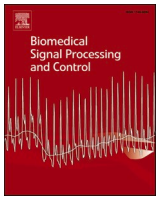
Biomedical Signal Processing and Control 71 (2022) 103178

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



ARF-Net: An Adaptive Receptive Field Network for breast mass segmentation in whole mammograms and ultrasound images

Chunbo Xu, Yunliang Qi, Yiming Wang, Meng Lou, Jiande Pi, Yide Ma *

School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu, China

ARTICLE INFO

Keywords:
Segmentation
Mammogram
Ultrasound image
Deep learning
Encoder-decoder

ABSTRACT

UNet adopting an encoder-decoder structure has been used widely in medical image segmentation tasks for its outstanding performance. However, in our work, we find that UNet has the worse segmentation performance of small masses. The reason behind this is that the sizes of receptive fields are limited. In this work, to address this issue, we develop a novel end-to-end model, Adaptive Receptive Field Network (ARF-Net), for the precise breast mass segmentation in whole mammographic images and ultrasound images. ARF-Net composes of an encoder network and a corresponding decoder network, followed by a pixel-wise classifier. In ARF-Net, a Selective Receptive Filed Module (SRFM) is proposed to allocate the suitable sizes of receptive fields to the breast masses of different sizes. SRFM consists of a Multiple Receptive Field Module (MRFM) for generating multiple receptive fields of different sizes and a Multi-Scale Selection Module (MSSM) for selecting the suitable sizes of receptive fields based on the objects' size. The proposed ARF-Net achieves the dice index of 86.1%, 85.75%, and 88.12% on the two mammographic databases (INbreast and CBIS-DDSM) and one ultrasonic database (UDIAT), respectively. Moreover, extensive ablation experiments show that ARF-Net transcends several state-of-the-art segmentation networks, and the developed MSSM exceeds several counterparts.

1. Introduction

Due to its extensiveness and high mortality rate, breast cancer has become one of the most common cancer types diagnosed among women. It is the second-highest cause of death among women and leads to the second-largest number of deaths in women worldwide. The statistics released by the International Agency for Research on Cancer (IARC) in December 2020 reveal that breast cancer is responsible for one in six of all cancer deaths among women. According to the American Cancer Society, if breast cancer is detected in the early stage, the 5-year relative survival rate is 99%. Hence, to decrease the mortality of breast cancer, the early diagnosis and detection of breast cancer is essential. For early diagnosis, medical imaging plays a crucial role. Digital mammography is primarily utilized for the diagnosis of breast cancer. To read mammograms requires experienced and well-trained radiologists. However, even a well-trained expert might have an inter-observer variation rate. Thus, computer-aided diagnosis (CAD) could be developed to assist radiologists in breast classification and detection tasks [14,35].

Breast masses in mammograms are a common clinical sign of breast cancer. The nature of breast masses can offer a considerable reference for follow-up treatment. Generally speaking, compared with the breast mass with regular shape, the breast mass with irregular shape is more likely to be malignant. Hence, the masses segmentation in whole mammograms is a crucial procedure of breast cancer CAD systems and is the key to further qualitative analysis of breast cancer. But the breast masses have various sizes, shapes, and locations, bringing a lot of challenges to accurate mass segmentation.

In the past few decades, many traditional breast mass segmentation approaches have been developed. These conventional methods can be divided into region-based algorithms [17,25], contour-based algorithms [42,58], and clustering algorithms [7]. It is noteworthy that the features used in conventional algorithms are hand-crafted. For example, Mencattini et al. [33] developed a modified region-based segmentation approach for reducing the computational cost and improving efficiency. In the work of Patel et al. [41], they proposed an adaptive k-means clustering method for breast lesion segmentation. In the work of Dalmiya et al. [13], the wavelet transformation and k-means clustering were used for breast mass segmentation. Hao et al. [18] proposed a hybrid method combining a random walks method and Chan-Vese active contour for breast mass segmentation in the mammographic images.

* Corresponding author.
E-mail address: ydma@lzu.edu.cn (Y. Ma).

https://doi.org/10.1016/j.bspc.2021.103178
Received 1 June 2021; Received in revised form 30 July 2021; Accepted 7 September 2021
Available online 17 September 2021

However, the performance of the traditional methods is restricted by hand-crafted features which are dependent on the designer's professional skills and knowledge.

We are in the Deep Learning (DL) era. With the help of DL, especially Convolution Neural Network (CNN), computer vision has achieved great development. Object classification, object detection, and semantic segmentation are three fundamental tasks in computer vision. There are many DL models proposed for solving these tasks. For example, ResNet [21], ResNeXt [60], Res2Net [16], and Inceptionv1-v4 [24,55–57] for image classification; FCN [31], SegNet [3], DeepLabv1-v3plus [9–12], and PSPNet [65] for semantic segmentation; Faster RCNN [48], SSD [30], Mask RCNN [20], and YOLOv1-v4 [4,45–47] for object detection.

Recently, DL techniques have been widely applied to medical image processing due to their excellent performance. For example, Mohamed et al. [34] proposed a CNN-based approach for the classification of mammogram density. In the work of Al-Masni [2], a YOLO-based CAD system is proposed for the detection and classification of breast cancer lumps simultaneously. In the work of Abdelhafiz et al. [1], for the segmentation of breast mass in mammograms, they designed RU-Net by integrating the residual attention modules into the UNet. Moreover, a ResNet classifier following RU-Net was used for the classification of segmented masses as malignant or benign. Rampun et al. [43] proposed a modified holistically-nested edge detection model which was equipped with multi-scale and multi-level learning. Xu et al. [61] developed a multi-channel, multi-scale FCN for mass segmentation in mammograms.

As we all know, YOLOv3 [47] employs the feature layers of three different scales to predict boxes. The feature layer of the large scale is used to detect small objects and the feature layer of the small scale is used to detect large objects. The reason behind is that the size of the receptive field at the large-scale feature layer is small and the size of the receptive field at the small-scale feature layer is large. The receptive field of large size should be utilized to detect objects of large size. On the contrary, the receptive field of small size should be utilized to detect objects of small sizes. For the breast mass segmentation in whole mammograms and breast ultrasound images, the breast mass has a variety of sizes and shapes. Therefore, the idea of multiple receptive fields is useful for boosting the performance of breast mass segmentation.

In this paper, the main contributions of our work mainly include:

- An Adaptive Receptive Field Network (ARF-Net) is proposed for precise breast mass segmentation in whole mammographic images and ultrasound images.
- A Selective Receptive Field Module (SRFM) in ARF-Net is proposed to select the approximate size of receptive field for the breast masses of different sizes. A SRFM is composed of a Multiple Receptive Field Module (MRFM) and a Multi-Scale Selection Module (MSSM). The MRFM is proposed to generate multiple receptive fields of different sizes. The MSSM is proposed to allocate the approximate size of the receptive field for the breast masses of different sizes.
- The ARF-Net is evaluated based on two mammographic databases (CBIS-DDSM [27] and INbreast [36]) and one ultrasonic database (UDIAT [63]). The ARF-Net obtain the dice index of 85.75%,86.1%, and 88.12% on CBIS-DDSM, INbreast, and UDIAT benchmarks respectively, which outperforms the state-of-the-art approaches.
- The ablation experiments demonstrate that our designed model is equipped with robustness and can learn generalizable representations.

The organizational structure of this paper is as follows. The related works are depicted in Section 2. Section 3 elaborates on our developed approaches. The experiments are detailedly depicted in Section 4. Section 5 elaborates on the ablation studies. The conclusions are presented in Section 6.

2. Related works

The mass segmentation in the whole mammographic images mainly suffers from the challenges of two aspects: the breast mass with various shapes and sizes and the interference of the normal parenchymal region. Last few years, several works have been developed for solving the aforementioned two problems. These works are elaborated as follows.

In the work of Sun et al. [54], they designed an attention-guided dense-upsampling network to segment the mass in whole mammographic images directly. Moreover, the attention-guided dense-upsampling block, which used a channel-attention approach to refine the features, was also proposed. However, the channel-attention method used in their work only used a single-scale feature, namely, global average-pooling feature content. Ravitha et al. [44] developed a deeply supervised U-Net (DS-UNet). They first used a contrast-limited adaptive histogram equalization approach to pre-process the input mammograms. Then the pre-processed images were fed into the DS-UNet. Finally, the output segmentation map was refined by dense conditional random fields. However, their method was not end-to-end. In the work of Chen et al. [8], they proposed a multi-scale adversarial network including a segmentation network (an improved U-Net) for generating the suspicious regions' mask and a discrimination network for discriminating the input masks to segment the breast mass in the whole mammograms. The integration of Earth-Mover distance and weighted cross-entropy loss was employed as object function during training to generate better segmentation results and alleviate the unbalanced class issue. In the work of Wang et al. [59], a developed effective attention method was incorporated into U-Net being as a generator network and made U-Net more focus on the breast mass in whole mammograms. A CNN with a multi-scale pooling module was used as a discriminator network for learning punctilious features from the breast masses with various sizes and shapes. Nevertheless, the last two methods had the disadvantage of training difficultly. For example, in the work of Chen et al. [8], the spectral normalization was used to stabilize the training process of their model.

The receptive field of large size is used suitably to recognize the large object. On the contrary, the receptive field of small size is used suitably to recognize the small object. If the receptive field of large size is used to recognize the small object, the irrelevant information is introduced; If the receptive field of small size is used to recognize large objects, most of the information about objects is lost. Therefore, choosing the appropriate receptive field's size for the breast masses of different sizes is crucial, which can boost the performance of breast mass segmentation in whole mammograms.

In this work, an Adaptive Receptive Field Network (ARF-Net) is proposed for precise mass segmentation in whole mammographic images. In ARF-Net, we propose a Selective Receptive Field Module (SRFM), including Multiple Receptive Field Module (MRFM) and an effective Multi-Scale Selection Module (MSSM). The MRFM is designed for obtaining spatial information of different sizes. An effective MSSM is designed to select the suitable receptive field's size for the breast masses of different sizes.

3. Proposed approach

In this part, we first introduce the overview of ARF-Net. Then, we first depict the designed SRFM in detail. Finally, we will elaborate on the overall architecture of the developed ARF-Net.

