Signal, Image and Video Processing (2023) 17:2297–2303
https://doi.org/10.1007/s11760-022-02446-2

ORIGINAL PAPER

RNN-combined graph convolutional network with multi-feature fusion for tuberculosis cavity segmentation

Zhitao Xiao[1,3] · Xiaomeng Zhang[2] · Yanbei Liu[1] · Lei Geng[1] · Jun Wu[4] · Wen Wang[1] · Fang Zhang[1]

Received: 29 August 2022 / Revised: 16 November 2022 / Accepted: 10 December 2022 / Published online: 5 January 2023

Abstract
Tuberculosis is a common infectious disease in the world. Tuberculosis cavities are common and an important imaging signs in tuberculosis. Accurate segmentation of tuberculosis cavities has practical significance for indicating the activity of lesions and guiding clinical treatment. However, this task faces challenges such as blurred boundaries, irregular shapes, different location and size of lesions and similar structures on computed tomography (CT) to other lung diseases or tissues. To overcome these problems, we propose a novel RNN-combined graph convolutional network (R2GCN) method, which integrates the bidirectional recurrent network (BRN) and graph convolution network (GCN) modules. First, feature extraction is performed on the input image by VGG-16 or ResNet-50 to obtain the feature map. The feature map is then used as the input of the two modules. On the one hand, we adopt the BRN to retrieve contextual information from the feature map. On the other hand, we take the vector for each location in the feature map as input nodes and utilize GCN to extract node topology information. Finally, two types of features obtained fuse together. Our strategy can not only make full use of node correlations and differences, but also obtain more precise segmentation boundaries. Extensive experiments on CT images of cavitary patients with tuberculosis show that our proposed method achieves the best segmentation accuracy than compared segmentation methods. Our method can be used for the diagnosis of tuberculosis cavity and the evaluation of tuberculosis cavity treatment.

Keywords Tuberculosis cavity · Computed tomography · Graph convolutional networks · Recurrent neural network

1 Introduction

Tuberculosis is a chronic infectious disease caused by the bacterium Mycobacterium tuberculosis. Tuberculosis is also one of the most common infectious diseases in China. Tuberculosis cavity is an important imaging sign of tuberculosis [1], which has practical significance to indicate the lesion activity and guide clinical treatment. The treatment of tuberculosis usually requires the combination of anti-tuberculosis drugs with different mechanisms. The changes of tuberculosis cavity before and after drug use can help doctors to administer drugs. Therefore, tuberculosis cavity segmentation could reduce the burden on doctors.

The use of computed tomography (CT) imaging to visualize and evaluate cavitary lesions in tuberculosis is critical for the diagnosis and treatment of this disease [2]. Since manual segmentation of cavitary lesions from CT images is time-consuming, labor-intensive, and faces inter- and intra-observer variability, automatic segmentation of cavitary lesions from CT images is ideal. For the application of medical image segmentation, the correct segmentation of the structure by the model in the medical image field is an important key to success [3]. To measure cavitation accurately, it is necessary to segment cavitary lesions from CT scans. Segmentation results can provide detailed spatial distribution of lesions and precise volumetric measurements, which have important implications for treatment decision-making, treatment effect evaluation, and prognosis prediction.

