



AMS-PAN: Breast ultrasound image segmentation model combining attention mechanism and multi-scale features

Yuchao Lyu, Yinghao Xu, Xi Jiang, Jianing Liu, Xiaoyan Zhao^{*}, Xijun Zhu

College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China



ARTICLE INFO

Keywords:

Ultrasound image
Multi-scale features
Attention mechanisms
Breast tumor segmentation
Feature pyramid

ABSTRACT

Breast ultrasound medical images are characterized by poor imaging quality and irregular target edges. During the diagnosis process, it is difficult for physicians to segment tumors manually, and the segmentation accuracy required for diagnosis is high, so there is an urgent need for an automated method to improve the segmentation accuracy as a technical tool to assist diagnosis. This study designed an improved Pyramid Attention Network combining Attention mechanism and Multi-Scale features (AMS-PAN) for breast ultrasound image segmentation. On the encoding side, the model adopts the depthwise separable convolution strategy to achieve a multi-scale receptive field with cumulative small-size convolution, which performs multi-dimensional feature extraction and forms a feature pyramid. The model uses Global Attention Upsample (GAU) feature fusion on the decoding side. In order to further process the fused feature information, the proposed method uses a Spatial and Channel Attention (SCA) module to shift the model's segmentation focus to the edge texture information. The good segmentation performance of our method is verified through experiments on BUSI and OASBUD. All the designed parts have contributed to the segmentation performance in practical applications. Compared with the traditional non-deep learning methods and the current mainstream deep learning methods, the improvement of the model in Dice and IoU metrics is pronounced. AMS-PAN has high computational efficiency, and its good performance has been proven to play a role in ultrasound detection tasks of breast tumors for physicians to specific auxiliary diagnostic roles to guide the subsequent diagnosis and treatment services for patients.

1. Introduction

According to the latest global cancer data report [1], breast cancer is women's most commonly diagnosed cancer. Due to its high mortality rate is also the deadliest cancer in the female population, second only to lung cancer. However, breast cancer can be triggered by various factors, such as obesity and alcohol consumption, and the incidence of breast cancer increases with age. Although the tumor is not incurable, the associated surgery can have an irreversible impact on the patient's psyche and life. It is difficult to treat at advanced stages of the tumor,

and its early detection and accurate diagnosis remain the most critical part of the diagnostic process. In addition to self-examination, Digital Mammography (DM) and Magnetic Resonance Imaging (MRI) are central to detecting tumor lesions, but physicians and patients are often wary of the radioactivity of these tests. After the validity and feasibility of breast ultrasound screening and cancer diagnosis were proven [2], ultrasound imaging has rapidly become widely accepted as a means to detect breast tumors due to its non-radiation, non-invasive, real-time, and inexpensive advantages. In order to improve the objectivity and accuracy of diagnosis, computer-aided diagnostic systems for breast

Abbreviations: AMS-PAN, Pyramid Attention Network model combining Attention mechanism and Multi-Scale features; GAU, Global Attention Upsample; SCA, Spatial and Channel Attention; BUSI, Breast Ultrasound Image dataset for segmentation; OASBUD, Open Access Series of Breast Ultrasonic Dataset; DM, Digital Mammography; MRI, Magnetic Resonance Imaging; SoTA, State-of-The-Art; PAN, Pyramid Attention Network; FPA, Feature Pyramid Attention; MW, Marker Watershed; MS, Morphological Snake; AMSMW, Adaptive Morphological Snake based on Marker Watershed; SSD, Single Shot multi-box Detector; FPN, Feature Pyramid Network; CBAM, Convolutional Block Attention Module; ReLU, Rectified Linear Units; IoU, Intersection over Union; TP, True Positives; TN, True Negatives; FP, False Positives; FN, False Negatives; ROC, Receiver Operating Characteristic; PR, Precision-Recall; AUC, Area Under Curve; mAP, mean Average Precision; Att U-Net, U-Net with Spatial and Channel'Squeeze Excitation' Attention.

* Corresponding author.

E-mail addresses: 4020110029@mails.qust.edu.cn (Y. Lyu), xyh0609@163.com (Y. Xu), 609793553@qq.com (X. Jiang), 872174423@qq.com (J. Liu), zhaoxiaoyan@qust.edu.cn (X. Zhao), zhuxj990@163.com (X. Zhu).

ultrasound images have gradually become a hot topic of research. Among them, the automatic image segmentation work will directly affect the accuracy and reliability of diagnostic results. Therefore, it is essential to study the segmentation technology of breast ultrasound images. Fig. 1 shows the ultrasound imaging results and marks the segmented area contours for benign and severe breast tumors, respectively. It can be seen that ultrasound images have the inherent problem of poor imaging quality and low image definition. This problem makes manual segmentation of the lesion area challenging, even for professional pathologists.

Most of the traditional segmentation methods only deal with images with sizeable target-background distinction, which have the problems of high data requirements and poor robustness, making it challenging to meet the segmentation needs of breast ultrasound images. Since then, researchers have started to improve such methods. Wang et al. [3] proposed a random forest-based method that significantly enhanced its segmentation accuracy while maintaining interpretability. On the other hand, Inan et al. [4], on the other hand, presented a hybrid machine learning-based model that could integrate the prediction results of different methods by XGboost. In addition, Ak et al. [5] tried to exploit the current State-of-The-Art (SoTA) machine learning methods to obtain better performance. However, these methods inevitably have poor generalization ability and are difficult to be applied under different datasets. Therefore, the performance on the segmentation task of breast ultrasound images remains poor.

However, with the development of deep learning methods, their application in medical image processing has become the focus of current research. Especially after the emergence of U-Net [6], with its excellent performance, the application prospect of codec models on medical image segmentation has been significantly broadened. It has sufficient good performance in tasks such as retinal vessel segmentation [7], liver tumor segmentation [8], and brain tumor segmentation [9]. Its improvement has also become an instant research focus, and models such as Att U-Net [10], U-Net++ [11], and Trans-UNet [12] have emerged in recent years. Because of the characteristics of breast ultrasound images, this study proposes a Pyramid Attention Network model combining Attention mechanism and Multi-Scale features (AMS-PAN) as a codec model for breast ultrasound image segmentation tasks. Unlike previous deep learning methods, we design multi-scale receptive fields in the coding part to extract multi-scale feature maps using different receptive fields and use them to form a feature pyramid. In the decoding part, we improve the Pyramid Attention Network (PAN) [13] structure, following the Global Attention Upsample (GAU) part of it, and design the Spatial and Channel Attention (SCA) module based on it. Compared with the original PAN model, the SCA module can further improve the effectiveness of utilizing the model for different scale features and enhance its segmentation effectiveness. The main contributions of this study can be summarized as the following three points:

This study uses a multi-scale feature extraction module on the encoding side. The module is implemented using the idea of depth separable convolution [14], which achieves multi-scale perceptual fields by stacking multiple layers of small-sized convolutional kernels to extract high-dimensional feature structures further while

maintaining low-dimensional information. In the segmentation process, the low-dimensional features can provide more global information about the basic shape of the target. In comparison, the high-dimensional features can provide more local information about its edges and texture.

This study designs a PAN-based decoder at the decoding end for fusing feature maps of different dimensions. The original PAN structure consists of two parts, GAU and Feature Pyramid Attention (FPA). The GAU part, which benefits from its more vital scalability, is continued in the design, while the proposed SCA module replaces the FPA for deeper feature processing. The SCA module is an attention mechanism that combines spatial information with channel information. This structure enables the network to pay more attention to the spatial and channel dimension information. Furthermore, further, integrate them to highlight the practical part of them.

Ablation experiments and comparison experiments were conducted on the Breast Ultrasound Image dataset for segmentation (BUSI) [15] and the Open Access Series of Breast Ultrasonic Dataset (OASBUD) [16]. The validity of AMS-PAN and the degree of contribution of each part of the module were validated. Furthermore, based on this, its contribution to the development of the segmentation tasks of the breast ultrasound image and the scalability of related ideas are thoroughly discussed.

2. Related work

Currently, the most widely used auxiliary segmentation methods are still improved based on traditional theories, such as the level set method proposed earlier by Sussman et al. [17]. They have undergone several innovations for different images since then. Another class of commonly used methods is the marker-washed-based method [18]. For the characteristics of breast ultrasound images, Shen et al. combined the advantages of Morphological Snake (MS) [19] and designed an Adaptive Morphological Snake based on Marker Watershed (AMSMW) to enhance its segmentation capability [20]. Furthermore, inspired by automated segmentation methods, deep learning models applicable to breast ultrasound image segmentation tasks began to gradually enter the public eye with the development of deep learning. In addition to U-Net [6] and its derivative models [10,11], network structures such as DeepLab [21] have been fully demonstrated in the field of medical image segmentation. On breast ultrasound images, Yap et al. [22] designed a pre-trained FCNAlexNet model, which is more targeted than other segmentation models but performs significantly worse for irregularly shaped target regions. Almajalid et al. [23] effectively improved the segmentation accuracy of the model by enhancing contrast and denoising operations, but the method corrupted the original image. On the other hand, Chen et al. [24] achieved a reliable feature description of breast tumors by integrating U-Nets with different depths and shared weights, further improving the robustness of such methods. In addition, Chen et al. [25] proposed a network using bi-directional perceptual guidance, which has stronger robustness and more accurate segmentation results. However, the attention mechanism used by the model has less influence on the feature dimension, resulting in mediocre results on complex images.

2.1. Multi-scale features

Extracting and utilizing multi-scale feature information in images is a mainstream idea in target detection tasks. Under its influence, improving segmentation accuracy by processing feature information of different dimensions has gradually become a research direction in segmentation tasks. In the segmentation task of breast tumor images, the more novel STAN [26] and ARF-Net [27] adopt the idea of a multi-scale receptive field. Among them, dilated convolution and feature pyramid structures perform best in practical applications. The former is the main improvement in the feature extraction stage by using different expansion rates in the convolutional structure to achieve different sizes of



Fig. 1. Schematic diagram of different stages of breast ultrasound images, where the red outline part is the segmented target area.

perceptual fields with less computation and thus obtain feature information of different dimensions. At the same time, the latter is the main improvement in the combination feature information stage by fully integrating the obtained multi-dimensional feature information to get better output results.

