2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Image Segmentation of Thyroid Nodule and Capsule for Diagnosing Central Compartment Lymph Node Metastasis

Xiandong Liao[1], Keru Lin[1], Donghao Chen[1], Honggang Zhang[1], Yingying Li[2] and Bo Jiang[2]

***Abstract*— Thyroid ultrasound (US) image segmentation is of great significance for both doctors and patients. However, it is a challenging task because of the low image quality, low contrast and complex background in each US image. In recent years, some researchers have done thyroid nodule segmentation tasks, but the results achieved are not particularly satisfactory. In this paper, we have broadened the targets of interest and included both thyroid nodules and capsules into our research scope. We propose a method that implements a C-MMDetection to detect and extract the region of interest (ROI), and a modified salient object detection network U^2-RNet to segment nodules and capsules respectively. Experiments show that our method segments nodules and capsules in US images more effectively than other networks, which is very helpful for doctors to diagnose central compartment lymph node metastasis (CLNM).**

I. Introduction

According to the global cancer statistics of 2019, the incidence rate of thyroid is increasing [1]. Papillary thyroid carcinoma (PTC) is the most common and typical thyroid malignancy [2]. Ultrasound (US) imaging technology is used to observe the state of the thyroid and has become an important way for PTC diagnosis and treatment. Many deep learning methods have been applied to thyroid US image segmentation [3] - [5], and most of them just care about segmentation of thyroid nodules or judgment of benign and malignant. However, about 30%–80% of PTCs are associated with lymph node (LN) metastases, and the precise identification of central compartment lymph node metastasis (CLNM) is crucial for the optimal management of patients with PTC. CLNM is hard to detect preoperatively, and there are many factors to affect it, including nodule site, nodule size, aspect ratio, distance between the nodules and the adjacent capsules, etc. Doctors need to know these messages in order to make the right treatment plan [6]. Therefore, it is significantly meaningful to segment thyroid nodules and capsules in the US image of PTCs and determine the positional relationship between them.

Convolutional neural network (CNN) has been greatly developed in recent years and applied to image semantic segmentation. For example, U-net [7] is symmetrically U-shaped to obtain edges of objects and performs well on some biomedical datasets. SegNet [8] shares information between encoder and decoder. PSPnet [9] proposes spatial pyramid pooling and produces multi-level semantic feature. Deep-Labv3+ [10] uses atrous convolution and deeper network. However, these networks are not good enough for thyroid image segmentation when the dataset is small and limited.

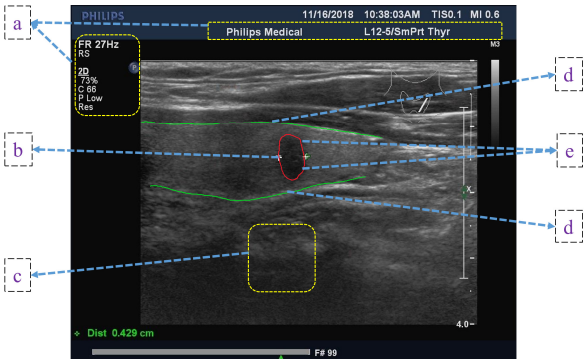


Fig. 1. Features of thyroid US image: (a) highlighted letters and symbols of the US machine; (b) white measurement marks drawn by doctor; (c) speckle noise in US images; (d) the green line drawn by the doctor represents the capsule, it will be saved separately as a label image; (e) the red circle drawn by the doctor represents the nodule, it will also be saved as a label image.

As we all know, it is challenging for accurate segmentation in thyroid US images because of the following reasons: (1) highlighted letters and symbols on the background; (2) manually induced artefacts; (3) lots of speckle noise; (4) low contrast and irregular boundary of the nodule; (5) long and narrow shape of the capsule, as shown in Fig. 1.

In this paper, we propose a method for segmentation of thyroid nodules and capsules in US images. The method mainly includes region of interest (ROI) extraction and image segmentation as shown in Fig. 2. For the ROI extraction assignment, we put forward an object detector named C-MMDetection, it is modified from MMDetection [11] which has been proven efficient in the field of object detection and instance segmentation. For image segmentation task, we innovatively consider advanced salient object detection networks and propose U^2-RNet which is built on the basis of U^2-Net [12] and BASNet [13]. U^2-RNet is aimed to greatly deal with the irregular boundary and size of thyroid nodules segmentation and the discontinuity of capsules segmentation, so as to provide more precise information for doctors to diagnose CLNM.

