Covid-19 Image Classification Using ResNet

Yu Yang
Tongji University
1850217@tongji.edu.cn

LiFan Zhou
Tongji University
1854141@tongji.edu.cn

YangSen Chen
Tongji University
1851350@tongji.edu.cn

Abstract

The image classification between the Covid-19 and the normal pneumonia is a meaningful problem under the background of pandemic of the Covid-19. Therefore, in this paper, we use some classical classification networks to classify the Covid-19, Non-infected and Cap and try to find the best networks for this problem on our datasets including the subject level and the slice level. By comparing these classic networks, we found that using ResNet can better perform classification operations on our datasets. On the final test set, an accuracy of 80% *is achieved. Finally, we add some training tricks on the basis of the classic network to ensure that our model will not overfitting, which is also a very critical point in the process of training the network.*

1. Introduction

Corona Virus Disease 2019 (COVID-19), referred to as ”New Coronary Pneumonia”, named by the World Health Organization as ”Coronavirus Disease 2019”, refers to pneumonia caused by the 2019 novel coronavirus infection. Since December 2019, some hospitals in Wuhan City, Hubei Province have successively discovered multiple cases of unexplained pneumonia with a history of exposure to the South China Seafood Market, which were confirmed to be acute respiratory infectious diseases caused by 2019 new coronavirus infection

COVID-19 has completely changed the world and affected several aspects of modern life. Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case was identified in Wuhan, China, in December 2019. It has since spread worldwide, leading to an ongoing pandemic.

The automated computer-assisted detection of new coronary pneumonia will help diagnosers quickly distinguish the type of patient's disease, assist doctors and nucleic acid detection, and quickly identify and classify CT images to obtain the status of the detected person, which is helpful In order to quickly identify high-risk personnel so that quarantine measures can be taken as soon as possible, this research will be of great help to the prevention and control of this new crown pneumonia epidemic.

By inputting a series of CT pictures of the patient's lungs, we can analyze whether the patient has new coronary pneumonia and distinguish between ordinary pneumonia and new coronary pneumonia. These are the goals of our research. Our research methods are mainly based on the training and learning of convolutional neural networks, the main model used is Resnet, and a comparative analysis with other deep learning models.

Image classification is the most basic task in computer vision. Basically, the development history of deep learning models is the development history of image classification tasks. However, image classification is not that simple and has not been completely solved.

Image classification is also a task for comparing almost all benchmark models. From the simpler 10-class grayscale image handwritten digit recognition task mnist, to the larger 10-class cifar10 and 100-class cifar100 tasks, to the later imagenet task, the image classification model is accompanied by the growth of the data set , Step by step up to today's level. Now, in a data set with more than 10 million images and more than 20,000 categories like imagenet, the level of computer image classification has exceeded that of humans.

medical image processing aims at processing medical images with different imaging mechanisms. The main types of medical imaging widely used in clinical practice are X-ray imaging (X-CT), magnetic resonance imaging (MRI), nuclear medicine imaging (NMI) and ultrasound imaging. (UI) Four categories.

In the current medical imaging diagnosis, it is mainly through observing a set of two-dimensional slice images to find the diseased body, which often needs to rely on the doctor's experience to determine. Use computer image processing technology to analyze and process two-dimensional slice images, realize segmentation, extraction, three-dimensional reconstruction and three-dimensional display of human organs, soft tissues and lesions, which can assist doctors in qualitatively identifying lesions and other areas of interest Even quantitative analysis can greatly improve the accuracy and reliability of medical diagnosis; it can also play an important auxiliary role in medical teaching, surgical planning, surgical simulation and various medical research. At present, medical image processing mainly focuses on four aspects: lesion detection, image segmentation, image registration and image fusion.

