A hybrid and fast deep learning framework for Covid-19 detection via 3D Chest CT Images

Shuang Liang
University of Science and Technology Beijing
30 Xueyuan Road, Haidian District, Beijing 100083 P. R.China
liangshuang@xs.ustb.edu.cn

Weicun Zhang
University of Science and Technology Beijing
30 Xueyuan Road, Haidian District, Beijing 100083 P. R.China
weicunzhang@263.net

Yu Gu *
Guangdong University of Petrochemical Technology
No. 139, Guandu 2nd Road, Maoming 525000, Guangdong
ygu@ccmu.edu.cn

Abstract

In this paper, we present a hybrid deep learning framework named CTNet which combines convolutional neural network (CNN) and transformer together for the detection of COVID-19 via 3D chest CT images. It consists of a CNN feature extractor module with SE attention to extract sufficient features from CT scans, together with a transformer model to model the discriminative features of the 3D CT scans. Compared to previous works, CTNet provides an effective and efficient method to perform COVID-19 diagnosis via 3D CT scans with data resampling strategy. Advanced results on a large and public benchmarks, COV19-CT-DB database, was achieved by the proposed CTNet with a macro F1 score of 88.21% on the validation set, which lead ten percentage over the state-of-the-art baseline approach proposed together with the dataset. Notably, the inference speed of the proposed framework is about ten times faster than that of the typical CNN frameworks which make it more promising in actual applications.

1. Introduction

The Coronavirus Disease 2019 (COVID-19), a highly infectious disease, has become a global pademic and posed serious threats to human worldwide. In order to prevent further spreading of COVID-19 and treat the infected patients instantly, various examination methods have been proposed for the diagnosis of COVID-19 [2, 5], of which rRT-PCR is usually considered as the golden standard for the diagnosis of COVID-19 [11]. However, due to the limitation of sample collection and transportation, the senstivity of rRT-PCR might encounter some problems. As reported by the World Health Organization (WHO), the lung infection was detected in the autosies [13]. Thus, medical imaging of chest radiography is proved to be useful for rapid COVID-19 detection [16]. Computed Tomography (CT) was considered as the precise examination tools since it provided a 3D view of organs especially the lung, which could be used to locate the lesion areas [3]. In the CT examination, large volume of CT scans are obtained from persons suspected of COVID-19, which poses heavy workload on physicians and radiologists to diagnose COVID-19. There is a great need to develop some auxilary reliable methods in the medical imaging analysis of COVID-19. As demonstrated by some studies, machine and deep learning methods might be a potential approach for the detection of COVID-19 [14].

However, in our pespective, there exists two main shortages for the rapid and accurate diagnosis of COVID-19 using current avaliable DL frameworks. First, the size of datasets used in these works is limited which makes it hard to avoid the overfitting problem and reduces the usability of their works. Second, the preprocess methods of 3D CT scans are not ideal. Some works only take the slices of CT

*Correspondence: ygu@ccmu.edu.cn; Tel.: +86-1850-008-7987 (Y.G.)

