13th International Conference on Advanced Computational Intelligence (ICACI)
Wanzhou, China, May 14-16, 2021

2021 13th International Conference on Advanced Computational Intelligence (ICACI) | 978-1-6654-1254-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACI52617.2021.9435874

COVID-19 Patients Detection in Chest X-ray Images via MCFF-Net

Wei Wang
School of Computer and Communication Engineering
Changsha University of Science and Technology
Changsha 410114, China
wangwei@csust.edu.cn

Yutao Li
School of Computer and Communication Engineering
Changsha University of Science and Technology
Changsha 410114, China
704545327@qq.com

Xin Wang*
School of Computer and Communication Engineering
Changsha University of Science and Technology
Changsha 410114, China
wangxin@csust.edu.cn

Ji Li
School of Computer and Communication Engineering
Changsha University of Science and Technology
Changsha 410114, China
hangliji@163.com

Peng Zhang
School of Electronics and Communications Engineering
Sun Yat-sen University
Shenzhen 518107, China
zhangpeng5@mail.sysu.edu.cn

***Abstract*—COVID-19 is a respiratory disease caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2). This paper proposes a deep learning model to assist medical imaging physicians in diagnosing COVID-19 cases. We designed the Parallel Channel Attention Feature Fusion Module (PCAF), and brand new structure of convolutional neural network MCFF-Net was put forward. The experimental results show that the overall accuracy of MCFF-Net66-Conv1-GAP model is 96.79% for 3-class classification. Simultaneously, the precision, recall, specificity and the sensitivity for COVID-19 are both 100%. Compared with the latest state-of-art methods, the experimental results of our proposed method indicate its uniqueness.**

***Keywords*—COVID-19, Deep Learning, MCFF-Net, Convolutional Neural Network, CXR images**

I. Introduction

Novel coronavirus pneumonia (hereafter referred as COVID-19) is a respiratory disease caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2). Currently, the most commonly used COVID-19 diagnostic technology is real-time reverse transcription polymerase chain reaction (RT-PCR) technology. Since the RT-PCR detection process takes several hours, and the sensitivity is as low as 60%-70%, leading to a high false-negative rate in the test results [1], so there are limitations in the diagnosis process of COVID-19. COVID-19 has obvious clinical imaging characteristics. Compared with CT, Chest X-Ray (CXR) has the advantages of convenient detection process, low cost, and low ionizing radiation intensity [2]. It causes less harm to patients and is easy to carry out large-scale promotion in remote and underdeveloped areas.

In recent years, deep learning has become one of the popular research fields in artificial intelligence. W Wang. et al. [3] improved a deep learning method for detecting colonoscopy polyp images, and implemented excellent effect. W Wang. et al [4] introduced the dense connection idea of the DenseNet model in the MobileNet model, and proposed a new type of image classification model Dense-MoblieNet. On the basis of the original model, the accuracy of the image classification task is improved and the complexity of the model is reduced. W Wang. et al [5] combined the dense connection idea with the full convolutional network FCN, proposed a dense full convolutional network DFCN, and used this model to perform semantic segmentation tasks on the Chenzhou remote sensing image dataset, and achieved good results. After the outbreak of COVID-19, the use of DCNN to detect COVID-19 has become a recognized research direction. Meanwhile, many outstanding research results have emerged in this direction. According to the characteristics of CXR images, W Wang. et al [6] devised a module named channel feature weight extraction (CFWE), and the network CFW-Net based on this module was proposed, which realized good classification effect. Based on the VGG19[7] network model, Ioannis et al. [8] performed a 3-class classification experiment on a dataset containing COVID-19 positive, COVID-19 negative, and normal CXR images, and the classification accuracy reached 93.48%. In another work, Wang and Wong et al. [9] proposed a COVID-Net model based on the PEPX structure, and introduced Depthwise separable convolution [10] into the structure, with a classification accuracy of 93.3%. While reducing the amount of model parameter, a good classification accuracy is obtained. Ozturk et al. [11] proposed the DarkCovid-Net network model in their research work. The model is improved on the basis of the Dark-Net19 model. The accuracy of 2-class classification is 98.08%, and the accuracy of 3-class classification is 87.02%. Recently, Khan et al. [12] raised the CoroNet network in view of Xception module [13], and conducted 2-class, 3-class, and 4-class classification experiments for CXR images. The classification accuracy were 99%, 95% and 89.6% respectively.

