TEST Engineering & Management

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

Research on Image Classification, Retrieval and Recognition Technology Based onConvolution Neural Network in Artificial Intelligence Environment

Zhenzhen Xing[1,*]

[1]Department of Computer Engineering, Taiyuan Institute of Technology, Taiyuan, Shanxi, China, 030008

Article Info
Volume 83
Page Number: 5488 - 5496
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *28 August 2020*

Abstract

With the gradual improvement of deep learning theory, deep learning has been applied to more and more fields, including image processing, video processing, audio processing and other fields. Among them, Convolution Neural Network (hereinafter referred to as CNN) is an important network model in deep learning, which is different from the traditional Neural Network (hereinafter referred to as NN), mainly introduces the concepts of convolution layer and pooling layer. With the rapid development of information technology in China, a large number of images are collected and used in modern applications, which is an important application way in artificial intelligence environment. However, the traditional image processing methods can't meet the needs of modern image classification, retrieval and recognition, which requires us to use modern information technology to achieve automatic classification and retrieval. By effectively mining image information, we can improve the management efficiency. Through the research of CNN, image classification retrieval and recognition technology gradually use the theory, which has gradually played an important role. Through CNN, we can better introduce into the field of image processing, which will improve the effect of better classification retrieval model.

Keywords: *Artificial Intelligence, Cnn, Image Classification, Retrieval And Recognition Technology;*

1. Introduction

The amount of network data increases exponentially every year, which requires us to improve the data processing capacity and speed of information[1]. With the improvement of data processing ability, machine learning algorithm plays an increasingly important role, which has become the most popular research field in recent years. Deep learning is a new classification and prediction method evolved from the traditional NN, which is a NN model for classification or prediction. Through different NN models, artificial intelligence technology can be applied to many different specific models, which has become the mainstream classification retrieval and recognition scheme of image, text, video and audio.

NN is a machine learning model derived from biology, which is a method of introducing neurons to simulate neurons in human brain. As the neuron receives the input, through a specific weight and activation function, we will get the corresponding output, which will become the input of the next neuron. After the emergence of NN, there have been many problems, such as the amount of data can't meet the training needs, network training over fitting and so on[2]. With the advent of the era of big data, the amount of network data increases exponentially

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

every year, which provides an important guarantee for the demand data of NN[3].

Through the CNN classification model, we greatly improve the recognition accuracy of Imagenet image dataset, which has become an important method in the field of image classification retrieval and recognition[4]. Subsequently, convolutional NNs have been applied to the fields of text, audio and video. In text processing, recurrent NN can deal with temporal recursive information, which has been widely used in text resource processing[5]. In the video field, based on the single frame convolutional NN classification algorithm, we can expand the processing of multi frame convolutional NN classification algorithm, which has achieved great success[6].

2. Constraints of CNN

2.1. Need a lot of acquired manuallylabel data

CNN is a supervised machine learning method, which faces a lot of training and raw data. Therefore, it is necessary to label a large number of data by artificial NN. At present, the trend of CNN is to build more and more layers of network model, which will make the structure more and more complex. Therefore, we need more and more tag data to complete the training process, which will cost a lot of labor and time[7]. Therefore, we can solve the problem of manual annotation information in deep learning by other methods, such as unsupervised learning and generative adversarial network (GAN).

2.2. Need high hardware resources

Deep learning method needs to consume a lot of material and time cost. At the same time, due to the large amount of data calculation, the CNN needs the computer to have enough hardware resources, which will complete the corresponding calculation and storage work[8]. For example, in the Alphago X44 model, in the training process, we need 50 image processing units to pass the training for 20 days, which is only one part of the training. Therefore, deep learning method needs a lot of computation and hardware resources, which is also a disadvantage of convolutional NN. Therefore, we need to design and manufacture hardware with higher performance or special adaptation to deep learning algorithm. By improving the algorithm or operating system, we can learn in depth and consume more resources, which has become an important direction of deep learning[9].

2.3. Needmore sample feature training

We need a variety of methods to make the deep learning method closer to the human learning method. However, the process of human learning is quite different from that of machine learning. Human beings can learn the main characteristics and differences of samples from very few samples. The ability of human learning and distinguishing comes from memory and knowledge base, which can also be acquired by human logical reasoning. Therefore, computer deep learning needs to be combined with knowledge mapping, logical reasoning and other methods, which requires small samples to complete the learning results[10]. By optimizing the network model structure, we can better extract the image features, which will better complete the classification, recognition, retrieval and other work.

