Automatic Thyroid Ultrasound Image Detection and Classification with Priori Knowledge

Mengdie Shi
School of Computer Science and Technology, Harbin
Institute of Technology, Harbin, China
smd2093@163.com

Jianrui Ding*
School of Computer Science and Technology, Harbin
Institute of Technology, Harbin, China
jrding@hit.edu.cn

Shili Zhao
School of Computer Science and Technology, Harbin
Institute of Technology, Harbin, China
854791028@qq.com

Zicheng Huang
School of Computer Science and Technology, Harbin
Institute of Technology, Harbin, China
847445563@qq.com

ABSTRACT

Medical ultrasonic imaging technology is currently the preferred method to detect and diagnose benign and malignant thyroid nodules, which is widely used because of their low cost and non-invasive damage to patients. But automatic lesion detection and classification on thyroid ultrasound image is quite challenging due to the poor image quality. To solve the problem, based on popular Faster R-CNN network for natural image detection, a ResAt-Faster R-CNN model was proposed in the paper according to the characteristics of thyroid ultrasound image, the residual module and attention mechanism. The medical prior knowledges such as location and distribution information are further introduced to constrain the model to reduce the interference of surrounding tissues. The experimental results demonstrated that our proposed method was effective in the discrimination of thyroid nodules.

CCS CONCEPTS

• **Computer systems organization**; • **Architectures**;

KEYWORDS

Thyroid ultrasound image, Faster R-CNN, Medical priori- knowledge

1 INTRODUCTION

Thyroid nodule disease is a common clinical disease with a higher incidence in young population [1]. If a patient can get reasonable diagnosis and treatment in time in the early stage, it can reduce the risk of further deterioration of thyroid nodules.

Ultrasonic technology is widely used in various clinical medical diagnostic tasks due to its low detection cost, real-time imaging and non-invasive damage to patients. However, the ultrasound image has low contrast, high labeling cost and different degrees of calcification and noise in the image [2]. Therefore, the diagnosis of thyroid nodules by ultrasound images is facing great challenges.

With the development of deep learning in the field of medical imaging, computer-aided diagnosis methods have rapidly emerged [3]. In particular, convolutional neural network has demonstrated strong feature extraction capabilities in the fields of detection and classification [4]. It is of great clinical value to use deep learning model to extract the features of thyroid ultrasound images and predict the development of the patient's condition. In recent years, some researchers have applied deep convolution model to automatic detection and segmentation of thyroid ultrasound images and achieved good results [5-13]. Due to the characteristics of ultrasound image, the direct application of the existing deep learning model will make the data set distributed in the space sparsely and lead to inaccurate positioning without any clinical knowledge.

To solve the above problem, this paper proposed the ResAt-Faster R-CNN that embedded residual substructures and attention block to effectively extract context information and autonomously learn the weight of the effective region feature map. In addition, this paper put forward a priori constraint based on thyroid position to further improve the accuracy of thyroid nodule detection and classification.

