EU-Net: Enhanced U-shaped Network for Breast Mass Segmentation

Chowdary, J., & Yogarajah, P. (2023). EU-Net: Enhanced U-shaped Network for Breast Mass Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 1-11. https://doi.org/10.1109/JBHI.2023.3266740

EU-Net: Enhanced U-shaped Network for Breast Mass Segmentation

G. Jignesh Chowdary, and Pratheepan Yoagarajah *Member, IEEE*.

***Abstract*— Segmentation of breast masses in digital mammograms is very challenging due to its complexity. The recent U-shaped encoder-decoder networks achieved remarkable performance in medical image segmentation. However, these networks have some limitations: a) The multi-scale context information is required to accurately segment mass but is not effectively extracted and utilized. b) The global context information is often ignored by the skip connection. To overcome these limitations and achieve better segmentation, we propose an Enhanced U-shaped Network (EU-Net). The proposed EU-Net comprises of 3 novel components: 1) dense-block, which is employed in the encoder and the decoder in place of convolutional layers to achieve the multi-scale features. 2) Multi-Scale Feature Extraction and Fusion, which is used in the junction between the encoder and the decoder for further extracting and fusing the multi-scale context information. 3) Skip Connection Reconstruction, which is inserted between the encoder and the decoder at each stage, to redesign the skip connection and emphasize the global context information. Extensive experimental results under different settings show that the proposed EU-Net achieves superior performances than the previous state-of-the-art segmentation models, and other existing approaches on IN-Breast and CBIS-DDSM mammogram datasets. The generalization ability of the proposed EU-Net is evidenced through cross-dataset and ternary dataset evaluation performance. In the ternary dataset evaluation, the model is trained and evaluated on the UDIAT breast ultrasound dataset without fine-tuning. The EU-Net achieves higher generalization performance in both evaluation experiments. These experiments collectively indicate the efficiency and high generalization ability of the proposed EU-Net.**

***Index Terms*— Atrous convolutions, U-shaped networks, Breast mass segmentation, Digital Mammogram, Ultrasound, Medical image segmentation.**

I. Introduction

BREAST cancer has the highest incidence rate among different types of cancers worldwide; in 2020, approximately 2.3 million new cases are reported, among which 685,000 women died from the disease [1]. Early screening can help to detect breast lesions at early stages, thereby enhancing the patient's survival chances and helping them recover faster [1]. Among the different screening facilities available like X-ray, ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), mammography is considered to be the best choice for the detection of breast cancer because they are safer and fast [2]. The accurate detection of breast masses in a mammogram is very challenging due to its complex characteristics and is also time-consuming. According to [3], the average misdetection rate among medical professionals is around 30%. In a recent survey [4], it was found that misdetection of breast cancer from mammograms is the main cause of legal suits against medical professionals. To overcome the above mention misdetection, computer-aided detection (CAD) systems have been developed by researchers to reduce false detection.

The CAD systems can detect several tissue abnormalities, including masses, bilateral asymmetries, structural distortions, and microcalcifications from mammograms [5]. Among these abnormalities, masses are crucial for diagnosing breast malignancies [6]. Thus accurate segmentation of masses is very important in a mammary CAD system. But, the irregularities in shape, size, boundary, and location with poor contrast make mass segmentation a complex task [7]. Many studies employed active contour, region growing, Markov random field, and Chan-Vese methods for mass segmentation. Since these methods rely on the prior knowledge of mass contour [7], [8], they cannot handle complex shape variations and masses surrounded by tissues with different densities [7], resulting in poor diagnostic performance.

With the rapid developments in deep learning, several breast mass segmentation approaches based on deep learning have been developed [9]–[12]. Singh et al. [13] designed an adversarial network, where the generator produces the initial breast tumor segmentation results. The discriminator learns to differentiate the produced result from the ground truth and forces the initial result to be close to the ground truth. Li et al. [14] proposed a dual-path network for segmenting the breast masses. One path hierarchically extracts features from the input image, and the other path focuses on yielding geometrical features of the image. With remarkable performance achieved by the U-shaped encoder-decoder networks in medical image segmentation [15] several studies employed such U-shaped networks for mass segmentation. For instance, Revitha et al. [16] developed an encoder-decoder network with deep supervision for segmenting masses from mammograms. And they also employed the conditional random fields method as a post-processing step to reduce the false negatives. Sun et al. [17] designed a segmentation approach with an attention-guided upsampling module to integrate low-level and high-level features to produce better segmentation results. However,

G.J. Chowdary was an PhD student at Stony Brook University, New York, USA. (e-mail: jigneshchowdary@gmail.com).

Pratheepan Yogarajah was a Lecturer in Computer Science at Ulster University, Northern Ireland, UK. (e-mail: p.yogarajah@ulster.ac.uk).

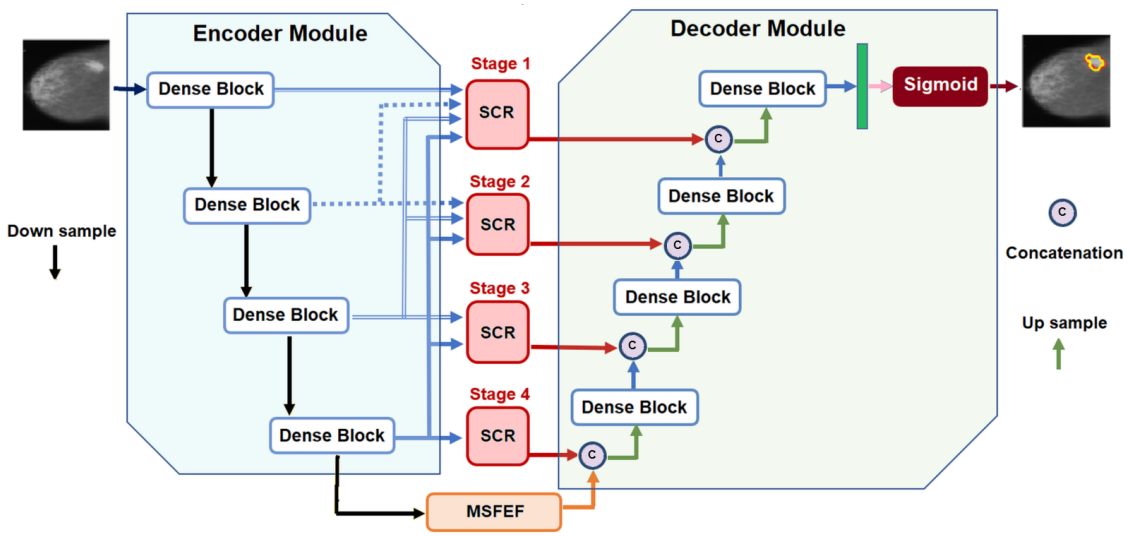

Fig. 1. Structure of the proposed EU-Net.

the U-shaped based methods have some limitations. Firstly, the multi-scale context information is not effectively extracted and utilized. This information is essential when dealing with targets with complex structures (breast masses) so that the structure's (masses) surroundings can be considered, and ambiguous decisions can be avoided [18]. Secondly, the skip connection at each stage of the U-shaped based methods [16], [17] ignores the global context information and is an indiscriminate combination of local information that introduces unnecessary clutter and causes pixel misclassification.

