

## RESEARCH ARTICLE

# Breast Cancer Segmentation From Ultrasound Images Using Deep Dual-Decoder Technology With Attention Network

ASMAA A. HEKAL<sup>1</sup>, (Member, IEEE), AHMED ELNAKIB<sup>1,2</sup>, (Senior Member, IEEE), HOSSAM EL-DIN MOUSTAFA<sup>1</sup>, (Senior Member, IEEE), AND HANAN M. AMER<sup>1</sup>

<sup>1</sup>Electronics and Communications Engineering (ECE) Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

<sup>2</sup>Electrical and Computer Engineering Department, School of Engineering, Penn State Erie, The Behrend College, Erie, PA 16563, USA

Corresponding author: Ahmed Elnakib (ahmed.elnakib@psu.edu)

**ABSTRACT** This paper introduces a deep learning approach for breast cancer segmentation from ultrasound imaging using a Dual Decoder Attention ResUNet (DDA-AttResUNet). DDA-AttResUNet utilizes a Dual Decoder Attention structure to simultaneously focus on tumor segmentation while also capturing supplementary contextual information, leading to enhanced segmentation accuracy. An Attention mechanism is incorporated to enhance the representation of segmented regions by effectively combining information from multiple sources. The model's performance is validated on a public challenging dataset of 780 Breast Ultrasound Images (BUSI), achieving a Dice similarity coefficient of  $92.92 \pm 0.69\%$ , Intersection over Union of  $87.39 \pm 1.10\%$ , Sensitivity of  $92.16 \pm 0.92\%$ , Precision of  $93.90 \pm 0.40\%$ , and Accuracy of  $98.82 \pm 0.10\%$ , using 10-fold cross-validation. These results, comparable to other leading methods, indicate that our DDA-AttResUNet can significantly advance breast tumor segmentation in BUS imaging, with implications for improved diagnosis and patient outcomes.

**INDEX TERMS** Breast ultrasound, tumor segmentation, deep learning, image segmentation, image detection.

## I. INTRODUCTION

Breast cancer represents a widespread disease and constitutes a foremost contributor to mortality among females on a global scale, as documented by the World Health Organization (WHO) [1]. In 2020, as of the last available update, there were 2.3 million reported cases and 685,000 deaths globally due to breast cancer [2]. In anticipation of the year 2040, predictive modeling forecasts a forthcoming rise, with an estimated annual incidence of over 3 million reported cases of breast cancer and an approximate annual mortality rate of 1 million worldwide [3]. Diverse imaging modalities, including mammography, computed tomography (CT), magnetic resonance imaging (MRI), and breast ultrasound (BUS), are pivotal in the context of breast cancer detection. Mammography, the standard clinical modality for breast cancer diagnosis, has limitations due to its ionized radiation, which makes it

unsuitable for pregnant women [4]. In contrast, BUS imaging is cost-effective and radiation-free, making it a valuable screening tool [5]. It provides information about breast tissue characteristics and the existence of malignancy tissues [5].

Segmenting breast tumors from breast images is a critical step in computer-aided diagnostic (CAD) systems for treatment planning [6]. This task is challenging for BUS imaging due to variable tumor shapes, ambiguous contours, low contrast, and inherent noise [7]. Advancements in CAD systems have led to investigations of its impact on breast cancer detection and segmentation using BUS images [8], [9]. Despite the difficulty of the task, CAD systems have shown promising progress in detecting and segmenting lesions.

Automated methodologies for breast cancer segmentation and classification can be categorized into two distinct classes: traditional and deep learning (DL) based techniques. Traditional methods, such as thresholding, region-growing, and handcrafted feature-based methods, may exhibit complexity

The associate editor coordinating the review of this manuscript and approving it for publication was Riccardo Carotenuto<sup>1b</sup>.

and potentially result in imprecise lesion segmentation owing to constraints in feature representations.

In contrast, deep learning-based segmentation methodologies have demonstrated notable performance achievements in the realm of breast lesion detection [10], [11], [12], [13], [14], [15]. Numerous deep learning techniques have been employed for breast cancer segmentation utilizing BUS images, leveraging automated feature learning through convolution layers.

The current research proposes a DL architecture, named the Dual Decoder Attention model with Attention ResUNet (DDA-AttResUnet). It integrates the concepts of the Dual Decoder Attention Network (DDANet) [16], Attention mechanisms [17], and the ResUNet architecture [18] to achieve precise and robust BUS cancer segmentation results.

The ResUNet combines ResNet and UNet architectures, utilizing residual blocks with skip connections from ResNet to address vanishing gradient problems in deep networks. The encoder-decoder architecture of UNet, augmented with skip connections, effectively captures both local and global contextual information, enhancing segmentation tasks.

The DDANet adheres to an encoder-decoder paradigm featuring a shared encoder shared between two parallel decoders. The initial decoder performs as the segmentation network, whereas the second decoder performs as an autoencoder network. The autoencoder strengthens the encoder's feature maps as an auxiliary task and generates attention maps used in each decoder to enhance the semantic representation.

In addition, the DDA-AttResUnet incorporates attention mechanisms that enable selective focus on relevant regions in input images. These mechanisms prioritize informative regions while suppressing less relevant areas, effectively utilizing both local and global contextual information, resulting in improved segmentation accuracy and robustness for lesion segmentation.

Overall, the paper's contributions/features are as follows:

- Introducing DDA-AttResUnet for accurate breast cancer segmentation from BUS imaging.
- Incorporating attention mechanisms with both spatial-wise and channel-wise attention to selectively focus on relevant regions in the input image. This attention mechanism assigns higher importance to informative areas while suppressing less relevant ones, enhancing ResUNet's ability to capture essential features.
- The incorporation of autoencoder attention maps is employed to enhance the feature maps within the encoder network.
- Competitive tumor segmentation performance in comparison to the current state-of-the-art works on the same challenging non-ionizing, safe BUSI dataset.

The remainder of this manuscript is structured as follows: Section II provides a comprehensive review of prior research efforts pertaining to breast cancer segmentation and enhancement. Section III expounds upon the architectural details of the proposed DDA-AttResUnet model. Section IV presents the experimental results and subsequent discussions, while

Section V offers the concluding remarks of the paper and delineates prospective avenues for future research.

## II. RELATED WORK

### A. RELATED METHODS FOR CANCER SEGMENTATION FROM BUS IMAGING

In the existing literature, the segmentation of breast lesions from ultrasound images has been thoroughly investigated through the application of diverse algorithms. Methods in the literature can be categorized into early methods and machine learning methods. Early methods include region-growing methods, deformable models, and graph models, whereas machine-learning methods include traditional handcrafted methods and deep learning (DL) approaches.

Region-growing methods start the segmentation process from manually or automatically selected seeds, gradually expanding to capture target region boundaries based on predefined growing criteria. The method was utilized by Shan et al. [19] for breast cancer segmentation based on incorporating contour smoothness and region similarity criteria.

On the other hand, deformable models initialize a foundational model and subsequently undergo deformations to converge towards object boundaries while accounting for internal and external energy factors. For instance, Madabhushi et al. [20] initiated the deformable model by utilizing boundary points and incorporating balloon forces to define the external energy. Chang et al. [21] utilized the stick filter to reduce speckle noise in ultrasound images before employing a 3D discrete active contour model to deform the model and accurately segment breast lesion regions. Initial graph-based models employ streamlined energy optimization methodologies within the context of Markov random fields or graph-cut frameworks. Chiang et al. [22] employed a pre-trained Probabilistic Boosting Tree (PBT) classifier to calculate the data term within the graph cut energy, whereas Xian et al. [23] formulated the energy function by integrating information from both frequency and spatial domains. However, these pre-existing models exhibit constraints in their ability to capture intricate semantic features and discern faint boundaries within areas of ambiguity, consequently resulting in boundary inaccuracies in ultrasound images with low contrast.

Traditional machine-learning techniques utilize designed handcrafted features to train machine-learning classifiers for segmentation tasks. Methods proposed by Liu et al. [24] and Jiang et al. [25] extracted various local image features and trained SVM or Adaboost classifiers for breast lesion segmentation. On the other hand, more recent machine-learning techniques for breast cancer segmentation are based on DL. More specifically, the recent advances in convolutional neural networks (CNNs) have shown excellent performance by learning high-level semantic features from labeled data. For example, Zhuang et al. [26] implemented the Residual Dilated Attention Gate UNet (RDAU-Net), a U-Net-derived

model tailored for the segmentation of tumors in Breast Ultrasound (BUS) images. Hu et al. [27] utilized the dilated fully convolutional network with a phase-based active contour model to segment breast tumors from ultrasound images. Byra et al. [28] employed a deep learning segmentation process based on entropy parametric maps. Zhu et al. [29] utilized an approach for breast lesion segmentation in ultrasound images, leveraging second-order statistics from multiple feature subregions.

