An Original Neural Network for Pulmonary Tuberculosis Diagnosis in Radiographs

Junyu Liu, Yang Liu, Cheng Wang, Anwei Li, Bowen Meng(✉), Xiangfei Chai, and Panli Zuo

Huiying Medical Technology (Beijing) Co., Ltd., Beijing, China
liujunyu@huiyihuiying.com, liuyang@huiyihuiying.com, wangcheng@huiyihuiying.com, mengbowen@huiyihuiying.com

Abstract. Tuberculosis (TB) is a widespread and highly contagious disease that may lead serious harm to patient health. With the development of neural network, there is increasingly attention to apply deep learning on TB diagnosis. Former works validated the feasibility of neural networks in this task, but still suffer low accuracy problem due to lack of samples and complexity of radiograph information. In this work, we proposed an end-to-end neural network system for TB diagnosis, combining preprocessing, lung segmentation, feature extraction and classification. We achieved accuracy of 0.961 in our labeled dataset, 0.923 and 0.890 on Shenzhen and Montgomery Public Dataset respectively, demonstrating our work outperformed the state-of-the-art methods in this area.

Keywords: Tuberculosis · Classification · DNN

1 Introduction

Tuberculosis is a highly contagious disease that may lead serious harm to patient health. According to the World Health Organization (WHO) [1], until the end of 2015, nearly 10 million people in the world suffered from tuberculosis and more than 1.5 million died. The WHO pointed out that early diagnosis and appropriate treatment can avoid the majority of tuberculosis deaths, and millions of people are saved each year. Nonetheless, huge number of people still suffers for high cost and lack of professional doctors. Therefore, reliable tuberculosis diagnosing system is an urgent demand.

At present, a large number of medical image data has not yet been digitized, and the level of data sharing and interoperability among hospitals is still at a low level. It is a dilemma that advanced method usually requires big data, which is impossible for medical dataset. Also, it is difficult to obtain reliable labeling data in the medical imaging field for the interdisciplinary gap. In addition, medical images contain more difficult samples and pixel-scale features, making AI image analysis in the medical field more challenging than natural image recognition. This work proposes a neural network specialized for pulmonary tuberculosis diagnosis in radiographs, to solve all above difficulties.

V. Kůrková et al. (Eds.): ICANN 2018, LNCS 11140, pp. 158–166, 2018.
https://doi.org/10.1007/978-3-030-01421-6_16

2 Related Works

In 2012, Hinton's team [2] first adopted convolutional neural network into the ImageNet classification challenge and achieved astonishing results, drastically reducing the Top5 error rate from 26% to 15%. This opened up a boom in deep learning. At present, deep learning has achieved remarkable results in the fields like image recognition, detection, segmentation, and so on [3, 5].

Deep learning technology was first officially applied to medical image analysis in 2015. Convolutional neural networks (CNN) soon gained increasingly popularity due to their ability to learn mid and high-level image representations. Bar Y et al. explore the ability of a CNN to identify different types of pathologies in chest x-ray images [6]. They used a pre-trained CNN on the ImageNet dataset as the first descriptor, and the second descriptor is PiCoDes, which is a compact high-level representation of popular low-level features (SIFTs [6], GIST, PHOG, and SSIM) which is optimized over a subset of the ImageNet dataset containing approximately 70,000 images. They found that the best performance was achieved using a combination of features extracted from the CNN and a set of low-level features. Of course, the capacity of system will be limited for lack of training.

U.K. Lopes et al. used a pre-trained CNN as a feature extractor, combining with traditional machine learning methods for tuberculosis detection [8]. They first used detached networks to extract features, then integrated CNN features and finally created an ensemble classifier by combining the SVMs trained using the features extracted from GoogLenet [9], ResNet [10], and VggNet [11]. The author of [12] proposed a novel method to detect pulmonary tuberculosis. The method is divided into two steps. The first step is to use pre-trained networks to make a two classification on chest X-rays. For classification, the chest X-rays are resized to respectively corresponding network, and the results of the prediction of all classification networks are averaged as the final classification result. The second step is that the sensitivity of softmax score to occlusion of a certain region in the chest X-Ray is used to find which region in the image is responsible for the classification decision. But the over-resize process will sharply reduce the accuracy of system.

