




A hybrid attentional guidance network for tumors segmentation of breast ultrasound images

Yaosheng Lu^{1,2} · Xiaosong Jiang² · Mengqiang Zhou² · Dengjiang Zhi² · Ruiyu Qiu² · Zhanhong Ou² · Jieyun Bai^{1,2} 

Received: 1 November 2022 / Accepted: 31 January 2023
© CARS 2023

Abstract

Purpose In recent years, breast cancer has become the greatest threat to women. There are many studies dedicated to the precise segmentation of breast tumors, which is indispensable in computer-aided diagnosis. Deep neural networks have achieved accurate segmentation of images. However, convolutional layers are biased to extract local features and tend to lose global and location information as the network deepens, which leads to a decrease in breast tumors segmentation accuracy. For this reason, we propose a hybrid attention-guided network (HAG-Net). We believe that this method will improve the detection rate and segmentation of tumors in breast ultrasound images.

Methods The method is equipped with multi-scale guidance block (MSG) for guiding the extraction of low-resolution location information. Short multi-head self-attention (S-MHSA) and convolutional block attention module are used to capture global features and long-range dependencies. Finally, the segmentation results are obtained by fusing multi-scale contextual information.

Results We compare with 7 state-of-the-art methods on two publicly available datasets through five random fivefold cross-validations. The highest dice coefficient, Jaccard Index and detect rate ($82.6 \pm 1.9\%$, $74.2 \pm 2.1\%$, $92.1 \pm 2.2\%$ and $77.8 \pm 3.1\%$, $66.8 \pm 3.2\%$, $91.9 \pm 5.0\%$, separately) obtained on two publicly available datasets (BUSI and OASUBD), prove the superiority of our method.

Conclusion HAG-Net can better utilize multi-resolution features to localize the breast tumors. Demonstrating excellent generalizability and applicability for breast tumors segmentation compare to other state-of-the-art methods.

Keywords Breast tumors · Attentional mechanisms · Ultrasound images · Long-range dependences

Introduction

Breast cancer is the biggest threat to women's health. According to a global study [1], breast cancer account for 11.7% of the total cancer incident rate, has surpassed lung cancer and becomes the most common cancer. Early diagnosis can assist

physicians in providing clinical decisions, treatments and rehabilitation plans, which is important in reducing mortality. Ultrasound imaging technology is a non-invasive, inexpensive and fast method that is widely used in various types of medical examination. For radiologists, the diagnosis of breast tumors in ultrasound images is time-consuming, challenging and more subjective. In recent years, computer-aided diagnostic (CAD) systems have been developed to simplify the operation, and reliable results can be obtained.

Segmentation of breast tumors is important for the diagnosis of breast cancer. As shown in Fig. 1, due to scattered noise, poor image quality and variable shape and size of tumors, the task of breast tumors segmentation is faced with great challenges. Many studies have paid off in the application of breast tumors segmentation and their approaches are divided into two categories: traditional methods [2,3] and deep learning methods [4–16].

Xiaosong Jiang contributed equally to this work.

✉ Jieyun Bai
bai_jieyun@126.com

Xiaosong Jiang
awen666@stu2020.jnu.edu.cn

¹ Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Information Technology, Jinan University, Guangzhou 510632, China

² College of Information Science and Technology, Jinan University, Guangzhou 510632, China

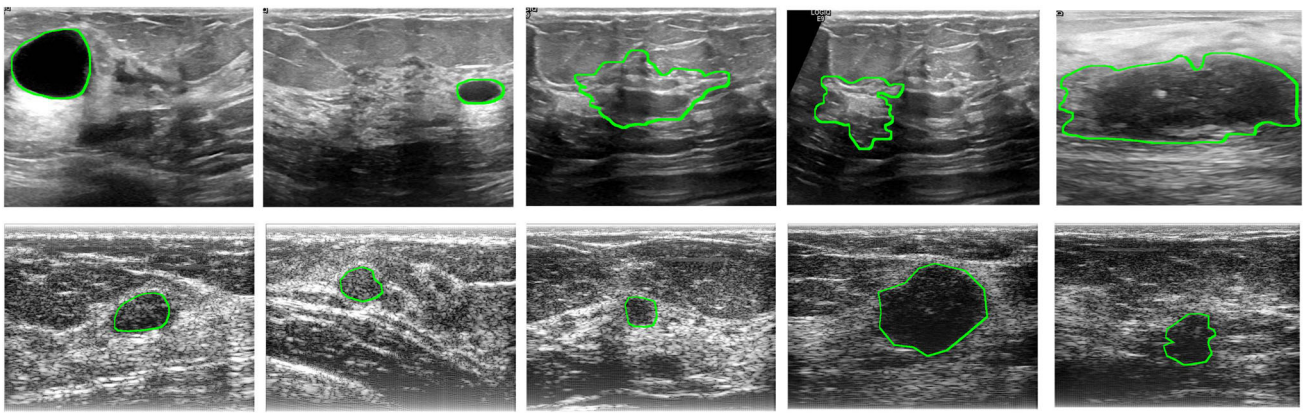


Fig. 1 In breast ultrasound, the size and shape of the tumor is irregular. At the same time, ultrasound images are characterized by low contrast and scattered noise. Therefore, it poses a high challenge for automatic

identification of breast tumor areas.(The first row of data comes from BUSI, while the second row of data comes from OASBUD)

Table 1 We review some studies using deep learning methods in breast tumors segmentation, summarize their experimental methods and segmentation performance metrics

Dataset	Model Name	Training method	Dice	Jaccard
BUSI	ESTAN [4]	5-fold cross validation	78	70
	SK-U-Net [5]	Train:test = 1:1	70.9	N
	GG-Net [8]	3-fold cross validation	82.1	73.8
	NAT-Net [10]	Train:valid:test = 6:2:2	N	77.5
	RCA-IUNet [11]	Train:valid = 7:3	91.4	89.9
	FPNN-TMEL [12]	Train:valid:test = 562:86:132	84.7	78.1
	BGM-Net [13]	Train:test = 661:119	83.97	75.97
	VEU-Net [14]	5-fold cross validation	89.62(B) 89.73(M)	78.43(B) 78.87(M)
	MSSA-Net [15]	10-fold cross validation	80.65	71.90
	SHA-MTL [16]	7-fold cross validation	81.42	N
	CTG-Net [6]		79	70
OASBUD	SK-U-Net [5]	Train:test = 1:1	67.6	N
	SEG-BM[7]	5-fold cross validation	73.37	63.13

Therefore, it can be used as an important reference for the study of this paper

The traditional methods are based on image texture features, which requires more professional skills. They are difficult to be widely applied and are less robust, although they are proved to be effective. Recently, the use of deep learning methods for breast tumors segmentation has worked effectively. However, convolutional layers are biased to extract local features and tend to lose global and location information as the network deepens, which will lead to a decrease in breast tumors segmentation accuracy. So, we propose a hybrid attention-guided network (HAG-Net).