More recently, Shareef et al. [30] devised the Enhanced Small Tumor-Aware Network (ESTAN), a Convolutional Neural Network architecture specifically tailored for the segmentation of diminutive tumors in Breast Ultrasound (BUS) images. ESTAN employed a dual-encoder architecture to collaboratively extract and integrate contextual information from images at multiple scales. Punn and Agarwal [31] utilized the RCA-IUnet (residual cross-spatial attention-guided inception U-net) model designed for tumor segmentation within breast ultrasound imaging. RCA-IUnet adopted the U-Net architecture, integrating elements such as residual inception depth-wise separable convolution, a blend of pooling layers (comprising max pooling and spectral pooling), and cross-spatial attention filters. Yang et al. [32] implemented the Cross-task guided network (CTG-Net), which encompassed a framework that combined two core tasks within computerized breast lesion analysis: lesion segmentation and tumor classification. Wu et al. [33] utilized the Boundary-Guided Multiscale Network (BGM-Net) designed for breast lesion segmentation within ultrasound images. The architecture of BGM-Net was rooted in the Feature Pyramid Network (FPN) framework and integrated boundary guidance mechanisms. Zhang et al. [34] implemented BO-Net, a specialized boundary-oriented network devised to improve breast tumor segmentation in ultrasound images. BO-Net incorporated a two-step process: first, a boundary-oriented module (BOM) was employed to delineate weak tumor boundaries through the acquisition of supplementary boundary maps; subsequently, feature extraction is carried out by deploying the Atrous Spatial Pyramid Pooling (ASPP) module, and developed InvUNET, a detection method for breast tumors in ultrasound images. InvUNET combined the UNET architecture with involution layers and lightweight kernels that enabled location-specific and channel-agnostic representation learning. Chen et al. [35] utilized AAU-net, an adaptive attention U-net designed for the automatic and stable segmentation of breast lesions in ultrasound images. They utilized AAU-net with the Hybrid Adaptive Attention Module (HAAM), which integrated both a channel self-attention block and a spatial self-attention block. Umer et al. [36] developed the Dual-Decoded Attention Mechanism (DDA-Net), utilizing an autoencoder architecture with a U-shaped structure and incorporating a dual-decoded attention mechanism to enhance performance. Lyu et al. [37] developed a pyramid attention network for breast ultrasound image segmentation, integrating Attention mechanisms and

multi-scale features. This architecture incorporated separable convolutions to generate a multi-scale receptive field through the aggregation of incremental small-size convolutions. It was further augmented by the inclusion of a Spatial and Channel Attention (SCA) module. Zhang et al. [38] implemented a computational model tailored for BUS screening, comprising a dual-branch architecture. This architecture encompassed a classification branch, dedicated to discerning between normal and tumor-present images, and a segmentation branch, tasked with delineating tumor regions. Notably, these two branches were endowed with a shared encoder network to facilitate information sharing and feature extraction. Zhang et al. [39] implemented the SaTransformer, a semantic-aware model explicitly engineered to concurrently address breast cancer classification and segmentation tasks within an integrated framework. This distinctive approach enabled mutually beneficial information exchange and synergy between the two tasks during the process of feature representation learning.

Table 1 presents a summary of related work of breast cancer segmentation including method and their data. While early methods and traditional machine learning-based approaches have been explored for breast lesion segmentation, recent advancements in deep learning techniques, particularly CNN-based models have shown promising results. However, the related works for breast cancer segmentation have the following limitations:

- Region growing methods struggle with handling inhomogeneities and noise present in ultrasound breast images, which can lead to inaccurate segmentations.
- Graph models have limitations in capturing high-level semantic features suitable for tumor segmentation from low-contrast ultrasound images.
- Deformable models, while flexible, are sensitive to initialization, and their performance may vary depending on the starting point, potentially resulting in suboptimal results.
- Handcrafted models rely on carefully engineered features and may not capture all the relevant information necessary for accurate segmentation.
- Despite the rich information extracted by deep learning methods, their segmentation accuracies still require further improvement.

To overcome the limitations of the existing works, an advanced Deep learning approach, DDA-AttResUNet integrates DDA-Net technology with Attention ResUNet architecture, leading to superior segmentation performance. The related work to this model is presented in the following subsection.

## B. RELATED WORK TO DDA-AttResUnet MODEL

The DDA-AttResUnet model is built upon three fundamental features: residual learning, incorporation of attention mechanisms, and the utilization of dual decoder attention technology. These components work together to enhance the model's performance in medical image segmentation tasks.

**TABLE 1. Summary of breast cancer segmentation related work: Author, methodologies and dataset used.**

Author(s)	Methodology	Dataset Used
Shan et al. [19]	Region-growing	120 Breast Ultrasound (BUS) images were obtained from the Second Affiliated Hospital of Harbin Medical University.
Madabhushi et al. [20]	Deformable models with boundary points and balloon forces.	42 breast BUS for 42 patients.
Xian et al. [23]	Energy function modeling	184 BUS images (93 benign and 91 malignant).
Liu et al. [24]	Machine-learning with SVM	112 BUS images (60 malignant and 52 benign).
Jiang et al. [25]	Machine-learning with handcrafted features	112 BUS images (80 diseased and 32 normal).
Zhuang et al. [26]	Residual Dilated Attention Gate UNet	1062 BUS images from GelderseVallei Hospital, Shantou University, and Breast Ultrasound Lesions Dataset.
Hu et al. [27]	Dilated fully convolutional network with active contour	1) 570 BUS images from 89 patients. 2) 128 BUS images (66 malignant and 62 benign)
Byra et al. [28]	Attention gated U-Net convolutional neural network	269 BUS masses (123 malignant masses and 146 benign masses)
Zhu et al. [29]	Second-Order Subregion Pooling network	1) <b>BUSI</b> dataset (780 images with 487 benign, 210 malignant, and 133 normal) from Baheya Hospital, 2) 632 BUS images from Shenzhen People's Hospital.
Shareef et al. [30]	Enhanced Small Tumor-Aware Network	1) <b>BUSIS</b> dataset, comprising 562 images, was sourced from multiple institutions, including Qingdao University, Harbin Medical University, and Hebei Medical University, 2) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 3) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell.
Punn and Agarwal [31]	Residual cross-spatial attention guided inception U-Net model	1) <b>BUSIS</b> dataset, comprising 562 images, was sourced from multiple institutions, including Qingdao University, Harbin Medical University, and Hebei Medical University, 2) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital.
Yang et al. [32]	Cross-task guided network	1) <b>THH</b> dataset, comprising 2718 images categorized as normal and 2022 as lesion ultrasound (US) images, was sourced from the privately owned Takamatsu Heiwa Hospital, 2) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell, 3) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital.
Wu et al. [33]	A boundary-guided multiscale network	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>BUSZPH</b> dataset, encompassing 632 breast ultrasound images, was obtained from Shenzhen People's Hospital dataset.
Zhang et al. [34]	A boundary-oriented network	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell.
Chavan et al. [40]	A UNET architecture with involution layers	<b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital.
Chen et al. [35]	An adaptive attention.	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell.
Umer et al. [36]	A Dual-Decoded attention mechanism.	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell.
Lyu et al. [37]	Pyramid Attention Network	1) 1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>OASBUD</b> dataset comprises 100 Breast Ultrasound (BUS) images, with 48 classified as benign and 52 as malignant.
Zhang et al. [38]	Classification and Segmentation branches network.	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>SYUSI</b> dataset, consisting of 1600 Breast Ultrasound (BUS) images, includes 212 images categorized as normal, 217 as benign, and 196 as malignant. The dataset was sourced from the Third Affiliated Hospital of Sun Yat-sen University.
Zhang et al. [39]	SaTransformer, a semantic-aware model.	1) <b>BUSI</b> dataset, comprising 780 images categorized as 133 normal, 487 benign, and 210 malignant, was acquired from Baheya Hospital, 2) <b>UDIAT</b> dataset, consisting of 163 images, with 110 classified as benign and 53 as malignant, was procured from the Diagnostic Centre of the Parc Tauli Corporation, Sabadell.

Residual learning, introduced by He et al. [41], addresses the degradation problem caused by increasing network depth during training [42]. This technique enables the training of very deep neural networks by utilizing skip connections and residual blocks, effectively mitigating the vanishing

gradient issue and facilitating a better flow of gradients during backpropagation. As a result, deeper architectures can be trained, leading to improved representation learning and higher accuracy in tasks like image segmentation. Residual learning has been pivotal in advancing the

state-of-the-art in deep learning and has become a foundational element in modern neural network architectures. In the context of ResUNet, full residual units are employed prior to activation [18].

Self-attention is an attention mechanism that can be effectively used as a modular component in basic CNN architectures with reduced computation and parameter overhead for generating attention maps [43]. This modular extension introduces supplementary neural network modules capable of dynamically assigning weights to features either across spatial dimensions or within channel-wise representations [43].

Spatial attention focuses on the importance of each pixel in the spatial location, while channel attention assigns weights to each channel of a feature map [44].

Channel attention, on the other hand, focuses on identifying and selecting important feature dimensions by assigning weights to each channel [45]. This mechanism is represented by a 1D vector. Both spatial and channel attention modules can be incorporated through parallel or sequential combinations. For example, Fu et al. [46] introduced the Dual Attention Network, which applies both types of attention in parallel and fuses their output features. On the contrary, Woo et al. [47] implemented the Convolutional Block Attention Module (CBAM), a mechanism that sequentially applies both channel and spatial attention modules.

The Dual-Decoded Attention Network (DDA-Net) is a novel architecture specifically designed for BUS image cancer segmentation. It follows an encoder-decoder structure with a shared encoder and two concurrent decoders [16]. The first decoder serves as the primary segmentation network, responsible for generating the segmentation mask, while the second decoder functions as an autoencoder network. The autoencoder network plays a crucial role in enhancing the encoder's feature mappings by performing an auxiliary task of creating an attention map. This attention map highlights important regions in the encoder's feature maps, which are then utilized by both decoders to enhance the semantic representation of the feature maps, resulting in improved segmentation performance. By incorporating the dual-decoder attention network and attention mechanism with ResUNet, DDA-AttResUnet excels at capturing high-level semantic features and accurately identifying weak boundaries in ambiguous regions.

This capability is particularly beneficial in challenging scenarios, such as low-contrast ultrasound images with artifacts and intensity inhomogeneity.

