IOP Conf. Series: Materials Science and Engineering **261** (2017) 012021 doi:10.1088/1757-899X/261/1/012021

Exhibits Recognition System for Combining Online Services and Offline Services

He Ma[1, a], Jianbo Liu[1], Yuan Zhang[1] and Xiaoyu Wu[1]

[1]School of Information Engineering, Communication University of China, Beijing, China

E-mail: [a]mahe@cuc.edu.cn; [b]wuxiaoyu@cuc.edu.cn

Abstract. In order to achieve a more convenient and accurate digital museum navigation, we have developed a real–time and online-to-offline museum exhibits recognition system using image recognition method based on deep learning. In this paper, the client and server of the system are separated and connected through the HTTP. Firstly, by using the client app in the Android mobile phone, the user can take pictures and upload them to the server. Secondly, the features of the picture are extracted using the deep learning network in the server. With the help of the features, the pictures user uploaded are classified with a well-trained SVM. Finally, the classification results are sent to the client and the detailed exhibition's introduction corresponding to the classification results are shown in the client app. Experimental results demonstrate that the recognition accuracy is close to 100% and the computing time from the image uploading to the exhibit information show is less than 1S. By means of exhibition image recognition algorithm, our implemented exhibits recognition system can combine online detailed exhibition information to the user in the offline exhibition hall so as to achieve better digital navigation.

1. Introduction

With the awakening of people's national consciousness, they are increasingly aware of the importance and the charm of traditional culture.[1] More and more people have a strong interest in Chinese traditional culture.[2] Compared with other leisure venues, many people prefer to go to the Chinese Traditional Culture Museum. In the Chinese Traditional Culture Museum, the most attractive is the exhibits which is the representative of the Chinese Culture.

With the development of technology, people's needs seem to be unable to meet by manual interpretation. People want to understand the exhibits in a more convenient way. In order to meet the needs of people, a variety of digital navigation system is coming. Until now, there are two major kinds of digital navigation systems on the market:

Firstly, digital key. It is the most popular way used in the Chinese Traditional Museum. The client is a hand-held device, and the user can choose the goal explained by the speaker by pressing the key. Then the device will play the prepared description. This method does not require too much manual intervention, and it gives the users so complete freedom that the user can choose the exhibits he likes to be introduced. So it's very convenient.

Secondly, automatic induction. This method is based on the digital keys and add the auto play function. At present, the main way of induction is infrared automatic induction and wireless automatic induction. The kind of the induction requires the museum to equip the facilities. Such facilities not

only require a certain amount of engineering construction but also may not be used due to the impact of the storage conditions of the exhibits. Therefore, the promotion of this approach has been greatly restricted.

The existing digital navigation mode has many disadvantages, such as the help of other equipment, the information provided cannot be saved and so on. Based on the shortcomings of the existing digital navigation, in this paper, we propose an online and offline interconnection system based on the recognition of the exhibits. Visitors can use the mobile phone to identify the kind of exhibits and know about the detailed information of this exhibit at anytime. The system uses the way of separating the client and server, which reduces the requirement of the equipment, and can not only reduce the recognition time, but also improve the efficiency of the recognition.

2. Implementation of the Exhibits Recognition

In this system, the most important steps is recognizing image. We use the features from the deep learning network and send them to the SVM to classify the pictures. And before setting up the network, we should prepare lots of pictures which are used to train the network.

2.1. Data Preparation

The exhibits mainly present in the system through the form of image, so the core technology of the system is image recognition. This system combines SVM with depth hierarchy features to achieve the identification of exhibits. In this system, we have to complete the model training offline, in order to make the model more robust and acquire better generalization ability. The training data is very important, we need to collect the field data to finish the training model.

We go to Jiading Museum, Shanghai Museum of art, Shanghai Museum and other museums. Finally, we choose the porcelain exhibit in the Jiading Museum as resource of the image data. In the process of acquiring the experimental data in the museum, we take the diversity of user equipment into consideration, so we use SLR camera, mobile phone, micro single, pad and other devices for data collection. In the process of collecting, we consider the complexity of real shooting situation, and do not have a single shot of exhibits: considering the diversity of the angle for taking the exhibits picture, we take photos of exhibits from different angles, covering the shooting angle the users use as much as possible; Taking into account the proportion and position that exhibits in the final picture. We take photos of exhibits from different locations, different distances. Taking different size exhibits photos, and making the location of the exhibits is not the same. Unfortunately, due to the special provisions of the museum, we cannot use our own light source to take photographs under different light conditions. Therefore, we get the photos under different lighting conditions by processing the image after taking photographs. In fact, the museum's light is relatively constant, the impact of light on the picture is relatively small. Therefore, the influence on the results of the identification is very small. The final selection of the types of porcelain and some of the pictures is shown in figure 1.

