Garbage Classification Based On Deep Residual Weakly Supervised Learning Model

Zhijie Yang[1], Hongbin Huang[1]

[1]Department of Electronic Engineering, College of Information Science and Technology, Jinan University, Guangzhou, 510632, China

Abstract

The realization of garbage classification has become a hot topic in the society, but today's garbage processing plants use the manual pipeline sorting method for waste sorting. This kind of work method has harsh working environment, high labor intensity and low sorting efficiency, etc. Moreover, for the treatment of large amounts of garbage, manual sorting can only sort out a very limited part, and the vast majority of the remaining garbage can only be landfilled, which undoubtedly brings great waste of resources and environmental pollution risks. With the application and development of deep learning technology in the field of computer vision, it is possible to use AI technology to automatically sort waste: using cameras to take pictures of waste and then detecting the type of waste in the pictures, so that the machine can automatically sort waste. This can greatly save huge labor costs and improve waste sorting efficiency. This paper is based on deep residual weakly supervised learning ResNext series networks to classify garbage images, researches and explores AI technology for garbage classification, and contributes to the whole society in garbage classification.

Keywords - *AI, Deep Learning, Garbage Classification, Manual sorting, ResNeXt*

I. INTRODUCTION

In daily life, each of us produces a lot of garbage and throws out a lot of garbage. It can be said that we are making garbage every day, and the amount of garbage produced is particularly huge. In some areas with good garbage management, most of the garbage will be harmlessly treated, such as sanitary landfill, incineration, and compost, etc. But in more places, the way to deal with garbage is just simple stacking or landfilling, which will seriously cause the spread of odor and contaminate the soil and groundwater. However, the cost of harmless treatment of garbage is also very high. According to different treatment methods, the cost of processing one ton of garbage ranges from about one hundred to several hundred yuan. We consume a large amount of resources for large-scale production and consumption, and thus produce a large amount of waste[1]. In the long run, if we do not adopt an effective waste classification method to deal with it, the consequences will be unimaginable.

Garbage classification is to classify and release garbage at the source, and to put it back into resources through clearing and recycling the classified garbage. Today's garbage classification method is a reform of traditional garbage collection and treatment methods. It is an effective and scientific method of garbage disposal. In the face of increasing garbage production and deteriorating environmental conditions, how to maximize the utilization of garbage resources, reduce the amount of garbage disposal, and improve the surrounding environment through effective garbage classification management are currently the issues of common concern to all countries in the world.

Garbage classification can not only reduce the cost of waste disposal and the consumption of land resources, but also reduce pollution and turn waste into treasure. If the garbage can be sorted and recycled in time[2], it will greatly solve the problem of garbage.

II. APPLICATION AND DEVELOPMENT OF DEEP LEARNING IN IMAGE CLASSIFICATION

As an image processing problem, image classification tasks mainly learn the features of images to determine whether a certain type of object is included in the picture. Traditional image classification algorithms generally use manual features or feature learning methods to learn the image to infer the types of objects in the picture. However, this type of feature extraction method is manually designed, which is not only a tedious process, but also can only represent the low-level information of the image. This is a kind of shallow learning, and cannot fully characterize the specific information of the image. Due to well good at deep network structure, deep learning[3] can better

extract higher-dimensional image features from the data and learn more abstract information of the image, so as to better characterize the image. Deep learning shows its excellent ability on the image classification task, and its feature extraction process is completely automatic, without the need for manual feature description and extraction.

Convolutional neural network(CNN) is the main research hotspot in the field of image classification[4], and it has shown good capabilities in image feature extraction and representation. With the increase of the depth of the CNN model, the image feature information extracted by it is more and more advanced and abstract. Therefore, it can better represent the image subject semantics and has a good image classification effect.

Since the design of AlexNet[4] successfully won the championship of ImageNet image classification in 2012, research results on deeper convolutional neural networks have continuously appeared. In 2014, GoogLeNet[5] designed the Inception module structure from the perspective of designing the network structure to capture the features of different scales. The VGG[6] network model in the same year further proved the importance of the depth of the network in improving the model effect. ResNet[7], a deep residual network in 2015, proposed a method for fitting residual networks to better train deeper networks. Subsequent classification networks such as Google's inception series[8] and 2017 mainstream models DenseNet[9], which won the best paper award, all learned from ResNet's design ideas.

III. INTRODUCTION TO THE USE MODEL

A. ResNet model

Proposed in 2015, ResNet[7] won the championship in the ImageNet image classification competition that year. Thanks to relatively simple and powerful, it can be said to be the most widely used deep learning feature extraction network at present, such as applied to common objects detection, image segmentation, recognition and other tasks.

