Segmentation Algorithm for Cancer Regions in Breast Cancer MRI Images Based on the Improved U^2-Net Network

Ye Lin[1], Zongyan Dai[2,*], Qi Jing[3], Rui Shi[4]
{994514145@qq.com[1], 2848474375@qq.com[2], 13701391909@163.com[3], 15611123330@163.com[4]}

Department of computer science and technology, Zhejiang Normal University, Jinhua,321004,China[1]
School of Software, International School of Information Science & Engineering , Dalian,116024,China[2]
Department of computer science and technology, Beijing Institute of Technology, Beijing, 100081, China[3]
School of Control and Computer Engineering, North China Electric Power University, Beijing,102206, China[4]
*corresponding author

Abstract. Around the world, breast cancer is the most common cancer in women disease. In this paper, an enhanced segmentation algorithm based on improved U ^{2} -Net network is proposed, which integrates the convolutional block attention module and dense connection to enhance the efficiency of feature extraction and segmentation of the breast cancer MRI images. Additionally, the model in this paper was rigorously evaluated using a comprehensive dataset of breast cancer MRI images, including both privately labeled and publicly labeled data. Our experimental results and the ablation experiments demonstrate that the integration of CBAM module and dense connection present the significant optimization improvement. Besides, the model in this paper provides a robust framework on the basis of achieving high segmentation accuracy, which can have a larger optimization space in the field of medical image segmentation as well.

Keywords: U^2-Net Network; Attention Mechanism; Dense Connections; MRI Medical Image Segmentation

1 Introduction

1.1 Research background and significance

Breast cancer has the highest proportion of new cancer cases among women globally. It is the most common cancer in women and also the main cause of female deaths. According to data from the World Health Organization (WHO) and the International Agency for Research on Cancer (IARC), every year, approximately more than 2 million women worldwide are diagnosed with breast cancer, and about 500,000 to 600,000 people die from breast cancer. The harm of breast cancer is

CONF-MLA 2024, November 21, Adana, Turkey

DOI 10.4108/eai.21-11-2024.2354611

not only limited to its high incidence and mortality rate but also includes its profound impact on the quality of life of patients. Patients may experience physical pain and discomfort, psychological stress and anxiety, as well as social and economic burdens. Therefore, in clinical practice, accurate segmentation of breast cancer lesion areas and normal human tissues plays a decisive role in the treatment process.

At present, the imaging examination methods for breast cancer mainly include mammography, breast ultrasonography, breast CT scan, and breast MRI.

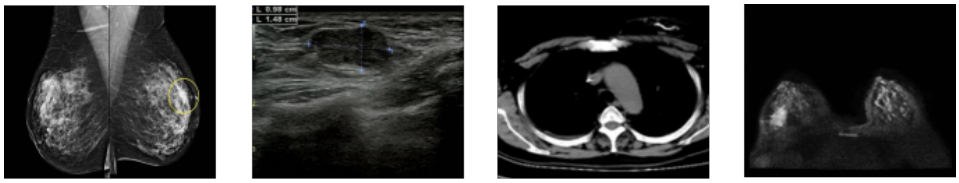


Fig. 1. From the first picture to the fourth picture in turn are Mammography, breast ultrasonography, breast CT scan, and breast MRI

Among them, breast MRI has better soft tissue resolution and spatial resolution. Compared with CT, mammography, and breast ultrasonography, its lesion detection sensitivity, accuracy, positive predictive value, negative predictive value, and specificity are better[1] .With the development and in-depth research of the clinical application of breast MRI, its application value in the diagnosis of breast cancer, breast-conserving treatment, neoadjuvant chemotherapy (NAC), and follow-up monitoring has increased.Among them, accurate segmentation of breast cancer lesion areas can provide better help. Based on the high soft tissue resolution and multi-parameter and multi-sequence imaging capabilities of MRI, MRI can display the information of breast tissue and axillary lymph nodes in three dimensions and all directions. It can observe the inside, edge, and axillary lymph nodes of tumors. Especially for lymph nodes with small volume and regular shape, it provides a more accurate judgment method for doctors. Secondly, through dynamic contrast-enhanced MRI (DCE-MRI), tumor angiogenesis and microvessel density can be evaluated, reflecting the blood supply of tumors, which is helpful for the diagnosis of benign and malignant tumors. However, the boundary of breast cancer in MRI images may not be clear, and the tumor size and shape are diverse.Especially when the lesion area of breast cancer is close to the surrounding normal human tissues, it is difficult to accurately define the boundary of the tumor with the naked eye. Secondly, the tumor area contains necrotic areas, fibrous tissues and tumor cells of different grades. Different tissue types will lead to different signal intensities on MRI images, greatly increasing the difficulty of segmenting the lesion area of breast cancer.