Therefore, this work proposes an Enhanced U-shaped Network (EU-Net) that comprises of Multi-Scale Feature Extraction and Fusion (MSFEF) and Skip Connection Reconstruction (SCR) modules to overcome these two limitations. The illustrated version of the EU-Net is shown in Fig 1. For extracting the multi-scale context information, the convolution stack used in the standard U-shape network is replaced with dense-block to extract the multi-scale feature at each stage in our proposed EU-Net. After the multi-scaled feature extraction from the last layer of the dense-block, the features are then passed to the MSFEF module. The MSFEF module is embedded in the junction between the encoder and the decoder to effectively capture and fuse multi-scale context information. The MSFEF module can then dynamically select appropriate receptive field for masses through self-learning and fuse multi-scale context information more effectively. To solve the second problem, the SCR module merges the multi-stage global information for reconstructing the skip connection. The SCR module is placed at each stage of the network to fuse the global context information extracted by the deeper layers of the network with the local information extracted at the same stage. The SCR module also guides global information flow to the decoder. The main contributions of our work are summarized below:

- We propose a novel EU-Net that includes SCR and MSFEF for segmenting breast masses efficiently in mammogram and ultrasound imaging modalities.
- The MSFEF module, and the dense-block are proposed to effectively exploit the multi-scale context information.
- The SCR module promotes global context information flow to the decoder through skip connection.
- The extensive experimental results under different settings indicate that our EU-Net not only achieves higher segmentation performance than the state-of-the-art (SOTA) models and other existing approaches on INBreast [19] and CBIS-DDSM [20] mammogram datasets, but also generalizes well on UDIAT [21] ultrasound dataset.

The rest of the article is structured as follows: Section II illustrates the proposed methodology, Section III presents the results and discussion, and finally, Section IV concludes the proposed work.

II. Proposed Methodology

The proposed EU-Net consists of four main components: Encoder module, SCR module, MSFEF module, and Decoder module. The encoder, and the decoder modules consists of dense-block for capturing multi-scale features. These four main components are explained below:

A. Encoder

With inspiration from the densenet [22], the proposed encoder module is designed to: (a) extract more efficient features from the input mammograms, (b) avoid the problem of vanishing-gradients, and (c) reduce the number of training parameters and over-fitting. Therefore, to achieve these, we propose to use dense-blocks instead of convolutional blocks at each stage of the standard UNet. The major difference between the standard convolution block in UNet and the dense-block is that the convolutional layers in the standard convolution block take feature maps as input from the previous one, whereas in dense-block the input to any layer is the concatenated output of the preceding layers enabling feature-reuse which is more advantageous. The proposed dense-block consists of 5 convolutional layers as shown in Fig. 2, where the first 4 convolutional layers use 3×3 kernels for extracting

features layer-by-layer. And we use a 1×1 kernel in the last convolutional layer to restrict the number of output feature maps from the dense-block to M, where M represents the number of input, and output features of the dense-block. After each convolutional layer, we use a Batch normalization and Rectified Linear Unit (ReLU) layers for simplifying the training process. The five convolutional layers in the dense-block take M, 2M, 3M, 4M, and 5M feature maps as input, and the output of the dense-block is M. In our case, in the first dense block, the input is a mammogram image with one channel, thus M = 1. As shown in Fig. 1, the encoder module contains 4 dense-blocks. Instead of using a pooling operation to downsample the feature map, we applied a stride of 2 to the first convolutional layer of each dense-block in the encoder module to reduce the feature map by half.

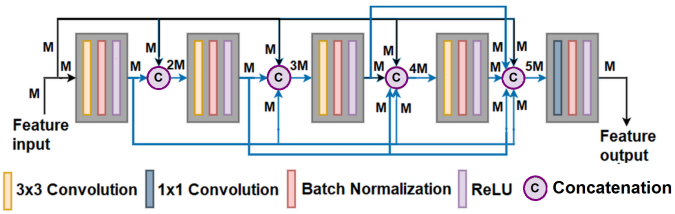

Fig. 2. Structure of the proposed dense-block.

B. SCR Module

The encoder extracts global context information from the input mammogram images, including the surrounding objects. However, such information is progressively diluted when gradually transmitted to the shallow layers [23]. Furthermore, the skip connection between the encoder and the decoder is prone to introduce unnecessarily clutters and have segmentation gaps due to the inconsistency of receptive fields. Therefore, to address this problem, we design a SCR module (see Fig. 1). In the SCR module, the skip connections are redesigned by merging the feature maps of the deeper stages with the same stage. For instance, Fig. 3 presents the SCR module of the first stage.

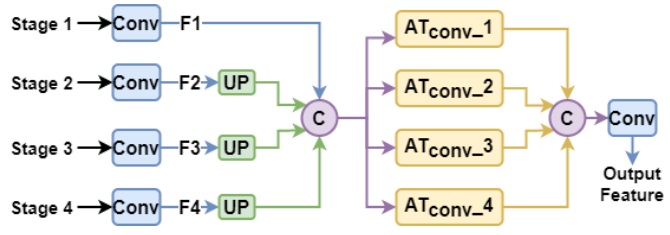

Fig. 3. Structure of the proposed SCR module for stage 1.

This module works by employing 3×3 convolution to map all stages' feature maps into the same channel space as stage 1. Then the generated feature maps F2, F3, and F4 are upsampled to the size of F1 for concatenation. For extracting the global context information from various levels of feature maps, four dilated convolutions (AT_{conv}_1, AT_{conv}_2, AT_{conv}_3, AT_{conv}_4) [24] with various atrous rates (2, 4, 8, 16) are employed in a parallel fashion. Finally, a convolution operation is applied to obtain the final feature map. It has to be noted that the number of parallel dilated convolutions and the atrous rates vary depending on the number of stages to be fused. The SCR module at any stage can be formulated as follows:

$$SCRO_s = ©_{j=s}^{j=4}(AT_{conv}\_2^{j-s}(©_{j=s}^{j=4}(\mathbb{F}_s U 2^{j-s}))) \quad (1)$$

In Equation 1, © represents the concatenation operation, $\mathbb{F}_s$ represents the feature map from encoder at s^{th} stage, $AT_{conv}\_2^{j-s}$ represents the atrous convolution with dilation rate of 2^{j-s}, $U2^{j-s}$ represents the upsampling operation with rate of 2^{j-s}, and $SCRO_s$ represent the ouput of SCR module at s^{th} stage. In this work, the proposed EU-Net employs only four SCR modules to reduce computational costs. By inserting multiple SCR modules between the encoder and the Decoder, the global context information from deep stages can be guided to various stages.

C. MSFEF module

The multi-scale context information enhances the segmentation performance. However, effectively integrating such information is difficult, but it is worth exploring. Therefore, we design a MSFEF module to exploit multi-scale information. This module is shown in Fig. 4.

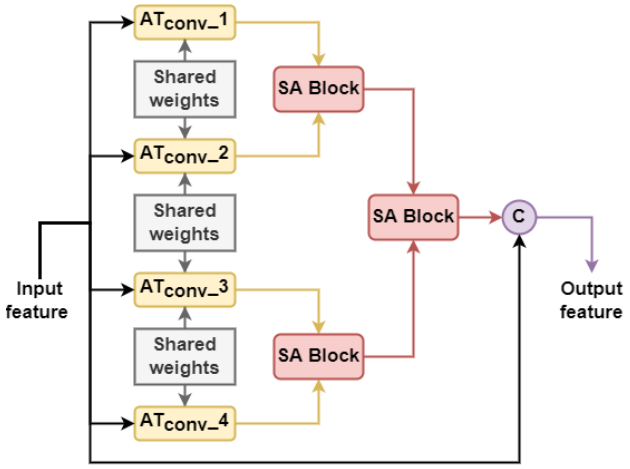

Fig. 4. Structure of the proposed MSFEF module.

