Automatic Thyroid Ultrasound Image Segmentation Based on U-shaped Network

Jianrui Ding*, Zichen Huang, Mengdie Shi
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China

Chunping Ning
Department of Ultrasound
The Affiliated Hospital of Qingdao University
Qingdao, China

Abstract—**Automatic tumor segmentation of thyroid ultrasound image is quite challenging due to the poor image quality. Recently the U-shaped network, especially U-Net, has achieved good results in medical image segmentation. In this paper, we proposed a modified U-Net model (ReAgU-Net), which embedded the improved residual units into the skip connection among the encoding and decoding path and introduce the attention gate mechanism to multiply the weight feature maps obtained from shallow layers and deep layers. Also, a hyperparameter is introduced to combine Focal-Tversky Loss, Dice Loss and Cross-entropy Loss to jointly guide the model optimization process. The experimental results demonstrate that the proposed approach outperforms the other U-shaped models.**

Keywords-thyroid ultrasound image; automatic segmentation; U-Net; ReAgU-Net

I. INTRODUCTION

Thyroid nodules are one of the most common diseases in adults. Although most nodules are benign, in recent years, the thyroid cancer has increased quickly. In the statistics of global cancer population in 2018, the incidence and mortality of thyroid cancer rank ninth and sixth respectively [1]. The computer-aided diagnosis (CAD) system can describe the nodules objectively and quantitatively, eliminate the subjectivity of doctors, and provide a useful reference for doctors. Automatic thyroid ultrasound image segmentation is a key step in CAD and also is very challenging due to low contrast, speckle noise, weak boundary and artifacts.

With the great success in natural scene image analysis, more and more deep learning methods have been applied to medical image segmentation [2-8], including thyroid ultrasound image segmentation [9-11]. The major strategy of the approaches is to apply Convolutional Neural Network (CNN) to encoding the image and upsample the deep features to decoding the image. Although some results have been achieved, there are still some problems to be studied in depth.

Natural scene images are easy access, easy labeling and have large data sets. The deep learning model designed for them is usually having deep hierarchy and large parameters. While medical images are difficult to obtain and label, and also are small data sets. Therefore, the direct application of the existing deep learning model will make the training data distributed in the space sparsely. This will bring the over-fitting and affect the generalization ability of the model. Therefore, we need to improve the existing depth learning model to fit for medical images.

To solve the above problem, this paper proposed an improved U-shaped model. On the basis of U-Net, residual substructures and attention gates are embedded in the jump connection to narrow the semantic gap between shallow and deep features. In addition, considering the small object in medical images, this paper improves the loss function so that the model can promote the sensitivity while retaining the attention of overlap.

II. MATERIALS AND METHODS

A. Patients and imaging acquisition

192 patients (148 females, mean age 46.31±9.79 years; range 11~67 years; and 44 males, mean age 54.9±11.7 years, range 29~81 years) and totally 1936 images were evaluated in the study. The mean size of the nodules was 1.74cm (range 0.77~2.64cm).

Ultrasonography (US) acquisition was performed with the HITACHI Vision 900, HIVISION Preirus (Hitachi Medical System, Tokyo, Japan) and Siemens S2000 (Siemens Medical Solutions) equipped with a liner probe with central frequency of 7.5~14.0MHz. All the examinations were conducted by two experienced sonographers and all the nodules used in this study were delineated by them as the ground truth. Both of them have more than 6 years' experience. The sample images were shown in Fig.1.

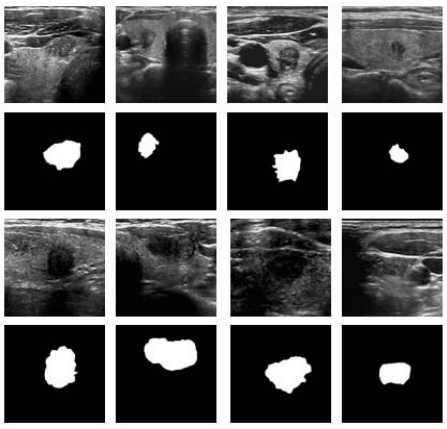
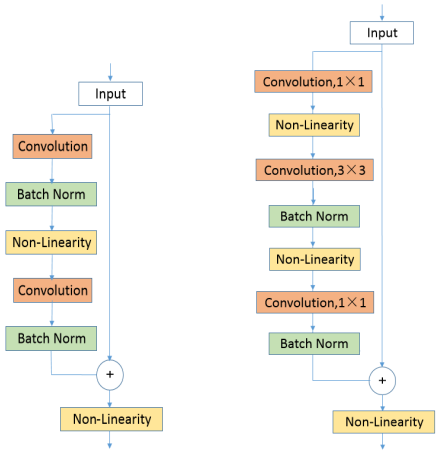


