Breast cancer segmentation using UNet and Global Convolutional Networks

Anand Thyagachandran[1] and Yeruru Asrar Ahmed[1]

Department of Computer Science and Engineering,
Indian Institute of Technology Madras, Chennai, India
{tanand,asrar}@cse.iitm.ac.in

Abstract. Breast ultrasound (BUS) imaging techniques have become efficient tools for cancer diagnosis. Convolutional neural network (CNN) based encoder-decoder architectures have been widely used for the automated segmentation of tumours in BUS images, assisting in breast cancer diagnoses. However, these models have limitations in capturing long-range dependencies. To overcome this limitation, various deep learning techniques, such as atrous convolution, attention mechanisms, and transformer encoder-based models, have been introduced to capture long-range dependencies in feature maps, improving segmentation accuracy by considering larger receptive fields and global context. As modelling techniques evolve, there is a shift towards more complex and intricate designs. This study proposes a simple yet effective model that combines UNet and Global Convolutional Network (GCN) architectures for breast lesion segmentation. By leveraging the GCN block, our model captures broader receptive fields with a simpler design strategy. We have demonstrated the efficacy of our approach through various experiments, including kernel size analysis, model component evaluation, and data preprocessing assessment. The proposed model has been evaluated using four-fold cross-validation with BUSI and Dataset-B datasets. Additionally, models trained on both datasets have been validated with a blind test dataset, where our model demonstrates better performance compared to state-of-the-art methods, achieving a 4.9% and 6.7% improvement in Intersection over Union (IoU) score, respectively. The robustness analysis and external validation experiments underscore the superior generalization performance of our model in breast lesion segmentation tasks.

Keywords: Breast cancer · segmentation · ultrasound imaging

1 Introduction

Breast cancer is a significant global health challenge, especially among women, and it has high mortality rates [1]. Early detection of symptoms is crucial for effective treatment. Breast ultrasound (BUS) imaging techniques have emerged as efficient tools for cancer diagnosis. They are cost-effective, non-invasive, provide real-time results, and do not involve ionizing radiation [2]. Breast cancer is classified into benign and malignant breast lesions, where benign breast lesions

pose no threat to health, while malignant breast lesions are cancerous growths that can spread throughout the body. The diagnosis of breast cancer involves detection, segmentation, and classification stages. This study mainly focuses on segmenting the lesion regions from BUS images to aid in the diagnosis of breast cancer.

The medical field has seen significant advancements in the automation of medical image segmentation, providing valuable assistance to clinicians in quantitative pathological assessment and diagnosis. Segmentation methods can broadly divided into semi-automated and fully automated procedures based on the manual intervention to fine-tune the breast lesion regions in BUS images [3,4]. A comprehensive review of earlier methods is available in Xian *et al.* [5]. Fully automatic methods, exemplified by recent works such as Chen *et al.* [6] and Yan *et al.* [7], eliminate the need for user intervention. These methods primarily leverage convolutional operations at each layer to extract local image features from neighbouring pixels, enabling these models to predict the semantics of objects in medical images. Among them, fully convolutional encoder-decoder-based models like UNet [8] are extensively utilized. The UNet architecture comprises encoder and decoder layers to extract features for predicting maps at image resolution. Additionally, it uses skip connections between the encoder and decoder layers to preserve the spatial structure, enabling precise object localization. The effectiveness of the UNet architecture is evident in its state-of-the-art performances, particularly with small medical image datasets, owing to its compact parameterization and encoder-decoder design. Variants of the UNet model are widely adopted in biomedical image segmentation tasks [9,10,11,12,13].

Almajalid *et al.* introduced UNet for breast lesion segmentation from BUS images [14]. Later, various UNet variants were proposed to refine segmentation accuracy. These variants can be broadly categorized into four classes: multiscale UNet [15,6,16,17], attention-based UNet [18,19,7], deep supervised UNet [20,21,22], and multi-module hybrid UNet [23,24]. Multiscale UNet models utilize diverse convolutional kernel sizes to capture context information across different receptive fields. Attention-based UNet models aim to capture global context information for improved segmentation, including hybrid dilated convolution-based attention UNet [7] and channel attention module [19]. Deep supervised UNet ensures each stage contributes to the loss function, enforcing feature learning proximity to the ground truth at every stage of the model. Multi-module hybrid UNet architectures integrate disparate, independent modules—such as Tversky loss functions [20], residual inception depth-wise separable convolutions, and hybrid pooling strategies (combining max pooling and spectral pooling)—alongside cross-spatial attention filters [23] to further refine segmentation predictions.

