A Method of Ultrasonic Image Recognition for Thyroid Papillary Carcinoma Based on Deep Convolution Neural Network

Yonghua Wang*, Wei Ke, Pin Wan

ABSTRACT

Thyroid cancer is a malignant tumor that occurs in the thyroid gland and is the most common malignant tumor in the endocrine system. Ultrasound examination is the most important method to diagnose thyroid cancer. The accuracy of ultrasound examination for thyroid cancer is closely related to doctors' cognition and understanding of ultrasound images, and there are subjective judgment and misjudgment. The ultrasound images of thyroid papillary carcinoma are mostly represented by two-dimensional gray scale, and with lower resolution, complicated internal tissue structure, and not obvious features of the cancer, it is difficult to distinguish and diagnose the thyroid papillary carcinoma. In this paper, we introduce the theory of convolution neural network (CNN) in view of the difficulty in recognizing the ultrasound image of thyroid papillary carcinoma, and propose a method which can automatically recognize the ultrasound image of thyroid papillary carcinoma. In terms of the need of ultrasonic image recognition of thyroid papillary carcinoma, the Fast Region-based Convolutional Network method (Faster RCNN) network is improved and normalized by connecting the fourth layer and the fifth layer of the shared convolution layer in the Faster RCNN network. Then, a multi-scale ultrasound image is used at the time of input. Finally, according to the main features of the ultrasound images of thyroid papillary carcinoma, they are classified so as to output detailed ultrasound image diagnosis reports. The experimental results show that compared with the original Faster RCNN network, the proposed method has higher recognition accuracy, shorter training time and higher efficiency in ultrasonic image recognition of thyroid papillary carcinoma.

Key Words: Thyroid papillary carcinoma, Ultrasound image, Convolutional neural network

DOI Number: 10.14704/nq.2018.16.5.1306 **NeuroQuantology 2018; 16(5):757-768**

Introduction

Thyroid carcinoma is the most common malignant tumor of endocrine system and its incidence is the first in head and neck malignant tumors. Thyroid papillary carcinoma is the most common, and its incidence accounts for about 85% of all thyroid cancer types (Davies *et al.*, 2006). Therefore, the diagnosis of thyroid papillary carcinoma is very important. The features of thyroid papillary carcinoma are as follows: unclear boundary, fibrous pseudomembrane, slight calcification of nodule or gravel inside the tubercle, non-uniform echo, etc. It is a challenging task to determine whether a lesion has these characteristics. The accuracy of ultrasound examination of thyroid cancer is closely related to the doctor's cognition and understanding of ultrasound images, there are often subjective judgments and misjudgments, and even experts with rich experience may misjudge. In addition, because of the different degree of experience and knowledge, the doctors of ultrasound images will have a certain degree of difference in their cognition of ultrasound images.

Corresponding author: Yonghua Wang
Address: School of Automation, Guangdong University of Technology, Guangzhou, 510006, China
e-mail ✉ sjzwyh@163.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 14 March 2018; **Accepted:** 13 April 2018

Therefore, it is necessary to be able to accurately judge and identify ultrasound images with thyroid cancer lesions. In addition, the traditional methods of recognizing ultrasound images with human eyes have been unable to keep up with the development of the times, mainly because it is difficult to accurately judge the lesions of complicated cancers in the ultrasound images, and it is easy to misjudge, which will bring serious consequences to the patients. Therefore, the ultrasonic medical image diagnosis can be realized by using the computer image recognition technology. By collecting a large number of medical image information of the same kind of diseases, carrying out the classification processing, and then combining the computer-aided diagnosis with the doctor's diagnosis, it's possible to quickly and accurately obtain the patients' diseases.

Convolution Neural Networks (CNN) (Hinton *et al.*, 2012) is one of the most rapidly developing fields of machine learning in recent years. As a kind of artificial neural network, it has become a hot topic in the field of speech analysis and image recognition. Convolution neural network is a multi-layer perceptron specially designed to recognize two-dimensional shapes. This network structure is highly invariant to translation, scaling, tilting or deformation of other forms (Hinton *et al.*, 2012). The structure is shown in Figure 1. This advantage is more obvious when the inputs of the network are multi-dimensional images, so that the images can be directly used as the input of the network, avoiding the complicated process of feature extraction and data reconstruction in the traditional recognition algorithm. CNN has so many advantages, but its application in ultrasound image recognition is rare, mainly because the small number of ultrasound image samples, the lower resolution and the large number of monochromatic images. In addition, because the tissue structure of ultrasound images is extremely complex, and the pathological features, the tissue features and their similarities are difficult to distinguish, which is also the key factor that causes the rare application of CNN in this field.

