Proceedings of CONF-SEML 2025 Symposium: Intefrating AI into Software Engineering
DOI: 10.54254/2755-2721/156/2025.CH23275

The Segmentation Model for Breast Cancer Ultrasound Image Based on Attention U-Net

Jun Cheng

Mongolia University of Technology Hohhot, Hohhot, China
1809615941@qq.com

Abstract: The segmentation of ultrasound image for breast cancer is an important task in the field of biomedical research. The traditional U-Net model, with its simple structure and remarkable performance, this approach has found extensive application in the segmentation of medical images. However, U-Net tends to be affected by background noise when handling images with complex backgrounds or blurry boundaries, which may impact the segmentation accuracy. To address this issue, the Attention U-Net model incorporates an attention mechanism, enabling the model to selectively focus on critical target areas within the image, thereby improving segmentation accuracy. This paper further optimizes the Attention U-Net architecture by increasing the depth of both the encoder and decoder sections, enhancing the model's capacity for feature extraction and image reconstruction. Consequently, both the accuracy and robustness of segmentation are enhanced. The experimental findings indicate that the proposed modified Attention U-Net model significantly outperforms traditional methods in breast ultrasound image segmentation tasks. It effectively handles various types of breast images, particularly those with complex backgrounds, blurred targets, or small sizes, maintaining high segmentation accuracy. This study offers an effective solution for the automated segmentation of breast ultrasound images, with substantial implications for enhancing both the automation and diagnostic efficiency in medical image analysis.

Keywords: Attention U-Net, image segmentation, ultrasound image, breast cancer

1. Introduction

Within the domain of medical imaging, with continuous advancements in technology, imaging techniques are playing an increasingly crucial role in the early disease diagnosis and treatment. Ultrasound imaging (US), as a widely used, real-time, non-invasive, convenient, and cost-effective imaging technique, has become an essential tool for diagnosing various diseases, especially in cancer screening and diagnosis, where it shows tremendous potential [1]. Compared to other imaging technologies, ultrasound offers unique advantages in providing high-resolution soft tissue contrast, which makes it a critical tool in the diagnosis of many types of cancer.

Breast cancer, being one of the most prevalent cancers among women globally, necessitates early detection and precise diagnosis to enhance cure and survival rates. As reported by the World Health Organization (WHO), breast cancer has become the primary cause of cancer-related mortality among women worldwide. With the increasing incidence of breast cancer, particularly in developing countries, early screening and diagnosis have become increasingly urgent. Due to its characteristics, such as no radiation and high reproducibility, ultrasound imaging has become an key tool for breast

DOI: 10.54254/2755-2721/156/2025.CH23275

cancer screening. Through ultrasound images, physicians can assess the shape, size, and location of lesions, providing crucial information for subsequent diagnosis and treatment. However, visual observation of breast ultrasound images does not always allow for accurate differentiation between benign and malignant tumors, necessitating the use of image segmentation techniques. Image segmentation is a core task in the analysis of medical images, aiming to precisely extract regions of interest (e.g., tumor regions) from the background. In breast cancer diagnosis, accurate image segmentation not only helps doctors delineate the boundaries of tumors and determine their nature but also provides important references for surgical planning and treatment evaluation. However, the noise, low contrast, and variability in tumor shape and size in breast ultrasound images pose significant challenges for segmentation. Tumor boundaries are often unclear, and the tumor appearance varies considerably across patients, making it difficult for traditional segmentation methods to achieve ideal results. Therefore, overcoming these challenges and developing efficient and accurate segmentation techniques are key to enhancing the diagnostic capability of breast cancer ultrasound imaging.