In our study, we compare related techniques and hold an ablation study that showed that incorporating the Attention Mechanism and Dual Decoder Attention Network with the ResUnet model achieves superior segmentation results.

### III. MATERIALS AND METHODS

#### A. DATASET

The Breast Cancer Ultrasound Images (BUSI) dataset encompasses a collection of 780 ultrasound images acquired

at the Baheya Hospital in Cairo, Egypt, utilizing LOGIQ E9 and LOGIQ E9 Agile ultrasound scanners [48]. Each individual image within the dataset possesses an average resolution of  $500 \times 500$  pixels and is stored in the PNG file format. The images have been classified into three distinct categories: normal (133 images), benign (437 images), and malignant (210 images), with each image accompanied by a binary ground truth mask that identifies tumor regions. The BUSI dataset is a valuable and challenging resource for breast lesion segmentation research, with a substantial number of ultrasound images categorized into normal, benign, and malignant cases of different nodules' sizes and shapes. The data sample is shown in Figure 1.

#### B. PREPROCESSING

Preprocessing involves normalization, resizing, and augmentation of the training images. Normalization is a process that entails the transformation of pixel values to conform to a predetermined range, conventionally bounded between 0 and 1, to facilitate a smoother training process and improve model generalizability. The images are then resized to a standard size of  $128 \times 128$  pixels to match the standard input size required by the chosen model.

Data augmentation is systematically applied to the training dataset with the objective of enriching the diversity of training data, thereby enhancing the model's performance and its capacity for generalization. We apply flip transformations (both vertical and horizontal flipping) and random rotations up to  $90^\circ$  to each input image. This process generates ten augmented images for every original input image.

#### C. METHODOLOGY

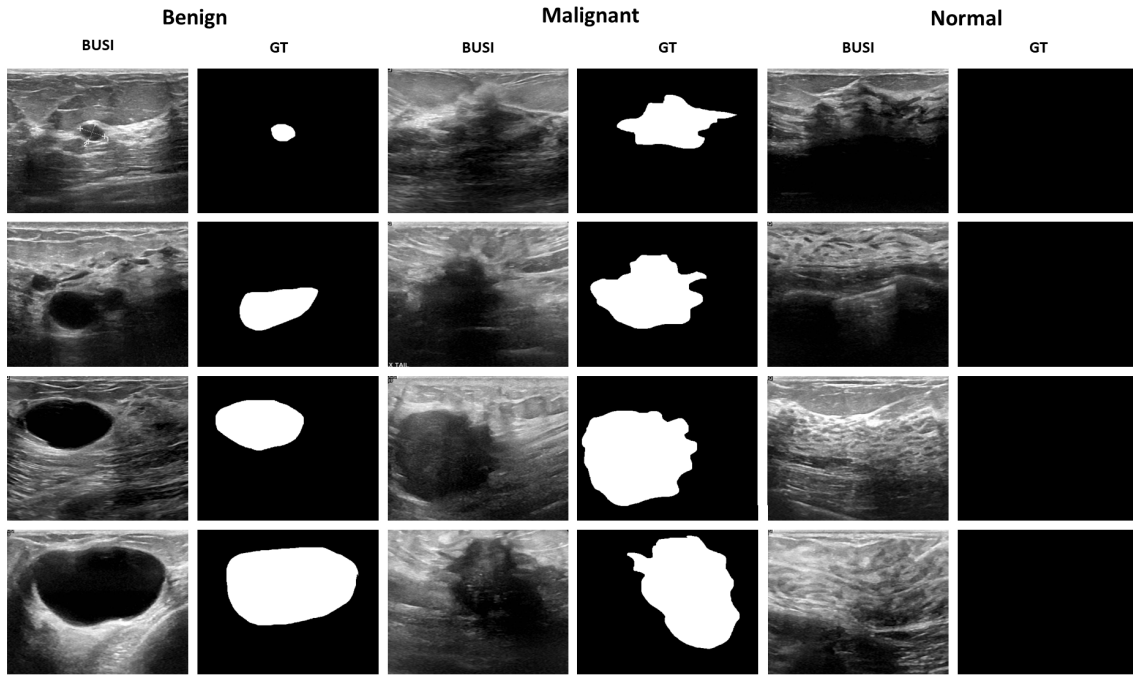
This section illustrates in detail our proposed encoder-decoder-based architecture.

##### 1) ARCHITECTURE OVERVIEW

The Dual Decoder and Attention Mechanism ResUNet architecture illustrated in Figure 2 is tailored for tumor segmentation in breast cancer ultrasound images. It operates by taking a grayscale BUS and its corresponding mask as input, processing the image through an encoder based on the ResUNet architecture.

##### 2) DDA-Net

The DDA-Net architecture includes an encoder for standard feature extraction, and its outputs are directed into two decoders: the Main Decoder (segmentation branch) and the Auxiliary Decoder (autoencoder branch). In the Auxiliary Decoder, the output is connected to an Autoencoder attention map block, which includes a Conv2D layer with a sigmoid activation function. Meanwhile, the output of the Main Decoder is multiplied by the output of the Autoencoder attention map block, resulting in an attention map that highlights significant regions for segmentation. This attention mechanism emphasizes areas where the features from



**FIGURE 1.** Breast cancer ultrasound images (BUSI), including benign, malignant, and normal cases, along with their corresponding binary ground truth masks for tumor segmentation.

the Main Decoder and the reconstructed image from the Auxiliary Decoder agree, indicating their importance for achieving accurate segmentation.

### 3) ATTENTION MECHANISM

In deep learning-based image processing, the convolutional layer is crucial for feature extraction, but it may overlook contextual information in input feature maps. To address this limitation, pooling is used to increase the dependency of each output pixel on a larger neighborhood of input pixels. However, convolutional layers may still struggle to incorporate contextual information effectively. Attention mechanisms are introduced to summarize input information and influence the main information flow in the network, considering variations in importance among pixels or channels. The integration of attention mechanisms, such as spatial-wise and channel-wise attention, enhances the representation of segmented regions by combining information from multiple sources [47]. This approach leads to significant advancements in breast tumor segmentation, leveraging deep learning and attention mechanisms for improved accuracy and performance. In our architecture, we incorporate two attention modules, based on the work by Zhao et al. [43], to further enhance breast tumor segmentation:

#### 1) Channel Attention:

Likewise, we integrate the channel attention module at the terminal layer of the encoder, given that the high-level feature map predominantly encapsulates intricate features characterized by a broad receptive field and an abundance of channels. Through the utilization of global information, this procedure

empowers the network to engage in feature recalibration, selectively amplifying significant attributes while dampening less consequential ones. Figure 3 depicts the organizational structure of the channel attention module.

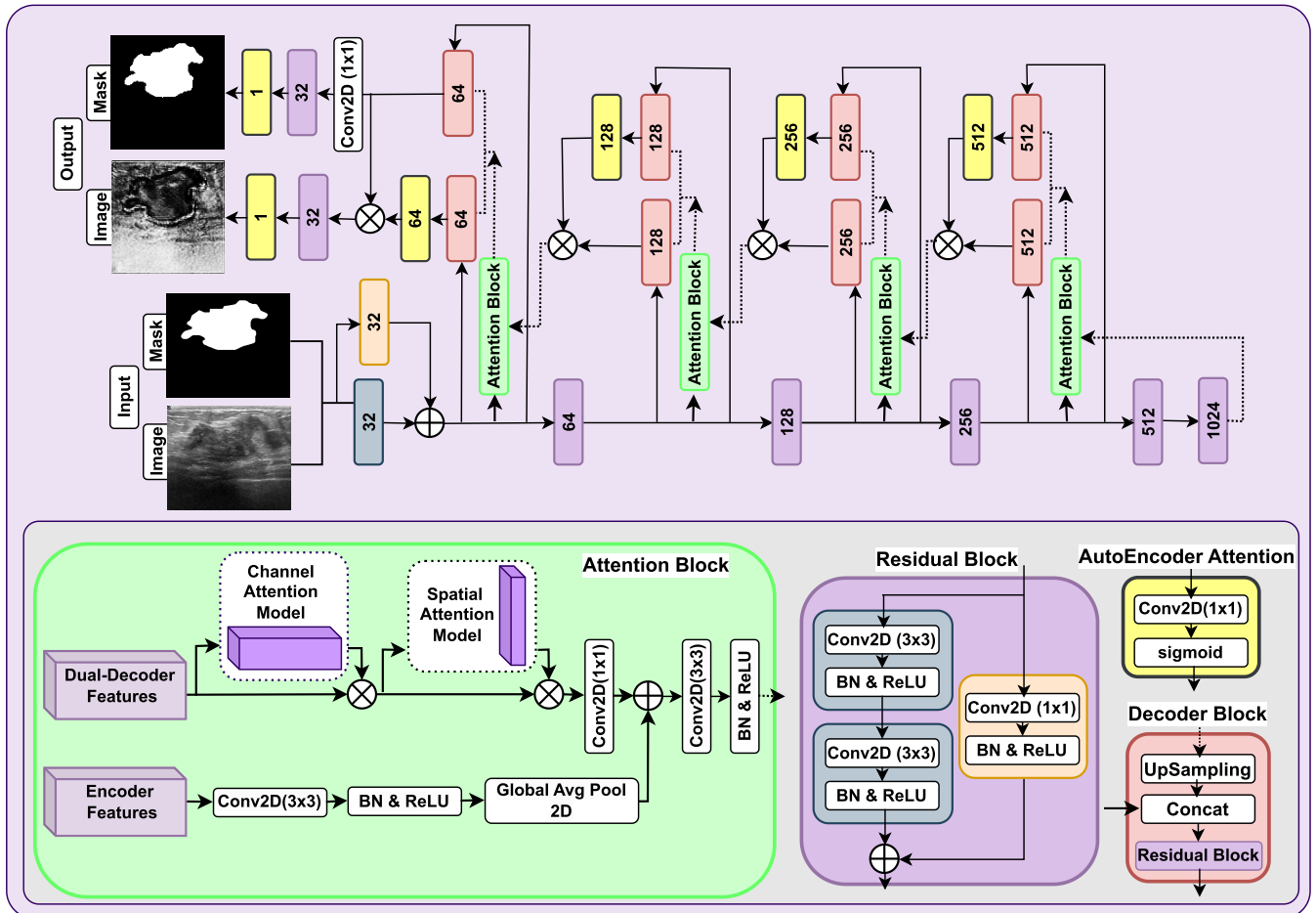
First, we send the feature map  $X \in \mathbb{R}^{H \times W \times C}$  to the aggregation operation, which aggregates the feature map in its spatial dimensions ( $H \times W$ ) to produce a channel descriptor  $v \in \mathbb{R}^C$ . The channel aggregation operation produces a global distribution of channel-specific features by means of calculating the channel descriptor. The calculation formula for this is as follows [43]:

$$v_c = Z_{ac}(x_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

Here,  $x_c \in \mathbb{R}^{H \times W}$  represents the local feature map of channel  $c$ . We use global average pooling for the spatial dimensions in the aggregation operation  $Z_{ac}$ . This produces a channel descriptor  $v \in \mathbb{R}^C$ , which represents a global distribution of channel features.

Following the aggregation operation, a self-learning weight computation process is executed using fully connected layers. The function  $Z_1(v, a)$  is designed to capture inter-channel dependencies and dynamically generate the channel weight map  $y \in \mathbb{R}^C$ . The calculation formula is as follows [43]:

$$y = Z_1(v, a) = \sigma(v, a) = \sigma(a_2 \delta(a_1 v)) \quad (2)$$



**FIGURE 2.** The proposed DDA-AttResUNet architecture is designed for breast tumor segmentation. It utilizes a ResUNet architecture with convolutional filters for hierarchical feature extraction. The architecture includes one stem block, consisting of a  $3 \times 3$  convolutional block (blue) added to a  $1 \times 1$  convolutional block (orange), followed by five encoder blocks. Each encoder block contains a Residual Block (purple), which consists of two  $3 \times 3$  convolutional blocks (blue) with the addition of a  $1 \times 1$  convolutional block (orange). The architecture also features five dual decoder blocks, with each block containing both the Main Decoder and the Auxiliary Decoder (pink). Each decoder includes an Upsample block and a Concatenate block, along with a ResNet Block (purple). The Auxiliary Decoder connects to a Conv2D layer with a sigmoid activation function, known as the Autoencoder Attention block (yellow), and its output is multiplied by the output of the Main Decoder. Additionally, the model incorporates an attention fusion mechanism represented by the Attention Block (green), which includes both spatial-wise and channel-wise components. This attention mechanism enhances the representation of segmented regions. After the last decoder, the Main Decoder connects to a  $1 \times 1$  Conv2D layer, and the output of the last decoder is combined with the ResNet Block (purple) and a Conv2D layer with a sigmoid activation function, referred to as the AutoEncoder Attention map block (yellow). This configuration generates two outputs: a segmentation mask highlighting the boundaries and locations of tumors and an enhanced ultrasound image.

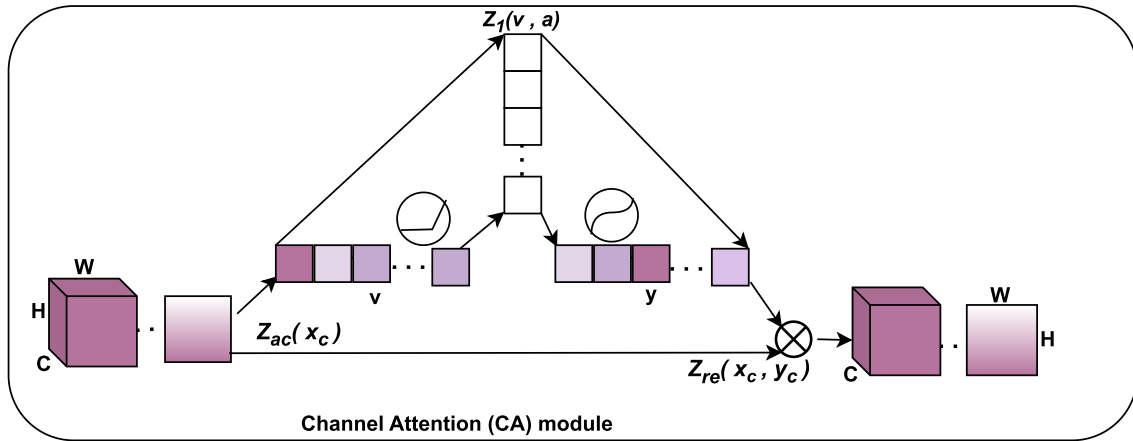
Here,  $a_1 \in \mathbb{R}^{L \times C}$ ,  $a_2 \in \mathbb{R}^{L \times C}$ , where  $L$  is the number of hidden neurons. The symbol  $\sigma$  represents the sigmoid activation function that generates the channel weight  $y_c \in (0, 1)$  for channel  $c$ . Fully connected hidden layers possess the capability to capture non-linear interactions among channels. The feature map  $X$  is multiplied by the weight calculated in the previous step. The Channel Attention module produces output  $X'$  by multiplying the feature values of different channels in  $X$  with different weights, achieved through channel-wise recalibration  $Z_{re}(x_c, y_c)$  [43].

$$x'c = Z_{re}(x_c, y_c) = x_c \cdot y_c \quad (3)$$

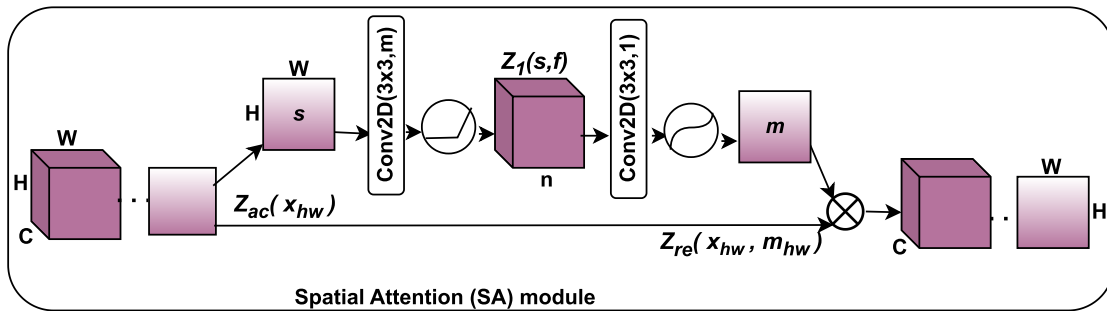
2) Spatial Attention:

Spatial attention has been incorporated into convolutional neural networks as an attention module and

has demonstrated strong performance in tasks such as classification and detection [47]. Spatial attention in convolution neural networks captures positional information between images, which helps to depict the spatial relationship between the input features. Spatial attention, by design, does not take into account channel information and treats all features from different channels equally. To address this limitation, we utilize a spatial attention module to the low-level feature map which primarily extracts spatial features such as contours and edges, and contains fewer channels. This module learns the interactions between spatial points, amplifies important regions, and suppresses irrelevant ones. Figure 4 depicts the architecture of the spatial attention module. Initially, the feature map  $X \in \mathbb{R}^{H \times W \times C}$  is fed into the aggregation operation that



**FIGURE 3.** The channel attention (CA) module comprises three key functions:  $Z_{ac}$ ,  $Z_1$ , and  $Z_{re}$ . Specifically,  $Z_{ac}$  produces a channel descriptor  $v \in \mathbb{R}^C$ ,  $Z_1$  generates the channel weight map  $y \in \mathbb{R}^C$ , and  $Z_{re}$  employs  $y$  to generate the CA module’s output. Notably,  $Z_1$  is realized through the utilization of two fully connected layers and is responsible for self-learning [43].



**FIGURE 4.** The Spatial Attention (SA) module comprises four primary functions. The aggregate function  $Z_{ac}$  is responsible for generating a spatial descriptor  $s \in \mathbb{R}^{H \times W}$  by aggregating the feature map across the channel dimension. The self-learning function  $Z_1$ , realized through the utilization of two convolutional layers, is tasked with generating the spatial weight map  $m \in \mathbb{R}^{H \times W}$  by adaptively capturing spatial correlations. Lastly, the function  $Z_{re}$  employs  $m$  to enact spatial-wise recalibration, culminating in the generation of the SA module’s output, denoted as  $X'$  [43].

produces a spatial descriptor  $s \in \mathbb{R}^{H \times W}$ . is descriptor is derived through the aggregation of the feature map along its channel dimension ( $n$ ). The aggregation operation yields a comprehensive distribution of spatial features at a global scale [43]:

$$s_{hw} = Z_{ac}(x_{hw}) = \frac{1}{n} \sum_{i=1}^n x_{hw}(i) \quad (4)$$

Here, the notation  $x_{hw} \in \mathbb{R}^C$  represents the local feature at spatial position  $(h, w)$  with  $C$  channels. The aggregate function  $Z_{ac}$  uses global average pooling to aggregate information across the channel dimension. Subsequently, a self-learning weight computation process ensues, realized through the utilization of convolutional layers. The aim of the function  $Z_1(s, f)$  is to capture the spatial correlations fully and generate the spatial weights map  $m \in \mathbb{R}^{H \times W}$  adaptively. The calculation formula for this is as follows [43]:

$$m = Z_1(s, z) = \sigma(s, z) = \sigma(z_2 \delta(z_1 s)) \quad (5)$$

Here,  $z_1$  refers to a  $3 \times 3$  convolution, which is denoted as  $\text{Conv}(3 \times 3, n)$ , while  $z_2$  refers to another  $3 \times 3$  convolution, denoted as  $\text{Conv}(3 \times 3, 1)$ .

