Automatic Diagnosis of COVID-19 Using a tailored Transformer-Like Network

Chengeng Liu\textsuperscript{1,a}, Qingshan Yin\textsuperscript{2,b}\textsuperscript{*}

\textsuperscript{1}Department of Software and Computing Systems, the University of Melbourne, Melbourne, Victoria, Australia
\textsuperscript{2}Shandong New Generation Information Industry Technology Research Institute, Inspur Group, Jinan, Shandong, China
\textsuperscript{a}email: chengengl@student.unimelb.edu.au, \textsuperscript{b}email: yingsh@inspur.com

Abstract. The emergence of the novel coronavirus(COVID-19) has left disastrous effect on global health and individuals. Even though in most areas, the RT-PCR test used as the dominant approach for diagnosis of COVID-19 has shown good accuracy, the test requires equipment, personnel and it is time-consuming. Researches have shown the effectiveness of X-ray images for predicting COVID-19. In this study, we applied a transformer-like deep-learning model on this problem with transfer learning technique, to diagnose X-ray images as COVID-19 or normal. The model outperformed the CNN SOTA. The model achieved a classification accuracy of 99.7% on the targeting dataset.

1. Introduction
A novel coronavirus disease(COVID-19) started in late December, 2019, quickly spread out around the world\cite{1,2}. By 20 July 2021, there were 190,169,833 confirmed cases and 4,086,000 dead cases\cite{3}. The fatality rate of COVID-19 is 2%-3\%\cite{4} and its contiguous nature and rapid spreading made it a global health crisis of the time. There were many challenges in some countries’ public health systems, such as limited medical resources and personnel, which may further increase the infection rate of the healthcare workers\cite{5}. Considering its rapid spreading nature, the early screening of COVID-19 is becoming important to avoid further spread.

Real-time polymerase chain reaction(RT-PCR) is the current solution for the diagnosis of COVID-19\cite{6}. However, the method is laborious, time-consuming and requires expert medical workers\cite{7}, it may not be feasible in some countries where medical resources are limited. Some researchers have used CT-scan images for diagnosis of COVID-19\cite{32,33}. While CT-scan can provide more features and better performance than X-ray images in detecting COVID-19, CT-scan is generally more costly than X-ray images.

Researchers have shown that lungs are one of the first affected organs\cite{8}. Deep learning with CNN has been long used in disease detection such as cancer\cite{9,10,11}, brain tumor and heart failure\cite{14,15}. The emergence of various datasets provides the possibility of employing deep learning in diagnosis of COVID-19. Recent studies\cite{12,13} showed the effectiveness of applying deep learning in assisting the diagnosis of COVID-19.

1.1 Related Work
Apostolopoulos and Mpesiana\cite{16} conducted research evaluating five classical deep learning models, including VGG19, MobileNet v2, Inception, Xception and Inception ResNet. The CNN models are
trained and evaluated on two datasets. The first dataset is Cohen’s covid-chestxray-dataset[[17]], which consists of 224 COVID-19 X-ray images, 700 common pneumonia images and 504 normal X-ray images. The second dataset[[18]] contains 224 COVID-19 X-ray images, 714 common pneumonia images and 504 normal images. MobileNet v2 achieved the highest accuracy of 96.78% on Top2 score and 94.72% on Top3 score. However, the study[[16]] suffers from the limited sample size of the dataset.

A study conducted by Hamdan et al.[[19]] proposed COVIDX network, which contains comparative analysis of seven deep learning models. The models were VGG19, DenseNet201, ResNetV2, InceptionV3, InceptionResNetV2, Xception and MobileNetV2. The models were trained and evaluated on Cohen’s dataset(early version), with 25 COVID-19 X-ray images and 25 normal images[[17]]. The study showed that VGG19 and DenseNet have the best accuracy of 90%. The analysis is limited due to its small data size. Similarly, Ahammed et al.[[23]] proposed a CNN network with accuracy of 94.03%. The size of the dataset is limited, with 285 images each for the three classes. Another research done on Cohen’s dataset[[17]] conducted by Kumar et al.[[25]] used ResNet50 and Support Vector Machine(SVM) for image classification. The study used two datasets: Cohen’s dataset[[17]] and Chest X-ray images(Pneumonia)[[26]]. The highest accuracy achieved by the study was 95.38%, which outperformed Hamdan et al.[[19]] on the same dataset. Similarly, Narin et al.[[27]] trained and evaluated on Cohen’s dataset[[17]] and Chest X-ray images(Pneumonia)[[27]]. The study achieved an accuracy of 96.1%.

