IEEE ITNEC(ISSN:2693-3128)

Classification of Dunhuang Mural Image Based on Small-sample and Semi-supervised Learning

Hui Cui[1,2] ,Zhibin Su[1,2] ,Luyue Zhang[1,2] ,Liyang Bai[3]
1. School of Information and Engineering, Communication University of China
2. Key Laboratory of Acoustic Visual Technology and Intelligent Control System
3. Center for Ethnic and Folk Literature and Art Development, Ministry of Culture and Tourism, P.R.C.
860182065@qq.com, suben@cuc.edu.cn, 3094706375@qq.com
Corresponding Author: Zhibin Su Email:suben@cuc.edu.cn

2023 IEEE 6th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC) | 978-1-6654-6004-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITNEC56291.2023.10082614

Abstract—**Dunhuang mural images classification belongs to the research task in the field of image recognition. In this paper, the semi-supervised model is established with multidimensional features extracted by transfer learning. A small number of labeled samples were used to obtain a large number of unlabeled data, combined with Active Learning and iterative strategy for multiple rounds of label transfer of selected samples. After several rounds of iterations, we can get a more powerful classification learner. Experiments on the self-built Dunhuang mural dataset have shown that the results can approach or even exceed some supervised learning methods when the number of known label samples is less than 4%. Our research implements the cultural resource image classification algorithm based on small samples, which is conductive to improving the accuracy when label samples are scarce.**

Keywords—Dunhuang Mural images; Label Propagation; Active Learning Strategy; Semi-supervised Model

I. INTRODUCTION

In recent years, with the rapid development of digital protection technology of cultural resources, the need for the establishment and management of digital museums or digital libraries of Chinese painting images has increased. Especially the image processing technology of Chinese painting has become the key problem to be solved. There are more and more researches on automatic semantic annotation, retrieval and automatic classification of Chinese painting images. Traditional cultural relic image classification is mainly through heavy manual classification work, but artificial intelligence means it can easily extract the features of the image and classify, to help the cultural relic protection workers reduce the classification burden.

The automatic classification of mural painting belongs to the research content of image recognition. There are usually two ways to identify images: one is based on traditional image processing. For example, when an image is given, the system retrieves it through the feature description vector of the image. If the feature of the image to be retrieved has a high similarity with the feature of the known category, the image belongs to this category. The advantage of traditional image processing method is convenient and rapid, but the disadvantage is that the accuracy of recognition is not high and it is easy to be disturbed by the outside world[1]. Another approach is based on machine learning[2]. Using the most advanced machine learning technologies such as regression analysis, Apriori algorithm, K-means algorithm and convolutional neural network, the accuracy of image processing can be greatly improved on the premise of ensuring the processing speed[3]. At present, with the rise of deep learning, under the premise of big data, image recognition methods based on deep learning have also been born[4]. However, the deep learning technology relies on abundant sample resources, and the digital mural image, as the object of cultural resources protection, has the problems of missing image types and serious shortage of sample resources. Therefore, in view of the current situation of insufficient sample resources, it is necessary to study the machine learning method based on small sample data[5].

Semi-supervised learning is a relatively emerging method in the field of machine learning. Its combining supervised learning and unsupervised learning, that is, a machine learning method using labeled sample data and unlabeled sample data at the same time.

Today, the research on semi-supervised learning covers a wide range. On the one hand, various traditional supervised and unsupervised learning algorithms are constantly modified or extended to improve their learning ability under semi-supervised conditions. On the other hand, new mathematical methods are constantly introduced into semi-supervised learning. At the same time, the research object of semi-supervised learning has expanded from the simple training of semi-supervised data to the prevalence analysis of semi-supervised data, the relationship between semi-supervised data and graph model, etc. In the case of the scarcity of marked samples, semi-supervised learning plays the role of unmarked samples to jointly optimize the model, and solves the problems such as the difficulty in acquiring marked samples and the consumption of large amount of resources in real life, which has high application value[6]. But the deficiency is easy to appear overfitting problem[7]. At the same time, the accuracy of image classification is difficult to reach the expected target due to the inadequacy of feature extraction and training methods[8]. There are many ways of semi-supervised learning. Inspired by literature [9], we adopt a semi-supervised learning algorithm based on the combination of Active Learning and Label Propagation. A small number of typical samples selected by Active Learning are used as known labels for Label Propagation of the remaining large number of blank samples, and the combination and optimization methods in its structure are rearranged and designed according to task requirements. The difference from the literature is, our method focuses on the iterative process of label propagation with selected samples rather than the iteration of Active Learning itself.

