



# PDF-UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network

Ahmed Iqbal<sup>\*</sup>, Muhammad Sharif

*Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan*



## ARTICLE INFO

**Keywords:**

Synthetic images  
U-shaped network  
Breast cancer detection  
Pyramid-dilated network  
Computer-aided diagnosis  
Breast tumor segmentation  
Convolutional neural network  
Generative adversarial network

## ABSTRACT

Rapid and precise segmentation of breast tumors is a severe challenge for the global research community to diagnose breast cancer in younger females. An ultrasound system is a non-invasive and efficient way of breast screening. The area, shape, and texture of different breast tumors play a vital role for clinicians in making accurate diagnostic decisions. Furthermore, the limited availability of breast tumor annotated datasets is another challenge for properly training deep neural networks. This research proposes a semi-supervised learning-based method, which incorporates a Data expansion network (DEN), Probability map generator network (PMG), and U-shaped pyramid-dilated fusion network (PDF-UNet) for accurate breast tumor segmentation. The first DEN network is trained on breast unannotated tumor images and generates synthetic images for the data expansion task. The second PMG network generates corresponding probability map images against synthetic unannotated images. Finally, we proposed a segmentation network (PDF-UNet), a modified variant of UNet, to segment the breast tumor images. The results demonstrate that compared with classical UNet, our proposed PDF-UNet achieves an increment of DSC (2.42%) on the Mendeley dataset and an increment of DSC (1.52%) observed on the SIIT dataset. The results reflect that the proposed method is effective when annotated breast ultrasound data is insufficient to train the network. Furthermore, the proposed method can be helpful in relieving the annotation burden of radiologists. The implementation source code is available at GitHub: <https://github.com/ahmedeqbal/PDF-UNet>.

## 1. Introduction

Breast cancer in females is a life-threatening type of disease, and more than 8% of females may suffer during their lifespan (Cheng, Shan, Ju, Guo, & Zhang, 2010). According to American cancer statistics, 287, 850 breast cases are reported in women, and 2710 cases are reported in men (American Cancer Society, 2022; Dag et al., 2022). Early diagnosis of a malignant breast tumor and immediate treatment could improve five years survival rate of the patient (Eisenbrey et al., 2016). Breast cancer is also the dominant cause of death among younger females. The major factors of breast cancer in females remain unrevealed; identifying the early signs of breast cancer is the only hope of decreasing the death rate. The ultrasound modality is widely accepted for breast cancer screening because it is non-invasive, non-radioactive, affordable, painless, and easily available (Vijayalakshmi & Savita, 2019). The manual screening of breast ultrasound images is costly and time-consuming, even for an expert radiologist (Iqbal, Sharif, Yasmin, et al., 2022). For quick and efficient breast cancer screening, computer-aided diagnosis

(CAD) systems are developed for breast tumor detection (Xian et al., 2018). The segmentation task of the breast tumor areas from the background image is the ultimate goal of any CAD system to follow up on breast cancer treatment plans and diagnosis. The precise segmentation of breast tumors is important in classifying ultrasound images between malignant and benign tumors. Fig. 1 (Case 1) shows ambiguous boundaries of breast tumor region in low contrast, low signal, and speckle noises. Similarly, Fig. 1 (Case 2) represents the large variability of breast tumor structures and shapes. Fig. 1 (Case 3) shows the smallest tumor size is also challenging to segment by the network.

The method for breast tumor segmentation can be subcategorized into conventional and deep learning methods. Many conventional methods are proposed for accurate breast tumor segmentation, such as active contour-based segmentation methods (Yu-Len Huang & Dar-Ren Chen, 2005; Liu et al., 2009), thresholding-based segmentation methods (Mughal et al., 2019; Díaz-Cortés et al., 2018; Hu et al., 2011; Xian et al., 2015; Aja-Fernández et al., 2015), clustering-based segmentation methods (Boukerroui et al., 1998; Prabusankarlaral et al., 2016;

\* Corresponding author.

E-mail address: [ahmedeqbal@gmail.com](mailto:ahmedeqbal@gmail.com) (A. Iqbal).

Huang et al., 2020), graph-based segmentation methods (McClymont et al., 2014; Zhou et al., 2014; Daoud et al., 2019), and watershed-based segmentation methods (Gu et al., 2016; Gómez et al., 2010; Sadat et al., 2020). In previous literature (Amjadian & Gharaei, 2022; Baradaran Rezaei et al., 2022; Gharaei et al., 2021), machine learning techniques such as Logistic regression, Random forest, and Deep feedforward networks are already used for Gastric cancer and Breast cancer-related disease diagnosis. These conventional approaches are simple but require extensive domain knowledge and expertise to extract color, shape, and texture features. Furthermore, conventional methods are not scalable and are considered highly sensitive to noise. The deep learning-based models are broadly incorporated in medical image segmentation for their outstanding performance in recent years. Different networks are proposed for the accurate segmentation of breast tumor images, such as FCN (Shelhamer et al., 2017), UNet (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018). The UNet and its different variants consist of encoder-decoder architectures and use fixed receptive fields. Due to the sequential optimization strategy of multiscale features, classical UNet is experiencing vanishing gradient problems with different optimization issues. Furthermore, the contextual information of multiple scales does not converge appropriately into the reconstruction of the segmentation mask. Moreover, other attention mechanisms have been incorporated with conventional networks, and highly promising results have been demonstrated with more contextual information feature space (Sinha & Dolz, 2020). In previous research, MDA-Net (Iqbal & Sharif, 2022) proposed with multiscale fusion block with composed convolution sequences and a dual attention mechanism with channel and lesion-based attention blocks. The proposed MDA-Net (Iqbal & Sharif, 2022) was tested against breast ultrasound images and simultaneously validated on an MRI dataset. Despite the outstanding performance, these networks are based on supervised-based learning that requires many annotated images for the training dataset, which is a time-taking and labor-intensive task for radiologists. The different researchers focused on using semi-supervised learning, which reduced the

annotation data burden from radiologists (Zhang et al., 2017; Iqbal, Usman, et al., 2022). In semi-supervised based learning, limited numbers of label images are required, and a large number of un-labeled images are utilized. The proposed study mainly focused on three challenges of breast tumor segmentation. First, the limited availability of breast tumor images creates a barrier to properly training the deep learning model. Secondly, ultrasound image annotated data is challenging and requires a highly experienced radiologist to perform manual annotation. It is a time-consuming and costly task. Third, the classical UNet structure is purely based on a fixed receptive field and remains less efficient in extracting context information from breast tumor regions.

This research introduced a semi-supervised learning-based method, incorporating three different networks for breast tumor segmentation with the highest accuracy.

- The first Data expansion network (DEN) is based on two generator and discriminator networks. The main objective of the DEN network is to train on breast unannotated tumor images and generate synthetic images for data expansion tasks.
- The second Probability map generator network (PMG) is modifications in the DeepLabv3+ network by using lower atrous rates in Atrous Spatial Pyramid Pooling (ASPP) module for better segmentation. The objective of the PMG network is to generate corresponding probability maps using unannotated breast ultrasound images.
- Finally, we proposed a segmentation network (PDF-UNet), a modified variant of UNet with some necessary modifications in down-sampling, up-sampling, and embedding pyramid-dilated fusion block (PDF) for better feature extraction. Furthermore, PDF-UNet achieves a large view of the receptive field with small trainable parameter size during the training process while conserving spatial resolution to accurately capture small and large tumors.
- The proposed method implementation source code, synthetic dataset, and annotated Mendeley dataset (renamed BUSC dataset) are released for the breast cancer research community.

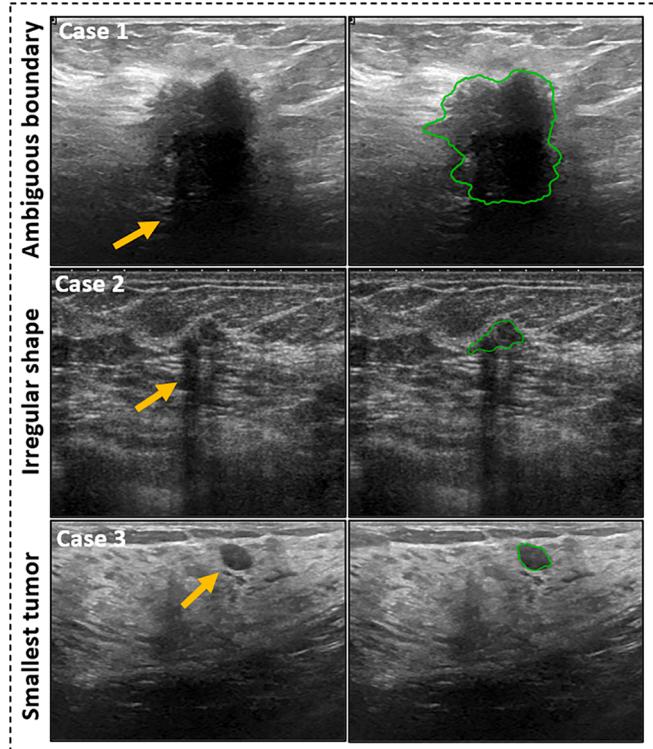
The rest of the work is organized as follows: Section 2 mentioned the related works closed to our proposed work. The detail of methodology is mentioned in Section 3, with a complete mathematical explanation and graphical representation. Section 4 describes the experimental setup and results. Finally, a detailed discussion and conclusions of the proposed research are given in Section 5 and Section 6.

## 2. Related work

For many years automated segmentation of breast tumors has been studied. The existing methods can be classified into conventional and deep learning-based methods. Here, we overview the method most relevant to our proposed work, especially state-of-the-art deep learning networks.

### 2.1. Active contour-based segmentation

The active contouring-based methods are widely used for determining the boundary of a different object. In (Yu-Len Huang & Dar-Ren Chen, 2005), proposed a watershed transformation and active contouring-based method clinically applied for breast ultrasound tumor segmentation. In (Rodtook & Makhanov, 2013), proposed two additions to the previously proposed method generalized gradient vector flow (GGVF). The first diffusion part is added as a polynomial function of intensity of edge map. Secondly, addition is the orientation value of vector fields. The ground truth images validate the proposed algorithm, and results demonstrate better segmentation accuracy than the existing GGVF method. In another work, (Liu et al., 2009) present two steps approach for breast tumor segmentation process. In the first step, texture classifier categorized tissues into different classes, and background



**Fig. 1.** Few ultrasound images with challenging shapes, textures, and boundaries. The green contours represent the ground truth image, and the yellow arrows indicate the ambiguous or missing boundaries in ultrasound images.

information is used for region of interest selection. In the second step, probability distance active contour is used for segmentation region of interest and localize accurate location of tumors. Active contour-based methods initialization process and different parameters tuning is a complicated task in contour convergence. Suppose the initialization process is far from actual boundaries; local minima presence may causes convergence issues. In another scenario, imbalance parameter tuning may produce inaccurate results.

## 2.2. Thresholding-based segmentation

In (Mughal et al., 2019), proposed an automatic thresholding-based approach for breast tumor segmentation. First, an adaptive threshold and relevant templates were created to preserve tumor salient features related to malignant and benign tumor regions in mammograms. Furthermore, the presented method adopted the curve stitching method to remove pectoral muscles. These hybrid technique results are better than a previously proposed method for tumor segmentation using mammogram images. In (Díaz-Cortés et al., 2018), a novel method was introduced for thermographic images that include spatial information on breast tumor regions. This technique adopts Dragonfly Algorithm as an optimization method to compute the best thresholds for accurate breast tumor segmentation. The proposed work could provide an effective clinical tool to help medical experts diagnose and treat breast cancer in women. In another work, (Hu et al., 2011) proposed a novel technique for suspicious tumor detection in mammogram images. The wavelet transform is utilized with adaptive global thresholding and local thresholding method to segment mammogram images. Experimental results are evaluated on the mini-MIAS dataset, which reflects the proposed method segment various types of tumors with minimum false-positive rates. Researchers in (Xian et al., 2015), proposed a model based on spatial frequency knowledge to operate lesion segmentation using a probability-dependent framework. However, all frequency limitations are based on edge detectors, and the cost function is graph-based. A novel multi-region thresholding-based method (Aja-Fernández et al., 2015) is proposed to tackle the most common issues of thresholding when ultrasound images contain noises, artifacts, and corrupted regions. Reliable brain atlases are difficult to achieve for preterm infant brains, which undergo fast structural changes. To address such gaps, (Mukherjee et al., 2018) presented a novel approach for hemorrhage detection without use of a brain atlas. The proposed method can effectively detect ventricles and sample regions of white matter around the ventricles. The results reflect the usefulness of the proposed approach as compared with the expert-annotated dataset. In most cases, global thresholding proves insufficient to extract the region of interest and causes high false positive and false negative rates. In local thresholding, it usually fails to separate the pixels correctly into suitable regions. Selecting threshold value is another hit and tries practice, leading to random and unstable results in most cases.

