ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414947

COVID-19 DIAGNOSTIC USING 3D DEEP TRANSFER LEARNING FOR CLASSIFICATION OF VOLUMETRIC COMPUTERISED TOMOGRAPHY CHEST SCANS

Shuohan Xue and Charith Abhayaratne

Department of Electronic and Electrical Engineering,
The University of Sheffield, Sheffield S1 3JD, United Kingdom.
e-mail: sxue6@sheffield.ac.uk and c.abhayaratne@sheffield.ac.uk

ABSTRACT

Deep learning-based algorithms provide an efficient and reliable diagnosis for medical imaging. This paper proposes COVID-19 diagnosis based on analysis of Computerised tomography (CT) chest scans. In recent years, deep learning-based analysis of CT chest scans has demonstrated competitive sensitivity for pneumonia prognosis. This paper presents our submission for the 2021 ICASSP Signal Processing Grand Challenge (SPGC). We exploit a 3D Network-based transfer learning approach to classify volumetric CT scans with a novel pre-processing method to render the volume with salient features. This work uses the pre-trained 3D ResNet50 as the backbone network. The 3D network is trained on a dataset consisting of 3 classes: Community Acquired Pneumonia (CAP), COVID-19 and Normal patient. The final testing results have shown an overall accuracy of 85.56% with the COVID-19 sensitivity attaining 82.86%.

***Index Terms*—** COVID-19, CT scans, Deep Learning, 3D CNN, Transfer Learning, Fully-Automated Classification

1. INTRODUCTION

The pandemic of COVID-19 has caused 122.7 million infections worldwide [1] to date. Considering the continuous impacts of the pandemic and the fact that this virus is extremely transmissible, a timely, fast and sensitive prognosis technique is of crucial importance, to help government authorities with the efficient allocation of medical resource and halting the transmission of the virus. The current gold standard diagnostic technique of COVID-19 is Reverse Transcription Polymerase Chain Reaction (RT-PCR). However, many recent studies claimed that RT-PCR is time-consuming and prone to a high false-negative rate [2] [3]. Additionally, processing a RT-PCR test is manual labour work and it is likely to cause a heavy workload to healthcare professionals. Hence, a fast and fully automated technique is needed as a complementary diagnostic approach to mitigate the pressure on the medical systems.

Recently, many studies have investigated the utilisation of deep learning-based Computed Tomography (CT) chest scans for COVID-19 detection and diagnosis. The method in [4] uses a decision fusion approach by combining predictions from 5 CNN architectures and yielded an overall accuracy of 86%. The study in [5] proposes a fully-automated framework named "COVID-FACT" that is capable of capturing spatial information of the CT images by using U-Net-based lung segmentation, achieving an accuracy of 90.82% for binary classification and achieving improved COVID-19 sensitivity. The work in [6] proposes a hybrid network architecture to achieve a trade-off between computational costs and spatial information and uses depth-wise convolution and spatial pyramid pooling layers to extract the spatial features. The method in [7] uses the pre-trained 3D ResNet-18 [8] as the backbone network and stacked the 2D CT images for the 3D input of the network and achieves an excellent overall accuracy.

Although the previous studies demonstrated a satisfactory accuracy in 2D-based or hybrid COVID-19 detection, few of them investigated using the whole lung volume as the 3D network input. This may be due to the bulkiness of the 3D volume. Inspired by the gap of most previous work, we employ a 3D Convolution Neural Network (CNN) transfer learning approach for COVID-19 screening using chest volumetric CT scans by adapting the 3D ResNet [9]. This paper presents our solution for the 2021 IEEE ICASSP Signal Processing Grand Challenge (SPGC). Noticeably, our algorithm does not require volumetric information by utilising a novel pre-processing method. Its input data is just patient-wise 3D volume.

2. METHODOLOGY

2.1. Dataset Description

As required by the 2021 IEEE ICASSP SPGC, the SPGC-COVID dataset, referred as "COVID-CT-MD" [10] is used for this study. The whole dataset comprises 3 classes of volumetric chest CT scans, which includes 171 patients diagnosed