In order to make the system has the ability to reject, we also need to collect negative samples. The negative samples is other data which are not belong to sample categories. We take samples is divided into two parts, one part is the samples collected in Shanghai museum. Some of them is shown in figure 2. The other part is the pictures we are randomly selected from the data in Imagenet.

Figure 1. Some of the positive sample set.

Figure 2. Some of the negative sample set.

Due to the limited manpower, the photos taken in the museum is difficult to meet the experimental requirements. In order to make the training model have greater generalization ability and avoid overfitting of the convolutional neural network, we need to expand the data. The extension methods used in the system are the actual extension and the virtual expansion. The actual expansion refers to the expansion that the data will be saved to the local, while the virtual expansion will made images only for the network and will not save to the local. Therefore, we will make each class which have the 200 original images rotated, as well as the light change to expand the data. So each class of sample is expanded to 8000. And on the base of the actual expansion, the virtual expansion is mainly used with extended data in Caffe mode. The angle and light have random changes to complete the expansion of the data. Finally the proportion that data from negative samples and positive samples is almost 1:1.

2.2. The exhibition images classification based on CNN

After the data preparation, we can process the image with our own method. We use the CNN to extract the features of the image. Then send the features to the SVM. This method take the advantage of the CNN and SVM. The specific processing is shown in Figure 3.

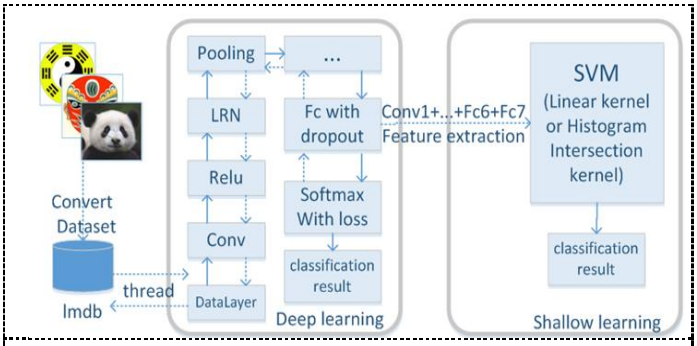

Figure 3. The process of the image processing.

2.2.1. Feature Extraction Using Deep Learning. With the introduction and development of the concept of deep learning, deep learning has played an important role in many areas. It is similar to the brain's deep cognitive process, which achieve multi-level data feature learning. Taking the image data used in the system as an example: Deep learning use the combination of the feature of the edge, the initial shape and other initial features to form a more abstract high-level representation. There are many models help us to apply the deep learning algorithms. Convolution Neural Network (CNN) is one of them, which is widely used in pattern recognition, image segmentation, and target detection.[3] And Alexnet is the classic model of image classification in CNN. Alexnet is the network made by Geoffrey and his student Alex in 2012. [4] Alexnet consists of five convolutions, two full layers and the last layer is the output layers. The server is mainly using the Caffe, to adjust the good mature model with a lot of data training, that is, fine-tuning.[5] Because the convolution network is to simulate the human brain's vision mechanism to identify different images. The more high-level features, the more abstract information we can get. The features are more able to represent the image. Thus, high-level features such as fc6, fc7 can better express the characteristics of the image. Therefore, we select fc6, fc7's output as the characteristics of the image.

The traditional network send feature information after fc7 directly into the softmax layer for classification.[6] But, whether the information has been abandoned is useless information or not? Considering the missing features, the server chooses to extract the features of the fc6 and fc7 layers. Each is characterized by a 4096 dimension, combining them into a large vector as a characteristic expression for various types of images. At the same time, we sent the features not into the softmax but the SVM classifier that we explains in the following chapters.