2 MATERIALS AND METHODS

2.1 Image Acquisition

The thyroid ultrasound data set used in our study was provided by the affiliated hospital of Qingdao University. The patient's private information has been hidden, and the contour of the nodules in the image was marked by doctors with years of clinical experiences. The image acquisition came from three ultrasound image scanners, namely HIVISION Preirus, HIVISION 900 and Siemens S2000 equipped with a liner probe with central frequency of 7.5~14.0MHz. There were 157 cases included in the study (62 Benign cases included 6 males and 56 females, 95 Malignant cases included 10 males and 85 females). All the diagnosis results were confirmed by surgery and pathological examinations or biopsy. The dataset consists of 157 folders, each folder includes an image sequence
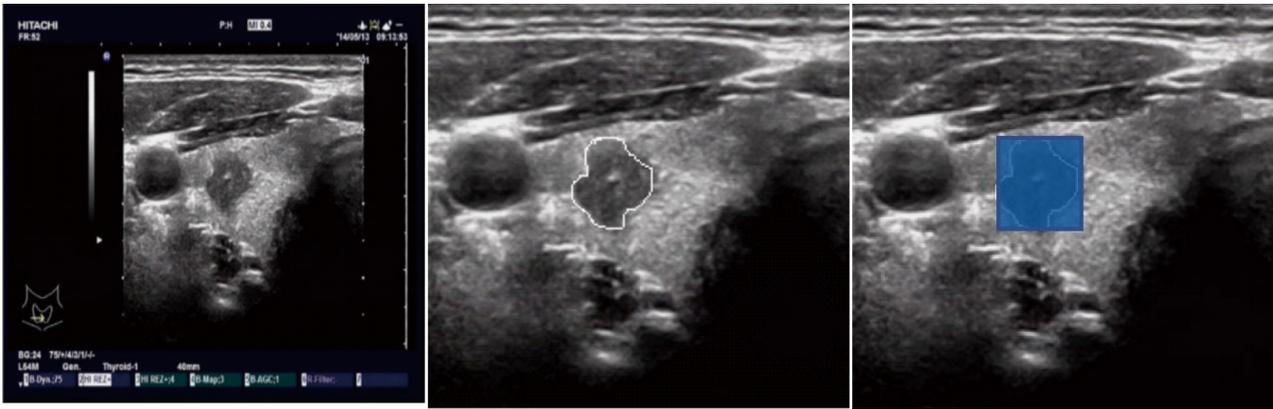
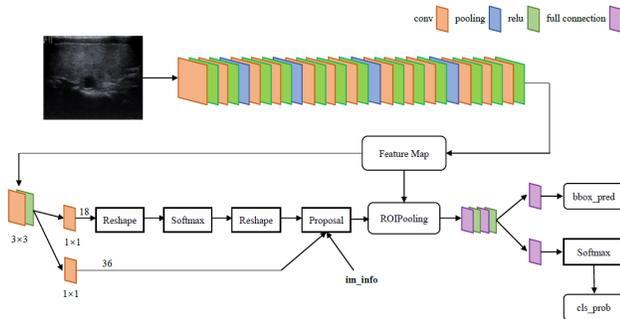**Figure 1: Sample Images. (a)Original Image, (b)Labeling Image, (c)Ground Truth.**

Figure 2: Faster R-CNN Model.

(ranging from 1 to 37 images). Finally, for data balance, there were totally 1324 images selected for this study, which the ratio of benign nodules and malignant nodules was about 1: 1.

Each image was labeled using professional software (according to the nodule outline drawn by doctors, the ground truth completely includes the smallest circumscribed rectangle of each nodule) and was made in accordance with the PASCAL (Pattern Analysis, Statistical Modelling and Computational Learning) format.

The sample images are shown in Figure 1. We can observe that the overall contrast of thyroid ultrasound image is very low, especially in the nodule area, where the echo is darker than the surrounding tissue. The features of benign and malignant nodules are not obvious, and the nodule area is small relative to the background. The basic model of Faster R-CNN is shown in Figure 2

2.2 Embedding Residual Substructure

Figure 3(a) is the prototype of residual unit [14] and Figure 3(b) is the improved residual substructure. The latter adds 1×1 convolution, which can reduce the calculation complexity to a certain extent. And it introduces more nonlinear mapping so that even if there is a slight change in the input image, it will cause a large change in the loss resulting in an increase in the gradient of back propagation. This can avoid the problem of gradient disappearance due to deepening of the network.

