Deep Learning for Ultrasound-Based Breast Cancer Early Detection

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Computer Science
in the
College of Graduate Studies
University of Idaho
by

Bryar Shareef

Major Professor: Min Xian, Ph.D.
Co-Major Professor: Aleksandar Vakanski, Ph.D.
Committee Members: Audrey Fu, Ph.D.; Xiaogang Ma, Ph.D.
Department Administrator: Terry Soule, Ph.D.

August 2023

Abstract

Breast cancer is a pervasive health issue that affects millions of women worldwide. Early detection of breast cancer is crucial for reducing mortality and improving patient prognosis. By identifying cancer at an early stage, treatment options can be initiated promptly, leading to more successful outcomes. Breast ultrasound imaging is a valuable tool in the early detection of breast cancer. It offers several advantages, including painless and noninvasive imaging, the absence of ionizing radiation, and affordability. Ultrasound imaging provides detailed visualization of breast tissue, allowing healthcare professionals to identify suspicious lesions, and assess tumor characteristics. Moreover, it is particularly effective in evaluating dense breast tissue, which may pose challenges for other modalities such as mammography. Despite its advantages, the interpretation of breast ultrasound images presents certain challenges. One of the major difficulties is the presence of speckle noise, which can obscure subtle abnormalities and make accurate tumor identification challenging. Additionally, variations in image quality, tumor shapes, and sizes further complicate the analysis. Computer-aided diagnosis (CAD) systems have emerged as crucial tools in breast cancer detection and diagnosis. These systems employ techniques from machine learning and image processing to assist healthcare professionals in analyzing breast ultrasound images. CAD systems can aid in operator-independent tumor segmentation, feature extraction, and precise tumor quantification, thereby enhancing diagnostic accuracy and efficiency. By leveraging the power of artificial intelligence, CAD systems can assist in early cancer detection, reduce false-positive rates, and improve overall patient care.

In this dissertation, I built a suite of deep learning approaches to enhance breast cancer early detection using ultrasound images.

First, I proposed two novel deep learning approaches, Small Tumor-Aware Network (STAN) and Enhanced STAN (ESTAN), to detect and segment small breast tumors. STAN addressed the challenges posed by speckle noise, poor image quality, and variable tumor shapes and sizes in breast ultrasound images. A multiscale feature extraction architecture was proposed to learn and fuse context information at different scales. Building upon the STAN network, the ESTAN model incorporated breast anatomy into STAN to address the aforementioned challenges.

Second, I built a benchmark for BUS image classification that consists of a large public dataset with 3,641 B-mode BUS images, provided open-source code of state-of-the-art approaches, and identified the best strategies for deep learning-based BUS classification. I proposed a comprehensive evaluation methodology that incorporates multiple performance metrics and compares the effectiveness of different classification algorithms.

The benchmark dataset and evaluation framework serve as valuable resources for researchers and practitioners, facilitating the development and assessment of robust classification models.

Third, I proposed a Multitask-Enhanced Small Tumor Aware Network (MT-ESTAN) to perform breast tumor classification and segmentation simultaneously. It incorporates a small-tumor aware network as its backbone, and leverages information from segmentation and classification tasks to enhance the overall performance for breast cancer classification.

Finally, I proposed a hybrid multitask CNN-Transformer network for breast ultrasound tumor classification. The proposed approach combines the strengths of convolutional neural networks (CNNs) and transformer networks to capture both local and global context effectively. The network is trained using a multitask learning framework, simultaneously performing tumor classification and segmentation.

Acknowledgments

I am incredibly grateful to the members of my dissertation committee who have significantly impacted the successful completion of my dissertation.

First, I sincerely appreciate Dr. Min Xian, my major supervisor, for his continuous guidance, invaluable insights, and constant support throughout my journey. His expertise, dedication, and encouragement have been instrumental in shaping the direction of my research and enhancing the quality of my work.

I would also like to extend my gratitude to Dr. Aleksandar Vakanski, my second supervisor, for his continuous support, thoughtful feedback, and valuable contributions to my research. His insightful perspectives and scholarly guidance have immensely enriched my dissertation.

I am grateful to Dr. Marshal and Dr. Fu for their constructive views and feedback, which have significantly improved the overall quality of my research. Their critical insights and expertise have been truly invaluable.

I would like to acknowledge the Department of Computer Science for providing me with a conducive research environment and access to invaluable resources. Their commitment to academic excellence has been instrumental in shaping my academic journey.

I would also like to express my sincere appreciation to my friends and colleagues in the Machine Intelligence and Data Analytics (MIDA) Lab. Their collaboration, stimulating discussions and constant support have contributed significantly to my growth as a researcher. I am grateful for the camaraderie and shared experiences that have made this journey both intellectually rewarding and enjoyable.

This work was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Number P20GM104420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Dedication

I dedicate this work to my beloved wife Hero Salih. Your unwavering support, understanding, and belief in my abilities have been a constant source of inspiration. Your presence by my side has given me the strength to overcome obstacles and persist in my endeavors. I am grateful for your unwavering faith in me and for being my pillar of strength throughout my Ph.D. journey.

To my dear father, Mustafa Shareef, your guidance, wisdom, and endless support have been instrumental in shaping my academic and personal growth. Thank you for standing by me through every step of this journey and being my inexhaustible source of inspiration.

To the loving memory of my late mother, who bravely fought breast cancer until her last breath. Your strength, resilience, determination, and the immense love you showered upon me have shaped the person I am today. Your untimely passing fueled my determination to contribute to the field of early breast cancer detection. I dedicate this research to you in your memory and with deep admiration. Your journey has ignited a fire within me, inspiring me to make a meaningful impact and ensure that others do not suffer the same fate.

Lastly, to my sister, Skala, your love, encouragement, support, and understanding have given me the strength to overcome challenges and stay focused on my goals. I am grateful for your presence in our life and for always being there to lift me.

Table of Contents

List of Tables

List of Figures

Statement of Contribution

This dissertation is submitted to the University of Idaho in partial fulfillment of the requirements for the degree of Doctor of Philosophy. I hereby declare that this work is my own and original work, and has not been previously submitted for obtaining an academic degree. All sources of information used have been appropriately acknowledged.

Chapters 2-5 of this dissertation are the portion of the published research. I would like to state the following contributions that any co-authors have made to the research published in this dissertation:

Chapter 2: Bryar Shareef: conceptualization, methodology, software, validation, formal analysis, investigation, writing original draft, visualization. Aleksandar Vakanski: investigation, writing – review & editing. Min Xian: conceptualization, methodology, investigation, writing – review & editing, supervision.

Chapter 3: Bryar Shareef: conceptualization, methodology, software, validation, formal analysis, investigation, writing original draft, visualization. Min Xian: conceptualization, methodology, investigation, writing – review & editing, supervision. Aleksandar Vakanski: investigation, writing – review & editing.

Chapter 4: Bryar Shareef: conceptualization, methodology, software, validation, formal analysis, investigation, writing original draft, visualization. Min Xian: conceptualization, methodology, investigation, writing – review & editing, supervision. Aleksandar Vakanski: investigation, writing – review & editing. Shoukun Sun: investigation and website development.

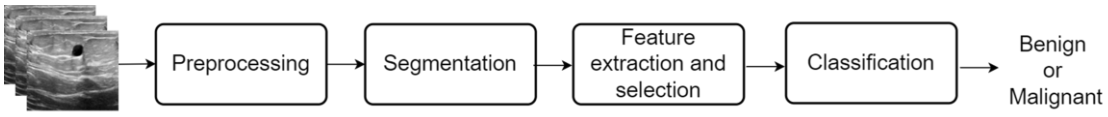
Chapter 5: Bryar Shareef: conceptualization, methodology, software, validation, formal analysis, investigation, writing original draft, visualization. Min Xian: conceptualization, methodology, investigation, writing – review & editing, supervision. Aleksandar Vakanski: investigation, writing – review & editing.

Chapter 1: Introduction

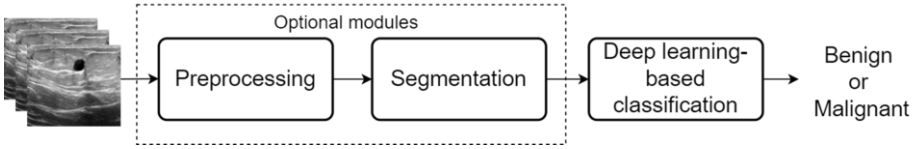
1.1 Background

Breast cancer holds the highest incidence rate among all cancers and remains the primary cause of cancer-related mortality among women across the globe [53]. Early detection plays a crucial role in reducing mortality rates and expanding treatment options. Mammography and ultrasound are the two commonly used imaging modalities for breast abnormality detection, although mammography implementation is limited in many low- and middle-income countries due to infrastructure costs [54]. Additionally, mammography exhibits high false-positive rates in women with dense breasts, leading to increased anxiety and additional biopsy procedures [55]. Studies have shown that ultrasound can detect around 40% more cancer cases than mammography in women with dense breasts [56].

Breast ultrasound (BUS) imaging is a standard and successful clinical procedure because it is painless, noninvasive, nonradioactive, and cost-effective. However, medical analysis of BUS images is challenging due to speckle noise, low contrast, weak boundary, artifacts, and varying tumor shapes and sizes among patients. To address these challenges and aid radiologists in breast tumor diagnosis, deep learning-based computer-aided diagnosis (CAD) systems have been developed.



(a) Conventional CAD systems



(b) Deep learning-based CAD systems

Figure 1-1 Key modules in conventional and deep learning-based CAD systems.

1.2 Computer Aided-Diagnosis (CAD) Systems

Computer-Aided Diagnosis (CAD) systems are tools designed to assist radiologists in the interpretation and analysis of ultrasound images. CAD systems have emerged as a powerful approach for early breast cancer detection, with their ability to automatically learn and extract meaningful features from large amounts of data.

These systems aim to improve diagnostic accuracy by providing additional support and automated analysis of ultrasound images.

Traditional CAD systems for breast cancer ultrasound typically rely on rule-based algorithms and handcrafted features that are manually designed by domain experts. These features capture various characteristics of abnormalities in ultrasound images, such as shape, texture, and intensity. The algorithm analyzes the input ultrasound image by comparing the extracted features against predefined rules or patterns to generate a diagnosis or highlight areas of suspicion, see Figure 1-1 (a).

In contrast, deep learning-based CAD systems for breast cancer ultrasound employ deep neural networks, specifically convolutional neural networks (CNNs), to automatically learn hierarchical representations directly from the raw ultrasound images. Deep learning models are trained on large, annotated datasets, allowing them to automatically extract relevant features without the need for explicit feature engineering, see Figure 1-1 (b).

The primary distinction between traditional CAD systems and deep learning-based CAD systems lies in the feature extraction process. Traditional CAD systems heavily rely on human experts to handcraft features, which can be time-consuming and subjective. Deep learning-based CAD systems offer several advantages over traditional CAD systems. They can learn complex representations directly from the data, potentially capturing subtle patterns and features that may be challenging to identify manually. Deep learning models also demonstrate scalability, as once trained, they can be applied to new datasets without extensive modifications, making them suitable for diverse clinical settings. However, the interpretability and explainability of deep learning models remain ongoing research challenges, whereas traditional CAD systems often provide more transparent decision-making processes based on explicitly defined rules and features.

1.3 Major Challenges

Despite significant advancements in machine learning (ML)-based techniques for enhancing breast ultrasound image analysis and processing, several challenges persist in the development of computer-aided diagnosis systems. The major challenges are summarized as follows:

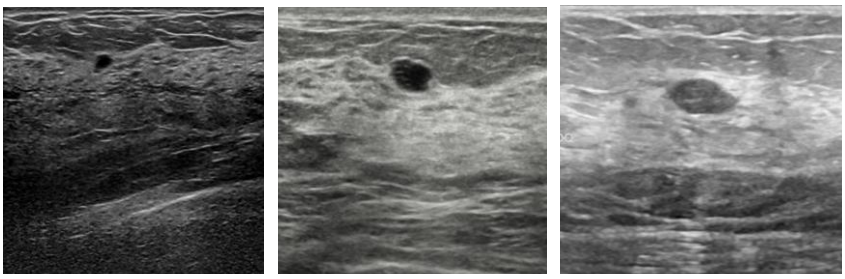


Figure 1-2 Breast ultrasound (BUS) images.

a) **Accurate small tumor detection and segmentation**: Accurately segmenting small tumors in breast ultrasound (BUS) image is a critical challenge in breast ultrasound image analysis. Small tumors can be challenging to differentiate from surrounding normal tissue due to their size, subtle appearance, and variable shapes. Accurately segmenting these small tumors is crucial for assessing tumor size, monitoring tumor progression, and guiding treatment decisions. If small tumors are missed or inaccurately segmented by radiologists, there is a risk of delayed diagnosis and potential implications for early breast cancer detection. Developing advanced segmentation algorithms that can effectively identify and delineate small tumors, even in challenging cases, can help improve the accuracy and reliability of breast ultrasound analysis, reducing the possibility of missing these tumors and facilitating early detection and intervention.

b) **Data availability and annotation**: Collecting a significant amount of annotated breast ultrasound images is time-consuming and resource intensive. Annotating ultrasound images requires expert knowledge and manual efforts to identify and label specific regions of interest, such as tumors or normal tissues. The scarcity of annotated data slows the development and optimization of accurate breast ultrasound image analysis algorithms.

c) **Diversity and representativeness:** Breast ultrasound images can vary significantly regarding patient demographics, imaging protocols, and equipment characteristics, Figure 1-2 shows three BUS images from different datasets. To develop robust algorithms, it is essential to have diverse and representative datasets encompassing various breast tissue types, tumor sizes, and pathologies.

d) **Reproducibility**: Reproducibility ensures the validity and reliability of research findings. By reproducing the results of an approach, researchers can verify the reported findings, build upon existing work, and contribute to cumulative knowledge in the field of breast cancer early detection. However, using private datasets, lack of standardized data preprocessing, and evaluation, in-house source code implementation, and algorithm complexity make it challenging to reproduce results.

e) **Poor generalizability**: Despite their superior performance, deep learning models can sometimes struggle with generalization when the data is acquired from different institutions, patient populations, and imaging protocols.

   The poor generalization of such models occurs due to the challenging nature of breast ultrasound images, inherent limitations of CNN, the sensitivity of these approaches to noise

and artifacts, limited data availability, overfitting and biased learning, Inter- and Intra-variability due to variations in patient characteristics, imaging settings, and pathological conditions.

These challenges present researchers with opportunities to develop innovative solutions. In the dissertation, we focus on these five significant challenges, which are addressed in the following chapters of the study. Chapters 2 and 3 address challenge a, Chapter 4 addresses challenges b, c, and d, and finally, chapter five addresses challenge e.

1.4 Dissertation Objectives and Contributions

The primary goal of this dissertation is to design deep learning approaches to detect, segment, and classify BUS images despite the innate challenging nature of such images.

Our main contributions are summarized as follows:

Contribution 1: Built a deep learning approach named STAN to segment breast tumors of various sizes. We proposed a two-encoder approach that can learn features using different kernel sizes.

Contribution 2: Extended the STAN approach and built an approach named ESTAN that can detect and segment breast tumors effectively. ESTAN designed a new kernel, named row-column-wise kernel, which targets learning from breast anatomy layers.

Contribution 3: Built the first and largest benchmark for breast ultrasound classification. The benchmark datasets consist of 3,641 B-mode images from five public datasets. The benchmark website provides excellent documentation and source code for the researcher in the field to experiment with and compare their results to the top-10 listed approaches.

Contribution 4: Building a multitask approach, namely MT-ESTAN, to perform simultaneous breast tumor segmentation and classification tasks. The approach learns shared features between the two tasks to improve the performance of the main task.

Contribution 5: Building a hybrid multitask learning approach to exploit the advantages of CNN and Vision Transformers in learning local and global features using a multitask learning approach.

Chapter 2: STAN: Small Tumor-Aware Network for Breast Ultrasound Image Segmentation

B. Shareef, M. Xian and A. Vakanski, "Stan: Small Tumor-Aware Network for Breast Ultrasound Image Segmentation," *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 2020, pp. 1-5, doi: 10.1109/ISBI45749.2020.9098691.

2.1 Introduction

According to the National Center for Health Statistics [1], in 2019, the United States is expected to have 891,480 new women cancer cases, where 30% of all cases will be breast cancer. Early detection is the key to improving the survival rate of breast cancer; the five-year relative survival rate is 98% if the breast cancer is detected and treated at the early stages, and only 22% in cases with advanced-stage cancers. Computer-aided diagnosis (CAD) systems have been proposed to detect breast cancer automatically. In these systems, breast tumor segmentation is a key step that helps accurate tumor quantification. A tremendous number of breast tumor segmentation approaches have been proposed in the last two decades; and some approaches have achieved promising overall performance on their private datasets. However, most approaches cannot segment small tumors accurately.

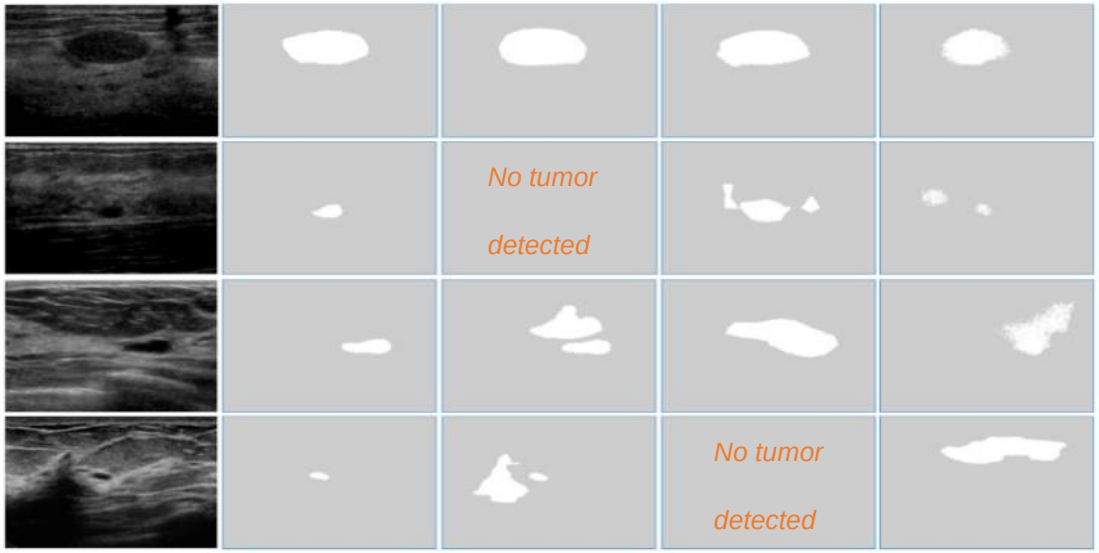


Figure 2-1 Performance of state-of-the-art approaches for segmenting breast tumors with different sizes.

Breast ultrasound (BUS) images are used in this study since ultrasound imaging is noninvasive, painless, nonradioactive and cost-effective.

In the last two decades, breast tumor segmentation has been an active research area. Existing approaches can be classified into traditional approaches and deep learning approaches. Various traditional image processing approaches have been applied to BUS image segmentation, such as thresholding [2-5], region growing [6,7], and watershed [4]. However, the traditional methods are not robust due to poor scalability and sensitivity to noise. Refer to [20] for a detailed review of BUS segmentation approaches.

Deep learning approaches [9-12,21] have recently demonstrated state-of-the-art performance for breast ultrasound segmentation. Cheng et al. [6] employed a stacked denoising auto-encoder (SDAE) to diagnose breast ultrasound lesions and lung CT nodules. The information extension strategy was used in [11], where the wavelet feature was added to the original image to train a fully convolutional network (FCN). Breast anatomy information was applied to the Conditional Random Fields (CRFs) to enhance the segmentation performance. In addition, Huyanh et al. [7] used transfer learning for classification of BUS images, however, the proposed model does not perform tumor segmentation. Similarly, Yap et al. [5] used three different deep learning methods, a patch-based LeNet, a U-Net, and a transfer learning approach with a pre-trained FCN-AlexNet on two different datasets to segment BUS images. However, they failed to achieve good performance for segmenting small tumors. Furthermore, a very deep CNN architecture GoogleNet Inception v2 in [8] is used for the classification task, to distinguish between benign and malignant tumors. The results showed that the CNN model had better, or equal diagnostic performance compared to radiologists. Moreover, in order to focus on regions with high saliency values, the method in [21] integrates radiologists' visual attention for BUS segmentation.

In this paper, our results indicate that the three state-of-art models (FCN-AlexNet, SegNet, and regular Unet) have difficulty in detecting small tumors, as shown in Figure 2-1. We propose a novel architecture based on the core of U-Net architecture to solve the current issue of segmenting small tumors in breast ultrasound images. The method is validated using two public datasets. The experimental results demonstrate enhanced ability of the proposed model for small tumor detection in comparison to existing methods.