ResNet has a variety of different layer structures, including two main types, respectively, for ResNet-18/34 and ResNet50/101/152, as the following two structures:

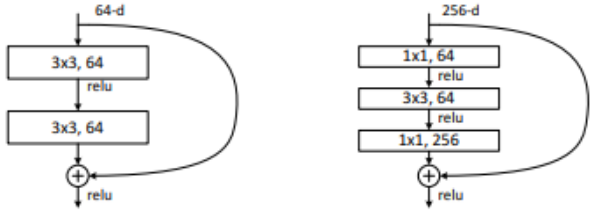


figure.1 Two ResNet structures[7]. By drawing on the 1x1 convolution layer structure of the Inception network, the design of the right part can not only save calculation time and shorten the training time of the entire model, but also can significantly reduce the number of calculations and parameters. Contrast to the left, the number of structural parameters can be reduced by almost 17 times.

B. ResNext model

ResNeXt[10] is an upgraded version of ResNet. In contrast, the ResNeXt network structure is concise, easy to understand, and powerful enough. There are fewer hyperparameters that need to be manually adjusted, and the results of ResNeXt are better than ResNet with the same number of parameters.

Xie et al. [10] analyzed the standard paradigm of neural networks in the ResNeXt paper and concluded that they conform to the mode of splitting-transforming-aggregating. For example, in a simplest fully connected network,

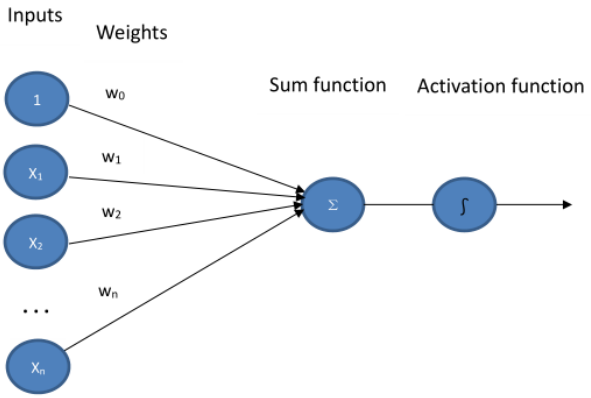


figure.2 Fully connected network. A general unit of a neural network can be represented by the following formula:

$$\sum_{i=0}^{D} W_i X_i \qquad (1)$$

Here $X = [X1, X2, \ldots, Xn]$ is a n-dimensional input vector, W is a weight vector, and W_i is the weight coefficient of the i-th dimension of the input vector. Here splitting first, the input vector X is divided into multiple single dimensions X_i ; then transforming, adding weights W_i to each input vector dimension; finally aggregating to “equation 1” . The author made the following substitutions in the ResNeXt paper,

$$f(x)=\sum_{i=1}^{C} T_i(x) \quad (2)$$

, according to the identity mapping relationship of the ResNet network[7], the structure with residual can be expressed by the following formula:

$$f(x)=x+\sum_{i=1}^{C} T_i(x) \quad (3)$$

Here T(x) represents any mapping transformation function, C refers to cardinality, which is a new concept of constructing a CNN network introduced by the author in addition to the depth and width of the network. On the basis of not deepening the depth and width of the network, increasing the cardinality can not only reduce hyperparameters, but also effectively improve the accuracy of the model; C also representing the number of branches in a block, which can be used to measure model complexity.

In the comparison of the basic block units constituting ResNeXt and ResNet, we can see:

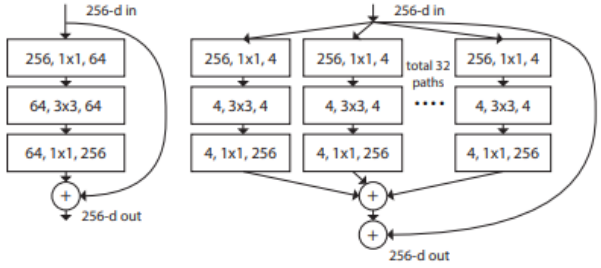


figure.3 Basic block comparison[10]. The residual connection that is directly connected from the input to the output is x in "equation 3" above. The remaining part on the right side of "equation 3" is the transformation of the branch structure, and then two transformations are aggregated. This mode is consistent with the splitting-transforming-aggregating mentioned above, which has stronger feature extraction capabilities, higher abstraction levels, and is more effective than ResNet networks. For example, in the comparison between ResNet-50 and ResNeXt-50, although their network structure have the same number of parameters, the latter has stronger feature extraction capabilities and higher accuracy.

