Nile Journal of Communication & Computer Science

Volume 8, December 2024

Journal Webpage: https://njccs.journals.ekb.eg

A Dual-Attention-ResUNet++ for Breast Tumor Segmentation using Ultrasound Images

Asmaa A. Hekal[1,2], Hanan M. Amer[1], Hossam El-Din Moustafa[1]

[1]Electronics and Communications Engineering (ECE) Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt

[2]Electrical Department, Faculty of Engineering, Modern University for Technology and Information (MTI University), Cairo, 11439, Egypt

Abstract

This study introduces DAtt-ResUNet++, an advanced deep learning model specifically designed to enhance breast tumor segmentation in ultrasound images. The model integrates a Dual-Attention mechanism within the ResUNet++ framework, significantly improving its ability to focus on tumor regions while capturing relevant contextual information from surrounding tissue. By combining both spatial and channel-based attention, DAtt-ResUNet++ achieves higher segmentation precision. For evaluation, the model was rigorously tested on a public dataset of 780 breast ultrasound images, utilizing a robust 10-fold cross-validation approach. It achieved impressive results with a Dice similarity coefficient of 90.40 ± 0.88%, Intersection over Union (IoU) of 84.62 ± 1.12%, Sensitivity of 89.82 ± 0.75%, Precision of 93.44 ± 0.56%, and Accuracy of 98.73 ± 0.12%. These results position DAtt-ResUNet++ as a competitive tool against state-of-the-art methods, showcasing its potential to improve breast tumor segmentation in ultrasound imaging. Future research will explore further optimizations and validation on additional datasets.

Keywords: Deep Learning, Dual Attention Networks, Image Segmentation, Breast Ultrasound, Computer-Aided Diagnosis (CAD).

1. Introduction

Breast cancer is a major global health concern, affecting millions of women worldwide. In 2020, the most recent data reported 2.3 million new cases and 685,000 deaths worldwide caused by breast cancer [1]. Predictions for 2040 anticipate a notable uptick, projecting over 3 million new cases and approximately 1 million deaths annually [2]. Early detection of breast cancer is significantly aided by various imaging technologies. These include X-ray mammography, CT scans, magnetic resonance imaging (MRI), and breast ultrasound (BUS). Each modality offers a unique perspective on breast tissue, allowing healthcare professionals to identify potential abnormalities. Mammography, the standard diagnostic tool for doctors, has drawbacks due to its use of ionizing radiation, making it unsuitable for pregnant women[3]. On the other hand, BUS exams are cost-effective and radiation-free, making them a vital tool for cancer screening [4]. It offers essential insights into breast tissue characteristics and the presence of malignant tissues [4].

Unlike some screening methods that can be expensive or involve radiation, BUS imaging emerges as a cost-accessible and radiation-free tool, expanding the options for breast cancer screening programs [4]. BUS can reveal details about the composition of breast tissue and identify the presence of cancerous growths [4].

Segmentation of breast tumors in images is a critical step for computer-assisted diagnosis (CAD) systems designed to aid in treatment decisions [5]. However, accurately separating tumors in BUS images remains challenging due to the wide variety of tumor shapes, poorly defined borders, weak contrast, and built-in image noise [6]. Progress in CAD systems has spurred investigations into their impact on detecting and segmenting breast cancer using BUS images [7],[8]. Despite these challenges, CAD systems have demonstrated promising advancements in lesion detection and segmentation. Breast cancer segmentation and classification methods can be broadly classified into two categories: conventional techniques and deep learning (DL)-based approaches.

Conventional methods such as thresholding, region-growing, and feature-based approaches crafted by hand can be intricate and limitations in their feature representations can potentially lead to inaccurate lesion segmentation. Deep learning approaches have revolutionized breast lesion segmentation, achieving impressive accuracy in the detection of these abnormalities [9] , [10], [11], [12], [13], [14]. The field of breast cancer detection is actively exploring various deep learning techniques for segmenting lesions in BUS images. These techniques leverage convolutional layers to enable the automatic identification of important features within the images.

This study unveils a deep learning architecture named Dual-Attention-ResUNet++ (DAtt-

ResUNet++). It integrates attention mechanisms, inspired by the work of Oktay et al. [15], with the ResUNet++ framework introduced by Jha et al. [16]. This innovative design aims to achieve robust and precise segmentation of breast cancer in ultrasound (BUS) images. ResUNet++ merges elements from ResNet and UNet architectures. It utilizes residual blocks and borrows ResNet's skip connections to tackle issues like vanishing gradients that plague deep networks. U-Net's architecture leverages an encoder-decoder structure with crucial skip connections. These connections bridge the gap between low-level (local) and high-level (global) features, allowing the network to retain crucial spatial details.

Dual-Attention-ResUNet++ incorporates focus mechanisms that enable it to concentrate on significant areas within the images being analyzed. These mechanisms highlight informative regions while minimizing the importance of less relevant ones. This allows the model to effectively utilize both local details and broader contextual information from the image. As a result, Dual Attention ResUNet++ achieves improved precision and resilience in identifying lesions.

The key contributions of this paper are:

- A novel deep learning model, DAtt-ResUNet++, has been developed for the accurate segmentation of breast cancer in ultrasound (BUS) images.
- Spatial and channel-wise attention mechanisms are incorporated into DAtt-ResUNet++, enabling critical tumor regions in the image to be prioritized while less relevant areas are minimized. This dual attention allows for effective capture of both local details and broader contextual information.
- State-of-the-art segmentation accuracy is achieved by DAtt-ResUNet++ on the BUSI dataset, a widely recognized, complex dataset for breast ultrasound images. Its non-ionizing, radiation-free nature makes it well-suited for safe and accessible breast cancer screening applications.

The Manuscript Structure: Section 2: Related Work: This section provides a comprehensive review of existing research on breast cancer segmentation and enhancement techniques. Section 3: DAtt-ResUNet++ Architecture: This section delves into the intricate details of the proposed Dual-Attention-ResUNet++ model's architecture. Section 4: Experimental Results and Discussion: This section presents the experimental findings, explores their significance, and discusses their implications. Section 5: Conclusion and Future Work: This concluding section summarizes the key takeaways of the research, highlights the potential impact of DAtt-ResUNet++, and outlines potential avenues for future research directions.

2. Related work

A significant body of research has explored the segmentation of breast lesions in ultrasound imagery using a diverse range of algorithms. Breast cancer segmentation techniques can be categorized into two main approaches: traditional methods and those driven by machine learning. Traditional segmentation approaches typically rely on techniques like a region-growing, graph-based methods, and deformable models. In contrast, machine learning methods encompass a broader range, including both traditional hand-crafted feature engineering techniques and the more recent advancements in deep learning (DL). Region-growing segmentation starts by identifying seed points, either manually or automatically. These seeds act as starting points for the algorithm, which then progressively expands outwards, incorporating neighboring pixels that meet predefined criteria until the entire target region is segmented. For instance, Shan et al. [17] utilized this approach for breast cancer segmentation, incorporating criteria that considered both smooth contours and similar image intensity within the region. In contrast, deformable models employ a starting template that progressively adapts to conform to the object's boundaries. This adaptation process considers forces acting within the model itself (internal energy) and forces arising from the image data (external energy). One approach utilizes deformable models with boundary points and balloon forces, as introduced by Madabhushi et al. [18], to define the external energy influencing the model's adaptation. Leveraging prior advancements in the field, Chang et al. [19] proposed a two-stage approach specifically designed for breast lesion segmentation in ultrasound images. Their method first utilizes a stick filter to mitigate speckle noise, a common challenge in ultrasound data. Following this, a three-dimensional discrete-active contour method is utilized to achieve precise segmentation of the lesion. Early graph-based models relied on optimizing energy within frameworks like Markov random fields or graph cuts. Chiang et al. [20] explored a different approach, employing a pre-trained Probabilistic Boosting Tree (PBT) model to assist with breast lesion segmentation. Their method integrates the output of the PBT model within the energy function used by a graph cut algorithm., while Xian et al. [21] integrated frequency and spatial domain information into the energy function. However, these models often struggle with capturing intricate semantic details and faint boundaries in low-contrast ultrasound images, which can lead to inaccuracies.

Traditional machine-learning approaches depend on features manually designed by researchers to train classifiers for segmentation. For instance, studies by Liu et al. [22] and Jiang et al. [23] extracted specific features from local image regions and then trained classifiers like Support Vector Machines (SVMs) or Adaboost to perform breast lesion segmentation. The recent surge in machine learning has been fueled by deep learning (DL) techniques, particularly convolutional neural networks (CNNs). CNNs possess a remarkable ability to automatically extract high-level, meaningful features from large datasets of labeled data. Deep learning approaches have achieved remarkable accuracy in segmenting breast tumors from BUS images. Zhuang et al. [24] introduced the Residual-Dilated-Attention-Gate-U-Net (RDAU-Net) architecture, specifically tailored for breast tumor segmentation in BUS images.

