Journal of Imaging Science and Technology® 66(4): 040409-1–040409-10, 2022.

Hierarchical Deep Learning Networks for Classification of Ultrasonic Thyroid Nodules

Bo Wang
Computer Engineering Technical College (Artificial Intelligence College), Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China
E-mail: hust_wb@hrbust.edu.cn

Fengqiang Yuan, Zhiwei Lv, and Ying He
School of Automation, Harbin University of Science and Technology, Harbin 150080, China

Zongren Chen
Computer Engineering Technical College (Artificial Intelligence College), Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China
Faculty of Information Technology, Macau University of Science and Technology, Macau 999078, China

Jianhua Hu, Jun Yu, Shuzhao Zheng, and Hai Liu
Computer Engineering Technical College (Artificial Intelligence College), Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China

Abstract. *Thyroid nodules classification in ultrasound images is actively researched in the field of medical image processing. However, due to the low quality of ultrasound images, severe speckle noise, the complexity and diversity of nodules, etc., the classification and diagnosis of thyroid nodules are extremely challenging. At present, deep learning has been widely used in the field of medical image processing, and has achieved good results. However, there are still many problems to be solved. To address these issues, we propose a mask-guided hierarchical deep learning (MHDL) framework for the thyroid nodules classification. Specifically, we first develop a Mask RCNN network to locate thyroid nodules as the region of interest (ROI) for each image, to remove confounding information from input ultrasound images and extract texture, shape and radiology features as the low dimensional features. We then design a residual attention network to extract depth feature map of ROI, and combine the above low dimensional features to form a mixed feature space via dimension alignment technology. Finally, we present an AttentionDrop-based convolutional neural network to implement the classification of benign and malignant thyroid nodules in the mixed feature space. The experimental results show that our proposed method can obtain accurate nodule classification results, and hierarchical deep learning network can further improve the classification performance, which has immense clinical application value.* © *2022 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2022.66.4.040409]

1. INTRODUCTION

Thyroid nodule is a common endocrine system frequently-occurring disease. Early detection and effective classification of benign and malignant are beneficial to reduce the incidence of thyroid cancer. Ultrasound imaging technology is real-time, safe, inexpensive and non-invasive, which can not only detect thyroid nodules, but also quantitatively evaluate the malignant risk of nodules, and has become the most common method for the diagnosis of thyroid diseases [1]. However, in clinical practice, there are subjective factors such as doctors' subjective operating habits and experience-dependent judgment bias in both ultrasonic examination and feature discrimination. Therefore, it is of great clinical significance to improve the diagnoses and treatment of thyroid nodules by in-depth understanding of ultrasonic image characteristics and automatic classification of benign and malignant thyroid nodules.

Most studies mainly focus on the design of various features, such as morphological features and texture features, etc., but these feature methods require accurate manual labeling information to obtain the contour of nodules, and are not applicable to practical clinical applications [2]. With the wide application of machine learning in the medical field, researchers hope to help doctors improve the accuracy of clinical diagnosis by using learning-based methods to classify benign and malignant thyroid nodules. Classification methods based on traditional machine learning [3–9] first extract and select the features of various nodules in ultrasonic images, and then use specific machine learning methods to achieve classification. Although such methods extract more diverse imaging features and compensate for the deficiency of features to some extent, it is very difficult and complex to extract high-quality effective features, and most of them still rely on manual selection, with heavy workload and low efficiency. In recent years, deep learning has become a hot topic in the study of thyroid ultrasound image classification, especially the deep learning method based on

Received Jan. 17, 2022; accepted for publication Apr. 12, 2022; published online May 11, 2022. Associate Editor: Hamad Hamad Naeem.
1062-3701/2022/66(4)/040409/10/$25.00

convolutional neural Network (CNN) [10]. By introducing technologies such as local connection weight sharing pooling operation and nonlinear activation, this kind of method enables the network to automatically learn features from data, thus avoiding the constraints of traditional machine learning manual extraction of features and achieving good results [11–17].

To this end, we propose a mask-guided hierarchical deep learning (MHDL) network model for benign and malignant classification of ultrasonic thyroid nodules. This model divides the original classification problem into several sub-problems and uses different networks to solve each sub-problem in a hierarchical manner. Firstly, a Mask RCNN network [18] is developed as the first layer backbone network to extract the region of interest (ROI) of thyroid nodules from the original ultrasound image and obtain the corresponding low-dimensional features of the image. Then, an attention residual network [19] is designed as the second layer network to generate depth feature maps of thyroid nodules, and dimension alignment technology is used to fuse with low-dimensional features to form mixed feature space. Finally, an AttetionDrop-based convolutional neural network (ADCNN) is presented to achieve benign and malignant classification of thyroid nodules in the mixed feature space. The main contributions of this paper include four aspects:

(1) A mask-guided hierarchical deep learning (MHDL) network model is proposed to achieve benign and malignant classification of thyroid nodules by multi-dimensional feature fusion.
(2) The method based on mask guidance is developed to locate thyroid nodule ROI to reduce irrelevant information in ultrasound images and improve the accuracy and effectiveness of thyroid nodule feature extraction.
(3) Multi-dimensional feature fusion combines deep feature with low-dimensional feature, which makes semantic feature fully learned and details feature added, thus improving the accuracy of classification.
(4) The experimental results show that the hierarchical deep learning network proposed in this paper can obtain accurate results of nodular classification and further improve the classification performance.

This paper is organized as follows. Section 1 describes the background significance of the research and the relevant contributions of this paper. Section 2 reviews the related research work of ultrasonic thyroid classification. Section 3 describes our proposed network. Section 4 reports the experimental results and analysis of the model design. Section 5 concludes our work.