## 2.3. Clustering-based segmentation

In research, (Boukerroui et al., 1998) propose a two-dimensional  $k$ -means clustering-based method that incorporates textual and grayscale features of ultrasound images. This method is utilized to address the highest posteriori estimation problem. The algorithm Multi resolution implementation is designed using a wavelets basis. In another research (Prabusankarlar et al., 2016), discrete wavelet transform is used to perform multi-resolution experiments on ultrasound images. With an increment in resolution level, the noise level is reduced accordingly. By utilizing inter and intra-resolution of the image feature vector for each pixel is constructed. Furthermore, fuzzy-c-means technique is utilized to segment two disjoint clusters. The fuzzy membership function is utilized to relate every image pixel to a different output centroid. Additionally, the method uses spatial information using local summation, where the degree of membership of per pixel is updated by local details

of neighboring pixels. Segmenting the three dimensions of 3D - Ultrasound volume is a resource-intensive and time-consuming task. (Huang et al., 2020) proposed a semantic classification and merging patched method for accurate breast tumor segmentation. The two diagonal points are cropped the inter-region on the actual image in the first step. In the second step, the histogram equalization method, pyramid mean shift filter, and bilateral filters are simultaneously applied to improve the quality of the image. The third step performs the image cropped process to further divide superpixels using simple linear iterative clustering. Finally, a backpropagation neural network uses bag-of-words and superpixels features for classification tasks. Furthermore, the  $k$ -nearest neighbor is also used for reclassification to refine preliminary results. In clustering techniques,  $k$ -means, and fuzzy c-mean clustering is frequently used as image segmentation methods. These methods require some cluster numbers and conditions before generating a new partition of segmentation images. However, accurate segmentation results are highly dependent and perceptive to cluster centroid and numbers of assignments, which are manually difficult to identify correctly before assignments.

## 2.4. Graph-based segmentation

In (McClymont et al., 2014), proposed a technique that is dependent on mean shifting clustering algorithms and graph cuts at region adjacency graphs. This method is tested on 35 MRI subjects and regions of interest compared with provided ground truth images. In (Zhou et al., 2014), proposed a semi-automatic hybrid method was proposed for accurate breast tumor segmentation, which is based on histogram equalization, Gaussian filtering, pyramid shift, and graph-cuts. Furthermore, according to ROI size, seeds of backgrounds for graph-cuts are generated automatically. Researchers (Daoud et al., 2019) decompose the input image into superpixels to detect breast tumor outlines. The custom graph-cut technique is utilized to label the superpixels regions as a tumor or background tissues. In the second step, the breast lesion outline accuracy is enhanced by using the decomposed method for superior pixels that have the highest boundary recall ratios. Various graph-based methods are proposed for image segmentation, such as radius R,  $k$  nearest neighbors, and 2D image grid-based. The R parameter requires manual adjustment in parameter tuning, and graph connectivity is sensitive to local point density. In the KNN case, parameter adjustment is also a challenge, and uncontrolled distances among various neighbors affect segmentations.

## 2.5. Watershed-based segmentation

In research (Gu et al., 2016), an automated algorithm for 3D ultrasound volume is proposed, which categorizes images into cyst, fatty tissue, and fibro-glandular tissues. Results indicate that the introduced method differentiates between fat and non-fatty issues and achieves good results on cyst/mass images. In (Gómez et al., 2010), introduced a composition of different preprocessing techniques for breast tumor segmentation to gain more accurate contour. The first process is CLAHE, and anisotropic diffusion filtering is implemented on the given image. In the second step, the proposed Watershed transformation method is employed. The results have shown that the presented method is efficient in drawing contours on highly irregular tumor shapes.

In (Sadad et al., 2020), Hilbert transform is used to reconstruct B-mode images from the raw data, followed by the marker-controlled watershed transformation to segment the breast tumor. The texture analysis approaches are highly sensitive to speckle noise and other abnormalities. As a result, after extracting texture and shape-based features of breast tumors, a hybrid features set is constructed. In (Xu et al., 2011), the proposed method retrieved the breast lesion border in the belt between the exterior and internal markers, the smoothed morphological gradient image was transformed using the conventional watershed transformation. The interior and external markers are automatically

determined using a template matching and thresholding method to establish the rough location of the lesion. The internal marker is then calculated using a distance transform, while the exterior marker is determined using morphological dilation. In (Shen et al., 2021), proposed a two step based breast lesion segmentation method. In the initial step, the CLAHE algorithm with a side window filter (SWF) are utilized to perform the prepossessing operations. Secondly, an Adaptive morphological snake (AMS) adjusts the parameters adaptively according to lesion size. The marked watershed (MW) selects the marked area and achieves candidate contours. The average radial derivate (ARD) is utilized to select the breast lesion contour. Watershed-based methods are sensitive to detailed textures and various kinds of noises in input images. Sometimes, low-contrast boundaries of images are unstable, and oversegmentation is considered a great challenge for the breast cancer segmentation research community.

## 2.6. Deep learning-based segmentation

In (Xing et al., 2021), proposed a new deep-learning method based on semi-pixel-wise cGANs for breast tumor segmentation. The proposed scheme incorporates a fully conventional network (FCN) and GAN advantages to segment tumor regions by using prior knowledge. The semi-supervised learning-based BUS-GAN model is proposed, where the segmentation network is known as BUS-S, and the evaluation network is called BUS-E network. The network BUS-S extracts multi-scale features from breast tumor regions and improves segmentation accuracy. Similarly, the BUS-E network is based on a dual-attention fusion module which has a spatial attention path to predict segmentation maps. The results demonstrate that BUS-GAN can efficiently use unannotated breast ultrasound images to enhance segmentation results. In (Singh et al., 2020), proposed a two-step breast lesion classification and segmentation approach that segments the breast lesion region using cGAN and classify binary masks using a CNN, which is based on shape descriptor. Furthermore, CGAN enforced learning to distinguish between real and synthetic images and create a binary mask as realistic as possible. Results demonstrate that cGAN performs well even with a limited training dataset and outperforms different state-of-the-art approaches. In (Negi et al., 2020), proposed a GAN-based algorithm for accurate segmentation of lesion breast ultrasound images. The generator block of the presented method is based on Residual dilated attention gate UNet. Similarly, the discriminator block is dependent on the CNN network, and it acts as a classifier. Researchers in study (Xiangyuan Ma, 2017) proposed an automated deep learning-based method, which utilizes a GAN network to identify fibroglandular tissue (FGT) regions from MRI images. As compared with the traditional UNet network presented to achieve more accurate and stable FGT segmentation results. The receptive fields play an essential role in extracting context information from tumor regions. The classical UNet is also suffered from limited receptive field challenges in convolutions operations. In research (Iqbal, Sharif, Khan, et al., 2022), a modified variant of FF-UNet is proposed by adjusting the fixed receptive field via a feature-fused module and attention-gate mechanism. In medical imaging, vision transformers are an emerging method for the accurate segmentation of disease. In other research, (Iqbal & Sharif, 2023) proposed a BTS-ST - Swin transformer network for the segmentation and classification of breast tumor images task. The structure of BTS-ST network incorporates Spatial Interaction block (SIB), Feature Compression block (FCB), and Relationship Aggregation block (RAB), respectively. In (Iqbal, Usman, & Ahmed, 2023) proposed TB-UNet, which is based on dilated fusion block (DF) and Attention block (AB) for segmentation task. TB-DenseNet is also proposed, which is based on five dual convolution blocks, a DenseNet-169 layer, and a feature fusion block for accurate classification task. In (Wu et al., 2022), proposed an adaptive feature transformer that is based on encoder-decoder architecture. The researcher also employs a memory-efficient decoder to improve the feature fusion capability of the network. The proposed architecture is based on the pretrained

'beit-tiny-distilled-patch16-224' model, and the encoder is partly based on ResNet-34. Similarly, TransUNet (J. Chen et al., 2021) and TransFuse (Zhang et al., 2021) networks also proposed for medical image segmentation. The standard transformer networks produce results that are below average because they only focus on global modeling and ignore positioning ability. The major limitation of vision transformer networks are complex structures and require huge parameter sizes, and most of the encoders are based on the pre-trained architecture of ResNet. As compared with traditional encoder-decoder architecture UNet (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018), we proposed dual pooling concatenation of max pooling and average pooling to obtain lower information loss. To alleviate the semantic gap, traditional skip connection is replaced to map the low-level features from the encoder to a lower dimension before concatenating them with high-level features from the decoder. In the bottleneck of encoder-decoder, Atrous Spatial Pyramid Pooling (ASPP) module is added to with different dilation rates, and their outputs are concatenated so that the network can better capture multi-scale features for segmentation of small and large tumors. Additionally, we use the residual connection in each convolutional block to facilitate the training and add a Squeeze-and-Excitation block to calibrate different channels for optimum results. In (Yang & Yang, 2023), proposed a pyramid-shaped network for feature encoding and decoding task, where CNNs and a Swin-Transformer are connected as a backbone for feature extraction tasks. In order to identify important features, first construct an interactive channel attention (ICA) module that uses channel-wise attention. Second, based on the gating mechanism, a supplemental feature fusion (SFF) module is introduced. The performance of breast lesion segmentation can be enhanced by adding the SFF module to the features during feature fusion.

## 2.7. Research gaps and contributions of this research

The basic limitation of conventional methods (Active contour-based, Thresholding-based, Clustering-based, Graph-based, and Watershed-based methods) are greatly influenced by image feature definitions, seed points, and various threshold parameters. Similarly, deep learning methods do not need precise image feature definitions compared with conventional feature-oriented methods. The deep neural network can learn a thorough relationship between the foreground pixels and the background pixels by organizing the global contextual information of the target. These learned relationship patterns help the deep neural network to identify the targeted regions. In biomedical imaging tasks, rich global contextual information is also crucial for the accurate segmentation of breast tumors. In previous research, global contextual information significance is ignored by various studies. In this research, our contribution is to propose three different networks to overcome data availability issues, unlabeled data issues, and a proposed segmentation network for accurate breast tumor images. In the DEN network pooling layer is replaced with stride convolution, and filter size and kernel size are modified in the generator network. The discriminator network is also modified using the LeakyReLU activation function. Furthermore, batch normalization is adopted in both the generator and discrimination network. The PMG network is modification in the DeepLabv3+ (Chen et al., 2018) network by using lower atrous rates in Atrous Spatial Pyramid Pooling (ASPP) module for better segmentation. We also used to use the ResNet-101 network as the backbone for feature extraction tasks. Finally, we proposed a segmentation network (PDF-UNet), a modified variant of UNet with some important changes in the down-sampling path, up-sampling path, and embedding pyramid-dilated fusion block (PDF) for better feature extraction.

## 3. Methodology

The proposed method comprises three networks: DEN network, PMG network, and PDF-UNet. The DEN network is inspired by the deep convolutional generative adversarial network DCGAN (Radford et al.,

2015) architecture and trained on unannotated breast tumor images using unsupervised-based learning. The PMG network is inspired by DeepLabv3+ (Chen et al., 2018) architecture, a pre-trained model that uses unannotated breast ultrasound images to generate corresponding probability maps. Subsequently, we proposed a U-shaped segmentation network that takes original breast ultrasound images and GANs-based generated synthetic images with corresponding probability maps to train the model. The explanation of each method is given in the subsequent sections. The complete structure of the proposed semi-supervised scheme for the segmentation process of breast tumor images is shown in Fig. 2. In first step, the unannotated images ( $x_u$ ) set is given to DEN network and synthetic images ( $x_s$ ) are generated. In second step, we tackle the unannotated synthetic images issue and generated the probability maps. The PMG network is trained on an additional annotated dataset that contains images ( $x_l$ ) and their corresponding ground truth images ( $y_l$ ). When PMG network is trained, probability maps ( $y_p$ ) are generated against  $x_s$  set, which are synthetic images. In third step, the training dataset batch is ready to fed to the proposed PDF-UNet, which is containing the annotated images ( $x_l, y_l$ ) and synthetic images set with their probability maps ( $x_s, y_p$ ). The synthetic images with probability maps and annotated images are combined and fed to PDF-UNet for the training and final segmentation results.

### 3.1. Data expansion network (DEN)

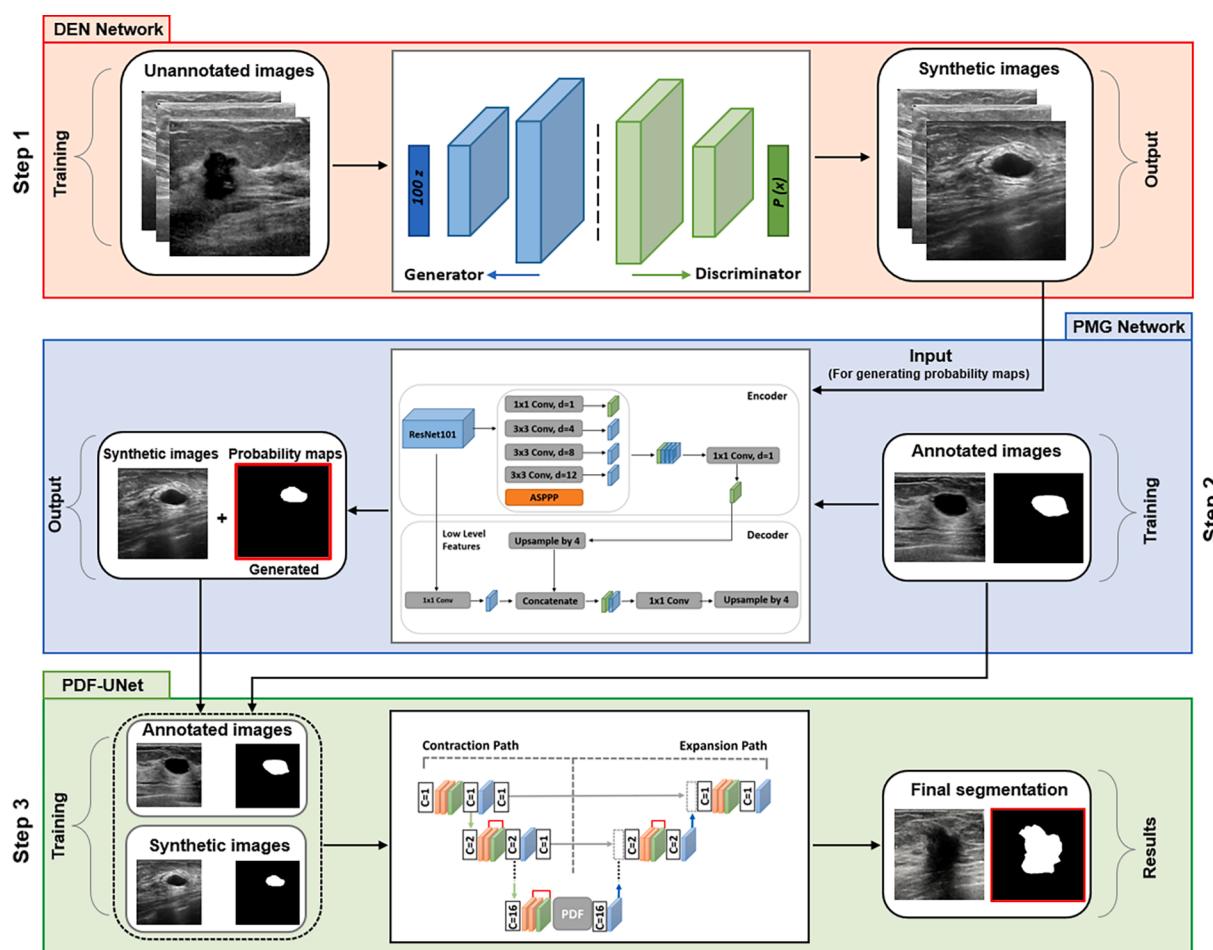
Generative adversarial networks (GANs) are composed of two different parts named generator and discriminator networks to estimate