Leveraging prior advancements, Hu et al. [25] proposed a segmentation method for breast cancer in (US) ultrasound images. Their approach innovatively combined a dilated fully convolutional network (FCN) with a phase-based active contour model. Byra et al. [26] employed a DL approach using entropy parametric maps for segmentation, while Zhu et al. [27] explored a method that leverages second-order statistics derived from various image subregions for segmenting breast lesions in ultrasound imagery. Shareef et al. [28] introduced the Enhanced Small Tumor Segmentation Network (ESTAN), a convolutional neural network (CNN) specifically designed to address the challenge of identifying small tumors in breast ultrasound (BUS) images. Yang et al. [29] introduced the Cross-task Guided Network (CTG-Net), which combines lesion segmentation and tumor classification in breast lesion analysis. Wu et al. [30] presented a novel approach to breast lesion segmentation in ultrasound images. Their method, named Boundary-Aware Multi-Resolution Network (BAMR-Net), leverages the Feature Pyramid Network (FPN) framework while incorporating new techniques for boundary guidance. This boundary guidance improves the network's ability to identify lesion borders, especially in unclear regions. Zhang et al. [31] developed BONet, a boundary Oriented network aimed at improving breast tumor segmentation, using a two-step process with a boundary-oriented module (BOM) and Atrous Spatial Pyramid Pooling (ASPP) for feature extraction. Chen et al. [32] introduced a deep learning method called AAU-net for segmenting breast lesions in ultrasound images. AAU-net is a variation of U-net that incorporates a novel module named the Hybrid-Adaptive-Attention Module (HAAM). This HAAM module combines two functionalities: channel-based attention and spatial attention. Hekal et al. [12] the Dual Decoder Attention ResUNet (DDA-AttResUNet), which features a more complex dual decoder attention structure designed to produce both segmentation images and enhanced ultrasound images.

Table 1 presents a summary of related work of breast cancer segmentation including method and their data. Despite past efforts using traditional techniques and early machine-learning approaches, segmenting breast lesions remains an evolving field. While Convolutional Neural Networks (CNNs) have demonstrated significant potential for breast tumor segmentation, certain limitations, such as generalizability or precision, warrant further exploration for optimal performance. Here's a breakdown of limitations in existing methods:

- Region-growing algorithms often struggle with the variability and noise present in breast ultrasound images, which can lead to inaccuracies in segmentation results.

- Graph-based approaches face challenges in extracting complex semantic features that are essential for accurately segmenting tumors in low-contrast ultrasound images.
- Deformable models provide flexibility but are sensitive to initialization, potentially resulting in suboptimal outcomes depending on the starting conditions.
- Handcrafted feature models rely on manually designed features, which may be insufficient for capturing the nuances necessary for high-fidelity segmentation.
- Deep learning methods, while capable of extracting rich information, still require further improvements to enhance segmentation accuracy.

TABLE 1

SUMMARY OF BREAST CANCER SEGMENTATION RESEARCH

Author(s)	**Methodology**	**Dataset Used**
Shan et al. [17]	Identified seed points for segmentation based on smooth contours and similar image intensity.	120 Breast Ultrasound (BUS) images
Madabhushi et al. [18]	Adapted starting template conforming to object boundaries using internal and external forces.	42 BUS images from 42 patients.
Chang et al. [19]	Utilized a stick filter for speckle noise reduction followed by a 3D discrete-active contour for lesion segmentation.	3-D ultrasound images.
Chiang et al.[20]	Integrated PBT output within energy function for breast lesion segmentation.	60 patients with malignant or benign breast lesions in BUS images.
Xian et al. [21]	Incorporated frequency and spatial domain information into the energy function for segmentation.	184 BUS images (93 benign and 91 malignant).
Liu et al. [22]	Extracted specific features from local image regions for classification.	112 BUS images (60 malignant and 52 benign).
Jiang et al. [23]	Trained classifiers using features extracted from ultrasound images.	112 BUS images (80 diseased and 32 normal).
Zhuang et al. [24]	Tailored architecture for breast tumor segmentation in BUS images.	1062 BUS images
Hu et al. [25]	Combined dilated FCN with a phase-based active contour model for segmentation.	570 BUS images from 89 patients and 128 BUS images (66 malignant and 62 benign).
Byra et al. [26]	Employed parametric maps for segmentation in ultrasound images.	269 BUS masses (123 malignant masses and 146 benign masses).
Zhu et al. [27]	Leveraged statistics from image subregions for segmenting breast lesions.	BUSI dataset (780 images categorized as normal, benign, malignant) and 632 BUS images.

Shareef et al. [28]	CNN designed for identifying small tumors in BUS images.	BUSIS dataset (562 images from multiple institutions), BUSI dataset (780 images categorized as normal, benign, malignant), UDIAT dataset (163 images).
Yang et al. [29]	Combined lesion segmentation with tumor classification in breast lesion analysis.	THH dataset (2718 images categorized as normal and lesion US images), UDIAT dataset, BUSI dataset.
Wu et al. [30]	Utilized Feature Pyramid Network (FPN) framework with boundary guidance for improved segmentation.	BUSI dataset (780 images categorized as normal, benign, malignant), BUSZPH dataset (632 images from Shenzhen People's Hospital).
Zhang et al. [31]	Developed a two-step process for tumor segmentation using a boundary-oriented module and Atrous Spatial Pyramid Pooling (ASPP).	BUSI dataset (780 images categorized as normal, benign, malignant), UDIAT dataset.
Chen et al. [32]	Introduced Hybrid-Adaptive-Attention Module (HAAM) for improved segmentation accuracy.	BUSI dataset, UDIAT dataset.
Hekal et al. [12]	Introduced the Dual Decoder Attention ResUNet (DDA-AttResUNet)	BUSI dataset (780 images categorized as normal, benign, malignant)

To address these challenges, the advanced deep learning model DAtt-ResUNet++ integrates attention mechanisms with the ResUNet++ architecture, significantly enhancing segmentation performance. The following subsection discusses related work to this model.

The DAtt-ResUNet++ model is built on two foundational components: residual learning and the integration of attention mechanisms. The interplay between these elements fosters a more robust and effective model for tasks involving medical image segmentation. The concept of residual learning, introduced by He et al. [33], tackles the challenge of diminishing performance with deeper networks during training. This approach, as highlighted by Zhang et al. [34] in their work on road profile reconstruction, allows for the creation of highly effective models even with many layers. Residual learning tackles the challenge of training very deep neural networks by introducing a new concept: skip connections. These connections allow the network to directly pass information from earlier layers to later layers, helping the model learn more complex features. This approach directly tackles a common hurdle in training deep neural networks: the vanishing-gradient problem [33]. Residual learning allows for the training of much deeper neural networks by enabling a more efficient gradient flow during backpropagation. This facilitates the network's ability to learn complex representations from the data. This smoother gradient flow is crucial for training deeper networks effectively. As a result, this approach enables the training of much deeper networks. This, in turn, facilitates stronger learning of image features and translates to improved accuracy in tasks like image segmentation. The introduction of residual learning stands as a major breakthrough, enabling the development of significantly deeper and more powerful neural networks. This concept has become a cornerstone of modern network architectures. ResUNet leverages residual units in a specific way. These units are placed before the activation function within the network [35].

Attention mechanisms are critical in augmenting CNN architectures by dynamically assigning importance weights to features. Self-attention, a powerful attention mechanism variant, can be seamlessly integrated as a modular component within existing CNN architectures. This strategic inclusion offers a computationally efficient approach, minimizing additional parameters required for the network to generate attention weights [36]. This modular extension introduces new building blocks for neural networks. These blocks can dynamically assign significance scores (importance weights) to features within an image (spatial dimensions) or across different feature channels [36].

Spatial attention mechanisms prioritize the significance of specific locations within an image, whereas channel attention assigns weights to different feature channels within a feature map [37]. Channel attention acts like a filter, highlighting important feature aspects within the data. It achieves this by assigning weights (importance scores) to each feature channel, forming a compact 1D vector [38]. There are two main approaches to integrating spatial and channel attention modules: parallel and sequential. In the parallel approach, both types of attention are applied simultaneously, with their outputs merged. For instance, Fu et al. [39] proposed the Dual Attention Network which utilizes this strategy. In contrast, the sequential approach applies spatial and channel attention one after the other. The Convolutional Block Attention Module (CBAM) by Woo et al. [40] exemplifies this sequential approach.

