Ultrasound in Med. & Biol., Vol. 49, No. 1, pp. 31–44, 2023

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2022.07.006

● *Original Contribution*

DGANET: A DUAL GLOBAL ATTENTION NEURAL NETWORK FOR BREAST LESION DETECTION IN ULTRASOUND IMAGES

Hui Meng,* Xuefeng Liu,† Jianwei Niu,† Yong Wang,‡ Jintang Liao,§ Qingfeng Li,¶ and Chen Chen†

* School of Intelligent Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, 1 Sub-lane Xiangshan, Hangzhou 310024, China; † State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Haidian District, Beijing, China; ‡ Department of Diagnostic Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; § Department of Ultrasound, Xiangya Hospital of Central South University, Changsha, China; and ¶ Research Center of Big Data and Computational Intelligence, Hangzhou Innovation Institute of Beihang University, Hangzhou, China

(*Received* 31 *December* 2021; *revised* 20 *June* 2022; *in final form* 13 *July* 2022)

Abstract—Deep learning-based breast lesion detection in ultrasound images has demonstrated great potential to provide objective suggestions for radiologists and improve their accuracy in diagnosing breast diseases. However, the lack of an effective feature enhancement approach limits the performance of deep learning models. Therefore, in this study, we propose a novel dual global attention neural network (DGANet) to improve the accuracy of breast lesion detection in ultrasound images. Specifically, we designed a bilateral spatial attention module and a global channel attention module to enhance features in spatial and channel dimensions, respectively. The bilateral spatial attention module enhances features by capturing supporting information in regions neighboring breast lesions and reducing integration of noise signal. The global channel attention module enhances features of important channels by weighted calculation, where the weights are decided by the learned interdependencies among all channels. To verify the performance of the DGANet, we conduct breast lesion detection experiments on our collected data set of 7040 ultrasound images and a public data set of breast ultrasound images. YOLOv3, RetinaNet, Faster R-CNN, YOLOv5, and YOLOX are used as comparison models. The results indicate that DGANet outperforms the comparison methods by 0.2%–5.9% in total mean average precision. (E-mail: niujianwei@buaa.edu.cn)

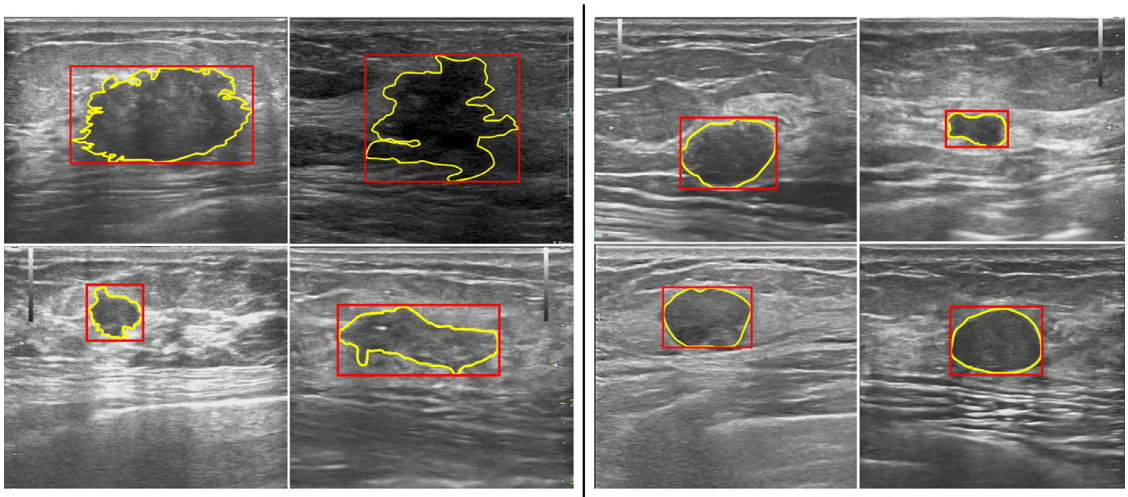
Key Words: **Ultrasound image, Breast cancer, Attention mechanism, Deep learning, Lesion detection.**

INTRODUCTION

Ultrasound imaging is one of the most common techniques for breast cancer diagnosis because of its non-ionizing radiation, low cost, real-time imaging and high sensitivity (Giger et al. 2013; Alcantara et al. 2014; Xing et al. 2021). However, image acquisition and analysis are often performed concurrently by radiologists with handheld ultrasound devices, which highly depend on the radiologists' experience. Additionally, the noises and artifact in ultrasound images interfere with the diagnosis of breast cancer. Therefore, many researchers have developed different deep learning models (Shin et al. 2019; Liao et al. 2020; Yap et al. 2020), which are supposed to alleviate the subjectivity of breast ultrasound (BUS) imaging and improve radiologists' accuracy in the diagnosis of breast cancer.

The low quality of ultrasound images and the heterogeneous appearances of breast lesions influence the discriminability of network features, which limits the performance of the deep learning models. Specifically, the low signal-to-noise ratio and areas similar to lesions (Fig. 1) introduce interference information into network features. Additionally, the significant shape variability of breast lesions (Fig. 1) makes it difficult for neural networks to extract discriminative network features on a single scale. Therefore, it is necessary to enhance the discriminability of features to improve the performance of the deep learning models.

Address correspondence to: Jianwei Niu, State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, 37 Xueyuan Road, Haidian District, Beijing 100191, China. E-mail: niujianwei@buaa.edu.cn



(a) Malignant (b) Benign

Fig. 1. Typical breast lesions in ultrasound images. Malignant (a) and benign (b) lesions of different size are shown. *Yellow curves* delineate the boundaries of the lesions. *Red rectangles* indicate the ground-truth bounding boxes of the lesions.

Recently, incorporation of medical domain knowledge into deep learning has been regarded as an effective approach to feature enhancement. One way is to utilize the hand-crafted features, which are designated by radiologists. For example, some researchers (Hagerty et al. 2019; Saba et al. 2020) have fused handcrafted features with deep learning features to improve the performance of classifiers. And others (Hussein et al. 2017; Murthy et al. 2017) have used handcrafted features as labels of multi-task networks to guide feature learning of neural networks. Additionally, handcrafted features are represented as image patches and taken as inputs to neural networks (Tan et al. 2019). Although the fusion of handcrafted features and deep learning help to improve the performance of network models, the calculation of handcrafted features influences the efficiency of deep learning models.

Another way is to design network structure based on the diagnostic experience of radiologists. One common method is guiding networks to extract features in regions on which radiologists focus. Considering radiologists generally read X-ray images from whole to local, a three-branch attention-guided convolution neural network is proposed to extract global and local features (Guan et al. 2018). Additionally, the work (Li et al. 2019) uses attention maps of ophthalmologists to guide feature learning of neural network. The experimental results of the aforementioned studies indicate that domain knowledge has the potential to enhance features of neural networks.

Inspired by these studies, we propose a novel dual global attention neural network (DGANet) for breast lesion detection in ultrasound images, which is illustrated in Figure 2. It uses attention mechanisms to enhance features in spatial and channel dimensions, respectively. Specifically, we observed that radiologists generally analyze breast lesions and surrounding issues in ultrasound images to make diagnoses. To integrate this domain knowledge into our network, we designed a bilateral spatial attention module (BSAM). It integrates global information by a weighted sum, where the weight is multiplied by feature similarity and physical Gaussian distance. That is, the feature similarity of any two positions in neighboring region is enhanced by physical Gaussian distance, whereas the feature similarity of two positions far away is weakened by physical Gaussian distance. Thus, the BSAM can enhance feature fusion in the neighboring region and suppress interference from distant noises. According to the Breast Imaging Reporting and Data System (BI-RADS) (Liberman and Menell 2002), radiologists generally analyze several important features of breast lesions to make diagnoses. Considering that a channel map in high-level features represents one kind of abstract image feature, we propose a global channel attention module (GCAM) to enhance features of important channels. Finally, the outputs of the BSAM and the GCAM are fused with the original input to further enhance features.

