ISSN 2320 - 2602

Volume 11 No. 8, August 2022

International Journal of Advances in Computer Science and Technology

Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst011182022.pdf

https://doi.org/10.30534/ijacst/2022/011182022

Multi-layer high-precision image classification technology embedded in SE modules

Xuejia GAO[1], Gejian DING[2]

[1]Zhejiang Normal University, China,1197505621@qq.com

[2]Zhejiang Normal University, China, dgj@zjnu.cn

Received Date : July 06, 2022 Accepted Date : July 29, 2022 Published Date : August 07, 2022

ABSTRACT

Due to the problems of model overfitting and gradient changes to reduce model performance in deep networks, the operation of improving the accuracy of image classification models by superimposing the number of layers of the network cannot be applied to all models. The Squeeze-and-Excitation (SE) module is a plug-and-play attention module in the field of computer vision that focuses on channel relationships. Experiments show that embedding SE modules in ResNet models of different scales brings much higher test accuracy improvement than increasing the depth of the original model; SE modules are extremely generalizable, and their embedding is universal to greatly improve the accuracy of different original models. Experimental results on the CIFAR-10 and Dogs-vs-Cats datasets show that the larger the amount of data, the more it can avoid the overfitting phenomenon of the model. A comparison experiment with the GoogLeNet model showed SENet being superior. According to the published research data, the application of SE modules accounts for 57.59% of the top 30 disciplines such as medical health, automation technology, telecommunications technology, electric power, light industry, automobiles, and transportation.

Key words : squeeze-and-excitation module; attention module; image classification; deep learning.

1. INTRODUCTION

Image classification is the product of the era of artificial intelligence, is the basic task of image recognition, behavior detection and other computer vision fields, has a high academic research value, its task is to complete the classification of images with the smallest error. In recent years, a large number of studies have shown that deep learning models have significant classification capabilities, which can integrate the three modules of feature extraction, feature screening, and result classification in image classification tasks, accelerate the classification of images, and are suitable for solving problems in the field of computer vision, which is better than traditional image classification methods [1]. It is no exaggeration to say that the development history of deep learning models is the evolutionary history of image classification tasks.

Squeeze-and-Excitation Networks (SENet) [2] focuses on channel relationships, and with its novel SE architecture, it won the Image Classification Mission for its outstanding performance in the last ImageNet competition in the field of computer vision[3].

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is one of the most sought-after and authoritative academic competitions in the field of computer vision in recent years, representing the highest level in the field of imagery. It was held from 2010 until the last edition in 2017, and every year a different network architecture won the championship with its unique advantages. For example, the GoogLeNet[4] network structure embeds multi-scale information in the Inception architecture, the Inside-Outside Net[5] network structure combines contextual information in space, and innovations such as integrating the Attention mechanism into the spatial dimension to obtain performance gains have all promoted the development of deep convolutional neural networks in the field of images [6]-[8]. SeNet, the winner of the last competition, put aside the fact that CNN processed the feature layer spatially, extracted the feature layer for the first time in the channel dimension, proposed a plug-and-play attention module with simple structure but strong functions: SE Module, and defined a strategy called "feature recalibration", which achieved great success in the image classification task. Since then, a simple plug-and-play attention module has taken the stage of deep learning and also caught the eyes of image classification researchers.

2. PRINCIPLE OF SE MODULE

SENet, which focuses on channel relations, has an architectural unit named "squeeze and excitation" (SE) block, the SE architecture is not a network structure that can be directly applied experimentally, it is a substructure that can be directly embedded in existing models and used in combination. The authors[2] embedded it in the ResNeXt[9] model, and used the combination to get a good score of 2.251% in the Image Classification task, hitting a 2.991% lower than the original best score, proving its strength.

