IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 February 2024, accepted 1 March 2024, date of publication 5 March 2024, date of current version 28 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373551

RESEARCH ARTICLE

DDRA-Net: Dual-Channel Deep Residual Attention UPerNet for Breast Lesions Segmentation in Ultrasound Images

JINGBO SUN[1,2], BAOXI YUAN[1,2], JUAN TIAN[3], ZIQI YUAN[4], YIXUAN WU[1], XUEXUE HE[1], CHUANWEI YANG[1], AND ZEXIANG TAO[1]

[1]School of Electronic Information, Xijing University, Xi'an, Shaanxi 710123, China
[2]Xi'an Key Laboratory of High Precision Industrial Intelligent Vision Measurement Technology, Xijing University, Xi'an, Shaanxi 710123, China
[3]School of Humanities and Education, Xijing University, Xi'an, Shaanxi 710123, China
[4]School of Economics, Minzu University of China, Beijing 100081, China

Corresponding author: Baoxi Yuan (ybxbupt@163.com)

This work was supported in part by the Foundation of Shaanxi Key Laboratory of Integrated and Intelligent Navigation under Grant SKLIIN-20190102; in part by the Natural Science Foundation of Shaanxi Province under Grant 2021JM-537, Grant 2019JQ-936, and Grant 2021GY-341; and in part by the Research Foundation for Talented Scholars of Xijing University under Grant XJ20B01, Grant XJ19B01, and Grant XJ17B06.

ABSTRACT Automated segmentation of breast tumors in breast ultrasound images has been a challenging frontier issue. The morphological diversity, boundary ambiguity, and heterogeneity of malignant tumors in breast lesions constrain the improvement of segmentation accuracy. To address these challenges, we propose an innovative deep learning-based method, namely Dual-Channel Deep Residual Attention UPerNet (DDRA-net), for efficient and accurate segmentation of breast tumor regions. The core of DDRA-net lies in the Dual-Channel Deep Residual Attention module (DDRA), which integrates depth-wise separable convolution and Convolutional Block Attention Module (CBAM). This design aims to enhance the extraction of crucial features within the receptive field to better capture subtle details of breast lesions. Through extensive experimental evaluation, DDRA-net demonstrates remarkable performance on a publicly available breast ultrasound datasets, exhibiting higher segmentation accuracy and stability compared to contemporary mainstream deep learning methods. Importantly, it is worth emphasizing that the flexibility of this method allows easy integration with other network structures to further improve the performance and applicability of breast tumor segmentation. In the segmentation of the Breast Ultrasound Image dataset, our precision, recall, IoU, F1 score, Dice, and Hausdorff Distance achieved the following values: 95.31%, 90.79%, 88.00%, 92.39%, 95.46%, and 3.02, respectively. Compared to the original UPerNet, DDRA-net demonstrated improvements of 2.71%, 4.03%, 4.61%, 4.38%, 3.14%, and 24.5% in these six metrics on the Breast Ultrasound Image dataset.

INDEX TERMS Breast tumors, ultrasound images, deep residual attention, deep learning.

I. INTRODUCTION

Breast cancer is one of the most common and serious types of cancer among women, posing a global health concern. According to the World Health Organization (WHO), breast cancer is reported as the most prevalent disease worldwide, with approximately 626,700 women succumbing to breast cancer-related ailments each year, and over 2 million new cases diagnosed in 2018 [1]. Among the contributing factors to breast cancer, some potential pathogenic factors include age, genetics, obesity, smoking, medications, and contraceptive measures. Due to an incomplete understanding of the causes behind breast cancer, effective prevention of the disease remains elusive. However, early detection of breast

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

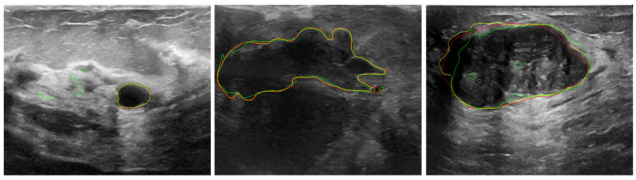

FIGURE 1. Various breast ultrasound images along with segmentation outcomes obtained using both UPerNet and our proposed method. The red curves denote the true boundaries of the lesion regions. The yellow and green curves represent the segmentation results produced by the method and UPerNet, respectively.

cancer at its initial stages can significantly reduce mortality rates and treatment expenses [2]. This approach is more suitable for patients as it eliminates the need for invasive biopsies. Moreover, studies reveal that due to the expanding population base, breast cancer cases are on the rise, generating a substantial volume of ultrasound images daily. Given the diverse appearance of breast lesions and the ambiguity of boundaries, radiologists may misdiagnose breast cancer, and in certain instances, breast tumors may remain undetected, as illustrated in Fig. 1. However, the availability of radiologists for analyzing these medical images is limited, necessitating the establishment of more specialized screening centers and medical experts.

Breast cancer is commonly detected through breast X-ray examinations, ultrasound imaging, and MRI scans. These techniques greatly contribute to early breast cancer detection. Among these methods, breast X-ray exams are frequently employed and prove highly effective in early tumor detection [3]. The diagnosis of early-stage breast cancer increases treatment opportunities and significantly reduces a related mortality rate by 25% [4]. However, about one-third of women who undergo breast X-ray exams receive negative results. Nevertheless, this technique has inherent limitations. Due to the similarity in attenuation coefficients between surrounding tissues and breast tumors, high rates of misdiagnosis occur, raising the risk of radiation exposure for patients. In past studies on breast cancer diagnosis, breast ultrasound examinations have notably improved the detection rate accuracy of breast cancer. As a safer alternative to breast X-ray exams, ultrasound imaging serves as a preliminary method for breast cancer detection. The effectiveness of diagnosing breast cancer through ultrasound images depends on the experience and skills of radiologists, who must interpret speckle noise [5], image complexity, and presence, which can vary among observers. In prior relevant research, high-performance Computer-Aided Diagnosis (CAD) systems for breast tumors have offered accurate and reliable second-opinion diagnoses to differentiate between benign and malignant breast conditions [6], reducing reliance on radiologists [7].

II. RELATED WORK

In the past few decades, traditional computer vision and Convolutional Neural Networks (CNN) have been extensively applied in medical segmentation, and various ultrasound image analysis methods have been used for diagnosing breast cancer, ranging from initial filtering to later, more complex deep learning-based methods. Drukker et al. [8] proposed the use of radial gradient index filtering for automated detection of breast ultrasound lesions. Fateh et al. [9] introduced a model designed for the recognition of multiscript images, comprising components for both language and digit recognition. Within this system, the authors utilized a transfer learning approach to elevate image quality, thereby enhancing the overall performance of recognition. Yap et al. [10] introduced a novel approach for detecting breast cancer lesions using a combination of filtering, multifractal processing, and threshold segmentation. Liu et al. [11] proposed a new active contour model based on level sets for breast ultrasound image segmentation. Morais et al. [12] presented a semi-automatic segmentation technique for three-dimensional transesophageal ultrasound echocardiograms. Huang et al. [13] tested a new approach for breast cancer segmentation through semantic classification and patch merging. While 'their method performed well in segmentation results, its performance was lower when dealing with larger tumors due to the limitations of the simple linear iterative clustering approach. Shan et al. [14] introduced a lesion detection method that simultaneously considers texture and spatial features, detecting and locating breast lesions through ultrasound imaging. Felzenszwalb et al. [15] proposed an effective object detection system based on a multiscale deformable part model mixture, which is one of the efficient methods in object detection. Pons et al. [16] evaluated this method and demonstrated its feasibility. Despite the widespread application and success of deep learning methods in many medical imaging and other domains, limitations still exist in achieving accurate segmentation of breast cancer in ultrasound images. For example, traditional deep learning architectures are composed of simple convolutional filters and can only leverage local information. However, given that most shape priors require manual acquisition, further research is still needed to achieve automated segmentation of breast lesions in ultrasound images.