Related work

In order to improve the capability of deep learning methods for segmentation, existing studies improved pixel-level

prediction through the following measures: (1) adopting attention mechanisms[5,6,8,9,15]; (2) transferring learning methods [8,12,15]; (3) fusing multi-scale contextual information [4,8,10–12,14–16]; (4) using boundary-guided methods[8,13]; (5)capturing long-range dependencies[8,12]; (6) multi-task learning approaches[6]. Two publicly available breast tumor datasets were used in this paper (BUSI [17] and OASBUD [18]), so we summarize the relevant studies in Table 1. It is important to note that there is a large variability in the experimental methods used by each researcher. Therefore, they cannot be fairly compared with each other. What is clear is that they all obtained better performance than the compared methods in their experimental results.

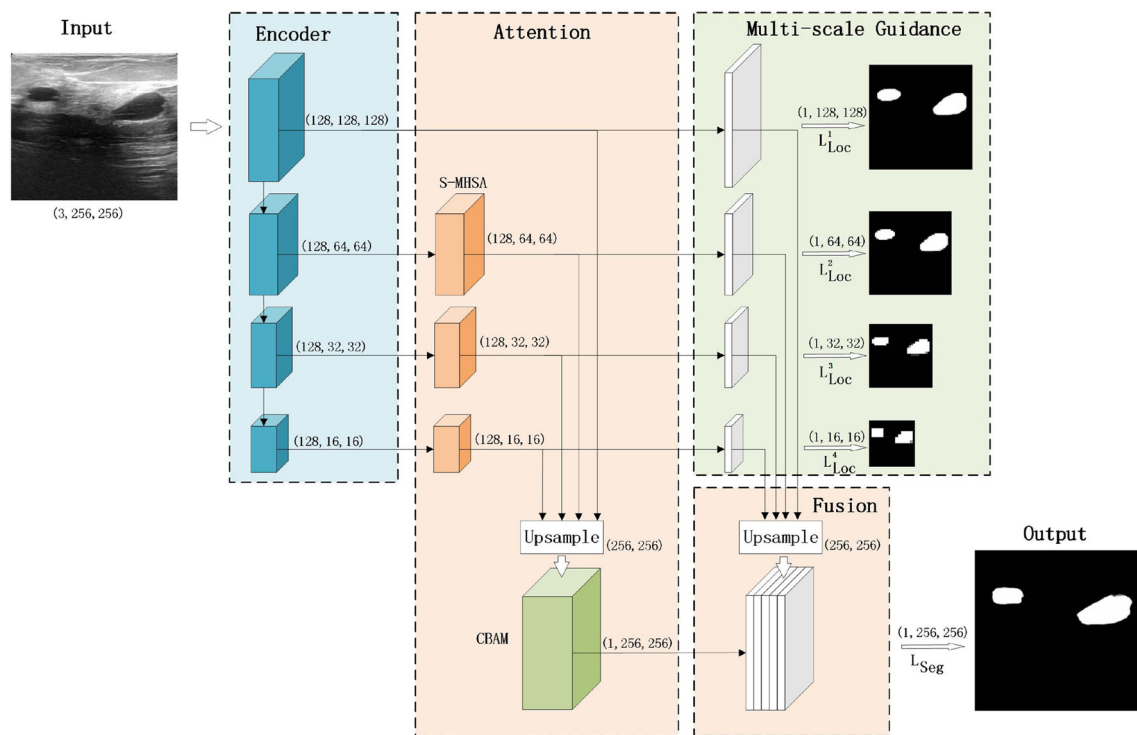


Fig. 2 The architecture of HAG-Net. It is broadly divided into four parts: (1) encoder: extract multi-scale feature maps; (2) attention: apply S-MHSA and CBAM to add attention mechanisms to capture long-range dependencies and global features; (3) multi-scale guidance: direct the network to pay more attention to tumors location information; (4)

fusion: integrate multi-scale feature maps. We use “(C,H,W)” to denote the dimension of output by each block, “C” denotes the number of channels, “H” denotes the height of feature maps, and “W” denotes the width of feature maps

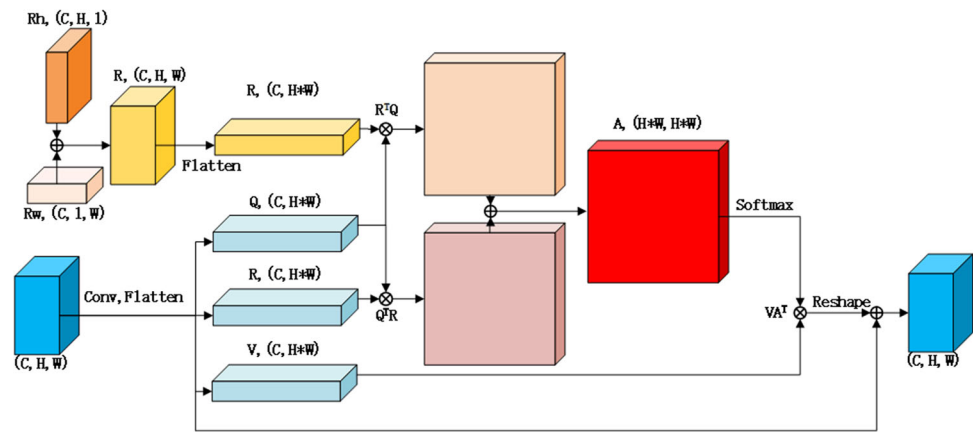
Methodology

To take full advantage of contextual information, spatial information of low-resolution features and capture long-range dependencies, we propose hybrid attentional guidance network (HAG-Net). The structure of network is shown in Fig. 2. Firstly, we use a pre-training network to extract feature maps, which can provide more prior knowledge to accelerate the convergence of the model. Then, short multi-head self-attention (S-MHSA) and convolutional block attention module (CBAM) are added to capture global feature and long-range dependencies, allowing the network to perform self-attention, spatial-attention and channel-attention learning. In addition, multi-scale guidance (MSG) method has been adopted, guiding the extraction of low-resolution features to pay more attention to tumors location information, further reducing the loss of spatial information due to the depths of the network. Finally, the features of different resolutions are up-sampled and fused using a regular convolution layer to obtain the segmentation result.