The value of  $n$  corresponds to the channel number of the hidden feature map. The symbol  $\delta$  refers to the ReLU activation function, while  $\sigma$  represents the sigmoid activation function that generates the spatial weight  $m_{hw} \in (0, 1)$  at position  $(h, w)$ . In essence, the convolutional operation, which accepts the original spatial descriptor as input, can be conceptualized as a spatial-wise self-attention function. By doing so, it can capture non-linear inter-spatial relationships. The spatial weights determined in the preceding step are subsequently applied to the feature map  $X$ . Through spatial-wise recalibration,  $Z_{re}(x_{hw}, m_{hw})$ , the feature values of different positions in  $X$  are multiplied by different spatial weights to generate the output  $X'$  of the SA module [43]:

$$x'_{hw} = Z_{re}(x_{hw}, m_{hw}) = x_{hw} \cdot m_{hw} \quad (6)$$

The proposed architecture incorporates two attention modules to enhance feature representation and capture rich contextual relationships. The first module, known as the channel-wise attention module, operates on the input feature maps resulting from the multiplication of the Main Decoder and the Auxiliary Decoder in the Dual Decoder. It assigns



a specific value to each channel through element-wise multiplication between the input feature maps and the output of the channel-wise attention module.

The output of the channel-wise attention module then undergoes further processing by the spatial-wise attention module. In this module, specific weights are assigned to each pixel in the feature maps, further emphasizing important regions. The resulting output is once again multiplied by the output of the previous channel-wise attention module.

To process the output of the encoder part in the model, we apply a 2D convolutional layer ( $3 \times 3$ ) followed by a ReLU activation function and Global Average Pooling 2D. The result is achieved through the summation of the outcome from this layer with the output of the attention module, which amalgamates the outcomes of both the channel-wise and spatial-wise attention modules. The resulting combined output is subsequently passed through a 2D convolutional layer ( $3 \times 3$ ), followed by a ReLU activation function and Batch Normalization (BN) layers. This process generates the ultimate output that is then directed to the next Dual Decoder block.

Figure 2 illustrates the block diagram of the attention unit process (depicted as the green block). We have incorporated Attention Units into the decoder segment of our architecture, enabling the model to direct its attention toward pivotal regions within the feature maps.

#### 4) RESIDUAL BLOCK

Training a deep neural network with increased network depth has the potential to improve overall performance. However, simply stacking CNN layers could also hamper the training process, leading to issues like exploding or vanishing gradients during backpropagation [41].

To address this problem, residual connections are introduced, facilitating the training process by directly routing the input information to the output and preserving the flow of gradients [41]. The residual function simplifies the optimization objective without introducing additional parameters and boosts the performance, inspiring the use of deeper residual-based networks [41].

The working principle of residual units is represented by the equation [41]:

$$y_n = F(x_n, W_n) + x_n$$

where  $y_n$  represents the output of the residual unit,  $F(\cdot)$  is the residual function operated on the input  $x_n$  using the weights  $W_n$ , and the result is added to the original input  $x_n$  to produce the final output  $y_n$ . The weights  $W_n$  represent the weights associated with the layers within the residual function  $F(\cdot)$ . The addition of the original input  $x_n$  is the key idea behind residual connections, which enables the network to learn the residual or the difference between the input and output. This approach simplifies the learning process and helps mitigate the vanishing gradient problem, thereby enabling the training of deeper neural networks.

The residual units are composed of combinations of Batch Normalization (BN), Rectified Linear Unit (ReLU), and convolutional layers. Detailed descriptions of these combinations and their impact can be found in the work of He et al. [41].

In the ResUNet architecture, every Residual Block is structured with two successive  $3 \times 3$  convolutional blocks and one  $1 \times 1$  convolutional block. Each convolutional block includes a batch normalization layer, a Rectified Linear Unit (ReLU) activation layer, and a convolutional layer. The Residual Block's operation involves adding the output of the two successive  $3 \times 3$  convolutional blocks to the output of the  $1 \times 1$  convolutional block. This design enables the model to create a more robust and accurate representation of the features.

Additionally, the ResUNet architecture includes a special Residual Block called the stem block. The operation of the stem block involves adding the output of one successive  $3 \times 3$  convolutional block to the output of the  $1 \times 1$  convolutional block. Like the regular Residual Block, this design empowers the model to create a more robust and accurate representation of the features.

#### 5) DDA-AttResUNet

The DDA-AttResUNet architecture consists of an encoder and two decoder branches with an autoencoder attention map, utilizing attention mechanisms to achieve improved medical image segmentation.

In the encoder, each block contains a Residual Block with the number of filters specified as [32, 64, 128, 256, 512, 1024]. The encoder comprises one stem block and five encoder blocks, capturing hierarchical features using identity mappings.

To reduce the feature map size, a stridden convolutional layer is applied to the first encoder block, allowing the model to extract abstract features. The Residual Block and downsampled feature maps enhance the model's ability to learn intricate patterns and hierarchical information to improve segmentation performance.

The DDA-AttResUNet architecture processes the ultrasound input image through the encoder network, creating an abstract feature representation while downsampling it.

The decoder branches, Main Decoder (segmentation branch), and Auxiliary Decoder (autoencoder branch) employ reversed filters [1024, 512, 256, 128, 64, 32] for tumor segmentation. The Main Decoder focuses on segmentation using upsampling and Residual blocks. Meanwhile, the Auxiliary Decoder captures contextual information and enhances the segmentation using an autoencoder network.

The encoder network's output is simultaneously directed to both decoders, wherein a  $2 \times 2$  transpose convolution operation is employed to double its spatial dimensions. Subsequently, the image is concatenated with a suitable feature map derived from the encoder network, employing the residual connections. These connections retrieve features from preceding layers at their native resolution, amplifying

their representation potency and offering an alternative pathway for gradient flow, which is advantageous for model convergence.

The result produced by the Auxiliary Decoder block (autoencoder branch) is subject to processing through an AutoEncoder attention block, comprising a  $1 \times 1$  convolutional layer and a sigmoid activation function, leading to the generation of an attention map.

The attention map thus generated is employed as a multiplier for the output of the Main Decoder block (segmentation branch), functioning as the input for the subsequent decoder block within the segmentation branch. Attention mechanisms, such as spatial and channel attention, refine segmented regions by combining information from different sources.

Finally, the last decoder and the last output of the Main Decoder with the  $1 \times 1$  convolutional layer are each connected to the Residual Block and an Autoencoder Attention block. The network produces two images: a segmentation mask image for precise tumor segmentation from the Main Decoder and an enhanced ultrasound image from the Auxiliary Decoder for improved visualization.

## 6) PERFORMANCE METRICS

After training the models on the training set, they are applied to predict tumor segmentation masks for the test set. These predicted masks are then evaluated using several performance metrics that involve true positive ( $TP$ ), false positive ( $FP$ ), true negative ( $TN$ ), false positive ( $FP$ ), and false negative ( $FN$ ) values. These evaluation metrics include:

The Mean Intersection over Union ( $IoU$ ), also referred to as the Jaccard index, quantifies the degree of overlap between the predicted segmentation mask and the ground truth mask. This metric computes the ratio of the intersection to the union of the two masks, yielding a comprehensive measure of segmentation accuracy.

$$IoU = \frac{TP}{TP + FP + FN} \quad (7)$$

The Dice Coefficient ( $Dice$ ) is employed to assess the similarity between the predicted and ground truth masks. This metric quantifies the ratio of twice the intersection to the sum of the sizes of the two masks, offering insight into the degree of agreement between them.

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (8)$$

Accuracy ( $Acc$ ): Accuracy measures the overall correctness of the predicted segmentation compared to the ground truth. It computes the ratio of pixels that are correctly classified to the total number of pixels encompassed by the mask.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Sensitivity ( $Sen$ ), alternatively referred to as recall or true positive rate, quantifies the fraction of genuine positive

**TABLE 2. The split of the data into train, validation, and test.**

Data	Training	Validation	Testing
Without Augmentation	648	65	72
With Augmentation	6480	65	72

instances (tumor pixels) that the model correctly identifies.

$$Sen = \frac{TP}{TP + FN} \quad (10)$$

Precision ( $Pre$ ): Precision measures the proportion of predicted positive cases (tumor pixels) that are truly positive. It quantifies the model's ability to accurately identify tumor pixels without including too many false positives.

$$Pre = \frac{TP}{TP + FP} \quad (11)$$

By evaluating the models using these metrics, we can assess the quality and effectiveness of the tumor segmentation predictions by the presented model.

