

Research on Dynasty Classification of Dunhuang Murals Based on Convolutional Neural Networks

Yingdi Wan *

Academy of Arts and Design, Qingdao City University, Qingdao, Shandong, 266106, China

* Corresponding author Email: tinawan0821@163.com

Abstract: Dunhuang Murals are one of China's intangible cultural heritages, which have received extensive attention. Identifying and classifying the dynasties of ancient murals quickly and accurately is extremely important for the study of Dunhuang Murals and the digital protection and inheritance. A method to classify Dunhuang Murals dynasties based on convolutional neural network (CNN) is proposed in this paper. First, the Dunhuang Murals data set is constructed from the mural materials; then, four convolutional neural network models with different depths and structures are constructed and trained; finally, the network model is tested using the test set and an appropriate classification model is selected. The experimental results show that for the Dunhuang mural image data sets of five different dynasties in this paper, the four models obtain high classification accuracy, and the accuracy of VGG11 and VGG19 reach 96%. Among them, the classification accuracy of murals in the Sui Dynasty and the Five Dynasties and Song Dynasties is lower than that of other dynasties in this paper.

Keywords: Intangible Cultural Heritage; Digital Protection; Convolutional Neural Network (CNN); Image Classification; Dynasty Identification.

1. Introduction

The Mogao Grottoes, as one of the world cultural heritages, located in Dunhuang City, Gansu Province, China, are the largest and most abundant Chinese Buddhist art sanctuary in the world. The grottoes contain some of the finest examples of Buddhist art spanning a period of 1000 years. These art works are treasures of human civilization, providing valuable material for studies of the foreign exchanges of China in different dynasties. But the dynasties of Dunhuang Murals are difficult to classify for ordinary people as the murals seem to be so much alike. Even the experts need analysis of a vast amount of historical documents. Therefore, murals classification has many challenges compared with other types of images classification.

1) Compared with general images, mural images have huge size, rich content and complex composition. Fig. 1(a) shows that every Bodhisattva has different facial expressions, actions, clothes and accessories in the mural of the late Tang Dynasty. In addition to the figures, a large number of buildings, moires, animals and flowers also can be seen in the murals, which can create high background noise and make the images classification more difficult.

2) The features of the same kind of objects are quite different in the same dynasty. As shown in Fig. 1(b) and Figure. 1(c), the figures in the murals are the patrons in the Yuan Dynasty, but they look so different, because one was painted with clear lines and the other was painted with bright colors.

3) The features of the same object have a high similarity from different dynasties. Fig. 1(d) shows the dancers in the middle Tang Dynasty, and Fig. 1(e) shows the dancers in the Song Dynasty, but they have great similarities in face features.

4) Fig. 1(f) shows the murals have been discoloured with the passage of time, that will also make image classification more difficult.

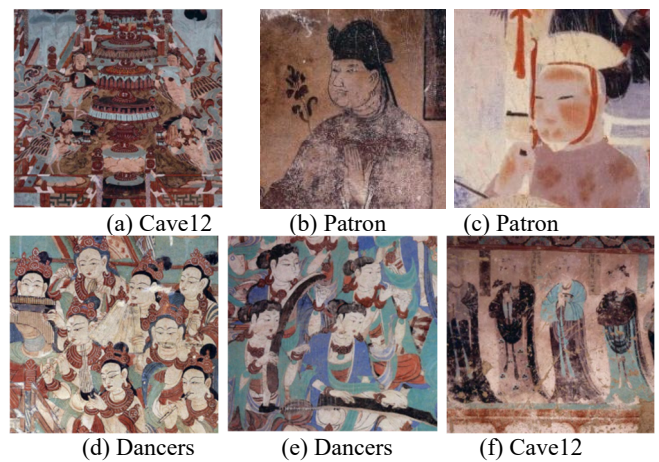


Fig 1. Examples of some images of Dunhuang Murals

The main contributions are as follows: First, the Dunhuang Murals data set is constructed from the mural materials; Second, four convolutional neural network models with different depths and structures are constructed and trained; finally, the network model is tested using the test set and an appropriate classification model is selected. It affords advantages for researchers on classifying Dunhuang Murals dynasties which can bring the digital protection on Dunhuang Murals to a wider audience.

2. Background

In recent years, a large number of studies have used computer technology to aid mural classification. The research methods in this field are mainly divided into traditional methods and Deep learning-based methods.