Short multi-head self-attention (S-MHSA)

Higher performance can be achieved if the model takes full advantage of long-range interactions. This was verified in a number of tumors segmentation studies [8,12]. Convolutional neural networks (CNN) can provide rich low-level features in shallow layers, but it is less suitable for modeling long-range dependencies, while long-range dependencies are better captured by the attention mechanism rather than by stacking convolutional layers [19]. Therefore, in our work, S-MHSA is developed to capture long-range dependencies. Multi-headed self-attention allows the model to focus on multiple subspaces of information at multiple locations simultaneously. Srinivas et al. [20] proposed a powerful backbone network, BoTNet, which employs relative distance encodings, allowing the network to notice not only content information but also the relative distance between features maps at different locations. S-MHSA is based on BoTNet by removing the coder and decoder layers, shorten the path between input and output, which can achieve higher

Fig. 3 The architecture of S-MHSA Block. A: attention; R: relative position encodings; Q: query; K: key; V: value; F: flatten; \oplus : Elemental Addition; \otimes : Matrix multiplication. Noted that the kernel size of Conv is 1×1



performance in our network by ablation experiments. S-MHSA is used in low-resolution feature extraction to perform supervised learning for low-resolution features, allowing the network to focus more on tumors location information rather than boundaries in low-resolution feature extraction, thus effectively compensating for the loss of spatial information among low-resolution features. Figure 3 illustrates our proposed S-MHSA. First, let Rh , Rw denote the relative position encodings of length and width. R can be obtained by summing Rh and Rw through broadcasting mechanism.

$$R^{(C,H,W)} = Rh^{(C,H,1)} + Rw^{(C,1,W)} \quad (1)$$

Three convolutional layers with a convolutional kernel size of 1×1 are used separately to extract features, which subsequently flatten according to the number of channels to obtain query, key and value, and denoted by $W_{Q(x)}, W_{K(x)}, W_{V(x)} (W \in R^{C,H*W})$ separately. And then, we calculate $W_{Q(x)}^T W_{K(x)} + R^T W_{Q(x)}$ and use the softmax activation function to obtain the attention matrix, which can be represented by $A (A \in R^{H*W,H*W})$.

$$A^{(H*W,H*W)} = \text{Softmax}(W_{Q(x)}^T W_{K(x)} + R^T W_{Q(x)}) \quad (2)$$

Once the attention matrix obtained, we will use matrix multiplication to compute $W_{V(x)} A^T$, and reshape it to (C, H, W) . Finally, the input and feature maps are summed by a skip connection to obtain the final output Y .

$$Y^{(C,H,W)} = x + W_{V(x)} A^T \quad (3)$$

Convolutional block attention module (CBAM)

Woo et al. [21] investigated how to improve the performance of the network from the attention perspective, and applied average pooling and maximum pooling in both channel attention and spatial attention to improve the representation capability of the network. In other words, let the

network automatically select the important channels and spatial information. The feature maps from different resolutions contain different levels of semantic information. In HAG-Net, we have up-sampled the features at different resolutions to the same size, which preserve the semantic information in the original resolution. They are then stacked together and entered into the CBAM module. The CBAM module automatically learns the importance of different channels and assigns a weight to them. The architecture of the CBAM block is shown in Fig. 4.

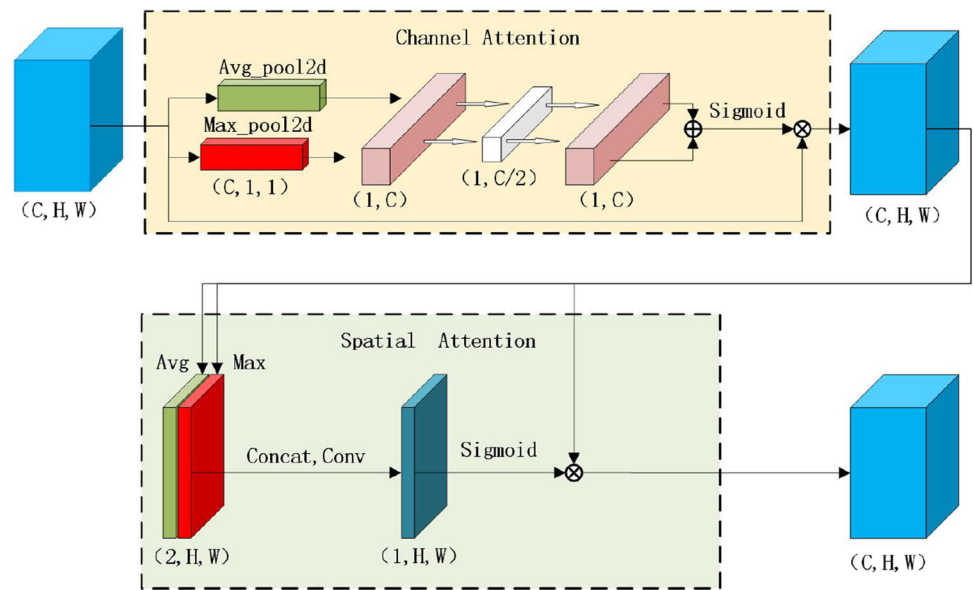
Multi-scale guidance (MSG)

High-resolution features contain more details compared to low-resolution features, but the latter have better target recognition capability. When the depth of the network deepens and the resolution decreases, the feature map will lose the spatial information of the target. So, we use a FPN [22] structure for capturing multi-scale location information, fusing multi-layers features for final prediction which contributing greatly to the detection of breast tumors. Considering the computational cost, S-MHSA is equipped in the low resolution feature extraction, directing the network paying more attention to the location information in low resolution features. Besides, Max-pooling is used for the ground truth as a way to obtain masks at different resolution levels. Figure 5 shows the results of our comparison experiments for MSG, and it can be clearly seen that there is less spatial information loss after applying MSG.