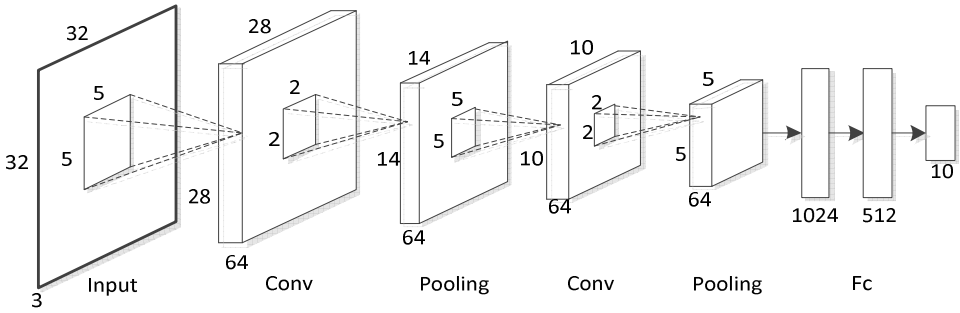

Figure 1. CNN Structure

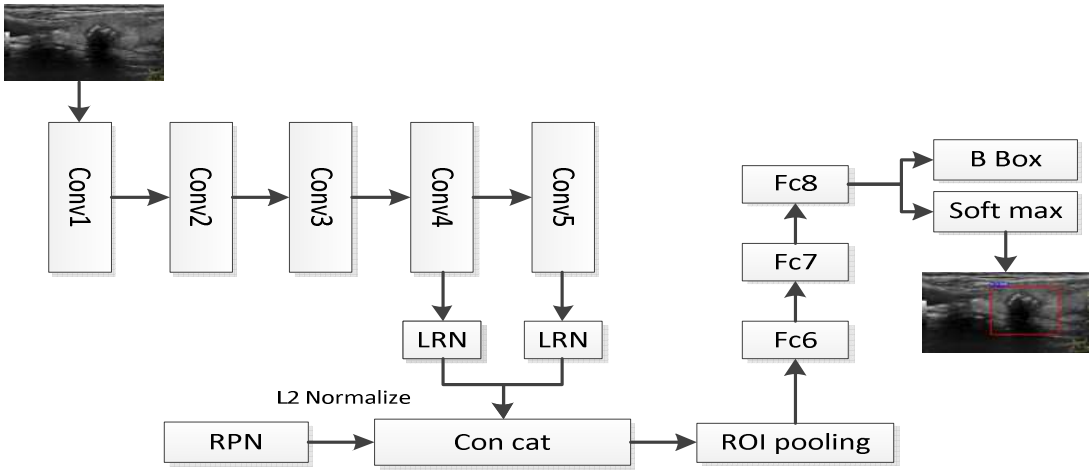

Figure 2. Faster RCNN Framework

Inspired by the object detection method of Faster RCNN (Fast Region-based Convolutional Network method) network based on CNN, this paper designs an improved Faster RCNN by layer linking, multi-scale input, multi-classification and fine tuning, with the framework shown in Figure 2 (Ke *et al.*, 2017). We use a large number of ultrasound images of the existing cases to extract the features of thyroid papillary carcinoma, and establish a stable, effective, accurate and specific diagnostic system, which is helpful to assist clinicians in the diagnosis and interns in training of the feature extraction of this kind of cancers. In addition, this can also help non-professionals have a preliminary understanding of their diseases.

Related Work

With the development of pattern recognition and machine learning, deep learning has become a hot topic in the field of machine learning (Zhang *et al.*, 2014; Längkvist *et al.*, 2014). Image recognition is a hot research topic, which mainly focuses on the classification and description of some objects or processes, and aims to develop a machine vision system which can automatically process some information and replace the traditional task of manual classification and recognition. However, there are relatively few applications in medical ultrasound images. The automatic classification of medical ultrasound images is essentially the comparison of similarities of images(Destrempes *et al.*, 2009), or called pattern recognition. Before the appearance of CNN, the main methods of image pattern recognition include SIFT (Hwang *et al.*, 2013), SUFR (Hwang *et al.*, 2013), BOW (Wu *et al.*, 2010), SVM (Prakosa *et al.*, 2014), KAZE (Alcantarilla *et al.*, 2012) and other algorithms. In the aspect of histopathology, for the medical ultrasound imaging system is influenced by its surroundings and imaging mechanism, the generation and transmission of ultrasound medical images will be disturbed by various noise sources, so that the acquired medical ultrasound images have poor image quality. Because of the non-uniformity of the imaged tissue and the coherence of the ultrasound imaging, a kind of peculiar noise, namely speckle noise, is formed in the medical ultrasound images. Speckle noise may obscure or even mask some important detailed information in medical ultrasound images, resulting in low signal-to-noise ratio, poor image quality, and difficult to conduct semi-automatic or automatic identification and medical diagnosis. To solve this problem, Zhang R et al. proposed an improved P-M model filtering method for determining diffusion threshold based on local information of images (Zhang *et al.*, 2015). This method can filter speckle noise more effectively. In addition, for incomplete gland features, Toki Y et al. used the SIFT method to extract features of images to identify prostate cancer (Toki *et al.*, 2012). The accuracy of this method is 6.3% - 13.3% higher than that of previous methods. Least Square Support Vector Machine (LS-SVM) classification algorithm is used to diagnose breast cancer for the color and texture features of biopsy specimens of breast cancer tissue (Niwas *et al.*, 2011). The above algorithms rely on limited resources of manual annotation, and can only match patterns for limited features. If the features change (such as distortion, flipping, illumination change, destruction and so on), the effects of these algorithms will be degraded. Therefore, the applicability is not strong.

At present, the form of convolution neural network (CNN) has been proved to be almost the best deep learning structure (Hinton *et al.*, 2012). CNN is also one of the most suitable methods for image feature extraction. In recent years, there have been some studies using CNN to classify histologic pathological images. Cruz-Roa et al. automatically segment pathological images of invasive breast cancer using CNN techniques and finally generate a feature map of the cancer (Cruz-Roa *et al.*, 2015). Kallenberg et al. implement breast density segmentation and breast risk assessment using a combination of supervised learning and unsupervised learning (Kallenberg *et al.*, 2011). In order to get a more accurate model, CNN often needs to train a large number of images (Moeskops *et al.*, 2016; Litjens *et al.*, 2016; Tajbakhsh *et al.*, 2016).In the field of medicine, it is difficult to get so many pictures. Because the features of histopathologic images are much more complex than those of natural images (Litjens *et al.*, 2016), and the pathological features are often very similar or not obvious to those of the surrounding tissues, so that it is sometimes difficult to distinguish by professional physicians (Russakovsky *et al.*, 2015) .

While CNN develops rapidly in the image recognition (Krizhevsky *et al.*, 2012; He *et al.*, 2016; Zeiler *et al.*, 2014; He *et al.*, 2015), Girsharp *et al.* turn the detection problem into classification problem and propose R-CNN structure (Girshick *et al.*, 2014). Then, for multiple-step operation of redundancy calculation and model training in R-

CNN (Girshick *et al.*, 2014), Girpick further proposes a Fast R-CNN frame structure, which integrates the whole detection process and performs only one feature extraction for each picture, greatly reduces the redundancy calculation and thereby improves the detection speed and the detection performance (Girshick *et al.*, 2015). In Fast R-CNN, the speed of extracting candidate region is slow, which is the bottleneck of the whole detection network. Ren et al. propose the Fraster R-CNN frame structure (Ren *et al.*, 2017). The work of obtaining candidate region is also accomplished by CNN. In addition, the candidate region extraction network and the target detection network share the feature extraction layer and achieve better detection performance.

