Xie XZ, Niu JW, Liu XF *et al.* DG-CNN: Introducing margin information into convolutional neural networks for breast cancer diagnosis in ultrasound images. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 37(2): 277–294 Mar. 2022. DOI 10.1007/s11390-020-0192-0

DG-CNN: Introducing Margin Information into Convolutional Neural Networks for Breast Cancer Diagnosis in Ultrasound Images

Xiao-Zheng Xie[1] (解晓政), Jian-Wei Niu[1,2] (牛建伟), *Senior Member*, *IEEE*, Xue-Feng Liu[1,*] (刘雪峰)
Qing-Feng Li[2] (李青锋), Yong Wang[3] (王　勇), Jie Han[3] (韩　洁), and
Shaojie Tang[4] (唐少杰), *Member*, *IEEE*

[1] *State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering Beihang University, Beijing 100191, China*

[2] *Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Hangzhou 310051, China*

[3] *Department of Diagnostic Ultrasound, National Cancer Center, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100021, China*

[4] *Naveen Jindal School of Management, The University of Texas at Dallas, Richardson, TX 75080-3021, U.S.A.*

E-mail: {xiexzheng, niujianwei, liu_xuefeng, liqingfeng}@buaa.edu.cn; drwangyong77@163.com; hanjiexh@126.com
shaojie.tang@utdallas.edu

Received November 26, 2019; accepted June 2, 2020.

Abstract Although using convolutional neural networks (CNNs) for computer-aided diagnosis (CAD) has made tremendous progress in the last few years, the small medical datasets remain to be the major bottleneck in this area. To address this problem, researchers start looking for information out of the medical datasets. Previous efforts mainly leverage information from natural images via transfer learning. More recent research work focuses on integrating knowledge from medical practitioners, either letting networks resemble how practitioners are trained, how they view images, or using extra annotations. In this paper, we propose a scheme named Domain Guided-CNN (DG-CNN) to incorporate the margin information, a feature described in the consensus for radiologists to diagnose cancer in breast ultrasound (BUS) images. In DG-CNN, attention maps that highlight margin areas of tumors are first generated, and then incorporated via different approaches into the networks. We have tested the performance of DG-CNN on our own dataset (including 1485 ultrasound images) and on a public dataset. The results show that DG-CNN can be applied to different network structures like VGG and ResNet to improve their performance. For example, experimental results on our dataset show that with a certain integrating mode, the improvement of using DG-CNN over a baseline network structure ResNet18 is 2.17% in accuracy, 1.69% in sensitivity, 2.64% in specificity and 2.57% in AUC (Area Under Curve). To the best of our knowledge, this is the first time that the margin information is utilized to improve the performance of deep neural networks in diagnosing breast cancer in BUS images.

Keywords medical consensus, domain knowledge, breast cancer diagnosis, margin map, deep neural network

1 Introduction

Last few years have witnessed tremendous progress in computer-aided diagnosis (CAD) in medical imaging and diagnostic radiology, primarily thanks to the advancement of deep convolutional neural networks (CNNs). Deep CNNs (such as VGG [1] and ResNet [2]) have demonstrated their great potential to be applied to the detection and diagnosis of different kinds of diseases ranging from breast cancer, lung cancer to skin cancer [3–6].

However, the medical datasets remain to be one of the major bottlenecks for these CNN-based CAD systems. In contrast to natural image applications

Regular Paper

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61976012 and 61772060, the National Key Research and Development Program of China under Grant No. 2017YFB1301100, and China Education and Research Network Innovation Project under Grant No. NGII20170315.

*Corresponding Author

where many large-scale and well-annotated datasets are available (e.g., ImageNet), the medical domain has no comparably large datasets.

To address the problem, researchers start looking for extra information besides the currently available medical datasets. For example, it has been a common practice that the model fine-tuned on medical datasets is first trained on some natural image datasets [7, 8]. The above transfer learning process implicitly leverages the information from natural images to improve the performance of deep models [9]. More recent research work leverages the information from the medical domain itself. For example, the network structures designed in [10] and [11] simulate the patterns of radiologists when they read medical images. In [12–14], the attention information when radiologists read each medical image is explicitly incorporated into the training process of the deep neural networks. The experimental results of the above work show the benefit of introducing the medical domain knowledge into deep neural networks.

In this paper, we focus on the cancer diagnosis in breast ultrasound (BUS) images. We attempt to incorporate the domain knowledge of breast radiologists into the deep neural networks. The experiences of radiologists are formally described as the consensus of BI-RADS (Breast Imaging Reporting and Data System) [15], which provides standardized terminology to depict features of the tumors for radiologists. Doctors use the BI-RADS system to place abnormal findings into different categories. Parts of the features mentioned in BI-RADS are shown in Table 1.

Table 1. Features in the BI-RADS Guidelines to Classify Benign and Malignant Tumors in Breast Ultrasound Images [15]

Feature	Benign	Malignant
Margin	Smooth, thin, regular	Irregular, thick
Shape	Round or oval	Irregular
Microcalcification	No	Yes
Echo pattern	Clear	Unclear

In Table 1, we can see that the margin attribute is particularly important: the smoothness, the thickness and the regularity of margins are directly related to the tumor categories.

To incorporate margin information, we design a scheme named as DG-CNN (Domain Guided CNN). In DG-CNN, various attention maps that highlight margin areas of tumors are first generated, and then incorporated into the networks. With these attention maps, the network will pay more attention to the margin areas of tumors.

We test DG-CNN on two BUS datasets, one of them is collected from our cooperative hospital which includes 1 485 BUS images, and the other is a public BUS dataset. Experimental results manifest that DG-CNN boosts the diagnostic performance when compared with the baseline model without integrating domain knowledge. Our contributions are three-fold as follows.

Firstly, we find that the margin feature information of BI-RADS can boost the performance of the network. To the best of our knowledge, this is the first time that the margin information is utilized to improve the diagnostic performance of breast cancer in ultrasound images.

Secondly, we design a scheme DG-CNN to integrate the above information into the network. In this scheme, four different simple but effective ways are designed to integrate this information into the network.

Thirdly, DG-CNN can be applied to different network structures like VGG and ResNet to improve the performance of the baseline networks. For example, experimental results on 1 485 ultrasound images show that for a certain integrating mode, the improvement of using DG-CNN over ResNet18 [2], a popular network structure, is 2.17% in accuracy, 1.69% in sensitivity, 2.64% in specificity and 2.57% in AUC (Area Under Curve).

2 Related Work

Many researchers start looking for extra information to improve the performance of CNNs with limited datasets. Generally speaking, the extra information comes either from natural images or from the medical domain itself.

To incorporate the information from natural images, many CNN models are pre-trained on natural image datasets like ImageNet and then are fine-tuned on given medical datasets [9, 16].

More recent research work leverages the information from the medical domain itself. According to the types of the medical information, the work can be divided into four categories: 1) the training process of radiologists, 2) patterns of radiologists when they read images, 3) the attention information of radiologists for medical images, and 4) additional diagnostic labels. The work in this paper belongs to the third category.

2.1 Training Process of Radiologists

In the training process of radiologists, the trainees are generally required to solve tasks with increasing difficulty. This process is simulated in [17], in which meta-training is utilized to model a classifier based on a series of tasks with increasing difficulty. This training method shows better performance (AUC = 0.90) when compared with its baselines on the weakly labeled DCE-MRI dataset.

2.2 Patterns of Radiologists When Reading Images

Besides the training process, experienced practitioners generally follow some patterns when they read medical images. For example, when radiologists read chest X-ray images for thorax diseases, they generally first browse the whole image, then concentrate on the local lesion areas, and finally combine the global and the local information to make decisions. This pattern is simulated as an attention-guided CNN (AG-CNN) for thorax disease classification[10]. AG-CNN has three branches to mimic the above three-staged diagnostic process, which achieves the state-of-the-art accuracy on the ChestX-ray14 dataset, and improves the average AUC to 0.868.

In addition, the DermaKNet simulates how dermatologists diagnose skin lesions[11]: first locating lesion areas, then identifying some dermoscopic features, and finally making a conclusion. In DermaKNet, a lesion segmentation network firstly segments the image into areas corresponding to the lesion and surrounding skin. Then a dermoscopic structure segmentation network segments each lesion view into a set of pre-defined dermoscopic features of special interest for dermatologists. The final diagnosis is given based on the features and the available non-visual meta-data about the lesion. DermaKNet ranks first in the Seborrheic Keratosis category and average AUCs, and achieves competitive results when compared with existing methods.

2.3 Attention Information of Radiologists for Medical Images

Besides using patterns of radiologists, some studies incorporate the attention information when radiologists read each image. For example, an attention-based CNN (AG-CNN) was proposed for glaucoma detection based on fundus images[12]. AG-CNN explicitly incorporates the attention areas of ophthalmologists on each image, which are labeled by them when reading the images.

Another example in this category is the lesion-aware CNN (LACNN)[13]. As ophthalmologists always focus on local lesion-related regions when analyzing the OCT image, LACNN designs a lesion detection network, trained on the segmentation labels of the training images, to detect these lesion-related regions and guide the following classification task. Experimental results on two clinical OCT datasets demonstrate the LACNN method has 8.3% performance gain when compared with the baseline model.

Similarly, an attention branch network (ABN) proposed in [14] allows the attention maps that highlight attention regions of the network to be manually modified on the basis of human knowledge. It achieves 93.73% classification accuracy on the disease grade recognition of retina images.