## IV. RESULTS AND DISCUSSION

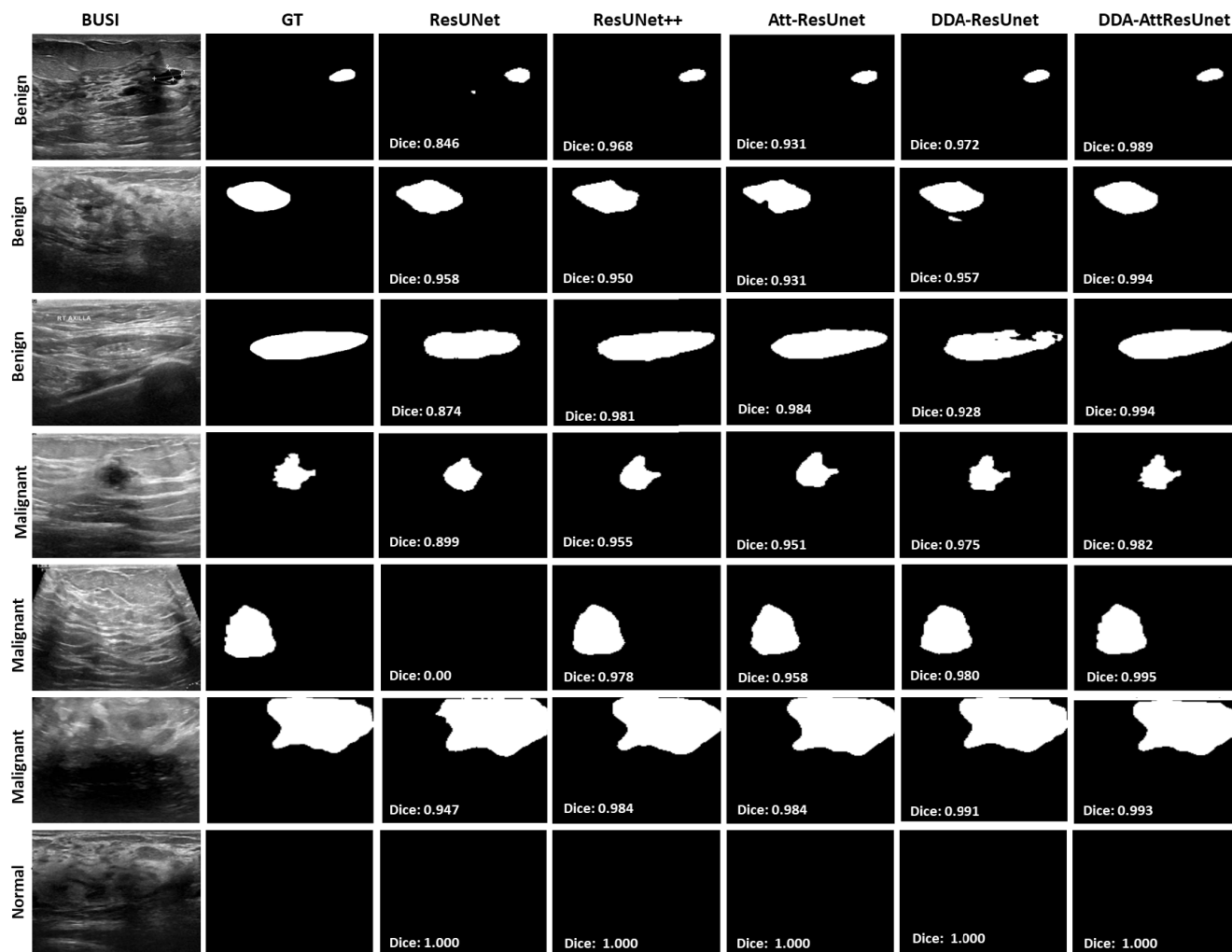
### A. EXPERIMENTAL SETTING

In this study, the BUSI dataset, composed of 720 images, is preprocessed to handle variations in image sizes. All images are normalized and resized to a standardized size of  $128 \times 128$  pixels for consistency across the segmentation models. The dataset is then randomly split into 10-fold cross-validation. This split remains consistent throughout the experimentation process. Each of the ten experiments divides the data into 90% (648 images) training and validation data, and 10% (72 images) test data which is kept separate and untouched throughout both the training and validation processes. In turn, the training and validation data are divided into 90% for training the model (which is subjected to augmentation) and 10% for validation, which involves hyperparameter tuning and model selection. The split of the data into train/validation/test data are illustrated on Table 2. The reported results in our paper reflect the average performance metrics obtained from the test set within the framework of 10-fold cross-validation, ensuring a robust and unbiased evaluation of the model's performance.

During the training phase, an NVIDIA Tesla P100 GPU accelerates the training process. The Adam optimizer is utilized to optimize the model, and an initial learning rate of 0.0001 is configured. The training process extends over 150 epochs, with the learning rate undergoing reduction by a factor of 2 when the learning progress reaches a plateau. This adjustment promotes enhanced model performance. To counteract overfitting, an early stopping technique is implemented. This technique halts the training process when validation error ceases to improve, preventing the model from becoming excessively tailored to the training data.

### B. QUALITATIVE RESULTS

In Figure 5 the input ultrasound image represents the original scan of the breast, while the ground truth image represents the manually annotated mask created by medical experts, serving



**FIGURE 5.** A qualitative assessment of tumor segmentation outcomes across different models is presented. The initial column depicts the original ultrasound image, while the second column exhibits the ground truth mask. Subsequent columns showcase the predicted masks produced by each model, specifically ResUNet, ResUNet++, Att-ResUnet, DDA-ResUnet, and DDA-AttResUnet. The corresponding Dice scores for each segmentation outcome are displayed beneath their respective predicted masks.

as the reference for tumor regions. The predicted segmented masks are generated by the DDA-AttResUnet model and other segmentation approaches, illustrating their respective performance in accurately delineating tumor regions.

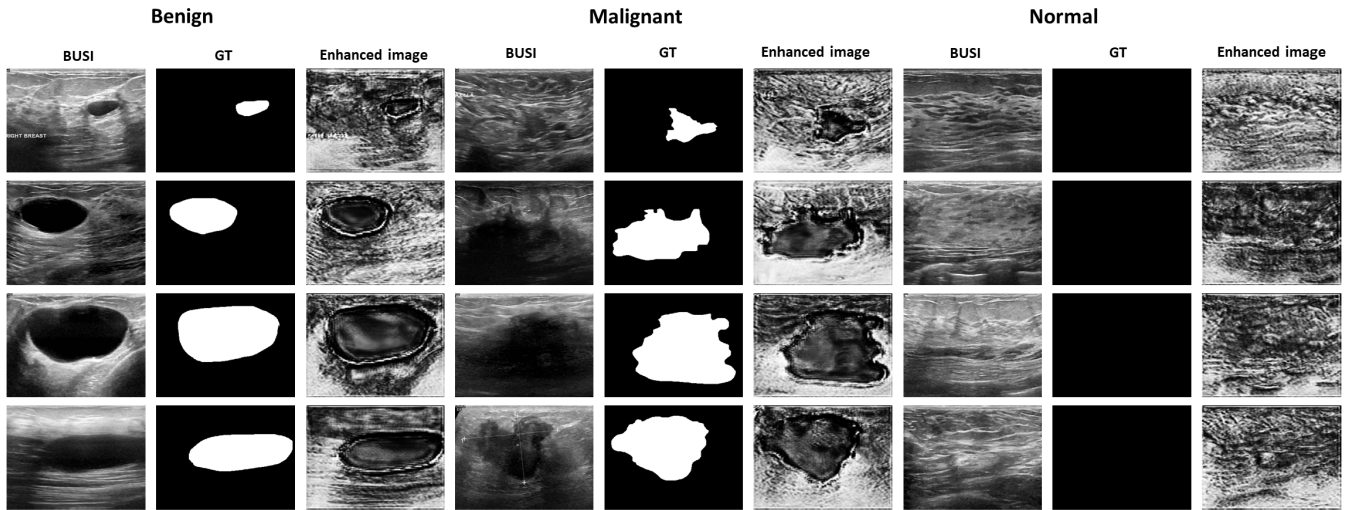
By visually comparing the predicted segmented masks to the ground truth masks, Figure 5 provides a comprehensive view of the segmentation performance of the DDA-AttResUnet model and its effectiveness in producing accurate and precise tumor segmentations. In the context of breast ultrasound image segmentation, the proposed Dual Decoder Attention model with Attention ResUNet architecture plays a crucial role in accurately identifying and delineating tumor regions. The model takes advantage of advanced deep learning techniques, attention mechanisms, and fusion operations to enhance its segmentation performance and generate enhanced ultrasound images.

The results produced by the network consist of two key components: a segmentation mask image and an improved ultrasound image, as depicted in Figure 6. The segmentation

mask offers crucial insights into tumor boundaries and positions, thereby facilitating precise tumor segmentation and diagnosis. Meanwhile, the enhanced ultrasound image, generated through the utilization of the autoencoder network (AutoEncoder attention block) within the Dual Decoder Attention, presents an upgraded rendition of the initial input ultrasound image. This enhanced image provides enhanced features, enabling medical practitioners to gain a clearer visualization of tumor regions and make more accurate clinical evaluations.

### C. QUANTITATIVE RESULTS

**Ablation experiments** are systematically conducted to analyze the individual components of the DDA-AttResUnet model. The study evaluated the performance of simplified versions, including ResNet, ResNet++, Att-ResUnet (utilizing Attention with ResNet), and DDA-ResUnet (incorporating Dual Decoder models). These experiments provided valuable insights into each component's contributions to



**FIGURE 6.** The comparative visualization displays breast tumor segmentation results involving benign, malignant, and normal breast images. Each case is presented in three columns. The leftmost column shows the input ultrasound image (Image) of the breast. The middle column presents the ground truth mask (Mask), while the rightmost column exhibits the predicted Enhanced Image generated by the DDA-AttResUnet model.

**TABLE 3.** Comparative analysis of DDA-AttResUnet with ablation experiments on the BUSI datasets (MEAN ± STD) by ten-fold cross-validation.