At present, the segmentation methods for breast cancer lesion areas include multimodal segmentation methods[2], segmentation methods combined with prior knowledge[3], and deep learning methods[4]. Among them, the multimodal segmentation method combines the PET/CT imaging technology of positron emission tomography and computed tomography, and proposes a segmentation method for breast cancer lesion areas based on collaborative learning and Transformer. By taking advantage of bimodal data, the segmentation accuracy is improved. The segmentation method

combined with prior knowledge combines the active contour model of Markov random field (MPF) energy and fuzzy speed function for the segmentation of breast cancer lesion areas in DCE-MRI images. By using MRF energy to enhance the contrast between the target area and the background and superimposing the consideration of the spatial correlation between pixels, the accuracy of segmentation is improved.This method uses MRF energy to enhance the contrast between the target area and the background and simultaneously considers the spatial correlation between pixels, improving the accuracy of segmentation. Deep learning models, especially convolutional neural networks (CNN), can automatically learn complex feature representations from images without the need to manually design feature extractors, which is an essential step in traditional methods. After training, deep learning models not only perform well on specific datasets but also can generalize well to other similar datasets. In particular, models pre-trained on large-scale datasets can adapt to different medical image segmentation tasks through transfer learning or fine-tuning.

With the development of precision medicine, deep learning algorithms have been widely recognized in medical images, and more and more algorithms have been put into clinical practice research . Breast cancer detection and diagnosis technology based on deep learning can automatically analyze images. Through learning and training on a large amount of breast image data, automatic detection and accurate diagnosis of breast cancer lesion areas can be achieved. Deep learning algorithms can provide doctors with more accurate information and formulate more precise treatment plans in judging whether breast masses are benign or malignant and predicting the development trend of breast cancer. It can be said that deep learning has great potential and advantages in the diagnosis and treatment of breast cancer and has achieved remarkable scientific research results in multiple fields. However, due to the certain complexity of human anatomical structure, deep learning models still have some difficulties in accurately segmenting breast cancer lesion areas. Therefore, optimizing the algorithm for segmenting breast cancer lesion areas to assist doctors in accurate diagnosis and treatment has great research significance.At present, various networks have been proposed for segmentation. The Res-Unet[5] network proposed by Zhang et al. in 2017 is mainly for semantic segmentation.For medical segmentation, the Unet++ model[6]proposed by Zhou et al. in 2018, the TransUnet[7]model proposed by Chen et al. in 2021, and the Swin-Unet[8] model proposed by Cao et al. have all greatly improved the segmentation accuracy of medical images.Compared with these, U^2-Net has more advantages in segmentation accuracy, generalization ability, practicability, stability of the training process and fitting on small samples.This paper conducts further research on the segmentation of breast cancer lesion areas by improving the U^2-Net neural network.

1.2 Theoretical basis of medical image segmentation research

1.2.1 Imaging principle of MRI images

Magnetic Resonance Imaging (MRI) is an advanced medical imaging technology. Its imaging principle is based on the characteristics of atomic nuclei in a strong magnetic field. The human body contains a large number of hydrogen protons. When there is no external magnetic field acting on them, the direction of their magnetic moments is random. When the human body is placed in a strong magnetic field, the magnetic moments of hydrogen protons will align along the direction of the main magnetic field, generating a longitudinal magnetization vector. At this time, according to

the Larmor equation:

where ω is the precession frequency, that is, the frequency at which hydrogen protons precess around the direction of the main magnetic field; γ is the gyromagnetic ratio, which is about 42.58 MHz/T for hydrogen protons; B_0 is the main magnetic field strength. This means that the stronger the main magnetic field strength, the higher the precession frequency of hydrogen protons. When a radio frequency (RF) pulse of a specific frequency is applied, and this frequency is the same as the precession frequency of hydrogen protons, hydrogen protons absorb energy, the longitudinal magnetization vector deflects, and a transverse magnetization vector is generated. After the radio frequency pulse stops, the transverse magnetization vector begins to decay, and at the same time, the longitudinal magnetization vector gradually recovers. The energy released in this process is received by the receiving coil and a magnetic resonance image is formed after computer processing.

Table 1: MRI values (ms) of normal human tissues.

biological tissue	T1	T2
adipose	80-130	80-100
bone	2000	10-30
lung	800-1500	30-50
thyroid gland	400-800	50-100
muscle	700-1200	40-50
mammary glandular tissue	700-1000	60-120

1.2.2 Theoretical basis of deep learning

The theoretical basis of deep learning mainly includes the principles of neural networks and convolutional neural networks.

A neural network is a computing model that imitates the biological nervous system. It consists of a large number of neurons. Each neuron receives multiple input signals, performs weighted summation through certain weights, and then generates an output signal after being processed by an activation function. Multiple neurons are interconnected to form a network structure.

The learning process of a neural network is mainly achieved by adjusting the connection weights between neurons. Usually, the backpropagation algorithm is used. Combined with the given training data, after calculating the error between the output and the expectation, the error is transmitted back to the network in reverse. Then, it is necessary to gradually adjust the weights to continuously reduce the error. By repeating this process continuously, the neural network can learn the complex mapping relationship between input data and output results.