3. Image enhancement technology

3.1. Theory of image enhancement technology

The goal of image enhancement is to improve the image clarity, which will improve the image quality. Image enhancement technology can be divided into frequency domain processing method and spatial domain processing method. Frequency domain processing is a method to enhance the image signal through two-dimensional Fourier transform, which requires high-pass filtering method to enhance the high-frequency signal of image edge in the changing process. In this way, we can make the blurred image clear[11]. At the same time, through the low-pass filtering method, we can filter the noise in the image. Among them, the spatial domain processing method is mainly the median filtering method, which is usually used to remove or weaken the noise[12].

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

3.2. Low pass filter and high pass filter technology

Low pass filtering and high pass filtering are realized by two-dimensional Fourier transform and inverse transform, which can eliminate image noise and enhance image edge[13]. There are many varieties of high pass filter, and the common filters are exponential filter and Butterworth filter. Compared with the linear filter, the exponential filter attenuates faster and the smoothed image has no obvious transition band. Butterworth filter reduces the degree of image blur, it has no obvious sharp cut-off, and the image clarity can be significantly improved after processing. The processing flow of low pass filter and high pass filter is shown in Figure 1.

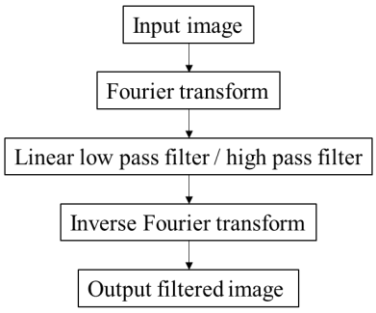

Figure 1. Low pass filter / high pass filter process flow chart.

The conversion function of low-pass filter is shown in Formula 1. Where LX represents the output of the low-pass filter function and the filtering threshold is t. If the output of the low-pass filter function is less than the threshold value, the output is 1, otherwise the output is 0.

$$L(x,y)=\begin{cases}1 & T(x,y)\le T_0\\ 0 & T(x,y)>T_0\end{cases} \quad (1)$$

The high pass filter conversion function is shown in Formula 2. If the output of the high pass filter function is less than the threshold value, the output is 0, otherwise the output is 1.

$$L(x,y)=\begin{cases}0 & T(x,y)\le T_0\\ 1 & T(x,y)>T_0\end{cases} \quad (2)$$

3.3. Median filtering technology

Median filtering uses the median value of sliding window to replace the value of window center point, which is a common image smoothing technology. Median filtering technology can effectively deal with impulse interference and particle noise, which can solve the image blur problem caused by linear filter. However, the median filtering technology can't deal with the images with complex details and lines. Median filtering is one of the most commonly used filtering methods, which can reduce the interference of impulse noise and random noise, while preserving the details of the image. In recent years, weighted median filtering, linear combination of median filtering and high-order median filtering have been proposed. Combined with deep learning algorithm, median filtering technology has achieved good filtering effect.

4. Image segmentation technology

Image segmentation technology is to divide the image into several regions according to the texture, color or content of the image. Through this segmentation method, we can extract foreground background separation and regions of interest. At present, the commonly used image segmentation techniques are: image segmentation based on edge, threshold and region.

Edge based image segmentation generally uses the edge gradient detection method based on gray image, which is a kind of image segmentation technology based on the change characteristics of image texture, color or gray value. Image edge points can be obtained by first or second derivative. If the gray value of the image changes in the form of roof or step, the point is on the edge of the image. Image segmentation technology based on threshold is classified according to the threshold, including adaptive threshold method and fixed threshold method. According to the threshold value range, we can divide into local threshold method and global threshold method.

5. Image retrieval technology

Published by: The Mattingley Publishing Co., Inc.

5.1. Theory of image retrieval technology

With the rapid popularization of smart phones and digital cameras, the types and sources of images are more and more extensive, which also accelerates the speed of image production and transmission. Therefore, how to retrieve the required images from massive images has become a hot topic in the field of multimedia, which also adds an important research content of image retrieval.

5.2. Image retrieval technology based on text

The traditional text retrieval technology is used in text-based image retrieval, which is based on information such as image keyword description. Text based image retrieval technology is based on mature text retrieval technology, which has many advantages, such as small difficulty, fast speed and so on. However, there are still many problems in the text-based image retrieval technology in China, which will be difficult to fully retrieve the required content, such as the image without text description, the limitation of description vocabulary, and the slow updating of vocabulary.