Method	Dice [%]	IOU [%]	Sen [%]	Pre [%]	Acc [%]
ResUnet	83.80 ± 1.22	74.36 ± 1.68	84.15 ± 1.66	85.99 ± 0.96	97.54 ± 0.16
ResUnet++	85.00 ± 0.97	75.29 ± 1.33	84.25 ± 0.88	86.23 ± 1.02	97.57 ± 0.14
Att-ResUnet	89.10 ± 1.03	81.87 ± 1.49	86.42 ± 1.27	92.39 ± 0.69	98.31 ± 0.14
DDA-ResUnet	90.15 ± 0.90	83.34 ± 1.41	87.21 ± 1.29	94.22 ± 0.36	98.48 ± 0.11
<b>DDA-AttResUnet</b>	<b>92.92 ± 0.69</b>	<b>87.39 ± 1.10</b>	<b>92.16 ± 0.92</b>	<b>93.90 ± 0.40</b>	<b>98.82 ± 0.10</b>

the overall segmentation performance, as demonstrated in Table 3. The evaluation metrics used in the analysis included the Dice coefficient, Jaccard index (IOU), sensitivity, precision, and overall accuracy. The results of the analysis demonstrated that the DDA-AttResUnet model outperformed the other segmentation approaches in terms of segmentation accuracy. It yielded a superior Dice coefficient of (92.92 ± 0.69%) and Jaccard index (IOU) of (87.39 ± 1.10%), signifying an enhanced degree of correspondence between the predicted segmentation masks and the ground truth masks. The model also showed improved sensitivity (92.16 ± 0.92%) and Precision (93.90 ± 0.40%), suggesting its ability to accurately detect tumor regions while minimizing false positives. Additionally, the DDA-AttResUnet model achieved a higher overall accuracy of (98.82 ± 0.10%). In the test images, the average time that each image takes is 40ms.

Table 4 provides an exhaustive comparative analysis of the proposed DDA-AttResUnet model in contrast to several contemporary state-of-the-art methodologies, all evaluated on the BUSI dataset. The table provides valuable insights into the performance of different techniques for segmenting breast tumors from ultrasound images. Notable methods, including Zhu et al. [29], Wu et al. [33], Shareef et al. [30], Zhang et al. [34], Chavan et al. [40], Yang et al. [32], Zhang et al. [38], Zhang et al. [39], Lyu et al. [37],

Umer et al. [36], and Punn and Agarwal [31], are included in this comparative analysis.

DDA-AttResUnet emerges as a competing method, achieving the highest scores in several critical performance metrics. It impressively attains a Dice coefficient of 92.92%, signifying a substantial overlap between its predicted segmentation and the ground truth. Additionally, DDA-AttResUnet demonstrates a sensitivity of 92.16%, indicating its capability to accurately identify true positive cases, along with an impressive accuracy of 98.82%, affirming its overall effectiveness in precisely classifying both positive and negative cases.

While DDA-AttResUnet excels in numerous aspects, it's important to acknowledge that its IOU (Jaccard index) measures at 87.39%, representing the proportion of intersections over the union of its predicted and ground truth regions. Furthermore, its precision stands at 93.90%, highlighting the accuracy of true positive predictions. In comparison, the method presented by Punn and Agarwal [31] achieves a slightly higher IOU of 89.9% and a precision of 94%, showcasing competitive performance in these specific metrics.

The remarkable performance of DDA-AttResUnet can be attributed to its sophisticated architecture, which combines the Dual Decoder Attention model with the Attention ResUnet segmentation model. By synergistically utilizing advanced deep learning techniques alongside dual decoder

**TABLE 4.** Comparative outcomes between the proposed methodology and contemporary state-of-the-art approaches are presented on the identical dataset, denoted as (BUSI).

Authors	Dice [%]	IOU [%]	Sen [%]	Pre [%]	Acc [%]
Zhu et al. [29]	84.70	76.39	85.51	87.62	NA
Wu et al. [33]	83.97	75.97	83.45	89.31	NA
Chen et al. [35]	77.21	67.97	80.63	78.66	NA
Shareef et al. [30]	80.00	72.00	85.00	NA	NA
Zhang et al. [34]	80.27	NA	80.42	85.24	NA
Chavan et al. [40]	80.09	71.46	82.77	83.15	96.09
Yang et al. [32]	82.00	74.00	84.00	NA	NA
Lyu et al. [37]	80.71	68.53	79.30	83.50	97.13
Zhang et al. [39]	86.34	83.12	89.51	88.32	NA
Zhang et al. [38]	89.8	79.1	85.9	NA	NA
Umer et al. [36]	88.68	80.12	87.79	95.42	NA
Punn and Agarwal [31]	91.40	<b>89.9</b>	89.00	<b>94.00</b>	97.00
Proposed <b>DDA-AttResUnet</b>	<b>92.92</b>	87.39	<b>92.16</b>	93.90	<b>98.82</b>

attention and Attention mechanisms, this model attains extraordinary accuracy and robustness in segmentation outcomes. This unique amalgamation of features sets DDA-AttResUnet apart in the pursuit of superior segmentation results.

#### D. DISCUSSION

The segmentation of breast cancer within breast ultrasound (BUS) images presents a formidable challenge, owing to the intricate and inherently ambiguous characteristics inherent in the data. Manual diagnosis of breast cancer from BUS images is subjective and difficult, leading to the development of CAD systems.

This study focuses on breast cancer segmentation from BUS images and proposes a method called DDA-AttResUnet. The proposed method achieves 92.92%, 92.16%, 93.90%, and 98.82% in Dice score, Sensitivity, Precision, and Accuracy, respectively, using the BUSU dataset. Comparative analysis with existing deep learning-based segmentation techniques demonstrates the competitive performance of DDA-AttResUnet in terms of segmentation DSC scores, highlighting its potential for reliable breast cancer diagnosis using BUS images.

Recent researches by Zhang et al. [39], Zhang et al. [38], Lyu et al. [37], Umer et al. [36], Shareef et al. [30], Zhang et al. [34], Yang et al. [32], and Punn and Agarwal [31] have concentrated on employing deep learning techniques for the segmentation of breast cancer, specifically on the challenging BUSI images. Punn and Agarwal [31] documented the highest DSC of 91.4%, slightly lower than the reported Dice score achieved by the proposed DDA-AttResUnet.

One of the limitation of the proposed system is that the IOU and Precision can still be improved. Pun and Agarwal [31] achieved higher values for IOU (89.9%) and Precision (94%) compared to the proposed DDA-AttResUnet. This indicates the need for investigating more areas for improvement.

Overall, the integration of the dual decoder attention and attention mechanism contributes to exceptional segmentation DSC scores compared to existing models. Additionally, qualitative results (Fig. 5) further validate its effectiveness in segmenting breast cancer from ultrasound images.

Furthermore, Figure 6 showed enhanced ultrasound images generated by the DDA-AttResUnet model with improved visual clarity when compared to the original inputs, which enables a clearer identification of tumor regions.

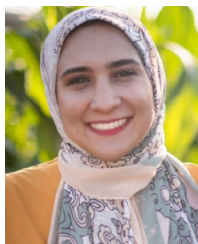
#### V. CONCLUSION

The present study introduces a deep-learning breast cancer segmentation method for breast ultrasound (BUS) images, utilizing the DDA-AttResUnet model with dual decoder attention and attention mechanisms. Our method achieves a competitive segmentation dice coefficient of 92.92% on the BUSI dataset, a set of non-invasive, safe, cheap, and non-ionizing-radiation ultrasound images. The dual decoding attention and attention mechanism in ResUNet significantly improved segmentation results. This mechanism enhanced localization accuracy by focusing on important regions within feature maps, leading to superior segmentation performance, especially in terms of the dice coefficient. To address the generalizability of the proposed system, we plan to test it on other BUS datasets and investigate its possible refinement. Additional research in breast cancer segmentation from BUS images holds potential implications for better cancer diagnosis and treatment planning.

#### REFERENCES

- [1] WHO. (2023). *Breast Cancer*. Accessed: Jan. 21, 2023. [Online]. Available: <https://www.who.int/news/item>
- [2] F. Islami, K. D. Miller, R. L. Siegel, Z. Zheng, J. Zhao, X. Han, J. Ma, A. Jemal, and K. R. Yabroff, "National and state estimates of lost earnings from cancer deaths in the United States," *JAMA Oncol.*, vol. 5, no. 9, Sep. 2019, Art. no. e191460.
- [3] *Global Breast Cancer Initiative Implementation Framework: Assessing, Strengthening and Scaling-up of Services for the Early Detection and Management of Breast Cancer*, World Health Organization, Geneva, Switzerland, 2023.
- [4] P. Skaane, A. I. Bandos, L. T. Niklason, S. Sebuødegård, B. H. Østerås, R. Gullien, D. Gur, and S. Hofvind, "Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: The Oslo tomosynthesis screening trial," *Radiology*, vol. 291, no. 1, pp. 23–30, Apr. 2019.
- [5] G. Carlino, P. Rinaldi, M. Giuliani, R. Rella, E. Bufi, F. Padovano, C. Ciardi, M. Romani, P. Belli, and R. Manfredi, "Ultrasound-guided preoperative localization of breast lesions: A good choice," *J. Ultrasound*, vol. 22, no. 1, pp. 85–94, Mar. 2019.
- [6] K. Askaripour and A. Zak, "Breast MRI segmentation by deep learning: Key gaps and challenges," *IEEE Access*, vol. 11, pp. 117935–117946, 2023.

