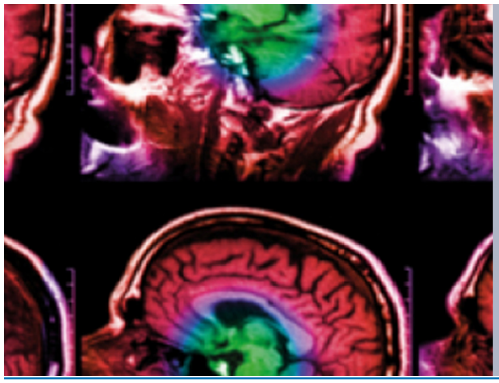
PAPER

A novel approach with dual-sampling convolutional neural network for ultrasound image classification of breast tumors



IOP Publishing Phys. Med. Biol. 65 (2020) 245001 https://doi.org/10.1088/1361-6560/abc5c7

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
2 July 2020

REVISED
21 October 2020

ACCEPTED FOR PUBLICATION
29 October 2020

PUBLISHED
2 December 2020

A novel approach with dual-sampling convolutional neural network for ultrasound image classification of breast tumors

Jiang Xie[1], Xiangshuai Song[1], Wu Zhang[2], Qi Dong[3], Yan Wang[3], Fenghua Li[3] and Caifeng Wan[3]

[1] School of Computer Engineering and Science, Shanghai University, Shanghai, People's Republic of China
[2] Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai, People's Republic of China
[3] Department of Ultrasound, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, People's Republic of China

E-mail: wancaifengky@sina.com

Keywords: breast tumors, computer-aided diagnosis, convolutional neural networks, deep learning, ultrasound imaging, BI-RADS

Abstract

Breast cancer is one of the leading causes of female cancer deaths. Early diagnosis with prophylactic may improve the patients' prognosis. So far ultrasound (US) imaging has been a popular method in breast cancer diagnosis. However, its accuracy is bounded to traditional handcrafted feature methods and expertise. A novel method, named dual-sampling convolutional neural networks (DSCNNs), was proposed in this paper for the differential diagnosis of breast tumors based on US images. Combining traditional convolutional and residual networks, DSCNN prevented gradient disappearance and degradation. The prediction accuracy was increased by the parallel dual-sampling structure, which can effectively extract potential features from US images. Compared with other advanced deep learning methods and traditional handcrafted feature methods, DSCNN reached the best performance with an accuracy of 91.67% and an area under curve of 0.939. The robustness of the proposed method was also verified by using a public dataset. Moreover, DSCNN was compared with evaluation from three radiologists utilizing US-BI-RADS lexicon categories for overall breast tumors assessment. The result demonstrated that the prediction sensitivity, specificity and accuracy of the DSCNN were higher than those of the radiologist with 10 year experience, suggesting that the DSCNN has the potential to help doctors make judgements in clinic.

1. Introduction

Breast cancer is the most common cause of cancer-related death among women worldwide (Siegel *et al* 2018). Because of the high risk of breast cancer, it is primarily important to precisely detect the disease at its early stage, which will contribute to effective prognosis and treatment (Gallego-Ortiz and Martel 2016). There are several methods to detect breast tumors such as digital mammography, ultrasound (US) imaging and magnetic resonance imaging. Among those methods, US is commonly used for breast cancer classification and detection due to its versatility, sensitivity, and safety (Stavros *et al* 1995, Hu *et al* 2019, Moon *et al* 2020, Pi *et al* 2020, Zhang *et al* 2020a).

Recent applications of deep learning in medical US analysis have involved various tasks (Liu *et al* 2019), such as multi-needle detection for US-guided prostate brachytherapy (Zhang *et al* 2020c), segmentation of common carotid intima-media complex in US images (Qian *et al* 2020), and differentiating benign and malignant lesions for detecting lung cancer (Chen *et al* 2019). However, the classification of breast tumors based on ultrasonography images faces several challenges. (1) The quality of US images is influenced by several factors including signal attenuation, speckle noise, artifacts of shadowing and enhancement, poor contrast and low signal-to-noise ratio (Wu *et al* 2019). Those factors will further influence sonographers with varying experience to judge if breast masses are benign or malignant. (2) If an automated model is to learn the features of breast lesions from ultrasonography images, it requires a large quantity of annotated data, the collection and annotation of which are laborious (Pi *et al* 2020). (3) The assessment of US image requires well-trained and experienced radiologists (Cheng *et al* 2010). To address those challenges, various

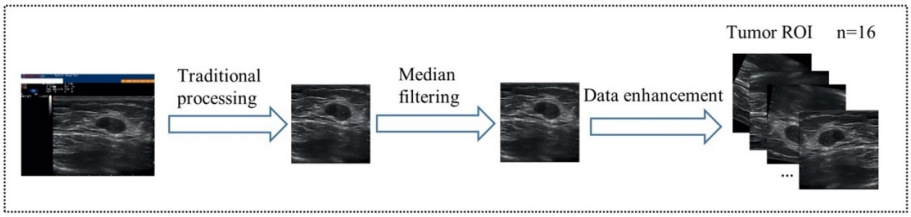
computer-aided diagnosis (CAD) systems have been proposed for radiologists to objectively assess breast masses (Gomez Flores *et al* 2015). In general, a CAD system automatically classifies breast tumors into malignant and benign, providing a more accurate diagnosis for doctors (Cho *et al* 2018). Deep learning based CAD systems is recently gaining increasing attention as an additional objective opinion to assist medical image diagnosis and interpretation (Akkus *et al* 2019).

Classification of breast tumors are based on their sonographic features from breast ultrasound (BUS) images such as margin, shape, echo patterns and posterior features in the early stages (Gomez-Flores *et al* 2019). Several texture analysis approaches have been used for breast lesion classification, including gray level co-occurrence matrix (Gomez-Flores *et al* 2019), gray-level run length matrix (Prabusankarlal *et al* 2015), local binary patter (Liu *et al* 2014) and histogram of oriented gradients (Abdel-Nasser *et al* 2017). In general, texture analysis processes grayscale images to obtain texture features with tumor types distinguished. Differential morphological features of benign and malignant tumors are also an indicator for their feature extraction and selection. For example, a CAD system was proposed to distinguish breast tumors according to their morphologies (Huang *et al* 2008). In addition, other features such as acoustic descriptors and multi-resolution grayscale invariant textures were also adopted for tumor type or stage identification (Nayeem *et al* 2014, Moon *et al* 2015).

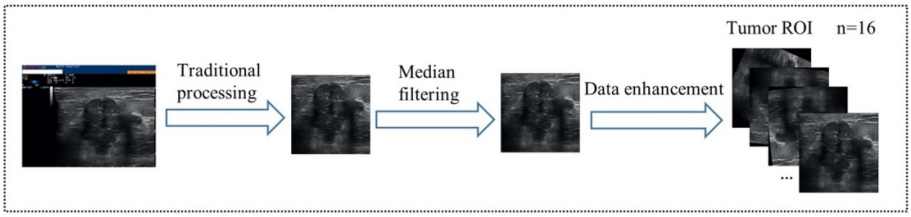
Recently, various machine learning methods, including support vector machine (SVM) and convolutional neural networks (CNNs), have attracted considerable attention in the field of building CAD systems (Fujioka *et al* 2019). Traditional machine learning methods focus on the process of efficient features extraction and lesions classification from US images. A method combining unsupervised technology fuzzy c-means clustering with back-propagation artificial neural network was proposed for benign and malignant classification of BUS (Singh *et al* 2016). Kim *et al* used Otsu's threshold method first to detect lesions on US images and then SVM classifier to classify breast tumors from two-dimensional BUS images (Kim *et al* 2014). In addition, Zhang *et al* suggested an approach called biclustering learning technique for tumor classification. Biclusters obtained from this method were used as training dataset for the *k*-nearest neighbor (KNN) classifier (Zhang *et al* 2014).