the generative result via the adversarial process (Goodfellow et al., 2014). The generative network produces a similar distribution of data  $G$  as the given training dataset. In this dataset  $G$ , there are noisy image set which is denoted  $N$ . These image sets were obtained by a sampling of uncorrelated Gaussian position of each image  $x$ , which is part of  $\Omega$ , where image lie in to rectangular domain, here  $\mathbb{R}^2$  is a subset of all supported images. The generative model is denoted by  $G : \mathbb{R}^N \rightarrow X(N, L_1(\Omega))$ . Then given variable weights  $w \in \mathbb{R}^N$  it constructs a continuous mapping  $G_w$  from the set of noise images  $N$  to a new set. The new set of images contains the fake distribution of images  $f = G_w(N) \subseteq L_1(\Omega)$ . The corresponding set of true images is represented as  $T \subseteq L_1(\Omega)$ . During the training process, true labels are labeled as 1. The discriminative model  $\mathbb{R}^M \ni v \rightarrow D_v$ , where operator  $D_v : L_1(\Omega) \rightarrow \{0, 1\}$  estimates the originality of sample data acquired from train set, and  $v$  are weights of the discriminator network. Here,  $D_v$  represents the discriminative operator weights and  $\mathbb{R}^M$  shows the sample datasets used for training purpose. The loss function is applied on generative network  $G$  and discriminative network  $D$ , whether provided input samples are realistic or fake. The network  $G$  and  $D$  influence are used to evaluate the performance of each other, and the loss function is based minimax game and represented as follows:

$$\min_{w \in \mathbb{R}^N} \max_{v \in \mathbb{R}^M} l_{GAN}(D_v, G_w, T, N) \quad (1)$$

$$l_{GAN}(D_v, G_w, T, N) = E_p[\log D_v(\cdot)] + E_p[\log(1 - D(G_w(\cdot))] \quad (2)$$

$$= \frac{1}{\tau} \sum_{f \in T} \log(D_v f) + \frac{1}{N} \sum_{z \in N} \log(1 - D_V(G_w z)) \quad (3)$$



**Fig. 2.** Overview of semi-supervised breast tumor segmentation method (a) The DEN network uses unannotated breast ultrasound images to generate synthetic images. (b) The PMG network is trained using annotated images and generates probability maps for synthetic images. (c) The PDF-UNet uses annotated, and probability map images for the training process and predicts segmentation results on provided test datasets.

Here the weights for generator network  $G$  are represented as  $w$  and  $v$  are the weights of discriminator network  $D$ . The probability the data is  $D_v f$  and the  $f \in L_1(\Omega)$  is a representation of realistic images, not a fake image. In addition,  $p_{Data} f = \frac{1}{N}$  is continuous distribution over real images, and  $p$  is the continuous distribution of noisy images  $f \in L_1(\Omega)$ , where  $N$  represent the overall numbers of noisy images. The discriminative model  $D$  is trained on increased probability to allocate true labels to input data  $f \in L_1(\Omega)$ . However, generative network  $G$  is trained to interpret the understanding by discriminative network  $D$ . The Eq (1) is individual from  $G$  is done by reducing the Eq (2), which is an assumption  $E_p[\log(1 - D_v(G_w(.)))]$ .

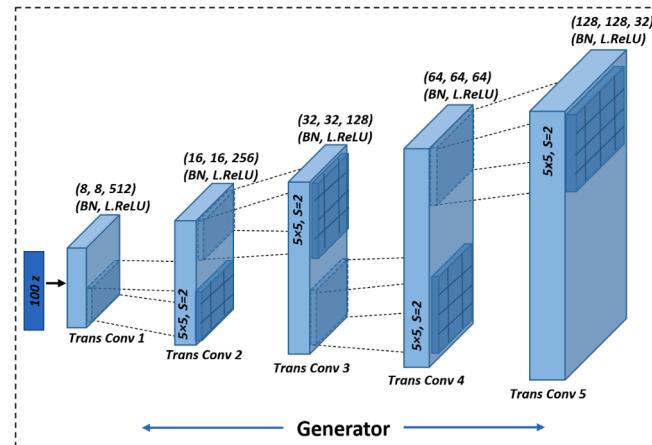
The training of the original GAN model is complex, and synthetic images suffer from the diversity of samples. Therefore, DCGAN (Radford et al., 2015) was proposed to improve the quality of samples and speed of convergence. Inspired by DCGAN (Radford et al., 2015), our proposed DEN is modified in five different aspects: 1) The pooling layers are skipped, and fractionally-strided convolution is used for up-sampling in generative part  $G$ , with a large string convolution layer used instated of the pooling layer in the discriminative network. 2) Batch normalization is adopted in discriminative and generative networks. 3) The FC layer is skipped to create the model completely fully convolution network. 4) In generative  $G$ , the ReLU function is utilized as activation in all layers expected final layer, Tanh function is adopted in the last layer. 5) LeakyReLU is used as activation function in the  $D$  network. The complete architectural overview is presented in Fig. 3 and Fig. 4.

In the proposed DEN, the encoding path is based on five different convolutional filters after batch normalization and LeakyReLU activation. In contrast, the decoder is based on deconvolutional kernels, which come after batch normalization, dropout, and LeakyReLU. The input of the generative network is a 100-dimensional uniform distribution  $Z$ . The convolutional and deconvolutional kernels are designed with  $5 \times 5$  and  $3 \times 3$  with a step size of 2 with zero padding, which up-sampled and down-sampled activation maps with a factor of 2. The generator ( $G$ ) is based on five different encoding layers with convolution filters size of  $32 \times 32$  and stride size of  $2 \times 2$ . Similarly, discriminator ( $D$ ) convolution uses filter size  $64 \times 64$ , with a step size of 1 with zero padding.

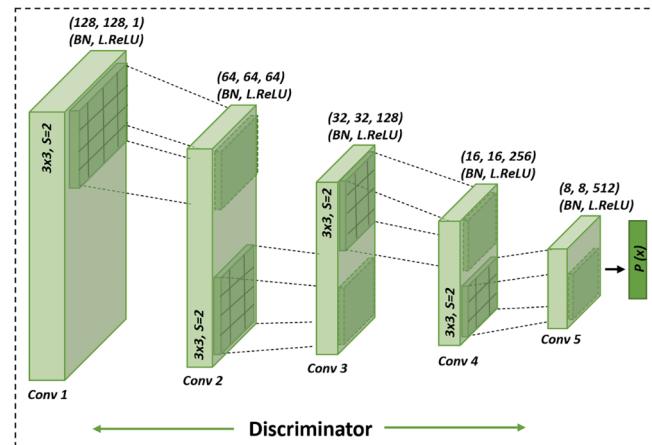
Finally, a  $128 \times 128$  pixel synthetic image is generated by network  $G$ . Fig. 5 shows a few sample images of breast tumors generated by the DEN network.

### 3.2. Probability map generator (PMG)

In this research, we used the PMG network as a probability map generator network inspired by DeepLabv3 (Chen et al., 2017) by adding an improved decoder module to recover the boundaries of objects. The DeepLabv3+ (Chen et al., 2018) used atrous convolutions with pyramid pooling to enlarge the field of view of kernels to integrate high context and control resolution of the capture feature. The use of atrous convolutions in parallel or cascade extracts the multi-scale content due to different sizes of atrous rates. DeepLabv3+ (Chen et al., 2018) used the backbone as ResNet-101 (Wang et al., 2017), and the feature extractor expected for the last block is updated to use atrous convolution with different dilation rates. For effectively extracting high multiscale information, four different Atrous spatial pyramid pooling (ASPP) blocks are applied with atrous rates of 6, 12, 18, respectively. But large atrous rates are only feasible for a large object or structured environment. Because our research focuses on breast tumor segmentation tasks, we used a lower atrous convolution rate to make receptive small for tumor regions. Optimizing atrous rates to 4, 8, and 12 enhanced the performance of breast tumor segmentation on the test dataset. Furthermore, the ASSP block is applied as depth-wise separable convolution instead of the standard convolution operation. The final segmentation result is then produced by performing a  $3 \times 3$  convolution followed by another up-sampling by 4 to produce the final segmentation output. Finally, the



**Fig. 3.** The generator network architecture is based on five different Transposed Convolutions operations followed by kernel size of  $5 \times 5$ , and stride rate of 2, respectively.



**Fig. 4.** The discriminator network architecture is based on five different convolution operations followed by kernel size of  $3 \times 3$ , and stride rate of 2, respectively.

decoder block enables detailed object boundary recovery. Fig. 6. shows a few synthetic images with their corresponding probability maps generated by the PMG network. The synthetic dataset is available online<sup>1</sup>.

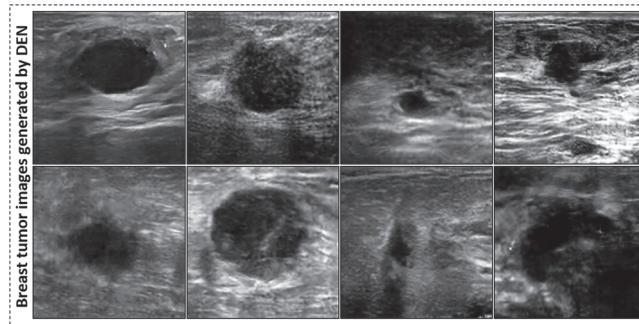
### 3.3. Pyramid-dilated fusion network (PDF-UNet)

The basic structure of the proposed Pyramid-dilated fusion network (PDF-UNet) follows the encoder-decoder structure and is inspired by UNet (Ronneberger et al., 2015; Oktay et al., 2018; Huang et al., 2019). The symmetrical structure of UNet (Ronneberger et al., 2015) is composed of low-level and high-level semantic information and obtained excellent effects on medical segmentation. The proposed segmentation network is shown in Fig. 7. To reduce the information loss, we have modified the down-sampling block of traditional UNet (Ronneberger et al., 2015) architecture. Furthermore, skip connections are modified before concatenating high-level features of the decoder block to reduce the semantic gap between low-level and high-level features. The proposed modifications are mentioned in the below sub-sections.

#### 3.3.1. Down-sampling block

The previous work used max pooling in the down-sampling path of

<sup>1</sup> BUS synthetic dataset: <https://data.mendeley.com/datasets/r4phnt49r7/1>.



**Fig. 5.** The DEN network generated a few synthetic images of breast tumors.

the network. In our work, we obtained lower information loss using max pooling and average pooling and then concatenating both pooling in the down-sampling path.

$$Z_c = (Z_{\max} + Z_{\text{avg}}) \quad (4)$$

In Eq (4),  $Z_{\max}$  and  $Z_{\text{avg}}$  represents the max pooling and average pooling, then concatenation is performed using  $Z_c$ .

$$C_c = \sigma_c(\beta_c(\phi_c^*(Z_c \cdot W_c + \lambda_c))) \quad (5)$$

To perform LeakyReLU, batch normalization, and convolution operation on max-pooling and average-pooling ( $Z_c$ ). In Eq (5)  $W_c$  are weights of convolution operations,  $\lambda_c$  bias factors, which are multiple with a product of  $Z_c$ , and  $\phi_c$  convolution operation is applied.  $\beta_c$  is denoted as batch normalization and  $\sigma_c = f(x) = \max(x, \text{leak } x)$  represents the Leaky ReLU activation functions.

$$S_c = \sigma_r(\phi_c^*(\sigma_l(\phi_c^*(Z_{a,\text{avg}}(C_c \cdot W_c + \lambda_c)))))) \quad (6)$$

Eq (6) is used for the different channel calibration and squeeze-and-excitation is performed. Similarly,  $W_c$  are weights of convolution operations  $\lambda_c$  denote the bias factors multiple with a product of  $Z_c$ . The  $Z_{a,\text{avg}}$  is average-pooling operation and  $\phi_c$ ,  $\sigma_l$ , and  $\sigma_r$  are convolution operations, LeakyReLU and simple ReLU activation functions.

