HHS Public Access

Author manuscript

IEEE Access. Author manuscript; available in PMC 2024 November 15.

Published in final edited form as:

IEEE Access. 2024 ; 12: 60541–60555. doi:10.1109/access.2024.3394808.

DMAeEDNet: Dense Multiplicative Attention Enhanced Encoder Decoder Network for Ultrasound-Based Automated Breast Lesion Segmentation

MANALI SAINI[1] **[(Member, IEEE)]**, **HUMAYRA AFRIN**[2], **SETAYESH SOTOUDEHNIA**[1], **MOSTAFA FATEMI**[2] **[(Life Fellow, IEEE)]**, **AZRA ALIZAD**[1] **[(Senior Member, IEEE)]**

[1]Department of Radiology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA

[2]Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA

Abstract

Automated and precise segmentation of breast lesions can facilitate early diagnosis of breast cancer. Recent research studies employ deep learning for automatic segmentation of breast lesions using ultrasound imaging. Numerous studies introduce somewhat complex modifications to the well adapted segmentation network, U-Net for improved segmentation, however, at the expense of increased computational time. Towards this aspect, this study presents a low complex deep learning network, i.e., dense multiplicative attention enhanced encoder decoder network, for effective breast lesion segmentation in the ultrasound images. For the first time in this context, two dense multiplicative attention components are utilized in the encoding layer and the output layer of an encoder-decoder network with depthwise separable convolutions, to selectively enhance the relevant features. A rigorous performance evaluation using two public datasets demonstrates that the proposed network achieves dice coefficients of 0.83 and 0.86 respectively with an average segmentation latency of 19*ms*. Further, a noise robustness study using an in-clinic recorded dataset without pre-processing indicates that the proposed network achieves dice coefficient of 0.72. Exhaustive comparison with some commonly used networks indicate its adeptness with low time and computational complexity demonstrating feasibility in real time.

INDEX TERMS

Ultrasound; breast lesion segmentation; deep learning; U-Net; convolution neural network

Corresponding author: Azra Alizad (Alizad.Azra@mayo.edu).
Humayra Afrin and Setayesh Sotoudehnia contributed equally to this work.

DISCLOSURE OF CONFLICT OF INTEREST

The authors do not have any potential financial interest related to the technology referenced in this paper.

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

I. INTRODUCTION

Breast cancer is the most common cancer among women, in the United States [1], [2]. Timely diagnosis of malignancy in the breast lesions can help prevent progression to an advanced stage of cancer and subsequently reduce the mortality rate [3]. Self-examination and mammography are common practices for screening of breast cancer [4], [5]. The use of Ultrasound (US) in conjunction with mammography increases the diagnosis accuracy, particularly in patients with dense breasts [6], [7], [8]. Furthermore, US is preferred over other imaging modalities due to its non-invasiveness, cost-effectiveness and real-time capability [9], [10], [11], [12]. Locating and interpreting of a breast mass is a time-consuming, operator dependent and difficult task that involves expert radiologist [13]. In past few decades, attempts have been made to develop computer-aided systems and machine learning approaches to make the complex medical tasks easier and help time efficiency in clinical practice [14]. To that end, deep learning that has potential use in medical applications and can offer expertise in resource limited-settings [15], [16]. Therefore, automated segmentation of breast lesions using deep learning can address the current challenges by reducing the time complexity and improving decision making in clinical practice [17], [18], [19].

Researchers have developed a variety of techniques including, conventional and machine learning algorithms for the automatic segmentation of breast lesions [20], [21], [22]. However, the segmentation performance of these algorithms is poor due to the inability of high-level feature extraction [23]. To this end, researchers have recently explored deep learning (DL) algorithms with the capability of automated high-level feature extraction from the raw images [24], [25], [26], [27], [28], [29]. These studies utilize different types of DL algorithms, such as, convolutional neural networks (CNNs), attention-based networks, encoder-decoder-based networks, etc. to aid cancer diagnosis, which can further reduce the workload of the physicians [27], [28].

A. LITERARY STUDIES

One of the most extensively used DL architectures for breast lesion segmentation is U-Net, which possesses an encoder-decoder-based structure, with several convolution layers [30]. Development of U-Net variants for breast ultrasound image segmentation has been a primary research focus in the recent studies [31], [32], [33], [34]. Significant improvements have been reported in the segmentation performances with respect to the conventional U-Net. For example, Byra et al. used selective kernels to adjust receptive fields by using attention mechanisms in U-Net [32], however, the validation on a limited dataset is insufficient to demonstrate the effectiveness of segmentation performance. Kumar et al. proposed a multi-U-Net algorithm without using placement of an initial seed to automatically segment the breast masses in real-time [33]. The authors demonstrated the superiority of the proposed algorithm w.r.t. the seeded U-Net algorithm, however, the performance could be improved with a larger dataset [33]. Guo et al. established an expanded U-Net with addition of dropout layers for preventing over-fitting, while preserving the surface and edge attributes at the output [34]. Punn et al. introduced a residual cross-spatial attention-directed inception U-Net (RCA-IUnet) utilizing the basic training parameters to enhance the segmentation

performance of different lesion sizes [35]. However, the trade-off between time complexity and performance limits the model capability in real-time. Yan et al. proposed an attention-based U-Net with hybrid-dilated convolution and introduced a novel loss function to focus on the lesion region for segmentation [36]. Tong et al. implemented a modified version of U-net based on mixed attention loss function to segment the breast US lesions, however, the use of images with ambiguous edges is not clear [37]. Authors in [38] addressed this issue by developing a multi-scale fusion-based version of U-Net to derive the surface and edge features of the image and overcome the substantial variations in the breast lesions. In addition to this, the authors proposed a novel focal loss for handling class imbalance problems in breast lesion segmentation, the results were validated on a public dataset, which is insufficient for performance validation. To handle the data-insufficiency issues, authors in [39] proposed and validated a saliency-guided morphology-aware U-Net on five different datasets. More recently, Chen et al. developed a hybrid adaptive attention-based U-Net by replacing the convolution layers with channel and self-attention modules, for capturing more relevant features with varying receptive fields [31]. However, the false prediction rate could be reduced by further optimizing the network.

Few recent research studies propose transfer learning-based fully convolutional network (FCN)-AlexNet [2] and boundary guided and region aware network with global scale adaptive [9]. These studies demonstrate outstanding segmentation performance on several datasets. However, the complex networks used in these works require a large number of floating point operations, which leads to increased output prediction time. Further, the researchers in [2] reported a significant decrease in the performance when inter-dataset images were used in testing. The number of images in the multiple datasets used in these works seem to be less, which indicates lack of sufficient validation.

B. OBJECTIVE AND KEY CONTRIBUTIONS

The foregoing literary works employ a variety of CNNs along with diverse alterations of the base U-Net architecture, such as, introduction of attention units. The attentions mechanisms adopted in these works help in enhancement of the pertinent features and diminishing the extraneous features which leads to improved segmentation performance. However, it is achieved at the expense of increased computational complexity of the networks. Furthermore, it is noted that the validation results are dataset-dependent, which leads to less robust networks with lack of explain-ability. To address these limitations, this study proposes a novel dense multiplicative attention enhanced encoder decoder network (DMAeEDNet) for ultrasound-based automated breast lesion segmentation. The proposed architecture uses encoding and decoding layers with depth-wise separable convolutions and self-explanatory multiplicative attention-based mechanism for the first time to effectively segment the breast lesions with low time and computational complexity. The novelty of the proposed work is mainly about introducing two dense multiplicative attention components after the depth-wise separable convolutions, where the first component helps in decreasing the number of depth-wise separable convolution operations after two downsampling steps in encoder to preserve both generic and local features. Further, it introduces the dense and multiply layers into the encoder-decoder structure for performing custom attention-based feature extraction which further enhances the segmentation performance, and has not been

explored in existing studies in the context of breast lesion segmentation. The performance of the proposed network is evaluated using two different public datasets along with an exhaustive analysis with respect to cross-dataset images and noisy images taken from an in-clinic recorded dataset. The key contributions of this study are summarized here as follows.

- Introduction of novel dense multiplicative attention mechanisms in the decoding layers of an encoder-decoder network for effective segmentation of breast lesion using ultrasound.
- A comprehensive performance assessment and comparison of the proposed network using two public datasets.
- Test-only performance analysis of the proposed network with respect to inter-dataset images acquired from different ultrasound machines.
- A noise robustness analysis of the proposed network using an in-clinic data without pre-processing.
- Comparison of the time and computational complexity of the proposed network with respect to existing methods for demonstrating its feasibility in real-time segmentation of breast lesions.

The study is presented and organized as follows. Section II describes the datasets used and the proposed automated segmentation network. Section III presents the performance assessment results along with the comparison studies. Section IV discusses the results and future directions of the proposed study. Finally, section V concludes the study.

II. DATASETS

This section provides a description of the two public datasets and an in-clinic dataset utilized in this study.

A. UDIAT DATASET

This dataset was acquired at the UDIAT Diagnostic Centre of the Parc Taul´ı Corporation, Sabadell (Spain) [2]. A total of 163 images were collected from various women using Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer at a frequency of 8.5 MHz. Each image had one or more lesions, and a mean dimension of 760 × 570 pixels. The lesions in 53 images were malignant, and the remaining images had benign lesions. From the malignant images, The lesions were manually segmented by experienced radiologists. This dataset is accessible upon request from the author in [2] for research purposes. The images were processed for noise removal.

B. BREAST ULTRASOUND IMAGES (BUSI) DATASET

The breast ultrasound images data was acquired at Baheya hospital, Cairo, Egypt, using LOGIQ E9 and LOGIQ E9 Agile ultrasound system with 1–5 MHz transducers on ML6–15-D matrix linear probe [40]. It is publicly accessible at https://scholar.cu.edu.eg/?=afahmy/pages/dataset. The data was collected from 600 female participants aged between 25 and

75 years [40]. The total number of images were reduced to 780, after removing the replicated images, and stored in PNG format. The mean pixel dimension of each image was approximately 500 × 500. The images were pre-processed for artifact removal.

In addition to the above mentioned datasets, a few images taken from the clinical dataset recorded at Mayo Clinic, using LOGIQ E9, are utilized for evaluation of the proposed study on inter-machine data.

C. IN-CLINIC RECORDED (ICR) DATASET

This work analyzes the noise-robustness of the proposed network, by utilizing an in-clinic recorded dataset without prior processing for any type of artifact removal. For the *ICR* dataset, the institutional review board approval (IRB:12–003329 and IRB: 19–003028)) was received in compliance with the Health Insurance Portability and Accountability Act along with signed written informed consent with permission for publication from each enrollee prior to this study. The breast ultrasound imaging data was acquired at Mayo Clinic, using the Supersonic Imagine Aixplorer (SSI, Aix-en-Provence, France) system equipped with a 4–15 MHz array transducer. The data was collected from 655 female participants on different days, with each patient having a pre-biopsy visit, and three post-biopsy visits. During each visit, six B-mode ultrasound images were acquired in both longitudinal and transverse orientations. After removing the replicated images, 2066 images were obtained and stored in PNG format. The pixel dimension of each image was 1400 × 1050. The lesions were segmented manually for preparing the masks, i.e., ground truth images for training the proposed segmentation network.

III. PROPOSED METHODOLOGY

This section presents a description of the proposed network, i.e., DMAeEDNet, for the automated segmentation of breast lesions in ultrasound images. Figures 1 and 2 represent the high-level and detailed overviews of the proposed network's structure, consisting of corresponding structural components: Ultrasound image acquisition and pre-processing, DMAeEDNet layers, and predicted segmentation output. As observed from the high-level structural overview in Figure 1, the input ultrasound images are pre-processed prior to be fed to the proposed network layers which are structured into encoding (left) and decoding (right) parts with dense multiplicative attention components (DMAC). The encoding layers consist of the depth-wise separable convolutional layers for feature extraction, followed by batch normalization and max-pooling layers, which extract the more generic features and perform downsampling while going deeper into the network. The output of the encoder is fed to the proposed DMAC component I, which applies a custom attention mechanism by using a dense and a multiply layer to further refine the encoder feature map. The corresponding DMAC-I output is fed to the decoding layers which reconstruct the segmentation mask, by using the depth-wise separable convolutional layers with half channels as compared to decoding layers and skip connections followed by batch normalization and upsampling layers. The decoder parts receive the feature maps from the corresponding encoder parts through skip connections which use concatenation of corresponding features, and incorporate that for reconstruction. The upsampling layers increase the size of down-

sampled encoder feature maps to retain the original input dimensions for reconstructing the segmentation mask. The output of decoder is again refined by DMAC-II which extracts the mask-relevant feature map and eventually leads to the predicted segmented mask after 1 × 1 convolution operation, as shown in the figure. The detailed description of all structural components is provided in this section as follows.

