



# A pre-trained convolutional neural network based method for thyroid nodule diagnosis



Jinlian Ma<sup>a</sup>, Fa Wu<sup>a</sup>, Jiang Zhu<sup>b</sup>, Dong Xu<sup>c</sup>, Dexing Kong<sup>a,\*</sup>

<sup>a</sup> School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> Department of Ultrasound, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310020, China

<sup>c</sup> Department of Ultrasound, Zhejiang Cancer Hospital, Hangzhou 310022, China

## ARTICLE INFO

### Article history:

Received 10 March 2016

Received in revised form 26 July 2016

Accepted 11 September 2016

Available online 12 September 2016

### Keywords:

Ultrasound image

Thyroid nodule

Convolutional neural network

Feature extraction

Classification

Diagnosis

## ABSTRACT

In ultrasound images, most thyroid nodules are in heterogeneous appearances with various internal components and also have vague boundaries, so it is difficult for physicians to discriminate malignant thyroid nodules from benign ones. In this study, we propose a hybrid method for thyroid nodule diagnosis, which is a fusion of two pre-trained convolutional neural networks (CNNs) with different convolutional layers and fully-connected layers. Firstly, the two networks pre-trained with ImageNet database are separately trained. Secondly, we fuse feature maps learned by trained convolutional filters, pooling and normalization operations of the two CNNs. Finally, with the fused feature maps, a softmax classifier is used to diagnose thyroid nodules. The proposed method is validated on 15,000 ultrasound images collected from two local hospitals. Experiment results show that the proposed CNN based methods can accurately and effectively diagnose thyroid nodules. In addition, the fusion of the two CNN based models lead to significant performance improvement, with an accuracy of  $83.02\% \pm 0.72\%$ . These demonstrate the potential clinical applications of this method.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Thyroid nodules are one of most commonly nodular lesions in the adult population. The National Cancer Institute has estimated that 64,300 new thyroid cancer cases will occur and the deaths related to thyroid cancer will be about 1980 in the world [1]. Moreover, the incidence of thyroid cancer has increased 2.4-fold over the last 30 years [2]. This increase is among the largest increases that has happened in all types of cancers [3]. Although the incidence of thyroid cancer is high, in more than 50% of the adults, most thyroid nodules are benign and only 7% turn out to be malignant [4]. However, 15–30% of thyroid nodules are diagnosed as indeterminate or suspicious for malignancy [5]. The non-surgical diagnosis of thyroid nodules is fine needle aspiration (FNA) biopsy, which is a critical diagnostic test for determining the proper management of thyroid nodules [6]. However, many physicians are confused about the various echo patterns of thyroid nodules. Approximately, 10–20% of thyroid biopsies by the FNA are nondiagnostic [7]. A surgical approach is generally recommended in such nodules even if more than 70% of them are benign at histological

examination, excisional biopsy is the best for differentiation [8]. However, both FNA biopsy and excisional biopsy are too labor intensive for large scale screenings. Moreover, unnecessary biopsies will make the patients more anxious and increase the health-care costs.

Being a real-time and non-invasive diagnosis that provides an image of the characteristics of thyroid nodules, ultrasound technology becomes a widely employed imaging method for the diagnosis and follow-up of thyroid nodules [9,10]. However, ultrasound images are easily affected by echo perturbations and speckle noise. Different types of thyroid nodules tend to have various internal echogenicities in the sonogram. Moreover, many benign and malignant nodules are similar in internal characteristics (see Fig. 1). Therefore, accurate visual interpretation of the differentiation of these thyroid nodules can only be done by experienced physicians. Otherwise, it results in subjective interpretations and inter-observer variabilities [11].

To eliminate the operator dependency and improve the diagnostic accuracy, computer aided diagnosis (CADx) based methods have been developed to detect and classify thyroid nodules recently [11,12]. The conventional designs of CADx are often composed of three main steps: pre-processing [13–15], feature extraction and selection [13,16–18], and classification. These three

\* Corresponding author.

E-mail address: [dkong@zju.edu.cn](mailto:dkong@zju.edu.cn) (D. Kong).

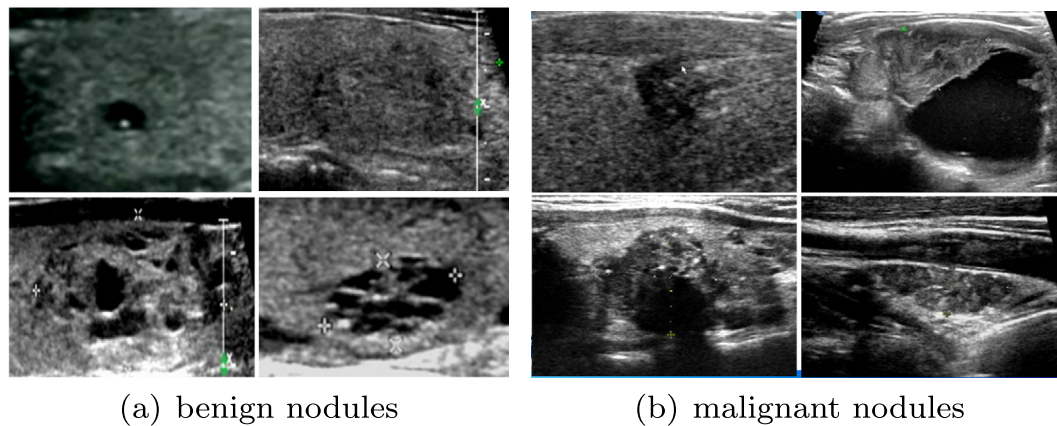


Fig. 1. Typical instances of thyroid nodules.

steps need to be well-addressed separately and then integrated together for the overall CADx performance tuning. In general, the pre-processing includes denoising, contrast enhancement, edge enhancement, segmentation and so on, these pre-processing steps are not only tedious, but also have a great effect on the following processing. The diagnostic image features can be categorized into morphological and textural features. The classifiers can be support vector machine (SVM), k-nearest neighbors, AdaBoost, gaussian mixture model, probabilistic neural network, decision tree and so on [11,19–21]. Particularly, SVM based methods with textural features are commonly used for classification of thyroid nodules [13,22,15]. For instance, Ding et al. [22] extracted statistical and textural features from thyroid elastograms, and then trained SVM to detect malignancy of thyroid nodules with a maximum classification accuracy of 95.2%. However, the classification accuracy is effected by a hard threshold. Singh and Jindal [15] first extracted 13 gray level co-occurrence matrix (GLCM) features, and then utilized SVM to classify thyroid nodules with a maximum classification accuracy of 84.62%. Acharya et al. [17] extracted fractal dimension, local binary pattern, laws texture energy features, and then also used SVM to classify thyroid nodules. Although these studies have gained encouraging results, they are mostly based on handcrafted features extracted from images with a series of pre-processing. The extraction of effective features is a challenge task that requires assistance from the latter steps of feature selection and feature integration by classifier. Moreover, it is also very difficult to select the most significant features and integrate different kinds of features.