2.2.2. SVM. The nature of the identification problem is actually a classification problem, the last layer of the traditional neural network is connected to the softmax. The softmax regression model using the gradient descent to update the parameters. Because of the existence of this process, the speed of convergence will slow down. At the same time, because of the interference from the interference group data, the classification accuracy will decline. However, the traditional SVM classifier does not need to calculate the cost function. So the speed of convergence will be better than the softmax. SVM classifier has another great advantage that it has good performance on anti-noise.[7] Therefore, the server does not use the traditional network to classify the feature. But send the feature that the network extracts to the traditional SVM classifier for classification.

The main idea of SVM is increase dimension of input space to ensure it is Linearly Separable in high dimension space by using the method of nonlinear mapping. SVM uses the expansion theorem of the kernel function to avoid the curse of dimensionality. Different kernel functions are selected to generate the different SVM. The kernel functions that we use is the histogram intersection kernel function.

SVM classifier is mainly used to solve bipartition. In the real world most of the data is multi-class data, so the simple SVM need further expansion, so that it can solve the multi-class classification. There are many methods can realize the expansion of SVM. The system uses the one to one , that is, design a classifier between any two samples, so the K types need to design K (k-1) /2 classifier. When an unknown sample is classified, the class with the highest number of votes is the sample.

The use and construction of SVM classifier consists of two parts: training and recognition

Firstly, offline training process: the feature that extracted by the deep learning network will be sent to the SVM for training. Finally, we will get a trained SVM classifier.

Secondly, online identification process: similarly, after convolution neural network extract the feature of the test image's fc6 and fc7 layer, the features will be sent into the SVM classification has been trained. And SVM will give the final classification results.

3. Design and Implementation of the system

In order to meet the requirements that we can identify more kinds of exhibits with less time and less resource. In our system, we use the communication between the client and the server to realize the identification of the exhibits. The photo of the exhibit will be uploaded to the server through the Internet. The server return the identification result to the client after identification. The client show the result with the text.

3.1. The Client

The client is based on Android 6.0.1 or above systems. Like the ordinary app, you can install software after downloading the installer. The client's interface is easy to understand and the operation is easy. From the main interface of the client we can clearly see all the function of the software. The software can not only recognizes the captured picture, but also realize the recognition of two-dimensional code. And with the help of communication of the Baidu Server, the function of text recognition can be realized.

As the most important function of the software, the interactive logic of the real-time image capture and recognition is very simple. When users decide to take a photo, they just need to press the camera button in the bottom left corner. The moment they press the camera button, the client will set up connections with the server and send the picture data to the server. After the server recognizes the picture, the number of the class that the picture belongs to will be returned to the client. After the client accepts the number, the name of the porcelain corresponding to the number and its related description are displayed on the interface.

At the same time, we add some auxiliary functionality to help the users use the app more convenient. Clicking the library button in the bottom left corner, the user can select photos from the gallery. And clicking the record button in the bottom right corner, it will show the picture has been

identified. The users can delete records and view the recognition results of the picture has been identified.

3.2. Communication between Client and Server

After the completion of the exhibits image's collection in the client, the image information must be uploaded to the server. And the result can be displayed after processing by the server. Therefore, the communication between the client and the server is an important step in the implementation of the system. There are many ways that can realize the communication between the client and the server.

Finally we choose the Http protocol for transmission. The Http protocol is the object oriented application layer's protocol. Because only the request method and path transmission are needed to, the Http server program is very small. Therefore, the communication speed is very fast. Http protocol is flexible enough to allow the transmission of any type of data object. The process using Http protocol to achieve the connection between the client and the server mainly includes the following steps:

Firstly, the client sends the request to the server. Because in our system, the client needs to upload the image data, and the server needs to return the information, we choose post (Post ask the serve receive the data attached to the request). We attach the data in the form to the request.

Secondly, server response the request and send the result to client. The server will process the image data that sent by the client. The server will get the number of the category in the image after recognition. The server will attach the results of recognition to the response to the request of the client. When the client receives the message sent by the server, the connection ends. The Http transport process is complete.