1) ULTRASOUND IMAGE ACQUISITION AND PRE-PROCESSING

The image data can be acquired either while scanning in real-time, using an ultrasound machine, or from pre-recorded and public data repositories. In this work, the B-mode images are acquired from a number of subjects using SSI machine as described previously. Prior to be fed as inputs to the proposed DMAeEDNet, all images are converted to gray-scale and resized to the dimension of 512 × 512 for maintaining the computation efficiency of the proposed network. The pixel intensity values are normalized to the range of 0 – 1 using the empirical method, for preserving the uniformity across all images. The corresponding mathematical representation for each pixel value p at location (i, j) of an image I is obtained as:

$$\frac{|p|_{(i,j)}}{\max |I|} \tag{1}$$

where |.| represents the absolute value, and max represents the maximum intensity value of the image, i.e., 255 in this case. The rationale behind this resizing and normalization is explained as follows. As observed from the dataset description above, there is a diverse variation in the image sizes across all the datasets used in this work. Therefore, resizing to one size and aspect ratio is needed to maintain the uniformity across all datasets like also done in existing image segmentation methods [30], [41]. Further, since the model trains faster on smaller images due to low resource requirements, it is essential to resize them to a lower dimension, specifically for real time tasks. Thereby, all images are resized to 512 × 512 which is neither too small nor too large, and helps to maintain a fair resolution without affecting the overall performance. Additionally, choosing even width and height dimension allows a consistent and smooth tiling of the predicted segmentation mask since the encoder uses 2 × 2 max-pooling operations [41]. Normalization of the image pixels to the range of 0–1 is an empirical method for re-scaling all images and maintaining the consistency among them, since the image pixel values have diverse and dynamic ranges across the different subjects and datasets used in this work [42], [43]. Also, this normalized range of pixel values is suitable for the activation functions and learning algorithm, which eventually leads to adequate model training with stable optimization process, faster convergence, and improved performance [43].

2) PROPOSED AUTOMATED SEGMENTATION NETWORK: DMAeEDNet

As stated earlier, this work proposes the use of depth-wise separable convolutional layers structured in an encoder-decoder manner along with DMAC to effectively segment the lesions with low time and computational complexity. The proposed DMAeEDNet for automated breast lesion segmentation is built in TensorFlow Keras framework in Python, as

shown in Figure 1 and detailed in Figure 2. It comprises encoding (left) and decoding (right) layers with dense multiplicative attention components (DMAC) as illustrated in Figure 2. The description of these layers is as follows.

- Encoding layers: The encoding layers consist of nine depth-wise separable convolutional (DSepConv2D) layers with batch normalization after each layer. These layers perform non-padded depth-wise and point-wise convolution operation in two steps, which helps in reducing the computation complexity as compared to the conventional convolution operation [44], [45]. These layers enhance the segmentation performance by generalizing better than standard convolutions. In addition, since these layers perform the spatial and cross-channel interactions separately, they can extract more abstract and diverse feature representations, enhancing the capability of the proposed network to learn complex patterns. The basic operation of these layers is explained as follows. Figure 3 illustrates the standard convolution and depth-wise separable convolution operations for a C channel input image with size $I \times I$ and filters with size $k \times k$ [46]. The standard convolution operation is performed for each input channel with a particular filter and consequently, the final outcome is obtained as the sum of the convolution outputs from all channels [46]. All channels are convolved at the same time. The number of weights are obtained from the number of kernels and their dimensions, i.e., $k \times k \times C \times N$. Accordingly, the total number of operations Op_{Conv2D} are estimated as the total number of multiplications [46]:

$$Op_{Conv2D} = I \times I \times k \times k \times N \times C \tag{2}$$

  The depth-wise separable convolution operation is performed in two steps including depth-wise convolution and point-wise convolution. As observed from Figure 3 (b), depth-wise convolution is performed as the convolution with each input channel at a time, by splitting each of the previous filters into C $k \times k \times 1$ filters. Subsequently, point-wise convolution is performed as the standard convolution operation using 1×1 filters, as shown in Figure 3(c). This leads to the number of weights as $k \times k \times C + C \times N$, and the total number of depth-wise separable convolution operations $Op_{DsepConv2D}$ are obtained as the multiplications [46]:

$$Op_{DsepConv2D} = I \times I \times k \times k \times C + I \times I \times C \times N \tag{3}$$

  For comparison of the complexities of standard and depth-wise separable convolution operations, the following ratio is computed:

$$\frac{Op_{DsepConv2D}}{Op_{Conv2D}} = \frac{1}{N} + \frac{1}{k^2}$$

It is evident from the ratio that the complexity of the depth-wise separable convolution operation is considerably reduced by the factor above. Thereby, the number of weights and parameters are significantly low for DSepConv2D layers, as compared to the standard convolution layer.

Figure 1 depicts that the first layer, i.e., DSepConv2D layer 1 receives B-mode images as inputs, the corresponding output is fed to the second layer, and so on. Each DSepConv2D layer uses the rectified linear unit (ReLu) activation function and varying number, i.e., 2^n ($n \in [6, 9]$) of 3 × 3 filters/kernels. Every three convolution layers in the first two encoding parts use the same number of kernels and are followed by a maxpooling2D layer with a stride of 2 for down-sampling. In the next two encoding parts, these layers are reduced to two and one respectively. For reducing the effect of overfitting, dropout regularization with a factor of 0.5 is used after the final encoding layer.

To illustrate the contribution of the encoding depth-wise separable convolution layers towards the overall segmentation process, the feature maps of each layer are generated in this work. As an example, a test image is given input to the trained model and the corresponding encoding layer output feature maps are extracted, as shown in Figure 4. For ease of representation, single filter map is presented at each step. It can be observed from the figure that the input image is consecutively encoded to extract the activations corresponding to the lesion area by each DsepConv2D layer. The first layer convolves the kernel with the input image to obtain its varied intensity activation map, which is further fed to the second layer. The second layer starts to extract the feature mappings towards the segmentation mask of the lesion. These feature/activation mappings are represented by the pixel intensity variations across the lesion area. The third layer further enhances the features, as seen from the figure. The layers 4–6 further extract the lesion area by refining the intensity variations for the lesion area. The deeper encoding layers 7–8 extract the boundary features of the lesion and finally, the layer 9 enhances the intensity activations relevant to the lesion boundary. Therefore, it can be demonstrated that the depth-wise separable convolution layers effectively extract the feature maps which eventually lead to the final segmented mask after passing through the DMAC and decoding layers. Further, it indicates that the depth-wise separable convolution layers can enhance the overall segmentation performance with computationally less intensive convolutions.

- Dense multiplicative attention component (DMAC): The output of the encoding layer nine is fed to a dense multiplicative attention component (DMAC-I). DMAC was proposed by Saini et al. in [44] to classify mental tasks using one-dimensional electroencephalogram signals. However, this work proposes the use of DMAC for the first time to segment breast lesions using the two-dimensional ultrasound images. This component consists of a fully connected/dense layer with the same number of neurons as the filters used in the preceding layer, i.e., 512, followed by an element-wise multiplication operation, as shown in Figure 1. It can be observed that the output of encoding layer nine is multiplied with

the output of the dense layer to enhance the high pixel intensities and suppress the lower pixel intensities in the output. This helps in attaining more valuable features with further reduction in computational complexity due to the reduction in the corresponding relevant weights. The working of the DMAC module is demonstrated by extracting and analyzing the corresponding activation/feature maps of a test image, as shown in Figure 5. It can be observed that the feature map of dense layer in Figure 5 (b)is more selectively refined after the multiplication operation in (c), which helps in capturing the lesion area with enhanced efficacy. Further, the less number of lighter pixel intensities in (c) as compared to (b) demonstrates the reduction in the weights after using the multiply operation in DMAC-I module. Finally, the corresponding DMAC-I output is upsampled and used as the input to the decoding layers.

- Decoding layers: These layers are analogous to the encoding layers, with 2 × 2 up-sampling convolution operations after two 3 × 3 DSepConv 2D layers with ReLu activation functions in the first decoding part, as illustrated in Figure 1. The number of features are thereby reduced by halves after every up-sampling operation. Analogous to the encoding part, the next two decoding parts consist of three 3 × 3 DSepConv 2D layers. The decoding layers also consist of skip connections for concatenating the reduced feature outputs with the corresponding encoding layer outputs as depicted in the figure.

  The output of the final DSepConv2D layer is fed to a second DMAC prior to the final decoding convolution layer for segmentation. DMAC-II refines the output feature map by retaining the more relevant intensities which further helps in attaining robust segmentation output/mask. The working of DMAC-II is illustrated in Figure 6 for the same test image (a) as used previously for DMAC-I explanation. The output of dense layer in (b) in DMAC-II module shows the lesion boundary, however, it is not well defined along with an ambiguity with the background represented by darker pixels. This is resolved by using the multiply operation in DMAC-II in Figure 6 (c) which enhances the features specific to the boundary of the lesion, as observed from the darker pixels specific to the lesion area only. This demonstrates that DMAC-II can effectively discriminate the background from the lesion by selectively capturing the lesion-specific features.

  For segmentation, the final layer is a convolution layer which receives the DMAC-II feature map and uses a sigmoid activation function to predict the mask.

The aforementioned layers are used to build he proposed network. The data is split into 70% training, 20% for validation, and 10% for testing. The selection of 70 – 20 – 10% train-test-validation split percentages is drawn from the statistical modeling methodology [47] and empirical studies [48]. Further, since the datasets have less number of images, as a rule of thumb, we take 70% of that for training, and remaining 30% for testing and validation purposes. This split allows for more number of images in training, and it is reported to be a good starting point for many deep learning algorithms [48]. We have experimentally assessed that the performance of the proposed network with this split is higher, as compared

to that with other data splits such as 60–20–20, 80 – 10 – 10, etc. For the training set, data augmentation is applied using standard geometric transformations such as translation, random flipping and rotating. The rationale behind the augmentation techniques is derived from empirical studies which use geometric transformations to alter the images and increase the number of images [30], [49]. These transformations such as random flipping and rotating used in this work are traditional image augmentation approaches which are simple and easy to execute [49]. They help in increasing the distribution density and variability of the data points in the training set, which eventually leads to improved performance and better generalization. The training hyper-parameters such as batch size, epochs, steps per epoch, learning rate and algorithm, are discussed in the later section. For building the proposed network, TensorFlow Keras 2.2.1 deep learning framework is used in Python. The pseudo-code for the proposed methodology is summarized in Algorithm 1, including the pre-processing, training the proposed DMAeEDNet, and segmented mask prediction.

Algorithm 1 Pseudo-Code for the Proposed Methodology

begin
for each subject **do** // Loading and pre-processing images
  Load inputdata // *BUSI*, *UDIAT*, and *ICR* datasets
  for each subject **do**
    $resized\_rdata \leftarrow resize(data, [512, 512])$
    $save\_image(rdata)$
  end for
  for each $resized\_rdata$ **do**
    $img \leftarrow normalize(resized\_rdata)$
    $save\_image(imgn)$
  end for
end for
for each proposed scheme **do**
  Load all pre-processed images and ground truth masks: $\{imgn_n, gt_n\}_{n=1}^{N}$
  $imgn_{train}, imgn_{test}, imgn_{val}, gt_{train}, gt_{test}, gt_{val} \leftarrow Split(imgn, gt)$
  $Batch\_sizes \leftarrow 2, 4$
  $Learning\_rate \leftarrow 0.0001$
  $No.\_of\_epochs \leftarrow 100$
  $Steps\_per\_epoch \leftarrow Size(img_{train})/Batchsize$
  for $i \leftarrow No.\_of\_epochs$ **do** // DMAeEDNet training using the train and validation sets
    for $j \leftarrow Steps\_per\_epoch$ **do**
      $Augment(batch\_j)$
      $gt'_{train} \leftarrow DMAeEDNet.fit(batch\_j)$
      $loss_{train} \leftarrow 1 - Dice\_coefficient(gt_{train}, gt_{train})$
      $Adam\_optimizer(DMAeEDNet, loss_{train})$
      $gt'_{val} \leftarrow DMAeEDNet.fit(img_{val})$
      $loss_{val} \leftarrow 1 - Dice\_coefficient(gt_val, gt'_{val})$
      $save\_model(DMAeEDNet)$
    end for
  end for
  $loss_{test} \leftarrow 0$ // Predict the segmented masks for the test set using the trained model
  for $t \leftarrow 1$ *to* $size(img_{test})$ **do**
    $gt^{p}_{test}[t] \leftarrow DMAeEDNet.predict(img_{test}[t])$
    $loss_{test}[t] \leftarrow 1 - Dice\_coefficient(gt_{test}[t], gt^{p}_{test}[t])$
    $loss_{test} \leftarrow loss_{test} + loss_{test}[t]$
    $save\_predicted\_segmented\_mask(gt^{p}_{test}[t])$
  end for
  $average\_loss_{test} \leftarrow loss_{test}/size(img_{test})$
end for

IV. RESULTS

This section discusses about the performance assessment of the proposed DMAeEDNet for automated segmentation of breast lesions in ultrasound images. The different learning parameters and segmentation metrics are elaborated for analysis of the proposed network.

A. TRAINING HYPER-PARAMETERS

The proposed network utilizes different learning parameters during training phase. In this work, Keras tuner is used for optimal hyper-parameter selection. Table 1 lists these parameters. It can be observed from the table that an extensive analysis is performed for network performance evaluation, by using different batch sizes as also used in some literary works [50], [51], [52], epochs based on early stopping criteria, and steps per epoch. For optimization and handling the local minima issue during training, Adam optimizer, i.e.,

adaptive moment estimation is used with a small learning rate of 0.0001 as listed in the table. Adam optimizer combines momentum with the adaptive learning algorithm, root mean square propagation, RMSprop [53]. It controls the rate of gradient descent for minimum oscillation at reaching the global minima as well as incorporating high enough step size to surpass the local minima problems along the path. Thereby, smaller learning rate is required [50] than the usual of 0.001 used mostly. This helps in reaching efficient global minimum for convergence. This work also uses batch normalization, dropout regularization as mentioned previously and also used in existing works [50], [51], [52], and He uniform weight initialization techniques for better convergence.