Figure 1. Sample images from dataset

978-1-7281-4852-6/19/$31.00 ©2019 IEEE

B. *Embedding residual unit in skip connections*

The prototype of residual unit [12] is shown in Fig.2(a). The structure of the new residual learning unit proposed in this paper is shown in Fig.2(b), which reduces the parameters while increasing the depth of the model. It can effectively combine the semantic level features from the encoder with the abstract features from the decoder, thus solving the semantic gap to a certain extent.



(a) original residual unit (b) proposed residual unit

Figure 2. residual unit

C. *Using attention gate mechanism*

In order to narrow the semantic gap and embedding multiple residual units in skip connections, we deepen the network horizontally as shown in Fig.2. It will inevitably lead to the loss of spatial information and the location shift of abstract features. Attention Gate (AG) [13] mechanism extracts context information from low-level features, gets weighted feature map and multiplies the abstract features. This process can autonomously learn and focus on the target structure without additional supervision, so it has the effect of position correction. The AG mechanism is shown in Fig.3.

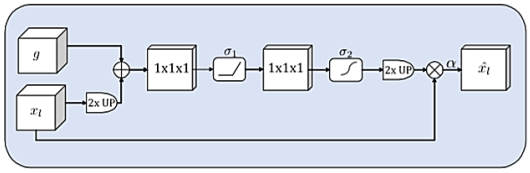


Figure 3. AG mechanism [13]

D. *Design of loss function*

In thyroid ultrasound images, nodules are usually much smaller than the background. Deep learning regards image segmentation problem as pixel classification. Therefore, the difference between foreground and background area reflects the imbalance of positive and negative categories of data. Moreover, the cost of incorrectly classifying foreground into background and background into foreground cannot be treated equally. Cross Entropy (CE) and Dice loss cannot solve the problem of category imbalance.

In the task of object detection, Lin et al. [14] compared single-stage and multi-stage detection methods, it is found that single-stage detector is often faster and simpler, but its accuracy lags behind cascade detector. The research shows that the most important problem is the serious imbalance of foreground-background categories. In this paper, a single-stage segmentation method is used, so Focal-Tversky loss is introduced. And the three loss (Cross Entropy, Dice and Focal-Tversky) are combined together as equation (1).

$$L_c = \varepsilon * FTL_c + (1-\varepsilon) * (CE_c + DL_c) \tag{1}$$

The super parameter ε was determined by grid search.

E. *Training ReAgU-Net*

On the basis of U-Net, residual unit and attention gates are embedded in the skip connection to narrow the semantic gap between shallow and deep features. The whole network structure of our model ReAgU-Net is shown in Fig.4.

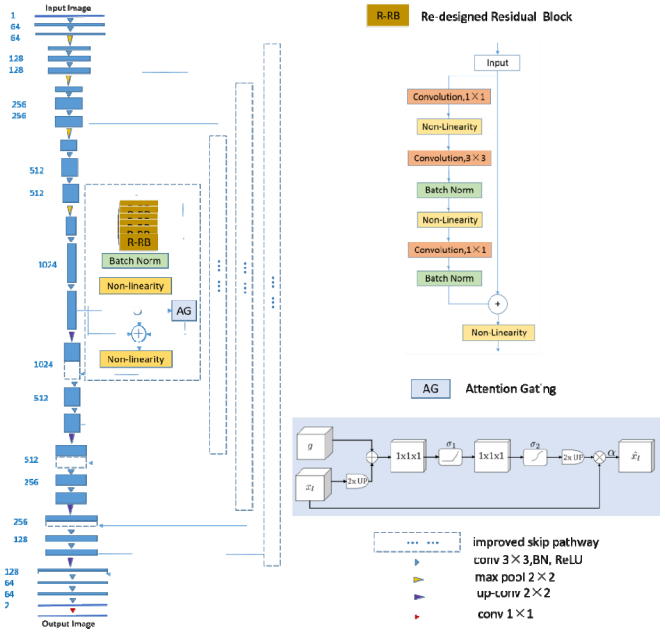


Figure 4. Network Structure of ReAgU-Net

The ReAgU-Net model uses the Adam optimization algorithm [15] to iteratively update the weight of the model. Its main parameters are: α (learning rate, used to control parameter update step), β_1 (exponential attenuation rate of first-order moment estimation, usually set to 0.9), β_2 (exponential attenuation rate of second-order moment estimation, usually set to 0.999), ϵ (very small constant, such as 10^{-7}). The training process is shown in Fig 5.