3.1. Overview

Our proposed ARF-Net adopts a U-shaped architecture consisting of an encoder network and a corresponding decoder network. In addition, we propose SRFM to better segment the masses of different sizes. SRFM is composed of MRFM and MSSM. MRFM offers receptive fields of different sizes to recognize the masses of various sizes. MSSM offers

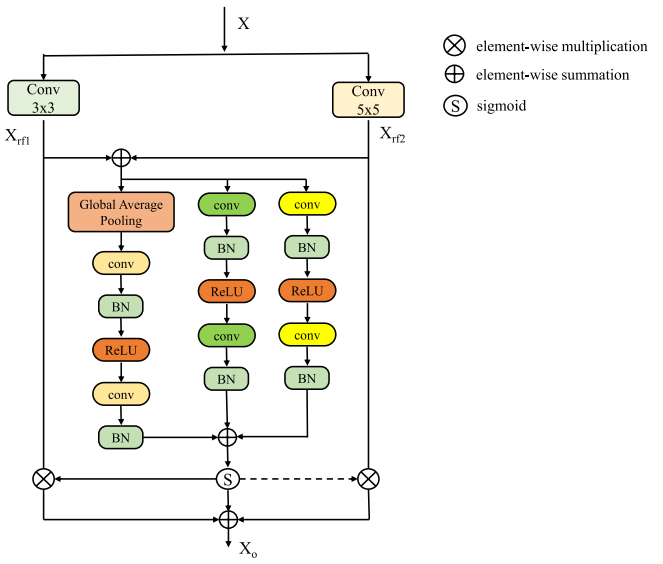


Fig. 1. The architecture of SRFM.

multi-scale attention to make ARF-Net better focus on the masses of different sizes. Because the max-pooling operation can cause the losses of spatial detail formation that is conducive to improve the segmentation performance, it is replaced by our proposed SRFM.

As we all know, high-resolution feature maps play a crucial role in medical image segmentation tasks. In addition, there have been many works [19,38,49,50] that prove that the encoder-decoder architecture is specifically suitable for medical image segmentation. Therefore, our proposed ARF-Net adopts the U-shaped architecture, namely encoder-decoder architecture. The encoder extracts high-level semantic features by using a sequence of convolutional layers, while the decoder leverages skip connections to re-use high-resolution feature maps from the encoder in order to recover lost spatial information from high-level representations.

It is well known that the diversity of the receptive field plays a vital role in medical image segmentation [28,29,64]. In addition, the limited receptive field of UNet can cause its performance bottleneck. The masses have different sizes in our study. Therefore, for the sake of better recognizing masses with different sizes, we propose MRFM consisting of two parallel convolutional layers with different kernel sizes which can generate two feature maps with different semantic information.

The output of MRFM contains two feature maps with different semantic information. The masses of different sizes correspond to different semantic information. If the receptive field of large size is used to segment the mass of small size, the model is likely to pay attention to the noise area; If the receptive field of small size is used to segment the mass of large size, a lot of regions that are more relevant to the semantic information of the masses may not be concerned. Thus, in order to make the network better segment the masses of different sizes, we propose MSSM to offer multi-scale attention.

3.2. SRFM

As shown in Fig. 1, the SRFM is composed of MRFM and MSSM. The MRFM consists of K parallel convolutional blocks, including a batch normalization (BN) layer and a rectified linear unit (ReLU) activation function layer in sequence. K is set to 2 in the following experiments. Each convolution in MRFM uses different kernel sizes. one is 3, the other is 5. It is noteworthy that the convolution that the kernel sizes are 5 is replaced by the atrous convolution with the dilation rate of 2 and the kernel size of 3 for improving efficiency. The MRFM makes the model competent for focusing on objects of different sizes by producing multiple receptive fields of diverse sizes. Fig. 2(a) shows the structure of MRFM.

The channel attention in SE-Net [22] re-weights each channel of feature maps by learning channel-wise weighting vectors from the global contextual contents. The channel attention can be defined as:

$$Y = \sigma(f_2(f_1(g(X)))) \tag{1}$$

where σ denotes the sigmoid function. $X \in R^{C\times H\times W}$ refers to the input feature maps with the spatial size of $H \times W$ and the channel number of C. $Y \in R^{C\times H\times W}$ refers to the output feature maps. The $g(X) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{[:,i,j]}$ denotes the global average pooling (GAP). f_1 and f_2 denote fully-connected (FC) layers which weights are $W_1 \in R^{\frac{C}{r}\times C}$ and $W_2 \in R^{C\times\frac{C}{r}}$ respectively. The first FC layer f_1 is followed by a BN layer and a ReLU layer. The second FC layer is only followed by a BN layer. The f_1 denotes a dimensionality reduction layer and the f_2 refers to a dimensionality increasing layer. The r refers to the channel reduction ratio and is set to 4 in our experiments.

The squeeze operation in the SE module [22] transforms the input feature maps of spatial size $C \times H \times W$ into a weighting vector of length C. This extremely coarse channel attention vector biases the model

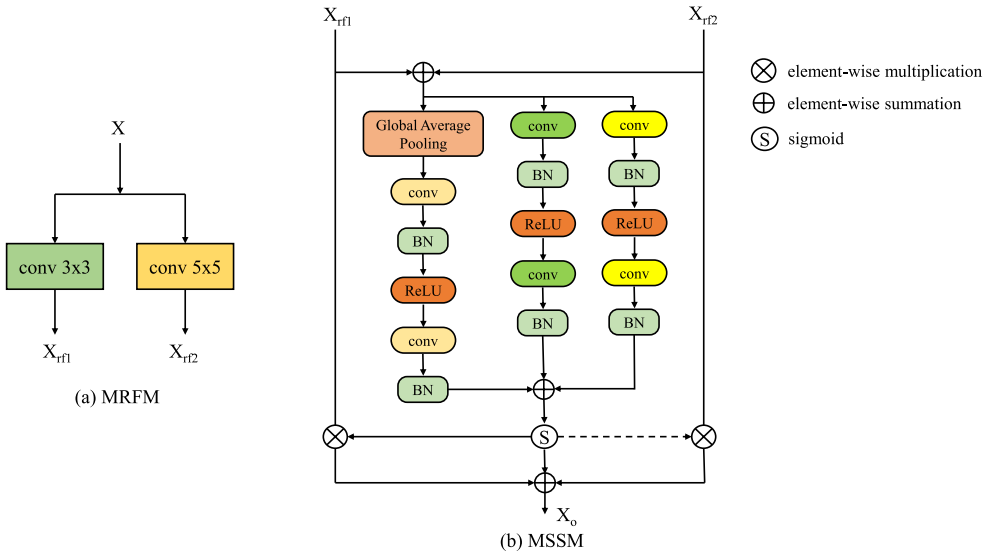


Fig. 2. The architectures of MRFM and MSSM.

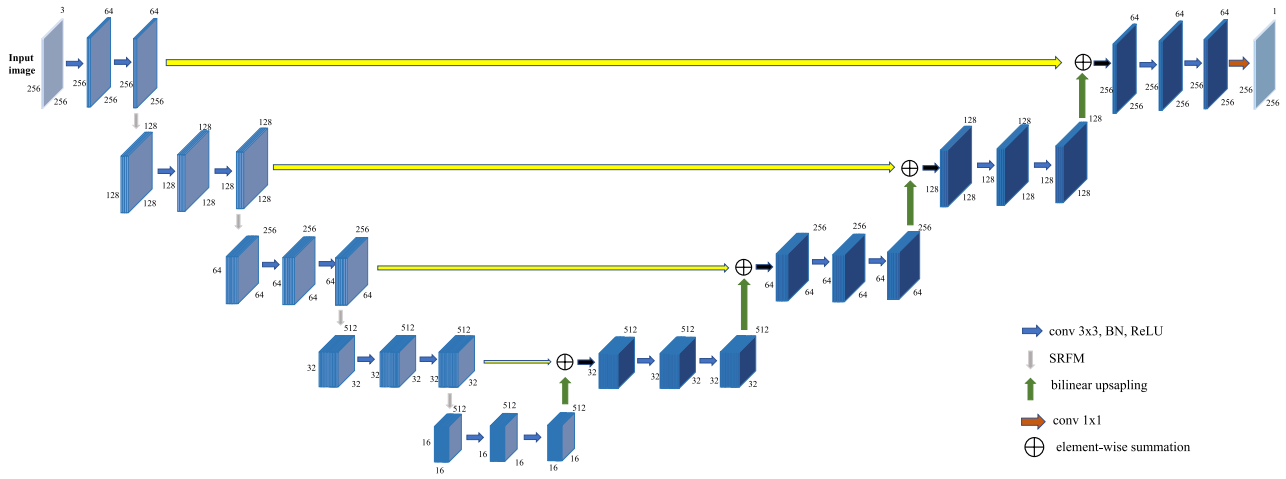


Fig. 3. The architecture of ARF-Net.

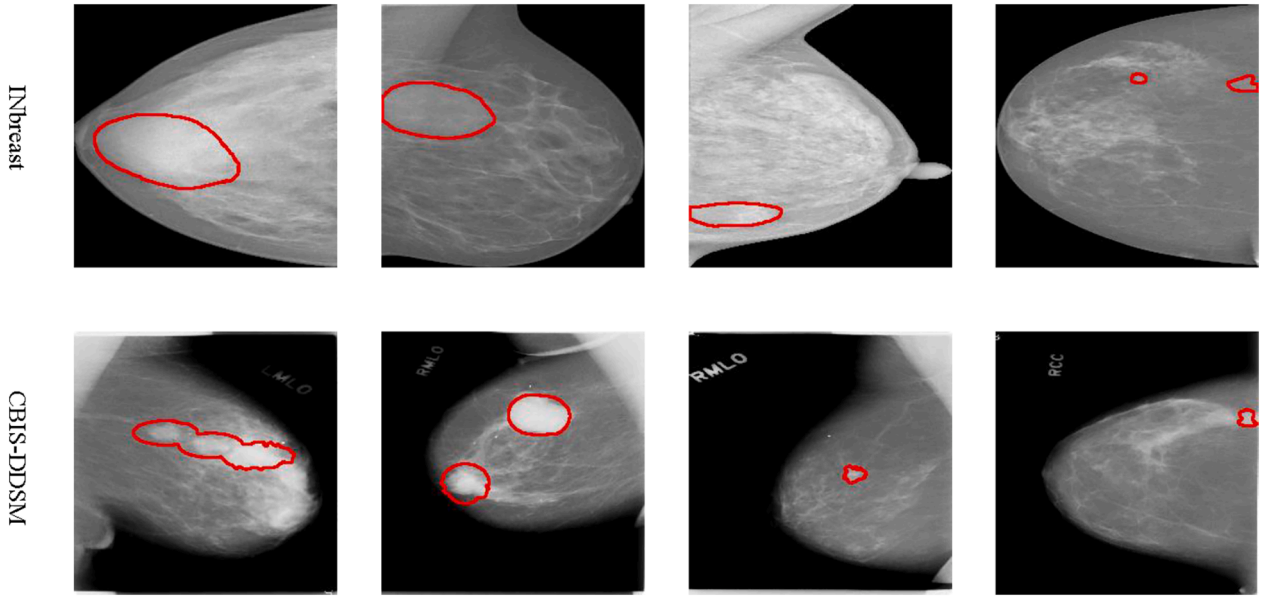


Fig. 4. Several mammograms. The region surrounded by a red closed curve represents the mass. Because breast masses have various sizes, the multiple receptive fields of different sizes should be adopted.