2.2 Method

The proposed method is based on one key observation: the size of breast tumors varies dramatically among patients; and existing deep neural networks that use fixed kernel size cannot detect small breast tumors accurately. To overcome this problem, we propose the Small Tumor-Aware Network (STAN)

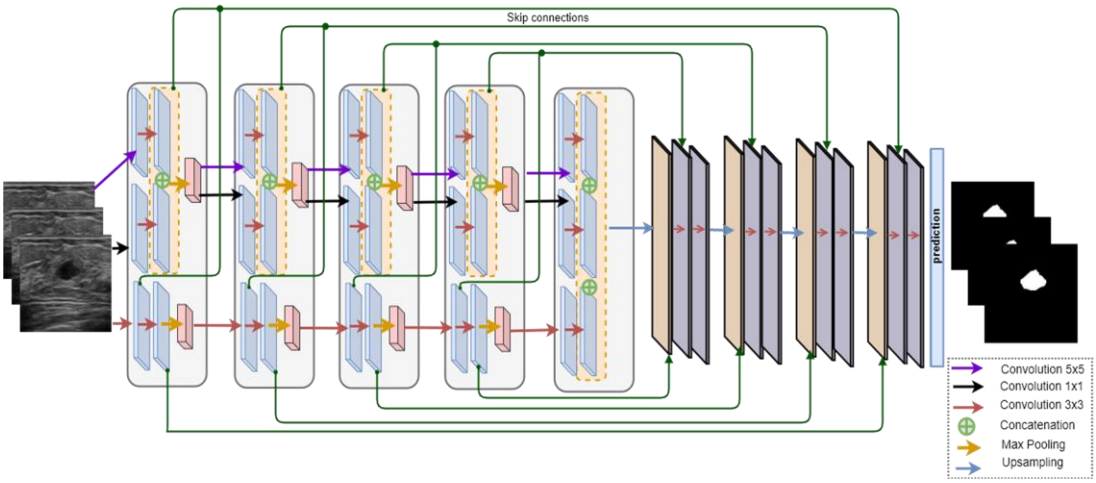


Figure 2-2 The STAN architecture. The blocks do not represent the actual feature maps.

to extract and fuse image context information at different scales. STAN constructs feature maps using kernels with three different sizes at each convolutional layer in the encoder. Such feature maps carry multiscale context information and preserve fine-grained tumor location information. Consequently, STAN improves the performance of breast tumor segmentation, especially for small tumors. Figure 2-2 illustrates the overall architecture of STAN.

2.2.1 STAN Architecture

The size of the receptive field is a crucial issue in deep neural networks, because the output must response to an appropriate size of regions to capture objects with different sizes. There are two main ways to tune the size of the receptive field: 1) downsampling; and 2) stacking more layers. The two methods can only increase the receptive field, and are suitable for segmenting large objects. In BUS image segmentation, a large receptive field will result in high false positives. Therefore, our goal is to avoid stacking too many layers with large kernel size, and design an architecture that has different sizes of the receptive field.

The proposed approach has a similar architecture as the general U-Net: i.e., it contains a contracting and expanding stage with skipping links. Unlike the U-Net architecture, where the contracting stage has only one branch, the proposed network comprises two encoder branches. In addition, the proposed network has three skipping links (the green links in Figure 2-2) between the encoder and decoder blocks, which allows retaining and propagating high-resolution features to the decoder. E.g., for the ith block, we denote the output of the two encoder branches as Ci,1 and Ci,2, and the next block will output

$$C_{i+1,1} = p\left(conv_3\left(conv_3\left(C_{i,1}\right)\right)\right) \tag{2.1}$$

$$C_{i+1,2} = p\left(conv_3\left(conv_1\left(C_{i,2}\right)\right) \oplus conv_3\left(conv_5\left(C_{i,2}\right)\right)\right) \tag{2.2}$$

where $conv_n$ denotes the convolutional operation with kernel size $n \times n$. C0,1 and C0,2 are used to denote an input image to the network, where C0,1 = C0,2; p denotes the max pooling operation; and, for the central layer, C5,1 and C5,2 are

$$\begin{aligned} C_5 = C_{5,1} = C_{5,2} = conv_5\left(conv_5\left(C_{4,1}\right)\right) \oplus \\ conv_1\left(conv_1\left(C_{4,2}\right)\right) \oplus \\ conv_3\left(conv_3\left(C_{4,2}\right)\right) \end{aligned} \tag{2.3}$$

In Eqs. (2.1-2.3), $\oplus$ denotes the concatenation operation. From the blocks one to four, each block applies kernels with three different sizes, that is 1×1, 3×3 and 5×5, and captures image features at three different scales. In general, when the dimensions of the input images to the neural network are reduced extremely via down-sampling layers, the network performs poorly because the network loses vast amount of information, recognized as a representational bottleneck [9]. To solve the representational bottleneck issue, the network-in-network architecture [9] used convolutional kernels of size 1×1 followed by a ReLU layer to introduce more no-linearity. Motivated by this approach, in the second branch of the encoder, we introduced 1×1kernels to increase the representational power of the model.

The original U-Net architecture copies features after the second convolutional layer in the encoder part and concatenates the features to the corresponding layer in the decoder section. In our proposed model, the skipping links involve the output of the first convolution in each layer merged to the result of the first convolution in the corresponding decoder part. In addition, a skipping layer from the merging of the two new layers after the second convolution in the encoder merges to the result of the second convolution in the decoder part. Accordingly, the expanding stage is enriched by fusing feature maps from the blocks in the two encoders. Let Ui (i = 5, 4, 3, 2, 1) be the output of ith up-sampling block; and the output of the next bock is given by

$$U_{i-1} = conv\left(conv\left(DeConv\left(U_i \oplus C_{i-1,1}\right)\right) \oplus conv_5\left(C_{i,1}\right) \oplus C_{i-1,2}\right) \tag{2.4}$$

In Eq.(2.4), U_5 is equal to C_5 from the central layer, and $DeConv$ denotes the deconvolution operation. In addition, since the layer five does not involve pooling, we discarded the pooling layers from the skipping block. The original skipping layers stay the same, where we combine it to the up-sampling layer before the first convolutional layer.

2.2.2 Implementation and Training

The input images and their corresponding ground truths are resized to 256×256. Since the datasets are of small size, we applied image width and height shift to augment the training set. The batch size is 4, and the number of training epochs is set to 50. Adam optimizer [9] is utilized for training the proposed network, and the initial learning rate is set to 0.0001.

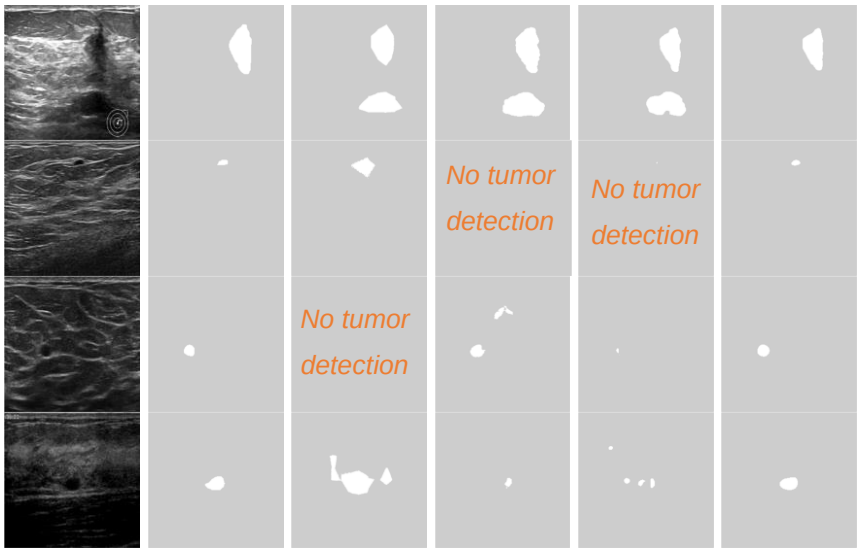
Let $P = \{p_i\}_{i=1}^{N}$ and $G = \{g_i\}_{i=1}^{N}$ be the output of the final pixel-wise sigmoid layer and the ground truth, respectively. The loss function is computed by using discrete dice loss [10]:

$$L_{dice} = 1 - \frac{1 + 2\sum_i^N p_i g_i}{1 + \sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2.5}$$

2.3 Experimental Results

2.3.1 Datasets, Evaluation Metrics, Setup

We use two publicly available datasets to validate the performance of the proposed approach, BUSIS dataset [11] and Dataset B [9]. The BUSIS dataset contains 562 images from three hospitals using GE VIVID 7, LOGIQ E9, Hitachi EUB-6500, Philips iU22, and Siemens ACUSON S2000. The Dataset B has 163 breast ultrasound images, and the UDIAT Diagnostic Centre of the Parc Taul´ı Corporation, Sabadell (Spain) collected the images using Siemens ACUSON Sequoia C512 system with 17L5 linear array transducer.



(a) BUS images (b) Ground Truth (c) FCN-AlexNet (d) SegNet (e) U-Net (f) STAN

Figure 2-3 Small tumor segmentation.

Both area and boundary metrics are used to evaluate the segmentation results. The metrics are true positive ratio (TPR), false positive ratio (FPR), Jaccard index (JI), dice's coefficient (DSC).

2.3.2 Overall Performance

The overall quantitative results are shown in Table 2-1, where the proposed STAN method outperformed the other three approaches in six metrics on the two datasets. FCN-AlexNet, SegNet, and U-Net produced high TPRs on the BUSIS dataset, and FCN-AlexNet and SegNet obtained higher TPRs than the proposed approach on the Dataset B. However, they achieved high TPR at the cost of large false positive ratio (FPR) shown in the fourth column of Table 2-1.

Table 2-1 Segmentation performance of four approaches on two datasets.

Datasets	Methods	TPR	FPR	JI	DSC	AER	AHE	AME
BUSIS	FCN AlexNet	**0.950**	0.336	0.736	0.841	0.386	25.1	7.1
	SegNet	0.938	0.158	0.820	0.895	0.220	21.7	4.5
	U-Net	0.920	0.138	0.825	0.897	0.218	26.8	4.9
	STAN	0.917	**0.093**	**0.847**	**0.912**	**0.176**	**18.9**	**3.9**
Dataset B	FCN AlexNet	**0.868**	1.167	0.469	0.610	1.299	40.8	14.5
	SegNet	0.852	0.834	0.595	0.708	0.982	41.6	11.4
	U-Net	0.776	0.406	0.653	0.745	0.630	39.6	10.8
	STAN	0.801	**0.266**	**0.695**	**0.782**	**0.465**	**35.5**	**9.7**

Figure 2-3 compares the segmentation results of SegNet, FCN-AlexNet, U-Net, and the proposed STAN. Figure 2-3(b) shows the corresponding ground truth of the original BUS images in Figure 2-3(a). As shown in the first row, FCN-AlexNet, SegNet, and U-Net produce high false positives, while the proposed STAN can accurately segment the tumors. In the second row of Figure 2-3, the FCN-AlexNet has high false positives compared to the ground truth; and both the SegNet and U-Net fail to detect the tumor.

2.3.3 Small Tumor Segmentation

In this section, we evaluate the performance of four approaches in segmenting small tumors. The criterion to select small tumors is the length of the longest axis of a tumor region, and the length threshold is set to 120 pixels. The physic sizes of tumors are not used because they are unavailable for most images in the two datasets. 76 and 49 images are selected form the BUSIS and Dataset B, respectively.

Table 2-2 Small Tumor Segmentation.

Dataset	Method	TPR	FPR	JI	DSC	AER	AHE	AME
BUSIS	FCN-AlexNet	**0.947**	0.767	0.603	0.732	0.821	26.3	9.6
	SegNet	0.923	0.251	0.747	0.841	0.328	22.4	6.2
	U-Net	0.920	0.296	0.756	0.843	0.376	44.2	8.3
	STAN	0.902	**0.165**	**0.791**	**0.870**	**0.263**	**21.3**	**5.2**
Dataset B	FCN-AlexNet	**0.868**	1.863	0.353	0.492	1.995	49.2	18.4
	SegNet	0.854	1.452	0.495	0.619	1.598	50.1	14.2
	U-Net	0.768	0.682	0.593	0.681	0.913	43.1	13.8
	STAN	0.814	**0.400**	**0.673**	**0.759**	**0.586**	**35.9**	**11.1**

As shown in Table 2-2, on the two datasets, all metrics except the TPR of the proposed STAN are better than those of FCN-AlexNet, SegNet, and U-Net. The FPR of the FCN-AlexNet on the small dataset (0.767) of is more than twice as its original FPR in Table 2-1(0.336). All other three approach generate high FPRs (FCN-AlexNet: 1.86, SegNet: 1.45 and U-Net: 0.68) for small tumors in the Dataset B. The third and fourth rows of Figure 2-3 show segmentation results of a small tumor, the FCN-AlexNet and U-Net detect no tumor, while the SegNet produced high false positive. In the fourth row, the FCN-AlexNet and U-Net generated high false positive, and the SegNet only found a small part of the tumor.

2.4. Conclusion

In this work, we proposed the Small Tumor-Aware Network (STAN) to overcome challenges in breast tumor early detection. The STAN has two encoder branches that extract and fuse image context information at different scales. The model constructs feature maps using kernels with three different sizes at each convolutional layer. These feature maps carry multiscale context information and preserve fine-grained tumor location information. The proposed STAN achieved the state-of-the-art overall performance on two public datasets, and outperformed the other three segmentation approaches in segmenting small tumors. In the future, we will focus on improving the robustness of the proposed STAN.

Chapter 3: ESTAN: Enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation

Shareef, B., Vakanski, A., Freer, P.E., Xian, M. ESTAN: Enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation. *Healthcare* 2022, *10*, 2262. https://doi.org/10.3390/healthcare10112262

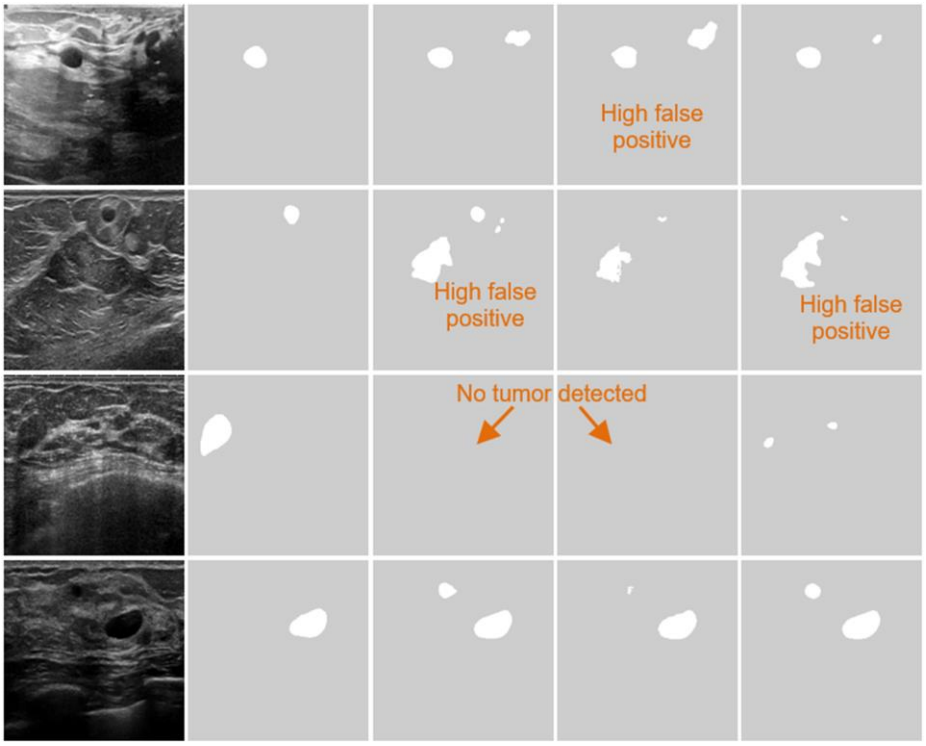
3.1. Introduction

Breast ultrasound (BUS) imaging is an effective screening method due to its painless, noninvasive, nonradioactive, and cost-effective nature. BUS image segmentation aims to extract tumor region(s) from normal breast tissues in images. It is an essential step in BUS computer-aided diagnosis (CAD) systems. However, because of the speckle noise, poor image quality, and variable tumor shapes and sizes, accurate BUS image segmentation is challenging.

According to the National Cancer Institute, in the United States, the relative survival is 99% if breast cancer is detected and treated at the early stages, and only 27% if cancer has spread to other organs of the body [12]. Early detection of breast tumors is the key to reducing the mortality rate. However, in the early stages, most tumors are small and occupy a relatively small region in BUS images. It is challenging to distinguish them from normal breast tissues. Therefore, accurate detection of small tumors is critical for breast cancer early detection and can improve clinical decisions, treatment planning, and recovery.

The approaches of BUS image segmentation can be classified into traditional approaches and deep learning-based approaches. Numerous traditional approaches have been used for BUS image segmentation, such as thresholding [13-18], region growing [3][19], and watershed [4][20]. Despite their simplicity, these methods require knowledge and expertise in extracting features, and they are not robust due to poor scalability and high sensitivity to noise. Refer to [21] for a comprehensive review of BUS image segmentation.

Recently, several deep learning approaches [5, 22-33] have been developed for BUS image segmentation; Table 3-1 lists the most recent deep learning approaches for BUS image segmentation.



(a) BUS Images (b) GT (c) DenseU-Net (d) CE-Net (e)

Figure 3-1 Performance of state-of-the-art approaches for segmenting breast tumors with different sizes. GT: Ground truth.

Huang et al. [22] proposed a fuzzy fully convolutional network to perform BUS image segmentation. Fuzzy logic is adopted to solve the uncertainty issue in the BUS images and feature maps. Contrast enhancement and wavelet features were applied as preprocessing techniques to augment the training data. The augmented training image set and features from convolutional layers were transformed into a fuzzy domain by a fuzzy membership function. The context information and the human breast structure were integrated into Conditional Random Fields (CRFs) to enhance the segmentation results. Yap et al. [5] evaluated the performance of three different deep learning approaches: a patch-based LeNet, a U-Net, and transfer learning with a pretrained AlexNet on two BUS datasets (Dataset A and Dataset B). The transfer learning AlexNet outperformed all others on Dataset A for true positive and F-measure metrics and patch-based LeNet achieved the best results on Dataset B for false positive per image metric. Although the results show that the different deep learning approaches designed for other tasks can be adopted and trained on BUS datasets, all the approaches could not achieve the best results for all the evaluation metrics on both datasets. Amiri et al. [23] studied transfer learning and the significance of fine-tuning configurations of U-Net architecture to solve the issue of scarce ultrasound

Table 3-1 Deep learning-based bus segmentation approaches.

Article	Year	Method	Dataset Size	Evaluation Metrics
Huang *et al.* [22]	2018	FCN + Wavelet features + CRFs	325	TPR, FPR, JI
Shareef *et al.* [33]	2020	U-Net + Two encoders	725	TPR, FPR, JI, DSC, AER, MAE, HD
Yap *et al.* [5]	2018	Patch-based LeNet, U-Net, and AlexNet	469	TPR, FPR, F1
Ameri *et al.* [35]	2020	Transfer learning	163	DSC
Nair *et al.* [24]	2020	Deep Neural Networks + Two Decoders +Simulated Data	22230	DSC
Zhuang *et al.* [25]	2019	U-Net+ Attention gate	1062	TPR, Sp, F1, Pr, JI, Acc, DSC, AUC
Hu *et al.* [26]	2019	Dilated FCN + Active contour model	570	DSC, MAD, and HD
Vakanski *et al.* [27]	2020	U-Net + Attention blocks	510	TPR, FPR, DSC, JI, Pr, AUC-ROC
Byra *et al.* [28]	2020	U-Net + Attention gate + Entropy maps	269	DSC, JI
Moon *et al.* [14]	2020	Ensemble CNNs	246	TPR, FPR
Lee *et al.* [30]	2020	U-Net + Channel attention module	163	FPR, F1, JI, AUC, Pr, Sp, TPR
Chen *et al.* [31]	2022	U-Net + Bidirectional attention + refinement residual net	780	Acc, DSC, Sens, Sp, Pr, JI
Hussain *et al.* [32]	2022	U-Net + level set	349	Acc, DSC, JI

*TPR: true positive rate, FPR: false positive rate, JI: Jaccard indices, IoU: intersection over union, Acc: Accuracy, Pr: precision, Sp: specificity, MCC: mattews correlation coefficient, AUC: area under curve, AER: area error rate, MAE: mean area error, HD: average Hausdorff distance, DSC: dice similarity coefficient, CRFs: conditional random fields, and FCN: fully convolutional network.

image data. Fine-tuning the shallow layers of U-Net for small BUS datasets achieved the best results; however, there is no significant difference in fine-tuning the whole network or shallow layers for large BUS dataset. Refer to [21][34] for more deep learning approaches for medical image segmentation.

In addition, Nair et al. [24] proposed a DNN with two decoders to create BUS images and segmentation masks from raw single plane wave channel data. This approach shows promising results where both the segmentation masks and B-mode images were generated in a single network using raw data. Zhuang et al. [24] proposed an RDAU-Net model, based on U-Net architecture, to perform the tumor segmentation task on BUS images. The dilated residual blocks and attention gates were used to replace the basic blocks and original skip connections in U-Net, respectively. The RDAU-Net design improves the overall sensitivity and accuracy of the model. Similarly, Hu et al. [26] proposed a DFCN method that combines the dilated fully convolution network with a phase-based active contour (PBAC) model to automatically segment breast tumors. The DFCN with PBAC network is more robust to noise and blurry boundaries, and successfully segments tumors with a large volume of shadows.

Moreover, Vakanski et al. [27] integrated radiologists' visual attention with a U-Net model to perform BUS segmentation task. The model designs attention blocks to ignore regions with low saliency and emphasize more on regions with high saliency. This study outperformed the U-Net model and has successfully combined prior knowledge information into a convolutional neural network. Byra et al. [36] proposed a deep learning segmentation approach for BUS images based on entropy parametric maps with the attention-gated U-Net network. The model achieved a good improvement; however, there are insufficient results and analysis to show the significance of entropy maps.

Furthermore, Moon et al. [14] proposed an ensemble CNN architecture for a CAD system comprising multi-models trained on original BUS images, segmented image tumors, tumor masks, and fused images. The fused images were prepared by combining an original image, segmented tumor, and tumor shape information (TSI). The results show that the fused images achieved the best results among all others, and the study provides a clear guide to choosing an approach for a specific dataset size. Lee et al. [30] proposed a channel attention module with multi-scale grid average pooling for segmenting BUS images. The approach utilizes both local and global information and achieves good overall segmentation performance. Chen et al. [31] proposed bidirectional attention and refine network that they added on top of the U-net to accurately segment breast lesions. However, training such a network on a small dataset is challenging to deal with overfitting/underfitting issues. These methods achieved good overall performance. However, as shown in Figure 3-1, they failed to achieve good performance for segmenting small tumors. First, these methods were designed to improve the overall performance using general-purpose square kernels which were developed for learning features in natural images. Second, all currently available BUS datasets are small, and most deep learning-based approaches require a large and high-quality training set.