II. Method

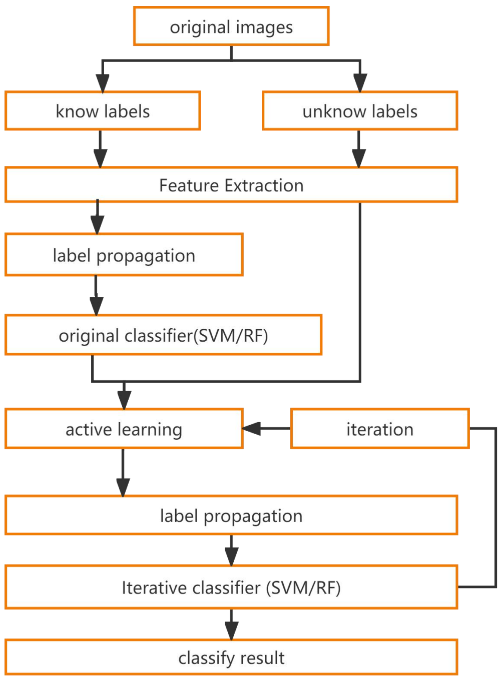

Fig. 1 The framework of our proposed approach

The classification method combining Active Learning and Label Propagation mainly includes three steps: the first is feature extraction, then the classification algorithm combining Label Propagation and SVM classifier, and the third is the iterative algorithm combining Active Learning and Label Propagation. For feature extraction, we extracted the CLIP features from the surface layer. For the classification algorithm combining Label Propagation and SVM classifier, we select the known label samples of the label transfer algorithm by sampling randomly, and bring the results after Label Propagation into the SVM classifier as a training set. For the iterative Active Learning and Label Propagation strategy, we adopted the RS algorithm to filter out several typical data into the strong classifiers of the previous round of classification, and used the prediction results of the strong classifier as new labels for the typical data. The typical data is passed as the known labels for the remaining data. After the end of Label Propagation, we took the whole samples of Label Propagation as the training set to bring the results into the classifier. And repeat the iterative process as described above. The framework of figure 1 represents our massive structure.

A. Feature Extraction

The CLIP pre-training model is the latest model developed by OpenAI company[5], which is developed across text and image dimensions and can map the images to the category of text description. After the comparison of the traditional features such as the colors, textures, SIFT, DSIFT and BOW, CLIP has shown its excellent image content recognition capabilities. Therefore, we took advantage of the CLIP model to perform CLIP feature extraction on the dataset.

B. Label Propagation

In supervised learning, when the numbers of known labels are relatively low, and the vast majority of the data is not labeled, relying solely on the acquired feature information cannot classify the image sets very well. At this time, we need to use some special spatial relationship between the marked data and the unmarked data to divide these data, so as to improve the classification accuracy.

Label Propagation is a semi-supervised learning algorithm, with two assumptions of the semi-supervised learning algorithm: 1. Nearby sample points have the same label. 2. Points on the same stream structure can have the same label. The basic principle of LP algorithm is from the marked node label information to predict unlabeled node label information, using the relationship between samples, nodes including marked and unlabeled data, the edge represents the similarity of the two nodes, through the edge between the nodes, the weight of the edge, said the more similar the two nodes, then the label is easier to spread in the past.

We define a probabilistic transition matrix of the N N, the matrix P:

$$P_{ij} = P(i \to j) = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}} \qquad (1)$$

$$w_{ij} = \exp\left(-\frac{\|x_i - y_i\|^2}{\alpha^2}\right) P_{ij} \qquad (2)$$

In the above formula, P_{ij} represents ote the probability of transfer from node i to node j.

w_{ij} represents ote the weight of edges composed of nodes i and j, α is a super parameter.

$\sum_{k=1}^{n} w_{ik}$ represents ote the weight sum of edges of node i and all nodes.

Community refers to some dense groups in the network.

The advantages of LP algorithm are: 1. The algorithm has simple logic and is close to the linear complexity; 2. There is no need to define the optimization function or specify the number of communities in advance. The algorithm will use its own network structure to guide the label propagation. The disadvantages are: 1. Community results are unstable and have strong randomness. Since when multiple adjacent nodes have the same label proportion, one is randomly taken as the final result. This leads to a small error in the early stages of the transmission, with no appropriate results. Especially when the asynchronous update, the different update order will also lead to the different final community division results; 2. Community results will fluctuate back and forth, but do not converge. When the transmission mode is in the synchronous update, it is easy to appear the result of the shock phenomenon.