Deep learning algorithms, as a branch of machine learning, are getting increasing attention for analyzing medical images because of its ability of automatic and effective feature extraction (Litjens *et al* 2017). In particular, CNN has been applied to automatically classify tumor types from various medical images including ultrasonographic scans, mammograms, fundus and skin images (Tanaka *et al* 2019). Numerous previous studies followed have been developed based on CNNs to distinguish benign and malignant tumors from US images. Since deep learning requires a large training dataset, image data enhancement and GoogLeNet were performed by Han *et al* (2017) to train ImageNet for transferring learning to achieve classification. The networks were trained on the data with and without augmentation. Area under curve (AUC) of the both were over 0.9. The networks showed 0.91 accuracy, 0.86 sensitivity and 0.93 specificity. However, significant data is needed to train this model. Xiao *et al* tried to address the problem of discriminating benign cysts from malignant masses in US images based on five models, including three CNNs transferred models, a CNN model with three convolutional layers and a traditional machine learning-based model with handcrafted features (Xiao *et al* 2018). However, this strategy did not show much improvement. Byra *et al* (2019) proposed to utilize color-dependent representational capacity of a deep CNN to improve breast mass classification in US images. They made use of public datasets and achieved good results. Tanaka *et al* constructed an ensemble network by combining two CNN models fine-tuned on balanced training data with augmentation. A mess-level classification method was adopted to enable CNN to classify a given mass using all views (Tanaka *et al* 2019). Although each of the above methods are generally more powerful than traditional methods, they were all fine-tuned or applied from existing network, which has serious limitations. As far as we know, the advanced learning models including Alexnet, VGG16 (Byra *et al* 2019), ResNet (Xiao *et al* 2018), GoogLeNet (Han *et al* 2017, Fujioka *et al* 2019) and EfficientNet (Tan and Le 2019) were trained on ImageNet, which has much difference with US images dataset. It is due to those disadvantages mentioned that a suitable model is required to efficiently extract features on US images dataset.

In this study, we proposed a method with dual-sampling convolutional neural network (DSCNN) for classification of breast tumors based on US images. Instead of using transfer models from existing deep learning techniques, the DSCNN combining traditional convolutional and residual networks is developed. We looked for the optimal parameters and architecture of the model by experiments primarily. Then, it turned out that our approach performed better by comparing with existing methods and traditional handcrafted feature methods. Additionally, we employed a public dataset to demonstrate the effectiveness of our classification method. To show potential clinical value of the employed methods, evaluation from three radiologists utilizing BI-RADS lexicon categories were compared with our model for overall breast tumors assessment.



(a) Benign tumor image preprocessing architecture



(b) Malignant tumor image preprocessing architecture

Figure 1. Image preprocessing framework of benign and malignant tumors. The framework contains three parts: traditional processing, median filtering, and data enhancement.

2. Materials and methods

2.1. Datasets

A cohort of 373 patients (female; mean age, 49.21 years; range 17–87) was collected from the Renji Hospital-affiliated Shanghai Jiao Tong University School of Medicine between 2018 and 2019 in this study. Each volunteer has signed informed consent and this study has passed an ethical certificate from the local hospital. Some patients have multiple tumors while some patients have only one tumor. In the cases of the benign tumors, the set contained fibroadenoma (58.82%), adenopathy (22.17%), intraductal papilloma (9.95%) and other carcinoids (9.06%). The distribution of malignant tumors is as follows: invasive ductal carcinoma (77.16%), intraductal carcinoma (15.43%) and other malignancies (7.41%). In total, the main dataset called Renji Breast Ultrasound Images (RJBUS) containing 1272 BUS images (800 × 680 pixels). Several images were captured from each patient from different tumor orientation. The RJBUS dataset includes 673 benign and 599 malignant lesions. The main dataset was divided into training and test sets. The training set contained 609 benign and 543 malignant lesions, while the test set contained 64 benign and 56 malignant lesions. All tumor types and properties were confirmed by pathological tissue section. In order to compare with the tradition handcrafted feature methods, the contours of tumors were manually delineated by an experienced radiologist.

To prove the validity of the model proposed here, we also employed a public dataset named UDIAT consisting of 163 breast masses US images (110 benign and 53 malignant) (Yap *et al* 2018). The UDIAT set was collected by Siemens ACUSON scanner from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell (Spain), in 2012. Initially, the authors used this dataset to develop deep learning algorithms for breast masses segmentation. So the original images were tailored to different shapes. A detailed description of the data can be found in the paper (Yap *et al* 2018).

2.2. Data preprocessing

The preprocessing of breast US images contains three steps including traditional processing (cropping, resizing, grayscaling and normalization), denoising and data enhancement. The preprocessing of US images is shown in figure 1. The purpose of image clipping is to remove the nonrelevant information out of region of interest (ROI) such as model number of instrument, time of scanning or imaging, and information of patients. Meanwhile, this operation imitates the process how a radiologist reads a US image. In order to input the US images into the deep neural network (Cao *et al* 2019), we resized it to a voxel size of 128 × 128 × 1, where '1' represents image grayscale.

Medical US images are more advantageous over CT, MIR and x-ray, but they have lower image resolution and serious speckle noise interference. Speckle noise, which affects recognition of ROI from medical US images, is an inherent phenomenon (Jirik and Taxt 2006). In order to improve the performance of neural network, we used the median filter (Gallagher and Wise 2003) to denoise US images. The principle of median filtering is as follows:

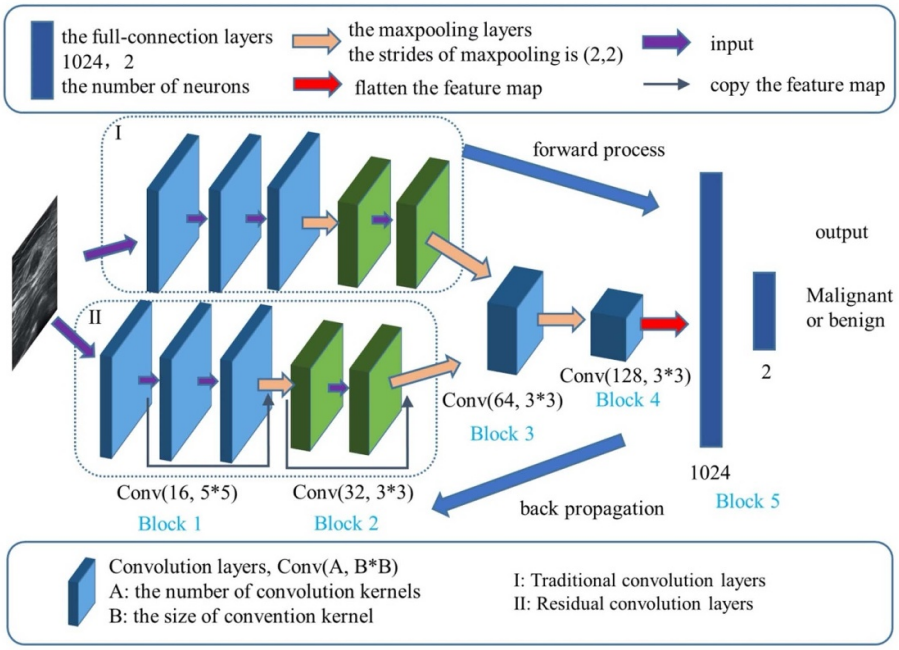


Figure 2. The framework of dual-sampling convolutional neural network (DSCNN). Each convolution layer is followed by batch normalization (BN) layer and rectified linear unit (RELU).