From RCNN to Fast RCNN, and then to Faster RCNN, the four basic steps of target detection (candidate region generation, feature extraction, classification, and location refinement) are unified into a deep network framework. All calculations are not repeated and completed in the GPU, greatly improving the speed of operation. Faster RCNN can be simply considered as a "region generation network + Fast RCNN" system, and the region generation network is used to replace the Selective Search method in Fast RCNN.

In order to meet the needs of ultrasound image recognition in thyroid papillary carcinoma, Faster RCNN is improved as follows:

(1) In view of the low resolution features of ultrasound images, the Faster RCNN is not very effective for low resolution image recognition, so we connect the fourth layer and the fifth layer of the Faster RCNN to extract the cancer features using the combination of the deep layer and the shallow layer. In this way, the mean average precision (mAP) and the accuracy of ultrasonographic feature recognition of thyroid papillary carcinoma is improved by 7.8% and 1.7% respectively.

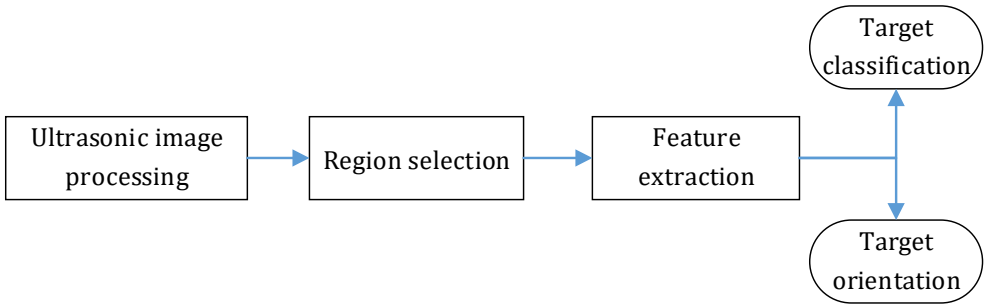

Figure 3. Flow chart of ultrasonic image recognition

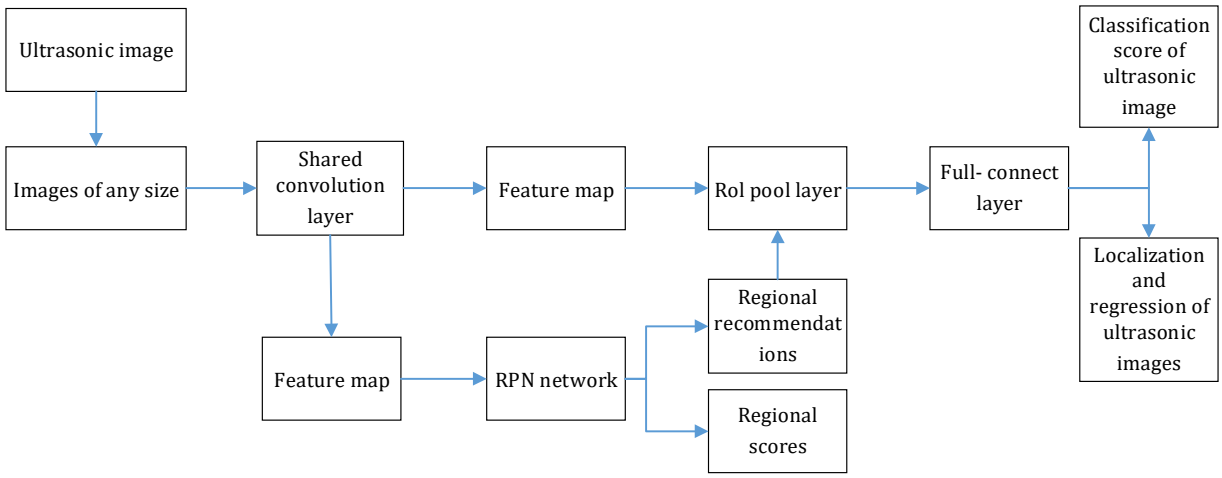

Figure 4. Ultrasonic image recognition scheme of thyroid papillary carcinoma

(2) The multi-dimension ultrasound image input improves the accuracy of local feature extraction and the efficiency of ultrasound image recognition in thyroid papillary carcinoma. The mAP is increased by 3.4% and the accuracy is increased by 9.2%.

(3) By means of multi-classification, the cancer images are labeled with multi-classification, so that the system can automatically generate simple ultrasonic image diagnosis report.

Faster R-CNN Network Model for Ultrasonic Image Feature Recognition of Thyroid Papillary Carcinoma

Overall plan

When identifying a complete ultrasound image, the andidate regions are firstly extracted from the image and include cancer feature targets in the ultrasound image as much as possible. A small number of multi-scale candidate regions can be obtained by selectively searching the whole image by the features such as texture and contour of thyroid papillary carcinoma, but this needs to be run on the CPU first. In this paper, in order to quickly identify the features of cancers, it's necessary to introduce the Region Proposals Network (RPN) to use GPU to accelerate, so that the end-to-end deep network structure can be realized, which is beneficial to network training. In this paper, the advantages of RPN network and Fast RCNN network are used to design the ultrasonic image recognition scheme of thyroid papillary carcinoma as shown in Figure 4.