We aim to solve the challenge of small tumor segmentation in BUS images. The work is inspired by current progress in small object detection and/or segmentation which is an important task in computer vision, as it forms the foundation of many image-related tasks, such as remote sensing, scene understanding, object tracking, instance and panoptic segmentation, aerospace detection, and image captioning. Chen et al. [37] proposed an augmented technique for the R-CNN algorithm with a context model and small region proposal generator; which was the first benchmark dataset for small object detection. Krishna et al. [38] designed a Faster R-CNN model with a modified upsampling technique to improve the performance of small object detection. Guan et al. [39] proposed a semantic context aware network (SCAN), which integrates a location fusion module and context fusion module to detect semantic and contextual features. The DenseU-Net architecture was proposed by Dong [40] for semantic segmentation of small objects in urban remote sensing images.

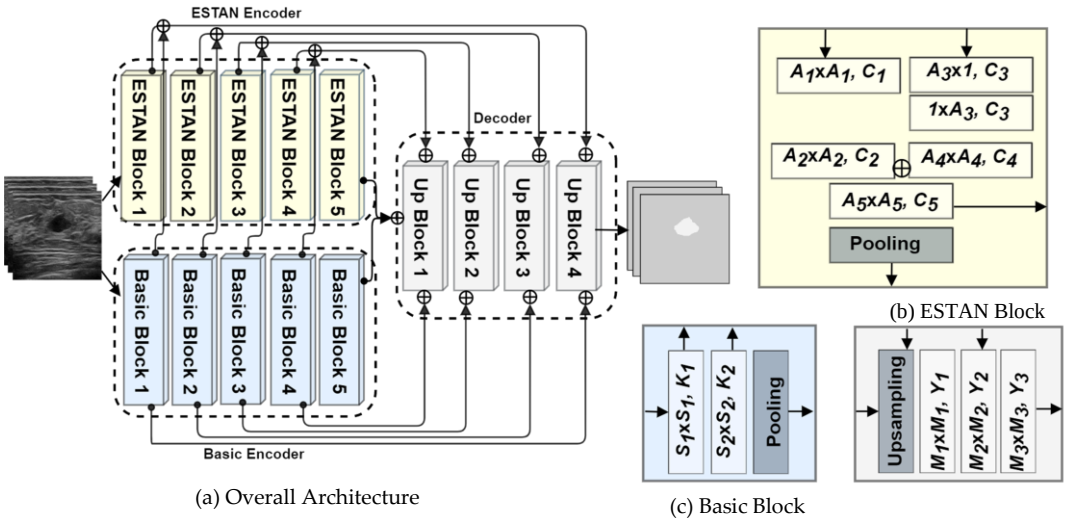
It uses residual connections and a weighted focal loss function with median frequency balancing to improve the performance of small object detection.

To the best of our knowledge, STAN [33] was the first deep learning architecture designed specifically for small tumor segmentation. Three skip connections and two encoders were employed to extract multi-scale contextual information from different layers of the contracting part. STAN outperformed other deep learning approaches for segmenting small tumors in BUS images. However, its average FPR on small tumors is much larger than the FPR on large tumors. In this paper, we extend STAN and propose a new architecture, namely Enhanced Small Tumor-Aware Network or ESTAN, to achieve robust segmentation for tumors of different sizes. The new architecture has two encoder branches. The basic encoder has five blocks and learns features at different scales. The ESTAN encoder applies row-column-wise kernels to adapt to the breast anatomy during the feature learning. Specifically, the human breast anatomy consists of four main layers: skin, premammary (subcutaneous fat), mammary, and retromammary layers [41] (Figure 3-3). Each layer is characterized by a distinct texture and corresponding echo patterns in ultrasound images. The tissue layers in BUS images appear vertically stacked, with similar echo patterns propagating horizontally across images. Breast pathology originates predominantly in the mammary layer. The row-column-wise kernels were designed to learn the breast tissue structure and thus improve detecting small tumor representations in BUS images. In the decoder, each block has three skip connections that fuse rich contextual features from the two encoders. The contextual features are robust to different tumor sizes and help distinguish tumor regions from normal regions.

The rest of the paper is organized as follows: Section 3.2 presents the proposed architecture and implementation details; Section 3.3 demonstrates experimental results; and Section 3.4 provides the conclusion.

3.2 Proposed Method

In this section, we introduce the proposed Enhanced Small Tumor-Aware Network (ESTAN) for solving the issue of small tumor segmentation in BUS images. ESTAN builds upon two observations: 1) BUS images contain tumors of a broad range of sizes, and current state-of-the-art approaches have poor performance on segmenting small tumors; and 2) the current deep learning-based approaches used square-shape kernels and have difficulty utilizing context information of BUS images, e.g., breast tissue anatomy. To alleviate these challenges, we propose ESTAN to extract and fuse image context information at different scales. ESTAN constructs feature maps using both square and large row-



(a) Overall Architecture (b) ESTAN Block (c) Basic Block

Figure 3-2 ESTAN architecture. ⊕ is the concatenation operator, Ai, Si, Mi, denote kernel sizes, and Ci, Ki, Yi define number of kernels.

column-wise kernels. These feature maps transmit multi-scale context information and preserve fine-grained tumor location information. Therefore, the new design enables ESTAN to accurately segment breast tumors of different sizes, and it is especially efficient with small size tumors. ESTAN consists of two encoders and one decoder where each decoder block (UP Block) receives three skip connections. The overall architecture of the proposed approach is shown in Figure 3-2.

3.2.1 Basic Encoder

The basic encoder downsamples the input feature maps to extract low-level spatial and contextual information. Both convolution and pooling operations with strides greater than 1 are employed for downsampling the feature maps in the encoder blocks. The basic encoder comprises of five blocks, where each block contains two convolutional layers and a max pooling layer; except the fifth block, which has no pooling layer. The basic blocks in the encoder are different from the original U-Net [42] encoder blocks since the new architecture uses two skip connections to copy feature maps from the encoder blocks to the corresponding upsampling layers in the decoder module. Figure 3-2(c) illustrates the architecture of the basic encoder. Let denote the input images as X ∈ $R^{h\times w\times c}$, where h, w, and c are the height, width, and number of channels, respectively.

Let f be the convolution function for square kernels, K_i be the number of kernels and S_i be kernel size in the ith convolution layer, followed by a rectified linear unit (ReLU) activation function. The output of the jth block of the basic encoder is defined by

$$B_j = \phi\left(f_{S_2,K_2}\left(f_{S_1,K_1}(X)\right)\right) \tag{3.1}$$

where B_j is the output of a given block, and ϕ is the pooling operatio n the jth block. Additionally, the kernel size S_1 and S_2 in Basic Block 1, 2, 3, 4, and 5 are all set to 3. The number of kernels K_1 and K_2 in Basic Block 1, 2, 3, 4, and 5 have values 32, 64, 128, 256, and 512, respectively.

3.2.2 ESTAN Encoder

The receptive field in CNNs has an important role in building effective feature maps. It defines the input image region that produces the output features, and image regions outside the receptive field of a feature will not contribute to the computation of the feature. To ensure the coverage of all relevant image regions and achieve enhanced performance, many dense prediction tasks used large receptive fields [43][44]. There are several techniques for increasing the size of the receptive field such as stacking more layers, sub-sampling, and dilated convolutions [45]. However, in BUS images, a large receptive field can result in poor performance for small tumor segmentation [33]. The goal of the ESTAN encoder is to effectively produce feature maps and avoid the large receptive field.

STAN [33] proposed a two-encoder architecture and applied kernels of sizes 1×1, 3×3, and 5×5. The small kernel size can avoid a large receptive field. The two encoders fused contextual information at different scales by producing features using different sizes of receptive fields. This design improved the overall performance for small breast tumor segmentation. However, STAN produced high false positives for some BUS images with small tumors.

To overcome this problem, we redesigned the encoder by applying row-column-wise kernels. The small square kernels in STAN constructed feature maps using only square image regions. The motivation for the design is because BUS images are composed of vertically stacked tissue layers (Figure 3-3). Applying row-column-wise kernels in CNNs can avoid calculating features using image regions from multiple anatomical layers and produce more accurate and meaningful feature maps. In addition, in this study, ESTAN is compared to nine state-of-the-art approaches on three datasets, while STAN was compared with only three state-of-the-art approaches on two datasets.

ESTAN encoder comprises five blocks, named ESTAN blocks, which are parallel with the basic encoder blocks. Each block has four square kernels and two row-column-wise kernels in two parallel branches. Such kernels can efficiently extract contextual and fine-grained details of small tumors in the BUS images.

Furthermore, ESTAN blocks add one extra non-linearity to each encoder block. Figure 3-2(b) illustrates the design of each ESTAN block. Let C_i be the number of kernels, and A_I be the kernel size. The output of j-th ESTAN block is defined by

$$E_j = \phi\left(f_{A_5,C_5}\left(f_{A_2,C_2}\left(f_{A_1,C_1}(X)\right) + f_{A_4,C_4}\left(h_{1,A_3,C_3}\left(h_{A_3,1,C_3}(X)\right)\right)\right)\right) \quad (3.2)$$

where E_j is the output of the jth ESTAN block, and ϕ is the pooling operatio h is the row-column-wise convolution function followed by a rectified linear unit (ReLU) activation function with the size of $A_3 \times 1$ and $1\backslash timesA_3$, respectively. The size of A_3 in ESTAN Block 1, 2, 3, 4, and 5 are 15, 13, 11, 9, and 7, respectively. The size of A_5 in ESTAN Block 2 and 5 is 5, and in the rest is 1. Furthermore, block 5 has no pooling operation for both encoders. Moreover, the number of kernels (C_I) in each ESTAN Block 1, 2, 3, 4, and 5 have values 32, 64, 128, 256, and 512, respectively.

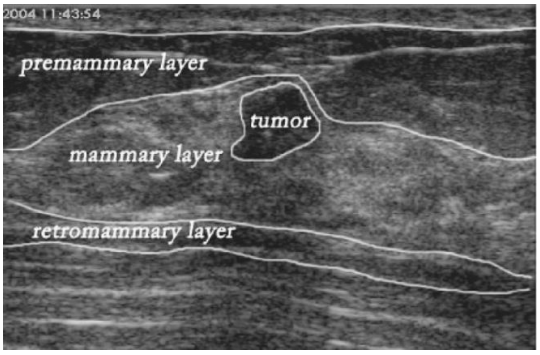


Figure 3-3 Major breast layers of a sample BUS image.

3.2.3 Decoder and Skip Connections

The decoder module comprises four upsampling blocks, where each has one upsampling layer followed by three convolution layers. Unlike the U-Net architecture, where the decoder has two convolution layers, the ESTAN adds an additional kernel after the first convolution kernel to control the post concatenation channels. Let f be the convolution function followed by a rectified linear unit (ReLU) activation function, Y_i be the number of kernels, and M_I be the kernel size. The output of the jth block of the decoder is defined by

$$U_j = f_{M_3,Y_3}\left(f_{M_2,Y_2}\left(f_{M_1,Y_1}(\Psi)\right)\right) \quad (3.3)$$

where Ψ is the upsampling layer. Kernel sizes M_1 and M_3 in all blocks are 3 and M_2 in blocks 1,2, and 3 is 1, and M_2 in block 4 is 5. In addition, $Y_1, Y_2,$ and Y_3, which represent the number of kernels in j-th Up Block has the same values in each block, and their values are 256, 128, 64, and 32 in Up Block 1, 2, 3, and 4, respectively.

We have introduced three skipping connections to copy feature maps at different scales from both encoders to the decoder. The first two skip connections come from combining the result of f_{S_1,K_1}

in the basic encoder and the result of f_{A_5,C_5} in the ESTAN encoder concatenates to the upsampling layer. The second skip connection that comes from the result of f_{S_2,K_2} combines to the f_{M_2,Y_2} in the decoder part. Afterward, the output layer utilizes a 1×1 convolution layer followed by a sigmoid activation function to predict the final outputs. Figure 3-2(d) illustrates the decoder module.

3.3 Experimental Results

3.3.1 Datasets, Evaluation Metrics and Setup

We use three public BUS datasets: BUSIS [21][17][46][47], BUSI [47] and Dataset B [48]. The BUSIS dataset contains 562 images collected from three hospitals using GE VIVID 7, LOGIQ E9, Hitachi EUB-6500, Philips iU22, and Siemens ACUSON S2000. The BUSIS dataset includes 306 benign and 256 malignant breast ultrasound images. The BUSI dataset is from Baheya Hospital for Early Detection & Treatment of Women's Cancer in Egypt using the LOGIQ E9 ultrasound system and the LOGIQ E9 Agile ultrasound system with ML6-15-D Matrix linear probe transducers. The BUSI dataset has 780 images, of which there are 133 normal, 487 benign, and 210 malignant images collected from 600 women patients aged 25 to 75 years old. In addition, radiologists from Baheya Hospital reviewed and modified the ground truth masks. The Dataset B has only 163 breast ultrasound images, and the UDIAT Diagnostic Centre of the Parc Taul'ı Corporation, Sabadell (Spain) collected the images using a Siemens ACUSON Sequoia C512 system with a 17L5 linear array transducer (8.5 MHz). Dataset B consists of 53 malignant, and 110 benign images from different women with a mean image size of 760 × 570 pixels. The Dice loss [48] function is used in this work.

The tumor size is an important variable, and Figure 4 illustrates the histograms of tumor size distributions of the three datasets based on their original resolution. The physical sizes of most tumors in the three datasets are unavailable; therefore, we define the tumor size as the length (in pixels) of the longest axis of a tumor region in the original BUS image. The distributions of BUSI and Dataset B show positive skewness where many tumors are smaller than 150 pixels. The BUSI dataset has more large tumors compared to the other datasets, and the sizes of most tumors are between 150 and 250 pixels. In addition, the images in the BUSIS dataset were collected with five different BUS workstations; thus, the image quality has large variations. To evaluate the segmentation results, both area and boundary metrics are employed. The metrics are true positive rate (TPR), false positive rate

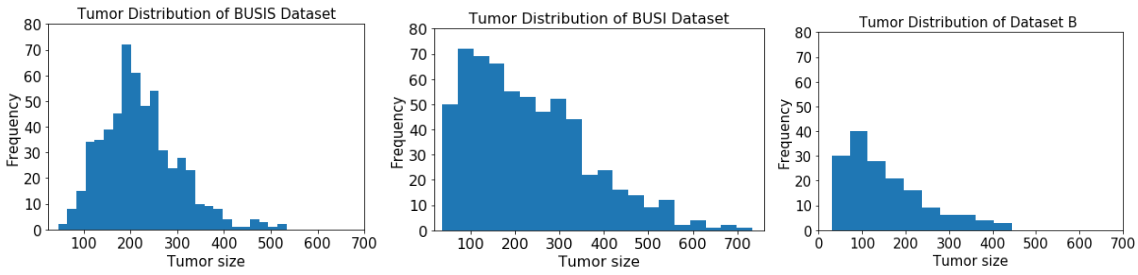


Figure 3-4 Histogram of tumor size (number of pixels) distribution per dataset.

(FPR), Jaccard index (JI), Dice similarity coefficient (DSC), area error rate (AER), Hausdorff distance (HD), and mean absolute error (MAE). For detailed information about the seven metrics, refer to [21]. We perform five-fold cross-validation individually for each dataset to evaluate the test performance of all methods, and the input image size is 256 × 256 pixels for all the approaches.

In this study, we compare the proposed method with nine state-of-the-art approaches: AlexNet [49], SegNet [50], U-Net [42], CE-Net [51], MultiResUNet [52], RDAU-Net [25], SCAN [39], DenseU-Net [40], and STAN [33]. These approaches have different backbone networks and different training strategies. We employ a transfer learning technique for AlexNet, which is pretrained on ImageNet. SegNet, U-Net, CE-Net, MultiResUNet, RDAU-Net, SCAN, and DenseU-Net are trained from scratch.

3.3.2 Overall Performance

In this section, we compare the proposed approach with AlexNet, SegNet, U-Net, CE-Net, MultiResUNet, RDAU-Net, SCAN, DenseU-Net, and STAN. The results are shown in Figure 3-6 and Table 3-2.

Figure 3-6 shows the segmentation results of four sample BUS images. In the first row, the tumor in the BUS image is small, and AlexNet, U-Net, MultiResUNet, SCAN, and DenseU-Net have poor segmentation performance. In the second and third samples (2nd and 3rd rows), all approaches, except the proposed ESTAN, produce high false positives, which demonstrates that they have difficulty distinguishing tumor regions from tumor-like regions. In Figure 3-6(k), STAN can segment small tumors accurately but still produces false tumor regions. Figure 3-6(l) shows that ESTAN segments the four images accurately without any false tumor regions.

Table 3-2 presents the quantitative results of all approaches on the three datasets. The proposed ESTAN achieved the best overall performance on all three datasets. AlexNet and SegNet obtained high TPRs, but at the cost of high FPRs.

To investigate the statistical significance of all the proposed approaches, Wilcoxon signed-rank test was employed to compare ESTAN against all other approaches for FPR, JI, DSC, AER, HE, and MAE metrics on the three datasets. The significance level is defined as p-value < 0.05. The obtained p-values from the Wilcoxon signed-rank test were corrected using Holm-Bonferroni method for multiple comparisons. The results indicate a statistically significant difference for the six metrics on the three datasets, except for the cases that are marked with (*) in Table 3-2.

STAN has 22 million parameters while ESTAN uses 30 million. The average training time of STAN for each fold is 21, 24, 23 minutes for DSB, BUSI, and BUSIS datasets, respectively, while ESTAN takes 34, 32, 34 minutes for training DSB, BUSI, and BUSIS datasets, respectively, with batch size of 4 and maximum 50 epochs. The average testing time of STAN for segmenting each image on the trained models is 150, 66, 61 milliseconds for DSB, BUSI, and BUSIS datasets, respectively, while ESTAN needs 205, 80, 85 milliseconds for segmenting each image.

3.3.3 Small Tumor Segmentation

The physical size for all images of the three datasets is not available. Therefore, the length of the longest axis of a tumor region in the original BUS image (non-resized) is chosen to be a criterion to select small tumors, and the length threshold is set to 120 pixels. BUSIS, BUSI, and Dataset B contain 49, 151, and 76 small tumors, respectively. Figure 3-5 illustrates the FPR comparison between the overall and small tumor segmentation. All ten approaches have higher FPR for small tumors on BUSIS and for both overall and small tumor segmentation. Table 3-3 shows all-inclusive results of all approaches on the three datasets using the selected seven quantitative metrics. ESTAN outperforms all other nine approaches for small tumor segmentation on the three datasets. AlexNet Dataset B datasets. The FPR of AlexNet increased dramatically for small tumor segmentation. The ESTAN approach is superior in comparison to all nine approaches and achieves the lowest false positive and SegNet obtain high TPRs, but at the cost of high FPR.

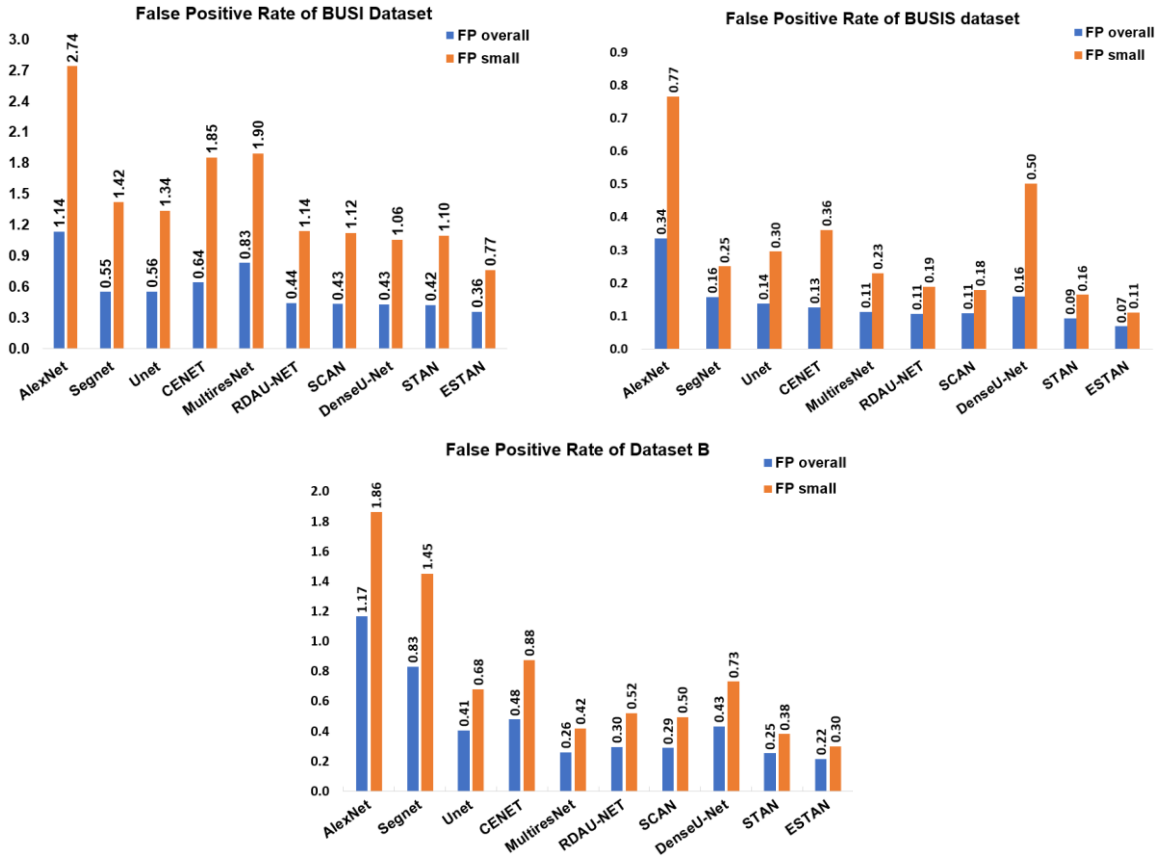


Figure 3-5 False positive rates of overall and small tumor segmentation on the three datasets.