To assess the performance of the DGANet, we implemented breast lesion detection experiments on our collected data set of 7040 breast ultrasound images. The results revealed that employing the BSAM and the GCAM yields a result of 0.840 in total mean average precision, which outperforms the baseline by 4.3%,

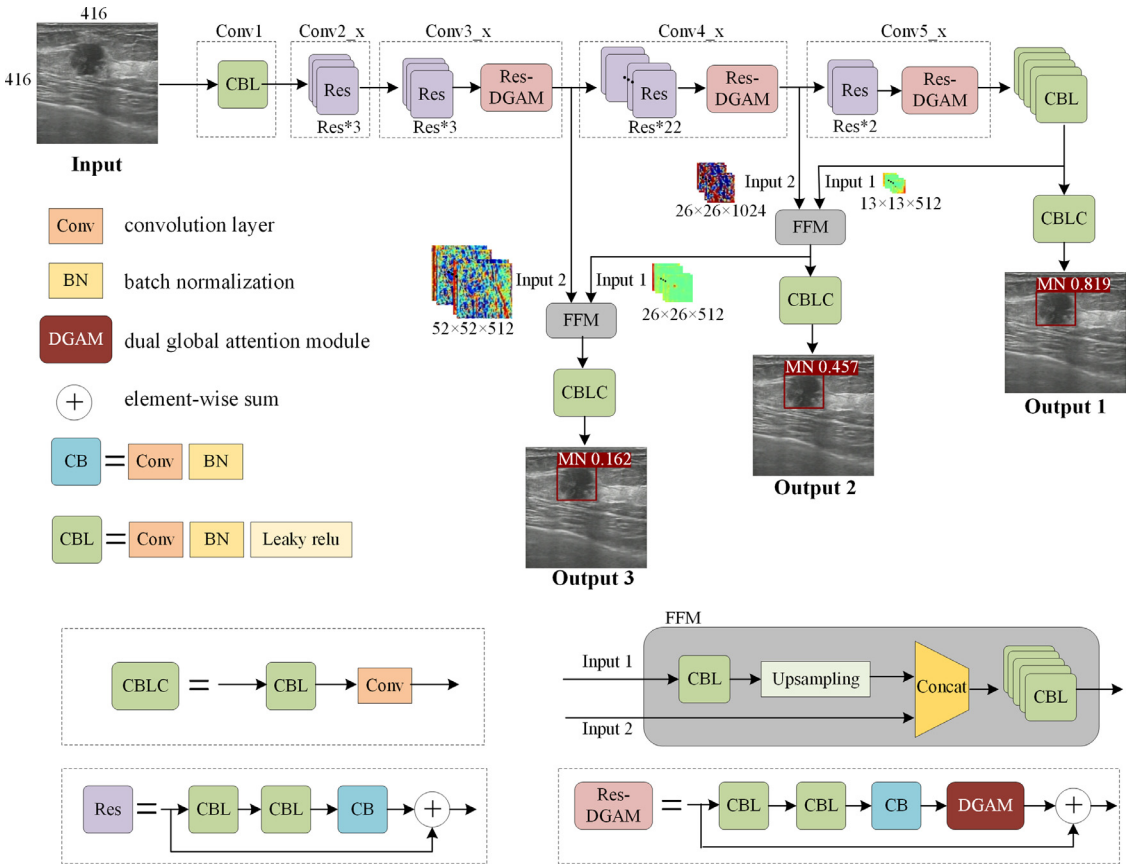


Fig. 2. Overview of the dual global attention neural network (DGANet). It consists of three branch networks, which correspond to three outputs. The backbone of the DGANet is ResNet-101, where three DGAMs are inserted into three residual blocks. *Red rectangular boxes* in ultrasound images represent detection boxes. Numbers in ultrasound images represent the confidence of detection boxes. Numbers below feature maps indicate the dimensions of the feature maps (H × W × C). The FFM achieves feature fusion of feature maps with different scales. C = number of channels; FFM = feature fusion module; H = height of feature map; MN = malignant lesion; W = width of feature map.

although using the BSAM or the GCAM alone can improve detection performance over the baseline by 2.5% or 3.2% in total mean average precision, respectively.

Our main contributions in this work can be summarized as follows:

- We propose a novel dual global attention neural network (DGANet) with global attention mechanism to enhance features for breast lesion detection in ultrasound images.
- A bilateral spatial attention module is proposed that integrates context information in regions neighboring breast lesions and suppresses interference from distant noises. A global channel attention module is designed based on global context to enhance features of important channels. Both attention modules improve the detection accuracy by modelling global context over local features.
- We evaluate the detection performance of the proposed DGANet on our collected data set of 7040 ultrasound images and a public data set of breast ultrasound images (BUSIs). Compared with state-of-the-art methods, DGANet achieves better lesion detection in breast ultrasound images.

RELATED WORK

Breast lesion detection

Deep learning-based methods have made great progress in breast lesion detection in ultrasound images. These methods can be classified based on whether localization and classification of breast lesions are achieved simultaneously. For methods to locate breast lesions, Cao et al. (2017) and Yap et al. (2017) evaluate the performance of object detection methods based on deep learning for breast lesion localization in ultrasound images. Experimental results in Yap et al. (2017)

revealed that deep-learning methods achieve overall improvement compared with conventional methods. In addition, Yap et al. used Faster-RCNN with Inception-ResNet-v2 to locate breast lesions in ultrasound images and achieve comparable results. For methods to locate and classify lesions simultaneously, a fully convolutional network (FCN-AlexNet) is proposed to achieve breast lesion localization and recognition in ultrasound images (Yap et al. 2018).

FCN-AlexNet achieves localization and classification simultaneously based on semantic segmentation, but it requires lesion segmentation annotations, which increases the burden of training data collection. Additionally, Shin et al. (2019) proposed a joint weakly and semi-supervised deep learning method for breast lesion detection in ultrasound images, which achieves accurate lesion detection with less annotation effort.

Self-attention mechanisms

To capture context information and model long-range dependencies, many researchers have developed different self-attention mechanisms. Non-local network (NLNet) aggregates features at all positions by a weighted sum, with the weights defined by pairwise relation maps (Wang et al. 2018). The pairwise relation maps are dependent on query positions, which are determined by feature similarities between positions. In addition, many researchers propose extensions of NLNet to optimize specific tasks. Specifically, Yue et al. simplified NLNet and proposed a global context network (GCNet) (Cao et al. 2019, 2020). The global context features are calculated by the attention map, which is independent of query positions. GCNet significantly reduces computation cost, but background noise aggregated in the global context may limit the performance. Additionally, Jun and co-workers proposed a dual attention network for scene segmentation, which achieves state-of-the-art segmentation performance based on self-attention mechanisms (Fu et al. 2019).

METHODOLOGY

In this section, we first provide an overview of the DGANet and introduce the design of the BSAM and the GCAM, which enhance features in spatial and channel dimensions, respectively. Finally, we introduce image acquisition, implementation details and evaluation indexes of our experiments. Code is available at https://github.com/huimeng16/DGANet.

Overview

Breast lesions in ultrasound images are diverse in shape, scale and brightness because of the heterogeneity of breast cancer and the differences in imaging environments. Because convolution operations tend to generate a local receptive field, features of lesions in the same category may have differences. These differences increase the difficulty in locating and recognizing breast lesions in ultrasound images. To address this issue, we enhanced the discriminability of features based on attention mechanisms by mimicking the diagnostic patterns of radiologists. Specifically, we propose a dual global attention module (DGAM) to enhance features in spatial and channel dimensions, respectively. In the spatial dimension, we designed a BSAM to capture supporting information in neighboring region and suppress interference from distant noises. In the channel dimension, we proposed a GCAM to enhance features of important channels.

The overall framework of the DGANet is illustrated in Figure 2. The network structure of YOLOv3 (Farhadi and Redmon 2018) was adopted as the framework of the DGANet, and ResNet-101 (He et al. 2016) was chosen as the backbone. Additionally, three DGAMs are inserted into three residual blocks of the backbone, respectively. Noted that the DGAMs are added to right before the last residual blocks of three stages (Conv3_x, Conv4_x, and Conv5_x), which is inspired by the insertion mode of attention modules in GCNet (Cao et al. 2020). The DGAM is constructed of a BSAM and a GCAM (Fig. 3), which enhance features in spatial and channel dimensions, respectively. The detailed structures of the BSAM and the GCAM are described in the following sections.

Bilateral spatial attention module

Information on the neighboring regions of breast lesions in ultrasound images is essential for diagnosis because it reflects echo pattern and acoustic shadowing (Shan et al. 2016). However, not all the information in ultrasound images is positive for improving lesion detection. Specifically, the meaningless background noises such as speckle artifact may affect lesion detection in ultrasound images. To aggregate context in the neighboring region and suppress the interference of background noises, we propose a BSAM. The BSAM integrates the

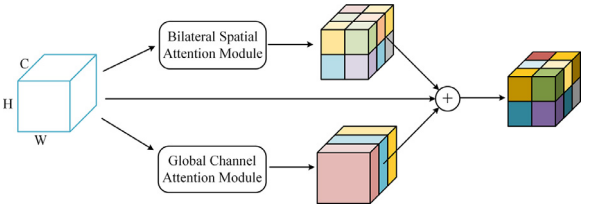


Fig. 3. Structure of the dual global attention module (DGAM). The features generated by the bilateral spatial attention module (BSAM) and global channel attention module (GCAM) are summed with the original input as the output of the DGAM. C = number of channels; H = height of input feature map; W = width of input feature map.

global information by a weighted sum, where the weight is calculated by multiplying feature similarities and physical Gaussian distance. With the help of the physical Gaussian distance, the BSAM can enhance feature integration in the neighboring region and weaken fusion of features given by distant noises.