Wang et al.[[20]] proposed a COVID-net model and compared the model with VGG19 and ResNet50. The model was pre-trained on ImageNet dataset and it achieved the accuracy of 93.3% on the three-class classification task. A combined dataset were created from multiple sources, with total 13974 X-ray images[[17],[19],[21],[31]]. The dataset suffers from imbalanced distribution, though Wang et al.[[20]] used data augmentation to reduce the impact of imbalanced dataset. Similar to Wang et al.[[20]], Apostolopoulos and Mpesiana[[16]] conducted comparative analysis on the combination of multiple datasets[[17], [21], [22]]. The total number of images used was 1427, including 224 COVID-19, 700 pneumonia and 504 normal images. The research showed that VGG19 outperformed other models with an accuracy of 98.75%. Another research conducted by Khan and Aslam[[28]] overcome the issue of imbalanced dataset by selecting images across different datasets to make data distribution evenly. The dataset contains 630 COVID-19 images and 642 Normal images. They performed comparative analysis using ResNet50, VGG16, VGG19 and DenseNet121, where VGG outperforms other networks and achieved an remarkable accuracy of 99.38%, though the size of the dataset is limited.

In the above related work, most researches used CNN network to perform either two-class or three-class image classification. Most datasets used were at limited sizes. Very few researches used large dataset. Wang et al.[[20]] used a dataset containing 13974 X-ray images but the distribution of the dataset was imbalanced. Moreover, most evaluations were done on classical CNN networks, but not evaluating transformer networks such as Vision Transformers. Recently visual recognition has been revolutionized by Vision Transformers(ViT) [[29]]. ViT has outperformed CNN networks on ImageNet classification task with extra training data.

In this study, we have proposed a fine-tuned deep learning network tailored for diagnosis COVID-19, which is based on Transformer-like network, named VOLO[[30]]. VOLO declared SOTA on ImageNet classification task without extra training data. The novelty and originality of proposed study is as follows:

1. The study does not suffer from imbalanced data distribution.
2. The network is fed with a relatively large number of X-ray images, compared with other related work.
3. One of the first studies on applying VOLO for specific image classification task and verifying the generality of VOLO.
4. The network outperforms the benchmark analysis on the targeting datasets.
2. Materials and Methods

2.1 Dataset
There are two datasets used in the study. Both are available on Internet. The first dataset originally
contained 219 COVID-19 X-ray images, 1341 Normal X-ray images and 1345 X-ray viral pneumonia
images\cite{31}. As mentioned in Ranman et al.\cite{33}, there are three researches conducted on the
dataset\cite{31-33}. Chowdhury et al.\cite{31} achieved the highest accuracy of 98.3%.

The dataset after updates now contains 3616 COVID-19 X-ray images, 10192 normal X-ray
images\cite{31}. To balance the dataset, in this study we randomly shuffle normal images and pick up
3600 normal images from it. So the data distribution of the first dataset(namely Dataset-1) is as Figure
1.

Dataset-1 is further divided into three parts: 2900 images for each class as training data, 300
images for validation and 400 images for testing. Data augmentation was applied on training dataset.
The size of training data of Dataset-1 after data augmentation is as follows:

Figure 1. Data distribution of Dataset-1

The second dataset used in the study, namely Dataset-2, was based from Cohen’s dataset\cite{17},
which contains 340 COVID-19 images and 340 normal images(Figure 4). Dataset-2 was used as
testing only, to prove that VOLO has generality given that it is not trained on Dataset-2. Some early
studies used Cohen’s vanilla dataset\cite{19,25], with 25 COVID-19 images. Ioannis et al.\cite{16}
conducted analysis on Cohen’s dataset with 224 COVID-19 images and 504 normal images.

2.2 Deep neural networks and Transfer Learning
Soon after the publication of ResNet\cite{34} in 2015, deep neural networks on image classification task
made remarkable progress. Recently, the emergence of Vision Transformers(ViT) challenged the
dominant position of CNN models. Until July 2021, ViT-G declared SOTA on ImageNet classification
task, with Top1 accuracy of 90.45%\cite{35]. However, both CNN-based
models and transformer-based models require large training dataset, and CNN models generally
outperform transformers without extra training data\cite{30}. There is a high correlation between the size
of the training data and the performance of the deep neural network. Deep neural networks can be
applied to a task where limited size of data is provided, by utilizing the technique of transfer
learning\cite{16,24]. The idea of transfer learning is to transfer the knowledge learning from previous
tasks to a target task where the latter lacks of high-quality data. In this study, we used pre-trained
models of VOLO, which was pre-trained on ImageNet-1K\cite{30}. The structure of the model is tailored
at the final stage and fine-tuned to the binary class classification task.
2.3 VOLO model
Vision Outlooker (VOLO) was developed based on transformer. VOLO aims to resolve the bottlenecks that restricted the performance of ViT by applying a new attention mechanism named outlooker attention [30]. VOLO declares SOTA on ImageNet classification with Top1 accuracy of 87.1% (no extra training data) on Jun 2021 and it is the first model that exceeds 87% without extra training data. In this study, we use VOLO-D3 because D3 outperforms D4 and D5 on the targeting dataset. Yuan et al. [30] also stated that D5 and D4 with parameters exceeding 100M may lead to overfitting on small dataset.