5.3. Image retrieval based on image content

Image content-based image retrieval is a retrieval method with image semantics as the main feature, which can find similar images in the image database. Therefore, we can divide it into simple and complex semantic retrieval. Among them, simple semantic retrieval mainly uses image features such as color histogram, color set, color matrix to establish a relatively simple image set for retrieval, which can't reflect the image features at a higher level and describe the essential features of images. Therefore, it is difficult for simple semantic retrieval to completely retrieve the required inner cylinder. Complex semantic retrieval usually uses SIFT algorithm, deep CNN and other methods to extract image features, which can overcome the feature changes caused by image rotation, stretching, scaling and other operations. Through complex semantic retrieval technology, we can reflect the characteristics of images from a higher level, which has become the trend of image retrieval. With the research of scale invariance, deep learning and other related algorithms, image complex semantic retrieval technology has been developed rapidly, which has been widely used in image retrieval technology.

5.4. Image retrieval based on combination of text and image content

Image retrieval based on the combination of text and image content is a research on the fusion of text and image content, which can give full play to the advantages of text retrieval and image retrieval. Through the fusion technology, image retrieval can achieve the purpose of high speed and accuracy.

6. Theoretical knowledge of CNN

6.1. Theory of artificial NN

Artificial NN (ANN) is a calculation method based on the structure and operation principle of biological NN. According to the topological structure of the network, an abstract mathematical model is constructed to simulate the neural behavior characteristics of human brain. Through the interconnection of a large number of basic neurons, we can form an intelligent NN system, which will have strong nonlinear, high fault tolerance and self-learning characteristics. Therefore, artificial NN can deal with all kinds of complex problems, which has been applied to many fields, such as intelligent control, pattern recognition, medicine and finance. Among them, perceptron model is regarded as the beginning of NN; restricted Boltzmann machine model is applied to energy function modeling; Hopfield NN is mainly used for associative memory.

6.2. Neurons

Biological NN is based on the principle of neuron. Therefore, we introduce the concept of "perceptron", which is the earliest artificial NN model. Neurons are feedforward NNs of living functions. After continuous learning of network parameters, neurons can classify and recognize input data. Neuron is a kind of method that uses threshold excitation sensor to respond to output junction after input data

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

operation. It is the basic processing unit of NN model, which mainly includes three elements: connection, summation node and activation function. The basic neuron model is shown in Figure 2. The structure of neurons can be expressed by mathematical expressions, such as formula 3 and formula 4.

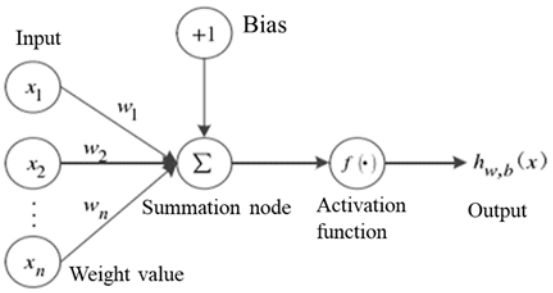

Figure 2. Neuron model.

$$z = \sum_{i=1}^{n} w_i x_i + b \quad (3)$$

$$h_{w,b}(x) = f(z) = f(\sum_{i=1}^{n} w_i x_i + b) \quad (4)$$

6.3. NN

NN model is a model with complex structure, which is composed of several basic neurons connected with each other according to certain rules. The connection rule is the output of one or more neurons in the upper layer of the layer as the input of the layer, which forms the input layer ($Layer\,L_1$), hidden layer ($Layer\,L_2$) and output layer ($Layer\,L_3$), as shown in Figure 3.

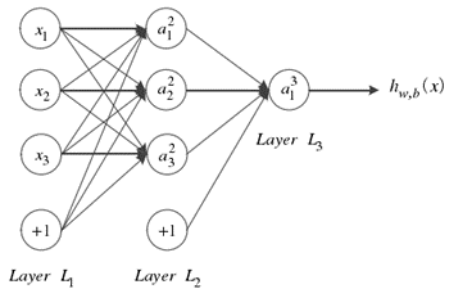

Figure 3. NN model.