B. SEGMENTATION PERFORMANCE METRICS

The performance of the proposed network is evaluated using a train-validation-test split of 70%−20%−10% respectively for both the datasets used in this work. For assessing the generalization capability of the proposed network, ten-fold cross-validation is used in both the datasets. To overcome the class-imbalance in segmented masks, dice loss function and dice coefficient metric are used during the training process for performance assessment. The test performance of the proposed network is evaluated using different well-adapted segmentation metrics such as, dice coefficient (*DC*), jaccard index (*JI*), precision (*PR*), sensitivity/recall (*SE*), false positive rate (*FPR*), and Matthews correlation coefficient (*MCC*). Description of these metrics is provided in [54].

1) PERFORMANCE ASSESSMENT SCHEMES—In this work, different evaluation schemes are proposed for an exhaustive analysis of the proposed segmentation network with respect to intra-dataset, inter-dataset, and inter-machine images. For all schemes, data augmentation is used by randomly generating multiple images from the data presented to the model. This helps in better generalization of the model by learning diverse feature representations. A description of these schemes is provided as follows.

- Scheme I- Training and testing using *UDIAT* dataset: This scheme utilizes 70%–20% training-validation split for ultrasound images from the *UDIAT* dataset during training phase, and the remaining 10% images during testing phase.
- Scheme II- Training and testing using *BUSI* dataset: This scheme utilizes 70%–20% training-validation split for ultrasound images from the *BUSI* dataset during training phase, and the remaining 10% images during testing phase.
- Scheme III- Training and testing using both *BUSI-UDIAT* datasets: This scheme utilizes 70% – 20% training-validation split for ultrasound images from both the above mentioned datasets combined together during training phase, and the remaining 10% from both datasets during the testing phase.
- Scheme IV- Training using *UDIAT* dataset and testing using *BUSI* dataset: This scheme utilizes 70% – 30% training-validation split for ultrasound images from the *BUSI* dataset during training phase, and all images from the *BUSI* dataset during the testing phase.
- Scheme V- Training using *BUSI* dataset and testing using *UDIAT* dataset: This scheme utilizes 70% – 30% training-validation split for ultrasound images from

BUSI dataset during the training phase, and all images from the *UDIAT* dataset during the testing phase.

- Scheme VI- Training using *BUSI-UDIAT* datasets and testing using inter-machine clinical data: This scheme utilizes 70% – 30% training-validation split for ultrasound images from the combined *BUSI-UDIAT* dataset during the training phase, and the clinical ultrasound images taken from different machine during the testing phase.

2) PERFORMANCE ANALYSIS—The evaluation metrics for different schemes and batch sizes are presented in Table 2. It can be observed from the table that for schemes I and II, the overall segmentation performance is superior to that achieved in the other schemes. This implies that training and testing the network using the individual datasets helps in prominent feature extraction resulting in effective segmentation of lesion area, with overall dice coefficients of 0.86 and 0.83 for batch size of 4. This attributes to the presence of pre-processed images in these datasets, and training and testing on the images captured using same machine and protocol. For scheme III, the performance is quite comparable with the previous two schemes, with dice coefficients 0.80 and 0.81, indicating the effectiveness of the proposed network in segmenting the images acquired from the combined datasets. A slight drop in the performance attributes to the training process, which uses combinations of images from all the datasets leading to increased variations in the data. This helps in better generalization but slight decrease in the metrics. Although, the change in performance is not too significant, which demonstrates that the proposed network is capable of learning variable feature representations of images taken from different machines.

As mentioned previously, the cross/inter-dataset and inter-machine ultrasound images are also used to analyze the robustness of the proposed network. It can be observed from Table 2 that for scheme IV and V, the overall performances are comparable with the schemes II and I respectively. It can be observed that for scheme V, the performance is better than scheme IV, which attributes to the test images taken from the *UDIAT* dataset correlating well with the performance in scheme I. Since the images used in training and testing belong to different datasets collected from different machines, there is a slight decline in the performance. However, the network is fairly able to segment the images with dice coefficients of above 0.79, which demonstrates its effectiveness in segmenting inter-dataset/machine images. This also indicates that the proposed network can segment the breast lesion for a new patient, without being trained on it prior. Furthermore, it can be observed from the table that in scheme VI, although the test images used are taken from different machines and datasets, the proposed network is still able to achieve comparable dice coefficients above 0.78, This attributes to the less number and low clarity of the test images taken from that machine (LE9). Finally, this demonstrates the robustness of the proposed network with respect to inter-machine and quality of data.

3) PROBABILITY DISTRIBUTION ANALYSIS OF PERFORMANCE—In this sub-section, the dice coefficients obtained by the proposed segmentation network are investigated for the variations in probability distributions. For this purpose, the *DC* of each test image is plotted for both datasets used in this work. Since scheme I and II utilize the two

datasets individually, these schemes are selected for analyzing the dataset-wise performance distribution. Figure 7 represents the probability distribution of the *DC* values in the datasets. It is observable that best performance is achieved for the *UDIAT* dataset in Figure 7 (a), with a mean *DC* of 0.86. Further, for the *DC* values of above 0.9, the highest probability is achieved as represented by the width of the curve. This demonstrates the maximum likelihood of segmenting the images with above 90% dice score in this dataset. Similar interpretations can be drawn for the *BUSI* datasets in Figure 7 (b). It can be perceived that the proposed DMAeEDNet is able to achieve about 85% and 90% dice scores for most of the images in the *BUSI* and *UDIAT* datasets respectively. Finally, this indicates a high probability of DMAeEDNet in segmenting a lesion with a dice score of above 0.85 in both the datasets, demonstrating its adaptability to the data.

4) SEGMENTATION PERFORMANCE EXEMPLIFICATION—This sub-section presents a demonstration of the segmented masks by the proposed network, for the different schemes mentioned previously. Based on the dataset used for testing, the segmented masks are illustrated for distinct combinations of the schemes in Figure 8. For each dataset, the true and predicted segmented masks are exemplified in the figure.

Figure 8(a) represents the masked true and predicted segmented lesions in (ii) and (iii)-(iv) respectively for a test image in (i) taken from the *BUSI* dataset. It can be observed that the lesion predicted using scheme II in (iii) is closely matched to the ground truth with respect to shape and boundary. The same is indicated by the dice coefficient value of 0.93, which is higher in comparison to that using schemes III and IV in (iv) and (v) respectively. The decrease in performance is due to the variation in the training images in this scheme which causes the network to learn more diverse feature representations rather than specific ones. Figure 8(b) illustrates the masked true and predicted segmented lesions in (ii) and (iii)-(v) respectively for a test image in (i) taken from the *UDIAT* dataset. Similar interpretation can be drawn for this dataset, however, the inter-dataset variability is minimal in this case. This owes to the high contrast and clarity of the images in the *UDIAT* dataset, as also observed from its best overall performance as mentioned previously.

C. NOISE ROBUSTNESS ANALYSIS

This sub-section analyses the performance of the proposed network on the *ICR* dataset, which consists of artifactual images, without any pre-processing. For this purpose, 70% – 20% training-validation split is used for ultrasound images from the *ICR* dataset during training phase, and the remaining 10% images are used during testing phase. An overall average *DC* of 0.73 is achieved, along with a sensitivity of 0.82. This indicates that the proposed network is able to fairly segment the breast lesions in the low contrast and artifactual images, demonstrating its robustness with respect to noise. Figure 9 shows the probability distribution graph for the *DC* values obtained for the test data. The highest probability is achieved for the *DC* values of about 0.8, as represented by the width of the curve. This indicates the maximum likelihood of segmenting the images with approximately 80% dice score in this dataset. Finally, this demonstrates the capability of the proposed network in learning the feature representations in the artifactual images, and effectively segmenting them.

D. EFFECTIVENESS OF DMAC

This work proposes the use of dense multiplicative attention components for efficient segmentation of lesion in breast ultrasound imaging. To demonstrate the effectiveness of DMAC, an ablation analysis of the proposed network is presented in this sub-section. For this purpose, the proposed network is analyzed after removing the DMAC for *ICR* dataset. An overall average dice coefficient of 0.62 is achieved, which is low in comparison to that obtained with DMAC. Figure 10 represents the true and predicted masks using network with and without DMAC for a B-mode test image in *ICR* dataset. It can be observed that the segmented mask with DMAC is significantly appropriate in comparison to that without DMAC. This indicates that the dense multiplicative attention enhances the segmentation performance by extracting more relevant feature representations from the input images.

E. LEARNING CURVES

This sub-section elaborates the training and validation performance of the proposed segmentation network to demonstrate its learning process with respect to the iterations/ epochs. Figure 11 illustrates the learning curves of the proposed network for all schemes during training phase. It can be observed that for the network convergence is slower for schemes I, III, IV, and V, and faster for the other two schemes. However, the overall training and validation performance indicates that the proposed network does not over-fit irrespective of the speed of convergence.

F. PERFORMANCE COMPARISONS AND COMPUTATIONAL COMPLEXITY ANALYSIS

This sub-section describes the performance comparison of the proposed network with different recent state-of-art deep learning networks for breast lesion segmentation. For this purpose, some of the previously proposed segmentation networks based on encoder-decoder structure are implemented and evaluated using all schemes in this study. Based on the similarity of implementation with the proposed network, the following networks are used for experimentation: U-Net [30], adaptive attention U-Net (AAU-Net) [31], selective kernel U-Net (SKU-Net) [32], multi-U-Net [33], and attention enhanced U-Net (AEU-Net) [36]. Figure 12 shows the different performance metrics estimated after implementing these networks using all six schemes proposed in this work. It can be observed that the proposed network outperforms all other networks for all schemes, with a comparable performance with the recently proposed AAU-Net, and U-Net for scheme I. Further, the proposed network uses separable convolutions which reduce the overall complexity, thereby, indicating its effectiveness with respect to U-Net and AAU-Net. Further, the performance for the last three schemes demonstrates the robustness of the proposed network with respect to inter-machine data in comparison to the existing networks. In this work, an analysis of the time complexities involved in prediction of the segmentation masks by existing and proposed networks is performed. All the network implementations including training and testing are done on Precision 7920 tower workstation with NVIDIA® RTX™ A4000, 16 GB GDDR6, 4 DP GPU. Figure 13 shows the average execution times of the existing networks and the proposed network during test phase, for each image. The execution time is shown after averaging over all schemes, and denotes the prediction time to estimate the segmented mask/lesion for each image. It is apparent from the figure that the proposed

network outperforms the existing networks in terms of time complexity, with a considerably low latency of 19*ms*.

This work elaborates the comparative analysis of the proposed DMAeEDNet by including dataset-wise metrics reported by existing works in table 3. The values of all metrics are adopted from a recent work in [9]. The methods which use both *BUSI* and *UDIAT* datasets are selected and listed in the table. It is evident from the table that the proposed DMAeEDNet outperforms all the reported methods for both datasets in terms of all metrics, and is in close proximity to the recently proposed method in [9]. Furthermore, DMAeEDNet uses significantly low number of model parameters and floating point operations (FLOPs) as compared to all methods reported in the table. This demonstrates the effectiveness of the proposed network in terms of segmentation performance and low computational complexity of the model.

V. DISCUSSION

The performance assessment and analysis schemes used in this work demonstrate the capability of the proposed network towards effective lesion segmentation. The purpose of using diverse schemes is to investigate the behavior of network with respect to inter-dataset and inter-machine ultrasound images. It is clearly observable from Table 2 and Figure 8 that the proposed network exhibits a minuscule trade-off between generalization and performance.

In this study, the learning graphs of the proposed network in Figure 11 demonstrate its ability to achieve similar performance for both training and validation data with respect to the iterations. This further indicates its appropriate learning behaviour without any over-fitting. Subsequent illustration using network ablation without DMAC in Figure 10 shows the significance of the proposed dense multiplicative attention components. Additionally, it performs computationally less expensive operations by reducing the overall number of weights in the feature matrices. Further feature map-based analysis demonstrates the activations of the proposed network before and after both DMAC modules I and II in Figures 5 and 6 respectively. It is demonstrated in these figures that DMAC-I performs enrichment of the encoder feature map by selecting the lesion-relevant features with multiplication operation, and DMAC-II refines the reconstructed the feature map for more precise segmentation output. It signifies the relevance of DMAC modules in selectively capturing the lesion-specific boundary features. Therefore, the proposed network with DMAC is capable of performing effective segmentation with low computational complexity.

The exhaustive comparison of the proposed network with select state-of-art encoder-decoder-based systems indicates that the existing attention-based networks are comparable to the proposed network in terms of performance. While, the attention components in the proposed work are light-weight, in terms of implementation and computation, as opposed to these existing attention-based networks AAU-Net and AEU-Net which are complex and require more explain-ability. Therefore, the effectiveness of proposed DMAeEDNet with respect to the existing networks is demonstrable. Further, the time complexity analysis in Figure 13 indicates the superiority of the proposed network in terms of efficient

segmentation with low latency, and further demonstrates its applicability in real-time. It is also observable that the U-Net architecture exhibits a trade-off between time complexity and performance, however, the proposed network overcomes this issue by obtaining better performance with low time complexity. The significant reduction in computation time is crucial for real-time automatic segmentation. Additional comprehensive comparative evaluation of the proposed method with recently reported existing methods in [9] further signifies its superiority in segmentation performance with considerably low computational complexity in terms of FLOPs and model parameters. Finally, the meticulous comparative analysis demonstrates the reduction in overall complexity of the proposed network with the use of depthwise separable convolution layers and its effectiveness in feature extraction with the introduction of DMAC modules.