Olaf Ronneberger et al. proposed a network called U-Net [13] for small-sample segmentation. The network consists of two parts, a contracted path is used to obtain contextual information and a symmetrical expansion path for precise positioning. At the same time, in order to make more efficient use of the annotation data, they also use a variety of data enhancement methods. In 2016, Milletari et al. proposed an extension to the U-Net layout that incorporates ResNet-like residual blocks and a Dice loss layer, rather than the conventional cross-entropy [14].

Inspired by all the mentioned works, we propose a combination of segmentation and classification deep neural network through the chest X-rays to detect tuberculosis. All chest X-rays were preprocessed to emphasize lung features. Main body of the network has two branches: one is a designed lung segmentation network to obtain chest masks, and the other a classification network. We achieve accuracy of 0.965 in our dataset, 0.923 and 0.890 on Shenzhen and Montgomery Public Dataset respectively, proving us the state-of-the-art in this area.

3 Proposed Methods

3.1 Method Overview

We proposed an end-to-end network for tuberculosis judgement. The whole system consists of a Lung Segmentation Network, a classification backbone and an output head. Heat maps are generated for further analysis and algorithm verification. This is the first work to combine all the steps of tuberculosis detection in a whole network, making a compromise between computational speed and preservation of image information. The whole system is demonstrated in Fig. 1.

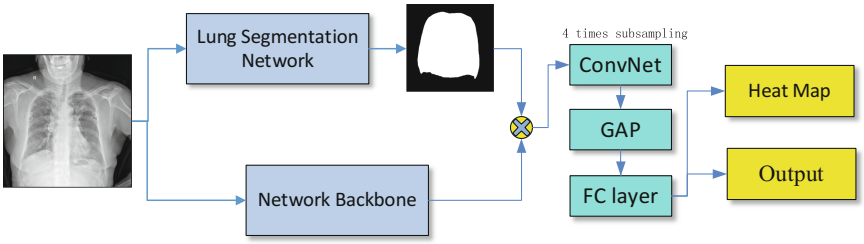

Fig. 1. The block diagram of the proposed network.

3.2 Lung Segmentation Network

According to [14], lung segmentation is necessary for automatic tuberculosis diagnosing. In this paper, we designed a simple and effective CNN with atrous convolutional layers [18] to segment the chest from X-rays referring to U-net. Basic feature extraction part has 3 conv-pooling blocks with different number of channels. Each conv-pooling block contains a pooling layer after a few convolutional blocks, while each convolutional block consists of a convolutional layer followed by a Batch-Norm layer and a ReLU activation layer. Totally 8 times subsampling was implemented and the network structure is shown in Fig. 2.

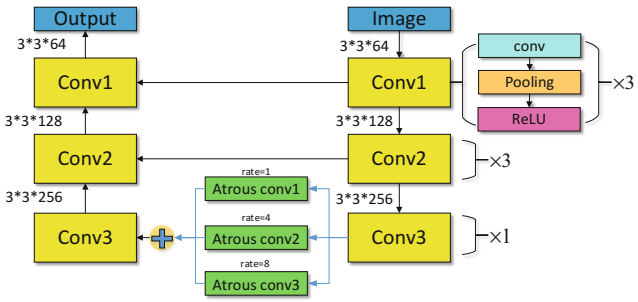

Fig. 2. ConvNet configuration for feature extraction.

Lungs in radiographs are of different sizes due to individual difference and other factors. Therefore, multi-scale segmentation was also taken into consideration. We used 3 atrous convolutional layers with different sample rates respectively. All the feature maps obtain by dilated convolution are added together and connected with the decoder of the network. Segmented results are generated by continuous up-sampling. In order to overcome the problem of low resolution after down-sampling in the FCN [17] method, we fused the feature map of each down-sampled feature with that of the corresponding up-sampling part. Chest segmentation results are shown in Fig. 3.

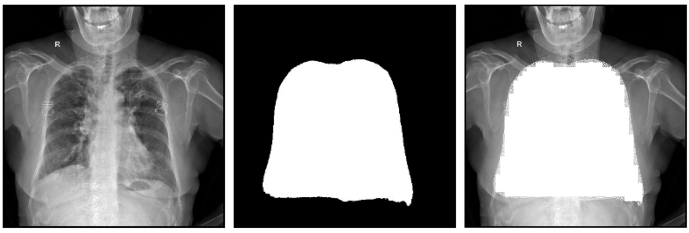

Fig. 3. Chest segmentation results. Left: original picture; Middle: segmentation result; Right: evaluation result.

