

Chejiao at ImageCLEFmed Tuberculosis 2020: CT Report Generation Based on Transfer learning

Jiao Che, Haiyan Ding, and Xiaobing Zhou*

School of Information Science and Engineering,
Yunnan University, Kunming 650091, P.R.China
Corresponding author: zhouxb@ynu.edu.cn

Abstract. Tuberculosis is a very common chronic infectious disease, so the study of CT report of tuberculosis has an important impact on clinical treatment. For the 3D CT image, it is easier to process the 3D images by transforming it into the 2D image along three dimensions and preserving the information of each dimension. We participate in the 2020 TB task: CT report generation. In order to improve the effect, we used the Mixup for data enhancement during training. Due to the small amount of data, we choose the ShufflenetV2 network model to use the transfer learning method for training and then fine-tune the model in training. Our team ranks the third in the ImageCLEF med-Tuberculosis 2020 challenge, it achieves mean_auc score and min_auc score of 0.791 and 0.682, respectively.

Keywords: Transfer Learning · ShufflenetV2 · Tuberculosis.

1 Introduction

Tuberculosis is a chronic infectious disease caused by Mycobacterium tuberculosis. According to the World Health Organization, about 130 years after the discovery of the disease, the disease is still a persistent threat and the main cause of death worldwide. It can invade many organs, especially pulmonary tuberculosis. Human infection with tubercle bacilli does not necessarily cause disease. When the resistance is reduced or the cell-mediated allergy is increased, it may cause clinical disease. If it can be diagnosed in time and treated reasonably, most of them can be cured clinically. Therefore, the correct detection and diagnosis of tuberculosis are very important.

At present, the clinical diagnosis of tuberculosis is mainly based on the comprehensive analysis of the patients' clinical manifestations, imaging data, and laboratory examination results. The imaging diagnosis method includes chest X-ray and CT examination. Chest X-ray examination can show the lymphadenopathy or miliary change of the lung and mediastinum. CT has a high resolution and

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.

can make a detailed evaluation of tuberculosis. It can also provide the basis for drug-resistant tuberculosis and active tuberculosis. Epidemiological investigation shows that there is a big delay between the occurrence of disease symptoms and clinical manifestations or diagnosis of tuberculosis. Finding fast and convenient means of early diagnosis will help to improve the efficiency of early diagnosis of tuberculosis.

ImageCLEF 2020 is an evaluation campaign organized as part of the CLEF initiative labs [1]. The campaign offers several research tasks and the ImageCLEFmedical is an area of great interest to us. The ImageCLEFmedical 2020 challenge includes a task based on tuberculosis CT [2]: To generate automatic lung-wise reports based on the CT image data. This task has an important impact on the clinical diagnosis of doctors and has practical value.

The rest of this paper is organized as follows: In the second part, we describe the research we did in this competition. In the third part, the specific tasks of this competition and the analysis of the data set are described in detail. Then, in the fourth part, we introduce the network model we used and the transfer learning method. In the fifth part, we introduce the relevant details of the experiment. In the sixth part, we report on our results and those of other teams. Finally, in the seventh part, we conclude and discuss the current and future directions.

2 Related Work

In recent years, deep learning has made exciting achievements in the field of medical image analysis and has rapidly become the current research hotspot. Machine learning and deep learning are also explored in the diagnosis of tuberculosis. Deep learning [3] is also used to diagnose CT images of lung diseases. The deep belief network and convolution neural network are used to classify pulmonary nodules in 3D tomography, and the effect of deep belief network is better than the convolution neural network and sift [4]. Lakhani et al showed that the combination of Alexnet and Googlenet performs best [5].

In 2018, Xiaoyun Hu et al proposed an artificial intelligence diagnosis and treatment system for CT TB detection, which first preprocessed CT images, used a two-dimensional convolution neural network to segment the lesion area, and then used the residual network to predict the true and false positive. It realized the automatic diagnosis of tuberculosis in CT images and provided doctors with diagnosis and treatment suggestions. This scheme had high recognition accuracy, greatly reduced the missed detection rate, and reduced the cost of manual screening. However, when segmenting the pulmonary tuberculosis lesions, only the two-dimensional characteristics of the image are used, and the influence of the three-dimensional characteristics of the CT image on the final segmentation effect is not considered. It can be seen that there is still much space for research and development in the automatic generation of CT reports of tuberculosis.