The proposed network is able to segment the breast lesions with superior performance and low complexity, however, it has certain limitations with respect to the average performance for inter-dataset images in schemes IV and VI, as well as the noise robustness test for *ICR* dataset. Further, there is a slight decrease in the segmentation performance for images with ambiguous and hazy lesion boundaries. Therefore, the future direction of the proposed study shall be towards further improvement of the segmentation performance of the inter-dataset images, the images which have unclear lesion boundaries, and noise robustness. This shall be achieved by exploring some modifications in the proposed attention components and the pooling operations. Some recently proposed hyper-parameter tuning schemes may be explored for optimization of the proposed network [65]. Further, the noise robustness analysis will be performed for the public datasets by artificially contaminating the images, along with more images acquired from different machines. Finally, the segmented masks will be utilized in conjunction with the original images for breast lesion categorization.

VI. CONCLUSION

In this study, a novel dense multiplicative attention enhanced encoder decoder network is proposed for ultrasound-based automated breast lesion segmentation. An extensive evaluation analysis using different training schemes demonstrates the adeptness of the proposed network toward segmentation of breast lesions. The relevance of the dense multiplicative attention components is also demonstrated through the network ablation and feature-map-based analysis.The comparative studies using implementation and evaluation of existing networks with different schemes indicate the supremacy of the proposed network in terms of all segmentation metrics and latency. Further comparisons with recently reported works indicates its efficacy in segmenting the lesions with significantly low computational complexity. Finally, the low time and computational complexity, and inter-machine/dataset performance of DMAeEDNet demonstrates its suitability for real-time breast lesion segmentation.

ACKNOWLEDGMENT

The authors would like to thank Cindy Andrist, Patricia O’Neil, and Lesley Riess for their valuable help in patient recruitment. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The NIH did not have any additional role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

This work was supported in part by the National Institute of Health under Grant R01CA239548, Grant R01CA1-48994, and Grant R01CA195527; and in part by the National Science Foundation under Grant NSF1837572.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Mayo Clinic Institutional Review Board, (Mayo Clinic IRB) under IRB: 12-003329 and IRB: 19-003028, and performed in line with the Health Insurance Portability and Accountability Act.

Biographies

MANALI SAINI (Member, IEEE) received the B.Tech. degree in electronics and communication engineering (ECE) from Guru Gobind Singh Indraprastha University (GGSIPU), Dwarka, India, in 2013, the M.Tech. degree in mobile and pervasive computing from Indira Gandhi Delhi Technical University for Women, New Delhi, India, in 2016, and the Ph.D. degree in ECE from the Shiv Nadar Institution of Eminence, Greater Noida, India, in 2022.

She was a Postdoctoral Research Associate with the Department of Electrical Engineering, Indian Institute of Technology Delhi (IITD). She is currently a Postdoctoral Research Fellow with the Department of Radiology and the Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science, Rochester, MN, USA. Her research interests include biomedical signal processing, tiny machine learning for IoT, and applied deep learning.

HUMAYRA AFRIN received the Medical Doctorate degree from the Shaheed Suhrawardy Medical College, Dhaka, Bangladesh, in 2018. Since 2022, she has been a Visiting Postdoctoral Research Fellow with the Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science. Her research interests include the application of novel imaging modalities using machine learning approaches. She has been an ECFMG Certified with the U.S., since 2023.

SETAYESH SOTOUDEHNIA received the Medical Doctorate degree from the Hormozgan University of Medical Sciences, in 2021. Since 2022, she has been a Visiting Postdoctoral Research Fellow with the Department of Radiology, Mayo Clinic College of Medicine and Science. Her research interests include the application of novel imaging modalities using machine learning approaches.

MOSTAFA FATEMI (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from Purdue University. In 1994, he joined the Mayo Clinic College of Medicine, Rochester, MN, USA, where he is currently a Professor of biomedical engineering with the Department of Physiology and Biomedical Engineering. At the Mayo Clinic, he is also an active member of the Mayo Clinic Cancer Center and the Center for Clinical and Translational Science. His current research interests include developing novel ultrasonic methods for cancer imaging, microvasculature imaging, and bladder function evaluation. His research program has been continuously supported by multiple major federal grants from the National Institute of Health (NIH), the National Science Foundation (NSF), the Department of Defense, and the Komen Breast Foundation for the Cure. He holds fellow membership at several institutions, such as American Institute of Medical and Biological Engineering (AIMBE), Acoustical Society of America (ASA), and American Institute of Ultrasound in Medicine (AIUM). He was a recipient of the IEEE-UFFC Distinguished Lecturer Award.

AZRA ALIZAD (Senior Member, IEEE) received the Medical degree and fellowship from Tehran University Medical Sciences. She is currently a Professor of radiology with the Department of Radiology and a Professor of biomedical engineering and an Associate Professor of medicine with the Mayo Clinic College of Medicine, Rochester, MN, USA. At the Mayo Clinic, she is also an active member of the Mayo Clinic Cancer Center and the Center for Clinical and Translational Science. She is a Principal Investigator of multiple major federal grants funded by the National Institute of Health (NIH). In addition to her training in medicine, she has a broad background and expertise in medical ultrasound research. She directs a translational research laboratory, involving the development and application of novel ultrasound technologies, such as microvasculature imaging, functional ultrasound, elastography, and vibro-acoustic imaging for diagnosis of abnormalities in breast, thyroid, prostate, neck masses, axillary lymph nodes, brain, and bone. She is elected as a fellow of American Institute for Medical and Biological Engineering (AIMBE) and the American Institute of Ultrasound in Medicine (AIUM) and a Senior Member of the Institute

of Electrical and Electronics Engineers-UFFC. She was a recipient of a grant funded by the Komen Breast Foundation for the Cure.

REFERENCES

[1]. Siegel RL, Miller KD, Fuchs HE, and Jemal A, “Cancer statistics, 2022,” CA, Cancer J. Clin, vol. 72, no. 1, pp. 7–33, Jan. 2022. [PubMed: 35020204]

[2]. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, Davison AK, and Martí R, “Automated breast ultrasound lesions detection using convolutional neural networks,” IEEE J. Biomed. Health Informat, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.

[3]. Winters S, Martin C, Murphy D, and Shokar NK, “Breast cancer epidemiology, prevention, and screening,” Prog. Mol. Biol. Transl. Sci, vol. 151, pp. 1–32, Jan. 2017. [PubMed: 29096890]

[4]. Ghodsi Z and Hojjatoleslami S, “Breast self examination and mammography in cancer screening: Women health protective behavior,” J. Preventive Med. Hygiene, vol. 55, no. 2, p. 46, 2014.

[5]. Gerbasi A, Clementi G, Corsi F, Albasini S, Malovini A, Quaglini S, and Bellazzi R, “DeepMiCa: Automatic segmentation and classification of breast MIcroCAlcifications from mammograms,” Comput. Methods Programs Biomed, vol. 235, Jun. 2023, Art. no. 107483.

[6]. Zhou Y, Chen H, Li Y, Cao X, Wang S, and Shen D, “Cross-model attention-guided tumor segmentation for 3D automated breast ultrasound (ABUS) images,” IEEE J. Biomed. Health Informat, vol. 26, no. 1, pp. 301–311, Jan. 2022.

[7]. Aristokli N, Polycarpou I, Themistocleous SC, Sophocleous D, and Mamais I, “Comparison of the diagnostic performance of magnetic resonance imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient’s history: A review,” Radiography, vol. 28, no. 3, pp. 848–856, Aug. 2022. [PubMed: 35148941]

[8]. Kim YE, Cha JH, Kim HH, Shin HJ, Chae EY, and Choi WJ, “The accuracy of mammography, ultrasound, and magnetic resonance imaging for the measurement of invasive breast cancer with extensive intraductal components,” Clin. Breast Cancer, vol. 23, no. 1, pp. 45–53, Jan. 2023. [PubMed: 36328930]

[9]. Hu K, Zhang X, Lee D, Xiong D, Zhang Y, and Gao X, “Boundary-guided and region-aware network with global scale-adaptive for accurate segmentation of breast tumors in ultrasound images,” IEEE J. Biomed. Health Informat, vol. 27, no. 9, pp. 4421–4432, Sep. 2023.

[10]. Webb JM, Adusei SA, Wang Y, Samreen N, Adler K, Meixner DD, Fazzio RT, Fatemi M, and Alizad A, “Comparing deep learning-based automatic segmentation of breast masses to expert interobserver variability in ultrasound imaging,” Comput. Biol. Med, vol. 139, Dec. 2021, Art. no. 104966.

[11]. Webb JM, Meixner DD, Adusei SA, Polley EC, Fatemi M, and Alizad A, “Automatic deep learning semantic segmentation of ultrasound thyroid cineclips using recurrent fully convolutional networks,” IEEE Access, vol. 9, pp. 5119–5127, 2021. [PubMed: 33747681]

[12]. Chen G, Dai Y, and Zhang J, “C-Net: Cascaded convolutional neural network with global guidance and refinement residuals for breast ultrasound images segmentation,” Comput. Methods Programs Biomed, vol. 225, Oct. 2022, Art. no. 107086.

[13]. Ranjbarzadeh R, Dorosti S, Ghoushchi SJ, Caputo A, Tirkolaee EB, Ali SS, Arshadi Z, and Bendechache M, “Breast tumor localization and segmentation using machine learning techniques: Overview of datasets, findings, and methods,” Comput. Biol. Med, vol. 152, Jan. 2023, Art. no. 106443.

[14]. Cheng Z, Li Y, Chen H, Zhang Z, Pan P, and Cheng L, “DSGMFFN: Deepest semantically guided multi-scale feature fusion network for automated lesion segmentation in ABUS images,” Comput. Methods Programs Biomed, vol. 221, Jun. 2022, Art. no. 106891.

[15]. Zhong S, Tu C, Dong X, Feng Q, Chen W, and Zhang Y, “MsGoF: Breast lesion classification on ultrasound images by multi-scale gradational-order fusion framework,” Comput. Methods Programs Biomed, vol. 230, Mar. 2023, Art. no. 107346.

[16]. van Sloun RJG, Cohen R, and Eldar YC, “Deep learning in ultrasound imaging,” Proc. IEEE, vol. 108, no. 1, pp. 11–29, Jan. 2020.

[17]. Khaled R, Vidal J, Vilanova JC, and Martí R, "A U-Net ensemble for breast lesion segmentation in DCE MRI," Comput. Biol. Med, vol. 140, Jan. 2022, Art. no. 105093.

[18]. Chen G, Liu Y, Dai Y, Zhang J, Cui L, and Yin X, "BAGNet: Bidirectional aware guidance network for malignant breast lesions segmentation," in Proc. 7th Asia–Pacific Conf. Intell. Robot Syst. (ACIRS), Tianjin, China, Jul. 2022, pp. 112–116.

[19]. Kumar V, Webb J, Gregory A, Meixner DD, Knudsen JM, Callstrom M, Fatemi M, and Alizad A, "Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning," IEEE Access, vol. 8, pp. 63482–63496, 2020. [PubMed: 32995106]

[20]. Ak MF, "A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications," Healthcare, vol. 8, no. 2, p. 111, Apr. 2020. [PubMed: 32357391]

[21]. Xu Y, Wang Y, Yuan J, Cheng Q, Wang X, and Carson PL, "Medical breast ultrasound image segmentation by machine learning," Ultrasonics, vol. 91, pp. 1–9, Jan. 2019. [PubMed: 30029074]

[22]. Ilesanmi AE, Chaumrattanakul U, and Makhanov SS, "Methods for the segmentation and classification of breast ultrasound images: A review," J. Ultrasound, vol. 24, no. 4, pp. 367–382, Dec. 2021. [PubMed: 33428123]

[23]. Ma Z, Qi Y, Xu C, Zhao W, Lou M, Wang Y, and Ma Y, "ATFE-Net: Axial transformer and feature enhancement-based CNN for ultrasound breast mass segmentation," Comput. Biol. Med, vol. 153, Feb. 2023, Art. no. 106533.

[24]. Lindsay GW, "Convolutional neural networks as a model of the visual system: Past, present, and future," J. Cognit. Neurosci, vol. 33, no. 10, pp. 2017–2031, Sep. 2021. [PubMed: 32027584]

[25]. Mohamad DNFP, Mashohor S, Mahmud R, Hanafi M, and Bahari N, "Transition of traditional method to deep learning based computer-aided system for breast cancer using automated breast ultrasound system (ABUS) images: A review," Artif. Intell. Rev, vol. 56, no. 12, pp. 15271–15300, Dec. 2023.

[26]. Huang R, Lin M, Dou H, Lin Z, Ying Q, Jia X, Xu W, Mei Z, Yang X, Dong Y, Zhou J, and Ni D, "Boundary-rendering network for breast lesion segmentation in ultrasound images," Med. Image Anal, vol. 80, Aug. 2022, Art. no. 102478.

[27]. Meiburger KM, Acharya UR, and Molinari F, "Automated localization and segmentation techniques for B-mode ultrasound images: A review," Comput. Biol. Med, vol. 92, pp. 210–235, Jan. 2018. [PubMed: 29247890]

[28]. Lyu Y, Xu Y, Jiang X, Liu J, Zhao X, and Zhu X, "AMS-PAN: Breast ultrasound image segmentation model combining attention mechanism and multi-scale features," Biomed. Signal Process. Control, vol. 81, Mar. 2023, Art. no. 104425.