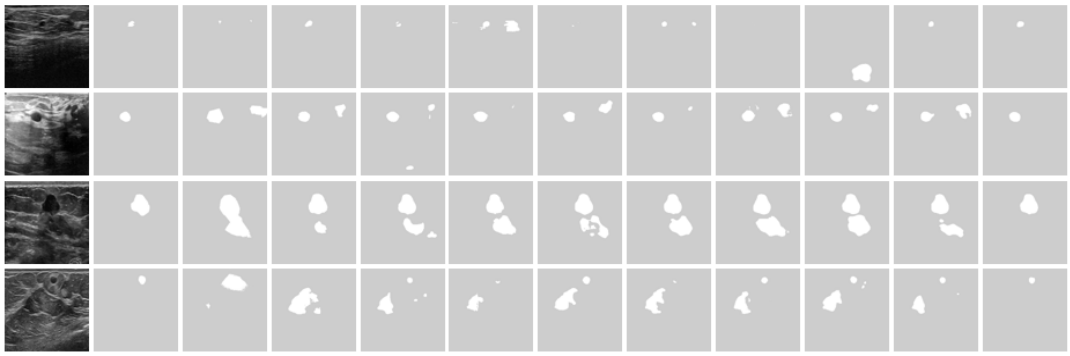


Figure 3-6 Tumor segmentation examples. (a) BUS Image, (b) ground truth, (c) AlexNet, (d) SegNet, (e) U-Net, (f) CE-Net, (g) MultiResUNet, (h) RDAU-Net , (i) SCAN, (j) DenseU-Net, (k) STAN, and (l) ESTAN.

Table 3-2 Overall performance

Datasets	Methods	TPR	FPR	JI	DSC	AER	HD	MAE
BUSIS [35]	AlexNet	**0.95**	0.77	0.60	0.73	0.82	26.3	9.6
	SegNet	0.92	0.25	0.75	0.84	0.33	22.4	6.2
	U-Net	0.92	0.30	0.76	0.84	0.38	44.2	8.3
	CE-Net	0.91	0.36	0.73	0.82	0.46	34.8	9.0
	MultiResUNet	0.91	0.23	0.77	0.84	0.33	27.7	8.5
	RDAU-NET	0.89	0.19	0.78	0.86	0.30	22.0	7.3
	SCAN	0.88	0.18	0.77	0.85	0.30	27.4	6.2
	DenseU-Net	0.90	0.50	0.72	0.81	0.60	34.5	8.2
	STAN	0.90	0.17	0.79	0.87	0.26	21.3	5.2
	ESTAN	0.90	**0.11**	**0.82**	**0.89**	**0.21**	**14.9**	**3.0**
Dataset B [16]	AlexNet	**0.87**	1.86	0.35	0.49	2.00	49.2	18.4
	SegNet	0.85	1.45	0.50	0.62	1.60	50.1	14.2
	U-Net	0.77	0.68	0.59	0.68	0.91	43.1	13.8
	CE-Net	0.72	0.88	0.53	0.63	1.15	50.0	14.4
	MultiResUNet	0.79	0.42	0.62	0.71	0.62	39.3	11.5
	RDAU-NET	0.78	0.52	0.62	0.71	0.73	34.1	8.8
	SCAN	0.75	0.50	0.61	0.70	0.74	48.7	11.2
	DenseU-Net	0.70	0.73	0.54	0.63	1.02	56.0	20.0
	STAN	0.81	0.40	0.67	0.76	0.59	35.9	11.1
	ESTAN	0.85	**0.30**	**0.72**	**0.80**	**0.44**	**21.5**	**6.3**
BUSI [31]	AlexNet	**0.94**	2.74	0.41	0.56	2.81	52.5	15.4
	SegNet	0.81	1.42	0.55	0.66	1.61	52.1	16.6
	U-Net	0.86	1.34	0.63	0.73	1.48	61.0	13.0
	CE-Net	0.83	1.86	0.59	0.69	2.03	50.9	13.3
	MultiResUNet	0.85	0.83	0.67	0.76	0.99	34.7	10.6
	RDAU-NET	0.87	0.99	0.68	0.77	1.13	33.9	9.9
	SCAN	0.80	1.13	0.63	0.73	1.33	42.4	12.5
	DenseU-Net	0.81	1.06	0.65	0.73	1.26	40.9	13.7
	STAN	0.86	1.10	0.67	0.76	1.25	49.2	11.3
	ESTAN	0.89	**0.77**	**0.72**	**0.81**	**0.88**	**24.2**	**6.1**

Table 3-3 Performance of small tumor segmentation.

Datasets	Methods	TPR	FPR	JI	DSC	AER	HD	MAE
BUSIS [35]	AlexNet	**0.95**	0.34	0.74	0.84	0.39	25.1	7.1
	SegNet	0.94	0.16	0.82	0.90	0.22	21.7	4.5
	U-Net	0.92	0.14	0.83	0.90	0.22	26.8	4.9
	CE-Net	0.91	0.13	0.83	0.90	0.22	21.6	4.5
	MultiResUNet	0.93	0.11	0.84	0.91	0.19	18.8	4.1
	RDAU-NET	0.91	0.11	0.84	0.91	0.20	19.3	4.1
	SCAN	0.91	0.11	0.83	0.90	0.20	26.9	4.9
	DenseU-Net	0.91	0.16	0.81	0.88	0.25	25.3	5.5
	STAN	0.92	0.09	0.85	0.91	0.18	18.9	3.9
	ESTAN	0.91	**0.07**	**0.86**	**0.92**	**0.16**	**16.4**	**3.2**
Dataset B [16]	AlexNet	**0.87**	1.17	0.47	0.61	1.30	40.8	14.5
	SegNet	0.85	0.83	0.60	0.71	0.98	41.6	11.4
	U-Net	0.78	0.41	0.65	0.75	0.63	39.6	10.8
	CE-Net	0.74	0.48*	0.61	0.72	0.74	40.1	10.5
	MultiResUNet	0.79	0.26	0.66	0.75	0.48	37.1	10.7
	RDAU-NET	0.78	0.30*	0.67	0.77	0.52	32.4	8.3
	SCAN	0.75	0.29*	0.65	0.74	0.54	43.7	9.9
	DenseU-Net	0.71	0.43	0.60	0.69	0.72	48.9	15.5
	STAN	0.80	0.27*	0.70*	0.78	0.47*	35.5	9.7*
	ESTAN	0.84	**0.22**	**0.74**	**0.82**	**0.38**	**25.5**	**7.0**
BUSI [31]	AlexNet	**0.87**	1.14	0.55	0.68	1.27	47.4	14.1
	SegNet	0.77	0.55	0.62	0.72	0.78	46.5	13.3
	U-Net	0.77	0.56	0.63	0.73	0.78	59.0	13.7
	CE-Net	0.77	0.64	0.64	0.73	0.88	43.9	12.4
	MultiResUNet	0.78	0.37	0.67	0.75	0.59	41.2	12.0
	RDAU-NET	0.80	0.42*	0.68	0.76	0.62	39.2	12.0
	SCAN	0.73	0.43	0.63	0.72	0.70	47.0	13.8
	DenseU-Net	0.74	0.43	0.64	0.72	0.69	47.4	15.5
	STAN	0.76	0.42*	0.66	0.75	0.66	46.5	12.1
	ESTAN	0.80	**0.36**	**0.70**	**0.78**	**0.56**	**34.8**	**9.9**

3.3.4 Segmentation Tumors with Different Sizes

To demonstrate the effectiveness of the proposed ESTAN model, we split the BUSIS [17][21][47][46] dataset into four tumor size groups. We chose the BUSIS dataset for the following reasons: 1) the images were collected from three hospitals using five ultrasound devices operated by different radiologists; 2) the ground truths of the BUSIS dataset have less bias because they were prepared by four experienced radiologists, where three radiologists generated tumor boundaries for each BUS image separately, and the fourth radiologist—a senior expert—judged and adjusted the majority voting results; and 3) all ten approaches achieved the best performance on the BUSIS dataset compared to BUSI and Dataset B. We chose the length of the longest axis of a tumor as a criterion for selecting tumor groups in the original BUS image. The first group contains 19 images with tumor sizes from 0 to 100 pixels, the second group has 30 images from 100 to 120 pixels, the third group consists of 81 images from 120 to 160 pixels, and the fourth group has 432 images from 160 to 533 pixels.

Table 3-4 Performance of four tumor size groups of BUSIS dataset.

Tumor size groups	(0-100)		(100-120)		(120-160)		(>160)	
Number of Images	19		30		81		432	
	JI	FP	JI	FP	JI	FP	JI	FP
AlexNet	0.57	0.97	0.63	0.64	0.68	0.44	0.76	0.27
SegNet	0.71	0.28	0.77	0.23	0.79	0.21	0.83	0.14
U-Net	0.72	0.34	0.78	0.27	0.80	0.18	0.84	0.11
CE-Net	0.62	0.63	0.80	0.19	0.80	0.16	0.84	0.09
MultiResUNet	0.71	0.34	0.80	0.16	0.82	0.17	0.86	0.09
RDAU-NET	0.72	0.26	0.82	0.14	0.81	0.17	0.85	0.09
SCAN	0.71	0.24	0.81	0.14	0.81	0.16	0.80	0.09
DenseU-Net	0.67	0.77	0.75	0.34	0.78	0.21	0.83	0.11
STAN	0.76	0.25	0.81	0.11	0.83	0.12	0.86	0.08
ESTAN	0.79	0.15	0.83	0.09	0.85	0.10	0.87	0.06

Table 3-4 lists the values of JI and FPR for the four tumor groups. AlexNet has poor performance for segmenting small tumor group with JI of 0.57 and FPR of 0.97, while FPR and JI improve

dramatically in the other three groups. The results of segmenting tumors in both mid-size groups (100-120) and (120-160) are very close to each other, e.g., CE-NET and SCAN have achieved the same JI with 0.81 and 0.80 in both groups, respectively. The results show that the tumor size between (0-100) are the most difficult cases, and all ten approaches cannot achieve as good performance as segmenting large tumors. On the other hand, for the fourth group containing the large tumor sizes (>160 pixels) all approaches achieved better results than the other tumor size groups. The proposed ESTAN achieved the highest JI and lowest FPR values on all tumor size groups.

3.4 Discussion

BUS images were obtained from different ultrasound devices with non-uniform settings, and they vary in resolution, depth, and contrast. As shown in our experimental results (Table 3-2), the performance of all approaches differs on images from different datasets. Therefore, to precisely evaluate the performance of BUS image segmentation approaches, it is recommended to involve large and diverse BUS datasets collected from different resources.

The results indicate that despite the absence of a large high-quality dataset, designing a better feature extractor is an effective approach to improving the segmentation performance of tumors of different sizes. (Tables 3-2,3-3 and 3-4).

The strengths of this study include (a) utilizing the human breast anatomical layers to design convolution kernels, (b) using two encoders to learn features and three skip connections to transfer contextual information to the decoder to locate tumors more accurately, and (c) validating the efficacy and weakness of the proposed approach using extensive experiments on three publicly available datasets. Although ESTAN achieved remarkable results for segmenting tumors of various sizes on the three datasets, it failed to detect tumors in 29 extremely challenging BUS cases, because these cases have high speckle noise, low contrast, and no clear tumor boundaries. To extract features at different scales, ESTAN uses two encoders instead of one. Despite its success, these encoders require more parameters, memory, and computational power. Therefore, optimizing ESTAN to eliminate unnecessary parameters and operations is significant, specifically for resource-constrained systems such as mobile devices.

3.5 Conclusion

To improve the segmentation of small tumors in BUS images, this paper proposed the Enhanced Small Tumor-Aware Network (ESTAN), which comprises of two encoder branches that extract and fuse image context information at different scales. The ESTAN blocks apply row-column-wise kernels to

adapt to the breast anatomy. The decoder has three skip connections from the two encoders to fuse features. The new design enhances the performance by incorporating multi-scale features and breast anatomy into the encoder layers. The proposed architecture is sensitive to small breast tumors, and segments small tumors accurately with a low FPR. In addition, the approach achieves state-of-the-art performance in segmenting tumors of different sizes. We validate the proposed approach extensively using three datasets and compare it with the other nine breast tumor segmentation approaches. The results demonstrate that ESTAN achieves the state-of-the-art performance on all datasets. In the future, we plan to test the proposed approach using large datasets and focus on developing domain-enriched deep architectures for small object detection.

Chapter 4: A Benchmark for Breast Ultrasound Image Classification

Bryar Shareef, Min Xian, Aleksandar Vakanski, Jianrui Ding, Chunping Ning, Heng-Da Cheng, A benchmark for breast ultrasound image classification, Ultrasound in Medicine &Biology, under review.

4.1 Introduction

Breast cancer has become one of the most common cancers worldwide, accounting approximately for 12% of all new cancer cases [53]. In the U.S., it is estimated that breast cancer affected 30% of all new female cancer cases in 2021 [53]. Early detection of breast cancer can significantly reduce mortality and expand treatment options. Among the different imaging modalities, mammography and ultrasound are the two most popular imaging tools for detecting breast abnormality. However, mammography is less commonly implemented in most low- and middle-income countries, because of the high costs of the required infrastructure [54].

Furthermore, mammography produces high false-positive rates in women with dense breasts, which leads to anxiety and additional examination steps, such as biopsy [55]. Rebolj et al. [56] reported that ultrasound detected approximately 40% more cancer cases than mammography in women with dense breasts. According to [5-9], women with dense breasts have a four to six times greater risk of breast cancer than those with fatty breast tissue. Asian women of age < 45 have 1.2 more dense breasts than white women of that age, and the ratio increases to 1.6 for age 65 and older. In contrast, black women have 1.7 more dense breasts than white women for age 65 and younger, while black, Hispanic, and white women have a similar breast density for ages>65.

BUS image processing is challenging due to the presence of speckle noise, low contrast, weak boundary, and artifacts [21]. Therefore, analyzing ultrasound images requires extensive experience and training. To alleviate this challenge, computer-aided diagnosis (CAD) systems have been developed to assist radiologists with breast tumor diagnosis. The idea of CAD was first introduced in the 1960s [62]. These systems can reduce operator dependency and identify breast tumors/cancers more accurately [47]. CADs can be broadly classified into conventional and deep learning-based systems [63]. The conventional BUS CAD systems typically comprise four modules: image preprocessing, tumor segmentation, feature extraction and selection, and tumor classification [47] (see Figure 1-1(a)). In deep learning-based CAD systems, the modules of preprocessing [11,12] and segmentation [16,17] become optional (see Figure 1-1(b)).

Table 4-1 Deep learning approaches for BUS image classification.

References	Approaches	Year	Dataset/Availability	Performance	Pretrained dataset
Huynh, et al. [7]	Feature extractor (AlexNet) + SVM	2016	1,125 cases/private	AUC: 88%	ImageNet
Shia, et al. [71]	Fine-tuned (ResNet101) + SVM	2021	2,099/private	Sen: 94%, Spec: 93%, AUC: 94%	ImageNet
Liang, et al. [67]	Feature extractor (Mask-R-CNN)	2019	150 cases/private 163 cases/public	Acc : 80%, TPR: 63%, TNR: 87%	Coco datasets
Liao et al. [77]	Fine-tuned (VGG19, ResNet50, DenseNet121, Inceptions V3) + Elastography images + B-mode images	2020	256/private	AUC:98%, Acc:93%, Sen:91%, Spec: 95%, F_1: 93%	ImageNet
Fei et al. [78]	Designed DL network (SVM + Elastography) + Transfer learning	2020	265/private	Acc:87%, Sen: 86%, Spec: 87%, Youden index (YI): 73%	--
Yap et al. [116]	Fine-tuned (FCN-AlexNet)	2018	306/private 163/public	Sen (Benign:83%, Malignant: 57%)	ImageNet
Zhang, et al. [72]	Fine-tuned (VGG16, ResNet50, InceptionV3, VGG19)	2020	6,007/private	Sen: 85%, AUC: 91%, PPV:64%, Acc:83%, NPV: 93.7%, Spec: 81.5%	ImageNet
Hijab et al. [69]	Fine-tuned (VGG16) + ROIs	2019	1,300/private	Acc: 97%, AUC: 98%	ImageNet
Cao et al. [70]	Fine-tuned (4 ROIs on five networks)	2019	1,041/private	APR: 97%, ARR:67%, F_1:79%, Acc: 87.5%	ImageNet
Xie et al. [73]	Network design (Dual-sampling (2 Encoders) network)	2020	1,272/private 163/public	Acc: 92%, Sen: 95%, Spec: 89%, PPV: 88%, NPV: 95%, AUC: 94%	ImageNet
Xing et al. [75]	Prior knowledge (BI-RADS + CNN)	2020	Training: 9,373/private Tested: 810/public	AUC: 91%, Acc: 87%, Sen: 82%, Spec: 89%, Precision: 80%	ImageNet
Zhuang et al. [76]	Prior knowledge (hand crafted features +SVM +DL)	2021	1,682/public	Acc: 93%, Precision: 91%, Sen: 95%, F_1: 93%, Spec: 91%	ImageNet
Han et al. [74]	Adopting modified network (GoogleNet) +ROIs	2017	7,408/private	AUC: 96%, Acc: 91%, Sen:84%, Spec: 96%	ImageNet
Al-Dhabyani et al. [64]	Data augmentation (GAN to produce data)	2019	780/public	Acc: 99%	ImageNet
Tanaka et al. [117]	Ensemble Learning (VGG19 +ResNet152)	2019	1,543/private	Acc: 86%, Precision: 85%, Sen: 89%, F_1 : 87%, Spec: 83%, AUC:94%	--
Byra et al. [65]	Preprocessing (Input Channel)	2019	Training:882/ private Tested: 163/public	AUC: 94%, Acc: 89%, Sen: 85%, Spec: 90%	ImageNet
Zhuang et al. [79]	Preprocessing (Decomposition of BUS images)	2020	2,280/public	AUC: 98%, Acc: 92%, Sen: 98%, Spec: 86%, F_1: 93%	ImageNet
Zhang et al. [81]	Multitask learning + attention mechanism	2021	647/public	Acc:94%, Sen: 89%, Spec: 96%, F_1:93%	--
Moon et al. [118]	Ensemble learning (BUS + tumor masks + segmented tumor + fused images)	2020	647/public 1,687/private	AUC: 95%, Acc:91%, Sen: 97%, Spec: 95%, F_1: 83%, Precision: 73%	--

Acc: Accuracy, AUC: area under curve, Sen: sensitivity, Spec: Specificity, TNR: true negative rate, TPR: true positive rate, PPV: positive predictive value, NPV: negative predictive value, APR: average precision rate, ARR: average recall rate,

Automatic feature learning without human intervention is a substantial advantage of deep learning-based approaches over conventional approaches [63].

On the other hand, conventional approaches rely on radiologists' knowledge to extract and select meaningful features [46].

Given recent advancements in deep learning approaches for medical image applications, prior work demonstrated the effectiveness of deep learning to classify breast tumors in ultrasound images (see Table 4-1). However, due to the lack of large, publicly available, high-quality BUS datasets, and unified quantitative metrics, a fair evaluation of the current approaches and strategies is impossible. Furthermore, most existing deep learning architectures for BUS image classification are simply adopted from general-purpose image classification tasks, and there is limited research on identifying the best architectures and strategies of deep learning for BUS image classification. In this paper, the focus is on benchmarking deep learning-based CAD systems for BUS image classification. Refer to [21] for more details on a BUS benchmark for breast tumor segmentation.

The paper is organized as follows. Section 2 discusses the fundamentals of BUS image classification using deep learning; Section 3 describes the benchmark setup. Section 4 illustrates the proposed approach; Section 5 presents comprehensive experimental results. Finally, Sections 6 and 7 provide a discussion and conclusion, respectively.

4.2 Fundamentals of BUS image classification using deep learning

4.2.1 Transfer Learning

Deep learning typically requires large and high-quality labeled data. However, many medical applications have scarce data due to expensive data collection, high labeling costs, and privacy issues. To address these issues, many approaches have adopted the transfer learning strategy. In transfer learning approaches, a deep learning network, which is previously pretrained for another task on a large-scale dataset is employed for BUS classification. For example, the ImageNet [66] dataset is widely used by deep learning approaches for learning feature representations. The pretrained model can be used as 1) a fixed feature extractor or 2) an initial model for fine-tuning.

For the fixed feature extractor, the pretrained layers are kept unchanged, and the prediction layers are trained based on the target task. Huynh et al. [7] employed a pretrained model (AlexNet) as a feature extractor and combined it with a support vector machine (SVM) algorithm to classify BUS images by using 1,125 whole images and 2,393 regions of interest (ROIs). Liang et al.

[67] proposed using Mask R-CNN to segment and classify breast tumors simultaneously, where a ResNet50 pretrained on the COCO [68] dataset was used as a backbone to extract features.

In models for fine-tuning, the whole network including the pretrained layers and the prediction layers is retrained using new data. The fine-tuning approach uses the pretrained weights to initialize the network and tune it to a target task. Hijab et al. [69] adopted transfer learning to train VGG16 for classifying BUS images. The authors studied three different training techniques, and the results demonstrated that the fine-tuned network outperformed both training from scratch and transfer learning without fine-tuning. Cao et al. [70] studied breast tumor detection and classification using five models with and without transfer learning. Moreover, [71] used a pretrained deep residual network as a feature extractor and a support vector machine (SVM) algorithm to classify BUS images, and their classification performance on 2,099 BUS images outperformed physicians. Zhang et al. [72] used a balanced training set and compared four pretrained classifiers (InceptionV3, VGG16, ResNet50, and VGG19), and pretrained InceptionV3 with fine-tuning outperformed all other three models.

4.2.2 Network Architectures

Developing network architectures based on domain knowledge can enhance the generalizability of deep learning-based approaches. Xie et al. [73] proposed the DSCNN to combine convolutional and residual layers for BUS image classification. DSCNN outperformed pretrained and fine-tuned AlexNet, ResNet18, VGG16, GoogleNet, and EfficientNet, and the three experienced radiologists. Han et al. [74] modified GoogleNet with different regions of interest (ROIs) which accepted single-channel images and removed two auxiliary classification branches. The proposed approach achieved a sensitivity of 86% and an AUC of 90% on a private dataset.

