A Hybrid Deep Learning and Handcrafted Features based Approach for Thyroid Nodule Classification in Ultrasound Images

Jiahao Xie¹, Lehang Guo², Chongke Zhao², Xiaolong Li², Ye Luo¹*, Jianwei Lu¹*
¹Tongji University, Shanghai, China
²Shanghai Tenth People's Hospital, Shanghai, China
yeluo@tongji.edu.cn

Abstract. With the increasing incidence rate of thyroid cancer, the diagnosis of thyroid nodules has become an important task. In this paper, we designed a deep neural network (DNN) to classify whether a thyroid nodule is benign or malignant, and proposed a structure which combines local binary pattern (LBP) with deep learning. Our method mitigates the effects of overfitting in medical image diagnosis tasks. With well-designed transfer learning, we achieve an accuracy of 85% on our own ultrasound thyroid dataset. To ensure the reliability of our experiments, all examples are estimated by experts in Shanghai Tenth People's Hospital using fine needle analysis (FNA), which is a gold standard for thyroid nodules diagnosis. The experimental results show that combinations of the traditional medial image features can help the deep learning network get more semantic information from low-level inputs.

1. Introduction
The incidence of Thyroid cancer continues rising around the world. More than 2.4 times increase during 29 years since 1973, form 3.6 per 100,000 in 1973 to 8.7 per 100,000 in 2002 [2]. [1] estimated that more than 62,000 new cases occurred among people in 2015, and about 300,000 new case in 2012 [8]. Thyroid cancer becomes the most threatening disease in the world, during 2008 to 2012, and about 0.2~0.4 per 100 000 men and 0.2~0.6 per 100 000 women died because of thyroid cancer [8].

To decrease the damage and the treatment costs, computer-aided diagnosis (CAD) has been developed to help doctors discriminate nodules from benign or malignancy. Some CAD diagnoses diseases by measuring specific hormone content in the human body [11], while others help doctors find the nodule [15] or compute the probability of malignant in medical images [9, 12, 16]. Many researchers use machine learning to learn the patterns of certain features, however, the feature is very hard to seek out, and some useful features are difficult to extract manually. In recent years, deep learning gives models the ability to extract feature automatically through the back propagation (BP) algorithm. As we all know that deep learning based methods are heavily relied on the big amount of data. Enough data and good algorithm make it possible to train a model which can classify thyroid nodules from benign or malignant effectively. However, thyroid image data is very rare and expensive in the field of medical image processing. Even though some hospitals start to collect data, few of them release the dataset and available dataset for deep neural network training is always in shortage. Moreover, due to the low quality of ultrasound image, thyroid nodules are usually difficult to recognize even using DNN. Some examples are visualized in Fig.1.
Fig. 1 Examples in our dataset. The four nodules on the left is benign, and the four nodules on the right is malignant. Those nodules are very similar and hard to discriminate.

Meanwhile, the roles of manually extracted image features and the ability of CNN on learning more general and invariant feature from fine processed images (e.g. feature maps of the first several layers) are under-estimated in existing work. On one hand, manually extracted image features such as local binary pattern (LBP) contain rich texture information and are robust to different exposure which commonly occurs to ultrasound images. On the other hand, in deep learning, the convolution in the first layer extracts low-level features like edge, contrast, color or some other simple information hence the appearance of the feature map at the first layer usually looks similar. However, high-level information for various specific tasks can be well extracted at the following deep layers. In other words, CNN has the ability to learn more general and invariant feature from fine processed images but this ability is under-estimated by most of existing research work.

