LOBENET: A GLOBAL POSITION RESERVATION AND FISSURE-AWARE CONVOLUTIONAL NEURAL NETWORK FOR PULMONARY LOBE SEGMENTATION

Mingjian Chen[1,2], *Yun Gu*[1,2], *Yulei Qin*[1,2], *Hao Zheng*[1,2], *Jie Yang*[1,2]

[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China
[2]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

Pulmonary lobe segmentation is vital for clinical analysis and preoperative planning, which is essential for Computer-Aided Diagnosis systems (CADs). The fissures, separating the lung parenchyma into five pulmonary lobes, are hard to detect by the classical method with combining anatomical information. The deep learning method based on convolutional neural networks (CNNs) has made great success in pulmonary lobe segmentation. As we know, the position of the pulmonary voxel is helpful for pulmonary lobe segmentation. However, 3D CNN can not take the whole scan, only smaller cubes cropped from the scan as input due to the limitation of GPU memory, which will make the position information lost. In this paper, we proposed a 3D residual CNN called LobeNet for pulmonary lobe segmentation with global position reservation and fissure-aware property. Our LobeNet makes up for the loss of position information by incorporating the 3D global position as high-level features making the network converge to a better local minimum. Moreover, we use a fissure-aware training strategy to lead the network to concentrate on the fissures, which forces the predicted border of pulmonary lobes more accurate. Meanwhile, the focal loss is leveraged for hard example mining. Our proposed network is validated on a public dataset and achieved the state-of-the-art performance with a 93.35% dice coefficient. The ablation study illustrates the global position and fissure-aware training strategy can produce a more accurate segmentation result.

Index Terms— Pulmonary Lobe Segmentation, Convolutional Neural Network, Computed Tomography

1. INTRODUCTION

Pulmonary cancer is one of the major diseases leading to death in the world. Computed Tomography (CT) has become the principal method to diagnose pulmonary diseases. The pulmonary parenchyma can be divided into five lobes (left upper lobe, left lower lobe, right upper lobe, right middle lobe, right lower lobe) by the pulmonary fissures. Lobe segmentation in CT scans is an important task in Computer-Aided Diagnosis systems (CADs) since it can provide a quantitative analysis of lobe [1]. Moreover, lobe segmentation can help to reduce unnecessary lung parenchyma excision in pulmonary nodule resection, which will greatly improve the life quality of patients after surgery. However, it's a time-consuming task for manually annotating different pulmonary lobes in a chest CT scan. The lobe segmentation is a challenging task since the border of the lobe varies in thickness, location and shape [2]. Also, the pulmonary fissure is not always completely visual in a chest CT.

Traditional methods for lobe segmentation mainly base on pulmonary fissures detection since pulmonary fissures are the boundary of lobes [3]. However, the pulmonary fissures vary very much in different patients and can be incomplete, which may cause pulmonary fissure detection failure. In addition, extra anatomical information such as the analysis of airway and vessels is essential for traditional lobe segmentation methods. A pulmonary fissure filter is proposed to extract the fissures and pulmonary lobes based on the position of lung parenchyma voxel relative to the fissures [4]. [1] develops a probabilistic segmentation method to segment pulmonary fissures with a non-parametric surface fitting algorithm, achieving a relatively high segmentation performance. However, the applied probabilistic model is complex and time-consuming.

With the development of deep learning, especially the success of convolutional neural networks (CNNs), many deep learning based methods are developed for pulmonary nodule detection and segmentation [5, 6]. For lobe segmentation, [7] proposes a modified V-Net with some regularization techniques including batch normalization, dropout, deep supervision and multi-task learning. However, the method needs to resize the CT scan into a much smaller size, which may cause the pulmonary fissures invisible. [8] introduces a convolutional neural network that takes the whole CT scan as input. Unfortunately, the network contains only one downsampling, which lacks in enough receptive filed. Therefore, the network fails in utilizing context information in the CT scan.