[29]. Ru J, Lu B, Chen B, Shi J, Chen G, Wang M, Pan Z, Lin Y, Gao Z, Zhou J, Liu X, and Zhang C, "Attention guided neural ODE network for breast tumor segmentation in medical images," Comput. Biol. Med, vol. 159, Jun. 2023, Art. no. 106884.

[30]. Ronneberger O, Fischer P, and Brox T, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent Cham, Switzerland: Springer, 2015, pp. 234–241.

[31]. Chen G, Li L, Dai Y, Zhang J, and Yap MH, "AAU-Net: An adaptive attention U-Net for breast lesions segmentation in ultrasound images," IEEE Trans. Med. Imag, vol. 42, no. 5, pp. 1289–1300, May 2023.

[32]. Byra M, Jarosik P, Szubert A, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, and Andre M, "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," Biomed. Signal Process. Control, vol. 61, Aug. 2020, Art. no. 102027.

[33]. Kumar V, Webb JM, Gregory A, Denis M, Meixner DD, Bayat M, Whaley DH, Fatemi M, and Alizad A, "Automated and real-time segmentation of suspicious breast masses using convolutional neural network," PLoS One, vol. 13, no. 5, May 2018, Art. no. e0195816.

[34]. Guo Y, Duan X, Wang C, and Guo H, "Segmentation and recognition of breast ultrasound images based on an expanded U-Net," PLoS One, vol. 16, no. 6, Jun. 2021, Art. no. e0253202.

[35]. Punn NS and Agarwal S, "RCA-IUnet: A residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging," Mach. Vis. Appl, vol. 33, no. 2, p. 27, Mar. 2022.

[36]. Yan Y, Liu Y, Wu Y, Zhang H, Zhang Y, and Meng L, "Accurate segmentation of breast tumors using AE U-Net with HDC model in ultrasound images," Biomed. Signal Process. Control, vol. 72, Feb. 2022, Art. no. 103299.

[37]. Tong Y, Liu Y, Zhao M, Meng L, and Zhang J, "Improved U-Net MALF model for lesion segmentation in breast ultrasound images," Biomed. Signal Process. Control, vol. 68, Jul. 2021, Art. no. 102721.

[38]. Li J, Cheng L, Xia T, Ni H, and Li J, "Multi-scale fusion U-Net for the segmentation of breast lesions," IEEE Access, vol. 9, pp. 137125–137139, 2021.

[39]. Ning Z, Zhong S, Feng Q, Chen W, and Zhang Y, "SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image," IEEE Trans. Med. Imag, vol. 41, no. 2, pp. 476–490, Feb. 2022.

[40]. Al-Dhabyani W, Gomaa M, Khaled H, and Fahmy A, "Dataset of breast ultrasound images," Data Brief, vol. 28, Feb. 2020, Art. no. 104863.

[41]. Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, Wood BJ, Roth H, Myronenko A, Xu D, and Xu Z, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," IEEE Trans. Med. Imag, vol. 39, no. 7, pp. 2531–2540, Jul. 2020.

[42]. Gonzalez RC and Woods RE, Digital Image Processing. Upper Saddle River, NJ, USA: Prentice-Hall, 2007, p. 85.

[43]. Pei X, Zhao YH, Chen L, Guo Q, Duan Z, Pan Y, and Hou H, "Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences," Mater. Design, vol. 232, Aug. 2023, Art. no. 112086.

[44]. Saini M, Satija U, and Upadhayay MD, "DSCNN-CAU: Deep-learning-based mental activity classification for IoT implementation toward portable BCI," IEEE Internet Things J, vol. 10, no. 10, pp. 8944–8957, May 15, 2023.

[45]. Saini M, Satija U, and Upadhayay MD, "One-dimensional convolutional neural network architecture for classification of mental tasks from electroencephalogram," Biomed. Signal Process. Control, vol. 74, Apr. 2022, Art. no. 103494.

[46]. Bai L, Zhao Y, and Huang X, "A CNN accelerator on FPGA using depthwise separable convolution," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 65, no. 10, pp. 1415–1419, Oct. 2018.

[47]. Dangeti P, Statistics for Machine Learning. Birmingham, U.K.: Packt Publishing, 2017.

[48]. Cabrera M, Ninic J, and Tizani W, "Fusion of experimental and synthetic data for reliable prediction of steel connection behaviour using machine learning," Eng. Comput, vol. 39, no. 6, pp. 3993–4011, Dec. 2023.

[49]. Xu M, Yoon S, Fuentes A, and Park DS, "A comprehensive survey of image augmentation techniques for deep learning," Pattern Recognit., vol. 137, May 2023, Art. no. 109347.

[50]. Akbar S, Hayat M, Tahir M, Khan S, and Alarfaj FK, "CACP-DeepGram: Classification of anticancer peptides via deep neural network and skip-gram-based word embedding model," Artif. Intell. Med, vol. 131, Sep. 2022, Art. no. 102349.

[51]. Akbar S, Raza A, Shloul TA, Ahmad A, Saeed A, Ghadi YY, Mamyrbayev O, and Tag-Eldin E, "PAtbP-EnC: Identifying anti-tubercular peptides using multi-feature representation and genetic algorithm-based deep ensemble model," IEEE Access, vol. 11, pp. 137099–137114, 2023.

[52]. Raza A, Uddin J, Almuhaimeed A, Akbar S, Zou Q, and Ahmad A, "AIPs-SnTCN: Predicting anti-inflammatory peptides using fastText and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks," J. Chem. Inf. Model, vol. 63, no. 21, pp. 6537–6554, Nov. 2023. [PubMed: 37905969]

[53]. Buduma N, Buduma N, and Papa J, Fundamentals of Deep Learning. Sebastopol, CA, USA: O'Reilly Media, 2022.

[54]. Chen G, Dai Y, Li R, Zhao Y, Cui L, and Yin X, “SDFNet: Automatic segmentation of kidney ultrasound images using multi-scale low-level structural feature,” Expert Syst. Appl, vol. 185, Dec. 2021, Art. no. 115619.

[55]. Badrinarayanan V, Kendall A, and Cipolla R, “SegNet: A deep convolutional encoder–decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell, vol. 39, no. 12, pp. 2481–2495, Dec. 2017. [PubMed: 28060704]

[56]. Zhao H, Shi J, Qi X, Wang X, and Jia J, “Pyramid scene parsing network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2881–2890.

[57]. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, and Rueckert D, “Attention U-Net: Learning where to look for the pancreas,” in Proc. MIDL, 2018, pp. 1–10.

[58]. Chen LC, Zhu Y, Papandreou G, Schroff F, and Adam H, “Encoder–decoder with atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis, 2018, pp. 801–818.

[59]. Shareef B, Xian M, and Vakanski A, “Stan: Small tumor-aware network for breast ultrasound image segmentation,” in Proc. IEEE 17th Int. Symp. Biomed. Imag, Apr. 2020, pp. 1–5.

[60]. Shareef B, Vakanski A, Freer PE, and Xian M, “ESTAN: Enhanced small tumor-aware network for breast ultrasound image segmentation,” Healthcare, vol. 10, no. 11, p. 2262, Nov. 2022. [PubMed: 36421586]

[61]. Huang K, Zhang Y, Cheng HD, and Xing P, “Shape-adaptive convolutional operator for breast ultrasound image segmentation,” in Proc. IEEE Int. Conf. Multimedia Expo. (ICME), Jul. 2021, pp. 1–6.

[62]. Xu M, Huang K, Chen Q, and Qi X, “MSSA-Net: Multi-scale self-attention network for breast ultrasound image segmentation,” in Proc. IEEE 18th Int. Symp. Biomed. Imag, Apr. 2021, pp. 827–831.

[63]. Huang K, Zhang Y, Cheng HD, and Xing P, “MSF-GAN: Multi-scale fuzzy generative adversarial network for breast ultrasound image segmentation,” in Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc, Nov. 2021, pp. 3193–3196.

[64]. Xu M, Huang K, and Qi X, “A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification,” IEEE Access, vol. 11, pp. 5377–5392, 2023.

[65]. Kaur M, Singh D, Kumar V, and Lee H-N, “MLNet: Metaheuristics-based lightweight deep learning network for cervical cancer diagnosis,” IEEE J. Biomed. Health Informat, vol. 27, no. 10, pp. 5004–5014, Oct. 2022.

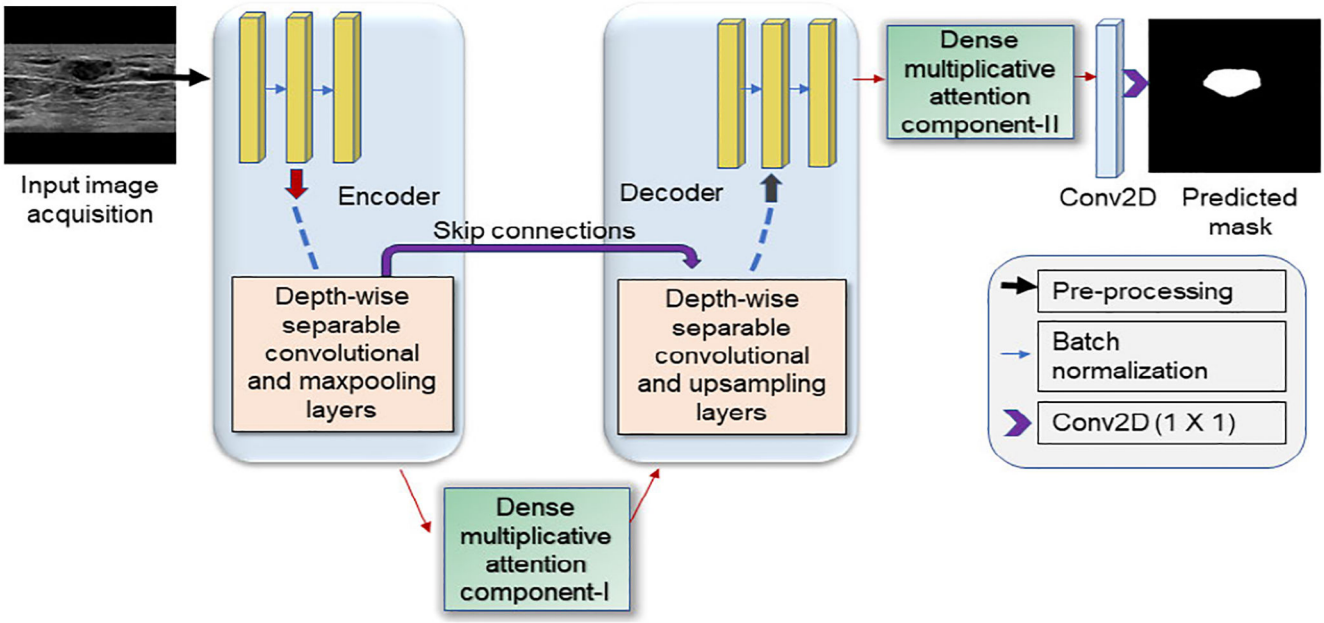

FIGURE 1.
High-level overview of the proposed segmentation network, DMAeEDNet.

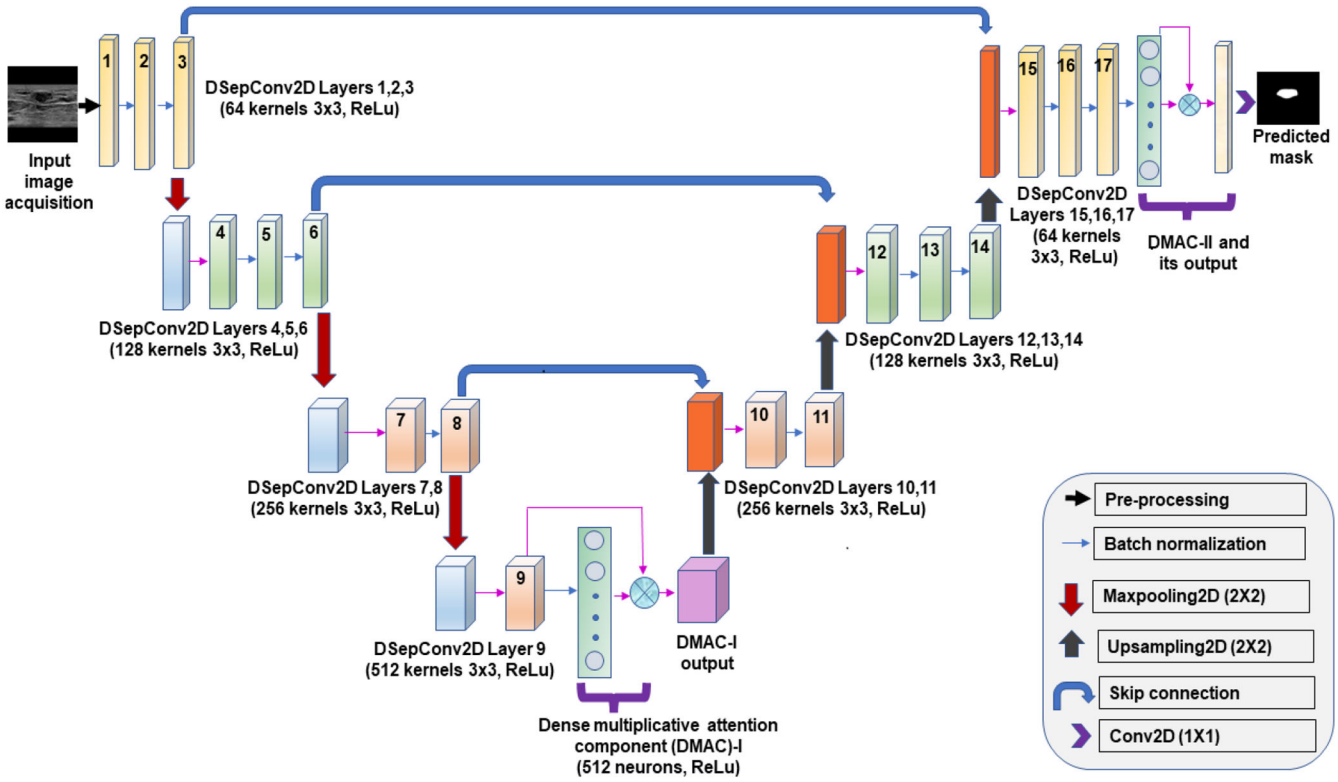

FIGURE 2.
Proposed segmentation network: DMAeEDNet.