3.3 Specialized Innovations

Preprocessing. Radiographs need preprocessing before checking. The grayscale of chest X-ray pixels usually range from tens to thousands, and it's impossible for human eyes to distinguish this huge change. Also, too large scales tend to cause the diagnosing network to divergent. Therefore, the original pixel values need adjustment according to WW (window width) and WP (window position). Because not all graphs are given guidance values of WW and WP, a standard set of WW and WP was generated from samples accompanied with WW and WP guidance values using cluster algorithm. We also found that histogram equalization operation can emphasize the features in lung while not significantly changing the gray level in other organs and background. Original radiographs often have as many as two thousand pixels in length, which is a huge burden for computation. But considering that some granule infections can be really small, input images are bilinear interpolated to 1024×1024.

Two Branches. The main body of proposed network has two branches, one for lung segmentation and the other for feature extraction with the network backbone. We choose 6 different popular and practical backbones in total for this work. To limit computation memory and time, we subsampled the feature map by 32 instead of the original picture masked by the output of segmentation branch, allowing main body of two branches to work simultaneously.

Network Head. There are two heads in the last part of network. The classification head of the network is specialized for this task. As input of our system is much larger than normal classification competitions, we need more times of subsampling than the original

networks. In practical, we adopted 128 times down sampling in our network. High similarity is a dangerous character of radiographs in this task, tending to cause over-fit. Therefore, we added a heat map head to analysis if the correct feature of graphs has been learned. For heat map generation, the second to last fully connected (FC) layer is replaced by a global average pooling (GAP) [18] layer, also reducing parameters in the network. Considering the imbalance of positive and negative samples, and also false negative (FN) is much more harmful in medical area, focal loss [4] is introduced into this work, giving positive samples a higher loss during training.

4 Experiments

4.1 Database

Database used in this paper comes from 2 sources. The first dataset was provided by *Huiying Medical Technology (Beijing) Co., Ltd*., containing 2443 frontal chest X-ray images (DICOM format), with labels marked by a reliable expert network. In the dataset, 2000 were randomly chosen as training set and the rest divided into validation and test ones. There are two public datasets [20] available on the Internet. Shenzhen Hospital dataset, which includes 662 frontal chest x-rays, was acquired from Shenzhen No. 3 People's Hospital in Shenzhen, China. Montgomery County chest X-ray set (MC) was collected in collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA, consisting of 138 frontal X-rays.

4.2 Experimental Results

To test the performance of network with different backbones, parallel comparisons were made on our test dataset. Accuracy, sensitivity, specificity, AP, and AUC results are shown in Table 1. Inception-v4 backbone without mask branch was also tested.

Table 1. Parallel comparisons of each method for our dataset

Backbone	AUC	Accuracy	AP	Sensitivity	Specificity
VGG-19	0.974	0.893	0.981	**0.988**	0.765
ResNet-50	0.983	0.875	0.992	0.979	0.892
ResNet-101	0.989	0.879	0.992	0.972	0.932
ResNet-152	0.991	0.923	**0.994**	0.960	0.945
Inception v4	**0.995**	**0.961**	**0.994**	0.966	**0.955**
ResNet-Inception v2	0.982	0.934	0.984	0.948	0.915
Inception v4 (no mask)	0.953	0.908	0.947	0.821	0.954

To be intuitive, the P-R curves and ROCs are shown in Fig. 4.

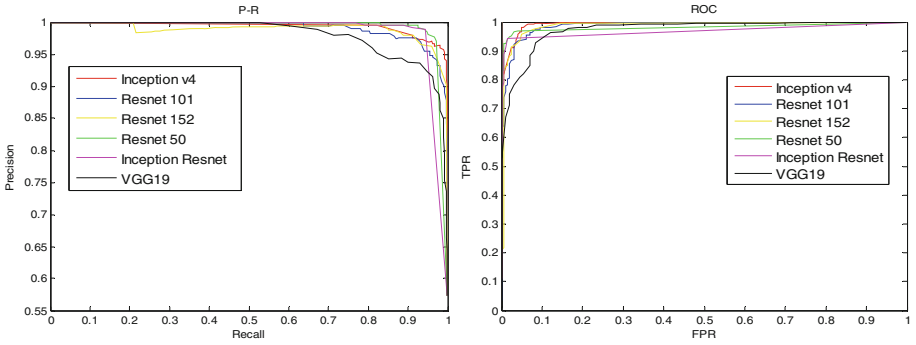

Fig. 4. P-R curves (left) and ROCs (right).