First, an odd-pixel square window F_{jk} centered on pixel point $G(i,y)$ is defined. Then pixel values in the window are placed in an array $\{F_{jk}\}$ in ascending or descending order. Finally, the value in the center of the window is replaced by the median value in the sequence:

$$G(i,y) = \mathrm{Mid}\{F_{jk}\} \tag{1}$$

where $\mathrm{Mid}\{F_{ik}\}$ represents the median of the gray scale in the window.

The effectiveness of deep neural network models is particularly dependent on the amount of training datasets (Salamon and Bello 2017), but the acquisition of medical datasets is very difficult (Al-Dhabyani *et al* 2019). Therefore, to improve the robustness of model and reduce overfitting problem, we used data augmentation methods to extend small data (Krizhevsky *et al* 2017). In this study, we enhanced RJBUS dataset by traditional data augmentation methods and enhancement strategy adopted in the paper (Zhou *et al* 2018). As introduced in the section 2.1, the RJBUS contains 1272 BUS images extracted from 373 patients, including 1152 images for training and 120 images for testing. The training dataset was rotated at four random angles from 0° to 180°, and then vertically and horizontally flipped. Finally, a training dataset of 18 432 ($1152 \times 4 \times 2 \times 2$) US images was generated.

2.3. Dual-sampling convolutional neural network (DSCNN)

2.3.1. The framework of DSCNN

Conventionally, the classification networks of breast tumors are implemented by CNN (Han *et al* 2017, Xiao *et al* 2018, Byra *et al* 2019, Tanaka *et al* 2019, Yao *et al* 2019). These networks automatically extract features for classification, and deep CNNs have led to a series of breakthroughs (He *et al* 2016). The research showed that deep CNNs naturally integrate different levels of features, which generally lead to outstanding performance in classification tasks (Zeiler and Fergus 2014). However, the disadvantages, such as poor network convergence, are gradually exposed to the increase of network depth. Especially, deep CNNs can cause gradient disappearance and degradation on US images. In this study, in order to solve the above problems, we proposed a CNN framework called DSCNN, which used a parallel dual-sampling structure combining traditional CNN and residual network. Gradient disappearance and degradation may affect feature extraction of the network. Since the residual network structure was another branch of DSCNN, the model could prevent gradient disappearance and degradation (Zhuang *et al* 2019, Zhang *et al* 2020b). Therefore, the dual-sampling structure combining with CNN can better extract features. The architecture of DSCNN is shown in figure 2.

The DSCNN was divided into five blocks, i.e. 14 weight layers (12 convolution layers and two fully connected layers) and five max-pooling layers for the basic DSCNN configuration in this framework. In

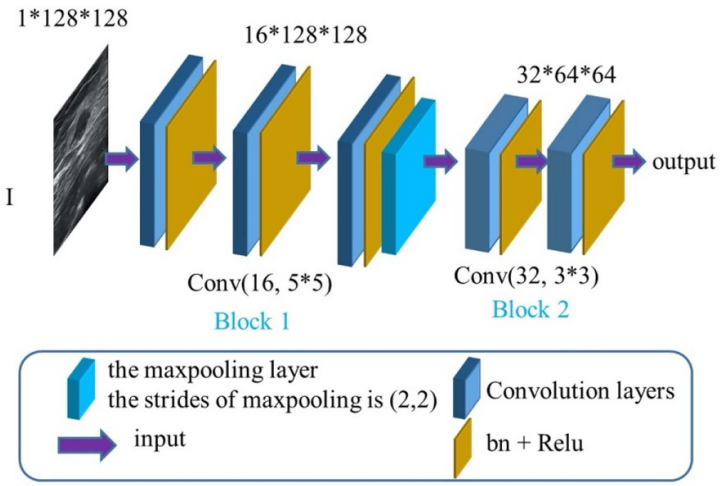


Figure 3. The CNN framework of DSCNN. In $A * B * B$, A represents the dimension of the feature map, and B represents the width and height of the feature map. Block 1 consists of the first three convolution layers, while Block 2 consists of the last two convolution layers.

order to accelerate the convergence speed of the network, we added a batch normalization (BN) layer after each convolution layer (Ioffe and Szegedy 2015). In addition, each BN layer was followed by a Rectified Linear Unit (RELU) function to increase the nonlinear (see section 2.3.2).

In the proposed model, the CNN (figure 2(I)) and residual network (figure 2(II)) were used to extract features from US images, which generates the name of the network DSCNN. Meanwhile, the initial dimension of the network is one-dimensional. The reason for this design is to extract more features without losing network gradient and to save time for model training. The architecture and parameters were designed by experiment (see section 3.4).

2.3.2. The implementation details of DSCNN

In this section, details and parameters including CNN (figure 2(I)), residual network (figure 2(II)), splicing layer (block 3, 4, 5 in figure 2), and model hyper-parameters, are explained according to the designed architecture of the model.

As shown in figure 3, the CNN architecture, which is part I of DSCNN, is described in detail. In this part, there are five convolution layers, and each layer is followed by BN layer and RELU activation function, which was designed for better network convergence. First, the network started from the input layer, i.e.

$$C_0 = X \tag{2}$$

where X denotes the input (Zhou *et al* 2018). Then C_0 is inputted into the convolution layer, and features are extracted by the convolution kernel. Finally, the feature map C_1 is generated. In this CNN model, C_0 is the pre-processed US images:

$$C_1 = \sigma_1 (W_1 * C_0 + b_1) \tag{3}$$

where σ_1 is the nonlinear activation function RELU, w_1 stands for network weight, and b_1 represents network bias for the first layer and $*$ denotes the convolution operation. In the same way C_1 is used as the input for the next layer to compute C_2. After three convolutions, we get C_3 and put C_3 into max pooling layer to reduce dimensions of feature map:

$$C_3 = \text{maxpooling}(C_3). \tag{4}$$

Finally, the feature extraction of the CNN model is completed by convolution of the last two layers.

Residual network (He *et al* 2016) model (the II part from DSCNN) and CNN have the same weight layers and pooling layers. The difference between two networks is whether to copy feature maps by residual blocks. The details of residual network are shown in figure 4. In this model, we can see that there are two feature map replication channels. The network can map the features of different layers identically through replication channels. Since the dimension of feature map is different when the second replication channel is connected, a $1 * 1$ convolution was added to change the dimensions. The details of residual blocks can be seen in right part.

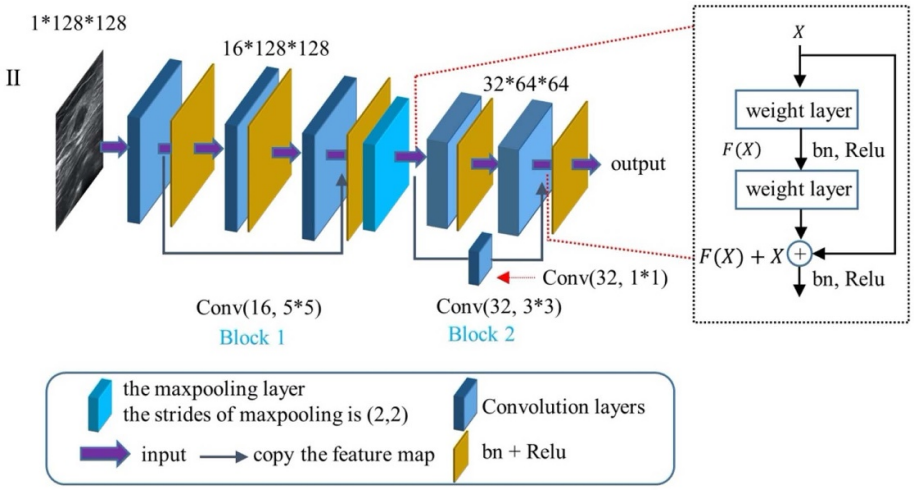


Figure 4. The residual network framework of DSCNN. The right part shows the detail of residual blocks.

