LRTHR-Net: A Low-Resolution-to-High-Resolution Framework to Iteratively Refine the Segmentation of Thyroid Nodule in Ultrasound Images

Huai Chen¹, Shaoli Song², Xiuying Wang³, Renzhen Wang⁴, Deyu Meng⁴, and Lisheng Wang¹(✉)

¹ Institute of Image Processing and Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China lswang@sjtu.edu.cn
² Department of Nuclear Medicine, Fudan University Shanghai Cancer Center, Shanghai, China
³ School of Computer Science, The University of Sydney, Sydney, Australia
⁴ School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, China

Abstract. The thyroid nodule is quickly increasing worldwide and the thyroid ultrasound is the key tool for the diagnosis of it. For the subtle difference between malignant and benign nodules, segmenting lesions is the crucial preliminary step for diagnosis. In this paper, we propose a low-resolution-to-high-resolution segmentation framework for TN-SCUI2020 challenge to alleviate the workload of clinicians and improve the efficiency of diagnosis. Specifically speaking, in order to integrate multi-scale information, several low-resolution segmenting results are obtained firstly and combined with a high-resolution image to refine them and obtain high-resolution results. Secondly, iterative-transfer is proposed to effectively initialize network based on previous trained one on small-scale images. Finally, ensemble refinement is introduced to utilize multiple models to refine the segmentation again. Experimental results showed the effectiveness of the proposed framework. And we won the 2nd place in the segmentation task of TN-SCUI2020.

Keywords: Thyroid nodule segmentation · Thyroid ultrasound · TN-SCUI2020 · Low-resolution-to-high-resolution · Iterative refinement

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]. The term thyroid nodule refers to an abnormal growth of thyroid cells that forms a lump within the thyroid gland [2].

© Springer Nature Switzerland AG 2021
N. Shusharina et al. (Eds.): ABCs 2020/L2R 2020/TN-SCUI 2020, LNCS 12587, pp. 116–121, 2021.
https://doi.org/10.1007/978-3-030-71827-5_15
Statistical studies show that the incidence of this disease increases with age, extending to more than 50% of the world’s population. Until recently, thyroid cancer was the most quickly increasing cancer in the United States. It is the most common cancer in women 20 to 34 [3]. Although the vast majority of thyroid nodules are benign (noncancerous), a small proportion of thyroid nodules contains thyroid cancer. 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 experience of clinicians [4].

The challenge of Thyroid Nodule Segmentation and Classification in Ultrasound Images (TN-SCUI 2020) [4] aims to provide a benchmark to validate all of the state-of-the-art computer-aided diagnosis (CAD) systems for thyroid nodule diagnosis.

In this paper, based on the large public dataset provided by TN-SCUI2020 [4], we propose a framework to complete the segmenting of thyroid ultrasound to assist the doctor in diagnosis. Firstly, due to the great change of lesions in size, we propose a small-resolution-to-high-resolution framework, where several low-resolution segmentation networks are firstly trained to gain results of ultrasound images under low scales. Then, these predictions will be concatenated with high-resolution images to make final results. Secondly, to reuse the power of pre-trained low-resolution networks, we propose iterative-transfer to iteratively transfer them to high-resolution ones. Finally, we propose ensemble refinement to combine outputs and high-level features of multiple models to further refine results again.

2 Methods

As illustrated in Fig. 1, our framework can be divided into two parts. The first part aims to refine high-resolution segmentation by combining multi-scale results. The second part is to effectively refine final results based on multiple models.

2.1 Refining High-Resolution Segmentations Based on Multi-scale Results

The main reasons for us to firstly get low-resolution segmentations and combine them to get final high-resolution predictions are shown as followed:

1) Thyroid nodules’ lesions in ultrasound varies greatly in size, it is not suitable to set network with fixed receptive field to cope all image instances. Thus, preparing several low-resolution segmentations by low-scale images for final high-resolution results is an expected strategy. By which, network with same receptive field in each step can perform well for targets with varied size. And the final high-resolution results can refer to previous low-resolution ones.
Fig. 1. An illustration of the framework. Figure (a) shows the single model, where several low-resolution segmentations will be obtained firstly and promote the final high-resolution segmenting. The iterative-transfer is utilized to initial networks based on pre-trained low-resolution networks. Figure (b) shows the ensemble refinement strategy.

2) Transferring pre-trained networks, especially models based on ImageNet, is a common trick to improve new tasks. However, when transferring models to ultrasound images, two channels of input will be redundant due to the fact that ultrasound images are gray ones. Therefore, combining two previous
results (low-resolution segmentation and medium-resolution ones) with high-scale image as the input is a good strategy to release model’s performance.

It worth noting that we train networks for varied resolutions one by one (low-resolution to medium-resolution to high-resolution). And the pre-trained models for lower resolution will be the initialization for high-resolution ones, which named as iterative-transfer.

