Classification of picture art style based on VGGNET

Zihan Yang
Confucius International School Qingdao, Qingdao, China

Corresponding author and e-mail: Zihan Yang, 15121402058@stumail.sdut.edu.cn

Abstract. Aiming at the classification of various painting art styles in modern society, a method based on convolution neural network is proposed. The classification of various painting art styles is completed by using convolutional neural network. In VGG-19 (visual geometry) Based on Group-19 (Group-19) network, this paper proposes a picture style recognition algorithm based on visual geometry group 19 mixed transfer learning model. By using the feature extraction ability of convolution neural network, the convolution layer parameters of VGG-19 model on the source data are transferred to the target data model, and then the sparse automatic encoder SAE (sparse auto) is used. Finally, softmax classifier is used to classify the artistic style of pictures. The artistic style of painting can be roughly divided into: characters, flowers and birds, landscape. The VGG-19 model is used to predict the three kinds of picture art styles, and then the proportion of the predicted results is compared with the actual naked eye proportion, and the conclusion that VGG-19 model is more suitable for classification of painting art styles is obtained.

1. Introduction
Artistic style is a relatively stable and integral artistic feature presented by the interaction between the artist's creative personality and the language and situation of the artistic work. Style is the symbol of the maturity of the artist's creative personality and the symbol of the work reaching a higher artistic level. The same is true of the painting style. Style includes not only the artist's personal style, but also the genre style, era style and national style. Artistic style can be divided into artist style and artistic work style. Due to the differences of artists' world outlook, life experience, personality, cultural education, artistic talent and aesthetic taste, they have different artistic characteristics and creative personalities, forming different artistic styles. Therefore, our life will have a lot of picture art style classification.

With network and the development of the society, VGG model has gradually been known to more people and applied to the scientific research, such as VGG model and Google Net model jointly attended the 2014 Image Net image classification and positioning the challenge, though, with 0.6% of the (Google Net, but its focuses on research network depth affect the performance of the model, and successfully constructed the 16-19 layer deep convolution neural network, which makes the classification error rate dropped significantly (Alex Net error rate reduce doubled compared) and enhance the generalization ability of the model. So far, many networks still use VGG to extract image features. VGG Net is also composed of convolutional layer and full connection layer, which can be regarded as a deepened version of Alex Net. Moreover, the author believes that the small size convolutional kernel with only 3*3 can significantly deepen the depth of the network. VGG model is widely used in the recognition and classification of various kinds of pictures. For example, the rock
thin image recognition based on VGG model [5] and the clothing image recognition based on VGG-19 mixed migration learning model [4].

Artistic style is also a kind of comprehensive general characteristic in literary creation. The concept of style has been used in literary criticism. In China, it originated from Liu Xie’s literary mind and carving of dragons, while in the West it originated from the theory of style written by Bufeng in France. Style has a variety of meanings, such as the style of works, the style of writers, the style of the times, and the style of nationality. Due to the differences in world outlook, life experience, personality, cultural education, artistic talent and aesthetic taste, an artist often presents distinctive characteristics in the selection of themes, theme extraction, plot arrangement, characterization, structural layout, language use and other aspects, forming an unrepeatable style. Style is the mark of the spiritual characteristics of writers and artists. It is not reflected in some works or some artistic factors, but in the unity of content and form of all creation. Style is a sign of maturity of writers and artists. Not every writer has his own style. Only those writers who are hard to explore, find their own way and dare to innovate can form their own style.

In this paper, the author discusses and studies various kinds of painting art styles in modern society. The VGGNET model and questionnaire are used to classify the artistic styles of pictures.

2. Principle and algorithm
The main work of VGG is to prove that increasing the depth of network can affect the final performance of network to a certain extent.

VGG-19 convolutional neural network is used as the preprocessing model. Compared with the traditional convolution neural network, VGG-19 improves the depth of the network and uses the alternating structure of multiple convolution layers and nonlinear activation layers, which can extract image features better than a single convolution layer structure. VGG network is a deep convolution neural network developed by Oxford University Computer Vision Group and researchers from Google deep mind company. It explores the relationship between the depth of convolution neural network and its performance. By repeatedly stacking 3 * 3 small convolution kernels and 2 * 2 maximum pooling layer, the convolution neural network with 16 ~ 19 layers is successfully constructed. VGG network won the second place in ILSVRC 2014 and the champion of positioning project, with an error rate of 7.5% on top 5.