2.4 Additional Diagnostic Labels

Another type of information comes from additional diagnostic labels. Note that these labels are not the direct labels for the tasks at hand (e.g., classification), but extra labels indicating some properties of images.

For example, in the ultrasonic diagnosis of breast cancer, the BI-RADS category classifies the tumors into 0–6 for the grained explanation. These labels are generally used in an auxiliary task to help the network to distinguish among normal images, benign and malignant tumors in multi-task learning structures[18]. Reasonable results are obtained in differentiating malignant tumors and benign lumps.

Furthermore, the clinical free-text radiological reports can also be incorporated as they reflect the findings of radiologists from the images. As an example of using this information, a Text-Image embedding network (TieNet) was designed to classify the common thorax disease in chest X-rays[19]. By using the image and text attention modules, TieNet achieves a high accuracy (over 0.9 on average in AUCs) in assigning disease labels.

3 Proposed Scheme

3.1 Basic Idea

As one of the most important indicators in BI-RADS, the margin attribute has already been widely used by radiologists to distinguish between benign and malignant tumors[20, 21]. In this paper, we propose a scheme named as DG-CNN to integrate this information into the networks. DG-CNN allows the networks to pay more attention to the margin areas of tumors by using a variety of different patterns, and hence good diagnostic results can be achieved.

In DG-CNN, the way of integrating margin information can be divided into two steps: we first build the margin-wise attention map for each image, and then integrate this map into the network. The margin-wise attention map highlights the margin areas of tumors in each image, and is used to help the network to pay more attention to these areas. For convenience, the margin-wise attention map is abbreviated to the "margin map" hereinafter. The way of generating margin maps will be elaborated in Subsection 3.2.

After obtaining the margin maps, we present several approaches to integrate them into DG-CNN. The first approach is to insert them as part of the input to some convolution layers, which enables DG-CNN to learn this feature directly. The second approach is using a multi-task learning structure, in which predicting the margin map and predicting the category label (benign or malignant) for each image are taken as the auxiliary task and the main task, respectively.

3.2 Generating Margin Maps

There are three different margin maps utilized in DG-CNN. According to the generating methods, these maps can be divided into the label-based margin maps and the model-based ones, and the latter one can be further divided into two subcategories. The former ones are generated from the segmentation annotations of tumors, while the latter ones are the prediction results of some edge detection models.

To generate the label-based margin maps, we first extract the tumor edge from the segmentation annotation of each image, and then expand some pixels inside and outside the edge to incorporate the margin area. Thus, these maps are also named as the edge-expanded maps, and the number of pixels expanded inside and outside the tumor edge is both set to 10 according to the experience of radiologists.

On the other hand, the margin areas of different tumors may have different widths, which may not be well described by a fixed value. Thus we also adopt the model-based margin maps, in which we think these maps can better fit the specificity of tumor margins. To generate the model-based margin maps, the popular Richer Convolutional Features (RCF) model [22] is adopted. More specifically, the tumor edge is first extracted from the segmentation annotation of each image and taken as the edge label. Then the RCF model pretrained on the BSDS500 dataset [23] is fine-tuned on our dataset. At last, the trained model is used to predict the possible margin areas for each unseen image.

According to the dependence on the segmentation labels of our dataset, the margin maps generated by the RCF model can be divided into two categories: the semi-RCF maps and the full-RCF ones. The former ones are the prediction results of the RCF model obtained by the semi-supervised method (in which only part of the images with segmentation labels are needed). In contrast, the full-RCF maps are obtained by the fully-supervised training methods (in which all of the images with segmentation labels are required). More details of the training process will be introduced in Subsection 4.2.

Fig.1 shows an example of our dataset and its margin maps. Specifically, Fig.1(a) is the original image with a benign tumor, and Figs.1(b)–1(d) are the edge-expanded margin map, the semi-RCF margin map and the full-RCF margin map, respectively.

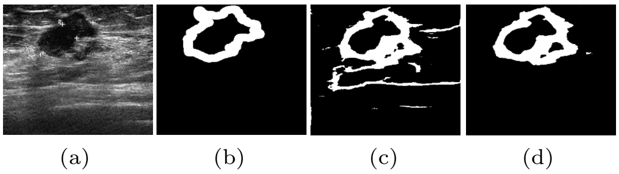


Fig. 1. An example and its margin maps of our dataset. (a) Original image with a benign tumor. (b) Edge-expanded margin map. (c) Semi-RCF margin map. (d) Full-RCF margin map.

The edge-expanded map describes the transition area with a fixed width around the edge of the tumor (10 pixels both inside and outside). On the contrary, the semi-RCF and the full-RCF margin maps are generated from the RCF model, and hence can better fit the boundary of tumors. In addition, we can see that when compared with the semi-RCF map shown in Fig.1(c), the full-RCF map in Fig.1(d) more accurately describes the margin area of the tumor, as non-tumor regions are not highlighted. The diagnostic performance when using these three margin maps will be described in detail in Subsection 4.3.4.

3.3 Integrating Margin Maps into the Networks

In this subsection, we first introduce our baseline, and then describe two approaches used in DG-CNN, namely the insert mode and the multi-task learning mode, to integrate the margin maps into the networks. The baseline model and different integrating modes are shown in Fig.2, where the ResNet18 structure [2] is used as the backbone. The details are described as follows.

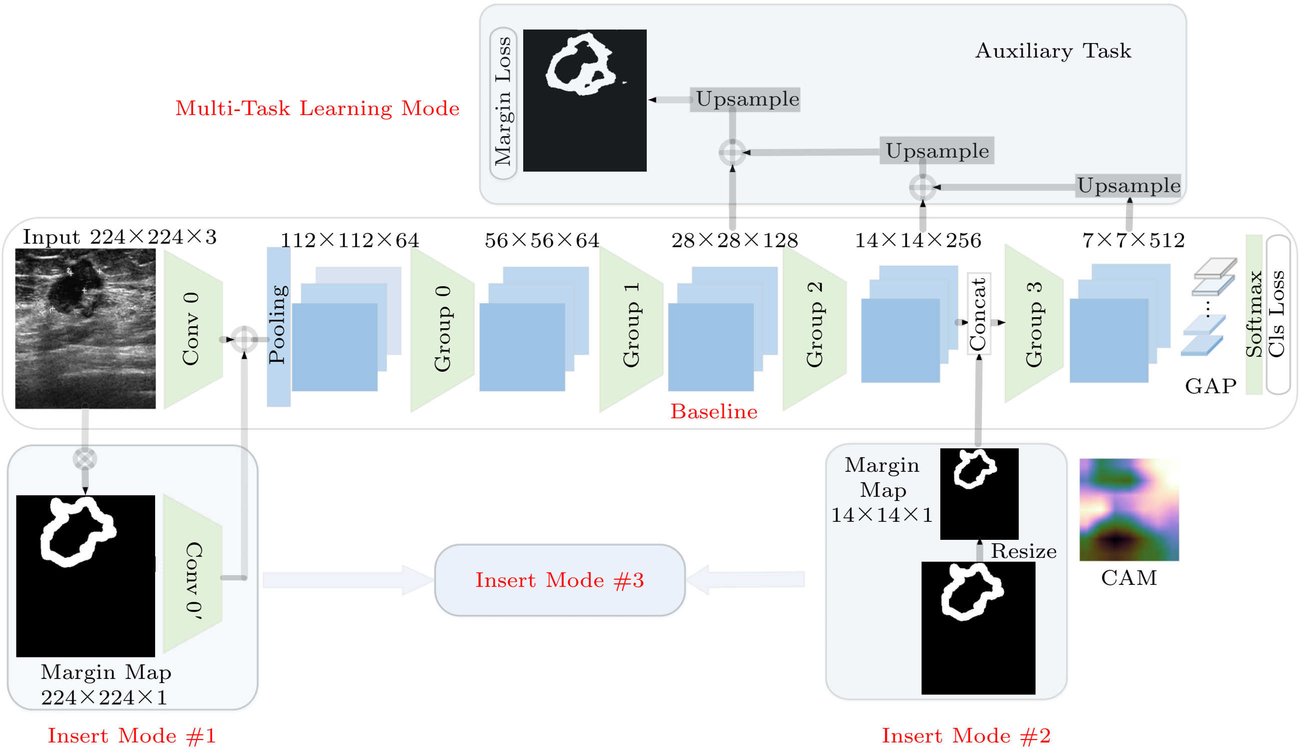


Fig.2. DG-CNN with three insert modes and the multi-task learning mode, where a popular network structure, ResNet18, is used as the backbone.

3.3.1 Baseline

Our baseline model follows the classification structure of ResNet18 [2] without incorporating any external information (illustrated as the rectangle in the middle of Fig.2). The original image is fed into the network as the input and the classification result for each image (benign or malignant) is the output. More specifically, this structure mainly contains one convolution layer (conv 0 with a 7×7 kernel), one max pooling layer (stride = 2) and four convolutional groups (group 0–group 3, containing two convolutional layers with 3×3 kernels), followed by a global average pooling (GAP) layer [24] and a fully-connected layer. Besides the feature vectorization, the GAP layer is also used to calculate the class activate map (CAM), which indicates the saliency areas in each image that are important for the final classification results [25]. The weighted sum of the feature maps of the last convolutional layer is computed to generate CAM for each image. The number of the output of the fully-connected layer is set to 2 for the two target categories. The loss function of our baseline model is the softmax cross entropy of the two categories defined as follows:

$$L_{\text{cls}} = -\sum_{i=1}^{K} y_i \log{(p_i)}, \tag{1}$$

$$p_i = \frac{e_{x_i}}{\sum_{j=1}^{n} e_{x_j}}, \tag{2}$$

where y_i, p_i are the true label and the prediction probability for target class i, respectively. $y_i = 1$ if the target class is i and $y_i = 0$ otherwise. K is set to 2 representing two target categories. In particular, p_i is calculated using the softmax function in (2), and n is set to 2 as there are two outputs of the network.

