



# Article RMAU-Net: Breast Tumor Segmentation Network Based on Residual Depthwise Separable Convolution and Multiscale Channel Attention Gates

Sheng Yuan, Zhao Qiu \*, Peipei Li and Yuqi Hong

School of Computer Science and Technology, Hainan University, Haikou 570228, China; 21210812000018@hainanu.edu.cn (S.Y.); 2121081200007@hainanu.edu.cn (P.L.); 21220854000052@hainanu.edu.cn (Y.H.) \* Correspondence: qiuzhao@hainanu.edu.cn

Abstract: Breast cancer is one of the most common female diseases, posing a great threat to women's health, and breast ultrasound imaging is a common method for breast cancer diagnosis. In recent years, U-Net and its variants have dominated the medical image segmentation field with their excellent performance. However, the existing U-type segmentation networks have the following problems: (1) the design of the feature extractor is complicated, and the calculation difficulty is increased; (2) the skip connection operation simply combines the features of the encoder and the decoder, without considering both spatial and channel dimensions; (3) during the downsampling phase, the pooling operation results in the loss of feature information. To address the above deficiencies, this paper proposes a breast tumor segmentation network, RMAU-Net, that combines residual depthwise separable convolution and a multi-scale channel attention gate. Specifically, we designed the RDw block, which has a simple structure and a larger sensory field, to overcome the localization problem of convolutional operations. Meanwhile, the MCAG module is designed to correct the low-level features in both spatial and channel dimensions and assist the high-level features to recover the up-sampling and pinpoint non-regular breast tumor features. In addition, this paper used the Patch Merging operation instead of the pooling method to prevent the loss of breast ultrasound image information. Experiments were conducted on two breast ultrasound datasets, Dataset B and BUSI, and the results show that the method in this paper has superior segmentation performance and better generalization.

**Keywords:** breast tumor segmentation; U-Net; residual depthwise separable convolution; multi-scale channel attention gate

# 1. Introduction

Breast cancer is a common malignant tumor that poses a serious health risk to women [1] and is known to be one of the deadliest cancers, causing the highest number of deaths globally [2], making early diagnosis and treatment crucial. Combining breast ultrasonography with computer-aided diagnostic (CAD) systems [3] is one of the most efficient and effective methods of cancer detection due to its painless, cost-effective, non-invasive, and nonradioactive properties [4]. However, accurate breast ultrasound image segmentation remains a challenging problem [5] due to the presence of artifacts and noise in various breast ultrasound images, including high speckle noise [6], low signal-to-noise ratios, and intensity inhomogeneities [7]. In clinical practice, the segmentation task is usually accomplished via manual annotation by a medical professional, which is very time-consuming, and the annotation accuracy varies widely. Therefore, it is of great significance to study the automatic segmentation techniques of breast ultrasound images.

Traditional automatic segmentation algorithms for breast ultrasound typically work directly on the images themselves, using methods such as image processing-based segmentation algorithms [8,9]. While these approaches can achieve effective segmentation



Citation: Yuan, S.; Qiu, Z.; Li, P.; Hong, Y. RMAU-Net: Breast Tumor Segmentation Network Based on Residual Depthwise Separable Convolution and Multiscale Channel Attention Gates. *Appl. Sci.* 2023, *13*, 11362. https://doi.org/10.3390/ app132011362

Academic Editor: Jan Egger

Received: 19 September 2023 Revised: 14 October 2023 Accepted: 15 October 2023 Published: 16 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). results for benign tumor regions with clear and well-defined boundaries, they often fall short when confronted with the challenges posed by irregular tumor regions and blurred boundaries. Similar traditional segmentation methods encompass threshold segmentation [10,11], cluster-based segmentation [12,13], watershed algorithm [14], graph-based segmentation [15,16], and more. These techniques encounter similar challenges when applied to the segmentation of complex ultrasound images of breast tumors.

With the continuous advancement of machine learning methods, artificial intelligence has revolutionized traditional approaches to solving a wide range of problems across various fields, particularly in the realm of biomedicine [17]. The utilization of artificial intelligence algorithms allows for more precise and efficient solutions to challenges like protein structure prediction, gene sequence data mutation recognition, diagnosis of bone diseases [18,19], and organ lesion segmentation. The introduction of convolutional neural networks (CNNs) has further catalyzed progress in biological research. CNNs are versatile in handling diverse types of data. One-dimensional CNNs slide in a single direction and are well-suited for data with only one spatial dimension, such as text or biological sequences. In contrast, two-dimensional CNNs can process data with two spatial dimensions, such as CT images, while 3D CNNs are designed for volumetric data like magnetic resonance imaging scans. Biological data often exhibit a pronounced local structure, and the recognition of these structures or patterns is crucial for analysis. CNNs inherently excel at capturing local features and possess robust feature extraction capabilities. Consequently, the integration of interdisciplinary knowledge and methods, along with the application of CNNs, is poised to drive further breakthroughs and innovations in the field of biological research.

