2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Multi-label Garbage Image Classification Based on Deep Learning

Kang Yan, Wenyu Si, Jin Hang, Hong Zhou, Quanyin Zhu*

Faculty of Computer & Software Engineering, Huaiyin Institute of Technology, Huaian, China
*Corresponding author's email: hyitzqy@126.com

Abstract—In recent years, with the development of deep learning technology, the accuracy of image recognition has been significantly improved. Deep learning has been widely used in the recognition of single-label images. This project aims to intelligently classify domestic garbage images as application scenarios based on depth. Learn to carry out multi-label classification research on images containing multiple visual objects, and design and build a multi-label garbage image classification model to improve recognition accuracy and speed as the main research goal to conduct classification research on multi-label garbage images.

Keywords−−Convolutional Neural Network; Garbage classification

I. INTRODUCTION

With the rapid economic development in recent years, and more and more garbage is generated in people's daily lives. The idea of secondary recycling of garbage[1] has become more and more popular. However, waste recycling is also facing more manpower requirements, and the efficiency of manual waste classification is low, and the cost of workers is also increasing. Relying on traditional garbage collection methods can not solve the current problems. Garbage identification technology has a positive effect on garbage classification.

The main technology used in garbage classification is a convolutional neural network based on deep learning[2] Therefore, how to efficiently and automatically identify garbage by using advanced computer vision technology has become the current focus. Traditional garbage classification requires people to classify themselves or manually. The former cannot ensure the accuracy of garbage classification[3], and the latter requires a lot of labor costs. It is more difficult to implement when the amount of garbage is large. This will waste a lot of manpower and material costs

In this case, the feature can be extracted by convolutional neural network[4] Use different feature scales to predict different scale targets to achieve accurate garbage identification.

II. IMAGE FEATURE EXTRACTIO

Computer vision technology is to extract features from images to identify images. Therefore, the choice of features is particularly important. Here, three characteristics are mainly explained.

Color characteristics: Color characteristics include color histogram, color moment, color set, color aggregation vector, etc. The color histogram can simply describe the global distribution of colors in an image, and it is also difficult to describe automatically divided images and images that do not need to be used to evaluate the local state of objects. The most commonly used color space is RGB color space and HSV color space.

In all color space models, the HSV model[5] can better reflect the human eye's ability to perceive and discern colors. It directly corresponds to the three elements of human eye color visual characteristics, namely hue H, saturation S and brightness, and the channels are independent of each other. Hue H represents the wavelength of light reflected from an object or transmitted through the object, that is, the color of the light. Light of different wavelengths exhibit different colors and have different hues. Saturation s represents the shade or shade of the color. The saturation is related to the proportion of white added to the color. It reflects the degree to which a certain color is washed out by white. When the white component is 0, the saturation is 100%, only white , The saturation is 0.

Brightness V represents the degree of darkness of light perceived by human eyes. The greater the energy of the light wave, the greater the brightness. The hue and saturation of the color indicate the depth of the color, collectively called chroma. The HSV color model can be shown in the figure1.

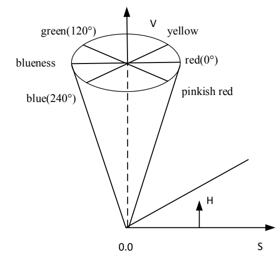

Figure l. HSV color model

HSV is a unified color space. His transformation is non-linear transformation. RGB values (R, G, b) in an image can be converted to HSV spatial values (H, S, V) by nonlinear transformation

The color moment is achieved through mathematical methods. Any color distribution in an image can be represented by color moments, or it can be used to describe color distribution. The color difference can be calculated in three channels. The color histogram and color moment only consider the overall distribution of image colors, and do not involve location information.

The color histogram can be represented by a color set. First, convert the image from three-channel RGB color to a visually balanced color space, and then quantize the color space into multiple handles. Then the image is divided into several regions, and each region can be indexed by a certain color component, thereby representing the image as a binary color index set. Generally, there are two types of representation methods for shape features, one is contour

DOI 10.1109/DCABES50732.2020.00047

features and the other is regional features. The contour feature of the image is mainly for the outer boundary of the object, while the regional feature of the image is related to the entire shape area.

Feature Border Method This method captures image shape parameters by describing feature borders. The basic idea of the Fourier Form Descriptor is to use Fourier to modify an object boundary as a shape description and to use the proximity and regularity of the boundary domain to convert a two-dimensional problem into a one-dimensional problem.