3.3.2 Insert Mode

In this approach, margin maps are inserted as the part of the input of some convolution layers. Specifically, margin maps are first combined with the feature maps extracted from a certain convolution layer, and then fed as the input to the next layer.

According to the insert positions and how margin maps and feature maps are combined, the insert mode can have many variants. We design three different insert methods which are denoted as "insert mode # 1", "insert mode # 2" and "insert mode # 3" (shown in the red fonts in Fig.2).

Insert Mode #1. In this mode, the margin map is inserted at conv 0, where two types of feature maps are combined and fed into the next layer, one extracted from the image masked by the corresponding margin map, and the other from the original image. Specifically, the element-wise multiplication (denoted as $\otimes$) is firstly implemented between the margin map (denoted as "$margin\_map$") and the original image (denoted as "img"). Then, the product "$img \otimes margin\_map$" and the original image "img" are fed into a pair of convolutional layers (conv 0′ and conv 0). At last, two outputs, $f_{\text{conv0}}(img)$ and $f_{\text{conv0}'}(img \otimes margin\_map)$, are added pixel by pixel (denoted as $\oplus$) to highlight the features of the margin areas of tumors. And the sum, denoted as I_1 (shown in (3)), is fed into the next max pooling layer,

$$I_1 = f_{\text{conv0}}(img) \oplus f_{\text{conv0}'}(img \otimes margin\_map)\,. \quad (3)$$

During the backpropagation process, the weights of the conv 0′ layer are updated to learn the important features from the margin map.

Insert Mode #2. In this mode, the margin map is inserted between "group 2" and "group 3", where the margin map and the feature maps extracted from "group 2" are concatenated together and then fed into "group 3". Specifically, the margin map is first resized to the same size with the feature maps generated from "group 2" (denoted as f_{group2}). The resized margin map, denoted as $r(margin\_map)$, is then concatenated with f_{group2}. Finally, as the new output of "group 2" (denoted as I_2 shown in (4)), the concatenation result is fed into the "group 3",

$$I_2 = \text{concat}\left(f_{\text{group2}}, r(margin\_map)\right). \quad (4)$$

Based on this concatenation, "group 3" can learn the distinguishing features from different channels automatically and the margin information can also be integrated into the higher layers of the network. Similarly, the weights of these channels can also be modified to measure the importance of the margin map during the backpropagation process.

Insert Mode #3. This mode simply combines insert mode #1 and insert mode #2 mentioned above, where the feature maps generated from the conv 0 layer and the group 2 convolution block are all modified by using (3) and (4) respectively. With the integration of margin maps in the lower and the higher layers of the network via the above two modes, the margin information can be enhanced directly and learned by the network automatically.

It should be mentioned that in these three insert modes, the margin maps are all integrated in the feature level, and the same loss functions as the baseline model are adopted. Different from the baseline model, the input of these three modes also includes the margin maps of the original images. In addition, the weights of the layers connected with the margin maps can also be modified during the backpropagation process.

3.3.3 Multi-Task Learning Mode

Besides the insert modes, another approach to integrating margin maps is using the multi-task learning structure. In this mode, the DG-CNN consists of two tasks, with the main classification task for predicting the category label, and the auxiliary task (called as the margin-wise attention generation task) for generating the margin map for each image. As mentioned in Subsection 3.1, these different margin maps (including the edge-expanded map, the semi-RCF map and the full-RCF map) are generated firstly, and then integrated into networks as the labels in the auxiliary task. As the auxiliary task highlights the margin areas of tumors, the network will learn to pay more attention to these areas during the training process.

Fig.2 also shows the example of the multi-task learning mode, where the classification task is used as the main task, and the auxiliary task is shown in the upper part. In the auxiliary task branch, to obtain better margin prediction results, we adopt the skip-layer connection structure [26] to fuse the feature maps generated from the shallow and deep layers.

In particular, the feature maps generated from "group 3" are upsampled and added with those generated from "group 2" pixel by pixel. The added results are then upsampled and added with the feature maps generated from "group 1". At last, the fused feature maps are upsampled to form the predicted margin map for each image. The size of each predicted margin map is $2 \times 224 \times 224$, with 2 and 224 being the number of channels and the size of feature maps, respectively.

The final loss of DG-CNN consists of the margin attention loss and the classification loss, which are calculated as follows respectively:

$$L_{\text{multi}} = L_{\text{m\_att}} + L_{\text{cls}}, \quad (5)$$

$$L_{\text{m\_att}} = -\sum_{i=1}^{K} \bar{y}_{pix\_i} \log\left(y_{pix\_i}\right), \quad (6)$$

where $L_{\text{m\_att}}$ and L_{cls} are the loss functions of margin-wise attention generation branch and classification branch, respectively. Concretely, $L_{\text{m\_att}}$ depicts the cross entropy loss between the ground truth label $\bar{y}_{pix\_i}$ and the prediction label y_{pix_i} for each pixel. K is set to 2 as there are two possible categories for each pixel, namely, the margin areas and the non-margin areas. L_{cls} is the classification loss and is calculated in (1).

The margin attention generation task is integrated as the auxiliary task and is trained first. Then the trained model is used to fine-tune the whole training process of the two tasks. More details about the training process will be described in Subsection 4.2.

3.4 Summary

We summarize all types of DG-CNN described in Subsections 3.3.2 and 3.3.3. DG-CNN can be classified according to the following two dimensions: 1) the types of margin maps and (2) the approaches to integrating the margin maps. Firstly, we design three types of margin maps, namely the edge-expanded map, the semi-RCF map and the full-RCF map. The details are described in Subsection 3.2. Secondly, the two approaches to integrating margin maps, namely the insert mode and the multi-task learning mode, are introduced in Subsection 3.3.

4 Experiments and Results

4.1 Dataset and Evaluation Metrics

The dataset used in this paper is collected from 953 patients from the Cancer Hospital of Chinese Academy of Medical Sciences during 2018. This dataset contains 1 485 BUS images sampled from different systems including PHILIPS, SIEMENS and HITACHI.

Within this dataset, 303 images from 60 patients are with benign tumors, while 1 182 images from 893 patients are with malignant ones. All the 1 485 images contain at least one tumor. In addition, all images are color ones with 3-channel (RGB) and their sizes vary from 321×335 to 676×437 pixels. Each image has a category label indicating whether it is benign or malignant, and a segmentation label indicating the location and shape of tumor in the image. All segmentation labels are normalized to 0 and 1, corresponding to the background and the tumor areas, respectively. The category labels are proved histopathologically by biopsy and segmentation labels are marked collaboratively by at least three experienced radiologists to reduce the inter observer variance. The dataset will be made public for academic research in the future.

Some BUS images and their segmentation labels in the dataset are illustrated in Fig. 3. In particular, Fig. 3(a) and Fig. 3(b) are the image with a malignant tumor and its segmentation label, respectively, and Fig.3(c) and Fig.3(d) are the benign ones.

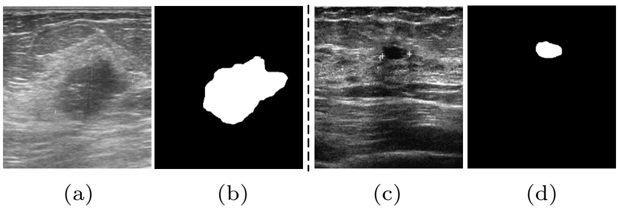


Fig.3. (a) Image containing a malignant tumor. (b) Segmentation label of (a). (c) Image with a benign tumor. (d) Segmentation label of (c).

We adopt the accuracy, sensitivity, specificity, and F_β score to quantify the diagnostic performance, and some of them are defined as:

$$sensitivity = \frac{TP}{TP + FN}, \tag{7}$$

$$specificity = \frac{TN}{TN + FP}, \tag{8}$$

$$F_\beta = \frac{(1+\beta^2) \times TP}{(1+\beta^2) \times TP + \beta^2 \times FN + FP}, \tag{9}$$

where TP, FP, TN and FN are the numbers of true positives (correctly identified malignant tumors), false positives (benign tumors reported as malignant), true negatives (correctly identified benign tumors) and false negatives (malignant tumors reported as benign) respectively. In addition, β in F_β score is set to 2, which places more emphasis on the sensitivity, as the capability to identify malignant tumors is more important in radiology. Furthermore, ROC curve and area under ROC (AUC) are also adopted to compare the performance of different methods.

4.2 Implementation Details

To generate semi-RCF and full-RCF maps, the pretrained BSDS500 model is first fine-tuned based on our dataset, and then the obtained models are used to predict the margin maps on all images. Specifically, when fine-tuning on our dataset, the numbers of training images used in the generation of semi-RCF and full-RCF maps are 800 and 1 485, respectively. The models in both of the two conditions are trained for 30 epochs with the learning rate of 1.0×10^{-6}.

As our dataset has more malignant cases than benign ones, we horizontally mirror and rotate 15 degrees clockwise for all images with benign tumors. In addition, 5-fold cross validation is utilized to validate the performance of different methods. Since multiple images are obtained from the same patient, the images from the same patient are divided into the same fold, and the distribution of two classes in each fold is maintained as same as possible. In addition, all images are resized to 224 × 224 and fed into the network without any pre-processing.

