

Transfer learning-based algorithm for classification of lung X-ray images of Covid-19

Hui Zhang*, Kun Liu

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

* Corresponding Author

Abstract

To address the problem of how to effectively classify the images for a small number of samples with a small amount of data available for the lung X-ray images of Covid-19, this paper investigates a transfer learning-based method for classifying the X-ray images of Covid-19. First, a feature extractor is trained using X-ray images of the lungs of ordinary pneumonia patients and normal people with relatively large amounts of data; then, it is fine-tuned and used for the classification of Covid-19 X-ray images, while the lung region of X-ray images is extracted using the semantic segmentation model UNet to reduce the effect of noise. The final prediction stage uses the visualization method Grad-CAM to locate the network decision region and enhance the persuasive power of the network model. The experimental results show that the proposed method achieves an average accuracy of 91.8% on the task of four classifications (Covid-19, normal, bacterial pneumonia, viral pneumonia) and 96.4% on the task of three classifications (Covid-19, pneumonia, normal).

Keywords

Covid-19, transfer learning, image classification, ResNet50.

1. Introduction

Starting in late 2019, a novel coronavirus COVID-19 (also known as SARS-CoV-2) has been identified in humans as a new class of disease, and outbreaks are affecting the world in an unprecedented manner, and once infected, patients are likely to experience a variety of signs and symptoms of infection, including fever, cough, fatigue, dyspnea, and headache. The main screening method currently used to detect COVID-19 is the real-time reverse transcription polymerase chain reaction (RT-PCR)[1], a test performed on swab samples collected from the respiratory tract; however, RT-PCR evaluation is very time-consuming and a repetitive manual procedure, which often leads to a high degree of subjectivity[2]. Therefore, chest scans, such as X-ray and computed tomography (CT) scans, are performed on all individuals with potential signs of pneumonia in order to diagnose and isolate infected individuals more quickly, although CT scans provide finer detail, X-ray is faster, easier to photograph, less invasive and more economical. Yet because COVID-19-infected pneumonia has many similarities to other types of viral or bacterial infections, differentiating it from other pneumonias by radiologists can be a time-consuming and challenging task. The use of computer-aided diagnosis (CAD) techniques combined with deep learning models for the classification of COVID-19 patients is necessary, and the use of deep learning methods as an auxiliary screening tool for physicians has attracted a great deal of interest in the last few years, such as the use of deep learning techniques for the diagnosis of lung diseases[3] and skin cancer classification, and deep learning CAD systems have been shown to be competent and reliable.

Due to the privacy of medical images and the specialized knowledge required for labeled data, it is difficult to collect large amounts of labeled data to train neural networks, which affects the

performance of deep learning techniques. To alleviate this problem, the literature [4] proposed Covid-Net, one of the first open-source networks designed specifically for detecting neocoronary pneumonia from lung X-ray images. The literature [5] proposed a patch-based convolutional neural network method, which uses random patch cropping with relatively few trainable parameters and can effectively prevent overfitting for less data. The literature [6] proposes a semi-supervised learning (SSL) method using a small amount of data for training, the main idea of this method is to use the information provided by labeled and unlabeled data to improve the learning performance by assembling a large amount of unlabeled data with a small amount of labeled data. The literature [7] proposes CoroNet: This model is based on the Xception architecture pre-trained on the ImageNet dataset. The literature [18] proposed the use of five pre-trained convolutional neural network based models for lung X-ray images for the detection of neocoronary pneumonia and found experimentally that the highest classification performance was obtained with the ResNet50 model.

A common mechanism in the field of transfer learning is to fine-tune existing models pre-trained using large-scale natural image datasets such as ImageNet. However, this approach has not achieved satisfactory performance in medical image applications due to the significant differences between lung X-ray images and natural images. Inspired by this, this paper investigates a deep learning approach based on homologous migration learning based on the previous study. First, images of the lungs of ordinary pneumonia patients and normal people with a large amount of data compared to the X-ray images of new coronary pneumonia are collected, and a feature extractor is trained using these images. Then, using the migration learning method, it is fine-tuned and used on the classification task of new coronary pneumonia X-ray images with a small amount of data. To suppress the influence of useless features, this paper uses the semantic segmentation model UNet to remove irrelevant parts of the X-ray images and keep the lung region. Finally, the experimental results show that the proposed method in this paper can effectively improve the classification accuracy of new coronary pneumonia X-ray images, and the performance is better than the current models often pre-trained using ImageNet datasets, and the method can be applied to medical image classification tasks targeting small amounts of data.

