Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Global guidance network for breast lesion segmentation in ultrasound images

Cheng Xue[a], Lei Zhu[c,1,*], Huazhu Fu[d], Xiaowei Hu[a], Xiaomeng Li[b], Hai Zhang[e,1], Pheng-Ann Heng[a,f]

[a] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
[b] Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[c] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
[d] Inception Institute of Artificial Intelligence, Abu Dhabi, UAE
[e] Shenzhen People's Hospital, The Second Clinical College of Jinan University, The First Affiliated Hospital of Southern University of Science and Technology, Guangdong Province, China
[f] Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 5 May 2020
Revised 28 January 2021
Accepted 29 January 2021
Available online 4 February 2021

Keywords:
Non-local features
Breast lesion segmentation
Deep neural network

ABSTRACT

Automatic breast lesion segmentation in ultrasound helps to diagnose breast cancer, which is one of the dreadful diseases that affect women globally. Segmenting breast regions accurately from ultrasound image is a challenging task due to the inherent speckle artifacts, blurry breast lesion boundaries, and inhomogeneous intensity distributions inside the breast lesion regions. Recently, convolutional neural networks (CNNs) have demonstrated remarkable results in medical image segmentation tasks. However, the convolutional operations in a CNN often focus on local regions, which suffer from limited capabilities in capturing long-range dependencies of the input ultrasound image, resulting in degraded breast lesion segmentation accuracy. In this paper, we develop a deep convolutional neural network equipped with a global guidance block (GGB) and breast lesion boundary detection (BD) modules for boosting the breast ultrasound lesion segmentation. The GGB utilizes the multi-layer integrated feature map as a guidance information to learn the long-range non-local dependencies from both spatial and channel domains. The BD modules learn additional breast lesion boundary map to enhance the boundary quality of a segmentation result refinement. Experimental results on a public dataset and a collected dataset show that our network outperforms other medical image segmentation methods and the recent semantic segmentation methods on breast ultrasound lesion segmentation. Moreover, we also show the application of our network on the ultrasound prostate segmentation, in which our method better identifies prostate regions than state-of-the-art networks.

1. Introduction

Breast cancer is one of the dreadful diseases that affect women globally. According to the statistic information reported in American Cancer Society (2019), an estimated 42,260 breast cancer deaths would occur in 2019. An accurate breast lesion segmentation from the ultrasound images helps the early diagnosis of breast cancer. However, the automatic breast lesion segmentation in a 2D ultrasound image is a challenging task, since there are the speckle noise, and strong shadows in the ultrasound, inhomogeneous distributions in the breast lesion regions, and ambiguous boundaries between the breast lesion and non-lesion regions, as well as the irregular breast lesion shapes; see Fig. 1 for the examples.

Segmenting breast lesion in ultrasound images has been widely studied in the research community. Early attempts, e.g., (Shan et al., 2012; Madabhushi and Metaxas, 2002; Shan et al., 2008; Kwak et al., 2005; Madabhushi and Metaxas, 2003; Yezzi et al., 1997; Chen et al., 2002; Xian et al., 2015; Ashton and Parker, 1995; Boukerroui et al., 1998; Xiao et al., 2002) detected the breast lesion boundaries mainly based on the hand-crafted features. These features, however, have the limited feature representation ability, leading to misrecognize the breast lesions in a complex environment. Recently, the convolutional neural networks (CNNs) have achieved impressive progress on breast ultrasound segmentation task. For examples, Yap et al., adopted U-Net, FCN-AlexNet, and

[*] Corresponding author.
E-mail address: lz437@cam.ac.uk (L. Zhu).
[1] Lei Zhu and Hai Zhang are the co-corresponding author of this work.

https://doi.org/10.1016/j.media.2021.101989
1361-8415/© 2021 Elsevier B.V. All rights reserved.

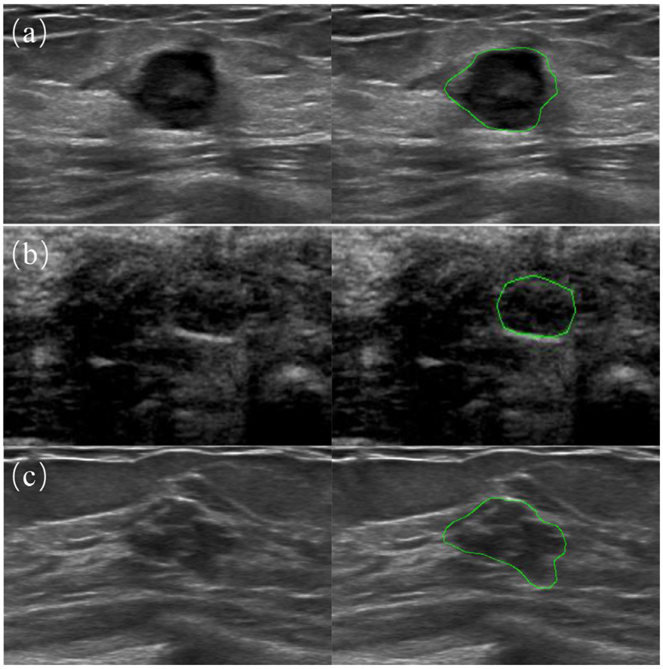


Fig. 1. Examples of challenging cases in breast ultrasound lesion segmentation. The green contour denotes the breast lesion boundary. Left: the input ultrasound images. Right: the lesion region. (a) Inhomogeneous distributions inside the breast lesion region. (b) Ambiguous boundary due to similar appearance between lesion regions and non-lesion backgrounds. (c) Irregular breast lesion shapes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

patch-based LeNet for 2D ultrasound image breast lesion detection (Yap et al., 2017). Lei et al., employed a deep neural network with the supervision signals on the boundary to address the whole breast ultrasound image (Lei et al., 2018). Xu et al., adopted an eight-layer CNN to segment 3D breast in the ultrasound data (Xu et al., 2019).

The ultrasound image has many distant pixels, which have the similar appearance as the breast lesions. Incorporating these pixels could provide long-term non-local features to learning discriminative features for the ultrasound breast lesion segmentation. Capturing the global contextual information for ultrasound image segmentation is a long-standing topic in the medical image community. Previous studies proposed to enlarge the receptive field with dilated convolutions, pooling operations (Chen et al., 2018; 2017b; 2014); or fuse the middle level and high level features with more task-related semantic features (Ronneberger et al., 2015; Lin et al., 2017). However, these methods fail to capture the contextual information in a global view and only consider the inter-dependencies among spatial domains. In medical image analysis community, most previous approaches rely on local region operation for segmentation task (Ronneberger et al., 2015; Dou et al., 2016; Lin et al., 2017). However, capturing the long-range dependencies information holds promising potentials but has not been well explored yet. Traditional non-local blocks in these networks (Qi et al., 2019; Dou et al., 2018) are only embedded into the deep CNN layers to learn long-range dependencies for network predictions. However, due to the relatively larger receptive fields than shallow CNN layers, the deep layers of a segmentation network are responsible for capturing cues of the whole breast lesions and somehow lack parts of breast lesion regions, degrading the segmentation performance.

In this work, we develop a convolutional neural network (CNN) to integrate features at all CNN layers (including deep and shallow CNN layers) to produce multi-level integrated features (MLIF) as a guidance information of the non-local blocks in spatial and channel manners to complement more breast lesion boundary details, which are usually neglected by deep CNN layers. Moreover, we propose to predict additional breast lesion boundary map such that the predicted boundary map is regularized to be as similar as the underlying ground truth. By doing so, our network can produce a segmentation result with more accurate breast lesion boundaries. In summary, our contributions are four-fold:

- First, we present a CNN (denoted as GG-Net) with a global guidance block (GGB) to aggregate non-local features in both spatial and channel domains under the guidance of multi-layer integrated features for learning a powerful non-local contextual information.
- Second, we develop a breast lesion boundary detection (BD) module in shallow CNN layers to embed additional boundary maps of breast lesions for obtaining the segmentation result with high-quality boundaries.
- Third, the experimental results on two ultrasound breast lesion datasets show that our network outperforms the state-of-the-art medical image segmentation methods on breast lesion segmentation.
- Moreover, we also show the application of our network on the ultrasound prostate segmentation, where our network obtains satisfactory performance.

2. Related works

Breast lesion segmentation from ultrasound images is very challenging due to the speckle artifacts, low contrast, shadows, blurry boundaries, and the variance in lesion shapes (Kirberger, 1995). A variety of breast lesion segmentation algorithms have been proposed and these methods can be broadly classified into four categories, including region based approach (Shan et al., 2012; Madabhushi and Metaxas, 2002; Shan et al., 2008; Kwak et al., 2005), deformable models (Madabhushi and Metaxas, 2003; Yezzi et al., 1997; Chen et al., 2002), graph-based approaches (Xian et al., 2015; Ashton and Parker, 1995; Boukerroui et al., 1998; Xiao et al., 2002) and learning based approaches (Liu et al., 2010; Huang et al., 2008; Lo et al., 2014; Moon et al., 2014; Othman and Tizhoosh, 2011). These approaches usually employed texture features to represent the local variation of pixel intensities and then detect abnormal regions in the ultrasound image. However, these methods rely on hand-crafted features and have limited representation capacity.