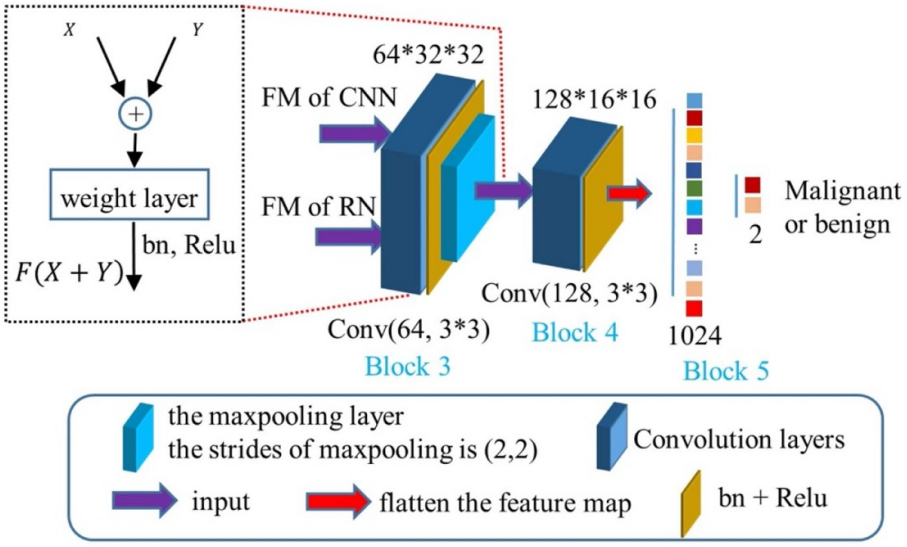


Figure 5. Splicing layer and classification layers. The box in black dash lines shows the details of the splicing layer. X and Y represent CNN feature map and residual network feature map, respectively. FM means feature map, while RN represents residual network.

X represents feature map of the previous layer, as the input of residual block. X was obtained from $F(X)$ after convolution layer operation. Finally, we got result of adding X and $F(X)$ as the input of next residual block.

After part I and part II operations of DSCNN, we got two kinds of feature maps, including CNN and residual network. The details of block 3, block 4 and block 5 are shown in figure 5. Next, we spliced the feature maps that obtained from the two parts according to the idea of the residual block. The spliced feature maps entered the fully connected layer after two convolution layers. To prevent overfitting, a dropout (Dropout probability = 0.5) function is placed in front of each fully connected layer (Srivastava *et al* 2014). Finally, the proposed network ended with softmax activation after two fully connected layers:

$$C_l = \text{softmax}(C_{l-1}) \tag{5}$$

where $\text{softmax}(x)$ is an activation function, C_{l-1} is input of last fully connected layer and C_l is output of last layer (the probability of benign and malignant).

The performance of machine learning algorithms is primarily affected by their hyper-parameters (Guo *et al* 2019). In particular, the deep learning model relies on the setting of hyper-parameters, which can make the model converge and achieve the optimal performance. The detailed implementation hyper-parameters obtained by some experiments were depicted in table 1.

Table 1. The hyper-parameters of DSCNN architecture.

Hyper-parameter	Value
Optimizer	Adam
Learning rate	0.001
Loss function	Cross-entropy
Batch size	8
Epochs	1000

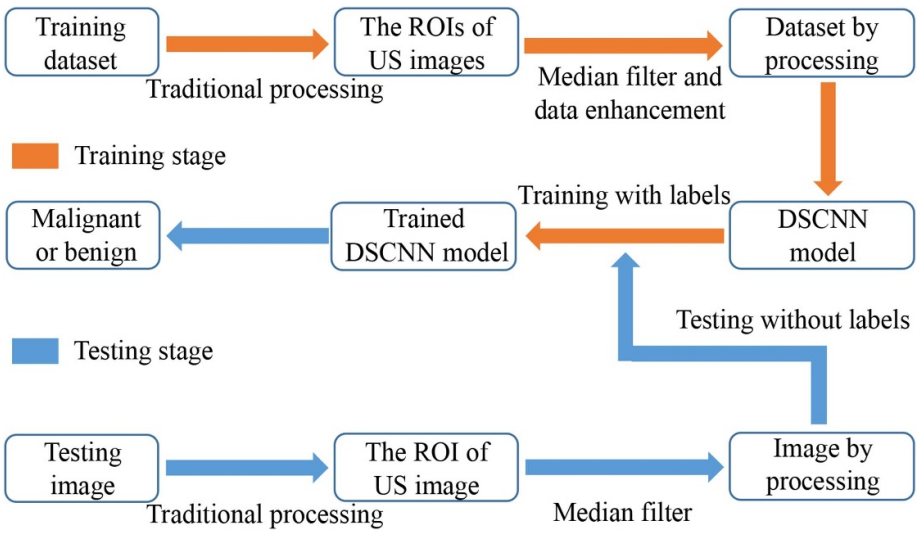


Figure 6. Classification framework of benign and malignant tumors images.

3. Experiments and results

3.1. The experimental flowchart of DSCNN

The dataset was divided into training and test sets, both of which required images preprocessing. Traditional processing and median filter were applied to training and testing, but data enhancement strategy was used in training set only. After the data preprocessing, the breast US images were classified by DSCNN model. As shown in figure 6, the classification of breast tumors images went through training stage and testing stage. In the training stage, large-scale processed US images and corresponding tumor type (benign or malignant) were used to directly train the DSCNN model. In the testing stage, the results can be predicted by feeding the preprocessed test US image into the trained DSCNN model. The training stage and testing stage were automatically processed without manual feature extraction (Huang *et al* 2018).

3.2. Experimental environment

Experiments were performed on the Shanghai University Machine Learning Platform, which has four NVidia Tesla V100 GPUs, 18 NVidia Tesla P100 GPUs, and 2 Intel Xeon Gold 6130 GPUs at 2.1 GHz, 192 GB RAM. The platform of experiments was also under the environment of Python 3.6 version. DSCNN was implemented by Pytorch, a deep learning tool. In addition, traditional feature extraction methods and classifiers were implemented by Scikit-learn package, a common machine learning tool.

3.3. Performance measurement and statistical analysis

In order to evaluate the performance of different methods more comprehensively, we employed six indicators for computation, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and area under the curve of receiver operating characteristic curve (ROC). The higher number of those indicators means the better performance of the model. Here are the formulas for calculating these measurements:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{8}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{10}$$

where TP, true positive, means the number of cases correctly identified as malignant. TN, true negative, is the number of cases correctly identified as benign. FP, false positive, denotes the number of cases incorrectly identified as malignant. And FN, false negative, represents the number of cases incorrectly identified as benign.

Statistical analysis was implemented on Scipy package, a Python-based open-source data processing tool. In order to demonstrate whether there is a significant difference between DSCNN and advanced deep learning techniques, Mann–Whitney *U* test was performed.

3.4. Performance with different network configurations

As mentioned above, neural networks of different depths will extract different features, which will then determine the classification performance of the network. Here, four similar networks were designed to explore the effects of DSCNN with different layers. Among these, DSCNN with seven layers (DSCNN-7) performed the best on the RJBUS dataset (table 3). Moreover, in order to compare the performance between CNN and DSCNN, the CNN with seven layers was also trained on the RJBUS dataset. Table 2 showed configurations of the five models.