2. RELATED WORK

2.1 *Machine Learning-based Method*

In recent years, the methods based on machine learning and deep learning have become the research focus of ultrasonic thyroid lesion classification. Among them, typical traditional machine learning methods include K-nearest Neighbor Classifier (K-NN), Support Vector Machine (SVM), Random Forests (RF), logistic regression model and fuzzy classifier etc. The common classification methods involve two steps, i.e., (1) getting ROI, and (2) extracting predefined features from each ROI for diagnosis with a certain classifier [20]. Literature [3] realized the risk assessment of thyroid nodule in K-NN and SVM by using boundary features, but the feature information such as texture and image frame sequence was not extracted, which is not suitable for medical decision support of thyroid nodule recognition. Literature [4] proposed a thyroid ultrasonic mode characterization method, which realized semi-supervised classification of thyroid nodules by using texture features and SVM classifier, but the study did not provide quantitative criteria for tumor risk assessment. Literature [5] proposed a hierarchical SVM classifier model, which realized automatic classification, and the diagnostic accuracy reached more than 96%, but only the gray-level run-length in the horizontal direction was calculated, which lacked robustness. Literature [6] proposed a quantitative elastic map measurement method, which achieved good predictability and effectiveness through SVM, but required manual selection and mapping of the lesion region, lacking automaticity. In addition, literature [7] used three-dimensional color Doppler ultrasound quantitative parameters combined with two-dimensional ultrasonic thyroid nodule characteristics to establish a Logistic regression model to evaluate the benign and malignant nodules, which improved the accuracy of thyroid nature diagnosis. However, due to the small sample size, the results were biased to some extent. Literature [8] proposed an automatic system for tumor classification in 3D contrast-enhanced ultrasound data set, and realized automatic real-time classification of benign and malignant thyroid nodules by using texture features in fuzzy classifier, but there is still a lot of room for improvement in classification accuracy. In literature [9], dual-threshold binary decomposition method was used to extract texture features unrelated to the imaging direction, and patch-based benign and malignant classification of ultrasonic thyroid nodules was carried out by RF and SVM. It solves the limitation, that the computer aided system can only be compatible with static images of one plane, but it does not locate the ROI quickly and effectively, which makes the spot selection lack accuracy. In general, traditional machine learning methods can achieve automatic classification of thyroid nodules to a certain extent and achieve good results. However, most of the features selected by these methods focus on low-dimensional information of local details, while deep semantic features are not learned, and feature extraction still relies on a large number of manual prior intervention with great limitations.

2.2 *Deep Learning-based Method*

Different from traditional machine learning methods, deep learning-based methods allow the network to train learning features automatically and achieve end-to-end classification.

Literature [11] for the first time migrates the CNNs model in ImageNet [21] to ultrasonic image dataset for depth feature extraction of thyroid nodules, and proposes a classification method of thyroid nodules combining traditional low-level features and deep learning features. This method significantly improved the classification accuracy, but the study did not make further improvement on CNN, and the classification accuracy of malignant nodules was not ideal. Literature [12] used fine-tuned GoogleNet [22] model to extract thyroid ultrasonic image features, and achieved benign and malignant classification of nodules by a cost-sensitive random forest classifier, and achieved high accuracy in large databases. However, the ROI of thyroid nodules was specified by experts and lacked automaton. Literature [13] designed a CNNs framework guided by clinical knowledge to achieve semi-automatic detection and classification of thyroid nodules in ultrasound images, achieving better classification results than previous automatic methods, but the detection speed of nodules is slow and there are limitations in clinical application. Literature [15] constructed a knowledge-guided auxiliary classifier generative adjoint network (KAC-GAN) for medical image enhancement, which improved the classification performance of ultrasonic thyroid nodules with good generalization ability and robustness. However, the model can only generate images of 64 pixels in size, and while experiments show that this helps enhance medical data and improve classification results, the synthesis of high pixel images is a good extension. Literature [17] proposed a multi-branch structure mixed feature clip-out network for feature extraction and benign and malignant classification of thyroid nodule ultrasound images, which achieved better classification effect than the existing mainstream network, but data marking is still the bottleneck problem of network training. Literature [23] proposed a new computer aided diagnosis (CAD) system based on deep learning for automatic detection and classification of nodules in ultrasonic images, and designed a detection network based on multi-scale regions to learn pyramid features to detect nodules at different feature scales, significantly improving the accuracy of classification. However, the data set of this study was class-unbalanced, nodular edges were poorly labeled, and region-based proposed detection methods were slower than similar SSD methods. Literature [24] proposed a hybrid network consisting of global classification branches and feature clipping branches for feature extraction and classification of thyroid nodule ultrasound images. In recent years, the attention network [19, 25] has been widely used because it makes the network more focused on core information. Literature [26] proposed a cascade network, which uses detection and segmentation tasks to achieve excellent classification prediction, and designed a two-step attention network to use segmentation results to obtain better classification results. However, the accuracy of classification is greatly affected by segmentation. Literature [27] proposed a CAD ultrasonic diagnosis system for thyroid and breast nodules based on deep learning, and developed a multi-organ CAD system based on convolutional neural network.

Considering the existing studies, traditional methods can obtain the low-dimensional feature of the image with a small amount of calculation through targeted feature selection and finally get good results. And deep learning-based method has achieved better results in both classification accuracy and classification performance in the classification task of thyroid nodules than before by extracting deep features.

In this paper, a mask-guided hierarchical deep learning network model is proposed to realize benign and malignant classification of thyroid nodules by multi-dimensional feature fusion technology. This method uses network layering strategy to obtain different types of feature information. Multi-scale feature fusion combines low-dimensional features with deep features to ensure local details of images while fully learning deep semantic features, which can not only improve the accuracy of classification, but also further improve the classification performance of hierarchical deep learning network.