As illustrated in Figure 4, the attention weight of the BSAM is calculated by multiplying the feature similarity matrix and Gaussian distance matrix. It is then normalized and multiplied by the input features to obtain the bilateral spatial attention. Given an input feature map $X \in R^{C\times H\times W}$, we use a 1 × 1 convolution layer W_k to aggregate channel information. Then, we reshape the aggregated information as $R^{(H*W)\times}$ 1 and apply a softmax layer to obtain the normalized feature similarity vector $M_{iv} \in R^{(H*W)\times}$ 1. Inspired by Kim et al. (2020), the absolute values of the aggregated features are applied before being fed into the softmax function. M_{iv} is calculated as

$$M_{iv} = \mathrm{SoftMax}\left(\left|W_k X\right|\right) \tag{1}$$

where $|\cdot|$ denotes the operator of absolute value. After that, M_{iv} is expanded as $R^{(H*w)\times(H*W)}$ to obtain the feature similarity matrix M_i. Each column in M_i represents feature similarities in the spatial dimension.

As for the Gaussian distance matrix, we calculate it based on physical spatial distances between positions. The Gaussian distance matrix is presented as

$$M_d = \begin{bmatrix} 1 & d_{1,2} & d_{1,3} & \cdots & d_{1,N} \\ d_{2,1} & 1 & d_{2,3} & \cdots & d_{2,N} \\ d_{3,1} & d_{3,2} & 1 & \cdots & d_{3,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{N,1} & d_{N,2} & d_{N,3} & \cdots & 1 \end{bmatrix} \tag{2}$$

where $N = H \times W$ is the number of positions in the spatial dimension. The diagonal elements of M_d are 1. $d_{i,j}$ represents the physical Gaussian distance between positions P_i and P_j and $i, j \in [1, N]$. $d_{i,j}$ is calculated as

$$d_{i,j} = e^{-\frac{(r_i-r_j)^2+(c_i-c_j)^2}{2\sigma^2}} \tag{3}$$

where r_i and c_i represent the row and column of position P_i, respectively. Similarly, r_j and c_j denote the row and column of position P_j, respectively. σ is Gaussian kernel parameter.

According to the computing methods of echo pattern and acoustic shadowing (Shan et al. 2016), the surrounding region containing the lesion in its center is about twice the size of the lesion. To capture context in the surrounding region, we define the Gaussian kernel parameter based on statistical information on lesion size. Considering the multiscale structure of the YOLOv3, we set different Gaussian kernel parameters in three DGAMs. To obtain statistical information on breast lesion size, we divide training samples into three groups (*i.e.,* <100, 100−200 and >200) based on lesion edge length calculated by the square root of lesion areas. For each group, we calculate the edge ratio of lesions to the corresponding ultrasound images

$$\mathrm{ratio} = \frac{\mathrm{EL}_{\mathrm{lesion}}}{\mathrm{EL}_{\mathrm{ultrasound\_image}}} \tag{4}$$

where $\mathrm{EL}_{\mathrm{lesion}}$ and $\mathrm{EL}_{\mathrm{ultrasound\_image}}$ are the edge lengths of lesion and ultrasound image, respectively. As illustrated in Figure 5, the medium edge ratios of three groups were 0.17, 0.33 and 0.52, respectively. Considering the characteristics of Gaussian function and calculation manner of BI-RADS features together, we set the Gaussian kernel parameters as half of

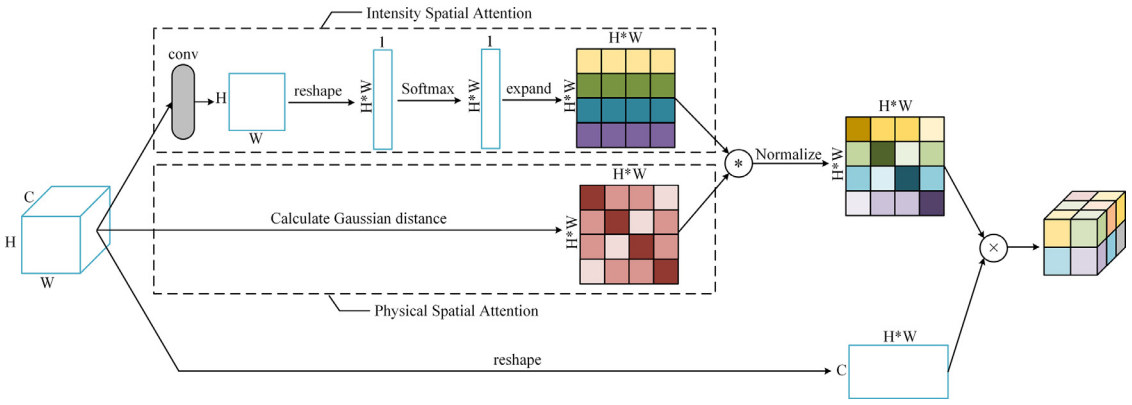


Fig. 4. Structure of the bilateral spatial attention module (BSAM). Its attention map is generated by multiplying feature similarity matrix and Gaussian distance matrix by elements. The output of the BSAM is obtained by matrix multiplication of normalized attention map and input features. C = number of channels; H = height of input feature map; W = width of input feature map.

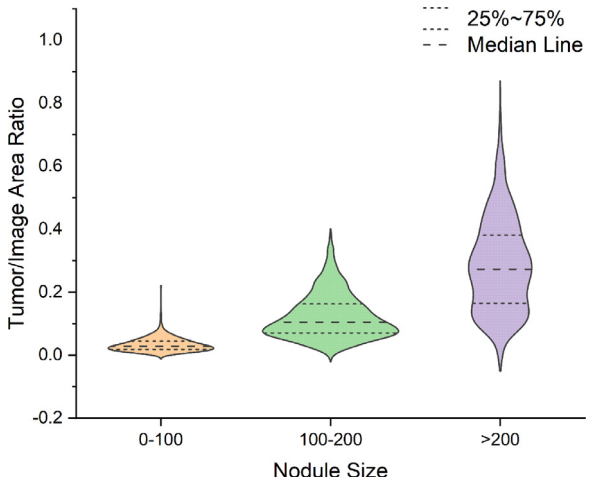


Fig. 5. Distribution of the edge ratios of lesions to the corresponding ultrasound images. The edge ratio is calculated by taking the square root of the area ratio. The *dotted lines* denote the median and quartile positions.

the lesion edges. The Gaussian kernel parameters are calculated as

$$\sigma = \frac{1}{2} * \mathrm{EL}_{\mathrm{lesion}} \tag{5}$$

Because the edge ratio of the lesion in the feature map to the corresponding feature map is the same as that in eqn (4), we set Gaussian kernel parameters as multiplication of side length of feature map and the edge ratio. The final definition of the Gaussian kernel parameter is

$$\sigma = \frac{1}{2} * \mathrm{ratio} * \mathrm{EL}_{\mathrm{feature_map}} \tag{6}$$

where $\mathrm{EL}_{\mathrm{feature\_map}}$ represents the edge length of the feature map. For the convenience of calculation, the Gaussian kernel parameters of three DGAMs in Conv3_x, Conv4_x and Conv5_x are set as 1/10, 1/6 and 1/4 of the edge length of the corresponding feature maps, respectively. Note that the ratio (*i.e.,* 1/10, 1/6 or 1/4) is set as a hyperparameter in the design of the BSAM. As illustrated in Figure 6, the larger the ratio of the Gaussian kernel parameter to the edge length of the corresponding feature map, the wider is the high-energy

