Ultrasound Image Segmentation of Thyroid Nodules Based On Joint Up-sampling

Junxia Wu¹, Zhenqing Zhang¹, Juanjuan Zhao¹ and Yan Qiang*¹

¹Taiyuan University of Technology, Taiyuan, Shanxi, 030600, China
²wjunjxia0296@link.tyut.edu.cn ⁴Corresponding author’s e-mail: qiangyan99@foxmail.com

Abstract. In view of the problem of inaccurate segmentation of thyroid nodule ultrasound images, a method of thyroid nodule ultrasound image segmentation based on joint up-sampling is proposed. In order to solve the problem of serious speckle noise in the original ultrasound data, preprocessing is performed to enhance the contrast between the nodule area and the background area. In order to achieve accurate positioning of the nodule target area, a joint up-sampling module is designed to fuse the context information of ordinary standard convolution and expansion convolution with different expansion coefficients. Experimental results show that the proposed method achieves 93.19% accuracy and dice similarity coefficient is 0.8558, which is better than other existing thyroid nodule segmentation network models.

1. Introduction

Thyroid tumor is one of the fastest growing malignant tumors in the world¹ and this poses a serious threat to human health. In recent years, the incidence of thyroid disease has increased year by year. At present, there are multiple methods for clinical detection of thyroid. Ultrasound image has the advantages of real-time detection, harmless to human body, low price, repeatable detection, etc². It has become the most commonly used imaging method in thyroid diagnosis. The contrast between the nodule area and the normal tissue area of the original thyroid ultrasound image data is low, and the speckle noise is serious. The use of traditional manual segmentation methods is time-consuming and cumbersome, which mainly depends on experienced doctors. This will cause the doctor's segmentation workload to be large, the diagnosis result is subjective, and problems such as nodule boundaries are difficult to define³. Therefore, the accurate and rapid segmentation of nodule regions with the help of computer image processing technology has high clinical research value⁴.

In recent years, dilated convolution has outstanding performance in the field of image segmentation with its unique advantages. Dilated convolution expands the receptive field of the model without changing the size of the convolution kernel⁵, so that it can better combine the context information to locate the target area better. Then a model based on joint upsampling is proposed aiming at solving the problem of inaccurate thyroid nodule segmentation in ultrasound images. By changing the size of the sampling rate, the context information based on different domains is aggregated, and the joint anti-noise strategy is used to suppress the unrelated areas, thereby obtaining a more accurate feature map. Finally, according to the similarity of the image scale, the shallow connection is used to fuse the shallow information and the global information to obtain a more accurate segmentation result.
2. Proposed Methods

This paper has conducted in-depth research on common ultrasound image segmentation methods of thyroid nodules. To properly combine global features, a network model of thyroid nodule segmentation based on joint upsampling is proposed. And the ability to mine global context information is enhanced by aggregating context information from different domains. The method proposed in this paper has more features and parameters, in order to speed up network calculations, we introduced separable convolution to decouple spatial information and depth information, so as to make more effective use of parameters. When up-sampling the feature maps, a jump connection is used to combine the deep prediction with the shallow prediction, adding local details while considering the global information, and finally achieving the purpose of quickly and accurately segmenting the thyroid nodules.

2.1. The architecture of our model

The thyroid nodule segmentation network framework of combined upsampling ultrasound proposed in this paper is shown in Figure 1. Taking U-Net as the backbone network, the coding layer network is used to extract the depth features of the image and obtain the global information of the segmented image. In general, downsampling 8-fold (8x) feature maps can retain most of the image data. In the network model of this paper, the joint upsampling module is used to replace the original upsampling process to obtain a more accurate 8-fold upsampling feature map. In the joint upsampling process, we adopted an anti-noise fusion strategy to suppress uncorrelated regions. Finally, according to the scale self-similarity of the image data, jump connection is used to add shallow detail information during upsampling, so as to obtain the segmentation results of the nodules.

![Figure 1. Combined with upsampled ultrasound thyroid nodule segmentation network model.](image)

Different from the U-Net network, our network model abandons the last three convolutional layers (Conv3-Conv5), which is up-sampling and skip connections, and replaces it with the structure shown in Figure 1. After conventional convolution block, we adjust the convolution output of the last three layers equal to the Conv3 size (8x), and after splicing up the sampling layer through the pyramid. Then the convolution operations are performed using separable expansion convolutions with expansion ratios of 1, 2, 4, and 8, respectively. Then use the jump connection to up-sample the resulting high-precision feature map to the original image size.

2.1.1. Joint Up-sampling

Up-sampling, also known as image interpolation, readjusts input pictures of any scale to a higher resolution by using resampling or interpolation. In the joint up-sampling ultrasound thyroid nodule segmentation network model, the spatial resolution of the five convolutional feature maps (2x-32x) of the coding layer was successively reduced by two times, and finally a 32x size feature map was obtained. In order to obtain a high-resolution feature map without introducing more complicated calculations, this paper proposes to use a joint up-sampling method to obtain a more accurate 8x size feature map with most of the global information.