The diagnosis and treatment of breast cancer is an important application in biomedical research. Through automatic segmentation of breast ultrasound images, doctors can evaluate disease conditions more efficiently and give diagnostic opinions. At present, more and more studies are combining CNNs to achieve segmentation of breast ultrasound images. The research shows that the image segmentation method based on deep neural networks has better performance in automatic feature extraction and segmentation accuracy [20,21]. Full Convolutional Networks (FCNs) [22], Semantic Segmentation Networks (SegNet) [23], and U-net networks [24] are commonly used image segmentation methods. U-Net network, in particular, has achieved great success in the field of medical image segmentation. It is an encoder-decoder network architecture which uses standard convolution and continuous downsampling to complete image feature extraction, as well as up-sampling operation and skip connection to complete feature image recovery and finally produce binary segmentation results. It only needs a small amount of medical data to achieve a good segmentation effect and has become a benchmark in the field of medical image segmentation. In recent years, many U-Net-based medical segmentation networks have been proposed, such as U-Net++ [25], Attention U-Net [26], ResU-Net [27], MultiResUNet [28], Unet3+ [29], and UNeXt [30]. However, these U-shaped networks still have the following problems: (1) the design of the feature extractor is complicated, and the calculation difficulty is increased; (2) the skip connection operation simply combines the features of the encoder and the decoder, without considering both spatial and channel dimensions; (3) during the downsampling phase, the pooling operation results in the loss of feature information. In addition, some Transformer-based [31] networks have also been applied to medical image segmentation tasks [32–34]; however, the Transformer model is not suitable for breast ultrasound image segmentation due to its high demand for medical data and large computational consumption due to its attention mechanism.

Therefore, basing itself on the Unet model and aiming at the shortcomings of the above U-shaped networks, this paper proposes a breast tumor segmentation network, RMAU-Net, which combines residual depthwise separable convolutions and multi-scale channel attention gates. This work has the following contributions:

(1) The feature extraction module RDw block was designed, which is simple in structure and can capture more global breast tumor feature information.

- (2) It proposes a multi-scale channel attention gate module to better localize irregular breast tumors by portraying low-level features in both spatial and channel dimensions.
- (3) It uses the Patch Merging operation for downsampling so that breast ultrasound image information will not be lost.
- (4) Experiments were conducted on two breast ultrasound datasets, Dataset B and BUSI, and the results show that the method in this paper has superior segmentation performance and better generalization.

## 2. Related Work

With the continuous development of deep learning, more and more deep learning models have been used to achieve breast ultrasound image segmentation. To solve various problems in breast ultrasound images, researchers have designed many segmentation networks. In order to accurately segment small tumors from breast ultrasound images, Shareef et al. [35] designed a small tumor perception network. This method uses different multi-scale convolutional blocks to integrate the context information of breast tumors with high-resolution features, thus improving the accuracy of small breast tumor segmentation. Lei et al. [36] proposed a boundary-regularized deep convolutional encoder-decoder network to alleviate the challenge of whole breast ultrasound image segmentation. Xue et al. [37] developed a deep CNN with a global guide block and a breast lesion boundary detection module to enhance breast lesion segmentation. Huang et al. [38] proposed a boundary rendering network for breast lesion segmentation through the differentiable boundary selection module and the GCN-based boundary rendering module. However, obtaining accurate boundaries from heavily cascaded or shaded areas remains challenging. Tong et al. [39] used residual convolutional blocks instead of AttU-net convolutional blocks to segment breast tumors, and then Zhuang et al. [40] introduced extended convolutional layers on this basis to capture features under different acceptance fields. Se Woon et al. [41] proposed a multi-stage breast tumor segmentation technique based on ultrasound image classification and segmentation, which firstly classifies the images correctly and then uses RFS-UNet to exclusively segment the images classified as abnormal. We summarize the common problems of these U-shaped networks and design our breast tumor segmentation model from the perspective of reconstructing convolutional blocks and enhancing skip connections.