Since the position of pulmonary voxel contributes to lobe segmentation a lot, the aforementioned two CNN-based

Corresponding authors: Jie Yang (jieyang@sjtu.edu.cn) and Yun Gu (yungu@ieee.org). This research is partly supported by NSFC, China (No: 61977046, 61806125, 61802247), Committee of Science and Technology, Shanghai, China (No. 19510711200). Thank the Student Innovation Center of Shanghai Jiao Tong University for providing GPUs.

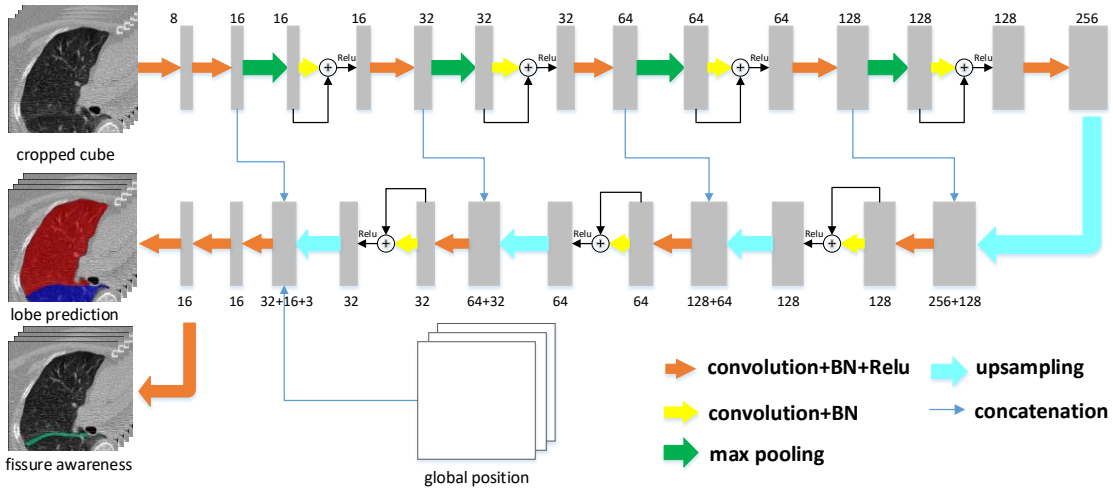

Fig. 1. The architecture of our proposed LobeNet. The digit above indicates the number of the feature maps.

methods try to make the network able to accommodate the whole CT scan. However, there are always some trade-offs between GPU memory and feature representation capacity, which causes these two methods low accuracy. In this paper, we propose a 3D residual CNN called LobeNet trained by end-to-end manner for lobe segmentation with strong feature representation ability in limited GPU memory. Unlike the previous method [8], our LobeNet is trained by small cubes cropped from the CT scans by sliding window method. Small cubes can provide local anatomical features, but the global position information is weakened, which can cause the pulmonary parenchyma difficult to segment. To resolve this problem, we calculate every voxel's position in the cube relative to the whole lung cube and feed this position information into the network. Since the fissures play an important role in separating the pulmonary parenchyma, we proposed a fissure-aware training strategy to lead the network to focus on the lobe border.

Our main contributions can be summarized as follows: (1) Global position reservation is introduced in the proposed LobeNet without any trade-off between GPU memory and feature representation capacity. (2) The fissure-aware training strategy is proposed to enhance the ability of feature representation of our LobeNet. (3) Experiments in public dataset [8] show that our proposed network achieves the state-of-the-art segmentation performance.

2. METHOD

In this section, we first introduce the preprocessing step, which can remove the irrelevant regions in the CT scans. Then the detail of the proposed LobeNet is presented.

