SPECT bone imaging thyroid lesion segmentation

Mingyang-Luo\textsuperscript{1,3}, Qiang-Lin\textsuperscript{2,3}, Tongtong-Li\textsuperscript{2,3}

\textsuperscript{1}Key Laboratory of China’s Ethnic Languages and Information Technology of Ministry of Education, Lanzhou 730030;
\textsuperscript{2}School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730030;
\textsuperscript{3}Key Laboratory of Streaming Data Computing and Application, Northwest Minzu University, Lanzhou 730124

Corresponding author:1031216317@qq.com

Abstract. Nuclear medicine SPECT is the main functional imaging modality, which plays an important role in the diagnosis and treatment of thyroid diseases. Hyperthyroidism is a common thyroid disease with symptoms such as exophthalmos, eyelid edema, and vision loss. In order to accurately segment thyroid lesions from SPECT thyroid data, a U-net-based thyroid lesion segmentation model was constructed. First, mirror, rotate, and translate the existing data. At the same time, use the generative adversarial network to expand the data. Then build a segmentation model based on U-net. Finally, perform experimental evaluation and analysis based on a set of real SPECT thyroid data. The results show that: in the traditional data expansion, the IoU value of the improved RS-U-net model is 66.83.

1. Introduction

With the development of medical imaging technology, X-ray, magnetic Resonance Imaging (MRI), Computerized Tomography (CT), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) have been gradually applied to examine human tissue, organs and pathological changes\cite{1,2}. Compared with traditional medical imaging techniques, although SPECT functional imaging has low resolution and is susceptible to noise, it not only captures the structure and morphology of organs, but more importantly, it can obtain information about the molecular characteristics of organ cells and tissues, which can display the functional and metabolic status of organs and tissues. With its unique advantages, SPECT has been widely used in the diagnosis and treatment of mental diseases such as Parkinson’s disease\cite{3,4}, brain diseases\cite{5}, thyroid diseases\cite{6}, etc.

In recent years, machine learning and deep learning have developed rapidly, with unique functions that can automatically extract features from images. In the field of image segmentation, there are a series of deep segmentation models, including full convolution network with semantic segmentation(FCN)\cite{7}, Mask R-CNN\cite{8} network with instance segmentation, and U-net\cite{9} network based on FCN in medical field. At present, deep learning is widely used in medical image, and a large number of applied research results have been produced, such as magnetic resonance image\cite{10,11}, CT image\cite{12,13}, ultrasound image segmentation\cite{14}.

Hyperthyroidism is a common thyroid disease. Its clinical manifestations include exophthalmos, eyelid edema, vision loss and so on. SPECT nuclear medicine imaging plays an important role in the diagnosis of thyroid diseases because of its unique advantages of early detection of diseases\cite{15}. Taking
SPECT thyroid data as the research object, this paper constructs a thyroid lesion segmentation model based on U-net. Firstly, in order to meet the data requirements of the model, the data is mirrored, rotated and translated. Then the model is improved through experiments. Finally, the model is tested with real data. The experimental results show that the IOU of the improved RS-U-net model is 66.83.

2. Materials and methods

The collected experimental data comes from the diagnosis records of patients in the Nuclear Medicine Department of Gansu Provincial People’s Hospital. Each diagnosis record is composed of diagnosis results and SPECT imaging data. Each SPECT imaging data is a data matrix composed of radiation values, represented by a 16-bit unsigned integer, with a size of 1024 (height) x 256 (width). Under normal circumstances, the SPECT whole body bone scan data will contain two data matrices, which are the anterior and posterior imaging matrices.

2.1 Data normalization

Min-Max standardization was used to make a linear transformation of the original radiation values, taking into account the individual differences in radiation doses and simplifying the processing. Specifically, assuming \([a, b]\) is a transformed interval, Max and Min represent the maximum and minimum values in the given original radiation matrix, respectively, the transformation coefficient \(K\) can be solved by formula (1).

\[
k = \frac{b - a}{\text{Max} - \text{Min}}
\]

Each SPECT data matrix corresponds to a coefficient \(k\), which is multiplied by the elements in the matrix to convert matrix elements from intervals \([\text{Min}, \text{Max}]\) to \([a, b]\). The converted SPECT matrix is referred to as the SPECT image.