In recent years, a plethora of studies have adopted deep learning methods for segmenting breast ultrasound images to identify and locate breast lesions. Deep learning methods such as Fully Convolutional Networks (FCN), U-Net, and Unified Perceptual Parsing Networks(UPerNet) [17], [18], [19] have made significant strides in the field of medical image segmentation, attributed to their exceptional nonlinear learning capabilities. Shi et al. [20] propose a lightweight multiscale feature fusion network (LMFFNet), consisting of Segmentation-Extraction-Merge Bottleneck (SEM-B) blocks, Feature Fusion Module (FFM), and Multiscale Attention Decoder (MAD). This innovative design efficiently extracts features with limited parameters, intricately fuses multiscale semantic features, and successfully restores rich details through a unique attention mechanism, significantly improving image segmentation performance. Lin et al.)) [21] introduced a deep dual attention network (D2ANet)

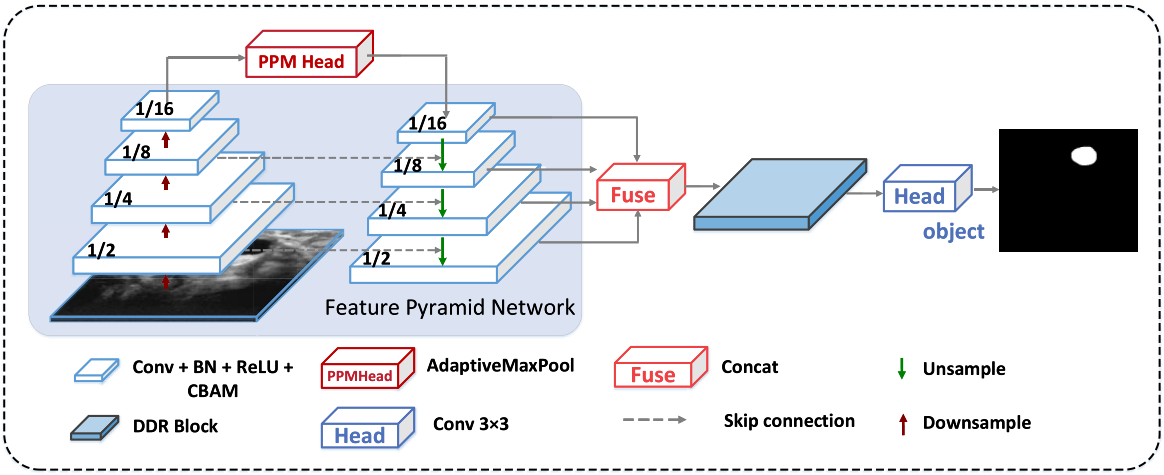

FIGURE 2. Explanation of network architecture.

for COVID-19 diagnosis using chest CT images, skillfully integrating dual attention modules(DAM) and multi-scale feature extractors to automatically detect lesion areas and extract discriminative radiological features, achieving high diagnostic accuracy. Woo et al. improved the performance of the U-Net model for breast cancer segmentation by applying contrast enhancement and speckle reduction preprocessing techniques to breast ultrasound images [22]. Howard et al. introduced a Channel Attention Module with Multi- Scale Grid Average Pooling (MSGRAP) [23], which enhances semantic segmentation for breast cancer diagnosis in ultrasound imaging. Unlike channel attention mechanisms, spatial attention mechanisms prioritize processing in regions of interest, extracting image information with more target-specific features. Xie et al. extended the receptive field of breast cancer segmentation models using dilated convolutions in deeper network layers [24]. However, Jiang et al.'s study pointed out that convolutional operations in Convolutional Neural Networks (CNNs) often focus only on local areas, limiting their ability to capture long-range dependencies in input breast ultrasound images. This limitation hampers the accuracy of CNNs in breast lesion segmentation tasks, resulting in decreased segmentation precision [25]. Dong et al. designed a set of mixed dilated convolutions applied to D2U-Net to address challenges posed by low signal-to-noise ratio, significant artifacts, and variations in breast tumor shape and size [26]. Al-Dhabyani et al. proposed an improved Attention U-Net model by combining image pyramids and attention mechanisms to capture context features at different levels for breast cancer segmentation [27]. Zhao et al. introduced a Convolutional Neural Network called Selective Kernel (SKNet), which was applied to the U-Net model. SKNet aims to adjust the network's receptive field through the introduction of attention mechanisms, combining feature maps extracted through dilated and traditional convolutions [28]. Jiang et al. introduced a Global Guided Network integrating Channel Attention Modules, Spatial Attention Modules, and Boundary Detection Modules for breast cancer lesion segmentation [25]. While dilated convolutions and atention mechanisms can enhance the performance of segmentation networks, they still have some limitations. One limitation is the fixed receptive field size, preventing the network from fully capturing the global contextual information of breast cancer lesions. Another one is the single attention operation, which cannot adapt to the diversity of different lesion regions. To overcome the complexities of ultrasound images, Badrinarayanan et al. [29] introduce an Adaptive Attention U-net (AAU-net) for automated and stable breast lesion segmentation. The approach incorporates a Hybrid Adaptive Attention Module (HAAM) to replace conventional convolution operations, encompassing both channel-wise and spatial self-attention blocks for effective feature capture across diverse receptive fields.

To address the aforementioned challenges, we have devised a novel Dual-Channel Deep Residual Attention (D-DRA) module, which extracts features through distinct branches and employs depth-wise separable convolutions to reduce the model's parameter count, followed by dimension concatenation of the two branches. Furthermore, we also introduce the Convolutional Block Attention Module (CBAM), injecting attention maps independently into feature maps' channel and spatial dimensions. The attention maps are then multiplied with the feature maps, enabling adaptive feature refinement on input breast cancer feature maps and enhancing CNN performance in ultrasound image segmentation [22]. Leveraging the DDRA module, we propose an improved segmentation network based on UPerNet, achieving precise segmentation of breast tumors. The significance of this research lies in the crucial role of achieving accurate breast cancer segmentation in ultrasound images for treatment planning and determining lymph node metastasis. The method introduced here exhibits innovation in its algorithmic approach, adapting well to the complexity of breast cancer feature maps and providing accurate and efficient

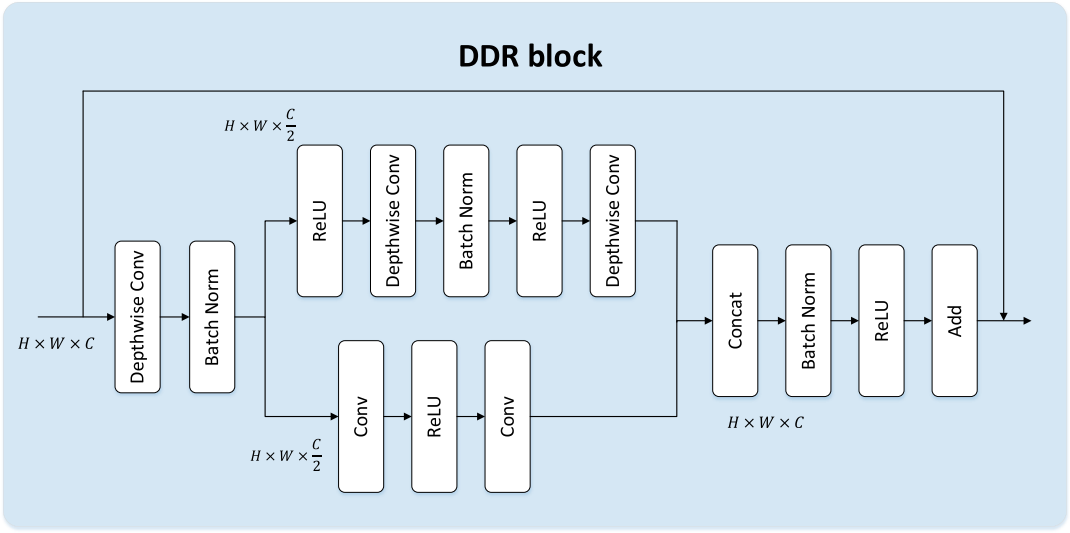

FIGURE 3. Illustration of the structure of the DDR block.