2.3 Introducing Attention Mechanism

We introduced the squeeze-and-excitation attention module [15], which can effectively extract context information. After obtaining the feature weight diagram, it will be multiplied by the abstract features and then focus on target regions of the thyroid images through network learning. In other words, the attentional mechanism can improve the sensitivity and accuracy of Faster R-CNN's prediction of dense pixel categories by suppressing the activation values of non-correlated regional features. The SE block is shown in Figure 4

2.4 Constraints Based on Medical Priori Knowledge

Based on the fact that the thyroid nodules are distributed in the thyroid region, firstly, the closed operation of the morphological filtering was used to eliminate the black hole in the closed region of ultrasound images and the external contours of thyroid regions were depicted, so as to roughly locate the thyroid nodules, as is shown in Figure 5(a) and Figure 5(b).

Secondly, it can be seen that the nodules are generally distributed in the center of the thyroid area from the analysis of the experimental data set. Therefore, the restriction of the central constraint was introduced and the scores of the constraint and the model candidate box were supplemented by different weights to obtain the final evaluation value. As is shown in equation 1), W_1 represents the weight based on the priori of the center position and W_2 represents the weight of the candidate box score. Distance in the formula is the distance between the center position of the candidate box and the center of image. MAX and MIN respectively represent the maximum and minimum distances in all the candidate boxes generated in one image, as is shown in Figure 5(c). The value of $\frac{MAX-distance}{MAX-MIN}$ tends to 0 when the distance is larger, which indicates the candidate box is off center. And the value of $\frac{MAX-distance}{MAX-MIN}$ tends to 1 when the distance is smaller, which indicates the candidate box is more towards the center. Finally, a group of weight values with the best effect were obtained through a series of experiments.
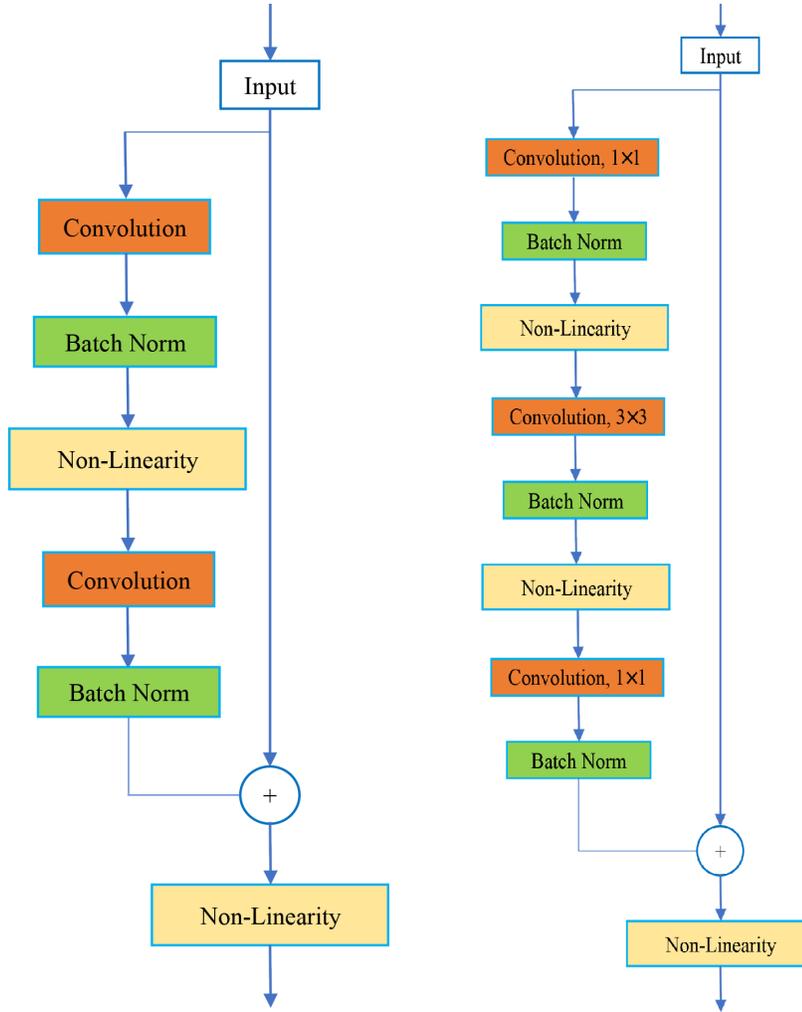
Figure 3: Residual Substructure. (a)Original Residual Unit, (b)Proposed Residual Unit.
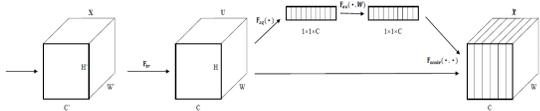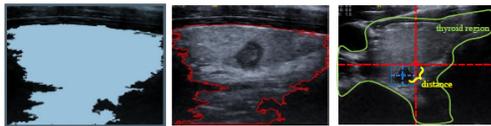