The convolution based encoder-decoder models encounter challenges associated with capturing long-range dependencies between pixels in the feature maps. The convolutional operation often leads to inductive biases [25], limiting architectures' ability to model long-range feature dependencies effectively. Two strategies are commonly employed to address the limitations, such as enlarging the recep-

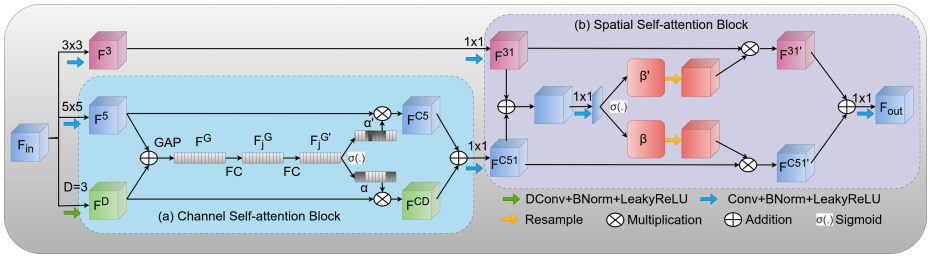

Fig. 1. The HAAM block architecture

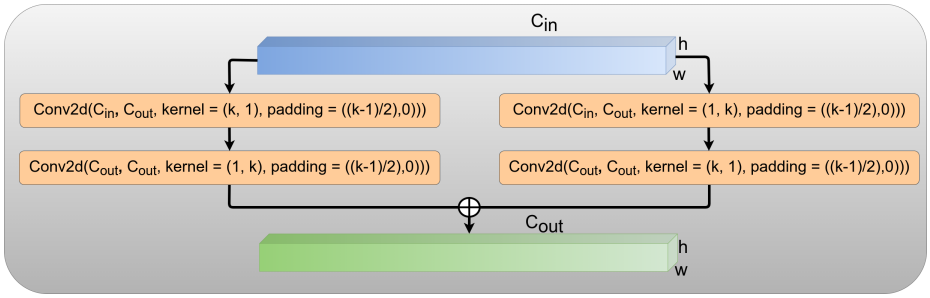

Fig. 2. The GCN block architecture

tive field [26,27,28,29,30,31] and incorporating attention mechanisms [32,33,34]. Atrous convolution operations [25] are utilized to insert holes into convolution kernels, preserving resolution and enlarging the receptive field. However, relying solely on atrous convolution operations may not fully address challenges posed by surrounding tissues and indistinct boundaries [15]. Attention mechanisms have also been integrated to exploit long-range dependencies in CNNs [20,19,11,34]. These mechanisms enhance models' capacity to capture intricate details and disregard irrelevant features by dynamically focusing on relevant regions within the input image [11]. As modelling choices evolve, there is a shift towards more complex and intricate designs. Several studies have utilized a combination of atrous convolution and attention mechanisms [6,7,24] to enhance tumour segmentation in Breast Ultrasound (BUS) images, categorizing them as hybrid models.

A recent development in hybrid models for breast cancer segmentation is the Adaptive Attention UNet (AAUNet)[6]. In this model, Chen *et al.* replaced the conventional convolution layers in the UNet encoder and decoder blocks with a Hybrid Adaptive Attention Module (HAAM) [6]. The intricate design of the AAUNet model has demonstrated superior performance compared to state-of-the-art semantic segmentation models with BUS images. The HAAM block, a key component of AAUNet, effectively captures a larger receptive field by employing multiple convolutions with varying receptive fields. The channel and spatial attention modules use these output features to enhance the segmentation accuracy. The HAAM block diagram is shown in Fig. 1. In contrast, our pro-

posed approach presents a simpler architecture, leveraging the UNet framework augmented with a Global Convolutional Network (GCN) [35]. The GCN block primarily employs separable large filters to capture extensive receptive fields while minimizing the number of parameters compared to standard convolutions. The GCN has a symmetric structure to capture broader and better receptive fields and is shown in Fig. 2. By integrating large receptive field information, the GCN blocks allow superior segmentation predictions [35]. The combination of UNet, Global Convolution Network (GCN), and Boundary Refinement (BR) Module has been previously proposed for segmenting tongue medical images, as demonstrated by [36]. GCN and BR blocks are employed to reduce the gap between localization and classification. Moreover, the GCN block is incorporated randomly in the UNet model. In contrast, the purpose of the GCN block in our model is to increase the receptive field.

The remainder of this paper is organised as follows: Section 2 discusses the proposed segmentation model for BUS images. Section 3 outlines the datasets and evaluation metrics utilized in the study. Experimental results and inferences are presented in Section 4. Finally, Section 5 outlines the future scope and conclusions drawn from the work.

2 Methodology

Our proposed model presents a UNet-based architecture including a Global Convolutional Network (GCN) block. Utilizing the UNet as its foundation, the proposed architecture comprises five encoder layers, five decoder layers, and skip connections between encoder and decoder for preserving spatial structure. GCN block is used to expand its receptive field, empowering it to extract contextual information effectively. The GCN block integrates into the skip connection structure through empirical analysis, enhancing its capacity to capture spatial relationships and semantic context in medical imaging data. This simple approach utilises the strengths of GCN to improve segmentation performance, particularly in tasks requiring capturing spatial and semantic relations. The proposed model architecture is shown in Fig. 3.