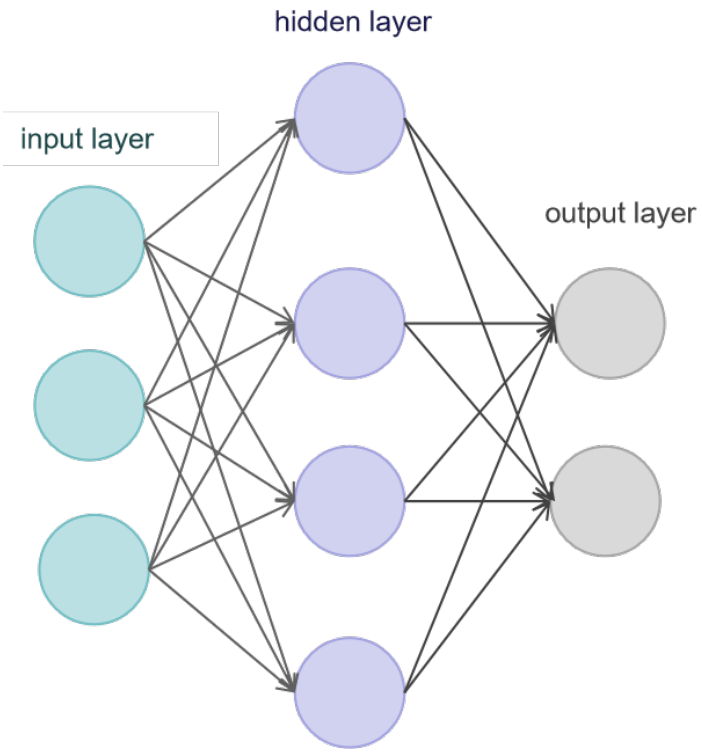


Fig. 2. neural network

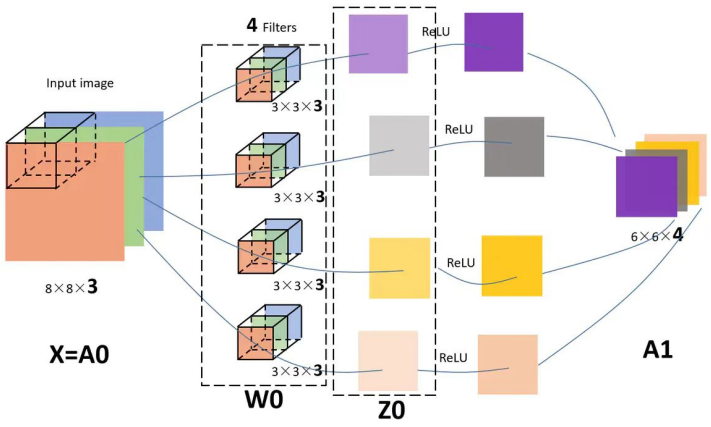


Fig. 3. Structure diagram of convolutional neural network

A convolutional neural network is a neural network specially used for processing data with grid structures such as images and videos. Its core idea is local connection and weight sharing. Local connection means that each neuron is only connected to a local area of the input data, rather than being fully connected as in traditional neural networks. This can greatly reduce the number of parameters, improve computational efficiency, and be able to extract local features of the input data. Weight sharing refers to the fact that multiple neurons in the convolutional layer use the same weight. This enables convolutional neural networks to automatically learn features with translation invariance, that is, for objects in images, no matter where they appear, they can be effectively recognized.

A convolutional neural network usually consists of convolutional layers, pooling layers, and fully connected layers. The convolutional layer extracts the features of the input data through convolution operations; the pooling layer performs downsampling on the features to reduce the feature dimension and improve the robustness of the model at the same time; the fully connected layer integrates the extracted features and outputs the final classification or prediction result.

2 Related Work

2.1 Medical Image Segmentation

Medical image segmentation is a crucial step in the diagnosis and localization of breast tumors. However, this task is highly challenging due to the diversity of tumor morphology, the fuzziness of boundaries, and the similar intensity distribution with surrounding tissues. Despite the extensive application of U-Net and its variants in this field, existing architectures still face two major limitations:

(1) Neglecting the feature representation capabilities of the baseline network.

(2) Introducing additional complex operations, increasing the difficulty of understanding and replicating the network.[9]

Based on these observations, we have also conducted related research on MRI breast cancer imaging and medical image segmentation.

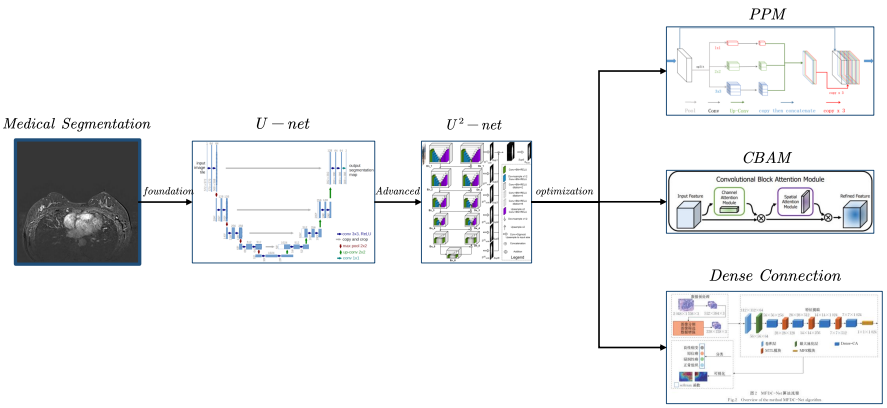


Fig. 4. related work

2.2 Application of the U^2-Net Basic Model in Image Segmentation

It can be seen in the Figure 4.In recent years, the deep convolution network has made breakthroughs in many visual recognition tasks. The success of these networks is part of the big network, which is done in large data sets. However, in biomedical image processing, the lack of large amounts of training images often limits the wide application of deep learning techniques.

Ronneberger et al.[10] introduced U-Net, a pioneering network architecture specifically designed for segmentation tasks with limited annotated samples. U-Net's architecture is distinguished by its U-shaped structure, and an expanding path. This combination of skip connections between the contracting and expanding paths ensures that fine details are preserved and accurately segmented.