The ultrasound image of thyroid papillary carcinoma is extracted simply to obtain a single frame image without normalization. Then it is input to the shared network and propagated forward to the last layer of the shared convolution layer through the convolution neural network. The feature map obtained at this time needs to be input into the RPN network as well as to continue to propagate forward to the RoI pooling layer. The feature response map input to the RPN network is processed to obtain scores of candidate regions and corresponding regions, and some candidate regions with inaccurate positions are deleted by non-maximum suppression, and Top-N is output (N can be adjusted, with a generally 200-300 score) for region of interest (RoI) pool layer.

The feature map of the shared convolution layer output and the region suggestion of the RPN network output are input to the RoI pool layer at the same time to extract higher-level features of the corresponding candidate region. Finally, the classification scores of the cancer features of the ultrasound images in the region and the positional localization after regression are output.

Ultrasonic feature recognition of thyroid papillary carcinoma based on Faster R-CNN

The traditional CNN is generally a neural network which consists of a linear convolution layer, a pool layer and a fully connected layer. The convolution layer performs a convolution operation through a linear filter, then performs a nonlinear operation on the result, and finally generates a feature map. At present, Faster RCNN has achieved remarkable results in target detection. However, our experiments have proven that it is not very effective to recognize thyroid papillary carcinoma directly from ultrasound images. The mAP (mean average precision) is only 0.62. The main reasons are as follows: The internal structure of the ultrasound image is complex and contains multiple individual tissues, and the tissues overlap each other, with unclear boundary, and the ultrasound image with malignant cancer features is very similar to some tissues; secondly, the medical ultrasound image is a monochromatic image formed by light spots with different brightness, and the resolution is low; third, as shown in Figure 5, overlapping covered portions between the respective tissues are difficult to distinguish from portions with cancer characteristics. Most of the existing CNN pre-training models are trained under large image database, such as ImageNet and VOC2007. These databases are made up of pictures in life, and their features differ greatly from those of cancers and are not suitable for ultrasound image training.

Figure 5. Artificial labeling of thyroid papillary carcinoma

In the aspect of cancer ultrasound image recognition, since the deep learning model can directly learn the hidden high-level features from the image, the shallow learning algorithm with other low-level features can obtain better effects. Generally, the training time required for deep learning model is longer than that of various shallow structure algorithms, but it depends on the model design and the selection of training set. In this study, we improve the FASTER RCNN network model, combine deep and shallow learning and multi-scale input to extract features, and effectively improve the recognition rate of ultrasound images.

(1) Layer connection

In order to extract the features better, according to the features of the low resolution of ultrasound images, feature extraction is performed by connecting the deep layer and the shallow layer of the shared convolution layer, and then the connected layer is normalized to construct the final network model. The experiment proves that we can connect the fourth layer and the fifth layer of the shared convolution layer to improve the accuracy of cancer feature recognition effectively.

(2) L2 Normalization

As shown in Figure 2(Ke *et al.*, 2017), when connecting the fourth and fifth layers of the shared convolution layer, in order to extend the depth feature of the defined objects on the plurality of convolution layers, it is necessary to pool the region of interest by combining two feature tensors to reduce the dimension thereof. In fact, the size, number, and pixel values of each layer of features are different, often with smaller values in the deep layers. As a result, random connection of deep and shallow layers may result in poor performance because the difference in size has too much impact on the following weights. Therefore, a direct solution to this problem is to standardize each ROI pool tensor before joining (Liu, *et al.*, 2015). Also, in this method, the system network is able to learn the value of the scale factor for each layer. The recognition accuracy is improved by this method.

Each tensor is normalized using L2, and normalization is accomplished within each pixel of the set feature tensor. After normalization, scaling is applied separately on each tensor, as follows:

$$\hat{x}=\frac{x}{\|x\|_2} \tag{1}$$

$$\|x\|_2=\left(\sum_{i=1}^{d}|x_i|\right)^{\frac{1}{2}} \tag{2}$$

where, x represents the original pixel vector, $\hat{x}$ represents a normalized pixel vector and d represents the number of channels in each ROI pool tensor.

The scaling factor γ_i is then applied to the ROI pool tensor for each channel:

$$y_i=\gamma_i\hat{x}_i \tag{3}$$

During training, the scaling factor γ is continuously updated and the input x is calculated using the back propagation and chain rule:

$$\frac{\partial l}{\partial \hat{x}}=\frac{\partial l}{\partial y}\cdot\gamma \tag{4}$$

$$\frac{\partial l}{\partial x}=\frac{\partial l}{\partial \hat{x}}\left(\frac{I}{\|x\|_2}-\frac{xx^T}{\|x\|_2^3}\right) \tag{5}$$

$$\frac{\partial l}{\partial \gamma_i}=\sum_{y_i}\frac{\partial l}{\partial y_i}\hat{x_i} \tag{6}$$

Where, $y=[y_1,y_2,\cdots,y_d]^T$.

(3) Multi-scale

CNNs are typically followed by a full - connect layer or classer, both of which require a fixed input scale. Therefore, the input data has to be cropped or warped, and these preprocessing can cause the loss of data or geometric distortion. The Faster R-CNN is also trained with fixed-size pictures. Our experiments have proven that the feature in different size range can be learned through multi-scale image input, which increases the robustness, reduces the influence of down sampling on the feature representation, improves the extraction efficiency of the original feature of the image, and raises the accuracy of cancer feature recognition.

(4) Multi classification

The Faster R - CNN output is a rectangular box with a category name and score. For the diagnosis of thyroid cancer, it is not enough to judge the presence or absence of cancer features only in this way. The output of more information may provide more diagnostic reference to doctors. Therefore, by studying the diagnosis report of tumor patients and the guidance of the chief physician provided by the hospital, it is concluded that the main basis

for judging the features of cancer in ultrasound images include unclear boundary, non-uniform echo, irregular morphology, and strong echo spot (calcification). In order to facilitate analysis, these features are represented by b, h, x, and q respectively. At the same time, in order to output these cancer features, each cancer area is marked multiple times in marking. Each area is marked with at least one "c", representing a region with cancer characteristics. If any of the above b, h, x and q features are met, then the appropriate label is used to label it. At the time of testing, a simple diagnostic report is automatically generated based on the label name and score, as shown in Figure 6(Ke *et al.*, 2017).