Unlike ordinary image classification tasks, CXR images have the characteristics of high inter-category similarity and low intra-category variability. This kind of data characteristics is easy to cause problems like model deviation and over-fitting, which may lead to a decline in the generalization performance

*Corresponding author
This research was funded Changsha Science and Technology Project (kq2004071), The Hunan Graduate Student Innovation Project (CX20200882).

of the network and increase the difficulty of model recognition in image classification tasks. To solve this problem, this paper designs a parallel channel attention feature fusion module (Parallel channel attention feature fusion module, PCAF). Based on the PCAF module, a new convolutional neural network structure MCFF-Net is proposed.

II. PCAF-MODULE

In order to relieve the pressure of current medical staff and improve the diagnosis speed of COVID-19, we adopt a convolutional neural network that can adaptively learn feature information to identify and classify CXR images. CXR images have high inter-category similarity and low intra-category variability. These problems will lead to model deviation and over-fitting as well as reduce the recognition ability and generalization performance of the model. Hence, we designed a PCAF module, whose structure is shown in Fig. 1. C is the number of channels related to the input feature map. H and W represent the height and width of the feature map. "r" represents the channel compression ratio. "GAP" [14] represents the global average pooling. "PWConv" represents the 1×1 Point convolution. "BN" [15] means batch normalization. "Leaky-ReLU" [16], "Sigmoid" are activation functions. " $\oplus$ " represents the feature matrix bitwise addition operation. " $\otimes$ " represents the feature matrix bitwise multiplication operation.

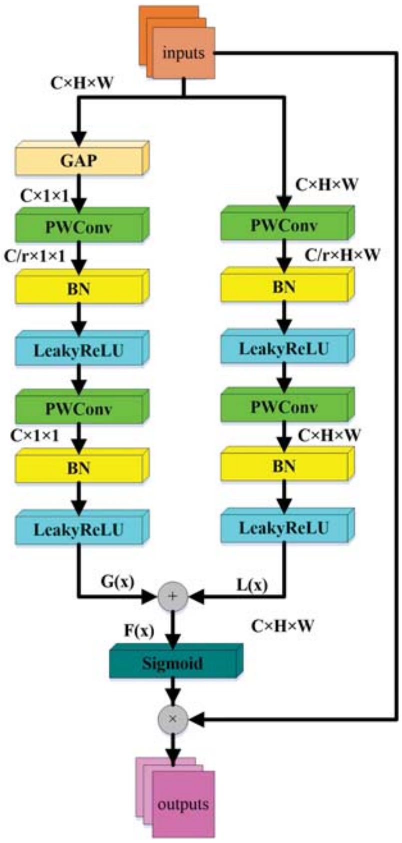

Fig. 1. The Structure of PCAF module

The PCAF module is composed of two parallel branches, namely the global feature extraction branch and the local feature extraction branch. Input feature map X, whose size is $C\times H\times W$, is imported into the two branches for feature extraction. Local feature extraction branch is composed of double point convolution "PWConv". The size of convolution kernel for the first PWConv is $C/r\times 1\times 1$, compressing the channels of feature map to C/r, reducing the dimension of the feature map. The size of the second PWConv convolution kernel is $C\times 1\times 1$, restoring the channels of the feature map to C, raising the dimension of the feature map.

On the basis above, the global feature extraction branch introduces the global average pooling layer (Global Average Pooling, GAP), which consists of one "GAP" and two "PWConv". The GAP operation can compress the global information around the feature map into a real number, which has the receptive field of global information to a certain extent.

Therefore, the global feature extraction branch focuses on extracting widely distributed global information in the feature map. The size of feature map in local feature extraction branch remains $H\times W$. It has not been compressed by the global average pooling from beginning to end. Consequently, more attention is paid to extract the local subtle information of the feature map.