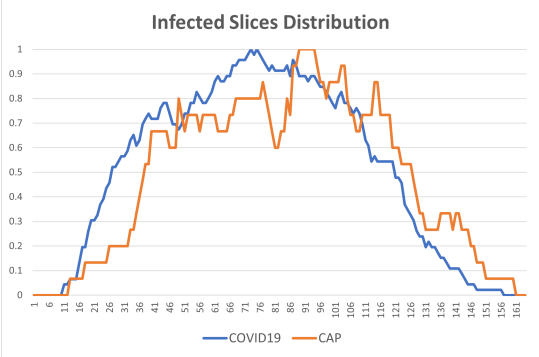


Fig. 1: Slice Distribution in Labelled Dataset

with COVID-19, 60 patients diagnosed with Community Acquired Pneumonia (CAP), and 76 normal patients. Patients with COVID-19 were collected from February 2020 to April 2020; whilst CAP and Normal cases were acquired from April 2018 to December 2019. For each patient, the dataset contains all slices of a volumetric chest CT scan in the Digital Imaging and Communications in Medicine (DICOM) format. The pixel information of the CT slices is provided in the Hounsfield Unit (HU) and the size of each slice image is 512×512. Note that the CT scans are obtained by multiple medical institutions, thus the scanning parameters may vary in slice thickness, pixel spacings, effective mAs and exposure time.

Besides the patient-level labels, slice-level labels are also provided with a subset of 55 COVID-19 and 25 CAP cases. Slice labels were identified by one radiologist to exhibit slices with infection and slices without infection. Fig.1 illustrates the distribution of infected slices for labelled cases. The x and y-axis respectively represent the index of slice and the proportion of infected slices on that specific index of the slice. This figure indicates the particular segment of the chest volume tends to have salient features, thus it is good to exploit these features and improve the training efficiency and model performance.

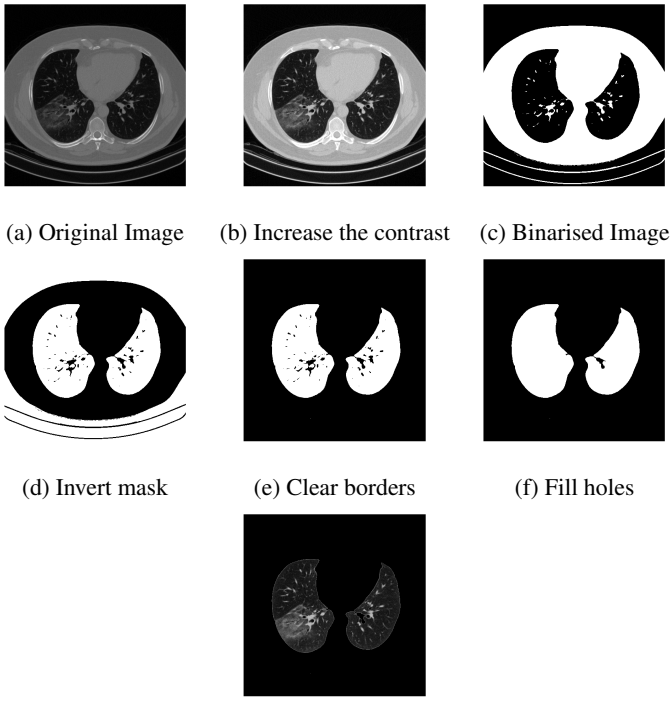
2.2. Data Preprocessing

2.2.1. Pixel Data Extraction

The pixel data of the chest volume per patient can be extracted by using the built-in DICOM read function in MATLAB, the pixel data of each chest volume are then converted to 3D grayscale ($pixel \in [0\ 1]$) volumetric images. Note that the dimensions of each volumetric lung is $512 \times 512 \times 1 \times N$ (N denotes the total number of slices).

2.2.2. 3D Image Resampling

The second step is to resample the volume by adjusting the pixel spacings and slice thickness. The resampling step aims



(a) Original Image (b) Increase the contrast (c) Binarised Image

(d) Invert mask (e) Clear borders (f) Fill holes

(g) The masked original image

Fig. 2: Pulmonary Segmentation Process

to create more slices per patient and better visualise the chest volume. In this dataset, the slice thickness is constantly $2mm$; the pixel spacing value can be obtained from the textual information of the DICOM file. The adjustment includes $pixel\ spacing \times 2.3$ and $slice\ thickness \times 0.5$. The filter of the 3D interpolation is 'cubic'. After the resampling process, the dimensions of the chest volume is $224 \times 224 \times 2N$ (N is doubled since the slice thickness is half; The number of dimensions is squeezed to 3).