![Figure 2. VOLO outlooker attention structure][30]

2.4 Model Hyper-parameters and Set-up
All images are scaled to size of $224 \times 224$ pixels before loaded into the model. We used a cosine learning rate scheduler, with initial learning rate of $1.6 \times 10^{-7}$, decay epoch of 30, decay rate of 0.1. Total number of epoch is 300. The optimizer used is AdamW. The training was performed on GPU Tesla V100 * 8 parallelly with memory 32GB each.

3. Results & Discussion
Figure 4 shows the training process of VOLO-D3 on Dataset-1, which consists of 11.6k images for each class after data augmentation. The best Top1 accuracy achieved is 99.67%. According to Rahman et al. [33], this has outperformed the SOTA on the dataset, which was 98.3% [31].

![Figure 3. Top1 accuracy during training](image)
![Figure 4. Evaluation loss during training](image)

To further investigate and verify the generality of VOLO, we applied the trained model from Dataset-1 to be tested on Dataset-2. The model with the best test score on Dataset-1 after cross-
validation was selected to be tested on Dataset-2 and no extra training data from Dataset-2 was applied on the model. Top1 accuracy of the model on Dataset-2 is 98.98%, which outperforms Apostolopoulos[[16]] 98.75%. A more detailed comparison is shown in Table 1:

| Studies                           | Model             | Top 1 Accuracy |
|----------------------------------|-------------------|----------------|
| Apostolopoulos et al.[[16]]      | VGG-19            | 98.8%          |
| Narin[[27]]                      | ResNet50/RestNet101| 96.1%          |
| Lv et al. [[36]]                 | DenseNet169       | 97.1%          |
| Punn et al. [[37]]               | NASNet            | 98.0%          |
| **Ours**                         | VOLO-tailored     | **99.0%**      |

Table 2. Performances of models in general

| Studies                           | Model             | Top 1 Accuracy |
|----------------------------------|-------------------|----------------|
| Hemdan et al. [[19]]             | COVIDX-Net        | 90.0%          |
| Wang et al. [[20]]               | Tailored CNN      | 93.3%          |
| Kumar et al. [[25]]              | ResNet + SVM      | 95.4%          |
| Minaee et al. [[24]]             | Deep-COVID        | (Sensitivity 97.5%) |
| Oh et al.[[39]]                  | ResNet18          | 88.9%          |
| Farooq et al.[[38]]              | COVID-ResNet      | 96.2%          |
| **Ours**                         | **VOLO-tailored** | **99.7%**      |

The proposed VOLO-tailored model outperforms some studies in the benchmark. Most studies had limited data and suffered from imbalanced data distribution.

4. Conclusions
In this study we trained and validated a new proposed model VOLO tailored with transfer learning technique using a larger X-ray dataset, compared with the related work. Early studies on detecting COVID-19 generally suffered from insufficient data and imbalanced data distribution. VOLO outperforms SOTA on the targeting datasets and an accuracy of 99.67% was achieved. The generality of VOLO was verified across datasets. In light of the study, it is believed that the model can assist medical workers during the diagnosis of COVID-19 with the support from more valid data.

Acknowledgements
Thanks Shandong New Generation Information Industry Technology Research Institute(INGIIT) for providing GPU services for training and testing.

References
[1] Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The lancet*, 395(10223), 470-473.
[2] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The lancet*, 395(10223), 507-513.
[3] World Health Organization. 2021,Weekly Operational Update on COVID-19: https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19---20-july-2021
[4] Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., ... & Zhong, N. S. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine*, 382(18), 1708-1720.
[5] Ng, K., Poon, B. H., Kiat Puar, T. H., Shan Quah, J. L., Loh, W. J., Wong, Y. J., ... & Raghuram, J. (2020). COVID-19 and the risk to health care workers: a case report. Annals of internal medicine, 172(11), 766-767.