The MSFEF module consists of 4 parallel atrous convolutions with dilation rates of 1, 2, 3, and 4 for capturing multi-scale context information, followed by 3 cascaded Scale Attention (SA) blocks employing the spatial-attention mechanism. These atrous convolutions have shared weights to achieve faster convergence, and reduce the number of parameters. Then we used SA blocks presented in Fig. 5 for fusing the multi-scale information. These SA blocks employ spatial attention mechanisms, as presented in Fig. 5, for dynamically choosing the appropriate multi-scale features and fusing them through self-learning. When two features, F_S and F_Q, with different scales are passed to the SA block, a series of convolutions are applied and features maps S, and $Q \in \mathbb{R}^{H \times W}$ are obtained, where H, W represents the height, and width of the feature

map. Secondly, a softmax operator is used to construct pixel-wise attention maps A_S, A_Q. Then element-wise product operations are performed between the multi-scale features maps (F_S and F_Q) and their attention maps (A_S and A_Q). The final fused feature map (F_{fused}) is obtained by adding the resulting products. As shown in Fig. 4, we employ three SA blocks to obtain the fusion feature of the four parallel atrous branches. Finally, the output of the MSFEF module is obtained by concatenating the fusion feature with the input feature map through an auxiliary connection.

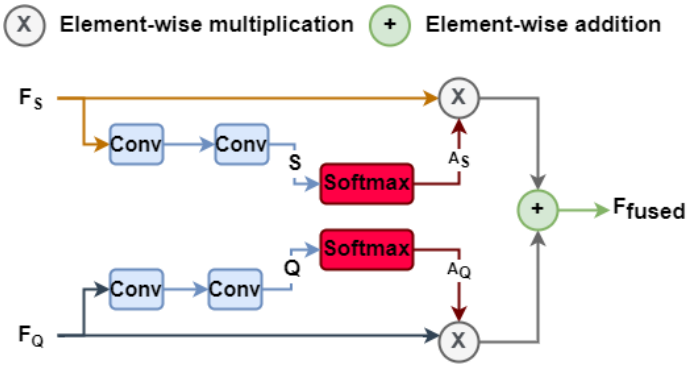

Fig. 5. Structure of the proposed SA block.

D. Decoder

As seen in Fig. 1, the decoder module consists of 4 dense-blocks. At each stage, the global context information guided by the SCR module is fused with the upsampled feature maps from the succeeding stage. In this work, we used deconvolution instead of upscaling. Because the deconvolution employs a convolution operation to enlarge the image, which could learn a self-adaptive mapping for restoring feature maps with more detailed information. At the end of the last dense-block in the decoder module, a 1×1 convolution layer and a sigmoid layer are employed to project the multi-channel feature-maps into desired segmentation mask.

E. Loss function

The proposed segmentation model is trained by minimizing the joint loss function ($Joint_{Loss}$) consisting of dice loss and the binary cross-entropy. This joint loss function is presented as $Joint_{Loss} = L_{BCE} + L_{DICE}$ where, L_{BCE}, and L_{DICE} are the binary cross-entropy, and dice loss. These loss functions are formulated below:

$$L_{BCE} = -\frac{1}{N}\sum_{j=1}^{N}[x_j log(y_j) + (1 - x_j) log(1 - y_j)] \quad (2)$$

$$L_{Dice} = 1 - \frac{2\sum_{j=1}^{N} y_j x_j}{\sum_{j=1}^{N} y_j + \sum_{j=1}^{N} x_j} \quad (3)$$

In Equation (2)-(3), the $x_j \in [0, 1]$ represent the ground-truth label of the j^{th} pixel, where 1 means that pixel belongs to the breast mass and 0 means that pixel belongs to the background. The $y_j \in [0, 1]$ represent the class predicted by the proposed EU-Net for the j^{th} pixel. And N represents the total number of pixels in an image.

F. Implementation details

Our network is programmed in Python using the PyTorch library. All the experiments are conducted using NVIDIA SCRU with 11 GB of on-chip memory. For optimization, the Adam [25] optimizer is used. The initial learning rate is set to 0.00001, which is further decreased by 1/25 for every 20 epochs. The proposed EU-Net is trained by 200 epochs, with a batch size of 4. All the input images are resized to 256×256 resolution using inter-area interpolation in our experiments. Since the datasets employed in this work are small, we employed several augmentation techniques, including horizontal flipping, vertical flipping, random rotation, and translation, to avoid the problem of over-fitting.

III. Results and Discussion

A. Datasets

In this work, we employ two publicly available mammogram datasets, namely CBIS-DDSM [20] and INBreast [19], for evaluating our model. The CBIS-DDSM is a subset of the Digital Database for Screening Mammography (DDSM) [20], and it consists of 303 mammograms in DICOM format with corresponding ground truths. The dataset provider has already split the dataset into training and test sets consisting of 242 and 62 images. These images are also categorized based on the Breast-Imaging-Reporting-and-Data-System (BI-RADS) scores. The number of images per BI-RADS score is tabulated in Table I. We also employed the same split for evaluating the proposed EU-Net. For configuring the hyper-parameters, we used 10% of the training set to perform validation experiments.

The INBreast dataset [19] is a Full Field Digital Mammography (FFDM) dataset. This dataset consists of 410 mammograms, among which 303 are normal cases, 72 are malignant cases, and 35 are benign cases. We employed only mammograms with breast masses (benign and malignant cases) and corresponding ground truths in this work. The number of images according to the BI-RADS score is presented in Table I. Since there is no official data-split provided by the dataset providers. We employed the 7:1:2 split (70% for training, 10% for validation, and 20% for testing) for evaluating the proposed EU-Net. In addition to these two mammogram datasets, we employed one ultrasonic image dataset (UDIAT [21]) to understand the generalization ability of the proposed EU-Net. The UDIAT dataset consists of 163 ultrasound images

TABLE I
BI-RADS score in CBIS-DDSM, and INBreast datasets.

Dataset	BI-RADS score	No. of images
CBIS-DDSM [20]	1	91
	2	159
	3	40
	4	13
INBreast [19]	2	23
	3	13
	4	20
	5	43
	6	8

with masses (110 benign, 53 malignant). All these images have an average resolution of 760×570 pixels.

B. Performance metrics

In this work, we employed both area-based and distance-based metrics for evaluating the proposed EU-Net. The area-based metrics, namely recall (REC), intersection over union (IoU), dice coefficient (DC), and pixel accuracy (PA), are used to compare the predicted segmentation result with the ground truths. The metrics are formulated below:

$$REC = \frac{TP_{seg}}{TN_{seg} + FP_{seg}} \tag{4}$$

$$IoU = \frac{TP_{seg}}{TP_{seg} + FP_{seg} + FN_{seg}} \tag{5}$$

$$DC = \frac{2 \times TP_{seg}}{2 \times TP_{seg} + FP_{seg} + FN_{seg}} \tag{6}$$

$$PA = \frac{TP_{seg} + TN_{seg}}{TP_{seg} + TN_{seg} + FP_{seg} + FNseg} \tag{7}$$

In Equations 4-7, the TP_{seg} represents the number of pixels that are correctly classified as mass region, TN_{seg} represents the number of pixels that are correctly classified as background, the FP_{seg}, and FN_{seg} represents the number of pixels that are wrongly classified as mass region and background.

The distance-based metrics are used to evaluate the segmentation result regarding the predicted region's shape accuracy. The distance-based metrics, namely Hausdorff Distance (HD) [26] and Mean Surface Distance (MSD) [27], are used in this work. The metrics are formulated below:

$$MSD = \frac{1}{2}\left(\frac{1}{n_b}\sum_{q \epsilon s_b} \min_{\mathbf{q} \,\epsilon\, \mathbf{s_b}} E_d(q,w) + \frac{1}{n_a}\sum_{w \epsilon s_a} \min_{\mathbf{w} \,\epsilon\, \mathbf{s_a}} E_d(q,w)\right) \tag{8}$$

$$HD = \max\left(\max_{q \,\epsilon\, s_b} \min_{w \,\epsilon\, s_a} E_d(q,w),\ \max_{w \,\epsilon\, s_a} \min_{q \,\epsilon\, s_b} E_d(q,w)\right) \tag{9}$$

where the s_a and s_b represent the number of surface points, $E_d()$ represents the Euclidean distance, n_a and n_b indicate the surfaces of the predicted mask and ground truth.

