International Core Journal of Engineering
ISSN: 2414-1895

Volume 9 Issue 9, 2023
DOI: 10.6919/ICJE.202309_9(9).0022

A Transfer Learning Algorithm for Terrain Type Recognition

Shuoyi Wen*, Li Yang, Hailong Duan, and Tingting Zhang

Tianjin University of Technology and Education, School of Automation and Electrical Engineering, Tianjin 300222, China

*1315282586@qq.com

Abstract

In order to achieve fast and accurate terrain type classification, a combination of migration learning and convolutional neural network based terrain type type image recognition method is proposed. First, after pre-processing the terrain type images by translation, rotation and scaling, the terrain type images are manually classified into five categories according to their characteristics: asphalt, mud, gravel, grass and snow and ice. Based on the migration learning method, the extraction layer of features is the convolutional layer and pooling layer of vgg16 for optimizing and improving the top layer design. The problems of model overfitting and insufficient sample size are solved, and the training time is greatly reduced by fine-tuning the network parameters, and the average recognition accuracy can reach about 98%.

Keywords

Deep Learning; Migration Learning; Convolutional Neural Network; Image Classification.

1. Introduction

In the era of progressive development of smart vehicles, environment sensing technology serves as the basis for vehicle decision making and route planning. As people's living standards continue to improve, the number of automobiles has increased significantly. While cars bring comfort to people, they also cause many traffic accidents, which in turn generate huge economic losses.[1] In recent years, automotive intelligent driving assistance technology is booming, with the help of active full control technology, advanced intelligent driving assistance system (ADAS) not only improves the comfort of the driver's ride, but also greatly enhances the safety performance of driving. Affected by the weather, accurate recognition of road images is particularly important. The vehicle active safety control system can pre-adjust the system threshold and alert the driver's maneuvering, which can effectively improve the safety, smoothness and comfort of vehicle driving without increasing the hardware cost.

Current terrain type classification methods can be divided into two categories: traditional machine learning-based classification methods and deep learning (CNN)-based classification methods. Traditional machine learning methods are usually based on manually extracted features. For example, commonly used image features include visual features such as shape, texture, color, and transformations of scale-invariant features, local binary patterns, histograms, traditional classifiers, vector machine support, k-nearest neighbors, and other classifiers to accomplish the classification goal. The effect is self-explanatory for complex images that are difficult to identify accurately. With the deepening of learning image recognition based on Convo-lutional Neural Network (CNN) is one of the hot spots of current research. With the deepening of science, Convolutional Neural Network (CNN) based image recognition is one of the hot spots of current research. Convolutional Neural Network (CNN) has richer function expressions because it contains a deeply hidden hierarchical network structure and does not require manual extraction of functions. It automatically learns image features from sample images through forward transmission and backward propagation processes, and extracts more accurate, higher dimensional and abstract features through the deep network structure. Finally, these features are fed into the classifier to achieve good classification results[2].

In this paper, a study is done for the recognition of five kinds of terrain type pictures. A convolutional neural network based learning method is proposed, a database creation, data augmentation method, network model using vgg16 for improvement and migration learning is performed, and the recognition rate reaches 98%.

2. Principle and Method

2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a deep learning model or a multi-layer perceptron similar to artificial neural networks. In recent years, Convolutional Neural Networks are used in assistive technologies, medical applications, autonomous driving, etc., which largely accelerate the rapid development and popularity of deep learning.