### 3.3.2. Up-sampling block

To improve the spatial resolution prior to the concatenation operation of high-level and low-level features, we use an up-sampling layer in the decoder block followed by a bilinear operation, where a previous feature channel number decreased to match the output of the bridge layer.

$$P_c = \phi_c^*(Y_c + (U_b^*(X_c \cdot W_c + \lambda_c))) \quad (7)$$

In Eq (7).  $X_c$ ,  $W_c$ , and  $\lambda_c$  shows that convolution operation is followed by weights of convolution operations and bias factors, respectively.  $U_b$  denote the up-sampling using bilinear operation,  $Y_c$  is the concatenation of convolution operations, and up-sampling followed by  $\phi_c$   $1 \times 1$  convolution operation.

### 3.3.3. Pyramid-dilated fusion block (PDF)

The concept of dilated-convolution operation was first introduced as “algorithme à trous” for wavelet decomposition (Holschneider et al., 1990). The primary concept is to add a number of zeros in pixels in convolutional filters. Consequently, higher image resolution is achieved with dense feature extraction from convolutional networks. For better segmentation of breast tumor regions at different scales, we embedded a Pyramid-dilated fusion (PDF) block in the bottleneck of encoder-decoder architecture. In Fig. 8, pyramid dilated fusion block architecture is illustrated. The PDF block includes four different aligned layers of dilated convolution with multiple dilation rates and corresponding concatenate so the proposed network can better extract multiscale

features for the segmentation of large and small size of breast tumors.

$$P_d = CAT(\phi_{d=1}^{k=1} + \phi_{d=2}^{k=3} + \phi_{d=4}^{k=3} + \phi_{d=6}^{k=3}) \quad (8)$$

The pyramid-dilated operation is defined in Eq. (8), where *CAT* represents the concatenation operation,  $\phi$  symbol shows the convolution operation,  $k$  and  $d$  represent the convolution kernel size and dilation size, respectively.

$$\text{PDF}_b = \sigma_l(\beta_c(\phi_c^*(P_d \cdot W_c + \lambda_c))) \quad (9)$$

In Eq. (9).  $W_c$  are weights of convolution operations,  $\lambda_c$  bias factors, which are multiple with a product of  $P_d$ , and  $\phi_c$ ,  $\beta_c$ , and  $\sigma_l$  represents the  $1 \times 1$  convolution, batch normalization, and LeakyReLU activation functions, respectively.

## 4. Experimental results

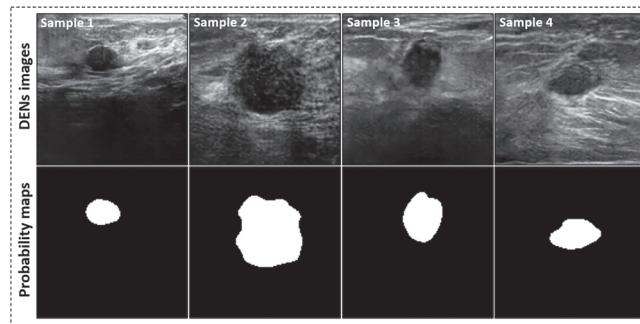
This section detailed the experimental results on three different breast ultrasound datasets to validate the usefulness of the proposed work.

### 4.1. Dataset description

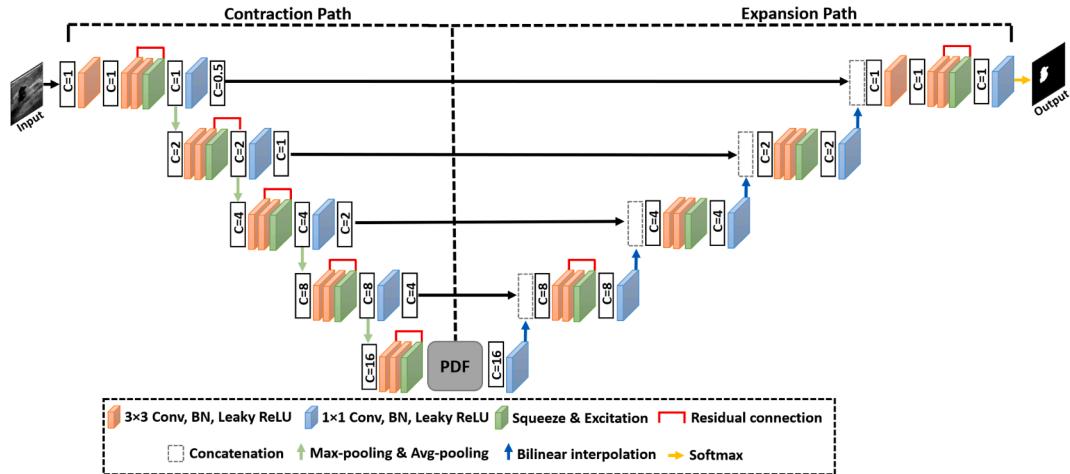
To assess the performance of our proposed method, we used three different 2D breast cancer ultrasound datasets.

- The BUSI dataset is provided by Bahya Hospital, Cairo Egypt (Al-Dhabayani et al., 2020). All Ultrasound images are acquired by LGIQ E9 ultrasound and LGIQ E9 Agile ultrasound system. The dataset includes 780 patients images of benign and malignant tumors. The ground truth images are also provided by expert radiologists, and images are available in  $1280 \times 1024$  resolution.
- The Mendeley (Paulo Sergio Rodrigues, 2017) ultrasound dataset includes 100 benign images and 150 malignant cancer images. The original resolution of ultrasound images is  $64 \times 64$  pixels, later transformed to  $128 \times 128$  pixels. The dataset is basically classification-based, and no ground truth images are provided. Therefore, with the help of an experienced radiologist, benign and malignant tumor images are annotated for the model training process. The BUSC dataset is freely available for download<sup>2</sup>.
- The SIIT (Karunananayake et al., 2020) ultrasound images are provided by Thammasat University Hospital, which were taken by Philips IU 22 ultrasound system. The dataset includes 180 patients images of benign and malignant tumors. The unannotated ultrasound images are manually annotated by three highly experienced radiologists using a Samsung tablet. The ground truth images were gathered by the majority voting rule (2 / 3) scheme. The original resolution range is between 200 and 400 pixels.

<sup>2</sup> BUSC Dataset: <https://data.mendeley.com/datasets/vckdnhtw26/1>.



**Fig. 6.** Few synthetic images with their corresponding probability maps.



**Fig. 7.** The visual representation of PDF-UNet over breast tumor images. The network structure is based on the encoder-decoder block with necessary modifications. The pyramid dilated fusion block is introduced to collect more detailed information using different dilation rates.

#### 4.2. Experimental setup

The proposed scheme is implemented in PyTorch deep learning library, and experiments are performed on 10<sup>th</sup> Generation Intel Core i5 Processor, 16 GB RAM, and RTX 2060 Supper NVIDIA single graphic card. All four datasets images were available in PNG or JPG format. In pre-processing steps, the train, validation, and test dataset resolution were uniformly resized into 128 × 128. For the experimental purpose, the whole dataset is divided into the training dataset (70%), validation dataset (15%), and testing dataset (15%). The 5-fold cross-validation scheme is used to achieve the stable experimental results. The batch size is set as 4, and the maximum number of 100 epochs are used for training. We utilized ADAM optimizer using learning rate '1e-4', learning scheduler ReduceLROnPlateau with patience size 10.

#### 4.3. Evaluation metrics

For quantitative comparison with the previous state-of-the-art networks, we used different evaluation metrics such as the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), and Hausdorff Distance (HD<sup>95</sup>). Jaccard similarity coefficient, also known as intersection over the union, determines a ratio of intersection union of two sets of ground truth images and prediction results. The DSC is a similarity measure to calculate the similarity of the ground truth image and prediction results, which can define as Eq. (10), and Eq. (11):

$$JSC(G\tilde{T}, P\tilde{M}) = \frac{|G\tilde{T} \cap P\tilde{M}|}{|G\tilde{T} \cup P\tilde{M}|} \quad (10)$$

$$DSC(G\tilde{T}, P\tilde{M}) = \frac{2|G\tilde{T} \cap P\tilde{M}|}{|G\tilde{T}| + |P\tilde{M}|} \quad (11)$$

Hausdorff distance (HD) is used as a performance measure to compute the distance between two point sets. Where  $P\tilde{M}$  and  $G\tilde{T}$  represents the probability maps and the ground truth. HD mathematical formulation is defined in Eq. (12) and Eq. (13).

$$HD'(G\tilde{T}, P\tilde{M}) = \max_{i \in G^t} \min_{j \in P^t} \|i - j\|_2 \quad (12)$$

$$HD(G\tilde{T}, P\tilde{M}) = \max(HD'(G\tilde{T}, P\tilde{M}), HD'(P\tilde{M}, G\tilde{T})) \quad (13)$$

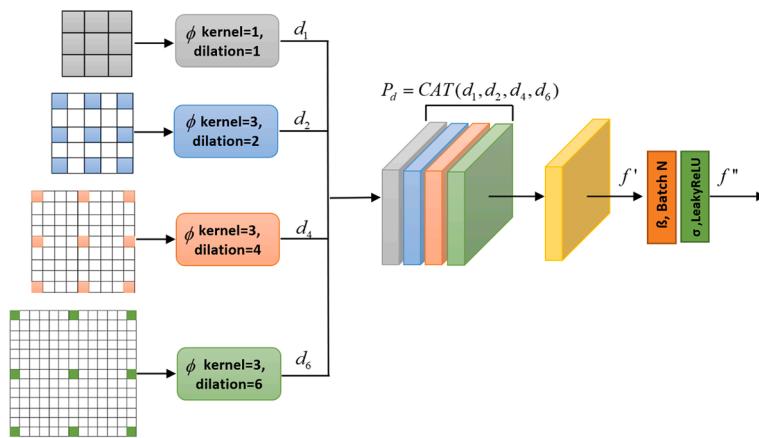
In our research, In HD<sup>95</sup> is used, where 95-th percentile are used rather than using the maximal value.

#### 4.4. Loss function

The Tversky index was introduced by (Salehi et al., 2017) for the optimum generalization of Dice scores by balancing the false positives and false negative rates. Similarly, Focal Tversky loss (Abraham & Khan, 2019) was introduced to add more emphasis on hard training and improve accurate predictions. The loss function is defined in Eq. (14) and Eq. (15).

$$T_{loss} = \sum_{i=1}^N (T_p + \delta) / (T_p + \alpha^* F_p + \beta^* F_N + \delta) \quad (14)$$

$$\tilde{T}_{loss} = \sum_{i=1}^N (1 - T_{loss})^* \gamma \quad (15)$$



**Fig. 8.** The pyramid dilated fusion block architectures with multiple dilation rates. The dilation function is performed in the second stage, followed by convolution, batch normalization, and leakyReLU operations.

The  $T_{loss}$  and  $\tilde{T}_{loss}$  represent the Tversky loss and Focal Tversky loss. Here  $T_p$ ,  $F_p$ , and  $F_N$  are representing the True positive, False positive, and False negative values of probability pixels. The hyperparameters  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  values are set to 0.5, 0.5, 1, and 1 respectively.

#### 4.5. Quantitative results

The ablation study is carried out to test the usefulness of our PDF-UNet over two different datasets. The performance of the best variant is compared with previous state-of-the-art methods.

##### 4.5.1. Ablation study

The BUSI dataset is only used for the Data expansion network (DEN) that contains 3120 images of benign and malignant breast cancers. The Mendeley (original) and SIIT (original) datasets include 400 and 546 images, respectively. The Expansion network (DEN) is used to generate 500 synthetic images, and the probability maps generator network (PMG) is used to generate their corresponding probability maps. The Mendeley (DEN) and SIIT (DEN) datasets also include 400 and 546 images with additional 500 DENs-generated images. During the testing process, we used only original images with their expert annotated ground truth images. The data augmentation is achieved using five different augmentation techniques: horizontal flipping, vertical flipping, transpose, zoom, and gaussian noise, respectively. First, we performed ablation-based studies and validated the overall effectiveness of our proposed Pyramid dilated fusion (PDF) block by using the same training configuration and datasets.

**Table 1** shows proposed network results with and without Pyramid dilated fusion block on the Mendeley dataset. In the original dataset, PDF-UNet without Pyramid dilated fusion block achieved JSC (0.5000), DSC(0.6218), and HD<sup>95</sup> (26.5379). Comparatively, PDF-UNet with Pyramid dilated fusion block achieved an improvement of JSC (9.5%), DSC (9.59%), and a decrease in HD<sup>95</sup> (4.1%) which is significantly large. In the DEN-based synthetic dataset, PDF-UNet without Pyramid dilated fusion block achieved JSC (0.7199), DSC(0.8090), and HD<sup>95</sup> (10.2241), respectively. Comparatively, PDF-UNet with Pyramid dilated fusion

block achieved an improvement of JSC (4.66%), DSC (4.2%), and a decrease in HD<sup>95</sup> (5%) which is significantly higher.

**Table 2** shows proposed network results with and without Pyramid dilated fusion block on the SIIT dataset. In the original dataset, PDF-UNet without Pyramid dilated fusion block achieved JSC (0.7235), DSC (0.8173), and HD<sup>95</sup> (11.0885). Comparatively, PDF-UNet with Pyramid dilated fusion block achieved an improvement of JSC (1.51%), DSC (0.87%), and a decrease in HD<sup>95</sup> (2.76%), which is significantly large. In the DEN-based synthetic dataset, PDF-UNet without Pyramid dilated fusion block achieved JSC (0.7616), DSC(0.8398), and HD<sup>95</sup> (6.9629), respectively. Comparatively, PDF-UNet with Pyramid dilated fusion block achieved an improvement of JSC (2.41%), DSC (2.65%), and a decrease in HD<sup>95</sup> (2.55%), which is significantly higher.

