ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

# Pavement Recognition Based on Multiple Deep Learning

**Fei Xie[1,*], Li Yang[1,2], Hailong Duan[1,2]**

*[1]Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjing, China*
*[2]School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjing, China*
**Corresponding author: 842399421@qq.com*

***Abstract:*** *In order to meet the development of The Times and enhance the intelligence of new energy vehicles, and solve the problem that the traditional image processing is too demanding for the feature quantity and the recognition effect is poor, a road recognition research based on deep learning is designed. Four deep learning networks (VGGNet, ResNet, GoogLeNet and AlexNet) were used in the study. A large number of images of different road surfaces were collected as experimental data, and the data were trained by the deep learning network. By comparing the results of four kinds of networks, the network which can be used in intelligent vehicles and has better recognition rate is finally analyzed. The research achieved the expected effect and laid the foundation for the subsequent design.*

***Keywords:*** *Deep learning, road recognition, Convolutional neural networks, Smart cars*

## 1. Introduction

The recognition technology in the intelligent vehicle refers to the sensor to sense the surrounding environment, and the identified information is transmitted back to the controller in real time, which can make it complete the specified operation of a technology. However, the recognition technology in the current intelligent vehicle industry is still in the initial stage, which needs a period of research and development. At the same time, with the continuous innovation of intelligent vehicle technology, road environment sensing technology has played a very important role in the decision-making control and path planning of vehicle driving. In the environment perception, the role of road surface recognition is particularly critical. The realization of this function can not only effectively improve the stability of intelligent vehicles in different road environments, but also ensure that the output torque of the engine in the car can be effectively adjusted. Pavement recognition technology has developed with the progress of The Times. As early as 2005, Xiao Wangxin [1] proposed a new algorithm for pavement damage classification based on the density factor of pavement damage, aiming at the problem of pavement damage classification. Through this algorithm, five pavement damage conditions were classified, and an ideal effect was achieved. Shi Weijia[2] from Hebei University of Technology designed a road recognition system that uses millimeter wave radar sensor group network technology to recognize the road surface. Zhang Xiaolong [3] studied the road surface coefficient and achieved the effect of road surface recognition through the friction coefficient of seven different road surfaces. However, Gao Wei [4] innovatively proposed a road recognition method based on machine vision, which realized the transition from algorithm recognition to deep learning. With the continuous development of deep learning networks, more industries begin to realize the importance of deep learning. Jing Weipeng [5] of Harbin University of Science and Technology classifies the recognized remote sensing images by using deep convolutional neural network. On the basis of deep learning, Lv Haoyuan's [6] team classifies a variety of different images by introducing semi-supervised learning. Zhang Xiaoling [7] combined deep learning with EEG to identify epilepsy. The application of these technologies shows that deep learning technology has been widely used not only in the field of intelligent vehicles, but also in all walks of life. This paper will realize road recognition and classification based on Pytorch environment. Four different neural networks, VGGNet, ResNet, GoogLeNet and AlexNet, are used to recognize and classify four kinds of road images, and the results generated by different networks are evaluated. On the basis of ensuring the feasibility of the method, the experiment compares the four networks with better stability and higher recognition rate, which has a good reference and reference value.

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

## 2. Deep Learning Network Model

### *2.1. Convolutional Neural Network Structure*

In the process of applying deep learning, convolutional neural network plays a major role [8-9]. A simple neural network mainly includes convolutional layer, activation layer, pooling layer and fully connected layer [10]. Its main structure is shown in Figure 1:

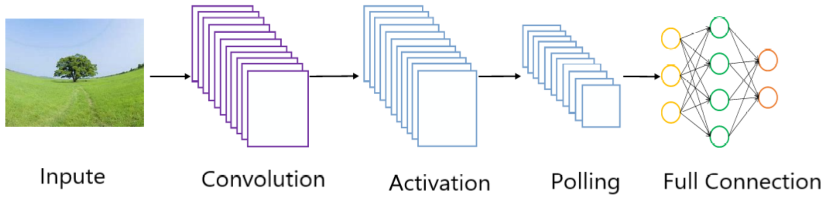


*Figure 1: Convolutional neural network architecture diagram.*

The convolutional layer plays an important role in the convolutional neural network. When the images of different input road types enter the convolutional layer, the convolutional kernel in the convolutional layer starts to extract the features of different roads. Therefore, it can be classified as the feature extraction stage. Meanwhile, in the process of feature extraction, it is completely completed by itself, which reduces the error caused by artificial feature extraction.