By integrating a novel dual attention mechanism, DAtt-ResUNet++ significantly improves upon the capabilities of the ResUNet++ architecture. By prioritizing informative regions, spatial attention empowers the model to effectively extract crucial semantic features. This enhanced feature extraction capability is particularly beneficial for accurately delineating faint tumor boundaries, especially in challenging ultrasound images marked by low contrast, artifacts, and inconsistencies. DAtt-ResUNet++ leverages a dual attention mechanism to create an attention map. This map acts as a spotlight, illuminating crucial regions within the encoded image features. The decoder capitalizes on this attention map, refining the semantic understanding of the encoded features. This enhanced semantic representation ultimately translates to improved segmentation accuracy.

This study delves deeper by comparing DAtt-ResUNet++ to relevant methodologies. Additionally, we employ an ablation study to conclusively demonstrate that incorporating the dual attention mechanism into the ResUNet++ model leads to superior segmentation performance.

3. Material and method

3.1. Dataset

The BUSI dataset offers a rich resource for researchers developing breast lesion segmentation algorithms. Its complexity presents a valuable challenge to further advance segmentation techniques in this domain. The dataset consists of 780 ultrasound images acquired from Baheya Hospital in Cairo, Egypt [41]. The BUSI dataset consists of ultrasound images in PNG format, with an average size of 500x500 pixels. The dataset is categorized into three classes: normal (containing 133 images), benign (containing 437 images), and malignant (containing 210 images). The dataset includes a corresponding binary ground truth mask for each image. These masks provide pixel-level annotations, precisely delineating tumor regions. The BUSI dataset offers a rich spectrum of ultrasound images featuring a variety of nodule sizes and shapes, making it a valuable resource for researchers in this field. An example of the dataset is depicted in Fig. 1.

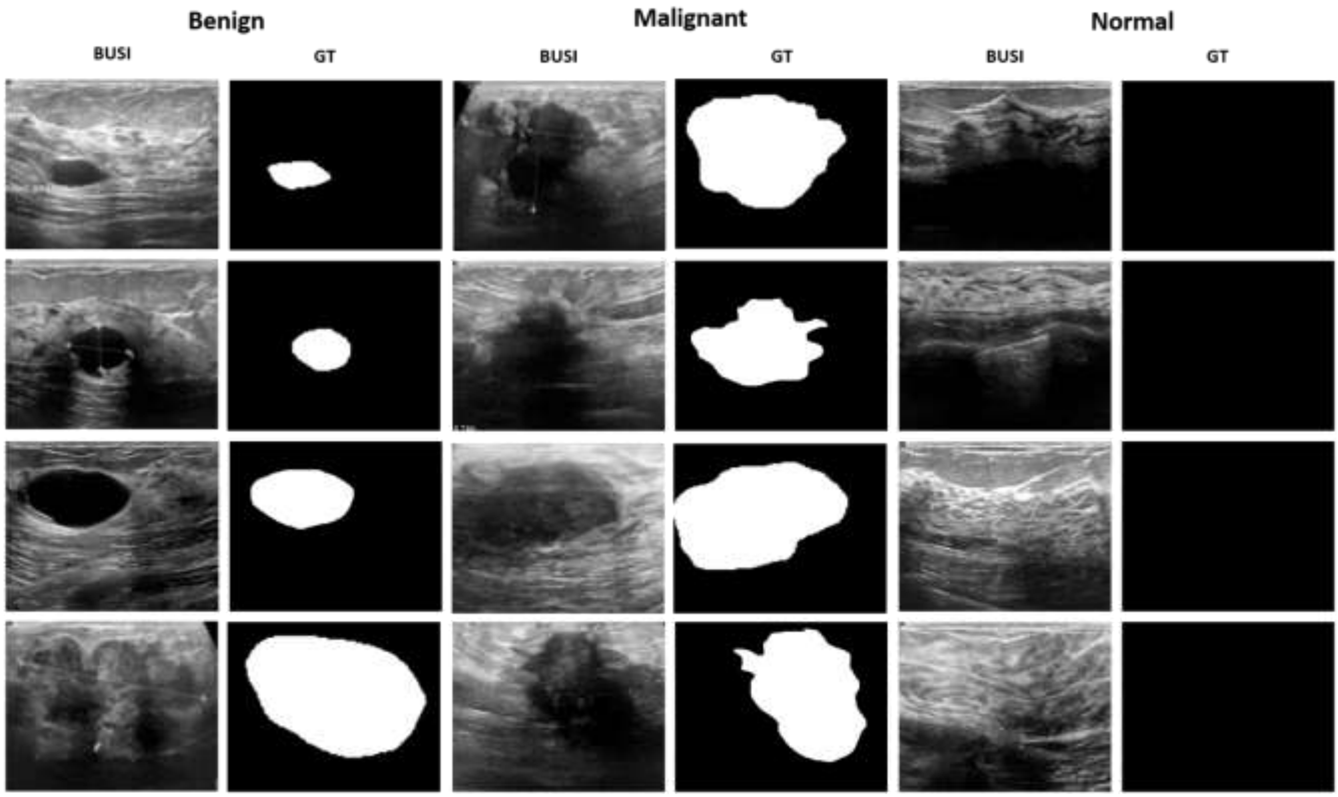

Fig. 1. Breast ultrasound (BUSI) images exhibit significant variability, encompassing normal, benign, and malignant tissues. Each BUSI image has a paired binary mask for accurate tumor segmentation.

3.2. Preprocessing

The Preprocessing prepares the training data for model training. This process involves three key steps:

- **Normalization**: Pixel intensities are scaled to a common range, often 0 to 1. This promotes smoother training and improves model generalizability across different datasets.
- **Resizing**: Images are resized to a uniform dimension of 128x128 pixels to conform to the model's input requirements, while resizing can lead to the loss of fine details, steps were taken to balance this with the model's ability to capture essential features. Mitigating potential information loss can be addressed through interpolation techniques and preserving the original aspect ratio during resizing.
- **Data Augmentation**: To diversify the training dataset and improve the model's performance and generalization capability, systematic data augmentation is employed. Techniques include flipping vertically and horizontally, as well as random rotations up to 90°. This strategy expands the training set by generating ten additional images for each original image.

3.3. Methodology

The intricate details of DAtt-ResUNet++, the novel encoder-decoder architecture, are unveiled in this section.

3.3.1. Overview of the Architecture

Central to this approach is the Dual-Attention-ResUNet++ architecture (DAtt-ResUNet++) (Fig.2).

Designed for breast tumor segmentation using ultrasound images (US), DAtt-ResUNet++ ingests a grayscale BUS image along with its segmentation mask as input. Following the input stage, the grayscale BUS image is fed into a ResUNet++ encoder network.