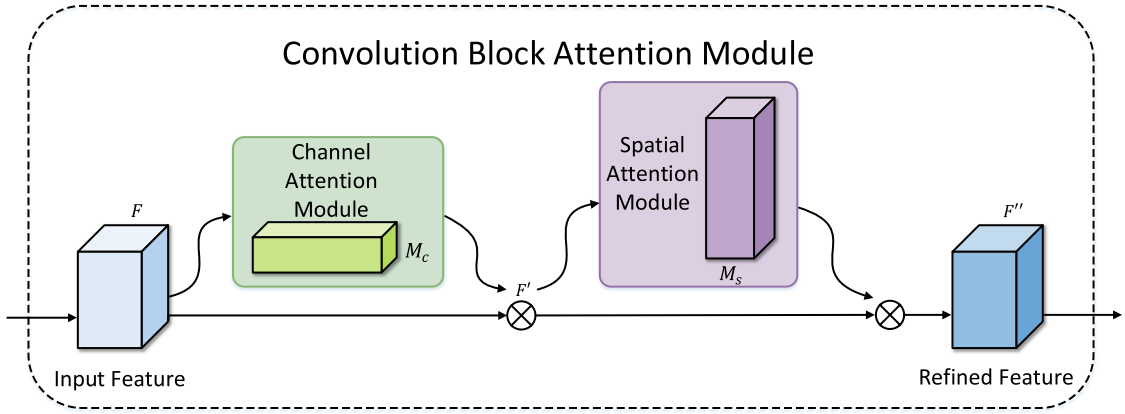

FIGURE 4. Convolution block attention module.

segmentation outcomes. It effectively annotates and localizes lesion regions in breast cancer images, offering vital information to physicians thereby supporting more accurate diagnosis and treatment decisions.

This research holds practical significance in enhancing treatment outcomes and survival rates for breast cancer patients. Extensive experiments demonstrate the outstanding performance and robustness of the method in breast cancer segmentation tasks. Through thorough experimental validation, the DDRA-net showcases substantial and consistent improvements in breast lesion segmentation tasks. The key contributions of our research include:

(1)We propose a novel DDRA module that enables dynamic allocation of attention based on the features of breast lesions. This module accurately focuses on crucial regions, enhancing segmentation precision.

(2)We integrate the Dual-Channel Deep Residual Attention module into the UPerNet network, harnessing its powerful feature extraction and contextual understanding capabilities. This combination allows our model to better comprehend and segment breast lesion areas.

(3)We conduct extensive experimental evaluations, utilizing breast lesion datasets and various evaluation metrics. Consistently, experimental results demonstrate that the D-DRA-net significantly enhances accuracy and stability in breast lesion segmentation tasks, outperforming state-of-the-art medical image segmentation methods.

In the remaining sections of this paper, we will provide a comprehensive overview of our segmentation network and its individual components in Section III. Section IV will delve into the details of the publicly available ultrasound dataset, experimental setup, and evaluation metrics employed. The experimental results are presented in Section V. Finally, we offer a detailed discussion in Section VI and draw conclusions in Section VII.

III. METHODS

Based on the comprehensive findings presented in Table 3, UPerNet showcased exceptional performance in breast tumor segmentation. This compelling superiority led us to designate UPerNet as the fundamental network for our subsequent detailed analyses. Fig. 2 illustrates the network architecture of

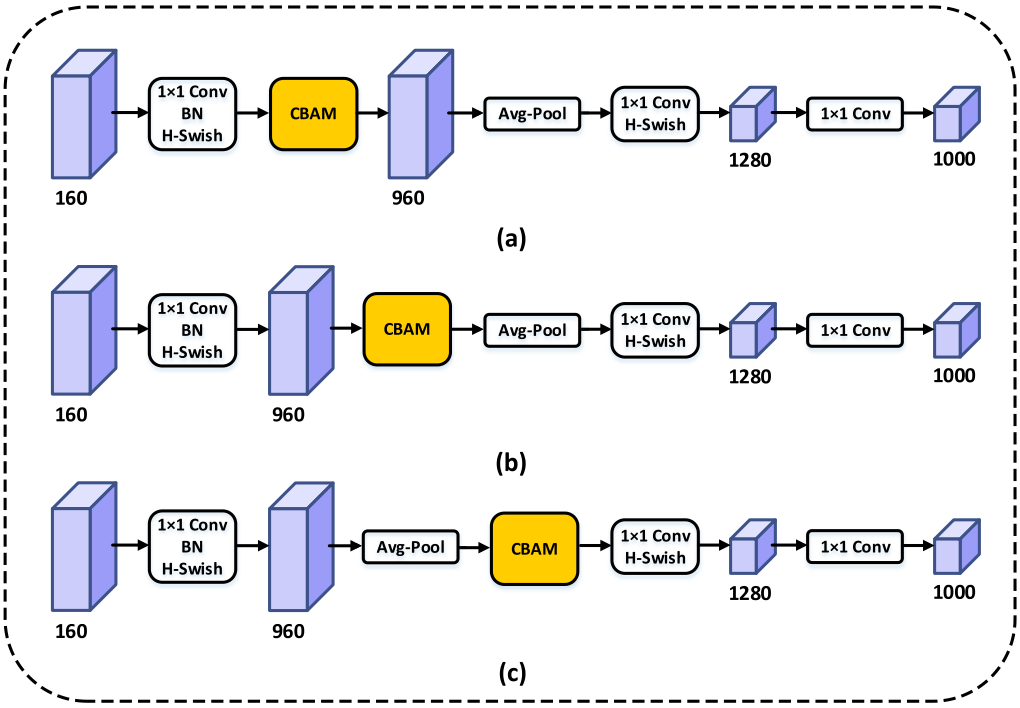

FIGURE 5. CBAM embeded position design.

our developed Dual-Channel Deep Residual Attention UPerNet (DDRA-net) for breast cancer segmentation. The core architecture of DDRA-net bears resemblance to UPerNet [23], including four down-sampling, four up-sampling, and four skip-connection stages. However, a novel Dual-Channel Depthwise Separable Residual module (DDR) is introduced in the feature fusion stage to better adapt to breast lesion segmentation tasks. Additionally, the architecture incurporates the Channel and Spatial Attention Module (CBAM) to automatically extract essential features from the images.

A. ARCHITECTURE OF THE PROPOSED NETWORK

As shown in Fig. 2, the UPerNet architecture consists of two main components: the encoder and the decoder. The encoder in our network is based on the MobileNetV3 architecture [31]. Within the decoder, a symmetrical encoder architecture is established, allowing our network to improve segmentation results by introducing global contextual information through up-sampling and skip connections. This involves multi-scale feature map fusion, effectively utilizing mechanisms such as shared parameters and weights. This design enables the network to maintain high segmentation accuracy while also possessing notable computational and parameter efficiency. In contrast to current popular transformer networks, the UPerNet demonstrates higher computational efficiency and delivers enhanced performance, making it highly practical, particularly in scenarios with limited computational resources. Moreover, the UPerNet's overall structure is based on the segmentation network utilizing convolutional neural networks (such as ResNet), integrating numerous advanced network design techniques to further enhance overall performance. We apply the proposed DDR block and CBAM block to this architecture and conduct ablation experiments using the Breast Ultrasound Image dataset. Table 1 presents the experimental results, demonstrating that the architecture outperforms the UPerNet architecture in semantic segmentation.

B. DDR BLOCK

In the quest for improved accuracy, the strategy of augmenting cardinality surpasses the effectiveness of expanding network depth or width. This conceptual innovation was originaly introduced in the influential ResNeXt framework [24], [25]. Drawing inspiration from this insight, the current study introduces a dual-channel structure in the UPerNet block. In one channel, it adheres to the original modular design while the other channel embraces the design of two consecutive convolutions from the Wide Residual Network. In these dual channels, each with dimensions set to half of the main channel, the study adopts depth-wise separable convolution as a substitute for conventional convolution, effectively reducing the model's parameter count. Ultimately, the model tactically averts overfitting and enhances overall performance by seamlessly integrating the output dimensions of two channels. This is achieved by selectively discarding information directed to the main branch via operations such as Batch Normalization. Through the strategic integration of the initial output with the final layer's output, the model's depth is notably enhanced, elevating both its expressiveness and overall performance. The incorporation of residual connections not only augments depth but also streamlines the model's adeptness in efficiently capturing crucial features. This is achieved by facilitating

direct connections across layers, culminating in a further refinement of the model's comprehensive performance.