III. MCFF-NET

Based on the PCAF module, three convolutional neural networks with different depths are proposed: Multi-scale Channel Feature Fusion Network (MCFF-Net), as shown in TABLE I. When calculating the network depth in TABLE I, a PCAF module is recorded as one layer, and the depth of the classifier in MCFF-Net is uniformly recorded as one layer. The "Conv" structure in TABLE I can be expressed as an integrated structure containing "convolution", "Batch Normalization" and "ReLU activation function". The value after "Conv" represents the number of channels corresponding to the structure. The network diagram is shown in Fig. 2.

In traditional convolutional neural networks, such as AlexNet [17] or VGGNet [7], three fully connected layers (3 Full Connection layer, 3-FC) are used as classifiers. This can increase the non-linear expression ability of network, accompanied by a large amount of memory occupation and high calculation overhead, which has caused a substantial increase in the amount of network parameters. In order to reduce the network parameters, our network uses a fully connected layer (1-FC) as the classifier to convert the computational overhead of the image recognition task to the convolutional layer, which reduces the burden of the fully connected layer.

Due to the extremely large number of features output by the convolutional layer, one Fully-Connected layer which presents as the classifier will cause tremendous parameters. Therefore, we first reduce the output feature map size of the convolutional layer to 1×1 through the GAP operation, and then the fully connected layer realizes the function of classifying, which reduces the parameter of the network model to a certain extent. "GAP-FC" is used to represent this structure.

TABLE I. MCFF-NET CONFIGURATION

MCFF-Net50	MCFF-Net66	MCFF-Net134

Conv7×7-64, stride 2					
3×3 Maxpooling, stride 2					
Conv3×3-64 Conv3×3-64 **PCAF-64**	×3	Conv1×1-64 Conv3×3-64 Conv1×1-256 **PCAF-256**	×3	Conv1×1-64 Conv3×3-64 Conv1×1-256 **PCAF-256**	×3
Conv3×3-128 Conv3×3-128 **PCAF-128**	×4	Conv1×1-128 Conv3×3-128 Conv1×1-512 **PCAF-512**	×4	Conv1×1-128 Conv3×3-128 Conv1×1-512 **PCAF-512**	×4
Conv3×3-256 Conv3×3-256 **PCAF-256**	×6	Conv1×1-256 Conv3×3-256 Conv1×1-1024 **PCAF-1024**	×6	Conv1×1-256 Conv3×3-256 Conv1×1-1024 **PCAF-1024**	×23
Conv3×3-512 Conv3×3-512 **PCAF-512**	×3	Conv1×1-512 Conv3×3-512 Conv1×1-2048 **PCAF-2048**	×3	Conv1×1-512 Conv3×3-512 Conv1×1-2048 **PCAF-2048**	×3
average pooling					
classifier,softmax					

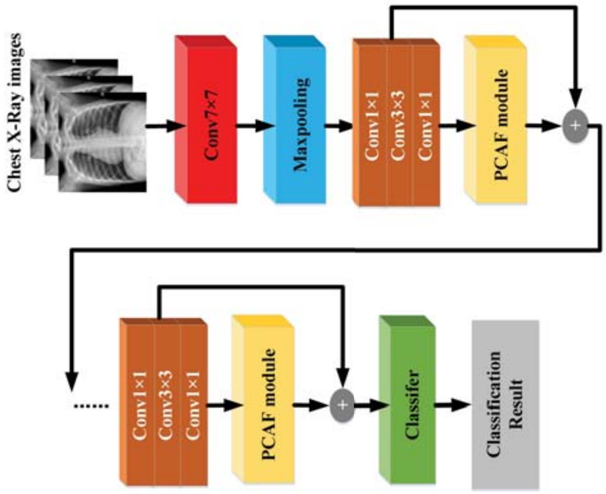

Fig. 2. The Network structure of MCFF-Net

Besides, 1×1 point convolution is considered to be inserted in front of the GAP structure, reducing the dimensionality of the output feature map at the end of the network. The classifier designed under this idea is nothing to do with the fully connected layer, thereby further reducing the amount of parameter. "Conv1-GAP" is used to represent this structure.

IV. Network complexity

When MCFF-Nets with different depths use different classifiers, the parameters are shown in TABLE II. Comparison of floating-point of operations(FLOPs) are shown in TABLE III.