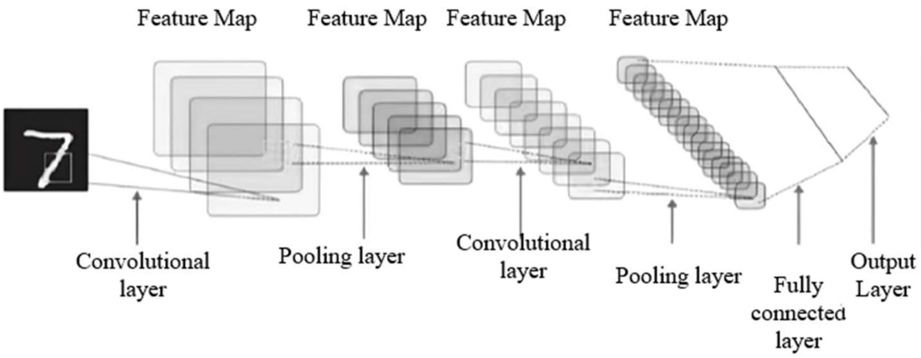


Figure 1. Convolutional neural network structure

As shown in Figure 1, a convolutional neural network generally consists of a convolutional layer, a pooling layer, a fully connected layer, and a Softmax layer. In image recognition, the input is a picture, a three-dimensional matrix to extract features through the convolutional layer, pooling layer to reduce feature complexity, and finally a fully connected layer to connect the feature values of the last convolutional layer or pooling layer into the Softmax classifier to calculate the probability of each class to complete the image classification. After the convolutional operation of the first convolutional layer of the convolutional neural network, the feature map obtained from the convolutional layer is the image edge information, line contours and other shallow features. For image recognition, deep layer features are needed. The shallow information is not enough to express. So to obtain deeper features, it is necessary to perform multi-layer convolution kernel for feature information extraction to form a feature map with multiple information[3]. In the field of image recognition, the feature hierarchy of the input image itself has. As shown in Figure 2, it goes from original pixels to forming lines from pixels, then from lines to patterns, and finally forming objects from patterns. The whole process finds the shallowest feature through the original pixel, and then explores the middle feature in it for the shallow feature, and after that obtains the deep feature. Therefore, it is not possible to go directly from the original to the deeper layers, so it is possible to obtain deeper features by increasing the number of layers of convolution.[4]Therefore, it is possible to obtain deeper features by increasing the number of layers of convolution.

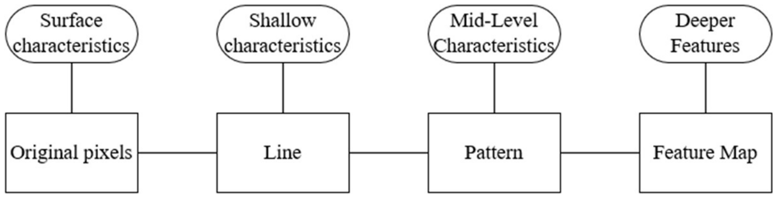


Figure 2. Schematic diagram of feature extraction process

2.2 Transfer Learning

The role of transfer learning is mainly to solve the small data problem. The second is to solve personalized problems. Since the 1990s, transfer learning has received increasing attention as a way to solve problems faster and more effectively by applying what has been learned in other areas to new problems.[5]The problem can be solved faster and more effectively by applying what has been learned in other areas to new problems. Traditional data training is done from scratch, which can be time-consuming. The ideal scenario for training is to have a large amount of labeled data, but in reality collecting large amounts of data is time consuming and even impractical.[6] In contrast, migration learning can transfer the network weights obtained from other similar data sets to the target network without having to train from scratch.[7] The problem of less data can be solved to some extent. According to Wikipedia, transfer learning is a new machine learning method that uses existing knowledge to solve problems in different but related domains[8]. The main idea is to train a convolutional neural network on an existing large data set, then migrate the pre-trained weights to the model to be used and fine-tune it.[9] . The use of migration learning can produce good results even with small sample data and can prevent the network from overfitting.

2.3 Vgg16 Network

The commonly used convolutional neural networks are LeNet[10], Alexnet[11], VGG[12][13] and GoogleNet[14] etc. In general, to obtain better performance, training convolutional neural networks requires large amounts of data and training time. In general, training convolutional neural networks requires a large amount of data and training time in order to obtain better performance. VGGNet is a model proposed by the Visual Geometry Group at the University of Oxford, which achieved excellent results in the ImageNet Image Classification and Localization Challenge ILSVRC-2014 by placing second in the classification task and first in the localization task. The outstanding contribution of VGGNet is to demonstrate that a small convolution can effectively improve the performance by increasing the network depth. That is, the network is deeper. As shown in Figure 3, VGG16 has 16 layers, including 13 convolutional layers and 3 fully connected layers, through repeated 3×3 convolutional kernels and 2×2 maximum pooling layers. After three convolutional layers with 256 kernels, a maximum pooling layer is connected, three convolutional kernels with 512 kernels are convolved twice and then a maximum pooling operation is performed, and finally the extracted features are passed into the three fully connected layers.