✉ Yanbei Liu
liuyanbei@tiangong.edu.cn

1 School of life Sciences, Tiangong University, Tianjin 300387, China

2 School of Artificial Intelligence, Tiangong University, Tianjin 300387, China

3 Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin 300387, China

4 School of Electronic and Information Engineering, Tiangong University, Tianjin 300387, China

Segmentation of tuberculosis cavities is a great challenge for the following reasons. First, the size and shape of cavitary lesions in tuberculosis varies widely with the severity of cavitary lesions in tuberculosis. Small lesions may contain only a few pixels, while large lesions may occupy an entire segment of the lung. Second, lesions have a complex spatial distribution that can be dispersed in different parts of the lung. Third, tuberculosis cavitary lesions may also have some similarities in appearance with other lung diseases or lung structures. These factors, combined with the low contrast of CT images, make it difficult to delineate the boundaries of cavitary lesions in tuberculosis. Figure 1 shows two examples of CT scans of cavitary lungs in tuberculosis, illustrating the difficulty of accurate segmentation. In recent years, with the deepening of graph neural network research, graph neural network has also been successfully applied in the field of image segmentation [4]. Some researchers apply RNN to the field of image segmentation [5,6], making the segmentation task flexible and efficient. However, to our knowledge, segmentation of the tuberculosis cavity has been rarely studied to date.

To exploit the advantages of GCN and RNN, we propose a novel RNN-combined graph convolutional network (R2GCN) method. As shown in Fig. 2, the proposed R2GCN consists of two key modules: graph convolutional network (GCN) module and bidirectional recurrent network (BRN) module. Specifically, we first use VGG-16 or ResNet-50 to extract features from the input images to obtain feature map. We take the vector for each location in the feature map as input nodes. Graphs are constructed by computing the similarity between nodes. Then, GCN is used to extract the relationship between nodes to improve the detection accuracy. Compared with other models, R2GCN have more flexible skip connections and thus, can explore multiple relationships between nodes in the graph. In addition, BRN was used to extract the sequence information of nodes, obtain the context information of features, and make full use of features to accurately segment the boundary of tuberculosis cavity. Finally, the features learned from these two modules are fused to obtain subsequent segmentation results. The strategy of RNN-combined graph convolutional network can make full use of the local and global information of feature maps and the relevant information between nodes, thereby further improving the accuracy of segmentation results. The main contributions of this paper are summarized as follows:

(1) We propose a novel RNN-combined graph convolutional network to improve the ability of feature extraction. It not only obtains the information between similar nodes, but also captures the contextual information of feature map.
(2) We present a parallel graph convolutional network framework module to extract topological information. It could utilize the correlation between similar nodes in the feature map. Simultaneously, we combine a BRN module to obtain the contextual information of the feature map. Two modules could work together to obtain richer features.
(3) By fusioning the features of the GCN and BRN modules, the segmentation of tuberculosis cavity can be effectively improved. The experimental results verify the superiority and effectiveness of R2GCN compared with existing segmentation methods.

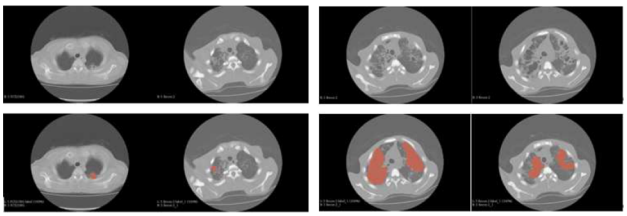

Fig. 1 Examples of cavitary lesions in tuberculosis. The first row shows lung CT images, and the second row shows manual segmentation results of cavitary lesions in tuberculosis

2 Related work

In the deep learning era, many segmentation methods based on convolutional neural network (CNN) have been proposed to accurately segment target objects in the fields of natural and medical images. A fully convolutional network (FCN) [7] with skip layers is proposed for semantic segmentation to preserve spatial localization information. Later, FCN was also used for many medical image segmentation tasks, and achieved good results [8]. Inspired by FCN, U-Net [9] proposes a method to predict accurate output by combining high-resolution features with up-sampled features by exploiting high-level contextual information. After U-Net appeared, it was widely used for medical image segmentation. Inspired by u-net, many deep learning segmentation networks were proposed [10]. A new network pyramid scene parsing network (PSPNet) [11] is proposed, which uses pyramid pooling to capture more scale context information and aggregate context information in different regions, thereby improving the ability to obtain global information.