The function of the activation layer is to nonlinear process the pavement features output from the previous layer through the activation function, so as to enhance the processing ability of the trained network for nonlinear data.

The function of pooling layer is data dimensionality reduction. After the images of different road surfaces are input to the convolutional layer, a large number of features will be extracted due to the large size of the images themselves, including redundant features, which will bring great interference to the following training. By pooling the layers, redundancy can be removed and important features can be retained, which can most truly reflect the differences between different road types.

The purpose of the fully connected layer is to reorganize and classify the input features of the pooling layer. After several iterations of convolution and pooling, a prominent feature map can be obtained. Before it is input into the fully connected layer, features need to be flattened and mapped into the feature space. Finally, each feature can be classified after internal processing of the fully connected layer.

### *2.2. Convolutional neural network model*

#### *2.2.1. AlexNet Model*

AlexNet [11] as the world's first recognized convolutional neural network, not only showed its innovation in recognition and classification to the outside world, but also was gradually accepted and used for deep model training. But it was in 2012 that AlexNet became widely used. With its powerful computing power, it won the championship in the ImageNet Image recognition Challenge and changed the conditions of deep learning application in machine learning. AlexNet has eight layers of convolutional neural network, among which the first five layers (CONV1 to CONV5) are convolutional layers, and the last three layers are fully connected layers. At the same time, a large number of image enhancement technologies are added, which can solve the feature redundancy caused by the large amount of data. On this basis, local normalization is added to the image preprocessing stage to improve the generalization ability of the model.

#### *2.2.2. VGGNet Model*

VGGNet [12] was proposed by the famous research Group VGG (Visual Geometry Group) of Oxford University in 2014, and up to now, it is still used as an important model for image feature extraction. The network construction includes 13 convolutional layers (CONV1-1 to CONV5-3), 5 pooling layers (pool1 to pool5) and 3 fully connected layers (TC6, TC7 and TC8). Compared with AlexNet network, VGGNet uses smaller 3*3 convolution kernels instead of 11*11, 7*7 and 5*5 kernels, which can reduce the parameters used on the one hand and increase the nonlinear changes of the network on the other hand. At the same time, VGG also adopts the multi-scale method to train and predict the target samples. Through this method, image features can be enhanced and overfitting can be reduced, so as to improve the accuracy in the recognition process.

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

#### *2.2.3. GoogLeNet Model*

GoogLeNet [13-14] is a convolutional neural network built based on the Inception framework. Compared with VGGNet, although the classification of VGG on ImageNet is more accurate, there are more parameters to be adjusted by users. Therefore, by sparse network structures, GoogLeNet connects convolutional layers in a sparse way, which can improve computational efficiency. At the same time, GoogLeNet abandons the commonly used fully connected layer and replaces it with the global mean pooling layer, thus reducing the time of network parameter adjustment. GoogLeNet has also made innovative improvements to the network, as shown in the figure below. On the one hand, it uses a 1*1 convolution kernel to carry out dimensionality reduction, and on the other hand, it can convolve images of multiple sizes at the same time and then aggregate them together.

#### *2.2.4. ResNet Model*

ResNet [15] is also known as deep residual neural network, combines the idea of residual with deep network, which can not only change the number of layers of the network, but also solve the problem of gradient disappearance caused by the increase of layers, and prevent the decline of recognition rate caused by reverse learning. ResNet network is composed of multiple residual blocks. Let F(x) be the resulting output of one residual, H(x) be the input of the following residual, and X be the initial identity mapping. The relationship between the two is $F(x) = H(x) -x$. This one layer of residual structure can be in the transfer of different characteristics at the same time reduces the difficulty of deep learning, and through this method, can make the residual network of 18 layer, 50 and 152 layer, by continuously increasing network layer, make the network on the basis of the original features constantly learning new characteristics, which can effectively improve the performance of the model.