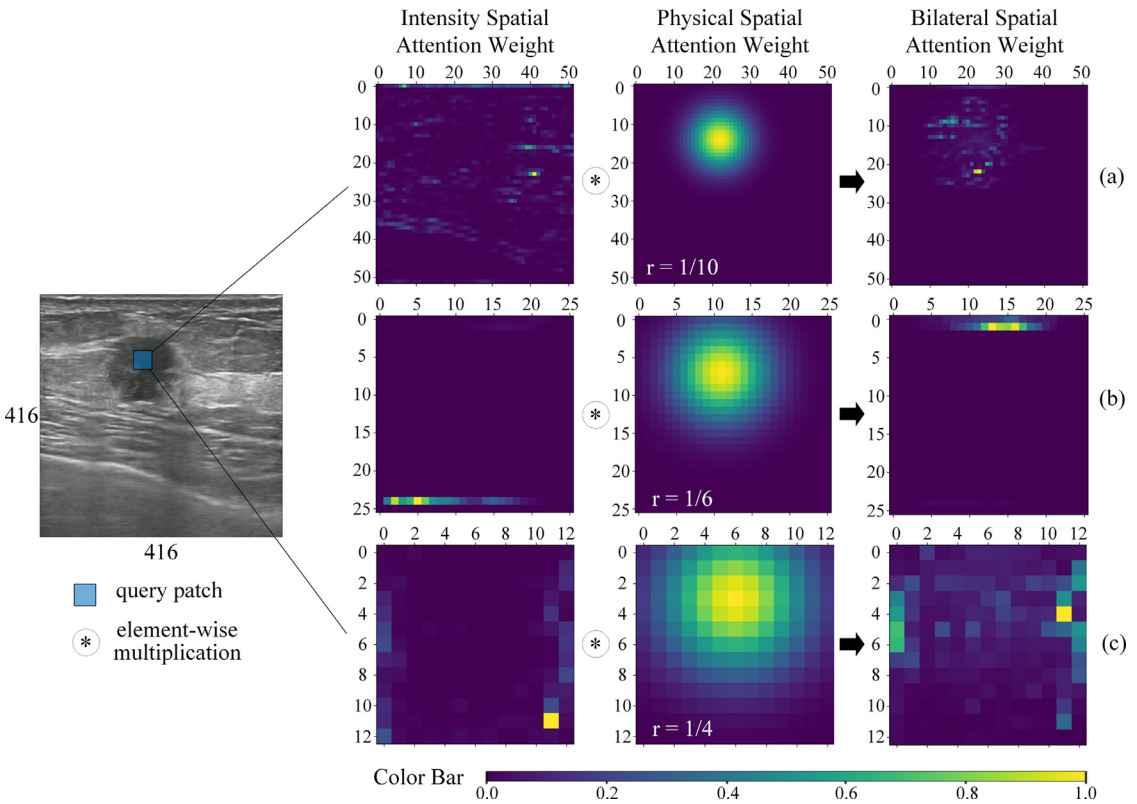


Fig. 6. Visualization of intensity spatial attention weight, physical spatial attention weight, and bilateral spatial attention weight in the bilateral spatial attention modules (BSAMs) at the given query patch (*blue*). (a–c) Present attention weight of the BSAMs in Conv3_x, Conv4_x, and Conv5_x, respectively. The edge length of the ultrasound image is 416. The edge length of weight maps in (a)–(c) is 52, 26 and 13, respectively. *r* defines the ratio of *a* to edge length of the corresponding feature map.

region in the physical spatial attention weight map. Additionally, the Gaussian distance can strengthen the weight of positions in the neighboring region, while weakening the weight of positions far from the query patch (Fig. 6a).

Global channel attention module

In deep learning-based breast lesion detection, each channel map of high-level features can be regarded as an abstract representation of a medical feature. According to BI-RADS (Liberman and Menell 2002), different medical features are associated with each other, and radiologists generally make diagnoses based on several important medical features. Therefore, we propose a GCAM to model channel-wise interdependencies and enhance features of important channels.

Figure 7 illustrates the structure of the GCAM. The global channel attention is calculated in two steps. Given a local feature map $X \in R^{C\times H\times W}$, we first use a 1×1 convolution layer W_t to aggregate channel information. Then, we reshape the aggregated information as $R^{N\times 1}$ and apply a softmax layer to generate the spatial attention map $S \in R^{N\times}$ 1:

$$S = \mathrm{SoftMax}(W_t X) \tag{7}$$

Meanwhile, we reshape X to $R^{C\times N}$. A spatial squeeze is performed by multiplying X and S, generating vector $z \in R^{C\times}$ 1, with its kth element:

$$z_k = \sum_i^N X_i^k S_i \tag{8}$$

This operation integrates the global spatial information into vector z. To model channel-wise interdependencies, we pass the vector z through two fully connected layers, ReLU operation (Nair and Hinton 2010) and sigmoid function (Menon et al. 1996). The transformed vector z_t is calculated as

$$z_t = \sigma(W_1(\delta(W_2 z))) \tag{9}$$

where $W_1 \in R^{C\times C/r}$ and $W_2 \in R^{C/r\times C}$ are weights of two fully connected layers, respectively. r is the bottleneck ratio, which is set as 16 in our experiments. $\delta(\cdot)$ and $\sigma(\cdot)$ represent the ReLU function and the sigmoid function, respectively.

Finally, we perform element-wise multiplication of the transformed vector with the original feature map to obtain the global channel attention $E \in R^{C\times H\times W}$

$$E = z_t \circ X \tag{10}$$

where $\circ$ denotes elementwise multiplication.

Image acquisition

To evaluate the performance of DGANet in breast lesion detection, we collect ultrasound images from our collaborators and construct a data set for breast lesion detection. This study has attained ethical approval from the institutional review board in all participating hospitals and waived the informed consent requirement. The data set consists of 7040 breast ultrasound images collected from 3759 patients, of whom 1476 have benign lesions and 2283 have malignant lesions. One or more images are collected from each patient, and each ultrasound image contains at least one lesion. The ultrasound images in the data set are acquired from different systems including those from Siemens, Philips, Aloka and Hitachi. Typical breast lesions are illustrated in Figure 1. For each lesion in ultrasound images, the ground-truth outline is drawn by experienced radiologists, such as the *yellow curve* in Figure 1. For the convenience of lesion detection, the bounding boxes of lesions (*red rectangle* in Fig. 1) are generated based on the ground-truth outlines automatically. Additionally, the ground-truth labels (*i.e.,* malignant/benign) of breast lesions are determined using fine-needle aspiration (FNA) biopsy and pathological examination. Furthermore, to validate the performance of different network models, the data set is divided into three groups (*i.e.,* training set, validation set and test set) in the ratio of 6:2:2. In particular, ultrasound

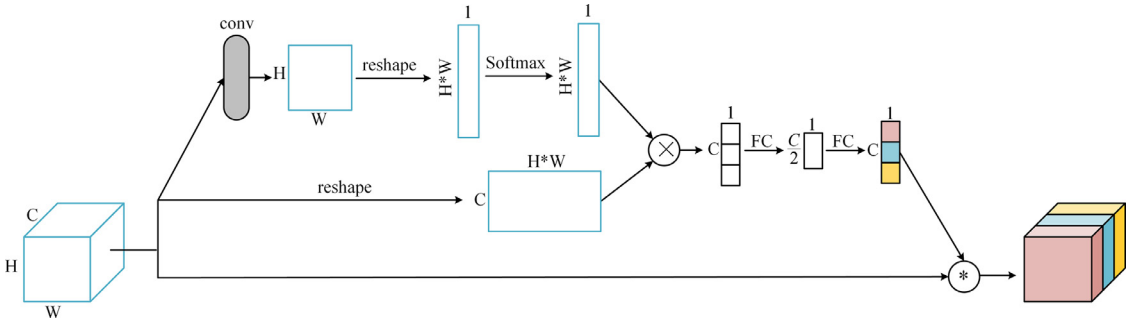


Fig. 7. Structure of global channel attention module (GCAM). The global channel information is calculated by multiplying input features and spatial weights. Fully connected layers are used to learn channel-wise interdependencies. The input features are weighted by channel attention weight to generate the output of the GCAM.

Table 1. Distributions of patients and images

Set	Patients	Images
Training	MN: 1389	MN: 2112
	BN: 891	BN: 2112
Validation	MN: 465	MN: 704
	BN: 300	BN: 704
Test	MN: 429	MN: 704
	BN: 285	BN: 704

MN = malignant lesion; BN = benign lesion.

images belonging to the same patient are divided into the same group. The data distribution of the three groups is provided in Table 1. To evaluate the performance of different methods in detecting breast lesions of different size, the test set is divided into three groups (<100, 100−200, >200) based on the edge length of breast lesions.