All VGG networks use 3 * 3 convolution cores and 2 * 2 pooling cores to improve the performance by deepening the network structure. The increase of network layers will not bring the explosion of parameters, because the parameters are mainly concentrated in the last three full connection layers. At the same time, the series connection of two 3 * 3 convolution layers is equivalent to a 5 * 5 convolution layer, and the series connection of three 3 * 3 convolution layers is equivalent to a 7 * 7 convolution layer, that is, the receptive field size of three 3 * 3 convolution layers is equivalent to a 7 * 7 convolution layer. However, the parameters of three 3 * 3 convolution layers are only about half of 7 * 7. At the same time, the former can have three nonlinear operations, while the latter has only one nonlinear operation, which makes the former more capable of learning features. Using 1 * 1 convolution layer to increase the linear transformation, the number of output channels does not change [1].

Compared with other methods, VGG-19 network has advantages and disadvantages in image recognition

Advantages: the method has great advantages in training time and verification time, and the verification accuracy is high. Several 3 * 3 convolution kernels are used to replace the larger convolution kernels (11 * 11, 7 * 7, 5 * 5) in Alex network. The whole network uses the same size convolution kernel size and maximum size. The performance can be improved by continuously deepening the network structure [2].

Disadvantages: VGG consumes more computing resources, uses more parameters, and takes up more memory.
Conclusion: by comparing the advantages and disadvantages of the appeal, it can be concluded that under the existing experimental conditions, the advantages of VGG-19 network are greater than the disadvantages. Therefore, this paper uses VGG-19 network to classify the art style of pictures.

2.1. VGG-19 network structure

VGG-19 convolutional neural network is used as the preprocessing model. Compared with the traditional convolution neural network, VGG-19 improves the depth of the network and uses the alternating structure of multiple convolution layers and nonlinear activation layers, which can extract image features better than a single convolution layer structure.

![Figure 1. The network structure of VGG-19.](image)

As shown in Fig. 1, VGG-19 is mainly composed of 5 convolutions, 5 down sampling layers and 3 fully connected layers. Each convolution has 2 to 3 convolution layers, in which the convolution kernel size is 3 × 3 and the convolution step size is 1. Maxtooling is used for down sampling, and the corrected linear unit (ReLU) is modified as the activation function. The number of nodes in the three full connection layers are 4096, 4096 and 1000 respectively. The input image size is 224 × 224, and the last layer of the network is a soft-max regression classifier, which classifies the input images by probability.

2.2. Convolution layer and activation function of network

In convolution neural network, convolution layer is used to simulate the characteristics of visual path, and different convolution check images are convoluted to obtain the response of different features. Let the j-th characteristic graph of layer 1 in the network: [4]

\[
X(j) = f(U(j)) \quad (1)
\]

\[
U(j) = \sum_{l-1} X(l-1) * K(l-1) + B(l) \quad (2)
\]

Where \(U(j)\) represents the net activation value of the jth node of convolution layer L, which is obtained by convolution with the output characteristic graph \(X(l-1)\) of the previous layer, and is the activation function; MJ Denotes the subset of the input eigengraph used to calculate \(U(j)\); \(K(l)\) represents the convolution kernel matrix; \(B(l)\) is the offset of the convoluted eigengraph; and "*" represents the convolution symbol.

In order to increase the nonlinear mapping ability of the whole network, activation function needs to be added to the network. ReLU is used as the activation function in VGG-19 network, which can effectively improve the gradient saturation problem of deep network in back propagation:

\[
f(x) = \max(0, x) \quad (3)
\]

Activate the image of the function, as shown in the following figure:
Figure 2. Shows the ReLU activation function.

2.3. Lower sampling layer and full connection layer of network
The lower sampling layer is mainly used to improve the anti-distortion ability of the network, while retaining the main features of the samples and reducing the number of parameters. In VGG-19 network, maxpooling is used as the down sampling function, that is, the maximum value in the image area is selected as the pooled value of the region. The expression of lower sampling layer is: [4]

\[
x^{(l)}_j = f(\tau^l_j \text{down}(x) + b^{(l)})
\]

(4)

Where, down represents the maximum pooled sampling function, \(\tau^l_j\) represents the coefficient corresponding to the \(j\)-th characteristic graph of layer \(L\), and \(f(\cdot)\) represents the ReLU activation function.