2.2 Data augmentation.

Currently, there are two main data extension methods used in the field of medical image, one is traditional data extension and the other is data extension based on the generation of antagonistic network. Traditional data expansion mainly generates images that are very close to the real image through geometric transformation (image translation, image rotation, image mirroring, etc.), while generation of antagonistic network generates similar images through deep learning.

2.3 Traditional data augmentation

2.3.1 Image mirror

SPECT nuclear medicine detection produces two images at a time, representing the front image and the back image. However, image loss may occur in the inspection process and in the data transmission and storage process. In order to avoid the possible impact of image loss, this paper does image processing for SPECT image. The common image images have horizontal and vertical images. However, due to the vertical symmetry of human body, this paper only uses the horizontal image processing method.

Suppose the height of the original image is \(h\) and the width is \(w\), then \((x_0, y_0)\) is used to represent the coordinates of the original image, and \((x_I, y_I)\) is used to represent the coordinates of the image after horizontal mirror transformation, as shown in Formula 2. Figure 1b shows an example of image mirroring.
2.3.2 Image translation

In the process of disease examination, the patient's lying posture is not standard enough, and the collected images will inevitably include the phenomenon of center offset. In order to make up for this phenomenon, it is necessary to translate the image appropriately. Image translation is to move the image up and down, left and right on the basis of preserving the focus area. For SPECT bone imaging, the radiation value of the lesion area is very high, the concentration point is very obvious, and the lesion site is relatively small and easy to distinguish from other non lesion sites. Therefore, SPECT bone imaging is very suitable for image translation, which not only retains the information of lesion location, but also expands the data.

Suppose that the coordinates of the image pixels are \((x, y)\), the translation amount is \(\Delta x, \Delta y\), and the coordinates of the translated pixels are \((x', y')\). The image translation calculation formula (3) is shown in Figure 2C.

\[
\begin{bmatrix}
    x' \\
    y
\end{bmatrix} =
\begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix}
\begin{bmatrix}
    x \\
    y
\end{bmatrix} + 
\begin{bmatrix}
    \Delta x \\
    \Delta y
\end{bmatrix}
\]

(3)

2.4 Image rotation

Due to the individual differences of patients, there is a deviation of focus imaging in the process of disease examination, and the movement of patients will lead to the deviation of imaging. In order to make up for the deviation and offset of the image focus, the original image is rotated appropriately, and the rotation amplitude of the image is set as \(\theta \in (0^\circ, 5^\circ)\).

Image rotation is to rotate an angle in the direction of anticlockwise or clockwise with a point in the image as the origin. The coordinate point \((x, y)\) is set to \((x_I, y_I)\) after the rotation of theta angle, as shown in Formula 4. Fig. 2D shows an example of the image rotating to the right.

\[
\begin{bmatrix}
    x_I \\
    y_I
\end{bmatrix} =
\begin{bmatrix}
    \cos \theta & \sin \theta \\
    -\sin \theta & \cos \theta
\end{bmatrix}
\begin{bmatrix}
    x \\
    y
\end{bmatrix}
\]

(4)

2.5 Generative Adversarial Networks

Generative Adversarial Networks (GAN) consists of two main networks: one is Gnerator Network, called Generator, which generates simulated data; the other, Discriminator Network, is called a discriminator, which determines whether the simulated data generated by the generator is true or simulated.

Since convolutional neural networks are suitable for processing image data, this article uses Deep Convolutional Generative Adversarial Networks (DCGAN) as the generative confrontation model. Like GAN, the network structure is mainly composed of the generator Gnerator (G) and the discriminator Discriminator (D) is composed, and its core idea is to achieve balance through G and D competing with
each other. Among them, G collects and learns the distribution of real sample data, and finally generates new sample data with given noise input. The discriminator D is mainly to discriminate whether the input sample is a real sample or a generated sample. Through the mutual game and confrontation learning of the two networks, G improves the generation ability of its own sample data, and finally achieves that the generator can generate the discriminator and cannot judge whether it is good or bad. Good and bad data. On the contrary, the discriminator D improves its discrimination ability and can judge that the input data is a generated sample instead of a real sample. Its basic structure is shown in Figure 1, Z is the input random noise, G is the generator, G(z) is the generated picture, X is the real data, and D is the discriminator. The structure of DCGAN is shown in Figure 2:

![Figure 2. Basic structure of DCGAN](image)

The experimental data are respectively expanded by the above traditional image data expansion and generated by the generated countermeasure network. The generated data sets are shown in Table 1 below.

| Dataset                          | Total |
|---------------------------------|-------|
| Original data set               | 56    |
| Traditional data augmentation   | 1020  |
| Generative Adversarial Networks | 1020  |

2.6 Segmentation models

2.6.1 U-Net based segmentation.

The U-net network model is an encoder decoder structure, which consists of the compression path on the left side and the expansion path on the right side. Compression channel is an encoder, which is used to extract image features by pixel by advanced down-line sampling operation; the extended channel is a decoder, and then carries out up sampling operation to restore the position information of image. Finally, the shallow feature and deep feature are fused by skip connection, so as to obtain more context information.

