Image Style Conversion using Deep Convolutional Neural Network

LINGLING WANG, XINGGUANG DONG
School of management science and Engineering, Anhui University of Finance and Economics
Bengbu 233000, CHINA

Abstract: At present, research on image style conversion based on deep learning is increasing, different from the conventional style conversion, this paper is based on convolutional neural networks, using the InceptionV3 model trained under ImageNet dataset. By using Deep Dream technology, which gives a dull and ordinary background picture a warm color, makes the picture content richer, the texture is very softer and more exquisite.

Keywords: deep learning, convolutional neural network, deep dream, image style conversion.

Received: March 9, 2022. Revised: August 13, 2022. Accepted: September 2, 2022. Available online: September 21, 2022.

1. Introduction

Image style conversion is an image processing method that converts its background and overall tone (style) without changing the overall framework of its own source image, to obtain a new aesthetic image, which perfectly combines the local characteristics of the source image with the style. However, based on the limitations of the style conversion algorithm, the traditional style conversion algorithm cannot extract the high-level abstract features in the target image, and can only synthetize the pictures of the abstract painting style. With the rapid development of artificial intelligence algorithms and the investment of funds in experimental facilities, scientists from Germany such as Gatys and others proposed the use of convolutional neural networks to achieve style conversion of images[1]. They discovered the convolutional neural network through research, dividing it into deep convolutional layers and shallow convolutional layers, the former can obtain the overall framework of the image, while the latter can obtain the style characteristics of the image, based on this discovery, to achieve the separation of style and content in the image, and then the overall content and style of the image are combined to achieve style conversion. Although the method successfully implements style conversion, the gradient descent method needs to be used to repeatedly update the pixel values of the source image in the implementation, which greatly increases the resource occupation and cost, and brings a series of problems of slow generation progress. Subsequently, researchers have made many improvements based on Gatys' research, such as Justin et al. proposed a rapid stylized conversion algorithm[2], which can make the source image transform the style after only one optimization by establishing a feed-forward network in advance, which greatly improves the generation speed.

Based on CNN's real-time style conversion algorithm[3], this paper designs a more efficient style conversion algorithm to effectively improve image quality and reduce the time of processing data redundancy and nonparametric algorithms to a certain extent. Then use crawler technology, by extracting the sensitive words of the information required by the user, obtain the required images from the website or user-side files, and transform the style of the image through deep learning, to obtain their ideal exquisite pictures, or get the important information they want to extract from the pictures. By combining the principle of deep dream technology to use it to optimize the random noise image, the pixel value randomly generated by the random noise picture is fixed by the convolutional neural network, and the random noise is adjusted at the beginning, and optimize the adjustment, so that the noise picture presents a certain feature distribution and thus optimizes the output picture.

2. Convolutional Neural Network

2.1 Feature extraction stage

In the premise of the image style conversion, about how to obtain the data information of the image reasonably and perfectly[4]. we cannot directly use the image as input to process it, for a colored picture, its basic characteristics mainly include two parts, depth and pixels. The main idea of machine learning is to transform real-world information into vectors, that computers can process. In the image style conversion, we can convert the image data into the form of a pixel matrix through the CNN, for black and white pictures, each point represents only one pixel value, if it is a color picture, each point will have three-pixel values representing RGB.

We know that the cross-correlation operation of the digital image saved as a matrix, which meanings for each pixel in the image, with the gray value of the pixels around it weighted to adjust the gray value of this point[5]. First need to define a convolutional kernel, which is also known as the convolutional template or convolution window, is an N x N matrix, the size of the convolutional kernel determines the scope of the operation, it should be a cardinality, so that there is a central point, the number of the convolutional kernel is the weight of this point and the points around it[6]. The example is described as follows Fig.1.