In **Table 3** and **Table 4**, the overall impact of different loss functions is evaluated on two different datasets Mendeley ultrasound dataset and the SIIT ultrasound dataset. **Table 3** results show that the Focal Tversky Loss function achieved JSC (1.7%), and JSC (1.3%) increment compared with BCE Loss and Dice loss. Similarly, Focal Tversky Loss function achieved DSC (2.07%), and DSC (1.7%) increment compared with BCE Loss and Dice loss. In the case of DEN-based synthetic dataset, Focal Tversky Loss function achieved JSC (1.76%), and JSC (0.67%) increment compared with BCE Loss and Dice loss. Similarly, Focal Tversky Loss function achieved DSC (2.1%), and DSC (1.1%) increment compared with BCE Loss and Dice loss. Furthermore, HD<sup>95</sup> is achieved 5.0392 mm, which is comparatively lower than other loss functions.

**Table 4** results show that the Focal Tversky Loss function achieved JSC (1.56%), and JSC (0.89%) increment compared with BCE Loss and Dice loss. Similarly, the Focal Tversky Loss function achieved DSC (0.61%), and DSC (1.51%) increments compared with BCE Loss and Dice loss. Focal Tversky Loss achieved HD<sup>95</sup> (8.3265 mm), which is comparatively lower than other loss functions. In the case of DEN-based synthetic dataset, the Focal Tversky Loss function achieved JSC (2.54%), and JSC (2.47%) increment compared with BCE Loss and Dice loss. Similarly, the Focal Tversky Loss function achieved DSC (1.73%), and DSC (0.83%) increment compared with BCE Loss and Dice loss. Furthermore, the Focal Tversky HD<sup>95</sup> achieved 4.4081 mm, which is

**Table 1**

Ablation results of PDF-UNet with pyramid dilated fusion block on Mendeley ultrasound dataset.

Method	PDF block	Results with the original dataset			Results with addition of DEN-based synthetic dataset		
		JSC ↑	DSC ↑	HD <sup>95</sup> ↓	JSC ↑	DSC ↑	HD <sup>95</sup> ↓
PDF-UNet	✗	Mean	0.5000	0.6218	26.5379	0.7199	0.8090
	✓	StdDev	0.0189	0.0222	1.7400	0.0158	0.0140
PDF-UNet	✗	Mean	<b>0.5950</b>	<b>0.7177</b>	<b>22.4332</b>	<b>0.7665</b>	<b>0.8510</b>
	✓	StdDev	0.0267	0.0252	2.2900	0.0125	0.0122

**Table 2**

Ablation results of PDF-UNet with pyramid dilated fusion block on SIIT ultrasound dataset.

Method	PDF block	Results with the original dataset			Results with addition of DEN-based synthetic dataset		
		JSC ↑	DSC ↑	HD <sup>95</sup> ↓	JSC ↑	DSC ↑	HD <sup>95</sup> ↓
PDF-UNet	✗	Mean	0.7235	0.8173	11.0885	0.7616	0.8398
		StdDev	0.0163	0.0137	2.7900	0.0035	0.0045
PDF-UNet	✓	Mean	<b>0.7386</b>	<b>0.8260</b>	<b>8.3265</b>	<b>0.7857</b>	<b>0.8663</b>
		StdDev	0.0266	0.0215	0.0000	0.0066	0.0055

**Table 3**

Ablation results of PDF-UNet with combination different loss functions on Mendeley ultrasound dataset.

Loss function	Results with the original dataset			Results with addition of DEN-based synthetic dataset		
	JSC ↑	DSC ↑	HD <sup>95</sup> ↓	JSC ↑	DSC ↑	HD <sup>95</sup> ↓
BCE	Mean	0.5780	0.6970	22.4392	0.7489	0.8300
	StdDev	0.0169	0.0151	2.2400	0.0125	0.0122
Dice	Mean	0.5820	0.7007	23.4233	0.7598	0.8400
	StdDev	0.0245	0.0242	2.1000	0.0129	0.0112
Focal Tversky	Mean	<b>0.5950</b>	<b>0.7177</b>	<b>22.4332</b>	<b>0.7665</b>	<b>0.8510</b>
	StdDev	0.0267	0.0252	2.2900	0.0120	0.0124

**Table 4**

Ablation results of PDF-UNet with combination different loss functions on SIIT ultrasound dataset.

Loss function	Results with the original dataset			Results with addition of DEN-based synthetic dataset		
	JSC ↑	DSC ↑	HD <sup>95</sup> ↓	JSC ↑	DSC ↑	HD <sup>95</sup> ↓
BCE	Mean	0.7230	0.8199	11.0985	0.7603	0.8490
	StdDev	0.0163	0.0137	1.7800	0.0030	0.0041
Dice	Mean	0.7297	0.8109	10.0478	0.7610	0.8580
	StdDev	0.0167	0.0135	2.7800	0.0033	0.0040
Focal Tversky	Mean	<b>0.7386</b>	<b>0.8260</b>	<b>8.3265</b>	<b>0.7857</b>	<b>0.8663</b>
	StdDev	0.0266	0.0215	0.1000	0.0066	0.0055

comparatively lower than other loss functions.

To analyze the visual significance of each BL, PDF, BCE loss function, Dice loss function, and Focal Tversky Loss (FTL) function, we presented results in Fig 9. Visual results are projected on four sample images. The results demonstrate that embedding a PDF block with baseline results improves. However, the PDF + BCE loss function and PDF + Dice loss function combination do not provide outperforming results. The PDF + FTL results are comparatively better than the other two loss functions.

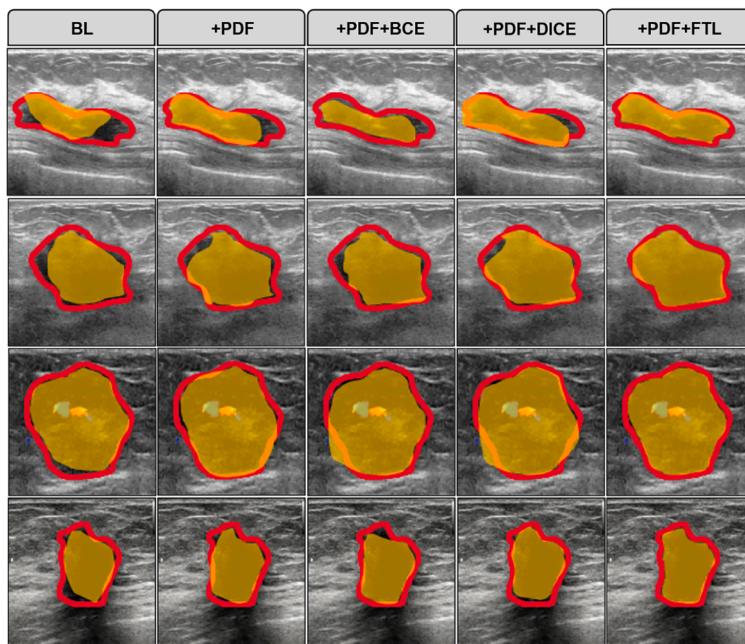
#### 4.5.2. Comparison with other deep-learning methods

Table 5. reflects the performances of PDF-UNet with other state-of-the-art methods on the Mendeley dataset. It can be noticed our network outperforms all the networks by a substantial margin in all the metrics. In the original dataset, the proposed network PDF-UNet achieved a 7.64% improvement in JSC measure, 8.46% improvement in DSC measure, and achieved 17% lower HD<sup>95</sup> scores as compared with the competitor UNet (Ronneberger et al., 2015) network. The second-best results are observed in the case of DeepLabv3+ (Chen et al., 2018), 0.5194 results in the JSC measure and 0.6506 results in the DSC measure. Furthermore, FAT-Net (Wu et al., 2022) achieved second-best results, 26.3024 in the HD95 performance measure. The worst-case results are observed in the case of TransUNet (Chen et al., 2021) JSC (0.2530), DSC (0.3380), and HD<sup>95</sup> (36.76), respectively. In the DEN-based dataset, the proposed network achieved a 2.87% improvement in JSC measure, 2.42% improvement in DSC measure, and achieved 34% lower HD<sup>95</sup> scores as compared with competitor UNet (Ronneberger et al., 2015). The second best results are observed in case of MSU-Net (Su et al., 2021) 0.7597 results in the JSC measure, 0.8441 results in the DSC measure, and 21.3417 in HD<sup>95</sup> measure. The worst-case results are also observed in the case of TransUNet (Chen et al., 2021) JSC (0.5977), and DSC (0.7058), respectively.

Table 6 indicate the performances of PDF-UNet with other state-of-

the-art methods on the SIIT ultrasound dataset. We can see that our network outperforms all the networks compared by an appreciable margin in all the metrics. In the original dataset, MSU-Net (Su et al., 2021) achieved a 1.68% improvement in the JSC measure, and DeepLabv3+ (Chen et al., 2018) achieved 1.59% improvement in the DSC measure as compared with competitor UNet (Ronneberger et al., 2015) network. In addition, SK-U-Net (Byra et al., 2020) achieved 23.1032 HD<sup>95</sup> scores as compared with traditional UNet (Ronneberger et al., 2015). The second-best results are observed in the case of UNet (Ronneberger et al., 2015) 0.7498 results in the JSC measure, 0.8397 results in the DSC measure, and 30.4592 HD<sup>95</sup> scores, respectively. The proposed PDF-UNet obtained the third position in all performance measures with JSC (0.7386), DSC (0.8260), HD<sup>95</sup> (20.4591), respectively. The worst-case results are observed in the case of TransUNet (Chen et al., 2021) JSC (0.5550), DSC (0.6568), and HD95 (31.6160), respectively. In the DEN-based dataset, the proposed network (PDF-UNet) achieved a 1.78% improvement in JSC measure, 1.52% improvement in DSC measure, and achieved 23% lower HD<sup>95</sup> scores as compared with competitors UNet (Ronneberger et al., 2015). The second best results are observed in the case of DeepLabv3+ (Chen et al., 2018) 0.7790 results in JSC measure and 0.8653 results in DSC measure. Furthermore, SK-U-Net (Byra et al., 2020) achieved second-best results, 22.062 in the HD<sup>95</sup> performance measure. The worst-case results are observed in the case of TransUNet (Chen et al., 2021) JSC (0.6365), DSC (0.7250), and HD<sup>95</sup> (28.4800), respectively.

In Table 7, quantitative comparison is also performed with non-deep learning methods such as MSGC (Zhou et al., 2014), DRLSE (Chunming Li et al., 2010), PAORGB (Huang et al., 2014), MOORGB (Luo et al., 2017), and (Huang et al., 2020). The proposed method has achieved the highest results in the term in F1-Score (0.940), TP (0.950), FP (0.076). The second-best results were obtained by (Huang et al., 2020), where the highest results are F1-Score (0.898), TP(0.914), FP(0.122).



**Fig. 9.** The visual effectiveness of BL, PDF, BCE loss, DICE loss, and FTL loss function. The red contour represents the actual ground truth, and the yellow region represents the predicted results.

**Table 5**

Quantitative comparison of PDF-UNet with other state-of-the-art methods on Mendeley ultrasound dataset.

Method	Results with the original dataset			Results with addition of DEN-based synthetic dataset		
	JSC ↑	DSC ↑	HD <sup>95</sup> ↓	JSC ↑	DSC ↑	HD <sup>95</sup> ↓
UNet (Ronneberger et al., 2015)	Mean	0.5186	0.6331	39.849	0.7378	0.8268
	StdDev	0.0179	0.0210	0.4000	0.0081	0.0081
UNet++ (Zhou et al., 2018)	Mean	0.4982	0.6175	42.3485	0.6767	0.7748
	StdDev	0.0186	0.0210	0.4600	0.0282	0.0299
DeepLabv3+ (Chen et al., 2018)	Mean	0.5194	0.6506	35.9891	0.7442	0.8365
	StdDev	0.0161	0.0147	1.6600	0.0305	0.0267
SK-U-Net (Byra et al., 2020)	Mean	0.4966	0.6476	30.1730	0.7184	0.8345
	StdDev	0.1014	0.0889	1.8900	0.0629	0.0441
MSU-Net (Su et al., 2021)	Mean	0.4460	0.5567	32.1830	0.7597	0.8441
	StdDev	0.0373	0.0426	2.9900	0.0104	0.0100
TransUNet (Chen et al., 2021)	Mean	0.2530	0.3380	36.7650	0.5977	0.7058
	StdDev	0.0730	0.0810	1.3400	0.0548	0.0553
TransFuse (Zhang et al., 2021)	Mean	0.5130	0.6290	35.8050	0.7229	0.8241
	StdDev	0.0320	0.0320	1.7000	0.0198	0.0149
FAT-Net (Wu et al., 2022)	Mean	0.5080	0.6380	26.3020	0.7186	0.8142
	StdDev	0.0180	0.0180	1.8000	0.0118	0.0090
PDF-UNet (our)	Mean	<b>0.5950</b>	<b>0.7177</b>	<b>22.4332</b>	<b>0.7665</b>	<b>0.8510</b>
	StdDev	0.0267	0.0252	2.2900	0.0125	0.0122

However, MSGC (Zhou et al., 2014) method achieved the worst results as compared to all other methods. The comparative results with the non-deep learning methods have validated that our PDF-UNet is competitive and shows better results than conventional methods.