C. Comparison with State-of-the-art models

We compared the performance achieved by our model with 7 SOTA models namely UNet [28], DeepLabV3+ [29], UNet++ [30], CE-Net [31], FCN [32], Attention U-Net [33], and SENet [34] on the CBIS-DDSM, and INBreast datasets. For a fair comparison, all these models are individually trained and evaluated with the same hyperparameters, loss function, and evaluation protocol on the two datasets.

1) Quantitative Comparison: The quantitative segmentation results achieved by the proposed EU-Net and the SOTA models on the INBreast dataset are tabulated in Table II. Our model achieved superior performance in terms of PA, REC, HD, and MSD over the SOTA models. Our model enhanced the DC and IoU by 12.24% and 4.5% than the CE-Net [31] (second-best performing model). Also, there is an enormous reduction of HD and MSD by 6.82 mm and 3.37 mm than the CE-Net [31]. The UNet [28] achieved the least segmentation performance with DC, IoU, and REC scores being lower than 75%, 70%, and 70% respectively.

TABLE II
QUANTITATIVE RESULTS ON INBREAST DATASET.

Model	HD (mm)	MSD (mm)	PA (%)	REC (%)	IoU (%)	DC (%)
SENet [34]	31.46	7.69	98.23	81.21	71.25	78.86
Attention U-Net [33]	19.52	5.77	98.86	79.19	72.79	82.22
FCN [32]	24.59	6.42	98.64	75.92	73.47	79.28
CE-Net [31]	16.24	4.72	99.15	83.59	80.12	81.81
UNet++ [30]	17.93	5.25	98.94	77.84	72.46	80.14
UNet [28]	32.61	8.26	97.63	69.64	68.43	74.22
DeepLabV3+ [29]	29.57	7.01	98.47	74.81	70.69	76.92
Ours	**9.42**	**1.35**	**99.85**	**86.30**	**84.62**	**94.05**

Table III presents the segmentation results achieved by the models on the CBIS-DDSM dataset. Our model achieved the highest performance in terms of all the performance metrics considered. This accounts for an increase of 0.67% PA, 3.26% REC, 9.07% DC, and 18.29% IoU, and a decrease in HD and MSD by 7.05 mm and 0.78 mm than the CE-Net [31]. The least performance is achieved by UNet [28].

TABLE III
QUANTITATIVE RESULTS ON CBIS-DDSM DATASET.

Model	HD (mm)	MSD (mm)	PA (%)	REC (%)	IoU (%)	DC (%)
SENet [34]	23.89	2.76	98.02	81.71	69.37	75.88
Attention U-Net [33]	19.53	1.57	99.25	87.34	72.39	79.98
FCN [32]	20.74	2.06	99.39	82.03	70.42	78.63
CE-Net [31]	18.52	1.05	99.22	86.39	73.18	85.51
UNet++ [30]	19.12	1.09	99.18	87.12	71.57	77.95
UNet [28]	24.29	2.95	97.05	79.56	68.49	71.63
DeepLabV3+ [29]	21.54	2.51	98.63	83.49	71.42	78.18
Ours	**11.47**	**0.27**	**99.89**	**89.65**	**91.47**	**94.58**

In these 2 experiments our model achieved very high DC values (i.e., 11.83% and 9.07%) compared to other models. The experimental results from Tables II and III indicate that the proposed EU-Net can work better and is more robust and effective than the given 7 SOTA models.

2) Qualitative comparison: The visual segmentation results achieved by the proposed EU-Net and the SOTA models on mammograms from INbreast [19] and CBIS-DDSM [20] datasets are presented in Fig. 6. The red and yellow contours indicate the ground truth and the region predicted by the models respectively. From Fig. 6, we can observe that the proposed EU-Net can segment the breast masses of smaller sizes, whereas the other SOTA models failed. Among the SOTA models, CE-Net, and UNet++ produced better segmentation results, but Attention U-Net, DeepLabV3+ slightly over-segmented the background, and SENet slightly under-segmented the breast lesion. The UNet produced very poor segmentation results by over-segmenting the background and under-segmenting the lesion in several cases. From Fig. 6, we can understand that the proposed EU-Net is not sensitive to glandular tissue and high-density mass compared to SOTA models. These results show that the proposed EU-Net can

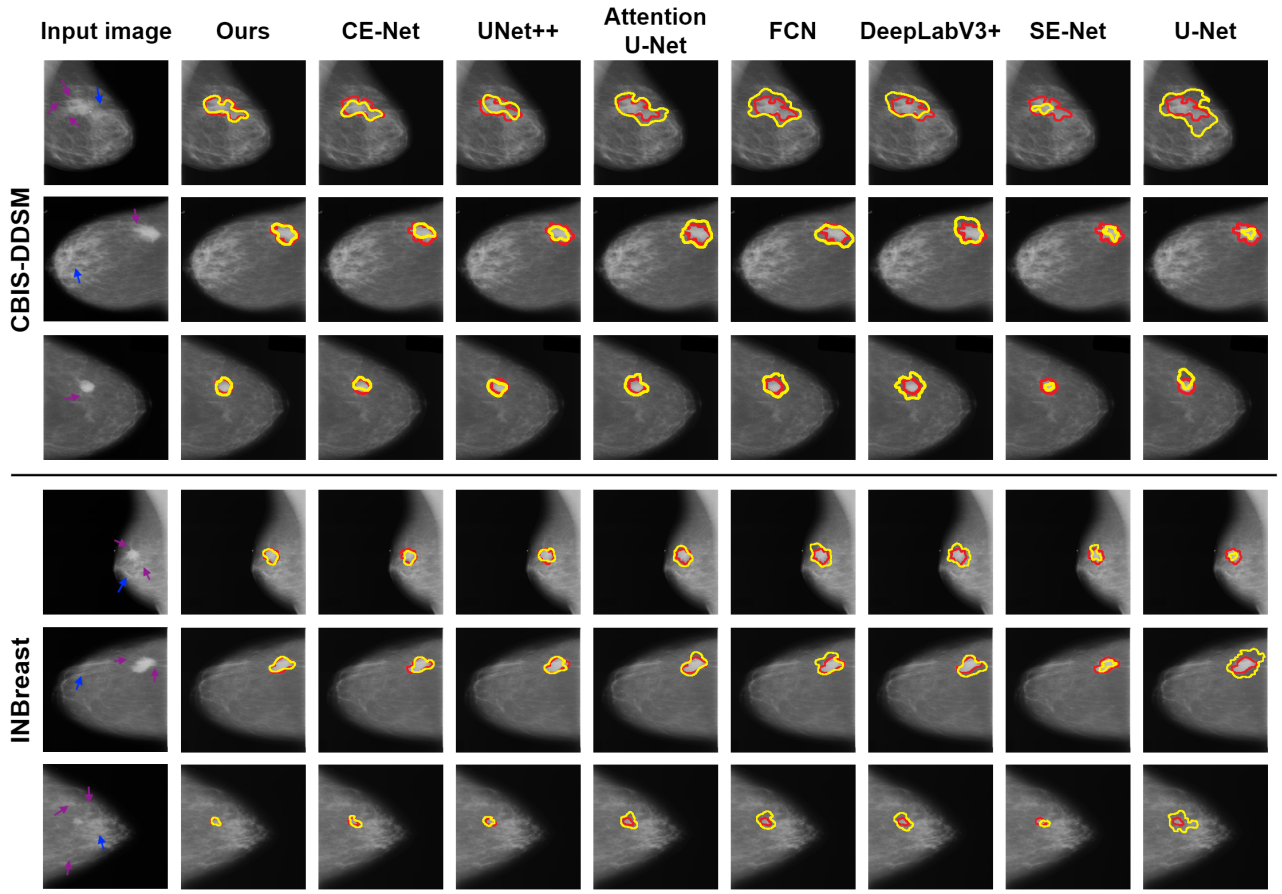

Fig. 6. Qualitative comparison with SOTA models on CBIS-DDSM, and INBreast datasets; the pink, and the blue arrows indicates the glandular tissue, and dense mass in the breast mammograms.

produce high precise segmentation results by adapting to complex characteristics of masses, including varying sizes and fuzzy boundaries.