2. Basic Network Theory

In this paper, we use ResNet50 [8], which is commonly used for classification tasks, as the base model. ResNet is a convolutional neural network model proposed by Kaiming He et al. Its main idea is to introduce a residual network, which can introduce the data output of a layer in a number of leaning layers directly skipping multiple layers to the input part of the later data layers, and the structure is shown in Figure 1.

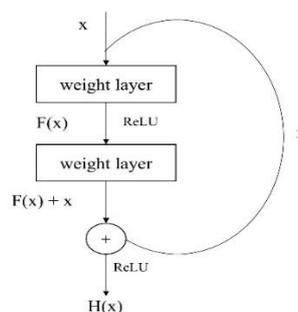


Figure 1: Structure of residual block

Where ReLU is the activation function, the residual block includes the weight layer, the input x is directly connected to the output through the skip layer, $F(x)$ is the residual mapping, $H(x)$ is the original mapping, the residual network makes the stacked weight layer fit the residual

mapping $F(x)$ instead of $H(x)$, then $F(x)=H(x)-x$, this design is to overcome the problem that the learning efficiency becomes low and the accuracy cannot be effectively improved due to the deepening of the network. ResNet50 network has two basic structural blocks, Conv Block and Identity Block, where Conv Block is used to change the dimension of the network; Identity Block serves to deepen the network. The structure of Conv Block is shown in Figure 2 left, we can see that Conv Block can be divided into two parts, the left part is the main part, there are two convolution, normalization, activation function and one convolution, normalization; the right part is the residual edge part, there is one convolution, normalization, because the residual edge part has convolution, so you can use Conv Block to change the width and height of the output feature layer and the number of channels. The structure of Identity Block is shown in Figure 2 right, Identity Block can be divided into two parts, the left part is the main part, there are two convolution, normalization, activation function and one convolution, normalization; the right part is the residual part, directly connected with the output, because the residual edge part does not exist convolution, so the dimension of the input feature layer and the output feature layer of Identity Block is the same, which can be used to deepen the network.

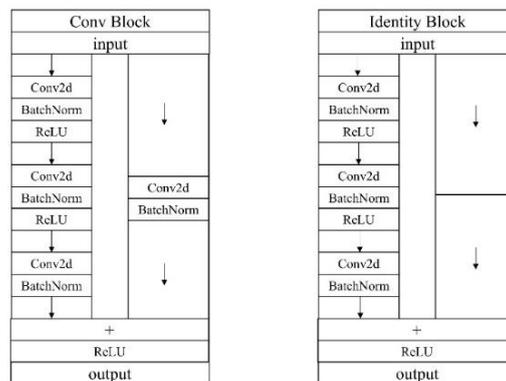


Figure 2: Structure of Conv Block and Identity Block

The overall network structure is shown in Figure 3, where Conv2D is the convolutional layer, BatchNorm is the batch normalization layer, MaxPool is the maximum pooling layer, AveragePooling2D is the average pooling layer, and FC is the fully connected layer.

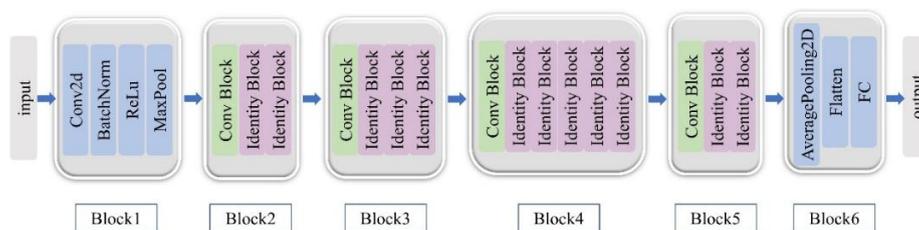


Figure 3: Network framework of ResNet50

3. Transfer learning-based algorithm for lung X-ray classification

A transfer learning-based classification algorithm is proposed for the sparse data volume of Covid-19 X-ray images in order to improve the classification performance. First, the dataset required for the experiment is obtained and analyzed, and the dataset is divided into dataset one containing a large number of common pneumonia and normal lung X-ray images and dataset two containing lung X-ray images of Covid-19, normal, bacterial pneumonia, and viral pneumonia; then, dataset one is input to the ResNet50 network to train the source task, and the best classification effect is saved by adjusting the parameters of the network training weights as pre-training weights for the next transfer learning step; next, the model weights trained in the previous step are migrated to dataset II for training using the idea of transfer learning. In