After stacking the convolution layer and the down sampling layer, the network generally has multiple fully connected layers. The full connection layer transforms the two-dimensional feature map of the convolution layer into a one-dimensional vector, and then uses softmax regression classifier to achieve the final classification.

2.4. Deep training of VGG-19 network model

2.4.1. VGG-19 network model pre training. Suppose that there are \(m\) samples in the source training set, and a single input sample is \((x^{(1)}, y^{(1)})\), \(x^{(1)}\) \(n\)-dimensional vector, which \(y^{(1)}\) is the real label of the sample. The overall cost function of the network model is: [4]

\[
J(W, b) = \frac{1}{m} \sum_{i=1}^{m} J(W, b; x^{(1)}, y^{(1)}) + \frac{\lambda}{2} \sum_{l=1}^{h} \sum_{i=1}^{j(l)} \sum_{j=1}^{j(l)} (W^{(l+1)}_{j})^2;
\]

The results show that the update amount of weight \(W^{(l)}_{j}\) and bias \(B^{(l)}_{i}\) are respectively: [4]

\[
\frac{\partial J(W, b)}{\partial W^{(l)}_{ij}} = \left[ \frac{1}{m} \sum_{i=1}^{m} \delta^{(l+1)}_{j} \delta^{(l+1)}_{i} \right] + \lambda w^{(l+1)}_{ij};
\]

(5)

\[
\frac{\partial J(W, b)}{\partial b^{(l)}_{i}} = \frac{1}{m} \sum_{i=1}^{m} \delta^{(l+1)}_{i};
\]

(6)

After many iterations, when the value of the contemporary value function is the smallest, the pre training on the source data set is completed, and then the convolution layer parameters obtained are transferred to the clothing image data set.
2.4.2. Retraining sparse automatic encoder. KL is the Kullback Leibler difference between $\rho$ and $\rho_J$ relative entropy; $\beta$ is the weight of controlling sparsity; $S$ is the total number of nodes in the network output end layer; $\rho$ is the expected sparse parameter close to 0; $\rho_J$ is the average output value of the jth node in the hidden layer.

The update quantity of sparse automatic encoder is as follows:

$$\frac{\partial J_{\text{sparse}}}{\partial b_{l+1}^{(i)}} = \frac{1}{m} \sum_{i=1}^{m} \left[ \delta_i^{(l+1)} + \beta ( - \frac{\rho}{\rho_j} + 1 - \rho_j - \rho ) \cdot j \cdot f'(z_i^{(l+1)}) \right]$$  \hfill (7)

$$\frac{\partial J_{\text{sparse}}}{\partial W_{l+1}^{(i,j)}} = \frac{1}{m} \sum_{i=1}^{m} \left[ \delta_i^{(l+1)} - \rho_j \cdot \frac{1}{1 - \rho_j} \cdot f'(z_i^{(l+1)}) \right]$$  \hfill (8)

2.4.3. Fine tuning of convolution layer parameters. Convolution layer parameter fine tuning in order to make the convolution layer of the new model better extract the clothing image features and avoid over fitting phenomenon, it is necessary to fine tune the convolution layer parameters in the network.

3. Experimental design

3.1. Advantages and disadvantages of VGG-19 network in image recognition compared with other methods:

Advantages: the method has great advantages in training time and verification time, and the verification accuracy is high. Several 3 * 3 convolution kernels are used to replace the larger convolution kernels (11 * 11, 7 * 7, 5 * 5) in Alex network. The whole network uses the same size convolution kernel size and maximum size. The performance can be improved by continuously deepening the network structure.

Disadvantages: VGG consumes more computing resources, uses more parameters, and takes up more memory

Conclusion: compared with the advantages and disadvantages of appeal, it can be concluded that under the existing experimental conditions, the advantages of VGG-19 network are greater than the disadvantages. Therefore, this paper uses VGG-19 network to classify the art style of pictures.

3.2. Experimental design process

This paper proposes a picture style recognition algorithm based on visual geometry group 19 hybrid transfer learning model. By using the feature extraction ability of convolution neural network, the convolution layer parameters of VGG-19 model on the source data are transferred to the target data model, and then the sparse automatic encoder SAE (sparse auto) is used Finally, softmax classifier is used to classify the artistic style of pictures.
Figure 3. Algorithm flow chart.

