Classification of Famous Paintings Based on Convolutional Neural Network

Qiqi Cai
Lanzhou University, Lanzhou, Gansu, China
**Corresponding Author.*

Abstract

Neural network is a method to realize machine learning. In recent years, it has achieved good performance in completing many tasks. Convolutional Neural Network (CNN) is one of the most important algorithms in neural networks, and it performs well in the field of image classification. This paper studies the classification of famous paintings based on CNN. Firstly, I analyze the influence of network depth and iteration number on classification accuracy. Then, I propose an optimal network for paintings classification. To solve the problem of over-fitting, I use a data augmentation method. The processed data set is used as the input to the neural network, and then the input through convolution layers, subsampling layers, and two fully connected layers connected by the a dropout layer. Besides, I use the combination of Sigmoid and LeakyReLU functions as the activation function. Finally, compared with the traditional deep learning methods, the model proposed in this paper achieves better results with the accuracy of about 0.82.

Keywords: *Classification, CNN, data augmentation, dropout, activation function*

I. Introduction

With the rapid development and wide application of the computer technology, getting information from images has become the focus of many research. As a part of human civilization, art precipitates the civilization of The Times and human wisdom, contains a lot of information, and reflects the social ideology and life under the background of creation. Classifying famous paintings can help people explore the world of art. But the source and number of famous paintings are limited. There are a number of problems such as difficult collection, unbalanced distribution and so on. Traditional studies on the classification of famous paintings are mostly based on shallow learning, while there are still many problems to be explored in this task based on deep learning.

Convolutional neural networks originated from the concept of "receptive field" of individual neurons in the primary visual cortex of cats, which was proposed by Hubel and Wiesel [1]. Yann Lecun et al. proposed the CNN algorithm LeNet based on gradient learning and successfully applied it to handwritten digit character recognition [2]. Later, AlexNet was proposed by Krizhevsky et al, it is a classical CNN structure, and made a major breakthrough in image recognition tasks [3]. Kaiming He, Zhang X et al. proposed the structure of ResNet neural network [4], which is a great breakthrough in deep learning. The structure of CNN is a simulation of real biological neural network. The weight sharing can reduce the complexity of the network efficiently, and the feature that images can be directly input into the neural network, can avoid the complexity of data reconstruction in the process of feature extraction and classification [5]. In the field of painting, paper [6] uses Bayes Classifier, FLD Classifier and SVMs Classifier to classify according to the artists by extracting color, texture and other characteristic. Paper [7] proposes a method based on hybrid sparse CNN to analyze the brush features of authors with different styles.

According to the research, the application of CNN in the task of classifying famous paintings is still in its infancy. Therefore, this paper explores the factors that affect the classification accuracy of famous paintings based on CNN, and proposes an improved CNN model. First, I use a data augmentation method for the original data set to solve the problem of over-fitting. Then, I explore the influence of network depth and iteration number on the classification accuracy. Finally, I choose the combination of Sigmoid and LeakyReLU as activation functions, and apply the improved CNN model for feature extraction and classification.

ISSN: 0010-8189

II. Data Augmentation

2.1 Data set

I select the Kaggle competition dataset "Best Artworks of All Time" (BAAT) for researching and evaluating. BAAT has 25 artists with 3821 paintings. I select the ten categories with the largest number of samples as the original data set. The specific statistical information is shown in Table 1.

Table 1 The original data set

Artist	Rene magritte	Vincent van gogh	Pablo picasso	Pierre-auguste renoir	Francisco goya
Number	194	877	439	336	291
Name	Alfred sisley	Marc chagall	Edgar degas	Rembrandt	Paul gauguin
Number	259	239	702	262	311

2.2 Data augmentation method

Data augmentation is primarily used to reduce over-fitting. For the neural network model, with the deepening of the neural network, the parameters that need to be learned will also increase, which will more easily cause over-fitting [7]. Too many parameters will fit all the characteristics of the data set if it is small, but the final model does not fit the commonality between the data [7]. Besides, for classification problem, if there is a serious sample imbalance problem, it will also affect the performance of the model. There are many methods for data augmentation. In this paper, I mainly use geometric translations and color space translations.