In recent years, graph neural networks have been actively researched. Many methods have been successfully applied to the field of image segmentation. Gao et al. [12] combined a graph neural network with a U-Net architecture to achieve the task of image pixel-level prediction. Li et al. [13] proposed an instance co-segmentation method based on graph convolutional networks. Zhang et al. [14] presented an affinity fusion graph framework that combines adjacency graphs and graphs based on nuclear spectral clustering for natural image segmentation. Lu et al. [15] adopted an adjacency matrix of a Gaussian kernel function for image semantic seg-

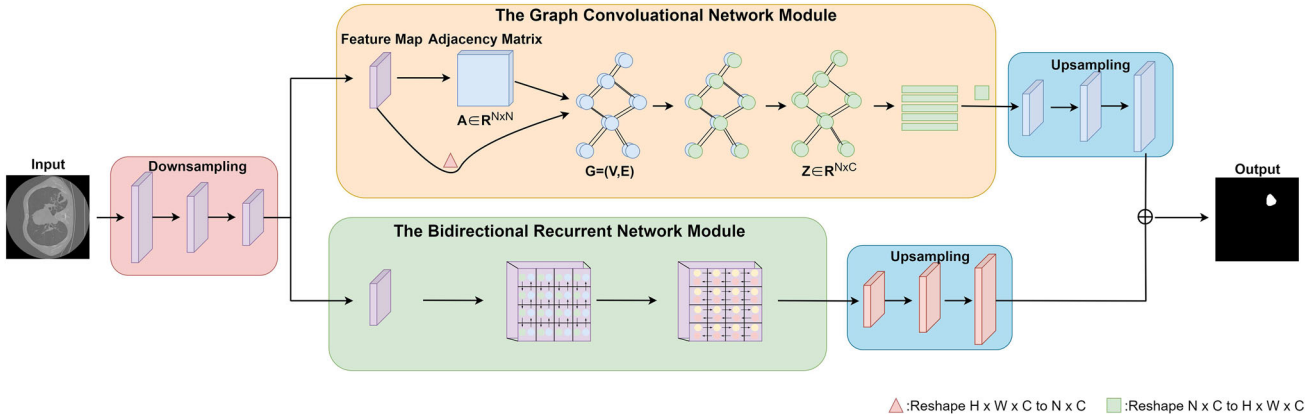

Fig. 2 The structure of our proposed R2GCN network. The upper part is to learn the relationships between nodes through GCN, and the lower part is to learn the contextual information of the feature map through the BRN

mentation. There is a natural correspondence between graph embedding tasks and instance embedding-based image segmentation tasks [16].

Convolutional neural networks have been widely used to segment lung structures from CT images[17]. Cascaded neural networks with non-local modules [18] were proposed to exploit structured relations for lobe segmentation. For lung lesions, a central focused CNN [19] was proposed to segment lung nodules from heterogeneous CT images, and Fan et al. [20] employed reverse attention and edge attention to segment COVID-19 lung infection. Despite extensive work to date on lung structure and lesion segmentation, there is a lack of deep learning models for the challenging task of tuberculosis cavity segmentation.

3 Method

In this section, we illustrate the R2GCN network by constructing the graph convolutional network and bidirectional recurrent network modules in detail.

3.1 The overall R2GCN framework

In this subsection, an RNN-combined graph convolutional network exploits broader feature map information and is designed for tuberculosis cavity segmentation. It not only takes into account the contextual information of feature map but also learns its feature of topological structure. Concretely, the entire R2GCN framework mainly consists of several structures, which are carried out according to the following steps. First, the feature map is obtained by extracting the features of the tuberculosis cavity image with VGG-16 or ResNet-50. Subsequently, we take the vector for each location in the feature map as input nodes, a graph is constructed by calculating the similarity between nodes to obtain an adjacency matrix and adjust the size of the feature map. The adjusted feature map is used as the input of the graph convolutional network to extract the topology information. Meanwhile, the contextual information of the feature map in the BRN is extracted to further generate more accurate segmentation boundaries. Finally, the features extracted by these two modules are fused by Eq. 1 to obtain the final experimental results.

$$P = \mu \frac{1}{1 - e^{O^{\leftrightarrow}}} + (1 - \mu) \frac{1}{1 - e^{Z}} \tag{1}$$

where P represents the feature after models fusion, Z represents the features extracted by the GCN model, $O^{\leftrightarrow}$ represents the features extracted by the BRN, and μ the fusion coefficient of the two models.