To overcome the aforementioned difficulties, we consider the deep convolutional neural networks (CNNs) in this study. CNNs are a type of multi-layer, fully trainable models that can capture highly nonlinear mappings between inputs and outputs. In this study, we employ a hybrid convolutional neural network (CNN) for thyroid nodule classification in 2D ultrasound images. One appealing advantage of CNNs is that it is data driven learning of features. Our CNNs take thyroid nodule image patches as inputs and then generate the feature maps as outputs. The multiple intermediate layers apply convolution, pooling, and normalization operations to transform the inputs to the outputs. The networks contain millions of trainable parameters that are adjusted on a set of manually delineated thyroid nodule image data. Moreover, CNNs automatically learn a hierarchy of effective features from thyroid nodule ultrasound images by building high-level features from low-level ones and train a softmax to classify thyroid nodules. Fig. 2 shows the framework of our CNN based methods. Firstly, the regions of interest (ROIs) for thyroid nodules are coarsely delineated by physicians. Secondly, two CNNs with different convolutional layers and

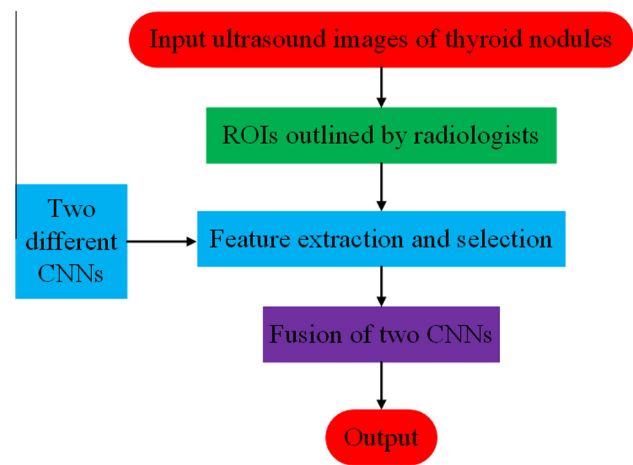


Fig. 2. Flowchart of the proposed CADx technique for thyroid nodule classification.

fully-connected layers are trained, separately. Finally, fused feature maps learned by the two CNNs are used to classify thyroid nodules based on a softmax. The main contributions of this study include the following several aspects:

1. To the best of our knowledge, it is the first attempt to apply CNNs to the task of thyroid nodule diagnosis.
2. A special CNN based method is developed to classify thyroid nodules, which is a fusion of two different CNN architectures. Moreover, these two CNNs use a pre-training strategy for initialization to avoid local optimum and employ a multi-view strategy to improve the performance.
3. Many clinical thyroid nodule images without any complex pre-processing are used in the study. Moreover, an objective and encouraging performance without any user intervention is achieved.
4. Novel single-valued integrated indices called TMI are determined to identify benign and malignant conditions more objectively. It is easy and transparent for the end users to diagnose thyroid nodules using this technique.

The rest is organized as follows. In Section 2, materials and the proposed method are described in details. Section 3 provides the implementation details and the results. In Section 4, we discuss the performance and comparison with the results of related techniques in the literature. The advantages and limitations of our technique are also presented. Conclusions are given in Section 5.

## 2. Materials and methods

### 2.1. Image acquisition and pre-processing

The methods are carried out in accordance with the approved guidelines. All experimental protocols are approved by two local hospitals, i.e. Sir Run Run Shaw Hospital, Zhejiang University School of Medicine and Zhejiang Cancer Hospital. 8148 fully-anonymized thyroid nodules (single or multiple) acquired from 4782 patients (with mean age 52) are studied. These thyroid nodules are referred for FNA biopsy following the physicians in ultrasound guidelines. The mean nodule size is  $2.5 \times 1.8 \times 2.1 \text{ cm}^3$ . For all patients with malignant nodules and some patients with benign nodules who underwent surgery, the final diagnosis is based on the histopathological examination of the excised thyroid nodule. Unless a patient underwent surgery, FNA results are used as the ground truth. In this study, 4126 thyroid nodules are benign and 4022 ones are malignant. Each thyroid nodule has several longitudinal cutting maps or crosscutting maps. In total, 15,000 thyroid nodule images are obtained from different sonographic systems (Philips, GE-Healthcare, Esaote, Toshiba, Siemens, mindray and Hitachi) before surgery or FNA. Details are shown in Table 1. The boundary of thyroid nodule in each image is manually delineated by physicians. So we can get 15,000 masks. These thyroid nodule images and their corresponding masks (Fig. 3) are in the same size and are both used for training our CNNs simultaneously.

### 2.2. CNN-based methods for diagnosing thyroid nodules

CNNs are a typical type of deep learning models [23–25], which are originally motivated from computer vision problems and intrinsically suitable for image-related applications [26–29]. In general, CNNs consist of standard structure-stacked convolutional layers (denoted conv, optionally followed by contrast normalization and pooling), followed by one or more fully-connected layers (denoted Fc) as in a standard multi-layer neural network, and have pointwise nonlinear function at the end of or after each layer [29]. Trainable convolutional filters, local neighborhood pooling and normalization operations are alternately applied on the input images, resulting in a hierarchy of increasingly high-level features. When trained with appropriate regularization [30–32], CNNs can achieve superior performance on visual objection and image classification tasks [29,25,33]. Recently, CNNs have been explored in detection tasks from medical domain, such as mitosis detection in breast cancer histology images by means of supervised CNNs [34,35], lung tissue classification with an unsupervised pre-trained CNN [36], and lung nodule classification by means of multi-scale CNNs [37].

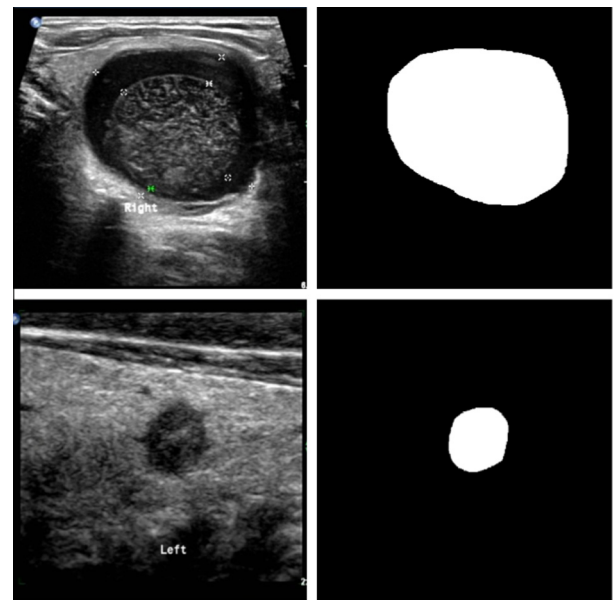
### 2.3. Architectures of CNNs

In this study, based on two proposed CNN architectures, a fused CNN is designed to diagnose thyroid nodules. A CNN based CADx framework could be realized in two steps: feature learning and classification. In the following, we provide details on one of the CNN architectures with input patch size of  $225 \times 225$  to explain the techniques used in this study. The detailed architecture is shown in Fig. 4. This CNN (referred to as CNN1) architecture contains 2 input feature maps corresponding to benign and malignant thyroid nodule image patches of  $225 \times 225$ . In feature learning step, it applies three convolutional layers with normalization and pooling operations to extract features. In classification step, softmax is adopted with the final fully-connected layer feature maps to classify thyroid nodules.

**Table 1**

Details of thyroid nodule data studied in our experiments. Most of them are jpg data, only 151 are DICOM data. Total format: number of benign nodules + number of malignant nodules.