Convolutional neural networks (CNNs) have shown remarkable performance in many medical image analysis tasks, including image classification (Yu et al., 2018; 2017), semantic segmentation (Ronneberger et al., 2015; Dou et al., 2016; Yu et al., 2016; Li et al., 2018). These methods utilized the superior learning capability of neural network and outperformed other traditional segmentation methods. For breast image analysis, recent works have featured CNN based methods (Yap et al., 2017; Lei et al., 2018; Xu et al., 2019; Dhungel et al., 2017; Mordang et al., 2016a; Ahn et al., 2017; Mordang et al., 2016b; Hu et al., 2019; Mishra et al., 2018). Yap et al. adopted pacth-based LeNet, U-Net, and FCN-AlexNet for breast lesion detection Yap et al. (2017). Leiet al. proposed a ConvEDNet for whole breast ultrasound image segmentation with the deep boundary supervision and adaptive domain transfer knowledge (Lei et al., 2018). Some works adopted CNNs with different layers to detect mass, estimate the breast density, and segment breast ultrasound images (Dhungel et al., 2017; Ahn et al., 2017; Xu et al., 2019). Mordang et al., adopted OxfordNet for mammography microcalcification detection (Mordang et al., 2016b). Hu et al., proposed a dilated fully convolutional network for breast tumor segmentation (Hu et al., 2019). Mishra et al., developed a fully convolutional neural network with deep supervision for lu-

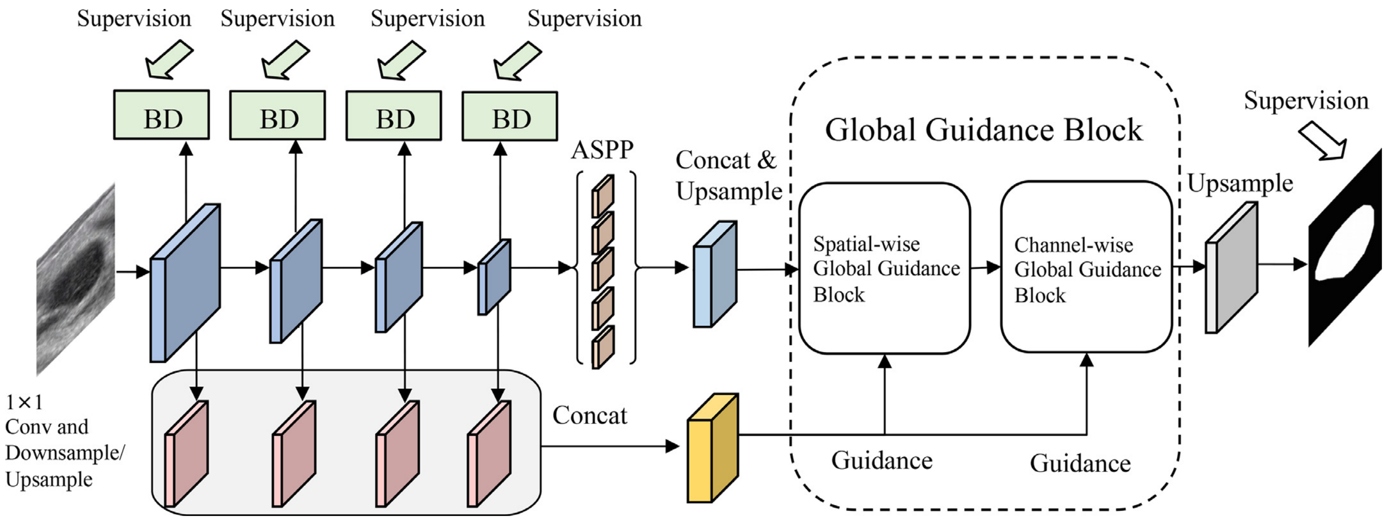


Fig. 2. The schematic illustration of the proposed breast lesion segmentation network (GG-Net) in this work. (i) We first use a convolutional neural network (CNN) to produce a set of feature maps with different scales, followed by a ASPP module to enlarge the receptive field. (ii) In each CNN layer, we pass its feature map to a breast lesion boundary detection (BD) module (see Section 3.2) to detect breast lesion boundaries. (iii) We concatenate features at all CNN layers and use it as the guidance to the developed global guidance block (GGB), which includes a spatial-wise global guidance block and a channel-wise global guidance block, to learn long-range dependencies for each pair of positions on the feature maps over spatial and channel domains. (iv) We use the output feature map of the GGB to predict the segmentation result of our network.

men segmentation and liver lesion segmentation (Mishra et al., 2018).

To improve the pixel-wise prediction accuracy, many researchers considered incorporating the long-range dependencies and contextual information in the network, thus enhancing the feature representation for pixel-wise prediction. For example, atrous spatial pyramid pooling (ASPP) was designed to embed the global contextual information, and it was widely adopted in DeepLabv2 (Chen et al., 2014) and DeepLabv3 (Chen et al., 2018). Similarly, Zhao et al., designed a pyramid pooling module to collect the effective contextual prior with different scales (Zhao et al., 2017). Besides, an EncNet was introduced a channel attention mechanism to capture the global context (Zhang et al., 2018). Peng et al., argued that large kernel plays an important role in semantic segmentation tasks, and a global convolutional network was proposed to learn the context information (Peng et al., 2017). In medical image analysis field, there are some recently work that also considered the context information, such as the encoder-decoder structures (Ronneberger et al., 2015) fused the mid-level and high-level features to obtain different scale context. In OBELISK-Net (Heinrich et al., 2019), sparse deformable convolutions were formulated to learn large context information. However, these methods mostly stacked a series of convolutional layers to capture the context information. Several works have been proposed to alleviate this issue by implicitly utilizing attention mechanisms or non-local operations to increase the receptive fields and capture contextual information (Wang et al., 2018a; Vaswani et al., 2017; Schlemper et al., 2019; Zhang et al., 2017; Roy et al., 2018; Joutard et al., 2019). However, the meticulous features in the multi-layer features and the long range dependencies between feature channels are ignored. In this regard, we introduce a network that gracefully unifies the approaches mentioned above, which not only consider the long-range dependencies spatial-wisely and channel-wisely, but also embed contextual information from different layers.

3. Methodology

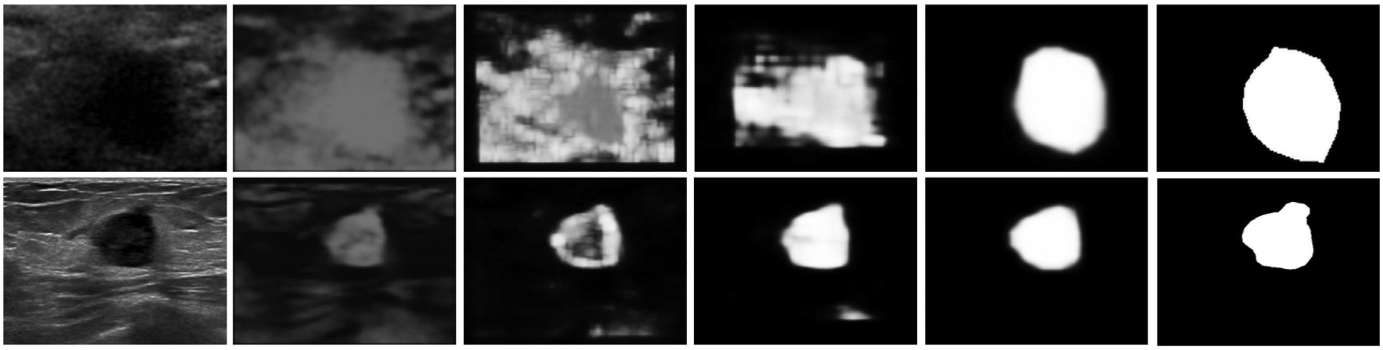
Fig. 2 illustrates the architecture of the developed network (denoted as GG-Net). Our network takes a breast ultrasound image as the input and produces a segmented mask in an end-to-end manner. Specifically, our GG-Net starts by using a CNN to generate multi-level feature maps with different spatial resolutions and adopting the ASPP (Chen et al., 2018) to enhance the receptive field of features. In order to utilize the complementary information among different CNN layers, the GGB is introduced to refine the features by learning long-range feature dependencies under the guidance of an integrated feature map from the shallow CNN layers. Moreover, the BD module is embedded in the shallow CNN layers to capture the breast lesion contour and provide a strong cue for better segmenting breast lesions and refining lesion boundaries. Finally, the prediction map is produced as the segmentation result of our network. In the following subsections, we will introduce details of the developed GGB and BD in our method.

3.1. Global guidance block

Convolutional and recurrent operations of CNNs only capture the spatial dependencies within a local neighborhood. Although stacking convolutional layers can learn the long range dependencies, such repeating local convolutions is time-consuming and leads to the optimization difficulties that need to be carefully addressed (Wang et al., 2018a). Moreover, breast ultrasound images usually contain speckles and shadows that tend to be recognized as breast lesion due to the limited receptive fields of local convolutions. In this regard, we develop a global guidance block (GGB), which leverages a guidance feature map to learn the long range dependencies by considering spatial and channel information.

3.1.1. Spatial-wise global guidance block

The feature maps from the shallow CNN layers provide detailed information but contain more non-lesion regions, while the deep CNN layers with larger reception fields eliminate the non-lesion regions, but tend to lose the local details. In this regard, we argue that feature maps at different CNN layers contain the complementary information, as shown in Fig. 3. In our method, we first resize the feature maps of the first four CNN layers to the size of feature map from the second CNN layer, and then concatenate them to one multi-layer integrated feature (MLIF) map. After that, a spatial-wise global guidance block (spatial-wise GGB) is proposed to learn



(a) inputs (b) 1st layer (c) 2nd layer (d) 3-th ayer (e) 4-th layer (f) ground truth

Fig. 3. Two examples are shown to illustrate the learned breast lesion feature on different layers. (a) Input images. (b)–(e) Segmentation maps predicted from the feature map from the 1st layer to the 4th layer. (f) Ground truths. The shallow layers (b), (c) and (d) contains more detail features compared to (e).