4.2.3 Incorporating Prior Knowledge

Xing et al. [75] integrated BI-RADS information into a three-layer residual network. The proposed approach showed promising results and outperformed all other transfer learning and non-transfer learning approaches on two public datasets and one private dataset. Zhuang et al. [76] extracted four characteristic semantic features (i.e., orientation, characteristics of posterior shadowing region, shape complexity, and edge indistinctness) and combined them with computational features learned from VGG16. The proposed approach outperformed the general-purpose-designed deep learning approaches. Liao et al. [77] extracted computational features using two VGG19 models from B-mode BUS images and strain elastography images, respectively; and all features are concatenated and input into a 3-layer network to conduct classification. The results showed that the proposed approach can achieve better

sensitivity and specificity compared to deep learning approaches trained solely on B-mode images. Similarly, [78] transferred knowledge from elastography ultrasound through transfer learning to improve the diagnostic accuracy of breast cancer.

4.2.4 Preprocessing

The image quality and size of a BUS dataset have a significant impact on deep learning models. Researchers have employed a variety of preprocessing techniques to enlarge, standardize, and enhance datasets. Al-Dhabyani et al. [64] implemented a new augmentation approach by combing generative adversarial networks (GANs) with traditional augmentation methods; and the classification accuracy of VGG16, Inception, ResNet, and NasNet was improved by 16%, 17%, 16%, and 15%, respectively. Byra et al. [65] introduced a matching layer to rescale grayscale BUS images to RGB images. The results showed that this technique improved the performance of a pretrained VGG19 network. Zhuang et al. [79] used fuzzy enhancement, bilateral filtering, and image morphology operation to produce a set of decomposed images which were combined to feature maps using three deep learning models. The approach showed promising results, with the specificity and sensitivity reaching 98% and 94%, respectively.

4.2.5 Multitask Learning

Multitask learning has been proved to be an effective approach to improve the generalizability of deep learning approaches by learning shared representations from multiple tasks. Vakanski et al. [80] implemented a deep multitask network that comprised both tumor segmentation and classification subnetworks, and the performance of tumor classification was significantly improved by learning representations focused on tumor regions. Zhang et al. [81] employed soft and hard attention mechanisms to perform tumor classification and segmentation simultaneously; and the classification accuracy increased by 2.45% compared with the single task model. Shi et al. [82] proposed the EMT-NET, a light-weighted multitask learning approach for both breast tumor classification and segmentation to replace the single task MobileNet; and its sensitivity increased by 18.81%.

4.2.6 Challenges

Conclusively, despite the potential of deep learning approaches for accurately classifying BUS images, considerable challenges still need to be addressed: 1) most deep learning approaches require large and high-quality labeled datasets, but most publicly available BUS datasets are small. It is time-consuming and expensive to collect a large BUS dataset. 2) The end-to-end learning scheme of deep learning approaches makes BUS image classification a black box, which leads to poor explainability. 3) Existing

deep learning approaches have poor robustness and are vulnerable to adversarial attacks. 4) Most deep learning approaches are computationally intensive, which makes it impossible to deploy them to devices with limited resources. To the best of our knowledge, there is an absence of benchmarking studies focusing on deep learning approaches in classifying breast ultrasound images. Therefore, we are introducing a BUS benchmark to identify the most useful strategies for classifying breast tumors using a combined dataset of 3,641 BUS images.

4.3 Benchmark Setup

This section provides a detailed description of the BUS image datasets, deep learning approaches, experimental setup, and evaluation metrics.

4.3.1 BUS Image Dataset

Existing public BUS datasets are small. We prepared a large and diverse BUS dataset from five sources, HMSS [83], BUSI [84], BUSIS [21], Thammasat [85], and Dataset B [5]. It contains a total of 3,641 B-model BUS images, of which 1,854 contain benign tumors and 1,763 have malignant tumors. Detailed information on the five datasets is shown in Table 4-2. We develop a set of scripts to prepare the images which are publicly available at http://busbench.midalab.net. Note that we do not own the images, and researchers need to obtain permissions to use the datasets from the original authors.

Table 4-2 Five public BUS datasets.

BUS dataset	BUS images	Class distribution	Ground truth availability	Country
HMSS [83]	2,006	B: 846, M: 1,160	Classification: Yes Segmentation: No	Netherlands
BUSI [84]	647	B: 437, M: 210	Classification: Yes Segmentation: Yes	Egypt
BUSIS [21]	562	B: 306, M: 256	Classification: Yes Segmentation: Yes	China
Thammasat [85]	263	B:120, M: 143	Classification: Yes Segmentation: No	Thailand
Dataset B [5]	163	B: 109, M: 54	Classification: Yes Segmentation: Yes	Spain
Total # of images	3,641	Total # of Benign (B): 1,823 (50.06%) Total # of Malignant (M): 1,818 (49.94%)		

A total of 2,006 BUS images are from the HMSS [83] dataset, of which 882 images have benign tumors and 1,100 have malignant tumors. HMSS was collected by Dr. Geertsma, an experienced radiologist at Gelederse Vallei hospital in Netherland, in a collaboration with Hitachi Medical Systems Europe. BUSI [84] dataset was collected from Baheya Hospital for Early Detection & Treatment of Women's Cancer (Cairo, Egypt) using LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system with the ML6-15-D Matrix linear probe transducers. The dataset has a total of 780 images, of

which 133 are normal, 437 are benign, and 210 are malignant. It was collected from 600 women patients aged between 25 and 75 years old. We excluded the normal cases, resulting in a total of 647 BUS images. BUSIS [21] dataset was collected from the Second Affiliated Hospital of Harbin Medical University, the Affiliated Hospital of Qingdao University, and the Second Hospital of Hebei Medical University using the GE VIVID 7, LOGIQ E9, Hitachi EUB-6500, Philips iU22, and Siemens ACUSON S2000 systems. It contains 562 images, of which there are 306 benign and 256 malignant images. Thammasat dataset [85] was collected by the Biomedical Engineering Unit at the Thammasat University Hospital, and Philips iU22 ultrasound workstation was used. We get a total number of 263 (120 benign and 143 malignant) BUS images from the Thammasat dataset. Dataset B [5] consists of 163 breast ultrasound images (53 malignant and 110 benign), provided by the UDIAT Diagnostic Centre of the Parc Taul´ı Corporation, Sabadell (Spain). The images were collected using the Siemens ACUSON Sequoia C512 system with a 17L5 linear array transducer (8.5 MHz). Refer to the original publications of the datasets for more details.

Table 4-3 The sizes of the selected classifiers.

List of generic deep learning classifiers			
	Classifiers	Number of parameters (million)	Size of trained models (megabytes)
1	MobileNet	4.2	29 MB
2	EfficientNetB0	5.3	37 MB
3	DenseNet121	8	59 MB
4	Xception	22.9	168 MB
5	InceptionV3	23	176 MB
6	ResNet50	25	189 MB
7	VGG16	138.3	172 MB
List of BUS-specific deep learning classifiers			
1	Shi, et al. [82]	5.1	60 MB
2	Zhang, et al. [81]	8.2	130 MB
3	Vakanski, et al. [80]	27.3	312.6 MB

Because most deep learning approaches require square images as input, all BUS images in the benchmark dataset are zero-padded and reshaped to form square images without distortions. Note that directly reshaping an original BUS image to a square shape will result in morphologic changes in breast tumors and their surrounding tissues. Refer to our scripts for preparing the benchmark dataset.

4.3.2 Deep Learning Approaches and Setup

In this study, we evaluate seven generic widely used deep learning-based classifiers [41-47] and three recently published state-of-the-art approaches [34-36] for BUS image classification (see Table 4-3). The generic approaches include MobileNet V1 [86], EfficientNet [87], DenseNet121 [88], ResNet50

[89], VGG16 [90], Xception [91], and InceptionV3 [92]. These classifiers are among the most commonly used architectures in medical image applications, thus, providing new insights into their performance will benefit the development of CAD systems and the research community. In addition, the approaches range from lightweight to heavyweight models, and evaluating them could help build applications with hardware limitations. The 5-fold cross-validation is utilized to assess the performance of all approaches. The maximum number of training epochs is set to 50, and the batch size is 32. In addition, a validation set that comprises 20% of the training set is used, and all BUS images of the benchmark dataset are resized to the original classifier's input size. In the benchmark dataset, multiple images may come from one patient/case. To prevent data leakage and bias, we split the train and test set based on the cases, i.e., all images from one case are assigned to only one of the training, validation, and test sets. The approaches are implemented in Keras and TensorFlow using Python (version 3.7) programming language. All experiments were performed on a GPU server with seven NVIDIA Quadro RTX 8000 GPUs, two Intel Xeon Silver 4210R CPUs (2.40GHz), and 512 GB of RAM.

4.3.3 Evaluation Metrics

To evaluate the performance of the classifiers, we use the following quantitative metrics: accuracy (Acc), sensitivity (Sens), specificity (Spec), F_1 score, false positive rate (FPR), false negative rate (FNR), and Area Under the Receiver Operating Characteristic Curve (AUC).

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{4.1}$$

$$\mathrm{Sens} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4.2}$$

$$\mathrm{Spec} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{4.3}$$

$$\mathrm{F_1} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + (\mathrm{FP} + \mathrm{FN})} \tag{4.4}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4.5}$$

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}} \tag{4.6}$$

In Eqs. (4.1-4.6), TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

4.3.4 Loss Functions

We explore three different loss functions to improve the overall performance and identify the best strategy that can better balance the sensitivity and specificity for breast cancer detection. The adopted loss functions include binary cross-entropy loss, focal loss [93], and weighted cross-entropy loss. The binary cross-entropy is widely employed in binary classification, and it is defined by

$$L_{BCE} = -\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}[(t_i \cdot log(p_i) + (1 - t_i) \cdot log(1 - p_i)] \tag{4.7}$$

where N denotes the number of image samples; t_i is the target label of the *i*th training sample; p_i denotes the prediction. Cross-entropy loss calculates the difference between two probability distributions and all classes are treated equally. To reduce the risk of false negatives, we employed the weighted cross-entropy function. The normal weighted cross-entropy is given by

$$L_{WBCE} = -\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}[(w_z \cdot t_i \cdot log(p_i) + (1 - t_i) \cdot log(1 - p_i)] \tag{4.8}$$

where w_z is the weight parameter that penalizes the false-negative predictions and could also mitigate the issue of imbalanced classes. To avoid overflow issues and produce stable results, we utilized a numerically stable weighted cross-entropy which was implemented in [36] and is defined by

$$L_{NS-WBCE} = -\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left((1 - t_i) \cdot l_i + s_i \cdot \log\left(1 + e^{-l_i}\right)\right) \tag{4.9}$$

where l_i is the logits of the predicted probability p_i, and s_i is from the positive weight coefficient. They defined as $l_i = \log\left(\frac{p_i}{1-p_i}\right)$ and $s_i = 1 + t_i \cdot (w_z - 1)$.

Furthermore, to focus more on difficult predictions, we utilized the focal loss function [50]. In the focal loss, a factor $(1 - p_i)^{\gamma}$ is added to the cross-entropy loss, where γ is a focusing parameter that makes the model focus on hard samples. The focal loss is defined by

$$L_{Focal} = -\frac{1}{\mathrm{N}}\sum_{i=1}^{N}[(\alpha \cdot t_i \cdot (1 - p_i)^{\gamma} \cdot \log(p_i) + (1 - t_i) \cdot (1 - \alpha) \cdot p_i \cdot \log(1 - p_i)] \tag{10}$$

where α is a weighting factor, and takes values from [0, 1]. We use nine combination of focal loss weights (γ = {2, 3, 4}, and α = {0.25, 0.50, 0.8}) and five weights for L_{WBC} (1, 2, 3, 4, and 5).

4.3.5 The Proposed Method

Multitask learning (joint BUS segmentation and classification) can significantly improve the generalization ability of deep learning approaches trained using datasets with limited sizes. The

performance of the primary task could be improved using better representations regularized by a secondary task. In BUS images, tumor categories are determined by features inside or around a tumor; if we could regularize a deep neural network to learn representations of tumor regions, a more accurate and robust model could be trained. Inspired by this, we propose a new deep multitask network, namely MT-ESTAN, which consists of both tumor segmentation and classification tasks.

The network architecture is shown in Figure 4-2.

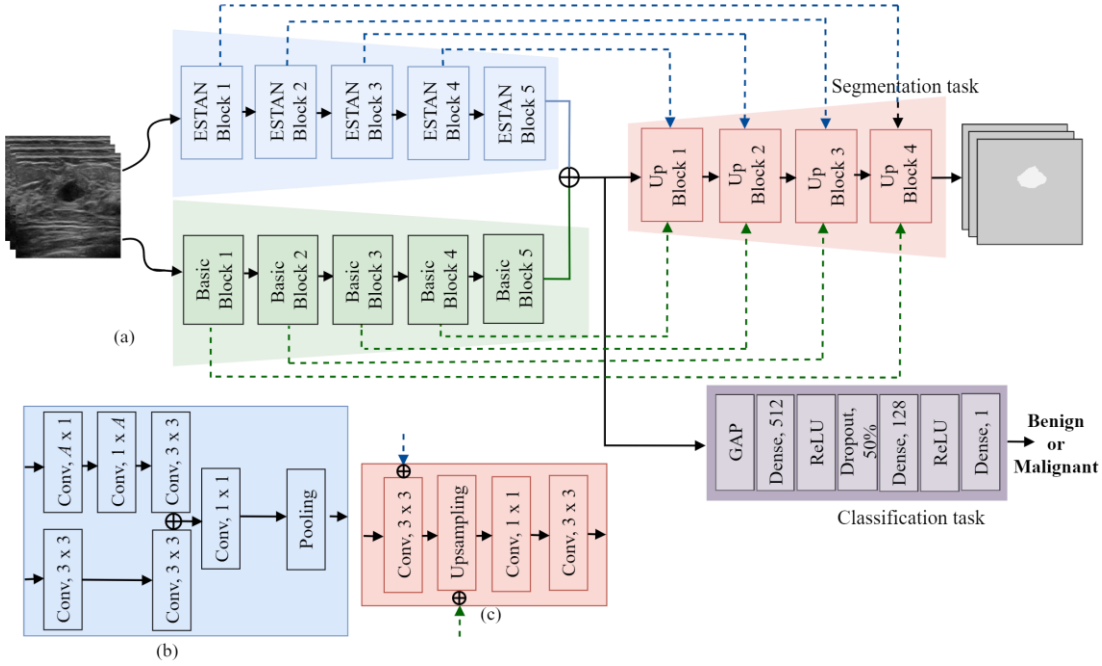


Figure 4-1 MT-ESTAN architecture. (a) Overall architecture; (b) the ESTAN block; and (c) the upsampling (Up) block. ⊕ denotes the concatenation operator, and A denotes kernel size.

In our previous work [16, 17], small-tumor aware networks were proposed to accurately segment tumors with different sizes. [16] used row-column-wise kernels to extract and fuse BUS context information at different scales. It consists of two parallel encoder branches: the enhanced small-tumor aware network (ESTAN) and basic encoders. In this work, we use the network in [16] as the backbone of MT-ESTAN to ensure sensitivity to tumors with different sizes; and ResNet50 is used as the building blocks of the basic encoder. Refer to [12] for the implementation details of ESTAN. There are several major differences between the proposed MT-ESTAN and our ESTAN in [16]: 1) MT-ESTAN performs tumor classification and segmentation simultaneously, and tumor classification is the primary task. ESTAN [16] only has a tumor segmentation task; 2) the loss function of MT-ESTAN is a balanced combination between $L_{NS-WBCE}$ and Dice loss, while ESTAN only has the Dice loss; and 3) the basic encoder was pretrained on ImageNet in MT-ESTAN, but trained from scratch in ESTAN.

Segmentation Task. The segmentation task is supplementary to the classification task. The

segmentation branch comprises four blocks, and each has an upsampling layer and three consecutive convolution kernels (see Figures. 4-2(a) and (c)). Each block receives two skip connections from blocks in the two encoders, i.e. a skip connection from the basic encoder and another from the ESTAN encoder.

Classification Task. The primary task of the proposed MT-ESTAN is to classify BUS tumors into benign and malignant. The classification branch receives input from the combined basic and ESTAN encoders. It consists of a Global Average Pooling (GAP) layer followed by two dense layers using ReLU activation with 512, and 128 nodes, respectively. A dropout layer with a rate of (50%) is added after the first dense layer. The final prediction consists of a single node employing a sigmoid activation function.

Loss function. In disease diagnosis, the models that produce higher sensitivity are more vital than that vice versa. We utilize the weighted cross-entropy loss function for the classification task to perform a trade-off between sensitivity and specificity with minimum sacrifice of overall accuracy. A numerically stable weighted cross-entropy from [82] is adopted and is defined in Eq. (4.9). The final multitask loss (L_{mtl}) function is defined by

$$L_{mtl} = \mathrm{w} \cdot L_{NS-WBCE} + L_{Dice} \tag{4.11}$$

where the weight (w) of the classification task is set to 3, and the positive weight of $L_{NS-WBCE}$ is set to 3. In addition, the best model with the minimum

where the weight (w) of the classification task is set to 3, and the positive weight of $L_{NS-WBCE}$ is set to 3. In addition, the best model with the minimum validation loss will be saved during training.

The proposed approach and [34-36] share the same two tasks. However, two major differences exist. 1) [80], [81], and [82] used U-Net, DenseNet, and MobileNet, respectively, as the backbone network. The proposed multitask network applies the ESTAN as the backbone, and is more robust to tumors of different sizes. 2) [80] and [81] used the cross-entropy function as the loss of the classification loss, and have no control on the balance of sensitivity and specificity. For example, [81] obtained high specificity but relatively low sensitivity. The proposed network utilizes the numerically-stable weighted cross-entropy loss that enables the flexibility to balance sensitivity and specificity.

4.4 Experimental Results

In this section, we evaluate the proposed approach and 10 deep learning-based approaches for BUS image classification using the proposed benchmark dataset. The five most useful strategies in deep learning are validated by experiments in Sections 5.1 and 5.2.1; and the effectiveness of the proposed approach is validated and discussed in Section 5.2.2.

4.4.1 Evaluate Useful Strategies in Deep Neural Networks for BUS Image Classification

Table 4-4 Training from Scratch (S) vs. Transfer learning (TL).

Classifiers	**Accuracy (%) ↑**		**Sensitivity (%) ↑**		**Specificity (%) ↑**		**F_1 ↑**		**AUC (%) ↑**		**FPR (%) ↓**		**FNR (%) ↓**	
	S	TL	S	TL	S	TL	S	TL	S	TL	S	TL	S	TL
DenseNet121	64.8	73.3	69.3	70.9	59.8	75.9	0.66	0.72	64.5	73.4	40.2	24.1	30.7	29.1
InceptionV3	64.5	71.6	69.0	62.8	59.6	**80.5**	0.66	0.69	64.3	71.7	40.4	**19.5**	31.0	37.2
MobileNet	61.7	75.3	74.5	76.9	49.1	74.2	0.66	**0.76**	61.8	75.5	50.9	25.8	25.5	23.1
ResNet50	62.0	70.3	74.2	**79.1**	50.6	61.8	0.66	0.73	62.4	70.4	49.4	38.2	25.8	**20.9**
VGG16	68.9	**76.7**	75.7	75.8	62.2	77.8	0.70	**0.76**	68.9	**76.8**	37.8	22.2	24.3	24.2
Xception	63.1	72.7	75.4	73.0	52.1	72.6	0.67	0.73	63.8	72.8	47.9	27.4	24.6	27.0
EfficientNetB0	59.7	74.0	75.6	73.0	43.8	75.4	0.65	0.74	59.7	74.2	56.2	24.6	24.4	27.0

Training from scratch versus transfer learning. In the transfer learning setup, all classifiers are pretrained on ImageNet, and the last prediction layer is replaced with two dense layers with 512 and 64 units, respectively. ReLU is used as the activation. All model parameters are trainable in the fine-tuning stage. For training from scratch, all seven models are trained from scratch using BUS images. Additionally, all experiments were conducted without using regularization, augmentation, and postprocessing techniques.

The results presented in Table 4-4 show that all seven models with transfer learning outperform those with training from scratch. It is worth noting that transfer learning significantly enhances the performance of the less complex classifiers with small model sizes. The reason could be that small models are prone to underfit when trained from scratch on a limited number of images. For example, the EfficientNetB0 model is a lightweight classifier with only 5.3 million parameters, and its accuracy, F_1 score, and AUC improved by 19.3%, 12.1%, and 19.5%, respectively. On the other hand, VGG16 is

Table 4-5 Augmentation (Aug.) vs. no augmentation (No Aug.).

Classifiers	**Accuracy (%) ↑**		**Sensitivity (%) ↑**		**Specificity (%) ↑**		**F_1 ↑**		**AUC (%) ↑**		**FPR (%) ↓**		**FNR (%) ↓**	
	No Aug.	Aug.	No Aug.	Aug.	No Aug.	Aug.	No Aug.	Aug.	No Aug.	Aug.	No Aug.	Aug.	No Aug.	Aug.
DenseNet121	73.3	76.9	70.9	72.2	75.9	**81.9**	0.72	0.76	73.4	77.0	24.1	**18.1**	29.1	27.8
InceptionV3	71.6	75.7	62.8	73.4	80.5	78.4	0.69	0.75	71.7	75.9	19.5	21.6	37.2	26.6
MobileNet	75.3	**77.2**	76.9	75.1	74.2	79.6	0.76	**0.77**	75.5	**77.4**	25.8	20.4	23.1	24.9
ResNet50	70.3	76.2	79.1	74.0	61.8	78.5	0.73	0.75	70.4	76.3	38.2	21.5	20.9	26.0
VGG16	76.7	76.6	75.8	**77.6**	77.8	75.8	0.76	**0.77**	76.8	76.7	22.2	24.2	24.2	**22.4**
Xception	72.7	76.0	73.0	72.1	72.6	79.9	0.73	0.75	72.8	76.0	27.4	20.1	27.0	27.9
EfficientNetB0	74.0	76.7	73.0	74.0	75.4	79.6	0.74	0.76	74.2	76.8	24.6	20.4	27.0	26.0

a heavyweight classifier with 138 million parameters, and its accuracy, F_1 score, and AUC improved by 10.1%, 7.8%, and 10.2%, respectively. The pretrained VGG16 classifier outperformed all other classifiers by achieving the best F_1 score and AUC. Because transfer learning improves the overall classification performance, it is used in the remaining sections.