Loss function

We adopt MSG for guiding the network extract more multi-scale breast tumors location information. L_{Loc} is used to calculate the loss of breast tumors location information in each layer of feature maps. For the final segmentations, we use L_{Seg} to evaluate the similarity of the ground truth and seg-

Fig. 4 The architecture of CBAM. The two main components are channel attention and spatial attention



mentations. In other words, the loss function can be expressed as L_{final} :

$$L_{\text{final}} = \sum_{n=1}^4 L_{\text{Loc}}^i + L_{\text{Seg}} \quad (4)$$

To make it easier to understand, L_{Loc}^i can be found in Fig. 2, which is used to evaluate the distance of the feature from the low-resolution masks. i denotes resolution level. $P_{m,n}$ represents the coordinates of the prediction images, while $G_{m,n}$ represents ground truth. Specifically, L_{Loc} can be expressed as Eq. 5.

$$L_{\text{Loc}}^i = \frac{\sum_{m=1}^M \sum_{n=1}^N (P_{m,n} - G_{m,n})^2}{MN} \quad (5)$$

For the final segmentation results, we use the dice coefficient loss function, which is widely used in segmentation studies. That is to say, the expression for L_{Seg} is:

$$L_{\text{Seg}} = 1 - \frac{\sum_{m=1}^M \sum_{n=1}^N (P_{m,n} \times G_{m,n})}{\sum_{m=1}^M \sum_{n=1}^N P_{m,n} + \sum_{m=1}^M \sum_{n=1}^N G_{m,n}} \quad (6)$$

Experimental part

Data set

We use two publicly available datasets to evaluate the performance of the method: BUSI [17] and OASBUD [18]. BUSI was obtained from Baheya Hospital and acquired by using LOGIQ E9 and LOGIQ E9 Agile ultrasound system. The data consisted of 780 images from 600 women aging from

25 to 75, of which 437 are benign, 210 are malignant and 133 are normal cases, respectively. To evaluate the performance of the method for tumors segmentation, we excluded normal cases and kept the ultrasound images containing tumors. Please note that some images in the dataset have multiple labels, which we merged into a new label. OASBUD was recorded in the Department of Ultrasound, Institute of Fundamental Technological Research Polish Academy of Sciences. The data were collected from 78 women between the ages of 24 and 75. A total of 100 cases are included in the data, of which 52 are malignant and 48 are benign. Each case is scanned twice, meaning there are a total of 200 ultrasound images.

Experimental setup

We review the previous studies and select true positive ratio (TPR), false positive ratio (FPR), Jaccard Index (JI), dice coefficient (DICE), area error ratio (AER), Hausdorff distance (HD), and detection rate (DR) as the evaluation metrics, the details of which are listed in Table 2.

Our experiments are conducted by using the SGD optimizer, with the batch size setting to 4, epochs to 100, and the initial learning rate to 0.001. To obtain the global optimal solution, cosine descent is used after 50 epochs. The tumors images are resized to 256×256 . During the training process, training will be stopped if the validation dice does not change for more than 30 epochs to prevent over-fitting. Moreover, we perform random data enhancement on the data, including rotation, horizontal inversion, and vertical flip ($p = 0.5$). Due to the large variation in the shape of tumors, especially malignant tumors. This will cause the data distribution in the training and validation sets to be not exactly the same, which

Table 2 The definition of metrics

Metrics	Definition
TPR	$\frac{TP}{TP + FN}$
FPR	$\frac{FP}{TP + FN}$
JI	$\frac{2TP}{TP + FN + FP}$
DICE	$\frac{2TP + FN + FP}{2TP + FN + FP}$
AER	$\frac{FN + FP}{TP + FN}$
HD	$\max\{\max_{x \in P} \{\min_{y \in G} (x - y)\}, \max_{x \in G} \{\min_{y \in P} (x - y)\}\}$
DR	$\begin{cases} 1 & dice \geq 0.5 \\ 0 & dice < 0.5 \end{cases}$

TP True Positive, TN True Negative, FP False Positive, FN False Negative, P Prediction, G Ground truth

will lead to fluctuating results. So we used five random five-fold cross validations to conduct ablation and comparison experiments. Specifically, we set a random order, and then divide it into fivefold. Each of them is used as the validation set in turn, while the others are used as the training set. We repeat this process five times, meaning that there are five random orders. The experiment was run separately on two publicly available datasets. One more thing to add is that all the experiments above are based on a NVIDIA Geforce RTX 3090 graphics card.

Ablation experiment

Ablation experiments have been conducted to study the influence of single component on the performance of breast tumor segmentation. Specifically, six sets of ablation experiments are conducted to compare the performance of segmentation after the combination of MSG, S-MHSA, BoTNet and CBAM. In addition, in order to compare the generalization performance of the methods in this paper, we conducted ablation experiments on two publicly available datasets respectively.

Table 3 Results of ablation experiments on BUSI

Model Name	TPR	FPR	JI	DIC	AER	DR	HD
Basic	83.9 ± 1.4	35.4 ± 9.9	73.4 ± 2.1	81.9 ± 1.9	51.5 ± 9.7	91.5 ± 2.6	48.0 ± 1.7
Basic-S	84.0 ± 1.7	35.1 ± 10.6	73.1 ± 2.2	81.8 ± 2.0	51.1 ± 10.6	91.6 ± 2.8	48.3 ± 2.2
Basic-C	84.0 ± 2.0	32.8 ± 9.2	73.6 ± 2.1	82.1 ± 2.0	48.8 ± 9.3	91.9 ± 2.8	47.9 ± 2.0
Basic-CS	84.5 ± 1.6	33.8 ± 8.7	73.8 ± 1.9	82.3 ± 1.7	49.3 ± 9.0	91.7 ± 2.3	47.9 ± 1.8
Basic-CB	84.0 ± 1.7	34.3 ± 9.4	73.5 ± 2.1	82.0 ± 2.0	50.3 ± 9.8	91.4 ± 2.8	48.1 ± 1.8
Basic-CSM	84.7 ± 1.5	35.1 ± 12.6	74.2 ± 2.1	82.6 ± 1.9	50.3 ± 12.7	92.1 ± 2.2	47.4 ± 1.9

“M”: MSG; “C”: CBAM; “S”: S-MHSA; “B”: BoTNet

The results of the experiment are expressed as “mean ± standard deviation”. The best results have been bolded for easier viewing

Table 3 shows the results of ablation experiments. We set a basic structure as a comparison, which is the same topology as the HAG-Net, but the individual components are removed. First, we compare “Basic”, “Basic + S-MHSA”, “Basic + CBAM”, “Basic + CBAM + S-MHSA”, “Basic + CBAM + BoTNet” and “Basic + CBAM + S-MHSA + MSG” in turns. It is found that although MSG does not add additional parameters, it can effectively improve segmentation results and obtain higher DICE, JI, and DR. It is verified that MSG guides the network focus on location information in low-resolution features and reduces the loss of spatial information. Secondly, comparing “Basic” and “Basic + CBAM + S-MHSA”, we find that by introducing self-attention, channel-attention and spatial-attention, global features and long-range dependencies can be effectively captured, improve the network’s predictive ability at pixel level. Moreover, we also try “Basic + CBAM + BoTNet”, and the experimental results can show that “Basic + CBAM + S-MHSA” is more suitable for the current application scenario.

