Cascade UNet and CH-UNet for Thyroid Nodule Segmentation and Benign and Malignant Classification

Yiwen Zhang, Haoran Lai, and Wei Yang

School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, Guangdong, China
weiyanggm@gmail.com

Abstract. The thyroid gland secretes indispensable hormones that are necessary for all the cells in your body to work normally. In order to diagnose and treat thyroid cancer at the earliest stage, it is desired to characterize the nodule accurately. We proposed cascade UNet and CH-UNet to segment thyroid nodules and classify benign and malignant thyroid nodules, respectively. Cascade UNet consists of UNet-I and UNet-II, which segment the nodules in the image at uniform resolution and original resolution, respectively. CH-UNet takes segmentation as an auxiliary task to improve classification performance. We verified our method on the test set of the TN-SCUI 2020 Challenge. We achieved 81.73% IoU on segmentation and 0.8551 F1 score on classification, which won the first place in the classification track and was only 0.81% IoU away from the first place in the segmentation track.

Keywords: Thyroid nodule · Segmentation · Classification

1 Introduction

The thyroid gland is a butterfly-shaped endocrine gland that is normally located in the lower front of the neck. It secretes indispensable hormones that are necessary for all the cells in your body to work normally [1]. Until recently, thyroid cancer was the most quickly increasing cancer diagnosis in the United States. It is the most common cancer in women 20 to 34 [2]. In order to diagnose and treat thyroid cancer at the earliest stage, it is desired to characterize the nodule accurately. Thyroid ultrasound is a key tool for thyroid nodule evaluation. It is non-invasive, real-time and radiation-free. However, it is difficult to interpret ultrasound images and recognize the subtle difference between malignant and benign nodules. The diagnosis process is thus time-consuming and heavily depends on the knowledge and the experience of clinicians.1

In recent years, deep learning has been widely used in medical image segmentation, classification, detection and other fields. A large number of practices show that deep learning has good performance in medical image segmentation and classification. UNet [3] and ResNet [4] are one of the most popular networks for medical image segmentation

1 https://tn-scui2020.grand-challenge.org/Home/.
and classification. SE-ResNet is added to the SE-block [5] based on ResNet, and its performance has been improved. In this paper, we use UNet and SE-ResNet to segment thyroid nodules, and classify benign and malignant thyroid nodules.

2 Method

We designed two independent schemes for thyroid nodule segmentation and benign and malignant classification. For segmentation, we proposed a cascade UNet (see Fig. 1). The size of images provided by TNSCUI2020 is inconsistent, the data is first resized to a uniform size and fed to the first UNet. The initial segmentation result of the first stage is used to crop out the ROI region and the ROI is fed to the second UNet. The input image for the second stage keeps the original resolution as much as possible, which can improve the segmentation performance. For classification, we take segmentation as an auxiliary task to improve the classification performance (see Fig. 3).

![Fig. 1. The pipeline of thyroid nodule segmentation. Two cascaded networks are called UNet-I and UNet-II. All images are resized to 512 × 512 and input into UNet-I. Then the 512 × 640 region centered on the initial segmentation result is input into UNet-II.](image)

2.1 Cascade UNet

UNet is the most popular network in medical image segmentation. The encoder-decoder architecture and skip-connection in UNet can capture multi-scale information in medical images. UNet is the preferred network in various medical image segmentation challenges. In TNSCUI2020 challenge, the size of the original images and the size of the thyroid nodule vary greatly, so we proposed a Cascade UNet to deal with these issues. UNet-I aims to locate nodules and predict the approximate size and shape of nodules. UNet-II aims to refine the prediction results of UNet-I in order to obtain more accurate nodule boundaries. TNSCUI2020 provides the ultrasound images with a width ranging from 247 to 1280 pixels and a height ranging from 206 to 818 pixels. Firstly, all the images are resized to 512 × 512 to train UNet-I. There is a trade-off between model performance and computing resource consumption for training UNet-II. Larger input size can keep more information at original resolution, but it also need more GPU memory and time consumption to train the model. Experiments show that 512 × 640 is considered as the best choice. For images which original size is less than 512 × 640, we perform padding...
operation. For images whose ROI in the UNet-I segmentation result is larger than 512 × 640, we resize the ROI to 512 × 640. For the rest images, we crop out the area of 512 × 640 with ROI as the center as input. Therefore, only the image resolution in the second case above is changed. According to statistics, this method can ensure that more than 90% of images keep their original resolution in UNet-II.