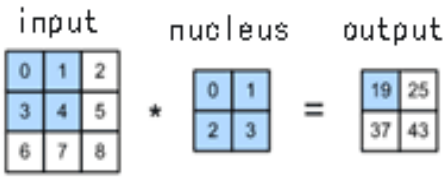

Figure 1. Convolutional calculation process with depth 1

For the example in the figure above, the convolution is calculated as follows:

$$0\times0+1\times1+3\times2+4\times3=19,$$
$$1\times0+2\times1+4\times2+5\times3=25,$$
$$3\times0+4\times1+6\times2+7\times3=37,$$
$$4\times0+5\times1+7\times2+8\times3=43.$$

In addition, a non-black and white color picture, that is also need to consider the basic characteristics of the depth of the picture, then if the depth of a picture is 2, a pixel is composed of 2 values, in the operation, if the above basic convolution operation is regarded as a two-dimensional cross-correlation operation, then the convolution operation with depth would be a three-dimensional cross-correlation operation[7], for example, a depth of 2 operation process is as described in Fig. 2.

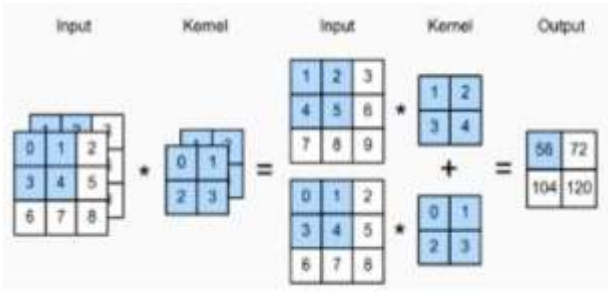

Figure 2. Convolutional calculation process with depth 2

Step size is equal to the side length of the convolutional kernel, equivalent to reducing the image by N times, the convolution operation of the step is also a way to reduce dimensions, the size of the picture after convolution[8]: If the step size is S, the original picture size is [N1, N1], the convolutional kernel size is [N2, N2], then the size of the graph after convolution:[(N1N2)/S+1,(N1N2)/S+1].

After the transformation of the image data, we will find that in the convolutional layer, a convolutional kernel can only extract one feature, which is obviously not enough, but through multiple convolutional nuclei at the same time to obtain multiple sets of features can be combined, judged, or classified. With the support of a large amount of data, it is possible to achieve a freer image style conversion after a certain degree of mechanical learning. Convolution in the image conversion more emphasis on certain features, for the picture, pixels and depth is the most significant features[9], and the significance of convolution is to extract different characteristics after calculation and strengthening, as the computer can understand the "image features". When the number of convolutional kernels reaches a certain level, a convolutional layer is formed, so the convolutional layer is also called the feature layer[10], the characteristic of the convolutional layer is to obtain many convolutional kernels (features) through convolutional operations, as the operation is repetitive, so it often contains many redundant data, most data cannot be used, so it still needs to "pool"[11].

The pooling layer, also known as the feature mapping layer, as the name suggests, is a certain processing of a large amount of feature data in the convolutional layer, to filter and integrate many duplicate data or similar data, reduce the overall data volume, and improve the availability of data. From the perspective of hierarchy, the pooling layer as a whole is characterized by downward sampling, if the meaning of the convolutional layer or convolution lies in the conversion and acquisition of the characteristics of the image data[12], then the significance of the pooling layer is to screen and refine the feature information extracted by the convolutional layer, and select the most representative features, which can lower the repeatability of the pixels, make the subsequent convolution more meaningful, and calculate more conveniently[13].

Pooling has a maximum pooling, average pooling, etc. different ways, the use of that way may depend on the situation, if the adjacent pixels of an image are very similar, and the number is large, then often the use of maximum pooling can achieve better results[14], if the picture composition is more complex, needs to extract more features, then you can consider the average pooling, the following Fig. 3 is the largest pooling of simple ideas and calculation process:

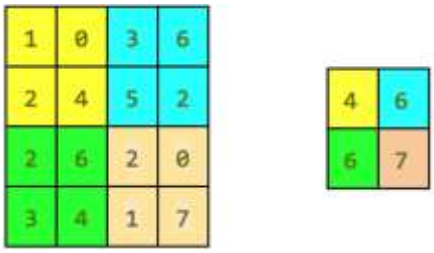

Figure 3. The process of maximum pooling

Convolutional layer and pooling layer together constitute the feature extraction stage of convolutional neural network[15], a complete convolutional neural network includes many convolutional layers and pooling layers, to achieve image style conversion through the support of a large amount of data, this part is completed by the software or machine itself, and constantly repeat the convolution, pooling process, making the data more and more useful. The overall process is shown in the following Fig. 4:

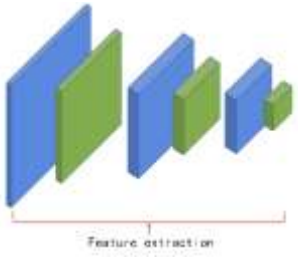

Figure 4. Feature extraction stage process

2.2 Classification identification stage

Classification recognition stage by one or more fully connected layers, convolutional neural network processing of the image of the basic process, can be seen as the image data for continuous convolutional calculation and pooling calculation[16], by changing the size of the convolutional kernel, the number, extract more features, finally, throws the features which go through multiple convolution and pooling processing into one or more fully connected layers, using of softmax function to classify them, so as to achieve the role of identifying different categories of objects, its overall flow as shown in the following Fig. 5:

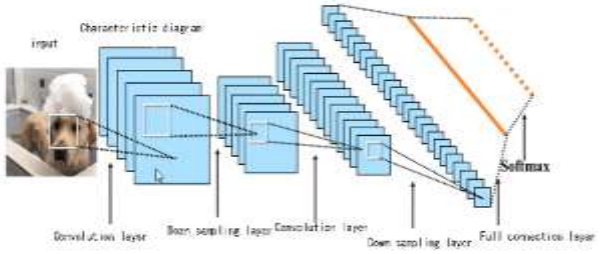

Figure 5. Convolutional neural network process diagram

Therefore, the focus of convolutional neural networks is the fully connected layer network supported by a large amount of data, and the convolutional layer and pooling layer can be regarded as a tool for machine learning from the result[17], and the image style conversion based on the convolutional neural network, first, takes out the features from the image, enters the network, Secondly, undergoes step-by-step transformation, and samples downward .Finally, repeat this process repeatedly, the longer the "learning" time, the more data can be used, the more accurate the classification, and the more diverse, accurate, and rapid the final transformation of the image style will be.

3. Proposed Method

3.1 Project introduction

Deep dream is a convolutional neural network based on Google in 2015, can make artistic modifications to the image technology[18], it can make the modified picture produce a fantastic artistic effect, just like the dream of people in the dream, strange abstraction, so it is called deep dream, with this technology, photos with an artistic style can be obtained. The images generated by this technique can not only impress people, but also help us understand what convolutional neural networks are learning[19].

3.2 Fundamentals

In the past practice, we used convolutional neural networks for image recognition, input a large amount of sample data into the network, test the features extracted by neurons, train the gradient of the neural network, reverse update the convolutional neural network weights, and iterate repeatedly until the network converges to the expected accuracy to stop training, then the resulting neural network can be used to classify in Fig. 6.

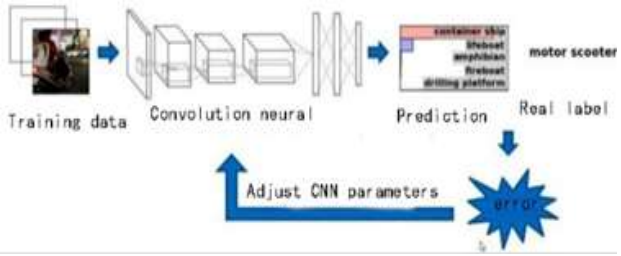

Figure 6. Training process

In contrast to the convolutional neural network's practice of testing the features extracted by neurons by entering pictures, the deep dream model is to randomly select some neurons to observe what their simulated pictures may look like, and this information is deep dream model updated back to the content of the network in reverse, so that the features or enhanced patterns that each neuron most want to represent can be obtained, assuming that we continue to iterate the output and constantly activate the features it wants to represent, and the final output result will be closer and closer to the target image. The essence of the deep dream model is to visualize the characteristics of each layer in the neural network through the gradient ascending method, and the difference from the convolutional neural network[20], that is the reverse feedback updates not the convolutional neural network weights, but the pixel values in Fig. 7.