order to suppress the influence of unimportant features on the model, this paper uses the semantic segmentation network UNet to extract the lung region in the dataset II images. The overall flow of the migration learning-based lung X-ray image classification method in this paper is shown in Figure 4.

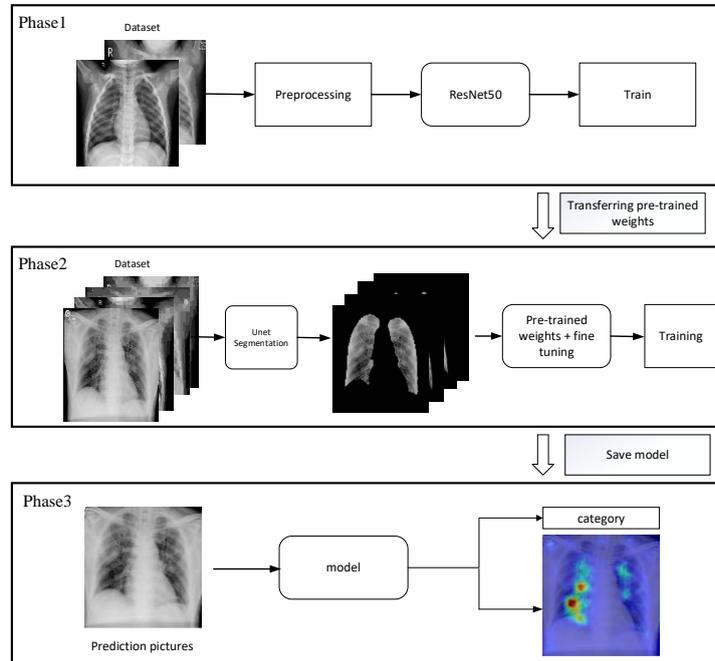


Figure 4: Overall flow chart

3.1. Database

The dataset of this paper consists of two parts. The first part of which is the lung X-ray images collected from Guangzhou Medical Center [9], with a total of 5856 (1583 normal X-ray images, 1493 non-COVID viral pneumonia X-ray images and 2780 bacterial pneumonia X-ray images), which is used to train the feature extractor for migration learning; the second part of the dataset II joins the COVID-19 patients' lung X-ray images from the publicly available GitHub repository [10]. Four types of small data volume datasets were produced, including normal X-ray images, viral pneumonia X-ray images, bacterial pneumonia X-ray images, and COVID-19 pneumonia X-ray images, the numbers of which are shown in Table 1, and the training sample example images are shown in Figure 5.

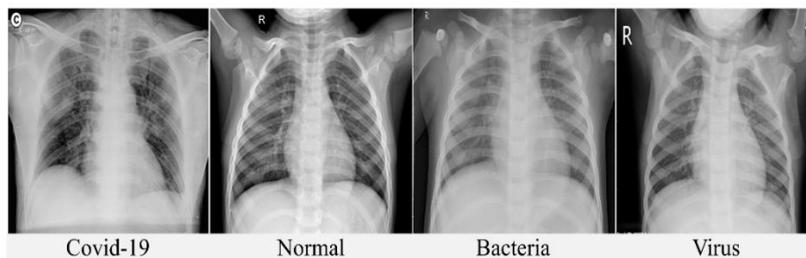


Figure 5: Examples of a sample data set

Table 1: Data set classes and number

Classifications	Numbers
COVID-19	255
Normal	234
Bacterial pneumonia	242
Viral pneumonia	148