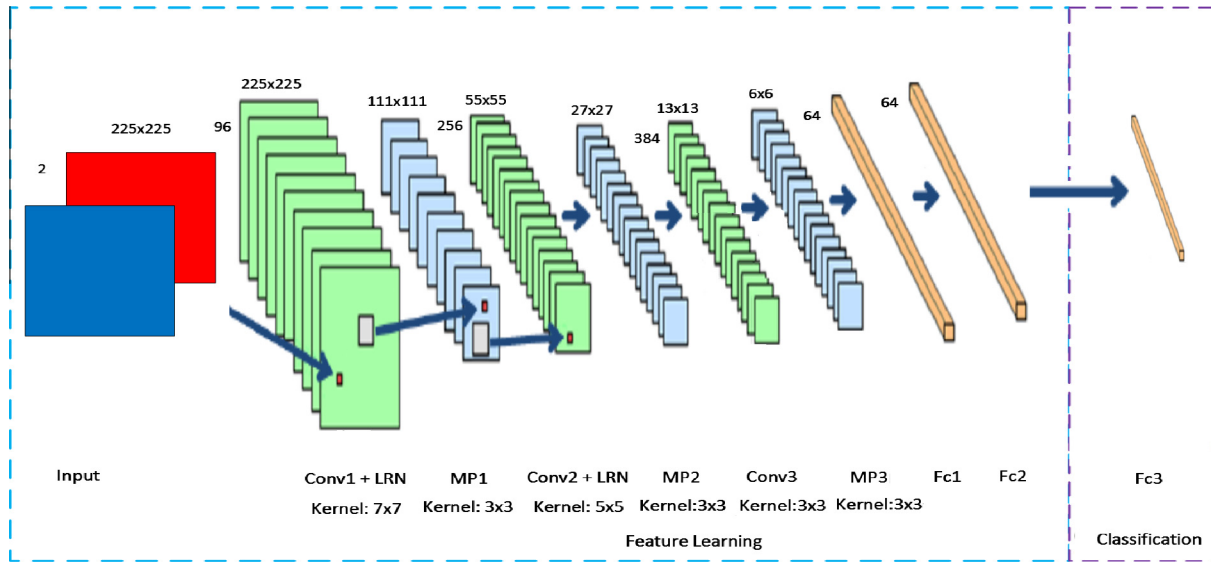
Hospital System	Sir Run Run Shaw Hospital Number of nodules	Zhejiang Cancer Hospital
Philips	5075	771
GE-Healthcare	230	1030 (151 DICOM)
Esaote	95	189
Toshiba	136	–
Siemens	118	117
mindray	119	–
Hitachi	268	–
Total	3260 + 2781	866 + 1241



**Fig. 3.** Typical sub-images of thyroid nodules (left) and their corresponding binary masks (right) used for training the CNNs.

In detail, the first convolutional layer generates 96 feature maps of size  $111 \times 111$ . Each feature map is connected to both of the two input feature maps through filters of size  $7 \times 7$ . A stride size of two pixels and a padding size of three pixels are used in this layer. The second convolutional layer generates 256 feature maps of size  $55 \times 55$ . Each of the feature maps is connected to all of the feature maps in the previous layer through filters of size  $5 \times 5$ . A stride size of two pixels is used in this layer. The third convolutional layer generates 384 feature maps of size  $13 \times 13$ . They are connected to all feature maps in the previous layer through filters of size  $3 \times 3$ . A stride size of one pixel is used in this layer. The function parametric rectified linear unit (PReLU) [38] is used as the activation function.

In addition, a few other types of layers have been used in the CNN1. Specifically, the max-pooling with window size of  $3 \times 3$  is applied after each convolutional layer. A stride size of two pixels is used in all the max-pooling layers. The local response normalization scheme is applied after each of the first two PReLU operations to enforce competitions between features at the same spatial location across different feature maps [25]. The three fully-connected layers following after the third max-pooling layer have 64, 64, 1 outputs, respectively. A 2-way softmax layer is used to generate a distribution over the labels after the output of the third fully-connected layer. Moreover, the network minimizes the cross entropy loss between the predicted labels and ground truth labels.



**Fig. 4.** Detailed architecture of the CNN1, taking patches of size  $225 \times 225$  as inputs (Conv: convolutional layer; MP: max-pooling layer; LRN: local response normalization; Fc: fully-connected layer).

The second CNN (referred to as CNN2) architecture consists of five convolutional layers, three fully-connected layers with 4096, 4096, 1 outputs, respectively and the other layers are the same as the CNN1. The filter size, the stride and the number of feature maps for the first two convolutional layers are also the same as the CNN1. In addition, the former two convolutional layers and the last convolutional layer are followed by max-pooling with a stride of 2, window size of 3. The complete details of the two CNN architectures can be seen from Table 2.

To the best of our knowledge, different CNN architectures can learn different features, shallow network can be suitable for learning low-level features and deep network can take full advantage of learning high-level features. Moreover, with the more network layers, the more complex features can be learned. So our fused CNN architecture (see Fig. 5) is based on CNN1 and CNN2, which consists of all the layers of the two CNNs. The size of filters and strides of these layers are all the same as the corresponding layers of two CNNs are, respectively. This structure can take advantage of CNN1 to capture the subtle low level features, and CNN2 to capture the complex high-level features from thyroid nodules so that it can learn multiple levels of features from thyroid nodules. Feature maps from six fully-connected layers are fused by a sumlayer to train a softmax.

#### 2.4. Training different CNN architectures

CNNs are built from the composition of several layers of non-linearities, with the suspected presence of many local minima. A gradient based optimization should thus end in the local minimum of whatever basin of attraction we start from. Nevertheless, the advantage of pre-training could be that it puts us in a region of parameter space where basins of attraction run deeper than when picking starting parameters at random [39]. So our CNN architectures are pre-trained with a set of 1.3 million natural images from the ImageNet database [40]. The parameters of 101 iterations are used to initialize the starting parameters of our CNNs. Then we train the networks using data consisting of patches extracted from the thyroid nodules. The image patches of size  $225 \times 225$  cropped sampled randomly from thyroid nodule images are the inputs of our CNN based models, whose centers are in their corresponding masks. Then the whole networks is fine-tuned by backpropagation with stochastic gradient descent [25], and simultaneously computes

**Table 2**

Architectures of CNNs used in this study. Here, “ $\frac{1}{2}$ ”, “ $\frac{1}{1}$ ” denote a stride of 2 and of 1, respectively. conv format: filter size, number of input, number of output feature maps, stride. max-pooling format: filter size, stride. fc format: number of output feature maps.

Layer	Input size	CNN1	CNN2
conv1	$225 \times 225$	$7 \times 7, 2, 96, \frac{1}{2}$	$7 \times 7, 2, 96, \frac{1}{2}$
max-pooling1	$111 \times 111$	$3 \times 3, \frac{1}{2}$	$3 \times 3, \frac{1}{2}$
conv2	$55 \times 55$	$5 \times 5, 96, 256, \frac{1}{2}$	$5 \times 5, 96, 256, \frac{1}{2}$
max-pooling2	$27 \times 27$	$3 \times 3, \frac{1}{2}$	$3 \times 3, \frac{1}{2}$
conv3	$13 \times 13$	$3 \times 3, 256, 384, \frac{1}{1}$	$3 \times 3, 256, 384, \frac{1}{1}$
conv4	$13 \times 13$	–	$3 \times 3, 384, 384, \frac{1}{1}$
conv5	$13 \times 13$	–	$3 \times 3, 384, 256, \frac{1}{1}$
max-pooling3	$6 \times 6$	$3 \times 3, \frac{1}{2}$	$3 \times 3, \frac{1}{2}$
fc1	–	64	4096
fc2	–	64	4096
fc3	–	1	1

discriminative feature maps. Finally, these feature maps are used to train a softmax for classifying thyroid nodules. In order to reduce overfitting, we use a regularization method, named split Dropout (sDropout) [41], which is a new extension of Dropout. This method trains a network in two sub-networks, through a random sample vector, rather than drops half of units. Then the sub-networks are taken as two different inputs for the consequent weight connections and activations. In this study, all the weights are trained while the co-adapting between units is broken with probability 0.5.

Training our fusion of CNN structures is based on the two CNN architectures above. Firstly, as the way of training above, two networks (CNN1 and CNN2) with different convolutional layers and fully-connected layers are trained, separately. Secondly, we fuse the multi-level feature maps learned by trained convolutional filters, pooling and normalization operations of the two CNNs. Finally, a softmax classifier with the fused feature maps is used to diagnose thyroid nodules (see Fig. 6).

#### 2.5. Performance measurement

In this study, diagnosis performance is evaluated by accuracy, sensitivity, specificity and receiver operating characteristic (ROC).