#### 4.5.3. Computational comparison

Table 8 reflect a computational comparison of PDF-UNet with other state-of-the-art networks. The network size of the trainable parameter is 29 million and requires 40 MB space, making it more convenient to deploy in mobile devices to provide an accurate and fast assessment of breast tumors. Furthermore, PDF-UNet training time is 29 m with inference time of 2.15 s and mean FPS (33.20), respectively. The ideal parameters, model size, training time, inference, and mean FPS is observed in case of UNet (Ronneberger et al., 2015). Similarly, UNet++ (Zhou et al., 2018) are also using a small range of parameters, training time, inference, and FPS rate than other networks. It should be noted

that UNet (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018) are designed on a simple encoder-decoder model and support fixed convolution size with no attention or pyramid dilation mechanism. However, PDF-Net also required fewer parameters as compared with TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021), FAT-Net (Wu et al., 2022), DeepLabv3+ (Chen et al., 2018), respectively. Our network required 40 MB model size which is 10 Mb larger than UNet (Ronneberger et al., 2015) network but comparatively lesser than all other state-of-the-art networks. The network training time is greater than UNet (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018), but lesser than TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021), FAT-Net (Wu et al., 2022), and DeepLabv3+ (Chen et al., 2018). The inference time and mean FPS is 2.15 ms and 33.20, slightly better than TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021), FAT-Net (Wu et al., 2022), and DeepLabv3+ (Chen et al., 2018).

**Table 6**

Quantitative comparison of PDF-UNet with other state-of-the-art methods on SIIT ultrasound dataset.

Method		Results with the original dataset			Results with addition of DEN based synthetic dataset		
		JSC $\uparrow$	DSC $\uparrow$	HD <sup>95</sup> $\downarrow$	JSC $\uparrow$	DSC $\uparrow$	HD <sup>95</sup> $\downarrow$
UNet (Ronneberger et al., 2015)	Mean	<b>0.7498</b>	<b>0.8397</b>	<b>30.4592</b>	0.7679	0.8511	27.436
	StdDev	0.0091	0.0081	2.5900	0.0047	0.0065	3.4100
UNet++ (Zhou et al., 2018)	Mean	0.6606	0.7725	32.4322	0.7616	0.8470	26.448
	StdDev	0.0227	0.0192	3.5900	0.0130	0.0089	2.4100
DeepLabv3+ (Chen et al., 2018)	Mean	0.7614	<b>0.8556</b>	36.7270	<b>0.7790</b>	<b>0.8653</b>	35.960
	StdDev	0.0154	0.0115	0.7200	0.0079	0.0065	3.8800
SK-U-Net (Byra et al., 2020)	Mean	0.7330	0.8450	<b>23.1032</b>	0.7605	0.8636	<b>22.0620</b>
	StdDev	0.0484	0.0319	1.3700	0.0322	0.0208	1.6400
MSU-Net (Su et al., 2021)	Mean	<b>0.7666</b>	0.8495	24.1032	0.7725	0.8632	26.0720
	StdDev	0.0056	0.0058	1.3400	0.0045	0.0034	2.1400
TransUNet (Chen et al., 2021)	Mean	0.5550	0.6560	31.6160	0.6365	0.7250	28.4800
	StdDev	0.0480	0.0480	0.0011	0.0615	0.0620	1.6121
TransFuse (Zhang et al., 2021)	Mean	0.7240	0.8230	30.9150	0.7306	0.8270	25.1140
	StdDev	0.0140	0.0120	1.3411	0.0196	0.0150	2.5511
FAT-Net (Wu et al., 2022)	Mean	0.6650	0.7630	29.3400	0.7291	0.8170	28.1210
	StdDev	0.0260	0.0220	1.9310	0.0203	0.0190	0.4201
PDF-UNet (our)	Mean	0.7386	0.8260	20.4591	<b>0.7857</b>	<b>0.8663</b>	<b>4.4081</b>
	StdDev	0.0266	0.0215	0.0000	0.0066	0.0055	0.3500

**Table 7**

Quantitative comparison of PDF-UNet with other non-deep learning methods on 100 testing BUS images provided by (Huang et al., 2020).

Method	F1-score	True Positive	False Positive
MSGC (Zhou et al., 2014)	$0.812 \pm 0.6270$	$0.715 \pm 0.8900$	$0.405 \pm 0.4780$
DRLSE (Li et al., 2010)	$0.860 \pm 0.4770$	$0.939 \pm 0.5280$	$0.249 \pm 0.1329$
PAORGB (Huang et al., 2014)	$0.788 \pm 0.1252$	$0.723 \pm 0.1708$	$0.931 \pm 0.1944$
MOORGB (Luo et al., 2017)	$0.896 \pm 0.5120$	$0.863 \pm 0.8370$	$0.604 \pm 0.9260$
(Huang et al., 2020)	$0.898 \pm 0.4050$	$0.914 \pm 0.5040$	$0.122 \pm 0.9420$
PDF-UNet (our)	<b><math>0.940 \pm 0.0006</math></b>	<b><math>0.950 \pm 0.0010</math></b>	<b><math>0.076 \pm 0.0002</math></b>

**Table 8**

The computational comparison of presented PDF-UNet with previous state-of-the-art methods.

Method	Parameters	Model Size	Training time	Inference time	Mean FPS
UNet (Ronneberger et al., 2015)	$7.76 \times 10^6$	30 MB	9 m 28 s	0.95 ms	52.37
UNet++ (Zhou et al., 2018)	$9.04 \times 10^6$	35 MB	21 m 55 s	1.16 ms	42.92
DeepLabv3+ (Chen et al., 2018)	$5.93 \times 10^7$	227 MB	30 m 14 s	2.75 ms	28.54
SK-U-Net (Byra et al., 2020)	$3.93 \times 10^6$	46 MB	16 m 30 s	0.08 ms	11.35
MSU-Net (Su et al., 2021)	$4.70 \times 10^7$	179 MB	56 m 54 s	3.70 ms	30.66
TransUNet (Chen et al., 2021)	$9.62 \times 10^7$	646 MB	98 m 10 s	7.10 ms	14.41
TransFuse (Zhang et al., 2021)	$2.61 \times 10^7$	84.2 MB	78 m 20 s	4.11 ms	22.90
FAT-Net (Wu et al., 2022)	$3.00 \times 10^7$	75.1 MB	41 m 28 s	2.81 ms	27.60
PDF-UNet (our)	$1.05 \times 10^7$	40 MB	29 m 18 s	2.15 ms	33.20

#### 4.6. Visual comparative results

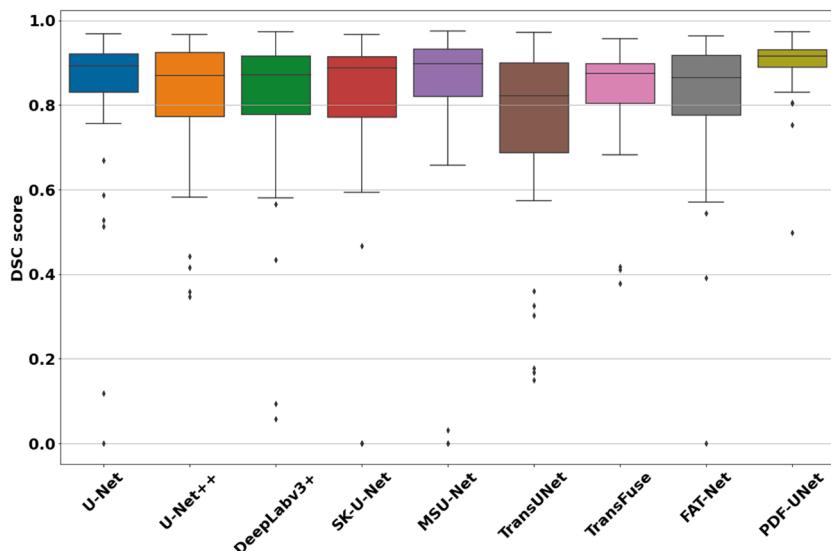
The proposed PDF-UNet performance is compared with previous state-of-the-art methods qualitatively. Fig. 10. box plots show the DSC score obtained for 50 testing images of breast tumor segmentation. Comparatively, PDF-UNet (Ronneberger et al., 2015) provides the best performance in segmenting breast tumor regions with 0.972 highest and 0.498 lowest DSC scores. The competitor UNet (Ronneberger et al., 2015) network achieved the highest 0.968 and 0 lowest DSC scores, and six different outliers were observed in the case of the UNet (Ronneberger et al., 2015) network. The worst-case results are observed with the TransUNet (Chen et al., 2021) network with a 0.7058 mean DSC score and six multiple outliers.

Fig. 11. box plots depict the DSC score achieved for 50 testing images of breast tumor segmentation. The PDF-UNet comparatively shows the best performance in segmenting breast tumor regions, with 0.970 highest and 0.555 lowest DSC scores. The competitor UNet (Ronneberger et al., 2015) network achieved the highest 0.968 and 0 lowest DSC scores, and four different outliers were observed in the case of UNet

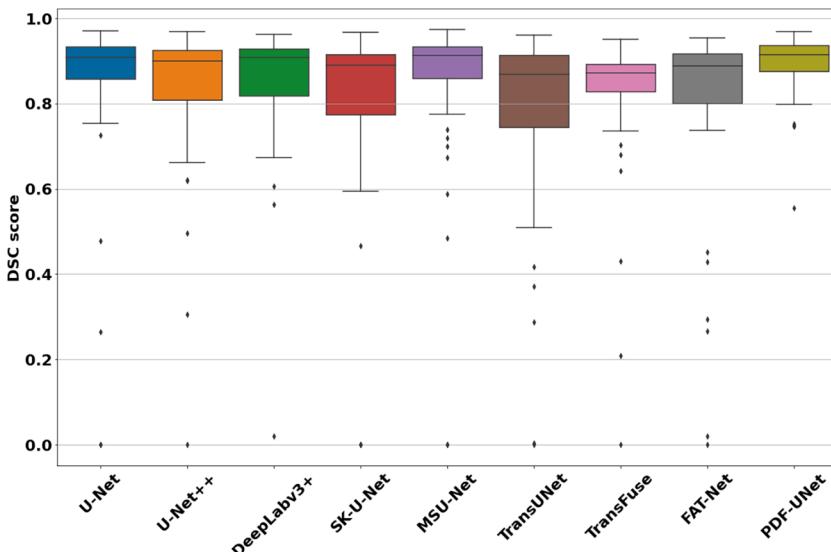
(Ronneberger et al., 2015) network. The worst-case results are observed with TransUNet (Chen et al., 2021) network with a 0.7058 mean DSC score and four outliers.

Fig. 12. reflect the Hausdorff Distance (HD<sup>95</sup>) results by PDF-UNet and others state-of-the-art methods. The dark grey line shows that the presented network obtained the lowest HD<sup>95</sup> (mm) scores on all datasets. We can observe that the presented network HD<sup>95</sup> (mm) values < 22 (mm). SK-U-Net (Byra et al., 2020) achieved the second-best result on Mendeley (Original) dataset. Similarly, SK-U-Net (Byra et al., 2020) obtained second-lowest HD<sup>95</sup> results on Mendeley (DEN), SITT (original), and SITT (DEN) synthetic datasets, respectively. The MSU-Net (Su et al., 2021) obtained the worst results on Mendeley (original), and UNet (Ronneberger et al., 2015) achieved the worst result on Mendeley (DEN) synthetic dataset. Furthermore, DeepLabv3+(Chen et al., 2018) obtained the worst result with the highest HD<sup>95</sup> scores on both SITT (original), and SITT (DEN) synthetic datasets.

Some challenging visual results are depicted in Fig. 13. It can be seen in the Mendeley dataset that our proposed PDF-UNet more accurately segmented tumor regions as compared with other networks. In 1<sup>st</sup>



**Fig. 10.** The Mendeley dataset breast tumor segmentation results are illustrated as boxplots. The y-axis shows the DSC scores, while the x-axis illustrates the multiple networks. The small dot symbols denote the suspected outliers.