Specifically, the input feature map to the DDR block undergoes processing through depthwise separable convolution and BatchNorm. The feature map captured by depthwise separable convolution and BatchNorm can be represented as follows:

$$F^7 = B(W_{7\times7} \times F_{input}) \tag{1}$$

$F_{input} \in \mathbb{R}^{c\times h\times w}$ represents the input feature map, $W_{7\times7}$ denotes the 7×7 convolutional matrix, $B(\cdot)$ represents batch normalization. Subsequently, the feature map is integrated into each branch. To extract valuable target features from the feature map, we initially design two convolutional modules to extract features, guiding the network to learn more robust feature representations. The obtained features can be represented as:

$$F^L = W_{fc}(\sigma_r(W_{fc} \times F^7)) \tag{2}$$

Firstly, we take the processed combined feature map $F^7 \in \mathbb{R}^{\frac{c}{2}\times h\times w}$ and input it into the left branch, where it is compressed into a new feature map $F^L \in \mathbb{R}^{\frac{c}{2}\times h\times w}$. W_{fc} and $\sigma_r(\cdot)$ representing the fully connected layer matrix and the ReLU activation operation, respectively. Moreover, in this study, we employ a depthwise separable convolution in place of conventional convolution to reduce the model's parameter count. Additionally, ReLU activation operations and BatchNorm are utilized for further processing,

$$F^R = W_{3\times3}(\sigma_r(B(W_{3\times3} \times \sigma_r(F^7)))) \tag{3}$$

It's worth noting that these two channel maps aid us in extracting more representative feature maps from different scales of receptive fields. Finally, the feature maps F1 and F2 are concatenated, and subsequently added to our input feature map. Ultimately, the resulting output feature map can be represented as:

$$F^D = \sigma_r(B(F^L \parallel F^R)) \oplus F_{input} \tag{4}$$

where $\parallel$ and $\oplus$ represent channel-wise concatenation and element-wise addition, respectively. Subsequently, $F^D \in \mathbb{R}^{c\times h\times w}$ serves as the input for the next stage and also serves as the output of the entire DDR block. The architecture of the DDR block is depicted in Fig. 3.

C. THE CHANNEL AND SPATIAL ATTENTION MODULE (CBAM)

In order to capture useful target features from various receptive fields, we introduce the CBAM block to replace the original feature fusion layer. This allows the network to dynamically guide itself to learn more meaningful feature representations. CBAM, proposed by Woo et al. in 2018, is a feedforward Convolutional Block Attention Module that infers attention maps sequentially along both channel and spatial dimensions. Subsequently, the attention maps are multiplied with the input feature map for adaptive feature refinement. Additionally, CBAM can be seamlessly integrated into any CNN architecture for end-to-end training. Fig. 4 illustrates the structure of CBAM, which primarily comprises channel attention and spatial attention modules.

As depicted in Fig. 4, the feature map first undergoes the channel attention module, which generates corresponding channel attention maps by considering relationships among different feature channels. Subsequently, the input feature map is element-wise multiplied with the channel attention maps, achieving adaptive weighting of different channel features. The output feature map is then fed into the spatial attention module. The spatial attention module generates spatial attention maps by considering relationships between different spatial positions in the feature map. Following this, the output from the channel attention module is element-wise multiplied with the spatial attention maps, resulting in the final output feature map of the CBAM module. The mathematical expressions for these operations are represented by (1) and (2), where $F(C \times H \times W)$ denotes the input feature map, and $M_cF(C \times 1 \times 1)$ represents the one-dimensional channel attention map, $M_S(F')(1 \times H \times W)$ for the two-dimensional spatial attention map. Here, $\otimes$ signifies element-wise multiplication, while $F'(\mathrm{C} \times \mathrm{H} \times \mathrm{W})$ is the output after the channel attention module, and represents the final output of the CBAM block.

The paper delves into the embedding strategy of CBAM in the Efficient Last Stage of the backbone network MobileNetV3 and explores three different structures [25], [26], as illustrated in Fig. 5: (a) applying CBAM before the convolutional layers in the efficient last stage; (b) introducing CBAM after the convolutional layers in the efficient last stage; (c) applying CBAM to the entire network after adaptive average pooling. Through experimental validation in Section IV-A, we observed that the effect of (b) was significantly more pronounced. Consequently, our model adopts CBAM embedded at the position designed by (b).

$$F' = M_C(F) \otimes F \tag{5}$$

$$F'' = M_S(F') \otimes F' \tag{6}$$

IV. EXPERIMENTS

A. DATA SET

To assess the performance of the proposed network, we employed a widely used breast cancer ultrasound dataset from previous studies. The Breast Ultrasound Image dataset (BUSI) [27] was curated by Dhabyani et al, comprising a total of 780 images acquired using the LOGIQ E9 and LOGIQ E9 Agile ultrasound systems at Baya Hospital. These images were obtained from 600 distinct female subjects, with an average size of 500×500 pixels per image. The dataset is categorized into three classes: normal, benign, and malignant. Specifically, there are 487 benign images, 210 malignant images, and 133 normal images. The dataset is divided into training, testing, and validation sets with a distribution ratio of 7:2:1. Sample data instances are depicted in Fig. 6.

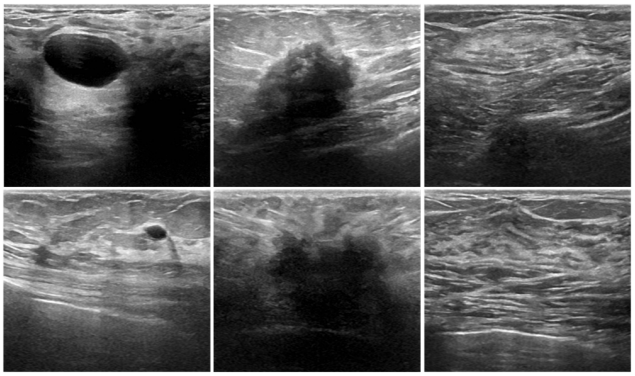

FIGURE 6. Representative images from the breast ultrasound image dataset.

B. EXPERIMENTAL SETTINGS

To evaluate the performance of the proposed network, we trained and compared several commonly used semantic segmentation models, including UPernet [23], FCN [17], PSPNet [28], SegNet [29], U-Net [18], U-Net++ [30], UNet3+ [31], and BASNet [32]. Training was accomplished by using the same parameters and configuration. The size of training images was fixed at 512 × 512, and the batch size was set to 4. We utilized the cross-entropy loss function and the AdamW optimizer for model training. In the initial training phase, a warm-up of 5 epochs was performed, gradually updating the learning rate for each iteration step using linear interpolation. After the warm-up, a polynomial decay learning rate schedule was adopted, with an initial learning rate of 0.001 [33]. To ensure a fair comparison of different models' performance, we refrained from using transfer learning and trained them from scratch. Each model was trained for 300 epochs.

Specifically, in addition to the network model, the choice of loss function plays a crucial role in achieving the desired model performance. The following loss functions are minimized:

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i \cdot \log(\sigma(p_{ic})) + (1 - y_{ic}) \cdot log(1 - \sigma(p_i))) \tag{7}$$

where $\sigma(\cdot)$ represents the Sigmoid function, p_{ic} is the probability of sample i belonging to class c, and y_{ic} is the indicator variable that is 1 if the class of sample i is the same as class c, and 0 otherwise.

All experiments were conducted on a high-performance deep learning server with the following specifications: an Intel Xeon Silver 4210 CPU with a clock frequency of 2.20GHz, an NVIDIA GeForce RTX 2080 Ti graphics processing unit with 11GB of memory, and 128GB of RAM. The deep learning framework used was Python 3.8.10, Cuda 10.2, torch 1.8.1, and torchvision 0.9.1. The operating system was Windows 10.

C. EVALUATION METRICS

To quantitatively evaluate the segmentation performance of breast lesions, we employed eight commonly used segmentation metrics. These metrics include precision, recall, intersection over union (IoU), Dice, F1 score and Hausdorff distance (HD), model parameters and Frames Per Second(FPS). Precision is a fundamental metric widely used in image detection and segmentation tasks; the recall rate is the percentage of the model correctly predicted as a positive sample to the total number of positive samples; IoU is a metric that evaluates the accuracy of segmentation by calculating the ratio of the intersection to the union of the true breast cancer lesion region and the predicted lesion region; Dice is a metric function used to evaluate the similarity between two samples, it operates within a range of 0 to 1, where larger values indicate a greater degree of similarity; F1 score is the harmonic mean of precision and recall; We use true positive (TP), false positive (FP) and false negative (FN) to calculate these metrics:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

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

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{11}$$

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{12}$$

$$HD = \max(h(pred, gt), h(gt, pred)) \tag{13}$$

where $h(A, B) = max_{a \in A}\{min_{b \in B} \parallel a - b \parallel\}$ and $\parallel \cdot \parallel$ denotes the Euclidean distance, measuring the spatial separation between two pixels. *pred* represents the predicted segmentation results, while *gt* designates the ground-truth masks.