Se modules are named in two extremely important operations, Squeeze and Exposition. The "feature rescale" strategy in the SE module is to realize the information fusion between the feature channels in the spatial dimension, that is, to obtain the importance of each feature channel first, and guide the machine to focus on the main features affecting the current task according to the level of the importance value, and accelerate the promotion of the main task; Ignore features that have little impact on the current task and save more resources. This is called the attention mechanism, focusing on where it needs more attention, so the SE module is also called a plug-and-play attention module.

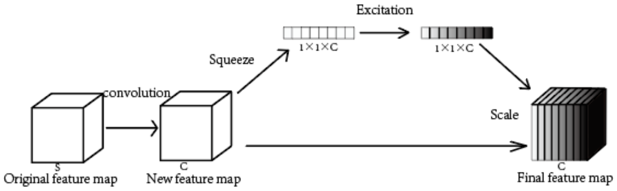


Figure 1: An extrusion and excitation (SE) block

Figure 1 is an extrusion and excitation (SE) block, known as the SE architecture schematic. Enter a raw feature map, a new feature map will be generated after convolution, and then an SE schema will begin. The SE architecture is to add a bypass branch to the normal flow direction, and the bypass branch mainly includes three parts: Squeeze, Excitation, and Reweight.

2.1 SQUEEZE

The Squeeze operation is a compression module that compresses features based on each binary channel according to the spatial dimension of the input image. Its main operation is to pool the characteristic channels of the current input feature map globally and compress them into a globally representative real number, which solves the problem of using channel dependence in the processing image classification task.

2.2 EXCITATION

Excitation is an excitation branch with two fully connected layers that is used to learn the importance of each channel. It is very similar to the mechanism of gates in recurrent neural networks, mainly using parameters to generate weights for each feature channel for the input image. It not only leverages the information aggregated in the Squeeze operation, but also captures the channel correlations needed to process image classification tasks.

2.3 REWEIGHT

Reweight is a rescale of the scale branch. Its main operation is to treat the weights of the output of the Exposition operation as a numerical description of the importance of each feature channel, and then use multiplication to weight the original feature map of the previous input channel by channel, and rescale the original feature map on the channel dimension.

Channel relationships for convolutional modeling are implicit and local in nature, and the SE module enhances the learning of convolutional features by explicitly modeling channel interdependencies so that the network can improve its sensitivity to information features and is therefore suitable for image classification tasks.

In order to prove that the excellent performance of embedding the SE module is not limited to the ImageNet dataset[10] and ResNeXt used in the competition, two additional datasets of CIFAR-10[11] and Dogs-vs-Cats are selected in this paper and the SE module is embedded in the ResNet[12] model for deep network experiments. The image classification task evaluation standard in this paper adopts the top5 error rate, that is, the answer of five categories is predicted for each test picture, as long as one answer is the same as the manually labeled category.

3. SOME COMMON MISTAKES

3.1 EXPERIMENTAL ANALYSIS BASED ON CIFAR-10 DATASET

3.1.1 INTRODUCTION TO THE CIFAR-10 DATASET

CIFAR-10 is a small dataset that is currently used in the field of deep learning images and is suitable for identifying general objects. The dataset includes a total of 3-channel color RGB images of ten animal species, including airplane, bird, cat, deer, dog, car, truck, ship, horse, frog, etc. The dataset has a fixed size and does not require pre-processing, with a total of 60,000 images in the dataset, 50,000 for training, and 10,000 for testing.

3.1.2 CIFAR-10 DATASET RESULTS

In this paper, the SE module is embedded in resNet models of different scales, and the accuracy comparison experiment of SEResNet (SEResNet20, SEResNet56, SEResNet110) models of different scales is implemented on the CIFAR-10 test set to deal with image classification problems. Table 1 shows the top5 test error rates on the CIFAR-10 test set for the SEResNet model with network depths of 20, 56, and 110, respectively.