2.1 UNet

The UNet, introduced by Ronneberger *et al.* in 2015 [8], is an encoder-decoder architecture based on fully convolutional neural networks. In the UNet, the encoder layer initially captures high-frequency features and gradually refines them for semantic extraction across subsequent encoder layers. Multiple encoder layers with max-pooling downsample the image to low-resolution feature maps, which are then passed to a bottleneck layer. These features are upsampled using decoder layers during the decoding process, and features from corresponding skip connections are incorporated. Skip connections help preserve the spatial structure, while the upsampled features from decoder layers capture more semantic

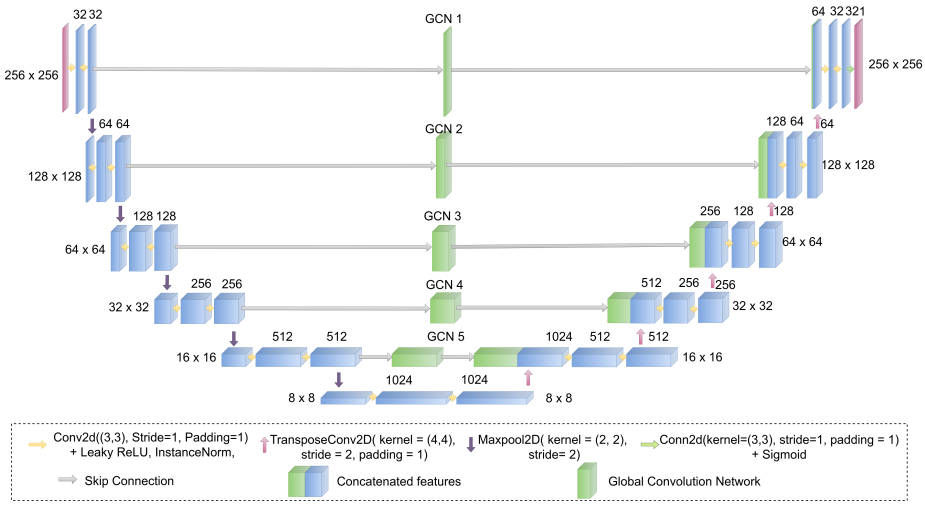

Fig. 3. The proposed UNet-GCN architecture

information, facilitating precise identification of regions of interest within medical images. Each encoder and decoder layer consists of two convolution layers (3x3 kernel), followed by Instance Normalization [37] and LeakyReLU activation functions.

2.2 Global Convolutional Network (GCN)

Medical image segmentation models typically use 3×3 kernels in convolutional layers [8,13,14], primarily capturing local information and limiting larger receptive fields in initial layers. While atrous convolution addresses this limitation by employing dilated convolutions [6], it often provides only large, sparse receptive fields [12]. A simpler approach would be using larger kernels in convolution, which increases the receptive field and aids in handling significant variations in lesion transformations in BUS images [15]. However, using large kernels increases exponentially the number of parameters and GPU memory usage in each convolution layer. GCN block [35] captures a larger receptive field with linear growth in parameters and can easily be incorporated into existing architectures.

GCN approximates $k \times k$ convolutions using four low-rank convolutions in two parallel branches. Each branch consists of two low-rank convolution kernel sizes of $k \times 1$ followed by $1 \times k$ and vice versa. The dual branch gives equal precedence to the horizontal and vertical kernels to capture the information. Traditional $k \times k$ convolutions require k^2 parameters, whereas parameters increase linearly in GCN and need only $4k$ parameters. GCN replicates dense connections to the input feature map within the receptive field of $k \times k$, which helps handle large variations of transformations. GCN approach helps to increase the receptive field in the early stages while reducing overall parameter growth in the model. Self-

Table 1. Different image preprocessing methods employed in the Data augmentation process.

Image Preprocessing	Description
Identity	Returns the original image.
Gaussian blur	Blur the image with a Gaussian kernel.
Equalize	Histogram Equalization
Contrast	Adjusts the contrast of the image by [0.05, 0.95].
Brightness	Adjusts the brightness of the image by [0.05, 0.95]
Random Flip	Horizontally flips the image with a probability of 0.5.

attention cannot be used in early layers to capture larger contexts due to the exponential increase in GPU memory, a challenge mitigated by GCN blocks. The optimal position of GCN within the backbone network and the kernel sizes of the GCN block are determined empirically, as detailed in Section 4.1, and Section 4.2. The block diagram of GCN is shown in Fig. 2.