Figure 4: Squeeze-and-Excitation Block.

Figure 5: Priori Knowledge Based on Thyroid Position. (a)Obtain Outline Points, (b)Outline, (c)Calculate Distance.

$$evaluate\ value = W_1 \times \frac{MAX - dis\tan ce}{MAX - MIN} + W_2 \times score,$$
$$MAX, MIN = \max_{1 \le i \le n} dis\tan ce_i, \min_{1 \le i \le n} dis\tan ce_i, \quad (1)$$
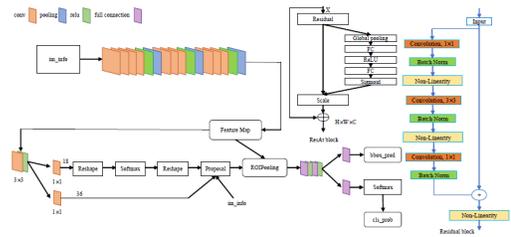$$W_1 + W_2 = 1$$

Figure 6: Residual Attention Faster R-CNN.

2.5 Training ResAt-Faster R-CNN

On the basis of Faster R-CNN [16], residual unit and attention mechanism were embedded in the network. In addition, the priori knowledge is proposed to further improve the model performance. The whole network structure of final model ResAt-Faster R-CNN is shown in Figure 6

Before training, we performed histogram equalization and data enhancement on the data set. After histogram equalization, the image had high contrast and variable gray tones, which would be helpful for feature extraction. Due to the fact that thyroid ultrasound image showed the characteristics of tissue stratification and thyroid nodules would change the corresponding category label if its size or shape became different, so only symmetric method was used to amplify the data set about twice.

3 RESULT

3.1 Performance Measurement

The performance of the model in terms of detection and classification can be evaluated by comparing the differences between the experimental results and the ground truth. In the experiment, AP (Average Precision) was used to quantitatively evaluate the detection performance. Specifically, the detection ability of the model was evaluated through AP of benign and malignant nodules and mAP of mean Average Precision of two types of nodules detection, shown in equation 2)-(5).

$$precision = \frac{true\ object\ detection}{all\ detected\ boxes} \quad (2)$$

$$recall = \frac{true\ object\ detection}{all\ ground\ truth\ boxes} \quad (3)$$

$$AP = \int_0^1 p(r)\,dr \quad (4)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (5)$$

where p(r) is precision – recall curve and N is class categories, for our problem, N =2.

The classification ability of the model was quantitatively evaluated using the following three indicators which are shown in equation 6)-(8) [17].

$$Classification\ accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

TP and TN respectively represent the number of positive and negative samples correctly classified by the model, while FP and FN respectively represent the number of negative and positive samples wrongly classified by the model. In our experiments, a positive sample refers to a malignant nodule and vice versa.