3 Task and Dataset

The Tuberculosis task of the ImageCLEF 2020 Challenge is to generate automatic lung-wise reports based on the CT image data. Each report should include the probability scores (ranging from 0 to 1) for each of the three labels and each of the lungs (resulting in 6 entries per CT). The result list of entries includes LeftLungAffected, RightLungAffected, CavernsLeft, CavernsRight, PleurisyLeft, PleurisyRight.

The 2020 ImageCLEFmed Tuberculosis task continues the 2019 Tuberculosis Subtask #2: CTR - CT report since it has an important outcome that can have a major impact in the real-world clinical routines. In order to make the task both more attractive for participants and practically valuable, the task of CT report generation is lung-based rather than CT-based, which means the labels for left and right lungs will be provided independently. Also, the dataset size is increased compared to the last year. This task includes 283 images in the training set and 120 in the test set. The task is to predict the presence of six types of findings in CT scans. Information about the corresponding labels is listed in Table 1.

Table 1: Presence of labels in the Training dataset

Filename	In Training set
LeftLungAffected	211
RightLungAffected	233
CavernsLeft	66
CavernsRight	79
PleurisyLeft	7
PleurisyRight	14

The images for the ImageCLEF tuberculosis task were provided as NIfTI 3D datasets. The 3D CT images which were provided have a slice size of 512×512 pixels and a number of slices varying from 50 to 400. The file format stores raw voxel intensities in Hounsfield units (HU) as well the corresponding image metadata such as image dimensions, voxel size in physical units, slice thickness, etc. Four slices of a CT were randomly selected to show in the Fig.1

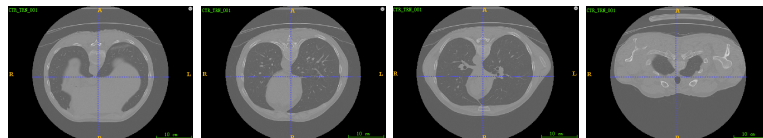


Fig. 1: Different slices of the same CT image

Both versions of the mask are available on the website for all patients. The first version of segmentation is retrieved using the same technique as the last years. The second version of segmentation is retrieved using non-rigid image registration scheme. The first version of segmentation provides more accurate masks, but it tends to miss large abnormal regions of lungs in the most severe TB cases [6]. The second segmentation on the contrary provides more rough bounds, but behaves more stable in terms of including lesion areas [7]. Two different versions of mask slices are shown in the Fig.2

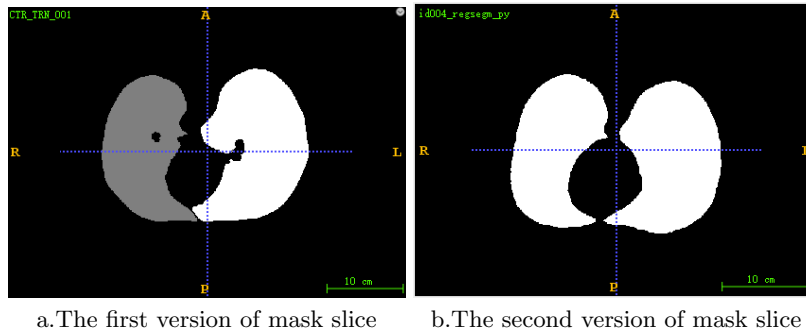


Fig. 2: Different versions of segmentation methods

4 Methods

4.1 Model

In recent years, deep CNN networks such as Resnet [8] and Densenet [9] have greatly improved the accuracy of image classification. But in addition to accuracy, computational complexity is also an important index to be considered in the CNN network. Over complex network may be slow, and some specific scenarios need low latency. In order to meet these needs, some lightweight CNN networks such as Mobilenet [10] and Shufflenet [10] were proposed, which made a good balance between speed and accuracy.