2. Related work

CNN is a highly parallelizable algorithm. Compared with single-core CPU processing, the graphics processing unit (GPU) computer chips used today have achieved substantial acceleration (about 40 times). In medical image processing, GPUs were first introduced for segmentation and reconstruction, and then for machine learning. Due to the development of new variants of CNN and the emergence of efficient parallel network frameworks optimized for modern GPUs, deep neural networks have attracted commercial interest. Training a deep CNN from scratch is a challenge. First, CNN requires a large amount of labeled training data, which may be difficult to meet in the medical field where expert annotation is expensive and diseases are scarce. Secondly, training deep CNN requires a lot of computing and memory resources, otherwise the training process will be very time-consuming. Third, the deep CNN training process is complicated by overfitting and convergence problems, which usually requires repeated adjustments to the network's framework structure or learning parameters to ensure that all layers are learned at a comparable speed. In view of these difficulties, some new learning solutions, called ”transfer learning” and ”fine-tuning”, have been proven to solve the above problems and are becoming more and more popular.

Resnet[1] presents a residual learning framework to ease the training of networks that are substantially deeper than those used previously. It explicitly reformulates the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. It provides comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth.

VGG[3] makes a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. The representations of VGG generalise well to many datasets, where it achieves state-of-the-art results.

DenseNet[2] connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections - one between each layer and its subsequent layer - the network has $L(L+1)/2$ direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

3. Our method

Plain Nework. The plain baselines of resnet are mainly inspired by the philosophy of VGG nets.The convolutional layers mostly have 3×3 filters and follow two simple design rules: (i) for the same output feature map size, the layers have the same number of filters; and (ii) if the feature map size is halved, the number of filters is doubled so as to preserve the time complexity per layer. We perform downsampling directly by convolutional layers that have a stride of 2. The network ends with a global average pooling layer and a 1000-way fully-connected layer with softmax. Resnet has fewer filters and lower complexity than VGG nets:The 34-layer baseline has 3.6 billion FLOPs (multiply-adds), which is only 18% of VGG-19 (19.6 billion FLOPs)

Residual Network. In figure 1, based on the above plain network, shortcut connections are inserted into the network which turn the network into its counterpart residual version. The identity shortcuts can be directly used when the input and output are of the same dimensions. When the dimensions increase, two options are considered: (A) The shortcut still performs identity mapping, with extra zero entries padded for increasing dimensions. This option introduces no extra parameter; (B) The projection shortcut is used to match dimensions (done by 1×1 convolutions). For both options, when the shortcuts go across feature maps of two sizes, they are performed with a stride of 2.

50-layer ResNet. The network is presented in figure 2. Each 2-layer block in the 34-layer net is replaced with a 3-layer bottleneck block, resulting in 50-layer ResNet.The model has 3.8 billion FLOPs.

4. Experiments

In this section, we evaluate our method to classify images on two-dimensional datasets. The first one is on the subject level. The another is on the slice level. Furthermore, we evaluated the effect of the method on the datasets using the prediction accuracy as evaluation metrics. At the same time, the corresponding implementation details will be clarified in this section.

4.1. Datasets

Subject level dataset. In this dataset, patients are classified into three categories, representing normal CT images of patients infected with covid-19, and patients infected with normal pneumonia. In each category, all pictures are placed

