IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 3, 2020, accepted March 20, 2020, date of publication March 23, 2020, date of current version April 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982767

Thyroid Nodule Ultrasound Image Classification Through Hybrid Feature Cropping Network

RUONING SONG[1], (Student Member, IEEE), LONG ZHANG[1], CHUANG ZHU[1], (Member, IEEE), JUN LIU[1], (Member, IEEE), JIE YANG[1], (Member, IEEE), AND TONG ZHANG[2]

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Ultrasonic Medical Center, The First People's Hospital of Chenzhou, Chenzhou 423000, China

Corresponding authors: Chuang Zhu (czhu@bupt.edu.cn) and Tong Zhang (18973575600@189.cn)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4182044, in part by the Research Foundation of Science and Technology Department of Hunan Province, China, under Grant 2018SK50303, and in part by the Research Foundation of Hunan Health Commission under Grant C20190072.

ABSTRACT With the increasing cases of thyroid malignant tumors, the diagnosis of thyroid nodule has attracted more and more attention. Deep learning has achieved promising results in computer-aided diagnosis due to the advantages of obtaining high-dimensional features. In this paper, we proposed a hybrid multi-branch convolutional neural network based on feature cropping method for feature extraction and classification of thyroid nodule ultrasound images. Firstly, we designed a backbone convolutional neural network to extract shared feature maps and a classification network as global branch. Next, we added a feature cropping branch in the network to perform multi-cropping on batch feature maps, to reduce the impact on classification caused by the similarity of local features between benign and malignant thyroid nodule images. Finally, based on softmax predictions of different branch feature maps, we utilize a weighted cross-entropy loss function to train our proposed binary-classification network. Experimental results show that our proposed method has achieved 96.13% accuracy, 93.24% precision, 97.18% recall, and 95.17% F1-measure in public dataset and local dataset, outperforming other models.

INDEX TERMS Thyroid nodule classification, ultrasound image, hybrid multi-branch network, feature cropping, weighted loss function.

I. INTRODUCTION

Thyroid nodules are solid or cystic lumps in the thyroid gland, which is the most common clinical symptom of thyroid surgery. According to statistics, in 2018, nearly 129,500 cases of thyroid cancer were diagnosed in China, accounting for 4.46% of the incidence of malignant tumors [1]. Because thyroid nodule is quite common (more than 60% of lifetime risk) and malignant tumors are extremely harmful, early diagnosis of malignant thyroid nodule can greatly reduce the degree of potential risks. In the past two decades, the examination technique of thyroid nodules have developed significantly. High-resolution ultrasound examination, relying on its advantages of non-surgical, rapid, accurate and convenient, has become an indispensable method for early screening and early diagnosis of thyroid nodule. Radiologists have listed some crucial features of thyroid nodule ultrasound image as important indicators for judging thyroid nodule malignancies, such as composition, shape, echo, edge, etc. [2], [3]. Because thyroid nodule is quite common (more than 60% of lifetime risk) and malignant tumors are extremely harmful, but early diagnosis of a malignant thyroid nodules can greatly reduce the degree of potential risks, so it is significant to accurately distinguish between benign and malignant based on ultrasound image.

In recent years, some computer-aided diagnoses (CAD) based machine learning method have also been applied to feature extraction and classification for ultrasound images. In 2012, Acharya *et al.* extracted the features of thyroid ultrasound image based on the discrete wavelet transformation (DWT) method and classified them using *k*-Nearest Neighbor (KNN) classifier, achieving a classification accuracy of 98.9% on their dataset [4]. Gopinash *et al.* designed a text-based automated diagnostic system in 2013, employed a

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

Gabor filter to extract the features of thyroid ultrasound image and support vector machine (SVM) classifier for classification, achieving an accuracy rate of 96.7% [5].

However, traditional machine learning is not suitable for all datasets and all tasks (for example feature extraction in computer vision), while convolutional neural networks (CNNs) gradually manifest advantages in the image classification tasks because their robustness to image translations and rotations. Different CNNs are adopted to process ultrasound image [6], [7]. Although the CNN-based method is a more advanced approach than traditional machine learning, the networks generally need more model parameters and higher computing burden in the others inference stage, and thus they are more complex. This will cause several problems: big store space requirement, large run-time memory consumption during inference.

This paper built a convolutional neural network to extract features and classify thyroid nodule ultrasound image. In detail, to address the problem of excessive parameters, we designed a convolutional network with four convolutional layers and take it as a backbone feature extraction network. This backbone network is to extract a set of feature maps that can distinguish the difference between benign and malignant nodule in ultrasound image as much as possible. However, because of biological characteristics and ultrasound imaging methods, some areas in malignant nodule ultrasound image are not much more different from benign nodule ultrasound images. Therefore, the feature vectors obtained in these areas will be also very similar, which is not conducive to the subsequent classification task. To reduce this problem, we designed the feature cropping method to crop the shared feature maps from backbone network into a series of local feature maps in different positions and combine independent classification results of each local feature map. So, our proposed hybrid network consists of two branches: one global branch and one feature cropping branch containing feature cropping method. Finally, we combine two predicted results of two branches to get the final result. On the four metrics of accuracy, precision, recall and F1-measure, the results are 96.13%, 93.24%, 97.18%, and 95.17%, respectively, better than other compared networks.

Therefore, the main contributions of this paper are summarized:

1) We propose a two-branch architecture-based hybrid network with feature cropping branch as a subnetwork. Using feature cropping reduces the amount of calculation compared to using image cropping directly. In implementing feature cropping branch, we proposed two feature cropping methods. And after experiments, we found that boundary feature cropping method has the best effect for thyroid nodule diagnosis.
2) For a medical diagnosis, we propose the weighted loss function based on cross-entropy for this binary-class classification. Because the consequences of malignant judgment as benign in the nodule diagnosis are intensely serious. Using this loss function can control the prediction to bias towards a certain class by adjusting different weights to ensure that malignant class has a higher recall rate.

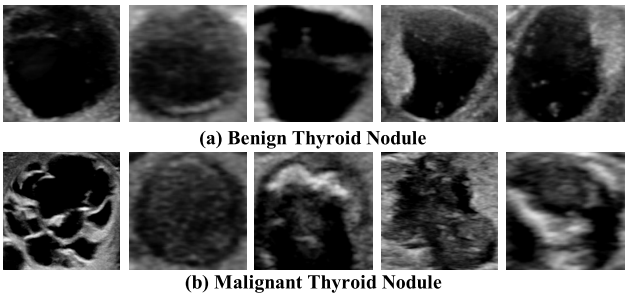


FIGURE 1. Some examples of benign and malignant thyroid nodules in ultrasound image. Images in first line are benign, and others are malignant.