By analyzing experimental results, our proposed method has achieved great performance in segmentation of thyroid nodules and capsules. We can obtain information that doctors need including nodule site, nodule size, aspect ratio, distance between the nodules and the adjacent capsules more accurately than other networks.

[1]Xiandong Liao, Keru Lin, Donghao Chen, Honggang Zhang are with Faculty of Information and communication Engineering, Beijing University of Posts and Telecommunications, CHINA
E-mail: {liaoxiandong, krlin, cdh0699, zhhg }@bupt.edu.cn.

[2]Yingying Li and Bo Jiang are with Department of Ultrasound, First medical center of Chinese PLA General Hospital, CHINA.
E-mail: lyy19900325@sina.com, Jiangbo301yy@163.com

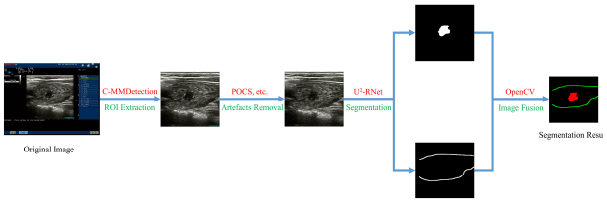


Fig. 2. The pipeline of our method. The red words above the arrow represent the method used at each step, and the green words below the arrow represent its function.

II. Methodology

A. C-MMDetection for ROI Extraction

ROI is generally in the central area of a thyroid US image, which is rectangular or approximately trapezoidal and contains pathological information that this research focuses on. ROI is obviously different from other regions of US images, which provides us with good convenience for extracting it using deep neural networks. Ying [4] extracts the ROI by a simple FCN model based on U-net, but the accuracy of the model is only 96% and the model has low robustness when the ROI is not a standard rectangle. In this paper, we propose an object detector C-MMDetection based on Faster R-CNN [14] for ROI extraction of thyroid US images, the framework of our detector is shown in Fig. 3. Our detector expands the functions of MMDetection [11], it includes: (1) ResNet-50 [15] without the last fully connected layer as the Backbone to transform an image to feature maps; (2) Feature Pyramid Network (FPN) as Neck to refine or reconfigure the original feature map generated by the backbone; (3) Region Proposal Network (RPN) as the DenseHead to operate on dense locations of feature maps; (4) Bounding Box (BBox) classification as the RoIHead to receive ROI features and make predictions of ROI; (5) Image Cropping Module (ICM) as the final image processing module to obtain the accurate coordinate information of the ROI and crop it out .

Therefore, our object detector C-MMDetection can not only detect the ROI, but also crop the ROI as a separate image according to the detection result. The ROI image is much smaller than the original US image, which will be very meaningful for our next image segmentation work.

B. U^2-RNet for Thyroid Image Segmentation

Automatic thyroid image segmentation is a difficult task, because of the complex content, low image quality and contrast of US images [5]. Besides, the variable positions and irregular shapes of nodules and capsules will also cause trouble for us to find them. We have applied some state-of-the-art image semantic segmentation networks to segment them at first. However, for one thing, the position and boundary of the nodules segmentation results are not satisfactory. For another, the continuity and accuracy of the capsules segmentation results also have many problems. After extensive comparison and research, we put more attention on the salient object detection networks that perform well in complex images and finally propose U^2-RNet based on U^2-Net [12] and BASNet [13]. U^2-RNet consists of three parts: multi-layer encoder and decoder, saliency map fusion module, refinement module (RM) as shown in Fig. 4.

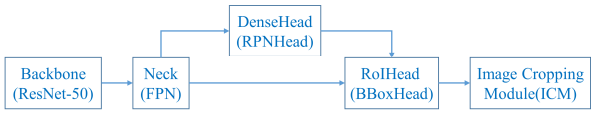


Fig. 3. Framework of the detector based on Faster R-CNN.

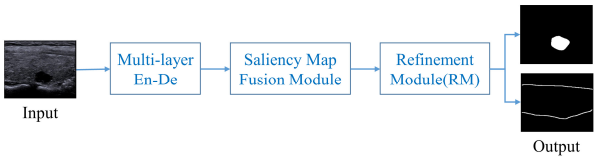


Fig. 4. Basic framework of U^2-RNet.