V. RESULTS

In this section, we commenced by conducting module ablation experiments to assess the segmentation performance of the individual components of our network. Subsequently, we compared the network with state-of-the-art deep learning segmentation methods. Lastly, we conducted an analysis of the network's robustness.

A. ABLATION STUDY

To assess the performance of different network components, we conducted ablation experiments on BUSI. In these experiments, we utilized UPerNet as the baseline network and evaluated its performance on BUSI. Table 1 presents the performance of different components on BUSI. Method A, repressenting Baseline UPerNet, reveals impressive performance metrics with precision at 92.60%, recall at 86.76%, IoU measuring 83.39%, F1 score achieving 88.01%, Dice coefficient at 92.32%, and HD recorded at 4.00mm. The introduction of the CBAM module enhanced UPerNet in

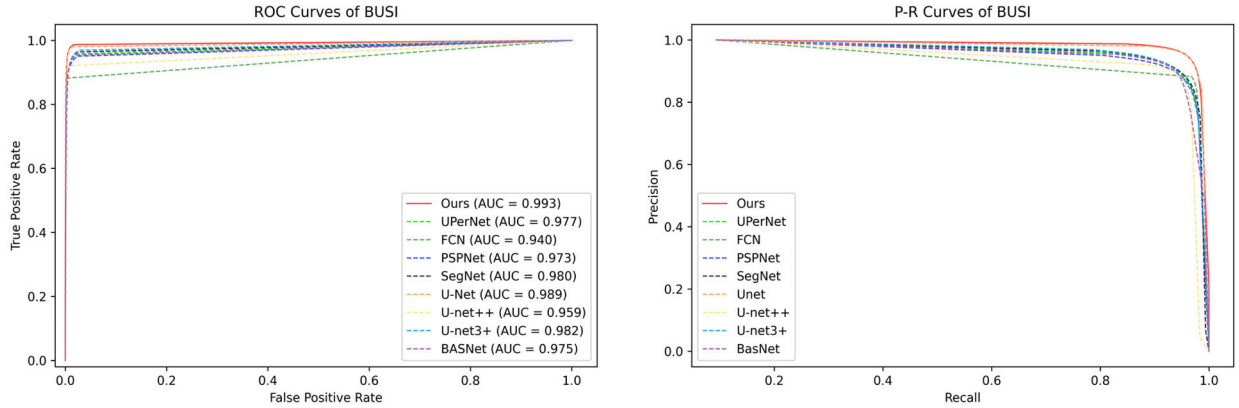

FIGURE 7. P-R and ROC curves of different segmentation methods on BUSI.

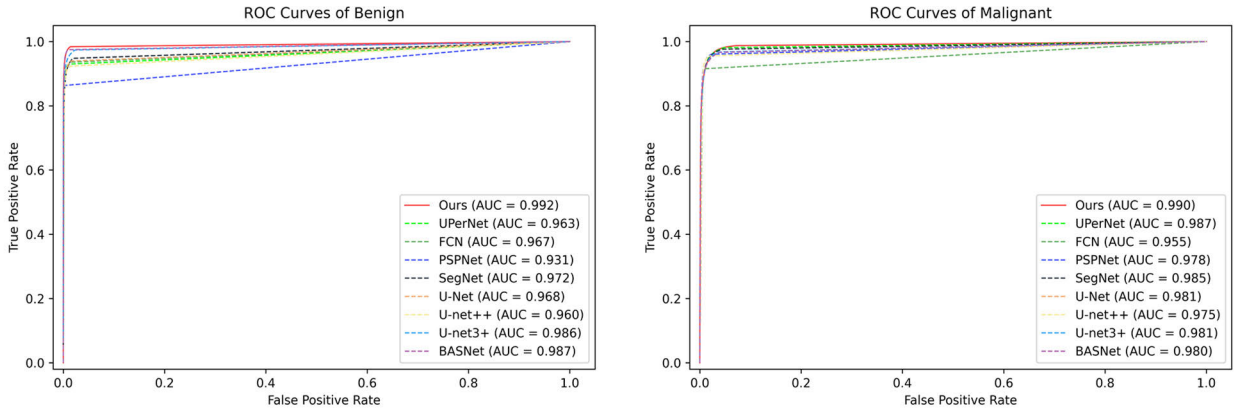

FIGURE 8. ROC curves of different segmentation methods on benign and malignant breast lesions.

Method B, resulting in improvements across six evaluation indices: 0.33%, 1.00%, 1.13%, 1.18%, 0.46%, and 2.75%, respectively. It is noteworthy that these optimizations were achieved without an increase in computational complexity. On the other hand, method C introduces the DDR module on the foundation of benchmark method A. Despite a slight increase in the model's parameter count by 0.01M, there is a corresponding decrease of 13 FPS due to the introduction of additional calculations. Method D incorporates the CBAM module into Method C, thereby introducing the proposed DDRA-NET in this paper. In comparison with Baseline UPerNet, despite a successful reduction of 0.02M in model parameters, DDRA-NET exhibits significant enhancements across the first six indices: 2.71%, 4.03%, 4.61%, 4.38%, 3.14%, and 24.5%, respectively. Although there is a decrease of 12 FPS, this performance adjustment remains more than sufficient to fulfill the practical requirements of breast tumor segmentation applications. The results of the ablation experiments demonstrate the positive impact of the designed network components in enhancing network performance. Particularly, the DDRA module exhibited the best segmentation results on BUSI. This indicates that through the integration of the DDRA module and spatial self-attention mechanism, the network is better able to learn and process BUSI images, leading to improved segmentation robustness.

B. CBAM EMBEDDING POSITION EXPERIMENT

Table 2 elucidates the nuanced experimental results pertaining to the three CBAM embedding strategies delineated in Section 2 of this manuscript. Upon careful examination, UperNet, harnessing MobileNetV3 as its backbone network, manifests robust proficiency in segmentation tasks. Of particular significance is the palpable performance enhancement achieved through the embedding of CBAM in scheme (b). Accordingly, we opt to deploy CBAM after the convolutional layers in the Efficient Last Stage for optimal model performance.

C. COMPARISON WITH STATE-OF-THE-ART METHODS

To validate the robustness and effectiveness of the proposed method, we compared it with the state-of-the-art deep learning methods in medical image segmentation tasks. The comparative methods include UPerNet, FCN, PSPNet, SegNet, U-Net, UNet++, UNet3+ and BASNet. The quantitative evaluation results of different segmentation methods are shown in Table 3. As seen from Table 3, the method achieved its purpose in all six evaluation metrics. The values of the six

TABLE 1. The segmentation results of different network components on BUSI. The best results are highlighted in bold text.

	Precision(%)	Recall(%)	IoU(%)	F1 score(%)	Dice(%)	HD(mm)	Params(M)	FPS
Baseline UPerNet(A)	92.60	86.76	83.39	88.01	92.32	4.00	5.48	**62**
UPerNet with CBAM(B)	92.99	87.76	84.52	89.19	92.78	3.89	5.48	52
UPerNet with DDR block(C)	94.08	89.54	85.73	90.92	94.66	3.25	5.49	49
UPerNet with DDRA(D)	**95.31**	**90.79**	**88.00**	**92.39**	**95.46**	**3.02**	**5.46**	50

TABLE 2. The comparative results on the performance of different CBAM embedding positions. The best results are highlighted in bold text.

	Precision(%)	Recall(%)	IoU(%)	F1 score(%)	Dice(%)	HD(mm)	Params (M)	FPS
UPerNet with (c)	91.78	86.19	82.37	87.54	91.95	4.03	5.48	58
UPerNet with (a)	91.86	86.96	83.13	88.17	92.39	4.00	**5.46**	**60**
UPerNet with (b)	**92.99**	**87.76**	**84.52**	**89.19**	**92.78**	**3.89**	5.48	57

TABLE 3. The comparative results of different segmentation networks on BUSI. The best results are highlighted in bold text. * indicates data from the original paper.