2.1. Preprocessing

The CT scans contain tissues and scanning bed that are irrelevant to the pulmonary lobe. The first step of preprocessing is to extract the lung mask, which can eliminate irrelevant noise and reduce computational complexity of the LobeNet. The lung mask can be extracted by binarization and connected component analysis. The largest two connected components are extracted as the lung mask. Then the lung mask is refined by closed operation. The extracted lung region is cropped into smaller cubes using the sliding window method, where the window size is $32 \times 224 \times 224$ and the sliding stride is $[8, 56, 56]$. Through the sliding window method, the required GPU memory can be dramatically reduced and the network architecture can be designed more flexibly. To train our proposed LobeNet, the CT value is clipped into $[-1000, 600]$ Hounsfield Unit and then normalized into $[0, 1]$.

2.2. LobeNet

In this subsection, the detail of our LobeNet is introduced including network architecture shown in Fig. 1, training loss and inference process.

2.2.1. Architecture

Our proposed LobeNet is based on 3D U-Net [9], which can capture more spatial feature than 2D U-net [10] in the CT scan. Given an input cropped cube, LobeNet encodes the input to the high-level semantic feature by the contracting path, which consists of four residual convolutional layers [11] and four down-samplings. The residual convolutional layer encourages the reuse of features and makes the network more trainable in a small dataset. After the contracting path, the semantic features are expanded to the same size of the input

cube in the expanding path, where the corresponding features in the contracting path are concatenated to fuse multi-level features. Then each voxel is classified into six classes (five lobes plus background). In detail, each convolutional block consists of a convolutional layer, batch normalization (BN) and rectified linear unit (ReLu). The $3 \times 3 \times 3$ convolution kernel is applied in the whole network. Following the common setting, we employ $2 \times 2 \times 2$ max pooling in the down-sampling, while for up-sampling, we repeat every voxel twice followed by a convolutional block.

It is easy to classify the left lung and the right lung via the global position in the CT scan. But within the cropped cube, the loss of the global position makes it hard to distinguish right and left lung and makes the lobe segmentation impossible. To make the network capable to capture global information, we encode the coordinate of each voxel into a three-channel position cube. This position cube is fed into the last expanding path by concatenation, which endows the LobeNet with global context information. In this way, the performance of our proposed LobeNet is greatly improved and the network converges fast. We can not feed the position cube as the input of the network since the position cube has already included underlying features for lobe segmentation. The convolutional neural network can not learn proper feature representation for both lung cube and position cube simultaneously.

From anatomy, the fissure is the most important tissue for radiologists to segment pulmonary lobes. The lobe segmentation is very difficult in the border, especially in the border of two lobes. We introduce another fissure-aware path from the main network. The fissure prediction path shares the backbone network with the lobe prediction, which leads to a better generalization for lobe segmentation. Here we only predict the inter-lobe border excluding the border between the lobe and background, since these two kinds of borders are definitely inconsistent. There is high contrast in the border between the lobe and background, while it is hard to recognize the inter-lobe border.

2.2.2. *Training Loss*

Dice coefficient loss [12] is a good choice for training a two-class segmentation network, which can solve the class-imbalance problem in the segmentation task. We extend the dice loss to multi-class loss, which can be described as the average dice coefficient loss across six classes. Let x represents the voxel from input cube X, $p_i(x)$ means the prediction probability belonging to the class i of the voxel x, while $g_i(x)$ is the ground truth. The multi-class dice loss can be represented as follows:

$$L_{dm} = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{2\sum_{x\in X} p_i(x)g_i(x)}{\sum_{x\in X} p_i(x) + \sum_{x\in X} g_i(x)},$$

where N is the number of classes.

For fissure-aware strategy, two-class dice coefficient loss is employed to train the fissure-aware path, which describes as follows:

$$L_{db} = 1 - \frac{2\sum_{x\in X} p(x)g(x)}{\sum_{x\in X} p(x) + \sum_{x\in X} g(x)},$$

where $p(x)$ is the prediction probability belonging to the fissures. Therefore, the total training loss can be written as $L_d = w_1 L_{dm} + w_2 L_{db}$. w_1 and w_2 are the balance weights between L_{dm} and L_{db}.