3. PROPOSED METHOD

3.1 *The Overall Architecture of Proposed Model*

In order to effectively classify benign and malignant thyroid nodules, a mask-guided hierarchical deep learning (MHDL) model is proposed in this paper. Figure 1 shows the network framework of the model. The whole model consists of three layers. The first layer uses Mask RCNN as the backbone network to achieve thyroid nodule ROI location and low-dimensional feature extraction. The second layer uses residual attention network to extract depth feature information. The third layer integrates low-dimensional features and depth features to form a mixed feature space, and an AttentionDrop based convolutional neural network (ADCNN) is used to realize benign and malignant classification of thyroid nodules in the mixed feature space.

3.2 *ROI Localization and Low-dimensional Feature Extraction*

In order to extract the characteristics of thyroid nodules effectively, we first detect and locate the ROI of nodules. Figure 2 presents the model framework for ROI localization of thyroid nodules. This model uses Mask RCNN as the backbone network. The main task is to locate the ROI of thyroid nodules from the original ultrasound image as the input Mask of the subsequent network, and at the same time extract the low-dimensional characteristic information such as shape and texture from the ROI.

The specific process of network learning is as follows. Firstly, the thyroid ultrasound image is preprocessed and the VGG16 network [28] is used to extract the image feature map, and the ROI of each point of the feature map is preset to obtain multiple candidate ROI. Then, the region proposal Network (RPN) is used to perform binary classification and boundary box regression for candidate ROI, and part of candidate ROI is filtered out to obtain the proposed

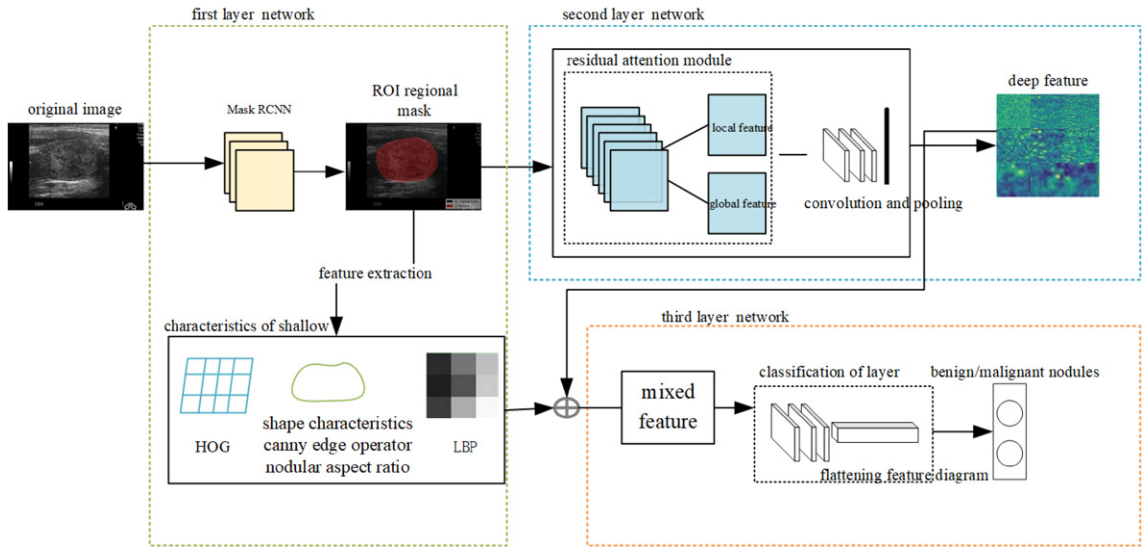

Figure 1. Illustration of mask-guided hierarchical deep learning (MHDL) framework for thyroid nodules classification. There are three components; (1) ROI generation via a Mask RCNN, (2) a residual attention module for deep feature extraction, and (3) benign and malignant classification module via an AttentionDrop-based convolutional neural network (ADCNN).

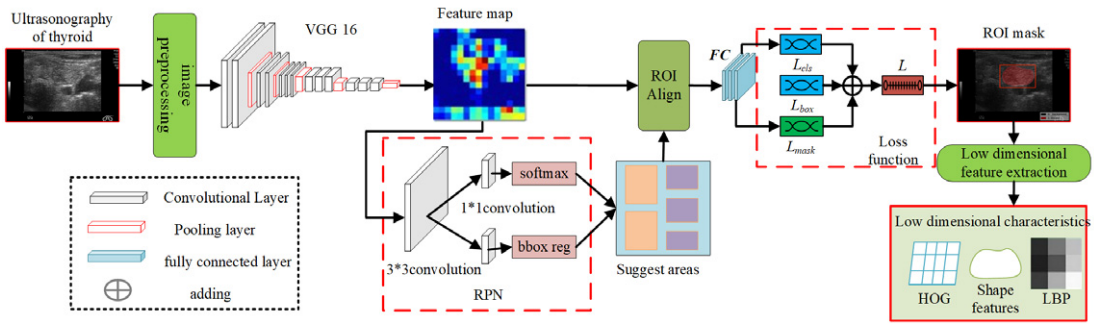

Figure 2. ROI localization framework for thyroid nodules based on Mask RCNN.

region. Meanwhile, RoIAlign operation is performed on the proposed region through the original feature map to obtain the fixed-size feature map. Finally, the nodule mask is generated through the full junction layer.

The total loss function L of each ROI in the network consists of the classification loss L_{cls}, the regression loss L_{box} and the mask loss L_{mask}, that is,

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}, \tag{1}$$

where L_{cls} uses two categories logarithmic losses via target and non-target, L_{box} controls the position of the detection frame by adjusting the offset to correct the position of the detection frame, and L_{mask} is used for segmentation error which is calculated by Sigmoid function via per pixel.

In addition, in order to obtain the local details of the image, a group of low-dimensional features are extracted from the nodule mask and expressed as feature group vectors. For feature map I, let the low-dimensional feature group vector at unit pixel point x in I be $L(x)$, then

$$L(x) = \{P(x), G(x), S(x), C(x), a\}, \tag{2}$$

where $P(x)$, $G(x)$, $S(x)$, $C(x)$, and a represent the local binary feature, gradient histogram feature, shape structure tensor, edge operator and nodule aspect ratio of pixel x, respectively.