ShufflenetV2 is an improved version of ShufflenetV1, and ShufflenetV2 improves the deficiencies of ShufflenetV1. The experimental results show that ShufflenetV2 is much faster and more accurate on both platforms than the previous network. ShufflenetV1 uses two technologies: point-by-point group convolution kernel and bottleneck-like structure; and then introduces channel shuffle operation to enable information exchange between different groups of channels and improve accuracy. The point-by-point group convolution and bottleneck structure both increase the memory access cost. This cost cannot be ignored, especially for lightweight models. In addition, using too many packets to generate network

fragmentation will reduce the degree of parallelism. Element-level addition operations in shortcut connections are also undesirable. Therefore, in order to achieve higher model capacity and efficiency, the key issue is how to maintain a large number of channels with the same width, neither dense convolution nor too many groups. So our model chooses the lightweight neural network ShufflenetV2 [11]. The structure of the network model is shown in the following figure.

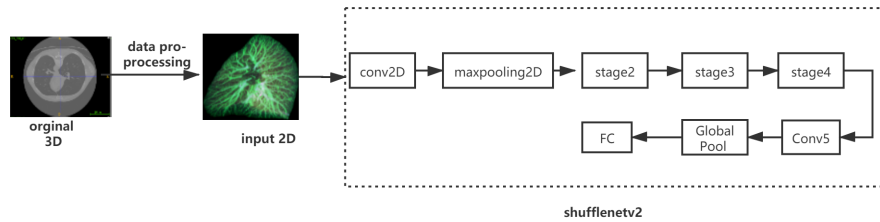


Fig. 3: The model we used in the competition

The stage part consists of one spatial down sampling and multiple basic unit. The two different block structures of Shufflenetv2 are shown in the following figure.

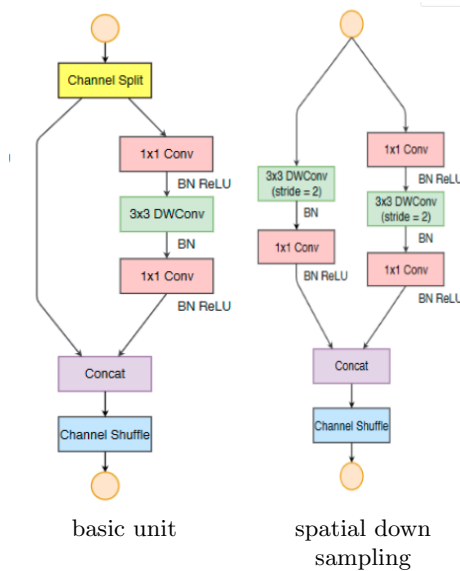


Fig. 4: Two different block structures of ShufflenetV2 (DWConv:depthwise convolution)

At the beginning of each unit, the input of the c characteristic channel is divided into two branches by Channel Split, with $c - c'$ and c' channels respectively. One branch is a shortcut stream, and the other branch contains three convolutions (and the three branches have the same number of channels). Two 1×1 convolutions are used. After convolution, the two branches are concatenated (Concat), so that the number of channels remains unchanged. Then Channel Shuffle operation is carried out to ensure the information exchange between the two branches. Depthwise convolution is used to minimize memory accesses with the same channel size. After shuffling, the next unit begins. Note that the Add operation no longer exists in ShufflenetV1, and ReLU and depthwise convolution only exist on one branch. Moreover, three consecutive element-by-element operations "Concat", "Channel Shuffle" and "Channel Split" are merged into one element-wise operation, and these changes are conducive to improving network efficiency. The complete network structure is made up of blocks, and the overall network structure is similar to Shufflenet, except that an additional convolutional layer is added to blend the characteristics.

4.2 Transfer Learning

Transfer learning is a machine learning method, which uses the parameters of the trained model developed for task A as the initial point, and re-transfers it to the model developed for task B to help the training of the new model. Using migration parameters to initialize the network can improve generalization performance. Even migrating from tasks that are not particularly similar is better than using random filters (or random parameters).