Table 3 summarized some test indicators of the five different models. We can intuitively see from table 3 that the DSCNN-7 model has the best performance in accuracy, sensitivity and NPV among DSCNN with different layers, while its specificity and PPV are very close to those of DSCNN-8. Along with deepening of layers, the DSCNN tends to degrade after reaching the best point with seven layers. The possible explanation is the 'fragility' of US images themselves. Once the network reaches the optimal architecture, useful features fail to be extracted from the US images as the network layer deepens.

It also can be found from table 3 that since model CNN-7 has no residual blocks, its performance was worse than DSCNN-7. Thus, it was proved that the residual network has an excellent classification improvement for US images. In the following experiments, we chose DSCNN-7 as the standard and referred it as DSCNN.

3.5. Comparison with advanced methods on RJBUS dataset

At present, there are many studies using CNNs to classify breast tumors as benign or malignant on US images. However, most of these studies were performed on existing network architectures, e.g. GoogLeNet, VGG and ResNet. In order to demonstrate the effectiveness of our approach, we compared those methods with DSCNN. Since the datasets of previous researchers are unavailable, we cannot compare our results with previous studies. We can only compare with original CNN algorithms that supported previous studies on the RJBUS dataset. The difference between our method and five CNNs in terms of accuracy, sensitivity, specificity, PPV, NPV and AUC are shown in table 4.

As we can see from table 4, the DSCNN obviously outperformed other techniques in accuracy, sensitivity, NPV and AUC with statistical significance, and its specificity and PPV were very close to those of GoogLeNet. These ROC curves were illustrated in figure 7.

As illustrated in section 2.2, DSCNN uses grayscale input for the images and adds BN layers, which will extensively save model's training time and speed up its convergence. Figure 8 reported the loss value trends in the training stage of five models and DSCNN on the RJBUS dataset. We can see that the convergence speed of DSCNN is faster than Alexnet, VGG16, ResNet18, and GoogLeNet, especially between DSCNN and Alexnet. Meanwhile, with the increasing of iteration number, the plots of training loss almost converge to a low point. However, the convergence speed of DSCNN is slower than EfficientNet.

Table 2. Different network configurations.

CNN-7	DSCNN-6		DSCNN-7		DSCNN-8		DSCNN-9	
7 layers CNN	6 layers DSCNN		7 layers DSCNN		8 layers DSCNN		9 layers DSCNN	
Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)
Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)
Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)	Conv(16, 5*5)
Max-pooling								
Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)
Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)	Conv(32, 3*3)
Max pooling-2								
Conv(64, 3*3)	Conv(64, 3*3)		Conv(64, 3*3)		Conv(64, 3*3)		Conv(64, 3*3)	
Max-pooling	—		Max-pooling		Max-pooling		Max-pooling	
Conv(128, 3*3)	—		Conv(128, 3*3)		Conv(128, 3*3)		Conv(128, 3*3)	
—	—		—		Max-pooling		Max-pooling	
—	—		—		Conv(256, 3*3)		Conv(256, 3*3)	
—	—		—		—		Max-pooling	
—	—		—		—		Conv(512, 3*3)	
FC(1024)								
FC(2)								

Table 3. Performance with different network configurations. Bold black fonts represent the best performance in each column.

Models	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%
CNN-7	85.83	85.71	85.94	84.21	87.30
DSCNN-6	87.50	85.71	89.06	87.27	87.69
DSCNN-7	**91.67**	**94.64**	89.06	88.33	**95.00**
DSCNN-8	88.33	83.93	**92.19**	**90.38**	86.76
DSCNN-9	86.67	85.71	87.50	85.71	87.50

Table 4. Comparison of DSCNN and other techniques on RJBUS dataset.

Models	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%	AUC
Alexnet	80.83	80.36	81.25	78.95	82.54	0.855[b]
VGG16	84.17	82.14	85.94	83.64	84.62	0.901[a]
ResNet18	82.50	83.93	81.25	79.66	85.25	0.839[a]
GoogLeNet	83.33	73.21	**92.19**	**89.13**	79.73	0.892[a]
EfficientNet	89.17	92.86	85.94	85.24	93.22	0.911
DSCNN	**91.67**	**94.64**	89.06	88.33	**95.00**	**0.939**

The value with 'a' and 'b' means the p-values of the comparisons between DSCNN and other techniques in AUC are <0.05 and <0.01, respectively. Mann–Whitney U test was used to calculated p-value.

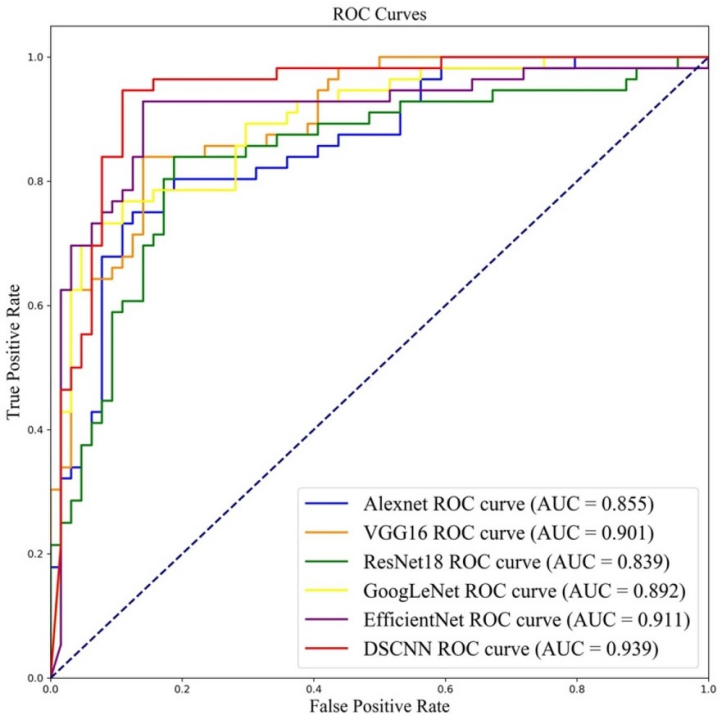


Figure 7. ROC curves of DSCNN and other techniques.

3.6. Verifying the generalization power of DSCNN

In order to prove the robustness and the model generalization ability of DSCNN, we employed the UDIAT dataset for verification. Unfortunately, due to the small size of UDIAT dataset, we were unable to perform DSCNN in an efficient way. Therefore, we trained DSCNN model by combining RJBUS and UDIAT datasets. As in the case of RJBUS, all US images from the UDIAT were first pre-processed to reach the same standard as our dataset. Finally, we obtained 1435 (1272 + 163) US images from two datasets, of which 1283 were used for training and 152 were used for testing. In table 5, the result showed that the overall indicators of DSCNN on the mixed dataset are still higher than those of other advanced methods.

The performance of DSCNN is better than other models with mixed dataset, but is inferior with RJBUS dataset alone. We attribute this result to the difference in the quality of RJBUS and UDIAT datasets. The images of the two datasets were acquired by different machine types in different time periods. As we all known, deep learning models are particularly dependent on data. The performance of the model will

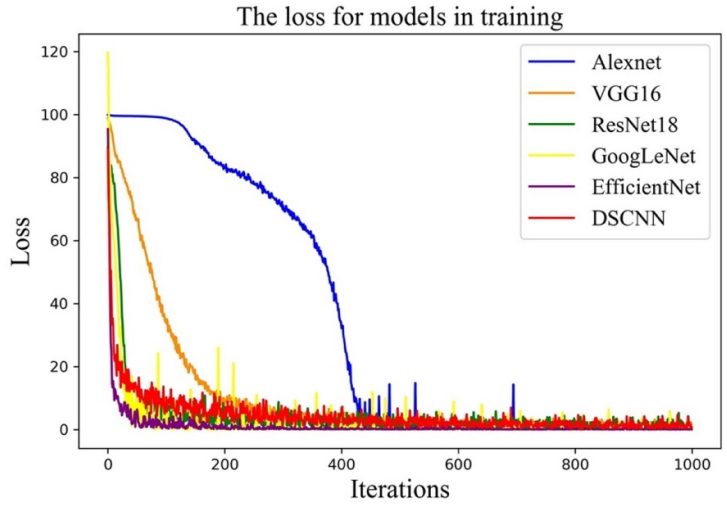


Figure 8. The loss over iterations during the training process.