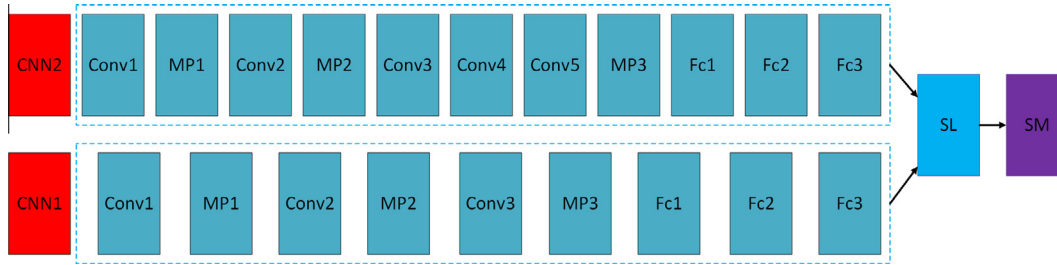


Fig. 5. Architecture of the fused CNN (Conv: convolutional layer; MP: max-pooling layer; Fc: fully-connected layer; SL: sumlayer; SM: softmax layer).

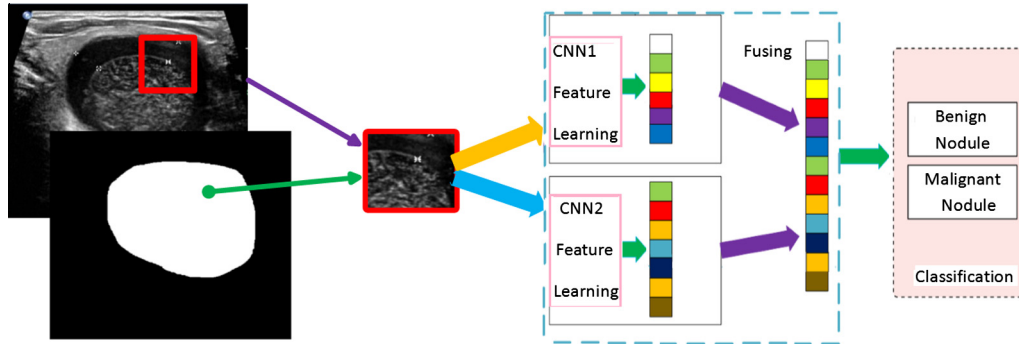


Fig. 6. Training of the fused CNN based on two CNNs with different convolutional layers and fully-connected layers.

Sensitivity measures the percentage of benign nodules correctly identified as benign. Specificity indicates the percentage of malignant nodules correctly identified as malignant. Accuracy measures the percentage of correctly diagnosing thyroid nodules. ROC is a 2D-plot with x-axis indicating '1-specificity' and the y-axis 'sensitivity'. The area under the ROC curve (AUC) indicates the classifier performance across the entire range of cut-off points and the value of AUC falls in the range between 0.5 and 1, which represents the probability of correctly distinguishing between benign and malignant nodule patterns. A good classifier will have an AUC close to unity [42].

### 2.6. Thyroid malignancy index (TMI)

In most classifiers, the end user would be unable to comprehend how the classifier arrived at the output class label. In this study, to make the classification process more transparent to the end user, it would be beneficial to use single-valued indices formed using a combination of the features learned by CNN1 and CNN2. This thyroid malignancy index (TMI) for each image can be represented as:

$$TMI = \frac{CNN(fc1) \times CNN(fc2)}{CNN(fc3)} \quad (1)$$

where

$$\begin{aligned} CNN(fc1) &= CNN1(fc1) \times CNN2(fc1) \\ CNN(fc2) &= CNN1(fc2) \times CNN2(fc2) \\ CNN(fc3) &= CNN1(fc3) \times CNN2(fc3) \end{aligned} \quad (2)$$

These indices are non-dimensional, and hence, do not have any units. In the clinical context, a threshold can be set for these indices to enable objective classification by the doctors as to whether the thyroid nodule is benign or malignant. Table 3 shows the ranges of the TMI indices for the two classes. These indices will help in a faster and more objective diagnosis of benign and malignant thyroid

nodules. They can be incorporated into a software application at no cost. Moreover, as can be seen from Table 3, that they are distinctly different for both the classes without any overlap.

## 3. Experiment results

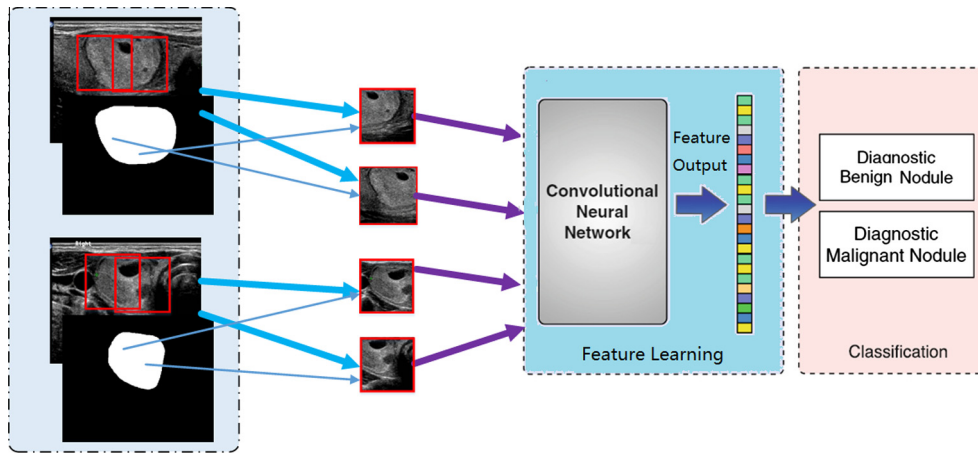
### 3.1. Implementation details

In this study, we focused on evaluating our CNN architectures for thyroid nodule classification. For CNN1 and CNN2, we performed a single testing in the training step. To improve the performance, we adopted a strategy of multi-view in the testing step. Specifically,  $225 \times 225$  cropped 256 views sampled randomly from each thyroid nodule image were the inputs of trained CNNs and the outputs were the average results of 256 views (see Fig. 7). Our fused CNN used the fused feature maps, learned by the two CNNs in multi-view testing, to train a softmax for thyroid nodule classification. For comparison purposes, we also implemented one commonly used classification method, namely SVM. The SVM with radial basis function (RBF) kernel was used in our experiments.

To evaluate the reliability and stability of the classifier, we performed 10-fold cross validation in all experiments with the same dataset. In detail, the original dataset was randomly divided into ten non-overlapping groups, carefully avoiding intersection of data from the same patient across folds. Nine folds were for training and one for testing. One fold was used to tune parameters of CNNs and then they were set as follows. Both the weight decay and bias decay were 0.0005. The learning rate was set 0.02 (applied to the average gradient on a minibatch 10). The momentum was 0.9 linearly over 10 epochs. In addition, for PReLU, the coefficient controlling the slope of the negative part with initial value 0.9 decreased at a speed of 0.01. For local response normalization, the parameters  $k$ ,  $n$ ,  $\alpha$ ,  $\beta$  [25] were set  $k = 96$ ,  $n = 24$ ,  $\alpha = 0.0005$ ,  $\beta = 0.75$  in the first normalization layer, and  $k = 256$ ,  $n = 64$ ,  $\alpha = 0.0005$ ,  $\beta = 0.75$  in second normalization layer, respectively. At the same time, we obtained a coarse approximation of the

**Table 3**  
Range of TMI for benign and malignant thyroid nodules.

CNN features	Benign	Malignant	<i>p</i> -value
TMI	$2.5308e + 16 \pm 6.3198e + 18$	$3.8651e + 15 \pm 9.1116e + 17$	<0.0001



**Fig. 7.** An overview testing of CNNs. Our CNN based approaches first extract multiple nodule patches to capture the wide range of nodule variability from 2D ultrasound images. The obtained patches are then fed into the networks simultaneously to compute discriminative features. Finally, a softmax is applied to label the input nodule.

optimal epoch by minimizing the testing data of this fold. All the parameters above and the epoch number were used to train the CNNs on other ninefold cross validations.