Table 1: Top5 error rates for SEResNet models with different network depths

SEResNet model	Top5 error rate
SEResNet20	7.37%
SEResNet56	6.77%
SEResNet110	5.33%

As the network depth increases and the test error rate decreases, it is possible to improve the accuracy of the model test by increasing the network depth within a certain range, and in recent years, the operation of improving the accuracy of the image classification task model by superimposing the number of network layers has been widely used, and ResNet [12] is a very typical deep neural network training model. However, the classic deep neural network of more than 30 layers is only the ResNet series, VGGNet [13] the deepest 19 layers, GoogLeNet [4] the deepest 22 layers, MobileNet [14] the deepest 28 layers, which shows that the network is not the deeper the better, the deep network brought by the model overfitting, gradient changes and other issues are always there, so when the network deepens to a certain extent, the performance does not rise and fall.

Compared with ResNet101 and ResNet152, which have a 7.1% error rate in the top5 test error rate in Table 2, and ResNet152, which have a test error rate of 6.7%, the embedding of the SE Module makes the test accuracy of SEResNet56 greatly exceed the test accuracy of model ResNet101, and even close to the test accuracy of model ResNet152, and the test accuracy of SEResNet110 also far exceeds the test accuracy of ResNet152. Experimental results show that embedding SE module can not only bring great performance improvement to the original network model at different depths, but also bring about a much higher test accuracy improvement than the accuracy improvement brought about by increasing the depth of the original model, and the excellent performance of the SE module is not only manifested in the ImageNet dataset and ResNeXt model, it is extremely generalizable. Therefore, embedded SE modules can be used to replace excessive network depth, avoid network degradation, and provide direction for researchers of subsequent related problems.

Table 2: Top5 error rates for ResNet models with different network depths [2]

ResNet model	Top5 error rate
ResNet50	7.8%
ResNet101	7.1%
ResNet152	6.7%

Figure 2 is a diagram of the loss function of the SEResNet56 model during training and testing on the CIFAR-10 dataset. The smoother blue curve represents the loss function on the training set, the more jittery orange curve represents the training function on the test set, the horizontal axis is the number of trainings, the unit is epoch, and one epoch is equivalent to all the samples in the training set participating in the training once. The training set curve shows that when the model reduces the learning rate at the 150th and 225th epochs, the model performance is improved, and the loss function is significantly reduced, but the loss function of the test set curve does not fall in the 225th epoch, and the model may appear slightly overfitting.

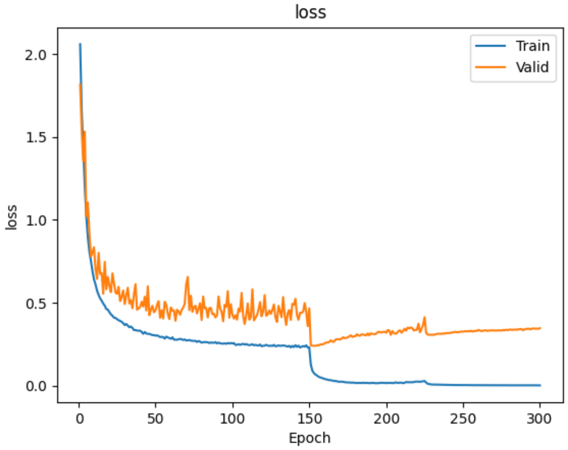


Figure 2: Graph of the loss function of SEResNet56

3.2 EXPERIMENTAL ANALYSIS BASED ON DOGS-VS-CATS DATASETS

3.2.1 INTRODUCTION TO THE DOGS-VS-CATS DATASET

Since there are many types of objects in the CIFAR-10 dataset and fewer pictures of the same kind, this paper uses the Dogs-vs-Cats dataset with a large data volume to conduct a comparative experiment.

The Dogs-vs-Cats dataset is a large cat and dog dataset used for picture recognition tasks for cats and dogs. The dataset has a total of 25,000 pictures, including cats and dogs, and 12,500 cats and dogs. The size of the picture is uncertain, the picture preprocessing is required before use, the number of pictures of cats and dogs in the training set is equal, 12500 pictures each and stored in order, and a total of 12500 pictures of cats and dogs are mixed and disordered in the test set. Due to the small variety of pictures and the abundance of data samples, this dataset is suitable for computer vision classification experiments.