C. ResNext-101 series network

In June 2019, Facebook's He Kaiming team open sourced ResNext WSL, the strongest ResNext pre-training model, where WSL said weakly supervised learning[11]. It first used 940 million pictures on Instagram to do weakly supervised pre-training (these Instagram dataset images have not gone through Special labeling), and then fine-tuning with ImageNet, the model has a total of more than 800 million parameters. The following are the names of ResNext-101 WSL 4 models, parameter size, floating point of operations and accuracy:

Table 1. ResNext-101 WSL[11]

Model	Params	FLOPs	Top1 Acc	Top5 Acc
32 x 8d	88M	16B	82.2%	96.4%
32 x 16d	193M	36B	84.2%	97.2%
32 x 32d	466M	87B	85.1%	97.5%
32 x 48d	829M	153B	85.4%	97.6%

Among them, Top-1 Accuracy is the accuracy rate of the category with the highest predicted probability, and Top-5 Accuracy is the accuracy rate of the correct category with the top five predicted probability; The experimental results show that the Top1 accuracy of ResNext-101_32x48d_WSL achieves the latest accuracy of 85.4%, refreshing the record of ImageNet.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset

This dataset has a total of 14,802 pictures, and each picture contains common trash in life, such as fruit peels, disposable snack boxes, cigarette butts, etc. The garbage classification standard of this paper adopts the Shenzhen in China garbage classification standard, which are four major types of garbage, recyclables, kitchen waste, hazardous waste, and other garbage, and subdivided into 43 sub-categories. For example, leftovers and fruit peels are kitchen waste, old clothes and cans are recyclables, dry batteries and expired drugs are hazardous wastes, and cigarette butts and towels are other wastes, etc.

An example of the result of the final output of the picture is:

figure.4 Other garbage / cigarette butts.

In this paper, the original data is divided into a training set and a test set with a ratio of 9:1, where the number of training sets is 13321 pictures and the test set is 1481 pictures;

B. Experimental

In this experiment, ResNext-101_32x8d_WSL and ResNext-101_32x16d_WSL were used as baseline model, and several sets of comparative experiments were designed. Due to the large amount of model parameters and insufficient server computing power, ResNext-101_32x32d_WSL and the more powerful ResNext-101_32x48d_WSL can only be reduced the parameter size to adapt to the model, such as reduce the image input size, reduce train and test batch size, etc. However, such results lead to longer training time, slower convergence, and the final effect is not very well, not elaborate here.

In this experiment, the GPU uses 4 TeslaP100, each graphics card memory is 16G; the operating system environment is Ubuntu 18.04.4 LTS 5.3.0-40-generic GNU/Linux; the CPU is Intel(R) Xeon(R) Gold 6130 CPU, and use Pycharm as the development environment for this experiment.

Experiment one is a comparison of two baseline model experiments of ResNext-101_32x8d_WSL and ResNext-101_32x16d_WSL. The results table are as follows:

Table3. Baseline 32x8d and 32x16d

Model	Train time	Top1	Top5
baseline_32x8d	1h16m30s	93.77%	99.45%
baseline_32x16d	2h2m59s	94.05%	99.52%

The graph is as follows:

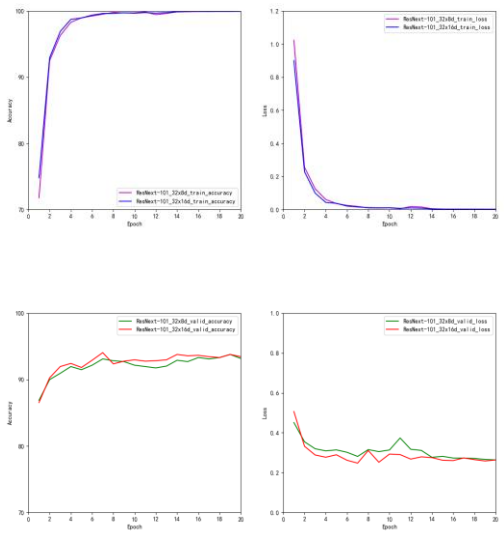


figure.5 Baseline 32x8d and 32x16d.

The ResNext-101_32x16d model has more parameters and is more complex than the ResNext-101_32x8d model and the final validation accuracy is higher and the validation loss is lower, but the corresponding training time is also longer.