2.2.3. Pulmonary Segmentation

Lung segmentation is an essential step for any CT scan-based pulmonary detection. In this study, automatic pulmonary segmentation is executed with successive procedures. Note that this segmentation is 2D-based and the purpose is to extract only the thorax region for each CT slice. Fig.2 illustrate the explicit process of the segmentation. Firstly, the contrast of the image is increased via saturating the bottom 1% and the top 1% of pixel values. Next, an experimentally selected threshold ($\geqslant 0.502$) is operated and then the image is binarised. The following operations of inverting, borders clearing and holes filling are executed sequentially, a clear mask of the lung region is generated as shown in Fig.2 (f). The final step is to multiply the lung region mask with the original image. Thus, the image of the Region of Interest (ROI) is extracted for the training input. Fig.2 (g) demonstrates the

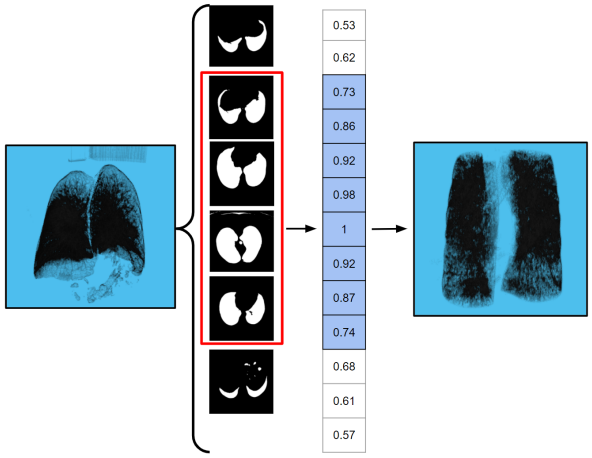


Fig. 3: Volume Range Selection Mechanism

output segmented lung image. After pulmonary segmentation, the 3D volume is constructed with the segmented lung images of each patient. Note that the segmentation does not change the dimensions of the volumes so the dimensions remain the same to the last step ($224 \times 224 \times 2N$).

2.2.4. Volume Range Selection

As stated, the input data type for this study is 3D patient-wise volume. Therefore, a slice range selection is operated before constructing the 3D volume with the segmented lung images. The aim of this step is to extract the slices with salient features for pathological identification so that the training algorithm can be more efficient without compromising much volumetric information. As previously shown in Fig.1, the slices with salient features are mainly located in the middle of the lung. Hence, by calculating the area of the lung regions using segmentation masks, the range of slices with salient features can be acquired. Fig.3 presents the slice selection mechanism. Firstly, the percentages of lung area of all slices in one volume are calculated and it returns a column of percentages. Next, the column of values is normalised and only the "$\geqslant 0.7$" slices are extracted for volume construction. It is noticeable that the normalisation method is by dividing the maximum value of the column so that even the areas of different patients vary, the range of salient slices can be extracted consistently. Finally, the volume with salient features is generated and resized into $[224 \times 224 \times 224]$ using 3D cubic interpolation to meet the required input size of the 3D ResNet. It is worth mentioning the chest volumes are saved in ".mat" format in 3 different folder based on its class, thus the 3D image can be read by MATLAB for deep learning input. Fig.4 demonstrates several examples of patient-wise volumes for input data(visualised by MATLAB Volume Viewer).

2.2.5. Data Augmentation

As the original dataset is an imbalanced dataset, classes of CAP cases and normal patients have insufficient samples

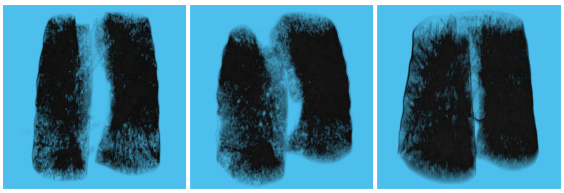


Fig. 4: 3D Volumes with Salient Features

Table 1: Dataset for Training (After Data Augmentation)

Dataset	CAP	COVID-19	Normal	Total
Training	82	116	65	263
Validation	19	55	24	98
Total	99	171	89	361

compared to the class of COVID-19 cases. Hence, we augmented the samples in the class of CAP and Normal patient respectively. The augmentation methods include two aspects: (1) Using different parameters for volume resampling, the pixel spacing is multiplied by 1.75 and the slice thickness is multiplied by 0.7; (2) Random image centre crop for slices in one volume, the crop window size is in a range of $[0.75\ 0.9]$ of the original image size. The image crop is 2D and slice-wise so it is independent of the slice selection step. Note that only the class of CAP and the class of Normal patient are augmented, Table.1 exhibits the details of the dataset for training the deep learning model after data augmentation.