3.2.2 DOGS-VS-CATS DATASET RESULTS

Figure 3 is an accuracy graph of the 56-scale SEResNet model tested in the Dogs-vs-Cats dataset, and the results show that when the model reduces the learning rate at 150 epochs, the performance of the model is improved, the accuracy has increased significantly and tends to be stable, and finally it can reach 94%. Figure 4 is a loss function diagram of the testing process of the SEResNet model with a scale of 56 in the Dogs-vs-Cats dataset, and comparing the test results of the SEResNet model of the same size on the CIFAR-10 test set, it is clear that on the Dogas-vs-Cats dataset with a larger amount of data, the model has less loss, higher accuracy, and finally approaches 0.2, which can have better performance.

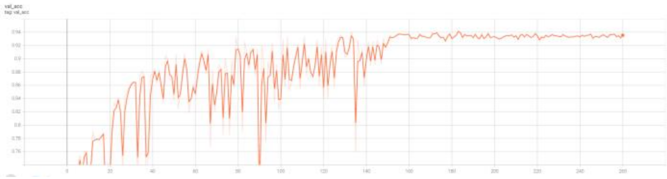


Figure 3: Test set accuracy graph for Dogs-vs-Cats-SEResNet56

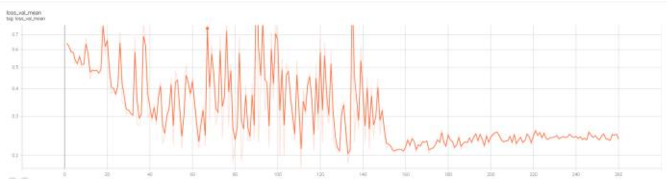
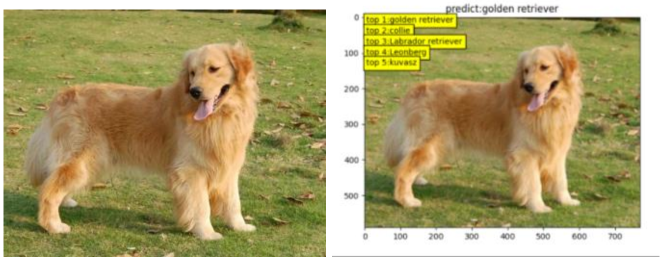


Figure 4: Plot of the test set loss function for Dogs-vs-Cats-SEResNet56

3.2.3 IMAGES CLASSIFICATION TEST FOR THE DOGS-VS-CATS DATASET

In this paper, an image classification test is performed using a 56-layer SEResNet model on the Dogs-vs-Cat dataset, and Figure 5 is the test result. Figure 5 (a) is the original test figure, Figure 5 (b) is the test result graph, the five yellow labels in the upper left corner of the test result graph show the first five answers with the highest predicted result values, and the first shown gold retriever is the golden retriever, which can be seen that the accuracy of the prediction is quite high.



(a). Test the original image (b). Test result graph

Figure 5: SEResNet56 image classification test original and result diagrams

3.3 EXPERIMENTAL ANALYSIS COMPARED WITH THE GOOGLENET MODEL

The GoogLeNet network model is a new deep learning network model proposed by Christian Szegedy in 2014. Before 2014, AlexNet [15], VGG and other networks are by superimposing the number of network layers to improve the accuracy of the model, it turns out that the increase in the number of network layers can indeed bring better experimental results, but the number of layers of the network is not the more the better, excessive increase will reduce the accuracy of the model, such as the occurrence of model overfitting, gradient disappearance, gradient explosion and so on. The solution to these fatal problems can be to increase the number of network layers while reducing the parameters of the model, the same period of scholars will replace the full connection with a sparse connection to try, but the conclusion is that the full connection into a sparse connection after the amount of operation will not be greatly reduced, because most of the current hardware equipment are based on the dense matrix to use optimization operations, although the sparse matrix contains less data, but in the context of no optimization calculations, the time spent is still long. In this context, the GoogLeNet network proposed by Christian Szegedy's team can overcome this problem very well, the network replaces the full connection with a sparse connection, and constructs the Inception network structure on the basis of using the dense matrix high computing performance, that is, the structure called "base neuron" is proposed, and a sparse, high-performance network structure is constructed. The model won the ImageNet Challenge with a 6.7% error rate that year and is a very good high-performance image classification model.