#### 2.1. Depthwise Separable Convolution

Depthwise separable convolution was proposed in Xception [42], which splits a complete convolution operation into two steps, namely depthwise convolution and pointwise convolution. Each convolution kernel is responsible for all channels, directly mixing information from each channel. Different from conventional convolution operations, each convolutional kernel of depthwise convolution is only responsible for one channel, and each convolutional kernel learns in its own feature space, failing to make effective use of the feature information of different channels in the same spatial position. Therefore, pointwise convolution is required to combine feature maps to generate new feature maps. Pointwise convolution combines previous maps weighted in depth to generate new feature maps to ensure the fusion of feature information. The decomposition of depthwise separable convolution greatly saves the calculation cost and reduces the number of parameters, which newly enlightens the design and reconstruction of convolutional blocks. Dar et al. [43] proposed EfficientU-Net, which used depthwise separable convolution to minimize training parameters and capture relevant texture features to accurately locate tumor boundaries in the breast feature extraction module of this paper. The breast feature extraction module in this paper is also based on depthwise separable convolution.

# 2.2. Skip Connection

The reason why U-Net performs well in medical segmentation tasks is the skip connection operation. Based on the encoder–decoder structure, the skip connection is designed to merge the upper convolutional feature layer with the lower convolutional feature layer with richer semantic information. In the process of network communication, as the network goes deeper and deeper, the receptive field of the corresponding feature map becomes larger and larger, but less and less detailed information is retained. For semantic segmentation tasks, spatial domain information is very important, and only features containing high-level semantic information can generate an accurate segmentation mask. With skip connection, shallow convolution layer features can be introduced. These features have higher resolution and a shallower level and contain relatively rich low-level information, which can help high-level semantic features generate more accurate masks. Attention U-Net [26], MultiResU-Net [28], and other networks extend skip connection and achieve a good segmentation effect. This module is also improved to meet the needs of breast tumor segmentation.

## 3. Method

As shown in Figure 1, the RMAU-Net is structured as a U-shaped encoder–decoder network. In the encoder stage, the input image has dimensions of  $256 \times 256 \times 3$ . After undergoing five RDw operations and four Patch Merging operations, the output size is adjusted to  $16 \times 16 \times 1024$ . In the decoder stage, the process begins with an up-sampling operation, which reduces the number of channels by half while doubling the resolution. Subsequently, the features from the MCAG module are fused with the up-sampled features in the channel dimension, facilitating precise localization of breast lesions. This fusion is repeated four times through consecutive up-sampling layers, ultimately restoring the feature map to its original size. Finally, the output is fed into a softmax layer for binarization, producing the breast lesion segmentation results.



Figure 1. Overview of the proposed RMAU-Net architecture.

#### 3.1. RDw Block

As illustrated in Figure 2a, when conducting feature operations on the same feature graph, depthwise separable convolution entails approximately one-third of the parameters and computational workload compared to conventional convolution. This property allows neural networks employing depthwise separable convolution to become deeper and larger while maintaining the same number of parameters. However, it is not recommended to entirely substitute standard convolution with depthwise separable convolution. This is

because depthwise separable convolution computes features for spatial and channel dimensions independently, which can lead to the loss of certain spatial interaction information.



Figure 2. (a) Depthwise separable convolution. (b) Residual depthwise separable convolution.

To address this limitation, we have made modifications to the depthwise separable convolution, as depicted in Figure 2b.

The residual depthwise separable convolution first deeply convolved the input features. Instead of using a  $3 \times 3$  convolution kernel, we used a  $7 \times 7$  convolution kernel to have a larger receptive field during feature extraction, which can effectively alleviate the locality of the convolution operation and allow us to see more comprehensive features of breast tumors. Then, the features went through the normalization operation and activation function. Here, we no longer directly carried out point-to-convolution but introduced residual operation to connect the features with the original residual input, reduce the spatial interaction information lost via deep convolution, and make the training results proceed in a more stable direction without degradation. Finally, point-to-convolution fusion channel information was carried out. In addition, we used LeakyReLU instead of ReLU to avoid negative inputs causing the neural network to not learn, making the network have a more stable gradient. The whole process is shown in Equations (1) and (2), which represents the LeakyReLU activation function.

$$X_1 = \sigma_1(BN(DepthwiseConv(X)))$$
(1)

$$X_2 = \sigma_1(BN(PConv(X + X_1))) \tag{2}$$

As shown in the extended part of Figure 1, the RDw block was formed by a separable convolutional stack of L residuals depth, and the numbers of L we set were (2,2,3,3) from top to bottom.