2.3 Preprocessing BUS images

Ultrasound images often suffer from low signal-to-noise ratio (SNR) and various artefacts like speckle noise, reverberations, and acoustic shadowing, which degrade image quality [38]. Image preprocessing techniques are commonly employed to enhance BUS images and their quality [38]. Image preprocessing methods like contrast enhancement, brightness adjustment, Gaussian blurring and histogram equalization are employed in our work. Gaussian blurring removes the high-frequency noise and preserves the structure and edges in the image. Histogram equalization [39] redistributes the intensity values across the histogram to enhance the quality of the image. This process effectively stretches the intensity levels, making the image appear more visually appealing with improved contrast and detail. These preprocessing methods are integrated into the data augmentation, as detailed in Table 1. Specifically, we employ six image transformations, with three randomly selected transformations applied to each image in the batch during training. Such augmentation strategies have been demonstrated to enhance model performance in semantic segmentation tasks significantly [40].

2.4 Model settings

All the images are reshaped into 256x256 pixels before being input into the proposed model. During model training, a stratified batch (equal number of images from the breast lesion class and normal class) is used to avoid bias in the class imbalance dataset [41]. Binary cross entropy [42] is employed as the loss function.

$$\mathcal{L}_{BCE} = -\sum_{(i,j)} GT(i,j) * \log(PD(i,j)) + (1 - GT(i,j)) * (1 - \log(PD(i,j)))$$

where $GT(i,j) \in [0,1]$ denotes the ground-truth mask (i,j), $PD(i,j) \in [0,1]$ represents the predict masks. Adam optimizer with a learning rate of 0.0001 is used for model optimization. The model is trained for 100 epochs with a batch size of 16. The parameters of the model are optimized on a validation set. Balanced data from both normal and lesion classes are used for training to maintain fairness. The code is written in PyTorch [43], and all experiments are conducted using two GeForce GTX 1080 Titans with an overall 24 GB GPU memory.

3 Datasets and Evaluation metrics

3.1 BUS Dataset

The BUSI dataset, collected at Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt, in 2018 [44], consists of 780 BUS images obtained from 600 patients aged 25 to 75 years. The dataset encompasses three distinct classes of BUS images: benign (487 images), malignant (210 images), and normal (133 images). Imaging data was captured using the LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound systems. Following the acquisition, skilled radiologists preprocessed the images to delineate lesion regions and eliminate extraneous areas. Subsequently, the images were converted to PNG format for standardized analysis.

Dataset-B [45] consists of 163 images, including 110 benign images and 53 malignant images. This dataset was captured using the Siemens ACUSON Sequoia C512 system at the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell, Spain. Additionally, the STU dataset [24] contains 42 BUS images and corresponding masks. These images were acquired using the GE Voluson E10 Ultrasound Diagnostic System at Shantou First Affiliated Hospital, Guangdong Province, China. While all images in the STU dataset depict lesions, they are not explicitly classified as benign or malignant. The STU dataset is an external validation (test) dataset for evaluating model performance.

3.2 Evaluation Metrics

Image segmentation evaluation metrics are helpful in assessing the effectiveness of segmentation models. Five widely recognized metrics are used in our work: Intersection over Union (IoU), Dice similarity coefficient (Dice), Precision (Prec.), Sensitivity (Sen.), and Specificity (Spec.). IoU is also known as the Jaccard index, which estimates the ratio of the intersection area between the prediction and ground truth mask. The dice score is also referred to as the F1 score, which estimates the ratio of twice the overlap between the prediction with ground truth mask to the sum of their areas. IoU and DSC evaluate the spatial correspondence between the predicted and ground truth masks, with higher values indicating superior segmentation accuracy. Precision estimates the proportion of correctly classified lesion pixels to the total number of lesion pixels predicted

in the prediction mask, while Sensitivity measures the proportion of correctly classified lesion pixels in the prediction mask to the ground truth mask. Moreover, Specificity assesses the proportion of correctly classified background pixels in the prediction mask to the ground truth mask.

4 Experimental Settings and Results

This section presents a series of experiments to evaluate the performance of the proposed and baseline models in breast cancer segmentation using BUS images. An ablation study is conducted to understand the importance of GCN components within the model architecture. We also investigate kernel size's impact on breast lesion segmentation in the GCN block. Another ablation study evaluates the effect of data preprocessing techniques on BUS medical image segmentation. We then present and discuss the segmentation results obtained with state-of-the-art models using the BUSI and Dataset-B. All experiments are conducted using four-fold cross-validation on the sorted dataset and employ internal shuffling for uniformity. Finally, we assess our proposed and baseline models' generalizability using the unseen (Test) STU dataset. The STU dataset consists of two classes, tumour and normal, and the trained models predict whether each pixel in the BUS image is normal or a tumour.