2.1.2. An Anti-noise Fusion Strategy

Generally speaking, the true positive area is shared by multiple dilated convolutional feature maps, but the false positive area will be different due to different
expansion coefficients. In the separation dilation convolution, we use two kinds of convolution operations. First, we use the ordinary standard convolution, that is, when $D=1$, so that we can get the precise positioning of the target area, but also ignore many areas related to the object. In the second way, we change the expansion rate $D=2, 4, 8...$ to expand the acceptance range of the field of view, so as to transfer the discriminative knowledge of the sparse highlight area to other target areas. As we all know, choosing a larger expansion coefficient will introduce some unrelated regions, but when considering convolution at different scales of expansion rate, the adjacent differentiated object parts will be used to highlight some truly negative regions. To solve this problem, we adopt a joint anti-noise strategy to avoid introducing more negative regions to reduce the segmentation effect when the coefficient of expansion is large.

In order to anneal false positive regions, we performed an average operation on the convolution ($D=2, 4, 8...$) results of large expansion coefficients. Then the feature map of the standard convolution block $\psi_0$ is added to the averaged feature map $\psi_1$ to obtain the final feature map $\psi$. Here we define as follows: $\psi = \psi_0 + \psi_1 = \psi_0 + \frac{1}{n_d} \sum_{i=1}^{n_d} \psi_i (i = 1,...,n_d)$. Among them, $n_d$ is the number of dilated convolutional blocks. The final feature maps obtained in this way will make the segmentation result more accurate.

2.2 Binary Cross-entropy Loss Function
The binary cross-entropy loss function is a special case of softmax cross-entropy loss. In multi-classification, when there are only two types of categories 0 and 1, it is a two-class classification, which belongs to the problem of logistic regression. The cross-entropy loss function that defines the two classifications is as follows:

$$\text{loss} = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - \hat{y}_i) \log(1 - y_i)]$$

Among them, $y_i$ is the real value of segmentation, and $\hat{y}_i$ is the predicted value of the network model segmentation. The loss function is a logarithmic function, while loss is an integer. When $y_i$ and $\hat{y}_i$ are exactly the same, $\text{loss}=0$. The training goal is to infinitely approach the value of loss to zero. Otherwise, when $y_i$ and $\hat{y}_i$ differ greatly, the loss will be bigger.

3. Experiment
3.1 Datasets
The thyroid ultrasound image data used in the experiment came from a cooperative hospital in Shanxi Province and contained 231 cases. It is a private database, and the ultrasound results are all experienced physicians outline the nodular boundary area. To test our segmentation model, we used 524 thyroid nodule images and their respective gold standard binary masks, and the size of each image was 256*256 pixels.

3.2 Image Preprocessing
The background of thyroid ultrasound images is mostly complicated, and the gray distribution inside the nodules is uneven. Therefore, it is necessary to preprocess the ultrasound image data before performing the segmentation experiment. Among the closed operation, the radius of the disk is selected as 1, and the effect of the preprocessing experiment is shown in Figure 2. The picture (a) is the original thyroid ultrasound image, and the picture (b) is the effect after the preprocessing operation. Compared with the a picture, the (b) picture enhances the contrast between the nodule area and the
background area while preserving the nodule area information as much as possible, and at the same time effectively suppresses the speckle noise in the ultrasound image.

3.3. Segmentation experiment
In this paper, the SVM model, CNN model, U-Net model and our method in this paper are compared in terms of running time, jaccard difference and dice similarity. The PC processor used in the experiment is Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80GHz, memory RAM is 8GB, 64-bit operating system, operating system is Windows10, running on PyCharm2019.1, the programming language is Python3.7, the network uses Keras2.1 framework structure. The Adam optimizer is used for model training, the learning rate is 0.001, the binary cross entropy is used as the loss function, the batch size is set to 5, and the validation split is set to 0.3. The following will analyze from both qualitative and quantitative perspectives. We selected a total of 5 representative sets of original ultrasound thyroid nodule images. Each nodule has its own characteristics to illustrate the advantages and disadvantages of each algorithm in areas with different morphological sizes and uneven grayscale. And it will highlight the robustness of the network segmentation model proposed in this paper.

3.3.1. Qualitative analysis. In Figure 3, the first column is the original thyroid ultrasound image, the second column is the result of preprocessing, the third column is the effect of the nodule region segmented using the SVM method, and the fourth column is the segmentation using the CNN method. Nodule area effect, the fifth column is the nodule area effect segmented using the U-Net method, the sixth column is the nodule area effect segmented using the pyramid joint upsampling method mentioned in this article, the last column is experienced the standard result marked by the doctor.