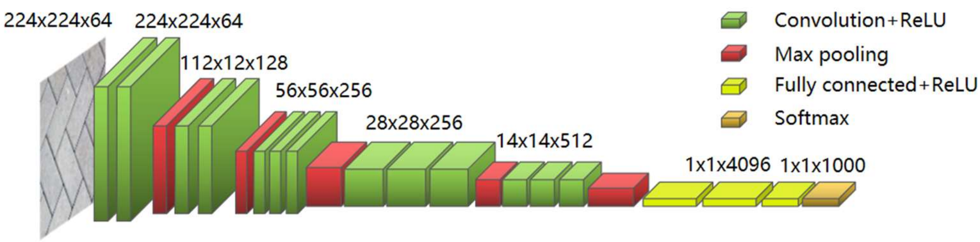


Figure 3. VGG16 network structure diagram

2.4 Loss Function and Optimization Algorithm of the Model

The aim of this paper is to implement the task of terrain type species classification using a cross entropy loss function and adding a regularization term using the L2 parametrization, which is calculated as follows:

$$\mathrm{L(w)} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{y_i} \cdot \widehat{\log \mathrm{y_i}} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \cdot \sum_{\mathrm{k=1}}^{\mathrm{N}} \mathrm{y_i} \cdot \widehat{\log \mathrm{y}}_{\mathrm{ik}} \tag{1}$$

$$\mathrm{L(W)_{str}} = \mathrm{L(W)} + \partial \|\mathrm{w}\|^2 \tag{2}$$

Adam is a first-order optimization algorithm that can replace the traditional stochastic gradient descent (SGD) process by iteratively updating, based on training dataNeural Networks weights. It has the advantages of efficient computation; less memory required; invariance of the diagonal scaling of the gradient; suitability for solving optimization problems with large-scale data and parameters; and applicability to non-stationary (non-stationary) objectives.

$$\mathrm{m_t} = \mu * \mathrm{m_{t-1}} + (1-\mu) * \mathrm{g_t}$$
$$\mathrm{n_t} = \mu * \mathrm{n_{t-1}} + (1-\nu) * \mathrm{g_t}^2$$
$$\widehat{m}_t = \frac{m_t}{1-\mu^t}$$
$$\widehat{n}_t = \frac{n_t}{1-\nu^t}$$
$$\Delta\theta_t = \frac{\widehat{m}_t}{\sqrt{\widehat{n}_t + \theta}} * \eta \tag{3}$$

3. Experimental Section

3.1 Data Sources

In this paper, images of five more common terrain type types, namely grass, mud, asphalt, ice and snow, and gravel, are used as the main research objects, and multiple road sections of each corresponding terrain type type are collected to ensure the relative richness and perfection of the data. A total of five types of images based on online selection and field shooting are used, and a 9:1 training set 5067 test set 563. contains asphalt terrain type, mud terrain type, gravel terrain type, grass terrain type, and snow and ice terrain type.

(a) Asphalt terrain type

(b) Dirt terrain type

(c) Gravel terrain type

(c) Grass terrain type

(d) Snow and ice terrain type

Figure 4. Data set

3.2 Data Pre-processing

Pre-processing of the images, in order to improve the image quality, the images are first cleaned to remove the damaged and problematic images, and the road images are expanded in the second step. Data set enhancement makes transformation of training images to improve sample quality, which can get a network with stronger generalization ability and better adaptability to apply to different scenarios. Common data enhancement methods: random cropping, random inversion, random contrast enhancement, color change, etc. The third step of adding random Gaussian noise to the images can make the trained model more robust and help to improve the performance of the model. Appropriate addition of noise is helpful for image generalization ability. The so-called Gaussian noise is a class of noise whose probability density function obeys a Gaussian distribution (i.e., normal distribution). If a noise. Its amplitude distribution obeys the Gaussian distribution, and its power spectral density is uniformly distributed, it is called Gaussian white noise. Gaussian white noise includes thermal noise and scattered particle noise.

$$p(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(s-u)^2}/2\sigma^2 \tag{4}$$

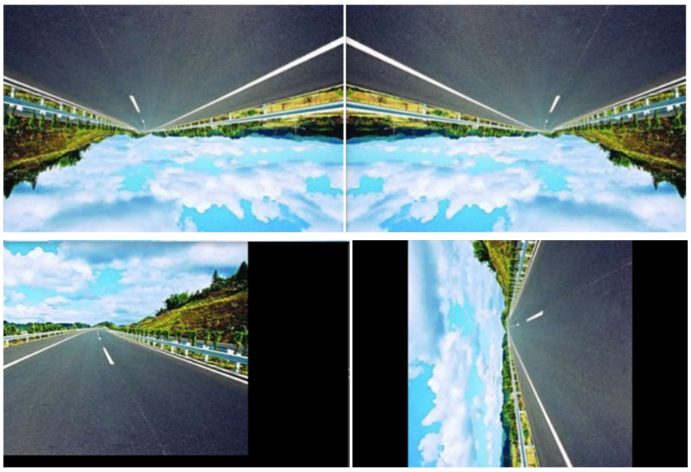


Figure 5. Pre-processing

3.3 Software Environment

The operating system is Windows 10, the programming language is python, and the open source framework is Tensorflow.