Figure 5 shows the feature maps extracted at different levels. To make the comparison more obvious, we compare “Basic” and “Basic + CBAM + S-MHSA + MSG” on two publicly available datasets. The feature maps are the output map via S-MHSA. In the high resolution feature maps, the receptive field is small. So, more shallow features are extracted, such as colour, texture, edge and corner information. For low resolution feature maps, the receptive field is larger compare to the former, contains more advanced semantic information. Therefore, “Basic + CBAM + S-MHSA + MSG” received the best assessment compared to “Basic”.

Comparative experiment

In this subsection, we reproduce other state-of-the-art methods, including U-Net [23], FCN [24], Attention U-Net [25], Nested U-Net [26], Deeplabv3 + [27], SK-U-Net [5] and FPN-TMEL [12]. We obtained the code for these networks and then ran them all on the same GPU. For FPN-TMEL, we followed their settings for data preprocessing, hyperparameterizers, optimizers, etc., to get the best performance.

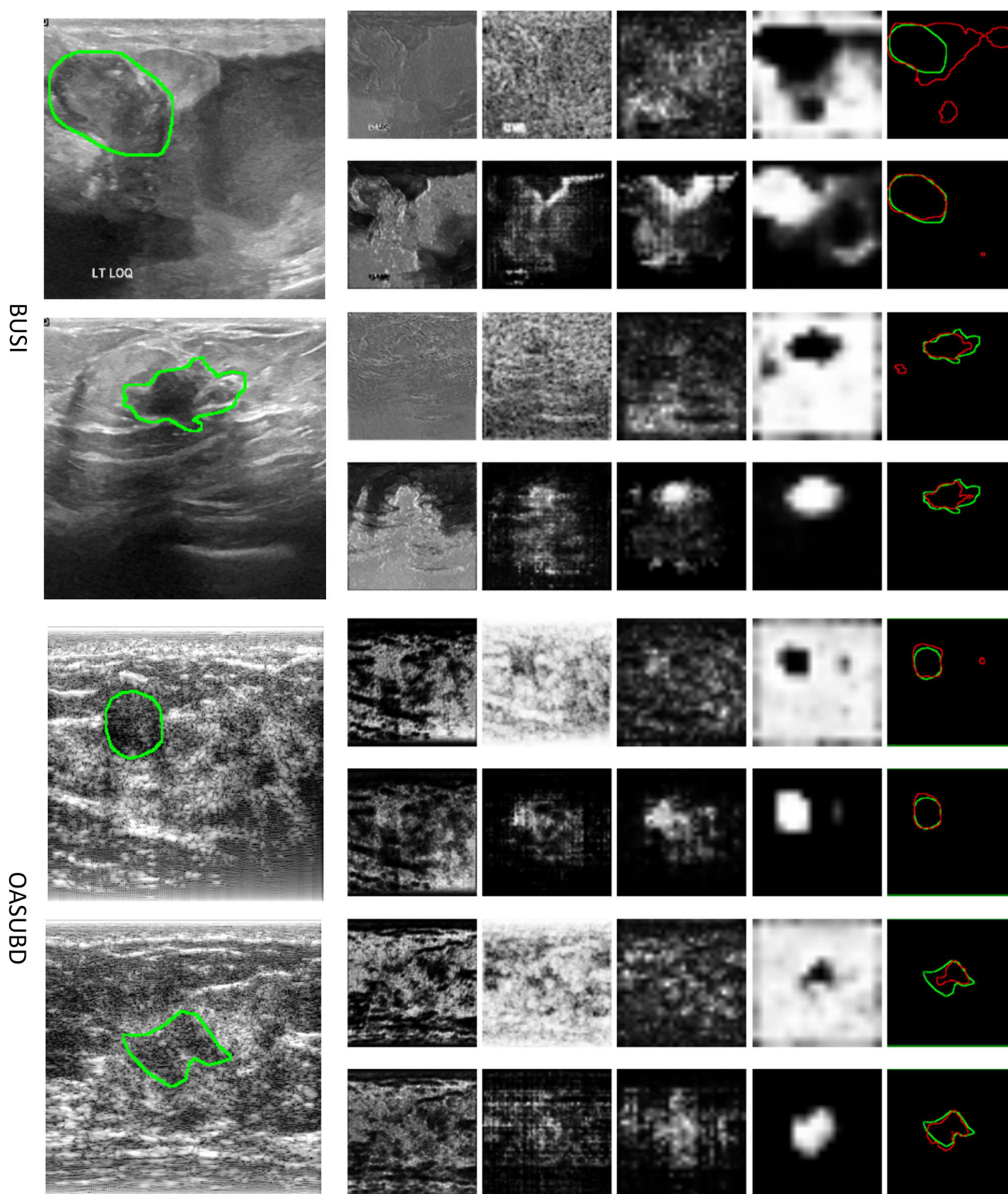


Fig. 5 The figure shows the feature maps extracted for “Basic” and “Basic + CBAM + S-MHSA + MSG” at different resolutions in BUSI and OASUBD. First, we show the original image, where the green

markers are the original masks. The size of the feature maps contains 128×128 , 64×64 , 32×32 and 16×16 . In the final image, the red markers are the segmentation boundaries we obtained from HAG-Net

Table 4 Results of ablation experiments on OASUBD

Model Name	TPR	FPR	JI	DIC	AER	DR	HD
Basic	76.5 ± 3.9	35.1 ± 23.2	65.2 ± 4.1	76.2 ± 3.8	58.6 ± 24.4	89.2 ± 4.0	45.0 ± 3.6
Basic-S	79.7 ± 4.3	41.9 ± 15.1	66.0 ± 4.4	77.0 ± 4.0	62.2 ± 17.4	90.5 ± 4.2	45.3 ± 4.0
Basic-C	78.6 ± 4.0	41.9 ± 18.1	64.8 ± 3.8	76.0 ± 3.6	63.3 ± 19.7	89.8 ± 3.9	46.2 ± 3.0
Basic-CS	79.9 ± 3.8	42.9 ± 22.8	66.7 ± 3.3	77.5 ± 3.2	63.0 ± 22.9	91.2 ± 4.5	45.3 ± 3.5
Basic-CB	77.3 ± 3.9	38.1 ± 18.9	64.6 ± 4.0	75.6 ± 3.8	60.9 ± 20.0	88.5 ± 5.4	46.2 ± 3.4
Basic-CSM	79.6 ± 3.3	38.5 ± 19.3	66.8 ± 3.2	77.8 ± 3.1	58.9 ± 20.1	91.9 ± 5.0	45.2 ± 2.6

“M”: MSG; “C”: CBAM; “S”: S-MHSA; “B”: BoTNet

The best results are marked with the significance of bold.