The rest of this paper is organized as follows. In section II, we present some related works about thyroid nodule ultrasound diagnoses, deep learning classification methods, image multiple cropping methods. Section III provides the detailed description of our proposed method, introducing the overall architecture and sub-branch design. Section IV gives our thyroid ultrasound image datasets and experimental details. Conclusions and future works are drawn in section V.

II. RELATED WORK

In this section, we briefly review the thyroid computer-aided diagnoses systems, the deep learning networks built for image classification and image multiple cropping method in state-of-the-art deep learning.

A. THYROID NODULE ULTRASOUND DIAGNOSES

Medical ultrasound is a widely used non-invasive imaging modality that reveals internal anatomy. Ultrasound uses a transducer to emit ultra-high frequency sound waves that will change direction when encountering a reflective surface. The precise timing of the transmitted sound signals and its observed echo is used to determine the anatomy, as shown in Fig. 1. The paper [8] presents a novel active contour model, named joint echogenicity-texture (JET) for precise delineation of thyroid nodules of various shapes according to their echogenicity and texture, as displayed in ultrasound images. Kollorz *et al.* reported a semi-automatic segmentation approach for the classification, and analysis of the thyroid gland based on 3-D US data [9]. Based on the ultrasound image features of thyroid nodules, the American College of Radiology (ACR) has published a professional thyroid imaging report and data system (TI-RADS) based on thyroid nodule ultrasound indicators [10]. TI-RADS points are in the range from 0 to 7, respectively, expressed as ''benign'', ''not suspicious'', ''mildly suspicious'', ''moderately suspicious'' and ''highly suspicious'', are used to rank the risk of a malignant thyroid nodule. Besides, ultrasound image is also widely used in the detection of various diseases, such as the breast cancer [11]–[13], the liver cancer [14], the gastrointestinal disease [15], the cardiovascular diseases [16],

spine curvature [17], and the muscle disease [18], [19]. These medical CADs based on ultrasound images have achieved excellent results and have good inspiration for our research.

B. DEEP LEARNING CLASSIFICATION METHOD

With the development of GPU and distributed frameworks [20]–[22], the bottleneck of time-consuming training has been effectively resolved, and deep learning has become the state-of-the-art method for computer vision, specially image classification [23]. In ILSVRC, deep convolutional neural networks, such as AlexNet, VGGNet, Inception and ResNet, greatly improved the accuracy of image classification through optimization of structure and depth [24]–[27]. Recently, because fine-grained image classification needs more discriminative details, attention mechanisms of deep neural networks have been studied, which was proven useful in many vision tasks. Thus various attention mechanisms have been proposed for image recognition and salient object detection in recent years [28]–[33]. In paper [30], authors encoded the proposed network with two novel attention mechanisms, i.e., landmark-aware attention and category-driven attention, for enhancing clothing category classification.

Typically, deep learning method requires a large number of annotated samples to perform the training task, so there are some very common challenges in medical image processing [34]. One challenge is to collect such huge dataset and perform the annotation on new images. Another challenge to overcome is to collect a suitable set of negative samples which will cause class imbalance. Data augmentation is the most commonly adopted method to increase the size of the training dataset, such as cropping, rotating, and flipping input images [35]. There are other solutions to address these issues, e.g., transfer learning [36], [37], weakly supervised learning [38], sparse annotation [39]. Meng *et al.* employed VGGNet transferred from ImageNet Dataset [40] to distinguish the level of liver fibrosis [41]. Zhu *et al.* [42] researched breast cancer histopathology image classification by assembling multiple compact CNNs. Chi *et al.* utilized GoogLeNet fine-tuned by Caffe framework to differentiate malignant thyroid nodules and benign nodules [6]. Reference [7] employed cascade deep CNNs to develop and evaluate a fully automatic detection of thyroid nodules from ultrasound images.

C. IMAGE MULTIPLE CROPPING IN DEEP LEARNING

In medical, all researchers participated in the challenge with no enough data used for training and testing. The easiest and most common method to reduce overfitting on image data is to artificially enlarge the dataset using label-preserving transformations [24], [35]. And multiple cropping is an effective method for additional data transformations in both training and testing [43]. One example of image multiple cropping is shown in Fig. 2. In AlexNet design, translated image crops of 224 × 224 pixels were selected from a original image of 256 × 256. While the cropped pixels are most likely less informative than the middle pixels we found that making use

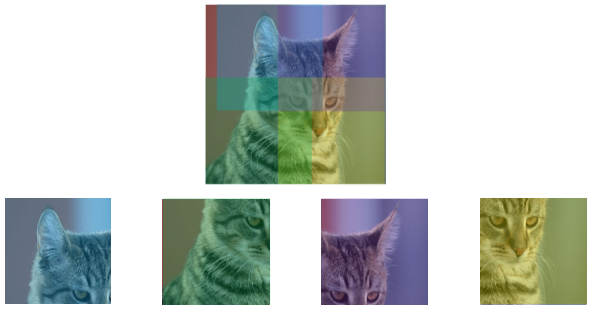


FIGURE 2. One example of image multi-crop. In AlexNet, Alex *et al.* used this method to extract a series of random 224×224 patches from one 256×256 image and trained network on these extracted patches.

of these additional pixels improved the model. This yields a large number of additional training examples and helps the net learn more extensive translation invariance. Howard [43] found that predicting at three different scales improved the joint prediction. Szegedy *et al.* [26] adopted multi-crop methods during testing to obtain a higher performance. Specifically, this paper expanded the image to 144 crops per image by a set of transformations. In VGGNet, Simonyan *et al.* converted one image to 150 crops for testing, and found that using multiple crops performs slightly better than dense evaluation.

However, the image cropping method applies convolutional neural network to each cropping image for computing features repeatedly. Feature sharing is a suitable method to compute only once per image. Wang *et al.* designed a hybrid network in which two sub-networks shared the same full-image convolutional feature map, and thus are computationally efficient [44]. To reduce the bottleneck on region proposal computation, Faster RCNN [45] also introduced novel region proposal networks (RPNs) that share convolutional layers with state-of-the-art object detection networks.

III. PROPOSED METHOD

In this section, we propose our thyroid nodule ultrasound image classification scheme. Firstly, we introduce the proposed hybrid feature cropping network architecture that includes global branch and feature cropping branch. Then, we present the details of global branch, feature cropping branch and weighted loss function.

