GARBAGE CLASSIFICATION BASED ON CNN

HOU An	**WANG Qijun**	**XU Wenjie**	**YE Junqi**	**ZHOU Wen**
20616544	20636271	20648406	20659728	20617964

ABSTRACT

Convolutional neural network is a hot topic today. It can be applied in many aspects, and get a good performance. In this report, we apply convolutional neural networks in garbage classification, which actually has practical significance.

1 INTRODUCTION

1.1 PROJECT BACKGROUND

Today, garbage classification has become a hot topic in society. Sorting and placing artificial waste is the first step in waste treatment, However, domestic garbage processing plants basically use the manual pipeline sorting method for garbage sorting at present, which has the disadvantages of harsh working environment, high labor intensity, and low sorting efficiency. In the face of massive garbage, manual sorting can only sort out a very limited part of recyclable garbage and hazardous garbage. Most of the garbage can only be landfilled, which brings great waste of resources and environmental pollution risks.

With the application and development of deep learning technology in the visual field, we have seen the possibility of using AI to automatically sort waste, taking pictures of garbage through a camera, and detecting the type of garbage in the picture, so that the machine can automatically sort waste. Greatly improve the efficiency of garbage sorting.

1.2 PROJECT DESCRIPTION

This project aims to do garbage classification. We will take 40 different types of garbage images as our input data, and return the corresponding types of garbage through our model. And we predefined `recyclables, kitchen waste, hazardous waste` and `other waste` as four categories. Therefore, this project is a classic image classification. For picture problems, the best model in machine learning is CNN Talo Muhammed (2019). And due to time and hardware issues, we will use transfer learning in this project at the same time, using the CNG model of the trained VGG16 as our CNN layers, focusing on training our final FC layers. In terms of the main tool in this project, we will use TensorFlow.

1.3 INDIVIDUAL CONTRIBUTION

The individual contribution of the project is shown as Table 1.

Table 1: Contribution

Name	Contribution
HOU An	Coding, Model training&evaluation, Image Transform
WANG Qijun	Collect data, Model building, Image preprocessing
XU Wenjie	Coding, Model building, Image preprocessing
YE Junqi	Coding, Model training&evaluation, Image Transform
ZHOU Wen	Video Making, Data Analysis, Model building, Image preprocessing

2 Related Work

2.1 Convolutional Neural Nework(CNN)

Convolutional neural network (CNN) is a kind of feed-forward neural network which has excellent performance for large image processing. Its artificial neurons can respond to a part of the surrounding cells in the coverage area. A convolutional neural network consists of one or more convolutional layers and a fully connected layer at the top corresponding to a classic neural network. It also includes association weights and a pooling layer. This structure enables the convolutional neural network to take advantage of the two-dimensional structure of the input data.

Compared with other deep learningSonia Phene (2019) structures, convolutional neural networks can give better results in terms of image and speech recognition. This model can also be trained using a back-propagation algorithm. Compared with other deep and feedforward neural networks, convolutional neural networks need to consider fewer parameters, making it an attractive deep learning structure. The architecture is refer to Figure 1. CNN has two major advantages. First of all, the

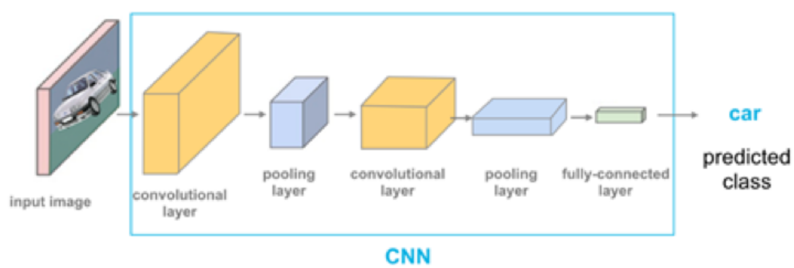


Figure 1: CNN Architecture

local features in the image are extracted through the filtering of the convolution kernel, which is similar to the feature extraction of human vision. Therefore, CNN can effectively retain the characteristics of the images. Secondly, CNN can effectively reduce the dimensionality of pictures with a large amount of data into a small amount of data via subsampling which can avoid overfitting.