For SVM performance, there were two important parameters: regularization factor  $C$  and RBF kernel parameter  $\sigma$ . A grid search [43] was applied to tune the two parameters of the SVM in the range  $C \in \{0.1 \times 2^{-4}, 0.1 \times 2^{-3}, \dots, 0.1 \times 2^4, 0.1 \times 2^5\}$  and  $\sigma \in \{0.1 \times 2^{-4}, 0.1 \times 2^{-3}, \dots, 0.1 \times 2^4, 0.1 \times 2^5\}$ . We performed on the testing set of one fold cross-validation, and chose the parameters ( $C = 0.1$ ,  $\sigma = 0.05$ ) that lead to the lowest error rate. Then we used this pair of parameters to the rest folds. The impact of using different values of this two main parameters is shown in Fig. 8. When testing on one parameter, the other kept as the default value. In order to ensure a fair comparison, the parameters of SVMs were all the same in this study: regularization factor  $C = 0.1$  and RBF kernel parameter  $\sigma = 0.05$ .

In addition, all the experiments were carried out on a server having one Intel Xeon E5-2630L 0 Central Processing Unit (CPU) at 2.00 GHz, 32 GB RAM, two Graphics Processing Units (GPUs) with Nvidia Geforce GTx780 Ti at 3 GB and running the 64-bit Windows 8 operating system. Our operation platforms were C++ and Matlab 2014. Our CNN based methods were based on the cuda-convnet package.<sup>1</sup> All trainings of CNNs were performed using the two GPUs for 60 (CNN1) or 100 (CNN2) optimization epoches. The training time was approximately 7.38 h for CNN1 and 15.57 h for CNN2. Testing 128 image patches from one thyroid nodule image on each GPU for classification only took about 0.3–0.6 s. The SVM methods were all based on the LibSVM package.<sup>2</sup>

### 3.2. Performance of different CNN structures

We first studied the impact of different architectures, and the detailed configurations have been described in Table 2. The classification performance of these architectures are reported in terms of

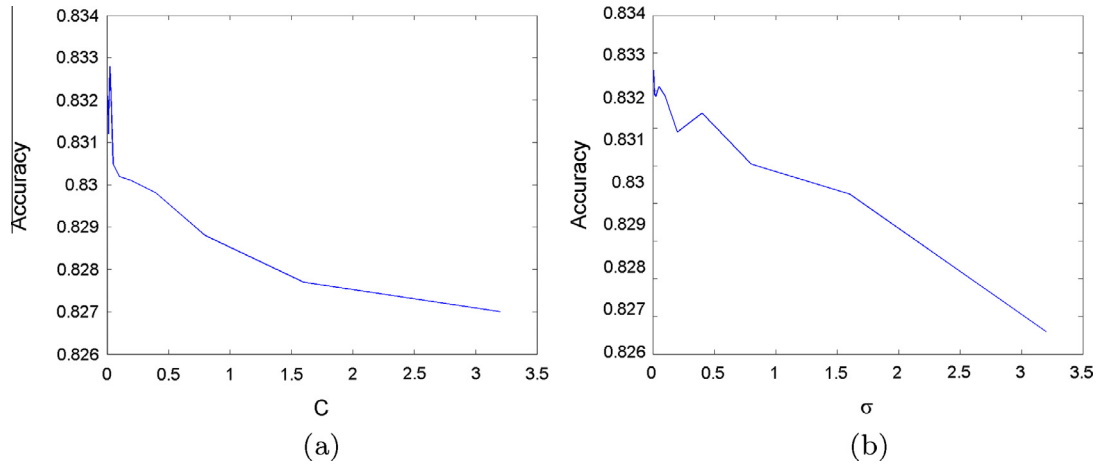
average classification accuracy, sensitivity and specificity, with standard deviation (see Table 4). It can be observed from the results that the predictive performance of the fused CNN method is generally higher. To examine the patterns captured by the CNN methods, we visualized the 96 filters in the first convolutional layer for the two CNNs with an input patch size of  $225 \times 225$  in Fig. 9. These results show different architectures can train different filters. Similar to the observation in [44], although these filters can capture primitive image features such as edges and corners, different filters can capture different image features. In order to show CNN methods can effectively classify thyroid nodules, we used the method in [45] to visualize the feature maps learned by CNN2 (see Fig. 10). The result shows that the features learned by CNN are effective to discriminate malignant thyroid nodules from benign ones. In addition, to examine the impact of multi-view, the single and multi-view testing errors are shown in Table 5. It can be observed that a clear improvement of multi-view compared to single view, demonstrating that the high-level features of the individual CNN based models provide a robust representation of the input images.

### 3.3. Comparison with different methods

In order to provide a comprehensive and quantitative evaluation of feature maps learned by CNN, we compared the performance of SVM (with RBF kernel) based methods with feature maps learned by CNN1, GLCM features, histogram features, respectively. In detail, firstly, we adopted feature maps of the first full-connected (fc1) layers of CNN1 to train SVM to classify thyroid nodules. Principal component analysis (PCA) was performed to reduce the number of the statistical parameters and to calculate a linear combination of the main characteristic features that adequately describe the sample. 23 components of the features generated by PCA and all the feature maps of the fc1 layer were input of SVM to train the SVM classifier, respectively. Then, a commonly used texture feature, GLCM features, were extracted to train SVM classifier. To eliminate the attenuation effect during the transmission and reflection of the ultrasound, the angle was set to  $0^\circ$  and the distance was set to 1. 8 features were extracted from each

<sup>1</sup> <https://code.google.com/p/cuda-convnet/>.

<sup>2</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.



**Fig. 8.** Illustration of parameter analysis of SVM. The average accuracies are evaluated on different values of  $C$  and  $\sigma$ . When testing on one parameter, the other keeps as the default value. The default values for  $C$  and  $\sigma$  are 0.1 and 0.05, respectively.

**Table 4**

Performance measures for the thyroid nodule classification based on different CNNs (Ac: Accuracy; Sn: Sensitivity; Sp: Specificity) are shown. The bold values are the best performance in this study. Numbers format: mean value (standard deviation).

Method	Ac (%)	Sn (%)	Sp (%)
CNN1	81.94 (1.19)	80.57 (1.69)	83.20 (2.32)
CNN2	82.34 (1.01)	81.94 (1.96)	82.82 (2.23)
Fusion of CNN1 and CNN2	<b>83.02</b> (0.72)	<b>82.41</b> (1.35)	<b>84.96</b> (1.85)

GLCM, including autocorrelation (Gactoc), dissimilarity (Gdissi), energy (Geneg), entropy (Gentro), local homogeneity (Glhomo), correlation (Gcorre), contrast (Gcontr) and cluster prominence (Gcprom) [46]. Finally, as a baseline, histogram features were used to train SVM classifier. We computed the histogram of intensity levels, and characterised the texture of thyroid nodules by mean (Hmea), variance (Hvar), skewness (Hske), kurtosis (Hkur), energy (Hene) and entropy (Hent) [47,48]. The results for SVM based methods are presented in terms of average classification accuracy, sensitivity and specificity, with standard deviation, as shown in Table 6. These results show that feature maps learned by CNN are more effective for thyroid nodule classification than GLCM and histogram features are. Moreover, PCA can improve the performance of SVM with feature maps learned by CNN.

In addition, the box-plots in Fig. 11 graphically depicts the classification accuracy of different methods in this study. It can be observed from the results above that CNN based methods outperform other methods in classifying thyroid nodules. Specifically, the fusion of CNN1 and CNN2 can achieve classification accuracy as

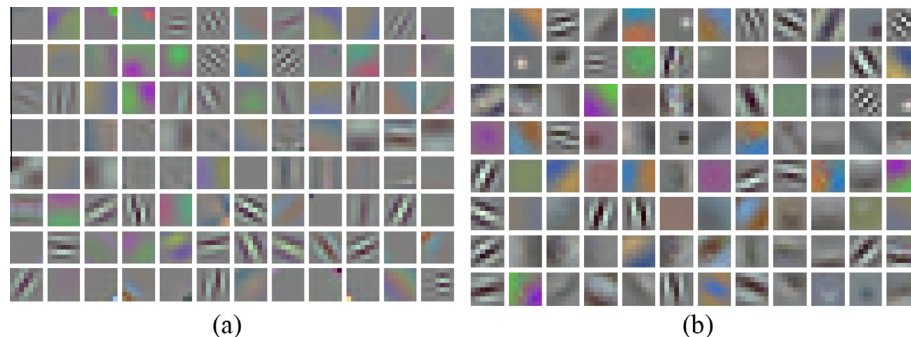
$83.02\% \pm 0.72\%$ , sensitivity as  $82.41\% \pm 1.35\%$ , and specificity as  $84.96\% \pm 1.85\%$ .