3.4 Model Construction

The pre-trained network model is on the ImageNet dataset.[15] The VGG16 network is trained using hundreds of thousands of images with 1000 categories, but this paper is performing 5 classifications without the complex fully connected layers of the 1000 classification network. The fully connected layer is removed in the last pooling layer, replaced with a new fully connected layer designed in turn, and then fed into the fully connected layer to integrate features, and the output of the last fully connected layer is fed into the somfmax classifier to calculate the corresponding probability value of each label for classification.

In this experiment, the parameters obtained from the training of VGG16 completed on ImageNet are migrated to the terrain type recognition modelto achieve the classification task for the water terrain type type dataset. Fine-tuning some of the network layers allows the model's partial convolutional layers to participate in training simultaneously with the top layer, which refers to several convolutional layers near the top layer.

First remove the fully connected layer behind VGG16; then add two fully connected layers with 512 and 5 nodes after the model to classify the custom dataset by 5. The global average pooling layer is

 Volume 9 Issue 9, 2023
DOI: 10.6919/ICJE.202309_9(9).0022

added so that the network parameters are reduced to avoid the occurrence of overfitting. The activation function is selected as Rule. Dropout regularization is added after the activation function, and the rounding probability value is set to 0.5, half of the parameters are discarded as they are trained to reduce the impact of parameter changes on the output results. The BN layer is built to batch normalize the irregular data with Batch Normalization technique to normalize the network output, reduce the model overfitting, make the model converge faster, and further improve the model performance. Retain block1-block5 of VGG16 for redesigning the fully connected layer module. The terrain type prediction category is obtained. The flow chart of the algorithm is shown in Figure 6.

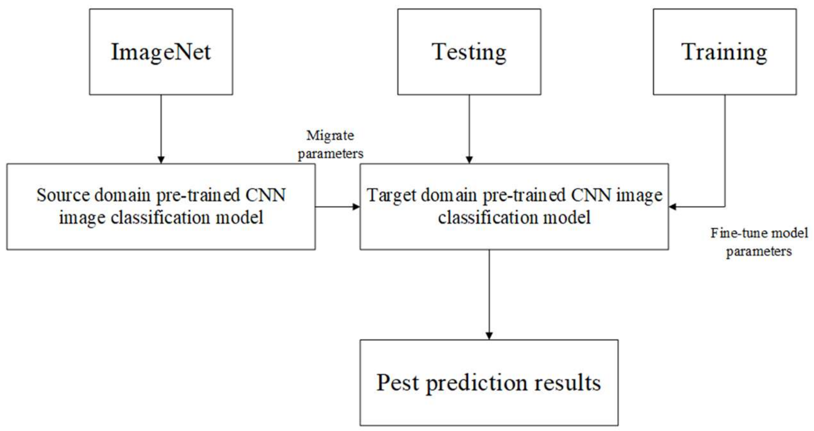


Figure 6. Algorithm flow chart

3.5 Comparison of Different Improvement Options

Experiment 1: Network selection comparison, in vgg16 and AlexNet.

Table 1. Experimental comparison of different network schemes

Models	Training set accuracy	Training set loss rate	Test Set Accuracy	Test set loss rate
Vgg16	92.7%	0.180	89.2%	0.226
AlexNet	85.9%	0.388	89.7%	0.288

After vgg16 network and AlexNet network, training, testing and loss comparison, it is obvious that vgg16 is better than AlexNet, so next, vgg16 is used as the experimental model to improve vgg16.

Experiment 2: Design 4 scenarios based on improved vgg16 for comparative analysis.

Table 2. Experimental comparison of different top-level design schemes

Programs	Global average pooling layer	Number of neurons	softmax output layer
1	without	64	Use
2	without	256	Use
3	without	512	Use
4	Join	512	Use

The experimental results obtained for the four top-level design scenarios are shown in Table 3.

