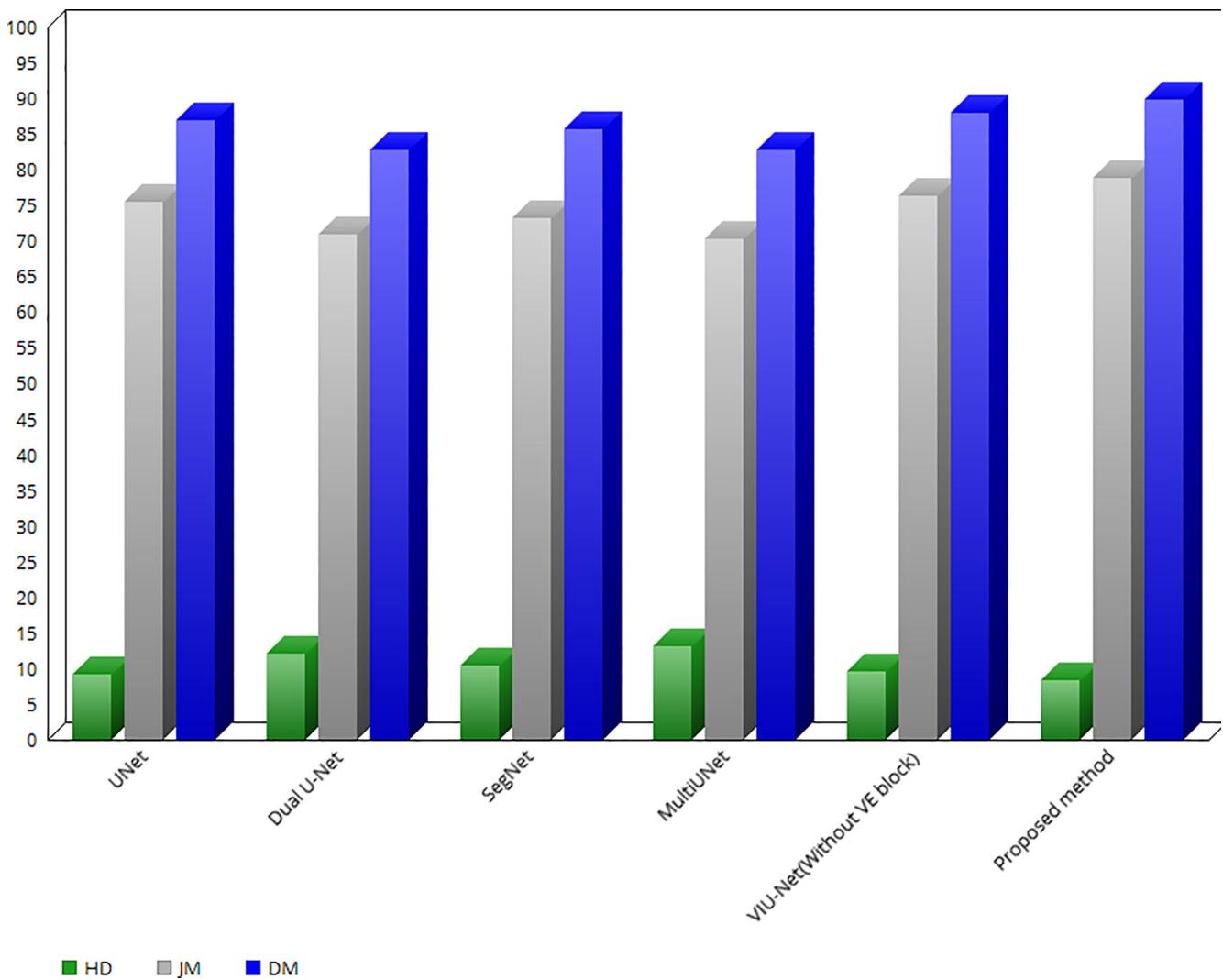


Fig. 16 – Results of malignant images obtained with LOGIQ E9 and LOGIQ E9 Agile ultrasound machine.

Quantitative results of ablation experiments with augmentation and without augmentation are shown in Tables 7–10. The proposed method produced better results than other methods. The visual experiments for ablation and augmentation experiments are depicted in Figs. 11 and 12.