2.2.1 Rotation
This method rotates the image to get new samples. People specify the center and the angle of rotation to rotate the image on the coordinate axes. Taking Fig 1 as an example, take the lower left corner of the original image as the center, rotate 90 degrees counterclockwise to get the new sample.

Fig 1: Rotation

2.2.2 Mirror symmetry
This method makes the image symmetrical left and right or up and down. In Fig 2, I get the new sample by making image symmetrical left and right.

ISSN: 0010-8189

Fig 2: Mirror symmetry

2.2.3 Change the contrast

Contrast plays a critical role in enhancing image texture. I use the method of histogram equalization (HE) to adjust the contrast of images in my augmentation method. HE is an effective image contrast enhancement method. In Fig 3, I enhance the contrast.

Fig 3: Contrast

2.2.4 Change the brightness

Brightness reflects how bright or dark an image is. In this paper, I use the Gamma transform to improve the dark details of the image (Fig 4).

Fig 4: Brightness

2.3 Sample generation

Table 1 describes the original data set, and the data set after data augmentation is shown in the Table 2. I increase the number of samples for each category to 900 and the total number of paintings in the new data set is 9000.

Table 2 The new data set

Artist	Rene magritte	Vincent van gogh	Pablo picasso	Pierre-auguste renoir	Francisco goya
Number	900	900	900	900	900
Name	Alfred sisley	Marc chagall	Edgar degas	Rembrandt	Paul gauguin
Number	900	900	900	900	900

ISSN: 0010-8189

III. Neural Network Architecture

3.1 Basic knowledge

Basic CNN has three main components which are convolution, activation, and pooling [9]. The output result of CNN is the specific feature space of each image [10]. In the process of image classification, firstly, input the space output to the fully connected layer. Then, it maps the input image to the label set. The final output is the category of the image [10].

The research shows that the selection of CNN structure will affect the performance of the classification results, and there are many factors that affect the classification accuracy, such as network depth, data size, iteration number, etc. In this paper, I just discuss the effects of network depth, number of iterations, and dropout layer on network performance.

3.1.1 Convolution layer

The function of the convolution layer is extracting features from each small part of the image. The input of convolution layer is an array. Then predefined convolution kernel (equivalent to the weight) extracts features from the image. Finally, multiple and add the convolution kernel and the corresponding bits of the matrix to obtain a result which is the output of the convolution layer [11]. When all the pixels are covered at least once, a convolution layer output can be generated. The convolution formula is

$$y(n) = \sum_{i=-\infty}^{\infty} x(i) * h(n-i) = x(n) * h(n) \quad [12] \tag{1}$$

*The sequences x(n) and h(n) are convolution variables, and * represents for convolution.*

The machine does not initially know what features the part to identify have, and it compares the output values obtained by acting with different convolutional kernel to determine which of the kernel represents the features of the image well. A higher convolution layer output value indicates a higher matching degree. And the more features of the image can be represented.

3.1.2 Pooling layer

The pooling layer mainly compresses the image and retain important information in the image by the way of downsampling on the premise of not affecting the quality of image. So, it is a good method of reducing the number of parameters, which can make the network simpler, make computational process easier and so on. It also plays an important role in reducing the over-fitting phenomenon.

3.1.3 Fully connected layer

The functions of convolutional and pooling layers are extracting features and reducing the parameters. And, they cannot map the input into the final label. So, the fully connected layer is applied to generate a classifier. Assuming that there are N categories in a classification task. So, the fully connected layer outputs a N dimension vector. Each number in this N dimension vector represents the probabilities of a particular category. The category of the image is equal to the category corresponding to the maximum probability.