## 3.2. Multi-Scale Channel Attention Gate

One of the key reasons for the excellent performance of U-Net in medical segmentation tasks is the skip connection operation, which concatenates low-level features with highlevel features along the channel dimension. This fusion assists in feature recovery by incorporating essential low-level spatial information. However, since channels in the feature maps contain redundant and diverse feature information, there is a desire to emphasize the information most crucial for feature recovery. To address this concern, the SE (Squeeze-and-Excitation) attention mechanism was introduced as a solution [44]. The structure of the SE attention mechanism is depicted in Figure A1 (Appendix A).

The input features  $X \in \mathbb{R}^{H \times W \times C}$  are first turned into features of size  $1 \times 1 \times C$  after global average pooling, and then the number of feature channels is reduced to 1/r through a fully connected layer  $F_{c1}$ . Then, after function ReLU activation, the fully connected layer  $F_{C2}$  is used to recover the feature channels, the attention coefficients are normalized through the sigmoid layer, and finally, the attention weights are multiplied by the input features X to obtain the attention-weighted feature map X*t*. The SE attention calculation process is shown in the following equation:

$$X' = X \otimes Sigmoid(F_{C2}(\text{ReLU}(F_{c1}(AvgPooling(X)))))$$
(3)

Nevertheless, breast tumors exhibit varying sizes and irregular shapes, and simply enhancing attention in the channel dimension may not suffice. To address this, we have designed the MCAG (Multi-scale Channel Attention Gate) module, the structure of which is depicted in Figure 3.





Our approach starts by conducting a multi-scale fusion of breast tumor features across various spatial locations. Subsequently, we obtain multi-scale coefficients  $\lambda$  after normalization and applying an activation function. These coefficients are then multiplied with the initial features, followed by a residual operation. This process enables the initial features to learn the spatial scale information that is most relevant for the task. Finally, we filter the valuable information within the channel dimension, further enhancing the segmentation network's adaptability to irregular breast tumor lesions. The calculation process for the MCAG module is outlined in Equations (4) and (5).

$$\lambda = Sigmoid(PConv(\sigma_1(concat\{BN\{(DConv(X_i)), (PConv(X_i)), (SConv(X_i))\}\})))$$
(4)

$$X_i' = SE(X_i + \lambda X_i) \tag{5}$$

# 3.3. Patch Merging

Image segmentation networks frequently employ average pooling or maximum pooling for feature downsampling. However, these operations inevitably result in the loss of resolution information. If a deep network struggles to learn effective features, it can significantly impair the model's performance. Taking inspiration from the Swin-Transformer [45], we adopted the Patch Merging operation to guarantee that no information was sacrificed during the downsampling process. This approach enabled us to adjust the number of network channels while decreasing resolution, as illustrated in Figure 4.



Figure 4. Schematic diagram of Patch Merging operation.

Since each downsampling operation reduced the resolution by half in both the row and column directions, we adopted a specific approach. We selected every other element at a 2-pixel interval in both directions and combined them to form new patches. These patches were then concatenated to create a unified tensor. Next, each patch was expanded in each channel, resulting in a fourfold increase in the original channel dimensions. Afterward, we applied a normalization operation to the tensor. Finally, we adjusted the channel dimensions back to twice the original size through a fully connected layer. This process ensured that we preserved resolution while simultaneously incorporating spatial information. This enables us to make full use of pixel feature information, especially in grayscale images like breast ultrasound scans.

# 3.4. Loss Function

The loss function we used was a hybrid loss function consisting of Binary Cross Entropy (BCE) loss and Dice loss, both of which are widely used loss functions in image segmentation tasks. Dice loss focuses on global observation while BCE loss focuses on micro-pixel by pixel comparison, and the two complement each other. When the segmentation content is unbalanced, such as a 512 × 512 picture with a 10 × 10 and a 200 × 200 segmentation example, Dice loss tends to learn large samples while ignoring small samples, but BCE still learns small samples, so it is necessary to combine the two for loss calculation.