Dilated convolution is more widely used in practical tasks of breast ultrasound imaging. Irfan et al. [28] proposed a semantic segmentation network using dilated convolution for end-to-end tasks. Zhuang et al. [29] applied dilated convolution and residual blocks to U-Net (RDAU-Net) and achieved considerable improvements in various metrics. Byra et al. [30] used dilation convolution to control the receptive field and conducted an in-depth study on its effect on the segmentation effect of breast tumors based on (SK) U-Net. On the other hand, Yan et al. [31] designed an attention-enhanced U-Net based on hybrid dilation convolution to reduce the loss of spatial information in the segmentation task of breast ultrasound images. However, the performance of the dilation convolution-based algorithm is not robust enough in terms of robustness and accuracy, and its performance is often unsatisfactory in the face of blurred target boundaries and low-resolution image subjects. The reason is that in most of these works, a one-sided view is held that the larger the receptive field contributes more the accuracy of the model. In contrast, the impact of the detailed information held by small receptive fields on the segmentation effect is ignored.

After extracting multi-scale features, how to integrate high-frequency local information with low-frequency global information becomes the focus of subsequent research. Liu et al. [32] proposed a Single Shot multi-box Detector (SSD) obtain different features. In order to focus more on the utilization of low-frequency information, Shrivastava et al. [33] abandoned the jump connection. They adopted the detection head structure of the model to the actual task. Feature Pyramid Network (FPN) has outstanding performance in the feature integration stage, and its related improvement work is gradually enriched. Li et al. [13] used the FPA module and GAU to construct a PAN. However, the development of feature pyramid structure in segmentation tasks has been relatively slow because most of the research on segmentation tasks has focused on the dramatically changing boundary parts of the image and less on low-frequency global information.

2.2. Attention mechanism

Inspired by the animal visual system's behavioral patterns and neuronal structures, attention mechanisms have been developed significantly in computer vision. Itti et al. [34 35] proposed an attention model based on image saliency. The model integrated features in three dimensions of color, luminance, and orientation in images with the support of the Gaussian pyramid algorithm. Woo et al. [36] suggested a Convolutional Block Attention Module (CBAM) as a plug-and-play module for integrating the channel information in the image space. In addition to the classical idea of applying attention mechanisms to enhance image feature information, attention mechanisms are also used for specific fine-tuning target problems for specific tasks in different domains. Ren et al. [37] designed a pyramidal self-attention module for target detection. On the other hand, Li et al. [38] designed an AttentionFGAN for spectral problems. In addition, Fan et al. [39] also introduced the reverse attention module for extracting a sufficient amount of edge information when segmenting the lung infection region of COVID-19 patients.

In the ultrasound image segmentation problem, Lee et al. [40] utilized a channel attention module to produce better segmentation results. On the other hand, Xue et al. [41] combined a spatial attention mechanism with a channel attention mechanism to cope with the irregular boundaries of breast ultrasound images. Punn and Agarwal [42] developed an inception U-Net guided by cross-spatial attention-guided inception U-Net, which significantly improved the segmentation ability in breast ultrasound images. Lei et al. [43] integrated various attention mechanisms and designed a more efficient network model to segment

tumors. However, this model still has a significant error on the edges of the segmentation target. Chen et al. [44] used a hybrid adaptive attention module instead of convolutional operations to construct the U-Net, which improved segmentation accuracy and generalization ability.

3. AMS-PAN for breast ultrasound image segmentation

To address the problem of poor clarity and blurred edges of breast ultrasound images that are difficult to segment, this study proposes a pyramid structure network, AMS-PAN, that combines attention mechanisms with multi-scale features. Fig. 2 illustrates the main structure of AMS-PAN. The network consists of two main parts encoder and decoder components. In the encoder part, depthwise separable convolution is used to increase the depth of the model while decreasing the number of model parameters, forming different size perceptual fields by stacking different scale convolutional layers and using them to extract feature information of the image. In the decoder part, GAU is used to fuse the feature maps of adjacent layers, and SCA is used to further highlight the primary information in the fused feature maps to achieve better performance in the final segmentation.

As far as the composition of the encoder is concerned, multi-scale receptive fields are realized by convolutional layers of different sizes for extracting feature information at different levels. However, the depth of the network formed by sequentially connecting each scale receptive field from the original image is too shallow to obtain in-depth feature information with richer dimensionality. To solve this problem and increase the network depth, we split the underlying convolutional layers into the channel and spatial convolution and use residual connections between adjacent feature layers, as shown by the red connecting lines in Fig. 2. With the residual structure, the model can further extract high-dimensional features on top of low-dimensional features and obtain a feature pyramid composed of different levels of feature information.

The decoder is composed of two parts, GAU and SCA. As shown in the orange part in Fig. 2, the primary role of GAU is to integrate the effects of different feature layers on the segmentation results by fusing the features with similar levels in both parts into a new intermediate feature layer. As shown in the pink part in Fig. 2, the fused middle feature layer will be passed to the deeper decoder from top to bottom. The SCA module is used to focus the attention of the network model on the high-frequency feature information in each layer, which has a more substantial discriminative effect on the fuzzy edges in breast ultrasound images.

3.1. Multi-receptive fields encoder

In AMS-PAN, the core of the encoder part lies in the multi-scale feature extraction module, which uses four different sizes of perceptual fields for convolution, namely 3×3 , 5×5 , 7×7 , and 9×9 . as shown in Fig. 3. In order to deepen the network structure and reduce the model computation, all the perceptual fields implemented in this study are made by stacking 3×3 and 5×5 sized convolutional layers. In addition, each convolution operator is also disassembled into spatial and channel convolution under the guidance of depthwise separable convolution.

Specifically, in the encoder part, the underlying convolution operator is split into a channel convolution of 1×1 size and a spatial convolution of 3×3 or 5×5 size, and the size of the spatial convolution determines the receptive field of the convolution operator in it. On this basis, the perceptual areas of different scales can be formed by combining convolution operators of various sizes. In this design, two convolution operators of size 3×3 and one convolution operator of size 5×5 are used to form a receptive field of actual size 7×7 . In contrast, a 9×9 receptive field is realized by a 3×3 sized convolution operator and two 5×5 sized convolution operators. The specific structures and parameters corresponding to the different receptive fields in the actual design are given in Table 1. After each convolution operator, we add a

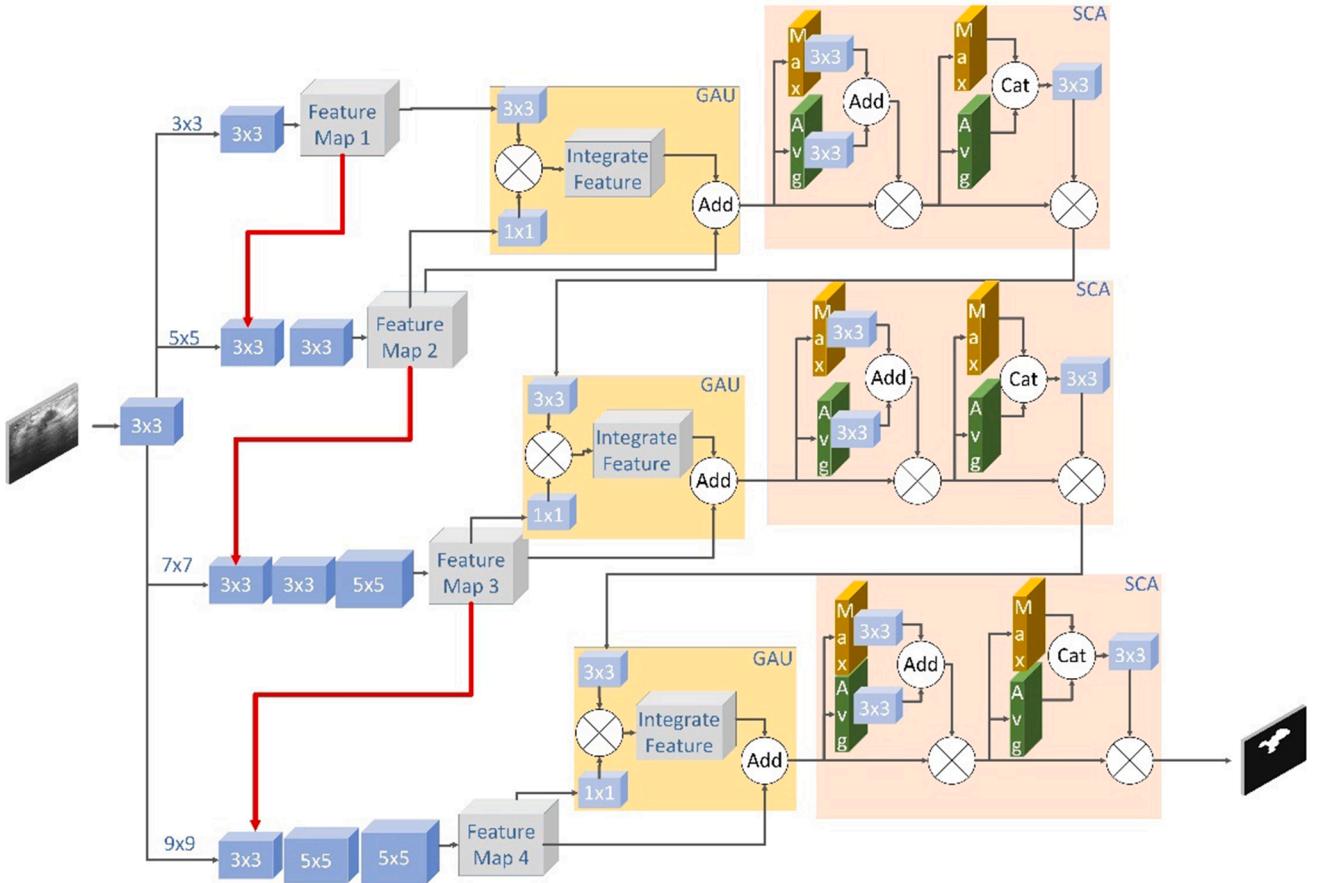


Fig. 2. Schematic diagram of the overall structure of the model. The orange part is the Global Attention Upsample (GAU), and the pink part is the Spatial and Channel Attention (SCA) module.

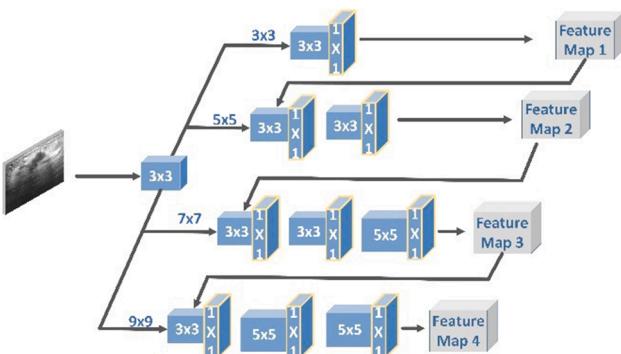


Fig. 3. Schematic diagram of the feature extraction module for multi-scale receptive fields.