Implementation details and evaluation index

Similar to the original YOLOv3 (Farhadi and Redmon 2018), our network uses anchors to predict region proposals for breast lesion detection. On the basis of the training data, we determine our bounding box priors using *k*-means clustering (Redmon and Farhadi 2017). The sizes of nine anchors are chosen: (50 × 31), (75 × 47), (111 × 61), (113 × 90), (155 × 114), (169 × 80), (220 × 125), (274 × 175) and (444 × 238). Considering fundamental differences between object detection in natural images and breast lesion detection in ultrasound images (Raghu et al. 2019), we initialize all parameters of our network by random noise rather than use the pre-trained backbone network on ImageNet (Farhadi and Redmon, 2018). To alleviate the overfitting problem and improve the generalization ability of our network, affine transformation, horizontal flipping, color jittering and multi-scale training are adopted. The SGD algorithm (Ketkar 2017) is used to optimize our network with a weight decay of 0.0005, a momentum of 0.9, a batch size of 8 and 500 epochs. The initial learning rate is set at 0.002. The loss function of our network consists of classification loss, confidence loss and bounding box loss. Cross-entropy loss (Farhadi and Redmon 2018) is used to calculate classification loss and confidence loss. Different from YOLOv3, GIoU loss (Rezatofighi et al. 2019) is adopted to calculate bounding box loss. Additionally, we implement our network using the Pytorch library and conduct all experiments on a single NVIDIA Tesla V100 GPU.

To quantitatively assess the performance of breast lesion detection using different methods, mean average precision (mAP) (Redmon and Farhadi 2017), recall and precision are chosen as the evaluation indexes. Recall reflects how many lesions have been detected, whereas precision denotes how many lesions have been correctly classified. Recall R and precision P are calculated as

$$R = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{11}$$

$$P = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \tag{12}$$

where TP (true positives) represents the number of lesions detected with correct labels, FN (false negatives) is the number of undetected lesions and FP (false positives) denotes the number of detected lesions with wrong labels. The mAP is calculated using precision−recall curves, and the threshold of intersection over union (IoU) used in our experiments is 0.5.

Results

In this section, we assess the performance of DGANet using our collected data set and a public data set of breast ultrasound images (BUSIs) (Al-Dhabyani et al. 2019). This section is arranged as follows: First, an ablation study for attention modules is implemented to verify the effectiveness of the BSAM and GCAM, and the BSAMs with different ratio parameter settings are evaluated. Second, we evaluate our DGANet on different backbones by replacing ResNet-101 with ResNet-50 (He et al. 2016) and Darknet-53 (Farhadi and Redmon 2018). Third, the performance of the DGAM is compared with that of other attention modules. Last, we compare our method with existing detection methods on our collected data set and the public BUSIs.

Ablation study for attention modules

The ablation study is implemented first to evaluate the effectiveness of the components of the DGAM, that is, the BSAM and GCAM. The ablation study is conducted mainly on our collected data set. The baseline of the ablation study is the original YOLOv3 network with ResNet-101 as the backbone.

Figure 8 illustrates the detection results for four samples (S1−S4) obtained by different networks. Although all networks could detect lesions in ultrasound images, networks with attention modules obtain higher confidence than the baseline. It should be noted that the DGANet obtains significantly higher confidence than the network with the GCAM (GCAM-Net) in S3, which indicates the effectiveness of the BSAM. In lesion detection of S2, the DGANet improves confidence by 3.7% compared with that of the network with the BSAM (BSAM-Net). These results suggest that the BSAM and GCAM benefit the localization and recognition of breast lesions in ultrasound images.

The quantitative results for the different networks are outlined in Table 2. Compared with the baseline, employing the GCAM individually yields a result of 0.829 in total mAP, which represents 3.2%

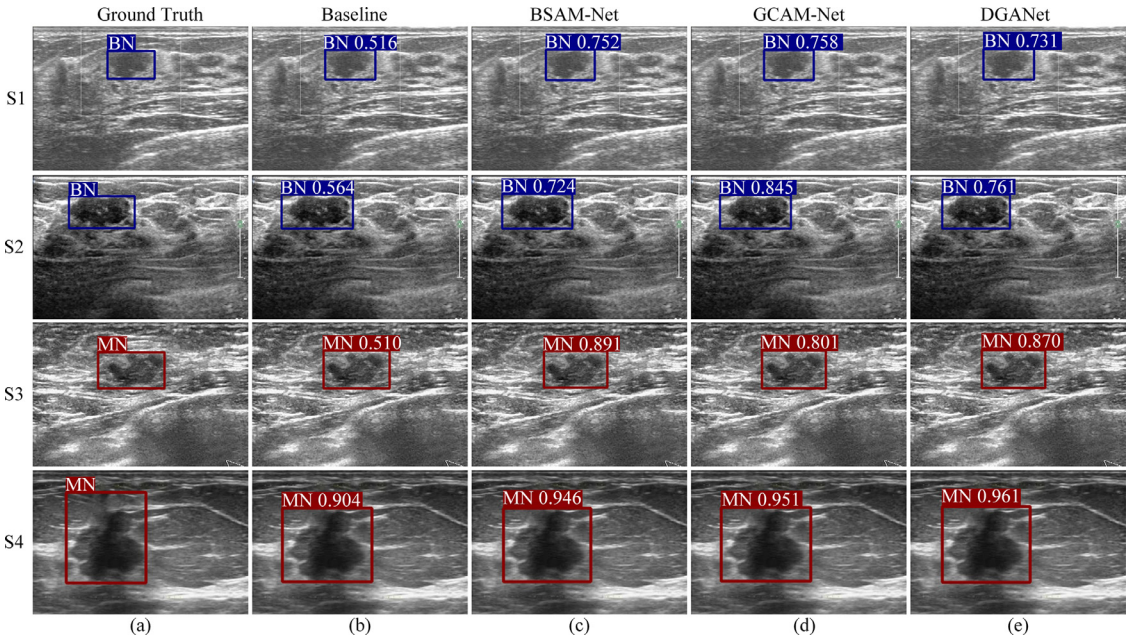


Fig. 8. Detection results of the four samples in our data set (S1−S4) in the ablation study. (a, b) The ground-truth regions of interest (a) and detection results (b) of the baseline. (c, d) Detection results for networks with part of the dual global attention module. (e) Detection results for the dual global attention neural network (DGANet). Numbers in ultrasound images represent the confidence of detection boxes. BN = benign lesion; MN = malignant lesion.

improvement. Additionally, integration of BSAM and GCAM (*i.e.,* DGANet) yields the highest total mAP (0.840), the highest total precision (0.754) and the highest total recall (0.845), which outperform the baseline by 4.3%, 5.0% and 6.3%, respectively. Furthermore, the DGANet significantly improves the precision and recall of medium lesion (100−200) detection compared with the BSAM-Net. These comparisons illustrate that our attention modules have the potential to improve the accuracy of breast lesion detection in ultrasound images.

Table 2. Effectiveness of BSAM and GCAM in our method

Attention module		Size	mAP	Precision	Recall
BSAM	GCAM				
		All	0.797	0.704	0.782
		<100	0.687	0.655	0.692
		100−200	0.863	0.749	0.855
		>200	0.802	0.681	0.781
√		All	0.822	0.744	0.833
		<100	0.765	**0.722**	0.776
		100−200	0.845	0.765	0.857
		>200	**0.866**	0.724	**0.862**
	√	All	0.829	**0.754**	0.834
		<100	0.770	0.721	**0.789**
		100−200	0.860	0.782	0.859
		>200	0.845	**0.737**	0.849
√	√	All	**0.840**	**0.754**	**0.845**
		<100	**0.775**	0.710	**0.789**
		100−200	**0.872**	**0.790**	**0.875**
		>200	0.853	0.727	0.853

BSAM = bilateral spatial attention module; GCAM = global channel attention module; mAP = mean average precision.

Furthermore, we compare the performance of DGANet with different ratio settings in the three BSAMs. Table 3 outlines the quantitative results of the DGANet with six groups of ratio settings (groups 1−6). The ratios in group 1 are obtained based on the statistical information of breast lesion size and diagnostic experience of radiologists. The ratios in groups 2−6 are generated from the ratios in group 1. As outlined in Table 3, the DGANet with ratio settings in group 1 obtains the highest recall (0.845), which is 0.3%−1.9% higher than those of the other models. Additionally, the DGANet with ratio settings in group 1 achieved the highest mAP (0.840), which outperforms other models by 1.2%−1.8%. These results verify that the ratio setting in group 1 is suitable for breast lesion detection. Furthermore, all these results illustrate that incorporating medical domain knowledge into hyperparameter design is an effective approach.