3.1.4 Activation function

In a neural network, for each layer, the output is a linear combination of the input [13, [14], regardless of the structure of the network. However, data collected in the real life is often not linear and separable. And when dealing with such data, we need to introduce a nonlinear function or make a linear transformation. In neural networks, in order to avoid simple linear combinations, an activation function is added to the output of each layer, and the output of each layer becomes more complex than before, thus enhancing the expressive power of the neural network model.

Common activation functions are Sigmoid, Tanh, ReLU etc. Sigmoid is one of the most widely used activation function, and the shape of it is like an exponential function [15]. Additionally, it is similar to biological neurons in the physical sense [15].

$$f(z) = \frac{1}{1+e^{-z}} \tag{2}$$

It can be seen that the Sigmoid function (2) is continuous, smooth, strictly monotonic [16], and centrally symmetric, and is an excellent threshold function [17]. However, Sigmoid also has some flaws, and the biggest one is saturation. Sigmoid networks produce vanishing gradient within small number of layers. Moreover, the Sigmoid function may cause the offset phenomenon. Tanh is also a very common activation function [18]. Compared to Sigmoid, its output average is 0, so that it converges faster than sigmoid and reduces the number of iterations. However, tanh has equally soft saturation, thus it will lead to the vanishing gradient.

ReLU can reduce the vanishing gradient problem as compared to the above two activation functions to some certain extent [19]. As is seen from Fig 5, the ReLU is hard saturated when x is bigger than 0, and when x is smaller than 0, there is no saturation problem. Therefore, gradient is unable to vanish at x>0, thus reducing the vanishing gradient problem. This function allows to train deep neural networks directly in a supervised way without relying on unsupervised layer by layer pretraining [20]. Nevertheless, ReLU can cause neuronal "death" problems if the value of the input activation function is always negative, then the gradient of the back propagation process is constant to 0 [21], and the model cannot update the corresponding weight and bias parameters. In view of this, the LeakyReLU was introduced. LeakyReLU(Fig 6) is similar to ReLU, and the only difference is the function in the part that the value of input is less than 0. The output of ReLU is always 0, but for LeakyReLU the value is negative, and there is a small gradient.

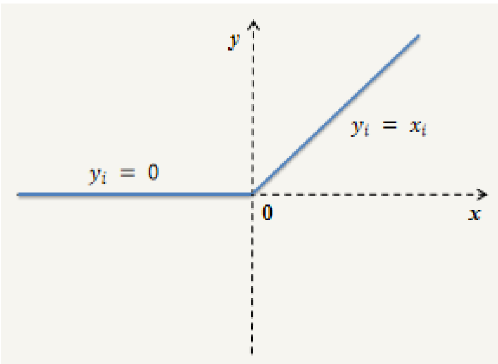

Fig 5: ReLU

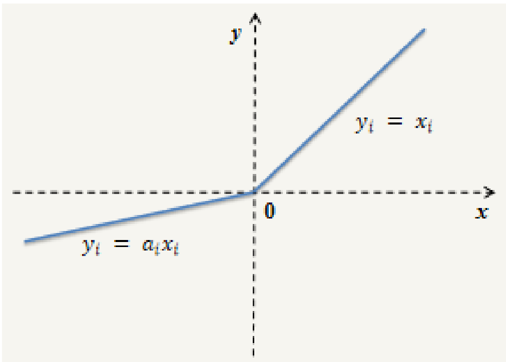

Fig 6: LeakyReLU

3.2 Network structure

According to the discussion of 3.1, using the LeakyReLU activation function can both alleviate the gradient vanishing and solve the problem of neurons "death". However, the output of the classification problem is a vector that all elements are probability, so the values are within the (0,1) interval. One advantage of the Sigmoid activation function is having the output range of (0,1), and therefore can be used in output layer, where the output represents probabilities of each category. Therefore, the convolutional layer uses LeakyReLU as the activation function, and the activation function of the output layer is Sigmoid.