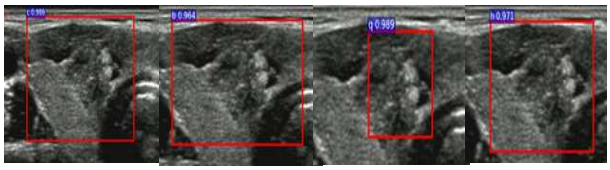

Figure 6. Detection report of thyroid papillary carcinoma

Experiments

Construction Of İmage Data Set

We collect ultrasound images of 307 people from the Cancer Center of Sun Yat-sen University from 2010 to 2014, including 54 men and 253 women, all of whom had their privacy removed. Each people has 5 to 30 ultrasound images, with a total of 4,738 ultrasound images. Of these patients, 256 are diagnosed with thyroid papillary cancer and have surgery, and 51 are diagnosed with normal thyroid. 51 people with normal thyroid have 1,153 images in total. A total of 1,367 ultrasound images are taken from 200 patients with thyroid papillary carcinoma confirmed as training samples, and the ultrasound images of the remaining 107 people (including 51 confirmed and 56 normal people) are used for testing. In addition, it is known from the diagnostic report that each ultrasound image of the diagnose people includes 1 to 3 cancer characteristics. According to the experimental needs, we randomly divide the trained ultrasound images into three groups, averaging 66 people in each group, keeping the original long-to-wide ratio, and setting the widths of the three groups to 800px, 600px and 400px respectively.

Annotation and classification of images

In order to make the cancer features of the labeled images accurate, an experienced chief physician uses the labeling software to label the training images of all confirmed patients. In addition, the ultrasonic image used for training should be in accordance with the xml format file required by Foster R-CNN training, and the ultrasonic image used for testing should be labeled as Ground True. Another chief physician reviews the images that have been labeled. A rectangular frame is used to label the image. The rectangular frame should completely surround the cancer area, so as to ensure the accuracy of the cancer area. In labelling, each cancer region is labeled "c" to indicate the cancer category. If the boundary is not clear, this area will continue to be labelled with "b"; if the shape is irregular, this area will continue to be labelled with "x"; if the echo is not uniform, this area will continue to be labelled with "h" and if a calcified area or spot is present, this area will continue to be labelled with "q". Finally, when we detect an image, as long as a region outputs one or more labels in {c b x h q}, the region is determined to be a cancer feature region. In the end, our training cancer images are labeled with 3,347 cancer features in total, and the cancer images used for testing are labeled with 805 cancer features. An xml file for region location coordinate and classification can be generated.

Introduction of sample training

The hardware environment of the training: Operating system: ubutun 14.04 64 digits; CPU:Intel® Xeon(R) CPU E5-1630 v3 @ 3.70GHz × 4; Memory: 64G; GPU: Graphics: Quadro K2200. We use the deep learning python version of the Foster R - CNN network framework, which can be accelerated using GPU. The CNN model uses ZF, and we use the VOC2007 database for pre-training. First, draw the object bounding box, which uses the opencv dynamic library encapsulated by others. Then produce the xml file and save the xml file to Annotations, put the training sample images into JPEG images, then overwrite the folder corresponding to the VOC2007 database, and then train it. The fourth and fifth layers of the ZF are connected and normalized by modifying the source code of the Faster R-CNN.

Training and results

Training 1: Using the original Faster RCNN network, the number of iterations is 10000, 5000, 10000, and 5000;

Training 2: Using the original Faster RCNN network, the number of iterations is 40000, 20000, 40000, and 20000;

Training 3: Using multi-scale input, the number of iterations is 40000, 20000, 40000, and 20000;

Training 4: Using layer connect (connecting the third layer and the fifth layer), the number of iterations is 40000, 20000, 40000, and 20000;

Training 5: Using layer connect (connecting the fourth layer and the fifth layer), the number of iterations is 40000, 20000, 40000, and 20000;

Training 6: Using multi-scale input and connecting the fourth layer and the fifth layer, the number of iterations is 40000, 20000, 40000, and 20000;

Training 7: Using multi-scale input and connecting the fourth layer and the fifth layer, the number of iterations is 80000, 40000, 80000, and 40000;

The learning rate is set to be 0.001;

The mAP training for each model is shown in Table 1 (Ke *et al.*, 2017).

Table 1. mAP training results

	Training						
	1	2	3	4	5	6	7
mAP	0.614	0.618	0.652	0.633	0.696	0.738	0.721

Test and result analysis

The results of the corresponding model test after each training are shown in Table 2:

Table 2. Test results for each model

	P	PR	N	NR
Ground Truth	805			
Model 1	138	0.171	667	0.829
Model 2	604	0.750	201	0.250
Model 3	678	0.842	127	0.158
Model 4	548	0.681	257	0.319
Model 5	618	0.768	187	0.232
Model 6	715	0.888	90	0.112
Model 7	686	0.852	119	0.148

Where, TP: the number of identified correct; the number of correctly identified thyroid papillary carcinomas, or true positives, is shown in this paper. TPR: the recognition rate of true positives. FN: indicating missing report, the number of matches not found correctly; the number of thyroid papillary carcinomas that are not recognized, i.e., false negatives, is represented herein. FNP: the recognition rate of false negatives.

Experiment 1: It can be seen from Training 1 and Training 2 in Table 1 that the mAP is higher when the number of iterations reaches 40000. After testing, it is found from Table 2 that the detection effect of Model 2 is good, the recognition rate of true positives is 75%, while Model 1 with the iteration of 10000 times is not different greatly from Model 2, but its recognition rate of true positives is very low. It's mainly because of the relatively large sample size, when the number of iterations is less, partial local feature extraction will be ignored, resulting in low recognition accuracy.