3.2 Graph convolutional networks module

Graphs are ubiquitous in the real world, and common graph data include social networks, citation networks and biological networks. The graph structure is usually denoted as $G = (V, E)$, V represents the node set, and E represents the edge set of the graph. Each node v has its own feature vector, and the connections between nodes contain important information. In graph data, they usually use a single sample as a node and use the connection between samples as a connection to construct a graph.

After extracting features, the vector for each location in the feature map will be regarded as a node in the graph, rather than a pixel in the original input images. Through this method, the number of nodes can be decreased significantly, thereby reducing computational complexity as well as improving computational efficiency. We design an algorithm to represent whether nodes are connected or not. First, we calculate the similarity between each node and other nodes through Eq. 2 and then, judge whether the nodes are con-

nected with each other by comparing with the λ.

$$a_{i,j} = \frac{\sum_{c=1}^{C} \left(x_i - x_j\right)^2}{C} \tag{2}$$

$$a_{i,j} = \begin{cases} 0, a_{i,j} > \lambda \\ 1, a_{i,j} \leq \lambda \end{cases} \tag{3}$$

where C represents the number of channels, $a_{i,j}$ represents the similarity of node x_i and node x_j, and the similarity of the node is less than the λ as the two nodes are connected. Among them, $a_{i,j}$ ($i, j \in 1, \cdots, N$) constitutes the adjacency matrix A, and the self-connection relationship is added to the matrix A to obtain the matrix $\tilde{A}$.

We use GCN to classify the nodes of the graph model that we have established. The GCN is one of the deep learning methods to process graph structure. The GCN model expression in this paper is as follows:

$$Z = \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X \Theta \tag{4}$$

where $\tilde{A} = A + I_N, \widetilde{D} = \sum_j \widetilde{A}_{ij}$, $X \in R^{N \times C}$, N represents the number of nodes, C represents the feature vector dimension of the node, $\Theta \in R^{N \times F}$ represents the trainable parameter matrix, F represents the output dimension, and $Z \in R^{N \times F}$ represents the output of the graph convolution. After the output Z of GCN is obtained, the graph data are converted into image data.

3.3 Bidirectional recurrent neural network module

RNN is a special kind of neural network that can be used to capture various patterns present in the data by maintaining state variables [21]. RNNs can model sequence data and share parameter sets over time, which is one of the main reasons why RNNs can learn patterns at each moment of the sequence. Bidirectional recurrent neural network (BRNN) is a neural network composed of two RNNs in opposite directions, so it can process information in both directions at the same time. Based on this feature of BRNN, we introduce two BRNN networks to form the BRN to process image data. By processing the information in the horizontal and vertical directions, the contextual information of the feature map can be captured and the target area can be better segmented.

In our method, the BRN is introduced to capture horizontal and vertical information in feature map. Concretely, we take the extracted features $X \in R^{H \times W \times C}$ as input, where H, W and C represent the height, width and the number of channels. We think of it as $H \times$ W block $p_{h,w} \in R^{H_P \times W_P \times C}$. First, we process the vertical direction of the data. In the vertical direction, the RNN will read information of a node at each step to generate a prediction $o^{\star}_{h,w}$ and updates its state $z^{\star}_{h,w}$:

$$o^{\downarrow}_{h,w} = f^{\downarrow}\left(z^{\downarrow}_{h-1,w}, p_{h,w}\right), \text{ for } h = 1, \cdots, H \tag{5}$$

$$o^{\uparrow}_{h,w} = f^{\uparrow}\left(z^{\uparrow}_{h+1,w}, p_{h,w}\right), \text{for } h = H, \cdots, 1 \tag{6}$$

