Research on Data Enhanced Ancient Pictogram Recognition Method Based on Convolutional Neural Network

Lily Tian
Minzu University of China
Beijing, China
86-15210661285
1185214958@qq.com

Yutong Zheng
Minzu University of China
Beijing, China
86-15911039377
zhengyutong68@aliyun.com

Qiao Cui
Minzu University of China
Beijing, China
86-15201616033
521071300@qq.com

ABSTRACT
As the carrier of national culture, words and pictograms record the unique culture and history of each nation, but the number of existing ancient pictogram is very small, and it is difficult to collect them, which makes it difficult for the academic research of ancient pictogram and the recognition by deep learning. In addition, due to the preservation environment and their own particularities, the traditional data enhancement methods will cause problems such as wrong data label, inability to simulate real scenes, etc. So, it can't effectively expand the large-scale data. To solve these problems, this paper proposes a set of data enhancement methods for small data sets and natural scenes. For the small data set enhancement method, firstly, we use artificial data enhancement to enhance original data, and then a limited random affine transform is used to limit the extent and extent of the enhancement. For natural scenes, we use the DCGAN to fuse the natural scene image and the ancient pictogram to simulate the natural environment. Finally, the paper designs a neural network model to recognize the ancient pictogram. It is proved that the data enhancement method can solve the problem of insufficient data, and finally achieve 99% accuracy.

CCS Concepts
• Computing methodologies → Object identification.

Keywords
Data Enhancement; Convolutional Neural Network; Ancient Pictogram Recognition

1. INTRODUCTION
The Poya songs book is the third batch of national intangible cultural heritage in China. It is a collection of folk songs circulating in the Zhuang area of Funing County, Yunnan Province. It records the Zhuang folk songs on the coarse cloth with the original pictogram. It is the only document that used pictograms to record folk songs. At the same time, the very popular deep learning technology in recent years is based on the data itself to extract and learn, and finally to classify and recognize. However, its core requirement is the amount of data. Therefore, it is particularly important to explore an effective method for data enhancement and to combine the deep learning to recognize ancient pictograms.

This paper attempts to use the Poya songs book [1] (Figure 1) as an example, use image segmentation technology to remove redundant information, and design an effective method for data enhancement. We selected 10 kinds of pictograms as research objects, using convolutional neural network technology to process the pictogram, and finally achieved a higher recognition rate.

The main contributions of this paper are as follows: 1) Proposes a set of data enhancement methods for small datasets, including artificially extended datasets and limited random affine transformations, which lays the foundation for the combination of ancient pictogram and convolutional neural networks; 2) proposes the ancient pictogram data enhancement scheme in natural scenes, using DCGAN to generate some of the ancient pictogram data pictures in the original scene, which is beneficial to improve the generalization of the model; 3) Constructed the first ancient pictogram database of the Poya songs book in Funing County, Yunnan Province; 4) For the first time, a classification and recognition model of ancient pictogram based on convolutional neural network is proposed and the accuracy rate can reach 99%.

The content of this paper is arranged as follows: the first section briefly describes the current research status of ancient pictograms in the field of computer vision; the second section focuses on the data enhancement methods in small data sets and natural scenes; the third section describes the design and parameter setting of the whole experimental model for Poya songs book. Finally, we summarize the methods we proposed and the areas that can be improved, and propose the next research direction.

Figure 1. The Poya songs book are simple and vivid, and each pictogram represents a song.

HPCCT 2019, June 22–24, 2019, Guangzhou, China

ACM ISBN 978-1-4503-7185-8/19/06…$15.00
DOI: https://doi.org/10.1145/3341069.3342993

2. Research status

For the exploration of character recognition, people have never stopped. In 1929, German scientist Taushechk proposed the concept of OCR, which was the earliest character recognition technology that can be used to quickly and accurately recognize images[2]. In the 1950s, Frank et al. first proposed the concept of a perceptron model, and artificial neural networks began to rise[3]. In 1998, Yann LeCun et al. proposed the LeNet model, which uses convolutional neural networks to process images and extract features, which is excellent in the recognition of handwritten digit sets[4]. In 2006, Geoffrey Hinton published a paper on science to prove the excellent performance of deep learning, which made people pay attention to the convolutional neural network again[5]. The field of character recognition also began to pay attention to deep learning[6][7][8][9][10]. In 2014, some scholars proposed a method of migration learning, which solved the problem of insufficient sample of small dataset to some extent[11][12]. In 2014, Ian Goodfellow proposed the GAN model, which allowed the generator and discriminator to constantly confront each other to generate the desired result, which gave a new idea in the field of image processing[13].