2.2 VGG

VGG Xu Xuemiao (2019) was proposed by the Visual Geometry Group of Oxford. This network is a related work at ILSVRC 2014. The main work is to prove that increasing the depth of the network can affect the final performance of the network to a certain extent. VGG has two structures, namely VGG16 and VGG19. There is no essential difference between the two, but the network depth is different.

There are three main advantages of VGG. First, the structure of VGGNet is very simple. The entire network uses the same size convolution kernel size (3x3) and the maximum pooling size (2x2). Second, a combination of several small filter (3x3) convolution layers is better than a large filter (5x5 or 7x7) convolution layer. Third, performance can be improved by continuously deepening the network structure. The main disadvantages of VGG are that it consumes more computing resources and uses more parameters, resulting in more memory consumption. Most of these parameters come from the first fully connected layer.

2.3 VGG16

VGG16 (Figure 2) is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous models submitted to ILSVRC-2014.It makes the improvement over AlexNet Wang Shui-Hua (2019) by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 33 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU.

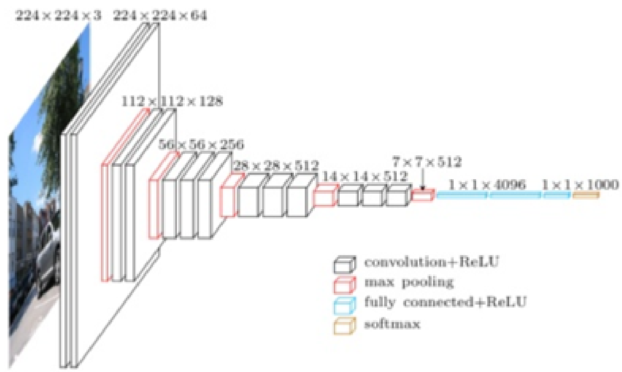


Figure 2: VGG 16 Architecture

2.4 Transfer Learning

Transfer learning Song (2017) is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although formal ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a reinforcement learning agent.

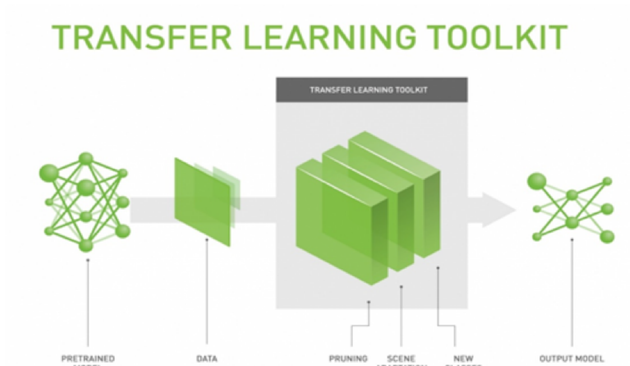


Figure 3

3 Experiments and Results

3.1 Datasets

There are about 15000 garbage images in the dataset. Every image has a related text file to describe its label. We divide the dataset to training set, validation set and test set. Their proportion is 81%, 9%, 10%.

In original datasets, there are 40 classes (class 0 to class 39) of images. Image data categories distribution is refer to Figure 4. According to the graph, we can see that the number of images as class 3 is much smaller than others. The number of images as class 11 and class 21 is much larger than others, etc.