- [7] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, and D. Shen, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918.
- [8] K. Drukker, N. P. Grusauskas, C. A. Sennett, and M. L. Giger, "Breast U.S. computer-aided diagnosis workstation: Performance with a large clinical diagnostic population," *Radiology*, vol. 248, no. 2, pp. 392–397, Aug. 2008.
- [9] M. H. Yap, E. Edirisinghe, and H. Bez, "Processed images in human perception: A case study in ultrasound breast imaging," *Eur. J. Radiol.*, vol. 73, no. 3, pp. 682–687, Mar. 2010.
- [10] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "A distinctive approach in brain tumor detection and classification using MRI," *Pattern Recognit. Lett.*, vol. 139, pp. 118–127, Nov. 2020.
- [11] J. Amin, M. Sharif, S. L. Fernandes, S. Wang, T. Saba, and A. R. Khan, "Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model," *Microsc. Res. Technique*, vol. 85, no. 5, pp. 1926–1936, May 2022.
- [12] J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Facial expressions classification and false label reduction using LDA and threefold SVM," *Pattern Recognit. Lett.*, vol. 139, pp. 166–173, Nov. 2020.
- [13] M. Sharif, M. A. Khan, M. Faisal, M. Yasmin, and S. L. Fernandes, "A framework for offline signature verification system: Best features selection approach," *Pattern Recognit. Lett.*, vol. 139, pp. 50–59, Nov. 2020.
- [14] M. J. Umer, J. Amin, M. Sharif, M. A. Anjum, F. Azam, and J. H. Shah, "An integrated framework for COVID-19 classification based on classical and quantum transfer learning from a chest radiograph," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 20, p. e6434, Sep. 2022.
- [15] M. J. Umer, M. Sharif, S. Kadry, and A. Alharbi, "Multi-class classification of breast cancer using 6B-net with deep feature fusion and selection method," *J. Personalized Med.*, vol. 12, no. 5, p. 683, Apr. 2022.
- [16] N. K. Tomar, D. Jha, S. Ali, H. D. Johansen, D. Johansen, M. A. Riegler, and P. Halvorsen, "DDANET: Dual decoder attention network for automatic polyp segmentation," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*. Cham, Switzerland: Springer, Jan. 2021, pp. 307–314.
- [17] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [18] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet—A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020.
- [19] J. Shan, H. D. Cheng, and Y. Wang, "Completely automated segmentation approach for breast ultrasound images using multiple-domain features," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 262–275, Feb. 2012.
- [20] A. Madabhushi and D. N. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 155–169, Feb. 2003.
- [21] R.-F. Chang, W.-J. Wu, W. K. Moon, W.-M. Chen, W. Lee, and D.-R. Chen, "Segmentation of breast tumor in three-dimensional ultrasound images using three-dimensional discrete active contour model," *Ultrasound Med. Biol.*, vol. 29, no. 11, pp. 1571–1581, Nov. 2003.
- [22] H.-H. Chiang, J.-Z. Cheng, P.-K. Hung, C.-Y. Liu, C.-H. Chung, and C.-M. Chen, "Cell-based graph cut for segmentation of 2D/3D sonographic breast images," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2010, pp. 177–180.
- [23] M. Xian, Y. Zhang, and H. D. Cheng, "Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains," *Pattern Recognit.*, vol. 48, no. 2, pp. 485–497, Feb. 2015.
- [24] B. Liu, H. D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images," *Pattern Recognit.*, vol. 43, no. 1, pp. 280–298, Jan. 2010.
- [25] P. Jiang, P. Peng, G. Zhang, E. Cheng, V. Megalooikonomou, and H. Ling, "Learning-based automatic breast tumor detection and segmentation in ultrasound images," in *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2012, pp. 1587–1590.
- [26] Z. Zhuang, N. Li, A. N. Joseph Raj, V. G. V. Mahesh, and S. Qiu, "An RDAU-NET model for lesion segmentation in breast ultrasound images," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0221535.
- [27] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, and C. Chang, "Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model," *Med. Phys.*, vol. 46, no. 1, pp. 215–228, Jan. 2019.
- [28] M. Byra, P. Jarosik, K. Dobruch-Sobczak, Z. Klimonda, H. Piotrkowska-Wroblewska, J. Litniewski, and A. Nowicki, "Breast mass segmentation based on ultrasonic entropy maps and attention gated U-Net," 2020, *arXiv:2001.10061*.
- [29] L. Zhu, R. Chen, H. Fu, C. Xie, L. Wang, L. Wan, and P.-A. Heng, "A second-order subregion pooling network for breast lesion segmentation in ultrasound," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*. Cham, Switzerland: Springer, 2020, pp. 160–170.
- [30] B. Shareef, A. Vakanski, P. E. Freer, and M. Xian, "ESTAN: Enhanced small tumor-aware network for breast ultrasound image segmentation," *Healthcare*, vol. 10, no. 11, p. 2262, Nov. 2022.
- [31] N. S. Punn and S. Agarwal, "RCA-IUnet: A residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging," *Mach. Vis. Appl.*, vol. 33, no. 2, p. 27, Mar. 2022.
- [32] K. Yang, A. Suzuki, J. Ye, H. Nosato, A. Izumori, and H. Sakanashi, "CTG-Net: Cross-task guided network for breast ultrasound diagnosis," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0271106.
- [33] Y. Wu, R. Zhang, L. Zhu, W. Wang, S. Wang, H. Xie, G. Cheng, F. L. Wang, X. He, and H. Zhang, "BGM-Net: Boundary-guided multiscale network for breast lesion segmentation in ultrasound," *Frontiers Mol. Biosciences*, vol. 8, pp. 1–12, Jul. 2021.
- [34] M. Zhang, A. Huang, D. Yang, and R. Xu, "Boundary-oriented network for automatic breast tumor segmentation in ultrasound images," *Ultrasound. Imag.*, vol. 45, no. 2, pp. 62–73, Mar. 2023.
- [35] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, "AAU-Net: An adaptive attention U-Net for breast lesions segmentation in ultrasound images," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1289–1300, May 2023.
- [36] M. J. Umer, M. Sharif, and S. Wang, "Breast cancer classification and segmentation framework using multiscale CNN and U-shaped dual decoded attention network," *Expert Syst.*, vol. 2022, Nov. 2022, Art. no. e13192.
- [37] Y. Lyu, Y. Xu, X. Jiang, J. Liu, X. Zhao, and X. Zhu, "AMS-PAN: Breast ultrasound image segmentation model combining attention mechanism and multi-scale features," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104425.
- [38] S. Zhang, M. Liao, J. Wang, Y. Zhu, Y. Zhang, J. Zhang, R. Zheng, L. Lv, D. Zhu, H. Chen, and W. Wang, "Fully automatic tumor segmentation of breast ultrasound images with deep learning," *J. Appl. Clin. Med. Phys.*, vol. 24, no. 1, Jan. 2023, Art. no. e13863.
- [39] J. Zhang, Z. Zhang, H. Liu, and S. Xu, "SaTransformer: Semantic-aware transformer for breast cancer classification and segmentation," *IET Image Process.*, vol. 17, no. 13, pp. 3789–3800, Nov. 2023.
- [40] T. Chavan and K. Prajapati, "InvUNET: Involved UNET for breast tumor segmentation from ultrasound," in *Proc. Int. Conf. Artif. Intell. Med. Cham*, Switzerland: Springer, 2022, pp. 283–290.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [42] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [43] P. Zhao, J. Zhang, W. Fang, and S. Deng, "SCAU-Net: Spatial-channel attention U-Net for gland segmentation," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 670, Jul. 2020.
- [44] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
- [45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [46] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.
- [47] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [48] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863.



**ASMAA A. HEKAL** (Member, IEEE) received the M.Sc. degree in electronic and communication engineering from Mansoura University, Egypt, in 2021, where she is currently pursuing the Ph.D. degree in electronic and communication engineering. She is an Assistant Lecturer with the Department of Communications Engineering, Modern University for Technology and Information (MTI), Cairo. She is a highly skilled Computer Vision Engineer. Her research interests

include proficiently identifying and implementing pre-processing techniques, conducting thorough testing of deep learning libraries, enhancing model accuracy, devising algorithmic solutions, and contributing to the development of architectures and models for image recognition, artificial intelligence (AI), deep learning (DL), machine learning (ML), image processing, medical data analysis, and computer vision (CV). She has authored several research articles within her areas of expertise.



**AHMED ELNAKIB** (Senior Member, IEEE) is currently an Assistant Professor with the Electrical and Computer Engineering Department, School of Engineering, Penn State Behrend, The Behrend College, Erie, PA, USA. He holds the position of an Associate Professor of electronics and communications engineering with Mansoura University, Mansoura, Egypt. His extensive research portfolio boasts over 50 journal articles, seven book chapters, and more than 50 conference

papers, published in prestigious journals, such as *Scientific Reports* and *IEEE TRANSACTIONS*, alongside presentations at IEEE-affiliated conferences. He serves as a respected reviewer and an editor for leading medical signal analysis journals. He has been honored with numerous accolades, including the prestigious John M. Houchens Prize for the Best Dissertation across all schools at the University of Louisville (UofL). Furthermore, his contributions to science and advanced technology were acknowledged with the State Encouragement Award in Engineering, in 2019, and the Mansoura University Engineering Encouragement Award for Distinguished Young Professors, from 2019 to 2020.



**HOSSAM EL-DIN MOUSTAFA** (Senior Member, IEEE) is currently a Professor with the Department of Electronics and Communications Engineering and the Founder and the Former Executive Manager of the Biomedical Engineering Program (BME), Faculty of Engineering, Mansoura University. His research interests include biomedical imaging, image processing applications, and bioinformatics.



**HANAN M. AMER** received the Ph.D. degree from Mansoura University, Egypt, in 2018. She is currently an Associate Professor with the Electronics and Communications Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt. She is the second Supervisor of more than 35 Ph.D. and master's students. She has published nine articles. Her research interests include signal and image processing, medical image analysis, and lung cancer.

...