3.2. Training source task models

This phase uses dataset one to train the source task for the next phase to be used as pre-training weights using the migration learning method. The data is first preprocessed by converting the images to three-channel color images suitable for classification network input and resizing the images to a uniform size (224, 224), and for images with different pixel ratios, the top and bottom edges of the images are filled with a gray background to avoid distortion and then scaled in size. And augment the dataset using random rotation, scaling, and color gamut distortion. The image data are input into the ResNet50 network to first perform a convolution, normalization, and activation function, and then generate a feature layer after multiple Conv Block and Identity Block operations, and then use the Softmax function after averaging pooling and fully connected layers to obtain the probability that the image belongs to each category with probability. Due to the class imbalance in the dataset, the number of images of bacterial pneumonia is significantly more than the other two species. In this paper, we use the method of adjusting the class weights to reduce the impact of imbalanced data, even though the class weights are inversely proportional to the number of input data, as shown in equation (1).

$$w_i = \frac{TotalNumber}{numberofclasses \times numberofcaseinclass(i)} \quad (1)$$

w_i is the weight of the class of i , $TotalNumber$ is the total number of samples, $numberofclasses$ is the total number of sample classes, and $numberofcaseinclass(i)$ is the number of samples of the class of i . By adding class weights to the loss function, the model increases the penalty for the majority class and takes more care of the classes with fewer samples. 100 iterations are performed to train the resnet50-based convolutional neural network model using the cross-entropy loss function, whose expression is shown in equation (2).

$$Loss(x, label) = -x_{label} + \log \sum_{j=1}^N \exp(x_j) \quad (2)$$

where x is the result generated by the model, $label$ is the label corresponding to the data, N is the number of samples in a batch, using the Adam algorithm to optimize the loss function, by observing the training effect set the learning rate to 0.001, every 5 epoch to reduce the learning rate by a factor of 0.9, each batch (batch) of 32 samples, in order to speed up convergence, using the network on the ImageNet dataset pre Training weights, save the model with the lowest loss value of the validation set, after pre-training, the powerful feature extraction function of the convolutional neural network has extracted certain features of the X-ray image, which is beneficial to fit the data in the next step.

3.3 Training a classification model for X-ray images of Covid-19

Based on stage 1, the method of transfer learning is utilized. Transfer learning is a machine learning method that migrates the knowledge structure learned in the source domain to the target domain, enabling the target domain to achieve better learning results. It includes instance-based migration, feature-based migration, and migration based on model shared parameters. Among them, the migration learning method based on model shared parameters can embed a network model pre-trained in a large-scale dataset into other task models as a feature extractor, thus effectively extracting deep feature information from small sample data [11]. Firstly, data processing is performed, as there are irrelevant objects in the whole X-ray image, in order to suppress unimportant features, the lung region is extracted using the semantic segmentation model UNet [12], the UNet model consists of a backbone feature extraction part, an enhanced feature extraction part and a prediction part, the image data is

firstly passed through the backbone feature extraction part to obtain the feature information, and then through the enhanced feature extraction part to upsample the feature information and fuse the features to obtain an effective feature layer that fuses all the features, and finally the prediction stage uses the effective feature layer to classify each pixel point. The specific approach is to obtain the semantic segmentation dataset from kaggle [13], including 800 lung X-ray images and corresponding lung masks, input to Unet model training, and save the model with the highest Dice coefficient. The evaluation index of semantic segmentation is the Dice coefficient, which is an ensemble similarity coefficient, usually used to calculate the similarity of two samples. The calculation formula is shown in equation (3).

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

Where X represents the true result and Y represents the predicted result, the model effect was evaluated by calculating the percentage of the area of the mutually overlapping part between the two contours and their total area. The best Dice coefficient achieved in this experiment is 0.958. Subsequently, the image data from dataset II was input into the model for prediction in batch, and the pixel points labeled as background in the image were set to black to obtain the image with only the lung region, and the specific results are shown in Figure 6