Once the first two vertical RNNs have processed the whole input X, we concatenate their projections $o^{\downarrow}_{h,w}$ and $o^{\uparrow}_{h,w}$ to obtain a composite feature map $O^{\updownarrow}$. After obtaining the concatenated feature map $O^{\updownarrow}$, we sweep over each of its rows with a pair of new RNNs, $f^{\rightarrow}$ and $f^{\leftarrow}$. Use $f^{\rightarrow}$ and $f^{\leftarrow}$to process $O^{\updownarrow}$ in the horizontal direction. The process is same with the previous step, and the cascaded feature map is obtained $O^{\updownarrow} = \left\{h^{\leftrightarrow}_{h,w}\right\}^{w=1,\cdots,W}_{h=1,\cdots,H}$, Each element for $O^{\updownarrow}$ represents the prediction of each node combined with the entire feature map information.

4 Experiments

In this section, the results of a series of comparative experiments are undertaken to validate the effectiveness of the R2GCN framework on tuberculosis cavity datasets. All the experiments are implemented on the python 3.6.12 software, and the system is Ubuntus 16.04.7 based on NVIDIA RTX 2080 Ti GPU.

4.1 Dataset

The dataset used in this paper contained a total of 59 sensitive patients and 373 lesion images is provided by Tianjin Haihe Hospital and was annotated by three experts from the hospital. The dataset mainly contains four types of voids: thin-walled voids, thick-walled voids, worm-eaten-like voids, and consolidation voids. Thin-walled voids and thick-walled voids have clear boundaries, while worm-eaten voids and consolidation voids have blurred boundaries. In this paper, four types of tuberculosis cavity are regarded as one type to be segmented, and the final segmentation results are analyzed.

4.2 Evaluation indices

For the image segmentation, two commonly used indices, the dice similarity coefficient (DICE) and intersection over union (IOU), are capitalized on computing the quantitative evaluation.

DICE represents the similarity between the predicted result and the true value. The mathematical formula is defined as follows:

$$DICE = \frac{2TP}{FP + 2TP + FN} \tag{7}$$

Table 1 Segmentation results of different methods

Methods	VGG-16		ResNet-50	
	DICE	IOU	DICE	IOU
FCN	87.89	78.65	83.37	71.49
U-net	57.62	47.45	59.80	49.91
PSPNet	60.01	49.21	55.47	45.88
FPN	93.75	88.24	88.45	79.29
TransUnet	–	–	89.96	81.75
R2GCN	**96.93**	**94.05**	**91.99**	**85.17**

Bold values indicate the best results

IOU represents the ratio of the intersection and union of the model to the predicted result and the true value, which is defined as follows:

$$IOU = \frac{TP}{FP + TP + FN} \tag{8}$$

where TP: True Positive, is judged as a positive sample, which is actually a positive sample, FP: False Positive, is judged as a positive sample, but is actually a negative sample, FN: False Negative, is judged as a negative sample, but in fact is a positive sample. The larger the $DICE$ and IOU index values, the better the segmentation effect of the model.

4.3 Results

In the experiments, all the methods involved use VGG-16 and ResNet-50 to extract features. We will compare the segmentation results in this paper with other segmentation methods, including FCN [7], U-Net [9], PSPNet [11], FPN [22], TransUNet [23]. Table 1 shows the quantitative comparison of $DICE$ and IOU between the proposed R2GCN and other different methods. From this tables, it can be seen that R2GCN achieves the best results in terms of DICE and IOU. For a more intuitive comparison, we provide visualization results, as shown in Fig. 3. Figure 3 shows the predicted segmentation plots for R2GCN and comparison methods. From top to bottom are patient images, ground truth, U-Net segmentation results, PSPNet segmentation results, FCN segmentation results, FPN segmentation results and R2GCN segmentation results. As can be seen from the fifth prediction map in the third row, U-Net has a problem of false detection. As can be seen from the first three prediction maps in the forth row, FCN is not stable enough and is susceptible to interference from other substances in the lungs. It can be seen from the whole segmentation results that there is a certain gap between the edges of most network segmentation results and the ground truth edges. The edges of the segmentation results of our model are closer to the edges of the ground truth, and our method has a better segmentation performance. A possible explanation is that the method we proposed in this paper combines the features between similar nodes of GCN and the contextual features of the BRN, and the use of GCN makes nodes with similar information not easily ignored and ensures the accuracy of segmentation. Another point is that the contextual features makes the segmentation boundary more consistent with the actual boundary.