3.3 *Depth Feature Extraction Based on Attention Residual Network*

In order to obtain the depth information of nodules, we design a learning network model based on attention residual [29], whose main purpose is to learn deeper features from nodules ROI and provide more effective depth semantic information for classification. Figure 3 shows the network framework of this model.

The network learning process contains three steps. Firstly, each nodule ROI mask goes through a 7×7 convolution and a 3×3 maximum pooling operation to

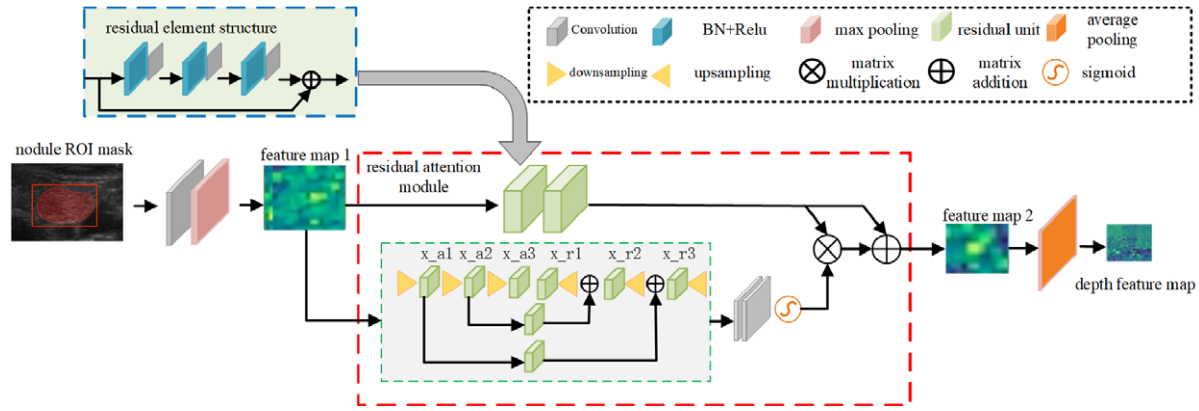

Figure 3. Depth feature extraction framework based on residual attention network.

generate the initial feature map. Then, the initial feature images are extracted by deep information learning through residual attention module. Finally, a 7×7 average pooling process is carried out on the secondary feature map to generate the depth feature map.

The attention residual module in the network consists of the main branch and the attention branch. The main branch is to process the initial features and contains two residual units. Each unit consists of three continuous convolution operations by the size of 1×1, 3×3 and 1×1, respectively. Additional, Batch Normalization (BN) and ReLu operation are conducted before each convolution. After that, the results of three convolution are added to the original input to obtain the main feature map. In the attention branch, three successive down-samples are taken to reduce the dimension of the initial feature map, and then three up-samples are taken by bilinear interpolation. After two 1×1 convolutions, the Sigmoid activation function is used to generate the attention feature map with normalized weight. Finally, the main feature map and the attention feature map are multiplied by elements and added to obtain the quadratic feature map.

Suppose that the input is x, the output characteristic diagrams of the main branch and attention branch are denoted as $F_{i,c}(x)$ and $M_{i,c}(x)$, respectively. Then the quadratic feature map $H_{i,c}(x)$ output by the attention residual module can be described as:

$$H_{i,c}(x) = \left(1 + M_{i,c}(x)\right) * F_{i,c}(x), \tag{3}$$

where i and c represent the number of space and channel, respectively. $M_{i,\,c}(x)$ has a value range of [0, 1] and is used as a selector for $F_{i,\,c}(x)$, thereby enhancing the effective features in $H_{i,c}(x)$ to suppress noise.

To this end, the depth feature $H(x)$ is obtained by average pooling of the quadratic feature map $H_{i,c}(x)$, that is:

$$H(x) = \text{AVGPooling}\left(H_{i,c}(x)\right), \tag{4}$$

where AVGPooling$(\cdot)$ is the global average pooling function.

3.4 *Nodule Classification Based on Mixed Feature Space*

3.4.1 *Mixed feature space*

The mixed feature space is a fusion of low-dimensional features and depth features. In this study, these two kinds of features are normalized, and the multi-scale features are obtained by using tensor concatenation method.

Suppose that the input is x, let $L_{bn}(x)$ and $H_{bn}(x)$ be the normalized low-dimensional feature and depth feature, respectively. Then the multi-scale feature $T(x)$ can be described as follows:

$$T(x) = TC\left(L_{bn}(x), H_{bn}(x)\right), \tag{5}$$

where $TC(\cdot)$ is the tensor concatenation function.

3.5 *Classification Network Training and Optimization*

The training and optimization of classification network includes loss function selection and regularization.

(1) *Loss function selection:* We adopt the cross entropy as the loss function for network training. Given a dataset containing N samples, its cross entropy L can be described as follows:

$$L = \frac{1}{N}\sum_{i}^{N} -\left[y_i \cdot \log\left(p_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - p_i\right)\right], \tag{6}$$

where y_i is the binary label of sample i, which defines positive sample as 1 and negative sample as 0. p_i is the probability when the sample i is positive.

(2) *Regularization:* To effectively prevent the over-fitting phenomenon of network training, we design a recursive regularization algorithm based on AttentionDrop [30], named AttentionDrop-based Regularization (ADR). Firstly, each pixel in the nth layer feature map is sorted according to the feature value, and the feature threshold is set to generate a shape mask for the pixels to cover the nodule region. Then, the shape mask is controlled by Bernoulli distribution to obtain the adaptive mask. Finally, the $(n+1)$th layer feature map is obtained by adjusting the activation value through the original feature map and the adaptive mask.