Experiments on different backbones

To evaluate the improvement in performance of DGANet over different backbones, we replaced ResNet-101 with ResNet-50 (He et al. 2016) and Darknet-53 (Farhadi and Redmon 2018) and implemented experiments on our collected data set. Similar to our network, three DGAMs are inserted immediately before the last residual block of stage Conv3_x, Conv4_x, and Conv5_x

Table 3. Quantitative results for DGANet with different ratio settings

Group	Conv_3 Ratio	Conv_4	Conv_5	mAP	Precision	Recall
1	1/10	1/6	1/4	**0.840**	0.754	**0.845**
2	1/5	1/3	1/2	0.824	0.754	0.842
3	1/20	1/12	1/8	0.823	0.757	0.829
4	1/10	1/10	1/10	0.828	0.755	0.828
5	1/6	1/6	1/6	0.822	0.744	0.826
6	1/4	1/4	1/4	0.827	**0.758**	0.833

DGANet = dual global attention neural network; mAP = mean average precision.

in ResNet-50, respectively. The Gaussian kernel parameters in Conv3_x, Conv4_x, Conv5_x are set as 1/10, 1/6, and 1/4 of the side length of the corresponding feature maps, respectively. For Darknet-53, three DGAMs are added to right before the last residual block of res8, res8, res4, respectively. The Gaussian kernel parameters of the DGAMs in res8, res8 and res4 are set as 1/10, 1/6 and 1/4 of the side lengths of the corresponding feature maps, respectively.

The quantitative results of networks with different backbones are outlined in Table 4. The detection results are close among the networks with original Darknet-53, ResNet-50 and ResNet-101. Compared with the networks with original backbones, networks with the DGAMs significantly improve detection performance. In the networks with Darknet-53 as the backbone, the DGAMs yield 2.8%, 5.4% and 4.9% improvement in total mAP, total precision and total recall, respectively.

Table 4. Effectiveness of different backbones and DGAM on our method

Backbone	DGAM	Size	mAP	Precision	Recall
Darknet-53		All	0.804	0.724	0.795
		<100	0.688	0.671	0.683
		100−200	0.869	0.769	**0.877**
		>200	0.828	0.699	0.814
Darknet-53	✓	All	0.832	**0.778**	0.844
		<100	**0.789**	**0.756**	0.798
		100−200	0.855	**0.796**	0.867
		>200	0.848	**0.760**	0.860
ResNet-50		All	0.807	0.725	0.800
		<100	0.694	0.682	0.700
		100−200	**0.872**	0.757	0.865
		>200	0.819	0.695	0.831
ResNet-50	✓	All	0.829	0.763	0.843
		<100	0.777	0.746	**0.799**
		100-200	0.867	0.776	0.869
		>200	0.844	0.749	**0.862**
ResNet-101		All	0.797	0.704	0.782
		<100	0.687	0.655	0.692
		100-200	0.863	0.749	0.855
		>200	0.802	0.681	0.781
ResNet-101	✓	All	**0.840**	0.754	**0.845**
		<100	0.775	0.710	0.789
		100-200	**0.872**	0.790	0.875
		>200	**0.853**	0.727	0.853

DGAM = dual global attention module; mAP = mean average precision.

Additionally, the network with ResNet-50 and the DGAM outperforms the baseline by 2.2%, 3.8% and 4.3% in total mAP, total precision and total recall, respectively. Furthermore, the network with ResNet-101 and the DGAMs achieves the highest total mAP (0.840) and the highest total recall (0.845), which are 4.3% and 6.3% higher than the baseline values. These comparisons suggest the DGAMs significantly improve performance in detecting breast lesions.

Comparison with different attention modules

The detection experiments for networks with different attention modules are implemented to verify the superiority of the DGAM. The networks are constructed by replacing the DGAM in the DGANet with a squeeze-and-excitation (SE) block (Hu et al. 2018), convolutional block attention module (CBAM) (Woo et al. 2018), non-local (NL) block (Wang et al. 2018), global context (GC) block (Cao et al. 2020) and dual attention (DA) block (Fu et al. 2019), respectively. For each attention mechanism, three attention modules are inserted immediately before the last residual block of stage Conv_3, Conv_4, and Conv_5 in ResNet-101, respectively. The parameters of attention modules are set as the default values described in Hu et al. (2018), Wang et al. (2018), Woo et al. (2018), Fu et al. (2019) and Cao et al. (2020). All networks are trained and evaluated on our collected training set and test set, respectively.

Table 5 outlines the quantifications of mAP, precision and recall given by different networks. The original YOLOv3 with ResNet-101 as the backbone is used as the baseline. Compared with the baseline, all attention modules yield improvement in total mAP, total precision and total recall, respectively. The networks with the SE block, CBAM or DA block obtain close mAP values in all lesion detection. In addition, the networks with NL or GC blocks achieve close total mAP. Furthermore, the DGANet achieves the highest total mAP (0.840), which is 1.0%−4.3% higher than those of the other networks. All these comparisons validate the superiority of DGANet in breast lesion detection.

Table 5. Quantitative comparisons of different attention modules

Attention module	Size	mAP	Precision	Recall
	All	0.797	0.704	0.782
	<100	0.687	0.655	0.692
	100−200	0.863	0.749	0.855
	>200	0.802	0.681	0.781
SE	All	0.829	0.750	0.836
	<100	0.775	0.715	0.791
	100−200	0.861	0.765	0.860
	>200	0.840	**0.755**	0.848
CBAM	All	0.829	0.743	0.836
	<100	0.766	0.708	0.781
	100−200	0.857	0.771	0.856
	>200	**0.859**	0.721	**0.870**
NL	All	0.823	0.748	0.843
	<100	0.775	0.730	**0.804**
	100−200	0.855	0.766	0.863
	>200	0.819	0.719	0.851
GC	All	0.821	0.755	0.828
	<100	0.774	0.719	0.779
	100−200	0.848	0.778	0.855
	>200	0.834	0.738	0.835
DA	All	0.830	**0.756**	0.845
	<100	**0.779**	**0.734**	0.802
	100−200	0.861	0.780	0.878
	>200	0.827	0.720	0.847
DGAM	All	**0.840**	0.754	0.845
	<100	0.775	0.710	0.789
	100−200	**0.872**	**0.790**	0.875
	>200	0.853	0.727	0.853

CBAM = convolutional block attention module; DA = dual attention block; DGAM = dual global attention module; GC = global context block; NL = non-local block; SE = squeeze and excitation block.

Comparison with state-of-the-art networks

To further assess the performance of the DGANet, we compare our network with Faster R-CNN (Ren et al. 2017), YOLOv3 (Farhadi and Redmon 2018), RetinaNet (Lin et al. 2017), YOLOv5 and YOLOX (Ge et al. 2021) by conducting detection experiments on our collected data set. Anchors used in all networks are generated with the *k*-means clustering algorithm based on training data. For statistical analysis, we implement five experiments for each comparison method. On the basis of data division in Table 1, we further divide the training set into three groups based on the ratio of 1:1:1. Thus, we obtain five parts of our collected data set. In the five experiments, each part of our collected data set is set as test data in turn, and remaining data are set as training data and validation data in the ratio of 3:1.

For quantitative analysis, mAP, recall and precision of the different methods are calculated and listed in Table 6. The mean ± standard deviation (SD) of the mAP given by the DGANet is 0.831 ± 0.011, which is significantly higher than that of the YOLOv3 ($p < 0.05$), the RetinaNet ($p < 0.01$), the Faster R-CNN ($p < 0.01$) and the YOLOX ($p < 0.01$). Note that the mean recall obtained by the Faster R-CNN (0.831) and the YOLOX (0.856) are close to that of the DGANet (0.841), but the DGANet achieves significantly higher precision than the Faster R-CNN ($p < 0.001$) and YOLOX ($p < 0.001$). These results indicate that the Faster R-CNN and