In our baseline model and the models of using DG-CNN with insert modes, the learning rate is initialized to 0.1, and then decreased by 10 at the 30th, 60th, 90th and 100th iterations respectively. However for the DG-CNN with the multi-task learning mode, the margin-wise attention generation task is firstly trained for 300 epochs with the learning rate of 0.1. Then in the following joint training process of multi-task learning, the trained model in the last step is used to initialize the parameters of the two tasks. It should be mentioned that the ImageNet pre-activation model is used to fine-tune the baseline model, the models of using DG-CNN with insert modes, and the model of the multi-task learning mode in the first stage. The batch size is set to 256 for 105 epochs, and the scale gradient optimization with scale factor 0.1 is used in all convolution layers. All experiments are trained on 2 GPUs of NVIDIA Tesla V100-PCIE-16GB.

4.3 Effectiveness of DG-CNN

In this subsection, we test the diagnostic performance of DG-CNN. As we design three different margin maps and four integrating methods, we first fix the types of the margin map as the full-RCF margin map, and test the performance of DG-CNN using different integrating methods. The detailed results are shown in Subsection 4.3.1. Besides, the diagnostic performance of DG-CNN is also compared with that of other methods, and the results are analyzed in Subsection 4.3.2.

Then we evaluate the generalization of DG-CNN in other three different network structures (ResNet34, VGG16 and VGG19). Note that the full-RCF margin map is used here. Specifically, we first choose the VGG16 structure, to manifest the diagnostic performance after integrating margin information. The performance of DG-CNN on some deeper network structures including ResNet34 and VGG19 is also evaluated. The detailed results are shown in Subsection 4.3.3.

At last, we test the performance of DG-CNN using different margin maps including the full-RCF map, the semi-RCF map and the edge-extended map. They are tested at some certain integrating methods (insert mode #1 and multi-task learning mode) when DG-CNN is applied to ResNet18. The details are shown in Subsection 4.3.4.

4.3.1 *Performance of DG-CNN Using Different Integrating Methods*

The quantitative results of baseline and different integrating methods are listed in Table 2, where the best diagnostic performance for each metric is highlighted. In Table 2, we can see that generally speaking, DG-CNN with different integrating methods outperforms the baseline on different extents. This demonstrates the effectiveness of introducing margin information.

In addition, in terms of different integrating methods, DG-CNN with the three insert modes achieves better performance when compared with the multi-task learning mode, and outperforms the corresponding baseline in most metrics. For example, DG-CNN (insert mode #3) has the highest accuracy, specificity and AUC, which outperforms the baseline by 2.17%, 2.64% and 2.57%, respectively. On the other hand, DG-CNN (insert mode #2) achieves the highest sensitivity and

Table 2. Comparing the Diagnostic Performance of DG-CNN with Different Integrating Methods and the Baseline in Different Metrics

Method	Metrics (%)				
	Accuracy	Sensitivity	Specificity	AUC	*F*2
Baseline (ResNet18)	77.53	81.81	73.35	84.91	80.34
DG-CNN (insert mode #1)	79.45	84.60	74.42	86.78	82.81
DG-CNN (insert mode #2)	79.24	**90.78**	67.99	87.27	**86.69**
DG-CNN (insert mode #3)	**79.70**	83.50	**75.99**	**87.48**	82.17
DG-CNN (multi-task)	79.07	90.19	68.23	86.82	86.25
Mask R-CNN [27]	77.78	87.73	68.07	85.12	84.28
LACNN [13]	79.24	88.41	70.30	88.28	85.47
Han *et al.* [28]	78.40	89.68	67.41	84.60	88.58

$F2$ score, and the improvements over the baseline are 8.97% and 6.35%, respectively. Furthermore, although being not so good as the insert modes on average, DG-CNN (multi-task) also performs better when compared with the baseline on most metrics.

We can see in Table 2 that after incorporating margin maps, not all the models achieve higher performance in all metrics compared with the baseline. In particular, the insert mode #2 and the multi-task learning mode have lower specificity values than the baseline. Note that low specificity is associated with the high sensitivity, which can be confirmed by very high sensitivity rates of these two methods. This is not surprising as a model with high sensitivity rarely misses malignant tumors, but the price to pay is that it may classify a benign tumor as malignant easily, generating low specificity.

In addition, although the margin information has been widely utilized by radiologists, we find that the margin information is more effective in identifying malignant tumors rather than the benign ones. This is confirmed in our dataset, as we find that while most malignant tumors have irregular margins, some benign tumors (about 8%) also have irregular margins. This means that if a model overly relies on this information to classify tumors, the model is likely to obtain high sensitivity but low specificity, which matches the results of DG-CNN (insert mode #2) and DG-CNN (multi-task).

In addition, the ROC curves of DG-CNN with different integrating methods and the baseline are shown in Fig.4. It can be seen that the ROC curve of DG-CNN (insert mode #3) is closer to the upper-left corner and has the highest AUC value. Moreover, from the ROC curves, we can also see that using DG-CNN with all these integrating methods can achieve higher sensitivity when compared with the baseline at the same specificity value.

Furthermore, the CAMs of these different methods are shown in Fig.5, where these images are wrongly predicted in baseline and correctly predicted in our DG-CNN with four different integrating methods. It should be mentioned that the highlighted areas are where the networks pay more attention to when making predictions. The first two rows are the images with malignant tumors, while the bottom two rows are benign ones. Fig.5(a) shows the original images, and Figs.5(b)–5(f) are the CAMs of the baseline and the DG-CNN using insert mode #1–insert mode#3 and multi-task learning mode, respectively.

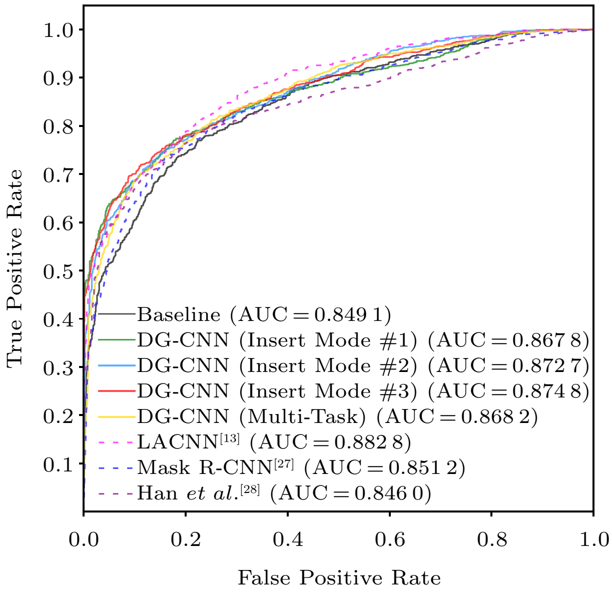


Fig.4. ROC curves of baseline and DG-CNN with different integrating methods.

We can see from the CAMs of the baseline, the network does not pay attention to the tumor margins or even the tumor itself in most cases. In contrast, the DG-CNNs using four integrating methods respectively focus more on the tumor margins and therefore give correct predictions.

4.3.2 *Performance Comparison of DG-CNN with Other Deep Learning Methods*

In this subsection, we compare the performance of DG-CNN and three popular methods, LACNN [13], Mask R-CNN [27] and the method proposed by Han *et al.* [28]. Note that all these methods are evaluated on our dataset for a fair comparison.

More specifically, LACNN [13] is originally designed for the classification of retinal optical coherence tomography (OCT) images. It designs an attention module to incorporate the information of the whole lesion area. Here, we apply LACNN to our BUS images but for a better comparison, we modify the attention module such that the information of the margin area instead of the whole lesion area is incorporated. As LACNN first generates the attention maps and then integrates them into the network, it is similar to the three insert modes of DG-CNN. Therefore, the hyper-parameters of LACNN are set as the same with DG-CNN with insert modes.

The structure of Mask R-CNN [27] is intrinsically a multi-task learning architecture. When applied to our dataset, the Mask R-CNN model has two tasks: the