ISSN: 0010-8189

The dropout layer is a structure that reduces the over-fitting. In the process of training deep learning network, it discards neural network units temporarily from the network according to a certain probability [22]. In this paper, I add a dropout layer between two fully connected layers.

The discussion and research suggest that the selection of CNN structure affects the performance of the classification results. After completing the influencing factor analysis, this paper optimizes the network by adjusting the network depths and the iterations number, and finally selects the optimal structure of CNN with 10 layers: an input layer, three convolutional layers, three pooling layers, two fully connected layers and a dropout layer [23]. And the iteration number is 30. The basic structure of the relatively optimal model can be seen in Fig 7.

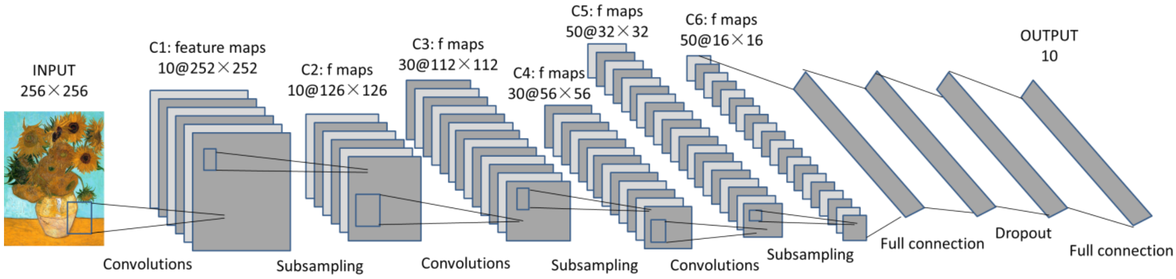

Fig 7: The optimal CNN model

Ⅳ. Experiments and Results

In order to explore the factors affecting the classification accuracy of famous paintings based on CNN, this paper observes the accuracy of models by changing the network depth and the number of iterations, and explores the factors that affect the model performance. To verify the strength of the method proposed in this paper, I use the new data set(Table 2) as the input of the model; I test the model without the dropout layer, and I train the model using common activation functions. Then, I compare the result of my method with them. Besides, I also compare the result with the traditional deep learning model to verify its effectiveness and feasibility.

4.1 Sample division

When dividing the data set, I use the random division method to select 80% as the train set and 20% as the test set.

4.2 Network depth

To explore the effect of network depth on the classification model, I change the number of convolutional and pooling layer. I try four sets of models with different network depth: 5, 7, 9 and 11. The specific information can be seen in Table 3.

Table 3 Depths of networks

Total layers	Input	Convolution	Pooling	Fully connected
5	1	1	1	2
7	1	2	2	2
9	1	3	3	2
11	1	4	4	2

By the Table 4, it is clear that the accuracy of the model with 5 layers is about 0.64. The accuracy is improved with the number of network layers increase, with approximate 0.68 and 0.69 at 7 layers and 9 layers, respectively. However, when the number of network layers reaches 11, the accuracy decreases to about 0.67. For neural networks, the deeper the network (the more number of layers), the better abstraction ability and the higher accuracy of the

ISSN: 0010-8189

model. It can be found that with the network gets deeper continuously, the model accuracy constantly increases. And when the number of network layers increases to the critical value, the accuracy begins to decrease, which means that it is difficult to train the deep network when the network becomes very deep.

Table 4 Different depths

Layer	5	7	9	11
Accuracy	0.64	0.68	0.69	0.67

Combined with the principle of back propagation, the output is calculated by forward propagation and then compared with the sample to get the error value.

$$E_{total} = \sum \frac{1}{2}(t\arg et - output)^2 \tag{3}$$

After continuous iteration of the parameter matrix, the error of the output is smaller and makes the output closer to the real category. The above process shows that the neural networks should constantly propagate the gradient in the back propagation process [24]. And with the increase of the number of network layers, the gradient will gradually vanish in the propagation process, which makes hard to adjust the weights of the previous network layers effectively.