2.2 Refining Results Based on Multiple Models

Multi-model ensemble is a common strategy to unite outputs of multiple models to get final results. To better utilize models, [5] propose parallel training to fuse the medium features to make final predictions. In this paper, referring to this idea, we combine models’ outputs with medium features and build a new simple network, constructed by several convolution layers, to refine predictions again.

As shown in Fig.1(b), we build three models (based on pre-trained VGG16 [7], Res18 [6], Efficientnet-b0 [8]) and the fusion net is constructed by three ’3×3’ convolution layers.

3 Experiments

3.1 Dataset

TN-SCUI2020 provides 3644 training data and 910 test data, each of which is labeled by experienced doctors with pixel-level and image-level labels. We focus on the segmentation task, and we divide 360 images as validate data from the training data during the training stage.

3.2 Evaluation Metrics

Segmentation IoU score [4]: IoU score is calculated by the area of the intersection of two regions divided by the area of their union set. It is a good indicator of whether the prediction is consistent with the label.

\[
IoU(Y, \bar{Y}) = \frac{TP}{TP + FN + FP}
\]

(1)

Where \(Y\) is the ground truth and \(\bar{Y}\) is the prediction. And TP, FN, FP are true positives, false negatives and false positives respectively. Considering the evaluation metric, we define the loss function as followed:

\[
loss = \frac{Y \times \bar{Y}}{Y \times \bar{Y} + Y \times (1 - \bar{Y}) + (1 - Y) \times \bar{Y}}
\]

(2)
3.3 Implementation Details

**Low-Resolution Segmenting:** The pre-trained model in ImageNet is utilized as the initial encoder. Under the strategy of fine-tune, the encoder is firstly frozen and only decoder is trained in the first 40 epochs and then both encoder and decoder are jointly trained in the following 100 epochs. Optimizers on these two stages are both Adam. And the learning rate is set as $10^{-3}$ for frozen stage and $10^{-4}$ for fine-tune stage. The batch size is 16.

**Medium-Resolution Segmenting:** The training epoch is set as 40 and the total model is initialized by the pre-trained model in Step1. The optimizer is Adam with learning rate of $10^{-4}$. The batch size is 3 due to the limitation of GPU memory.

**High-Resolution Segmenting:** The training details are similar to Step2. However, the batch size is set as 1 due to the limitation of memory.

**Others:** For the training of models, the soft IoU is set as the loss function and the training rate is reduced to half if the loss does not decrease in five consecutive epochs.

4 Results

| Basebone         | Initialization                | Output scale | Mean IoU(%) |
|------------------|-------------------------------|--------------|-------------|
| VGG16 (step1)    | Trained on ImageNet           | 112 × 112    | 77.2        |
| VGG16 (step2)    | Iterative – transfer          | 224 × 224    | 78.4        |
| VGG16 (step3)    | Iterative – transfer          | 448 × 448    | 79.4        |
| Efficientnet-b0 (step1) | trained on ImageNet   | 112 × 112    | 75.8        |
| Efficientnet-b0 (step2) | Iterative – transfer        | 224 × 224    | 79.3        |
| Efficientnet-b0 (step3) | Iterative – transfer        | 448 × 448    | 79.9        |
| Res18 (step1)    | Trained on ImageNet           | 112 × 112    | 77.1        |
| Res18 (step2)    | Iterative – transfer          | 224 × 224    | 78.6        |
| Res18 (step3)    | Iterative – transfer          | 448 × 448    | 79.7        |
| Ensemble refinement | –                             | 448 × 448    | 80.4        |

The segmentation results on validate data are shown on Table 1. It worth noting that the input of single model with output $448 \times 448$ is the combination of original image and two previous outputs with varied resolutions. We can come to conclusions as followed:

1) Based on the proposed iterative-transfer and low-resolution-to-high-resolution strategy, the segmentation can be improved whatever base bone is.
2) Ensemble refinement, utilizing multiple features and predictions from models, can further improve the performance.
5 Conclusion

In this paper, we propose a novel framework for the segmenting of thyroid nodule in ultrasound image. The proposed low-resolution-to-high-resolution is an effective strategy to refine the predicted results by merging low-resolution segmentations with high-scale image to predict high-resolution output. The proposed iterative-transfer is a good initialization strategy, reusing the power model trained on low-resolution images. Additionally, the ensemble refinement is a convenient mechanism to fuse features and outputs of multiple models to further refine the segmentation.

References

1. https://www.btf-thyroid.org/what-is-thyroid-disorder
2. http://www.thyroid.org/wp-content/uploads/patients/brochures/Nodules_brochure.pdf
3. https://www.cancer.net/cancer-types/thyroid-cancer/statistics
4. https://tn-scui2020.grand-challenge.org/Home/
5. Chen, H., Wang, X., Huang, Y., Wu, X., Yu, Y., Wang, L.: Harnessing 2D networks and 3D features for automated pancreas segmentation from volumetric CT images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 339–347. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_38
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, pp. 770–778 (2016)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014)
8. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks, pp. 6105–6114 (2019)