Building on this,Qin et al.[11] developed the U^2-Net model, which features a two-layer nested U-structure and RSU modules. This architecture captures multi-scale contextual information while maintaining computational efficiency, making it particularly robust for salient object detection (SOD).The U^2-Net is an advanced variant of the U-Net architecture, and its structure serves as the foundational basis for our model. We are committed to continuously improving its capabilities to achieve even better performance.

Then, by integrating a Pyramid Pooling Module (PPM) and a Convolutional Block Attention Module ($CBAM$), Fu et al.[12] proposed An enhanced U-network which has strengthened the connectivity between the encoder and decoder subnetworks, effectively enhancing segmentation precision.The insights from their work on attention mechanisms are being utilized to refine our own model.

Additionally, Fang et al.[13] advanced breast cancer pathological image classification by incorporating multi-scale features and coordinate attention mechanisms within DenseNet. Their approach demonstrates the potential of dense connections and attention mechanisms, which we are integrating to improve both the classification and segmentation performance of our model.

3 Processing and limitations of the dataset

3.1 Preprocessing of the dataset

The dataset we used primarily consists of two parts. One part is the unlabeled dataset, and the other part is the dataset that has already undergone marking processing. The unlabeled dataset totals $2,340$ images. After performing batch format conversion, uniform resolution, noise reduction, and normalization operations on these images, we obtained a total of 430 high-visibility JPEG images. By learning medical knowledge and understanding the location of breast cancer lesion areas, we imported the jpg files into the labelme software to manually annotate the suspected areas of breast cancer tumors. The resulting JSON files were converted into labeled PNG images to facilitate subsequent model training. The other part of the dataset, which had already been labeled, included a total of $7,089$ MRI images after screening. We selected $6,380$ images for model training and 709 images for model testing. We carried out corresponding renaming operations on them to ensure the availability of the model training. Finally, we performed cropping and standardization operations on a total of $6,810$ images for the final model training, with 709 images used for model testing and evaluation.

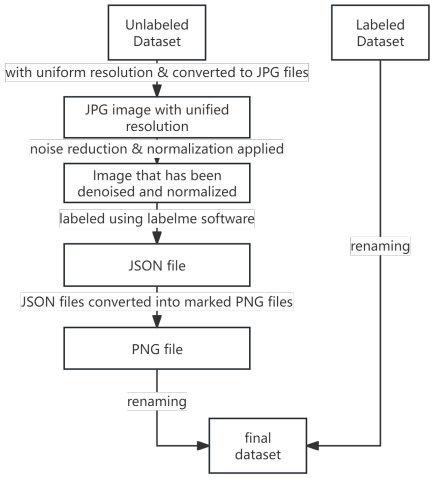
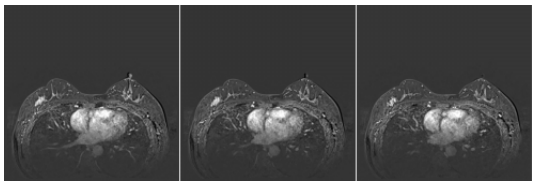
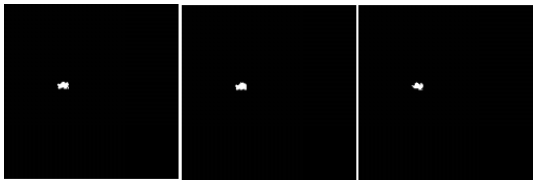


Fig. 5. Flow chart



(a) Examples of Breast MRI Data Images



(b) Examples of Annotated Data Images

3.2 Limitations of the dataset

However, in the dataset we selected, each patient provided 2-26 images. Since the clear MRI slicing areas are concentrated in almost the same lesion area, this results in a large number of nearly identical MRI images in the dataset. In addition to the low training efficiency, training the model

Fig. 6. Similar images that are likely to lead to poor model testing

multiple times on similar images will lead the model to learn the noise and details in the data, resulting in good performance on the training set but poor performance on unseen data, leading to overfitting problems. Overfitting issues will reduce the model's generalization ability for new data, making the model perform poorly in practical applications.

4 Methodology

This paper presents a novel image segmentation model that builds upon the U^2-Net network, specifically optimized for MRI image segmentation. The proposed model introduces the CBAM ($Convolutional\ Block\ Attention\ Module$) to the RSU-block and incorporates dense connection modules within the RSU-block to enhance feature extraction and information transmission.

4.1 Model Structure

The enhanced model retains the foundational encoder-decoder structure of U^2-Net, known for its multi-level feature extraction capabilities. However, the new model addresses some limitations in U^2-Net, particularly in handling fine details and boundary predictions. The improvements focus on two key areas:

1. CBAM Attention Mechanism: CBAM modules are embedded at the input and output sections of each RSU-block. The CBAM module consists of two sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). CAM enhances the weighting of important channel features, while SAM focuses on highlighting relevant spatial features, thus reducing the impact of irrelevant features.

2. Dense Connection Module: Added independently in both the encoder and decoder sections to ensure effective information transmission without redundancy. The dense connections are not added within the central RSU-block, as it already has skip connections that prevent repetitive information flow.