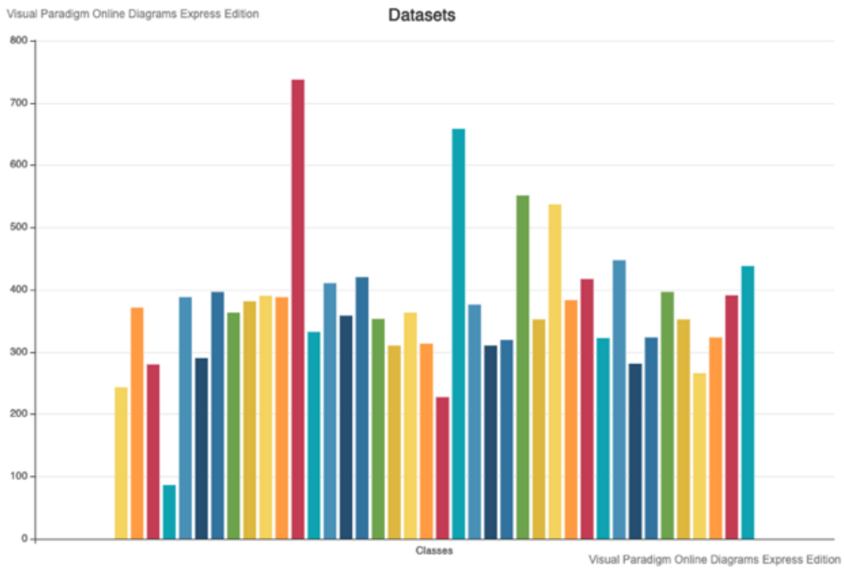


Figure 4: Dataset distribution

3.2 Data Preprocessing

Before building an effective neural network model, careful consideration of the input procession is significant. After getting the data, we firstly did the data analysis, such as statistical data sample distribution, size distribution, picture morphology, etc. Based on the analysis, we removed some images that have strong noise and reclassified some misclassified images in order to achieve a good train model.

Also, there are certain differences in picture aspect ratios. As we know, for the image data, the most common parameters are the number of image height, image width, channels and the number of levels per pixel. Therefore, in order to achieve a uniform format of input, simple data transformation operations are performed on the image, including image filling, scaling, cropping, horizontal flipping, and Gaussian noise. In our data set, we are considering two ways to resize the image. The first way is to resize the image without padding. In this case, we use the Bilinear Interpolation Wang Xiuhua (2019) algorithm. The Bilinear Interpolation is performed with linear interpolation first in one direction and again in the other direction. As a whole, the interpolation is not linear but rather quadratic in the sample location. In this case, the image is scaled with the larger edge as the reference and the smaller edge is broaden to build a square image. The Figure 5 shows that we change the image to a square image of 128*128*3.

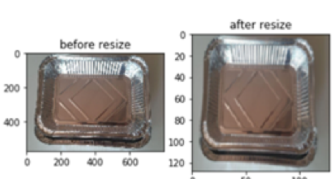


Figure 5

The second way we try is to resize the data that the original image is scaled proportionally with the largest edge as the reference, and the shortcomings are padding with black. As the Figure 6 shows, we resize the image to the scale of 128*128*3.

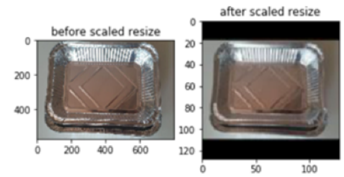


Figure 6

After resizing the image, some image enhancement operations are used to outstand the features of the image. We mainly consider the algorithm called the Adaptive histogram equalization (AHE) Zhiming (2006). The Figure 7 shows the comparison of the ordinary histogram equalization and the Adaptive histogram equalization (AHE). Generally, unlike ordinary histogram equalization algorithms, the AHE algorithm calculates the local histogram of an image and then redistributes the brightness to change the contrast of the image. Therefore, this algorithm is more suitable for improving the local contrast of images and obtaining more image details.

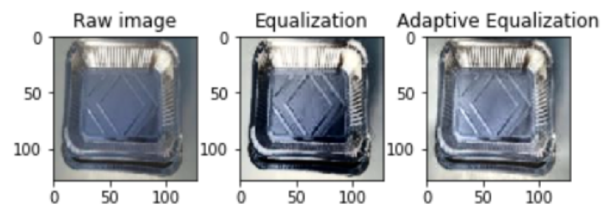


Figure 7

3.3 Network Structure

After data procession, we built the network model. However due to the huge number of parameters of VGG16, as well as the limit of the hardware issues, we can not include VGG16 in our model through the training time. Therefore, we firstly input all resized photos into the VGG model and output a feature map with (4,4,512) as the inputs of our afterward fully-connected layers, as Figure 8.

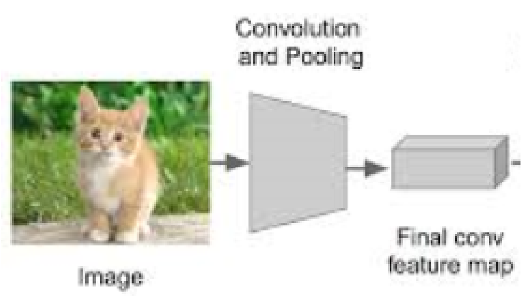


Figure 8

Then, We define three fully-connected layers and mainly train these three layers(Figure 9). Also among the fully-connected layers, we add batch normalization layers to modify the distribution of input of each layer and improve our results. Finally, we use SoftMax loss as our loss, and use the Adam algorithm to optimize the parameters.