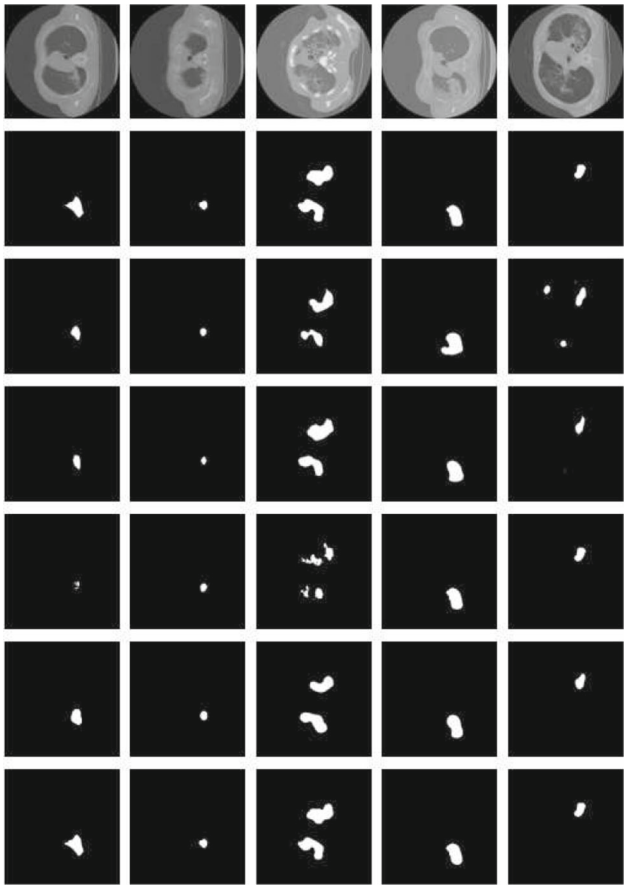

Fig. 3 Visualizations of the dataset. Top-down: original images, ground truth, U-Net, PSPNet, FCN, FPN and R2GCN

Table 2 Results of the ablation experiments

Methods	VGG-16		ResNet-50	
	DICE	IOU	DICE	IOU
BRN (without GCN)	80.21	67.19	89.38	80.80
GCN (without BRN)	59.56	47.24	78.31	64.35
R2GCN	**96.93**	**94.05**	**91.99**	**85.17**

