

Office Garbage Intelligent Classification Based on Inception-v3 Transfer Learning Model

Jie-wen Feng and Xiao-yu Tang*

School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou, China

E-mail: tangxy@scnu.edu.cn

Abstract. With the increase in garbage production, the problem of garbage pollution is becoming more and more serious. Garbage recognition and classification can reduce the environmental burden, but there are still some challenges. Image classification, an image processing method that separates different categories of objects according to different characteristics reflected in the image information. This paper collected and produced an image data set with 2313 photos of different office garbage, and proposed an intelligent classification garbage can to solve the realistic problems. That method based on transfer techniques to retain the excellent feature extraction ability of the Inception-v3 model of TensorFlow, which can recognize objects through the convolutional neural network model. The experimental results showed that the garbage classification effect was obvious and the average accuracy rate reached 95.33%.

1. Introduction

1.1. Research background

Image classification, an image processing method that distinguishes different categories of objects according to different characteristics reflected in the image information, in order to in place of the human visual interpretation. It is the basic problem in the field of computer vision and the foundation of other high-level visual tasks such as behaviour recognition[1], object tracking[2], and image super-resolution[3]. In general, image classification requires describing the whole image through feature learning and then using a classifier to distinguish the object categories in the image.

In the early stage, the primitive method based on the traditional word bag model was mostly used for image recognition, and sparse representation of features was formed by quantifying local features, but it still failed to solve the semantic gap between low-level features and high-level concepts. KNN[4], SVM[5], BP neural network[6] and MLP model have been successfully applied to image recognition. However, because of the mode of full connection between the nodes, the recognition rate is not ideal for high-resolution images.

Nevertheless, image classification method based on deep learning[7] which describes the hierarchical features of images in a supervised or unsupervised way to replace the manual design or learn of image features. An effective method for image recognition based on the convolutional neural network developed in recent years, which can avoid the complex pre-processing of the image. Deep convolutional neural network consists of multiple layers, with feature maps of many two-dimensional planes being formed. Directly using the image pixel information as the input, it retains all the



information of the image that needs to be distinguished to the greatest extent, then combining the low-level features of the image into high-level abstract features through convolution operation, and finally outputs the results of image recognition directly. This end-to-end learning method based on "input-output" way has achieved satisfactory results in many visual tasks.

1.2. Research content and significance

With the increasing production of waste, the problems of garbage pollution in the environment have become increasingly prominent. Therefore, it is necessary to take effective measures to manage waste so as to facilitate the recyclable materials into new resources. Garbage classification is the basic link in the process of the garbage disposal, and it has been an essential work advocated and carried out by governments at all levels. Since 1st July 2019, it has implemented a compulsory garbage classification system in Shanghai, China, through legislation. However, because of the imperfect supporting facilities and people's insufficient awareness and ability of garbage classification, the progress of garbage classification is not ideal. The collected garbage usually is mixed garbage, which seriously limits the operational efficiency of the post-treatment process. Consequently, it is meaningful to design an intelligent trash can that can automatically complete garbage classification to help reduce the cost of garbage classification, which will undoubtedly reduce the pressure of late classification effectively and improve the efficiency of garbage treatment.

Nowadays, increasing researches has been conducted regarding the garbage classification. Chen, et al.[8] proposed a multi-task detection system for garbage classification based on the region proposal network (RPN) and the VGG-16 model. The experimental results of sorting the bottle in the garbage obtain 71% of accuracy. Another research conducted by Salimi et al.[9] designed a trash detection and classification system by using Support Vector Machines (SVM) method to classify the features. However, offline testing of the classification system into organic waste obtains 82.7% of accuracy.

In this paper, in order to deal with garbage classification problems in the actual office scene and provide convenience for managing office environment, we implemented the currently popular transfer learning technology to retrain the Inception-v3 model on a garbage image dataset in the TensorFlow environment[10]. Meanwhile, exploring the convolutional neural network whether it can be limited in the current performance on embedded devices to achieve a higher accuracy of garbage identification in a short training time. With an effective garbage classification system, we provide workable solutions for an ordinary garbage can through a low-cost way to transform into an intelligent automatic classification trash can. The remainder of this paper is organized as follows: part II introduces the overall design and implementation; part III shows transfer learning method and the model of garbage classification process construction; part IV verifies the effectiveness of the model by experiments.