Table 1
Parameters of different receptive field.

| Receptive field | kernel size | stride | padding |
|-----------------|-------------|--------|---------|
| 3×3 | 3 | 1 | 1 |
| 5×5 | 3 | 1 | 1 |
| 7×7 | 3 | 1 | 1 |
| | 3 | 1 | 1 |
| | 5 | 1 | 2 |
| 9×9 | 3 | 1 | 1 |
| | 5 | 1 | 2 |
| | 5 | 1 | 2 |

max pooling layer to reduce the image size for subsequent computations and use Rectified Linear Units (ReLU) to avoid the gradient vanishing problem.

In this stage, different sizes of receptive fields are used to extract feature maps of different dimensions. 3×3 and 5×5 size receptive fields can focus more on global low-frequency information. In contrast, 7×7 and 9×9 size receptive fields can focus more on local high-frequency information. However, a new problem caused by using different scales of receptive fields is that it is difficult to obtain enough local feature information as the basis for image segmentation. Therefore, we use residual connectivity in the encoder to use the shallow feature map as the input of the deep feature extraction module. This structure can obtain a richer representation of local features and then obtain a multi-scale feature layer with more distinct dimensional distinction to construct a feature pyramid.

3.2. Decoder with attention mechanism

In AMS-PAN, the decoder has four layers. Each layer can be divided into two main modules: GAU and SCA. The former is mainly used to fuse the results of adjacent feature layers to generate a more comprehensive and complete segmentation basis. The latter is a spatial attention module and a channel attention module designed to extract high-frequency feature information from the feature layers. Fig. 4 shows the organization of the decoder for each layer. The GAU module will process the shallow feature map and the deep feature map into a fused intermediate feature result. The fused feature maps are processed by the SCA module, in turn, by the spatial attention module and the channel attention module to further mine their edge information.

In the GAU module, 3×3 and 1×1 convolutional layers can extract

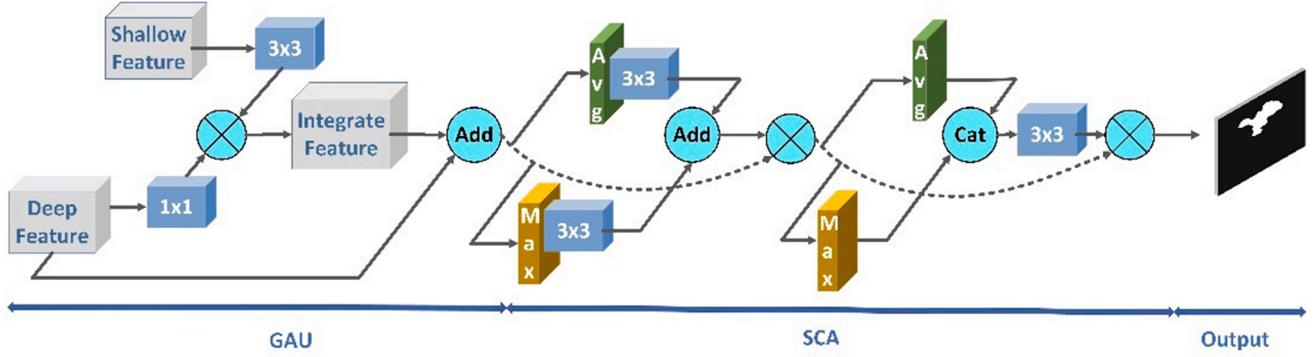


Fig. 4. PAN structure decoder with SCA module added.

high-dimensional information from shallow feature maps and deep feature maps, respectively. For the deep feature map, the smaller convolution kernel can effectively prevent over-abstraction of the extracted features, thus avoiding network overfitting and accelerating the gradient convergence. In contrast, for shallow feature maps, there is no need to consider the problem caused by the level of feature abstraction. We use a larger convolution kernel to extract as much local high-frequency information as possible from the feature map with global low-frequency information. After the convolution operation, we multiply the two feature maps together as the fused feature map. A residual connection is added between the deep feature map and the fused feature map to avoid gradient disappearance. This way, the relatively more abstract high-dimensional information can be better preserved, and the high-frequency feature information can be more fully expressed in the final output of GAU.

To describe the computational process intuitively, we use F_G is used to denote the output of GAU, F_d for the deep feature map, and F_s for the shallow feature map. Meanwhile. We use $\text{Cov}_{k,s,p}(\cdot)$ to represent the convolution operation with a kernel size of k , the stride of s , and padding of p . \otimes denotes the matrix multiplication. The calculation procedure is shown in Eq (1):

$$F_G = F_d + (\text{Cov}_{1,1,0}(F_d) \otimes \text{Cov}_{3,1,1}(F_s)) \quad (1)$$

As shown in equation (1), the shallow feature map F_s needs to go through a convolution block $\text{Cov}_{3,1,1}(\cdot)$ with a convolution kernel of 3×3 , stride of 1, and padding of 1, while the deep feature map F_d needs to go through a convolution block $\text{Cov}_{1,1,0}(\cdot)$ with a convolution kernel of 1×1 , stride of 1, and padding of 0. The final output F_G is obtained by multiplying the two convolved feature maps and adding them to the original deep feature map F_d through residual concatenation.

For the SCA module, the attention mechanisms are organized so that the spatial attention module comes first, and the channel attention mechanism comes second. Firstly, the equation (2) aims to illustrate the computation process of the spatial attention part in the SCA part as in Fig. 4, and we write F_{SA} as the output feature map of the spatial attention part. Then the feature map F_G is obtained in GAU, which needs to go through the AvgPool layer $\text{Avg}(\cdot)$ and MaxPool layer $\text{Max}(\cdot)$, respectively, to obtain the global and texture information in spatial dimensions. The feature results after the pooling part need to be further processed using a convolutional block $\text{Cov}_{3,1,1}(\cdot)$ with a convolutional kernel of 3×3 , the stride of 1, and padding of 1. This operation could obtain the relative position information on the two-dimensional space within the feature map. After that, in order to integrate the low-dimensional and high-dimensional information in the spatial dimension, the two convolutional results need to be added together, and the summed result is multiplied with the original F_G through the residual join to obtain the final output F_{SA} .

$$F_{SA} = F_G \otimes (\text{Cov}_{3,1,1}(\text{Avg}(F_G)) + \text{Cov}_{3,1,1}(\text{Max}(F_G))) \quad (2)$$

Secondly, the equation (3) aims to illustrate the computation process of the channel attention part in the SCA part as in Fig. 4, and we will note F_{CA} as the output feature map of the channel attention part. Then, the feature map F_{SA} is obtained in the spatial attention part, which needs to pass through the AvgPool layer $\text{Avg}(\cdot)$ and MaxPool layer $\text{Max}(\cdot)$, respectively, to reduce the feature map size and highlight the core information in it. We concatenate the feature results after the pooling part and further abstract them using a convolutional block $\text{Cov}_{3,1,1}(\cdot)$ with a kernel of 3×3 , stride of 1, and padding of 1. This operation could extract the deep information hidden in the channel dimension. $\text{Cat}(\cdot, \cdot)$ presents the concatenation. After that, the final output F_{CA} is obtained by multiplying it with the original F_{SA} through residual concatenation.

$$F_{CA} = F_{SA} \otimes \text{Cov}_{3,1,1}(\text{Cat}(\text{Avg}(F_{SA}), \text{Max}(F_{SA}))) \quad (3)$$

4. Experiments results and discussion

The AMS-PAN proposed in this study was trained and tested on BUSI and OASBUD, respectively. The components of the two datasets are shown in Table 2. The former includes 780 breast ultrasound images, divided into 133 normal images with ground truth as the label, 210 benign tumor images, and 437 malignant tumor images. The latter included 48 benign tumor images and 52 malignant tumor images.

The original sizes of the BUSI and OASBUD are not uniform; the BUSI image size is around 550×460 , and the OASBUD images are generated from the original mat files using MATLAB, whose image sizes are more diverse. Since the model cannot be uniform for pictures of different sizes, we choose to scale all pictures to 256×256 size considering the hardware conditions. Furthermore, the dataset is divided into training and test sets according to the 85% and 15% ratios. This study performs data augmentation on the breast ultrasound images in the training set, in which the images are randomly rotated, flipped, or contrast adjusted to improve the robustness of the model and the training performance.

4.1. Experiment setting

We implemented AMS-PAN using PyTorch v1.10 and trained and tested it on an NVIDIA GeForce RTX A4000 GPU. The hyperparameter settings for training are shown in Table 3.

4.2. Evaluation metrics

Currently, medical image segmentation tasks are evaluated using six

Table 2
The introduction of datasets.

| Dataset | Normal | Benign | Malignant |
|---------|--------|--------|-----------|
| BUSI | 133 | 210 | 437 |
| OASBUD | 48 | 52 | |

Table 3

Hyperparameters of AMS-PAN training.

| Settings | Parameters |
|-----------------------|-------------------|
| Epoch | 50 |
| Learning rate | 0.01 |
| Batch | 16 |
| Gradient decay policy | ReduceLROnPlateau |
| Patience epoch | 3 |
| Decay factor | 0.2 |

categories of metrics [442925]: Accuracy, Recall, Specificity, Precision, Dice, and Intersection over Union (IoU). The primary operator of these is the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) in the segmentation results. Eq (4) - Eq (9) shows the specific calculation process of the above six indicators.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (8)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (9)$$

In addition to IoU and Dice, which are essential visual metrics for the segmentation task, other metrics have their focus. Specifically, Accuracy refers to the proportion of pixels in the segmentation result that is correctly segmented as foreground or hard ground; Precision refers to the proportion of pixels in the segmentation result that is correctly segmented as foreground; Recall refers to the proportion of all

foreground pixels in the ground truth that are correctly segmented as foreground by the model; Specificity refers to the proportion of all hardground pixels in the ground truth that the model correctly segments as rear view.

4.3. Training procedure

In this study, initial screening of optimizers is performed for the network structure of AMS-PAN. The candidate optimizers include the first-order momentum-based SGD iterator, the second-order momentum-based RMSprop iterator, and the Adam iterator, combining both characteristics. Before training, we perform the same preprocessing on the two datasets and combine them into one, forming a mixed dataset with 880 underlying data. Fifty training rounds are performed on each optimizer to verify the performance of different optimizers in the segmentation task of breast ultrasound images. The variation of the metrics of other optimizers during the training process is shown in Fig. 5.

As seen in Fig. 5, the training process of the SGD optimizer is challenging to maintain stability. Under the constraint of this optimizer, the model's ability to discriminate incorrect samples is not stable, resulting in its volatility in Accuracy and Specificity, which first decreases and then increases. In addition, the SGD constraint makes AMS-PAN's segmentation on correct samples unsatisfactory, and its performance on other metrics is far inferior to the other two types of optimizers. In contrast, the second-order momentum constraint is better. The Adam and RMSprop optimizers converge significantly and perform much better than SGD on various metrics. In terms of Dice, IoU, Precision, and Recall metrics, the training effect of Adam are better than that of the RMSprop method. Therefore, in this study, with the Adam method as the training optimizer, the model's main metrics variation with epoch on the training and test sets is shown in Fig. 6.