Method	UPerNet	Lee et al.*	AAU-Net*	CSwin-PNet*	FCN	PSPNet	SegNet	U-net	U-net++	U-net3+	BASNet	Ours
Precision(%)	92.60	74.59	79.61±1.07	88.61±1.34	90.91	91.84	92.50	92.77	91.04	92.31	93.09	**95.31**
Recall(%)	86.76	80.41	81.10±0.52	88.10±1.61	85.13	85.25	87.00	88.35	86.10	88.15	86.96	**90.79**
IoU(%)	83.39	62.26	-	**78.61±1.23**	80.93	82.15	84.32	85.15	82.03	84.61	83.21	**88.00**
F1 score(%)	88.01	76.58	-	88.33±0.74	86.75	87.04	88.18	89.50	87.33	89.20	88.44	**92.39**
Dice(%)	92.32	-	77.51±0.68	87.25±1.19	91.64	91.83	91.68	93.23	92.12	92.78	91.95	**95.46**
HD(mm)	4.00	-	-	9.42±1.33	4.25	4.24	4.42	3.76	3.79	4.01	4.12	**3.02**
Params(M)	5.48	-	-	-	134.27	51.30	29.48	**3.45**	47.19	26.97	87.06	5.46
FPS	62	-	-	-	58	65	**88**	53	28	52	52	50

TABLE 4. The comparative results of different segmentation networks with DDRA module on BUSI. The best results are highlighted in bold text.

Method	FCN	PSPNet	SegNet	U-net	U-net++	U-net3+	BASNet	UPerNet(Ours)
Precision(%)	91.53	92.57	93.20	93.01	92.49	93.62	94.77	**95.31**
Recall(%)	86.38	86.88	87.71	89.25	88.04	90.83	**91.28**	90.79
IoU(%)	82.45	83.18	85.76	86.78	84.73	86.69	87.66	**88.00**
F1 score(%)	85.33	88.76	89.17	91.02	89.58	90.15	91.86	**92.39**
Dice(%)	89.55	89.47	89.49	89.11	92.51	90.47	88.05	**95.46**
HD(mm)	4.13	3.95	3.66	3.40	3.57	3.89	3.92	**3.02**
Params(M)	134.27	51.30	29.48	**3.45**	47.19	26.97	87.06	5.46
FPS	50	44	**73**	40	24	45	41	50

evaluation metrics for BUSI using our method are 95.31%, 90.79%, 88.00%, 92.39%, 95.46% and 3.02 respectively. The results of our various metrics are higher than BASNet's results by 2.22%, 3.83%, 4.79%, 3.95%, 3.51%, 26.70%, respectively. The results presented above vividly demonstrate the commendable performance of our method in effectively segmenting breast lesions. With a lean parameter count of just 5.46M, our model excels beyond the majority of split networks, promising an efficient footprint for both storage and runtime in practical applications. Furthermore, our proposed model operates at a speed of 50 FPS. While it may be marginally slower than alternative segmentation algorithms, excluding U-Net++, its prioritized segmentation accuracy proves paramount in the context of breast tumor analysis. The consistent 50FPS segmentation speed guarantees the fulfillment of practical application requirements. Among various segmentation algorithms, the FCN model stands out with the highest parameter count, yet its compromised accuracy leads to suboptimal performance. Both PSPNet and FCN exhibit similar proficiency in evaluation metrics. However, PSPNet gains a significant edge in speed, boasting a mere 38.20% of the parameters found in FCN. With a segmentation speed of 65FPS, PSPNet ranks second only to SegNet. SegNet has delivered impressive results in segmentation performance, coupled with an outstanding segmentation speed boasting 88 FPS. In comparison, U-net closely parallels SegNet in terms of evaluation metrics. Furthermore, U-net distinguishes itself with a lightweight design attributed to its straightforward structure and a minimal model parameter count of merely 3.45M. U-net++ stands out as a prevalent semantic segmentation network in the realm of medical image segmentation, distinguished by its incorporation of a deeper encoder-decoder structure and enhanced deep supervision. While the network attains a notable balance between precision and recall, it does exhibit lower IoU performance, coupled with only FPS of 28. UPerNet, U-Net3+, and BASNet consistently demonstrate robust performance among various segmentation methods, emphasizing their compelling

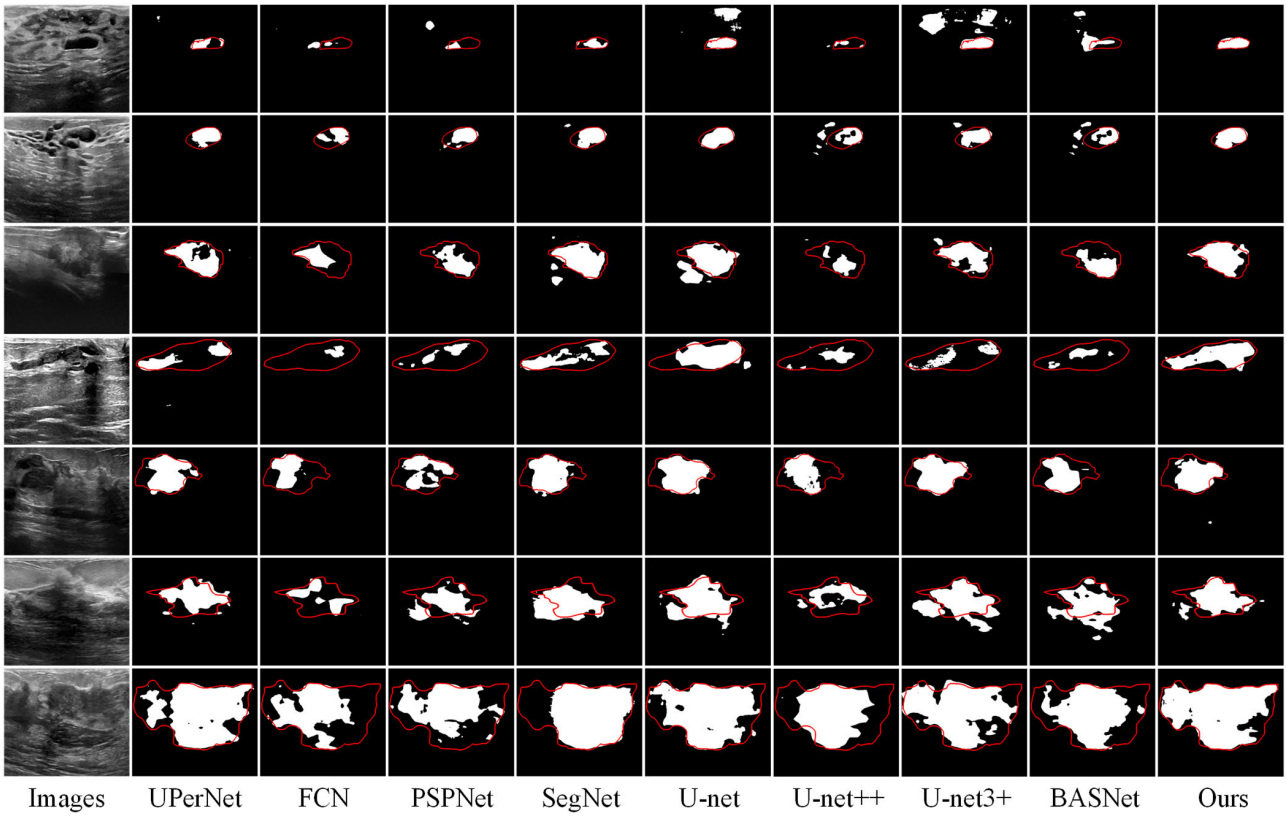

FIGURE 9. Segmentation results of various methods applied to BUSI. The red curves represent the boundaries of breast lesions.

capabilities in the domain of breast lesion segmentation. Moreover, as illustrated in Table 4, we have incorporated the DDRA module into each network for comparative analysis, underscoring the robust efficacy of our innovative D-DRA module.

In Fig. 7, we present the Precision-Recall (P-R) curves and Receiver Operating Characteristic (ROC) curves obtained by using different segmentation methods on BUSI. The P-R curve evaluates model performance by plotting the ratio of true positives to false positives at different confidence levels. The ROC curve measures the correct prediction rate of the method at different confidence levels. We further quantify the comprehensive performance of the ROC curve by calculating the Area Under the Curve (AUC) score. Importantly, in comparison, our method achieved the highest AUC value on BUSI, indicating its superior overall performance compared to other methods. Through the comparison of P-R and ROC curves, the method demonstrates the highest confidence and accuracy in image segmentation tasks on BUSI.