After training with dice coefficient loss, the network has learned almost proper feature representation. Then we apply focal loss [13] to fine-tune the network, after which the network can generalize better in hard examples. The low prediction confidence can produce high weight in the focal loss. This rule can be represented as:

$$L_f = -\frac{1}{N}\sum_{i=1}^{N}\sum_{x\in X}(1 - p_i(x))^{\gamma} g_i(x)\log(p_i(x)),$$

where γ is the hyper-parameter. The focal loss is both used for lobe prediction and fissure-aware strategy.

2.2.3. *Inference*

In the inference phase, we should get the segmentation results of a whole CT scan. The trained LobeNet predicts the probability of each voxel in all the cropped cubes that are generated by sliding window with stride $[16, 112, 112]$. Since there are overlapped regions between cropped cubes, the final prediction probability is the average of all cropped cubes. Finally, the segmentation results are set to the class with the max probability.

3. EXPERIMENT

We use the public dataset from [8] to validate our proposed LobeNet, which consists of 40 training scans and 10 testing scans. The dataset is a subset of LUNA16 [14] and the labels are manually annotated by the radiologist. The axial size of all CT scans is 512×512 and the dimension of z axis ranges from 113 to 561 due to different scanners.

Our experiments are conducted on the keras with 2 NVIDIA Titan X GPUs. The Adam [15] is used to optimize our LobeNet. We use dice loss L_d to optimize our LobeNet with 15 epochs in a learning rate of 1×10^{-3}, then focal loss fine-tunes the LobeNet with 1 epoch in a learning rate of 1×10^{-5}. The batch size is 6. Data augmentation can increase the diversity of the training data and the generalization ability of the network, especially in a small dataset. Our argumentation methods include XY-Plane flipping and gaussian filter with $\sigma = 1$. The hyper-parameter γ is set to 2. The balance weight w_1 and w_2 are 0.7 and 0.3, respectively.

Table 1. The comparison results on the testset. The dice coefficient results for right upper lobe (RU), right middle lobe (RM), right lower lobe (RL), left upper lobe (LU), left lower lobe (LL) and their average result are presented.

	RU	RM	RL	LU	LL	AVG
[7]	92.76	**84.68**	94.33	88.10	94.78	90.93
[8]	92.53	80.60	93.05	96.10	95.30	91.48
Ours	**94.05**	82.62	**96.39**	**97.06**	**96.65**	**93.35**

Table 2. The results of the ablation study about our proposed method and its variations. GP and FA mean global position reservation and fissure-aware strategy, respectively.

Method	DICE	TPR	PPV
LobeNet w/o GP&FA	87.67	88.08	88.36
LobeNet w/o FA	92.44	92.83	92.59
LobeNet	**93.35**	**93.68**	**93.37**

3.1. Comparisons with Other Approaches

In this experiment, we compare our proposed method with two previous state-of-the-art methods, which are the method of Ferreira et al. [7] and Tang et al. [8]. Following the metric in [7, 8], we apply the average dice coefficient to evaluate the segmentation performance in the experiments. The segmentation results are reported in Table 1. It can be concluded that our proposed LobeNet achieves highest dice coefficient of 93.35% with an improvement of 1.87% compared to other methods. For per-lobe analysis, our method achieves the best performance in RL, RU, LU and LL, which validates that our proposed method with global position reservation and fissure-aware strategy can model the pulmonary lobe and anatomical information better and has a great generalization ability.

3.2. Ablation Study

To further validate the effect of global position reservation and fissure-aware strategy, we compare our proposed method with its variations. The variations include (1) the LobeNet without these two components and (2) the LobeNet without fissure-aware strategy. In addition, we calculate true positive rate (TPR) and positive predictive value (PPV) to further analyse our proposed method.

According to Table 2, our proposed method performs much better than other variations in DICE, PPV and TPR, which illustrates that global position reservation and fissure-aware strategy are helpful for the network to capture the essential features for the pulmonary lobe segmentation. The great gain of performance from global position reservation means that the position information is vital for classifying different lobes. With the position information, the network can converge faster and reach a lower local minimum. Since the position of the voxel is directly related to the class of pulmonary lobes, the network can spare more attention to excavate the anatomical information in the CT scans that has been proved to be an important prior to the lobe segmentation. The further improvement by employing the fissure-aware strategy proves that this strategy leads the features representation learned by the network more accurate.