Given the nth layer feature map $F^{(n)}$, let S and M be the shape mask and adaptive mask, respectively. Then the $(n+1)$th layer feature map $F^{(n+1)}$ can be described as follows:

$$\gamma_{i,j} = \text{Bernoulli}(\alpha) \tag{7}$$

$$S_{i,j} = F_{i,j}^{(n)} \subseteq F_{\text{top\_}\beta}^{(n)} \tag{8}$$

$$M_{i,j} = \begin{cases} S_{i,j}, & \gamma_{i,j} = 1 \\ 1, & \gamma_{i,j} = 0 \end{cases} \tag{9}$$

$$F^{(n+1)} = \frac{F^n \odot M}{sum(M)}, \tag{10}$$

where (i, j) is the subscript of pixel points in the feature map, $\gamma_{i,j}$ is the sampling value of Bernoulli distribution subject to probability $\alpha \in (0, 1)$, $F_{\text{top\_}\beta}^{(n)}$ is the set of pixel points before $\beta\%$ of the feature value in $F^{(n)}$, $\odot$ is the point-by-point multiplication operation, and $1/\text{sum}(M)$ is the scaling parameter. The detailed steps are shown in Algorithm 1.

3.5.1 *Classification Algorithm*

Combining the feature space fusion and the training and optimization process of classification network, we design a hybrid feature space based classification algorithm for thyroid nodules (HFSCTN). Algorithm 2 gives the specific process of the algorithm.

4. EXPERIMENTS

4.1 *Dataset and Experimental Setup*

The open data set DDTI (Digital Database Thyroid Image) [31] provided by Columbia National University is used in this experiment. The dataset contains 480 ultrasound images from 400 patients with thyroid diseases, and 427 ultrasound images are remained after removing invalid images. Each image is classified by experts into 6 different grades according to TI-RADS standard [1], which are {2, 3, 4A, 4B, 4C, 5}, respectively. In order to classify benign and malignant nodules, class {2, 3} is defined as benign nodules and class {4A, 4B, 4C, 5} is defined as malignant nodules according to the classification principle of TI-RADS. The experimental data set is randomly divided into training set, validation set and test set in the ratio of 6:2:2.

The detailed information of the experimental platform are Ubuntu 18.04, Intel i7-9700K 8-core 8-thread processor, 32G memory, and NVIDIA GeForce GTX 1080 GPU. All experiments are carried out using Python 3.6 language and based on Keras deep learning framework.

Algorithm 1: AttentionDrop-based Regularization (ADR)

Input: The n^{th} layer feature map $F^{(n)}$.

Step 1: Initialization: Set the initial value of Bernoulli distribution parameter α and pixel feature threshold β;

Step 2: Perform max-min normalization on the n^{th} layer features:

$$F_{i,j}^{(n)} \leftarrow \frac{F_{i,j}^{(n)} - F_{\min}^{(n)}}{F_{\max}^{(n)} - F_{\min}^{(n)}}$$

Step 3: Sort the eigenvalues and select the first $\beta\%$ pixels:

$$S_{i,j} \leftarrow SortAndSelect(F_{i,j}^{(n)}, \beta)$$

Step 4: **For** each pixel in $F^{(n)}$ **do**

4.1: $\gamma_{i,j} \leftarrow Bernoulli(\alpha)$;

4.2: **If** $\gamma_{i,j} == 1$ **then**

4.3: $M_{i,j} \leftarrow S_{i,j}$;

4.4: **Else**

4.5: $M_{i,j} \leftarrow 1$;

4.6: **EndFor**

Step 5: Compute the scaled $F^{(n+1)}$:

$$F^{(n+1)} \leftarrow \frac{F^{(n)} \odot M}{sum(M)}$$

Output: The $(n+1)^{th}$ layer feature map $F^{(n+1)}$.

Algorithm 2: Hybrid feature space based classification algorithm for thyroid nodules (HFSCTN)

Input: Feature matrix D, low-dimensional feature $L_{bn}(x)$, and depth feature $H_{bn}(x)$.

Step 1: Initialization: Set the initial value of network weight w and offset θ;

Step 2: **For** each characteristic data X_i in D **do**

2.1: Perform multi-scale feature fusion according to Formula (5);

2.2: Activate with softmax function:

$$Y_i = softmax(w^T X_i + \theta);$$

2.3: Calculate the loss function according to Formula (6);

2.4: Forward propagation:

2.5: **For** each layer unit j **do**

2.6: $y_j = \Sigma w_{ij} x_i + \theta_j$;

2.7: Call the algorithm 1:

$$D^{(j+1)} = ADR(D^{(j)});$$

2.8: Calculation error:

$$Err_j = T_j - Y_j$$

where T_j is the label value, Y_j is the predicted value;

2.9: **EndFor**

2.10: Update the weight w and offset θ according to the error;

2.11: **EndFor**

Step 3: Repeat Step 2 until the loss remains unchanged or the maximum number of iterations;

Output: Classification result softmax(Y).

4.2 *Metrics*

In this experiment, the proposed method is compared with the residual attention network and the methods in [3, 4, 24, 26] and [27] in terms of classification accuracy. Accuracy rate (ACC), sensitivity (SENS), specificity (SPEC) and receiver operating characteristic (ROC) area under curve (AUC) are used to evaluate performance. ACC, SENS and SPEC are defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FR + FN} \tag{11}$$

$$SENS = \frac{TP}{TP + FN} \tag{12}$$

$$SPEC = \frac{TN}{TN + FP}, \tag{13}$$

where TP, TN, FP and FN represent the number of pixels of true positive, true negative, false positive and false negative in the classification results, respectively.

AUC is defined as the area under the ROC curve. The closer AUC value is to 1, the better the performance of the classification algorithm is. Wherein, the ROC curve is drawn with the true positive rate ($SENS$) as the ordinate and the false positive rate (1-$SPEC$) as the abscissa.