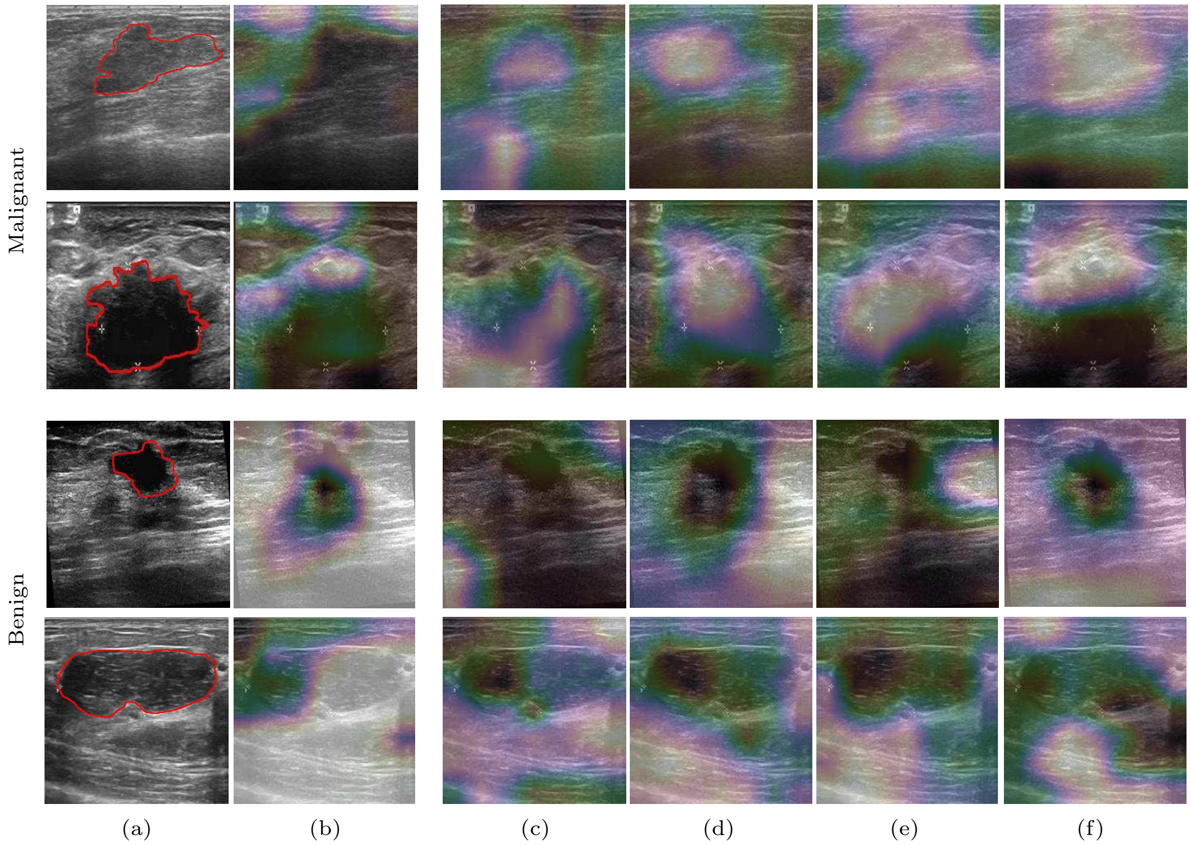


Fig.5. (a) Original images in which the tumors are annotated by the red curves. The upper two rows are the images with malignant tumors, while the bottom two rows are the benign ones. (b) CAMs of the baseline. (c) CAMs of DG-CNN (insert mode #1). (d) CAMs of DG-CNN (insert mode #2). (e) CAMs of DG-CNN (insert mode #3). (f) CAMs of DG-CNN (multi-task).

classification task (as the main task) and the margin segmentation task (as the auxiliary task). The hyper-parameters of Mask R-CNN are set to the same with the multi-task learning mode of DG-CNN.

The method proposed by Han *et al.*[28] highlights the margin information by augmenting the cropped images with some fixed numbers of margin pixels. We adopt the original settings in [28].

The quantitative results of our DG-CNN and the comparative methods are also listed in Table 2. We give the detailed analysis below.

For Mask R-CNN, we can firstly see that compared with the baseline method which does not incorporate any margin information, Mask R-CNN has much higher improvement in sensitivity (about 6%), a slight improvement in accuracy (about 0.2%) and a decrease in specificity (about 5%). The high improvement in sensitivity, as well as the decrease in specificity shows that by incorporating the margin information, the sensitivity of the model for classifying malignant tumors increases significantly. This matches well with our discussion in Subsection 4.3.1 that although the margin information is helpful, it should not be overly emphasized.

As the parameters settings of Mask R-CNN are the same with the multi-task learning mode of DG-CNN, we compare these two methods. From the result, we can see that DG-CNN (multi-task) outperforms Mask R-CNN in all evaluation metrics, with 1.29% improvement in accuracy, 2.46% in sensitivity, 0.16% in specificity, and 1.70% in AUC.

For LACNN, we can see that DG-CNN achieves comparable diagnostic performance. In terms of the accuracy, DG-CNN (insert mode #1) and DG-CNN (insert mode #3) have a slightly higher accuracy (with 0.21% and 0.46% improvement respectively) than LACNN, while DG-CNN (insert mode #2) has the same accuracy with LACNN. However, compared with LACNN, the three insert modes of DG-CNN have much simpler architectures and therefore are more light-weighted.

For the method proposed by Han *et al.*[28], we can see that although not significantly, the proposed DG-CNN for all four integration methods outperforms the

Han *et al.*'s method [28] in terms of the accuracy, specificity and AUC. The results show that using the margin maps to incorporate the margin information can obtain better results than using a simple data augmentation.

The ROC curves of these different methods are also shown in Fig. 4, where the DG-CNNs with different integrating methods show higher AUC values when compared with Mask R-CNN [27] and Han *et al.*'s method [28]. In addition, from the ROC curves, we can see that DG-CNN achieves comparable performance when compared with LACNN [13], but with the simple integrating methods.

4.3.3 Performance of DG-CNN Using Different Network Structures

In addition to ResNet18, we also evaluate DG-CNN with other network structures including VGG16, ResNet34 and VGG19. It should be mentioned that when DG-CNN is applied to VGG structures, the first two fully-connected layers are replaced by the GAP layer. Note that here the "baseline" refers to the network model without integrating margin maps. The experimental results are shown in Table 3. The best performance for each metric in different backbones is highlighted in bold.

When the VGG structure is adopted, the baseline model and the models of using DG-CNN with insert modes have the initial learning rate 0.002 5 during the training process. Then the learning rate decreases by 10 at the 30th, 60th and 80th iterations, respectively. In the multi-task learning mode, the learning rate and epochs of margin-wise attention generation task are 0.05 and 300, respectively. The batch size is set to 32 for 100 epochs. ImageNet pre-trained VGG model is used to initialize the network parameters, and the Stochastic Gradient Descent (SGD) optimization method is used with the momentum of 0.9. The other settings are kept the same with the condition when DG-CNN is applied to the ResNet structure. As for ResNet34, the same settings with that in ResNet18 are adopted.

As shown in Table 3, when DG-CNN is applied to these structures, almost all networks can achieve distinguishable improvements over the corresponding baselines, which proves the generalization of DG-CNN. For example, in terms of accuracy, the highest improvements are 2.00%, 0.46% and 1.30% when DG-CNN is applied to ResNet34, VGG16 and VGG19, respectively. In particular, DG-CNN (insert mode #3) is prone to improve the accuracy on the shallower network structures (VGG16), while DG-CNN (insert mode #1) performs better on deeper network structures like ResNet34 and VGG19.

As the same with that in ResNet18, DG-CNN (insert mode #2) also seems to be able to achieve good sensitivity and $F2$ score on VGG structures, with the improvements of 3.05% and 1.67% on VGG16, and 2.71% and 1.62% on VGG19, respectively. In terms of AUC, DG-CNN with all insert modes also has the better performance when compared with DG-CNN (multi-task). The improvements are 1.43%, 2.08% and 1.71%, respectively.

We can see in Table 3 that some integrating methods perform better in shallow networks (i.e., VGG16) than in deeper ones (i.e., VGG19 and ResNet34). For

Table 3. Diagnostic Performance of DG-CNN with Different Integrating Methods When Using Other Network Structures

Backbone	Method	Metrics (%)				
		Accuracy	Sensitivity	Specificity	AUC	$F2$
ResNet34	Baseline	78.49	82.15	74.92	86.73	80.88
	DG-CNN (insert mode #1)	**80.49**	83.08	**77.97**	**88.16**	82.15
	DG-CNN (insert mode #2)	77.69	82.49	73.02	85.96	80.85
	DG-CNN (insert mode #3)	79.04	81.39	76.73	87.42	80.54
	DG-CNN (multi-task)	75.98	**93.15**	59.24	85.55	**87.06**
VGG16	Baseline	78.49	89.76	67.49	86.75	85.80
	DG-CNN (insert mode #1)	78.82	89.59	68.32	**88.83**	85.80
	DG-CNN (insert mode #2)	78.61	**92.81**	64.77	88.70	**87.47**
	DG-CNN (insert mode #3)	**78.95**	88.24	69.88	88.49	84.99
	DG-CNN (multi-task)	78.40	80.80	**76.07**	86.15	79.94
VGG19	Baseline	78.78	90.27	67.57	87.31	86.22
	DG-CNN (insert mode #1)	**80.08**	91.12	69.31	88.71	87.18
	DG-CNN (insert mode #2)	78.61	**92.98**	64.60	88.46	**87.84**
	DG-CNN (insert mode #3)	79.45	92.47	66.75	**89.02**	87.81
	DG-CNN (multi-task)	77.57	81.90	**73.35**	85.58	80.41

example, when ResNet34 is used as the backbone, only DG-CNN (insert mode #1) and DG-CNN (insert mode #3) achieve better results when compared with the baseline model. While with VGG19, only DG-CNN (insert mode #1) has the better performance than the baseline in all metrics. Another observation is that the improvement of DG-CNN over the baselines on VGG structures is not so significant as on ResNet. For example, when DG-CNN (multi-task) is applied to VGG16 and VGG19, the obtained networks only outperform the corresponding baseline in specificity, and have comparable or even a lower accuracy, sensitivity, AUC and $F2$ score. The reason may come from the imbalanced dataset and the training process of multi-task learning.

4.3.4 *Performance of DG-CNN Using Different Margin Maps*

In this subsection, we analyze the effect of different margin maps on the performance of DG-CNN using two integrating methods. The margin maps include the full-RCF map, the semi-RCF map and the edge-expanded map. The two modes are DG-CNN (insert mode #1) and DG-CNN (multi-task).