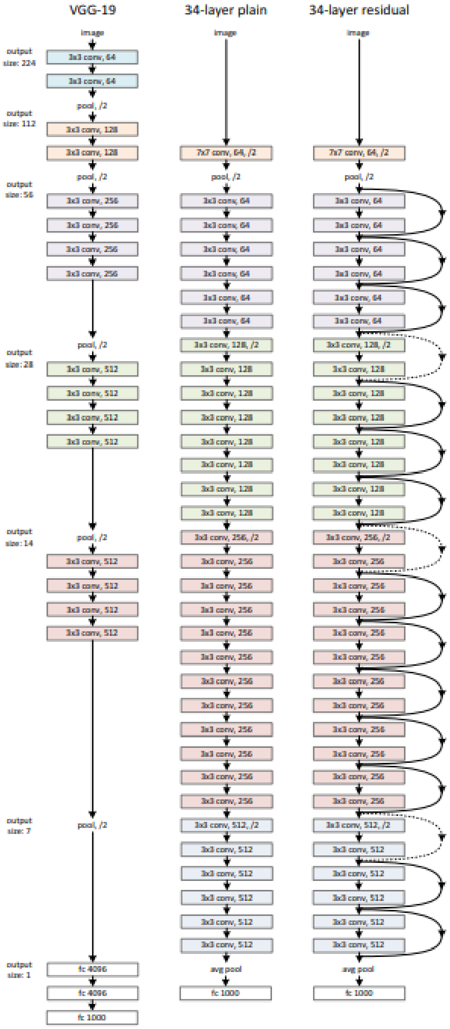
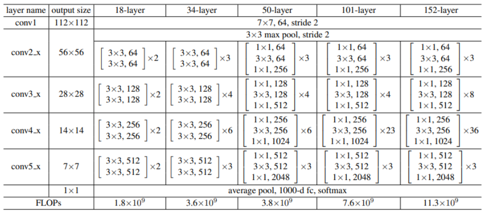


Figure 1. Residual Network



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 2. 50-layer ResNet

in the same directory file according to the patients they belong to. Therefore, the data set describing the subject-level should be described in terms of individual patient cases. This means that in this data set, each patient is a data sample. The patient catalog contains all the CT pictures of the patient, and the number of CT pictures of the patient is not all the same. More than 100 sheets, less than 50 sheets. In the pictures of these patients, because the data sets are divided according to whether the corresponding patients are sick. In other words, for a patient, the CT picture in his catalog file will contain the non-infected part and the infected part. Therefore, it can be said that there will be a lot of noise in this data set if the image classification is directly performed. Using pictures directly for classification will cause a lot of disturbance. In the data set, the picture is given in a size of 512 × 512, and contains detailed lung pictures.

Slice level dataset. In this dataset, the patients given are cases of normal pneumonia and covid-19, and their description information. This descriptive information shows in detail the corresponding lesion location of a patient. The pictures of the patient's lungs are also classified into three categories, which represent CT pictures not infected, CT pictures infected with new coronary pneumonia, and CT pictures infected with normal pneumonia. In each category, all the pictures are not placed in the same directory file according to the patients they belong to as mentioned above, but the pictures of all patients are extracted and mixed, so the data set describing the slice-level is based on the ct picture description. Therefore, this data set has undergone certain preprocessing when it is obtained from the original picture. That is, to determine which category the picture belongs to by a given label. In this data set, the pictures are also given in the size of 512 × 512, and contain detailed lung pictures.

4.2. Performance metrics

In order to obtain the effect of the model we adopted, in the training process, we divided a part of the data set into the corresponding verification set to test the prediction accuracy of the corresponding model three classification. Generally speaking, for a classification model, we already have a good way to verify the accuracy of the corresponding prediction model. We usually adopt a measurement index of dividing the correct prediction by the total number of predictions as the corresponding evaluation standard, which is also logical.

Forecast accuracy rate. The mathematical expression of prediction accuracy is shown in the formula. The ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{\sum_i C_{i,i}}{\sum_{i,j} C_{i,j}} \tag{1}$$

where $C_{i,j}$ represents the number of predicted Class j for the actual Class i, where $i, j \in \{1, 2, 3\}$ in three-way, and $i, j \in \{1, 2\}$ in binary classifications.

In this process, in tabel 1, in order to test the effect of our model on the three classifications of pictures. At the

Method	Accuracy	Loss	Valid Accuracy
DenseNet201	0.82	0.823	0.67
ResNet50	0.86	0.533	0.83

Table 1. Results. ResNet50 is better.