D. Comparison with existing approaches

We compared the performance achieved by the proposed EU-Net with the existing approaches, namely ARF-Net [35], FS-Unet [36], Li et al. [14], ACMSCNet [37], Connected ResUNets [38], DS U-Net [16], and MNPNet [12] on the INBreast, and CBIS-DDSM datasets. These approaches employed different evaluation protocols: ARF-Net [35], Li et al. [14], and DS U-Net [16] employed the official train/test split provided in the dataset, whereas FS-Unet [36], ACMSCNet [37], Connected ResUNets [38] employed customs data-splits. And all these approaches [14], [16], [35]–[38] used the full mammogram image as input, except the MNPNet [12]. The MNPNet used the mass-centered RoI's as input. Therefore we re-evaluated our model with those evaluation protocols, and input types for a fair comparison.This comparison is presented in Table IV, and Table V. As seen in Table IV, and V, the proposed EU-Net outperformed all the existing approaches in their respective evaluation protocols. As recall be considered to be the most important metric in medical image segmentation, our model phenomenally enhanced the recall on both the datasets. This shows the effectiveness of the proposed EU-Net in segmenting breast masses from mammogram images.

E. Ablation study

To evaluate the efficiency of the proposed SCR and MSFEF modules, we performed several ablation experiments on the test set on CBIS-DDSM [20]. For convenience, we term the standard UNet [28] as the baseline model. The results of all the ablation experiments are tabulated in Table VI.

TABLE IV
QUANTITATIVE COMPARISON (%) WITH EXISTING APPROACHES EMPLOYING INBREAST DATASET. THE * REPRESENTS THAT THE MODEL IS TRAINED, AND TESTED USING THE MASS CENTERED ROI'S.

Methods	Evaluation protocol	PA	REC	DC
ARF-Net [35]	5-fold	99.11	83.00	85.06
FS-UNet [36]	5-fold	99.19	79.31	84.57
Li et al. [14]	70:10:20	-	-	93.69
MS-ResCU-Net [39]	3 fold	94.16	93.11	91.78
ACMSCNet [37]	5-fold	99.66	85.91	84.11
Connected ResUNets [38]	70:10:20	93.03	-	89.52
DS U-Net [16]	5-fold	99.00	81.10	78.60
MNPNet [12] *	50:10:40	-	-	91.10
Ours	**5-fold**	**99.73**	**89.49**	**87.63**
	70:10:20	**99.85**	**86.30**	**94.05**
	3-fold	**96.29**	**93.62**	**93.58**
	50:10:40	**93.57**	**92.83**	**94.31**

TABLE V
QUANTITATIVE COMPARISON (%) WITH EXISTING APPROACHES ON CBIS-DDSM DATASET. THE * REPRESENTS THAT THE MODEL IS TRAINED, AND TESTED USING THE MASS CENTERED ROI'S.

Methods	Evaluation protocol	PA	REC	DC
ARF-Net [35]	Official split	99.80	88.91	85.75
FS-Unet [36]	80:10:10	99.80	85.47	84.19
Li et al. [14]	Official split	-	-	92.27
ACMSCNet [37]	70:10:20	99.84	84.77	82.81
Connected ResUNets [38]	70:10:20	86.91	-	89.52
DS U-Net [16]	Official split	99.70	83.10	82.70
MNPNet [12] *	50:10:40	-	-	91.69
Ours	**Official split**	**99.89**	**89.65**	**94.58**
	80:10:10	**99.90**	**86.96**	**85.72**
	70:10:20	**99.89**	**85.69**	**90.42**
	50:10:40	**93.54**	**90.41**	**95.26**

TABLE VI
RESULTS OF THE ABLATION EXPERIMENTS.

Model	REC (%)	DC (%)	IoU (%)
baseline (UNet [28])	79.56	71.63	68.49
baseline + Dense-block	81.62	76.59	79.18
baseline + SCR no-atrous	82.48	80.31	83.54
baseline + SCR	83.58	88.15	87.18
baseline + MSFEF no-atrous	80.17	81.49	80.57
baseline + MSFEF no-SA	81.58	81.62	82.59
baseline + MSFEF no-weight sharing	82.06	83.14	82.63
baseline + MSFEF	84.79	86.35	83.15
baseline + MSFEF + SCR	83.95	81.46	84.51
baseline + Dense-block + MSFEF	85.98	88.57	87.95
baseline + Dense-block + SCR	84.26	89.21	89.67
baseline + Dense-block + SCR + MSFEF (no-weight sharing)	85.27	92.16	87.38
baseline + Dense-block + SCR + MSFEF (Ours)	**89.65**	**94.58**	**91.47**

1) Evaluating the effectiveness of the dense-block: To demonstrate the effectiveness of the proposed dense-block, we conducted an experiment by adding dense-block to the baseline and compared the performance achieved by this baseline + dense-block with the baseline. As seen in Table VI, after embedding the proposed dense-block there is an enhancement of 2.06% of REC, 4.96% of DC, and 10.69% of IoU.

2) Evaluating the effectiveness of the SCR module: For evaluating the effectiveness of the proposed SCR module, we conducted several experiments by adding the SCR module to the baseline model and by adding the SCR module without parallel atrous convolutions (SCR no-atrous in Table VI). As shown in Table VI, the addition of the SCR module to the baseline model has shown considerable improvements in terms of all the performance metrics considered. And the performance achieved by the SCR no-atrous model is poor than the baseline with the SCR module. This shows that the parallel atrous convolutions with variant atrous rates are more advantageous for extracting global information.

3) Evaluating the effectiveness of the MSFEF module: To evaluate the MSFEF module's effectiveness, we conducted several experiments by adding the MSFEF module to the baseline, adding the MSFEF module without atrous convolution (MSFEF no-atrous in Table VI), adding MSFEF without SA block (MSFEF no-SA), and adding the MSFEF module without shared-weights (MSFEF no-weight sharing in Table VI). Compared with the baseline, the addition of the MSFEF module resulted in the enhancement of 5.23% of REC, 14.72% of DC, and 14.66% of IoU, respectively. This indicates that the proposed MSFEF module can fuse multi-scale context information dynamically. While employing MSFEF no-atrous in the baseline, there is a depreciation of 4.62% of REC, 4.86% of DC, and 2.58% of IoU, respectively, when compared with the baseline with MSFEF. This shows the necessity of extracting multi-scale information. The addition of MSFEF no-SA in the baseline decreased the REC by 3.21%, DC by 4.73%, and IoU by 0.56%, respectively, when compared with the baseline with MSFEF. In addition to these experiments, we also evaluated the effect of weight-sharing in the MSFEF module. As seen in Table VI, we observe that the model with MSFEF module with weight-sharing enhanced the REC, DC, and IoU by 2.73%, 3.21%, 0.52% respectively, when compared with the baseline with MSFEF no-weight sharing. This implies that the dynamic selection of multi-scale context information, and weight-sharing between the layers is advantageous for accurate breast lesion segmentation. In addition to these experiments, we also evaluated the impact of weight-sharing on the entire network. We observed that when the entire network is trained without weight sharing in the MSFEF module, it achieved an REC, DC, and IoU of 85.27%, 92.16%, and 87.38% respectively. After using weight sharing, the network had an improvements of 4.38% of REC, 2.42% of DC, and 4.09% of IoU. All these experiments indicate that the proposed MSFEF module enhances the segmentation performance by taking advantage of the scale-attention and multi-scale information fusion.