Algorithm 1 Improved U^2-Net with CBAM and Dense Connections

Input: MRI image X
Output: Segmented mask Y
procedure U^2-NET WITH CBAM AND DENSE CONNECTIONS
    1. Encoder:
    for each encoder layer l **do**
        Apply convolution to extract features: $F_l = \text{Conv}(X)$
        Apply Dense Connection: $F_l = [F_0, F_1, ..., F_{l-1}]$
        Apply CBAM: Only in input section.
    - Compute channel attention: $F_l = \text{CAM}(F_l)$
    - Compute spatial attention: $F_l = \text{SAM}(F_l)$

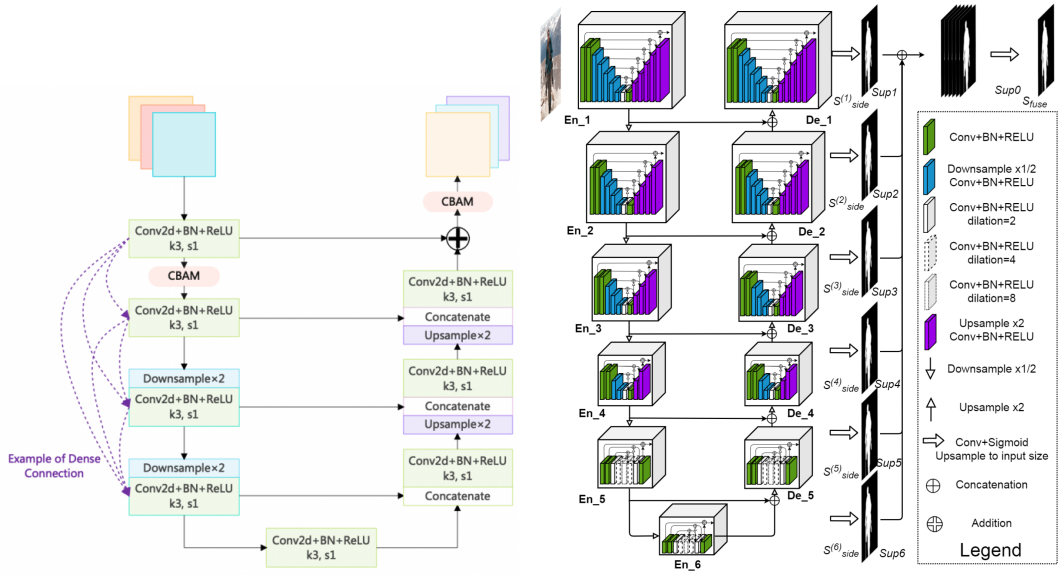
    end for
    2.RSU-block:
    Apply RSU-block with Dense connection to extract deep features in the encoder part
    3. Decoder:
    for each decoder layer l **do**
        Apply upsampling to the previous layer
        Apply convolution and dense connection: $F_l = \text{Conv}([F_0, F_1, ..., F_{l-1}])$
        Apply CBAM: Only in ouput section.
    end for
    4. Final Output:
    Apply final convolution to obtain segmented mask Y
end procedure

4.2 CBAM Module

The attention mechanism, as introduced in Vaswani et al.'s work on transformers [14] , plays a crucial role in enhancing feature learning by focusing on important aspects of the feature maps. This idea is extended in the CBAM module, where both channel and spatial attention mechanisms are utilized to refine the feature maps.

4.2.1 Channel Attention Mechanism (CAM)

In the Channel Attention Mechanism (CAM), the input feature map is first subjected to Global Average Pooling and Global Max Pooling operations along the channel dimension. Then, the pooled features are then processed by a Multilayer Perceptron (MLP) to compute channel attention weights. Through this process, the original feature map is weighted according to the computed attention values, thereby enhancing the significance of more important channels. Specifically, the steps are followed. Let X represent the input feature map with C channels. After performing Global Average Pooling (GAP) and Global Max Pooling (GMP), the resulting features are passed through the MLP to



(a) Optimized residual U-block (RSU-4). (Only an example for encoder section.)

(b) Illustration of U^2-Net architecture[11]. (Adapted from Qin et al., 2020)

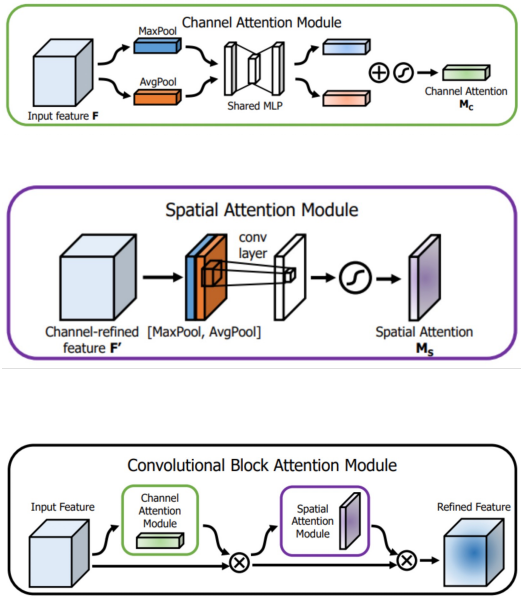
Fig. 7. Overall model structure.

generate channel-wise attention weights M_c. These weights are subsequently applied to the original feature map X to produce the weighted output feature map X'.