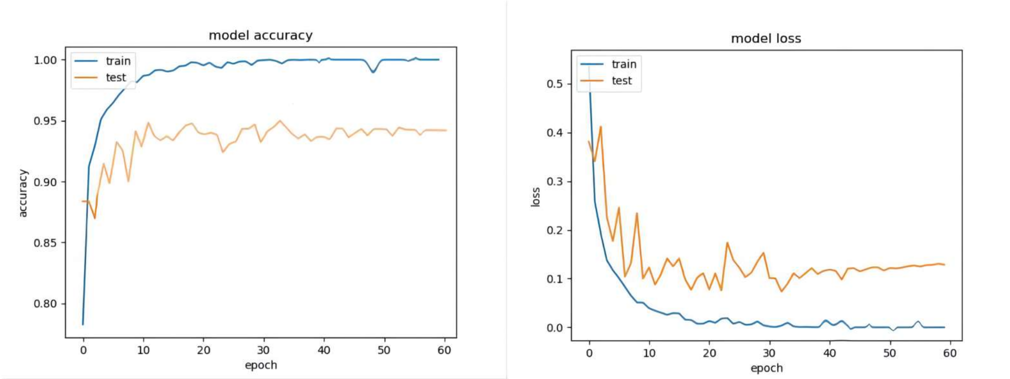
Table 3. Comparison of experimental results of different top-level design schemes

Programs	Global average pooling layer	Number of neurons	softmax output layer
1	without	64	Use
2	without	256	Use
3	without	512	Use
4	Join	512	Use

The comparison of the global average pooling layer, the number of neurons in the dense layer, is shown in Table 2. From Table 3, the results of the four different top layer designs are compared in terms of accuracy, loss rate, and model size, respectively. The loss rate is the difference between the true value and the predicted value, and the smaller the loss indicates the better the robustness of this model.

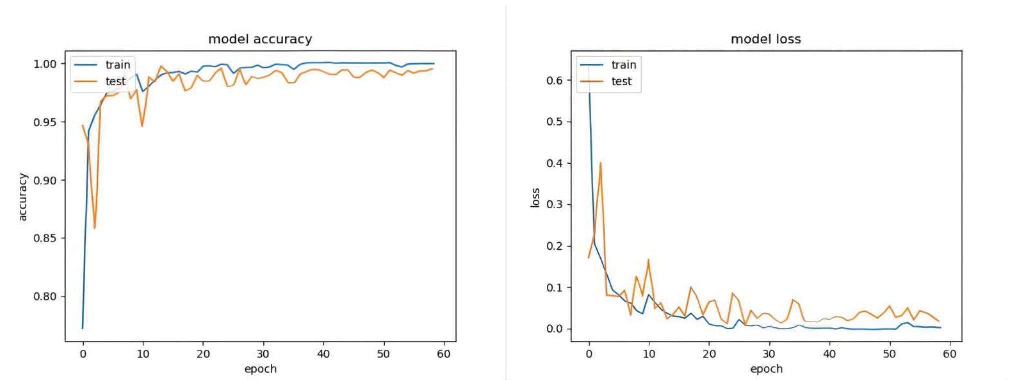
Scheme I, Scheme II and Scheme III are comparisons of the number of neurons in the dense layer. As the number of neurons in the fully connected layer increases, Scheme II and Experiment III outperform Scheme I in terms of accuracy and loss, but the size of the generated model also becomes larger, with Scheme III having the largest model. Scheme IV differs from the first three by adding a global average pooling layer, which produces a significant effect on the model size and an increase in accuracy and loss.

Comprehensive analysis, the top layer is designed as a softmax output layer, a dense layer with 512 neurons, and the addition of a global average pooling layer can obtain a more desirable effect in the combination of shift learning and VGG16, and scheme four is set as the final improvement scheme for this experiment.



(a) Accuracy curve (b) Loss curve

Figure 7. Recognition effect of VGG16 based on new learning



(a) Accuracy curve (b) Loss curve

Figure 8. Transfer learning vgg16

The final result plots of the original method and the improved scheme run on the terrain type dataset are shown in Figure 7and Figure 8 respectively. a shows the accuracy curve and b shows the loss value curve with 60 iterations.