2.6.2 Residual module

Because deep convolution network can enrich features by deepening the depth of the network, researchers tend to use deeper network structure to extract deeper features. However, with the increase of training iterations, the deeper network structure is prone to gradient disappearance and network degradation.

For an ordinary stacked convolution network structure, when the input is x, the characteristic is \( h(x) = f(x) \). Then the residual mapping obtained by residual unit training is \( f(x) = H(x) - X \). The residual unit calculation formula is as follows (5) and (6):

\[
y_i = F(x_i, w_j) + h(x_i) \tag{5}
\]

\[
x_{i+1} = f(y_i) \tag{6}
\]

Where \( x_i \) and \( x_{i+1} \) represent the input and output of the i-th residual cell, \( F \) represents the learned residual cell characteristics, \( w_i \) is the i-th layer parameter, \( f \) represents the relu activation function, and
\( h(x_i) = x_i \) represents the identity mapping. The structure of residual unit is shown in Figure 3:

![Residual Mapping Diagram](image)

**Fig 3.** The structure of a residual module.

Due to the insufficient depth of U-net network structure, in order to obtain more deep feature information from SPECT data set, residual block is added to replace the ordinary convolution layer. Residual block can deepen the depth of U-net network and effectively solve the problem of gradient disappearance with the deepening of network layers. Improved RS-U-Net is shown in Figure 4.

![RS-U-Net Structure Diagram](image)

**Fig 4.** RS-U-Net structure diagram

2.6.3 Focal loss

Since there are only one or two lesion areas in an image, and the pixel ratio of the lesion area is small, the background is too large compared to the lesion area, resulting in imbalance of positive and negative samples, making network training more difficult. In the training phase, focal loss function is used as the loss function of the segmentation part. Focal loss adds a gamma factor and an alpha balance factor to the binary cross loss function. Among them, the gamma factor reduces the loss of easy to classify samples, and makes the model focus more on the difficult and wrong samples; while the alpha factor is used to balance the uneven proportion of positive and negative samples. The formula of focal loss function is shown in formula (7):
\[ L = \begin{cases} -\alpha(1 - y') \log y', y = 1 \\ -(1 - \alpha)y' \log(1 - y'), y = 0 \end{cases} \]  

(7)

Where \( y \) is the label of the real sample (1 positive and 0 negative), and \( y' \) is the predicted output through the sigmoid activation function (the value is between 0-1).

3. Experiment and evaluation

3.1 Experimental design

In this experiment, the commonly used evaluation indexes in medical image segmentation are used to evaluate the segmentation results quantitatively. It includes class pixel accuracy (CPA), Recall, intersection over Union (IoU) and pixel accuracy (PA). For any pixel, it will be divided into one of the following four categories:

- **True Positive (TP)**: The area predicted by the model is the focus area, and the actual area is the focus area;
- **False Negative (FN)**: The normal area predicted by the model is actually the focus area;
- **False Negative (FP)**: The model is predicted to be the focus area, and the actual area is normal;
- **True Negative (TN)**: It is predicted that the model is a normal area, but it is actually a normal area.

Based on the above definition, this paper defines the evaluation indexes CPA, Recall, IoU and PA as follows:

\[ CPA = \frac{TP}{TP + FP} \]  

(8)

\[ Recall = \frac{TP}{TP + FN} \]  

(9)

\[ IoU = \frac{TP}{TP + FN} \]  

(10)

\[ PA = \frac{TP + TN}{TP + TN + FP + FN} \]  

(11)

The experiment runs on Ubuntu 16.04 operating system, the CPU is inter Xeon (R) silver 4110, the number of CPU cores is 16, the memory is 62g, and the graphics card is geforcrtx2080ti x 2. The main network is resnet50 and resnet101, and the input image size is 256 x 256, The RS-U-net learning rate is 0.0001, the learning momentum is 0.9, the weight decay rate is 0.0001, and the optimizer is Adam. The model trains 250 epochs, and each epoch trains 720 times.