The weights of BCE loss and Dice loss were 0.5 and 1, which were consistent with the results of breast tumor segmentation networks such as unext. To be fair, we did not adjust parameters separately for their specific weights. It was defined as follows:

$$L = 0.5BCE(\hat{y}, y) + Dice(\hat{y}, y)$$
(6)

#### 4. Experiment and Analysis

## 4.1. Dataset and Preprocessing

In this paper, two widely used public breast ultrasound datasets are used to evaluate the performance of segmentation networks. The first is Dataset B, collected by Yap et al. [46]. The second breast ultrasound dataset used in this paper is BUSI, constructed by Al-Dhabyani et al. [47], with specific information shown in Table 1.

The dataset was split into 80% for training and 20% for validation. During the data preprocessing stage, we resized all images to a  $256 \times 256$  resolution. To address the sample imbalance between benign and malignant tumor data, we applied data augmentation techniques specifically to the tumor samples. This augmentation involved contrast stretching

and flipping operations. You can see the results of the data augmentation for breast tumor ultrasound in the Figure 5.

Table 1. Details of the Dataset B and BUSI.

	Equipment	Benign	Malignant	Total
Dataset B	Siemens ACUSON Sequoia C512	110	53	163
BUSI	LOGIQ E9 and LOGIQ E9 Agile	437	210	647



Figure 5. From left to right, original image, contrast stretch, and flip.

#### 4.2. Experimental Settings

In this experiment, we utilized the Adam optimizer to optimize our network. We set the initial learning rate to 0.0001 and employed a momentum value of 0.9. The batch size was configured to 8, and we ran the training for a total of 300 epochs. Our network was executed on a system running Ubuntu 20.04 with Python version 3.8, PyTorch version 1.11.0, and powered by NVIDIA RTX 3090 GPUs.

We show the training flow of our proposed segmentation network in detail in Algorithm 1, and the algorithm pseudo-code is shown as follows.

Algorithm 1. The detailed training process of the RMAU-Net **Input**: Augmented training sample  $S = \{X_1, X_2, \dots X_n\}$ , where  $X \in \mathbb{R}^{256 \times 256 \times 3}$ 1. Begin

- 2. Randomly initialize the model parameters
- 3. While  $\varepsilon$  have not converged **do**
- 4. **For** epoch = 0, 1, ..., 300 **do**
- 5. The parameters were retrained on the target dataset
- The Adam optimizer was used to update the weights, as expressed by 6.  $\varepsilon_{c+1} = \varepsilon_c - \alpha \times \frac{s_c}{(\sqrt{\hat{r}_c} + \theta)}$ 
  - $\alpha$  (learning rate),  $\hat{s}_c$  (the first corrected bias),  $\hat{r}_c$  (the second corrected bias)
- 7. The BDice loss function was used to update the weights, as expressed by  $G_{\varepsilon} \leftarrow \nabla_{\varepsilon} Loss_{BDice}(S)$ , where "BDice" means BCE Loss and Dice Loss
- 8. Apply the cosine annealing strategy to adjust  $\boldsymbol{\alpha}$
- 9. Continuously update the weighs using the S
- 10.  $\varepsilon \leftarrow \varepsilon - \mu G_{\varepsilon}$
- 11. End for
- 12. End while
- 13. End
- **Output**: the best weight parameter  $\varepsilon_{best}$

# 4.3. Evaluation Indicators

In this paper, five commonly used segmentation metrics are used to evaluate the effectiveness of different methods for the segmentation of breast lesions; they are Dice, IoU, recall, precision, and accuracy. IoU and Dice are the two most important indexes of the image segmentation task. Recall refers to the proportion of all foreground pixels in

the ground truth that are correctly segmented as the foreground by the model, precision refers to the proportion of pixels in the segmentation result that are correctly segmented as foreground, and accuracy refers to the proportion of pixels in the segmentation result that are correctly segmented as foreground or hard ground.

## 4.4. Results and Discussion

To validate the effectiveness of our network, we conducted a comparison with current open-source and widely used medical image segmentation networks, including U-Net, UNet++, SegNet, Attention U-Net, UNeXt, and ResU-Net. All models were deployed and executed locally, with all experimental variables being consistent except for the network architecture.

Figure 6 showcases the outcomes of breast tumor segmentation alongside actual manual annotations. When assessing the segmentation results as a whole, our model demonstrates the capability to accurately locate lesion regions and determine their shape and size. In terms of segmentation intricacies, our model outperforms manual annotations, delivering more detailed and refined results. Notably, even in the case of fuzzy and unclear tumor regions, as highlighted in the red box, our model still achieves precise segmentation.



Figure 6. The segmentation results of ours and ground truth.