TABLE II. The parameters comparison of MCFF-Net(in millions)

	MCFF-Nets50	**MCFF-Net66**	**MCFF-Net134**
GAP-FC	23.08	45.79	82.74
Conv1-GAP	23.08	45.78	82.73
1-FC	46.37	144.1	181.04

It can be seen from TABLE II that the choice of the classifier has a great influence on the amount of network parameters. In the case of the same network depth, the networks using the "1-FC" classifier are obviously more than those using other classifiers. Therefore, when the hardware conditions and memory conditions are limited, the "1-FC" classifier should be avoided as much as possible under the premise of ensuring the classification accuracy. In addition, the network depth also has a huge impact on the amount of network parameters. The number of parameters of MCFF-Net134-GAP-FC is 3.90 times that of MCFF-Net50-GAP-FC, and the parameters of MCFF-Net134-GAP-FC is 1.26 times that of MCFF-Net66-GAP-FC.

TABLE III. The Comparison of floating points of operations(FLOPs, in millions)

	MCFF-Nets50	**MCFF-Net66**	**MCFF-Net134**
GAP-FC	3781.22	5792.03	11286.62
Conv1-GAP	3805.82	5890.28	11384.86
1-FC	3695.81	5890.23	11384.81

According to TABLE III, the computational cost is mainly determined by the depth of network. MCFF-Net134 is very computationally intensive. Compared with MCFF-Net66, MCFF-Net134 has a FLOPs increase of 94.87%. MCFF-Net66 has an increase of 53.18% compared to MCFF-Net50. Compared with MCFF-Net66, MCFF-Net134 has an increase of 194.87%, which is the largest increase in calculations. In conclusion, when the difference of experimental result is not obvious, for the purpose of saving computational cost, MCFF-Net66 has the highest cost performance.

V. Experiments and Results

A. Data sets

Since COVID-19 is a new type of disease, there is a lack of datasets suitable for this study. In this paper, we have constructed a dataset-A by collecting CXR images from public image databases.

Dataset-A combines the images of two public databases [18,19], which contains three types of CXR images, namely COVID-19, pneumonia, and normal, with a total of 6252 images. The training set has a total of 5526 images, containing 310 images of COVID-19 patients, 1341 normal images, and 3875 normal pneumonia images. There are 726 images in the test set, including 102 images of COVID-19 patients, 234 normal images, and 390 ordinary pneumonias. The sample images from the dataset are shown in Fig. 3.

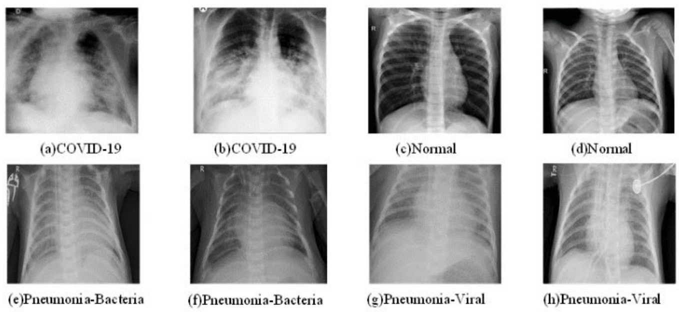

Fig. 3. Chest X-ray images

During the training process, in order to allow the model to learn more feature information from images, we first resize the training set and the test set images to a fixed resolution of 224×224 size. Then perform data enhancement processing on the training set images. The data enhancement technique was achieved by the frame of deep learning named fastai. Obtaining the data-enhanced training set, the amount of data in the training process is equivalent to 4 times that before processing.

B. Experimental setup

As the precondition, the whole experiments are conducted on unanimous software platform and environment, which guarantee the credibility of the comparison results among various network models. TABLE IV indicates the software configuration detail of the experimental platform. The batch size of the training set and the test set are both 16.

