COVID-19 Detection in Chest X-ray Images Using Swin-Transformer and Transformer in Transformer

1st Juntao Jiang

Hangzhou, China

jj2910@nyu.edu

2nd Shuyi Lin

Khoury College of Computer Sciences

Northeastern University

Boston, United States

lin.shuyi@northeastern.edu

Abstract—The Coronavirus Disease 2019 (COVID-19) has spread globally and caused serious damages. Chest X-ray images are widely used for COVID-19 diagnosis and Artificial Intelligence method can assist to increase the efficiency and accuracy. In the Challenge of Chest XR COVID-19 detection in Ethics and Explainability for Responsible Data Science (EE-RDS) conference 2021, we proposed a method which combined Swin Transformer and Transformer in Transformer to classify chest X-ray images as three classes: COVID-19, Pneumonia and Normal (healthy) and achieved 0.9475 accuracy on test set.

Index Terms—COVID-19, Chest X-ray Images, Swin-Transformer, Transformer in Transformer, Model Ensembling, Image Classification

I. INTRODUCTION

The Coronavirus Disease 2019 (COVID-19), which caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2, 2019-nCoV) has become a global pandemic and brought unprecedented damages seriously worldwide.As Chest X-ray tests are typically has a high sensitivity diagnosis of COVID-19 [1], [2], Chest X-ray images can be used for not only following up the effects of COVID-19 on lung tissue, but also early detection of COVID-19, thus the immediate isolation and treatment for the suspected can be achieved.

Past years witnessed the growing use of AI techniques [3]–[6], especially deep-learning-based methods, in disease detection on chest X-ray images, successfully increasing the accuracy and efficiency for early diagnosis. After the outbreak of pandemic, many approaches have also been proposed to detect COVID-19. Shervin Minaee et al. [7] trained 4 state-of-the-art convolutional networks for COVID-19 detection on a dataset of around 5000 X-ray images, and achieved higher than 90 percent of sensitivity and specificity rate. Asif IqbalKhan et al. [8] designed a deep convolutional neural network model based on Xception architecture to classify Normal, Pneumonia-bacterial, Pneumonia-viral and Covid-19 chest X-ray images. Linda Wang et al. [9] introduced a deep convolutional neural network called COVID-Net to detect COVID-19 cases from chest X-ray images, which is open source and available to the general public. Rachna Jain et al. [10] used Inception net V3, XCeption net and ResNeXt to do the classification. Sanhita Basu et al. [11] gave a new concept called domain extension transfer learning (DETL) with pre-trained deep convolutional neural network on a related large
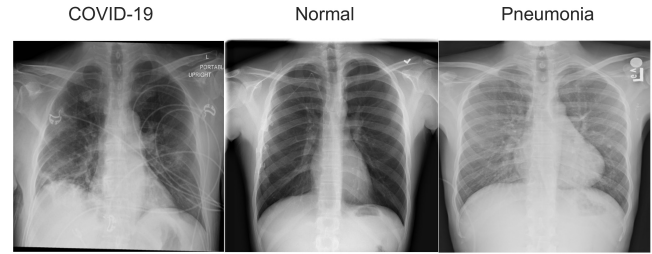

Fig. 1: Example Chest X-ray Images for classes of COVID-19, Normal and Pneumonia

chest X-Ray dataset. We can conclude that the fundamental idea for most current methods are based on multi convolutional neural network and transfer learning from large dataset.

Special attention should be paid to the truth that Transformer methods have outperform convolutional neural network in some datasets of different scenes recently. Swin Transformer [12] and Transformer in Transformer [13] are successful works to adapt Transformer from language to vision and achieved the state of the art in different tasks, but we haven't seen a lot of application in chest X-ray images classification and COVID-19 detection. This paper is a technical report in Chest XR COVID-19 Detection Challenge [14], a part of Ethics and Explainability for Responsible Data Science (EE-RDS) conference. We combined Swin-transformer and Transformer in Transformer to classify chest X-ray images as three classes as COVID-19, Pneumonia and Normal (healthy) on the dataset offered by the challenge organizer and achieved 0.9475 accuracy on test set.

II. INTRODUCTION TO THE DATASET AND THE TASK

The dataset of this challenge contains three parts, which are used to train, validate and test respectively:

- Train = 17,955 chest X-ray images
- validation = 3,430 chest X-ray images
- test = 1,200 chest X-ray images

There are three types of chest X-ray images in this dataset: COVID-19, Pneumonia and Normal (healthy). The example of three classes are shown as Figure 1.

Our task is to design an algorithm to auto classify chest X-ray images into these three classes.
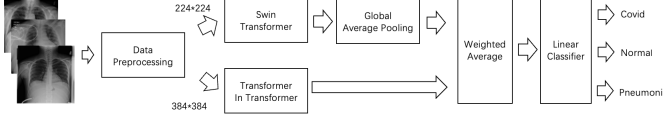
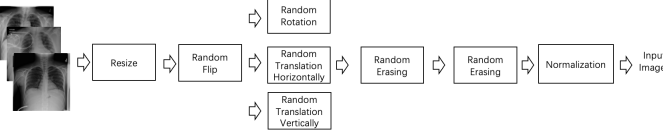Fig. 2: Workflow for Chest X-ray Images Classification with Swin Transformer and Transformer in Transformer

Fig. 3: Data Preprossing in Training

III. METHODS

The workflow of our classification method is shown as the Figure 2. We trained the Swin Transformer and Transformer in Transformer separately and did results ensemble by using weighted average methods.

A. Data Preprossing

The workflow of our Preprossing method for training is shown as the figure 3 and the workflow of our Preprossing method for training is shown as the figure 4.