The experimental data further validate the efficacy of our model. Results from various models tested on both Dataset B and the BUSI dataset are summarized in Tables 2 and 3. Notably, on the Dataset B dataset, our model exhibits a 1.72% improvement in IoU and a 0.84% improvement in the Dice score. On the BUSI dataset, the improvements are even more pronounced, with a 2.25% increase in IoU and a significant 2.58% improvement in the Dice score. While our model may not have achieved the highest scores in individual metrics, overall, it has achieved a notably high level of segmentation quality. Importantly, our model excels in the two most critical segmentation indicators, Dice and IoU, which underscores its effectiveness in accurately delineating breast tumor lesions.

Tab	le 2.	Comparison	results	with ot	her mode	ls on	Dataset	В.
-----	-------	------------	---------	---------	----------	-------	---------	----

Model	Dice	IoU	Recall	Precision	Accuracy
U-Net [24]	86.28	75.89	87.15	85.75	99.15
U-Net++ [25]	83.84	72.20	85.17	82.79	98.99
SegNet [23]	83.51	71.86	80.33	87.49	99.05
Attention U-Net [26]	84.27	72.86	84.46	84.70	99.05
UNeXt [30]	81.09	68.57	80.09	81.66	98.81

Table 2. Cont.

Model	Dice	IoU	Recall	Precision	Accuracy
ResU-Net [27]	84.37	73.71	85.35	83.98	99.04
Ours	87.12	77.61	86.04	88.55	99.22

Table 3. Comparison results with other models on BUSI.

Model	Dice	IoU	Recall	Precision	Accuracy
U-Net [24]	75.65	62.33	71.97	81.90	95.95
U-Net++ [25]	75.03	61.59	69.50	85.15	96.04
SegNet [23]	77.21	64.72	72.10	86.81	96.33
Attention U-Net [26]	74.65	61.52	68.85	85.60	95.99
UNeXt [30]	72.77	58.59	67.57	82.05	95.67
ResU-Net [27]	75.20	61.72	69.69	84.09	95.95
Ours	79.79	66.97	79.63	84.77	96.43

The visual comparison results for breast tumor ultrasound segmentation can be observed in Figures 7 and 8. When examining these visual results in conjunction with the data presented in the table above, it becomes evident that our model excels in segmenting tumor lesions with irregular and indistinct boundaries. It accurately delineates lesions of varying scales. These findings underscore the effectiveness of our proposed multi-scale channel attention gate.



**Figure 7.** The segmentation results for Dataset B are shown in the figure, from left to right: U-Net, U-Net++, SegNet, Attention U-Net, UNeXt, ResU-Net, ours, ground truth.



**Figure 8.** The segmentation results for BUSI are shown in the figure, from left to right: U-Net, U-Net++, SegNet, Attention U-Net, UNeXt, ResU-Net, ours, ground truth.

# 5. Conclusions

In this study, our paper addresses several issues related to the segmentation of breast tumors using U-shaped networks. Firstly, U-shaped networks often suffer from a complex feature extraction module. Additionally, the skip connection operation fails to consider both spatial and channel dimensions simultaneously, and pooling operations result in the loss of valuable feature information. To tackle these challenges, we introduced a novel breast tumor segmentation network known as RMAU-Net, which combines the power of residual depthwise separable convolution and multi-scale channel attention gates. We have designed the RDw block, which boasts a straightforward structure capable of capturing more comprehensive global characteristics of breast tumors. Simultaneously, the MCAG block is devised to rectify low-level features across both spatial and channel dimensions, aiding in the effective learning of non-regular breast tumor features and facilitating high-level feature recovery during up-sampling. Furthermore, our approach replaces traditional pooling methods with Patch Merging operations to prevent the loss of critical breast ultrasound image information. Our experimental results on two distinct datasets demonstrate that RMAU-Net outperforms existing methods in terms of segmentation accuracy. In the future, we plan to further enhance RMAU-Net to handle more challenging breast lesion images and explore lightweight model designs to achieve more efficient breast tumor segmentation.

**Author Contributions:** Conceptualization: S.Y.; methodology: S.Y.; formal analysis and investigation: S.Y.; writing—original draft preparation: S.Y., Z.Q., and P.L.; writing—review and editing: Z.Q., P.L., and Y.H.; funding acquisition: Z.Q.; resources: Z.Q.; supervision: Z.Q. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Education Department of Hainan Province (Grant No. Hnjg2021ZD-10).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The datasets used in this paper are public datasets. Datasets are available in the relevant studies of the authors mentioned in Section 4.1.

