Diagnosis of COVID-19 Based On Slice-level CT Scans Classification

Wentao Pan
School of Information and Software Engineering
UESTC
Chengdu, China
panwentao1301@gmail.com

Jiangpeng Yan
Department of Automation
Tsinghua University
Beijing, China
yanjp17@mails.tsinghua.edu.cn

Xiu Li
Shenzhen International Graduate School
Tsinghua University
Beijing, China
li.xiu@sz.tsinghua.edu.cn

***Abstract*—In this paper, we proposed a novel coronavirus pneumonia diagnosis framework based on slice level classification network, which was responsible for COVID-19 Identification Signal Processing Grand Challenge. We trained the slice-level classification model using slice labeled data and finally obtain patient-level predictions through the fusion block. The performance of our workflow on the validation set is as follow: accuracy 0.898, normal sensitivity 0.917, COVID-19 sensitivity 0.964, CAP sensitivity 0.684.**

Keywords—COVID-19, slice-level, fusion block

I. Introduction

Since the outbreak of Novel Coronavirus(COVID-19), over 100 million people have been infected worldwide. Existing studies have shown that the virus is transmitted through droplet transmission, fecal-oral route, conjunctiva, and fomites, with a long incubation period, powerful infectivity[1]. It is necessary to quickly diagnose suspected patients and carry out isolation treatment, to block the spread of the virus and control the development of the epidemic situation. It is possible to use computer vision technology for rapid diagnosis based on CT due to COVID19's imaging manifestations, such as interlobular septal thickening and an obvious crazy-paving pattern, etc.[2].

The goal of IEEE SPGC Grand Challenge COVID-19 Diagnosis in 2021 is to develop an automated framework for identifying COVID-19 infection with CT scans. The official data set includes three types of patients: normal, COVID-19, and community pneumonia. Community pneumonia and Novel Coronavirus pneumonia are similar in CT imaging manifestations, which may lead to misjudgment in practical diagnosis. The dataset includes continuous CT images of lung and patient-level label for each patient, in addition a slice-level labeled subset. The CT scans are acquired by different scanners in multiple medical centers, using different settings for slice thickness, effective mAs, and exposure time.

We proposed a novel coronavirus pneumonia diagnosis framework based on slice-level classification network. In our framework, All CT slices of a patient are taken as input. Firstly, the slices are preprocessed to reconstruct the lung's 3D structure and to unify voxel size. Then a slice selection algorithm is used to filter closed lung slices. The selected slices are preprocessed and then fed into the slice-level classification network to get the score of each category and combined into the case's feature vector. Finally, the case level prediction is obtained by the fusion block. The final performance on the validation set is as follow: accuracy 0.898, normal sensitivity 0.917, COVID-19 sensitivity 0.964, CAP sensitivity 0.684.

II. Method

Our overall framework is shown in Fig. 1. A series of continuous lung CT slices of the case is taken as the input. First of all, all lung CT slices are processed by Preprocess Module 1 to reconstruct the 3D structure of the lung and to unify the size of the voxel. Lung segmentation is performed slice by slice after Preprocess module 1. Then use lung segmentation mask and slice selection algorithm to do slice selection. Before fed into slice-level convolutional neural network classifier, the selected slices are processed by Preprocess Module 2. Finally, the output of slices through slice-level convolutional neural network classifier are fed into fusion block to get patient-level prediction. The detailed description of each module is as follows.

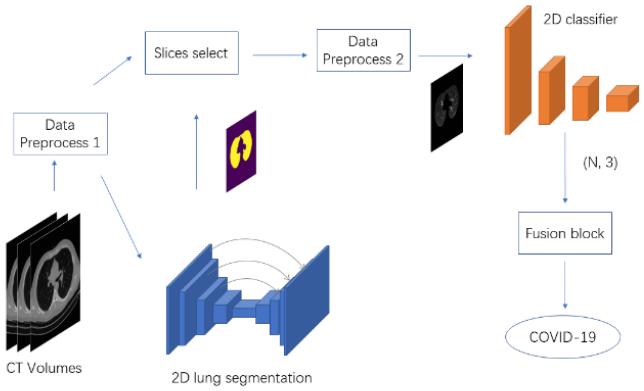

Fig. 1. novel coronavirus pneumonia detection framework.

A. *Preprocess Module 1*

CT scans are obtained through continuous scanning, and can be restored to the 3D structure of organs by reconstruction. Due to different scanning machine settings, the interval between adjacent slices (slice thickness) is different, that means the voxels represented by CT scans are different. The generalization of the model may be poor when the model is applied to test samples with different voxel. We resampled all slices for each patient to ensure the same slice thickness.