In this paper, the method of combining neural networks with transfer learning ideas is adopted. First, transfer learning is used to train the model. After a certain batch of the epoch, the method of fine-tuning is used instead. When applying deep learning in the field of image processing, it will be observed that the features extracted in the first layer are basically similar to Gabor filters and color blobs. Usually, the first layer is not particularly related to the specific image data set, and the last layer of the network is closely related to the selected data set and its task goal; the first layer feature is called general feature, the last layer is called specific feature. Because the data set is small and the data similarity is not high. So we choose to freeze the weight of the first k layers in the pre-training model and then retrain the next $n - k$ layers. Of course, the last layer also needs to be modified according to the corresponding output format. Because the similarity of data is not high, the process of retraining becomes very important. The shortage of the size of the new data set is made up by freezing the front k layer of the pre-training model.

5 Experiment

In this challenge, we use projection [12] to preprocess the original image. So the 3D CT image which is difficult to process can be converted into 2D

image processing. It retains the information of each dimension of the 3D image. According to the requirements of the competition, all labels are marked on the left and right lungs, so each CT image is divided into three axes (sagittal, frontal and axial) for projection. Additional enhancement operations are performed on the training image: Mixup data enhancement [13], random rotation within 10 degrees, changes in width and height, and random scaling. Data enhancement helps to prevent overfitting and memorizing the details of training images. If the training mode of Mixup is adopted, then two input images are read each time, assuming that they are represented by (X_i, Y_i) and (X_j, Y_j) , then a new image (x, y) can be synthesized by the following two formulas, and then the new image is used for training. It should be noted that more epoch should be trained when training the model in this way. λ in the formula is a super parameter, which is used to adjust the specific gravity of the synthesis. The value range is $[0,1]$.

$$\hat{x} = \lambda x_i + (1 - \lambda)x_j \quad (1)$$

$$\hat{y} = \lambda y_i + (1 - \lambda)y_j \quad (2)$$

Because this year’s CT report is based on lungs, the projected image is directly divided into the left and right lungs. Due to the small amount of data, we use the ratio of 3:1 to divide the training set and validation set, i.e., there are 213 images in the training set and 70 in the validation set.

We treat this task as multiple binary classification tasks. For different lesions, we adjust the CT value to obtain different two-dimensional projections. Then the projected pictures are fed into our model for training. In order to generate a CT report finally, we first classify leftlungaffed and rightlungaffed by using the binary classification method, then cavernsleft and cavernsright, pleurisyleft and pleurisyright. Six examples are classified three times, and then the six examples are averaged auc. In this experiment, we compare the Mobilenet model with the ShufflenetV2 model. On the validation set, Mobilenet performance results are shown in Table2.

Table 2: Mobilenet prediction results on the Validation set

lesion	auc
LeftLungAffected	0.783163
RightLungAffected	0.843391
CavernsLeft	0.770299
CavernsRight	0.798762
PleurisyLeft	0.625000
PleurisyRight	0.867188
Mean auc	0.781300

A GeForce RTX2080Ti GPU is used in this experiment. We use the method of transfer learning to train 100 epochs in the dataset and then use the method of fine-tuning to modify and compile the model. After the model is built, we

need to train 200 epochs. The optimizer we used is Adam, and the learning rate is set to 1e-05, the batch size is set to 16. The best results are obtained with the above parameters. The prediction results of the model we used on the validation set are shown in Table 3 below.

Table 3: ShufflenetV2 prediction results on the Validation set

lesion	auc
LeftLungAffected	0.887755
RightLungAffected	0.926724
CavernsLeft	0.818376
CavernsRight	0.859649
PleurisyLeft	0.625588
PleurisyRight	0.867188
Mean auc	0.83088

Through experiments we can conclude that the model we used have achieved very good results.

6 Results

This section shows final performance results of submitted runs in the CT report challenge. Imageclef med-Tuberculosis 2020 competition implements two evaluation methods, mean_auc and min_auc. The ID of our team is “chejiao”. During the competition, 120 testing data are supplied and are ranked based on mean_auc. Our work is ranked 3 with 0.791 mean_auc value and 0.682 min_auc. The main reason for our good results is that we used projection to process the data, and used the Mixup method to enhance the data, and finally combined with transfer learning to fine-tune the model. From the experimental results, the ShufflenetV2 model we used also improves the final result. The final submitted result is shown in Fig.9.