Conflicts of Interest: The authors declare no conflict of interest.

## Appendix A



Figure A1. SE Attention Module.

## References

- Matsumoto, R.A.E.K.; Catani, J.H.; Campoy, M.L.; Oliveira, A.M.; Barros, N.D. Radiological findings of breast involvement in benign and malignant systemic diseases. *Radiol. Bras.* 2018, *51*, 328–333. [CrossRef]
- Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2015. CA Cancer J. Clin. 2015, 65, 5–29. [CrossRef] [PubMed]
- Sahiner, B.; Chan, H.P.; Hadjiiski, L.M.; Helvie, M.A.; Wei, J.; Zhou, C.; Lu, Y. Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: A 3D approach. *Med. Phys.* 2012, 39, 28–39. [CrossRef] [PubMed]
- Cheng, H.D.; Shan, J.; Ju, W.; Guo, Y.; Zhang, L. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognit.* 2010, 43, 299–317.
- 5. Noble, J.A.; Boukerroui, D. Ultrasound image segmentation: A survey. IEEE Trans. Med. Imaging 2006, 25, 987–1010. [PubMed]
- 6. Wells, P.N.T.; Halliwell, M. Speckle in ultrasonic imaging. *Ultrasonics* **1981**, *19*, 225–229. [CrossRef]
- Xiao, G.; Brady, M.; Noble, J.A.; Zhang, Y. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. *IEEE Trans. Med. Imaging* 2002, 21, 48–57. [CrossRef]
- 8. Liu, Y.; Ren, L.; Cao, X.; Tong, Y. Breast tumors recognition based on edge feature extraction using support vector machine. *Biomed. Signal Process. Control* **2020**, *58*, 101825.

- Inoue, K.; Yamanaka, C.; Kawasaki, A.; Koshimizu, K.; Sasaki, T.; Doi, T. Computer aided detection of breast cancer on ultrasound imaging using deep learning. Ultrasound Med. Biol. 2017, 43, S19. [CrossRef]
- Drukker, K.; Giger, M.L.; Horsch, K.; Kupinski, M.A.; Vyborny, C.J.; Mendelson, E.B. Computerized lesion detection on breast ultrasound. *Med. Phys.* 2002, 29, 1438–1446. [CrossRef]
- Horsch, K.; Giger, M.L.; Venta, L.A.; Vyborny, C.J. Automatic segmentation of breast lesions on ultrasound. *Med. Phys.* 2001, 28, 1652–1659. [CrossRef] [PubMed]
- 12. Moon, W.K.; Lo, C.M.; Chen, R.T.; Shen, Y.W.; Chang, J.M.; Huang, C.S.; Chen, J.; Hsu, W.; Chang, R.F. Tumor detection in automated breast ultrasound images using quantitative tissue clustering. *Med. Phys.* **2014**, *41*, 042901. [CrossRef] [PubMed]
- 13. Shan, J.; Cheng, H.D.; Wang, Y. A novel segmentation method for breast ultrasound images based on neutrosophic l-means clustering. *Med. Phys.* 2012, *39*, 5669–5682. [CrossRef] [PubMed]
- Mitra, A.; De, A.; Bhattacharjee, A.K. MRI Skull Bone Lesion Segmentation Using Distance Based Watershed Segmentation. In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Bhubaneswar, Odisha, India, 14–15 November 2014.
- 15. Zhou, Z.; Wu, W.; Wu, S.; Tsui, P.H.; Lin, C.C.; Zhang, L.; Wang, T. Semi-automatic breast ultrasound image segmentation based on mean shift and graph cuts. *Ultrason. Imaging* **2014**, *36*, 256–276. [CrossRef]
- Huang, Q.H.; Lee, S.Y.; Liu, L.Z.; Lu, M.H.; Jin, L.W.; Li, A.H. A robust graph-based segmentation method for breast tumors in ultrasound images. *Ultrasonics* 2012, 52, 266–275. [CrossRef]
- Sohail, A.; Arif, F. Supervised and unsupervised algorithms for bioinformatics and data science. *Prog. Biophys. Mol. Biol.* 2020, 151, 14–22. [CrossRef]
- Sohail, A.; Younas, M.; Bhatti, Y.; Li, Z.; Tunç, S.; Abid, M. Analysis of trabecular bone mechanics using machine learning. *Evol. Bioinform.* 2019, 15, 1176934318825084. [CrossRef]
- 19. Al-Utaibi, K.A.; Idrees, M.; Sohail, A.; Arif, F.; Nutini, A.; Sait, S.M. Artificial intelligence to link environmental endocrine disruptors (EEDs) with bone diseases. *Int. J. Model. Simul. Sci. Comput.* **2022**, *13*, 2250019. [CrossRef]
- Hu, G.; Du, Z. Adaptive kernel-based fuzzy c-means clustering with spatial constraints for image segmentation. Int. J. Pattern Recognit. Artif. Intell. 2019, 33, 1954003. [CrossRef]
- Xu, M.; Huang, K.; Chen, Q.; Qi, X. Mssa-net: Multi-scale self-attention network for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021.
- Mannem, R.; Ca, V.; Ghosh, P.K. A SegNet based image enhancement technique for air-tissue boundary segmentation in real-time magnetic resonance imaging video. In Proceedings of the 2019 National Conference on Communications (NCC), Bangalore, India, 20–23 February 2019.
- Ben-Cohen, A.; Klang, E.; Raskin, S.P.; Soffer, S.; Ben-Haim, S.; Konen, E.; Amitai, M.M.; Greenspan, H. Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. *Eng. Appl. Artif. Intell.* 2019, 78, 186–194. [CrossRef]
- 24. Amiri, M.; Brooks, R.; Rivaz, H. Fine tuning u-net for ultrasound image segmentation: Which layers? In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*; Springer: Cham, Switzerland, 2019; pp. 235–242.
- 25. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018; Springer International Publishing: Cham, Switzerland, 2018.
- 26. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
- Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted res-unet for high-quality retina vessel segmentation. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018.
- 28. Ibtehaz, N.; Sohel Rahman, M.M. Rethinking the U-Net architecture for multimodal biomedical image segmentation. *arXiv* 2019, arXiv:1902.04049. [CrossRef] [PubMed]
- Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020.
- 30. Valanarasu, J.M.J.; Patel, V.M. Unext: Mlp-based rapid medical image segmentation network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30. [CrossRef]
- 32. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
- Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021.