The pixel matrix of each slice is transformed into HU value before resample. Then HU value matrix of each slice is stacked according to the actual relative position of the physical world to form a 3D structure. We set the standard slice thickness to be 2 mm due to the slice thickness of slice-level labeled data is 2 mm. If the slice thickness of the current patient is not 2mm, interpolation is used for scaling and resampling to make the voxels the same.

B. *Lung segmentation and slice select algorithm*

As for CT scan, the lesions of the two diseases, COVID-19 and CAP, are all located in the lung area, and other information such as background is meaningless for the current task. Therefore, we use the lung segmentation results combined with CT images to restrict the attention of the network to the lung area.

All sections of the lung are shown in Fig. 2. The lung area of the front and back CT images is smaller, which belongs to closed lung. Compared with the middle part of the open lung, the lung area is too small or even no, such slices have little useful information, which may cause noise for training. So we discard these slices with small lung area. The existing work is based on the observation of the data set, using the hard coding method for slice selection[3]. If the lung position of another test sample changes, the performance will be worse. We propose a slice selection algorithm based on lung segmentation.

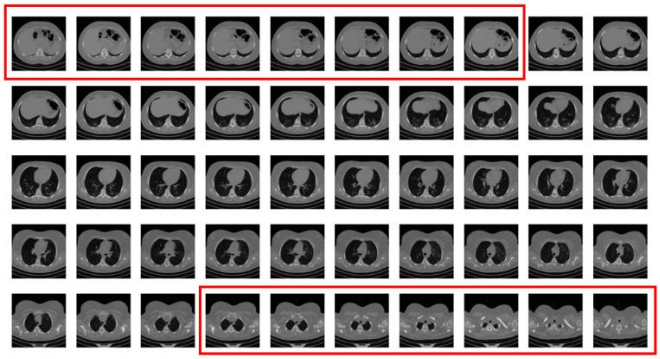

Fig. 2. Closed lung and open lung. (*We removed these scans in the red box those are closed lung slcies.*)

The lung selection algorithm selects according to the area of the lung region, inputs a single slice, obtains the lung mask by using the lung segmentation network, and calculates the number of pixels occupied by the lung without distinguishing the left and right lung. If the number is lower than the threshold T, the slice is deleted. T, a hyperparameter, was set to 5000 through experiments.

As for the lung segmentation task, we use U-net[4], which achieves state-of-the-art performance on very different biomedical segmentation applications. We use the pre-trained model provided by Johannes et al. The model was trained on a large and diverse dataset that covers a wide range of visual variability[5].

C. *Preprocess Module 2*

The HU values of different organs and tissues have different ranges. We can filter out unnecessary tissues by windowing. We window Hu values to -1200 to 700 and Normalized into 0 to 1. Then use lung mask to set the matrix outside the lung to 0. Repeatedly stack image as input..

D. *Slice level CNN classifier*

We use slice-level labeled data to train our classification model. First, we select the slice through slice select algorithm and filter out normal slices from COVID-19 and CAP cases. Then, the CT and lung segmentation mask are used as input, training the model, and we use densenet121[6] as the backbone.

E. *Fusion Block*

Due to the error of the classification model, it is difficult to transform the prediction results of the slices to the case-level. Under normal circumstances, the number of infected slices in COVID-19 and CAP cases is uncertain. It is unreasonable to predict case level label by the number of infected slices in each category. Refer to existing work [7], we use the fusion block to transform the prediction results of the slices to the case-level prediction.

Through the previous steps, we can get the output of three categories of a series of slices of a case. Apply softmax to each slice's output then fed to fusion block. Fused scores of 3 classes (Normal, COVID-19, CAP) are computed in the module. The fused score of Normal class is obtained by averaging all slices in a patient, but the fused scores of COVID-19 and CAP are obtained by averaging top-K highest scores of each class. Through experiments, choose 3 as the value of K.

III. RESULTS

Our final performance is shown in the TABLE I. .

TABLE I. OUR WORKFOLW PERFORMANCE

Method	level	ACC.	Sen. Normal	Sen. COVID-19	Sen. CAP
Train	Patient	0.885	1.0	0.897	0.707
Valid	Patient	0.898	0.917	0.964	0.684

In our workflow, we set Hyperparameter T 5000 in Slice selected part, not using normal slices in COVID-19 and CAP patients for training, and repeat stack CT multiply mask of lung area as the input. The detail information for dataset D, is shown in the following table. We used densenet121 as backbone. We trained our model use mini-batch stochastic gradient descent with momentum 0.9 and batch size 13 on 1 GPU(GeForce RTX 1080Ti). The learning rate was set to 0.01 at the beginning and used exponential learning rate decay with the alpha of 0.9. The weight decay parameters was set to 0.0001 and. Randomly rotated -20 degrees to 20 degrees and randomly cropped after padding 50 pixels for data augmentation.