#### 3.4. Statistical analysis

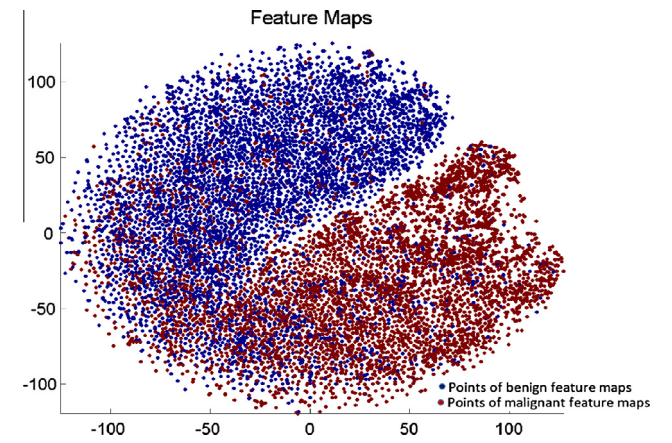
Table 8 presents the summary statistics of the features for thyroid nodule dataset. The last column of this table shows the p-values of the features. The features having p-values below 0.0001 are clinically significant. The diagnostic performance of different methods are evaluated by ROC curves, plotted in Fig. 12. To assess the statistical significance of the performance differences, non-parametric Mann Whitney U test [49] was performed for the values of AUC. 95% confidence interval of AUC are reported in Table 7,  $p < 0.05$  is considered to be statistically significant. As can be seen, not only our CNN based methods are statistically significant, but also significantly outperform SVM with GLCM or histogram features.

#### 4. Discussion

In this study, we address the problem of thyroid nodule diagnosis with CNN based methods. CNNs are a typical type of deep learning based method for differential diagnosis of nodules, which can avoid the potential errors caused by inaccurate image preprocessing results (e.g., boundary segmentation), as well as the classification bias resulting from a less robust feature set. To the best of our knowledge, although there have been some studies on the deep learning based CADx for medical images, it is the first



**Fig. 9.** Visualization of the 96 filters in the first convolutional layer of the different CNN structures with an input patch size  $225 \times 225$  and a filter size  $7 \times 7$ . (a) Shows filters in the first convolutional layer of the CNN1, (b) Shows filters in the first convolutional layer of the CNN2.



**Fig. 10.** Feature maps extracted from CNN2 architecture, visualized in a 2-dimensional space. The positive samples (feature maps of benign thyroid nodules) and negative samples (feature maps of malignant thyroid nodules) are shown in blue and red points in the 2D plane, respectively. Note the clear separation between categories. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 5**  
Errors of a single testing or a multi-view testing used in this study.

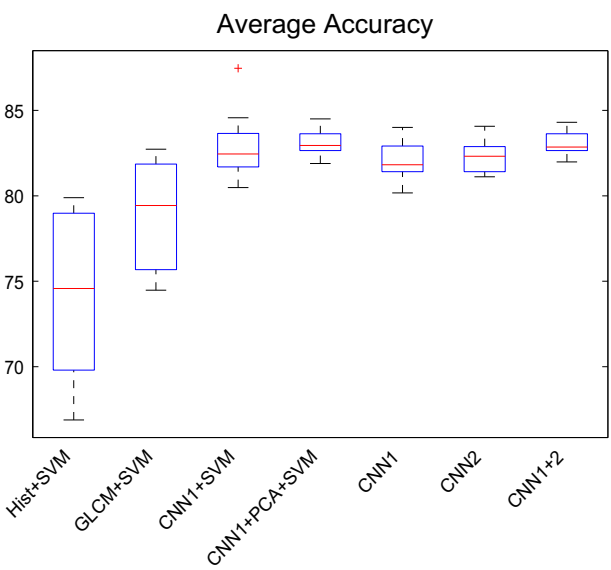
Method	Single error (%)	Multi-view error (%)
CNN1	19.33	18.06
CNN2	18.86	17.66

**Table 6**  
Performance measures for thyroid nodule classification based on SVMs with different features (Ac: Accuracy; Sn: Sensitivity; Sp: Specificity) are shown. The bold values are the best performance in this study. Numbers format: mean value (standard deviation).

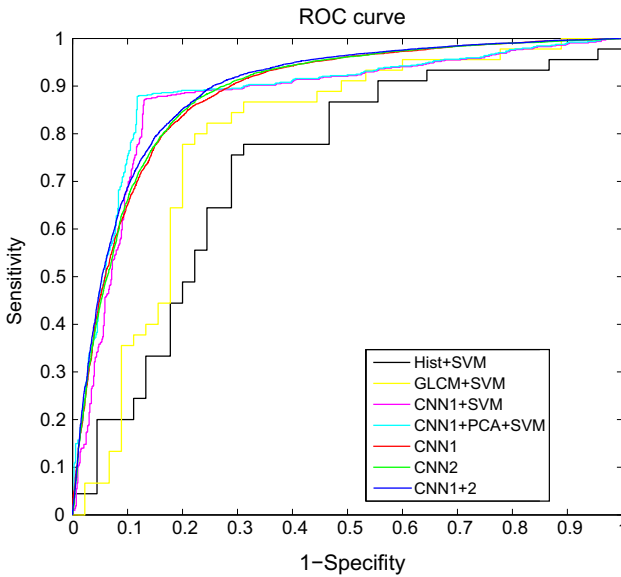
Method	Ac (%)	Sn (%)	Sp (%)
CNN1(fc1)+SVM	82.93 (1.98)	82.06 (1.62)	<b>83.88</b> (2.36)
CNN1(fc1)+PCA + SVM	<b>83.05</b> (0.81)	<b>82.34</b> (1.59)	83.86 (2.26)
GLCM + SVM	78.89 (3.02)	77.78 (3.90)	61.69 (3.03)
Histogram + SVM	65.89 (4.99)	63.78 (4.72)	59.69 (4.22)

attempt to use CNNs for thyroid nodule classification. Compared with the conventional methods [16,50] that use handcrafted features designed from human experience, the CNNs can automatically extract effective features from 2D thyroid nodule ultrasound images without any assumptions about the relevant visual features. With well noise tolerance, CNNs can be suitable to deal with the intrinsically noisy property of ultrasound image data from various sonographics systems. CNNs can also take full advantage of the 2D structure of an input image. Boundary, edge and different texture features can be learned by different convolutional filters, pooling and normalization operations. In addition, CNNs are invariant for geometry transformation, deformation and illumination in a certain degree. Moreover, Fig. 9 shows convolutional layer filters learned by different CNNs are different, so different CNNs can learn different features.

To evaluate the generalization of the proposed method, 15,000 images from different sonographics systems are studied in the experiments. 10-fold cross-validations are conducted to illustrate the efficacy of CNN based methods. As can be seen from Fig. 10, the CNN based methods can effectively classify thyroid nodules. The results are consistent with the fact that CNNs with more convolutional layers and feature maps tend to have a deeper hierarchical structure and more trainable parameters. Thus, these networks are capable of capturing the complex relationship between inputs



**Fig. 11.** Box plots of performance measures for classifying between benign and malignant thyroid nodules. In each box plot, the center red line is the median and the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 12.** ROC curves for performance of different configurations on thyroid nodule classification. Specificity indicates the percentage of malignant nodules correctly identified as malignant, and sensitivity measures the percentage of benign nodules correctly identified as benign.