Despite the significant success of traditional U-Net in medical image segmentation [2], it still faces some limitations when handling images with complex backgrounds or small target regions, especially in breast ultrasound image segmentation tasks. Breast ultrasound images often contain noise, low contrast, and irregular tumor shapes, which make U-Net prone to background interference, leading to imprecise segmentation results that affect accurate tumor diagnosis and detection. To address this issue, Attention U-Net was introduced, incorporating attention mechanisms (Attention Gates, AGs) into the traditional U-Net architecture, significantly improving the model's performance in complex images [3]. Unlike traditional U-Net, Attention U-Net can dynamically assign weights to different regions of the image through the attention gate mechanism, automatically focusing on parts of the image relevant to the target, thus avoiding interference from irrelevant backgrounds. This allows Attention U-Net to maintain the details of the target regions in the presence of noise and complex backgrounds, without producing overly smoothed segmentation results as traditional U-Net tends to do. This is especially advantageous when processing small and complex tumor regions. Furthermore, the attention mechanism in Attention U-Net can adaptively learn the shape, size, and location of target regions without the need for additional supervisory information. This makes the model more flexible and robust in segmentation tasks, particularly in cases where tumor shapes change significantly and boundaries are unclear. In contrast, traditional U-Net may confuse the target regions with the background in such tasks, whereas Attention U-Net significantly improves segmentation accuracy by precisely focusing attention. In comparison, traditional U-Net is more likely to produce blurry boundaries when handling complex breast ultrasound images. By introducing the attention mechanism, Attention U-Net not only effectively suppresses background noise but also provides higher accuracy in segmenting small targets, particularly in the early diagnosis of breast tumors, where it can more clearly and accurately identify tumor boundaries. Therefore, Attention U-Net has greater advantages and broader application prospects than traditional U-Net, especially in medical image segmentation tasks involving complex scenes and small target segmentation.

This study aims to propose an enhanced Attention U-Net model to further enhance the segmentation accuracy of breast ultrasound images while reducing interference from background noise. Through experimental validation, our model demonstrates good robustness and applicability, effectively segmenting breast cancer lesions.

2. Previous works

Traditional image segmentation methods typically rely on low-level features such as grayscale, edges, and texture, as well as heuristic algorithms. These methods have shown promising results in simpler scenarios but often face significant challenges in more complex images. Common traditional image

DOI: 10.54254/2755-2721/156/2025.CH23275

segmentation techniques include threshold-based segmentation, edge detection, and the watershed algorithm, each with its own strengths and limitations.

Threshold-based segmentation is one of the most basic and widely used methods [4]. It works by setting one or more grayscale thresholds to divide the pixels in the image into different regions. In simple terms, the thresholding method compares the pixel values in the image to a fixed threshold, classifying pixels above the threshold as one category and those below the threshold as another. The advantage of this method is its low computational cost and ease of implementation, making it particularly useful for images where the background and target are clearly distinguishable. However, threshold-based segmentation performs poorly when dealing with images that have complex backgrounds or uneven lighting.

Edge detection methods identify the edges of objects by detecting changes in pixel intensity in the image [5]. Widely edge detection algorithms include Sobel, Canny, and Prewitt. Edge detection is effective at extracting the contours of objects and is often used for segmenting structured areas within an image. Its main advantage is its ability to effectively extract prominent edge information, especially for images where object boundaries are clear and the background is relatively simple. However, edge detection is highly sensitive to noise, and in images with significant noise or where object boundaries are blurry or overlapping, it can lead to inaccurate edge detection.

The watershed algorithm is an image segmentation technique based on gradient information, inspired by the process of water flowing over a terrain [6]. This algorithm first calculates the gradient of the image and simulates water flowing from lower to higher points to segment different regions of the image. The watershed algorithm performs well when there are clearly defined contours between objects, allowing for precise segmentation. However, one of its inherent drawbacks is over-segmentation. In regions with noise or weak gradient variations, the algorithm may produce excessive segmentation, leading to undesirable results.

Although these traditional methods can achieve good results in simple or idealized scenarios, they often fail to deliver satisfactory segmentation in the face of more complex medical or natural images. Factors such as image noise, complex backgrounds, and irregular object shapes frequently impact their performance. In cases with severe noise or when the target and background share high similarity, these methods are particularly prone to suboptimal results. As a response to these issues, deep learning-based image segmentation methods have gradually become the prevailing, especially in applications like medical image analysis. Deep learning methods are better equipped to handle these challenges, providing more accurate and robust segmentation results.

With the widespread use of convolutional neural networks (CNNs) in computer vision, many CNN-based image segmentation methods have been proposed. These methods, compared to traditional segmentation techniques, offer enhanced automation and higher accuracy. Common CNN-based segmentation methods include U-Net, FCN, and Fast R-CNN.