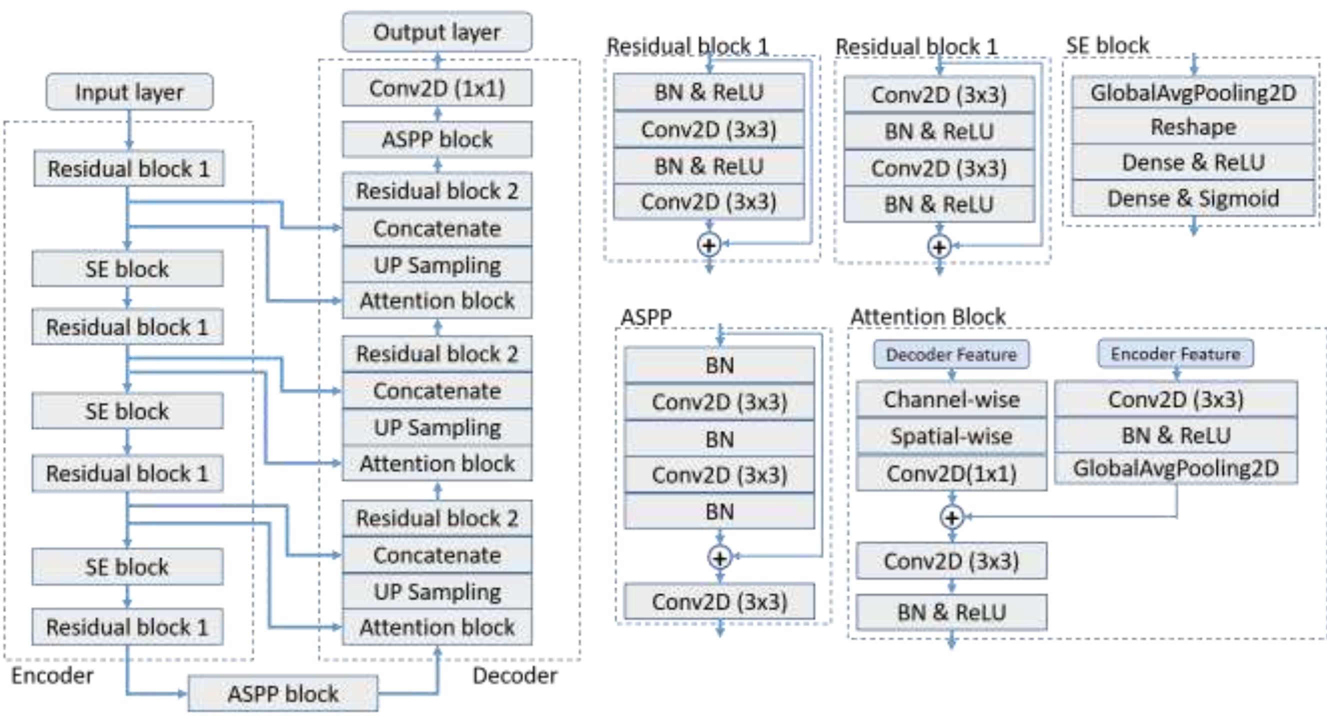

Fig. 2. Architecture of DAtt-ResUNet++

3.3.2. ResUNet++

As data is a grayscale image, the first layer is read The ResUNet++ architecture is an extension of the Deep Residual U-Net (ResUNet) [34], combining the benefits of deep residual learning [33]. and the U-Net model [42]. To achieve superior segmentation performance, the encoder leverages a synergistic combination of residual blocks, squeeze-and-excitation blocks (SE), Atrous Spatial Pyramid Pooling block (ASPP), and attention blocks. Residual blocks facilitate the training of deeper networks, SE blocks enhance feature importance, ASPP captures multi-scale information, and attention blocks selectively focus on crucial image regions. This combined strategy empowers DAtt-ResUNet++

to extract richer and more informative features from the input images.

3.3.3. Residual Units

Building upon the concept of residual learning established by He et al. [33] to address challenges with deep networks, ResUNet++ utilizes skip connections and residual blocks [43]. By incorporating residual connections, a cornerstone of residual learning [33], DAtt-ResUNet++ facilitates the training of significantly deeper neural networks. This enables the training of deeper networks, ultimately leading to enhanced feature representation and improved segmentation accuracy.

In the ResUNet++ architecture, complete residual units are implemented before activation functions [43]. Each encoder block consists of two consecutive 3x3 convolutional blocks followed by an identity mapping. Each convolutional block within the encoder consists of three key elements: batch normalization (BN), a ReLU activation layer, and a convolutional layer (Conv2D). The identity mapping, a core concept of residual learning, ensures the original input information is preserved as it flows through the block. Additionally, a strided convolution operation is employed to reduce the feature map size by half. This combination of techniques allows for deeper networks while maintaining informative features and facilitating efficient processing. The encoder block's output then passes through a squeeze-and-excitation block. The decoder path also incorporates residual units, and attention blocks are employed to enhance feature maps before up-sampling and concatenation with corresponding encoder path features. Finally, the decoder block output undergoes ASPP and a 1x1 convolution with sigmoid activation to generate the segmentation map. A detailed illustration of these components can be found in Fig.2.

3.3.4. Squeeze and Excitation Units

The squeeze-and-excitation network [38] enhances the network's feature representation by re-calibrating the feature responses through detailed modeling of channel inter-dependencies. This process involves two main steps:

- **Squeeze** (Global Information Embedding): Global average pooling is applied to each channel to generate channel-wise statistics.
- **Excitation** (Active Calibration): This step fully captures channel-wise dependencies network [38].

DAtt-ResUNet++ integrates SE blocks within its residual blocks. This inclusion strengthens feature recalibration, enhancing generalization across datasets.

3.3.5. Atrous Spatial Pyramid Pooling (ASPP)

ResUNet++ leverages Atrous Spatial Pyramid Pooling (ASPP) as a core component. This technique builds upon the concept of spatial pyramid pooling [44], which efficiently gathers features at various scales. ASPP captures multi-scale contextual information using parallel atrous convolutions with different dilation rates. DAtt-ResUNet++ leverages this capability by employing parallel atrous convolutions with varying dilation rates within the feature map. These dilation rates govern the field of view for each convolution, enabling the extraction of contextual information at different scales [45], [46]. Atrous convolutions [45] offer precise control over the field of view, allowing the model to capture multi-scale information effectively. As illustrated in Fig.2, ASPP acts as a bridge between the encoder and decoder in the architecture. The effectiveness of ASPP in capturing multi-scale information has been consistently demonstrated across a range of segmentation tasks. By integrating ASPP, our model is able to effectively gather essential features from multiple scales, significantly enhancing the accuracy and robustness of the semantic segmentation process.

3.3.6. Attention Mechanisms

Attention mechanisms, which initially gained traction in Natural Language Processing (NLP) [47], have also proven valuable in semantic segmentation tasks [48]. These mechanisms prioritize significant portions of the input data, enhancing feature quality and reducing computational demands. In ResUNet++, attention blocks are incorporated into the decoder to emphasize crucial areas in the feature maps, thereby improving segmentation precision.

Within the realm of deep learning for image processing, convolutional layers play a central role in extracting meaningful features, yet they might miss contextual details in the input feature maps. Pooling layers expand the influence of each output pixel by incorporating a wider area of input pixels. However, convolutional layers by themselves might not be sufficient to fully capture contextual information within an image. Attention mechanisms address this by summarizing input data and guiding the primary information flow within the network, acknowledging the varying importance of pixels or channels. Introducing attention mechanisms, like channel and spatial-wise attention, improves the model's understanding of segmented regions. This is achieved by integrating information from various sources within the feature maps [40]. This approach leverages deep learning and attention mechanisms to achieve remarkable results in breast tumor segmentation, surpassing previous methods in terms of accuracy and overall performance. To further enhance breast tumor segmentation performance, two novel attention modules inspired by the work of Zhao et al. [36] have been integrated into the model. The DAtt-ResUNet++ architecture incorporates both spatial and channel attention mechanisms, allowing the model to selectively focus on critical spatial regions and essential feature channels. This capability is particularly beneficial in ultrasound imaging, where a significant class imbalance is present—specifically, the number of tumor pixels is vastly smaller than that of background pixels. The channel attention mechanism prioritizes key feature channels, while the spatial attention mechanism directs focus to relevant spatial regions. This dual approach enhances the model's ability to manage the challenges associated with tumor pixel representation, leading to improved segmentation accuracy. By embedding these attention mechanisms within the decoder, DAtt-ResUNet++ is better equipped to target critical areas, resulting in superior segmentation outcomes.

3.3.7. Dual Attention Modules in ResUNet++

DAtt-ResUNet++ introduces two novel attention modules specifically designed to enhance feature representation and capture intricate contextual relationships within the data contributing to improved segmentation accuracy. These modules act as a spotlight, selectively focusing on informative regions within the feature maps. By emphasizing these crucial areas, the attention mechanisms empower the network to extract richer and more informative features, ultimately leading to improved segmentation accuracy. DAtt-ResUNet++ incorporates a channel-wise attention module specifically designed to prioritize informative channels within the decoder's input features. The attention module refines the initial features by assigning weights (significance scores) to each channel. This is achieved through a process where the initial features are multiplied with corresponding values from the attention module's output, on a channel-by-channel basis. This strategic filtering process effectively highlights the most relevant channels within the data, directing the decoder's focus toward crucial features for accurate segmentation.

Following the channel-wise attention module, DAtt-ResUNet++ incorporates a spatial-wise attention module. This module delves deeper, assigning weights not only to channels but also to individual pixels within the feature maps. This strategic weighting process empowers the network to pinpoint crucial spatial regions within the data, directing its focus towards areas most pertinent to accurate segmentation. The final feature map is generated by combining the channel-wise refined features with the output of the spatial attention module. This is achieved through element-wise multiplication.

The encoder's output undergoes a specific processing pipeline before being integrated with the attention modules. First, a 2D convolutional layer (3x3) followed by a ReLU activation function is applied. This is then followed by Global Average Pooling (GAP) 2D, which effectively summarizes the spatial information within the feature maps. The resulting output is then merged with the combined output from both the spatial-wise and channel-wise attention modules. This merged output serves as a rich representation that incorporates both high-level features extracted by the encoder and the crucial information identified by the attention mechanisms. Finally, the combined feature map undergoes further refinement through a final 2D (Conv2D) convolutional layer with a 3x3 kernel size. This layer incorporates ReLU activation and Batch Normalization (BN) to enhance the learned representation before it's passed on to the later decoder block.