**Table 7**  
AUC and 95% confidence interval (CI) of the methods used in this study are shown. AUC value format: mean (standard deviation). The bold value is the best AUC in this study.  $p < 0.05$  is considered to be statistically significant.

Method	AUC (%)	CI
CNN1	88.51 (1.96)	[0.8789 0.8904]
CNN2	88.64 (1.79)	[0.8804 0.8917]
Fusion of CNN1 and CNN2	<b>89.30</b> (1.68)	[0.8879 0.8991]
CNN1(fc1)+SVM	89.15 (1.90)	[0.8856 0.8962]
CNN1(fc1)+PCA + SVM	89.22 (1.88)	[0.8860 0.8968]
GLCM + SVM	79.31 (3.22)	[0.6999 0.8572]
Histogram + SVM	65.69 (3.98)	[0.6356 0.8132]



**Table 8**Numbers of various features and mean  $\pm$  standard deviation of the various features for the benign and malignant thyroid nodules.

Feature	Number	Benign	Malignant	p-value
CNN1(fc1)	64	93.4927 $\pm$ 36.4191	72.2919 $\pm$ 24.9439	0.0000
CNN1(fc2)	64	79.9557 $\pm$ 175.5972	290.1176 $\pm$ 103.8371	0.0000
CNN1(fc3)	1	1.0881 $\pm$ 19.5002	−1.0176 $\pm$ 19.5359	0.0000
CNN2(fc1)	4096	22.6523 $\pm$ 53.8742	104.1771 $\pm$ 48.6115	0.0000
CNN2(fc2)	4096	236.9267 $\pm$ 150.1257	11.4805 $\pm$ 113.2010	0.0000
CNN2(fc3)	1	0.2035 $\pm$ 23.3071	−0.1777 $\pm$ 23.2735	0.0000
Gactoc	1	238.0075 $\pm$ 46.9706	242.5355 $\pm$ 36.6721	0.1433
Gdissi	1	10.0284 $\pm$ 1.0690	9.9800 $\pm$ 0.6516	0.3263
Generg	1	0.0015 $\pm$ 0.0015	0.0013 $\pm$ 0.0003	0.0094
Gentro	1	6.6710 $\pm$ 0.2355	6.7217 $\pm$ 0.1092	0.0002
Glhomo	1	0.0962 $\pm$ 0.0175	0.0969 $\pm$ 0.0097	0.0428
Gcorre	1	−0.0082 $\pm$ 0.1327	0.0102 $\pm$ 0.0895	0.0153
Gcontr	1	151.3747 $\pm$ 26.0910	150.2936 $\pm$ 16.2399	0.4864
Gcprom	1	5.1339e + 04 $\pm$ 1.4334e + 04	5.5193e + 04 $\pm$ 1.0694e + 04	0.0000
Hmea	1	2.1616 $\pm$ 2.0406	2.0959 $\pm$ 1.7213	0.3384
Hvar	1	2.1949 $\pm$ 2.7581	2.0579 $\pm$ 1.6662	0.2707
Hske	1	2.4353 $\pm$ 5.6227	2.0494 $\pm$ 1.7080	0.2100
Hkur	1	2.1655 $\pm$ 1.9362	2.1075 $\pm$ 1.7294	0.4664
Hene	1	2.0699 $\pm$ 1.7775	2.1460 $\pm$ 1.8043	0.2991
Hent	1	2.0856 $\pm$ 1.7534	2.0654 $\pm$ 1.5191	0.3756

and outputs. Moreover, the results shown in Table 4 suggest that the fused CNN method is more effective than CNN1 and CNN2, respectively. This is because of the fused configuration is more suitable for effectively extracting multi-level features from thyroid nodules by different convolutional layers.

In addition, to assess the outperformance of CNN learning features based methods over the conventional scheme, GLCM and histogram features based methods are implemented for comparison. The detail of features are shown in Table 8. Our experiment results shown in Table 6 indicate that SVM with features learned by CNN1 is more effective than SVM with GLCM or histogram features. These mean that CNNs can learn effective features to represent the image content associated with different natural thyroid nodules.

To make the classification process more transparent to the end user, TMI, a single-valued indices formed using a combination of the features learned by CNN1 and CNN2 is determined for each thyroid nodule image. Once our proposed methods are evaluated on larger databases, they can be used as adjunct diagnostic tools in daily clinical practice as the techniques do not incur any extra cost of installation. This technique can offer physicians an objective and valuable second opinion, consequently assist them in ascertaining thyroid nodules before FNA and reduce their heavy workload so as to avoid misdiagnosis causes because of excessive fatigue. Moreover, it is easy, fast, and reproducible for a person without medical expertise to diagnose thyroid nodules.

Currently, there are no analytic approaches to design hyperparameters (such as number of layers and units, filter size, and so on) in CNNs and they are mainly obtained empirically, as we perform in this study. Particularly for deep CNNs, it is necessary for more data to help reach a better generalization and reduce the overfitting problem. Our data set is not enough to get a higher accuracy because of the insufficient training samples when exploiting the deep CNNs to represent high-level features. Thus, the higher performance level need to be further studied in the future.

## 5. Conclusion

Thyroid nodule diagnosis based on ultrasound images plays an important role for ultrasound guided needle aspiration. However, it is a challenging task due to low signal-to-noise ratio, low contrast and so on. In this study, we propose a hybrid method to classify thyroid nodules, which is a fusion of two pre-trained CNNs.

The results shown in Tables 4 and 6 demonstrate CNN based methods are new comprehensive ways of addressing the thyroid nodule classification problem. Moreover, our fused CNN based method can improve the classification performance. We recognize that this is a preliminary study, and our dataset cannot be adequately cover or represent the actual population base in the clinical practice. In our experiments, we employed CNNs with a few hidden layers. Recent studies showed that CNNs with many hidden layers yielded very promising performance on visual recognition tasks when appropriate regularization was applied [25]. In the future, we will explore CNNs with many hidden layers as more clinical data become available, so that our method can implement fully automatic detection and classification of thyroid nodules in ultrasonography.

## 6. Conflict of interest

No conflicts of interest.

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 11271323, 91330105, 11401231) and the Zhejiang Provincial Natural Science Foundation of China (Grant No. LZ13A010002). The authors would like to thank PHD Jialin Peng, College of Computer Science and Technology, Huaqiao University, for carefully reading our manuscript.

## References

- [1] National cancer institute. (2016). Thyroid cancer. Information available at, <http://www.cancer.gov/cancertopics/types/thyroid>.
- [2] C. Zhu, T. Zheng, B.A. Kilfoy, X. Han, S. Ma, Y. Ba, et al., A birth cohort analysis of the incidence of papillary thyroid cancer in the United States, 1973–2004, *Thyroid* 19 (10) (2009) 1061–1066.
- [3] J.A. Sipos, Advances in ultrasound for the diagnosis and management of thyroid cancer, *Thyroid* 19 (12) (2009) 1363–1372.
- [4] C.D. Lansford, T.N. Teknos, Evaluation of the thyroid nodule, *Cancer Contr.* 13 (2) (2006) 89.
- [5] C. Vorlander, J. Wolff, S. Saalabian, R.H. Lienenlke, R.A. Wahl, Real-time ultrasound elastography – a noninvasive diagnostic procedure for evaluating dominant thyroid nodules, *Langenbeck's Arch. Surg.* 395 (7) (2010) 865–871.
- [6] Z.W. Baloch, S. Fleisher, V.A. LiVolsi, P.K. Gupta, Diagnosis of follicular neoplasm: a gray zone in thyroid fine-needle aspiration cytology, *Diagn. Cytopathol.* 26 (1) (2002) 41–44.
- [7] T. Rago, M. Scutari, F. Santini, V. Loiacono, P. Piaggi, G. Di Coscio, et al., Real-time elastosonography: useful tool for refining the presurgical diagnosis in