Table 5 Quantitative experimental comparison of HAG-Net with the state-of-the-art methods on BUSI

Model Name	TPR	FPR	JI	DICE	AER	DR	HD
U-Net	74.6 ± 2.5	38.3 ± 15.1	63.6 ± 2.2	72.7 ± 2.2	63.7 ± 14.8	82.2 ± 3.1	55.1 ± 2.0
FCN	84.7 ± 1.7	39.0 ± 14.1	74.0 ± 2.0	82.2 ± 1.9	54.3 ± 14.3	91.6 ± 2.2	47.4 ± 2.0
Deeplabv3+	86.9 ± 1.6	40.2 ± 14.0	73.3 ± 2.5	81.9 ± 2.3	53.3 ± 13.9	91.3 ± 2.9	47.6 ± 2.1
Attention U-Net	77.5 ± 2.8	40.2 ± 18.4	66.6 ± 2.4	75.4 ± 2.4	62.7 ± 17.5	85.1 ± 2.7	53.0 ± 1.9
U-Net++	76.6 ± 2.9	37.1 ± 13.7	65.8 ± 2.3	74.8 ± 2.3	60.5 ± 12.9	84.8 ± 3.3	53.6 ± 2.3
SK-U-Net	80.2 ± 2.6	41.6 ± 16.8	68.1 ± 2.3	77.5 ± 2.2	61.4 ± 16.7	87.9 ± 2.7	52.4 ± 2.1
FPNN-TMEL	82.6 ± 2.0	28.7 ± 9.5	73.6 ± 2.2	81.7 ± 2.1	46.1 ± 9.7	90.5 ± 2.8	48.7 ± 1.7
HAG-Net (our)	84.7 ± 1.5	35.1 ± 12.6	74.2 ± 2.1	82.6 ± 1.9	50.3 ± 12.7	92.1 ± 2.2	47.4 ± 1.9

The best results are marked with the significance of bold.

For the other networks, we used a unified scheme, as indicated in the experimental setup. They are both proven to have excellent image segmentation capabilities. The quantitative analysis results of each methods are shown in Tables 5 and 6. First, it can be seen that our method obtain the highest DICE, JI and DR on BUSI ($82.6 \pm 1.9\%$, $74.2 \pm 2.1\%$, $92.1 \pm 2.2\%$) and OASUBD ($77.8 \pm 3.1\%$, $66.8 \pm 3.2\%$, $91.9 \pm 5.0\%$). These metrics are among the most concerned metrics with the segmentation task. Besides, other metrics of HAG-Net also have significant advantages. The high performance in both public datasets illustrated the excellent segmentation and generalization ability.

Each network generally performed better on BUSI than on OASUBD. We analyze that this may be due to the lower contrast of the OASUBD data compared to BUSI, which places a higher demand on the segmentation method. Even so, HAG-Net still maintain excellent performance, further suggesting the effectiveness of HAG-Net.

Discussion

Accurate segmentation of breast tumors is an important component of CAD. For this purpose, many studies have been done and their ideas have been verified. In this work, we review the past methods and propose HAG-Net, and then perform ablation and comparison experiments. As is shown in the ablation experiments (Tables 3, 4), HAG-Net exhibit

excellent segmentation performance by adding individual components. Breast tumors are variable in size, making it difficult to accurately localize tumors areas at a single resolution. This problem can be solved by expanding the receptive field through a multi-scale fusion. But this also has some problems, as the depth of the network deepens, the boundary and location information of the tumors are lost. In low-resolution feature maps, boundary features are more difficult to capture and less useful for segmentation compare to location information due to the blurring boundary of breast tumors. Therefore, we adopt the MSG method to remedy this deficiency by guiding the extraction of multi-scale features to focus more attention on the tumors location information, which explains why our method can obtain the highest DR. The convolutional layer extracts features based on the receptive field of the convolutional kernel, resulting in a limited area for scanning. So, it is more biased to extract local information and is not good for capturing long-range dependencies. Therefore, we introduce multiple attention mechanisms, S-MHSA and CBAM. S-MHSA is adopted in the low-resolution layer and CBAM is adopted in the fusion layer as a way to capture global features and long-range dependencies.

The results compared to other 7 state-of-the-art methods (Tables 5, 6) shows that our method obtained the highest DICE, JI and DR on BUSI and OASUBD, which are the most common types of segmentation performance evaluation

Table 6 Quantitative experimental comparison of HAG-Net with the state-of-the-art methods on OASBUD

Model Name	TPR	FPR	JI	DICE	AER	DR	HD
U-Net	74.8 ± 5.6	52.4 ± 18.9	57.6 ± 3.1	69.9 ± 3.2	77.6 ± 17.3	84.4 ± 5.3	51.2 ± 2.9
FCN	78.2 ± 3.8	39.5 ± 20.0	65.9 ± 3.8	76.6 ± 3.7	61.3 ± 21.8	90.0 ± 5.0	44.7 ± 3.0
Deeplabv3+	82.5 ± 3.6	53.2 ± 20.9	64.9 ± 3.6	75.8 ± 3.4	70.7 ± 21.9	88.3 ± 4.5	45.6 ± 2.7
Attention U-Net	77.0 ± 3.1	47.0 ± 18.9	61.9 ± 3.4	73.3 ± 3.1	70.0 ± 19.7	87.2 ± 3.6	48.7 ± 3.0
U-Net++	75.5 ± 4.2	50.3 ± 23.7	60.3 ± 3.8	71.9 ± 3.4	74.8 ± 23.6	84.7 ± 4.9	49.7 ± 3.2
SK-U-Net	75.8 ± 4.1	50.9 ± 19.7	59.0 ± 3.3	71.0 ± 3.4	75.1 ± 19.7	84.0 ± 5.4	50.2 ± 3.3
FPNN-TMEL	79.8 ± 4.1	54.0 ± 23.4	63.2 ± 4.1	74.2 ± 3.7	74.2 ± 24.0	86.2 ± 5.3	49.0 ± 37.1
HAG-Net(our)	79.6 ± 3.3	38.5 ± 19.3	66.8 ± 3.2	77.8 ± 3.1	58.9 ± 20.1	91.9 ± 5.0	45.2 ± 2.6

The best results are marked with the significance of bold.

metrics. Deeplabv3 obtained the highest TPR, however, the DICE was lower than that of HAG-Net. We infer that this is due to over-segmentation. In other words, Deeplabv3 has a larger prediction range for tumors regions, which allows the real tumors to be identified more accurately, but also misidentifies non-tumors regions.