As can be seen from the figure, the accuracy and loss rate curves of and of the traditional VGG16 based on brand-new learning start to show a convergence trend at about 20 iterations, but the loss rate curve shows a local abrupt change, showing the instability of the model, and the accuracy and loss rate curves of Figure 5 after convergence show a large distance from the corresponding training curve, indicating that the traditional VGG16 model based on brand-new learning does not fit well on the roadway dataset is not very well fitted, and the phenomenon of underfitting appears.

The accuracy and loss rate curves of migration learning combined with VGG16 start to converge around 20 times, which is better than the original model. The accuracy and loss rate curves after convergence are close to the corresponding training curves, indicating that the model fits the terrain type dataset well, and it can also be seen from the figure that the accuracy and loss rate curves of migration learning VGG16 introduced in the late training period oscillate less compared to the original model amplitude becomes smaller, indicating that the model is more stable.

The specific values in terms of accuracy, loss value, and model size are compared, and as shown in Table 3, the introduction of migration learning and fine-tuning the neural network can significantly improve the network performance compared to the original method, where the accuracy is 98.44% on the terrain type dataset, the loss value is 0.05, and the model size is 85.2 MB, which is much smaller than the original method.

4. Conclusion

The training time of the model in this paper is smaller than the training time required by the VGG16 network model, and the recognition time of a single image only takes 0.347s, which meets the practical use requirements. The experimental results prove that the results are good, the generalization ability of the model is strong, and the recognition accuracy is high, which can reach about 98%. However, there is still room for improvement, such as too few classification categories and not enough complex classification backgrounds. In the subsequent work, we will continue to do optimization and improvement along the direction of these potential problems.

References

[1] Liao Luming,Zhang Wei. Hybrid batch training traffic sign recognition based on improved VGG16 network[J]. Electronic Science and Technology,2021,34(8):8-13. doi:10.16180/j.cnki.issn1007-7820. 2021.08.002.

[2] Ke Yie,Lu Yaling. Pet dog identification method based on VGG16[J]. Electronic Production, 2020(21):4.

[3] Wang JH. Research on a semantic web-based recommendation system for agricultural learning resources[J]. Computer Application and Software,2013,30(08):233-235+257.

[4] Fan, Chun-Quan, He, Bin-Bin. Migration learning-based rice pest identification[J]. China Agricultural Information, 2020, 32(2):9.

[5] PAN S J Q Y. A Survey on Transfer Learning [J].IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10) : 1345-1359.

[6] ZHUANG F Q Z, DUAN K, ET AL. A Comprehensive Survey on Transfer Learning [J].Proceedings of the IEEE, 2020, pp(99) : 1-34.

[7] Wu Biqiao, Xing Yongxin, Wang Tianyi. High-resolution palm print image recognition based on VGG16 and migration learning [J]. Intelligent Computers and Applications, 2021, 11(05) : 37-42.

[8] Zhuang Fuzhen, Luo Ping, He Qing, et al. Advances in migration learning research [J]. Journal of Software, 2015, 26(01) : 26-39.

[9] Zhang Fangyan, Zhao Meng, Zhou Yizhi, et al. A method for detecting diseased fish of redfin oriental triggerfish based on Res Net50 and migration learning [J]. Fishery Modernization, 2021, 04) : 1-10.

[10] Wu Z, Chen H, Lei Y. Recognizing Non-Collaborative Radio Station Communication Behaviors Using an Ameliorated LeNet[J]. Sensors, 2020, 20(15): 4320.

[11] He M, Zhao X, Lu Y, et al. An improved AlexNet model for automated skeletal maturity assessment using hand X-ray images[J]. Future Generation Computer Systems, 2021, 121: 106-113.

[12] Qiu J, Lu X, Wang X, et al. Research on Rice Disease Identification Model Based on Migration Learning in VGG Network[C]//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2021, 680(1): 012087.

[13] Ding Yng J J, Li M Z. Identification of lily diseases based on convolutional capsule network [J].

[14] Huang Xuehua,Chen Weihong,Yang Wangdong. Improved Algorithm Based on The Deep Integration of Googlenet and Residual Neural Network[J]. Journal of Physics: Conference Series,2021,1757(1).Moon J, Hossain M B, Chon K H. AR and ARMA model order selection for time-series modeling with ImageNet classification[J]. Signal Processing, 2021, 183: 108026.

[15] Xiao Y, Huang X, Liu K. Model Transferability from ImageNet to Lithography Hotspot Detection[J]. Journal of Electronic Testing, 2021, 37(1): 141-149.