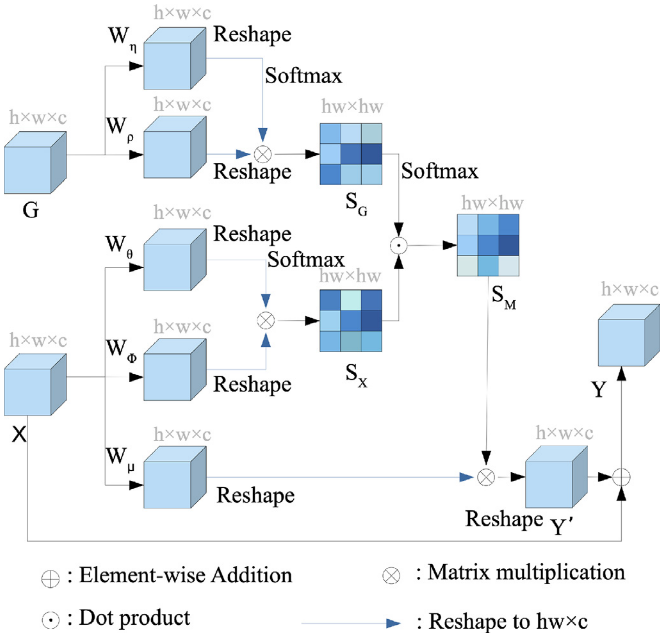


Fig. 4. The schematic illustration of the details of spatial-wise GGB, where G is the guidance map, and X is the input feature map.

the long-range position dependencies by taking MLIF as a guidance map.

Fig. 4 shows the schematic illustration of our spatial-wise GGB. Specifically, let X ($x \in \mathbb{R}^{h \times w \times c}$) denote the output feature map of the ASPP module (see Fig. 2), and G ($g \in \mathbb{R}^{h \times w \times c}$) denotes the guidance map. The spatial-wise GGB first feeds X into three 1×1 convolution layers with different parameters, $W_{\theta(x)}$, $W_{\phi(x)}$, and $W_{\mu(x)}$), to generate three feature maps, $\theta(x)$, $\phi(x)$, and $\mu(x)$, respectively. After that, we reshape $\theta(x)$, $\phi(x)$, and $\mu(x)$ as $\mathbb{R}^{hw \times c}$ matrices, multiply the reshaped $\phi(x)$ with the transpose of the reshaped $\theta(x)$, and apply a softmax layer on the multiplication result to compute a $hw \times hw$ spatial-wisely position similarity map S_x:

$$S_x = Softmax(X^T W_{\theta(x)}^T W_{\phi(x)} X) \qquad (1)$$

where $softmax$ follows the traditional sigmoid function and it is applied on each element of the $hw \times hw$ $X^T W_{\theta(x)}^T W_{\phi(x)} X$. On the other side, two 1×1 convolution layers with parameters, $W_{\eta(g)}$, and $W_{\rho(g)}$), are applied on guidance map G to obtain two feature maps, $\eta(x)$ and $\rho(x)$, reshape $\eta(x)$ and $\rho(x)$, multiply the reshaped $\eta(x)$ to the transpose of the reshaped $\rho(x)$, and apply a softmax layer for producing another $hw \times hw$ position similarity matrix (denoted as S_g) from the guidance map G:

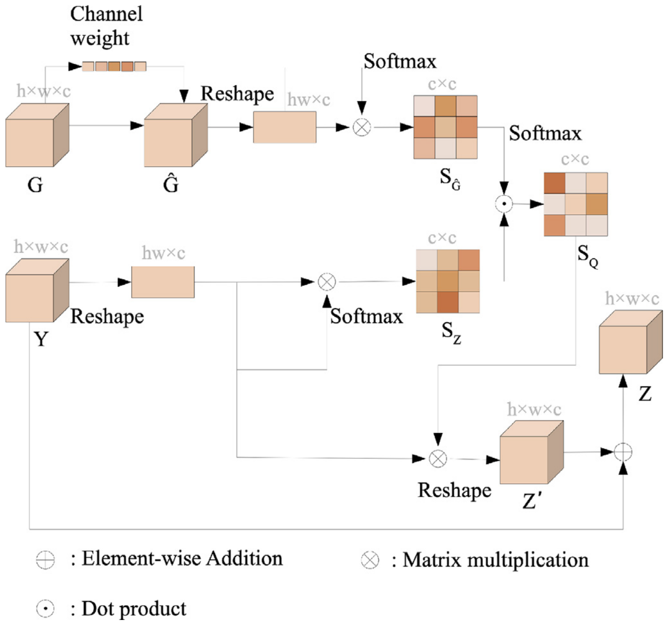


Fig. 5. The schematic illustration of the channel-wise GGB, where G is the guidance map and Y is the input feature map.

$$S_g = Softmax(G^T W_{\rho(g)}^T W_{\eta(g)} G)) . \qquad (2)$$

Once obtaining two similarity matrices S_X and S_G, we use a softmax layer on the element-wise multiplication result of S_X and S_G to generate a guided similarity matrix S_M. Then, we multiply S_M with the features $\mu(x)$ to obtain a new feature map Y', which is then added with the input features X to generate the output feature map Y:

$$Y = \mu(x)\ Softmax(S_x \cdot S_g) + X . \qquad (3)$$

3.1.2. Channel-wise global guidance block

Our spatial-wise GGB treats each feature channel equally when learning the long range dependencies, resulting in neglecting the correlations among different feature channels. Recently, allowing varied contributions from different feature channels has achieved superior performance in many computer vision tasks (Hou et al., 2019; Chen et al., 2017a; Hu et al., 2018). Motivated by these, we develop a channel-wise global guidance block (channel-wise GGB) to further learn the long range inter-dependencies between different feature channels. Fig. 5 illustrates the schematic details of the

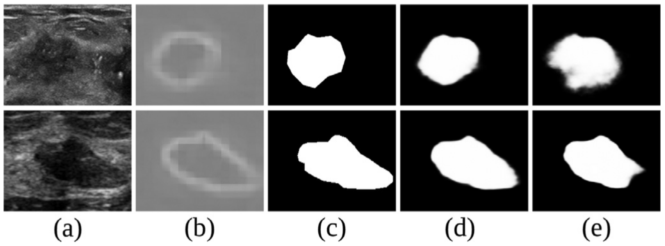


Fig. 6. An analysis of segmentation improvement based on detected boundaries. (a) Input images. (b) Detected boundary map at BD module at the fourth CNN layer. (c) Ground truths of breast lesion segmentation. (d) Segmentation results of our method. (e) Our results without the BD module. Apparently, learning additional boundary maps of breast lesion incurs a better segmentation result.

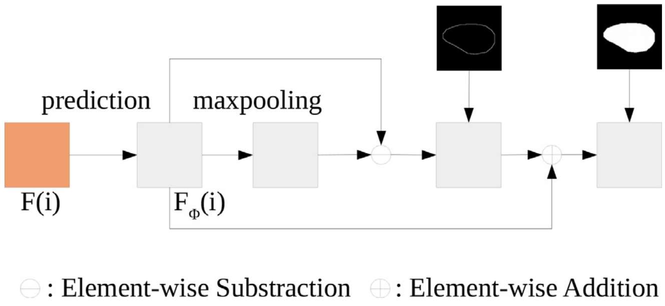


Fig. 7. The schematic illustration of the breast lesion boundary detection (BD) module. $F(i)$ is the feature map at the ith CNN layer.

proposed channel-wise GGB, which takes a feature map Y and a guidance map G as two inputs and generates a refined feature map Z. Specifically, we reshape Y to $\mathbb{R}^{c\times hw}$, multiply the reshaped Y and the transpose of the reshaped Y, and use a softmax layer to obtain a channel-wise similarity map $S_Z \in \mathbb{R}^{c\times c}$. Regarding the input guidance feature map G, we first use squeeze-and-excitation block to emphasis informative feature channels of G and suppress less useful ones. To achieve this, we use a global average pooling to generate the channel-wise statistics β, and the kth element of the descriptor (β) is given by

$$\beta_k = \frac{1}{h\times w}\sum_{i=1}^{h}\sum_{j=1}^{w} G(i,j,k) \tag{4}$$

where $G(i,j,k)$ denotes the element at the position (i,j,k) of guidance map G. After that, we use two fully connected (fc) layers and a sigmoid activation function on the channel-wise statistics β to generate a coefficient vector V_λ:

$$V_\lambda = \Phi(W_2\Omega(W_1\beta)) \tag{5}$$

where W_1 and W_2 denote the parameters of the two fully connected layers, Ω and Φ are the ReLU and the sigmoid activation function, respectively. Then, we multiply V_λ with G to assign different weights on channels of G and obtain a refined feature map (denoted as $\hat{G}$). Once obtaining $\hat{G}$, we reshape it to $\mathbb{R}^{c\times hw}$, multiple the reshaped $\hat{G}$ and the transpose of the reshaped $\hat{G}$, and use a softmax layer to generate a $c\times c$ similarity map $S_{\hat{G}}$. Later, a softmax layer is applied on the multiplication of S_Z and $S_{\hat{G}}$ to obtain a guided similarity map S_Q. Finally, we multiply S_Q with the input Y to produce a new feature map Z', which is then added to the input features Y to obtain the output feature map Z of our channel-wise GGB.