Table 4: Final result of the ImageCLEFmed Tuberculosis challenge

Participants	min_auc	mean_auc
SenticLab.UAIC	0.924	0.885
agentili	0.875	0.811
chejiao	0.791	0.682
CompElecEngCU	0.767	0.733
KDE-lab	0.753	0.698
Waqas_shelkh	0.705	0.644
uaic2020	0.659	0.562
JBTTM	0.601	0.432
sztaki_dsd	0.595	0.546

7 Perspectives For Future Work

In this competition, there are only seven cases of left pulmonary pleurisy, and the accuracy of the prediction verification set for pleurisy is low. As a result, the average accuracy is low. For the problem of too little data of pleurisy, we plan to add other authoritative data sets to better training models in future research. The results show that there is still space for improvement.

In this challenge, due to our lack of medical knowledge, the understanding of CT images is not deep enough. In future work, we will improve our medical knowledge so as to better understand and process data. In order to obtain high accuracy, it is recommended that medical knowledge should be embedded. In future experiments, we plan to use the current popular Efficientnet net model [14] and the new network design paradigm Regnet [15] proposed by the Kaiming and others in 2020 to carry out experiments, so as to obtain better results.

Acknowledgments

This work was supported by the Natural Science Foundations of China under Grants 61463050, the NSF of Yunnan Province under Grant 2015FB113.

References

1. Bogdan Ionescu, Henning Müller, Renaud Péteri, Asma Ben Abacha, Vivek Datta, Sadid A. Hasan, Dina Demner-Fushman, Serge Kozlovski, Vitali Liauchuk, Yashin Dicente Cid, Vassili Kovalev, Obioma Pelka, Christoph M. Friedrich, Alba García Seco de Herrera, Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Riegler, Pål Halvorsen, Minh-Triet Tran, Mathias Lux, Cathal Gurin, Duc-Tien Dang-Nguyen, Jon Chamberlain, Adrian Clark, Antonio Campello, Dimitri Fichou, Raul Berari, Paul Brie, Mihai Dogariu, Liviu Daniel Ștefan, and Mihai Gabriel Constantin. Overview of the ImageCLEF 2020: Multimedia retrieval in lifelogging, medical, nature, and internet applications. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, volume 12260 of *Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020)*, Thessaloniki, Greece, September 22-25 2020. LNCS Lecture Notes in Computer Science, Springer.
2. Serge Kozlovski, Vitali Liauchuk, Yashin Dicente Cid, Aleh Tarasau, Vassili Kovalev, and Henning Müller. Overview of ImageCLEFtuberculosis 2020 - automatic CT-based report generation. In *CLEF2020 Working Notes*, CEUR Workshop Proceedings, Thessaloniki, Greece, September 22-25 2020. CEUR-WS.org <<http://ceur-ws.org>>.
3. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
4. Kai-Lung Hua, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Yu-Jen Chen. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8, 2015.
5. Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582, 2017.

6. Yashin Dicente Cid, Oscar Alfonso Jiménez del Toro, Adrien Deppeursinge, and Henning Müller. Efficient and fully automatic segmentation of the lungs in ct volumes. In Orcun Goksel, Oscar Alfonso Jiménez del Toro, Antonio Foncubierta-Rodríguez, and Henning Müller, editors, *Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI*, CEUR Workshop Proceedings, pages 31–35. CEUR-WS, May 2015.
7. Vitali Liauchuk and Vassili Kovalev. Imageclef 2017: Supervoxels and co-occurrence for tuberculosis ct image classification. In *CLEF2017 Working Notes, Series = CEUR Workshop Proceedings, Year = 2017, Publisher = CEUR-WS, Month = September 11-14, Address = Dublin, Ireland*.
8. 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*, pages 770–778, 2016.
9. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
10. Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and H Adam Mobilenets. Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
11. Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018.
12. Vitali Liauchuk. Imageclef 2019: Projection-based ct image analysis for tb severity scoring and ct report generation. In *CLEF (Working Notes)*, 2019.
13. Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2019.
14. Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.
15. Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10428–10436, 2020.