Table 5. Comparison of DSCNN and other techniques on the combination of RJBUS and UDIAT datasets.

Models	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%
Alexnet	78.28	76.39	80.00	77.46	79.01
VGG16	80.26	73.61	86.25	82.81	78.41
ResNet18	79.61	75.00	83.75	80.59	78.82
GoogLeNet	81.58	77.78	85.00	82.35	80.95
EfficientNet	84.21	**84.72**	83.75	82.43	85.90
DSCNN	**85.53**	83.33	**87.50**	**87.27**	**87.69**

Table 6. Distribution of BI-RADS categories under predictions of three radiologists (radiologist 1 with 2 year, radiologist 2 with 5 year and radiologist 3 with 10 year experience in breast imaging) for test set. Fleiss' kappa was equal to 0.31.

	BI-RADS			
	2	3	4	5
Radiologist 1				
Benign	0	28	36	0
Malignant	0	1	44	11
Radiologist 2				
Benign	0	48	14	2
Malignant	0	9	7	40
Radiologist 3				
Benign	0	48	16	0
Malignant	0	5	42	10

decrease as the data quality deteriorates. Despite this, the performance of DSCNN still exceeded other deep learning techniques.

3.7. Comparison with radiologist readout methods and traditional handcrafted feature methods

In order to demonstrate the clinical significance of our study, three radiologists with different working experiences (radiologist 1 with 2 year, radiologist 2 with 5 year and radiologist 3 with 10 year experience in breast imaging) evaluated US images of test data in RJBUS dataset. They knew nothing about the data before they took part in this experiment. Table 6 shows the BI-RADS distribution of breast tumors categorized by radiologists. In order to verify the consistency of experimental labeling data, we employed Fleiss' kappa coefficient to participate in measurement. The Fleiss' kappa was 0.31, indicating fair consistency in radiologists' final assessment of BI-RADS category.

The traditional handcrafted feature methods extract texture features (e.g. angular second moment, contrast, and dissimilarity) and morphological features (e.g. long axis to short axis ratio, perimeter, and form factor), which are adopted by machine learning. After comparing KNN (Xie *et al* 2017, Huang *et al* 2020), random forest (Xie *et al* 2020), AdaBoost (Xiao *et al* 2018) and SVM on RJBUS dataset, we selected SVM that

Table 7. Performance of the radiologists employing BI-RADS for the benign cutoff set to BI-RADS 3. SVM represents quantitative classification results based on texture features and morphological features.

	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%	AUC
Radiologist1	69.17	**98.21**	43.75	60.44	**96.55**	0.710
Radiologist2	79.17	83.93	75.00	74.60	84.21	0.795
Radiologist3	82.50	91.07	75.00	76.12	90.57	0.830
SVM	74.17	69.64	78.13	73.58	74.63	0.698
DSCNN	**91.67**	94.64	**89.06**	**88.33**	95.00	**0.939**

had the best performance here as the classifier of traditional model. Furthermore, it was compared with radiologist readout methods and DSCNN in table 7.

Here, US-BI-RADS category three was used as benign cutoff. Category two and three were considered as benign tumors, while category four and five were seen as malignant tumors. First three rows showed that the prediction performance was increasing with higher experience of radiologist. In this case, sensitivity of the radiologists was excellent (mean 91.07%), but specificity was lower (mean 64.58%). The accuracy of positive prediction among radiologists is much higher than that of the negative prediction rate, which may be caused by the more obvious features of malignant tumors.

Interestingly, the performance of the traditional features is not as good as that of a radiologist with five years' experience (radiologist 2 and radiologist 3), possibly because of the limitation of the extracted features. However, the predictive indicators of traditional features are more balanced than those of radiologists. Although radiologist with ten years' work experience (radiologist 3) predicted the best results in the accuracy (82.50%) and AUC (0.830) (table 7), the performances were still lower than those of DSCNN model (accuracy: 91.67%, AUC: 0.939) under the same test set.

4. Discussion and conclusions

US imaging is applied as a complementary approach for breast cancer classification and clinical diagnosis reference. However, the low quality of US images adversely influenced doctors' diagnosis. The challenges of breast tumor classification based on US images were already remained. Our study proposed a novel method of DSCNN for classification of breast tumors at US images on RJBUS and UDIAT datasets. In the case of the RJBUS dataset, we designed a series of similar networks and selected the best one through experimental comparisons. The DSCNN achieves the best performance compared with advanced methods and traditional handcrafted feature methods. Additionally, we employed a public dataset named UDIAT to demonstrate the robustness of this model, which also obtained good results. Moreover, three radiologists utilized US-BI-RADS lexicon categories for overall breast tumors assessment. We found that the DSCNN was helpful in distinguishing between benign and malignant masses. It showed better diagnostic performance than other methods and experienced radiologists.

Classification of breast tumors was mainly based on traditional handcrafted features (e.g. texture features, morphological features, and first-order features) in early stage. A CAD system based on handcrafted features generally includes feature extraction, feature selection, and classification. Although this process is more complicated, it built foundation for computer-assisted diagnosis of benign and malignant breast tumors. Our experiments also included the establishment of traditional models based on handcrafted features, but the results of the models were mediocre. The most possible explanation is the lack of useful features. In this model, only texture features and morphological features were extracted. Therefore, traditional classification models that are particularly feature-dependent have certain limitations on the establishment of classification model and are not conducive to the auxiliary diagnosis.

As a part of deep learning, CNN has been favored by some scholars in the process of US image classification of breast tumors (Han *et al* 2017, Xiao *et al* 2018, Fujioka *et al* 2019, Byra *et al* 2019, Tanaka *et al* 2019). To our best knowledge, CNN-based breast tumor classification models are based on fine-tuning or previously developed neural network models, which were originally trained on ImageNet with fixed architectures. This has a certain limitation on the expansion and optimization of the models, which are not well applicable to medical images such as US images. However, with continuous adjustment we experimentally designed a model architecture named DSCNN that is more suitable for US images of breast tumors. This design concept is also reflected in the literature (Zhou *et al* 2018). Similar to the method in literature (Byra *et al* 2019), public dataset was applied to demonstrate the robustness of the model. Fortunately, the model is still excellent compared with other advanced deep learning techniques. Interestingly, we have designed radiologist readout experiments as a diagnostic effect for clinicians.