The expression and matching of shapes adopt a simpler method of describing regional features, such as the shape parameter method related to quantitative measurement of

shapes (such as moment, area, perimeter, etc.). Texture features: texture features include statistical methods, geometric methods, model methods, signal processing methods.

The geometric method is a texture feature analysis method based on texture primitive theory. The theory of texture primitives believes that complex textures can be composed of simple texture primitives. At present, the development of geometric methods is limited, and there are not many follow-up studies.

This model method is based on the image structure model. At the same time, the parameters of the model are used as texture features.

The extraction and matching of texture features mainly include: gray level co-occurrence matrix[6], Tamura texture features, auto-regressive texture model, wavelet transform, etc.

III. MULTI-LABLE IMAGE RECOGNITION

Multi-label image recognition[7] is a very important and difficult task in the field of computer vision. It is generally used in recommendation systems and search engines. Multi-label image recognition solution: In the early days, the multi-label problem was decomposed into multiple single-label problems to solve. This method of treating each label independently ignored the correlation between the labels, which tended to lead to poor training results; then A method based on a classifier chain is proposed, which captures the dependence of tags through the conditional product of probabilities, but this method will generate a high calculation cost when processing a large number of tags, and its ability to capture the correlation between tags is still lacking; in recent years With the development of neural networks, more complex neural networks such as Deep Neural Network (DNN) and Recurrent Neural Network(RNN) began to be applied to the field of multi-label image recognition[8]. Neural structure, weight loss function to optimize training.

This study intends to design and implement a multi-label classification[9] model based on Graph Convolutional Network (Graph Convolutional Network GCN). The model can build a directed graph and map different categories through GCN, so as to make better use of local dependencies and highlight the correlation between models to effectively solve the complex layout and partial occlusion in multi-label image recognition and other issues.

Graph convolutional networks[10] can be used for semi-supervised classification tasks, and the core idea is to update the representation of nodes through the propagation of information between nodes.

Unlike the standard convolution method that operates on the local Euclidean structure of an image, the goal of GCN is to learn the function f(.,.) of a graph G. The input of this function is the feature description $H^l \in \mathbb{R}^{n\times d}$ and the correlation coefficient matrix$A \in \mathbb{R}^{n\times n}$, there by updating the node feature as to a Equation 1

$$H^{l+1} \in \mathbb{R}^{n\times d'} \tag{1}$$

Each GCN layer can be written as a nonlinear Equation 2:

$$H^{l+1} = f(H^l, A). \tag{2}$$

f(.,.) can be expressed as Equation 3:

$$H^{l+1} = h(H^l \hat{A} w^l), \tag{3}$$

In this way, it is possible to model the complex interwoven relationship between nodes by stacking multiple GCN layers.

The matrix form of the Fourier transform of the graph is shown in Equation 4.

$$gF\{x\} = U^T x \tag{4}$$

For the Fourier transform of graph x, write the Fourier transform base multiplied by Equation 5

$$x = \big(f(1) \dots f(n)\big) \in \mathbb{R}^n \tag{5}$$

(Fourier transform bases are also the eigenvectors of Laplace matrix) use the variety of complex Fourier transform and Laplace transform quality, we have reached a concise and very important conclusion. Just

The above formula is the Fourier transform of the graph.Similarly, the inverse Fourier transform is as shown in Equation 6

$$IgF\{\hat{f}\}(\mathrm{i}) = \sum_{l=0}^{n-1} \hat{f}(\lambda_i) U_l(\mathrm{i}) \tag{6}$$

Image feature learning: In principle, any CNN-based model can be used to learn image features. This paper expects to use ResNet-101 as the experimental basic model in the experiment; then apply the global maximum pooling to obtain the image-level features x: x is as shown in Equation 7

$$\chi = f_{GMP}\big(f_{Cnn}(I; \theta_{cnn})\big) \in \mathbb{R}^D \tag{7}$$

GCN classifier learning: learning an interdependent target classifier $w = \{\omega_i\}_{\mathrm{i}=1}^{C}$ from a label feature through a GCN-based mapping function. The disregarded researcher uses stacked GCNs, where the input of each GCN layer I takes the node feature of the previous layer (H^l) as input, and then outputs the new node feature H^{l+1}. The input of the first layer is the word embedding vector $Z \in \mathbb{R}^{C\times d}$, and the output of the last layer of the matrix is the classifier $W \in \mathbb{R}^{C\times D}$.By applying the learned classifier to the image features, the prediction score is obtained: $\hat{y} = wx$. Assuming that the true label of an image is $y \in \mathbb{R}^C$, then the entire network can be trained using the loss function of traditional multi-label classification, as follows Equation 8:

$$\mathcal{L} = \sum_{c=1}^{C} y^c \, log\big(\sigma(\hat{y}^c)\big) + (1 - y^c) \, log\big(1 - \sigma(\hat{y}^c)\big) \tag{8}$$