The results show that our method made a highest accuracy of over 96.1% on our test dataset, achieving by Inception v4. Mask branch contributed about 5.3% in accuracy. We also reselected training set and retrained our networks from the beginning to exclude the possibility of coincidence. We also checked the heat maps generated by our network, finding it reasonable although slight bias and blur happens due to 128 times subsampling. The visualized results are shown in Fig. 5.

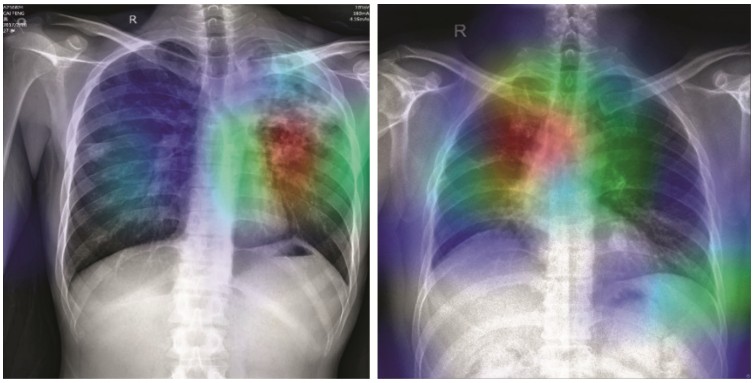

Fig. 5. The heat map acquired in our network. Although slight positioning bias happens due to totally 128 times subsampling, the red area roughly reflects position of infection. (Color figure online)

Longitudinal comparisons with former works [8, 12, 15, 16] were also accomplished. To be fair and objective, we compared the results of proposed method and the other works on two public datasets. All the data of former works cited in this paper are the best results the authors claimed. The models we used were still the ones we trained on our dataset. Figure 6 shows the visualized results of our networks on Shenzhen Dataset. Comparison with former works are shown in Table 2.

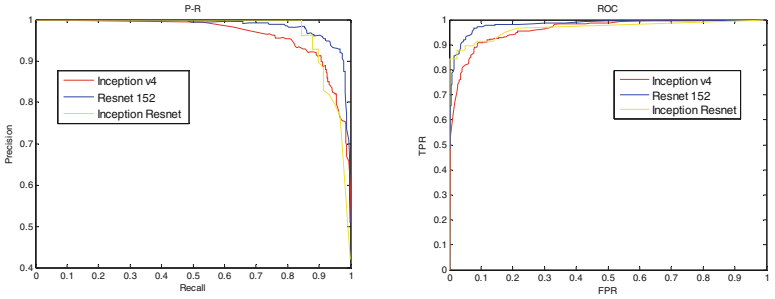

Fig. 6. P-R curves (left) and ROCs (right) of our networks on Shenzhen Dataset.

Table 2. Performance for Shenzhen Dataset. Last three are proposed methods.

Method	AUC	Accuracy	AP	Sensitivity	Specificity
U.K. Lopes et al.	0.894	0.837	-	-	-
Mohammad et al.	0.940	0.900	-	0.960	0.960
Sangheum et al.	0.926	0.837	0.940	-	-
ResNet-152	0.967	**0.923**	0.971	**0.978**	**0.986**
Inception v4	0.979	0.897	0.965	0.923	0.937
Inception-ResNet v2	**0.983**	0.917	**0.985**	0.857	0.981

Results on Montgomery Dataset are shown in Fig. 7 and Table 3. We found that many radiographs in the MC Dataset has large scale of black blocks and seriously disturbed histogram equalization, making the background of preprocessed graphs lighter than usual. We cut off the black blocks and resized the images, and saw an incredible improvement in results.