Embedded within the decoder portion of the architecture, the attention unit process serves as the key mechanism for directing focus towards crucial areas of the feature maps. As illustrated in the block Fig.2, this process enables the model to selectively attend to these important areas, ultimately leading to improved segmentation performance.

The primary distinction between the architecture presented in this work and that of Hekal et al. [12] lies in the design of the decoder and encoder blocks, despite both methods utilizing the same Attention UNet structure. In Hekal et al.'s work, a dual decoder attention structure is implemented, where the Attention UNet is applied within the dual decoders to generate both segmentation masks and enhanced ultrasound images. The encoder in their model consists solely of residual blocks without additional mechanisms.

In contrast, the proposed model modifies the architecture by integrating residual blocks with Squeeze and Excitation (SE) units in the encoder, thereby enhancing feature extraction. Additionally, in the decoder, instead of employing a dual attention structure, the proposed method utilizes a combination of residual blocks and Atrous Spatial Pyramid Pooling (ASPP) blocks, which are also added in the bottleneck. This design captures a wider range of spatial information while enhancing feature maps, improving both segmentation accuracy and image quality, without increasing the complexity in the decoder, thus maintaining lower computational cost.

3.3.8. Performance metrics

Following the training phase on the designated dataset, the models are evaluated on unseen breast ultrasound images from the test dataset. This evaluation process involves predicting tumor segmentation masks and assessing their accuracy using various performance metrics. These metrics rely on correctly identified pixels categorized as true positive ($\boldsymbol{T_P}$), incorrectly identified pixels as false positive ($\boldsymbol{F_P}$), correctly identified negative pixels (background) as true negative ($\boldsymbol{T_N}$), and incorrectly identified negative pixels as false negative ($\boldsymbol{F_N}$).

- Accuracy (**Acc**) reflects the proportion of correctly classified pixels in the predicted segmentation, compared to the ground truth. It's calculated as the proportion of pixels that are correctly classified (both tumor and background) divided by the total number of pixels in the entire mask. This metric provides a general idea of how well the model performed in segmenting the tumor region.

$$\textbf{Acc} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (1)$$

- The Dice coefficient (**Dice**) is a metric used to assess the similarity between the predicted segmentation mask and the ground truth mask. It achieves this by calculating twice the area of overlap between the two masks divided by their total combined area. This provides a measure of how well the model's predictions agree with the actual tumor regions.

$$\textbf{Dice} = \frac{2T_P}{2T_P + F_P + F_N} \quad (2)$$

- Jaccard index, also known as the Intersection over Union (**IoU**), quantifies the degree of overlap between the predicted segmentation mask and the ground truth. It calculates the proportion of pixels belonging to both the predicted tumor area and the actual tumor area, divided by the total number of pixels encompassing either the predicted tumor or the actual tumor. This metric provides a comprehensive measure of segmentation accuracy.

$$\textbf{IOU} = \frac{T_P}{T_P + F_P + F_N} \quad (3)$$

- Sensitivity (**Sen**), also known as Recall or True Positive Rate, evaluates the model's capability to correctly identify actual tumor pixels. It calculates the proportion of true positive instances (correctly identified tumor pixels) divided by the total number of actual tumor pixels in the image. This metric essentially tells us how well the model avoids missing true tumor regions.

$$\textbf{Sen} = \frac{T_P}{T_P + F_N} \quad (4)$$

- Precision (**Pre**) quantifies the model's ability to avoid false positives. It calculates the proportion of pixels identified as tumor pixels by the model that actually correspond to tumors in the ground truth mask. This metric helps us understand how accurate the model is in identifying tumor regions and avoiding mistakenly classifying background pixels as tumors.

$$\textbf{Pre} = \frac{T_P}{T_P + F_N} \quad (5)$$

Leveraging these evaluation metrics enables a comprehensive assessment of the quality and effectiveness of the proposed model's tumor segmentation predictions. This assessment provides valuable insights into the model's strengths and weaknesses in identifying tumor regions within breast ultrasound images.

4. Results

4.1. Experimental Setting

This study leverages a publicly available medical ultrasound image dataset (BUSI) for breast cancer analysis. To ensure compatibility with the segmentation models, a preprocessing step is implemented to address image size variations. All images undergo standardization through normalization and resizing to a uniform dimension of 128x128 pixels.

To ensure a reliable and impartial assessment of the model's generalization capabilities, we utilize a 10-fold cross-validation approach. The dataset is randomly divided into 10 folds. Within each fold, 90% (648 images) is allocated for training and validation purposes, while the remaining 10% (72 images) serves as the unseen test set. The training and validation set is further split, with 90% dedicated to training (incorporating data augmentation) and 10% reserved for validation. The validation set is crucial for hyperparameter tuning and model selection. The performance metrics reported in our study represent the average values obtained from the test sets across all 10 folds of the cross-validation process.

To facilitate efficient training, we utilize a high-performance graphics processing unit (NVIDIA Tesla P100 GPU). The model's learning process is guided by the Adam optimization algorithm, starting with an initial learning rate of 0.0001. Training is conducted over 150 epochs, with the learning rate halved when the model's performance levels off. This adjustment is implemented to refine the model's performance. Additionally, to prevent overfitting, we incorporate early stopping. To prevent overfitting and enhance generalizability, DAtt-ResUNet++ employs a technique that strategically halts the training process when the validation error stops improving. This approach safeguards against the model memorizing training data, ultimately leading to better performance on unseen images.

4.2. Qualitative Results

In Fig.3, a comparative visualization is provided to illustrate the segmentation results achieved by different models. The original breast image is depicted in the input ultrasound image. The corresponding ground truth segmentation mask, meticulously delineated by medical experts, precisely identifies the tumor regions within the image. The predicted segmentation masks are generated by various segmentation approaches, including our proposed DAtt-ResUNet++ model. This visualization allows for a qualitative assessment of each model's capability to precisely identify tumor regions within the ultrasound images.

In Fig. 3, a visual comparison of predicted segmentation masks with ground truth masks allows for a clear evaluation of the DAtt-ResUNet++ model's performance. This comparison reveals the model's proficiency in generating accurate and well-defined tumor segmentations. However, challenges are observed in accurately segmenting small tumors in both benign and malignant cases, as illustrated in raw 2 and 3 of the Fig.3. Despite this limitation, the proposed Dual-Attention-ResUNet++ architecture demonstrates its value in the domain of breast ultrasound image segmentation, enabling precise tumor identification and delineation. The model's success stems from its utilization of advanced deep learning techniques, strategically incorporated attention mechanisms, and effective fusion operations. This combination of elements significantly enhances segmentation performance and yields superior results.

To provide a comprehensive analysis of the impact of attention mechanisms on the segmentation process, visualizations of the attention maps generated by the proposed DAtt-ResUNet++ model are presented. These visualizations offer a clearer understanding of how attention mechanisms contribute to feature enhancement and improved segmentation accuracy. Fig. 4 displays a series of images for comparison. The first column contains the original Breast Ultrasound (BUSI) images. The second column presents the corresponding ground truth segmentation masks, serving as a benchmark. The third column shows the attention maps generated by DAtt-ResUNet++, illustrating the regions of the ultrasound images where the model concentrates to improve feature extraction. The fourth column depicts the predicted segmentation masks produced by DAtt-ResUNet++. These attention maps provide valuable insights into the model’s ability to capture critical spatial features, which in turn leads to more precise segmentation. The visualizations demonstrate the role of attention mechanisms in enhancing the performance of DAtt-ResUNet++ for ultrasound image segmentation tasks.