4. Experimental Results

In this experiment, we use a PC as a server, and huawei honor 7i smart-phone as the client, whose configuration is 8 core 1.5 GHz processor, memory is 3.0 GB and mobile phone system is android 6.0.1 system. To achieve the communication between server and client, we use the campus network. In a relatively stable environment, we do the test, using the phone's inbuilt speed test tool to test the speed of the network. Of course, the background software was clean. The average upload speed is about 63.7 KB/S, and the average download speed is about 371 KB/S.

Due to the limited conditions, we choose to take photos of the picture taken from a museum from different perspectives. We randomly selected five photos of each porcelain to do this experiment. And the number of taken photos was 100. Part of the picture are shown in figure 4.

In the end, the recognition accuracy is 100%.Because our system is real-time, recognition time is what we concern most. In the experiment, the average recognition time for 599 ms. Among them, the shortest recognition time is 329 ms. All recognition time within 1000 ms, the longest recognition time is 975 ms. The distribution of recognition time is shown in figure 5.

Figure 4. Some of the pictures which have been used in the experiments.

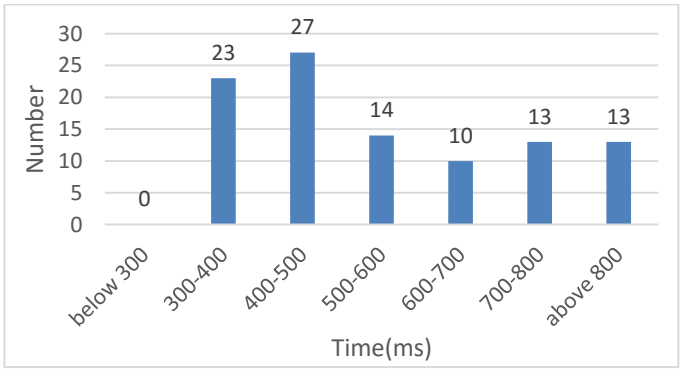

Figure 5. The proportion of the recognition time distribution

5. Summary

In this paper, we introduce the exhibits recognition system for combining online services and offline services. The client of this system is the phone. The visitors can use the phone to take the pictures of the exhibits. Then the data will be sent to the server. The server is used to recognize the image. We use the deep learning network to extract the features of the image and send features to the SVM to classify. What is worth mentioning is, in this system, in order to make the model more robust and acquire better generalization ability, we have to complete the model training offline. After recognition, the result will be returned to the client and will be shown in the phone.

To reach a higher speed and use less time to complete the communication between the client and the server, we use the HTTP transport. Finally, in our experiment, we find the recognition accuracy is 100% and the time of one process is very short.

The online and offline interconnection system based on exhibits identification proposed in this paper has the following advantages compared with the existing navigation system:

Firstly, the server uses a combination of deep learning and SVM classifier, with a higher recognition rate.

Secondly, the use of the way of separating the client and server makes the requirements of the client greatly reduced.

Finally, using a smart phone as a client, more simple and convenient.

Experimental results show that the system is feasible and practical, in the era of smart phones, the navigation based on smart phone will bring the difference to the museum's navigation system.

Acknowledgments

The work on this paper was supported by the National Key Technology R&D Program (2015BAK22B00) and the Program (JXJYG1606) of Communication University of China.

References

[1] Tan X, Wu X, Yang C, Shen Y 2016 Proceedings of the 20165th IEEE International Conference on Image, Vision and Computing Chinese Traditional Visual Cultural Symbols Recognition based on SPM Muti-feature Extraction C

[2] Tan X, Wu X, Yang C 2015 Proceedings of the 2015.8th International Symposium on Computational Intelligence and Design -Volume 02. IEEE Computer Society 2015.304 Visual Cultural Symbol Recognition Based on Muti-feature Extracting C

[3] Dahl J V, Koch K C, Kleinhans E, et al 2010 Convolutional networks and applications in vision J 14(5):253-256 345

[4] Krizhevsky A, Sutskever I, Hinton G E 2012 Advances in Neural Information Processing Systems ImageNet Classification with Deep Convolutional Neural Networks J 25(2) 7442

[5] Jia Y 2013 http://caffe.berkeleyvision.org Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding 2435

[6] Deng J, Dong W, Socher R, et al. 2009 ImageNet: A large-scale hierarchical image database C 248-255 3212

[7] Zhang X 2000 Acta Automatica Sinical, Vol.26, No.1, pp 32-42 1481