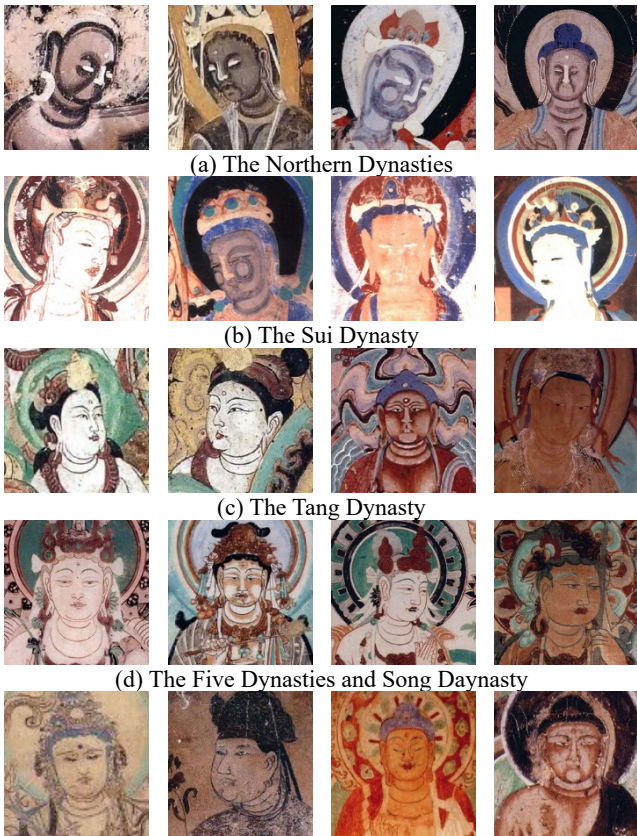
Most of the traditional classification methods are based on the bag-of-words (BoW) model which, as a middle level feature, can narrow the gap between low-level visual features and high-level semantic features [1]. How to choose the keywords of feature extraction affect the final classification results. In 2013, a method was introduced based on the

essential contour structure of the mural images, which leads to high classification accuracy [2]. In 2017, another method was introduced based on the mean, variance, etc of the portrait character eigenvalues, which can classify the murals into different dynasties [3].

Since the concept of deep learning was put forward in 2006, it has been widely used in computer vision technology. Convolutional neural network (CNN) based on deep learning models has a good performance in natural image classification, and has also been applied to the classification of murals. In 2012, Alex Krizewski and others proposed AlexNet network in the Image Net Large Scale Visual Recognition Challenge (ILSVRC) which is widely used in natural image classification [4]. In [5], the Res Net50 was selected as the basic framework and a new network model was proposed. The new network layer was used to replace the 49th layer of Res Net50, which improved the dynasty recognition accuracy of Dunhuang Murals. In [6], A prototype system of ancient mural intelligent classification based on AlexNet and TFNet was developed. On the other hand, it completed the automatic classification function of mural images, and saved the classification results to the database to upload murals, which greatly facilitated the unified management of murals. In 2021, an Inception-v3 model integrated with transfer learning was proposed to identify and classify the dynasties of ancient murals which could effectively identify the dynasty to which the murals belong [7].

3. Materials and Methods

3.1. Experimental Data



(e) The Western Xia Regime and the Yuan Dynasty
Fig 2. Examples of Dunhuang mural dataset

The data set is obtained from *a series of books for appreciating Dunhuang art and the entire works of Dunhuang murals in China*. By the content depicted, these murals, spanning a period of 1000 years are divided into Buddha statue paintings, sutra converted paintings, narrative paintings and donor portraits, etc. According to the historical documents, five data sets are adopted to validate the proposed methods. They are the Northern Dynasties, the Sui Dynasty, the Tang Dynasty, the Five Dynasties and Song Daynasty, the Western Xia Regime and the Yuan Dynasty. As shown in Fig. 2, the images in each row belong to the same Dynasty.

To improve the quality of the data sets, we intercept 881 clear images of faces from Buddha, Bodhisattva, Buddhist monk, supporters. For all five data sets, we split the labeled samples into two subsets, i.e., 831 samples for training set and 50 samples for testing set.

3.2. Experimental Method

3.2.1. Convolutional Neural Network (CNN)

Convolutional neural network (CNN) based on deep learning method has many advantages, including local connections and shared weights, which make it easier to train with lower complexity [8]. Additionally, CNN is invariant with respect to translation, scale and rotation, which make the network with better robustness and fault-tolerance.

A CNN layer contains a convolution layer and a pooling layer. The convolution layer use filters to filter the matrix information to extract features from the images. The pooling layer always appears between two convolution layers, and can effectively reduces the size of the matrix to achieve sampling. In addition, it can also reduces the number of parameters in the input layer of the fully connected layer, and reduces the amount of data operations to avoid over-fitting.