Image augmentation. Several augmentation techniques are explored to improve models' generalizability. An optimal augmentation technique should not distort the BUS images, because tumor shapes, boundaries, echo patterns, and margins in breast cancer classification are essential in etermining the tumor type. The classifiers are trained on six different augmentation techniques individually: horizontal flip, height shift, width shift, zoom, shear, and rotation. A combination of the four best-performed techniques including the horizontal flip, height shift (0.2), width shift (0.2), and rotation (20%), is chosen to augment the training set. The results in Table 4-5 demonstrate that the augmentation combination improves the overall performance of DenseNet121, InceptionV3, MobileNet, ResNet50, Xception, and EffienetNetB0 classifiers except for VGG16.

Table 4-4 Results of different loss functions.

Classifiers	**Loss**	**Accuracy (%) ↑**	**Sensitivity (%) ↑**	**Specificity (%) ↑**	**F_1 ↑**	**AUC (%) ↑**	**FPR (%) ↓**	**FNR (%) ↓**
DenseNet121	L_{BCE}	76.9	72.2	**81.9**	0.76	77.0	**18.1**	27.8
	L_{WBCE} (w_z=4)	72.7	90.1	55.7	**0.77**	72.9	44.3	9.90
	$L_{Focal}(\gamma = 3, \alpha = 0.8)$	70.3	88.6	52.6	0.75	70.6	47.4	11.4
InceptionV3	L_{BCE}	75.7	73.4	78.4	0.75	75.9	21.6	26.6
	$L_{WBCE}(w_z = 3)$	71.6	86.8	57.1	0.75	71.9	42.9	13.2
	$L_{Focal}(\gamma = 2, \alpha = 0.8)$	68.3	90.3	47.4	0.74	68.8	52.6	9.70
MobileNet	L_{BCE}	**77.2**	75.1	79.6	**0.77**	**77.4**	20.4	24.9
	$L_{WBCE}(w_z = 3)$	74.0	87.6	60.8	**0.77**	74.2	39.2	12.4
	$L_{Focal}(\gamma = 3, \alpha = 0.8)$	72.3	87.2	57.6	0.76	72.4	42.4	12.8
ResNet50	L_{BCE}	76.2	74.0	78.5	0.75	76.3	21.5	26.0
	$L_{WBCE}(w_z = 3)$	72.6	86.2	59.4	0.76	72.8	40.6	13.8
	$L_{Focal}(\gamma = 3, \alpha = 0.8)$	71.3	88.3	54.4	0.75	71.4	45.6	11.70
VGG16	L_{BCE}	76.6	77.6	75.8	**0.77**	76.7	24.2	22.4
	$L_{WBCE}(w_z = 3)$	74.5	86.7	62.6	**0.77**	74.7	37.4	13.3
	$L_{Focal}(\gamma = 2, \alpha = 0.8)$	70.3	90.2	50.9	0.75	70.5	49.1	9.80
Xception	L_{BCE}	76.0	72.1	79.9	0.75	76.0	20.1	27.9
	$L_{WBCE}(w_z$=3)	72.9	88.7	57.7	**0.77**	73.2	42.3	11.30
	$L_{Focal}(\gamma = 2, \alpha = 0.8)$	68.2	**91.9**	45.1	0.74	68.5	54.9	**8.10**
EfficientNetB0	L_{BCE}	76.7	74.0	79.6	0.76	76.8	20.4	26.0
	$L_{WBCE}(w_z$=3)	73.8	86.8	61.2	**0.77**	74.0	38.8	13.2
	$L_{Focal}(\gamma = 2, \alpha = 0.8)$	69.6	91.3	48.5	0.75	69.9	51.5	8.70

This is because the VGG16 without augmentation has less overfitting than other approaches, and extra augmented images do not improve its performance significantly. The proposed combination of augmentation techniques is utilized for all classifiers to expand the dataset size in the remaining experiments.

Loss functions. As described in section 3.4, the binary cross-entropy loss(LBCE), focal loss [93] (LFocal), and weighted cross-entropy loss (LWBCE) are evaluated. Table 4-6 shows the performance of different models with the loss parameter(s) that leads to the best overall and sensitivity values. By utilizing the L_{WBCE}, the sensitivity of DenseNet121, InceptionV3, MobileNet, ResNet50,

VGG16, Xception, and EfficientNet improved by 19.8%, 15.4%, 14.2%, 14.1%, 10.4%, 18.7%, and 14.7%, respectively. Additionally, with the Focal loss, the sensitivity has further improved, but the overall performance degrades considerably. For example, the sensitivity of InceptionV3 and Xception has increased by 18.7%, and 21.5%, respectively; however, the AUC is reduced by 9.3%, and 9.8%, respectively. The best trade-off between sensitivity and specificity is achieved by MobileNet and VGG16 when L_{WBCE} is used.

Optimizers. We compare three popular optimizers: Adaptive Moment Estimation (ADAM) [94], Stochastic Gradient Descent (SGD) with momentum, and Nesterov-accelerated Adaptive Moment Estimation (NADAM) [95]. In the experiments, ADAM is applied with a learning rate of 0.00001, SGD with a learning rate of 0.002 and momentum of 0.9, and NADAM with a learning rate of 0.00001, beta_1 of 0.9, beta_2 of 0.999, and epsilon of 1e-08. All other parameters take default values in Keras.

As shown in Table 4-7, DenseNet121, ResNet50, VGG16, and EfficientNet classifiers achieved better F_1 scores and AUC values using the ADAM optimizer. On the other hand, InceptionV3,

Table 4-5 Results of different optimizers.

Classifier	Optimizer	Accuracy (%) ↑	Sensitivity (%) ↑	Specificity (%) ↑	F_1 ↑	AUC (%) ↑	FPR (%) ↓	FNR (%) ↓
DenseNet121	**ADAM**	72.7	**90.1**	55.7	**0.77**	72.9	44.3	**9.9**
	SGD	71.6	89.0	54.8	0.76	71.9	45.2	11.0
	NADAM	71.1	87.7	55.0	0.75	71.3	45.0	12.3
InceptionV3	ADAM	71.6	86.8	57.1	0.75	71.9	42.9	13.2
	SGD	73.0	88.4	57.6	**0.77**	73.0	42.4	11.6
	NADAM	70.5	86.5	54.6	0.74	70.6	45.4	13.5
MobileNet	ADAM	74.0	87.6	60.8	**0.77**	74.2	39.2	12.4
	SGD	74.0	87.4	61.3	**0.77**	74.4	38.7	12.6
	NADAM	72.4	83.6	61.5	0.75	72.5	38.5	16.4
ResNet50	**ADAM**	72.6	86.2	59.4	0.76	72.8	40.6	13.8
	SGD	70.8	87.6	54.6	0.75	71.1	45.4	12.4
	NADAM	70.8	85.2	56.7	0.74	70.9	43.3	14.8
VGG16	**ADAM**	**74.5**	86.7	**62.6**	**0.77**	**74.7**	**37.4**	13.3
	SGD	70.2	89.7	51.1	0.75	70.4	48.9	10.3
	NADAM	71.5	86.3	57.0	0.75	71.7	43.0	13.7
Xception	ADAM	72.9	88.7	57.7	**0.77**	73.2	42.3	11.3
	SGD	73.7	88.5	59.6	**0.77**	74.0	40.4	11.5
	NADAM	69.1	87.6	50.0	0.74	68.8	50.0	12.4
EfficientNetB0	**ADAM**	73.8	86.8	61.2	**0.77**	74.0	38.8	13.2
	SGD	73.8	86.2	61.7	**0.77**	73.9	38.3	13.8
	NADAM	72.4	85.1	59.7	0.75	72.4	40.3	14.9

MobileNet, and Xception achieved better results using the SGD optimizers. It is worth mentioning that the optimizers have the slightest impact on the generalization performance among all the strategies that we tested. DenseNet121 achieved the best sensitivity with 90.1% by using ADAM optimizers, and EfficientNetB0 attained the lowest sensitivity with 85.1% by using the NADAM optimizer. In addition, the VGG16 using Adam and MobileNet using SGD achieved the best AUC by 74.7% and 74.4%, respectively.

4.5 Multitask Learning

The multitask learning approaches need ground truth labels for both tumor class and tumor boundaries, and a combined dataset (BUSI and BUSIS) with a total of 1,209 BUS images is used. BUSI and BUSIS are chosen because they have accurate annotations for both tumor boundaries and classes. The 5-fold cross-validation is utilized to evaluate the performance of all approaches. The max epoch is set to 70, and the batch size is 32. We optimize all approaches using ADAM [94].

Table 4-6 Results of five deep NNs using multitask learning.

Classifiers	**Accuracy (%) ↑**		**Sensitivity (%) ↑**		**Specificity (%) ↑**		**F_1 ↑**		**AUC (%) ↑**		**FPR (%) ↓**		**FNR (%) ↓**	
	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi
DenseNet121	82.2	85.0	75.3	79.1	87.1	88.9	0.76	0.80	81.2	84.0	12.9	11.1	24.7	20.9
MobileNet	85.1	87.0	78.1	81.1	90.2	91.0	0.81	**0.83**	84.1	**86.1**	9.8	9.0	21.9	18.9
ResNet50	85.1	86.1	78.5	80.1	89.2	89.0	0.80	0.81	83.8	85.0	10.7	10.9	21.5	21.3
VGG16	**86.1**	87.1	81.0	**81.3**	**91.2**	90.9	**0.82**	**0.83**	**86.1**	**86.1**	**8.8**	9.1	19.0	**18.7**
EfficientNetB0	84.2	**87.5**	**81.2**	81.0	86.9	**91.2**	0.80	**0.83**	84.0	**86.1**	13.1	**8.8**	**18.8**	19.0

4.5.1 The Effectiveness of Multitask Learning Using Generic Deep Learning Models

Many previous studies [34-36] have demonstrated the effectiveness of integrating tumor segmentation tasks into tumor classification networks. In BUS images, the shared representations between tumor classification and segmentation tasks include tumor morphology, size, shape, and echo pattern. We evaluate multitask learning networks with five different pretrained (ImageNet) backbone networks, DenseNet121, MobileNet, ResNet50, VGG16, and EfficientNetB0. A subnetwork [80] is added to perform breast tumor segmentation at the end of the convolutional layers of the backbone network. The subnetwork consists of four blocks, each of which contains one upsampling layer, and two consecutives 3×3 convolution layers with batch normalization and ReLU activation.

The loss function is a combination of both the Dice loss and binary cross-entropy loss. The weight for the binary cross-entropy loss is set to 1.5 by experiments.

Table 4-7 Results of three multitask learning approaches developed for BUS image classification.

Approaches	**Accuracy (%) ↑**	**Sensitivity (%) ↑**	**Specificity (%) ↑**	**F_1 ↑**	**AUC (%) ↑**	**FPR (%) ↓**	**FNR (%) ↓**
Zhang, et al. [81]	87.4	81.4	**91.4**	0.83	86.4	**8.6**	18.6
Vakanski, et al. [80]	83.6	77.4	87.8	0.78	82.6	12.2	22.5
Shi, et al. [82]	83.9	87.3	81.7	0.80	84.5	18.3	12.6
MT-ESTAN	**90.0**	**90.4**	89.8	**0.88**	**90.1**	10.2	**9.6**

As shown in Table 4-8, with the additional segmentation task, the overall performance of the five approaches can be improved. VGG16, MobileNet, and EfficientNetB0 achieve the best AUC of 86.1% among all the approaches. The sensitivity of DenseNet121 is improved by 5%. It is worth noticing that, in all approaches, the specificity values are significantly higher compared to the sensitivity values. We observed the same outcome in [80] and [81]. This issue could be addressed by choosing the weighted binary cross-entropy function.

4.5.2 The Effectiveness of the Proposed MT-ESTAN

In this section, we compare the proposed MT-ESTAN with three multitask learning approaches [34-36]. We obtained the source code from the authors of [34, 36], and implemented the approach in [35], all model parameters were adopted from the papers.

As shown in Table 4-9, the AUC of the proposed MT-ESTAN is significantly higher than those of [80], [82], and [81], and MT-ESTAN outperforms all approaches reported in Table 4-8. For example, compared to the best performed multitask network (VGG16) in Table 4-8, the proposed MT-ESTAN improves the sensitivity, F_1 score, and AUC by 11.2%, 7.3%, and 4.6%, respectively. However, [34-36] are not significantly better than the multitask learning approaches reported in Table 4-8. [34-36] achieves high specificity values, but at the cost of low sensitivity values, which leads to high false negative rates (FNRs), e.g., the FNR of [36] is 18.6%. In addition, all multitask learning approaches have low sensitivity values and high FNRs. The proposed MT-ESTAN achieves a better balance between sensitivity and specificity and has a low FNR of 9.6%.

4.6 Discussion

The experiments and similar outcomes in [36, 50, 51] demonstrate that the transfer learning (TL) strategy consistently outperforms training from scratch for deep learning approaches for BUS image classification, which implies that knowledge learned from a different domain (e.g., nature images) could be transferred and used to improve BUS image classification. BUS images share common image elements in natural images, e.g., object boundaries, image contrast, and texture, and deep neural networks learning the representations of those elements from nature images can also contribute to BUS image classification. Inspired by this, medical image datasets sharing common features with BUS images could be applied to further improve the performance of deep learning approaches for BUS image classification. For example, ultrasound images from other organs and breast images from other modalities (e.g., MRI, CT, and Mammogram) can be used to pretrain BUS image classifiers.

Our results and previous studies [64] suggest that image augmentation techniques could improve the generalizability of most deep learning approaches for BUS image classification. Augmentation techniques introduce variations and enlarge the training set size, and could prevent overfitting [97]; and model training using an augmented dataset alleviates the issue of the small size of the medical dataset. To further increase the generalizability of deep learning models, the simplest way is to add more images from different sources to the model training. The additional images could be either new real BUS images or synthetic images generated using algorithms [98].

Many BUS image classification approaches have achieved promising overall performance (e.g., accuracy and F_1 score), but failed to balance the sensitivity and specificity. They used the binary cross-entropy as the loss function and treat cancer and non-cancer cases equally, which makes predictions that favor the dominant class, e.g., benign class, and produce low sensitivities. Sensitivity is the most important assessment metric in breast cancer detection because missing malignant cases may risk patients' lives; and a well-balanced model should achieve both high overall performance and high sensitivity.

One solution is to utilize the numerically-stable weighted cross-entropy function discussed in Section 3.4 to achieve a better balance between the sensitivity and specificity.

Multitask learning (MTL) is a promising future direction to improve the robustness and generalization of deep learning approaches for BUS image classification. Table 4-8 demonstrates that MTL networks with a primary BUS tumor classification task and a secondary segmentation task outperform single-task networks with only the classification task. The segmentation task incorporates semantic information, i.e., tumor region, during the training, which enables an MTL network to learn meaningful and focused representations in tumor regions rather than random features from a whole

BUS image. This secondary task performs as a regularizer that could also improve models' convergence using small or medium datasets. Inspired by this finding, researchers can further advance BUS image classification by incorporating other semantic knowledge, e.g., breast anatomy and BI-RADs descriptors, into MTL networks.

Last but not least, to improve the adoption and trustworthiness of CAD systems for breast cancer detection, the explainability of approaches should be improved. Existing deep learning-based methods still have a black-box nature in which limited information is provided to help understand the BUS image classification process [54-55]. This gap discourages radiologists from using BUS CADs in clinical practice. Therefore, solving this gap by introducing explainability into models [99] is a promising direction for BUS image classification.

4.7 Conclusion

In this work, we build a public benchmark for the classification of B-mode BUS images which consists of a diverse dataset, useful strategies, and findings for developing deep learning-based approaches, and a novel MTL network, MT-ESTAN, for accurate BUS image classification.

The benchmark dataset comprises 3,641 B-mode BUS images from five countries, and a set of public software tools for data preparing and preprocessing. The BUS images were collected with different ultrasound devices and patient populations, and have a wide variation in image contrast, brightness, level of noise, etc. We highlight three major findings by evaluating 10 deep learning-based approaches using the benchmark dataset: 1) Transfer learning and image augmentation are effective strategies to significantly improve the overall performance of deep learning-based BUS image classifiers; 2) the numerically-stable weighted cross-entropy loss function offers a better balance between the sensitivity and specificity; 3) MTL networks with both the breast tumor segmentation and classification tasks is one of the most useful strategies to improve the generalization of deep learning approaches for BUS image classification.

The newly proposed MT-ESTAN incorporates a small-tumor aware network as the backbone network and consists of one primary task (tumor classification) and a secondary task (tumor segmentation). The results show that MT-ESTAN achieves state-of-the-art performance, and significantly improved the sensitivity of the model.

In the future, we will be continuously adding more BUS images, new findings, and emerging approaches to the benchmark.

Chapter 5: Breast Ultrasound Tumor Classification Using a Hybrid Multitask CNN-Transformer Network

Bryar Shareef, Min Xian, Aleksandar Vakanski, Haotian Wang, Breast Ultrasound Tumor Classification using a Hybrid Multitask CNN-Transformer Network: MICCAI, 2023.

5.1 Introduction

Breast cancer is the leading cause of cancer-related fatalities among women. Currently, it holds the highest incidence rate of cancer among women in the U.S., and in 2022 it accounted for 31% of all newly diagnosed cancer cases [53]. Due to the high incidence rate, early breast cancer detection is essential for reducing mortality rates and expanding treatment options. BUS imaging is an effective screening option because it is cost-effective, nonradioactive, and noninvasive. However, BUS image analysis is also challenging due to the large variations in tumor shape and appearance, speckle noise, low contrast, weak boundaries, and occurrence of artifacts.

In the past decade, deep learning-based approaches achieved remarkable advancements in BUS tumor classification [101][102]. The progress has been driven by the capability of CNN-based models to learn hierarchies of structured image representations as semantics. To extract deep context features, CNNs apply a series of convolutional and downsampling layers, frequently organized into blocks with residual connections. Nevertheless, one disadvantage of such architectural choice is that the feature representations in the deeper layers become increasingly abstract, leading to a loss of spatial and contextual information. The intrinsic locality of convolutional operations hinders the ability of CNNs to model long-range dependencies while preserving spatial information in images effectively.

Vision Transformer (ViT) [103] and its variants recently demonstrated superior performance in image classification tasks. These models convert input images into smaller patches and utilize the self-attention mechanism to model the relationships between the patches. Self-attention enables ViTs to capture long-range dependencies and model complex relationships between different regions of the image. However, the effectiveness of ViT-based approaches heavily relies on access to large datasets for learning meaningful representations of input images. This is primarily because the architectural design of ViTs does not rely on the same inductive biases in feature extraction which allow CNNs to learn spatially invariant features.

Accordingly, numerous prior studies introduced modifications to the original ViT network specifically designed for BUS image classification [104] [105][106][107]. In addition, several works

proposed network architectures that combined Transformers and CNNs [108-110]. For instance, Mo et al. [108] proposed a hybrid CNN-Transformer incorporating BUS anatomical priors. Qu et al. [109] employed squeeze and excitation blocks to enhance the feature extraction capacity in a hybrid CNN-based VGG16 network and ViT. Similarly, Iqbal et al. [110] designed two hybrid CNN-Transformer networks intended either for classification or segmentation of multi-modal breast cancer images. Despite the promising results of such hybrid approaches, effectively capturing the local patterns and global long-range dependencies in BUS images remains challenging [110][103][111].

Multitask learning leverages shared information across related tasks by jointly training the model. It constrains models to learn representations that are relevant to all tasks rather than learning task-specific details. Moreover, multitask learning acts as a regularizer by introducing inductive bias and prevents overfitting [112] (particularly with ViTs), and with that, can mitigate the challenges posed by small BUS dataset sizes. In [102], the authors demonstrated that multitask learning outperforms single-task learning approaches for BUS classification.

In this study, we introduce a hybrid multitask approach, Hybrid-MT-ESTAN, which encompasses tumor classification as a primary task and tumor segmentation as a secondary task. Hybrid-MT-ESTAN combines the advantages of CNNs and Transformers in a framework incorporating anatomical tissue information in BUS images. Specifically, we designed a novel attention block named Anatomy-Aware Attention (AAA), which modifies the attention block of Swin Transformer by considering the breast anatomy. The anatomy of the human breast is categorized into four primary layers: the skin, premammary (subcutaneous fat), mammary, and retromammary layers, where each layer has a distinct texture and generates different echo patterns. The primary layers in BUS images are arranged in a vertical stack, with similar echo patterns appearing horizontally across the images. The kernels in the introduced AAA attention blocks are organized in rows and columns to capture the anatomical structure of the breast tissue. In the published literature, the closest approach to ours is the work by Iqbal et al. [110], in which the authors used hybrid single-task CNN-Transformer networks for either classification or segmentation of BUS images. Conversely, Hybrid-MT-ESTAN employs a multitask approach and introduces novel architectural design. The main contributions of this work are summarized as:

a) The proposed architecture effectively integrates the advantages of CNNs for extracting hierarchical and local patterns in BUS images and Swin Transformers for leveraging long-range dependencies.
b) The designed Anatomy-Aware Attention (AAA) block improves the learning of contextual information based on the anatomy of the breast.

c) The multitask learning approach leverages the shared representations across the classification and segmentation tasks to improve the model performance.