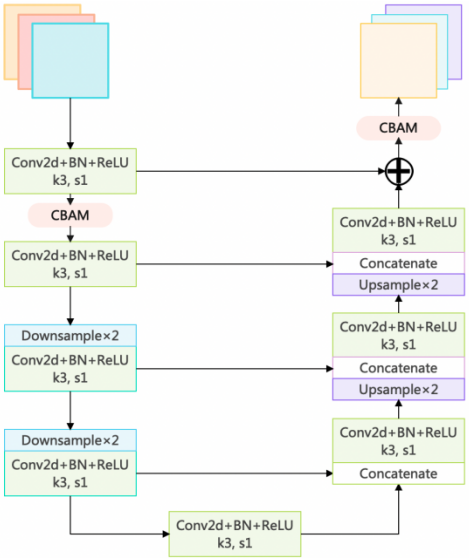
4.2.2 Spatial Attention Mechanism (SAM)

In the Spatial Attention Mechanism (SAM), Global Average Pooling and Global Max Pooling are applied across the spatial dimensions to capture global information. Then, the pooled feature maps are then concatenated and processed through a convolution layer to compute spatial attention weights. In addition, the original feature map is weighted according to these spatial attention values to enhance spatially significant features. Overall, the detailed steps are followed. Let X denote the input feature map. Spatial pooling operations are applied to generate pooled feature maps, which are then concatenated and passed through a convolution operation to compute spatial attention weights M_s. The final output feature map X' is obtained by applying these weights to the original feature map.

This series of processes demonstrates how attention mechanisms effectively enhance key information in feature maps by weighting operations across different dimensions.



(a) (a) CBAM Module Architecture



(b) (b) RSU-4 with CBAM

Fig. 8. Illustration of CBAM in RSU-block. (b) is an example of RSU-4; other RSU-blocks are similar.

4.3 Dense Connection Module

In this model, the dense connection module[15] is primarily added independently to the encoder and decoder sections. Given that the RSU-block already contains skip connections in the middle part, the dense connections are employed to further optimize feature extraction and information transmission, without duplicating information.

4.3.1 Independent Dense Connections in Encoder and Decoder Sections

In both the encoder and decoder sections, the output of each layer is not only used as the input for the subsequent layer but is also directly connected to all subsequent layers, forming a dense connection pattern. These independent dense connections promote the reuse of features and enhance the flow of information effectively.

4.3.2 Feature Fusion

Before concatenating feature maps, we adjust the size of each feature map to ensure consistent height (H) and width (W) across layers. This resizing step is crucial for maintaining the structural integrity of the network. The input of each layer is conducted by concatenating the outputs of all previous layers $F_0, F_1, \cdots F_{i-1}$ through the concatenate operation:

where F_l is the output of the l-th layer, and H_l represents the operation performed by the l-th layer.

4.3.3 Output Feature Map

The final output feature map of the RSU block is generated by combining the output maps of all layers, enhancing the diversity and expressiveness of the extracted features.

4.4 Loss Function

For binary classification tasks, the Binary Cross Entropy (BCE) loss function, as introduced in the original U2Net paper, is utilized to maintain efficiency:

$$BCE_{Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \log(p_i) + (1-y_i) \cdot \log(1-p_i)] \tag{1}$$

where y_i is the true label, p_i is the predicted probability, and N is the number of samples. By minimizing the BCE loss, the model's performance in segmentation tasks can be effectively enhanced.

4.5 Benefits and Disadvantages

The modifications to the model offer several advantages alongside some drawbacks. One notable advantage is improved feature extraction due to the inclusion of CBAM modules, allowing the model to focus on key features and significantly enhancing segmentation accuracy. Additionally,

dense connections aid in transmitting information across layers, ensuring that important features are preserved throughout the network.

However, these improvements also come with certain trade-offs. The addition of CBAM and dense connections increases the model's complexity, leading to a larger parameter count and potentially higher computational demands. While dense connections help maintain information flow, they may also introduce redundancy if not carefully controlled.

By incorporating the CBAM and Dense Connection modules within the RSU-block, the enhanced model not only retains the strengths of the original U^2-Net architecture but also significantly improves feature extraction and information transmission. Although these modifications increase the model's complexity, they lead to notable improvements in segmentation accuracy, making the model particularly well-suited for challenging scenarios. Looking forward, as computational resources continue to advance, models with higher parameter counts but superior performance, like this one, will likely become more prevalent, especially in fields such as medical image segmentation[16].

5 Experimental results of breast cancer MRI dataset

5.1 Evaluation index

To quantify the segmentation results of this model, this paper will evaluate the quality of the segmentation results from three dimensions: loss train, mean absolute error (MAE), and maximum F1 score (maxF1).

Train loss is a comprehensive indicator that measures the overall error during training, providing insight into how well the model is learning. However, it does not give detailed information about the prediction accuracy. The calculation formula is shown in equations 2.

$$\text{Train}_{loss} = -\sum_{i=1}^{N} y_{ij} \log(p_{ij}) \tag{2}$$

where y_{ij} is the true label, p_{ij} is the predicted probability, and N is the total number of samples.[17]

Therefore, to provide a more precise assessment of prediction accuracy, MAE is introduced to evaluate the average magnitude of errors between the predicted and actual values, offering a more precise assessment of prediction accuracy. The calculation formula is shown in equations 3.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{3}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the total number of samples.[18]

Additionally, the maxF1 score is introduced to judge the balance between precision and recall, indicating the model's performance in terms of correctly identifying true positives while minimizing false positives and false negatives. The smaller the loss train and MAE values are and the larger the maxF1 value is, the better the performance of the model is. Their calculation formulas are shown in equation 4.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

where precision denotes precision and recall denotes recall.[19]

5.2 Ablation experiment

To measure the impact of the fusion of channel attention mechanism and dense connection module on the experimental results, a model ablation experiment was designed. Based on U^2-Net as the basic network, four comparative experiments were conducted: (1) without adding any modules; (2) adding dense connection module, without adding channel attention module; (3) adding channel attention module, without adding dense connection module; (4) adding dense connection module and channel attention module.

