International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 1255-1262 www.ijaem.net ISSN: 2395-5252

Research on Image Classification Algorithm Based on Convolutional Neural Network[1]

Haoning Pu, Zhan Wen* ,Yiquan Li,Yanhe Na
School of Communication Engineering, Chengdu University of Information Technology, Chengdu, China;

Date of Submission: 07-06-2020 Date of Acceptance: 23-06-2020

ABSTRACT: At present, the application of convolutional neural network (CNN) in image classification has become one of the hotspots in the field of intelligent vision, and it has excellent performance in image processing. Deep learning often requires a lot of time and computing resources for training, which is also a major reason for bothering the development of deep learning algorithms. Although CNN has achieved good results in the field of image classification, its algorithm has a great impact on the effect and efficiency of image classification. Only the optimization algorithm that requires fewer computational resources and faster model convergence can fundamentally accelerate the learning speed and effect of the machine. Prevention and diagnosis of COVID-19 is a hot issue during this pandemic. Identifying COVID-19 from lung CT images can assist diagnosis of patients, which is of certain significance to improve the accuracy of COVID-19 diagnosis. In this paper, three classical CNN model are used to analyze the lung CT image data, which are divided into normal conditions, common pneumonia and COVID-19, so as to realize the recognition of COVID-19. In this scheme, image preprocessing is firstly carried out. Due to the small amount of open data set on the Internet, data expansion method is adopted to expand the data, and then training set and test set are divided into 4:1. Finally, AlexNet, VGG and ResNet CNN are used to extract the features of lung CT images respectively. The detection accuracy of three kinds of networks for pneumonia was obtained. The final test results show that VGG network has a better recognition effect on pneumonia detection. At the same time, VGG network can converge faster in training with calibrated data set. Tests on the data set show that VGG CNN can successfully classify CT images as COVID-19, normal and viral pneumonia with 94.47% accuracy. It can be proved that the model is feasible in pneumonia recognition and classification.

KEYWORDS: CNN; image classification; COVID-19; CT images

I. INTRODUCTION

The COVID-19 epidemic, which began in December 2019, has caused serious harm to human health, and accurate diagnosis of the severity of COVID-19 and timely treatment is an important response to the outbreak. Among them, lung CT images are an important basis for the diagnosis of COVID-19. Therefore, accurate identification of COVID-19 based on CT images has become an important research direction. In the early stage of the onset of this coronavirus pneumonia, blood routine test results are almost worthless, while nucleic acid detection is an important basis for pathogen diagnosis and the most important basis for diagnosis at present. However, due to the shortage of nucleic acid test reagents, an effective method is needed to assist diagnosis. Imaging diagnosis includes CT and chest X-ray. For the diagnosis of pneumonia, it can be seen whether there is inflammation in the lung, which is of great value for subsequent research. As imaging diagnosis is not a quantitative index, it is difficult to be as clear as blood results, and manual judgment is also difficult. If there is no quantitative result of blood test, only artificial imaging screening to determine whether further diagnosis is needed may bring uncertain factors [1]. Therefore, the auxiliary diagnosis of CT images in the diagnosis of novel coronavirus can, to a certain extent, help front-line medical staff to classify CT images of COVID-19 patients and

[1] the Network and Data Security Key Laboratory of Sichuan Province, UESTC (NO.NDS2021-7).

effectively improve the diagnostic accuracy and efficiency.

In the past, CT images were often processed through manual diagnosis by medical personnel. For small data, manual processing may be better, but in the face of large data, not to mention the efficiency of diagnosis, even the accuracy is difficult to guarantee. Therefore, introducing CNN into the identification of COVID-19 is expected to solve the above problems.

