Automatic detection of thyroid nodules with ultrasound images: Basing on semi-supervised learning

Qingsong Wang¹, Jie Zheng², Hui Yu¹✉, Jingqi Zhang³ and Jie Zhang⁴

¹ Tianjin International Engineering Institute, Tianjin 300072, China
² Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin 300072, China
³ Graduate School of Tianjin Medical University, Tianjin 300050, China
⁴ Tianjin Medical University General Hospital, Tianjin 300052, China

*Corresponding author. Email: yuhui@tju.edu.cn

Abstract. Ultrasonography is one of the most common method for the diagnosis of thyroid nodules clinically. But the lack of experienced radiologists has become the biggest problem which reduce the accuracy and effectiveness of thyroid nodules' detection. What's more, although artificial intelligence (AI) has been widely used to solve the problem, the performance is limited by the size of the database. In this paper, an improved U-Net is designed for the segmentation of thyroid nodules, following a fully convolution network for the classification of thyroid nodules. During the training, pseudo-label is used for semi-supervised learning. The whole work is trained in 2032 ultrasound images (1000 nodules) and tested in 400 ultrasound images (200 nodules). The average dice coefficient is 89.7% in the test set for segmentation, while the precision and recall is 93.2% and 96% separately for classification. These results are higher than the method which only use the images with true labels, proving the effectiveness of pseudo-label.

Keywords: thyroid nodules, pseudo-label, ultrasound images.

1. Introduction
Thyroid nodules are very common diseases of the human endocrine system. It is pointed out in the literature that the prevalence of thyroid nodules in the United States accounts for 27.3% [1], and the incidence is increasing every year [2]. In clinical diagnosis, ultrasonography is one of the most widely used method for early detection of thyroid nodules on account of its advantages of real-time dynamic multi-section observation, non-radiation, and non-invasive inspection [3,4]. However, the diagnosis’s accuracy is easily influenced by the experience of radiologists and their working status. What’s more, the lack of experienced physicians has become more and more serious with the increase of patients.

In recent years, computer aided diagnosis (CAD) has attract great attention of scholars in order to reduce the stress of radiologists. V. Kumar [5] proposed a CNN based method which can automatically segment thyroid nodules and glands but the accuracy is limited. L. Wang [6] designed a deep learning model for automatic diagnosis for thyroid nodules in ultrasound images and got an excellent result.

However, the data sets of the above works are all small. And the annotations for the boundary of thyroid nodules will take physicians a lot of time, making it hard to get a big data set. In 2013, pseudo-
label [7] is proposed as a method of self-training. The method is used in the MINST dataset and showed comparable level with supervised methods.

In this work, an improved U-Net is designed for the segmentation of thyroid nodules. Then a fully convolutional ResNet50 is used to classify benign and malignant thyroid nodules. Our main contribution is as followed:
1. We improved the structure of U-Net by adding ResNet34 for feature extraction and replacing the Softmax layer to global average pooling. The results proved the effectiveness of these improvements.
2. To solve the problem of insufficient samples, a semi-supervised learning method, pseudo-label, was applied during training.
3. The classification model, ResNet50, was improved to a fully convolutional model in order to meet the inputs with different sizes.

The proposed method was trained in the training set with 2032 images and evaluated in the test set with 400 images. For segmentation, the average dice coefficient is 89.7% in the test set. For classification, the precision is 93.2%, while the recall is 96%.

The remain of this paper is organized as followed: Section 2 gives the detail of the proposed method. Section 3 explains the dataset and Section 4 shows the results of comparative experiments. In Section 5, a concise conclusion is given.

2. Methods

The whole method for automatic detection of thyroid nodules is shown in Fig. 1. In Fig.1 (a), firstly, ultrasound images with correct annotations are entered into the segmentation model and then ultrasound images with no annotations are entered into the model to get pseudo-label for the next training. Secondly, the classification model is trained in the similar thought. Fig.1 (b) gives the test procession.

Figure 1. The whole method for automatic detection of thyroid nodules. (a) is for training and (b) is for classification.