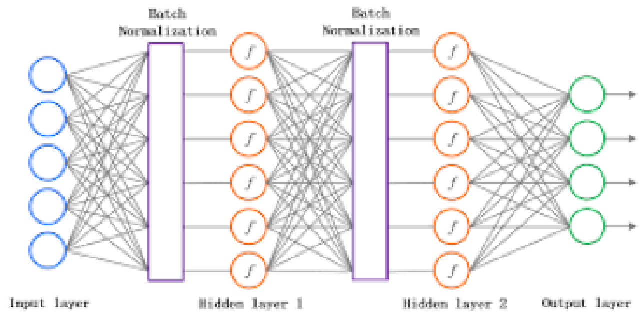
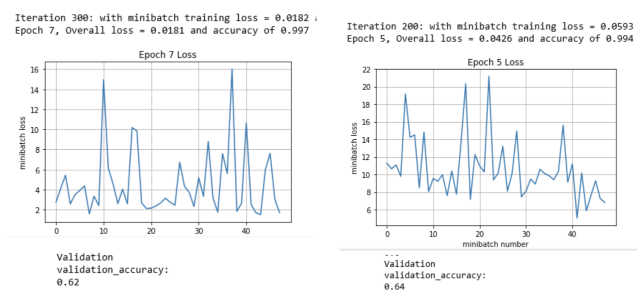


Figure 9

3.4 Training and Results

For the complexity of our model, we firstly define 2048 units in the first fully-connected layer Oncology (2019), 1024 units in the second layer and 40 units in the last one.

During the whole training, we set epoch and batch size to different values and use validation set to verify our model. Figure 10 is part of our result:

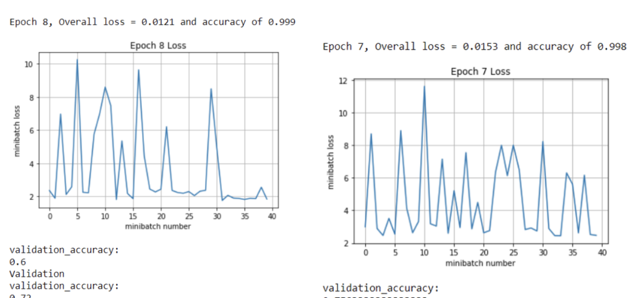


(a) (b)

Figure 10: Result

From the above results, we can find that the accuracy of training set nearly closes to 1, but that of validation set is quite low. Definitely, our fully-connect layers have too many parameters and lead to overfitting problems.

Therefore, we modify the number of units in each layer. We change 2048 units and 1024 units to 1024 units and 512 units in the first two layers and try different values in epoch and batch size variables. Figure 11 is part of our result:



(a) (b)

Figure 11: Result

From the above results, we can know that although our model still has overfiting problems, our validation accuracy increases a lot and the highest one reaches to 76%.

After trying different values in layer units, epoch and batch size variables, we realize that maybe our bottleneck is our training images and the complexity of VGG16.

4 ANALYSIS AND FUTURE WORK

4.1 ANALYSIS

According to section 3.4, we output the misclassified images and their wrong predicted labels, to find the reason why they are wrong and how to approve the performance of our model.

Figure 12 is one example of misclassification (the following figure), we randomly choose 100 images in validation datasets, and plot their real labels (in blue) and wrong prediction labels (in red).

After the whole training process, we have statistics misclassified images.