4.1 GCN Position

The GCN is an independent block used to capture larger receptive fields and can be easily integrated into the UNet architecture. We explore three variants: Model A, where the GCN block is within the skip connection; Model B, where it's placed between each encoder and decoder block; and Model C, where it replaces each convolution in both encoder and decoder blocks (except for the upsampling convolution). Table 2 shows the performance of these variants, with skip connections proving to be the optimal choice in terms of performance and is employed for further analysis.

Table 2. Segmentation results for GCN at different positions in the proposed network with BUSI dataset. Models A, B, and C are defined in Section 4.1.

Models	BUSI				
	IoU	**Dice**	**Sensitivity**	**Precision**	**Specificity**
Model A	61.05 ±1.31	75.69 ±1.00	72.13 ±1.35	78.29 ±2.98	98.28 ±0.31
Model B	60.28 ±1.55	75.07 ±1.21	72.12 ±1.39	78.41 ±3.39	98.26 ±0.40
Model C	59.96 ±1.07	74.79 ±0.83	70.98 ±2.97	79.06 ±3.47	97.36 ±0.44

4.2 Kernel Size

Using larger kernels enables the model to have larger receptive fields, enhancing its ability to predict lesions effectively. In an ablation study, we tested three different kernel sizes (3, 5, and 7) for the k parameter in the GCN, ensuring uniformity across all GCN kernels. We opt for a maximum kernel size of 7, restricted by the smallest feature size of 8×8 within the proposed network. It is observed in Table 3 that the larger kernel is progressively improving the model's segmentation prediction.

Table 3. Segmentation results for different kernels used in the GCN block with BUSI dataset.

Kernel Size (k)	**BUSI**				
	IoU	**Dice**	**Sensitivity**	**Precision**	**Specificity**
7	61.05 ±1.31	75.69 ±1.00	72.13 ±1.35	78.29 ±2.98	98.28 ±0.31
5	60.51 ±0.81	75.26 ±0.66	72.04 ±3.09	79.08 ±3.65	98.29 ±0.49
3	59.44 ±1.10	74.72 ±0.84	72.37 ±2.64	76.69 ±1.53	98.06 ±0.23

4.3 Data augmentation

BUS images are characterised by noise and low quality, often exhibiting low contrast. We apply domain knowledge-based data augmentation methods to address these issues and enhance image perception to improve contrast and reduce noise. Rather than adding domain-based augmentation directly, random augmentation settings are used for a superior augmentation approach [40]. To verify the claim that such a data augmentation method improves the perception quality and aids in the prediction of maps by the network, we perform an ablation study involving data augmentation techniques. We train our model with and without augmentation approaches and report its results in Table 4. The model trained with data augmentation achieves superior Dice and IoU scores compared to methods that do not utilise augmentation.

Table 4. Segmentation results of the proposed model with and without data augmentation using BUSI dataset.

Data Augmentation	**BUSI**				
	IoU	**Dice**	**Sensitivity**	**Precision**	**Specificity**
✓	61.05 ±1.31	75.69 ±1.00	72.13 ±1.35	78.29 ±2.98	98.28 ±0.31
✗	59.31 ±0.59	74.28 ±0.50	70.59 ±1.29	78.53 ±2.24	98.29 ±0.22

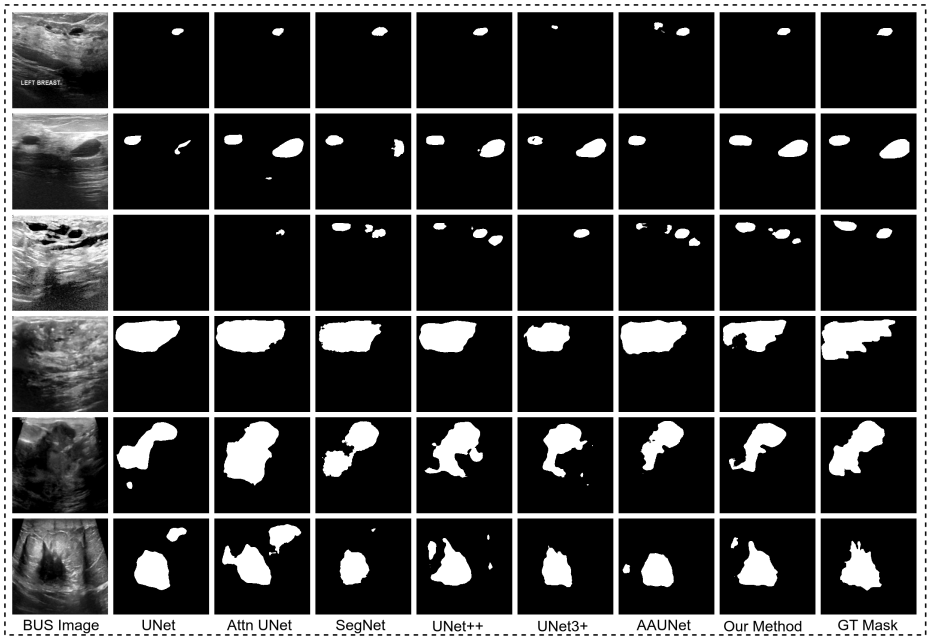

Fig. 4. The segmentation results of different methods on breast ultrasound images. The first column represents the input image. The remaining columns represent the corresponding mask predicted by the models.