U-Net is a deep learning architecture based on convolutional neural networks, specifically designed for medical image segmentation tasks [2]. Since its inception in 2015, U-Net has rapidly become one of the standard methods in the field of medical image analysis due to its excellent segmentation performance and adaptability. The architecture of U-Net consists primarily of a contraction path and a symmetric expansion path, forming a U-shaped structure. This design provides U-Net with unique advantages for processing medical images, particularly in tasks that require handling complex structures and detailed information. U-Net effectively addresses common issues in medical images, such as noise, low contrast, and complex structures, enabling outstanding performance in segmentation tasks involving breast ultrasound, brain MRI, liver CT, and other imaging data. Its precise image reconstruction capabilities and high sensitivity to small objects make U-Net particularly suitable for segmenting complex and tiny structures in medical images. For example, in breast cancer ultrasound images, the shape, size, and boundaries of tumors can exhibit

DOI: 10.54254/2755-2721/156/2025.CH23275

significant variation, but U-Net can effectively enhance the segmentation accuracy of tumor regions through multi-level feature fusion and detailed recovery processes.

Fully Convolutional Networks (FCNs) are a deep learning approach specifically designed for image segmentation [7]. The key feature of FCNs is replacing the fully connected layers of traditional neural networks with convolutional layers, enabling the network to handle input images of arbitrary sizes while producing segmentation maps of the same size as the input. FCNs offer significant advantages in image segmentation, particularly in pixel-level precision, as they can automatically learn features and structures from the image, reducing the need for manually designed features. However, due to the layer-by-layer propagation of information, FCNs may lose some fine details, leading to poor performance when dealing with complex structures and small targets.

Fast R-CNN is another CNN-based image segmentation method initially designed for object detection tasks [8]. Fast R-CNN works by extracting candidate regions (Regions of Interest, RoIs) from the image and performing feature extraction on each region to detect objects. Its main advantage lies in accelerating computation through shared convolutional features, which significantly improves efficiency and accuracy compared to traditional sliding window methods. However, Fast R-CNN has its limitations, particularly when applied to dense target segmentation tasks, where its performance may be restricted.

It is noteworthy that the introduction of attention mechanisms has had a significant impact on image segmentation, especially in U-Net-based models. In the segmentation of medical images, the target regions are often irregularly shaped and similar to the background, particularly in tasks such as breast ultrasound imaging, where small tumor boundaries can be easily blurred. The attention mechanism automatically adjusts the focus of the network based on the characteristics of the target, improving segmentation accuracy. This is especially beneficial for tumor detection and early diagnosis, as it allows for more accurate identification of tumor boundaries. The adaptive learning capability of attention mechanisms makes attention-based U-Net models more robust and efficient than traditional U-Net models when dealing with complex images.

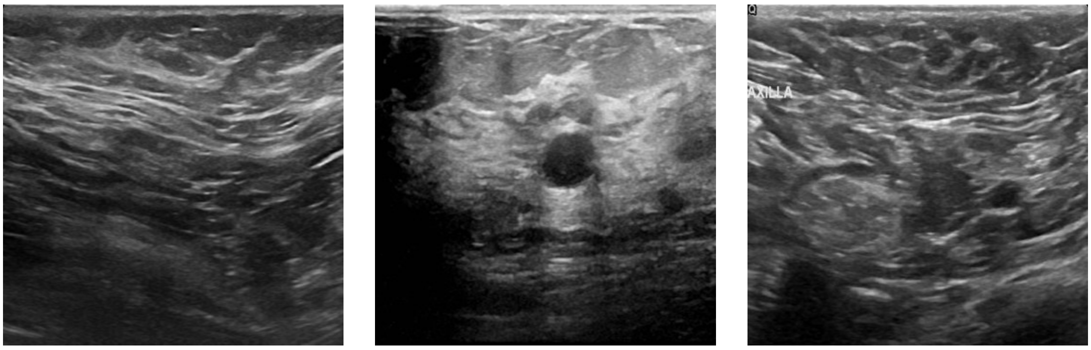
3. Dataset and preprocessing

Table 1: The three categories of breast cases and the number of images in each category