The matrix form of the inverse Fourier transform is shown in Equation 9

$$IgF\{x\} = Ux \tag{9}$$

Graph convolution network Equation [11]10

$$Z = f(X, A) = softmax(\hat{A}\ ReLU(\hat{A}\ XW^{(0)})W^{(1)}) \quad (10)$$

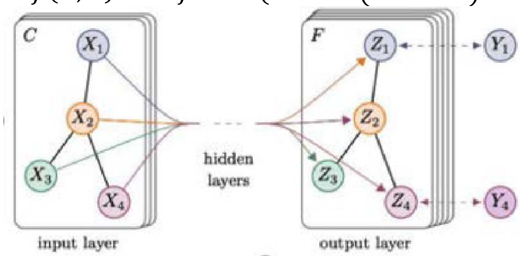

Figure 2.Graph Convolutional Network

The number of hidden layer of feature maps for H, $W^{(1)} \epsilon \mathbb{R}^{H \times F}$ is the weight matrix.

Use Softmax to normalize it to, As shown in Formula 11:

$$p(y|X) = \frac{e^{S(X,y)}}{\sum \tilde{y} \in Y X\, e^{S(X,\tilde{y})}} \quad (11)$$

Number of input layer to C, output layer is one of the symmetric adjacency matrix.The formula is shown in formula 12

$$\hat{A} = \tilde{D}^{-\frac{1}{2}}\ \hat{A}\tilde{D}^{\frac{1}{2}-} \quad (12)$$

Similarly, the weight $W^{(0)} \epsilon \mathbb{R}^{C \times H}$ is t he weig ht matr ix from the input layer to the hidden layer.

IV. Web Management Module

Mainly use HTML5, CSS3 and JavaScript as the language for building web pages. At the same time, link the information in the database with the information on the web page. Users can log in to the garbage classification system through the web platform. By logging in to the system, personal information can be obtained from the database and displayed on the interface. The system will provide garbage classification[12]. At the same time, the results can be returned to the user. The realization of the system is shown in Figure 3

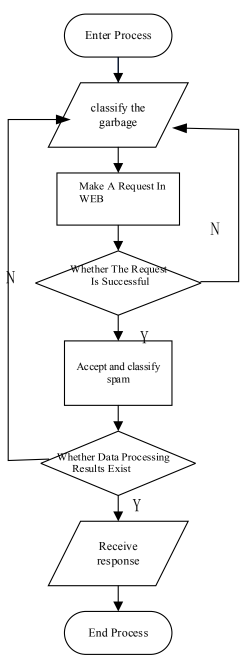

Figure 3. Web Management

V. System Operation Flowchart

Constructing a garbage image information extraction system using the Web to make the garbage classification system easier for users to use. After the user successfully logs into the garbage image information extraction system, the information can be dumped in the trash and visualized. Figure 4 below shows the operation flow chart of this system.You can split the whole process as shown in Figure 3 above. To the next step:

1) Obtain junk training images. If the garbage image is successfully obtained, go to step 2.

2) Tag the junk images. If the image label is finished, go to step 3.

3) Build a neural network and train a garbage classification model. If building the neural network is successful, proceed to step 4.

4) Determine whether the model can correctly distinguish the types of garbage. If you can successfully distinguish the type of garbage, go to step 5, otherwise go to step 2.

5) After the model training is completed, save the model. If the save is successful, go to step 6.