Bold values indicate the best results

Table 3 Segmentation results of different upsampling layers

Method	DICE	IOU
One upsampling	80.42	67.25
Two upsampling	**96.93**	**94.05**

Bold values indicate the best results

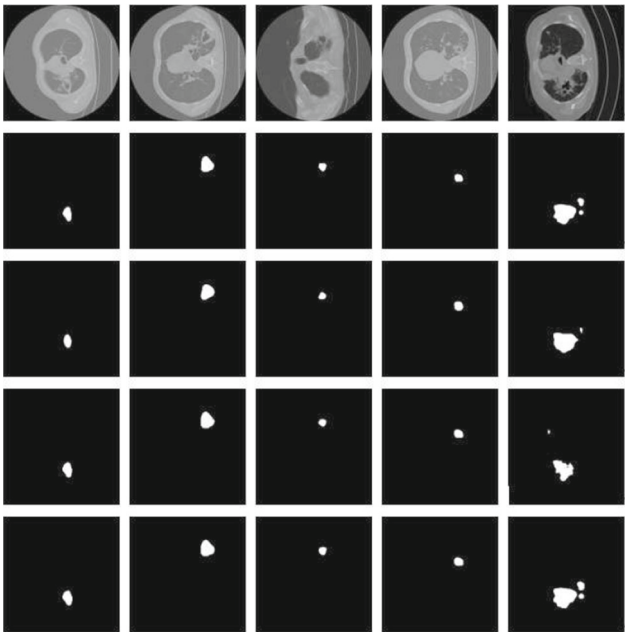

Fig. 4 Visualization of ablation experiments. Top-down: original images, ground truth, GCN, BRN and R2GCN

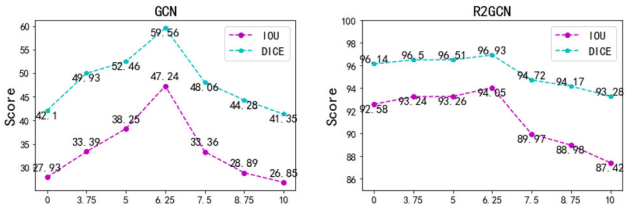

Fig. 5 Comparison under different similarity λ

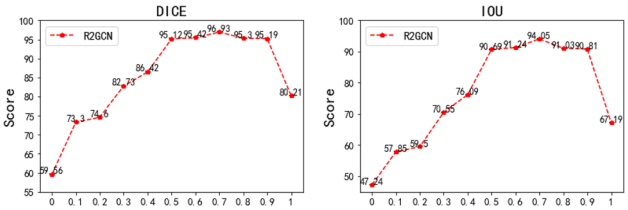

Fig. 6 Comparison under different degrees of fusion μ

Table 2 shows the results of ablation experiments. We can find the BRN module and the GCN module play an active role. Compared with two variants, the fused model performs better. Figure 4 shows the comparative segmentation results. From top to bottom are patient images, ground truth, GCN segmentation results, the BRN segmentation results and R2GCN segmentation results. As can be seen from the figure, there is a certain gap between the edge of the GCN segmentation results and the edge of the ground truth. Most of the edges of the BRN segmentation results are similar to the edge of the ground truth, but there is a problem of missed detection. The fusion of the two models compensates for the disadvantage of a single module. From Table 3, we can

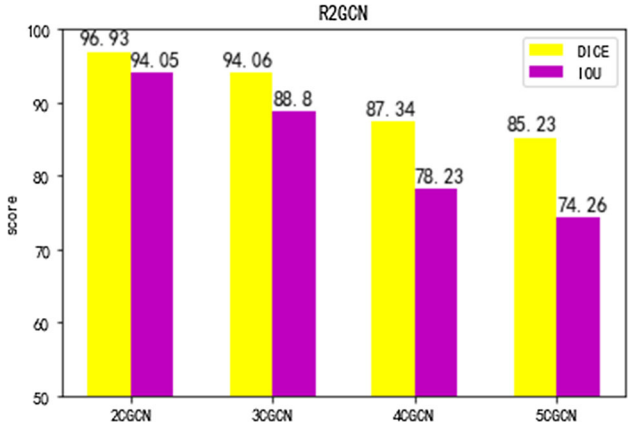

Fig. 7 Comparison of different layers of GCN

see that the proposed model can achieve 96.93 and 94.05 in terms of DICE and IOU by using two up-sampling blocks, which are significantly better than one up-sampling manner. Therefore, we adopt two up-sampling layers in our model.