Case	Number of images
Benign	487
Malignant	210
Normal	133
Total	780

The dataset used in this study is derived from the breast cancer ultrasound image dataset, which was provided by The Baheya Hospital in Cairo, Egypt [9]. The dataset contains three categories of images: Normal, Benign, and Malignant, sourced from 600 female patients aged 25 to 75 years. During the data collection process, some images were found to be duplicates. The original images had a resolution of 1280×1024 pixels; after deduplication and preprocessing, the average resolution of the images was adjusted to 500×500 pixels. The distribution of sample quantities for each category in the dataset is shown in Table 1. Sample breast ultrasound images from the dataset are illustrated in Figure 1, while ultrasound images of different categories along with their corresponding mask images are displayed in Figure 2. The mask images clearly outline the boundaries of normal regions, benign masses, and malignant masses in the breast tissue, providing accurate target regions for training the segmentation model proposed in this study.

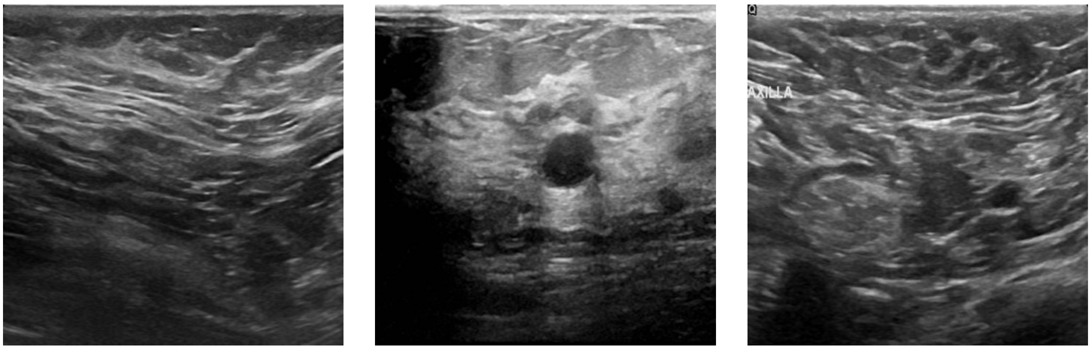
DOI: 10.54254/2755-2721/156/2025.CH23275



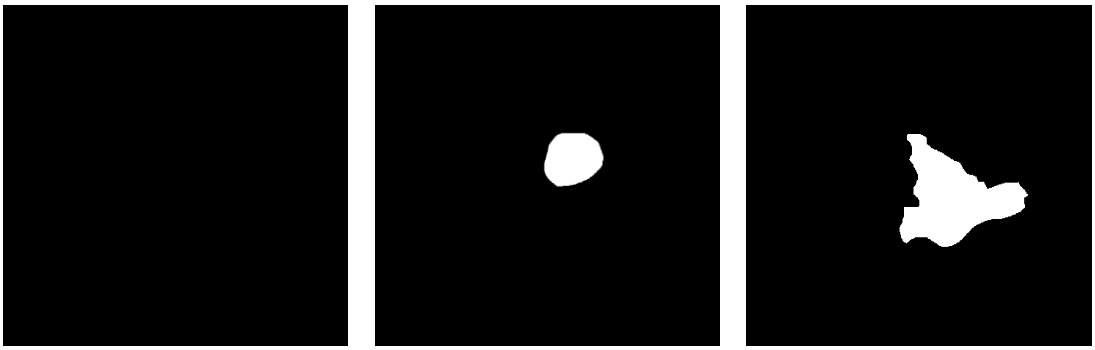
Normal Benign Malignant

Figure 1: Ultrasonic breast image data sample

Further preprocessing was performed on the images, including resizing, normalization, and batch loading, to ensure consistent input for the Attention U-Net model proposed in this paper. All images were resized to 256×256 pixels, and normalization was applied to convert the images into arrays and standardize them within the range [0, 1]. This normalization helps mitigate issues such as gradient explosion during model training. Additionally, batch processing techniques were employed to automatically load, crop, and transform both the original images and the mask images, thereby enhancing training efficiency.