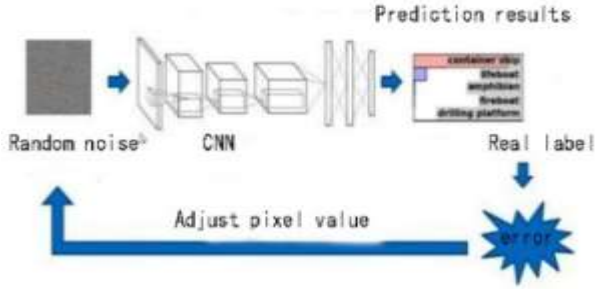

Figure 7. Feedback network

Suppose the image of the input network is x, the N-dimensional vector [P1, P2,...,Pn] represents that there are N classifications, assuming that the probability the image is of class A as PA, the higher the value of PA, the higher the probability that the image is of class A, the higher the probability that the image is of class A, and PA as our optimization goal, constantly adjust the parameters, so that the value of PA is maximized, the image has the characteristics of class A more prominent, and finally achieve the effect of generating a deep dream picture. The result after the Deep Dream model run is to maximize the probability of a certain category (or the image of various categories in the CNN) to obtain a picture, of course, we can also achieve the final goal by maximizing the activation layer of a certain channel of the convolutional layer, that is, visualizing the features of the convolutional layer.

3.3 Implementation

The network parameters of the convolutional neural network model, which involved in the Deep Dream model are fixed and are already trained models, so a convolutional neural network image classification model should be imported first. The starting point for the development of convolutional neural networks was the neurocognitive machine model, at that time there was already a convolutional structure, the first neural network was LeNet, but as the technology of convolutional neural networks was not very mature at that time, it was replaced by other hand-designed feature classifiers. With the advent of Dropout, ReLu, and GPU+ big data, convolutional neural networks ushered in an epic breakthrough in Fig. 8.

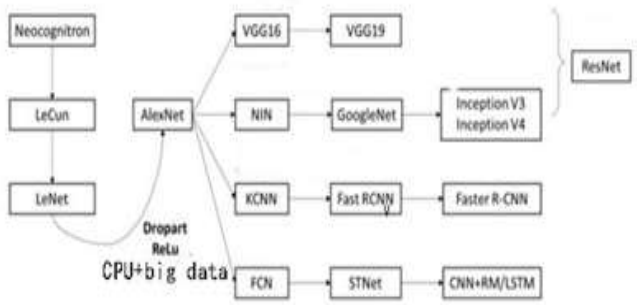

Figure 8. Evolution of convolutional neural networks

The evolution of convolutional neural networks can be clearly seen from the graph. After Alex Net, people through the network increased the functionality of the convolutional layer, the transition from classification tasks to detection tasks, and the addition of new functional modules, which made convolutional neural networks evolve into deep learning networks such as VGG, Inception, ResNet, etc. Here we use the Inception as the model we imported.

In the case of TensorFlow, there are two ways to store and load a model, one is to generate a checkpoint file, and the other is to generate a graph protocol file. Here the model is imported as a diagram file. After storing the model as a graph file, first define the placeholders for the input picture, and then preprocess the picture, subtract the mean, and increase the dimension. Subtraction is due to the subtraction averaging preprocessing done during the training of inception, so the same value needs to be subtracted here to maintain consistency. The dimension is increased because when the image is entered into the network, the image is often not one, but a batch, so it is necessary to add a batch dimension so that multiple pictures can be entered into the network at the same time. Finally, the model is imported, and the preprocessed image is fed into the network.