3.2. Breast lesion boundary detection module

Although GGB generates a breast lesion segmentation result, we find that there are many failed segmented regions in the results, as shown in Fig. 6(e), which have inaccurate boundary maps of the breast lesion. To alleviate this, we develop a breast lesion boundary detection (BD) module to identify multi-level boundary maps of the breast lesions and enhance the segmentation result with an additional boundary prediction loss. Fig. 7 shows the schematic illustration of the developed BD module at the i-th CNN layer to detect breast lesion boundaries. It takes the feature map of i-th CNN layer as the input and outputs a boundary map of the breast lesion and a breast lesion segmentation result. Specifically, we first use a 1×1 convolutional layer on the input features $F(i)$ to obtain a new feature map $F_\phi(i)$ with one channel. Then, we shift $F_\phi(i)$ with one pixel via a maxpooling operation (stride = 1, padding = 1, kernel size = 3×3; see (Feng et al., 2019) for details) and subtract the shifted result from $F_\phi(i)$ to obtain a boundary map E of the breast lesions. After that, we add $F_\phi(i)$ with E to obtain a breast lesion segmentation map.

3.3. Loss function

As shown in Fig. 2, we add a BD module for the shallow CNN layer to jointly locate breast lesions and detect a boundary map from feature map at the CNN layer. Hence, our network generates four boundary maps and four breast lesion segmentation results at four CNN layers. Moreover, our network generates a final segmentation result of breast lesions from the GGB. With an annotated breast lesion mask, we apply a canny operator (Canny, 1986) to obtain the boundary mask as the ground truth of the boundary prediction. Finally, we compute the total loss of our network as:

$$L_{total} = \sum_{i=1}^{N_{layer}}(\lambda_1\cdot L^i{}_{seg} + \lambda_2\cdot L^i{}_{boundary}) + L^f{}_{seg} \tag{6}$$

where N_{layer} is the number of CNN layers, and we empirically set N_{layer} as four in our implementation. $L^i{}_{seg}$ and $L^i{}_{boundary}$ denote the segmentation loss and the boundary loss in the BLBD module of i-th CNN layer, respectively. $L^f{}_{seg}$ is the loss function of the final segmentation result. The weights λ_1 and λ_2 are to balance $L^i{}_{seg}$, $L^i{}_{boundary}$, and $L^f{}_{seg}$, and their values are empirically set as $\lambda_1 = 1$ and $\lambda_2 = 10$.

Let $\mathcal{P}^i$ denote the predicted breast lesion segmentation result at i-th CNN layer and $\mathcal{G}$ is the ground truth of the annotated breast lesion mask. $L^i{}_{seg}$ combines a dice coefficient loss and a binary cross-entropy loss to compute the difference between $\mathcal{P}^i$ and $\mathcal{G}$:

$$L^i{}_{seg} = 1 - \frac{2\sum_{j=1}^{N_p}(\mathcal{P}^i)_j\times(\mathcal{G})_j}{\sum_{j=1}^{N_p}(\mathcal{P}^i)_j^2 + \sum_{j=1}^{N_p}(\mathcal{G})_j^2} - \frac{1}{N_p}\sum_{j=1}^{N_p}(\mathcal{P}^i)_j log(\mathcal{G})_j \tag{7}$$

where N_p is the number of pixels in the $\mathcal{P}^i$;

$L^i{}_{boundary}$ is computed as the mean square error (MSE) between the predicted breast lesion boundary map (denoted as $\mathcal{D}^i$) and ground truth of the boundary map (denoted as $\mathcal{B}_G$):

$$L^i{}_{boundary} = \sum_{j=1}^{N_p}\{(\mathcal{D}^i)_j - (\mathcal{B}_G)_j\}^2 \tag{8}$$

where N_p is the number of pixel in $\mathcal{D}^i$; $(\mathcal{D}^i)_j$ is the jth pixel at $\mathcal{D}^i$; and $(\mathcal{B}_G)_j$ is the jth pixel at $\mathcal{B}_G$.

Moreover, following $L^i{}_{seg}$, $L^f{}_{seg}$ also combines the dice coefficient loss and binary cross-entropy loss to compute the difference between the predicted segmentation map (denoted as $\mathcal{F}$) and $\mathcal{G}$ (see Eq. (7)):

$$L^f{}_{seg} = 1 - \frac{2\sum_{j=1}^{N_p}(\mathcal{F})_j\times(\mathcal{G})_j}{\sum_{j=1}^{N_p}(\mathcal{F})_j^2 + \sum_{j=1}^{N_p}(\mathcal{G})_j^2} - \frac{1}{N_p}\sum_{j=1}^{N_p}(\mathcal{F})_j log(\mathcal{G})_j\,. \tag{9}$$

3.4. Implementation

3.4.1. Training parameters

To accelerate the training process, we initialize the parameters of the feature extract network using the pre-trained ResNext on ImageNet while other parameters are initialized by random noise. The SGD algorithm is used to optimize the whole network with a momentum of 0.9, a weight decay of 0.0001, a mini-batch size of 4, and 100 epochs. We set the initial learning rate as 0.001 and reduce it by multiplying 0.1 after finishing every 50 epochs. Random rotation and horizontal flip operations are adopted for performing the data augmentation on the training set. We implement the whole network using PyTorch library and train our network on a single NVIDIA TITIAN Xp GPU.

3.4.2. Inference

In the testing stage, we take the segmentation result predicted from the refined features of the dual guided non-local block as the output of our segmentation network, and then pass the result to the conditional random fields (CRF) (Krähenbühl and Koltun, 2011) for obtaining the final segmentation result. The network has 55 M trainable parameters. The inference time was 0.039 s per image.

4. Experiments

We first introduce two datasets on breast ultrasound lesion segmentation and evaluation metrics, then conduct ablation studies to verify the major components of our network, as well as quantitatively and qualitatively compare our method against the state-of-the-art segmentation methods.

4.1. Datasets

We evaluate our segmentation network on two datasets including a public benchmark dataset (i.e., BUSI in Al-Dhabyani et al., 2020) and our collected dataset. BUSI collected 780 images from 600 female patients, with 437 benign cases, 210 benign masses, and 133 normal cases. Note that the main purpose of breast lesion segmentation in the clinical usage is for the lesion assessment, tracking the lesion change, and identifying distribution and seriousness of lesions. As a result, clinicians usually screen the input ultrasound sample firstly to identify the lesion, and then conduct the breast lesion segmentation for clinical measurement. As a result, we remove the normal cases without breast lesion masks to form the benchmark dataset, and adopt the three-fold cross-validation to test each segmentation method.

Our collected dataset has 632 clinical breast ultrasound images in total from 200 patients. The images are captured by different ultrasound imaging systems from Shenzhen Peoples Hospital and the Second Affiliated Hospital of Jinan University. We follow the widely-used annotation procedure of the medical image segmentation for annotating breast lesions. Firstly, three experienced radiologists are invited to annotate the breast lesion regions of each ultrasound image using a software interface developed via Matlab. Each radiologist used about two weeks to delineate all the breast lesion regions, and the segmentation ground truths of each image were then obtained based on inner- and intra-observer agreement of the three radiologists. Then, the final ground-truths were further refined by a senior radiologist with more than 10-year experience for quality control. To make the comparisons fair, we adopt the seven-fold cross-validation to test each segmentation method on this dataset.

4.2. Evaluation metrics

We adopt seven commonly used metrics to quantitatively compare different methods on the breast lesion segmentation. They are Dice coefficient (denoted as Dice), Jaccard index, Recall, Precision, Accuracy, Hausdorff distance (denoted as HD) and average boundary distance (denoted as ABD).

4.3. Ablation analysis of our GG-Net

In this section, we show the effectiveness of the principal components of our network, i.e., sptial-wise GGB, channel-wise GGB, and BD module in our network. And the ablation study experiments are mainly conducted on our collected dataset. The baseline (i.e., first row of Table 1) is constructed by removing both GGB and the BD module from our network. It is the original DeeplabV3+ network with ResNeXt as the backbone.

Table 1 shows the comparison results of our method with different components. By comparing 'SNLB', 'CNLB' and baseline (first row of Table 1), we can see that learning the long-range dependencies has a superior performance in segmenting the breast lesion regions from the ultrasound images. Then, 'SNLB + Guidance ' (i.e., spatial-wise GGB) and 'CNLB + Guidance' (i.e., channel-wise GGB) have better results than 'SNLB' and 'CNLB', showing that adding our MLIF guidance information into the spatial non-local and the channel non-local can help to capture the long-range position dependencies for the breast lesion segmentation. Moreover, the combination of the spatial-wise GGB and the channel-wise GGB has superior segmentation results over only using spatial-wise GGB or channel-wise GGB, demonstrating that combining the spatial and channel information into learning guided non-local features can enhance the breast lesion segmentation performance. Finally, our method with full components has the best segmentation accuracy, which means that the detected breast lesion boundaries in the BD module of our network also contribute to the superior breast lesion segmentation performance.

Fig. 8 visually compares the segmentation results produced by the baseline, "basic + GGB" and our method. From the visual results, we can easily find that "basic + GGB" has a higher segmentation accuracy than "basic", showing that the developed GGB can learn the long-range position dependencies to boost the breast lesion segmentation performance. Moreover, as shown in Fig. 8(e) and (d), our method (i.e., "basic + GGB + BD") can more accurately detect breast lesion regions than "basic+GGB". It means that adding the BD module into our method can further improve the segmentation accuracy by generating refined boundaries.

BD on the network output branch Our network applies the BD module on different CNN layers; see Fig. 2. Here, we modify our network by applying the BD module on the output branch for detecting breast lesions and the modified network is denoted as "Ours-BD". Table 2 lists different metric values of our method and "Ours-BD", showing that our method has only slightly better metric results than "Ours-BD". It means that adding the BD module on the network output branch reaches a similar segmentation accuracy as our network.