Firstly we resized the images to a certain scale (224*224 or 384*384), then applied random flipping, then selected a policy from from rotation, translation horizontally or vertically, and then we applied random erasing. Finally we did the normalization. All these changes are applied with small probability.

Different from the natural images, the overall characteristics of medical images are significant to classify, so we avoid using crop or center crop as data augmentation methods. We only use horizontal flipping, and the rotation or translation augmentation methods are also controlled within a small range.

We did not use prepossessing methods related to brightness or contrast to avoid destroying the difference between the three classes.

B. Models

1) Swin Transformer: Swin Transformer [12] constructs hierarchical feature maps to better model objects of different
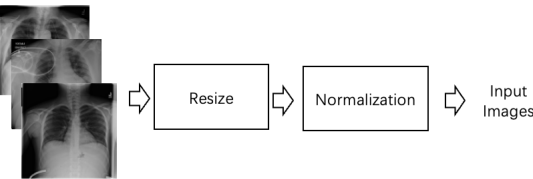

Fig. 4: Data Preprossing in Testing

TABLE I: Results for Image Classification

Models	Weights	Accuracy
Swin Transformer	/	0.9467
Swin Transformer, TNT	1:1	0.9467
Swin Transformer, TNT	2:1	0.9475

sizes and has linear computational complexity to image size. Swin Transformer has different types according to the size of the models and we selected Swin-B(ase) for the task.

2) Transformer in Transformer: Transformer in Transformer (TNT) [13], which regards the local patches as "visual sentences" and then present to further divide them into smaller patches as "visual words", and features of both words and sentences will be aggregated to improve the representation ability.

Transformer in Transformer has different types according to the size of the models and we selected the Transformer in Transformer small for the task. The larger models should have better performance but we failed to try them due to the limitation of time.

C. Training Setting

Training settings. Our implementation is based on PyTorch [15] and mmclassification [16], and one NVIDIA Tesla-V100 GPUs are used for training. Model training is done by using the AdamW, where initial learning rate, batch size and weight decay set to 0.001, 64 and 0.05 respectively respectively. For learning rate scheduler, cosine annealing [17] and warmup [18] is used. We also applied label smoothing in loss.

In the training for Swin Transformer, the input image size is 224*224 and for Transformer in Transformer, the input image size is 384*384. The design is due to the limitation of computation power, and also for convenience to use the pre-trained model directly from ImageNet [19].

D. Ensemble

We obtained the probabilities for each classes from two models, then calculated the weighted average for these probabilities. Finally, a linear classifier was used to get the results for classification.

IV. EXPERIMENTS AND RESULTS

We trained the two models on training set and validate on validation set. Then we selected the best model on validation set to inference on test set.

Due to the limitation of time, the experiments are not enough for comparison but may still give some information. The results are shown as Table 1.

More specifically, our final results are:

- Accuracy Score: 0.9475,
- Sensitivity Score: 0.9475
- Specificity Score: 0.9509

Our results ranks 10 at the leaderboard for this challenge.

V. Conclusion

In this paper, we applied Swin Transformer and Transformer in Transformer to classify the chest X-ray images in In the Challenge of Chest XR COVID-19 detection and did model ensemble by using weighted average method. We achieved 0.9475 accuracy on test set and ranks 10 on the leaderboard.

References

[1] Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, Chuansheng Zheng,Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study, The Lancet Infectious Diseases, 2020, Pages 425-434

[2] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology (2020), 10.1148/radiol.2020200642

[3] Y. Dong, Y. Pan, J. Zhang, W. Xu, Learning to read chest X-ray images from 16000+ examples using CNN Proceedings, 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, IEEE (2017), pp. 51-57

[4] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, ..., B.N. Patel, Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, PLoS Medicine, 15 (11) (2018)

[5] V. Chouhan, S.K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, ..., V.H.C. de Albuquerque, A novel transfer learning based approach for pneumonia detection in chest X-ray images, Applied Sciences, 10 (2) (2020)

[6] C. Liu, Y. Cao, M. Alcantara, B. Liu, M. Brunette, J. Peinado, W. Curioso, TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network 2017 IEEE international conference on image processing (ICIP), IEEE (2017), pp. 2314-2318.

[7] Shervin Minaee, Rahele Kafieh, Milan Sonka, Shakib Yazdani, Ghazaleh Jamalipour Soufi, Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning, Medical Image Analysis, Volume 65, 2020

[8] Asif Iqbal Khan, Junaid Latief Shah, Mohammad Mudasir Bhat, Coro-Net: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, Computer Methods and Programs in Biomedicine, Volume 196,2020

[9] Wang, L., Lin, Z.Q.,Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 10, 19549 (2020).

[10] Jain, R., Gupta, M., Taneja, S. et al, Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl Intell 51, 1690–1700 (2021).

[11] S. Basu, S. Mitra and N. Saha, Deep Learning for Screening COVID-19 using Chest X-Ray Images,2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 2521-2527.

[12] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, arXiv:2103.14030

[13] Kai Han, An Xiao, Enhua Wu, Jianyuan Guo, Chunjing Xu, Yunhe Wang, Transformer in Transformer, arXiv:2103.00112

[14] Akhloufi, Moulay A. and Chetoui, Mohamed, Chest XR COVID-19 detection,https://cxr-covid19.grand-challenge.org/,August,2021

[15] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, Soumith Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, NeurIPS 2019

[16] MMClassification Contributors,OpenMMLab's Image Classification Toolbox and Benchmark, https://github.com/open-mmlab/mmclassification, 2020

[17] I. Loshchilov and F. Hutter. SGDR: stochastic gradient descent with warm restarts. ICLR, 2017.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, arXiv:1512.03385

[19] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009