Table 5. The cross-fold validation segmentation results for the baseline and proposed model. Sen, Prec, and Spec represent sensitivity, precision and specificity, respectively. NoP represents the number of parameters of the model, and the values are expressed in millions. FLOPS represents the Floating-point operations per second, and values are denoted in gigabits per second (Gbps).

Models	BUSI					Dataset-B					NoP	FLOPS
	IoU	Dice	Sen.	Prec.	Spec.	IoU	Dice	Sen.	Prec.	Spec.		
UNet	53.82 ±2.59	69.75 ±2.10	65.96 ±4.67	74.07 ±4.49	97.98 ±0.57	60.83 ±3.29	75.62 ±2.62	68.32 ±3.35	84.70 ±5.85	99.37 ±0.35	39	27.78
Attention UNet	57.08 ±1.12	72.57 ±0.90	70.89 ±3.15	74.65 ±3.99	97.87 ±0.57	69.98 ±2.68	82.29 ±1.82	78.48 ±5.03	86.96 ±5.75	99.41 ±0.28	34	66.69
UNet ++	57.14 ±0.88	72.60 ±2.21	69.15 ±2.82	76.53 ±3.59	98.16 ±0.44	68.14 ±2.17	80.99 ±1.61	80.35 ±5.75	82.25 ±5.86	99.14 ±0.37	47	199.85
UNet 3+	56.93 ±0.95	72.43 ±0.79	68.45 ±2.09	77.27 ±2.15	98.20 ±0.29	69.86 ±2.29	82.23 ±1.53	78.02 ±4.03	87.06 ±4.01	99.39 ±0.17	26	198.03
SegNet	57.55 ±1.44	72.93 ±1.18	68.04 ±2.85	78.73 ±2.47	98.34 ±0.30	68.38 ±2.54	81.20 ±1.88	77.75 ±2.97	85.23 ±3.72	99.32 ±0.20	29	40.82
AAUNet	59.90 ±2.24	74.72 ±1.76	69.62 ±2.99	80.68 ±5.59	98.52 ±0.62	70.02 ±2.83	82.34 ±1.93	78.42 ±3.76	86.70 ±4.15	99.40 ±0.26	43	85.33
Proposed Method	61.05 ±1.31	75.69 ±1.00	72.13 ±1.35	78.29 ±2.98	98.28 ±0.31	72.11 ±1.92	83.77 ±1.29	83.96 ±3.46	83.72 ±2.82	99.02 ±0.18	80	37.08

4.4 Comparison with state-of-the-art models

We have compared approaches like AAUNet, designed explicitly for breast lesion segmentation, with our proposed model performance. The other state-of-the-art medical segmentation methods include UNet [8], SegNet [46], Attention UNet [11], UNet++ [13], and UNet3+ [9] are also assessed. We have used the officially available repositories of these models to reproduce the results on the BUSI and other datasets. All models are performed four-fold cross-validation with data augmentation, and the results are shown in Table. 5. Models were trained separately with four-fold cross-validation with datasets BUSI and Dataset-B. Our proposed model performs better than other state-of-the-art models regarding IoU, Dice score, and sensitivity with BUSI and Dataset-B. Though the number of parameters is large, the number of Floating-point operations per second (FLOPS) is lower, suggesting that our approach is simpler and requires no intricate design. Visual outputs of the proposed model and other state-of-the-art models are shown in Fig. 4. Our approach captures the better spatial structure of the breast lesions when compared to other state-of-the-art models.

Table 6. The External validation segmentation results for the STU dataset with baseline and proposed model trained with BUSI and Dataset-B. Sen, Prec, and Spec represent sensitivity, precision and specificity, respectively.