The fully connected layer is similar to a part of the traditional neural network and is used to exports result. The data processed by convolution layer and pooling layer can be input to the fully connected layer. The output result is to combine all data into global features to obtain the final classification result. The activation function is used to provide nonlinear modeling for the network by introducing the nonlinear elements into the neural network, so that the neural network can complete the nonlinear mapping, and allow the deep neural network to have the ability to learn nonlinearity. Therefore, it is very important to add activation function in the process of deep neural network learning. The common activation functions include Sigmoid, Elu, Tanh and ReLU [9]. In this paper, the adopted ReLU is a simple nonlinear operation that take the maximum value of 0 to x which can improve the performances. The ReLU function is shown in formula (1):

$$f(x) = \max(0, x) \quad (1)$$

3.2.2. Dynasty Classification Model Structure

The VGG network architecture was introduced by Simonyan and Zisserman from the University of Oxford in 2014 [10]. This network use 3×3 small convolutional kernels stacked on top of each other and 2×2 max pooling layers to build the CNN with 16 to 19 layers. It is generally believed that deeper networks have better expression ability than shallow networks, and can complete more complex tasks.

Table 1. The structural parameters of different VGG networks

VGG11		VGG13		VGG16		VGG19	
Layer number	Size	Layer number	Size	Layer number	Size	Layer number	Size
C1-1	224*224*64	C1-1	224*224*64	C1-1	224*224*64	C1-1	224*224*64
P1	112*112*64	C1-2	224*224*64	C1-2	224*224*64	C1-2	224*224*64
C2-1	112*112*128	P1	112*112*64	P1	112*112*64	P1	112*112*64
P2	56*56*128	C2-1	112*112*128	C2-1	112*112*128	C2-1	112*112*128
C3-1	56*56*256	C2-2	112*112*128	C2-2	112*112*128	C2-2	112*112*128
C3-2	56*56*256	P2	56*56*128	P2	56*56*128	P2	56*56*128
P3	28*28*512	C3-1	56*56*256	C3-1	56*56*256	C3-1	56*56*256
C4-1	28*28*512	C3-2	56*56*256	C3-2	56*56*256	C3-2	56*56*256
C4-2	28*28*512	P3	28*28*256	C3-3	56*56*256	C3-3	56*56*256
P4	14*14*512	C4-1	28*28*512	P3	28*28*256	C3-4	56*56*256
C5-1	14*14*512	C4-2	28*28*512	C4-1	28*28*512	P3	28*28*256
C-2	14*14*512	P4	14*14*512	C4-2	28*28*512	C4-1	28*28*512
P5	7*7*512	C5-1	14*14*512	C4-3	28*28*512	C4-2	28*28*512
FC-1	1*1*4096	C5-2	14*14*512	P4	14*14*512	C4-3	28*28*512
FC-2	1*1*4096	P5	7*7*512	C5-1	14*14*512	C4-4	28*28*512
FC-3	1*1*6	FC-1	1*1*4096	C5-2	14*14*512	P4	14*14*512
SM	1*1*6	FC-2	1*1*4096	C5-3	14*14*512	C5-1	14*14*512
		FC-3	1*1*6	P5	7*7*512	C5-2	14*14*512
		SM	1*1*6	FC-1	1*1*4096	C5-3	14*14*512
				FC-2	1*1*4096	C5-4	14*14*512
				FC-3	1*1*6	P5	7*7*512
				SM	1*1*6	FC-1	1*1*4096
						FC-2	1*1*4096
						FC-3	1*1*6
						SM	1*1*6

Because the choice of network architecture has a great influence on the effect and efficiency of image classification. Therefore, this Dynasty Classification model adopts the network architecture to the depth of 11 to 19 layers on the basis of VGG network. The structural parameters of VGG networks are shown in Table 1.

We choose RGB 3-channels mural images with the size of 224×224 as the input to the model. In order to retain the image information more completely and improve the accuracy of mural classification, 3×3 kernel and 1 step size are selected to run convolution, 2×2 kernel and 2 step size are selected for pooling. The rectified linear unit (ReLU) is applied as activation function. Dropout is applied in the fully connected layer by dropping neurons randomly to avoid over-fitting. Softmax is used in the fully connected layer to classify the mural images, which can map the fully connected layer to values from 0 to 1, that can represent the probability value of each category. Its definition function is shown in equation (2):

$$S_i = \frac{e^{V_i}}{\sum_i^c e^{V_i}} \quad (2)$$

The mural classification network models use cross entropy as the loss function. As shown in formula (3), the loss function use the distribution probability q to express the accuracy of probability distribution P . The proximity of the two probability distributions is expressed by calculating the distance between the two probability distributions. Where P is the tag value and q is predicted value. The shorter the distance between the two probability distributions, the closer the predicted value is to the tag value.