Table 1
Segmentation performance on the CBIS-DDSM benchmark. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
Sun et al. [54]	81.8 ± 0.0	–	84.9 ± 0.3	–	26.9 ± 0.3	2.96 ± 0.03
Chen et al. [8]	82.16	99.81	85.23	99.86	–	
Ravitha et al. [44]	82.7	99.7	84.1	99.8	22.7	2.43
Wang et al. [59]	84.49	–	–	–	–	5.01
Ours	$\mathbf{85.75 \pm 0.22}$	99.81 ± 0.0	$\mathbf{88.91 \pm 0.76}$	$\mathbf{99.89 \pm 0.01}$	$\mathbf{14.36 \pm 1.96}$	$\mathbf{0.93 \pm 0.46}$

towards highlighting the large objects and ignore the small objects. However, the area ratio that most of the breast masses take up the whole mammograms is smaller than 1% in the INbreast and CBIS-DDSM database. Hence, the global-average attention may not be the optimal choice for mass segmentation in whole mammograms. To mitigate the problem resulting from scale variation and small objects, the method of aggregating the multi-scale feature information should be adopted.

Based on the idea of aggregating the multi-scale feature information, apart from the global context path, we add the two local context paths in MSSM. The local context $L(X) \in R^{C\times H\times W}$ is computed via a bottleneck structure as follows:

$$L(X) = f_2^{1\times1}(f_1^{k\times k}(X)) \tag{2}$$

where $f_2^{1\times1}$ and $f_1^{k\times k}$ are two different convolutions and their kernel sizes are $C \times \frac{C}{r} \times 1 \times 1$ and $\frac{C}{r} \times C \times k \times k$ respectively. The first convolution layer $f_1^{k\times k}$ is followed by a BN layer and a ReLU layer and the second convolutional layer $f_2^{1\times1}$ is followed by a BN layer. The only difference

Table 2
Segmentation performance on the INbreast database. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
Sun et al. [54]	79.1 ± 6.0	–	80.8 ± 7.1	–	37.6 ± 15.4	4.04 ± 0.33
Chen et al. [8]	81.64	**99.43**	82.72	99.56	–	–
Ravitha et al. [44]	79	98.11	81	98.4	31.1	1.86
Wang et al. [59]	83.92	–	–	–	–	5.81
Ours(w/o)	85.06 ± 2.64	99.11 ± 0.58	83 ± 1.99	**99.62 ± 0.22**	**11.56 ± 3.79**	1.81 ± 0.85
Ours(w/)	**86.1 ± 2.75**	99.14 ± 0.52	**85.82 ± 2.97**	99.54 ± 0.39	14.34 ± 5.08	**1.8 ± 0.9**

w/o – train from scratch.
w/ – with pre-training on the CBIS-DDSM benchmark.

Table 3
Segmentation performance on UDIAT database.

Method	DSC(%)↑	ACC (%)↑	SEN(%)↑	SPE(%)↑
Osman et al. [39]	73.14 ± 29.17	–	73.21 ± 31.35	**99.49 ± 0.83**
Huang et al. [23]	82.4	–	–	–
Ning et al. [37]	85.553 ± 1.718	–	85.211 ± 1.342	–
Singh et al. [52]	86.82	98.22	**91.55**	99.49
Lee et al. [26]	76.58	97.794	80.41	98.66
Shareef et al. [51]	82	–	84	–
Byra et al. [6]	79.1	98.5	–	–
Su et al. [53]	82.7 ± 0.8	–	–	–
Gao [15]	85.39 ± 1.63	–	88.39 ± 1.12	-
Ours	**88.12**	**98.84**	89.44	99.36

between the first local context path and the second context path is the kernel size of the first convolution. One is 3 × 3, and the other is 5 × 5. It is noteworthy that the convolution with 5 × 5 is replaced by the dilated convolution with the kernel size of 3 and the dilation rate of 2 for improving efficiency. The structure of MSSM is described in Fig. 2(b).

MSSM consists of three different parallel paths, including a global context path and two local context paths. The MSSM can be formulated as:

$$X_{\text{sum}} = X_{\text{rf1}} + X_{\text{rf2}} \tag{3}$$

$$att = \sigma(\text{G}(\text{X}_{\text{sum}}) + \text{L}_1(\text{X}_{\text{sum}}) + \text{L}_2(\text{X}_{\text{sum}})) \tag{4}$$

$$\text{X}_{\text{o}} = \text{att} \otimes \text{X}_{\text{rf1}} + (1 - \text{att}) \otimes \text{X}_{\text{rf2}} \tag{5}$$

where X_{o} is the output feature maps of SRFM. X_{rf1} and X_{rf2} are the output feature maps of MRFM. $\otimes$ refers to the element-wise multiplication.

3.3. ARF-Net

The ARF-Net comprises two networks, including an encoder network and a corresponding decoder network, followed by a 1 × 1 convolutional layer for projecting each 64-dimensional feature map to the number of classes and a sigmoid layer for the pixel-wise classification. The encoder network is composed of the repeated application of two 3 × 3 convolutional layers, each followed by a BN layer, a ReLU layer, and a SRFM for down-sampling the feature maps (halving the size of feature maps). Each decoder in the decoder network is composed of an upsampling layer for up-sampling the summation of the corresponding feature maps from the encoder network and the output feature maps of the previous decoder, and two convolutional layers, each followed by a BN layer and a ReLU layer. The overall framework of ARF-Net is depicted in Fig. 3.

4. Experiments

In this section, we will first introduce two mammographic databases and one ultrasonic dataset used in our study. Secondly, the experimental configurations are described in detail. Thirdly, we will elaborate on the experimental results. Finally, the discussions and analyses are depicted detailedly.

4.1. Databases

Two publicly available mammographic databases, CBIS-DDSM [27] and INbreast [36], are used to conduct an evaluation of our designed ARF-Net in this work. CBIS-DDSM database contains curated mammograms from the largest publicly available mammographic dataset, DDSM [5]. Because the data provider has divided the CBIS-DDSM dataset into a training dataset and a testing dataset, we do not need to split the database anymore. In our experiments, 303 mammograms containing 242 mammograms for training and 61 mammograms for validation are used. The INbreast database has 115 cases containing a total of 410 mammographic images with the normal mammograms of 303, the cancerous mammograms of 72, and the benign mammograms of 35 [36]. The five-fold cross-validation experiments are conducted on INbreast database. Several mammographic images are displayed in Fig. 4.

The ultrasonic database, UDIAT [63] is also used to evaluate the developed ARF-Net. The UDIAT dataset [63] has a total of 163 breast ultrasound images with 53 cancerous images and 110 benign images. Each image from the UDIAT datasets only contains one mass.

4.2. Implementation details

Our all experiments are performed in the publicly available deep learning framework PyTorch [40] and run on NVIDIA RTX 2080 Ti GPU with 11 GB on-chip memory. Optimization is performed using AdamW optimizer [32]. The initial learning rate is set to 0.0001, and the iterations are set to 70. It is noteworthy that all the segmentation models are trained from scratch in the following experiments.

In our experiments, all of the mammographic images are resized 256 × 256 before being fed into the model. Several data augmentation approaches with horizontal flipping, vertical flipping, mirroring, transposition, and random rotation are adopted to avoid over-fitting and enhance the model's generalization ability. The integration of dice loss and binary cross-entropy (BCE) loss is used as the unified cost function for training the model. It is defined as:

$$\text{L} = \text{L}_{\text{Dice}} + \alpha \text{L}_{\text{BCE}} \tag{6}$$

where α, a weight constant, is used to adjust the trade-off between the dice loss and the BCE loss. The α is set to 0.8.

The dice similarity coefficient (DSC), sensitivity (SEN), specificity (SPE), accuracy (ACC), Hausdorff distance (HAU), and relative area difference (ΔA) metrics are adopted to conduct the quantitative evalu-