4.4. Ablation experiment

This study mainly designs a coding and decoding network applied to breast ultrasound image segmentation. On the coding side, we designed a multi-scale feature extraction method. We have improved the PAN

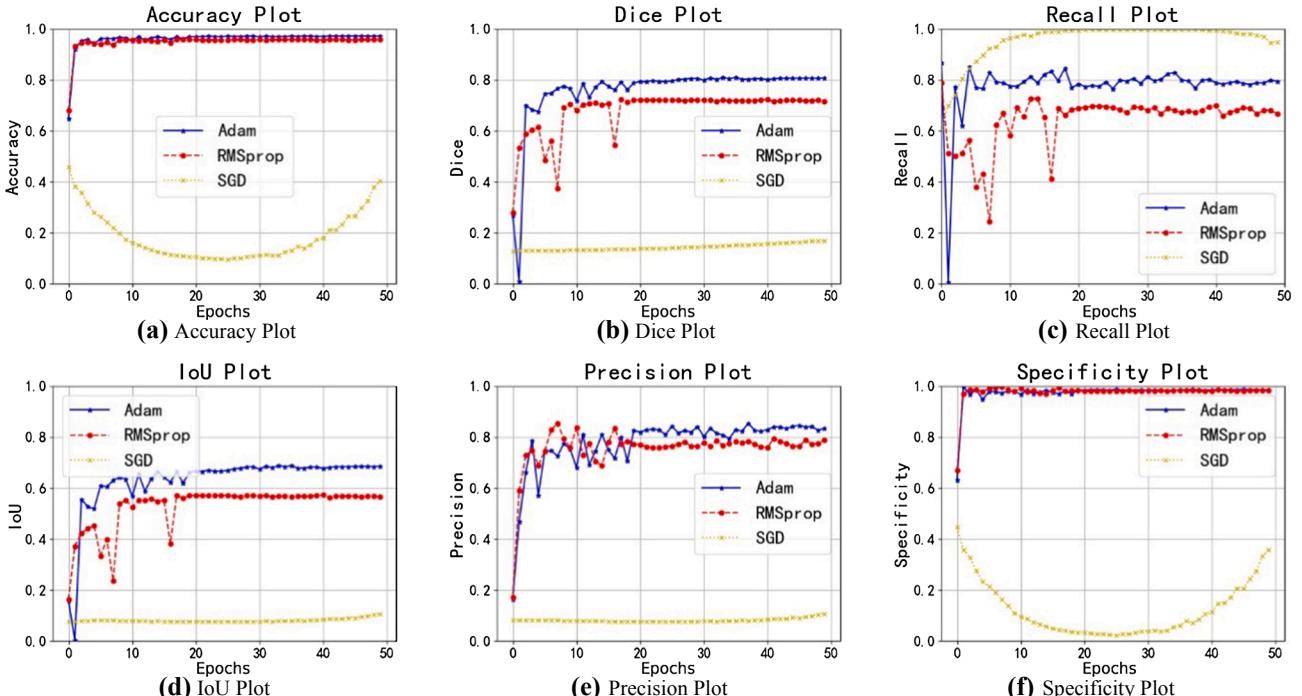


Fig. 5. Comparison chart of the variation of segmentation metrics for different optimizers.

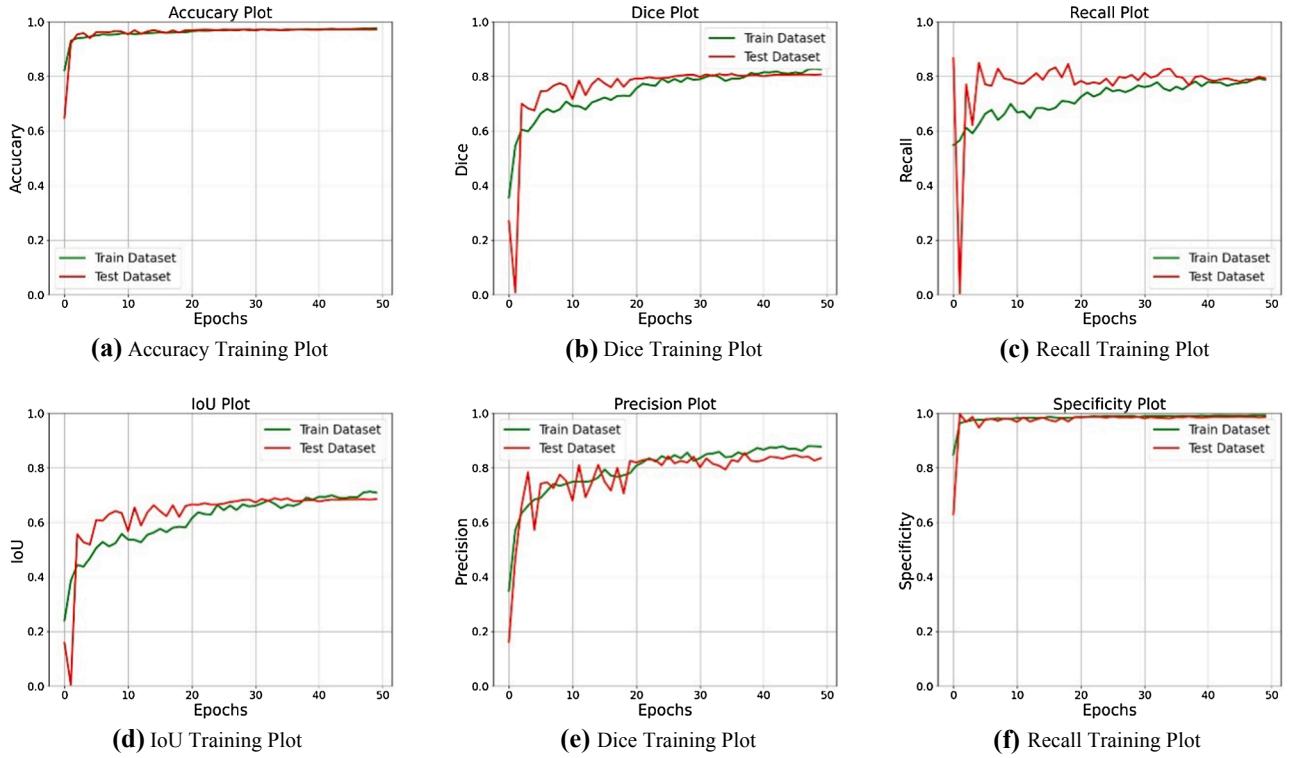


Fig. 6. AMS-PAN model training process different metrics change process.

structure by replacing the original FPA module with an SCA module that places more emphasis on high frequency features. Ablation experiments are designed for each part of this study to verify the effectiveness and advancedness of the above structure. For the encoder, a model without the multi-scale feature extraction module is designed as a comparison in this study. For the decoder, a model without the SCA module is designed for the SCA module as a comparison. Since the GAU structure in the decoder is derived from the underlying structure of PAN, this study uses the models of the underlying structure of PSPNet [45] and the underlying structure of FPN as comparisons, respectively, as a way to verify the effectiveness of the multi-scale feature fusion approach of GAU. The results of the ablation experiments on BUSI and OASBUD are shown in Table 4, respectively. The multi-scale feature extraction module, the SCA module, and the GAU module all improved on both datasets. On BUSI, AMS-PAN is only not optimal in terms of Specificity, but its value of 98.54% still indicates that it performs well for segmentation specificity in the anterior and posterior views and that it is better at excluding non-focal parts from the segmentation target.

Explicitly analyzing the performance of the ablation experiments, in Table 4, for BUSI, the multi-scale feature extraction module and the SCA module showed evident improvements in Dice, IoU, and Recall metrics. The multi-scale feature extraction module improves the model by

20.21%, 34.69%, and 34.61% in Dice, IoU, and Recall, respectively. The SCA module, on the other hand, improves 7.00%, 12.75%, and 7.92% in the above three metrics, respectively. The feature fusion method of GAU improves 12.05%, 20.63%, and 9.80%, and 11.69%, 19.52%, and 14.41%, respectively, compared to the feature fusion approach of PSPNet. In the above three metrics, the advantages of AMS-PAN are more evident in different metrics due to the lower contrast of OASBUD. In Fig. 7, the segmentation effect plots of other models in the ablation experiment for both types of datasets are given.

The Receiver Operating Characteristic (ROC) and Precision-Recall (P-R) plots of the ablation experiments models in two different data sets are given in Fig. 8. Among them, Fig. 8 (a) and (b) show the ROC and PR plots on BUSI, respectively, and Fig. 8 (c) and (d) show the ROC and PR plots on OASBUD, respectively.

As seen in Fig. 8 (a) and (c), all three core modules, including the GAU feature fusion approach, contribute to the model performance improvement. In BUSI, the Area Under Curve (AUC) value of the AMS-PAN model is 0.7517, and the GAU feature fusion approach brings an improvement of 0.0188 to 0.02, which is 1.72% compared to the model without the addition of the multi-scale feature extraction module, and 2.64% compared to the model without the SCA mechanism. In OASBUD, the AUC value of the AMS-PAN model improves by 9.17% compared to

Table 4

Comparison of segmentation metrics of different ablation experiments on BUSI and OASBUD, where the optimal performance has been shown in bold.

| Datasets | Ablation Models | Accuracy | Dice | IoU | Recall | Precision | Specificity |
|----------|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BUSI | No attention | 96.13 | 75.43 | 60.78 | 73.48 | 82.15 | 98.53 |
| | No multi scale | 95.76 | 67.14 | 50.88 | 58.91 | 82.27 | 98.93 |
| | FPN as decoder | 96.20 | 72.03 | 56.81 | 72.22 | 72.75 | 97.91 |
| | PSPNet as decoder | 95.75 | 72.26 | 57.34 | 69.31 | 76.23 | 98.09 |
| | AMS-PAN | 97.13 | 80.71 | 68.53 | 79.30 | 83.50 | 98.54 |
| OASBUD | No attention | 96.90 | 77.56 | 64.26 | 72.99 | 84.30 | 98.88 |
| | No multi scale | 96.16 | 57.08 | 41.08 | 53.76 | 63.52 | 98.38 |
| | FPN as decoder | 97.27 | 74.28 | 59.17 | 69.92 | 79.37 | 98.91 |
| | PSPNet as decoder | 96.15 | 68.09 | 52.59 | 69.26 | 68.62 | 97.91 |
| | AMS-PAN | 97.97 | 79.62 | 67.52 | 74.43 | 87.92 | 99.38 |