TABLE IV. EXPERIMENTAL PLATFORM CONFIGURATION

Attributes	Configuration information
operating system	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz
GPU	GeForce RTX 2080
CUDNN	CUDNN 7.5.0
CUDA	CUDA 10.0.130
Frame	Fastai
IDE	PyCharm
Language	Python

The learning rate annealing algorithm is introduced in the training process, and a larger learning rate is used in the initial stage of training. As the number of iterations increases, the learning rate is gradually reduced. This algorithm can avoid large fluctuations of classification accuracy in the later stage of training, so as to get closer to the optimal solution. After repeated experiments, we finally adjusted the parameter settings as follows: the initial learning rate is set to 0.001. Since the first 50 epochs, the learning rate decays twice as much as before, and then decreases by 2 times every 50 epochs. A total of 300 epochs are used for training. In order to evaluate the performance of the model more objectively, we take the recognition accuracy of the last 10 epochs on test set to calculate the average value, which is used as the final classification accuracy.

C. Experimental results

In this section, we will explain the evaluation indicators used to quantify the classification performance of the model: accuracy, precision, sensitivity, specificity, and F1-score. In order to study the depth and the classifier of MCFF-Net's influence on CXR image classification performance, 9 types of MCFF-Net are used to execute experiments on Dataset-A with 3 types of CXR images. For each experiment, the average value of the last 10 epochs is used as the experimental result, which is shown in TABLE V.

In the light of TABLE V, the recognition performance of MCFF-Net using the "1-FC" classifier is totally lower than the network using the other two classifiers. MCFF-Net134-Conv1-GAP performs best on Dataset-A, with the highest experimental result in the TABLE V, the various indicators are respectively 97.70%, 98.40%, 97.62%, 98.45% and 98.01%. The overall accuracy of MCFF-Net66-Conv1-GAP is 96.79%, which is about 0.5% higher than MCFF-Net50 and about 0.5% lower than MCFF-Net134. The result shows that as the network deepens, the classification capability of the network does not change notably. Considering the complexity of the model, the calculation amount of MCFF-Net134 is 1.933 times and 2.99 times that of MCFF-Net66 and MCFF-Net50, respectively. And the amount of parameter is 1.807 and 3.584 times that of MCFF-Net66 and MCFF-Net50, respectively. Compared with the shallower network, the overall accuracy of the deep network is only less than 1% higher. The overall accuracy of MCFF-Net66-Conv1-GAP is higher than that of MCFF-Net66-1FC and MCFF-Net66-GAP-FC. Considering the classification performance and model complexity, and has the best cost performance.

TABLE V. PERFORMANCE OF DIFFERENT DEPTH MCFF-NET (%)

Model	Accuracy	Precision	Recall	Specificity	F1-Score
MCFF-Net50-1FC	95.96	97.18	95.74	97.27	96.45
MCFF-Net50-GAP-FC	96.32	97.3	95.99	97.48	96.64
MCFF-Net50-Conv1-GAP	96.61	97.29	96.40	97.76	96.84
MCFF-Net66-1FC	96.19	97.01	95.91	97.48	96.46
MCFF-Net66-GAP-FC	96.49	97.30	96.44	97.69	96.87
MCFF-Net66-Conv1-GAP	96.79	97.72	96.54	97.85	97.12
MCFF-Net134-1FC	96.68	97.46	96.54	97.77	97.00
MCFF-Net134-GAP-FC	97.52	98.21	97.42	98.33	97.81
MCFF-Net134-Conv1-GAP	**97.70**	**98.40**	**97.62**	**98.45**	**98.01**

In TABLE VI, more detailed results are given for the recognition performance of MCFF-Net66-Conv1-GAP.

TABLE VI. PRECISION, RECALL, SPECIFICITY AND F1-SCORE OF MCFF-NET66-CONV1-GAP (%).

Class	Precision	Recall	Specificity	F1-score
COVID-19	100	100	100	100
Normal	97.73	91.88	98.98	94.71
Pneumonia	95.30	98.72	94.34	96.98
Average	97.68	96.87	97.77	97.23

According to the TABLE VI, our proposed model MCFF-Net66-Conv1-GAP can efficiently help classify CXR images of COVID-19 positive patients, normal and ordinary pneumonia patients. What's more, the overall accuracy, sensitivity, specificity and F1-score of COVID-19 images have reached 100%. We compare MCFF-Net66-Conv1-GAP with the existing classic deep learning methods such as ResNet50 and DenseNet121, etc. The comparative experimental results are presented in TABLE VII.