In order to verify the excellent performance of the SE module, this paper performs a comparison experiment on the CIFAR-10 dataset with the GoogLeNet model that solves the shortcomings of deep networks. Figure 6 shows the accuracy of the 20-scale GoogLeNet model tested in the CIFAR-10 dataset, and the results show that the accuracy of the GoogLeNet model is significantly more than 85%, but it is still some distance from 90%, which is lower than the SENet model's nearly 94% accuracy. Figure 7 shows a loss function plot of the 20-scale GoogLeNet model tested on the CIFAR-10 dataset, and the results show that the loss function of the GoogLeNet model is significantly more than 0.6, which is higher than the loss value of the SENet model nearly 0.2. Obviously, the SENet model is superior in dealing with image classification tasks.

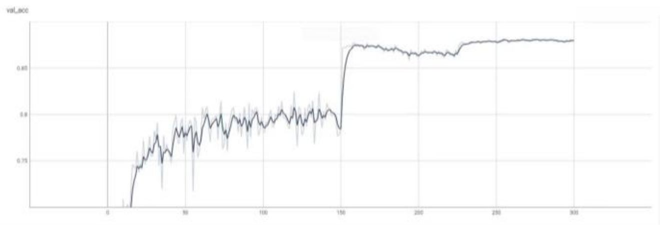


Figure 6: Test set accuracy graph for CIFAR-10-GoogLeNet20

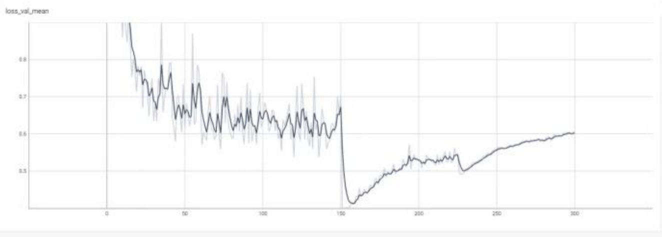


Figure 7: Plot of the test set loss function for CIFAR-10-GoogLeNet20

4. MULTI-LAYER HIGH-PRECISION IMAGE CLASSIFICATION TECHNOLOGY FEATURES EMBEDDED IN SE MODULE

By explicitly modeling the dependencies between each channel, the SE module adaptively recalibrates the feature response in each channel direction at the expense of a slightly small computational cost, resulting in extremely high performance improvements for the most advanced convolutional neural networks available today. Also affectionately known as "a simple plug-and-play attention module", its features are summarized as follows:

First, it forms the basis for the ILSVRC 2017 classification submission. SENets stacked by SE schemas have better results on multiple datasets. It won the Image Classification competition with a good score of 2.251% top5 error rate that year, and in the last authoritative evaluation imagenet competition in the field of computer vision in 2017, it won the image classification competition, which is about a 25% relative improvement compared to the previous year's winner, Trimps-Soushen with a 2.991% error.

Secondly, it has a simple structure, strong portability and high generalization. SE Module is now the most classic and commonly used attention module, whether it is image classification, image recognition, behavior detection and other computer vision tasks, as long as the SE module is added, the effect is improved. The development and application of new convolutional neural network architectures is a long-term and arduous engineering task, which requires repeated testing of different hyperparameters and layer configuration parameters to obtain optimal results. In contrast, the SE module structure is simple, high compatibility, can be seamlessly embedded in the mainstream network structure, integrated with other architectures, so as to effectively improve the performance of the original model, the current other attention mechanism is an improvement on SENet. Its high generalizability means that it can be directly applied to a wider range of architectures and data sets, and tasks of different scales and original models can be safely selected.