Figure 3 is a flow chart designed by using VGG model algorithm
① Create an experimental model of VGG-19, sample, modify the linear unit, improve the model.
② Establish a ReLU activation function and find the nodes of each layer of this function
③ The convolution neural network uses convolution layer to simulate the characteristics of the visual path, and convolutes the images with different convolution check to get the response of different features
④ Assuming the characteristic graph of each layer, the value corresponding to its function is obtained.
⑤ The activation function is added to the neural network to get the activation function curve.
⑥ The lower sampling layer is mainly used to improve the anti distortion ability of the network, while retaining the main features of the samples and reducing the number of parameters.
Then, according to the full connection layer, the two-dimensional feature map output by the convolution layer is transformed into a one-dimensional vector, and then the final classification is realized by using softmax regression classifier according to this vector.

Finally, input the data set that we created, and use the back error propagation algorithm to adjust the weight $W_i^{(l)}$ and bias $b_i^{(l)}$ in the convolution layer to minimize the overall cost function. When the cost error of the whole network structure is minimum, the parameters of VGG-19 network model are adjusted.

Finally, the recognition results of VGG model for picture art classification are obtained. After using the VGG-19 model to get the predicted proportion of the three kinds of art styles, and then using the way of questionnaire survey and comparison, we can further judge the differences between VGG-19 model and the naked eye, and get the differences and imitation accuracy of the three art styles.

Finally, the accuracy of VGG net based classification is judged, and the prediction result is good.

4. Results and discussion

VGG-19 convolutional neural network is used as the preprocessing model. Compared with the traditional convolution neural network, VGG-19 improves the depth of the network and uses the alternating structure of multiple convolution layers and nonlinear activation layers, which can extract image features better than a single convolution layer structure. VGG network is a deep convolution neural network developed by Oxford University Computer Vision Group and researchers from Google deep mind company. It explores the relationship between the depth of convolution neural network and its performance. By repeatedly stacking 3 * 3 small convolution kernels and 2 * 2 maximum pooling layer, the convolution neural network with 16 ~ 19 layers is successfully constructed. VGG network won the second place in ILSVRC 2014 and the champion of positioning project, with an error rate of 7.5% on top 5.

Therefore, VGG model is very suitable for the classification results of this type of pictures, and the conclusion drawn from the questionnaire survey can be very intuitive compared with the prediction proportion of VGG-19 model, which verifies that the VGG model has strong classification effect for this kind of art pictures and has stronger accuracy than other methods.

The following results are a comparison of the VGG model with the other three methods commonly used in image recognition.

| Methods                                    | The test of time/s | Verify the time/s | Training set accuracy (%) | Accuracy of verification set (%) |
|--------------------------------------------|--------------------|-------------------|---------------------------|---------------------------------|
| An appearance-based approach               | 2.0                | 1.0               | 30.0                      | 24.5                            |
| Retrieval via Deep Learning of Binary Codes | 548.0              | 1.0               | 77.5                      | 58.5                            |
| Using Convolutional Neural Networks        | 3540.0             | 5.2               | 81.5                      | 60.5                            |
| VGG-19                                    | 5348.5             | 17.2              | 98.5                      | 97.2                            |

It can be seen from the table that although the former two methods have more advantages in test time 2s, 548s and reaction time 1s, the recognition accuracy of the former two methods is only 24.5%. Compared with the second method, the recognition accuracy of the second method and the third method is significantly improved, and the accuracy on the verification set reaches 58.5% and 60.5% respectively. Compared with the previous two methods in literature, the third method can effectively
improve the recognition accuracy, but its recognition accuracy is lower than that of this method. The accuracy of this method is 3.7% higher than that of the third method.

The following are the results of VGG-19 model for classification and recognition of three types of picture art:

**Table 2.** Results of VGG-19 model on classification and recognition of three kinds of picture art.

| Type test             | Predicted class                        | Recognition rate (%) |
|-----------------------|----------------------------------------|----------------------|
|                       | Landscape painting                      |                      |
| Landscape painting    | 192                                    | 0                    | 96.0 |
| Figure painting       | 3                                      | 3                    | 96.5 |
| Flower and-bird painting | 189                              | 2                    | 94.5 |

It can be seen from table 1 that the prediction accuracy of VGG-19 model for various painting art styles is very high. In the test of 200 pictures in each of the three categories, the prediction accuracy of VGG-19 model for landscape painting, figure painting and flower and bird painting reached 96.0%, 96.5% and 94.5%, respectively. According to the prediction of the same number of pictures in the input set, the prediction ratio of 96% of landscape painting and 96.5% of figure painting by VGG model is more accurate than 94.5% of flower and bird painting. It can be reasonably speculated that VGG model is more likely to make errors in the prediction of art pictures with insignificant features.