In the above process and the following experiments, the training round is set to 20 epochs, because the model has almost converged when it is trained to 20 epochs. In 30, 40, and even higher epochs experiments, the model's effect is not obvious improved and increasing the training epoch blindly will only increase the time cost and consume server resources.

Experiments two, three, and four are based on the baseline pretraining models and have been optimized as follows:

(1) Data augmentation[12], do horizontal flip, vertical flip, etc. to increase the number of samples and improve the overall effect of the model;

(2) Smooth labeling[13] to prevent incorrect labeling and sample imbalance issues;

(3) Dropout[14] is added at the last fully connected layer of the model to reduce the risk of overfitting.

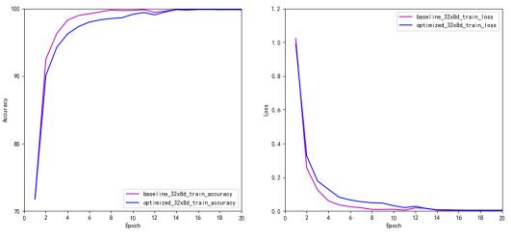
Furthermore, more attempts have been made in the selection of other hyperparameters, such as the loss function, using CrossEntropyLoss and FocalLoss[15]; the optimizer refers to SGD[16], NAdam, adabound, adam, RAdam[17] and RMSProp; the initial learning rate is 0.001, and the learning rate strategy chooses ReduceLROnPlateau method.

Among them, Experiment two is a comparison between ResNext-101_32x8d baseline and ResNext-101_32x8d after optimization, as shown below:

Table4. Baseline 32x8d and optimized 32x8d.

Model	Train time	Top1	Top5
baseline_32x8d	1h16m30s	93.77%	99.45%
optimized_32x8d	1h3m8s	93.98%	99.59%

The graph is as follows:



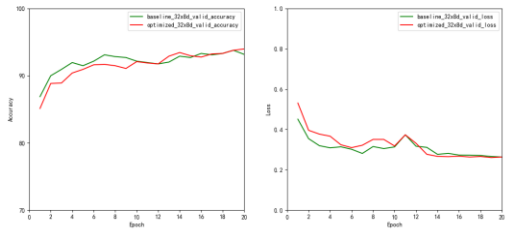


figure.6 Baseline 32x8d and optimized 32x8d.

From the experimental results, it can be seen that the training time of the optimized ResNext-101_32x8d model is shortened, the accuracy of validation is improved more obviously, the accuracy of top1 and top5 is higher, and the loss value is also significantly reduced.

Experiment three is a comparison between ResNext-101_32x16d baseline and ResNext-101_32x16d after optimization, and the experimental results are as follows:

Table5. Baseline 32x16d and optimized 32x16d.

Model	Train time	Top1	Top5
baseline_32x16d	2h2m59s	94.05%	99.52%
optimized_32x16d	2h2m44s	94.25%	99.66%

The graph is as follows:

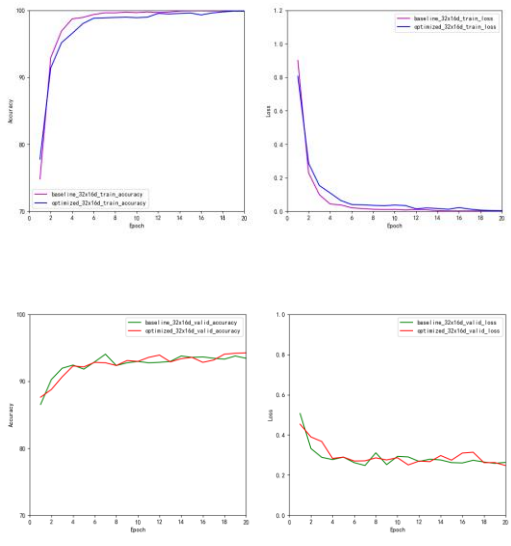


figure.7 Baseline 32x16d and optimized 32x16d.

It can be seen from the graph that the optimized ResNext-101_32x16d model can achieve higher accuracy and lower loss, and the overall trend of the model is better than the baseline model.

Experiment four is the comparison between optimized ResNext-101_32x8d and ResNext-101_32x16d, as shown below:

Table6. Optimized 32x8d and 32x16d.