A. HYBRID CNN ARCHITECTURE

A challenge for us is that the malignant nodule images include healthy tissues, and these regions are likely to get similar features with benign nodules after convolutional neural networks. To acquire more useful information from convolutional feature map, our proposed network is a hybrid network which consists of global branch and feature cropping branch. For a thyroid nodule ultrasound image, a feature extending network is performed first to generate feature map. A series of feature cropping maps and the origin feature map are predicted by a separate branch. Because the central region is the most critical for judging malignancy and benignity

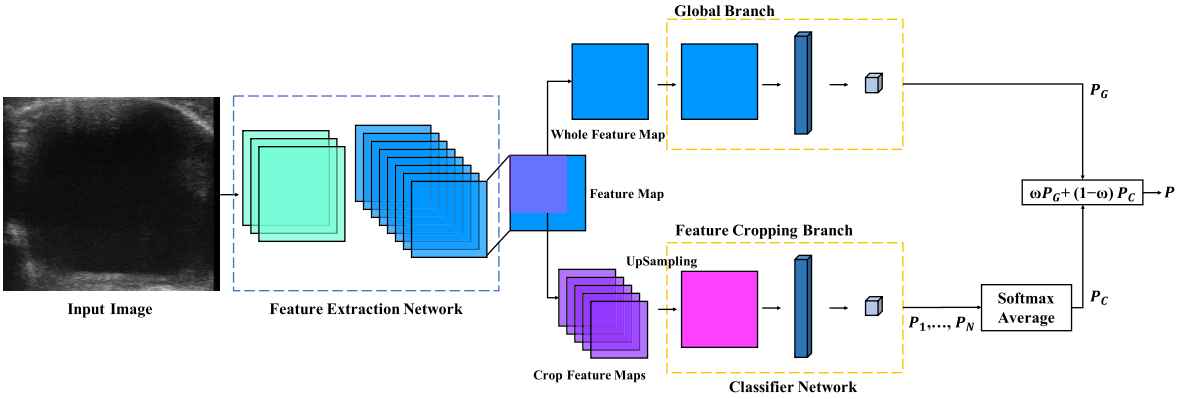


FIGURE 3. The structure of our proposed hybrid feature cropping network with global branch and feature cropping branch. Global branch and feature cropping branch share the feature maps of the backbone feature extraction network. Global branch classifies the original feature map directly. And feature cropping branch generate a series of sub-maps by cropping origin feature map, classifies by the same network as the global branch after upsampling convolution, and predicts branch result by averaging multiple softmax results. Finally, we utilize correlation weight method on the two branch results to get the final prediction.

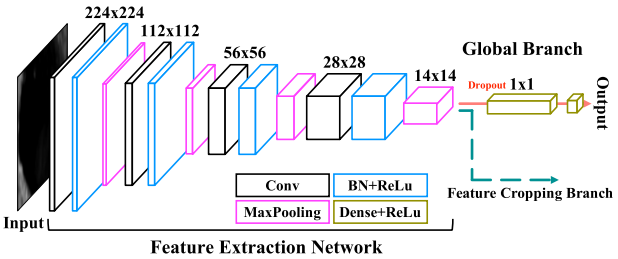


FIGURE 4. Our proposed feature extraction network. Global branch and feature cropping branch share feature maps.

while the surrounding area will sometimes generate similar features, we propose a hybrid cutting network to avoid this adverse effect on the classification and get multi-scale features in the meanwhile.

The proposed framework of our hybrid CNN architecture is shown in Fig. 3. The produced crop maps are passed to the classifier method, and N softmax predictions $(P_1, P_2, \ldots, P_N)$ are yielded for the N feature crop maps. This final softmax output P_C for the feature crop branch is generated by averaging N softmax predictions. On the other hand, the origin feature map is put into the global branch and the prediction P_G is obtained. Finally, the feature crop branch prediction P_C and the global prediction P_G are weighted together by weight ω, as shown in (1).

$$P = \omega P_G + (1-\omega)P_C \quad \omega \in [0,1] \tag{1}$$

B. FEATURE EXTRACTION NETWORK AND GLOBAL BRANCH

The global branch is commonly used for providing global feature representations in multi-branch network architectures. As the ultrasound images are textures-like images, there are no obvious large scale or high level features for the network to learn. The network with less convolutional layer performs even better than networks with multiple convolution layers. In this network, Batch Normalization (BN) method [46] and

TABLE 1. The outline of the proposed network architecture without feature cropping branch.

Layers	Parameter	Output size
Convolution	3×3,32,stride=1	32@224×224
Max_Pooling	3×3,stride=2	32@112×112
Convolution	3×3,64,stride=1	64@112×112
Max_Pooling	3×3,stride=2	64@56×56
Convolution	3×3,128,stride=1	128@56×56
Max_Pooling	3×3,stride=2	128@28×28
Convolution	3×3,128,stride=1	128@28×28
Max_Pooling	3×3,stride=2	128@14×14
Fully_Connected		1×1×512
Fully_Connected		1×1×2
Softmax		1×1×2

Dropout method [47] are also combined together to ensure classification performance. The less-layer network architecture also reduced the number of parameters to be learned, which helps to reduce the over-fitting problem. So we employ a 4-layer 3 × 3 convolutional neural network to extract features, as shown in Fig. 4. Table 1 illustrates the details of our proposed CNN. The three prominent feature of CNN are localized receptive fields, weight sharing and spatial pooling. By combing the three ideas above convolutional neural networks can represent learning, and achieve some extend of drift, scaling and deformation invariability. Pooling is another important concept in convolutional neural networks. It is actually a form of down-sampling. In our work, we have adopted max-pooling as our pooling method, which divides the input image into several rectangular areas and outputs a maximum value for each sub-area. The reason this mechanism can be effective is that, after a feature is discovered, its precise location is far less important than its relative position to other features. The pooling layer will continuously reduce the size of the data space, so the number of parameters and the amount of calculation will also decrease, which controls

the over-fitting to some extent. Besides, the BN method is adopted to allow the utilization of much higher learning rates and be less careful about initialization by normalizing layer inputs, which ensures a high robustness of our model. Then, we use a 50% dropout to reduce the inter-node co-dependence to achieve regularization of the neural network and reduce its structural risk. Finally, extracted features are input into an excitation operation, which is implemented with two fully-connected layers and two corresponding activation layers (ReLu [48]).