[6] Tahamtan, A., Ardebili, A., 2020. Real-time rt-pcr in COVID-19 detection: issues affecting the results. Expert review of molecular diagnostics 20, 453– 454.

[7] Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in different types of clinical specimens. Jama, 323(18), 1843-1844.

[8] Li, M. (2020). Chest CT features and their role in COVID-19. Radiology of Infectious Diseases, 7(2), 51-54.

[9] Li, Y., & Shen, L. (2018). Skin lesion analysis towards melanoma detection using deep learning network. Sensors, 18(2), 556.

[10] Liao, Q., Ding, Y., Jiang, Z. L., Wang, X., Zhang, C., & Zhang, Q. (2019). Multi-task deep convolutional neural network for cancer diagnosis. Neurocomputing, 348, 66-73.

[11] Yoo, S., Gujrathi, I., Haider, M. A., & Khalvati, F. (2019). Prostate cancer detection using deep convolutional neural networks. Scientific reports, 9(1), 1-10.

[12] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology. 2020;296(2):200905.

[13] Aridakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Computers in biology and medicine, 121, 103795.

[14] Podnar, S., Kukar, M., Gunčar, G., Notar, M., Gošnjak, N., & Notar, M. (2019). Diagnosing brain tumours by routine blood tests using machine learning. Scientific reports, 9(1), 1-7.

[15] Gjoreski, M., Gradišek, A., Budna, B., Gams, M., & Poglajen, G. (2020). Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. IEEE Access, 8, 20313-20324.

[16] Apostolopoulos, I. D., & Mpiesiana, T. A. (2020). Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine, 43(2), 635-640.

[17] Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988.

[18] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122-1131.

[19] Hemdan, E. E. D., Shouman, M. A., & Karar, M. E. (2020). Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055.

[20] Wang, L., Lin, Z. Q., & Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports, 10(1), 1-12.

[21] Radiological Society of North America, 2018, RSNA Pneumonia Detection Challenge. Available online: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data (accessed on 16 June 2020).

[22] Italian Society of Medical and Interventional Radiology (SIRM), 2020. Available online: https://www.sirm.org/en/italian-society-of-medical-and-interventional-radiology/

[23] Ahammed, K., Satu, M. S., Abedin, M. Z., Rahaman, M. A., & Islam, S. M. S. (2020). Early Detection of Coronavirus Cases Using Chest X-ray Images Employing Machine Learning and Deep Learning Approaches. medRxiv, 2020-06.

[24] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. Medical image analysis, 65, 101794.
[25] Kumar, P.; Kumari, S. Detection of coronavirus Disease (COVID-19) based on Deep Features. Preprints 2020.
[26] Paul Mooney, 2018, Chest X-ray Images (Pneumonia). Available online: https://www.kaggle.com/paulimothymooney/chest-xray-pneumonia (accessed on 16 June 2020).
[27] Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications, 1-14.
[28] Khan, I. U., & Aslam, N. (2020). A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images. Information, 11(9), 419.
[29] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
[30] Yuan, L., Hou, Q., Jiang, Z., Feng, J., & Yan, S. (2021). VOLO: Vision Outlooker for Visual Recognition. arXiv preprint arXiv:2106.13112.
[31] Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., ... & Islam, M. T. (2020). Can AI help in screening viral and COVID-19 pneumonia?. IEEE Access, 8, 132665-132676.
[32] Bassi, P. R., & Attux, R. (2021). A deep convolutional neural network for COVID-19 detection using chest X-rays. Research on Biomedical Engineering, 1-10.
[33] Rahman, S., Sarker, S., Al Miraj, M. A., Nihal, R. A., Haque, A. N., & Al Noman, A. (2021). Deep Learning–Driven Automated Detection of COVID-19 from Radiography Images: a Comparative Analysis. Cognitive Computation, 1-30.
[34] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[35] Paper With Code, 2021, Image Classification on ImageNet, https://paperswithcode.com/sota/image-classification-on-imagenet
[36] Lv, D., Qi, W., Li, Y., Sun, L., & Wang, Y. (2020). A cascade network for detecting covid-19 using chest x-rays. arXiv preprint arXiv:2005.01468.
[37] Punn, N. S., & Agarwal, S. (2021). Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. Applied Intelligence, 51(5), 2689-2702.
[38] Farooq, M., & Hafeez, A. (2020). Covid-resnet: A deep learning framework for screening of covid19 from radiographs. arXiv preprint arXiv:2003.14395.
[39] Oh, Y., Park, S., & Ye, J. C. (2020). Deep learning covid-19 features on cxr using limited training data sets. IEEE transactions on medical imaging, 39(8), 2688-2700.