Compared with other criteria, the HD, JM, DM of the proposed method in the ablation experiment are 7.81, 80.07, 90.82 for the first dataset, and 7.71, 79.49, 90.67 for the second dataset respectively. The results suggest that the proposed method is effective for segmenting BUS images. Results of ablation experiments with augmentation produced significant improvement when compared with results without augmentation (compare Tables 7, 9 and 8, 10). For all ablation experiments, Criteria 3 and the proposed method produced better results than other criteria. Although other criteria produce good results, they are not satisfactory. Overall, the VE block, concatenated convolutions, and noise removal/image enhancement methods were significant for improving segmentation performance (see Figs. 11 and 12).

5. Discussion

This paper presents a semantic segmentation procedure to segment BUS images. We proposed a BUS segmentation method using the VE block and the concatenated convolutions. To evaluate the performance of the proposed method, the HD, JM and DM were adopted. As shown in Tables 2–5, the proposed method achieves higher HD, JM, and DM in quantitative analysis. For visual inspection, the proposed method effectively segment BUS images (see Figs. 9–12). The VE block, concatenated convolutions, and noise removal/image enhancement methods were instrumental for effective segmentation of BUS in the ablation experiment. Augmentation was conducted on the images, and it has proven to be instrumental for effective segmentation when compared to experiments without augmentation (compare Figs. 9, 10, and Figs. 11 and 12). Results indicate that the proposed method performs better than the competing methods. Visual inspection of segmentation methods is available in Figs. 9 and 10. Results by visual inspection indicate that the proposed method was better than other methods.

The proposed method and others CNN methods were investigated on two datasets. The results showed that the proposed method is robust producing the highest measures for JM and DM, and the lowest value for HD (Figs. 13–16).

6. Conclusion

Segmentation of tumors in BUS images is important to clinicians because it can assist doctors to diagnose and make on-time decisions. We present a semantic segmentation procedure for BUS images using the VEU-Net. First, the CLAHE method was used to enhance the image and the BF remove noise. Then semantic segmentation was used to segment tumors. Our method used the VE block and a concatenated convolution method for segmentation. The proposed method consists of an encoder and decoder path. Two datasets of real breast ultrasound images were used in this paper. The HD, JM, and DM values were quantitatively

measured from each experiment. Four state-of-the-arts methods were used to compete with the proposed method. Results indicate that the VEU-Net produced better segmentation when compared with other competing CNN methods. The proposed method is simple to understand and straightforward.

In future research, advanced deep learning methods to segment BUS images will be developed. This model is expected to be useful in clinical medicine such as ultrasound surgery navigation and diagnosis.

CRediT authorship contribution statement

Ademola Enitan Ilesanmi: Investigation, Methodology, Software, Writing - original draft, Conceptualization, Formal analysis, Visualization, Resources, Validation. **Utairat Chaumrattanakul:** Data curation, Formal analysis. **Stanislav S. Makhanov:** Funding acquisition, Project administration, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by Thailand Research Fund grant RSA6280098 and the Center of Excellence in Biomedical Engineering of Thammasat University.

REFERENCES

- [1] Zhang Zi, Rao R, Mango VL, Wilson-Gardner P, Vempalle S, Ojutiku O. Presentation of self-detected breast mass in minority women with limited access to care: Can self-examination assist in early cancer detection? *Clin Imaging* 2021;70:89–92.
- [2] Nishikawa RM, Gur D. CADe for early detection of breast cancer—current status and why we need to continue to explore new approaches. *Acad Radiol* 2014;21:1320–1.
- [3] Kriti, Virmani J, Agarwal R. Effect of despeckle filtering on classification of breast tumors using ultrasound images. *Biocybernetics Biomed Eng* 2019;39(2):536–60.
- [4] Ramadan H, Lachqar C, Tairi H. Saliency-guided automatic detection and segmentation of tumor in breast ultrasound images. *Biomed Signal Process Control* 2020;60:101945.
- [5] Maes F, Vandermeulen D, Suetens P, Marchal G. Automatic image partitioning for generic object segmentation in medical images. *Comput Imaging Vision* 1995.
- [6] Lu R, Marziliano P, Thng CH. Liver tumor volume estimation by semi-automatic segmentation method. In: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. p. 3296–9.
- [7] Zhao B, Yankelevitz D. Two-dimensional multi-criterion segmentation of pulmonary nodules on helical CT images. *Med Phys* 1999;26(6):4678–89.
- [8] Xian M, Zhang Y, Cheng HD. Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. *Pattern Recogn* 2015;48:485–97.