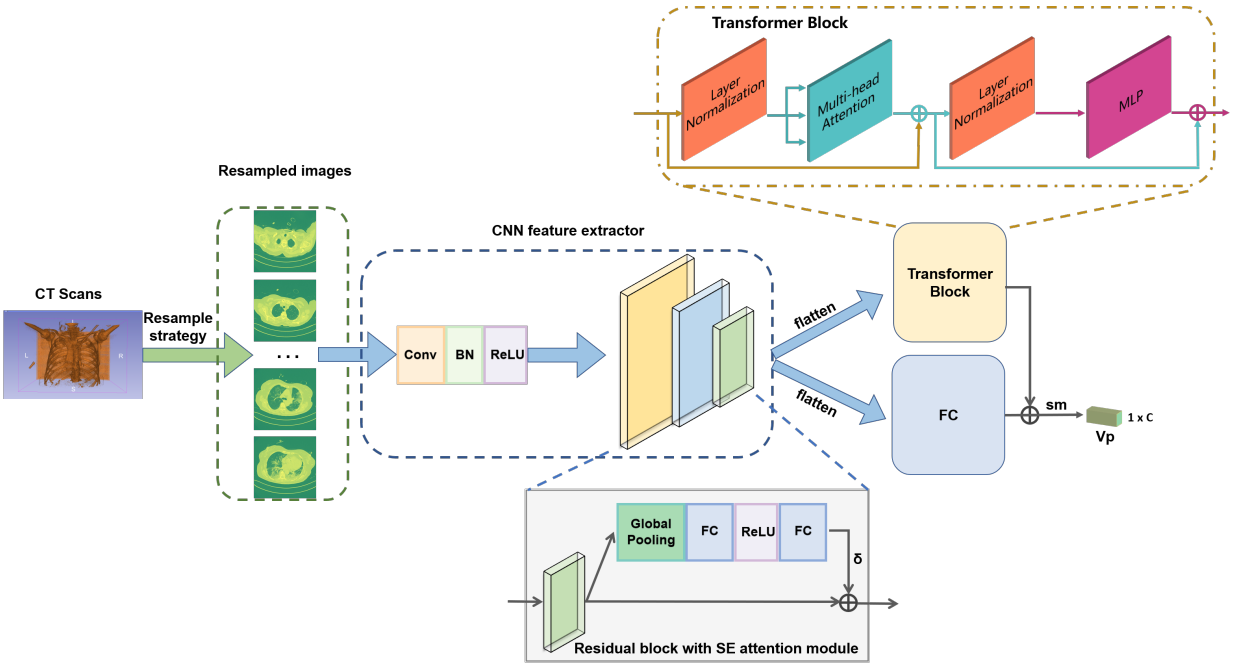

Figure 1. The framework of the proposed CTNet.

scans that are suspected to have related sympotoms into account [12, 6, 1, 15]. But in actual use case, these methods require physicians and radiologists to first screen all the slices and provided the suspected slices to the DL frameworks which is not ideal to incorporate the method into clinical practice. The other works take full volumn of CT scans into consideration. Although this method ensures the discriminative features were included in the input data, the training and inference processes are quite time-cosuming. Moreover, given the large size of each input data, the batch size is limited which might make the training process unstable and hard to converge. Therefore, it is important to develop DL frameworks on a large and clean COVID-19 detection database and make use of 3D chest CT data properly and effeciently.

In this paper, we proposed a hybrid deep learning framework named CTNet that combines CNN and transformer together for the detection of COVID-19 via 3D chest CT images. We also proposed a hypothese that the relevant lesion area is continuous in lung and the features of lesion area is discriminative for the detection of COVID-19. Based on the hypothese, we proposed a data resampling method to reduce the size of input data while preserving useful information. The contribution of our paper could be summarized as following:

1. A effective data resampling method was proposed for 3D CT scans to reduce the size of input while preserving sufficent information.

2. A hybrid deep learning framework is proposed for the detetion of COVID-19 which combines a CNN to mining effective features and a Transformer to conduct feature aggregation.

3. Good performances were achieved by the proposed framework with a macro F1 score of 88.24% on the validation set of COV19-CT-DB dataset, which is more than 18 percent lead of that of the baseline results released along with the dataset.

The rest of this paper mainly focuses on the introduction of the proposed method and the corresponding results and conclusion.

2. Implementation details

As we know, peformances of DL methods highly rely on the size and quality of samples. If the size of samples is quite small, it might be difficult to obtain a framework with high generative ablity. On the other hand, if the size of samples is large, it is also important to make use of the samples properly and effectively. The proposed framework, called CTNet, provides a novel pespective on how to make effective use of samples. The pipeline of the CTNet is shown in Fig.1, from which we can have a intutive sight of the whole framework. First of all, the proposed DL framework, called CTNet, is a hybrid DL method consisting of three componets: the input module with data resampling strategy, the CNN feature extractor module with SE attention module [4], and the information aggregation module with the transformer and fully connected (FC) layers.

Given the 3D CT scans as the original input, the input module applies a resampling strategy following the uniform distribution to ensure slices in different positions are included in the input data for a comprehensive representation of the original 3D CT scans with a fixed number. The resampled images generated by the input module are then sent to the CNN feature extractor to extracted sufficient and effective features. The extracted features were then flatten to vectors and sent to two branches, transformer brach and FC branch, and the vectors generated by the two branches were fused using element-wise-add (EWA) operation. The vector is then sent to the softmax function to predict the probability vector for the detection of COVID-19.

2.1. Data resampling strategy

As we introduced in Introduction, the preprocess method of 3D CT scans matters a lot considering the actual application in clinical. Unlike the previous works that only take the annotated slices or take all the volumns of CT scans, we proposed a data resampling strategy that following the uniform distribution. Since the total number of CT scans varies from different cases, we firstly set a variable named totalnum to deal with this situation. The resampled number of CT scans is controled by the variable renumCT which could be adjusted according to the actual use case (considering the variance of computation power and memory size). The algorithm contains two branches, one for resampling and the other for oversampling which is switchable according to the total number of CT scans and the resampled number of CT scans. The output is the indexes array of the 3D CT images. In our study, the value of the renumCT is set to 32 as default. The process is summarized as following:

Algorithm 1 Data resampled strategy for 3D CT scans

Input: $\mathrm{totalnum}, \mathrm{n}$
Output: indexes(The indexes array of the resampled CT scans)

```
renumCT ← n
i ← 0
indexes ← []
if totalnum >= renumCT then
    diff = totalnum/(renumCT + 1)
    while i < renumCT do
        indexes.append(int(i * diff))
        i+ = 1
    end while
else
    while i < renumCT do
        indexes.append(random.choice(totalnum))
        i+ = 1
    end while
end if
```