The detailed results are shown in Table 4 where the best performance for each metric in different methods is shown in bold. We can see that using all these three margin maps (semi-RCF, full-RCF and edge-expanded maps) can achieve higher performance in most of the metrics, especially in the accuracy, sensitivity and $F2$ score. Among the three margin maps, using semi-RCF and full-RCF margin maps seems to be able to achieve better performance than using the edge-expanded map. This may be attributed to the fact that the semi-RCF and the full-RCF margin maps are generated from the RCF model, and therefore can better fit the margin of each tumor than the edge-expanded map with the fixed width. For example, when using the full-RCF margin map, DG-CNN (insert mode #1) achieves the highest accuracy (79.45%), sensitivity (84.60%) and $F2$ score (82.81%), improved by 1.92%, 2.79% and 2.47% over the baseline, respectively. In contrast, when the semi-RCF map is integrated, the specificity and AUC are improved by 3.22% and 1.95%, respectively.

It should be noted that when considering the overall performance of these two integrating methods, using the semi-RCF margin map achieves comparable or even better performance when compared with that using full-RCF one. For example, DG-CNN (insert mode #1) using semi-RCF and full-RFC maps achieves 79.07% and 79.45% in diagnostic accuracy respectively, and 86.86% and 84.10% in AUC respectively.

For DG-CNN (multi-task), the accuracy and AUC values using semi-RCF and full-RCF maps are 79.16% vs 79.07% and 86.40% vs 86.82%, respectively. In addition, the specificity of using the semi-RCF map is even higher than that in the full-RCF map in both the insert mode #1 and the multi-task mode.

The good performance of using the semi-RCF margin map is highly preferable, as it is generated from the model trained that only requires a part of segmentation annotations of tumors on the dataset.

The ROC curves of different margin maps when using DG-CNN (insert mode #1) and DG-CNN (multi-task) are shown in Figs.6(a) and 6(b), respectively. We can see in Fig.6(a) that using DG-CNN (insert mode #1), the best performance is achieved when the semi-RCF margin map is utilized. When DG-CNN (multi-task) is adopted, using the full-RCF margin map can obtain the best performance. The AUC values of them are 0.868 6 and 0.868 2, respectively, and are 1.95% and 1.91% higher than those in the baseline.

4.4 Comparison of Effect of Tumor Margin and Tumor Itself

In this paper, we design DG-CNN to incorporate the margin information of tumors. The experimental results in Subsection 4.3 demonstrate the effect of introducing this information.

Table 4. Diagnostic Performance of DG-CNN (Insert Mode #1) and DG-CNN (Multi-Task) When Integrating Different Margin Maps

Method	Margin Map	Metrics (%)				
		Accuracy	Sensitivity	Specificity	AUC	$F2$
Baseline	—	77.53	81.81	73.35	84.91	80.34
DG-CNN (insert mode #1)	Semi-RCF	79.07	81.64	**76.57**	**86.86**	80.73
	Full-RCF	**79.45**	**84.60**	74.42	84.10	**82.81**
	Edge-expanded	75.56	82.40	68.89	80.02	80.11
DG-CNN (multi-task)	Semi-RCF	**79.16**	86.89	71.62	86.40	84.19
	Full-RCF	79.07	90.19	68.23	**86.82**	86.26
	Edge-expanded	78.11	**90.95**	65.59	85.31	**86.41**

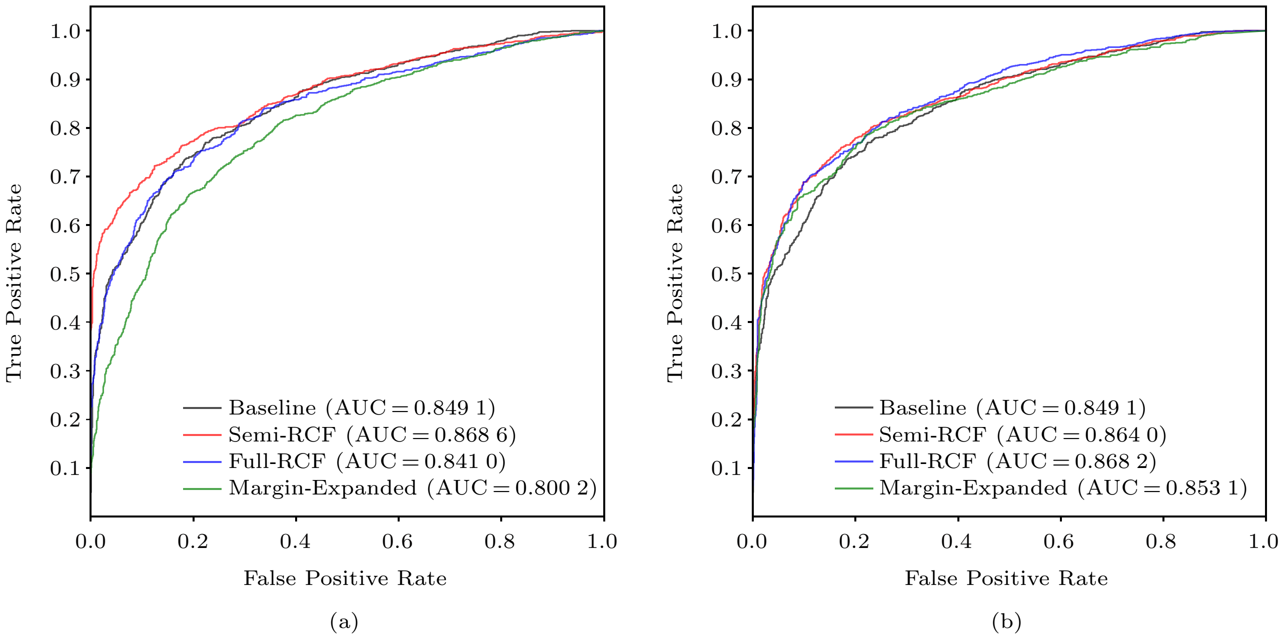


Fig.6. (a) ROC curves when different margin maps are integrated using DG-CNN (insert mode #1). (b) ROC curves when different margin maps are integrated using DG-CNN (multi-task).

It should be noted that, the importance of margin of tumors in diagnosing breast cancer comes from the medical consensus and guidelines.

However, it is widely recognized that in object classification tasks, if the network focuses on the object itself, the performance can generally be improved [10, 29]. The knowledge comes from the community of computer vision.

Therefore, it is interesting to compare the effect of introducing these two types of knowledge. More specifically, can simply letting the network focus on the tumor itself achieve comparable or even better results than using the margin information?

To answer this question, we integrate the information of segmentation annotations of tumors into the network. Specifically, the mask map, which indicates the whole area of tumors, is utilized to replace the original margin map in DG-CNN. In addition, the performance of introducing the mask-expanded map, which contains both tumor and the corresponding margin area, is also tested. The mask-expanded map is generated by expanding the edge of the mask map to a certain number of pixels. The number of expanded pixels is also set to 10, the same with that in the edge-expended map.

Having obtained the mask map and the mask-expanded map, we simply utilize them to replace the margin map in DG-CNN. The margin map to be compared in DG-CNN is the semi-RCF margin map, as it has demonstrated its effectiveness in Subsection 4.3.4. In addition, we test the three types of maps under two integrating methods mentioned before, 1) DG-CNN (insert mode #1) and 2) DG-CNN (multi-task).

The experimental results are shown in Table 5 where the best performance in each metric of different meth-

Table 5. Diagnostic Performance When Using Mask and Mask-Expanded Margin Map in DG-CNN

Method	Margin Map	Metric (%)				
		Accuracy	Sensitivity	Specificity	AUC	*F*2
Baseline	—	77.53	81.81	73.35	84.91	80.34
DG-CNN (insert mode #1)	Semi-RCF	**79.07**	81.64	**76.57**	**86.86**	**80.73**
	Mask	76.52	77.83	75.25	79.47	77.34
	Mask-expanded	77.40	80.96	73.93	81.75	79.74
DG-CNN (multi-task)	Semi-RCF	**79.16**	86.89	**71.62**	**86.40**	84.19
	Mask	74.27	90.78	58.17	85.17	85.05
	Mask-expanded	74.52	**92.56**	56.93	85.02	**86.22**

ods is shown in bold. We can see that generally speaking, introducing the mask map and the mask-expanded map cannot achieve comparable results as the semi-RCF margin map. It is surprising that the performance in some metrics is even inferior to the baseline.

More detailed analyses show that compared with the mask-expanded map and the semi-RCF margin map, introducing the mask map generally achieves the lowest performance in both integrating methods. Using the mask-expanded map achieves a higher accuracy than using the mask map, but the accuracy is inferior to the baseline. In particular, the diagnostic accuracy of DG-CNN (insert mode #1) and DG-CNN (multi-task) when using the mask map is declined by about 1.01% and 3.26% when compared with the baseline.

When using the mask and the mask-expanded map in DG-CNN (insert mode #1), the sensitivity and specificity values are similar to those in the baseline. On the other hand, when using these two maps in DG-CNN (multi-task), although they have higher sensitivity, their specificity values are the lowest among all the methods. For example, when the mask map is integrated in DG-CNN (multi-task), the sensitivity is improved by 8.97% compared with the baseline, while the specificity is decreased by 15.18%. When using the mask-expanded map, the increase of sensitivity and the decrease of specificity are 10.75% and 16.42%, respectively.