In this paper, we proposed a method to combine high-level features extracted by a two-stream deep neural network (DNN) on the source image and its corresponding LBP feature map, respectively. By this way, we can make full use of the low-level features (i.e. LBP) and the ability of CNN to extract high-level features from the first-layer like features (i.e. LBP). Moreover, it helps a lot for training a robust model especially when the training data is small. Specifically, for each image, its LBP feature is extracted beforehand. Then, we use a pair of LBP and its source image as the input, and adopt two Resnets as the two branches of the two-stream DNN. Next, we concatenate the features from the two streams in the fully connected layer to get the final result. To validate the effectiveness of the proposed method, we conduct the experiments on a small but challenging dataset, which is collected by professional doctors and validated by the gold standard. Also, we use transfer learning to initialize our model and then training it by fine-tuning.

The main contributions of our method are two folds: (1) Instead of using low-level features for classification directly, we design a network to learn deep features on extracted low-level features. (2) We propose a method that combines high-level features extracted from the source image and its corresponding LBP feature map, and the accuracy boosts about 4.5% through this method.

Fig. 2 Source image and LBP image: left is the cropped source image, and the right is corresponding LBP image. We can see that the texture of the LBP image is more obvious.

2. Relate Work

2.1. Deep learning models

Since Alex built AlexNet [7] using the convolutional neural network (ConvNet) and won the challenge of ImageNet in 2012, more and more researchers dived into improving the structure of ConvNet. VGG [14] built deeper ConvNet through stacking sample structure and down-sampling by pooling operation. ResNet [3] proposed a novel structure named residual block, which is implement by skip connections.
Inception [6, 17-19] assumed that the various size of filters focus on different parts of a feature map, and gathered different sizes of filters in ConvNets can boost the performance. MobileNet [4, 5, 13] decoupled a convolutional operation to a depth-wise convolution and a point-wise convolution, which can decrease calculation cost while can remain powerful expression. All of those researches validate that ConvNets is powerful and can be robust to classify medical images successfully. In this paper, we use ResNet as our baseline.

2.2. Recent Work about Thyroid Nodule
As deep learning developed so fast, more and more works about thyroid nodule diagnose focus on ConvNets. Recent work [9] trained a ConvNet by combining the source image and the segmentation of thyroid nodule, and got 83.02% ± 0.72% accuracy, which was 2% higher than only using source image. [12] used particle swarm optimization to extract features from higher-order spectral (HOS) of image and then classified features through support vector machine (SVM). They achieved a maximum accuracy of 97.71%. [16] found a way to combine conventional features such as LBP, HOG, SIFT with convolutional features, and the highest accuracy they got is 93.1%. Their method is similar to ours, but they used statistic histogram directly. On the contrary, the conventional features we used are first sent into a ConvNet, and the hidden semantic information of conventional features extracted makes our model more robust and effective to the thyroid nodule classification. It is worth mentioning that numerically the aforementioned methods got higher accuracy than ours, but we use totally different dataset. Our dataset is more rigorous and challenging, and many difficult examples are in the dataset. Moreover, in this paper we propose a novel method to combine conventional features and deep features via ConvNets, which can boost the classification performance and generalize the ConvNet model to analyze on a small ultrasound image dataset.

3. The Proposed Method
We designed a network structure to learn high-level features from low-level features. As shown in Fig.2, first we extract the low-level feature i.e. LBP from the input image, and then train both LBP image and the source image in two parallel ConvNets (i.e. ResNet). Different from other methods, we propose to learn deep features from the LBP feature map by a ConvNet together with learning from the source image. LBP there actually acts as a filter with good generalization and can help our model overcome the risk of over-fitting caused by data shortage.