2.2. CNN feature extractor with SE attention module

The CNN feature extractor is a multi-stage residual neural network that incorporate the SE attention module to increase the discriminative ability. As shown in Figure. 1, the CNN can be seperated to two parts, the input convolution layer and the residual block. The number of the input channels for the input convolution layer is the same as the number of the resampled CT images. The following part is the residual block that adopted the SE module. In order to achieve a good trade-off between performance and speed, we adopted the Resnet-18 architecture as the backbone structure. The structure of the SE module is also demonstrated in Fig.1, which consists of a global pooling, two FC layers, the ReLU function, and the sigmoid function to generate attention vectors with the value range from 0 to 1. Given the resampled CT images as input, the extractor generates feature maps with N channels. In our study, the size of the generated feature maps is $512 \times 28 \times 28$. Global pooling and flatten operations are performed to transform the feature maps to a feature vector with a size of 1×2048.

2.3. Information aggregation module

As introduced previously, the 3D CT scans were resampled and sent to the CNN feature extractor. The output feature vector is then sent to two braches, the transformer block and the FC block, to fuse information and generate the final prediction. The structure of the transformer is shown in Fig.1, which consists of two layer normalization, a multi-head attention module and the multilayer perception module. The probability vectors generated by the two branches are fused using EWA operation to predict the diagnosis result. Here, the transformer block mainly acts as a feature encoder module that owns the ability to capture long range dependecies between differnt elements of the feature vectors, thus bring the framework a nonlinear discriminative

ability in the prediction. And the FC layer provide a linear transform from the feature vectors to the predictive vectors. The combination of the linear predictions and nonlinear predictions offers a fesible information aggregation approach which helps improve the discriminative ability of the proposed framework.

2.4. Training details

The framework is trained in an end-to-end and from-scratch manner on a local machine with two 1080ti GPUs. The learning rate is set to 0.01, and is decreased following the step adjust strategy. The total epoches is set to 120, and the step epoches are 50 and 100. The number of resampled CT scans is set to 32 in our experiments and the batch size is set to 32. The size of each CT scan is resized to 224 $\times$ 224. The size of the parameters memory for the CTNet is 21.93 MB.

3. Results

3.1. COV19-CT-DB database

The dataset used in this paper for training and validation is the COV19-CT-DB dataset [7, 8, 10, 9], which consists of 5,000 chest CT scans that are annotated as COVID-19. Data was collected in the period from September 1, 2020 to March 31, 2021. The data were aggregated from different hospitals, containing anonymized human lung CT scans with signs of COVID-19 and without signs of COVID-19. The database was seperated into three set, the training set, the validation set and the test set, of which the training and validation set are publicly avaliable currently while the test dataset is preserved for competition purpose. The training set consists of 1560 3D CT scans, including 690 COVID-19 cases and 870 Non-COVID-19 cases. The validation set contains 374 3D CT scans, including 165 COVID-19 cases and 209 Non-COVID-19 cases.The test set contains 3455 3D CT cases. Each CT scan owns different numbers of CT scans, ranging from 50 to 700. The details were shown in Table. 1.

Table 1. Distribution of the COV19-CT-DB database.

KACD	train	val	test	total
COVID-19	690	870	-	1560+
Non-COVID-19	165	209	-	374+
Total	855	1079	3455	5389

3.2. Evaluation Metrics and experiments

As introduced in the database, the 'macro' F1 score is used as the evaluation metric and the score of 0.70 is reported as the baseline score. The 'macro' F1 score is defined as the unweighted average of the class-wise/label-wise F1-scores. In our experiments, we conduct two experiments to validate the proposed framework, one only uses the probability vector generated by the FC layer and the other uses the fused probability vector to obtain the diagnosis results. The performance on the validation set and the test set can be seen from Table 2. Good performance was achieved by the CTNet, with a best 'macro' F1-score of 88.23% on validation set, which is 18.23 percent higher than that of the baseline approach. Notably, the number of resampled CT scans is set to 32 in our experiments while that for the baseline is 700. Besides, the proposed CTNet ranked 7th in the competition and the inference process of the CTNet cost only 380ms in diagnosing each case while that for the other methods involved in this competition are at least ten times slower than CTNet. Therefore, the CTNet might be more applicable in actual use cases considering its good performance and high inference speed. Besides, as shown in Table 2, the combination of FC and Transformer brings the CTNet better generalization ability. Despite the achievements made by the proposed CTNet, we can also observe the performance gap between validation set and test set. Based on a naive analysis, the reason of the performance gap might be attributed to the incomplete sampling strategy given the distributions might be different between validation set and test set. The sampling strategy could be considered a context-free pre-process method while there also exists context-related pre-process methods including lung segmentation. However, the context-related pre-process methods need consume additional computing resources which is time-consuming and parameter inefficient. Therefore, the effective approach with good priori knowledge in the pre-process of the CT images is remain an open challenge.