Figure 5 shows the segmentation results of GCN and R2GCN at different degrees of similarity. As can be seen from the figure, the performance raises with larger λ. When λ achieves 6.25, promising performance can be expected. But when λ exceeds 6.25, performance starts to degrade. Figure 6 shows the segmentation results of R2GCN under different fusion coefficients. As can be seen from the figure, when μ=0.7, R2GCN obtains the best performance. When the model fusion coefficient is greater or less than 0.7, the effect of the model will gradually deteriorate. Figure 7 shows the segmentation results of R2GCN with different GCN layers. It can be seen that R2GCN performs best by using two layers of GCN. When the number of GCN layers increases, the model will gradually decrease.

5 Conclusions

In this paper, an RNN-combined graph convolutional network is proposed for the segmentation of tuberculosis cavities. The segmentation task of tuberculosis cavities is achieved by using patient lesion images, obtaining information about similar nodes using GCN and aggregating the contextual information of the feature map using BRN. It achieved better performance in terms of DICE and IOU metrics on the dataset compared to other segmentation networks.

Nevertheless, the model's segmentation of tuberculosis cavities is also interfered by some other structures or diseases in the lung. This leads to some false detections. There is room for further improvement in the segmentation of tuberculosis cavities, such as prioritising the segmentation of lung structures before the segmentation of tuberculosis cavities, which would greatly reduce the interference of lung struc-

tures and other diseases in the segmentation of tuberculosis cavities.

Author Contributions ZX, XZ and YL wrote the main manuscript text. All authors reviewed the manuscript.

Funding Funding is not applicable.

Data availability Data are a non-public data provided by a hospital.

Declarations

Conflict of interest I declare that the authors have no conflict of interest as defined by Springer, or other interests that might be perceived to influence the results and discussion reported in this paper.

Ethical approval Ethical approval is not applicable.

References

1. Hui, G.: To explore the clinical value of ct in the diagnosis and clinicopathological classification of peripheral small lung cancer. China Continuing Med. Educ. **8**(21), 48 (2016)
2. Lange, C.: Advances in the diagnosis of tuberculosis. Respirology **15**(2), 220–240 (2010)
3. Sahiner, B.: Deep learning in medical imaging and radiation therapy. Med. Phys. **46**(1), 1–36 (2019)
4. Yanbei Liu, L.F.: Incomplete multi-modal representation learning for Alzheimer's disease diagnosis. Med. Image Anal. **69**, 101953 (2021)
5. Zheng, S.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1529–1537 (2015)
6. Visin, F.: Reseg: a recurrent neural network-based model for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 41–48 (2016)
7. Long, J.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
8. Yu, Q.: Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8280–8289 (2018)
9. Ronneberger, O.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015)
10. Dalmış, M.U.: Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med. Phys. **44**(2), 533–546 (2017)
11. Zhao, H.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
12. Gao, H.: Graph u-nets. In: International Conference on Machine Learning, pp. 2083–2092 (2019)
13. Li, T.: Image co-saliency detection and instance co-segmentation using attention graph clustering based graph convolutional network. IEEE Trans. Multimedia **24**, 492–505 (2021)
14. Zhang, Y.: Affinity fusion graph-based framework for natural image segmentation. IEEE Trans. Multimedia **24**, 440–450 (2021)
15. Lu, Y.: Graph-fcn for image semantic segmentation. In: International Symposium on Neural Networks, pp. 97–105 (2019)
16. Li, T.: Image co-saliency detection and instance co-segmentation using attention graph clustering based graph convolutional network. IEEE Trans. Multimedia **24**, 492–505 (2021)
17. Van Rikxoort, E.M.: Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. Phys. Med. Biol. **58**(17), 187 (2013)
18. Wang, X.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803 (2018)
19. Wang, S.: Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med. Image Anal. **40**, 172–183 (2017)
20. Fan, D.-P.: Inf-net: automatic Covid-19 lung infection segmentation from ct images. IEEE Trans. Med. Imaging **39**(8), 2626–2637 (2020)
21. Gatta, C.: Unrolling loopy top-down semantic feedback in convolutional deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 498–505 (2014)
22. Lin, T.-Y.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
23. Chen, J., et al.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)