In the end, it lifts a lot and comes at a small cost. SE architecture in the calculation is a lightweight level, it through the weighted calculation of each channel, highlight the key information, inhibit the omission of invalid information, although it will slightly increase the complexity of the model and computation, but it brings a great improvement in performance, experiments have proved that the test accuracy of embedded SE module is much higher than the accuracy improvement brought about by increasing the depth of the original model, the technology provides a guide for the optimization of the image classification algorithm.

5. TRENDS IN SE MODULE EMBEDD APPLICATIONS

Articles based on the CNKI database (China Periodicals Full-text Database) show that its total database contains a total of 194 articles about SE modules. The SE module has advanced to deep learning[16], convolutional neural networks[17], algorithm research,[16] attention mechanisms,[18] and methodological research[19] by virtue of its own strengths.

Figure 8 is a Chinese of the main discipline distribution of articles on "SE modules" (taken from the top 30), and there is no doubt that SE modules are used the most in computer software and computer application disciplines, accounting for 36.68% of the total disciplines. It is gratifying that the SE module has penetrated into the field of people's livelihood. The embedded application of SE modules in industry-related fields including automation technology, telecommunications technology[20], power industry[18], light industry handicraft industry[21], automobile industry[22], road and water transportation[22], radio electronics and other industrial-related fields accounted for 47.82% of the top 30 disciplines; The proportion of medical and health fields including oncology[23], clinical medicine[24], biomedical engineering[25], ophthalmology and otolaryngology[26], urology, traditional Chinese medicine, respiratory diseases, endocrine gland and systemic diseases, and cardiovascular system diseases accounted for 9.77% of the top 30 disciplines. And since the advent of SENet in 2017, the number of articles on SE module research has increased exponentially from 6 in 2018, 20 in 2019, 52 in 2020, and 81 in 2021.

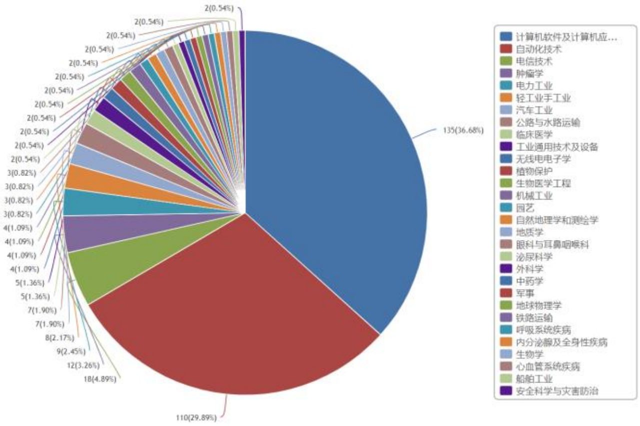


Figure 8: Chinese contains a map of the main discipline distribution of the articles on "SE Modules"

6. CONCLUSION

With the development and application of deep networks, people are aware of the problems of model overfitting and gradient changes brought about by deep networks that reduce model performance, so more and more people have begun to conduct alternative research. This article first introduces the characteristics of the SE module. Secondly, the SE module is embedded in different scale ResNet models for experimental research, and it is found that the test accuracy improvement brought by embedding SE module is much higher than the accuracy improvement brought about by increasing the depth of the original model, and the excellent performance of the SE module is not only manifested in the ImageNet dataset and the ResNeXt model, it has high generalization, simple module structure, and its embedding is universal to greatly improve the accuracy of image classification models of different scales. It can be widely used to avoid fatal problems caused by deep and deep networks. Experimental results on the CIFAR-10 and Dogs-vs-Cats datasets show that the larger the amount of data, the more it can avoid the overfitting phenomenon of the model. A comparison experiment with the Same GoogLeNet model, which addresses the shortcomings of deep networks, shows that SENet is superior. Finally, based on the articles collected in the CNKI database, it is found that in the past five years, the research and application of SE modules by researchers has increased exponentially and SE modules have penetrated into the field of people's livelihood and are widely used in industrial and medical and health-related fields. We look forward to it being introduced into more different fields, combining in-depth mining applications and embedding, so that its performance can be fully illuminated, and then benefit human society.