In recent years, deep learning methods have not only demonstrated excellent performance in computer vision tasks that take natural images as analysis objects, but also made breakthroughs in the field of medical images. At present, most histopathological and microscope image segmentation methods are based on CNN. CNN framework is widely used in lesion classification. Anthimopoulos et al. used CNN to design a multi-classification framework to distinguish the modes of interstitial lung disease, such as ground-glass lesions, honeycomb lesions, calcification and pulmonary nodules, with an accuracy of about 85.5%[2]. Jiao et al. used CNN to extract depth features at different levels to improve the classification accuracy of breast cancer [3]. Tajbakhsh et al. Compared the large-scale training of artificial neural network (Massive - trainingartificialneuralnetworks MTANNs) with CNN both the end-to-end performance of artificial neural network training, the experimental results show that it is only with less training data, Performance of MTANN is significantly higher than that of CNN[4]. In 2015, KaimingHe and other four Chinese from Microsoft research institute proposed ResNet (ResidualNeuralNetwork) [5,6], successfully trained 152-layer neural network by ResNetUnit, and won the champion in ILSVRC2015 competition. The error rate of top5 is 3.57%, and the number of parameters is lower than that of VGGNet, so the effect is very outstanding. The structure of ResNet can accelerate the training of neural network very quickly, and the accuracy of the model is also greatly improved.

Some scholars also combine CNN with other basic models to achieve classification. Such as Kallenberg according to the characteristics of the CNN and SAE using unsupervised training convolution sparse automatic coding machine (Convolutionalsparselyautocoder, CSAE) model, realize the segmentation and the risk of breast and breast density evaluation [15]. Shi et al. used the new deep polynomial network to classify tumors in the Z-small sample ultrasound data set, and the classification accuracy in the chest and prostate data sets was 92.4% and 90.28%, respectively [7]. There are also a small number of works using other deep learning methods to achieve the detection of interested targets or lesions. For example, Xu et al from Nanjing University of Information Science and Technology used SSAE network to learn depth features to identify the nuclei of breast cancer in histopathological images and determine the staging of breast cancer [8].

In 2014, RossGirshick proposed R-CNN and successfully applied it to the problem of target detection [9]. This model for the first time combined candidate region with CNN, trained the model through a large amount of marked data, and finally obtained the boundary box coordinates and classification information of the target to be detected. On the basis of r-cnn, RossGirshick proposed a fastr-cnn in 2015 [10], which is faster than the r-cnn proposed last year, but this r-cnn also has defects. In 2019, DanielBolya et al. proposed a simple full-convolutional model for real-time instance segmentation -- YOLACT[11], which divides instances into two parallel sub-tasks :(1) generating a set of prototype masks and (2) predicting the mask coefficients of each instance. The instance mask is then generated by linear combination of the prototype and mask coefficients, which is much faster than any previous method.

Therefore, this paper adopts the image classification algorithm based on CNN to construct the CT image classification model based on CNN. Through comparative analysis of Alexnet, VGG and RESNET network models, the optimal model is obtained. Auxiliary diagnosis of COVID-19 can effectively improve the accuracy and efficiency of diagnosis and, to a certain extent, help frontline medical staff to classify CT images of COVID-19 patients.

CONVOLUTIONAL NEURAL NETWORK MODEL

AlexNet

Due to the influence of computer performance, although LeNet has achieved good results in image classification, it has not attracted much attention. Until 2012, the AlexNet network proposed by Alex et al. won the ImageNet contest by far surpassing the second place, and CNN and even deep learning attracted extensive attention again. AlexNet deepens the network structure on the basis of LeNet and learns richer and higher-dimensional image features.

As shown in Figure 1, AlexNet can recognize 1000 categories and has 60 million parameters. Its first layer uses 96 11-by-11 filters, which are greatly reduced in width and height by four steps, from 227 to 55. This is followed by a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 1255-1262 www.ijaem.net ISSN: 2395-5252

maximum pooling layer. Then we use the same convolution layer, which uses a certain amount of padding to keep the width and height of the output matrix the same as the width and height of the input matrix. The input matrix is 27 by 27, and the output matrix is 27 by 27. The same convolution layer is followed by a maximum pooling layer. And then it's followed by three successive same convolution layers, all of which are 13 by 13. They are followed by a maximum pooling layer. The pooling layer is followed by the full connection layer. Softmax was used in the input layer to predict a category from 1000 categories. Alexnet uses the RELu activation function.

