Garbage recognition and classification system based on convolutional neural network VGG16

Wang Hao
Wuhan University of Technology
Wuhan, 430070, China
Corresponding author's e-mail: 1065120476@qq.com

***Abstract*—To study the application of deep learning in the field of environmental protection, the convolutional neural network VGG16 model is used to solve the problem of identification and classification of domestic garbage. This solution first used the OpenCV computer vision library to locate and select the identified objects and preprocessed the images into 224×224 pixel RGB images accepted by the VGG16 network. Then after data enhancement, a VGG16 convolutional neural network based on the TensorFlow framework is built, by using the RELU activation function and adding BN layer to accelerate the model's convergence speed, while ensuring recognition accuracy. This project finally classifies domestic garbage into recyclable garbage, hazardous garbage, kitchen waste and other garbage. After actual tests, the correct classification rate of the garbage classification system based on VGG16 network proposed in this paper is 75.6%, the result meets the needs of daily use.**

Keywords-component; Deep Learning; VGG16 Network; Garbage Recognition

I. INTRODUCTION

At this stage, due to the large population of China, China adds about 1.1 billion tons of domestic garbage each year, and with the improvement of living standards, the rate of garbage generation is still rising. The reason is that China's garbage generation is rapid, and people's awareness of garbage classification is weak. Therefore, according to the severe status of domestic garbage disposal, the "Ministry of Housing and Urban-Rural Development" issued the "Notice on Comprehensively Carrying Out Domestic Waste Classification Work in Cities at Prefecture Levels and Above in China" in 2019: Starting from 2019, Major cities at the prefecture level and above across the country must carry out the work of classifying domestic garbage into designated categories in accordance with requirements. The resource classification and recycling treatment mechanism of urban domestic garbage is an effective solution to solve the environmental damage caused by urban domestic garbage in China.

Due to the large increase in computer operating speed, the efficiency of computer processing of images has greatly improved. Deep learning models with CNN (Convolutional Neural Network) as the core have begun to play a pivotal role in the field of image recognition and classification. Convolutional neural network combines feature extraction and image recognition in the traditional image recognition process, and uses the original image as input data, it avoids the complex image processing process of traditional recognition methods, and it can extract the most essential features of data that cannot be extracted manually from a limited set of data samples, so the recognition result of deep learning is more reliable, and its ability to recognize images is also stronger. Because of this, this paper proposes a deep learning-based VGG16 network to identify and classify domestic waste.

II. SCHEME DESIGN PROCESS

My solution is that I first use web crawling and collection to build a dataset with a total of 2650 images in 4 categories and label each image. According to the ratio of the training sample set to the number of test sets is 8: 2, that is, the training sample set has 8300 pictures and the test set has 2300 pictures for classification. The data is enhanced in image preprocessing to increase the number of data sets, which can reduce the possibility of overfitting the training model due to lack of data, and improve the accuracy of test results. Before the data is sent to the network for training, in order to obtain an RGB image of 225× 225 pixels to meet the format requirements of the input data of the VGG16 network, I use the OpenCV library to target the identified object, then binarize the image, and then Frame and crop the image. After modifying the VGG16 model's activation function and adding the BN layer to improve the convergence rate, I input the test sample set for model training, and use the test set for model testing until the test correct rate is acceptable, then the model can be applied to the identification and classification of domestic waste to meet the intended purpose.