F. Performance achieved for different BI-RADS scores of mammograms.

Breast density plays an important role in mass segmentation [40]. In practice, a higher BI-RADS score indicates a higher percentage of glandular tissue in the breast, making the segmentation process more difficult [41]. For presenting the efficiency of the proposed work, we compared the DC achieved for each BI-RADS score with the SOTA models (see Table VII). From Table VII, it can be noted that with the increase in the BI-RADS score, there is a gradual decrease in the DC. This indicates that the mass density significantly affected the segmentation performance. But when compared with SOTA models, the DC achieved by our EU-Net model in all BI-RADS scores is higher on both datasets.

We also compared the DC evaluation with existing approaches. However, in the existing approaches, only the FS-Unet [36] presented the BIRADS score specific to DC evaluation, therefore we only employed this method for comparison. This comparison is presented in Table VIII. The same pattern of decrease in DC with increase in BI-RADS score is found in DC achieved by the FS-Unet [36]. However, the DC achieved by our model is higher than the ARF-Net. Especially, the DC was enhanced by 9.65% for mammograms with BI-RADS

TABLE VII
BI-RADS SCORE SPECIFIC DC (%) COMPARISON WITH SOTA MODELS.

Dataset	BI-RADS score	SENet [34]	Attention U-Net [33]	FCN [32]	CE-Net [31]	UNet++ [30]	UNet [28]	DeepLabV3+ [29]	Ours
INBreast [19]	2	83.97	84.51	83.42	85.91	82.38	76.79	80.19	**97.92**
	3	80.52	84.22	80.18	83.42	81.54	75.92	78.85	**95.39**
	4	78.47	83.69	78.94	81.24	80.67	74.36	76.72	**95.04**
	5	76.79	81.32	77.07	80.06	78.49	73.24	75.62	**94.62**
	6	74.56	77.37	76.83	78.45	77.65	70.19	73.26	**87.29**
CBIS-DDSM [20]	1	81.32	82.35	81.94	87.79	80.13	74.59	81.49	**93.32**
	2	80.63	80.62	80.18	86.94	78.92	72.31	79.82	**92.31**
	3	77.91	78.91	76.83	85.40	77.06	70.52	76.78	**89.03**
	4	75.67	78.05	75.59	81.92	75.69	69.13	74.65	**87.04**

TABLE VIII
BI-RADS SCORE SPECIFIC DC (%) COMPARISON WITH EXISTING APPROACHES.

Dataset	BI-RADS	FS-Unet [36]	Ours
INBreast [19]	2	-	97.92
	3	-	95.39
	4	-	95.04
	5	-	94.62
	6	-	87.29
CBIS-DDSM [20]	1	84.65	**93.32**
	2	86.61	**92.31**
	3	81.82	**89.03**
	4	77.39	**87.04**

score of 4, compared with FS-Unet. This shows the efficiency of our model in segmenting breast masses from mammograms than the other models, and existing approaches.

G. Statistical significance assessment

We further performed the paired t-test to understand the performance enhancement by the proposed EU-Net over the other SOTA models on the two datasets. The $p-values$ are presented in Table IX. In this work, we used the Bonferroni correction [42] in our experiments to adjust the α value. The orginal α was 0.05, but after performing 7 experiments, the adjust α equals to 0.007, i.e., $(\approx \frac{0.05}{7})$. Since all the $p-values$ of DC, and IoU on the two datasest are less that 0.007, the improvements are statistically significant. These results show the effectiveness of the proposed EU-Net.

TABLE IX
STATISTICAL SIGNIFICANCE ANALYSIS OF THE PROPOSED EU-NET WITH SOTA MODELS.

Network	CBIS-DDSM		INBreast	
	DC	IoU	DC	IoU
EU-Net-SENet	0.0004	0.0002	0.0005	0.0006
EU-Net-Attention U-Net	0.0014	0.0016	0.0019	0.0015
EU-Net-FCN	0.0010	0.0009	0.0014	0.0013
EU-Net-CE-Net	0.0023	0.0027	0.0037	0.0030
EU-Net-UNet++	0.0017	0.0018	0.0025	0.0020
EU-Net-UNet	0.0001	0.00005	0.0002	0.0003
EU-Net-DeepLabV3+	0.0009	0.0006	0.0005	0.0006

TABLE X
COMPUTATIONAL COMPLEXITY COMPARISON WITH SOTA MODELS.

Model	FLOPS (G)	Parameter (M)	Inference time (ms)	Memory (MB)
SENet	35.8	23.4	190	301.5
Attention U-Net	45.3	32.3	372	406.1
FCN	27.6	18.6	166	217.7
CE-Net	41.6	29.0	259	385.8
UNet++	38.4	26.2	231	341.6
UNet	33.7	21.4	185	274.6
DeepLabV3+	56.0	40.9	462	486.5
Ours	94.0	62.1	521	566.4

H. Computational complexity

We computed the computational complexity of our model in terms of floating-point operations per second (FLOPS), the number of model parameters, and the average interface time per input mammogram image. As we have re-trained, and evaluated the SOTA models, we have computed the computational complexity of those models as well. And compared them with the proposed model, this comparison is presented in Table X. As seen in Table X, our model was more computationally expensive compared to the SOTA models. Among the SOTA models, FCN has the least computational complexity. We agree that our model is computationally expensive compared to other SOTA models, but in medical informatics, the performance of the models is very important [43], [44] as they deal with the lives of the patients. When compared with FCN, our model achieved significant performance improvements in terms of HD, MSD, PA, REC, IoU, and DC on both mammogram datasets.

I. Generalization ability

To evaluate the proposed EU-Net's generalization ability, we performed two sets of experiments. The first one is cross-dataset evaluation, where the model is trained on the training set of dataset_1 and tested on the test of dataset_2 and vice-versa. And the second one is training and testing the model on the tertiary dataset with different modality without any hyper-parameter tuning or structural modifications. The experimental results are illustrated below:

1) Cross-dataset evaluation: As discussed above, the model trained on the training set of CBIS-DDSM is tested on the test set (i.e., 20% of the images) of the INBreast dataset. And

the model trained on the train set of INbreast (70% of the images) is tested on the test set of the CBIS-DDSM dataset. To verify the proposed EU-Net's effectiveness, we compared its performance with SOTA models with the same train-test setting. This comparison is tabulated in Tables XI, and XII. Our model achieved better performance than the SOTA models in both settings.

2) Tertiary dataset evaluation: We used breast ultrasound UDIAT dataset [21] to evaluate our model on a different modality as done in [36]. The UDIAT dataset didn't have any official data split, so we employed the standard 5-fold cross-validation. We compared the quantitative performance of the SOTA models on the UDIAT dataset employing the same evaluation protocol. The quantitative comparison is presented in Table XIII. The proposed EU-Net outperformed the SOTA models in all four performance metrics.

TABLE XI
CROSS-DATASET EVALUATION ON INBREAST.

Model	HD (mm)	MSD (mm)	PA (%)	REC (%)	IoU (%)	DC (%)
SENet [34]	18.48	4.93	97.11	83.66	78.94	83.35
Attention U-Net [33]	27.04	6.32	96.29	82.39	77.35	81.46
FCN [32]	23.38	5.86	96.74	81.44	79.93	83.01
CE-Net [31]	13.24	3.56	97.15	84.56	79.54	83.42
UNet++ [30]	26.91	5.99	96.38	87.35	77.47	81.63
UNet [28]	27.81	6.93	96.15	85.36	74.69	79.55
DeepLabV3+ [29]	14.69	4.24	97.13	84.04	79.34	82.34
Ours	**6.82**	**2.96**	**98.85**	**89.38**	**83.59**	**85.32**

TABLE XII
CROSS-DATASET EVALUATION ON CBIS-DDSM.