In our experiment, the first round of Label Propagation is to randomly select a small amount of data as the known labels on the remaining more than 1,000 images. Label Propagation in the iterative process is the whole composition of the typical data selected by Active Learning and the known label data from the first round of Label Propagation as the known labels of the new round of Label Propagation, and the label spread to the remaining data. Through experiments, we found that the more balanced the number of samples in each category before Label Propagation, the better the effect of subsequent prediction was.

C. Active Learning Based on The Random Sampling

Active Learning refers to the method of machine learning method modeling with the participation of human beings, to select some suitable candidate sets for the next step of artificial annotation. Its general idea is through the machine learning method from the data set actively get some sample data without labels, let manual annotation and audit, after each annotation model again on the labeled data training, and then actively choose unlabeled data annotation, and then keep repeating the process. The main study of the active learning strategy in our experiment is the RS strategy. The RS strategy is a random selection strategy, which is implemented as follows:

$$X_{RS} = random_{x_i \in D_U}\{P(y_i = m|x_i), m \in L\} \quad (3)$$

$$F_{RS}(n) = F_{RS}(n-1) + X_{RS}(n), \quad n \geq 1 \quad (4)$$

$$F_{RS}(0) = X_{original} \quad (5)$$

In the above formula, D_U is a candidate sample set that is not tagged, y_i is the label, m is the type of label, n is the number of iterations, L is the total category of the image set (0,1,2), $X_{original}$ are the total samples of known labels prior to the iteration, X_{RS} are the typical samples screened out by Active Learning, $F_{RS}(n)$ are the total samples of known labels in the of the round n iteration.

D. Iteration

Iteration is the activity of the repeated feedback process. The purpose is usually to make the final result closer to the goal. Each repetition of the process is called one iteration. The result of each iteration serves as the initial value of the next iteration.

The specific implementation process of the iteration is as follows:

1） The CLIP features were extracted from the dataset;

2） One hundred samples were randomly selected as the test set for each round of the classification;

3） Several data were randomly selected as known label samples for Label Propagation to the remaining samples;

4） Classify the data as the training set and predict the test set accuracy;

5） Active Learning screens out several typical data;

6） The labels of the typical data were predicted by the classifier of the previous round of classification;

7） The predicted results in the classifier and the initially selected samples of known labels were collectively transferred as known label samples to the remaining samples;

8） Classify the data as the training set and predict the test set accuracy;

9） Repeat Step 5) 6) 7) 8), and compare the accuracy predicted by each round of the classification until the accuracy is unchanged or all of the unknown samples are labeled.

III. Experimental Results

A. Experimental Dataset

In this section, we will illustrate the details of the classification algorithms that combine Active Learning and Label Propagation. All experiments were performed on our dataset of 1813 Dunhuang mural images. Our dataset consists of three categories: Buddha painting, story painting and decorative painting. Since the initial decorative painting categories were few and some image categories were not clear, we had used image cutting and image rotation to enhance the image set. The number of the Buddha statue painting, story painting and decorative painting is 578,754 and 481, respectively. The figure 2 below shows some samples from the image dataset.

Fig. 2 some experimental pictures from the our database.

B. Comparison of The Supervised Classifiers

We experimented with comparing the classification effects of different classifiers without adding Label

Propagation and Active Learning algorithms when the number of known samples was 100,500,100, respectively. four common classifiers of SVM, Random Forest, Rusboost, and Bayes were used for comparison. Due to the small amount of data in machine learning, the results of 100-labeled samples are only used as a reference and are not necessarily universal. The following table Ⅰ shows our experimental results.

TABLE Ⅰ. RESULTS FOR DIFFERENT CLASSIFIERS WITH 100,500,1000 KNOWN SAMPLES.

Classifiers	100	500	1000
SVM	79.1%	86.8%	87.9%
Bayes	75.8%	78.6%	79.1%
Rusboost	79.7%	82.4%	84.1%
Random Forest	68.7%	77.5%	79.1%

C. *Comparison of The Number of Iterations*

First, we compared the classification effect of the classifier by changing the number of Active Learning samples at 20 known samples, with no iteration, 1 iteration, and multiple iteration (within 10 times). At the beginning of the iteration, the model did not perform very well. However, as the number of iterations increased, the accuracy also increased fast. The more iterations, the better the effect when fewer labeled samples were known. The figure 3 and figure 4 below show our experimental results.