C. FEATURE CROPPING BRANCH

In this paper, we propose the feature cropping branch to process the feature maps of thyroid nodule ultrasound image. In practice, multiple crops would increase computation time [25] because it performs a CNN forward pass for each cropping images, without sharing computation. SPPnet [49] and Fast RCNN [50] were proposed to speed up training and testing by sharing computation. Feature cropping method is also proposed to speed up image crop by sharing computation. This method computes a convolutional feature maps for the entire input image and then classifies each cropped feature maps using a feature vector extracted from the shared feature map. In our network, feature cropping branch is designed to share feature extract network with global branch. The implementation of feature cropping method is similar to the batch feature erasing method proposed in the paper [51], which randomly crops away a block of the learned feature maps for all images in the same batch, achieving impressive results in the person re-identification task. For the width w and height h of origin maps, a constant value α from 0 to 1 is defined to represent the ratio between the cropped and the origin feature map. So the size of the cropping map is $\alpha w \times \alpha h$.

In this method, the original feature matrix obtained by feature extraction network is cropped to a certain scale to get different feature maps. Therefore, the features of relatively small scale are obtained, so that the discriminating results are no longer completely dependent on the global features. More importantly, because the central region has a relatively important influence on the discrimination, when the value of factor α is greater than 0.5, the all cropped feature regions retain the part features of the central area and select the edge area of a specific position, while the features of the rest edge areas are removed. In this way, the weight of the selected edge area is increased and the interference of other edge area is avoided during training and classification, leading to higher accurate results. At training and testing time, feature cropping branch predicts by averaging the predictions made by softmax layer on all crops, which examples are visualized in Fig. 5.

In particular, we have designed two different methods to crop feature maps.

1) **Random Feature Cropping (RFC)**: Similar to the traditional image multi-cropping method, given the feature maps computed by feature extract network from one batch of input images, the feature cropping method randomly crops the same matrices with R_c $(\alpha w \times \alpha h)$, which examples are visualized in Fig. 6(a).
2) **Boundary Feature Cropping (BFC)**: Unlike the random cropping feature area, we design the boundary feature cropping (BFC) method to crop the feature map in feature maps from one batch of input images. Select the sub-map R_c $(\alpha w \times \alpha h)$ as the first cutting position as the first image shown in Fig. 6(b). Then we use the cropping window to slide along the edge of the original feature map in steps $(S_h,0)$ and $(0,S_w)$ to select the cropping feature map, and obtain the feature map subset $(R_1, R_2, R_3, R_4, \ldots, R_N)$. The size N of subset is shown in (2). For example, the size N will be 4, while S_h is set to $(1\text{-}\alpha)h$ and S_w is set to $(1\text{-}\alpha)w$. Together with the central map cropped in the middle position, there will be five available sub-feature maps, which examples are visualized in Fig. 6(b).

$$N = (1 + \frac{(1-\alpha)h)}{S_h}) * (1 + \frac{(1-\alpha)w)}{S_w}). \quad (2)$$

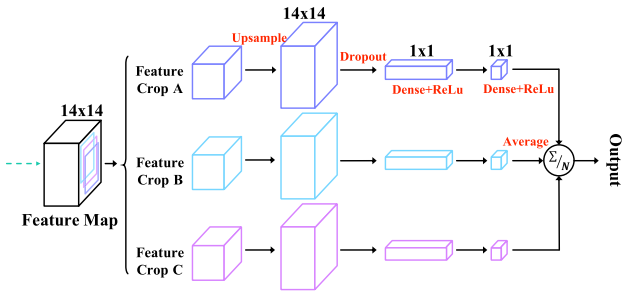


FIGURE 5. Example of our proposed feature cropping branch with three feature crops. Global branch and feature cropping branch share feature maps.

D. LOSS FUNCTION

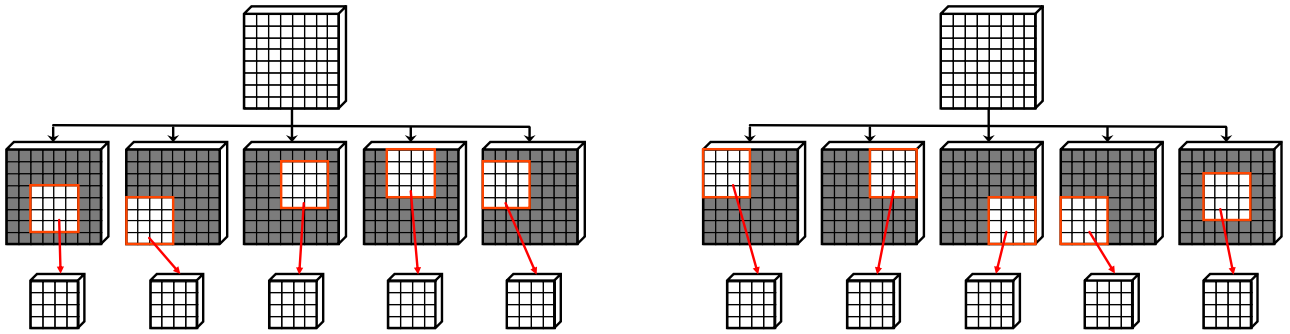
The loss function that is optimized during the training of classification CNN models is the cross-entropy loss [52]. Firstly, the cross-entropy loss for one class is given by:

$$£_{c_n} = -y_{c_n} log\widehat{y_{c_n}} \quad (3)$$

where n is the class label assumed to be in the range {1,2, . . ., N}, c_n is the n-th label class, $\widehat{y_{c_n}}$ is the n-th label prediction value of the final softmax layer of convolutional neural network.

Because of the particularity of medical diagnosis, the consequences of malignant judgment as benign in the nodule determination are intensely serious, the patient will likely miss the optimal treatment opportunity and cause irreversible influence. Nevertheless, if the benign is judged as malignant, the doctor will make a second judgment and correct the error, so the influence is relatively acceptable.

However, it may be desirable to weight the malignant nodule image higher in the training. One simple way of achieving this is by weighting the loss function for a frame higher if the label for the frame belongs to malignant. In reference [53], Panchapagesan *et al.* defined a weight vector $\vec{\omega}$



(a) Random Feature Cropping (b) Boundary Feature Cropping

FIGURE 6. Examples of random feature cropping (RFC) while $\alpha = 0.5$. Every graph represents one batch feature maps. White region shows the position of feature cropping region. Examples of boundary feature cropping (BFC) while $\alpha = 0.5$, $S_w = w/2$ and $S_h = h/2$. White region shows the position of feature cropping region.

with elements $\omega_{c_n} > 0$ defined over the range of class labels $n \in \{1, 2, ..., N\}$ for class-weighted cross-entropy. We also define another weighted cross-entropy as follows:

$$\pounds_{c_n} = -\omega_{c_n} y_{c_n} log\widehat{y_{c_n}} \tag{4}$$

Therefore, the loss function of multiple classification network is defined as

$$\pounds = \sum_{n=1}^{N} \pounds_{c_n} = -\sum_{n=1}^{N} \omega_{c_n} y_{c_n} log\widehat{y_{c_n}} \tag{5}$$