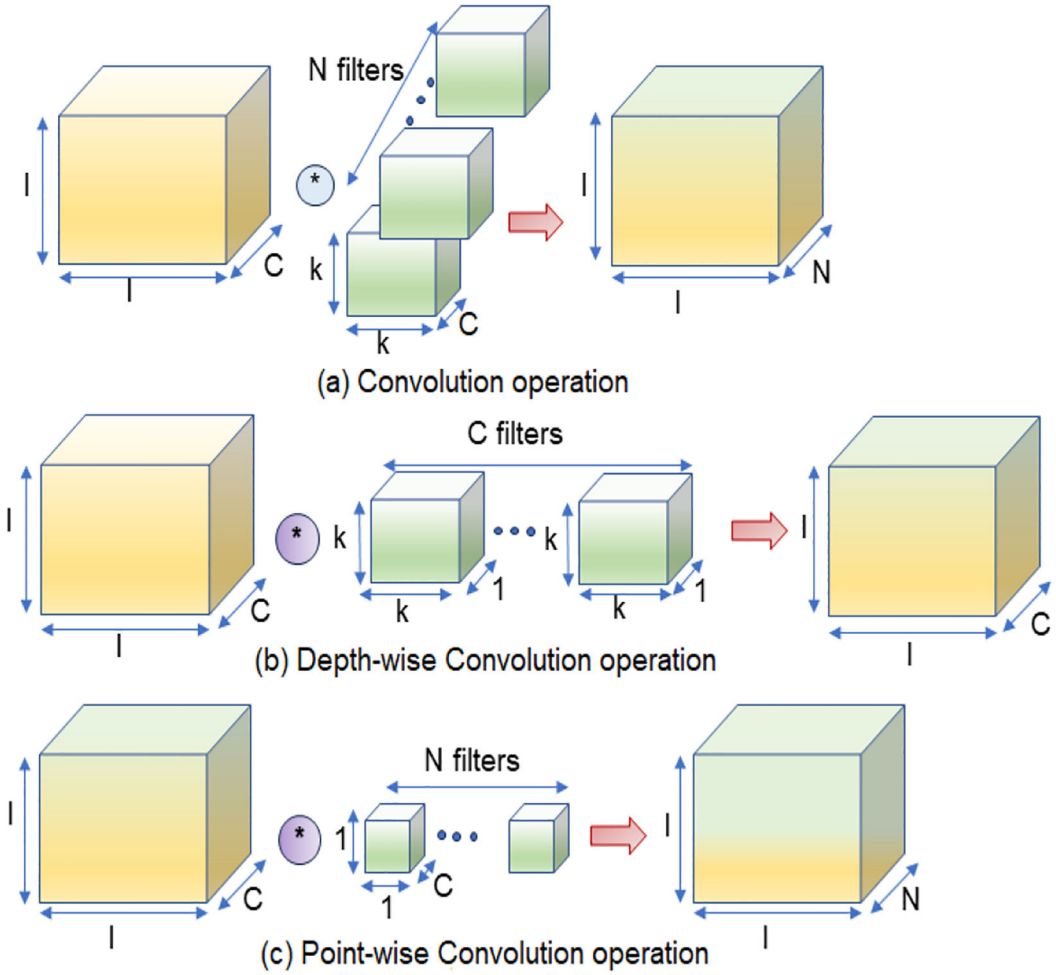

FIGURE 3.

Illustration of convolution and depth-wise separable convolution operations: (a) Standard convolution operation between $I \times I \times C$ input and N filters/kernels each with a size of $k \times k \times C$, (b) Depth-wise convolution operation for each channel at a time with C filters/kernels each with a size of $k \times k \times 1$, (c) Point-wise convolution operation for all C channels at a time using N filters each with a size of $1 \times 1 \times C$.

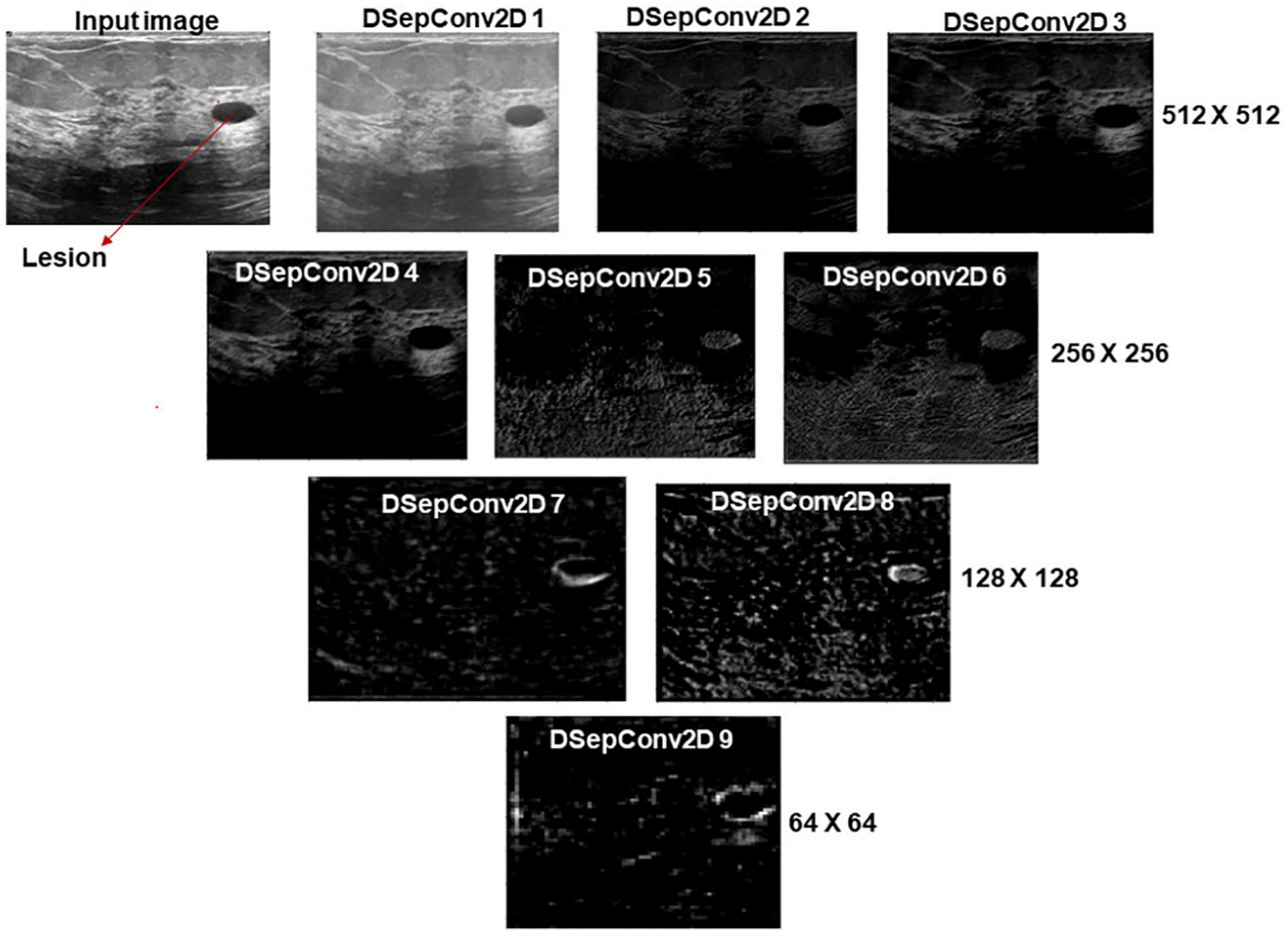

FIGURE 4.

Encoding layer output feature maps for a test input image in *BUSI* dataset. For illustration, single kernel/filter maps are presented. Note the dimensions at each level is halved by the maxpooling operation as shown in Figure 2.

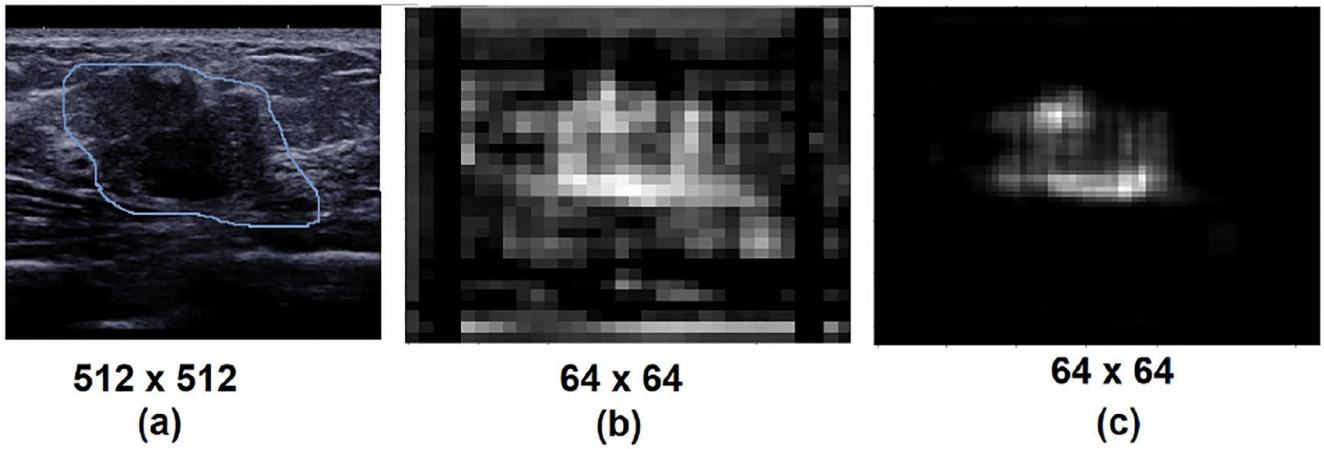

FIGURE 5.

Illustration of feature maps of DMAC-I module for a test image in *ICR* dataset: (a) A test B-mode image with the selected lesion area, (b) Feature map of the dense layer in DMAC-I, (c) Feature map of the multiply layer in DMAC-I.

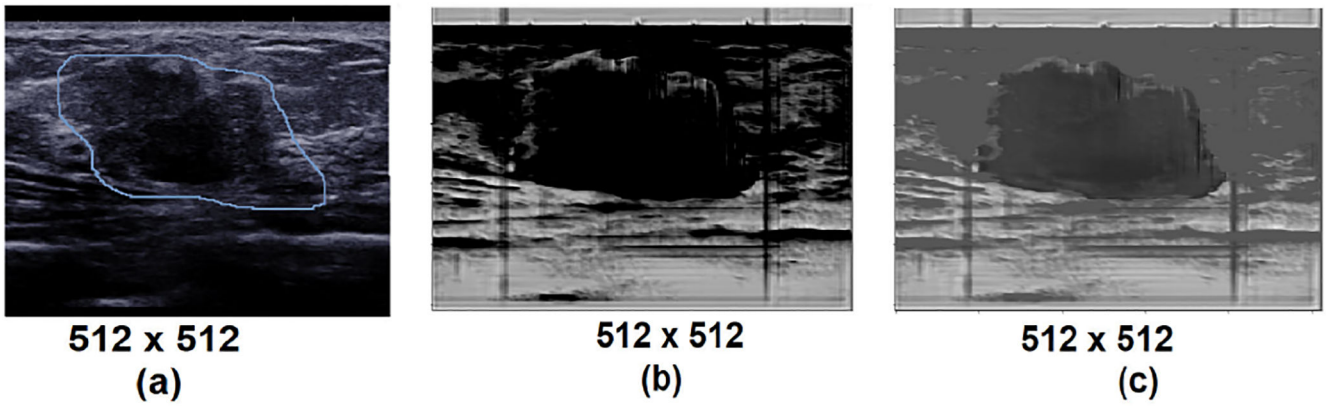

FIGURE 6.

Illustration of feature maps of DMAC-II module for a test image in *ICR* dataset: (a) A test B-mode image with the selected lesion area, (b) Feature map of the dense layer in DMAC-II, (c) Feature map of the multiply layer in DMAC-II.