## 3. Identify Processes and Build Models

### *3.1. Recognition of the Road*

The first task for the whole road surface recognition is image acquisition. This paper mainly collects the road conditions commonly encountered by four kinds of vehicles in the process of driving, namely grass, snow, mud and gravel, by means of camera shooting, and takes them as an important research object. Thousands of pictures of each road surface were taken to ensure sufficient data. On this basis, in order to realize the expansion of data volume, rotation, translation and scaling were also adopted to double the database, which can effectively improve the generalization ability of the model during training.

After completing the establishment of the database, you need to collect the image preprocessing, due to the existence of different sizes of images taken, will bring some obstacles to the back of the handle, therefore, by means of normalization to the images of different size all converted into 100 * 100 size, so that we can bring back image processing is convenient.

In the following process, the processed images need to be input into different networks. After continuous convolution, pooling, activation and return, the extracted features are input into the fully connected layer and finally outputed and classified by Softmax. The complete flow chart is shown in Figure 2:

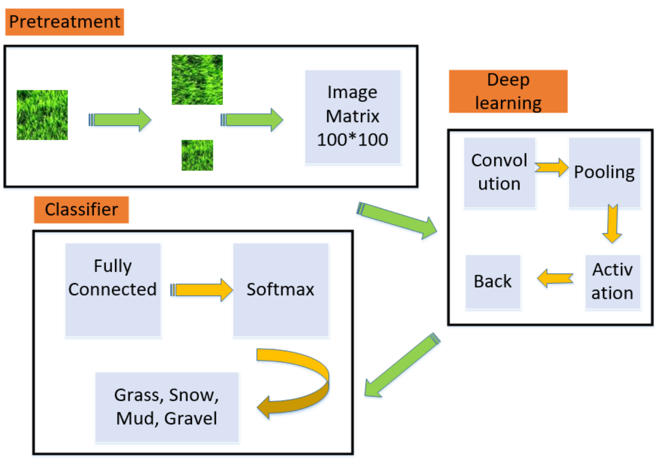


*Figure 2: Pavement recognition flow chart.*

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

### *3.2. Construction of Convolutional Neural Network*

#### *3.2.1. Training of the Model*

The training process of convolutional neural network is as follows:

Step 1: If we want to use this network for road surface recognition, when modifying the code, we take into account that we want to achieve four classifications, but there are too many classifications corresponding to ImageNet in the original model, so we need to remove the fully connected layer in the last layer of the model, and adopt the fully connected layer in line with the classification number of road surface recognition. Adding the Dropout layer while adding the full connection layer can prevent the overfitting phenomenon during the running of the program.

Step 2: The establishment of training set and test set data, and enhance the data. The collected data will be allocated to the training set and the test set according to the standard 9:1. At the same time, before testing the training set, we will enhance our data samples by flipping, scaling and translation, etc. In this way, we can enrich our samples.

Step 3: Adjust the main parameters according to the network. Among them, the more important parameters, for example, Batch_size is set to 32, which means that the data in the training is extracted in the form of 32 pictures each time for training; If the Epoch is set to 20, the number of iterations is 20. Learning is set to 0.001 of the standard, which represents his initial Learning rate and so on.

Step 4: VGG16, RESNET-101, GoogLeNet and AlexNet are respectively used as feature extractors in the network to input the established dataset into the model for training and return different weights.

#### *3.2.2. Test of the Model*

The test process of convolutional neural network is as follows:

Step 1: Load the previously saved weights separately.

Step 2: Start to run the program, select the test set and training set established before, input it into the network.

Step 3: The test set is backpropagated through the network and calculated layer by layer through the network, and finally the network output value is obtained.

Step 4: Compare the output results with the samples of the test set, calculate the correct rate and make statistics.

Step 5: The experiment was repeated for four networks, and the results obtained were compared to get the best network.