Table 6. Quantitative results for different methods in our collected data set

Method	Size	mAP	Precision	Recall
YOLOv3	All	0.812 ± 0.008	0.737 ± 0.013	0.806 ± 0.010
	<100	0.711 ± 0.023	0.694 ± 0.026	0.698 ± 0.028
	100−200	0.863 ± 0.013	0.758 ±0.020	0.864 ± 0.015
	>200	0.846 ± 0.024	0.751 ± 0.032	0.854 ± 0.026
RetinaNet	All	0.772 ± 0.019	0.736 ± 0.009	0.771 ± 0.030
	<100	0.667 ± 0.041	0.685 ± 0.018	0.696 ± 0.043
	100−200	0.813 ± 0.018	0.758 ± 0.014	0.801 ± 0.024
	>200	0.813 ± 0.036	0.757 ± 0.028	0.821 ± 0.038
FasterR-CNN	All	0.784 ± 0.018	0.597 ± 0.023	0.831 ± 0.005
	<100	0.669 ± 0.034	0.550 ± 0.030	0.774 ± 0.013
	100−200	0.827 ± 0.021	0.619 ± 0.014	0.854 ± 0.011
	>200	0.831 ± 0.029	0.615 ± 0.029	0.867 ± 0.025
YOLOv5	All	0.829 ± 0.009	0.786 ± 0.014	0.796 ± 0.016
	<100	0.814 ± 0.012	0.792 ± 0.026	0.761 ± 0.030
	100−200	0.839 ± 0.023	0.793 ± 0.030	0.812 ± 0.028
	>200	0.845 ± 0.029	0.803 ± 0.033	0.803 ± 0.047
YOLOX	All	0.796 ± 0.006	0.372 ± 0.073	0.856 ± 0.028
	<100	0.758 ± 0.022	0.346 ± 0.066	0.842 ± 0.034
	100−200	0.813 ± 0.008	0.386 ± 0.075	0.865 ± 0.032
	>200	0.825 ± 0.018	0.389 ± 0.081	0.865 ± 0.030
DGANet	All	0.831 ± 0.011	0.762 ± 0.011	0.841 ± 0.010
	<100	0.780 ± 0.028	0.727 ± 0.030	0.788 ± 0.025
	100−200	0.859 ± 0.017	0.781 ± 0.015	0.866 ± 0.021
	>200	0.845 ± 0.025	0.763 ± 0.027	0.863 ± 0.012

DGANet = dual global attention neural network; mAP = mean average precision.

YOLOX generate more wrong detection boxes, whereas the DGANet achieves more accurate location and classification. Additionally, the mAP given by the YOLOv5 (0.829) is close to that of the DGANet (0.831), and there is no significant difference between the two methods ($p = 0.860$). These comparisons suggest the DGANet has the capability to accurately detect breast lesions in ultrasound images.

In addition to the quantitative results, the ground truth and detection results of four samples (S1−S4) are illustrated in Figure 9. Consistent with the quantitative results, the Faster R-CNN generates some bounding boxes without lesions other than right bounding boxes with lesions in S1, S3 and S4 (Fig. 9d). Similarly, a bounding box without a lesion is detected by the RetinaNet and the YOLOX in S3 (Fig. 9c, 9f). Also, the lesion in S2 is assigned two category labels by the RetinaNet, which reduces the precision of the RetinaNet. Compared with the Faster R-CNN and RetinaNet, the YOLOv3 and DGANet generate fewer detection boxes without lesions. However, the confidence of bounding boxes given by the DGANet is significantly higher than that generated by the YOLOv3. These results illustrate that the DGANet achieves accurate location and classification of breast lesions in ultrasound images.

Furthermore, we performed five-fold cross-validation experiments on our collected data set. To generate training data and validation data, 7040 samples were randomly divided into five groups. Note that samples belonging to the same patient were divided into the same groups. Quantitative results given by different methods are listed in Table 7. The mean ± SD of the mAP given by the DGANet is 0.836 ± 0.004, which is significantly higher than that of the YOLOv3 ($p < 0.05$), the RetinaNet ($p < 0.001$), the Faster R-CNN ($p < 0.001$) and the YOLOX ($p < 0.05$). Additionally, the mAP achieved by the YOLOv5 is close to that of the DGANet ($p = 0.217$). These results further reveal the superior performance of the DGANet in breast lesion detection in ultrasound images.

To further evaluate the performance of the DGANet in breast lesion detection in ultrasound images, we implemented breast lesion detection experiments on a public data set of BUSIs (Al-Dhabyani et al. 2019); 437 benign samples and 210 malignant samples in the public data set were used for breast lesion detection. Considering the amount of data in the public data set is small, five-fold cross validation experiments were implemented to assess the performance of different methods. The quantitative results of different methods are outlined in Table 8. The mean ± SD of the mAP obtained by the DGANet is 0.778 ± 0.014, which is close to that of the YOLOv5 ($p = 0.763$). Additionally, the mAP generated by the DGANet outperforms those of the YOLOv3, the RetinaNet, the Faster R-CNN and the YOLOX by 7.0%, 3.2%, 1.7% and 3.8%, respectively. Furthermore, the SD of the mAP generated by the DGANet is the smallest (0.014), which demonstrates the superiority of the DGANet with respect to robustness.

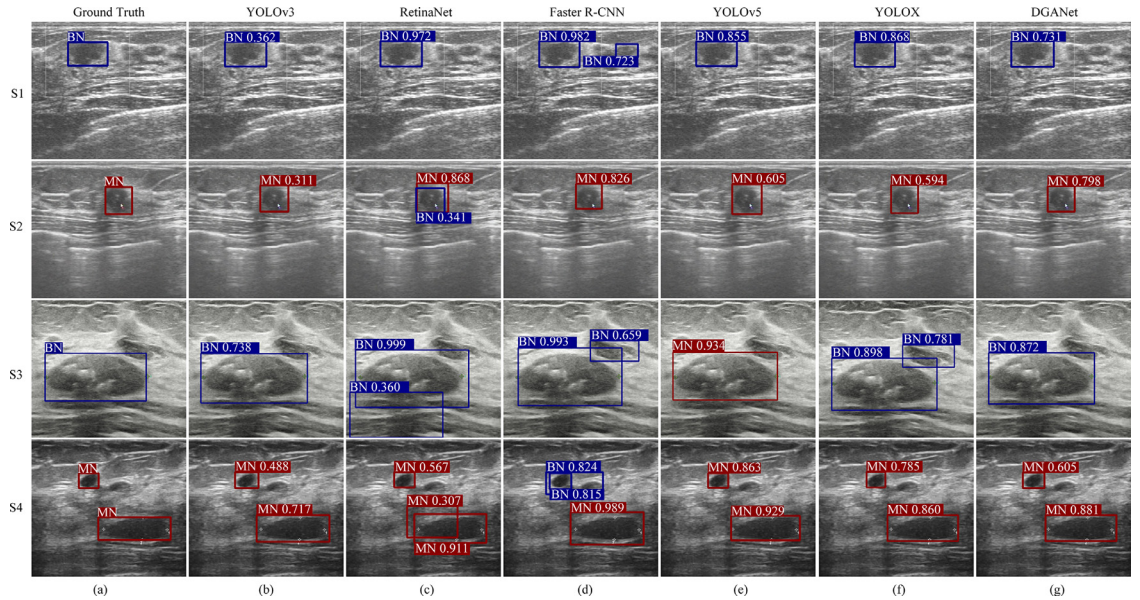


Fig. 9. Detection results for the four samples in our collected data set (S1−S4) given by different methods. (a) Ground-truth regions of interest. (b−f) Detection results for YOLOv3, RetinaNet, Faster R-CNN, YOLOv5, YOLOX and the dual global attention neural network (DGANet). BN and MN denote benign and malignant lesion, respectively. Numbers in ultrasound images represent the confidence of detection boxes.

Table 7. Quantitative results for different methods in five-fold cross validation experiments

Method	mAP	Precision	Recall
YOLOv3	0.819 ± 0.009	0.745 ± 0.012	0.806 ± 0.012
RetinaNet	0.785 ± 0.008	0.744 ± 0.011	0.765 ± 0.016
Faster R-CNN	0.793 ± 0.014	0.614 ± 0.012	0.829 ± 0.023
YOLOv5	0.843 ± 0.011	0.823 ± 0.024	0.802 ± 0.012
YOLOX	0.805 ± 0.012	0.370 ± 0.105	0.868 ± 0.042
DGANet	0.836 ± 0.004	0.749 ± 0.012	0.850 ± 0.007

DGANet = dual global attention neural network; mAP = mean average precision.

DISCUSSION

Here we proposed a novel DGANet to achieve accurate breast lesion detection in ultrasound images. Based on the domain knowledge that both lesion regions and the areas surrounding the lesions are considered by doctors in the diagnosis of breast lesions, we created a BSAM. With the help of the BSAM, the DGANet integrates the supporting context in neighboring region and suppresses the interference from distant noises. In deep learning-based breast lesion detection in ultrasound images, each channel map of high-level features can be regarded as an abstract representation of a medical feature. Considering that radiologists generally make diagnoses on the basis of several important medical features, we built the GCAM to capture the channel-wise interdependencies and enhance features of important channels.