![Flow chart of generating pseudo UNet-I segmentation results](image)

**Fig. 2.** Flow chart of generating pseudo UNet-I segmentation results. Firstly, the fitting ellipse of the label is generated. The blue box is the bounding box of the ellipse. The green box is expanded 0.3 times on the basis of the blue box. Then do geometric transformation and cutout on the ellipse in the green box. Finally, put the transformed image back to the original image and crop the area of 512 × 640 as the input of UNet-II.

### 2.1.1 Generating Pseudo UNet-I Segmentation Results from Labels

When training UNet-II, the segmentation result of UNet-I will be used as input. We use labels to generate pseudo UNet-I segmentation results (see Fig. 2). We only hope that UNet-I segmentation results can provide the approximate location and size information of thyroid nodules. It is found through experiments that in the process of training UNet-II, pseudo UNet-I segmentation results should not be too similar to the label, and UNet-II will take a shortcut and copy UNet-I results directly to the output instead of learning how to recognize the original image. Therefore, ellipse fitting, geometric transformation and cutout are used to erase the detailed information in UNet-I results. Geometric transformation includes small scale, translation and 180° random rotation.

### 2.2 Classification with Auxiliary Task

We take segmentation as an auxiliary task to classify benign and malignant thyroid nodules. Segmentation and classification tasks use the same encoder. The classification head (CH) generates a branch from the bottom of UNet. The classification head includes an adaptive average pooling layer, a dropout layer, and a full connection layer. All images are resized to 512 × 512 and then fed to the network. Our experiments show that the segmentation auxiliary task can improve the classification accuracy. However, the performance improvement of classification to segmentation task is limited.
2.3 Multi-scale Test and Test Time Augmentation

Multi-scale test (MST) and test time augmentation (TTA) are common post-processing tricks. MST means transforming the images to multiple different scales in the test phase and then averaging their prediction results. TTA refers to the use of data augmentation and averaging of prediction results in the test phase. They make the prediction results more robust. In the segmentation task, we know that there are targets in each image. Therefore, we can delete the prediction results without any targets in TTA and MST, and only average the remaining results, which can improve the segmentation accuracy.

3 Experiment

3.1 Implementation Details

Training and testing of the network were done in PyTorch\(^2\), all network frameworks are built by segmentation_models_pytorch\(^3\). The encoders of UNet-I and UNet-II are ResNet34 and ResNet101 respectively, and the encoders of CH-UNet is SE-ResNet50. The network was trained using one Nvidia-RTX 2080Ti GPU. We perform on-the-fly data augmentation including scaling, translation, rotation, flipping, elastic deformation, gamma transformation and Gaussian noise. We performed network and hyperparameter evaluation on initial five-fold cross-validation experiments using the training set of the TNSCUI 2020 challenge\(^4\). UNet-I and UNet-II are trained using Binary-cross-entropy loss and Dice loss \((0.4 \times L_{bce} + 0.6 \times L_{dice})\) and CH-UNet is trained using cross-entropy loss in addition to above loss functions. All networks are trained with Adam optimizer with maximum learning rate 0.0002 (with 1cycle learning rate strategy) and L2 weight regularization factor of 0.0001. UNet-I, UNet-II and CH-UNet are trained respectively with a mini-batch size of 16, 7 and 10 while they are all trained for 300 epochs.

3.2 Ablative Evaluation on Segmentation

In order to verify effectiveness of Cascade UNet, Multi-scale test (MST) and test time augmentation (TTA), we did ablation experiments (see Table 1). We use \(384 \times 384\),

---

\(^2\) https://pytorch.org/.