## 4. Experimental Results and Analysis

A total of 6,788 images were used, including four different types of road surfaces. VGG16, ResNET-101, GoogLeNet and AlexNet convolutional neural networks were used to train the image set and establish the road recognition and classification model. Then, the accuracy [16] and cross-entropy loss function [17] of the pavement recognition model are analyzed to evaluate the four models. Among them, the accuracy is between 0-1, the closer it is to 1, the higher the accuracy of the model for recognition, while the cross-entropy loss function is on the contrary. The closer the value is to 0, the closer the two probability distributions are, and the more accurate the test results are. In order to reduce the experimental interference caused by external factors in the experiment, the same test set and training set were used in the training and testing process of the four models, and the ratio was distinguished according to 9:1. When setting important parameters, the iteration Epoch is set to 20 times, and the training number Batch_size is set to 32. Other important parameters are kept the same.

According to the steps in 3-2, the data set is trained into 4 label files, and 32 images are trained each time, which can maximize the performance of computer GPU and reduce the training time. After 20 iterations, the final result is shown in the figure below. The figure shows the accuracy and loss values of the four network training and test sets respectively.

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

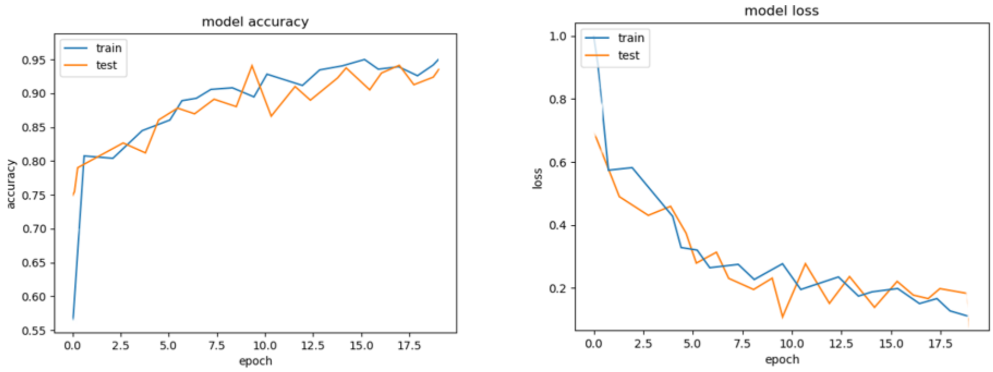


*Figure 3: Accuracy and loss of AlexNet model.*

As shown in figure 3, AlexNet, as the earliest neural network, can obtain a good result with an accuracy around 0.9 when it is used for image recognition and training. With the increase of training times, it can be seen that its growth trend is on the rise. In addition, the loss value began to be stable after the number of iterations reached 10, but the loss value of the test set showed an upward trend with the increase of the number of iterations. Therefore, it can be concluded that attention should be paid to the selection of training times when training the model.

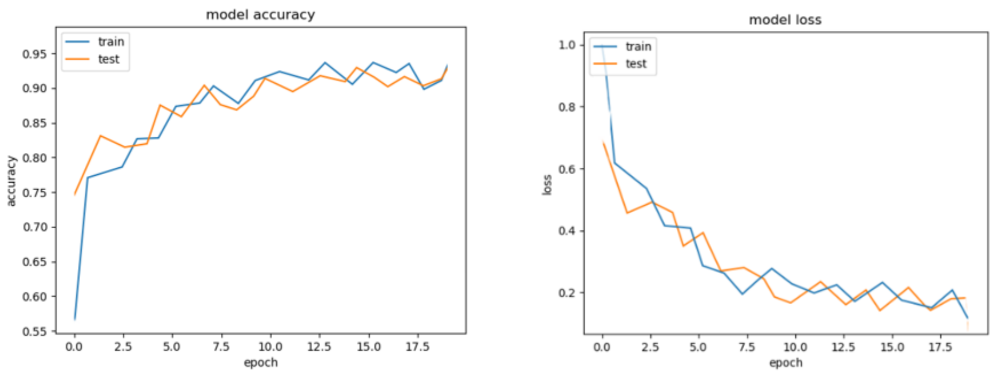


*Figure 4: Accuracy and loss of GoogLeNet model.*

During the model training of GoogLeNet network, it can be seen from the figure 4 that both the correct rate and the loss value remain relatively stable on the curve without large data offset. In terms of the accuracy of the training set and the test set, there are many repetition points between them, and the final value remains between 0.9-0.95. The loss value also maintains a stable downward trend, thus ensuring the stability of the model.