3.2 Experimental results and analysis

The preprocessed data is divided into two data sets: traditional data expansion and generation countermeasure network - Net and RS_ U - Net. The loss and PA curves of two different data sets and different network training are shown in Figure 5 and Figure 6.
Figure 5. Curve of traditional data augmentation loss and AP (a is RS-U-net model B is u-net model)

Figure 6. DCGAN loss and AP curve (a is RS-U-net model B is u-net model)

It can be seen from Figure 7 and figure 8 that the training performance of RS-U-net model is better than that of U-net model curve on both traditional data sets and confrontation generated network data sets, and the PA value of RS-U-net model is higher and the loss value is lower on the training set. This is because the depth of the original U-net model is slightly insufficient, and the residual block is added to the RS-U-net network structure, which not only deepens the depth of the model, but also carries out feature fusion, so as to obtain multi-level features. However, the original SPECT thyroid data focus area is small, the background is too large, there is a problem of positive and negative sample imbalance. RS-U-net introduces the loss function of focal loss to increase the number of positive and negative samples $\gamma$ Factor sum $\alpha$ In this way, the model can train the positive samples more fully and improve the segmentation accuracy. Table 2 shows the quantitative results of the two models on different data sets.

Table 2. Model segmentation results

| Data set         | Model    | PA   | IoU   | CPA  | Recall |
|------------------|----------|------|-------|------|--------|
| Traditional data | U-Net    | 99.18| 59.42 | 70.12| 77.32  |
| expansion        | RS-U-Net | 99.61| 66.83 | 75.73| 81.82  |
|                  | U-Net    | 98.98| 52.37 | 66.18| 71.44  |
| GAN              | RS-U-Net | 99.11| 54.15 | 67.98| 73.33  |

It can be seen from Table 2 that the results of the traditional data expansion of the two segmentation models are better than the results of generating the data expansion of the confrontation network. The data generated by the traditional data expansion method is very close to the original data. This is because the traditional data expansion method expands the data through geometric transformation, which not only preserves the lesion area of the original data, but also expands the data. The generation of the confrontation network is to use the generator and the discriminator to game and fight each other to generate images similar to the original data. Due to the lack of original data, the features learned by the generator in the generation confrontation network are insufficient. The generated data only retains the
lesion area, and the overall details of the image are severely lost, resulting in model performance degradation and reduced model segmentation accuracy.

In order to compare the optimization results of different loss functions on the model, Table 3 shows the quantitative results of using three loss functions on traditional data expansion. It can be seen that the RS-U-Net model uses the Focal loss loss function to achieve the best results. This is because the purpose of the Focal loss loss function is to solve the problem of a serious imbalance in the proportion of positive and negative samples. By introducing the $\gamma$ and $\alpha$ balance factors, the weight of the background can be suppressed to increase the weight of the target. Although Dice loss and BCEWithLogitsLoss also solve the scenario of imbalanced positive and negative samples by increasing the weight of the foreground area, they all have problems such as severe gradient transformation and difficulty in training. Experiments have shown that Focal loss is more suitable for segmentation of thyroid data.

| Loss function          | PA  | IoU  | CPA | Recall |
|------------------------|-----|------|-----|--------|
| Focal loss             | 99.61 | 66.83 | 75.73 | 81.82  |
| Dice loss              | 99.09 | 60.63 | 69.22 | 74.89  |
| BCEWithLogitsLoss      | 98.25 | 57.87 | 68.22 | 70.43  |

In order to visually observe the segmentation results of the two models on the two data sets, figure 7 shows the segmentation results of thyroid in the test set. The green area is the result of manual annotation, and the red area is the result of model prediction.