Alternative deep supervision in BD modules Note that the BD module of our network imposes the deep supervision on two predictions, i.e., the breast lesion segmentation and the breast lesion boundary detection. To really verify the contribution of the BD module, we conduct an experiment by constructing a network (denoted as 'Ours-ADS') by using alternative deep supervision methods in the BD module, which means that we only impose the deep supervisions on the breast lesion segmentation and remove the supervisions on breast lesion boundary predictions in each BD module. Table 3 summarizes the quantitative results of our method and 'Ours-ADS' on our collected dataset. From the results, we can easily conclude that our method has achieved superior quantitative results than 'Ours-ADS' on all the seven evaluation metrics, demonstrating that utilizing an alternative deep supervision method (i.e., removing breast lesion boundary detection supervision) in the BD

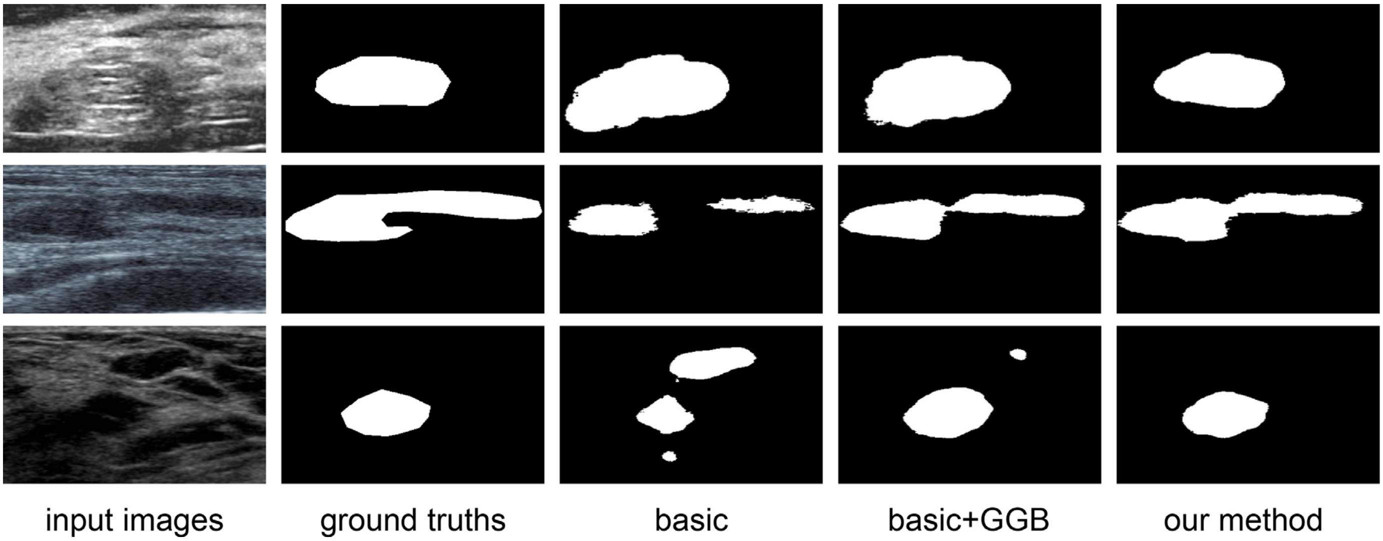


Fig. 8. Visual results of ablation study. (a) Input images; (b) Ground truths; (c)–(e) are the segmentation produced by basic, “basic + GGB”, and our method (i.e., “basic + GGB + BD”) respectively.

Table 1
Quantitative results on our collected dataset and the number of parameters for all networks constructed in ablation study on our collected dataset.. The first row is Deeplabv3+ with ResNeXt as the feature extraction backbone. “Guidance” denotes guidance information. “SNLB” denotes the traditional spatial-wisely non-local block while “SNLB+Guidance” is our spatial-wise GGB (see Fig. 4). “CNLB” is the traditional channel-wisely non-local block while “CNLB + Guidance” is our channel-wise GGB (see Fig. 5).

SNLB	CNLB	Guidance	BD	Parameters	Jaccard %	Dice %	Accuracy %	Recall %	Precision %
				53.4 M	73.4 ± 2.5	81.5 ± 2.6	97.0 ± 0.3	78.9 ± 2.4	88.7 ± 3.0
✓				53.5 M	77.6 ± 1.3	84.5 ± 1.2	97.2 ± 0.3	83.1 ± 1.6	90.7 ± 1.7
	✓			53.5 M	77.5 ± 1.6	84.3 ± 1.3	97.2 ± 0.4	83.5 ± 1.7	90.8 ± 1.6
✓		✓		53.9 M	78.1 ± 1.4	85.0 ± 1.3	97.3 ± 0.4	84.1 ± 0.2	91.0 ± 1.5
	✓	✓		53.9 M	78.2 ± 1.4	85.2 ± 1.2	97.3 ± 0.4	84.5 ± 1.2	90.9 ± 1.5
✓	✓			55.2 M	78.4 ± 1.6	85.4 ± 1.4	97.2 ± 0.4	84.9 ± 1.8	90.9 ± 1.7
✓	✓	✓		55.4 M	78.8 ± 1.7	86.7 ± 1.2	97.3 ± 0.3	86.1 ± 1.7	91.2 ± 1.2
✓	✓	✓	✓	55.4 M	**79.1 ±1.6**	**87.1 ±1.4**	**97.4 ±0.3**	**86.6 ±1.7**	**91.3 ±1.0**

Table 2
Quantitative results on our method and that with the BD module on the network output branch on our collected dataset.

	Dice %	Jaccard %	Accuracy %	Recall %	Precision %
Our method	**87.1 ±1.4**	**79.1 ±1.6**	**97.4 ±0.3**	**86.6 ±1.7**	**91.3 ±1.0**
Ours-BD	87.0 ± 1.3	**79.1 ±1.5**	97.3 ± 0.4	86.4 ± 1.0	91.0 ± 1.5

Table 3
Quantitative comparisons of our network with and without an alternative deep supervision in BDBL.

	Jaccard %	Dice %	Accuracy %	Recall %	Precision %	HD	ABD
Ours-ADS	78.5 ± 1.7	86.6 ± 1.5	97.3 ± 0.3	86.3 ± 1.2	86.1 ± 1.5	16.4 ± 2.3	5.5 ±0.8
GG-Net (our method)	**79.1 ±1.6**	**87.1 ±1.2**	**97.4 ±0.3**	**86.6 ±1.7**	**91.3 ±1.0**	**16.2 ±2.4**	**5.3 ±0.7**

module reduces the breast lesion segmentation accuracy of our network.

4.4. Comparison with the state-of-the-arts

Compared methods We compare our network against several deep-learning-based segmentation methods, including *context-based methods*: feature pyramid network (FPN) Lin et al. (2017), U-Net Ronneberger et al. (2015), U-Net++ Zhou et al. (2018), pre-trained TernausNet Iglovikov and Shvets (2018), SK-U-Net Byra et al. (2020), DeeplabV3+ Chen et al. (2018); as well as *attention-based methods*: AG-Unet Schlemper et al. (2019), and DAF Wang et al. (2018b). To provide fair comparisons, we obtain the segmentation results of compared methods by downloading their public implementations and re-training their networks on our dataset. Similarly, we also use the CRF Krähenbühl and Koltun (2011) to post-process the predicted segmentation maps of compared methods.

Quantitative comparisons Table 4 reports the mean and standard deviation values of the seven metrics among our method and all the competitors on our collected dataset, while Table 5 summarizes the mean and standard deviation scores of seven metrics on BUSI. Compared to other segmentation methods, our method has larger Jaccard, Dice, Accuracy, recall, and prediction scores, as well as smaller HD and ABD values. It indicates that our GG-Net can more accurately identify breast lesions from ultrasound images than all the competitors.

Table 4
Comparing our method (GG-Net) with the state-of-the-art methods for beast lesion segmentation on our collected dataset.

	Jaccard %	Dice %	Accuracy %	Recall %	Precision %	HD	ABD
U-Net Ronneberger et al. (2015)	69.3 ± 2.4	78.0 ± 2.4	96.5 ± 0.3	76.9 ± 0.3	85.6 ± 2.4	25.1 ± 2.4	8.1 ± 0.9
U-Net+ Zhou et al. (2018)	73.3 ± 2.1	82.1 ± 2.2	96.6 ± 0.4	81.1 ± 1.7	87.9 ± 2.6	25.6 ± 4.0	8.4 ± 1.2
TernausNet Iglovikov and Shvets (2018)	73.7 ± 1.5	82.2 ± 1.5	96.8 ± 0.3	82.1 ± 1.2	86.9 ± 0.2	21.6 ± 2.6	7.5 ± 0.9 5
FPN Lin et al. (2017)	77.2 ± 1.9	85.4 ± 1.7	97.1 ± 0.4	85.6 ± 1.8	89.1 ± 2.4	18.1 ± 2.7	6.1 ± 1.0
DeepLabv3+ Chen et al. (2018)	73.4 ± 2.5	81.5 ± 2.6	97.0 ± 0.3	78.9 ± 2.4	88.7 ± 3.0	22.3 ± 4.1	7.9 ± 1.3
AG-Unet Schlemper et al. (2019)	74.1 ± 1.9	82.8 ± 1.9	96.6 ± 0.4	82.5 ± 2.3	87.3 ± 1.9	24.1 ± 3.0	7.8 ± 1.0
DAF Wang et al. (2018b)	75.4 ± 1.9	83.6 ± 2.1	97.1 ± 0.4	84.5 ± 2.3	86.6 ± 2.4	17.1 ± 2.3	5.8 ± 0.9
GG-Net (our method)	**79.1 ±1.6**	**87.1 ±1.2**	**97.4 ±0.3**	**86.6 ±1.7**	**91.3 ±1.0**	**16.2 ±2.4**	**5.3 ±0.7**

Table 5
Comparing our method (GG-Net) with the state-of-the-art methods for beast lesion segmentation on the BUSI dataset. Best results are marked with bold texts.