2.1. Segmentation and classification model

Since U-Net [8] has been greatly used and applied in the segmentation of medical images, the basic structure of our model is U-Net. The details of the proposed segmentation model are shown in Fig. 2.
Figure 2. The structure of the improved U-Net for segmentation.

The whole model follows down-sampling to up-sampling and can be divided into three parts: feature extraction (down-sampling), image restoration (up-sampling) and skip connection. For feature extraction, ResNet34 is used and divided into 4 blocks. The details of these 4 blocks are shown in the Table 1. After each convolution operation, ReLU is used for activation. For image restoration, up-sampling is used to restore feature maps of different scales. However, if only the last feature maps are restored, there will be a lot of loss of feature. So skip connection is added in each scale to reduce feature loss. For example, the feature map, which is calculated by block1, is added to the corresponding features during the up-sampling. After up-sampling, global average pooling is used to classify each pixel of the images.

Table 1. The details of 4 blocks

| Blocks | Operations |
|--------|------------|
| Block1 | $\begin{align*} &\text{conv}=3 \times 3, \text{ channel}=128 \\
| Block2 | $\begin{align*} &\text{conv}=3 \times 3, \text{ channel}=128 \\
| Block3 | $\begin{align*} &\text{conv}=3 \times 3, \text{ channel}=256 \\
| Block4 | $\begin{align*} &\text{conv}=3 \times 3, \text{ channel}=256 \\

The difficulty for classifying benign and malignant thyroid nodules is the different sizes of thyroid nodules, which makes it impossible to use original classification models with fully connected layers. So a fully convolutional network is designed on the basis of ResNet50 [9]. Firstly, a $7 \times 7$ kernel is used to activate images in a relatively large vision. Then the following model is divided into 4 blocks. In each block, $1 \times 1$ kernels and $3 \times 3$ kernels are used. The improvement is that the Softmax layers are replaced by two global average pooling layers, which make the model possible to get inputs with different sizes. Then batch normalization (BN) is used in order to reduce over fitting. As for the activation function, leaky-ReLU is applied. The advantage of leaky-ReLU is that it can avoid the death of kernel.
2.2. Semi-supervised learning by pseudo-label
It is usually the insufficient samples that make it hard to get a credible result of automatic detection of medical images. So we managed to use semi-supervised learning method instead of traditional supervised learning method. The models are firstly trained by the images which have labels by experienced radiologists. After one epoch, the images without labels are entered to the trained models and pseudo-label is get. Then before the next epoch’s training, these images with pseudo-label are mixed with images which have real labels. The pseudo-labels are updated during the training progress. Since the model is trained in the direction of a better one, these pseudo-labels are more and more real.

However, traditional loss function is not possible in this training way because in the beginning of the training, the pseudo-labels are unreliable. So a new loss function is applied. For segmentation, a new dice loss function is calculated by:

\[
s = \begin{cases} 
1 \times \frac{2(y \cap z') + 1}{y \cup y' + 1} \\
1 \times \frac{2(z \cap z') + 1}{z \cup z' + 1}
\end{cases} 
\]

where \(s\) is the value of dice loss, \(y\) is the output of the model when the input is images with real labels, \(z\) is the output of the model when the input is images with pseudo-labels. \(n\) is the number of the epoch.

For classification, we choose cross-entropy function and improved it in the similiar way. The equation is:

\[
s = \begin{cases} 
1 \times (y \times \log_e y' + (1 - y) \times \log_e (1 - y')) \\
1 \times (z \times \log_e z' + (1 - z) \times \log_e (1 - z'))
\end{cases} 
\]

where \(s\) is the value of cross-entropy loss, \(y\) is real label and \(z\) is pseudo-label.

3. Dataset and training environments
The ultrasound images are collected by the Tianjin Medical University General Hospital. There are 1000 nodules in the train set, which have 2032 images, and the test set has 200 nodules with 400 images. In the train set, 1000 images are annotated by two experienced radiologists while the remaining are not annotated. All 400 images of the test set are annotated. The type of the nodules is judged according to pathological results. There are 427 benign nodules and 573 malignant nodules in the training set. The test set has 100 benign nodules and 100 malignant nodules.