Similarly, the loss function of the binary-class network is defined as

$$\pounds = \sum_{n=1}^{2} \pounds_{c_n} = -(\omega_{c_1}(y_{c_1} log\widehat{y_{c_1}}) + \omega_{c_2}(y_{c_2} log\widehat{y_{c_2}})) \tag{6}$$

However, for this special case of binary classification, we define binary-class weighted cross-entropy as

$$\pounds = -(\omega_m(y_m log\widehat{y_m}) + \omega_b(y_b log\widehat{y_b})) \tag{7}$$

where m is the class label 'malignant' and b is the class label 'benign', ω_m and ω_b are the weight for malignant and benign, y_m and y_b are the label assumed to be malignant and benign, $\widehat{y_m}$ and $\widehat{y_b}$ are the malignant and benign label prediction value of the final softmax layer of convolutional neural network.

For one training batch, we define batch loss function as

$$\pounds_{batch} = -\sum_{batch} (\omega_m(y_m log\widehat{y_m}) + \omega_b(y_b log\widehat{y_b})) \tag{8}$$

IV. EXPERIMENT AND ANALYSIS

A. DATASET

The thyroid nodule ultrasound images used in our research work were from the following two databases:

1) PUBLIC DATASET

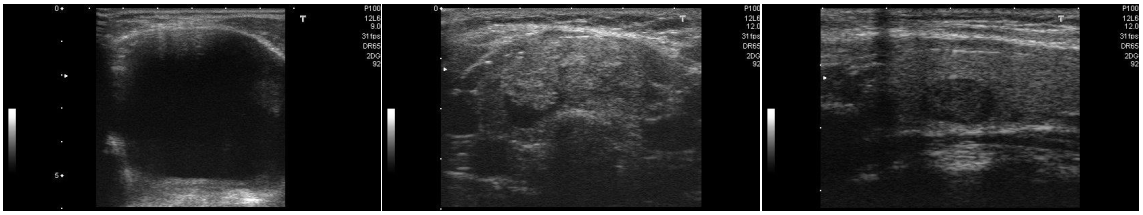
The Digital Database of Thyroid Ultrasound Images (DDTI) [54] are adopted in the experiments. DDTI consists of a set of B-mode Ultrasound images, including a complete annotation and diagnostic description of suspicious thyroid lesions by at least two expert radiologists. Currently, DDTI contains 299 cases (270 women and 29 men) with 347 images. All of the cases with relevant thyroid disorders were collected from the IDIME Ultrasound Department, one of the largest diagnostic imaging centers in Colombia. Ultrasound images were extracted from thyroid ultrasound video sequences captured by Ultrasound devices (TOSHIBA Nemio 30/TOSHIBA Nemio MX). Ultrasound images were saved in an uncompressed JPEG format. The nodule polygon and annotation information were saved in an alone XML file per one patient. The patients were classified by the experts using the thyroid imaging, reporting and data system (TI-RADS) [10]. TI-RADS give points for all features of five ultrasound categories in a nodule, with more suspicious features being awarded additional points. TI-RADS scores sums feature points of all the categories, which range from TR1 (benign) to TR5 (high suspicion of malignancy). These TI-RADS descriptions were also included in the XML file. Some samples of DDTI are illustrated in Fig. 7.

2) LOCAL DATASET

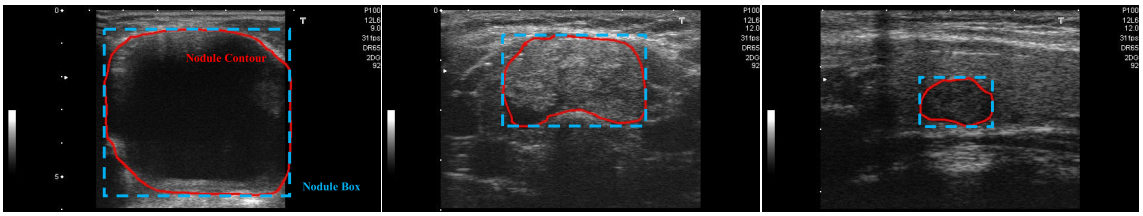
Another database is a local database, consisting of 377 thyroid ultrasound images which images are with the sizes 1024×695. 269 images in the database are labeled as malignant, and 108 cases are labeled as benign. This database is collected by Ultrasonic Medical Center of The First People's Hospital of Chenzhou. The nodule polygon and annotation information were also saved in an alone XML files per one patient.

3) DATA PREPROCESS

We divide all thyroid nodule ultrasound images into training and testing sets in a 3:1 ratio. The training set contains 330 benign cases and 213 malignant cases while the testing set contains 110 benign cases and 71 malignant cases. Accurate contours of all nodules are provided in XML format, so firstly we extracted the region of interest (RoI) of each nodule by doctors' demarcation, which region is marked as a blue box in Fig. 7. Given that some images were grayscale



(a) Thyroid Ultrasound Image In DDTI



(b) Thyroid Nodule Contour In DDTI

FIGURE 7. Some samples of the thyroid ultrasound images in the public dataset. In (b), we illustrate that both nodule contour (red curve) and nodule box (blue rectangle) from the nodule polygon and annotation information in XML files.

TABLE 2. Basic metrics description.

Description for Metrics		Predict Result	
		Malignant	Benign
Ground Truth	Malignant	**TP**	**FN**
	Benign	**FP**	**TN**

while others were 3-channel, we convert all the images to RGB mode, and then resize to the same scale 224 × 224.

B. EXPERIMENTAL IMPLEMENTATION DETAILS

1) METRICS

For the performance evaluation in the experiment. We choose the different networks and test the performance of the model with different networks. We use the accuracy and F-measure as the classification performance metrics to examine different network. First, we define malignant nodule as positive sample. So we denote TP as true positive (the number of instances correctly predicted as malignant), FP as false positive (the number of instances incorrectly predicted as malignant), TN as true negative (the number of instances correctly predicted as benign) and FN as false negative (the number of instances incorrectly predicted as benign), which is shown in Fig. 2. Then, we can calculate four metrics: accuracy, precision, recall and F-Measure as follows:

1) *Accuracy*:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{9}$$

2) *Precison*:

$$Precison = \frac{TP}{TP + FP}. \tag{10}$$

3) *Recall*:

$$Recall = \frac{TP}{TP + FN}. \tag{11}$$

TABLE 3. Comparison of accuracy of different α.