REFERENCES

[1] Zhang H.L ,2014, " Research on image fine classification based on traditional methods and deep learning ", master thesis, Hefei University of Technology, CHINA.

[2] Jie, et al,2018, "Squeeze-and-Excitation Networks. " IEEE transactions on pattern analysis and machine intelligence, 2018: 7132-7141.

[3] Jia, D., et al. "ImageNet: A large-scale hierarchical image database." CVPR 2009:248-255.

[4] Szegedy, Christian , et al. "Going Deeper with Convolutions." IEEE Computer Society, CVPR 2015:1-9.

[5] Bell, S., et al. "Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2016.

[6] Zhang K.et al. " A review of deep convolutional neural network models for image classification ". Chinese Journal of Image and Graphics, 2021 ,26(10):2305-2325.

[7] Hou Y ,2020, " Fine-grained image classification based on deep convolutional neural networks and two-domain attention mechanism ", master thesis, Northwestern University, CHINA.

[8] Xu J,2020, " Research on fine-grained image classification based on deep convolutional neural networks ", master thesis, Chongqing University of Posts and Telecommunications, CHINA.

[9] Xie, S., et al. "Aggregated Residual Transformations for Deep Neural Networks." IEEE, CVPR 2017:5987-5995.

[10] Russakovsky, O., et al. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision 115.3(2015):211-252.

[11] Krizhevsky, A., et al. "Learning multiple layers of features from tiny images". Handbook of Systemic Autoimmune Diseases, 2009, 1(4).

[12] He, K., et al. "Deep Residual Learning for Image Recognition." IEEE 2016:770-778.

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

[14] Howard, A. G., et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." (2017).

[15] Krizhevsky, A., I. Sutskever , and G. Hinton . "ImageNet Classification with Deep Convolutional Neural Networks." Advances in neural information processing systems 25.2(2012: 1097-1105).

[16] Li S.Z ,2021, " Research on image description algorithm based on deep learning ", master thesis, Beijing Jiaotong University, CHINA.

[17] Qiu N,J,et al. " Research on text topic recognition algorithm of improved Convolutional neural network ", Computer Engineering and Applications,2022,58(02):161-168.

[18] Liu G.T.et al. " An improved X-DR image detection method for GIS internal subtle defects based on superresolution attention mechanism ",High voltage technology,2021,47(11):3803-3809.

[19] Zhang M,Y ,2021, " Research on traffic flow statistics method of two-way lane based on traffic surveillance video ", master thesis, Chang'an University, CHINA.

[20] Wu P,J, et al. " Recognition algorithm of multi-digit phase modulation signal based on convolutional neural network ", Computer applications and software,2019,36(11):202-209. [J].

[21] Li L ,2019, " Research on fabric defect detection method based on convolutional neural network ", master thesis, Huazhong University of Science and Technology, CHINA.

[22] Zhang X,L et al . " License plate character recognition based on improved LeNet-5 network ", Journal of Shenyang University (Natural Science Edition),2020,32(04):312-317.

[23] Xu X,B, et al " High precision breast cancer classification based on fusion space and channel characteristics ", Journal of Computer Applications,2021,41(10):3025-3032.

[24] Xia K,L ,2019, " Research on 3D medical image segmentation method based on U-Net ", master thesis, South China University of Technology, CHINA.

[25] Wang H, 2021," Research on automatic medical image segmentation algorithm based on attention mechanism ", master thesis, University of Electronic Science and Technology of China, CHINA.

[26] Mu G,R ,2020, " Research on multi-organ segmentation of head and neck based on 3D convolutional neural network ", master thesis, Southern Medical University, CHINA.