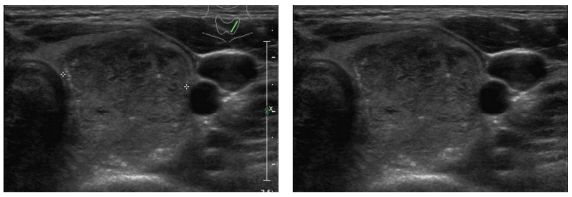
1) Multi-layer Encoder and Decoder: The main encoding and decoding structure is the same as U^2Net [12], that includes a six stages encoder and a five stages decoder, each stage consists of different depth residual U-block (RSU) proposed by Qin [12]. The special structure can ensure that our network has a larger range of receptive fields and extracts richer local and global features from input feature maps with arbitrary spatial resolution, and it is very suitable for our ROI image characteristics. Finally, we can obtain six side output saliency probability maps from the Multi-layer Encoder and Decoder.

2) Saliency Map Fusion Module: Referring to the saliency map fusion module of U^2-Net [12], first, we fuse the six side output saliency probability maps from last step with a concatenation operation followed by a 1×1 convolution layer and a sigmoid function to generate the synthesized saliency probability map M_1. M_1 is grayscale image and the brightness value of each pixel of M_1 is between 0 and 255. Second, in order to highlight our most sure area, we set a brightness threshold to perform operations on each pixel in M_1 to get a new and clearer binary map M_2 by (1).

$$p_2^i = \begin{cases} 1, & p_1^i \geq T_h \\ 0, & p_1^i < T_h \end{cases}, \tag{1}$$

where p_1^i represents brightness value of each pixel in M_1, p_2^i represents each corresponding pixel in M_2, T_h is the brightness threshold we set. When our salient object is nodule, T_h is 192, and when salient object is capsule, T_h is 128.

3) Refinement Module (RM): Experiments show that the M_2 often contains some isolated small regions that are incorrectly segmented due to the speckle noise of the ROI images. Inspired by the Residual Refinement Module (RRM) of BASNet [13], we design a Refinement Module (RM) to solve this problem based on Python-OpenCV. First, we find the contour information of each independent region in M_2. Then we calculate their area based on the contour information. Finally, for nodules segmentation results, we only keep the largest area on the image, because our current US images are all single malignant nodules. For capsules segmentation results, considering that most capsules are long and narrow, we remove all areas that are less than one-third of the largest area on the image.



(a) (b)

Fig. 5. Example of artefacts removal. (a) The original ROI image; (b) the artifacts-removed ROI image.

III. EXPERIMENTS

A. Dataset and Experiment Settings

1) Dataset: The US data of PTCs were collected in Chinese People's Liberation Army (PLA) General Hospital. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. There are two experimental datasets in the study: (1) dataset A is mainly used for training and verification of C-MMDetection. We randomly select 989 US images from the hospital and use the LabelMe tool [16] to mark the ROI with a rectangular box. These marks are stored in the XML file format, and each image corresponds to an XML file as label. We split the image and the label into two parts: the training set and the validation set according to the ratio of 8:2, and make a simple VOC-style dataset A; (2) dataset B is mainly used for training and testing of U^2-RNet. Up to now, dataset B is composed of 632 US images from 308 patients in the study. These images were collected and marked by experienced radiologists in Chinese PLA General Hospital. The sample image and marks are shown in Fig. 1, and the marks of nodules and capsules will be saved as label images separately. We randomly split the images into training set and test set with the ratio of 8:2. Therefore, the training set contains 506 images and the test set contains 126 images. By flipping images horizontally and vertically, we increase the number of training set to 1012.

In addition, before the ROI images of the dataset B enter U^2-RNet, we remove the artefacts on the ROI image using the method proposed by Narayan [17] that includes the steps below: (1) obtain a binary image according to the pixel values; (2) apply 2D connected component algorithm on the binary image and generate a labelled image with K components; (3) define the texture region R1 with the maximum number of pixels in the labelled image; (4) calculate R2 as the complement of R1; (5) plot the histogram of R2 and divide it into three parts: [0,100], [101,200] and [201,255], respectively; (6) take the histogram peaks in each of the three parts as the intensity levels of the artifacts; (7) suppress artefact intensities to generate the artifacts-removed image; (8) restore the image using the method of Projection onto Convex Sets (POCS) [18]. The method is effective for removing artefacts, and Fig. 5 shows the example of artifacts-removed ROI image.