Require: α: Learning rate

$\beta_1 \in [0,1)$: Exponential decay rate for 1st moment estimate

$\beta_2 \in [0,1)$: Exponential decay rate for 2nd moment estimate

ϵ: Infinite decimal

m: Batch size

$f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0: Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$v_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged **do**

  $t \leftarrow t + 1$

  $I_i, M_i, i = 1,2,\dots,m$ (Randomly selecting m pairs with I_i and M_i)

  $S_i \leftarrow f(\theta_t; I_i), i = 1,2,\dots,m$ (Computing segmented image S_i for I_i)

  $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t stochastic objective at timestep t)

  $m_t \leftarrow \beta_1 * m_{t-1} + (1-\beta_1) * g_t$ (Update biased 1st moment estimate)

  $v_t \leftarrow \beta_2 * v_{t-1} + (1-\beta_2) * {g_t}^2$ (Update biased 2nd raw moment estimate)

  ${m_t}' \leftarrow m_t/(1-{\beta_1}^t)$ (Compute bias-corrected 1st moment estimate)

  ${v_t}' \leftarrow v_t/(1-{\beta_2}^t)$ (Compute bias-corrected 2nd raw moment estimate)

  $\theta_t \leftarrow \theta_{t-1} - \alpha * {m_t}'/(\sqrt{{v_t}' + \epsilon})$ (Update parameters)

end while

return θ_t (Resulting parameters)

Figure 5. Training the ReAgU-Net

III. RESULTS

A. Evaluation Criteria

In this paper, four commonly used indexes are used to evaluate the segmentation algorithm. They are mean Intersection over Union (mIoU), Dice Similarity Coefficient (DSC), Precision (Precision) and Recall (Recall). The computation of the indexes is shown in equation (2) – (5).

$$\text{mIoU} = \frac{1}{c}\sum_{i}^{c}\frac{TP_i}{TP_i + FN_i + FP_i} \tag{2}$$

$$\text{DSC} = \frac{2*TP}{2*TP + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

where c is the class (foreground or background), TP (True Positive), TN (True Negative), FP (False Positive) an FN (False Negative) are refer to pixels predicted by the algorithm.

B. Exprimental Results

The dataset was divided into training set, verification set and test set according to the ratio of 7:2:1. The training set is used to train the model iteratively, and the convergence of the model is judged by the error of verification set. Then, the trained model is used to segment the test set. The segmentation performance is shown in Table 1.

Table 1. Comparison of segmentation results

model	mIoU	DSC	Precision	Recall
U-Net	0.722	0.820	0.829	0.811
UNet++	0.765	0.854	0.872	0.837
ReAgU-Net	**0.788**	**0.869**	**0.873**	**0.865**

From the table, we can see that ReAgU-Net model has 6.6%, 4.9%, 4.4% and 5.4% improvement in mIoU, DSC, Precision and Recall compared with U-Net model, and 2.3%, 1.5%, 0.1% and 2.8% improvement compared with UNet+. This shows that ReAgU-Net model can recognize the location and contour of nodules better and has higher accuracy.

Some results are shown in Fig. 6.

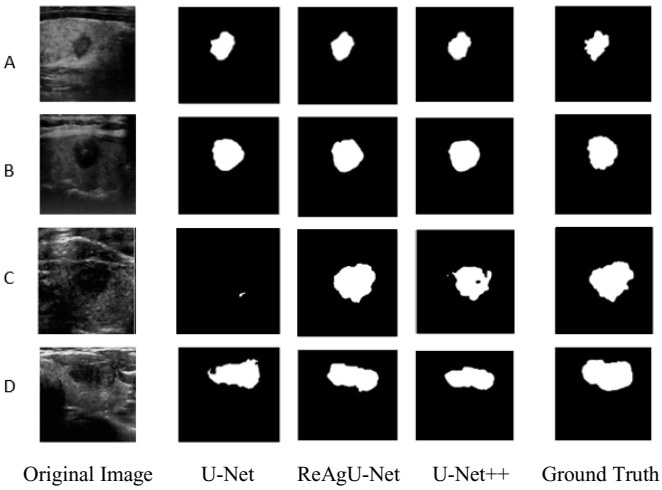


Figure 5. Training the ReAgU-Net

In order to better demonstrate the effectiveness of each improvement point and compare their roles in the model, we add each improvement point to the U-Net model in turn to compare their contributions. The results are shown in Table 2.