3. Data Enhancement Research ideas

The collection of ancient pictogram is different from other pictograms. It appears relatively early, mostly used to represent a specific scene. Because it is not retained as historical materials, and the preservation form is relatively random, so the damage is serious and the collection is difficult. Moreover, the pictograms of ethnic minorities are only passed down as a national culture, and collection is more difficult. Due to this special background, the same pictogram in the retained literature has a small amount of data, often with only a single digit, or even only one sample.

In this background, the most direct way for small sample data sets is artificial data enhancement. We have designed an efficient artificial data enhancement method that allows different people to draw the same pictogram in different external factors, and then use the limited random affine transformation to perform simple deformation on the data. Considering that the small sample data enhancement method does not change the background of the pictogram, we combine the pictogram with the natural scene to simulate the natural scene. The data enhancement method will be described in detail later.

3.1 Traditional method

Traditional methods generally flip, crop, rotate, translation, zoom, change saturation, change color channels, add noise, etc., mainly to change the numerical information of the image or to delete the information. These methods have some drawbacks. For example, for cropping, it is very likely that part of the original image is cropped to another image, causing the label to change; and rotation is very easy to appear as the error of pictogram 2 after the rotation of the pictogram 1. A similar problem occurs when noise is added to the original image.

In addition, these methods are based on the preservation of the original structural features of the image, essentially without changing the existing form of the features. The feature extraction of deep learning is entirely from the image itself. If the training data structure features are similar, then the multi-round training only strengthens the sensitivity of the network to the features in the image, which is not conducive to improving the generalization ability of the model.

Moreover, the traditional data enhancement method can't change the background of the pictogram, nor can it simulate the pictogram in the natural scene, which is not conducive to the actual needs. Obviously, for the ancient pictogram, the traditional data enhancement method is not applicable. For the above problems, we have designed a data enhancement scheme.

3.2 Small data set

3.2.1 artificial data enhancement

In order to avoid the influence of redundant information in

the original image, we use image segmentation to remove irrelevant information in the original image. The original image and the segmented image are shown in Figure 2 and Figure 3.

Figure 2. Original Picture of 10 Kinds of Poya Songs Book

Figure 3. 10 class segmented pictures

Initially we used a simple method, such as changing the lighting, simple deformation, or hand-drawn photos, as shown in Figure 4, but this kind of data set is not ideal when tested. Therefore, we have designed a complete data set enhancement solution: draw different pictograms on different backgrounds by different people using different colors and different thickness pens, as shown in Figure 5.

This method not only retains the pictogram features, but also randomly increases the deformation characteristics of the pictograms. In addition, due to the different background materials, the image features of the pictograms are randomly added after taking pictures with a high-definition camera, which makes the training data produces a certain amount of noise and improves the robustness of the system.

Figure 4. original expansion effect

Figure 5. Expanded new data set

In terms of the number of enhancements, we need to allocate the sample size equally for each type of pictogram based on the number of types of pictograms and the number of samples included in each data set. We have created a total of 10 categories, totaling more than 2,300 data samples, and after data cleaning; the remaining 2000 pictures are reserved for use.

3.2.2 A limited random affine transform

The artificial data enhancement method has achieved good results, but it requires a lot of time and labor costs. We propose a method of data enhancement using a limited random affine transformation.