2.3. 3D ResNet50 for Transfer Learning

The pre-trained 3D ResNet50 [9] is used as the backbone CNN for this challenge. The utilisation of 3D CNN is due to its capability of extracting more robust volumetric representations, while less annotated data is needed, in comparison to 2D and 2.5D networks [11] [12]. The architecture of 3D ResNet50 is generally similar to the 2D ResNet50, the main difference is that 3D modules are used in each essential section such as 3D convolution layers and 3D maxpooling layers. These 3D modules allow the network to learn the spatial information of the voxel.

In this study, we use the convolution base of the 3D ResNet to for feature extraction. The original fully connected layers are trimmed off and replaced by new ones, the output of the last fully connected layer is set to 3 in corresponding to this challenge. A Dropout layer [13] is added to mitigate the overfitting problem. Table.2 illustrates the details of the classifier.

2.4. Training Implementation

Table.3 presents the options for the training process. In this work, the categorical cross-entropy loss function is utilised. Moreover, we choose Adam optimiser for faster convergence. The initial learning rate is set to $2e-4$ and multiplied with

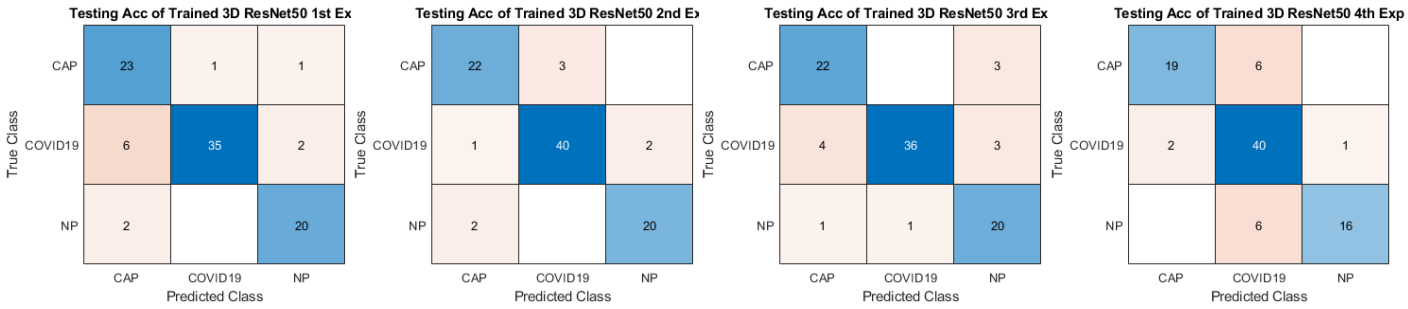


Fig. 5: Confusion Charts of 4-fold Cross Validation

Table 2: Classification Layer for Transfer Learning

Layer	Activations
GlobalAverage Pooling	$1 \times 1 \times 1 \times 512$
FullyConnected-1	$1 \times 1 \times 1 \times 512$
ClippedRelu	-
Dropout	0.5
Fully-Connected-2	$1 \times 1 \times 1 \times 3$
Softmax-Prob	$1 \times 1 \times 1 \times 3$
Classification	Output Label

Table 3: Training Options

Training options	Parameters
Loss Function	CrossEntropy Loss
Optimiser	Adam
Initial learning rate	$2e-4$
Lr descendant factor	0.5
Lr descendant rate	per 40 epochs
Total epochs	120

a factor of 0.5 per 40 epochs. The number of total epochs is 120. Since the testing set does not provide the categorical labels as it is a grand challenge, we employ a 4-fold cross validation approach to test the model performance. Meanwhile, for each one experiment, we split 25% of the training and validation set is randomly to test the trained model and generate a confusion chart illustrating the testing accuracy. Note that the model for the final submission is trained with the whole dataset. The implementation of the training is achieved by using MATLAB on a RTX 2070 GPU with 8GB of memory.

3. PERFORMANCE EVALUATION

In this work, we employ a straightforward training approach that the classification is based on patient-wise prediction. This section reports our initial evaluation of the proposed method. Since the labels of the test data set are not available, we report the 4-fold cross validation using the training test set. The accuracy rate, specificity and the sensitivity for each of the classes are reported in Table.4. Fig.5 illustrates the confusion tables of the 4-fold cross validation experiments. Results show an overall Accuracy of 86.94%, sensitivity of 87.79% and specificity of 89.88% for the COVID-19 class.