6.4. CNN

6.4.1. CNN theory

CNN is a multi-layer network model under supervised learning, which is composed of convolution layer and sampling layer alternately. After the input image is mapped by nonlinear activation function for many times, we can map the feature representation learned from the full connection layer to the sample label space. The output layer is composed of radial basis function units, which can output the probability of corresponding categories respectively. The uniqueness of CNN is mainly manifested in two aspects. First, the connection between neurons is a local connection, which is not a full connection. Secondly, some neurons in the same layer adopt the mechanism of weight sharing, which leads to different neurons sharing the same weight parameters. In this way, we can effectively reduce the complexity of the model, which will greatly reduce the number of parameters to be learned in the network.

6.4.2. Basic structure

The basic structure of CNN is shown in Figure 4. When an image is loaded into the input, we can transfer the image feature information layer by layer through the convolution pooling (down sampling) structure between layers, which decodes, deduces, converges and maps the original image feature information to the hidden layer. Finally, the full connection layer matches the extracted features and outputs the classification results.

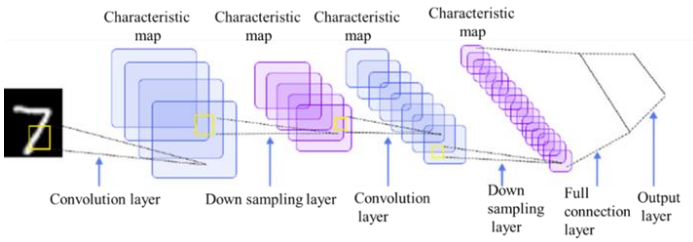

Figure 4. Basic structure of CNN.

6.4.3. Convolution layer

In the convolution layer, we can extract the features of the upper layer by convolution operation of one or more convolution kernels, which will get the feature map of the next layer. At the same time,

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

through the activation function, we can map the output of convolution operation, which will get the characteristic mapping relationship between input and output. The convolution kernel in each layer traverses the whole feature graph based on the sliding window mechanism. The number of convolution kernels is equal to the number of feature graphs. The specific image convolution operation process is shown in Figure 5.

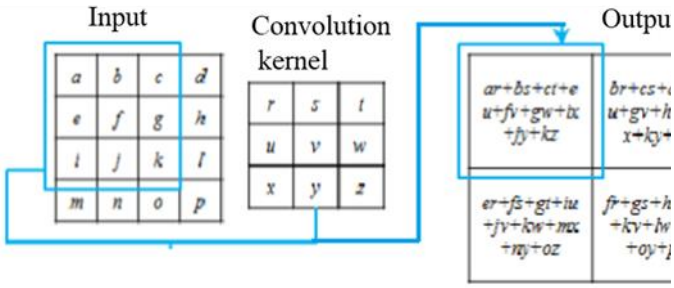

Figure 5. Convolution process.

The convolution process is a 4x4 two-dimensional image feature map. After a 3x3 convolution kernel convolution operation, we get a 2x2 two-dimensional image feature map. The expression definition of convolution operation is shown in formula 5. Among them, x_i^l represents the set of input characteristic graphs. *Is the convolution operation, b_i^{l-1} is the corresponding offset term.

$$x_i^l = f\left\{\sum_{j\in M_i} x_i^{l-1} * k_i^{l-1} + b_i^{l-1}\right\} \tag{5}$$

6.4.4. Activation layer

The activation layer is the nonlinear activation output of the upper layer's output feature information. CNN is often deep. If sigmoid function and Tanh hyperbolic tangent function are used, gradient vanishing problem will appear. Therefore, the excellent nonlinear activation functions in CNN include ReLU function, PReLU function and ELU function. Among these nonlinear activation functions, the ReLU function can solve the gradient dispersion problem of the network, which will reduce the training time of the network and accelerate the convergence speed. Therefore, the current convolution network generally uses ReLU as the activation function. The image of the ReLU function is shown in Figure 6. We can identify it by mathematical expression, as shown in formula 6.

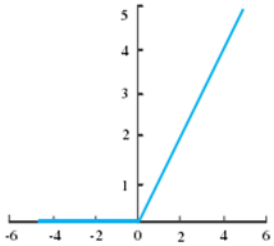

Figure 6. Image of ReLU function.

$$f(x) = \begin{cases} x, if & x \geq 0 \\ 0, if & x < 0 \end{cases} \tag{6}$$

6.4.5. Pool layer

Pooling layer usually appears between continuous convolution layers in CNN, which mainly simplifies a pixel area in the upper feature map into a pixel. By reducing the dimension of feature graph and extracting features, we can reduce the amount of calculation and learning parameters in the network. The pooling layer only reduces the dimension of a single feature map. Therefore, after pooling, the number of output feature maps will be consistent with the number of upper feature maps. At present, three common pooling methods are Max, mean and random pooling. The maximum pooling operation selects the maximum value as the output value of the current image area. Average pooling is to calculate the average value of all pixels as the output after pooling. Random pooling is to select pixels randomly according to the size of probability matrix as output. The operation process of pooling layer is shown in Figure 7. The pooling layer can be abstractly represented by mathematical expression, as shown in formula 7.