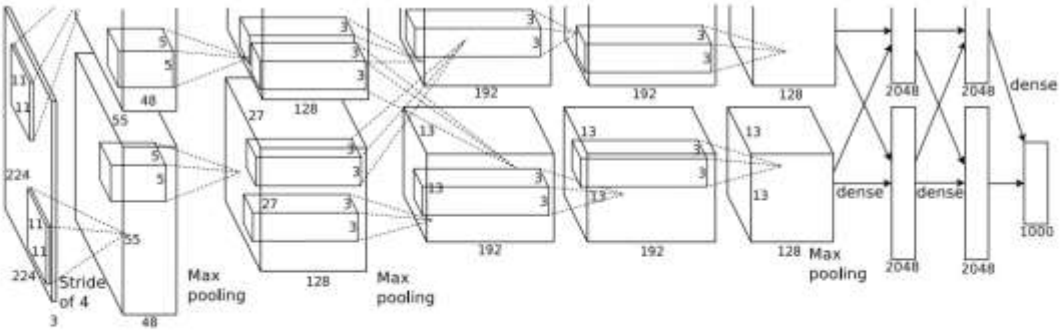

Figure 1 The AlexNet network structure

VGG16

In 2014, researchers at the University of Oxford's Visual Geometry Group and Google DeepMind developed a new deep CNN: VGGNet and won the 2nd place in ILSVRC2014 category (1st place is GoogLeNet). In the same year, he published a paper [12], which mainly focused on the influence of the depth of CNN on the recognition accuracy of large-scale image sets. The main contribution was to construct CNN structures of various depths using small convolutional kernels (3×33×3), and evaluate these network structures. Finally, it was proved that the network depth of 16-19 layer, It can obtain better recognition accuracy. These are also vGG-16 and VGG-19, which are commonly used to extract image features.

In Figure 2, VGG has 138 million parameters. However, its structure is relatively simple. Although VGG constructs the network by constantly combining the convolutional layer and the pooling layer, in addition, each convolutional layer of VGG uses a 3×3 filter with step size of 1 and is the same convolution. In addition, each of its pooling layers is also a maximum pooling of 2x2 with a step of 2. So the structure looks relatively simple. For example, the first CONV64×2 represented two convolution layers, each of which had 64 filters. Because same convolution was used, the width and height of the two convolution layers were still 224, followed by a maximum pooling layer. Because it was 2x2 pooling with step size of 2, the width and height became 112. Then I continued to combine the convolution layer and the pooling layer, and finally made the matrix 7×7×512. It is followed by two fully connected layers. Finally, there is a Softmax layer, which is used to identify 1000 categories. VGG network is very neat. The width and height are constantly shrinking twice, from 224 to 112 to 56...... The depth is continuously doubled from 64 to 128 to 256...... This makes it look simpler and clearer. VGG is also commonly referred to as VGG-16 because it has 16 neural network layers with parameters, 13 convolutional layers, 3 fully connected layers, and pooling layer is not counted because it has no parameters.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2 VGG-16 network structure

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 1255-1262 www.ijaem.net ISSN: 2395-5252

ResNet18

ResNet, proposed by He Keming et al in 2015, won the first prize in the ImageNet competition classification task and was awarded the CVPR2016 Best paper. Since it is "simple and practical", many target detection and image classification tasks are completed on the basis of ResNet, which has become an important cornerstone structure in the field of computer vision. ResNet comes in five main forms: Res18, Res34, Res50, Res101, and Res152. Resnet18 differs from other resnet networks mainly in layer1 to layer4,

Due to the problems of gradient explosion and gradient disappearance, the deeper the neural network, the more difficult it is to train well. Therefore, even if you have enough computing power and data, it is difficult to train a very deep and excellent neural network. A jump connection that avoids gradient explosion and gradient extinction by passing the previous activation value directly to the later network layer through the middle network layer. The neural network constructed using this jump connection is called a residual network (ResNet). Using this technique can help us train very deep neural networks. Sometimes more than 100 layers deep. The network structure of ResNet is shown in Figure 3.