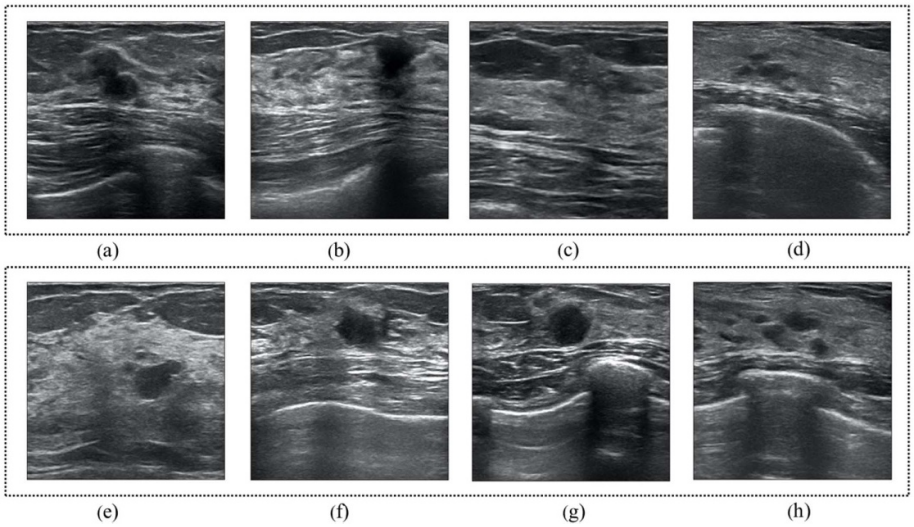


Figure 9. The samples of benign and malignant tumors images that were incorrectly classified. (a), (b), (e), and (f) The images of benign tumors; (c), (d), (g), and (h) the images of malignant tumors. The first row shows the failed images classified by DSCNN and also incorrectly classified by the radiologist. The second row shows the failed images classified by radiologist but correctly classified by the DSCNN.

As is known well, the different pathological and histological characteristics of breast tumors result in different gray-scale ultrasonographic manifestations of breast tumors. The boundary of the malignant tumor is irregular and the internal echo is uneven. It is revealed in the experiments that DSCNN performs better than some advanced deep learning models in accuracy and sensitivity. The possible reason is that the dual-sampling architecture is more sensitive to the edge and the echo distribution of the malignant lesions. However, when identifying benign tumors with smooth edges and uniform internal echoes, the performance is not as good as GoogLeNet (the specificity in table 4). Different CNN architectures can learn different features (Tanaka *et al* 2019). However, GoogLeNet was better at identifying benign tumors.

Although the performance of DSCNN is better than some other advanced deep learning models, there are still cases of failure in US image classification of breast tumors. The analysis of the causes of the failure of classification may help radiologists to make a more accurate diagnosis. Therefore, we analyzed two sets of images here. The first set was misclassified by the DSCNN and incorrectly classified by the radiologist with 10-year experience (as shown in figures 9(a)–(d)). The second was correctly classified by the DSCNN and misclassified by the radiologist(as shown in figures 9(e)–(h)). Among them, (a), (b), (e), and (f) are the images of benign tumors, while (c), (d), (g), and (h) are the images of malignant tumors. In the images ((a)–(d)) of failed DSCNN classification, we can see that the lesion boundaries of the tumors are not very obvious. And what they have in common is that both benign and malignant lesions are relatively small. In particular, the boundary of benign tumors is irregular ((a) and (b)), and the internal echoes of malignant tumors are relatively uniform ((c) and (d)). This seems to be different from the ultrasonographic features of breast tumors. In the images where radiologist failed to classify ((e), (f), (g), and (h)), the features are more consistent with the US features of breast tumors, but the same aspect as before is that the lesion area is also too small.

There are several aspects of this study that we need to explore. First, the input to the DSCNN is grayscale (single-channel) images. As far as we know, some research input three-channel color US images into their models (Han *et al* 2017, Byra *et al* 2019). The study (Han *et al* 2017) investigated the effects of single-channel and three-channel images on the model output. Although the accuracy of the single-channel (90.21%) was slightly lower than that of the three-channel (90.24%), the author applied the single-channel to the following experiments. The benefit of inputting grayscale images is saving training time, which is also the reason we use single-channel US images to train the model. Considering little difference in their accuracy, we improved the model performance by reducing the time required. Second, DSCNN model uses a parallel architecture. There are two reasons for this design. First, US images have a series of defects, including speckle noise, shadowing and enhancement artifacts, poor contrast, and low signal-to-noise ratio (Wu *et al* 2019). We can use double sampling techniques to extract more useful features in the image to overcome the above disadvantage. Second, feature extraction has been well developed previously. Future research direction is to improve the training speed of the model with deep learning technology development (Feng *et al* 2020).

There is no doubt that the parallel architecture can improve the training speed of the model and accelerate model's convergence. This achievement is also reflected in the convergence curve in figure 8.

This study has several limitations. First, the ROI area needs to be marked from the US image by the radiologist before it can be put into the model. If the US image is directly inputted, it may generate some useless information and thus reduce the performance of the model. Nevertheless, cropping ROI is easy to implement and convenient for doctors. Second, this study only utilized US images of breast tumors. Whether DSCNN can be generally applied to other tumor images needs further verification.

The contribution of this work is mainly the development of a novel deep learning model called DSCNN for breast tumors classification. Compared with traditional deep learning model, traditional handcrafted feature methods, and radiologist readout method, the DSCNN model adopts the dual multi-sampling architecture to complete the classification task accurately of breast tumor US images. We demonstrated that DSCNN was more efficient than Alexnet, VGG16, ResNet18, GoogLeNet and EfficientNet on different datasets. The accuracy of the five CNNs in the classification of RJBUS dataset were 80.83%, 84.17%, 82.50%, 83.33%, and 89.17% respectively, while the accuracy of our model is up to 91.67%.

In the future, our work will focus on automatic identification of ROI regions, enhancement of automation, intelligence of model recognition, and application of our models to other types of disease.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61873156).

References

Abdel-Nasser M, Melendez J, Moreno A, Omer O A and Puig D 2017 Breast tumor classification in ultrasound images using texture analysis and super-resolution methods *Eng. Appl. Artif. Intell.* **59** 84–92

Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston A D, Philbrick K A and Erickson B J 2019 A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow *J. Am. Coll. Radiol.* **16** 1318–28

Al-Dhabyani W, Fahmy A, Gomaa M and Khaled H 2019 Deep learning approaches for data augmentation and classification of breast masses using ultrasound images *Int. J. Adv. Comput. Sci.* **10** 618–27

Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'boyle M, Comstock C and Andre M 2019 Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion *Med. Phys.* **46** 746–55

Cao Z, Duan L, Yang G, Yue T and Chen Q 2019 An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures *BMC Med. Imaging* **19** 51

Chen C H, Lee Y W, Huang Y S, Lan W R, Chang R F, Tu C Y, Chen C Y and Liao W C 2019 Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network *Comput. Methods Programs Biomed.* **177** 175–82

Cheng H D, Shan J, Ju W, Guo Y H and Zhang L 2010 Automated breast cancer detection and classification using ultrasound images: a survey *Pattern Recogn.* **43** 299–317

Cho E, Kim E-K, Song M K and Yoon J H 2018 Application of computer-aided diagnosis on breast ultrasonography: evaluation of diagnostic performances and agreement of radiologists according to different levels of experience *J. Ultrasound Med.* **37** 209–16

Feng T Q, Choy M and Laik M N 2020 Predicting book sales trend using deep learning framework *Int. J. Adv. Comput. Sci.* **11** 28–39

Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kasahara M, Oda G, Ishiba T, Nakagawa T and Tateishi U 2019 Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network *Japan. J. Radiol.* **37** 466–72

Gallagher N and Wise G 2003 A theoretical analysis of the properties of median filters *IEEE Trans. Acoust. Speech Signal Process.* **29** 1136–41

Gallego-Ortiz C and Martel A L 2016 Improving the accuracy of computer-aided diagnosis for breast MR imaging by differentiating between mass and nonmass lesions *Radiology* **278** 679–88

Gomez Flores W, De Albuquerque Pereira W C and Catelli Infantosi A F 2015 Improving classification performance of breast lesions on ultrasonography *Pattern Recogn.* **48** 1125–36