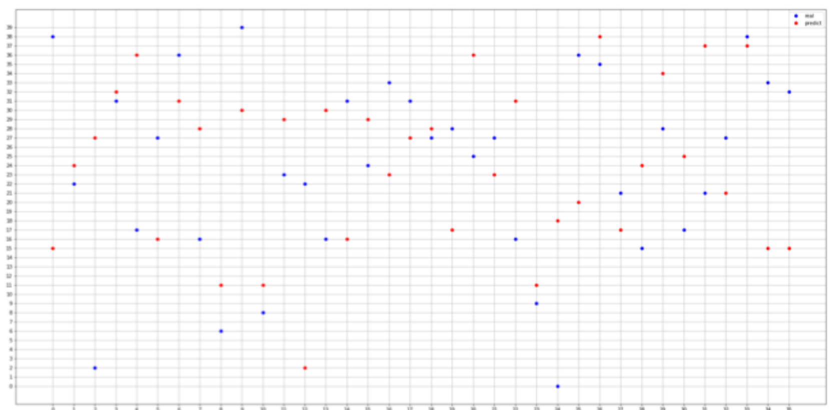


Figure 12: Misclassification samples

4.1.1 MISCLASSIFICATION CASE ONE

One of the most frequent wrong prediction happened in class 0 and class 18. Examples are refer to Figure 13. Class 0 is disposable food box. Most images of class 0 are cubic boxes in various colors, but there are some else images which are round. Class 18 is plastic bowl. Images of class 18 are round bowls. The similarities between the two classes of images' features in shape, might be the reason of the misclassification.

(a) Class 0 (b) Class 18

Figure 13

4.1.2 MISCLASSIFICATION CASE TWO

The other wrong prediction happened in class 27, class 31, class32, and class 36. These four classes are glass bottles, spice bottles, wine bottles, drink bottles. Here are 4 samples(Figure 14). Actually,

(a) Class 27 (b) Class 31 (c) Class 32 (d) Class 36

Figure 14

we can see that shapes of objects in above four images have high similarities. That's why images of these four classes are misclassified.

4.2 FUTURE WORK

According to analysis in section 4.1, the problem of low accuracy maybe since the feature extraction is not appropriate enough. In order to solve this problem, we plan to try different feature extraction models as our future work, to improve our performance. For the solution, we have the following ideas.

- Don't froze all layers of vgg16. We can unfreeze top one or two layers of vgg16 model, train the weights to fit our image better. Then the extracted feature could be more appropriate.
- Try other pretrained models. For example, ResNeXt-50 model. The ResNeXt architecture is an extension of the deep residual network which replaces the standard residual block with one that leverages a "split-transform-merge" strategy. It was discovered that using grouped convolutions led to a degree of specialization among groups where separate groups focused on different characteristics of the input image.

5 CONCLUSION

We have four predefined categories of garbage refer to section 1.2. So after obtaining a best result through training the model and tweaking hyperparameters, we map the 40 different classes labels to four predefined categories labels, and then calculate the accuracy. Finally, the accuracy of garbage classification for four categories can reach a maximum of 88%.

```
37]: nt("validation accuracys:", (map_real == map_predict).sum()/real_label.shape[0]
validation accuracys: 0.8888888888888888
```

(a)

```
print("validation accuracys:", (map_real == map_predict).sum()/real_label.shape
validation accuracys: 0.8828828828828829
```

(b)

Figure 15: Final accuracy

References

Oncology. Oncology - breast cancer; new breast cancer findings from beijing university of posts and telecommunications reported (breast cancer classification based on fully-connected layer first convolutional neural networks. *Computers, Networks Communications*, 2019.

Yang Song. Disease prediction based on transfer learning in individual healthcare. icycsee steering committee.abstracts of the third international conference of pioneering computer scientists, engineers and educators,icpcsee 2017 parti. *ICYCSEE Steering Committee*, 3, 2017.

Naama Hammel Yun Liu Sonia Phene, R. Carter Dunn. Deep learning and glaucoma specialists. *Ophthalmology*, 126(12), 2019.

Baloglu Ulas Baran Aydin Galip Acharya U Rajendra Talo Muhammed, Yildirim Ozal. Convolutional neural networks for multi-class brain disease detection using mri images. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 78, 2019.

Chen Xianqing Wang Shui-Hua, Xie Shipeng. Alcoholism identification based on an alexnet transfer learning mode. *Frontiers in psychiatry*, 10, 2019.

Zhou Wei Qin Xiaoyun Guo Hanming Wang Xiuhua, Jia Xinyue. Correction for color artifacts using the rgb intersection and the weighted bilinear interpolation. *Applied optics*, 58(29), 2019.

Miao Peiqi Qu Wei Xu Xuemiao, Xie Minshan. Perceptual-aware sketch simplification based on integrated vgg layers. *IEEE transactions on visualization and computer graphics*, 2019.

WANG Zhiming. A fast implementation of adaptive histogram equalization. *The Chinese Institute of Electronics(CIE):IEEE BEIJING SECTION*, 4, 2006.