Model	HD (mm)	MSD (mm)	PA (%)	REC (%)	IoU (%)	DC (%)
SENet [34]	13.84	5.73	98.86	83.59	71.95	83.81
Attention U-Net [33]	15.69	6.37	98.41	84.22	71.35	83.53
FCN [32]	20.48	7.03	97.69	85.15	71.33	81.16
CE-Net [31]	11.29	4.59	99.12	86.79	72.81	84.49
UNet++ [30]	23.57	7.41	97.55	83.46	70.49	80.14
UNet [28]	25.62	7.89	96.39	79.63	67.58	77.37
DeepLabV3+ [29]	17.41	6.94	98.25	85.98	70.49	82.46
Ours	**5.37**	**3.51**	**99.84**	**89.41**	**74.51**	**87.38**

TABLE XIII
QUANTITATIVE COMPARISON: SOTA MODELS ON UDIAT.

Model	HD (mm)	MSD (mm)	PA (%)	REC (%)	IoU (%)	DC (%)
SENet [34]	15.47	3.59	97.81	86.69	77.29	84.23
Attention U-Net [33]	13.27	3.21	98.78	88.42	80.44	88.18
FCN [32]	15.04	3.53	98.03	84.06	78.37	82.39
CE-Net [31]	18.29	3.69	97.35	82.46	76.81	80.63
UNet++ [30]	14.62	3.35	98.16	85.28	79.05	81.45
UNet [28]	20.59	4.91	98.31	85.66	79.36	82.51
DeepLabV3+ [29]	23.15	5.67	97.23	80.52	76.13	77.41
Ours	**8.35**	**2.48**	**99.20**	**90.57**	**83.23**	**89.71**

In addition to the quantitative comparison, the qualitative comparison of tertiary dataset evaluation is presented in Fig. 7. The red and green contours represent the ground truth and the predicted masks. As seen in Fig. 7, the predicted masks of the proposed EU-Net are close to the ground truths. In contrast, all 7 SOTA models reported over-segmentation or under-segmentation of breast masses. The Attention U-Net [33] achieved the second-best qualitative performance of the 7 SOTA models. Major over-segmentation is produced by SENet [34], CE-Net [31]. Among the 7 SOTA models, the FCN [32], DeepLabV3+ [29], and UNet [28] produced under-segmentation results. And the least performance was achieved by the U-Net.

TABLE XIV
QUANTITATIVE COMPARISON: EXISTING APPROACHES ON UDIAT.

Methods	Evaluation protocol	DC (%)	REC (%)
Singh et al. [45]	70:10:20	86.82	91.55
Haung et al. [46]	65:15:20	82.40	-
Lee et al. [47]	10 fold	76.58	80.41
ARF-Net [35]	5 fold	88.12	89.44
Ours	**5 fold**	**89.71**	**90.57**
	10 fold	**90.47**	**92.18**
	70:10:20	**90.52**	**93.68**
	65:15:20	**88.62**	**88.41**

We also compared the performance achieved by our EU-Net with the existing approaches. This comparison is tabulated in Table XIV. For a fair comparison we re-evaluated our work with the evaluation protocol used by the existing approaches. As seen in Table XIV, the proposed EU-Net outperformed all the existing approaches. This shows the powerful generalization ability of our model.

J. Limitation of the proposed EU-Net

Even though the proposed EU-Net has achieved excellent performance in mass segmentation on INBreast and CBIS-DDSM datasets, it has achieved limited segmentation accuracy on a few individual cases. For example, two cases are presented in Fig 8. In Fig 8(A), the breast mass occupies very little space in the whole mammogram, in Fig 8(B), the brightness of the surrounding fibroglandular tissue is the same as the breast mass. These cases are very challenging not only for the proposed EU-Net model but also for the other segmentation models reported in the literature.

IV. CONCLUSION

In this work, we propose an efficient deep learning framework, EU-Net, for the segmentation of masses from digital mammograms. The proposed EU-Net is designed by embedding dense-blocks and MSFEF, into the standard U-shaped network to extract, and fuse rich multi-scale information about the masses from the input mammograms. In addition, the SCR module is embedded into EU-Net for redesigning the skip connection for avoiding the ignorance of the global context information which is efficient for accurate mass segmentation. We have experimented EU-Net on 2 publicly available datasets namely INBreast, and CBIS-DDSM. We have compared the

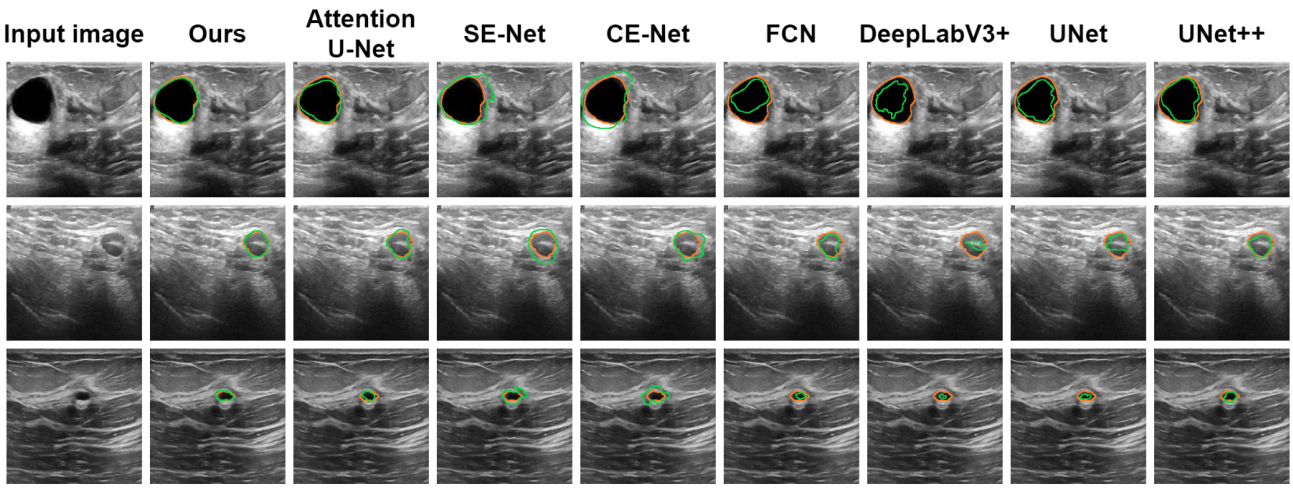

Fig. 7. Qualitative comparison with SOTA models on UDIAT dataset.

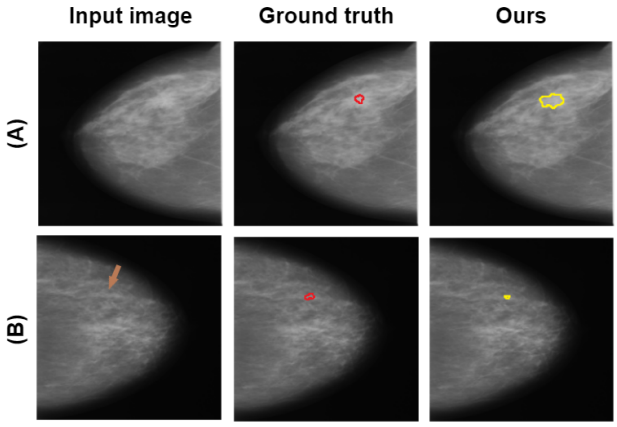

Fig. 8. Limited segmentation performance of the proposed EU-Net; the brown arrow shows the fibroglandular tissue.

performance achieved by EU-Net with 7 SOTA models including CE-Net, FCN, Attention U-Net, UNet++, SENet, DeepLabV3+, and UNet, and also with the existing works reported in the literature. The experimental results indicate that the EU-Net has outperformed the SOTA models and the existing approaches on both datasets. It is to be noted that the EU-Net outperformed the SOTA models with a huge difference in DC by 11.83% on the INBreast, and 9.07% on the CBIS-DDSM datasets.