Normal US image Benign US image Malignant US image



Ground truth image (Normal) Ground truth image (Benign) Ground truth image (Malignant)

Figure 2: Samples of various breast Ultrasound images and Ground Truth images

DOI: 10.54254/2755-2721/156/2025.CH23275

4. Model

This study employs an Attention U-Net model for the segmentation of breast ultrasound images. The Attention U-Net model introduces an attention mechanism based on the traditional U-Net, enabling the model to focus more effectively on the target regions and thus improving segmentation accuracy. U-Net, due to its simple yet effective structure and widespread application in medical image segmentation tasks, has become a classic model for image segmentation. However, traditional U-Net tends to be affected by background noise when handling images with complex backgrounds or blurry boundaries, which leads to imprecise segmentation results. To address this issue, Attention U-Net incorporates a self-attention mechanism between the contraction and expansion paths of the U-Net. This mechanism adaptively selects important regions while suppressing irrelevant information, thereby optimizing the transmission and fusion of feature maps and significantly enhancing segmentation accuracy. In particular, Attention U-Net performs better in identifying and segmenting target regions in complex backgrounds.

During the model training process, we used a breast ultrasound image dataset that contains a large number of accurately labeled breast images, covering different types of tumors and their surrounding tissues. To ensure the model's generalization capability, we set the training and test set split to 80% and 20%, respectively, with a batch size of 4. The number of convolutional filters in the encoder is set as 64, 128, 256, and 512, with dropout rates of 0.25, 0.25, 0.3, and 0.3 for each layer, respectively. The number of convolutional filters in the decoder is set to 256, 128, 64, and 32, with dropout rates of 0.3, 0.3, 0.25, and 0.25, respectively. The convolutional filter sizes in the attention mechanism part are the same as those in the decoder. The architecture of the improved Attention U-Net model is illustrated in Figure 3.

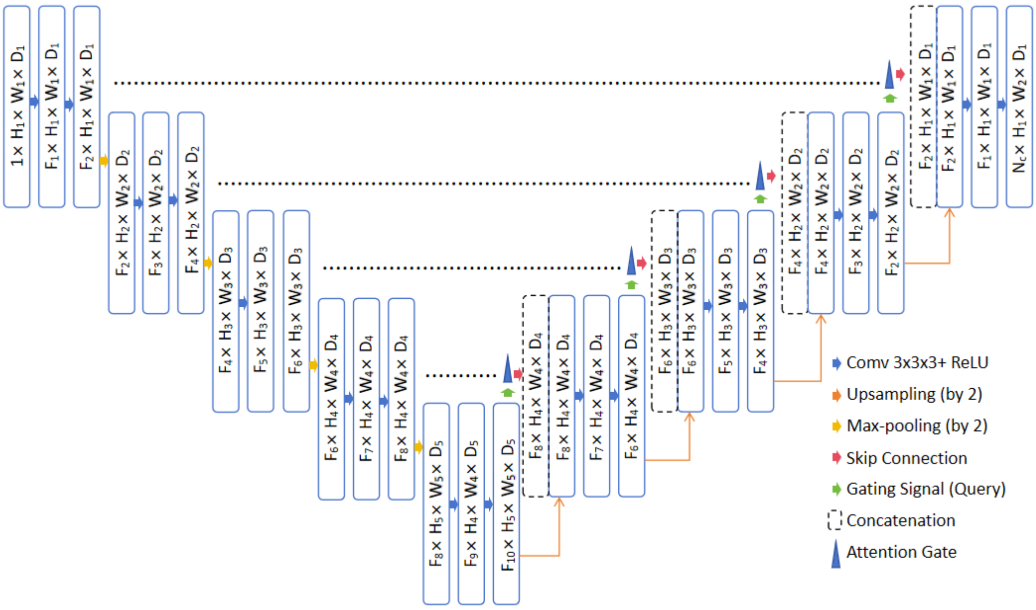


Figure 3: Model structure diagram

Based on the original Attention U-Net model, we made improvements to its network architecture and information integration method. Specifically, we deepened both the encoder and decoder, adding more convolutional and pooling layers. The original Attention U-Net model had a depth of three layers, but by adding an additional layer, the model is better able to extract image features, thus improving the image segmentation performance. The model is trained using the binary crossentropy loss function to optimize the segmentation results. During the optimization process, we used the

DOI: 10.54254/2755-2721/156/2025.CH23275

Adam optimizer, which combines the advantages of momentum and adaptive learning rates to efficiently handle large-scale data and high-dimensional parameter spaces. The initial learning rate for Adam is set to 0.001, with a learning rate decay strategy applied to progressively reduce the rate. This approach facilitates finer adjustments to the model parameters as optimization approaches the optimal solution, thereby further improving segmentation accuracy.