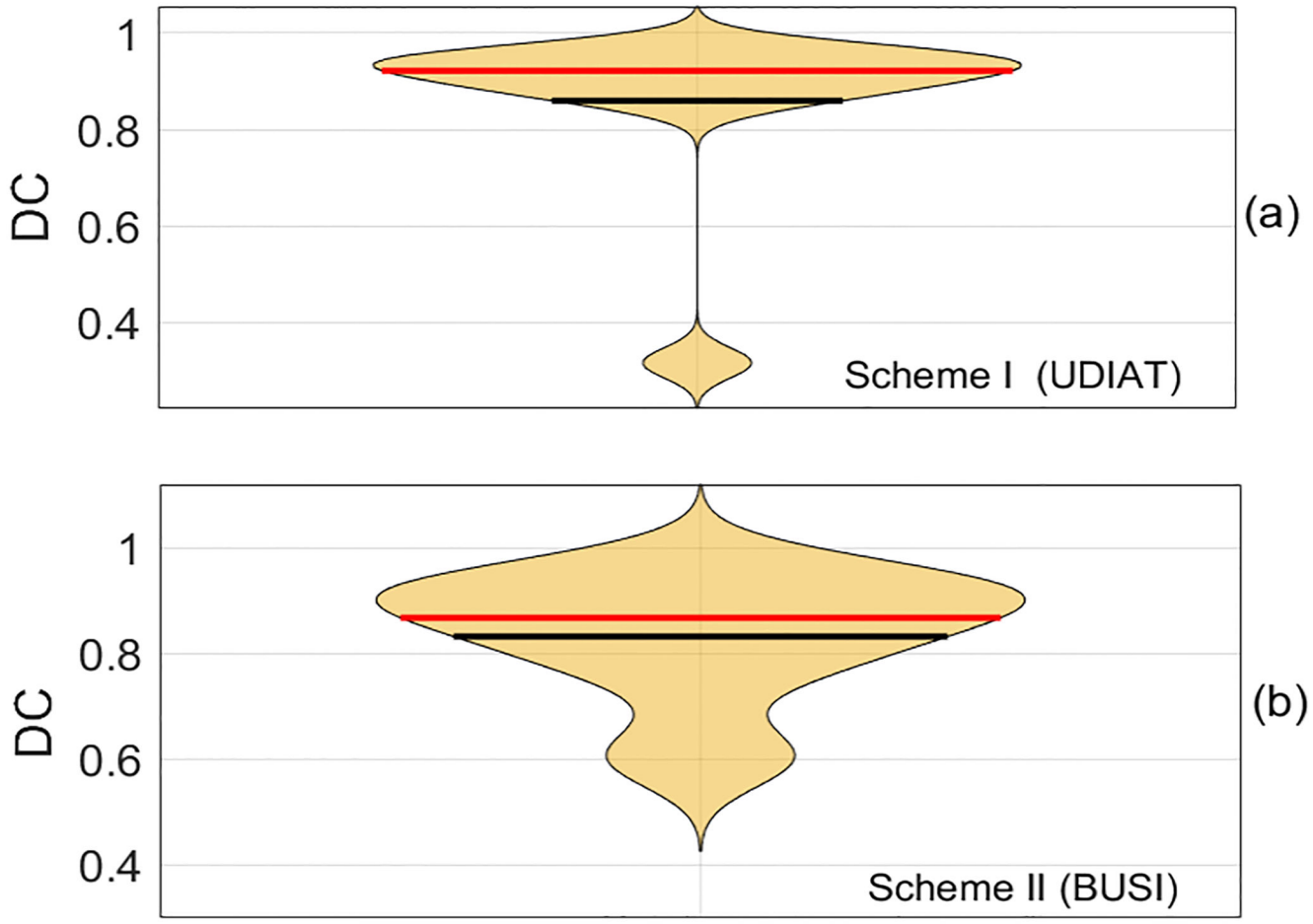

FIGURE 7.
Probability distribution analysis of dice coefficients for both datasets.

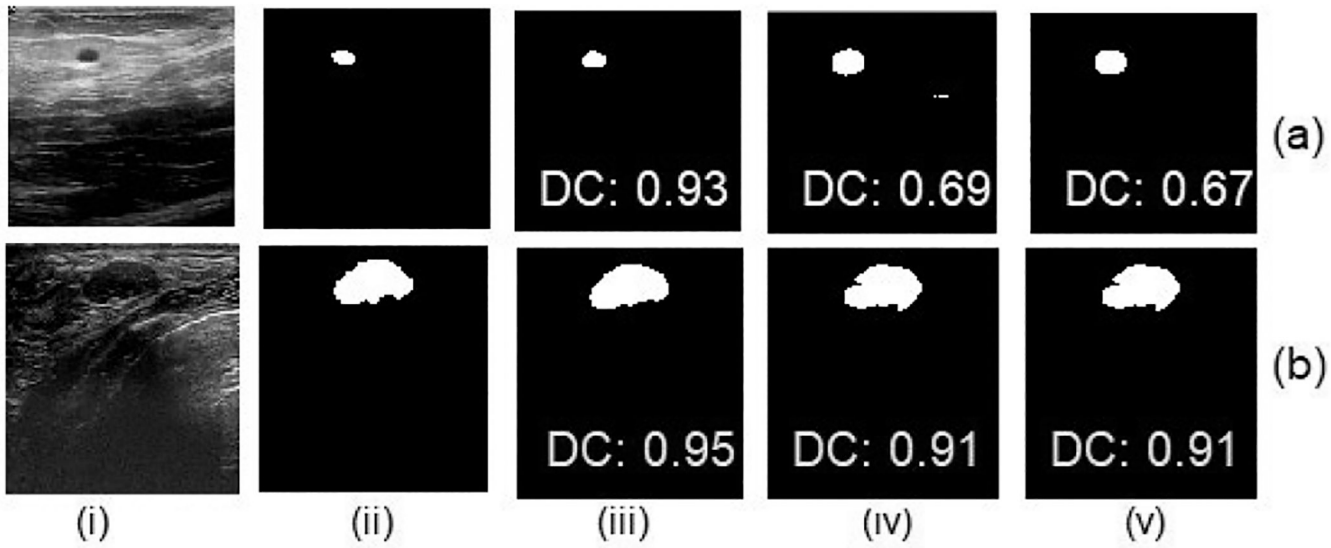

FIGURE 8.

Illustration of actual and network-predicted segmented masks using different schemes in test datasets (a) *BUSI*: (i) A test B-mode image (ii) Ground truth mask, (iii), (iv), and (v) Predicted masks using schemes II, III, and IV respectively, and (b) *UDIAT*: (i) A test B-mode image (ii) Ground truth mask, (iii), (iv), and (v) Predicted masks using schemes I, III, and V respectively.

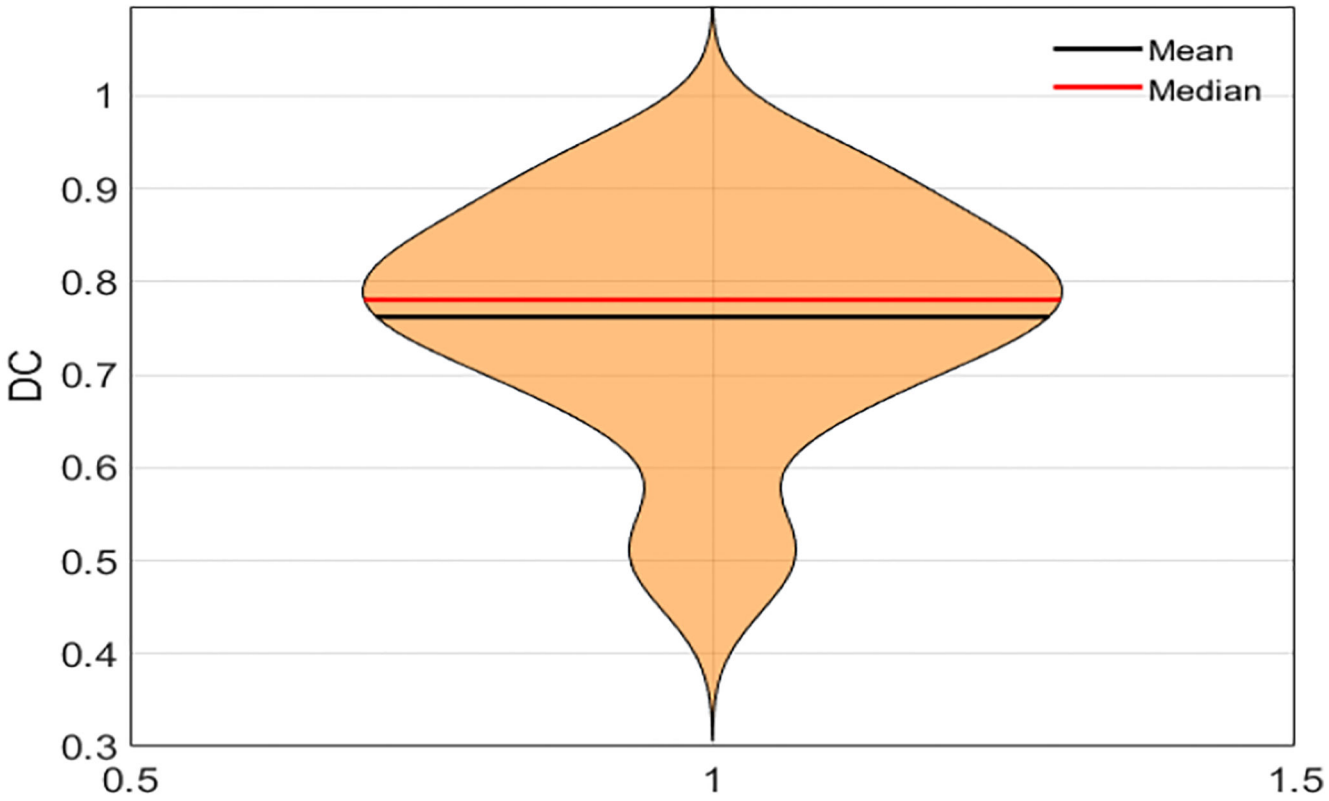

FIGURE 9.
Probability distribution analysis of dice coefficients for *ICR* dataset.

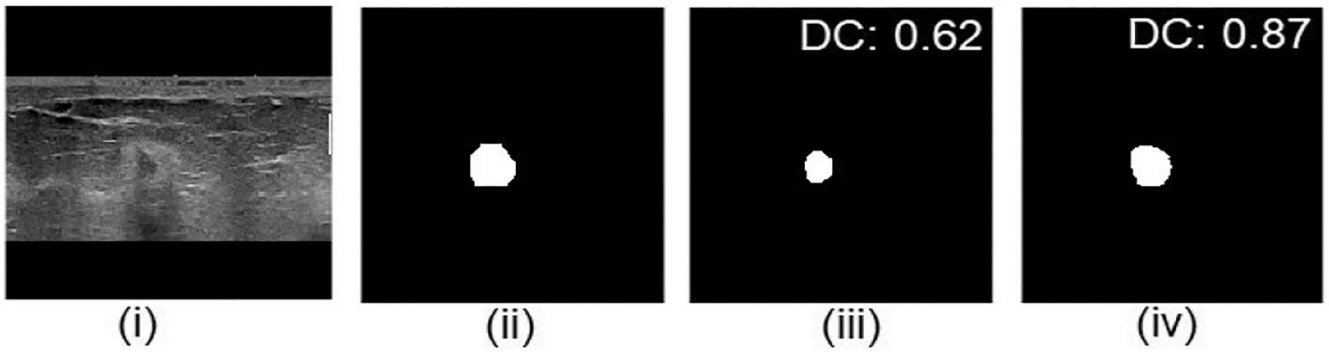

FIGURE 10.

Illustration of actual and predicted segmented masks using network ablation for scheme I in test dataset *ICR*: (i) A test B-mode image in *ICR* dataset, (ii) Ground truth mask, (iii), and (iv) Predicted masks using the proposed network without and with DMAC respectively.

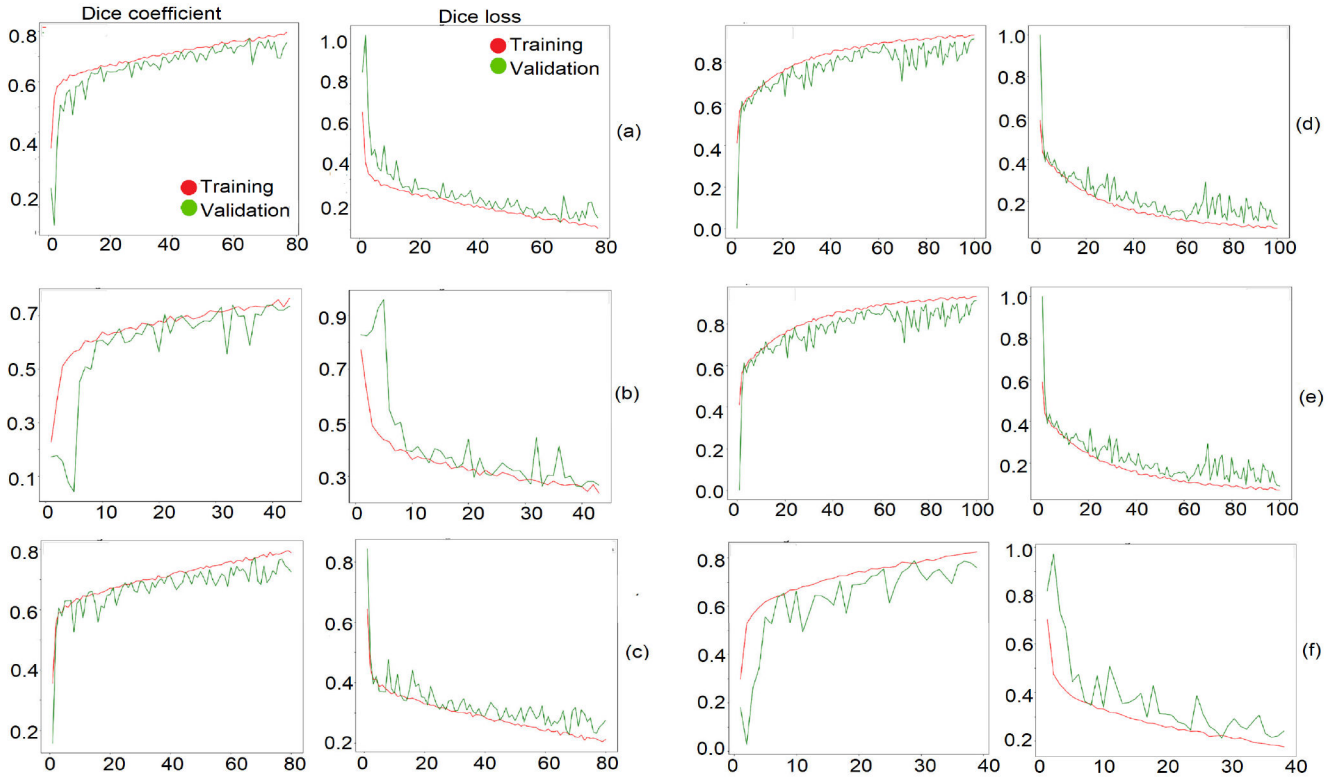

FIGURE 11.
Training curves for all schemes I-VI in (a)-(f) respectively.