(1)

(2)

Figure 7. Thyroid segmentation results ((1) the best segmentation results, (2) the worst segmentation results; a traditional data expansion U-net segmentation results; B traditional data expansion RS-U-net partition result c generates countermeasure network extension, U-net partition result D generates countermeasure network extension RS-U-net segmentation results)

It can be seen from Figure 10 that the results of traditional data expansion segmentation are better than those of the generation of adversarial network segmentation. Among them, the RS-U-Net model has the best segmentation effect in the traditional data expansion data set, and the IoU value reaches 66.83. This is because the data generated by traditional data expansion is the closest to the original data. The expansion of the data through the traditional geometric transformation method not only preserves the lesion area but also expands the data. However, the generation of the confrontation network
expansion data requires a sufficient amount of raw data to provide training. Due to the lack of raw data, the generator that generates the confrontation network cannot learn enough SPECT image features, and the generated data samples are single, insufficient details, and high similarity. Experimental results show that for small data thyroid data, the segmentation effect of the traditional data expansion method is better than that of the generated adversarial network data expansion method, and the improved RS-U-Net model is better than the original U-Net model.

4. Summary and Prospect
Aiming at the automatic segmentation of thyroid lesions in SPECT imaging, this paper constructs an automatic segmentation model based on U-Net network, and proposes an RS-U-Net model that adds residual blocks to the U-Net model and modifies the focal loss function. First, explain the preprocessing and expansion process of SPECT bone scan data; then explain the built-in depth segmentation model in detail; finally, based on the original SPECT data, the built segmentation model is experimentally verified, and the experimental results show the improvement proposed in this paper. The feasibility of the method for segmentation of thyroid lesions.

In the future, the work of this paper will be expanded from the following aspects:
Firstly, more real SPECT data are obtained to evaluate the authenticity and usability of the depth segmentation model.
Secondly, optimize the model to improve the segmentation accuracy.
Finally, expand the research field, build the segmentation model for multiple diseases and multiple lesions, and develop the computer-aided diagnosis system of SPECT.

Acknowledgments
This work is supported by the special funds of basic scientific research expenditure of the central colleges and universities of Northwest Nationalities University(Yxm2021101).

References
[1] Antony J, Mcguinness K, Moran K, et al. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks[J]. Machine Learning and Data Mining in Pattern Recognition, 2017, 10358: 376-390.
[2] Liu Fang, Zhou Zhaoye, Samsonov A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection[J]. Radiology, 2018, 289(1): 160-169.
[3] Iizuka T, Fukasawa M, Kameyama M. Deep learning-based imaging classification identified cingulate is-land sign in dementia with Lewy bodies[J]. Scientific Reports, 2019, 9(1): 8944
[4] Martinezmurcia F J, Ortiz A, Gorriz J M, et al. A 3D Convolutional Neural Network Approach for the Diag-nosis of Parkinson’s Disease[C]//Proceedings of Inter-national Work-conference on the Interplay between Natural and Artificial Computation. 2017: 324-333.
[5] Spier N, Christoph R, Rupprech C, et al. Defect detection in cardiac SPECT using graph-based convolutional neural networks[J]. Journal of Nuclear Medicine, 2018, 59(1): 1541
[6] Ma L, Ma C, Liu Y, et al. Diagnosis of Thyroid Diseases Using SPECT Images Based on Convolutional Neural Network[J]. Journal of Medical Imaging and Health Informatics, 2018, 8(8):1684-1689
[7] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
[8] He Kaiming, Gkioxari G, Piotr Dollár, et al. Mask R-CNN [C]// Proc of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 2980-2988.
[9] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[J]. Medical Image Computing and Computer-Assisted Intervention, 2015, 9351: 234-241.
[10] Poudel R P K, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation[J]. Lecture Notes in Computer Science, 2017, 10129: 83-94.
[11] Cui Zhipeng, Yang Jie, Qiao Yu. Brain MRI segmentation with patch-based CNN approach [C]// Proc of the 35th Chinese Control Conference. Piscataway, NJ: IEEE Press, 2016: 7026-7031.
[12] Christ P F, Elshaer M E, Ettlinger F, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields [C]// Proc of Medical Image Computing and Computer Assisted Intervention, 2016: 415-423.
[13] Vorontsov E, Tang A, Pal C, et al. Liver lesion segmentation informed by joint liver segmentation [C]// Proc of the 15th International Symposium on Biomedical Imaging. Piscataway, NJ: IEEE Press, 2018: 1332-1335.
[14] Ghesu F C, Krubasik E, Georgescu B, et al. Marginal space deep learning: efficient architecture for volumetric image parsing[J]. IEEE Trans on Medical Imaging, 2016, 35(5): 1217-1228.
[15] Ma L, Ma C, Liu Y, et al. Thyroid Diagnosis from SPECT Images Using Convolutional Neural Network with Optimization[J]. Computational Intelligence and Neuroscience, 2019: 1-11