Experiment 2: We connect the deep and shallow layers of the shared convolution layer to improve the Faster RCNN network. The shared convolution layer has five layers, so which layer should we connect? In this experiment, we compare Model 4 with Model 5, and the mAP of model 5 is higher. After testing, we find that connecting the fourth layer and the fifth layer of the shared convolution layer is better than connecting the third layer and the fifth layer of the shared convolution layer, and the recognition rate of true positives is 76.8%, with the TPR improved by 8.7% compared to connecting the third lay and the fifth layer. It's is mainly because more aggregated feature information extracted by connecting the third layer and the fifth layer of the shared convolution layer is outside the ROI region, so that if the ROI region is small, the ratio of the useful feature information extracted is small. Thus, we use a method of connecting the fourth and fifth layers of the shared convolution layer in this study.

Experiment 3: It can be seen from Model 6 and Model 7 that mAP with 40000 iterations is higher. After testing, we find that the detection effect of Model 6 is better, reaching 88.8%, which is 3.6% higher than that of Model 7.

Ablation test:

In order to study the effect of applying our improved Faster RCNN on medical images, we use a cross-test to verify the method we use. From Experiment 1, 2, and 3, we select Model 2, Model 3, Model 5, and Model 6 as the study subjects. Corresponding marks are as follows: ID1, ID3, ID2, and ID4.

Table 3. Cross-validation

ID	Multi-scale input	Layer connect
1	No	No
2	No	Yes
3	Yes	No
4	Yes	Yes

For the use of different strategies and the original Faster RCNN network model, we use a cross-validation analysis as shown in Table 3.

With strategy methods, we get the results of the test and the relevant data as shown in Table 4 (Ke *et al.*, 2017).

This study focuses on the automatic identification of ultrasound images of thyroid papillary carcinoma, which is closely related to medicine. We rely on the working feature curve of the subjects, i.e. the ROC curve, which is a comprehensive index of continuous variables reflecting the sensitivity and specificity, reveals the relationship between the sensitivity and specificity by using a conformational diagram, and calculates a series of sensitivities and specificities by setting the continuous variables to a plurality of different critical values, and then takes the sensitivity as the ordinate and (1-specificity) as the abscissa to draw a curve. The larger the area under the curve is, the higher the diagnostic accuracy is. On the ROC curve, the point closest to the upper left of the coordinate graph is the critical value of both sensitivity and specificity. The method is simple and intuitive, and the clinical accuracy of the analysis method can be observed through the illustration, and can be judged with the naked eye. ROC curve combines sensitivity and specificity by the method shown in the diagram, which can accurately reflect the relationship between specificity and sensitivity of an analytical method, and is a comprehensive representative of testing the accuracy.

Compare two kinds of ROC curves for disease recognition. First, when comparing two or more diagnostic methods of the same disease, the ROC curve of each test can be drawn into the same coordinate to visually identify the merits and demerits, and the ROC curve close to the upper left corner represents the most accurate test worker. Second, by calculating the area (AUC) under the ROC curve of each test, the higher the AUC is, the higher the accuracy is.

Figure 7 shows the ROC curve for all methods (Ke *et al.*, 2017):

As can be seen from Figure 7, with the improvement of the method we use, the effect of ROC becomes better and better. ID4 uses all of our methods and achieves the best results. When TPR < 0.5 or FPR > 0.5, it is of no medical significance. Therefore, it can be seen from Figure 8 that the model obtained has a higher resolution for the features such as unclear edge, irregular shape, non-uniform echo, calcification or light spot of the ultrasound images of thyroid papillary carcinoma. Next, we analyze and compare ID1-ID4 of Table 3 respectively.

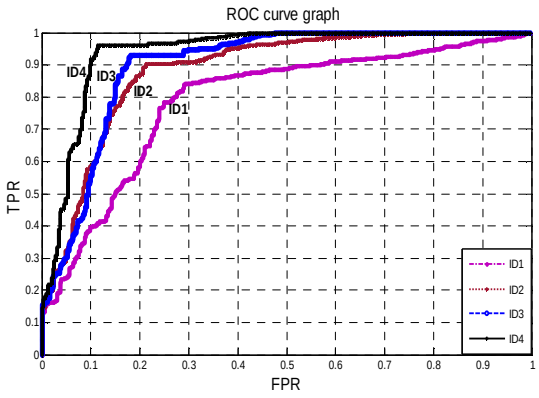

Figure 7. ROC curve graph

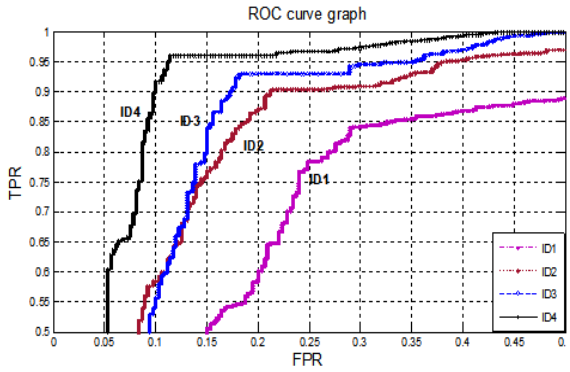

Figure 8. ROC curve graph

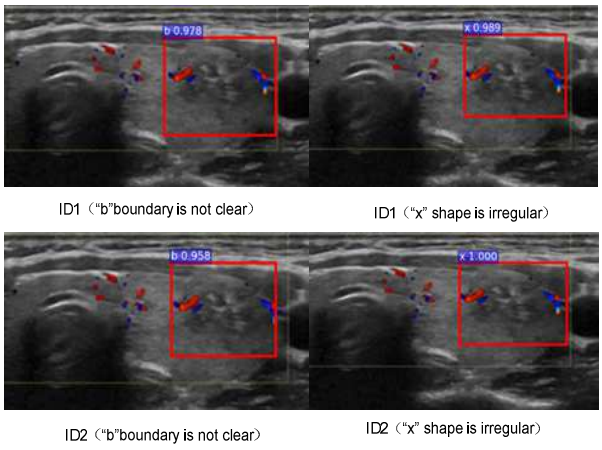

Figure 9. ID1 VS ID2

(1): ID2 VS ID1 (layer connection) As can be seen from Figure 8, due to the use of the layer connection, ID2 is significantly better than ID1 and its ROC curve is closer to the upper left corner. It can be seen from Table 4, TP and TN are

increased and FP and FN are decreased by connecting the fourth layer and the fifth layer. In addition, the layer connection makes the model more effective in identifying cancer features with irregular shape and low resolution but slightly less effective in identifying cancer features with unclear boundaries, as shown in Figure 9(Ke *et al.*, 2017).