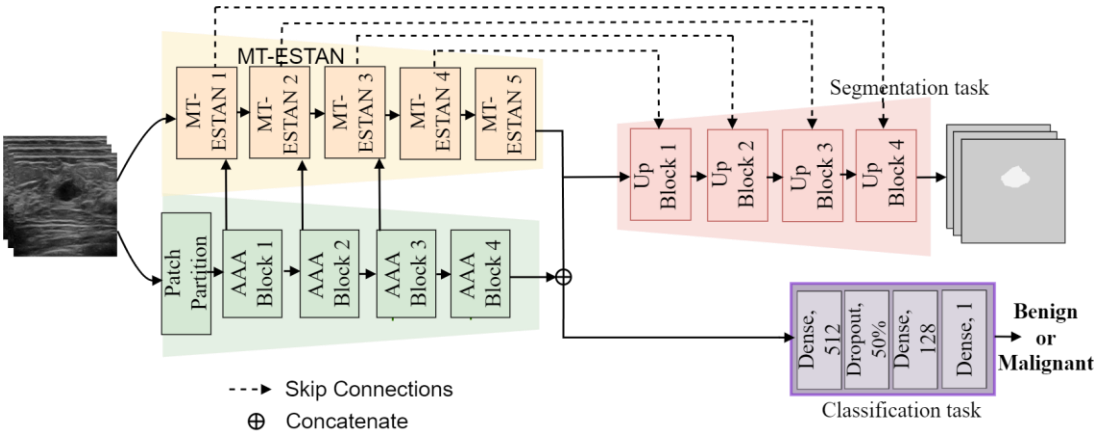


Figure 5-1 Hybrid-MT-ESTAN consists of MT-ESTAN and AAA encoders, a segmentation decoder, and a classification branch.

5.2 Proposed Method

5.2.1 Hybrid-MT-ESTAN

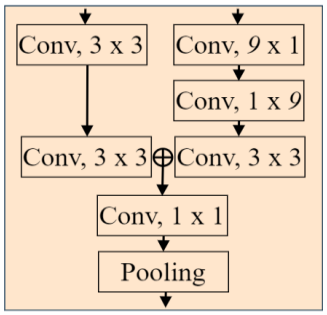


Figure 5-2 MT-ESTAN blocks include parallel convolutional branches with different kernel size, followed by 1x1 convolution and a pooling layer.

The architecture of Hybrid-MT-ESTAN is shown in Figure 5-1, and it consists of (1) a CNN-based encoder MT-ESTAN, and a Swin Transformer-based encoder with Anatomy-Aware Attention (AAA) blocks, (2) a decoder branch for the segmentation task, and (3) a branch with fully-connected layers for the classification task. MT-ESTAN [102] is a CNN-based multitask learning network that simultaneously performs BUS classification and segmentation.

The encoder sub-network of MT-ESTAN is ESTAN [12], which employs row-column-wise kernels to learn and fuse context information in BUS images at different context scales (See Figure 5-2). Specifically, each MT-ESTAN block is composed of two parallel branches consisting of four square convolutional kernels and two consecutive row-column-wise kernels. These specialized convolutional kernels effectively extract contextual information of small tumors in BUS images. Refer to [12][33] and [33]for the implementation details of ESTAN and MT-ESTAN. The source codes of the proposed work are available at http://busbench.midalab.net.

5.2.2 Anatomy-Aware Attention (AAA) Block

Swin Transformer [33] is a hierarchical Transformer-based approach that uses shifted windows to model global context information. Swin Transformer partitions an input image into non-overlapping patches of siz4 × 4e, where each patch is treated as a "token." A linear layer receives the patches and projects them into an arbitrary dimension. Each Swin Transformer block consists of a LayerNorm layer (LN) layer, a multi-head self-attention module (MSA), and a multi-layer perceptron (MLP) with GELU activation. To model long-range dependencies, the original Swin Transformer relies on shifted windows, where the window-based multi-head self-attention ($\mathrm{W_{MSA}}$) and shifted window-based multi-head self-attention ($\mathrm{SW_{MSA}}$) modules are employed in each consecutive Swin block. The Swin block is formulated as follows:

$$\hat{f}^{l} = W\_MSA\left(LN\left(f^{l-1}\right)\right) + f^{l-1} \tag{5.1}$$

$$f^{l} = MMLP\left(LN\left(\hat{f}^{l}\right)\right) + \hat{f}^{l} \tag{5.2}$$

$$\hat{f}^{l+1} = SW\_MSA\left(LN\left(f^{l}\right)\right) + f^{l} \tag{5.3}$$

$$f^{l+1} = MLP\left(LN\left(\hat{f}^{l+1}\right)\right) + \hat{f}^{l+1} \tag{5.4}$$

Where f^{l} and $\hat{f}^{l}$ are the output features of the MLP module and the (S)$\mathrm{W_{MSA}}$module for block 1, respectively; In the proposed Anatomy-Aware Attention (AAA) block, we redesigned the Swin blocks to enhance their ability to model global and local features by adding an attention block based on the breast anatomy (see Figure 5-3). The additional layers are defined as

$$y^{i} = M\left(f^{l+1}\right) \tag{5.5}$$

$$B\hat{}i = U(MAX\_P(y\hat{}i\,) + AVG\_P(y\hat{}i) \tag{5.6}$$

$$O^i = y^i \cdot \left(\sigma\left(A(B)\right)\right) \tag{5.7}$$

Concretely, we first reconstruct the i^{th} feature map (y^i) by merging all image patches (M), and afterward, we applied average pooling ($\mathrm{AVG_P}$) and max pooling ($\mathrm{MAX_P}$) layers with size (2, 2). The outputs of ($\mathrm{AVG_P}$) and ($\mathrm{MAX_P}$) layers are concatenated and up-sampled (U) with size (2, 2) and stride (2, 2). Row-column-wise kernels (A) with size (9 x 1) and (1 x 9) are then employed to adapt to the anatomy of the breast, and finally, a sigmoid function (σ) is applied to the output of (A) multiplied by the input feature map (y^i).

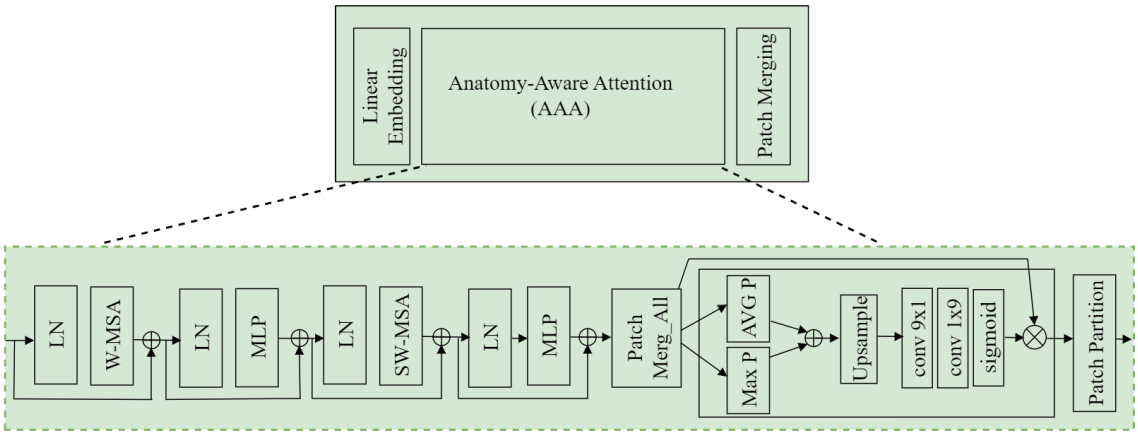


Figure 5-3 Anatomy-Aware Attention (AAA) block.

The segmentation branch in Figure 5-1 outputs dense mask predictions of BUS tumors. It consists of four blocks, Up Blocks 1-4, each with three convolutional layers and one upsampling layer(size (2, 2) and stride (2, 2)). The settings of the convolutional layers are adopted from [102] for the details of the convolutional layers. In addition, the blocks receive four skip connections from the MT-ESTAN encoder, i.e., there is a skip connection from each MT-ESTAN block 1 to 4. The classification branch consists of three dense layers, a dropout layer (50%), and the final dense layer that predicts the tumor class into benign or malignant.

5.2.3 Loss Function

We applied a multitask loss function (L_{mt}) that aggregates two terms: a focal loss L_{Focal} for the classification task and dice loss L_{Dice} for the segmentation task. Therefore, the composite loss function is $L_{mt} = w_1 \cdot L_{Focal} + L_{Dice}$, where the weight coefficient w_1 is set to apply greater importance to the classification task as the primary task. Since in medical image diagnosis achieving high sensitivity places emphasis on the detection of malignant lesions, we employed the focal loss for the classification task to trade-off between sensitivity and specificity. Because malignant tumors are more challenging to

detect due to greater differences in margin, shape, and appearance in BUS images, focal loss forces the model to focus more on difficult predictions. Specifically, focal loss adds a factor $(1-p_i)\hat{}\gamma$ to the cross-entropy loss where γ is a focusing parameter, resulting in $L_{Focal} = -1 / N \sum_{i=1}^{N}[(\alpha \cdot t_i \cdot (1-p_i)^{\gamma}) \cdot \log(p_i) + (1-\alpha) \cdot p_i \cdot \log(1-p_i)]$. In the formulation, α is a weighting coefficient, N denotes the number of image samples, t_i is the target label of the i^{th} training sample, and p_i denotes the prediction. The segmentation loss is calculated using the commonly-employed Dice loss (L_{Dice}) function.

5.3 Experimental Results

In this section, we describe the datasets, evaluation metrics, and various implementation details. Then we present ablation studies to verify the effectiveness of each component in the proposed architecture.

5.3.1 Datasets

We evaluated the performance of the proposed approach on four public datasets, HMSS [113], BUSI [84], BUSIS [11], and Dataset B [11]. We combined all four datasets to build a large and diverse dataset with a total of 3,320 B-mode BUS images, of which 1,664 contain benign tumors and 1,656 have malignant tumors. Table 5-1 shows the detailed information for each dataset. HMSS dataset does not provide the segmentation ground truth masks, and for this study, they were prepared by a group of experienced radiologists. Refer to the original publications of the datasets for more details.

Table 5-1 Four public breast ultrasound (BUS) datasets. B denotes a benign tumor, and M is a malignant tumor.

BUS dataset	No. of images	Distribution	Source
HMSS	1948	B:812, M:1136	Netherlands
BUSI	647	B:437, M:210	Egypt
BUSIS	562	B:306, M:256	China
Dataset B	163	B:109, M:54	Spain
Total	3320	B: 1664, M: 1656	

5.3.2 Evaluation Metrics

For performance evaluation of the classification task, we used the following metrics: accuracy (Acc), sensitivity (Sens), specificity (Spec), F1 score, Area Under the Curve of Receiver Operating Characteristic (AUC), false positive rate (FPR), and false negative rate (FNR).

To evaluate the segmentation performance, we used the dice score coefficient (DSC) and Jaccard index (JI).

5.3.3 Implementation Details

The proposed approach was implemented with Keras and TensorFlow libraries. All experiments were performed on a machine with NVIDIA Quadro RTX 8000 GPUs and two Intel Xeon Silver 4210R CPUs (2.40GHz) with 512 GB of RAM. All BUS images in the dataset were zero-padded and reshaped to form square images. To avoid data leakage and bias, we selected the train, test, and validation sets based on the cases, i.e., the images from one case (patient) were assigned to only one of the training, validation, and test sets. Furthermore, we employed horizontal flip, height shift (20%), width shift (20%), and rotation (20 degrees) for data augmentation. The proposed approach utilizes the building blocks of ResNet50 and Swin-Transformer-V2, pretrained on ImageNet dataset. Namely, MT-ESTAN uses pretrained ResNet50 as a base model for the five encoder blocks (the implementation details of MT-ESTAN can be found in [102]). The encoder with AAA blocks uses the SwinTransformer_V2_Base_256 pretrained model as a backbone. For the composite loss function, we adopted a weight coefficient $w_1 = 3$, and in the focal loss $\alpha = 0.5$ and $\gamma = 2$. For model training, we utilized Adam optimizer with a learning rate of 10^{-5} and mini batch size of 4 images.

5.3.4 Performance Evaluation and Comparative Analysis

We compared the performance of Hybrid-MT-ESTAN with with eight deep learning approaches commonly used for medical image analysis, which include CNN-based, ViT-based, and hybrid approaches. CNN-based networks include SHA-MTL [81], MobileNet [86], DenseNet121 [88], and EMT-Net [88]. ViT-based approaches include the original ViT [88], Chowdery [114], and Swin Transformer [115]. VGGA-ViT [109] is a hybrid CNN-Transformer network. The values of the performance metrics are shown in Table 5-2, indicating that the proposed Hybrid-MT-ESTAN outperformed all eight approaches by achieving the best accuracy, sensitivity, F1 score, and AUC with 82.8%, 86.4, 86.0%, and 82.8, respectively. Although SHA-MTL [81] obtained the highest specificity of 90.8% and FPR of 9.2%, the trade-off between sensitivity and specificity should be taken into consideration, as that approach had a sensitivity of 48.1%. The preferred trade-off in medical image analysis typically is high sensitivity without significant degradation in specificity.

We evaluated the segmentation performance of Hybrid MT-ESTAN and compared the results to five multitask approaches, including SHA-MTL [81], EMT-Net [88], Chowdery [114], MT-ESTAN [102], and VGGA-ViT [109]. Table 5-2 presents the quantitative results. The proposed Hybrid MT-

ESTAN achieved the highest performance and increased DSC and JI by 5.9% and 6.4%, respectively, compared to MT-ESTAN. Note that the models in Table 5-2 for which the segmentation results are not provided are single-task models.

In our experiments in Table 5-2, the proposed approach is compared to four singletask approaches, including MobileNet, DenseNet121, ViT, and Swin Transformer, and five multitask approaches including SHA-MTL, VGGA-ViT, EMT-Net, Chowdery, and MT-ESTAN.

Table 5-2 Performance metrics of the compared BUS image classification and segmentation methods.

	Classification							Segmentation	
Methods	Acc	Sens.	Spec.	F1	Auc	FNR	FPR	DSC	JI
SHA-MTL	69.6	48.1	**90.8**	0.58	69.5	51.9	9.2	72.2	60.7
MobileNet	71.0	82.0	61.0	0.74	71.5	18.0	39.0	-	-
VGGA-ViT	73.6	61.8	79.8	0.61	70.8	38.2	20.2	74.9	64.9
DenseNet121	73.0	74.0	71.0	0.73	72.5	26.0	29.0	-	-
EMT-Net	74.1	79.4	69.1	0.75	74.3	20.6	30.9	76.7	67.0
ViT	72.1	74.1	69.3	0.73	71.7	25.9	30.7	-	-
Chowdery	77.4	77.3	77.3	0.77	77.3	22.7	22.7	77.0	67.9
Swin Transformer	77.4	72.6	82.5	0.74	77.6	27.4	17.5	-	-
MT-ESTAN	78.6	83.7	72.6	0.83	78.2	16.3	27.4	78.2	69.3
Ours	**82.8**	**86.4**	79.2	**0.86**	**82.8**	**13.6**	20.8	**84.1**	**75.7**

Table 5-3 Effectiveness of the Anatomy-Aware Attention (AAA) Block

	Classification							Segmentation	
Methods	Acc	Sens.	Spec.	F1	Auc	FNR	FPR	DSC	JI
MT-ESTAN	78.6	83.7	72.6	0.83	78.2	16.3	27.4	78.2	69.3
Swin Transformer	77.4	72.6	82.5	0.74	77.6	27.4	17.5	-	-
MT-ESTAN + Swin Transformer	80.3	84.2	76.3	0.83	80.2	15.8	23.7	82.3	73.6
Hybrid MT-ESTAN + AAA (Ours)	**82.8**	**86.4**	79.2	**0.86**	**82.8**	**13.6**	20.8	**84.1**	**75.7**

5.3.5 Effectiveness of the Anatomy-Aware Attention (AAA) Block

To verify the effectiveness of the Anatomy-Aware Attention (AAA) block, we conducted an ablation study that quantified the impact of the different components in Hybrid-MT-ESTAN on the classification and segmentation performance. Table 5-3 presents the values of the performance metrics

for MT-ESTAN (pure CNN-based approach), Swin Transformer (pure Transformer network), a hybrid architecture of MT-ESTAN and Swin Transformer, and our proposed Hybrid-MT-ESTAN with AAA block. According to the results in Table 5-3, MT-ESTAN achieved better sensitivity and F1 score than Swin Transformer, with 83.7% and 83%, respectively. The hybrid architectures of MT-ESTAN with Swin Transformer improved the classification performance and has higher accuracy, sensitivity, F1 score, and AUC with 80.3%, 84.2%, 83%, and 80.2%, compared to MT-ESTAN and Swin Transformer individually. The proposed approach, Hybrid-MT-ESTAN with AAA block, further improved accuracy, sensitivity, F1 score, and AUC by 2.5%, 2.2%, 3%, and 2.6%, respectively, relative to the hybrid model without the AAA block. We compared the proposed approach with and without the AAA block and Swin Transformer to evaluate the segmentation performance. As shown in Table 5-3, MT-ESTAN combined with Swin Transformer improved DSC and JI by 4.1% and 4.3%, respectively, compared to MT-ESTAN. Employing the proposed AAA block further improved DSC and JI by 1.8% and 2.1%, respectively.

5.4 Conclusion

In this paper, we introduce Hybrid-MT-ESTAN, a multitask learning approach for BUS image analysis that alleviates the lack of global contextual information in the low-level layers of CNN-based approaches. Hybrid-MT-ESTAN concurrently performs BUS tumor classification and segmentation with a hybrid architecture that employs CNN-based and Swin Transformer layers. The proposed learning approach exploits multi-scale local patterns and global long-range dependencies provided by MT-ESTAN and AAA Transformer blocks for learning feature representations that resulted in improved generalization. Experimental validation demonstrated significant performance improvement of Hybrid-MT-ESTAN compared to current state-of-the-art models for BUS image classification.

Chapter 6: Conclusion and Future Work

In this dissertation, we have made significant contributions to the field of breast cancer early detection through the development of deep learning approaches. Firstly, we proposed novel deep learning models for breast ultrasound image segmentation. We addressed the current challenges and developed innovative approaches to segment breast ultrasound tumors accurately. These segmentation models provide a crucial step in computer-aided diagnosis (CAD) systems, enabling more precise tumor quantification and assisting healthcare professionals in making informed decisions. Secondly, we established the first and largest breast ultrasound image classification benchmark. The benchmark provides a diverse dataset and serves as a foundation for evaluating the effectiveness of different classification algorithms. Furthermore, we proposed a novel multitask learning approach to perform classification and segmentation. Lastly, a hybrid multitask CNN-Transformer network was proposed for breast ultrasound tumor classification. The approach effectively captures both local and global information in breast ultrasound images, leading to improved tumor classification performance.

The outcomes of this dissertation have far-reaching implications for the medical community and breast cancer research. Our work provides a solid foundation for further advancements in deep learning-based approaches for breast cancer detection. It opens doors to developing more sophisticated algorithms, improved CAD systems, and personalized medicine approaches tailored to individual patients.

While our research has made significant strides, there are still challenges and opportunities for future investigation. Continued efforts in refining segmentation algorithms, enhancing classification models, interpretability and explainability, and dataset expansion.

6.1 List of Publications

- **B. Shareef,** M. Xian, A. Vakanski, J.Ding, C. Ning, H.D. Cheng, “A benchmark for breast ultrasound image classification,” Ultrasound in Medicine & Biology, 2023, under review.
- **Bryar Shareef,** Min Xian, Aleksandar Vakanski, Haotian Wang, Breast Ultrasound Tumor Classification using a Hybrid Multitask CNN-Transformer Network, 2023, in MICCAI 2023. Accepted.
- H. Wang, M. Xian, A. Vakanski, and **B. Shareef**, “SIAN: style-guided instance-adaptive normalization for multi-organ histopathology image synthesis,” in IEEE International Symposium on Biomedical Imaging, 2023, pp. 1-5. Accepted.
- **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, pp. 2262, 2022.
- Y. Zhang, M. Xian, H.-D. Cheng, **B. Shareef**, J. Ding, F. Xu, K. Huang, B. Zhang, C. Ning, and Y. Wang, “BUSIS: A Benchmark for Breast Ultrasound Image Segmentation,” *Healthcare*, vol. 10, no. 4, pp. 729, 2022-04-14, 2022.
- **B. Shareef**, M. Xian, and A. Vakanski, “STAN: Small tumor-aware network for breast ultrasound image segmentation,” in IEEE International Symposium on Biomedical Imaging, Iowa City, Iowa, USA, 2020, pp. 1-5.
- R. E. Hiromoto, M. Haney, A. Vakanski, and **B. Shareef**, “Toward a Secure IoT Architecture,” *Advanced Control Techniques in Complex Engineering Systems: Theory and Applications: Dedicated to Professor Vsevolod M. Kuntsevich*, pp. 297-323, 2019.
- **B. Shareef,** E. de Doncker and J. Kapenga, "Monte Carlo simulations on Intel Xeon Phi: Offload and native mode," 2015 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 2015, pp. 1-6, doi: 10.1109/HPEC.2015.7322456.

Bibliography

[1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2019,” CA. Cancer J. Clin., vol. 69, no. 1, pp. 7–34, 2019, doi: 10.3322/caac.21551.

[2] Y. Ikedo et al., “Development of a fully automatic scheme for detection of masses in whole breast ultrasound images,” Med. Phys., vol. 34, no. 11, pp. 4378–4388, 2007, doi: 10.1118/1.2795825.

[3] A. Madabhushi and D. N. Metaxas, “Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions,” IEEE Trans. Med. Imaging, vol. 22, no. 2, pp. 155–169, 2003, doi: 10.1109/TMI.2002.808364.

[4] Y. L. Huang and D. R. Chen, “Automatic contouring for breast tumors in 2-D sonography,” Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., vol. 7, pp. 3225–3228, 2005, doi: 10.1109/IEMBS.2005.1617163.

[5] M. H. Yap et al., “Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks,” IEEE J. Biomed. Heal. Informatics, vol. 22, no. 4, pp. 1218–1226, 2018, doi: 10.1109/JBHI.2017.2731873.

[6] J. Z. Cheng et al., “Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans,” Sci. Rep., vol. 6, no. October 2015, pp. 1–13, 2016, doi: 10.1038/srep24454.