Random affine transformation refers to a vector in space, which is randomly mapped to another vector space by a certain method. The formula is expressed as:

$$\vec{y}=A\vec{x}+\vec{b} \qquad (1)$$

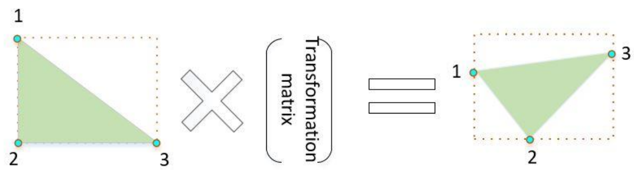

Figure 6. Schematic diagram of affine transformation

It can retain the main features of the pictogram well when the pictogram shape is changed, and the original pictogram is deformed to some extent. But for the deformation of pictograms under various factors, we need to control the deformation range of the affine transformation so that it is not too exaggerated. We propose a limited random affine transformation to constrain the randomness of the affine transformation. This method allows the pictogram to add the random deformation property while retaining the key features, and some of the pictograms have information missing. The forward noise data improves the robustness of the model. The process is shown in Figure 7. The enhanced pictogram is shown in Figure 8:

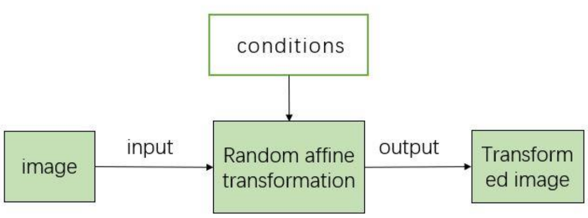

Figure 7. Schematic diagram of defining an affine transformation

Figure 8. original pictogram(up) and data enhanced pictogram(down)

3.3 Natural scene

In order to make the pictogram background closer to the natural scene, we use a combination network (DCGAN) of convolutional networks and GAN to generate images of the fusion of pictograms and natural scenes.

The training of DCGAN is a process of continuous confrontation and iteration of the generation model and the discriminant model. The generated data introduces environmental noise while retaining the main features of the original data. In order to ensure the performance of the generated model and the discriminant model, training can be started from a very small learning rate of 0.0002. Because the background of the pictogram in the natural scene can be traditional paper, stone wall, wood products, masonry products, or the leaves of plants, we have selected three representative backgrounds in natural scenes (rock, wood, green natural scenery). we use DCGAN to fuse the pictogram and the environment background (as shown in Figure 9), and finally generated the pictogram with a natural scene background. We divided the data set into a training set and a test set in a 5:1 ratio. For the three fused pictures, we used 7730 pictures for training and 1537 pictures for testing. Figure 10 shows the process of fusing the pictograms with rock, wood, and green natural scenery. The four images in each row represent the results of the 100th, 200th, 300th, and 400th iterations.

Figure 9. Original pictogram (left) and background (right)

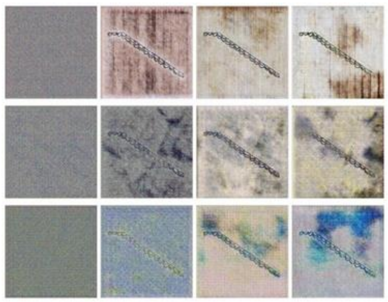

Figure 10. Image generated by three pictograms

As can be seen from Figure 10, the image at the 100th epoch has only color information and noise; the 200th epoch can extract the main features of the pictogram; at the 400th epoch, the two fused well and generated a pictogram with a background texture.

After the iteration is completed, the network is tested, and Figure 11 is the output of the test network. It can be seen that not only the pictogram and the background are well blended, but also the generated image has clear pictogram features, and some places even have some occlusion on the pictogram. This is enough to show that the method we use largely restores the pictogram state in natural scenes.

Figure 11. Image generated after 500 iterations

3.4 chapter summary

The three data enhancement methods we proposed have a good test effect on the design of the ancient pictogram recognition model (hereinafter detailed description of the model). The following is a comparison of the performance of these methods.