Table 2. Performance of the CTNet framework

Method	Validation set	Test set
Baseline	70.00%	67.00%
CTNet (FC)	87.96%	74.74%
CTNet (Fused)	88.23%	78.86%

4. Conclusion

We present a hybrid deep learning framework called CTNet for the detection of COVID-19 via 3D CT scans. The proposed CTNet provides a novel resampling strategy and network architecture for the effective mining of 3D CT scans. Experimental results demonstrated that our resampling strategy and framework achieved a good trade-off between speed and accuracy with a lead of 18.23 percent on the macro F1 score and ten times faster than that of the SOTA baseline method. Ablation studies showed CTNet's superior ability to deal with COVID-19 detection with less requirement of computation power and data.

References

[1] Parnian Afshar, Shahin Heidarian, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Faranak Babaki Fard, Kaveh Samimi, Konstantinos N Plataniotis, and Arash Mohammadi. Covid-ct-md, covid-19 computed tomography scan dataset applicable in machine learning and deep learning. *Scientific Data*, 8(1):1–8, 2021.

[2] Roohallah Alizadehsani, Zahra Alizadeh Sani, Mohaddeseh Behjati, Zahra Roshanzamir, Sadiq Hussain, Niloofar Abedini, Fereshteh Hasanzadeh, Abbas Khosravi, Afshin Shoeibi, Mohamad Roshanzamir, et al. Risk factors prediction, clinical outcomes, and mortality in covid-19 patients. *Journal of medical virology*, 93(4):2307–2320, 2021.

[3] Wei-cai Dai, Han-wen Zhang, Juan Yu, Hua-jian Xu, Huan Chen, Si-ping Luo, Hong Zhang, Li-hong Liang, Xiao-liu Wu, Yi Lei, et al. Ct imaging and differential diagnosis of covid-19. *Canadian Association of Radiologists Journal*, 71(2):195–200, 2020.

[4] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[5] Adil Khadidos, Alaa O Khadidos, Srihari Kannan, Yuvaraj Natarajan, Sachi Nandan Mohanty, and Georgios Tsaramirsis. Analysis of covid-19 infections on a ct image using deepsense model. *Frontiers in Public Health*, 8, 2020.

[6] Hoon Ko, Heewon Chung, Wu Seong Kang, Kyung Won Kim, Youngbin Shin, Seung Ji Kang, Jae Hoon Lee, Young Jun Kim, Nan Yeol Kim, Hyunseok Jung, et al. Covid-19 pneumonia diagnosis using a simple 2d deep learning framework with a single chest ct image: model development and validation. *Journal of medical Internet research*, 22(6):e19569, 2020.

[7] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian, and Stefanos Kollias. Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis. *arXiv preprint arXiv:2106.07524*, 2021.

[8] Dimitrios Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and S Kollias. Deep transparent prediction through latent representation analysis. *arXiv preprint arXiv:2009.07044*, 2020.

[9] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias, and Georgios Tagaris. Deep neural architectures for prediction in healthcare. *Complex & Intelligent Systems*, 4(2):119–131, 2018.

[10] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and Stefanos D Kollias. Transparent adaptation in deep medical image diagnosis. In *TAILOR*, page 251–267, 2020.

[11] Chunqin Long, Huaxiang Xu, Qinglin Shen, Xianghai Zhang, Bing Fan, Chuanhong Wang, Bingliang Zeng, Zicong Li, Xiaofen Li, and Honglu Li. Diagnosis of the coronavirus disease (covid-19): rrt-pcr or ct? *European journal of radiology*, 126:108961, 2020.

[12] Vruddhi Shah, Rinkal Keniya, Akanksha Shridharani, Manav Punjabi, Jainam Shah, and Ninad Mehendale. Diagnosis of covid-19 using ct scan images and deep learning techniques. *Emergency radiology*, 28(3):497–505, 2021.

[13] Catrin Sohrabi, Zaid Alsafi, Niamh O'neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). *International journal of surgery*, 76:71–76, 2020.

[14] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Ruixuan Wang, Huiying Zhao, Yunfei Zha, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

[15] Xing Wu, Cheng Chen, Mingyu Zhong, Jianjia Wang, and Jun Shi. Covid-al: The diagnosis of covid-19 with deep active learning. *Medical Image Analysis*, 68:101913, 2021.

[16] Ying Xiong, Dong Sun, Yao Liu, Yanqing Fan, Lingyun Zhao, Xiaoming Li, and Wenzhen Zhu. Clinical and high-resolution ct features of the covid-19 infection: comparison of the initial and follow-up changes. *Investigative radiology*, 2020.