2) Experiment Settings: The training schedule of C-MMDetection is similar to MMDetection [11] based on GPUs, there are 12 training epochs in total. We use Adam optimizer [16] to train our U^2-RNet with a batch size of 12 and set its hyper parameters to default (initial learning rate =1e-3, betas= (0.9, 0.999), eps=1e-8, weight decay=0). We train the network until the loss converges. The whole experiment process was operated on PyTorch and lasted about 4 days.

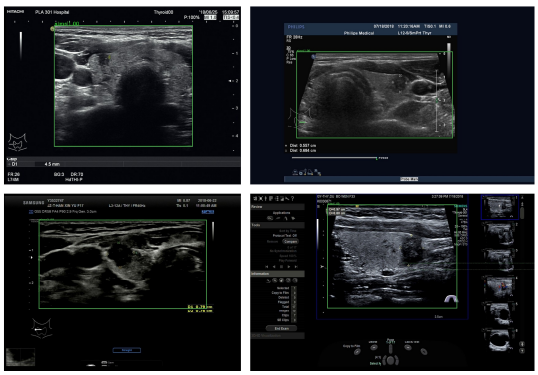


Fig. 6. Examples of ROI extraction. The green rectangle in the middle of each image represents the ROI detected by our model

B. Evaluation Metrics

In this paper, the evaluation Metrics mainly include Precision, Recall, Dice Similarity Coefficient (DSC) and Intersection over Union (IOU). Their calculation formulas are present in (2), (3), (4) and (5).

$$Precision = \frac{TP}{TP + FP} \times 100\%, \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%, \quad (3)$$

$$DSC = \frac{2 \times TP}{FP + 2 \times TP + FN} \times 100\%, \quad (4)$$

$$IOU = \frac{TP}{FP + TP + FN} \times 100\%, \quad (5)$$

where TP is True Positive, FP is False Positive, FN is False Negative. They depend on the relationship between the pixels of ground truth and predicted result.

C. Result Analysis

1) ROI Extraction: The C-MMDetection model we trained performs very well on the validation set of the VOC-style dataset A, and its precision is 98.5% and recall is 100%. It means that we can find the ROI of various US images accurately. The test performance of the model on dataset B is illustrated in Fig. 6. We can see that no matter what type of US images, the effect of the ROI extraction model is excellent.

2) Image Segmentation: The nodules segmentation performance of the dataset B is shown in Fig. 7 and Table I. U^2-RNet segments nodules of different sizes and characteristics very well and has achieved 1.8%, 1.9%, 1.7% and 1.1% improvement at least in DSC, IOU, Recall and Precision compared with other networks. Besides, U^2-RNet also performs better in capsules segmentation than U^2-Net, it removes some useless small areas and improves 3.3%, 3.9%, 3.5%, 2.8% in DSC, IOU, Recall and Precision as shown in Fig. 8 and Table II. We will merge the results of nodules segmentation and capsules segmentation in the last step, as shown in Fig. 9. From final fused image, we can easily get the information doctors want by analyzing pixels, including nodule site, nodule size, aspect ratio, distance between the nodules and the adjacent capsules, etc.

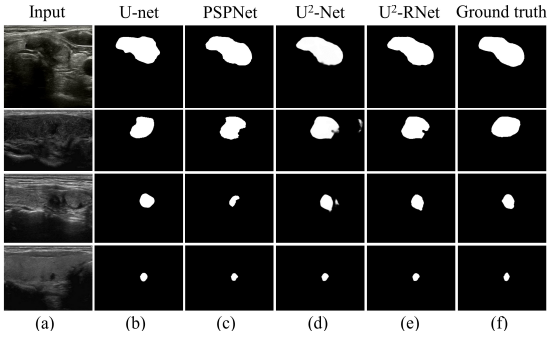


Fig. 7. Examples of nodules segmentation on different networks.