Table 4: Classification performance for 4-fold CrossValidation

Model	Accuracy	Sensitivity	Specificity
EXP 1	86.67%	81.39%	97.22%
EXP 2	91.11%	93.02%	93.02%
EXP 3	86.67%	83.72%	97.30%
EXP 4	83.33%	93.02%	76.92%
Summary	86.94%	87.79%	89.88%

It is possible that the model performance is mainly constrained by the limited and imbalanced dataset. Moreover, the slice-level labels are not involved in this deep learning algorithm as this approach has minimal requirements for annotated data. Furthermore, this study does not use a deep learning-based lung segmentation may also slightly affect the model performance. It is likely that the model performance can be improved with more samples available. Also, this approach does not modify the architecture of the 3D ResNet50 due to the complexity. The final testing results demonstrate that this model has achieved an overall 85.56% accuracy. The sensitivity for COVID-19, CAP and NP is 82.86%, 80.00% and 91.43% respectively.

4. CONCLUSIONS

In this paper, we have presented a novel data preprocessing approach and a 3D Network-based transfer learning algorithm to fully automatically detect and classify the chest volumetric CT scans, submitted as a solution for the 2021 ICASSP Signal Processing Grand Challenge. The model was trained with a limited dataset and a fairly good result has been achieved. An overall classification accuracy of 85.56% was achieved. The sensitivity for COVID-19, CAP and NP was 82.86%, 80.00% and 91.43% respectively. As there has not yet been many studies that investigate 3D networks for patient-wise COVID-19 prognosis, our solution has provided a promising approach. Moreover, the heavy computational costs for this algorithm can be mitigated via continuously optimising the algorithm in future works.

5. REFERENCES

[1] "Coronavirus disease (COVID-19)," Accessed on: Mar 20, 2021. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[2] A. T. Xiao, Y. X. Tong, and S. Zhang, "False negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence," *Journal of Medical Virology*, vol. 92, no. 10, pp. 1755–1756, Jul 2020.

[3] I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R. Del Campo, A. Ciapponi, O. Sued, L. Martinez-García, A. W. Rutjes, N. Low, P. M. Bossuyt, J. A. Perez-Molina, and J. Zamora, "False-negative results of initial RT-PCR assays for COVID-19: A systematic review," *PLOS ONE*, vol. 15, no. 12, pp. 1–19, Dec 2020.

[4] A. K. Mishra, S. K. Das, P. Roy, and S. Bandyopadhyay, "Identifying COVID19 from chest CT images: A deep convolutional neural networks based approach," *Journal of Healthcare Engineering*, vol. 2020, p. 8843664, Aug 2020.

[5] S. Heidarian, P. Afshar, N. Enshaei, F. Naderkhani, A. Oikonomou, S. F. Atashzar, F. B. Fard, K. Samimi, K. N. Plataniotis, A. Mohammadi, and M. J. Rafiee, "COVID-FACT: A fully-automated capsule network-based framework for identification of COVID-19 cases from chest CT scans," pp. 1–16, Sep 2020.

[6] B. Khaled, H. Fayçal, and M. Abdellatif, "Hybrid-COVID: a novel hybrid 2D/3D CNN based on cross-domain adaptation approach for COVID-19 screening from chest X-ray images," vol. 43, no. 4, pp. 1415–1431, Dec 2020.

[7] Z. Li, Z. Zhong, Y. Li, T. Zhang, L. Gao, D. Jin, Y. Sun, X. Ye, L. Yu, Z. Hu, J. Xiao, L. Huang, and Y. Tang, "From community-acquired pneumonia to COVID-19: a deep learning–based method for quantitative analysis of COVID-19 on thick-section CT scans," *European radiology*, vol. 30, no. 12, pp. 6828–6837, Jul 2020.

[8] K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3D residual networks for action recognition," *CoRR*, vol. abs/1708.07632, Aug 2017.

[9] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer learning to 3D resnet-18 for Alzheimer's disease detection on MRI images," in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, Nov 2020, pp. 1–6.

[10] P. Afshar, S. Heidarian, N. Enshaei, F. Naderkhani, M. J. Rafiee, A. Oikonomou, F. B. Fard, K. Samimi, K. N. Plataniotis, and A. Mohammadi, "COVID-CT-MD: COVID-19 computed tomography (CT) scan dataset applicable in machine learning and deep learning," pp. 1–9, Sep 2020.

[11] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D deep learning on medical images: A review," Oct 2020.

[12] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, pp. 582 – 596, May 2019.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, p. 1929–1958, Jan. 2014.