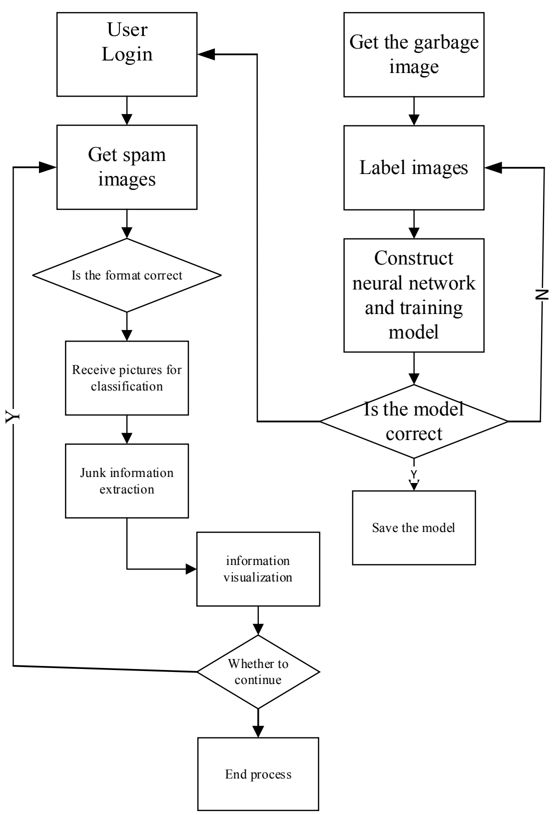

Figure 4. operation procedure

6) The user can log into the garbage classification system. If the system shows that the login is successful, please skip to step 7

7) Obtain the garbage images invested by users. If the acquisition is successful, g o to step 8

8) Next, it will be judged whether the format is correct. If the format is correct, go to step 9

9) Receive junk images, classify junk images, and extract junk image information. If the information is successfully extracted, go to step 10

The junk image information. Present visual information to users. If the display is successful, go to step 11.

10) Choose whether to continue using the system according to actual needs. If you continue, go to step 6, otherwise, go to step.

11) Exit the system

VI. CONCLUSIOJ

Through this experiment, we hope to have higher accuracy and faster speed for garbage classification technology. We have fully used image processing, machine learning and other technologies on the system. Based on single-label image classification, we studied multi-label image garbage classification, which effectively improved the accuracy and speed of recognition. Although the recognition efficiency of the stage proposed in the test set is very low, it is still improving. At the same time, for some garbage with unobvious characteristics, their recognition rate is often not so satisfactory. In the course of further research, the accuracy of the recognition requires further data mining and the use of manually annotated data sets for correction and improvement. Through special training to solve the problem that certain types of garbage are difficult to be identified.

REFERENCES

[1] Costa B S, Bernardes A C S, Pereira J V A, et al. Artificial intelligence in automated sorting in trash recycling[C]//Anais do XV Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2018: 198-205.

[2] Olugboja Adedeji,Zenghui Wang. Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network[J]. Procedia Manufacturing,2019,35.

[3] J. Xiong, W. Cui, W. Zhang and X. Zhang, "YOLOv3-Darknet with Adaptive Clustering Anchor Box for Intelligent Dry and Wet Garbage Identification and Classification," 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 2019, pp. 80-84.

[4] C. Zhihong, Z. Hebin, W. Yan, W. Yanbo and L. Binyan, "Multi-task Detection System for Garbage Sorting Base on High-order Fusion of Convolutional Feature Hierarchical Representation," 2018 37th Chinese Control Conference (CCC), Wuhan, 2018, pp. 5426-5430.

[5] Xu Guoyi. Research on Image Retrieval Based on Color Features[D]. Hunan University, 2009.

[6] Chen Zhi. Multi-label scene classification based on convolutional neural network[D]. Shandong: Shandong University, 2015. DOI: 10.7666/d. Y2792471.

[7] Anya Li. Research on multi-label scene image classification based on deep learning[D]. Xi'an Shiyou University, 2019.

[8] Di Jiang.Research on multi-label image classification method based on deep learning[D].Anhui:Hefei University of Technology,2019.

[9] S. Sudha, M. Vidhyalakshmi, K. Pavithra, K. Sangeetha and V. Swaathi, "An automatic classification method for environment: Friendly waste segregation using deep learning," 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, 2016, pp. 65-70.

[10] Chen, Z.-M., Wei, X.-S., Wang, P., & Guo, Y. (2019). Multi-Label Image Recognition With Graph Convolutional Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[11] Chen Mianshu, Yu Lulu, Su Yue, Sang Aijun, Zhao Yan. Multi-label image classification based on convolutional neural network[J].Journal of Jilin University (Engineering Science Edition),2020,50(03):1077-1084.

[12] Shihui.G Research on image multi-label classification algorithm based on deep learning [D]. Beijing Jiaotong University, 2019.