	Jaccard %	Dice %	Accuracy %	Recall %	Precision %	HD	ABD
U-Net Ronneberger et al. (2015)	64.1 ± 1.8	73.3 ± 1.7	95.9 ± 0.6	70.4 ± 1.9	83.3 ± 1.3	65.2 ± 4.7	24.4 ± 2.3
U-Net+ Zhou et al. (2018)	56.2 ± 1.7	66.0 ± 1.4	95.4 ± 0.4	62.8 ± 1.5	78.2 ± 1.2	78.6± 6.1	31.8 ± 4.0
FPN Lin et al. (2017)	72.2 ± 1.6	80.4 ± 1.6	95.9 ± 0.6	79.3 ± 1.3	85.1 ± 1.5	47.6± 5.8	18.9 ± 2.6
DeepLabv3+ Chen et al. (2018)	68.2 ± 1.8	77.2 ± 1.6	96.3 ± 0.6	74.4 ± 2.5	84.8 ± 1.8	54.4± 5.9	22.4 ± 2.9
SK-U-Net Byra et al. (2020)	–	70.9	95.6	–	–	–	–
DAF Wang et al. (2018b)	68.4 ± 3.1	77.1 ± 3.1	96.4 ± 0.6	76.7 ± 3.8	82.2 ± 3.1	46.9± 8.1	17.9 ± 4.7
GG-Net (our method)	**73.8 ±1.1**	**82.1 ±1.1**	**96.9 ±0.5**	**81.2 ±1.6**	**86.5 ±0.5**	**43.9±4.8**	**16.4 ±2.2**

Table 6
Comparing our method (GG-Net) with the state-of-the-art methods for beast lesion segmentation on the BUSI dataset. (include normal data). Best results are marked with bold texts.

	Jaccard %	Dice %	Accuracy %	Recall %	Precision %	HD	ABD
U-Net Ronneberger et al. (2015)	51.2 ± 1.9	58.8 ± 1.5	96.3 ± 0.7	56.1 ± 2.3	68.1 ± 1.7	67.1 ± 6.1	24.7 ± 3.1
U-Net+ Zhou et al. (2018)	44.5 ± 3.5	52.1 ± 3.7	95.9 ± 0.2	48.8 ± 4.7	63.6 ± 2.4	73.5 ± 5.0	27.8 ± 2.0
FPN Lin et al. (2017)	55.4 ± 2.1	63.0 ± 2.3	96.2 ± 0.4	62.1 ± 3.4	68.3 ± 1.9	56.8 ± 8.9	21.2 ± 4.6
DeepLabv3+ Chen et al. (2018)	54.3 ± 2.1	62.1 ± 2.5	96.4 ± 0.5	59.2 ± 2.4	63.6 ± 2.5	55.5 ± 10.7	21.3 ± 5.5
DAF Wang et al. (2018b)	55.8 ± 1.5	62.8 ± 1.8	96.6 ± 0.6	62.8 ± 2.3	66.5 ± 1.1	52.8 ± 4.2	20.3 ± 2.3
GG-Net (our method)	**56.6 ±1.9**	**64.1 ±2.1**	**96.6 ±0.3**	**63.3 ±3.6**	**69.7 ±0.4**	**48.6 ±7.2**	**18.8 ±3.3**

On the other hand, when further looking into the metric results in Tables 4 and 5, we can find that the segmentation performance on our collected dataset (see Table 4) is better than the results on the public BUSI dataset (see Table 5) with respect to all seven evaluation metrics. The reason behind is that the ultrasound image quality in our dataset is better than that in BUSI, thereby making the better segmentation performance.

Utilizing BUSI's normal cases The general purpose of breast lesion segmentation in the clinical usage is mainly for the lesion assessment, tracking the lesion change, and identifying distribution and seriousness of lesions. As a result, people usually assume that the input ultrasound samples possess one or more lesions, and then conduct the breast lesion segmentation for clinical analysis. Here, we conduct another experiment by including the normal cases of BUSI into the training data and re-training all the compared methods and our network to obtain their new results. Tables 5 and 6 report the results of each method with and without the BUSI's normal cases. According to the results, we can easily find that the quantitative results of all the competitors and our network tend to be worse when considering normal cases in the network training. Among all the segmentation methods, our network still achieves the best performance of all seven metrics even though the normal cases are added into the training set and the testing set.

Visual comparisons We also visually compare the breast lesion segmentation results produced by our network and compared methods; see Fig. 9 for examples. U-Net, U-Net++, FPN, and DeeplabV3+ tend to neglect breast lesion details or wrongly classify non-lesion regions as breast lesions into their predicted segmentation maps, while our method produces more accurate segmentation results on breast lesion regions. Furthermore, our results are most consistent with ground truths (see Fig. 9(b)) among all segmentation results. This proves the effectiveness of long-range dependencies and breast lesion boundaries in our method.

5. Application

Note that our network can be retrained for other ultrasound image segmentation tasks. Hence, we further evaluate the effectiveness of our network by testing it on the ultrasound prostate segmentation task. To conduct fair comparisons, we follow the same experimental setting of a recent prostate segmentation work, i.e., DAF Wang et al. (2018b), to obtain the prostate segmentation results of our network. We use the DAF's training set to train our network, test our method on the DAF's testing set, and report the results of same four metric (i.e., Jaccard, Dice, Recall and Precision; see Wang et al. (2018b) for their definitions) for comparisons. Table 7 summarizes the comparison results on four metrics between our method and state-of-the-art networks, including U-Net Ronneberger et al. (2015), FCN Lin et al. (2017), BCRNN Yang et al. (2017), and DAF Wang et al. (2018b); see Wang et al. (2018b) for details of these compared methods. Apparently, our method outperforms all the competitors on almost all the four metrics, demonstrating that our method can also identify prostate regions better from ultrasound images. It further verifies the effectiveness of the developed segmentation network in our work.

6. Discussions

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related death among women worldwide. The automatic breast lesion segmentation from ultrasound images

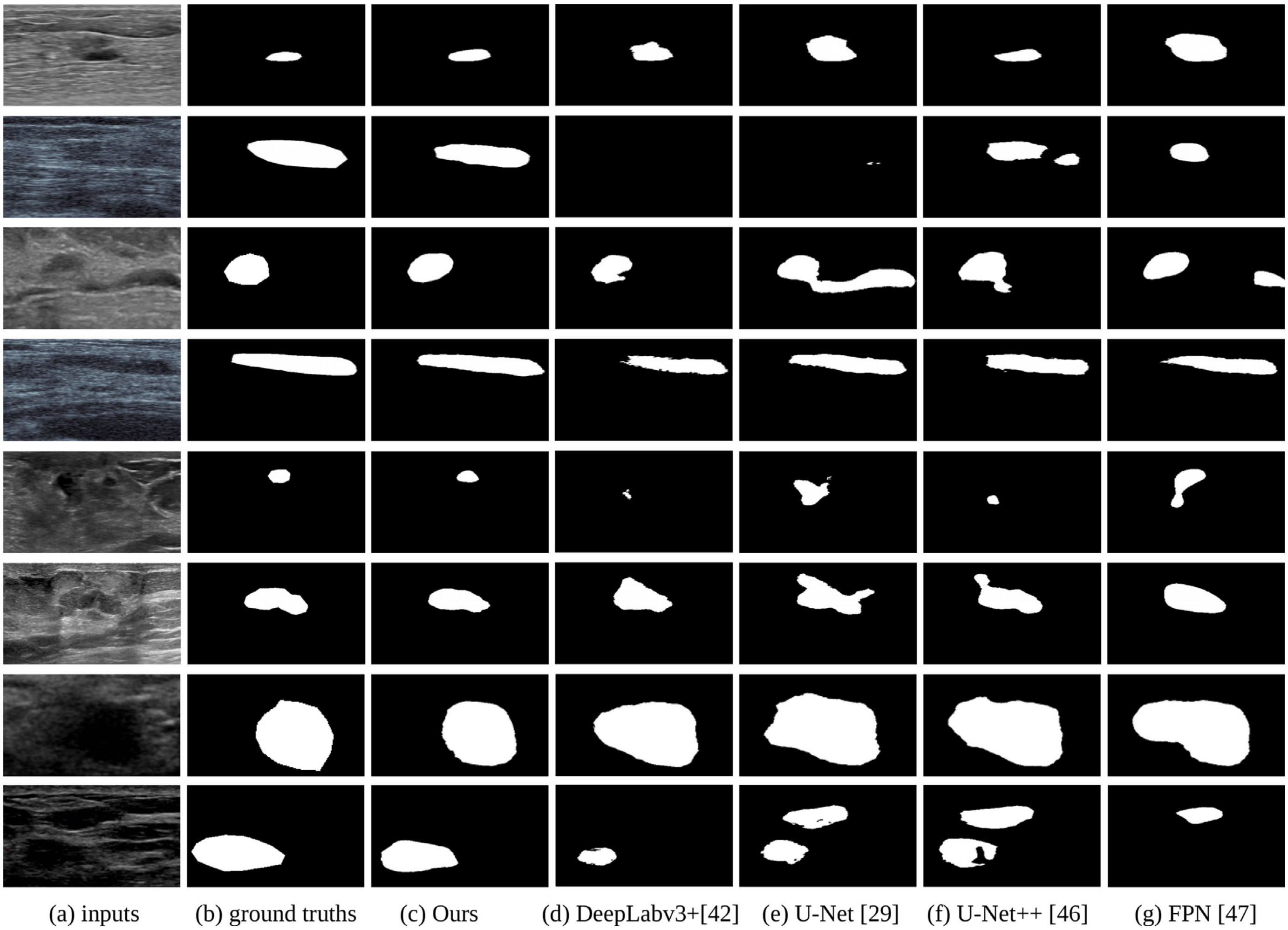


Fig. 9. Visual comparison of the breast lesion segmentation maps produced by different methods. (a) input breast ultrasound images; (b) ground truths; (c)–(g) are segmentation results produced by our method, DeeplabV3+ Chen et al. (2018), U-Net Ronneberger et al. (2015), U-Net++ Zhou et al. (2018), and FPN Lin et al. (2017).