4. Results
4.1. Segmentation results
The trained model is tested on the test set. Several segmentation results are shown in Fig. 3. According to the results, the proposed model can segment thyroid nodules very accurately. For quantitative evaluation, the results are shown in Table 2. According to Table 2, the dice coefficient is 89.7% and the sensitivity is 88.3%, meaning that the model can segment thyroid nodules correctly. The false positive rate is only 8.8% and the false negative rate is only 11.7%. all these parameters are beyond the existing methods or equal to these methods.

To prove the effectiveness of the semi-supervised method, a comparative experiment is designed and only images with real labels are used for training. The results are shown in Table 2. According to the result, our proposed training method exceed the supervised method. There is 4.2% improvement in dice coefficient and the false positive rate reduces from 12.5% to 8.8%.
Table 2. Performance parameters of segmentation.

| Parameters      | Semi-supervised method | Supervised method |
|-----------------|------------------------|-------------------|
| dice            | 89.7%                  | 85.5%             |
| sensitivity     | 88.3%                  | 83.6%             |
| false positive  | 8.8%                   | 12.5%             |
| false negative  | 11.7%                  | 16.4%             |

Figure 3. The examples for the results of the segmentation.

4.2. Classification results

The results for the classification of malignant and benign nodules are shown in Table 3. Table 3 is the confusion matrix of the test set. Assume that malignant nodules are positive samples while benign nodules are negative samples. According to the confusion matrix, the precision is 93.2% and the recall is 96%. The quantitative index proves that the classification model can distinguish benign and malignant nodules correctly.

Table 3. The confusion matrix of the test set.

| Confusion matrix | Prediction value |
|------------------|------------------|
| malignant        | 96               |
| benign           | 4                |
| malignant        | 7                |
| benign           | 93               |

5. Conclusion

In this paper, we proposed an improved U-Net model and a fully convolutional ResNet50 for automatic segmentation and classification of thyroid nodules in ultrasound images. To solve the problem of insufficient samples, a semi-supervised method, pseudo-label, is applied during the training progress while the dice loss function and the cross-entropy loss function are both improved in order to make the training method effective. The test results show that our method can segment and classify thyroid nodules very correctly, which has great significance to reduce the pressure of radiologists and assist them detect nodules quickly and accurately.
Acknowledgement
The author would thank the respiratory doctors of Tianjin Medical University General Hospital for their help and medical guidance in the process of clinical data collection. This project is sponsored by National Key Research and Development Project of China (2019YFC0119402) and Major science and technology projects in Tianjin, China (18ZXZNSY00240).

References
[1] Brander, A. , Viikinkoski, P. , Nickels, J. , & Kivisaari, L. . (1991). Thyroid gland: us screening in a random adult population. Radiology, 181(3), 683.
[2] Carolina, Ferraz, Markus, Eszlinger, Ralf, & Paschke. (2011). Current state and future perspective of molecular diagnosis of fine-needle aspiration biopsy of thyroid nodules. The Journal of clinical endocrinology and metabolism, 96(7), 2016-26.
[3] Bushberg, J. T. , Anthony, S. J. , Leidholdt, E. M. , Boone, J. M. , & Goldschmidt, E. J. . (2003). The essential physics of medical imaging. Medical Physics.
[4] Toru Chikui and Kazutoshi Okamura and Kenji Tokumori and Seiji Nakamura and Mayumi Shimizu and Makoto Koga and Kazunori Yoshiura. (2006). Quantitative analyses of sonographic images of the parotid gland in patients with sjögren’s syndrome. Ultrasound in Medicine & Biology.
[5] Kumar, V. , Webb, J. , Gregory, A. , Meixner, D. D. , & Alizad, A. . (2020). Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning. IEEE Access, PP(99), 1-1.
[6] Wang, L. , Zhang, L. , Zhu, M. , Qi, X. , & Yi, Z. . (2020). Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks. Medical Image Analysis, 61, 101665.
[7] Lee, D. H. (2013, June). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML (Vol. 3, No. 2).
[8] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).