In addition to this, we also evaluated the generalization ability of our model by performing the cross-dataset evaluation, and tertiary dataset evaluation. For terinary dataset evaluation, we used the UDIAT breast ultrasound dataset. Experimental results indicates that the proposed model out-performed the SOTA models, and existing approaches in ternary dataset evaluation, and SOTA models in cross-dataset evaluation. These experiments collectively indicate the improved efficiency and high generalization ability of the proposed EU-Net. In the future, we will be extending our work to segment other organs from different modalities.

REFERENCES

[1] S. W. Duffy, L. Tabár, A. M.-F. Yen, P. B. Dean, R. A. Smith, H. Jonsson, S. Törnberg, S. L.-S. Chen, S. Y.-H. Chiu, J. C.-Y. Fann *et al.*, "Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women," *Cancer*, vol. 126, no. 13, pp. 2971–2979, 2020.

[2] K. Ng and M. Muttarak, "Advances in mammography have improved early detection of breast cancer," *Journal-Hong Kong College of Radiologists*, vol. 6, pp. 126–131, 2003.

[3] A. G. Mulley, C. Trimble, and G. Elwyn, "Stop the silent misdiagnosis: patients' preferences matter," *Bmj*, vol. 345, 2012.

[4] M. Kamba, M. Manabe, S. Wakamiya, S. Yada, E. Aramaki, S. Odani, I. Miyashiro *et al.*, "Medical needs extraction for breast cancer patients from question and answer services: Natural language processing-based approach," *JMIR cancer*, vol. 7, no. 4, p. e32005, 2021.

[5] R. Hou, Y. Peng, L. J. Grimm, Y. Ren, M. A. Mazurowski, J. R. Marks, L. M. King, C. C. Maley, E. S. Hwang, and J. Y. Lo, "Anomaly detection of calcifications in mammography based on 11,000 negative cases," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 5, pp. 1639–1650, 2021.

[6] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," *Annual review of biomedical engineering*, vol. 15, pp. 327–357, 2013.

[7] Y. Wang, D. Tao, X. Gao, X. Li, and B. Wang, "Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions," *Pattern Recognition*, vol. 44, no. 9, pp. 1903–1915, 2011.

[8] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 605–612.

[9] E. Michael, H. Ma, H. Li, F. Kulwa, and J. Li, "Breast cancer segmentation methods: current status and future potentials," *BioMed Research International*, vol. 2021, 2021.

[10] Z. Rezaei, "A review on image-based approaches for breast cancer detection, segmentation, and classification," *Expert Systems with Applications*, vol. 182, p. 115204, 2021.

[11] E. Justaniah, A. Alhothali, and G. Aldabbagh, "Mammogram segmentation techniques: A review," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.

[12] R. Wang, Y. Ma, W. Sun, Y. Guo, W. Wang, Y. Qi, and X. Gong, "Multi-level nested pyramid network for mass segmentation in mammograms," *Neurocomputing*, vol. 363, pp. 313–320, 2019.

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

[14] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. I. Laurenson, "Dual convolutional neural networks for breast mass segmentation and

diagnosis in mammography," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 3–13, 2021.

[15] J. Chowdary, P. Yogarajah, P. Chaurasia, and V. Guruviah, "A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images," *Ultrasonic Imaging*, vol. 44, no. 1, pp. 3–12, 2022.

[16] N. Ravitha Rajalakshmi, R. Vidhyapriya, N. Elango, and N. Ramesh, "Deeply supervised u-net for mass segmentation in digital mammograms," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 59–71, 2021.

[17] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. D. Feng, and S. Wang, "Aunet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Physics in Medicine & Biology*, vol. 65, no. 5, p. 055005, 2020.

[18] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi, and A. Nori, "Autofocus layer for semantic segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 603–611.

[19] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[20] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, no. 1, pp. 1–9, 2017.

[21] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[23] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3917–3926.

[24] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] J. Henrikson, "Completeness and total boundedness of the hausdorff metric," *MIT Undergraduate Journal of Mathematics*, vol. 1, no. 69-80, p. 10, 1999.

[27] V. Yeghiazaryan and I. D. Voiculescu, "Family of boundary overlap metrics for the evaluation of medical image segmentation," *Journal of Medical Imaging*, vol. 5, no. 1, p. 015006, 2018.

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

[29] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[30] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[31] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[32] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 640–651, 2016.

[33] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[34] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 421–429.

[35] C. Xu, Y. Qi, Y. Wang, M. Lou, J. Pi, and Y. Ma, "Arf-net: An adaptive receptive field network for breast mass segmentation in whole mammograms and ultrasound images," *Biomedical Signal Processing and Control*, vol. 71, p. 103178, 2022.

[36] J. Pi, Y. Qi, M. Lou, X. Li, Y. Wang, C. Xu, and Y. Ma, "Fs-unet: Mass segmentation in mammograms using an encoder-decoder architecture with feature strengthening," *Computers in Biology and Medicine*, vol. 137, p. 104800, 2021.

[37] W. Zhao, M. Lou, Y. Qi, Y. Wang, C. Xu, X. Deng, and Y. Ma, "Adaptive channel and multiscale spatial context network for breast mass segmentation in full-field mammograms," *Applied Intelligence*, vol. 51, no. 12, pp. 8810–8827, 2021.

[38] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, "Connected-unets: a deep learning architecture for breast mass segmentation," *NPJ Breast Cancer*, vol. 7, no. 1, pp. 1–12, 2021.

[39] T. Shen, C. Gou, J. Wang, and F.-Y. Wang, "Simultaneous segmentation and classification of mass region from mammograms using a mixed-supervision guided deep model," *IEEE Signal Processing Letters*, vol. 27, pp. 196–200, 2019.

[40] H. Bulu, A. Alpkocak, and P. Balci, "Uncertainty modeling for ontology-based mammography annotation with intelligent bi-rads scoring," *Computers in Biology and Medicine*, vol. 43, no. 4, pp. 301–311, 2013.

[41] Y. Qi, Z. Yang, J. Lei, J. Lian, J. Liu, W. Feng, and Y. Ma, "Morph_spcnn model and its application in breast density segmentation," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2821–2845, 2021.

[42] E. W. Weisstein, "Bonferroni correction," *https://mathworld. wolfram. com/*, 2004.

[43] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on u-net: A review," *Journal of Imaging Science and Technology*, vol. 64, pp. 1–12, 2020.

[44] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The lancet digital health*, vol. 1, no. 6, pp. e271–e297, 2019.

[45] V. K. Singh, M. Abdel-Nasser, F. Akram, H. A. Rashwan, M. M. K. Sarker, N. Pandey, S. Romani, and D. Puig, "Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework," *Expert Systems with Applications*, vol. 162, p. 113870, 2020.

[46] H. Huang, H. Chen, H. Xu, Y. Chen, Q. Yu, Y. Cai, and Q. Zhang, "Cross-tissue/organ transfer learning for the segmentation of ultrasound images using deep residual u-net," *Journal of Medical and Biological Engineering*, vol. 41, no. 2, pp. 137–145, 2021.

[47] H. Lee, J. Park, and J. Y. Hwang, "Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 7, pp. 1344–1353, 2020.