Table 7
Metric results of different methods on ultrasound prostate segmentation.

	Jaccard %	Dice %	Recall %	Precision %
FCN Lin et al. (2017)	85.1	91.9	90.8	93.3
BCRNN Yang et al. (2017)	86.0	92.4	90.5	94.5
U-Net Ronneberger et al. (2015)	87.1	93.0	96.8	89.9
DAF Wang et al. (2018b)	91.0	95.3	**97.0**	93.7
GG-Net (ours)	**91.2**	**95.4**	95.7	**95.1**

assists the doctors in finding early signals of breast cancer, which is of great importance in clinical practice. Traditional CNN-based methods (Chen et al., 2018; 2017b; 2014; Ronneberger et al., 2015; Lin et al., 2017) conducted convolutional operations in local regions to learn deep discriminative features for medical image analysis and thus suffered from unsatisfactory segmentation accuracy due to the limited receptive fields of their local convolutions.

Recently, capturing non-local long-range pixel dependencies has achieved superior prediction performance in many medical imaging community (Qi et al., 2019; Dou et al., 2018) by devising non-local blocks. However, these non-local blocks are only embedded into the deep CNN layers for network predictions. However, the deep layers of a segmentation network are responsible for capturing cues of the whole breast lesions and somehow lack parts of breast lesion regions due to the relatively larger receptive fields than shallow CNN layers. In this regard, we integrate all CNN layers to produce multi-level integrated features (MLIF) as a guidance information of the non-local blocks to complement more breast lesion boundary details (neglected by deep CNN layers).

This project presented a global guidance network (denoted as “GG-Net”) with a spatial guidance block and a channel guidance block to leverage guidance information for improving long-range dependency feature learning in spatial and channel manners. Moreover, a breast lesion boundary detection module is devised to learn boundary details for futher refining the breast lesion segmentation performance. Compared with state-of-the-art methods, our network achieves a significant (p-value <0.05, see Table 8) improvement on two datasets, which proves the effectiveness of the developed spatial and channel guidance block as well as boundary detection block. Moreover, compared with other segmentation networks, our method has better performance on relatively less obvious lesion segmentation. This is crucial in the clinical practice, especially for breast ultrasound, where most of the lesions have low contrast, shadows, and blurry boundaries.

Table 8
P-values between our method and other compared methods on different evaluation metrics.

Metrics	U-Net vs. Ours	U-Net+ vs. Ours	TernausNet vs. Ours	FPN vs. Ours	AG-Net vs. Ours	DAF vs. Ours	DeepLabv3+ vs. Ours
Jaccard	1.62×10^{-7}	1.64×10^{-5}	9.41×10^{-5}	4.60×10^{-2}	3.45×10^{-4}	4.20×10^{-3}	3.00×10^{-6}
Dice	3.94×10^{-8}	2.81×10^{-5}	5.17×10^{-5}	4.00×10^{-2}	4.19×10^{-4}	1.10×10^{-3}	7.86×10^{-6}
Accuracy	1.74×10^{-3}	3.28×10^{-3}	3.48×10^{-2}	3.70×10^{-2}	1.33×10^{-3}	4.80×10^{-2}	3.57×10^{-2}
Recall	9.80×10^{-11}	2.31×10^{-4}	3.02×10^{-3}	3.50×10^{-2}	4.55×10^{-3}	1.20×10^{-3}	3.97×10^{-9}
Precision	6.70×10^{-3}	2.36×10^{-3}	3.24×10^{-4}	6.60×10^{-3}	2.06×10^{-3}	8.90×10^{-3}	2.70×10^{-3}
HD	1.64×10^{-6}	6.10×10^{-6}	7.35×10^{-4}	4.00×10^{-4}	2.89×10^{-6}	8.40×10^{-3}	4.00×10^{-4}
ABD	1.01×10^{-6}	1.73×10^{-7}	8.07×10^{-5}	2.00×10^{-3}	6.03×10^{-5}	5.50×10^{-3}	7.40×10^{-3}

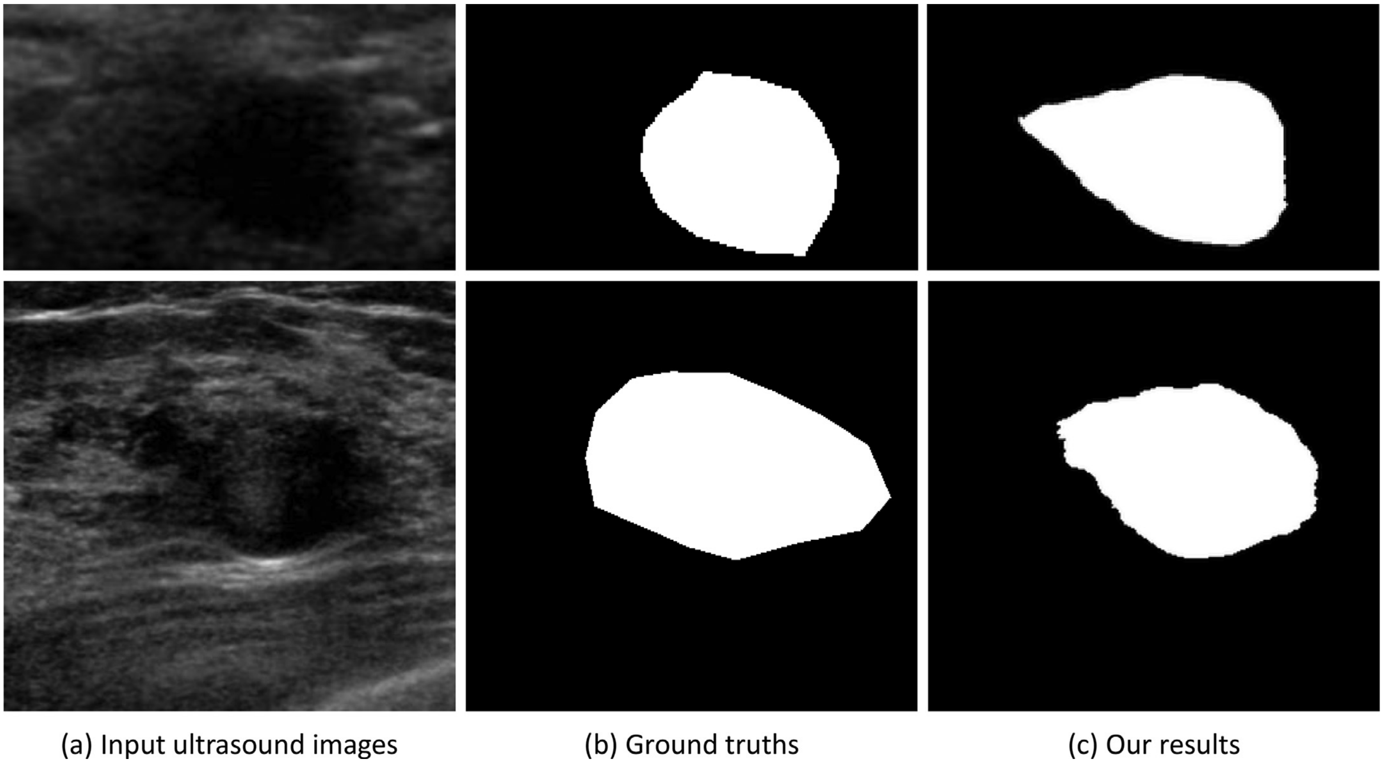


Fig. 10. Failure cases. (a) Input ultrasound images. (b) Ground truths of the breast lesion segmentation. (c) Segmentation results produced by our network.

Note that the elastography image encodes the density of the tissue in the screen. In our future work, we will leverage elastography images for further boosting breast lesion segmentation results in ultrasound.

Failure cases Like other breast lesion segmentation methods, our network tends to fail in fully detecting breast lesion regions when the target breast lesion has a very large size and a complicated intensity distribution inside it, or unclear boundaries. Fig. 10 shows two examples, where our results in (c) wrongly identify non-lesion regions as lesion ones, or neglect a part of breast lesion regions of the input ultrasound image when comparing to the ground truths (see (b)).

Statistical test To investigate the statistical significance of the proposed network over compared methods on different quantitative metrics, we conduct a statistical analysis of p-values and show the p-values of our network against compared methods in terms of different metrics in Table 8. As shown in Table 8, we can find that the p-values of all the seven paired methods are almost smaller than 0.05 for all the seven metrics, demonstrating that our method can be regarded as reaching a significant improvement over the other six compared methods on these evaluation metrics. Note that the Accuracy p-values of our method over TernausNet, FPN, DAF, and DeepLabv3+ are 3.48×10^{-2}, 3.70×10^{-2}, 4.80×10^{-2}, and 3.57×10^{-2}, which are closer to 0.05. It indicates that our method has a similar Accuracy performance to TernausNet, FPN, DAF, and DeepLabv3+. Generally, the superior metric performance of our method in Tables 4–6 shows that our network can better segment breast lesions from ultrasound than other compared segmentation methods.