To assess the performance of the DGANet, we implemented breast lesion detection experiments on our collected data set and the public BUSIs. First, we evaluated the effectiveness of the DGAM in ablation study and comparison experiments with different backbones. The results indicated that networks with the DGAM outperform the baseline, and the DGANet achieves the highest total mAP (0.840). Additionally, the superiority of the DGAM is further verified by comparison with the CBAM, SE, NL, GC, and DA blocks. The DGAM achieves the highest total mAP (0.840), which is 1.0%−1.9% higher than those of other attention modules. Furthermore, we compared our network with YOLOv3, RetinaNet, Faster R-CNN, YOLOv5 and YOLOX. The DGANet significantly outperformed YOLOv3, RetinaNet, Faster R-CNN and YOLOX with a dominant advantage. These results verify that DGANet achieves accurate lesion detection in breast ultrasound images.

Table 8. Quantitative results for different methods in breast lesion detection on public breast ultrasound images

Method	mAP	Precision	Recall
YOLOv3	0.708 ± 0.032	0.736 ± 0.035	0.644 ± 0.045
RetinaNet	0.746 ± 0.041	0.748 ± 0.038	0.785 ± 0.017
Faster R-CNN	0.761 ± 0.036	0.709 ± 0.059	0.826 ± 0.019
YOLOv5	0.782 ± 0.030	0.808 ± 0.039	0.732 ± 0.013
YOLOX	0.740 ± 0.024	0.521 ± 0.034	0.751 ± 0.028
DGANet	0.778 ± 0.014	0.730 ± 0.019	0.779 ± 0.032

DGANet = dual global attention neural network; mAP = mean average precision.

To our best knowledge, this is the first study integrating domain knowledge into attention mechanisms to enhance features for breast lesion detection in ultrasound images. Benefitting from the superiority of DGAM, DGANet holds great potential for small lesion detection. Additionally, DGANet is constructed under the YOLOv3 framework, which achieves detection of breast lesions in a short time. Our comparison experiments illustrate the feasibility of using DGANet for breast lesion detection in ultrasound images.

There are, however, some limitations to the DGANet. The major drawback of the DGANet is the calculation of Gaussian distance matrix, which is time-consuming and redundant. Specifically, the calculation of Gaussian distances between pixels in the background might be useless for lesion detection. Additionally, the Gaussian kernel parameter of the DGAM is defined manually based on the domain knowledge. A trainable strategy for selecting the Gaussian kernel parameter is necessary to further improve the performance of the DGANet, and will be studied in future work. Furthermore, there were more patients with malignant lesions than those with benign lesions in our collected data set, which induces a category imbalance problem to some extent. A solution to this problem of category imbalance will be studied in our future work.

CONCLUSIONS

A novel DGANet is proposed for lesion detection in breast ultrasound images. Compared with the existing detection networks, it achieves more accurate detection in both location and classification. We believe that this method holds great potential for improving efficiency of breast lesion detection in ultrasound images.

Acknowledgments—We thank the doctors from the cooperative hospital for their help in data annotation and experience sharing. We also thank Ningbo Gu and Xiaozheng Xie for their valuable suggestions. This article was supported in part by the National Natural Science Foundation of China under Grant Nos. 61976012, 62273009 and 61772060, in part by the National Key R&D Program of China under Grant No. 2017YFB1301100, in part by China Education and Research Network Innovation Project under Grant No. NGII20170315, and in part by the Postdoctoral Research Startup Fund of Hangzhou Innovation Institute, Beihang University, under Grant No. 2020Y4A004.

REFERENCES

Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data Brief 2019;28 104863.

Alcantara D, Leal MP, Garca-Bocanegra I, Garca-Martin ML. Molecular imaging of breast cancer: Present and future directions. Front Chem 2014;2:112.

Cao Z, Duan L, Yang G, Yue T, Chen Q, Fu H, Xu Y. Breast tumor detection in ultrasound images using deep learning. In: Wu G, Munsell B, Zhan Y, Bai W, Sanroma G, Coupé P, (eds). Patch-based Techniques in Medical Imaging. Patch-MI 2017. Cham: Springer; 2017. Lecture Notes Comput Sci10530:121−128.

Cao Y, Xu J, Lin S, Wei F, Hu H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). New York: IEEE1971–1980.

Cao Y, Xu J, Lin S, Wei F, Hu H. Global context networks. IEEE Trans Pattern Anal Mach Intell 2020; Published online December 24.

Farhadi A, Redmon J. Yolov3: An incremental improvement. Computer Vision and Pattern Recognition. Springer Berlin/Heidelberg: Springer; 2018 1804−1802.

Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H. Dual attention network for scene segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE3146–3154.

Ge Z, Liu S, Wang F, Li Z, Sun J. Yolox: Exceeding YOLO series in 2021. : arXiv preprint; 2021 https://arxiv.org/abs/2107.08430.

Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. Annu Rev Biomed Eng 2013;15:327–357.

Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y. Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. arXiv preprint arXiv:1801.09927, 2018.

Hagerty JR, Stanley RJ, Almubarak HA, Lama N, Kasmi R, Guo P, Drugge RJ, Rabinovitz HS, Oliviero M, Stoecker WV. Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images. IEEE J Biomed Health Inform 2019;23:1385–1391.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE; 2016. p. 770–778.

Hu J, Shen L, Sun G. Squeeze-and-excitation networks. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE; 2018. p. 7132–7141.

Hussein S, Cao K, Song Q, Bagci U. Risk stratification of lung nodules using 3D CNN-based multi-task learning. Cham: Springer; 2017.

Ketkar N. Stochastic gradient descent. Deep learning with Python. Cham: Springer; 2017. p. 113–132.

Kim J, El-Khamy M, Lee JT-GSA. Transformer with Gaussian-weighted self-attention for speech enhancement. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE6649–6653.

Li L, Xu M, Liu H, Li Y, Wang X, Jiang L, Wang Z, Fan X, Wang N. A large-scale database and a CNN model for attention-based glaucoma detection. IEEE Trans Med Imaging 2019;39:413–424.

Liao WX, He P, Hao J, Wang XY, Yang RL, An D, Cui LG. Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model. IEEE J Biomed Health Inform 2020;24:984–993.

Liberman L, Menell JH. Breast imaging reporting and data system (birads). Radiol Clin 2002;40:409–430.

Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. IEEE International Conference on Computer Vision. New York: IEEE2980–2988.

Menon AR, Mehrotra K, Mohan CK, Ranka S. Characterization of a class of sigmoid functions with applications to neural networks. Neural Networks 1996;9:819–835.

Murthy V, Hou L, Samaras D, Kurc TM, Saltz JH. Center-focusing multitask CNN with injected features for classification of glioma nuclear images. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). New York: IEEE834–841.

Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. 27th International Conference on International Conference on Machine Learning. IMLS807–814.

Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding transfer learning for medical imaging. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, BC, Canada.

Redmon J, Farhadi A. Yolo9000: Better, faster, stronger. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE7263–7271.

Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137–1149.

Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE.

Saba T, Sameh Mohamed A, El-Affendi M, Amin J, Sharif M. Brain tumor detection using fusion of hand crafted and deep learning features. Cogn Syst Res 2020;59:221–230.

Shan J, Alam SK, Garra B, Zhang Y, Ahmed T. Computer-aided diagnosis for breast ultrasound using computerized bi-rads features and machine learning methods. Ultrasound Med Biol 2016;42:980–988.

Shin SY, Lee S, Yun ID, Kim SM, Lee KM. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. IEEE Trans Med Imaging 2019;38:762–774.

Tan J, Huo Y, Liang Z, Li L. Expert knowledge-infused deep learning for automatic lung nodule detection. J X-Ray Sci Technol 2019;27:17–35.

Wang X, Girshick R, Gupta A, He K. Non-local neural networks. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE7794–7803.

Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, (eds). Computer Vision—ECCV 2018. Cham: Springer; 2018. Lecture Notes Comput Sci 2018;11211:3−19.

Xing J, Chen C, Lu Q, Cai X, Yu A, Xu Y, Xia X, Sun Y, Xiao J, Huang L. Using BI-RADS stratifications as auxiliary information for breast masses classification in ultrasound images. IEEE J Biomed Health Inform 2021;25:2058–2070.

Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, Davison AK, Marti R. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform 2017;22:1218–1226.

Yap MH, Goyal M, Osman F, Ahmad E, Mart R, Denton E, Juette A, Zwiggelaar R. End-to-end breast ultrasound lesions recognition with a deep learning approach. Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. Proc SPIE. 10578, 1057819.

Yap MH, Goyal M, Osman F, Marti R, Denton E, Juette A, Zwiggelaar R. Breast ultrasound region of interest detection and lesion localisation. Artif Intell Medic 2020;107 101880.