5. Results

In this study, we conducted multiple experiments on the improved Attention U-Net model to validate its effectiveness and feasibility. Some of the results are shown in Figure 4, where the predicted mask, processed mask, and GradCAM heatmap (used to explain the regions attended to by the attention mechanism) [10] are highlighted. The experimental results confirm that the model is capable of accurately segmenting the target regions even in images with complex backgrounds and small targets, showcasing its high accuracy and adaptability.

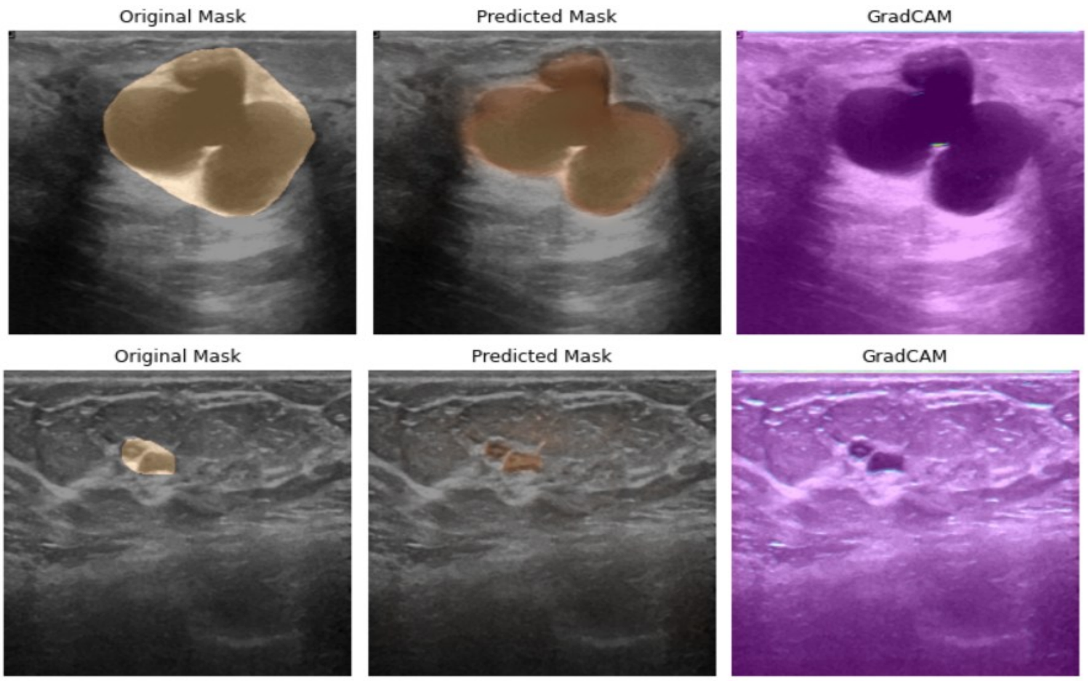


Figure 4: The results of segmentation

Through repeated experiments, we found that when the number of epochs was configured to 50, the model's accuracy stabilized at around 96%, eventually converging. This indicates that after 50 epochs, the model had effectively learned and extracted image features, achieving high segmentation precision and accuracy. Moreover, during training, the value of the model's loss function gradually decreased and stabilized, further confirming the model's convergence and stability. The high segmentation accuracy achieved by the Attention U-Net model can be attributed to the introduction of the attention mechanism. Compared to the traditional U-Net model, Attention U-Net enables the model to autonomously learn the shape and features of the image while effectively suppressing irrelevant regions and minimizing the interference from background noise. This significantly improves segmentation accuracy. Additionally, the attention mechanism also helps save computational resources, enhancing the model's operational efficiency. These advantages enable Attention U-Net to overcome some of the limitations of the original U-Net, performing excellently in breast ultrasound image segmentation tasks and making it an ideal tool for image segmentation.