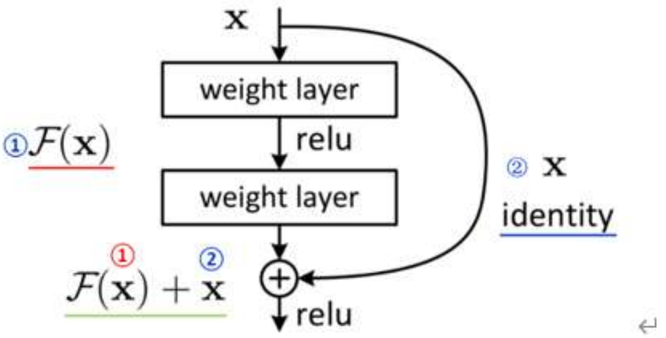

Figure 3 resnet network structure

A summary of the three models

Compared with the other two models, AlexNet has the following characteristics: (1) Deeper network structure (2) use cascading convolutional layers, That is, convolutional layer + convolutional layer + pooling layer to extract image features (3) suppressed overfitting with Dropout, (4) suppressed overfitting with Data Augmentation, (5) replaced sigmoID with Relu as activation function, and (6) multi-GPU training. And VGG structure is relatively simple: it is composed of 5 convolutional layers, 3 fully connected layers and Softmax output layer. Max-pooling is used to separate layers and ReLU function is used to activate units of all hidden layers. In addition, small convolutional kernels and multi-convolutional sub-layers are adopted: VGG uses multiple smaller convolutional kernels (3x3) to replace the larger convolutional layer of one convolution kernel. On the one hand, parameters can be reduced, and on the other hand, more nonlinear mapping can be carried out, which can increase the fitting and expression ability of the network. Resnet residual structure is proposed to solve the problem of network degradation. It has the following characteristics :(1)learning results are more sensitive to the fluctuation of network weight. (2) Residual results are more sensitive to data fluctuations.

II. DATA PREPROCESSING AND PARAMETER SELECTION

The experiment of this paper is to configure python environment under Windows 10 operating system, use PyCharm3.7 software, and install the deep learning framework as Keras-2.1.6 + Tensorflow-2.5.0. In this paper, three classical network models are used for comparative experiments, alexnet network, VGG16 network, Resnet18 network.

Data preprocessing

Although the CT image data set available publicly on the Internet is used, the CT image data set of COVID-19, which is a precious medical image, is still relatively small, so it is still difficult to reach the amount of data required by the training model. Training the deep learning model on such a small data set is prone to over-fitting: The model performs well on training data, but does not generalize well on test data. Therefore, I adopted a data augmentation approach to solve this problem.

The purpose of data augmenting is to combine approximately correct image-tag groups, for example, in most combined image tag groups, the tag is the correct annotation of the image. That is, a new image-label group is created from the limited training data, and the combined group is added to the original training set. When creating

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 1255-1262 www.ijaem.net ISSN: 2395-5252

new groups, I used random affine transformation, random clipping and flipping to expand each training image. Random affine transformations include translation and rotation (angles 5,15,25, in order).

Finally,as shown in figure 4, a data set containing 3886 lung CT images was made, including 1200 CT images of COVID-19, 1341 CT images of normal lungs, and 1345 CT images of viral pneumonia. According to the ratio of 4 to 1, 3109 were divided into training sets and 777 test sets.

Figure 4 The COVID-19 dataset

Selection of the activation function and the loss function

This paper design takes the relu function, and the loss function is the cross-entropy loss function.The relu is a rectified linear unit.It is also one of the most widely used activation functions.The reason for the most widely used activation function is that it is nonlinear and means that the backpropagation algorithm is available.And a very good feature of relu is that it does not activate all neurons simultaneously.Relu function expression:

$$f(x) = \max(0,x) \quad (3.1)$$

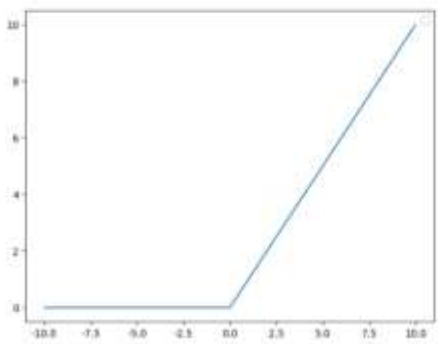

Figure 5 The relu function image

The relu function is seen in Figure 5, where the input is negative, it would output 0, and then the neurons will not be activated.This means that only some neurons are activated at the same time, thus making the network sparse and in turn very computationally efficient.Common activation functions also have sigmoid and tanh, but they have a common drawback —— When the value of the input data is relatively large, the neural network learns very slowly.

Cross-entropy loss functions are often used in neural network classification problems, and this paper is designed for 3 classification with the expression of:

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N} - \sum_{c=1}^{M} y_{ic}\log(p_{ic}) \quad (3.2)$$

Where: M represents the number of categories; y_{ic} represents the indicator variable (0 or 1), if the category and the same category of sample i is 1, otherwise 0; p_{ic} is the predic ted probability that belongs to category c for the observed sample i.

Selection of the learning rate

The design learning rate in this paper is 0.5.In general, specific learning rates should depend on the model.Different data models, the specific learning rate is different, which needs to be constantly adjusted to find the appropriate learning rate, to ensure that the final output value of the system is the optimal solution, and to find the most accurate learning rate that can make the neural network learn the fastest.

Selection of the initial weight value

In this paper, the initial weight value is set to random number.There are two main ways to select the initial weight value. One is by setting the weight values to either 1 or 0, and the other is by setting the initial weight value to any random

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 9 Sep 2021, pp: 1255-1262 www.ijaem.net ISSN: 2395-5252

number. However, when the weight values are all set to 0 or 1, the parameters updated by each layer may be the same value, which cannot reflect the learning ability of the neural network. When the initial weight value is randomly selected, the neural network will automatically distinguish the important and unimportant features according to the data, so as to update the different size of the weight value. Most values of the normal distribution are concentrated near 0, which conforms to the learning characteristics of neural network and is conducive to improving learning efficiency.

Selection of optimization algorithm

Adam optimizer is used in this design.Adam algorithm is actually the combination of momentum gradient descent and RMSprop, which can adaptively change the learning rate, requires less resources, and the model converges faster (find the minimum loss value) to accelerate the learning speed and effect of the machine.

III. THE EXPERRIMENTAL RESULT

Loss function change process of convolutional neural network

As shown in Figure 6, the abscissa is the training times, the ordinate is the loss function value, the dotted line represents the training loss function value of the neural network, and the solid line represents the test loss function value of the neural network. Visible alexnet convolution neural network training loss function value to start falling, falling test loss function values, learning network, loss appeared several times in the process of training oscillation phenomenon, but the overall downward trend, indicate that the network is still in learning can continue to training, but in the end there is still a rising curve, the effect is not very good.

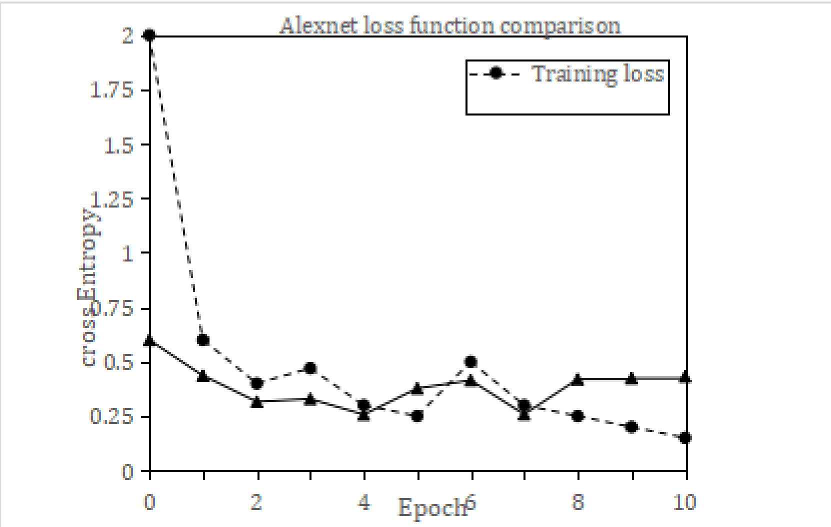

Figure 6 Alexnet network loss function

As shown in Figure 7, it can be seen that the training loss function value of VGG16 CNN begins to decline continuously, and the test loss function value continues to decline, indicating that the network is learning. Training loss tends to remain constant, Validation Loss tends to remain constant, loss value converges, and network learning is completed.