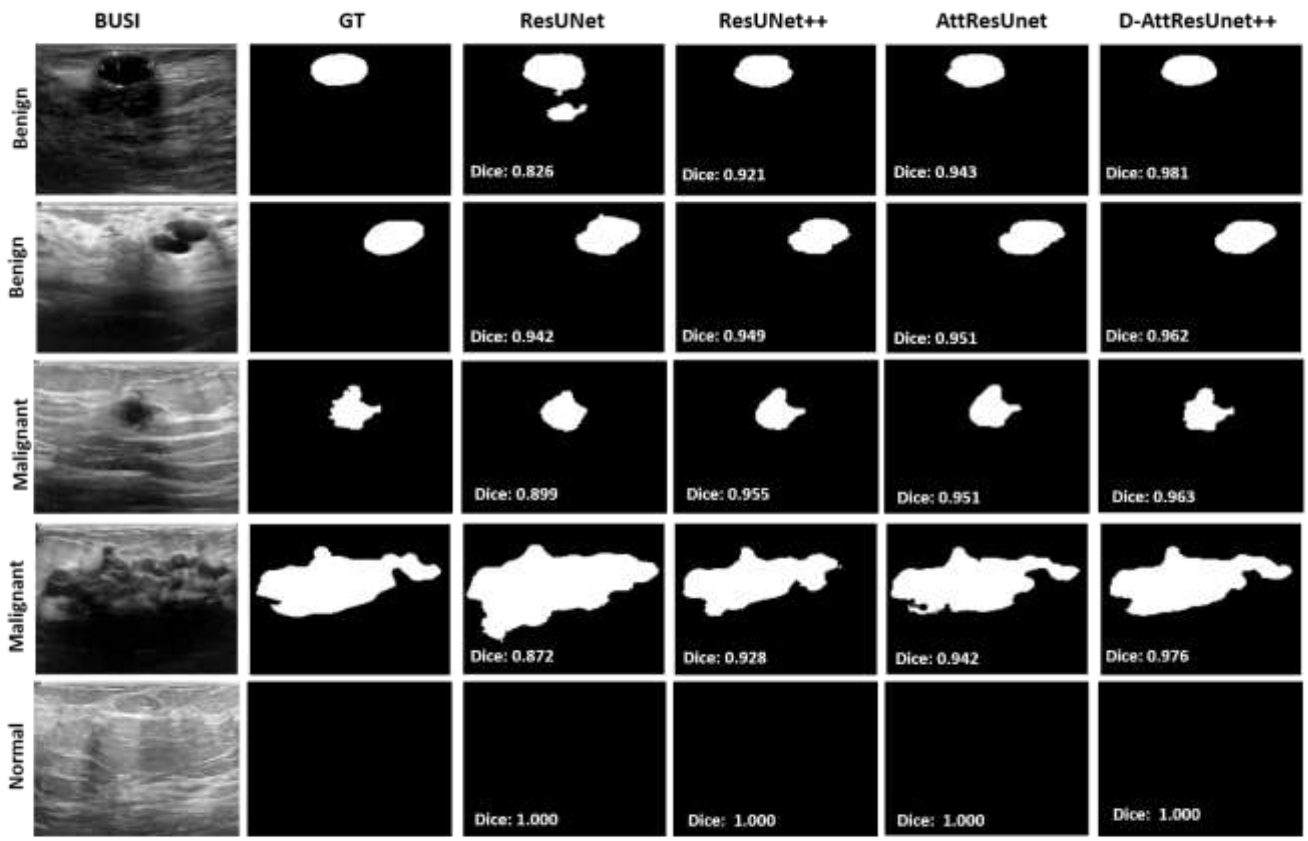

Fig. 3 To offer a visual assessment of the segmentation performance across different models, qualitative results are presented. The original ultrasound images are showcased in the first column. Corresponding ground truth masks, which precisely identify tumor regions, are displayed in the second column. The segmentation masks predicted by various architectures—ResUNet, ResUNet++, Att-ResUNet, and our proposed DAtt-ResUNet++ are shown in the following columns. To provide a quantitative measure of accuracy alongside the visual representation, the Dice score is included below each predicted mask. The Dice score serves as a metric to quantify the similarity between the predicted segmentation and the ground truth mask.

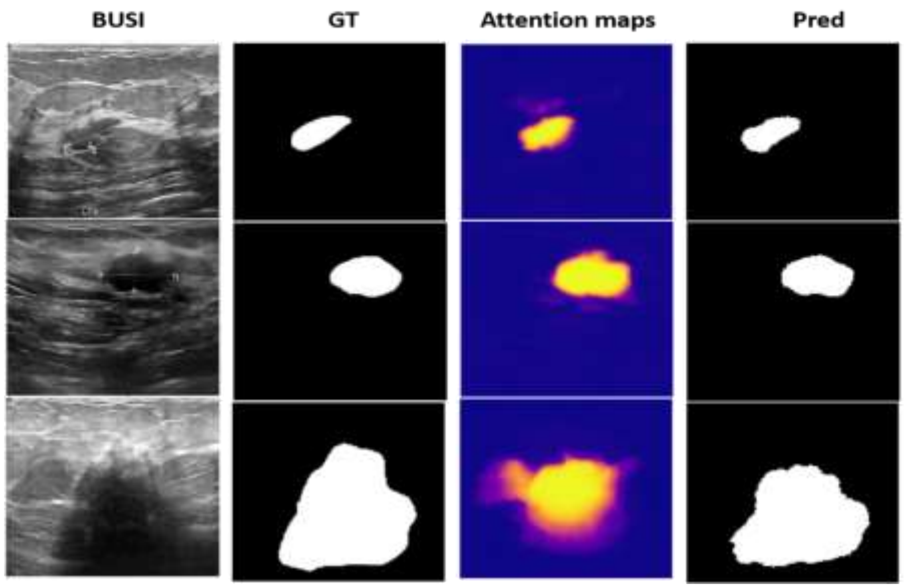

Fig. 4 Visualizing Attention Maps to Illustrate Segmentation Process Improvement

TABLE 2

PERFORMANCE COMPARISON OF DATT-RESUNET++ AND ABLATION VARIANTS ON **BUSI** DATASET (MEAN ± STD) (10-FOLD CROSS-VALIDATION)

Method	**Dice [%]**	**Acc [%]**	**IOU [%]**	**Pre [%]**	**Sen [%]**
UNet	80.64 ± 1.88	96.22 ±1.72	73.23 ±2.32	91.04 ±2.42	72.88 ± 1.11
Vgg16-UNet	80.01 ± 2.31	96.81±1.22	77.12 ± 1.32	83.84 ± 1.2	85.14 ± 2.1
MobileNetV2-UNet	65.12 ± 1.88	92.21 ± 1.22	70.22 ± 1.47	79.22 ± 1.61	80.33 ± 1.22
ResUnet	83.80 ± 1.22	97.54 ± 0.16	74.36 ± 1.68	85.99 ± 0.96	84.15 ± 1.66
ResUNet++	85.00 ± 0.97	97.57 ± 0.14	75.29 ± 1.33	86.23 ± 1.02	84.25 ± 0.88
AttResUnet	89.10 ± 1.03	98.31 ± 0.14	81.87 ± 1.49	92.39 ± 0.69	86.42 ± 1.27
DAtt-ResUNet++	**90.40** ± 0.88	**98.73** ± 0.12	**84.62** ±1.12	**93.44** ± 0.56	**89.82** ± 0.75

TABLE 3

PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS ON **BUS**I DATASET

Authors	**Method**	**Split Dataset**	**Dice [%]**	**Acc [%]**	**IOU [%]**	**Pre [%]**	**Sen [%]**
Zhu et al. [27]	Second-Order Subregion Pooling network	Five-folder cross-validation	84.70	NA	76.39	87.62	85.51
Wu et al. [30]	A boundary-guided multiscale network	661 img train 119 img test	83.97	NA	75.97	89.31	83.45
Chen et al. [32]	An adaptive attention	Four-fold cross-validation	77.21	NA	67.97	78.66	80.63

Shareef et al. [28]	Enhanced Small Tumor-Aware Network	Five-folder cross-validation	80.00	NA	72.00	NA	85.00
Zhang et al. [31]	A boundary-oriented network	Four-fold cross-validation	80.27	NA	NA	85.24	80.42
Chavan et al. [49]	A UNET architecture with involution layers	517 img train 130 img test	80.09	96.09	71.46	83.15	82.77
Yang et al. [29]	Cross-task guided network	80% training 20% test	82.00	NA	74.00	NA	84.00
Lyu et al. [50]	Pyramid Attention Network	85% training 15% test	80.71	97.13	68.53	83.50	79.30
Zhang et al. [51]	Classification and Segmentation branches network.	Five-folder cross-validation	86.34	NA	83.12	88.32	89.51
Zhang et al. [52]	SaTransformer, a semantic-aware model.	Ten-fold cross-validation	89.80	NA	79.10	NA	85.90
Umer et al. [53]	A Dual-Decoded attention mechanism.	70% training 30% test	88.68	NA	80.12	**95.42**	87.79
Proposed DAtt-ResUnet++	**Dual Attention ResUnet++**	Ten-fold cross-validation	**90.40**	**98.73**	**84.62**	93.44	**89.82**

4.3. Quantitative Results

The performance comparison of various UNet-based architectures, including DAtt-ResUNet++, UNet (standard architecture), VGG16-UNet (leveraging VGG16 as a backbone), MobileNetV2-UNet (with MobileNetV2 backbone for lightweight modeling), ResUNet (with no attention), ResUNet++, AttResUNet (incorporating attention with ResNet), and DAtt-ResUNet++, is presented in Table 2**.** To isolate the influence of each component within the DAtt-ResUNet++ architecture, ablation studies were performed. This involved evaluating simplified variations of the model, such as ResNet (with no attention), ResNet++, Att-ResUNet (incorporating attention with ResNet), and the final DAtt-ResUNet++ (Daul attention with ResNet++). As shown in Table 2, this analysis provided valuable insights into the contributions of each element to the overall segmentation performance. The evaluation metrics employed in this analysis encompassed the Dice Coefficient, overall Accuracy, Intersection over Union (IoU), Sensitivity, and Precision, offering a comprehensive assessment of model performance.