Models	BUSI					Dataset-B				
	IoU	**Dice**	**Sen.**	**Prec.**	**Spec.**	**IoU**	**Dice**	**Sen.**	**Prec.**	**Spec.**
UNet	69.77 ±2.21	82.18 ±1.54	79.11 ±4.31	85.72 ±2.82	98.16 ±0.50	57.82 ±6.78	73.09 ±5.55	59.76 ±7.54	94.92 ±1.04	99.54 ±0.14
Attention UNet	73.36 ±3.69	84.60 ±2.49	85.83 ±0.65	83.48 ±4.55	97.62 ±0.79	68.43 ±3.35	81.23 ±2.32	72.71 ±2.10	88.67 ±3.48	99.12 ±0.40
UNet ++	74.02 ±2.33	80.05 ±10.90	83.18 ±2.47	87.04 ±1.16	98.29 ±0.17	68.11 ±2.77	81.01 ±1.95	73.83 ±4.41	90.08 ±3.95	98.80 ±0.57
UNet 3+	71.63 ±1.23	83.46 ±0.84	81.65 ±2.15	85.42 ±1.62	98.07 ±0.28	66.38 ±1.90	79.78 ±1.39	71.11 ±1.49	92.25 ±1.64	99.17 ±0.21
SegNet	75.13 ±0.62	85.80 ±0.41	85.26 ±2.22	86.45 ±2.40	98.14 ±0.44	68.30 ±2.63	78.99 ±1.92	69.30 ±3.37	91.97 ±1.90	99.15 ±0.23
AAUNet	75.41 ±3.22	85.96 ±2.11	84.72 ±2.52	87.29 ±3.08	98.28 ±0.47	70.06 ±2.66	82.38 ±1.86	74.35 ±3.30	92.45 ±0.59	99.16 ±0.10
Proposed Method	79.08 ±1.28	88.32 ±0.80	87.43 ±0.71	89.23 ±1.12	98.59 ±0.16	74.79 ±0.39	85.58 ±0.25	79.48 ±1.94	92.78 ±2.02	99.13 ±0.28

4.5 External Validation

External validation is a critical step in assessing the generalizability and robustness of segmentation models. In our study, we used the STU dataset as an external validation set. This dataset, acquired by different imaging systems and from different geographical locations compared to BUSI and Dataset-B, serves as an essential benchmark for evaluating our proposed model's performance in

real-world scenarios. Testing our model on this external dataset ensures its effective generalization to unseen data and different acquisition conditions. We trained models using BUSI and Dataset-B separately and tested them with the STU dataset to predict whether each pixel in the BUS image is normal or a tumour. Our proposed model demonstrates better generalizability to the unseen datasets than other models, with the results shown in Table 6.

5 Conclusions

Our study introduces a novel UNet-based model integrating GCN blocks in skip connections to facilitate breast lesion segmentation in BUS images. Our proposed model demonstrates superior performance compared to existing state-of-the-art methods in this domain. Through several ablation studies, we explain the significance of individual model components, providing insights into their contributions to segmentation accuracy. Moreover, we emphasize the pivotal role of image preprocessing in enhancing segmentation performance for BUS images. Our model showcases robustness across unseen datasets. Looking ahead, we aim to extend our model's capabilities beyond segmentation to encompass comprehensive tasks such as cancer detection, identification, and segmentation within a unified framework.

Acknowledgements

The authors would like to express their sincere gratitude to Prof. Hema A. Murthy, Emeritus Professor, Department of Computer Science and Engineering, IIT Madras, for her invaluable guidance and support throughout the completion of this research. The authors also extend their heartfelt thanks to IITM Pravartak Technologies Foundation, a Technology Innovation Hub of the Indian Institute of Technology, Madras, funded by the Department of Science and Technology, Government of India, under its National Mission on Interdisciplinary Cyber-Physical Systems, for supporting Anand Thyagachandran through a fellowship grant.

References

1. Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D.: Global cancer statistics. CA: a cancer journal for clinicians **61**(2), 69–90 (2011)
2. Chan, V., Perlas, A.: Basics of ultrasound imaging. Atlas of ultrasound-guided procedures in interventional pain management pp. 13–19 (2011)
3. Zhou, Z., Wu, W., Wu, S., Tsui, P.H., Lin, C.C., Zhang, L., Wang, T.: Semi-automatic breast ultrasound image segmentation based on mean shift and graph cuts. Ultrasonic imaging **36**(4), 256–276 (2014)
4. Pons, G., Martí, J., Martí, R., Ganau, S., Noble, J.A.: Breast-lesion segmentation combining b-mode and elastography ultrasound. Ultrasonic imaging **38**(3), 209–224 (2016)