1.3. Research innovation of this paper

- Multiple type of garbage classification. Under traditional thinking, the intelligent classification trash can often rely on various sensors to complete the identification and classification with general accuracy and relatively single identification type.
- Able to transplant a smart garbage classification system through low-cost embedded equipment. Since the transfer learning of convolution neural network model enables higher accuracy in the case of few or even not any labelled data in target domains.
- High accuracy. Proposing a system for garbage classification, which obtains high-order features by using deep convolutional neural network model.

2. Overall design

This system is mainly composed of embedded master control, which can recognize the types of garbage and realize the automatic classification control. Full consideration of cost and volume, we selected the hardware platform is based on a common embedded device Raspberry Pi 3B+ motherboard as the experimental device. Due to the limit of the embedded equipment performance and training the convolution of the neural network needs more computer memories and resources, building

and training the convolutional neural network by learning of feature extraction to acquire precise recognition in PC. When the garbage is put into the system, Raspberry Pi starts the camera to acquire the garbage image, then runs the model of the convolutional neural network to obtain the garbage category information, which has been transplanted to the embedded device. Finally, parameter optimization of Raspberry Pi and controls the garbage fall into the corresponding bucket according to the classification of the garbage category. Figure 1 shows the design of intelligent classification garbage.

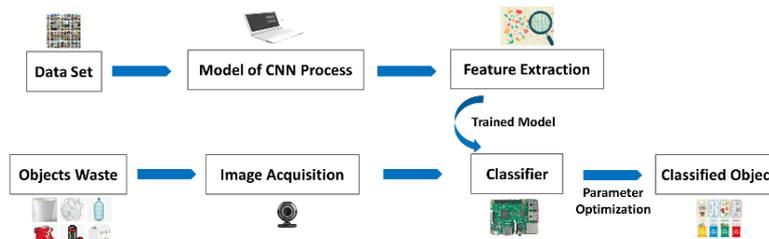


Figure 1. The design of intelligent classification garbage can.

3. Methods and networks

3.1. Main methods

Transfer learning[11] solves cross-domain learning problems and being used in discrimination tasks through extracting favourable information from data in a relevant field. More importantly, transfer learning could greatly improve the performance of learning by solving the situation when the training data is identical to the data distribution or limited label, the network cannot be guaranteed to be sufficient enough to avoid the over-fitting of data. In recent years, transfer learning techniques have been mainly well applied to various branches of pattern recognition, such as object recognition, fine-grained recognition, domain adaptability, and image classification.

In 2014, Donahue J. et al. [12] proposed that using an auxiliary large-scale object database to train convolutional neural network that can learn features to achieve sufficient representation and generalization capabilities compared with using simple linear classifiers. According to that, the main idea of this paper is extracting the features through taking the input layer and convolution-pooling layer of the model trained from the source data. Besides, retaining the parameters of all convolutional layers in the trained convolutional neural network model and only replacing the last fully connected layer. Finally, adjust the model of the output layer to adapt to the task of office garbage identification and classification. Based on the parameter-based transfer, this transfer techniques retained the excellent feature extraction ability of the pre-training model, so that the recognition accuracy and generalization ability can be improved compared with based on sophisticated multi-kernel learning with traditional hand-engineered features. The main diagram of implementing the transfer learning technology to retrain the Inception-v3 model as shown in figure 2.

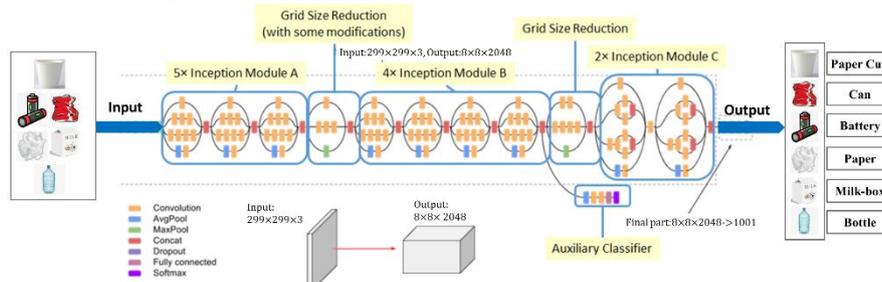


Figure 2. The main diagram of the Inception-v3 model.