978-1-7281-8143-1/20/$31.00 ©2020 IEEE
DOI 10.1109/AEMCSE50948.2020.00061

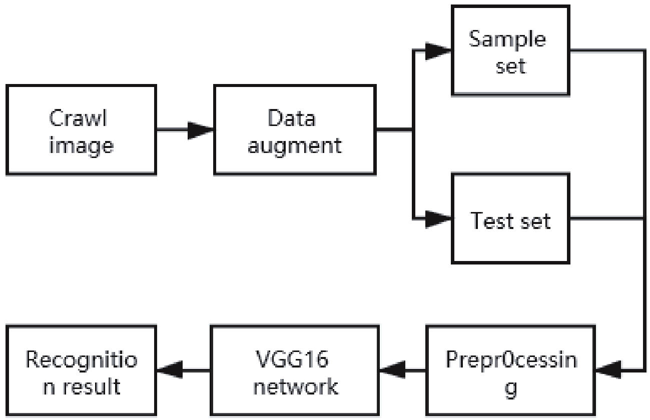

Figure 1 System overall flowchart

III. IMAGE PREPROCESSING

As there is no public garbage image dataset at present, the dataset used for the training model in this paper is built by camera shooting and web crawling. Due to limited manpower and resources, a large number of garbage images cannot be obtained for learning. The complex VGG16 model requires a large number of labeled training data sample sets. Otherwise, the over-fitting problem will occur due to the small amount of input data. In order to solve this problem, this article adopts the method of image data enhancement of the dataset: (1) Change the saturation, contrast, brightness and other parameters of domestic garbage pictures to simulate the difference in light intensity during real-life testing. (2) Reverse the image by 45 degrees, 90 degrees, 135 degrees, and 180 degrees to imitate the difference in the angle of the article when taking pictures

The convolutional neural network VGG16 deep learning model requires that the input image data is an RGB image format of 224×224 pixels, and the pixel size of the image obtained in reality is much larger than the input data requirement of the model. In image preprocessing, the OpenCV computer vision library is used for target recognition and localization of the identified objects, and the ROI sensing area is selected and cropped into a 224×224 pixel image. The specific implementation steps are as follows:

- Use the Cvtcolor function to convert the input RGB three-channel image into a single-channel grayscale image with a pixel value of 0 to 255..
- Use the GaussianBlur function to perform Gaussian blur operation on the grayscale image to obtain the image shown in Figure 2(a). That is, each pixel of the grayscale image is convolved to extract important features of the image to weaken reflective points and noise.
- Use the Threshold function to perform global threshold binarization on the Figure 2(a) after Gaussian blurring, and invert the image in black and white to obtain an image as shown in Figure 2(b) below to facilitate facilitate the contour extraction of the image in the next step.
- Use the FindContours function to perform contour recognition on Figure 2(b) to obtain the contour list and its index. The bounding rectangle of the contour is obtained according to the BoundingRect function, and the image is drawn as shown in Figure 2(c) by adjusting the recognized object to be located in the center of the box.
- Use the scale_image operator to scale the image to an RGB image of 224×224 pixels as shown in Figure 2(d) below.

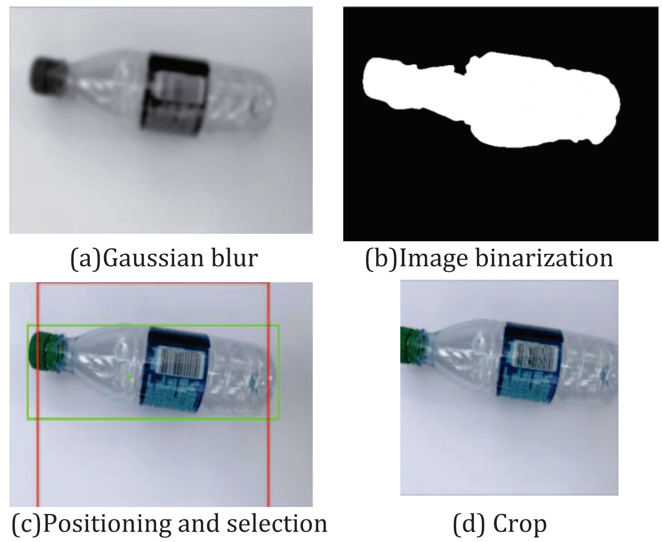

Figure 2 Image preprocessing

IV. CONSTRUCTION OF CONVOLUTIONAL NEURAL NETWORK VGG16

A. Convolutional neural network

Convolutional neural network is a feedforward neural network structure that includes a complex network. The core idea of image processing is weight sharing, local connection, and pooling. It reduces the large number of parameters that need to be set during the training of the neural network model through weight sharing and local receptive fields, and it can learn the most essential characteristics of the data from a large number of image data sets. It has a wide range of applications in image recognition and classification. The classic convolutional neural network structure is mainly composed of five parts: input layer, convolution layer, pooling layer, fully connected layer, and output layer.