Gomez-Flores W, Rodriguez-Cristerna A and De Albuquerque Pereira W C 2019 Texture analysis based on auto-mutual information for classifying breast lesions with ultrasound *Ultrasound Med. Biol.* **45** 2213–25

Guo B, Hu J, Wu W, Peng Q and Wu F 2019 The tabu_genetic algorithm: a novel method for hyper-parameter optimization of learning algorithms *Electronics* **8** 579

Han S, Kang H K, Jeong J Y, Park M H, Kim W, Bang W C and Seong Y K 2017 A deep learning framework for supporting the classification of breast lesions in ultrasound images *Phys. Med. Biol.* **62** 7714–28

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *IEEE Conf. on Computer Vision and Pattern Recognition* (https://arxiv.org/pdf/1512.03385.pdf)

Hu Y, Guo Y, Wang Y, Yu J, Li J, Zhou S and Chang C 2019 Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model *Med. Phys.* **46** 215–28

Huang Q, Huang Y, Luo Y, Yuan F and Li X 2020 Segmentation of breast ultrasound image with semantic classification of superpixels *Med. Image Anal.* **61** 101657

Huang Q, Zhang F and Li X 2018 Machine learning in ultrasound computer-aided diagnostic systems: a survey *Biomed. Res. Int.* **2018** 5137904

Huang Y L, Chen D R, Jiang Y R, Kuo S J, Wu H K and Moon W K 2008 Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound *Ultrasound Obstet. Gynecol.* **32** 565–72

Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *arXiv e-prints* (available at: https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I) (Accessed: 1 February 2015)
Jirik R and Taxt T 2006 High-resolution ultrasonic imaging using two-dimensional homomorphic filtering *IEEE. T. Ultrason. Ferroelectr.* **53** 1440–8
Kim J H, Cha J H, Kim N, Chang Y, Ko M-S, Choi Y-W and Kim H H 2014 Computer-aided detection system for masses in automated whole breast ultrasonography: development and evaluation of the effectiveness *Ultrasonography* **33** 105–15
Krizhevsky A, Sutskever I and Hinton G E 2017 ImageNet classification with deep convolutional neural networks *Commun. ACM* **60** 84–90
Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, Van Der Laak J A W M, Van G B and Sanchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88
Liu H, Tan T, Van Zelst J, Mann R, Karssemeijer N and Platel B 2014 Incorporating texture features in a computer-aided breast lesion diagnosis system for automated three-dimensional breast ultrasound *J. Med. Imaging* **1** 024501
Liu S F, Wang Y, Yang X, Lei B Y, Liu L, Li S X, Ni D and Wang T F 2019 Deep learning in medical ultrasound analysis: a review *Engineering* **5** 261–75
Moon W K, Huang Y-S, Lo C-M, Huang C-S, Bae M S, Kim W H, Chen J-H and Chang R-F 2015 Computer-aided diagnosis for distinguishing between triple-negative breast cancer and fibroadenomas based on ultrasound texture features *Med. Phys.* **42** 3024–35
Moon W K, Lee Y W, Ke H H, Lee S H, Huang C S and Chang R F 2020 Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks *Comput. Methods Programs Biomed.* **190** 105361
Nayeem M A R, Joadder M A M, Shetu S A, Jamil F R and Al H A 2014 Feature selection for breast cancer detection from ultrasound images *2014 Int. Conf. on Informatics, Electronics & Vision (ICIEV)* (https://www.computer.org/csdl/proceedings-article/iciev/2014/06850813/12OmNBLdKRB)
Pi Y, Chen Y, Deng D, Qi X F, Li J L, Lv Q and Yi Z 2020 Automated diagnosis of multi-plane breast ultrasonography images using deep neural networks *Neurocomputing* **403** 371–82
Prabusankarlal K M, Thirumoorthy P and Manavalan R 2015 Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound *Hum-Centric. Comput. Inf.* **5** 12
Qian C, Su E and Yang X 2020 Segmentation of the common carotid intima-media complex in ultrasound images using 2-D continuous max-flow and stacked sparse auto-encoder *Ultrasound Med. Biol.* **46** 3104–24
Salamon J and Bello J P 2017 Deep convolutional neural networks and data augmentation for environmental sound classification *IEEE Signal. Proc. Lett.* **24** 279–83
Siegel R L, Miller K D and Jemal A 2018 Cancer statistics, 2018 *CA-Cancer. J. Clin.* **68** 7–30
Singh B K, Verma K and Thoke A S 2016 Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images *Expert Syst. Appl.* **66** 114–23
Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929–58
Stavros A T, Thickman D, Rapp C L, Dennis M A, Parker S H and Sisney G A 1995 Solid breast nodules: use of sonography to distinguish between benign and malignant lesions *Radiology* **196** 123–34
Tan M and Le Q V 2019 EfficientNet: rethinking model scaling for convolutional neural networks (arXiv:1905.11946)
Tanaka H, Chiu S-W, Watanabe T, Kaoku S and Yamaguchi T 2019 Computer-aided diagnosis system for breast ultrasound images using deep learning *Phys. Med. Biol.* **64** 235013
Wu T, Sultan L R, Tian J, Cary T W and Sehgal C M 2019 Machine learning for diagnostic ultrasound of triple-negative breast cancer *Breast Cancer Res. Treat.* **173** 365–73
Xiao T, Liu L, Li K, Qin W, Yu S and Li Z 2018 Comparison of transferred deep neural networks in ultrasonic breast masses discrimination *Biomed. Res. Int.* **2018** 4605191
Xie J, Sun J, Feng J, Yang F, Wang J, Wen T and Nie Q 2020 Kernel differential subgraph analysis to reveal the key period affecting glioblastoma *Biomolecules* **10** 318
Xie J, Wang H, Zhang J, Chao M, Kong Y, Mao S, Xu L and Wu Z 2017 A novel hybrid subset-learning method for predicting risk factors of atherosclerosis *2017 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)* (https://doi.org/10.1109/BIBM.2017.8217987)
Yao H, Zhang X, Zhou X and Liu S 2019 Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification *Cancers* **11** 1901
Yap M H, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, Davison A K and Marti R 2018 Automated breast ultrasound lesions detection using convolutional neural networks *IEEE J. Biomed. Health* **22** 1218–26
Zeiler M D and Fergus R 2014 Visualizing and understanding convolutional networks (Cham: Springer International Publishing) pp 818–33
Zhang E L, Seiler S, Chen M L, Lu W G and Gu X J 2020a BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis *Phys. Med. Biol.* **65** 125005
Zhang Q, Chang H, Liu L, Li A, and Huang Q 2014 A computer-aided system for classification of breast tumors in ultrasound images via biclustering learning *Int. Conf. on Machine Learning and Cybernetics* pp 24–32
Zhang Q C, Bai C C, Liu Z, Yang L T, Yu H, Zhao J Y and Yuan H 2020b A GPU-based residual network for medical image classification in smart medicine *Inf. Sci.* **536** 91–100
Zhang Y P *et al* 2020c Multi-needle detection in 3D ultrasound images using unsupervised order-graph regularized sparse dictionary learning *IEEE Trans. Med. Imaging* **39** 2302–15
Zhou Y, Xu J, Liu Q, Li C, Liu Z, Wang M, Zheng H and Wang S 2018 A radiomics approach with CNN for shear-wave elastography breast tumor classification *IEEE Trans. Bio-Med. Eng.* **65** 1935–42
Zhuang Z M, Li N, Raj A N J, Mahesh V G V and Qiu S M 2019 An RDAU-NET model for lesion segmentation in breast ultrasound images *Plos One* **14** e0221535