7. Conclusion

This paper presents a global guidance network (GG-Net) equipped with a global guidance block and a breast lesion boundary detection module for breast lesion segmentation in ultrasound images. The global guidance block aims to combine the multi-layer context information as guidance information to learn the long-term non-local features in spatial and channel manners. The breast lesion boundary detection predicts additional breast lesion boundary map to assist in improving the segmentation performance. We evaluate our network on a public dataset and our collected dataset of breast lesion segmentation in ultrasound images by comparing it against state-of-the-art methods, and the experimental results show that our network can more accurately segment the breast lesions than all the competitors. We also show the application of our network on the ultrasound prostate segmentation task and our network also has a higher segmentation accuracy than state-of-the-art methods.

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.

CRediT authorship contribution statement

Cheng Xue: Methodology, Software, Writing - original draft, Formal analysis. **Lei Zhu:** Conceptualization, Methodology, Investigation, Writing - review & editing, Data curation. **Huazhu Fu:** Conceptualization, Writing - review & editing, Investigation. **Xiaowei Hu:** Writing - review & editing, Visualization. **Xiaomeng Li:** Writing - review & editing, Visualization. **Hai Zhang:** Validation, Resources, Data curation. **Pheng-Ann Heng:** Supervision, Project administration, Funding acquisition.

Acknowledgements

The work described in this paper was supported by Key-Area Research and Development Program of Guangdong Province, China under Project No. 2020B010165004, Hong Kong Innovation and Technology Fund under Project No. ITS/311/18FP, National Natural Science Foundation of China (Grant No. 61902275), National Natural Science Foundation of China under Project No. U1813204, Natural Science Foundation of SHENZHEN City NO: JCYJ20190806150001764, Natural Science Foundation of GUANGDONG Province No:2020A1515010978, and The Sanming Project of Medicine in Shenzhen training project No: SYJY201802.

References

Ahn, C.K., Heo, C., Jin, H., Kim, J.H., 2017. A novel deep learning-based approach to high accuracy breast density estimation in digital mammography. In: Medical Imaging 2017: Computer-Aided Diagnosis, 10134. International Society for Optics and Photonics, p. 101342O.

Al-Dhabyani, W., Gomaa, M., Khaled, H., FahmyaH, A., 2020. Dataset of breast ultrasound images. Data Brief 28.

American Cancer Society, 2019. Cancer facts & figures 2019.

Ashton, E.A., Parker, K.J., 1995. Multiple resolution Bayesian segmentation of ultrasound images. Ultrason. Imaging 17 (4), 291–304.

Boukerroui, D., Basset, O., Guerin, N., Baskurt, A., 1998. Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation. Eur. J. Ultrasound 8 (2), 135–144.

Byra, M., Jarosik, P., Szubert, A., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., Andre, M., 2020. Breast mass segmentation in ultrasound with selective kernel U-net convolutional neural network. Biomed. Signal Process. Control 61, 102027.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. pattern Anal. Mach. Intell. (6) 679–698.

Chen, C.-M., Lu, H.H.-S., Huang, Y.-S., 2002. Cell-based dual snake model: a new approach to extracting highly winding boundaries in the ultrasound images. Ultrasound Med. Biol. 28 (8), 1061–1073.

Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., Chua, T.-S., 2017a. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5659–5667.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.. Semantic image segmentation with deep convolutional nets and fully connected CRFs.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017b. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Dhungel, N., Carneiro, G., Bradley, A.P., 2017. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. Image Anal. 37, 114–128.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.-A., 2016. 3D deeply supervised network for automatic liver segmentation from CT volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 149–157.

Dou, T., Zhang, L., Zheng, H., Zhou, W., 2018. Local and non-local deep feature fusion for malignancy characterization of hepatocellular carcinoma. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 472–479.

Feng, M., Lu, H., Ding, E., 2019. Attentive feedback network for boundary-aware salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1623–1632.

Heinrich, M.P., Oktay, O., Bouteldja, N., 2019. Obelisk-Net: fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions. Med. Image Anal. 54, 1–9.

Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., Chen, X., 2019. Interaction-and-aggregation network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9317–9326.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.

Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., Chang, C., 2019. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Med. Phys. 46 (1), 215–228.

Huang, S.-F., Chen, Y.-C., Moon, W.K., 2008. Neural network analysis applied to tumor segmentation on 3D breast ultrasound images. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1303–1306.

Iglovikov, V., Shvets, A.. Ternausnet: U-Net with vgg11 encoder pre-trained on imagenet for image segmentation.

Joutard, S., Dorent, R., Isaac, A., Ourselin, S., Vercauteren, T., Modat, M., 2019. Permutohedral attention module for efficient non-local neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 393–401.

Kirberger, R.M., 1995. Imaging artifacts in diagnostic ultrasound—Areview. Vet. Radiol. Ultrasound 36 (4), 297–306.

Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with gaussian edge potentials. In: Advances in Neural Information Processing Systems, pp. 109–117.

Kwak, J.I., Kim, S.H., Kim, N.C., 2005. RD-based seeded region growing for extraction of breast tumor in an ultrasound volume. In: International Conference on Computational and Information Science. Springer, pp. 799–808.

Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.-H., Cheng, J.-Z., 2018. Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. Neurocomputing 321, 178–186.

Li, H., He, X., Zhou, F., Yu, Z., Ni, D., Chen, S., Wang, T., Lei, B., 2018. Dense deconvolutional network for skin lesion segmentation. IEEE J. Biomed. Health Inform. 23 (2), 527–537.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125.

Liu, B., Cheng, H.-D., Huang, J., Tian, J., Tang, X., Liu, J., 2010. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recognit. 43 (1), 280–298.

Lo, C., Shen, Y.-W., Huang, C.-S., Chang, R.-F., 2014. Computer-aided multiview tumor detection for automated whole breast ultrasound. Ultrason. imaging 36 (1), 3–17.

Madabhushi, A., Metaxas, D., 2002. Automatic boundary extraction of ultrasonic breast lesions. In: Proceedings IEEE International Symposium on Biomedical Imaging. IEEE, pp. 601–604.

Madabhushi, A., Metaxas, D.N., 2003. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Trans. Med. Imaging 22 (2), 155–169.

Mishra, D., Chaudhury, S., Sarkar, M., Soin, A.S., 2018. Ultrasound image segmentation: a deeply supervised network with attention to boundaries. IEEE Trans. Biomed. Eng. 66 (6), 1637–1648.

Moon, W.K., Lo, C.-M., Chen, R.-T., Shen, Y.-W., Chang, J.M., Huang, C.-S., Chen, J.-H., Hsu, W.-W., Chang, R.-F., 2014. Tumor detection in automated breast ultrasound images using quantitative tissue clustering. Med. Phys. 41 (4), 042901.

Mordang, J.-J., Gubern-Mérida, A., den Heeten, G., Karssemeijer, N., 2016a. Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications. Med. Phys. 43 (4), 1676–1687.

Mordang, J.-J., Janssen, T., Bria, A., Kooi, T., Gubern-Mérida, A., Karssemeijer, N., 2016b. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: International Workshop on Breast Imaging. Springer, pp. 35–42.

Othman, A.A., Tizhoosh, H.R., 2011. Segmentation of breast ultrasound images using neural networks. In: Engineering Applications of Neural Networks. Springer, pp. 260–269.

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters–improve semantic segmentation by global convolutional network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361.

Qi, K., Yang, H., Li, C., Liu, Z., Wang, M., Liu, Q., Wang, S.. X-Net: brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 234–241.

Roy, A.G., Navab, N., Wachinger, C., 2018. Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks. IEEE Trans. Med. Imaging 38 (2), 540–549.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D., 2019. Attention gated networks: learning to leverage salient regions in medical images. Med. Image Anal. 53, 197–207.

Shan, J., Cheng, H., Wang, Y., 2012. Completely automated segmentation approach for breast ultrasound images using multiple-domain features. Ultrasound Med. Biol. 38 (2), 262–275.

Shan, J., Cheng, H.-D., Wang, Y., 2008. A novel automatic seed point selection algorithm for breast ultrasound images. In: 2008 19th International Conference on Pattern Recognition. IEEE, pp. 1–4.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in neural information processing systems, pp. 5998–6008.

Wang, X., Girshick, R., Gupta, A., He, K., 2018a. Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803.

Wang, Y., Deng, Z., Hu, X., Zhu, L., Yang, X., Xu, X., Heng, P.-A., Ni, D., 2018b. Deep attentional features for prostate segmentation in ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 523–530.

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 Recognit. 48 (2), 485–497.

Xiao, G., Brady, M., Noble, J.A., Zhang, Y., 2002. Segmentation of ultrasound b-mode images with intensity inhomogeneity correction. IEEE Trans. Med. Imaging 21 (1), 48–57.

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L., 2019. Medical breast ultrasound image segmentation by machine learning. Ultrasonics 91, 1–9.

Yang, X., Yu, L., Wu, L., Wang, Y., Ni, D., Qin, J., Heng, P.-A., 2017. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: Thirty-First AAAI Conference on Artificial Intelligence.

Yap, M.H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A.K., Martí, R., 2017. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform. 22 (4), 1218–1226.

Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., 1997. A geometric snake model for segmentation of medical imagery. IEEE Trans. Med. Imaging 16 (2), 199–209.

Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A., 2016. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging 36 (4), 994–1004.

Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T., 2018. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Trans. Biomed. Eng. 66 (4), 1006–1016.

Yu, Z., Tan, E.-L., Ni, D., Qin, J., Chen, S., Li, S., Lei, B., Wang, T., 2017. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. IEEE J. Biomed. Health Inform. 22 (3), 874–885.

Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., Agrawal, A., 2018. Context encoding for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7151–7160.

Zhang, Z., Xie, Y., Xing, F., McGough, M., Yang, L., 2017. MDNet: a semantically and visually interpretable medical image diagnosis network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6428–6436.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: a nested U-Net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 3–11.