α	0.5	0.6	0.7	0.8	0.9	1
Accuracy (%)	92.27	95.03	94.48	93.37	93.37	89.50
F1-Measure (%)	90.54	93.88	93.24	91.89	91.89	87.07

4) *F-Measure*:

$$F - Measure = \frac{(1 + \beta^2) * Precision * Recall}{\beta^2 * (Precision + Recall)}. \tag{12}$$

β in (12) is a coefficient that adjusts the proportion of precision and recall and is usually set to 1 [55], which is defined as *F1-Measure*.

2) LEARNING CONFIGURATION

We implement the proposed network on Tensorflow [21] framework. We train our models using stochastic gradient descent. We use an initial learning rate equal to 10^{-5}, the momentum of 0.9, weight decay of 0.0005, and compute the batch normalization value with a batch size of 16. And we set malignant weighted factor ω_m to 3, benign weighted factor ω_b to 2 for loss function (8), branch weight ω to 0.5 in (1). All training images are resized to a maximum size of 224 × 224.

3) CHOOSING THE CROPPING RATIO α IN FEATURE CROPPING BRANCH

Under the condition that malignant class recall must be as close to 100% as possible, we will increase accuracy as much as possible, and select the optimal α values as factor of feature cropping branches. Based on boundary feature cropping method, we tried $\alpha = 0.5, 0.6, 0.7, 0.8, 0.9$ and 1 (1 means only global branch actually), respectively, and calculated the accuracy that a single branch can achieve when the factor α is set to 0.6, as shown in Table 3. As α is set to 0.6, the accuracy and F1-Measure single feature cropping branch can reach the best result 95.03% and 93.88%.

TABLE 4. Network performance comparison.

Architecture	*Feature Extraction Network*	*Accuracy*	*Precision*	*Recall*	*F1-Measure*
Global Branch	Inception_v3	88.95%	82.28%	91.55%	86.67%
	VGG19	88.40%	82.05%	90.14%	85.91%
	Resnet50	88.95%	84.00%	88.73%	86.30%
	Xception	88.95%	83.12%	90.14%	86.49%
	Our proposed CNN	**89.50%**	**84.21%**	**90.14%**	**87.07%**
Global Branch + RFC Branch	Inception_v3	94.48%	90.67%	95.77%	93.15%
	VGG19	92.82%	88.16%	94.37%	91.16%
	Resnet50	92.27%	85.19%	97.18%	90.79%
	Xception	94.48%	90.67%	95.77%	93.15%
	Our proposed CNN	**95.58%**	**92.00%**	**97.18%**	**94.52%**
Global Branch + BFC Branch	Inception_v3	94.48%	89.61%	97.18%	93.24%
	VGG19	93.37%	88.31%	95.77%	91.89%
	Resnet50	94.48%	90.67%	95.77%	93.15%
	Xception	95.03%	90.79%	97.18%	93.88%
	Our proposed CNN	**96.13%**	**93.24%**	**97.18%**	**95.17%**

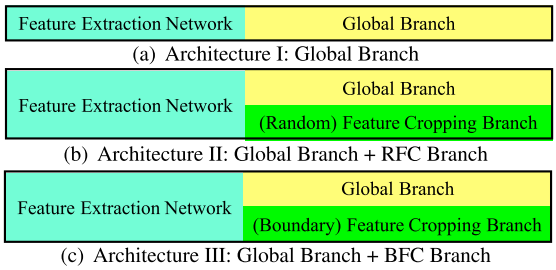


FIGURE 8. Three architectures of our comparison experiment. In every architecture, feature extraction network will use the backbone network from inception_v3 [56], VGG19 [25], ResNet50 [27], Xception [57] and our proposed network.

C. NETWORK PERFORMANCE COMPARISON

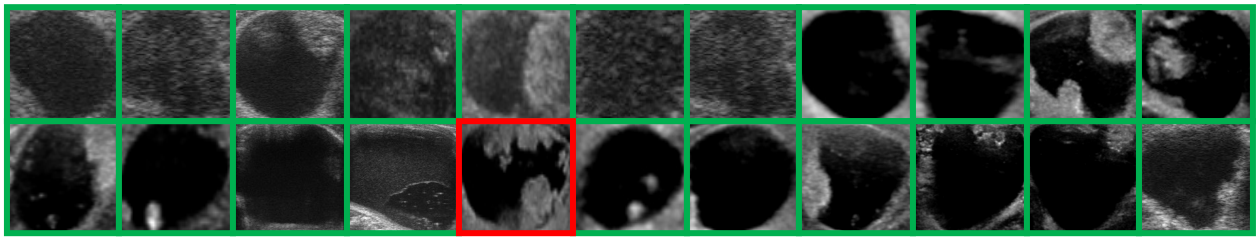
In this section, with the cropping ratio parameter α equal to 0.6 and the size of cropping maps N in (2) equal to 5, we compare the classification performance and calculated the above four indicators with different backbone networks, i.e., inception_v3 [56], VGG19 [25], ResNet50 [27], Xception [57] and our proposed network. Because our proposed network is based on a hybrid CNNs with two independent branches and feature cropping branch can adopt two different methods respectively, we performed a comparison of three architectures in different combinations. Three architectures are illustrated in Fig. 8.

1) **Global Branch**: Firstly, we only use the structure of feature extraction network and global branch, as introduced in Section III-B.
2) **Global Branch+RFC Branch**: Outside of feature extraction network, we will use proposed hybrid network of both global branch and feature cropping branch, and feature cropping branch uses the random feature cropping method mentioned in Section III-C.
3) **Global Branch+BFC Branch**: Correspondingly, feature cropping branch with boundary feature cropping method will be used in this architecture.

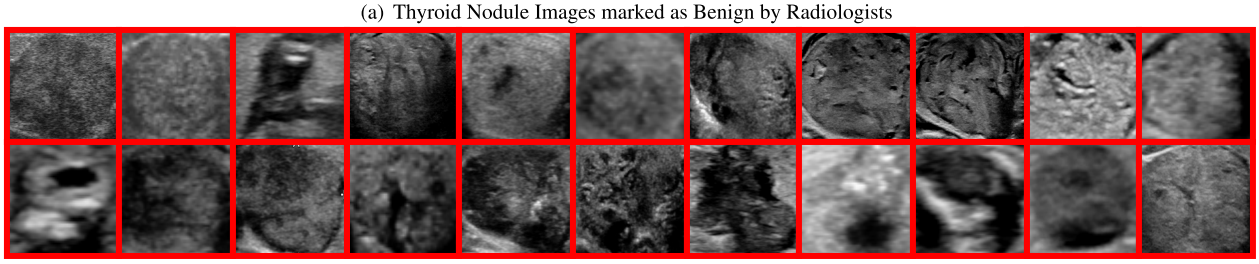
In addition, feature extraction network can be replaced by different networks, and the feature extraction section is all the networks before the last fully connected layer in each network. With our proposed network, we also compared five networks as feature extract network, while adding different feature cropping branches. The results are given in Table 4 for the two datasets.