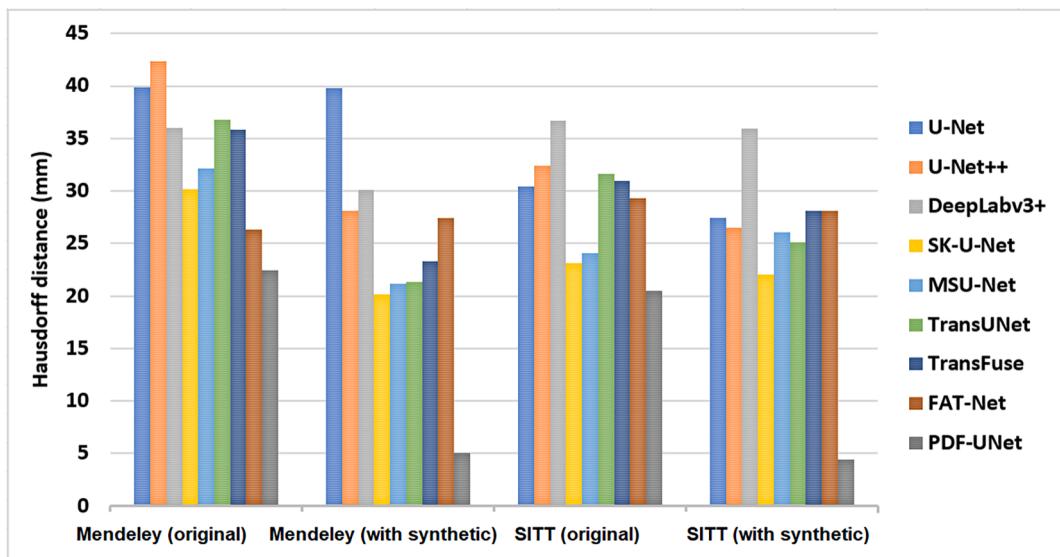
**Fig. 11.** The SIIT dataset breast tumor segmentation results are depicted as boxplots. The y-axis shows the DSC scores, while the x-axis depicts the multiple networks. The small dot symbols denote the suspected outliers.

image, except for PDF-UNet, other networks do not perform well and face under segmenting challenge. In the 2<sup>nd</sup> image, U-Net (Ronneberger et al., 2015) and FAT-Net (Wu et al., 2022) slightly perform well as compared with UNet++ (Zhou et al., 2018), TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021) and DeepLabv3+ (Chen et al., 2018) networks. In the 3<sup>rd</sup> image, second good results are observed in the case of TransFuse (Zhang et al., 2021), but an over-segmentation problem is observed in all other networks. In the 4<sup>th</sup> of SITT image, our network performs with smaller gaps in prediction and ground truth results. In the 5<sup>th</sup> image, our proposed network shows excellent results, but the worst prediction results are observed in the case of MSU-Net (Su et al., 2021). Furthermore, MSU-Net (Su et al., 2021) over segmenting tumor regions with large marginal gaps. The 6<sup>th</sup> image, does not seem challenging for any network, and all network results show average results in the case of prediction compared to ground truth.

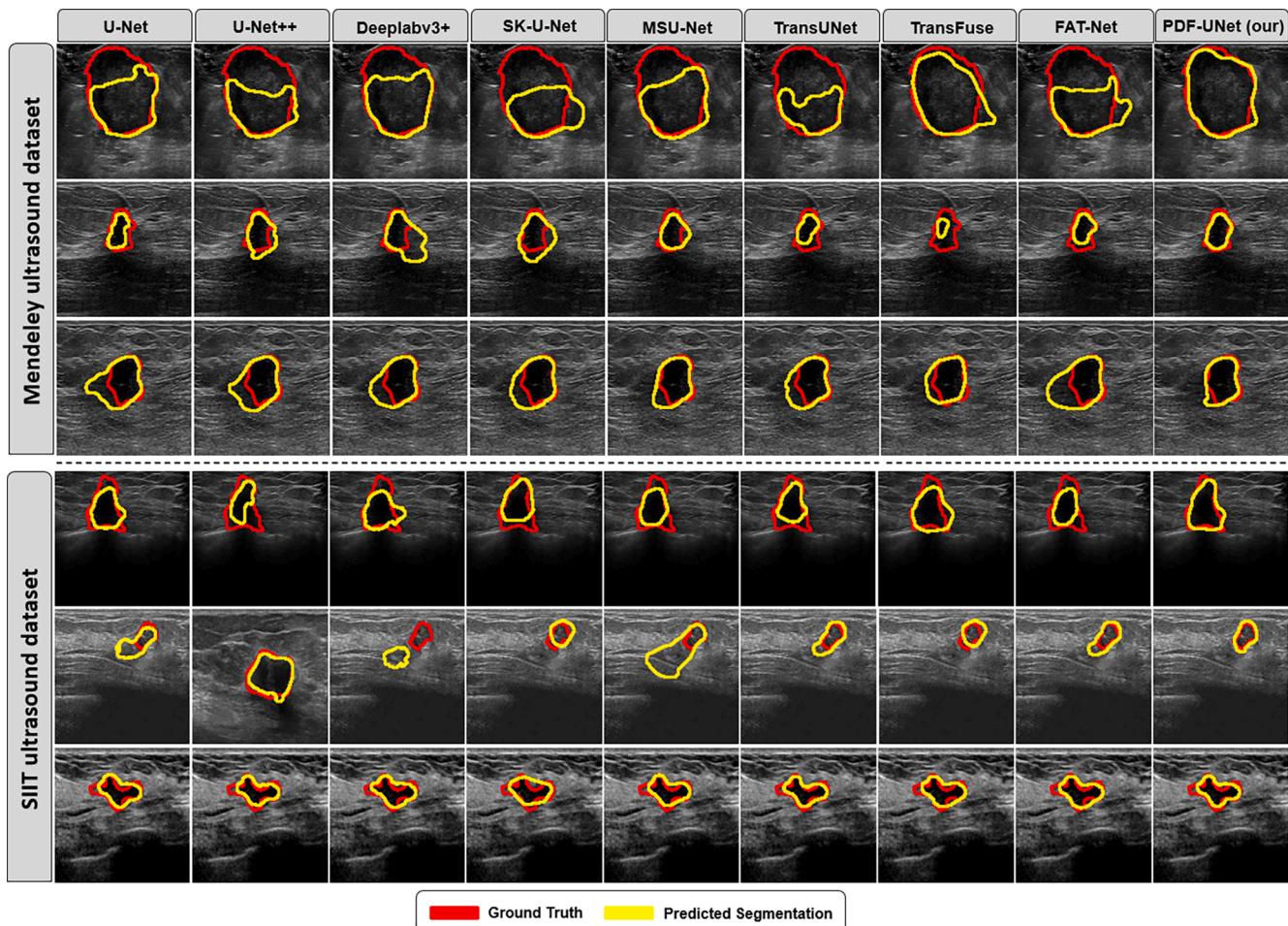
## 5. Discussion

Breast tumor segmentation is a crucial procedure for early diagnosis and treatment planning. Developing an effective CAD system is difficult due to tumor size, texture, and ambiguous boundaries of tumor regions. For a breast tumor segmentation task, a few targets of tumors may have a very large variation of position, texture, shape, and scale, network awareness of the spatial size and position for accurate segmentation is a more important task for any network. Furthermore, the limited availability of breast tumor annotated datasets are another challenge for proper training of deep learning networks. Third, traditional UNet architecture is based on the encoder-decoder mechanism due to fixed receptive fields and more restrictive convolutional operation experience vanishing gradient problems with different optimization issues.

In this work, a novel method is presented with three different networks, such as DEN network, PMG network, and PDF-UNet are introduced for breast tumor segmentation with the highest accuracy. The DEN network is trained on unannotated breast tumor images and



**Fig. 12.** The results of HD<sup>95</sup> scores on Mendeley (original), Mendeley (with synthetic), SIIT (original), and SIIT (with synthetic) datasets. The corresponding HD<sup>95</sup> scores are shown at the left axis of the plot.



**Fig. 13.** The visualization of the segmentation results on two different breast tumors Mendeley and SIIT datasets. Three sample images are shown from the left to right competitor and our proposed PDF-UNet. The solid red contour represents the ground truth regions, and the yellow contour draws the predicted segmentation result.

generates synthetic images for the data expansion process. The PMG network is inspired by DeepLabv3+ (Chen et al., 2018) architecture to generate corresponding probability maps using unannotated breast ultrasound images. Finally, the third PDF-UNet is a modified variant of classical UNet (Ronneberger et al., 2015) with some necessary modifications in down-sampling, up-sampling path, and embedding PDF block for better extraction features from small and larger size tumor regions. The common usage of average-pooling with max-pooling decreases the information loss in down-sampling, and bridge layers are used to reduce the semantic gaps of features in the encoder and decoder path. Specifically, a Pyramid-dilated fusion block (PDF) is embedded in the bottleneck of the encoder and decoder path with different dilation rates to extract multiscale features from small and large-size breast tumors.

The ablation study shows that using PDF block in our network achieved an improvement of DSC (4.2%) on Mendeley (DEN) dataset and DSC (2.65%) on SITT (DEN) dataset. Similarly, we have observed a reduction in HD<sup>95</sup> (50%) on Mendeley (DEN) dataset, and HD<sup>95</sup> (25%) on SITT (DEN) dataset, respectively. The experimental results demonstrate that the proposed (PDF-UNet) provides an outstanding DSC (0.8510) Mendeley dataset and DSC (0.8663) SITT ultrasound dataset. In Table 5, Table 6, we have compared our PDF-UNet results with other state-of-the-art methods on Mendeley and SITT ultrasound dataset. Similarly, In Table 5, PDF-UNet results are compared with few state-of-the-art non-deep learning methods on 100 BUS test images. In Table 8, the computational efficiency of different state-of-the-art networks are compared with our proposed PDF-UNet. The comparative results demonstrate that the presented network is parameters size, and model size is better than TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021), FAT-Net (Wu et al., 2022), DeepLabv3+ (Chen et al., 2018), SK-U-Net (Byra et al., 2020) and MSU-Net (Su et al., 2021). Similarly, the proposed PDF-UNet only requires 29 m for the training process, which is slightly better than TransUNet (Chen et al., 2021), FAT-Net (Wu et al., 2022), DeepLabv3+(Chen et al., 2018), MSU-Net (Su et al., 2021). The inference time and mean FPS is comparatively better than TransUNet (Chen et al., 2021), TransFuse (Zhang et al., 2021), FAT-Net (Wu et al., 2022), DeepLabv3+(Chen et al., 2018), SK-U-Net (Byra et al., 2020), and MSU-Net (Su et al., 2021), respectively.

Fig. 10 and Fig. 11. depict that our PDF-UNet overall performance on Mendeley and SITT datasets is comparatively better than other state-of-the-art network. Fig. 12 shows our network HD<sup>95</sup> results are slightly better on all four datasets with the lowest HD<sup>95</sup> score. The proposed method not only provides a better diagnosis for breast tumor segmentation but is also capable of fairly diagnosing breast cancer with mild tumors with the highest precision. The overall results demonstrate that the proposed method is not only effective for breast tumor segmentation tasks using ultrasound images but also require less computational cost as compared with other state-of-the-art methods. Therefore, the proposed scheme can be further optimized to reduce training and inference time and increase the mean FPS rate.

Our DEN network architecture is based on playing a min–max game between generator and discriminator network, improving performance with reducing adversarial losses; however, it is difficult to develop an effective network with balance capability for learning different variations and shapes of tumor images. The researcher can usually face non-convergence, mode collapse, diminished gradient; and highly hyper-parameter sensitivity during the training process of GANs inspired networks. The limitation of our work is that we are comparing our segmenting result with provided annotation from one medical expert. Although, annotation provided by multiple experts is a better approach for a better segmentation process. Furthermore, GANs inspired networks are capable of learning from prior knowledge, for example, the shape and different textures of breast tumors. It can be expected it learns the style of a single reader. For generating probability maps, we used the PMG network, which is inspired by DeepLabv3+ (Chen et al., 2018) architecture, but probability maps are estimated prediction results not comparatively equal to ground truth images. However, PMG network

training is a more important process to generate probability maps with the best estimates, and it requires a sufficient amount of annotated datasets to train the network.

## 6. Conclusion

In this research, we present a semi-supervised learning-based method that incorporates a DEN network, PMG network, and PDF-UNet for accurate breast tumor segmentation. To validate the proposed method, two publicly available Mendeley datasets and the SIIT dataset are used for experimental purposes. These datasets are highly complex and heterogeneous, resulting in a challenging task in realistic scenarios. The proposed PDF-UNet achieved a JSC (0.7665), DSC (0.8510), HD<sup>95</sup> (5.0392) on the Mendeley ultrasound dataset, and a JSC (0.7857), DSC (0.8663), HD<sup>95</sup> (4.4081) on the SIIT dataset. The experimental results show that the proposed method not only tackles the data limitations but also produces promising results for segmenting breast tumor ultrasound images. The primary limitation of our proposed DEN and PMG networks is that they are sensitive to filter size, stride size, and different hyper-parameters. Furthermore, our PMG network requires ResNet-101 as the backbone to produce accurate probability maps for synthetic images. The actual motive to propose a novel PDF-UNet is that compared to the PMG network, it's lightweight, requires fewer parameters, and is easily deployable for the CAD system. The other limitation of DEN and PMG networks is required a powerful GPU system and long training time. Overall, the results revealed that the proposed semi-supervised method could be embedded with a CAD system to provide complementary support to healthcare experts. In the future, our method can be tested against other modalities, such as Mammograms and MRI-based images, to validate the usefulness of our method.

## CRediT authorship contribution statement

**Ahmed Iqbal:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Muhammad Sharif:** Supervision, Project administration, Investigation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The author gratefully acknowledges Bahya Hospital, Cairo Egypt (Al-Dhabayani et al., 2020) and Thammasat University Hospital, Thailand (Karunananayake et al., 2020) for providing access to Ultrasounds datasets.