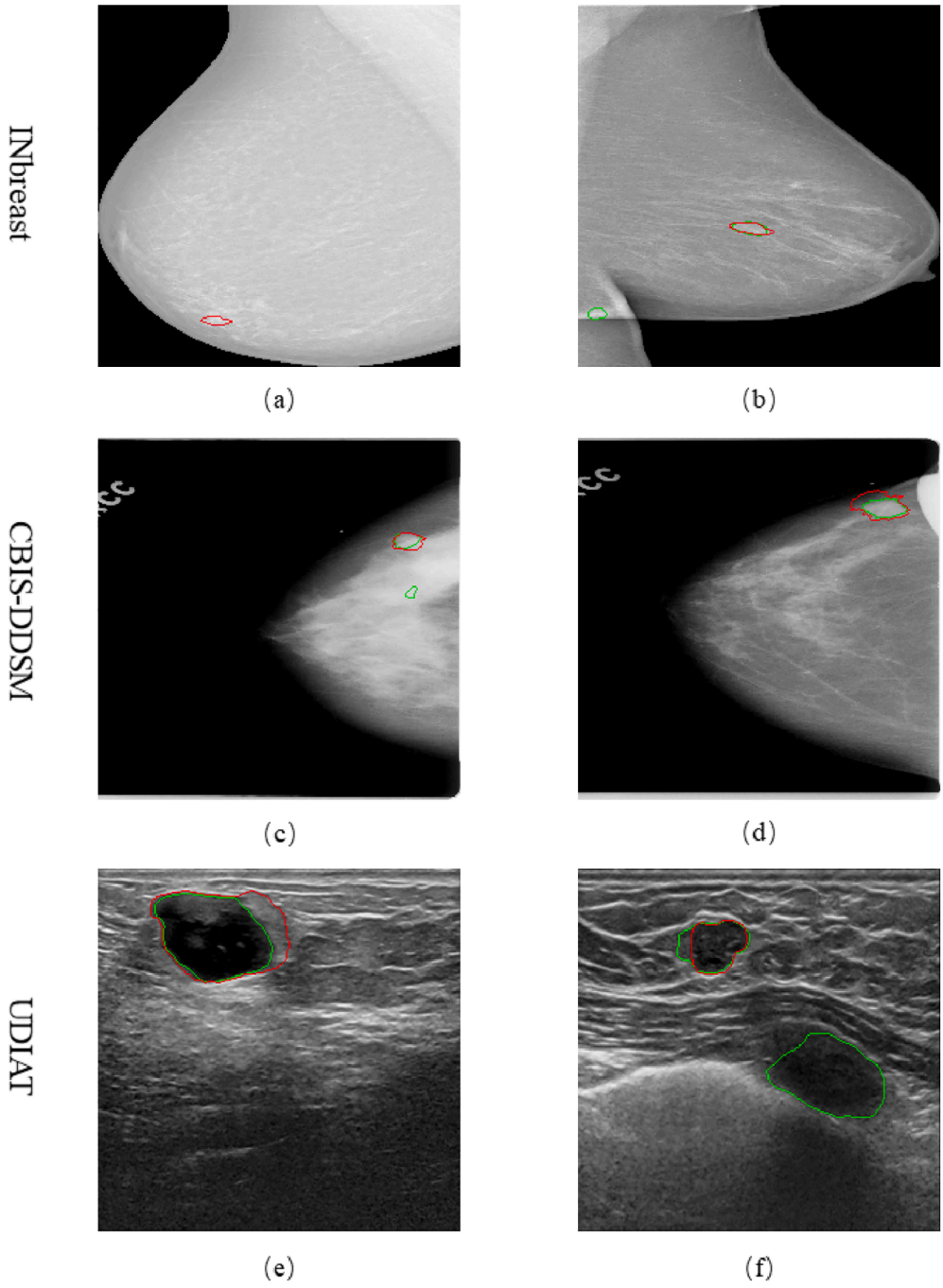


Fig. 5. Several bad segmentation cases. (a) and (b) are from the INbreast dataset; (c) and (d) are from CBIS-DDSM dataset; (e) and (f) are from UDIAT dataset. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

ation in our study. Their mathematical definitions are as follows:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{7}$$

$$SEN = \frac{TP}{TP + FN} \tag{8}$$

$$SPE = \frac{TN}{TN + FP} \tag{9}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\Delta A = \frac{|(TP + FP) - (TP + FN)|}{TP + FN} \tag{11}$$

$$HAU = \max(\mathrm{h}(pred, gt), \mathrm{h}(gt, pred)) \tag{12}$$

where $\mathrm{h}(A, B) = \max_{a \in A}\{\min_{b \in B}||a - b||\}$ and $||\cdot||$ denotes the euclidean distance between two pixels. *pred* refers to the predicted segmentation results, and *gt* is the ground-truth masks.

4.3. Experimental results

4.3.1. Results on CBIS-DDSM database

We establish a comparison between our proposed network and

Table 4
Comparison between the developed network and several state-of-the-art segmentation network on the INbreast database. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
UNet [49]	70.05 ± 3.24	98.43 ± 0.63	70.87 ± 3.7	99.31 ± 0.28	30.27 ± 12.63	5.12 ± 1.44
Attention-UNet [38]	82.8 ± 5.21	98.97 ± 0.76	79.81 ± 4.07	**99.64 ± 0.29**	16.06 ± 5.12	2.83 ± 1.25
DeepLabv3Plus [12]	76.4 ± 4.36	98.76 ± 0.67	72.1 ± 5.01	99.54 ± 0.34	22.38 ± 5.88	4.07 ± 1.74
FCN [31]	79.3 ± 5.74	98.76 ± 0.97	76.04 ± 3.09	99.51 ± 0.54	18.46 ± 7.54	2.47 ± 0.56
UNetPlusPlus [19]	81.75 ± 4.8	98.94 ± 0.72	78.25 ± 5.02	99.63 ± 0.28	15.99 ± 5.84	2.8 ± 1.16
scSENet [50]	83.31 ± 4.6	98.93 ± 0.85	80.79 ± 1.67	99.54 ± 0.51	14.65 ± 6.9	2.28 ± 0.94
Ours(w/o)	85.06 ± 2.64	99.11 ± 0.58	83 ± 1.99	99.62 ± 0.22	**11.56 ± 3.79**	1.81 ± 0.85
Ours(w/)	**86.1 ± 2.75**	**99.14 ± 0.52**	**85.82 ± 2.97**	99.54 ± 0.39	14.34 ± 5.08	**1.8 ± 0.9**

w/o – train from scratch.
w/ – with pretraining on CBIS-DDSM.

Table 5
Comparison between the developed network and several state-of-the-art segmentation network on the CBIS-DDSM database. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
UNet [49]	79.41 ± 2.08	99.74 ± 0.02	80.48 ± 4.99	99.88 ± 0.02	20.66 ± 2.33	2.66 ± 0.34
Attention-UNet [38]	84.3 ± 0.36	99.8 ± 0.0	87.14 ± 0.85	99.89 ± 0.0	18.75 ± 2.0	1.91 ± 0.47
DeepLabv3Plus [12]	78.59 ± 1.75	99.74 ± 0.02	80.25 ± 2.82	99.88 ± 0.01	21.27 ± 3.29	2.44 ± 2.85
FCN [31]	80.98 ± 0.24	99.76 ± 0.01	81.7 ± 1.03	99.89 ± 0.01	20.44 ± 0.81	1.61 ± 0.11
UNetPlusPlus [19]	83.05 ± 0.37	99.78 ± 0.0	86.99 ± 0.89	99.88 ± 0.0	18.99 ± 0.69	2.02 ± 0.18
scSENet [50]	84.34 ± 0.13	99.8 ± 0.0	88.67 ± 1.39	99.88 ± 0.01	19.39 ± 0.31	1.51 ± 0.47
Ours	**85.75 ± 0.22**	**99.81 ± 0.0**	**88.91 ± 0.76**	99.89 ± 0.01	**14.36 ± 1.96**	**0.93 ± 0.46**

Table 6
Comparison between the developed ARF-Net and several state-of-the-art segmentation network on the UDIAT database.

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑
UNet [49]	80.09	98.34	81.31	99.2
Attention-UNet [38]	80.64	98.35	81.09	99.12
DeepLabv3Plus [12]	77.36	98.08	78.09	98.99
FCN [31]	83.39	98.43	83.94	99.1
UNetPlusPlus [19]	79.45	98.06	86.27	98.64
scSENet [50]	84.86	98.52	86.36	99.09
Ours	**88.12**	**98.84**	**89.44**	**99.36**

several related approaches on the CBIS-DDSM benchmark. Table 1 depicts the experimental results in detail.

It can be found from Table 1, our proposed network exceeds the rest on the three evaluation metrics, with very significant DSC, SEN, and SPE (85.75%, 88.91%, and 99.89%, respectively), and has a substantial improvement of 1.26%, 3.68%, and 0.03%. It is noteworthy that our proposed model can remarkably enhance the true positive rate (sensitivity), which is essential to breast mass detection tasks. Moreover, our proposed approach decreases the ΔA and HAU by 8.34% and 1.5, separately. These experimental results show that our designed ARF-Net can work better for mass segmentation in whole mammograms than other methods and also achieve high breast mass detection accuracy.

4.3.2. Results on INbreast database

To confirm our proposed network's effectiveness, it is compared with several related approaches on the INbreast benchmark. The experimental results are presented in Table 2.

Table 2 illustrates that without any pre-training process, the DSC, ACC, SEN, SPE, ΔA, and HAU of our proposed model are 85.06%, 99.11%, 83%, 99.62%, 11.56%, and 1.81. With pre-training on the CBIS-DDSM database, our proposed approach acquires the 86.1%, 99.14%, 85.82%, 99.54%, 14.34%, and 1.8 on the DSC, ACC, SEN, SPE, ΔA, and HAU metrics. Therefore, we can conclude that transfer learning can enhance the performance of mass segmentation in whole mammograms. Our proposed model has an 2.18%, 3.1%, and 0.06% improvement in the DSC, SEN, and SPE when compared to other methods. Notably, our proposed approach improves the true positive rate significantly. There is an enormous reduction of the ΔA and HAU by 19.54% and 0.06. The experimental results demonstrate that ARF-Net is more robust and effective when compared against other networks.

4.3.3. Results on UDIAT database

Table 3 summarizes the segmentation results of our designed ARF-Net and several related approaches.

Our proposed ARF-Net obtains a better performance of DSC and ACC than several related approaches and has the improvements of 1.3% and 0.34%. Because a false negative will decrease the possibility of early diagnosis, detection, and treatment of breast cancer and the true positive rate (sensitivity) is essential to breast mass detection tasks, the true positive rate is vitally crucial. On the sensitivity measurement, our proposed ARF-Net attains 89.44% and is only second to the approach of Singh et al. [52]. In conclusion, our proposed ARF-Net also generalizes well on UDIAT.

4.4. Discussions and analyses

Fig. 5 shows six bad segmentation cases. Two bad segmentation cases (a) and (c) arise from the high-density region in whole mammograms. Breast masses are covered with high-density areas in whole mammographic images. The bad case (b) is caused by the pectoral muscle. The brightness of some breast masses is the same as the pectoral muscle in the mammograms. The bad case (d) results from the incorrect ground truth mask. There are many coarse ground-truth labels in the CBIS-DDSM dataset. The bad segmentation example (f) is likely caused by the breast fatty tissue (Breast fat appears dark gray on breast ultrasound images). The cysts and ducts are likely the reasons for the bad segmentation case (e).