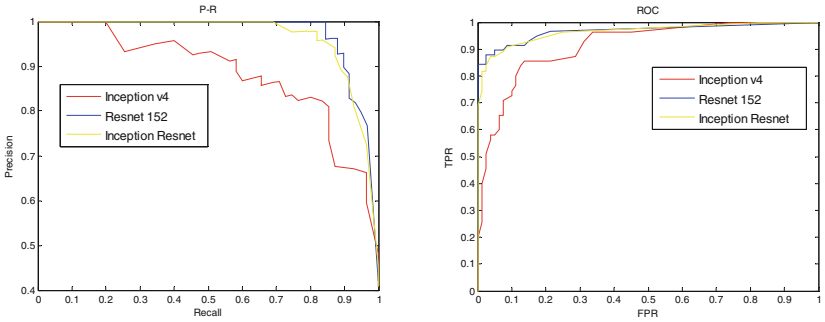

Fig. 7. P-R curves (left) and ROCs (right) of our networks on MC Dataset.

Table 3. Performance for MC Dataset. Last three are proposed methods.

Method	AUC	Accuracy	AP	Sensitivity	Specificity
U.K. Lopes et al.	0.926	0.810	-	-	-
Stefan Jaeger et al.	0.831	0.75	-	~0.5	~0.9
Sangheum et al.	0.884	0.674	0.890	-	-
ResNet-152	0.951	**0.890**	0.935	**0.711**	**0.955**
Inception v4	0.914	0.822	0.884	0.654	0.938
Inception-ResNet v2	**0.957**	0.844	**0.965**	0.618	0.913

Longitudinal and parallel experimental results show the superiority of our proposed network. The models achieved relatively good results on our own test set. It's hard to explain why ResNet 152 seems to do better than other network backbones on the public datasets. But our models undoubtedly showed adaptability to public datasets, outperforming the state-of-the-art results.

5 Conclusion and Future Work

We proposed an end-to-end network for pulmonary tuberculosis classification, including preprocessing, lung segmentation and classification. The system optimized the inference time, while guaranteeing the accuracy.

Future work will include (1) making specialized optimization on network backbones (2) optimization of preprocessing to increase adaptability of network (3) extending this system to the detection of focus of infection.

Acknowledgement. We would like to thank *Huiying Medical Technology (Beijing) Co., Ltd*. for providing essential resource and support for us.

References

1. World Health Organization (WHO): Global tuberculosis report (2017). http://www.who.int/tb/publications/global_report/en/. Accessed 26 May 2018
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1097–1105. Curran Associates Inc. (2012)
3. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
4. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection, pp. 2999–3007. In: IEEE Computer Society (2017)
5. He, K., Gkioxari, G., Dollár, P., et al.: Mask R-CNN. In: Computer Vision and Pattern Recognition (CVPR) (2017)
6. Bar, Y., Diamant, I., Wolf, L., et al.: Deep learning with non-medical training used for chest pathology identification. In: Medical Imaging 2015: Computer-Aided Diagnosis, p. 94140V (2015)

7. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: SIFT flow: dense correspondence across different scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5304, pp. 28–42. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88690-7_3
8. Lopes, U.K., Valiati, J.F.: Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput. Biol. Med. **89**, 135–143 (2017)
9. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: IEEE Computer Society, pp. 1–9 (2014)
10. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition, pp. 770–778. IEEE Computer Society (2015)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014)
12. Islam, M.T., Aowal, M.A., Minhaz, A.T., et al.: Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv (2017)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
14. Drozdzal, M., et al.: The importance of skip connections in biomedical image segmentation. In: Carneiro, G. et al. (eds.) Deep Learning and Data Labeling for Medical Applications
15. Jaeger, S., Karargyris, A., Antani, S., et al.: Detecting tuberculosis in radiographs using combined lung masks. Conf. Proc. IEEE. Eng. Med. Biol. Soc. **2012**(4), 4978–4981 (2012)
16. Hwang, S., et al.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: Medical Imaging 2016: Computer-Aided Diagnosis, p. 97852W. International Society for Optics and Photonics (2016)
17. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2014)
18. Chen, L.C., Papandreou, G., Kokkinos, I., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018)
19. Zhou, B., et al.: Learning deep features for discriminative localization. In: Computer Vision and Pattern Recognition, pp. 2921–2929. IEEE (2016)
20. Stefan, J., et al.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med. Surg. **4**(6), 475–477 (2014)