Model	Train time	Top1	Top5
optimized_32x8d	1h3m8s	93.98%	99.59%
optimized_32x16d	2h2m44s	94.25%	99.66%

The graph is as follows:

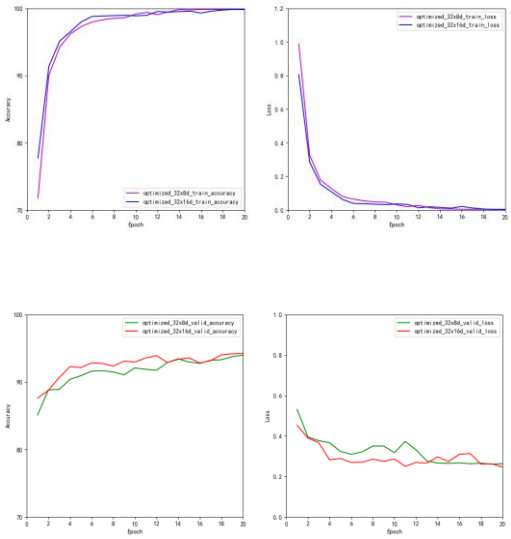


figure.8 Optimized 32x8d and 32x16d.

Similar to the conclusion of the first experiment, the overall effect of ResNext-101_32x16d is still better than ResNext-101_32x8d after optimization, which can achieve higher accuracy and lower loss. However, due to the relatively complicated model, ResNext-101_32x16d training time is still longer than ResNext-101_32x8d.

The above experimental results suggests that the optimized model has a higher accuracy for the classification task of garbage images.

V. CONCLUSION

Garbage classification has become a hot topic for everyone now. In this era where we produce a large amount of garbage every day, classifying garbage, integrating resources, and protecting the environment have become an indispensable part of our daily life. It is of great significance to use the most advanced AI technology to help people sort waste, which can realize the practical application of the latest technology.

Additionally, make people aware that the development of new technologies has promoted and improved the quality of people's daily, which will undoubtedly spur Continued innovation of technology and practical application innovation.

REFERENCES

[1] Hoornweg D, Bhada-Tata P. "*What a waste: a global review of solid waste management[J]"*. 2012.

[2] Johansson N, Corvellec H. "*Waste policies gone soft: An analysis of European and Swedish waste prevention plans[J]"*. Waste Management, 2018, 77: 322-332.

[3] LeCun Y, Bengio Y, Hinton G. "*Deep learning[J]. nature"*, 2015, 521(7553): 436-444.

[4] Krizhevsky A, Sutskever I, Hinton G E. "*Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems"*. 2012: 1097-1105.

[5] Szegedy C, Liu W, Jia Y, et al. "*Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition*". 2015: 1-9.

[6] Simonyan K, Zisserman A. "*Very deep convolutional networks for large-scale image recognition[J]"*. arXiv preprint arXiv:1409.1556, 2014.

[7] He K, Zhang X, Ren S, et al. "*Deep residual learning for image recognition[C]*"//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Szegedy C, Vanhoucke V, Ioffe S, et al. "*Rethinking the inception architecture for computer vision[C]"//*Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826..

[9] Huang G, Liu Z, Van Der Maaten L, et al. "*Densely connected convolutional networks[C]"//*Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708..

[10] Xie S, Girshick R, Dollár P, et al. "*Aggregated residual transformations for deep neural networks[C]//*"Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500..

[11] Mahajan D, Girshick R, Ramanathan V, et al. "*Exploring the limits of weakly supervised pretraining[C]"//*Proceedings of the European Conference on Computer Vision (ECCV). 2018: 181-196.

[12] Cubuk E D, Zoph B, Mane D, et al. "*Autoaugment: Learning augmentation strategies from data[C]"//*Proceedings of the IEEE conference on computer vision and pattern recognition. 2019: 113-123.

[13] Müller R, Kornblith S, Hinton G E. "*When does label smoothing help?[C]*"//Advances in Neural Information Processing Systems. 2019: 4696-4705.

[14] Srivastava N, Hinton G, Krizhevsky A, et al. "*Dropout: a simple way to prevent neural networks from overfitting[J]*". The journal of machine learning research, 2014, 15(1): 1929-1958.

[15] Lin T Y, Goyal P, Girshick R, et al. "*Focal loss for dense object detection[C]*"//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[16] Goyal P, Dollár P, Girshick R, et al. "*Accurate, large minibatch sgd: Training imagenet in 1 hour[J]*". arXiv preprint arXiv:1706.02677, 2017.

[17] Liu L, Jiang H, He P, et al. "*On the variance of the adaptive learning rate and beyond[J]"*. arXiv preprint arXiv:1908.03265, 2019.

[18] Dr.V.V.Narendra Kumar, T.Satish Kumar, "*Smarter Artificial Intelligence with Deep Learning*", SSRG International Journal of Computer Science and Engineering, Vol-5,Iss 6,2018