After completing the basic operation of the graph, you can output the number of convolutional layers, you can also output the names of all convolutional layers, and the parameters of the specified convolutional layer, it should be noted here that when outputting parameters, these three dimensions are the batch, height and width of the image, because at this time the image has not been entered, it is not clear the size and quantity of the input image, and the last dimension is CHANNEL, that is, the number of channels of the convolutional layer, due to the ImageNet image classification mode imported by the deep Dream model is trained, so its network parameters are fixed values and never change.

4. Experiments and Analysis

4.1 Features of the technology

The main idea of Deep dream is to select a channel or convolutional layer of a convolutional layer (or multiple network layers, change the image pixels --the biggest difference from the trainer classifier) to maximize the activation value of the channel or the layer.

4.2 Training the model

In this experiment, the parameters of deep dream are optimized by the InceptionV3 model trained from the ImagNet dataset.

. The architecture of the lnceptionV3 model is quite large, with 11 layers from 'mixed0' to 'mixed10'. Using different layers produces different images, with the darker layers responding to higher-level features such as eyes and faces, while the earlier layers respond to simpler features such as edges, shapes, and textures.

The layers that can be freely tried to choose, deeper layers (layers with higher indexes) take longer to train because of the gradients are calculated more deeply in Fig. 9.

Model: "inception_v3"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, None, None, 3)]	0	[]
conv2d (Conv2D)	(None, None, None, 32)	864	['input_1[0][0]']
batch_normalization (BatchNormalization)	(None, None, None, 32)	96	['conv2d[0][0]']
activation (Activation)	(None, None, None, 32)	0	['batch_normalization[0][0]']
conv2d_1 (Conv2D)	(None, None, None, 32)	9216	['activation[0][0]']
mixed9_1 (Concatenate)	(None, None, None, 768)	0	['activation_87[0][0]', 'activation_88[0][0]']
concatenate_1 (Concatenate)	(None, None, None, 768)	0	['activation_91[0][0]', 'activation_92[0][0]']
activation_93 (Activation)	(None, None, None, 192)	0	['batch_normalization_93[0][0]']
mixed10 (Concatenate)	(None, None, None, 2048)	0	['activation_85[0][0]', 'mixed9_1[0][0]', 'concatenate_1[0][0]', 'activation_93[0][0]']

Total params: 21,802,784
Trainable params: 21,768,352
Non-trainable params: 34,432

Figure 9. InceptionV3 model

4.3 Single-layer network compared with multi-layer network

A single loss layer single channel is used to extract the characteristics of the specified channel for the style conversion of the background image, which generate a deep dream image. By maximizing the average value of a certain channel can get a

meaningful image, compared with the original background picture, the texture of the picture becomes visible to the naked eye, and the color of the image has also changed significantly, which is a wonderful place of deep dream technology. But its effect is only slightly changed on the basis of the original image The experimental demonstration is described in Fig. 10 – Fig. 15:

Figure 10. Background image A

Figure 11. Deep dream single-layer channel feature conversion A

Figure 12. Background image B

Figure 13. Deep dream single-layer channel feature conversion B

Figure 14. Background image C

Figure 15. Deep dream single-layer channel feature conversion C

Using multiple loss layers full-channel network to extract multiple features comprehensively applied to the background layer, generating a Deep Dream image. Multi-layer network loss is the sum of the output of the selected layer activation function, the loss is normalized at each layer, so the difference in the impact of each layer on the result will not be very large. (The purpose of this experiment is mainly to compare the effects of single-channel extraction features and multi-channel extraction features on the same background picture, to obtain a more suitable optimization method for deep dream technology. According to the results of this experiment, the use of all channel features to generate a Deep dream image is more abstract, the color of the image changes significantly from the use of a single channel to extract features, with changes in brightness, edges, and some details on the picture that have a noticeable effect. The experimental demonstration is described in Fig. 16 – Fig. 18:

Figure 16. Deep dream multilayer full-channel feature conversion A


Figure 17. Deep dream multilayer full-channel feature conversion B


Figure 18. Deep dream multilayer full-channel feature transformation C

3.4 Problems and Solutions

Through the results of deep dream processing of the original image, it is not difficult to find that the image has the following problems:

First, the output picture is relatively rough and noisy. Although the picture has undergone obvious changes under the feature processing of the above single channel and multi-channel, overall, the texture and pixel values of the image are disorganized, and this study aims to generate an abstract beauty image, and there are still some places that need to be improved and optimized to produce images that are accepted by the public

Second, the resolution of the image is relatively low. Both of the above methods have a significant drawback, the resolution of the image is low, such an image looks blurry, some details are still not very clear after processing, it is difficult to attract the viewer's attention, so it is still necessary to find ways to improve the deep dream, so that the image is more easily recognized and liked.

Third, the texture of output feature mode is similar between the various parts. Texture features are represented by the grayscale distribution of pixels and their surrounding spatial neighborhoods, i.e., local texture information. In addition, the repetitiveness of local texture information to varying degrees is global texture information. While a texture feature reflects the nature of a global feature, it also describes the surface properties of the scene corresponding to the image or image area. Unlike color features, texture features are not pixel-based features, which require statistical calculations in areas that contain multiple pixels. When retrieving texture images with large differences in thickness, density, etc., using texture features is an effective method. However, when there is little difference between the thickness and density of textures and other easily distinguishable information, it is difficult for the usual texture features to accurately reflect the difference between different textures of people's visual perceptions.

For the deep dream has shortcomings when using existing training models, this study reduces the image into a different proportion of the size, in order to more effectively shrink the image, removed the number of channels for the image , only retain the width and height of the image of the two pixel values, and then through the loop iteration, under the existing set scale of the zoom, the image shape size continues to change in a fixed proportion, and then for each changed picture, with the defined channel extraction feature of the deep dream technology optimized, after the end of the loop, the iteratively optimized image is adjusted to the original scale. The conclusion of the experiment is shown in Fig .19 – Fig. 21.


Figure 19. Deep dream optimization feature A

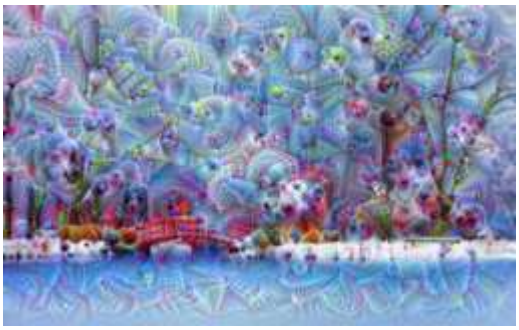
Figure 20. Deep dream optimization feature B


Figure 21. Deep dream optimization feature C

5. Conclusion

This paper is mainly to study the use of deep dream method to generate artistic images, and further optimize the algorithm

in the technology of deep dream, and compare the pictures generated by the original deep dream method with the optimized effect diagram, which shown in abstract artistic renderings with softer textures, clearer images, and richer content. The experimental results show that the style conversion using technology optimized by Deep Dream makes the demonization effect of the converted background picture more significant, the resulting image is softer, and the image texture is clearer. However, there are still some shortcomings in this technology that need to be improved, such as the short training batch of the algorithm, the small number of trainings, and the relative redundancy of the algorithm, which are important topics worth talking about in the future.

Acknowledgement

This work was supported in part by the Science Research Project of Anhui University of Finance and Economics under grant No. ACKYC20085.

References

[1] Gatys L.A., Ecker A.S., Bethge Texture Synthesis Using Convolutional Neural Networks ,2015.

[2] GATYS L A , ECKER A S, BETHGEM. Image style transfer using convolutional neural networks[C]/Proc of IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 2414-2423

[3] JOHNSON J , ALAHI A , FEI-FEI LI. Perceptual losses for real-time style transfer and super resolution[C]//Proc of European Conference on Computer Vision.Cham: Springer,2016:694-711.