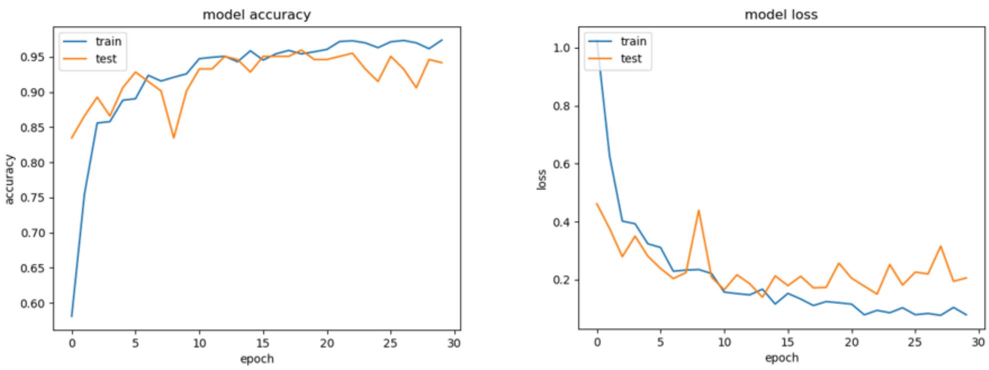


*Figure 5: Accuracy and loss of Resnet-101 model.*

As shown in figure 5, the accuracy of the training set of RESNET-101 model reaches a high of 0.95, and the training set can also achieve a good accuracy when the number of iterations is about 20. However, with the increase of the number of iterations, the accuracy of the test set decreases, but it still remains around 0.92. Meanwhile, in terms of the loss function, with the increase of the number of iterations, the loss value of the training set continues to decrease with good effect, but the test set shows an increasing trend after 15 iterations and finally stabilizes around 0.2.

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

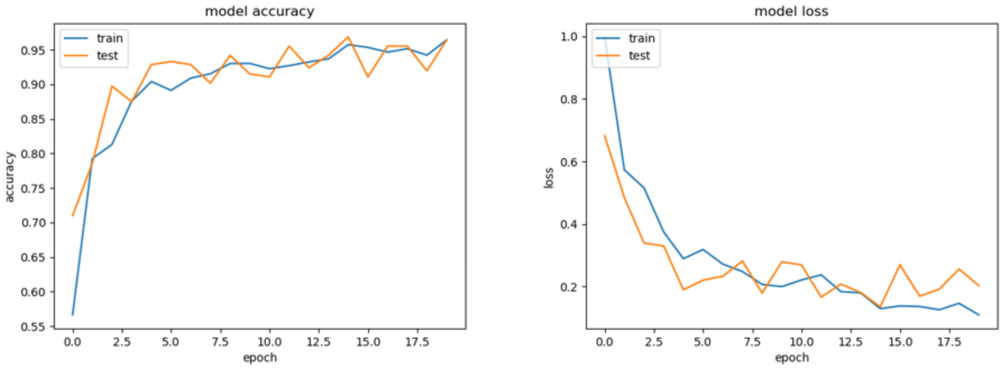


*Figure 6: Accuracy and loss of VGG16 model.*

Compared with ResNET-101 model, VGG16 has a certain improvement in accuracy and loss function. As shown in figure 6. In terms of accuracy, the test set and training set are relatively stable, which can be stabilized around 0.95 after about 20 iterations. Although the loss value of the training set is still higher than that of the test set, it can be guaranteed to remain below 0.2 without increasing.

In addition, by comparing the time consumed in model training, on the basis of ensuring the same equipment and database, the four kinds of neural networks have their own unique operation process, so the time consumed can also reflect the real-time performance of the model. The results are shown in Table 1:

*Table 1: Time consumption.*

| Morels | The number of iterations and times | | | |
|---|---|---|---|---|
| | 5 | 10 | 15 | 20 |
| AlexNet | 113 | 256 | 373 | 464 |
| VGG16 | 95 | 190 | 303 | 401 |
| GoogLeNet | 155 | 352 | 503 | 677 |
| Resnet-101 | 160 | 317 | 473 | 633 |