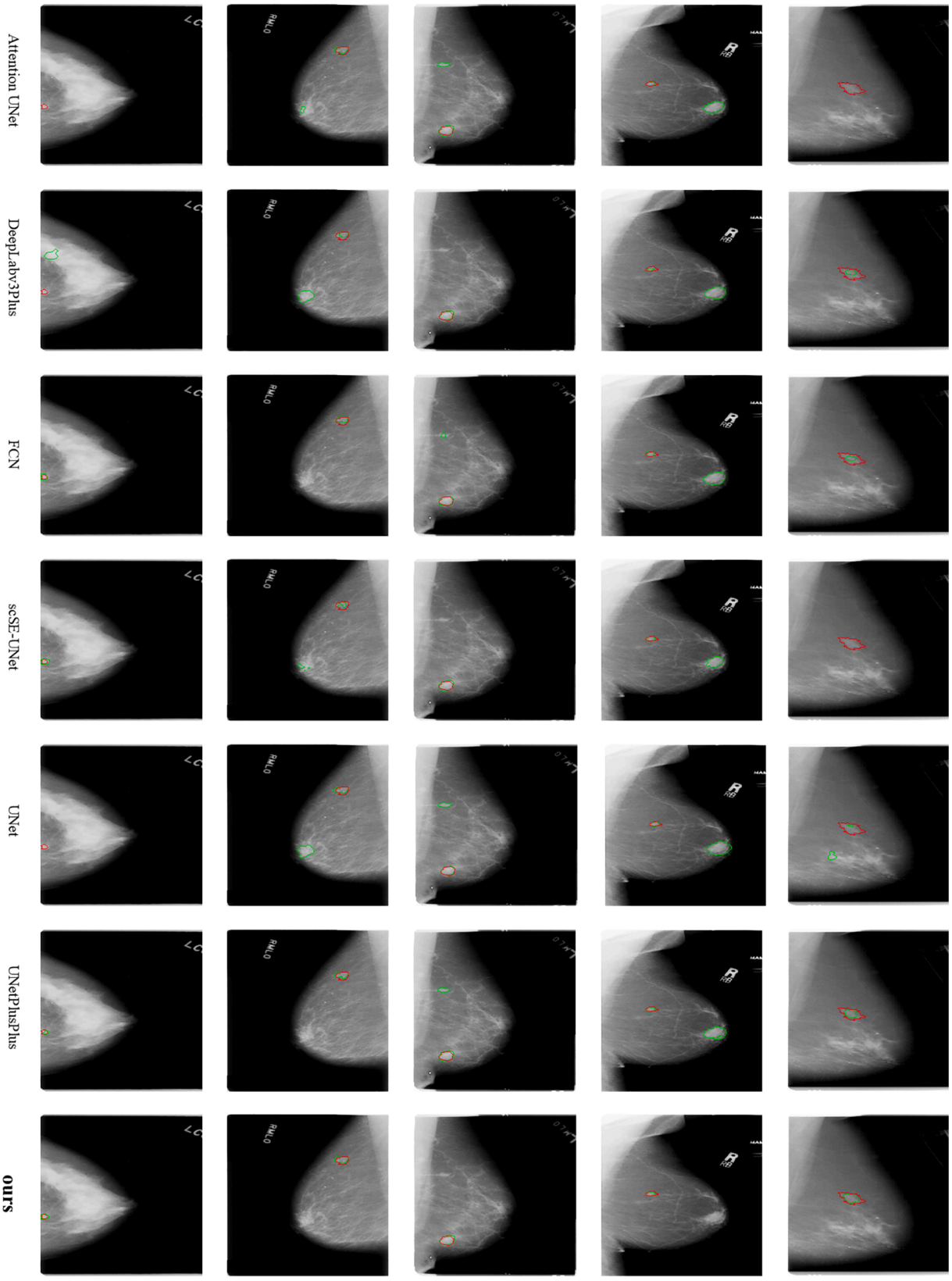


Fig. 6. The visualization results of our proposed ARF-Net and several well-known segmentation models on the CBIS-DDSM benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

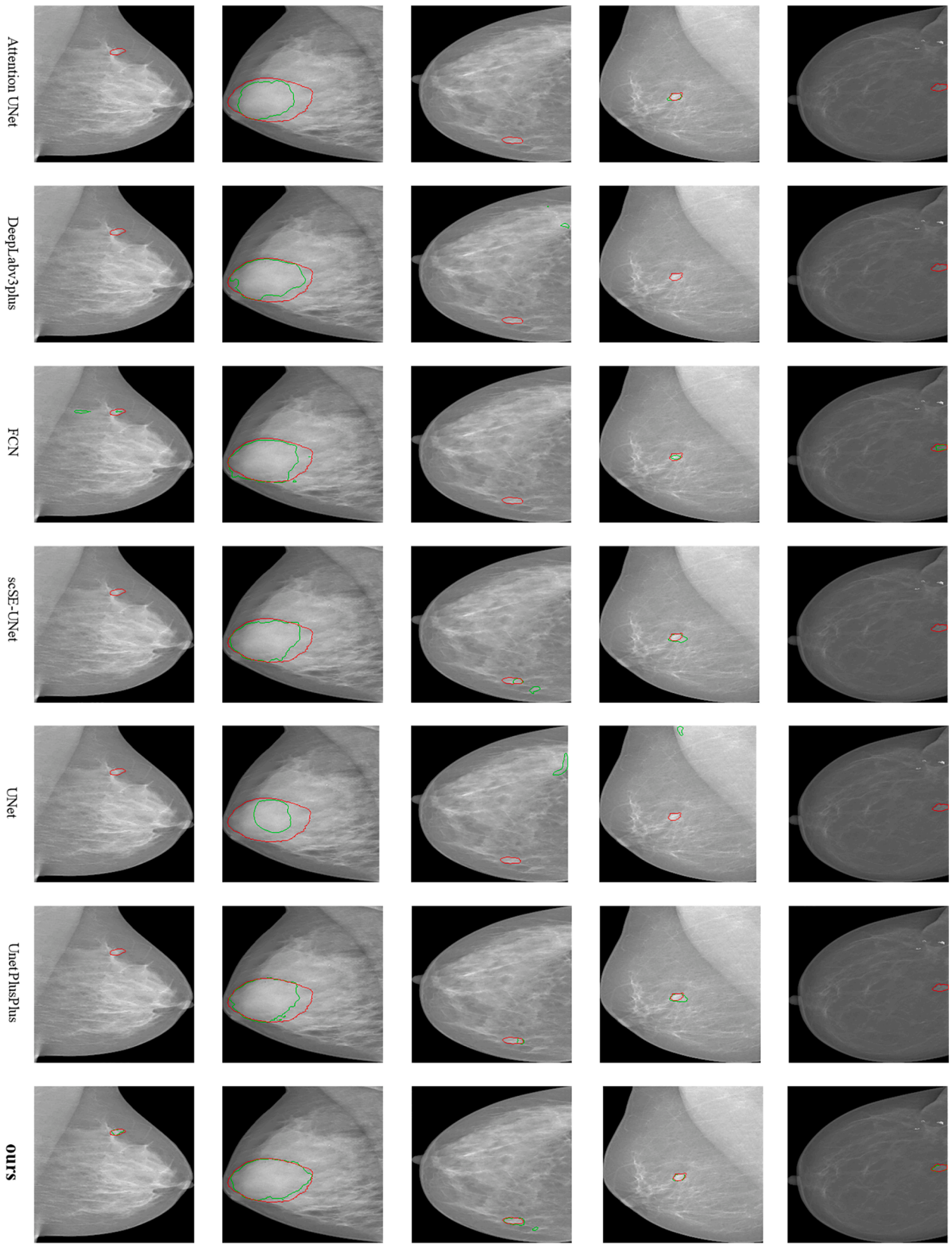


Fig. 7. The visualization results of our proposed ARF-Net and some excellent segmentation models on the INbreast benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

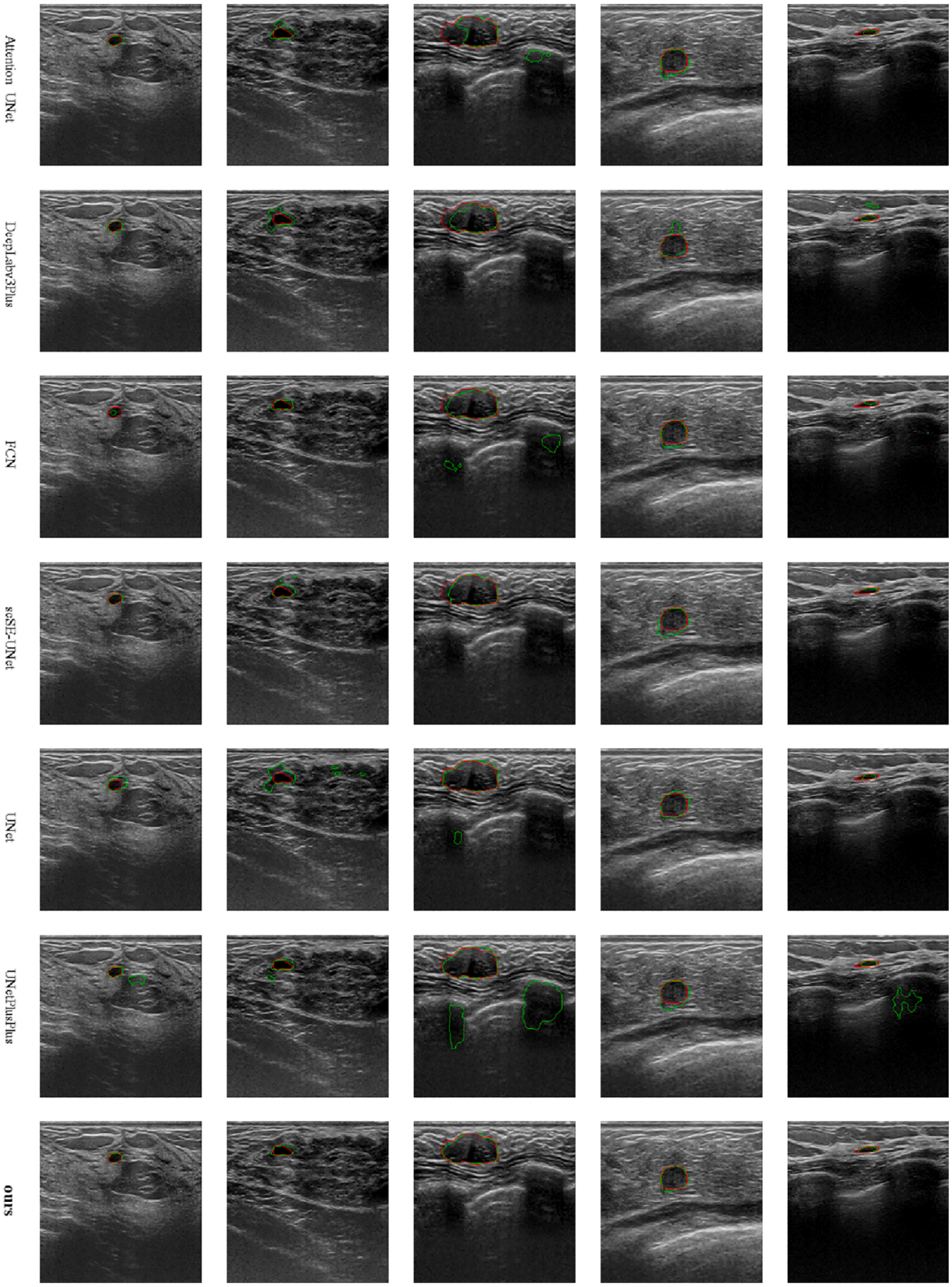


Fig. 8. The visualization results of our proposed ARF-Net and several well-known segmentation models on the UDIAT benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