3.1. Local Binary Pattern (LBP)
LBP algorithm generates an 8-bit binary number in every pixel by using other pixels around the center. LBP algorithm compute values under grayscale image, and the formula is:

$$LBP(x_i) = \sum_{p=0}^{7} 2^{7-p} \cdot \text{sign}(i_p - i_c)$$

where $i_c$ means the center pixel, and $i_p$ is the pth pixel around the center pixel starting from the left-up clock-wisely. Advanced LBP algorithm such as the circle LBP [10] using a circle around the center to compute the feature value of the center pixel. In this paper, we focus on demonstrating the improvement by adopting LBP features to the ConvNet model, hence the simplest algorithm of LBP is used. After applying the LBP algorithm to the source image (e.g. Fig.2 left), the source image can be converted to a LBP image as shown in Fig.2 right. The reason why using LBP to extract low-level features is that LBP contains rich texture features, and previous work once verified the superior performance in face recognition tasks. On the other hand, LBP is a type of visual descriptor used for classification. It encodes one pixel by comparing the pixel around it. This method is robust to different exposure, which is a common problem when taking ultrasound images.
3.2. Network Structure
The structure of our model is shown in Fig.3. Firstly, we generate a LBP image for every input image before training. Second, we extract deep features from the two images (i.e. the LBP image and the source image) using two individual ReaNet34. From each image, we extract 512 features, then concatenate them together. Finally, we use a 1024 * 2 fc layer and a softmax function to get the final probability to identify the class labels. For the whole network, we use pretrained weights from ImageNet challenge to initialize our ResNet34 before training.

3.3. Training Process
The training process is critical, especially in the region of medical image.

(1) Extract regions of interest(ROI). The full image of thyroid contains blood vessel, muscle, and bone. It's hard to classify thyroid nodules through entire image, so we crop the ROI in the image, and use thyroid nodules to train our model.

(2) Heavy data augmentation. The most severe problem in the domain of medical image analysis is the lack of data. Labeling medical image is very expensive, and the alternative method is data augmentation. Choosing adaptive transformation is important in that thyroid nodule has different sizes and doctors can diagnose cancer from parts of the nodule. Hence, cropping the image into tiles is rational. Due to different view angle doesn't change the property of nodule, flipping and rotation can get different characteristics of one image. Different environments and operations also cause different strengths of ultrasounds. We apply several types of transformations to generate different images, such as changing exposure, flipping, rotation, cropping, changing saturability, adding random noise and so on.

(3) Do augmentation respectively. Source image and LBP pattern should do random transform respectively. This is because if a transform enlarges the input space by N times, we combine two individual transforms to make the input space N^2 time larger. Enough data helps models easier to capture special pattern in tremendous space.

![Fig. 3 Structure of our proposed method. Both the source image and the corresponding LBP image are input to the network together.](image)

![Fig. 4 In the vertical image(left) we can see four signals marking the location of the nodule, and the smooth area around the nodule is tissue, which can be distinguished from the nodule, and the horizontal stripes upon the nodule are muscles, which are also different to the nodule. In the horizontal](image)
image, we can see vessels in the left and the throat in the right. Both vessel and throat are similar to nodules and hard to distinguish.

4. Experiment

We choose exactly one ultrasound image from one patient. Finally, we get 623 images contains 305 benign thyroid nodule images and 318 malignant thyroid nodule images. We split the 623 images into three parts, 85% as train data, 5% as validation data, and 10% as test data. Our method is based on ultrasound images which are labeled by fine-needle aspiration (FNA), and all images are collected from the same type of machines from Shanghai Tenth People's hospital. All cases have been classified to benign or malignant by FNA biopsy or surgery.

We first explore the difference between vertical and horizontal ultrasound images as shown in Fig.4. Different view angle shows various appearance about nodules. We expect that combine vertical and horizontal images would boost the performance of the ConvNet. But the experiment in Table.1 shows that only using vertical images achieves the highest accuracy. We think the reason is that vessel around the nodules effects the discrimination of thyroid nodules. It hurts more than benefiting the models to learn general information from the data.