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

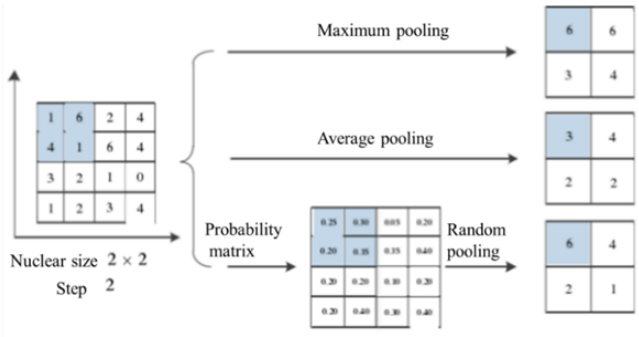

Figure 7. Calculation process of pooling layer.

$$x_i^l = f(\beta_i^l down(x_i^{l-1}) + b_i^{l-1}) \tag{7}$$

Among them, $down$ is the down sampling function, x_i^l is the sampling result, b_i^{l-1} is the corresponding offset term.

6.4.6. Full connection layer

The last few layers of convolutional NNs are usually one or more full connection layers. In front of the full connection layer is the process of two-dimensional image feature extraction. The main task of the full connection layer is to transform the two-dimensional feature vector extracted in front of it into one-dimensional feature vector, and the classifier completes the classification. The calculation formula of full connection layer is shown in formula 8.

$$x_i^l = f\left\{\sum_{j \in M_i}^{n} x_j^{l-1} w_{i,j}^{l-1} + b_i^{l-1}\right\} \tag{8}$$

7. Image classification, retrieval and recognition technology based on convolutional NN

7.1. Image classification algorithm

There are two kinds of image classification algorithms. One is the bag of words model based on classical features, which is a training mode through support vector machine or other classifiers. In the classification operation, we need to get the image features first, and then get the corresponding vector according to the bag of words model. Then, we can use the vector as the input of the classifier. Finally, we can output the classification results for the image. The other is a classification model based on convolutional NN, which can directly use convolutional NN for image processing. By using the image as the input of the NN, our output can be used as the classification result of the image. The feature-based classification method mainly includes several features: gist feature, SIFT feature and hog feature. The algorithm and process of image classification using convolutional NN are different from those of traditional image classification algorithms. In the implementation of image classification algorithm, the input is all pixels of the image, which is $WIDTH \times HEIGHT \times CHANNEL$ pixel matrix. The final output is the classification result.

7.2. Analysis of image classification results

Based on the principle of multi-layer CNN classification, the accuracy of classification results has been greatly improved. Through the experiment of keras framework, we found that the training time is relatively long and the classification accuracy is low. Then we use the Caffe framework for experiments, because Caffe frameworks have ready-made training models. Therefore, $WIDTH \times HEIGHT \times CHANNEL$ training and testing is more convenient, and its test results have made great progress. The flow chart of classification principle based on multilayer CNN is shown in Figure 8.

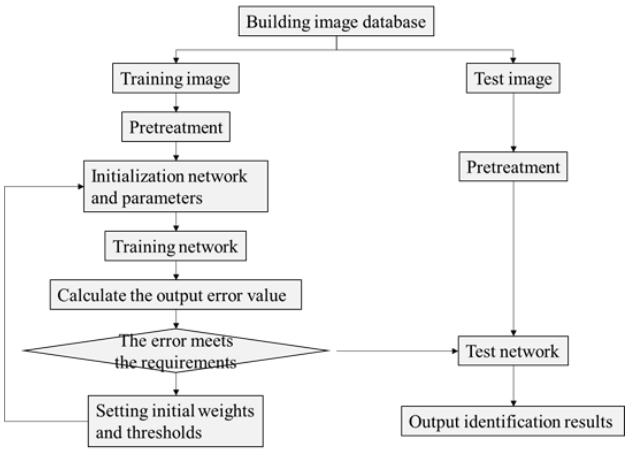

Figure 8. Flow chart of classification principle.