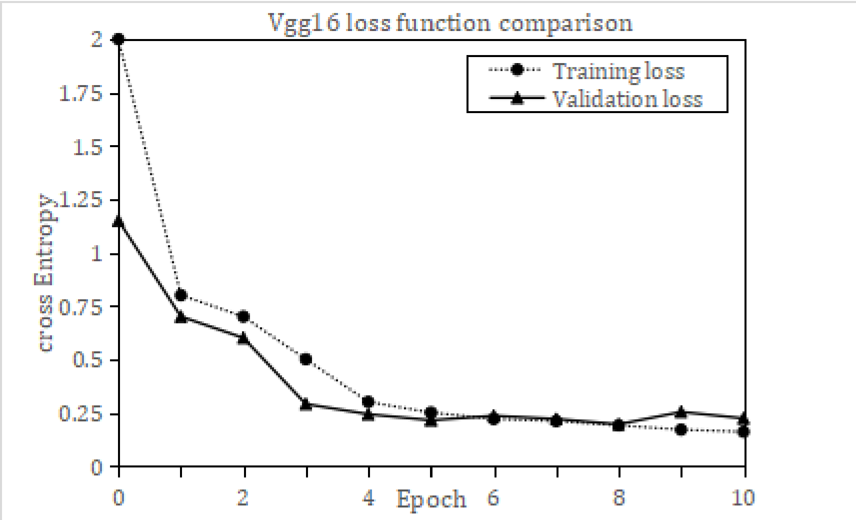

Figure 7 VGG16 network loss function

As shown in Figure 8, the training loss function value of RESNet18 CNN and the test loss function value continue to decline. The network is learning. Several oscillations of loss occur in the training process with a large range, indicating that the network effect is not ideal.

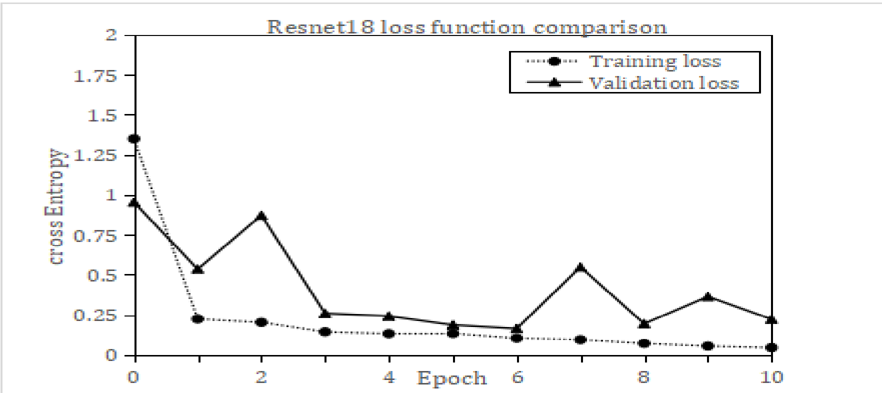

Figure 7 VGG16 network loss function

Comprehensive comparison shows that the change process of loss function of VGG16 network model presents a good trend, loss value is approaching 0 and stable, and the network learning state is good.

The process of accurate value change of convolutional neural network

In the following picture, the abscissa is the training times, and the ordinate is the accuracy, representing the test accuracy of the neural network. Figure 9 shows three CNNs learning and the changing process of test accuracy.