5.2.1 Comparison of deep learning capabilities of different networks

From the value of Train Loss shown in 9, all four models dropped significantly after one epoch and can be maintained below 0.02 after five epochs. This shows that these models can quickly fit the known image information and they all have a good learning accuracy. From the value of MAE shown in 10, we can know that the U^2-Net with CBAM channel attention module has a low initial value, which shows that this model suits the prediction task very well. When we focus on prediction, the U^2-Net with CBAM channel attention module also has great advantages in prediction accuracy and stability. Furthermore, according to the value of maxF1 shown in 11, the fifth epoch of U^2-Net with CBAM attention mechanism reached the highest among all the training models. The indicates that the U^2-Net with CBAM channel attention module excels in balancing precision and recall, which is crucial for tasks requiring accurate pixel-wise predictions.

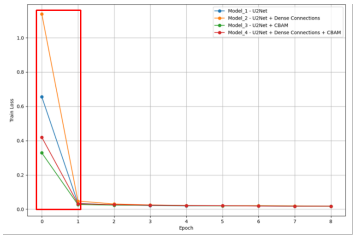


Fig. 9. Train loss comparison

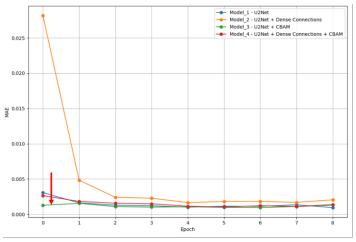


Fig. 10. MAE comparison

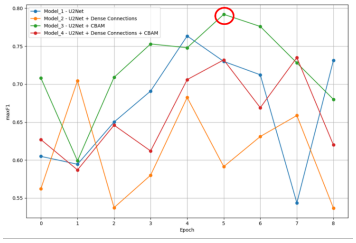


Fig. 11. maxF1 comparison

5.2.2 MAE result analysis between the best model of each network

After training each network for eight epochs to gain the best model, we analyzed the values of MAE to test its accuracy in predicting tumor locations in breast cancer MRI images. The experimental results are shown in 2.

The results show that compared with the original U^2-Net network, the averages of MAE from the three models have increased, while the variance and median have decreased. From average, the U^2-Net with the CBAM attention mechanism increased by 9.16% compared with the traditional U^2-Net, the U^2-Net with the CBAM attention mechanism and dense connections increased by 21.58% compared with the traditional U^2-Net, and the U^2-Net with dense connections increased by 27.83% compared with the traditional U^2-Net.

In terms of variance, the U^2-Net with dense connections has a 57.50% decrease compared to the traditional U^2-Net, while the U^2-Net with the CBAM attention mechanism has a 12.05% decrease compared to the traditional U^2-Net. In terms of the median, U^2-Net with CBAM attention mechanism and dense connection decreased by 18.89% compared with the traditional U^2-Net, the U^2-Net with CBAM attention mechanism decreased by 16.68% compared with the traditional U^2-Net, and the U^2-Net with dense connection decreased by 9.82% compared with the traditional U^2-Net.

Table 2: Results of ablation experiments

Enhancement	Average	Variance	Median
U^2-Net	0.000862	0.0000010155	0.0007172305
With CBAM	0.000941	0.0000008931	0.0005975913
With Dense	0.001102	0.0000004316	0.0006468121
With Both	0.001048	0.0000012165	0.0005816931

All four models have considerable superiority in accurate image segmentation and lesion localization, and its efficient diagnostic ability provides strong support for doctors. However, although the traditional U^2-Net performs well in terms of average, it has a large variance and is prone to extremely inaccurate examples, which means hidden dangers for the application of medical image segmentation. To more intuitively demonstrate the differences between various prediction methods, the visualization results are presented in 12. The MRI image represents the original image, and the Ground Truth represents the manually annotated tumor area. We can see that the original U^2-Net has a large absolute error, incorrectly labeling the tumor area on the right side, while the other three models did not exhibit such errors.