Table 7
Comparison between our proposed MSSM and several counterparts on the INbreast dataset. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
SE Module [22]	83.54 ± 4.16	98.96 ± 0.81	82.85 ± 1.56	99.45 ± 0.64	17.71 ± 10.01	1.9 ± 0.79
GCT [62]	83.11 ± 4.19	99 ± 0.77	79.82 ± 3.71	99.58 ± 0.45	15.11 ± 5.43	2.74 ± 1.06
Ours(w/o)	85.06 ± 2.64	99.11 ± 0.58	83 ± 1.99	**99.62 ± 0.22**	**11.56 ± 3.79**	1.81 ± 0.85
Ours(w/)	**86.1 ± 2.75**	**99.14 ± 0.52**	**85.82 ± 2.97**	99.54 ± 0.39	14.34 ± 5.08	**1.8 ± 0.9**

w/o – train from scratch.
w/ – with pre-training on the CBIS-DDSM benchmark.

Table 8
Comparison between our proposed MSSM and several counterparts on the CBIS-DDSM benchmark. (mean±std)

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑	ΔA(%)↓	HAU↓
SE Module [22]	84.04 ± 0.23	99.79 ± 0.01	86.19 ± 0.51	99.89 ± 0.01	17.53 ± 1.01	1.71 ± 0.08
GCT [62]	84.04 ± 0.1	99.8 ± 0.01	86.49 ± 1.04	99.89 ± 0.01	18.2 ± 0.98	2.36 ± 0.01
Ours	**85.75 ± 0.22**	**99.81 ± 0.0**	**88.91 ± 0.76**	99.89 ± 0.01	**14.36 ± 1.96**	**0.93 ± 0.46**

Table 9
Comparison between our proposed MSSM and several counterparts on the UDIAT benchmark.

Method	DSC(%)↑	ACC(%)↑	SEN(%)↑	SPE(%)↑
SE Module [22]	84.43	98.56	84.8	99.15
GCT [62]	84.87	98.56	87.3	99.06
Ours	**88.12**	**98.84**	**89.44**	**99.36**

5. Ablation studies

In this section, we will first compare the designed model against several excellent segmentation models. Then, we compare the developed MSSM with several well-known channel attention modules. Finally, we will analyze the speed of the proposed model.

5.1. Comparison with other state-of-the-art methods

We compare the developed ARF-Net against several state-of-the-art semantic segmentation networks including UNet [49], Attention-UNet [38], DeepLabv3Plus [12], FCN [31], UNetPlusPlus [19], and scSENet [50] to further prove its effectiveness and performance. In order to provide a fair comparison, we first access their public implementations and then re-train their models on our databases to obtain the segmentation results of compared methods. Tables 4–6 depict the results of the quantitative comparison on the INbreast, CBIS-DDSM, and UDIAT databases, respectively.

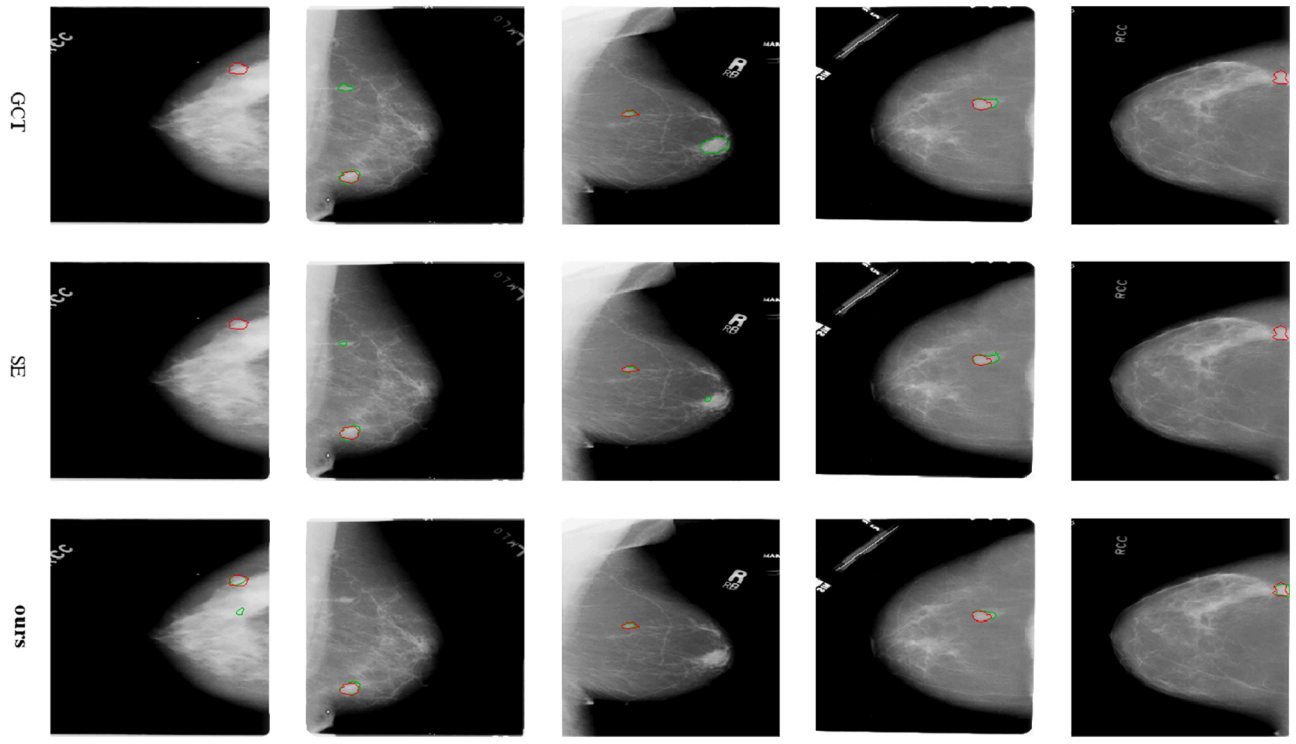


Fig. 9. The visualization results of our proposed MSSM and several counterparts on the CBIS-DDSM benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

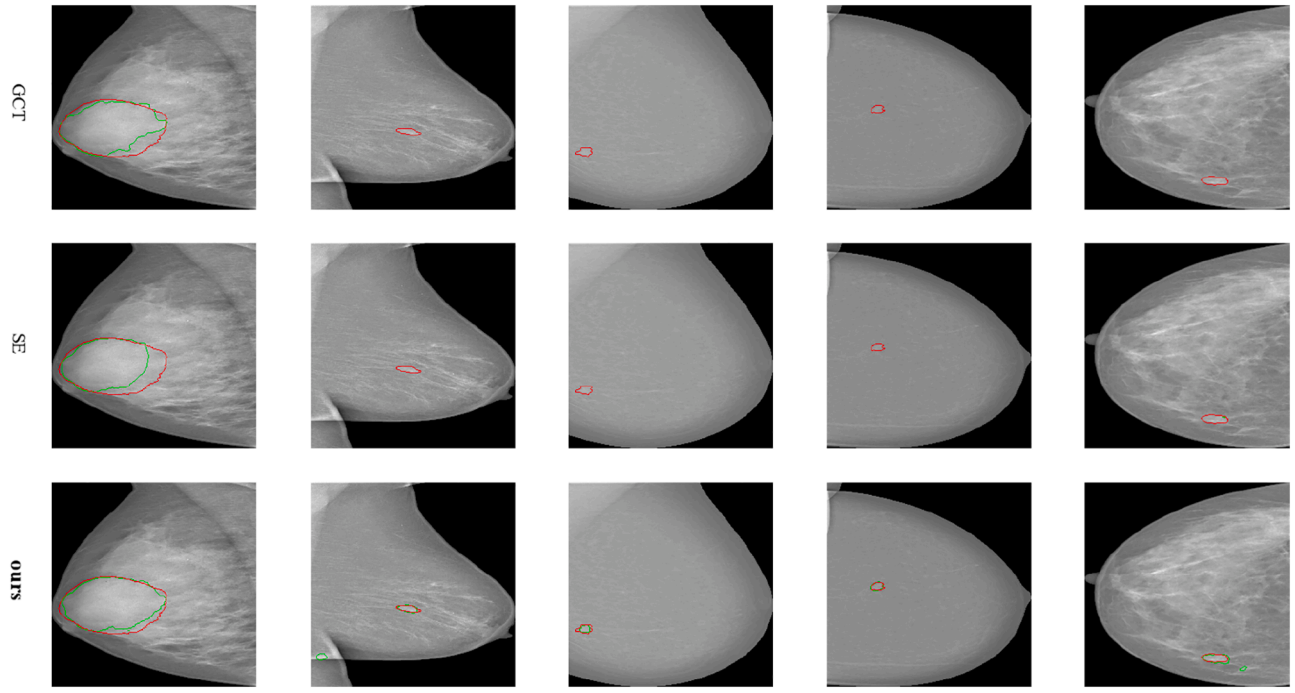


Fig. 10. The visualization results of our proposed MSSM and several counterparts on the INbreast benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