[4] Shang Ronghua, Meng Yang, Zhang Weitong, Shang Fanhua, Jiao Licheng, Yang Shuyuan. Graph Convolutional Neural Networks with Geometric and Discrimination information[J]. Engineering Applications of Artificial Intelligence,2021,104.

[5] Zou Yan, Zhang Linfei, Liu Chengqian, Wang Bowen, Hu Yan, Chen Qian. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections[J]. Optics and Lasers in Engineering,2021,146.

[6] Ozkok Fatma Ozge, Celik Mete. Convolutional neural network analysis of recurrence plots for high resolution melting classification[J]. Computer Methods and Programs in Biomedicine,2021,207.

[7] Rekha Rajagopal. Comparative Analysis of COVID-19 X-ray Images Classification Using Convolutional Neural Network, Transfer Learning, and Machine Learning Classifiers Using Deep Features[J]. Pattern Recognition and Image Analysis,2021,31(2).

[8] Wei Bingzhe, Feng Xiangchu, Wang Kun, Gao Bian. The Multi-Focus-Image-Fusion Method Based on Convolutional Neural Network and Sparse Representation. [J]. Entropy (Basel, Switzerland),2021,23(7).

[9] Chen Yonghao, Cheng Ming Bao. Sports Sequence Images Based on Convolutional Neural Network[J]. Mathematical Problems in Engineering,2021, 2021.

[10] Tokarev K E, Zotov V M, Khavronina V N, Rodionova O V. Convolutional neural network of deep learning in computer vision and image classification problems[J]. IOP Conference Series: Earth and Environmental Science,2021,786(1).

[11] Liu Xun, Chen XiaoLin, Xia GuoQing, Chen HuaZhen, Ye HeZhong, Chen ZhanChi. A Multi-view Image Sets Classification Based on Graph Convolutional Neural Network[J]. Journal of Physics: Conference Series,2021,1948(1).

[12] Jiaqi Wang, Yibo Fan. Predictive fire image recognition based on convolutional neural networks[J]. Scientific Journal of Intelligent Systems Research,2021,3(6).

[13] Xiao Haixia, Zhang Feng, Shen Zhongping, Wu Kun, Zhang Jinglin. Classification of Weather Phenomenon from Images by Using Deep Convolutional Neural Network[J]. Earth and Space Science,2021,8(5).

[14] Wang Hao, Jiao Kaijie. Blind guidance system based on image recognition and convolutional neural network[J]. IOP Conference Series: Earth and Environmental Science,2021,769(4).

[15] Ji Zou, Chao Zhang, Zhongjing Ma. An Image Classification Algorithm for Plantar Pressure Based on Convolutional Neural Network[J]. TS,2021,38(2).

[16] Yiyue Luo, Yu Fan, Xianjun Chen. Research on optimization of deep learning algorithm based on convolutional neural network[J]. Journal of Physics: Conference Series,2021,1848(1).

[17] Chen Yuanyi. Research on Convolutional Neural Network Image Recognition Algorithm Based on Computer Big Data[J]. Journal of Physics: Conference Series,2021,1744(2).

[18] Kazuya URAZOE, Nobutaka KUROKI, Yu KATO, Shinya OHTANI, Tetsuya HIROSE, Masahiro NUMA. Multi-Category Image Super-Resolution with Convolutional Neural Network and Multi-Task Learning: Regular Section[J]. IEICE Transactions on Information and Systems,2021, E104.D(1).

[19] Zhang Qinghua. CNNA: A study of Convolutional Neural Networks with Attention[J]. Procedia Computer Science,2021,188.

[20] Ademola E. Ilesanmi, Taiwo O. Ilesanmi. Methods for image denoising using convolutional neural network: a review[J]. Complex & Intelligent Systems,2021,7(5).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Lingling Wang and Xingguang Dong designed the experiments, implemented the deep learning models, performed the experiments, analyzed the experiment results and wrote the paper.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work was supported in part by the Science Research Project of Anhui University of Finance and Economics under grant No. ACKYC20085.

Conflict of Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.