From the segmentation results, we can intuitively find that when using the support vector machine SVM method for segmentation, there is a more obvious under-segmentation phenomenon, and the segmentation effect is poor; when using the CNN method for segmentation, in the second and When the segmentation boundary of the fourth ultrasound image is not much different from the background pixels, the nodule area cannot be segmented well, and there are
### 3.3.2. Quantitative analysis.

From Figure 4 we can clearly see that the pixel accuracy of the network model increases with the number of training iterations and finally tends to converge in this experiment. Due to the machine learning SVM model itself and the characteristics of the image data, multiple calculations are required to find the optimal hyperplane, and the running speed is slow, so it takes a long time to reach the convergence state. Compared with the U-Net model, the method in this paper has a higher amount of parameters, so it will reach a stable state later than the U-Net model. However, because the method of this paper merges the context information of different domains, it has a higher value than the U-Net model. Accuracy. The CNN model has no jump structure, so the accuracy of the model is poor compared to the U-Net model.

| Original | Processing | SVM | CNN | U-Net | Our Method | Ground Truth |
|----------|------------|-----|-----|-------|------------|--------------|
| ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) | ![Image](image5.png) | ![Image](image6.png) | ![Image](image7.png) |

Figure 3. Comparison of thyroid nodule segmentation results by four methods.
Figure 4. Four methods of training pixel accuracy comparison curve.

In order to verify the effectiveness and feasibility of the method in this paper, Table 1 gives the comparative evaluation values of SVM, CNN, U-Net and the segmentation experiment effect of the method in this paper. Experimental results show that the method obtained in this paper achieves an accuracy rate of 93.19%, which is higher than the other three methods. It is precisely because the pyramid upsampling module designed by this method fuse the positioning results at different sampling rates.

| Methods    | Accuracy | Dice    | Jaccard difference | Time/s |
|------------|----------|---------|--------------------|--------|
| SVM        | 0.7667   | 0.6742  | 0.2553             | 27.15  |
| CNN        | 0.8333   | 0.7208  | 0.1673             | 0.667  |
| U-Net      | 0.9052   | 0.7476  | 0.1496             | 0.678  |
| Our method | 0.9319   | 0.8558  | 0.0824             | 5      |

In general, the effect of the pyramid joint upsampling network framework in this paper is benefited from its deeper network and more network parameters. Through the training of massive ultrasound images, learning to extract more elaborate features, the segmentation effect is more accurate. At the same time, in the dice similarity and jaccard distance, the method of this paper also showed the best effect, indicating that the similarity between the segmentation results of this paper and the gold standard is higher. In terms of test time, because the SVM method calculates the hyperplane at the time of segmentation, the calculation of the secondary planning is large, and the time is much higher than that of several other deep learning methods. The CNN model is more lightweight than U-Net and the model trained by the method in this paper, with the least network parameters and the shortest calculation time. The method in this paper is based on the U-Net model, which is more complicated than the U-Net model, so the calculation time is longer, but we think such a time is acceptable.

4. Summary
Ultrasonic accurate segmentation of the thyroid nodule area is an indispensable prerequisite for the diagnosis of thyroid cancer. To solve this problem, this paper proposes an ultrasound thyroid nodule segmentation network model based on joint upsampling. Before segmentation, we first preprocess the
noisy data to obtain high contrast and low noise processing results. Subsequently, the design uses a deep network to extract the deep features of the image, and combined with the expansion convolution of different sampling rates to accurately locate the nodule area. Among them, a joint anti-noise strategy is used to suppress areas that are not related to the target nodules. Finally, the advantage of the skip structure in segmentation tasks is used to fuse the deep and shallow information of the image to refine the segmentation edge information. Experiments show that the segmentation network model proposed in this paper can accurately segment thyroid nodules, and the optimal segmentation effect is obtained in the existing thyroid nodule segmentation depth model.

Acknowledgements
This work is supported by the National Natural Science Foundation of China under grant no. 61872261.

References
[1] Horvath, E., Majlis, S., Rossi, R., Franco, C., Niedmann, J. P., & Castro, A., et al. (2009). An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab, 94(5), 1748-1751.
[2] Weinstein, S. P., Conant, E. F., & Sehgal, C.. (2006). Technical advances in breast ultrasound imaging. Seminars in Ultrasound Ct & Mri, 27(4), 273-283.
[3] Chen, J., You, H., & Li, K. (2020). A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Computer Methods and Programs in Biomedicine, 185.
[4] Xin, W., & Wenjie, X.. U.. (2016). Ultrasonic image segmentation algorithm for thyroid nodules. Video Engineering.
[5] Yu, F., & Koltun, V.. (2015). Multi-scale context aggregation by dilated convolutions.