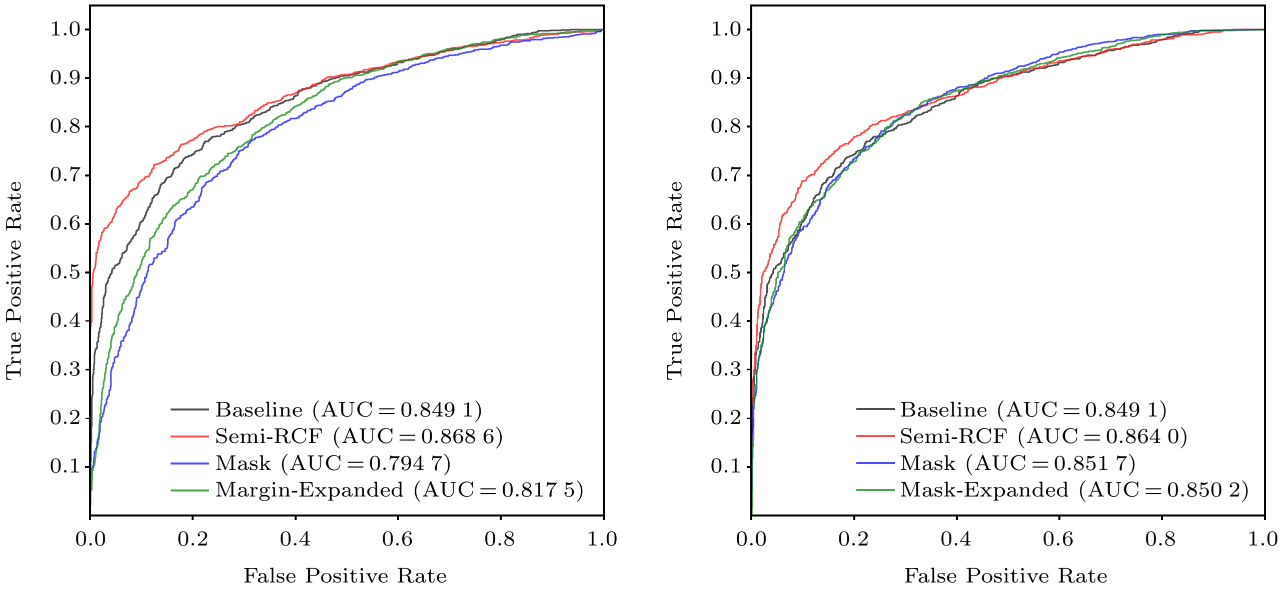
The ROC curves of DG-CNN with integrating the mask and the mask-expanded map along with the semi-RCF map and the baseline are shown in Fig.7. We can see that the ROC curves of the mask and the mask-expanded map are at the bottom-right corner which indicates the low AUC values.

Furthermore, CAMs when integrating these different maps using DG-CNN (insert mode #1) are shown in Fig.8. In particular, two images in the first column are the images containing the malignant tumor (upper figure) and the benign one (lower figure), respectively. The locations of these two tumors are also annotated by the red curves. Figures in the second column to the last one are the CAMs of the baseline, and the DG-CNN (insert mode #1) using the semi-RCF margin map, the mask map and the mask-expanded map respectively. Note that except using the semi-RCF margin map, all the other three conditions fail to correctly identify the category of the two tumors.

4.5 Evaluation on a Public Dataset

In this subsection, we test DG-CNN on a public BUS dataset named “dataset B” [8]. This dataset is abbreviated to BUS-B for convenience. BUS-B consists of 163 images in total, in which 110 images are with benign tumors while 53 are with malignant ones. All the images are annotated with the segmentation labels at the pixel level.

Due to the imbalanced classes, before training on



(a) (b)

Fig.7. (a) ROC curves when mask and mask-expanded maps are integrated using DG-CNN (insert mode #1). (b) ROC curves when mask and mask-expanded maps are integrated using DG-CNN (multi-task).

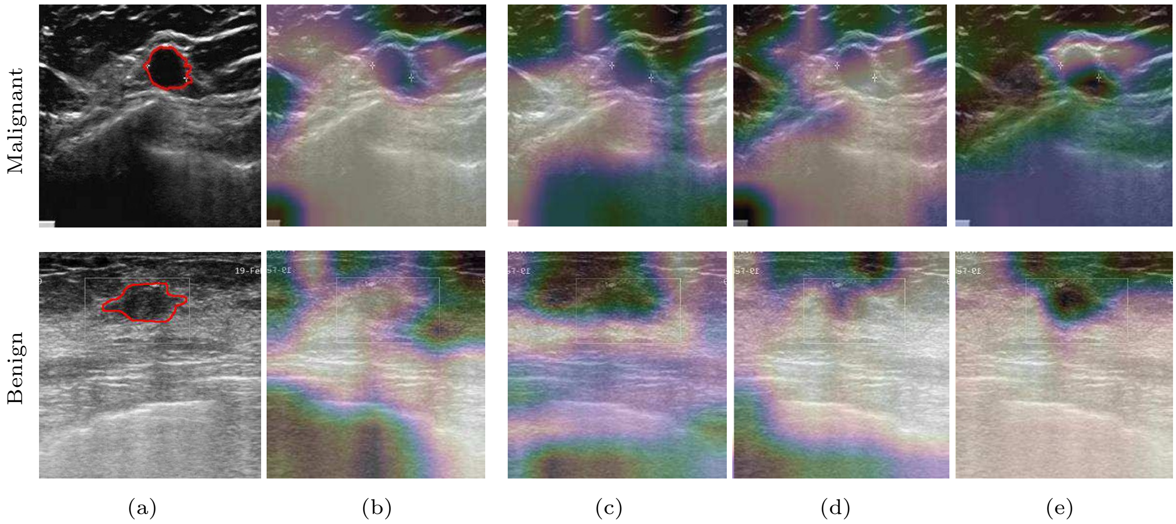


Fig.8. (a) Original images (upper: malignant; lower: benign). (b) CAMs of the baseline. (c)–(e) CAMs of DG-CNN when integrating the semi-RCF map, the mask map and the mask-expanded margin map, respectively, by using DG-CNN (insert mode #1).

the BUS-B dataset, we first augment all images with malignant tumors by horizontally flipping. Then, the 10-fold cross validation method is used to test the diagnostic performance. The performance of the baseline model with the ResNet18 backbone is compared with that of using DG-CNN with insert mode #1 and DG-CNN with multi-task learning mode. In particular, the margin maps of the BUS-B dataset are the prediction results by using the semi-RCF model trained on our dataset.

During the training process, in our baseline model and the model of using DG-CNN with insert mode #1, the batch size is set to 32 for 65 epochs. The learning rate is initialized to 0.0625, and then decreased by 10 at the 20th, 40th and 60th iterations. On the other hand, for the multi-task learning mode, the auxiliary task (i.e., the margin-wise attention generation task) is trained firstly for 100 epochs with the learning rate of 0.01. Other settings are the same with those when training on our dataset.

The quantitative results of the baseline model and using DG-CNN with these two integrating methods based on the BUS-B dataset are listed in Table 6, where the best diagnostic performance for each metric is highlighted. We can see that for almost all evaluation metrics, the diagnostic performance of DG-CNN with these two integrating methods outperforms the baseline model. For example, when compared with the baseline model, the improvements of diagnostic accuracy are 1.38% and 2.76% for DG-CNN (insert mode #1) and DG-CNN (multi-task), respectively. In addition, DG-CNN (insert mode #1) achieves the highest sensitivity, while DG-CNN (multi-task) achieves the highest specificity.

Furthermore, the ROC curves of the baseline model and the DG-CNN with two integrating methods are shown in Fig. 9. We can see in Fig. 9 that using the DG-CNN with insert mode #1 can achieve better performance than the baseline model.

5 Conclusions

In this paper, we showed that the information of medical consensus and guidelines can be integrated into deep neural networks to improve their performance. In particular, we utilized the margin information highlighted in BI-RADS for radiologists when they diagnosed breast cancer in BUS images. We proposed

Table 6. Evaluating the Diagnostic Performance of DG-CNN on BUS-B Dataset

Method	Metric (%)				
	Accuracy	Sensitivity	Specificity	AUC	*F*2
Baseline	80.65	72.22	88.99	84.67	74.71
DG-CNN (insert mode #1)	82.03	**78.70**	85.32	**87.22**	**81.63**
DG-CNN (multi-task)	**83.41**	76.85	**89.91**	85.27	78.90

a scheme named DG-CNN to incorporate this margin information. Within DG-CNN, different methods have been designed to integrate the margin information. We tested the effectiveness of DG-CNN on different datasets and with different network structures, and the results demonstrated the effectiveness of the margin information as well as DG-CNN. To the best of our knowledge, this is the first time that the margin information is utilized to improve the performance of deep neural networks in diagnosing breast cancer in BUS images, and we believe introducing medical domain knowledge into the networks bears great promise for the CAD in medical images.