5. Xian, M., Zhang, Y., Cheng, H.D., Xu, F., Zhang, B., Ding, J.: Automatic breast ultrasound image segmentation: A survey. Pattern Recognition **79**, 340–355 (2018)
6. Chen, G., Li, L., Dai, Y., Zhang, J., Yap, M.H.: Aau-net: an adaptive attention u-net for breast lesions segmentation in ultrasound images. IEEE Transactions on Medical Imaging (2022)
7. Yan, Y., Liu, Y., Wu, Y., Zhang, H., Zhang, Y., Meng, L.: Accurate segmentation of breast tumors using ae u-net with hdc model in ultrasound images. Biomedical Signal Processing and Control **72**, 103299 (2022)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
9. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 1055–1059. IEEE (2020)
10. Huang, R., Lin, M., Dou, H., Lin, Z., Ying, Q., Jia, X., Xu, W., Mei, Z., Yang, X., Dong, Y., et al.: Boundary-rendering network for breast lesion segmentation in ultrasound images. Medical image analysis **80**, 102478 (2022)
11. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
12. Zhang, X., Xiao, Z., Fu, H., Hu, Y., Yuan, J., Xu, Y., Higashita, R., Liu, J.: Attention to region: Region-based integration-and-recalibration networks for nuclear cataract classification using as-oct images. Medical Image Analysis **80**, 102499 (2022)
13. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)
14. Almajalid, R., Shan, J., Du, Y., Zhang, M.: Development of a deep-learning-based method for breast ultrasound image segmentation. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 1103–1108. IEEE (2018)
15. Byra, M., Jarosik, P., Szubert, A., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., Andre, M.: Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. Biomedical Signal Processing and Control **61**, 102027 (2020)
16. Chen, G., Li, L., Zhang, J., Dai, Y.: Rethinking the unpretentious u-net for medical ultrasound image segmentation. Pattern Recognition **142**, 109728 (2023)
17. Shareef, B., Xian, M., Vakanski, A.: Stan: Small tumor-aware network for breast ultrasound image segmentation. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). pp. 1–5. IEEE (2020)
18. Chen, G., Liu, Y., Dai, Y., Zhang, J., Cui, L., Yin, X.: Bagnet: bidirectional aware guidance network for malignant breast lesions segmentation. In: 2022 7th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS). pp. 112–116. IEEE (2022)
19. Lee, H., Park, J., Hwang, J.Y.: Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. IEEE transactions on ultrasonics, ferroelectrics, and frequency control **67**(7), 1344–1353 (2020)

20. Abraham, N., Khan, N.M.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). pp. 683–687. IEEE (2019)
21. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7479–7489 (2019)
22. Wang, Y., Wang, N., Xu, M., Yu, J., Qin, C., Luo, X., Yang, X., Wang, T., Li, A., Ni, D.: Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. IEEE transactions on medical imaging **39**(4), 866–876 (2019)
23. Punn, N.S., Agarwal, S.: Rca-iunet: a residual cross-spatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging. Machine Vision and Applications **33**(2), 27 (2022)
24. Zhuang, Z., Li, N., Joseph Raj, A.N., Mahesh, V.G., Qiu, S.: An rdau-net model for lesion segmentation in breast ultrasound images. PloS one **14**(8), e0221535 (2019)
25. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
26. Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., Cheng, L.: Dilated densely connected u-net with uncertainty focus loss for 3d abus mass segmentation. Computer methods and programs in biomedicine **209**, 106313 (2021)
27. Chen, G., Dai, Y., Zhang, J.: C-net: Cascaded convolutional neural network with global guidance and refinement residuals for breast ultrasound images segmentation. Computer Methods and Programs in Biomedicine **225**, 107086 (2022)
28. Chen, G., Yin, J., Dai, Y., Zhang, J., Yin, X., Cui, L.: A novel convolutional neural network for kidney ultrasound images segmentation. Computer methods and programs in biomedicine **218**, 106712 (2022)
29. Irfan, R., Almazroi, A.A., Rauf, H.T., Damaševičius, R., Nasr, E.A., Abdelgawad, A.E.: Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. Diagnostics **11**(7), 1212 (2021)
30. Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., Chang, C.: Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Medical physics **46**(1), 215–228 (2019)
31. Li, C., Wang, X., Liu, W., Latecki, L.J., Wang, B., Huang, J.: Weakly supervised mitosis detection in breast histopathology images using concentric loss. Medical image analysis **53**, 165–178 (2019)
32. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
33. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
34. Xue, C., Zhu, L., Fu, H., Hu, X., Li, X., Zhang, H., Heng, P.A.: Global guidance network for breast lesion segmentation in ultrasound images. Medical image analysis **70**, 101989 (2021)
35. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters–improve semantic segmentation by global convolutional network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4353–4361 (2017)

36. Li, M.Y., Zhu, D.J., Xu, W., Lin, Y.J., Yung, K.L., Ip, A.W.: Application of u-net with global convolution network module in computer-aided tongue diagnosis. Journal of healthcare engineering **2021**(1), 5853128 (2021)
37. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6924–6932 (2017)
38. Hooley, R.J., Scoutt, L.M., Philpotts, L.E.: Breast ultrasonography: state of the art. Radiology **268**(3), 642–659 (2013)
39. Gonzalez, R.C.: Digital image processing. Pearson education india (2009)
40. Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J.: Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11350–11359 (2023)
41. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. European Radiology Experimental **4**, 1–13 (2020)
42. De Boer, P.T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. Annals of operations research **134**, 19–67 (2005)
43. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
44. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
45. Yap, M.H., Goyal, M., Osman, F., Martí, R., Denton, E., Juette, A., Zwiggelaar, R.: Breast ultrasound region of interest detection and lesion localisation. Artificial Intelligence in Medicine **107**, 101880 (2020)
46. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)