In addition, Deeplabv3 and FCN we used embedded Resnet-101 as the backbone network. FPNN-TMEL has embedded SEResnext-50 as backbone network. Using the pre-trained model as the backbone network not only speeds up the convergence of the network model, but also improves the accuracy of the model segmentation. Besides, SK-U-Net and Attention U-Net, which are based on the attention mechanism, achieve significantly higher performance compared to U-Net, demonstrating the effectiveness of attentional mechanisms for the segmentation of breast tumors.

Also, we show the results of successful and failed segmentation of the compared networks separately (Fig. 6). In successful cases, various methods are more effective in the segmentation of benign tumors. However, they are significantly worse in malignant tumors. The borders and location of malignant tumors are more complex than those of benign tumors, and their borders have significant irregularities. Nevertheless, HAG-Net still have a higher similarity to the real area. For successful cases, the contrast of the tumors margins is higher, while the failure cases are lower. Therefore, all the methods in the failed cases are not very good for boundary identification. The higher sensitivity of our method for tumors identification can be seen not only from the image segmentations, but also from the highest detection rate in two datasets. To summarize, experimental results above can

show that our method will be useful for computer-aided diagnosis of breast cancer and provide reference value to other fields of segmentation methods.

Conclusion

In this article, we synthesize previous approaches for breast tumors segmentation and propose HAG-Net. HAG-Net compensate for the loss of spatial information due to network depth by using the MSG method. Besides, S-MHSA and CBAM are adopted to introduce attention mechanisms for better capturing global features and long-range dependencies. Ablation and comparative experiments have been conducted to verify the effectiveness and superiority of the method. The results on two publicly available datasets show that our method obtained the highest DICE, JI and DR, superior to other 7 state-of-the-art methods we compare with. Meaning that our method is more useful to assist computerized diagnosis of breast cancer.

The research in this work focuses on improving the segmentation performance of benign and malignant tumors. However, the reality also includes ultrasound images without tumors. These images contain some tumor-like shadows that can mislead the network to learn the wrong features. Therefore, we will explore how to improve the network's ability to recognize those ultrasound images that do not contain tumors.

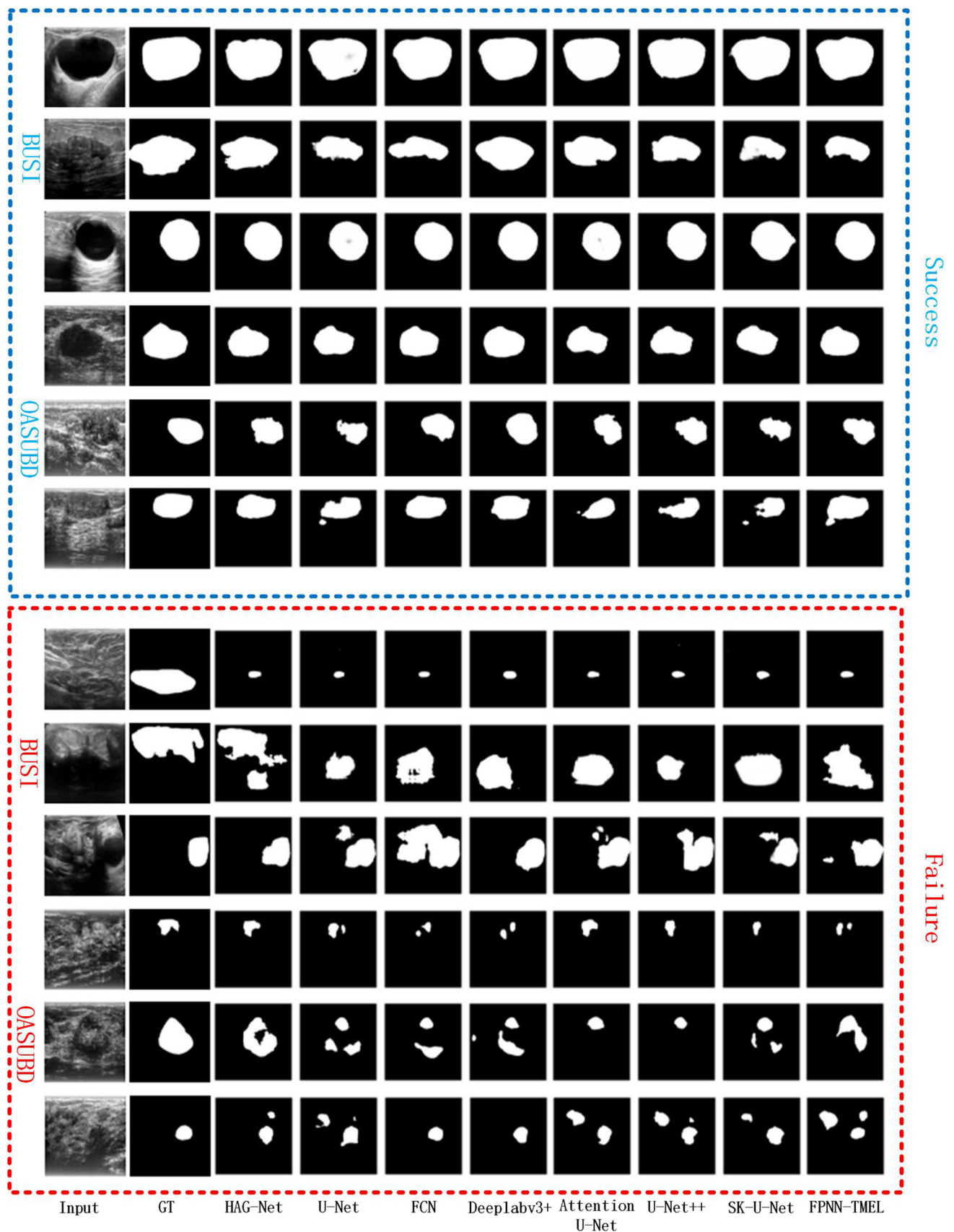


Fig. 6 Comparing the results of each method for breast tumors segmentation. We show the successful and failed cases from BUSI and OASUBD, respectively. In each section, the first three lines are from BUSI, while the last three lines are from OASUBD

Author Contributions YL and XJ contributed to the design of the entire work. MZ: Investigation, Conceptualization, Writing, Review & editing. DZ and RQ: Writing, Funding acquisition, Writing, Review & editing. ZO and JB: Writing, Review & editing, investigation.