Table 2. Segmentation Results of Different Improvement Points

mdel	mIoU	DSC	Precision	Recall
U-Net	0.722	0.820	0.829	0.811
U-Net + AG	0.754	0.844	0.838	0.850
U-Net + R-RB	0.787	0.867	0.871	0.863
U-Net + AG + R-RB + Dice loss	0.766	0.851	0.858	0.844
ReAgU-Net	**0.788**	**0.869**	**0.873**	**0.865**

From the table, we can see that compared with the attention gate mechanism, the improved residual units contributes greatly to the performance and compared with the Dice loss, the loss function proposed in this paper can further improve the performance.

C. Performance comparison under different data sets

In order to test the generalization ability of the model, in addition to the data provided by the Affiliated Hospital of Qingdao University, an open dataset of 428 thyroid ultrasound images from DDTI (Digital Database Thyroid Image) [16] was used. These images are collected with TOSHIBA Nemio 30

and TOSHIBA Nemio MX. The frequency of linear detector is 12MHz. The location of the nodule in each image is recorded by an XML file. So, the corresponding mask image can be obtained by parsing the XML file.

Some examples are shown in Fig. 6.

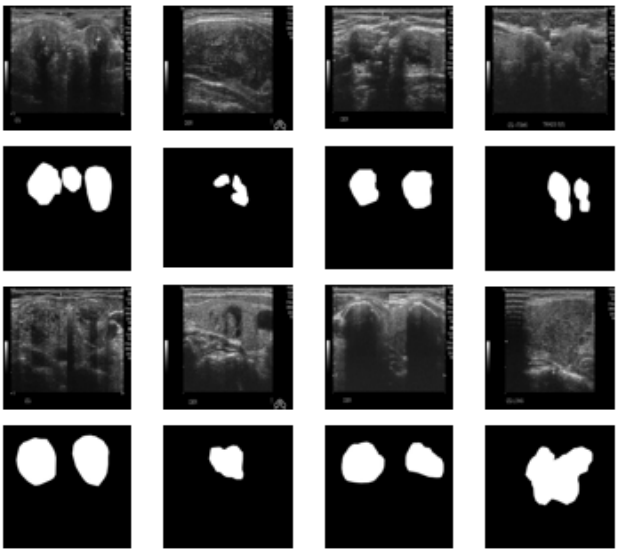


Figure 6. Sample image from DDTI

From the figure, we can see that multi-nodule exists in most images in DDTI datasets. It greatly increases the difficulty of segmentation.

The mIoU index of segmentation results are listed in Table 3.

Table 3. mIoU index of different datasets

model	Original dataset	DDTI
U-Net	0.722	0.201
UNet++	0.765	0.258
ReAgU-Net	**0.788**	**0.260**

The low performance on DDTI datasets may be attributed to the multi-nodule first, and then to the small number of images in DDTI datasets compared with the original datasets (428 vs. 1936). Deep learning algorithms require a large number of samples for training to get satisfactory results.

IV. DISCUSSIONS AND CONCLUSIONS

The paper proposed an improved U-Shaped model for thyroid ultrasound image segmentation. The advantages of the proposed method are: (1) embedding residual units in skip connections to lessen the semantic gap by reducing the parameters and increasing the depth of the model. Batch standardization layer is introduced to increase the backpropagation gradient, which avoids the disappearance of gradient and improves the convergence speed. (2) Introducing attention gate mechanism to enable the model to learn independently and focus on the object structure without additional supervision, thus solving the problem of spatial information loss caused by increasing the horizontal depth of the network. (3) Combining the advantages of Dice loss, cross-entropy loss and Focal-Tversky loss, a new loss function is designed to effectively solve the imbalance of foreground and background categories in medical image segmentation. Experiments show that ReAgU-Net has 6.6%, 4.9%, 4.4% and 5.4% improvement in mIoU, DSC, precision and recall compared with U-Net, and it also has outstanding performance in different data sets.

But automatic segmentation of thyroid ultrasound image is a very challenging task. There are still some images which cannot segment well by the algorithm. Some examples are shown no Fig. 7.