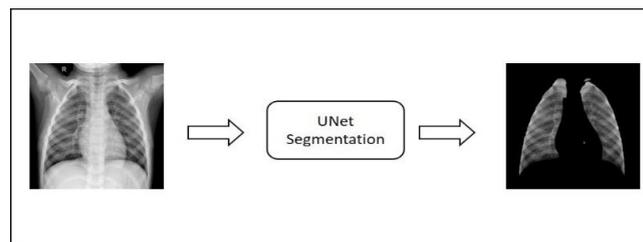


Figure 6: UNet semantic segmentation results

In this paper, we use a migration learning method based on model sharing parameters, and classify common pneumonia and normal human lung X-ray images with a large amount of data as the source domain task, migrate the model weights of the source domain task to the classification task of new coronary pneumonia images, and use the model weights obtained from the training of the source task as the feature extractor to extract features, which can avoid training the neural network from scratch, the specific steps are to delete the original fully connected layer and replacing it with a fully connected layer suitable for this task. Adding Dropout regularization to the fully connected layer can improve the generalization ability, and Dropout will randomly delete neurons during the training process to achieve the effect of reducing overfitting. The obtained images of the lung region are used as input to the classification network, and the network structure is fine-tuned. Due to the small amount of data, in order to prevent overfitting, the first five modules of the network are first frozen, keeping the weights unchanged, and only the fully connected layers suitable for this task are trained. The Adam optimizer will make the network learn more and more information and reduce the gap between the training set and the validation set, and then a part of the layers are unfrozen for training, and some of the initial layers of the model are frozen, because the initial layers of the network extract generic features of the images, using a smaller The learning rate 0.0001, which will control the size of the adjusted network weights, will lead to the loss of the previously accumulated knowledge when using a relatively large learning rate. In the final classification stage, Softmax classification is used for the multi-classification task. After migration learning, a good performance is also achieved for the classification task of new coronary pneumonia X-ray images with small data size. In this study, four types (covid-19, viral

pneumonia, bacterial pneumonia, normal) and three types (covid-19, pneumonia, normal) of image data were experimented using the proposed model.

3.4 Results and discussions

The experiments were conducted on a computer with NVIDIA RTX 2060 GPU with 16GB memory and Pytorch framework. To verify the effectiveness of the method in this paper, the same data were used to conduct experiments using the pre-training weights of this paper and ImageNet pre-training weights, respectively. For medical image classification, Accuracy, Precision, Recall, and F1-score are used as evaluation metrics, where F1-score is the summed average of Precision and Recall, and the larger the value is, the better the specific formula is shown in (4).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

For the four classification tasks, the precision, recall, and F1-score of the proposed model for COVID-19 classes are 93.78%, 98.25%, and 95.96%, respectively, improving the classification performance for small data volumes, as shown in Table 2.

Table 2: Classification results of the four categories

Classification	Precision(%)	Recall(%)	F1-score(%)
COVID-19	93.78	98.25	95.96
Normal	98.21	98.67	98.43
Bacterial pneumonia	89.27	88.45	88.86
Viral pneumonia	88.9	86.58	87.72

Figure 7 shows the training set and validation set loss descent curve during the training process, the red curve is the train loss, the orange curve is the validation loss.

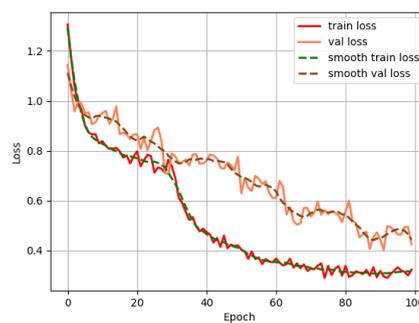


Figure 7: Training set and validation set loss descent curve

Also this paper was compared with the methods of literature [7] and literature [14], which also performed 4 classification, and the average accuracy performance is shown specifically in Table 3, the proposed method improved the classification performance of the Covid-19 X-ray images.

Table 3: Comparison of accuracy of each model on four classifications

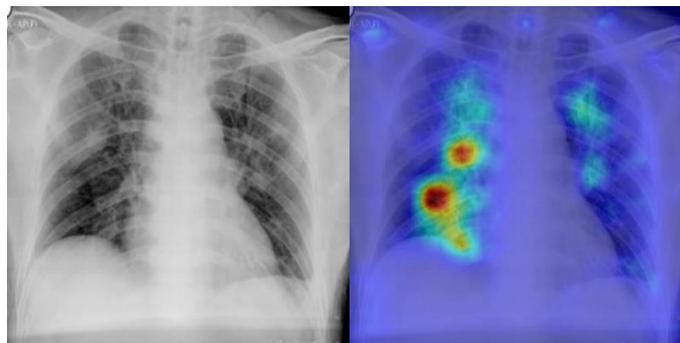
Method	Accuracy(%)
CoroNet[7]	89.9
CovXNet[14]	90.2
Proposed	91.8

From the data in Table 2, it is concluded that the accuracy of the model for the classification of bacterial pneumonia and viral pneumonia is low, and after observing the image data, it is found that there is a great similarity between these two types of images. In order to compare with the triple classification experiment in the previous papers, the next study classified bacterial pneumonia and viral pneumonia as common pneumonia and selected 200 for the experiment, and the experiment obtained a triple classification (normal, pneumonia, COVID-19) with an accuracy of 96.4%, which is shown in Table 4 for comparison with other methods in the literature.