In order to get the best setting of K, we have tried the different values of K on the training set. TABLE II. presents the experimental results on a different value of K. We set K to 3, due to its best performance. Finally, we evaluated out workflow on the validation set, results are shown in TABLE I.

TABLE II. THE PERFORMANCE OF DIFFERENT K

Top-K	ACC	Sen. Normal	Sen. COVID-19	Sen. CAP
1, 2	**0.876**	**0.962**	**0.914**	**0.659**
3	**0.885**	**1.0**	**0.897**	**0.707**
4, 5	**0.880**	**1.0**	**0.888**	**0.707**
6, 7, 8	**0.876**	**1.0**	**0.879**	**0.707**
9	**0.870**	**1.0**	**0.879**	**0.683**

IV. EXPERIMENT AND ANALYSIS

The default settings for our experiment are as follow. Densenet121 was used as the backbone. We used stochastic gradient descent with momentum of 0.9 for optimization, a weight decay of 0.0001, and exponential learning rate decay with alpha equal to 0.9. Randomly rotated -20 degrees to 20 degrees and randomly cropped after padding 50 pixels for data augmentation. Our experiments were performed on a single GPU, GeForce GTX 1080Ti.

A. *Slice Level Dataset Processing.*

In order to alleviate the problem of data imbalance, I tried not to use normal slices from sick patients. We experimented with two methods to form slice level dataset. 1. All slices in normal patients and uninfected slices in COVID-19 and CAP patients are regarded as normal slices; Infected slices in COVID-19 patients and CAP patients are regarded as COVID-19 slices and CAP slices. 2.Only slices in normal patients are regarded as normal slices; Others are the same with method 1. The slice select part is not used here, and only use CT scans as input for classifier. The results are shown in the TABLE III. , method 2 shows a better performance. Method 2 not use normal slices in COVID-19 patients or CAP patients, but not change the COVID-19 and CAP slices, leading the performance for each class improving. So we think there are some annotation error in slice level label. We used method 2 in the following experiments.

TABLE III. PERFORMANCE OF METHOD 1 AND METHOD 2

Method	level	ACC.	Sen. Normal	Sen. COVID-19	Sen. CAP
Method1	Patient	0.571	0.667	0.655	0.211
Method2	Patient	0.704	0.458	0.836	0.631

B. *Use Lung segmentation.*

As for CT scan, the lesions of the two diseases, COVID-19 and CAP, are all located in the lung area, We think using a lung segmentation mask will help the network focus on the lung area, which is easier to learn the features of lesions. We verified this by changing the input of the slice level classification network. a. Stack CT, CT, CT as input; b. Stack CT*mask, CT*mask, CT*mask as input; c. Stack CT, mask, CT*mask as input. In order to avoid the interference of slices without lungs, we set T 0 in slice select part. The results are shown in TABLE IV. No significant difference between the three methods was evident in slice-level validation. But there was a significant difference in patient-level validation, the performance of method b is much better. So we used the method b in our method.

TABLE IV. A

Method	level	ACC.	Sen. Normal	Sen. COVID-19	Sen. CAP
a	Slice	0.818	0.892	0.621	0.776
b	Slice	0.835	0.934	0.627	0.557
c	Slice	0.811	0.917	0.570	0.625
a	Patient	0.694	0.458	0.836	0.579
b	Patient	0.837	0.875	0.945	0.474
c	Patient	0.714	0.542	0.909	0.368

C. *The setting of T*

Observing the data set, we found that some slices with small lung are labeled as infected slices. This will inevitably generate noise if we use these slices for training. But If the threshold is too large, more slices will be filtered out, resulting in too little data. In order to choose a better setting of T for the slice select algorithm, we conducted a rough search. In order to ensure patients have enough slices for diagnosis, we set the upper bound of T to 20000. For each case, we perform patient-level validation on the validation set. From the TABLE V. we can see that its performs best when T is set to 5000. So we choose 5000 as the default setting for T.