$$H(p, q) = -\sum_x p(x) \log q(x) \quad (3)$$

4. Experimental Verification and Result Analysis

4.1. Experimental Environment

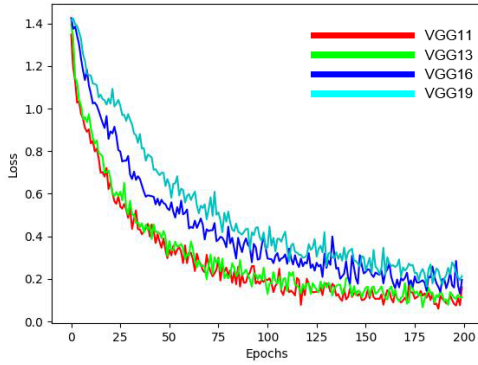
Model training and testing are carried out on Windows10 operating system, and Python is selected as the programming language. I7-10700 processor, NVIDIA GTX3060 GPU, Memory 16GB.

4.2. Model Comparison Experiment

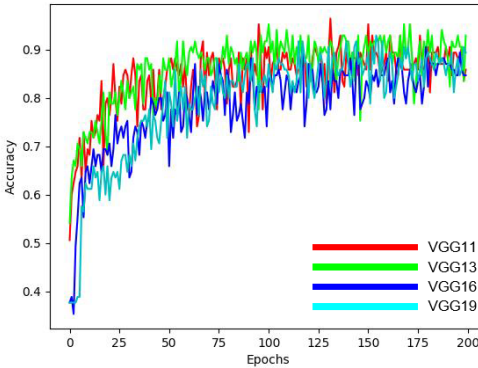
831 images of each dynasty of murals are in the training set. After training and testing the models for many epochs, The learning rate is set to 0.0001, the batch size is set to 20 and the number of epochs is set to 200.

Four models are trained to classify the mural data sets of five dynasties, and the loss value and accuracy are visualized to analyze the convergence performance of the model, as shown in Fig. 3(a), X-axis is the number of epochs, Y-axis is the training loss. In Fig. 3(b), X-axis is the number of epochs, Y-axis is the training accuracy.

It can be seen from Fig. 3(a) that as the number of epochs increases, the VGG11 model training loss is gradually decreasing. After roughly 100 epochs, the training loss curve become stable. In Fig. 3(b), the VGG11 model training accuracy is gradually increasing and converges between 0.85 and 0.9. After 100 epochs, the curve stabilizes, and its accuracy is about 0.9.



(a) Comparison of training loss



(b) Comparison of classification accuracy

Fig 3. Experimental results of network training

Compared to VGG11, although VGG16 model and VGG19 model have deeper layers, but they have similar convergence

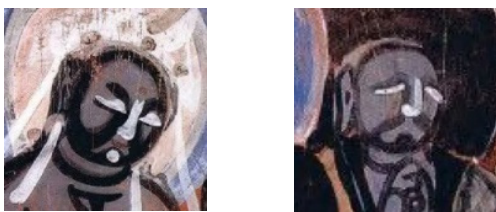
trend of loss curve and higher training loss. The convergence values and the accuracy of VGG13 model are basically the same as the VGG11 model. In the training process, when the loss curve fluctuates, its training accuracy will also decline. By comparing four networks with different structures, it is known that the deeper the network is, the greater the convergence value of the loss function is and the lower the training accuracy is. Therefore, when the network depth is enough, continuing to increase the network depth does not help the network convergence.

4.3. Model Accuracy Analysis

In this experiment, we compare and analyze the accuracy of VGG model in the category recognition of each dynasty on the 11th to 19th layers of network structure. The accuracy of each model is shown in Table 2. The classification accuracy of VGG11 in the Sui Dynasty is lower than that in the Five Dynasties and Song Dynasty and the classification accuracy of other dynasties is more than 95%. In VGG13, only the classification accuracy of the Northern Dynasties is more than 95%, and that of other dynasties is about 90%. The lowest classification accuracy of VGG16 is about 60% in the Five Dynasties and Song Dynasty, and the classification accuracy of other dynasties is more than 90%. The classification accuracy of VGG19 in the Sui Dynasty is lower than that in the Five Dynasties and Song Dynasty, and the classification accuracy of other dynasties is more than 90%. These results indicate that VGG11 achieves a high accuracy of 96% on fewer training parameters.