Table 1. Accuracy of different methods

methods	traditional	artificial	affine transform	natural scene
accuracy	16%	89%	99%	99%

In the enhanced scheme with limited random radiation transformation, when the iteration reaches 1300 times, the test accuracy tends to be stable, reaching 99%. The data enhancement scheme in the natural scene, after 35 iterations, the test accuracy rate has risen to over 90%. In 140 iterations, the accuracy rate tends to reach 99%. It is obvious that we have proposed a set of data enhancement methods. It is very feasible in practical engineering.

4. Experimental algorithm

4.1 Method principle

One important point to consider for the recognition of ancient pictograms is how to effectively use existing data for enhancement. We prepared the data according to the method described above, and then put the data into the network for training. In the network structure, we adopt the structure of the classic convolutional layer and pooling layer. After the data is input into the network in lmdb format, it will pass through the three convolutional layer and pooling layer, and then the feature dimension reduces to 100 through a fully connected layer. Because the difference between the pictograms is not large, we should increase the class spacing as much as possible, and use SoftMax as the loss function to output the classification results through the SoftMax classifier.

We use a 5*5 convolution kernel in all three convolutional layers, and set the pad to 2 to ensure that the feature size after calculation is the same as the original data. After the convolution operation of each layer, it will be normalized by the BN layer to improve the training speed of the network and prevent the gradient explosion and gradient disappearing.

For learning algorithms, its pros and cons are directly related to the performance of the network. After SGD, AdaGrad, and AdaDelta were respectively tested, it can be seen from Figure 12 that although the SGD algorithm has obvious fluctuations, the highest accuracy is obtained. The AdaGrad algorithm converges faster; the AdaDelta algorithm fluctuates greatly. Based on the comprehensive experimental results, our network uses the SGD algorithm and periodically attenuates the learning rate.

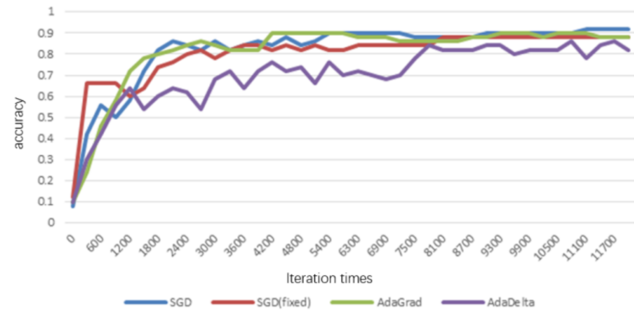

Figure 12. Comparison of gradient descent algorithms

The whole model was developed based on the deep learning framework caffe and GPU. The ancient pictogram recognition model used in the experiment is shown below.

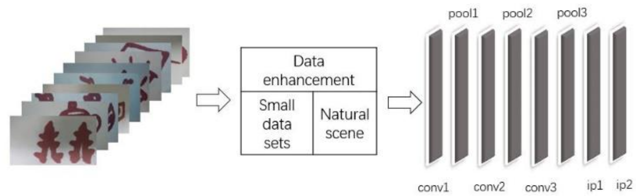

Figure 13. Schematic diagram of the ancient pictogram recognition model

As far as the current situation is concerned, the existing minority pictograms of ethnic minorities have a small amount of data, and the research on their pictogram recognition is still immature. The ancient pictogram recognition model proposed in this paper can achieve 99% accuracy based on a large amount of data expansion. The offline Yi pictogram recognition method proposed by Jiwa Aying of Sichuan University has an accuracy of 99% for offline pictogram recognition. After applying the data set of this paper, the performance is average, and the test accuracy is only 41%. It can be seen that the method of the ancient pictogram studied in this paper is worthy of recognition in the field of ancient pictogram recognition.

4.2 Result Analysis

Based on the above experimental results, the system is debugged and a network with high accuracy and good robustness is obtained. After the network training, we made a test picture simulating the natural scene to test the network performance and achieved good results. The test pictures and results are as follows.