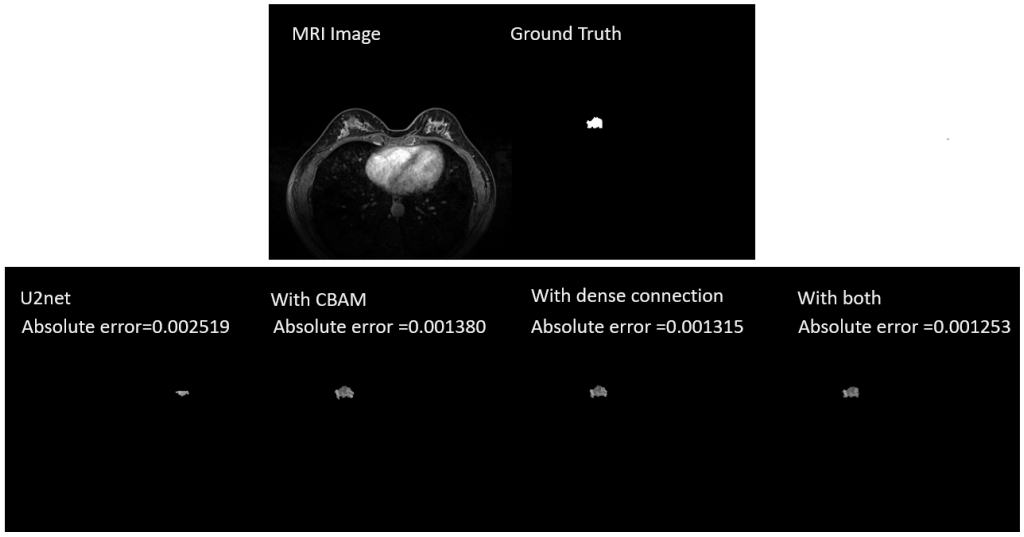


Fig. 12. Visualized results

By providing more dependable predictions, the U^2-Net with CBAM attention offers clear improvements for doctors utilizing deep learning models for diagnosis. It achieves the smallest variance and a lower median while maintaining a consistently low average error, which indicates that compared to traditional U^2-Net this model offers greater stability where more images are predicted with a higher accuracy. In the area of predictions, the reduction in variance also implies fewer extreme outliers, which enhances the reliability of the model in breast cancer MRI image settings. The level of consistency and precision is critical in clinical practice, where even a single inaccurate prediction can lead to great damages. This model's ability to minimize variability in results allows clinicians to have more confidence, supporting more effective and timely decision-making in patient care.

In summary, while the traditional U^2-Net model still demonstrates strong performance in certain metrics, its higher variance makes it less reliable for real-world breast cancer MRI images segmentation applications. The U^2-Net with CBAM attention mechanism, by contrast, balances precision, accuracy, and stability, making it a more robust tool for assisting in the early detection and diagnosis of breast cancer.

6 Summary

In this study, we improved the segmentation performance of breast cancer MRI images by integrating convolutional block attention modules (CBAM) and dense connections into the U ^{2} -Net network. Compared with the traditional U ^{2} -Net, the improved model performs better in feature extraction and information transfer, and significantly improves the accuracy of segmentation and robustness of the model. Although these enhancements increase the computational complexity of

the model, the resulting improved accuracy is of great significance in the field of medical image processing. In the future, we will further expand and optimize our model, which means that the future research will focus on optimizing the model structure to maintain or further enhance the existing segmentation effect while reducing the need for computing resources. In addition, we plan to explore lighter versions of the model to achieve more efficient performance in clinical applications.There is still a long way to go, but we will continue to work towards research.

Acknowledgements

Ye Lin, Zongyan Dai, Qi Jing, and Rui Shi contributed equally to this work and should be considered co-first authors.

References

[1] Huang Chongquan Chen Lijun. Comparison of the diagnostic value of different imaging methods for breast cancer[j](in chinese). In *Chinese Journal of Woman and Child Health Research*, pages 29(08):1031–1035, 2018.

[2] Zhang Jing Shen Jing. Diagnostic value of high-frequency color doppler ultrasound and magnetic resonance imaging for early breast cancer[j](in chinese). In *Famous Doctors*, pages (02):9–11, 2024.

[3] Li Jie Zhang Xiaopeng. Application value and evaluation of mri in comprehensive treatment of breast cancer[j](in chinese). In *Chinese Journal of Practical Surgery*, pages 031(10):922–926, 2011.

[4] C. Sahaya Pushpa Sarmila Star, T.M. Inbamalar, and A. Milton. Automatic semantic segmentation of breast cancer in dce-mri using deeplabv3+ with modified resnet50. *Biomedical Signal Processing and Control*, 99:106691–106691, 2025.

[5] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, PP(99):1–5, 2017.

[6] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. 2018.

[7] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. 2021.

[8] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. 2021.

[9] Gongping Chen, Lei Li, Jianxun Zhang, and Yu Dai. Rethinking the unpretentious u-net for medical ultrasound image segmentation. *Pattern Recognition*, 142:109728, October 2023.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[11] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, October 2020.

[12] Zhongming Fu, Hejian Chen, Mengsi He, and Li Liu. An enhanced u-network by combining ppm and cbam for medical image segmentation. *IEEE Access*, 12:107098–107112, 2024.

[13] Y. Fang and F. Ye. Mfdc-net: A multi-scale feature and attention mechanism fusion algorithm for breast cancer pathology image classification(in chinese). *Journal of Zhejiang University: Science Edition*, 50(4):455–464, 2023.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. A. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[16] A. Khosla and L. Riss. Review of deep learning methods for image segmentation in medical applications. *IEEE Access*, 8:162337–162354, 2020.

[17] Ben Claydon Lucia Vadicamo Richard Connor, Alan Dearle. Correlations of cross-entropy loss in machine learning. *Entropy (Basel, Switzerland)*, page 491, 2024.

[18] Timothy O. Hodson. Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not. *Geoscientific Model Development*, pages 5481–5487, 2022.

[19] Shivam Mishra, Amit Vishwakarma, and Anil Kumar. Multi-headed u-net: an automated nuclei segmentation technique using tikhonov filter-based unsharp masking. *Smart Science*, 2024.