3.2. Main network

Convolutional neural network is a feedforward neural network, including input layers, convolution layers, pooling layers, fully connected layers, classifier layers and output layers. With the development of technology, the structure of CNN has gradually transformed. Inception-v3 is a pre-training model on the TensorFlow, one of the convolutional neural networks. This is a rethinking of the initial structure of Inception-v1, Inception-v2 in the field of computer vision. Although inception networks are 42 layers deep, its computational cost is rarely about 2.5 higher than GoogLeNet and its efficiency is still much higher than VGGNet[13].

3.2.1. Convolution layers. The intention of convolution operation is to extract different input features, and reduce the number of parameters through the way of using the convolution kernel. This convolution layers design 3×3 and 1×1 convolution kernel with step sizes of 1. Besides, the small pixel area is called local sensor field, and the weight of the area is the convolution kernel. After the convolution operation, it combines the image with the offset value to form the feature graph. The transformation process is shown in the formula (1).

$$X_j^l = \int (\sum_{i \in M_j} X_i^{l-1} \cdot K_{ij}^l + b_j^l) \quad (1)$$

3.2.2. Pooling layers. The pooling layer achieves the purpose of preserving the effective key information of the image data by effectively reducing the characteristic dimension of the output layer and the number of parameters of the fully connected layer. In addition, the pooling layer can speed up calculations and prevent overfitting. Our pooling layers are mainly designed in 3×3 and 8×8 sizes, and it retains only the maximum value of the specified size filter each time, so as to achieve the effect of feature reduction and compression data. The process is shown in formula (2).

$$X_j^l = \text{down}(X_j^{l-1}) \quad (2)$$

3.2.3. Fully connected layers. Multiple inception module groups through multiple 1×1 convolution kernel and multi-branch convolution aggregation greatly reduce the parameters and layers of the network and reduce the operation amount of the convolutional neural network. Defining our data set as the input layer of the network and use the weight sharing of the Inception-v3 model, which can vastly reduce the number of parameters in the convolution kernel. By redefining in the full connection layer and connecting all the layer variation through formula (3), it is successfully integrated into a one-dimensional vector.

$$\begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} = \left(\begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} \right) \quad (3)$$

3.2.4. Classifier layers. Selecting the softmax classifier as the output classifier, which the calculation way through formula (4). Its function is outputting multiple neurons and mapping certain item categories into 0 to 1 range, so as to output the probability distribution and realize multiple classification process.

$$\text{Softmax}(y_i) = y_i' = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (4)$$

3.2.5. Cross-entropy loss. Entropy in information theory represents the measure of uncertainty of random variables. Cross-entropy is to determine the distance between the expected output and the actual output probability. The smaller the value of cross-entropy represents the closer the two probabilities between the actual and the expected distributions are. The cross-entropy loss can help to

adjust the weight parameter by calculating the error between the output of the softmax layer and the label vector for the given sample category. The process is shown in formula (5), where probability distribution p and q separately represent expected output and actual output, and the value of cross-entropy is $H(p,q)$.

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (5)$$

3.2.6. Activation function. Because introducing non-linear function as the excitation function, the ability of expression in deep neural network is more powerful. The activation function currently includes sigmoid, tanh, ELU, Relu. In this experiment, selecting the most commonly used Relu function as the activation function, which not only the convergence speed is much faster than sigmoid and tanh, but also perform well on the non-linear. Its expression is shown in formula (6), where x is the input of activation functions, $f(x)$ is the activation function.

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (6)$$

The outline of the Inception-v3 network module architecture is shown in figure 3.

Type	Patch size/stride or remarks	Input size
conv	3*3/2	299*299*3
conv	3*3/1	149*149*32
conv padded	3*3/1	147*147*32
pool	3*3/2	147*147*64
conv	3*3/1	73*73*64
conv	3*3/2	71*71*80
conv	3*3/1	35*35*192
3+Inception		17*17*768
5+Inception		8*35*1280
2+Inception		8*8*1280
pool	8*8	8*8*2048
linear	logits	1*1*2048
Softmax	classifier	1*1*1000

Figure 3. The outline of the Inception-v3 network architecture.