- [9] Sapate S, Talbar S, Mahajan A, Sable N, Desai S, Thakur M. Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms. *Biocybernetics Biomed Eng* 2020;40(1):290–305.
- [10] Atkins MS, Mackiewich BT. Fully automatic segmentation of the brain in MRI. *IEEE Trans Med Imaging* 1998;17(1):98–107.
- [11] Campadelli P, Casiraghi E, Pratissoli S. Fully automatic segmentation of abdominal organs from CT images using fast marching methods. In: 2008 21st IEEE International Symposium on Computer-Based Medical Systems, Jyvaskyla, Finland. p. 554–9. DOI: 10.1109/CBMS.2008.9.
- [12] Xian M, Zhang Y, Cheng HD, Xu F, Zhang B, Ding J. Automatic breast ultrasound image segmentation: A survey. *Pattern Recogn* 2018;79:340–55.
- [13] Ilesanmi AE, Idowu OP, Makhanov SS. Multiscale superpixel method for segmentation of breast ultrasound. *Comput Biol Med* 2020;125:103879.
- [14] Kucybala I, Tabor Z, Ciuk S, Chrzan R, Urbanik A, Wojciechowski W. A fast graph-based algorithm for automated segmentation of subcutaneous and visceral adipose tissue in 3D abdominal computed tomography images. *Biocybernetics Biomed Eng* 2020;40(2):729–39.
- [15] Gamarrá M, Zurek E, Escalante HJ, Hurtado L, San-Juan-Vergara H. Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. *Biomed Signal Process Control* 2019;53:101575.
- [16] Yu H, Yang Z, Tan L, Wang Y, Sun W, Sun M, et al. Methods and datasets on semantic segmentation: A review. *Neurocomputing* 2018;304:82–103.
- [17] Ilesanmi AE, Chaumrattanakul U, Makhanov SS. Methods for the segmentation and classification of breast ultrasound images: a review. *J Ultrasound* 2021. <https://doi.org/10.1007/s40477-020-00557-5>.
- [18] Gómez-Flores W, Coelho de Albuquerque Pereira W. A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound. *Comput Biol Med* 2020;126:104036.
- [19] Yang J, Faraji M, Basu A. Robust segmentation of arterial walls in intravascular ultrasound images using Dual Path U-Net. *Ultrasonics* 2019;96:24–33.
- [20] Li C, Tan Y, Chen W, Luo X, He Y, Gao Y, et al. ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. *Computers Graphics* 2020;90:11–20.
- [21] Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S. Attention guided U-Net for accurate iris segmentation. *J Vis Commun Image Represent* 2018;56:296–304.
- [22] Huang Q, Huang Y, Luo Y, Yuan F, Li X. Segmentation of breast ultrasound image with semantic classification of superpixels. *Med Image Anal* 2020;61 101657.
- [23] Kumar V, Webb JM, Gregory A, Denis M, Meixner DD, Bayat M, et al. Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PLoS ONE* 2018;13(5):e0195816.
- [24] Hiramatsu Y, Muramatsu C, Kobayashi H, Hara T, Fujita H. Automated detection of masses on whole breast volume ultrasound scanner: false positive reduction using deep convolutional neural network. In: Proceedings of the SPIE Medical Imaging; 2017 Feb 11–16; Orlando, FL, USA. Bellingham: SPIE; 2017.
- [25] Reza SMS, Bradley D, Aiosa N, Castro M, Lee JH, Lee B-Y, et al. Deep learning for automated liver segmentation to aid in the study of infectious diseases in nonhuman primates. *Acad Radiol* 2020. <https://doi.org/10.1016/j.acra.2020.08.023>.
- [26] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
- [27] Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE; 2016. p. 565–71.
- [28] T-Y. Lin, P. Doll_ar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017:2117–2125.
- [29] Vakanski A, Xian M, Freer PE. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. *Ultrasound Med Biol* 2020;46(10):2819–33.
- [30] Byra M, Jarosik P, Szubert A, Galperin M, Ojeda-Fournier H, Olson L, et al. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomed Signal Process Control* 2020;61 102027.
- [31] Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial intelligence in medicine: Where are we now? *Acad Radiol* 2020;27(1):62–70.
- [32] von Landesberger T, Bremm S, Kirschner M, Wesarg S, Kuijper A. Visual Analytics for model-based medical image segmentation: Opportunities and challenges. *Expert Syst Appl* 2013;40(12):4934–43.
- [33] Rodtook A, Kirimasthong K, Lohitvisate W, Makhanov SS. Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities. *Pattern Recogn* 2018;79:172–82.
- [34] Al-Dhabyan W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. *Data in brief* 2020;28 104863.
- [35] Wang H, Wang S, Qin Z, Zhang Y, Li R, Xia Y. Triple attention learning for classification of 14 thoracic diseases using chest radiography. *Med Image Anal* 2021;67:1094–109.
- [36] Wang S, Cheng H, Ying L, Xiao T, Ke Z, Zheng H, et al. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn Reson Imaging* 2020;68:136–47.
- [37] Hua W, Xiao T, Jiang X, Liu Z, Wang M, Zheng H, et al. Lymph-vascular space invasion prediction in cervical cancer: exploring radiomics and deep learning multilevel features of tumor and peritumor tissue on multiparametric MRI. *Biomed Signal Process Control* 2020;58:101869.
- [38] Ali Y, Janabi-Sharifi F, Beheshti S. Echocardiographic image segmentation using deep Res-U network. *Biomed Signal Process Control* 2021;64:102248.
- [39] Sonali, Sahu S, Singh AK, Ghrera SP, Elhoseny M. An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Opt Laser Technol* 2019;110:87–98.
- [40] Balocco S, Gatta C, Pujol O, Mauri J, Radeva P. SRBF: Speckle reducing bilateral filtering. *Ultrasound Med Biol* 2010;36:1353–63.
- [41] M.F. Rachmadi, M.D.C. Valdés-Hernández, M.L.F. Agan, T. Komura, Deep Learning vs. Conventional Machine Learning: Pilot Study of WMH Segmentation in Brain MRI with Absence or Mild Vascular Pathology, *Journal of Imaging*, 3 (2017), 66.
- [42] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G.V. Hernandez, L. Krpalkova et al., Deep learning vs. Traditional Computer Vision. In *Proceedings of the Science and Information Conference*, Tokyo, Japan; Springer: Berlin/Heidelberg, Germany, 2019; 128–144.
- [43] Andy Lee. Comparing deep neural networks and traditional vision algorithms in mobile robotics. Swarthmore College, 2015.
- [44] Tang KJW, Ke En Ang C, Theodoros C, Rajinikanth V, Acharya UR, Cheong KH, Artificial Intelligence and machine learning in Emergency Medicine, *Biocybernetics Biomed Eng*, 2020.