For more than 90% of the subjects, the accuracy of the VGG model was more than 19%. Using the way of questionnaire survey, a large number of questionnaires are used to simulate the recognition accuracy of human naked eye for three kinds of art pictures, so as to verify the accuracy and simplicity of VGG model by comparing with the results of VGG model.

First of all, the data were normalized, and 200 questionnaires were obtained, 198 were recalled, 196 were valid, the recall rate was 96.0%, and the effective rate was 96.875%. The results of human naked eye resolution are normalized. The results are as follows:

**Table 3.** Questionnaire results.

| Type test             | Predicted class                        | Recognition rate (%) |
|-----------------------|----------------------------------------|----------------------|
|                       | Landscape painting                      |                      |
| Landscape painting    | 178                                    | 14                   | 90.81 |
| Figure painting       | 9                                      | 6                    | 92.35 |
| Flower and-bird painting | 174                              | 13                   | 88.77 |

Among 196 valid questionnaires that have been successfully recovered, the recognition accuracy of human naked eye for figure painting is the highest, reaching 92.35%, but it is still 4.15% compared with VGG-19 model, while the recognition accuracy of flower and bird painting and landscape painting by human naked eye is lower than that of mountain and water painting, only 90.81% and 88.77%, which are different from the predicted results of VGG-19 model by 5.19% and 5.19% respectively 5.73%.

5. Conclusions

In this paper, an art image recognition method based on VGG-19 hybrid transfer learning model is proposed. The combination of migration learning and unsupervised learning of sparse automatic encoder can solve the problem of feature extraction ability reduction caused by data set differences in the migration process. The experimental results show that, when extracting the features of art image, this method obtains better detail features of art pictures, and the recognition ratio of landscape painting, flower and bird painting and figure painting is more than 94%. Performance comparison experiments show that the accuracy of art image recognition in this paper is higher in training set and test set [1].

Compared with questionnaire survey and other models, VGG-19 model has the following advantages: less operation. That can save a lot of time and steps just by creating a model. Results the
accuracy is high and the probability of error is small. The model has been simplified into a few parts to reduce the difficulty of operation. In this paper, by establishing VGG-19 model, aiming at the problems of large amount of calculation and over fitting in the process of art picture classification training based on VGG-19 model, this paper proposes to optimize the network with batch normalization, global pooling and joint loss function while increasing the amount of data based on VGG-19 model. The average accuracy of the optimized network is increased by 1.4% compared with the original network.

This paper discusses and designs the classification of picture art styles based on VGG net. Through the model, get the prediction proportion of VGG-19 model for landscape painting, figure painting, flower and bird painting. 5%, 96. 0%, 94. 5%, respectively. Then through the questionnaire survey, we get the prediction ratio of the three types of painting art style by human naked eye, which are 90.81%, 92.35% and 88.77%. Then we compare the two and draw a conclusion: the classification of picture art based on VGG-19 is more accurate than that of the naked eye. The differences of the three types of picture styles were 4.15%, 5.19% and 5.73%, respectively. It can be concluded that VGG-19 network has more advantages, for the processing of a large number of data, has a higher accuracy, has a strong assistance for mass production.

References
[1] Huang Wenjian, Tang yuan. Tensorflow actual combat [M]. Beijing: Electronic Industry Press, 2017
[2] Yue Youjun, Tian Bokai, Wang Hongjun, Zhao Hui. Application of improved VGG model in Apple appearance classification [J]. Science, technology and engineering, 2020,20 (19): 7787-7792
[3] Ziqi Yin. A study on the classification of painting styles in pictures and anecdotes [J]. National art research, 2018,31 (03): 5-15
[4] Haoyang Tang, sunziwei, Wang Jing, Qian Meng. Clothing image recognition based on VGG-19 hybrid transfer learning model [j]. Journal of Xi'an University of post and telecommunications, 2018,23 (06): 87-93
[5] Bai Lin, Wei Xin, Liu Yu, Chongyang Wu, Lihui Chen. Rock slice image recognition based on VGG model [J]. Geological bulletin, 2019, 38 (12): 2053-2058