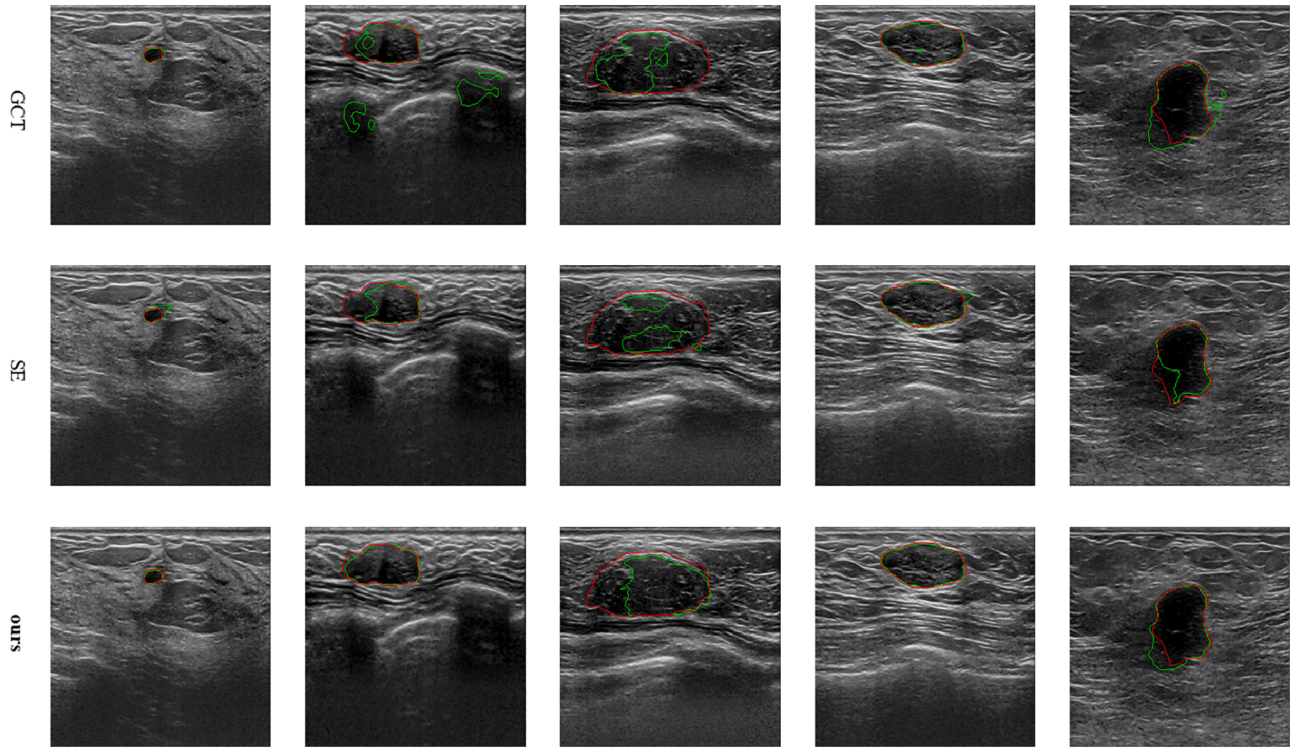


Fig. 11. The visualization results of our proposed MSSM and several counterparts on the UDIAT benchmark. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

As is depicted in Table 4, compared to several state-of-the-art segmentation approaches, the DSC, ACC, and SEN of our designed model have an improvement of at least 2.79%, 0.17%, and 5.03% on the INbreast database. Besides, the ΔA and HAU are reduced significantly by at least 3.09% and 0.48, respectively, in comparison with other segmentation methods for the INbreast dataset. Table 5 shows that our proposed network offers an improvement of at least 1.41%, 0.01%, 0.24% in terms of the DSC, ACC, and SEN metrics when compared with other state-of-the-art segmentation approaches on the CBIS-DDSM database. In addition, there is also an enormous decrease on the ΔA and HAU metrics by at least 4.39% and 0.58. It is clear from Table 6 that ARF-Net exceeds all the other models in all performance. Compared

Table 10
Speed Comparison of our method against other methods. Image size is 256 × 256.

Method	INbreast		CBIS-DDSM		UDIAT		ms↓	fps
	DSC(%)	SEN(%)	DSC(%)	SEN(%)	DSC(%)	SEN(%)		
UNet [49]	70.05	70.87	79.41	80.48	80.09	81.31	5.8	171
Attention-UNet [38]	82.8	79.81	84.3	87.14	80.64	81.09	12.0	83
DeepLabv3Plus [12]	76.4	72.1	78.59	80.25	77.36	78.09	17.1	58
FCN [31]	79.3	76.04	80.98	81.7	83.39	83.94	3.8	266
UNetPlusPlus [19]	81.75	78.25	83.05	86.99	79.45	86.27	8.2	122
scSENet [50]	83.31	80.79	84.34	88.67	84.86	86.36	6.5	153
Ours	**86.1**	**85.82**	**85.75**	**88.91**	**88.12**	**89.44**	8.0	125

with the other models, the proposed ARF-Net obtains improvements of at least 3.26%, 0.32%, 3.08%, and 0.16% on all evaluation metrics. It is more important that our proposed approach obtains a high true positive rate (sensitivity), as a false negative will reduce the possibility of early diagnosis, detection, and treatment of breast cancer.

To prove more directly that our proposed model is superior to other segmentation networks, we visualize the predicted segmentation results of all the models on the CBIS-DDSM, INbreast, and UDIAT databases. The red closed curve delineates the boundary of the ground-truth maps, and the green closed curve delineates the boundary of the predicted results. The visual contrasts of segmentation on the CBIS-DDSM, INbreast, and UAIDT benchmarks are displayed in Figs. 6–8, separately.

From Fig. 6, we can observe that our developed network is not very sensitive to the high-density region and the glandular tissue when compared with several segmentation approaches. From Fig. 7, we can find that our developed network can detect the breast masses of smaller size while other segmentation model fails to segment the breast masses of smaller size. Moreover, our proposed model can obtain better segmentation results of large-size breast masses than other segmentation methods. It can be observed from Fig. 8 that our proposed ARF-Net can obtain better segmentation results than other methods.

In conclusion, the quantitative results and qualitative visualization show that our designed approach can obtain a better breast mass segmentation performance than other segmentation models.

5.2. Comparison between MSSM and other counterparts

To demonstrate the effectiveness of the designed MSSM, we build a comparison between MSSM and several excellent counterparts including SE Module [22] and GCT [62] on the INbreast [36] and CBIS-DDSM [27] databases. In order to provide a fair comparison, we first access their public implementations and then re-train their models on our databases to obtain the segmentation results of compared methods. In the following experiments, the MSSM will be replaced by the SE Module and GCT, respectively. The experimental results on the INbreast [36], CBIS-DDSM [27], and UDIAT [63] benchmarks are displayed in Tables 7–9, respectively.

From Table 7, we can find that the designed MSSM exceeds other counterparts and yield an improvement of at least 5.26%, 0.18%, 2.97%, and 0.04% in the field of the DSC, ACC, SEN, and SPE on the INbreast dataset. Moreover, the ΔA and HAU metrics of our proposed MSSM reduce by at least 3.55% and 0.1 over other counterparts. Table 8 shows that our proposed model increases by at least 1.71%, 0.01%, and 2.42% in terms of the DSC, ACC, and SEN on the CBIS-DDSM database. Our proposed MSSM obtains a massive reduction with the ΔA of 3.17% and the HAU of 0.78. From Table 9, we can observe that the MSSM increases DSC, ACC, SEN, and SPE by at least 3.25%, 0.28%, 2.14%, and 0.21%. In particular, compared to several counterparts on the INbreast, CBIS-DDSM, and UDIAT databases, our developed MSSM can significantly improve the true positive rate.

On the CBIS-DDSM, INbreast, and UDIAT benchmarks, the qualitative improvements of our proposed network over other segmentation methods are displayed in Figs. 9–11, respectively. The red closed curve delineates the boundary of the ground-truth mask, and the green closed curve delineates the boundary of the predicted results.

From Fig. 9, we can find that the designed MSSM can detect the masses with smaller size better and is not more susceptible to interference from the high-density region when compared to the GCT and SE module for the CBIS-DDSM database. It can be found from Fig. 10 that, on the INbreast database, our developed MSSM can successfully detect the breast masses with smaller sizes, but the GCT and SE modules hardly detect the small breast mass. However, there are two cases in which the pectoral muscle and glandular tissue are mistakenly recognized as the masses. From Fig. 11, we can observe that the proposed MSSM can obtain better segmentation results than the GCT and SE modules. However, there is one case in which the breast fatty tissue is mistakenly recognized as the masses.

In summary, our proposed MSSM is effective for segmentation.

5.3. Speed analysis

The frames per second (FPS) is used to measure the speed of the model. For a fair comparison, we chose the 256 × 256 as the resolution of the input image. We analyze the speed of the model through the variations of dice score and sensitivity index. Table 10 presents the speed comparison between our method with other approaches on the RTX 2080Ti GPU.

Compared with Attention-UNet, DeepLabv3Plus, and UNetPlusPlus, our proposed ARF-Net has lower latency, while having better performance. For the INbreast, CBIS-DDSM, and UDIAT datasets, the proposed ARF-Net obtains improvements of at least 3.3% and 6.01%, 1.45% and 1.77%, and 7.48% and 3.17% on the DSC and SEN index, respectively. Compared with FCN, in spite of higher latency, our proposed ARF-Net has better performance. The reasons behind high latency are that our proposed method adopts the symmetric encoder-decoder architecture, while FCN adopts the asymmetric encoder architecture. On the index of the DSC and SEN, ARF-Net obtains improvements of 6.8% and 9.78%, 4.77% and 7.21%, and 4.73% and 5.5% on the INbreast, CBIS-DDSM, and UDIAT. Compared with UNet and scSENet, in spite of a little higher latency, our proposed method has better performance. The proposed ARF-Net obtains improvements of the DSC and SEN of at least 2.79% and 5.03%, 1.41% and 0.24%, and 3.26% and 3.08% on the INbreast, CBIS-DDSM, and UDIAT databases.