Method	Accuracy	Loss	Valid Accuracy
Vgg16	0.86	0.623	0.82
ResNet50	0.93	0.201	0.91

Table 2. Results. ResNet50 is better.

subject level, we used densenet201 as one of our comparative experiments. At the slice level, we used vgg16 as one of our comparative experiments. During the experiment, we found that the accuracy of training on the subject level dataset is not lower than that of the slice level, which is in line with our previous discussion about noise in the subject level dataset. For the subject level, we can see that the accuracy of the training set using densenet201 converges quickly, but it converges to a not high position. In the table, we can also clearly see that densenet201 has a lower accuracy than the resnet50 training set we used. At the same time, we can also see the same rule on the verification set. On the verification set, the accuracy of our densenet201 is much lower than resnet50. At the same time, we can also find that during the training process, the corresponding curve of densenet201 will oscillate back and forth. According to speculation, this is probably due to the noise in the subject-level during the training process. But in the process of using resnet50, there is no such concern, resnet50 correspondingly provides a good anti-interference ability.

As a refined slice-level, in tabel 2, there is no such concern. Because there is no corresponding noise to interfere in this dataset. Through the table, we can see that in the training process, we used vgg16 as a comparison test, and the experimental results also show that the network model using vgg16 is not as good as resnet50. This undoubtedly shows that we choose the correctness of the experimental model. During the training process, we found that the parameters of vgg16 are too large for the network, which leads to a very long training process, but the effect is not as good as resnet50. This may be related to our data set and the characteristics of the network itself. During the training process, the accuracy of the training set was maintained at a high level, and the corresponding verification set was also maintained at a reasonable level. But in the actual test, we found that the corresponding accuracy rate is far lower than the result on the slice-level. Therefore, we made a speculation that during the training process, the learned model was not the target of the pneumonia lesion, but the target of the lung contour. Therefore, we added random cropping in the preprocessing stage to ensure the training of the image content.

4.3. Implementation details

Image enhancement. During the implementation of the network model, we adopted some image preprocessing methods to help us obtain a better data set. Through normalization, to process the pictures in the input batch. By observing the data set, we also found that there are some images with too high or too low gray levels inside the data set. Through the normalization method, we can keep them to a consistent level. At the same time, in order to prevent the model mentioned in the above process from training the thoracic contour instead of the lesion. We have enhanced the data. In the process, we cropped, translated and rotated the input of the picture to the model. In this way, the external thoracic cavity contour will be cut off by our image processing method, highlighting the internal lesion part, we have completed the image preprocessing process, and achieved the effect of image enhancement.

Label smoothing. The label smoothing operation is a regularization strategy, which is mainly to add noise through soft one-hot, which reduces the weight of the real sample label category in the calculation of the loss function, and finally has the effect of suppressing overfitting. In the actual operation, it is assumed that there may be errors in the label during training, so as to avoid excessive trust in the label of the training sample. This also fits the corresponding noise problem we encountered in the process of training the subject level. Through the label smoothing method, the accuracy of our final project prediction has been further improved.

Perturbation factor. The perturbation factor is added in the input image part of the code. Adding a disturbance factor to this part can play a role in interference to the training set and prevent the occurrence of model overfitting. The impact of the disturbance factor on the model is mainly reflected in the accuracy of training. By adding new man-made noise, our model can play a corresponding role in the actual implementation of the prediction process.

Early termination. Finally, an early termination method is adopted to ensure the generalization ability of the model. In the process of model training, when our adaptive learning rate drops to a certain level, we can use this method to end the training of the model, otherwise, too long training will definitely cause the model to fall into the local optimal value and cause the pan The chemical capacity is far below the expected level.

Through the above several image enhancement and overfitting operations, we finally completed the training of our three-category model, and obtained certain effects. The training curve during the model process is shown in the figure.

5. Conclusion

In this work, we classify the CT images among the Non-infected, Covid-19 and Cap by using a classical classification model. We present ResNet50 to solve this problem. The proposed method is compared with the DenseNet201 and Vgg16. We evaluate the proposed method on our datasets on subject level and slice level. Furthermore, we propose some tricks such as label smoothing, perturbation factor and early termination to prevent the overfitting. In the end, our project achieved an accuracy of 80% during the entire test process, and the effect is very impressive.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 2

[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2