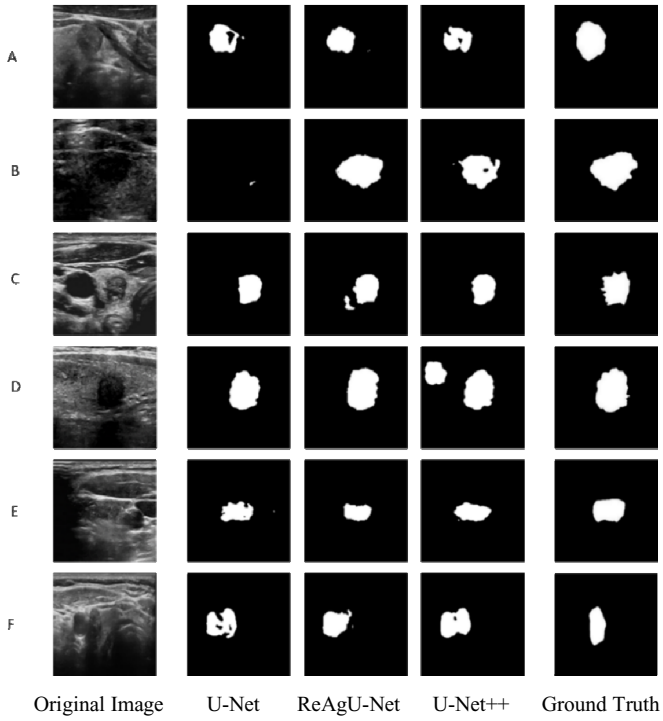


Figure 7. examples of unsatisfactory segmentation results

When the contrast between nodules and background is low, as shown in the first row, the performance of the three models is not good. The dark areas of blood vessels, muscle folds and low echo in the image are similar to the gray level of the lesion. Therefore, when facing these problems, the probability of misjudgment increases. And when multi-nodule appears, such as the case in DDTI dataset, the algorithm also needs to improve.

We believe that besides continuing to collect a large number of samples and enrich training data, we should make full use of the prior knowledge of thyroid pathology and integrate it into the model. This will further improve the robustness of the model.

ACKNOWLEDGMENT

This work is supported, in part, by National Science Foundation of China; the Grant numbers is 81501477.

REFERENCES

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

[2] O. Ronneberger, P. Fischer, T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp. 234-241, 2015.

[3] J. Long, E. Shelhamer and T. Darrell. "Fully Convolutional Networks for Semantic Segmentation" Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431-3440, 2015.

[4] S. Zheng, S. Jayasumana, B. Romeraparedes, et al. "Conditional Random Fields as Recurrent Neural Networks" Proceedings of the IEEE international conference on computer vision. pp. 1529-1537, 2015.

[5] M. Alom, M. Hasan, C. Yakopcic, et al. "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," arXiv preprint arXiv:1802.06955, 2018.

[6] Z. Zhou, M. Siddiquee, N. Tajbakhsh, et al. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation" Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, pp. 3-11, 2018.

[7] Y. Xue, T. Xu, H. Zhang, et al. "SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation" Neuroinformatics, vol. 16, pp. 383-392, 2018.

[8] M. Rezaei, K. Harmuth, W. Gierke, et al. "A Conditional Adversarial Network for Semantic Segmentation of Brain Tumor" International MICCAI Brainlesion Workshop. Springer, Cham, pp. 241-252, 2017.

[9] H. Li, J. Weng, Y. Shi, et al. "An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images" Scientific Reports, vol. 8, pp. 1-12, 2018.

[10] P. Poudel, A. Illanes, D. Sheet, et al. "Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches" Journal of Healthcare Engineering, pp. 1-13, 2018.

[11] J. Ma, F. Wu, T. Jiang, et al. "Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks" International Journal of Computer Assisted Radiology and Surgery, vol. 12, pp. 1895-1910, 2017.

[12] K. He, X. Zhang, S. Ren, et al. "Deep Residual Learning for Image Recognition" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[13] J. Fu, H. Zheng and T. Mei. "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4438-4446, 2017.

[14] T. Lin, P. Goyal, R. Girshick, et al. "Focal Loss for Dense Object Detection" Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.

[15] D. Kingma, J. Ba. "Adam: A Method for Stochastic Optimization" arXiv preprint arXiv:1412.6980, 2014.

[16] L. Pedraza, C. Vargas C , Fabián Narváez, et al. "An open access thyroid ultrasound image database" 10th International Symposium on Medical Information Processing and Analysis. International Society for Optics and Photonics, 2015.