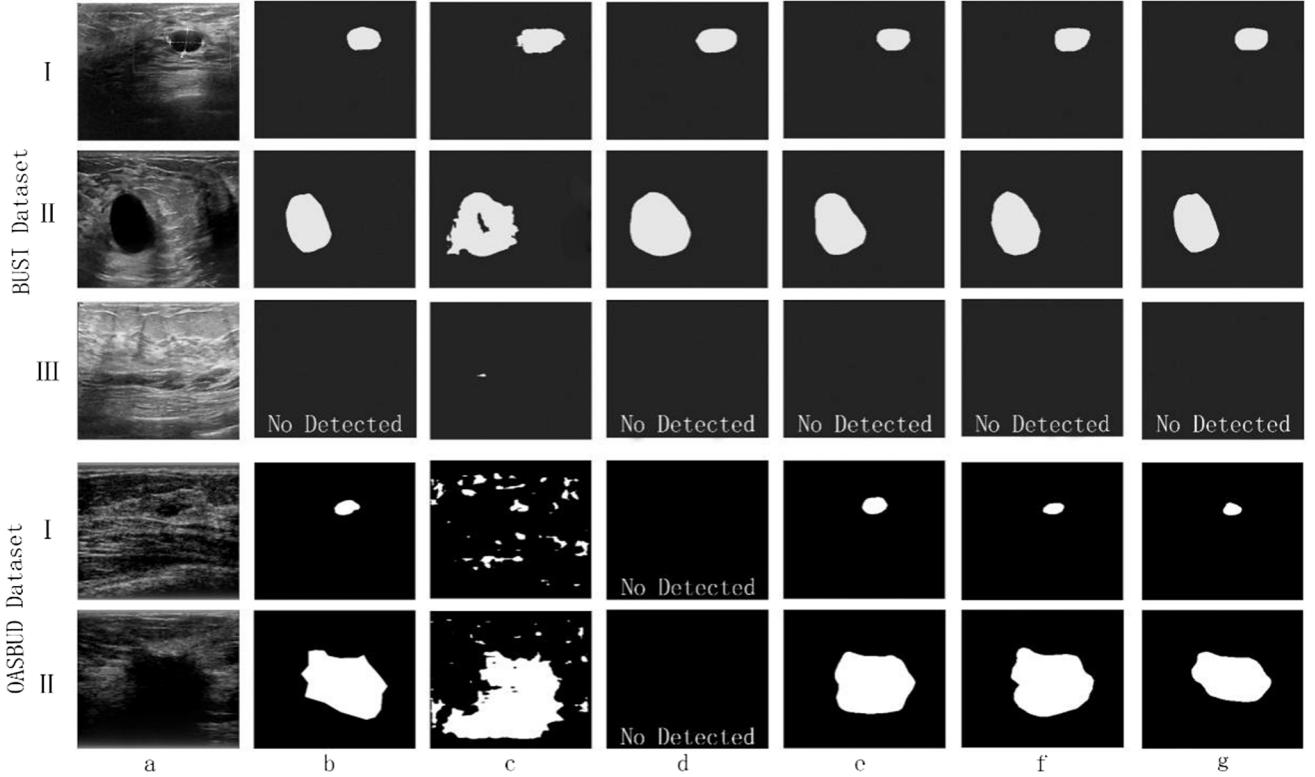


Fig. 7. The results of ablation experiments for different improvements. Row I shows the benign breast tumor images and model results from two datasets, row II shows the malignant breast tumor images and model results from two datasets, and row III shows the normal breast images and model results. Columns a-g represent the original images, ground truth, the results of No SCA module, No multi-scale feature extraction, FPN as decoder, PSPNet as decoder and AMS-PAN, respectively.

the model without adding the SCA mechanism, and the contribution made is much higher than the other components. As can be seen from Fig. 8 (b) and (d), In BUSI, the PR curves and mAP values are approximately the same for all experimental ablation models. Still, in OASBUD, the P-R curve of AMSPAN performs significantly better than the other models, especially for the model without the addition of the SCA module. Specifically for the mean Average Precision (mAP) metrics, on BUSI, the multi-scale feature extraction module, the feature fusion method of GAU, and the SCA module can improve by 0.26%, 0.38%, and 0.38%, respectively. On the OASBUD, the multi-scale feature extraction module and the feature fusion approach of GAU contribute 0.18% and 0.10%, respectively. In comparison, the SCA module brings an improvement of 1.36%, which plays a more significant role in images with lower contrast.

4.5. Comparison experiments

Deep learning-based methods have started to gain more attention with the development of deep learning in recent years. In medical image segmentation, the more mature deep learning methods mainly include U-Net [6], U-Net with Spatial and Channel'Squeeze Excitation' Attention (Att U-Net) [10], U-Net++ [11], DeepLabV3 [21], and LinkNet [46]. These methods have been improved considerably in past research. They have good and superior performance in most medical image segmentation tasks and gradually become the paradigm methods and the first choice for deep learning applications in the medical image domain. The segmentation metrics of the above methods for BUSI and OASBUD are shown in Table 5. AMS-PAN outperforms the current mainstream deep learning segmentation methods in all metrics for both datasets. Among the six metrics, the only underperformance of AMS-PAN is the Recall metric on the OASBUD, which is the next best method on that value with a value of 74.43%, and this Recall performance is sufficient to show that for the OASBUD, AMS-PAN still has some ability to identify

real lesions out of the box accurately.

As can be seen in Table 5, the AMS-PAN model does not show remarkable improvements in Accuracy, Recall, Precision, and Specificity compared to the more mature deep learning models, with improvements of no more than 1.00%. In contrast, Dice and IoU upgrade dramatically. For the Dice metric, the best performer on BUSI was the DeepLabV3 model, and the best performer on the OASBUD was the U-Net++ model, while AMS-PAN improved 8.73% and 7.39%, respectively, compared to the above two. For the IoU metric, AMS-PAN improves by 9.84% over the LinkNet model, which is the best performer on BUSI, and by 4.83% over the U-Net++ model, which is the best performer on the OASBUD and has better segmentation capability. The only underperformance of AMS-PAN was the Recall metric on the OASBUD, which was the best performer on the OASBUD. With a value of 74.43%, it was the following best method for this value. This Recall performance is sufficient to show that AMS-PAN still has some ability to identify natural lesions for the OASBUD accurately. Fig. 9 shows the experimental results of comparing AMS-PAN with other deep learning methods.

The ROC and PR plots for comparing the experimental models in two different data sets are given in Fig. 10. Among them, Fig. 10(a) and (b) show the ROC and PR plots for BUSI, respectively, and Figure Fig. 10(c) and (d) show the ROC and PR plots for the OASBUD, respectively.

Fig. 10(a) and (b) show that AMS-PAN outperforms the other deep learning models on BUSI. Its mAP value also improves by 0.0021% to 0.0052% compared with other models, which is not significantly improved. Nevertheless, combined with the fact that its AUC value also has increased from 1.11% to 3.50%, it can be seen that its segmentation accuracy on BUSI is more robust than other deep learning models. Fig. 10(c) and (d) show that the AMS-PAN model performs significantly better than other models on the ROC curve and PR curve for the OASBUD with poor clarity. Among the other deep learning models, the best performer is the U-Net++ model, and the worst is the DeepLabV3

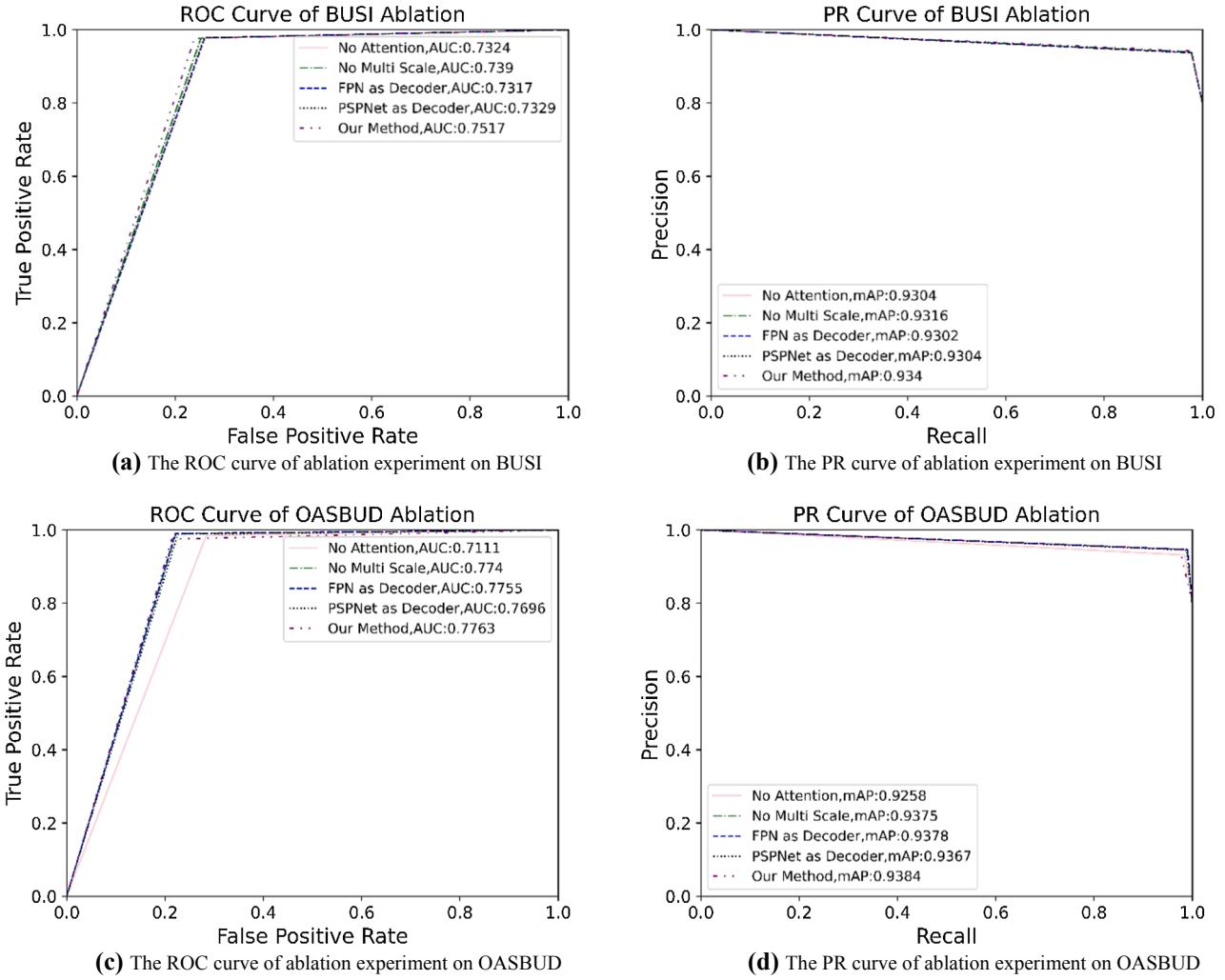


Fig. 8. ROC curves and PR curves of different ablation experimental models on BUSI and OASBUD.