B. VGG16 network

The VGG16 network structure is one of the VGG NET networks proposed by Zisserman and Simonyan in 2014. It is a deeper network based on the AlexNet network. It can more accurately express the characteristics of the data set when identifying and classifying images. It performs better on large-scale data sets and complex background recognition tasks. The network structure includes 13 convolutional layers, 3 fully connected Layers, and 5 pool layers. Compared to other networks, the convolution kernel used in the 13 convolutional layers included in VGG16 is a medium-sized 3 × 3 matrix with a moving step of 1. The number of convolution kernels gradually increased from 64 in the first layer to 128 to 256 and then to the last 512. The size of the convolution kernel of the pooling layer is 2 × 2, and the step size is 2. Compared to other networks with a convolution kernel size of 5× 5, it has better performance on the extracted features.

However, the VGG16 network model has the advantages of accurate identification, but also has disadvantages. As we deepen the network structure, the number of model parameters and the complexity of calculations during training have also increased, resulting in long training time and low training efficiency. In this paper, in order to keep the main features of the model extraction without reducing the accuracy of recognition, and at the same time improve the rate of model training and reduce the time it takes to train the model, the following methods are used to improve the VGG16 model.

(1) Replaced activation functions such as Sigmoid and Tahh with rectified linear units (RELU). When Sigmoid or Tahh functions are used as training activation functions, because these activation functions belong to the saturated non-linear function category, the model converges very slowly during training, and the efficiency will also decrease. At the same time, when the Sigmoid function approaches the saturation region in the model training, the value of the function changes gently, leading to the appearance of vanishing gradient. This also easily causes the algorithm to fall into the local optimal solution, which can make it impossible to complete the training of the deep network model and to achieve the expected recognition accuracy.

Compared to non-linear activation functions such as Sigmoid or tahh, the linear correction unit is a non-saturated and non-linear activation function. As can be seen from the function expression, the slope of the part whose independent variable is less than 0 is a certain constant, so it will not cause the vanishing gradient problem during training. In addition, the RELU activation function can more sparsely express extracted image features, it can also reduce the amount of calculation and speed up the model's convergence speed during training. The formula for the RELU activation function is:

$$RELU(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1}$$

(2) A batch normalized BN (Batch Normalization) layer is added after the fully connected layer of the VGG16 network, before the non-linear transformation operation, the activation input value of the deep neural network will gradually move

during the training process as the network depth deepens or changes, that is, for each hidden layer neuron, the input is mapped by a non-linear function, and its value interval is close to saturation. After passing the BN layer, the input distribution is forced to return a standard normal distribution with an average of 0 and a variance of 1. Therefore, the input value after the non-linear transformation function falls into a region more sensitive to input changes to avoid the problem of gradient disappearance.

The BN layer operation in this paper is to calculate the activation value of each neuron in the hidden layer as follows:

$$x^{(k)} = \frac{x^{(k)} - E\left[x^{(k)}\right]}{\sqrt{Var\left[x^{(k)}\right]}} \tag{2}$$

After the BN layer is added, the model training time is reduced and the convergence rate is accelerated. Similar to how Dropout is calculated to kick out neurons, the BN layer can also avoid overfitting during the training process, thereby increasing the accuracy of recognition.

V. Analysis Of Experimental Results

A. test environment

The training test environment of this paper's domestic garbage recognition system is: CPU processor: Intel Cor I5-8250U; RAM is 8 GB; graphics card is GeForce MX150; operating system is Windows 10; compilation tool is Python 3.5 of Pycharm. The tool for processing computer vision is the OpenCV4.1 library, and The construction of the VGG16 network is based on the TensorFlow background.

B. Experimental results

According to the experimental process described in this paper, the collected garbage image data is subjected to data enhancement and image preprocessing. After changing the activation function and increasing the BN layer after the fully connected layer to increase the convergence rate, the sample set is sent to the vgg16 network model for convolutional neural network model training, then the test data set containing 2300 garbage images are used to verify the accuracy of the generated convolutional neural network to identify domestic garbage. In training, by comparison, the learning rate is set to 0.0001, Bitch size is set to 100, the sample set iterative training after about 100×10,000 times. After testing, it was found that the accuracy rate of the final recognition result reached the highest. We can find that the convergence speed of the VGG16 network training process in this paper is accelerated, thereby reducing the training time. After testing the test data set, it is found that the correct rate of the scheme in this paper is 75.6%. The result can meet the needs of the daily waste identification and classification. An example of some test set recognition results is shown in Figure 3, in this figure, “YHLJ” means hazardous waste, “CYLJ” means kitchen waste, “HSLJ” means recyclable garbage, “QTLJ” means other garbage.