From each architecture in Table 4, we can find that the hybrid network we designed has achieved the better performance in all four indicators than other deep network based on same architecture. Meanwhile, the number of layers is less, reducing the number of nodes simplifying the complexity of the model, and saving the computing time. This result indicates that deep and complex networks may not be suitable for thyroid nodule classification task. Perhaps more powerful network is more likely to overfit, which isn't good for processing two small datasets from different ultrasound devices.

Comparing the two different methods random feature cropping and boundary feature cropping from Table 4, we also find that most networks using boundary feature cropping method can achieve better performance. Our proposed hybrid CNN with boundary feature cropping method achieves the best performance in terms of the accuracy, precision, and F1-Measure. Learning from many medical articles [10], [58], [59], the guideline of thyroid nodule diagnosis using ultrasound image is mostly related to the edge contour and central region (such as shape, margin, composition), which is also consistent with our experimental results. Since our goal is to obtain a set of feature vectors that can best distinguish between benign and malignant nodules, vectors obtained from marginal regions have a positive impact.



(a) Thyroid Nodule Images marked as Benign by Radiologists



(b) Thyroid Nodule Images marked as Malignant by Radiologists

FIGURE 9. Classification samples by our proposed hybrid CNN with boundary feature cropping branch. The images in the upper part are marked by the radiologists as benign nodules, and another part is a set of malignant thyroid nodules. In these benign-labeled images, there is one image (red border) is classified to be malignant by network. And all of malignant-labeled images have the same predicted results with radiologists.

D. CASE STUDY

We select 44 thyroid nodule ultrasound images, randomly 22 benign images and 22 malignant images based on labeled by radiologists, from two datasets in this paper and depict the retrieval results in Fig. 9.

In the upper panel of Fig. 9(a), we assessed what the network has predicted on 22 test images labeled benign. Most of the classification results appear reasonable. We can find that only one benign test image (red border) has been identified as malignant. In this thyroid nodule ultrasound image, the irregularity of the cystic tissue may prevent the image from being properly predicted. Other images are correctly recognized by the network.

In the below panel of Fig. 9(b), we also visualize 22 test images which have been labeled malignant by radiologists. The predicted results of these images are consistent with the doctor's annotations. This result is also consistent with the original intention of experimental network parameter selection. Since the consequences of malignant judgment as benign in the nodule determination are intensely serious, it is very likely that the patient will miss the optimal treatment opportunity and cause irreversible influence. Nevertheless, if the benign is judged as malignant, the doctor will definitely make a second judgment and correct the error, so the influence is relatively acceptable.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose hybrid feature cropping network with a multi-branch structure to extract features and classify the ultrasound images of thyroid nodule. The global branch is used to get global features. In the other branch, we add feature cropping method designed to reduce the negative effects of local similarities between benign and malignant nodules. At the same time, due to the different sizes of the two branch cropping frames, feature vectors at different scales are obtained. The final result proves it is superior to the existing mainstream networks.

The next step will be improving the method and collecting more thyroid images at the same time. More specifically, we will introduce new network modules, such as attention network, to improve the learning ability of the network. In addition, we will try weakly supervised learning or unsupervised learning to solve the bottleneck of labeled data. How to reduce the bottleneck of limited annotated data, as suggested in reference [34], is still a key issue.

ACKNOWLEDGMENT

This work was conducted on the platform of the Center for Data Science of Beijing University of Posts and Telecommunications.

REFERENCES

[1] W. Chen, K. Sun, R. Zheng, H. Zeng, S. Zhang, C. Xia, Z. Yang, H. Li, X. Zou, and J. He, "Cancer incidence and mortality in China, 2014," *Chin. J. Cancer Res.*, vol. 30, no. 1, p. 1, 2018.

[2] D. Gaitini, R. M. Evans, and G. Ivanac, "Thyroid ultrasound," in *EFSUMB Course Book on Ultrasound*. London, U.K.: EFSUMB, 2011, ch. 16.

[3] J. Y. Kwak, K. H. Han, J. H. Yoon, H. J. Moon, E. J. Son, S. H. Park, H. K. Jung, J. S. Choi, B. M. Kim, and E.-K. Kim, "Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk," *Radiology*, vol. 260, no. 3, pp. 892–899, Sep. 2011.

[4] U. R. Acharya, F. Molinari, R. Garberoglio, A. Witkowska, and J. S. Suri, "Automated benign & malignant thyroid lesion characterization and classification in 3D contrast-enhanced ultrasound," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2012, pp. 452–455.

[5] B. Gopinath and N. Shanthi, "Support vector machine based diagnostic system for thyroid cancer using statistical texture features," *Asian Pacific J. Cancer Prevention*, vol. 14, no. 1, pp. 97–102, Jan. 2013.

[6] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian, "Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network," *J. Digit. Imag.*, vol. 30, no. 4, pp. 477–486, Aug. 2017.

[7] W. Song, S. Li, J. Liu, H. Qin, B. Zhang, S. Zhang, and A. Hao, "Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1215–1224, May 2019.

[8] M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis, "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 519–527, Jul. 2009.

[9] E. N. K. Kollorz, D. A. Hahn, R. Linke, T. W. Goecke, J. Hornegger, and T. Kuwert, "Quantification of thyroid volume using 3-D ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 457–466, Apr. 2008.

[10] F. N. Tessler, W. D. Middleton, E. G. Grant, J. K. Hoang, L. L. Berland, S. A. Teefey, J. J. Cronan, M. D. Beland, T. S. Desser, M. C. Frates, L. W. Hammers, U. M. Hamper, J. E. Langer, C. C. Reading, L. M. Scoutt, and A. T. Stavros, "ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee," *J. Amer. College Radiol.*, vol. 14, no. 5, pp. 587–595, May 2017.

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

[12] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1292–1300, Oct. 2004.

[13] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," *Med. Phys.*, vol. 29, no. 2, pp. 157–164, Jan. 2002.

[14] J. Virmani, V. Kumar, N. Kalra, and N. Khandelwal, "SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors," *J. Digit. Imag.*, vol. 26, no. 3, pp. 530–543, Jun. 2013.