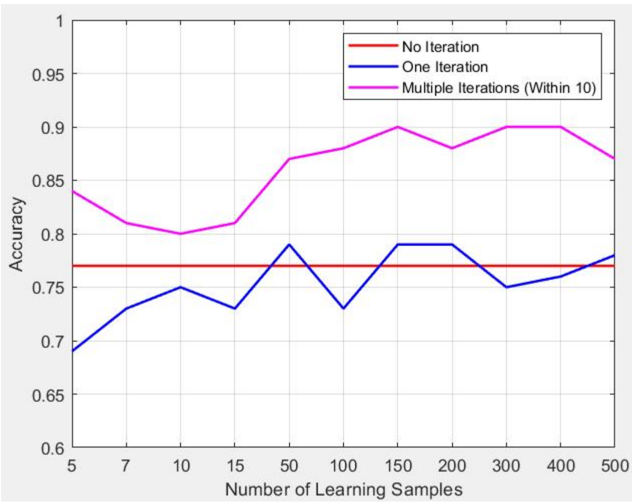

Fig. 3 The Number of Known Samples is 20 SVM Classifier

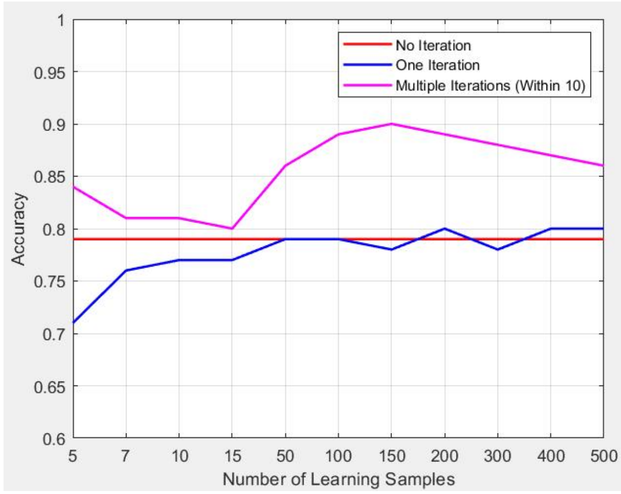

Fig. 4 The Number of Known Samples is 20 RF Classifier

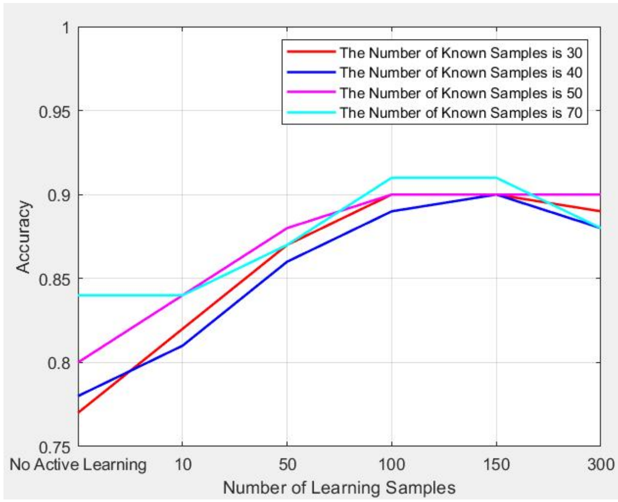

Fig. 5 The Number of Known Samples is 30-70 SVM Classifier

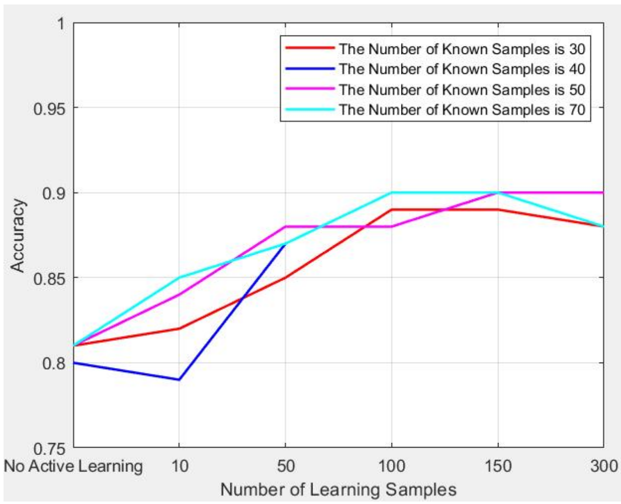

Fig. 6 The Number of Known Samples is 30-70 RF Classifier

In addition, we increased the number of original labeled samples to 30,40,50 and 70 to compare the effect on the accuracy of classification. Similar to the expected results, the accuracy would get better as the number of labeled samples increases. Another finding was that, the number of samples chosen by active learning would also affect the final result. In most cases, 100 or 150 samples chosen for reassignment with the iteration classifiers would get the highest accuracy. The figure 5 and figure 6 show our experimental results.