\(^3\) https://github.com/qubvel/segmentation_models.pytorch.

\(^4\) https://tn-scui2020.grand-challenge.org/.
512 × 512 and 640 × 640 as the multi-scale inputs size of TSM. For TTA, we use flipping and gamma transform. Experiments show that both MST and TTA can improve the segmentation performance, and they can improve segmentation performance most when used together.

Table 1. Ablations on segmentation. We validate the effectiveness of Cascade UNet, Multi-scale test (MST) and test time augmentation (TTA) on segmentation task. We show average accuracy (IoU) of five-fold cross-validation on the training set of the TNSCUI 2020 challenge. MST is only used for UNet-I, because the input of UNet-II must keep the original resolution.

| UNet-I | UNet-II | MST | TTA | IoU (%) |
|--------|---------|-----|-----|---------|
| ✓      |         | ✓   |     | 80.37   |
| ✓      |         |     | ✓   | 80.64   |
| ✓      |         | ✓   | ✓   | 80.59   |
| ✓      |         | ✓   | ✓   | 80.85   |
| ✓      | ✓       |     | ✓   | 81.03   |
| ✓      | ✓       | ✓   | ✓   | 81.20   |
| ✓      | ✓       | ✓   | ✓   | 81.19   |
| ✓      | ✓       | ✓   | ✓   | 81.34   |

Table 2. Ablations on classification. We validate the effectiveness of CH-UNet, Multi-scale test (MST) and test time augmentation (TTA) on classification task. We show average F1 score of five-fold cross-validation on the training set of the TNSCUI 2020 challenge.

| SE-ResNet50 | Auxiliary task | MST | TTA | F1 score |
|-------------|----------------|-----|-----|----------|
| ✓           |                | ✓   |     | 0.8130   |
| ✓           |                |     | ✓   | 0.8103   |
| ✓           |                | ✓   | ✓   | 0.8100   |
| ✓           |                | ✓   | ✓   | 0.8107   |
| ✓           | ✓               |     | ✓   | 0.8314   |
| ✓           | ✓               | ✓   | ✓   | 0.8289   |
| ✓           | ✓               | ✓   | ✓   | 0.8286   |
| ✓           | ✓               | ✓   | ✓   | 0.8296   |

3.3 Ablative Evaluation on Classification

In order to verify effectiveness of CH-UNet, Multi-scale test (MST) and test time augmentation (TTA), we did ablation experiments (see Table 2). Settings of MST and TTA
are consistent with the segmentation task (see Sect. 3.2). The results show that adding segmentation as an auxiliary task can improve the classification accuracy greatly. Experiments show that MST and TTA do not improve the classification accuracy. This may be caused by imperfect setting of MST and TTA. However, due to the limited time, we cannot do a more comprehensive search, and finally we did not do MST and TTA for the classification results.

3.4 TNSCUI 2020 Challenge Results

The TNSCUI 2020 Challenge distributed 3,644 ultrasound images as training set. Subsequently, 400 verification sets and 510 test sets were published. The final score depends on the results on the 510 test set. Our final result is obtained by averaging the 5-fold model results (Table 3).

| Track       | IoU (%) | F1 score |
|-------------|---------|----------|
| Segmentation| 81.73%  | /        |
| Classification| /      | 0.8551   |

4 Summary

We propose cascade UNet and CH-UNet for thyroid nodule segmentation and benign and malignant classification. Cascade UNet can ensure segmentation at the original resolution as much as possible even if the image size in the data set varies widely, which can improve the segmentation accuracy. We find that adding segmentation auxiliary task can improve the classification accuracy, while adding classification auxiliary task does not help to improve the segmentation accuracy. Multi-scale test (MST) and test time augmentation (TTA) are proved to be effective for segmentation, but ineffective for classification.

References

1. https://www.btf-thyroid.org/what-is-thyroid-disorder
2. https://www.thyroid.org/wp-content/uploads/patients/brochures/Nodules_brochure.pdf
3. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany (2015)
4. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Presented at the International Conference on the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA (2016)
5. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)