Figure 14. Pictures of simulated natural scenes

The left and middle images are the pictograms on the simulated bronze background, and the right image is the pictogram on the simulated rock background. The difference between the foreground and the background on the left is more obvious. The middle and right images try to fuse the pictogram and the background. The test results corresponding to the above three pictures are as follows:

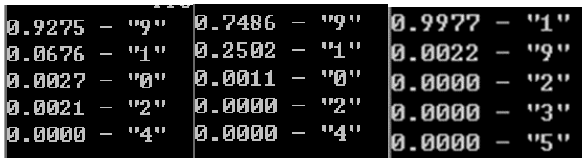

Figure 15. Classification results of simulated natural scenes

It can be seen that the network has strong robustness, which can not only recognize clear pictures, but also can recognize pictures with small differences in foreground and background colors and strong texture interference.

We tested the network with the enhanced data set, as shown in Figure 16.

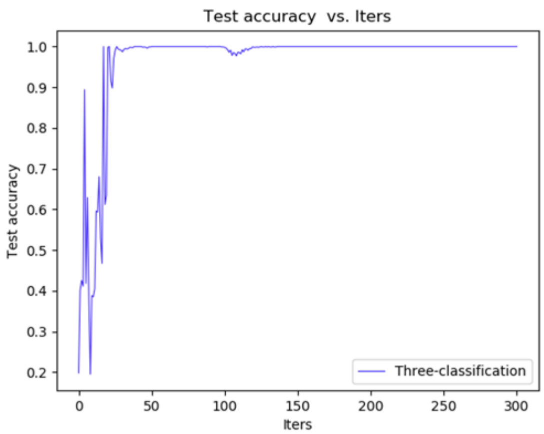

Figure 16. Graph of accuracy

When iterating to 140 times, the accuracy began to stabilize to 99%, which also proved the feasibility of our data enhancement scheme and the ancient pictogram recognition model.

5. Conclusion

`We proposed and summarized a set of data enhancement methods for ancient pictograms, solved the problem of small data amount of ancient pictogram recognition, and established an ancient pictogram recognition model and achieved good results. In the future, we plan to study the network multi-classification problem, continue to expand the types of pictograms, and finally realize the accurate identification of 81 types of pictograms in the Poya songs book.

6. REFERENCES

[1] Jinfang Li, Bingshan Liu, Binghui Huang, ect. Zhuang nationality's "Poya songs book" and its literal meaning [J]. *Journal of Minzu University of China: Philosophy and Social Sciences Edition*, 2010(1): 108-115.

[2] Smith R. An overview of the Tesseract OCR engine[C]//Document Analysis and Recognition, 2007. ICDAR 2007. *Ninth International Conference on.* IEEE, 2007, 2: 629-633.

[3] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. *Psychological review*, 1958, 65(6): 386.

[4] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.

[5] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural computation*, 2006, 18(7): 1527-1554.

[6] Sermanet P, Chintala S, LeCun Y. Convolutional neural networks applied to house numbers digit classification[C]//Pattern Recognition (ICPR), 2012 *21st International Conference on. IEEE*, 2012: 3288-3291.

[7] Cireşan D C, Meier U, Schmidhuber J. Transfer learning for Latin and Chinese characters with deep neural networks[C]//*Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE*, 2012: 1-6.

[8] Liu C L, Yin F, Wang D H, et al. Online and offline handwritten Chinese character recognition: benchmarking on new databases[J]. *Pattern Recognition*, 2013, 46(1): 155-162.

[9] Cireşan D, Meier U. Multi-column deep neural networks for offline handwritten Chinese character classification[C]//*Neural Networks (IJCNN), 2015 International Joint Conference on. IEEE*, 2015: 1-6.

[10] Shah M, Jethava G B. A literature review on hand written character recognition[J]. *Indian Streams Research Journal*, 2013, 3(2): 1-19.

[11] Oquab M, Bottou L, Laptev I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014: 1717-1724.

[12] Dai W, Yang Q, Xue G R, et al. Boosting for transfer learning[C]//*Proceedings of the 24th international conference on Machine learning*. ACM, 2007: 193-200.

[13] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//*Advances in neural information processing systems*. 2014: 2672-2680.