4. Experiments and analysis

4.1. Data set introduction

This system selects six kinds of office common garbage in daily life: can, bottle, milk-box, paper cup, paper and battery. In order to obtain the data set which is read and divided into three data sets: training, verification and testing, we obtained them through two ways: on the one hand, collected 1713 images which met the requirements from the Internet, on the other hand, used the raspberry PI wide-angle camera to take 600 pictures of waste items from different backgrounds and angles. Therefore, the original data set consisted of 2313 images, including 412 batteries, 453 bottles, 466 paper cups, 457 cans, 458 milk-boxes and 477 papers. Then, we need to label the data and store different kinds of images in different label folders in image pre-processing step. Finally, the data set was randomly divided into the training set and the test set according to the ratio of 4:1 for neural network training.

4.2. Training and testing of convolutional neural network

The scale of training set is set to 100 times per training stage, each stage of training output a result. Through testing each output set of average recognition accuracy, we can find that as the increasing training times, the number of training accuracy increases gradually and finally converges to a value. Finally, increasing the number of training times, the validation accuracy can be maintained at around 99% indicates the neural network training has been finished. The relationship between epochs and validation accuracy rate of recognition as shown in figure 4.

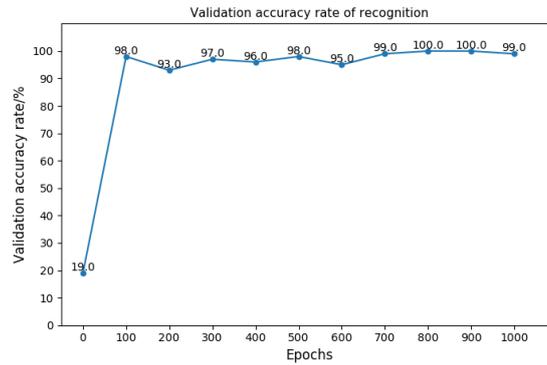


Figure 4. The relationship between epochs and validation accuracy rate of recognition.

4.3. Transplantation of convolutional neural network

Through the above steps, the convolution neural network with good recognition performance on PC is obtained. In order to obtain a good transplantation effect on the embedded device, it is necessary to ensure that the software environment in raspberry PI is consistent with PC, so as to ensure the error free operation of the neural network and obtain a better transplantation effect. Therefore, we set up the Linux system environment and TensorFlow software environment on raspberry PI. After the trained convolutional neural network model was transplanted into raspberry pie, then used the above test set to test the transplantation effect. The experimental results showed that the garbage classification effect was obvious and the average accuracy rate reached 95.33%. Table 1. shows the accuracy rate of recognition was obtained from each type. Figure 5. shows the average rate of recognition accuracy running on raspberry PI 3B+.

Table 1. The average accuracy rate of recognition accuracy running on raspberry PI.

Type	Graphics/numb	Average accuracy rate of recognition/%
Can	100	91.00
Bottle	100	95.00
Milk-box	100	92.00
Paper cup	100	95.00
Paper	100	100.00
Battery	100	99.00



Figure 5. The average rate of recognition accuracy running on raspberry PI 3B+.

4.4. Parameter optimization of convolutional neural network

Compared with PC, the embedded device raspberry PI has scarce memory resources and weak computing power. Therefore, it is necessary to adjust and optimize some parameters of the raspberry PI, so as to improve the accuracy of garbage type classification and processing speed.

4.4.1. Image resolution. However, after the convolution neural network is transplanted into raspberry PI, the images resolution acquired are different. For the trained neural network size, the optimal resolution size is different. Therefore, in order to choose the appropriate image resolution for the network, the identification accuracy and identification time were selected as dependent variables to explore the influence of different image resolutions on the used network. Modifying the multiple test sets to uniform resolution in order to keep the same variable in this round of experiment. From the experimental results, 640×480 is selected as the resolution of image pre-processing, which both the processing time and accuracy are reach a relatively ideal state. The specific experimental results are shown in the following Table 2.

Table 2. Image resolution and processing time and accuracy data table.

Image resolution	Single photo processing time/ms	Accuracy rate of recognition/%
360*240	2453	89.00
400*300	2833	90.00
640*480	2864	95.00
800*600	3491	93.00
1024*768	3532	92.00

4.4.2. Discrimination thresholds. What the classifier outputs is the probability distribution of the categories in the input layer, consequently, each category will have a certain recognition probability value. In order to strictly distinguish the recognition types and reduce the recognition error rate, it is necessary to set the discrimination threshold. When the recognition probability of the input item exceeds the threshold, it can be judged as the corresponding item, so as to improve the accuracy. Therefore, by changing the threshold several times and using the above test set for testing, the average recognition accuracy of the six types of garbage with the threshold change curve is shown as figure 6. According to the experimental results, when the discriminant threshold is set as 0.830, the classification accuracy is the highest. So far, we have obtained a convolutional neural network model that can run on low-cost embedded devices with high recognition accuracy.