TABLE VII. PERFORMANCE COMPARISON WITH OTHER CNNS (%)

Model	Accuracy	Precision	Recall	Specificity	F1-score	COVID-19ACC
ResNet50 [20]	93.53	96.01	93.15	96.53	94.56	98.4
DenseNet121[21]	93.11	95.98	92.75	96.38	94.34	99.02
GoogleNet [22]	92.56	95.29	91.56	95.78	93.37	95.1
VGG19[7]	93.11	96.09	92.93	96.47	94.49	**100**
MCFF-Net66-Conv1-GAP	**96.79**	**97.72**	**96.54**	**97.85**	**97.12**	**100**

ResNet50 constructs an identity mapping between the input and output of the convolutional layer, which solves the degradation problem caused by the deepening of the network to a certain extent, and has a good performance in CXR image recognition tasks. DenseNet121 realizes the repeated use of features by introducing dense connections in the network, and further deepens the network depth on the basis of ResNet. Good classification accuracy is also achieved on Dataset-A, but the calculation amount of DenseNet is much higher than our method. The accuracy of VGG19 is about 3.68% lower than that of MCFF-Net66-Conv1-GAP. Due to its superficial depth, it is hard for the network to amply extract the features from CXR image. Moreover, VGG19 uses three fully connected layer as a classifier, which greatly increases the parameter and calculation, improves the requirements for hardware resources, and reduces the timeliness of CXR image classification tasks. The PCAF module proposed in this paper can fuse the global feature information and local feature information in the CXR image, so that the network can learn more important features from the image in a targeted manner, enhance the classification performance and the generalization ability of the network.

VI. CONCLUSIONS

In this paper, a Parallel Channel Attention Feature Fusion (PCAF) module is designed according to the characteristics of CXR images. And based on this module, we propose a brand new convolutional neural network MCFF-Net to classify CXR images to diagnose and detect COVID-19 cases. Through the analysis and comparison of the experimental results, we believe that MCFF-Net66-Conv1-GAP has the highest application value. The overall accuracy of the 3-class classification experiment and the COVID-19 image recognition accuracy have reached 96.79% and 100%, respectively. Despite the excellent results have been realized, MCFF-Net still needs clinical research and testing. We will overcome the limitations of hardware conditions and train our proposed MCFF-Net method with a larger dataset to further improve its classification accuracy.

REFERENCES

[1] Zu Z Y, Jiang M D, Xu P P, et al. Coronavirus disease 2019 (COVID-19): a perspective from China[J]. Radiology, 2020: 200490.

[2] World Health Organization. Use of chest imaging in COVID-19: a rapid advice guide, 11 June 2020[R]. World Health Organization, 2020.

[3] Wang W, Tian J, Zhang C, et al. An improved deep learning approach and its applications on colonic polyp images detection[J]. BMC Medical Imaging, 2020, 20(1): 1-14.

[4] Wang W, Li Y, Zou T, et al. A novel image classification approach via dense-MobileNet models[J]. Mobile Information Systems, 2020, 2020.

[5] Wang W, Yang Y, Li J, et al. Woodland Labeling in Chenzhou, China, via Deep Learning Approach[J]. International Journal of Computational Intelligence Systems, 2020, 13(1): 1393-1403.

[6] Wang W, Liu H, Li J, et al. Using CFW-Net Deep Learning Models for X-Ray Images to Detect COVID-19 Patients[J]. International Journal of Computational Intelligence Systems, 2021, 14(1):199-207.

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[8] Apostolopoulos I D, Mpesiana T A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks[J]. Physical and Engineering Sciences in Medicine, 2020: 1.

[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[J]. Scientific Reports, 2020, 10(1): 1-12.

[10] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[11] Ozturk T, Talo M, Yildirim E A, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images[J]. Computers in biology and medicine, 2020, 121: 103792.

[12] Khan A I, Shah J L, Bhat M M. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images[J]. Computer Methods and Programs in Biomedicine, 2020, 196: 105581.

[13] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[14] Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.

[15] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015: 448-456.

[16] Xu B, Wang N, Chen T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv preprint arXiv:1505.00853, 2015.

[17] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[18] Cohen JP, Morrison P, Dao L. Covid-chestxray-dataset. 2020. 2003.11597. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[19] Paul Mooney. Chest-X Ray Images(Pneumonia). [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[21] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[22] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.