(2): ID3 VS ID1 (Multi-scale input) As can be seen from Figure 8, ID3 is significantly better than ID1 with its ROC curve closer to the upper left corner due to the use of multi-scale input (the original long-to-wide ratio is maintained here, and three scales of 800 px, 600 px and 400 px are used). It can be seen from Table 4, the multi-scale input increases TP and TN and decreases FP and FN. As shown in Figure 10(Ke *et al.*, 2017), multi-scale input is significantly better for recognizing cancer features with unclear boundaries than models that do not use this strategy. In addition, ID3 has stronger recognition capability for cancer features of calcification or strong spots, while ID1 is better for echo non-uniform.

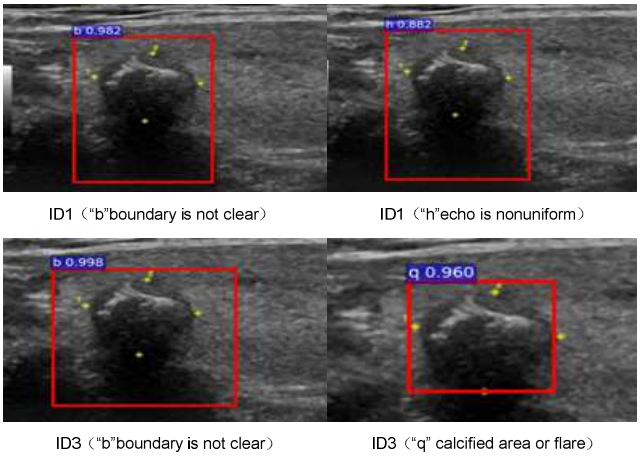

Figure 10. ID1 VS ID3

Table 4. Prediction results

	P	PR	P	PR	N	NR	N	NR
Ground Truth	805				196			
ID1	604	0.750	57	0.291	139	0.709	201	0.250
ID2	618	0.767	48	0.245	148	0.755	187	0.233
ID3	678	0.842	40	0.204	156	0.796	127	0.158
ID4	715	0.888	31	0.158	165	0.842	90	0.112

(3): ID2 VS ID3: As can be seen from Figure 8, the effect of multi-scale input is better than that of layer connection. Its ROC curve is closer to the upper left corner. However, when TPR is less than 0.6, the AUX of layer connection method is larger than that of ID3, but then the performance of ID2 is better than ID3. But when TPR is greater than 0.6, AUX of multi-scale input is larger than that of ID2, and its ROC curve is closer to the upper left corner, so the recognition effect is better. As shown in Figure 11, by selecting two of the detection results, ID2 recognizes a total of 3 cancer features, and ID3 recognizes a total of 5 cancer features. Both have relatively high accuracy from the score obtained.

(4): ID4 VS ID1, ID2 and ID3: As can be seen from Figure 8, ID4 detection is better than the others when all method strategies are used. Its AUX value is significantly greater than the other methods, and its ROC curve is closer to the upper left corner. Figure 12 shows the detection graph after using all the methods.

Experiments show that ID4 (improved Faster RCNN) works best with all methods, and we compare it with artificial recognition performance, as shown in Table 5:

Table 5. Performance comparison between improved Faster RCNN and Handcraft

		PR	P	PR	N	NR	N	NR
Ground Truth	805				196			
Hand-crafted	638	0.793	39	0.199	157	0.801	167	0.207
Faster RCNN	715	0.888	31	0.158	165	0.842	90	0.112

As can be seen from Table 5, for false positives, the difference between Faster RCNN and artificial recognition is not significant, reaching 19.9% and 15.8%, respectively. The reason is that pathological image features are very complex, for example, tissue folds, shadows, overlapping and other features are very similar to cancer characteristics, and it is difficult for both Faster RCNN and artificial recognition to distinguish. However, Faster RCNN had a higher number and a higher ratio of true positives and negatives than artificial recognition. False-negative ratio is one of the most concerned features in the medical field. The excessively high number and ratio of false negatives may cause serious misjudgment. Obviously, the cancer feature is judged to be a non-cancer characteristic,