Table 5

Comparison of segmentation metrics for deep learning methods on BUSI and OASBUD, where the optimal performance has been shown in bold.

| Datasets | Contrast Models | Accuracy | Dice | IoU | Recall | Precision | Specificity |
|----------|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BUSI | U-Net | 95.27 | 67.98 | 57.58 | 72.01 | 70.96 | 96.67 |
| | Att U-Net | 96.94 | 69.12 | 59.07 | 77.14 | 69.43 | 96.40 |
| | U-Net++ | 96.03 | 71.54 | 62.14 | 78.50 | 71.71 | 96.73 |
| | DeepLabV3 | 96.64 | 74.23 | 62.20 | 78.72 | 71.68 | 96.77 |
| | LinkNet | 96.21 | 72.04 | 62.39 | 75.21 | 71.38 | 96.99 |
| | AMS-PAN | 97.13 | 80.71 | 68.53 | 79.30 | 83.50 | 98.54 |
| OASBUD | U-Net | 96.95 | 66.15 | 53.76 | 72.37 | 67.65 | 98.22 |
| | Att U-Net | 95.01 | 54.62 | 30.29 | 35.88 | 66.04 | 98.81 |
| | U-Net++ | 97.13 | 74.14 | 64.41 | 76.07 | 83.95 | 98.83 |
| | DeepLabV3 | 96.59 | 63.95 | 48.52 | 55.62 | 79.15 | 99.10 |
| | LinkNet | 96.28 | 57.41 | 48.14 | 58.55 | 73.02 | 98.64 |
| | AMS-PAN | 97.97 | 79.62 | 67.52 | 74.43 | 87.92 | 99.38 |

model. Regarding AUC values, the AMS-PAN model improved by 22.43% compared to DeepLabV3 and 1.32% compared to U-Net++, which is a considerable increase. As for the mAP value, the AMS-PAN model improved by 0.0021% compared with U-Net++ and 0.0265% compared with DeepLabV3, and the segmentation effect is more pronounced.

On the other hand, specialized deep learning models have emerged for the segmentation task of breast ultrasound images. The current SoTA methods that perform well are RDAU-Net [29], (SK)U-Net [30], AE U-Net with HDC [31], GCNet [41], and AAU-Net [44]. This study designed

a comparison experiment on the performance of the above models on BUSI, which is more widely recognized, to illustrate the advancement of AMS-PAN as a new method. Moreover, the results are shown in Table 6.

As can be seen in Table 6, AMS-PAN is the leading performer in all six metrics. Compared to AAU-Net, which performs better overall, the proposed method improves by 0.05% in Accuracy and by 1.72% in Precision. As for Dice and IoU, the two most intuitive metrics in medical image segmentation tasks, AMS-PAN improves 1.89% and 0.82%, respectively, compared to AAU-Net in these two metrics. In Recall, the proposed method improves by 0.82 compared to RDAU-Net, which has

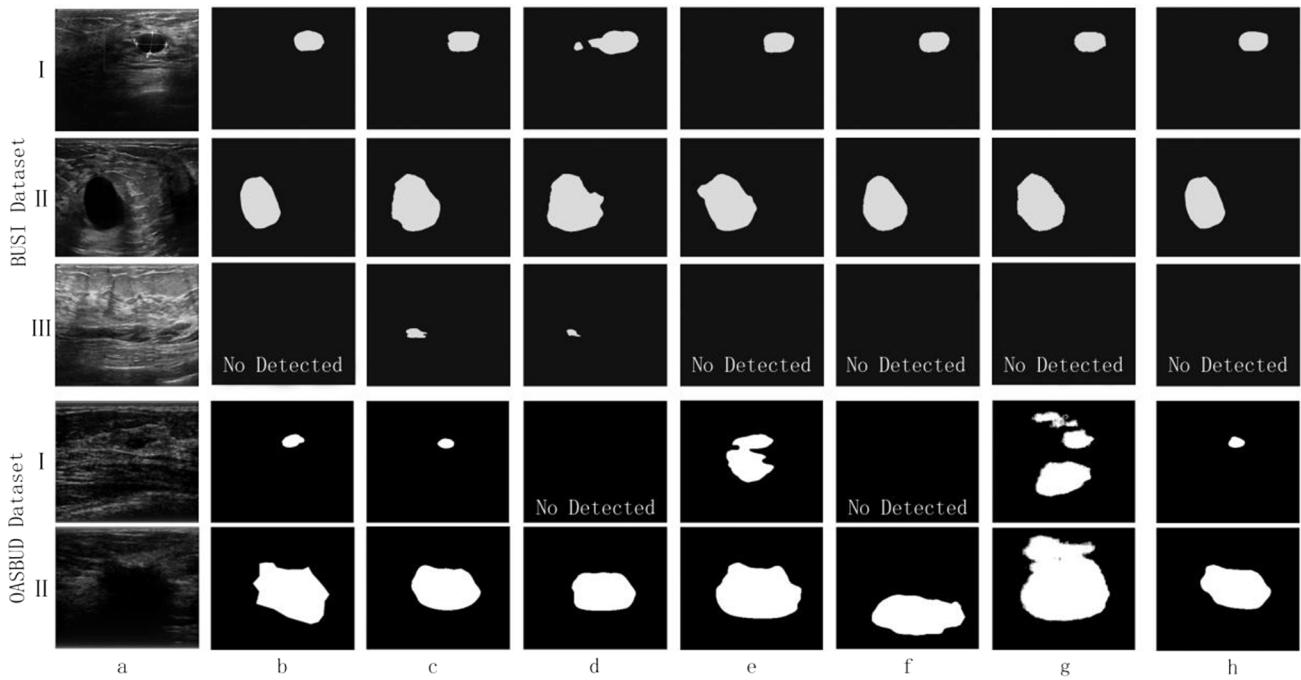


Fig. 9. The results of comparison experiments for deep learning methods. Row I shows the benign breast tumor images and model results from two datasets, row II shows the malignant breast tumor images and model results from two datasets, and row III shows the normal breast images and model results. Columns a-h represent the original images, ground truth, the results of U-Net, U-Net with scse, U-Net++, DeepLabV3, LinkNet and AMS-PAN, respectively.

the next best performance, and 0.28 compared to GCNet in Specificity.

In addition to the comparison mentioned above, experiments on deep learning and non-deep learning methods are still the most widely used segmentation methods in computer-aided diagnostic tools. In the segmentation task of breast ultrasound images, four main methods that have been developed soundly and widely used are Marker Watershed (MW) [18], level set [17], MS [19], and AMSMW [20]. In this study, the above methods were used as comparison experiments to verify the feasibility of the practical application of the model to assist diagnosis.

The results of the metrics for the comparison experiments against the non-deep learning methods are shown in Table 7. Combining the performance of the two datasets, AMS-PAN offers different degrees of improvement over the traditional non-deep learning methods in the metrics of Dice and IoU, with at least 2.31% and 3.19% revision in BUSI, respectively, and at least 3.37% and 3.39% improvement in the OASBUD. However, on the BUSI, the best-performing method for Recall was the MW method at 97.64%, while on the OASBUD, the best performer for Recall was the AMSMW method. There is a significant gap in our model in this metric due to its poor grasp of the overall shape of the segmented target in critical images. In addition, on the Precision metric, our model also performs sub-optimally on the BUSI, with a 2.13% drop-off compared to the AMSMW method, with this gap focusing on the model's more aggressive segmentation edge detail for small targets. The effect plots of the comparison experiments using different non-deep learning methods are shown in Fig. 11.

5. Discussion

Most of the existing studies have focused on how to extract an abstract representation of local high-frequency information, while the utilization of low-frequency features is relatively limited. On the other hand, this study constructs a feature pyramid through a multi-scale feature extraction module, which effectively uses high-frequency and low-frequency information and uses an attention mechanism to control the model to focus more on the critical parts of the feature information during decoding.

The encoder with a multi-scale receptive field and the decoder with

an increased attention mechanism are the core innovations of AMS-PAN. In the past, most encoders used deep sequential convolution or residual structure for feature extraction. Despite the deeper network level, the model structure is single and has a surplus of extraction capability for high-dimensional features and insufficient utilization capability for low-dimensional features. The encoder designed in this study can extract features of different dimensions by different perceptual fields and effectively balance low-frequency global information and detailed high-frequency information by using small-sized convolutional kernel stacking. Feature extraction at different scales enables the model to retain shallow features while mining deep features rich enough, forming a feature pyramid of multi-scale feature layers. This study improves the PAN model for the decoder by maintaining the GAU structure for integrating features between adjacent layers. However, the contribution of the FPA module in the original PAN structure to the local information is limited, and it isn't easy to further process the high-frequency information in it. Its segmentation capability is unsatisfactory in the face of poor clarity and low contrast of breast ultrasound images. Therefore, this study designs the SCA module in the decoder so that the model can still mine the irregular edges of the segmentation target region in the poor quality data, focus on the more critical high-frequency local information, and improve the segmentation effect of the model on the blurred edges in the images.