Table 2. Comparison of accuracy under different dynasties /%

Model	The Northern Dynasties	The Sui Dynasty	The Tang Dynasty	The Five Dynasties and Song Daynasty	The Western Xia Regime and the Yuan Dynasty	AVG
VGG11	100	90	100	90	100	96
VGG13	100	90	90	90	90	92
VGG16	100	90	100	60	90	88
VGG19	100	90	100	90	100	96



(a) Bodhisattva (b) Bodhisattva
Fig 4. Murals of the Northern Dynasties



(a)The Sui Dynasty (b) The Northern Dynasties (c) The Five Dynasties and Song Dynasty (d) The Tang Dynasty
Fig 5. Examples of misclassified murals

Through comparison, it can be seen that the classification accuracy of four models in the Northern Dynasty is generally high. As shown in Fig. 4, compared with other dynasties, the characters in Dunhuang murals in the Northern Dynasty are shown with white nose and eyes which made the characteristics of the characters are significant. Therefore, all

the models in the experiment can recognize the mural images of the Northern Dynasty. The data sets with the most misclassification examples are in the Sui and Tang Dynasties and the Song and Yuan Dynasties. As shown in Fig. 5, the feminine characteristics of the characters are quite obviously. Due to the similarity of colors and painting styles, it is difficult to extract features in Dynasty classification. As a whole, VGG11 and VGG19 have the highest classification accuracy, followed by VGG13, and VGG16 has the lowest classification accuracy.

In the misclassification examples, Fig. 5(a) shows that the mural of the Sui Dynasty is wrongly divided into the Northern Dynasty. We can see the color fades to gray which is similar to the color of the mural in the Northern Dynasty. In addition, the character's beard is easy to be confused with the rough line drawing style in the mural of the Northern Dynasty. Fig. 5(c) shows that the mural of the Five Dynasties and Song Dynasty is wrongly divided into the Tang Dynasty. The facial features, the colors and the painting styles of the the mural look very similar and are difficult to distinguish. To a certain extent, the above misclassification examples are due to the similarity of colors and painting styles, which lead to the reduction of accuracy in the classification of murals in the Sui Dynasty, the Five Dynasties and the Song Dynasty.

5. Conclusion

In this paper, we propose a method based on convolutional neural network to classify the dynasties of ancient murals, which is important for the research of Dunhuang murals and the digital protection of inheritance. From the results, we can see that the classification accuracy of VGG11 and VGG19 are both 96%, which are higher than the traditional artificial classification. The classification accuracy of murals in the Sui Dynasty and the Five Dynasties and Song Dynasties is lower than that of other dynasties.

Although the model in this paper has been implemented with a high classification accuracy. The examples of misclassification also show that the models need further research. This method can be applied to the classification of patterns, clothing and headwear in ancient murals which can save the human and material resources spent on manual classification.

References

- [1] Zhao Lijun, Tang Ping, Huo Lianzhi, et al. Review of the bag-of-visual-word models in image scene classification[J]. *Journal of Image and Graphics*, 2014, 19(3): 11.
- [2] Tang Dawei, Lu Dongming, Yangbing, et al. Similarity metrics between mural images with constraints of the overall structure of contours[J]. *Journal of Image and Graphics*, 2013, 18 (8): 968-975. 2014, 19(5): 708-715.
- [3] Hao Yabo. Research and Implementation on Classification Algorithm with People of ancient Chinese Murals Based on Style Characteristics[D]. Tianjing: Tianjin University, 2017.
- [4] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Image Net Classification with Deep Convolutional Neural Networks[J]. *Advances in Neural Information Processing*, 2012, 25(2), 1097–1105.
- [5] Li Xiangyu. Based on in-depth learning, Dunhuang mural era identification and damage detection[D]. Harbin: Harbin Institute of Technology, 2019.
- [6] Cui Hongyan. Research on Ancient Mural Classification Method Based on Deep Learning[D]. Taiyuan: Taiyuan University of Science and Technology, 2020.
- [7] Cao Jianfang, Yan Minmin, Jia yinming, Tian Xiaodong. Application of Inception-v3 model integrated with transfer learning in dynasty identification of ancient murals[J]. *Journal of Computer Applications*, 2021, 41(11): 3219-3227.
- [8] Cheng Bing. Research on Automatic Image Annotation Technology based on Convolutional Neural Network[J]. *Electronics World*, 2019(16): 124-126.
- [9] Zhang Lulu. Research on Facial Expression Recognition Method Based on Convolutional Neural Network[D]. Shijiazhuang: Hebei University of Science & Technology, 2019.
- [10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. *Computer Science*, 2014.