TABLE I. NODULES SEGMENTATION RESULTS

Model	Evaluation Metrics			
	DSC	***IOU***	***Recall***	***Precision***
U-net	76.70%	69.70%	78.06%	79.07%
PSPNet	83.74%	76.13%	84.29%	86.97%
U²-Net	86.23%	78.89%	87.26%	87.88%
U²-RNet	**88.03%**	**80.77%**	**89.00%**	**89.01%**

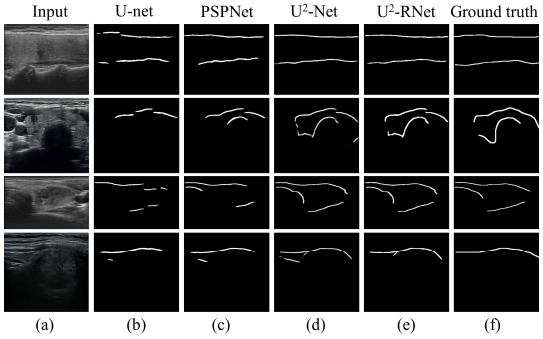


Fig. 8. Examples of capsules segmentation on different networks.

TABLE II. CAPSULES SEGMENTATION RESULTS

Model	Evaluation Metrics			
	DSC	***IOU***	***Recall***	***Precision***
U-net	32.35%	21.07%	28.39%	41.67%
PSPNet	40.80%	27.88%	38.66%	46.68%
U²-Net	44.11%	30.30%	41.06%	49.91%
U²-RNet	**47.42%**	**34.16%**	**44.60%**	**52.72%**

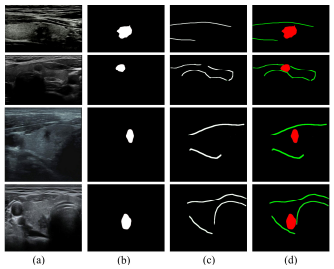


Fig. 9. Examples of segmentation results fusion. (a) The input ROI images, (b) the nodules segmentation results, (c) the capsules segmentation results, (d) the final fusion segmentation results.

IV. CONCLUSION

In this paper, we propose an efficient segmentation method for the US images. Both C-MMDetection and U²-RNet proposed by us have good performance on the US dataset, and experiments show that our segmentation results can provide important reference information for doctors to diagnose CLNM. Our US data are still increasing, we will pay more attention to optimizing the network structure and conducting more research on US images in the future.

ACKNOWLEDGMENT

This work was funded by the National Natural Sciences Foundation of China under Grant No.81771834.

REFERENCES

[1] Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. "Cancer statistics, 2019." CA: a cancer journal for clinicians 69.1 (2019): 7-34.

[2] Lee, Young Ki, et al. "The relationship of comorbidities to mortality and cause of death in patients with differentiated thyroid carcinoma." Scientific reports 9.1 (2019): 1-10.

[3] Ding, Jianrui, et al. "Automatic Thyroid Ultrasound Image Segmentation Based on U-shaped Network." 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2019.

[4] Ying, Xiang, et al. "Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network." International Conference on Neural Information Processing. Springer, Cham, 2018.

[5] Chen, Junying, et al. "A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images." Computer methods and programs in biomedicine 185 (2020): 105329.

[6] Tian, Xiaoqi, et al. "Papillary thyroid carcinoma: an ultrasound-based nomogram improves the prediction of lymph node metastases in the central compartment." European radiology 30.11 (2020): 5881-5893.

[7] Ronneberger, et al. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[8] Badrinarayanan, et al. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.

[9] Zhao, Hengshuang, et al. "Pyramid scene parsing network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[10] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." Proceedings of the European conference on computer vision (ECCV). 2018.

[11] Chen, Kai, et al. "MMDetection: Open mmlab detection toolbox and benchmark." arXiv preprint arXiv:1906.07155 (2019).

[12] Qin, Xuebin, et al. "U2-Net: Going deeper with nested U-structure for salient object detection." Pattern Recognition 106 (2020): 107404.

[13] Qin, Xuebin, et al. "Basnet: Boundary-aware salient object detection." Computer Vision and Pattern Recognition (CVPR). 2019.

[14] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." arXiv:1506.01497 (2015).

[15] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[16] Russell, Bryan C., et al. "LabelMe: a database and web-based tool for image annotation." International journal of computer vision (2008).

[17] Narayan, Nikhil S., et al. "Automatic removal of manually induced artefacts in ultrasound images of thyroid gland." 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013.

[18] Youla, Dan C., and Heywood Webb. "Image Restoration by the Method of Convex Projections: Part 1ꟾTheory." IEEE transactions on medical imaging 1.2 (1982): 81-94.

[19] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).