The ablation experiments show that the three types of core modules on AMS-PAN contribute more significantly to its segmentation capability. As can be seen in Fig. 7, the multi-scale feature extraction module, on the other hand, provides the model with richer feature information, which enables the AMS-PAN model to perform better on data sets with poor clarity represented by OASBUD. While the GAU feature fusion approach can refine the model's segmentation granularity for large-size targets, the SCA module benefits from its prominent representation of in-depth features, providing more substantial stability to the model and improving the segmentation accuracy. In the comparison experiments, the AMS-PAN model can achieve better segmentation results when comparing deep learning methods. Combined with Fig. 11, it can be seen that the advantage of the AMS-PAN model is reflected in the segmentation of large targets with irregular edges compared with non-deep

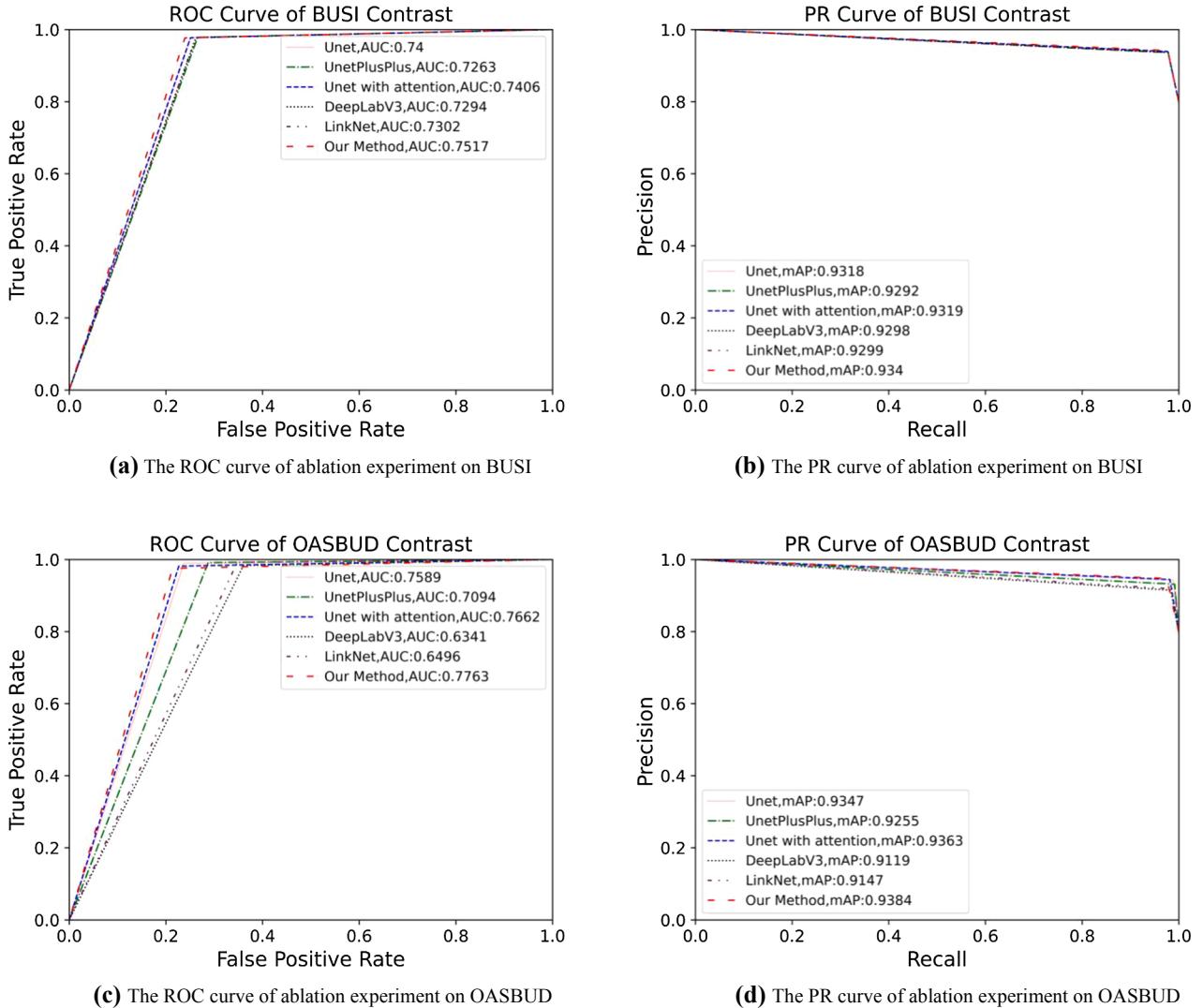


Fig. 10. ROC curves and PR curves of different comparison experimental models on BUSI and OASBUD.

Table 6
Comparison of segmentation metrics for SoTA methods on BUSI and OASBUD, where the optimal performance has been shown in bold.

| Contrast Models | Accuracy | Dice | IoU | Recall | Precision | Specificity |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| RDAU-Net | 93.91 | 70.58 | 61.17 | 78.48 | 71.09 | 97.59 |
| (SK)U-Net | 95.63 | 70.90 | 61.11 | 75.29 | 72.15 | 95.19 |
| AE U-Net with <i>HDC</i> | 90.28 | 70.36 | 60.67 | 75.25 | 69.70 | 96.13 |
| GCNet | 96.92 | 66.23 | 58.55 | 66.87 | 70.5 | 98.26 |
| AAU-Net | 97.08 | 79.21 | 67.97 | 77.27 | 81.78 | 97.75 |
| AMS-PAN | 97.13 | 80.71 | 68.53 | 79.30 | 83.50 | 98.54 |

learning methods, and the segmentation results of this model have more complete edge details.

Due to the application of depthwise separable convolution, AMS-PAN has better segmentation performance while retaining a relatively small computational volume. Combining Tables 5 and 8, it can be seen that, compared to the LinkNet model with the smallest volume, AMS-PAN trades a 3.97G increase in the number of parameters and a 3.43 billion increase in the number of operations for its Dice metrics of 12.03–38.69% and 25.87–40.26% on IoU. In fact, with the same hardware condition of RTX A4000 GPU, the segmentation time of AMS-PAN for a single image is 13.59 ms, which is only 1.57 ms more than the shortest LinkNet model, and such a time difference is negligible in real

applications. We consider that the primary time consumption of the deep learning method occurs in the pre-processing process of image data rather than the automated segmentation process. The time consumption of this process can be effectively reduced when the hardware conditions are satisfied.

Compared with the traditional segmentation methods based on non-deep learning, the model has no clear improvement in accuracy, specificity, or even mediocre performance in Recall and Precision. It is found that the segmentation results of AMS-PAN are somewhat different from ground truth but rich in edge details, and the segmentation results of non-deep learning methods are more similar to ground truth but poor in the continuity of edge details. In future work, the research will focus on how to absorb the advantages of non-deep learning methods and further improve the segmentation performance of the AMS-PAN model for irregular edge targets.

6. Conclusion

We propose a model called AMS-PAN for the segmentation task of breast ultrasound images in this study. As a codec network, the model mainly consists of two major parts, encoder, and decoder. The encoder mainly consists of a multi-scale feature extraction module. The module extracts global and local information in ultrasound images by combining the multi-scale receptive field with the idea of residuals. Based on PAN's

Table 7

Comparison of segmentation metrics for non-deep learning methods on BUSI and OASBUD, where the optimal performance has been shown in bold.

| Datasets | Contrast Methods | Accuracy | Dice | IoU | Recall | Precision | Specificity |
|----------|------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BUSI | MW | 94.37 | 71.80 | 56.68 | 97.64 | 82.27 | 96.58 |
| | level set | 94.69 | 69.58 | 54.83 | 95.56 | 80.14 | 97.21 |
| | MS | 95.76 | 71.57 | 56.72 | 63.05 | 79.36 | 96.36 |
| | AMSMW | 96.59 | 78.40 | 65.34 | 83.29 | 85.63 | 98.23 |
| | AMS-PAN | 97.13 | 80.71 | 68.53 | 79.30 | 83.50 | 98.54 |
| OASBUD | MW | 92.21 | 59.58 | 58.32 | 69.63 | 81.64 | 97.72 |
| | level set | 93.46 | 68.17 | 46.83 | 68.26 | 85.53 | 98.18 |
| | MS | 97.65 | 76.25 | 62.39 | 71.35 | 82.29 | 99.05 |
| | AMSMW | 98.12 | 74.93 | 64.13 | 75.29 | 86.33 | 98.84 |
| | AMS-PAN | 97.97 | 79.62 | 67.52 | 74.43 | 87.92 | 99.38 |

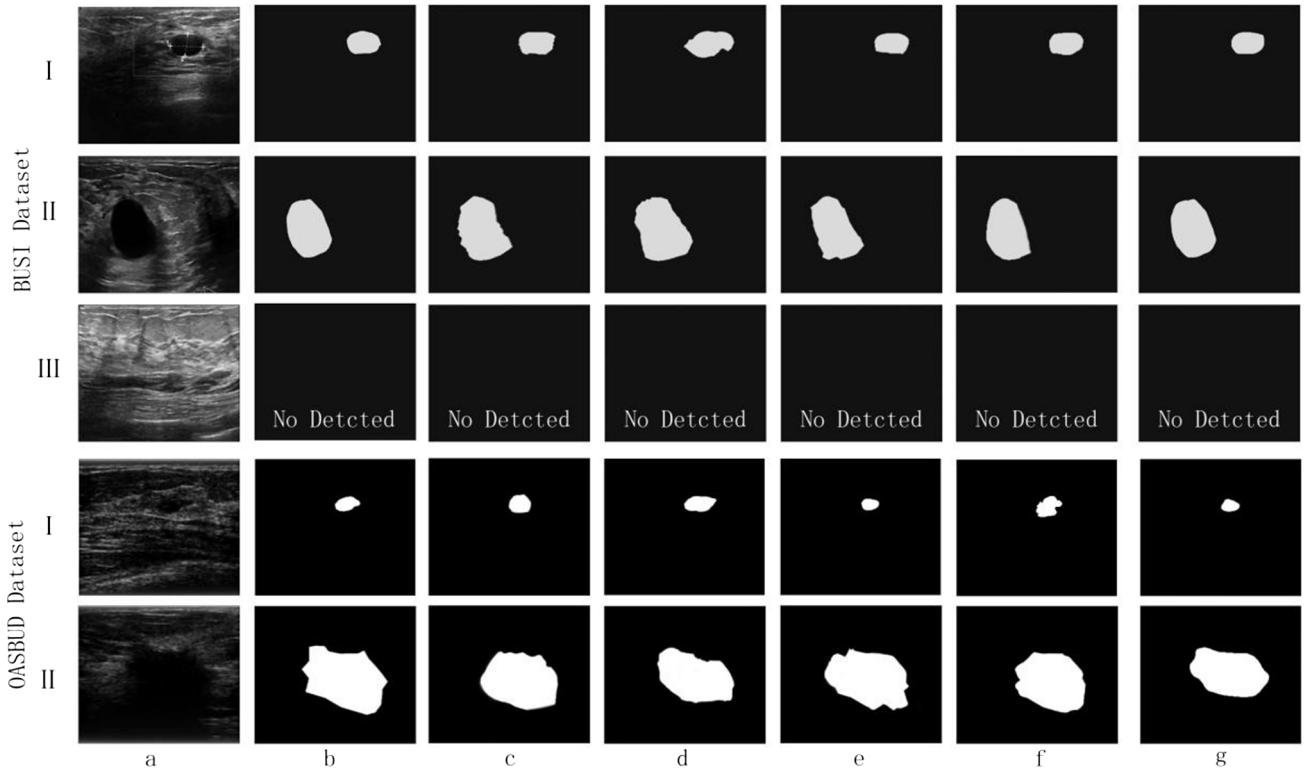


Fig. 11. The results of comparison experiments for non-deep learning methods. Row I shows the benign breast tumor images and model results from two datasets, row II shows the malignant breast tumor images and model results from two datasets, and row III shows the normal breast images and model results. Columns a-g represent the original images, ground truth, the results of MW, level set, MS, AMSMW and AMS-PAN, respectively.