Table 4: Accuracy comparison of the models on the three categories

Method	Accuracy(%)
Proposed	96.4
CoroNet[7]	95
Morteza et al[15]	94.5
Tulin et al[16]	87.02
Loannis et al[17]	94.72
Bhowal et al[2]	95.49

To enhance the persuasiveness of the network, the Grad-CAM visualization method was used in the prediction phase to display the decision part of the network, showing which part of the image the network paid the most attention to by means of a heat map. The results are shown in Figure 8.



(a) Prediction pictures (b) Grad-CAM Method Results

Figure 8: Visualization results

The darker the color in the graph, the more attention the network pays to the location, where the red part indicates the highest attention the network pays to the location, and the blue part indicates the low attention the network pays to the location.

4. Conclusion

In order to solve the problem of insufficient data volume of lung X-ray images of Covid-19, a transfer learning based classification model is proposed to pre-train on homologous tasks with similarity and extract lung regions using a semantic segmentation method compared to the traditional use of weights pre-trained on ImageNet dataset, and the proposed method performs better for other methods, and the visualization algorithm Grad-CAM is used to display the decision part of the network model and locate the lesion regions, which enhances the persuasive power of the network model.

Acknowledgements

Aviation Science Foundation(AFSC-20195501001)

References

- [1] Wang, W. et al. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA (2020).
- [2] Bhowal P, Sen S, Yoon JH, Geem ZW, Sarkar R. Choquet Integral and Coalition Game-based Ensemble of Deep Learning Models for COVID-19 Screening from Chest X-ray Images. IEEE J Biomed Health Inform. 2021 Sep 9;PP. doi: 10.1109/JBHI.2021.3111415. Epub ahead of print. PMID: 34499608.China National Standardization Management Committee. Specifications of Crane Design (China Standardization Press, China 2008), p. 16-19.
- [3] Wang H , Xia Y. ChestNet: A Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography[J]. 2018.
- [4] 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).
- [5] Y. Oh, S. Park and J. C. Ye, "Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2688-2700, Aug. 2020, doi:10.1109/TMI.2020.2993291.
- [6] Qi X , Noshier J L , Foran D J , et al. Multi-Feature Semi-Supervised Learning for COVID-19 Diagnosis from Chest X-ray Images[J].2021.
- [7] 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, 2020:105581.
- [8] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778 (2016).
- [9] Kermany, D; Zhang, Kang; Goldbaum, M(2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, V2, doi: 10.17632/rschbjbr9sj.2
- [10] J.P. Cohen, P. Morrison and L.Dao, COVID-19 image data collection, arXiv: 2003.11597 , 2020 <https://github.com/ieee8023/covid-chestxray-dataset>.
- [11] Donahue J ,Jia Y , Vinyals O , et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition[C]// International Conference on Machine Learning. JMLR.org, 2013.
- [12] Ronneberger O , Fischer P , Brox T . U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer International Publishing, 2015.
- [13] S.Jaeger et al., "Automatic Tuberculosis Screening Using Chest Radiographs," in IEEE Transactions on Medical Imaging, vol.33, no.2, pp.233-245, Feb.2014, doi:10.1109/TMI.2013.2284099. <https://www.kaggle.com/nikhilpandey360/chest-xray-masks-and-labels>.
- [14] Mahmud T, Rahman M A, Fattah S A. CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization[J]. Computers in Biology and Medicine, 2020, 122.
- [15] Mh A , Sm A , Azk B , et al. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms[J]. International Journal of Medical Informatics, 2020, 144.
- [16] To A , Mt B , Eay C , et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images[J]. Computers in Biology and Medicine, 121.
- [17] Apostolopoulos I D , Bessiana T . Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks[J]. 2020.
- [18] Narin A , Kaya C , Pamuk Z . Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks[J]. 2020.