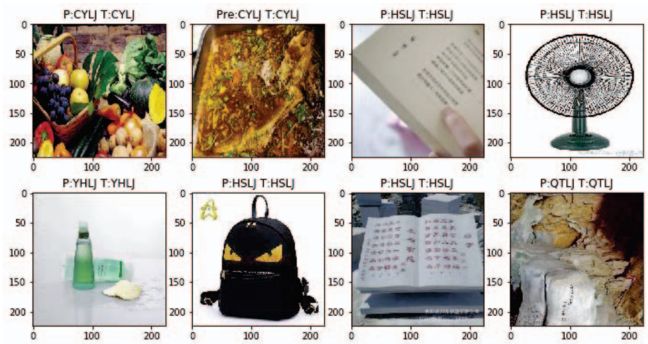

Figure 3 Test set identification example

VI. CONCLUSION

This paper uses the OpenCV computer vision library to perform data enhancement and image preprocessing on the collected garbage images. Using TensorFlow as the background for model training, a VGG-16 convolutional neural network is built, and by using the RELU activation function and adding a BN layer to improve the model convergence rate and increase the recognition accuracy rate. After testing on the test set, the correct rate of the scheme in this paper is 75.6%, the scheme described in this paper can effectively realize the classification of domestic garbage into hazardous garbage, kitchen waste, other garbage, and recyclable garbage, which can meet the needs of practical applications. However, compared with other projects in deep learning implements image recognition and classification, the accuracy of this project still needs to be improved.

REFERENCES

[1] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C] 2015.

[2] Mittal G, Yagnik KB, Garg M, et al 2016 Spot Garbage: smartphone app to detect garbage using deep learning. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing pp 940-945.

[3] Zhao Shifeng, He Xijian.An image binarization method in complex environment based on OpenCV [J] .Electronic Measurement Technology, 2018,41 (06): 55-59.

[4] Wei Shufa, Cheng Zhanglin 2017 Urban scence garbage automatic monitoring based image processing.J.Journal of international technology, 6, pp 40-52.

[5] Zhang Hui, Wang Kunfeng, Wang Feiyue. Application progress and prospect of deep learning in target visual detection [J].Acta Automatica Sinica, 2017, 43 (8): 1289-1305.

[6] Chao Duan. Image classification of fashion-mnist data set based on VGG Network[C]. International Informatization and Engineering Associations.Proceedings of 2019 2nd.

[7] Liao Ruyi, Huang Jianzhong, Yang Danrong. Status and countermeasures of urban domestic garbage classification and treatment in China [J]. Occupational Health and Disease, 2012 (01): 36-40.

[8] Lin Yunsen, Fan Wenqiang, Jiang Jialiang.Research on fruit recognition technology based on deep learning [J] .Application of Optoelectronic Technology, 2019,34(06):45-48+58.

[9] Zou Jinfan, Liu Yuhong, Zhang Rongfen, Ma Zhinan, Ge Zili, Lin Fuchun.Pneumonia detection system based on convolutional neural network [J] .Modern Electronic Technology, 2019,42 (23): 35-39.

[10] He Xiangyu. Research on machine vision recognition technology of plastic bottles for intelligent sorting and recycling [D] .Zhejiang University, 2017.

[11] Zhenzhen Song,Longsheng Fu,Jingzhu Wu,Zhihao Liu,Rui Li,Yongjie Cui. Kiwifruit detection in field images using Faster R-CNN with VGG16[J]. IFAC PapersOnLine,2019,52(30).

[12] Shuai Zhao.Research on the recognition of car models based on deep-learning networks[C].Proceedings of The 2nd Asia-Pacific Computer Science and Application Conference (CSAC 2017).2017:669-674