The proposed DAtt-ResUNet++ model is evaluated alongside several contemporary segmentation approaches, all assessed on the BUSI dataset. This comparative analysis, presented in Table 3, offers crucial information about the effectiveness of various methods for breast tumor segmentation in ultrasound (US) images. The analysis incorporates noteworthy methods from recent publications, including those by Zhu et al. [27], Wu et al. [30], Shareef et al. [28], Zhang et al. [31], Chavan et al. [49], Yang et al. [29], Zhang et al. [51], Zhang et al. [52], Lyu et al. [50], and Umer et al. [53]. By comparing DAtt-ResUNet++ to these established methods, we gain a comprehensive understanding of its effectiveness in tumor segmentation tasks.

The DAtt-ResUNet++ model emerges as a frontrunner, achieving exceptional scores across several crucial performance metrics. DAtt-ResUNet++ achieves an impressive Dice coefficient of 90.40%, indicating a strong concordance between its predicted tumor segmentations and the ground truth masks. In addition, DAtt-ResUNet++ exhibits a Sensitivity (Sen) of 89.82%, indicating its proficiency in precisely identifying true tumor pixels. Furthermore, the model achieves a remarkable accuracy of 98.73%, highlighting its overall efficacy in correctly classifying both tumor and background pixels within the images. These remarkable results emphasize the potential of DAtt-ResUNet++ for breast tumor segmentation tasks.

While DAtt-ResUNet++ demonstrates exceptional performance in several metrics, a closer look reveals a trade-off between certain evaluation criteria. The model achieves an Intersection over Union (IoU) of 84.62%, indicating the proportion of correctly segmented tumor pixels relative to the combined area of predicted and ground truth tumor regions. Additionally, its precision of 93.44% reflects the accuracy of its positive predictions, meaning a low rate of false positives (non-tumor pixels identified as tumors). It's worth noting that the method by Umer et al. [53] achieves a slightly higher precision of 95.42%, suggesting a potential advantage in this specific aspect.

The remarkable performance of DAtt-ResUNet++ stems from its innovative architecture, which strategically integrates the Dual-Attention mechanism with the well-established ResUNet++ segmentation model. This fusion proves to be highly effective in the domain of breast tumor segmentation within medical ultrasound images. DAtt-ResUNet++ leverages advanced deep learning techniques alongside a unique combination of dual decoder attention and attention mechanisms. This combination empowers the model to achieve exceptional accuracy and robustness in segmentation outcomes.

One noteworthy feature is the incorporation of a dual decoding fusion attention mechanism, which plays a significant role in considerably enhancing segmentation precision. This specific element, along with the overall design choices within DAtt-ResUNet++, sets it apart from other approaches and contributes to its superior segmentation results.

TABLE 4

Analysis of time performance time of the proposed system

	UNet	Vgg16-UNet	MobileNetV2-UNet	ResUnet	ResUnet++	AttResUnet	DAtt-ResUNet++
Training (hr/data)	5	5.2	5.1	6.15	6.8	8.1	9.2
Testing (ms/img)	11.2 ±3.2	12.1 ±4.4	12.5 ±2.5	13.3 ± 1.6	12.6 ± 3.2	16.3 ±2.2	17.4 ± 4.9

5. Discussion

Breast tumor segmentation in breast ultrasound (BUS) images presents a significant challenge. The inherent complexity and ambiguity of BUS data make accurate identification and delineation of tumors difficult. This has led to the exploration of computer-aided diagnosis (CAD) systems to augment and potentially improve upon subjective manual diagnosis based on BUS images.

This study proposes DAtt-ResUNet++, a novel deep-learning architecture for segmenting breast tumors in ultrasound images. Evaluated on the BUSI dataset, DAtt-ResUNet++ achieves impressive results, including a Dice score of 90.40± 0.88%, indicating a high level of agreement between predicted and ground truth segmentations. Furthermore, the model demonstrates strong performance across other metrics, such as Intersection over Union (IoU) at 84.62±1.12%, sensitivity at 89.82± 0.75%, precision at 93.44 ± 0.56 and overall accuracy at 98.73± 0.12%. Evaluation against existing deep learning techniques highlights the competitiveness of DAtt-ResUNet++. These findings suggest that DAtt-ResUNet++ has the potential to be a valuable tool for reliable breast tumor segmentation in clinical settings, potentially aiding in earlier diagnosis and improved patient outcomes.

Several recent studies have explored deep learning techniques for breast tumor segmentation in challenging BUS images. These investigations include works by Zhang et al. [52] on SA Transformer and Fully Convolutional approaches, Lyu et al. [50] on AMS, Umer et al. [53]on breast cancer segmentation, Shareef et al. [28] on ESTAN, Zhang et al. [51] on Boundary detection, and Yang et al. [29] on CTG. Notably, Zhang et al. [52] achieved a previously reported highest Dice score of 89.80%. The proposed DAtt-ResUNet++ surpasses this performance by achieving a Dice score of 90.40%. These findings demonstrate the promise of DAtt-ResUNet++ for improved breast tumor segmentation in BUS images.

A limitation of DAtt-ResUNet++ is that, while it demonstrates state-of-the-art performance across various metrics, its precision at 93.44% falls short compared to the 95.42% achieved by [53]. Additionally, the proposed method is somewhat weaker in segmenting small tumors in both benign and malignant cases (Fig. 3). This indicates a potential area for improvement, suggesting that further optimization of DAtt-ResUNet++ could enhance its precision while maintaining its high segmentation accuracy, as reflected by the Dice score. To provide a more detailed comparison of performance, computational cost, and inference time with existing state-of-the-art models for breast tumor segmentation, it is important to note that all results presented in this paper were obtained using a standard laptop equipped with an Intel Core i5-6200U @ 2.30 GHz processor and 8 GB of RAM. Additionally, a high-performance GPU, accessed through Colab Pro, was required for proposed model training to improve processing speed, because of the complexity of the DAtt-ResUNet++ model. This highlights the model's computational demands while showcasing its effectiveness in achieving higher accuracy. Furthermore, despite the increased accuracy, the developed model does not require any additional hardware upgrades in medical machines, ensuring its practicality for clinical implementation without the need for costly infrastructure changes. As illustrated in Table 4, the training time was calculated in hours for all 10-fold cross-validation, while the testing time was measured in milliseconds per image. DAtt-ResUNet++ required the longest time for both training and testing each image when compared to other models, attributed to its increased number of parameters. This increase in time presents a limitation, despite the model's superior Dice score and accuracy, as demonstrated in Table 2. The potential integration of DAtt-ResUNet++ into real-world clinical workflows will be explored by focusing on optimizing computational efficiency, improving model interpretability, and ensuring ease of use, all while maintaining accuracy for seamless clinical deployment, and this consideration will guide future work.

Overall, the incorporation of the dual attention mechanism in DAtt-ResUNet++ demonstrates its effectiveness in breast tumor segmentation. DAtt-ResUNet++ achieves superior Dice scores compared to existing models, as evidenced by both quantitative and qualitative results (refer to Fig.2 for visual examples). This study highlights the potential of DAtt-ResUNet++ as a valuable tool for computer-aided diagnosis systems in breast cancer detection using ultrasound images.

6. Conclusions

This study introduces DAtt-ResUNet++, a novel deep-learning architecture for segmenting breast tumors in ultrasound (BUS) images. DAtt-ResUNet++ integrates a dual attention mechanism with the ResUNet++ model, achieving a remarkable Dice score of 90.40% on the BUSI dataset. BUS images offer a non-invasive, safe, cost-effective, and radiation-free alternative for cancer screening. The core strength of DAtt-ResUNet++ lies in its dual attention mechanism, which refines feature maps by emphasizing critical regions. This targeted focus leads to superior segmentation performance, particularly evident in the Dice coefficient.

To ensure the generalizability of DAtt-ResUNet++, future endeavors will involve testing on additional BUS datasets and exploring further refinements to the model. Continued advancements in breast tumor segmentation using BUS images hold significant promise for improving the precision of oncological diagnostics and therapeutic planning methodologies. In future work, methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) Gharaibeh et al. [54] and additional attention map visualizations will be employed to further enhance the interpretability, ensuring that relevant regions are effectively targeted by the attention mechanism during segmentation, which is crucial for clinical adoption. ROC analysis will also be incorporated to enhance the evaluation framework of the approach. Furthermore, experimentation will be extended to include larger and more diverse datasets, incorporating multi-institutional sources to better validate the model's generalizability and performance across varied patient populations. Additionally, parallel computing will be employed to accelerate the training process and efficiently manage the increased computational demands of larger datasets.