Funding This research was funded by the Guangdong Basic and Applied Basic Research Foundation (2023A1515012833), the Science and Technology Program of Guangzhou (202201010544), Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Informatization (2021B1212040007) and National Key Research and Development Project (2019YFC0120100, and 2019YFC0121907).

Declarations

Conflict of interest The authors do not have any conflicts of interest.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
- Xian M, Zhang Y, Cheng HD (2015) Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. *Pattern Recogn* 48(2):485–497. <https://doi.org/10.1016/j.patcog.2014.07.026>
- Moon WK, Lo C-M, Chen R-T, Shen Y-W, Chang JM, Huang C-S, Chen J-H, Hsu W-W, Chang R-F (2014) Tumor detection in automated breast ultrasound images using quantitative tissue clustering. *Med Phys* 41(4):042901. <https://doi.org/10.1118/1.4869264>
- Shareef B, Vakanski A, Xian M, Freer PE (2020) Estan: enhanced small tumor-aware network for breast ultrasound image segmentation. *Healthcare*. <https://doi.org/10.48550/arXiv.2009.12894>
- Byra M, Jarosik P, Szubert A, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, Andre M (2020) Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. *Biomed Signal Process Control* 61:102027. <https://doi.org/10.1016/j.bspc.2020.102027>
- Yang K, Suzuki A, Ye J, Nosato H, Izumori A, Sakanashi H (2022) Ctg-net: cross-task guided network for breast ultrasound diagnosis. *PLoS ONE* 17(8):1–25. <https://doi.org/10.1371/journal.pone.0271106>
- Podda AS, Balia R, Barra S, Carta S, Fenu G, Piano L (2022) Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. *J Comput Sci* 63:101816. <https://doi.org/10.1016/j.jocs.2022.101816>
- Xue C, Zhu L, Fu H, Hu X, Li X, Zhang H, Heng P-A (2021) Global guidance network for breast lesion segmentation in ultrasound images. *Med Image Anal* 70:101989. <https://doi.org/10.1016/j.media.2021.101989>
- Yang K, Suzuki A, Ye J, Nosato H, Izumori A, Sakanashi H (2021) Tumor detection from breast ultrasound images using mammary gland attentive U-Net. In: *International forum on medical imaging in Asia 2021*, vol. 11792, p 1179202. <https://doi.org/10.1117/12.2590073>
- Zou H, Gong X, Luo J, Li T (2021) A robust breast ultrasound segmentation method under noisy annotations. *Comput Methods Prog Biomed* 209:106327. <https://doi.org/10.1016/j.cmpb.2021.106327>
- Punn NS, Agarwal S (2022) Rca-iunet: a residual cross-spatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging. *Mach Vis Appl* 33(2):1–10. <https://doi.org/10.1007/s00138-022-01280-3>
- Tang P, Yang X, Nan Y, Xiang S, Liang Q (2021) Feature pyramid nonlocal network with transform modal ensemble learning for breast tumor segmentation in ultrasound images. *IEEE Trans Ultrason Ferroelectr Freq Control* 68(12):3549–3559. <https://doi.org/10.1109/TUFFC.2021.3098308>
- Wu Y, Zhang R, Zhu L, Wang W, Wang S, Xie H, Cheng G, Wang FL, He X, Zhang H (2021) Bgm-net: Boundary-guided multiscale network for breast lesion segmentation in ultrasound. *Front Mol Biosci*. <https://doi.org/10.3389/fmolb.2021.698334>
- Ilesanmi AE, Chaumrattanakul U, Makhanov SS (2021) A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. *Biocybern Biomed Eng* 41(2):802–818. <https://doi.org/10.1016/j.bbe.2021.05.007>
- Xu M, Huang K, Chen Q, Qi X (2021) Mssa-net: Multi-scale self-attention network for breast ultrasound image segmentation. In: *2021 IEEE 18th international symposium on biomedical imaging (ISBI)*, pp 827–831. <https://doi.org/10.1109/ISBI48211.2021.9433899>
- Zhang G, Zhao K, Hong Y, Qiu X, Zhang K, Wei B (2021) Sha-mtl: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification. *Int J Comput Assist Radiol Surg* 16(10):1719–1725. <https://doi.org/10.1007/s11548-021-02445-7>
- Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020) Dataset of breast ultrasound images. *Data Brief* 28:104863. <https://doi.org/10.1016/j.dib.2019.104863>
- Piotrkowska-Wróblewska H, Dobruch-Sobczak K, Byra M, Nowicki A (2017) Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions. *Med Phys* 44(11):6105–6109. <https://doi.org/10.1002/mp.12538>
- Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: *Computer vision—ECCV 2014*, Cham, pp 818–833. https://doi.org/10.1007/978-3-319-10590-1_53
- Srinivas A, Lin T-Y, Parmar N, Shlens J, Abbeel P, Vaswani A (2021) Bottleneck transformers for visual recognition. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp 16519–16529. <https://doi.org/10.48550/arXiv.2101.11605>
- Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*. <https://doi.org/10.48550/arXiv.1807.06521>
- Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. <https://doi.org/10.48550/arXiv.1612.03144>
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI*, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. <https://doi.org/10.48550/arXiv.1605.06211>

25. Oktay O, Schlemper J, Folgoc LL, Lee MCH, Heinrich MP, Misawa K, Mori K, McDonagh SG, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention u-net: learning where to look for the pancreas. CoRR [arXiv:1804.03999](https://arxiv.org/abs/1804.03999)
26. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
27. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). <https://doi.org/10.48550/arXiv.1802.02611>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.