In Fig. 9, we showcase the visual segmentation results of different methods on BUSI. The method exhibits a significant advantage over other approaches in image segmentation tasks. It excels in eliminating interfering factors, preserving tumor morphology integrity, and aligning closely with the ground truth mask. Additionally, the proposed method effectively mitigates the impact of heterogeneous structures on the image segmentation outcome. Through a comprehensive evaluation of experimental results and visual effects, we observe that our method demonstrates outstanding performance in breast lesion segmentation tasks, particularly excelling in omission and error detection.

D. ROBUSTNESS ANALYSIS

To evaluate comprehensively the robustness of the proposed method, we conducted analyses and comparisons in the following two aspects. Firstly, we examined the segmentation performance of different methods specifically focusing on benign and malignant breast tumors. Furthermore, we assessed the impact of normal ultrasound images on the performance of the method.

1) ROBUSTNESS ON BENIGN AND MALIGNANT LESIONS

In this experiment, we conducted a comparison of benign and malignant images within BUSI to assess the robustness of our network in segmenting different lesion types. Malignant lesions typically exhibit characteristics like irregular shapes, unclear boundaries, and uneven intensity distributions. In Table 5, we present the results of different methods on the segmentation of malignant and benign breast lesions. It is evident from the results that our method achieved higher scores in segmenting benign lesions. Notably, it showed a significant improvement in segmentation accuracy compared to other methods. Fig. 9 displays the ROC curves of different segmentation methods for benign and malignant lesions,

TABLE 5. Segmentation results of benign and malignant lesions using different methods. The best results are highlighted in bold text.

	Method	UPernet	FCN	PSPNet	SegNet	U-net	U-net++	U-net3+	BASNet	Ours
Benign	Precision(%)	94.27	93.49	92.79	94.96	94.56	92.75	95.68	96.49	**97.48**
	Recall(%)	87.72	88.69	87.65	88.71	88.01	87.31	89.17	91.36	**91.71**
	IoU(%)	84.87	85.68	84.69	86.61	85.58	83.86	87.33	87.74	**88.65**
	F1 score(%)	89.51	90.14	89.37	90.48	89.82	88.87	92.07	92.99	**93.13**
	Dice(%)	92.18	92.68	92.12	92.26	92.73	91.87	92.58	96.28	**96.34**
	HD(mm)	2.77	3.91	4.55	4.13	3.87	3.64	2.80	2.37	**2.29**
Malignant	Precision(%)	92.39	91.43	91.24	92.74	92.25	91.35	93.65	92.00	**94.18**
	Recall(%)	89.18	87.23	88.34	88.47	87.85	87.32	86.33	87.49	**88.85**
	IoU(%)	85.67	82.54	84.23	85.55	84.85	83.68	84.33	84.18	**86.62**
	F1 score(%)	89.91	88.21	88.88	89.01	88.58	88.42	88.46	87.96	**90.47**
	Dice(%)	93.00	92.58	92.93	93.56	93.83	93.44	93.68	92.48	**95.48**
	HD(mm)	4.70	5.11	5.56	4.96	4.86	4.91	5.13	4.86	**4.35**

TABLE 6. Segmentation results of normal BUSI using different methods. The best results are highlighted in bold text.

Method	UPerNet	FCN	PSPNet	SegNet	U-net	U-net++	U-net3+	BASNet	Ours
Precision(%)	91.50	90.12	91.14	92.77	91.73	92.40	91.08	92.44	**95.02**
Recall(%)	86.69	84.76	85.19	85.22	87.73	87.76	88.30	86.13	**92.32**
IoU(%)	82.27	80.35	81.88	82.76	84.44	83.67	84.42	82.39	**89.23**
F1 score(%)	87.16	86.01	86.45	86.40	88.51	88.67	88.72	87.53	**93.11**
Dice(%)	73.35	72.97	73.14	72.69	73.75	73.31	73.68	73.23	**75.32**
HD(mm)	3.73	4.52	4.48	4.74	3.99	4.00	4.32	4.45	**3.31**

further demonstrating the confidence level of our method. Through the comparison of ROC curves, it clearly showed that the method's performance on benign lesions is convincing and exhibits the most competitive performance in segmenting malignant lesions.

2) COMPARISON ON BUSI WITH NORMAL IMAGES

Breast lesion segmentation in clinical applications generally serves the purpose of lesion assessment, tracking lesion changes, and identifying the distribution and severity of lesions. As a common practice, ultrasound samples are assumed to contain one or more lesions, and then breast lesions are segmented for clinical analysis. In this study, novel comparative experiments were conducted by introducing normal ultrasound images from the BUSI dataset. We validated the method on BUSI containing both normal and lesion images. Table 6 presents the segmentation results of various methods on normal ultrasound images from BUSI. A noticeable impact on the segmentation network's performance was observed with the introduction of normal ultrasound images in BUSI, as compared to Table 3. It can be observed from Table 6 and Fig. 9 that our method achieved the best segmentation performance on BUSI both with and without normal ultrasound images. This indicates that the proposed method mitigates the disturbance caused by surrounding tissues with similar intensity distributions to some extent.

VI. DISCUSSION

In this study, we introduced a novel dual-channel deep residual attention UPerNet (DDRA-net) to overcome the challenges of breast lesion segmentation. To assess the effectiveness of network components, we conducted ablation studies. Through the comparative analysis in Experimental Results Table 1, we distinctly observed that the specific configuration of network components we employed yielded the most superior performance for breast lesion segmentation tasks. The selection of these network components played a crucial role in the success of the method.

Based on our experimental results, we can draw conclusions as follows. Variant networks based on the U-Net architecture (such as U-Net3+) outperform the original U-Net in breast lesion segmentation tasks. This suggests that utilizing skip connections to merge low-level features from the encoding stage with high-level features from the decoding stage is beneficial for accurate breast lesion segmentation. The segmentation outcomes of SegNet indicate that leveraging positional information of features in the U shape network yields superior results compared to the majority of other segmentation methods. From the visual segmentation results depicted in Fig. 9, we can summarize the following three key points. In the segmentation results from the first to fourth rows of Fig. 9, various methods exhibit a certain degree of omission for smaller breast tumors, and in some cases, they might even fail to detect breast lesions. In the segmentation results from the third to sixth rows of Fig. 9, the presence of surrounding tissues with similar intensity distributions and the significant heterogeneity of breast tumors contribute to substantial false negatives and false positives in breast lesion detection. Furthermore, as illustrated in the sixth and seventh rows of Fig. 9, blurry or cascaded BUSI images fail to accurately capture tumor contours. Although our method still demonstrates error detection and missing detection, it achieves significant improvement compared to other methods.

In the robustness analysis experiments, our network demonstrated remarkable generalization capabilities, further highlighting the advantages of the DDRA module. In the

robustness experiments specifically targeting benign and malignant lesion segmentation, the segmentation performance of various methods remained relatively stable. Despite a slight reduction in our method's superiority in these experiments, it still achieved better segmentation performance compared to other methods in the comparative context. This demonstrates the adaptability of our approach to different types of input data, ensuring its consistency in performance.

VII. CONCLUSION

The method proposed in this study has consistently delivered outstanding segmentation results on BUSI. In comparison to existing attention modules, our DDRA module demonstrates superior performance in breast lesion segmentation tasks. Through the incorporation of DDRA module, our network is capable of learning more generalized representations of breast lesions from ultrasound images. This attribute empowers the method to exhibit excellence in handling diverse datasets and scenarios. In the comparative experiments, our DDRA module notably surpasses the performance of state-of-the-art models in breast lesion segmentation. This substantiates the effectiveness and superiority of DDRA module in enhancing network performance.

While our method has demonstrated good performance in breast lesion segmentation, as evident from Fig. 7, there are still some limitations that need to be addressed. For more complex BUSI segmentation, further optimization is required to reduce false positive and false negative rates. Accurately obtaining object contours remains a challenging task. We plan to introduce self-attention mechanisms to further enhance the feature extraction capabilities of the network and are considering the design of a suitable medical image enhancement algorithm.