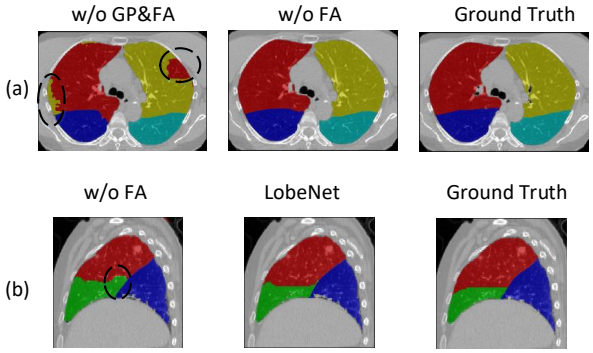

Fig. 2. The segmentation results of our proposed LobeNet and its variations. Row (a) is shown in the axial view, while row (b) is shown in the sagittal view.

For visual comparison, some segmentation results predicted by the proposed LobeNet and its variations are illustrated in Fig. 2. As shown in Fig. 2(a), the method without global position reservation will obtain totally wrong segmentation for right upper lobe and left upper lobe due to the loss of the position information. With the fissure-aware strategy, our proposed method achieves more accurate segmentation boundaries, especially between different lobes, as shown in Fig. 2(b).

4. CONCLUSION

In this paper, we proposed an automatic pulmonary lobe segmentation network named LobeNet. The position information contributes much to lobe segmentation. However, the loss of the global position in cropped cubes is not well solved in previous CNN-based methods. Our LobeNet possesses the property of global position reservation and the fissure awareness, thus achieving state-of-the-art performance compared to other methods. Experimental results further validate the effectiveness of our LobeNet.

5. REFERENCES

[1] Felix JS Bragman, Jamie R McClelland, Joseph Jacob, John R Hurst, and David J Hawkes, "Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior," *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1650–1663, 2017.

[2] Kevin George, Adam P. Harrison, Dakai Jin, Ziyue Xu, and Daniel J. Mollura, "Pathological pulmonary lobe segmentation from ct images using progressive holistically nested neural networks and random walker," in *International Workshop on Deep Learning in Medical Image Analysis International Workshop on Multimodal Learning for Clinical Decision Support*, 2017.

[3] Bianca Lassen, Jan-Martin Kuhnigk, Ola Friman, Stefan Krass, and Heinz-Otto Peitgen, "Automatic segmentation of lung lobes in ct images based on fissures, vessels, and bronchi," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 560–563.

[4] Van Rikxoort, Eva M, De Hoop Bartjan, Van De Vorst Saskia, Prokop Mathias, and Van Ginneken Bram, "Automatic segmentation of pulmonary segments from volumetric chest ct scans," *IEEE transactions on medical imaging*, vol. 28, no. 4, pp. 621–630, 2009.

[5] Yulei Qin, Hao Zheng, Yue-Min Zhu, and Jie Yang, "Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3d cnns," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1005–1009.

[6] Yulei Qin, Hao Zheng, Xiaolin Huang, Jie Yang, and Yue-Min Zhu, "Pulmonary nodule segmentation with ct sample synthesis using adversarial networks," *Medical physics*, vol. 46, no. 3, pp. 1218–1229, 2019.

[7] Filipe T Ferreira, Patrick Sousa, Adrian Galdran, Marta R Sousa, and Aurélio Campilho, "End-to-end supervised lung lobe segmentation," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[8] Hao Tang, Chupeng Zhang, and Xiaohui Xie, "Automatic pulmonary lobe segmentation using deep learning," *arXiv preprint arXiv:1903.09879*, 2019.

[9] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[13] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[14] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas van den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Medical image analysis*, vol. 42, pp. 1–13, 2017.

[15] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.