So, in practice, for simple CNN, not the more the number of layers, the higher the accuracy. Based on this, Kaiming He et al. proposed the ResNet model [4], in which the author introduces a residual structure. When the gradient disappears and new content cannot be learned, the original content can be retained, so as to ensure that the performance of the network is not worse than that of the shallow network.

4.3 Number of iterations

To explore the influence of the number of iterations on the famous paintings classification model, this paper tries four models with different iteration numbers: 10, 20, 30 and 40. Except for the different number of iterations, the other structures(the model mentioned in 4.2 with 9 layers) and parameters of the model are the same.

Table 5 Different iteration

Iteration	10	20	30	40
Accuracy	0.63	0.66	0.69	0.68

By the Table 5, it shows that with the number of iterations increases, the accuracy is improved. At 30 iterations, the accuracy achieves about 0.69. However, the accuracy decreases to 0.68 when the number of iteration increases to 40. In short, when the number reaches a specific value, the accuracy of the model cannot be further improved, which will decrease or fluctuate.

4.4 Data augmentation

To verify the effectiveness of the data augmentation method in this paper, I apply the final model mentioned in chapter 3. This model trains unprocessed and processed datasets respectively. Table 6 shows the results.

Table 6 Data augmentation

Data augmentation	NO	YES
Accuracy	0.76	0.82

Table 6 illustrates that the accuracy of the model is significantly improved after using the data augmentation method described in this paper. This demonstrates the availability of this method, and also illustrates the necessity and effectiveness of data augmentation when training models, especially with the problem of serious unbalanced data set.

4.5 Effect of dropout

The optimal model is compared with the results of removing the dropout layer in it. The results are shown in Table 7, and it illustrates that adding dropout layers is crucial when doing classification. It also shows the effectiveness of the dropout layer when solving the problem of over-fitting.

Table 7 Dropout layers

Dropout layer	NO	YES
Accuracy	0.77	0.82

4.6 Activation function

Unlike the traditional CNN model, the method proposed in this paper uses the combination of Sigmoid and LeakyReLU as activation functions. To verify the effectiveness of it, I change the activation function of the optimal model (Fig 7) and compare the results (Table 8).

Table 8 Activation function

Activation function	Sigmoid	Sigmoid + ReLU	Sigmoid + LeakyReLU
Accuracy	0.76	0.81	0.82

Table 8 illustrates that the activation function affects the performance of the model and the accuracy is significantly improved after using the activation functions described in this paper. This indicates the effectiveness of the activation function selected in this paper.

4.7 Comparison with experiments by CNN, KNN

I test the final model in this paper, traditional CNN algorithm and KNN algorithm on the data set respectively. The datasets for experimental tests are the original data set(Table1) and the new one(Table 2). The experimental results are described in the Table 9.

Table 9 Different CNN structure

Algorithms	Data set	Accuracy
My model	Original	0.76
	New	0.82
CNN	Original	0.71
	New	0.79
KNN	Original	0.76
	New	0.79

Table 9 shows that in either model, the accuracy of the model is improved by using the data augmentation method. Besides, the performance of my model is significantly better than others.

V. Conclusion

In this paper, I explore the influence of network depth and the number of iterations on the classification of famous paintings based on convolutional neural networks, and propose a optimal model based on CNN for famous paintings classification: making data augmentation to data set, adding dropout layer in the model, and selecting the combination of Sigmoid and LeakyReLU as activation functions. My method performs favorably against the traditional deep learning methods.

ISSN: 0010-8189

In the future, there are still some work to continue. First, this paper considers only basic data augmentation methods and, in the future, more and more effective methods can be explored to improve the performance of the model. In addition, for famous paintings, the historical background, the experience of artists and other information also have a certain influence on the classification results. These effects can be taken into account in the next stage to increase the accuracy of the model.