REFERENCES

[1] F. Bray, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, 2018. Accessed: Jul. 7, 2023. [Online]. Available: https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21492

[2] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognit.*, vol. 39, no. 4, pp. 646–668, Apr. 2006, doi: 10.1016/j.patcog.2005.07.006.

[3] A. Alizad, M. Fatemi, L. E. Wold, and J. F. Greenleaf, "Performance of vibro-acoustography in detecting microcalcifications in excised human breast tissue: A study of 74 tissue samples," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 307–312, Mar. 2004, doi: 10.1109/TMI.2004.824241.

[4] B. Hela, M. Hela, H. Kamel, B. Sana, and M. Najla, "Breast cancer detection: A review on mammograms analysis techniques," in *Proc. 10th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2013, pp. 1–6, doi: 10.1109/SSD.2013.6563999.

[5] P. N. T. Wells and M. Halliwell, "Speckle in ultrasonic imaging," *Ultrasonics*, vol. 19, no. 5, pp. 225–229, Sep. 1981, doi: 10.1016/0041-624x(81)90007-x.

[6] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Breast Cancer Res. Treatment*, vol. 89, no. 2, pp. 179–185, Jan. 2005, doi: 10.1007/s10549-004-2043-z.

[7] K. Drukker, N. P. Gruszauskas, C. A. Sennett, and M. L. Giger, "Breast U.S. computer-aided diagnosis workstation: Performance with a large clinical diagnostic population," *Radiology*, vol. 248, no. 2, pp. 392–397, Aug. 2008, doi: 10.1148/radiol.2482071778.

[8] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, and E. B. Mendelson, "Computerized lesion detection on breast ultrasound," *Med. Phys.*, vol. 29, no. 7, pp. 1438–1446, Jun. 2002, doi: 10.1118/1.1485995.

[9] A. Fateh, M. Fateh, and V. Abolghasemi, "Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning," *Inf. Sci.*, vol. 581, pp. 479–494, Dec. 2021, doi: 10.1016/j.ins.2021.09.051.

[10] M. H. Yap, E. A. Edirisinghe, and H. E. Bez, "A novel algorithm for initial lesion detection in ultrasound breast images," *J. Appl. Clin. Med. Phys.*, vol. 9, no. 4, pp. 181–199, Sep. 2008, doi: 10.1120/jacmp.v9i4.2741.

[11] B. Liu, H. D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Probability density difference-based active contour for ultrasound image segmentation," *Pattern Recognit.*, vol. 43, no. 6, pp. 2028–2042, Jun. 2010, doi: 10.1016/j.patcog.2010.01.002.

[12] P. Morais, S. Queirós, P. D. Meester, W. Budts, J. L. Vilaça, J. M. R. S. Tavares, and J. D'Hooge, "Fast segmentation of the left atrial appendage in 3-D transesophageal echocardiographic images," *IEEE Trans. Ultrason., Ferroelectr, Freq. Control*, vol. 65, no. 12, pp. 2332–2342, Dec. 2018, doi: 10.1109/TUFFC.2018.2872816.

[13] Q. Huang, Y. Huang, Y. Luo, F. Yuan, and X. Li, "Segmentation of breast ultrasound image with semantic classification of superpixels," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101657, doi: 10.1016/j.media.2020.101657.

[14] J. Shan, H. D. Cheng, and Y. Wang, "Completely automated segmentation approach for breast ultrasound images using multiple-domain features," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 262–275, Feb. 2012, doi: 10.1016/j.ultrasmedbio.2011.10.022.

[15] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010, doi: 10.1109/TPAMI.2009.167.

[16] G. Pons, R. Marti, S. Ganau, M. Sentis, and J. Marti, "Feasibility study of lesion detection using deformable part models in breast ultrasound images," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, vol. 7887, J. M. Sanches, L. Micó, and J. S. Cardoso, Eds. Berlin, Germany: Springer, 2013, pp. 269–276, doi: 10.1007/978-3-642-38628-2_32.

[17] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (Lecture Notes in Computer Science), vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[19] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified Perceptual Parsing for Scene Understanding," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jul. 2018, pp. 418–434. Accessed: Jul. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Tete_Xiao_Unified_Perceptual_Parsing_ECCV_2018_paper.html

[20] M. Shi, J. Shen, Q. Yi, J. Weng, Z. Huang, A. Luo, and Y. Zhou, "LMFFNet: A well-balanced lightweight network for fast and accurate semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 6, pp. 3205–3219, Jun. 2023, doi: 10.1109/TNNLS.2022.3176493.

[21] Z. Lin, Z. He, R. Yao, X. Wang, T. Liu, Y. Deng, and S. Xie, "Deep dual attention network for precise diagnosis of COVID-19 from chest CT images," *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 104–114, Jan. 2024, doi: 10.1109/TAI.2022.3225372.

[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jul. 2018, pp. 3–19. Accessed: Jul. 7, 2023. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html

[23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," 2019, *arXiv:1905.02244*.

[24] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995. Accessed: Jul. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/

[25] L. Jiang, B. Yuan, W. Ma, and Y. Wang, "JujubeNet: A high-precision lightweight jujube surface defect classification network with an attention mechanism," *Frontiers Plant Sci.*, vol. 13, Jan. 2023, Art. no. 1108437, doi: 10.3389/fpls.2022.1108437.

[26] H. Dong, K. Song, Q. Wang, Y. Yan, and P. Jiang, "Deep metric learning-based for multi-target few-shot pavement distress classification," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1801–1810, Mar. 2022, doi: 10.1109/TII.2021.3090036.

[27] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863, doi: 10.1016/j.dib.2019.104863.

[28] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239. Accessed: Jul. 26, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.html

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[30] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

[31] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059, doi: 10.1109/ICASSP40776.2020.9053405.

[32] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7471–7481. Accessed: Jul. 25, 2023. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Qin_BASNet_Boundary-Aware_Salient_Object_Detection_CVPR_2019_paper.html

[33] Y. Zhang, B. Yuan, Y. Wang, F. Wang, and J. Guo, "A real-time oil content analysis method of cuttings based on deep learning," *IEEE Access*, vol. 10, pp. 132083–132094, 2022, doi: 10.1109/ACCESS.2022.3229760.

[34] H. Lee, J. Park, and J. Y. Hwang, "Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 7, pp. 1344–1353, Jul. 2020, doi: 10.1109/TUFFC.2020.2972573.

[35] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, "AAU-net: An adaptive attention U-net for breast lesions segmentation in ultrasound images," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1289–1300, May 2023, doi: 10.1109/TMI.2022.3226268.

JINGBO SUN received the B.S. degree in electronic information engineering from the School of Economics and Management, Hebei University of Economics and Business, Hebei, China, in 2020. She is currently pursuing the M.S. degree in next-generation electronic information technology with Xijing University, Xi'an, Shaanxi. Her research interests include deep learning and computer vision.

BAOXI YUAN received the B.S. degree in fire control engineering and the M.S. degree in signal and information processing from Air Force Engineering University, Xi'an, China, in 2000 and 2006, respectively, and the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications, Beijing, China, in 2013. He is currently an Associate Professor with the School of Electronic Information, Xijing University, Xi'an. His current research interests include pattern recognition, machine learning, computer vision, and deep learning.

JUAN TIAN received the B.A. degree from Xi'an Foreign Language University, in 1998, and the M.A. degree in English language and literature (major) from Air Force Engineering University, in 2006, and the Ph.D. degree from PLA Information Engineering University, Zhengzhou, China, in 2017. Currently, she is an Associate Professor with the School of Humanities and Education, Xijing University, Xi'an. Her research interests include study on translation and cross-cultural communications.

ZIQI YUAN is currently pursuing the B.S. degree with the School of Economics, Minzu University of China. Her current research interests include deep learning, digital economics, and enterprise digitization.

YIXUAN WU is currently pursuing the B.S. degree with the School of Electronic Information, Xijing University, Xi'an. Her research interests include deep learning and computer vision.

XUEXUE HE is currently pursuing the B.S. degree with the School of Electronic Information, Xijing University, Xi'an. Her research interests include deep learning and computer vision.

CHUANWEI YANG is currently pursuing the B.S. degree with the School of Electronic Information, Xijing University, Xi'an. His research interests include deep learning and computer vision.

ZEXIANG TAO is currently pursuing the B.S. degree with the School of Electronic Information, Xijing University, Xi'an. His research interests include deep learning and computer vision.