To validate our proposed method outperform the conventional method and a single ConvNet, we design three experiments to compare the performance among our proposed method with others. The first method is the baseline, which only uses source images. The second method is to combine LBP histogram with source image by concatenate them in fully connect layer, and the results show F1 score has dropped down about 10%, which means LBP histogram is hard to learn and easy to overfit. The last method is our proposed method, which gets a 4.5% boost of F1 score as shown in Table.2. The result further validates that we can learn high-level information from the conventional features such as LBP, and the extracted high-level feature learned from the LBP makes the model to be more generalized and can be easier applied to small dataset especially in medical image processing domain.

| Image component         | Precisions | Recall  | F1 score |
|-------------------------|------------|---------|----------|
| Horizontal              | 79.49%     | 75.61%  | 77.50%   |
| Vertical                | 80.72%     | 81.70%  | 81.21%   |
| Horizontal + Vertical   | 78.87%     | 75.67%  | 77.24%   |

Table 1 Performance of models that trained by different aspect of nodules

| Method                                | Precisions | Recall | F1 score |
|---------------------------------------|------------|--------|----------|
| Source image                          | 80.72%     | 81.70% | 81.21%   |
| LBP histogram + Source image          | 68.54%     | 74.39% | 71.35%   |
| LBP image + Source image              | 80.64%     | 91.46% | 85.71%   |

Table 2 Performance of models that trained by different methods

5. Conclusion

In this paper, we proposed a novel method that combines LBP features and the source image, which boosts the model performance in a relatively small and challenging dataset. Additionally, our method validates that traditional extracted low-level feature can be used as the input of a ConvNet for further high-level feature extraction, and this can also generalize the proposed network to a small dataset on other deep learning tasks. In the future, other conventional image features and different ways of feature fusion methods can be explored.

References

[1] Cabanillas, M.E., Mcfadden, D.G., Durante, C.: Thyroid cancer.388(10061), 2783(2016).
[2] Davies, L., Welch, H.G.: Increasing incidence of thyroid cancer in the united states,1973-2002. Jama295(18), 2164.
[3] He, K., Zhang, X., Ren, S., Jian, S.: Deep residual learning for image recognition. In: IEEE
Conference on Computer Vision and Pattern Recognition (2016).

[4] Howard, A., Pang, R., Adam, H., Le, Q.V., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., Chu, G., Vasudevan, V., Zhu, Y.: Searching for mobilenetv3. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019. pp. 1314-1324. IEEE (2019).

[5] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR abs/1704.04861 (2017).

[6] Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift (2015).

[7] Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems (2012).

[8] La Vecchia, C., Malvezzi, M., Bosetti, C., Garavello, W., Bertuccio, P., Levi, F., Negri, E.: Thyroid cancer mortality and incidence: A global overview. International Journal of Cancer 136(9), 2187-2195.

[9] Ma, J., Wu, F., Zhu, J., Xu, D., Kong, D.: A pre-trained convolutional neural network based method for thyroid nodule diagnosis. Ultrascience 73, 221-230 (2016).

[10] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation-invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971-987 (2002).

[11] Ozyilmaz, L., Yildirim, T.: Diagnosis of thyroid disease using artificial neural network methods. In: International Conference on Neural Information Processing (2002).

[12] Raghavendra, U., Gudigar, A., Maithri, M., Gertych, A., Meiburger, K.M., Yeong, C.H., Madla, C., Kongmebhol, P., Molinari, F., Ng, K.H.: Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images. Computers in Biology and Medicine p. S0010482518300246

[13] Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 4510-4520. IEEE Computer Society (2018).

[14] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014).

[15] Song, W., Shuai, L., Ji, L., Hong, Q., Bo, Z., Zhang, S., Aimin, H.: Multi-task cascade convolutional neural networks for automatic thyroid nodule detection and recognition. IEEE Journal of Biomedical and Health Informatics pp. 1-1.

[16] Sun, W., Liu, T., Xie, S., Yu, J., Niu, L., Sun, W.: Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features (2017).

[17] Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning.

[18] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions.

[19] Szegedy, C., V.V.I.S.S.J., Wojna, Z.: Rethinking the inception architecture for com-puter vision. In: IEEE Conference on Computer Vision and Pattern Recognition (2016).