References

1. Islami, F., et al., *National and state estimates of lost earnings from cancer deaths in the United States*. JAMA oncology, 2019. **5**(9): p. e191460-e191460.
2. Organization, W.H., *Global breast cancer initiative implementation framework: assessing, strengthening and scaling-up of services for the early detection and management of breast cancer*. 2023: World Health Organization.
3. Skaane, P., et al., *Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: the Oslo Tomosynthesis Screening Trial*. Radiology, 2019. **291**(1): p. 23-30.
4. Carlino, G., et al., *Ultrasound-guided preoperative localization of breast lesions: a good choice*. Journal of Ultrasound, 2019. **22**: p. 85-94.
5. Askaripour, K. and A. Zak, *Breast MRI Segmentation by Deep Learning: Key Gaps and Challenges*. IEEE Access, 2023. **11**: p. 117935-117946.
6. Zhou, Y., et al., *Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images*. Medical Image Analysis, 2021. **70**: p. 101918.
7. Drukker, K., et al., *Breast US computer-aided diagnosis workstation: performance with a large clinical diagnostic population*. Radiology, 2008. **248**(2): p. 392-397.
8. Yap, M.H., E. Edirisinghe, and H. Bez, *Processed images in human perception: A case study in ultrasound breast imaging*. European journal of radiology, 2010. **73**(3): p. 682-687.

9. Amin, J., et al., *Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model*. Microscopy Research and Technique, 2022. **85**(5): p. 1926-1936.
10. Shah, J.H., et al., *Facial expressions classification and false label reduction using LDA and threefold SVM*. Pattern Recognition Letters, 2020. **139**: p. 166-173.
11. Sharif, M., et al., *A framework for offline signature verification system: Best features selection approach*. Pattern Recognition Letters, 2020. **139**: p. 50-59.
12. Hekal, A.A., et al., *Breast cancer segmentation from ultrasound images using deep dual-decoder technology with attention network*. IEEE Access, 2024.
13. Umer, M.J., et al., *An integrated framework for COVID-19 classification based on classical and quantum transfer learning from a chest radiograph*. Concurrency and Computation: Practice and Experience, 2022. **34**(20): p. e6434.
14. Umer, M.J., et al., *Multi-class classification of breast cancer using 6b-net with deep feature fusion and selection method*. Journal of Personalized Medicine, 2022. **12**(5): p. 683.
15. Oktay, O., et al., *Attention u-net: Learning where to look for the pancreas*. arXiv preprint arXiv:1804.03999, 2018.
16. Jha, D., et al. *Resunet++: An advanced architecture for medical image segmentation*. in *2019 IEEE international symposium on multimedia (ISM)*. 2019. IEEE.
17. Shan, J., H. Cheng, and Y. Wang, *Completely automated segmentation approach for breast ultrasound images using multiple-domain features*. Ultrasound in medicine & biology, 2012. **38**(2): p. 262-275.
18. Madabhushi, A. and D.N. Metaxas, *Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions*. IEEE transactions on medical imaging, 2003. **22**(2): p. 155-169.
19. Chang, R.-F., et al., *Segmentation of breast tumor in three-dimensional ultrasound images using three-dimensional discrete active contour model*. Ultrasound in medicine & biology, 2003. **29**(11): p. 1571-1581.
20. Chiang, H.-H., et al. *Cell-based graph cut for segmentation of 2D/3D sonographic breast images*. in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2010. IEEE.
21. Xian, M., Y. Zhang, and H.-D. Cheng, *Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains*. Pattern Recognition, 2015. **48**(2): p. 485-497.
22. Liu, B., et al., *Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images*. Pattern Recognition, 2010. **43**(1): p. 280-298.
23. Jiang, P., et al. *Learning-based automatic breast tumor detection and segmentation in ultrasound images*. in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. 2012. IEEE.
24. Zhuang, Z., et al., *An RDAU-NET model for lesion segmentation in breast ultrasound images*. PloS one, 2019. **14**(8): p. e0221535.
25. Hu, Y., et al., *Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model*. Medical physics, 2019. **46**(1): p. 215-228.
26. Byra, M., et al., *Breast mass segmentation based on ultrasonic entropy maps and attention gated U-Net*. arXiv preprint arXiv:2001.10061, 2020.
27. Zhu, L., et al. *A second-order subregion pooling network for breast lesion segmentation in ultrasound*. in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International*

Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. 2020. Springer.

28. Shareef, B., et al. *Estan: Enhanced small tumor-aware network for breast ultrasound image segmentation*. in *Healthcare*. 2022. MDPI.
29. Yang, K., et al., *CTG-Net: Cross-task guided network for breast ultrasound diagnosis*. PloS one, 2022. **17**(8): p. e0271106.
30. Wu, Y., et al., *BGM-Net: boundary-guided multiscale network for breast lesion segmentation in ultrasound*. Frontiers in Molecular Biosciences, 2021. **8**: p. 698334.
31. Zhang, M., et al., *Boundary-oriented network for automatic breast tumor segmentation in ultrasound images*. Ultrasonic Imaging, 2023. **45**(2): p. 62-73.
32. Chen, G., et al., *AAU-net: an adaptive attention U-net for breast lesions segmentation in ultrasound images*. IEEE Transactions on Medical Imaging, 2022. **42**(5): p. 1289-1300.
33. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
34. Zhang, Z., Q. Liu, and Y. Wang, *Road extraction by deep residual u-net*. IEEE Geoscience and Remote Sensing Letters, 2018. **15**(5): p. 749-753.
35. Diakogiannis, F.I., et al., *ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data*. ISPRS Journal of Photogrammetry and Remote Sensing, 2020. **162**: p. 94-114.
36. Zhao, P., et al., *SCAU-net: spatial-channel attention U-net for gland segmentation*. Frontiers in Bioengineering and Biotechnology, 2020. **8**: p. 670.
37. Jaderberg, M., K. Simonyan, and A. Zisserman, *Spatial transformer networks*. Advances in neural information processing systems, 2015. **28**.
38. Hu, J., L. Shen, and G. Sun. *Squeeze-and-excitation networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
39. Fu, J., et al. *Dual attention network for scene segmentation*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
40. Woo, S., et al. *Cbam: Convolutional block attention module*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.
41. Al-Dhabyani, W., et al., *Dataset of breast ultrasound images*. Data in brief, 2020. **28**: p. 104863.
42. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. 2015. Springer.
43. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
44. He, K., et al., *Spatial pyramid pooling in deep convolutional networks for visual recognition*. IEEE transactions on pattern analysis and machine intelligence, 2015. **37**(9): p. 1904-1916.
45. Chen, L.-C., et al., *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*. IEEE transactions on pattern analysis and machine intelligence, 2017. **40**(4): p. 834-848.
46. Chen, L.-C., *Rethinking atrous convolution for semantic image segmentation*. arXiv preprint arXiv:1706.05587, 2017.

47. Vaswani, A., *Attention is all you need*. Advances in Neural Information Processing Systems, 2017.
48. Li, H., et al., *Pyramid attention network for semantic segmentation.* arXiv preprint arXiv:1805.10180, 2018.
49. Chavan, T., K. Prajapati, and K.R. JV. *InvUNET: Involuted UNET for breast tumor segmentation from ultrasound*. in *International Conference on Artificial Intelligence in Medicine*. 2022. Springer.
50. Lyu, Y., et al., *AMS-PAN: Breast ultrasound image segmentation model combining attention mechanism and multi-scale features*. Biomedical Signal Processing and Control, 2023. **81**: p. 104425.
51. Zhang, J., et al., *SaTransformer: Semantic-aware transformer for breast cancer classification and segmentation*. IET Image Processing, 2023. **17**(13): p. 3789-3800.
52. Zhang, S., et al., *Fully automatic tumor segmentation of breast ultrasound images with deep learning*. Journal of Applied Clinical Medical Physics, 2023. **24**(1): p. e13863.
53. Umer, M.J., M. Sharif, and S.H. Wang, *Breast cancer classification and segmentation framework using multiscale CNN and U-shaped dual decoded attention network*. Expert Systems, 2022: p. e13192.
54. Gharaibeh, H., et al., *Improved Breast Cancer Detection using Modified ResNet50-Based on Gradient-Weighted Class Activation Mapping*. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2024: p. 190-208.