[7] B. Huynh, K. Drukker, and M. Giger, “Computer-Aided Diagnosis of Breast Ultrasound Images Using Transfer Learning From Deep Convolutional Neural Networks,” Med. Phys., vol. 43, no. 6, pp. 3705–3705, 2016.

[8] T. Fujioka et al., “Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network,” Jpn. J. Radiol., 2019, doi: 10.1007/s11604-019-00831-5.

[9] M. Lin, Q. Chen, and S. Yan, “Network In Network,” pp. 1–10, 2013, [Online]. Available: http://arxiv.org/abs/1312.4400.

[10] F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016, pp. 565–571, 2016, doi: 10.1109/3DV.2016.79.

[11] Y. Zhang et al., “A Benchmark for breast ultrasound image segmentation (BUSIS),” 2018, [Online]. Available: http://arxiv.org/abs/1801.03182.

[12] 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, 2022, doi: 10.3390/healthcare10112262.

[13] M. A. Elaziz, D. Oliva, A. A. Ewees, and S. Xiong, “Multi-level thresholding-based grey scale image segmentation using multi-objective multi-verse optimizer,” Expert Syst. Appl., vol. 125, pp. 112–129, 2019, doi: 10.1016/j.eswa.2019.01.047.

[14] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, “Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors,” Breast Cancer Res. Treat., vol. 89, no. 2, pp. 179–185, 2005, doi: 10.1007/s10549-004-2043-z.

[15] M. H. Yap, E. A. Edirisinghe, and H. E. Bez, “A novel algorithm for initial lesion detection in ultrasound breast images,” J. Appl. Clin. Med. Phys., vol. 9, no. 4, pp. 181–199, 2008, doi: 10.1120/jacmp.v9i4.2741.

[16] 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, 2012, doi: 10.1016/j.ultrasmedbio.2011.10.022.

[17] 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, 2015, doi: 10.1016/j.patcog.2014.07.026.

[18] J. Shan, H. D. Cheng, and Y. Wang, “A novel automatic seed point selection algorithm for breast ultrasound images,” in 19th International Conference on Pattern Recognition, 2008, pp. 1–4, doi: 10.1109/icpr.2008.4761336.

[19] J. Massich, F. Meriaudeau, E. Pérez, R. Martí, A. Oliver, and J. Martí, “Lesion segmentation in breast sonography,” in Lecture Notes in Computer Science, vol 6136, 2010, pp. 39–45, doi: 10.1007/978-3-642-13666-5_6.

[20] C. Lo et al., “Multi-Dimensional Tumor Detection in Automated,” IEEE Trans. Med. Imaging, vol. 33, no. 7, pp. 1503–1511, 2014.

[21] Y. Zhang et al., “BUSIS: A Benchmark for Breast Ultrasound Image Segmentation,” Healthc., vol. 10, no. 4, 2022.

[22] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, and B. Zhang, “Semantic segmentation of breast ultrasound image with fuzzy deep learning network and breast anatomy constraints,” Neurocomputing, vol. 450, pp. 319–335, 2021, doi: 10.1016/j.neucom.2021.04.012.

[23] M. Amiri, R. Brooks, and H. Rivaz, “Fine-tuning U-Net for ultrasound image segmentation: different layers, different outcomes.,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, no. June, pp. 1–1, 2020, doi: 10.1109/tuffc.2020.3015081.

[24] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. L. Bell, “Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, pp. 1–1, 2020, doi: 10.1109/tuffc.2020.2993779.

[25] Z. Zhuang, N. Li, A. N. J. 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, pp. 1–23, 2019, doi: 10.1371/journal.pone.0221535.

[26] Y. Hu et al., “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, 2018, doi: 10.1002/mp.13268.

[27] A. Vakanski, M. Xian, and P. E. Freer, “Attention-Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images,” Ultrasound Med. Biol., vol. 46, no. 10, pp. 2819–2833, 2020, doi: 10.1016/j.ultrasmedbio.2020.06.015.

[28] M. Byra and M. Andre, “Breast mass classification in ultrasound based on Kendall’s shape manifold,” arXiv, no. May, 2019.

[29] W. K. Moon et al., “Computer-aided tumor detection in automated breast ultrasound using a 3-D convolutional neural network,” Comput. Methods Programs Biomed., vol. 190, p. 105360, 2020, doi: 10.1016/j.cmpb.2020.105360.

[30] H. Lee, J. Park, and J. Y. Hwang, “Channel Attention Module with Multi-scale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 67, no. 7, pp. 1344–1353, 2020, doi: doi: 10.1109/TUFFC.2020.2972573.

[31] G. Chen, Y. Dai, and J. Zhang, “C-Net: Cascaded convolutional neural network with global guidance and refinement residuals for breast ultrasound images segmentation,” Comput. Methods Programs Biomed., vol. 225, p. 107086, 2022, doi: 10.1016/j.cmpb.2022.107086.

[32] S. Hussain et al., “A Discriminative Level Set Method with Deep Supervision for Breast Tumor Segmentation,” Comput. Biol. Med., vol. 149, no. April, p. 105995, 2022, doi: 10.1016/j.compbiomed.2022.105995.

[33] B. Shareef, M. Xian, and A. Vakanski, “STAN : Small Tumor-Aware Network for Breast Ultrasound Image Segmentation,” IEEE 17th Int. Symp. Biomed. Imaging (ISBI 2020), 2020.

[34] N. M. ud din, R. A. Dar, M. Rasool, and A. Assad, “Breast cancer detection using deep learning: Datasets, methods, and challenges ahead,” Comput. Biol. Med., vol. 149, no. January, p. 106073, 2022, doi: 10.1016/j.compbiomed.2022.106073.

[35] M. Amiri, R. Brooks, and H. Rivaz, “Fine tuning U-net for ultrasound image segmentation: Which layers?,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11795 LNCS, pp. 235–242, 2019, doi: 10.1007/978-3-030-33391-1_27.

[36] M. Byra et al., “Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network,” Biomed. Signal Process. Control, vol. 61, 2020, doi: 10.1016/j.bspc.2020.102027.

[37] J. Chen, Chenyi; Liu, Ming-Yu; Tuzel, C. Oncel; Xiao, “R-CNN for Small Object Detection,” Asian Conference on Computer Vision. LNCS, vol. 10115. Springer, Cham, pp. 214–230, 2016.

[38] H. Krishna and C. V. Jawahar, “Improving small object detection,” Proc. - 4th Asian Conf. Pattern Recognition, ACPR 2017, pp. 346–351, 2018, doi: 10.1109/ACPR.2017.149.

[39] L. Guan, Y. Wu, and J. Zhao, “SCAN: Semantic context aware network for accurate small object detection,” Int. J. Comput. Intell. Syst., vol. 11, no. 1, pp. 951–961, May 2018, doi: 10.2991/ijcis.11.1.72.

[40] R. Dong, X. Pan, and F. Li, “DenseU-Net-Based Semantic Segmentation of Small Objects in Urban Remote Sensing Images,” IEEE Access, vol. 7, pp. 65347–65356, 2019, doi: 10.1109/ACCESS.2019.2917952.

[41] F. Xu et al., “Breast Anatomy Enriched Tumor Saliency Estimation,” pp. 1–5, 2019, [Online]. Available: http://arxiv.org/abs/1910.10652.

[42] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Int. Conf. Med. image Comput. Comput. Interv. O., Fischer, P., Brox, T. (2015, October). Springer, Cham., vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[43] J. Araujo, André and Norris, Wade and Sim, “Computing Receptive Fields of Convolutional Neural Networks,” Distill, vol. 4, no. 11, p. e21, 2019, doi: 10.23915/distill.00021.

[44] W. Luo, Y. Li, R. Urtasun, and R. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” Adv. Neural Inf. Process. Syst., no. Nips, pp. 4898–4906, 2016.

[45] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc., 2016.

[46] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, B. Zhang, and J. Ding, “Automatic breast ultrasound image segmentation: A survey,” Pattern Recognit., vol. 79, pp. 340–355, 2018, doi: 10.1016/j.patcog.2018.02.012.

[47] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, “Automated breast cancer detection and classification using ultrasound images: A survey,” Pattern Recognit., vol. 43, no. 1, pp. 299–317, 2010, doi: 10.1016/j.patcog.2009.05.012.

[48] J. Jam, C. Kendrick, V. Drouard, K. Walker, G.-S. Hsu, and M. H. Yap, “Symmetric Skip Connection Wasserstein GAN for High-Resolution Facial Image Inpainting,” vol. 1, no. c, 2020, [Online]. Available: http://arxiv.org/abs/2001.03725.

[49] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, pp. 3431–3440, doi: 10.1109/TPAMI.2016.2572683.

[50] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[51] Z. Gu et al., “CE-Net: Context Encoder Network for 2D Medical Image Segmentation,” IEEE Trans. Med. Imaging, vol. 38, no. 10, pp. 2281–2292, 2019, doi: 10.1109/tmi.2019.2903562.

[52] N. Ibtehaz and M. S. Rahman, “MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation,” Neural Networks, vol. 121, pp. 74–87, 2020, doi: 10.1016/j.neunet.2019.08.025.

[53] “American Cancer Society.Cancer Facts & Figures,” 2022. http://cancerstatisticscenter.cancer.org.

[54] F. Z. Francies, R. Hull, R. Khanyile, and Z. Dlamini, “Breast cancer in low-middle income countries: abnormality in splicing and lack of targeted treatment options,” Am. J. Cancer Res., vol. 10, no. 5, pp. 1568–1591, 2020.

[55] S. H. Kim, H. H. Kim, and W. K. Moon, “Automated Breast Ultrasound Screening for Dense Breasts,” Korean J. Radiol., vol. 21, no. 1, pp. 15–24, 2020, doi: 10.3348/kjr.2019.0176.

[56] M. Rebolj, V. Assi, A. Brentnall, D. Parmar, and S. W. Duffy, “Addition of ultrasound to mammography in the case of dense breast tissue: Systematic review and meta-analysis,” Br. J. Cancer, vol. 118, no. 12, pp. 1559–1570, 2018, doi: 10.1038/s41416-018-0080-3.

[57] C. Byrne et al., “Mammographic features and breast cancer risk: Effects with time, age, and menopause status,” J. Natl. Cancer Inst., vol. 87, no. 21, pp. 1622–1629, 1995, doi: 10.1093/jnci/87.21.1622.

[58] G. Ursin et al., “Mammographic density and breast cancer in three ethnic groups,” Cancer Epidemiol. Biomarkers Prev., vol. 12, no. 4, pp. 332–338, 2003.

[59] R. M. Tamimi, C. Byrne, G. A. Colditz, and S. E. Hankinson, “Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women,” J. Natl. Cancer Inst., vol. 99, no. 15, pp. 1178–1187, 2007, doi: 10.1093/jnci/djm062.

[60] M. N. Linver, “4-19 Mammographic Density and the Risk and Detection of Breast Cancer,” Breast Dis., vol. 18, no. 4, pp. 364–365, 2008, doi: 10.1016/S1043-321X(07)80400-0.

[61] L. Yaghjyan et al., “Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics,” J. Natl. Cancer Inst., vol. 103, no. 15, pp. 1179–1189, 2011, doi: 10.1093/jnci/djr225.

[62] G. S. Lodwick, T. E. Keats, and J. P. Dorst, “The coding of roentgen images for computer analysis as applied to lung cancer.,” Radiology, vol. 81(2), pp. 185–200, 1963.

[63] Q. Huang, F. Zhang, and X. Li, “Machine Learning in Ultrasound Computer-Aided Diagnostic Systems: A Survey,” Biomed Res. Int., vol. 2018, 2018, doi: 10.1155/2018/5137904.

[64] W. Al-Dhabyani, A. Fahmy, M. Gomaa, and H. Khaled, “Deep learning approaches for data augmentation and classification of breast masses using ultrasound images,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, pp. 618–627, 2019, doi: 10.14569/ijacsa.2019.0100579.

[65] M. Byra et al., “Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion,” Med. Phys., vol. 46, no. 2, pp. 746–755, 2019, doi: 10.1002/mp.13361.

[66] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[67] Y. Liang, R. He, Y. Li, and Z. Wang, “Simultaneous segmentation and classification of breast lesions from ultrasound images using Mask R-CNN,” IEEE Int. Ultrason. Symp. IUS, vol. 2019-Octob, pp. 1470–1472, 2019, doi: 10.1109/ULTSYM.2019.8926185.

[68] T.-Y. Lin et al., “Microsoft COCO: Common Objects in Context,” in Computer Vision -- ECCV 2014, 2014, pp. 740–755.

[69] A. Hijab, M. A. Rushdi, M. M. Gomaa, and A. Eldeib, “Breast Cancer Classification in Ultrasound Images using Transfer Learning,” Int. Conf. Adv. Biomed. Eng. ICABME, vol. 2019-Octob, pp. 1–4, 2019, doi: 10.1109/ICABME47164.2019.8940291.

[70] Z. Cao, L. Duan, G. Yang, T. Yue, and Q. Chen, “An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures,” BMC Med. Imaging, vol. 19, no. 1, pp. 1–9, 2019, doi: 10.1186/s12880-019-0349-x.

[71] W. C. Shia and D. R. Chen, “Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine,” Comput. Med. Imaging Graph., vol. 87, no. October 2020, p. 101829, 2021, doi: 10.1016/j.compmedimag.2020.101829.

[72] H. Zhang, L. Han, K. Chen, Y. Peng, and J. Lin, “Diagnostic Efficiency of the Breast Ultrasound Computer-Aided Prediction Model Based on Convolutional Neural Network in Breast Cancer,” J. Digit. Imaging, vol. 33, no. 5, pp. 1218–1223, 2020, doi: 10.1007/s10278-020-00357-7.

[73] J. Xie et al., “A novel approach with dual-sampling convolutional neural network for ultrasound image classification of breast tumors,” Phys. Med. Biol., vol. 65, no. 24, 2020, doi: 10.1088/1361-6560/abc5c7.

[74] S. Han et al., “A deep learning framework for supporting the classification of breast lesions in ultrasound images,” Phys. Med. Biol., vol. 62, no. 19, pp. 7714–7728, 2017, doi: 10.1088/1361-6560/aa82ec.

[75] J. Xing et al., “Using BI-RADS Stratifications as Auxiliary Information for Breast Masses Classification in Ultrasound Images,” IEEE J. Biomed. Heal. Informatics, vol. XX, no. XX, pp. 1–1, 2020, doi: 10.1109/jbhi.2020.3034804.

[76] Z. Zhuang, Z. Yang, S. Zhuang, A. N. J. Raj, Y. Yuan, and R. Nersisson, “Multi-Features-Based Automated Breast Tumor Diagnosis Using Ultrasound Image and Support Vector Machine,” Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/9980326.

[77] W. X. Liao et al., “Automatic Identification of Breast Ultrasound Image Based on Supervised Block-Based Region Segmentation Algorithm and Features Combination Migration Deep Learning Model,” IEEE J. Biomed. Heal. Informatics, vol. 24, no. 4, pp. 984–993, 2020, doi: 10.1109/JBHI.2019.2960821.

[78] X. Fei et al., “Doubly supervised parameter transfer classifier for diagnosis of breast cancer with imbalanced ultrasound imaging modalities,” Pattern Recognit., vol. 120, p. 108139, Dec. 2021, doi: 10.1016/J.PATCOG.2021.108139.

[79] Z. Zhuang, Z. Yang, A. N. J. Raj, C. Wei, P. Jin, and S. Zhuang, “Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion,” Comput. Methods Programs Biomed., p. 106221, Jun. 2021, doi: 10.1016/j.cmpb.2021.106221.

[80] A. Vakanski and M. Xian, “Evaluation of Complexity Measures for Deep Learning Generalization in Medical Image Analysis,” 2021 IEEE 31st Int. Work. Mach. Learn. Signal Process., pp. 1–15.

[81] G. Zhang, K. Zhao, Y. Hong, X. Qiu, K. Zhang, and B. Wei, “SHA-MTL : soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification,” Int. J. Comput. Assist. Radiol. Surg., 2021, doi: 10.1007/s11548-021-02445-7.

[82] J. Shi, A. Vakanski, M. Xian, J. Ding, and C. Ning, “EMT-NET: Efficient multitask network for computer-aided diagnosis of breast cancer,” in 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5.

[83] T. Geertsma, “Ultrasoundcases.info,” FujiFilm, 2014. https://www.ultrasoundcases.info/.

[84] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” Data Br., vol. 28, p. 104863, 2020, doi: 10.1016/j.dib.2019.104863.

[85] A. Rodtook, K. Kirimasthong, W. Lohitvisate, and S. S. Makhanov, “Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities,” Pattern Recognit., vol. 79, pp. 172–182, 2018, doi: 10.1016/j.patcog.2018.01.032.

[86] A. G. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[87] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” 36th Int. Conf. Mach. Learn. ICML 2019, vol. 2019-June, pp. 10691–10700, 2019.

[88] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, vol. 2017-Janua, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[89] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[90] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.

[91] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1800–1807, 2017, doi: 10.1109/CVPR.2017.195.

[92] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[93] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318–327, 2018, doi: 10.1109/TPAMI.2018.2858826.

[94] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980, 2014, [Online]. Available: http://arxiv.org/abs/1412.6980.

[95] T. Dozat, “Incorporating Nesterov Momentum into Adam,” ICLR Work., no. 1, pp. 2013–2016, 2016.

[96] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, “Comparison of Transferred Deep Neural Networks in Ultrasonic Breast Masses Discrimination,” Biomed Res. Int., vol. 2018, 2018, doi: 10.1155/2018/4605191.

[97] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0276-2.

[98] H. Wang, M. Xian, A. Vakanski, and B. Shareef, “SIAN: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis,” 2022, [Online]. Available: http://arxiv.org/abs/2209.02412.

[99] B. Zhang, A. Vakanski, and M. Xian, “BI-RADS-Net: An Explainable Multitask Learning Approach for Cancer Diagnosis in Breast Ultrasound Images,” 2021.

[100] A. Singh, S. Sengupta, and V. Lakshminarayanan, “Explainable deep learning models in medical image analysis,” J. Imaging, vol. 6, no. 6, pp. 1–19, 2020, doi: 10.3390/JIMAGING6060052.

[101] Z. Zhuang, Z. Yang, A. N. J. Raj, C. Wei, P. Jin, and S. Zhuang, “Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion,” Comput. Methods Programs Biomed., vol. 208, p. 106221, 2021, doi: 10.1016/j.cmpb.2021.106221.

[102] B. Shareef et al., “A Benchmark for Breast Ultrasound Image Classification,” Available SSRN https//ssrn.com/abstract=4339660 or http//dx.doi.org/10.2139/ssrn.4339660, 2023.

[103] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” 2020, [Online]. Available: http://arxiv.org/abs/2010.11929.

[104] B. Gheflati and H. Rivaz, “Vision Transformers for Classification of Breast Ultrasound Images,” Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf., vol. 2022, pp. 480–483, Jul. 2022, doi: 10.1109/EMBC48229.2022.9871809.

[105] G. Ayana and S. Choe, “BUViTNet: Breast Ultrasound Detection via Vision Transformers,” Diagnostics, vol. 12, no. 11, p. 2654, Nov. 2022, doi: 10.3390/diagnostics12112654.

[106] W. Wang, R. Jiang, N. Cui, Q. Li, F. Yuan, and Z. Xiao, “Semi-supervised vision transformer with adaptive token sampling for breast cancer classification,” Front. Pharmacol., vol. 13, Jul. 2022, doi: 10.3389/fphar.2022.929755.

[107] M. A. Hassanien, V. K. Singh, D. Puig, and M. Abdel-Nasser, “Predicting Breast Tumor Malignancy Using Deep ConvNeXt Radiomics and Quality-Based Score Pooling in Ultrasound Sequences,” Diagnostics, vol. 12, no. 5, 2022, doi: 10.3390/diagnostics12051053.

[108] Y. Mo et al., “HoVer-Trans: Anatomy-aware HoVer-Transformer for ROI-free Breast Cancer Diagnosis in Ultrasound Images,” IEEE Trans. Med. Imaging, vol. 42, no. 6, pp. 1–1, 2023, doi: 10.1109/tmi.2023.3236011.

[109] X. Qu et al., “A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images,” Med. Phys., Sep. 2022, doi: 10.1002/mp.15852.

[110] A. Iqbal and M. Sharif, “BTS-ST: Swin transformer network for segmentation and classification of multimodality breast cancer images,” Knowledge-Based Syst., p. 110393, Feb. 2023, doi: 10.1016/j.knosys.2023.110393.

[111] S. Tang et al., “Transformer-based multi-task learning for classification and segmentation of gastrointestinal tract endoscopic images,” Comput. Biol. Med., vol. 157, no. February, p. 106723, 2023, doi: 10.1016/j.compbiomed.2023.106723.

[112] S. Ruder, “An Overview of Multi-Task Learning in Deep Neural Networks,” no. May, 2017, [Online]. Available: http://arxiv.org/abs/1706.05098.

[113] T. Geertsma and Fujifilm, “Ultrasound cases,” 2014. https://www.ultrasoundcases.info/.

[114] J. Chowdary, P. Yogarajah, P. Chaurasia, and V. Guruviah, “A Multi-Task Learning Framework for Automated Segmentation and Classification of Breast Tumors From Ultrasound Images,” Ultrason. Imaging, vol. 44, no. 1, pp. 3–12, Jan. 2022, doi: 10.1177/01617346221075769.

[115] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14030.

[116] M. H. Yap et al., “Breast ultrasound lesions recognition: end-to-end deep learning approaches,” J. Med. Imaging, vol. 6, no. 01, p. 1, Oct. 2018, doi: 10.1117/1.JMI.6.1.011007.

[117] H. Tanaka, S. W. Chiu, T. Watanabe, S. Kaoku, and T. Yamaguchi, “Computer-aided diagnosis system for breast ultrasound images using deep learning,” Phys. Med. Biol., vol. 64, no. 23, 2019, doi: 10.1088/1361-6560/ab5093.

[118] W. K. Moon, Y. W. Lee, H. H. Ke, S. H. Lee, C. S. Huang, and R. F. Chang, “Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks,” Comput. Methods Programs Biomed., vol. 190, 2020, doi: 10.1016/j.cmpb.2020.105361.