- thyroid nodules with indeterminate or nondiagnostic cytology, *J. Clin. Endocrinol. Metab.* 95 (12) (2010) 5274–5280.
- [8] C.M. Lewis, K.P. Chang, M. Pitman, W.C. Faquin, G.W. Randolph, Thyroid fine-needle aspiration biopsy: variability in reporting, *Thyroid* 19 (7) (2009) 717–723.
  - [9] J.T. Bushberg, J.M. Boone, *The Essential Physics of Medical Imaging*, Lippincott Williams and Wilkins, 2011.
  - [10] T. Chikui, K. Okamura, K. Tokumori, S. Nakamura, M. Shimizu, M. Koga, K. Yoshiura, Quantitative analyses of sonographic images of the parotid gland in patients with Sjogrens syndrome, *Ultras. Med. Biol.* 32 (5) (2006) 617–622.
  - [11] U.R. Acharya, G. Swapna, S.V. Sree, F. Molinari, S. Gupta, R.H. Bardales, et al., A review on ultrasound-based thyroid cancer tissue characterization and automated classification, *Technol. Cancer Res. Treat.* 13 (4) (2014) 289–301.
  - [12] D. Koundal, S. Gupta, S. Singh, Computer-aided diagnosis of thyroid nodule: a review, *Int. J. Comput. Sci. Eng. Surv.* 3 (4) (2012) 67.
  - [13] S. Tsantis, N. Dimitropoulos, D. Cavouras, G. Nikiforidis, Morphological and wavelet features towards sonographic thyroid nodules evaluation, *Comput. Med. Imag. Graph.* 33 (2) (2009) 91–99.
  - [14] J. Ma, S. Luo, M. Dighe, D.J. Lim, Y. Kim, Differential diagnosis of thyroid nodules with ultrasound elastography based on support vector machines, in: *Ultrasonics Symposium (IUS)*, 2010 IEEE, 2010, pp. 1372–1375.
  - [15] N. Singh, A. Jindal, Ultra sonogram images for thyroid segmentation and texture classification in diagnosis of malignant (cancerous) or benign (non-cancerous) nodules, *Int. J. Eng. Innov. Technol.* 1 (2012) 202–206.
  - [16] C.Y. Chang, S.J. Chen, M.F. Tsai, Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images, *Pattern Recogn.* 43 (10) (2010) 3494–3506.
  - [17] U.R. Acharya, S.V. Sree, M.M.R. Krishnan, F. Molinari, R. Garberoglio, J.S. Suri, Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan™ systems, *Ultrasonics* 52 (4) (2012) 508–520.
  - [18] D. Bibicu, L. Moraru, A. Biswas, Thyroid nodule recognition based on feature selection and pixel classification methods, *J. Digit. Imag.* 26 (1) (2013) 119–128.
  - [19] U.R. Acharya, O. Faust, S.V. Sree, F. Molinari, J.S. Suri, ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform, *Comput. Methods Prog. Biomed.* 107 (2) (2012) 233–241.
  - [20] U.R. Acharya, O. Faust, S.V. Sree, F. Molinari, R. Garberoglio, J.S. Suri, Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan™ algorithms, *Technol. Cancer Res. Treat.* 10 (4) (2011) 371–380.
  - [21] U.R. Acharya, S.S. Vinitha, F. Molinari, R. Garberoglio, A. Witkowska, J.S. Suri, Automated benign and malignant thyroid lesion characterization and classification in 3D contrast-enhanced ultrasound, in: *Engineering in Medicine and Biology Society (EMBC)*, Annual International Conference of the IEEE, 2012, pp. 452–455.
  - [22] J. Ding, H. Cheng, C. Ning, J. Huang, Y. Zhang, Quantitative measurement for thyroid cancer characterization based on elastography, *J. Ultrasound Med.* 30 (9) (2011) 1259–1266.
  - [23] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cybern.* 36 (4) (1980) 193–202.
  - [24] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
  - [25] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inform. Process. Syst.* (2012) 1097–1105.
  - [26] B.B. Le Cun, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Handwritten digit recognition with a back-propagation network, in: *Advances in Neural Information Processing Systems*, 1990.
  - [27] P.Y. Simard, D. Steinkraus, J.C. Platt, Best practices for convolutional neural networks applied to visual document analysis, in: *International Conference on Document Analysis and Recognition*, IEEE Computer Society, 2003, p. 958.
  - [28] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Unsupervised learning of hierarchical representations with convolutional deep belief networks, *Commun. ACM* 54 (10) (2011) 95–103.
  - [29] D. Ciresan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3642–3649.
  - [30] S.J. Nowlan, G.E. Hinton, Simplifying neural networks by soft weight-sharing, *Neural Comput.* 4 (4) (1992) 473–493.
  - [31] D.J. MacKay, Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks, *Netw.: Comput. Neural Syst.* 6 (3) (1995) 469–505.
  - [32] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, 2012. Available from: <1207.0580>.
  - [33] L. Wan, M. Zeiler, S. Zhang, Y.L. Cun, R. Fergus, Regularization of neural networks using dropconnect, in: *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013, pp. 1058–1066.
  - [34] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: *Advances in Neural Information Processing Systems*, 2012, pp. 2843–2851.
  - [35] D.C. Cirecan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Springer, Berlin Heidelberg, 2013, pp. 411–418.
  - [36] T. Schlegl, J. Ofner, G. Langs, Unsupervised pre-training across image domains improves lung tissue classification, in: *Medical Computer Vision: Algorithms for Big Data*, Springer International Publishing, 2014, pp. 82–93.
  - [37] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian, Multi-scale convolutional neural networks for lung nodule classification, in: *Information Processing in Medical Imaging*, Springer International Publishing, 2015, pp. 588–599.
  - [38] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
  - [39] D. Erhan, P.A. Manzagol, Y. Bengio, P. Vincent, The difficulty of training deep architectures and the effect of unsupervised pre-training, in: *International Conference on Artificial Intelligence and Statistics*, 2009, pp. 153–160.
  - [40] O. Russakovsky, J. Deng, H. Su, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
  - [41] F. Wu, P. Hu, D. Kong, Flip-rotate-pooling convolution and split dropout on convolution neural networks for image classification, 2015. Available from: <1507.08754>.
  - [42] T.J. Downey Jr., D.J. Meyer, R.K. Price, E.L. Spitznagel, Using the receiver operating characteristic to assess the performance of neural classifiers, in: *IEEE Neural Networks*, International Joint Conference, 1999, pp. 3642–3646.
  - [43] Y.W. Chen, C.J. Lin, Combining SVMs with various feature selection strategies, in: *Feature Extraction*, Springer, Berlin Heidelberg, 2006, pp. 315–324.
  - [44] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: *Computer Vision – ECCV 2014*, 2014, pp. 818–833.
  - [45] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2605) (2008) 2579–2605.
  - [46] L.K. Soh, C. Tsatsoulis, Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, *IEEE Trans. Geosci. Rem. Sens.* 37 (2) (1999) 780–795.
  - [47] F.S. Cohen, Z. Fan, M.A. Patel, Classification of rotated and scaled textured images using Gaussian Markov random field models, *IEEE Trans. Pattern Anal. Mach. Intell.* (2) (1991) 192–202.
  - [48] A. Materka, M. Strzelecki, *Texture Analysis Methods – A Review*, Technical University of Lodz, Institute of Electronics, COST B11 Report, Brussels, 1998, 9–11.
  - [49] G. Qin, L. Hotilovac, Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test, *Stat. Methods Med. Res.* 17 (2) (2008) 207–221.
  - [50] S. Tsantis, N. Dimitropoulos, D. Cavouras, G. Nikiforidis, A hybrid multi-scale model for thyroid nodule boundary detection on ultrasound images, *Comput. Methods Prog. Biomed.* 84 (2) (2006) 86–98.