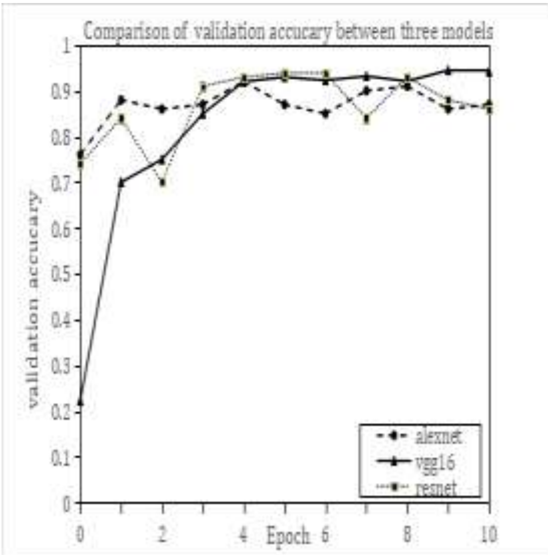

Figure 9 Comparison of test accuracy

According to Figure 9 AlexNet network model test accuracy, it can be concluded that the test accuracy rate keeps increasing to more than 90%, and the model construction is accurate. The test accuracy rate can reach 92% at the highest, but it will stabilize around 85%. The image classification accuracy of network model is not very high.

VGG16 network model test accuracy can be obtained that the test accuracy is improved to more than 90%, model construction is accurate, the highest test accuracy can reach 94.47%, stable around 95%, network model image classification accuracy is very high, good effect.

The test accuracy of ResNet18 network model can be concluded that the test accuracy is increased to more than 90%, the model is constructed accurately, and the highest test accuracy can reach 93.39%, but it is unstable, the network model image classification accuracy is not high, and the effect is poor.

Figure 9 Comparative analysis shows that vGG16 network training accuracy is relatively high, reaching 95%. The network training effect is very good, and the test accuracy is as high as 94.47%. The image classification effect is very good.It can be seen that VGG16 model is optimal for lung COVID-19CT image recognition.

IV. CONCLUSION

In this article, three neural network classification models were used to train 3109 CT images of pneumonia and test 777 CT images of pneumonia. Finally, these images were classified into normal condition, normal pneumonia and COVID-19. Through comparative experiments, analysis of loss function values of training and testing, and changes in training and testing accuracy, it can be found that VGG16 model is the optimal model, and the classification accuracy of CT images of pneumonia is up to 94.47%. In 2020, COVID-19 suddenly spread around the world, completely disrupting people's work and life. Lung CT images are of certain testing value in the detection of COVID-19, which can assist in the diagnosis of COVID-19 and, to a certain extent, help front-line medical staff to classify CT images of COVID-19 patients.

REFERENCES

[1]. Gregory L.Krauss.An introduction to neural networks[Z].1996.

[2]. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Transactions on Medical Imaging, 2016, 35(5): 1207−1216

[3]. Jiao Z C, Gao X B, Wang Y, Li J. A deep feature based framework for breast masses classification. Neurocomputing, 2016, 197: 221−231

[4]. Tajbakhsh N, Suzuki K. Comparing two classes of end-toend machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. Pattern Recognition, 2016, 63: 476−486

[5]. He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 770−778

[6]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, JianSun. Identity Mappings in Deep Residual Networks. IN ECCV, 2016.

[7]. Kallenberg M, Petersen K, Nielsen M, Ng A Y, Diao P F, Igel C, VachonC M, Holland K, Winkel R R, Karssemeijer N, Lillholm M. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Transactions on Medical Imaging, 2016, 35(5): 1322−1331

[8]. Shi J, Zhou S C, Liu X, Zhang Q, Lu M H, Wang T F. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. Neurocomputing, 2016, 194: 87−94

[9]. Xu J, Xiang L, Liu Q S, Gilmore H, Wu J Z, Tang J H, Madabhushi A. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging, 2016, 35(1): 119−130.

[10]. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[11]. R. Girshick . Fast R-CNN[J]. international conference on computer vision, 2015: 1440-1448.

[12]. Bolya,YOLACT Real-Time Instanc Segmentation . ICCV 2019

[13]. Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.