TABLE V. PERFORMANCE OF DIFFERENT SETTING T

T	level	ACC.	Sen. Normal	Sen. COVID-19	Sen. CAP
0	Patient	0.837	0.875	0.945	0.474
5000	Patient	0.898	0.917	0.964	0.684
10000	Patient	0.816	0.708	0.981	0.473
15000	Patient	0.857	0.792	0.982	0.579
20000	Patient	0.857	0.833	0.964	0.579

D. *Try to use unlabel data*

Slice-level labeled data is only a small part of the whole dataset. The labeled subset of the data contains 4, 993 number of slices demonstrating infection and 18, 416 number of slices without infection. Inspired by the work of semi-supervised[8], we tried to use semi-supervised learning to train slice-level classifier, to make better use of unlabeled data.

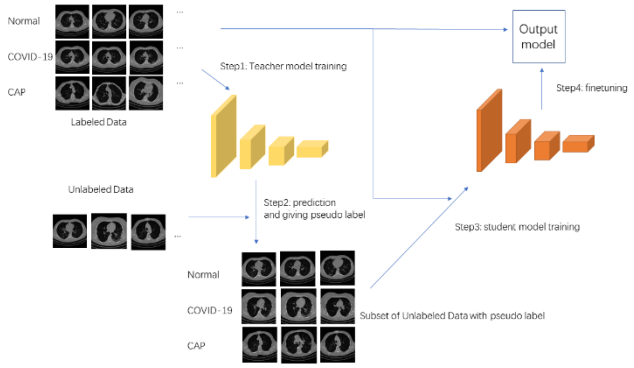

Fig. 3. Semi-supervised learning workflow.

Fig. 3provides the workflow of semi-supervised learning. the teacher model was obtained by training the model with labeled dataset D. The teacher model was used to predict unlabeled dataset U to get the pseudo labels. We selected slices, whose pseudo-label and case-level label are the same or pseudo label and ground truth label are the same in those normal slices in COVID-19 and CAP cases, to add to the dataset to construct a new dataset D^. Details of the dataset are shown in the TABLE VI. Then Trained a new student model on D^. The addition of data with pseudo label changed the distribution of the original labeled data. We used a weighted loss to compensate. Finally fine-tune the trained student on the labeled set D. The performance of the teacher model and student model are shown in the TABLE VII. , no increase in performance was detected. We analyzed that data in the dataset U can be classified correctly by the teacher model, and joining them in training caused the over-fitting problem.

TABLE VI. DETTAIL OF DATASETS.

Dataset	Total	Normal	COVID-19	CAP
D	8710	6188	1955	567
U	7784	1950	4982	825
D^	15909	8138	6379	1392

TABLE VII. PERFORMANCE OF STUDEN MODEL AND TEACHER MODEL.

Model	dataset	ACC.	Sen. Normal	Sen. COVID-19	Sen. CAP
teacher	Train	0.885	1.0	0.897	0.707
student	Train	0.909	1.0	0.905	0.805
teacher	Valid	0.898	0.917	0.964	0.684
student	Valid	0.857	0.792	0.927	0.737

E. Shortcoming

CT has a series of continuous scans. There is some information contained physical continuum, But our workflow did not use information between consecutive slices which is the major limitation.

CONCLUSION

In our workflow, We use lung segmentation to force limit the network's attention. We use the slice selection algorithm to remove the noise caused by closed lung slices on the basis of ensuring enough slices for diagnosis. The fusion block realized the prediction from slice level to patient-level. These tricks, proven effective in experiments, formed our workflow.

REFERENCES

[1] Kakodkar P, Kaka N, Baig M (April 06, 2020) A Comprehensive Literature Review on the Clinical Presentation, and Management of the Pandemic Coronavirus Disease 2019 (COVID-19). Cureus 12(4): e7560. DOI 10.7759/cureus.7560.

[2] Dai WC, Zhang HW, Yu J, et al. CT Imaging and Differential Diagnosis of COVID-19. Can Assoc Radiol J. 2020;71(2):195-200. doi:10.1177/0846537120913033.

[3] Rahimzadeh, M.; Attar, A.; Sakhaei, M. A Fully Automated Deep Learning-based Network For Detecting COVID-19 from a New And Large Lung CT Scan Dataset. Preprints 2020, 2020060031 (doi: 10.20944/preprints202006.0031.v3).

[4] Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Hofmanninger, J., Prayer, F., Pan, J. et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp 4, 50 (2020). https://doi.org/10.1186/s41747-020-00173-2.

[6] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[7] Jin, C., Chen, W., Cao, Y. et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 11, 5088 (2020). https://doi.org/10.1038/s41467-020-18685-1.

[8] Zeki Yalniz, I., Jégou, H., Chen, K., Paluri, M., and Mahajan, D., "Billion-scale semi-supervised learning for image classification", 2019

.