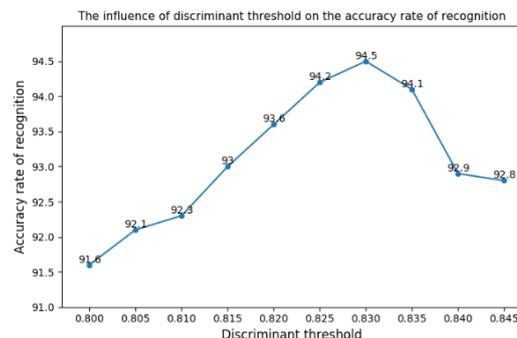


Figure 6. The influence of discriminant threshold on the accuracy of recognition.

5. Discussion and Conclusion

This paper proposed a method of transplant a smart garbage classification system through low-cost embedded equipment. Specifically, the method of garbage image classification based on Inception-v3 convolutional neural network and transfer learning techniques can efficaciously improve the accuracy.

Moreover, this method can provide an effective computer-aided detection as well and have high accuracy of recognition when image data is insufficiency. When transplanted trained model by the embedded device raspberry PI to realize an intelligent automatic classification trash can, the accuracy is well maintained.

Nevertheless, the situation of negative transfer may occur if the model of transfer learning is inappropriate, which will affect the effect of task learning in the target domain such as increasing the training time or decreasing the accuracy. Additionally, the embedded device raspberry PI has scarce memory resources and weak in computing power. Therefore, in the further research, the direction is improving the convolutional neural network model for garbage image classification to reach the higher accuracy of identification and selecting better embedded master control device to obtain faster processing speed.

Acknowledgments

In the process of experimental design and verification, we obtained the comments on programme implementation by Deng Xiang-en, Wu Bicheng, Zhang Zichong. Team members Zhao Shuhui and Qiu Zhixun also offered selfless help. We are grateful to them for their kindness.

References

- [1] Hu, Yaocong, Mingqi Lu, and Xiaobo Lu. "Driving behaviour recognition from still images by using multi-stream fusion CNN." *Machine Vision and Applications* 30.5 (2019): 851-865.
- [2] Feng, Weitao, et al. "Multi-object tracking with multiple cues and switcher-aware classification." *arXiv preprint arXiv:1901.06129* (2019).
- [3] Na, Bokyoung, and Geoffrey Fox. "Object Classification by a Super-resolution Method and a Convolutional Neural Networks." *International Journal of Data Mining Science* 1.1 (2019): 16-23.
- [4] Huang, Kunshan, et al. "Spectral-spatial hyperspectral image classification based on KNN." *Sensing and Imaging* 17.1 (2016): 1.
- [5] Foody, Giles M., and Ajay Mathur. "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM." *Remote Sensing of Environment* 103.2 (2006): 179-189.
- [6] Kai, Song, et al. "A research of maize disease image recognition of corn based on BP networks." 2011 third international conference on measuring technology and mechatronics automation. Vol. 1. IEEE, 2011.
- [7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [8] Zhihong, Chen, et al. "Multi-task Detection System for Garbage Sorting Base on High-order Fusion of Convolutional Feature Hierarchical Representation." 2018 37th Chinese Control Conference (CCC). IEEE, 2018.
- [9] Salimi, Irfan, Bima Sena Bayu Dewantara, and Iwan Kurnianto Wibowo. "Visual-based trash detection and classification system for smart trash bin robot." 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC). IEEE, 2018.
- [10] Martin Abadi, Ashish Agarwal, et al, TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR* abs/1603.04467, 2016.
- [11] Shao, Ling, Fan Zhu, and Xuelong Li. "Transfer learning for visual categorization: A survey." *IEEE transactions on neural networks and learning systems* 26.5 (2014): 1019-1034.
- [12] Donahue, Jeff, et al. "Decaf: A deep convolutional activation feature for generic visual recognition." *International conference on machine learning*. 2014.
- [13] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.