Table 8

Compares the number of parameters, computation, and computing time (in the environment of RTX A4000 GPU) for different deep learning models, where the optimal performance has been shown in bold.

| Method | Parameters (G) | MACs (billions) | Time (ms) |
|-----------|----------------|-----------------|--------------|
| U-Net | 24.44 | 7.84 | 12.54 |
| Att U-Net | 24.55 | 7.85 | 15.63 |
| U-Net++ | 26.08 | 18.43 | 13.89 |
| DeepLabV3 | 26.01 | 27.34 | 17.36 |
| LinkNet | 21.77 | 5.47 | 12.02 |
| AMS-PAN | 25.74 | 8.9 | 13.59 |

infrastructure, the AMS-PAN's decoder adopts the feature fusion method of GAU to integrate features in the feature fusion stage. Furthermore, to make the model better focus on the edge detail information of ultrasound images, the SCA module is designed in the decoder to enhance its ability to process high-frequency local information to improve the segmentation accuracy of the model for irregular tumor targets. In the

experiments, all three designed modules significantly enhanced the model's segmentation ability.

The AMS-PAN model proposed in this study performs well in the commonly used segmentation metrics, especially in the Dice and IoU metrics. Due to the relatively sharper images in BUSI, AMS-PAN achieves Dice and IoU of 80.71% and 68.53%, respectively, on this dataset, which is 6.48% and 6.14% higher than the mainstream deep learning models. The model improved 1.89% and 0.82% on Dice and IoU compared to the current SoTA approach (AAU-Net) on the breast segmentation task. For the poorly defined OASBUD, these metrics are 79.62% and 67.52%, respectively, improving 7.39% and 4.83% compared to the mainstream deep learning methods. In addition, the AUC values of the AMS-PAN model on the two datasets are 0.7517 and 0.7763, which are significantly improved compared with the mainstream methods. In practical applications, we believe that AMS-PAN can play a role in aiding diagnosis. However, the segmentation results of the model are still somewhat different from the ground truth. In future work, we need to absorb the advantages of non-deep learning methods to improve the segmentation accuracy of the model further, as well as to

process and experiment with a broader range of datasets to enhance the generalisability and robustness of the model.

CRediT authorship contribution statement

Yuchao Lyu: Conceptualization, Writing – original draft. **Yinghao Xu:** Writing – review & editing. **Xi Jiang:** Software, Investigation. **Jianing Liu:** Data curation. **Xiaoyan Zhao:** Supervision, Project administration. **Xijun Zhu:** Resources.

Funding

This work was supported by the National Natural Science Foundations of China (grant number 62171177 and 82174166) and Doctoral Fund of Qingdao University of science and technology.

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.

Data availability

The authors do not have permission to share data.

References

- [1] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, Ahmedin Jemal, Cancer statistics, 2022. CA: A Cancer J. Clin. 72(1) (2022) 7–33.
- [2] J. Geisel, M. Raghu, R. Hooley, The role of ultrasound in breast cancer screening: the case for and against ultrasound, in: Seminars in Ultrasound, CT and MRI, Elsevier, 2018, pp. 25–34.
- [3] S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, Y. Jin, An improved random forest-based rule extraction method for breast cancer diagnosis, Appl. Soft Comput. 86 (2020), 105941.
- [4] M.S.K. Inan, R. Hasan, F.I. Alam, A hybrid probabilistic ensemble based extreme gradient boosting approach for breast cancer diagnosis, in: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2021, pp. 1029–1035.
- [5] Muhammet Fatih Ak, A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications, in: Healthcare, MDPI, 2020, p. page 111.
- [6] 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.
- [7] Q. Jin, Z. Meng, T.D. Pham, Q.i. Chen, L. Wei, S.u. Ran, Duet: a deformable network for retinal vessel segmentation, Knowl.-Based Syst. 178 (2019) 149–162.
- [8] L.I. Song, K.F. Geoffrey, H.E. Kaijian, Bottleneck feature supervised u-net for pixel-wise liver and tumor segmentation, Exp. Syst. Appl. 145 (2020), 113131.
- [9] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, Automatic brain tumor detection and segmentation using u-net based fully convolutional networks, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2017, pp. 506–517.
- [10] 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.
- [11] M.d. Zongwei Zhou, M.R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (6) (2019) 1856–1867.
- [12] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, Yuyin Zhou, Transunet: transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306, 2021.
- [13] Hancho Li, Pengfei Xiong, Jie An, Lingxue Wang, Pyramid attention network for semantic segmentation, arXiv preprint arXiv:1805.10180, 2018.
- [14] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, Mobilenets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv: 1704.04861, 2017.
- [15] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data Brief 28 (2020), 104863.
- [16] H. Piotrkowska-Wróblewska, K. Dobruch-Sobczak, M. Byra, A. Nowicki, Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions, Med. Phys. 44 (11) (2017) 6105–6109.
- [17] M. Sussman, P. Smereka, S. Osher, A level set approach for computing solutions to incompressible two-phase flow, J. Comput. Phys. 114 (1) (1994) 146–159.
- [18] S.H. Lewis, A. Dong, Detection of breast tumor candidates using marker-controlled watershed segmentation and morphological analysis, in: 2012 IEEE Southwest symposium on image analysis and interpretation, IEEE, 2012, pp. 1–4.
- [19] L. Alvarez, L. Baumela, P. Henriquez, P.M. Neila, Morphological snakes, in: 2010 IEEE Computer Society conference on computer vision and pattern recognition, IEEE, 2010, pp. 2197–2202.
- [20] X. Shen, H.e. Ma, R. Liu, H. Li, J. He, W.u. Xinran, Lesion segmentation in breast ultrasound images using the optimized marked watershed method, Biomed. Eng. Online 20(1) (2021) 1–23.
- [21] Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam, Rethinking atrous convolution for semantic image segmentation, 2017. arXiv preprint arXiv: 1706.05587.
- [22] 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 J. Biomed. Health Inform. 22 (4) (2017) 1218–1226.
- [23] R. Almajalid, J. Shan, D.u. Yaodong, M. Zhang, Development of a deep-learning-based method for breast ultrasound image segmentation, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 1103–1108.
- [24] Gong-Ping Chen, Lei Li, Yu Dai, Jian-Xun Zhang, Nu-net: an unpretentious nested u-net for breast tumor segmentation, 2022. arXiv preprint arXiv:2209.07193.
- [25] Gongping Chen, Yuming Liu, Yu Dai, Jianxun Zhang, Liang Cui, Xiaotao Yin, Bagnet: bidirectional aware guidance network for malignant breast lesions segmentation, arXiv preprint arXiv:2204.13342, 2022.
- [26] B. Shareef, M. Xian, A. Vakanski, Stan: small tumor-aware network for breast ultrasound image segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1–5.
- [27] X.u. Chunbo, Y. Qi, Y. Wang, M. Lou, J. Pi, Y. Ma, Arf-net: An adaptive receptive field network for breast mass segmentation in whole mammograms and ultrasound images, Biomed. Signal Process. Control 71 (2022), 103178.
- [28] R. Irfan, A.A. Almazroi, H.T. Rauf, R. Damaševičius, E.A. Nasr, A.E. Abdelgawad, Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion, Diagnostics 11(7):1212 (2021).
- [29] Z. Zhuang, N. Li, A.N.J. Raj, V.G.V. Mahesh, S. Qiu, An rdau-net model for lesion segmentation in breast ultrasound images, PLoS One 14 (8) (2019) e0221535.
- [30] 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, Biomed. Signal Process. Control 61 (2020), 102027.
- [31] Y.u. Yan, Y. Liu, W.u. Yiyun, H. Zhang, Y. Zhang, L. Meng, Accurate segmentation of breast tumors using a u-net with hdc model in ultrasound images, Biomed. Signal Process. Control 72 (2022), 103299.
- [32] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, F.u. Cheng-Yang, A.C. Berg, Ssd: single shot multi-box detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.
- [33] Abhinav Shrivastava, Rahul Sukthankar, Jitendra Malik, Abhinav Gupta, Beyond skip connections: top-down modulation for object detection, arXiv preprint arXiv: 1612.06851, 2016.
- [34] L. Itti, C. Koch, E. Niebur, A model of saliencybased visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
- [35] L. Itti, C. Koch, Computational modelling of visual attention, Nat. Rev. Neurosci. 2 (3) (2001) 194–203.
- [36] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, Cbam: convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.
- [37] G. Ren, T. Dai, P. Barmpoutis, T. Stathaki, Salient object detection combining a self-attention module and a feature pyramid network, Electronics 9 (10) (2020) 1702.
- [38] J. Li, H. Huo, C. Li, R. Wang, Q.i. Feng, Attentionfgan: Infrared and visible image fusion using attention-based generative adversarial networks, IEEE Trans. Multimedia 23 (2020) 1383–1396.
- [39] C. Fan, Z. Zeng, L. Xiao, Q.u. Xilong, Gfnet: Automatic segmentation of covid-19 lung infection regions using ct images based on boundary features, Pattern Recogn. 132 (2022), 108963.
- [40] H. Lee, J. Park, J.Y. Hwang, Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (7) (2020) 1344–1353.
- [41] C. Xue, L. Zhu, F.u. Huazhu, H.u. Xiaowei, X. Li, H. Zhang, P.-A. Heng, Global guidance network for breast lesion segmentation in ultrasound images, Med. Image Anal. 70 (2021), 101989.
- [42] Narinder Singh Punn and Sonali Agarwal, Rca-iunet: a residual crossspatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging, Mach. Vis. Appl. 33 (2) (2022) 1–10.
- [43] B. Lei, S. Huang, H. Li, R. Li, C. Bian, Y.-H. Chou, J. Qin, P. Zhou, X. Gong, J.-Z. Cheng, Selfco-attention neural network for anatomy segmentation in whole breast ultrasound, Med. Image Anal. 64 (2020), 101753.
- [44] Gongping Chen, Yu Dai, Jianxun Zhang, Moi Hoon Yap, Aaunet: an adaptive attention u-net for breast lesions segmentation in ultrasound images. arXiv preprint arXiv:2204.12077, 2022.
- [45] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
- [46] A. Chaurasia, E. Culurciello, Linknet: exploiting encoder representations for efficient semantic segmentation, in: 2017 IEEE Visual Communications and Image Processing (VCIP), IEEE, 2017, pp. 1–4.