In the rest of this section, we evaluated the detection and classification performance of Residual Attention Faster R-CNN on our private data set, and then tested similar indicators on the improved model under the constraints of medical priori knowledge. Finally, we compared the benchmark model with the final model and the experimental results were analyzed and summarized. All the network models were trained on ImageNet data set in advance for initializing parameters [18].
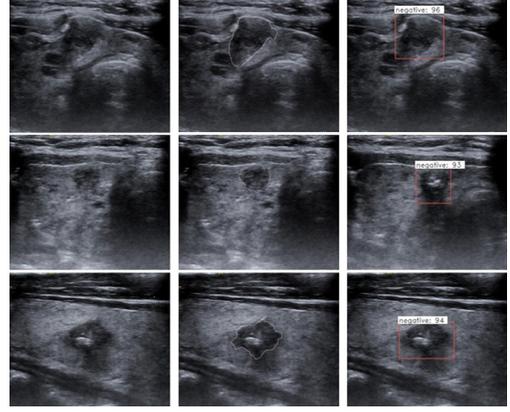

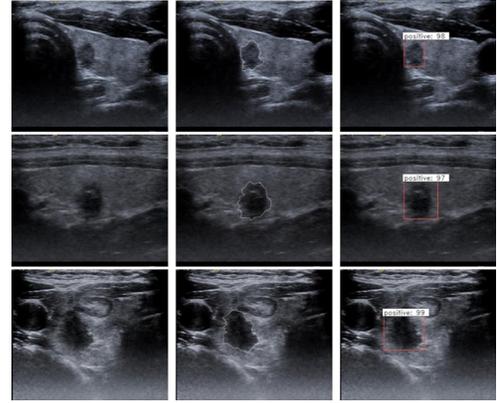**Figure 7: Results of Experiments. (a)Benign Nodule,(b)Labeled Image,(c)Output of Model**

Figure 8: Results of Experiments. (a)Malignant Nodule, (b)Labeled Image, (c)Output of Model

3.2 Experimental Results

The dataset was divided into training set and test set according to the ratio of 10:3. The performance of ResAt-Faster R-CNN is shown in Table 1, and it can be seen that the introduction of residual unit and attention mechanism greatly improve both detection and classification performance of thyroid nodules. Each row of experiment is based on the previous row.

From the table, we can see that Residual Attention Faster R-CNN has 20%, 29%, 11%, 13%,2% and 8% improvement in classification accuracy, specificity, sensitivity, malignant AP, benign AP and mAP compared with the base model. It turns out that ResAt-Faster R-CNN can achieve more accurate detection and classification.

Some results are shown in Figure 7 and Figure 8

Table 2 shows the experimental results obtained by using priori strategy based on thyroid position. We found the optimal combination of weights through experiments on W_1 and W_2 with ten groups of different values. As can be seen from the Table 2, the detection and classification evaluation indexes were the best when $W_1 = 0.6$ and $W_2 = 0.4$.

Table 1: Performance Indicators

Method	Accuracy	Specificity	Sensitivity	MalignantAP	BenignAP	MAP
Base model	0.61	0.54	0.70	0.41	0.43	0.42
Residual unit	0.69	0.89	0.60	0.52	0.43	0.48
Data expansion	0.73	0.92	0.64	0.51	0.42	0.47
Equalization	0.72	0.94	0.61	0.46	0.47	0.46
Attention	0.81	0.83	0.81	0.54	0.45	0.50