Through the analysis of the accuracy, loss and time of different networks, it can be found that the accuracy and loss of VGG16 network after 20 iterations have certain advantages, and the consumption of time is relatively less, which indicates that VGGNet can save our computer resources and make the convergence rate of the model faster。

## 5. Conclusions

This paper discusses four kinds of road recognition models of convolutional neural networks, and analyzes the advantages of different neural networks by comparing the accuracy, loss value and consumption time of the four networks. The experimental results show that VGG16, RESNET-101, GoogLeNet and AlexNet neural networks can achieve the goal of road recognition, and the four trained models can be used in the deeper research of road recognition field. Through the experimental analysis and the matching of results, it can be seen that the four networks can obtain higher accuracy and lower loss value. Among them, compared with the other three groups of models, the accuracy rate of VGG16 is higher, the loss value is relatively lower, and the consumption of time is less, indicating that VGG16 has a good network performance in the field of image recognition. In view of the above discussion, this paper studies the advantages and disadvantages of each network model, which provides ideas for the research of road recognition technology in the intelligent vehicle industry and lays a foundation for the following research.

## References

*[1] Xiao Wangxin, Zhang Xue, Huang Wei, et al. A New Algorithm for Automatic Identification of Pavement Damage [J]. Highway Traffic Technology, 2005, 3(11): 79-82.*
*[2] Shi Weijia, Ma Shui, Ma Bin, et al. Design and implementation of pavement recognition system [J]. Automation instrumentation, 2011, 32(04): 62-65.*
*[3] Zhang Xiaolong, Sun Renyun, Liu Changwei, et al. Research on road surface recognition method based on road surface characteristic coefficient [J]. China test and measurement, 2015, 41(08): 31-35.*

ISSN 2616-5775 Vol. 5, Issue 11: 90-96, DOI: 10.25236/AJCIS.2022.051114

*[4] Gao Wei. Research on Road Recognition Technology of Mobile Robot [D]. Hebei: Hebei University of Technology, 2014.*

*[5] Jing Weipeng, Zhang Mingwei, Lin Jingbo. A neural network framework search method for remote sensing image classification [J]. Journal of Harbin university of science and technology, 2021, 26(01): 25-29.*

*[6] Lu Haoyuan, Yu Lu, Zhou Xingyu, et al. A survey of semi-supervised deep learning image classification methods [J]. Computer science and exploration, 2021, 15(03): 1038-1048.*

*[7]* Zhang Xiaoling. *Study on epilepsy prediction based on deep learning and scalp electroencephalogram [D]. Tianjin: Tianjin Vocational and Technical Normal University, 2022.*

*[8] Zhang Ke, Feng Xiaohan, Guo Yurong, et al. A survey of deep convolutional neural network models for image classification [J]. Journal of image and graphics, 2021, 26(10): 2305-2325.*

*[9] Li H, Chen D. Complexity and Accuracy Analysis of Convolutional Neural Networks in Scene Text Detection [C]. International Conference on Electronic Information Technology and Computer Engineering (EITCE), 2019: 416-1419.*

*[10] Li Y D, Hao Z B, Lei H, et al. A survey of Convolutional Neural Networks [J]. Journal of Computer Applications, 2016, 36(9): 2508-2515.*

*[11] 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.*

*[12] Wang Z, Zheng X, Li D, et al. A VGGNet-like approach for classifying and segmenting coal dust particles with overlapping regions [J]. Computers in Industry, 2021, 132: 103506.*

*[13] Rajan S, Poongodi C, Devaraj S, et al. Novel Deep Learning Model for Facial Expression Recognition based on Maximum Boosted CNN and LSTM [J]. IET Image Processing, 2020, 14(7): 1373-1381.*

*[14] Xie S, Zheng X, Chen Y, et al. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction [J]. Scientific Reports, 2018, 8(1): 1-9.*

*[15] Wen L, Li X, Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50 [J]. Neural Computing and Applications, 2020, 32(10): 6111-6124.*

*[16] Zhao Jin. Theory and Application of Sparse Deep Learning [D]. Xi 'an: Xidian University, 2019.*

*[17] Ren Jinjun, wang Ning. Research on loss function in artificial neural networks [J]. Journal of gansu normal university, 2018, 23(02): 61-63.*