- [45] Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. *Workshop on Deep Learning Med Image Anal*, 2016. <https://arxiv.org/abs/1608.04117v2>.
- [46] Yu D, Wang H, Chen P, Wei Z, Mixed pooling for convolutional neural networks, in: *Rough Sets Knowl. Technol.*, Cham, Switzerland: Springer, pp. 364–375, 2014.
- [47] Krizhevsky A, Sutskever I, Hinton G, Imagenet classification with deep convolutional neural networks, in: *Proceedings of the NIPS*, 2012, pp. 1097–1105, DOI: 10.1145/3065386.
- [48] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., Going deeper with convolutions, in: *Proceedings of the CVPR*, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
- [49] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the CVPR*. p. 770–8. doi: 10.1109/CVPR.2016.90.
- [50] Arnett K, Kavukcuoglu K, Ranzato M, LeCun Y. What is the best multi-stage architecture for object recognition? In: *ICCV*, pp. 2146–2153 (2009).
- [51] Tong Z, Tanaka G. Hybrid pooling for enhancement of generalization ability in deep convolutional neural networks. *Neurocomputing* 2019;333:76–85.
- [52] Chang J, Zhang L, Gu N, Zhang X, Ye M, Yin R, et al. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *J Vis Commun Image Represent* 2019;58:316–22.
- [53] Ilesanmi AE, Idowu OP, Chaumrattanakul U, Makhanov SS. Multiscale hybrid algorithm for pre-processing of ultrasound images. *Biomed Signal Process Control* 2021;66:102396.
- [54] Karunanayake N, Aimmanee P, Lohitvisate W, Makhanov SS. Particle method for segmentation of breast tumors in ultrasound images. *Math Comput Simul* 2020;170:257–84.
- [55] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software Available from tensorflow.org, 2015.<<https://www.tensorflow.org/>>.
- [56] Kingma DP, Ba J, Adam: A Method for Stochastic Optimization. Available from: arXiv preprint arXiv:1412.6980.
- [57] Li X, Wang Y, Tang Q, Fan Z, Yu J. Dual U-Net for the segmentation of overlapping glioma nuclei. *IEEE Access* 2019;7:84040–52.
- [58] Chen T, Cai Z, Zhao X, Chen C, Liang X, Zou T, et al. Pavement crack detection and recognition using the architecture of SegNet. *J Ind Inf Integration* 2020;18:100144.
- [59] Ibtehaz N, Rahman MS, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121(2020), 74–87.
- [60] Wong T-T. Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets. *Pattern Recogn* 2017;65:97–107.