D. Comparison of Whether There is Label Propagation

At the end of the experiment, we had changed the number of known samples and compare the classification effect with and without Label Propagation, thus verifying the importance of Label Propagation. As for the classifiers we used still SVM and RF as a comparison. It could be seen from the experimental results that the Label Propagation significantly improved the accuracy of the classification. The figure 7 below shows our experimental results.

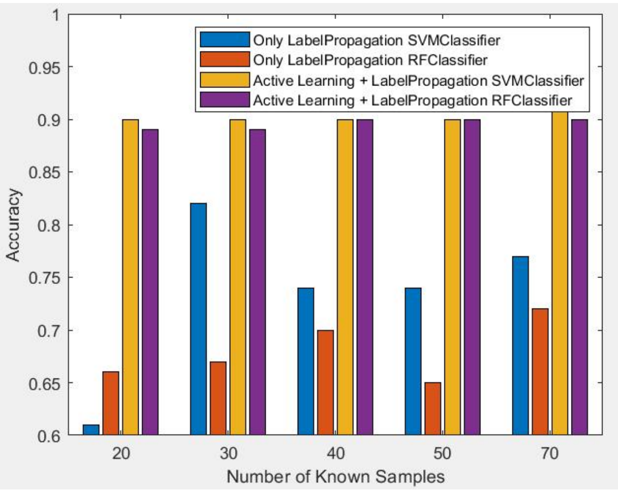

Fig. 7 Comparison Between Active learning and No Active Learning

IV. Conclusion

This paper studies an iterative learning model combining Active Learning and Label Propagation algorithm for classification task of Dunhuang mural images. The innovations in our work include: 1) the semi-supervised learning method is selected into the classification task of mural images, provided a stronger learner with the ability to predict on data sets of few samples; 2) the effective Active Learning strategy and iterative method are used to pass known labels to more reliable samples through multiple iterations, which can gradually enhance the prediction accuracy of the model and improve the performance of the classifier. Experimental results show that our model can achieve a high classification accuracy of Dunhuang murals with only dozens of known label samples. However, influenced by the random extraction strategy, when the quality of the initial sample is different, the accuracy of the subsequent algorithm will be affected. In the future work, we will extract more features related to the artistic characteristics of the mural itself, and optimize the Parallel label-delivery with iterative strategy for better generalization performance.

Acknowledgment

This work was sponsored by Key Laboratory of Audio and Video restoration and evaluation, Ministry of Culture and Tourism(2021KFKT005).

References

[1] Yong Zhang, Xuefeng Du, Yue Gao, Wei Yang. Traditional algorithms of image processing based on the lane line detection [J]. Journal of practical technology, 2022,47 (02) : 20-23, DOI: 10.16638 / j.carol carroll nki. 1671-7988.2022.002.005.

[2] DaFei Yu. Machine learning methods in the application of image classification [J]. Journal of electronic world, 2017 (23) : 57-58, DOI: 10.19353 / j.carol carroll nki DZSJ. 2017.23.023.

[3] Kai Wu. Application and Design of Machine Learning in image Processing [J]. Digital World,2019(03):146.

[4] Wei Jin, Xiaoman Meng, Yichao Wu. Deep learning in the application of image classification review [J]. Journal of modern information technology, 2022, 6 (16) : 29-31 + 35 DOI: 10.19850 / j.carol carroll nki. 2096-4706.2022.16.008.

[5] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In ICML, 2021

[6] Caihong Mu, Hyperspectral image classification based on Active learning and semi-supervised learning. Xidian University, Shaanxi Province,2016-02-01.

[7] Yuwan Gan,The image classification algorithm based on a semi-supervised learning research [D]. Guilin university of electronic science and technology, 2021. The DOI: 10.27049 /, dc nki. GGLDC. 2021.000060.

[8] Tao Wu, Image Classification Algorithm based on Feature Extraction and semi-supervised Learning [J]. Bonding, 201,48(11):92-97.

[9] Li-Guo Wang, Hui Shang, Yao Shi. Hyperspectral image classification based on Active Learning and label transfer algorithm [J]. Journal of Harbin Engineering University,20,41(05):731-737.