Table 2: Priori Experiment Based on Thyroid Position

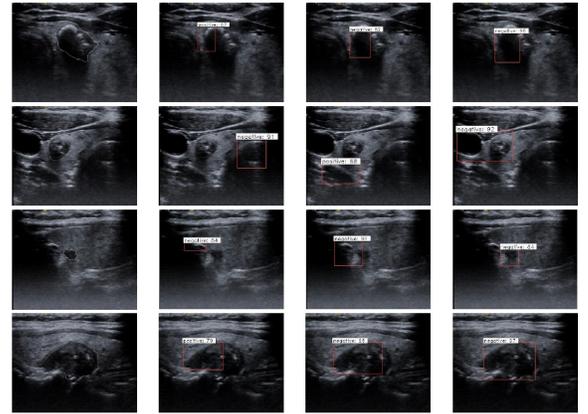| W1/W2	Accuracy	Specificity	Sensitivity	MalignantAP	BenignAP	mAP
0.1/0.9	0.79	0.87	0.76	0.59	0.45	0.52
0.2/0.8	0.79	0.89	0.76	0.57	0.46	0.52
0.3/0.7	0.80	0.89	0.76	0.57	0.46	0.52
0.4/0.6	0.81	0.90	0.78	0.58	0.47	0.53
0.5/0.5	0.82	0.90	0.78	0.59	0.47	0.53
0.6/0.4	0.83	0.90	0.80	0.59	0.48	0.54
0.7/0,3	0.82	0.60	0.79	0.59	0.47	0.53
0.8/0.2	0.82	0.89	0.79	0.60	0.46	0.53
0.9/0.1	0.78	0.88	0.73	0.56	0.45	0.50
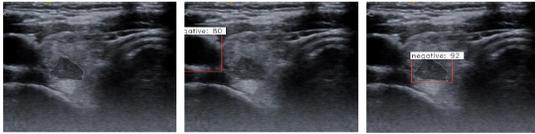
Figure 9: Effectiveness of Contrast. (a)Ground Truth, (b)Before Constraint, (c)After Constraint

Due to lower gray level of the carotid artery on the left edge whose appearance was a round shape, the model detected it as a nodule by mistake, as is shown in Figure 9(a) and 9(b). But Figure 9(c) shows that the nodule can be correctly detected when we used the priori knowledge and the score has been increased correspondingly. Compared with the previous results, the accuracy and detection rate of the final model were both improved, which indicated that the priori knowledge based on thyroid position was reasonable and effective.

It can be seen from the Table 3, the final model which priori knowledge added as constraint has a significant improvement in classification accuracy and nodule detection rate compared with the base model. The model can satisfy the requirements of the auxiliary doctors and confirm the validity of the proposed method.

Figure 10: Hard Samples. (a)Labeling Image, (b)Base Model, (c) ResAt Faster R-CNN, (d)Final Model

3.3 Performance Comparison of Hard to Classify Images

Although the above model has performed well in test data set, automatic nodule detection and classification is a very challenging clinical task. Here are some images which are not easy for model or

Table 3: Final Performance Comparison

Method	Accuracy	Specificity	Sensitivity	MalignantAP	BenignAP	mAP
Base Model	0.61	0.54	0.70	0.41	0.43	0.42
Final Model	0.83	0.90	0.80	0.60	0.48	0.54

doctors to recognize. The hard to recognize examples are shown in Figure 10The figure shows four representative samples cases with difficulty in detection and classification in the data set from top to bottom. The ROI of the first image has no obvious contrast with the background color, and the image is dark on the whole. In the second image, the ROI margin is blurred and there was interference of dark echo region in the carotid artery on the left of the image. The nodule in the third image is too small for the whole image and the nodule in the fourth image is too large, which is the opposite of the previous situation.

Taking the image in the first row for example, the contrast between the nodule and the background was so low that the edge of the nodule was almost integrated into the surrounding tissue. Therefore, the candidate box generated by the first two models did not contain the entire nodule region. When facing these complicated situations, the probability of misjudgment increased. Obviously, the model needs to be further improved to deal with these interferences in the ultrasound image.

4 DISCUSSIONS AND CONCLUSIONS

The paper firstly analyzed the shortcomings of applying Faster R-CNN directly in the field of medical image detection and classification, and proposed Residual Attention Faster R-CNN with priori knowledge for thyroid ultrasound image. The advantages of the proposed model were as follows: (1) Embedding improved residual units in the feature extraction network. Based on the residual mechanism in ResNet, the residual substructure was improved, which can not only reduce the calculation number of parameters during training, but also avoid the problem of gradient disappearance and the convergence speed of the network was improved. (2) Introducing attention mechanism to effectively extract context information. The model can improve the weight of effective regions in the image and accuracy of ROI classification through independent learning. (3) Proposing the priori strategy based on thyroid position. According to the priori knowledge that nodules were generally distributed near the center of thyroid images, the scoring rules of each candidate box were reformulated. The new indicators were calculated to reorder and screen when the model generated candidate boxes, so as to obtain a more appropriate and accurate detection box.