[15] X. Wang, V. Seetohul, R. Chen, Z. Zhang, M. Qian, Z. Shi, G. Yang, P. Mu, C. Wang, Z. Huang, Q. Zhou, H. Zheng, S. Cochran, and W. Qiu, "Development of a mechanical scanning device with high-frequency ultrasound transducer for ultrasonic capsule endoscopy," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1922–1929, Sep. 2017.

[16] W. Qiu, X. Wang, Y. Chen, Q. Fu, M. Su, L. Zhang, J. Xia, J. Dai, Y. Zhang, and H. Zheng, "Modulated excitation imaging system for intravascular ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1935–1942, Aug. 2017.

[17] G.-Q. Zhou, W.-W. Jiang, K.-L. Lai, and Y.-P. Zheng, "Automatic measurement of spine curvature on 3-D ultrasound volume projection image with phase features," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1250–1262, Jun. 2017.

[18] G.-Q. Zhou and Y.-P. Zheng, "Automatic fascicle length estimation on muscle ultrasound images with an orientation-sensitive segmentation," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 12, pp. 2828–2836, Dec. 2015.

[19] G.-Q. Zhou, P. Chan, and Y.-P. Zheng, "Automatic measurement of pennation angle and fascicle length of gastrocnemius muscles using real-time ultrasound imaging," *Ultrasonics*, vol. 57, pp. 72–83, Mar. 2015.

[20] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2014, pp. 675–678.

[21] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.

[22] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. Autodiff Workshop Future Gradient-Based Mach. Learn. Softw. Techn. (NIPSW)*, Dec. 2017. [Online]. Available: https://openreview.net/forum?id=BJJsrmfCZ

[23] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[28] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.

[29] W. Wang, S. Zhao, J. Shen, S. C. H. Hoi, and A. Borji, "Salient object detection with pyramid attention and salient edges," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1448–1457.

[30] W. Wang, W. Wang, Y. Xu, J. Shen, and S.-C. Zhu, "Attentive fashion grammar network for fashion landmark detection and clothing category classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4271–4280.

[31] W. Wang, J. Shen, X. Dong, A. Borji, and R. Yang, "Inferring salient objects from human fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 18, 2019, doi: 10.1109/TPAMI.2019.2905607.

[32] Q. Lai, W. Wang, H. Sun, and J. Shen, "Video saliency prediction using spatiotemporal residual attentive networks," *IEEE Trans. Image Process.*, vol. 29, pp. 1113–1126, 2020, doi: 10.1109/TIP.2019.2936112.

[33] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 38–49, Jan. 2018.

[34] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[35] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: http://arxiv.org/abs/1712.04621

[36] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[37] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[38] X. Feng, J. Yang, A. F. Laine, and E. D. Angelini, "Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 568–576.

[39] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[41] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," *IEEE Access*, vol. 5, pp. 5804–5810, 2017.

[42] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 198, Dec. 2019.

[43] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," 2013, *arXiv:1312.5402*. [Online]. Available: http://arxiv.org/abs/1312.5402

[44] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1531–1544, Jul. 2019.

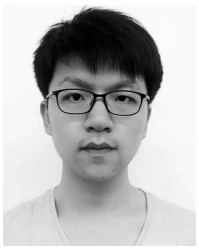
[45] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[47] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[48] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[50] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[51] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch DropBlock network for person re-identification and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3691–3701.
[52] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[53] S. Panchapagesan, M. Sun, A. Khare, S. Matsoukas, A. Mandal, B. Hoffmeister, and S. Vitaladevuni, "Multi-task learning and weighted cross-entropy for DNN-based keyword spotting," in *Proc. Interspeech*, Sep. 2016, pp. 760–764.
[54] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, "An open access thyroid ultrasound image database," *Proc. SPIE*, vol. 9287, Jan. 2015, Art. no. 92870W.
[55] Y. Qiao, Y. Wu, F. Duo, W. Lin, and J. Yang, "Siamese neural networks for user identity linkage through Web browsing," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 13, 2019, doi: 10.1109/TNNLS.2019.2929575.
[56] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[57] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[58] J. F. Sánchez, "TI-RADS classification of thyroid nodules based on a score modified according to ultrasound criteria for malignancy," *Rev. Argentine Radiol.*, vol. 78, no. 3, pp. 138–148, 2014.
[59] G. Russ, "Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: Description and reflections," *Ultrasonography*, vol. 35, no. 1, pp. 25–38, Jan. 2016.

RUONING SONG (Student Member, IEEE) is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China. His current research interests focus on image processing, deep learning, and cloud computing.



LONG ZHANG received the B.E. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. His research interests include artificial intelligence, image processing, and machine learning algorithm.

CHUANG ZHU (Member, IEEE) received the Ph.D. degree in microelectronics from Peking University, Beijing, China.

He is currently a Lecturer with the Beijing University of Posts and Telecommunications (BUPT), Beijing, where he also leads the Image Processing Group. He is also a member of the Center for Data Science, BUPT. Before that, he was a Postdoctoral Research Fellow with the School of Electronics Engineering and Computer Science, Peking University, from 2015 to 2017. His research interests include deep learning, image processing, multimedia content analysis, and machine learning algorithm optimization. He has published more than 30 publications in international magazines and conferences in these areas, including the IEEE TRANSACTIONS ON MULTIMEDIA, journal of *Signal Processing: Image Communication*, and the IEEE International Conference on Multimedia and Expo (ICME). He has served on the Technical Program Committee of the International Conference on Computing and Pattern Recognition (ICCPR 2019).

JUN LIU (Member, IEEE) received the B.E. and Ph.D. degrees from the Department of Information Engineering, Beijing University of Posts and Telecommunications (BUPT), in 1998 and 2003, respectively. He is currently an Associate Professor and the Director of the Center for Data Science. His research interests include deep learning, big data analysis, and stream data algorithms.

JIE YANG (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Beijing University of Posts and Telecommunications, China, in 1993, 1999, and 2007, respectively. She is currently a Professor and the Deputy Dean of the School of Information and Communication Engineering, BUPT. Her current research interests include broadband network traffic monitoring, user behavior analysis, and big data analysis in Internet and Telecom. She has published several articles on international magazines and conferences, including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS.

TONG ZHANG is currently a Chief Physician and the Deputy Director of the Ultrasonic Medical Center of Chenzhou No.1 People's Hospital. He has been engaged in ultrasound medicine for more than 20 years. He has been engaged in prenatal ultrasound diagnosis for more than ten years, and has a profound understanding of fetal medicine. In the past five years, his main focus has been on the diagnosis and ablation treatment of thyroid tumors. He is also a member of Thyroid Interventions Committee of Intervened Physicians Branch Committee of Chinese Medical Doctor Association.