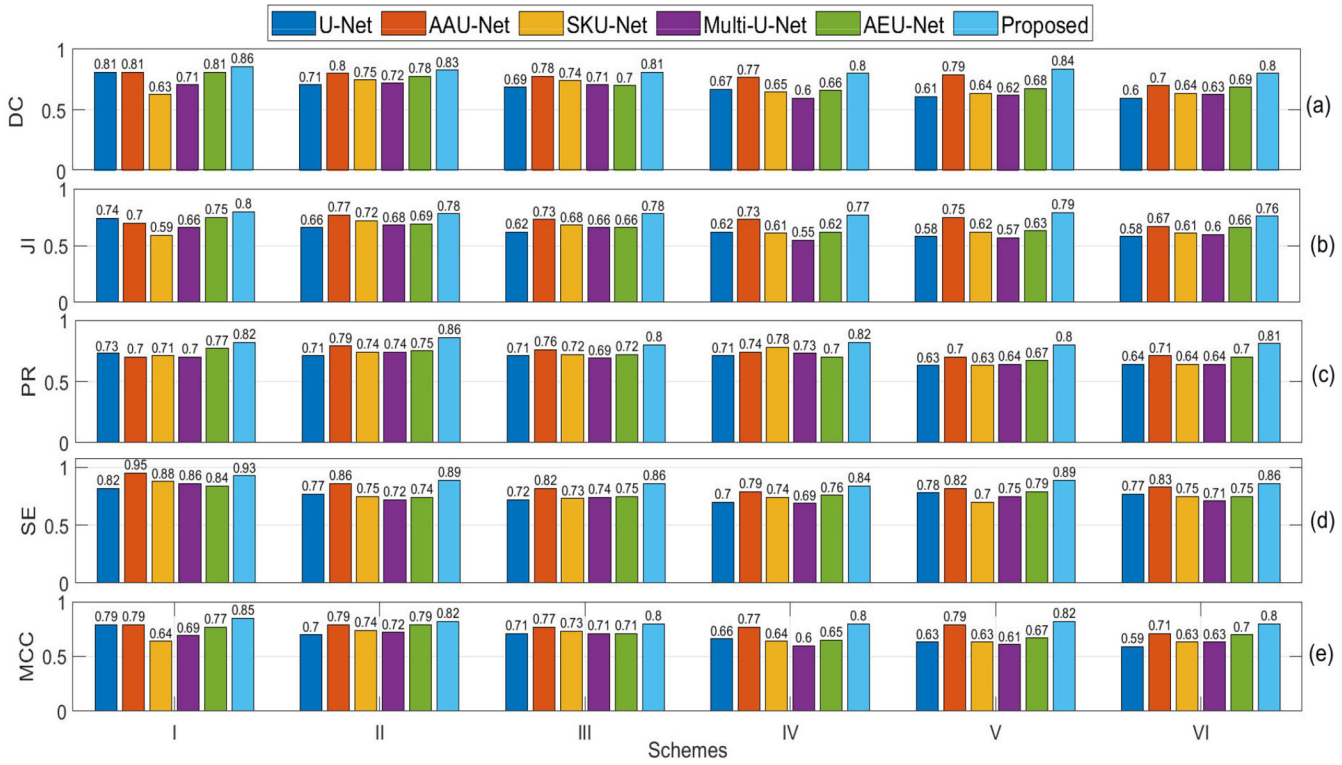

FIGURE 12.

Different segmentation evaluation metrics obtained by implementing the existing networks and proposed network for all schemes: (a) Dice coefficient, (b) Jaccard index, (c) Precision, (d) Sensitivity, and (e) Matthews correlation coefficient.

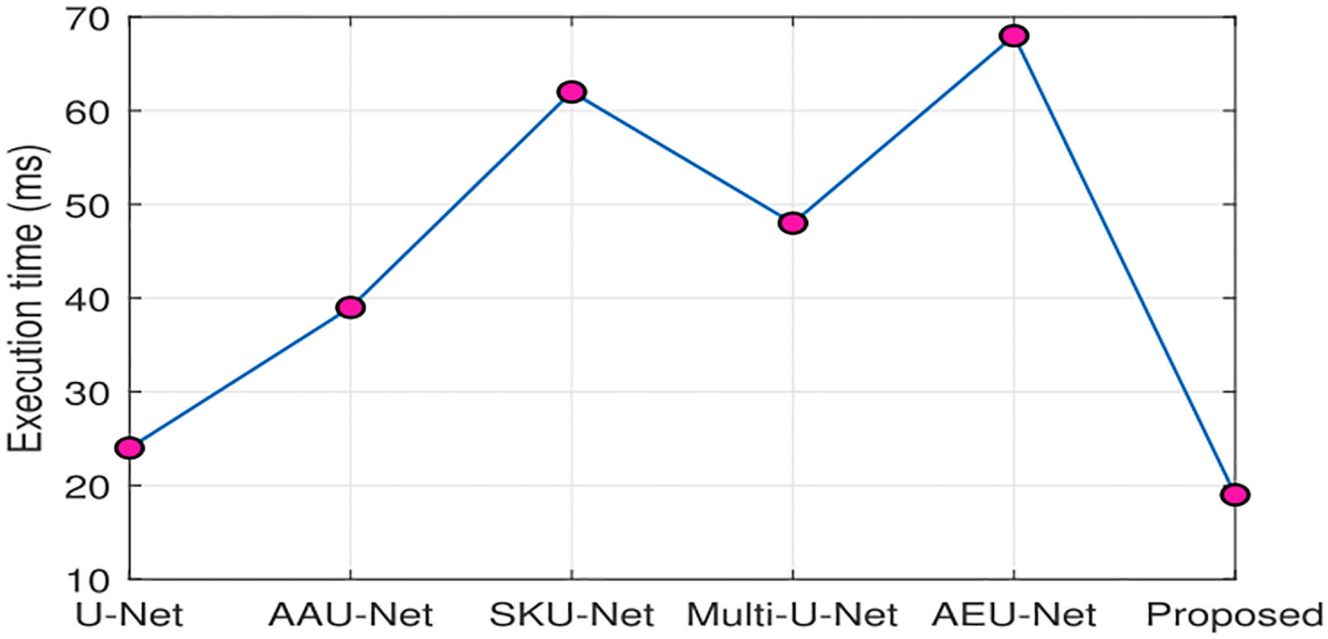

FIGURE 13.
Illustration of the test prediction times per image (average over all schemes) for existing and proposed networks.

TABLE 1.

Training hyper-parameters.

D_r	k_s	sr	lg_r	B_{sz}	e_i	s_e
0.5	3 × 3	1	0.0001	2, 4	50–100	250–500

D_i: Dropout rate, k_S: Kernel size, sr: Stride, lg_I: Learning rate, B_{SZ}: Batch size, e_I: Epochs, s_E: Steps per epoch.

TABLE 2.

Segmentation performance metrics of the proposed DMAeEDNet for all schemes (Mean over patients (Standard deviation)).

Scheme	Batch size	*DC*	*JI*	*PR*	*SE*	*FPR*	*FNR*	*MCC*
I	2	0.85 (0.09)	0.79 (0.03)	0.81 (0.06)	0.91 (0.02)	0.001 (0.001)	0.003 (0.001)	0.83 (0.01)
	4	0.86 (0.10)	0.80 (0.05)	0.82 (0.10)	0.93 (0.03)	0.001 (0.001)	0.002 (0.001)	0.85 (0.03)
II	2	0.82 (0.10)	0.76 (0.05)	0.79 (0.12)	0.88 (0.05)	0.05 (0.02)	0.03 (0.02)	0.80 (0.05)
	4	0.83 (0.08)	0.78 (0.04)	0.86 (0.12)	0.89 (0.02)	0.03 (0.01)	0.02 (0.02)	0.82 (0.03)
III	2	0.80 (0.09)	0.75 (0.06)	0.79 (0.10)	0.85 (0.03)	0.04 (0.01)	0.01 (0.01)	0.78 (0.03)
	4	0.81 (0.11)	0.78 (0.05)	0.80 (0.12)	0.86 (0.04)	0.02 (0.01)	0.008 (0.006)	0.80 (0.02)
IV	2	0.79 (0.10)	0.74 (0.03)	0.77 (0.11)	0.84 (0.04)	0.05 (0.01)	0.05 (0.03)	0.78 (0.05)
	4	0.80 (0.11)	0.77 (0.04)	0.82 (0.12)	0.84 (0.05)	0.04 (0.02)	0.02 (0.02)	0.80 (0.05)
V	2	0.83 (0.12)	0.78 (0.05)	0.79 (0.11)	0.88 (0.03)	0.06 (0.01)	0.04 (0.008)	0.80 (0.05)
	4	0.84 (0.10)	0.79 (0.04)	0.80 (0.13)	0.89 (0.03)	0.02 (0.01)	0.04 (0.05)	0.82 (0.03)
VI	2	0.78 (0.12)	0.74 (0.05)	0.79 (0.14)	0.84 (0.05)	0.06 (0.05)	0.05 (0.05)	0.78 (0.05)
	4	0.80 (0.10)	0.76 (0.06)	0.81 (0.11)	0.86 (0.05)	0.06 (0.05)	0.03 (0.03)	0.80 (0.03)

TABLE 3.

Comparison of the proposed DMAeEDNet with the existing methods for breast lesions segmentation [9] in *UDIAT*, *BUSI* datasets.

Method	**Year**	***DC***	***PR***	***JI***	***SE***	**FLOPs**	**Parameters**
U-Net [30]	2015	0.80, 0.71	0.82, 0.65	0.68, 0.55	0.79, 0.78	318.46	31.03
SegNet [55]	2017	0.79, 0.75	0.78, 0.73	0.66, 0.60	0.82, 0.77	340.11	41.84
PSPNet [56]	2017	0.80, 0.75	0.83, 0.73	0.68, 0.60	0.79, 0.78	354.69	49.07
Attention-U-Net [57]	2018	0.81, 0.71	0.79, 0.69	0.68, 0.55	0.84, 0.74	327.55	35.58
DeepLabv3+ [58]	2018	0.78, 0.77	0.78, 0.76	0.64, 0.62	60.78, 0.78	338.50	41.04
STAN [59]	2020	0.78, 0.75	-	0.70, 0.66	0.80, 0.76	-	-
ESTAN [60]	2020	0.82, 0.78	-	0.74, 0.70	0.84, 0.80	-	-
SAC [61]	2021	0.83, 0.79	-	0.78, 0.72	0.88, 0.83	-	-
MSSA-Net [62]	2021	0.84, 0.81	-	0.76, 0.72	0.86, 0.81	-	71.53
MSF-GAN [63]	2021	0.82, 0.80	-	0.73, 0.71	0.85, 0.78	-	-
MTL-COSA [64]	2022	0.84, 0.79	-	0.76, 0.71	0.87, 0.79	-	109.24
RMTL-Net [64]	2023	0.86, 0.80	-	0.78, 0.72	0.90, 0.82	-	93.51
BGRA-GSA [9]	2023	0.88, 0.81	0.86, 0.79	0.79, 0.69	0.90, 0.84	334.76	101.34
DMAeEDNet (Proposed)	-	0.86, **0.83**	0.82, **0.86**	**0.80, 0.78**	**0.93, 0.89**	**123.01**	**4.19**

PSPNet: Pyramid scene parsing network, STAN: Small tumor-aware network, ESTAN: Enhanced STAN, SAC: Shape-adaptive convolutional network, MSSA-Net: Multi-scale self-attention network, MSF-GAN: Multi-scale Fuzzy generative adversarial network, MTL-COSA: Multi-task learning with context-oriented self-attention, RMTL-Net: Regional-attentive multi-task learning network, BGRA-GSA: Boundary-guided and region-aware network with global scale-adaptive.

Note: FLOPs and parameters are in the order of $\times 10^6$.