References

[1] D.H. Hubel, T.N. Wiesel, "Republication of The Journal of Physiology (1959) 148, 574-591: Receptive fields of single neurones in the cat's striate cortex. 1959", Journal of Physiology, vol. 587, no. 12, pp. 2721-32, 2009.
[2] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[3] A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet classification with deep convolutional neura networks,” International Conference on Neural Information Processing Systems. Curran Associates Inc., pp. 1097-1105, 2012.
[4] K. He, X. Zhang, S. Ren, et al., “Deep residual learning for image recognition,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
[5] N. Kaushik, A.K. Sharma, “Cloud Computing and Security,” 2018.
[6] C. Liu, H. Jiang, “Classification of traditional Chinese paintings based on supervised learning methods,” IEEE International Conference on Signal Processing, Communications and Computing. IEEE, pp. 641-644, 2014.
[7] M.J. Sun, D. Zhang, J.C. Ren, et al., “Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-wash paintings,” IEEE International Conference on Image Processing, pp. 626-630, 2015.
[8] X.W. Jiang, S.C. Satapathy, L.X. Yang, S.H. Wang, Y.D. Zhang, "A Survey on Artifificial Intelligence in Chinese Sign Language Recognition,” Arabian Journal for Science and Engineering, 2020.
[9] N. Ketkar, “Convolutional Neural Networks”, Springer International Publishing, 2017.
[10] M. Matsugu, K. Mori, T. Suzuki, “Image processing apparatus for identifying an individual object, image processing method, and storage medium,” US, 2012.
[11] H.B. Cheng, L.Y. Yu, P.F. Li, et al., “Asynchronous Convolution Acceleration Algorithm Based on 4 × 4 Convolution Kernels,” 2018.
[12] G.S. Babu, P. Zhao, X.L. Li, “Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life,” International Conference on Database Systems for Advanced Applications. Springer, Cham, 2016.
[13] Nestor, L, SY, “Modelling the infiltration process with a multi-layer perceptron artificial neural network,” Hydrological Sciences Journal, 2006.
[14] X. Wu, H.J. Xu, X.L. Wei, Q. Wu, W.Y. Zhang, X.M. Han, “Damage Identifification of Low Emissivity Coating Based on Convolution Neural Network,” IEEE Access, 2020.
[15] H. Pan, H.X. Zhou, "Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce,” Electronic Commerce Research, 2020.
[16] L.F. Chen, B.Y. Peng, W.W. Gan, Y.Q. Liu, "Plaintext attack on joint transform correlation encryption system by convolutional neural network,” Optics Express, 2020.
[17] D. Xaveriayouh, “An Intelligent Knowledge-based Approach for the Management of Human Immunodeficiency Virus and Tuberculosis,” International Journal of Computer Applications, vol. 115, no. 2, pp. 30-37, 2015.
[18] P. Luo, “Research on Quantum Neural Network and its Applications Based on Tanh Activation Function,” Computer & Digital Engineering, 2012.
[19] H. Ide, T. Kurita, “Improvement of learning for CNN with ReLU activation by sparse regularization,” International Joint Conference on Neural Networks. IEEE, 2017.

ISSN: 0010-8189

[20] X. Glorot, A. Bordes, Y. Bengio, “Deep Sparse Rectifier Neural Networks,” Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 315-323, 2011.

[21] H. Ide, T. Kurita, “Improvement of learning for CNN with ReLU activation by sparse regularization,” International Joint Conference on Neural Networks. IEEE, 2017.

[22] H. Salehinejad, S. Valaee, “EDropout: Energy-Based Dropout and Pruning of Deep Neural Networks,” 2020.

[23] S. Jia, D. Yang, J. Liu, “Product Image Fine-grained Classification Based on Convolutional Neural Network,” Journal of Shandong University of Science and Technology (Natural Science), 2014.

[24] S. Rajakarunakaran, P. Venkumar, D. Devaraj, et al., “Artificial neural network approach for fault detection in rotary system,” Applied Soft Computing, vol. 8, no. 1, pp. 740-748, 2008.