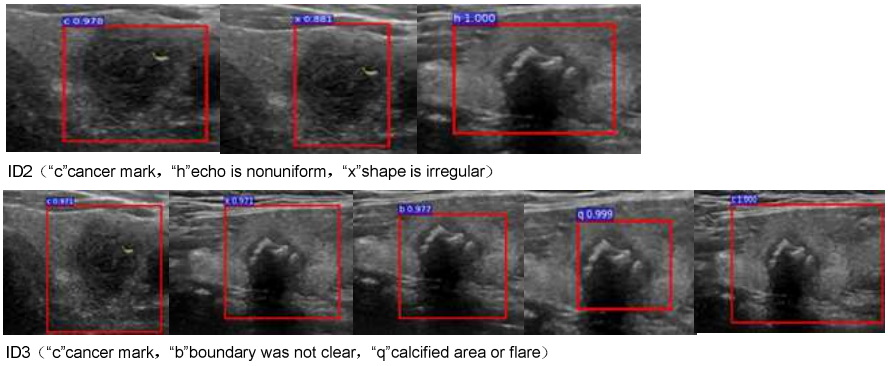

Figure 11. ID2 VS ID3

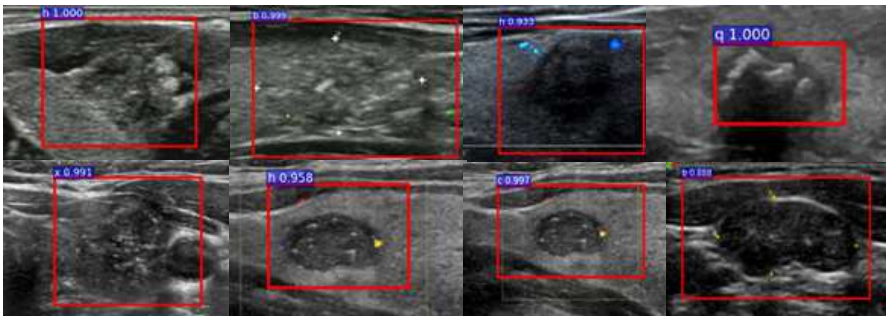

Figure 12. ID4

which is more harmful than the non-cancer feature is judged to be a cancer characteristic. In this respect, the effect of Faster RCNN is much better than that of artificial recognition. The FNR of Faster RCNN is 11.2% and that of artificial recognition is 20.7%. Obviously, the effect of Faster RCNN is better than that of artificial recognition.

Conclusions

In order to overcome the problems such as small number of images, non-obvious features of cancer, complex structure of internal tissues and low pixels in ultrasonic image recognition of thyroid papillary carcinoma, we improve the Faster RCNN network model based on CNN and adopt multi-scale input and layer connect to extract cancer features better. Through a large number of experiments, the detection model that can recognize the features of cancer with complex structure is obtained, with good recognition effect. The experimental results show that the recognition rate of true positives by the proposed network model is 88.8%, which can accurately judge whether there is cancer feature in the ultrasound image of thyroid papillary carcinoma. In the future, how to integrate the training of various medical image features to enlarge the number of samples, so as to improve the true positive rate of cancer features needs to be further studied.

Acknowledgments

This work was supported in part by the degree and graduate education reform project of Guangdong Province under grant No.2016JGXM-MS-26 and the higher education quality project of Guangdong University of Technology. This work is an extended version of our ICONIP paper(Ke *et al.*, 2017). Thanks Hailiang Li and Weiwei Liu for their suggestions on this work.

References

Alcantarilla P F, Bartoli A, Davison A J. KAZE Features. European Conference on Computer Vision. Springer, Berlin, Heidelberg 2012: 214-27.

Cruz-Roa A, Basavanhally A, Gonzalez F, Feldman M, Ganesan S, Shih N, Tomaszewski J, Gilmore H, Madabhushi A. A Feature Learning Framework for Reproducible Invasive Tumor Detection of Breast Cancer in Whole-Slide Images. Laboratory Investigation 2015: 40A.

Davies L, Welch HG. Increasing Incidence of Thyroid Cancer in the United States, 1973-2002. JAMA 2006; 295(18): 2164–67.

Destrempes F, Meunier J, Giroux M F, et al. Segmentation in Ultrasonic B-Mode Images of Healthy Carotid Arteries Using Mixtures of Nakagami Distributions and Stochastic Optimization. IEEE Transactions on Medical Imaging 2009; 28(2): 215-29.

Girshick R, Donahue J, Darrell T, Malik J.Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014: 580-87.

Girshick R. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015:1440-48.

He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016: 770-78.

He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015: 1026-34.

Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine 2012; 29(6): 82-97.

Hwang C L, Chou Y J, Lan C W. Comparisons Between Two Visual Navigation Strategies for Kicking to Virtual Target Point of Humanoid Robots. IEEE Transactions on Instrumentation & Measurement 2013; 62(11): 3050-63.

Kallenberg M, Karssemeijer N. Multi-class probabilistic atlas-based segmentation method in breast MRI. Iberian Conference on Pattern Recognition and Image Analysis. Springer-Verlag, 2011: 660-67.

Ke W, Wang Y, Wan P, Liu W, Li H. An Ultrasonic Image Recognition Method for Papillary Thyroid Carcinoma Based on Depth Convolution Neural Network. InInternational Conference on Neural Information Processing 2017; 10635: 82-91.

Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097-1105.

Längkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters 2014;42(1):11-24.

Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-Van De Kaa C, Bult P, Van Ginneken B, Van Der Laak J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific Reports 2016; 6: 26286.

Liu W, Rabinovich A, Berg A C. ParseNet: Looking Wider to See Better. arXiv:1506.04579v2.

Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Transactions on Medical İmaging 2016; 35(5):1252-61.

Niwas SI, Palanisamy P, Zhang WJ, Isa NA, Chibbar R. Log-gabor wavelets based breast carcinoma classification using least square support vector machine.IEEE International Conference on Imaging Systems and Techniques. IEEE 2011: 219-23.

Prakosa A, Sermesant M, Allain P. Cardiac electrophysiological activation pattern estimation from images using a patient-specific database of synthetic image sequences. IEEE Transactions on Biomedical Engineering 2014; 61(2): 235-36.

Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.IEEE Transactions on Pattern Analysis and Machine Intelligence 2017, 39(6):1137-49.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 2015;115(3): 211-52.

Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. Convolutional neural networks for medical image analysis: Full training or fine tuning?. IEEE transactions on medical imaging 2016; 35(5):1299-312.

Toki Y, Tanaka T. Image feature extraction method with SIFT to diagnose prostate cancer. Sice Conference. IEEE 2012: 2185-88.

Wu L, Hoi S C H, Yu N. Semantics-Preserving Bag-of-Words Models and Applications. IEEE Transactions on Image Processing 2010; 19(7): 1908-20.

Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks.European Conference on Computer Vision 2014: 818-33.

Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks. IEEE Winter Conference on Applications of Computer Vision 2014: 1036-41.

Zhang R,Liu Q,Liu Q. The Research on Preprocessing for the Gray-Scale Ultrasound Breast Tumor Images of 76 Cases. Hans Journal of Biomedicine 2015; 5: 9-16.