In conclusion, our proposed ARF-Net achieves the state-of-the-art segmentation performance with the relatively high speed.

6. Conclusion

A novel end-to-end network, ARF-Net for precise breast mass segmentation in whole mammograms and ultrasound images, is proposed in this work. The ARF-Net consists of an encoder network and a corresponding decoder network. The SRFM in ARF-Net can adaptively

allocate the appropriate sizes of receptive fields for objects of different sizes. The SRFM composes of a MRFM for generating multiple receptive fields of different sizes and a MSSM for selecting the suitable receptive field's size. The proposed approach is tested on two mammographic databases (INbreast and CBIS-DDSM) and one ultrasonic database (UDIAT). The experimental results show that the developed network exceeds several related methods. Extensive ablation experiments demonstrate that our developed ARF-Net exceeds several excellent segmentation networks, and our proposed MSSM also exceeds several counterparts. In conclusion, our proposed model is of effectiveness.

CRediT authorship contribution statement

Chunbo Xu: Conceptualization, Methodology, Software, Writing - original draft, Visualization, Validation, Writing - review & editing. **Yunliang Qi:** Visualization, Investigation. **Yiming Wang:** Investigation. **Meng Lou:** Data curation, Validation. **Jiande Pi:** Investigation. **Yide Ma:** Supervision, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Nos. 61961037] and the Natural Science Foundation of Gansu Province [No. 18JR3RA288]. The authors gratefully acknowledge the Breast Research Group, INESC Porto, Portugal for the INbreast database.

References

[1] D. Abdelhafiz, S. Nabavi, R. Ammar, C. Yang, J. Bi, Residual deep learning system for mass segmentation and classification in mammography, in: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019, pp. 475–484.
[2] M.A. Al-Masni, M.A. Al-Antari, J.M. Park, G. Gi, T.Y. Kim, P. Rivera, E. Valarezo, M.T. Choi, S.M. Han, T.S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system, Computer Methods and Programs in Biomedicine 157 (2018) 85–94.
[3] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495.
[4] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection, 2020, arXiv preprint arXiv:2004.10934.
[5] K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, M. Sallam, K. Chang, K. Woods, The digital database for screening mammography, in: Third International Workshop on Digital Mammography, 1996, p. 27.
[6] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network, Biomedical Signal Processing and Control 61 (2020), 102027.
[7] A. Cao, Q. Song, X. Yang, S. Liu, Breast mass segmentation on digital mammograms by a combined deterministic annealing method, in: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), IEEE, 2004, pp. 1303–1306.
[8] J. Chen, L. Chen, S. Wang, P. Chen, A novel multi-scale adversarial networks for precise segmentation of x-ray breast mass, IEEE Access 8 (2020) 103772–103781.
[9] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, 2014, arXiv preprint arXiv:1412.7062.
[10] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2017) 834–848.
[11] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587.
[12] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, 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.
[13] S. Dalmiya, A. Dasgupta, S.K. Datta, Application of wavelet based k-means algorithm in mammogram segmentation, International Journal of Computer Applications 52 (2012).
[14] J. Ding, H.D. Cheng, J. Huang, J. Liu, Y. Zhang, Breast ultrasound image classification based on multiple-instance learning, Journal of Digital Imaging 25 (2012) 620–627.
[15] C. Gao, H. Ye, F. Cao, C. Wen, Q. Zhang, F. Zhang, Multiscale fused network with additive channel–spatial attention for image segmentation, Knowledge-Based Systems 214 (2021), 106754.
[16] S. Gao, M.M. Cheng, K. Zhao, X.Y. Zhang, M.H. Yang, P.H. Torr, Res2net: A new multi-scale backbone architecture, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019).
[17] D. Guliato, R.M. Rangayyan, W.A. Carnielli, J.A. Zuffo, J.L. Desautels, Segmentation of breast tumors in mammograms using fuzzy sets, Journal of Electronic Imaging 12 (2003) 369–378.
[18] X. Hao, Y. Shen, S.r. Xia, Automatic mass segmentation on mammograms combining random walks and active contour, Journal of Zhejiang University SCIENCE C 13 (2012) 635–648.
[19] S. Hasan, C.A. Linte, U-netplus: a modified encoder-decoder u-net architecture for semantic and instance segmentation of surgical instrument, 2019, arXiv preprint arXiv:1902.08994.
[20] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[22] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[23] H. Huang, H. Chen, H. Xu, Y. Chen, Q. Yu, Y. Cai, Q. Zhang, Cross-tissue/organ transfer learning for the segmentation of ultrasound images using deep residual u-net, Journal of Medical and Biological Engineering, 1–9.
[24] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.
[25] L. Kinnard, S.C. Lo, P. Wang, M.T. Freedman, M. Chouikha, Automatic segmentation of mammographic masses using fuzzy shadow and maximum-likelihood analysis, in: Proceedings IEEE International Symposium on Biomedical Imaging, IEEE, 2002, pp. 241–244.
[26] H. Lee, J. Park, J.Y. Hwang, Channel attention module with multi-scale grid average pooling for breast cancer segmentation in an ultrasound image, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2020).
[27] R.S. Lee, F. Gimenez, A. Hoogi, K.K. Miyake, M. Gorovoy, D.L. Rubin, A curated mammography data set for use in computer-aided detection and diagnosis research, Scientific Data 4 (2017) 1–9.
[28] F. Li, W. Li, Y. Shu, S. Qin, B. Xiao, Z. Zhan, Multiscale receptive field based on residual network for pancreas segmentation in ct images, Biomedical Signal Processing and Control 57 (2020), 101828.
[29] L. Liu, F.X. Wu, Y.P. Wang, J. Wang, Multi-receptive-field cnn for semantic segmentation of medical images, IEEE Journal of Biomedical and Health Informatics 24 (2020) 3215–3225.
[30] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.
[31] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[32] I. Loshchilov, F. Hutter, Fixing weight decay regularization in adam, 2018.
[33] A. Mencattini, G. Rabottino, M. Salmeri, R. Lojacono, E. Colini, Breast mass segmentation in mammographic images by an effective region growing algorithm, in: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2008, pp. 948–957.
[34] A.A. Mohamed, W.A. Berg, H. Peng, Y. Luo, R.C. Jankowitz, S. Wu, A deep learning method for classifying mammographic breast density categories, Medical Physics 45 (2018) 314–321.
[35] W.K. Moon, Y.W. Shen, C.S. Huang, L.R. Chiang, R.F. Chang, Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images, Ultrasound in Medicine & Biology 37 (2011) 539–548.
[36] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. Cardoso, J.S. Cardoso, Inbreast: toward a full-field digital mammographic database, Academic Radiology 19 (2012) 236–248.
[37] Z. Ning, K. Wang, S. Zhong, Q. Feng, Y. Zhang, Cf2-net: Coarse-to-fine fusion convolutional network for breast ultrasound image segmentation, 2020, arXiv preprint arXiv:2003.10144.
[38] 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, 2018, arXiv preprint arXiv:1804.03999.
[39] F.M. Osman, M.H. Yap, Adjusted quick shift phase preserving dynamic range compression method for breast lesions segmentation, Informatics in Medicine Unlocked 100344 (2020).
[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, 2019, pp. 8026–8037.
[41] B.C. Patel, G. Sinha, An adaptive k-means clustering algorithm for breast image segmentation, International Journal of Computer Applications 10 (2010) 35–38.
[42] P. Rahmati, A. Adler, G. Hamarneh, Mammography segmentation with maximum likelihood active contours, Medical Image Analysis 16 (2012) 1167–1186.

[43] A. Rampun, K. López-Linares, P.J. Morrow, B.W. Scotney, H. Wang, I.G. Ocaña, G. Maclair, R. Zwiggelaar, M.A.G. Ballester, I. Macía, Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network, Medical Image Analysis 57 (2019) 1–17.
[44] N. Ravitha Rajalakshmi, R. Vidhyapriya, N. Elango, N. Ramesh, Deeply supervised u-net for mass segmentation in digital mammograms, International Journal of Imaging Systems and Technology (2020).
[45] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[46] J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271.
[47] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, 2018, arXiv preprint arXiv:1804.02767.
[48] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, 2015, arXiv preprint arXiv:1506.01497.
[49] O. Ronneberger, P. Fischer, 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.
[50] A.G. Roy, N. Navab, 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.
[51] B. Shareef, A. Vakanski, M. Xian, P.E. Freer, Estan: Enhanced small tumor-aware network for breast ultrasound image segmentation, 2020, arXiv preprint arXiv: 2009.12894.
[52] V.K. Singh, M. Abdel-Nasser, F. Akram, H.A. Rashwan, M.M.K. Sarker, N. Pandey, S. Romani, D. Puig, Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework, Expert Systems with Applications 162 (2020), 113870.
[53] R. Su, D. Zhang, J. Liu, C. Cheng, Msu-net: Multi-scale u-net for 2d medical image segmentation, Frontiers in Genetics 12 (2021) 140.
[54] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D.D. Feng, S. Wang, Aunet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms, Physics in Medicine & Biology 65 (2020), 055005.
[55] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[58] S. Timp, N. Karssemeijer, A new 2d segmentation method based on dynamic programming applied to computer aided detection in mammography, Medical Physics 31 (2004) 958–971.
[59] Y. Wang, S. Wang, J. Chen, C. Wu, Whole mammographic mass segmentation using attention mechanism and multiscale pooling adversarial network, Journal of Medical Imaging 7 (2020), 054503.
[60] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
[61] S. Xu, E. Adeli, J.Z. Cheng, L. Xiang, Y. Li, S.W. Lee, D. Shen, Mammographic mass segmentation using multichannel and multiscale fully convolutional networks, International Journal of Imaging Systems and Technology 30 (2020) 1095–1107.
[62] Z. Yang, L. Zhu, Y. Wu, Y. Yang, Gated channel transformation for visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11794–11803.
[63] M.H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A.K. Davison, R. Marti, Automated breast ultrasound lesions detection using convolutional neural networks, IEEE Journal of Biomedical and Health Informatics 22 (2017) 1218–1226.
[64] J. Yuan, Z. Deng, S. Wang, Z. Luo, Multi receptive field network for semantic segmentation, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2020, pp. 1883–1892.
[65] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.