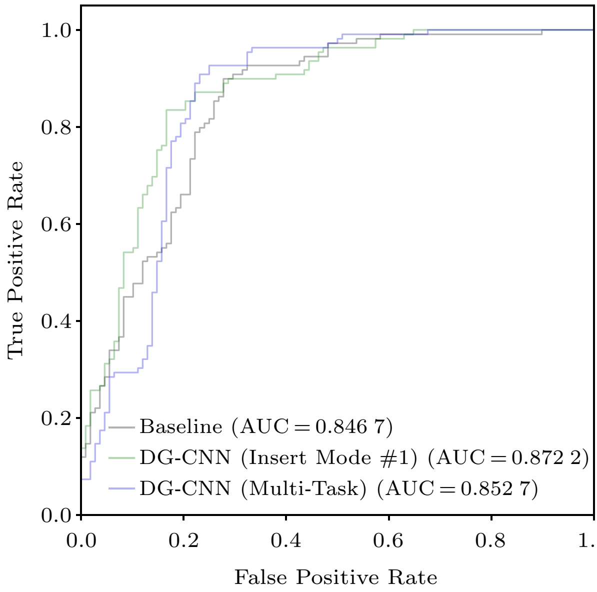


Fig.9. ROC curves of baseline and DG-CNN with two different integrating methods based on the BUS-B dataset.

It should be noted that the proposed DG-CNN has not been tested in a real CAD system. More experimental results on larger datasets are required to confirm its effectiveness, and practical problems like the runtime speed, the requirements for computation and storage resources need to be considered before it can be applied to real CAD systems.

References

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2015. https://arxiv.org/abs/1409.1556, Nov. 2021.

[2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp.770-778. DOI: 10.1109/CVPR.2016.90.

[3] Shin S Y, Lee S, Yun I D, Lee K M. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE Trans. Med. Imaging*, 2019, 38(3): 762-774. DOI: 10.1109/TMI.2018.2872031.

[4] Xu X, Lu Q, Yang L, Hu S X, Chen D Z, Hu Y, Shi Y. Quantization of fully convolutional networks for accurate biomedical image segmentation. In *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp.8300-8308. DOI: 10.1109/CVPR.2018.00866.

[5] Zhou Z, Shin J Y, Zhang L, Gurudu S R, Gotway M B, Liang J. Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp.4761-4772. DOI: 10.1109/CVPR.2017.506.

[6] Esteva A, Kuprel B, Novoa R A, Ko J M, Swetter S M, Blau H M, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 542(7639): 115-118. DOI: 10.1038/nature21056.

[7] Huynh B, Drukker K, Giger M. MO-DE-207B-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. *Med. Phys.*, 2016, 43(6): 3705-3705. DOI: 10.1118/1.4957255.

[8] Yap M H, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, Davison A K, Marti R. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.*, 2018, 22(4): 1218-1226. DOI: 10.1109/JBHI.2017.2731873.

[9] Tajbakhsh N, Shin J Y, Gurudu S R, Hurst R T, Kendall C B, Gotway M B, Liang J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging*, 2016, 35(5): 1299-1312. DOI: 10.1109/TMI.2016.2535302.

[10] 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:1801.09927, 2018. https://arxiv.org/abs/1801.09927, Nov. 2021.

[11] González-Díaz I. DermaKNet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. *IEEE J. Biomed. Health Inform.*, 2018, 23(2): 547-559. DOI: 10.1109/JBHI.2018.2806962.

[12] Li L, Xu M, Wang X, Jiang L, Liu H. Attention based glaucoma detection: A large-scale database and CNN model. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, June 2019, pp.10571-10580. DOI: 10.1109/CVPR.2019.01082.

[13] Fang L, Wang C, Li S, Rabbani H, Chen X, Liu Z. Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE Trans. Med. Imaging*, 2019, 38(8): 1959-1970. DOI: 10.1109/TMI.2019.2898414.

[14] Mitsuhara M, Fukui H, Sakashita Y, Ogata T, Hirakawa T, Yamashita T, Fujiyoshi H. Embedding human knowledge in deep neural network via attention map. arXiv:1905.03540, 2019. https://arxiv.org/abs/1905.03540, May 2021.

[15] Dorsi C, Bassett L, Feisg S, Lee C I, Lehman C D, Bassett L W. Breast Imaging Reporting and Data System (BI-RADS). Oxford University Press, 2018.

[16] Bian C, Lee R, Chou Y, Cheng J. Boundary regularized convolutional neural network for layer parsing of breast anatomy in automated whole breast ultrasound. In *Proc. the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Sept. 2017, pp.259-266. DOI: 10.1007/978-3-319-66179-7_30.

[17] Maicas G, Bradley A P, Nascimento J C, Reid I, Carneiro G. Training medical image analysis systems like radiologists. In *Proc. the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2018, pp.546-554. DOI: 10.1007/978-3-030-00928-1_62.

[18] Liu J, Li W, Zhao N, Cao K, Yin Y, Song Q, Chen H, Gong X. Integrate domain knowledge in training CNN for ultrasonography breast cancer diagnosis. In *Proc. the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2018, pp.868-875. DOI: 10.1007/978-3-030-00934-2_96.

[19] Wang X, Peng Y, Lu Y, Lu Z, Summers R M. TieNet: Text-image embedding network for common thorax disease classification and reporting in chest X-rays. In *Proc. the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp.9049-9058. DOI: 10.1109/CVPR.2018.00943.

[20] Berg W A, Cosgrove D O, Doré C J *et al.* Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. *Radiology*, 2012, 262(2): 435-449. DOI: 10.1148/radiol.11110640.

[21] Dobruch-Sobczak K, Piotrzkowska-Wróblewska H, Roszkowska-Purska K, Nowicki A, Jakubowsi W. Usefulness of combined BI-RADS analysis and Nakagami statistics of ultrasound echoes in the diagnosis of breast lesions. *Clin. Radiol.*, 2017, 72(4): 339-339. DOI: 10.1016/j.crad.2016.11.009.

[22] Liu Y, Cheng M M, Hu X, Wang K, Bai X. Richer convolutional features for edge detection. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp.5872-5881. DOI: 10.1109/CVPR.2017.622.

[23] Arbeláez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, 33(5): 898-916. DOI: 10.1109/TPAMI.2010.161.

[24] Lin M, Chen Q, Yan S. Network in network. arXiv:1312.4400, 2013. https://arxiv.org/abs/1312.4400, March 2021.

[25] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp.2921-2929. DOI: 10.1109/CVPR.2016.319.

[26] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, 39(4): 640-651. DOI: 10.1109/TPAMI.2016.2572683.

[27] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In *Proc. the IEEE International Conference on Computer Vision*, October 2017, pp.2980-2988. DOI: 10.1109/ICCV.2017.322.

[28] Han S, Kang H K, Jeong J Y *et al.* A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys. Med. Biol.*, 2017, 62(19): 7714-7728. DOI: 10.1088/1361-6560/aa82ec.

[29] 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(6): 1137-1149. DOI: 10.1109/TPAMI.2016.2577031.

Xiao-Zheng Xie is currently a Ph.D. candidate with School of Computer Science and Engineering, Beihang University. She received her M.S. and B.S. degrees in computer science and technology from Zhengzhou University, Zhengzhou, in 2014 and 2017, respectively. Her research interests include medical image processing and computer vision.

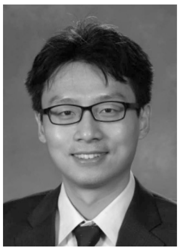
Jian-Wei Niu received his M.S. and Ph.D. degrees in computer science from Beihang University (BUAA), Beijing, in 1998 and 2002, respectively. He was a visiting scholar at School of Computer Science, Carnegie Mellon University, Pittsburgh, from Jan. 2010 to Feb. 2011. He is a professor in the School of Computer Science and Engineering, BUAA, Beijing, and an IEEE senior member. His current research interests include mobile and pervasive computing, and mobile video analysis.

Xue-Feng Liu received his M.S. and Ph.D. degrees in automatic control and aerospace engineering from the Beijing Institute of Technology, Beijing, and the University of Bristol, Bristol, in 2003 and 2008, respectively. He was an associate professor at the School of Electronics and Information Engineering in the Huazhong University of Science and Technology, Wuhan, from 2008 to 2018. He is currently an associate professor at the School of Computer Science and Engineering, Beihang University, Beijing. His research interests include wireless sensor networks, distributed computing and in-network processing. He has served as a reviewer for several international journals/conference proceedings.

Qing-Feng Li is an associate researcher at the Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC), Beihang University, Beijing. He received his M.S. and B.S. degrees in computer software from Zhengzhou University, Zhengzhou, in 2014 and 2017, respectively. His research interests include computer vision, image processing, and computer graphics.

Yong Wang received his M.D. degrees in medicine imaging and nuclear medicine from Peking Union Medical College, Beijing, in 2016. He was a visiting scholar in Department of Imaging and Interventional Radiology at the Chinese University of Hong Kong, Hong Kong, from May 2010 to May 2011. He is a professor in the Department of Diagnostic Ultrasound, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing. His research work mainly focuses on computer-aided diagnosis in medical imaging, the ultrasound (US) diagnosis of breast tumor and interventional ultrasound. He serves as an editor in Quantitative Imaging in Medicine and Surgery.

Jie Han received her M.D. degrees in medical imaging and nuclear medicine from Peking Union Medical College, Beijing, in 2015. She is an attending doctor since 2017 in the Department of Diagnostic Ultrasound, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing. She majors in the application of contrast-enhanced ultrasound in the pancreas and the ultrasound diagnosis of superficial organs. Her work mainly focuses on computer-aided diagnosis in medical imaging, the ultrasound (US) diagnosis of breast tumor, including segmentation and diagnostic modeling establishment.



Shaojie Tang is currently an assistant professor of Naveen Jindal School of Management at University of Texas at Dallas, Richardson. He received his Ph.D. degree in computer science from Illinois Institute of Technology, Illinois, in 2012. His research interest includes social networks, e-business and optimization. Tang served as chairs and TPC members at numerous conferences.