4.3 *Visualization Results and Analysis of Mask Extraction*

Figure 4 shows the localization and correction results of ultrasonic thyroid nodule mask in four groups. Fig. 4(a) is the original ultrasonic thyroid image. Fig. 4(b) is the gold standard of nodule contour annotated by experts, marked with red contour. Fig. 4(c) is the Mask obtained by Mask RCNN, marked with green rectangular box. Fig. 4(d) is the result after coordinate offset correction, marked with blue rectangular box.

From Fig. 4(c) and Fig. 4(d), we can see that the nodule region detected by traditional Mask RCNN is about 1/2 of the expert gold standard region, while the result of coordinate offset correction of RPN regression output in this study basically covers the entire nodule region. Traditional Mask RCNN requires a large amount of training data to obtain a relatively ideal result. Due to the limited data set in this experiment, the detection result is not ideal. Therefore, this study corrected the coordinate offset of the regression results, and the experiment proved that this operation can obtain a more accurate nodule mask and guide the classification network.

4.4 *Classification Results and Analysis of Different Methods*

Figure 5 shows the ROC curves and AUC results of two traditional classification methods on the data set of this study. Among them, Method 1 is proposed in reference [3], which uses shape features to achieve classification of thyroid nodules by K-NN and SVM. Method 2 is proposed in reference [4], which realize semi-supervised classification of thyroid nodules by SVM based on texture and shape features.

From Fig. 5, we can see that the AUC value of the residual attention network is only 86.94%, while the AUC value of

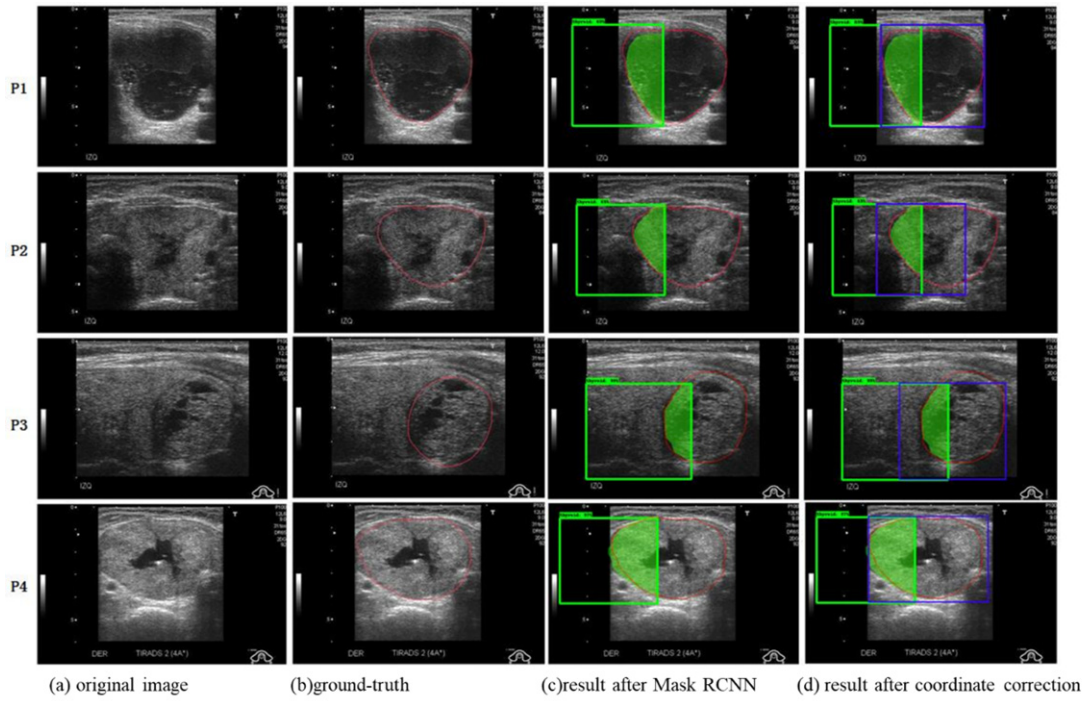

Figure 4. Localization and correction of thyroid nodule mask in ultrasound image.

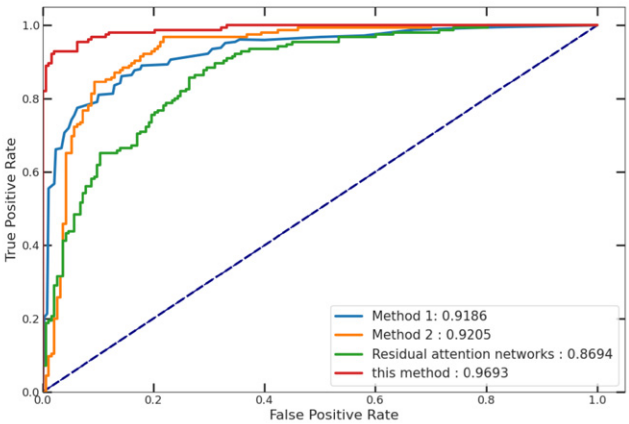

Figure 5. ROC curves of each classification methods.

Method 1 and Method 2 are 91.86% and 92.05% respectively. The AUC value of the proposed method is 96.93%, which is 11.40%, 5.52% and 5.30% higher than the previous three methods, respectively. Due to the low abstraction degree of depth feature and the lack of detail learning, the classification result of residual attention network is not ideal. Method 1 and Method 2 both selected a considerable number of low-dimensional feature combinations, which improved the fitting ability of the classifier to a certain extent, but did not abstract the higher-level image semantic features, so it did not achieve the best classification performance. However, the proposed method adopts the multi-dimensional hybrid feature that integrates deep semantic information with traditional image information, and gives consideration to both deep semantic feature and image detail feature to enhance the accuracy and robustness of the classification model. Experimental results show that our method obtain a better AUC value.

Table I gives the results of various classification evaluation indexes obtained by several different classification methods on the data set of this study. Method 1 and Method 2 are proposed in references [3] and [4], respectively.