DOI: 10.54254/2755-2721/156/2025.CH23275

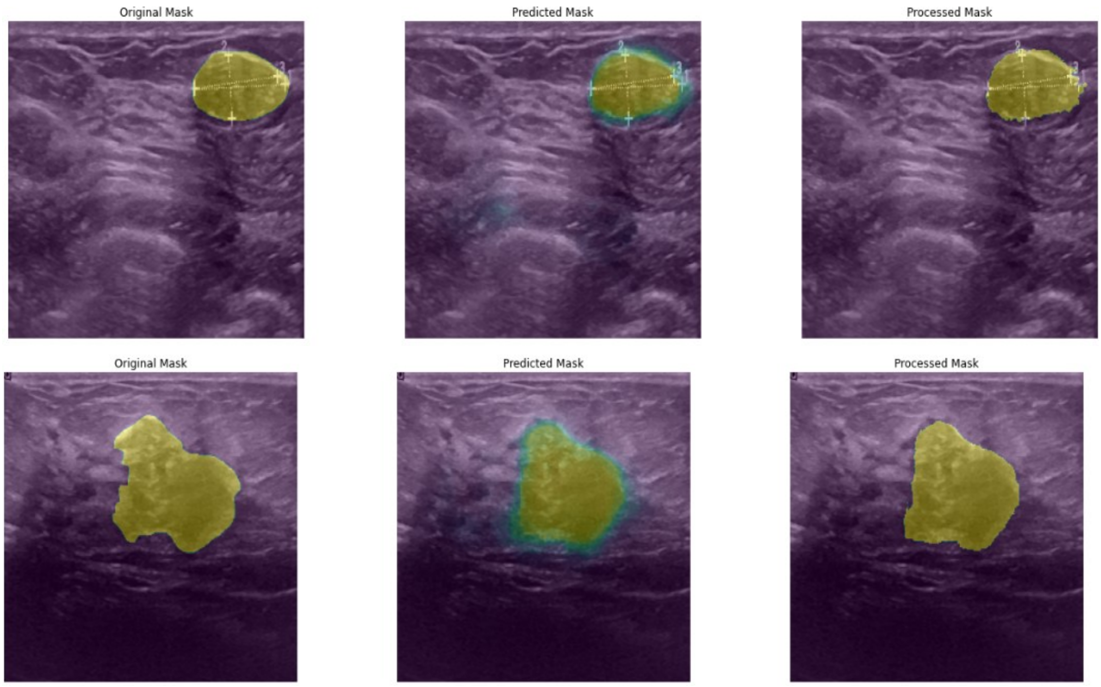


Figure 5: Samples with good segmentation results

We further analyzed the model's segmentation outcomes for different samples, as shown in Figures 5 and 6. Although the model achieved outstanding results in breast ultrasound image segmentation tasks, some limitations remain. As seen in Figure 6, segmentation errors occurred due to the small size of the target and significant background interference, as well as the relatively low image resolution. Although these segmentation errors occurred infrequently, further improvements and optimizations to the model are necessary to enhance its performance when handling such images. These improvements would further enhance its application effectiveness in the field of medical imaging.