## References

- Abraham, N., & Khan, N. M. (2019). 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). <https://doi.org/10.1109/ISBI50293.2019.8759329>.
- Aja-Fernández, S., Curiale, A. H., & Vegas-Sánchez-Ferrero, G. (2015). A local fuzzy thresholding methodology for multiregion image segmentation. *Knowledge-Based Systems*, 83(1), 1–12. <https://doi.org/10.1016/j.knosys.2015.02.029>
- Al-Dhabayani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, 28, Article 104863. <https://doi.org/10.1016/j.dib.2019.104863>
- American Cancer Society. (2022). *American Cancer Society: Cancer Facts and Figures 2022*. Atlanta, Ga: American Cancer Society.
- Amjadian, A., & Gharaei, A. (2022). An integrated reliable five-level closed-loop supply chain with multi-stage products under quality control and green policies: Generalised outer approximation with exact penalty. *International Journal of Systems Science: Operations & Logistics*, 9(3), 429–449. <https://doi.org/10.1080/23302674.2021.1919336>

- Baradaran Rezaei, H., Amjadian, A., Sebt, M. V., Askari, R., & Gharaei, A. (2022). An ensemble method of the machine learning to prognosticate the gastric cancer. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04964-1>
- Boukerroui, D., Basset, O., Guérin, N., & Baskurt, A. (1998). Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation. *European Journal of Ultrasound*, 8(2), 135–144. [https://doi.org/10.1016/S0929-8266\(98\)00062-7](https://doi.org/10.1016/S0929-8266(98)00062-7)
- Byra, M., Jarosik, P., Szubert, A., Galperin, M., Ojeda-Fournier, H., Olson, L., ... Andre, M. (2020). Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomedical Signal Processing and Control*, 61, Article 102027. <https://doi.org/10.1016/j.bspc.2020.102027>
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. <http://arxiv.org/abs/2102.04306>.
- Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv Preprint*, 1–20. <http://arxiv.org/abs/1706.05587>.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Pertanika Journal of Tropical Agricultural Science*, 34(1), 137–143.
- Cheng, H. D., Shan, J., Ju, W., Guo, Y., & Zhang, L. (2010). Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43(1), 299–317. <https://doi.org/10.1016/j.patcog.2009.05.012>
- Dag, A. Z., Akcam, Z., Kibis, E., Simsek, S., & Delen, D. (2022). A probabilistic data analytics methodology based on Bayesian Belief network for predicting and understanding breast cancer survival. *Knowledge-Based Systems*, 242, Article 108407. <https://doi.org/10.1016/j.knosys.2022.108407>
- Daoud, M. I., Atallah, A. A., Awwad, F., Al-Najjar, M., & Alazrai, R. (2019). Automatic superpixel-based segmentation method for breast ultrasound images. *Expert Systems with Applications*, 121, 78–96. <https://doi.org/10.1016/j.eswa.2018.11.024>
- Díaz-Cortés, M. A., Ortega-Sánchez, N., Hinojosa, S., Oliva, D., Cuevas, E., Rojas, R., & Demin, A. (2018). A multi-level thresholding method for breast thermograms analysis using Dragonfly algorithm. *Infrared Physics and Technology*, 93(May), 346–361. <https://doi.org/10.1016/j.infrared.2018.08.007>
- Eisenbrey, J. R., Dave, J. K., & Forsberg, F. (2016). Recent technological advancements in breast ultrasound. *Ultrasonics*, 70, 183–190. <https://doi.org/10.1016/j.ultras.2016.04.021>
- Gharaei, A., Amjadian, A., & Shavandi, A. (2021). An integrated reliable four-level supply chain with multi-stage products under shortage and stochastic constraints. *International Journal of Systems Science: Operations & Logistics*, 1–22. <https://doi.org/10.1080/23302674.2021.1958023>
- Gómez, W., Leija, L., Alvarenga, A. V., Infantosi, A. F. C., & Pereira, W. C. A. (2010). Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. *Medical Physics*, 37(1), 82–95. <https://doi.org/10.1118/1.3265959>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, 2017-Decem, 4089–4099.
- Gu, P., Lee, W.-M., Roubidoux, M. A., Yuan, J., Wang, X., & Carson, P. L. (2016). Automated 3D ultrasound image segmentation to aid breast cancer image interpretation. *Ultrasonics*, 65, 51–58. <https://doi.org/10.1016/j.ultras.2015.10.023>
- Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, Ph. (1990). *A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform* (pp. 286–297). [https://doi.org/10.1007/978-3-642-75988-8\\_28](https://doi.org/10.1007/978-3-642-75988-8_28)
- Hu, K., Gao, X., & Li, F. (2011). Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. *IEEE Transactions on Instrumentation and Measurement*, 60(2), 462–472. <https://doi.org/10.1109/TIM.2010.2051060>
- Huang, C., Han, H., Yao, Q., Zhu, S., & Zhou, S. K. (2019). 3D U2-Net: A 3D Universal U-Net for Multi-Domain Medical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11765 LNCS(2018135), 291–299. <https://doi.org/10.1007/978-3-03-32245-8-33>
- Huang, Q., Bai, X., Li, Y., Jin, L., & Li, X. (2014). Optimized graph-based segmentation for ultrasound images. *Neurocomputing*, 129, 216–224. <https://doi.org/10.1016/j.neucom.2013.09.038>
- Huang, Q., Huang, Y., Luo, Y., Yuan, F., & Li, X. (2020). Segmentation of breast ultrasound image with semantic classification of superpixels. *Medical Image Analysis*, 61, Article 101657. <https://doi.org/10.1016/j.media.2020.101657>
- Iqbal, A., & Sharif, M. (2022). MDA-Net: Multiscale dual attention-based network for breast lesion segmentation using ultrasound images. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 7283–7299. <https://doi.org/10.1016/j.jksuci.2021.10.002>
- Iqbal, A., & Sharif, M. (2023). BTS-ST: Swin transformer network for segmentation and classification of multimodality breast cancer images. *Knowledge-Based Systems*, 110393. <https://doi.org/10.1016/j.knosys.2023.110393>
- Iqbal, A., Sharif, M., Khan, M. A., Nisar, W., & Alhaisoni, M. (2022). FF-UNet: A U-Shaped Deep Convolutional Neural Network for Multimodal Biomedical Image Segmentation. *Cognitive Computation*. <https://doi.org/10.1007/s12559-022-10038-y>
- Iqbal, A., Sharif, M., Yasmin, M., Raza, M., & Aftab, S. (2022). Generative adversarial networks and its applications in the biomedical image segmentation: A comprehensive survey. *International Journal of Multimedia Information Retrieval*, 11 (3), 333–368. <https://doi.org/10.1007/s13735-022-00240-x>
- Iqbal, A., Usman, M., & Ahmed, Z. (2022). An efficient deep learning-based framework for tuberculosis detection using chest X-ray images. *Tuberculosis*, 136, Article 102234. <https://doi.org/10.1016/j.tube.2022.102234>
- Iqbal, A., Usman, M., & Ahmed, Z. (2023). Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach. *Biomedical Signal Processing and Control*. <https://doi.org/10.1016/j.bspc.2023.104667>
- Karunanayake, N., Aimateer, P., Lohitvisate, W., & Makhanov, S. S. (2020). Particle method for segmentation of breast tumors in ultrasound images. *Mathematics and Computers in Simulation*, 170, 257–284. <https://doi.org/10.1016/j.matcom.2019.10.009>
- Li, C., Chenyang, X., Gui, C., & Fox, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12), 3243–3254. <https://doi.org/10.1109/TIP.2010.2069690>
- Liu, B., Cheng, H. D., Huang, J., Tian, J., Liu, J., & Tang, X. (2009). Automated segmentation of ultrasonic breast lesions using statistical texture classification and active contour based on probability distance. *Ultrasound in Medicine & Biology*, 35(8), 1309–1324. <https://doi.org/10.1016/j.ultrasmedbio.2008.12.007>
- Luo, Y., Liu, L., Huang, Q., & Li, X. (2017). A novel segmentation approach combining region- and edge-based information for ultrasound images. *BioMed Research International*, 2017(3), 1–18. <https://doi.org/10.1155/2017/9157341>
- McClymont, D., Mehnert, A., Trakic, A., Kennedy, D., & Crozier, S. (2014). Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph. *Journal of Magnetic Resonance Imaging*, 39(4), 795–804. <https://doi.org/10.1002/jmri.24229>
- Mughal, B., Muhammad, N., & Sharif, M. (2019). Adaptive hysteresis thresholding segmentation technique for localizing the breast masses in the curve stitching domain. *International Journal of Medical Informatics*, 126(June 2018), 26–34. <https://doi.org/10.1016/j.ijimedinf.2019.02.001>
- Mukherjee, S., Cheng, I., & Basu, A. (2018). *Atlas-Free Method of Periventricular Hemorrhage Detection from Preterm Infants' T1 MR Images* (pp. 157–168). <https://doi.org/10.1007/978-3-03-04375-9-14>
- Negi, A., Raj, A. N. J., Nersisson, R., Zhuang, Z., & Murugappan, M. (2020). RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks. *Arabian Journal for Science and Engineering*, 45(8), 6399–6410. <https://doi.org/10.1007/s13369-020-04480-z>
- Oktay, O., Schlemper, J., le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., ... Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv Midl*. <https://arxiv.org/abs/1804.03999>.
- Paulo Sergio Rodrigues. (2017). *Breast Ultrasound Image*. 2017. <https://doi.org/10.17632/wmy84gzngw.1>
- Prabusankarla, K. M., Thirumoolthy, P., & Manavalan, R. (2016). Segmentation of breast lesions in ultrasound images through multiresolution analysis using undecimated discrete wavelet transform. *Ultrasonic Imaging*, 38(6), 384–402. <https://doi.org/10.1177/0161734615615838>
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 1–16.
- Rodtook, A., & Makhanov, S. S. (2013). Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer. *Journal of Visual Communication and Image Representation*, 24(8), 1414–1430. <https://doi.org/10.1016/j.jvcir.2013.09.009>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Sadat, T., Hussain, A., Munir, A., Habib, M., Ali Khan, S., Hussain, S., ... Alawairdi, M. (2020). Identification of Breast Malignancy by Marker-Controlled Watershed Transformation and Hybrid Feature Set for Healthcare. *Applied Sciences*, 10(6), 1900. <https://doi.org/10.3390/app10061900>
- Salehi, S. S. M., Erdogan, D., & Gholipour, A. (2017). *Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks* (pp. 379–387). [https://doi.org/10.1007/978-3-319-67389-9\\_44](https://doi.org/10.1007/978-3-319-67389-9_44)
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- Shen, X., Ma, H., Liu, R., Li, H., He, J., & Wu, X. (2021). Lesion segmentation in breast ultrasound images using the optimized marked watershed method. *BioMedical Engineering OnLine*, 20(1), 57. <https://doi.org/10.1186/s12938-021-00891-7>
- Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., ... Torrents-Barrena, J. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Systems with Applications*, 139, Article 112855. <https://doi.org/10.1016/j.eswa.2019.112855>
- Sinha, A., & Dolz, J. (2020). Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 25(April), 1–20. <https://doi.org/10.1109/JBHI.2020.2986926>
- Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation. *Frontiers in Genetics*, 12(February), 1–14. <https://doi.org/10.3389/fgene.2021.639930>
- Vijayalakshmi, S., & Savita.. (2019). Image-Guided Surgery Through Internet of Things. In *Internet of Things in Biomedical Engineering* (pp. 75–116). Elsevier. <https://doi.org/10.1016/B978-0-12-817356-5.00005-X>
- Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., & Tang, X. (2017). Residual attention network for image classification. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua(1)*, 6450–6458. <https://doi.org/10.1109/CVPR.2017.683>
- Wu, H., Chen, S., Chen, G., Wang, W., Lei, B., & Wen, Z. (2022). FATT-Net: Feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis*, 76, Article 102327. <https://doi.org/10.1016/j.media.2021.102327>

- Xian, M., Zhang, Y., & Cheng, H. D. (2015). Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. *Pattern Recognition*, 48(2), 485–497. <https://doi.org/10.1016/j.patcog.2014.07.026>
- Xian, M., Zhang, Y., Cheng, H. D., Xu, F., Zhang, B., & Ding, J. (2018). Automatic breast ultrasound image segmentation : A survey. *Pattern Recognition*, 79, 340–355. <https://doi.org/10.1016/j.patcog.2018.02.012>
- Ma, X. (2017). Automated fibroglandular tissue segmentation in breast MRI using generative adversarial networks. *Physics in Medicine & Biology*, 65(10), 1–31. <https://doi.org/10.1088/1361-6560/ab7e7f>
- Xing, J., Li, Z., Wang, B., Qi, Y., Yu, B., Zanjani, F. G., ... Tan, T. (2021). Lesion Segmentation in Ultrasound Using Semi-Pixel-Wise Cycle Generative Adversarial Nets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6), 2555–2565. <https://doi.org/10.1109/TCBB.2020.2978470>
- Xu, S., Liu, H., & Song, E. (2011). Marker-Controlled Watershed for Lesion Segmentation in Mammograms. *Journal of Digital Imaging*, 24(5), 754–763. <https://doi.org/10.1007/s10278-011-9365-2>
- Yang, H., & Yang, D. (2023). CSwin-PNet: A CNN-Swin Transformer combined pyramid network for breast lesion segmentation in ultrasound images. *Expert Systems with Applications*, 213, Article 119024. <https://doi.org/10.1016/j.eswa.2022.119024>
- Huang, Yu-Len, & Chen, Dar-Ren (2005). Automatic Contouring for Breast Tumors in 2-D Sonography. In . 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 7 VOLS (pp. 3225–3228). <https://doi.org/10.1109/IEMBS.2005.1617163>.
- Zhang, Y., Liu, H., & Hu, Q. (2021). *TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation*. <https://doi.org/2102.08005>.
- Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D. P., & Chen, D. Z. (2017). Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10435 LNCS (pp. 408–416). [https://doi.org/10.1007/978-3-319-66179-7\\_47](https://doi.org/10.1007/978-3-319-66179-7_47).
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-net architecture for medical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11045 LNCS (pp. 3–11). [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1).
- Zhou, Z., Wu, W., Wu, S., Tsui, P. H., Lin, C. C., Zhang, L., & Wang, T. (2014). Semi-automatic breast ultrasound image segmentation based on mean shift and graph cuts. *Ultrasonic Imaging*, 36(4), 256–276. <https://doi.org/10.1177/0161734614524735>