From Table I, we can see that *SENS*, *SPEC* and *ACC* obtained by using shape features in Method 1 are 90.73%, 83.34% and 83.75%, respectively. In Method 2, the *SENS*, *SPEC* and *ACC* obtained by texture and shape features are 89.69%, 80.33% and 89.21%, respectively. *SENS*, *SPEC* and *ACC* are 87.24%, 83.19% and 85.06%, respectively. In our method, the *SENSSPEC* and *ACC* obtained by using multidimensional features that combine depth features and traditional features are 96.01%, 86.79% and 93.01%, respectively. The results show that the proposed method achieves better results in all evaluation indexes. Compared with the residual attention network, the proposed method improves the *SENS*, *SPEC* and *ACC* by 10.05%, 4.32% and 9.34%, respectively. This suggests that mixed-features-based attention residual learning can fully learn the classification target, and has a higher recognition rate of benign and malignant nodules, as well as higher sensitivity and specificity. Compared with Method 1 and Method 2, which use traditional image information as classification feature, the indexes of the proposed method are also greatly improved. This indicates that the mask can be used as guidance to pay more attention to the effective area of nodular classification. After the fusion of depth characteristic information, it not

Table I. The results of evaluation index for each classification methods.

Method	SENS	SPEC	ACC	Feature
Method 1 [3]	0.9073±0.0248	0.8334±0.0315	0.8375±0.0695	Shape feature
Method 2 [4]	0.8969±0.0386	0.8033±0.0234	0.8921±0.0386	texture + shape features
Residual attention	0.8724±0.0325	0.8319±0.0420	0.8506±0.0683	depth feature
Proposed method	0.9601±0.0181	0.8679±0.0425	0.9301±0.0597	multidimensional fusion feature

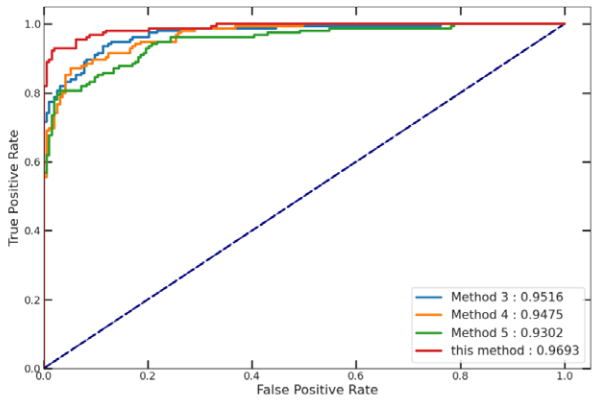

Figure 6. ROC curves of each classification methods.

Table II. Comparison of set metrics of different improved algorithms on ISBI dataset.

Method	SENS	SPEC	ACC
Method 3 [24]	0.9235±0.0278	0.8523±0.0285	0.9051±0.0658
Method 4 [26]	0.9187±0.0396	0.8478±0.0356	0.9014±0.0265
Method 5 [27]	0.9087±0.0316	0.8578±0.0240	0.8914±0.0159
Our method	0.9601±0.0181	0.8679±0.0425	0.9301±0.0597

only makes up for the limitation of low-dimensional characteristic information, but also deepens the learning of detail information, so that the classification performance is significantly improved.

Figure 6 shows the *ROC* curves and *AUC* results of three deep learning-based methods on the data set of this study. Method 3 is proposed in literature [24]; Method 4 is proposed in literature [26] and Method 5 is proposed in literature [27].

From Fig. 5, we can see that AUC values of Method 3, Method 4 and Method 5 are 95.16%, 94.75% and 93.02% respectively, while AUC value of our method is 96.93%, which are improved by 1.86%, 2.30% and 4.20% compared with the previous three methods, respectively. Experimental results show that the proposed method achieves better AUC values.

Table II gives the results of various classification evaluation indexes obtained by several different classification methods on the data set of this study.

From Table II, we can see that the SENS, SPEC and ACC obtained by Method 3 are 92.35%, 85.34% and 90.51%, respectively. The SENS, SPEC and ACC obtained by Method 4 are 91.87%, 84.78% and 90.14%, respectively. The SENS, SPEC and ACC obtained by Method 5 are 90.87%, 85.78% and 89.14%, respectively. The results show that the proposed method achieves better results in all evaluation indexes.

5. CONCLUSION

In this paper, a mask guided hierarchical deep learning network is proposed to classify thyroid nodules in ultrasound image. The model divides the original classification problem into several relatively easy sub-problems. Each sub-problem is solved with the strategy of network stratification, and finally realized the benign and malignant classification of thyroid nodules. In order to improve the accuracy and effectiveness of nodule feature extraction, the proposed network uses mask guidance to significantly reduce irrelevant information in ultrasound images by locating nodule ROI. Considering the impact of feature engineering on classification accuracy, this study integrates deep features and low-dimensional features to generate multi-dimensional features, and combines detail features to improve classification performance while fully learning semantic features. Finally, the experimental results show that the layering strategy of the proposed network not only achieves better classification results, but also greatly improves the classification performance.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Foundation of China, under Grant No. 61172167; the Foundation for Young Innovative Talents in Department of education of Guangdong province, under Grant No. 2019GKQNCX043; the Special projects in key fields of colleges and universities in Guangdong Province (new generation information technology), under Grant No. 2020ZDZX3094; the Special topic of school-level fund cultivation of Guangdong Polytechnic of Science and Technology, under Grant No. XJPY202005; Innovative Research Team in Universities of Guangdong Province of China, under Grant No. 2021KCXTD079; the Characteristic innovation project of Department of Education of Guangdong Province, under Grant No. 2019GKTSCX029.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

REFERENCES

[1] J. Zhou and W. Zhan, "Interpretation of the 2020 Chinese guidelines for ultrasound thyroid report and data system (C-TIRADS)," J. Diagnostics Concepts & Practice **19**, 350–353 (2020) (in Chinese).