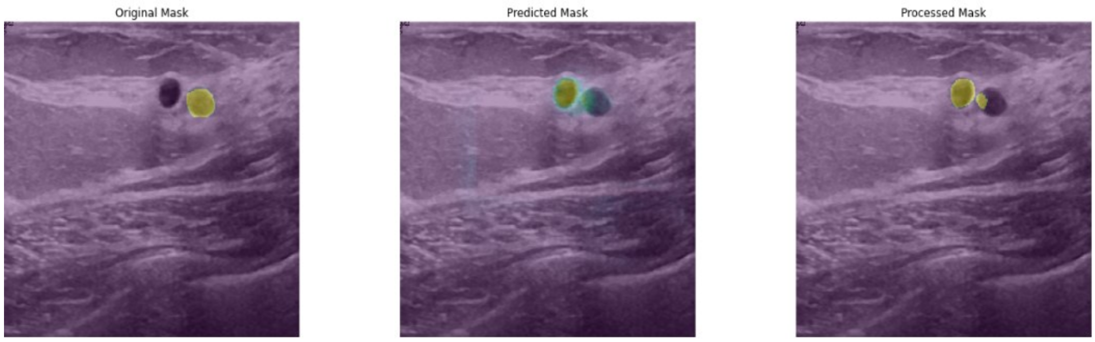


Figure 6: Samples with Substandard segmentation results

6. Discussion and conclusion

The Attention U-Net model, by incorporating the attention mechanism, has demonstrated significant advantages in image segmentation tasks. The model can more accurately focus on the target regions while effectively suppressing irrelevant background interference, thereby greatly improving segmentation precision. Even when faced with images that have varying target sizes, different shapes, or complex backgrounds, the Attention U-Net consistently delivers stable, efficient, and high-precision segmentation results. However, to further enhance segmentation performance, the model

DOI: 10.54254/2755-2721/156/2025.CH23275

still requires higher-quality data support. Additionally, when the model encounters strong noise interference or low-resolution images, the segmentation accuracy may be somewhat affected.

To address these potential limitations, this study suggests that future research should consider combining the Attention U-Net model with large model techniques. By leveraging the powerful representation and generalization capabilities of large models, the performance of Attention U-Net could be further enhanced. For example, pre-trained large models could be used as feature extractors, providing Attention U-Net with richer and more accurate feature representations and data support, thus reducing or even eliminating segmentation errors caused by severe image noise interference or low resolution.

The proposed Attention U-Net model holds significant guidance for the segmentation of breast ultrasound images and its application in clinical diagnosis and treatment. Thanks to its high-precision segmentation capabilities, the model can accurately identify lesion areas in breast tissue, providing reliable diagnostic references for doctors. Furthermore, the segmentation results assist in assessing tumor size, shape, and location, offering valuable information for surgical resection, radiation therapy, and other treatment options. Moreover, automated segmentation technology significantly reduces the workload of physicians, enhancing diagnostic efficiency and allowing them to focus more on patient communication, treatment, and care.

To further improve the performance of the Attention U-Net model and expand its application to other diseases, this study recommends integrating multimodal medical image data (such as MRI, CT, etc.) to enhance the model's segmentation accuracy and robustness. By adjusting and optimizing the model, it could be made adaptable to different disease needs. In particular, in clinical practice, the model should undergo extensive validation and continuous optimization, improving its performance and practicality based on real-world feedback. This will enable the model to play an even greater role in the field of medical image segmentation, offering robust technical support for disease diagnosis and treatment.

References

[1] R. Guo, G. Lu, B. Qin, and B. Fei, "Ultrasound imaging technologies for breast cancer detection and mana gement: A review," Ultrasound in Medicine & Biology, vol. 44, no. 1, pp. 37–70, Jan. 2018, doi: 10.1016/j. ultrasmedbio.2017.09.012.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[3] O. Oktay et al., "Attention U-net: Learning where to look for the pancreas," May 20, 2018, arXiv: arXiv:18 04.03999. doi: 10.48550/arXiv.1804.03999.

[4] M. Zeenath and M. A. Ali, "A review paper on image segmentation and its various techniques in image pro cessing," International Journal of Research, 2014, Accessed: Feb. 07, 2025. [Online]. Available: https://ww w.semanticscholar.org/paper/A-Review-Paper-on-Image-Segmentation-and-its-in-Zeenath-Ali/67ac7045ee6f5282 3929d5caeddb390d06451f0e

[5] H. G. Kaganami and Z. Beiji, "Region-based segmentation versus edge detection," in 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sep. 2009, pp. 1217–1221. doi: 10.1109/IIH-MSP.2009.13.

[6] A. S. Kornilov and I. V. Safonov, "An overview of watershed algorithm implementations in open source libraries," Journal of Imaging, vol. 4, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/jimaging4100123.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3 440. Accessed: Feb. 07, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Lon g_Fully_Convolutional_Networks_2015_CVPR_paper.html

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[9] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data in Brief, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int J Comput Vis, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.