7.3. Image retrieval based on semantic and content

By extracting the image feature vector, we can calculate and obtain the retrieval image according to

Published by: The Mattingley Publishing Co., Inc.

the similarity of the feature vector, which is a content-based retrieval method. Only relying on image feature vector needs a lot of computing resources. Therefore, this paper designs a way to combine semantic and content-based, which can improve the efficiency and accuracy of retrieval. The image retrieval process is shown in Figure 9.

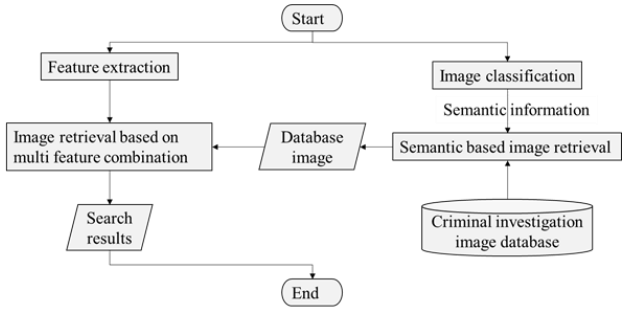

Figure 9. Image retrieval process based on semantic and content.

8. Conclusion

CNN is a kind of deep learning model which is formed by neuron signal transmission of visual mechanism in cerebral cortex. According to the characteristics of local feeling, parameter sharing, sparse connection and down sampling, we can recognize the strong learning and classification ability. Based on the combination of semantic and content retrieval method, we can get better semantic information of image classification, which will reduce the scope of retrieval. Through multi feature retrieval, we can improve the retrieval effect.

Acknowledgments

Science and technology innovation project of colleges and universities in Shanxi Province: "cancer-gene" relationship model research based on deep learning, (project No.2019L0931). Project leader: Liu Jie.

References

[1] Hu Qingwu, Wang Haiyin. Application of image-based case scene investigation [J]. Criminal technology, 2010 (6): 57-59.

[2] Zhang Mingyuan, Wang Hongli, Zheng Jiahua, et al. Application of improved median filter in image denoising [J]. Ordnance automation, 2007, 26 (8): 45-47.

[3] Li Zhi, sun Yubao, Wang Feng, Liu Qingshan. Clothing image classification and retrieval algorithm based on deep CNN [J]. Computer Engineering, 2016, 42 (11): 309-315.

[4] Li Xiangfeng, Wang Bin, Liu Feng, Hu Fuqiao. A color text image binarization method based on graph theory clustering and binary texture analysis [J]. Chinese Journal of image graphics, 2004, 9: 36-42.

[5] Hou Kejie. An algorithm structure for optimizing matching source and destination color space by color visual perception parameters [J]. China printing and packaging research, 2009, 1: 22-26.

[6] Jiao pengpeng. Matlab implementation of texture feature extraction based on gray level co-occurrence matrix [J]. Computer technology and development, 2012, 22: 169-171.

[7] Zhang Zhenghua, Fang Zhijun. A novel anti distortion image detection algorithm [J]. Computer applications and software, 2007 (04): 142-144.

[8] Zhu Mingdong, Xu Lixin, Shen Derong, Kou Yue, Nie Tiezheng. Cosine similarity query method for uncertain text data [J]. Computer science and exploration, 2018, 12 (01): 49-64.

[9] Han xingshuo, Lin Wei. Research and implementation of deep CNN in image recognition algorithm [J]. Microcomputer and application, 2017, 36 (21): 54-56.

[10] Xu Siwei, Chen Siyu. Image classification method based on deep learning [J]. Application of electronic technology, 2018,44 (06): 116-119.

[11] Zhang Qiang, Li Jiafeng, Zhuo Li. Vehicle color recognition in monitoring scene based on convolutional neural network [J]. Measurement and control technology, 2017, 36 (10): 11-14.

July - August 2020
ISSN: 0193-4120 Page No. 5488 - 5496

[12] Tian Juan, Li Yingxiang, Li Tongyan. Comparative study of activation function in convolutional neural networks [J]. Computer system applications, 2018, 27 (07): 43-49.

[13] Wang Qiang. Image classification based on convolutional neural network algorithm [J]. Journal of Chengdu University of information technology, 2017,32 (05): 503-507

Published by: The Mattingley Publishing Co., Inc.