In the future work, in-depth research can be carried out from the following aspects to provide directions for further improving performance of the model: (1) Observe the test samples with inaccurate detection and classification respectively, and summarize the commonalities through analyzing the characteristics of these sample images to make targeted improvements. (2) Study the influence of images of different quality collected by different equipment on the thyroid images, which can increase the amount of training data and reduce the impact caused by different distribution of images.

(3) In the actual scene, it is worth considering and studying how to make full use of images which are not labeled by doctors.

ACKNOWLEDGMENTS

REFERENCES

[1] B. Freddie, F. Jacques, S. Isabelle, *et al.* (2018). Global Cancer Statistics: GLOBO-CAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin, vol. 68, pp. 394-424.

[2] Tingting Zhang, Ning Qu, Pu Zheng, el al. (2017). Application of Machine Learning in Diagnosis and Treatment of Thyroid tumors. Chinese Journal of Cancer, 27(12), 992-995.

[3] Xuan Gao, Xiaolin Wang (2019). Preliminary Research Progress of Deep Learning in Imaging Medicine. Medical Journal of Fudan University, 46(3), 408-414.

[4] Bhattacharya, S., Maddikunta, P. K. R., Pham, Q. V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., & Piran, M. J. (2020). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 102589.

[5] Lucheng Zhu, Yunliang Ye, Wenhua Luo, *et al.* (2013). A Model to Discriminate Malignant from Benign Thyroid Nodules Using Artificial Neural Network. PLOS ONE, 8(12): 1-6.

[6] Ma, J., Wu, F., Zhu, J., *et al.* (2016). A pre-trained convolutional neural network based method for thyroid nodule diagnosis. Ultrasonics, 73, 221-230.

[7] Jianning Chi, Ekta Walia, Paul Babyn, *et al.* (2017). Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. Journal of Digital Imaging, 30(3), 477-486.

[8] Ma, J., Wu, F., Zhao, Q., *et al.* (2017). Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery, 12(11), 1895-1910.

[9] Tianjiao Liu, Shuaining Xie, Jing Yu, *et al.* (2017). Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. IEEE International Conference, 919-923.

[10] Hailiang Li, Jian Weng, Yujian Shi, *et al.* (2018). An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. Scientific Reports, 8(1), 6600-6612.

[11] Wenfeng Song, Shuai Li, Ji Liu, *et al.* (2018). Multi-task Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition. IEEE Journal of Biomedical and Health Informatics, 23(3), 1215-1224.

[12] Tianjiao Liu, Qianqian Geuo, Chunfeng Lian, el al. (2019). Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks. Medical Image Analysis, DOI: 10.1016/j.media.2019.101555.

[13] Fang, H., *et al.* (2021). Reliable Thyroid Carcinoma Detection with Real-Time Intelligent Analysis of Ultrasound Images. Ultrasound in Medicine & Biology, 47(3), 590-602.

[14] He K, Zhang X, Ren S, *et al.* (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2016.90.

[15] Hu J, Shen L, Albanie S, *et al.* (2017). Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, DOI: 10.1109/TPAMI.2019.2913372.

[16] Ren, S., He, K., Girshick, R., *et al.* (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis $\&$ Machine Intelligence, 39(6), 1137-1149.•

[17] Goutte C., Gaussier E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada D.E., Fernández-Luna J.M. (eds) Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-31865-1_25

[18] Tajbakhsh N, Shin J Y, Gurudu S R, *et al.* (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging, 35(5), 1299-1312.