[2] W. U. Kuan, Q. I. N. Pinle, C. H. A. I. Rui, and Z. E. N. G. Jianchao, "Benign and malignant diagnosis of thyroid nodules based on different ultrasound imaging," J. Computer Appl. **40**, 77–82 (2020) (in Chinese).

[3] M. Savelonas, D. Maroulis, and M. Sangriotis, "A computer-aided system for malignancy risk assessment of nodules in thyroid US images based on boundary features," Computer Methods and Programs Biomed. **96**, 25–32 (2009).

[4] D. K. Iakovidis, E. G. Keramidas, and D. Maroulis, "Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns," Artif. Intell. Med. **50**, 33–41 (2010).

[5] S. J. Chen, C. Y. Chang, K. Y. Chang, J. E. Tzeng, Y. T. Chen, C. W. Lin, W. C. Hsu, and C. K. Wei, "Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines," Ultrasound Med. Biol. **36**, 2018–2026 (2010).

[6] Ding J., Cheng H., Ning C., Huang J., and Zhang Y., "Quantitative measurement for thyroid cancer characterization based on elastography," J. Ultrasound Med. **30**, 1259–1266 (2011).

[7] C. Liu, Q. Mu, Y. Zhang, J. Gu, and T. Shi, "Logistic regression analysis of benign and malignant thyroid nodules by ultrasound features," Mod. Oncol. Med. **27**, 156–160 (2019) (in Chinese).

[8] U. R. Acharya, S. V. Sree, G. Swapna, S. Gupta, F. Molinari, R. Garberoglio, A. Witkowska, and J. S. Suri, "Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound," Proc. Inst. Mech. Eng. H **227**, 284–292 (2013).

[9] A. Prochazka, S. Gulati, S. Smutek, and D. Holinka, "Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition," Comput. Med. Imaging Graph. **71**, 9–18 (2019).

[10] M. Buda, B. Wildman-Tobriner, J. K. Hoang, D. Thayer, F.N. Tessler, W.D. Middleton, and M.A. Mazurowski, "Management of thyroid nodules seen on US images: Deep learning may match performance of radiologists," Radiology **292**, 181343 (2019).

[11] T. Liu, S. Xie, J. Yu, L. Niu, and W. Sun, "Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features," *Proc. Int'l. Conf. on Acoustics, Speech and Signal Processing* (IEEE, Piscataway, NJ, 2017), pp. 919–923.

[12] 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. Digital Imaging **30**, 477–486 (2017).

[13] X. Ying, Z. Yu, R. Yu, X. Li, M. Yu, M. Zhao, and K. Liu, "Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network," *Proc. 2018 Int'l. Conf. on Neural Information Processing (ICONIP 2018)* (Springer, Cham, LNCS, 2018), Vol. 11306, pp. 373–384.

[14] X. Liang, J. Yu, J. Liao, and Z. Chen, "Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging," Biomed. Res. Int. **2020**, 1–10 (2020).

[15] G. Shi, J. Wang, Y. Qiang, X. Yang, J. Zhao, R. Hao, W. Yang, Q. Du, and N.G.F. Kazihise, "Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification," Comput. Methods Programs Biomed. **196**, 1–14 (2020).

[16] X. Xinying, J. Junzhong, and Y. Yao, "Brain networks classification based on an adaptive multi-task convolutional neural networks," J. Comput. Res. Dev. **57**, 1449–1459 (2020).

[17] R. Song, L. Zhang, C. Zhu, J. Liu, J. Yang, and T. Zhang, "Thyroid nodule ultrasound image classification through hybrid feature cropping network," IEEE Access **8**, 64064–64074 (2020).

[18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proc. IEEE Int'l. Conf. on Computer Vision* (IEEE, Piscataway, NJ, 2017), pp. 2961–2969.

[19] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2017), pp. 6450–6458.

[20] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," Med. Image Anal. **43**, 157–168 (2018).

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural network," in *Proc. Advances in Neural Information Processing Systems*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates Inc., New York City, 2012), pp. 1097–1105.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proc. 2015 IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2015), pp. 1–9.

[23] T. Liu, Q. Guo, C. Lian, X. Ren, S. Liang, J. Yu, L. Niu, W. Sun, and D. Shen, "Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks," Med. Image Anal. **58**, 101555 (2019).

[24] R. Song, L. Zhang, C. Zhu, J. Liu, J. Yang, and T. Zhang, "Thyroid nodule ultrasound image classification through hybrid feature cropping network," IEEE Access **8**, 64064–64074 (2020).

[25] H. Guan, Y. Liu, E. Yang, P.T. Yap, D. Shen, and M. Liu, "Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification," Med. Image Anal. **71**, 102076 (2021).

[26] X. Shen, X. Ouyang, T. Liu, and D. Shen, "Cascaded networks for thyroid nodule diagnosis from ultrasound images," in *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data. MICCAI 2020*, edited by N. Shusharina, M. P. Heinrich, and R. Huang, Lecture Notes in Computer Science (Springer, Cham, 2021), vol. 12587.

[27] X. Liang, J. Yu, J. Liao, and Z. Chen, "Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging," BioMed. Res. Int. **2020** (2020).

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. 3rd Int'l. Conf. on Learning Representations (ICLR 2015)* (San Diego, 2015), pp. 1–14.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2016), pp. 770–778.

[30] Z. Ouyang, Y. Feng, Z. He, T. Hao, T. Dai, and S.T. Xia, "Attentiondrop for convolutional neural networks," *Proc. 2019 IEEE Int'l. Conf. on Multimedia and Expo* (IEEE, Piscataway, NJ, 2019), pp. 1342–1347.

[31] L. Pedraza, C. Vargas, F. Narvaez, O. Duran, E. Munoz, and E. Romero, "An open access thyroid ultrasound image database," Proc. SPIE **9287**, 92870W (2015).