- Wang, W.; Chen, C.; Ding, M.; Yu, H.; Zha, S.; Li, J. Transbts: Multimodal brain tumor segmentation using transformer. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021.
- Shareef, B.; Xian, M.; Vakanski, A. Stan: Small tumor-aware network for breast ultrasound image segmentation. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020.
- Lei, B.; Huang, S.; Li, R.; Bian, C.; Li, H.; Chou, Y.H.; Cheng, J.Z. Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. *Neurocomputing* 2018, 321, 178–186. [CrossRef]
- Xue, C.; Zhu, L.; Fu, H.; Hu, X.; Li, X.; Zhang, H.; Heng, P.A. Global guidance network for breast lesion segmentation in ultrasound images. *Med. Image Anal.* 2021, 70, 101989. [CrossRef] [PubMed]
- Huang, R.; Lin, M.; Dou, H.; Lin, Z.; Ying, Q.; Jia, X.; Xu, W.; Mei, Z.; Yang, X.; Dong, Y.; et al. Boundary-rendering network for breast lesion segmentation in ultrasound images. *Med. Image Anal.* 2022, *80*, 102478. [CrossRef]
- Tong, Y.; Liu, Y.; Zhao, M.; Meng, L.; Zhang, J. Improved U-net MALF model for lesion segmentation in breast ultrasound images. Biomed. Signal Process. Control 2021, 68, 102721. [CrossRef]
- Zhuang, Z.; Li, N.; Joseph Raj, A.N.; Mahesh, V.G.; Qiu, S. An RDAU-NET model for lesion segmentation in breast ultrasound images. *PLoS ONE* 2019, 14, e0221535. [CrossRef]
- Cho, S.W.; Baek, N.R.; Park, K.R. Deep Learning-based Multi-stage segmentation method using ultrasound images for breast cancer diagnosis. J. King Saud Univ. Comput. Inf. Sci. 2022, 34, 10273–10292.
- Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
- 43. Dar, M.F.; Ganivada, A. EfficientU-Net: A Novel Deep Learning Method for Breast Tumor Segmentation and Classification in Ultrasound Images. *Neural Process. Lett.* **2023**, 1–24. [CrossRef]
- Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
- Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
- 46. Yap, M.H.; Goyal, M.; Osman, F.; Martí, R.; Denton, E.; Juette, A.; Zwiggelaar, R. Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine* **2020**, *107*, 101880. [CrossRef] [PubMed]
- Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* 2020, 28, 104863. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.